



Développement et application de méthodes bioinformatiques pour la recherche de gènes de sRNA à boîtes H/ACA dans les génomes d'archaea et étude fonctionnelle des ARN et sRNP H/ACA correspondants

Sébastien Muller

► To cite this version:

Sébastien Muller. Développement et application de méthodes bioinformatiques pour la recherche de gènes de sRNA à boîtes H/ACA dans les génomes d'archaea et étude fonctionnelle des ARN et sRNP H/ACA correspondants. Biologie moléculaire. Université Henri Poincaré - Nancy 1, 2007. Français. NNT : 2007NAN10110 . tel-01748244

HAL Id: tel-01748244

<https://hal.univ-lorraine.fr/tel-01748244>

Submitted on 29 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

FACULTE DES SCIENCES & TECHNIQUES

**U.F.R. Sciences et Techniques Biologiques
Ecole Doctorale Biologie, Santé, Environnement**

Thèse

présentée pour l'obtention du titre de

**Docteur de l'Université Henri Poincaré, Nancy-I
en Biologie Moléculaire**

par **Sébastien MULLER**

**Développement et application de méthodes bioinformatiques pour la
recherche de gènes de sRNA à boîtes H/ACA dans les génomes
d'archaea et étude fonctionnelle des ARN et sRNP H/ACA
correspondants**

Soutenance publique le 27 Novembre 2007

Membres du Jury :

Président du Jury :	M. Alain JACQUIER	Directeur de Recherche CNRS, Paris
Rapporteurs :	M. Daniel GAUTHERET	Professeur, Université Paris XI, Orsay
	M. Yves HENRY	Directeur de Recherche CNRS, Toulouse
Examineur :	M. Bruno CHARPENTIER	Professeur, Université Henri Poincaré, Nancy
Directeur de thèse :	Mme Christiane BRANLANT	Directeur de Recherche CNRS, Nancy

Je tiens à remercier Christiane Branlant, pour m'avoir accueilli dans son laboratoire et permis de réaliser ce travail dans les meilleures conditions. Sa rigueur scientifique et son énergie m'ont considérablement stimulé tout au long de cette expérience formatrice. J'ai été particulièrement honoré de la confiance qu'elle a toujours eue en moi et mon travail. Elle a toujours su m'orienter vers les meilleures pistes de recherche et collaborations internes au sein du laboratoire. Elle m'a toujours témoigné son soutien lors des moments difficiles.

Je suis très sensible au grand honneur que me font Messieurs Daniel Gautheret et Yves Henry, d'accepter d'être rapporteurs de ce mémoire. Je remercie aussi sincèrement Messieurs Alain Jacquier et Bruno Charpentier qui ont accepté d'examiner cette thèse.

Je suis particulièrement reconnaissant envers Fabrice Leclerc, Bruno Charpentier et Yuri Motorin qui ont su me communiquer leur passion pour la recherche, leur savoir-faire dans des domaines aussi divers que la bioinformatique, la biochimie des protéines et la biologie moléculaire mais aussi pour leur rigueur. Les conseils qu'ils m'ont prodigués m'ont permis d'avancer efficacement dans mes travaux. J'espère, dans ma vie professionnelle à venir, pouvoir à mon tour atteindre ce niveau et pouvoir le transmettre.

J'exprime ma reconnaissance aux post-docs, thésards ou anciens thésards avec lesquels j'ai pu collaborer au cours de ces travaux : merci tout particulièrement à Isabelle qui m'a appris à manier la technique RT-CMCT, à Emmanuel, à Alan et Jean-Baptiste qui ont pu poursuivre des expériences quand j'ai débuté la rédaction de ce manuscrit. Nos discussions, scientifiques ou non, m'ont beaucoup apporté et j'espère que nos routes se recroiseront.

Merci à tous ceux qui m'ont aidé à rédiger ce manuscrit: principalement mes collaborateurs déjà cités, en particulier Isabelle et Bruno, et mon épouse dont les facilités en langue m'ont été très profitables.

Merci à tous ceux qui ont contribué à la bonne ambiance et au merveilleux environnement dans lequel j'ai eu la chance d'évoluer. Je pense notamment à Tadek, Zdenek, Benoit, Alan, Aileen, Nath MG, Nath R, Isabelle, Arnaud, Mathias, François, Bruno, Athanase, Lionel, François G, Sylvain, Claude, Denise, Valérie, Magali, Karine, Rémi, Hadidja, Jackie, Corinne, Delphine, Xavier M, Christophe G, Christophe C, Séverine B, Manu et les autres que je n'oublie pas.

Je tiens aussi à remercier Benoit, Alex O, Sylvain, Bruno, Alan, Xavier R, Laure, Nico, Christophe et les autres qui m'ont permis de me défouler grâce au Tennis, Badminton, Foot et autres sports.

Remerciements

Je remercie également mes amis exilés : Christophe, Franck, Jean-Thomas, les membres du cercle « Fear » dont tonton Jean-Paul et Ronan, et tous ceux qui se reconnaîtront sous l'appellation « mes amis ». Ils ont contribué à leur façon à cet ouvrage.

Je suis immensément reconnaissant envers mes parents. Ils m'ont toujours soutenu et encouragé en formidables et parfaits parents qu'ils sont. C'est certainement ma mère qui, inconsciemment, m'a guidé vers le monde de la recherche et ses longues études. Je ne sais pas pourquoi, et je ne chercherai pas à y répondre ici. Mon père lui, m'a toujours fait croire que la vie était un jeu et en cela je le remercie de me rappeler qu'il faut aussi en profiter.

Enfin, cette liste ne serait pas complète et juste si j'oubliais ma petite famille qui, tout au long de ma thèse, m'a toujours soutenu et encouragé. Mon épouse, Eline, par sa compréhension, son intelligence, sa patience et son amour a su créer un environnement dans lequel j'ai pu accomplir cette lourde tâche. Elle m'a supporté pendant ces années, et ce même si cela n'était pas facile. J'ai bien conscience des sacrifices consentis et je ne l'oublierai jamais.

C'est avec elle que j'ai réalisé la plus belle « chose » de ma vie : ma fille Mina. Elle est née pendant ce travail de thèse et constitue donc, en quelque sorte, mon travail de thèse. Cependant, cet aspect ne sera pas abordé dans la suite de l'exposé. C'est son incroyable joie de vivre et son amour qui m'ont apporté la confiance qui me manquait au début de cette thèse. Etant donné que cette confiance a été essentielle à la concrétisation de mes convictions scientifiques et, par conséquent, à la réalisation de ces travaux, je dois avouer que je n'aurais jamais réussi à mener aussi bien ma recherche sans elles. Par conséquent, je dédie cet ouvrage à Eline & Mina.

Je remercie également le reste de ma famille et de ma belle-famille. Malgré tous mes efforts, mes grands-parents n'ont toujours pas compris ce que je faisais en thèse, et ce n'est pas grave... Ce qui compte pour eux, et donc ce qu'ils espèrent, est que je commence enfin à travailler ! Ainsi, ce manuscrit ne serait pas le fruit d'un long et difficile travail mais plutôt le résultat d'une passion qui m'a pris quelques unes de mes soirées. Après tout, c'est peut-être la meilleure manière de considérer ces travaux...

Abréviations & Anglicismes

A	: adénine
aa	: acide aminé
ADN	: acide désoxyribonucléique
ADNc	: acide désoxyribonucléique complémentaire
ARN	: acide ribonucléique
ARNm	: acide ribonucléique messenger
ARNnc	: ARN non codant
ARN pol	: ARN polymérase
ARNr	: acide ribonucléique ribosomique
ARNt	: acide ribonucléique de transfert
ATP	: adénosine 5' triphosphate
C	: cytosine
°C	: degré celsius
Ci	: curie
CB	: « Cajal Bodies », corps de Cajal
CMCT	: 1-cyclohexyl-3-(2-morpholinoéthyl) carbodiimide méthyl-p-toluène
cpm	: coups par minute
cps	: coups par seconde
C-ter	: extrémité C-terminale
Da, KDa	: dalton, kilodalton
db	: double brin
ddNTP	: didésoxyribonucléotide triphosphate
DNase	: désoxyribonucléase
dNTP	: désoxyribonucléotide triphosphate
DO	: densité optique
DTT	: dithiothréitol
EDTA	: éthylène diamine tétraacétate
5FU	: 5-fluoro-uridine
g/ mg/ µg/ ng	: gramme/ milligramme/ microgramme/ nanogramme
G	: guanine
h	: heure
I	: inosine
IPTG	: isopropyl β-D-thiogalactoside
kan ^r	: gène de résistance à la kanamycine
Kb	: kilobase
l/ ml/ µl	: litre/ millilitre/ microlitre
LB	: Luria Bertani
M/ mM/ µM/ nM	: molaire/ millimolaire/ micromolaire/ nanomolaire
min	: minute
NMP	: ribonucléotide monophosphate
NTP	: ribonucléotide triphosphate
nt	: nucléotide
N-ter	: extrémité N-terminale
pb	: paire de base
PEG	: polyéthylène glycol
pGalS	: promoteur galactose dépendant
pré-ARNm	: précurseur de l'ARN messenger
pré-ARNr	: précurseur de l'ARN ribosomique
pré-ARNt	: précurseur de l'ARN de transfert

qsp	: quantité suffisante pour
R	: purine
RMN	: résonance magnétique nucléaire
RNase	: ribonucléase
RNasin	: inhibiteur des ribonucléases
rNTP	: ribonucléotide triphosphate
RT	: reverse transcriptase
RX	: rayons X
s	: seconde
S	: Svedberg
sb	: simple brin
SDS	: sodium dodécylsulfate
structure 2D	: structure secondaire
structure 3D	: structure tertiaire
T	: thymine
TLC	: chromatographie sur couche mince
U (UE)	: unité enzymatique
Ψ	: pseudouridine
Y	: pyrimidine

Anglicisme: certains termes ne sont pas ou sont rarement traduits :

ETS	: external transcribed sequence
GST	: glutathione S-transferase
LUCA	: Last Universal Common Ancestor ou dernier ancêtre commun universel
LSU	: large subunit
ORF	: phase ouverte de lecture (Open Reading Frame)
PCR	: amplification en chaîne par une polymérase (polymerase chain reaction)
SSU	: small subunit
SELEX	: Systematic Evolution of Ligands by EXponential enrichment
gRNA	: ARN guide
sRNA	: petit ARN (Small RNA)
sRNP	: particule ribonucléoprotéique contenant un ou des petits ARN (Small ribonucleoprotein particule)
scaRNA	: petit ARN contenu dans les Cajal Bodies (Small Cajal Bodies RNA)
scaRNP	: particule ribonucléoprotéique contenant un ou des petits ARN contenus dans les Cajal Bodies (Small Cajal Bodies ribonucleoprotein particule)
snoRNA	: petit ARN nucléolaire (Small nucleolar RNA)
snRNA	: petit ARN nucléaire (Small nuclear RNA)
snoRNP	: particule ribonucléoprotéique contenant un ou des petits ARN nucléolaires (Small nucleolar ribonucleoprotein particule)
snRNP	: particule ribonucléoprotéique contenant un ou des petits ARN nucléaires (Small nuclear Ribonucleoprotein Particule)
UsnRNA	: « uridine rich small nuclear ribonucleic acid »
UsnRNA	: « uridine rich small nuclear ribonucleoprotein particle »
spliceosome	: complexe nucléaire ribonucléoprotéique dans lequel ont lieu les réactions d'épissage
UTR	: untranslated region

Résumé

La conversion des résidus U en pseudouridine (Ψ) est la modification la plus fréquente dans les ARN. Cette réaction est catalysée par des ARN: Ψ -synthases qui fonctionnent soit seules, soit au sein de particules snoRNP H/ACA composées de 4 protéines et d'un snoRNA. Celui-ci assure la reconnaissance de l'ARN cible par appariement de bases. Les pseudouridylations des ARNr eucaryotes sont catalysées par des snoRNP alors que les modifications des ARNt sont catalysées par des enzymes dépourvues de guide. Aucun snoRNA H/ACA n'est décrit chez les Bactéries.

Trois équivalents des snoRNA H/ACA (sRNA H/ACA) avaient été découverts chez les Archaea lorsque j'ai débuté ma thèse, mais leur contribution aux modifications des ARNr n'avait pas été démontrée. Nous avons développé une approche bioinformatique permettant l'identification exhaustive de gènes de sRNA H/ACA et de leurs cibles. Nous avons ainsi identifié 7 sRNA H/ACA chez *Pyrococcus abyssi* et détecté 14 Ψ dans les ARNr.

Nous avons réussi la reconstitution de particules H/ACA actives *in vitro*, ce qui nous a permis d'étudier leur mode d'assemblage et d'assigner 12 des Ψ détectées aux 7 sRNA.

L'application à d'autres espèces d'archaea de l'approche bioinformatique développée nous a permis d'identifier environ 90 sRNA, nous avons ainsi pu étudier leur évolution. De plus, nous avons expérimentalement démontré pour la première fois qu'un sRNA H/ACA est impliqué dans la modification d'un ARNt.

Enfin, nous avons montré que l'enzyme aCBF5 des sRNP H/ACA peut catalyser la pseudouridylation en position 55 des ARNt sans utilisation de sRNA et nous avons identifié des résidus de aCBF5 impliqués dans les activités guidée et/ou non guidée.

Mots-Clés : archaea, *Pyrococcus abyssi*, sRNA H/ACA, pseudouridine, ARN guide, ARN: Ψ -synthase

Title :

Development and application of computational methods dedicated to the identification of H/ACA box sRNAs within archaeal genomes and functional study of the corresponding RNAs and H/ACA sRNPs.

Summary

The conversion of U into Pseudouridine (Ψ) residue is the most prevalent RNA modification. This reaction is catalysed either by stand-alone RNA: Ψ -synthases or by H/ACA snoRNPs composed of 4 proteins and a snoRNA. In snoRNPs the recognition of the U residue is carried out by base-pair interactions between the snoRNA and the RNA target. SnoRNPs are involved in eukaryal rRNA

modifications whereas stand-alone RNA:Ψ-synthase are involved in tRNA and bacterial rRNA modifications.

Three counterparts of the H/ACA snoRNAs had been discovered in Archaea when I started my PhD work, but their contribution to rRNA modification was not demonstrated. We developed a computational approach to identify the H/ACA sRNA genes and their rRNA targets. Its application to the *Pyrococcus abyssi* genome revealed 7 candidate genes. We verified their expression and experimentally identified 14 Ψ in the *P. abyssi* rRNAs.

We succeeded in the *in vitro* reconstitution of active H/ACA sRNPs and thereby studied the sRNP assembly mechanism, identified important H/ACA sRNA structural features and assigned 12 of the identified Ψ residues to the 7 sRNAs.

The application of the computational approach to the other archaeal genomes revealed 90 sRNA genes. This repertoire shed new light on the function and the evolution of this ncRNA class.

In addition, we experimentally demonstrated for the first time that a few H/ACA sRNAs are involved in tRNA modification.

We also showed that the enzyme of the H/ACA sRNP (aCBF5) is involved in Ψ55 formation in all archaeal tRNAs without the use of an H/ACA sRNA. We compared the guided and non-guided CBF5 activities and identified aminoacids which are important for these activities.

Key Words: archaea, *Pyrococcus abyssi*, H/ACA sRNA, pseudouridine, RNA guide, RNA:Ψ-synthase.

TABLE DES MATIERES

Table des Matières

Remerciements_____	I
Abréviations_____	III
Résumé_____	V
Table des matières_____	VII

✕ Introduction ✕

Avant-propos_____	1

Partie I : Le Domaine Archaea

1. Historique/Généralités _____	3
2. Classification et Evolution des Archaea _____	5
A. Habitats et Formes _____	5
B. Taxonomie _____	8
C. Phénotypes _____	11
D. Phylogénie et Evolution _____	14
3. Caractéristiques des Archaea _____	19
A. Organisation du génome _____	19
B. Réplication, Réparation et Chromatine _____	22
1. Réplication _____	22
2. Chromatine _____	26
3. Réparation de l'ADN _____	28
C. Transcription _____	28
1. La région promotrice _____	28
2. L'élongation et la terminaison de la transcription _____	32
D. Traduction _____	32
E. Membranes _____	36
F. Pathogénicité _____	38
G. Intérêt biotechnologique _____	40
H. Outils génétiques _____	41
I. Virus d'Archaea _____	43

Partie II : La recherche de gènes d'ARN non codant

1. Les ARN non codant	47
A. Généralités	47
1. Variété des ARNnc	47
2. Relecture du génome	47
3. ARNnc et complexité	48
4. Identification biochimique	49
5. Identification bioinformatique	49
B. Classification	50
1. Complexes avec des protéines et motifs ARN	50
2. Variété des modes d'action	50
3. Biogenèse	51
4. Localisation	52
5. Distinction selon leur taille	52
6. Distinction selon leur(s) rôles	52
7. Familles d'ARNnc	52
C. Distribution dans les trois Domaines	53
1. Bactéries	53
1.1. Interactions avec des protéines	54
1.1.1 Modulation de l'activité	54
1.1.2 Modulation du nombre de protéines disponibles	54
1.2. Action par appariement à un ARN	55
1.2.1. Action antisens en cis	55
1.2.2. Action antisens en trans	58
2. Eucaryotes	60
2.1. ARN antisens	60
2.2 miRNA et siRNA	62
2.3 gRNA	66
2.4 lRNA	68
2.5 snRNA	70
2.6 snoRNA et scaRNA	72
2.6.1. snoRNA à boîtes C/D	74
2.6.2 snoRNA à boîtes H/ACA	82
2.6.3 scaRNA	87
2.7 Recours aux guides ARN	88
3. Archaea	90
3.1. sRNA à boîtes C/D	90
3.2. sRNA à boîtes H/ACA	95
3.3. Evolution des sRNA / snoRNA	96
3.4. D'autres ARNnc chez les Archaea ?	99

2. Méthodes de recherche	102
A. Méthodes expérimentales	102
1. Séquençage ARN	102
2. Clonage et construction de banques de cDNA	102
3. Puces à ADN	104
4. SELEX génomique	105
B. Méthodes bioinformatiques	105
1. Identification non spécifique des gènes d'ARNnc d'un génome	106
1.1. Signaux de transcription et localisation des gènes d'ARNnc	107
1.2. Utilisation d'un biais dans la composition de la séquence	107
1.3. Utilisation de signaux thermodynamique	108
1.4. Génomique comparative	109
1.5. Combinaison de différents critères	111
2. Identification des représentants d'une famille de gènes d'ARNnc	113

Partie III : La formation de pseudouridines

1. Les modifications post-transcriptionnelles	115
2. Distribution et Localisation des pseudouridines	120
A. ARN de transfert	120
B. snRNA	121
C. ARN ribosomiques	121
3. Catalyse de la pseudouridylation	128
A. Caractérisation des ARN:Ψ-synthases	128
B. Mécanisme enzymatique	130
C. Structure tridimensionnelle des ARN:Ψ-synthases	133
4. Reconnaissance de la position à modifier	138
A. ARN:Ψ-synthases qui ne requièrent pas de guide ARN	138
1. Mécanisme d'extrusion de bases	138
2. Domaines périphériques et additionnels	140
3. Reconnaissance de la tige-boucle TΨC par TruB	142
4. Région-spécificité de TruA	144
5. Attribution des ARN:Ψ-synthases aux résidus Ψ détectés	146
5.1. Situation dans les ARNt	147
5.2. Situation dans les ARNr	151
B. ARN:Ψ-synthases associées à des guides ARN	155
1. Caractéristiques des 4 protéines des sno/sRNP à boîtes H/ACA	156
2. Interactions entre les constituants des complexes sRNP à boîtes H/ACA	159

2.1. Structure globale d'un sRNA associé aux 4 protéines	159
2.2. Association de L7Ae sur le motif K-turn	159
2.3. Interactions entre aCBF5 et aNOP10	164
2.4. Interactions entre aCBF5 et aGAR1	166
2.5. Interactions de aNOP10 avec à la fois le sRNA et aL7Ae	167
2.6. Interactions entre aCBF5 et le guide sRNA	167
2.7. Structure RMN d'un sRNA associé à l'ARN cible	169
3. Les snoRNP H/ACA eucaryotes	171
4. Attribution des sno/sRNA à boîtes H/ACA aux résidus Ψ détectés	173
4.1. Situation dans les snRNA	173
4.2. Situation dans les ARNt	173
4.3. Situation dans les ARNr	173
5. Rôle des résidus Pseudouridines dans les ARN	174
A. Considérations générales	174
B. Rôle dans les ARNt	174
1. Résidu Ψ en position 13	175
2. Résidu Ψ en position 27	175
3. Résidu Ψ en position 35	176
4. Résidu Ψ en position 55	176
5. Autres résidus Ψ des ARNt	176
C. Rôle dans les ARNr	177

✕ Objectifs de mon travail de thèse	183

✕ Matériel & Méthodes ✕

-Partie Bioinformatique -

1. Matériel	185
1.1. Matériel informatique	185
1.2. Logiciels utilisés	185
1.3. Outils informatiques utilisés	185
1.4. Programmes élaborés	186
1.5. Séquences et annotations des génomes	186
1.6. Autres sites d'intérêt	186
2. Méthodes	187

- Partie Biologie Moléculaire -

1. Matériel	189
1.1. Les souches utilisées	189
1.1.1. Souches de bactéries	189
1.1.2. Souches d'archaea	189
1.2. Les plasmides utilisés	189
1.3. ADN génomique de <i>P. abyssi</i> et <i>S. solfataricus</i>	190
1.4. Milieux de culture, tampons et solutions d'usage courant	190
2. Méthodes spécifiques	191
2.1. Extraction des ARN totaux	191
2.2. Identification de résidus modifiés dans les ARN par des méthodes basées sur l'extension par la transcriptase inverse	192

Chapitre d'ouvrage N°1 :

Identification of modified residues in RNAs by reverse transcription - based methods.

Yuri MOTORIN, Sébastien MULLER, Isabelle BEHM-ANSMANT, Christiane BRANLANT

Methods in Enzymology, 2007 volume 425:355-87. **193**

2.3. Mutagenèse dirigée	227
2.4. Extraits cellulaires	227
2.5. Northern Blot	227
2.6. Transcription in vitro	228
2.7. Analyse de la formation de complexes ARN/protéine	230
2.8. Evaluation de l'activité de pseudouridylation	230
2.9. Purification des protéines	234

✕ Résultats / Discussion ✕

1. Identification de 73 séquences pouvant coder un ou des ARNnc chez les <i>Pyrococcus</i>	235
A. Contexte	235
B. Démarche	235
C. Collection de séquences IRC	236
D. Sélection des fragments IRC présentant des éléments de séquence conservés entre les trois espèces de <i>Pyrococcus</i> .	238
E. Evaluation de l'approche	242
F. Comparaison de la méthode développée avec d'autres approches appliquées au génome de <i>P. abyssi</i>	246
G. Amélioration de la méthode développée	247

2. Identification de 5 sRNA H/ACA parmi les 73 ARNnc candidats _____ 249

- A. Contexte et Démarche _____ 249
- B. Résultats _____ 250

3. Etude de l'interaction des sRNA H/ACA avec les protéines associées et reconstitution de particules actives in vitro _____ 254

- A. Contexte et intérêts _____ 254
- B. Les approches utilisées, les questions posées et les réponses obtenues _____ 255
- C. Résultats _____ 256

Publication N°1 :

Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation

Bruno Charpentier, Sébastien Muller, Christiane Branlant.

Nucleic Acids Research, 2005, Vol. 33, No. 10 p3133–3144

257

- D. Résultats complémentaires _____ 269
 - 1. Mise en évidence d'une interaction directe de aNOP10 avec le sRNA H/ACA Pab21 _____ 269
 - 2. L'affinité de aCBF5 pour le sRNA Pab21 est supérieure à celle pour le sRNA Pab91 _____ 270
 - 3. Etude de l'assemblage des protéines sur des sRNA H/ACA à plusieurs motifs _____ 272
 - 4. Cas particulier du sRNA Pab35 comportant une structure tige-boucle supplémentaire en 5' _____ 274
- E. Discussion complémentaire & Perspectives _____ 275
 - 1. Trois protéines peuvent se fixer directement sur les sRNA H/ACA d'archaea : L7Ae, aCBF5 et aNOP10 _____ 275
 - 2. aNOP10 joue un rôle primordial pour l'activité _____ 276
 - 3. L7Ae et aGAR1 augmentent l'efficacité de la réaction catalytique _____ 276

4. Développement d'une méthode permettant de rechercher des sRNA à boîtes H/ACA dans tous les génomes d'Archaea_____ 279

A. Principe _____ 279

Chapitre d'ouvrage N°2 :

A Dedicated Computational Approach for the Identification of Archaeal H/ACA sRNAs.

Sébastien MULLER, Bruno CHARPENTIER, Christiane BRANLANT and Fabrice LECLERC

Methods in Enzymology, 2007 volume 425:21-53.

281

B. Discussion _____ 315

5. Etude de la contribution des sRNA H/ACA à la pseudouridylation des ARNr de *P. abyssi* _____ 317

Publication N°2 :

Combined *in silico* and experimental analyses of the *P. abyssi* H/ACA sRNAs and their target sites reveal the implication of RNA guided and non guided RNA:Ψ-synthase activities for Ψ residues formation in ribosomal RNAs

Sébastien Muller, Fabrice Leclerc, Isabelle Behm-Ansmant, Jean-Baptiste Fourmann, Fujihiko Matsunaga, Bruno Charpentier and Christiane Branlant.

Préparé pour soumission à **Nucleic Acids Res.**

321

6. Etude comparative de l'action de aCBF5 avec et sans sRNA H/ACA _____ 365

A. Contexte et démarche _____ 365

B. Résultats _____ 366

Publication N°3 :

Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Ψ55-synthase and RNA-guided RNA:Ψ-synthase activities.

Sébastien Muller, Jean-Baptiste Fourmann, Christine Loegler, Bruno Charpentier and Christiane Branlant.

369

C. Résultats complémentaires _____ 385

1. Le complexe aCBF5/aNOP10/aGAR1 est capable de modifier les ARNt de différentes espèces _____ 385

2. Etude de l'efficacité relative d'association des protéines aCBF5 et aNOP10 avec les ARNt ^{Asp} de <i>P. abyssi</i> et de <i>S. cerevisiae</i> et leurs variants	386
3. L'ARN :Ψ-synthase aPus10 peut agir en position 55 et 54 des ARNt	390
D. Discussion complémentaire et Perspectives	392
7. Identification de sRNA H/ACA capables de cibler la position 35 dans les ARNt^{Tyr} de certaines espèces du royaume des crénotes	398
A. Contexte et intérêts	398
B. Résultats	398
Publication N°4 :	
Deficiency of the tRNA^{Tyr}:Ψ35-synthase Pus7 in Archaea of the Sulfolobales order is rescued by the presence of a specific H/ACA sRNA	
Sébastien Muller, Alan Urban, Arnaud Hecker, Fabrice Leclerc, Christiane Branlant and Yuri Motorin.	
	401
8. Identification de gènes potentiels de sRNA à boîtes H/ACA dans l'ensemble des séquences de génomes d'Archaea disponibles	443
A. Contexte et intérêts	443
B. Démarche	443
C. Résultats et Discussion	443
1. Identification, dans les 46 génomes disponibles, d'un total de 162 motifs qui peuvent être regroupés en 35 familles	443
2. Localisation des gènes de sRNA H/ACA dans les génomes d'archaea	451
3. Lien entre sRNA H/ACA et modifications des ARNr d'archaea	454
4. Caractéristiques structurales des motifs H/ACA identifiés	456
5. Caractéristiques des appariements proposés entre les motifs H/ACA identifiés et leurs ARN cibles	459
6. Aspect modulaire et évolution des sRNA H/ACA	460
7. Relation entre nombre de motifs H/ACA et conditions de croissance	466
8. Présence de sRNA H/ACA dans les éléments extra-chromosomiques ?	470
✕ Conclusion	471
✕ Références	475
✕ Annexes	505

AVANT-PROPOS

Avant-Propos

Lorsque j'ai débuté ce travail de thèse, les résultats des campagnes de séquençage de génomes de nombreuses espèces vivantes devenaient disponibles. Ceci encourageait le développement d'approches nouvelles visant à exploiter cette grande quantité d'informations. Ces approches pouvaient notamment s'appuyer sur l'annotation et la comparaison des séquences (génomique), ainsi que sur la recherche d'éléments ou motifs particuliers dans ces séquences. Ainsi, la bioinformatique, basée sur des développements informatiques et l'analyse biologique, a pris une place primordiale dans ce domaine. D'autre part, à ce moment, la découverte des microARN et d'autres ARN régulateurs a montré que la communauté scientifique sous-estimait le rôle des ARN non codant dans l'ensemble des processus cellulaires. Il devenait évident que la compréhension des mécanismes cellulaires nécessitait une compréhension fine de l'ensemble des différentes classes d'ARN non codant. Des études portant sur les ARN non codant de Bactéries ou d'Eucaryotes avaient été initialisées par d'autres équipes. Cependant, un nombre limité de données était disponible pour le troisième domaine du vivant, celui des Archaea. Or, le domaine des Archaea présente un grand intérêt du fait de la ressemblance de plusieurs des processus cellulaires généraux de ces organismes avec ceux des eucaryotes et du caractère apparemment simplifié de ces processus chez ces organismes. De plus, les archaea se développent dans des conditions très variées et parfois extrêmes qui peuvent nécessiter le recours à certaines classes d'ARNnc.

C'est dans ce contexte que nous avons débuté le développement d'approches bioinformatiques dont le but était initialement la recherche de gènes d'ARN non codant (ARNnc) dans les séquences génomiques d'Archaea. Cette recherche nous a conduit à l'identification d'ARNnc candidats, appelés sRNA à boîte H/ACA, qui sont responsables de pseudouridylation dans les ARNr. Nous avons ensuite vérifié expérimentalement l'expression et la fonctionnalité de ces sRNA ainsi que plusieurs hypothèses émises sur leur mode d'action par l'analyse bioinformatique. Les résultats expérimentaux ont été utilisés en retour pour affiner la recherche de gènes de représentants de cette classe d'ARNnc dans toutes les séquences génomiques d'archaea disponibles. Ce travail de thèse a donc comporté différentes phases alternées de travaux bioinformatiques et expérimentaux.

Le chapitre d'introduction comporte trois parties en lien avec le travail de thèse : le domaine des archaea, les ARN non codant et la pseudouridylation des ARN. Nous présenterons dans un premier temps les caractéristiques des archaea ainsi que les relations phylogénétiques entre

espèces d'archaea. La seconde partie de l'introduction fera l'état des connaissances sur les différents ARN non codant identifiés à ce jour et présentera les méthodes expérimentales et bioinformatiques permettant de les rechercher de manière systématique. Enfin, la troisième partie sera dédiée aux différents systèmes permettant l'isomérisation de résidus U en pseudouridines (Ψ), ceci en distinguant les enzymes agissant seules, des enzymes agissant au sein de particules sno/sRNP à boîtes H/ACA. Ce dernier chapitre présentera également les rôles potentiels des pseudouridylations des ARN.

INTRODUCTION

Partie I : Le Domaine Archaea

1. Historique & Généralités

La découverte, dans les années 1970, de bactéries inhabituelles (Kushner, 1964 ; Brock *et al.*, 1972; Langworthy *et al.*, 1972) puis, plus tard, les études phylogénétiques ont conduit les scientifiques à modifier leur vision traditionnelle de la dichotomie procaryotes/eucaryotes proposée dès 1938 par Edouard Chatton (pour revue, Sapp, 2005). Ainsi, l'émergence des études phylogénétiques par comparaison des séquences des ARNr a joué un rôle majeur dans l'établissement, en 1977, d'un arbre phylogénétique universel du vivant divisé en trois branches : eubactéries, archaebactéries et eucaryotes, correspondant aux trois Domaines primaires (Fig. I.1 ; (Woese et Fox, 1977). En effet, Carl Woese a observé que tous les organismes sans noyau (procaryotes) ne partagent pas un ancêtre commun et qu'ils ne font pas tous partie d'un même groupe monophylétique (ou clade). Carl Woese et collaborateurs ont alors renommé les domaines eubactéries et archaebactéries en Bactéries et Archaea (Woese *et al.*, 1990). La longue liste des caractéristiques non-bactériennes mais également non-eucaryotes, i.e. propres aux Archaea, apporte aujourd'hui de nombreux arguments convaincants confortant cette division en trois domaines primaires (pour revue, Woese, 1987). Le fait que les Archaea comportent un grand nombre d'espèces extrêmophiles adaptées à une ou plusieurs conditions extrêmes de température, de pH, et de salinité a longtemps conduit les scientifiques résistants à cette idée à continuer à considérer qu'elles ont fait partie des premiers organismes qui se sont développés dans les conditions extrêmes existant originellement sur terre. L'idée opposée proposant que ces espèces soient issues d'un phylum bactérien ayant évolué de manière particulière est également défendue par certains. Bien que cette classification/nomenclature soit encore parfois rejetée (Cavalier-Smith, 2006), elle est maintenant admise dans la communauté scientifique, et ce même si le terme « archaebactéries » est encore souvent utilisé (pour rétrospective, Woese, 2004).

Il est à noter que la découverte récente du domaine Archaea explique l'existence d'un certain retard dans l'étude de ces organismes, d'autant plus qu'en raison des conditions extrêmes de croissance de beaucoup d'entre eux, la mise au point des conditions de culture en laboratoire a été difficile. Néanmoins, des progrès importants ont été réalisés sur ces organismes au cours des dernières années. Etant donné que les archaea partagent un certain nombre de systèmes avec les eucaryotes, ils représentent des modèles plus simples permettant d'accéder à la compréhension de systèmes eucaryotes, à une meilleure compréhension de l'évolution et à l'obtention de molécules biologiques stables dans des conditions extrêmes. Enfin, la stabilité de certains des composants cellulaires des archaea tels que les ribosomes ont permis d'établir des structures 3D à haute résolution par cristallographie de ces composants (Ban *et al.*, 2000).

L'objectif de ce chapitre est de dresser une liste non exhaustive des caractéristiques des Archaea qui les distinguent ou les rapprochent des Bactéries et des Eucaryotes. Nous débuterons par la classification en taxons et l'évolution des espèces d'archaea en nous appuyant sur les séquences génomiques complètes qui ont été exploitées au cours de mon travail de thèse.

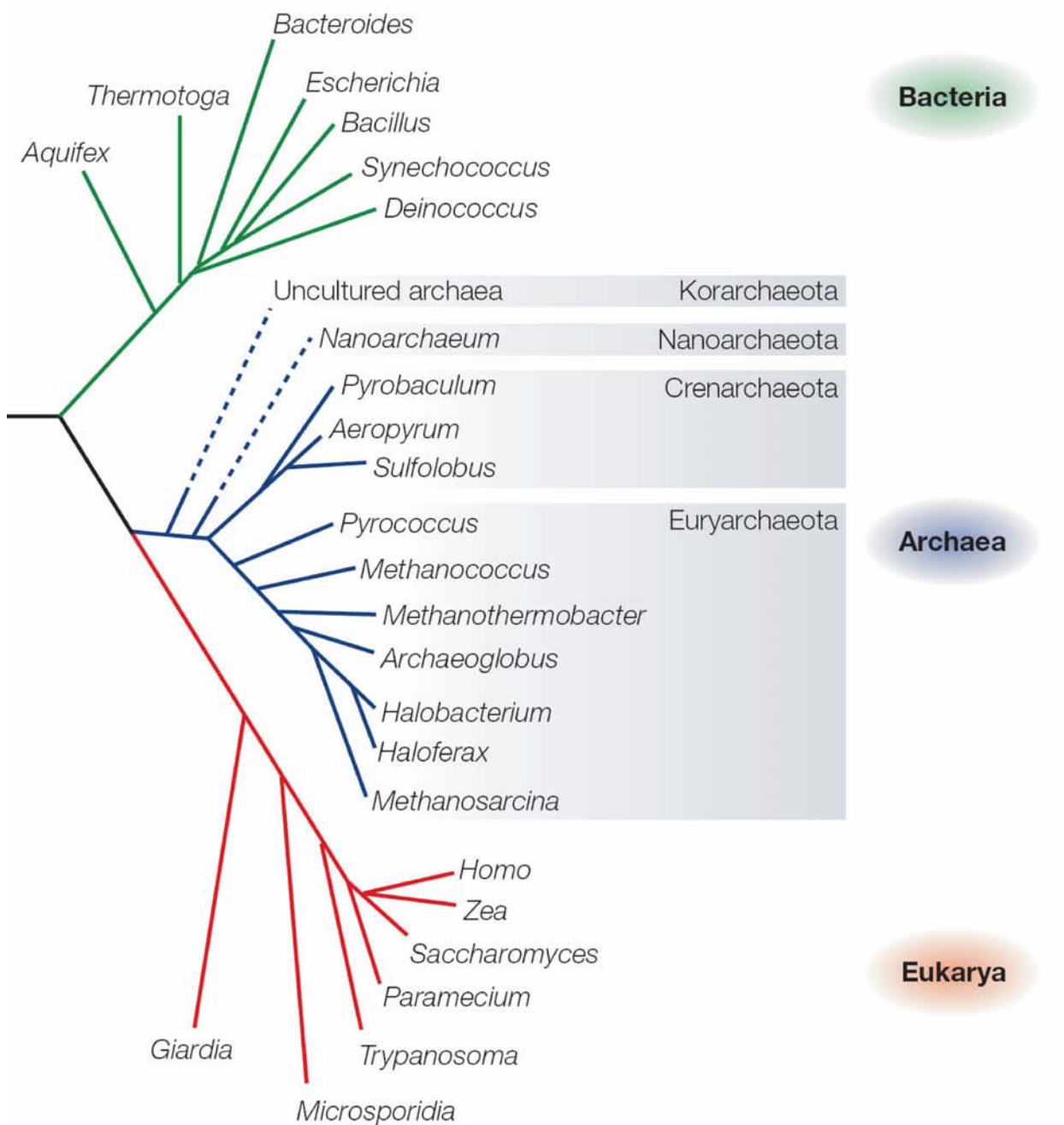


Figure I.1 : Arbre phylogénétique universel. Adapté des travaux de Carl Woese et collaborateurs (Allers et Mevarech, 2005).

Cet arbre est basé sur la comparaison des séquences des ARNr de la petite sous-unité ribosomale. Il décrit une division du vivant en trois domaines primaires : Bactéries, Archaea et Eucaryotes. Le domaine Archaea est divisé en deux domaines principaux, les euryotes (euryarchaeota) et les crénotes (crenarchaeota). L'existence de phyla supplémentaires (Korarchaeota et Nanoarchaeota) aux deux précédents n'est pas encore clairement établie.

2. Classification et Evolution des Archaea

A. Habitats et Formes

Les archaea se caractérisent par leur colonisation d'habitats très variés correspondant pour certains à des conditions extrêmes. Wolfram Zillig et son élève Karl Stetter ont fortement contribué à la découverte de nouvelles espèces d'archaea en parcourant le monde pour prélever des échantillons dont ils ont établi la séquence de l'ARNr 16S après amplification par PCR en utilisant un couple d'oligonucléotides. Ils ont ainsi isolé et caractérisé de nombreuses espèces d'archaea pour lesquelles ils ont développé les conditions de culture en laboratoire. Ils sont à l'origine de la découverte de la plupart des classes d'archaea dont les hyperthermophiles qui se développent à des températures supérieures à 95°C, autre caractéristique qui n'est décrite que chez les Archaea (pour revue, Stetter, 2006). Leurs résultats confirment la grande variété d'espèces d'archaea qui, d'après les travaux de Edward Delong et Norman Pace (DeLong et Pace, 2001), représentent environ 20% de la biomasse sur Terre. Leur abondance et leur distribution dans les sols et océans pourraient fortement contribuer aux cycles énergétiques du carbone, de l'azote et du soufre sur Terre (Hallam *et al.*, 2006 ; Ingalls *et al.*, 2006 ; Reysenbach *et al.*, 2006; Wuchter *et al.*, 2006), d'où l'intérêt des études encore peu nombreuses sur ces organismes.

La grande variabilité des archaea leur permet de s'adapter à une grande diversité de niches écologiques. Des archaea sont en effet présents dans les marais salants (Baliga *et al.*, 2004), les profondeurs des océans (DeLong, 2005), les zones volcaniques (pour revue, Stetter, 2006), les mines de fer (Dopson *et al.*, 2004), les zones arides et acides (Futterer *et al.*, 2004), les réservoirs de pétrole (Magot *et al.*, 2000) ou encore dans les lacs gelés de l'antarctique (DeLong *et al.*, 1994) (Fig. I.2). Les Archaea ne sont pas seulement retrouvés dans des conditions extrêmes puisque des représentants archaea colonisent également le sol et les eaux froides des lacs et des océans (Ochsenreiter *et al.*, 2003), les eaux stagnantes et nos déchets (Chouari *et al.*, 2005) ou encore le tractus digestif des animaux dont celui de l'Homme (Dore *et al.*, 1995 ; Lepp *et al.*, 2004 ; pour revue, Dethlefsen *et al.*, 2006). Les Archaea sont donc ubiquitaires (pour revue complète des différents habitats connus, Chaban *et al.*, 2006). Leur cohabitation avec des Bactéries et des Eucaryotes a autorisé des échanges de matériel génétique par transfert horizontal (pour revue, Jain *et al.*, 2002).

Les Archaea sont des organismes unicellulaires anucléés dont la morphologie est similaire à celle des Bactéries. Il s'agit soit de coques (ou coccus) plus ou moins réguliers ou de bâtonnets (ou bacilles) dont le diamètre ou la longueur est compris entre 0.6 µm et 8 µm. Il existe cependant des morphologies propres aux Archaea telles que, notamment, les coques très irréguliers, parfois en forme de disque, des hyperthermophiles (Stetter, 2006), les « branches » de *Thermoproteus* (Zillig *et al.*, 1981) ou les polygones des halophiles (Bolhuis *et al.*, 2004)

(Fig. I.3). Celles-ci pourraient refléter une rigidité particulière des enveloppes cellulaires ou encore la présence de mécanismes de morphogénèse inconnus à ce jour (Forterre *et al.*, 2002).

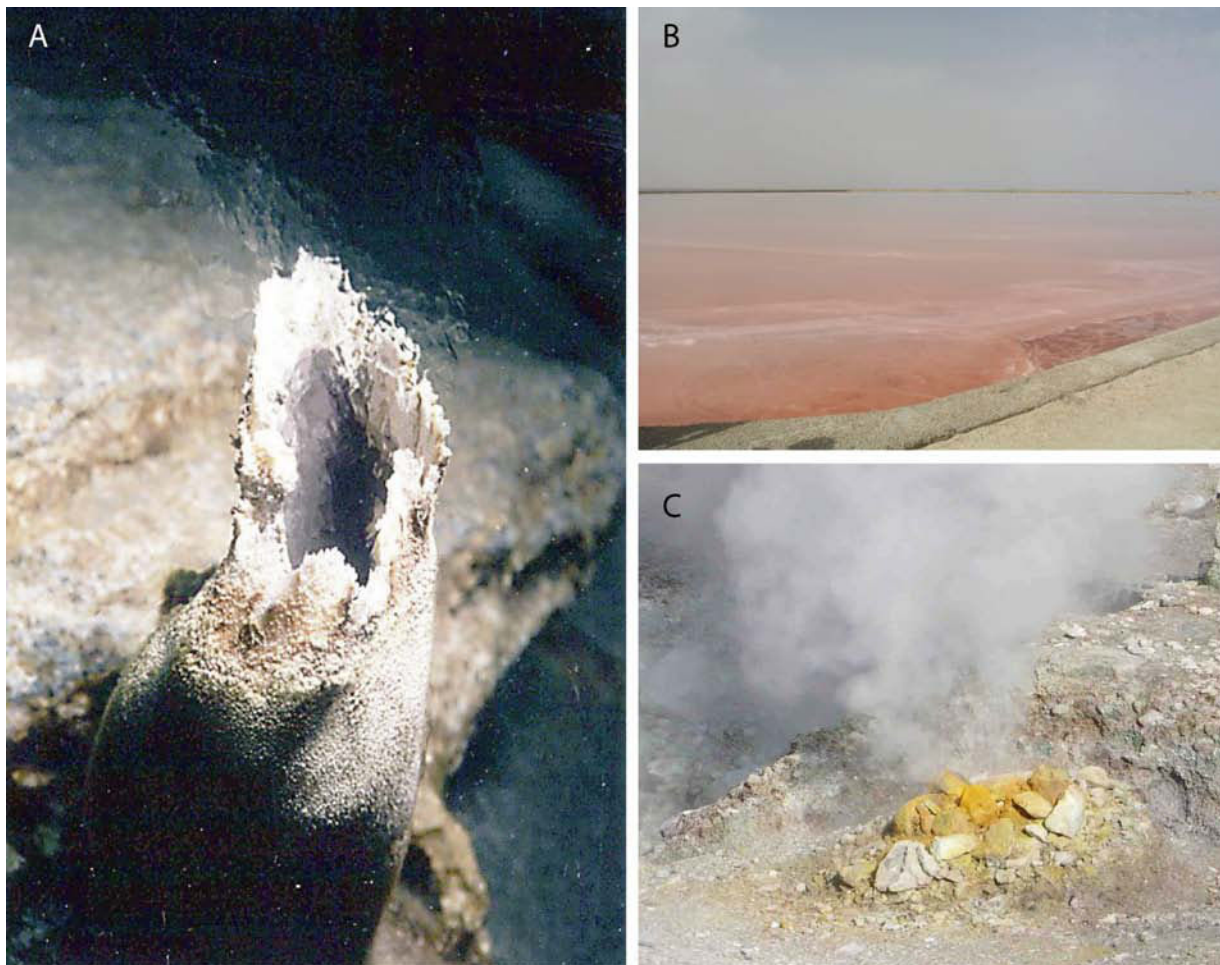


Figure I.2 : Exemples d'habitats des Archaea. **A.** Photographie d'une source chaude sous-marine : de l'eau minérale portée à ébullition par son passage à proximité du magma ressort à la surface d'un fond océanique. De nombreuses espèces hyperthermophiles peuvent être isolées à partir des filaments blancs observables au bout de cette "cheminée". **B.** Photographie d'un lac salé colonisé par des archaea halophiles. La couleur rose ou orangée du lac provient des pigments caroténoïdes synthétisés par ces archaea. **C.** Photographie d'un solfatare. Ce terme caractérise les terrains géologiques desquels se dégage, par des fissures, de la vapeur d'eau contenant de l'hydrogène sulfuré. On y observe les dépôts de soufre résultant de ces dégagements gazeux. Ces solfatares sont souvent colonisés par des archaea acidophiles thermophiles.

Le plus petit archaea est le symbiote *N. equitans* dont le diamètre n'excède pas 400 nm (Huber *et al.*, 2002) alors que le plus grand est l'atypique halophile *H. walsbyi* dont la forme carrée ou rectangulaire unique peut atteindre 40 à 100 μm de côté (Bolhuis *et al.*, 2004) (Fig. I.3). Nombreux sont ceux qui sont mobiles du fait de la présence d'un ou plusieurs flagelle(s) (pour revue, (Jarrell *et al.*, 1996) ; Fig. I.3). Ces observations sont en accord avec l'identification de transducteurs de phototaxie, aérotaxie et chemotaxie, i.e. le fait que certains archaea s'orientent en fonction de ou soient attirés par la lumière, l'oxygène ou certains produits chimiques (Gordeliy *et al.*, 2002 ; Kokoeva *et al.*, 2002). Les cellules archaea sont souvent associées en filaments, agrégats ou réseaux réticulés dont on ne connaît pas encore la fonction (Fig. I.3).

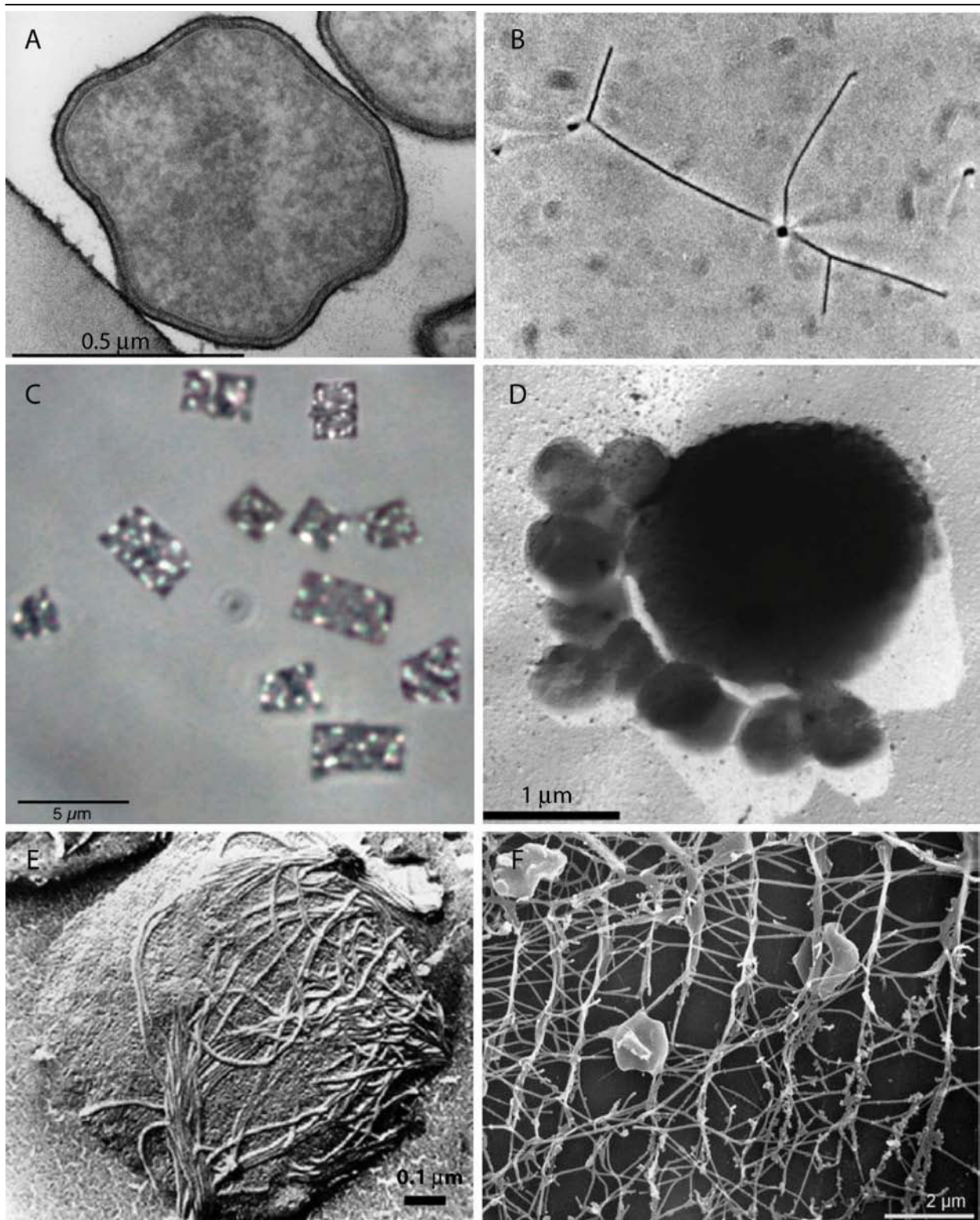


Figure I.3 : Exemples de formes cellulaires particulières aux archaea, observées en microscopie électronique. Motilité et associations des cellules archaea.

A. *Pyrolobus fumarii* : exemple de coque irrégulière de l'hyperthermophile cultivé présentant la température optimale de croissance la plus élevée (106°C) (tiré de Stetter, 2006). **B.** Structure en branche de *Thermoproteus*, la longueur des branches est de 100 µm pour 0.4 µm de diamètre (tiré de Stetter, 2006). **C.** Forme carrée ou rectangulaire unique de *Haloquadratum walsbyi* (photographie provenant du laboratoire de Dyall-Smith). **D.** Plusieurs cellules de *Nanoarchaeum equitans* fixées à la paroi d'*Ignicoccus hospitalis* (tiré de Huber *et al.*, 2002). **E.** Cellule de *Methanocaldococcus vulcanius* montrant plusieurs flagelles (tiré de Jeanthon *et al.*, 1999). **F.** Cellules de *Pyrodictium abyssi*, non flagellées, en forme de disque et impliquées dans un réseau de très fins tubules (tiré de Stetter, 2006).

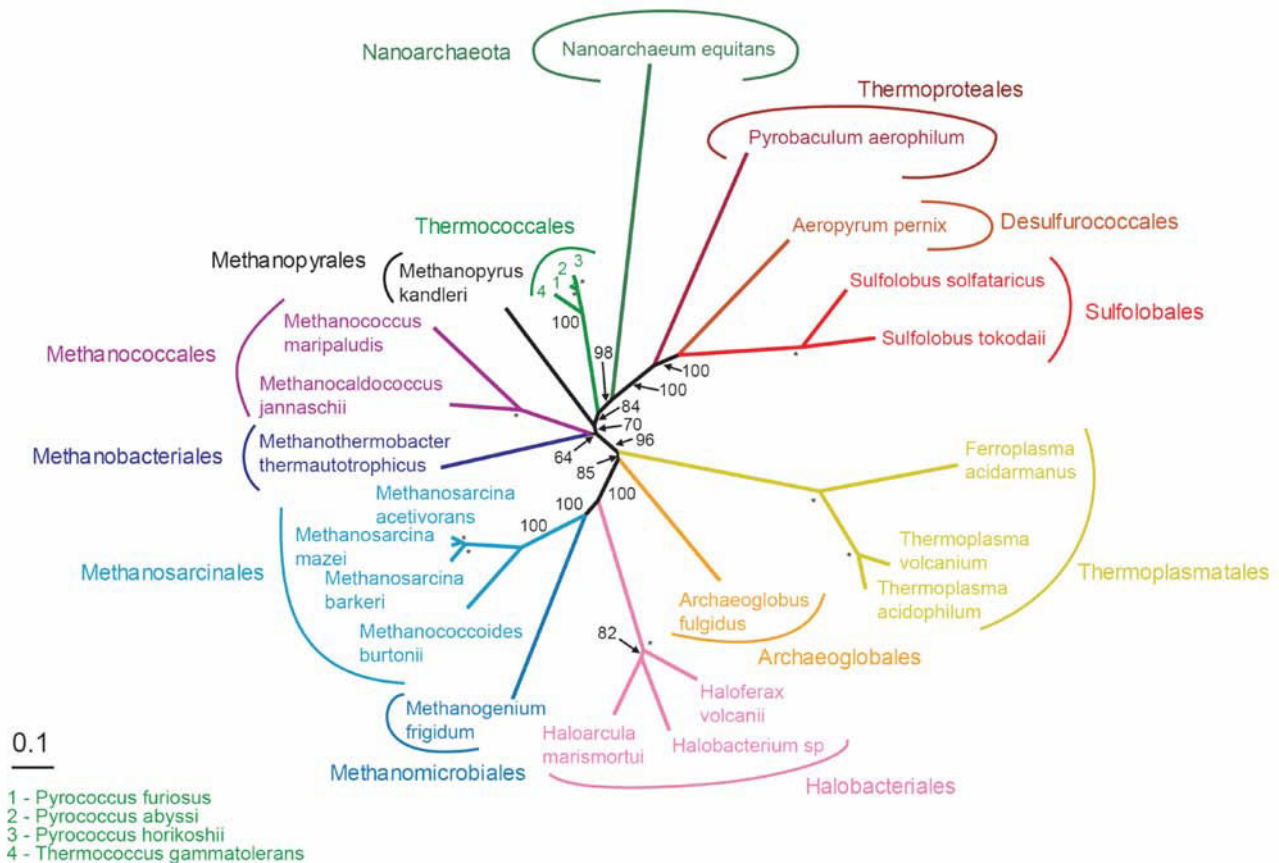


Figure I.4 : Arbre phylogénétique des Archaea basé sur la comparaison des protéines ribosomales (d'après Brochier *et al.*, 2005). Les nombres indiqués aux nœuds sont les valeurs de « bootstrap ». L'échelle correspond à 10 substitutions pour 100 positions.

B. Taxonomie

Actuellement plus d'une centaine d'espèces d'archaea ont pu être cultivées. Le séquençage complet des génomes d'une soixantaine d'entre elles est achevé ou en cours. J'ai eu accès à ce jour à 46 séquences génomiques complètes accessibles à partir du site du ncbi (http://www.ncbi.nlm.nih.gov/genomes/static/a_g.html), et une vingtaine sont en cours d'assemblage. L'étude comparative des séquences des petites sous-unités de l'ARNr révèle que le domaine comporte deux royaumes : les crénotes (crenarchaeota) et les euryotes (euryarchaeota) (Woese *et al.*, 1990). La rupture entre ces deux royaumes est très marquée, comme nous le verrons par la suite, puisque la distance phylogénétique qui les sépare serait plus importante que la distance qui sépare deux des 23 phyla bactériens. L'existence d'un troisième royaume qui ne regrouperait que des espèces non cultivées, appelé korarchaeota (Barns *et al.*, 1996), a été proposée. Elle n'est fondée que sur l'interprétation de données d'amplification par PCR des gènes des ARNr 16S à partir de prélèvements effectués dans différents environnements. Aucune preuve fiable de l'existence de ce royaume n'a encore été apportée (Forterre *et al.*, 2002). Récemment, il a également été suggéré qu'un nouveau royaume (nanoarchaeota) serait

représenté par le plus petit organisme hyperthermophile, *Nanoarchaeum equitans* (Boucher et Doolittle, 2002; Huber *et al.*, 2002). Il s'agit d'un symbiote d'Archaea, le seul identifié chez les hyperthermophiles. Il vit fixé à la paroi d'un crénote du genre *Ignicoccus* et ne peut être cultivé indépendamment, ce qui représente probablement un style de vie parasitaire (Huber *et al.*, 2003 ; Waters *et al.*, 2003 ; Das *et al.*, 2006). Cependant, il est vraisemblable que *N. equitans* ne soit qu'un représentant d'une lignée euryote. C'est ce que suggère une comparaison des protéines ribosomales et de la totalité des protéines du symbiote avec les protéines de toutes les autres espèces d'archaea et la prise en compte d'une vitesse évolutive probablement plus rapide liée au caractère parasitaire de cet organisme (Brochier *et al.*, 2005) (Fig. I.4).

Le royaume crénote est divisé en 6 ordres : les Caldisphaerales, les Cenarchaeales, les Nitrosopumilales, les Thermoproteales, les Desulfurococcales et les Sulfolobales (seules les trois dernières ont été suffisamment bien étudiées pour affirmer le caractère monophylétique de leur lignée ; Fig. I.4). Ces ordres regroupent environ 30 genres différents pour lesquels les séquences génomiques de 14 espèces sont disponibles : *Aeropyrum pernix* (noms de genre et d'espèce, le nom de la souche est donné dans le tableau I.1), *Cenarchaeum symbiosum*, *Hyperthermus butylicus*, *Ignicoccus hospitalis*, *Methallosphaera sedula*, *Pyrobaculum aerophilum*, *Pyrobaculum calidifontis*, *Pyrobaculum arsenaticum*, *Pyrobaculum islandicum*, *Staphylothermus marinus*, *Sulfolobus acidocaldarius*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, et *Thermophilum pendens* (Fig. I.4 et Tableau I.1).

Le royaume euryote est divisé en 10 ordres : Les Thermoplasmatales, les Archaeoglobales, les Halobacteriales, les Méthanomicrobiales, les Methanosarcinales, les Methanobacteriales, les Methanococcales, les Methanopyrales, les Thermococcales et les Nanoarchaea dont la position reste toutefois incertaine (Fig. I.4). Ces ordres regroupent environ 64 genres différents. Les séquences génomiques de 32 espèces sont disponibles : *Archaeoglobus fulgidus*, *Ferroplasma acidarmanus*, *Haloarcula marismortui*, *Halobacterium* sp. *NRC-1* (*halobium* ou *salinarium*), *Haloquadratum walsbyi*, *Methanobrevibacter smithii*, *Methanocaldococcus jannaschii*, *Methanococcoides burtonii*, *Methanococcus maripaludis*, *Methanococcus aeolicus*, *Methanococcus vanniellii*, *Methanocorpusculum labreanum*, *Methanoculleus marisnigri*, *Methanopyrus kandleri*, *Methanoregula boonei*, *Methanosaeta thermophila*, *Methanosarcina acetivorans*, *Methanosarcina barkeri*, *Methanosarcina mazei*, *Methanosphaera stadtmanae*, *Methanospirillum hungatei*, *Methanothermobacter thermautotrophicus*, *Nanoarchaeum equitans*, *Natronomonas pharaonis*, *Picrophilus torridus*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Thermococcus kodakarensis*, *Thermococcus gammatolerans* (génomique non disponible au public), *Thermoplasma acidophilum*, et *Thermoplasma volcanium* (Tableau I.1).

Le tableau I. 2 présente les caractéristiques des espèces dont le génome est en cours de séquençage.

Introduction - Partie I

Royaume	Ordre	Genre / espèce	T opt (°C)	Oxygène requis	Halophile	Acido-phile	Motilité	Habitat	Situation géographique
Crénotes	Cenarchaeales	<i>Cenarchaeum symbiosum</i>	10	Facultatif	Non	Non	Oui	Plancton marin	Multiple
	Desulfurococcales	<i>Aeropyrum pernix</i>	95	Aérobie	Non	Non	Oui	Fonds Marins volcaniques	Ile Japonaise
		<i>Staphylothermus marinus</i>	92	Anaérobie	Non	Non	Non	Sédiments marins zone volcanique	Italie
		<i>Ignicoccus hospitalis</i>	90	Anaérobie	Non	Non	Oui	Symbiote/Sources chaudes 100m de profondeur	Nord Islande
		<i>Hyperthermus butylicus</i>	95-106	Anaérobie	Modéré	Non	Oui	Fonds Marins volcaniques	Ile des Açores
	Thermoproteales	<i>Thermofilum pendens</i>	88	Anaérobie	Non	Oui	Non	Sources chaudes	Islande
		<i>Pyrobaculum islandicum</i>	100	Anaérobie	Non	Non	Oui	Eaux d'une centrale géothermique	Islande
		<i>Pyrobaculum caldifontis</i>	90-95	Facultatif	Non	Non	Oui	Sources chaudes	Philippines
		<i>Pyrobaculum arsenaticum</i>	95	Anaérobie	Non	Non	Oui	Sources chaudes et sulfatares	Italie
		<i>Pyrobaculum aerophilum</i>	100	Facultatif	Non	Non	Oui	Points d'eau chaude d'une plage	Italie
	Sulfolobales	<i>Metallosphaera sedula</i>	75	Aérobie	Non	Oui	Non	Sulfatares	Italie
		<i>Sulfolobus solfataricus</i>	87	Facultatif	Non	Oui	Oui	Sulfatares	Italie
		<i>Sulfolobus acidocaldarius</i>	75	Facultatif	Non	Oui	Non	Sources chaudes Yellowstone	Etats-Unis
		<i>Sulfolobus tokodai</i>	80	Facultatif	Non	Oui	Non	Sources chaudes	Japon
Euryotes	Nanoarchaeota	<i>Nanoarchaeum equitans</i>	90	Anaérobie strict	Non	Non	Non	Symbiote/Sources chaudes 100m de profondeur	Nord Islande
	Archaeoglobales	<i>Archaeoglobus fulgidus</i>	83	Anaérobie strict	Non	Non	Oui	Fonds Marins volcaniques	Ile Italienne
	Halobacteriales	<i>Halobacterium sp. NRC1</i>	42	Facultatif	Extrême	Non	Oui	Lacs salés et déchets salés	Multiple
		<i>Haloarcula marismortui</i>	50	Aérobie	Extrême	Non	Oui	Mers et lacs salés	Mer Morte
		<i>Haloquadratum walsbyi</i>	40	Facultatif	Extrême	Non	Non	Etangs salés	Egypte
		<i>Natronomonas pharaonis</i>	45	Aérobie	Modéré	Non	Non	Lacs salés à pH élevé	Egypte
	Methanococcales	<i>Methanocaldococcus jannashii</i>	85	Anaérobie strict	Non	Non	Oui	Sources chaudes en profondeur	Multiple
		<i>Methanococcus aeolicus</i>	42	Anaérobie	Non	Non	Oui	Sédiments en profondeur	Japon
		<i>Methanococcus vanniellii</i>	35-40	Anaérobie	Non	Non	Oui	Boue baie San Francisco	Etats-Unis (CAL)
		<i>Methanococcus maripaludis</i>	35-40	Anaérobie strict	Oui	Non	Oui	Marais salant	Multiple
	Methanopyrales	<i>Methanopyrus kandleri</i>	98-105	Anaérobie strict	Modéré	Non	Oui	Fonds Océaniques (1000-3650m) sources chaudes	Océan Atlantique Nord
	Methanosarcinales	<i>Methanococcoides burtonii</i>	23	Anaérobie strict	Modéré	Non	Oui	Hypolimnion des lacs	Lac Ace, Antarctique
		<i>Methanosaeta thermophila</i>	50-65	Anaérobie	Non	Non	Non	Rizières, égouts	Multiple
		<i>Methanosarcina acetivorans</i>	35	Facultatif	Non	Non	Non	Sédiments marins	Multiple
		<i>Methanosarcina barkeri</i>	35	Facultatif	Non	Non	Non	Boues marines, rumens ongués, lacs	Multiple
		<i>Methanosarcina mazei</i>	30	Facultatif	Non	Non	Non	Fumier, égouts, compost, faeces d'herbivores...	Multiple
	Methanomicrobiales	<i>Methanospirillum hungatei</i>	37	Anaérobie	Non	Non	Non	Rumens bovins, vase, boues, égouts	Multiple
		<i>Methanoculleus marisnigri</i>	21-25	Anaérobie	Non	Non	Oui	Sédiments marins des eaux froides	Mer Noire, Multiple
		<i>Methanoregula boonei</i>	37	Anaérobie	Non	Oui		Marais de tourbe	Etats-Unis (NY)
		<i>Methanocorpusculum labreanum</i>	37	Anaérobie	Non	Non	Non	Sédiments à la surface de Puits de goudron	Etats-Unis (CAL)
	Methanobacteriales	<i>Methanosphaera stadtmanae</i>	37	Anaérobie	Non	Non	Non	Tractus intestinal humain	Multiple
		<i>Methanobrevibacter smithii</i>	37-40	Anaérobie	Non	Non	Non	Sol, Sédiments, Déchets, Tractus digestif animaux	Multiple
		<i>Methanothermobacter thermoautotrophicum</i>	62-67	Anaérobie strict	Non	Non	Non	Egouts et déchets dans environnement chaud	Multiple
	Thermococcales	<i>Pyrococcus abyssi</i>	96	Anaérobie strict	Modéré	Non	Oui	Fonds Océaniques (3500m) sources chaudes	Océan Pacifique proximité îles Fidji
		<i>Pyrococcus horikoshii</i>	98	Anaérobie strict	Non	Non	Oui	Fonds Océaniques (1300m) sources chaudes	Océan Pacifique proximité îles du Japon
		<i>Pyrococcus furiosus</i>	100	Anaérobie strict	Non	Non	Oui	Fonds Marins volcaniques	Italie
		<i>Thermococcus kodakaraensis</i>	95	Anaérobie	Non	Non	Oui	Sulfatares	Japon
		<i>Thermococcus gammatolerans</i>	85-90	Anaérobie strict	Non	Non	Oui	Sources chaudes	Mexique
	Thermoplasmatales	<i>Ferropasma acidarmanus</i>	40	Facultatif	Non	Oui	Non	Mines de Fer	Etats-Unis (COL)
		<i>Picrophilus torridus</i>	60	Aérobie	Non	Oui		Sols volcaniques acides	Nord du Japon
		<i>Thermoplasma acidophilum</i>	55-60	Facultatif	Non	Oui	Oui	Charbon auto-chauffant acide, Yellowstone	Etats-Unis
		<i>Thermoplasma volcanium</i>	55-60	Facultatif	Non	Oui	Oui	Fonds Marins volcaniques	Italie

Tableau I.1 : Caractéristiques des espèces dont la séquence génomique a été utilisée au cours de ce travail de thèse. Les informations proviennent des publications relatives au séquençage de chacune des espèces dont les références sont indiquées dans le tableau I.3, ainsi que des publications relatives à l'isolement et la caractérisation de chacune de ces espèces. Lorsque la présence d'oxygène est requise l'espèce est dite aérobie.

Si l'espèce présente les systèmes permettant une croissance en aérobiose et en anaérobiose, le besoin en oxygène est dit facultatif. Lorsqu'une espèce ne présente pas le système permettant une respiration en aérobiose, elle est dite anaérobie voire anaérobie stricte ou microaérophile si elle ne tolère pas ou peu la présence d'oxygène.

Royaume	Ordre	Genre / espèce	T opt (°C)	Oxygène requis	Halophile	Acido-phile	Motilité	Habitat	Situation géographique
Crénates	Nitrosopumilales	<i>Nitrosopumilus maritimus</i>	28	Aérobie	Non	Non		Bassin marin tropical d'un aquarium de Seattle	Probablement Multiple
	Desulfurococcales	<i>Pyrodictum abyssi</i>	97	Anaérobie	Non	Non	Non	Sources chaudes des abysses	Ile Japonaise
		<i>Pyrolobus fumarii</i>	106	Facultatif	Non	Modéré	Non	Cheminées noires des abysses	Faïlle Atlantique
	Thermoproteales	<i>Caldivirga maquilingensis</i>	85	Microaérophile	Non	Non	Non	Sources chaudes	Philippines
		<i>Thermoproteus neutrophilus</i>	85	Anaérobie	Non	Non	Non	Sources chaudes	Islande
	Sulfolobales	<i>Acidianus brierleyi</i>	65	Facultatif	Non	Oui	Non	Solfatares et Drainage des mines	Italie
		<i>Sulfolobus islandicus</i>	75-85	Facultatif	Non	Oui	Non	Sources chaudes Yellowstone	Etats-Unis
		<i>Sulfolobus metallicus</i>	65	Facultatif	Non	Oui	Non	Sources chaudes et solfatares	Italie, Caraïbes
Euryotes	Halobacteriales	<i>Halobaculum gomorense</i>	40	Aérobie	Extrême	Non	Oui	Sédiments riches en Mg	Mer Morte
		<i>Halobiforma lacisalsi</i>	42-45	Facultatif	Extrême	Non	Oui	Lac à 3880m d'altitude	Chine
		<i>Haloferax volcanii</i>	45	Aérobie	Modéré	Non	Oui	Sédiments	Mer Morte
		<i>Halorubrum lacusprofundi</i>	31-37	Aérobie	Extrême	Non	Oui	Lac gelé	Antarctique
		<i>Natrialba asiatica</i>	30-40	Aérobie	Extrême	Non	Oui	Plage	Japon
	Methanococcales	<i>Methanothermococcus thermolithotrophicus</i>	65	Anaérobie	Non	Non	Oui	Sédiments des fonds marins à proximité sources chaudes	Multiple
		<i>Methanococcus voltae</i>	37	Anaérobie	Non	Non	Oui	Sédiments des estuaires	Multiple
	Methanosarcinales	<i>Methanosarcina thermophila</i>	50	Anaérobie	Non	Non	Non	Boue, déchets	Multiple
		<i>Methanosaeta concilii</i>	35	Anaérobie	Non	Non		Sols des rizières	Multiple
	Methanomicrobiales	<i>Methanogenium frigidum</i>	15	Anaérobie	Non	Non	Non	Lac gelé	Antarctique
	Methanobacteriales	<i>Methanothermus fervidus</i>	83	Anaérobie	Non	Non	Oui	Sources chaudes	Islande
	Thermoplasmatales	<i>Aciduliprofundum boonei</i>	70	Anaérobie	Non	Oui	Oui	Fonds Océaniques (1900m) sources chaudes	Océan Pacifique

Tableau I.2 : Caractéristiques des espèces dont la séquence génomique est en cours d'assemblage. Les informations proviennent des publications relatives à l'isolement et la caractérisation de chacune des espèces.

C. Phénotypes

Les archaea sont autotrophes ou hétérotrophes. Le caractère autotrophe correspond à la capacité qu'a l'organisme de se développer en milieu strictement minéral sans matière organique générée par un autre être vivant. L'autotrophie correspond donc à la possibilité de produire de la matière organique par la réduction du carbone inorganique (dioxyde de carbone), ce qui s'accompagne de l'absorption d'autres minéraux du milieu (ions nitrate, sulfates, phosphates...). L'énergie des archaea autotrophes provient soit des liaisons chimiques (chemio-lithoautotrophes) soit de la photosynthèse qui s'appuie sur la famille des gènes de la rhodopsine (et non sur la chlorophylle) présentant des équivalents bactériens (photo-lithoautotrophes) (Beja *et al.*, 2000 ; Beja *et al.*, 2001). Ils comportent une grande variété de donneurs et d'accepteurs d'électrons (Huber *et al.*, 2000).

Indépendamment de la classification phylogénétique on peut distinguer les archaea mésophiles dont la température optimale de croissance est comprise entre 20 et 50°C, les archaea thermophiles qui ont une température optimale de croissance supérieure à 50°C et les archaea

hyperthermophiles qui présentent une température optimale de croissance supérieure à 80°C. Ainsi, *Pyrolobus fumarii*, dont le séquençage du génome est en cours, présente une température optimale de croissance de 106°C et survit après une heure de traitement en auto-clave à une température de 121°C pour une pression de 200kPa (Blochl *et al.*, 1997). Le record est détenu par une souche dénommée « strain 121 » qui se développe encore à une température de 121°C et voit sa masse doubler après 24h de ce même traitement en autoclave (Kashefi et Lovley, 2003). Certaines souches sont au contraire psychrophiles, ce qui signifie qu'elles ont une température optimale de croissance inférieure à 20°C et le seul représentant dont la séquence génomique est disponible est *Cenarchaeum symbosium*, symbiote du spongiaire *Axinella mexicana* (Preston *et al.*, 1996). Enfin, certaines souches sont dites psychrotolérantes car elles supportent de longues périodes à des températures inférieures à 10°C (cas de *Methanococcoides burtonii*). De telles basses températures correspondent aux températures de la plupart des océans et des mers (2 à 3°C en moyenne). Certaines souches d'archaea semblent même capables de se développer à des températures inférieures à 0°C : *Methanococcoides burtonii* ($T_{\min}$: -2.5°C; T_{opt} : 23°C) (Franzmann *et al.*, 1992), *Methanogenium frigidum* ($T_{\min}$: -10°C; T_{opt} : 15°C) (Franzmann *et al.*, 1997). Par ailleurs, des espèces mésophiles voire thermophiles peuvent aussi occuper ces niches (Vorobyova *et al.*, 1997). Il est à noter que des températures supérieures à 90°C sont principalement observées dans les zones volcaniques. Les milieux correspondants sont pauvres en oxygène du fait de sa faible solubilité à température élevée. Les archaea hyperthermophiles colonisent en particulier les sources chaudes où l'eau minérale, portée à ébullition à proximité du magma remonte à la surface terrestre ou des sources chaudes sous-marines, chargée en sels, sulfates et carbonates qui lui donnent un caractère plus ou moins acide. Ces sources forment notamment des solfatares ou des cheminées sous-marines (pour revue Huber *et al.*, 2000 ; Stetter, 2006).

Les espèces archaea peuvent également être caractérisées en fonction de leur pH optimum de croissance. Pour les espèces acidophiles ce pH doit être inférieur à 3. Au contraire, pour les espèces alcalinophiles le pH doit être supérieur à 10. Les acidophiles se développent dans les zones volcaniques où le soufre est oxydé en acide sulfurique qui est concentré par évaporation (Schleper *et al.*, 2005). Les alcalinophiles colonisent principalement les lacs riches en sodium du fait d'un déséquilibre de la concentration en bicarbonate de sodium (Rees *et al.*, 2004).

Une autre distinction peut être faite en séparant les espèces qui supportent ou même requièrent de fortes pressions, appelées respectivement barotolérants ou barophiles. Les barophiles les plus extrêmes peuvent se développer à des pressions allant jusqu'à 700 atm (c'est le cas pour *P. abyssi*). De telles conditions sont rencontrées dans les profondeurs des mers et océans, dans les roches en formation ou encore dans les puits de pétrole.

Enfin, certaines espèces appartenant à l'ordre des *Thermococcales* dont *T. gammatolerans* (Tableau I.1) sont résistantes à des doses élevées de rayonnements gamma, de l'ordre de 30 kGy (pour *T. gammatolerans* ; Jolivet *et al.*, 2003a) alors que la dose létale est de 10 Gy pour l'homme et 100 Gy pour *E. coli*. Ces espèces classées comme radio-résistantes ont dû être

sélectionnées au cours de l'évolution dans les milieux naturels où l'exposition à des niveaux élevés de radioactivité est forte. Il a été démontré en effet que certaines cheminées (sources chaudes) sous marines présentent des taux de radioactivité 100 fois supérieurs à la radioactivité rencontrée à la surface terrestre (Cherry *et al.*, 1992). Ceci s'explique par la présence de métaux parmi lesquels certains isotopes radioactifs qui se concentrent autour de ces cheminées.

L'organisme sur lequel nous avons beaucoup travaillé, *P. abyssi* fait partie, en matière de taxonomie, du royaume des euryotes, de l'ordre des Thermococcales, du genre *Pyrococcus* et de l'espèce *abyssi* (la famille, ici Thermococcaceae, est peu utilisée) (Tableau I.1). Son phénotype est hyperthermophile, barophile et radio-résistant. On peut également mentionner qu'il fait partie des thermophiles réducteurs de composés soufrés et qu'il s'agit d'un halophile modéré.

Sur le plan du métabolisme, il est possible de distinguer 4 catégories d'archaea selon leurs propriétés physiologiques : -1) les Méthanogènes qui réduisent le dioxyde de carbone ou des formes oxydées de carbone en méthane lors de leur croissance en anaérobiose qui est plus ou moins stricte. Tous les noms de genre correspondant à des espèces méthanogènes commencent par « Methano ». On distingue 5 ordres différents correspondant à des espèces méthanogènes (pour revue, Garcia *et al.*, 2000), -2) les Thermophiles réducteurs de composés soufrés qui sont anaérobies et réduisent le soufre en H₂S, -3) les archaea Soufre dépendants sont aérobies facultatifs ou microaérophiles (capables de supporter une faible pression d'oxygène) et ils oxydent le soufre en acide sulfurique ou réduisent les composés soufrés en H₂S. Enfin -4) les Halophiles extrêmes qui requièrent une concentration élevée en sels (NaCl) (Tableau I.1 et 2). De telles concentrations sont rencontrées dans des zones sèches ou alimentées en eau mais avec un taux d'évaporation élevé. Elles sont en général comprises entre 1,5 M et 5.2 M NaCl soit 85 g/l à 300 g/l (pour revue, Oren, 2002). Ces concentrations sont observées dans des mers ou lacs salés et sont supérieures à la concentration moyenne observée dans les océans (inférieure à 30 g/l). Il est à noter que la plupart des bactéries non-halophiles ne survivent pas à une concentration en sels dépassant 0.85 M soit 50 g/l (Boch *et al.*, 1994 ; Lucht et Bremer, 1994). Cependant, il existe des espèces halophiles dans les trois domaines du vivant (*Salinibacter ruber*, bactérie halophile ou encore les eucaryotes du genre *Dunaliella*). A noter aussi qu'il existe des bactéries qui réduisent les composés soufrés. Ainsi, seuls les phénotypes méthanogène et hyperthermophile sont propres aux Archaea.

Le royaume des crénotes ne compte que des archaea Soufre dépendants alors que le royaume euryote comporte des espèces présentant l'une ou l'autre des trois autres catégories de métabolisme (Mathrani *et al.*, 1988). On observe plus d'homogénéité des propriétés des espèces au sein des différentes branches crénotes que parmi les euryotes (Woese *et al.*, 1990). D'ailleurs, « euryote » vient du grec signifiant « large » alors que « crénote » vient du grec signifiant « origine ». Ils étaient en effet initialement supposés être à l'origine non seulement du domaine Archaea mais aussi du monde du vivant. En fait, l'apparente homogénéité du domaine crénote pourrait être due au faible nombre de représentants crénotes étudiés. Ainsi, l'analyse récente des deux seuls crénotes non thermophiles caractérisés à ce jour, *Nitrosopumilus maritimus* (Konneke

et al., 2005) et *Cenarchaeum symbiosum* (Hallam *et al.*, 2006), symbiote d'une algue qui fait partie du plancton marin, révèle que ces organismes et, par conséquent plusieurs lignées archaea, pourraient utiliser l'ammoniaque comme source primaire d'énergie c'est-à-dire une stratégie différente des autres crénotes. Ce métabolisme particulier pourrait, du fait de l'abondance des crénotes dans les océans (estimée à plus de 20%) (Karner *et al.*, 2001), jouer un rôle important dans le cycle de l'azote des océans. Une autre lignée d'Archaea, les méthanotrophes, présente une autre stratégie métabolique de par leur faculté à convertir le méthane en dioxyde de carbone, soit la réaction inverse de la méthanogenèse (Orphan *et al.*, 2001 ; Hallam *et al.*, 2004).

Une étude en méta-génomique c'est-à-dire une étude qui est étendue aux génomes récupérés à partir d'échantillons environnementaux comportant de nombreuses espèces non-cultivées (environ 8000), met en en exergue une diversité beaucoup plus importante que celle précédemment évoquée. Ainsi, les deux royaumes compteraient plus de phyla ainsi que des stratégies métaboliques/propriétés physiologiques différentes des 4 catégories présentées précédemment (pour revue, Schleper *et al.*, 2005). Il est également envisageable que l'analyse des prélèvements par PCR ait apporté un biais dans la détection de nouvelles lignées puisque les séquences des sondes propres aux archaea déjà identifiés peuvent être trop différentes des séquences d'ARNr provenant de nouvelles lignées d'archaea. Les nouvelles techniques de recherche basées sur le séquençage de tous les ADN présents dans un prélèvement après clonage dans de grands vecteurs (type BAC, Bacterial Artificial Chromosome) ainsi que l'amélioration des conditions de culture, permettront de mieux apprécier cette diversité (pour revue, Robertson *et al.*, 2005).

Les caractéristiques des espèces dont la séquence génomique est disponible ou en cours d'assemblage, sont respectivement présentées dans les tableaux I.1 et I.2. La taxonomie, la température optimale de croissance, le besoin en oxygène, les caractères halophile, acidophile, et motile des espèces y sont indiqués avec le lieu où elles ont été isolées. Parmi les génomes en cours de séquençage on peut compter 8 espèces crénotes dont un représentant d'un nouvel ordre (Nitrosopumilales) et 12 espèces euryotes.

D. Phylogénie et Evolution

Etant donné que la taxonomie des Archaea a été établie récemment, elle est largement plus basée sur les études phylogénétiques récentes que sur la comparaison des caractéristiques morphologiques et phénotypiques encore souvent incomplètes. La phylogénie permettant de retracer l'évolution des espèces à partir d'un ancêtre hypothétique commun, il est intéressant de l'utiliser pour essayer de comprendre comment la grande variété de phénotypes des archaea a été générée au cours de l'évolution.

En se basant sur l'hypothèse de l'existence d'une constance de la vitesse d'évolution au cours du temps, on peut, à partir de la connaissance du taux d'accumulation des mutations dans les séquences nucléotidiques ou d'acides aminés, estimer le temps de divergence entre espèces par comparaison de leurs séquences. Le principe de base est de comparer des gènes homologues,

c'est à dire descendant d'un ancêtre commun, qui ont conservé une fonction identique au cours du temps.

Au départ, les ARNr ont été beaucoup utilisés dans ces études puisqu'ils sont universels et que certains de leurs domaines ont été très conservés au cours de l'évolution alors que d'autres présentent un plus fort degré de variation. De plus, ils sont abondants dans la cellule et faciles à purifier. Il est ainsi possible de définir des liens phylogénétiques entre espèces très éloignées et entre espèces proches. Cependant, plus récemment, d'autres gènes dont les produits sont très conservés ont aussi été utilisés comme marqueur génétique pour établir les liens phylogénétiques entre archaea.

Il est important dans ces études de tenir compte des transferts horizontaux possibles qui peuvent fausser la définition des clades et des distances entre espèces. Ainsi, des espèces qui colonisent les mêmes niches peuvent partager des gènes sans que ceux-ci aient été hérités d'un ancêtre commun. Cet échange de gènes peut conduire à l'acquisition de nouvelles fonctionnalités permettant par exemple la mise en place de nouvelles voies métaboliques. Par ailleurs, des mécanismes de « cell mating » observés chez certains halophiles peuvent favoriser un tel échange (Rosenshine *et al.*, 1989). En effet, un système de transfert naturel de matériel génétique qui semble propre aux Archaea a été observé chez *H. volcanii*. Celui-ci requiert un contact physique entre deux cellules parentales qui peuvent être sans distinction donneur ou récepteur et implique une continuité cytoplasmique entre deux ou plusieurs cellules par le biais de ponts cytoplasmiques. Il est à noter que l'échange est restreint au mouvement de certaines molécules telles que l'ADN chromosomique (Rosenshine *et al.*, 1989).

Une analyse du contexte génétique par génomique comparative peut permettre d'identifier des déplacements de groupes de gènes orthologues appelés COG (Clusters of Orthologous Genes) parmi des espèces éloignées d'un point de vue évolutif (pour revue, Koonin *et al.*, 2001 ; Makarova et Koonin, 2003). Au sein des archaea, de nombreux transferts latéraux de gènes ont été observés, parmi lesquels les mieux décrits sont ceux entre les Thermoplasmatales et les Sulfolobales, tous deux thermophiles acidophiles mais appartenant aux euryotes et aux crénotes, respectivement (Fig. I.4 ; Ruepp *et al.*, 2000). Ce gain de gènes par transfert horizontal ne se limite pas aux archaea, il peut aussi se faire à partir de génomes bactériens chez certaines espèces telles que *M. mazei* pour laquelle 30% du génome dont la taille est très grande (4.1 Mpb) provient du gain par transfert horizontal (Deppenmeier *et al.*, 2002). Cependant, des études montrent que les transferts horizontaux chez les Archaea restent des événements rares (Forterre *et al.*, 2002).

La vitesse de divergence d'une espèce peut aussi conduire à des erreurs de positionnement (Felsenstein, 1978). De telles situations ont été observées notamment chez *M. kandleri* (Brochier *et al.*, 2004) et *N. equitans* (Brochier *et al.*, 2005), en particulier, lorsque des groupes éloignés du point de vue évolutif sont utilisés pour l'étude.

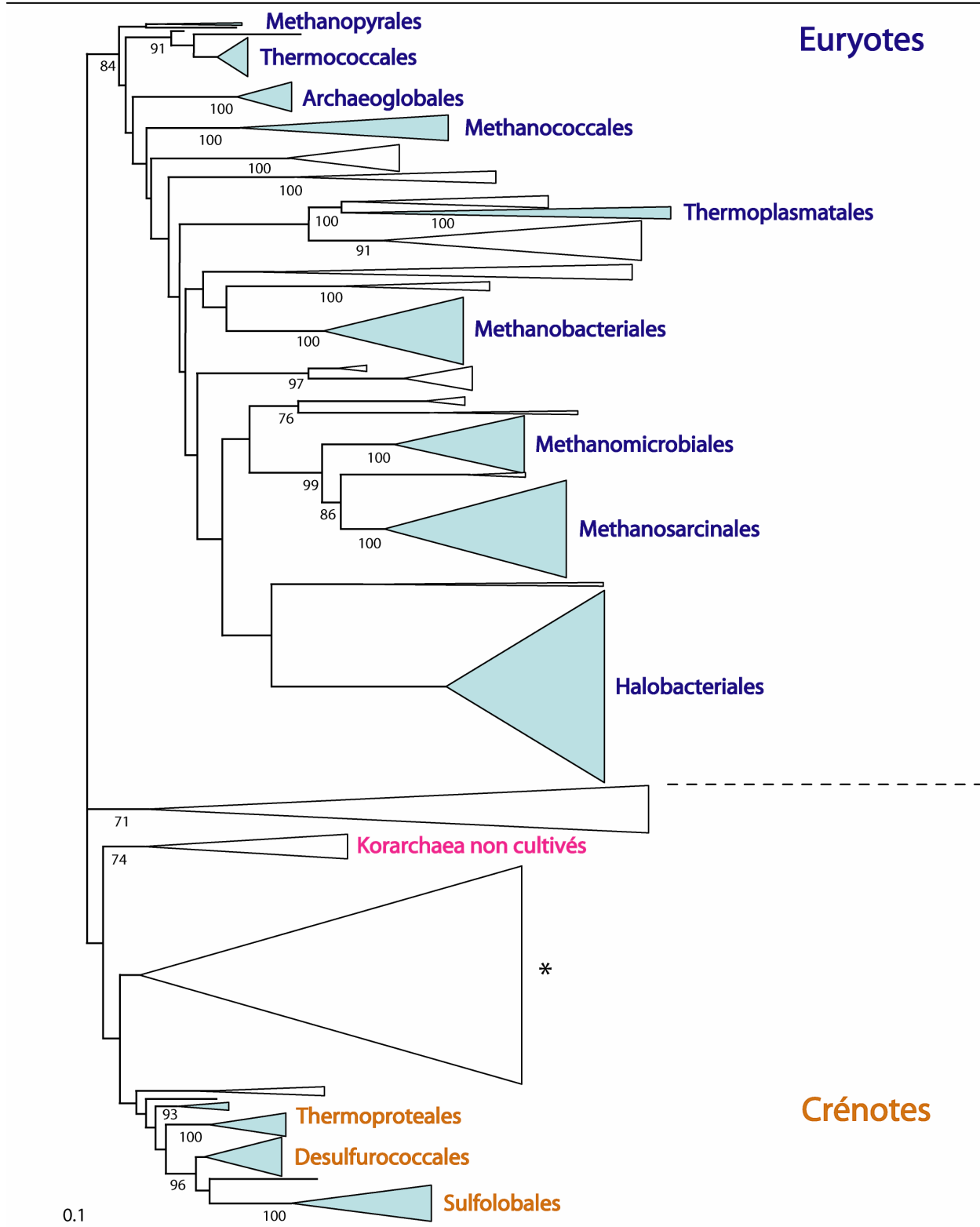


Figure I.5 : Arbre phylogénétique des Archaea basé sur la comparaison des séquences des ARNr 16S des espèces cultivées ou non cultivées (342 séquences presque complètes) (d'après Forterre *et al.*, 2002). La taille des triangles est proportionnelle au nombre de séquences analysées. Les triangles blancs n'incluent que des séquences d'espèces non cultivées provenant de prélèvements effectués dans différents environnements, tandis que les triangles bleus représentent des séquences provenant à la fois d'espèces cultivées et non cultivées. Les nombres indiqués aux nœuds sont les valeurs de « bootstrap ». L'échelle correspond à 10 substitutions pour 100 positions. * ce groupe comporte certainement l'ordre des Cenarchaeales, abondant dans les océans, dont un représentant (*Cenarchaeum symbiosium*) est étudié depuis peu de temps, voir figure I.6. A noter que l'ordre des Nanoarchaea n'est pas représenté du fait qu'un seul représentant est connu, voir figure I.4.

Du fait des difficultés à cultiver un grand nombre d'espèces, les arbres phylogénétiques construits à partir des seules données sur les espèces cultivables sont moins riches en terme de clades que ceux construits à partir des espèces non cultivées (Fig. I.5). Ainsi, les études des ARNr 16S d'Archaea non cultivés suggèrent la présence d'un phylum supplémentaire en plus des euryotes et des crénotes (Fig. I.6) (Forterre *et al.*, 2002). Les nouvelles lignées mises en évidence à partir des amplifications d'ADN semblent être largement distribuées dans le plancton, les eaux froides ou le sol.

La construction d'un arbre revient à grouper et rapprocher différentes classes entre elles parce qu'elles partagent un caractère commun qui n'est pas présent chez un ancêtre plus « primitif » (synapomorphie). A chaque fois que des espèces ou des groupes d'espèces sont rapprochés, une valeur, appelée « bootstrap », indiquée en général en pourcentage, traduit statistiquement le bien fondé de ce rapprochement. En comparant les positions relatives sur l'arbre phylogénétique ainsi que les phénotypes des espèces archaea, il est possible d'apprécier l'émergence des différentes caractéristiques au cours de l'évolution. Il ressort de ces analyses que la méthanogenèse est une stratégie métabolique complexe apparue très tôt et une seule fois au sein du royaume euryote après la divergence avec le phylum des Thermococcales mais avant la séparation avec les Archaeoglobales, Halobacteriales et Thermoproteales qui en présentent des reliquats (Fig. I.5 ; Klenk *et al.*, 1997 ; Matte-Tailliez *et al.*, 2002 ; Slesarev *et al.*, 2002 ; Baptiste *et al.*, 2005). Certains ordres des méthanogènes partagent avec l'ordre des Halobacteriales, phylogénétiquement proche, la présence d'organismes halophiles (Methanococcales), l'absence d'organisme thermophile (Methanomicrobiales) et la propriété de tolérer l'oxygène (Methanosarcinales) (tableau I.1). Les gènes impliqués dans la respiration en aérobiose de *H. sp. NRC1* ont pu être empruntés aux bactéries par transfert latéral de gènes à moins qu'un ancêtre commun aux méthanogènes et Halobacteriales ait présenté ce système de respiration perdu par la suite chez les méthanogènes (Kennedy *et al.*, 2001). Les crénotes comportent à la fois des espèces anaérobies, des espèces aérobies et des espèces aérobies facultatives réparties entre et parfois au sein des 5 ordres (Tableau I.1 et 2). Il semble donc que l'apparition de la respiration oxygénée y soit une adaptation secondaire (Forterre *et al.*, 2002). L'adaptation à une atmosphère oxygénée est donc apparue de manière indépendante au moins deux fois au sein des euryotes (Halobacteriales et Thermoplasmatales) et au moins une fois au sein des crénotes (Forterre *et al.*, 2002).

Ces études suggèrent que l'ancêtre commun aux Archaea était anaérobie voire microaérophile (supportant une faible pression en oxygène). Cet ancêtre était aussi vraisemblablement hyperthermophile.

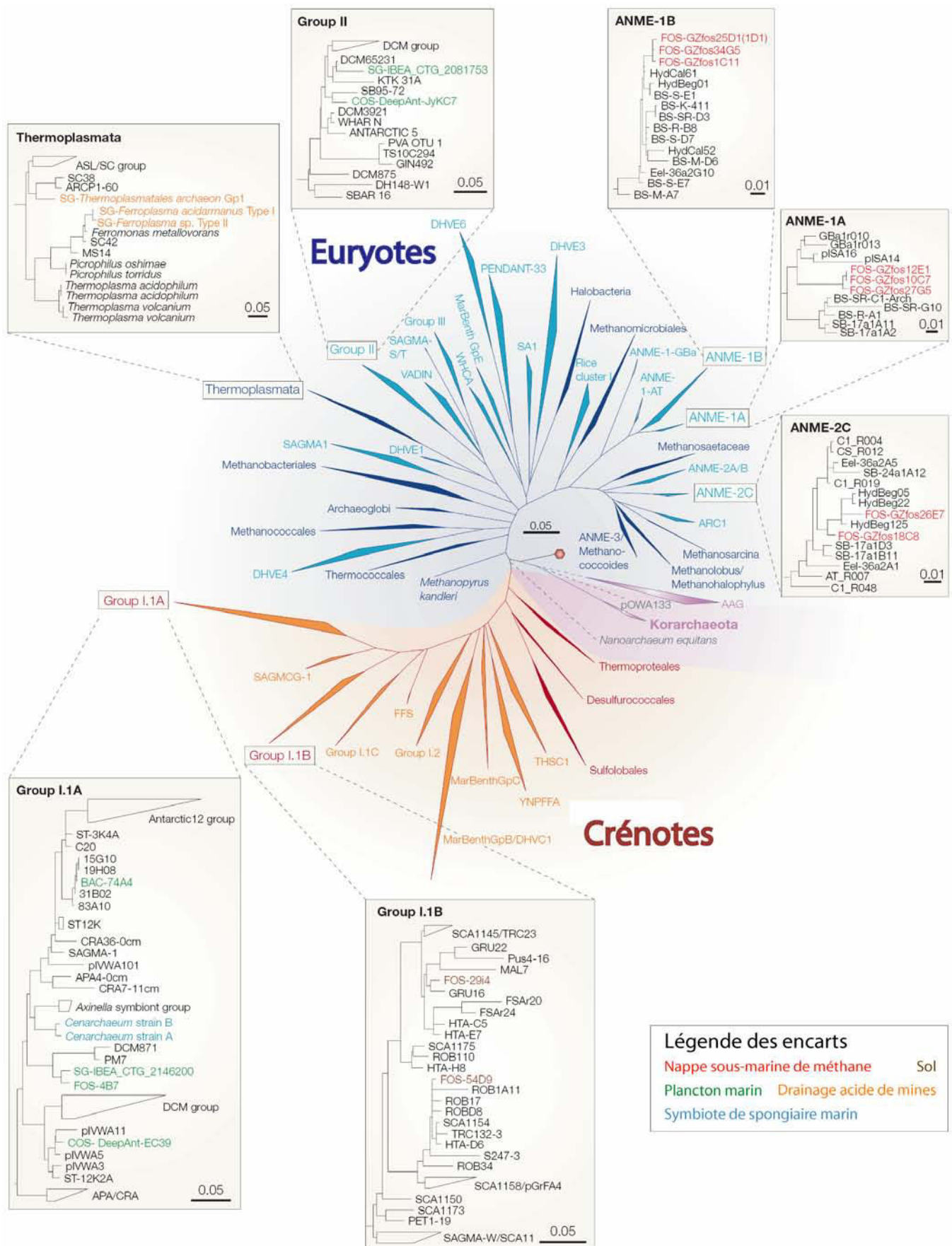


Figure I.6 : Arbre phylogénétique des Archaea basé sur la comparaison des séquences des ARNr 16S des espèces cultivées et non cultivées (1344 séquences presque complètes) (d'après Schleper *et al.*, 2005). Les préfixes indiquent qu'il s'agit de foscides (FOS), cosmides (COS), chromosomes bactériens artificiels (BACs)

ou séquences assemblées à partir de projets de séquençage shotgun (SG). Les triangles de couleur claire représentent les branches comportant exclusivement des espèces non cultivées tandis que les triangles de couleur foncée représentent les branches comportant des espèces cultivées. La taille des triangles est proportionnelle au nombre de séquences analysées. L'hexagone rouge représente l'enracinement dans le domaine bactérien. L'échelle correspond à 5 substitutions pour 100 positions et est différente selon les encarts.

En effet, la membrane particulière des archaea, qui semble être adaptée à ces conditions de vie anaérobies et à une température élevée est similaire dans l'ensemble des archaea. De plus, dans l'arbre phylogénétique, tous les hyperthermophiles émergent à la base de la séparation entre euryotes et crénotes avec des tailles de branche courtes suggérant qu'ils ont conservé plus de caractéristiques de l'ancêtre commun des Archaea que les autres espèces archaea (Fig. I.5 ; (Matte-Tailliez *et al.*, 2002). Le gène de la gyrase inverse est présent chez tous les procaryotes hyperthermophiles et absent chez les mésophiles et thermophiles modérés. Une analyse phylogénique révèle que ce gène a été transféré à partir des Archaea vers les Bactéries hyperthermophiles par transfert horizontal de gènes, et suggère par conséquent que le phénotype hyperthermophile est apparu chez les Archaea et a ensuite pu être transmis (Forterre *et al.*, 2000 ; Forterre, 2002).

3. Caractéristiques des Archaea

A. Organisation du génome

Les 46 génomes d'archaea dont la séquence complète est disponible sont constitués de un parfois deux chromosomes circulaire(s) (cas unique de *H. marismortui*). Par ailleurs, des plasmides de très grande taille sont présents et il n'est pas facile de distinguer les méga plasmides de mini chromosomes comportant des gènes essentiels. En effet, certaines espèces comme *H. sp. NRC1* présentent plusieurs plasmides dont la taille peut atteindre 400 kpb (Ng *et al.*, 1998 ; Ng *et al.*, 2000). En plus des plasmides, les archaea peuvent, comme les bactéries, contenir de nombreux virus (voir le paragraphe ci-après). La taille des génomes des virus d'archaea dont la séquence est disponible varie de 15 à 75 kpb.

Le tableau I.3 résume les caractéristiques des génomes disponibles. Leur taille varie de 1.5 Mpb (*P. torridus*) à 5.8 Mpb (*M. acetivorans*) avec une exception pour le symbiote *N. equitans* dont la taille est inférieure à 500 kpb (ce qui en fait le plus petit organisme séquencé à ce jour). La taille des génomes d'archaea est d'une manière générale plus petite que celle des génomes bactériens qui dépasse souvent 5 Mpb. Les variations de taille entre génomes d'archaea semblent intimement liées aux phénomènes de transferts latéraux de gènes, discutés ci-dessus, et à la présence de transpositions et de répétitions au sein des génomes.

Genre / espèce / souche	Taille (pb)	% codant	Nombre d'ORF	Nombre de gènes d'ARNnc	% G/C	Références
<i>Aeropyrum pernix</i> K1	1669696	88	1700	52	56	Kawarabayasi et al., 1999
<i>Archaeoglobus fulgidus</i> DSM 4304	2178400	91	2420	49	48	Klenk et al., 1997
<i>Cenarchaeum symbiosum</i>	2045096	91	2066	48	58	Hallam et al., 2006
<i>Ferroplasma acidarmanus</i> fer1	1935153	87	1896		36	Copeland et al., non publié
<i>Haloarcula marismortui</i> ATCC 43049	3419774	85	3416	59	62	Baliga et al., 2004
<i>Halobacterium</i> sp NRC-1 ATCC 700922	2014239	88	2075	52	67	Ng et al., 2000
<i>Haloquadratum walsbyi</i> DSM 16790	3132494	74	2610	52	47	Bolhuis et al., 2006
<i>Hyperthermus butylicus</i> DSM 5456	1667136	80	1602	52	54	Brugger et al., 2006
<i>Ignicoccus hospitalis</i> sp. Kin4-I	1297490	91	1434	50	57	Copeland et al., non publié
<i>Metallosphaera sedula</i> DSM 5348	2191517	87	2256	47	46	Copeland et al., non publié
<i>Methanocaldococcus jannaschii</i> DSM 2661	1664970	88	1729	43	31	Bult et al., 1996
<i>Methanococcoides burtonii</i> DSM 6242	2575032	80	2273	63	40	Copeland et al., soumis
<i>Methanococcus maripaludis</i> C5	1780761	86	1813	47	33	Copeland et al., non publié
<i>Methanococcus maripaludis</i> C7	1772694	86	1788	47	33	Copeland et al., non publié
<i>Methanococcus maripaludis</i> S2	1661137	88	1722	50	33	Hendrickson et al., 2004
<i>Methanococcus aeolicus</i> Nankai-3	1569500	84	1788	47	33	Copeland et al., non publié
<i>Methanococcus vannielii</i> SB	1720048	85	1678	51	31	Copeland et al., non publié
<i>Methanocorpusculum labreanum</i> Z	1804962	86	1739	63	50	Copeland et al., non publié
<i>Methanobrevibacter smithii</i>	1853150	90	1793	42	31	Samuel et al., 2007
<i>Methanoculleus marisnigri</i> JR1	2478101	87	2489	53	62	Copeland et al., non publié
<i>Methanopyrus kandleri</i> AV19	1694969	88	1687	38	61	Slesarev et al., 2002
<i>Methanoregula boonei</i> 6A8	2542943	85	2450	53	54	Copeland et al., non publié
<i>Methanosaeta thermophila</i> PT	1879471	82	1696	51	53	Copeland et al., soumis
<i>Methanosarcina acetivorans</i> C2A	5751492	73	4540	69	42	Galagan et al., 2002
<i>Methanosarcina barkeri</i> str. fusaro chromosome 1	4837408	69	3606	73	39	Copeland et al., non publié
<i>Methanosarcina mazei</i> Go1	4096345	74	3370	66	41	Deppenmeier et al., 2002
<i>Methanosphaera stadtmanae</i> DSM 3091	1767403	84	1534	52	27	Fricke et al., 2006
<i>Methanospirillum hungatei</i> JF-1	3544738	85	3139	66	45	Copeland et al., non publié
<i>Methanothermobacter thermautotrophicus</i> str. Delta H	1751377	90	1873	47	49	Smith et al., 1997
<i>Nanoarchaeum equitans</i> Kin4-M	490885	92	536	31	46	Waters et al., 2003
<i>Natronomonas pharaonis</i> DSM 2160	2595221	90	2661	51	63	Falb et al., 2005
<i>Picrophilus torridus</i> DSM 9790	1545895	91	1535	47	35	Futterer et al., 2004
<i>Pyrobaculum aerophilum</i> str. IM2	2222430	88	2605	96	51	Fitz-Gibbon et al., 2002
<i>Pyrobaculum arsenaticum</i> DSM 13514	2121028	86	2299	45	55	Copeland et al., non publié
<i>Pyrobaculum islandicum</i> DSM 4184	1826402	86	1978	48	49	Copeland et al., non publié
<i>Pyrobaculum caldifontis</i> JCM 11548	2009313	90	2149	26	57	Copeland et al., non publié
<i>Pyrococcus abyssi</i> GE5	1765118	92	1896	97	44	Genoscope, non publié
<i>Pyrococcus furiosus</i> DSM 3638	1908256	91	2125	104	40	Maeder et al., 1999
<i>Pyrococcus horikoshii</i> OT3	1738505	92	1955	50	41	Kawarabayasi et al., 1998
<i>Staphylothermus marinus</i> F1	1570485	86	1546	45	35	Copeland et al., non publié
<i>Sulfolobus acidocaldarius</i> DSM 639	2225959	84	2223	82	36	Chen et al., 2005
<i>Sulfolobus solfataricus</i> P2	2992245	83	2977	56	35	She et al., 2001
<i>Sulfolobus tokodaii</i> str. 7	2694756	83	2825	49	32	Kawarabayasi et al., 2001
<i>Thermococcus kodakarensis</i> KOD1	2088737	91	2306	52	51	Fukui et al., 2005
<i>Thermophilum pendens</i> Hrk 5	1781889	90	1824	40	57	Copeland et al., non publié
<i>Thermoplasma acidophilum</i> DSM 1728	1564906	87	1482	48	45	Ruepp et al., 2000
<i>Thermoplasma volcanium</i> GSS1	1584804	85	1499	49	39	Kawashima et al., 2000

Tableau I.3 : Caractéristiques des génomes dont les séquences ont été utilisées lors de ce travail de thèse.

Les conditions extrêmes de température ou de pH semblent limiter la plasticité des génomes (Tableaux I.1&2 ; Futterer *et al.*, 2004). Ainsi, les hyperthermophiles et acidophiles possèdent des génomes de plus petite taille.

Les pourcentages en G/C des différents génomes varient fortement entre 27% (*M. stadtmanae*) et 67% (*H. sp. NRC1*). Il est intéressant de noter que les génomes des archaea halophiles présentent les valeurs les plus élevées. Le pourcentage élevé en G/C constituerait une protection face aux lésions de l'ADN (de type dimères de thymines) provoquées par les irradiations solaires aux UV, plus fréquentes chez les halophiles (Goo *et al.*, 2004). Notons qu'il n'existe pas, chez les archaea, de corrélation entre pourcentage en G/C et température optimale de croissance.

La fraction de séquence codant des gènes varie de 69% à 92% selon les espèces, les valeurs les plus élevées étant bien sûr observées pour les génomes les plus compacts. Du fait de la plus grande taille des génomes bactériens la fraction codante est légèrement plus élevée que chez les bactéries (80-85%). Plusieurs introns ont été détectés dans des gènes d'ARN non codant (Lykke-Andersen *et al.*, 1997) alors qu'un seul exemple d'introns dans les gènes de protéines de aCBF5 a été relevé (Watanabe *et al.*, 2002). Quelques protéines présentent des intéines, c'est-à-dire des fragments de séquences en acides aminés retirés après la traduction (Gogarten *et al.*, 2002).

Le nombre de gènes d'ARN non codant annotés est très semblable d'une espèce à une autre. Le plus petit nombre a été attribué au plus petit des génomes, celui de *N. equitans*. Actuellement, ces gènes correspondent essentiellement aux gènes d'ARNr et d'ARNt, les gènes de sRNAs que nous décrirons plus loin n'ayant été identifiés que dans quelques espèces. Les gènes codant les ARNr sont uniques ou dupliqués deux à quatre fois, selon l'espèce et ce nombre est en rapport avec la taille du génome.

Les génomes d'archaea comportent environ 2000 ORF chacun, parmi lesquelles environ 313 COG sont communs à 13 représentants d'archaea (nombre de génomes disponibles au moment de l'étude) et codent des fonctions essentielles. Environ 200 de ces COG sont rencontrés à la fois chez les Bactéries et les Eucaryotes, 34 chez les Bactéries uniquement, 63 chez les Eucaryotes uniquement et 16 ne sont jamais retrouvés chez les Bactéries et les Eucaryotes (Makarova et Koonin, 2003). Les premières observations montraient déjà qu'une lignée de gènes codant des protéines intervenant dans le traitement de l'information génétique, dits gènes informationnels, semble s'apparenter aux équivalents eucaryotes. En effet, parmi les 63 COG partagés entre archaea et eucaryotes, seuls deux ne sont pas des gènes informationnels. Au contraire, les gènes d'archaea codant les protéines impliquées dans le métabolisme central, la conversion de l'énergie et la biosynthèse, qui sont dits être des gènes opérationnels, ressemblent à leurs équivalents bactériens et sont également organisés en opérons (Bult *et al.*, 1996 ; Smith *et al.*, 1997 ; Rivera *et al.*, 1998). Les opérons ont été définis comme correspondant à l'association d'au moins deux ORF espacées de moins de 40 nt. En l'absence de connaissance sur les signaux de transcription il est en effet difficile de se baser sur ces signaux. Ainsi, environ 50 à 70% des gènes des archaea semblent organisés en opérons contenant en moyenne 2,5 à 3 gènes (Torarinsson *et al.*, 2005).

Les 16 COG particuliers aux archaea sont principalement impliqués dans la traduction et la modification des ARN.

Il est à noter cependant que près de la moitié des gènes d'archaea identifiés à ce jour restent sans fonction évidente attribuée.

B. Réplication, Réparation et Chromatine

1. Réplication

La réplication est un processus précis et ponctuel faisant intervenir de nombreuses protéines qui assurent la duplication du matériel génétique. Son initiation marque le début de la phase S (par analogie avec les Eucaryotes) (pour revue, Lundgren et Bernander, 2005). Bien que l'étude de la réplication chez les archaea puisse constituer un modèle pour la compréhension de ces phénomènes chez les eucaryotes, l'état actuel des connaissances sur ce processus chez les archaea est très incomplet et, en fait, les études réalisées s'appuient sur les connaissances acquises chez les bactéries et les eucaryotes.

Beaucoup d'archaea ont, comme les bactéries, une seule origine de réplication. Néanmoins, certains crénotes possèdent de multiples origines de réplication (Myllykallio et Forterre, 2000 ; Lundgren *et al.*, 2004; Robinson et Bell, 2005). La réplication se fait de manière semi-conservatoire sur les deux brins. Ceci impose une machinerie capable de synthétiser à la fois un brin dit précoce et un brin dit retardé (synthèse semi-continue). Elle est aussi bidirectionnelle et semble débiter de manière synchrone au niveau des origines multiples (Lundgren *et al.*, 2004 ; pour revue, Kelman et White, 2005). Les origines correspondent, comme pour les bactéries, à des régions riches en A/T. La localisation de ces régions peut se faire par voie bioinformatique par identification d'un changement rapide de biais dans l'utilisation des codons ainsi que dans la fréquence des nucléotides et oligomères (Grigoriev, 1998 ; Lopez *et al.*, 1999). Ces biais reflèteraient une différence dans le taux d'erreurs ou de réparation du brin précoce par rapport au brin retardé (Karlin, 1999). Chez *S. cerevisiae*, contrairement aux autres eucaryotes, les origines multiples sont clairement définies par une séquence spécifiquement reconnue par le complexe ORC (Origin Recognition Complex). Ce dernier contient 6 protéines, Orc1-6 (pour revue, Bell et Dutta, 2002). Notons que Orc1 et Cdc6, un autre facteur de la réplication, appartiennent à la même famille. Chez les eucaryotes supérieurs, l'origine n'est pas clairement caractérisée par une séquence consensus et pourrait être définie par l'interaction de l'ADN avec des protéines annexes qui recrutent le complexe ORC (Coverley et Laskey, 1994). Les archaea présentent une ou plusieurs protéines homologues à Cdc6/Orc1 qui assureraient une fonction analogue à la protéine bactérienne DnaA (Tableau I.4 ; pour revue, Robinson et Bell, 2005). Il est à noter que l'origine de réplication des archaea est presque toujours localisée à proximité d'un gène codant une protéine de la famille Cdc6/Orc-1, ce qui les rapproche des bactéries pour lesquelles l'origine est souvent localisée à proximité du gène codant DnaA (Lundgren *et al.*, 2004). Des boîtes appelées ORB (Origin Recognition Boxes) qui sont constituées de séquences répétées

inversées (IR) définissent le point d'initiation de la réplication et sont reconnues par Cdc6/Orc1 (Robinson *et al.*, 2004). DnaA diffère principalement de Cdc6/Orc-1 par la reconnaissance de séquences consensus différentes (Erzberger *et al.*, 2002) et par le fait que DnaA ne reste pas fixé à l'origine et permet de dénaturer un segment du double brin, ce qui n'est le cas ni du complexe ORC eucaryote ni des protéines Cdc6/Orc-1 archaea (Geraghty *et al.*, 2000); pour revue (Bell et Dutta, 2002). Le rôle des différentes protéines Cdc6/Orc-1 présentes dans une espèce archaea n'est pas encore clairement défini (Robinson *et al.*, 2004).

	Bactéries	Eucaryotes	Archaea
Protéine se liant à l'origine de réplication (OBP)	DnaA (1 sous-type)	ORC (6 sous-types Orc1 à 6)	Cdc6/Orc1 (1 ou 2 sous-types)
Chargeur d'hélicase	DnaB (1 sous-type)	Cdc6 fixé à ORC (1 sous-type)	Cdc6/Orc1 (1 ou 2 sous-types)
Hélicase répllicative	DnaC (1 sous-type)	MCM (6 sous-types)	MCM (1 sous-types)
Protéine se liant à l'ADN simple-brin	SSB (1 sous-type)	RPA (3 sous-types)	RPA/SSB (1 ou 3 sous-types)

Tableau I.4 : Les protéines impliquées dans l'initiation de la réplication chez les bactéries, les eucaryotes et les archaea (d'après Kelman et Kelman, 2003).

L'association de Cdc6/Orc1 sur oriC conduit au recrutement de l'hélicase MCM (Mini Chromosome Maintenance) (Matsunaga *et al.*, 2001). Chaque archaea présente au moins un représentant de la famille MCM (jusqu'à 4 exemplaires pour *M. jannaschii*). MCM est multimérique (le nombre de sous-unités varie d'un organisme à l'autre). Elle semble être l'hélicase impliquée dans la réplication qui déplace le complexe de réplication dans le sens 3' vers 5' (Shin *et al.*, 2003). La protéine bactérienne DnaB est l'hélicase analogue à MCM, elle se fixe à DnaA grâce à DnaC qui ressemble à Cdc6 (Tableau I.4 ; pour revue, Robinson et Bell, 2005). En outre, les protéines Orc1/Cdc6 sont capables de s'auto-phosphoryler, ce qui inactiverait l'hélicase MCM et régulerait ainsi une nouvelle initiation de la réplication (Grabowski et Kelman, 2001). MCM interagit donc avec l'ADN, avec Orc1/Cdc6 mais également avec la primase via les protéines GINS (voir ci-après ; Marinsek *et al.*, 2006). Le caractère essentiel de MCM n'a pas été démontré, en effet d'autres hélicases pourraient avoir des fonctions équivalentes (Bolt, 2005 ; Fujikane *et al.*, 2005 ; Guy et Bolt, 2005). Les archaea présentent aussi des équivalents de protéines SSB bactériennes (Single-Stranded Binding protein) qui fixent et protègent l'ADN simple brin de la dégradation ou de modifications lors de certains processus. Cependant, les protéines SSB d'eucaryotes présentent plus de similarités avec

leurs équivalents eucaryotes, les protéines RPA (Replication Protein A) qu'avec les protéines SSB bactériennes (Kerr *et al.*, 2003).

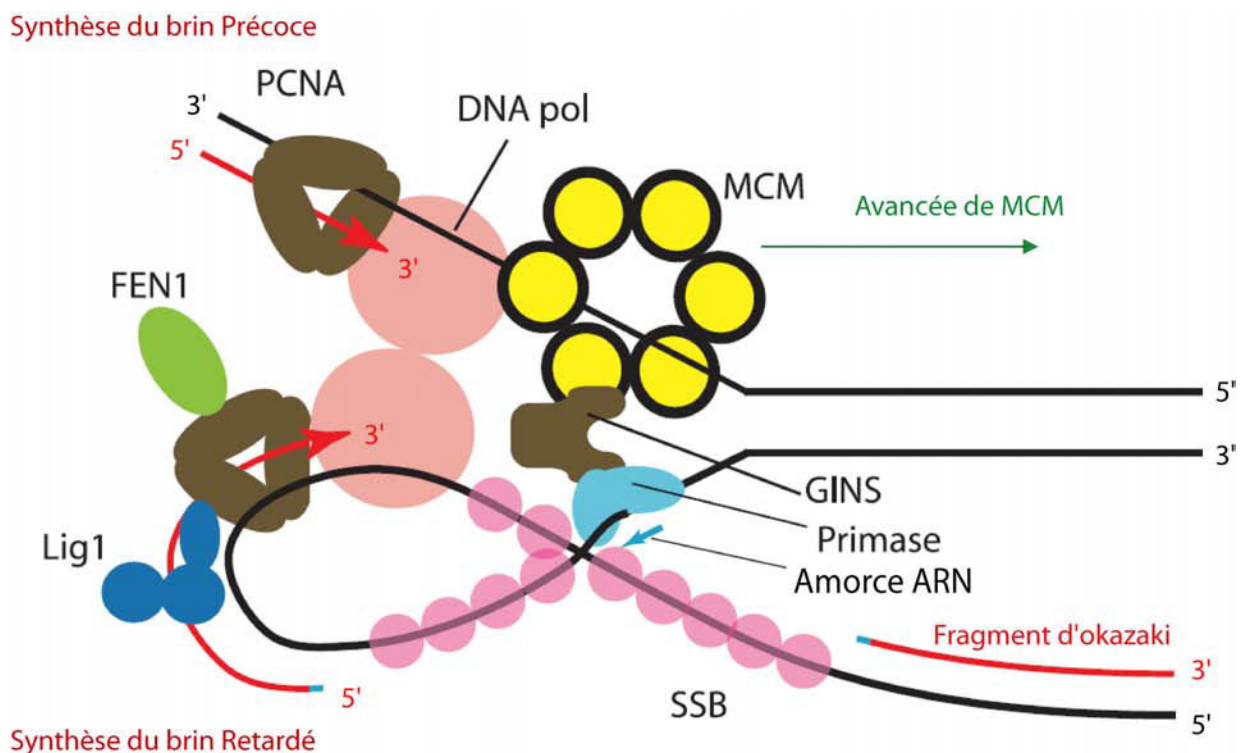


Figure I.7 : Modèle de l'architecture de la machinerie de réplication des archaea (d'après Barry et Bell, 2006). L'ADN parental est représenté en noir et l'ADN nouvellement synthétisé est représenté en rouge. L'amorce ARN synthétisée par la primase est représentée en bleu. MCM est l'hélicase répllicationnelle qui déplace la machinerie du 3' vers le 5' sur le brin précoce en déroulant le duplex d'ADN parental. Les deux ADN polymérases (DNA pol) sont représentées comme étant couplées, ce qui n'est pas encore prouvé dans le système archaea.

Dans les trois domaines du vivant la synthèse *de novo* du brin retard requiert la présence d'amorces qui sont synthétisées par la primase. L'amorce est ensuite étendue par l'ADN polymérase Pol α . Or l'absence d'équivalent de Pol α chez les Archaea suggère que la primase d'archaea joue également un rôle dans l'extension de l'amorce (Matsunaga *et al.*, 2003). Les amorces ARN sont ensuite dégradées par les nucléase RNase H et FEN1 (Flap EndoNuclease) et une ADN ligase I joint les produits de l'extension d'amorces maturés et dont les espaces ont été complétés par une ADN polymérase. Des équivalents des protéines eucaryotes GINS ont été caractérisés chez les archaea. Il est probable que ces protéines qui interagissent avec à la fois MCM et la primase, forment un pont entre l'hélicase MCM associée au brin précoce et la primase ancrée sur le brin retard (Fig. I.7 ; Marinsek *et al.*, 2006). Les euryotes possèdent deux types d'ADN polymérase répllicative, l'une ubiquitaire appartenant à la famille B et l'autre appartenant à la famille D qui est unique aux archaea appartenant au royaume euryote. Une étude biochimique réalisée chez *Pyrococcus* montre que la polymérase B1 synthétise le brin précoce

alors que la polymérase D synthétise le brin retardé (Henneke *et al.*, 2005). Les crénotes abritent différents types de polymérase B (au moins B1 et B2), contrairement aux euryotes, ce qui suggère une probable spéciation des rôles vis-à-vis des deux brins. Enfin, des protéines accessoires viennent « clamber » les différents intervenants et assurent ainsi la processivité de la réplication. Un équivalent archaea de la protéine eucaryote PCNA (Proliferating Cell Nuclear Antigen), est présent chez les archaea (Fig. I.8). Des protéines accessoires de type RFC (Replication Factor C) maintiennent PCNA sur l'ADN (Miyata *et al.*, 2005). Le trimer PCNA, via des motifs d'interaction particuliers, appelés PIP (PCNA Interacting Protein), lie FEN1, l'ADN polymérase B et l'ADN ligase I (Fig. I.7 ; Warbrick, 2000; Vivona et Kelman, 2003). Il a été démontré que la fourche de réplication avançait à une vitesse allant de 100 bp/s pour *Sulfolobus* (Lundgren *et al.*, 2004), ce qui est similaire à celle des eucaryotes, à 300 bp/s pour *P. abyssi* (Myllykallio et Forterre, 2000), ces vitesses sont inférieures à celles déterminées pour les bactéries (1000 bp/s). La réplication bidirectionnelle de l'ADN prend fin, comme chez les bactéries, dans une région située à l'opposé de l'origine de réplication (Myllykallio et Forterre, 2000).

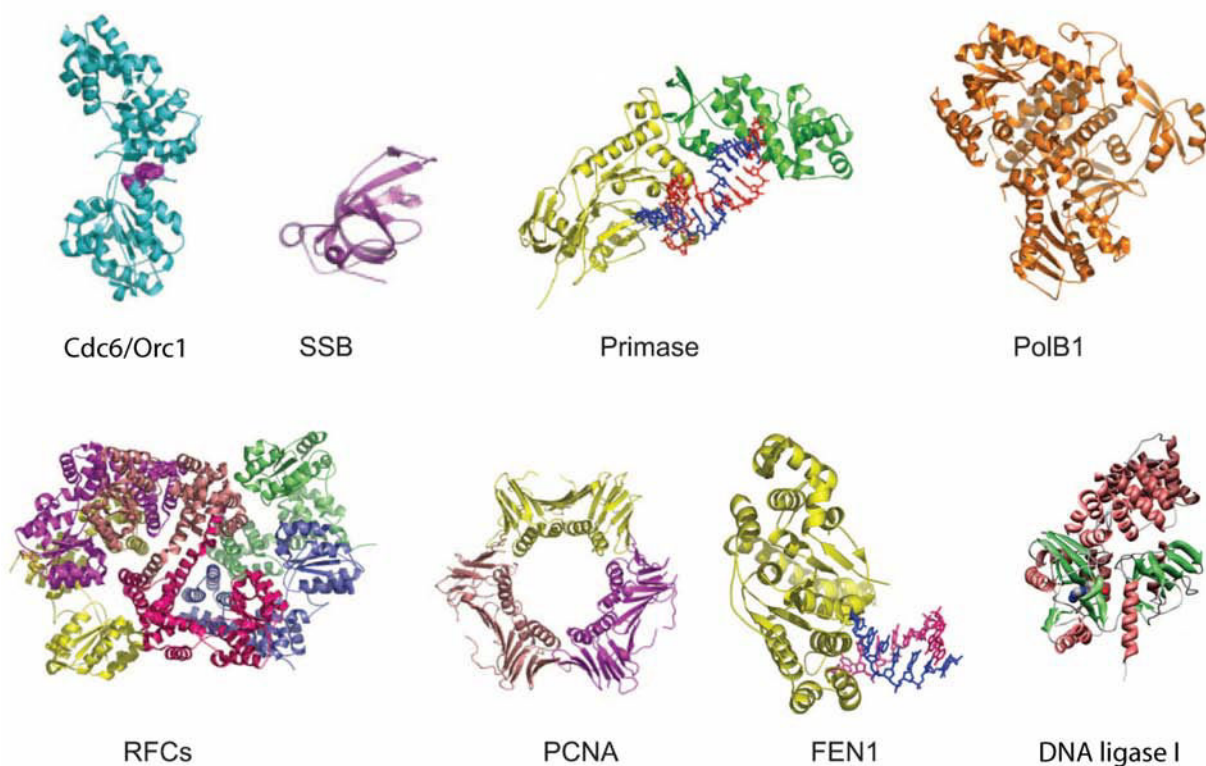


Figure I.8 : Structures 3D de facteurs protéiques intervenant dans la machinerie de la réplication des Archaea (d'après Barry et Bell, 2006).

Les structures 3D des protéines Cdc6/Orc1, SSB, primase, PolB1, RFC, PCNA, FEN1 et DNA ligase I sont présentées. Les coordonnées PDB sont les suivantes : *Pyrobaculum* Cdc6, 1FNN ; *Sulfolobus* SSB, 1O71 ; *Pyrococcus* RFC small subunit, 1IQP ; *Sulfolobus* Pol B1, 1S5J ; *Pyrococcus* PCNA, 1ISQ ; *Archaeoglobus* Fen1, 1RXW. Les structures de la ligase et de la primase proviennent de Y. Ishino et L. Pellegrini, respectivement.

2. Chromatine

Les régulations possibles de la réplication et de la transcription doivent impérativement prendre en considération le fait que l'ADN n'est pas nu mais associé à de nombreux types de protéines. En effet, une grande quantité d'ADN nu présent dans un petit volume de solvant précipiterait. De petites protéines basiques liées à l'ADN permettent d'éviter cette agrégation (pour revue, Sandman et Reeve, 2005) et ceci dans les trois domaines du vivant. Dans chacun de ces domaines du vivant, la chromatine représente un complexe dynamique de protéines et d'ADN qui permet de modifier l'encombrement de l'ADN en particulier en favorisant son super-enroulement et en empêchant son agrégation.

La plupart des archaea comporte des représentants de deux familles de protéines chromatinienne, Alba et les histones, et de manière moins conservée des représentants d'autres familles, c'est le cas de HTa, membre de la famille d'histone-like bactérienne HU identifié uniquement chez les Thermoplasmales, qui est certainement le résultat d'un transfert horizontal de gènes à partir de bactéries. Deux autres protéines, Sul7 et est MC1 ont été identifiées chez les Sulfolobales, et les Halobacteriales et les Methanomicrobiales, respectivement (pour revue, White et Bell, 2002).

Comme mentionné ci-dessus, des histones ont été identifiées chez les euryotes (à l'exception des Thermoplasmales) et, récemment, dans l'ordre des Cenarchaeales (Cubonova *et al.*, 2005) et l'ordre des Thermoproteales des crénates (Fig. I.9 ; Sandman et Reeve, 2006). Ceci suggère que la famille des histones est apparue avant la division archaea/eucaryotes mais après la division bactéries/archaea et qu'elle est remplacée par d'autres familles chez certains crénates hyperthermophiles au sein du genre *Thermoplasma* et chez les eucaryotes de la classe des dinoflagellés (Fig. I.9 ; Sandman et Reeve, 2005). Les archaea codent de 1 à 6 histones qui sont associées en homo- ou hétéro-dimères mais qui, en présence d'ADN, polymérisent en tetramères et hexamères (Marc *et al.*, 2002). Les nucléosomes des eucaryotes sont composés d'un tetramère (H3-H4)₂ central flanqué par deux dimères H2A-H2B. L'ADN au sein des nucléosomes eucaryotes est beaucoup moins accessible qu'au sein des architectures formées chez les archaea ce qui permet un accès rapide à l'ADN. Les nucléosomes d'archaea ne renferment en effet qu'un tetramère homologue aux tetramères H3-H4 eucaryotes et sont de ce fait plus flexibles et moins stables (Sandman et Reeve, 2005). Ainsi, la présence d'histones chez les archaea ralentirait mais ne stopperait pas l'élongation de la transcription contrairement aux nucléosomes eucaryotes dont la structure doit être modifiée par acétylation pour que la transcription soit facilitée (Xie et Reeve, 2004). D'ailleurs, les modifications post-traductionnelles des histones eucaryotes impliquées dans de nombreuses régulations sont localisées dans des régions généralement absentes dans les histones d'archaea (Sandman et Reeve, 2006). Jusqu'à présent aucune modification post-traductionnelle n'a été détectée dans ces dernières (Forbes *et al.*, 2004). Il n'existerait donc pas, pour les histones d'archaea, de mode de régulation similaire à celui des histones eucaryotes.

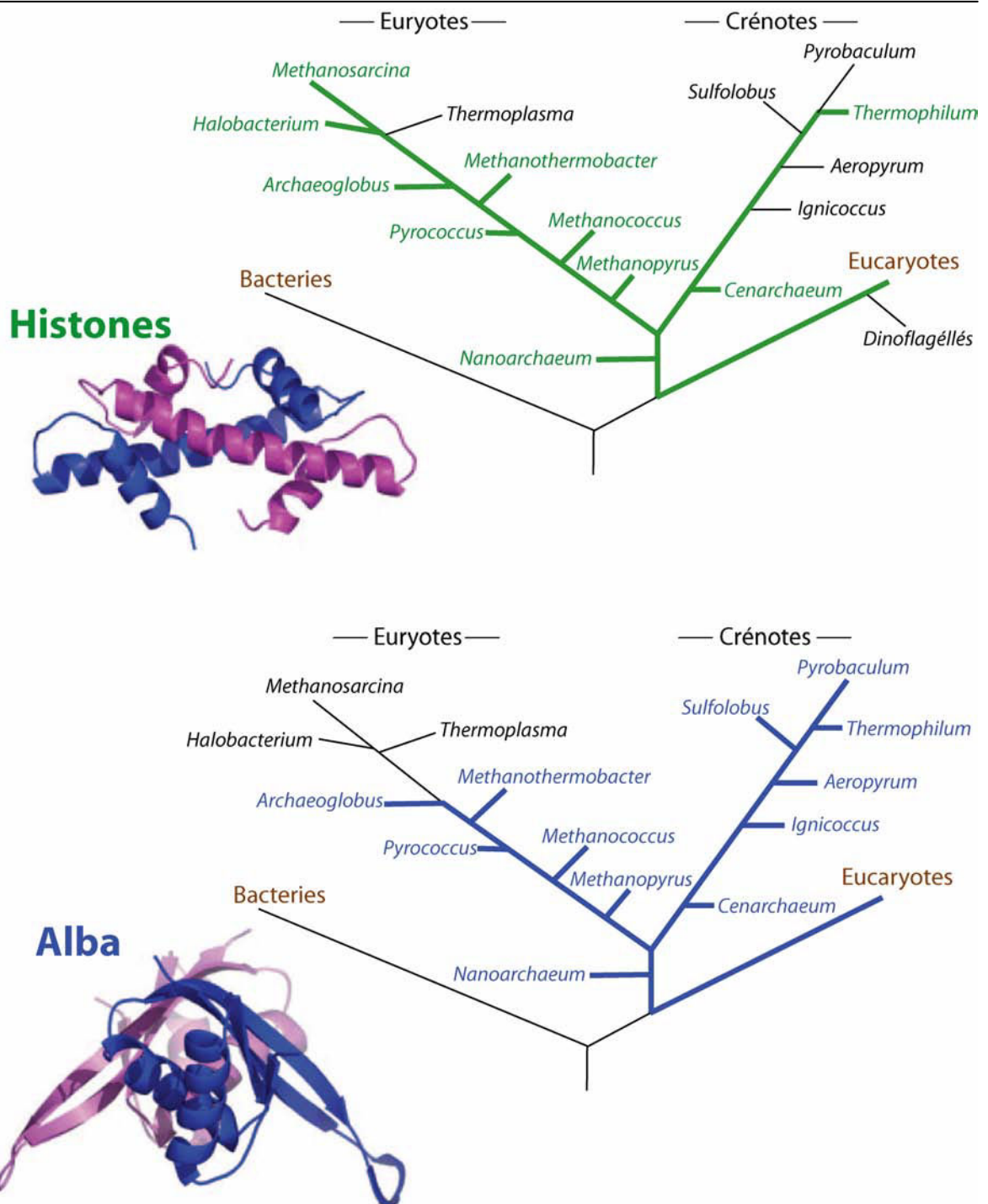


Figure I.9 : Distribution des protéines histones et Alba sur un arbre universel du vivant basé sur la comparaison des ARN ribosomiques (d'après Sandman et Reeve, 2005, 2006). Pour simplification les longueurs des branches ne représentent pas des distances évolutives. Les lignes épaisses relient les génomes dans lesquels les homologues des histones (A) ou des protéines Alba (B) sont présents. Les coordonnées PDB des structures 3D présentées sont les suivantes : histones de *Methanothermobacter* HmfB, 1BFM ; protéine Alba de *Archaeoglobus*, 1H0Y.

La famille Alba (Acetylation lower binding affinity) regroupe plusieurs protéines dont DBNP-B, Sac10b, Sso10b et Ssh10b (Dijk et Reinhardt, 1986 ; Reddy et Suryanarayana, 1989 ; Guo *et al.*, 2003). Ses représentants forment des dimères capables de fixer l'ADN et l'ARN (Guo *et al.*, 2003). Ils sont présents chez quelques eucaryotes et tous les archaea à l'exception des

Halobacterales et des Methanomicrobiales (Fig. I.9). Ces organismes présentent un à deux homologues de Alba (Alba1 et Alba2) permettant la formation d'homo- ou d'hétéro-dimères. Lorsque le rapport Alba/ADN est faible, des homodimères d'Alba 1 ou 2 sont liés à l'ADN en formant des structures de type filaments alors que lorsque ce rapport est élevé apparaissent des hétérodimères Alba1-Alba2 se fixant à l'ADN en formant des structures plus compactes et branchées (Jelinska *et al.*, 2005). L'équilibre entre les deux formes paraît donc être un premier niveau de régulation. Comme indiqué par le nom, les protéines Alba, comme les histones eucaryotes, sont susceptibles d'être acétylées, ce qui réduirait leur affinité pour l'ADN et/ou modulerait les interactions entre protéines Alba 1 et/ou Alba 2 (Bell *et al.*, 2002 ; Zhao *et al.*, 2003 ; Jelinska *et al.*, 2005). Cependant, ce mode de régulation reste encore peu compris.

3. Réparation de l'ADN

Bien que se multipliant dans des environnements extrêmes, les taux d'erreurs observés dans les génomes d'archaea sont tout à fait comparables à ceux observés dans les autres micro-organismes (Grogan *et al.*, 2001). Cependant, très peu d'informations concernant le maintien de l'intégrité des séquences génomiques d'archaea sont disponibles. Certaines protéines impliquées dans le mécanisme de base de la réplication jouent un rôle dans la réparation. Ainsi, les protéines SSB peuvent reconnaître des dommages de l'ADN (Cubeddu et White, 2005), certaines ADN polymérases peuvent reconnaître et cliver des nucléotides mésappariés (Jokela *et al.*, 2004) ou s'arrêter sur les résidus uracile (Connolly *et al.*, 2003). Le trimère PCNA peut recruter des protéines impliquées dans la réparation (Warbrick, 2000) et la primase d'archaea pourrait également jouer un rôle dans la réparation (Lao-Sirieix *et al.*, 2005). Les archaea ne présentent pas les gènes principaux des voies MutS/MutL et NER (Nucleotide Excision Repair) respectivement impliqués dans la réparation des mésappariements et dans le retrait des lésions de l'ADN chez les bactéries, ce qui suggère la présence de voies de réparation différentes (pour revues, (Grogan, 2004; Kelman et White, 2005). Il a été démontré, par exemple que, lorsque les irradiations aux rayons gamma produisent des cassures double-brin (DSB) dans l'ADN chromosomique de *P. abyssi*, celui-ci stoppe sa réplication et exporte l'ADN endommagé hors de la cellule, préservant ainsi l'accumulation d'erreurs (Jolivet *et al.*, 2003b). C'est un exemple d'adaptation à des conditions de vie très particulières dont le mécanisme n'est pas encore caractérisé.

C. Transcription

1. La région promotrice

Les Archaea n'utilisent qu'une ARN polymérase (ARN Pol) qui comporte entre 10 et 14 sous-unités selon l'espèce (Langer *et al.*, 1995 ; Soppa, 1999b). La plupart des sous-unités sont des homologues des sous-unités de l'ARN Pol eucaryote de type II (les ARN Pol eucaryotes de type I, II et III sont dédiées aux ARNr, ARNm et la plupart des UsnRNA, ARNt et autres sRNA,

respectivement). Seulement 4 de ces sous-unités sont des homologues des sous-unités $\alpha_2\beta\beta'\omega$ de l'ARN Pol bactérienne (pour revue, Bartlett, 2005). Contrairement aux bactéries pour lesquelles l'ARN Pol s'associe à un des différents facteurs σ , ce qui assure la reconnaissance de différents promoteurs, l'initiation de la transcription chez les eucaryotes et les archaea nécessite des facteurs extrinsèques (Murakami *et al.*, 2002a; Murakami *et al.*, 2002b ; Vassilyev *et al.*, 2002). En contraste avec les bactéries, l'initiation de la transcription chez les archaea comme chez les eucaryotes dépend donc de l'interaction de l'ARN Pol avec des facteurs transcriptionnels ou des protéines de régulation déjà liées à l'ADN (Soppa, 1999b). Chez les archaea, l'initiation requiert deux facteurs généraux de la transcription : TBP (TATA-box Binding Protein) et TFB respectivement équivalant aux facteurs TBP et TFIIB des eucaryotes (Hausner *et al.*, 1996). Le facteur TBP reconnaît la boîte TATA et provoque une courbure qui permet notamment le recrutement de TFB (Fig. I.10 ; Rowlands *et al.*, 1994; Littlefield *et al.*, 1999). TFB se lie à TBP et à l'ADN au niveau d'un élément appelé BRE (TFB Receptor Element). Il joue un rôle important dans le recrutement de l'ARN Pol en contactant certaines de ses sous-unités (Qureshi *et al.*, 1995 ; Littlefield *et al.*, 1999 ; Magill *et al.*, 2001 ; Bartlett, 2005). Le complexe TBP-TFB-ADN recrute ainsi l'ARN Pol. Il semblerait également que TFB contacte l'ADN à proximité voire à quelques nucléotides en amont du site +1 de début de la transcription (Fig. I.10 ; Bartlett *et al.*, 2004 ; Renfrow *et al.*, 2004). Les deux facteurs assurent donc la reconnaissance du promoteur en définissant le site +1 mais aussi la fréquence de l'initiation. Un troisième facteur, TFE, homologue de la sous-unité α de TFIIE des eucaryotes, a été identifié dans la plupart des génomes d'archaea et, même s'il ne paraît pas indispensable, il semble stimuler la transcription en facilitant ou en stabilisant les interactions entre TBP et une boîte TATA sous-optimale (Bell *et al.*, 2001; Hanzelka *et al.*, 2001). Aucun facteur similaire à TFIH, TFIIA ou TFIID, ni aucun homologue des TAF (TBP Associated Factor) eucaryotes n'a été détecté chez les Archaea (pour revue, Soppa, 1999b).

Les promoteurs d'archaea sont composés de deux éléments principaux, une boîte TATA (parfois appelé boîte A), un élément BRE, ainsi que deux éléments moins étudiés, IE (Initiator Element, parfois appelé boîte B) et un élément PPE (Proximal Promoter Element) (Fig. I.10). La boîte TATA localisée environ 25 pb en amont du site +1 est une séquence de 5 à 8 pb riche en A/T. L'élément BRE qui est situé directement en amont de la boîte TATA est composé de 4 à 7 pb. Il influencerait l'orientation de la transcription (Qureshi et Jackson, 1998). L'élément IE localisé au niveau du site +1 contient un dinucléotide Py-Pu avec initiation au niveau du résidu Pu (Palmer et Daniels, 1995 ; Soppa, 1999a). L'élément PPE localisé entre le site +1 et 12 pb en amont du site +1 est riche en A/T (Soppa, 1999a). Cependant, les éléments IE et PPE pourraient ne pas être universels et n'être présents que dans certains sous-groupes de promoteurs. De même, des éléments de séquence ADN supplémentaires pourraient interagir avec des protéines régulatrices (Kessler *et al.*, 2004).

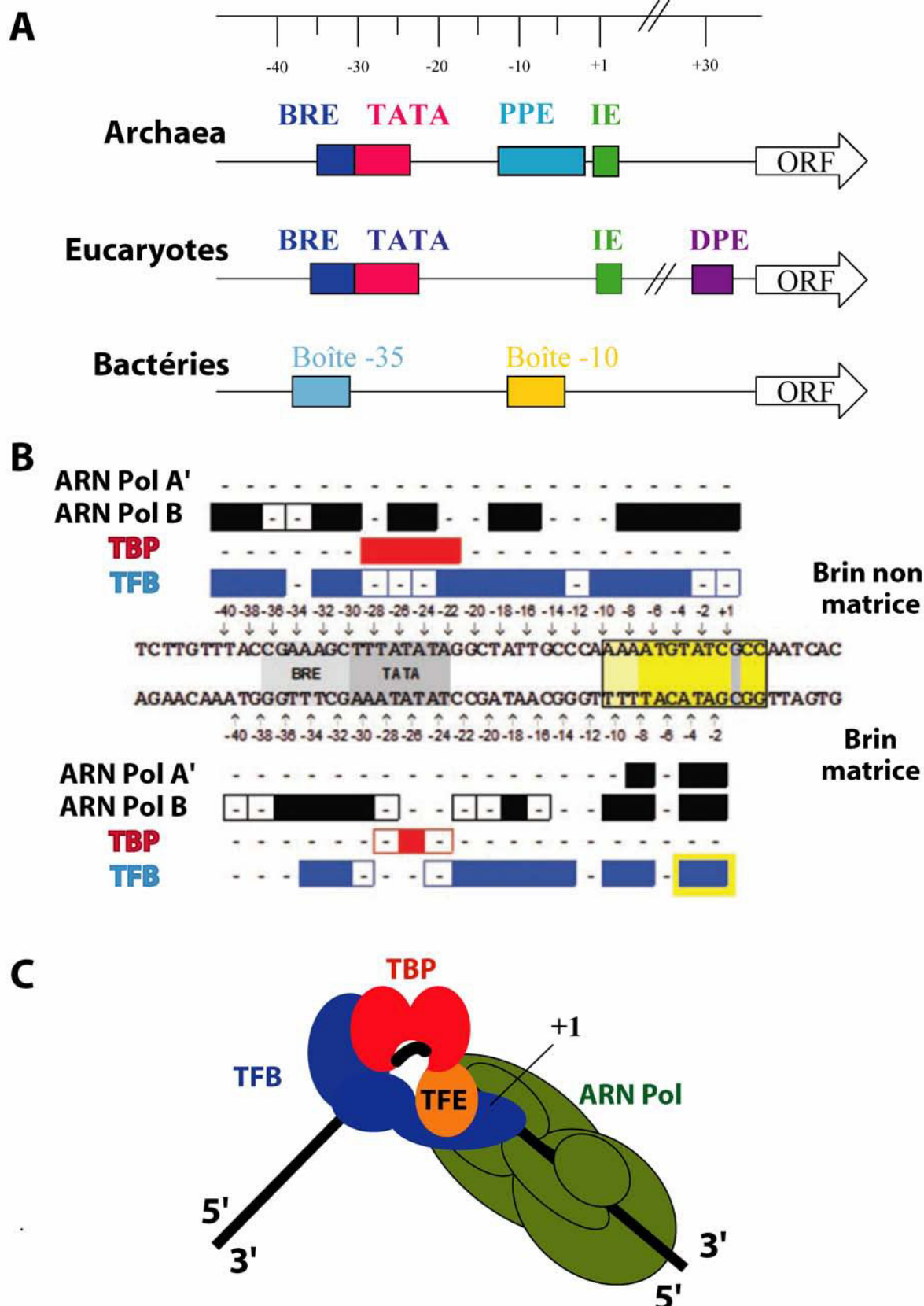


Figure I.10 : L'initiation de la transcription chez les Archaea.

A. Comparaison des promoteurs d'archaea, d'eucaryotes et de bactéries (adapté de Soppe, 1999). Représentation schématique des éléments composant les promoteurs pour chacun des trois domaines du vivant. Les éléments sont alignés avec l'échelle représentée ci-dessus. Le site +1 représente le point d'initiation de la transcription. BRE : TFB Recognition Element. PPE : Proximal Promoter Element. IE : Initiator Element. DPE : Downstream Element en interaction avec les TAF (TBP Associated Factors). ORF : Open Reading Frame. **B.**

Résultats des expériences de pontage covalent aux UV réalisées chez *P. furiosus* avec le complexe TBP-TFB-ARN Pol (d'après Renfrow *et al.*, 2004). Les phosphates analysés sont représentés par des flèches. Les pontages observés sont indiqués par des barres pleines ou vides selon qu'ils sont respectivement d'intensité forte ou faible. C. Représentation schématique du complexe TBP-TFB-ARN Pol-TFE sur le promoteur d'un archaea (adapté de Soppa, 1999). Le positionnement de TFE est incertain. L'ARN Pol se déplace rapidement lors de l'initiation, les connaissances sur son positionnement initial sont peu précises.

Le mécanisme de transcription « basale » des Archaea a donc de nombreux points communs avec celui des eucaryotes mais paraît moins complexe. Par ailleurs, les protéines intervenant dans la régulation de la transcription sont le plus généralement homologues à celles des bactéries (Bell et Jackson, 2001 ; Bell, 2005 ; pour revue Geiduschek et Ouhammouch, 2005) qu'à celles des eucaryotes (par exemple MBF1 ; Aravind et Koonin, 1999 ; Kyrpides et Ouzounis, 1999). Les protéines régulatrices se lient à l'ADN et favorisent ou empêchent le recrutement des facteurs généraux TFB, TBP et de l'ARN Pol. Certains activateurs de la famille Lpr peuvent, par exemple, s'associer à la région promotrice de l'ADN via un site UAS (Upstream Activator Site) (Ouhammouch et Geiduschek, 2005). Les protéines régulatrices peuvent également modifier la conformation de l'ADN. Un super-enroulement négatif stimule, par exemple, l'initiation de la transcription (Bell et Jackson, 1998 ; Hethke *et al.*, 1999). En outre, les protéines de la chromatine peuvent inhiber l'initiation de la transcription si les nucléosomes sont associés aux régions promotrices et les mécanismes de remodelage de la chromatine sont encore peu connus chez les archaea (Xie et Reeve, 2004 ; Geiduschek et Ouhammouch, 2005).

Sur la base de ce que l'on sait sur la conservation des éléments de séquence de la région promotrice des gènes d'Archaea, des algorithmes combinant la recherche de signaux potentiels de transcription et la présence d'ORF ont été créés pour identifier les unités transcriptionnelles (Suckow *et al.*, 1998 ; Soppa, 1999a). Cependant, le degré de conservation des séquences des promoteurs est trop faible pour qu'on puisse distinguer les sites +1. En effet, non seulement la boîte TATA ne requiert pas une stricte conservation de séquence mais, en plus, la position de cet élément par rapport au site +1 est variable (plus ou moins 2 nts). D'autre part, une analyse des signaux localisés en amont des ORF des génomes d'archaea disponibles démontre que les séquences consensus établies pour les éléments TATA, BRE et IE sont différentes d'un sous-groupe d'archaea à un autre (Soppa, 1999a ; Torarinsson *et al.*, 2005). Les différences dans le nombre d'éléments constituant les promoteurs, leurs positions relatives et leurs séquences qui sont observées entre sous-groupes d'Archaea sont peut-être la conséquence de l'utilisation de plusieurs variants de TBP et TFB, comme ceux identifiés chez certains méthanogènes et Halobacterales. Ces variants pourraient former des complexes alternatifs ayant des spécificités de reconnaissance différentes (Baliga *et al.*, 2000 ; Galagan *et al.*, 2002 ; Santangelo *et al.*, 2007). Une autre possibilité est l'utilisation dans certains cas des protéines SSB au lieu de TBP, comme cela a été suggéré récemment pour *Sulfolobus solfataricus* (Richard *et al.*, 2004). On peut aussi imaginer la présence de facteurs transcriptionnels spécifiques encore non identifiés (pour revues, Soppa, 1999b; Geiduschek et Ouhammouch, 2005).

Il est par conséquent très difficile d'identifier les unités transcriptionnelles d'un génome d'Archaea par recherche d'une séquence promotrice particulière. Ceci limite actuellement la recherche de gènes d'ARNnc pour lesquels aucun signal de transcription ne peut être utilisé.

2. L'élongation et la terminaison de la transcription

Les événements qui suivent l'initiation de la transcription chez les archaea ont fait l'objet de peu d'études. L'extrémité C-ter de la sous-unité RPBI de l'ARN Pol II eucaryote est phosphorylée lors de la transition de l'initiation vers l'élongation et elle recrute également les facteurs de terminaison ; or, la séquence correspondant à cette extrémité est absente dans la sous-unité équivalente (A) de l'ARN Pol d'archaea (Stiller et Hall, 2002). L'ARN Pol II eucaryote utilise plusieurs facteurs d'élongation et de terminaison. Son homologue d'archaea qui n'est pas bloqué en cours d'élongation par les nucléosomes (Xie et Reeve, 2004), serait assisté par un seul facteur d'élongation, TFS (Hausner *et al.*, 2000) et ne ferait appel à aucun facteur de terminaison (Muller *et al.*, 1985 ; Thomm *et al.*, 1994). Chez la plupart des bactéries, le motif qui paraît nécessaire et suffisant pour une terminaison ρ -indépendante (intrinsèque) de la transcription est une structure secondaire en tige-boucle située dans les 50 nt en aval du codon stop. Cette structure est suivie d'un enchaînement plus ou moins long et régulier de résidus T (U dans l'ARN). Une analyse de la distribution dans des génomes d'archaea de ce motif ayant cette double propriété, tige-boucle et séquence riche en U, suggère que ce motif n'est sans doute pas utilisé pour la terminaison de la transcription chez ces organismes (Ermolaeva *et al.*, 2000 ; Unniraman *et al.*, 2001). Les données actuelles montrent que la terminaison de la transcription chez les archaea dépend essentiellement de la présence de séquences riches en T (U dans l'ARN), que la présence de structures en tige-boucle stimule la terminaison mais n'est pas nécessaire et, enfin, que des facteurs de terminaison similaires au facteur de terminaison intrinsèque des bactéries, tels que des homologues bactériens de NusA et NusG et le facteur typiquement archaea TFS pourraient être impliqués (Lange et Hausner, 2004 ; Santangelo et Reeve, 2006 ; Shibata *et al.*, 2007).

Récemment il a été démontré que la polyadénylation joue un rôle dans la stabilité et le taux de traduction des ARNm bactériens (Joanny *et al.*, 2007). Chez les Archaea comme chez les Bactéries, une polyadénylation transitoire des ARNm stimule la dégradation d'ARNm présentant de fortes structurations secondaires en permettant le recrutement d'exoribonucléases (Portnoy *et al.*, 2005 ; Portnoy et Schuster, 2006). Cette dégradation est réalisée par une activité associée aux exosomes. Cependant, aucune polyadénylation ni aucun gène correspondant à des exosomes n'a été observé chez les halophiles et certains méthanogènes, suggérant l'implication de ribonucléases particulières (Portnoy et Schuster, 2006).

D. Traduction

Bien que longtemps envisagé, ce n'est que récemment que le couplage entre la transcription et la traduction a été prouvé chez les Archaea (French *et al.*, 2007). En effet, en microscopie

électronique, les polysomes sont observés à proximité de l'ADN et en sont rarement éloignés. Ainsi, la traduction débute avant la fin de la transcription voire dès le début de la transcription. Il est donc très probable que, comme pour les bactéries, les polysomes influencent la fin de la transcription et de ce fait l'expression des gènes distaux d'opérons (Pan et Sosnick, 2006).

Comme pour la transcription, la traduction chez les Archaea a des caractéristiques communes avec les Eucaryotes et les Bactéries, ce qui correspond sans doute à un mécanisme hybride (Saito et Tomita, 1999). De nombreux ARNm d'archaea comportent une séquence Shine-Dalgarno (SD), comme c'est le cas chez les bactéries, le plus souvent complémentaire avec 4 nt de l'extrémité 3' de l'ARNr 16S (UCACCUCC) (Shine et Dalgarno, 1974 ; Torarinsson *et al.*, 2005). Les séquences SD sont centrées environ 10 nts en amont du codon d'initiation de la traduction.

Il est possible de distinguer deux types de gènes. Les gènes de type I correspondent aux gènes monocistroniques et aux gènes transcrits les premiers dans les opérons, alors que les gènes de type II correspondent aux autres gènes en aval dans les opérons. Les gènes de type I pourraient être traduits par un mécanisme qui ne ferait pas intervenir de signal SD. La reconnaissance se ferait directement au niveau du codon initiateur ou quelques nt en amont de celui-ci. Pour certains de ces gènes la distance séparant la boîte TATA du codon d'initiation peut être inférieure à 30 nts, ce qui suggère l'absence de région 5' UTR (UnTranslated Region) et donc de séquence SD (Slupska *et al.*, 2001). Il a également été démontré que la délétion de la séquence SD abolit complètement la traduction mais que celle-ci peut être rétablie par la délétion de la région 5' UTR suggérant un mécanisme utilisé par défaut qui ne requiert pas d'interaction spécifique entre ARNm et ARNr 16S (Condo *et al.*, 1999). L'existence de ce type d'initiation sans recours aux séquences 5'UTR a également été mise en évidence chez les eucaryotes (Saito et Tomita, 1999). Par ailleurs, il est possible de distinguer deux groupes d'archaea se comportant différemment (Torarinsson *et al.*, 2005). Le groupe A comprend des espèces pour lesquelles les gènes de type I présentent, dans environ 50%, des cas une boîte TATA mais pas de séquence SD, et les gènes de type II qui présentent, dans environ 50 % des cas une séquence SD (mais sans région promotrice). Le groupe B comprend des espèces pour lesquelles les gènes de type I présentent peu de transcrits « sans séquence leader », la majorité des gènes de type I et II renfermant une séquence SD. C'est aussi dans ce groupe B que les taux les plus élevés de codons initiateurs alternatifs (UUG et GUG) sont observés. Il n'y a pas de corrélation entre la distribution des espèces au sein de ces groupes et leur positionnement phylogénétique, cependant, il a été remarqué que les espèces appartenant au groupe B semblent être apparues plus tôt dans l'évolution du domaine archaea (Fig. I.11). Ces organismes sont le plus souvent des hyperthermophiles (Stetter, 2006). Une corrélation entre la température de croissance et le temps de doublement est aussi observée (Torarinsson *et al.*, 2005). Ainsi les hyperthermophiles sont ceux pour lesquels le cycle cellulaire est le plus rapide (environ 30 min contre 600 min pour les Méthanogènes mésophiles) et sont aussi ceux qui ont le plus recours aux séquences SD qui accélèreraient le décodage des gènes polycistroniques (Londei, 2005). La mise en place de la

traduction sans séquence SD pourrait être apparue plus tardivement chez les archaea, comme suggéré par la figure I.11, (Torarinsson *et al.*, 2005). Le recours aux séquences SD pourrait être une adaptation à une lecture des gènes polycistroniques qui auraient pu apparaître très tôt dans l'évolution (Londei, 2005). Les eucaryotes auraient systématiquement perdu cette caractéristique lors de l'apparition de l'enveloppe nucléaire engendrant une absence de couplage entre transcription et traduction.

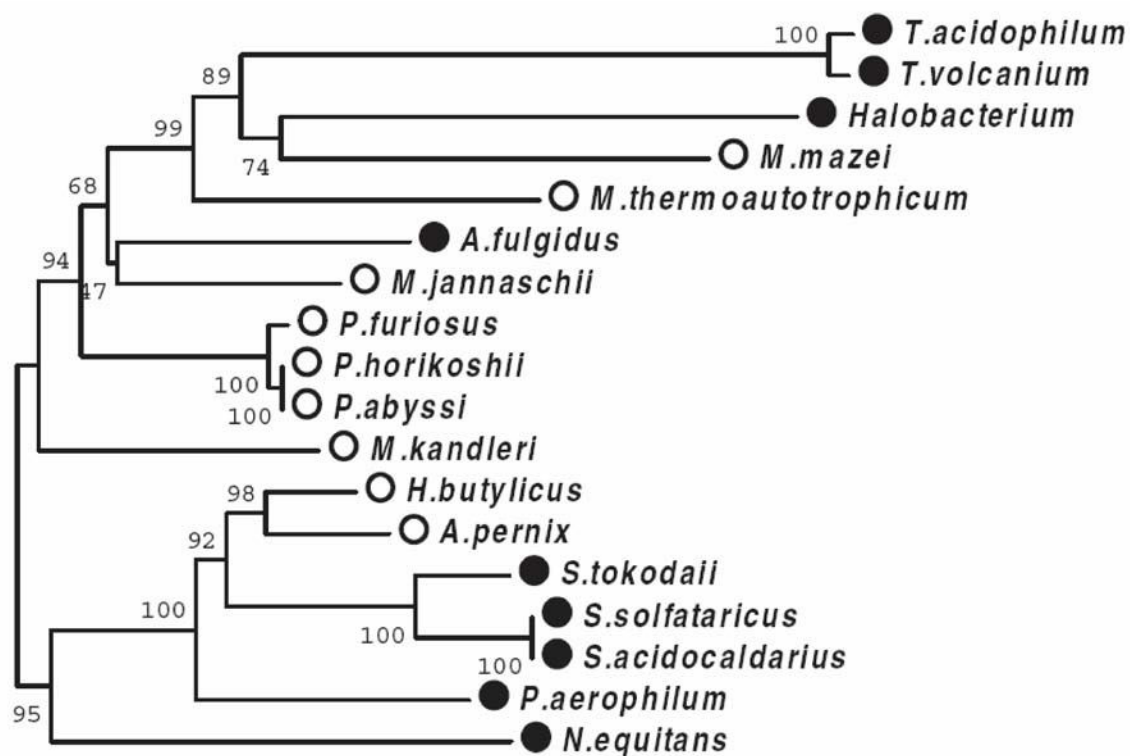


Figure I.11 : Séparation entre espèces du groupe A (cercles pleins) et espèces du groupe B (cercles vides) représentée sur un arbre phylogénétique construit à partir des séquences des ARNr 16S (d'après Torarinsson *et al.*, 2006). Les espèces du groupe A sont caractérisées par l'absence de séquence SD (transcrits dits « sans leader ») pour les gènes de type I, et par la présence de séquence SD pour les gènes de type II. Les espèces du groupe B sont caractérisées par la présence de séquences SD pour les gènes de type I et II soit un faible taux de transcrit sans leader, ainsi que par un taux d'utilisation de codons d'initiation alternatifs plus élevé.

Les facteurs d'initiation de la traduction ont été recherchés par similarité de séquence avec ceux des bactéries et des eucaryotes. Quatre des 6 facteurs archaea sont universels, les deux autres sont partagés avec les eucaryotes. Il n'y a donc pas, à ce jour, de facteur propre aux archaea (pour revue, Bell et Jackson, 1998 ; Londei, 2005). Les facteurs d'initiation de la traduction des eucaryotes absents chez les archaea sont ceux supposés interagir avec la coiffe en 5' des ARNm (eIF4F et eIF3), coiffe absente dans les ARNm d'archaea (Tableau I.5). Chez les bactéries, l'initiation de la traduction dépend de la reconnaissance des codons d'initiation (en général AUG) par trois facteurs IF1, IF2 et IF3 (Tableau I.5 ; Boelens et Gualerzi, 2002). IF1 et IF2 occupent le site A du ribosome et recrutent l'ARNt-formyl-^{Met} dans le site P jouant ainsi un rôle central dans l'initiation de la traduction (Roll-Mecak *et al.*, 2001 ; Marintchev *et al.*, 2003). Ces deux facteurs présentent des homologues dans les deux autres domaines du vivant : aIF1A et

aIF2 chez les archaea, et eIF1A et eIF5B chez les eucaryotes (Tableau I.5). Chez les eucaryotes les facteurs eIF2, eIF2B et eIF5 sont aussi impliqués. eIF2, sous sa forme liée au GTP, se lie à l'ARNt initiateur, eIF5 présente l'activité GTPase qui permet la dissociation de eIF2 après fixation de l'ARNt^{Met} à la sous-unité 40S et eIF2B permet de réactiver eIF2 (Kimball, 1999). Le chargement de l'ARNt initiateur sur la petite sous-unité 40S permet au pré-ribosome de scanner la région 5'UTR de l'ARNm pour rechercher le codon initiateur. eIF1 assure la fidélité de l'association de la sous-unité 40S avec le codon initiateur (Pestova et Kolupaeva, 2002). Quand ce dernier est reconnu, le facteur eIF5B est requis pour favoriser l'association des deux sous-unités en ribosomes 80S (Pestova *et al.*, 2000). eIF5A stimule la formation de la première liaison peptidique (Aoki *et al.*, 1997). Enfin, eIF6 interagit avec la sous-unité 60S et empêche de ce fait l'association avec la petite sous-unité. Son élimination requiert sa phosphorylation qui est régulée par de nombreuses cascades de kinases (Ceci *et al.*, 2003). Les archaea présentent des homologues des facteurs eIF2, eIF2B, eIF1, eIF5B, eIF5A et eIF6 suggérant un mode d'initiation similaire à celui des eucaryotes (Tableau I.5).

Bactéries	Archaea	Eucaryotes
Facteurs Universels		
IF1 Stimule IF2	aIF1A Fonction Inconnue	eIF1A Assiste le criblage du codon initiateur
IF2 Lie l'ARNt Met formylé	aIF2/aIF5B stimule l'association de l'ARNt Met initiateur sur le ribosome	eIF5B stimule l'association des sous-unités du ribosome
YCiH (absents de certains phyla) Fonction Inconnue	aSUI1 Fonction Inconnue	eIF1/SUI1 Facteur de fidélité
EFP Stimule la première liaison peptidique	aIF5A Fonction Inconnue	eIF5A Stimule la première liaison peptidique
Facteurs Spécifiques		
IF3 Assure la fidélité de IF1 et IF2	-	-
	-	eIF4F (trimer) lie la Cap et déplie l'ARNm
	-	eIF3 (octamer) intéragit avec eIF4F et le ribosome
	a/eIF2 (trimer) Lie l'ARNt Met initiateur	eIF2 (trimer) Lie l'ARNt Met initiateur
	sous-unités , et de eIF2B Fonction Inconnu	eIF2B (pentamer) facteur d'échange de Guanine
	-	eIF5 GTPase de eIF2
	aIF6 Fonction Inconnue	eIF6 inhibe l'association des sous-unités du ribosome

Tableau I.5 : Les facteurs d'initiation de la traduction identifiés dans chacun des trois domaines du vivant (d'après Londei, 2005).

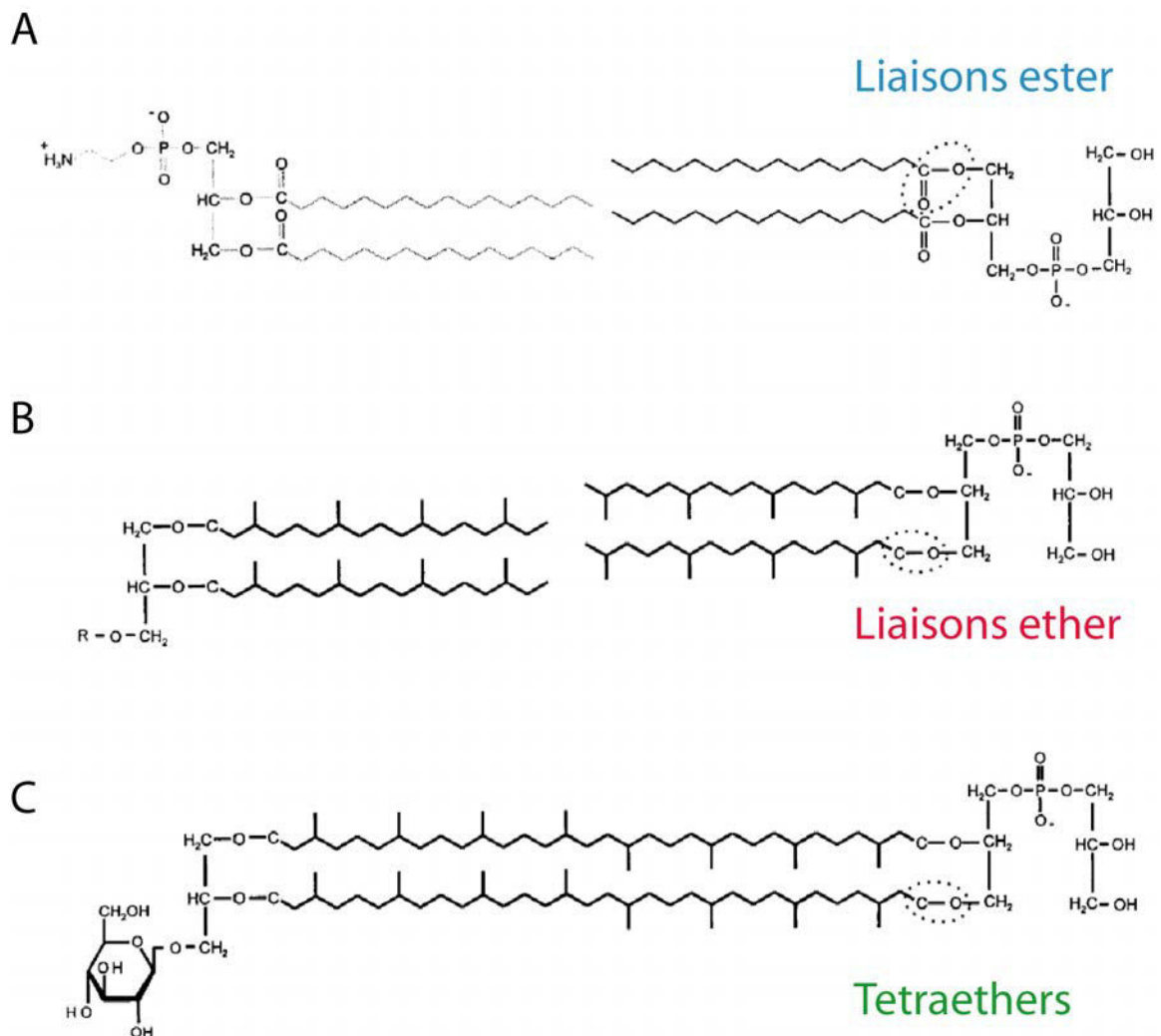
L'élongation et la terminaison de la traduction archaea nécessitent d'autres facteurs ressemblant aux facteurs eucaryotes (Inagaki et Ford Doolittle, 2000). Ainsi un facteur homologue à eEF1 α eucaryote permettrait de recruter l'ARNt dans le site A vacant des ribosomes et la translocation du ribosome sur l'ARNm serait assurée par un homologue de eEF2 (Bell et Jackson, 1998). Enfin, un seul facteur, un homologue du facteur eRF (Release Factor) eucaryote, serait responsable de la reconnaissance des trois codons stop et de l'hydrolyse de la liaison entre le peptide et le dernier ARNt (Dennis, 1997). L'initiation, l'élongation et la terminaison de la traduction chez les Archaea sont donc plus similaires aux mécanismes eucaryotes qu'aux mécanismes bactériens.

E. Membranes

Les membranes sont composées de monocouches ou bicouches phospholipidiques auxquelles sont associées de nombreuses protéines. La barrière ainsi formée entre cytoplasme et milieu extérieur permet notamment de contrôler les concentrations de solutés et d'ions, essentielles au métabolisme, à l'intérieur de la cellule. Des protéines membranaires dont les ORF représenteraient jusqu'à 20% de l'ensemble des ORF d'un génome procaryote participent au maintien de gradients de concentration de certains solutés et ions. Un grand nombre d'entre elles sont de type ATP-Binding Cassette (ABC) et consomment donc de l'énergie. Puisqu'une perméabilité trop importante engendre une dépense d'énergie supplémentaire pour maintenir ces gradients, il est primordial pour la cellule que la composition lipidique de sa membrane soit adaptée aux conditions du milieu et puisse être ajustée en fonction des variations de ce milieu (Konings *et al.*, 2002).

La composition de la membrane des archaea a été très tôt utilisée pour les distinguer des deux autres domaines du vivant. Les phospholipides bactériens et eucaryotes sont composés de deux acides gras associés au glycérol par liaison ester. Le troisième groupement hydroxyl du glycérol est associé à un phosphate souvent lui-même associé à des glucides (Fig. I.12). En contraste, les archaea présentent, à la place des deux acides gras, deux chaînes isoprénoïdes saturées associées par liaison éther au glycérol, ou à d'autres alcools tels que le nonitol, ceci dans une configuration stéréochimique unique aux archaea (Fig. I.12 ; DeRosa *et al.*, 1991). Les archaea évoluant dans des conditions dites modérées présentent des bicouches de chaînes isoprénoïdes en C₂₀ alors que des chaînes tetraethers en C₄₀ forment une monocouche chez les thermophiles et les acidophiles (Fig. I.12 ; DeRosa *et al.*, 1991 ; Kates, 1996). Dans cette dernière monocouche, les faces polaires sont donc connectées entre elles et ne peuvent pas être séparées (Choquet *et al.*, 1992 ; Elferink *et al.*, 1992). En raison d'un degré de rotation plus restreint des carbones des chaînes apolaires, les membranes composées de liaisons éthers sont plus stables que celles composées de liaisons esters (Elferink *et al.*, 1992 ; Thompson *et al.*, 1992). La stabilité des chaînes apolaires réduit ainsi la perméabilité des membranes d'archaea.

L'augmentation de la température provoque une augmentation irrémédiable de la fluidité membranaire qui peut entraîner une augmentation de la perméabilité aux protons et aux ions et ainsi une incapacité à maintenir une force proto-motrice suffisante (Russell et Fukunaga, 1990).



A. Les membranes des bactéries sont composées de bicouches de phospholipides reposant sur l'association du glycérol avec des acides gras par le biais de liaisons esters. **B.** Les phospholipides d'archaea sont composés de glycérol associé à des chaînes isoprénoides par des liaisons éthers. **C.** Les membranes archaea sont composées par une fraction variable de monocouche de phospholipides de type « tetraethers ».

Enfin, la plupart des archaea présente une couche de surface ou S-layer composée de glycoprotéines qui, contrairement à ce qui est observé chez les bactéries, est directement ancrée dans la membrane cytoplasmique et ne présente pas de muréine (Kandler et König, 1998). *Ignicoccus* est le seul archaea qui présente une membrane externe supplémentaire décrivant ainsi un espace périplasmique (Rachel *et al.*, 2002).

F. Pathogénicité

Compte tenu de l'ubiquité des archaea, de leur abondance évaluée à près de 20% de la biomasse sur Terre et de la présence de certains de ses représentants dans le tractus digestif de l'Homme et des autres animaux, il est objectivement pertinent de s'interroger quant à un rôle potentiel des archaea dans les pathologies humaines et animales. Plusieurs pré-requis semblent communs aux organismes pathogènes : ils sont capables de coloniser l'hôte i.e. interagir avec lui en évitant les réponses de son système de défense, ils provoquent des dommages par le biais de toxines et/ou par le biais de la réponse inflammatoire de l'hôte et ces dernières caractéristiques sont matérialisées par la présence de gènes de virulence et/ou de toxines (pour revue, Cavicchioli *et al.*, 2003; Eckburg *et al.*, 2003). Les archaea sont effectivement capables d'interagir avec les eucaryotes puisqu'ils sont détectés au sein de ciliés présents dans le rumen de mammifères, dans les intestins de certains arthropodes et mammifères où ils sont acquis par ingestion et maintenus durant la division cellulaire (Vogels *et al.*, 1980 ; van Hoek *et al.*, 2000), ou encore en symbiose avec une algue comme déjà évoqué (*C. symbiosum* ; Preston *et al.*, 1996). Les archaea sont doués de mobilité via leurs flagelles (pour revue, Jarrell *et al.*, 1996) et présentent des protéines d'adhésion semblables à celles de certains pathogènes bactériens (Kachlany *et al.*, 2000). Des toxines appelées archaeocines, uniques aux archaea ou ressemblant aux équivalents bactériens, ont été identifiées chez les Halophiles et les Sulfolobales (Prangishvili *et al.*, 2000; O'Connor et Shand, 2002). Ces toxines seraient dirigées contre d'autres espèces d'archaea (Kis-Papo et Oren, 2000 ; Haseltine *et al.*, 2001) et il est possible que certaines soient dirigées contre des espèces eucaryotes. Les gènes de virulence et de toxines sont regroupés au sein d'îlot de pathogénicité chez les bactéries, or, compte tenu des transferts latéraux déjà observés et de la proximité écologique entre archaea et bactéries, un échange reste envisageable entre espèces des deux domaines. Enfin, les lipides membranaires uniques aux archaea stimulent l'expression du complexe majeur d'histocompatibilité et provoquent ainsi une forte réaction inflammatoire en recrutant les cellules présentatrices d'antigènes à l'instar des liposaccharides bactériens (Krishnan *et al.*, 2001).

Trois principaux méthanogènes, *Methanobrevibacter smithii*, *Methanobrevibacter oralis* et *Methanosphaera stadtmanae* sont connus pour coloniser le colon, le vagin et la cavité orale de l'Homme (Karlin *et al.*, 1982 ; Miller et Wolin, 1982 ; Belay *et al.*, 1988 ; Belay *et al.*, 1990). Bien que la présence de certains d'entre eux soit parfois corrélée au développement de certaines pathologies, aucun de ces méthanogènes n'a été décrit comme étant l'agent responsable de ces pathologies (Cavicchioli *et al.*, 2003). Ainsi, les archaea peuvent potentiellement être des

pathogènes mais seules deux implications dans des procédés pathogéniques ont récemment été décrites.

Une implication d'une espèce d'archaea encore non définie en tant que co-facteur de déclenchement de l'athérosclérose a été proposée (Higuchi *et al.*, 2006). La formation de plaques d'athérome est initiée par un dépôt lipidique, riche en cholestérol, formé au début du processus et qui donne son nom à la maladie. Il s'agit de la première cause de mortalité dans les pays industrialisés. L'athérosclérose affecte l'aorte, les artères coronaires, carotides, fémorales, rénales et les artères des membres inférieurs et évolue silencieusement pendant plusieurs décennies avant d'être mise en évidence à l'occasion d'un accident grave, tel que infarctus du myocarde, embolie pulmonaire, accident vasculaire cérébral... Dès l'enfance, des dépôts graisseux s'accumulent sur la tunique interne de la paroi artérielle (l'intima) et entraînent la formation d'une plaque fibreuse. Celle-ci s'étend progressivement à la tunique intermédiaire (média), en provoquant la migration de cellules musculaires lisses vers la plaque qui augmente de volume et devient fibro-lipidique. La plaque fait alors davantage saillie dans la lumière artérielle et provoque un risque de sténose (rétrécissement du diamètre de l'artère) et d'ischémie (arrêt de la circulation). La plaque évolue en durcissant et en se calcifiant puis la média calcifiée perd son élasticité et devient fragile : la paroi peut se dilater (anévrisme) et risque de rompre. Des circonstances aggravantes qui ne sont pas encore clairement identifiées, entraînent l'ulcération de la plaque fibro-lipidique calcifiée : on évoque notamment les contraintes mécaniques liées à la sténose artérielle et l'agression par les LDL (Low Density Lipoprotein) oxydées (Steinberg, 1991). Des archaea associées à des bactéries aérobies du genre *Chlamydophila* et *Mycoplasma* sont observées au sein de ces plaques fibreuses (Nicolson *et al.*, 2003 ; Momiyama *et al.*, 2004; Higuchi *et al.*, 2006). Ces deux espèces bactériennes pourraient être impliquées dans l'oxydation des LDL (Kalayoglu *et al.*, 1999). L'association avec une espèce d'archaea anaérobie (microphile) qui exprime des enzymes telles que la SuperOxyde Dismutase (SOD) impliquée dans la défense contre la toxicité de l'oxygène, pourrait favoriser le développement de ces deux espèces bactériennes. Il pourrait donc y avoir une relation de cause à effet entre la présence d'archaea et l'athérosclérose bien que celle-ci implique des organismes bactériens.

Un autre exemple d'implication possible d'archaea dans les pathologies concerne *M. oralis* dans les endodontites et parodontites, qui sont, respectivement, des infections du canal radiculaire de la dent ou des tissus de la base de la dent tels que la gencive, le ligament alvéolaire, le ciment, et l'os alvéolaire (Kulik *et al.*, 2001 ; Lepp *et al.*, 2004 ; Vianna *et al.*, 2006). Des archaea y sont détectés associés à d'autres bactéries alors qu'ils sont absents des échantillons sains, ce qui suggère un rôle de *M. oralis* dans ces infections polymicrobiennes. Une fois encore, une espèce archaea est ainsi décrite comme partenaire syntrophique d'espèces bactériennes pathogènes. Toutefois, il est surprenant de constater qu'à ce jour aucun archaea pathogène en tant que tel n'a été identifié.

G. Intérêt biotechnologique

Comme décrit auparavant, les archaea présentent des protéines et des ARN qui leur sont spécifiques. Par ailleurs, certaines protéines sont très différentes de leurs homologues bactériens ou eucaryotes, ce qui peut être expliqué par l'adaptation nécessaire aux conditions extrêmes de température, de salinité, de pression, de radioactivité, de pH et de concentration en substances toxiques telles que certains solvants ou détergents. L'intérêt des enzymes d'archaea pourrait être multiple : simplification des procédés de purification, meilleure efficacité de la catalyse à température élevée, meilleure solubilité, viscosité diminuée, plus grande stabilité. Par conséquent, les protéines d'archaea représentent un potentiel important en matière d'application en industrie chimique, alimentaire ou pharmaceutique et en matière de protection de l'environnement (pour revues, Schiraldi *et al.*, 2002 ; van den Burg, 2003 ; Egorova et Antranikian, 2005). Certaines des enzymes d'archaea ont déjà des applications en biologie moléculaire, c'est le cas de la *Pfu* DNA polymérase extraite de *P. furiosus* qui est utilisée pour les réactions de PCR et l'industrie utilise déjà une variété d'enzymes d'espèces bactériennes ou d'archaea thermophiles pour, par exemple, l'hydrolyse de polymères tels que l'amidon ou la cellulose qui sont liquéfiés à température élevée (pour revue, van den Burg, 2003) ainsi que des enzymes d'archaea psychrophiles par exemple en tant que détergents utilisables à froid (pour revue, Cavicchioli *et al.*, 2002). Les perspectives sont diverses et nombreuses : les lipides bipolaires de type « tetraethers » pourraient être utilisés comme monomères en bioélectronique (De Rosa *et al.*, 1994) ; des enzymes de désulfuration et vulcanisation de *P. furiosus* pourraient permettre le recyclage du caoutchouc des pneus (Bredberg *et al.*, 2001) ; les acidophiles pourraient être exploités pour l'extraction de métaux (Norris *et al.*, 2000) ; des souches d'archaea pourraient être à l'origine de l'identification de nouveaux antibiotiques (O'Connor et Shand, 2002) ; les protéines chaperonnes d'archaea peuvent permettre d'améliorer le repliement et les propriétés de stabilisation et de solubilisation de protéines recombinantes (Ideno *et al.*, 2004 ; Maruyama *et al.*, 2004). Cette liste non exhaustive ne représente qu'une fraction des applications actuelles ou envisagées à court terme et dépend grandement de la compréhension des mécanismes moléculaires et cellulaires d'adaptation aux environnements particuliers des archaea. Il est à noter également que pour pallier à des problèmes d'expression hétérologue de ces protéines dans des systèmes bactériens, des systèmes d'expression ont été testés avec succès dans des espèces de *Methanococcus* et *Sulfolobus* (Gardner et Whitman, 1999). En matière d'écologie enfin, il est intéressant de préciser que les méthanogènes déjà utilisés pour coupler dégradation des déchets biologiques et production d'énergie (Scherer *et al.*, 2000) pourraient participer à l'effet de serre par l'émission de méthane dont l'effet de trappe des Infra Rouges serait 20 à 60 fois plus important que le CO₂. La maîtrise du procédé de méthanogenèse ou de sa réaction inverse fait donc partie des procédés qui sont des enjeux majeurs en terme de bioremédiation consistant à augmenter la biodégradation ou la biotransformation par des microorganismes (pour revue, Schiraldi *et al.*, 2002).

H. Outils génétiques

Les études génétiques conduisant à une étude *in vivo* d'un modèle sont le plus souvent réalisées à partir des espèces *E. coli* pour les Bactéries, ou *S. cerevisiae* pour les levures. Ces organismes permettent de réaliser des études *in vivo* qui peuvent compléter des études *in vitro*. Les cultures et productions y sont par ailleurs rapides et bien maîtrisées. Or chez les Archaea, les conditions de cultures sont difficiles à appréhender et à maîtriser : caractère acide, anaérobiose stricte, température élevée, pression élevée...et le temps de culture dépasse souvent la semaine. De plus, les outils génétiques, à savoir : protocoles de transformation et marqueurs sélectionnables, n'ont été développés que récemment. Ainsi l'accumulation de données structurales et biochimiques en provenance des archaea s'accompagne d'une quasi absence de données *in vivo*. Aujourd'hui, seuls quelques laboratoires « archae-ologistes » utilisent ces nouveaux modèles d'étude.

Différentes techniques de transformation ont été testées : la transfection d'archaea par un ADN nu en présence de PolyEthylèneGlycol (PEG) est efficace avec les archaea pour lesquels il est possible de retirer la couche de glycoprotéine de surface (S-Layer) (Lucas *et al.*, 2002) ; l'électroporation s'est montrée efficace pour certaines espèces (Patel *et al.*, 1994 ; Cannio *et al.*, 1998) et impossible pour d'autres telles que les halophiles qui ne supportent pas la faible concentration en sel requise pour l'application de cette technique (<1 M NaCl) ; la transformation des espèces du genre *Methanosarcina* ne pourrait se faire que via des liposomes, technique actuellement adoptée pour les méthanogènes (Metcalf *et al.*, 1997). L'ADN contenu dans une cellule d'archaea peut ensuite être transféré aux cellules voisines par transduction via des phages (Meile *et al.*, 1990) ou par « cell mating » (Rosenshine *et al.*, 1989).

La plupart des antibiotiques sont inefficaces contre les Archaea. En effet, les antibiotiques ont été sélectionnés pour affecter la survie ou le développement bactérien sans affecter le développement des cellules eucaryotes. Puisque les Archaea partagent de nombreuses caractéristiques avec les Eucaryotes, il n'est pas surprenant de ne voir une sensibilité qu'envers une petite portion des antibiotiques connus, par exemple, la Novobiocine qui inhibe la gyrase B (gyrB) présente chez les Bactéries et les Archaea, ou la Puromycine, analogue de substrat du site Peptidyl Transférase, efficace dans les deux domaines également. Dans chaque cas, un gène de résistance à l'antibiotique a été identifié soit chez les Bactéries soit chez les Archaea (Gernhardt *et al.*, 1990 ; Holmes et Dyall-Smith, 1990 ; Holmes *et al.*, 1991) permettant ainsi le développement de vecteurs sélectionnables (pour revue, Allers et Mevarech, 2005). Des marqueurs d'auxotrophie (leucine, histidine, proline, tryptophane, thymidine, uracil...) ont également été développés chez les hétérotrophes et autotrophes facultatifs qui sont capables de prélever ces éléments dans le milieu. L'auxotrophie à l'uracil est très utilisée car l'addition d'acide-5-fluoroacétique (5-FOA) permet de contre-sélectionner les cellules qui expriment le gène *ura3* dont le produit permet la conversion du 5-FOA en 5-fluorouracil toxique.

La sélection des clones recombinants peut s'effectuer sur milieu solide : il s'agit de boîte de Petri traditionnellement remplies d'agar pour les halophiles, ou de Gelrite® pour les

hyperthermophiles. Ce dernier est un polysaccharide déacétylé qui se solidifie en présence de cations divalents et conserve de ce fait un état solide jusqu'à 120°C. Pour les archaea prospérant en anaérobiose tels que les méthanogènes, des conditions anaérobies de culture sont requises et des systèmes particuliers de boîtes à gants sont utilisés.

Ainsi, des techniques visant à déléter, remplacer ou muter un gène ont vu le jour chez les halophiles, les thermophiles et les méthanogènes, et d'autres sont encore en cours de développement (Allers et Mevarech, 2005 ; Rother et Metcalf, 2005). Une stratégie basée par exemple sur l'utilisation de plasmides présentant deux marqueurs pour une sélection à la fois positive et négative ainsi qu'une région complémentaire à une partie d'un gène cible permet le remplacement ou la délétion du gène cible par deux étapes de recombinaison (Fig. I.13). De ce fait, de nombreuses questions qui n'ont pas pu être abordées par des approches moléculaires le seront à court terme par des approches génétiques.

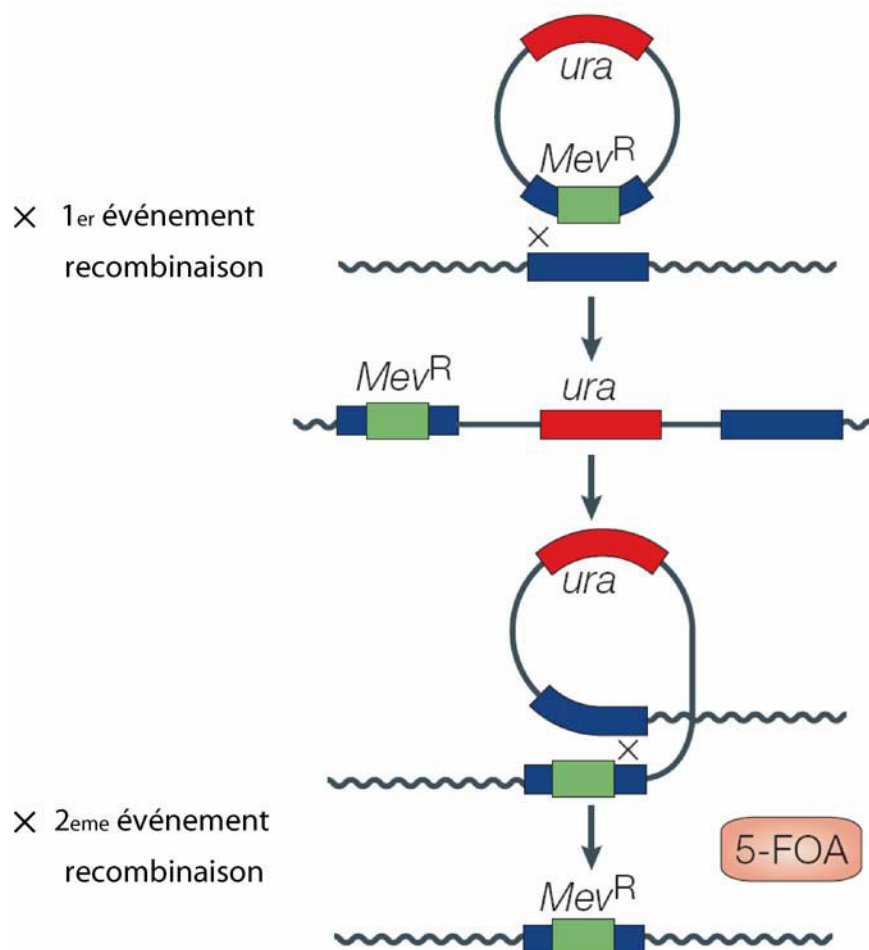


Figure I.13 : Exemple de stratégie de délétion d'un gène archaea.

Une souche *ura⁻* est transformée par un plasmide intégratif (sans origine de réplication) qui porte les gènes *Mev^R* (gène de résistance à la mévinoline) et *ura* ainsi qu'une région de complémentarité avec le gène cible. L'ADN plasmidique est intégré à l'ADN de l'hôte par un premier événement de recombinaison. Un second événement de recombinaison permet de reformer un plasmide qui contient une partie du gène ciblé chez l'hôte et de conserver le gène de résistance à la mévinoline. La sélection est effectuée en présence de Mévinoline qui favorise la récupération de souches ayant conservé le gène de résistance, et le 5-FOA permet de se débarrasser des cellules qui comporteraient encore le gène *ura*.

I. Virus d'Archaea

Les virus sont généralement classés selon le type d'acide nucléique de leur génome, la forme de leur capsid, la présence ou l'absence d'enveloppe, ou simplement par l'identité de leur hôte. Comme les Archaea se distinguent nettement des deux autres domaines du vivant, il est intéressant de savoir si les virus d'archaea sont eux aussi adaptés aux conditions extrêmes de croissance.

Les virus d'archaea ont tous un ADN double-brin comme génome, sont présents dans toutes les espèces d'archaea étudiées et sont très différents des virus des deux autres domaines du vivant, ce qui explique qu'ils ne sont pas capables d'infecter des membres d'un autre domaine. La forme singulière de certains des virus d'hyperthermophiles ainsi que la variété de ces formes sont surprenantes. Les premières observations en microscopie électronique ont révélé la présence de virus tout à fait communs chez les euryotes (halophiles et méthanogènes), ces virus ressemblant au bactériophage T4 et autres membres de la famille des *Myoviridae*. Par contre, les morphotypes des virus de crénotes hyperthermophiles et acidophiles ont des formes encore jamais observées pour des virus infectant des espèces des deux autres domaines du vivant : fusiformes (*Fuselloviridae*) dont bi-caudés (*Bicaudaviridae*), forme de goutte ou de citron (*Guttaviridae*), de bouteille ou d'ampoule (*Ampullaviridae*), de bâton (*Rudiviridae*), filamenteux (*Lipothrixviridae*) ou sphériques (*Globuloviridae*) (Fig. I.14) (pour revue, Prangishvili et Garrett, 2005; Prangishvili *et al.*, 2006a). Récemment, des virus présentant le même type de particularités ont également été détectés chez les euryotes hyperthermophiles. Ainsi, l'étude des morphotypes, des structures des protéines virales et des génomes des virus d'archaea révèle une variété supérieure à celle observée pour les virus ADN double-brin de Bactéries ou d'Eucaryotes (Ackermann, 2001 ; Prangishvili *et al.*, 2006a). Le fait que cette diversité soit plus particulièrement observée chez les archaea hyperthermophiles pourrait être expliquée par une réduction de la variété, au cours de l'évolution, due à l'expansion des virus de type bactériophage dans les habitats correspondant à des conditions modérées. L'absence actuelle de virus ARN infectant des archaea peut être expliquée par l'origine présumée hyperthermophile, discutée ci-avant, de ce domaine, et le caractère labile de l'ARN à haute température.

Il est intéressant de remarquer que tous les virus d'hôtes acidophiles et hyperthermophiles, à l'exception de l'un d'eux, maintiennent une relation stable avec leur hôte et leur cellules filles (Prangishvili *et al.*, 2001) ce qui limite leur exposition aux environnements extrêmes. Le virus ATV (Acidianus Two-tailed Virus) y est le seul virus lytique. Il a la particularité de présenter deux longues queues une fois en dehors de l'hôte, et ce uniquement dans un milieu où la température excède 75°C, ce qui pourrait constituer la première stratégie de survie virale en dehors d'un hôte (Haring *et al.*, 2005). Une fonction possible de ces queues flexibles pourrait être de favoriser l'ancrage à un nouvel hôte dans un environnement agressif peu pourvu d'hôtes (Prangishvili *et al.*, 2006b).

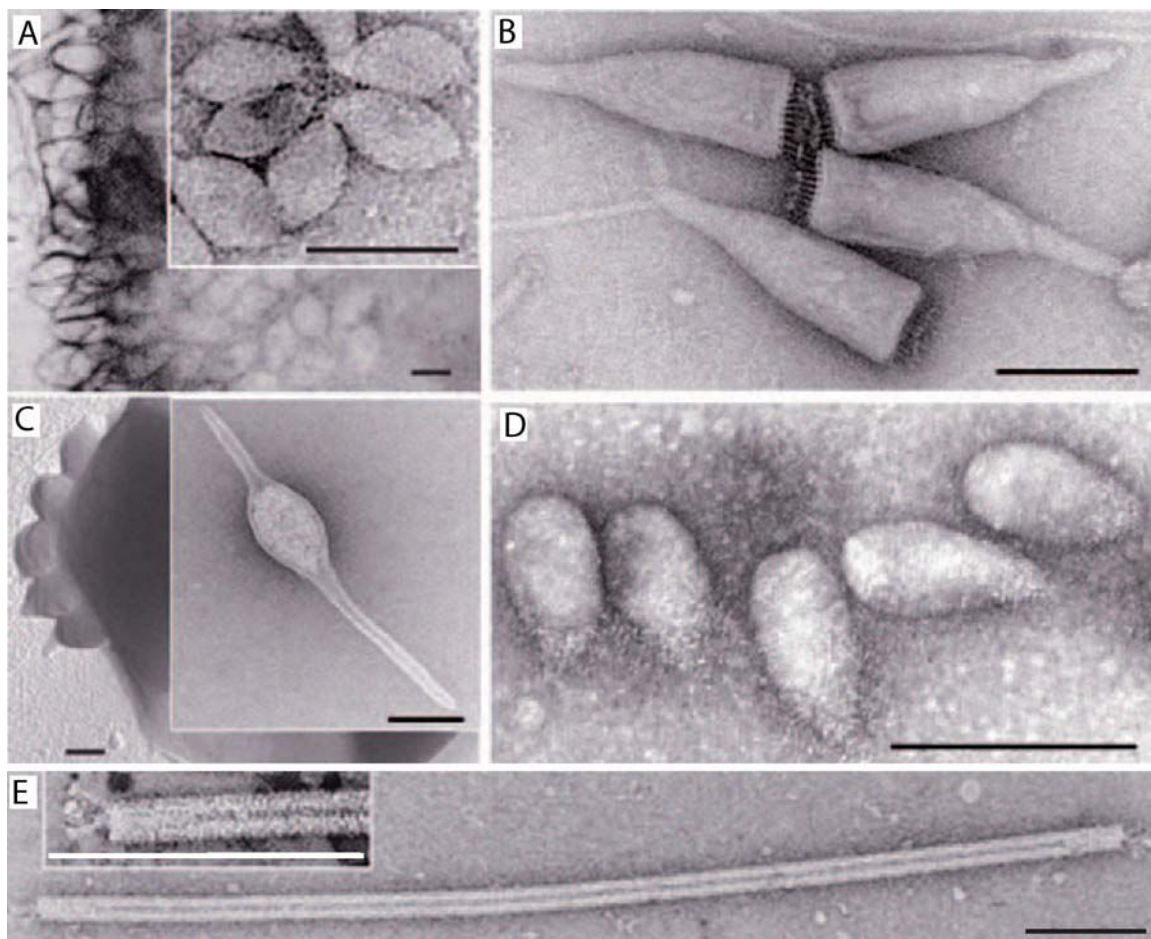


Figure I.14 : Morphologies particulières de certains virus d'archaea (d'après Prangishvili *et al.*, 2006). Visualisation en microscopie électronique. **A.** Virus fusiforme SSV1 (*Sulfolobus* spindle-shaped virus 1) et sa sortie de la cellule hôte. **B.** Forme de bouteille du virus ABV (*Acidianus* bottle-shaped virus). **C.** Forme bi-caudée du virus ATV (*Acidianus* two tailed virus) et sa sortie de la cellule hôte. **D.** Forme de goutte d'eau du virus SNVD (*Sulfolobus neozealandicus* droplet-shaped virus). **E.** Forme de bâton du virus SIRV1 (*Sulfolobus islandicus* rod-shaped virus 1). L'échelle représente 100 nM.

L'étude des virus d'archaea et de leurs relations avec les virus des autres domaines devrait conduire à une meilleure compréhension de l'émergence des premiers virus et de ce fait à l'origine des Archaea. En effet, Patrick Forterre a récemment suggéré que les virus pourraient avoir joué un rôle majeur dans l'apparition de l'ADN et de ses mécanismes de réplication ainsi que dans la séparation des trois domaines du vivant (Forterre, 2005 ; Forterre, 2006a). Les virus ARN sont considérés comme des reliques d'un monde ARN/protéines qui auraient précédé le monde ADN actuel (Woese, 1968 ; Gilbert, 1986); Fig. I.15A). Les virus proviennent probablement de formes de vie parasites ayant perdu leur capacité à produire indépendamment leurs protéines (Fig. I.15B). Or, les virus ARN auraient pu développer une stratégie de contournement des défenses de l'hôte conduisant, entre autres, à la réduction du ribose et à la méthylation des résidus uracile en thymine, i.e. à l'apparition de l'ADN (Fig. I.15C) (Forterre, 2005). L'ADN viral, plus stable chimiquement, répliqué de façon plus fidèle et autorisant des

mécanismes de réparation des cytosines désaminés en uracile, contrairement à l'ARN, aurait permis, après infection d'une cellule ARN et transfert dans le génome viral des gènes cellulaires rétro-transcrits, le développement de génomes d'une taille plus importante (Fig. I.15D).

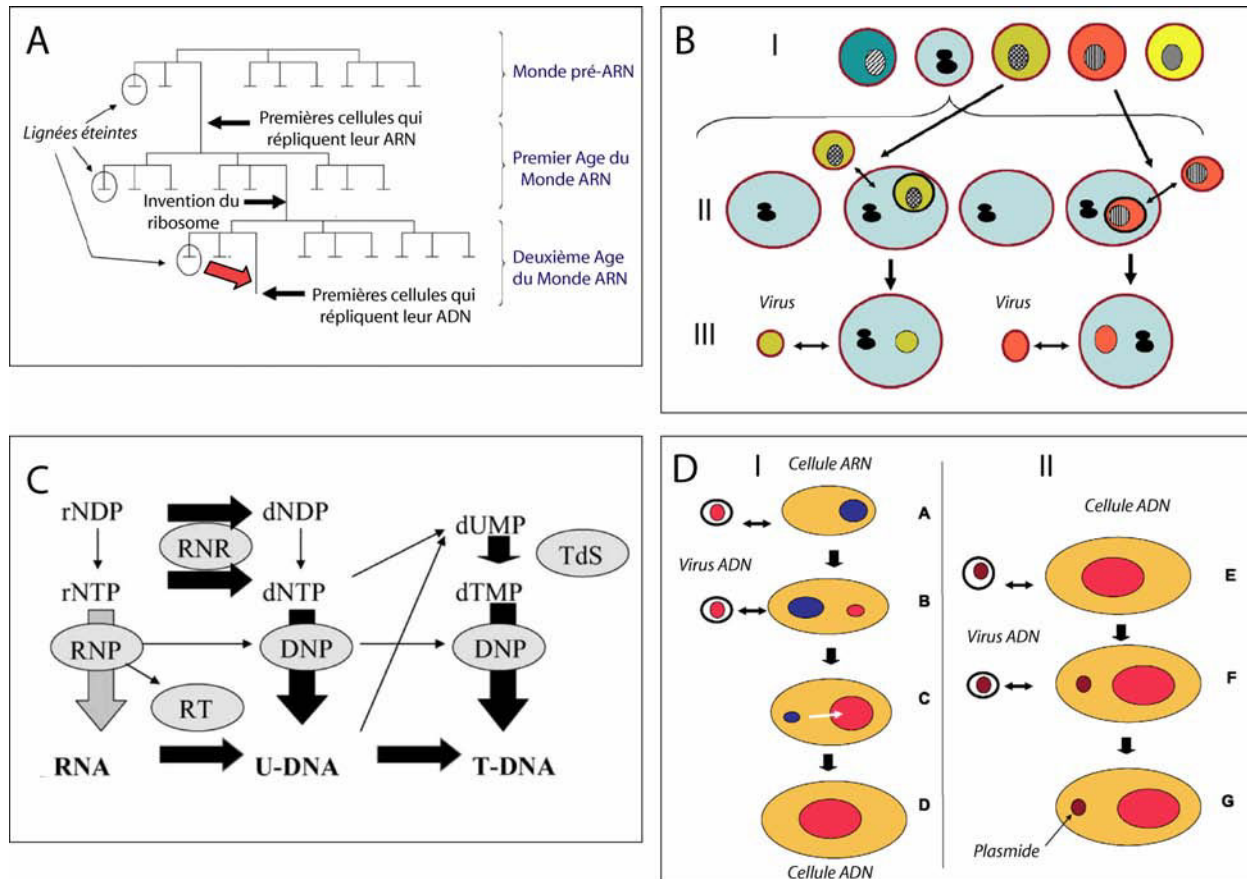


Figure I.15 : Théorie de l'émergence des virus et du monde ADN à partir du monde ARN (d'après Forterre, 2005). **A.** Distinction de 2 Mondes ARN révolus. Une évolution en trois étapes est suggérée : apparition d'une machinerie de réplication de l'ARN, invention du ribosome permettant la synthèse de protéines (d'autres machineries de production de peptides ou protéines peuvent avoir vu le jour au cours du premier âge) et apparition de l'ADN (flèche rouge) qui remplace l'ARN en tant que génome. Des lignées des deux mondes peuvent survivre en tant que parasites ou virus. **B.** Emergence des premiers virus. I. Les premiers ribosomes apparaissent dans un type cellulaire qui coexiste avec d'autres types présentant ou non des mécanismes alternatifs pour produire des protéines. II. Le type cellulaire présentant des ribosomes élimine, par compétition, les autres types dont certains survivent en tant que parasites intracellulaires. III. Les parasites perdent leur machinerie de synthèse de protéines et deviennent des virus. **C.** Transition de l'ARN vers l'ADN. Les différentes voies enzymatiques sont représentées. RNR : ribonucléase réductase ; TdS : thymidylate synthase ; RNP : RNA polymérase ; DNP, DNA polymérase ; RT : reverse transcriptase. **D.** Modèle de transfert de l'ADN à partir de virus vers une cellule ARN conduisant à l'émergence de génomes cellulaires à ADN et de plasmides. I. Un virus ADN infecte une cellule ARN et co-évolue avec elle à un état stable. Les gènes de la cellule ARN sont progressivement transférés vers l'ADN viral par rétrotranscription. L'ADN viral devient progressivement un plasmide qui entre en compétition avec le génome ARN jusqu'à devenir un chromosome ADN. II. L'infection d'une cellule ADN par un virus ADN peut conduire par un mécanisme similaire à l'obtention d'une cellule ADN comportant un plasmide en plus d'un chromosome.

Les cellules à génomes ADN plus complexes auraient ainsi pris le dessus sur les cellules ARN antérieures. Cette hypothèse séduisante expliquerait pourquoi 3 transferts d'ADN provenant de trois virus différents dirigés vers 3 types de cellules ARN différents auraient engendré 3 domaines primaires (Fig. I.16 ; Forterre, 2006a; Forterre, 2006b). Les ressemblances entre d'une part les 3 cellules ARN ciblées par ces transferts et d'autre part les 3 virus ADN impliqués sont actuellement les seuls arguments qui permettent d'expliquer la distribution des caractéristiques évoquées au cours de ce chapitre (Fig. I.16).

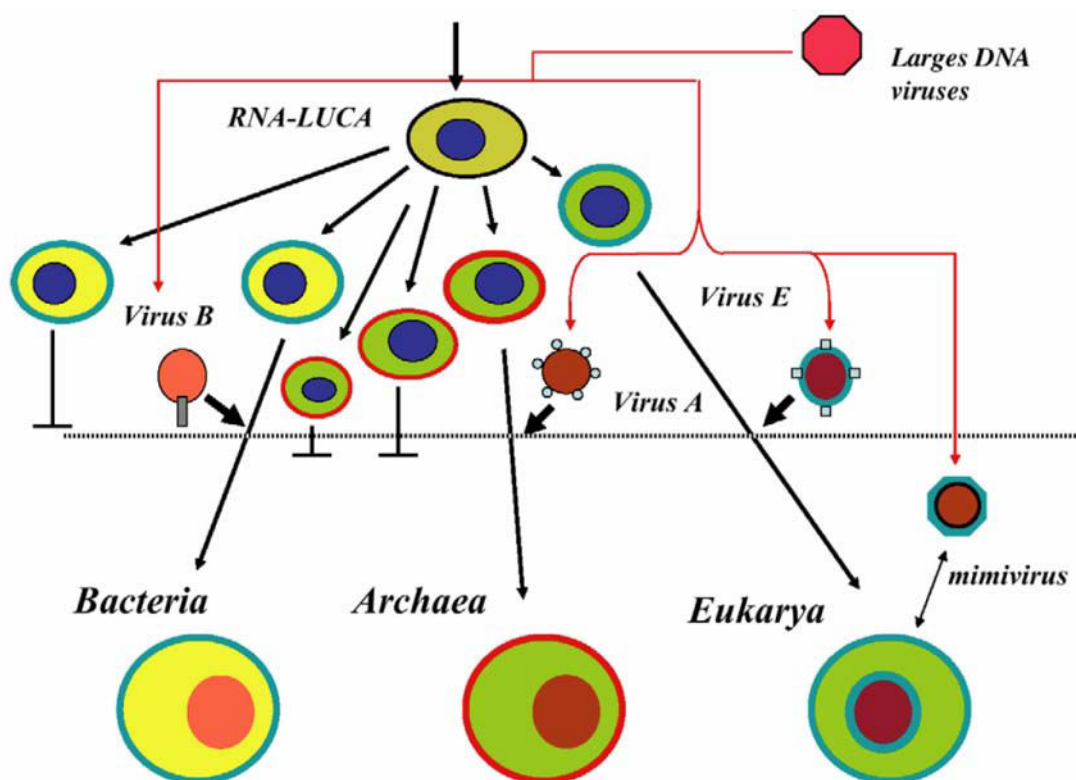


Figure I.16 : Hypothèse de l'émergence des trois Domaines du vivant par trois transferts d'ADN viral dans trois cellules ARN différentes (d'après Forterre, 2005).

Les génomes ARN et ADN sont représentés en bleu et en rouge respectivement. Les cytoplasmes sont représentés en jaune ou en vert selon le type de machinerie de la traduction. Dans trois des types cellulaires l'ARN génomique a été remplacé indépendamment par de l'ADN provenant de 3 virus à ADN double-brin différents. Les virus à ADN et cellules ARN à l'origine des Archaea et des Eucaryotes partagent plus de similarités dans leurs gènes informationnels (représenté par les couleurs similaires des génomes et cytoplasmes). Cependant, les cellules ARN à l'origine des Bactéries et des Eucaryotes partagent les mêmes types de lipides membranaires (représenté par une membrane verte). La ligne en pointillés indique la frontière entre « Monde ARN » et « Monde ADN ». Les virus géants de type mimivirus (Raoult *et al.*, 2004) ne sont pas positionnés sur ce modèle hypothétique.

INTRODUCTION

Partie II : Recherche de gènes d'ARNnc

1. Les ARN non codant

A. Généralités

Les gènes d'ARN non codant (ARNnc ou ncRNA) dirigent la synthèse de transcrits qui ne codent pas des protéines mais fonctionnent directement en tant qu'ARN (Brosius et Tiedge, 2004). Cependant, la séquence de ces ARN est soumise à des contraintes liées à leur mécanisme d'action et à leur stabilité particulière (Argaman *et al.*, 2001 ; Eddy, 2001 ; Costa, 2007). D'où les dénominations suivantes aussi employées pour distinguer les deux classes possibles de transcrits : la classe I correspondant aux transcrits non-traduits, i.e. ne présentant aucune ORF et la classe II correspondant aux ARNm traduits en protéines par la machinerie ribosomale (Brosius et Tiedge, 2004). Notons que les termes ARN fonctionnels (ARNf ou fRNA), ARN non traduits (utRNA : untranslated) ou petits ARN (sRNA) sont souvent utilisés pour désigner les ARNnc.

1. Variété des ARNnc

Les ARNnc sont présents au sein des trois Domaines primaires du vivant et régulent de nombreuses fonctions cellulaires de base, le développement des organismes ainsi que de nombreuses adaptations aux variations du milieu. Leurs représentants les plus étudiés sont les ARNr et les ARNt intervenant dans la traduction. Cependant, d'autres ARNnc sont impliqués dans l'orientation des protéines tronquées vers une voie de dégradation (ARNtm : transfert messenger ; Gillet et Felden, 2001), dans l'épissage (snRNA : small nuclear ; Will et Luhrmann, 2001), dans la modification post-transcriptionnelle des ARN (snoRNA : small nucleolar ; pour revue, Kiss, 2002) y compris l'édition (ARNg : guide), dans la sécrétion de protéines (ARN de la SRP, Signal Recognition Particule ; Althoff *et al.*, 1994), dans la régulation de la réplication de plasmides (système hok-sok ; Gerdes *et al.*, 1997) ou dans les réponses au stress (e.g. OxyS ; Argaman et Altuvia, 2000). A cette liste loin d'être exhaustive viennent s'ajouter la régulation de la traduction, de la stabilité de l'ARNm et de la transcription par les microRNA tels que lin-4 et let-7 (pour revue, Pasquinelli et Ruvkun, 2002 ; Carrington et Ambros, 2003) et les siRNA (mécanisme d'interférence ARN) qui conduisent à la dégradation spécifique d'ARNm (siRNA : small interfering, ou RNAi : interference ; Pickford et Cogoni, 2003). Dans les génomes eucaryotes, l'identification très récente d'autres types d'ARNnc, tels que les piRNA (car co-purifiés avec la protéine PIWI) impliqués dans la gamétogenèse (Lau *et al.*, 2001 ; Aravin *et al.*, 2006 ; Girard *et al.*, 2006) ou les PINC (Pregnancy-Induced ncRNA) qui jouent un rôle dans la modulation des cellules des glandes mammaires (Ginger *et al.*, 2006), suggère que de nombreuses classes d'ARNnc sont encore inexplorées.

2. Relecture du génome

Les ARNnc sont donc divers et nombreux. Ils ne sont pas des sous-produits des ARNm, néanmoins, certains sont produits à partir des introns des pré-ARNm (Costa, 2007). Des études

bioinformatiques et expérimentales réalisées chez l'Homme et la Souris indiquent que le nombre d'ARNnc est sans doute comparable au nombre d'ORF (soit estimés à une valeur comprise entre 20000 et 30000 ORF et autant voire plus d'ARNnc ; Consortium, 2004 ; Washietl *et al.*, 2005a ; Furuno *et al.*, 2006 ; Liu *et al.*, 2006 ; Mattick et Makunin, 2006). Ainsi, moins de 2% du génome humain abrite des ORF alors qu'au moins 60% d'un génome de mammifère est transcrit sur un des deux brins, cette dernière valeur étant sous-estimée car elle ne tient pas compte de l'analyse de différents types cellulaires à différents stades du développement et dans différentes conditions (Consortium, 2004 ; Carninci *et al.*, 2005 ; Frith *et al.*, 2005). L'idée que le génome humain soit constitué d'une masse d'ADN structural dans laquelle est noyée une faible quantité de gènes, est maintenant révolue au profit d'une vision où le génome est en grande partie transcrit (Kampa *et al.*, 2004). Cette prise de conscience récente est venue bouleverser la vision des interactions et régulations qui ont lieu dans la cellule (Costa, 2007). Il n'est donc pas surprenant de découvrir un nombre croissant de liens entre ARNnc et pathologies, en particulier cancers et maladies neurologiques (pour revue, Costa, 2005).

3. ARNnc et complexité

Il est difficile d'envisager une compréhension complète des mécanismes cellulaires sans considérer l'influence d'ARNnc ubiquitaires et multifonctionnels. D'abord considérés comme reliquats d'un monde ARN, il est de plus en plus évident que les ARN ont été sélectionnés dans de nombreuses innovations au cours de l'évolution. Récemment, il a été démontré que contrairement au nombre de gènes de protéines, le nombre de gènes d'ARNnc augmente avec la complexité d'un organisme eucaryote. Ceci pourrait ainsi expliquer la relation paradoxale observée entre un nombre de gènes de protéines quasi-constant entre ciliés, plantes, poissons, souris et humains et un degré de complexité croissant (Taft *et al.*, 2007). Il faut bien sûr tenir compte de l'accroissement de la complexité lié à l'augmentation des possibilités d'épissage alternatif. Néanmoins, ceci n'est pas suffisant pour expliquer l'évolution de la complexité, et le recours à des mécanismes de régulation reposant sur des ARNnc doit aussi largement y contribuer. Les ARNnc sont en effet bien adaptés à la réalisation d'une fonction bien définie par reconnaissance par complémentarité de séquence, ils permettent de répondre rapidement à un stimulus externe par une synthèse et une destruction rapide, ceci en économisant le coût correspondant à la production de protéines (Eddy, 2002). Par ailleurs, la facilité d'évolution des interactions entre ARN qui ne comportent que 4 nucléotides par rapport aux interactions entre protéines reposant sur des combinaisons complexes d'acides aminés au niveau des domaines de reconnaissance des protéines, a probablement favorisé l'apparition rapide de nouvelles régulations.

4. Identification biochimique

Bien qu'occupant une place majeure dans le métabolisme cellulaire, les premiers ARNnc ont été découverts par hasard (pour revue, Wassarman *et al.*, 1999). Ce n'est que depuis la fin des années 1990 que quelques analyses systématiques ont été réalisées en vue de les répertorier. Les études expérimentales se sont heurtées à différentes difficultés en particulier à leur faible taux d'expression qui ne permet parfois aucune détection en analyse par puces ADN (Bono *et al.*, 2003). Par ailleurs, du fait de leur implication dans des régulations fines ils ne sont parfois exprimés que dans des conditions très précises, en réponse à un signal environnemental, lors d'une étape du développement ou dans un tissu cellulaire particulier (Storz, 2002). En contrepartie, il reste envisageable qu'une partie des transcrits détectés expérimentalement ne soient que des produits de maturation, d'épissage ou de dégradation d'ARNm ou d'autres ARNnc (Kapranov *et al.*, 2005). Il est à noter qu'une distinction entre transcrits fonctionnels et transcrits non fonctionnels qui se baserait sur des études phénotypiques est aussi pour certains ARNnc limitée par le fait qu'ils seraient plus impliqués dans des fonctions d'adaptation ou de développement que dans des fonctions essentielles pour la cellule (Kapranov *et al.*, 2005 ; Costa, 2007). Des approches expérimentales relativement lourdes permettant une recherche systématique d'ARNnc dans les trois domaines du vivant ont été récemment mises en place et seront présentées ci-après.

5. Identification bioinformatique

La prédiction systématique des ORF par bioinformatique a pu s'appuyer sur la traduction des phases ouvertes, la présence de codons initiateurs et stop, séquences Shine-Dalgarno, séquences de régulation, biais dans l'utilisation des codons ou comparaison de séquences en acides aminés avec des ORF existantes. La recherche de gènes d'ARNnc ne peut s'appuyer sur des signaux statistiques aussi forts (Eddy, 2002). Le développement de stratégies bioinformatiques innovantes telles que celles présentées ci-après a donc été requis.

L'objectif de ce chapitre n'est pas de présenter une liste complète des nombreux ARNnc décrits puisque celle-ci est certainement amenée à être aussi longue que la liste de protéines décrites, mais uniquement de présenter les principaux modes d'action élucidés à ce jour, la distribution des différentes classes d'ARNnc dans les trois Domaines primaires du vivant et les méthodes de recherche systématique à la fois expérimentales et bioinformatiques qui ont été proposées.

B. Classification

Afin d'éclaircir le vaste ensemble des ARNnc, il convient de distinguer différentes familles. Il est envisageable de séparer les ARNnc selon leur mode d'action, leur taille, leur rôle cellulaire, leur implication dans un processus physiologique ou pathologique, leur biogenèse,

leur localisation, la présence de motifs particuliers ou encore par leur interaction avec des protéines particulières.

1. Complexes avec des protéines et motifs ARN

Les ARNnc sont en interaction permanente avec des protéines qui jouent un rôle catalytique, de chaperonne, de protection ou qui constituent un échafaudage pour l'interaction d'autres protéines ou ARNnc. Il est le plus souvent pertinent de parler de complexes ou de particules RiboNucleoProtéiques (RNP). Certaines protéines ont pour fonction de se lier aux ARNnc. Ainsi, par exemple, la protéine L7Ae archaea reconnaît un motif en K-turn (discuté par la suite) qui est présent dans de nombreux ARNnc : ARNr, ARN à boîtes C/D, ARN à boîtes H/ACA et autres ARN antisens (Zago *et al.*, 2005). Dans la plupart des cas, l'interaction avec des protéines dépend de la présence de motifs ARN spécifiques conservés au sein d'un type ou plusieurs types d'ARNnc. La présence de ces motifs ARN particuliers est le critère le plus simple pour distinguer des familles d'ARNnc.

Il convient, pour la suite de l'exposé, de préciser ce qu'est un motif ARN. Un motif se définit par une unité de repliement particulière de petite taille comportant des interactions de type Watson-Crick (W-C) mais aussi des appariements non W-C. Celle-ci est retrouvée plusieurs fois les ARN d'un organisme voire plusieurs fois dans un même ARN. Si la séquence et la structure secondaire constituent un premier et un deuxième niveau d'organisation, le motif représente donc un troisième niveau puisqu'il prend en compte la structure 3D. Il faut remarquer que le terme « motif » est également employé dans d'autres sens tels que pour désigner une séquence retrouvée plusieurs fois dans les ARN. Il est donc nécessaire lorsque l'on parle de motifs ARN de définir de quel type de motifs il s'agit.

2. Variété des modes d'action

Certains ARNnc sont directement impliqués dans la catalyse via leur groupement 2'OH du ribose, c'est le cas pour l'ARNr de la grande sous-unité ribosomale impliqué dans l'activité peptidyl transférase et pour l'ARN de la RNase P intervenant dans la maturation des ARNt.

D'autres ARNnc portent une activité très spécifique comme c'est le cas pour l'ARNtm (ou tmRNA) (pour revue, Keiler, 2007) qui possède à la fois une partie ARNt et une partie ARNm qui code un peptide particulier. Lorsqu'un ARNm est tronqué en amont du codon stop, l'intervention d'un ARNtm permet l'addition de ce peptide aux protéines tronquées. Ce peptide ainsi ajouté en C-ter oriente la protéine tronquée vers une voie de dégradation. L'ARNtm permet aussi, en recrutant les facteurs de terminaison via le codon stop qu'il porte, de libérer les ribosomes bloqués à l'extrémité 3' de l'ARNm tronqué (Fig. II.1).

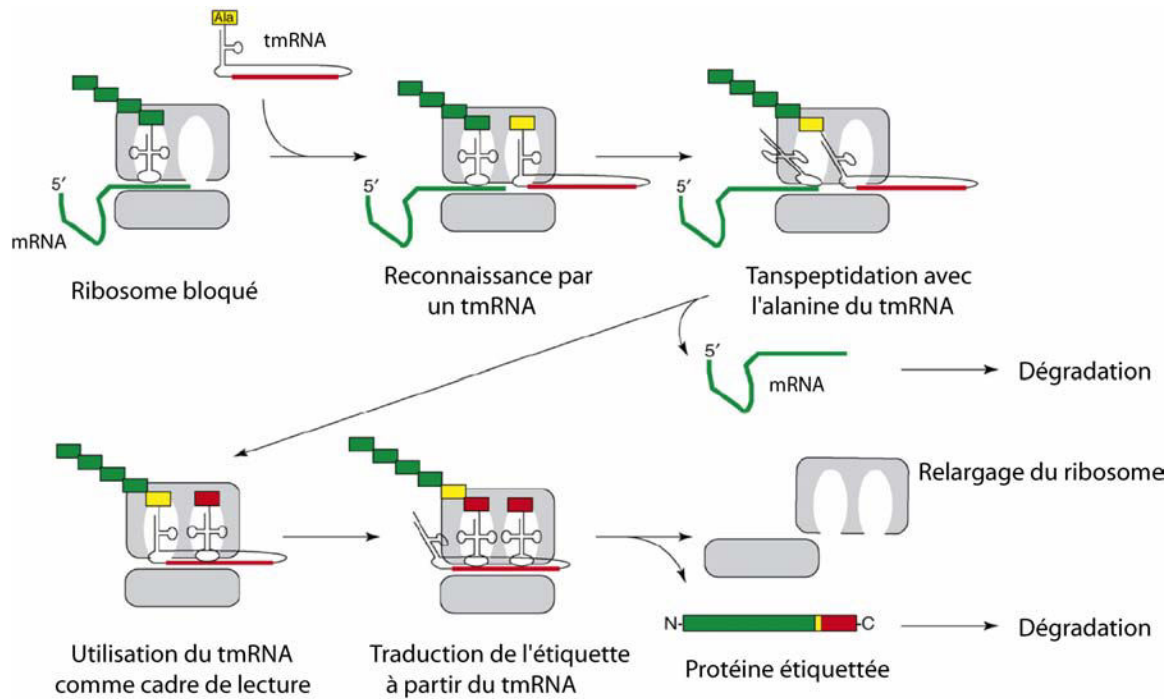


Figure II.1 : Mode d'action des tmRNA bactériens (adapté de Keiler, 2007). Un tmRNA chargé en alanyl reconnaît les ribosomes bloqués du fait de la troncation de l'ARNm. Il entre par le site A. Le peptide naissant est transféré sur le résidu Ala porté par le tmRNA. La séquence codante du tmRNA est utilisée, après élimination de l'ARNm, pour coder un peptide en C-ter du peptide naissant. Le peptide en C-ter permet d'adresser la protéine vers une voie spécifique de dégradation. La traduction se termine par un codon stop, ce qui libère le ribosome. La protéine naissante est protéolysée.

Plus fréquemment, des ARNnc interagissent avec des protéines en modulant l'activité enzymatique de la protéine, c'est le cas par exemple de l'ARN 6S bactérien (Wassarman et Storz, 2000), ou en réduisant le nombre de protéines régulatrices disponibles, c'est le cas de l'ARN CsrB (Liu *et al.*, 1997). Ces deux exemples sont discutés ci-après.

Cependant, le plus souvent, les ARNnc interviennent via une interaction séquence-spécifique par complémentarité complète ou partielle avec une séquence de l'ADN génomique, d'un ARNm ou d'un autre ARNnc (pour revue, Erdmann *et al.*, 2001). Cette action, qualifiée d'antisens, se divise en deux groupes selon que l'ARNnc est transcrit à partir du brin complémentaire de la région ciblée, action dite en *cis* (*cis*-antisens), ou selon que cet ARNnc est transcrit à partir d'un locus différent, action dite en *trans* (*trans*-antisens).

Notons que certains ARNnc agissent en utilisant plusieurs de ces modes d'action sur différentes cibles (ex. : OxyS, DsrA ; Altuvia *et al.*, 1997 ; Lease *et al.*, 1998).

3. Biogenèse

A l'exception des ARNnc appartenant aux *cis*-antisens, la plupart des ARNnc sont transcrits à partir de régions situées entre les ORF ou chevauchant légèrement ces régions (Olivas *et al.*, 1997 ; Dennis *et al.*, 2001). Leur transcription est soit indépendante via un promoteur propre, soit dépendante de celle d'un pré-ARNm ou d'autres ARNnc. En effet, il n'est pas rare d'observer une organisation des gènes d'ARNnc en groupes ou clusters (Gaspin *et al.*, 2000 ;

Omer *et al.*, 2000; Lau *et al.*, 2001 ; Cavaille *et al.*, 2002). Un grand nombre des transcrits précurseurs des ARNnc nécessite une maturation, voire une modification de certains des nucléotides. Par ailleurs, une grande proportion des introns eucaryotes (supérieure à 30%) est une source d'ARNnc (pour revue, Mattick et Makunin, 2006). En effet, une grande partie des snoRNA et miRNA sont codés par des introns (Ambros *et al.*, 2003 ; pour revue, Lin *et al.*, 2006).

4. Localisation

Chez les eucaryotes il est possible de distinguer les ARNnc cytoplasmiques des ARNnc nucléaires, nucléolaires ou encore contenus dans les corps de Cajal. Ces ARNnc eucaryotes présentent un motif permettant leur adressage. Les scaRNA responsables de modifications des snRNA contiennent, par exemple, dans leurs boucles terminales le motif UGAG appelé CAB Box responsable de leur localisation dans les corps de Cajal (Richard *et al.*, 2003).

5. Distinction selon leur taille

Il est possible de grouper les ARNnc en trois catégories selon la taille du transcrit maturé (Costa, 2007). Les plus petits d'entre eux sont les miRNA et siRNA dont la taille est d'environ 20 nt. Les petits ARN, généralement appelés sRNA ont une taille comprise entre 30nt et 300nt et comptent notamment les snoRNA, snRNA, piRNA ou autres ARN régulateurs encore appelés smRNA (small modulatory). Les grands ARN ou lRNA (large) ont des tailles supérieures à 300nt voire supérieures à 10000 nt. L'ARNnc Xist impliqué dans l'inactivation du chromosome X lors de l'embryogenèse ne porte pas d'ORF bien qu'ayant une taille proche de 15000 nt (Brockdorff *et al.*, 1992 ; pour revue, Brockdorff, 2002). D'autres exemples de lRNA incluent l'ARN Air ou encore H19 tout deux agissant *in cis* dans des régulations épigénétiques (pour revue, Andersen et Panning, 2003 ; Gabory *et al.*, 2006).

6. Distinction selon leur(s) rôles

Les rôles cellulaires, déjà énoncés en partie, sont nombreux : épissage, modifications, maturations, édition, inhibition post-transcriptionnelle par interférence, modulation de la structure de la chromatine, régulations de la transcription ou de la traduction, modulation de la fonction des protéines, régulation de la distribution cellulaire des ARN et des protéines. Une classification selon le rôle à la fois cellulaire et physiologique a été suggéré pour les ARNnc eucaryotes (Szymanski *et al.*, 2003).

7. Familles d'ARNnc

Le rythme incroyable des découvertes récentes n'a pas favorisé de concertation convergeant vers une terminologie adéquate des familles d'ARNnc, le terme ARN non codant étant lui-même amené à être remplacé. Par conséquent, la distinction des principales familles d'ARNnc s'appuie sur les différents critères évoqués ci-dessus. Une représentation de ces familles est suggérée dans la figure II.2.

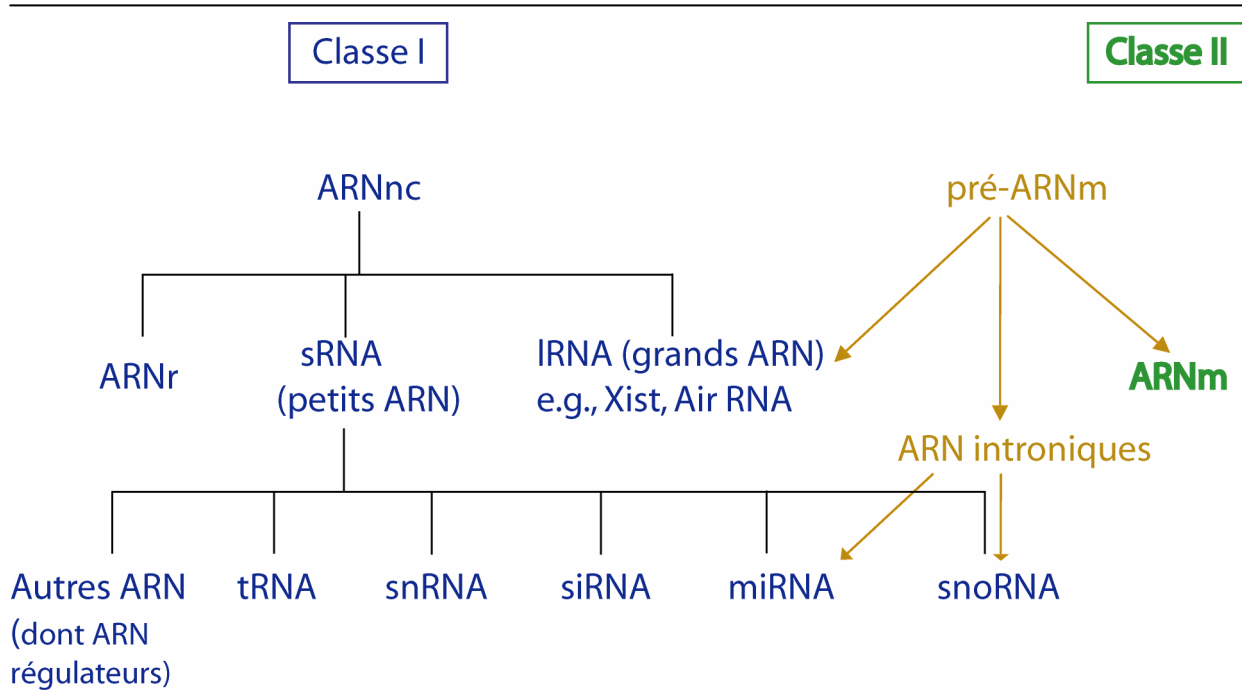


Figure II.2 : Classification des familles d'ARNnc (adapté de Brosius et Tiedge, 2004).

La classe I représente les ARN fonctionnels autres que les ARNm, i.e. les ARNnc qui ne présentent aucune ORF. La classe II comporte les ARNm qui sont traduits en protéines par la machinerie ribosomale. Les flèches indiquent l'existence d'étapes de maturation. Il est à noter que les ARNr, ARNt, snRNA et autres ARN nécessitent également des étapes de maturation non indiquée. sRNA : small RNA ; IRNA : large RNA ; siRNA : small interfering RNA ; miRNA : microRNA ; snRNA : small nuclear RNA ; snoRNA : small nucleolar RNA ; pré-ARNm principalement chez les eucaryotes où ils sont transcrits par l'ARN polymérase II (également nommé hnRNA : heterogeneous nuclear RNA).

C. Distribution dans les trois Domaines du vivant

Les ARNr, ARNt, ARN de la SRP (ARN 4.5S chez les Bactéries, ARN 7SL chez les eucaryotes et ARN 7S chez les Archaea) impliquée dans la translocation de protéines et l'ARN de la RNase P impliquée dans la maturation des ARNt, sont communs aux trois domaines du vivant, les autres ARNnc se répartissent différemment selon le domaine du vivant considéré.

1. Bactéries

Chez les bactéries le terme sRNA est synonyme d'ARNnc. Il n'existe pas de classe homogène mais une variété de petits ARN impliqués dans de nombreuses régulations transcriptionnelles et post-transcriptionnelles soit par appariement avec l'ADN génomique ou des ARNm soit par interaction avec des protéines. En dehors des cas particuliers de la RNase P, de l'ARN 4.5S ou du tmRNA propre aux bactéries (Fig. II.1) (Christensen et Gerdes, 2003), il est possible de distinguer deux classes d'ARNnc selon que le mode d'action repose sur l'interaction avec soit une protéine cible soit un ARN cible.

1.1. Interactions avec des protéines

1.1.1 Modulation de l'activité

Certains ARNnc se lient spécifiquement à une protéine enzymatique et en modulent l'activité. Ainsi, l'ARN 6S bactérien, en mimant un promoteur ADN lorsqu'il est à l'état de complexe ouvert lors de l'initiation de la transcription, interagit spécifiquement avec l'holoenzyme ARN polymérase comportant le facteur $\sigma 70$ et, de ce fait, inhibe son activité (Fig. II.3). Il en résulte une forte diminution de la transcription des gènes à promoteurs dépendant de $\sigma 70$ (Wassarman et Storz, 2000). L'expression des gènes dont le promoteur dépend du facteur σS , facteur alternatif utilisé en phase stationnaire lors de carences en nutriments, est alors favorisée (pour revue, Wassarman, 2002).

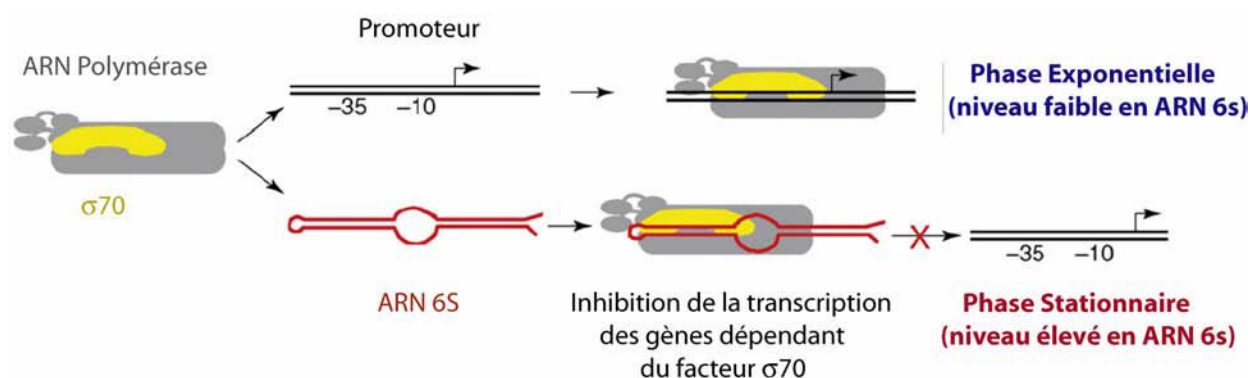


Figure II.3 : Mode d'action de l'ARN 6S bactérien (adapté de Wassarman, 2007).

L'abondance d'ARN 6S oriente la cellule vers une phase stationnaire de croissance par le biais de l'inhibition de la transcription de gènes dépendant du facteur σ^{70} .

1.1.2 Modulation du nombre de protéines disponibles

L'interaction d'un ARNnc avec une protéine impliquée dans la régulation de la traduction peut moduler le taux de traduction d'un ARNm particulier (Fig. II.5F) ou d'un ensemble d'ARNm, c'est le cas de l'ARN CsrB (Liu *et al.*, 1997). La protéine CsrA contrôle négativement l'expression d'opérons impliqués dans la biosynthèse du glycogène, mais intervient aussi dans le contrôle de la mobilité cellulaire ainsi que dans le contrôle des propriétés d'adhérence de certains pathogènes. L'ARN CsrB mime une région 5'UTR reconnue par la protéine CsrA et exerce de ce fait une action antagoniste (Fig. II.4). L'ARN CsrB comporte 360 nt et 22 sites potentiels de fixation de CsrA (pour revue, Babitzke et Romeo, 2007). CsrB réprime donc les fonctions régulatrices de CsrA en titrant cette protéine dans certaines conditions. Etant donné que la protéine cible est impliquée dans de nombreuses fonctions cellulaires de senseurs, dans la motilité, la chimiotaxie, les voies de signalisation du stress et la pathogenèse, en plus du flux du carbone, l'ARN régulateur a un rôle central même si ce dernier n'interagit qu'avec un seul type de molécule. Ainsi, ces études montrent que la synthèse d'une

seule molécule d'ARN peut permettre de générer des effets pléiotropiques et ce de façon rapide et coordonnée.

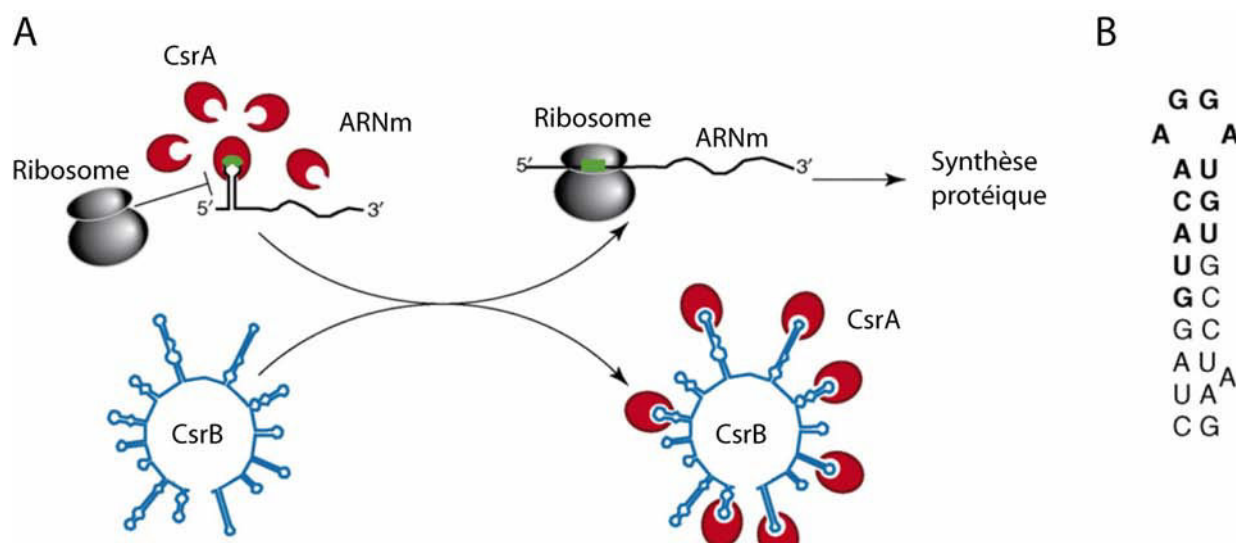


Figure II.4 : Mode d'action de l'ARN CsrB (adapté de Babitzke et Romeo, 2007 ; Toledo-Arana *et al.*, 2007). (A) Les protéines de la famille CsrA inhibent la traduction de certains gènes en se fixant au niveau des séquences Shine-Dalgarno (SD). L'ARN CsrB, mime des séquences SD et présente ainsi 22 sites d'interaction avec CsrA. La représentation circulaire de CsrB n'est pas significative. (B) Motif type reconnu par la protéine CsrA, dérivé d'expériences de Selex. La séquence consensus serait RUACARGGAUGU.

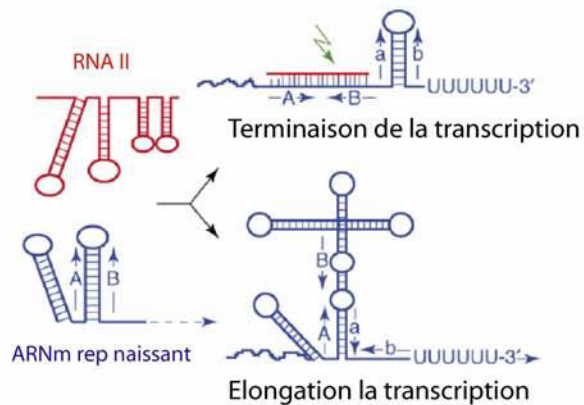
1.2. Action par appariement à un ARN

Les ARN antisens ont en général une taille inférieure à 100 nt. Ils présentent soit une séquence strictement complémentaire à leur cible car transcrits par le brin complémentaire, on parle d'action en *cis*, observé pour les plasmides, transposons et bactériophages mais aussi au sein du chromosome bactérien à partir duquel ils ne sont pas exprimés constitutivement mais dans certaines conditions (Brantl, 2007), soit une séquence partiellement complémentaire à leur cible lorsque l'ARN est synthétisé à partir d'un locus distinct, on parle d'action en *trans* (Kolb *et al.*, 2001).

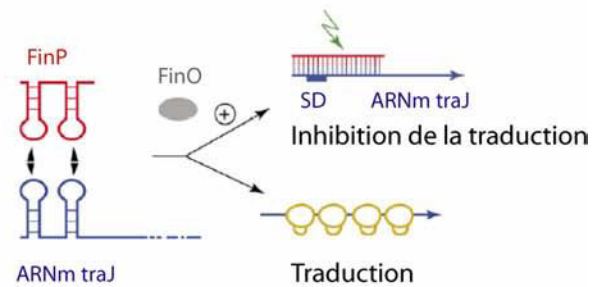
1.2.1. Action antisens en *cis*

Parmi les *cis*-antisens, le mécanisme d'action le plus fréquent est l'inhibition traductionnelle. L'appariement d'un ARN antisens à sa cible peut créer un site de coupure pour une endoribonucléase spécifique d'ARN double brin et ainsi diminuer la stabilité de l'ARNm ciblé. Ainsi, le système de destruction post-ségrégationnelle appelé « hok-sok » fait intervenir d'une part l'ARNm Hok (host-killing) dont une forme maturée d'ARNm en 3' code une toxine qui détruit la membrane de la cellule, et d'autre part l'ARN Sok (suppression of killing) de 67 nt, transcrit dans le sens opposé de Hok (Thisted *et al.*, 1994).

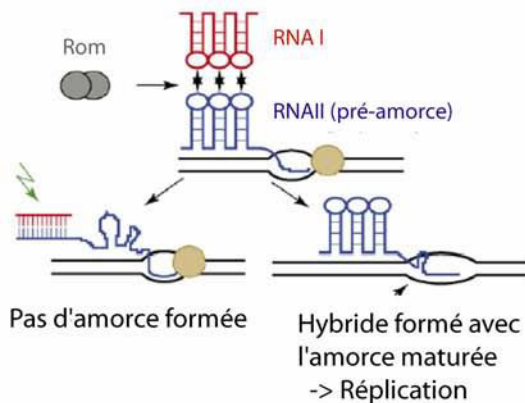
A Atténuation de la transcription



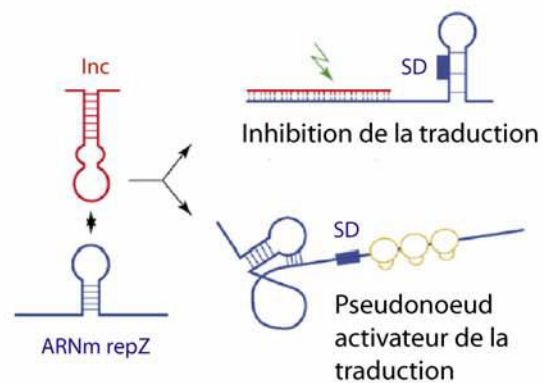
B Inhibition de la traduction



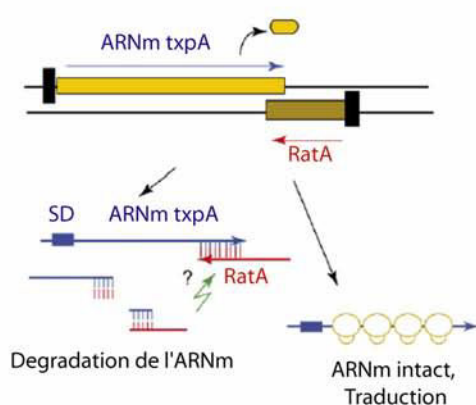
C Inhibition de la maturation d'amorce



D Inhibition de la formation de pseudo-noeuds



E Dégradation d'ARNm



F Stabilisation d'ARNm

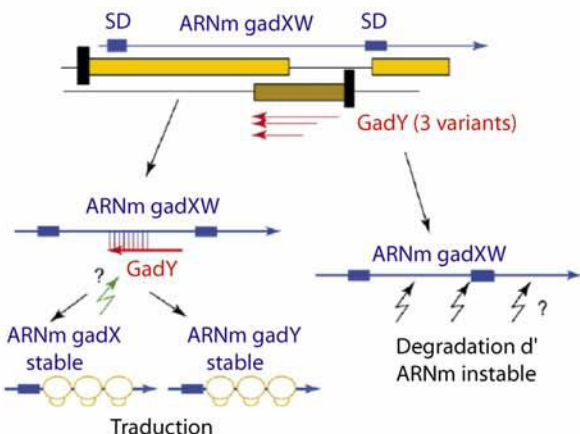


Figure II.5 : Aperçu des différents modes d'action utilisés par les ARN antisens agissant en *cis* (d'après Brantl, 2007). Différents exemples sont utilisés pour illustrer les 6 modes d'action proposés. Les ARN antisens et leur(s) cible(s) sont représentés en rouge et en bleu respectivement. Les rectangles noirs et bleus représentent des promoteurs et des séquences Shine-Dalgarno respectivement. Les flèches vertes et noires indiquent des dégradations par la RNase III ou autres RNase respectivement. «?» signale une absence de vérification expérimentale.

Ce système est porté notamment par le plasmide R1. Le transcrit Sok peut s'apparier à la partie 5'UTR de l'ARNm Hok (appelée SokT : Target) entraînant ainsi la dégradation d'un complexe Hok/Sok double brin par la RNase III. Cependant, l'ARNnc Sok a une demi-vie courte contrairement à l'ARNm Hok. Par conséquent, si le plasmide est perdu, l'ARNm Hok encore présent est maturé, conduisant de ce fait la cellule à se suicider. Ainsi les cellules qui survivent sont celles qui ont conservé le plasmide. Il semblerait que la plupart des gènes de toxines bactériennes soient soumis à ce type de régulation mais que certains chromosomes bactériens puissent également porter ce système (Gerdes et Wagner, 2007).

D'autres mécanismes d'action ont également été observés : atténuation de la transcription, inhibition de la maturation de l'amorce, inhibition de la formation de pseudo-nœuds et stimulation ou inhibition de la dégradation d'ARNm (pour revue, Brantl, 2007). Les différentes éventualités sont évoquées au travers d'exemples dans la figure II.5.

Dans tous les cas, la fixation de l'ARN antisens à sa cible est déterminante pour l'efficacité de son action. Le processus d'appariement du *cis*-antisens peut faire intervenir une ou plusieurs étapes. Plusieurs étapes sont requises, par exemple, lors de la régulation de l'initiation de la réplication du plasmide R1 par la concentration intracellulaire d'ARN *cis*-antisens CopA. La synthèse *de novo* de la protéine RepA est indispensable pour l'initiation de la réplication du plasmide R1 (pour revue, Wagner et Simons, 1994). CopT (« Target », cible de CopA) est une partie de la région 5'UTR de l'ARNm *repA*. L'interaction de l'antisens CopA avec la partie CopT inhibe la traduction de *repA*. Les contacts initiaux se font en général entre régions simple brin provenant soit de boucles terminales de structures tige-boucles soit de parties linéaires. Dans ce cas, la première interaction est de type boucle-boucle (Kolb et al., 2001b) et conduit à l'ouverture des hélices en formant un complexe transitoire appelé « kissing complexe » qui est étendu jusqu'à la formation d'une jonction à quatre hélices (Fig. II.6) (Ohman et Wagner, 1989 ; Kolb *et al.*, 2001). Il n'est pas nécessaire d'aboutir à un duplex étendu présentant un appariement strict d'une cinquantaine de nucléotides puisque la jonction à quatre hélices est reconnue par la RNase III et donc permet l'inhibition.

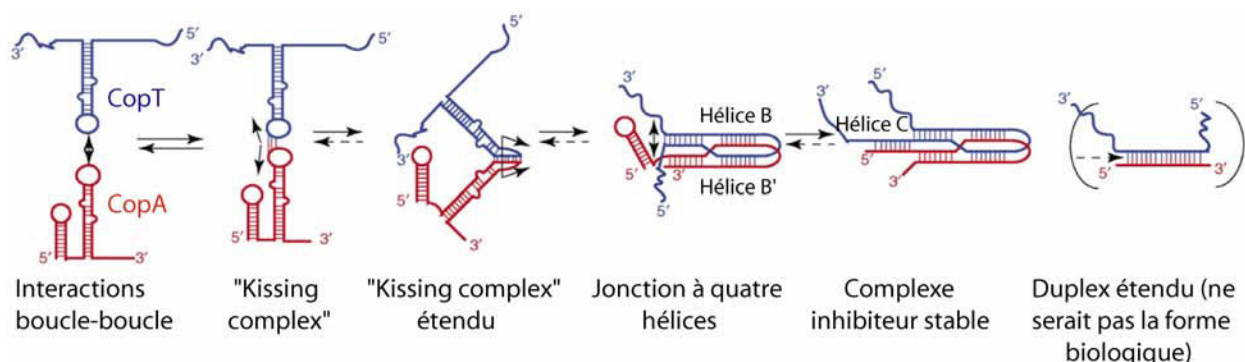


Figure II.6 : Exemple d'appariement en plusieurs étapes d'un antisens agissant en *cis* (d'après (Kolb *et al.*, 2000). L'ARN antisens CopA et sa cible dans l'ARNm *repA*, appelé CopT, sont représentés en rouge et en bleu respectivement. Les hélices inter-moléculaires sont indiquées. La première étape est une interaction boucle-boucle rapidement convertie en un complexe inhibiteur stable. La formation d'un appariement strict ou duplex étendu est un processus lent qui n'existerait pas *in vivo*.

1.2.2. Action antisens en trans

La reconnaissance entre l'ARN OxyS et l'ARNm *fhlA* est un exemple d'action d'un *trans*-antisens. L'interaction met en jeu 8 à 11 nucléotides présents dans des boucles formées sur OxyS et sur l'ARNm cible (Fig. II.7) (Altuvia et Wagner, 2000). L'ARN OxyS coordonne l'adaptation à la concentration en peroxyde d'hydrogène et protège ainsi la cellule des dommages oxydatifs. Cet ARN affecte l'expression de plus d'une quarantaine de gènes. Il réprime notamment la traduction de l'ARNm du gène *fhlA* codant un facteur de transcription en interagissant avec une région contenant la séquence Shine-Dalgarno (SD) de l'ARNm. L'interaction sens-antisens bloque de ce fait l'accès à la séquence SD par le ribosome (Fig. II.7). En général, les processus faisant intervenir un *trans*-antisens n'implique pas obligatoirement la dégradation de l'ARNm par une RNase (pour revue, Aiba, 2007).

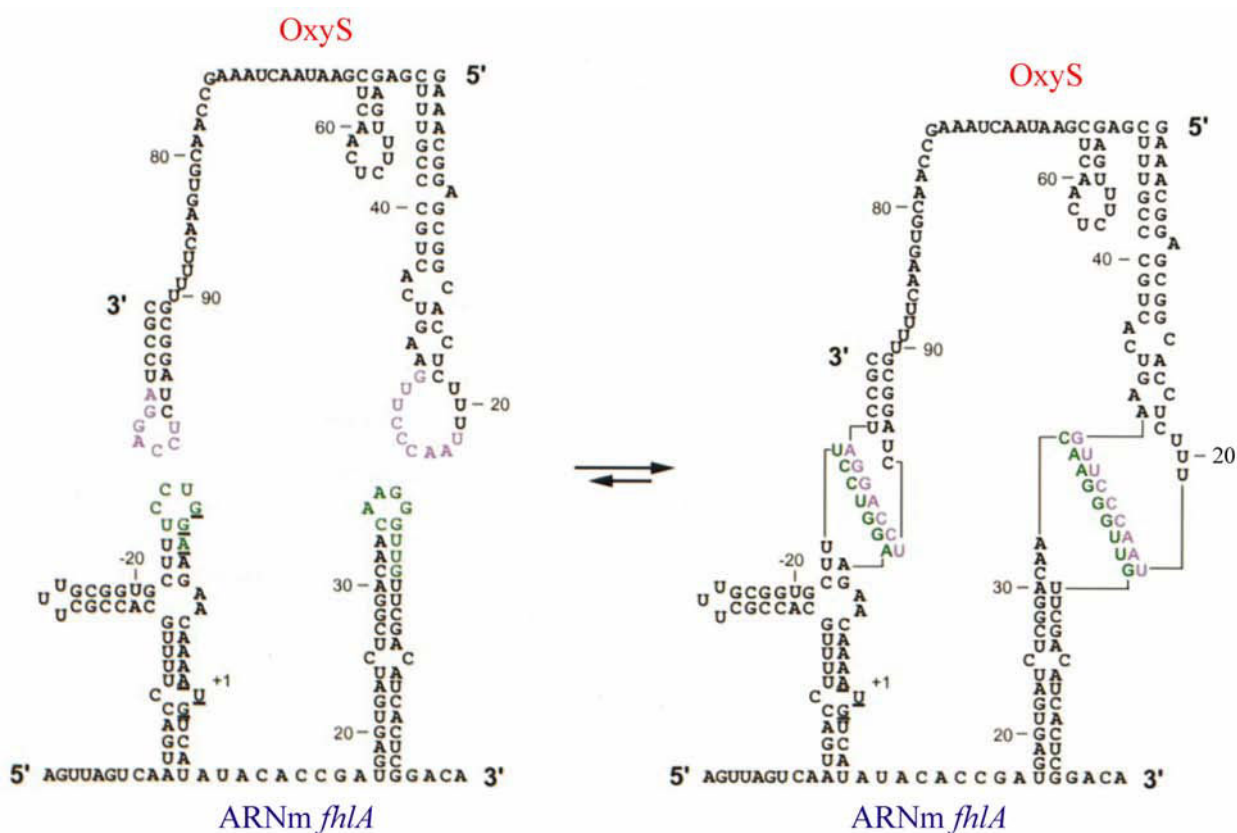


Figure II.7 : Modèle de répression de la traduction par interaction entre le *trans*-antisens OxyS et l'ARNm *fhlA* (d'après Argaman et Altuvia, 2000).

La séquence Shine-Dalgarno et le codon d'initiation (ici +1) sont soulignés. Les séquences complémentaires entre les ARN sont représentées en vert et en pourpre pour, respectivement, *fhlA* et OxyS. Une interaction boucle-boucle initie le contact.

L'hybridation d'un ARN *trans*-antisens dans la région 5'UTR de l'ARNm cible peut également engendrer une activation de la traduction de sa cible. Ainsi, l'interaction de l'ARN DsrA, également exprimée lors de conditions de stress, avec l'ARNm de *rpoS* permet de libérer le site d'initiation de la traduction séquestré dans une tige-boucle et donc de favoriser la production de σ^S (Fig. II.8) (Sledjeski *et al.*, 1996). Notons que DsrA agit comme inhibiteur

post-transcriptionnel auprès du gène *hns* qui code une histone-like qui représente un facteur de régulation global (Repoila *et al.*, 2003). Le rôle de DsrA est donc pléiotropique, comme c'est le cas pour la plupart des ARNnc bactériens. Remarquons que la régulation de l'expression de RpoS est particulièrement intéressante du fait qu'au moins trois ARN antisens, OxyS, DsrA et RprA y sont impliqués. Ainsi, ce recours à des ARN antisens révèle comment la bactérie intègre et répond rapidement aux signaux environnementaux sans utiliser un réseau complexe de facteurs protéiques (Repoila *et al.*, 2003).

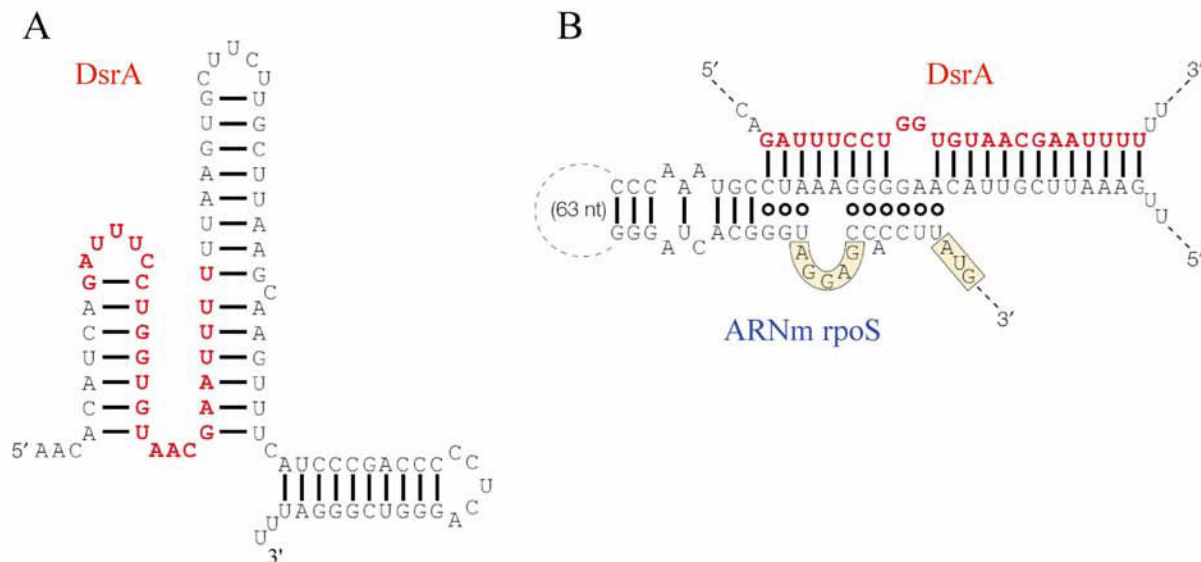


Figure II.8 : Modèle de répression de la traduction par interaction entre l'antisens DsrA agissant en trans et l'ARNm *rpoS* (d'après Eddy, 2001).

A. La structure en trois tige-boucles du *trans*-antisens DsrA s'ouvre pour interagir avec le site d'initiation de la traduction de plusieurs ARNm différents dont l'ARNm *rpoS*. **B.** L'ARNm *rpoS* se structure en masquant la séquence Shine/Dalgarno et le codon d'initiation encadrés. L'antisens DsrA peut être activateur de la traduction de l'ARNm *rpoS* en favorisant une structure secondaire qui libère ce site d'initiation.

On observe que l'action des antisens repose toujours sur la présence d'une séquence précise au sein d'une structure bien définie. L'interaction avec des protéines particulières contribue à des changements structuraux essentiels, pouvant augmenter leur stabilité, ou la stabilité d'un complexe antisens avec la cible. Ainsi, la plupart des *trans*-antisens dont l'action est limitée par le nombre de nucléotides appariés à leur cible, nécessitent la présence de la protéine Hfq (host factor for replication of QB phage). Cette dernière est très abondante, conservée chez la moitié des bactéries et agirait comme chaperonne. En plus de protéger le transcrit d'une dégradation car interagissant au niveau de séquences reconnues par des RNases (notamment la RNase E) et d'affecter la traduction d'ARNm, elle stabilise de nombreuses interactions ARN-ARN (Moller *et al.*, 2002 ; Folichon *et al.*, 2003 ; Rasmussen *et al.*, 2005). Hfq appartient à la famille des protéines Sm-like qui se lie à certains snRNA eucaryotes (Assens *et al.*, 1982). Sa forme en anneau homohexamérique se lie avec la cible au niveau de régions riches en A/U à proximité d'une tige-boucle ou au niveau de la queue polyA. Il est ainsi suggéré que Hfq est capable

d'augmenter le nombre de cibles d'un ARNnc *trans*-antisens (Wassarman *et al.*, 2001 ; Masse et Gottesman, 2002 ; Aiba, 2007; Brennan et Link, 2007). Notons que Hfq contacte des composants du dégradosome et qu'il est envisageable qu'elle ait d'autres partenaires encore inconnus (Storz *et al.*, 2004 ; Brennan et Link, 2007). En contraste, il semblerait que la plupart des *cis*-antisens ne requièrent pas la présence de protéines supplémentaires (pour revue, Brennan et Link, 2007). Cependant, un *cis*-antisens pourrait également agir en *trans* par le biais de protéines associées et d'un appariement partiel (Brantl, 2007). De nombreuses questions concernant le nombre d'appariement requis ou l'implication de protéines qui favoriseraient des appariements partiels restent encore sans réponse.

Le génome d'*E. coli* présente environ 60 gènes d'ARNnc dont l'expression a été mise en évidence expérimentalement mais dont la fonction particulière n'est pas établie. Cependant, des prédictions informatiques laissent entrevoir la présence de plus de 200 ARNnc (Rivas *et al.*, 2001 ; Hershberg *et al.*, 2003 ; Vogel *et al.*, 2003 ; Kawano *et al.*, 2005). Ce nombre de gènes d'ARNnc est à comparer au nombre de gènes de protéines voisin de 4000.

2. Eucaryotes

2.1. ARN antisens

Des ARN antisens naturels (NAT : Natural Antisense Transcripts) existent aussi bien chez les bactéries que chez les eucaryotes (Wagner et Flardh, 2002). En effet, on observe fréquemment, au niveau du locus d'un gène eucaryote, le chevauchement de deux unités transcriptionnelles codées en sens opposés par les deux brins. Ainsi, un brin code un ARNm, tandis que l'autre code un NAT. Notons qu'une partie non traduite d'un ARNm mais aussi un ARNm entier (qui n'est donc plus un ARN non codant *stricto sensu*) peut agir comme NAT. La régulation d'un ARNnc par un NAT est également envisageable. La proportion de gènes de protéines qui seraient soumis à ce type de régulation est estimée à plus de 15% chez l'Homme (pour revue, Lapidot et Pilpel, 2006). Remarquons qu'il existe aussi des NAT agissant en *trans*. Les *cis*-antisens sont classés en trois catégories sur la base de leur orientation par rapport à leur cible : les 5' à 5', les complètement chevauchant et les 3' à 3' (Fig. II.9A). Les appariements de 3' à 3' sont les plus fréquents.

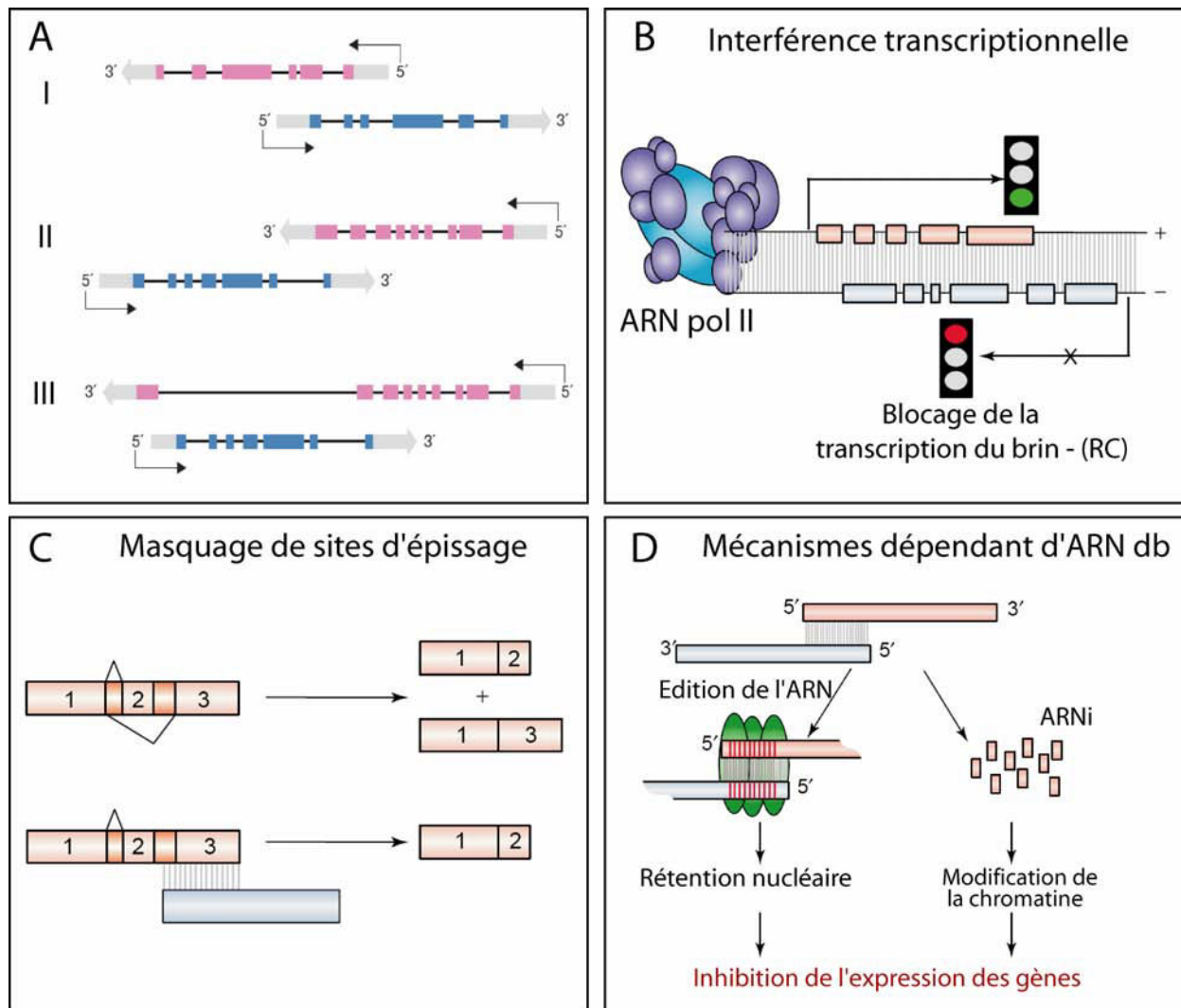


Figure II.9 : Orientations relatives des paires d'ARN antisens agissant en *cis* et différents modes d'action des NAT connus pour réguler l'expression de gènes eucaryotes (d'après Lapidot et Pilpel, 2006). (A.) **I.** Le chevauchement 5' à 5' (« head to head »). **II.** Le chevauchement 3' à 3' (« tail to tail »). **III.** Le chevauchement complet. La représentation de régions UTR en gris et d'exons en couleur souligne l'implication de l'épissage alternatif dans le mécanisme de production des NAT eucaryotes. Les ARN antisens sont représentés en bleu (B.) Phénomène d'interférence transcriptionnelle. Ce phénomène impliquerait la présence simultanée de deux complexes d'ARN Pol sur les brins opposés conduisant à l'inhibition de l'activité d'une ARN pol ou des deux. (C.) Phénomène de masquage d'éléments régulateurs d'un pré-ARNm. L'interaction d'un antisens avec sa cible masque par exemple des éléments requis pour la sélection des exons et des introns et influence ainsi l'épissage alternatif de certains gènes. (D.) Mécanismes dépendant de la production d'ARN db. Lorsque la complémentarité entre l'antisens et sa cible est de taille supérieure à 200 nt, les régions d'ARN db peuvent être éditées par la machinerie d'édition. Les transcrits hyper-édités peuvent être reconnus par des protéines (représentées en vert) qui empêchent leur export du noyau et ainsi empêchent leur traduction dans le cytoplasme. Ces ARN db peuvent également être maturés en petits fragments par la machinerie d'interférence ARN (ARNi).

Une recherche de *cis*-antisens par alignement des ADNc avec les séquences des banques EST montre, chez la souris par exemple, que le nombre de *cis*-antisens pourrait représenter jusqu'à 29% du nombre total de transcripts (Katayama *et al.*, 2005). Une recherche récente de *trans*-antisens dans le génome humain conduit à l'idée que la plupart d'entre eux impliquent des interactions avec des régions UTR qui peuvent être partiellement ou complètement éliminées par épissage alternatif, suggérant ainsi un lien entre régulation par *trans*-antisens et épissage alternatif (Li *et al.*, 2006). Au final, du fait de l'interaction avec plusieurs cibles, les *trans*-antisens ont la possibilité de former des réseaux complexes de régulations (Li *et al.*, 2006).

Deux mécanismes d'action des antisens eucaryotes sont caractérisés. L'interférence transcriptionnelle est le résultat d'une transcription qui se fait dans le même temps à partir des deux brins opposés. Ainsi, les deux ARN pol II impliquées se rencontrent, ce qui entraîne l'arrêt de la transcription des deux brins (Fig. II.9B). C'est le cas, par exemple, des gènes GAL10 et GAL7 (Prescott et Proudfoot, 2002). Les antisens peuvent également, lorsqu'ils sont appariés à leur cible, masquer un élément en *cis* important pour le transport de l'ARNm, sa polyadénylation, sa traduction, sa dégradation ou encore son épissage (Fig. II.9C). Par exemple, un antisens du gène *erbA α* codant un récepteur de l'hormone thyroïdienne fait basculer l'équilibre entre deux variants d'épissage en masquant un site d'épissage (Hastings *et al.*, 1997). Enfin, l'association sens-antisens est notamment à l'origine de phénomènes d'édition et de phénomènes d'interférence ARN (Fig. II.9D).

2.2 miRNA et siRNA

Deux ARN antisens, *lin-4* et *let-7*, identifiés chez le nématode *Caenorhabditis elegans*, sont à l'origine de la découverte d'une vaste famille de microARN (miRNA) présente dans tous les animaux et les plantes étudiés (Lee *et al.*, 1993 ; Lagos-Quintana *et al.*, 2001 ; Lau *et al.*, 2001 ; Lee et Ambros, 2001). Bien qu'initialement répertoriés comme les premiers représentants d'une classe d'ARN temporels (ou small temporal RNA, stRNA) impliqués dans la synchronisation du développement (Pasquinelli *et al.*, 2000), il s'est avéré que des transcrits équivalents sont également impliqués dans de nombreux phénomènes de différenciation cellulaire, dans l'apoptose ou certains métabolismes tels que celui des lipides (Brennecke *et al.*, 2003 ; Xu *et al.*, 2003 ; Yi *et al.*, 2006 ; pour revue, Bartel, 2004). Ces transcrits sont conservés entre espèces éloignées, par exemple au moins un tiers des miRNA de *C. elegans* présentent des homologues chez l'Homme (Lim *et al.*, 2003b). Les miRNA sont très abondants au sein de la cellule, atteignant pour certains 50 000 molécules par cellule de nématode, ce qui est supérieur au nombre de molécules de snRNA U6 impliquées dans les spliceosomes (Lim *et al.*, 2003b). Le nombre de gènes de miRNA différents est estimé chez l'homme à une valeur comprise entre 250 et 800 voire plus (Lim *et al.*, 2003a ; Bartel, 2004 ; Bentwich *et al.*, 2005 ; Costa, 2007).

Les miRNA correspondent à des transcrits de 20 à 25 nt dont une partie est parfaitement ou partiellement complémentaire à un ou plusieurs ARNm cible dont ils influencent la traduction ou la stabilité (Fig. II.10 B). Ils ont en commun la particularité de résulter de deux étapes de maturation spécifiques (Fig. II.10 C). Un premier transcrit appelé pri-miRNA (Lee *et al.*, 2002) dont la taille peut atteindre 1000 nt, est transcrit majoritairement par l'ARN pol II (Lee *et al.*, 2004) soit indépendamment à partir d'un promoteur propre, soit à partir de clusters de gènes de miRNA, soit à partir d'introns de gènes de protéines (dans un quart des cas chez l'homme) (Bartel, 2004). Le pri-miRNA est maturé en pré-miRNA dans le noyau par Drosha, une endonucléase de la famille des RNase III qui définit une extrémité du miRNA mature (Fig. II.10C) (Lee *et al.*, 2003). Le pré-miRNA présente une structure en tige-boucle de 60 à 70 nt couramment utilisée pour la recherche des miRNA. Le pré-miRNA est transporté dans le cytoplasme où il subit une seconde maturation par Dicer, une autre endonucléase de la famille des RNase III qui définit l'autre extrémité du miRNA (Fig. II.10C) (Hutvagner *et al.*, 2001). Notons que les pré-miRNA de plantes sont maturés différemment par une autre endonucléase appelée DCL1 (Dicer-like) avant transport dans le cytoplasme (Reinhart *et al.*, 2002). Chez les animaux comme chez les plantes, un des deux brins est ensuite préférentiellement incorporé (pour revue, Hutvagner, 2005) dans un complexe ribonucléoprotéique effecteur appelé miRNP souvent assimilé au complexe RISC (RNA Induced Silencing Complex). Les miRNP contiennent notamment une protéine de la famille Argonaute, nommée eIF2C2, ainsi que l'hélicase Gemin3 et la protéine Gemin4 (Mourelatos *et al.*, 2002). Le complexe RISC comprend également des protéines de la famille Argonaute, dont Ago2, qui présente un domaine de type RNase H. Elle serait chez les mammifères l'enzyme impliquée dans le clivage de l'ARNm ciblé lorsqu'un duplex parfait est formé entre le petit ARN et l'ARNm (Liu *et al.*, 2004).

Certains complexes RISC contiennent un siRNA (small interfering RNA). Les siRNA ont une taille d'environ 20 nt similaire à celle des miRNA (Elbashir *et al.*, 2001) et sont impliqués dans le phénomène de répression de gènes spécifiques appelé « ARN interférence » (ou RNAi ; (Fire *et al.*, 1998). Ils résultent de la maturation par Dicer (Bernstein *et al.*, 2001) de longs ARN double brin provenant, à la différence des miRNA, soit d'ARN db exogène viral ou introduit lors d'expériences (Fire *et al.*, 1998) soit d'ARN db endogène généré par deux ARN antisens transcrits à partir de deux brins opposés ou par l'activité d'une ARN polymérase ARN dépendant (RdRp) présente chez les nématodes et les plantes (pour revue, Bartel, 2004). Notons que de nombreux siRNA différents peuvent être générés à partir d'un ARN db alors que seuls deux types de miRNA (brin 5' ou 3') sont générés à partir d'un pri-miRNA. Les siRNA endogènes ne sont pas conservés entre espèces éloignées contrairement aux miRNA. Une dernière distinction est l'auto-inhibition caractéristique d'un siRNA capable de réprimer sa propre expression.

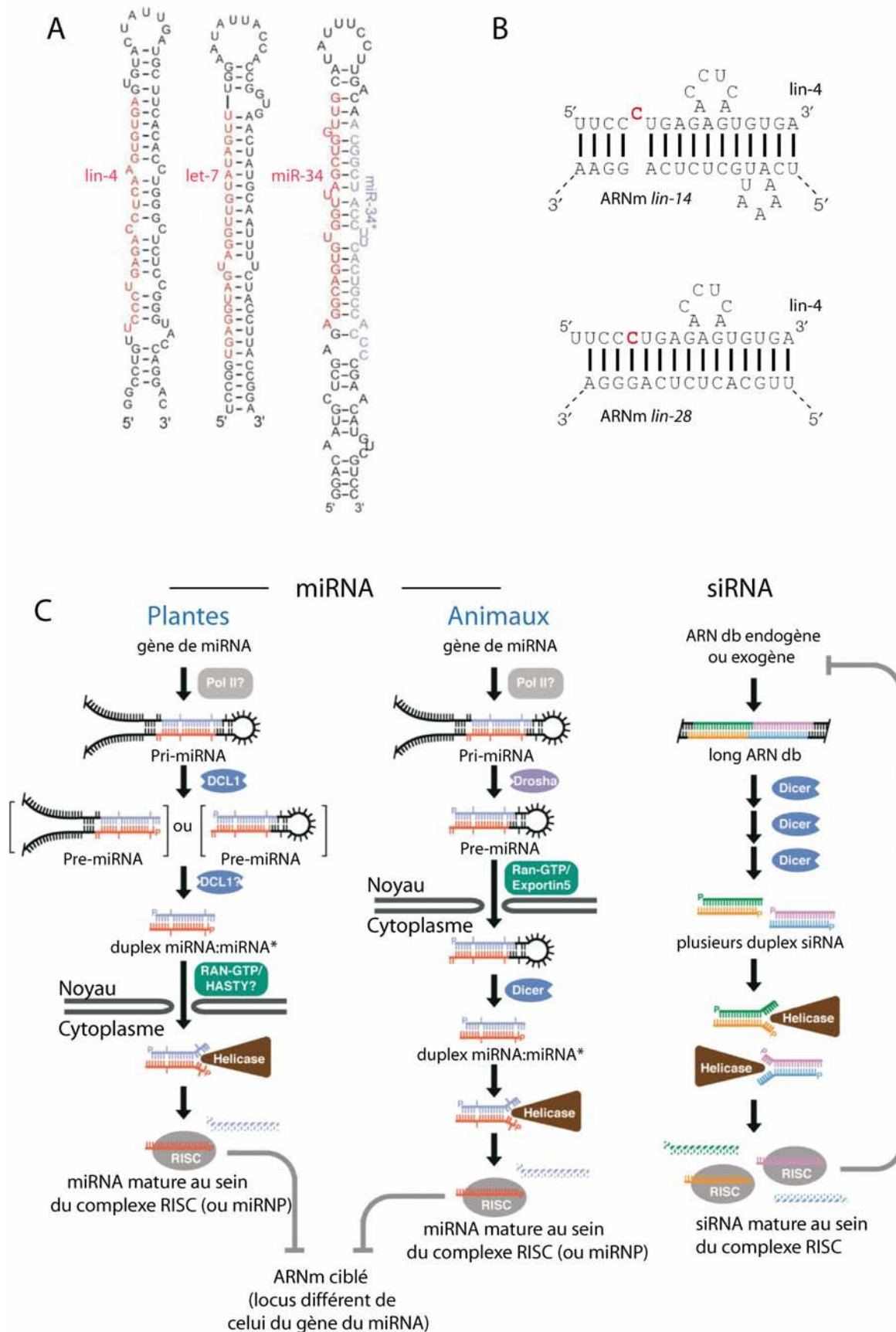


Figure II.10 : Biogenèse des miRNA et siRNA.

A. Exemples de structure en tige-boucle de pré-miRNA d'animaux. Les deux membres fondateurs lin-4 et let-7 ainsi que miR-34 de *C. elegans* sont représentés (d'après Bartel, 2004). **B.** Exemples d'interactions proposées entre le miRNA lin-4 de *C. elegans* et deux de ses ARNm cible (*lin-14* et *lin-*

28), (d'après Eddy, 2001). L'ARN *lin-4* peut s'apparier à sept sites différents dans la région 3' UTR de l'ARNm *lin-14* et peut aussi s'apparier à un site dans la région 3' UTR de l'ARNm de *lin-28*. On remarque la présence d'un résidu C parfois mésapparié qui pourrait être impliqué dans le mode d'action de l'ARN *lin-4*. C. Les différentes étapes de maturation lors de la biogenèse des miRNA et siRNA à partir, respectivement, de pri-miRNA et longs ARN double brin (db). Deux processus différents chez les plantes et les animaux sont indiqués pour les miRNA. Le brin miRNA qui apparaît en rouge est celui qui est préféré lors de l'incorporation aux complexes miRNP ou RISC (le complexe miRNA est souvent assimilé au complexe RISC) (d'après Bartel, 2004).

Néanmoins, les siRNA et les miRNA présentent de nombreuses propriétés communes au niveau de leur biogenèse et du mode d'appariement à leur cible. La définition des siRNA est plutôt fonctionnelle alors que la définition des miRNA est plutôt structurale et la frontière entre ces deux familles n'est pas stricte. En effet, bien que le terme ARN interférence soit associé à un clivage par le complexe RISC comprenant les siRNA, il a été montré que les miRNA sont également capables d'engendrer un clivage en association avec le complexe RISC (Hutvagner et Zamore, 2002). De même, des siRNA peuvent réprimer la traduction sans provoquer de clivage (Zeng *et al.*, 2003). Ainsi, miRNA et siRNA inhibent l'expression de gènes par deux mécanismes similaires, à savoir soit un clivage de l'ARNm ciblé, soit une inhibition de la traduction (Zeng *et al.*, 2003). Finalement, seule la qualité de l'appariement avec la cible et non la nature de l'ARNc (miRNA ou siRNA) déterminerait le mode de répression post-transcriptionnelle. Dans les deux cas, un appariement constant de 6 à 8 nt (« seed sequence » ou « core pairing ») parfaitement complémentaires est observé (Brennecke *et al.*, 2005). Lorsque la complémentarité de l'extension de cet appariement est imparfaite, généralement une dizaine d'appariements au plus interrompus par des mésappariements (Pasquinelli, 2002) la traduction est inhibée sans clivage de l'ARNm tandis qu'une complémentarité parfaite déclenche le phénomène dit d'interférence par clivage au niveau d'une position précise de l'ARNm (Fig. II.11) (Bartel, 2004). Dans ce dernier cas, seul l'ARNm est clivé, ce qui permet au complexe RISC comprenant un siRNA ou un miRNA intact d'agir sur d'autres ARNm (Fire *et al.*, 1998). Une action sans clivage n'empêche pas l'initiation de la traduction mais bloquerait l'avancée des ribosomes, il s'agit donc d'une inhibition plus fine de la traduction. Une recherche de cibles des miRNA a démontré que la région 3'UTR est deux à trois fois plus souvent ciblée que le reste de l'ARNm et il est intéressant d'observer que la taille des régions UTR augmente avec la complexité d'un organisme pour atteindre une taille voisine de la région codante (Brennecke *et al.*, 2005 ; Mattick et Makunin, 2006). Cette recherche a aussi permis de déduire qu'un miRNA présenterait plus d'une centaine de cibles potentielles (Brennecke *et al.*, 2005). Enfin, les nombreuses observations de différents miRNA qui ciblent un même ARNm suggèrent un nombre très élevé de combinaisons possibles à l'origine d'un réseau complexe de contrôles précis de nombreux processus (pour revue, Pasquinelli *et al.*, 2005).

Un troisième mécanisme reste cependant envisageable, il s'agit d'inhibition transcriptionnelle par interaction de complexes RITS (RNA Induced initiation of Transcriptional gene Silencing) avec l'ADN (Fig. II.11) (Verdel *et al.*, 2004). En effet, des siRNA et des

protéines Argonaute ont déjà été observés dans des processus de méthylation de l'ADN, formation de l'hétérochromatine et réarrangements d'ADN chez les plantes, levures et ciliés respectivement, (pour revue, Bartel, 2004).

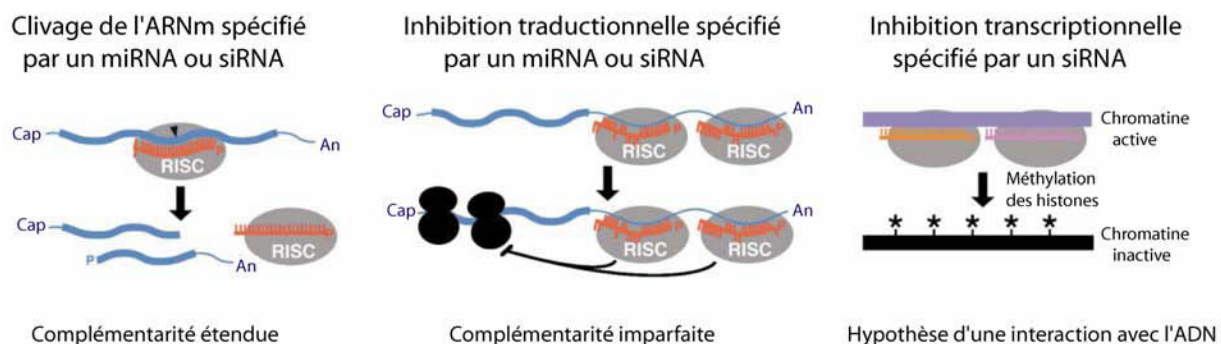


Figure II.11 : Les trois modes d'action proposés pour les miRNA et siRNA (d'après Bartel, 2004). La qualité de la complémentarité entre le siRNA ou miRNA et sa cible semble déterminer l'orientation vers une des deux premières voies.

2.3 gRNA

L'édition est un procédé par lequel un transcrit voit sa séquence altérée par substitution d'un nucléotide par un autre, par délétion ou par addition de nucléotides. Il ne s'agit ici que de résidus A, U, C ou G, contrairement au terme « modification » qui décrit des résidus autres que les 4 canoniques. Les mécanismes d'édition n'impliquent pas forcément un ARNnc. Un exemple bien étudié d'édition impliquant des petits ARN a été découvert dans les mitochondries de protozoaires. Ils sont impliqués dans l'ajout ou le retrait de résidus U dans les ARN mitochondriaux et ont été appelés simplement gRNA pour ARN guides (Blum *et al.*, 1990). Ce phénomène d'édition permet souvent de générer une ORF dans l'ARNm. Le mécanisme, présenté dans la figure II.12 repose sur l'appariement de 8 à 10 pb immédiatement en 3' de la position « éditée », ainsi que l'intervention d'un complexe multi-protéique pas encore complètement caractérisé appelé « éditosome » (pour revue, Stuart *et al.*, 2005). Une étape de clivage par une endonucléase est requise. Puis l'addition ou la délétion de résidus U fait apparaître des appariements supplémentaires avec le gRNA qui joue le rôle de « matrice ». L'inactivation de l'édition est létale pour le protozoaire. Par conséquent, l'édition est une cible pour de nombreuses drogues visant à combattre certains de ces parasites.

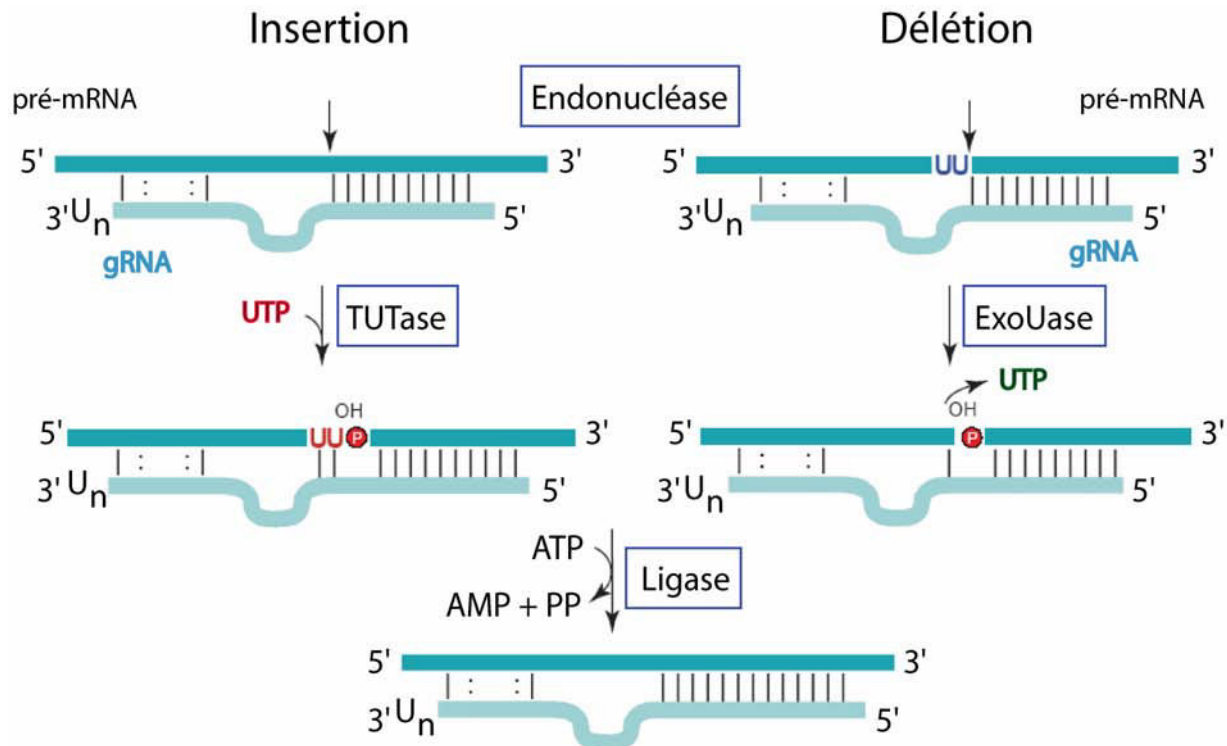


Figure II.12 : Mécanisme général de l'édition par des ARN guides ou gRNA (d'après Stuart *et al.*, 2005).

Deux séries d'évènements différents conduisent à l'insertion ou à la délétion de résidus U. L'édition est réalisée progressivement de 3' vers 5' au sein du pré-ARNm au niveau de sites définis par l'appariement avec le gRNA. Les gRNA possèdent une queue polyU en 3' en interaction avec le pré-ARNm et essentielle pour la réaction d'édition. Un clivage endonucléolytique est réalisé par une endonucléase en aval d'un appariement de 8 à 10 pb entre le pré-ARNm et le gRNA. Des résidus U sont soit rajoutés par une TUTase (Terminal Uridyl Transferase), soit retirés par une ExoUase, selon la séquence du gRNA. Ainsi, après addition ou délétion, l'appariement entre le pré-ARNm et le gRNA se trouve amélioré. Les fragments 5' et 3' sont enfin reliés par une ARN ligase. L'édition est réalisée en plusieurs cycles qui aboutissent à la création, au sein du pré-mRNA, d'une séquence complémentaire au gRNA.

2.4 lRNA

Des ARNnc de grande taille car supérieure à 1000 voire 100 000 nt, appelés « large RNA » ou lRNA, sont impliqués dans la régulation dite épigénétique de groupes de gènes. Une régulation épigénétique correspond à des modifications chimiques, héritées lors de la mitose, concernant l'ADN (méthylation des cytosines) ou les histones qui lui sont associées (notamment méthylation, acétylation, ubiquitinylation et phosphorylation de résidus particuliers). Ils conduisent à un changement de l'expression des gènes en absence d'altération de la séquence de l'ADN (Goldberg *et al.*, 2007). Un des exemples fondateurs de ce groupe est l'ARN Xist (X-inactive specific transcript). D'une taille d'environ 17000 nt, transcrit à partir d'une séquence de plus de 100 000 nt, il induit l'inactivation d'un des deux chromosomes X (Xa/Xi pour actif/inactif) chez les femelles de mammifères pour assurer des quantités égales de produits de gène de ce chromosome entre les femelles et les mâles (compensation de dosage) (pour revues, Heard, 2004 ; Pauler *et al.*, 2007). Ce processus nécessite le comptage des chromosomes X (Bacher *et al.*, 2006), la sélection aléatoire du futur Xi et enfin l'apposition de marques épigénétiques qui maintiennent définitivement l'inactivation des gènes au sein des cellules filles.

Ce processus est initié lors de la différenciation des cellules embryonnaires de la masse interne par le biais de l'ARN Xist, lequel n'a plus de fonction évidente après la mise en place de l'inactivation. Les mécanismes par lesquels l'ARN Xist conduit à la sélection et au silence transcriptionnel de la plupart des promoteurs de Xi ne sont pas encore résolus. Il semblerait que Xist joue un rôle d'échafaudage dans l'établissement de complexes ribonucléoprotéiques qui se propagent à partir de son locus dans les deux sens vers des régions distantes de plusieurs centaines de milliers de nt pour, au final, que ces complexes habillent pratiquement l'ensemble du chromosome Xi (Fig. II.13). La propagation de ces complexes ne ferait pas intervenir de complémentarité de séquence entre l'ARN Xist et le Xi. L'accumulation de l'ARN Xist sur le Xi engendre ainsi, via les protéines associées, le remodelage rapide et irréversible de la chromatine (hétéro-chromatinisation).

Le recours à un ARN, Xist ou autres lRNA, bien que systématique dans tous les processus d'empreinte, n'est pas encore clairement justifié. Il est possible que la grande taille du transcrit autorise de nombreuses interactions spécifiques avec des protéines, et que la transcription d'un ARN directement au niveau de sa zone d'action empêche la diffusion dans tout le noyau de cet élément qui est central au processus.

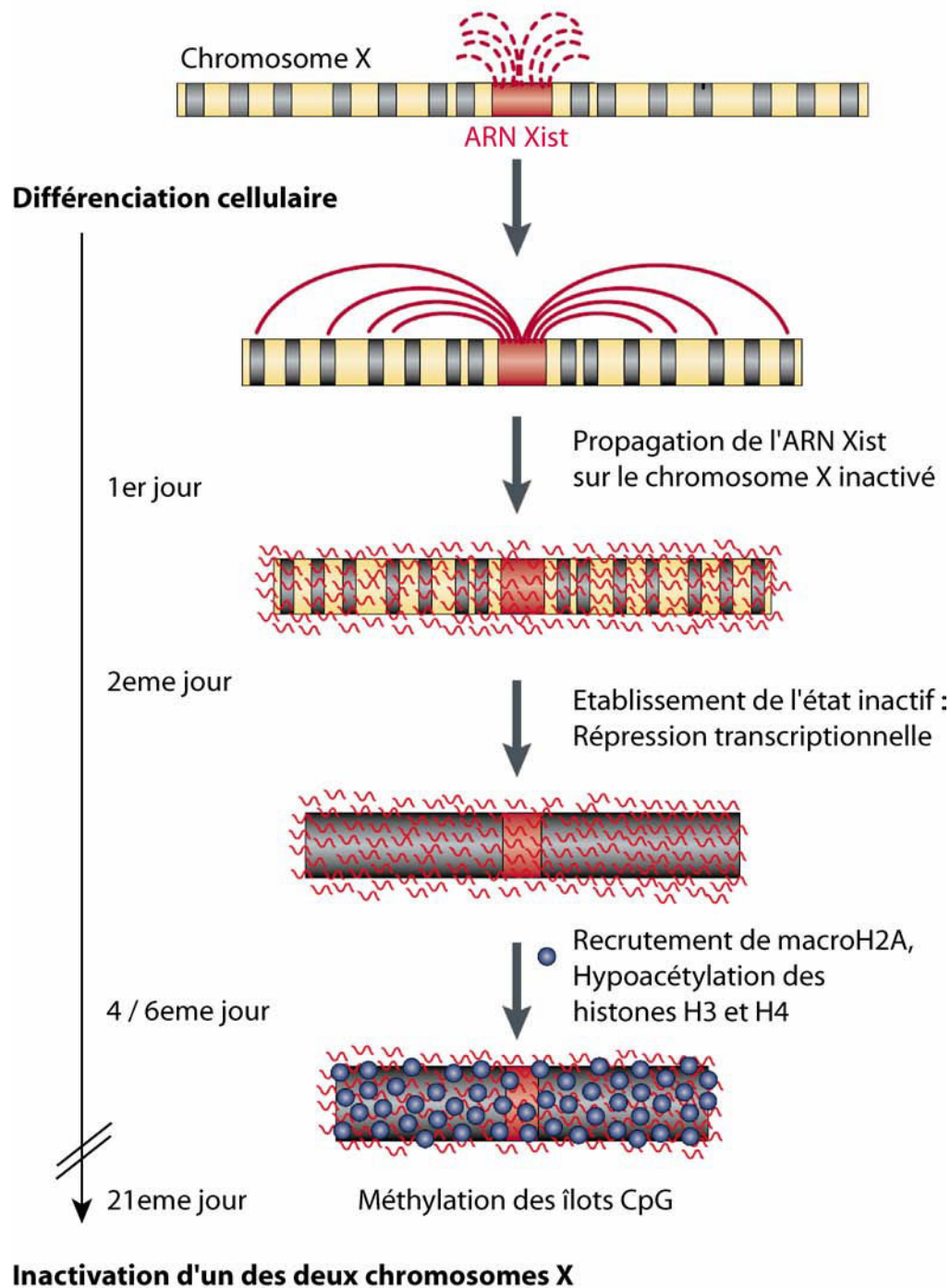


Figure II.13 : Modèle d'inactivation d'un des deux chromosomes X (adapté de Avner et Heard, 2001 ; Brockdorff, 2002).

Avant l'inactivation, l'ARN Xist serait exprimé sous une forme instable (pointillés rouges). Lors de l'initiation de la différenciation cellulaire une forme stabilisée de l'ARN Xist se propage et recouvre l'ensemble des régions du chromosome X à inactiver ce qui entraîne le silence transcriptionnel de la plupart des gènes du chromosome X. L'ARN Xist recrute un variant d'histone, macroH2A, et de nombreuses protéines impliquées dans la condensation de la chromatine en hétérochromatine. L'inactivation du chromosome X devient irréversible avec la méthylation des îlots CpG au 21^{ème} jour après la différenciation cellulaire.

2.5 snRNA

L'épissage est une maturation d'ARNm qui a lieu dans le noyau et qui est catalysée par une macrostructure appelée spliceosome (pour revue, Moore *et al.*, 1993). La réaction d'épissage est initiée par une première attaque nucléophile du phosphate situé à la jonction exon 1-intron en 5' par le groupement OH en 2' du ribose de l'adénosine de la boîte de branchement de l'intron. Ceci libère l'exon 1 avec une extrémité 3'OH et un intermédiaire en lasso constitué de l'intron lié à l'exon 2 (Fig. II.14). La seconde réaction de transestérification est réalisée par l'attaque nucléophile du groupement 3'OH libre de l'exon 1 sur le phosphate de l'extrémité 5' du deuxième exon, ce qui entraîne le clivage de la jonction intron-exon 2 et la ligation des exons 1 et 2. L'intron est libéré sous forme de lasso avant d'être dégradé par des RNases.

La réaction d'épissage fait intervenir 5 petits ARN nucléaires abondants, non polyadénylés, de taille comprise entre 100 et 200 nt et riche en résidus U (domaine de fixation des protéines Sm et Lsm) d'où le nom de U snRNA (pour revue, Valadkhan, 2005). Il s'agit des snRNA U1, U2, U4, U5 et U6, associés à des protéines spécifiques au sein de particules ribonucléoprotéiques appelées UsnRNP, mais également associés à de nombreuses autres protéines (certainement plus de 150 partenaires protéiques ; pour revue, Will et Luhrmann, 1997 ; Will et Luhrmann, 2001 ; Nilsen, 2003). Un autre système mis en évidence chez les métazoaires et les plantes est appelé spliceosome mineur. Ce dernier met en jeu une autre série de UsnRNP : les snRNP U11, U12, U4atac et U6atac, se substituant respectivement aux snRNP U1, U2, U4 et U6 (pour revues, Mount, 1996 ; Tarn et Steitz, 1997). Certains des UsnRNA spliceosomaux sont transcrits par l'ARN Pol II et présentent une coiffe triméthylée en 5', c'est le cas des snRNA U1, U2, U4, U4atac, U5, U7, U11 et U12 qui se lient aux protéines Sm (Matera *et al.*, 2007). D'autres sont transcrits par l'ARN Pol III, portent une coiffe monométhyl-phosphate en 5' et se lient aux protéines Lsm, c'est le cas de U6 et U6atac. Le snRNA U7 est impliqué dans la maturation de l'extrémité 3' des ARNm d'histones. Ce UsnRNA ne fait pas partie du spliceosome. Notons que les UsnRNA subissent des maturations post-transcriptionnelles et sont assemblés en UsnRNP grâce au complexe SMN (Survival of Motor Neurons) (Matera *et al.*, 2007).

Les séquences des UsnRNA spliceosomaux permettent la formation d'appariements intermoléculaires snRNA/ARN pré-messager ou snRNA/snRNA participant à la fois à la formation des complexes spliceosomaux et à la catalyse (Fig. II.14). La conservation dans toutes les espèces de la structure formée entre le snRNA U2 et le snRNA U6 et leurs homologues U12 et U6atac ainsi que leur ressemblance avec le site catalytique des introns de groupe II suggèrent que cette structure intervient directement dans la catalyse (Fig. II.15) (pour revue, Jacquier, 1998). En effet, les interactions entre U2 et U6 ainsi qu'entre U12 et U6atac reposent sur la formation de trois appariements hétérologues (hélices I à III) laissant libre une séquence de U6 (ou U6atac) et une séquence de U2 (ou U12) qui sont respectivement complémentaires d'une séquence à l'extrémités 5' de l'intron et de la boîte de branchement de l'intron.

Ainsi, les UsnRNA participent à l'assemblage des complexes spliceosomaux ainsi qu'à leur activité catalytique.

ARN pré-messager

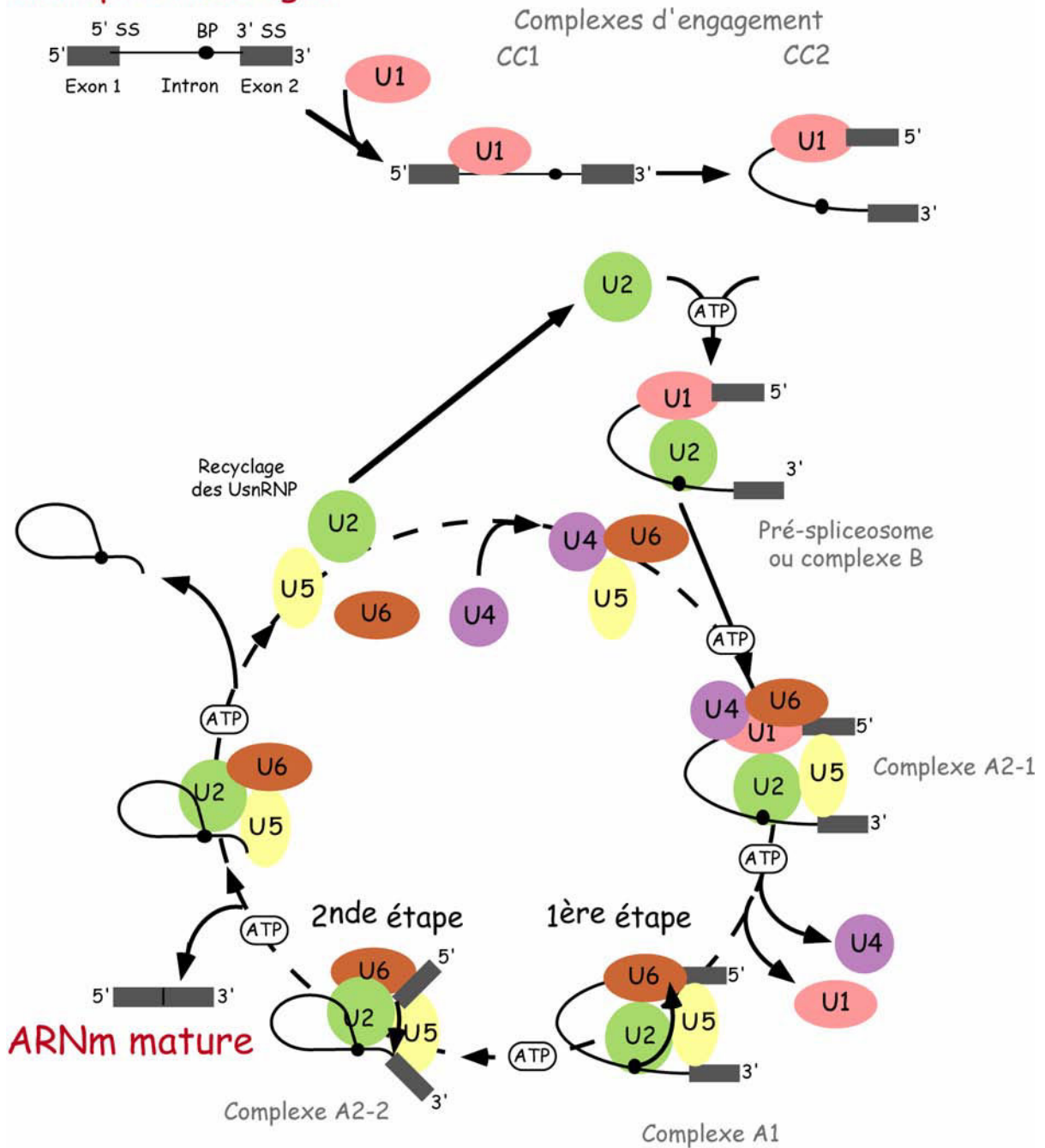


Figure II.14 : Représentation schématique du cycle d'assemblage, d'action et de dissociation du spliceosome chez *S. cerevisiae*.

5'SS et 3'SS représentent respectivement, les sites 5' et 3' d'épissage et BP, la boîte de branchement. Les étapes ATP dépendantes sont indiquées par ATP encadré. Les noms des complexes suivent la nomenclature de *S. cerevisiae*.

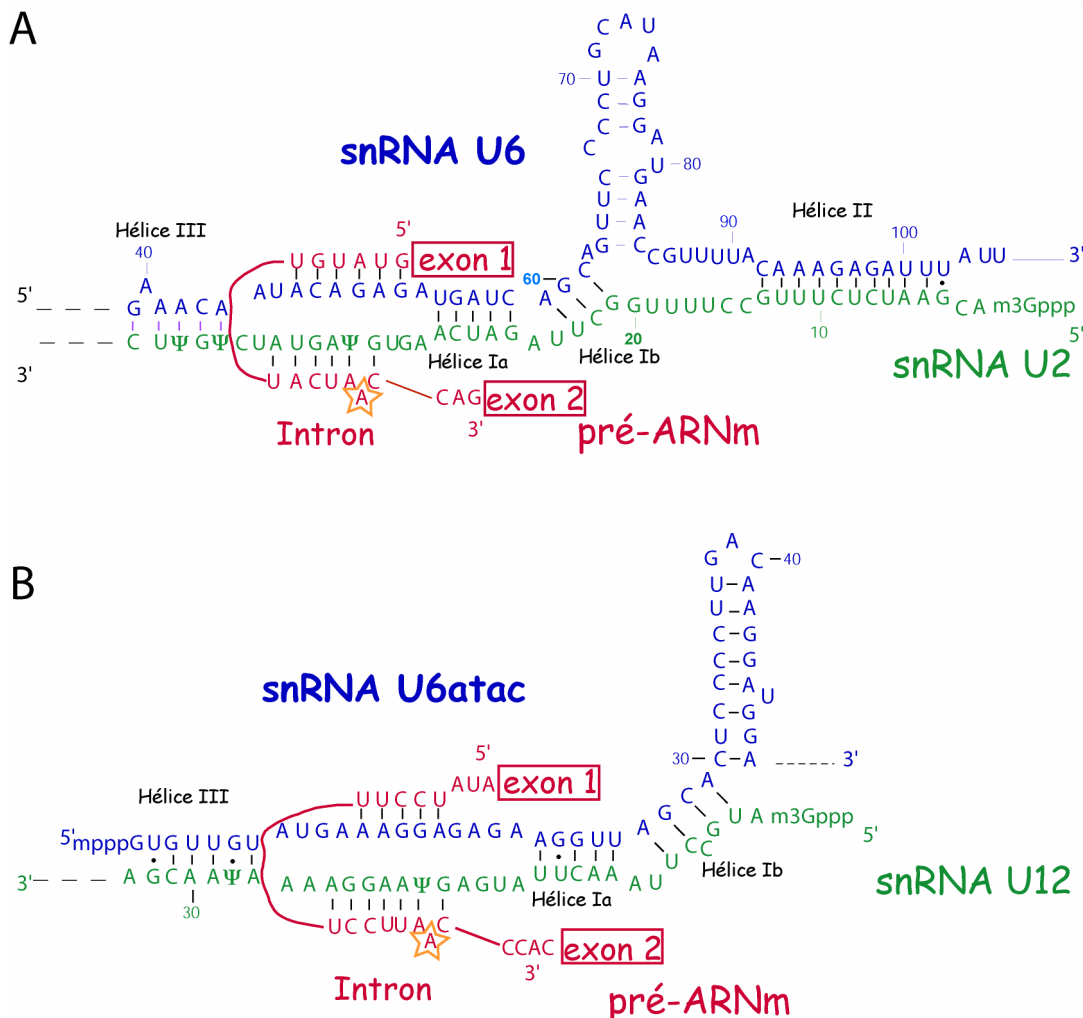


Figure II.15 : Conservation des interactions ARN/ARN impliquant le pré-ARNm et les snRNA U6 et U2 du spliceosome majeur (A) ou U6atac et U12 du spliceosome mineur (B) de *S. cerevisiae*. Les appariements mis en place entre les UsnRNA et le pré-ARNm ici favorisent le rapprochement du groupement 2'OH du résidu A de la boîte de branchement (représenté sous forme d'étoile) avec l'extrémité 3' de l'exon 1, mais aussi le rapprochement de l'extrémité 3' de l'exon 1 avec l'extrémité 5' de l'exon 2, jouant ainsi un rôle catalytique dans les deux étapes de trans-estérification.

2.6 snoRNA et scaRNA

Une abondante famille d'ARNnc d'eucaryotes, dont la taille est comprise entre 50 et 400 nt, est impliquée dans les modifications et la maturation des précurseurs des ARNr. Ces ARNnc sont localisés dans le nucléole à proximité des gènes d'ADNr d'où le nom de snoRNA (small nucleolar). Les ARNr constituent plus de 80% des ARNs totaux d'une cellule eucaryote. La transcription des pré-ARNr est très régulée via le remodelage de la chromatine, la régulation transcriptionnelle et la maturation post-transcriptionnelle évoquée ci-après. Chez tous les eucaryotes, 3 des 4 ARN des ribosomes cytoplasmiques eucaryotes (5,8S, 18S et 26/28S) proviennent de la transcription par l'ARN pol I d'un précurseur d'ARNr unique (pré-ARNr 35S chez la levure ou 45S chez les mammifères) (Fig. II.16).

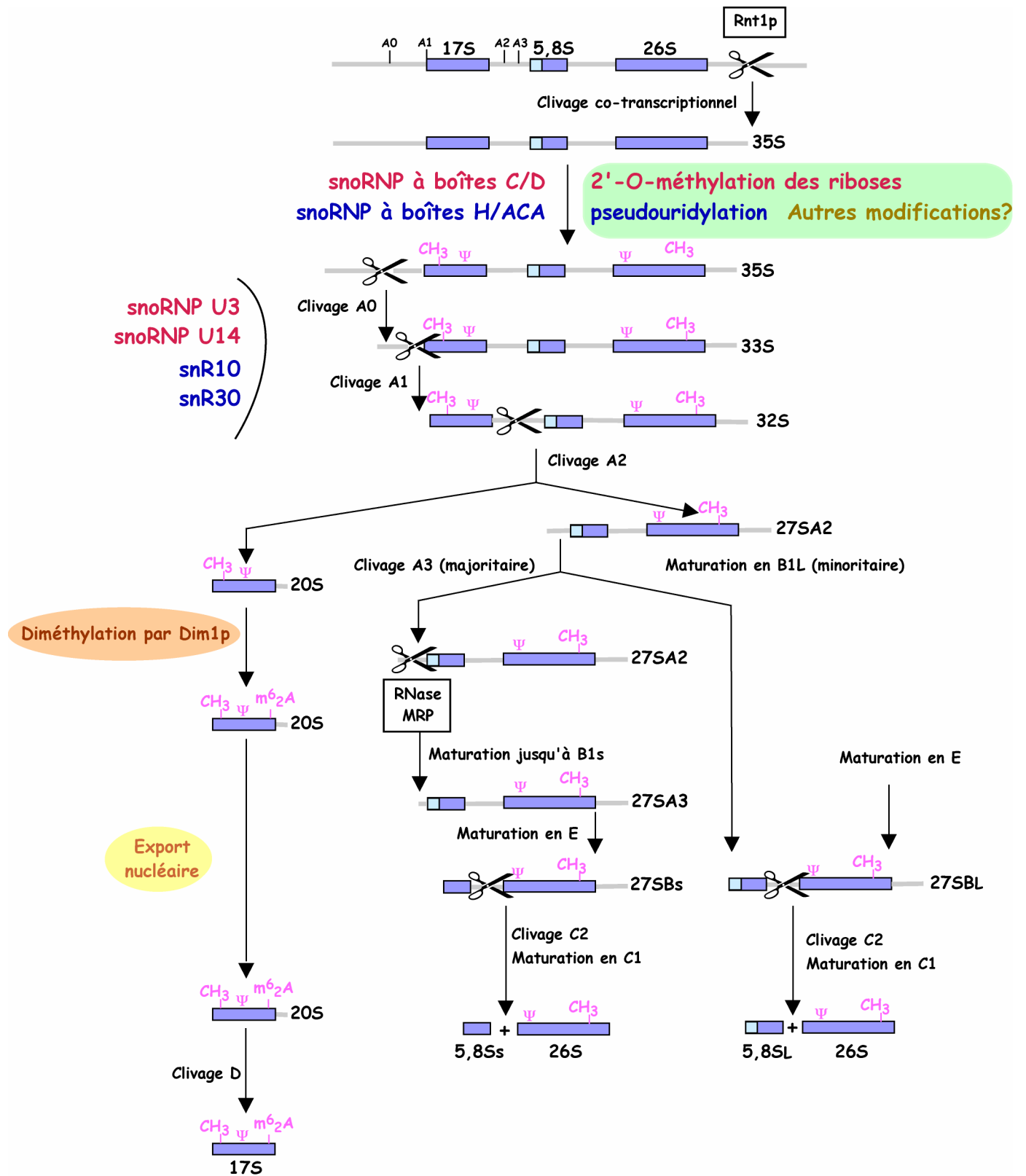


Figure II.16 : Maturation du précurseur pré-ARNr de *S. cerevisiae* (adapté de Kressler *et al.*, 1999; Venema et Tollervy, 1999).

Les nombreuses étapes permettant la maturation du pré-ARNr en trois ARNr, 17S, 5,8S et 26S sont présentées.

Ce précurseur est mûré par un processus complexe impliquant une cascade de clivages endonucléolytiques et de digestions par des exonucléases (pour revues, Kressler *et al.*, 1999; Venema et Tollervey, 1999). La majorité des composants du ribosome s'assemblent au niveau du nucléole pour former les particules pré-ribosomiques. Celles-ci traversent le nucléoplasme et sortent du noyau pour rejoindre le cytoplasme où les ribosomes finissent leur maturation.

De manière concomitante avec la série de clivages, le pré-ARNr subit dans le nucléole des modifications de ses résidus par, notamment, pseudouridylation et 2'-O-méthylation des riboses. Les snoRNA qui sont impliqués dans ces modifications sont nombreux (au moins 150 types différents chez un mammifère). Il s'agit, d'une part, des snoRNA à boîtes C/D qui guident des 2'-O-méthylations des riboses (pour revue, Eliceiri, 1999 ; Dennis *et al.*, 2001). D'autre part, des snoRNA à boîtes H/ACA dirigent des événements de pseudouridylation (pour revue, Bachellerie *et al.*, 2002 ; Kiss, 2002). Certains snoRNA présentant des boîtes C/D (snoRNA U3, U14 chez *S. cerevisiae*, U8 chez les vertébrés) ou des boîtes H/ACA (snR10 et snR30 chez *S. cerevisiae*) interviennent dans des clivages conduisant la maturation du précurseur pré-ARNr. Ainsi, les snoRNA U14 et snR10 sont connus pour intervenir à la fois dans les étapes de clivage et de modification du pré-ARNr (pour revue, Tollervey et Kiss, 1997).

Notons que l'ARN de la MRP agissant en tant que ribozyme dans ce processus, est un complexe ARN/protéine dont huit des neuf protéines sont aussi retrouvées dans la RNase P. Ce dernier appartiendrait donc à une troisième famille comportant l'ARN de la RNase P. Par ailleurs, des mutations de l'ARN de la RNase MRP sont associées à une maladie humaine appelée « cartilage-hair hypoplasia » dont les symptômes sont notamment des déficiences immunitaires et une prédisposition à l'apparition de certains cancers (Ridanpää *et al.*, 2001).

Les sous-unités ribosomiques bactériennes assemblées *in vitro* à partir d'ARNr hypo-modifiés sont beaucoup moins actives que celles assemblées à partir d'ARNr naturels, ce qui suggère un rôle primordial des modifications dans leur fonction (Green et Noller, 1996 ; Khaitovich *et al.*, 1999). Les découvertes récentes montrant que des mutations dans les composants des snoRNP sont responsables de maladies génétiques ont suscité un regain d'intérêt pour leur étude. Les rôles des 2'-O-méthylations des riboses et des pseudouridylations seront discutés dans cette partie mais aussi dans la partie III et dans la discussion de nos résultats.

2.6.1. *snoRNA à boîtes C/D*

Définition et mode d'action

Les snoRNA à boîtes C/D se définissent par la présence de deux séquences conservées : les boîtes C (consensus : 5'-RUGAUGA-3', où R est une purine) et D (consensus : 5'-CUGA-3'), respectivement situées à proximité des extrémités 5' et 3' de l'ARN (Fig. II.17). Ces boîtes sont requises pour la stabilité de ces snoRNA, leur localisation nucléolaire, leur biogenèse et leur fonction au sein de snoRNP (Caffarelli *et al.*, 1996 ; Cavaillé *et al.*, 1996 ; Kiss-Laszlo *et al.*, 1996 ; Watkins *et al.*, 1996 ; Xia *et al.*, 1997 ; Bachellerie et Cavaillé, 1998 ; Lange *et al.*, 1998

; Narayanan *et al.*, 1999 ; Speckmann *et al.*, 2000 ; Watkins *et al.*, 2000) ; pour revues, Tollervey et Kiss, 1997 ; Weinstein et Steitz, 1999 ; Bachellerie *et al.*, 2002 ; Kiss, 2002 ; Omer *et al.*, 2003 ; Henras *et al.*, 2004 ; Matera *et al.*, 2007). Les boîtes C et D sont le plus souvent rapprochées par la formation d'une hélice terminale caractéristique résultant d'un appariement de 4 à 10 pb des extrémités 5' et 3'. L'absence de cette hélice terminale dans quelques snoRNA pourrait être expliquée par la présence d'une hélice « externe » éliminée après maturation (Darzacq et Kiss, 2000 ; Villa *et al.*, 2000). Les snoRNA à boîtes C/D portent également en leur centre un second couple de boîtes C/D nommé couple C'/D' dont la séquence diverge souvent légèrement de la séquence consensus (Kiss-Laszlo *et al.*, 1998). La distance qui sépare les boîtes C' et D' varie entre 3 et 9 nt voire plus mais dans ce cas une structure tige boucle rapproche ces boîtes (Fig. II.17).

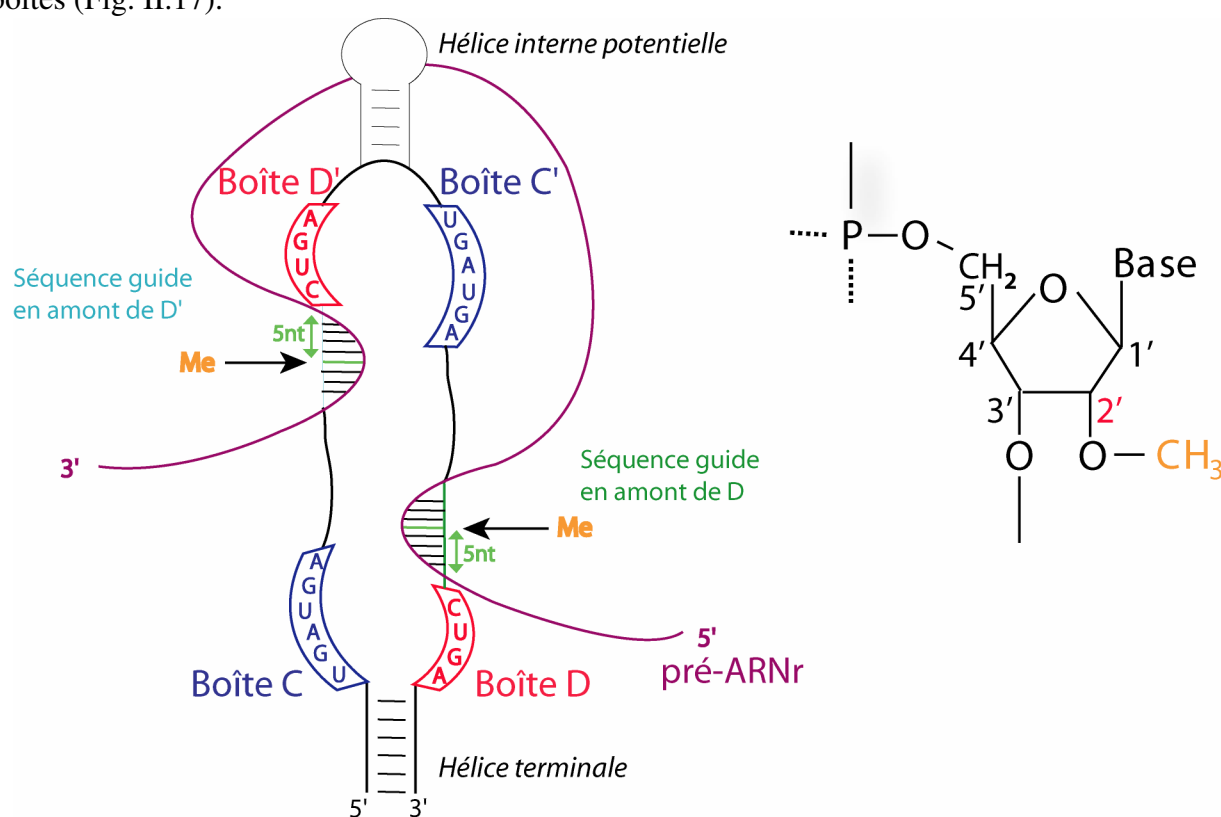


Figure II.17 : Représentation schématique d'un snoRNA à boîtes C/D guidant la méthylation en position 2'OH des riboses. Deux couples de boîtes, C/D et C'/D' sont présents aux extrémités 5' et 3' et au centre de la séquence, respectivement. Les séquences guides, en amont de la boîte D et en amont de la boîte D', complémentaires à deux séquences du pré-ARNr sont présentées. Le nucléotide dont le ribose sera méthylé est toujours celui qui s'apparie avec le 5^{ème} nucléotide en amont des boîtes D ou D'. La méthylation du ribose est indiquée en orange.

Les snoRNA à boîtes C/D contiennent une ou deux séquences guides potentielles qui sont complémentaires sur 10 à 21 nt à la séquence du pré-ARNr (Bachellerie et Cavaille, 1998). Un appariement canonique d'au moins 9 pb sans mésappariement semble indispensable à leur fonction de guide (Chen *et al.*, 2007). Les snoRNA à boîtes C/D dirigent la méthylation de telle sorte que le nucléotide substrat du pré-ARNr est toujours apparié avec le 5^{ème} nucléotide en

amont de la boîte D ou de la boîte D' du snoRNA (Fig. II.17) (Cavaillé *et al.*, 1996 ; Kiss-Laszlo *et al.*, 1996 ; Nicoloso *et al.*, 1996). Une mutation qui empêche la formation de la paire de bases entre le snoRNA et sa cible au niveau du nucléotide à modifier entraîne un défaut de méthylation (Omer *et al.*, 2002). La nature de la séquence qui précède les boîtes D et D' est donc le seul élément déterminant le ou les sites de méthylation. De ce fait, un snoRNA guide une ou deux méthylation(s) différente(s), selon la fonctionnalité des boîtes C'/D'. En effet, chez la levure *S. cerevisiae*, comme chez les autres eucaryotes, seuls 20% des snoRNA à boîtes C/D semblent avoir des boîtes C'/D' fonctionnelles (Kiss-Laszlo *et al.*, 1998 ; Lowe et Eddy, 1999). Les connaissances actuelles font état de 55 résidus 2' O-méthylés détectés dans les ARNr de la levure *S. cerevisiae* dont 51 seraient guidés par des snoRNA à boîtes C/D (Lowe et Eddy, 1999), contre 86 modifications de ce type guidées par 66 snoRNA à boîtes C/D chez *A. thaliana* (Barneche *et al.*, 2001 ; Brown *et al.*, 2001 ; Qu *et al.*, 2001) et 105 à 107 méthylations guidées par 93 snoRNA à boîtes C/D chez les mammifères (pour revue, Bachellerie *et al.*, 2002) suggérant un nombre de snoRNA croissant avec la complexité des eucaryotes. La majorité des snoRNA à boîtes C/D ciblent le pré-ARNr, cependant d'autres molécules d'ARN telles que les ARNm sont également des cibles potentielles. Ainsi, le snoRNA à boîtes C/D HBII-52, exprimé dans le cerveau humain et conservé chez la souris (MBII-52), présente une complémentarité de séquence avec l'ARNm du récepteur 5-HT_{2C} de la sérotonine au niveau d'une région subissant un phénomène d'édition (A en inosine) et un épissage alternatif.

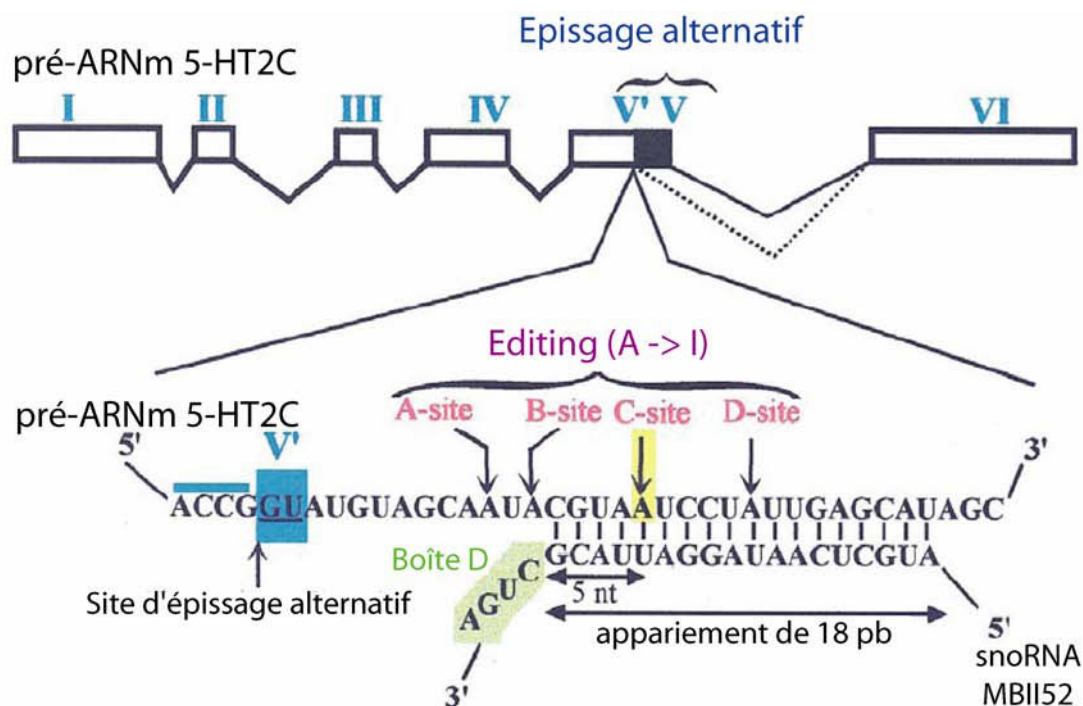


Figure II.18 : Le snoRNA à boîtes C/D MBII52 de souris cible un ARNm (d'après Bachellerie *et al.*, 2002). L'architecture de l'ARNm du récepteur de la sérotonine 5-HT_{2C} est représentée schématiquement. En dessous est représenté l'appariement suggéré entre la séquence guide en amont de la boîte D de l'ARN MBII52 et une séquence de l'exon V qui subit des éditions A->I à 4 sites distincts. La méthylation probable réalisée par le biais de l'ARN guide MBII52 a lieu à la position indiquée en jaune, le site d'épissage alternatif est indiqué en bleu.

Ceci suggère une régulation possible de ces étapes post-transcriptionnelles par les snoRNA (Fig. II.18) (Cavaille *et al.*, 2000 ; Huttenhofer *et al.*, 2001). A noter que l'édition entraîne un changement de la séquence peptidique du récepteur et donc une modification de ses propriétés biologiques. Le fait que, pour de nombreux snoRNA à boîtes C/D, aucune cible potentielle n'ait été détectée dans les ARNr suggère que plusieurs snoRNA à boîtes C/D pourraient être impliqués dans la modification d'ARNm, à moins que la fonction de ces snoRNA soit autre que de guider des événements de méthylation (pour revue, Bachellerie *et al.*, 2002).

Rôle des 2'-O-méthylations

Il est probable que les 2'-O-méthylations des riboses influencent la conformation locale de l'ARNr. Elles favorisent en effet la conformation 3' endo du ribose, ce qui facilite l'empilement des nucléotides et donc les appariements de bases et aussi les interactions avec les protéines associées (Davis, 1998). La localisation de résidus modifiés en position 2'-O des riboses dans les ARNr a montré leur forte concentration dans les régions fonctionnellement importantes et hautement conservées. Alors que l'absence d'un site ou de quelques sites de méthylation n'a pas de conséquence, l'absence de tous les sites provoque un défaut d'assemblage des sous-unités ribosomiques (pour revue, Omer *et al.*, 2003). Cette modification contribue donc à la modulation de la biogenèse et de l'activité des ribosomes. Il est également envisageable que la fonction des snoRNA soit associée à l'appariement en lui-même. Cet appariement permettrait de favoriser ou défavoriser un repliement donné en rapprochant ou empêchant le contact entre deux segments de l'ARNr. Dans ce dernier cas, les protéines associées aux snoRNA ne joueraient qu'un rôle de régulateur ou de chaperon pour la formation de cet appariement (pour revue, Omer *et al.*, 2003). Le rôle de guide ou d'antisens des snoRNA à boîtes C/D sera discuté en comparaison à celui des snoRNA à boîtes H/ACA tout au long de ce chapitre.

Production des snoRNA à boîtes C/D

La majorité des snoRNA à boîtes C/D décrits chez les vertébrés sont produits par maturation d'introns (pour revue, Filipowicz et Pogacic, 2002). Ces introns proviennent de gènes codant des protéines mais également de gènes non codant. Récemment, il a été proposé qu'un couplage entre maturation des introns par le spliceosome et la maturation des snoRNA à boîtes C/D dépendraient notamment de la protéine IBP160, facteur général d'épissage, ce qui expliquerait le positionnement de nombreux gènes de snoRNA à boîtes C/D à une distance quasi-constante de la boîte de branchement (Hirose *et al.*, 2006). Chez la levure, au contraire, la plupart des snoRNA à boîte C/D sont transcrits indépendamment à partir de précurseurs mono- ou polycistroniques. Le regroupement des gènes de snoRNA à boîtes C/D en clusters, parfois mêlés à des gènes de snoRNA à boîtes H/ACA, est également fréquent chez les plantes et les trypanosomes (pour revue, Bachellerie *et al.*, 2002). Avant de s'accumuler dans le nucléole, les snoRNA à boîtes C/D transitent par les corps de Cajal. Les corps de Cajal ou « cajal bodies » (CB ou encore « coiled body ») sont des organelles nucléaires localisées à proximité du nucléole et impliqués dans la biogenèse des snRNA et des snoRNA (Matera, 1999). La biogenèse des

snoRNA ainsi que leur parcours dans les différents compartiments nucléaires sont résumés en figure II.19.

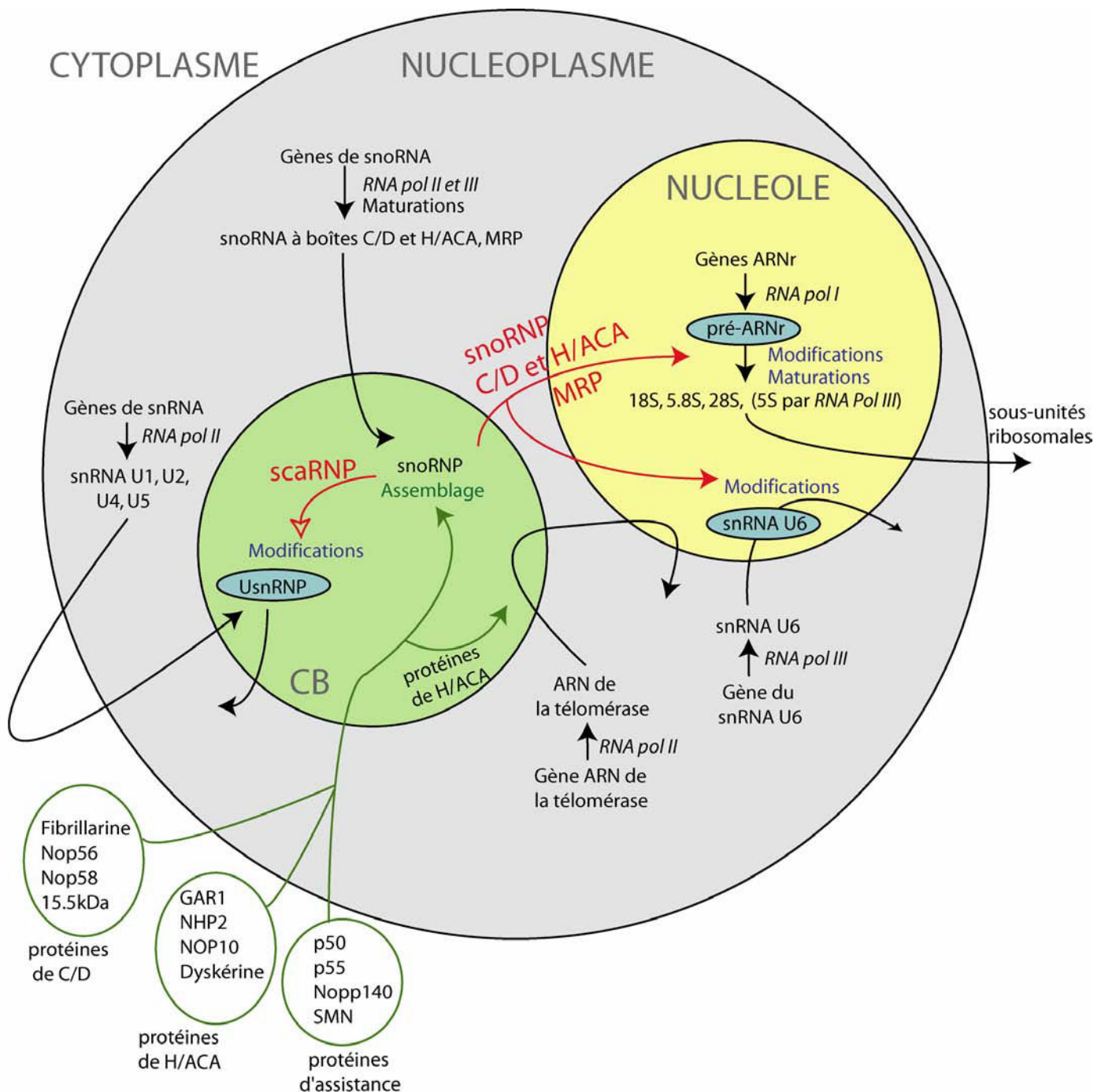


Figure II.19 : Biogenèse, trafic et fonctions des snoRNP chez les vertébrés (d'après Filipowicz et Pogacic, 2002).

Les snoRNP sont indiquées en rouge, leurs cibles et leurs actions sont indiquées en bleu. Des protéines jouant un rôle dans la biogenèse et le trafic des snoRNP, dont la liste n'est pas exhaustive, sont indiquées en vert.

Notons que certains snoRNA à boîtes C/D humains sont exprimés de manière tissu-dépendante (Huttenhofer *et al.*, 2001) et sont soumis à des régulations épigénétiques, c'est le cas de HBII-52 précédemment évoqué (Cavaille *et al.*, 2000). Le syndrome de Prader-Willi caractérisé par une insuffisance de tonus musculaire (hypotonie sévère), un développement sexuel incomplet, des difficultés d'apprentissage, et des problèmes comportementaux comme la recherche permanente de nourriture, due à l'absence de sensation de satiété, qui conduit à une obésité morbide sévère est associé à une délétion de la région du chromosome 15 paternel. Cette région comporte des gènes codant 6 snoRNA distincts à boîtes C/D, dont 47 répétitions en tandem du gène de HBII-52 au sein d'une région qui n'est pas exprimée par le chromosome maternel en raison de l'empreinte paternelle. Les patients atteints de ce syndrome n'expriment pas le snoRNA à boîtes C/D (ainsi que deux autres types de snoRNA) et n'expriment qu'une quantité réduite de récepteurs 5-HT_{2C} à la sérotonine fonctionnels (Fig. II.18).

K-turn

Dans la majorité des snoRNA à boîtes C/D, le couple de boîtes C/D a la capacité de se replier en un motif structural particulier : le motif en K-turn (Fig. II.20) (Watkins *et al.*, 2000 ; Klein *et al.*, 2001 ; Marmier-Gourrier *et al.*, 2003). Ce motif en K-turn est aussi présent plusieurs fois dans les ARNr mais aussi dans le snRNA U4. La structure 3D de ce motif a été établie par analyse radiocristallographique du complexe formé entre la protéine p15,5kD et le snRNA U4 (Vidovic *et al.*, 2000) et également grâce à l'étude structurale de la grande sous-unité ribosomale de *H. marismortui* dont l'ARN 23S comporte 6 motifs K-turn (Kt-7, Kt-15, Kt-38, Kt-42, Kt-46 et Kt-58) (Ban *et al.*, 2000) et de la petite sous-unité ribosomale de la bactérie *T. thermophilus* dont l'ARN 16S comporte deux K-turn (Kt-11 et Kt-23) (Wimberly *et al.*, 2000). L'utilisation de la nomenclature Leontis-Westhof permet de détailler plus précisément ce motif (Leontis et Westhof, 2001 ; Lescoute *et al.*, 2005). On distingue dans les K-turn des snoRNA à boîtes C/D, par analogie avec ceux observés dans le ribosome et le snRNA U4, deux hélices encadrant une boucle interne de trois nucléotides comprenant un résidu U, un résidu purinique R et n'importe quel nucléotide N (Fig. II.20C). L'hélice I est composée d'appariements canoniques de type *cis* Watson-Crick, alors que la seconde comporte des appariements non canoniques, en particulier, elle est terminée le plus souvent par deux paires G.A de type Sucre – Hoogsteen en *trans*. Dans la structure proposée pour un K-turn issu de sRNA à boîtes C/D, l'empilement du A13 avec le résidu N60 permet une interaction Sucre-Sucre en *trans* (Fig. II.20C). Il en est de même pour les résidus N9 et A59. Ces interactions induisent une courbure prononcée (entre 45 et 65°) de la chaîne ribose-phosphate de l'ARN (Moore *et al.*, 2004), à l'origine de la dénomination de « Kink-turn ». Dans cette architecture, la base du résidu U11 est rejetée vers l'extérieur (Klein *et al.*, 2001). Trois motifs composites identifiés dans la grande sous-unité ribosomale (Kt-4-5, Kt-77-78 et Kt-94-99) présentent les mêmes appariements et empilements mais font intervenir, comme leur nom l'indique, deux régions hélicoïdales différentes de l'ARNr.

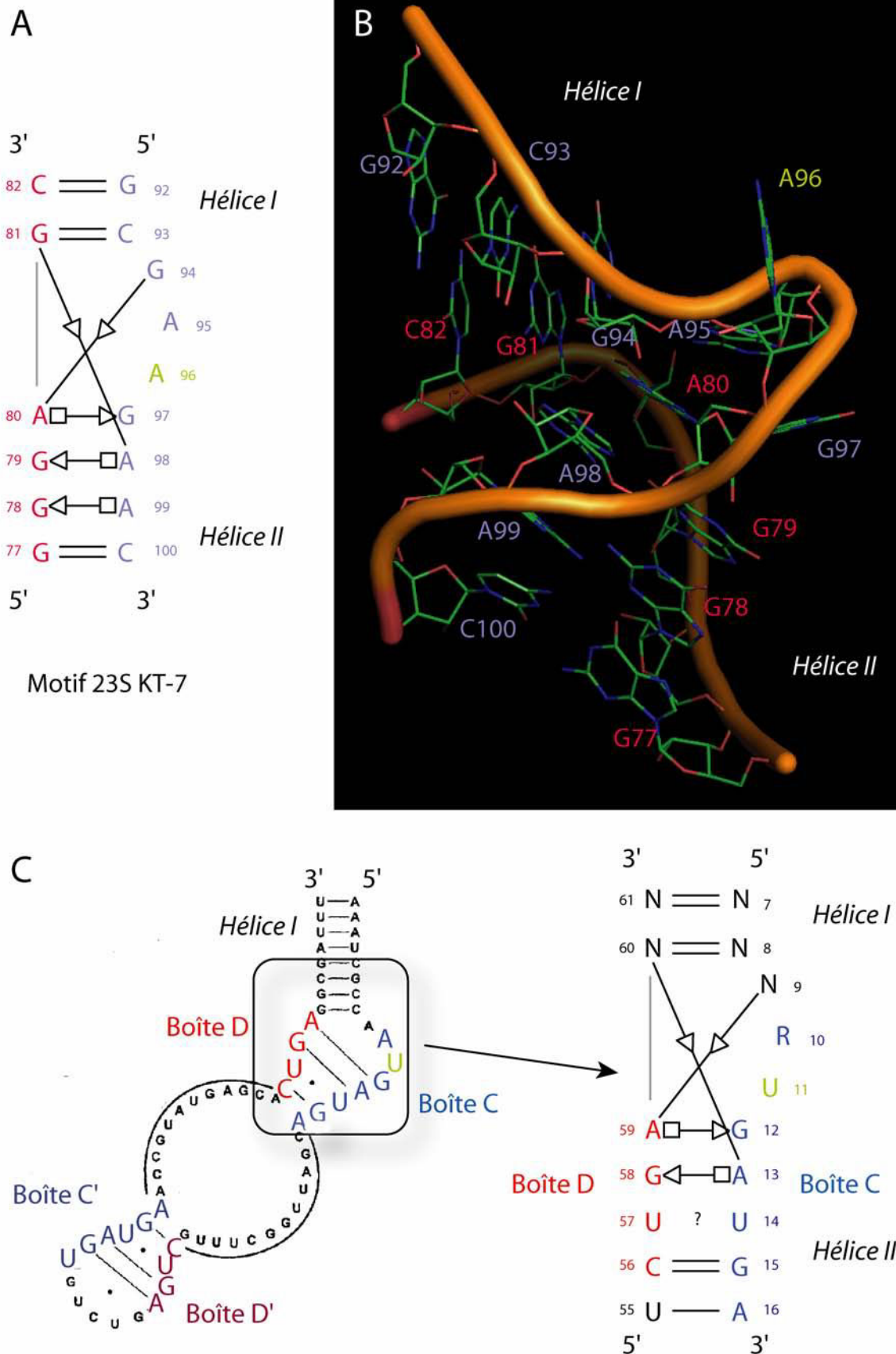


Figure II.20 : Structure d'un motif en K-turn.

A. Le motif en K-turn KT-7 provenant de la structure de la partie 23S de l'ARNr de *H. marismortui* est annoté selon la nomenclature Leontis-Westhof (Lescoute *et al.*, 2005). Les appariements canoniques *cis* Watson-Crick sont représentés par un trait pour une paire A-U et deux traits pour G-C. Un triangle représente l'implication, dans une interaction entre deux nucléotides, de la face sucre tandis qu'un carré

représente l'implication de la face Hoogsteen. Un symbole placé au milieu du trait représentant l'interaction entre deux nucléotides signifie l'implication des mêmes faces. Des symboles vides traduisent des interactions en *trans* et des symboles pleins traduisent des interactions en *cis*. **B.** Visualisation à l'aide du programme Pymol (Warren L. Delano) de la structure tridimensionnelle du KT-7. La chaîne phospho-ribose est représentée pour souligner la formation d'une courbure dans l'hélice ARN. Le nucléotide A96 indiqué en jaune est rejeté vers l'extérieur de la structure. **C.** Analogie entre un K-turn formé dans un snoRNA à boîtes C/D et les K-turn décrits dans les ribosomes. Aucune structure de K-turn provenant de snoRNA à boîtes C/D eucaryote n'est disponible.

Notons que le terme « kink turn » traduisant une courbure de la chaîne ribose-phosphate n'est pas adapté à ces motifs composites. Il est à noter également que des motifs appelés « K-loop » correspondent à des K-turn dépourvus de l'hélice I (Charron *et al.*, 2004 ; Nolivos *et al.*, 2005).

snoRNP

Les snoRNP à boîtes C/D sont composées, en plus du snoRNA, de 4 protéines nucléolaires essentielles : la fibrillarine (Nop1p chez *S. cerevisiae*) qui porte l'activité méthyltransférase (Niewmierzycka et Clarke, 1999), Nop56p, Nop58p (ou Nop5p) et p15,5kD (ou Snu13p chez *S. cerevisiae*) (pour revues, Kiss, 2002 ; Reichow *et al.*, 2007). Cette dernière protéine a d'abord été identifiée comme composant de la tri-snRNP (U4/U6.U5) (Gottschalk *et al.*, 1999 ; Nottrott *et al.*, 1999; Stevens et Abelson, 1999). Notre équipe a montré, en collaboration avec l'équipe de R. Lührmann, que Snu13p et son homologue de mammifères p15,5 kD, aussi nommé NHPX (Leung et Lamond, 2002) se lient aux snoRNA à boîtes C/D (Watkins *et al.*, 2000). Les protéines 15,5 kD et Snu13 appartiennent à une famille de protéines particulières, la famille L30/L7Ae dont les membres présentent un domaine conservé de fixation à l'ARN, plus précisément avec un motif K-turn (Vidovic *et al.*, 2000 ; Klein *et al.*, 2001 ; Charron *et al.*, 2004). Notre équipe a démontré que les couples de boîtes C/D des snoRNA forment une structure en K-turn qui recrute la protéine Snu13/15,5kD (Watkins *et al.*, 2000 ; Charron *et al.*, 2004) et que sa fixation favorise la formation du K-turn (Marmier-Gourrier *et al.*, 2003). Il a également été montré que la fixation de la protéine 15,5kD pourrait entraîner des réarrangements structuraux qui génèrent des sites de fixation pour Nop58p, Nop56p et la fibrillarine (Watkins *et al.*, 2002). La protéine Nop58 est plus spécifiquement associée au couple de boîtes C/D, alors que la protéine Nop56 est plus spécifiquement liée au couple de boîtes C'/D'. Ces protéines sont responsables de la formation d'un complexe asymétrique (Fig. II.21) (Cahill *et al.*, 2002). Il est surprenant de remarquer que, bien que le couple C'/D' interagisse avec Nop56 et la fibrillarine, celui-ci ne peut pas interagir avec la protéine 15,5kD du fait de l'absence d'hélice I et donc de l'absence d'un motif en K-turn canonique (Szewczak *et al.*, 2002).

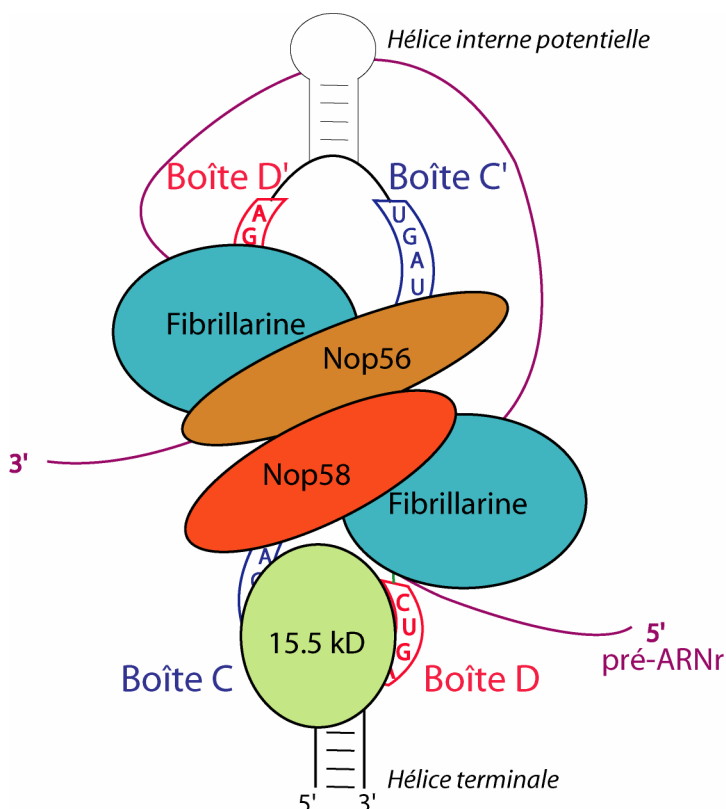


Figure II.21 : Représentation schématique de l'organisation des snoRNP à boîtes C/D (adapté de Reichow *et al.*, 2007).

2.6.2 snoRNA à boîtes H/ACA

Définition et Mode d'Action

Les snoRNA à boîtes H/ACA se définissent par la présence de boîtes H (5'-ANANNA-3') et ACA (5'-ANA-3') qui jouent un rôle majeur dans leur maturation, leur accumulation, leur localisation et leur fonction (Balakin *et al.*, 1996 ; Bortolin *et al.*, 1999 ; pour revues, Omer *et al.*, 2003 ; Kiss *et al.*, 2004 ; Matera *et al.*, 2007; Reichow *et al.*, 2007). Ils adoptent presque tous une structure secondaire en deux tige-boucles. La boîte H est localisée dans la région dite charnière entre les deux tige-boucles et la boîte ACA à l'extrémité 3' (Fig. II.22). Cependant, certains snoRNA à boîtes H/ACA découverts chez les trypanosomes et chez *Euglena* ne comportent qu'une seule tige-boucle présentant une boîte ACA à l'extrémité 3' dont la séquence est en fait AGA, mais aucune boîte H en 5' (Fig. II.23) (Liang *et al.*, 2001 ; Russell *et al.*, 2004; Liang *et al.*, 2007; pour revue, Uliel *et al.*, 2004). Un snoRNA particulier a été décrit chez *Euglena* : en plus de 4 structures à tige-boucles uniques sans particularité ce snoRNA comporte deux tige-boucles avec les boîtes H et ACA (Eg-h1). Une troisième tige-boucle caractéristique supplémentaire est localisée dans la région charnière en aval de la boîte H. Aucun ARN cible n'a encore été identifiée pour cet ARN particulier (Fig. II.23) (Russell *et al.*, 2004). A l'instar du snR30 de *S. cerevisiae* ce snoRNA pourrait être impliqué les clivages du pré-ARNr (Atzorn *et al.*, 2004). Des snoRNA à boîtes H/ACA présentant des tige-boucles de plusieurs centaines de nucléotides voire des tige-boucles supplémentaires ont été identifiés chez *S. cerevisiae* ou *G. lamblia* (Fig. II.23) (Torchet *et al.*, 2005 ; Yang *et al.*, 2005).

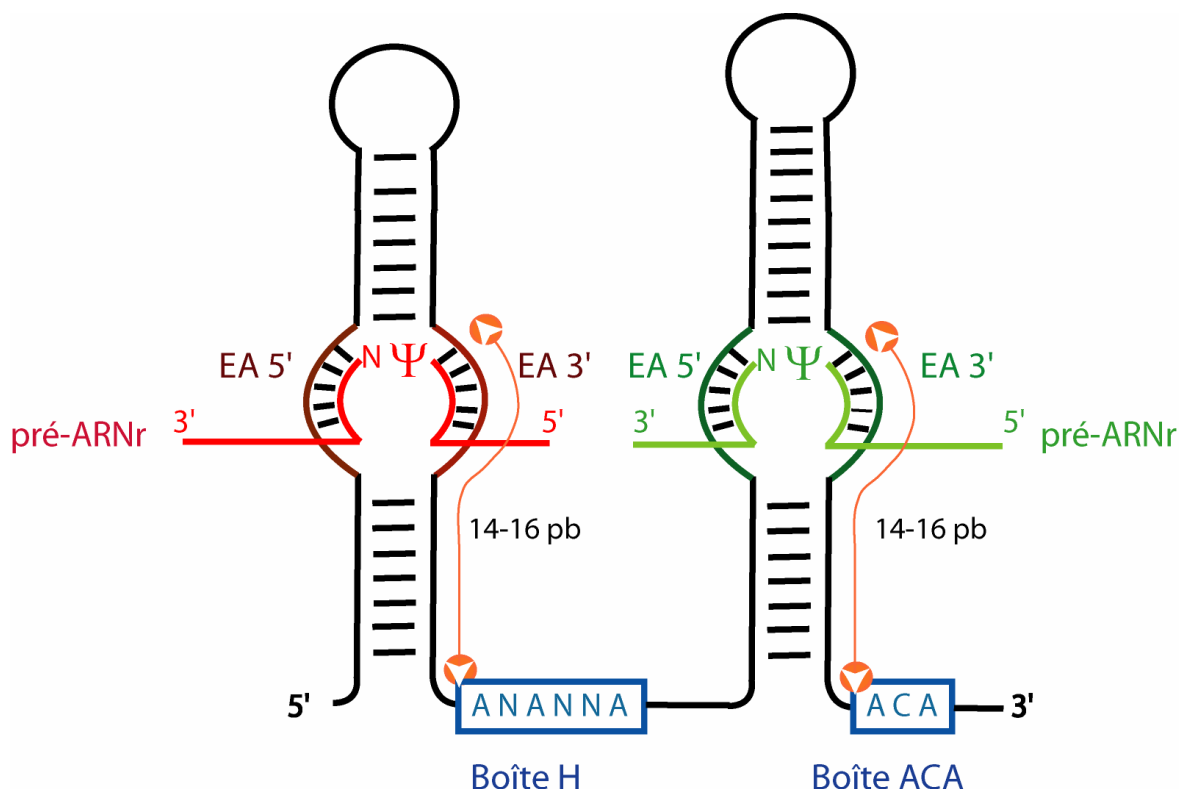


Figure II.22 : Représentation schématique de la structure secondaire des snoRNA à boîtes H/ACA. Les boîtes H et ACA sont encadrées. Les éléments antisens 5' et 3' notés EA 5' et EA 3' présents dans les boucles internes de chacune des deux tige-boucles présentent des complémentarités de séquence avec des régions du pré-ARNr. La distance entre le site de pseudouridylation et les boîtes H ou ACA représentée ici est invariablement comprise entre 14 et 16 nt.

Dans le cadre d'une architecture classique comportant deux structures tige-boucles présentant une poche de pseudouridylation, chacune des 2 tige-boucles comporte une boucle interne qui présente deux séquences « guides », séquences qui seront appelées élément antisens 5' (EA5') et élément antisens 3' (EA3') dans la suite de ce manuscrit. Les deux éléments antisens sont complémentaires à deux séquences du pré-ARNr situées respectivement 2 nt en aval et directement en amont du résidu U à convertir en résidu pseudouridine (Ψ) (Fig. II.22). Ces deux séquences déterminent donc avec précision la position de la modification. Par conséquent, un snoRNA à boîtes H/ACA guide potentiellement deux modifications différentes suivant la possibilité d'apparier les séquences de ses deux boucles internes avec une ou plusieurs cible(s). Comme pour l'action des snoRNA à boîtes C/D il existe une contrainte de distance entre le site de modification et une des deux boîtes. En effet, le site de pseudouridylation est toujours localisé 14 à 16 nt en amont de l'une des deux boîtes, H ou ACA (Balakin *et al.*, 1996 ; Ganot *et al.*, 1997a ; Ganot *et al.*, 1997b ; Ni *et al.*, 1997).

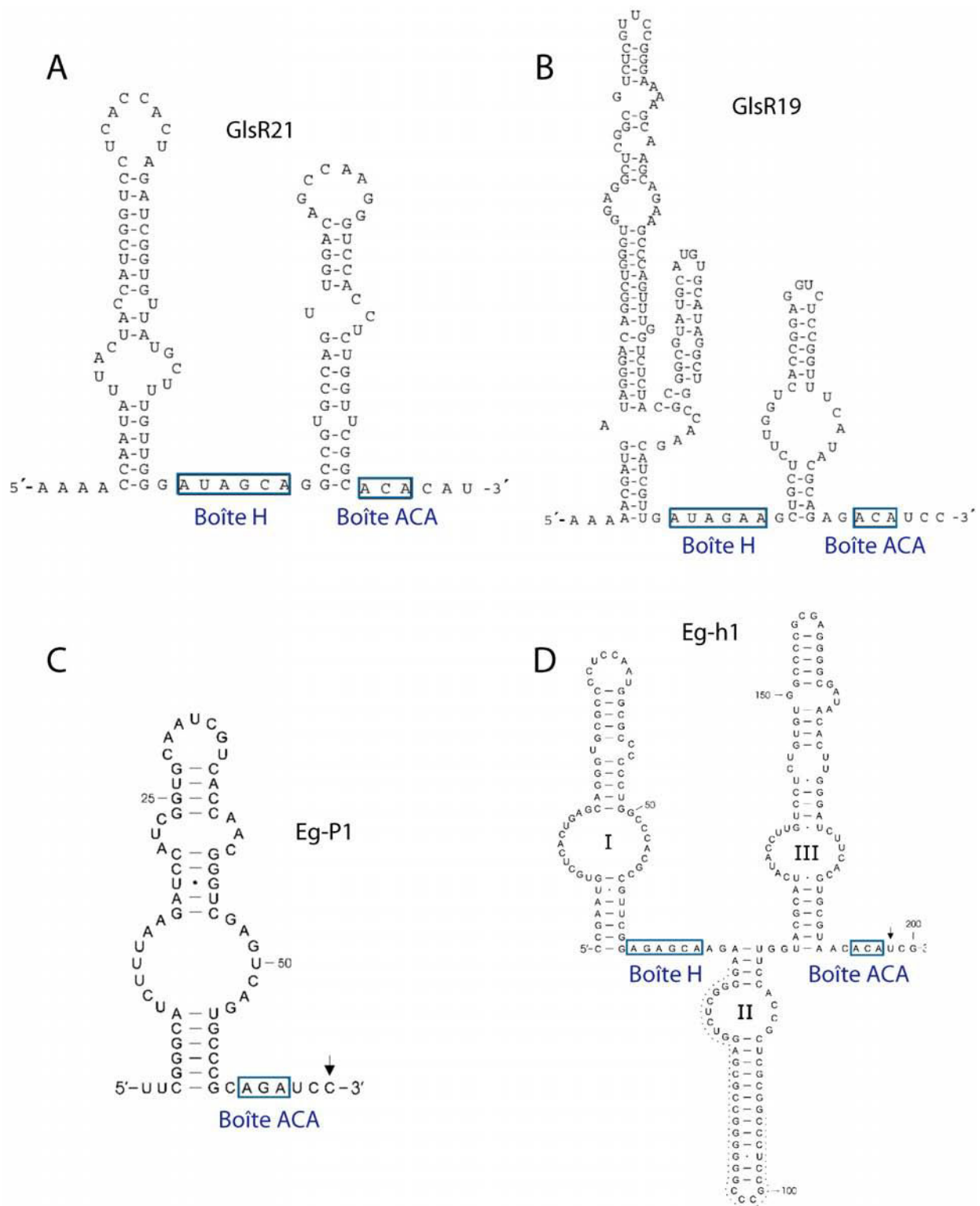


Figure II.23 : Exemples de snoRNA à boîtes H/ACA identifiés chez les eucaryotes (adapté de Russell *et al.*, 2004; Yang *et al.*, 2005).

A. Structure secondaire à deux tige-boucles dite canonique de l'ARN GlsR21 identifié chez *Giardia lamblia*. **B.** Structure secondaire à deux tige-boucle comportant une hélice supplémentaire au sein de la première tige-boucle, prédite pour l'ARN GlsR19 identifié chez *Giardia lamblia*. **C.** Structure secondaire à une seule tige-boucle comportant une seule boîte ACA, ici AGA, prédite pour l'ARN Eg-P1 identifié chez *Euglena gracilis*. **D.** Structure secondaire à trois tige-boucles comportant deux boîtes (H et ACA) observée pour l'ARN Eg-h1 identifié chez *Euglena gracilis*.

snoRNP et rôle

Quatre protéines sont associées aux snoRNA à boîtes H/ACA: la Dyskérine ou Cbf5p (Centromere Binding Factor-5) également appelée NAP57 chez les vertébrés, Nhp2p, Nop10p et Gar1p (pour revue, Reichow *et al.*, 2007). Toutes sont requises pour l'activité des snoRNP et toutes sauf Gar1p sont requises pour leur stabilité (Bousquet-Antonelli *et al.*, 1997). La structure, l'assemblage et la fonction de ces protéines ainsi que les caractéristiques des snoRNA à boîtes H/ACA seront détaillées dans la partie III.

Télomérase et Dyskératose Congénitale (DC)

La télomérase est un grand complexe (environ 1000 kDa) dont le complexe minimal actif comprend l'ARN de la télomérase (hTR pour human Telomerase RNA) et une ADN polymérase ARN-dépendante (hTERT pour human TELomerase Reverse Transcriptase) (Beattie *et al.*, 1998 ; pour l'historique, de Lange, 2006). Cette dernière enzyme réplique les séquences des télomères en utilisant comme matrice l'ARN TR. Cet ARN TR est composé d'un domaine 5' constituant la matrice pour l'action de l'enzyme et d'un domaine 3' ayant les mêmes caractéristiques qu'un snoRNA H/ACA (Fig. II.24) (Mitchell *et al.*, 1999a). Le domaine 3' n'est pas requis pour l'activité mais est indispensable pour la maturation, la stabilité et la localisation nucléaire (Mitchell *et al.*, 1999a ; Mitchell et Collins, 2000 ; Lukowiak *et al.*, 2001).

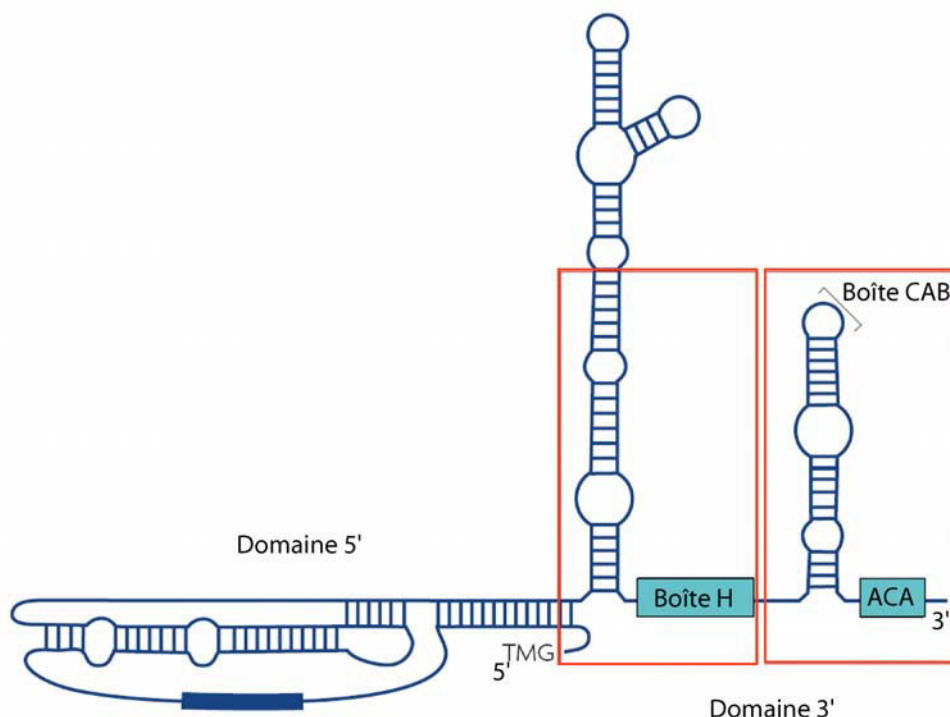


Figure II.24 : Structure secondaire schématisée de l'ARN de la télomérase de vertébrés (d'après Matera *et al.*, 2007).

Le domaine 3' comporte deux tige-boucles caractéristiques des snoRNA à boîtes H/ACA. Les boîtes H, ACA et CAB sont encadrées. Le domaine 5' constitue un pseudo-nœud requis pour la synthèse des télomères. L'élément antisens qui s'apparie au substrat ADN est représenté par un rectangle plein. TMG : TriMéthylphosphateGuanosine.

Les télomères protègent l'extrémité des chromosomes de la dégradation ou de la fusion entre chromosomes qui entraînerait de multiples recombinaisons. Ils comportent une répétition d'ADN double brin en tandem de plusieurs kilobases (10-30 kb), riche en GT (TTAGGG chez l'homme) et des facteurs protéiques (de Lange, 2005) qui permettent à la cellule de les distinguer des sites de cassure de l'ADN. La réplication de l'extrémité des chromosomes pose un problème à la cellule puisque, du fait de l'utilisation des fragments d'Okasaki sur le brin retardé, l'ADN polymérase ne peut pas complètement répliquer l'extrémité des chromosomes. De ce fait, près de 200 pb sont perdues à l'extrémité des chromosomes à chaque division cellulaire. Au-delà d'un certain nombre de divisions, les cellules somatiques de mammifères entrent en sénescence répllicative. Seules certaines cellules à réplication rapide telles que les cellules épithéliales échappent à ce phénomène en maintenant la longueur de leurs télomères grâce à un mécanisme impliquant la télomérase. Il est intéressant de noter que son activité est absente dans beaucoup des cellules somatiques alors qu'elle est présente dans plus de 90% des cancers.

Un syndrome appelé dyskératose congénitale est associé à un défaut dans les protéines associées aux snoRNA à boîtes H/ACA (Mitchell *et al.*, 1999b). Ces mutations concernent principalement la dyskérine (Li et Ye, 2006; Rashid *et al.*, 2006) mais aussi NOP10 (Walne *et al.*, 2007). Ce syndrome létal se caractérise par des altérations des tissus dont les cellules se renouvellent rapidement : cellules cutanées, de la muqueuse et de la moelle osseuse provoquant notamment une pigmentation anormale de la peau, une dystrophie des ongles et une leucoplasie des muqueuses ainsi qu'une prédisposition au cancer et une aplasie médullaire progressive dans 80% des cas. La dernière défaillance concernant la moelle osseuse est à l'origine de la mort dans la troisième décennie. Ce syndrome peut s'expliquer par un défaut de maintenance des télomères dû à la réduction de l'activité télomérase. En effet, une autre forme autosomale dominante de la maladie est liée à des mutations du gène codant l'ARN de la télomérase (hTR), (Vulliamy *et al.*, 2001). Cependant, l'issue fatale de cette maladie pourrait également être liée à des défauts de maturation et/ou de modification des ARNr entraînant des défauts de la traduction. A partir d'une lignée de souris génétiquement modifiées qui présente en première et deuxième générations tous les symptômes cliniques caractéristiques de la dyskératose congénitale humaine, il a été observé que la réduction de la taille des télomères et la diminution de l'activité télomérase n'apparaissent que tardivement, au niveau de la 4ème génération (G4) (Ruggero *et al.*, 2003). Par conséquent, les symptômes observés au niveau des générations G1 et G2 semblent être indépendants de la voie de maintenance des télomères. Or les auteurs ont remarqué, au niveau de la génération G2, une réduction du taux de pseudouridylation des ARNr 18S et 28S, un défaut de maturation des ARNr et une hypersensibilité aux inhibiteurs traductionnels ce qui reflète un défaut de fonctionnement des ribosomes. Ainsi, chez la souris, le premier effet des mutations est la dérégulation du fonctionnement des ribosomes qui est suffisante pour provoquer l'apparition des symptômes. Un défaut de la télomérase pourrait, ultérieurement, modifier ou exacerber la pathologie. Notons qu'une forme plus sévère de DC,

appelé syndrome de Hoyerall-Hreidarsson qui est marqué par un retard de croissance intra-utérin et de nombreuses malformations aboutissant à un retard mental sont également associés à des mutations dans le gène de la dyskérine (Vulliamy *et al.*, 2006). Il serait intéressant de déterminer si les acides aminés concernés dans cette forme sévère sont impliqués dans la catalyse de la réaction de pseudouridylation ou dans l'assemblage des complexes snoRNP et télomérase.

2.6.3 scaRNA

Il a été observé que les snRNA ainsi que des snoRNA ayant des boîtes C/D ou H/ACA qui guident des modifications dans les snRNA transcrits par l'ARN pol II à savoir U1, U2, U4 et U5 co-localisent dans les corps de Cajal (ou CB) suggérant que ces modifications ont lieu uniquement dans ces corpuscules (Darzacq *et al.*, 2002). Ces ARNnc particuliers sont appelés scaRNA (pour « small Cajal Bodies RNA ») du fait de leur localisation. Chez l'humain, 25 scaRNA identifiés seraient responsables de la formation d'au moins 17 résidus Ψ parmi les 24 résidus Ψ détectés (soit 17/24, dans U1 : 1/2, U2 : 9/14, U5 : 3/3, U6 : 2/ nd, U12 : 2/2) et 14 2'-O-méthylations des riboses parmi les 21 2'-O-méthylations détectés (soit 14/21 dans U1 : 1/2, U2 : 6/9, U4 : 3/4, U5 : 3/5, U12 : 1/1) (Xie *et al.*, 2007). Ils présentent une structure comparable aux snoRNA et tombent dans deux catégories selon qu'ils présentent soit uniquement des boîtes C/D soit des boîtes H/ACA. Ces derniers sont majoritaires et ont la particularité d'être parfois composés d'une association soit de deux snoRNA à boîtes H/ACA tel que le U93 humain (Kiss, 2002) soit d'un snoRNA à boîtes H/ACA combiné à un snoRNA à boîtes C/D tel que le U85 humain (Fig. II.25) (Jady et Kiss, 2001 ; pour revue, Henras *et al.*, 2004). Parmi les scaRNA à boîtes C/D certains présentent également des particularités telles que l'association de deux snoRNA à boîtes C/D parfois séparés de plus de 80 nucléotides (Tycowski *et al.*, 2004).

Les boucles apicales (ou terminales) des scaRNA à boîtes H/ACA présentent toujours une ou deux séquences conservées de 4 nt appelée « CAB box » ou boîte CAB dont le consensus est UGAG dans lequel les 3^{ème} et 4^{ème} nucléotides sont les mieux conservés (Richard *et al.*, 2003). Les boîtes CAB comprises dans une boucle terminale sont donc responsables de la rétention dans les corps de Cajal par le biais d'une interaction avec une ou plusieurs protéines encore inconnues mais qui requièrent la présence des protéines associées aux boîtes H/ACA (Fig. II.19). En effet, l'altération soit d'une des deux séquences de boîte CAB soit de la séquence des boîtes H/ACA entraîne une localisation dans le nucléole (Richard *et al.*, 2003). Notons que l'ARN de la télomérase comporte une boîte CAB au niveau de la boucle apicale de la seconde tige-boucle de son domaine 3' guide H/ACA, suggérant également un transit par les corps de Cajal (Fig. II.19) (Jady *et al.*, 2004). Cependant, l'existence de scaRNA à boîtes C/D qui ne présentent ni boîte H/ACA ni boîte CAB, tels que les scaRNA U90 et U91 humains, suggère la présence d'autres mécanismes de rétention dans les CB (Richard *et al.*, 2003).

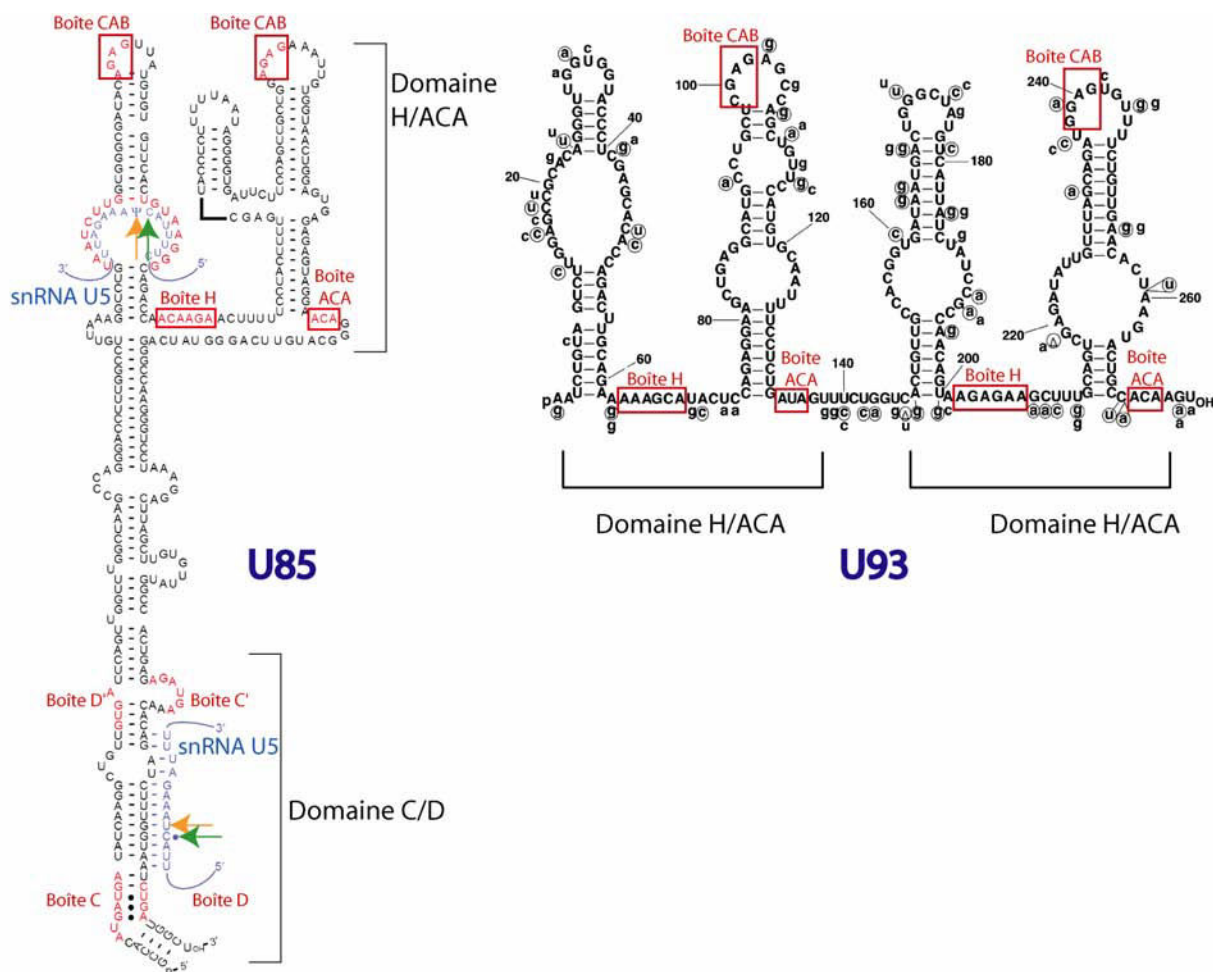


Figure II.25 : Exemples de structures secondaires de scaRNA humains (d'après Jady et Kiss, 2001 ; Kiss, 2002). Le scaRNA U85 illustre la combinaison d'un domaine C/D avec un domaine H/ACA alors que le scaRNA U93 illustre la combinaison de deux domaines H/ACA. Les boîtes C, D, C', D', H, ACA et CAB sont encadrées en rouge. Les différences de séquence entre le scaRNA U93 et ses homologues de *Mus musculus* (minuscules) et *Bos taurus* (minuscules encadrées) sont représentées. Le scaRNA U85 guide la méthylation et la pseudouridylation de deux positions juxtaposées du snRNA U5 qui sont indiquées par des flèches vertes et orange respectivement.

2.7 Recours aux guides ARN

Aucun ARN guide de modification n'a été décrit chez les Bactéries alors que, comme présenté par la suite, des guides ARN sont présents chez les Archaea. Il semblerait donc que ce système soit apparu avant la divergence Archaea/Eucaryotes mais après la divergence entre la branche des Bactéries et celle des Archaea/Eucaryotes à moins que de telles fonctionnalités aient été perdues chez les Bactéries (pour revue, Huttenhofer et Schattner, 2006).

Il est à noter que ce système guidé par des ARNnc concerne plus particulièrement les 2'-O-méthylations et les clivages, donc concerne plus la chaîne ribose-phosphate que les bases dont les modifications sont plus souvent catalysées par des systèmes enzymatiques de nature protéique. Les pseudouridylation constituent donc un cas particulier puisque d'une part le résidu U à convertir en Ψ n'est pas apparié et donc pas reconnu par le guide ARN et d'autre part

un appariement singulier entre l'ARN guide et l'ARN cible positionne le résidu U dans une conformation qui le rend accessible à l'enzyme (nos travaux ; Charpentier *et al.*, 2005 ; Jin *et al.*, 2007 ; Wu et Feigon, 2007 ; développé dans le chapitre suivant).

Il est très important de remarquer que les enzymes chimériques (ou RNP) coexistent avec des enzymes protéiques. Par exemple, trois pseudouridylation présentes dans le U2 snRNA de levure sont catalysées soit par des enzymes protéiques pour deux d'entre elles ($\Psi 35$ et $\Psi 44$) soit par un snoRNP ($\Psi 42$) (Ma *et al.*, 2005). Dans un second exemple, la 2'-O-méthylation de la position G2922 de l'ARNr 28S de levure est réalisée par une enzyme protéique, Sbp1p, alors que la plupart des méthylations de ce type se font via des snoRNP à boîtes C/D (Lapeyre et Purushothaman, 2004). Ainsi, une théorie propose que trois routes aient permis une évolution à partir des ribozymes d'un monde ARN vers les systèmes actuels (Fig. II.26) (Huttenhofer et Schattner, 2006).

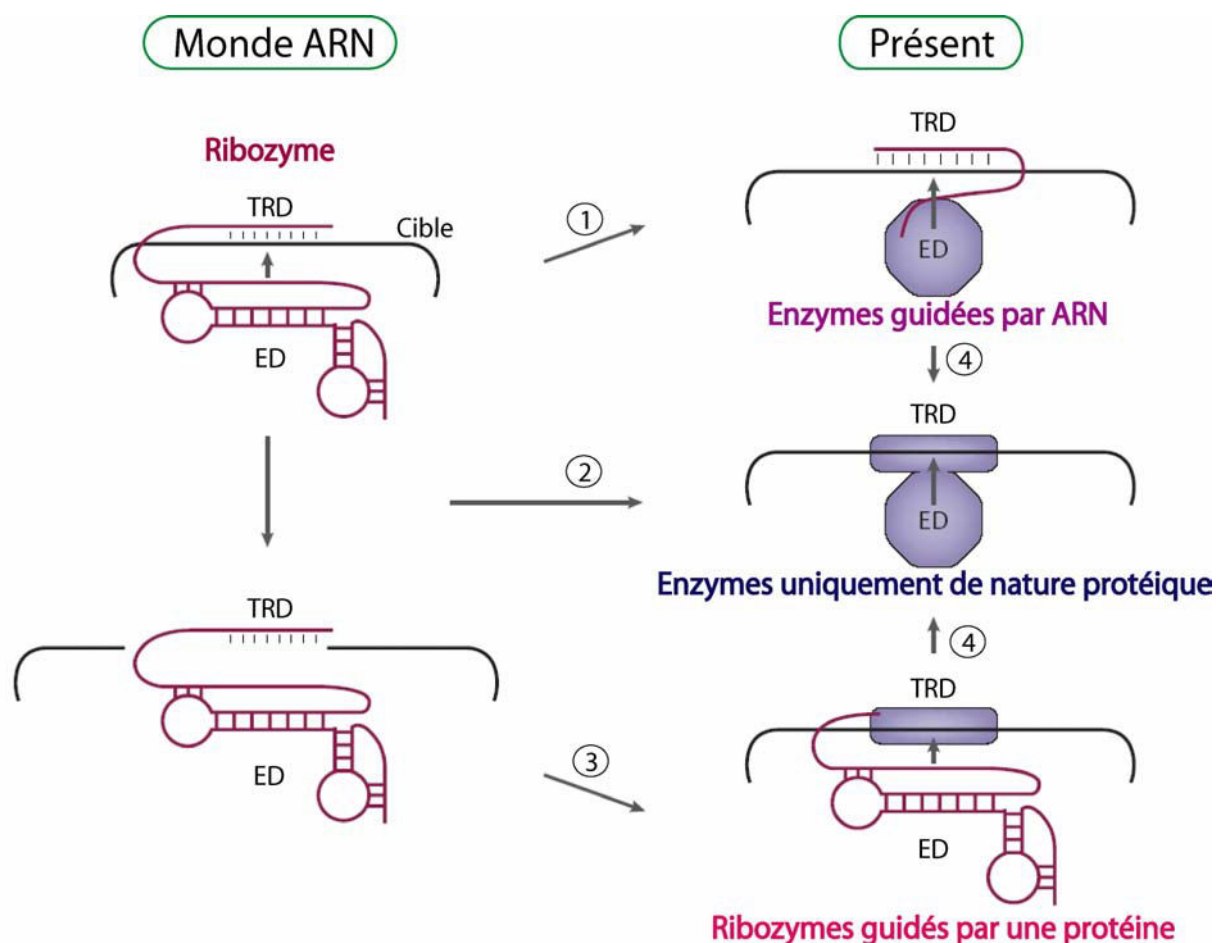


Figure II.26 : Modèle d'évolution des systèmes enzymatiques protéiques ou chimériques (ARN-protéine) (d'après Huttenhofer et Schattner, 2006).

Trois chemins différents sont envisagés à partir de ribozymes issus d'un monde ARN impliqués à la fois dans la reconnaissance par le domaine TRD (Target Recognition Domain) et dans le clivage d'une molécule ARN par le domaine ED (Enzyme Domain). Le remplacement du domaine ED par un domaine protéique (1) conduit à un système d'enzymes de nature protéique guidées par des ARN (ex. : miRNA). Sont également représentés le remplacement possible du ribozyme par une enzyme uniquement de nature protéique (2) (ex. : RNase A) ainsi que le remplacement de TRD par un domaine protéique conduisant à un ribozyme guidé par des acides aminés (3) (ex. : RNase P). La substitution des domaines TRD ou ED par des domaines protéiques (4) pourrait également conduire à l'apparition d'enzymes uniquement protéiques.

La reconnaissance de la cible (Target Recognition Domain ou TRD) peut être effectuée soit par un domaine d'acides aminés soit par une séquence d'ARN alors que le domaine enzymatique (Enzymatic Domain ou ED) est le plus souvent assuré par des acides aminés et rarement par des nucléotides. Il en résulte trois combinaisons différentes qui co-existent au sein des cellules d'Eucaryotes et d'Archaea. Le recours à un TRD de nature ARN a permis au cours de l'évolution, outre le coût moindre de la synthèse de transcrits par rapport à la production de protéines, la multiplication des cibles de régulations. Il est évident que les nombreuses modifications (principalement pseudouridines et 2'-*O*-méthylation du ribose) présentes dans les ARNr auraient requis des centaines voire des milliers de protéines différentes. Or une telle évolution d'un TRD de nature protéique requiert la duplication du gène de la première enzyme et de nombreuses mutations de ses acides aminés, alors que l'évolution d'un TRD de nature ARN ne requiert que la duplication d'une petite région d'ADN et la mutation d'un nombre très limité de nucléotides. Le recours aux guides ARN a sans doute été une formidable opportunité de développement d'un système de régulation très élaboré dont une grande partie reste inexploré.

3. Archaea

Les Archaea présentent un génome d'une taille comparable à celles des génomes bactériens, codant à plus de 80% des protéines. Aucune machinerie d'épissage des pré-ARNm ressemblant à celle des eucaryotes n'y a été décrite. Cependant, les archaea partagent de nombreuses caractéristiques avec les eucaryotes (présentées dans la première partie de l'introduction) dont des ressemblances de chromatine, ou de machineries de réplication, transcription et traduction. Etant donné qu'un impressionnant répertoire d'ARNnc régule ces processus eucaryotes, il était envisageable que les archaea contiennent de tels ARNnc.

3.1. *sRNA à boîtes C/D*

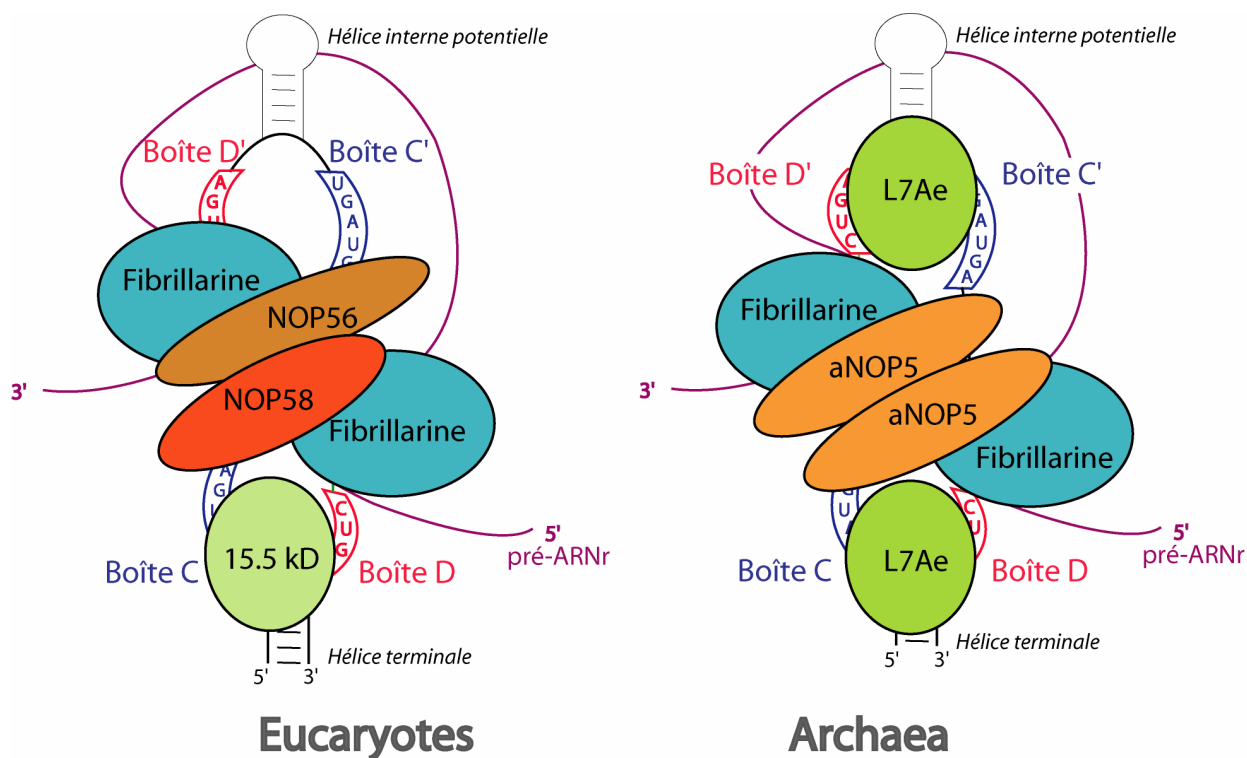
Un système d'ARN guides de méthylation équivalent au système eucaryote a été mis en évidence chez les archaea et implique des ARN guides comparables aux snoRNA à boîtes C/D appelés sRNA (pour « small RNA ») à boîtes C/D (Gaspin *et al.*, 2000; Omer *et al.*, 2000). La présence de ces ARN guides de 2'-*O*-méthylations va de paire avec la présence d'un grand nombre de 2'-*O*-méthylations dans les ARNr d'archaea. Par exemple, l'examen des ARNr de *S. solfataricus* a révélé 67 résidus 2' O-méthylés contre 4 résidus détectés chez *E. coli* (Noon *et al.*, 1998). Les sRNA à boîtes C/D d'archaea présentent cependant de nombreuses caractéristiques permettant de les distinguer des snoRNA eucaryotes.

Ils sont en général plus petits (50 à 60 nt) (Gaspin *et al.*, 2000; Omer *et al.*, 2000) que leurs homologues eucaryotes (75 à 100 nt) (Kiss-Laszlo *et al.*, 1998 ; Lowe et Eddy, 1999). Par ailleurs, le consensus observé pour la boîte C est un peu plus long et plus strict que chez les eucaryotes : 5'-RAUGAUGA-3' au lieu de 5'-RUGAUGA-3'. L'hélice terminale rapprochant

les extrémités 5' et 3' est nettement moins fréquente chez les archaea et, lorsque elle est présente, elle comporte un petit nombre d'appariements. Des phénomènes de circularisation par ligation des extrémités 5' et 3' ont été observés chez les sRNA à boîtes C/D extraits de *P. furiosus* (Starostina *et al.*, 2004). Un tel phénomène pourrait contribuer à rapprocher les extrémités 5' et 3', et peut-être à stabiliser le motif en K-turn formé par le couple de boîtes C/D. La distance entre les couples de boîtes C/D et C'/D', en d'autres termes la taille de la séquence qui est potentiellement complémentaire à une séquence cible varie peu, entre 11 et 14 nt (Tran *et al.*, 2005). La règle de la méthylation du nucléotide apparié au 5^{ième} nucléotide en amont de la boîte D ou D' est respectée. La majorité des sRNA C/D comportent deux régions potentiellement guides de 2'-O-méthylation, contre seulement 20% pour les snoRNA humains ou de levure (Kiss-Laszlo *et al.*, 1998 ; Lowe et Eddy, 1999). Par ailleurs, les 2 cibles d'un même sRNA C/D sont en général très proches spatialement (soit proches dans la séquence, soit proches dans la structure secondaire ou tertiaire) contrairement à ce qui est observé pour les snoRNA eucaryotes. Ceci suggère que les deux duplex sont formés simultanément au sein de l'ARNr. Cette particularité suggère une participation en tant que chaperons des sRNA à boîtes C/D au repliement des ARNr (Gaspin *et al.*, 2000; Omer *et al.*, 2000 ; pour revues, Terns et Terns, 2002 ; Bachellerie *et al.*, 2002). Une modélisation bioinformatique de l'effet des sRNA C/D sur le repliement des ARNr propose que ceux-ci facilitent l'appariement de régions éloignées, en particulier à température élevée (Schoemaker et Gultyaev, 2006).

Les protéines associées aux sRNA C/D d'archaea sont homologues à celles associées aux snoRNA C/D eucaryotes. Néanmoins leur nombre est réduit à trois (Fig. II.27) (pour revue, Reichow *et al.*, 2007). En effet, une de ces protéines présente des similarités avec à la fois les protéines eucaryotes Nop56 et Nop58, et, par conséquent, est appelé aNOP56/58 ou aNOP5. Cette protéine aurait donc un ancêtre commun avec NOP56 et NOP58. La présence de deux protéines distinctes chez les eucaryotes pourrait être liée à une duplication de gène. La fibrillarine (encore appelée aNOP1 ou aFIB1) qui porte l'activité catalytique est conservée. La protéine Snu13 ou 15.5kD est substituée par la protéine ribosomique L7Ae (Watkins *et al.*, 2000). Cette protéine est le prototype de la famille des protéines L7Ae/L30 précédemment décrite. L7Ae est connue pour se lier au motif K-turn Kt-15 de l'ARNr 23S au sein de la sous-unité 50S de *H. marismortui* (Klein *et al.*, 2001). L'association de la protéine Snu13 de *S. cerevisiae* avec un motif en K-turn prouve son lien de parenté avec L7Ae (Watkins *et al.*, 2000). Des travaux de notre équipe sur la protéine L7Ae de *P. abyssi* ont confirmé ses capacités d'interaction avec les sRNA à boîtes C/D, les motifs C/D contenus dans des produits de maturation des pré-ARNr d'archaea (Tang *et al.*, 2002b) et le motif appelé K-loop formé par le couple de boîtes C'/D' (Charron *et al.*, 2004). Ce motif en K-loop ne fixe pas Snu13 mais fixe L7Ae avec une affinité légèrement moindre, à condition toutefois que la boucle formée à la place de l'hélice I compte au moins 5 nucléotides (Charron *et al.*, 2004). La protéine L7Ae a donc un spectre de fixation plus large que son homologue eucaryote Snu13. La résolution de la structure de L7Ae associée avec des motifs K-turn montre que L7Ae contacte les G des deux

paires G-A et les trois nucléotides de la boucle interne (Vidovic *et al.*, 2000 ; Klein *et al.*, 2001 ; Charron *et al.*, 2004; Hamma et Ferre-D'Amare, 2004 ; Moore *et al.*, 2004).



Fibrillarine	Fibrillarine, aNOP1, ou aFIB
Snu13, 15.5kD ou NHPX	L7Ae
NOP56	aNOP5 ou aNOP56/58
NOP58	

Figure II.27 : Comparaison de l'association des snoRNA et sRNA à boîtes C/D d'eucaryotes et d'archaea. La formation d'un complexe asymétrique chez les eucaryotes et d'un complexe symétrique chez les archaea est représentée. Les relations d'homologie entre les protéines d'archaea et d'eucaryotes sont indiquées dans le tableau.

La reconstitution de particules sRNP C/D actives d'archaea a permis la détermination de l'ordre d'assemblage des trois protéines de la particule (Omer *et al.*, 2002 ; Bortolin *et al.*, 2003 ; Rashid *et al.*, 2003 ; Tran *et al.*, 2003). Il semble que l'interaction de L7Ae avec le motif en K-turn restructure l'ARN en créant des sites de fixation pour aNOP56/58 (ou aNOP5) (Schultz *et al.*, 2006). La protéine aNOP56/58 pourrait contacter la chaîne phosphate de la boîte C. Lorsque cette dernière est liée au guide, elle recrute la fibrillarine (Omer *et al.*, 2002 ; Rashid *et al.*, 2003 ; Tran *et al.*, 2003). L'assemblage des protéines est symétrique, en contraste avec le système

équivalent eucaryote, chaque couple de boîtes fixant indépendamment à la fois L7Ae, aNOP56/58 et aNOP1 (Tran *et al.*, 2003 ; Hardin et Batey, 2006).

L'étude structurale d'une particule sRNP C/D d'archaea a été réalisée (Wang *et al.*, 2000 ; Aittaleb *et al.*, 2003 ; Charron *et al.*, 2004 ; Deng *et al.*, 2004 ; Moore *et al.*, 2004 ; Suryadi *et al.*, 2005). Une étude de la protéine aNOP56/58 en interaction avec la fibrillarine et la SAM a révélé une association en hétérodimère fibrillarine-aNOP56/58-aNOP56/58-fibrillarine (Aittaleb *et al.*, 2003) qui a pour conséquence le positionnement d'une seconde molécule de fibrillarine sur le couple C'/D' (Szewczak *et al.*, 2002 ; Rashid *et al.*, 2003). Notons qu'il ne semble pas y avoir de contact direct entre l'hétérodimère et L7Ae.

Une dernière particularité qui serait propre aux sRNA C/D d'archaea est la possibilité de cibler des ARNt (Omer *et al.*, 2000 ; Clouet d'Orval *et al.*, 2001 ; Dennis *et al.*, 2001 ; Renalier *et al.*, 2005). En effet, une étude réalisée chez 9 espèces archaea a révélé 22 sites potentiels de méthylation dans les ARNt guidés par des sRNA à boîtes C/D (Fig. II.28) (Dennis *et al.*, 2001). Chez *P. aerophilum*, par exemple, environ 20 sRNA à boîtes C/D sont responsables des méthylations de 20 positions différentes dans les ARNt. Notons qu'un sRNA peut guider la modification d'une position précise dans une ou plusieurs espèce(s) d'ARNt différentes (jusqu'à 19). Chez les eucaryotes, le transit possible des ARNt par le nucléole avant leur export vers le cytoplasme suggère que les ARNt eucaryotes peuvent également être la cible de snoRNA à boîtes C/D. En effet, récemment, il a été proposé que 6 snoRNA C/D de nématodes soient impliqués dans la modification de résidus dans les ARNt (Zemann *et al.*, 2006). Il semblerait donc que, chez les archaea et eucaryotes, les s/snoRNA C/D soient indispensables à la modification de certains résidus des ARNr, ARNt voire d'autres types d'ARNc.

L'hypothèse d'une corrélation entre le nombre prédit de résidus potentiellement 2'-O-méthylés, d'après les sRNA identifiés, et la température optimale de croissance des archaea a été avancée (Omer *et al.*, 2000 ; Dennis *et al.*, 2001). En effet, il semble que plus la température de croissance est élevée, plus il y a de méthylations des riboses au sein des ARNr (Tableau II.1). Le nombre d'espèces d'archaea analysées est cependant trop limité pour valider cette hypothèse. Néanmoins, cette hypothèse est en accord avec des données anciennes montrant que les méthylations renforcent la stabilité des ARNr et ARNt chez les organismes hyperthermophiles. En effet, la méthylation empêche l'hydrolyse par attaque nucléophile en 2'-O des riboses qui est une réaction favorisée à température élevée en milieu aqueux contenant des ions divalents.

Nous avons participé à la recherche de nouveaux sRNA à boîtes C/D dans les génomes d'Archaea qui sera présentée dans le chapitre des résultats.

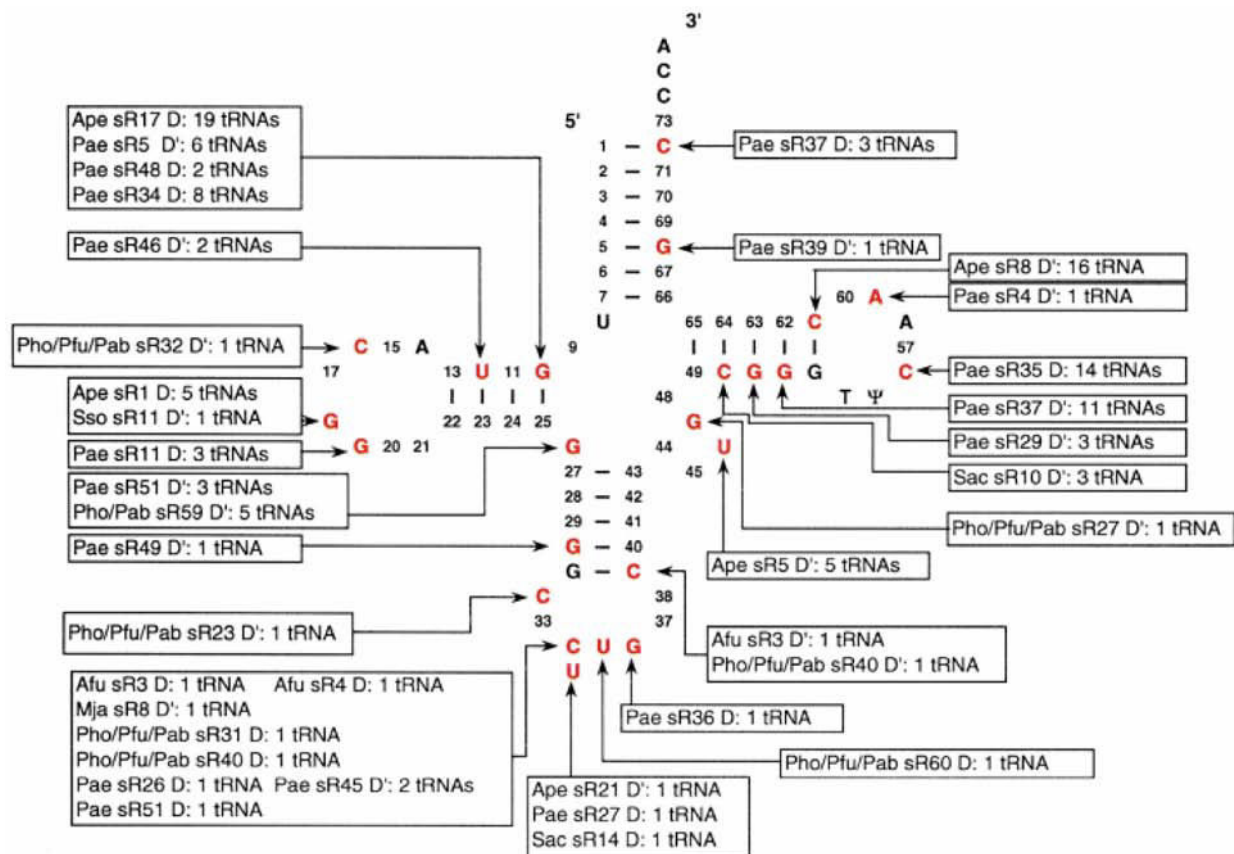


Figure II.28 : Prédiction des sites de 2'-O-méthylations des ARNt de 9 espèces d'archaea (d'après Dennis *et al.*, 2001). Les positions ont été prédites sur la base des complémentarités de séquence observées entre ARNt et sRNA C/D, elles sont indiquées en rouge. Les sRNA proposés pour chaque modification sont indiqués, en précisant la boîte impliquée (D ou D'), ainsi que le nombre d'ARNt potentiellement modifiés par ce guide (Afu : *Archaeoglobus fulgidus* ; Ape : *Aeropyrum pernix* ; Mja : *Methanocaldococcus jannaschii* ; Pab/Pfu/Pho : *Pyrococcus abyssi/furiosus/horikoshii* ; Pae : *Pyrobaculum aerophilum* ; Sac : *Sulfolobus acidocaldarius* ; Sso : *Sulfolobus solfataricus*). Les nucléotides conservés au sein des ARNt sont indiqués par des lettres, les nucléotides variables uniquement par des numéros.

Genre/Espèce	Nombre de gènes correspondant à des sRNA	Température optimale de croissance (°C)
Crenotes		
<i>Sulfolobus acidocaldarius</i>	29	75-80
<i>Sulfolobus solfataricus</i>	13	85-90
<i>Aeropyrum pernix</i>	23	95
<i>Pyrobaculum aerophilum</i>	51	95-100
Euryotes		
<i>Halobacterium sp. NRC1</i>	0	42
<i>Methanobacterium thermoautotrophicum</i>	0	62-67
<i>Archaeoglobus fulgidus</i>	4	83
<i>Methanocaldococcus jannaschii</i>	8	85
<i>Pyrococcus horikoshii</i>	51	98
<i>Pyrococcus abyssi</i>	52	96
<i>Pyrococcus furiosus</i>	56	100

Tableau II.1 : Nombre de gènes correspondant à des sRNA C/D potentiels ou confirmés dans les génomes de différentes espèces d'archaea (d'après Dennis *et al.*, 2001).

La température de croissance de ces espèces est précisée.

3.2. sRNA à boîtes H/ACA

La détection chez les archaea de protéines homologues aux protéines associées aux snoRNA à boîtes H/ACA d'eucaryotes suggérait la présence d'équivalents de snoRNA. Les protéines eucaryotes Nop10p, Cbf5p, et Gar1p présentent en effet des équivalents archaea, respectivement, aNOP10, aCBF5 et aGAR1 (Watanabe et Gray, 2000). En 2002, après le début de mon travail de thèse, une famille d'ARNnc équivalents aux snoRNA à boîtes H/ACA a été découverte chez *A. fulgidus* (Fig. II.29) (Tang *et al.*, 2002a). Depuis cette date, nous avons participé à l'étude des sRNA à boîtes H/ACA dans les génomes d'archaea. De ce fait, notre équipe a fortement contribué, avec les équipes de A. Huttenhofer, A. Ferré d'Amaré, G. Varani, M. Terns, P. Dennis, C. Gaspin, J. Feigon, P. Moore et K. Ye, à l'identification de nouveaux sRNA à boîtes H/ACA (Tang *et al.*, 2002a ; Rozhdestvensky *et al.*, 2003 ; Zago *et al.*, 2005 ; Thebault *et al.*, 2006 ; nos résultats en cours de publication), à l'étude de l'assemblage de particules sRNP H/ACA actives *in vitro* (Baker *et al.*, 2005 ; Charpentier *et al.*, 2005) et à l'étude structurale des sRNA (Wu et Feigon, 2007 ; Jin *et al.*, 2007) et de leurs protéines associées (Charron *et al.*, 2004 ; Hamma et Ferre-D'Amare, 2004; Hamma *et al.*, 2005 ; Li et Ye, 2006; Manival *et al.*, 2006 ; Rashid *et al.*, 2006).

L'étude structurale sera détaillée dans la troisième partie. La recherche de nouveaux guides H/ACA chez les archaea et l'étude de l'assemblage des particules seront présentées dans la partie *Résultats*.

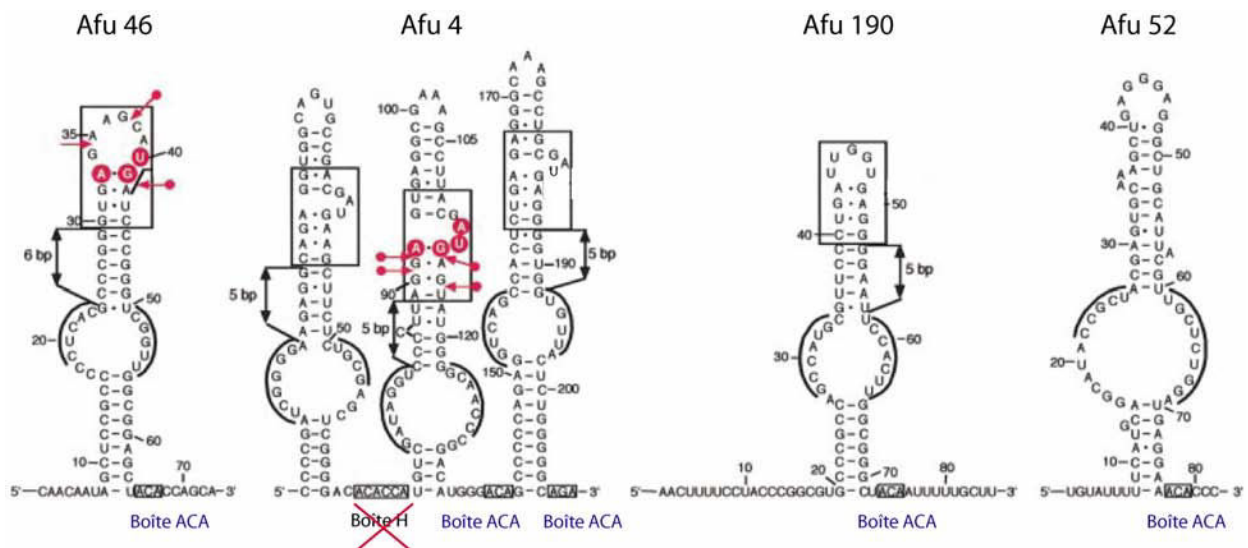


Figure II.29 : Structures secondaires proposées pour les 3 premiers sRNA à boîtes H/ACA identifiés chez un représentant des archaea : *A. fulgidus* (d'après Tang *et al.*, 2002a).

Les boîtes H et ACA ainsi que les motifs en K-turn ou K-loop sont encadrés. Il est à noter que la boîte H présentée ne serait qu'une boîte ACA. En effet, le reste de la séquence est complémentaire avec la partie 5' de la tige-boucle en 3' et participerait donc à l'hélice de cette tige-boucle. Le candidat Afu52 ne présente ni K-turn, ni K-loop, ni cible vérifiée et ne serait pas un véritable sRNA à boîtes H/ACA. Les bases et les liaisons phosphodiester protégées en présence de la protéine L7Ae sont représentées en rouge.

Il se dégage de ces études deux différences majeures entre sRNA et snoRNA à boîtes H/ACA eucaryotes qui seront approfondies dans la partie *Résultats*. D'une part, les sRNA à boîtes H/ACA d'archaea comportent une, deux ou trois tige-boucles contrairement aux homologues eucaryotes qui en comptent le plus souvent deux ou, dans quelques rares cas, une seule. On remarque l'absence de boîte H chez les sRNA archaea. Ainsi, toutes les tige-boucles sont suivies en 3' d'une boîte ACA dont la séquence consensus est ANA (nos résultats en cours de publication). D'autre part, les tige-boucles des sRNA archaea comportent soit un motif en K-turn soit un motif en K-loop en interaction avec la protéine L7Ae. La protéine Nhp2p eucaryote est donc substituée par L7Ae dans le système archaea. Il semble donc que l'assemblage de particules actives diffère entre les systèmes archaea et eucaryote.

3.3. Evolution des sRNA / snoRNA

Il a été proposé que des éléments d'abord présents en *cis* et impliqués dans la modification ou les clivages de l'ARN pourraient avoir évolué pour agir en *trans* au sein de particules RNP (Tran *et al.*, 2004). Une telle hypothèse a déjà été évoquée pour discuter de l'évolution des complexes d'épissage qui dériveraient d'introns « auto-épissés » du groupe II (Sharp, 1991). Les introns du groupe II nous donneraient ainsi une image de ce qu'ont pu être des étapes intermédiaires dans l'évolution de RNP agissant en *cis* vers des RNP agissant en *trans*. Un modèle potentiellement similaire a été découvert à partir d'un ARNt archaea. L'intron du pré-tRNA^{Trp} de pratiquement tous les archaea du royaume euryote présente un sRNA à boîtes C/D impliqué dans la méthylation de deux positions de l'ARNt^{Trp} qui le porte (Fig. II.30) (Clouet d'Orval *et al.*, 2001). Par ailleurs, l'une des deux modifications engendrées dans l'ARNt^{Trp} est indispensable pour qu'il y ait maturation de l'ARNt par reconnaissance d'une structure caractéristique BHB (pour « Bulge-Helix-Bulge » ou « mésappariement-hélice-mésappariement ») par une endonucléase (Arn et Abelson, 1996). Cette observation souligne une fois de plus le lien entre modification et maturation.

Au final, le fait que L7Ae soit commune aux ribosomes, aux sRNA à boîtes C/D et sRNA à boîtes H/ACA suggère que les guides C/D et H/ACA proviennent d'un appareil de traduction primitif (Fig. II.31). Dans cette hypothèse, les K-turn ribosomiques auraient pu d'abord aider au repliement de l'ARNr, puis par l'association à des protéines autres que les protéines ribosomales, certains ARN ou certaines parties d'ARN de l'appareil de traduction primitif auraient fonctionné comme RNP en *cis* responsables de modifications ou de clivages. En effet, des motifs en K-turn ont été observés dans les régions qui séparent les ARNr et flanquent le motif BHB caractéristique des introns d'archaea chez *A. fulgidus*, de *S. solfataricus* et *S. acidocaldarius*. Ces motifs ne sont formés que lors de l'appariement de deux régions responsables de la formation du motif BHB et sont de ce fait responsables de l'excision des ARNr 16S et 23S (Fig. II.32) (Tang *et al.*, 2002a). Il a été démontré expérimentalement, en collaboration avec notre équipe, que cette région est capable de former un complexe avec la protéine L7Ae et ce, malgré la présence d'un résidu G protubérant à la place de l'habituel U du

motif K-turn canonique. Cependant, aucun élément antisens complémentaire aux ARNr n'a pu être détecté en amont de ces motifs équivalents aux couples de boîtes C/D. Néanmoins, il reste envisageable que ceux-ci ciblent d'autres type d'ARN que les ARNr. Mais au final, il est plus probable que ces motifs en K-turn présents dans le pré-ARNr agissent en *cis* comme chaperons pour le repliement et/ou la maturation des ARNr.

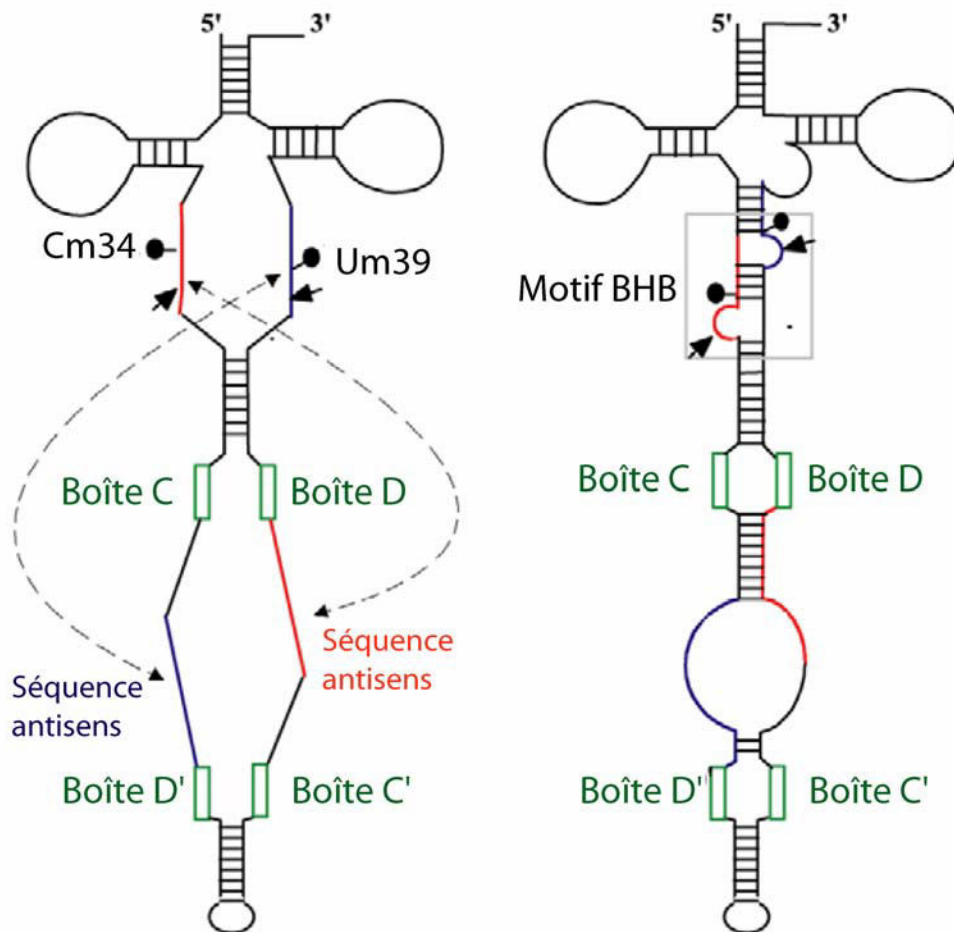


Figure II.30 : Présence, dans l'intron de l'ARNt^{Trp}, d'un sRNA à boîtes C/D qui est impliqué dans la méthylation de deux nucléotides de cet ARNt (Cm34 et Um39) (d'après Clouet d'Orval *et al.*, 2001).

Deux structures secondaires différentes sont représentées selon la formation ou non d'un motif BHB généralement reconnu en par une endonucléase lors de la maturation des ARNnc d'archaea.

Des RNP agissant initialement en *cis*, auraient pu devenir indépendantes en agissant en *trans* et évoluer vers les deux familles de guides C/D ou H/ACA. L'évolution a certainement conduit à « perfectionner » le système en produisant de nouvelles protéines associées telles que NOP56 et NOP58 résultats d'une duplication de aNOP56/58 d'archaea, en multipliant le nombre de cibles ainsi qu'en élargissant les types d'ARN ciblés (ARNm, ARNt, UsnRNA). Il est ainsi envisageable que les boîtes C'/D' des snoRNA ainsi que les snoRNA à boîtes H/ACA aient perdu la capacité de recruter les protéines Snu13 ou 15.5kD. Enfin, la séparation des

compartiments aurait requis l'interaction avec d'autres protéines telles que celles impliquées dans la reconnaissance des boîtes CAB des scaRNA ou celles non présentées ici qui sont impliquées dans leur biogenèse et leur transit.

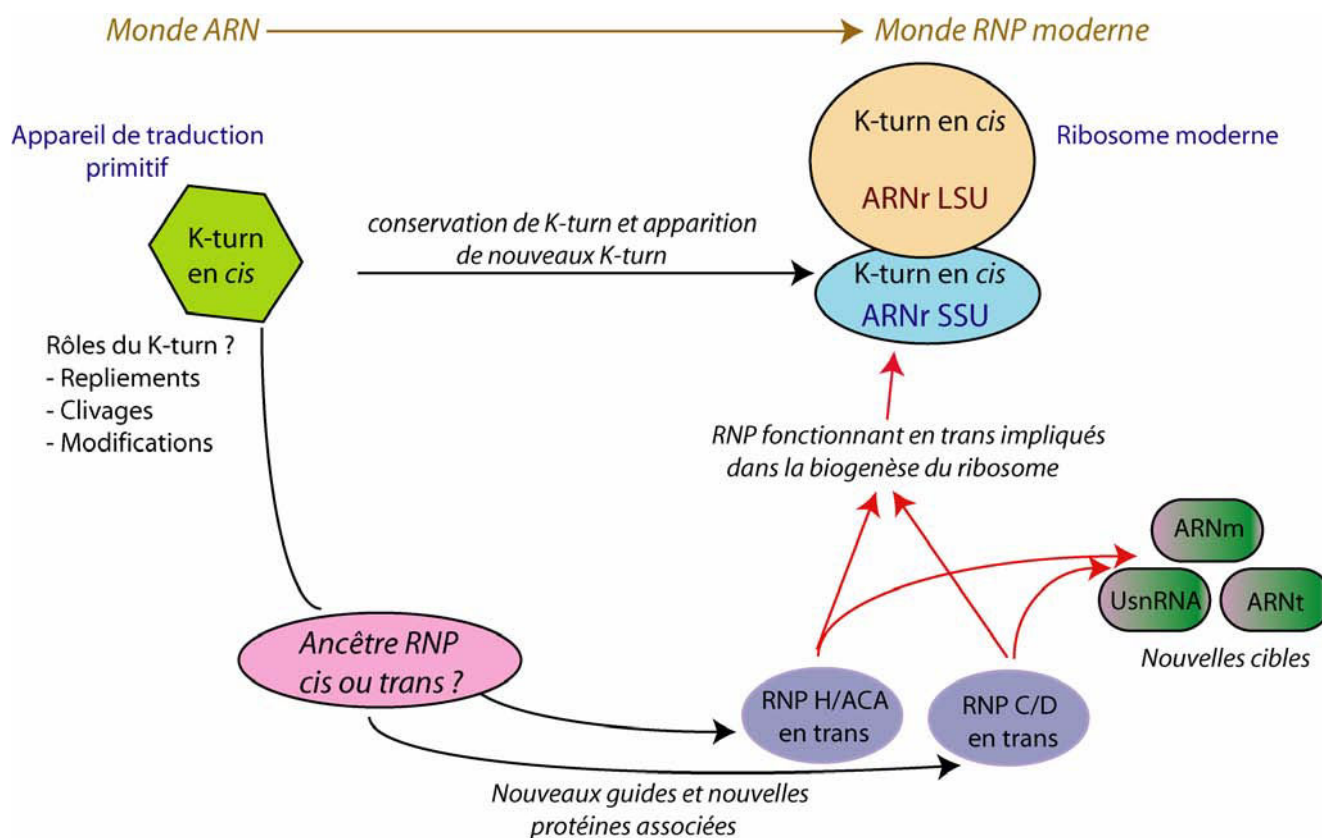


Figure II.31 : Hypothèse de l'évolution des ARN guides à partir d'un appareil de traduction primitif (d'après Tran *et al.*, 2004).

Des motifs en K-turn contenus dans les ARNr ont été conservés tandis que d'autres ont permis l'apparition de guides fonctionnant en *trans*.

La présence de K-turn dans les ARNr bactériens suggère que ce motif faisait déjà partie de l'appareil de traduction primitif de LUCA. La conservation des protéines associées aux sRNA et snoRNA C/D et H/ACA, telles que, par exemple CBF5 appartenant à la famille TruB impliquée dans la modification de la position 55 dans tous les ARNt bactériens, suggère que des sRNA à boîtes C/D ou H/ACA existaient peut-être déjà lors de l'apparition de LUCA (Lafontaine et Tollervey, 1998). Si cela est le cas, seule l'absence d'effet bénéfique explique l'absence totale de ces guides de modification dans les génomes bactériens (Watanabe et Gray, 2000 ; Tran *et al.*, 2004). Cette dernière idée attire l'attention sur le manque d'information concernant le rôle à la fois de ces modifications et des ARNnc qui guident leur formation.

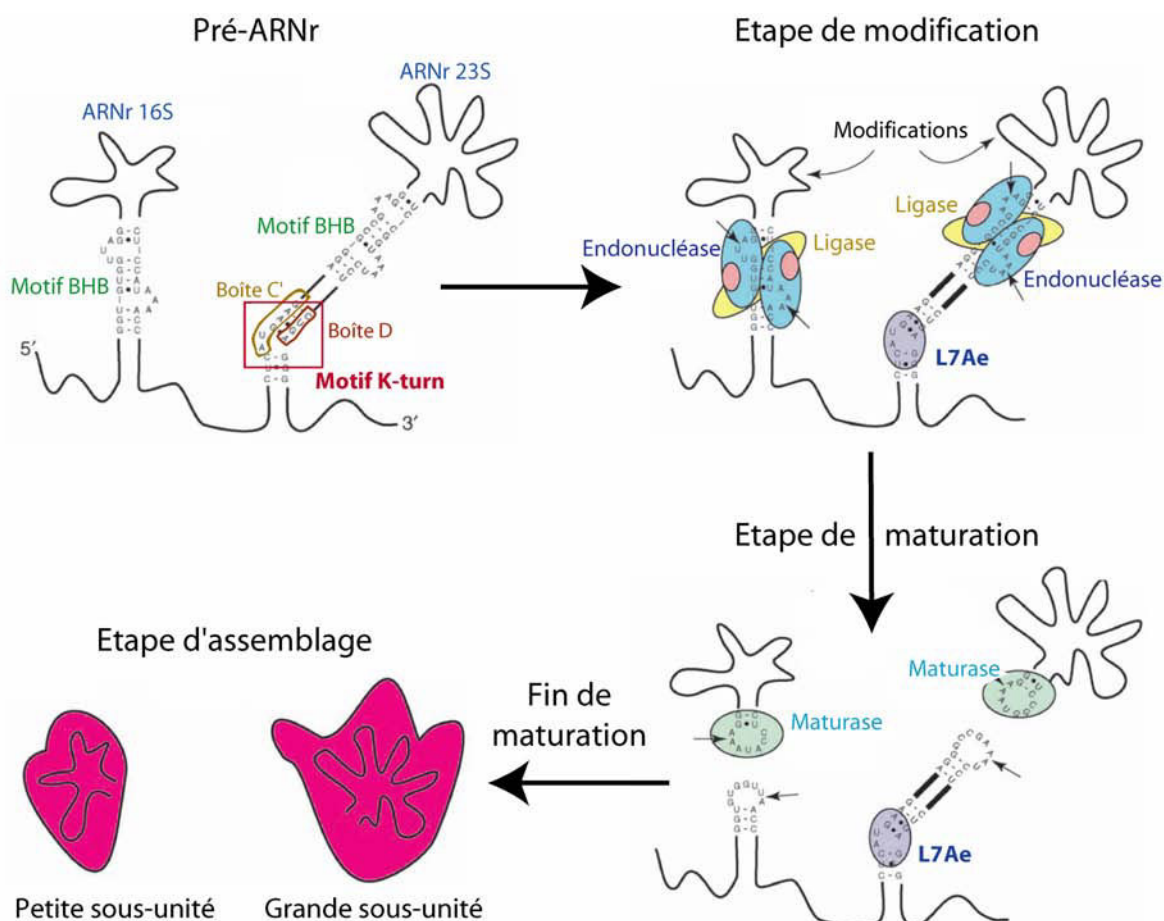


Figure II.32 : Formation de motifs en K-turn lors des étapes de maturation des pré-ARNr (adapté de Dennis et Omer, 2005).

Représentation schématisée des étapes de modification, maturation et assemblages des petite et grande sous-unités ribosomales proposées pour *S. solfataricus*. La maturation est initiée par excision des ARNr 16S et 23S au niveau des motifs BHB (bulge-helix-bulge). Les sites de clivage y sont indiqués par des flèches. Les boîtes C'/D présentes dans une région flanquant l'ARNr 23S forment un motif K-turn (encadré en rouge). Ce motif K-turn interagit avec la protéine L7Ae.

3.4. D'autres ARNnc chez les Archaea ?

Les archaea présentent les ARNnc communs aux trois domaines du vivant ainsi que des sRNA à boîtes C/D ou H/ACA découverts depuis peu. Par ailleurs, deux études très récentes proposent la présence d'autres familles d'ARNnc chez les archaea (Tang *et al.*, 2005; Zago *et al.*, 2005). Ces recherches d'ARNnc portent sur *S. solfataricus* et ont été réalisées par clonage/construction de banques d'ADNc ainsi que par co-immunoprécipitation avec L7Ae. La fonction des ARNnc identifiés reste encore hypothétique puisqu'elle a seulement été prédite sur la base de comparaison de séquences par voie bioinformatique. Les seules données expérimentales sont l'expression de ces ARNnc et l'interaction de certains d'entre eux avec L7Ae.

L'étude par clonage/construction de banques de cDNA révèle 57 nouveaux ARNnc répartis par les auteurs en 3 classes : les ARN antisens, les ARN antisens à boîtes (dont seule une partie de l'ARN est un antisens) et les ARNnc dont le mode d'action est indéterminé (Tang *et al.*,

2005). Parmi les 19 ARN antisens de taille comprise entre 45 et 100 nt, 8 présentent des complémentarités de séquence avec des ORF ou des régions non-traduites des transposases, 6 avec des ORF et 5 avec des ARNnc (ARNt ou sRNA à boîtes C/D). Ces ARNnc pourraient donc jouer un rôle de contrôle de la mobilité d'éléments d'insertion autonomes, d'inhibiteur traductionnel ou dans la régulation de la fonction d'autres ARNnc. Parmi les 24 ARN antisens à boîtes détectés qui sont tous exprimés à partir de régions intergéniques, 17 sont de nouveaux sRNA à boîtes C/D (à ajouter aux 13 prédits dans le Tableau II.1) et 6 ne présentent pas de séquence conservée mais présentent tous, à la manière des miRNA eucaryotes, une région de complémentarité particulièrement longue avec une partie 3'UTR d'une ou plusieurs ORF (Fig. II.33). Enfin, 14 autres ARNnc, dont le gène est localisé dans des éléments répétés (1), dans des régions entièrement intergéniques (11) ou légèrement chevauchantes avec des ORF (2), ne présentent aucun motif caractéristique et aucune complémentarité de séquence avec une région n'a été établie. Cependant, ces ARNnc présentent des homologues dans deux autres espèces de *Sulfolobus* et un nombre important de changements compensatoires permettent de proposer une structure secondaire conservée (Fig. II.33).

La seconde approche consistait à identifier des ARNnc de taille inférieure à 500 nt en interaction avec la protéine L7Ae (Zago *et al.*, 2005). Elle a permis le recensement de 45 nouveaux ARNnc qui ont été regroupés en 6 groupes dont certains ont été découverts dans l'approche décrite ci-dessus. Le premier groupe comporte 6 sRNA à boîtes C/D et 1 sRNA à boîtes H/ACA (Sso109). Nous avons également identifié ce guide H/ACA par approche bioinformatique (voir la partie *Résultats*). Le deuxième groupe comporte des parties d'ARNm et le troisième des séquences provenant de régions intergéniques. Le 4^{ième} groupe compte 7 ARN antisens dont deux présentent des motifs en K-turn. Les 5^{ième} et 6^{ième} groupes comportent des morceaux de l'ARN 7S et d'ARNr, respectivement. D'une part, ces résultats sont en accord avec ceux de la première approche et d'autre part, ils attirent l'attention sur la place majeure que pourrait occuper L7Ae en interaction avec des motifs en K-turn présents dans de nombreuses familles d'ARNnc (pour revue, Dennis et Omer, 2005). Les auteurs vont même jusqu'à proposer que L7Ae pourrait être un composant central de particules apparentées au complexe RISC. Néanmoins, aucun équivalent de protéine Argonaute, ni aucun événement de clivage d'ARNm par des ARN antisens n'a encore été démontré dans les génomes d'Archaea. Enfin, la présence, dans les génomes d'Archaea, d'équivalents de la protéine Hfq (Brennan et Link, 2007) ainsi que d'équivalents de protéines Sm (Thore *et al.*, 2003) suggèrent que de nombreux mécanismes impliquant des ARNnc associés à ce type de protéines sont encore inexplorés.

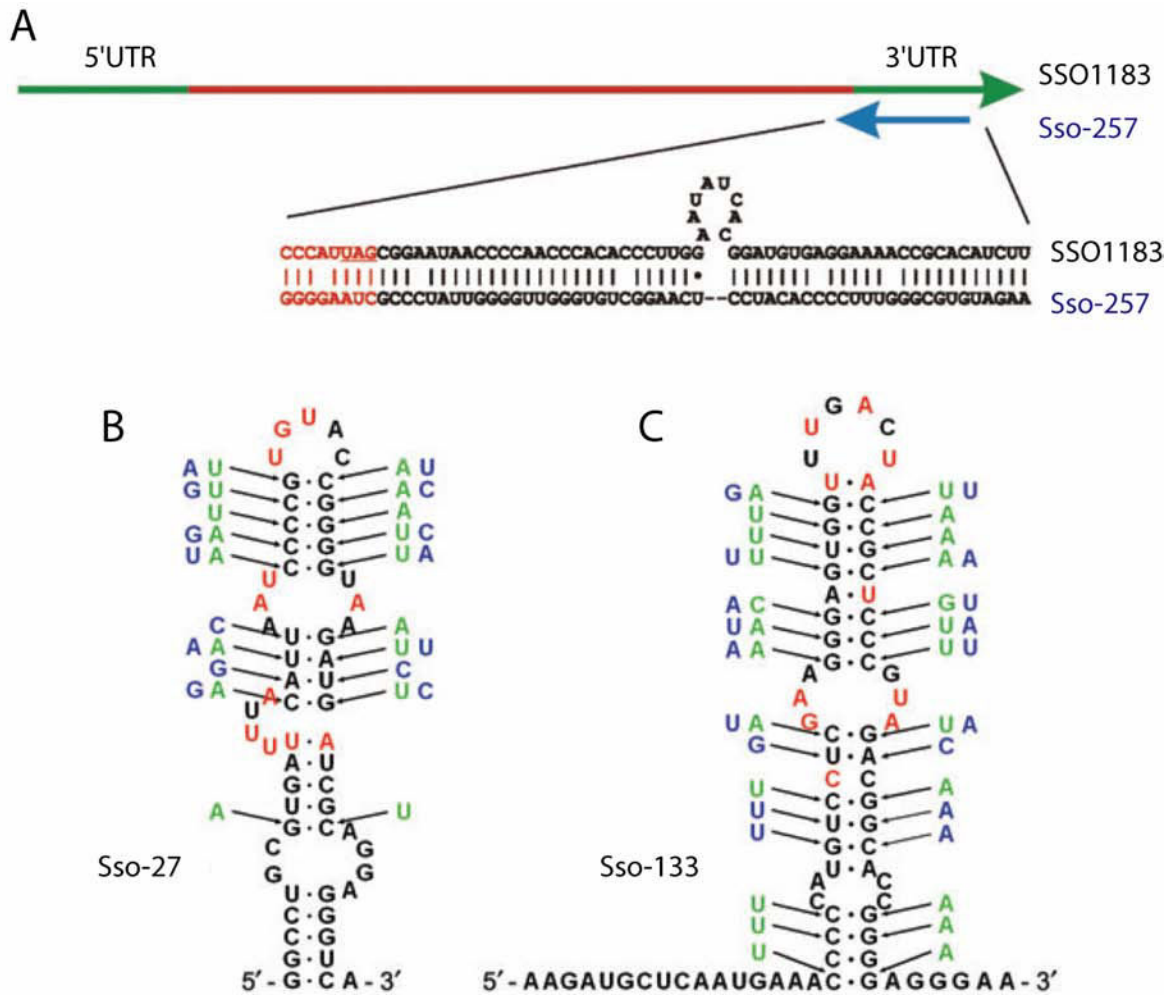


Figure II.33 : Exemples de nouvelles familles d'ARNnc chez *S. solfataricus* (d'après Tang *et al.*, 2005).

A. Exemple d'ARN à boîtes antisens : interaction suggérée entre l'ARN Sso-257 et l'ARNm SSO1183 codant un transporteur de phosphate inorganique. Les régions non traduites (UTR) sont représentées en vert. **B. et C.** Exemples d'ARNnc de classe inconnue et ne présentant pas de complémentarité de séquence avec une cible ARN : structures secondaires prédites pour deux ARNnc appelés Sso-27 et Sso-133. Les nucléotides marqués en rouge sont conservés dans les homologues identifiés chez *S. acidocaldarius* et *S. tokodaii*. Les changements compensatoires de nucléotides dans les homologues de *S. acidocaldarius* et de *S. tokodaii* sont représentés en bleu et en vert, respectivement.

2. Méthodes de recherche

Le fait que les ARNnc soient passés inaperçus pendant autant de temps est en grande partie dû à leur variété et à l'absence de signature génétique évidente permettant de les détecter *in silico*. Des méthodes sensibles et spécifiques de recherche systématique de tous les types ou d'un type donné d'ARNnc ont commencé à voir le jour, permettant ainsi le criblage des trois Domaines du Vivant. L'ensemble des méthodes conduisant à l'étude de tous les ARNnc d'un organisme donné relève de la transcriptomique.

A. Méthodes expérimentales

1. Séquençage ARN

Principe

Le séquençage ARN est une technique qui date de plus de 40 ans et qui a permis l'étude des ARNt, ARNr, UsnRNA et autres ARN tels que les ARN viraux. La technique consiste à séparer les ARN totaux le plus souvent sur gel dénaturant, à extraire et éluer la bande d'intérêt représentant idéalement un seul type d'ARNnc, à marquer radioactivement cet ARN puis à séquencer par voie chimique ou enzymatique cet ARNnc (Fig. II.34 panneau A) (Yarus et Barrell, 1971 ; Ehresmann *et al.*, 1977). L'étude de grands ARN est en général facilitée par la digestion préalable par la RNase T1 (Branlant *et al.*, 1981a ; Branlant *et al.*, 1981b).

Inconvénients et limites

Cette technique ne permet pas d'étudier des ARN peu abondants et non visibles après séparation sur gel. De plus, il faut que l'ARNnc d'intérêt ait une taille qui permette de le distinguer des autres ARN. Il faut ajouter à cela que les résultats de séquençage chimique ou enzymatique sont parfois difficiles à interpréter. Par conséquent, cette technique n'est pas à privilégier pour une étude systématique et fiable de toutes les familles d'ARNnc d'un type cellulaire.

2. Clonage et construction de banques de cDNA

Principe

Cette technique repose sur la synthèse d'un brin d'ADN complémentaire (ou ADNc) à tous les ARN extraits d'un type cellulaire donné et ceci grâce à une transcriptase inverse qui utilise comme amorce un oligonucléotide oligodT qui s'apparie aux queues polyA. Etant donné que les petits ARNnc sont en général dépourvus de queue polyA il est possible, parmi différentes stratégies, d'ajouter une queue polyA par l'utilisation d'une polyA-polymérase (pour revue, Huttenhofer et Vogel, 2006). Le résultat est donc comparable aux banques EST (Expressed Sequence Tag) représentant tous les ARNm codant des protéines à la différence ici que les ARN totaux sont préalablement séparés de telle sorte que seuls les ARN de taille inférieure à 500 nt, moins susceptibles d'être des ARNm, sont conservés (Fig. II.34 panneau B). Il est également possible, lors de ce traitement, de ne retenir que les ARN qui interagissent avec une protéine particulière par le biais d'anticorps dirigés contre cette dernière. On parle dès lors

d'immunoprécipitation. Dans ce cas, la séparation ne dépend plus de la taille mais de l'affinité pour une protéine et conduit au final à l'identification spécifique, et non plus généraliste, d'ARNnc d'une famille donnée voire de plusieurs familles qui partagent une interaction avec la même protéine. Cependant, les ARNnc doivent restés fortement liés avec cette protéine au cours de l'immunoprécipitation, ce qui suggère une détection incomplète des représentants de ces familles d'ARNnc.

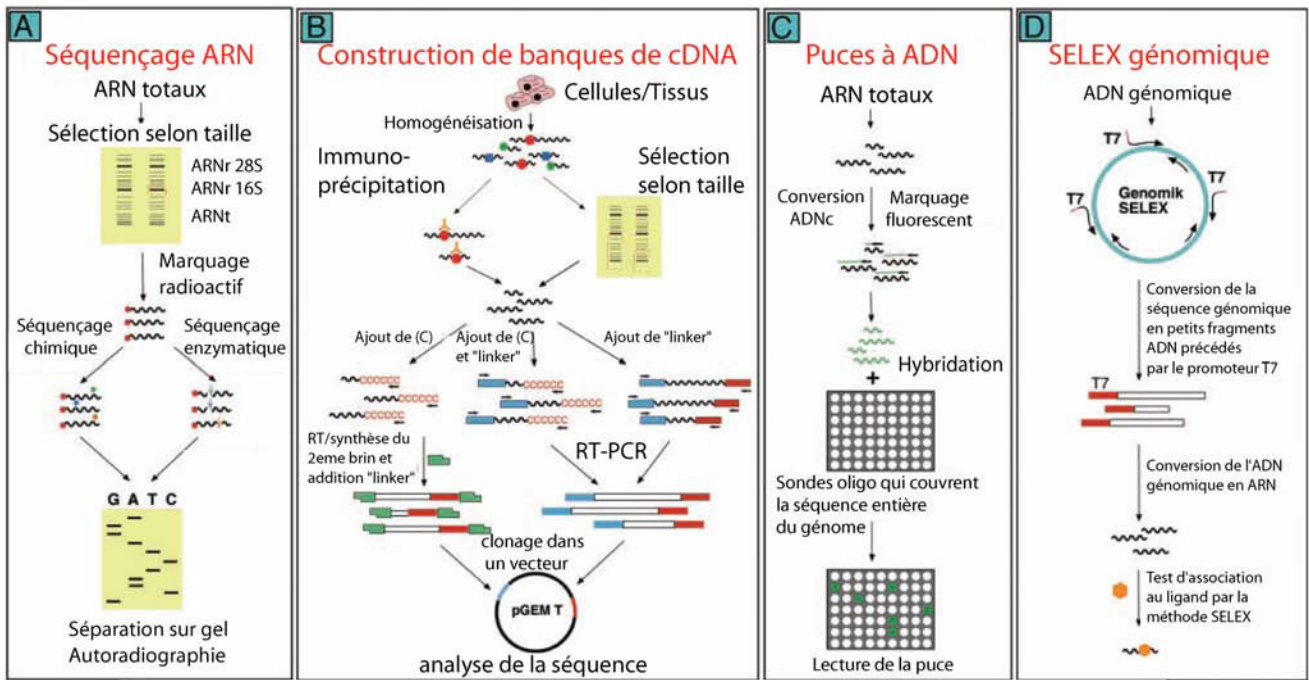


Figure II.34 : Approches permettant l'identification expérimentale d'ARNnc (d'après Huttenhofer et Vogel, 2006).

Quatre approches expérimentales sont présentées : le séquençage ARN (A), la construction de banques de cDNA (B), les pucés à ADN (C) et le SELEX génomique (D). L'immunoprécipitation est une variante de la construction de banques de cDNA (B). Dans cette dernière approche (B) trois méthodes sont indiquées pour la transcription d'ADnc à partir des ARNnc suivant l'ajout de queues de résidus C ou l'ajout d'oligonucléotides dits « linker » aux extrémités 5' ou à la fois 5' et 3' permettant une réaction de RT-PCR.

Les fragments d'ADnc sont ensuite clonés dans un vecteur puis séquencés. A ce stade un grand nombre de clones peut être obtenu pour un ARNnc abondant. Une analyse bioinformatique conduit ensuite à la localisation du locus de ces ARNnc dans le génome mais aussi à l'identification de structures ou motifs particuliers participant à la compréhension de leur fonction ainsi qu'à leur classification.

Cette méthode s'est montrée efficace pour l'identification de nombreux nouveaux ARNnc chez *Mus musculus* (Huttenhofer *et al.*, 2001), *Arabidopsis thaliana* (Marker *et al.*, 2002), *Drosophila melanogaster* (Yuan *et al.*, 2003), *Aquifex aeolicus* (Willkomm *et al.*, 2005) et chez deux archaea, *A. fulgidus* (Tang *et al.*, 2002a) et *S. solfataricus* (Tang *et al.*, 2005). L'immunoprécipitation, variante de cette méthode, a également permis l'identification de

nouveaux représentants d'une classe connue, notamment des snoRNA et scaRNA (Darzacq *et al.*, 2002; Vitali *et al.*, 2003).

Inconvénients et limites

Cette approche repose sur l'utilisation de la transcriptase inverse. Or, son efficacité est influencée par la présence de structurations, de modifications des bases et des riboses ainsi que par la taille de l'ARN à transcrire. Ainsi certains ARNnc seraient sous-représentés dans la banque d'ADNc voire en seraient complètement absents alors que certains ARNnc seraient sur-représentés. Notons que les longs ARNnc sont exclus de l'analyse puisque de taille supérieure à 500 nt. De plus, cette approche ne permet de détecter que les ARNnc exprimés dans un type cellulaire étudié, à un stade du développement donné et dans des conditions de milieu données. Pour cloner tous les ARNnc potentiellement présent dans la cellule il faudrait faire varier tous ces paramètres, ce qui est impossible. Enfin, cette approche, en plus d'être laborieuse est très coûteuse.

3. Puces à ADN

Principe

Les puces à ADN permettent de mesurer le taux d'expression de différents ARNm mais également, à l'échelle du génome, de détecter l'expression de séquences non décrites comme ORF. Les puces sont des supports de verre ou de silicone qui portent différentes sondes oligonucléotidiques de 25 à 70 nt arrangées sur la base d'une grille. Il est également possible de fixer des fragments PCR double-brin. Les sondes sont marquées avec des colorants fluorescents tels que Cy3 ou Cy5 (pour revue, Stoughton, 2005). Un échantillon contenant des ARN totaux est mis au contact de ces sondes. La fluorescence des différentes sondes apparaît sous forme de points mesurés par un scanner. L'intensité de la fluorescence reflète la quantité de transcrits présents. Deux marqueurs peuvent être utilisés pour comparer les taux d'expression de deux échantillons différents. Les premières puces à ADN incluant des sondes complémentaires à des régions intergéniques ont été utilisées pour *E. coli* et ont révélé la présence de nombreux ARN antisens (Selinger *et al.*, 2000 ; Wassarman *et al.*, 2001 ; Tjaden *et al.*, 2002). Cette technique a aussi été utilisée pour analyser les ARNnc associés à la protéine Hfq, combinant ainsi immunoprécipitation avec Hfq et puces à ADN (Zhang *et al.*, 2003). Les puces à ADN sont maintenant couramment utilisées pour étudier les ARNnc (pour revue, Huttenhofer et Vogel, 2006).

Inconvénients et limitations

Les principales limitations proviennent de la capacité d'une sonde ADN à s'apparier avec une séquence ARN emprisonnée dans une structure secondaire ainsi que de la distinction difficile entre signal et bruit de fond. Ainsi, les 3000 à 4000 ARN antisens présents sur le brin complémentaire de nombreuses ORF identifiées chez *E. coli* pourraient n'être que du bruit de fond (Tjaden *et al.*, 2002). L'utilisation de puces à ADN couvrant les régions interORF est très coûteuse. De plus, celles-ci peuvent conduire à des résultats qui ne sont pas toujours en accord

avec ceux obtenus par Northern Blot. Enfin, il faut rappeler ici encore que le profil d'expression ainsi obtenu ne traduit que l'expression d'ARNnc dans un type cellulaire, à un stade du développement, dans des conditions de milieu données. La répétition des expériences avec des échantillons différents traduisant différents états cellulaires est facilitée ici mais reste laborieuse.

4. SELEX génomique

Principe

La technique consiste à identifier tous les ARNnc qui pourraient interagir avec une protéine connue. L'approche SELEX est basée sur la transcription *in vitro* d'ARN qui dérivent d'une banque d'ADN d'un génome entier par utilisation d'amorces aléatoires (« random priming ») (Fig. II.34 panneau D). La masse d'ARN ainsi exprimée subit différents cycles d'association avec la protéine d'intérêt de telle sorte qu'à chaque cycle les ARN qui fixent le partenaire protéique soient enrichis. Après plusieurs cycles et une fois que les séquences des ARN liés à la protéine sont déterminées, des correspondances entre celles-ci et la séquence du génome sont établies. Des régions du génome sont ainsi prédites comme abritant de potentiels ARNnc dont l'expression est ensuite testée expérimentalement. Cette approche très récente a été utilisée pour rechercher des ARNnc liés à la protéine Hfq chez *E. coli* (Lorenz *et al.*, 2006).

Cette technique peut permettre de détecter des ARNnc non décelés par les trois approches précédentes. De plus, à l'inverse des deux précédentes elle ne requiert pas l'examen de tous les états de croissance, de développement ou de différenciation d'une cellule.

Inconvénients et limitations

Il y a encore trop peu de recul pour recenser les inconvénients. Cependant, une limitation provient du fait que peu de protéines sont connues pour former des complexes spécifiques avec différentes familles d'ARNnc. Par conséquent, cette méthode serait plus pertinente pour identifier tous les représentants d'une famille particulière d'ARNnc.

B. Méthodes bioinformatiques

Les techniques expérimentales de recherche des gènes d'ARNnc sont longues et ont un coût élevé. Par ailleurs, elles sont loin d'être exhaustives puisqu'elles ne permettent de décrire des ARNnc exprimés que dans un type cellulaire et des conditions données, ceci à moins de les répéter dans différents types cellulaires et dans différentes conditions. Des alternatives bioinformatiques plus performantes sont envisageables. Bien que de nombreux algorithmes et logiciels aient permis l'identification d'ORF et l'attribution d'une fonction potentielle à certaines des protéines codées, de tels programmes ne sont pas utilisables pour la prédiction de gènes d'ARNnc *in silico*. En effet, l'absence dans les gènes d'ARNnc de signaux statistiques forts tels que le biais résultant de la présence de codons, leur petite taille et le fait que leur séquence évolue parfois plus rapidement que celle des protéines sont autant de caractéristiques qui rendent leur identification *in silico* difficile.

La recherche systématique d'ARNnc par voie bioinformatique a débuté à la fin des années 1990 avec la prise de conscience de leur importance et l'aboutissement des premiers projets de séquençage de génomes entiers. Le développement d'outils d'analyse dédiés à la recherche d'ARNnc est donc un challenge de premier ordre dans l'ère post-génomique (Eddy, 2002). Les programmes existants sont loin d'être aussi performants que les programmes de détection d'ORF et ils nécessitent encore d'être améliorés. Par ailleurs, des programmes dédiés à la recherche de classes spécifiques d'ARNnc méritent d'être développés. Les génomes de procaryotes, du fait de l'absence d'épissage, de leur petite taille et d'annotations approfondies ont souvent été utilisés pour tester l'application de ces approches.

Les méthodes de prédiction actuelles peuvent être distinguées selon qu'elles s'appuient sur :

- la localisation des gènes d'ARNnc,
- la structuration des ARNnc et leur stabilité thermodynamique,
- le biais dans la composition de la séquence des ARNnc,
- la conservation de séquence ou de structure des ARNnc entre espèces proches d'un point de vue phylogénétique,
- la présence de signaux d'initiation et de terminaison de transcription,
- les homologues qui existent dans une même famille d'ARNnc.

Dans un premier temps, il faut choisir le critère de sélection et le principe de la méthode qui conduira à un signal permettant de distinguer significativement une séquence susceptible de contenir le gène d'un ARNnc du bruit de fond correspondant à des séquences ne présentant aucun ARNnc. Ces critères de recherche sont plus ou moins discriminatifs. Ils peuvent être combinés grâce à l'emploi de traitements algorithmiques parfois très sophistiqués, l'objectif étant dans chaque cas la prédiction d'un nombre restreint d'ARNnc dont l'expression et la fonction potentielle doivent être testées expérimentalement. Le choix primordial du ou des critères de sélection dépend du type de génome analysé et des informations disponibles. En effet, les caractéristiques d'ARNnc déjà identifiées orientent également le choix de ces critères.

Parmi les méthodes développées certaines visent l'identification générale des ARNnc quelque soit la famille à laquelle ils appartiennent alors que d'autres visent l'identification spécifique des représentants d'une famille d'ARNnc précise. Dans ce second cas, des connaissances préalables concernant la famille d'ARNnc sont requises.

1. Identification non spécifique de tous les gènes d'ARNnc d'un génome

Les approches dites généralistes ne nécessitent aucune connaissance préalable concernant les familles d'ARNnc potentiellement présentes. Il est possible de distinguer deux types d'approches différentes selon les informations requises. D'une part, des approches dites intrinsèques ne requièrent que la séquence génomique et les annotations de l'organisme étudié. D'autre part, des approches dites extrinsèques utilisent les séquences et annotations de plusieurs génomes. Ces dernières relèvent de la « génomique comparative ».

1.1. Signaux de transcription et localisation des gènes d'ARNnc

Les premiers travaux ont été focalisés sur l'examen des régions intergéniques de grande taille (Olivas *et al.*, 1997). En particulier, chez *S. cerevisiae* dont le génome avait été annoté, des régions intergéniques de taille supérieure à 2 kb étaient susceptibles de renfermer soit de petites ORF jusque là non détectées soit des gènes d'ARNnc. En effet, les gènes d'ARNnc déjà connus étaient en général localisés dans des régions qui ne comportaient aucune ORF sur les deux brins (Storz, 2002). Ainsi, l'identification, dans ces régions, de signaux de transcription qui n'étaient pas attribuables à des ORF adjacentes pouvait conduire à la détection de gènes d'ARNnc. En outre, l'utilisation de programmes performants pour la détection d'ORF, tels que par exemple GeneMarks (Besemer *et al.*, 2001) permet de différencier les gènes d'ARNnc d'ORF potentielles.

Un crible similaire a été réalisé chez *E. coli* (Argaman *et al.*, 2001). Les auteurs ont recherché des régions interORF d'*E. coli* comportant des séquences correspondant l'une à un promoteur, l'autre à un terminateur, avec la contrainte que les deux séquences soient peu distantes l'une de l'autre (de 50 à 500 nt). Dans ce but, la séquence consensus des promoteurs reconnus par l'holoenzyme $E\sigma^{70}$, ainsi que des motifs de terminaison rho indépendants ont été recherchés. L'emploi de cette dernière approche a conduit à proposer l'existence de 24 gènes d'ARNnc potentiels qui sont, de plus, conservés chez d'autres bactéries et dont l'expression de 14 d'entre eux a pu être mise en évidence. Une recherche similaire également restreinte aux régions interORF mais ne se limitant pas aux ARNnc conservés a révélé 7 nouveaux ARNnc certainement plus spécifiques à *E. coli* K12 (Chen *et al.*, 2002). Il faut cependant remarquer que cette approche nécessite une connaissance préalable des séquences reconnues par les différentes holoenzymes des ARN polymérase de l'espèce. Or, les promoteurs ne sont pas aussi bien décrits chez les Archaea (voir Partie I). Par ailleurs, la multiplicité d'ARN polymérases et de facteurs de transcription chez les Eucaryotes limite l'emploi d'une telle approche. De plus, dans ce dernier Domaine, certains ARNnc dont les miRNA et les snoRNA ne sont pas transcrits indépendamment mais sont des produits de maturation de polycistrons ou d'introns. Le fait que les ARNnc aient une localisation bien définie au sein d'introns (Hirose *et al.*, 2006) pourrait permettre de les identifier si les sites d'épissage étaient faciles à prédire, mais ce n'est pas le cas.

1.2. Utilisation d'un biais dans la composition de la séquence

Une séquence codant des protéines présente un biais du fait de l'attribution à chaque acide aminé (code à 21 lettres) à un ou plusieurs codon(s) composé(s) de trois nucléotides (soit 64 possibilités). Le maintien nécessaire de la fonction de la protéine et donc de certains éléments de séquence ou de structure crée des contraintes de séquence importantes au niveau de son ORF. Des contraintes existent aussi au niveau des séquences des ARNnc. Elles sont inhérentes au maintien de la structuration et de la présence de motifs fonctionnels au cours de l'évolution. Ainsi les possibilités de mutation à chaque position ne sont pas identiques. En particulier celles pouvant préserver la structure secondaire sont souvent observées par paires de façon à maintenir

les possibilités de former une hélice. Ceci se traduit par une composition et un enchaînement de nucléotides différents de ce à quoi on pourrait s'attendre dans le cas d'une évolution aléatoire non influencée par une pression de sélection. Ainsi, suivant la composition des séquences intergéniques non transcrites variable suivant le génome considéré, un gène d'ARNnc peut se distinguer du reste du génome par une composition en A, C, G et T différente. Du fait de la nécessité d'avoir une forte structuration secondaire, le pourcentage élevé en G/C des gènes d'ARNnc, en particulier chez les hyperthermophiles, a permis l'identification de nouveaux gènes candidats dans les génomes d'Archaea (Klein *et al.*, 2002 ; Schattner, 2002).

Ce biais a aussi été utilisé dans des approches basées sur l'apprentissage statistique. L'apprentissage statistique permet, à partir d'un échantillon d'éléments appartenant à au moins deux classes, de prévoir la valeur d'une variable capable de caractériser l'appartenance d'un élément à une de ces classes. Au final, il s'agit de trouver une fonction de décision qui permette de placer tout nouvel élément avec le moins d'erreur possible soit dans la classe des gènes d'ARNnc soit dans la classe des séquences purement intergéniques. Les machines à vecteurs support (SVM) sont des systèmes d'apprentissage automatique qui ont la possibilité de gérer plusieurs classes (Vapnik, 1998). Les SVM « apprennent », à partir de la connaissance acquise sur un échantillon d'apprentissage. Pour cela, elles définissent des variables permettant de séparer correctement en classes les éléments afin de pouvoir classer de nouvelles séquences dans les classes définies lors de l'apprentissage. Les SVM sont de plus en plus utilisés pour annoter les séquences génomiques et, récemment, un outil basé sur des SVM, PsoL, a été développé pour l'identification d'ARNnc chez les bactéries (Wang *et al.*, 2006). Il faut remarquer que cette approche dépend fortement du fichier d'apprentissage et donc des informations d'annotation disponibles sur le génome, ceci avec un risque que ces données soient entachées d'erreurs. Cette recherche requiert donc des informations préalables très précises.

1.3. Utilisation de critères thermodynamiques

Cette méthode est basée sur l'hypothèse que les produits de transcription des gènes d'ARNnc présentent des structures secondaires significativement plus stables que ce que serait des produits de transcription de séquences génomique ne contenant pas de gène d'ARNnc. Une approche basée sur la version probabiliste de l'algorithme de l'énergie minimum de repliement développé par M. Zuker (MFOLD, (Zuker et Stiegler, 1981) a été développé pour la détection systématique d'ARNnc (Rivas et Eddy, 2000). Néanmoins, il s'est avéré que le signal correspondant aux ARNnc ne peut être facilement distingué du bruit de fond, ce qui suggère qu'une méthode uniquement basée sur des calculs thermodynamiques n'est pas utilisable. Bien que la structuration de la plupart des ARNnc soit un critère important pour leur recherche, il est donc nécessaire de combiner ce critère avec d'autres critères tels que, par exemple, le biais dans la composition des nucléotides (Carter *et al.*, 2001). Cette combinaison, grâce notamment à l'utilisation d'un réseau neuronal d'apprentissage statistique, a permis de mettre en évidence 370 candidats ARNnc chez *E. coli*.

1.4. Génomique comparative

Principe

La plupart des approches présentées ci-dessus ont été associées à une recherche de la conservation phylogénétique des gènes candidats détectés, ceci avant vérification expérimentale que les ARNnc candidats sont bien exprimés. Il est aussi possible que la stratégie de recherche bioinformatique d'ARNnc débute par la recherche de séquences conservées entre espèces proches au niveau des régions interORF. En effet, les séquences génomiques conduisant à des produits de transcription fonctionnels évoluent moins vite que les séquences non-fonctionnelles (pour revue, Frazer *et al.*, 2003). Retrouver des séquences conservées revient donc à distinguer les gènes de protéines ou d'ARNnc et leurs séquences régulatrices, du reste de la séquence génomique.

Interprétation des résultats obtenus

En limitant cette recherche aux régions interORF elle conduit à la détection de gènes d'ARNnc mais aussi à la détection d'ORF non annotées et de séquences régulatrices (Wassarman *et al.*, 2001). Une limitation évidente de cette approche est donc de ne pas identifier les ARNnc *cis*-antisens transcrits sur le brin opposé aux ORF. Le développement d'une telle approche est plus complexe chez les eucaryotes du fait que la délimitation des ORF est compliquée par l'existence d'introns. Les séquences régulatrices impliquées dans la régulation de la réplication, la transcription, la traduction, les transpositions et autres processus cellulaires sont en général répétées plusieurs fois dans un organisme et ont des tailles inférieures aux tailles attendues pour les ARNnc. Ceci peut être utilisé pour distinguer ces séquences des gènes d'ARNnc

Quel critère de sélection retenir : la séquence ou la structure ?

Pour développer une approche de génomique comparative appliquée aux ARNnc il faut savoir si il est préférable de considérer que la conservation repose plus sur la séquence ou sur la structure secondaire de l'ARNnc. Si la fonction peut dépendre fortement de la structure de l'ARNnc avec la conservation de très courts segments, il est préférable de rechercher une conservation de structure. Si, au contraire, la fonction réside dans la conservation de séquences assez longues permettant d'établir des appariements canoniques avec un ARN cible, il est préférable de rechercher une conservation de séquence. Il est souvent souhaitable de pouvoir combiner ces deux types de recherche d'ARNnc.

Choix des génomes

Une conservation de séquence ou de structure entre deux gènes d'ARNnc peut résulter soit de l'existence de contraintes fonctionnelles des transcrits soit du manque de divergence entre les génomes considérés (Dubchak *et al.*, 2000). Le choix des génomes soumis à la comparaison est donc primordial. Il faut que ces génomes aient suffisamment divergé pour que l'action de la pression de sélection soit décelable et permette de distinguer les ARNnc des régions ne comportant aucun gène (Cliften *et al.*, 2001). Par contre, si les génomes ont trop divergé, les séquences ou structures des ARNnc peuvent avoir accumulé trop de différences pour que des

similarités de séquence ou de structure soient observées, à l'exception toutefois des ARNr et autres ARNnc universels dont la séquence de certaines régions varie peu au cours de l'évolution et dont la structure 2D est conservée (Branlant *et al.*, 1981c ; Eddy, 1999). S. Eddy a développé un modèle statistique permettant de prédire le nombre de séquences génomiques requises dans une comparaison visant à détecter des gènes d'ARNnc en fonction de la distance phylogénétique qui sépare ces génomes et de la taille des éléments conservés à détecter (Eddy, 2005).

Approches développées

Plusieurs algorithmes prenant en compte la conservation de la structure secondaire ont été développés (Rivas et Eddy, 2001 ; Coventry *et al.*, 2004; Di Bernardo *et al.*, 2004). Ils sont inspirés d'une approche utilisée pour la détection d'ORF (Badger et Olsen, 1999) qui consiste, à partir d'un alignement de séquences, à repérer des substitutions permettant d'échanger un codon donné par un codon synonyme. De manière similaire, un modèle dit de covariance opère, à partir d'un alignement, en mettant en évidence des changements de bases compensatoires vis-à-vis de la structure secondaire (Fig. II.35). Cette méthode présente l'avantage de ne pas définir de structure *a priori*. Une approche, exploitant des SCFG (Stochastic Context Free Grammars) permettant une modélisation des formations de structure secondaire, a conduit à l'identification de 275 candidats ARNnc dans des espèces bactériennes, l'expression de seulement 11 d'entre eux a été vérifiée (QRNA, Rivas *et al.*, 2001). Il faut noter que les séquences utilisées dans l'alignement sont limitées en taille, aussi ce type d'analyse n'est pas adapté aux grands génomes eucaryotes. Par ailleurs, là aussi, si les ORF ne sont pas éliminées lors de la recherche, de nombreuses structurations portées par les ARNm sont recueillies par cette approche (i.e. ORF plus régions 5' et 3' UTR). De plus, une telle approche ne permet pas la détection d'ARNnc peu structurés, tels que, par exemple les snoRNA à boîte C/D (Dennis *et al.*, 2001). En revanche, une analyse de conservation de séquences peut conduire à la détection de ces snoRNA (Ray *et al.*, 2001).

En effet, la comparaison des séquences primaires s'est avéré être un outil puissant de détection de certaines classes d'ARNnc. On remarque, par exemple, que la comparaison des séquences interORF d'*E. coli* et *Salmonella typhi* révèle un taux de conservation parfois supérieur à 85% alors que pour la plupart des ORF, il est inférieur à 70%. Ainsi, deux gènes d'ARNnc peuvent partager plus de similarités de séquence que deux gènes de protéines. Ainsi, il est possible d'identifier un signal associé à un ARNnc qui se distingue significativement du bruit de fond (Cliften *et al.*, 2001). Cependant, une recherche de gènes d'ARNnc, uniquement basée sur la comparaison de séquences, requiert la présence d'un haut degré de conservation entre ARNnc et donc un choix adéquate des séquences génomiques utilisées (Cliften *et al.*, 2001; Ray *et al.*, 2001 ; Wassarman *et al.*, 2001). Une approche de génomique comparative reposant uniquement sur les comparaisons de séquences intergénomiques de plusieurs espèces bactériennes a abouti à la détection de 60 candidats ARNnc chez *E. coli* et l'expression de 17 d'entre eux a été mise en évidence (Wassarman *et al.*, 2001).

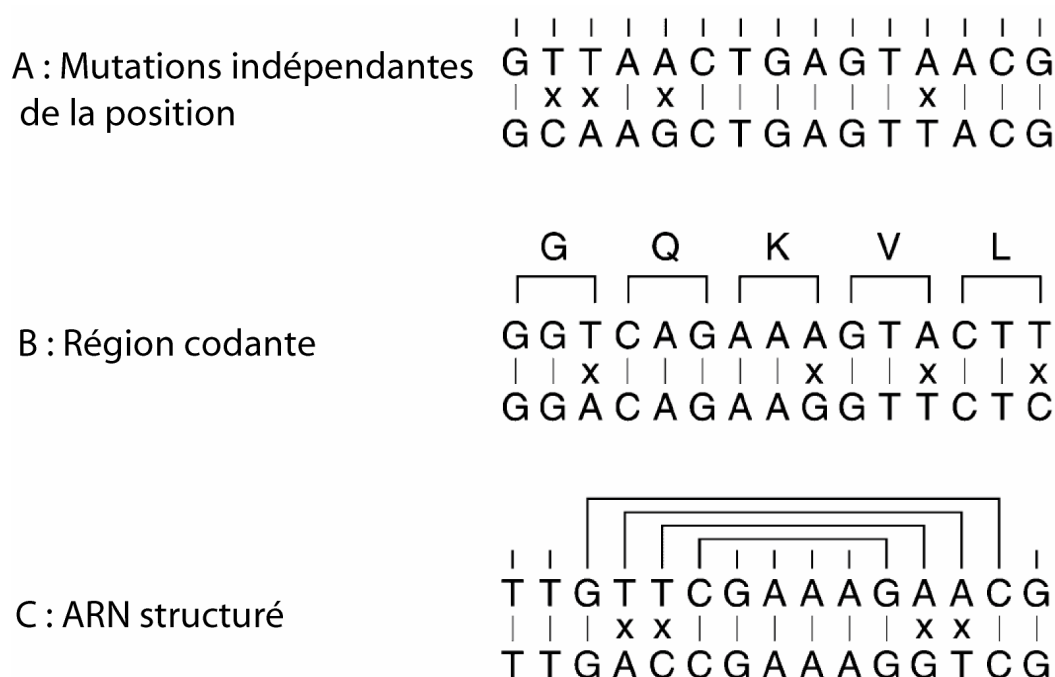


Figure II.35 : Modèles de substitution de bases traduisant soit la présence d'ORF soit la présence de gènes d'ARNnc (d'après Rivas et Eddy, 2001).

Ces trois alignements de séquence comportent le même nombre et le même type de substitution de nucléotides mais on peut y distinguer – (A) des séquences pour lesquelles les substitutions ont eu lieu indépendamment de leur position – (B) des séquences présentant des substitutions qui maintiennent la lecture des codons d'une ORF – (C) des substitutions qui se compensent pour préserver la possibilité de former une structure tige-boucle dans l'ARN.

1.5. Combinaison de différents critères

Les recherches bioinformatiques basées sur des critères différents ont souvent abouti à la prédiction d'ensemble d'ARNnc candidats présentant peu de chevauchement (Tableau II.2), ce qui suggère que ces approches sont soit peu sélectives soit non exhaustives (Eddy, 2002; Hershberg *et al.*, 2003). Aussi, de nombreuses approches combinant la recherche de plusieurs de ces critères ont été développées récemment (pour revue, Livny et Waldor, 2007). Elles combinent souvent la génomique comparative et la recherche de paramètres thermodynamiques (Washietl et Hofacker, 2004), avec parfois la recherche de signaux de transcription (Livny *et al.*, 2005) ou un biais dans la composition en nucléotide (Pichon et Felden, 2005). Certaines approches se sont avérées efficaces pour des génomes eucaryotes, c'est le cas par exemple de la méthode RNAZ qui, parmi les approches les plus récentes, présente les meilleures sensibilité et spécificité (Washietl *et al.*, 2005b). Cependant, la plupart de ces approches nécessitent un niveau élevé d'informations sur les génomes et les ARNnc de ces génomes, informations souvent uniquement disponibles chez les bactéries pour lesquelles un grand nombre de séquences génomiques phylogénétiquement proches sont accessibles.

Des programmes ayant pour vocation de regrouper ces différentes approches de prédiction ont été récemment établis. Par exemple, le programme sRNAPredict2 rassemble toutes les

prédictions et en élimine les redondances, ceci afin d'identifier rapidement tous les loci d'ARNnc d'un génome bactérien (Livny *et al.*, 2006). Ce programme identifie plus de 2700 candidats ARNnc dans 10 espèces bactériennes. De même des collections de gènes d'ARNnc sont accessibles sur le Web telles que Rfam (pour ncRNA family) (Griffiths-Jones *et al.*, 2003 ; Griffiths-Jones *et al.*, 2005). Elles rassemblent et classent en familles les ARNnc qui présentent des homologies. Enfin, une dernière approche consiste à retrouver, parmi les régions interORF ou les prédictions provenant d'autres approches, des séquences qui présentent des homologies avec des ARNnc déjà découverts (Saetrom *et al.*, 2005). Ces deux dernières approches sont pratiquement des approches spécifiques dédiées à la détection de tous les représentants d'une même famille d'ARNnc. Elles pourraient donc également être classées dans la catégorie d'approches que nous allons décrire dans le paragraphe suivant.

Stratégie	Références	Candidats	Testés	Exprimés	# / 31
Signaux de transcription	Argaman <i>et al.</i> , 2001	24	23	14	14
Signaux thermodynamiques et biais dans la composition	Carter <i>et al.</i> , 2001	370	-	-	18
Conservation de la structure secondaire	Rivas <i>et al.</i> , 2001	275	49	11	13
Conservation de la séquence primaire	Wassarman <i>et al.</i> , 2001	60	60	17	22
			Non redondants	31	

Tableau II.2 : Comparaison des résultats obtenus lors de la recherche systématique d'ARNnc chez *E. coli* (d'après Eddy, 2002).

Sont indiqués, en fonction de la stratégie utilisée, le nombre de candidats ARNnc prédits, le nombre de candidats testés par Northern blot et le nombre de candidats exprimés. Le nombre total de candidats ARNnc exprimés et non redondants identifiés par ces 4 stratégies est de 31. La dernière colonne montre, parmi ces 31 candidats, le nombre de candidats identifiés pour chacune des stratégies.

2. Identification des représentants d'une famille de gènes d'ARNnc

Une recherche exhaustive de tous les représentants d'une famille d'ARNnc qui s'appuierait uniquement sur l'utilisation d'outils d'alignement local ou global tels que Blast ou Fasta n'est pas suffisamment performante, à moins que la recherche concerne des ARNnc fortement conservés tels que le tmRNA bactérien ou les ARNr (pour revue, Vogel et Sharma, 2005). Il faut ajouter à cela que, même entre espèces proches d'un point de vue phylogénétique, certains ARNnc ne présentent que très peu de similarités de séquence à l'instar des ARNnc bactériens CsrB ou RyhB (Wilderman *et al.*, 2004; Lenz *et al.*, 2005).

Deux types de programmes alternatifs peuvent être distingués. D'une part des programmes spécialisés sont dédiés à la recherche d'une seule famille d'ARNnc. Un des premiers programmes développés dans ce sens a été tRNAscan SE qui permet la recherche quasi-exhaustive de tous les types d'ARNt présents dans une séquence génomique (Lowe et Eddy, 1997). De nombreux programmes dédiés sont aujourd'hui disponibles, notamment pour l'identification des snoRNA à boîtes C/D eucaryotes (SnoScan, Lowe et Eddy, 1999), de l'ARN 4.5S composant la SRP bactérienne (Regalia *et al.*, 2002), ou des snoRNA à boîtes H/ACA eucaryotes (SnoGPS, Schattner *et al.*, 2004).

Il existe aussi des programmes permettant à l'expérimentateur de spécifier lui-même le type d'ARNnc qu'il recherche en y précisant ses caractéristiques. On peut ainsi distinguer deux catégories de programmes ayant ces propriétés. Les premiers sont basés sur un descripteur. Editer un descripteur consiste, pour l'expérimentateur, à préciser l'ordre et la taille des structures simple/double brin ainsi que leur séquence primaire si celles-ci sont conservées. Un résultat positif respecte donc exactement les exigences précisées. RNAMot, utilisé au cours de notre travail (Gautheret *et al.*, 1993) est un descripteur de motifs appartenant à cette catégorie. Ce programme a déjà été appliqué à la recherche de plusieurs types de motifs tels que les motifs des ribozymes à têtes de marteau (ou Hammerhead) ou les structures tige-boucle SECIS (Selenocysteine Insertion Sequence) et les résultats ont été validés expérimentalement (Laferriere *et al.*, 1994 ; Bourdeau *et al.*, 1999 ; Lescure *et al.*, 1999; Lescure *et al.*, 2002). L'avantage de RNAMOT est que les contraintes de séquence et de structures sont définies indépendamment des détails de procédure propres au programme. Cependant, il ne permet pas la description de combinaisons de contraintes complexes en séquence et en structure, ni l'exigence de contraintes complexes telles que « si la taille de l'hélice I est supérieure à 4 pb il n'y a pas de séquence simple brin II...etc ». Cette approche a été améliorée par l'introduction de flexibilité dans la recherche ainsi que de la notion de score implémentées par plusieurs programmes tels que PALINGOL (Billoud *et al.*, 1996), PatScan (Dsouza *et al.*, 1997), PatSearch (Pesole *et al.*, 2000) ou RNAMotif (Macke *et al.*, 2001). Ces derniers programmes s'étaient avérés d'une efficacité limitée. Ainsi, bien que RNAMotif ait été appliqué avec succès à la recherche de boucles E dans les ARNr (Macke *et al.*, 2001) l'élaboration de descripteurs par ce programme reste subjective et des contraintes subtiles ne sont pas prises en compte.

Pour pallier à ces défauts une seconde catégorie regroupe les programmes qui ne s'appuient pas sur la description d'un motif mais sur l'utilisation d'un profil construit à partir d'un alignement de représentants d'une classe donnée d'ARNnc. Deux cas de figure peuvent encore être distingués. Lorsque la structure secondaire de la famille d'ARNnc recherchés est relativement bien connue, il est possible d'aligner les séquences selon la définition d'hélices ou de segments simple-brin, c'est le cas du programme ERPIN (Gautheret et Lambert, 2001). Ce dernier a été utilisé avec succès pour la recherche de plusieurs familles d'ARNnc dont certaines classes de miRNA ou les motifs SECIS impliqués dans l'incorporation de la sélénocystéine (Lambert *et al.*, 2002; Lambert *et al.*, 2004 ; Legendre *et al.*, 2005). Lorsque aucune information de structure secondaire n'est disponible, il est possible d'utiliser des programmes basés sur le modèle de covariance combiné à la minimisation de l'énergie libre de repliement tels que CMFinder (Yao *et al.*, 2006). Ceux-ci peuvent ainsi établir une structure secondaire commune aux séquences d'ARNnc soumis en entrée. Dans les deux cas, l'alignement aboutit à une signature propre à une classe d'ARNnc utilisée par les programmes pour rechercher des séquences comportant ces signatures.

Nous avons utilisé les programmes RNAMOT et ERPIN pour mettre en place une approche de recherche de gènes de sRNA à boîtes H/ACA dans les génomes d'Archaea.

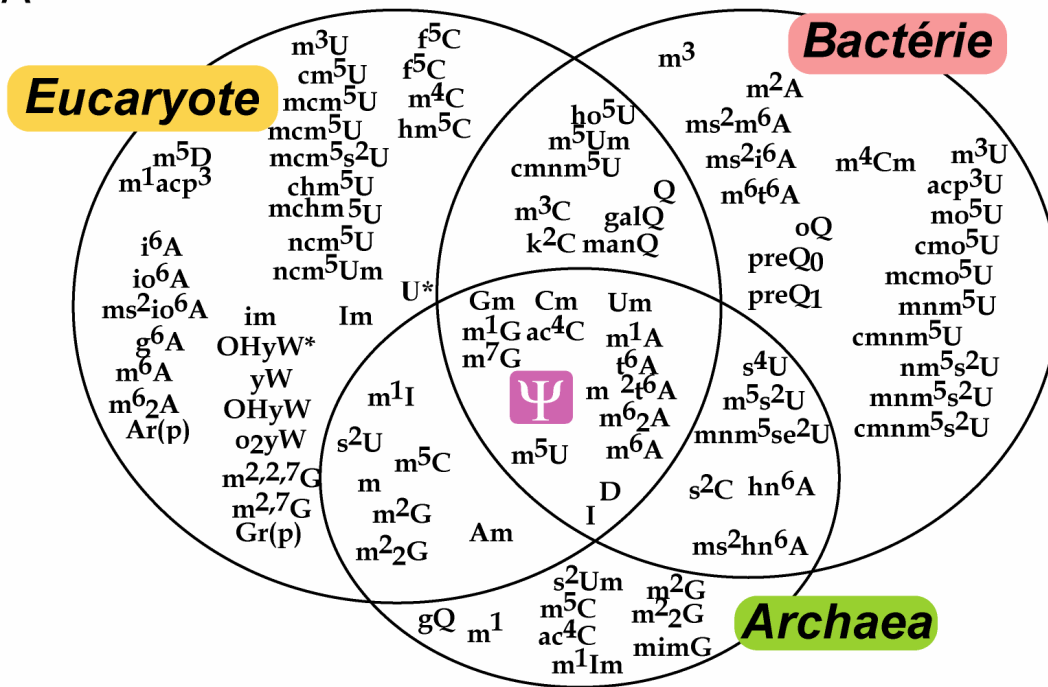
INTRODUCTION

Partie III : Formation des Pseudouridines

1. Les modifications post-transcriptionnelles

Les quatre bases standard ainsi que les riboses faisant partie d'un ARN peuvent subir de nombreux type de modifications après transcription. Actuellement, 107 modifications différentes sont recensées (pour revue, Motorin et Grosjean, 1998 ; Rozenski *et al.*, 1999). La structure et la nomenclature de l'ensemble des nucléotides modifiés sont disponibles sur la page web : <http://library.med.utah.edu/RNAmods/>. Les modifications se produisent au niveau de positions précises dans les trois domaines du vivant (Fig. III.1A). Ainsi, à ce jour, les Eucaryotes comportent 65 modifications différentes répertoriées contre 52 pour les Bactéries et 46 pour les Archaea (Fig. III.1B). Les ARNt sont les plus modifiés des ARNnc puisque leurs séquences peuvent compter jusqu'à un quart de résidus modifiés. Les ARNt comportent également le plus grand nombre de modifications différentes, soit 91 contre 31 pour les ARNr, 11 pour les snRNA et 13 pour les ARNm (Rozenski *et al.*, 1999). Notons que moins de 1% des résidus des ARNm seraient modifiés. Cependant, les analyses des ARNm comme celles d'ARNnc autres que ceux évoqués ci-avant sont encore incomplètes. Les modifications résultent soit de réactions enzymatiques simples ne faisant intervenir qu'une seule enzyme, pour la méthylation (de bases ou de riboses), la déamination, l'isomérisation, la thiolation et la réduction de base lors de la formation de résidus 5-méthylcytidine (m5C), 2'-O-méthylés (Nm), d'inosine (I), de pseudouridine (Ψ), de 4-thiouridine (s4U) et de dihydrouridine (D), soit de réactions complexes impliquant une cascade d'activités enzymatiques, c'est notamment le cas de la formation de la wybutosine (yW) (Fig. III.2) (pour revue, Garcia et Goodenough-Lashua, 1998). Par contre, la formation des résidus queunine (Q) et archaeosine (gG) est particulière parce qu'elle implique, au préalable, la biosynthèse d'hétérocycles modifiés incorporés par la suite dans l'ARN par un mécanisme d'échange de bases.

A



B

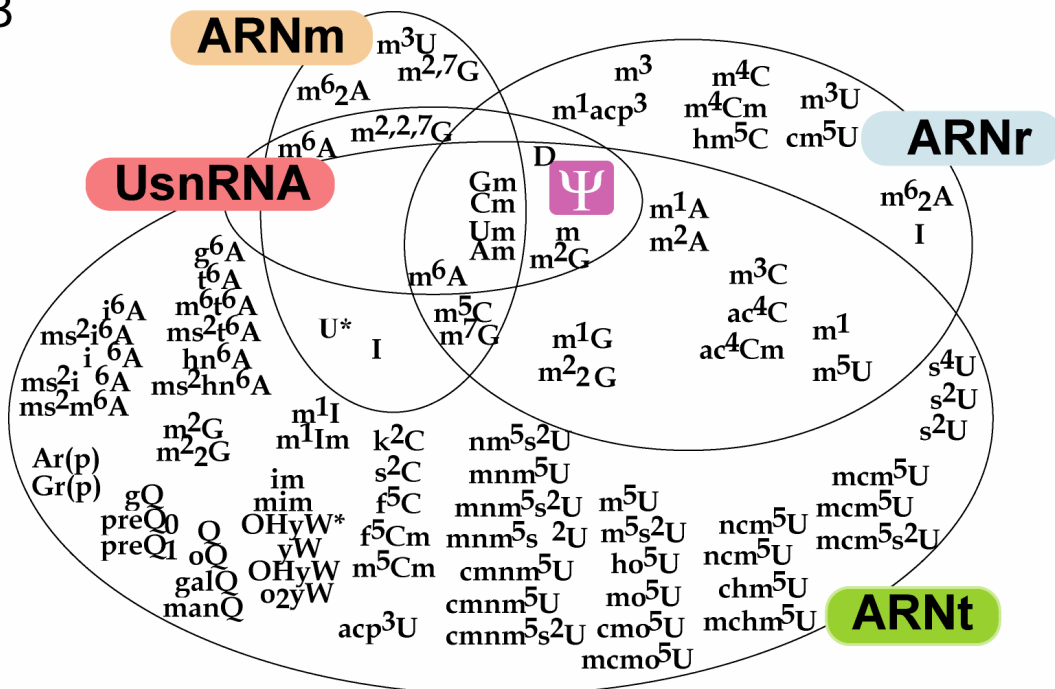


Figure III.1 : Distribution des nucléotides modifiés dans les trois domaines du vivant (A) et par type d'ARN (B) (d'après, Motorin et Grosjean, 1998).

U* correspond à l'uridine formée par déamination d'une cytosine par un mécanisme d'édition post-transcriptionnel.

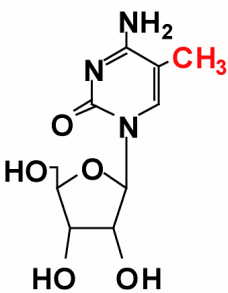
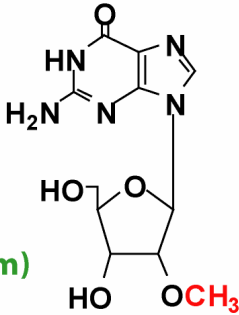
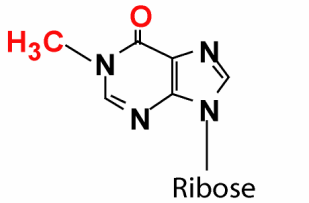
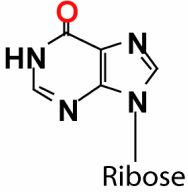
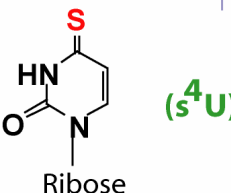
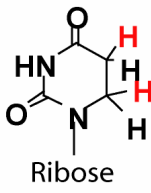
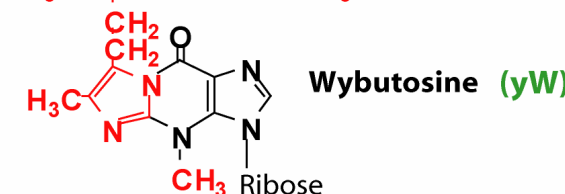
Types de réactions	Nucléotides modifiés
Méthylations	 (m⁵C) 5-méthylcytidine (méthylation de la base)  (Gm) 2'-O-méthylguanosine (méthylation du ribose)  1-méthylinosine (m¹I)
Désamination	 Inosine (I)
Thiolation	 4-thiouridine (s⁴U)
Isomérisation	 Pseudouridine (Ψ)
Réduction	 (D) Dihydrouridine
Réaction complexe	 Wybutosine (yW)

Figure III.2 : Structure chimique de quelques nucléotides modifiés (d'après Limbach *et al.*, 1994). Les altérations chimiques des nucléotides standard sont représentées en rouge et le symbole correspondant à la nomenclature officielle est indiqué en vert entre parenthèses.

La pseudouridylation ou conversion de résidus U en pseudouridine (5 β -D-ribofuranosyluracile ou Ψ) (Fig. III.3) et la 2'-O-méthylation des riboses (ou Nm), sont les des deux modifications les plus fréquentes (Cohn, 1951; Davis *et al.*, 1957). Les résidus Ψ sont présents dans les trois domaines du vivant (Fig. III.1). La pseudouridine est formée par isomérisation d'un résidu uridine, ce qui implique la rupture de la liaison N-glycosidique entre la base et le ribose, suivie d'une rotation de la base de 180° sur elle-même, aboutissant à la formation d'une liaison entre le C1' du ribose et le C5 du cycle pyrimidique (Fig. III.3) (Johnson et Soll, 1970 ; Ciampi *et al.*, 1977; Kammen *et al.*, 1988). La pseudouridylation ne requiert ni énergie ni co-facteur (Arena *et al.*, 1978 ; Green *et al.*, 1982 ; Kammen *et al.*, 1988 ; Samuelsson et Olsson, 1990 ; Nurse *et al.*, 1995 ; Wrzesinski *et al.*, 1995). Une pseudouridine présente une liaison carbone-carbone entre la base et le ribose à la place de la liaison N-glycosidique ainsi qu'un groupement imino (NH) supplémentaire qui autorise une liaison hydrogène supplémentaire via une molécule d'eau avec la chaîne phosphate. Par conséquent, une pseudouridine comporte des propriétés différentes de celles d'un résidu U. Bien que la liaison C-C confère une plus grande flexibilité par rapport à la liaison N-glycosidique, les études réalisées en RMN, cristallographie en RX et simulation en dynamique moléculaire, montrent une augmentation de la rigidité du squelette phospho-ribose. En effet, la liaison hydrogène supplémentaire restreint la mobilité et les conformations de bases présentes en amont du résidu Ψ (pour revue, Charette et Gray, 2000). De plus, la liaison hydrogène supplémentaire favorise une conformation 3'endo du ribose qui restreint une conformation *anti* de la base et donc favorise l'empilement des bases dans l'environnement proche du brin portant le résidu Ψ (Davis, 1998). Les résidus Ψ contribueraient ainsi à stabiliser une structure ARN.

L'isomérisation des résidus U en Ψ est catalysée par une classe d'enzymes particulières, les ARN:pseudouridine-synthases (ou ARN: Ψ -synthases). On peut distinguer deux systèmes de pseudouridylation différents selon que la reconnaissance du résidu U est effectuée par l'ARN: Ψ -synthase elle-même ou qu'elle repose sur un appariement ARN-ARN entre l'ARN cible et un snoRNA à boîtes H/ACA précédemment évoqué.

La proportion importante, dans un génome, de gènes consacrés aux modifications des ARN et en particulier à la pseudouridylation qui nous intéresse ici (Björk, 1995), et la conservation de certaines des modifications tant au niveau de l'identité qu'au niveau de la position des nucléotides modifiés au cours de l'évolution suggèrent un rôle essentiel de ces modifications dans le fonctionnement cellulaire. Un défaut de pseudouridylation dans les ARNr participe au syndrome de Dyskératose Congénitale comme évoqué précédemment. Cependant, en dépit de nombreuses études, l'effet exact des pseudouridylations comme des autres modifications est loin d'être clair (Lane, 1998).

Ce chapitre est principalement consacré aux pseudouridylations. Son objectif est de présenter, dans un premier temps la localisation et la distribution des résidus Ψ , principalement dans les ARNt, snRNA et ARNr, puis les mécanismes d'action des ARN: Ψ -synthases impliquées. Nous examinerons précisément la reconnaissance du résidu U ciblé en distinguant les deux systèmes qui

reposent ou non sur des snoRNA à boîtes H/ACA. Enfin, nous terminerons par les rôles prouvés ou suggérés des résidus Ψ détectés dans les différents ARN.

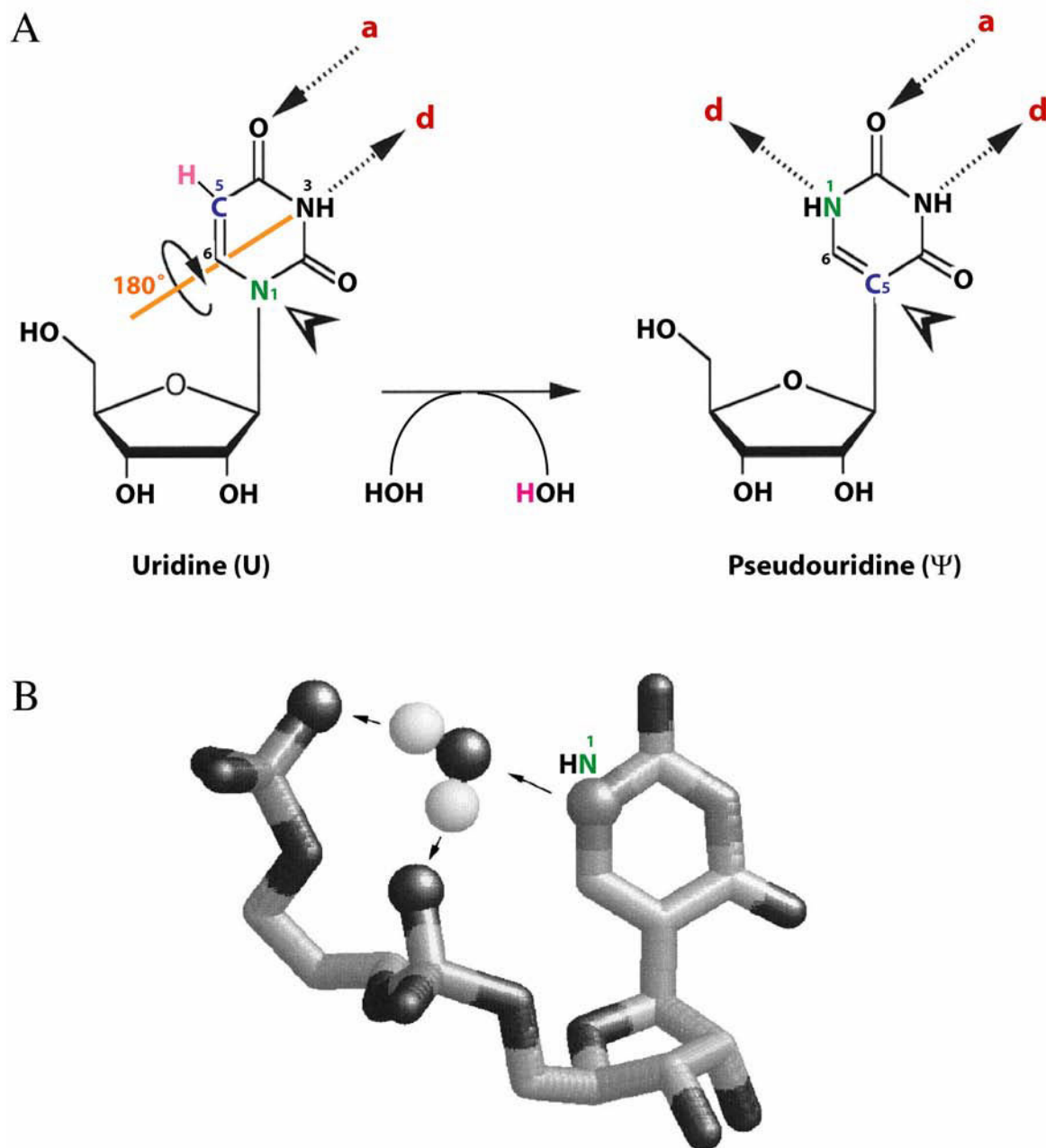


Figure III.3 : Conversion par isomérisation d'un résidu uridine en pseudouridine et modélisation de la liaison hydrogène qui peut se former entre la position N_1H de la base et la chaîne phosphate du brin qui comporte ce résidu (d'après Charette et Gray, 2000).

A. Les liaisons hydrogène possibles dans le sens accepteur (a) ou donneur (d) sont indiquées par des flèches. L'axe (N_3-C_6) autour duquel est effectuée la rotation de 180° est présenté en orange. Les différences entre résidus U (1- β -D-ribofuranosyluracil) et Ψ (5- β -D-ribofuranosyluracil) sont indiquées en couleur dans la configuration *anti* du groupe glycosyl. **B.** Formation d'une liaison hydrogène supplémentaire, via une molécule d'eau, entre la position N_1 de la base du résidu Ψ et la chaîne phosphate du résidu Ψ . La base du résidu en 5' n'est pas représentée. La structure correspond à celle du résidu Ψ_{39} des ARN^{t^{Phe}} de levure.

2. Distribution et Localisation des pseudouridines dans les ARN

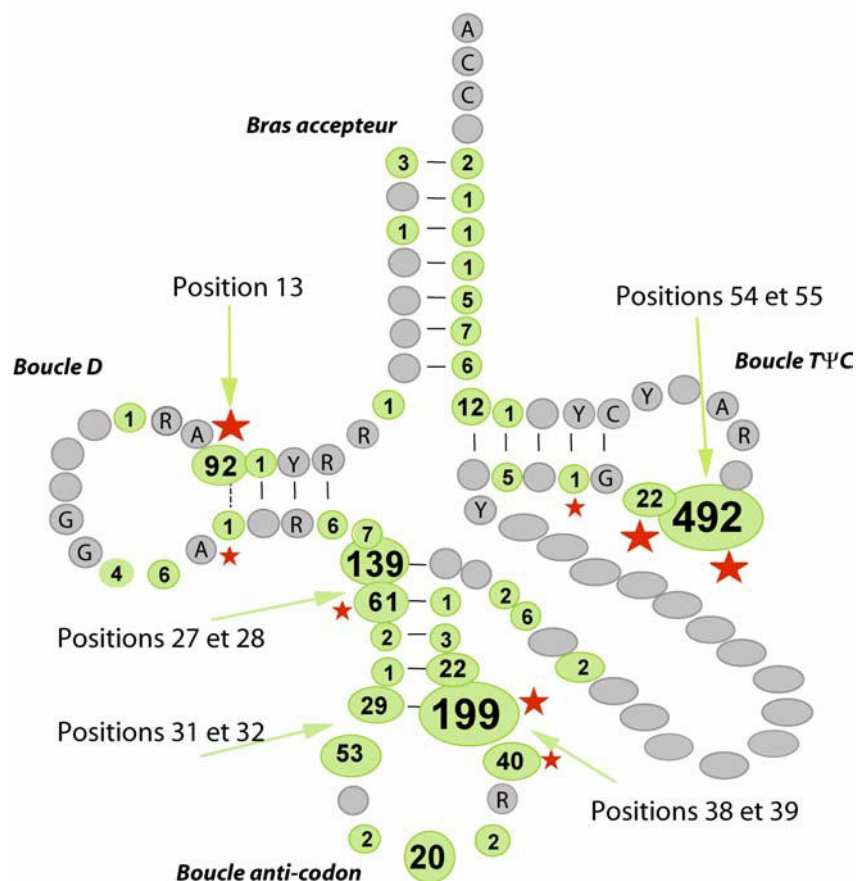
Les pseudouridines sont principalement présentes dans les ARNt, les snRNA et les ARNr. Il faut toutefois remarquer que des résidus Ψ ont également été détectés dans les ARNtm bactériens et certains snoRNA tels que les snoRNA à boîtes C/D U3 et U8 impliqués dans des événements de clivage ainsi que dans le snoRNA snR4, guide de méthylation chez la levure (Wise *et al.*, 1983; Reddy, 1988 ; Massenet *et al.*, 1998) ou le snoRNA à boîtes H/ACA snR8, guide de pseudouridylation chez la levure (Wise *et al.*, 1983).

A. ARN de transfert

Les ARNt des trois domaines du vivant présentent tous au moins un résidu Ψ . La modification du résidu U55 en Ψ 55 est presque universelle. D'autres nucléotides modifiés tels que la ribothymidine (T) (position 54) ou les résidus dihydrouridines (D) sont aussi présents dans la majorité des ARNt, ce qui justifie l'appellation de boucles T Ψ C et D de ces ARN. Des pseudouridines sont très fréquentes aux positions 13, 27, 28 et 39, et moins fréquentes aux positions 31, 32, 35, 38 et 54 (Fig. III.4) (Sprinzl *et al.*, 1998). Les Archaea, pour lesquels 59 ARNt provenant principalement d'*H. volcanii* (41 séquences) ont été analysés, présentent principalement des résidus Ψ aux positions 13, 39, 54 et 55 (la position 54 étant principalement une 1-méthylpseudouridine) mais également, dans des cas exceptionnels, aux positions 22, 28, 38 et 52 (Grosjean *et al.*, 1995).

Figure III.4 : Localisation et fréquences des résidus Ψ aux différentes positions des ARNt sur la base de l'ensemble des ARNt étudiés dans les trois domaines du vivant (d'après la banque de données de modifications dans les tRNA ; Rozenski *et al.*, 1999).

Ces données regroupent les informations concernant 571 séquences d'ARNt dont 59 seulement proviennent d'Archaea. Le nombre d'ARNt contenant un résidu Ψ est indiqué pour chaque position. Les positions pseudouridylées chez les Archaea sont indiquées par une étoile rouge dont la taille est proportionnelle au nombre de résidus Ψ détectés. Les nucléotides conservés dans plus de 90% des cas sont indiqués (R = purine, Y = pyrimidine).



B. snRNA

Chez les eucaryotes, des résidus Ψ ont été détectés dans les snRNA U1, U2, U4, U5 et U6 des spliceosomes majeurs (pour revue, Massenet *et al.*, 1998) ainsi que dans les snRNA U12, U4atac et U6atac des spliceosomes mineurs (Massenet et Branlant, 1999).

Les snRNA contiennent également des 2'-*O*-méthylations et quelques méthylations de bases (notamment en position 6 des adénosines : m6A). Ces modifications ont été étudiées dans un certain nombre d'espèces (pour revue, Massenet *et al.*, 1998) et une fois encore, la proportion de modification des snRNA est plus élevée chez les organismes plus complexes i.e. vertébrés, plantes et insectes que chez les levures.

La position de certains des résidus Ψ est souvent conservée entre organismes même éloignés sur le plan phylogénétique. Ces résidus conservés sont presque tous regroupés dans des régions fonctionnellement importantes qui participent aux interactions ARN-ARN ou ARN-protéines impliquées dans l'assemblage ou le fonctionnement des spliceosomes. Par exemple, plusieurs résidus Ψ sont détectés dans les régions d'interactions ARN-ARN entre snRNA U2 et U6 mais aussi dans les régions du snRNA U2 en interaction avec l'intron (Fig. II.15). Ainsi, une paire Ψ -A adjacente au résidu A de la boîte de branchement qui est responsable de la première attaque nucléophile de la réaction d'épissage est conservée au cours de l'évolution. Cette pseudouridine pourrait stabiliser l'interaction entre le U2 et l'intron et favoriser une conformation du résidu A favorisant l'attaque nucléophile de l'extrémité 3' de l'exon 1 par le groupement hydroxyl en 2' du résidu A (Yang *et al.*, 2005).

C. ARN ribosomiques

Les ARNr comportent une proportion de nucléotides modifiés comprise entre 1 et 3%. Ces modifications sont essentiellement de deux sortes : les méthylations en 2'-*O*- des riboses et les résidus Ψ . Le nombre de résidus Ψ présents dans les ARNr de la petite et de la grande sous-unité du ribosome varie fortement entre les bactéries (10 pour *E. coli*) et les eucaryotes (près d'une centaine pour l'Homme). Le nombre de résidus Ψ dans les ARNr mitochondriaux et chloroplastiques est légèrement inférieur à celui observé chez les bactéries. Par contre, la situation chez les Archaea semble être intermédiaire (Tableau III.1). En effet, notre analyse partielle des ARNr de *P. abyssi* a révélé la présence d'au moins 15 résidus Ψ . Le nombre de résidus Ψ présent chez les Archaea pourrait donc être supérieur au nombre observé pour les Bactéries mais nettement inférieur à celui des Eucaryotes. Notons que les ARNr 5S comportent également des résidus Ψ (Tableau III.1) (Szymanski *et al.*, 2000).

Il faut noter également que les différentes méthodes de détection des résidus Ψ n'aboutissent pas toutes à la localisation précise de leur position. Ainsi les méthodes de spectrométrie de masse, à moins qu'elles soient couplées à des étapes de digestion par des RNases, ne permettent de définir que la proportion de résidus Ψ dans la molécule entière. Au contraire, la méthode que nous avons utilisée, mise au point par l'équipe de J. Offengand et qui sera présentée par la suite, permet de

localiser précisément les résidus Ψ dans les ARNr. Par contre, son emploi pour un ARNr entier est laborieux (Tableau III.1, voir pourcentages entre parenthèses).

Organisme	Nombre de mole de Ψ /mole d'ARNr			
	Petite sous-unité	Grande sous-unité	Analyse	Références
ARNr cytoplasmique				
<i>E. coli</i>	1	9	Seq. (100)	Bakin et al., 1994
<i>B. subtilis</i>	1	5	Seq. (40)	Niu et Ofengand, 1999
<i>H. halobium</i>	0	4	Seq. (70)	Bakin et Ofengand, 1995
<i>S. solfataricus</i>	5	4	Total	Noon et al., 1998
<i>S. acidocaldarius</i>	nd	6	Seq. (?)	Massenet et al., 1999
<i>P. abyssi</i>	3	12	Seq. (25)	Nos résultats
<i>A. fulgidus</i>	1	5	Seq. (?)	Tang et al., 2002
<i>H. marismortui</i>	nd	3	Seq. et Total	Kirpekar et al., 2005
<i>S. cerevisiae</i>	14	30	Seq. (99)	Bakin et al., 1994
<i>D. melanogaster</i>	nd	57	Seq. (59)	Ofengand et Bakin, 1997
<i>X. laevis</i>	44	~52	Total	Maden, 1990
<i>M. musculus</i>	~36	~57	Seq. (~50)	Ofengand et Bakin, 1997
<i>H. sapiens</i>	~36	~57	Seq. (~50)	Ofengand et Bakin, 1997
ARNr mitochondrial				
<i>S. cerevisiae</i>	nd	1	Seq. (99)	Bakin et al., 1994
<i>M. musculus</i>	nd	1	Seq. (17)	Ofengand et Bakin, 1997
<i>H. sapiens</i>	nd	1	Seq. (41)	Ofengand et Bakin, 1997
<i>T. brucei</i>	nd	6	Seq. (98)	Ofengand et Bakin, 1997
ARNr chloroplastique				
<i>Z. mays</i>	nd	4	Seq. (29)	Ofengand et Bakin, 1997
ARNr 5.8S				
<i>S. cerevisiae</i>	1		Seq. (100)	Rubin, 1973
<i>M. musculus</i>	2		Seq. (100)	Nazar et al., 1976 ;
<i>H. sapiens</i>	2		Seq. (100)	Khan et al., 1977

Tableau III.1 : Nombre de résidus Ψ présents dans les ARNr des petites et grandes sous-unités des ribosomes cytoplasmiques, mitochondriaux et chloroplastiques ainsi que dans les ARNr 5S de différents organismes provenant des trois domaines du vivant (d'après Ofengand, 2002).

Les organismes appartenant au domaine Archaea sont présentés sur fond vert. L'analyse des ARNr a été faite soit par séquençage de l'ARN (Seq.) et détermination précise de la position (voir technique basée sur l'utilisation de CMCT et de la transcriptase inverse développée dans la partie Matériels & Méthodes), soit par l'étude de la composition totale en nucléotides modifiés reposant sur la spectrométrie de masse (Total). Le nombre indiqué entre parenthèse à côté de « Seq. » traduit le pourcentage de la séquence des ARNr analysée. « nd » signifie non déterminé.

L'idéal est donc l'utilisation conjointe des deux méthodes. Cette démarche a permis, chez *H. marismortui* de corriger le nombre de résidus Ψ initialement prédits, il est passé de 19 à 3 résidus Ψ (Ofengand, 2002 ; Del Campo *et al.*, 2005; Kirpekar *et al.*, 2005).

La superposition de la structure tridimensionnelle d'un ribosome avec la cartographie des résidus Ψ présents dans les ARNr révèle une concentration élevée dans les régions fonctionnellement importantes de l'ARNr de la grande sous-unité (Fig. III.5 et 6) (pour revues, Charette et Gray, 2000 ; Decatur et Fournier, 2002 ; Ofengand, 2002). Ces régions sont majoritairement impliquées dans les interactions avec les molécules d'ARNt au niveau des sites A, P et dans une moindre mesure au site E, à proximité du tunnel de sortie du peptide en cours de synthèse, au niveau du centre peptidyl-transférase directement impliqué dans la catalyse ou au niveau des sites de fixation des facteurs d'élongation. Le grand ARNr contient des résidus Ψ très conservés dans l'évolution (Branlant *et al.*, 1981) (Fig. III.5 et 6). Ainsi, au sein de l'ARN de la grande sous-unité des trois domaines du vivant, les modifications sont principalement regroupées dans les domaines II, IV et V (Fig. III.7) (Branlant *et al.*, 1981 ; pour revue, Decatur et Fournier, 2002). Les domaines IV et V comportent la majorité des résidus Ψ . Le domaine V est situé au niveau du centre peptidyl-transférase et participe à la catalyse. Le domaine IV est le centre de décodage en interaction avec à la fois l'ARNm et la tige-boucle anti-codon de l'ARNt. Il est à noter que l'hélice 69 du domaine IV présente les résidus Ψ les mieux conservés au cours de l'évolution (Fig. III.7). Or cette hélice, de même qu'une région du domaine II, est précisément à l'interface entre les deux sous-unités (Merryman *et al.*, 1999a; Merryman *et al.*, 1999b). Les domaines II et IV, bien que distants du domaine V sur le plan de la structure secondaire, sont en fait très proches du site de formation de la liaison peptidique. En fait, la localisation des positions des résidus Ψ dans les différentes espèces suggère un besoin de conserver des résidus Ψ dans des régions précises mais pas nécessairement à une position précise (Fig. III.7). La conservation de certains résidus Ψ dans des régions de haute importance suggère qu'ils ont un rôle important.

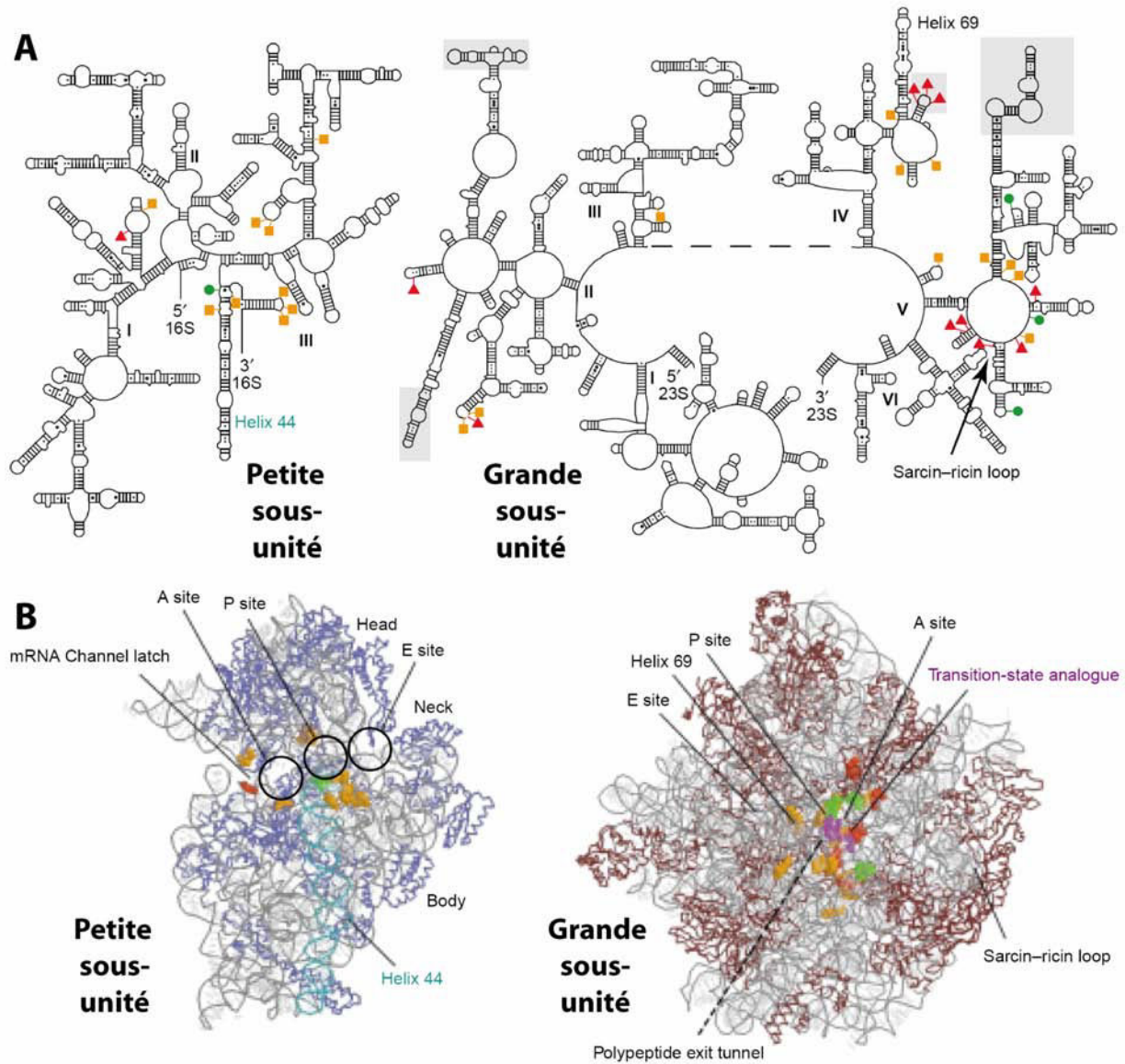


Figure III.5 : Distribution des nucléotides modifiés dans les ARNr d'*E. coli* (d'après Decatur et Fournier, 2002).

A. Structures secondaires des ARNr 16S et 23S. Les résidus Ψ sont représentés par des triangles rouges, les 2'-*O*-méthylations (Nm) par des cercles verts et les méthylations de bases (mN) par des carrés orange.

B. Structures tridimensionnelles des petite et grande sous-unités du ribosome d'*E. coli* modélisées par analogie avec, respectivement, les structures 3D disponibles de *T. thermophilus* (PDB : 1fjf) et de *H. marismortui* (PDB : 1ffk et 1ffz). Les nucléotides modifiés sont représentés selon le même code couleur qu'en (A). Les chaînes protéiques sont respectivement indiquées en bleu et en magenta pour la petite et la grande sous-unité du ribosome. L'analogue de l'état de transition est montré en violet.

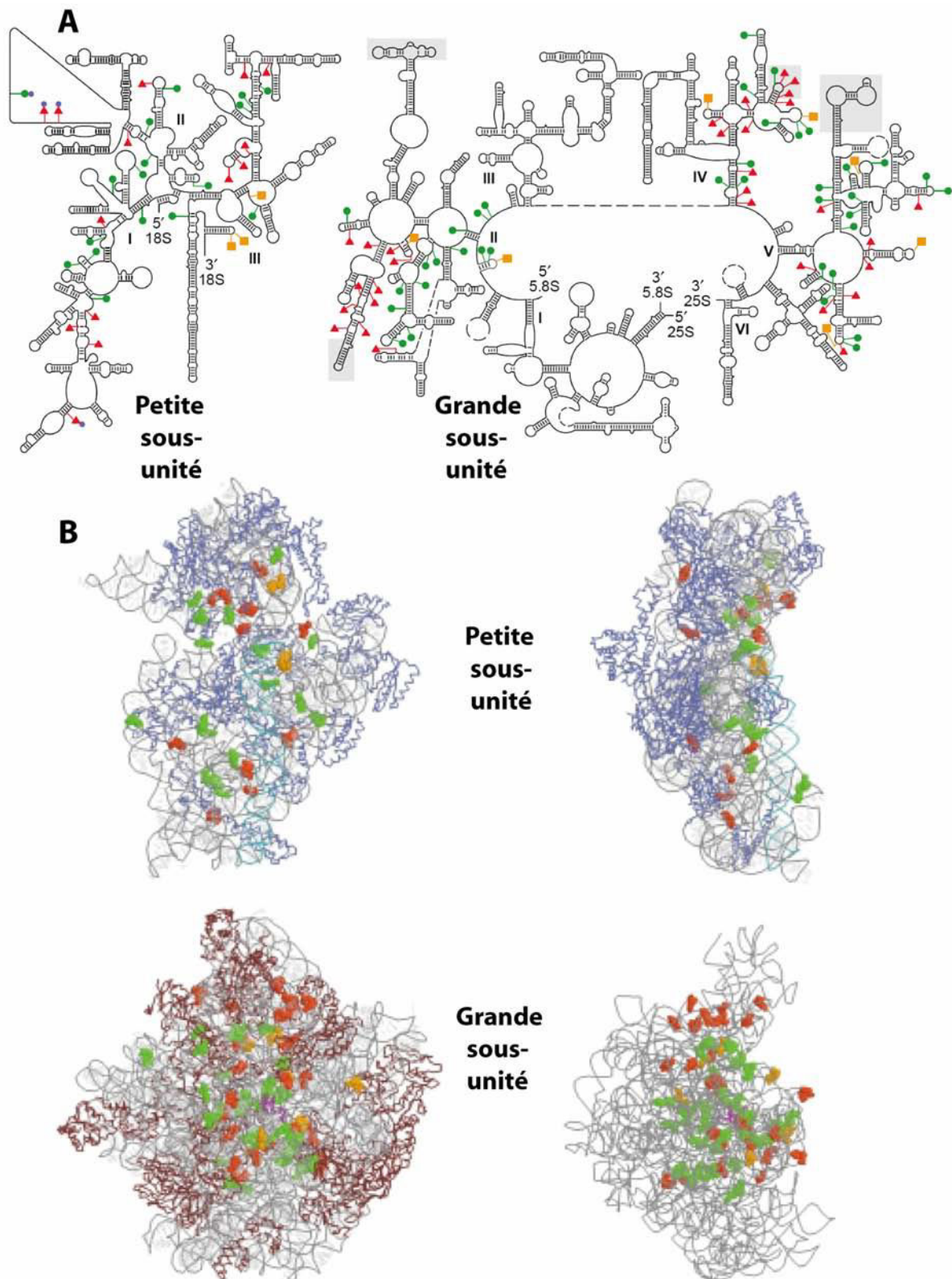


Figure III.6 : Distribution des nucléotides modifiés dans les ARNr de *S. cerevisiae* (d'après Decatur et Fournier, 2002).

Voir légende de la Figure 5. Les panneaux **B.** correspondent à deux vues différentes de la petite sous-unité ainsi qu'à deux vues de la grande sous-unité.

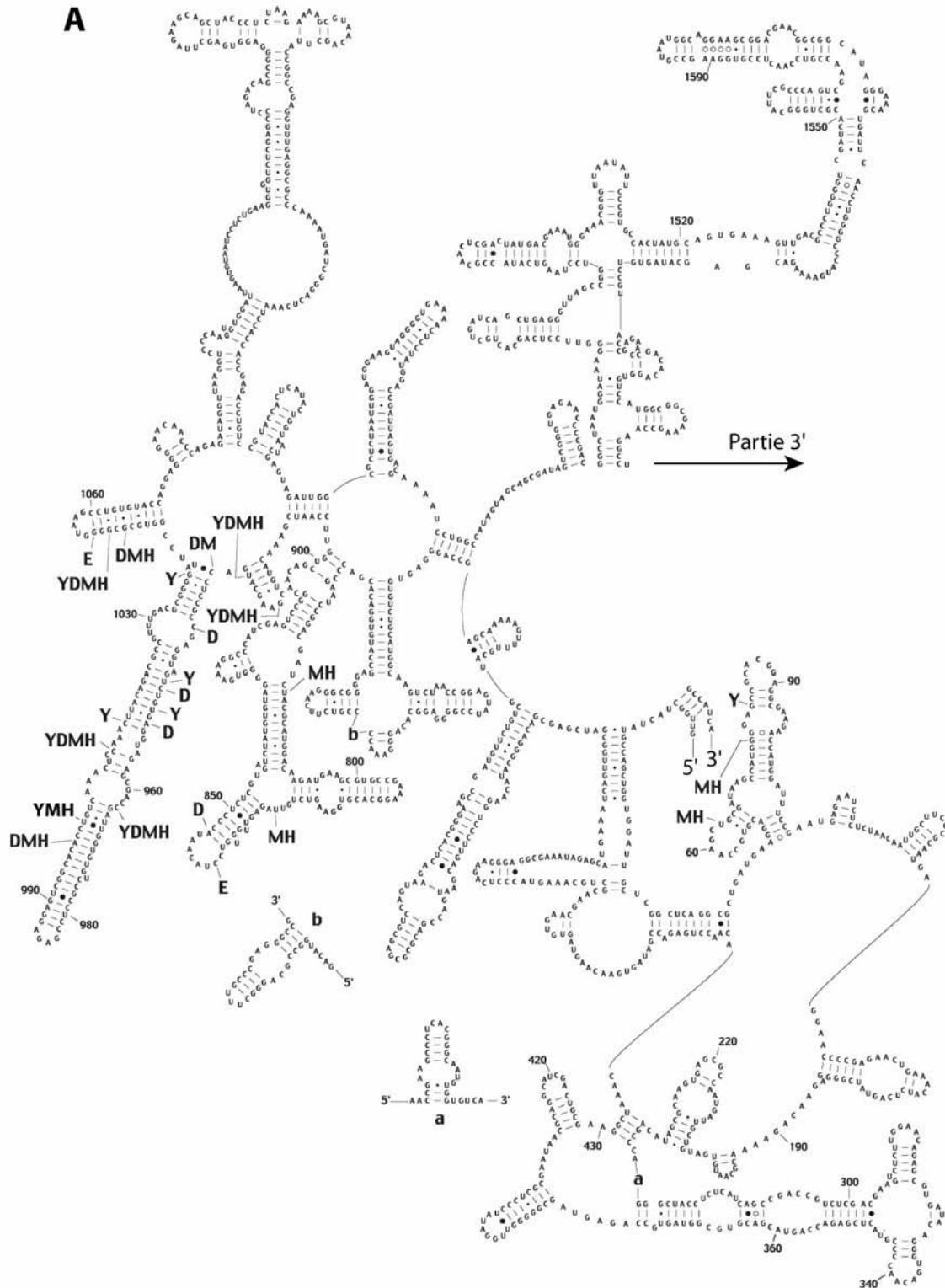


Figure III.7 : Localisation des résidus Ψ dans le grand ARN de la grande sous-unité ribosomique chez les diverses espèces où ils ont été étudiés (d'après Ofengand, 2002; corrigé par Del Campo *et al.*, 2005 ; Kirpekar *et al.*, 2005). La structure secondaire et la numérotation de *H. marismortui* sont utilisées (voir site web de Gutell). Les positions des résidus Ψ chez *H. marismortui* (Ψ), *E. coli* (E), *B. subtilis* (B), *Z. mays* (Z), *H. halobium* (A), *S. acidocaldarius* (S), *S. cerevisiae* (Y), *D. melanogaster* (D), *M. musculus* (M) et *H. sapiens* (H) ont été obtenues à partir des références indiquées en Tableau III.1. La séquence est divisée en région 5' (A) et région 3' (B).

3. Catalyse de la pseudouridylation

A. Caractérisation des ARN:Ψ-synthases

Les ARN:Ψ-synthases connues à ce jour ont été caractérisées par 3 approches distinctes : soit par voie biochimique, soit par homologie de séquence avec les ARN:Ψ-synthases préalablement caractérisées, soit par criblage de produits de gènes par des techniques à haut débit (pour revue, Ansmant et Motorin, 2001).

Les premières ARN:Ψ-synthases ont été identifiées par voie biochimique en testant, après fractionnement chromatographique, les activités ARN:Ψ-synthases grâce à une technique qui consiste à utiliser des ARN ayant incorporé des uraciles marqués au tritium (^3H) en position 5 (Cortese *et al.*, 1974). En effet, la réaction d'isomérisation entraîne l'élimination de tritium à cette position qui peut être dosé (Fig. III.3). C'est ainsi que l'enzyme TruA, a été caractérisée chez *Salmonella typhimurium* (Arena *et al.*, 1978) et chez *E. coli* (Kammen *et al.*, 1988). D'autres ARN:Ψ synthases d'*E. coli* ont ensuite été identifiées par la même approche (Tableau III.2). De nombreuses ARN:Ψ-synthases d'*E. coli* et d'autres organismes ont ensuite été détectées par homologie des séquences codées par les ORF avec les ORF précédentes. Deux approches expérimentales sont utilisées pour prouver la fonction d'une ARN:Ψ-synthase. D'une part, lorsque le gène est non essentiel, il est possible de le déléter et de repérer l'absence de résidus Ψ aux positions attendues. La modification de ces positions peut ensuite être rétablie en complétant la souche délétée par apport du gène par un plasmide. D'autre part, dans le cas de gènes essentiels, la région codante du gène n'est pas délétée mais placée sous contrôle d'un promoteur régulable. Dans ce dernier cas, les substrats potentiels sont analysés après extinction de l'expression du gène pendant un temps supérieur à la demi-vie des ARN étudiés.

La seconde approche consiste à cloner le gène, suspecté par le fait que sa séquence présente des similarités avec des ARN:Ψ-synthases connues, dans un vecteur d'expression qui permet la surproduction de la protéine chez *E. coli*. Elle doit ensuite être purifiée. Les tests d'activité consistent à modifier des substrats ARN produits par transcription *in vitro*. Ces différentes approches sont complémentaires et aboutissent à la caractérisation précise d'une ARN:Ψ-synthase y compris l'identification de ses substrats (pour revue, Ansmant et Motorin, 2001).

Une approche plus récente, basée sur le criblage de produits de l'expression de gènes à haut débit a permis l'identification de l'ARN:Ψ-synthase Pus7p de *S. cerevisiae* qui a la particularité de ne pas présenter d'homologie de séquences forte avec les autres ARN:Ψ-synthases connues (Ma *et al.*, 2003). Cette équipe cherchait à caractériser les protéines responsables de la pseudouridylation en position 35 et 42 du snRNA U2. Pour cela, une banque d'expression des ORF de *S. cerevisiae* qui avaient été individuellement fusionnées à la GST a été utilisée. La stratégie consistait à tester l'activité sur un snRNA U2 produit *in vitro* de différents ensembles de protéines de fusion purifiées. Le snRNA U2 utilisé comme substrat contenait un résidu U35 marqué au ^{32}P en 5'. Par recoupement entre plusieurs ensembles portant l'activité, et test de l'activité de sous-ensembles de clones, les auteurs ont caractérisé le gène Yor243 comme codant l'ARN:Ψ-synthase responsable de

	Organisme	Mode d'identification	Spécificité de substrat	Références
Famille TruA				
TruA (HisT, PSU-I)	<i>E. coli</i>	Biochimique	Ψ_{38} , Ψ_{39} , Ψ_{40} ARNt	Kammen <i>et al.</i> , 1988
PSU-I	<i>S. typhimurium</i>	Biochimique	Ψ_{38} , Ψ_{39} , Ψ_{40} ARNt	Arena <i>et al.</i> , 1978
PSU-I	<i>B. taurus</i>	Biochimique	Ψ_{38} , Ψ_{39} , Ψ_{40} ARNt	Green <i>et al.</i> , 1982
PusI	<i>S. cerevisiae</i>	Homologie de séquence TruA	Ψ_{26-28} , Ψ_{34-36} , Ψ_{65} , Ψ_{67} ARNt Ψ_{44} snRNA U2	Simos <i>et al.</i> , 1996/Motorin <i>et al.</i> , 1998 Massenet <i>et al.</i> , 1999
mPusI	<i>M. musculus</i>	Homologie de séquence PusI	Ψ_{27} , $\Psi_{34, 36}$, (Ψ_{32} ?) ARNt Ψ_1 ARNt, Ψ_{44} snRNA U2	Chen et Patton, 1999 Behm-Ansmant <i>et al.</i> , 2006
PusI	<i>C. elegans</i>	Homologie de séquence PusI	Ψ_{27} ARNt	Patton et Padgett, 2003
Pus2	<i>S. cerevisiae</i>	Homologie de séquence TruA	Ψ_{27} , Ψ_{28} ARNt mitochondriaux	Behm-Ansmant <i>et al.</i> , sous presse
Pus3 (Deg1)	<i>S. cerevisiae</i>	Homologie de séquence TruA	Ψ_{38} , Ψ_{39} ARNt cytoplasmiques et mitochondriaux	Lecointe <i>et al.</i> , 1998
Famille TruB				
TruB	<i>E. coli</i>	Biochimique	Ψ_{55} ARNt	Nurse <i>et al.</i> , 1995
Pus4 (Ynl292)	<i>S. cerevisiae</i>	Homologie de séquence TruB	Ψ_{55} ARNt cytoplasmiques et mitochondriaux	Becker <i>et al.</i> , 1997a
Cbf5	<i>S. cerevisiae</i>	Homologie de séquence TruB	ARNr cytoplasmiques	Lafontaine <i>et al.</i> , 1998a
Famille RluA				
RluA	<i>E. coli</i>	Biochimique	Ψ_{746} ARNr 23S + Ψ_{32} ARNt	Wrzesinski <i>et al.</i> , 1995a
RluC (YccC)	<i>E. coli</i>	Homologie de séquence TruA	Ψ_{955} , Ψ_{2504} , Ψ_{2580} ARNr 23S	Conrad <i>et al.</i> , 1998
RluD (YfiI, SfhB)	<i>E. coli</i>	Biochimique	Ψ_{1911} , Ψ_{1915} , Ψ_{1917} ARNr 23S	Raychaudhuri <i>et al.</i> , 1998
TruC (YqcB)	<i>E. coli</i>	Homologie de séquence RsuA	Ψ_{65} ARNt	Del Campo <i>et al.</i> , 2001
Pus5 (Ylr165)	<i>S. cerevisiae</i>	Homologie de séquence RluA	Ψ_{2819} (Ψ_{2580} <i>E. coli</i>) ARNr 21S mitochondrial	Ansmant <i>et al.</i> , 2000
Pus6 (Ygr169)	<i>S. cerevisiae</i>	Homologie de séquence RluA	Ψ_{31} ARNt cytoplasmiques et mitochondriaux	Ansmant <i>et al.</i> , 2001
Pus8 (Rib2)	<i>S. cerevisiae</i>	Homologie de séquence RluA	Ψ_{32} ARNt cytoplasmiques	Behm-Ansmant <i>et al.</i> , 2004
Pus9 (Ydl036)	<i>S. cerevisiae</i>	Homologie de séquence RluA	Ψ_{32} ARNt mitochondriaux	Behm-Ansmant <i>et al.</i> , 2004
Famille RsuA				
RsuA	<i>E. coli</i>	Biochimique	Ψ_{516} ARNr 16S	Wrzesinski <i>et al.</i> , 1995b
RluB (YciL)	<i>E. coli</i>	Homologie de séquence RluB B. subtilis	Ψ_{2605} ARNr 23S	Del Campo <i>et al.</i> , 2001
RluB	<i>B. subtilis</i>	Homologie de séquence RsuA	Ψ_{2633} (Ψ_{2605} <i>E. coli</i>) ARNr 23S	Niu et Ofengand, 1999
RluE (YmfC)	<i>E. coli</i>	Homologie de séquence RsuA	Ψ_{2457} ARNr 23S	Conrad <i>et al.</i> , 1999 Del Campo <i>et al.</i> , 2001
RluF (YjbC)	<i>E. coli</i>	Homologie de séquence RsuA	Ψ_{2604} ARNr23S	Del Campo <i>et al.</i> , 2001
Famille TruD				
TruD (YgbO)	<i>E. coli</i>	Biochimique	Ψ_{13} ARNt	Kaya et Ofengand, 2003
Pus7 (Yor243)	<i>S. cerevisiae</i>	Banque ORF fusionnées GST	Ψ_{35} du snRNA U2 Ψ_{13} , Ψ_{35} ARNt	Ma <i>et al.</i> , 2003 Behm-Ansmant <i>et al.</i> , 2003
Famille Pus10				
Pus10 / PsuX	<i>P. furiosus</i>	Homologie de séquence	Ψ_{55} ARNt	Watanabe and Gray 2000 Roovers <i>et al.</i> , 2006

Tableau III.2 : ARN: Ψ -synthases caractérisées à ce jour et approches utilisées pour leur caractérisation. Le code couleur utilisé pour le nom des enzymes correspond au mode d'identification : biochimique en violet, par homologie de séquence en vert et par utilisation d'une banque d'ORF fusionnées à la GST en orange.

la modification de la position 35 du snRNA U2. La démonstration définitive de l'activité de la protéine Yor243p a été obtenue par production et purification de cette protéine chez *E. coli* et test *in vitro* de son activité sur un transcript du snRNA U2. Une dernière démonstration de son activité en position 35 du snRNA U2 a été obtenue par comparaison de l'activité sur un snRNA U2 produit *in vitro*, d'un extrait de levure sauvage et d'un extrait de levure correspondant à une souche dans laquelle le gène YOR243 avait été délété. Parmi les trois résidus Ψ présents dans le snRNA U2 seul le résidu Ψ 35 est formé par l'enzyme Pus7p (Ma *et al.*, 2003). Mais notre équipe a démontré qu'en plus d'agir sur le snRNA U2, Pus7p agit aussi à la position 13 de certains ARNt cytoplasmiques et à la position 35 du pré-ARNt^{Tyr} (Tableau III.2) (Behm-Ansmant *et al.*, 2003).

Au final, les ARN: Ψ -synthases d'*E. coli* ont été regroupées, sur la base de leurs similarités de séquence, en 5 familles : les familles TruA, TruB, RluA, RsuA et TruD (Tableau III.2) (pour revue, (Hamma et Ferre-D'Amare, 2006). Une sixième famille pourrait être représentée par les enzymes Pus10/PsuX identifiées chez certains eucaryotes et archaea mais qui seraient absentes chez les bactéries (Watanabe et Gray, 2000 ; Roovers *et al.*, 2006). La fonction encore incertaine de cette probable sixième famille sera discutée par la suite.

Des équivalents de ces 6 familles d'ARN: Ψ -synthases ont été recherchés par similarités de séquence chez les Archaea (Tableau III.3) (Watanabe et Gray, 2000). Il semble que seuls les euryotes présentent un membre de la famille TruA. Les deux royaumes comportent un représentant pour la famille TruB qui a été annoté comme étant aCBF5 chez les Archaea et CBF5 chez les eucaryotes. Les familles TruD et Pus10 sont également représentées chez les Archaea. Cependant, aucun représentant des familles RsuA et RluA n'y a été observé.

	Euryotes	Crénotes
Famille TruA	aTruA	-
Famille TruB	aCBF5	aCBF5
Famille TruD	aPus7/TruD	aPus7/TruD
Famille Pus10	aPus10/PsuX	aPus10/PsuX
Famille RluA	-	-
Famille RsuA	-	-

Tableau III.3 : Liste des ARN: Ψ -synthases identifiées par similarités de séquences chez les Archaea (d'après Watanabe et Gray, 2000).

B. Mécanisme enzymatique

L'identification de signatures propres aux ARN: Ψ -synthases a permis de détecter, par similarités de séquence, plus de 150 gènes codant des ARN: Ψ -synthases putatives, mais aussi de mieux comprendre leur mécanisme d'action. Cette signature est matérialisée par des motifs conservés entre familles ou au sein d'une même famille. Ainsi, les familles RluA, RsuA, TruB, TruA et TruD comportent 3 motifs conservés communs (motifs I, II et III), et des motifs conservés supplémentaires, les motifs IIa et IIIa ont également été observés (Fig. III.8) (Del Campo *et al.*,

2004). Ces motifs IIa et IIIa ont été détectés plus tardivement en aval des motifs II et III, et correspondraient respectivement aux enchaînements « +xYx### » et « Lxxxx# » où + correspond à un acide aminé basique, x, à n'importe quel résidu et #, à un résidu hydrophobe. Bien que peu d'information soit disponible sur la famille Pus10, il semble qu'elle présente au moins le motif II. Le motif II commun aux différentes familles d'ARN:Ψ-synthases comporte deux acides aminés très conservés, il s'agit d'un acide aspartique strictement conservé et d'une glycine. Notons que ce motif est également présent chez les déoxycytidine triphosphate déaminases (Dcd) et deoxyuridine triphosphatases (Dut) qui lient également spécifiquement des résidus U (Koonin, 1996). Par conséquent, le motif II des ARN:Ψ-synthases a été proposé comme étant le site de fixation de l'uridine et il appartiendrait donc au site catalytique (Koonin, 1996). En effet, l'implication de l'acide aspartique conservé du motif II dans la catalyse a été démontré pour l'enzyme TruA (Huang *et al.*, 1998a) puis pour les enzymes RluA (Raychaudhuri *et al.*, 1998 ; Ramamurthy *et al.*, 1999), RsuA (Conrad *et al.*, 1999), TruB (Ramamurthy *et al.*, 1999) et RluC d'*E. coli* (Conrad *et al.*, 1999).

Deux mécanismes catalytiques différents impliquant tous deux l'acide aspartique conservé ont été proposés (Fig. III.9) (Huang *et al.*, 1998a) :

- le premier mécanisme catalytique repose sur l'attaque nucléophile de la position C6 de l'uridine par le carboxylate de l'acide aspartique conservé de l'enzyme (Fig. III.9A). Une attaque nucléophile de ce type a déjà été proposée pour une méthyltransférase responsable de la méthylation de la base U en position 54 des ARNt (RUMT) mais dans ce cas elle est réalisée par une cystéine (pour revue, Garcia et Goodenough-Lashua, 1998).

- le deuxième mécanisme catalytique possible (Fig. III 9B) repose sur l'attaque nucléophile du carbone C1' du ribose de l'uridine par le carboxylate de l'acide aspartique catalytique. Cette réaction a été suggérée sur la base du mécanisme enzymatique proposé pour la première étape de la formation de la queue à partir de la guanine par l'ARNt guanine transglycosylase (pour revue, (Romier *et al.*, 1998)

Bien qu'il soit difficile de distinguer lequel de ces mécanismes est utilisé, de nombreux arguments sont en faveur du premier. L'utilisation d'un tRNA uniformément marqué par incorporation de 5-fluorouracile (5FU) a montré pour TruA d'*E. coli* que le complexe formé entre TruA et l'ARNt marqué est un complexe covalent qui, après dénaturation de TruA, libère le 5FU complexé à TruA et que la séparation du complexe TruA-5FU-tRNA en ses composants ARN et protéine donne lieu à la libération de 6-hydroxy-5,6-dihydro-fluorouridine (Gu *et al.*, 1999). Ces résultats ne peuvent s'expliquer qu'avec le premier mécanisme reposant sur l'attaque nucléophile en position C6 de la base et non sur le ribose (Gu *et al.*, 1999 ; Hamilton *et al.*, 2006). Il faut ajouter à cela qu'aucun échange de base qui pourrait être envisagé avec le second mécanisme n'a été observé. Enfin, nous avons l'habitude d'identifier la présence de résidus Ψ après chromatographie sur couche mince et nous observons quelque fois la présence de taches aberrantes qui pourraient correspondre à des groupes phosphoribosyl, résultats de la digestion d'un intermédiaire réactionnel qui peuvent uniquement être obtenus avec le premier mécanisme.

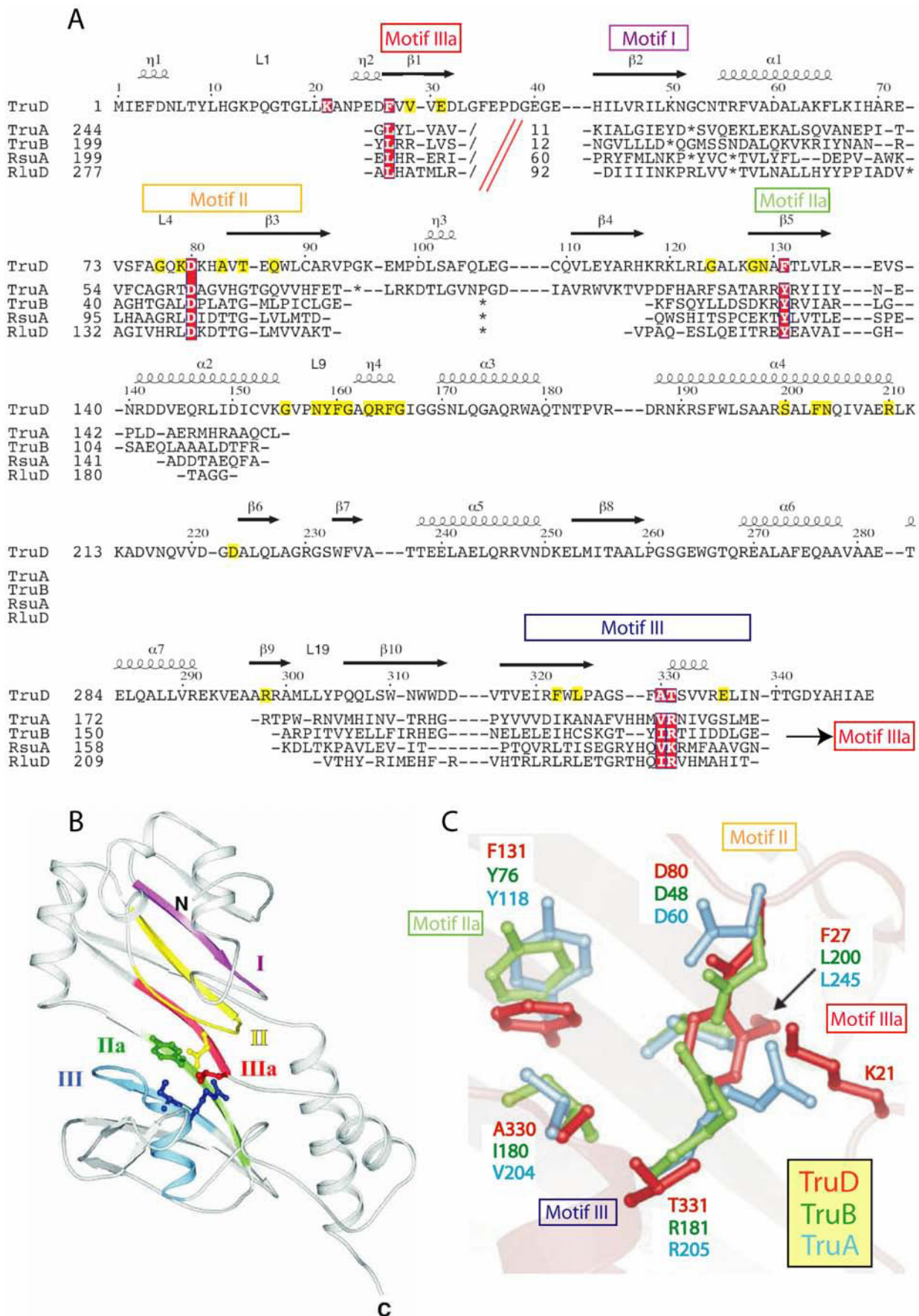


Figure III.8 : Les acides aminés conservés dans les 5 familles d'ARN : Ψ -synthases TruA, TruB, TruD, RluA et RluD (adapté de Kaya *et al.*, 2004 ; Del Campo *et al.*, 2004). **A.** L'alignement de la séquence en acides aminés a été réalisé selon les données de structure 3D à l'aide du programme DALI (Holm et Sander,

1993), en utilisant TruD comme référence. L'alignement multiple des séquences est réalisé à partir de ClustalX (Jeanmougin *et al.*, 1998). Les résidus présentant plus de 80% de similarité de séquence sont représentés en jaune. Les astérisques impliquent la présence de plus de deux acides aminés non alignés. Le motif IIIa présent en N-ter chez TruD est localisé en C-ter dans les autres familles. Les 5 résidus appartenant à 4 des 5 motifs conservés entre les 5 familles et jouant un rôle dans le site catalytique sont encadrés en rouge. Ces 5 résidus, D du motif II, Y/F du motif IIa, I/V/A et R/T du motif III et L/F du motif IIIa, sont replacés dans la structure 3D du domaine pseudouridine synthase commun aux 5 familles (**B**). La structure 3D présentée ici est celle de RluD (famille RluA). **C**. Superposition des chaînes latérales des 5 résidus jouant un rôle dans le site catalytique. Le code couleur pour chacune des trois familles est le suivant : en rouge TruD, en vert TruA, et en bleu TruB. L'action supplémentaire du résidu K21 de TruD pourrait se substituer à l'action du résidu R du motif III conservé chez les 4 autres familles qui est absent chez TruD.

Cependant, des résultats récents ne sont pas en cohérence avec le premier mécanisme. En effet, il a été observé que pour TruB liée à un substrat contenant du 5FU, il est possible de récupérer un produit d'isomérisation du 5FU (5-fluoro-6-hydroxypseudouridine) et qu'aucun complexe covalent entre l'acide aspartique et l'ARN n'est formé (Hoang et Ferre-D'Amare, 2001 ; Spedaliere et Mueller, 2004). Aucune conclusion claire ne peut donc être donnée sur le ou les mécanismes réellement utilisés. Il est possible que certaines familles d'ARN:Ψ-synthases utilisent le second mécanisme utilisé par des ADN-glycosylases (Labahn *et al.*, 1996).

Le rôle des autres résidus intervenant dans la catalyse a été en grande partie clarifié après étude structurale de certains représentants de ces ARN:Ψ-synthases.

C. Structure tridimensionnelle des ARN:Ψ-synthases

Les structures 3D des représentants de chacune des 6 familles ont été étudiées. La première structure 3D d'une ARN:Ψ-synthase, TruA d'*E. coli*, a été décrite en 2000 (Foster *et al.*, 2000). Les structures de TruB (Hoang et Ferre-D'Amare, 2001 ; Pan *et al.*, 2003), RsuA (Sivaraman *et al.*, 2002 ; Matte *et al.*, 2005), RluD (Del Campo *et al.*, 2004 ; Mizutani *et al.*, 2004 ; Sivaraman *et al.*, 2004), RluC (Mizutani *et al.*, 2004), RluE (Pan *et al.*, 2007), RluF (Sunita *et al.*, 2006), CBF5 (Hamma *et al.*, 2005 ; Li et Ye, 2006; Manival *et al.*, 2006 ; Rashid *et al.*, 2006), TruD (Ericsson *et al.*, 2004 ; Hoang et Ferre-D'Amare, 2004; Kaya *et al.*, 2004) et PUS10 (McCleverty *et al.*, 2007) ont été décrites par la suite. La structure 3D de la protéine RsuA a été étudiée en interaction avec un résidu uracile et avec une molécule d'UMP (Sivaraman *et al.*, 2002). Seules les structures 3D de TruB, RluA, TruA (Hur et Stroud, 2007) et aCBF5 ont été obtenues sous forme de complexes avec un ARN. Il s'agissait de segments d'ARN mimant respectivement la tige-boucle du bras TΨC, la tige-boucle anticodon, un ARNt entier et enfin un sRNA à boîtes H/ACA.

Les structures obtenues mettent en évidence des caractéristiques structurales communes aux ARN:Ψ-synthases. En effet, en dépit des faibles similarités de séquence, il est possible de distinguer un domaine structural commun aux 5 familles d'ARN:Ψ-synthases, qui renferme les 5 motifs caractéristiques (Fig. III.8) (Del Campo *et al.*, 2004). Ce repliement comporte au moins 8 brins β flanqués d'hélices et de boucles, qui sont organisés en deux feuillets, le plus souvent anti-

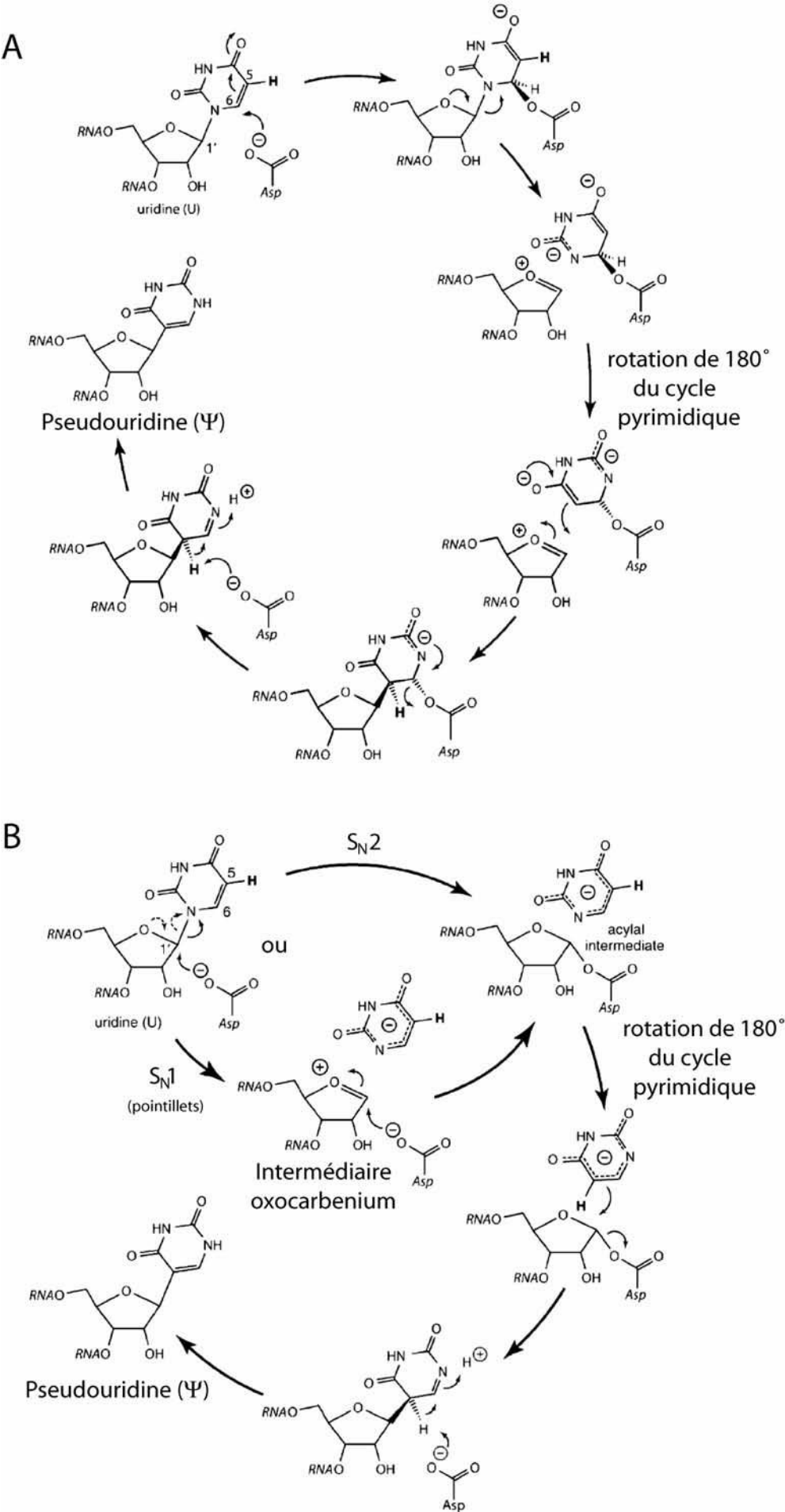


Figure III.9 :
Mécanismes proposés
pour la réaction de
pseudouridylation
(d'après Hamilton *et al.*, 2006).
Deux mécanismes
différents reposent l'un
sur l'attaque
nucléophile en position
C6 de l'uridine (**A**)
l'autre sur l'attaque
nucléophile de la
position C1' du ribose
de l'uridine (**B**).

parallèles centrés autour du site catalytique (Fig. III.8B & 10). L'aspartate catalytique est systématiquement au centre d'une crevasse dont la surface est couverte de résidus basiques. Il est orienté de telle façon que sa chaîne latérale soit dirigée vers le solvant. Des domaines présentant une variété de repliements sont observés aux extrémités N et C terminales (Fig. III.10). Parmi ces repliements particuliers, on note le domaine N-ter de la famille RsuA qui présente des homologies avec celui de la protéine ribosomale S4. La partie N-ter de la famille appelée TruD forme un domaine typique de cette famille (domaine TruD). Une partie de région C-ter des protéines de la famille RluA forme un domaine caractéristique de cette famille. Le domaine C-ter de TruB est appelé PUA car il est également présent dans d'autres enzymes de modification (PseudoUridine synthase and Archaeosine transglycosylases).

Il est probable que l'ensemble des familles actuelles d'ARN:Ψ-synthases dérivent d'un seul et même ancêtre commun (Fig. III.10C). Néanmoins, elles ont beaucoup divergé au cours de l'évolution. La conservation des acides aminés du site actif et de son voisinage est plus importante entre les enzymes des familles TruA et RsuA mais, en terme de séquence primaire totale, TruA et RsuA sont les enzymes les plus éloignées. De plus, les représentants de la famille TruA sont les seuls connus pour agir en tant que dimères. TruD présente le motif IIIa à l'extrémité N-ter contrairement aux autres ARN:Ψ-synthases qui le présentent du côté C-ter. L'enchaînement différent des motifs observés dans la famille TruD pourrait résulter d'une permutation circulaire (Hamma et Ferre-D'Amare, 2006). Les familles RluA et RsuA sont les deux familles qui partagent le plus de similarités.

Au final, la superposition des sites actifs des 5 enzymes révèle cinq acides aminés conservés : l'acide aspartique catalytique strictement conservé du motif II, la tyrosine (ou phénylalanine pour TruD) du motif IIa, et la leucine (ou phénylalanine pour TruD) du motif IIIa, et deux autres résidus dont les propriétés chimiques sont conservées : arginine ou lysine et isoleucine ou valine du motif III (Fig. III.8) (Del Campo *et al.*, 2004). Une exception est le remplacement de ces résidus par des résidus alanine et thréonine dans la famille TruD (Fig. III.8). Il semblerait que le résidu basique du motif III permette la formation d'un pont salin avec l'acide aspartique catalytique, ce qui entraînerait un abaissement du pKa et donc une accentuation du caractère nucléophile du résidu catalytique (Hoang et Ferre-D'Amare, 2001). L'absence de résidu basique à cette position dans la famille TruD serait compensée par la présence d'une lysine en position 21 (Fig. III.8) (Kaya *et al.*, 2004). La tyrosine du motif IIa pourrait jouer un rôle soit de transfert de protons soit dans le positionnement de la base de l'uridine cible (Hoang et Ferre-D'Amare, 2001 ; Phannachet *et al.*, 2005). Le Tableau III.4 résume le rôle, dans le site actif, des acides aminés conservés entre TruA, TruB, RsuA, RluD et TruD. Par ailleurs, il a également été démontré, par l'étude des enzymes TruB et RluA, que les résidus lysine et proline assez fortement conservés au sein du motif I ont une importance fonctionnelle (Spedaliere *et al.*, 2000). En fait, leurs mutations n'ont que peu d'effet sur l'efficacité de la catalyse mais provoquent une déstabilisation des protéines. Ces résidus joueraient donc plus un rôle structural que catalytique, et il a été proposé qu'ils pourraient ainsi renforcer l'activité catalytique (Hoang *et al.*, 2005).

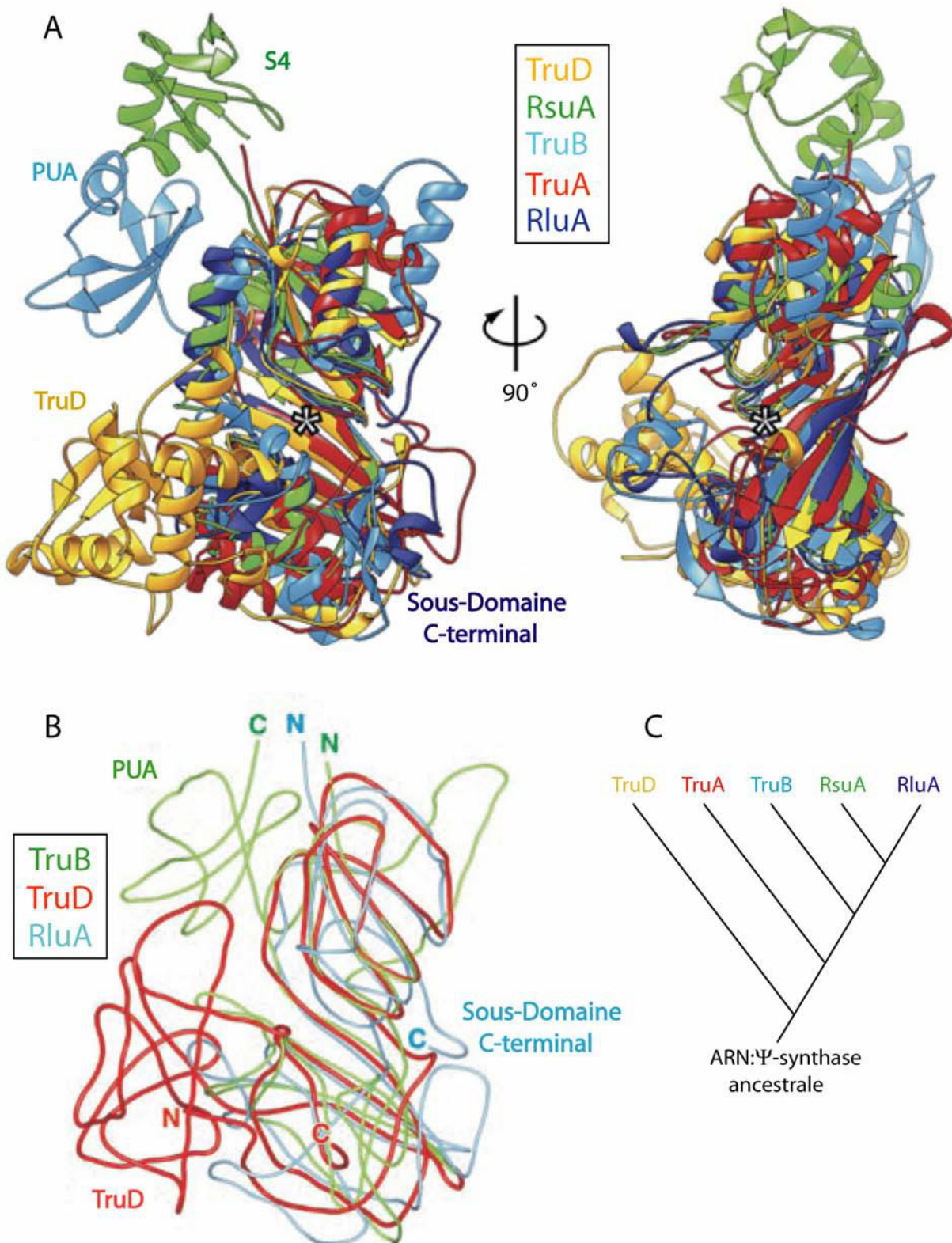


Figure III.10 : Comparaison des structures 3D de membres de chacune des 5 familles d'ARN :Ψ-synthases (d'après Hamma et Ferre-D'Amare, 2006 ; Kaya *et al.*, 2004).

A. Superposition des modèles en ruban des structures 3D de TruD (jaune), TruA (rouge), RsuA (verte), RluA (bleu foncé) et TruB (bleu clair). Le domaine central est très similaire et permet la mise en exergue des domaines périphériques TruD, PUA, S4 et sous-domaine C-terminal uniques aux familles indiquées. Le site catalytique est marqué par un astérisque. **B.** Représentation en fil de fer des chaînes peptidiques des protéines dont la structure 3D est représentée dans le panneau A, TruD (rouge), TruA (bleu clair) et TruB (vert). Les structures ont été alignées par le programme DALI. **C.** Evolution hypothétique des relations entre les 5 familles d'ARN :Ψ-synthases, basée sur la comparaison de leur séquence et de leur structure.

ARN:Ψ-synthases							
Motif	RluD	RsuA	TruA	TruD	TruB	aCBF5	Rôle suggéré par l'étude structurale
II	-	-	-	-	His43	His77	Stabilisation de l'extrusion de bases
	Arg137	Arg100	Arg58	-	-	-	Stabilisation de l'extrusion de bases
	Asp139	Asp102	Asp60	Asp80	Asp48	Asp82	Acide aspartique catalytique
	Gly144	Gly107	Gly65	Thr85	Gly53	Gly87	?
IIa	Arg169	Lys130	Arg116	Asn129	Lys74	Lys108	?
	Tyr171	Tyr132	Tyr118	Phe131	Tyr76	Tyr110	Empilement avec la base de l'uridine cible
III	His234	His184	His202	-	-		?
	-	-	-	-	Tyr 179	Tyr179	Empilement avec le ribose de l'uridine cible
	Ile236	Val186	Val204	Ala330	Ile180	Ile180	?
	Arg237	Lys187	Arg205	Thr331*	Arg181	Arg181	Formation d'un pont salin avec l'acide aspartique catalytique
IIIa	Leu279	Leu200	Leu245	Phe27	Leu200	Leu200	?
	-	-	-	Lys21*	-		Formation d'un pont salin avec l'acide aspartique catalytique

Tableau III.4 : Localisation et rôle suggéré des résidus appartenant aux cinq motifs conservés entre les cinq familles d'ARN:Ψ-synthases.

Les résidus représentés en bleu correspondent aux 5 résidus les mieux conservés. * : le résidu K21 de TruD compenserait l'absence de résidu basique à la position T331. La protéine d'archaea aCBF5 est un représentant de la famille TruB. Ses résidus sont indiqués sur fond vert selon la numérotation de l'équivalent provenant de *P. abyssi*. La protéine RluD est un représentant de la famille RluA. Les numérotations des protéines RluD, TruA, TruB, TruD et RsuA correspondent aux numérotations des équivalents d'*E. coli*.

Ainsi, les mécanismes réactionnels des membres des 5 familles s'appuieraient principalement sur 5 résidus conservés, et pourraient donc être identiques. Les subtiles divergences observées dans les 5 motifs mais aussi la présence et la nature des éléments additionnels et périphériques pourraient être à l'origine des différences de spécificité de reconnaissance de l'ARN cible.

4. Reconnaissance de la position à modifier

Les ARN:Ψ-synthases ne reconnaissent une uridine que si celle-ci est dans le contexte d'une séquence ARN. Le maintien de mécanismes de reconnaissance précise de l'uridine à l'intérieur de certains substrats est responsable de la conservation au cours de l'évolution de la localisation de certains résidus Ψ, en particulier dans les ARNr, les ARNt et les snRNA.

Deux systèmes différents sont distingués selon que la reconnaissance de la position de l'uridine repose sur une interaction ARN-Protéine ou sur un appariement ARN-ARN avec les snoRNA ou sRNA à boîtes H/ACA.

A. ARN:Ψ-synthases qui ne nécessitent pas de guide ARN

L'idée était que toutes les ARN:Ψ-synthases de ce type sont capables de reconnaître leur cible en absence de protéine accessoire. Ces enzymes assurent donc à elles seules, en plus de la catalyse, la reconnaissance d'un ou plusieurs type(s) d'ARNnc et l'accès au sein de cet ARNnc à une ou plusieurs positions précises. La reconnaissance des résidus U cibles passe souvent par un mécanisme d'extrusion de bases.

1. Mécanisme d'extrusion de bases

La résolution du co-cristal de TruB associé à une partie mimant le bras TΨC de l'ARNt a permis de mettre en évidence un mécanisme d'extrusion de bases (ou « flipping out ») (Hoang et Ferre-D'Amare, 2001 ; pour revue, Hamma et Ferre-D'Amare, 2006). La reconnaissance par TruB d'une tige-boucle TΨC induit un changement de conformation de trois nucléotides de la boucle TΨC. Les résidus U55, uridine cible, ainsi que C56 et G57 sont rejetés vers l'extérieur de l'empilement hélicoïdal auquel ils appartiennent au sein de la boucle TΨC. Ainsi, des liaisons hydrogène peuvent se former entre ces bases et des résidus de TruB (Fig. III.11&13). Cette conformation particulière de la boucle TΨC semble stabilisée par un résidu histidine (H43 chez *E. coli*) qui est conservé chez tous les représentants de cette famille. Ce résidu, placé 5 résidus en amont du résidu catalytique, interagit avec la paire reverse Hoogsteen U54-A58, (i.e. interaction en *trans* de la face W-C du U avec la face Hoogsteen du A) caractéristique des boucles TΨC et rend le résidu U55 accessible à l'aspartate catalytique (Fig. III.11). Ce mécanisme d'extrusion de base n'est pas unique à TruB. En effet, l'étude structurale du co-cristal de RluA associé à un segment d'ARN mimant la tige-boucle anticodon montre que cette association impose une nouvelle structure à la boucle anticodon (Hoang *et al.*, 2006). La fixation de RluA entraîne la formation d'une nouvelle paire de bases reverse Hoogsteen U33-A36 ainsi que l'extrusion de trois bases, U32 le site de modification, ainsi que G34 et A37 au sein de la boucle (Fig. III.11). Un résidu arginine présent deux résidus en amont du résidu catalytique et conservé chez tous les membres des familles RluA, RsuA et TruA stabiliserait cette conformation. Le rôle d'intercalant de cette arginine est donc analogue au rôle de l'histidine de TruB évoqué ci-avant. Le mécanisme d'extrusion de bases serait donc indispensable à la catalyse réalisée par chacune des familles d'ARN:Ψ-synthases

(Hoang et Ferre-D'Amare, 2001 ; Sivaraman *et al.*, 2002). Par ailleurs ce mécanisme d'extrusion est fréquemment observé pour des enzymes qui requièrent un accès à une base de l'ADN ou de l'ARN lors d'évènements de modification ou de réparation (pour revue, Cheng et Roberts, 2001). Il est à noter que la voie empruntée par TruD pour accéder au résidu U substrat est encore inconnue.

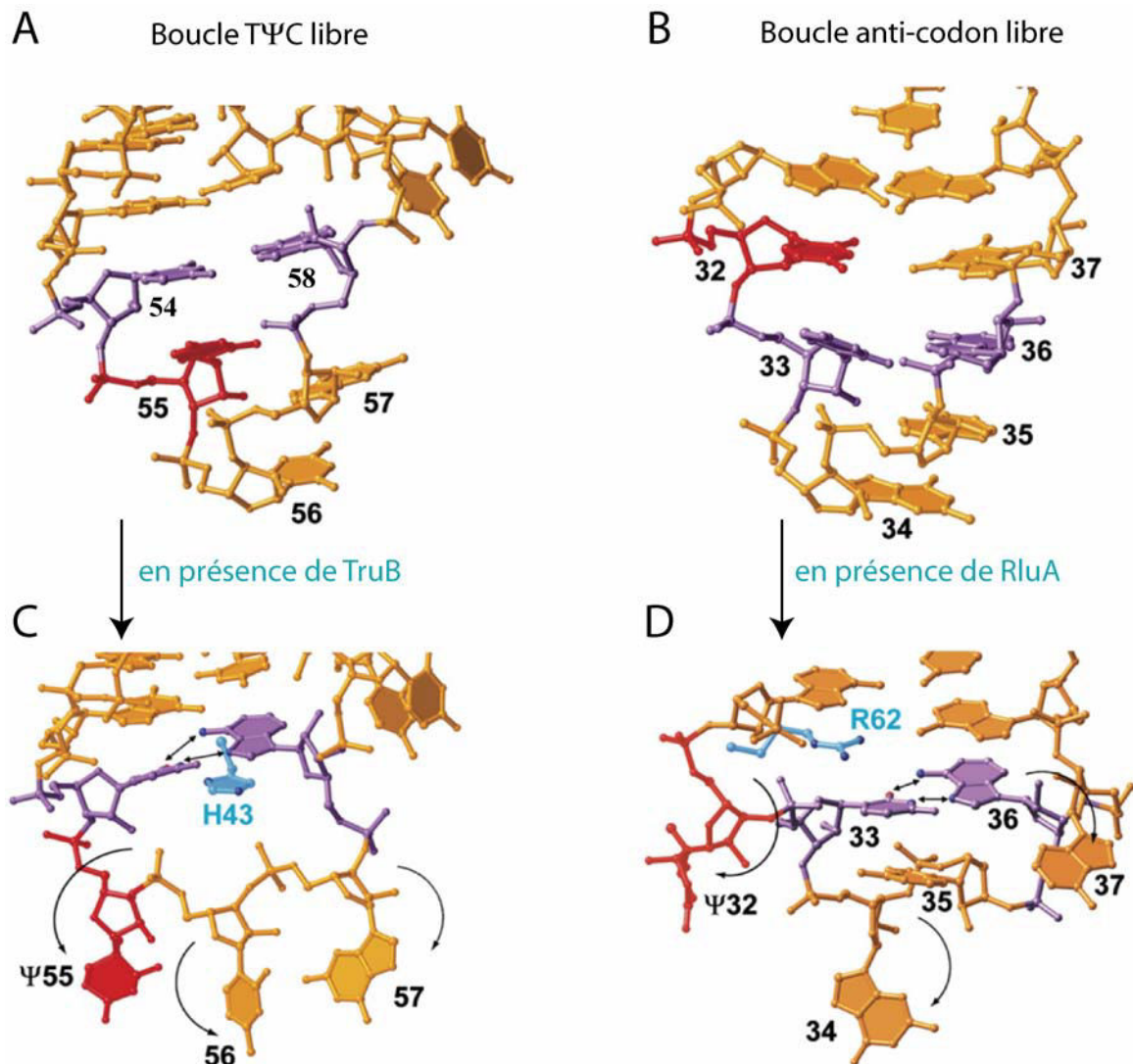


Figure III.11 : Mécanisme d'extrusion de bases mis en évidence par l'étude de la reconnaissance de l'ARN par TruB et RluA (d'après Hamma et Ferre-D'Amare, 2006).

A. et B. Structure 3D des boucles TΨC et anti-codon libre provenant de la structure 3D établie pour l'ARNt^{Phe} d'*E. coli* (Shi et Moore, 2000). Les résidus U ciblés par TruB et RluA sont représentés en rouge. **C.** Structure 3D de la boucle TΨC liée à TruB (Hoang et Ferre-D'Amare, 2001). Le résidu H43 indiqué en bleu provient de TruB. La formation de paires reverse-Hoogsteen entre les résidus U et A représentés est indiquée en violet. **D.** Structure 3D de la boucle anti-codon liée à RluA (Hoang *et al.*, 2006). Le résidu R62 indiqué en bleu provient de RluA. La formation de paires reverse-Hoogsteen entre les résidus U et A représentés est indiquée en violet.

Au final, le faible nombre d'interactions entre l'enzyme et les nucléotides avant réarrangement suggère que la reconnaissance de la cible reposerait plus sur la complémentarité structurale entre l'enzyme et la nouvelle structure induite du segment ARN ciblé et par conséquent, reposerait sur la capacité de la séquence ARN à adopter cette structure caractéristique. La possibilité de former la paire reverse-Hoogsteen U-A serait donc un pré-requis pour qu'il y ait reconnaissance de la cible. Cependant, le mode d'extrusion de bases exploité par l'enzyme TruA semble légèrement différent car il ne repose pas sur la présence d'une paire reverse-Hoogsteen (Hur et Stroud, 2007) et aucun motif spécifiquement reconnu par toutes les ARN:Ψ-synthases n'a encore été déterminé.

2. Domaines périphériques et additionnels

En dehors du site actif et du mécanisme d'extrusion de bases bien conservés entre les différentes familles, d'autres déterminants de la reconnaissance semblent spécifiques d'une famille voire d'une classe d'ARN:Ψ-synthases au sein d'une famille. L'accès à l'ARN substrat n'est pas le même pour chacune des ARN:Ψ-synthases. Il a été observé par exemple, que l'angle d'approche des deux substrats reconnus par TruB et RluA diffère d'environ 60°. Un domaine additionnel appelé « Thumb » ou « pouce » présent chez TruB subit un important changement de conformation lors de l'interaction avec le grand sillon de la molécule ARN substrat et joue donc un rôle important dans la reconnaissance de la cible (Pan *et al.*, 2003 ; Chaudhuri *et al.*, 2004 ; Phannachet et Huang, 2004). Ce domaine provient de l'association de 2 éléments dits d'insertion car absents des autres ARN:Ψ-synthases de la même famille (Ins1 et Ins2). Ainsi, l'association avec l'ARNt provoque un changement de conformation des deux éléments Ins1 et Ins2 qui n'interagissent plus entre eux mais avec la boucle TΨC (Phannachet et Huang, 2004). Par ailleurs, il a été démontré que des perturbations du site catalytique se propagent vers le domaine « pouce » (Hoang *et al.*, 2005).

La famille RluA présente une boucle qui joue le même rôle que le domaine « pouce » mais aussi un second domaine additionnel appelé domaine « forefinger » ou « index » en interaction avec le petit sillon de l'ARN substrat (Hoang *et al.*, 2006). Ensemble, les domaines « pouce » et « index » saisissent donc l'ARN substrat comme pourrait le faire une pince (Fig. III.12). Notons que ce domaine « index » bien que présent au sein des familles RsuA, RluA et TruA, est absent des familles TruB et TruD (Tableau III.5). Le domaine « pouce », quant à lui, est absent des représentants de la famille RsuA. Les membres de la famille TruD présentent un domaine « pouce » matérialisé par le domaine « TruD » qui est particulier puisque représentant 40% de la protéine, en contraste donc avec les petits domaines observés pour TruB ou RluA. Il est très probable que des résidus du domaine TruD jouent un rôle essentiel qui reste à élucider.

Certaines familles d'ARN:Ψ-synthases présentent également des domaines périphériques. Ainsi, le domaine PUA localisé en C-ter de TruB est responsable de contacts non spécifiques avec la tige acceptrice de l'ARNt (Hoang et Ferre-D'Amare, 2001). Les sous-domaines C-ter de RluB ou RluD ainsi que le domaine N-ter S4 de RsuA et RluD sont également impliqués à un degré qui n'est pas encore déterminé dans le mécanisme de reconnaissance de la cible.

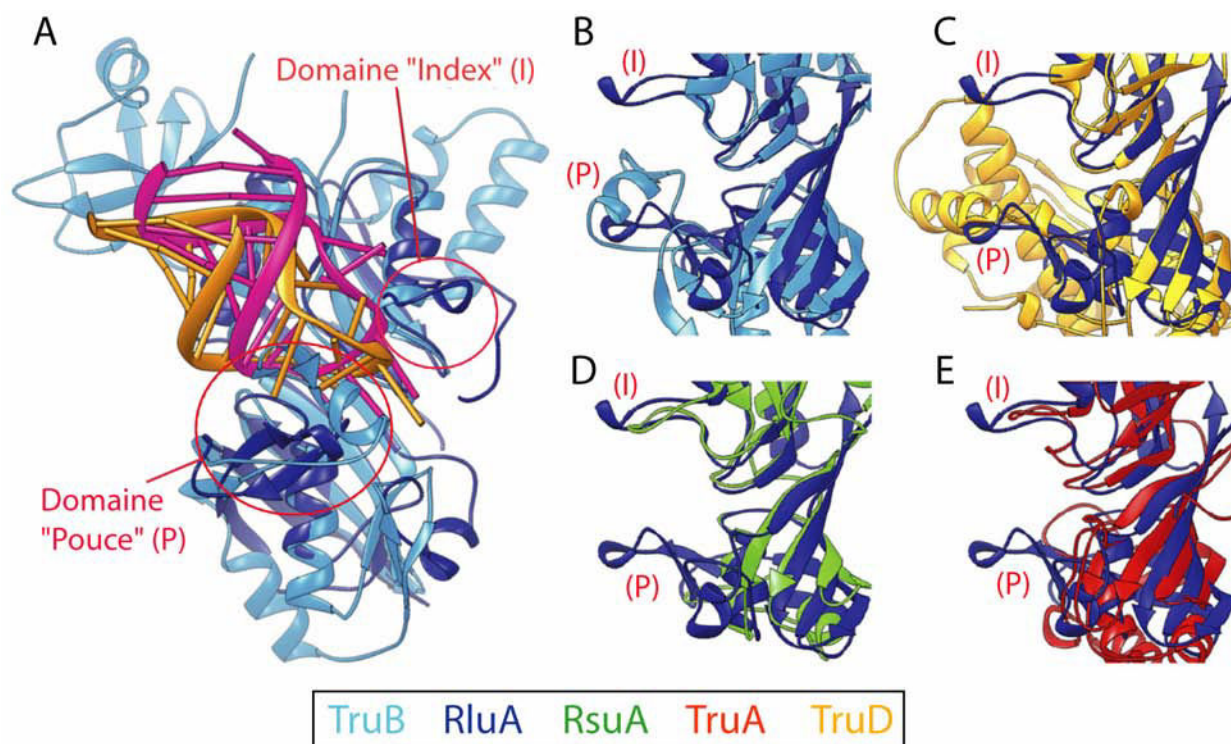


Figure III.12 : Comparaison des domaines additionnels « pouce » et « index » par superposition des structures 3D des 5 familles d'ARN:Ψ-synthases (d'après Hamma et Ferre-D'Amare, 2006).

A. Superposition des structures 3D des enzymes RluA (bleu foncé) et TruB (bleu clair) liées à leur substrat boucle anti-codon (orange) et boucle TΨC (rose). Les domaines « pouce » et « index » sont encadrés en rouge. **B.** Superposition de TruB (bleu clair) et RluA (bleu foncé). **C.** Superposition de TruD (jaune) et RluA (bleu foncé). **D.** Superposition de RsuA (vert) et RluA (bleu foncé). **E.** Superposition de TruA (rouge) et RluA (bleu foncé).

Familles	Représentants	Extrémité N-ter	Domaine "index"	Séquence consensus du site actif	Domaine "pouce"	Extrémité C-ter
TruD	TruD	-	-	XAGXKD	TruD	-
TruA	TruA	-	+	XXXRTD	+	-
TruB	TruB	-	-	HXGXLD	+	Domaine PUA
RsuA	RsuA/RluE	Domaine S4	+	XXGRLD	-	-
	RluB/RluF	Domaine S4	+	XXGRLD	-	Domaine C-ter
RluA	RluA	-	+	XXHRLD	+	-
	RluC/RluD	Domaine S4	+	XXHRLD	+	Sous-domaine C-ter

Tableau III.5 : Éléments caractéristiques des ARN:Ψ-synthases (d'après Hamma et Ferre-D'Amare, 2006).

Il semble donc que différentes stratégies associées aux combinaisons de domaines « pouce », domaines « index », domaines périphériques et substitutions dans la séquence consensus du site catalytique confèrent une remarquable spécificité d'action aux ARN:Ψ-synthases en dépit des nombreuses caractéristiques communes de leur domaine catalytique (Tableau III.5).

3. Reconnaissance de la tige-boucle TΨC par TruB

Il s'agit ici d'un exemple de reconnaissance d'un substrat ARN par une ARN:Ψ-synthase que nous réutiliserons plus loin, du fait que aCBF5 appartient à la même famille d'ARN:Ψ-synthases.

L'ARN:Ψ-synthase TruB d'*E. coli* est une enzyme dite site-spécifique de 314 acides aminés, qui modifie la position 55 de tous les ARNt à l'exception des ARNt initiateurs. Des études de spécificité des enzymes TruB d'*E. coli* et de son homologue Pus4p chez *S. cerevisiae* ont été réalisées par mutagenèse des positions des bras accepteur et TΨC (Becker *et al.*, 1997b ; Gu *et al.*, 1998). Ces études ont montré que TruB et Pus4 sont capables de modifier des fragments d'ARNt ne comportant que le bras TΨC. Ceci signifie d'une part que la structure 3D de l'ARNt n'est pas requise et d'autre part que la présence de la tige acceptrice n'est pas essentielle pour que la position U55 soit pseudouridylée. Il a été démontré également, pour TruB, que la longueur de la tige TΨC doit être au moins de 2 pb et que la paire 53-61 doit être impérativement une paire G-C (Fig. III.13E) (Gu *et al.*, 1998). Une boucle TΨC de 7 nt est optimale pour la modification mais une boucle de 6 ou 8 nt est également reconnue par TruB. Au sein de cette boucle TΨC, la nature de certains nucléotides, U54, C56, une purine en position 57 ainsi que A58 semble essentielle à la modification de la position U55. La spécificité de Pus4 semble pratiquement similaire à celle de TruB à l'exception d'une plus grande contrainte au niveau de l'hélice qui doit être d'au moins 4 pb G-C (Becker *et al.*, 1997b). Hormis les ARNt initiateurs qui comportent un résidu A à la place d'un U en position 54, les spécificités de TruB et Pus4 sont telles qu'elles peuvent reconnaître tous les ARNt de *E. coli* et de *S. cerevisiae* à part les ARNt initiateurs.

La structure de la protéine TruB complexée à un ARN synthétique mimant le bras TΨC d'un ARNt a permis de détailler les interactions entre TruB et son ARN cible (Hoang et Ferre-D'Amare, 2001). L'ARN cible utilisé a une taille de 22 nucléotides et comportait un résidu 5-fluorouridine en position équivalente à la position 55. On remarque que le domaine PUA C-terminal correspondant aux acides aminés 250 à 314 n'interagit qu'avec le squelette phospho-ribose de certains résidus de la tige acceptrice de l'ARNt substrat (Fig. III.13). La tige-boucle TΨC, plus précisément sa partie distale, est imbriquée dans la protéine TruB au niveau d'une crevasse principalement constituée de l'association du domaine « pouce » et de résidus provenant à la fois du motif II et de la région en aval du motif II (résidus 40 à 67). Le domaine « pouce » est constitué de résidus provenant des éléments Ins1 (résidus 90 à 105), Ins2 (résidus 135 à 150) et de quelques résidus en amont de Ins2 (résidus 128 à 135). Le résidu His43 interagit avec la paire reverse-Hoogsteen U54-A58 et de ce fait est, comme discuté ci-avant, impliqué dans l'extrusion des bases U55, C56 et G57. Le résidu His43, en prenant ainsi la place du résidu G18 de la boucle D, serait aussi responsable de la rupture des interactions tertiaires entre boucles D et TΨC.

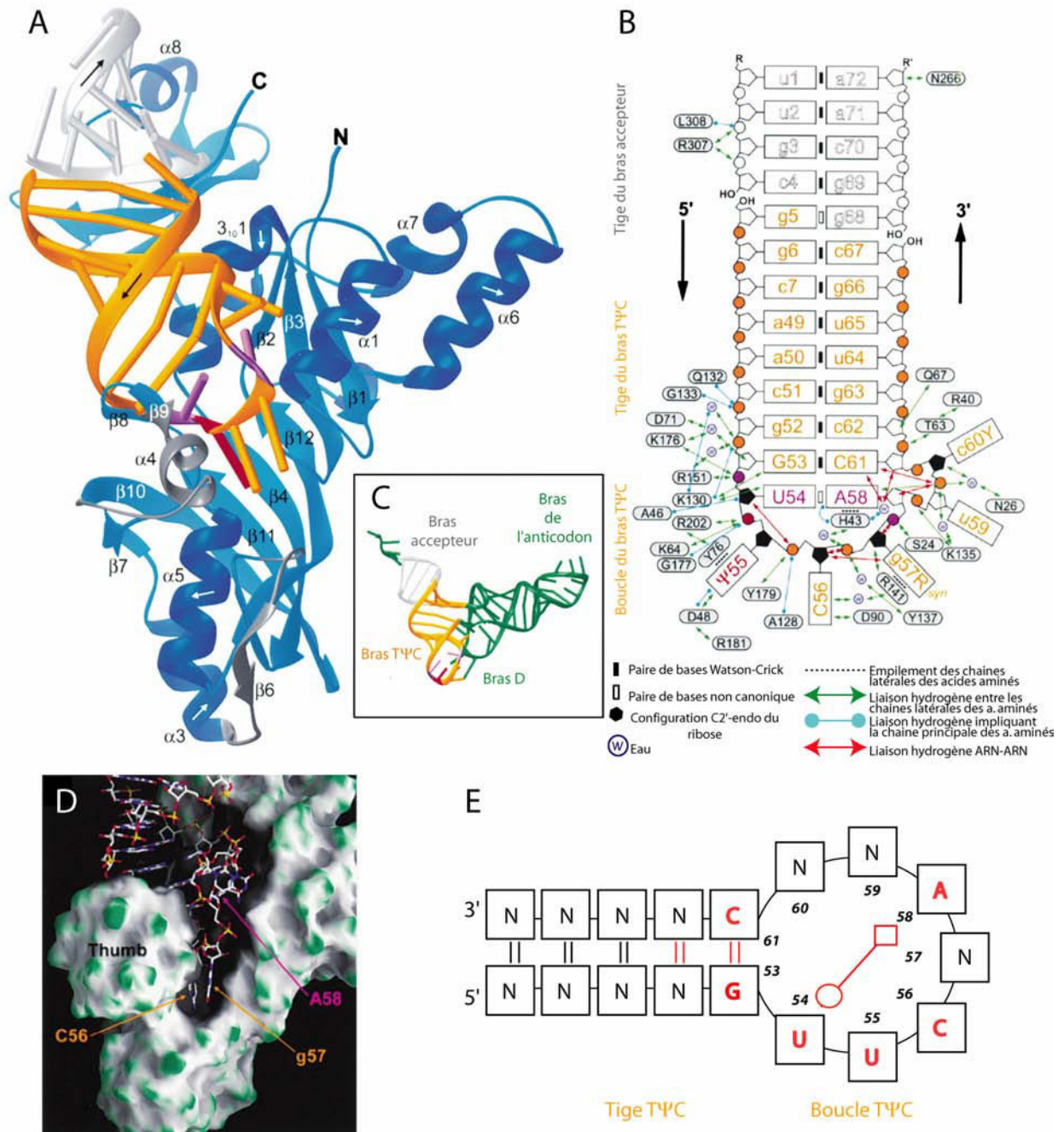


Figure III.13 : Structure de l'ARNt:Ψ55-synthase TruB d'*E. coli* complexée à un ARN synthétique mimant le bras TΨC de l'ARNt (d'après Hoang et Ferre-D'Amare, 2001).

A. Représentation de la structure 3D du complexe. Les hélices de la protéine sont représentées en bleu foncé, les brins et les boucles en bleu clair. Les deux segments protéiques (Ins1 et Ins2) caractéristiques des protéines du représentant TruB d'*E. coli* sont indiqués en gris. La tige-boucle TΨC de l'ARNt est représentée en jaune, à l'exception de l'uridine cible (U55) en rouge et des nucléotides U54 et A58 qui forment une paire de bases reverse Hoogsteen en rose. **B.** Schématisation de l'ensemble des interactions protéine-ARN et ARN-ARN. Les résidus conservés au sein de l'ARN sont écrits en majuscules et R et Y désignent respectivement la conservation d'un résidu purine ou pyrimidine. **C.** Représentation de la structure 3D de l'ARNt^{Ψ55} de *S. cerevisiae* (Shi et Moore, 2000). Les bras T et accepteur sont représentés selon le code couleur utilisé dans le panneau A, le reste de la molécule est en vert. **D.** Bases moléculaires de la fixation de la tige-boucle TΨC par l'enzyme TruB. Représentation 3D de l'interaction de TruB via son domaine « pouce » avec le grand sillon du bras TΨC. TruB accommode la tige-boucle TΨC de l'ARNt substrat dans une crevasse profonde formée notamment par le domaine « pouce ». La protéine TruB est ici représentée par sa surface moléculaire colorée selon l'incurvation (les zones vertes correspondant à des surfaces convexes et les zones grises à des surfaces concaves). **E.** Séquence

consensus reconnue par TruB obtenue grâce à l'étude de mutants des différentes positions (d'après Gu *et al.*, 1998). Les bases et les liaisons dont l'altération affecte significativement la pseudouridylation de la position 55 sont représentées en rouge.

Parmi les autres nucléotides de l'ARN, les bases U55, C56, G57 sont les seules qui sont en interaction spécifique avec, respectivement les résidus des motif II/III, Ins1 et Ins2. Les résidus du motif III jouent principalement un rôle dans l'environnement immédiat du U55. Un résidu tyrosine (Y179) uniquement conservé dans la famille TruB fait partie du motif III et jouerait un rôle dans le positionnement du ribose du résidu U55. De plus ses orbitales π riches en électrons pourraient compenser la charge positive de l'intermédiaire réactionnel oxocarbenium (Fig. III.9).

Par conséquent, les résidus de TruB responsables de la reconnaissance du substrat ARN sont principalement localisés dans le motif II et quelques nucléotides situés en C-ter de celui-ci, et dans les éléments Ins1 et Ins2. Le domaine PUA semble jouer un rôle mineur. Les nucléotides de la tige-boucle T Ψ C qui sont reconnus sont donc principalement G53, U55, C56, G57, C61 et la paire U54-A58. Ces derniers résultats sont en accord avec les données expérimentales de l'étude de spécificité de TruB et de Pus4 (Becker *et al.*, 1997b ; Gu *et al.*, 1998).

Il faut noter également que, dans un contexte naturel, TruB doit accommoder les modifications présentes aux positions 54 et 58 qui apparaissent précocement lors de la biogenèse des ARNt (Grosjean *et al.*, 1996). Par exemple, le ribose du nucléotide U54 peut être 2'-O-méthylé (Um55) et le C5 de la base peut aussi être méthylé (T54). Il existe en effet un espace libre à proximité du nucléotide 54 dans TruB qui autoriserait la présence de ces groupements méthyles (Hoang et Ferre-D'Amare, 2001).

4. Région-spécificité de TruA

Contrairement à TruB qui est site spécifique, et cible une seule position dans le bras T Ψ C, l'enzyme TruA modifie trois positions différentes : 38, 39 et 40 dans l'hélice du bras anti-codon. Elle est donc qualifiée de région spécifique. Il est à noter également que TruA modifie plusieurs ARNt (17 chez *E. coli*) comportant d'importantes divergences dans le bras anti-codon. Il était intéressant de déterminer comment TruA, avec un site actif unique, peut cibler des positions pourtant distantes de 15 Å. Le co-cristal du dimère de TruA associé à deux ARNt différents a été résolu très récemment (Hur et Stroud, 2007). C'est la première fois qu'un ARNt entier a été utilisé comme substrat pour l'étude structurale d'une ARNt: Ψ -synthase. Ceci était justifié par le fait que des expériences antérieures avaient démontré que le bras anti-codon seul n'est pas substrat de TruA (Huang *et al.*, 1998b). Trois structures différentes obtenues pour chacun des ARNt miment trois intermédiaires réactionnels différents qui ont conduit les auteurs à proposer un modèle de reconnaissance d'un ARNt par TruA dans lequel ni TruA ni l'ARNt ne subissent de changement conformationnel lors de l'association (Fig. III.14) (Hur et Stroud, 2007).

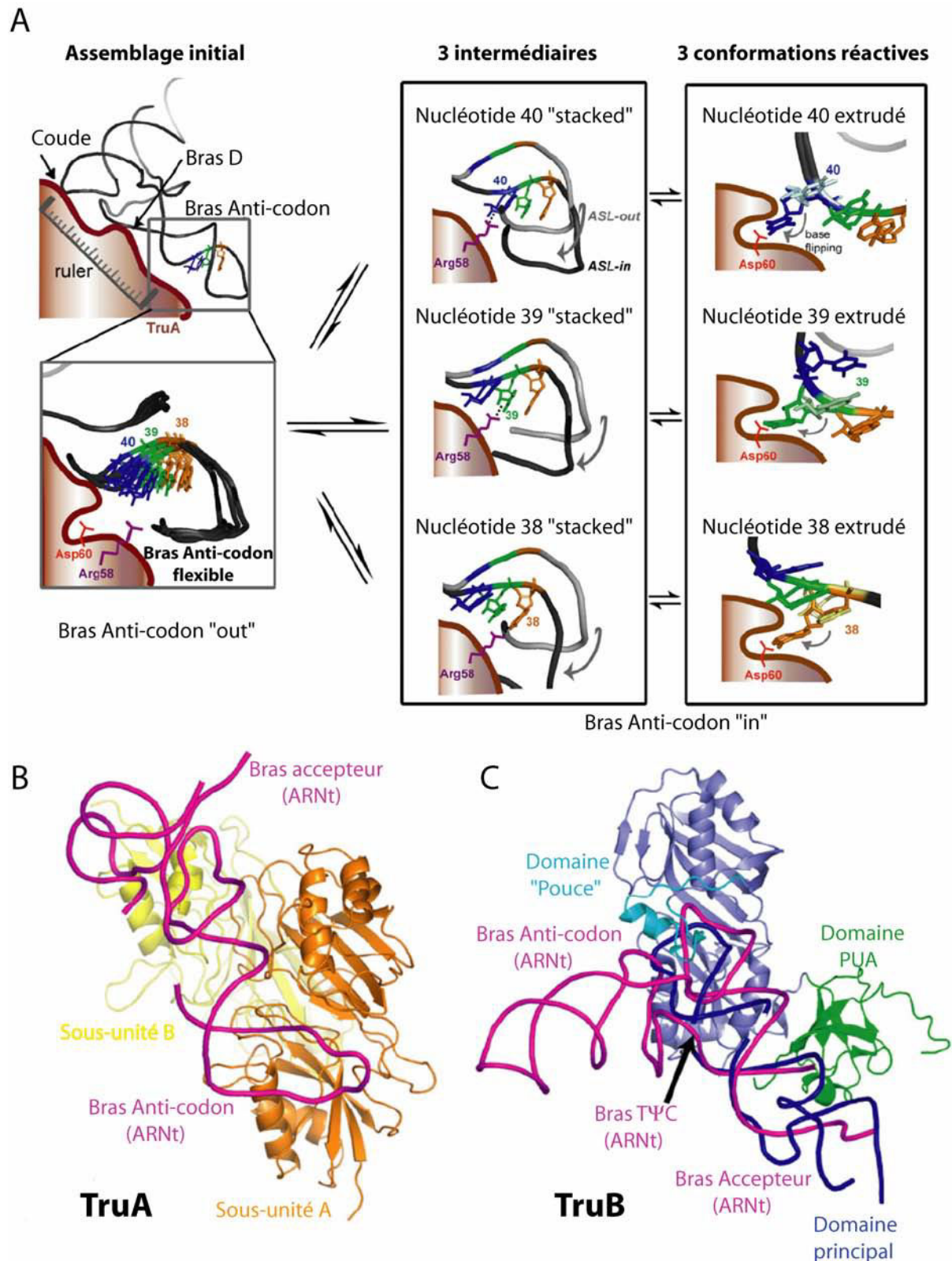


Figure III.14 : Modèle proposé pour la reconnaissance du nucléotide cible par TruA, et comparaison avec le mécanisme de reconnaissance de TruB (d'après Hur et Stroud, 2007). A. Modèle de reconnaissance par TruA. Compte tenu de la flexibilité du bras anti-codon lors de l'assemblage initial, trois intermédiaires peuvent être formés selon l'identité du nucléotide adjacent au résidu Arg58, ce qui conduit à trois conformations réactives. **B.** et **C.** Comparaison des structures de dimères de TruA avec TruB dans la même orientation. L'ARNt est représenté en rose dans les deux cas. En **C.** il s'agit d'une modélisation du positionnement de l'ARNt, le fragment utilisé pour la cristallisation y apparaît en bleu.

Ainsi, TruA interagirait, dans un premier temps, avec le coude (jonction entre les boucles D et TΨC) et l'hélice D de l'ARNt par le biais d'interactions électrostatiques et par reconnaissance de la forme de la chaîne ribose phosphate. Ces interactions orientent l'ARNt et positionnent le bras anti-codon à proximité du site actif (Fig. III.14A). Puis, en passant d'une conformation dite ouverte vers une conformation fermée, le résidu Arg58 vient interagir avec un des trois nucléotides : 38, 39 ou 40 et provoque son extrusion vers le résidu Asp60 catalytique. Le choix du nucléotide ne semble dépendre que de la flexibilité et donc des oscillations possibles de l'hélice du bras anti-codon. En effet, il a été démontré, qu'en stabilisant ou déstabilisant cette hélice, le taux de modification de chacun des trois nucléotides varie. La présence d'un résidu Ψ modifie la stabilité de l'hélice du bras anti-codon et affecte la lecture du codon par l'anti-codon (ce rôle est discuté ci-après). Un résidu Ψ altère donc le positionnement des nucléotides à proximité du résidu Arg58, et ainsi favorise ou au contraire empêche une seconde puis une troisième modification dans cette région de l'hélice (Hur et Stroud, 2007). Au final, les enzymes de la classe d'enzyme TruA ont évolué vers une reconnaissance de la flexibilité de l'hélice anticodon. Ce mode de reconnaissance permet de distinguer les hélices stables, leur évitant une stabilisation excessive, des hélices peu stables pour lesquelles il est possible de renforcer leur stabilité par la modification de résidus U en Ψ. Ce mode d'action permet donc d'ajuster finement la stabilité du bras anti-codon de chaque ARNt (Hur et Stroud, 2007).

En comparant TruA avec TruB (Fig. III.14B) on remarque que l'ARNt interagit avec la même crevasse du domaine principal de l'enzyme mais ceci avec des orientations différentes. TruA utilise la dimérisation pour reconnaître le coude de l'ARNt alors que TruB utilise son domaine PUA pour la reconnaissance du bras accepteur. La dynamique de l'assemblage de TruA sur l'ARNt repose principalement sur la reconnaissance du coude alors que celle de TruB repose sur le positionnement du bras TΨC dans le site actif grâce au domaine « pouce », le mécanisme d'extrusion de bases et les nombreuses interactions spécifiques avec les bases de la boucle TΨC. Au final, il existe donc un fort contraste entre ces deux enzymes ayant évolué indépendamment pour accommoder des substrats différents. TruB distingue le nucléotide substrat en employant des domaines périphériques et additionnels ainsi que des changements induits à la fois dans sa structure et celle de l'ARNt substrat tandis qu'un dimère de TruA distingue le coude des ARNt et positionne le bras anti-codon à proximité du site actif autorisant ainsi la modification de trois nucléotides dynamiquement accessibles grâce à la flexibilité du bras anti-codon.

5. Attribution des ARN:Ψ-synthases aux résidus Ψ détectés

Les ARN:Ψ-synthases présentent différentes potentialités de modification de cibles, qui sont expliquées par une grande variété de spécificités de reconnaissance. En effet, une ARN:Ψ-synthase peut reconnaître une ou plusieurs position(s) dans un ou plusieurs type(s) d'ARNnc. Par conséquent nous pouvons distinguer trois catégories d'ARN:Ψ-synthases (Tableau III.3):

- les enzymes dites site spécifiques ne modifient qu'un seul résidu U dans une classe d'ARNnc voire dans un seul ARN donné, c'est le cas de RsuA d'*E. coli* (Wrzesinski *et al.*, 1995), de TruB

d'*E. coli* (Nurse *et al.*, 1995) et de son homologue Pus4p chez la levure *S. cerevisiae* (Becker *et al.*, 1997a) ainsi que de RluB de *B. subtilis* (Niu et Ofengand, 1999).

- les enzymes dites région spécifiques modifient plusieurs positions proches au sein d'un même ARN, c'est le cas de RluD d'*E. coli* (Huang *et al.*, 1998a; Raychaudhuri *et al.*, 1998), de TruA d'*E. coli* et Pus3p de *S. cerevisiae* (Kammen *et al.*, 1988 ; Lecointe *et al.*, 1998).

- les enzymes à spécificités multiples modifient plusieurs résidus U dans des contextes différents au sein d'un ou plusieurs types d'ARNnc. C'est le cas, par exemple, de Pus1p de *S. cerevisiae* et de souris (Simos *et al.*, 1996 ; Motorin *et al.*, 1998 ; Chen et Patton, 1999), ainsi que de RluC d'*E. coli* (Conrad *et al.*, 1998).

Il est également possible de séparer, parmi les enzymes à spécificités multiples, celles qui reconnaissent un seul ou plusieurs type(s) d'ARNnc. On distingue dès lors une spécificité pour un substrat donné et non plus pour un résidu U localisé dans un contexte précis.

5.1. Situation dans les ARNt

La plupart des nucléotides modifiés sont communs à tous les ARNt d'un même organisme. Une ARN:Ψ-synthase impliquée dans ces modifications doit donc être adaptée aux différences qui existent entre les ARNt d'un organisme comme c'est le cas pour TruB.

Chez *E. coli*, cinq ARNt:Ψ-synthases différentes sont impliquées dans la modification de 7 positions différentes (Fig. III.15A) (Del Campo *et al.*, 2001). TruA est région spécifique comme évoqué précédemment. RluA est à spécificités multiples car elle modifie également une uridine appartenant à la grande sous-unité ribosomale. Les trois autres enzymes ne modifient qu'une seule position et sont donc site spécifiques.

Chez *S. cerevisiae*, six ARN:Ψ-synthases différentes sont impliquées dans la modification de 7 positions différentes des ARNt (Fig. III.15B). Trois d'entre elles sont site spécifiques, Pus4 l'équivalente de TruB, Pus6 et Pus8 les équivalents de RluA. Néanmoins, contrairement à son homologue bactérien, cette dernière ne serait pas impliquée dans la modification des ARNr. Bien que Pus7 soit l'équivalent de TruD qui est site spécifique, elle est à spécificités multiples puisqu'elle est impliquée dans la modification des positions 13 de certains ARNt et 35 du pré-ARNt^{Tyr} (Behm-Ansmant *et al.*, 2003) ainsi que dans celle de la position 35 du snRNA U2 (Ma *et al.*, 2003). Pus3 est l'équivalent de TruA qui est région spécifique. Enfin, Pus1 qui n'a pas d'équivalent bactérien, est à spécificités multiples car impliquée dans la modification à plusieurs positions des ARNt (Motorin *et al.*, 1998) et aussi dans la modification de la position 44 du snRNA U2 (Massenet *et al.*, 1999). Il faut aussi remarquer que, contrairement aux eucaryotes supérieurs, aucune pseudouridine n'est détectée en position 54 des ARNt de *S. cerevisiae*.

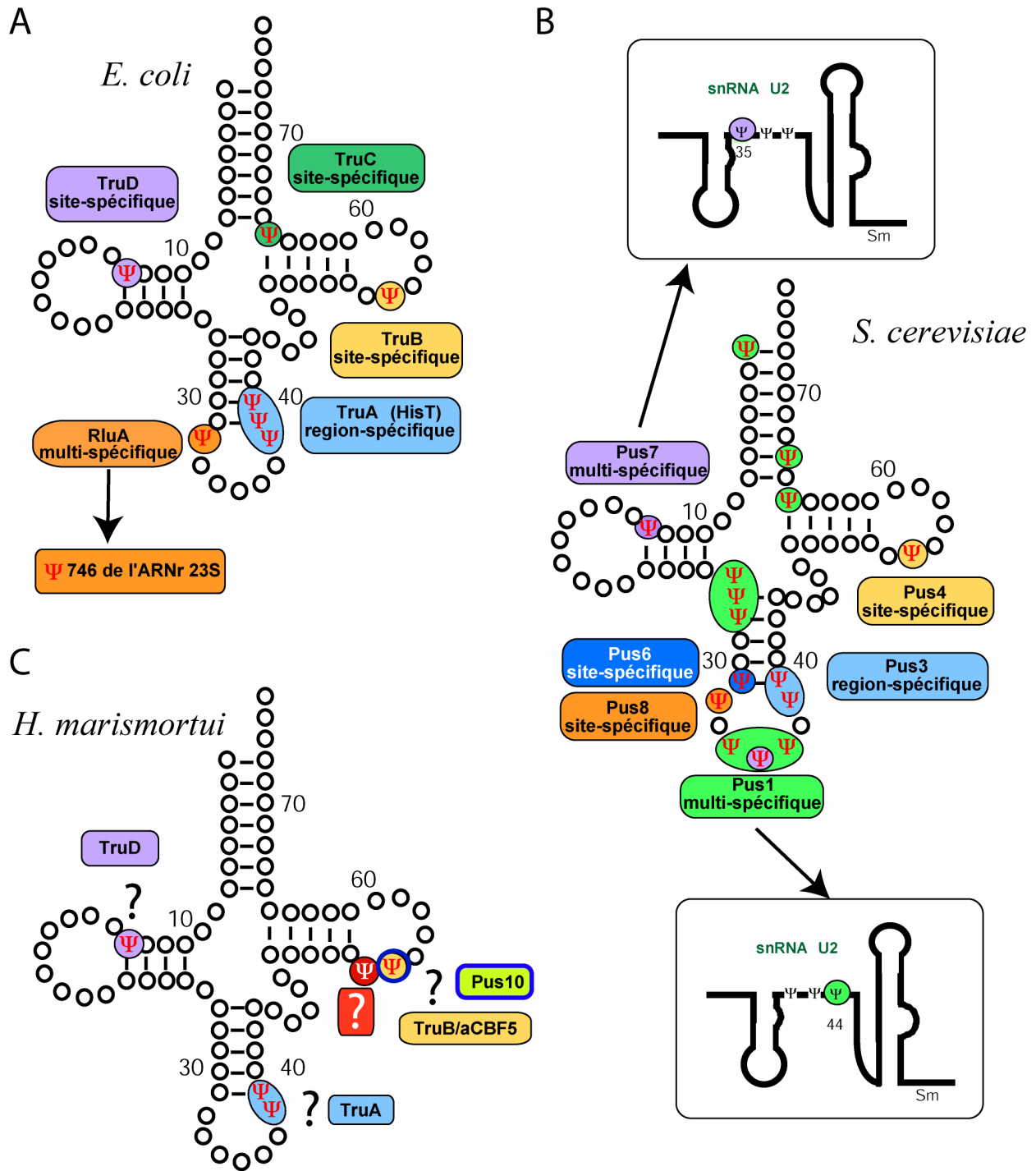


Figure III.15 : Schémas représentant l'état de l'identification des ARN:Ψ-synthases qui catalysent les différents résidus Ψ décrits dans les ARNt.

Les panneaux A, B et C représentent respectivement l'état des connaissances chez les Bactéries (*E. coli*), les Eucaryotes (*S. cerevisiae*) et les Archaea (*H. marismortui*). Les incertitudes chez les Archaea sont indiquées par des « ? ». Les références sont citées dans le tableau III.3. Les enzymes appartenant à une famille ou une classe commune sont représentées avec la même couleur.

Chez les Archaea, et particulièrement pour *H. marismortui* dont les modifications des ARNt ont été analysées, les attributions de positions modifiées aux enzymes connues restent hypothétiques (Fig. III.15C). Récemment, une étude a montré que TruB/aCBF5 et aPus10 de *P. furiosus* sont toutes deux capables de modifier la position 55 des ARNt (Roovers *et al.*, 2006). Cette redondance de fonction est surprenante et serait le premier exemple identifié. Nous avons pu vérifier ce résultat qui sera discuté dans la partie « Résultats ». L'attribution de la modification des positions 13 et 38/39 aux équivalents archaea des enzymes de TruD et TruA, respectivement, n'est qu'une prédiction logique réalisée par comparaison avec les situations chez les Bactéries et Eucaryotes et doit être vérifiée expérimentalement. Enfin, il est à noter que l'ARN:Ψ-synthase responsable de la pseudouridylation de la position 54 de tous les ARNt élongateurs n'a pas encore été identifiée. Cette position est également modifiée chez la plupart des eucaryotes, or aucune ARN:Ψ-synthase n'a encore été proposée comme responsable de sa modification.

Différentes classifications des ARN:Ψ-synthases ont été proposées sur la base de leur mode de reconnaissance :

- Une première classification propose de classer les enzymes de modification des ARNt en deux groupes en fonction de leur sensibilité ou non aux perturbations de structure globale de la molécule d'ARNt (Grosjean *et al.*, 1996).

Les enzymes dites du groupe I reconnaissent un élément structural local de l'ARNt (Fig. III.16). C'est le cas des enzymes de la famille TruB (Becker *et al.*, 1997b ; Gu *et al.*, 1998). La formation du résidu Ψ55 dépend de la structure de la boucle TΨC et en particulier du maintien de la paire de bases reverse Hoogsteen entre la position 54 et la position 58. La reconnaissance d'éléments structuraux localisés dans le bras TΨC est également une nécessité pour qu'il y ait formation d'autres nucléotides modifiés, comme le résidu m5U54 (T54) chez la levure (Becker *et al.*, 1997b). De la même façon, la méthylation sur le ribose du nucléotide Gm34 dans les ARNt de *X. leavis* (Droogmans *et al.*, 1986), la formation du résidu queueine à partir de la guanine en position 34 chez *E. coli* (Curnow *et al.*, 1993 ; Mueller et Slany, 1995) et la formation du résidu Ψ32 chez *X. laevis* (Grosjean *et al.*, 1996), notamment, ne dépendent que de la présence de la structure tige-boucle anticodon.

Les enzymes dites du groupe II sont sensibles à la structure tridimensionnelle globale de l'ARNt (Fig. III.16). Les enzymes TruA et Pus3p catalysant la formation des résidus Ψ en position 38, 39 et 40 chez *E. coli* et chez *S. cerevisiae*, respectivement, appartiennent au groupe II (Edqvist *et al.*, 1993 ; Grosjean *et al.*, 1996 ; Lecoite *et al.*, 1998). Cependant, Pus3p semble présenter des différences subtiles de reconnaissance de l'ARNt par rapport à TruA. En effet, l'enzyme de levure modifie *in vitro* et *in vivo* les positions 38 et 39, mais pas 40, contrairement à TruA (Lecoite *et al.*, 1998). Les enzymes catalysant la formation notamment des nucléotides modifiés m22G26 (Edqvist *et al.*, 1992 ; Edqvist *et al.*, 1994), m1G37 (Holmes *et al.*, 1992 ; Redlak *et al.*, 1997 ; Brule *et al.*, 1998), I34 (Auxilien *et al.*, 1996), i6A37 (Motorin *et al.*, 1997) et t6A34 (Morin *et al.*, 1998) font également partie du groupe II (Fig. III.16). L'hypothèse émise dans le cas de la formation des résidus I34 et i6A37 est que la reconnaissance de la cible requiert la reconnaissance de l'ARNt en

globalité et que sa fixation induit une torsion de la structure tige-boucle de l'anticodon permettant d'optimiser un contact avec le site actif de l'enzyme nécessaire à la formation de ces nucléotides modifiés (Auxilien *et al.*, 1996 ; Motorin *et al.*, 1997).

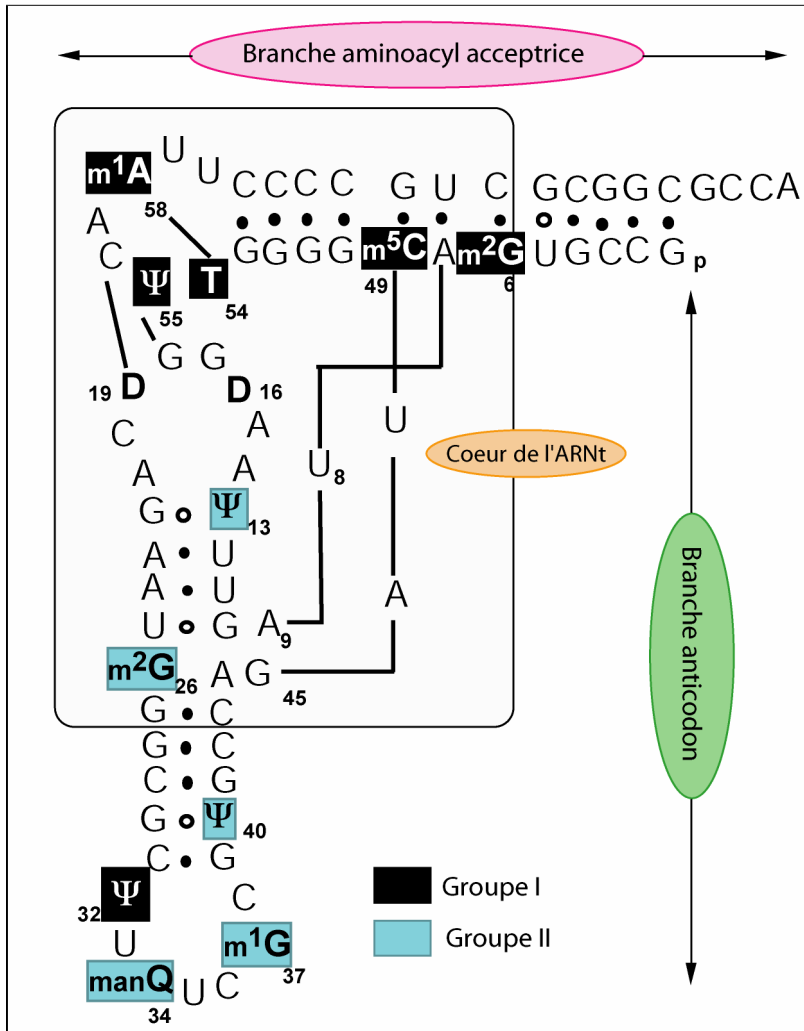


Figure III.16 : Classification des enzymes de modification des ARNt en deux groupes selon l'influence de la structure tertiaire de l'ARNt (adapté de Grosjean *et al.*, 1996).

Les enzymes du groupe I ne sont pas affectées par des perturbations de l'architecture de l'ARNt, tandis que les enzymes du groupe II sont dépendantes de la structure tridimensionnelle de l'ARN. La structure 3D de l'ARNt^{Asp} de *S. cerevisiae* et les modifications formées dans cet ARNt (ou ses variants) lorsque ces ARNt sont microinjectés dans le cytoplasme d'oocytes de *X. laevis* sont représentées ici.

- L'influence d'un intron sur la formation des modifications post-transcriptionnelles portées par un ARNt a été mise en évidence *in vivo* et *in vitro* dans plusieurs études (Jiang *et al.*, 1997 ; pour revue, Motorin *et al.*, 1997). Une seconde classification proposée pour les ARNt:Ψ-synthases eucaryotes, est basée sur l'importance des introns avec : les enzymes insensibles à la présence d'un intron, celles dont l'activité est intron-dépendante et celles qui nécessitent l'absence de l'intron (Fig. III.17). Il faut remarquer que la majorité des enzymes de modification des ARNt sont insensibles à la présence d'un intron. Le groupe des enzymes de modification intron-dépendantes est constitué uniquement d'enzymes modifiant des nucléotides appartenant au bras de l'anticodon, mais toutes les modifications situées dans cette région des ARNt ne sont pas intron-dépendantes. La nécessité de la présence des introns pour la formation de certains nucléotides modifiés n'est pas encore démontrée expérimentalement. Cependant, il a été montré récemment que la synthèse du

résidu pseudouridine en position 35 de l'ARNt^{Tyr} de plantes dépend probablement de la taille de l'intron plutôt que de sa séquence ou de sa structure (Pienkowska *et al.*, 2002). Enfin, en ce qui concerne le dernier groupe, l'effet inhibiteur de l'intron sur la formation de la modification serait expliqué par changement de structure local et global qui empêcherait la reconnaissance de la cible par les enzymes appartenant à ce groupe.

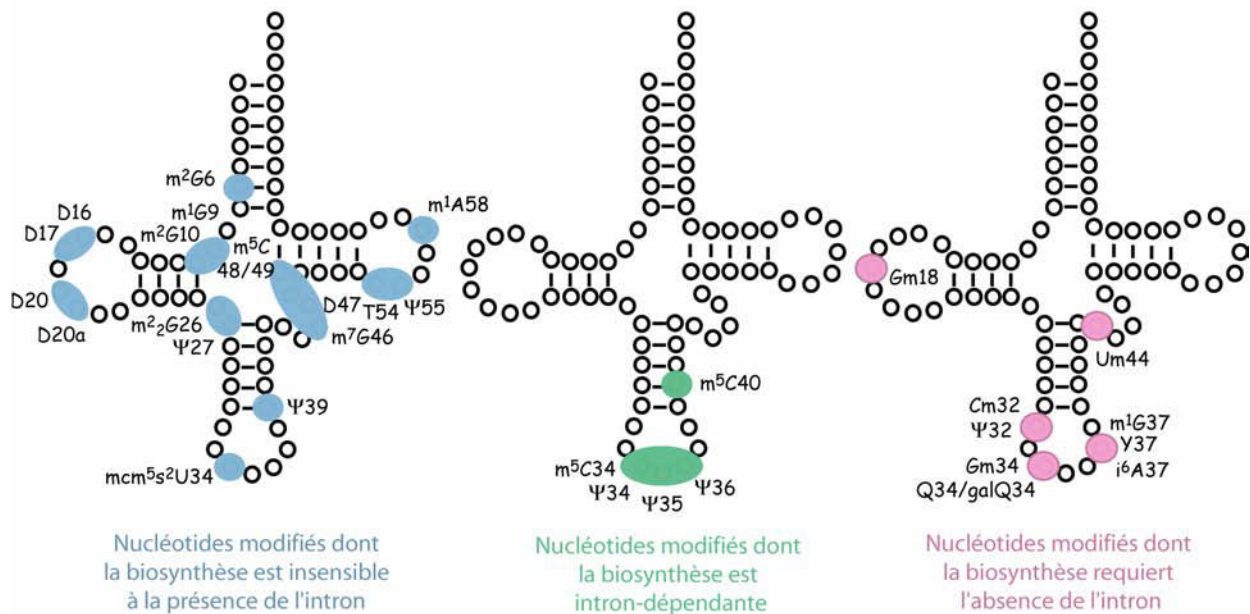


Figure III.17 : Classification des enzymes de modification des ARNt en trois groupes selon l'influence de la présence d'un intron au sein de l'ARNt (adapté de Grosjean *et al.*, 1997).

5.2. Situation dans les ARNr

Chez les bactéries, tous les résidus Ψ sont formés par des ARN:Ψ-synthases qui ne nécessitent pas d'ARN guide. Les 10 résidus Ψ identifiés chez *E. coli* résultent de l'action de 6 ARN:Ψ-synthases (Fig. III.18) (Del Campo *et al.*, 2001). Trois d'entre elles sont site spécifiques : RluB, RluE et RluF. L'enzyme RluA, comme précisé ci-avant, modifie également les ARNt et a donc une spécificité multiple. RluD modifie trois positions dans l'hélice 69 du domaine IV de l'ARN de la grande sous-unité ribosomale et est donc région spécifique. L'hélice 69 pourrait être l'élément reconnu et chaque résidu U situé à une distance compatible avec son positionnement dans le site actif de RluD pourrait être isomérisé en pseudouridine (Huang *et al.*, 1998a ; Raychaudhuri *et al.*, 1998). Il est à noter également que chacun des résidus U modifiés par RluD est suivi par un résidu A en 3'. Enfin, RluC reconnaît trois positions dans des contextes apparemment différents et a donc une spécificité multiple. Ces positions sont éloignées les unes des autres dans la séquence primaire mais sont en fait proches dans la structure tertiaire de l'ARNr 23S (Conrad *et al.*, 1998 ; Huang *et al.*, 1998a). En effet, la proximité des résidus A960 et C2475 a été démontrée par des expériences de pontage (Dokudovskaya *et al.*, 1996). RluC pourrait donc reconnaître un élément structural dans l'ARNr 23S et modifier dans cette région des résidus U qui, par ailleurs, sont toujours suivis par un

résidu G en 3' (Conrad *et al.*, 1998). Par conséquent, RluC et RluD semblent sensibles à la nature du nucléotide en 3' du site à modifier.

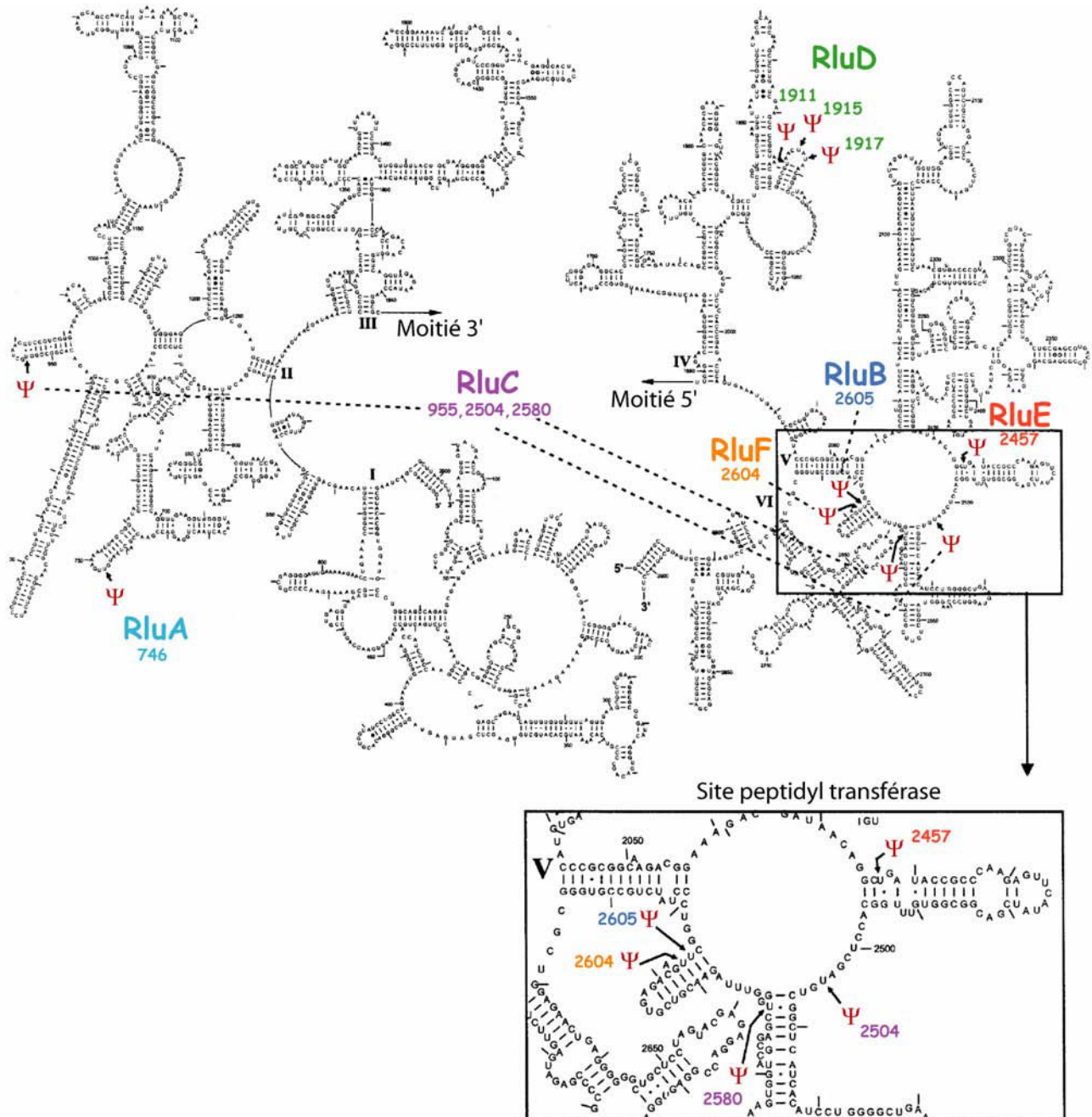


Figure III.18 : Les cibles des ARNr:Ψ-synthases d'*E. coli* au sein de l'ARNr 23S (d'après Del Campo *et al.*, 2001).

Il est intéressant de remarquer que les enzymes RluB et RluF d'*E. coli* reconnaissent des cibles espacées d'un nucléotide soit, respectivement les positions 2605 et 2604 de l'ARNr 23S (Fig. III.16) (Del Campo *et al.*, 2001). Ces deux protéines reconnaissent donc la même région de l'ARNr 23S, mais dirigent la conversion spécifique d'un seul résidu U.

Seul le mécanisme de reconnaissance de RluA a été étudié expérimentalement par le biais de l'analyse structurale d'un cocrystal correspondant à RluA associé à un fragment d'ARN qui mime le bras anticodon d'un ARNt (Fig. III.19) (Hoang *et al.*, 2006). Comme l'étude de TruB, celle-ci montre l'implication dans les interactions avec l'ARN cible de résidus appartenant aux motifs II, III, domaine « pouce » ainsi que, et contrairement à TruB, des résidus appartenants au motif I, en particulier de résidus constituant le domaine « index ». Cette étude montre également que seule la partie distale de la tige-boucle i.e. les bases des deux dernières paires de bases, ainsi que les bases de la paire de bases reverse-Hoogsteen U33-A36 et des nucléotides extrudés de la boucle, à savoir la cible U32, G34, A35 et A37 mais aussi A38 sont en interaction spécifique avec des résidus de l'enzyme. Le résidu U746 de l'ARNr de la grande sous-unité ribosomale est situé dans un contexte analogue. Bien que la boucle comportant la position U746 compte un nucléotide de plus, la formation d'une paire de bases reverse Hoogsteen U-A est possible. Il n'est pas actuellement possible de prédire si cette modification dans l'ARNr est la marque d'une activité résiduelle, la marque de l'évolution et de la divergence de RluA ou si cette modification présente une fonction au sein de l'ARNr.

Pour être complet il faut ajouter que la formation du seul résidu Ψ détecté dans l'ARNr 16S d'*E. coli* est catalysée par l'enzyme RsuA.

Chez *S. cerevisiae*, seule la formation du résidu Ψ 2819 présent dans les ARNr mitochondriaux est catalysée par une ARN: Ψ -synthase (Pus5) (Ansmant *et al.*, 2000). Enfin, jusqu'à nos travaux, aucune ARN: Ψ -synthase d'archaea n'avait été montrée comme directement responsable de la formation de résidus Ψ dans les ARNr. Les pseudouridylation des ARNr cytoplasmiques des eucaryotes étant catalysées par des systèmes impliquant des snoRNA à boîtes H/ACA, il était possible qu'il en soit de même chez les archaea, ce que nous avons démontré au cours de nos travaux.



B. ARN:Ψ-synthases associées à des guides ARN

Il était connu qu'une classe d'ARN:Ψ-synthases appartenant à la famille TruB et nommée Cbf5p chez *S. cerevisiae*, Dyskérine chez l'homme, NAP57 chez les rongeurs (Meier et Blobel, 1994 ; Lafontaine *et al.*, 1998b ; Watanabe et Gray, 2000; Wang et Meier, 2004) assure la pseudouridylation de nombreuses positions grâce à l'utilisation des ARN guides appelés snoRNA à boîtes H/ACA. Nous avons démontré expérimentalement que l'homologue aCBF5 d'archaea joue le même rôle. La détermination de la cible de cette classe d'ARN:Ψ-synthases ne dépend donc que de la complémentarité entre la séquence de la boucle interne de la tige-boucle du sno/sRNA à boîtes H/ACA et de la séquence de l'ARNnc ciblé. Il semblerait qu'un appariement reposant sur au moins 8 paires de bases soit requis. Cependant, les requis au niveau de la longueur et de la régularité de ces appariements n'étaient pas encore bien définis. En effet, aucune analyse expérimentale n'avait permis de définir précisément la longueur minimale des deux appariements permettant à CBF5 de modifier l'uridine ciblée par un guide ARN. Seule l'analyse des appariements prédits, par voie bioinformatique, entre la séquence comportant un résidu Ψ détecté et un guide à boîtes H/ACA proposé pour être impliqué apportait des renseignements sur les déterminants de la reconnaissance de la cible. Nos travaux qui ont contribué à une meilleure compréhension des requis existants au niveau de ces appariements ainsi qu'à une meilleure description des caractéristiques des sRNA à boîtes H/ACA du domaine archaea seront présentés dans la partie « Résultats ».

En plus des guides à boîtes H/ACA, l'enzyme CBF5 et ses homologues nécessitent l'intervention de trois autres protéines dites auxiliaires appelées Nop10p, Gar1p et Nhp2p chez la levure, NOP10, GAR1 et NHP2 chez l'Homme. On savait que des homologues aNOP10, aGAR1 et L7Ae étaient présents chez les Archaea (Henras *et al.*, 1998 ; Watkins *et al.*, 1998 ; Watanabe et Gray, 2000). Mais leurs implications dans la pseudouridylation des ARNr n'avaient pas encore été démontrées. Nos études et celles de l'équipe de Terns ont permis, par reconstitution de particules sRNPs actives *in vitro* à partir de transcrits produit *in vitro* et de protéines recombinantes, de démontrer le rôle de ces protéines comme cela sera présenté de manière plus approfondie dans la partie « Résultats ». L'absence de données correspondantes chez les eucaryotes, en particulier l'absence de reconstitution de particules snoRNP actives était liée à l'insolubilité des protéines recombinantes CBF5 et GAR1.

La partie qui suit vise à décrire les caractéristiques des 4 protéines impliquées dans les sRNA H/ACA d'archaea. En effet, des études structurales très récentes, réalisées à partir de différentes combinaisons des 4 protéines provenant de trois espèces eucaryotes : *P. abyssi*, *P. furiosus* et *M. jannaschii*, avec ou sans un fragment ARN mimant un sRNA à boîtes H/ACA et toujours sans une molécule d'ARN ciblée, ont permis de définir :

- les résidus acides aminés de aCBF5, L7Ae et aNOP10, mais pas aGAR1, interagissant avec les sRNA à boîtes H/ACA
- les nucléotides du sRNA à boîtes H/ACA impliqués dans les contacts avec ces 3 protéines

- les résidus importants pour les 3 interactions protéines-protéines entre aNOP10 et L7Ae, aNOP10 et aCBF5, aCBF5 et aGAR1 (Charron *et al.*, 2004 ; Hamma et Ferre-D'Amare, 2004 ; Hamma *et al.*, 2005 ; Li et Ye, 2006 ; Manival *et al.*, 2006 ; Rashid *et al.*, 2006).

Très récemment, l'équipe de Ye a réussi à cristalliser un complexe comportant un fragment d'ARN mimant une tige-boucle d'un sRNA à boîtes H/ACA associé aux protéines L7Ae, aCBF5, aNOP10 et aGAR1 provenant de *P. furiosus* (Li et Ye, 2006). Cette structure, bien qu'elle n'ait pas été réalisée en présence d'un fragment ARN substrat, est actuellement la plus aboutie et sera utilisée comme référence pour discuter des différentes interactions qui se produisent entre ces 4 protéines et le sRNA à boîtes H/ACA.

1. Caractéristiques des 4 protéines des sno/sRNP à boîtes H/ACA

CBF5 (Cbf5p / aCBF5 / NAP57 / Dyskérine)

L'ARN:Ψ-synthase CBF5 a une taille comprise entre 40 kD (archaea) et 65 kD (levure). La représentante archaea présente environ 35 % d'identité de séquence avec TruB et adopte une structure très similaire à celle de TruB, d'*E. coli* (Fig. III.20) (Hoang et Ferre-D'Amare, 2001 ; Hamma *et al.*, 2005 ; Manival *et al.*, 2006 ; Rashid *et al.*, 2006). Le domaine catalytique de aCBF5 est composé, comme TruB de deux sous-domaines (D1 et D2) qui se rejoignent au niveau d'un sillon comportant le site catalytique. Le site actif de aCBF5 se superpose bien avec celui de TruB et le résidu His responsable de l'extrusion de bases est bien conservé. Le domaine « pouce » de TruB, constitué par les deux éléments d'insertion Ins1 et Ins2, est absent ou raccourci dans les enzymes CBF5 d'archaea et d'eucaryotes. En effet, la boucle β7 de aCBF5 peut remplacer en partie le domaine d'insertion 2 de TruB. Le domaine PUA de CBF5 est plus grand que celui de TruB. Par comparaison avec TruB, une trentaine de résidus supplémentaires situés en N-ter entourent le domaine PUA, et quelques résidus supplémentaires sont aussi présents à l'extrémité C-ter (Fig. III.20) (Manival *et al.*, 2006). Ces extensions terminales sont importantes pour la stabilité et/ou le repliement de aCBF5. Il est à noter également que de nombreuses mutations responsables de la dyskératose congénitale sont localisées dans le domaine PUA localisé à l'extrémité C-ter ainsi que dans l'extrémité N-ter de aCBF5 qui entoure ce dernier. Le domaine PUA flanque une crevasse située du côté opposé au site actif et qui est impliquée dans l'interaction avec le sRNA.

Chez la levure, Cbf5p contient un domaine KKE/D dans sa région C-terminale, qui n'est pas retrouvé dans les protéines homologues de certains eucaryotes et archaea (Watkins *et al.*, 1998 ; Winkler *et al.*, 1998). Le rôle de ce domaine n'est pas élucidé et sa délétion n'a pas d'effet sur la viabilité des cellules de *K. lactis* (Winkler *et al.*, 1998). Néanmoins, il faut noter que le domaine KKE/D est retrouvé dans un certain nombre d'autres protéines associées à des petits ARN (Nop56p et Nop58p associées aux snoRNA à boîtes C/D, et Prp31p associée au snRNA U4 spliceosomal). Cependant, il ne semble pas impliqué dans des interactions avec des ARN.

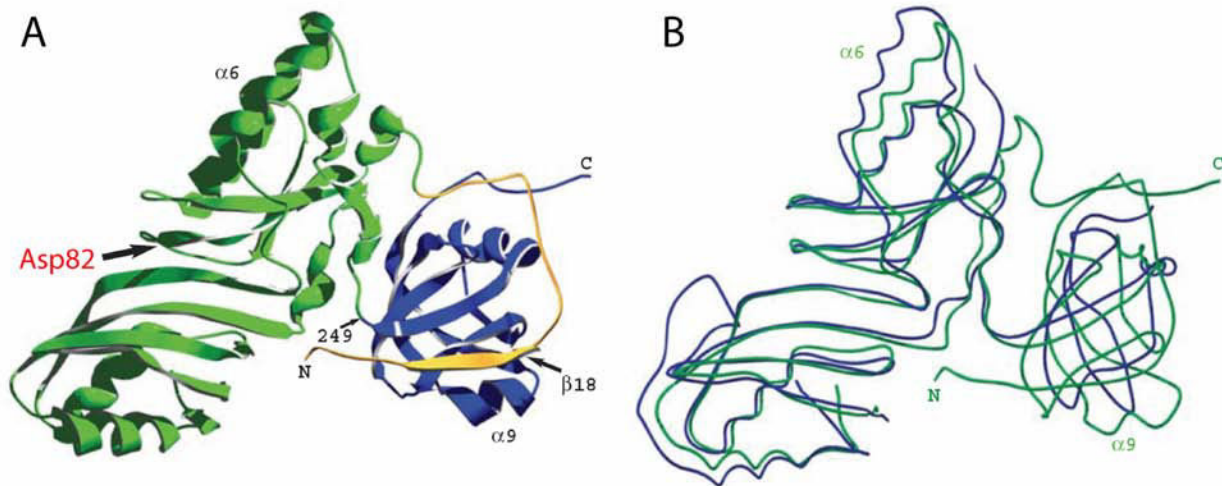


Figure III.20 : Structure 3D de aCBF5 établie à partir d'un cristal du complexe aCBF5-aNOP10 de *P. abyssi* et comparaison avec la structure 3D de TruB de *Thermotoga maritima* (d'après Manival *et al.*, 2006).

A. Représentation de la structure 3D de aCBF5. L'extension en N-ter est représentée en jaune. **B.** Superposition de la structure 3D de aCBF5 (en vert) avec la structure 3D de TruB libre (en bleu).

NOP10 (Nop10p / aNOP10)

Il s'agit d'une petite protéine (entre 6 et 10 kDa) essentielle aux clivages A1 et A2 des pré-ARNr et à l'activité de pseudouridylation des sno/sRNA H/ACA (Henras *et al.*, 1998 ; Charpentier *et al.*, 2005 ; nos résultats). Elle a été identifiée avec Nhp2p par purification des snoRNP à boîtes H/ACA de la levure grâce à une version étiquetée de Gar1 (Gar1-ZZ) (Henras *et al.*, 1998). Cette protéine, conservée chez tous les archaea et eucaryotes, ne présente aucun homologue en dehors des protéines NOP10. La comparaison des séquences de NOP10 d'eucaryotes et d'archaea révèle la conservation de la partie C-ter composée d'une hélice α mais révèle aussi de nombreuses différences dans le domaine N-ter. En effet, aNOP10 comporte un doigt à Zinc du côté N-ter chez les archaea qui est remplacé par une boucle chez les eucaryotes (Hamma *et al.*, 2005 ; Khanna *et al.*, 2006). Les deux domaines sont reliés par une région « linker » qui paraît désordonnée (Fig. III.21) (Hamma *et al.*, 2005 ; Manival *et al.*, 2006).

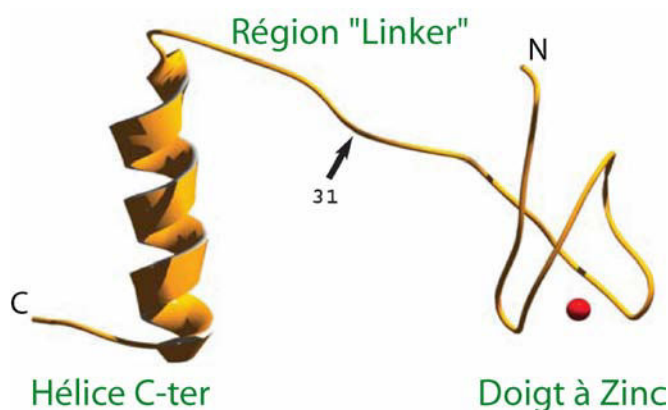


Figure III.21 : Structure 3D de la protéine aNOP10 établie à partir d'un cristal du complexe aCBF5-aNOP10 de *P. abyssi* (d'après Manival *et al.*, 2006). L'ion Zinc est représenté en rouge. La flèche indique la séparation entre les domaines N et C-ter.

Nhp2p / L7Ae

Les protéines Nhp2p et L7Ae présentes respectivement chez les eucaryotes et les archaea, appartiennent, comme les protéines Snu13p de levure et 15,5 kD humaine, à la famille L7Ae/L30. Leur taille varie entre 7 kD (archaea) et 22 kD (eucaryotes). La protéine L7Ae de *M. jannaschii* présente 45% d'identité de séquence avec la protéine NHP2 humaine (Henras *et al.*, 1998). Bien que Nhp2p ait des propriétés physico-chimiques semblables à celles des protéines HMG (High Mobility Group), elle ne présente aucune similarité de séquence significative avec ces protéines (Kolodrubetz et Burgum, 1991). La protéine Nhp2p est associée à tous les snoRNA à boîtes H/ACA et elle est essentielle à ces complexes puisque son absence conduit à la déstabilisation des snoRNP et à l'absence d'activité (Henras *et al.*, 1998 ; Watkins *et al.*, 1998). Contrairement à L7Ae, les protéines 15,5 kD et Snu13p qui s'associent aux motifs K-Turn, Nhp2p ne reconnaît pas spécifiquement ce motif. En effet, il semble que la présence d'une tige-boucle irrégulière est nécessaire et suffisante pour la fixation de Nhp2p (Henras *et al.*, 2001). Les sites d'interaction entre Nhp2p et le snoRNA ne sont pas encore déterminés. La spécificité de la liaison de Nhp2p pourrait résulter de l'interaction avec d'autres facteurs protéiques de la snoRNP. Ainsi, il est proposé qu'*in vivo*, Nhp2p se fixe à la boîte H après interaction avec CBF5 et NOP10 (Wang et Meier, 2004).

GAR1 (Gar1p / aGAR1)

Il s'agit d'une protéine nucléolaire de 25 kD chez les eucaryotes, essentielle à la viabilité cellulaire, mais non nécessaire à la stabilité *in vivo* des snoRNA H/ACA. Cette protéine contient deux domaines renfermant des répétitions de séquences glycine/arginine : les domaines GAR, localisés à chacune de ses extrémités. Ces domaines GAR, présents dans de nombreuses protéines nucléolaires comme la nucléoline ou la fibrillarine (pour revue, Ginisty *et al.*, 1999), participent à la localisation nucléolaire de ces protéines et ont la capacité de fixer aspécifiquement l'ARN. Cependant, dans le cas de Gar1p, ces domaines ne sont nécessaires ni à la localisation nucléolaire, ni à la fixation sur les snoRNA (Girard *et al.*, 1994 ; Bagni et Lapeyre, 1998). La protéine aGAR1 d'archaea d'une taille d'environ 11 kD ne présente pas ces domaines GAR mais partage avec les homologues eucaryotes un domaine central qui adopte une structure en tonneau à 6 brins β . Ce dernier ressemble à des domaines de fixation à l'ARN présents dans des facteurs bactériens d'initiation et d'élongation de la traduction. Chez les eucaryotes, il a été proposé que Gar1p se fixe au snoRNA (Bagni et Lapeyre, 1998 ; Wang et Meier, 2004) alors que chez les archaea aGAR1 n'interagit pas directement avec l'ARN (Li et Ye, 2006). La différence pourrait provenir de l'absence de domaine GAR.

2. Interactions entre les constituants des complexes sRNP à boîtes H/ACA

2.1. Structure globale d'un sRNA associé aux 4 protéines

Le complexe ressemble à un triangle équilatéral dont les côtés mesurent environ 80-90 Å de longueur et 20 Å d'épaisseur (Fig. III.22). Le site catalytique de aCBF5 se trouve au centre et les trois angles sont représentés par le domaine PUA, L7Ae et aGAR1 (Li et Ye, 2006). La protéine aNOP10 est prise en « sandwich » entre L7Ae et aCBF5. La tige-boucle de l'ARN s'étend sur 80 Å et se trouve encerclée par L7Ae, aNOP10 et aCBF5 au niveau de l'hélice apicale, et par le domaine PUA de aCBF5 au niveau de l'hélice basale. Leur fixation provoque également une courbure de 50° de l'axe de la tige-boucle. La poche de pseudouridylation se positionne au niveau du site catalytique de aCBF5 de telle sorte que l'élément antisens 5' se situe loin de la surface de aCBF5 alors que l'élément antisens 3' se situe à proximité de aCBF5. Les 4 protéines s'associent sur une seule face du sRNA autorisant ainsi un accès à l'ARN substrat qui peut entrer et sortir du complexe sans que les 4 protéines ne se dissocient du sRNA. Il est à noter que le sRNA utilisé pour l'étude de la structure du complexe ressemble au sRNA Afu46 d'*A. fulgidus* mais présente des modifications indiquées sur la figure III.22. L'ensemble des interactions relevées entre les 4 protéines et l'ARN est détaillé dans les figures III.23 et 24.

2.2. Association de L7Ae sur le motif K-turn

La structure des protéines 15,5 kD humaine associée au snRNA U4 (Vidovic *et al.*, 2000) et al7Ae de *H. marismortui* associée à l'ARNr (Ban *et al.*, 2000 ; Klein *et al.*, 2001) sont connues et notre équipe a étudié la structure de L7Ae libre afin de savoir si les protéines de la famille L7Ae/L30 subissent des changements de conformation importants lors de la fixation à l'ARN (Charron *et al.*, 2004). Notre équipe a ainsi montré que le squelette polypeptidique de la protéine L7Ae libre de *P. abyssi* est très similaire à celui de la protéine humaine 15,5kD associée au snRNA U4 et à celui de la protéine L7Ae de *H. marismortui* dans la sous unité 50S (Fig. III.24). Seuls de faibles réarrangements de chaînes latérales doivent avoir lieu lors de la fixation de l'ARN. En effet, la poche de fixation du nucléotide U proéminent dans la structure en «K-turn» est préformée dans la protéine libre. Il apparaît que l'ARN subit de forts changements de conformation lors de sa fixation à la protéine, contrairement à la protéine qui ne subit que très peu de modifications. C'est donc l'ARN qui adapterait sa conformation à celle de la protéine (Charron *et al.*, 2004).

L'interface entre la protéine L7Ae de *M. jannaschii* et le motif K-loop de Afu190 d'*A. fulgidus* (K-turn auquel il manque l'hélice comportant les appariements W-C, voir chapitre II) montre que les bases des nucléotides des deux paires reverse Hoogsteen, le résidu U projeté à l'extérieur ainsi que le résidu purine en 5' de ce dernier sont en interaction directe avec des acides aminés de L7Ae (Hamma et Ferre-D'Amare, 2004). Les résidus de L7Ae impliqués dans ces interactions sont montrés en figure III.23, 24 et 25.

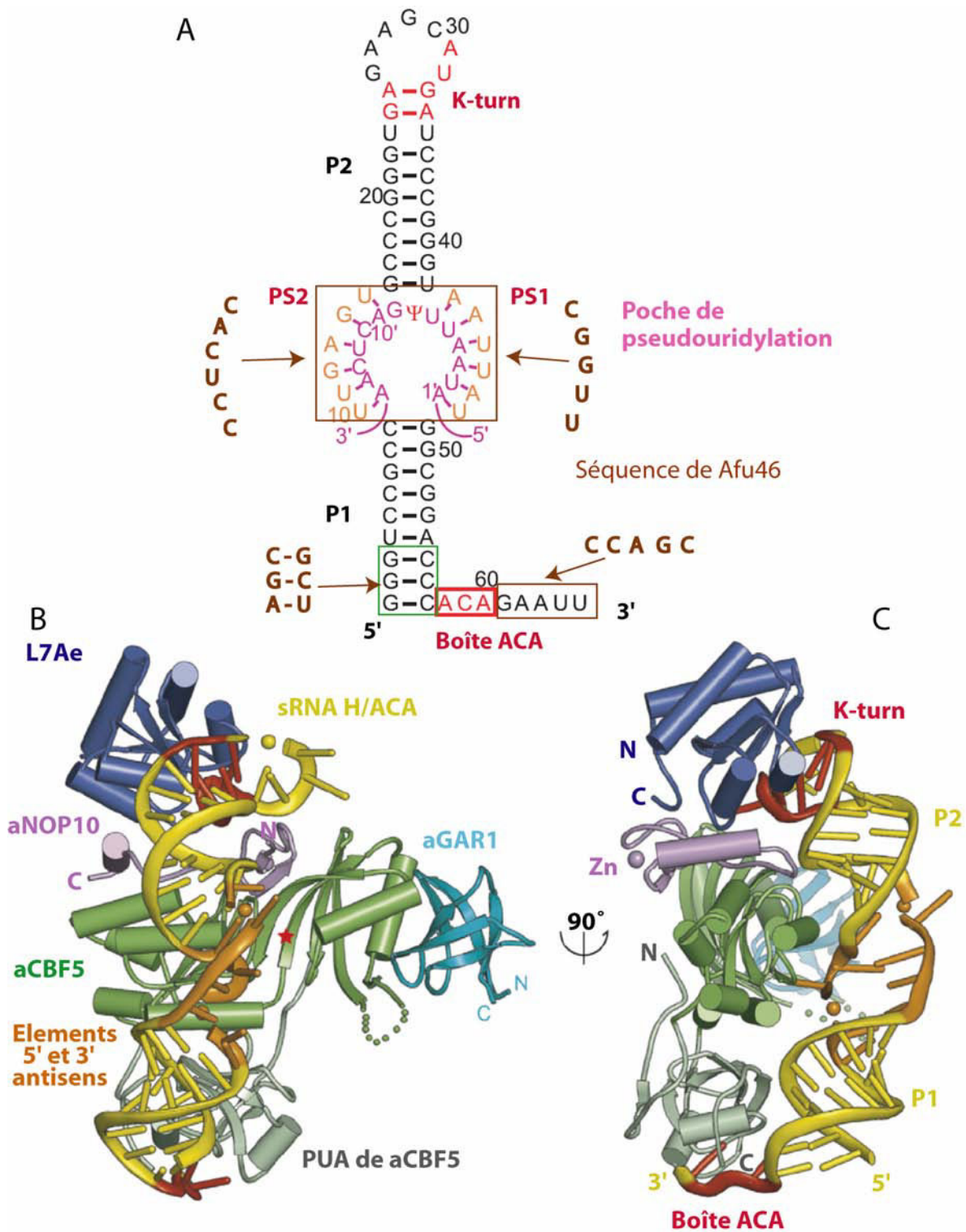
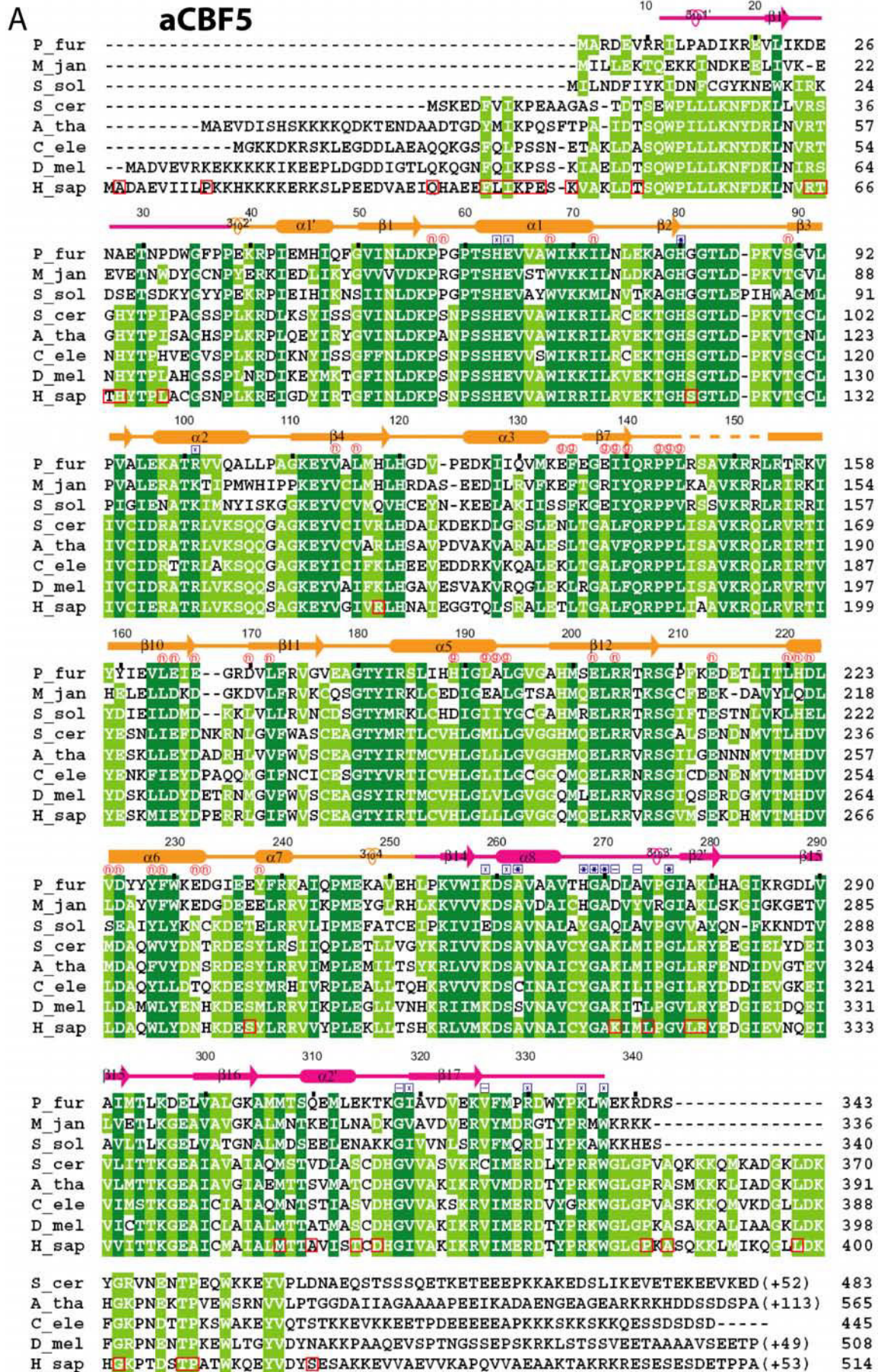
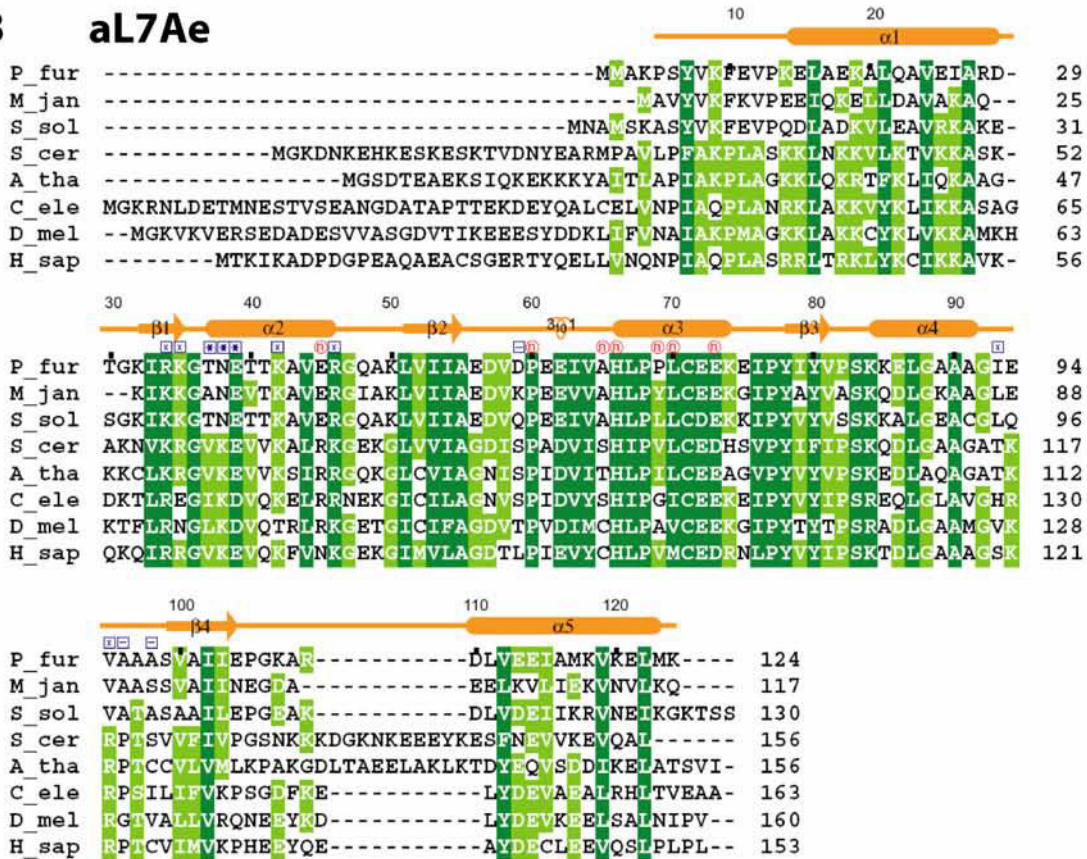


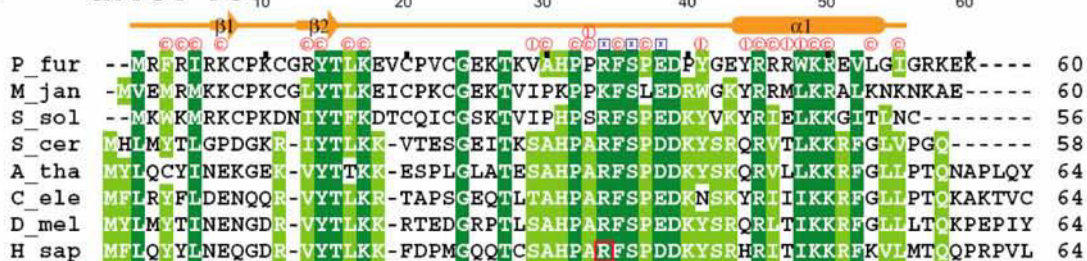
Figure III.22 : Structure 3D à haute résolution d'un complexe comportant un sRNA à boîtes H/ACA, aCBF5, aNOP10, aGAR1 et L7Ae établie par radiocristallographie (adapté de Li et Ye, 2006). **A.** Structure secondaire du sRNA à boîtes H/ACA utilisé pour cette étude. Il s'agit du sRNA Afu46 provenant de *A. fulgidus* auquel des modifications ont été apportées : les séquences de la poche de pseudouridylation, de la base de l'hélice basale (P1) et de la séquence terminale en 3' ont été modifiées (encadrées en marron). Les hélices formées lors de l'appariement de l'ARN substrat (rose) avec les éléments antisens (orange) sont appelées PS1 et PS2. Le motif K-turn et la boîte ACA sont représentés en rouge. **B.** et **C.** Structure 3D du complexe présentée selon deux angles de vue différents. L'étoile indique l'aspartate catalytique. Différentes couleurs sont utilisées pour le sRNA (jaune), le domaine principal de aCBF5 (vert), son domaine PUA (gris), L7Ae (bleu foncé), aNOP10 (rose) et aGAR1 (bleu clair). L'ARN substrat présenté en A n'est pas présent dans la structure 3D.



B aL7Ae



C aNOP10



D aGAR1



☒ Lie l'ARN par la chaîne latérale ☐ Lie l'ARN par la chaîne principale ☒ Lie aCBF5 ☒ Lie aGAR1
☒ Lie l'ARN par la chaîne principale et la chaîne latérale ☒ Lie aNOP10 ☒ Lie aL7Ae

(Pages précédentes) **Figure III.23 : Alignement multiple des séquences des protéines aCBF5 (A), L7Ae (B), aNOP10 (C) et aGAR1 (D) de différentes espèces avec celles de ces protéines chez *P. furiosus* (P_fur), et indication des résidus proposés pour être impliqués dans les interactions (d'après Li et Ye, 2006).**

Les séquences sont issues de *M. jannaschii* (M_jan), *S. solfataricus* (S_sol) ou *S. acidocaldarius* (S_aci), et 5 eucaryotes : *S. cerevisiae* (S_cer), *A. thaliana* (A_tha), *C. elegans* (C_ele), *Drosophila melanogaster* (D_mel) et *H. sapiens* (H_sap). Les éléments de structure secondaire sont indiqués au dessus de la séquence. Les résidus en interaction avec des résidus de l'ARN guide (carrés) ou des autres protéines (cercles) sont représentés comme dans l'encart. Les résidus complètement conservés sont représentés en vert foncé, et ceux qui présentent au moins 60% de conservation sont représentés en vert clair. Les mutations impliquées dans le syndrome de la Dyskératose Congénitale sont indiquées en rouge dans la séquence de la dyskérine humaine et de la protéine NOP10 humaine (Walne *et al.*, 2007).

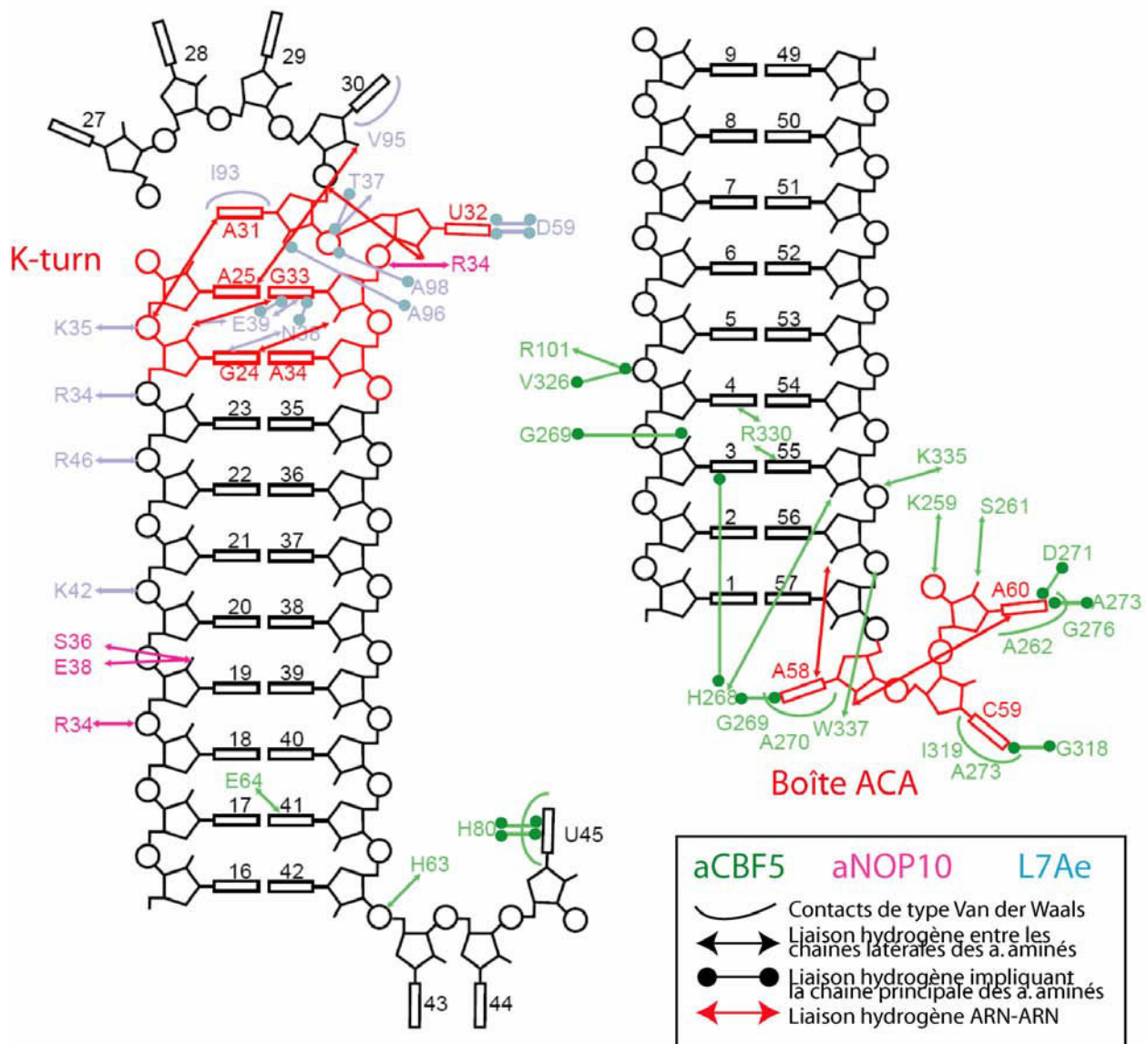


Figure III.24 : Représentation schématique des interactions ARN-protéines du complexe sRNA-aCBF5-aNOP10-L7Ae-aGAR1 (d'après Li et Ye, 2006).

Les nucléotides du sRNA qui ne contactent pas de résidus des protéines et qui sont désordonnés ne sont pas représentés.

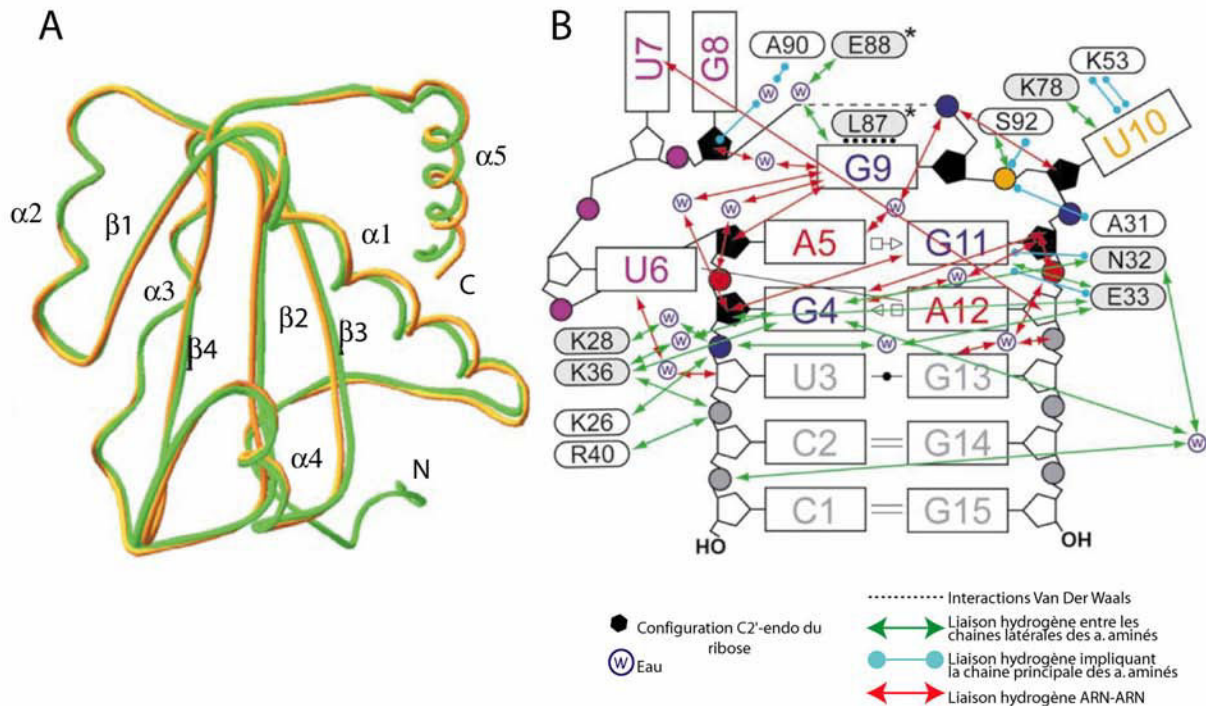


Figure III.25 : Structures des protéines L7Ae libre ou en interaction avec un motif K-turn (d'après Charron *et al.*, 2004 ; Hamma et Ferre-D'Amare, 2004).

A. Superposition des structures 3D des protéines L7Ae libre de *P. abyssi* (vert) et L7Ae lié au K-turn de *H. marismortui* (orange) (PDB : 1JJ2). **B.** Schématisation des interactions protéine-ARN et ARN-ARN mis en place lors de l'interaction de L7Ae de *M. jannaschii* et du motif K-turn de Afu190 (Tang *et al.*, 2002). Les astérisques indiquent des résidus qui ne sont conservés qu'au sein des archaea.

2.3. Interactions entre aCBF5 et aNOP10

aNop10 adopte une structure allongée et interagit sur toute une face avec le domaine catalytique de aCBF5 (Fig. III.22). Alors que les structures des domaines N-ter de aNOP10 et yNOP10 de levure diffèrent, il semblerait que les interactions avec aCBF5 ou yCbf5 soient de même nature et présentent une forte spécificité (Hamma et Ferre-D'Amare, 2006). Les résidus de aNOP10 qui interagissent avec aCBF5 proviennent des domaines N-ter et C-ter mais aussi de la région « linker » initialement désorganisée qui se trouve organisée par cette interaction (Fig. III.23). Il semblerait que l'interaction avec aNOP10 optimise le positionnement idéal de résidus de aCBF5 localisés à proximité du résidu Asp catalytique. Plusieurs réseaux d'interaction différents ont été proposés et nous avons contribué à la mise en évidence des acides aminés impliqués. Ces informations seront détaillées dans la partie « Résultats ».

Le potentiel électrostatique de surface du complexe aCBF5-aNOP10 est caractérisé par l'existence d'une ceinture de charges positives qui s'étend à partir du site actif vers à la fois une cavité formée par l'interaction avec aNOP10 et une cavité formée par le domaine PUA (Fig. III.26). Une petite zone supplémentaire (patch), également fortement chargée positivement, est présente à l'interface entre le domaine PUA et le domaine catalytique de aCFB5. Les résidus

responsables de cette région chargée sont conservés entre homologues d'archaea (Fig. III.26). Le positionnement de ces charges positives et les dimensions des cavités sont tels qu'elles peuvent permettre d'accommoder les hélices du sRNA, ainsi que les hélices formées lors de l'interaction du sRNA avec l'ARN substrat.

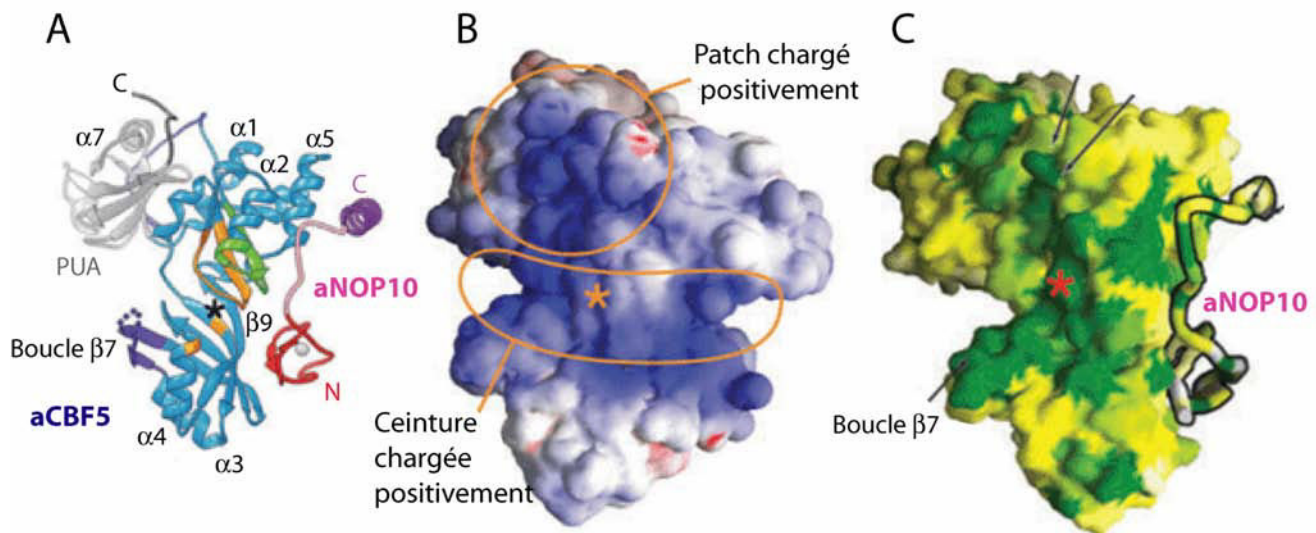


Figure III.26 : Potentiel électrostatique de surface du complexe aCBF5-aNOP10 de *M. jannaschii* (d'après Hamma *et al.*, 2005).

Représentation sous le même angle de la structure 3D globale du complexe aCBF5-aNOP10 (A), du potentiel électrostatique de surface (B) et de la conservation de séquence des résidus de aCBF5 et aNOP10 au sein des archaëa (C). Les surfaces représentées en bleu sont chargées positivement alors que celles représentées en rouge sont chargées négativement (B). La surface représentée en vert foncé correspond aux résidus fortement conservés alors que la surface représentée en blanc correspond à des résidus non conservés (C).

En produisant indépendamment chaque domaine de aNOP10, comme représenté sur la figure III.21, notre équipe a démontré que la partie N-ter est impliquée dans l'ancrage à aCBF5, elle-même associée à l'ARN guide tandis que la partie C-ter joue un rôle dans le recrutement de l'ARN cible (Manival *et al.*, 2006). Cependant la présence des deux fragments de aNOP10 ne rétablit pas l'action stimulatrice de la protéine entière et il en a été déduit et plus tard confirmé que la région « linker » joue un rôle important (Li et Ye, 2006 ; Manival *et al.*, 2006). En effet, alors que la partie N-ter de aNOP10 n'interagit qu'avec aCBF5, les parties C-ter et « linker » contactent L7Ae et la région « linker » interagit également avec le sRNA (Fig. III.23 et 24).

Enfin, notre équipe a suggéré une tétramérisation de l'hétéro-dimère aCBF5-aNOP10 via l'association de deux molécules de aNOP10 (Fig. III.27A) (Manival *et al.*, 2006). Deux types d'interactions entre ces deux molécules ont été observés. Le premier mode d'interaction implique les résidus Tyr41, Leu48 et Leu52 des deux hélices α en C-ter de aNOP10 de *P. abyssi* qui sont alors parallèles (Fig. III.27 panneaux B et D). Le deuxième mode d'interaction implique seulement des contacts entre les résidus Tyr41 et Tyr44 des extrémités N-ter des deux hélices qui sont alors perpendiculaires (Fig. III.27 panneaux C et E). Les résultats des expériences de mutagenèse sont en faveur du premier mode d'interaction (Manival *et al.*, 2006). Il est à noter cependant, que les

résidus Tyr41, Tyr44 et Leu/Trp48 sont impliqués dans des interactions avec L7Ae (Fig. III.23). Par conséquent, il est difficile d'envisager une dimérisation de aNOP10 au sein d'un complexe avec aCBF5 et L7Ae.

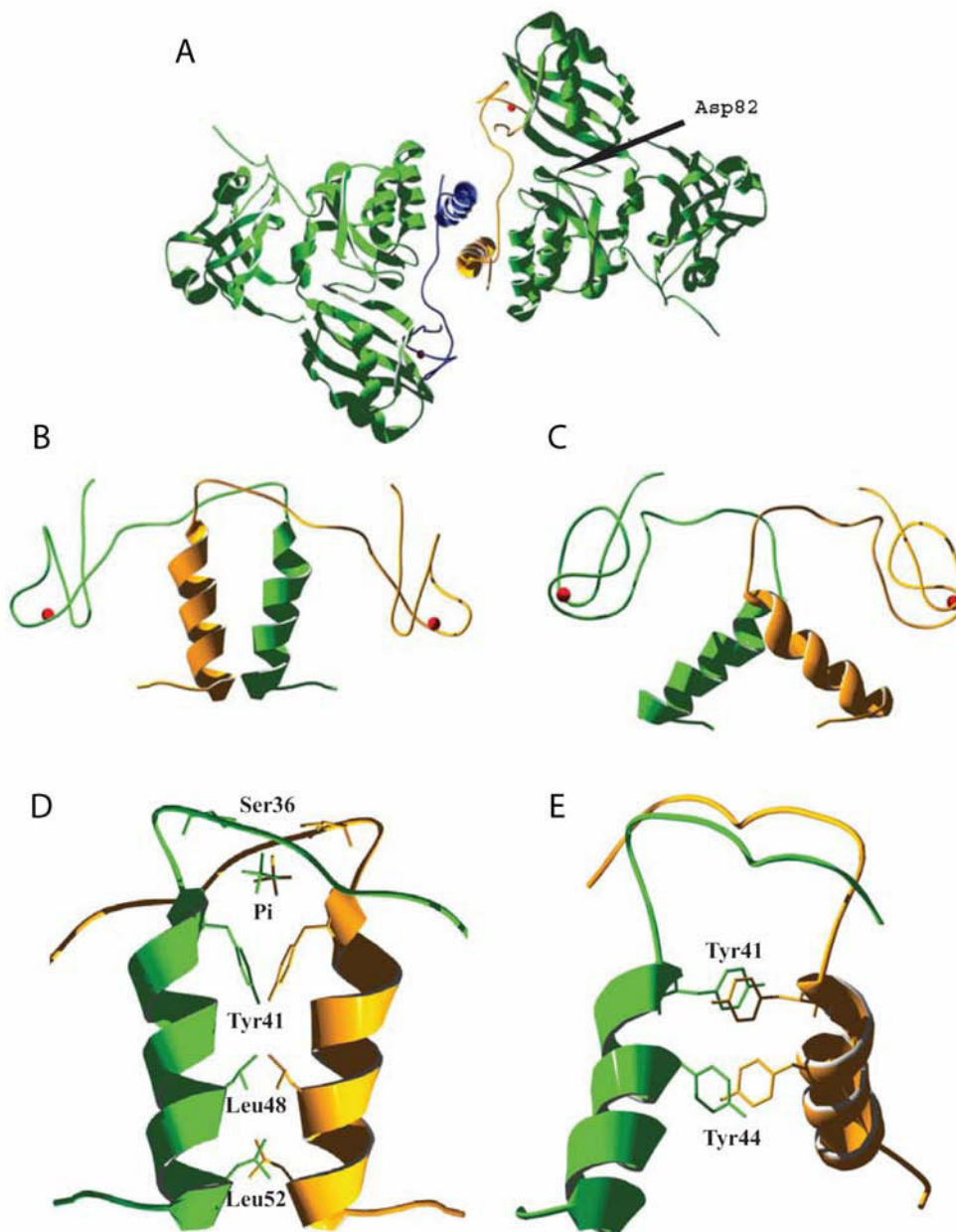


Figure III.27 : Hétéro-tétramères de aNOP10-aCBF5 de *P. abyssi* observés dans la structure cristalline (d'après Manival *et al.*, 2006).

A. Représentation de la structure 3D de l'hétéro-tétramère aNOP10-aCBF5. **B.** et **D.** ou **C.** et **E.** Représentation des deux types d'interactions différentes dans les dimères de aNOP10 présents dans le cristal.

2.4. Interactions entre aCBF5 et aGAR1

aGAR1 n'interagit qu'avec aCBF5 et pourrait jouer un rôle dans la modulation de la reconnaissance de la cible (Reichow *et al.*, 2007). En effet, aGAR1 interagit fortement avec la région $\beta 7$ de aCBF5. Une reconstitution par « docking » de la structure 3D du complexe sRNA

comportant les 4 protéines associées à un ARN cible montre que la boucle $\beta 7$ de aCBF5 est susceptible d'interagir avec l'hélice issue de l'interaction entre l'élément antisens 3' du sRNA et la séquence en 5' de l'uridine cible (Fig III.28) (Li et Ye, 2006 ; Rashid *et al.*, 2006). Ainsi, la boucle $\beta 7$ de aCBF5, dans une conformation dite « fermée », stabiliserait l'appariement entre le sRNA et sa cible. L'interaction de la boucle $\beta 7$ de aCBF5 avec aGAR1 aurait pour effet de maintenir une conformation « ouverte » favorable à la libération de l'ARN substrat et contribuerait ainsi à accélérer la formation du résidu Ψ (Fig III.23 et 28) (Charpentier *et al.*, 2005; Li et Ye, 2006 ; Rashid *et al.*, 2006).

2.5. Interactions de aNOP10 avec à la fois le sRNA et L7Ae

Récemment, la structure 3D du complexe du sRNA associé aux 4 protéines a démontré que aNOP10 n'interagit pas seulement avec aCBF5 mais aussi avec L7Ae et le guide sRNA (Fig. III.22, 23 et 24) (Li et Ye, 2006). Les interactions entre aNOP10 et L7Ae ne couvrent qu'une surface réduite mais impliquent des résidus conservés dans les protéines aNOP10 et L7Ae et qui sont situés dans les régions « linker » et C-ter de aNOP10 et de la région centrale ($\alpha 3$) de L7Ae. La région « linker » de aNOP10, déjà impliquée dans des interactions avec aCBF5 et L7Ae, établit donc également des contacts avec la tige apicale du sRNA. Un des trois résidus de NOP10 susceptibles d'interagir avec le sRNA, le résidu Arg34 est conservé dans la protéine NOP10 humaine et une mutation R34W serait responsable d'un type de Dyskératose Congénital Autosomale Dominante (Fig. III.23B) (Walne *et al.*, 2007). Ces observations soulignent le fait que aNOP10 pourrait jouer un rôle central dans ces particules. Etant donné que, dans la particule, l'affinité la plus forte est observée entre L7Ae et le K-turn (ou K-loop) du sRNA, il est probable que l'interaction avec aNOP10 constitue un des points d'ancrage de aCBF5.

2.6. Interactions entre aCBF5 et le guide sRNA

L'analyse structurale du domaine PUA de l'enzyme tRNA-guanine-transglycosylase (ArcTGT) a révélé une interaction avec la tige acceptrice et le trinuécléotide CCA des ARNt (Ferre-D'Amare, 2003 ; Ishitani *et al.*, 2003). Il a également été démontré, même si le mode de reconnaissance diffère entre les deux familles d'enzymes, que le domaine PUA de aCBF5 interagit avec le trinuécléotide ACA ainsi qu'avec le sillon mineur de l'hélice basale (P1) du sRNA (Fig. III.23 et 29) (nos résultats ; Li et Ye, 2006). La ressemblance entre le segment regroupant la tige basale et la boîte ACA avec le segment regroupant la tige acceptrice et la séquence CCA d'un ARNt sera discutée plus en détails dans la partie « Résultats ». La première et la troisième base de la boîte ACA interagissent spécifiquement avec des résidus du domaine PUA (Fig. III.29 panneaux A et B). La deuxième base n'est pas spécifiquement reconnue, ce qui explique pourquoi le résidu C n'est pas strictement conservé. L'enzyme aCBF5 interagit également, mais plus faiblement, avec l'hélice apicale et contacterait également les hélices formées entre l'élément antisens 3' du sRNA et l'ARN ciblé déjà discuté ci-avant (nos résultats ; Li et Ye, 2006 ; Rashid *et al.*, 2006).

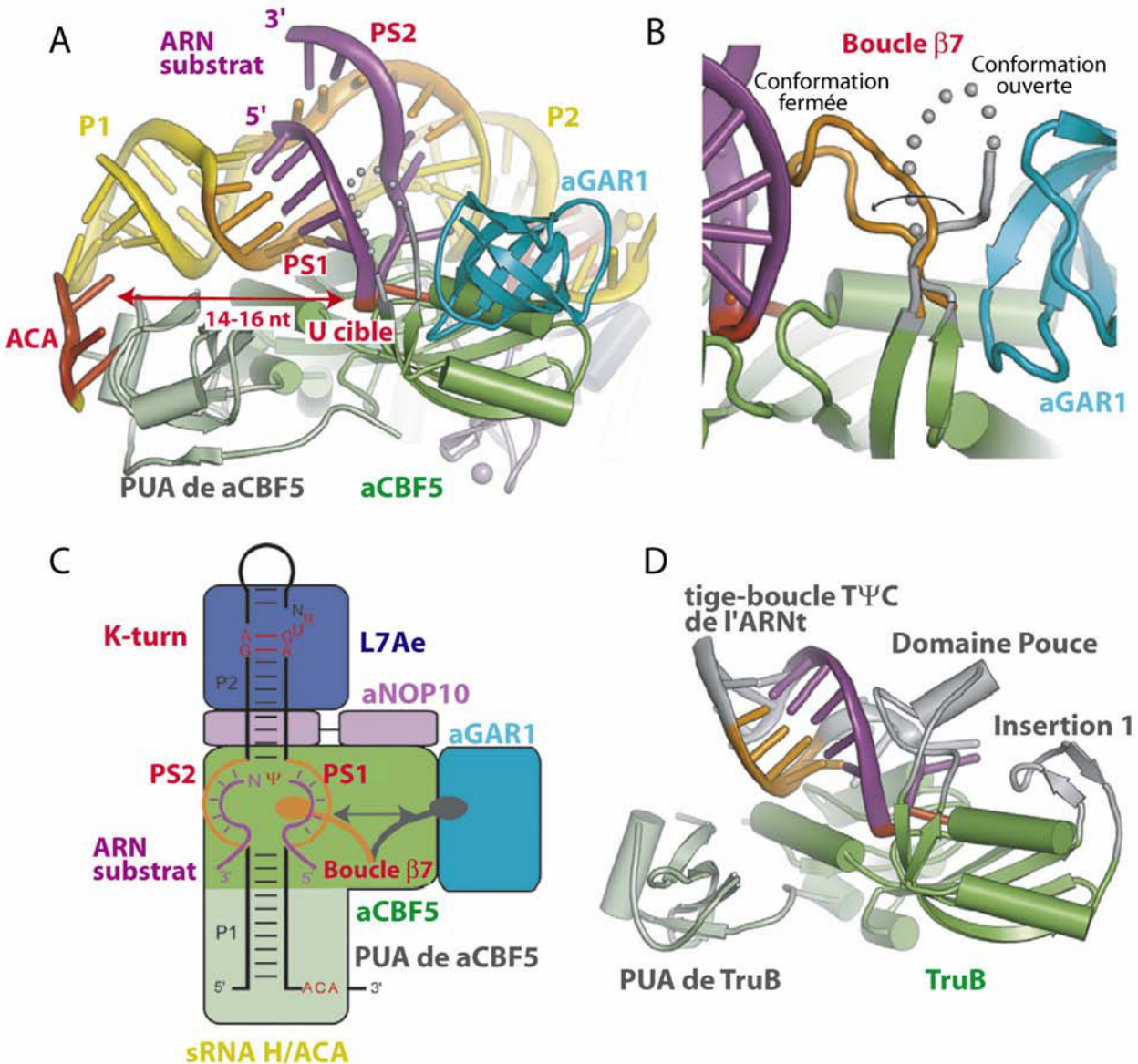


Figure III.28 : Rôle potentiel de aGAR1 dans la modulation de la reconnaissance de l'ARN cible (d'après Ye, 2007).

A. Modèle proposé pour l'association du complexe sRNP à boîtes H/ACA avec à un ARN substrat. **B.** La boucle $\beta 7$ mime l'action du domaine « pince » de TruB (**D**). Deux positionnements de la boucle $\beta 7$ semblent possibles : une conformation fermée (en orange) en interaction avec l'ARN substrat, ou une conformation ouverte (en gris) en association avec aGAR1. Dans ce dernier cas la boucle $\beta 7$ ne peut plus interagir avec l'ARN substrat. **C.** Représentation schématique de la structure du complexe sRNP à boîtes H/ACA précisant les deux conformations possibles de la boucle $\beta 7$. **D.** Structure 3D de TruB associé à la tige-boucle T Ψ C d'un ARNt. Le domaine « pince » de TruB est comparable à la boucle $\beta 7$ de aCBF5. L'insertion 1 pourrait jouer le même rôle que la protéine aGAR1.

Il est intéressant de remarquer que les positions dont les mutations sont responsables du syndrome de la Dyskératose Congénitale, qui sont localisées dans le domaine PUA de la dyskérine, ne correspondent pas à des résidus impliqués dans des interactions avec le sRNA d'après les données structurales sur le système archaea (Fig. III.23). Ceci suggère que ces positions pourraient soit être impliquées dans la reconnaissance des boîtes H soit jouer un rôle différent de ce qui est observé dans le système archaea. Par conséquent, seule l'étude des complexes snoRNP à boîtes H/ACA humain ou d'autres mammifères permettra d'expliquer le rôle de ces résidus.

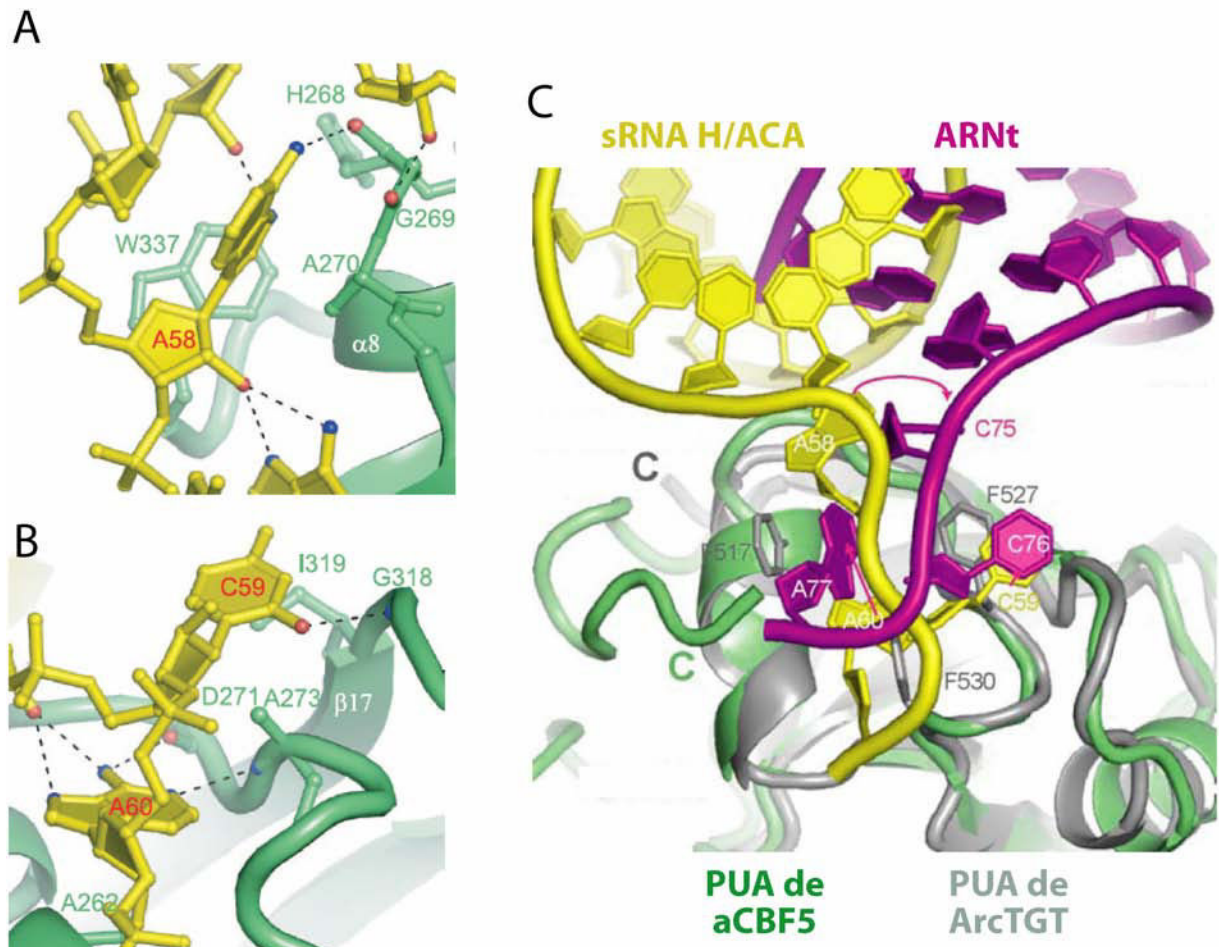


Figure III.29 : Reconnnaissance de la boîte ACA par le domaine PUA de aCBF5, comparaison à la reconnaissance du triplet CCA de la tige acceptrice d'un ARNt par le domaine PUA de Arc-TGT (Archaeosine tRNA-guanine transglycosylase) (d'après Li et Ye, 2006).

A. Interaction du résidu A58 (1^{er} résidu de la boîte ACA) avec les résidus du domaine PUA. **B.** Interaction des résidus C59 et A60 (2^{ème} et 3^{ème} résidus de la boîte ACA) avec les résidus du domaine PUA. **C.** Comparaison des interactions entre la boîte ACA d'un sRNA à boîtes H/ACA (en jaune) ou le triplet CCA terminal d'un ARNt (en violet) avec, respectivement, le domaine PUA de aCBF5 (en vert) ou le domaine PUA de ArcTGT (en gris).

2.7. Structure RMN d'un sRNA associé à l'ARN cible

La résolution par RMN de la structure en solution d'un snoRNA lié à un ARN substrat a été réalisée récemment par les équipes de J. Feigon et P. Moore (Jin *et al.*, 2007; Wu et Feigon, 2007).

Les résultats révèlent deux informations importantes qui sont applicables au système archaea. D'une part, dans de fortes concentrations utilisées pour les analyses RMN, le substrat peut se lier au sRNA en absence de protéine. D'autre part, l'hélice résultant de l'appariement de l'élément 5' antisens du sRNA et de la partie en 3' de l'uridine substrat, appelée P2S, s'empile de façon co-axiale avec l'hélice apicale du sRNA, appelée ici P2 ; alors que l'hélice résultant de l'appariement de l'élément 3' antisens du sRNA et de la partie en 5' de l'uridine substrat, appelée P1S s'empile de façon co-axiale avec l'hélice basale du sRNA, appelée ici P1 (Fig. III.30). On peut donc distinguer deux hélices discontinues qui imposent une courbure de 180° à l'ARN substrat qui prend une forme de U. Il en résulte deux jonctions triple-hélices caractéristiques (J1 et J2) qui positionnent l'uridine vers l'extérieur comme cela a été observé lors du mécanisme d'extrusion de bases par les ARN:Ψ-synthases. Il est à noter également que la situation au niveau de la jonction J2 mime exactement celle de la tige-boucle TΨC d'un ARNt en interaction avec TruB. Au final les 4 protéines ne seraient requises ni pour l'appariement de l'ARN substrat au sein du sRNA, ni pour l'extrusion de l'uridine.

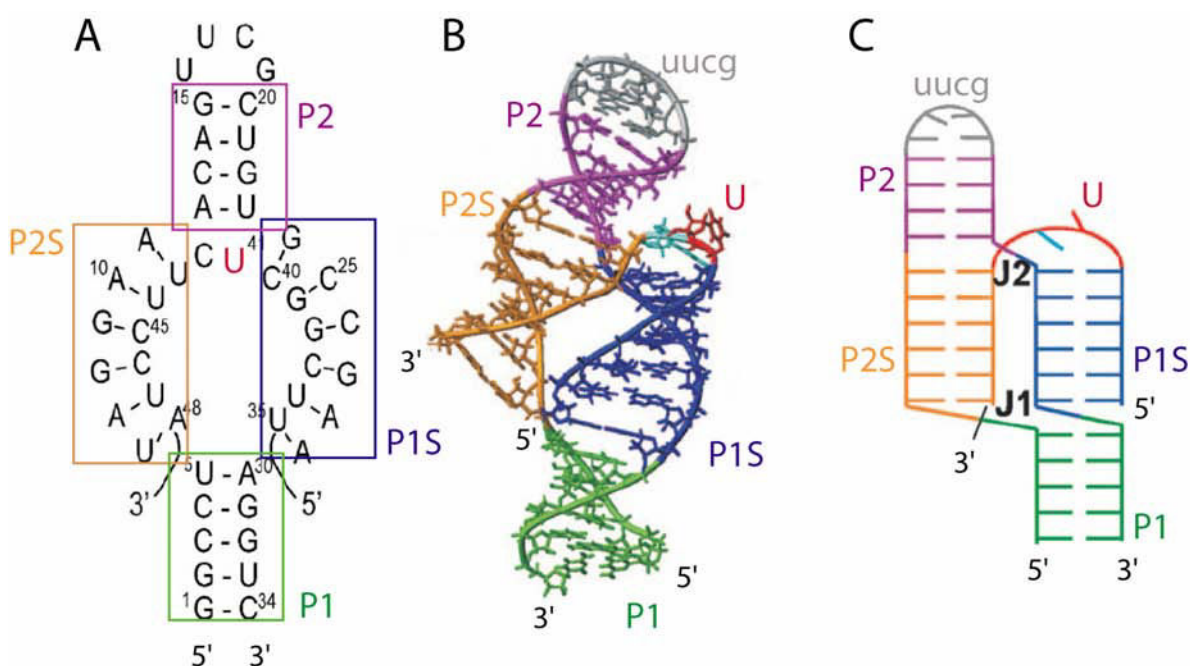


Figure III.30 : Etude par RMN de l'association du sRNA et de l'ARN substrat (d'après Wu et Feigon, 2007).

A. Structure secondaire du snoRNA U65hp associée à un fragment d'ARNr substrat (partie 4423-4435 de l'ARNr 28S humain). Le snoRNA U65hp correspond à la poche de pseudouridylation de la deuxième tige-boucle du snoRNA U65 comportant une partie des hélices basales (P1) et apicales (P2). Une boucle uucg a été placée au sommet de l'hélice apicale. **B.** Structure en solution du complexe U65hp associé au substrat ARN. Le résidu U substrat est représenté en rouge, le nucléotide en 3' du résidu U substrat est non apparié et représenté en bleu clair. Les hélices P1, P1S, P2S et P2 sont représentées respectivement en vert, bleu foncé, orange et rose. **C.** Schéma représentant l'empilement des hélices et la localisation des jonctions.

La fixation de aCBF5 ne reposerait donc pas sur le mécanisme d'extrusion de base conduisant au bon positionnement de l'uridine et du résidu catalytique. En effet, chez TruB, le positionnement du résidu catalytique par rapport à l'uridine cible repose sur l'interaction du résidu histidine, localisé deux résidus en amont de l'aspartate catalytique, avec la paire reverse Hoogsteen. Il apparaît donc que le positionnement de aCBF5 repose sur des interactions en partie différentes de celles de TruB. Les interactions réalisées, d'un côté via son domaine PUA avec la boîte ACA et l'hélice basale et de l'autre côté avec aNOP10 elle-même associée au sRNA et à L7Ae fortement fixée sur le sRNA permettent de fixer aCBF5 sur le guide H/ACA palliant peut-être une absence de liaison spécifique au niveau de la poche de pseudouridylation de aCBF5 seule.

3. Les snoRNP H/ACA eucaryotes

Les snoRNA diffèrent des sRNA d'archaea par l'absence de motifs en K-turn/K-loop, la présence d'une boîte H et leur architecture, le plus fréquemment, en deux structures tige-boucles. Les snoRNA eucaryotes sont capables de se lier à deux molécules de CBF5 (Watkins *et al.*, 1998 ; nos résultats). Or, la mutation de la boîte ACA terminale des snoRNA inhibe complètement la fixation de CBF5 (Baker *et al.*, 2005). Ceci suggère un couplage de la fixation de CBF5 sur chacune des structures tige-boucles et, certainement, des autres protéines associées à ces structures organisées en deux tige-boucles.

L'absence de reconnaissance spécifique des ARN par NHP2, l'association stable de NHP2 avec le complexe CBF5-NOP10 ainsi que la fixation directe de GAR1 sur le snoRNA sont trois autres différences déjà évoquées ci-avant qui ont conduit l'équipe de U. Meier à proposer un mécanisme d'association des particules snoRNP à boîtes H/ACA différent de celui qui est pressenti pour les archaea (Fig. III.31) (Wang et Meier, 2004 ; Darzacq *et al.*, 2006).

En effet, chez les mammifères, la formation du dimère NAP57-NOP10 est requise pour le recrutement de NHP2 sur le snoRNA. Il a également été démontré chez les eucaryotes que GAR1 est en compétition avec une autre protéine, NAF1, qui comporte d'importantes similarités structurales avec GAR1 (Leulliot *et al.*, 2007). NAF1 peut aussi s'associer au dimère NAP57-NOP10. Elle s'associerait précocement au cours de la biosynthèse des particules et maintiendrait les particules sous forme inactives dans les premières étapes de biogenèse et de leur transport dans les corps de Cajal (Darzacq *et al.*, 2006 ; Kittur *et al.*, 2006 ; Leulliot *et al.*, 2007). Par conséquent, un complexe protéique contenant NOP10, NAP57, NHP2 et NAF1 formé dans le cytoplasme pourrait être importé dans le noyau par l'intermédiaire de la protéine NAF1 avant que GAR1 remplace NAF1 (Fig. III.31). La protéine SHQ1 est comme NAF1 essentielle à la biogenèse des particules snoRNP à boîtes H/ACA de levure mais ne fait pas partie de la particule mature (Dez *et al.*, 2002 ; Fatica *et al.*, 2002 ; Yang *et al.*, 2002). L'interaction de NAF1 avec le domaine CTD de l'ARN pol II suggère que ces deux protéines pourraient recruter les protéines CBF5 et NOP10 sur un transcrit snoRNA naissant (Fig. III.31). Une interaction possible de SHQ1 avec RNT1, RNase double-brin homologue de la RNase III bactérienne qui clive en amont et en aval des pré-snoRNA (Chanfreau *et al.*, 1998), suggère aussi une implication de SHQ1 dans le couplage maturation/assemblage des snoRNP H/ACA. Par ailleurs, les interactions entre GAR1 et RNT-1

sont importantes pour la localisation nucléolaire de GAR1 (Tremblay *et al.*, 2002). La protéine SMN (Survival of Motor Neuron) absente de *S. cerevisiae* mais présente chez de nombreux eucaryotes interagit avec GAR1 via ses répétitions GAR et pourrait également être impliquée dans l'assemblage des particules snoRNP H/ACA en tant que protéine chaperonne (Pellizzoni *et al.*, 2001 ; Terns et Terns, 2001).

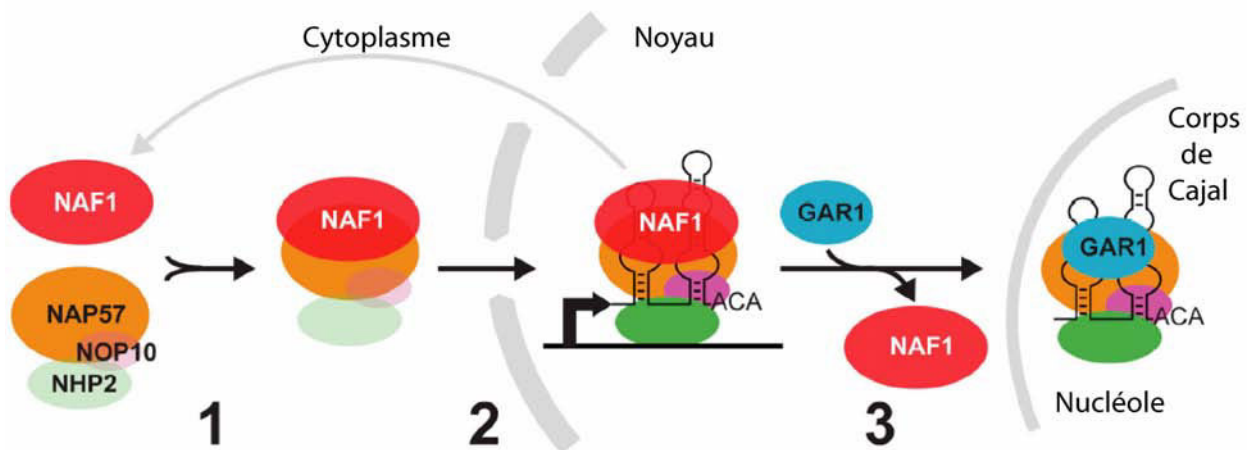


Figure III.31 : Modèles d'assemblage des snoRNP d'eucaryote (d'après Darzacq *et al.*, 2006).

Le modèle a été proposé pour l'assemblage des snoRNP H/ACA chez l'homme. **1.** La protéine NAF1 s'associe dans le cytoplasme avec la protéine NAP57/Dyskérine/CBF5 déjà liée aux protéines NOP10 et NHP2. **2.** Le complexe est recruté sur l'ARN en cours de transcription. **3.** La protéine NAF1 serait remplacée par la protéine GAR1 pour former la snoRNP H/ACA mature dans les corps de Cajal.

Enfin, les études menées par E. Bertrand et T. Kiss suggèrent que la biogenèse et le transit dans différentes régions du noyau fait intervenir un grand nombre de protéines autres que celles contenues dans les particules matures. Malgré ces différences, la conservation entre archaëes et eucaryotes de nombreux résidus importants pour les interactions présentées ci-avant suggère un type d'association de CBF5-NOP10-GAR1 similaire au sein des particules actives des deux domaines du vivant.

Récemment, deux analyses réalisées indépendamment par RMN des deux composants eucaryotes solubles, le snoRNA U65hp humain et la protéine NOP10 de *S. cerevisiae*, ainsi que des expériences de cartographie par modification chimique des interactions entre ces deux composants ont suggéré l'interaction directe de la protéine Nop10 de levure avec la région constituée de la jonction entre la tige basale (P1) et l'élément antisens 3' de la poche de pseudouridylation du snoRNA (Khanna *et al.*, 2006). La protéine Nop10p interagirait par son extrémité N-ter qui, contrairement à aNOP10 d'archaëes ne présente aucun doigt à Zinc. Les auteurs proposent que cette interaction permette une ouverture de cette poche facilitant l'hybridation de l'ARN cible. Cependant, il reste envisageable que les protéines NOP10, CBF5 et NHP2 se lient au snoRNA sous forme de complexes préformés. Une première analyse de l'association de CBF5 sur le snoRNA a

été réalisée grâce à une protéine Cbf5p de levure tronquée (Normand *et al.*, 2006). En effet, le domaine KKE/D en C-ter (138 aa des 483 aa) dont la fonction inconnue n'est pas essentielle pour la survie cellulaire a été retiré ce qui rend soluble la protéine recombinante Cbf5p. Cette étude a révélé que CBF5 s'associe au snoRNA d'une manière similaire à ses homologues d'archaea, ceci en interagissant principalement avec des résidus des boîtes H et ACA et de l'hélice basale (P1).

4. Attribution des sno/sRNA à boîtes H/ACA aux résidus Ψ détectés

4.1. Situation dans les snRNA

Alors qu'un seul snoRNA à boîtes H/ACA (snR81) a été décrit comme responsable de la modification d'un résidu Ψ d'un snRNA chez *S. cerevisiae* (la position 42 dans le U2) (Ma *et al.*, 2005), de nombreux guides appelés scaRNA sont responsables de nombreuses pseudouridylation dans les snRNA d'eucaryotes supérieurs. Ceux-ci ont déjà été décrits dans la partie II.

4.2. Situation dans les ARNt

Contrairement aux sRNA à boîtes C/D d'archaea, aucun sno/sRNA à boîtes H/ACA d'eucaryotes ou d'archaea modifiant un ARNt n'avait été mis en évidence. Cependant, nos travaux récents, présentés dans la partie « Résultats », prouvent qu'une famille de sRNA à boîtes H/ACA présent chez les crénotes est responsable de la modification de la position 35 des ARNt^{Tyr}.

4.3. Situation dans les ARNr

La caractérisation exhaustive des snoRNA à boîtes H/ACA de la levure *S. cerevisiae* a été réalisée récemment par deux approches, une approche bioinformatique et une approche expérimentale. L'approche bioinformatique s'est basée sur la description des snoRNA à boîtes H/ACA existants chez les levures ainsi que sur les positions précises des résidus Ψ détectés dans les ARNr de *S. cerevisiae* (Schattner *et al.*, 2004 ; Schattner *et al.*, 2005). L'approche expérimentale a consisté à produire les protéines communes Gar1p et Nhp2p sous forme fusionnée avec une étiquette TAP (tandem-affinity purification). Cette fusion a permis de réaliser une purification en tandem des deux protéines et des ARN associés. Les ARN associés ont été marqués puis hybridés sur une puce à ADN contenant des fragments recouvrant l'ensemble de l'ADN génomique de *S. cerevisiae* sous forme de fragments (Torchet *et al.*, 2005). Ces études complémentaires ont révélé que les 44 pseudouridines présentes dans les ARNr de *S. cerevisiae* sont guidées par 29 snoRNA à boîtes H/ACA. Il s'agit de la recherche la plus aboutie pour un génome eucaryote. Ainsi, ces résultats suggèrent que toutes les pseudouridylation formées dans les ARNr des eucaryotes le seraient par des systèmes guidés par des snoRNA à boîtes H/ACA et qu'aucune ARN: Ψ -synthase fonctionnant seule ne serait capable de reconnaître et modifier une position d'un ARNr.

La situation dans le domaine archaea n'est pas aussi claire puisqu'aucune recherche systématique, expérimentale ou bioinformatique, de gènes de sRNA à boîtes H/ACA dans les

espèces archaea n'avaient été réalisée avant nos travaux. Notre contribution à ce domaine sera décrite dans la partie « Résultats ».

5. Rôle des résidus Pseudouridines dans les ARN

A. Considérations générales

Compte tenu de l'abondance et de la présence dans les trois domaines du vivant des résidus Ψ ainsi que de la quantité d'information génétique qui est nécessaire à leur formation il est raisonnable de penser que les résidus Ψ jouent un rôle important dans l'ARN. Les conséquences locales de la présence d'un résidu Ψ , notamment les possibilités de rigidification du squelette ribose-phosphate par formation d'une liaison hydrogène supplémentaire qui stabilise les hélices (Auffinger et Westhof, 1998 ; Davis, 1998) ont déjà été évoquées au début de ce chapitre. Cependant, la fonction cellulaire de ces résidus Ψ reste encore peu comprise (pour revue, Charette et Gray, 2000).

Les tests expérimentaux visant à définir ce rôle reposent soit sur la comparaison *in vitro* des propriétés de molécules déficientes en résidus modifiés soit sur la comparaison *in vivo* des phénotypes de micro-organismes dont le système de modification responsable de la formation d'un ou plusieurs résidus Ψ a été inactivé. Ces études ont permis de mettre en évidence une influence de modifications des ARNt et ARNr sur la fidélité et l'efficacité de la traduction, mais aussi leur intervention dans des processus tels que l'initiation de la transcription inverse de l'ARN du virus HIV-1 (pour revue, Marquet, 1998). En effet, des souches mutées pour lesquelles certaines pseudouridines sont absentes montrent des défauts de traduction, une vitesse de croissance ralentie et sont supplantées par la souche sauvage lors d'expérience de compétition. Ceci suggère que les pseudouridylation apportent un avantage sélectif important dans un contexte biologique particulier.

Notre travail de thèse concerne principalement les résidus Ψ localisés dans les ARNr ainsi que trois résidus Ψ localisés dans les ARNt. Par conséquent, la partie suivante sera essentiellement consacrée au rôle connu ou envisageable des résidus Ψ dans les ARNr et dans les ARNt. Les fonctions potentielles de ces pseudouridylation seront encore discutées lors de la présentation de nos résultats.

B. Rôle dans les ARNt

L'absence de résidus Ψ dans les ARNt semble le plus souvent ne pas affecter la viabilité de la cellule (Kammen *et al.*, 1988 ; Simos *et al.*, 1996 ; Becker *et al.*, 1997a ; Lecointe *et al.*, 1998 ; Motorin *et al.*, 1998). En effet, seule la délétion des gènes codant TruA et Pus3 responsables de la formation des résidus Ψ en positions 38, 39 et 40 chez respectivement, *E. coli* et *S. cerevisiae* entraîne une diminution de la vitesse de croissance (Carbone *et al.*, 1991 ; Tsui *et al.*, 1991 ; Lecointe *et al.*, 1998). Par approche génétique il a été observé qu'aucune autre délétion d'ARN: Ψ -

synthase responsable de pseudouridylation dans les ARNt n'engendre un phénotype de croissance. Les résidus Ψ ne sont pas non plus essentiels pour la réaction d'amino-acylation. Les résidus Ψ affectent les structures locales, mais non globales, des régions dans lesquelles ils résident. L'influence des résidus Ψ sur la structuration a été mesurée dans de nombreuses études en comparant les caractéristiques structurales d'ARNt produits par transcription *in vitro* et donc dépourvus de modifications post-transcriptionnelles, à celles d'ARNt naturels par différentes méthodes biophysiques : la fusion thermique, la mesure du rayon de Stokes par gel filtration, le dichroïsme circulaire, l'analyse par sondes chimiques ou enzymatiques de la structure en solution, la diffraction des rayons X, la RMN ou bien encore la spectrofluorimétrie (Perret *et al.*, 1990 ; Chen *et al.*, 1993 ; Derrick et Horowitz, 1993 ; Arnez et Steitz, 1994 ; Yue *et al.*, 1994 ; Serebrov *et al.*, 1998). Au final, ces études ont révélé que la présence des nucléotides modifiés stabilise les structures secondaire et tertiaire des ARNt. Cet effet résulterait de la stabilisation de l'ancrage des ions Mg^{2+} au sein des ARNt et d'une réorganisation de la répartition spatiale des molécules d'eau à l'intérieur et à proximité des molécules d'ARNt comme décrit ci-dessus (pour revue, Agris, 1996). La présence de résidus Ψ au niveau des jonctions entre région simple-brin et région double-brin jouerait un rôle dans la stabilisation des hélices d'ARN (pour revue, Auffinger et Westhof, 1998).

Les paragraphes suivants détaillent plus précisément les rôles spécifiques attribués aux résidus Ψ présents dans différentes régions des ARNt.

1. Résidu Ψ en position 13

Ce résidu modifié est présent dans des ARNt des trois domaines du vivant. Sa fonction n'est pas établie mais il a été proposé que ce résidu renforce la stabilité de l'appariement non Watson-Crick Ψ -G ou Ψ -U entre les positions 13 et 22 du bras D. Notons que le nucléotide en position 22 est souvent impliqué dans un appariement triple avec la base en position 46 de la boucle variable (pour revue, Auffinger et Westhof, 1998).

2. Résidu Ψ en position 27

La paire de bases 27-43 constitue la jonction entre le bras de l'anticodon et le reste de la structure de l'ARNt. La fréquence d'apparition d'une paire de bases Ψ -A à ces positions est de l'ordre de 20% tandis que la paire de bases U-A n'est présente que dans 11% des cas (les paires G-C et C-G représentent 44% des appariements, les autres 25%). La pseudouridylation de la position 27 serait nécessaire pour stabiliser la paire de base U-A énergétiquement moins favorable (pour revue, Auffinger et Westhof, 1998). Schultz et Yarus (1994) ont étudié l'ARNt su7G36, cet ARN dérive de l'ARNt^{Trp}, la différence est qu'il possède un anticodon de séquence CUG, anticodon complémentaire du codon glutamine CAG. Cet ARN a la particularité de pouvoir se comporter en tant qu'ARNt supprimeur. En effet, la formation d'une paire de bases G-U au niveau de la première position du codon permet la lecture d'un codon UAG (ambre) de terminaison. Cette activité supprimeur peut être mesurée précisément et il a été noté que la substitution de la paire de bases C27-G43 par des paires de bases A-U, U-A ou G-C avait pour conséquence une diminution

de l'activité suppresseur de l'ARNt su7G36. Cette activité peut être restaurée en utilisant une paire de bases Ψ -A (Schultz et Yarus, 1994a ; Schultz et Yarus, 1994b). Il apparaît donc qu'aux positions les plus fréquemment pseudouridylées, une paire de bases Ψ -A ou A- Ψ peut être considérée comme une alternative pour remplacer une paire de bases C-G ou G-C.

3. Résidu Ψ en position 35

Ce résidu peut également jouer un rôle dans la stabilisation du bras anticodon et participe donc à l'amélioration de la fidélité de lecture du codon en stabilisant les interactions anticodon-codon (pour revue, Auffinger et Westhof, 1998). Il a également été vérifié expérimentalement que ce résidu augmente l'activité suppresseur de terminaison de l'ARNt^{Tyr} (U Ψ A) de *S. cerevisiae* en augmentant la stabilité de l'interaction anticodon-codon (Johnson et Abelson, 1983). Par ailleurs, la présence de ce résidu dans les ARNt^{Tyr} (G Ψ A) cytoplasmiques des plantes est obligatoire pour une reconnaissance efficace des codons stop UAG ou UAA et pour la discrimination entre codons UAA et UGA puisque l'anticodon GUA ne peut s'apparier aux codons stop UAA et UAG (Zerfass et Beier, 1992). Il est également concevable que ce résidu modifie le taux d'aminacylation. En effet, il a été démontré que la substitution Ψ 35 en U35 augmente de deux fois le K_m de l'aminacylation (Bare et Uhlenbeck, 1986).

4. Résidu Ψ en position 55

Ce résidu est présent dans plus de 90% des ARNt (soit la fréquence de résidu Ψ la plus élevée) et est à l'origine de la dénomination « bras T Ψ C ». Des travaux anciens ont montré un défaut de croissance des souches de *E. coli* pour lesquelles le gène de TruB a été délété (pour revue, Charette et Gray, 2000). Néanmoins, l'absence du gène homologue Pus4 chez *S. cerevisiae* ne se traduit par aucun phénotype de croissance (Becker *et al.*, 1997a). De plus, l'absence de ce résidu n'affecte ni l'aminacylation, ni la traduction *in vitro*. Ce résidu semble donc n'intervenir que dans la structuration de l'ARNt (pour revue, Auffinger et Westhof, 1998). Le groupement imino N3-H du résidu Ψ 55 est connu pour interagir avec un atome d'oxygène anionique du résidu 58. De plus, le proton du groupement imino N1-H du résidu Ψ 55 interagit avec une molécule d'eau, elle-même en interaction avec des atomes d'oxygène anioniques adjacents (Westhof *et al.*, 1988). Il apparaît donc que ce résidu pseudouridine universellement conservé est « bloqué » dans une conformation particulière de par les liaisons hydrogène qu'il établit et il pourrait contribuer ainsi à la stabilité de la boucle T Ψ C. Par ailleurs, ce résidu est également impliqué dans une interaction avec le nucléotide G17 de la boucle D de l'ARNt^{Asp} de *S. cerevisiae*.

5. Autres résidus Ψ des ARNt

Il est proposé que les résidus Ψ 34 et Ψ 36 renforcent les appariements entre codon et anticodon et particulièrement les interactions impliquant les paires de bases Wobble G-A et G-G (Davis *et al.*, 1998). Il apparaît également que la nature du nucléotide présent en position 34 a une influence sur

la réaction d'aminacylation et qu'un résidu Ψ à cette position est un déterminant de spécificité d'aminacylation (Senger *et al.*, 1997).

Les résidus Ψ 38, Ψ 39 et Ψ 40 se révèlent également importants pour l'aminacylation, probablement du fait qu'ils jouent un rôle dans la stabilisation de la structure du bras anticodon (Yarian *et al.*, 1999). Il semble également que la présence des résidus Ψ 38 et Ψ 39 chez la levure favorise la translecture de certains codons par les ARNt non-sens naturels, ainsi que le décalage du cadre de lecture au niveau de séquences « glissantes » (Lecoïnte *et al.*, 2002).

La présence de résidus Ψ en position 32 ou 38 contribue à la stabilisation de l'interaction 32-38. Les structures 3D d'ARNt obtenues par radiocristallographie révèlent qu'il s'agit en fait d'une pseudo-paire de bases avec une seule liaison hydrogène pour l'ARNt^{Asp} de levure (Westhof *et al.*, 1985) pour l'ARNt^{Phe} de levure (Quigley et Rich, 1976) et pour l'ARNtGln d'*E. coli* (Rould *et al.*, 1991). Cette interaction allonge l'hélice de la tige anticodon, et le résidu Ψ par la stabilisation qu'il apporte est impliqué dans la discrimination des codons par l'ARNt^{Gly} d'*E. coli* (Lustig *et al.*, 1993 ; Claesson *et al.*, 1995). Le résidu Ψ 32 et le résidu Ψ 39 pourraient, en synergie obligatoire avec d'autres modifications du bras anticodon, contribuer à la stabilisation du motif en U-turn de la boucle anticodon, modulant ainsi la lecture du codon (Tworowska et Nikonowicz, 2006).

Ainsi, le rôle des résidus Ψ localisés dans le bras anticodon ou au sommet de celui-ci serait principalement de stabiliser la structure de la boucle anticodon pour améliorer la lecture du codon. Les résidus Ψ mais aussi les nombreuses autres modifications dans la boucle anticodon permettent de moduler la reconnaissance du codon, favorisant ainsi l'appariement de l'anticodon à plusieurs codons synonymes, la suppression de la terminaison ou le décalage du cadre de lecture. Les autres résidus Ψ stabilisent des motifs structuraux spécifiques dans les autres régions de l'ARNt. Tous ces résidus Ψ peuvent aussi moduler les interactions avec l'ARNr, l'amino acyl ARNt synthétase ou d'autres facteurs protéiques (Eshete *et al.*, 2007).

Ainsi, les résidus Ψ sont impliqués dans de nombreuses fonctions des ARNt et contribuent à préserver la fidélité du processus de traduction. Leurs interventions sont cruciales ou subtiles, indépendantes ou concertées, et peuvent conférer un avantage à la cellule qui les contient.

C. Rôle dans les ARNr

En comparaison avec les ARNt, peu d'informations sur la relation structure-fonction des résidus Ψ des ARNr sont disponibles. Les fonctions des résidus Ψ ont donc principalement été proposées à partir d'informations concernant leurs rôles connus dans d'autres contextes tels que les ARNt évoqués précédemment, leur localisation dans des régions fonctionnelles et leur distribution au sein de plusieurs espèces étudiées.

Une première approche pour aborder l'étude de la fonction des modifications des ARNr a consisté à analyser les caractéristiques de ribosomes assemblés *in vitro* à partir d'ARNr non modifiés ou partiellement modifiés. Il s'est avéré que les grandes sous-unités de trois espèces bactériennes, reconstituées à partir de transcrits dépourvus de modifications, catalysent encore la synthèse de liaisons peptidiques. Cependant, ces ribosomes présentent une activité moindre en

comparaison aux sous-unités reconstituées avec des ARNr naturels (Green et Noller, 1996 ; Green et Noller, 1999 ; Khaitovich *et al.*, 1999). En dépit du nombre parfois élevé de résidus Ψ présents dans un ARNr, il semble qu'ils ne soient pas indispensables (pour revue, Ofengand *et al.*, 2001a). Par conséquent, les résidus Ψ confèreraient un avantage sélectif, mais ils ne seraient pas essentiels. Cette hypothèse est en accord avec le fait que certaines espèces présentent peu de résidus Ψ , principalement des espèces bactériennes et d'archaea, et que ces positions ne sont pas strictement conservées dans les trois domaines du vivant.

Chez *E. coli*, l'approche visant à montrer une importance fonctionnelle des résidus Ψ a consisté dans un premier temps à interrompre individuellement les gènes codant les ARN: Ψ -synthases responsables de la formation de résidus Ψ dans les ARNr, et à mesurer d'éventuels effets sur la croissance. Une série d'études analogues par délétions des snoRNA à boîtes H/ACA de *S. cerevisiae* a également été réalisée. La délétion de 6 des 7 ARN: Ψ -synthases présentes chez *E. coli* n'a aucun effet détectable sur la croissance cellulaire, ceci pour des tests effectués à plusieurs températures (Conrad *et al.*, 1998 ; Conrad *et al.*, 1999 ; Raychaudhuri *et al.*, 1999 ; Del Campo *et al.*, 2001). Seule la perte de l'ARN: Ψ -synthase RluD, responsable de la formation des résidus Ψ 1911, Ψ 1915 et Ψ 1917 dans l'ARNr 23S, provoque un défaut de la croissance (Raychaudhuri *et al.*, 1998 ; Gutgsell *et al.*, 2001). Cependant, dans ce cas, il a été prouvé que ce défaut de croissance peut être corrigé par l'apport d'un plasmide portant le gène codant RluD muté au niveau de l'acide aspartique catalytique. Ceci suggère que la protéine RluD est essentielle pour une croissance normale, mais pas son activité ARN: Ψ -synthase (Gutgsell *et al.*, 2001). Des études complémentaires ont montré que le défaut de croissance observé en absence de RluD est une conséquence directe d'un mauvais assemblage de la sous-unité 50S du ribosome, suggérant une implication de RluD dans le processus d'assemblage du ribosome, peut-être en tant que protéine chaperonne (Gutgsell *et al.*, 2001). Des délétions simultanées de plusieurs gènes ont été réalisées pour empêcher la formation de résidus Ψ appartenant à une même région. Ainsi, la délétion des gènes codant les protéines RluB, RluC, RluE et RluF qui abolit la formation de 6 résidus Ψ situés à proximité du centre peptidyl transférase (PTC) conduit à l'observation de certains défauts métaboliques (pour revue, Ofengand, 2002). Il a donc été suggéré que les résidus Ψ n'ont pas d'effet indépendant mais contribuent chacun, partiellement et de manière coopérative, à la stabilisation d'une conformation particulière de l'ARNr. De même, chez *S. cerevisiae*, la délétion individuelle de 5 snoRNA à boîtes H/ACA (snR10, snR34, snR37, snR42 et snR46) responsables de la formation de 6 résidus Ψ du centre PTC n'a pas d'effet sur la croissance (Fig. III.32) (Ofengand et Fournier, 1998), alors que la délétion combinée des 5 guides entraîne une réduction de la vitesse de croissance et du taux de synthèse protéique, une hypersensibilité aux antibiotiques et des anomalies dans la structuration de l'ARNr de la grande sous-unité (King *et al.*, 2003). Cependant, les analyses de l'effet de la perte d'expression de l'ARN snR10, responsable de la modification de la formation du résidu Ψ en position 2919, sont compliquées par le fait que ce snoRNA est également impliqué dans la maturation de l'ARNr de la petite sous-unité (Tollervey, 1987). Pour distinguer ces deux effets, un nucléotide de l'élément antisens 3' de snR10 a été muté

pour inhiber l'appariement du guide à sa cible (snR10 Δ C) (King *et al.*, 2003). Il a ainsi été observé que l'absence de formation du résidu Ψ 2919 aurait pour conséquences un profil de polysomes altéré et un défaut de synthèse protéique sans pour autant altérer la vitesse de croissance cellulaire alors que l'absence d'expression de snR10 entraîne une diminution plus forte de la synthèse protéique. Ainsi, la diminution de la croissance et l'hypersensibilité aux antibiotiques sont associées à la perte de snR10 et non à la perte du résidu Ψ 2919. Il est à noter que, dans la souche pour laquelle les 5 ARN guides ne sont pas exprimés, les défauts de synthèse protéique sont encore plus marqués, ce qui suggère une influence de la présence d'autres snoRNP dans la formation des ribosomes actifs.

L'élimination globale des résidus Ψ des ARNr de *S. cerevisiae* a été réalisée par des mutations de CBF5, mais l'interprétation des résultats est rendue difficile par l'existence d'un grand nombre d'interactions entre CBF5 et des protéines participant à la ségrégation des chromosomes, des kinases spécifiques de la méiose, des facteurs transcriptionnels de l'ARN pol. I, hTR le composant ARN de la télomérase et des snoRNA à boîtes H/ACA impliqués dans la maturation des ARNr (pour revue, Charette et Gray, 2000). Il n'est donc pas surprenant d'observer que l'absence d'expression de CBF5 est létale pour la cellule (Jiang *et al.*, 1993 ; Lafontaine *et al.*, 1998a). Des souches produisant une protéine CBF5 mutée au niveau de l'aspartate catalytique, ce qui abolit la formation de l'ensemble des résidus Ψ dans les ARNr, survivent mais présentent une sévère inhibition de croissance (Zebarjadian *et al.*, 1999).

Un rôle possible des résidus Ψ dans la structure du site peptidyl transferase (PTC) a été analysé par des modifications au DMS de la région de l'ARNr 23S de *S. cerevisiae* contenue dans ce site et provenant de souches n'exprimant pas les 5 snoRNA responsables des 6 résidus Ψ situés dans cette région (Fig. III.32). Des sites additionnels de réactivité au DMS apparaissent dans la souche mutante démontrant ainsi que les résidus Ψ du PTC influencent la conformation de cette région fonctionnelle de l'ARNr 23S (King *et al.*, 2003). L'étude de la structure 3D de la grande sous-unité ribosomique (Ban *et al.*, 2000) révèle qu'aucun de ces 6 résidus Ψ n'est localisé à moins de 13 Å du site de formation de la liaison peptidique et donc à priori aucun de ces résidus ne peut avoir une influence directe sur la formation de cette liaison (Ban *et al.*, 2000). Les données de cette étude sont donc en défaveur de l'hypothèse selon laquelle la fonction imino N1-H d'un résidu Ψ de cette région serve d'accepteur de l'extrémité C-terminale de la chaîne peptidique en formation, avant sa réaction avec l'acide aminé suivant (Lane *et al.*, 1992; Lane *et al.*, 1995).

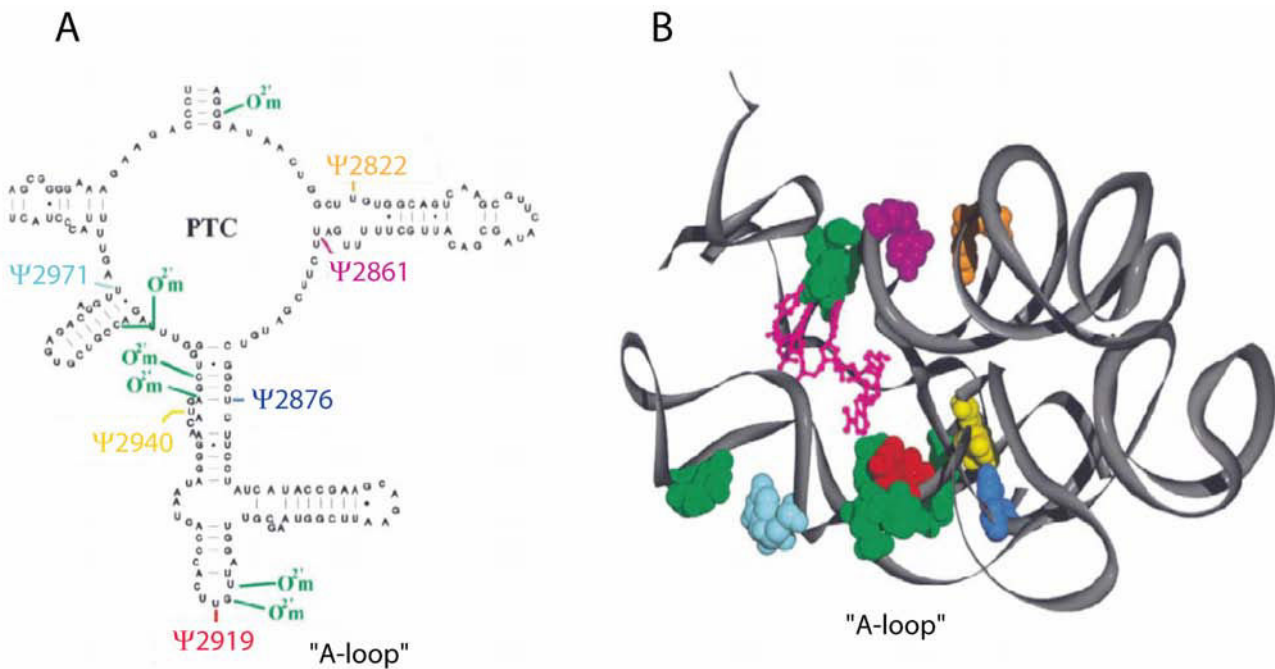


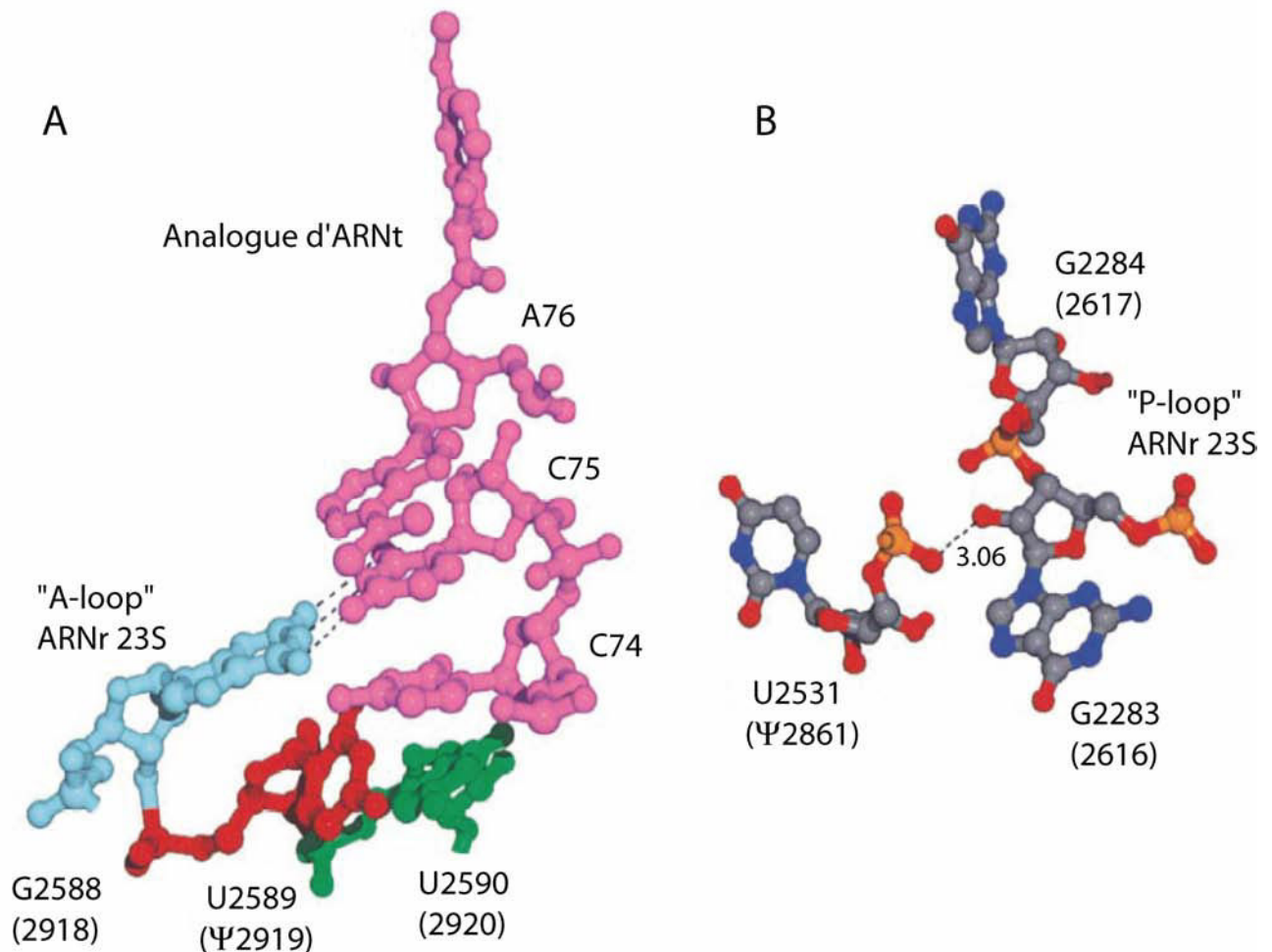
Figure III.32 : Le centre peptidyl transférase (PTC) de l'ARNr 23S de *S. cerevisiae* contient 6 résidus pseudouridine (d'après King *et al.*, 2003).

A. Structure secondaire de la région centrale du PTC de l'ARN 23S de *S. cerevisiae*. Les résidus 2'-O-méthylés sont représentés en vert. Les résidus pseudouridines sont indiqués par des couleurs différentes.

B. Prédiction de la localisation des nucléotides modifiés au sein de la structure 3D du PTC. Les modifications post-transcriptionnelles observées chez *S. cerevisiae* ont été remplacées sur la structure du PTC de *H. marismortui*, complexé à de la CCdA-p-putomycine (Yarus analog, en magenta) (Nissen *et al.*, 2000). Le code couleur est le même qu'en A.

Cependant, du fait de leur localisation à proximité des sites A et P, les résidus Ψ pourraient agir sur la fixation des ARNt, leur sélection, leur translocation et leur sortie. Il a été suggéré, d'après la structure 3D du ribosome, qu'une interaction entre le résidu Ψ2919 et le trinucléotide CCA à l'extrémité de l'ARNt ait lieu (Fig. III.33A) (King *et al.*, 2003). Les résidus adjacents en 5', G2918 et U2920 (2588 et 2590 chez *H. marismortui*) forment des appariements de type Watson-Crick avec le résidu C75 et C74 du CCA de l'ARNt fixé au site A (Kim et Green, 1999 ; Ban *et al.*, 2000). Le résidu Ψ2919 pourrait stabiliser ces appariements en augmentant le potentiel de formation de liaisons hydrogène. Par ailleurs, le résidu Ψ2861, situé à proximité du site P, pourrait de la même manière intervenir lors de la fixation de l'ARNt à ce site (Fig. III.33B). De même, il est envisageable que le résidu Ψ2492 (numérotation de *H. marismortui* ; (Ban *et al.*, 2000) conservé dans la plupart des espèces étudiées et localisé 6 nt en aval du résidu A2486 qui serait lui-même le nucléotide agissant directement dans la formation de la liaison peptidique, pourrait jouer un rôle dans la stabilisation de la conformation la plus efficace de ce dernier résidu (Nissen *et al.*, 2000; Ofengand *et al.*, 2001b). Une autre illustration est la présence de deux résidus Ψ localisés dans l'hélice 69 de l'ARN de la grande sous-unité ribosomale et conservés entre espèces éloignées. Ces deux résidus modifiés ne sont pas essentiels mais confèrent un avantage dans des expériences de compétition (Badis *et al.*, 2003). Ils seraient importants pour stabiliser une conformation spécifique

de la boucle de l'hélice 69 qui interagit avec des résidus de la petite sous-unité et des résidus de l'ARNt fixé au site A (Meroueh *et al.*, 2000 ; Sumita *et al.*, 2005). Cette région de l'ARN de la grande sous-unité comporte les résidus Ψ les mieux conservés dans les trois domaines du vivant.



La restriction des résidus Ψ au niveau des régions fonctionnelles et le fait que deux systèmes de pseudouridylation aient été maintenus au cours de l'évolution suggèrent que les résidus Ψ jouent un rôle important dans le fonctionnement des ribosomes, certainement en optimisant chaque aspect de la traduction. La fonction principale des résidus Ψ présents dans les ARNr pourrait être de jouer le rôle d'une « colle moléculaire » qui fortifie une conformation qui en soi n'est pas suffisamment rigide (Ofengand, 2002).

Les eucaryotes supérieurs, dont le nombre de résidus Ψ dans les ARNr dépasse parfois la centaine, semblent avoir tiré profit de la possibilité de placer des résidus Ψ à chaque fois que la stabilisation d'une conformation trop souple présente un avantage (Ofengand *et al.*, 2001b). Cependant, nous ne pouvons pas exclure une part de hasard dans la présence de certains résidus Ψ d'eucaryotes. De plus tous les résidus Ψ ne semblent pas avoir la même importance fonctionnelle. Ainsi, par exemple, les résidus Ψ du site PTC semblent jouer un rôle important dans le maintien de la conformation de cette région impliquée dans la catalyse de la formation de la liaison peptidique. Certains résidus Ψ n'apporteraient un effet bénéfique qu'associés à d'autres résidus Ψ . Il faut noter aussi que des études phénotypiques plus élaborées sont nécessaires pour mesurer l'importance de ces résidus. Il apparaît donc difficile d'étudier expérimentalement l'apport bénéfique des résidus Ψ chez les eucaryotes.

Au contraire, il est envisageable que l'étude expérimentale des ARN d'archaea qui évoluent dans des conditions variées et parfois extrêmes puisse nous renseigner sur leur rôle bénéfique. Par exemple, comme discuté ci-dessus, la présence de résidus Ψ dans le site PTC pourrait permettre de stabiliser une conformation favorable à la catalyse de la formation de la liaison peptidique dans des conditions particulières de température, de pH ou de concentration saline. En effet, il a déjà été démontré que la structure du site PTC est sensible au pH et à la concentration en ions (Bayfield *et al.*, 2004). Le positionnement de nucléotides impliqués directement ou indirectement dans cette catalyse mais aussi dans les interactions avec les ARNt ou l'ARN de la petite sous-unité ribosomale dépend donc de ces conditions. Ainsi, chez les archaea, il serait envisageable d'expliquer le nombre et la localisation des résidus Ψ en lien avec leurs conditions de croissance très particulières.

OBJECTIFS

Objectifs de mon travail de thèse

Les données présentées dans le chapitre d'introduction conduisent à plusieurs observations :

- les ARNnc jouent un rôle très important dans la plupart des processus cellulaires et ce rôle a longtemps été négligé voire complètement ignoré.
- les projets de séquençage ont conduit à l'accumulation de séquences génomiques contenant une énorme quantité d'information encore inexploitée.
- le domaine des archaea présente un vif intérêt mais les informations sur ce domaine, en particulier celles qui concernent les ARNnc, sont limitées.
- les caractéristiques des systèmes de pseudouridylation ainsi que le rôle des modifications qu'ils génèrent sont encore peu connus.

Nous avons essayé, au cours de ce travail, d'apporter des informations sur une classe particulière d'ARNnc chez les archaea : les sRNA H/ACA. Nos premiers travaux visaient à identifier de nouveaux ARNnc dans les génomes d'archaea. Il s'agissait d'une nouvelle thématique au laboratoire. F. Leclerc spécialiste de bioinformatique venait d'intégrer le laboratoire. Grâce à ses compétences, nous avons développé une nouvelle approche bioinformatique permettant de rechercher des ARNnc dans les séquences des génomes d'archaea disponibles, ceci en tenant compte des annotations de ces génomes et des caractéristiques générales des ARNnc. Lorsque j'ai débuté ce travail sous la direction de C. Branlant en étant encadré par F. Leclerc, trois séquences génomiques d'espèces du genre *Pyrococcus* venaient d'être rendues publiques. Ceci nous a conduit à envisager une approche de génomique comparative appliquée initialement à ces trois espèces et applicable par la suite à d'autres espèces d'archaea qui présentent des distances phylogéniques compatibles. Cette approche nous a permis de décrire 73 gènes candidats d'ARNnc conservés dans les génomes des trois *Pyrococcus*.

Lorsque je réalisais ce travail, l'équipe d'A. Huttenhofer a identifié les premiers sRNA à boîtes H/ACA guides de pseudouridylation chez *A. fulgidus*. En prenant en considération les caractéristiques structurales des 3 premiers sRNA à boîtes H/ACA d'Archaea ainsi identifiés, j'ai pu caractériser 7 ARN parmi les 73 candidats d'ARNnc pouvant correspondre à des guides de pseudouridylation. J'ai vérifié expérimentalement l'expression de ces 7 candidats sRNA H/ACA et, afin de définir quelles sont leurs cibles, analysé la localisation des résidus Ψ dans les ARNr de *P. abyssi*, ceci en collaboration avec I. Behm-Ansmant. Compte tenu du fait que les systèmes sRNP à boîtes H/ACA d'archaea n'avaient pas encore été étudiés, B. Charpentier a mis au point une technique de reconstitution de particules sRNP H/ACA à partir des protéines recombinantes et de sRNA transcrits *in vitro* qui a permis d'étudier leurs mécanismes d'assemblage et d'action, et j'ai participé à ce travail.

Aucune étude complète de tous les systèmes de pseudouridylation des ARNr d'une espèce d'archaea donnée n'avait été réalisée. La question était de savoir si, comme chez les eucaryotes, toutes les pseudouridylations des ARNr dépendent d'un système incluant un ARN guide à boîtes H/ACA ou si, comme chez les bactéries, certaines sont catalysées par des protéines à activité ARN: Ψ -synthases agissant sans ARN guide. Notre troisième objectif a donc été d'analyser la

contribution des sRNA à boîtes H/ACA à la pseudouridylation des ARNr de *P. abyssi*. Cette étude comportait la recherche exhaustive par voie bioinformatique de sRNA H/ACA, ainsi que la reconstitution de particules sRNP H/ACA actives et l'étude de leur spécificité. Cette étude devait également nous apporter des renseignements sur les caractéristiques intrinsèques des sRNA H/ACA, en particulier au niveau des contraintes structurales et des appariements requis pour qu'il y ait modification de la cible.

Sur la base des informations ainsi acquises, nous avons ensuite pu étendre la recherche bioinformatique de sRNA à boîtes H/ACA à tous les génomes d'archaea dont la séquence était disponible. Les caractéristiques définies ont été utilisées pour la mise au point d'une recherche bioinformatique de gènes de sRNA H/ACA dans les génomes entiers. A partir des données obtenues nous avons pu étudier la distribution et l'évolution des sRNA à boîtes H/ACA au sein du domaine archaea.

Cette recherche nous a conduit à l'identification de près d'une centaine de sRNA à boîtes H/ACA chez les archaea. Ils ont pu être regroupés en différentes classes ciblant une ou plusieurs positions dans les ARNr. Néanmoins, un petit groupe d'ARN H/ACA ne semblait pas cibler les ARNr mais la position 35 dans les ARNt^{Tyr} de crénotes. Or, jusqu'alors aucun sno/sRNA à boîtes H/ACA n'avait été montré participer à la modification d'ARNt. En parallèle, nous avons remarqué que l'ARN:Ψ-synthase aPus7 proposée comme étant responsable de la modification de la position 35 des ARNt était présente mais mutée dans les espèces comportant cette classe particulière de sRNA à boîtes H/ACA. Notre dernier objectif a donc été de prouver que les sRNP à boîtes H/ACA que nous avons détectées étaient responsables de la pseudouridylation en position 35 des ARNt chez certains crénotes. Cette démonstration a été obtenue en collaboration avec Y. Motorin, spécialiste des ARN:Ψ-synthases et d'A. Urban, doctorant.

L'enzyme aCBF5 est le seul représentant de la famille TruB chez les archaea. Or, un résidu Ψ est détecté en position 55 dans les ARNt élongateurs d'archaea. Ainsi, j'ai montré en parallèle avec les équipes de Terns et H. Grosjean (Roovers *et al.*, 2006), que aCBF5 est capable de catalyser la formation du résidu Ψ55 en absence d'ARN H/ACA. Avec B. Charpentier et J.B. Fourmann, doctorant, nous avons comparé les déterminants au niveau ARN H/ACA et au niveau aCBF5 et aNOP10 qui sont nécessaires à l'activité de aCBF5 dans un système comportant un ARN guide et sans ARN guide. Nous avons pu démontrer que des résidus de aCBF5 ont été conservés au cours de l'évolution car ils sont essentiels pour l'activité de sRNP à boîtes H/ACA, alors que d'autres résidus sont conservés car essentiels à l'activité enzymatique en absence de sRNA à boîtes H/ACA.

MATERIEL & METHODES

-Partie bioinformatique-

1. Matériel

1.1. Matériel informatique

Tous les alignements, les recherches de motifs et les prédictions de structure ont été réalisés à partir d'une station PC-Linux avec le système Redhat 7.0.

1.2. Logiciels utilisés

Blasta est un algorithme basé sur l'interprétation de comparaison de deux séquences (Altschul *et al.*, 1997).

Fasta est un algorithme qui, comme Blasta, est basé sur l'interprétation de deux séquences mais celui-ci aboutit à un alignement global de deux séquences (Pearson et Lipman, 1988). La version utilisée est Fasta 2.0.

Clustal-W (version 1.81) est un programme qui permet un alignement global de plusieurs séquences (Thompson *et al.*, 1994).

RNAMOT est un programme qui permet la recherche de motifs particuliers dans une séquence. Le logiciel comporte un descripteur de motifs permettant de rechercher ces motifs (Gautheret *et al.*, 1993).

GeneMarkS est un programme récent qui prédit la présence d'ORF dans les séquences soumises à examen (Besemer *et al.*, 2001). Ce programme n'avait pas été utilisé pour annoter les génomes dont la séquence complète était disponible avant que nous débutions cette étude. Ce programme est accessible sur le site :

<http://dixie.biology.gatech.edu/GeneMark/genemarks.cgi>.

MUMer est un programme permettant l'alignement global de deux séquences complètes de génome (Delcher *et al.*, 1999). La version 2.0 est accessible sur le site :

<http://www.tigr.org/tigr-scripts/CMR2/webmum/mumplot>.

Mfold (version 3.1) est un programme de prédiction de structures secondaires (Zuker et Jacobson, 1998; Mathews *et al.*, 1999).

ERPIN est un programme de recherche de motifs basé sur l'élaboration d'un profil à partir d'un alignement de séquences (Gautheret et Lambert, 2001). Son utilisation sera détaillée dans le chapitre d'ouvrage que nous avons publié dans « Methods in Enzymology », présenté dans la partie résultat.

1.3. Outils informatiques utilisés

Emacs (GNU Emacs version 1.1) est un éditeur de texte qui facilite la saisie et le formatage des informations.

Extract fait partie de la série de programmes « glimmer 2.02 ». C'est un utilitaire qui permet d'extraire, à partir d'un fichier de coordonnées, des séquences précises d'un organisme.

Revseq fait partie de la série de programmes « EMBOSS-1.13.3 ». Cet utilitaire convertit les séquences en leur séquence complémentaire.

Readseq (développé par le NCBI) transforme les séquences en différents formats.

Seqmatchall fait partie de la série de programmes « EMBOSS-1.13.3 ». Cet utilitaire permet la recherche d'identités strictes afin d'observer des répétitions de séquences.

Formatdb (développé par le NCBI) est indispensable pour la création de banques de données.

Freqbmk et **Freqsq** ont été utilisés pour calculer la fréquence de l'usage des codons des ORF détectées dans les IRC. Freqbmk et Freqsq sont accessibles sur le site <http://www.infobiogen.fr/services/menuserv.html>.

RNA Forester (Package Vienna 1.6.4 ; Hofacker *et al.*, 1994) a été utilisé afin de construire un arbre phylogénétique de motifs H/ACA. Le programme compare entre elles les structures secondaires précisées en entrée, ceci selon un modèle d'évolution basé sur des travaux récents (Jiang *et al.*, 1995 ; Hoechsmann *et al.*, 2003).

1.4. Programmes élaborés

ExtractIRC : ce programme, écrit en langage shell (tcsh), trie les positions des régions codantes (RC) de telle sorte qu'il peut extraire les séquences des IRC dont la taille est supérieure à 15 nt.

Multiblast : ce programme, écrit en langage shell (tcsh), permet la comparaison de la séquence d'un IRC à toutes les séquences des IRC appartenant à une banque de données. Il utilise l'option Blastall de Blasta dont les paramètres sont définis par défaut.

1.5. Séquences et annotations des génomes

Le tableau I.1 présente les séquences et annotations de génomes d'archaea utilisées lors de ce travail. Les séquences et annotations des génomes ont été téléchargées à partir du site ftp du NCBI (GenBank) : <ftp://ftp.ncbi.nih.gov/genomes/>,

1.6. Autres sites d'intérêt

Les séquences et positions des gènes des équivalents Archaea des snoRNA à boîte C/D sont accessibles sur le site : <http://rna.wustl.edu/snoRNAdb/>. La prédiction de leurs gènes a été effectuée notamment par snoscan accessible via le site : <http://lowelab.ucsc.edu/snoscan/>.

Les comparaisons rapides de séquences par l'utilisation de Blast et de ses variants ont été réalisées en ligne sur le site : <http://www-archbac.u-psud.fr/genomics/lbmgeblast.html>.

La prédiction de gènes d'ARNt dans les séquences génomiques non annotées a été effectuée en ligne par tRNA-scan-SE via le site : <http://lowelab.ucsc.edu/tRNAscan-SE/>.

Le serveur ERPIN sur lequel nous avons rendu accessibles des alignements de sRNA à boîtes H/ACA permettant de les rechercher dans les génomes d'archaea est accessible via le site : <http://tagc.univ-mrs.fr/erpin/>.

2. Méthodes

Identification de gènes d'ARNnc (Fig. M.1)

Les séquences et positions des régions codantes (RC) des génomes ont été saisies sous forme de tableaux à l'aide de l'utilitaire emacs, à partir des annotations de GenBank. Les séquences des IRC (inter régions codantes) ont ensuite été extraites par traitement de ces tableaux à partir du programme extractIRC. Les séquences des IRC ont été reformatées à l'aide de Readseq, puis regroupées. L'utilitaire formatdb a permis à partir de ces collections de séquences de former des banques de données comprenant, d'une part, toutes les séquences des IRC des onze génomes d'archaea, et d'autre part, trois banques de données comprenant chacune l'ensemble des IRC d'un des génomes de *Pyrococcus*.

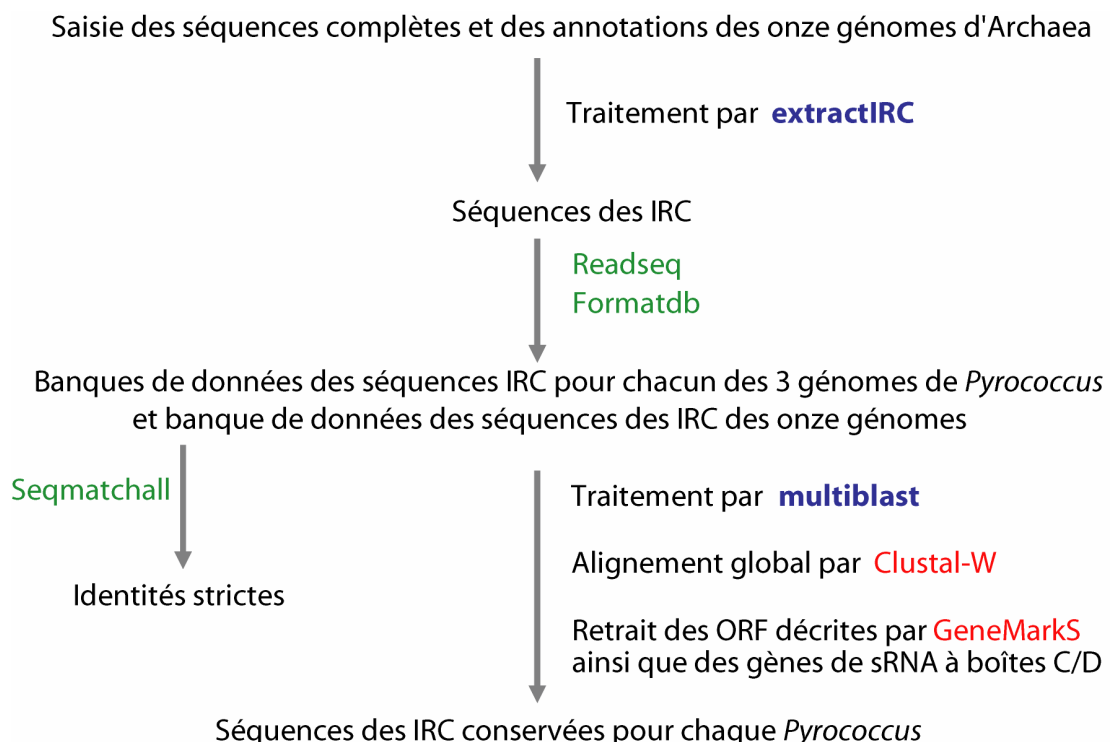


Figure M.1 : Les différentes étapes de l'identification de gènes d'ARNnc dans les génomes des trois *Pyrococcus*. Les programmes que nous avons élaborés apparaissent en bleu, les outils informatiques en vert et les logiciels en rouge.

Dans un premier temps, grâce à l'utilitaire Seqmatchall, des séquences strictement identiques de taille supérieure à 50 nts conservées à l'intérieur de chacun des génomes ont été recherchées. Ces séquences correspondent à des motifs répétés, elles seront notamment utilisées dans l'analyse des résultats de l'étape suivante.

Le programme multiblast a ensuite été utilisé pour comparer une banque de données comprenant toutes les séquences des IRC d'un génome de *Pyrococcus* à la banque de données des séquences des IRC des onze Archaea. De cette comparaison, nous avons extrait deux informations. La première, issue de la comparaison des séquences des IRC d'un génome de *Pyrococcus* avec les séquences des IRC de tous les Archaea, nous a permis d'apprécier la complexité de la séquence étudiée. Ainsi, par exemple, une séquence retrouvée une fois dans deux génomes de *Pyrococcus*, mais présente de nombreuses fois dans le génome d'*Archaeoglobus fulgidus*, pourrait en fait correspondre à une séquence qui serait peu informative si l'objectif est de rechercher des gènes d'ARN régulateurs, et a été de ce fait rejetée. La seconde information, issue de la comparaison des séquences des IRC d'un génome de *Pyrococcus* aux séquences des IRC des deux autres génomes de *Pyrococcus*, nous a permis d'analyser les similarités de séquence entre IRC issues des trois génomes de *Pyrococcus*. Les résultats ainsi obtenus ont été analysés pour ne retenir que les segments IRC issus d'une espèce de *Pyrococcus* dont l'alignement avec les IRC des deux autres *Pyrococcus*, réalisé à l'aide de Clustal-W, montrait une similarité de séquence significative. Les IRC comportant des ORF prédites par GeneMarkS ou des gènes de sRNA à boîtes C/D déjà décrits ont été retirées.

Identification de gènes de sRNA à boîtes H/ACA et de leurs cibles.

La méthode conduisant à l'identification des gènes de sRNA à boîtes H/ACA et de leurs cibles a été publiée dans « Methods in Enzymology » et sera présentée dans la partie « Résultats » (chapitre d'ouvrage N°2).

-Partie Biologie Moléculaire-

1. Matériel

1.1. Les souches utilisées

1.1.1. Souches de bactéries

Les souches d'*E. coli* **DH5 α** (*supE44*, Δ *lacU169* (Φ 80 *lacZ* Δ M15), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*), **XL1-blue** (*recA1*, *endA1*, *gyrA48*, *thi-1*, *hsdR17*, *supE44*, *relA1*, *lac F'* [*proAB*, *lacI^q*, *lacZ* Δ M15, Tn10(*Tet^r*)]), **TOP 10 F'** (Invitrogen) (*F'*, *endA1*, *recA1*, *hsdR17* (*r_k⁻*, *m_k⁺*), *supE44*, *thi-1*, *gyrA96*, *relA1*, Φ 80*lacZ* Δ M15, Δ (*lacZY A- argF*), U169 λ -) et **TG1** (*supE*, *hsd* Δ 5, *thi*, Δ (*lac-proAB*), *F'* [*traD36*, *proAB*+, *lacI^q*, *lacZ* Δ M15]) ont été utilisées pour la production et la purification de plasmides.

Les souches **BL21 CodonsPlus** (DE3) RIL d'*E. coli* (*F'*, *ompT*, *hsdS*(*r_B*.*m_B*-), *dcm*+, *Tet^r*, *gal* λ (DE3), *endA*, Hte[*argU ileY leuW Cam^r*]) ont été utilisées pour la production de protéines recombinantes aPus7.

1.1.2. Souches d'archaea

Les souches *Pyrococcus abyssi* GE5, *Pyrococcus furiosus* DSM 3638, *S. solfataricus* P2, et *Methanopyrus kandleri* AV19 correspondent aux souches dont le génome est disponible. Des culots cellulaires des souches de *P. abyssi*, *P. furiosus* et *S. solfataricus* cultivés en milieu VSM nous ont été aimablement transmis par F. Matsunaga et A. Hecker appartenant au laboratoire LBMGE de P. Forterre à Orsay. Des culots cellulaires de *M. kandleri* nous ont été aimablement transmis à partir du laboratoire de K. Stetter (Regensburg).

1.2. Les plasmides utilisés

Plasmide pT-Adv (Clontech) ou **pCR2.1** (Invitrogen) : Le clonage de fragments d'ADN obtenus par amplification par la technique PCR a été réalisé dans le plasmide pT-Adv fourni avec le kit "Advantage PCR cloning kit". Le plasmide est commercialisé sous forme linéarisée et porte un résidu thymine protubérant sur chaque extrémité 3'. Il contient des gènes de résistance à l'ampicilline et à la kanamycine. Il possède également le gène *lacZ* codant la β -galactosidase sous le contrôle du promoteur *Plac* et interrompu par la cassette de clonage. Ce plasmide a été utilisé pour vérifier la séquence des produits de PCR qui ont été utilisés pour exprimer la plupart des sRNA à boîtes H/ACA et de leur(s) cible(s).

Plasmide pUC18 : ce plasmide, linéarisé par *SmaI* et déphosphorylé a été utilisé pour cloner les fragments ADN amplifiés par PCR et phosphorylés correspondant aux protéines aPus7 WT

de *P. abyssi* et *S. solfataricus* ainsi que les ARNt (Asp) et (Tyr) provenant des deux espèces citées.

Plasmide pet28B : ce plasmide, linéarisé par *NheI/BamHI*, a été utilisé pour cloner les fragments correspondant aux protéines aPus7 insérés dans le plasmide pUC18 et extraits par coupure *NheI/BamHI* en vue d'une production *in vitro* avec une étiquette His et d'une purification sur résine de Nickel.

Plasmide pGEX-6P-1 (Amersham Pharmacia Biotech) : ce plasmide a été utilisé pour la production de protéines recombinantes TruB, aPus10, L7Ae, aCBF5, aGAR1 et aNOP10 fusionnées à l'étiquette GST en vue d'une purification sur résine glutation-sépharose. Les fragments ADN correspondant aux séquences de ces protéines ont été clonés entre les sites *BamHI* et *XhoI* de la cassette de clonage.

1.3. ADN génomique de *P. abyssi* et *S. solfataricus*

Des lots d'ADN génomique de ces deux espèces nous ont été transmis par F. Matsunaga du laboratoire de P. Forterre à Orsay (LBMGE).

1.4. Milieux de culture, tampons et solutions d'usage courant

Milieu LB (Luria Bertani) : peptone pancréatique animale 5 g/l ; peptone tryptique de caséine 5 g/l ; extrait de levure 5 g/l ; NaCl 10 g/l. Le pH est ajusté à 7,5 avec NaOH 10 N. Le milieu de culture est ensuite autoclavé 20 min à 121°C sous une pression de 1 bar.

Milieu solide LB : l'agar est ajouté à raison de 17 g par litre de milieu de culture LB avant stérilisation à l'autoclave.

Milieu LBA/LBK : l'ampicilline ou la kanamycine (100 mg/ml) est ajoutée après autoclave dans le milieu de culture LB à une concentration finale de 100 µg/ml.

Milieu SOC : Tryptone 2 % ; Extrait de levure 0.5 % ; NaCl 10 mM ; KCl 2,5 mM ; MgCl₂-6H₂O 10 mM ; Glucose 20 mM.

Milieu VSM (Vent Sulfothermophiles Medium): NaCl 20 g ; KCl 0,25 g ; NaBr 0,05 g ; SrCl₂-6xH₂O 0,5 g ; Acide borique 0,08 g ; PIPES 3 g ; Extrait de levure 1 g ; Peptone 4 g ; Resazurine 0,1 %. Le pH est ajusté à 6,8. Le milieu de culture est ensuite autoclavé puis sont ajoutés : 5 ml de MgSO₄ 20 % ; 1 ml de CaCl₂ 5 % ; 4 ml de KH₂PO₄ 5 %.

TE : Tris-HCl 10 mM, pH 7,4 ; EDTA 1 mM.

TBE 1 X : Tris borate 89 mM, pH 8,3 ; EDTA 2 mM.

Solution de polyacrylamide dénaturante : Acrylamide/Bisacrylamide (23,75/1,25) 25 % (p/v) ; TBE 1X ; urée 8 M.

Solution de polyacrylamide native : Acrylamide/Bisacrylamide (38/2) 40 % (p/v) ; TBE 1X.

Solution de bleu de dépôt agarose (6X) : EDTA 125 mM ; SDS 2 % (v/v) ; glycérol 25 % (v/v) ; bleu de bromophénol 0,12 % (p/v).

Solution de bleu de formamide : EDTA 20 mM ; bleu de bromophénol 0,05 % (p/v) ; bleu de xylène cyanol 0,05 % (p/v) dans de la formamide désionisée.

Solution de bleu de dépôt (non dénaturant) pour les expériences de retard sur gel (CSB) : HEPES-KOH 20 mM, pH 7,9 ; glycérol 40 % (v/v) ; bleu de bromophénol 0,05 % (p/v) ; bleu de xylène cyanol 0,03 % (p/v).

Solution de bleu de coloration de gel de protéine : Ethanol 20% (v/v) ; Acide Acétique 10% (v/v) ; 2 g/l Bleu de Coomassie R250.

Solution de décoloration de gel de protéine : Ethanol 20% (v/v); Acide Acétique 10% (v/v).

Tampon de gel de migration de gel de protéine : Glycine 14,25 g/l ; Tris 3 g/l ; SDS 1 g/l.

Tampon d'élution : Tris-HCl 10 mM, pH 7,5 ; NaCl 300 mM ; EDTA 1 mM ; SDS 1%.

Tampon D 10x : HEPES-KOH 20 mM, pH 7,9 ; KCl 150 mM ; MgCl₂ 1,5 mM ; EDTA 0,2 mM.

Tampon Psi 1x : Tris-HCl 100 mM, pH 8 ; AcNH₄ 100 mM ; MgCl₂ 5 mM ; DTT 2 mM ; EDTA 0,1 mM.

Tampon de migration de gel retard : Glycérol 87% 115 ml ; TBE 20x 50 ml ; H₂O qsp 2 l.

Tampon T2 2x : Acétate d'ammonium 100 mM, pH 4,6.

Tampon P1 2x : Acétate d'ammonium 100 mM, pH 5,2.

Tampon de migration N1 pour TLC : Acide Isobutyrique 25% 330 ml ; Ammoniac 25% 5 ml ; EDTA 0,5 M pH 8 ; H₂O qsp 500 ml.

Tampon de migration N2 pour TLC : Isopropanol 70 ml ; HCl 37% 15 ml ; H₂O qsp 100 ml.

Tampon de migration RII pour TLC : Isopropanol 10 ml ; Sulfate d'ammonium 300 g ; Na₂HPO₄/NaH₂PO₄ pH 6,8 0,1 M final (soit 127,5 ml Na₂HPO₄ mélangé à 123,5 ml NaH₂PO₄) ; H₂O qsp 500 ml.

Tampon de clivage pour la protéase PreScission : Tris-HCl 50 mM pH 7 ; NaCl 150 mM ; EDTA 1 mM ; DTT 1 mM.

2. Méthodes spécifiques

2.1. Extraction des ARN totaux

Les techniques d'extraction des ARN totaux de bactérie, levures et d'archaea sont présentées dans le chapitre d'ouvrage N°1 qui apparaît dans le paragraphe suivant.

2.2. Identification de résidus modifiés dans les ARN par des méthodes basées sur l'extension par la transcriptase inverse

Le chapitre d'ouvrage suivante présente différentes techniques basées sur l'extension d'amorce par la transcriptase inverse qui permettent d'identifier les nucléotides modifiés au sein des ARN totaux ou au sein des transcrits *in vitro*. La technique reposant sur l'utilisation de la carbodiimide CMCT nous a notamment permis de détecter la présence de résidus Ψ dans les ARNr et ARNt extraits de *P. abyssi* et *S. solfataricus*.

Chapitre d'ouvrage N°1 :

Identification of modified residues in RNAs by reverse transcription - based methods.

Yuri MOTORIN, Sébastien MULLER, Isabelle BEHM-ANSMANT, Christiane BRANLANT

Methods in Enzymology, 2007 volume 425:355-87.

CHAPITRE D'OUVRAGE N°1

IDENTIFICATION OF MODIFIED RESIDUES IN RNAs BY REVERSE TRANSCRIPTION-BASED METHODS

Yuri Motorin, Sébastien Muller, Isabelle Behm-Ansmant, *and* Christiane Branlant

Contents

1. Introduction	22
2. Reverse Transcription (RT)–Based Methods for Detection of Modified Residues	24
2.1. Reverse transcriptases used for RNA analysis	26
2.2. General protocol for primer labeling and reverse transcription of RNA	27
3. RNA Extraction from Various Cell Types	29
3.1. Protocol for the extraction of total RNA from archaeal or eukaryal cells	29
3.2. Protocol for the extraction of total RNA from yeast cells	31
4. RNA Modifications Detectable After Specific Chemical Treatment	32
4.1. Detection of inosine using glyoxal treatment followed by RNase T1 hydrolysis	32
4.2. Detection of pseudouridine (Ψ) residues	36
4.3. Protocol for the detection of pseudouridine residues in RNA using CMCT modification	36
4.4. Complementary approaches for Ψ detection	41
4.5. Detection of 5-methylcytosine (m ⁵ C)	41
4.6. Detection of 7-methylguanine (m ⁷ G)	42
4.7. Detection of 2'-O-methylated nucleotides using OH [−] cleavage or 2'-OH reactivity	43
4.8. Protocol for detection of 2'-O-methylated residues in RNA using OH [−] cleavage	43
4.9. Detection of 2'-O-methylated nucleotides by use of low dNTP concentrations	44

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, Nancy Université, Faculté des Sciences et Techniques, Vandoeuvre-les-Nancy, France

Methods in Enzymology, Volume 425

ISSN 0076-6879, DOI: 10.1016/S0076-6879(07)25002-5

© 2007 Elsevier Inc.

All rights reserved.

4.10. Protocol for detection of 2'- <i>O</i> -methylated residues in RNA by use of low dNTP concentrations	45
4.11. Other chemical modification methods specific for 2'- <i>O</i> -methylations	47
4.12. Detection of dihydrouridine (D) residues by alkaline hydrolysis	47
5. RNA Modifications Leading to RT Pauses Without Preliminary Chemical Treatment	48
6. Conclusion	48
References	49

Abstract

Naturally occurring modified residues derived from canonical RNA nucleotides are present in most cellular RNAs. Their detection in RNA represents a difficult task because of their great diversity and their irregular distribution within RNA molecules. Over the decades, multiple experimental techniques were developed for the identification and localization of RNA modifications. Most of them are quite laborious and require purification of individual RNA to a homogeneous state. An alternative to these techniques is the use of reverse transcription (RT)-based approaches. In these approaches, purification of RNA to homogeneity is not necessary, because the selection of the analyzed RNA species is done by specific annealing of oligonucleotide DNA primers. However, results from primer extension analysis are difficult to interpret because of the unpredictable nature of RT pauses. They depend not only on the properties of nucleotides but also on the RNA primary and secondary structure. In addition, the degradation of cellular RNA during extraction, even at a very low level, may complicate the analysis of the data. RT-based techniques for the identification of modified residues were considerably improved by the development of selected chemical reagents specifically reacting with a given modified nucleotide. The RT profile obtained after such chemical modifications generally allows unambiguous identification of the chemical nature of the modified residues and their exact location in the RNA sequence. Here, we provide experimental protocols for selective chemical modification and identification of several modified residues: pseudouridine, inosine, 5-methylcytosine, 2'-*O*-methylations, 7-methylguanosine, and dihydrouridine. Advice for an optimized use of these methods and for correct interpretation of the data is also given. We also provide some helpful information on the ability of other naturally occurring modified nucleotides to generate RT pauses.

1. INTRODUCTION

The presence of posttranscriptionally modified nucleotides derived from the canonical A, C, G, and U residues is a characteristic feature of most cellular RNAs (for review, Björk, 1995; Björk *et al.*, 1987; Massenet *et al.*, 1998;

Ofengand and Fournier, 1998; Internet resource: <http://library.med.utah.edu/RNAmods/>). These modified residues are formed during RNA maturation steps, and their formation is catalyzed by numerous specific enzymes (RNA: modification enzymes) or by RNP particles (for review, Ansmant and Motorin, 2001; Garcia and Goodenough-Lashua, 1998; Grosjean *et al.*, 1998; Internet resource: <http://genesilico.pl/modomics/index2.pt>).

Despite extensive studies of modified nucleotides in RNAs over more than 40 years, our knowledge on their presence, chemical identity, and localization remains limited to highly abundant cellular RNAs, like tRNAs (Sprinzl and Vassilenko, 2005; Internet resource: <http://www.uni-bayreuth.de/departments/biochemie/trna/>), rRNAs (McCloskey and Rozenski, 2005; Internet resource: <http://medlib.med.utah.edu/SSUmods>) and U snRNAs (Massenet *et al.*, 1998). In contrast, low abundance or recently discovered RNAs, such as snoRNAs or regulatory RNAs from eukaryotes and bacteria, as well as mRNAs, have not been studied yet from this point of view. In most cases, this lack of information is due to experimental difficulties in detecting and localizing modified residues in low-abundance RNAs.

Multiple experimental techniques allowing the detection and analysis of modified residues in RNAs have been developed in the past. The techniques of direct RNA sequencing and fingerprinting require the purification of homogeneous, pure RNA in sufficient amounts (at least 1–2 μ g). However, these methods allow the detection of noncanonical residues and their precise localization (Branch *et al.*, 1989; Gupta and Randerath, 1979; Stanley and Vassilenko, 1978; Tanaka *et al.*, 1980). The application of these and related techniques in 1970–1980 allowed an extensive analysis of tRNAs (more than 500 species) from various sources and the constitution of a tRNA modification database (Sprinzl and Vassilenko, 2005). Posttranscriptional modifications were also detected and localized in several rRNAs and U snRNAs (Branlant *et al.*, 1981; Krol *et al.*, 1981; Veldman *et al.*, 1981; see Massenet *et al.*, 1998, for review). Later on, the methods based on high-performance liquid chromatography (HPLC) allowed fragmentation and analysis of longer RNAs, like rRNAs from various sources (Bruenger *et al.*, 1993; Gehrke and Kuo, 1989; Gehrke *et al.*, 1982; Smith *et al.*, 1992). More recently, the approaches that use HPLC and mass spectrometry (MS) have been successfully used for direct RNA sequencing and the identification of modified residues (for review, Meng and Limbach, 2006; Thomas and Akoulitchchev, 2006; Douthwaite and Kirpekar, Chapter 1). Recently, the use of DNA chips was suggested for the detection of modified residues in RNAs (Hiley *et al.*, 2005). Despite a great interest in these powerful techniques for RNA analysis, the use of these methods is limited to highly purified RNAs and frequently requires significant amounts of analyzed material. These two impediments are quite important, because purification of low-abundance RNA species from total RNA always represents a

difficult task. In addition, the application of MS approaches requires particular expertise in data interpretation and cannot be done routinely and at high throughput.

An alternative group of techniques capable to detect and to identify some of the modified nucleotides in RNAs is derived from the RNA reverse-transcription (RT) sequencing method (Brownlee and Cartwright, 1977; Lane *et al.*, 1988; Qu *et al.*, 1983). In contrast to direct RNA sequencing, HPLC, or MS, these methods do not require purification of the analyzed RNA species and may be directly applied to total RNA fractions extracted from the cells. The selection of the analyzed RNA is done by specific annealing of DNA oligonucleotides selected for primer extension (Fig. 2.1).

2. REVERSE TRANSCRIPTION (RT)-BASED METHODS FOR DETECTION OF MODIFIED RESIDUES

Despite these advantages, the use of reverse transcription for detection of modified nucleotides in RNAs is not very straightforward. In fact, the reverse transcription profile, even for an unmodified RNA, depends on several parameters, especially the presence of Py-A (C-A and U-A) bonds, which are very sensitive to nuclease cleavage, and also on the RNA secondary structure that can generate RT pauses. Therefore, the simple presence of a pause in a primer extension profile does not necessarily indicate the presence of a modified residue. In addition, reverse transcription profiles of modified RNAs are often prone to some kind of stuttering (doubling of the band on the gel) at the modified residue that blocks the reverse transcription (Bakin and Ofengand, 1993; Denman *et al.*, 1988; Gustafsson and Persson, 1998). The reasons for such a phenomenon are not yet clear, but this possibility complicates the detection of nearby modified residues and should be considered during analysis of reverse transcription patterns.

A significant improvement of the RT-based techniques was the development of specific chemical reagents for several modified nucleotides found in RNAs. Because these chemicals specifically react with a given modified nucleotide, the intensity of the pause observed in the RT profile should depend on the chemical's concentration and the treatment time. For this reason, for an unambiguous identification of modified nucleotides, it is recommended to vary these two parameters. The comparison of the data obtained under these different conditions allows verification that the observed pause can, indeed, be attributed to the action of the selected chemical (Fig. 2.1A). These techniques generally give quite reliable results.

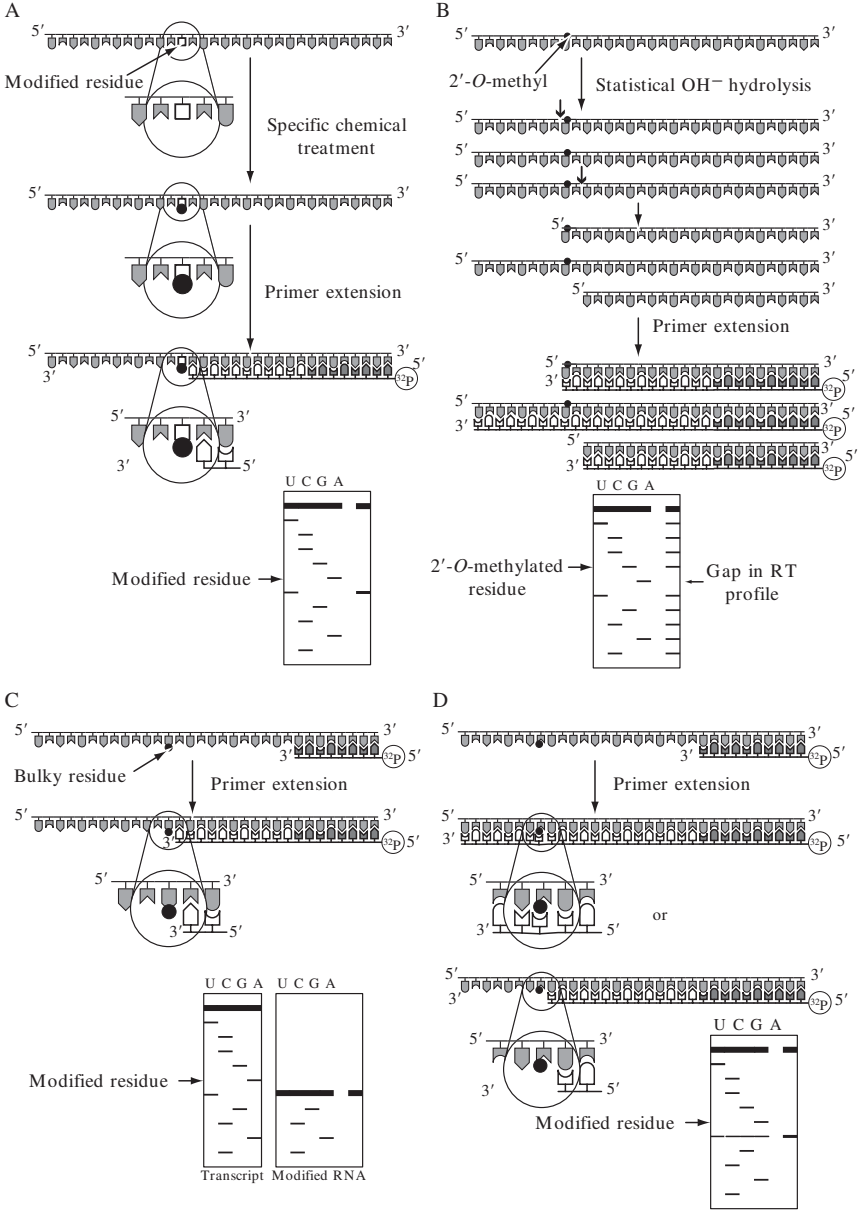


Figure 2.1 Schematic representation of the principle of reverse transcription-based approach used for the detection of modified nucleotides in RNA. (A) Detection of RNA modifications by use of specific chemical modification of modified residues present in natural RNAs. (B) Detection of 2'-O-methylated residues by statistical RNA cleavage by OH^- . (C) Detection of bulky (or non-base pairing) modified residues in RNA that block primer extension. (D) Detection of modified residues that induce a pause in RT profile.

Hence, we will first describe the RT-based techniques based on preliminary chemical treatment of RNA that convert “silent” and otherwise invisible modified nucleotides into chemical adducts detectable by pauses or blockage of the primer extension. To get such extension arrest, the adduct may abolish or limit the base-pairing property of the modified residue or create some steric hindrance in the active site of the RT. Therefore, reverse transcription generally stops one nucleotide downstream of the modified residue (see Fig. 2.1A).

Information on the presence of modified residues in RNA can also be obtained by the natural ability of some modified nucleotides to induce a pause (or a complete block) of the primer extension by reverse transcriptase (Fig. 2.1C,D). Here again, this phenomenon may be related either to the inability to form a base pair with any canonical nucleotide or to steric hindrance limiting or excluding the recognition of the modified residue by the RT. However, as mentioned previously, because there can be multiple reasons for RT pausing, unambiguous localization of modified nucleotide requires additional information. One simple complementary experiment is the comparison of the reverse transcription profiles obtained for the *in vitro* RNA transcript and the authentic cellular RNA. Pauses (or blocks) observed in the cellular RNA but absent in the corresponding *in vitro* transcript may indicate the presence of modifications at these positions. However, extreme care should be taken to avoid nucleolytic degradation of fragile Py-A linkages during extraction of cellular RNA. In addition, modifications located too close to 5′- or 3′-extremity of RNA are difficult to detect because of natural RT pauses at the 5′-end of RNAs and to the necessity of annealing the oligonucleotide primer at the 3′-extremity of the analyzed molecule.

2.1. Reverse transcriptases used for RNA analysis

Viral RNA-dependent DNA polymerases (reverse transcriptases) are most frequently used for RT-based analysis of RNA structure and sequence (including modified nucleotides).

Avian Myeloblastosis virus (AMV) reverse transcriptase available from numerous suppliers is most frequently used for primer extension analysis. However, other reverse transcriptases may also be used, like Moloney murine leukemia virus (MMLV, M-MuLV) reverse transcriptase or its mutants deficient in RNase H activity. The main advantage of AMV reverse transcriptase consists in its higher thermostability; indeed, in contrast to MMLV, it can be used up to 70° in the appropriate buffer (Fuchs *et al.*, 1999). In addition, AMV RT seems to be less sensitive to RNA secondary structures that may perturb primer extension.

2.2. General protocol for primer labeling and reverse transcription of RNA

2.2.1. Equipment

Water baths set to 37°, 42°, and 65°.

Heating block set to 96°.

Equipment for polyacrylamide gel electrophoresis (PAGE) and power supply: small vertical slab gels for primer purification and sequencing gels for primer extension analysis.

Vacuum drier for sequencing gels.

PhosphorImager (Typhoon 9410, GE Healthcare, formerly Amersham Biosciences, or equivalent) or screens, cassettes, and films for autoradiography.

2.2.2. Reagents

[γ -³²P]ATP (3000 Ci/mmol; 10 mCi/ml) (purchased from GE Healthcare, formerly Amersham Biosciences).

T4 Polynucleotide kinase (PNK) (10 U/ μ l) (MBI Fermentas, Lithuania or other source) and corresponding 10 $\times$ reaction buffer (500 mM Tris-HCl, pH 7.6, at 25°, 100 mM MgCl₂, 50 mM DTT, 1 mM spermidine, and 1 mM EDTA).

3 M Sodium acetate, pH 5.4.

Elution buffer (0.5 M sodium acetate, pH 5.4, 1 mM EDTA).

RNAse free 96% ethanol.

Hybridization buffer 5 $\times$ (250 mM Tris-HCl, pH 8.3, 300 mM NaCl, 50 mM DTT).

Mixture of dNTP (5 mM dATP, 5 mM dCTP, 5 mM dGTP, 5 mM dTTP).

ddNTP solutions (0.5 mM): ddATP, ddCTP, ddGTP, and ddTTP.

Primer extension buffer 10 $\times$ (500 mM Tris-HCl, pH 8.3, 600 mM NaCl, 100 mM DTT, 60 mM MgCl₂).

AMV reverse transcriptase (MP Biomedicals, formerly QBiogene) (20 U/ μ l).

TBE buffer (89 mM Tris-borate, pH 8.3, 2 mM Na₂EDTA).

Formamide blue loading solution (20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol blue in deionized formamide).

2.2.3. Procedure

1. The sequences of the oligonucleotides used as primers are chosen to hybridize to the target RNA at least 3-nt downstream from the region to be analyzed. Extension performed with one primer generally enables the analysis of an approximately 100–150-nt long RNA sequence. Thus, multiple primers have to be used to perform a complete analysis of a longer RNA. The region of the RNA targeted by primers must be

accessible and not hidden in a stable secondary structure, such as, for instance, a long hairpin. To optimize the hybridization, the lengths of primers are adjusted to obtain a T_m comprised between 55 and 70°. Furthermore, primer sequences have to be compared with genomic sequence databases, when available, to avoid nonspecific hybridization.

2. Approximately 100 ng of each primer are 5'-end labeled by use of 10 μCi of [γ - ^{32}P]ATP (3000 Ci/mmol) and 10 U of T4 PNK in T4 PNK buffer. The purification of radiolabeled primers on 10% polyacrylamide denaturing gels usually improves the quality of the resulting cDNA profile. However, specific care must be taken to avoid primer contamination by polyacrylamide that can inhibit the RT.
3. The primer is eluted from the gel slice in the elution buffer at 4° overnight and ethanol precipitated. Approximately one fifth of the eluted primer (20 ng) is then used for one extension reaction. Hybridization of the primer is performed with 2–50 μg of total RNA (depending on the concentration of the target RNA within total RNA) or 2–50 pmol of purified transcript, in the hybridization buffer, without MgCl_2 , to prevent stable folding of the target RNA and Mg^{2+} -dependent hydrolysis of RNA at 65°. The hybridization step is carried out in a final volume of 2.5 μl by incubation for 10 min at 65°, followed by a slow decrease to reach room temperature.
4. The extension step is performed by addition of a 2.5 μl mixture containing 40 μM of each dNTP (0.2 μl of stock solution), 0.5 U of AMV reverse transcriptase (0.025 μl), in 1 $\times$ primer extension buffer (0.25 μl of 10 $\times$ stock). The conditions are similar to the hybridization step except the presence of 3 mM MgCl_2 required for the enzymatic activity. The extension step is carried out in a 5 μl final volume by incubation for 1 h at 42°.

To prepare a sequencing ladder, four tubes are prepared as previously indicated, but 0.5 μl of either ddATP, ddCTP, ddGTP, or ddTTP (0.5 mM each) is added to the 5 μl final volume before incubation at 42°. The extension reaction is stopped by addition of 3 μl of formamide blue.

5. The extension mixture is heated for 2 min at 96° to denature secondary structure in the cDNA and placed on ice to prevent renaturation. Fractionation is achieved on a 7% denaturing (8 M urea) polyacrylamide (19:1 ratio acrylamide/bisacrylamide) gel by loading of 2 μl of the sample. The remaining 6 μl may be used later for another gel load. The electrophoresis is carried out in TBE buffer at 40–50° by use of a constant power of 100 W (approximately 30–50 V/cm). The migration time is set according to the distance between the primer-annealing site and the sequence to be analyzed. Gels are then transferred onto Whatman 3 MM paper sheets and dried at 80° with a vacuum drier.
6. Resulting stops in cDNA synthesis are visualized by autoradiography or by exposure to PhosphorImager screens for a few hours.

2.2.4. Comments

Several preliminary tests must be performed to define the best oligonucleotide primers for reverse transcription. To test the primer quality, RNA sequencing is first done by primer extension. Only primers that give clearly readable sequencing profiles should be used for further experiments.

It is noteworthy that the analysis of highly structured G/C-rich RNA regions by primer extension may be rather difficult because the extension temperature does not exceed 42°. At this temperature, RT cannot denature stable RNA structure during elongation. Other RT enzymes that can be used at higher temperatures (50–55°, HIV RT and other RT mutants available from various producers) bring only a moderate improvement.

Naturally modified nucleotides in RNAs may also generate partial or total RT stops; thus the use of different primers that are located upstream or overlap those modified residues can be required.

When a large amount of total RNA is taken for primer extension, the use of a DNase-free RNase treatment before loading on the gel is recommended to reduce smearing of the cDNA pattern.



3. RNA EXTRACTION FROM VARIOUS CELL TYPES

Numerous protocols exist for extraction of total RNA from different cell types. In this chapter, we will give two techniques: the first one uses guanidine thiocyanate and can be used for the RNA extraction from eukaryal and archaeal cells. The second one applies to the yeast *S. cerevisiae*. Extraction of bacterial (*Escherichia coli*) total RNA may be performed by hot phenol treatment (Aiba *et al.*, 1981). Other extraction procedures may also be used (see, for examples, Mangan *et al.*, 1997; Rivas *et al.*, 2001; Internet resource http://www.protocolonline.org/prot/Molecular_Biology/RNA/RNA_Extraction/Total_RNA_Isolation/index.html). Whatever protocol is used, particular care has to be taken to avoid possible degradation of RNAs by nucleases. For this reason, the use of strong denaturing agents like guanidine thiocyanate and strong detergents generally provides total RNA of better quality.

3.1. Protocol for the extraction of total RNA from archaeal or eukaryal cells

This protocol is adapted from Chomczynski and Sacchi (1987).

3.1.1. Equipment

Vacuum drier (SpeedVac).

UV spectrophotometer for small volumes.

3.1.2. Reagents

2 M Sodium acetate, pH 4.0.

0.75 M Sodium citrate, pH 7.0, sterilized (Merck).

10% Sarcosine (dissolved at 65° in distilled water) sterilized (Sigma).

RNase free 96% ethanol.

Guanidine thiocyanate (Fluka).

2-mercaptoethanol (Merck).

Phenol-chloroform-isoamyl alcohol (125/24/1, v/v/v) saturated with water.

3.1.3. Procedure

1. The stock solution for extraction (solution S) is prepared by mixing of 100 g guanidine thiocyanate dissolved in 117.2 ml of sterilized water, with 7.04 ml of 0.75 M sodium citrate and 10.56 ml of a 10% sarcosine solution. This mixture can be stored at room temperature. The denaturing solution (solution D) is obtained by adding 72 μ l of 2-mercaptoethanol to 10 ml of solution S. The solution D can be stored for 1 month.
2. A pellet of cells (from 1×10^8 – 1×10^{10} cells, depending on cell type), coming from cultures or tissues is resuspended in 4 ml of solution D. It is then separated into ~500- μ l aliquots in eight Eppendorf tubes. Then 50 μ l of 2 M sodium acetate, pH 4.0, and 500 μ l of a phenol-chloroform-isoamyl alcohol mixture (125/24/1) are added to each tube. The tubes are vigorously vortexed and kept on ice for 15 min. The samples are then centrifuged at 15,000g for 20 min at 4°.
3. The aqueous phases are transferred in new tubes, and 2 volumes of 96% ethanol are added. The tubes are placed for more than 30 min at –20° and then centrifuged at 15,000g for 20 min at 4°.
4. The supernatant is removed and the pellet is dissolved in 50 μ l of solution D. All the RNA solutions obtained are gathered into a single tube and 2 volumes of 96% ethanol are added. The tube is placed for more than 30 min at –20° and centrifuged at 4° (15,000g) for 10 min.
5. The pellet is washed by 500 μ l of 70% ethanol and centrifuged at 15,000g for 10 min at 4°. The supernatant is discarded, and residual ethanol is completely evaporated with a SpeedVac.
6. The RNA pellet is dissolved in 80–160 μ l of sterile water. The RNA concentration is determined by measuring the A_{260} and A_{280} with a UV spectrophotometer.

3.1.4. Comments

The presence of guanidine thiocyanate during extraction inhibits the activity of RNases. However, after the removal of guanidine thiocyanate one has to take special care to avoid contamination of samples by RNases coming from

human skin. We recommend wearing gloves and using, when possible, sterilized material and reagents. It is also important to avoid unnecessary incubation at room temperature and also at 37°, the optimal temperature for human and bacterial RNases.

The quality of the RNA sample is estimated using the A_{260}/A_{280} ratio.

RNA can be considered as pure for values between 1.8 and 2.

The use of guanidine thiocyanate requires particular precautions for safety reasons.

The use of an RNase-free DNase I can ensure the absence of DNA in the sample.

3.2. Protocol for the extraction of total RNA from yeast cells

3.2.1. Equipment

Vacuum drier (SpeedVac).

Vortex.

Water bath at 65°.

UV spectrophotometer.

3.2.2. Reagents

3 M Sodium acetate, pH 5.2.

RNase free 96% ethanol.

Phenol-chloroform-isoamyl alcohol (125/24/1, v/v/v) saturated with water.

5 M NaCl.

1 M Tris-HCl, pH 7.5.

0.5 M EDTA.

10% SDS.

Acid-washed glass beads (0.45–0.5 mm in diameter, SIGMA-ALDRICH ref# G8772).

3.2.3. Procedure

1. Prepare a stock solution of extraction buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM EDTA).
2. A pellet of yeast cells, from a 25-ml culture in YPD medium grown to an A_{600} of 0.8, is resuspended in 200 μ l of extraction buffer. Add 300 μ l of glass beads and vortex vigorously twice for 1 min with a break of 2 min on ice in between.
3. Add 200 μ l of extraction buffer, 50 μ l of 10% SDS and 400 μ l of phenol-chloroform-isoamyl alcohol mixture. Carefully close the tubes (with a special cap that clamp the lid and the body of the tube or use Safe-Lock

- Eppendorf Tubes), vortex 1 min and incubate 10 min at 65°. The tubes are then centrifuged at 15,000g for 15 min at 4°.
4. The aqueous phase is transferred in a new tube, and a second phenol-chloroform-isoamyl alcohol extraction is performed.
 5. The RNAs present in the aqueous phase obtained after the second extraction are then precipitated by addition of 1 ml of ethanol in the presence of 100 mM sodium acetate (14 μ l of a 3 M solution). The tube is placed for 30 min at -80° and centrifuged at 15,000g for 15 min at 4°.
 6. The pellet is washed with 500 μ l of 70% ethanol and centrifuged at 15,000g for 10 min at 4°. The supernatant is discarded, and residual ethanol is completely evaporated with a SpeedVac.
 7. The pellet is resuspended in 50 μ l of sterile water. The RNA concentration is determined by measuring the A_{260} and A_{280} with a UV spectrophotometer.

4. RNA MODIFICATIONS DETECTABLE AFTER SPECIFIC CHEMICAL TREATMENT

This group of techniques uses preliminary chemical treatment of RNA to achieve selective modification of specific RNA residues. Up to now, reliable methods of this type have been developed only for a limited number of modified residues (see [Figs. 2.2 and 2.3](#) and [Table 2.1](#)). However, these modified residues are rather frequently encountered in many cellular RNAs. It should also be mentioned that the development of new specific chemical modification methods for the detection of other types of modified nucleotides is an open field of research.

4.1. Detection of inosine using glyoxal treatment followed by RNase T1 hydrolysis

The modified nucleotide inosine (I) is derived from adenosine (A) by enzymatic deamination catalyzed by specific adenosine deaminases ([Bass, 1995](#); [Maas *et al.*, 2003](#); [Schaub and Keller, 2002](#); [Seeburg, 2002](#)). RNase T1 does not distinguish inosine residues present in RNA from guanosine residues and cleaves on their 3' side. However, preliminary specific reaction with glyoxal allows RNase T1 to distinguish G and I residues ([Morse and Bass, 1997](#)). Indeed, glyoxal reacts with N1 and N2 of G residues and thus abolishes their recognition by RNase T1. The resulting covalent adduct is stabilized by boric acid ([Fig. 2.3A](#)). In contrast, inosine cannot react in the same way with glyoxal because of the absence of NH₂-group at position 2 and the glyoxal-inosine adduct is unstable. Thus, RNase T1 cleavage of

glyoxal-treated RNA proceeds only at inosine residues and not at G residues. The cleavage positions may be detected by primer extension analysis or other methods.

When the genomic sequence is available, an alternative approach for inosine detection consists in direct RNA sequencing with RT and comparison of the cDNA sequence obtained with the corresponding genomic sequence. Because of the absence of amino-group at position 6, I base

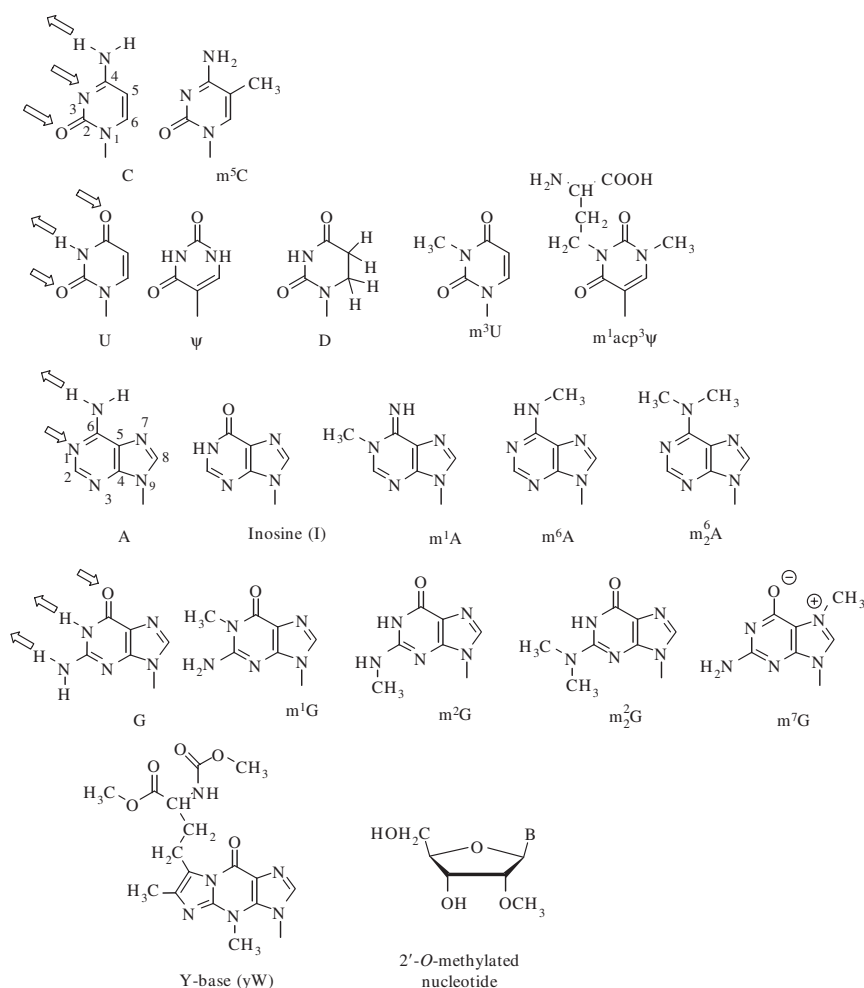
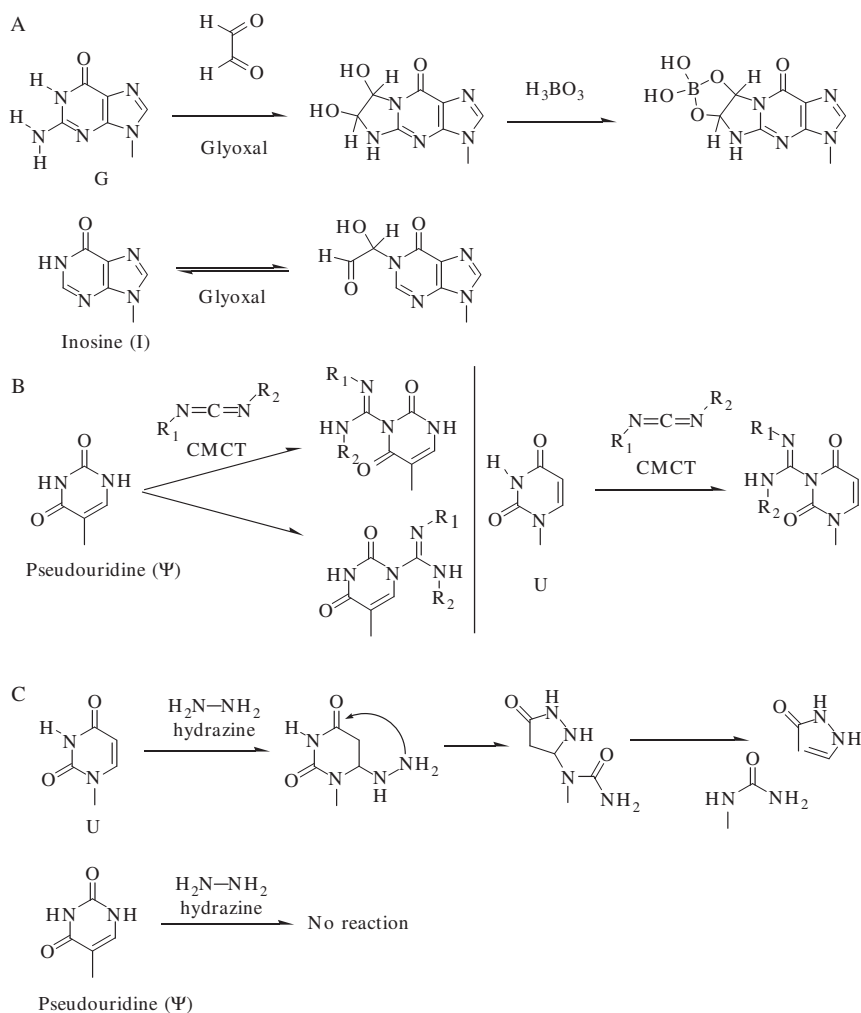


Figure 2.2 Chemical structures of selected modified nucleotides discussed in this review. Oxygen and nitrogen atoms used to form hydrogen bonds during reverse transcription are shown by arrows.

pairs with C residues instead of U residues and thus changes the sequence of the cDNA synthesized by RT. Thus one can detect unexpected A→G changes in the cDNA compared with the genomic sequence (Saccomanno and Bass, 1999). To confirm the data, it is also possible to compare the cDNA sequence obtained by direct sequencing of an unmodified transcript produced by *in vitro* transcription with the cDNA sequence obtained with the authentic RNA.



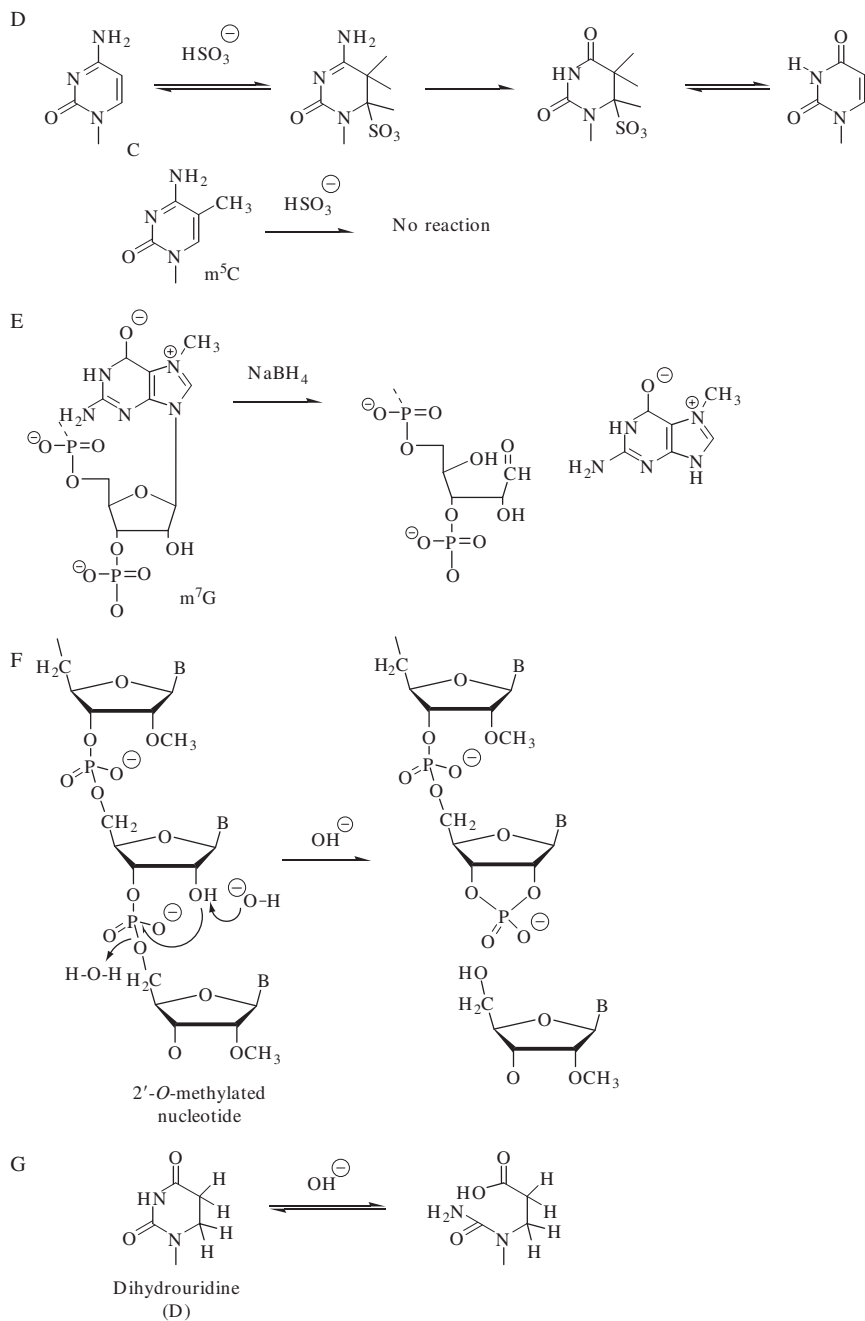


Figure 2.3 (continued)

4.2. Detection of pseudouridine (Ψ) residues

4.2.1. CMCT modification

Pseudouridine (5-[β -d-ribofuranosyl]uracil, abbreviation Ψ) residues are widespread in cellular RNAs. In contrast to all other known chemical modifications in RNAs, Ψ residues (and their derivatives) have an unusual C–C glycosidic bond linking the base to the ribose (Fig. 2.2, Cohn, 1960). Moreover, Ψ residues are formed by posttranscriptional isomerization of U residues, and thus both residues have the same molecular mass and base-pairing properties, which complicate experimental detection of Ψ residues. However, several methods of chemical treatment distinguishing U and Ψ have been developed.

The technique that is now generally used for detection of Ψ residues in RNA was initially developed for analysis of ribosomal RNA from various organisms (Bakin and Ofengand, 1993; 1998). RNA molecules are modified by a soluble carbodiimide (CMCT), which preferentially reacts with U and Ψ residues, and, to very limited extent, with G residues (Ho and Gilham, 1971; Metz and Brown, 1969a,b,c) (Fig. 2.3B). Extended incubation at alkaline pH (pH 10.3) of the CMCT-modified RNA allows the hydrolysis of U-CMT adducts, which are less stable than Ψ -CMC adducts. The remaining bulky CMC residues are then detected by primer extension using reverse transcriptase.

4.3. Protocol for the detection of pseudouridine residues in RNA using CMCT modification

4.3.1. Equipment

Water bath set to 37°.

Vacuum drier (SpeedVac).

4.3.2. Reagents

CMCT buffer (50 mM Bicine, pH 8.0, 7 M Urea, 4 mM EDTA).

CMCT solution (14 mg CMCT dissolved in 200 μ l of CMCT buffer).

Figure 2.3 Chemical reactions specific to modified residues in RNAs that may be used for their detection. (A) Reaction of G or inosine (I) with glyoxal. (B) Modification of pseudouridine (Ψ) and uridine (U) residues by carbodiimide (CMCT). (C) Reactions of U and Ψ with hydrazine. (D) Deamination of cytosine (C) residues by reaction with bisulfite followed by alkaline treatment. (E) Depurination of 7-methylguanosine residue (m^7G) by $NaBH_4$. (F) Cleavage of phosphodiester bond under alkaline conditions, the linkage at the 3'- of 2'-O-methylated residue is resistant to cleavage. (G) Scission of the dihydrouridine (D) base ring under alkaline conditions.

Table 2.1 Common chemical modifications in tRNA, rRNA, snRNA and mRNA and their possible detection by RT-based techniques

N°	Chemical name	Abbreviation	RT block	RT pause	Chemical treatment
1	2'-O-Methyladenosine	Am	—	+ (low dNTP)	+
2	N ¹ -Methyladenosine	m ¹ A	+	+	—
3	N ² -Methyladenosine	m ² A	—	+	—
4	N ⁶ -Methyladenosine	m ⁶ A	—	+	—
5	N ⁶ , N ⁶ -Dimethyladenosine	m ₂ ⁶ A	+	+	—
6	Inosine	I	—	—	+
7	2'-O-Methylguanosine	Gm	—	+ (low dNTP)	+
8	N ¹ -Methylguanosine	m ¹ G	+	+	—
9	N ² -Methylguanosine	m ² G	—	+	—
10	N ² ,N ² -Dimethylguanosine	m ₂ ² G	+	+	—
11	N ⁷ -Methylguanosine	m ⁷ G	—	—	+
12	2'-O-Methyluridine	Um	—	+ (low dNTP)	+
13	N ³ -Methyluridine	m ³ U	+	+	—
14	5-Methyluridine (ribothymidine)	m ⁵ U (T)	—	—	+ ? ^a
15	5-Carboxymethyluridine	cm ⁵ U	—	—	+ ? ^a
16	Dihydrouridine	D	—	+	+
17	Pseudouridine	Ψ	—	—	+
18	N ¹ -Methylpseudouridine	m ¹ Ψ	—	—	+
19	N ³ -Methylpseudouridine	m ³ Ψ	+	+	+
20	N ¹ -Methyl-3-(3-amino-3-carboxypropyl)-pseudouridine	m ¹ acp ³ Ψ	+	+	+
21	2'-O-Methylcytidine	Cm	—	+ (low dNTP)	+
22	N ³ -Methylcytidine	m ³ C	+	+	—
23	N ⁴ -Methylcytidine	m ⁴ C	—	+ ? ^b	—
24	N ⁴ -Acetylcytidine	ac ⁴ C	—	+ ? ^b	—
25	5-Methylcytidine	m ⁵ C	—	—	+
26	5-Hydroxymethylcytidine	hm ⁵ C	—	—	+ ? ^c

^a m⁵U and cm⁵U may be modified by CMCT and the stability of respective adducts seems to be higher than for unmodified U.

^b N⁴-modified cytidines may provoke a RT-pause.

^c hm⁵C probably has the same properties as m⁵C in the bisulfite modification reaction.

Comments: Complex modifications of both base and ribose, like Im, Ψm, m⁴Cm, and ac⁴Cm are not shown.

RNase-free 96% ethanol.

300 mM Sodium acetate, pH 5.2.

2 mg/ml Glycogen (Roche, USA).

Na₂CO₃ solution at pH 10.3 (500 mM; adjusted to pH 10.3 with 500 mM NaHCO₃). A precise adjustment of the pH to this value is very important for efficient hydrolysis of CMC-U adducts in RNA.

4.3.3. Procedure

1. Pseudouridine detection in RNA requires the comparison of the extension profiles obtained for four samples. One control experiment is done without CMCT treatment ("C") and another one without alkaline treatment ("C2"). We recommend performing two assays on samples treated with CMCT for 10 min ("C10") and 20 min ("C20"), respectively. Each reaction is carried out on 2–200 µg total RNA (depending on the abundance of the analyzed RNA in total RNA) or on 2–200 pmol of purified transcripts (an amount of 20 pmol is recommended).
2. The CMCT solution should be freshly prepared a few minutes before use. The handling of CMCT requires particular care: the use of a hood and wearing gloves are recommended; 30 µl of CMC buffer is added to the control tube C, whereas 30 µl of the CMC solution is added to the C2, C10, and C20 tubes. Samples are incubated at 37°: C2 for 2 min, C10 for 10 min, C and C20 both for 20 min.
3. The CMCT treatment is stopped by incubation on ice and by addition of 100 µl of 300 mM sodium acetate at pH 5.2, 700 µl of ethanol and 1 µl of glycogen at 2 mg/ml. Glycogen is used as a carrier for cDNA precipitation. The tubes are kept for at least 1 h at –80° and centrifuged at 15,000g for 15 min at 4°. The supernatant is then removed, and the pellet is washed with 500 µl of 70% ethanol. The tubes are placed once again at –80° for at least 1 h and then centrifuged at 15,000g for 15 min at 4°. The supernatant has to be carefully and completely removed before the next step to avoid remaining CMCT in the further steps.
4. The C2 assay is kept at –80°, whereas the three remaining tubes are subjected to alkaline treatment.
5. The alkaline treatment is achieved by resuspending the RNA pellet in 40 µl of 50 mM Na₂CO₃ (10-fold diluted 500 mM stock solution) and incubating at 37° for 3 h (it is important to completely take up the RNA-glycogen precipitate in the Na₂CO₃ solution by pipetting up and down). After incubation, the CMC-modified RNAs are precipitated by addition of 1 µl of glycogen solution, 100 µl of 300 mM sodium acetate and 700 µl of 96% ethanol. After a 1 h precipitation at –80°, the RNA pellet is collected by centrifugation at 15,000g for 15 min at 4°. The

- supernatant is completely removed and residual ethanol is evaporated with a SpeedVac.
6. The precipitated RNA is now ready for primer extension analysis (see previously). The extension products of the four reactions are fractionated in parallel on a denaturing 7% polyacrylamide-urea gel, together with a sequencing ladder obtained with the same primer.

4.3.4. Comments

Because the CMCT-modification protocol involves three successive ethanol precipitations, and because only small amounts of RNA are sometimes analyzed, particular care has to be taken to get an efficient RNA recovery at each precipitation step. To this end, the presence of glycogen, long precipitation times at -80° , centrifugation at 4° , handling in ice to avoid RNA degradation, the use of gloves and sterilized material are recommended. Note that the use of total tRNA instead of glycogen as precipitation carrier is not recommended because primers might anneal to tRNAs.

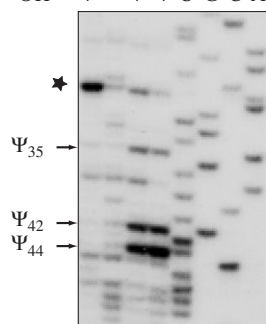
4.3.5. Analysis pitfalls

First, the sequencing ladder has to be clearly readable to correlate the RT stops to a position in the sequence. It is important to remember that a CMCT-modified base prevents any Watson-Crick base pairing with dNTPs, and, therefore, the RT extension stops one base downstream from the CMCT modified base (Fig. 2.1). For that reason, there is a one-nucleotide shift between the sequencing ladder by use of ddNTPs and the stops corresponding to CMC-U and CMC- Ψ adducts in CMCT-treated RNA. Primer extension can display strong pauses, for a given position, in each of the sequencing ladder lanes and also within the C, C2, C10, and C20 lanes. These pauses are natural RT pauses and might result from RNA degradation or from the presence of stable RNA secondary structure. Therefore, strong natural RT pauses may prevent the identification of pseudouridine. The comparison of lane C2 with the control lane C reveals the presence of U and Ψ residues, whereas the comparison of lane C2 with lanes C10 and C20 enables the distinction between Ψ and U residues. Note that Ψ residues are modified to a lesser extent than U residues by CMCT. For this reason, the pauses generated by Ψ residues are quite faint in lane C2. However, because of the increased stability of the CMC- Ψ adduct, these residues give a stronger RT stop after alkaline treatment. As the CMCT modification of Ψ residue is not complete after 10 min, a stronger RT stop should be observed in lane C20 corresponding to a 20 min-step reaction. Finally, it is essential to verify the reproducibility of the results to confirm the identification of Ψ residues in RNA. We recommend performing three independent experiments.

Several examples of Ψ detection by CMCT-RT are presented in Fig. 2.4. Primer extension was performed on U2 snRNA and two tRNAs from *S. cerevisiae* by use of specific oligonucleotides. As one can see in these three panels, natural pauses of RT are rather frequent and may considerably

A *S. cerevisiae* snRNA U2

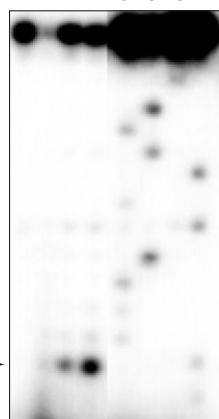
CMCT	-	+	+	+			
Time	20	2	10	20			
OH ⁻	+	-	+	+	U	G	C A



C C10
C2 C20

B *S. cerevisiae* tRNA^{Asp}

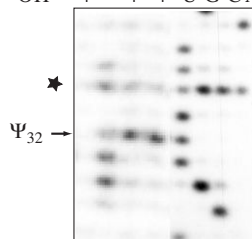
CMCT	-	+	+	+			
Time	20	2	10	20			
OH ⁻	+	-	+	+	U	G	C A



C C10
C2 C20

C *S. cerevisiae* tRNA^{Ser} mito

CMCT	-	+	+	+			
Time	20	2	10	20			
OH ⁻	+	-	+	+	U	G	C A



C C10
C2 C20

Figure 2.4 Examples of pseudouridine (Ψ) detection in RNA by the CMCT-RT approach. The primer extensions were performed on total *S. cerevisiae* RNA fractions that were treated with CMCT for 2, 10, and 20 min (lanes C2, C10, C20) followed or not by alkaline treatment (+ or -). Lane C is a control extension on RNA incubated for 20 min in CMCT buffer in the absence of CMCT and subjected to alkaline treatment. Oligonucleotide primers were specific for U2 snRNA (A) and two tRNAs (cytoplasmic tRNA^{Asp} and mitochondrial tRNA^{Ser}) (B and C). Sequencing ladders (lanes U, G, C, A) were prepared with the same primers. Natural RT pauses are indicated by stars. Arrows indicate the pauses corresponding to the identified Ψ residues.

complicate the analysis. However, the presence of Ψ residues is characterized by strong pauses in lanes C10 and C20 that are not detected in control lanes C and C2. Furthermore, in most cases, increased intensity of the RT pauses is seen in lane C20 compared with lane C10 (see Ψ_{44} in snRNA U2 or Ψ_{13} in tRNA^{Asp}). In some instances, the pauses in lanes C10 and C20 are accompanied by pauses in lanes C and C2. In these cases it may be difficult to conclude whether a Ψ residue is present at this position or not (see positions marked by an asterisk in primer extension profile for U2 snRNA and mitochondrial tRNA^{Ser}, Fig. 2.4A and C).

4.4. Complementary approaches for Ψ detection

Another approach has been used in the past for Ψ detection in RNA. After the development of CMCT-modification approach, its use is considerably reduced, but it still may be applied to confirm CMCT-RT results. The method is based on the different reactivities of U and Ψ residues to hydrazine (Peattie, 1979). Hydrazine reacts with the double C5=C6 bond in pyrimidine ring, leading to subsequent aniline cleavage of the RNA polynucleotide chain at every U residue (Fig. 2.3C). In contrast, Ψ residues do not react with hydrazine, and thus cleavages of the RNA are observed at U, but not at Ψ residues. Hydrazine cleavage was successfully used for the detection of two Ψ residues present at the 5'-end of U1 snRNA (Branlant *et al.*, 1980). Later, hydrazine treatment was combined with primer extension analysis to identify Ψ residues in U5 snRNA from *Physarum polycephalum* (Szkukalek *et al.*, 1996).

Alternative techniques for selective modification of pseudouridine residues in RNA have been developed recently (Emmereichs *et al.*, 2005; Mengel-Jorgensen and Kirpekar, 2002). In both cases, chemical modification is based on the particular reactivity of pseudouridine nitrogen N1 with the activated C=C bond of acrylonitrile (Mengel-Jorgensen and Kirpekar, 2002) or methylvinylsulfone (Emmereichs *et al.*, 2005). Both techniques were used for pseudouridine detection by mass spectrometry, but such a specific modification may probably also be used for primer extension analysis, although this has not been examined yet.

4.5. Detection of 5-methylcytosine (m⁵C)

The presence of base methylations at position 5 of pyrimidines in RNA is quite difficult to detect, because the CH₃ group at this position does not change the base-pairing properties of the nucleotide. The only methylation that is detectable using preliminary chemical treatment is that of m⁵C. This modified nucleotide is quite common in various cellular RNAs and is also

present in DNA. The techniques used for m^5C detection in DNA were developed a long time ago (for review, Rein *et al.*, 1998; Thomassin *et al.*, 1999). Besides m^5C -specific PCR and various HPLC- and MS-based techniques, DNA polymerase-based approaches that use primer extension were also proposed. These methods are based on the differential reactivity of m^5C compared with T (m^5U) and C. The detection of m^5C in DNA is performed using a specific oxidation of these residues by potassium permanganate ($KMnO_4$) (Hayatsu *et al.*, 1991) or by selective degradation of C and T, but not m^5C , residues by a hydrazine or bisulfite treatment. The reaction with bisulfite converts C (but not m^5C) into U residues. These techniques generally give reliable results for DNA, but their possible application to RNA has not been tested except for bisulfite treatment.

The bisulfite treatment technique is based on the low reactivity of m^5C compared with C residues with HSO_3^- ions in neutral or acidic conditions. Indeed, cytidine and some of its derivatives (like m^3C , Cm and ac^4C) form adducts, which are quite unstable and rapidly decompose at alkaline pH, concomitantly with chemical deamination of the cytosine ring. Such a chemical deamination converts these nucleotides to U or derivatives of U (Fig. 2.3D). The use of this technique for RNA was, until recently, limited by rather harsh reaction conditions (especially high pH), which were incompatible with the stability of phosphodiester bonds in RNA.

However, a technique based on bisulfite modification was recently proposed for m^5C mapping in RNA (Gu *et al.*, 2005). The bulk RNA is first incubated in the presence of sodium bisulfite under mildly acidic conditions (pH 5.1), followed by desalting and desulfonation at pH 9.0. These conditions lead to complete conversion of all C residues into U. The remaining m^5C residues in RNA are detected by primer extension by use of ddGTP. It should be noted that bisulfite treatment leads to complete conversion of all C residues in RNA into U residues. Thus, the sequence of the specific primer used for RNA analysis should be adapted to include the expected sequence changes because of this complete C→U conversion.

4.6. Detection of 7-methylguanine (m^7G)

To detect m^7G modifications in RNA, one can take advantage of their particular sensitivity to $NaBH_4$ reduction (Wintermeyer and Zachau, 1970, 1975). Such a treatment was used in the past for direct chemical RNA sequencing (Peattie, 1979). In this method, G residues in RNA are first methylated by dimethylsulfate (DMS) at position 7 and the resulting m^7G residues are cleaved by $NaBH_4$ reduction, followed by β -elimination with aniline (Fig. 2.3E). A similar technique without preliminary DMS methylation has been used for the detection of naturally present m^7G residues in yeast tRNAs (Alexandrov *et al.*, 2002).

4.7. Detection of 2'-O-methylated nucleotides using OH⁻ cleavage or 2'-OH reactivity

Alkaline hydrolysis of RNA polynucleotide chains proceeds by nucleophilic attack of the 2'-OH ribose group at the nearby 3'-phosphate. The resulting intermediate is unstable and rapidly decomposes into 2',3'-cyclic phosphate leading to phosphodiester bond cleavage (Fig. 2.3F). The methylation of ribose 2'-OH decreases the reactivity of the oxygen and thus prevents phosphodiester bond cleavage almost completely. This resistance to alkaline hydrolysis of the phosphodiester bonds on the 3'-side of 2'-O-methylated nucleotides can be detected by primer extension. The 2'-O-methylated residue appears as a "gap" in the regular ladder of OH⁻ cleavage (Fig. 2.1B). In practice, it is recommended that the primer extension profiles of both an unmodified RNA transcript and the natural modified RNA be compared. However, the presence of gaps in the RT profile is frequently hidden by natural pauses of RT at the corresponding positions (see Maden *et al.*, 1995; Maden, 2001 for discussion). Thus, an alternative approach for detection of 2'-O-methylations has been developed (see later).

4.8. Protocol for detection of 2'-O-methylated residues in RNA using OH⁻ cleavage

4.8.1. Equipment

Water bath set to 42°, heating block set to 90°.
Vacuum drier (SpeedVac).

4.8.2. Reagents

300 mM Sodium acetate, pH 5.2.
2 mg/ml Glycogen (Roche, USA).
Na₂CO₃ solution at pH 10.3 (500 mM; adjusted to pH 10.3 with 500 mM NaHCO₃).
RNase-free 96% ethanol.

4.8.3. Procedure

1. Each reaction is carried out on 5–200 µg of total RNA (depending on the concentration of the analyzed RNA in total RNA) or on 5–200 pmol of purified transcripts. The stock Na₂CO₃ solution is diluted five times to get a 100 mM final concentration. The volume of RNA solution is adjusted to 20 µl with RNase-free water, and 20 µl of the diluted Na₂CO₃ solution are added. The tube is then incubated at 96° for 5 min.

2. After incubation, the RNAs are precipitated by addition of 1 μ l of glycogen solution, 100 μ l of 300 mM sodium acetate and 700 μ l of 96% ethanol. After at least 1 h incubation at -80° , the RNA pellet is collected by centrifugation at 15,000g for 15 min at 4° . The supernatant is completely removed and residual ethanol is evaporated using a SpeedVac.
3. The RNA is now ready for primer extension analysis (see previous). The products of the primer extensions are fractionated in parallel on a denaturing 7% polyacrylamide-urea gel together with a sequencing ladder obtained with the same primer.

4.8.4. Comments

Statistical RNA hydrolysis can also be performed by incubation at high temperature (95 – 100°) in neutral solutions (H_2O at pH 7.0–7.5). Incubation time ranges from 2–10 min and has to be adapted to each individual case.

4.8.5. Analysis pitfalls

Natural RT pauses may sometimes hide the absence of OH^{-} cleavage at a given position. The results obtained by OH^{-} hydrolysis should be verified by other methods, for example by use of RT with low concentrations of dNTPs (see later).

4.9. Detection of 2'-O-methylated nucleotides by use of low dNTP concentrations

The alternative approach for detection of 2'-O-methylated residues in RNA uses primer extension by reverse transcriptase at low dNTP Concentrations (Maden, 2001; Maden *et al.*, 1995).

This method is based on the observation that reverse transcriptase frequently slows down (or pauses) at such nucleotides, especially at low dNTP concentrations. These dNTP-concentration dependent stops seem to be specific to 2'-O-methylation and are rarely observed for other modified nucleotides in RNA. Primer extension with normal (unmodified) residues is much less sensitive to reduced dNTP concentration. However, the pauses observed in these conditions may correspond to regions where RT has a tendency to pause and not only to 2'-O-methylations of nucleotides. The comparison of RT profile for natural (modified) RNA and unmodified RNA transcript may help to distinguish between these two possibilities.

It is noteworthy that pauses observed at 2'-O-methylated residues are sequence dependent and, depending on the sequence context, no pause may occur at some of the 2'-O-methylated residues. Therefore, the method does not allow an exhaustive identification of all 2'-O-methylated residues in RNA molecules.

4.10. Protocol for detection of 2'-*O*-methylated residues in RNA by use of low dNTP concentrations

4.10.1. Equipment

Water bath set to 42°.

4.10.2. Procedure

1. Each primer extension reaction is carried out on 2–50 μg total RNA or 2–50 pmol of purified transcripts. Oligonucleotide labeling and purification are done as described previously.
2. Primer extension reactions are performed under the same conditions, except for the variable concentrations of dNTP in the mixture. Four concentrations are usually tested, ranging from 100–2 nM, with two intermediate concentrations of 20 nM and 10 nM.
3. A sequencing ladder, performed with the same oligonucleotide is prepared as described previously.
4. After denaturation at 94° for 2 min, 2 μl of each sample are loaded to 7% PAGE-urea gel.

4.10.3. Comments

The experiment can be performed by use of only the two most extreme concentrations of dNTPs (100 nM and 2 nM) instead of the four concentrations; nevertheless the two intermediate concentrations can be very helpful to interpret the results.

4.10.4. Analysis pitfalls

First, the sequencing ladder has to be clearly readable to correlate the RT stops to a position in the sequence. Here again the RT extension stops one base downstream from the 2'-*O*-methylated residue; therefore, there is a one nucleotide shift between the sequencing ladder using ddNTPs and the stops corresponding to the 2'-*O*-methylated residue. RNA degradation at sensitive Py–A bonds may easily induce pauses in the RT profile. One should always compare the RT profile obtained for unmodified RNA transcript, sequencing ladder, and RT profile at low dNTP concentration. Only pauses absent in the RNA sequencing ladder and in unmodified RNA may serve as an indication of a 2'-*O*-methylation.

An example of such an analysis is given in [Fig. 2.5](#). Primer extension was performed on the naturally modified human U2 snRNA (panel B) and the corresponding unmodified RNA transcript (panel A). Multiple pauses are visible both in the T7 RNA transcript and the naturally modified RNA. However, additional strong RT stops that appear or are reinforced at low

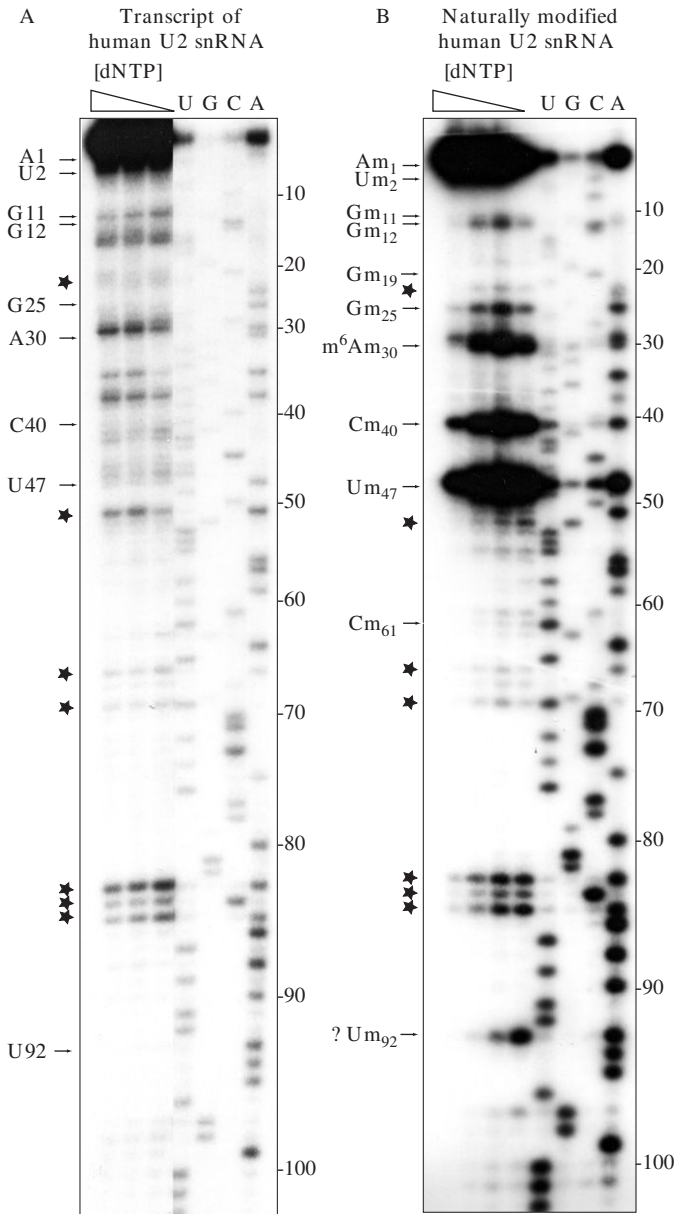


Figure 2.5 Examples of primer extension for the detection of 2'-O-methylated residues in human U2 snRNA. (A) Primer extension was performed on an unmodified U2 snRNA transcript produced with the T7 RNA polymerase. (B) The extension was performed with the same primer on the authentic U2 snRNA contained in nuclear RNA fraction from HeLa cells. The sequencing ladder is shown on the right of each panel. Natural RT pauses are indicated by stars. Locations of the known 2'-O-methylated residues

dNTP concentrations indicate the presence of 2'-O-methylated residues in human U2 snRNA. Note the great variability of the intensity of these stops; some of the 2'-O-methylations create very strong pauses whereas other ones give only a faint band in the RT profile.

4.11. Other chemical modification methods specific for 2'-O-methylations

Very recently, an alternative chemical treatment that should allow the distinction between unmethylated and 2'-O-methylated residues has been proposed (Merino *et al.*, 2005). Free unconstrained ribose 2'-OH easily reacts with *N*-methylisatoic anhydride (NMIA), and the resulting adduct can be analyzed by primer extension. For the moment, the method was only applied to RNA structural analysis. However, comparisons of NMIA modification and primer extension profiles for unmodified transcript and 2'-O-methyl-containing RNA may probably be used as a tool for 2'-O-methylation mapping.

4.12. Detection of dihydrouridine (D) residues by alkaline hydrolysis

Chemical treatment allowing the detection of D residues in RNA has been developed to study the yeast *S. cerevisiae* RNA/dihydrouridine-synthases (DUS) (Xing *et al.*, 2004). Dihydrouridine is particularly unstable at alkaline pH (pH ~13 for 5 min) and is converted into a β -ureidopropionic acid ribose derivative by pyrimidine cycle opening (Fig. 2.3G). The resulting product does not base pair with other nucleotides and may be easily detected by primer extension analysis. Another property of D residues that can be used for their detection is their stability to hydrazine treatment in contrast to uridine, but until now, this technique has not been used for primer extension analysis.

of human U2 snRNA are indicated on the left of (B). Altogether, 10 2'-O-methylated residues have been localized in human U2 snRNA (Reddy *et al.*, 1981; Westin *et al.*, 1984). Five of them can be unambiguously detected (Gm₁₂, Gm₂₅, m⁶Am₃₀, Cm₄₀, and Um₄₇). Two residues are located too close to the RNA 5'-end (Am₁ and Um₂) to be detected by this approach, and three remaining residues (Gm₁₁, Gm₁₉ and Cm₆₁) are not detected by RTat low (dNTP), because only very weak pauses are observed at the corresponding positions. Note the presence of (dNTP)-dependent stop at position 93, which may indicate the existence of an additional 2'-O-methylation at Um₉₂.

5. RNA MODIFICATIONS LEADING TO RT PAUSES WITHOUT PRELIMINARY CHEMICAL TREATMENT

Many modified nucleotides present in RNA molecules are not able to form base-pairing interactions with other nucleotides because of the presence of chemical groups on the Watson-Crick face of the base (positions 1, 2, and 6 in purines and 2, 3, and 4 in pyrimidines) (see Fig. 2.1 and Table 2.1). When the presence of such modified nucleotides is confirmed by the analysis of total RNA composition (HPLC or other techniques), the position of such nucleotides may be easily detected by primer extension. Examples include the localization of $\text{acp}^3\Psi$ in *D. melanogaster* 18S rRNA (Youvan and Hearst, 1981), m^6_2A in yeast 18S rRNA (Lafontaine *et al.*, 1995), m^1G_{745} in *E. coli* 23S rRNA (Gustafsson and Persson, 1998), and m^1A and m^3U in *H. marismortui* 23S RNA (Kirpekar *et al.*, 2005). Smaller RNAs, like tRNAs, were also analyzed by this method (for example, detection of m^2_2G in various yeast tRNAs [see Ansmant *et al.*, 2001; Behm-Ansmant *et al.*, 2004] and m^1G_9 or Y_{37} -base in yeast tRNA [Jackman *et al.*, 2003]).

Some other modified nucleotides may still base pair with complementary nucleotides, despite the presence of additional chemical groups at Watson-Crick positions. However, these additional groups may affect base-pairing stability, particularly when modification occurs at position 2 of G or 6 of A. Such modified residues are also detected by primer extension, but the intensity of the corresponding pause in the primer extension profile is much less pronounced than in the case of complete blockage of primer extension. Several examples of RT-based detection of this type have been described in literature, such as a kinetic pausing of RT by m^2G residues in rRNA (Brimacombe *et al.*, 1993; Youvan and Hearst, 1979) or various modifications of A at N6 (i^6A , m^6A , t^6A , etc.).

6. CONCLUSION

RT-based techniques applied to the analysis of RNA modification represent a powerful tool for the localization of modified residues in low-abundance cellular RNAs. At the current stage of development of these techniques, specific chemical treatment is only proposed for a few frequently encountered modified bases and 2'-O-ribose methylations. A specific treatment allowing the detection of other modified residues remains to be developed.

Despite a great interest of RT-based analysis, extreme care should be taken for unambiguous interpretation of the results. Indeed, the appearance of pauses in an RT profile may only serve as an indication of the possible occurrence of modified residues at these positions. The use of other complementary techniques of RNA analysis (like HPLC and MS) may be useful to confirm the RT results.

RT-based techniques for RNA analysis may also apply not only to the analysis of naturally occurring modified nucleotides but also to the detection and localization of chemically altered nucleotides in cellular RNAs. Like DNA, cellular RNA molecules may contain oxidized or alkylated residues resulting from the action of free radicals during oxidative stress. The presence of such spontaneous chemical modifications may alter RNA properties in mRNA translation. The development of RT-based techniques for these “nonnatural” modified residues may be extremely useful for better understanding these complex cellular phenomena.

REFERENCES

- Aiba, H., Adhya, S., and de Crombrughe, B. (1981). Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. *J. Biol. Chem.* **256**, 11905–11910.
- Alexandrov, A., Martzen, M. R., and Phizicky, E. M. (2002). Two proteins that form a complex are required for 7-methylguanosine modification of yeast tRNA. *RNA* **8**, 1253–1266.
- Ansmant, I., and Motorin, Y. (2001). Identification of RNA-modification enzymes using sequence homology. *Mol. Biol. (Moscow)* **35**, 206–223.
- Ansmant, I., Motorin, Y., Massenet, S., Grosjean, H., and Branlant, C. (2001). Identification and characterization of tRNA: Ψ_{31} -synthase (Pus6p) in yeast *S. cerevisiae*. *J. Biol. Chem.* **276**, 34934–34940.
- Bakin, A., and Ofengand, J. (1993). Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: Analysis by the application of a new sequencing technique. *Biochemistry* **32**, 9754–9762.
- Bakin, A. V., and Ofengand, J. (1998). Mapping of pseudouridine residues in RNA to nucleotide resolution. *Methods Mol. Biol.* **77**, 297–309.
- Bass, B. L. (1995). RNA editing. An I for editing. *Curr. Biol.* **5**, 598–600.
- Behm-Ansmant, I., Grosjean, H., Massenet, S., Motorin, Y., and Branlant, C. (2004). Pseudouridylation at position 32 of mitochondrial and cytoplasmic tRNAs requires two distinct enzymes in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**, 52998–53006.
- Björk, G. R. (1995). Biosynthesis and function of modified nucleosides. In “tRNA: Structure, Biosynthesis and Function” (D. Söll and U. RajBhandary, eds.), pp. 165–206. ASM Press, Washington.
- Björk, G. R., Ericson, J. U., Gustafsson, C. E. D., Hagervall, T. G., Jonsson, Y. H., and Wikstrom, P. M. (1987). Transfer RNA modification. *Ann. Rev. Biochem.* **56**, 263–287.
- Branch, A. D., Benenfeld, B. J., and Robertson, H. D. (1989). RNA fingerprinting. *Methods Enzymol.* **180**, 130–154.

- Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Gallinaro, H., Jacob, M., Sri-Widada, J., and Jeanteur, P. (1980). Nucleotide sequences of nuclear U1A RNAs from chicken, rat and man. *Nucleic Acids Res.* **8**, 4143–4154.
- Branlant, C., Krol, A., Machatt, M. A., Pouyet, J., Ebel, J. P., Edwards, K., and Kossel, H. (1981). Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. *Nucleic Acids Res.* **9**, 4303–4324.
- Brimacombe, R., Mitchell, P., Osswald, M., Stade, K., and Bochkariov, D. (1993). Clustering of modified nucleotides at the functional center of bacterial ribosomal RNA. *FASEB J.* **7**, 161–167.
- Brownlee, G. G., and Cartwright, E. M. (1977). Rapid gel sequencing of RNA by primed synthesis with reverse transcriptase. *J. Mol. Biol.* **114**, 93–117.
- Bruenger, E., Kowalak, J. A., Kuchino, Y., McCloskey, J. A., Mizushima, H., Stetter, K. O., and Crain, P. F. (1993). 5S rRNA modification in the hyperthermophilic archaea *Sulfolobus solfataricus* and *Pyrodicticum occultum*. *FASEB J.* **7**, 196–200.
- Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159.
- Cohn, W. E. (1960). Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: Isolation, structure, and chemical characteristics. *J. Biol. Chem.* **235**, 1488–1498.
- Denman, R., Colgan, J., Nurse, K., and Ofengand, J. (1988). Crosslinking of the anticodon of P site bound tRNA to C-1400 of *E. coli* 16S RNA does not require the participation of the 50S subunit. *Nucleic Acids Res.* **16**, 165–178.
- Emmereichs, G., Herdewijn, P., and Rozenski, J. (2005). Pseudouridine detection improvement by derivatization with methyl vinyl sulfone and capillary HPLC-mass spectrometry. *J. Chromat. B* **825**, 233–238.
- Fuchs, B., Zhang, K., Rock, M. G., Bolander, M. E., and Sarkar, G. (1999). High temperature cDNA synthesis by AMV reverse transcriptase improves the specificity of PCR. *Mol. Biotechnol.* **12**, 237–240.
- Garcia, G. A., and Goodenough-Lashua, D. M. (1998). Appendix 3: General properties of RNA-modifying and -editing enzymes. In “The Modification and Editing of RNA” (H. Grosjean and R. Benne, eds.), pp. 555–560. ASM Press, New York.
- Gehrke, C. W., and Kuo, K. C. (1989). Ribonucleoside analysis by reversed-phase high-performance liquid chromatography. *J. Chromatogr.* **471**, 3–36.
- Gehrke, C. W., Kuo, K. C., McCune, R. A., Gerhardt, K. O., and Agris, P. F. (1982). Quantitative enzymatic hydrolysis of tRNAs: Reversed-phase high-performance liquid chromatography of tRNA nucleosides. *J. Chromatogr.* **230**, 297–308.
- Grosjean, H., Motorin, Y., and Morin, A. (1998). RNA-modifying and RNA-editing enzymes: Methods for their identification. In “The Modification and Editing of RNA” (H. Grosjean and R. Benne, eds.), pp. 21–46. ASM Press, New York.
- Gu, W., Hurto, R. L., Hopper, A. K., Grayhack, E. J., and Phizicky, E. M. (2005). Depletion of *Saccharomyces cerevisiae* tRNA^{His} guanylyltransferase Thg1p leads to uncharged tRNA^{His} with additional m⁵C. *Mol. Cell. Biol.* **25**, 8191–8201.
- Gupta, R. C., and Randerath, K. (1979). Rapid print-readout technique for sequencing of RNA's containing modified nucleotides. *Nucleic Acids Res.* **6**, 3443–3458.
- Gustafsson, C., and Persson, B. C. (1998). Identification of the *rrmA* gene encoding the 23S rRNA m¹G₇₄₅ methyltransferase in *Escherichia coli* and characterization of an m¹G₇₄₅-deficient mutant. *J. Bacteriol.* **180**, 359–365.
- Hayatsu, H., Atsumi, G., Nawamura, T., Kanamitsu, S., Negishi, K., and Maeda, M. (1991). Permanganate oxidation of nucleic acid components: A reinvestigation. *Nucleic Acids Symp. Ser.* **25**, 77–78.

- Hiley, S. L., Jackman, J., Babak, T., Trocheset, M., Morris, Q. D., Phizicky, E., and Hughes, T. R. (2005). Detection and discovery of RNA modifications using microarrays. *Nucleic Acids Res.* **33**, e2.
- Ho, N. W., and Gilham, P. T. (1971). Reaction of pseudouridine and inosine with N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide. *Biochemistry* **10**, 3651–3657.
- Jackman, J. E., Montange, R. K., Malik, H. S., and Phizicky, E. M. (2003). Identification of the yeast gene encoding the tRNA m¹G methyltransferase responsible for modification at position 9. *RNA* **9**, 574–585.
- Kirpekar, F., Hansen, L. H., Rasmussen, A., Poehlsgaard, J., and Vester, B. (2005). The archaeon *Haloarcula marismortui* has few modifications in the central parts of its 23S ribosomal RNA. *J. Mol. Biol.* **348**, 563–573.
- Krol, A., Branlant, C., Lazar, E., Gallinaro, H., and Jacob, M. (1981). Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs. *Nucleic Acids Res.* **9**, 2699–2716.
- Lafontaine, D., Vandenhaute, J., and Tollervey, D. (1995). The 18S rRNA dimethylase Dim1p is required for pre-ribosomal RNA processing in yeast. *Genes Dev.* **9**, 2470–2481.
- Lane, D. J., Field, K. G., Olsen, G. J., and Pace, N. R. (1988). Reverse transcriptase sequencing of ribosomal RNA for phylogenetic analysis. *Methods Enzymol.* **167**, 138–144.
- Maas, S., Rich, A., and Nishikura, K. (2003). A-to-I RNA editing: Recent news and residual mysteries. *J. Biol. Chem.* **278**, 1391–1394.
- Maden, B. E. (2001). Mapping 2'-O-methyl groups in ribosomal RNA. *Methods* **25**, 374–382.
- Maden, B. E., Corbett, M. E., Heeney, P. A., Pugh, K., and Ajuh, P. M. (1995). Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. *Biochimie* **77**, 22–29.
- Mangan, J. A., Sole, K. M., Mitchison, D. A., and Butcher, P. D. (1997). An effective method of RNA extraction from bacteria refractory to disruption, including mycobacteria. *Nucleic Acids Res.* **25**, 675–676.
- Massenet, S., Mougin, A., and Branlant, C. (1998). Posttranscriptional modifications in the U small nuclear RNAs. In "The Modification and Editing of RNA" (H. Grosjean and R. Benne, eds.), pp. 201–227. ASM Press, New York.
- McCloskey, J. A., and Rozenski, J. (2005). The Small Subunit rRNA Modification Database. *Nucleic Acids Res.* **33**, D135–D138.
- Meng, Z., and Limbach, P. A. (2006). Mass spectrometry of RNA: Linking the genome to the proteome. *Brief Funct. Genomic Proteomic.* **5**, 87–95.
- Mengel-Jorgensen, J., and Kirpekar, F. (2002). Detection of pseudouridine and other modifications in tRNA by cyanoethylation and MALDI mass spectrometry. *Nucleic Acids Res.* **30**, e135.
- Merino, E. J., Wilkinson, K. A., Coughlan, J. L., and Weeks, K. M. (2005). RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231.
- Metz, D. H., and Brown, G. L. (1969a). The investigation of nucleic acid secondary structure by means of chemical modification with a carbodiimide reagent. I. The reaction between N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide and model nucleotides. *Biochemistry* **8**, 2312–2328.
- Metz, D. H., and Brown, G. L. (1969b). The investigation of nucleic acid secondary structure by means of chemical modification with a carbodiimide reagent. II. The reaction between N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide and transfer ribonucleic acid. *Biochemistry* **8**, 2329–2342.

- Metz, D. H., and Brown, G. L. (1969c). The tertiary structure of transfer ribonucleic acid investigated by a chemical modification technique. *Biochem. J.* **114**, 35P.
- Morse, D. P., and Bass, B. L. (1997). Detection of inosine in messenger RNA by inosine-specific cleavage. *Biochemistry* **36**, 8429–8434.
- Ofengand, J., and Fournier, M. J. (1998). The pseudouridine residues of rRNA: Number, location, biosynthesis, and function. In “The Modification and Editing of RNA” (H. Grosjean and R. Benne, eds.), pp. 229–254. ASM Press, New York.
- Peattie, D. A. (1979). Direct chemical method for sequencing RNA. *Proc. Natl. Acad. Sci. USA* **76**, 1760–1764.
- Qu, H. L., Michot, B., and Bachellerie, J. P. (1983). Improved methods for structure probing in large RNAs: A rapid ‘heterologous’ sequencing approach is coupled to the direct mapping of nuclease accessible sites. Application to the 5' terminal domain of eukaryotic 28S rRNA. *Nucleic Acids Res.* **11**, 5903–5920.
- Reddy, R., Henning, D., Epstein, P., and Busch, H. (1981). Primary and secondary structure of U2 snRNA. *Nucleic Acids Res.* **9**, 5645–5658.
- Rein, T., DePamphilis, M. L., and Zorbas, H. (1998). Identifying 5-methylcytosine and related modifications in DNA genomes. *Nucleic Acids Res.* **26**, 2255–2264.
- Rivas, R., Vizcaino, N., Buey, R. M., Mateos, P. F., Martinez-Molina, E., and Velazquez, E. (2001). An effective, rapid and simple method for total RNA extraction from bacteria and yeast. *J. Microbiol. Methods* **47**, 59–63.
- Saccomanno, L., and Bass, B. L. (1999). A minor fraction of basic fibroblast growth factor mRNA is deaminated in *Xenopus* stage VI and matured oocytes. *RNA* **5**, 39–48.
- Schaub, M., and Keller, W. (2002). RNA editing by adenosine deaminases generates RNA and protein diversity. *Biochimie* **84**, 791–803.
- Seeburg, P. H. (2002). A-to-I editing: New and old sites, functions and speculations. *Neuron* **35**, 17–20.
- Smith, J. E., Cooperman, B. S., and Mitchell, P. (1992). Methylation sites in *Escherichia coli* ribosomal RNA: Localization and identification of four new sites of methylation in 23S rRNA. *Biochemistry* **31**, 10825–10834.
- Sprinzl, M., and Vassilenko, K. S. (2005). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **33**, D139–D140.
- Stanley, J., and Vassilenko, S. (1978). A different approach to RNA sequencing. *Nature* **274**, 87–89.
- Szkukalek, A., Mougin, A., Gregoire, A., Solymosy, F., and Branlant, C. (1996). A unique U5→A substitution in the *Physarum polycephalum* U1 snRNA: Evidence at the RNA and gene levels. *Biochimie* **78**, 425–435.
- Tanaka, Y., Dyer, T. A., and Brownlee, G. G. (1980). An improved direct RNA sequence method; its application to *Vicia faba* 5.8S ribosomal RNA. *Nucleic Acids Res.* **8**, 1259–1272.
- Thomas, B., and Akoulitchiev, A. V. (2006). Mass spectrometry of RNA. *Trends Biochem. Sci.* **31**, 173–181.
- Thomassin, H., Oakeley, E. J., and Grange, T. (1999). Identification of 5-methylcytosine in complex genomes. *Methods* **19**, 465–475.
- Veldman, G. M., Klootwijk, J., de Regt, V. C., Planta, R. J., Branlant, C., Krol, A., and Ebel, J. P. (1981). The primary and secondary structure of yeast 26S rRNA. *Nucleic Acids Res.* **9**, 6935–6952.
- Westin, G., Lund, E., Murphy, J. T., Pettersson, U., and Dahlberg, J. E. (1984). Human U2 and U1 RNA genes use similar transcription signals. *EMBO J.* **3**, 3295–3301.
- Wintermeyer, W., and Zachau, H. G. (1970). A specific chemical chain scission of tRNA at 7-methylguanosine. *FEBS Lett.* **11**, 160–164.

- Wintermeyer, W., and Zachau, H. G. (1975). Tertiary structure interactions of 7-methylguanosine in yeast tRNA^{Phe} as studied by borohydride reduction. *FEBS Lett.* **58**, 306–309.
- Xing, F., Hiley, S. L., Hughes, T. R., and Phizicky, E. M. (2004). The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. *J. Biol. Chem.* **279**, 17850–17860.
- Youvan, D. C., and Hearst, J. E. (1979). Reverse transcriptase pauses at N2-methylguanine during *in vitro* transcription of *Escherichia coli* 16S ribosomal RNA. *Proc. Natl. Acad. Sci. USA* **76**, 3751–3754.
- Youvan, D. C., and Hearst, J. E. (1981). A sequence from *Drosophila melanogaster* 18S rRNA bearing the conserved hypermodified nucleoside amY: Analysis by reverse transcription and high-performance liquid chromatography. *Nucleic Acids Res.* **9**, 1723–1741.

2.3. Mutagenèse dirigée

Tous les variants d'ARNt, sRNA à boîtes H/ACA et protéines ont été obtenus par PCR à partir des plasmides pUC18, pet28B et pGEX-6P-1 comportant leurs gènes et des couples d'oligonucléotides permettant la substitution, la délétion ou l'addition de nucléotides. Les mutants obtenus ont été vérifiés par séquençage au laboratoire.

2.4. Extraits cellulaires

Les extraits cellulaires de *P. abyssi* et *S. solfataricus* sont obtenus par sonication de 10^{10} cellules de ces espèces dont la culture cellulaire a été stoppée en fin de phase exponentielle. Les culots cellulaires sont repris avec 200 ml de tampon d'extraction : Tris-HCl 25 mM, pH 7,5 ; KCl 25 mM ; $MgCl_2$ 5 mM ; Glycérol 10 % ; DTT 2 mM ; solution soumise à l'autoclave avant l'ajout de DTT. La sonication se fait par 5 séries de 20 sec en maintenant l'échantillon à froid. L'échantillon est centrifugé à 10 000 g puis le surnageant (S 10) est récupéré. La concentration en glycérol est ajustée à 20% en vue d'une conservation à $-20^{\circ}C$.

2.5. Northern Blot

Marquage de la sonde

[$\gamma^{32}P$] ATP (GE Healthcare, 10 mCi/ml, 3000 Ci/mmol)	1 μ l
Oligonucléotide à marquer (15 pmol/ μ l soit environ 100 ng/ μ l)	1 μ l
Tampon de phosphorylation 10X (Tris/HCl 0,5 M, pH 7,6 ; $MgCl_2$ 100 mM ; DTT 50 mM ; Spermidine 1 mM ; EDTA 1 mM)	1 μ l
T4 polynucléotide kinase (MBI Fermentas, 10U/ μ l)	1 μ l
H ₂ O stérile	qsp 10 μ l
Incuber 1h à $37^{\circ}C$.	

La T4 PNK est inactivée en ajoutant 10 μ l d'eau stérile et en incubant 10 min à $65^{\circ}C$.

Un marqueur de taille 100 pb+ ou un mélange d'oligonucléotides dont on connaît la taille sont également marqués. Le protocole est le même en substituant l'oligonucléotide à marquer par le marqueur de taille (1 μ g).

Séparation des ARN

10 μ g d'ARN totaux sont déposés sur un gel de polyacrylamide dénaturant 6% (acrylamide/bisacrylamide (38:2) ; urée 8M ; TBE 0,5X) en présence de deux volumes de bleu de formamide, après dénaturation 1 min à $96^{\circ}C$ et refroidissement rapide dans la glace. La migration a lieu à 12 mA à température ambiante durant 40 min dans le tampon TBE 0,5X stérile. Le marqueur de taille marqué peut être déposé en parallèle.

Transfert sur membrane de nylon Hybond N+ (Electroblotting)

Après la migration, les ARN sont transférés électriquement (Biorad Trans Blot SD SemiDry Transfer Cell) sur une membrane de nylon (Hybond N+, 0,45 μ m, Pharmacia) en tampon TBE 0,5x pendant 30 min à une intensité de 3 mA/cm². Les ARN sont fixés de façon covalente à la membrane par pontage photochimique aux UV pendant 5 min.

Pré-hybridation

La membrane est pré-hybridée 1 heure à 37°C sous agitation dans un tube à hybridation en présence de 4 ml de tampon de pré-hybridation composé de 30 ml de SSPE 20x (NaCl 175,3 g ; Na₂HPO₄, 2H₂O 27,6 g ; EDTA 7,4 g ; H₂O qsp 1 l, ajusté à pH 7,4) ; 5 ml de réactif de Denhardt 100X (Ficoll 5 g ; PolyVinylPyrrolidone 5g ; BSA en poudre 5 g ; H₂O qsp 250 ml) ; 5 ml de SDS 10% ; H₂O qsp 100 ml ; préchauffé à 37°C.

Hybridation

L'oligonucléotide marqué en 5' est dénaturé par incubation 2 min à 96°C puis refroidissement dans la glace. Puis, 1 ml de tampon de pré-hybridation est ajouté à l'oligonucléotide marqué et le mélange est filtré à l'aide d'une seringue et d'un filtre de 2 μ m (de façon à réduire le bruit de fond sur la membrane) puis est ajouté au tampon dans le tube à hybridation. L'hybridation est réalisée sur la nuit (16 heures à 37°C, d'autres températures sont testées) sous agitation dans un four à hybridation. La membrane est alors lavée 3 à 4 fois pendant 5 min avec 7 ml de tampon de lavage SSPE 6x chauffé à 42°C et placée sous agitation. La membrane humide est emballée dans un plastique hermétique avant d'être exposée en cassette d'un système de type PhosphorImager (Typhoon, Amersham Biosciences).

Déshybridation et réhybridation

Il est possible de réutiliser la membrane en éliminant la sonde fixée par deux lavages de la membrane avec une solution de SDS 0,1% portée à ébullition, sous agitation pendant 10 min. Une exposition en cassettes de PhosphorImager permet de vérifier que la sonde est éliminée. La membrane peut à nouveau être utilisée pour une hybridation avec une autre sonde en suivant le protocole décrit précédemment (pré-hybridation, hybridation, lavages et exposition).

2.6. Transcription *in vitro*

Obtention de la matrice

La matrice ADN comporte le promoteur de l'ARN polymérase T7 (5'-TAATACGACTCACTATAGGG-3') et est obtenue par trois méthodes différentes. Une première méthode, utilisée pour la production de matrices utilisées pour la transcription de petits ARN substrats, consiste à appairer deux oligonucléotides de séquences complémentaires en

quatre étapes d'incubation : 2 min à 96°C , 15 min à 65°C, 15 min à 37°C et 15 min à 20°C. L'ADN double brin formé est purifié par électrophorèse en gel de polyacrylamide à 10% non dénaturant puis élué. Une deuxième méthode également utilisée pour la production de matrices utilisées pour la transcription d'ARN substrats consiste à associer deux oligonucléotides partiellement complémentaires et à réaliser une élongation de ceux-ci par PCR. Une troisième méthode, utilisée pour la production de matrices utilisées pour la transcription des sRNA à boîtes H/ACA et des tRNA, consiste à amplifier par PCR une partie de l'ADN génomique grâce à un couple d'oligonucléotides qui permet également l'incorporation de la séquence du promoteur T7. Les fragments amplifiés par PCR sont ensuite clonés dans un plasmide pUC18, pT-ADV ou pCR2.1. La nature de ces fragments est vérifiée par un séquençage réalisé au laboratoire. Ces plasmides sont utilisés comme matrice après linéarisation.

Transcription non radioactive

Matrice ADN (produits PCR ou plasmide linéarisé 1 µg/µl)	1 µl
Tampon de transcription 10X (100 mM MgCl ₂ , 100 mM Tris-HCl pH 7.5)	2 µl
rNTP (10 mM de chaque rATP, rCTP, rGTP, UTP)	1 µl
RNasin (RNA guard Pharmacia Biotech, 35,7 U/ml)	1 µl
T7 RNA polymérase (Ambion) (20 U/µl)	1,5 µl
H ₂ O	qsp 20 µl

Le mélange est incubé 3 heure à 37°C, puis 1 µl de DNase I (Amersham Pharmacia, 7,5 U/µl) est ajouté et le mélange est à nouveau incubé 30 min à 37°C. Les transcrits sont purifiés sur gel de polyacrylamide dénaturant 8%. Les ARN sont visualisés par ombrage aux UV (254 nm), découpés puis élués une nuit à 4°C dans du tampon d'élution à forte concentration saline puis précipités avec de l'éthanol 96% en présence d'acétate d'ammonium. Les ARN sont ensuite repris dans de l'eau stérile.

Transcription radioactive uniforme au [$\alpha^{32}P$]-ATP

Matrice ADN (produit PCR ou plasmide linéarisé 1 µg/µl)	1 µl
Tampon de transcription 10X (100 mM MgCl ₂ , 100 mM Tris-HCl pH 7.5)	2 µl
rNTP (10 mM de chaque rCTP, rGTP, UTP)	1 µl
[$\alpha^{32}P$]-ATP (Amersham, 3000 Ci/mmol ; 10 µCi/µl)	2 µl
ATP 50 µM	1 µl
RNasin (RNA guard Pharmacia Biotech, 35,7 U/ml)	1 µl
T7 RNA polymérase (Ambion) (20 U/µl)	1,5 µl
H ₂ O	qsp 20 µl

Le mélange est incubé 3 heures à 37°C, puis 1 µl de DNase I (Amersham Pharmacia, 7,5 U/µl) est ajouté et le mélange est à nouveau incubé 30 min à 37°C. Les ARN sont purifiés sur gel de polyacrylamide dénaturant 8% puis visualisés par autoradiographie. Les étapes qui

suivent sont identiques à celles décrites pour la transcription non radioactive. Les transcrits sont repris dans de l'eau stérile à 100 000 cpm/ml.

2.7. Analyse de la formation de complexes ARN/protéine

Les expériences de gel retard sont réalisées comme détaillé dans la publication N°1 présentée dans la partie « Résultats » (Charpentier *et al.*, 2005; Charpentier *et al.*, 2007). Un mélange comprenant 50 fmol d'ARN guide uniformément marqué (10 000 cpm), 0,5 µg d'ARNt de levure (Roche Molecular Biochemicals) et potentiellement l'ARN substrat froid (100 ng, 15 pmol) est incubé dans 3,5 µl de tampon D 1x, 5 min à 65°C, puis 10 min à température ambiante. Après cette étape de renaturation de l'ARN, la formation des complexes est initiée par ajout d'1 µl du mélange de protéines (concentration finale : 200 nM chacune). Les échantillons sont incubés 10 min à 65°C. Puis 4 µl de bleu CSB sont ajoutés aux échantillons qui doivent être rapidement déposés sur gel. Les complexes ARN-protéines sont fractionnés par électrophorèse en gel polyacrylamide-glycérol 6% non dénaturant. L'électrophorèse se fait à 100 V pendant 3 h à température ambiante.

Une procédure alternative consiste à marquer l'ARN substrat et non le sRNA à boîtes H/ACA. Celle-ci requiert un mélange réalisé dans du tampon D 1x comprenant 200 ng (10 pmol) de sRNA, 0,5 µg d'ARNt de levure et 150 fmoles d'ARN substrat marqué (10 000 cpm) en présence d'un mélange de protéines (200 nM chacune).

2.8. Evaluation de l'activité de pseudouridylation

Ce test d'activité repose sur la méthode du plus proche voisin (Fig. M.2). L'ARN substrat uniformément marqué en présence de l'un des 4 radio-isotopes est incubé à 65°C (ou autre température signalée) en présence d'une ARN:Ψ-synthase ou d'un complexe sRNP (Behm-Ansmant *et al.*, 2004; Charpentier *et al.*, 2005; Charpentier *et al.*, 2007).

Dans le cas d'un complexe sRNP, environ 10 pmoles soit 200 ng de sRNA à boîtes H/ACA sont mis en présence de 150 fmoles d'ARN substrat marqué (10 000 cpm) dans du tampon D1x à température ambiante. Un mélange de protéines (200 nM chacune) est ajouté dans les mêmes conditions que pour les expériences de gel retard. Le volume final de 4,5 µl est incubé à 65°C (ou autre température signalée). Lors des expériences de cinétique, un aliquot de 4,5 µl est retiré d'un mélange comportant 45 µl.

Dans le cas d'une ARN:Ψ-synthase agissant sans sRNA à boîtes H/ACA, l'ARN substrat uniformément marqué est incubé dans du tampon Psi 1x pendant 5 min à 80°C. Lorsque la température est lentement redescendue à 37°C, 0,1 à 1 µg de l'ARN:Ψ-synthase est ajouté pour déclencher la réaction. L'incubation se fait à une température en accord avec l'origine de l'ARN:Ψ-synthase.

Dans les deux cas, la réaction est arrêtée en plaçant le mélange dans la glace et par ajout de 150 μl d'acétate de sodium 0,3 M, 1 μl d'ARNt de levure à 2 $\mu\text{g}/\mu\text{l}$ et 150 μl de phénol-chloroforme (1:1) (v/v). L'ARN présent dans la phase aqueuse est précipité à l'éthanol à 96%.

L'ARN substrat est ensuite totalement digéré par une RNase afin de recueillir des mononucléotides marqués en 3' qui sont fractionnés en Chromatographie sur Couche Mince (CCM ou TLC) soit en une dimension, soit en deux dimensions (Fig. M.2). Deux RNases différentes peuvent être utilisées. La RNase T2 coupe l'ARN en libérant une extrémité 3' phosphate. Il s'agit donc de placer un phosphate marqué dans la liaison phosphodiester en 3' de l'uridine qui sera modifiée en résidu Ψ . L'hydrolyse totale par la RNase T2 est réalisée dans du tampon T2 1x avec 0,5U de RNase T2 (Invitrogen) dans 8 μl final. L'incubation s'effectue toute la nuit à 37°C. Chaque échantillon est ensuite déposé sur une plaque de cellulose en plastique ou en verre pour chromatographie en couche mince (TLC). Notons que les migrations N1/N2 ne peuvent être effectuées sur des plaques en plastique. La migration est effectuée pendant 4 h dans une cuve et à saturation dans le solvant approprié pour la 1ère dimension (Tampon N1). Si une 2^{nde} dimension est réalisée, la plaque est retournée de 90°, séchée, puis placée en présence du solvant pour la 2^{nde} dimension (Tampon N2 ou RII). Les plaques sont séchées puis exposées au PhosphoImager. Les profils de séparation des taches correspondant aux différentes modifications sont donnés en figure M 3.

La RNase P1 est une enzyme qui clive l'ARN en libérant des nucléotides 5' monophosphate. Dans ce cas, un marquage uniforme doit être effectué au $[\alpha^{32}\text{P}]\text{-UTP}$ et permet de visualiser directement le résidu Ψ formé après digestion à la RNase P1. La digestion est réalisée dans du tampon P1 1x avec 0,3U de RNase P1 dans 5 μl final. L'incubation est réalisée la nuit à 37°C. Les échantillons sont déposés sur plaque TLC, la séparation se fait en une seule dimension et les taches sont quantifiées de la même façon que lors de la digestion par la RNase T2.

La radioactivité contenue dans chaque tache correspondant à chacun des nucléotides 3' monophosphate (Fig. M.3) (Keith, 1995) est quantifiée à l'aide du logiciel « ImageQuant Software Version 5.2 » (Molecular Dynamics). Puisque la composition de la séquence ARN substrat en nucléotide est connue, il est possible de mesurer le taux de résidus U convertis en Ψ (exprimé en mole de résidu Ψ par mole de substrat) selon le tableau M.1.

Les données obtenues par les études ont été traitées à l'aide du logiciel « SigmaPlot 2000 version 6.1 » (SPSS Inc.). L'équation utilisée, $y = a(1-e^{-bx})$ est la fonction la plus simple pour ajuster au mieux les données expérimentales à un modèle mathématique. Les paramètres entrés correspondent à : x = temps en minutes, y = mole de résidu Ψ formées par mole de substrat, a = nombre maximal de moles de résidu Ψ formées par mole de substrat et correspond à l'amplitude, notée A ; b = vitesse apparente du nombre de moles de résidu Ψ formées en min^{-1} et représente une constante de vitesse apparente, notée k_{obs} .

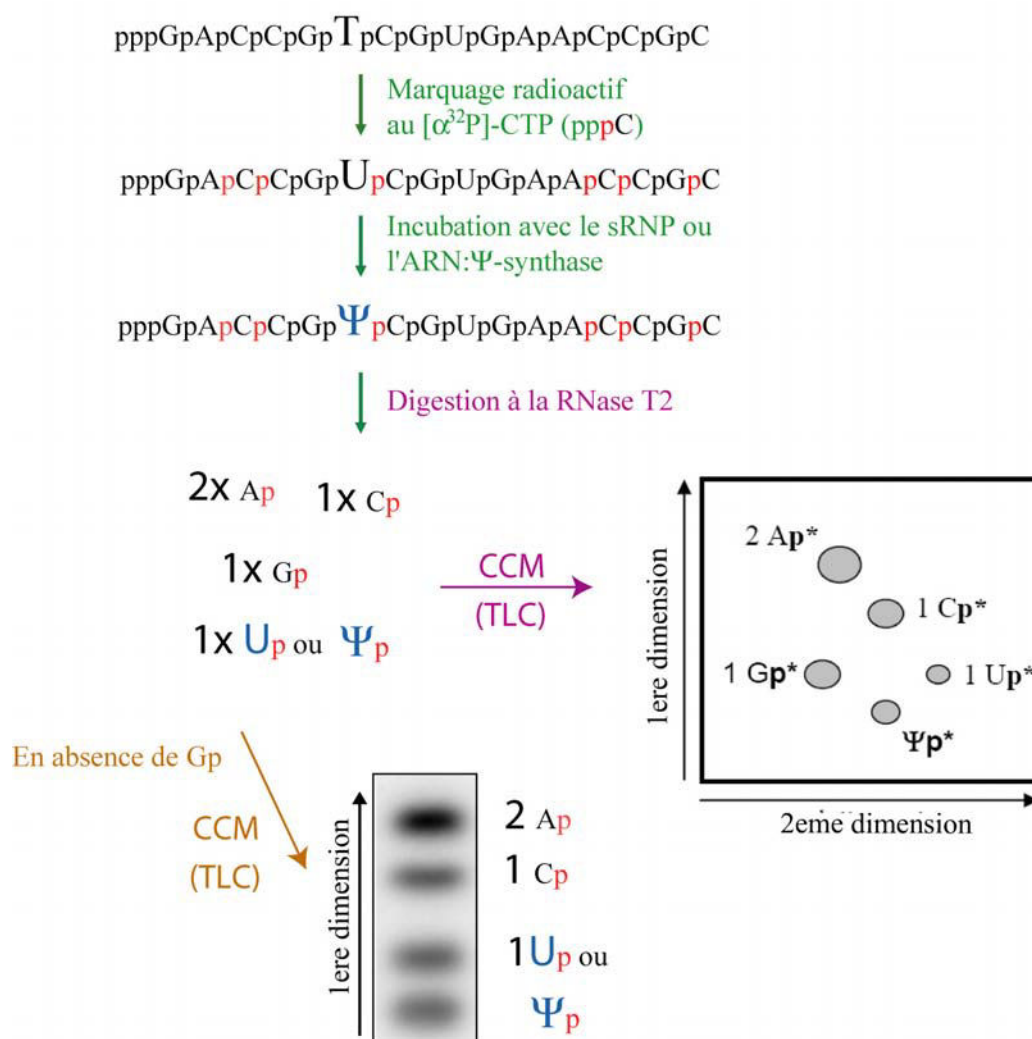


Figure M.2 : Détection des résidus Ψ par la méthode du plus proche voisin et chromatographie sur couche mince. Les phosphates marqués radioactivement sont indiqués en rouge. La chromatographie sur couche mince peut être utilisée en une seule dimension quand aucun des résidus 5'-Gp-3' marqué n'est obtenu après digestion à la RNase T2.

	A	B	C	D	E	F
	Nucléotides	Nombre de nucléotides marqués en 3'	Valeur de quantification (Image Quant)	Quantité relative	Comptage relatif (Ratio)	Rendement de la formation de pseudouridine
1	Blanc		C1			
2	A	B2	C2	$D2=(C2-C1)/B2$	$E2=D2/D6$	$F2=(C6-C1)/D2$
3	C	B3	C3	$D3=(C3-C1)/B3$	$E3=D3/D6$	$F3=(C6-C1)/D3$
4	G	B4	C4	$D4=(C4-C1)/B4$	$E4=D4/D6$	$F4=(C6-C1)/D4$
5	U	B5	C5	$D5=(C5+C6-2C1)/B5$	$E5=D4/D6$	$F5=(C6-C1)/D5$
6	Psi		C6	$D6=(D2+D3+D4+D5)/4$		$F6=(F2+F3+F4+F5)/4$

Tableau M.1 : Evaluation du taux de résidus Ψ formé à partir des quantifications des taches de CCM. Dans la colonne B sont déterminés les nombre de nucléotides A, C, G ou U dont le phosphate en 3' est marqué (voir Fig. M.2). La colonne C rassemble les quantifications provenant du logiciel Image Quant. Les ratios de la colonne E doivent être idéalement proches de 1. Une valeur qui s'écarte trop de 1 ne permet pas une quantification du taux de pseudouridylation au final car traduit un défaut dans la composition de l'ARN substrat marqué. La valeur F6 traduit le taux de modification de U en Ψ exprimé en moles U/ moles de substrat.

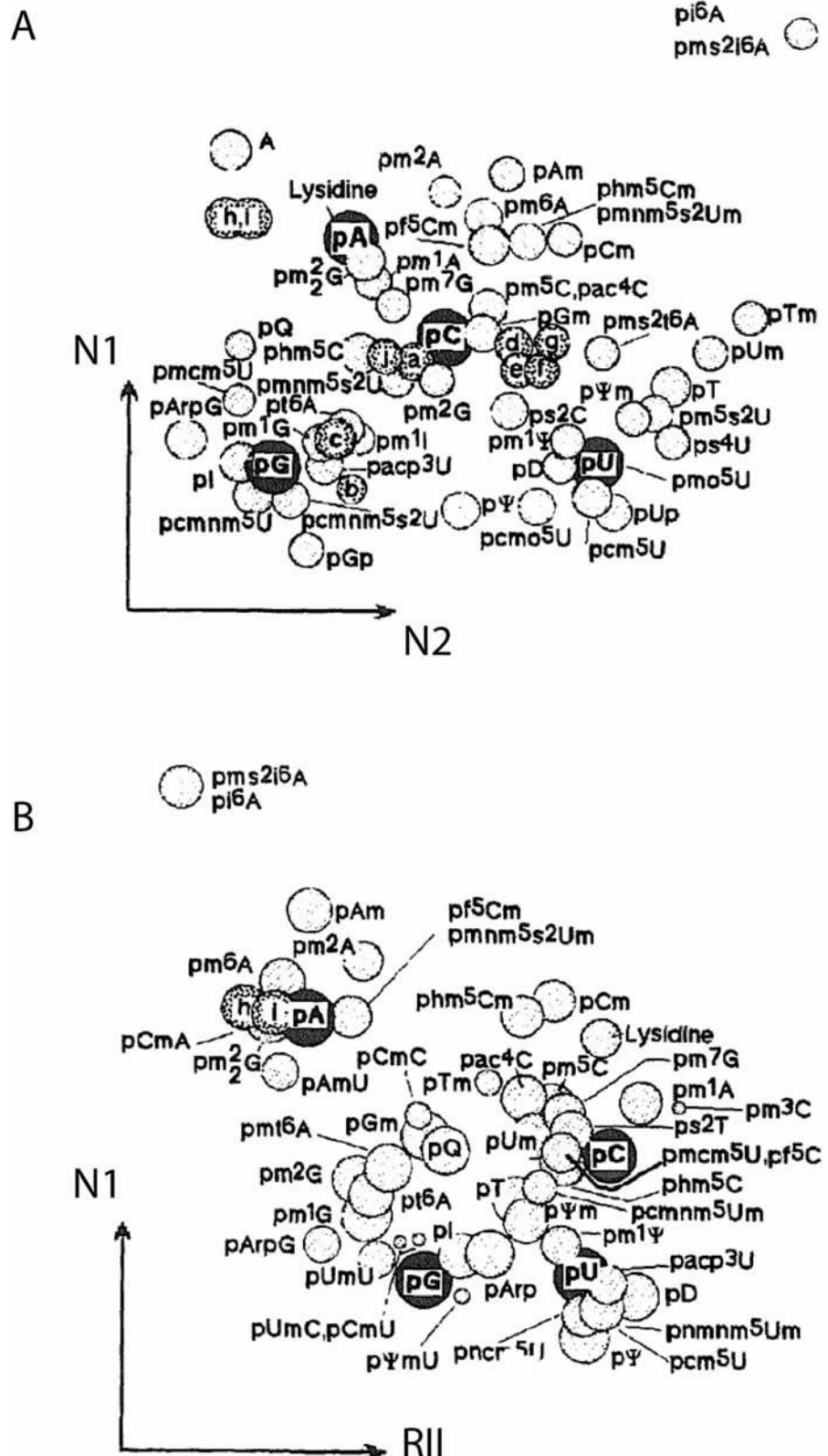


Figure M.3 : Positionnement relatif en chromatographie en deux dimensions sur couche mince des taches correspondant aux nucléotides majeurs et nucléotides modifiés de l'ARN (d'après Keith, 1995).

Deux distributions différentes sont observées selon l'utilisation de tampon N2 (A.) ou RII (B.) pour la seconde dimension de séparation par chromatographie. La distribution est ici celle de mononucléotides portant un phosphate marqué en 5' mais ne diffère pas significativement de la distribution de mononucléotides portant le phosphate marqué en 3'.

2.9. Purification des protéines

Deux stratégies différentes ont été utilisées pour produire et purifier les protéines. D'une part, les protéines associées aux sRNA à boîtes H/ACA, L7Ae, aNOP10, aCBF5, aGAR1 ainsi que TruB et aPus10 et tous les variants de ces protéines ont été produits avec une étiquette GST permettant une purification sur résine glutation-sépharose. Les protéines sont éluées avec l'enzyme PreScission dans son tampon permettant le clivage du groupe GST directement sur la colonne de résine. Les protéines purifiées sont conservées à une concentration de 5 μM et congelées à -80°C en présence du glycérol 10%.

D'autre part, les protéines aPus7 de *P. abyssi* et *S. solfataricus* ainsi que leur variants ont été produits avec une étiquette His permettant une purification sur résine de Nickel. Ces protéines ont été éluées de la colonne par l'utilisation d'un tampon comportant 250 mM d'imidazole. Les protéines purifiées sont conservées à une concentration de 0,5 $\mu\text{g}.\mu\text{l}^{-1}$ en présence de glycérol 50%.

RESULTATS & DISCUSSION

1. Identification de 73 séquences pouvant coder un ou des ARNnc chez les Pyrococcus

A. Contexte

Le traitement bioinformatique des séquences génomiques a surtout été consacré jusqu'à très récemment à l'identification des ORF. En effet, comme nous l'avons expliqué dans l'introduction, il est beaucoup plus difficile d'identifier les régions des génomes servant de matrice pour la transcription des régions non codantes. De ce fait, la recherche des ARNnc autres que les ARNr et ARNt a longtemps reposé sur une identification biochimique non systématique. Du fait de l'accumulation de séquences génomiques et de la prise de conscience de l'importance des ARNnc encore non identifiés, des méthodes bioinformatiques ont été développées depuis quelques années, comme nous l'avons décrit dans l'introduction de ce manuscrit. Elles sont dédiées à un type d'ARNnc, à un domaine du vivant, voire même à une espèce donnée. Ainsi, au début de ce travail de thèse, seules deux études conduisant à l'identification des premiers gènes de sRNA à boîtes C/D avaient été réalisées dans le domaine archaea (Gaspin *et al.*, 2000 ; Omer *et al.*, 2000). La recherche systématique de gènes d'ARNnc chez les archaea était une nouvelle thématique au laboratoire mais l'équipe avait une longue expérience de l'utilisation de la phylogénie en matière d'étude des ARN. La phylogénie paraissait plus prometteuse en matière de recherche d'ARN non connus que la composition en nucléotides. En effet, la composition en nucléotide des séquences génomiques varie fortement suivant les espèces (Schattner, 2002). Par ailleurs, les signaux de transcription n'étaient pas non plus de bons critères de sélection puisque les promoteurs transcriptionnels d'archaea sont mal décrits car encore trop peu étudiés (pour revue, Soppa, 1999 ; Slupska *et al.*, 2001). L'utilisation de critères thermodynamiques peut aussi être mal adaptée à la recherche de certains ARNnc qui peuvent être peu structurés et, de plus, la distinction entre les ARNnc et les ARNm qui présentent également une structuration n'est pas facile et requiert l'utilisation d'algorithmes compliqués combinant souvent plusieurs critères de sélection (Rivas et Eddy, 2000). En partant de l'hypothèse qu'entre espèces proches sur le plan phylogénétique, les séquences des gènes d'ARNnc doivent être plus conservées que les séquences environnantes non transcrites, nous avons recherché des gènes potentiels d'ARNnc dans le genre *Pyrococcus* en comparant les séquences génomiques de 3 espèces : *P. abyssi*, *P. furiosus* et *P. horikoshii*. En effet, ces trois espèces présentent les distances phylogénétiques adéquates pour ce type d'approche et constituaient le seul groupe de trois génomes présentant de telles distances parmi les 11 génomes d'archaea initialement disponibles.

B. Démarche

D'après les données disponibles sur les gènes de sRNA C/D chez les archaea, ces gènes sont situés entre ou très légèrement chevauchant avec les ORF (Omer *et al.*, 2000 ; Gaspin *et al.*, 2000). Il

en est de même pour les ARNnc identifiés chez *E. coli* (Wassarman *et al.*, 2001). Nous avons donc décidé de focaliser l'étude comparative sur les segments reliant des ORF déjà répertoriées, ainsi que des gènes d'ARNr et d'ARNt caractérisés. Nous excluons ainsi les ARNnc *cis*-antisens. Néanmoins, nous pouvions ainsi avoir accès aux gènes d'autres types d'ARNnc. Nous avons donc, à partir de l'annotation des génomes, défini des régions dites Inter-Régions-Codantes (IRC) ne comportant ni ORF, ni gène d'ARNnc déjà décrit. Afin de prendre en compte de légers chevauchements possibles entre gènes d'ARNnc et ORF adjacentes, les segments IRC ont été définis de façon à contenir 15 nt de part et d'autre de la séquence séparant les ORF. Il est évident qu'une conservation de la séquence de ces IRC entre plusieurs espèces pouvait avoir plusieurs causes possibles : la présence d'un gène d'ARNnc, mais aussi d'éléments intervenant dans la transcription ou la traduction et leurs régulations et enfin la présence d'ORF non détectées lors des annotations initiales des génomes.

L'approche comporte donc deux étapes pour lesquelles nous avons écrit des scripts : l'extraction des séquences IRC et leur regroupement en banques, et la recherche de similarités de séquence entre les IRC issues des banques des trois espèces de *Pyrococcus*.

C. Collection de séquences IRC

Dans un premier temps, nous avons mis au point un « script » permettant l'extraction de toutes les séquences IRC d'une séquence génomique. Ce script basé sur l'utilitaire « extract » (voir partie Matériel et Méthodes) a permis l'établissement de collections (ou banques) de fragments IRC pour chacun des 11 génomes d'archaea dont la séquence du génome était disponible. Cette banque de données est constituée de séquences de fragments IRC double brin, identifiés par un nom indiquant le génome d'appartenance, ainsi que leur position au sein de la séquence de ce génome. Nous avons choisi de répertorier systématiquement les séquences provenant du brin 5'→3' ou brin « Forward » (FW). Toutefois, ce choix n'interdit pas l'analyse ultérieure du brin complémentaire ou brin « Reverse Complement » (RC).

Pour établir cette banque, le programme édite, à partir d'un fichier répertoriant toutes les positions de début et de fin des régions codantes, un autre fichier répertoriant les positions de début et de fin des régions IRC. Les séquences correspondant à ces IRC sont ensuite extraites à partir de ces positions et de la séquence entière du génome concerné. Les séquences IRC d'un organisme sont ensuite regroupées au sein d'une même banque. Au final, les segments correspondant aux régions codantes présentes sur l'un ou l'autre des deux brins sont absents des séquences IRC recueillies dans cette banque (Fig. R.1).

Comme mentionné ci-dessus, les segments IRC ont été définis de façon à contenir 15 nt de part et d'autre de la séquence séparant les régions connues pour être traduites (ORF) ou transcrites (ARNnc connus), ceci afin de prendre en compte des chevauchements éventuels entre gènes d'ARNnc et ORF adjacentes (Fig. R.1).

Nous avons exclu les s quences IRC potentielles dont la taille est inf rieure   15 nt. De telles s quences sont en effet peu susceptibles de coder un g ne d'ARNnc. De ce fait, la taille minimum des s quences IRC est donc de 45 nt.

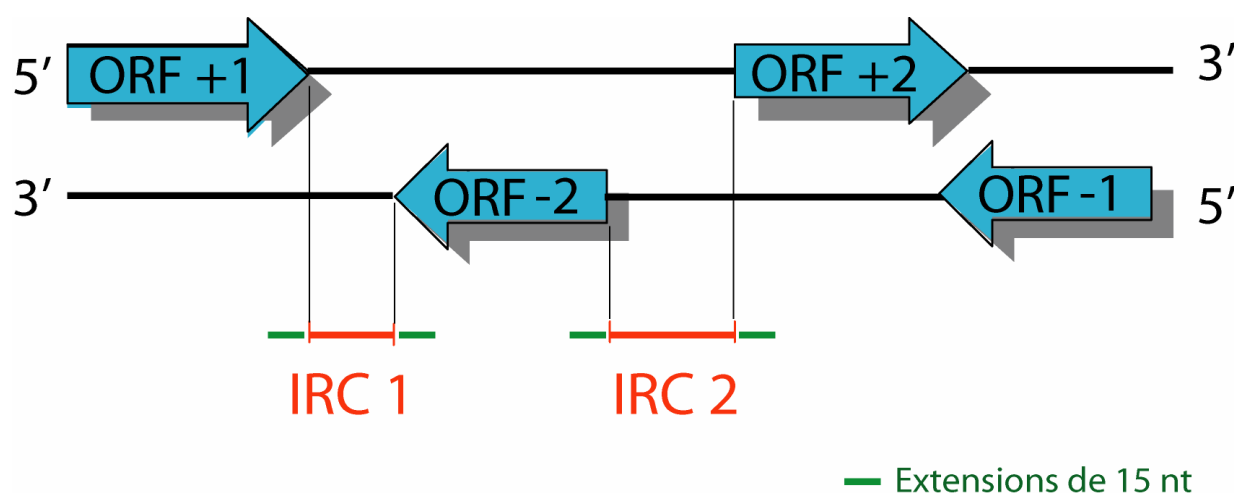


Figure R.1 : Extraction des fragments IRC   partir de la s quence g nomique et de ses annotations.

La banque de fragments IRC est compos e de fragments double brin ne comportant pas d'ORF d crite sur les deux brins d'ADN. 15 nt en amont et 15 nt en aval ont  t  ajout s   ces s quences IRC, permettant de rechercher des g nes d'ARNnc chevauchant l g rement les ORF.

Bien que d j  identifi s, les fragments correspondant aux sRNA   bo tes C/D provenant des trois esp ces de *Pyrococcus* (Gaspin *et al.*, 2000 ; Omer *et al.*, 2000) n'ont pas  t  consid r s comme r gions codantes. Notre objectif  tait de tester la validit  de l'approche en v rifiant que nous les retrouvions bien.

Au final, cette premi re  tape nous a conduit   extraire, respectivement, 912, 1054 et 929 s quences IRC dans les esp ces de *P. abyssi*, *P. furiosus* et *P. horikoshii* (Fig. R.2).

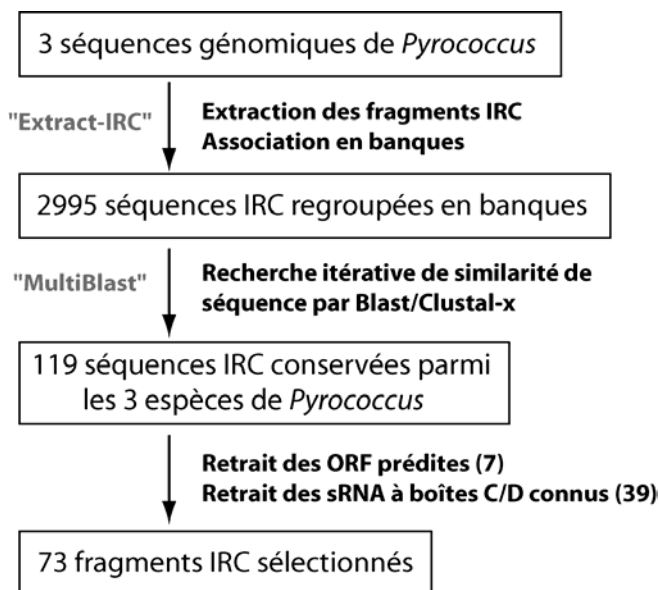


Figure R.2 : Représentation schématique des principales étapes menant à la sélection des fragments IRC.

D. Sélection des fragments IRC présentant des éléments de séquence conservés entre les trois espèces de *Pyrococcus*.

Pour rechercher des conservations de séquences entre IRC provenant des 3 espèces de *Pyrococcus*, nous avons mis au point une démarche bioinformatique basée sur le programme Blast, (voir Matériel et Méthodes). Ceci nous a permis de comparer un par un chaque fragment IRC d'une des trois souches de *Pyrococcus* à tous les fragments IRC de cette souche et des deux autres souches (Fig. R.3) mais aussi aux fragments IRC de tous les génomes d'archaea disponibles. Le script qui nous a permis d'automatiser cette comparaison de séquence a été appelé « multiblast » (Fig. R.2).

Le logiciel d'alignement local Blast recherche des similarités de séquences entre deux molécules, indépendamment du positionnement de ces séquences au sein des deux molécules comparées. Il nous a fallu définir, à partir de combien de résidus identiques successifs nous considérions que la similarité détectée était significative. Sur la base de la taille des génomes, 2 Mpb en moyenne, sans tenir compte du biais de l'usage de chacun des quatre nucléotides, une séquence de 11 pb est statistiquement présente une seule fois dans le génome. Cependant, il s'agit d'une valeur théorique. Sur la base des comparaisons des séquences réalisées pour les *Pyrococcus* (Maeder *et al.*, 1999 ; Lecompte *et al.*, 2001), il ressortait qu'une succession de 18 résidus identiques, ou 21 résidus identiques avec un mésappariement, une délétion ou une insertion entre séquences issues des 3 génomes traduisait une similarité de séquence significative. Nous avons testé ces dernières valeurs et finalement nous les avons choisies comme critères de sélection de cette étape. Dans le logiciel Blast, ceci se traduit par l'obtention d'un score (e-value) inférieur à 0,01 (Altschul *et al.*, 1997).

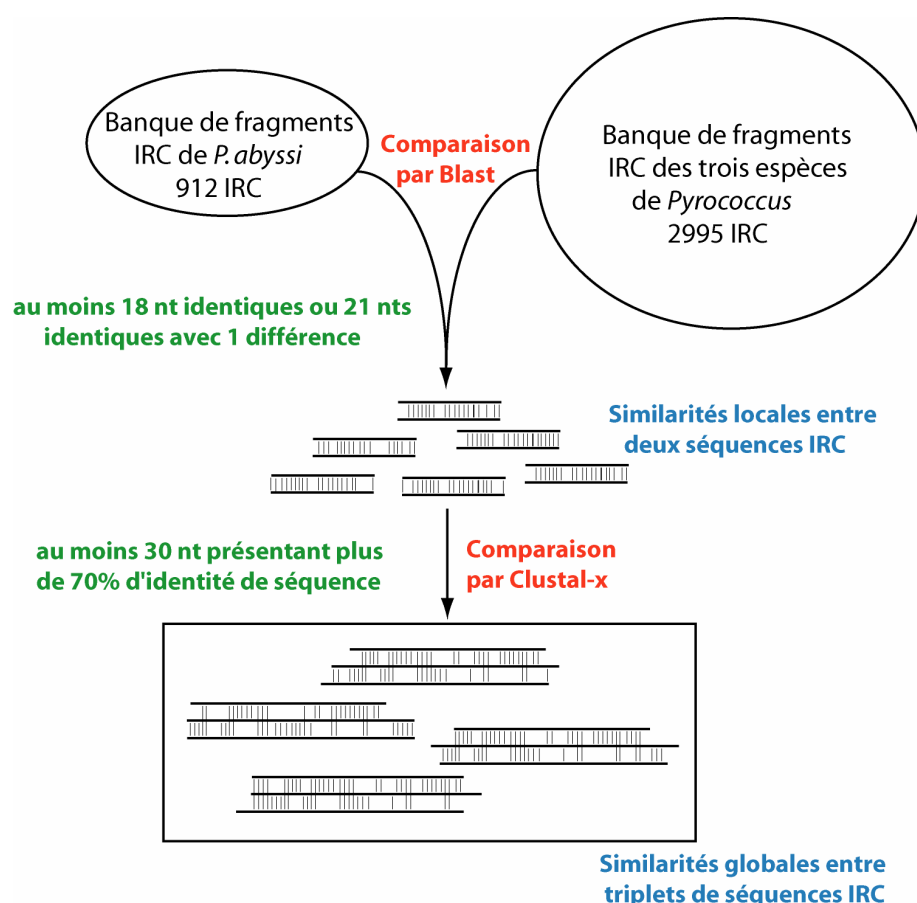


Figure R.3 : Repr sentation sch matique de la recherche de similarit  de s quence   partir des banques de fragments IRC. Les programmes utilis s sont indiqu s en rouge, les crit res de s lection en vert, et les r sultats des  tapes en bleu.

Par cons quent, nous avons s lectionn  les fragments IRC dont la s quence pr sentait, dans chacune des deux autres esp ces de *Pyrococcus*, une identit  de s quence sur au moins 18 nt successifs ou sur 21 nt successifs   l'exception de l'addition, la d l tion ou la substitution d'un r sidu. Certains de ces fragments IRC pr sentaient des similarit s avec de nombreuses s quences IRC issues du m me g nome ou d'un autre g nome de *Pyrococcus* et pouvaient correspondre de ce fait   des s quences r p t es. Nous avons r pertori  ces fragments, car ils pourraient faire l'objet d'une  tude ult rieure sur les  l ments r p t s dans les g nomes de *Pyrococcus*. N anmoins, nous n'avons pas pouss  plus loin l' tude sur ces fragments,  tant donn  que notre objectif  tait de rechercher des g nes d'ARNnc, g nes qui dans la plupart des cas, chez les procaryotes, sont d tect s   l' tat de copie unique. L'utilitaire Seqmatchall (voir Mat riel et M thodes) nous a permis de retrouver les r p titions parfaites de taille sup rieure   50 nt pr sentes dans les esp ces de *Pyrococcus* ou d'autres archaea. Les r p titions d tect es par ces deux m thodes ont  t   cart es de la poursuite de l'analyse.

Par ailleurs, il  tait important  galement de mesurer la complexit  de la s quence de l'IRC. Pour cela, nous avons utilis  la comparaison des s quences IRC de *Pyrococcus* avec celles des autres esp ces d'archaea. Nous avons estim  qu'une s quence pr sente une seule fois dans une des esp ces de *Pyrococcus*, mais qui pr sente des similarit s de s quence avec de nombreuses s quences IRC d'un g nome d'archaea donn , est certainement une s quence de faible complexit  et doit donc  tre  cart e.

Ainsi, nous avons sélectionné, dans un premier temps, les séquences IRC qui présentent des similarités avec une séquence IRC de chacune des deux autres espèces de *Pyrococcus*, sur la base des critères de sélection évoqués ci-dessus, et qui ne correspondaient ni à des répétitions, ni à des séquences de faible complexité.

Il paraissait important de compléter cette analyse en déterminant si la zone de similarité de séquence identifiée concerne une plus grande partie de l'IRC. Dans ce but, nous avons employé le logiciel Clustal-X qui permet de réaliser des alignements globaux. Les séquences regroupées par l'observation de similarités lors de la première étape d'alignement local ont été utilisées en entrée de Clustal-X. Afin d'éviter de sélectionner des éléments de régulation de l'expression des ORF adjacentes, nous avons choisi de ne retenir que les fragments ICR qui présentent, dans chacune des trois espèces de *Pyrococcus*, au moins un segment de taille supérieure à 30 nt dont l'identité de séquence entre les trois IRC dépasse 70% (Fig. R.3).

Par cette démarche, nous avons identifié 119 fragments IRC présentant un degré de conservation significatif chez les trois espèces de *Pyrococcus* (Fig. R.2).

Il était envisageable que des ORF non décrites lors de l'annotation initiale de ces génomes puissent expliquer la conservation de séquence relevée pour certains segments IRC. Pour tester cette possibilité, nous avons utilisé le logiciel GeneMarkS (Besemer *et al.*, 2001). Ce logiciel est le fruit d'une amélioration du logiciel GeneMark, créé pour prédire les positions des ORF dans les nouveaux génomes procaryotes pour lesquels très peu de données sont disponibles. GeneMarkS est basé sur l'analyse de la distribution des nucléotides au sein des séquences comprises entre un codon potentiel d'initiation et un codon potentiel de terminaison en phase. Sa particularité est l'intégration par l'utilisation de chaînes de Markov cachées, en plus d'un biais dans l'usage des codons, de la présence de motifs spécifiques au voisinage du codon d'initiation, tels que le site RBS. Il nous a paru intéressant d'explorer les séquences des trois génomes de *Pyrococcus* par ce logiciel, et de comparer les positions des ORF prédites avec les positions des fragments IRC. Dans un certain nombre de cas, nos données suggèrent que les annotations de *Pyrococcus horikoshii*, incluses dans GenBank, méritent d'être révisées. Pour *P. abyssi* et *P. furiosus*, GeneMarkS prédit plus d'ORF que le nombre actuellement répertorié dans GenBank à l'inverse de la situation chez *P. horikoshii* (1899 ORF prédites contre 1816 ORF référencées pour *P. abyssi* ; 2120 contre 2071 pour *P. furiosus* ; 1886 contre 2069 pour *P. horikoshii*). D'une manière générale, pour les trois génomes, GeneMarkS prédit souvent un début d'ORF différent de ce qui est donné dans GenBank. L'accord est meilleur en ce qui concerne la fin des ORF. Enfin, de très nombreux désaccords apparaissent entre les ORF de *P. horikoshii*, prédites par GeneMarkS et celles indiquées dans GenBank. De manière intéressante, les prédictions de GeneMarkS pour *P. horikoshii* conduisent à des propositions plus proches de celles qui ont été publiées pour *P. abyssi* et *P. furiosus*, que celles publiées pour *P. horikoshii* (Kawarabayasi *et al.*, 1998). Ceci est en accord avec une critique générale des spécialistes d'archaea vis-à-vis de l'annotation qui a été faite du génome de *P. horikoshii*. Un exemple est donné dans la figure R.4: les annotations de GenBank indiquent que le gène *rpoN* codant la sous-unité N de l'ARN polymérase est situé entre les positions 1446906 et 1446547 sur le brin négatif de *P. horikoshii*, ce qui correspond à une protéine de 119

acides aminés alors que les protéines prédites chez les autres archaea comportent entre 65 et 75 acides aminés. De plus, selon les données de GenBank, l'ORF de *rpoN* chevaucherait *rps9P* codant la protéine S9 de la petite sous unité ribosomale et ceci sur un même brin d'ADN, ce qui n'est pas le cas chez *P. abyssi* et *P. furiosus*. Ce chevauchement n'existe plus avec la prédiction d'ORF faite par GeneMarkS. Aussi nous proposons que chez *P. horikoshii*, le codon d'initiation de *rpoN* est localisé à la position 1446758 (soit 148 nts en aval) (Fig. R.4).

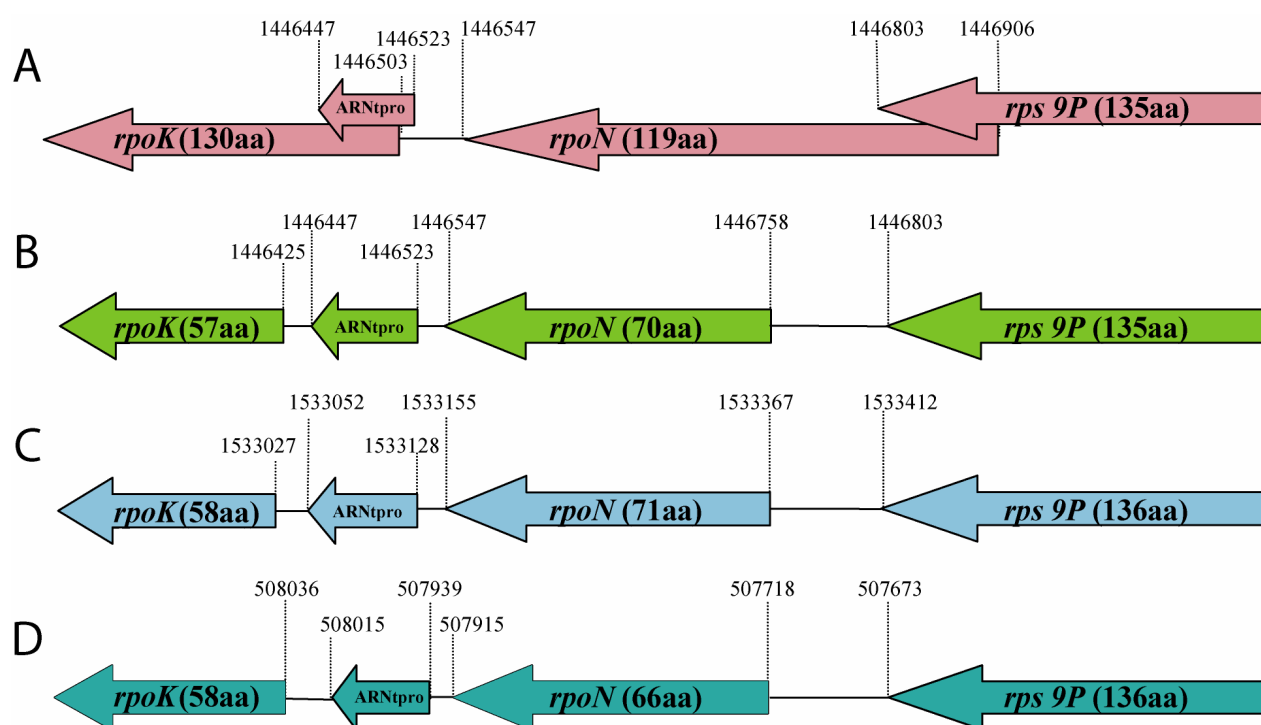


Figure R.4 : Exemple de désaccord entre les annotations de *P. horikoshii* et celles des deux autres *Pyrococcus*.

L'organisation précisée dans GenBank de cette région de *P. horikoshii* (A) montre d'importantes différences avec les organisations des gènes proposées pour, *P. furiosus* (C) et *P. abyssi* (D). L'utilisation du logiciel GenMarkS nous a permis de proposer pour *P. horikoshii* une organisation similaire (B) à celle des deux autres espèces (C et D).

Finalement, nous avons complété l'analyse faite par GenMarkS par un examen basé sur d'autres critères : la taille, la comparaison de la nature des codons d'initiation et de terminaison, la distribution des codons utilisés par rapport à la distribution observée dans l'ensemble des ORF, le degré de conservation entre les trois *Pyrococcus*, l'organisation de la région et la possibilité d'appartenance à un opéron (synthénie) et enfin, les homologues possibles avec des ORF du même génome ou d'autres génomes. De ce fait, 7 des fragments IRC sélectionnés, présentent des ORF qui, à elles seules

expliquent le fort degré de conservation entre les trois espèces de *Pyrococcus*. Ainsi, 112 séquences IRC conservées ont été retenues.

L'annotation du génome est donc une limitation importante. S'il est possible d'exclure des fragments IRC qui comportent d'éventuelles ORF, par contre il est évident que de mauvaises définitions des positions de début et de fin des ORF ont initialement contribué à annoter à tort des segments d'IRC en tant que ORF et donc à diminuer la taille ainsi que le nombre des séquences IRC. L'amélioration du nombre de faux négatifs nécessite donc une vérification initiale des annotations des génomes.

E. Evaluation de l'approche

Pour évaluer l'efficacité de cette approche nous avons recherché dans les 112 IRC conservées entre les trois espèces de *Pyrococcus* la présence de sRNA à boîtes C/D que nous avons volontairement choisi de ne pas considérer comme ARNnc connus. Les sRNA à boîtes C/D sont en effet, de par leur taille, leur positionnement dans le génome et leur degré de conservation, de bons représentants des ARNnc. Le génome de *P. abyssi* compte 59 gènes de sRNA à boîtes C/D (Gaspin *et al.*, 2000 ; Omer *et al.*, 2000) 9 n'étant pas conservés dans les deux autres espèces et 1 (sR40) étant codé dans l'intron de l'ARNt^{Tp} (Clouet d'Orval *et al.*, 2001). Parmi les 49 sRNA à boîtes C/D supposés présents dans les séquences IRC conservées, 45 ont effectivement été détectés, ceci au sein de 41 des fragments IRC. En effet, certains segments IRC comportent deux sRNA à boîtes C/D. Par conséquent, seuls 4 sRNA n'ont pas été détectés, ceci en raison de la localisation d'un d'entre eux (sR7) dans une ORF, ou en raison d'un trop faible degré de conservation entre les trois génomes de *Pyrococcus* (sR25, sR28 et sR59). Dans les autres cas, les IRC renfermant un gène de sRNA à boîtes C/D comportent un segment dont la séquence est très conservée entre les trois espèces. Les séquences correspondant aux sRNA sR4 dans les 3 espèces présentent par exemple 86% d'identité (Fig. R.5). La séquence de l'IRC qui encadre ce sRNA sR4 est nettement moins bien conservée puisqu'elle présente un peu moins de 50% d'identité entre les trois espèces. Par conséquent, il est possible de distinguer facilement les limites des séquences correspondantes à tous ces sRNA à boîtes C/D au sein des fragments IRC conservés. Cette observation révèle qu'en plus de permettre une identification efficace des sRNA à boîtes C/D, cette approche permet aussi de délimiter leurs extrémités. En effet, cet alignement montre que la conservation du sRNA sR4 s'étend jusqu'à l'hélice terminale qui rapproche en général les extrémités 5' (boîte C) et 3' (boîte D) des sRNA à boîtes C/D.

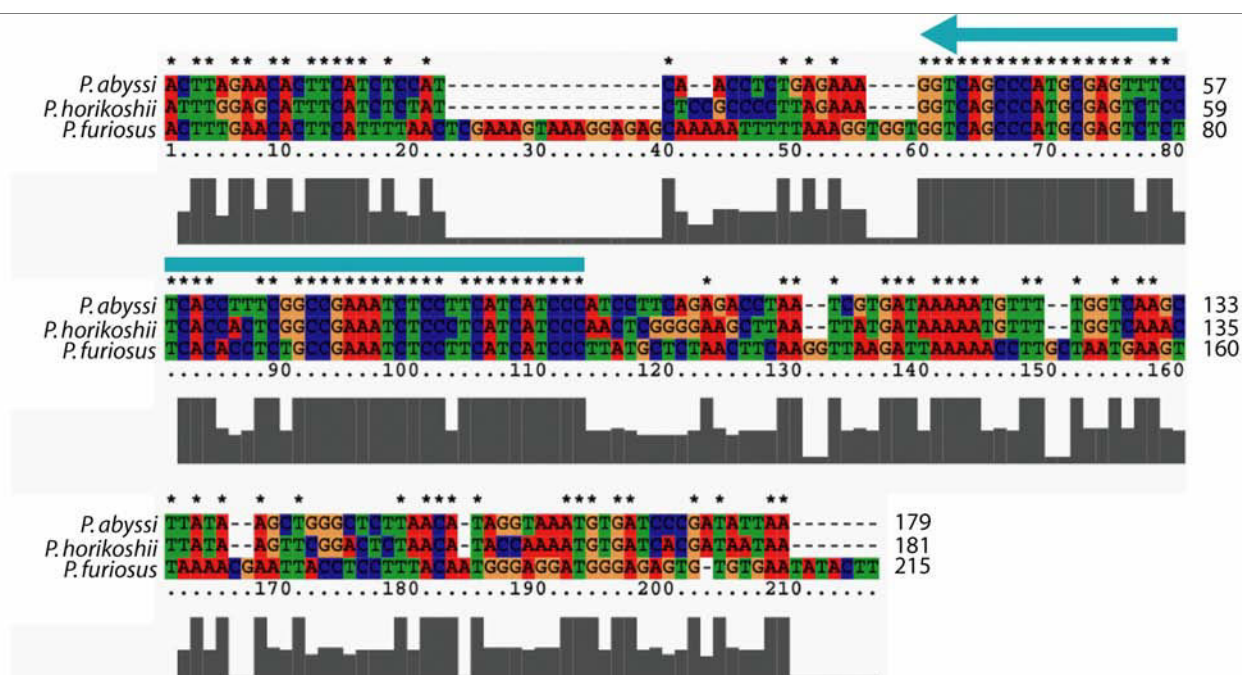


Figure R.5 : Alignement par Clustal-X des fragments IRC comportant le sRNA à boîtes C/D sR4.

La flèche représente le sRNA sR4 conservé dans les trois souches de *Pyrococcus* étudiées. Les nucléotides conservés dans les trois espèces sont indiqués par des étoiles.

Notre approche a donc permis de détecter plus de 90% des sRNA à boîtes C/D identifiés chez *P. abyssi*. Il est à noter que lorsque nous avons réalisé l'analyse, seulement 52 sRNA avaient été identifiés et que notre analyse avait permis de mettre en évidence certains sRNA avant qu'ils aient été caractérisés par clonage (Dennis *et al.*, 2001). Ces observations validaient notre approche.

Les 39 fragments IRC dont la conservation de séquence n'était expliquée que par la présence de sRNA à boîtes C/D ont été retirés. Ainsi, au final, nous avons donc sélectionné 73 IRC présentant des conservations de séquence dans les trois espèces de *Pyrococcus* (Fig. R.2). Les positions et les caractéristiques de ces 73 fragments IRC conservés sont données dans le tableau R.1.

Numéro du Fragment IRC	Position Début	Position Fin	Taille du fragment IRC (nt)	Taille du segment d'intérêt (nt)	Taux d'identité du segment d'intérêt	Taux d'identité du reste du fragment IRC	Commentaires
N1	16302	16484	183	70	76%	46%	
N3	58651	59441	178	102	75%	22%	
N9	86008	86155	131	78	74%	39%	
N11	122685	123515	825	82	63%	39%	PfQ1 (a)
				80	81%		
				56	79%		
N13	129263	129451	210	113	71%	28%	
N14	164337	164489	153	153	73%	-	
N15	191281	191372	92	27	81%	29%	
N16	197232	197551	290	83	65%	23%	
				59	88%		
N17	199087	199233	147	69	71%	37%	
N18	201371	201479	86	43	65%	16%	
N19	204695	205054	360	135	89%	25%	
N21	230312	230668	316	59	95%	23%	Pf1(a), présence sR2 et sR9
N25	256848	256934	74	49	63%	12%	
N28	266045	266203	140	51	75%	29%	
N29	279643	279734	85	85	60%	-	
N30	281039	281150	112	67	84%	18%	
N31	296588	296692	105	32	84%	45%	
N35	318095	318449	303	157	94%	42%	Pf6 (a), présence sR29
N36	337681	337851	171	86	79%	36%	
N37	340459	340784	245	84	79%	34%	
N39	371639	371799	161	62	71%	47%	
N40	382313	382638	271	163	91%	45%	Pf7 (a)
N41	383506	383624	119	82	79%	30%	
N42	385981	386111	98	32	72%	29%	
N44	403337	403477	121	61	75%	22%	
N50	522817	522911	95	40	80%	44%	
N51	541763	542080	307	103	83%	33%	
N52	543312	543450	122	59	78%	54%	
N55	559464	559615	151	27	74%	16%	
				18	100%		
				40	90%		
N56	579142	579287	145	38	92%	46%	
N57	605555	605722	168	77	78%	33%	
N58	628756	628839	84	45	91%	38%	

Tableau R.1 : Positions, tailles et taux d'identité de séquence des IRC sélectionnées et de leurs segments conservés. Les positions données sont celles dans le génome de *P. abyssi*. Une identité traduit le maintien d'un nucléotide dans les trois séquences IRC issues des trois espèces de *Pyrococcus*. L'indice (a) précise des candidats identifiés par les travaux de Klein *et al.* (2002).

Num�ro du Fragment IRC	Position D�but	Position Fin	Taille du fragment IRC (nt)	Taille du segment d'int�r�t (nt)	Taux d'identit� du segment d'int�r�t	Taux d'identit� du reste du fragment IRC	Commentaires
N61	636623	636935	170	70	63%	19%	PfQ11, pr�sence sR44
N62	638191	638258	68	31	74%	46%	
N63	639620	639791	131	99	76%	38%	
N65	644738	644842	105	64	88%	63%	
N66	683475	683686	180	159	73%	36%	
N67	690007	690250	244	165	83%	48%	Pf5 (a)
N69	711628	711831	194	37	84%	33%	
				68	87%		
N71	827652	827786	135	43	98%	29%	
N72	849537	849752	114	46	67%	26%	
N73	883964	884245	121	32	84%	28%	
				60	75%		
N76	938093	938175	83	40	80%	47%	
N80	966985	967096	112	51	73%	43%	
N83	1000575	1000780	159	83	72%	21%	
N85	1028990	1029110	121	50	80%	35%	
N86	1029877	1029959	82	82	55%	-	
N88	1056633	1056994	116	116	63%	-	
N91	1074484	1074608	68	32	84%	36%	Pf4 (a)
N92	1079738	1080246	55	44	93%	18%	
N93	1083263	1083399	70	70	69%	-	
N97	1198143	1198265	104	85	81%	38%	
N99	1219982	1220573	570	47	83%	27%	PfQ6 (a)
N100	1232286	1232335	50	47	72%	-	
N104	1330609	1330747	127	58	83%	29%	
N105	1335637	1335882	246	199	85%	45%	Pf3 (a)
N106	1386088	1386607	136	136	72%	-	
N107	1389104	1389257	154	60	75%	35%	
N109	1398113	1398243	129	56	68%	34%	
N112	1412516	1413224	39	39	87%	-	
N114	1458060	1458206	147	63	75%	39%	
N115	1468780	1468885	95	61	84%	18%	
N116	1478359	1478541	157	57	74%	25%	
N117	1483990	1484099	110	70	85%	17%	
N120	1606713	1606886	185	68	69%	25%	
N121	1640335	1640419	85	60	75%	28%	
N122	1642847	1642990	137	137	68%	-	
N125	1691472	1691626	106	76	67%	40%	
N126	1705458	1705637	112	72	72%	25%	
N127	1715251	1715366	112	66	68%	28%	
N128	1754402	1754788	345	68	78%	31%	
N130	1763426	1763498	65	65	73%	-	
N160	289007	289493	183	59	70%	34%	Pf9 (a)

F. Comparaison de la méthode développée avec d'autres approches appliquées au génome de *P. abyssi*

Notre approche a abouti à la mise en évidence de 73 fragments IRC conservés entre les trois espèces de *Pyrococcus*. Nous ne pouvons pas exclure qu'en plus de la présence de gènes d'ARNnc cette conservation puisse être liée à des contraintes au niveau de l'architecture du génome et à la fixation de protéines intervenant dans sa structuration, la conservation de séquence de promoteurs transcriptionnels et de leurs éléments de régulation, la conservation d'éléments dans les segments 5' ou 3' UTR des ARNm intervenant dans l'efficacité de la traduction, ou la stabilité des ARN et l'existence de gènes de petites protéines pas même détectées avec par emploi du logiciel GeneMarkS. Cependant, étant donné la taille des segments conservés au sein de ces fragments IRC ainsi que leur degré de conservation, nous pensons que de nombreux segments IRC sélectionnés contiennent des gènes d'ARNnc. L'examen des 73 IRC montre que leur taille pourrait atteindre jusqu'à 200 nt.

Un premier point à souligner est l'efficacité de notre approche par rapport à une autre approche de génomique comparative appliquée à ces trois mêmes espèces de *Pyrococcus*. Celle-ci (appelée Tricross) est basée sur le programme Fasta et a abouti à un nombre quasi-similaire de fragments intergéniques conservés (72) (Ray *et al.*, 2001). Parmi les 72 candidats proposés par l'approche Tricross, seuls 5 candidats sont présents dans les IRC que nous avons sélectionnées. Les candidats proposés par cette autre approche sont donc majoritairement différents des candidats que nous avons pu proposer. Il faut noter également que l'approche « tricross » n'a permis la description que de 29 des gènes de sRNA à boîtes C/D contre 45 par notre approche. La différence est notamment expliquée par le fait que nous avons couplé deux étapes reposant sur un alignement local puis un alignement global des séquences des IRC mais aussi par le fait que nous avons utilisé des séquences IRC prenant en compte 15 nt provenant des régions codantes adjacentes. En effet, nous avons remarqué que si nous n'avions pas inclus les 15 nt appartenant aux régions adjacentes, nous n'aurions pas été en mesure de détecter 8 des 45 sRNA à boîtes C/D dans les IRC conservées. Il serait peut-être intéressant d'utiliser des segments chevauchant plus longuement les ORF adjacentes (20 ou 25 nt) pour améliorer la procédure bioinformatique établie. En effet, 25 des 59 sRNA C/D de *P. abyssi* chevauchent les ORF adjacentes, et parmi ces 25 sRNA C/D 15 chevauchent sur plus de 15 nt les ORF adjacentes. Par ailleurs, les gènes de sRNA C/D sont synthétisés sur le même brin ou sur le brin complémentaire aux ORF adjacentes. On ne peut pas exclure que, dans ce dernier cas, en plus de leur rôle de guide de la méthylation des ARNr ou ARNt, ils aient un rôle régulateur de la traduction ou de la stabilité des ARNm (Zago *et al.*, 2005). Cette hypothèse mériterait d'être testée expérimentalement. Une autre hypothèse est que le chevauchement des gènes de sRNA à boîtes C/D, et par extension d'autres ARNnc, ne soit qu'une conséquence de la densité de gènes au sein des génomes d'archaea.

Une autre approche visant à rechercher des ARNnc dans le génome de *P. furiosus* s'est appuyée sur la composition en G/C. L'équipe de S. Eddy a ainsi proposé l'existence de 8 ARNnc candidats appelés Pf1 à Pf9 (Klein *et al.*, 2002). Parmi ces 8 candidats, 7 sont des équivalents de nos candidats (Tableau R.1). Cette étude a été complétée par la recherche de structures secondaires conservées par le biais de l'utilisation de QRNA (Rivas et Eddy, 2001). Ainsi, 13 nouveaux candidats (appelés PfQ1 à

PfQ13) ont été proposés (Klein *et al.*, 2002). Parmi ces 13 candidats, seuls 3 sont des équivalents de nos candidats (Tableau R.1). Enfin, cette dernière approche n'a permis la détection que de 6 gènes de sRNA C/D parmi les 55 présents chez *P. furiosus*. Par ailleurs, il est à noter que, dans notre étude, le pourcentage en G/C des fragments IRC que nous avons sélectionnés ne permet pas de les distinguer facilement de l'ensemble des fragments IRC. En effet, pour les trois espèces de *Pyrococcus*, la différence de pourcentage G/C entre l'ensemble des fragments IRC et les fragments IRC retenus ne dépasse pas 4 points (en faveur de ces derniers fragments).

G. Amélioration de la méthode développée

Un autre point intéressant est la possibilité de distinguer le gène potentiel de l'ARNnc du reste du fragment IRC sur la base du taux de conservation de séquence de son gène par rapport au reste de la séquence de l'IRC. En effet, on constate que le degré de conservation de ses gènes (Tableau R.1) dépasse souvent 70% alors que le degré de conservation de séquence du reste du fragment IRC est souvent inférieur à 50%. Par exemple, l'IRC N35 de 303 nt, renferme deux éléments : le gène du sRNA à boîtes C/D sR29 d'une taille de 58 nt, ainsi qu'un segment de 157 nt qui constitue un gène potentiel d'ARNnc. Le pourcentage d'identité de séquence entre les trois espèces de *Pyrococcus* est de 84% pour le gène du sRNA C/D sR29 et de 94% pour le segment de 157 nt. Le degré de conservation du reste de la séquence (88 nt) n'est que de 42 %. Ainsi, les gènes potentiels d'ARNnc candidats comme ceux de sRNA C/D peuvent être distingués au sein de leur contexte par le degré de conservation de séquence (Fig. R.6).

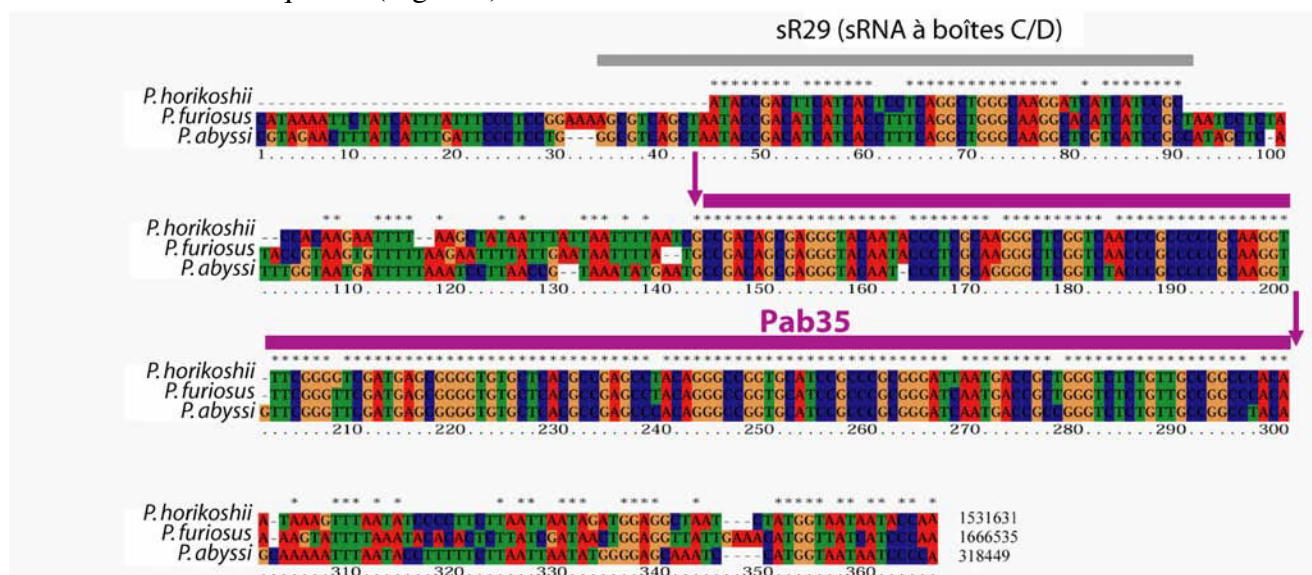


Figure R.6 : Alignement par Clustal-X des fragments IRC N35 comportant le candidat sRNA à boîtes H/ACA Pab35. La flèche représente le sRNA Pab35 conservé dans les trois souches de *Pyrococcus* étudiées. Les nucléotides conservés dans les trois espèces sont indiqués par des étoiles.

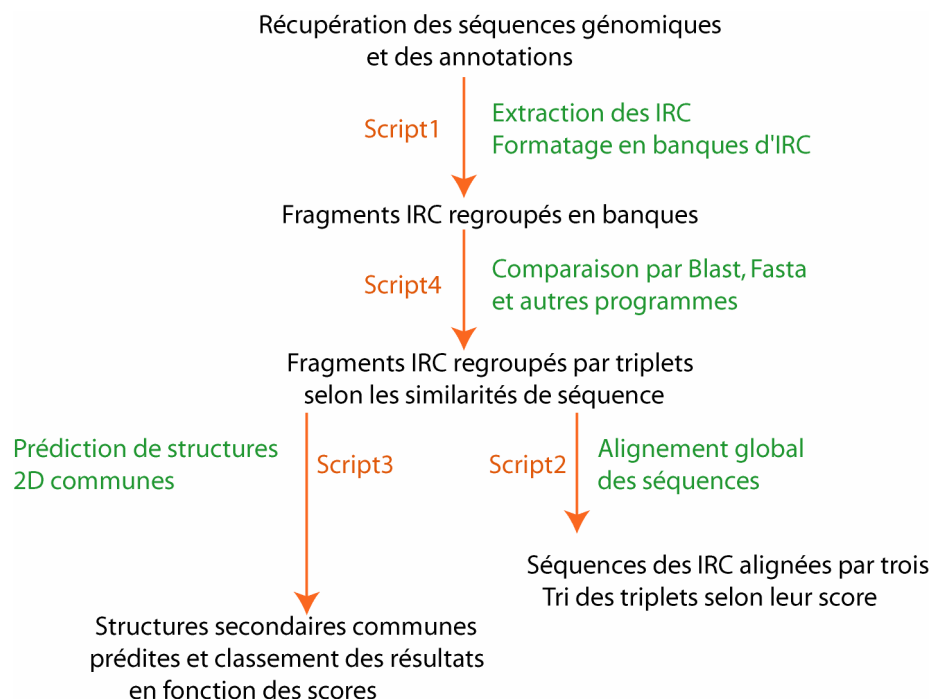
Cet aspect est particulièrement important en vue de l'application à d'autres espèces d'archaea. D'autres groupes de génomes pourraient être analysés par cette approche. En effet, l'ordre des

Sulfolobales, par exemple, présente au moins trois espèces dont les séquences génomiques sont disponibles et pour lesquelles les distances phylogénétiques permettraient l'application de cette approche. Cependant, pour appliquer cette approche à d'autres séquences génomiques et éventuellement réaliser quelques ajustements, il est important de développer des outils permettant son automatisation. En collaboration avec C. Bicep, stagiaire encadré par F. Leclerc, nous avons débuté l'écriture de plusieurs scripts. Deux scripts ont été préparés (Fig. R.7) : le premier automatise l'extraction de fragments IRC et leur regroupement en banques (script 1), le second automatise la recherche de similarités de séquence par des programmes d'alignement local de type Blast ou Fasta (script 4). Deux autres scripts sont en cours de préparation : l'un pourra aligner globalement les séquences des fragments IRC qui présentent des similarités de séquence et les trier selon la qualité de cet alignement (script 2), l'autre permettra la recherche de structurations secondaires à partir des séquences des fragments IRC qui présentent des similarités de séquence, ceci notamment en recherchant des paires de base compensatoires dans l'alignement de séquences qui lui est fourni (script 3). Ce dernier aboutira à un classement des résultats selon des critères thermodynamiques. Ce dernier script nous apportera des informations concernant la présence de motifs spécifiques portés par les ARNnc potentiels et donc des informations quant au rôle possible de ces ARNnc.

Par ailleurs, il est envisageable de comparer plus que trois génomes en ajoutant par exemple la séquence génomique de *Thermococcus kodakarensis* à celles des trois *Pyrococcus* et également d'ajuster le seuil de similarité de séquence (fixé à 18 identités) en fonction de la distance qui sépare les trois génomes regroupés pour cette analyse.

Figure R.7 : Représentation schématique de l'automatisation de la recherche de fragments IRC conservés entre espèces d'archaea.

Quatre scripts différents devraient permettre, à partir de fichiers comportant les séquences et les annotations de différents génomes, d'aboutir à deux fichiers de résultats contenant soit les séquences des IRC alignées par trois et triées selon un score traduisant la similarité de séquence, soit les structures secondaires communes prédites pour un triplet de séquences IRC et triées en fonction de scores thermodynamiques.



Au final, ces résultats ont été peu exploités, puisque seule la recherche de sRNA à boîtes H/ACA, après la découverte des premiers représentants archaea chez *A. fulgidus*, décrite ci-après,

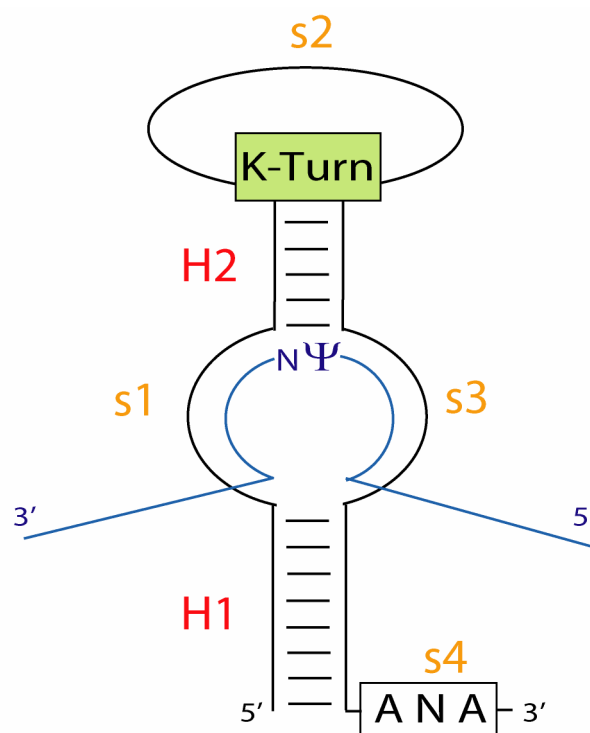
nous a permis d'expliquer la conservation entre les trois espèces de 7 de ces 73 ARNnc sélectionnés. La question de la présence de gènes d'ARNnc au sein des 66 ARNnc restants est donc toujours ouverte.

2. Identification de 5 sRNA H/ACA parmi les 73 ARNnc candidats

A. Contexte et Démarche

Comme énoncé dans la partie introduction, les sRNA à boîtes H/ACA d'archaea ont une architecture qui leur est spécifique. En effet, ils sont composés de 1, 2 ou 3 motifs en tige-boucles et chaque motif semble disposer de toutes les caractéristiques structurales essentielles à son mode d'action. Nous avons donc recherché, dans les 73 fragments IRC sélectionnés, la présence d'un ou plusieurs motifs H/ACA. Pour cela un descripteur du motif a été créé dans le programme RNAMOT (Gautheret *et al.*, 1993), il impose la présence de deux hélices, H1 de 6 à 10 pb et H2 de 4 à 10 pb, qui encadrent une poche interne composée de deux séquences simple-brin appelées s1 et s3 (Fig. R.8). Les tailles des séquences s1 et s3 ont été définies comme ayant entre 4 et 11 nt. Ce motif H/ACA est fermé en son sommet par une boucle considérée comme simple-brin (s2) et dont la taille a été fixée comme étant comprise entre 8 et 50 nt. Enfin, l'extrémité 3' est définie par un triplet ANA (s4). Le motif en K-turn présent au sommet de l'hélice apicale H2 n'a pas été utilisé lors de cette recherche, car il a été caractérisé ultérieurement (Rozhdestvensky *et al.*, 2003).

Figure R.8 : Description schématique d'un sRNA à boîtes H/ACA recherché avec le programme RNAMOT dans les séquences IRC sélectionnées. Ce motif caractéristique se définit par la présence de deux hélices H1 et H2 et de 4 parties simple-brin (s1 à s4) dont le segment s4 contenant une boîte ACA. L'existence de complémentarités de séquence entre les éléments antisens 5' et 3' de la boucle interne du sRNA avec les séquences des ARNr de l'espèce concernée a également été recherchée.



Pour chacun des sRNA H/ACA potentiels obtenus par cette première approche, nous avons recherché des cibles potentielles dans les ARNr et autres ARNnc des espèces de *Pyrococcus* étudiées. Pour cela, nous avons utilisé le programme RNAMOT (Gautheret *et al.*, 1993) permettant de décrire la séquence recherchée dans les ARN cibles en prenant en considération les appariements Watson-Crick et ceux de type Wobble (G.U). La séquence qui est recherchée est complémentaire aux éléments antisens 5' et 3' et comprend un doublet 5'-UN-3' comme décrit dans la figure R.9.

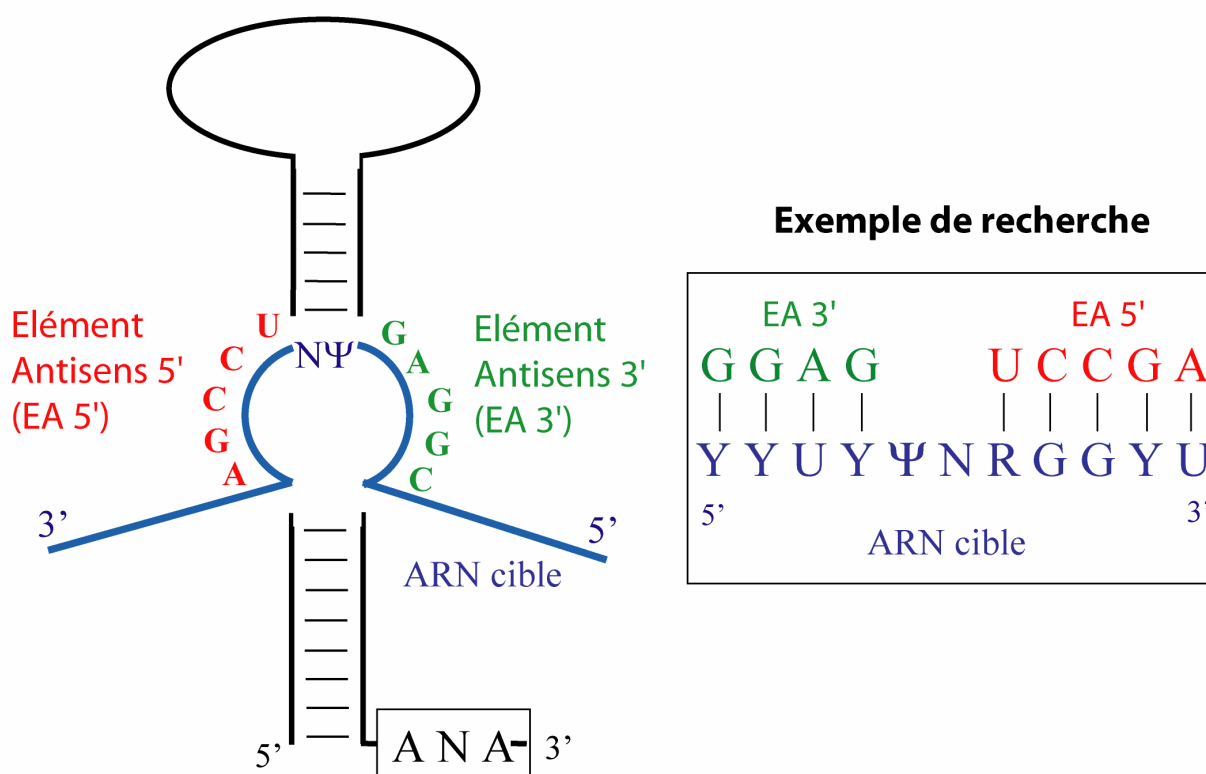


Figure R.9 : Représentation schématique de la méthode permettant de rechercher les cibles des sRNA à boîtes H/ACA candidats.

Les séquences des éléments antisens 5' et 3' sont représentées en rouge et en vert. La séquence de l'ARN cible recherchée est indiquée en bleu. La complémentarité tient compte des appariements de type Wobble G.U. Y : pyrimidine ; R : purine.

B. Résultats

L'utilisation du programme RNAMOT nous a permis d'identifier 9 séquences codant des segments d'ARN correspondant au motif défini. Ces motifs sont regroupés en 5 sRNA à boîtes H/ACA. Ceux-ci ont été identifiés dans les IRC N21, N35, N40, N91 et N105 de *P. abyssi*, *P. horikoshii* et *P. furiosus*. Le nom que nous avons attribué aux sRNA potentiels correspond aux numéros des fragments IRC sélectionnés : Pab21, Pab35, Pab40, Pab91 et Pab105 (les candidats de *P. abyssi* sont utilisés comme références). Le sRNA candidat Pab40 comporte trois motifs en tige-boucles, les deux candidats Pab35 et Pab105 en comportent deux, alors que les deux candidats

Pab21 et Pab91 n'en comportent qu'un seul. Les structures secondaires de ces 5 candidats sont présentées en figure R.10.

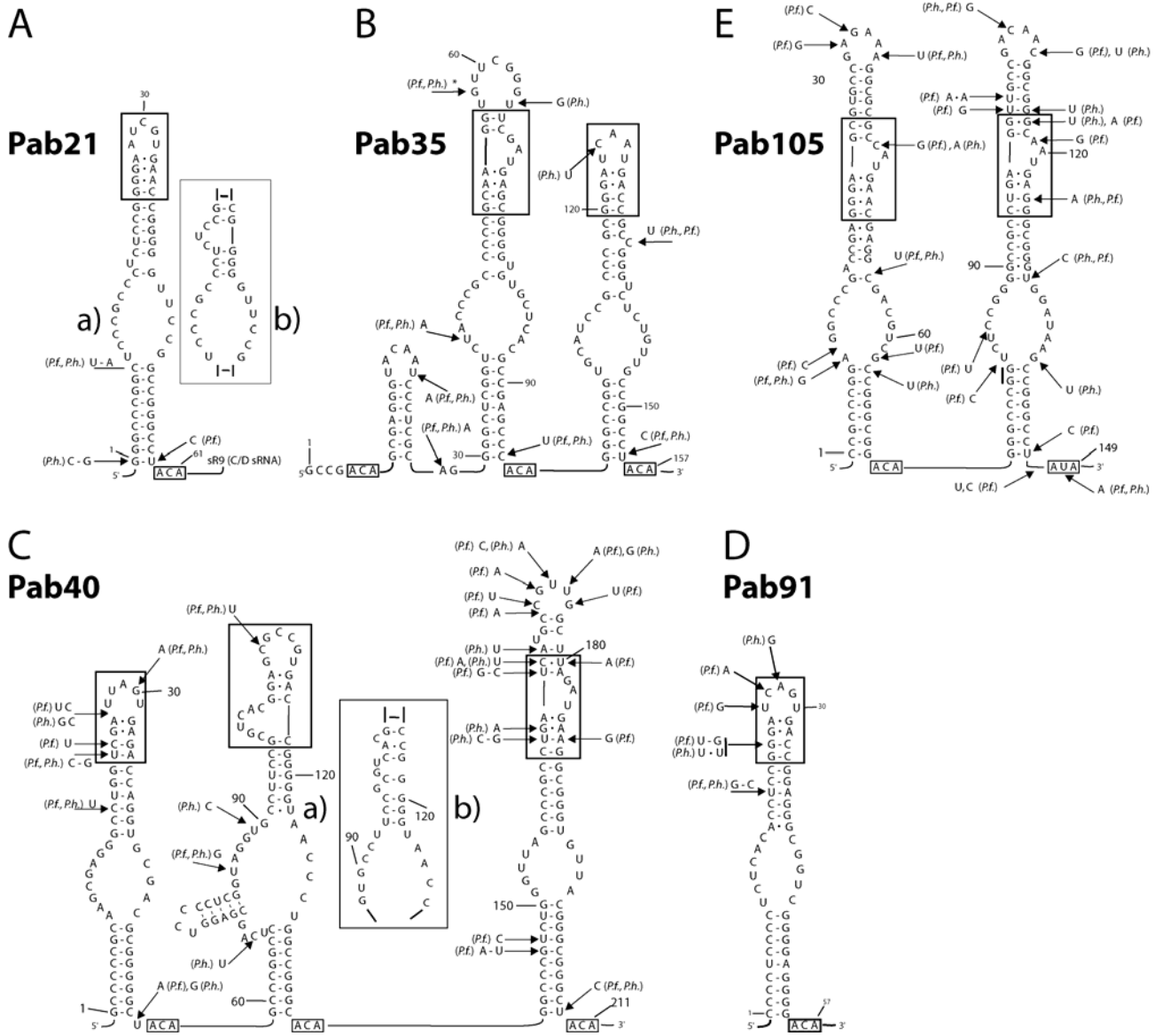


Figure R.10 : Structures secondaires proposées pour les 5 sRNA à boîtes H/ACA identifiés chez les trois espèces de *Pyrococcus*. Les séquences et les structures secondaires proposées par MFOLD sont présentées pour Pab21 (A), Pab35 (B), Pab40 (C), Pab91 (D) et Pab105 (E). Les boîtes ACA et motifs K-turn sont encadrés. Pour Pab21 et Pab40, deux repliements alternatifs sont proposés (a) et (b). Les différences et les mutations compensatoires entre les homologues provenant des espèces de *P. abyssi*, *P. horikoshii* (*P.h.*) et *P. furiosus* (*P.f.*) sont indiquées.

Par cette approche nous avons donc mis en évidence 5 sRNA H/ACA potentiels. Nous avons démontré que chacun de ces ARN est bien exprimé chez *P. abyssi* par des expériences de Northern Blot à partir d'extraits d'ARN totaux. Nous avons aussi pu prédire une ou plusieurs cibles potentielles pour chacun des motifs H/ACA qu'ils contenaient (Tableau R. 2).

sRNA H/ACA de <i>P. abyssi</i>	Cibles prédites dans les ARNr de <i>P. abyssi</i>
Pab21	16S 891
	16S 892 *
Pab35	stb 1 : 16S 1122
	stb 1 : 16S 922
	stb 1 : 23S 2930
	stb 1 : 23S 2672
	stb 2 : 23S 2250
	stb 2 : 23S 2549
	stb 2 : 23S 2697 *
Pab40	stb 1 : 23S 2588 *
	stb 2 : 23S 278 *
	stb 2 : 23S 1932
	stb 3 : 23S 2016 *
Pab91	23S 2685
Pab105	stb 1 : 23S 2552 *
	stb 1 : 23S 2554
	stb 2 : 16S 995
	stb 2 : 16S 27 *
	stb 2 : 23S 1377
	stb 2 : 23S 2794

Tableau R. 2 : Résumé des sRNA H/ACA identifiés et de leurs cibles prédites.

* : positions prédites par une autre étude pour Pf1, Pf3, Pf6, Pf7 qui sont les homologues respectivement de Pab21, Pab105, Pab35, Pab40 (Rozhdestvensky *et al.*, 2003). stb : structure tige-boucle.

Il est intéressant de noter que le fragment IRC N21 renferme le gène du motif Pab21 directement en amont du sRNA C/D sR9 déjà identifié (Fig. R.10A). Il n'est pas exclu que l'ARN exprimé contienne les deux motifs guidant l'un une 2'-O-méthylation et l'autre une pseudouridylation. En effet, d'après nos données de Northern Blot réalisés à partir d'extraits d'ARN totaux de *P. abyssi*, une sonde radioactive complémentaire au sRNA Pab21 révèle la présence de deux transcrits dont la taille correspond au sRNA Pab21 seul ou au sRNA Pab21 associé au sRNA C/D sR9 (Fig. R.11).

Par ailleurs, en parallèle à mes travaux, l'équipe d'A. Huttenhofer a aussi réalisé une étude expérimentale des sRNA H/ACA de *P. furiosus* (Rozhdestvensky *et al.*, 2003). Pour cela, ils se sont appuyé sur la recherche d'ARNnc que l'équipe de S. Eddy avait réalisée sur *P. furiosus* en utilisant le biais de la composition en G/C et que nous avons décrit ci-dessus (Klein *et al.*, 2002). Parmi les 5 ARNnc que l'équipe de S. Eddy avait montré être exprimés chez *P. furiosus*, l'équipe de A. Huttenhofer a détecté 4 sRNA H/ACA potentiels. Au final, parmi les 5 sRNA H/ACA que nous avons identifiés, 4 d'entre eux, Pab21, Pab105, Pab35, Pab40 se sont avérés être des homologues des sRNA identifiés chez *P. furiosus*, respectivement Pf1, Pf3, Pf6 et Pf7 (Rozhdestvensky *et al.*, 2003), et 2 d'entre eux, Pab40 et Pab91 sont des homologues respectivement de Afu4 et Afu46 détectés

chez *A. fulgidus* (Tang *et al.*, 2002). Il faut remarquer que les s quences des sRNA H/ACA de *P. abyssi* pr sentent de nombreuses similarit s avec les s quences des sRNA H/ACA homologues de *P. furiosus* (Fig. R.10), alors que les s quences des sRNA H/ACA de *P. abyssi* pr sentent tr s peu de similarit  de s quences avec les s quences des sRNA H/ACA homologues de *A. fulgidus*. Dans ce cas, les sRNA sont consid r s comme homologues parce qu'ils ciblent des positions  quivalentes dans les ARNr.

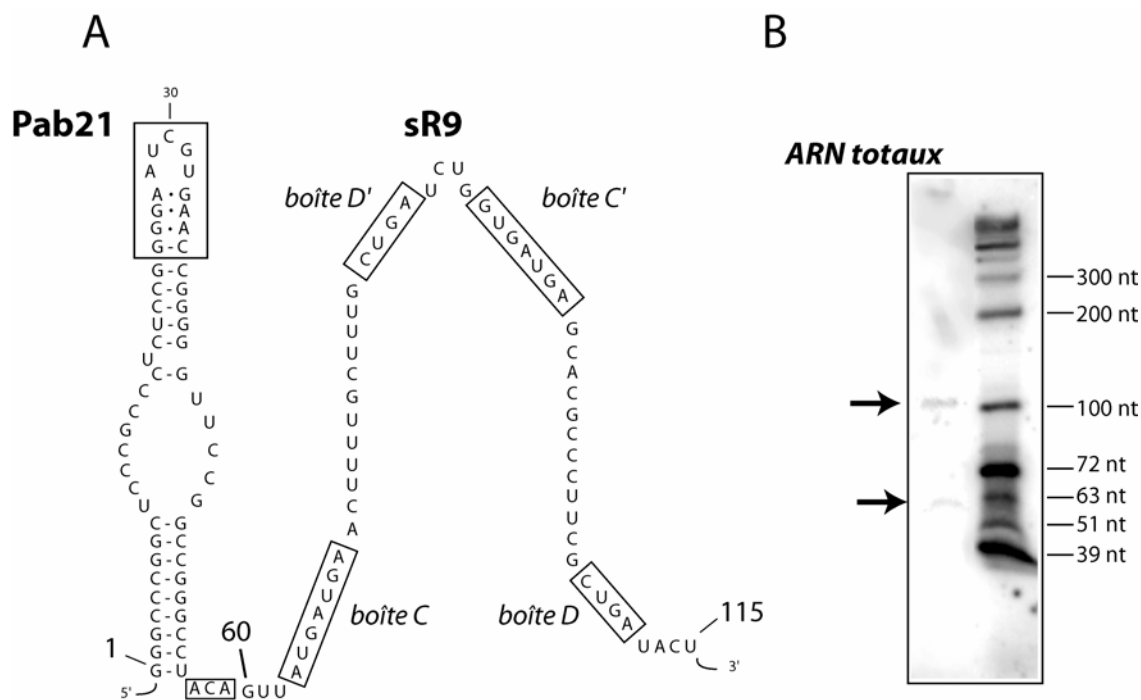


Figure R.11 : V rification par Northern Blot de l'expression du candidat sRNA Pab21.

A. S quence et structure secondaire propos e pour les sRNA H/ACA Pab21 et sRNA C/D sR9. **B.** R sultats de Northern Blot r alis    partir d'ARN totaux extraits de *P. abyssi* avec une sonde compl mentaire   une partie du sRNA Pab21 qui est marqu e au $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$ en 5'.

Au cours des travaux de l' quipe d'A. Huttenhofer, une structure en K-turn ou pseudo-K-turn a  t  mise en  vidence dans chaque structure tige-boucle de sRNA H/ACA. De plus, il a  t  d montr  la fixation de L7Ae sur le motif en K-turn (Rozhdestvensky *et al.*, 2003).

L'identit  des cibles des structures tige-boucles assurant la pseudouridylation d pend de la structure 2D propos e pour ces structures en tige-boucle. En effet, les diff rences de structures propos es ont un impact sur la proposition des cibles (Tableau R.2). Ainsi on remarque que des cibles diff rentes ont  t  pr dites pour une m me structure en tige-boucle. De plus, certaines modifications seraient r alis es par le biais de deux sRNA H/ACA diff rents. Pour v rifier ces hypoth ses il  tait indispensable d'avoir un moyen de v rifier exp rimentalement quelles  taient ces cibles.

Dans ce but, avec B. Charpentier nous avons choisi le sRNA Pab91 dont l'homologue n'avait pas été mis en évidence chez *P. furiosus* du fait de la non expression du candidat Pf4 (Klein *et al.*, 2002). Celui-ci présentait l'avantage de ne contenir qu'une seule structure tige-boucle. Nous avons développé une méthode de reconstitution de particules H/ACA actives à partir du sRNA Pab91 transcrit *in vitro* et des protéines produites sous forme recombinantes.

Comme nous disposions de la séquence de plusieurs homologues de sRNA dans 3 espèces de *Pyrococcus* nous pouvions nous appuyer sur la phylogénie pour proposer une structure 2D précise pour les structures tige-boucles des 5 sRNA que nous avons identifiés (Fig. R.10). Pour les sRNA Pab35 et Pab105 la structure que nous proposons est en accord avec celles proposées pour les homologues de *P. furiosus* (Rozhdestvensky *et al.*, 2003). Par contre pour les sRNA Pab40 et Pab21, nous proposons des structures différentes qui ont pour conséquences de nous conduire à des prédictions de cibles différentes de celles proposées par l'équipe d'A. Huttenhofer.

Ces résultats ont donc principalement été exploités au travers des chapitres suivants qui présentent, d'une part, l'étude de l'assemblage des sRNP à boîtes H/ACA et la reconstitution de particules actives *in vitro* et, d'autre part, l'étude de la contribution des guides sRNA à boîtes H/ACA à la formation de résidus Ψ dans les ARNr de *P. abyssi*.

3. Etude de l'interaction des sRNA H/ACA avec les protéines associées et reconstitution de particules actives in vitro

A. Contexte et intérêts

Alors que des snoRNP C/D d'eucaryotes n'avaient jamais pu être reconstituées *in vitro*, des particules de sRNP à boîtes C/D issues d'archaea avaient pu être reconstituées à partir de leur constituants (Omer *et al.*, 2002). Cette étude, comme nous l'avons vu dans la partie introductive de ce manuscrit, avait permis de réaliser des études fines sur l'assemblage et le mécanisme d'action des sRNP à boîtes C/D (Omer *et al.*, 2002 ; Bortolin *et al.*, 2003 ; Rashid *et al.*, 2003 ; Tran *et al.*, 2003 ; Singh *et al.*, 2004 ; Tran *et al.*, 2005). La reconstitution de particules sRNP H/ACA actives devait pouvoir offrir les mêmes possibilités d'étude puisque dans ce cas aussi, aucune reconstitution complète n'avait été réalisée du fait de l'insolubilité des protéines eucaryotes lorsque produites sous forme recombinantes. Comme nous l'avons déjà vu dans l'introduction, les reconstitutions réalisées sur le système eucaryote n'avaient pu être réalisées qu'à partir d'extraits cellulaires (Wang et Meier, 2004). On savait, chez les eucaryotes, que les particules snoRNP H/ACA contiennent la protéine Nhp2p qui appartient à la famille des protéines L30/L7Ae qui interagissent avec l'ARN. Cependant, Nhp2p n'a qu'une très faible

spécificité d'interaction (Wang et Meier, 2004). Vu l'homologie entre Nhp2p et L7Ae et la démonstration par l'équipe d'A. Huttenhofer de la fixation de L7Ae sur les sRNA H/ACA, on pouvait supposer que Nhp2p était l'homologue de L7Ae dans les snoRNP eucaryotes (Rozhdestvensky *et al.*, 2003). Les snoRNP eucaryotes contiennent par ailleurs la protéine Cbf5p/Dyskérine qui présente des motifs de séquence retrouvés dans les ARN:Ψ-synthases et deux autres protéines, Nop10p et Gar1p (Hamma *et al.*, 2006). Chez *S. cerevisiae*, les protéines Cbf5p, Nop10p et Nhp2p sont essentielles à la croissance cellulaire (Henras *et al.*, 1998). Il avait été montré que Cbf5p interagit à la fois avec Gar1p et Nop10p (Henras *et al.*, 2004). D'après les premiers travaux de reconstitution réalisés à partir d'extraits cellulaires par l'équipe de T. Mayer, il avait été proposé qu'un complexe soit formé entre CBF5, NOP10, GAR1 et NHP2 et que ce complexe cytoplasmique se fixe sur les snoRNA dans le noyau (Wang et Meier, 2004). La découverte de NAF1 par l'équipe de Y. Henry a remis ce modèle en question. L'idée est maintenant qu'un assemblage aurait lieu tôt après ou pendant la transcription et impliquerait les protéines CBF5, NOP10, NHP2 et NAF1, la protéine GAR1 remplaçant ensuite la protéine NAF1 (Darzacq *et al.*, 2006 ; Kittur *et al.*, 2006 ; Leulliot *et al.*, 2007). Par ailleurs, les travaux de collaboration de E. Bertrand, de B. Charpentier et T. Kiss suggèrent fortement que cet assemblage est un processus assisté par de nombreuses protéines qui ne sont pas retrouvées dans les snoRNP (article soumis à J. Cell. Biol.). C'est sans doute pour cette raison qu'il n'est pas possible de réaliser une reconstitution des snoRNP H/ACA eucaryotes à partir de leurs constituants. D'où l'importance que revêtait la réalisation de reconstitution de particules actives d'archaea. On savait en effet qu'en plus de L7Ae qui pouvait être l'homologue de Nhp2p, il existait chez les archaea des gènes pouvant coder les protéines aCBF5, aNOP10, aGAR1 homologues des protéines eucaryotes Cbf5p/Dyskérine/NAP57, Nop10p et Gar1p respectivement (Watanabe et Gray, 2000). Il fallait démontrer qu'elles faisaient bien partie des particules sRNP d'archaea, et si c'était bien le cas, disséquer l'assemblage de ces particules et décrire les protéines responsables de l'activité ARN:Ψ-synthase. C'est ce que B. Charpentier a décidé de tester et j'ai participé à ce travail.

B. Les approches utilisées, les questions posées et les réponses obtenues

Nous avons choisi de produire, par transcription *in vitro*, le sRNA Pab91 pour les raisons précédemment évoquées. Les 4 protéines L7Ae, aCBF5, aGAR1 et aNOP10 ont été produites sous forme recombinantes chez *E. coli*.

Nous avons mis en place les conditions d'assemblage qui ont permis, par expériences de retard sur gel, l'analyse des interactions existantes au sein des particules. De plus, des tests d'activité enzymatiques reposant sur la méthode du plus proche voisin ont été développés (voir Matériel & Méthodes) et mis au point grâce à l'expérience du laboratoire sur l'étude des

ARN:Ψ-synthases. Ces tests enzymatiques, couplés à des mutations dans le sRNA Pab91 et dans la protéine aCBF5 nous ont permis de définir les éléments essentiels à la formation et à l'activité de la sRNP Pab91.

La première question à laquelle il fallait répondre était de savoir lesquelles de ces protéines se fixent directement sur le sRNA. Les données sur le sRNA Pab91 ont montré qu'à la fois L7Ae, comme attendu, mais aussi aCBF5 sont capables de se lier directement à l'ARN. Les expériences que j'ai réalisées sur le sRNA Pab21, et qui sont présentées après l'article, montrent que aNOP10 peut également se fixer sur certains des sRNA.

La seconde question était de savoir dans quel ordre peut se faire l'assemblage. Les données ont mis en évidence la formation d'un complexe aCBF5-aNOP10. Dans le cas de la sRNP Pab91, aNOP10 ne peut se lier à l'ARN que sous la forme de ce complexe. Le complexe aCBF5-aNOP10 fixé sur le sRNA permet le recrutement de l'ARN substrat. La protéine aNOP10 mais pas aGAR1 est nécessaire à ce recrutement.

Enfin, nous avons démontré par les tests d'activité des particules reconstituées que l'ensemble minimal de protéines requises pour avoir une activité est le couple aCBF5-aNOP10. Les protéines L7Ae et aGAR1 renforcent cependant fortement l'activité.

C. Résultats

Les résultats ont été publiés dans le journal Nucleic Acids Research. Cet article est présenté dans ce chapitre :

Publication N°1 :

Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation

Bruno Charpentier, Sébastien Muller, Christiane Branlant.

Nucleic Acids Research, 2005, Vol. 33, No. 10 p3133–3144

PUBLICATION N°1

Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation

Bruno Charpentier*, Sébastien Muller and Christiane Branlant

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 UHP-CNRS, Université des Sciences et Techniques Henri Poincaré Nancy I, 54506 Vandoeuvre-Lès-Nancy cedex, France

Received April 26, 2005; Revised and Accepted May 15, 2005

ABSTRACT

Pseudouridine (Ψ) are frequently modified residues in RNA. In Eukarya, their formation is catalyzed by enzymes or by ribonucleoprotein complexes (RNPs) containing H/ACA snoRNAs. H/ACA sRNA and putative ORFs for H/ACA sRNP proteins (L7Ae, aCBF5, aNOP10 and aGAR1) were found in Archaea. Here, by using *Pyrococcus abyssi* recombinant proteins and an *in vitro* transcribed *P.abyssi* H/ACA sRNA, we obtained the first complete *in vitro* reconstitution of an active H/ACA RNP. Both L7Ae and the aCBF5 RNA: Ψ synthase bind directly the sRNA; aCBF5 also interacts directly and independently with aNOP10 and aGAR1. Presence of aCBF5, aNOP10 and a U residue at the pseudouridylation site in the target RNA are required for RNA target recruitment. In agreement, we found that the aCBF5–aNOP10 pair is the minimal set of proteins needed for the formation of a particle active for pseudouridylation. However, particles more efficient in targeted pseudouridylation can be formed with the addition of proteins L7Ae and/or aGAR1. Although necessary for optimal activity, the conserved ACA motif in the sRNA was found to be not essential.

INTRODUCTION

Conversion of uridines into pseudouridines (Ψ) and riboses 2'-O-methylation are the two most frequent RNA post-transcriptional modifications [for review, (1,2)]. Some of these modifications are highly important for the activity of RNAs in translation or pre-mRNA splicing (3–7). In Eukarya and Archaea, both modifications can be generated by two distinct systems. Either a single protein has the specific RNA modification activity: RNA: Ψ -synthase [for review,

(8)] or RNA:2'-O-methylase activity [for review, (9)]; or the proteins carrying these enzymatic activities are found within ribonucleoprotein complexes (RNP) containing small RNAs. These RNAs are designated as small nucleolar RNAs (snoRNAs) in Eukarya and sRNAs in Archaea. They define the positions to be modified by base-pair interaction with the targeted RNAs [for review, (10–13)]. Based on conserved structural features, the snoRNAs and sRNAs can be divided into two classes: RNAs with C and D boxes (eukaryal C/D snoRNAs and archaeal C/D sRNAs) that guide 2'-O-methylations (14,15), and RNAs with H/ACA boxes (eukaryal H/ACA snoRNAs and archaeal H/ACA sRNAs) that guide RNA pseudouridylations (16,17). Among the proteins associated with C/D box RNAs, the Snurl3p/15.5 kDa protein in Eukarya and its L7Ae archaeal counterpart play an essential role in the initiation of RNP assembly (18–20). C/D sRNPs active in 2'-O-methylation have already been reconstituted from *in vitro* transcribed RNAs and recombinant proteins (18,21–24). Interestingly, the archaeal L7Ae protein is also able to bind the H/ACA sRNAs (25). However, limited information is available on the H/ACA sRNP assembly and function.

Information on H/ACA RNPs largely comes from eukaryal H/ACA snoRNP analyses. All H/ACA snoRNAs contain two irregular hairpin structures linked by a hinge sequence and followed by a short 3' tail. The ANANNA sequence (Box H) present in the hinge region and the ACA sequence (Box ACA), which is located three nucleotides upstream of the RNA 3' end, are essential for snoRNA stability and function (16,26,27). Each of the hairpin structures contains a central loop that carries a 9–13 bp bipartite antisense element. After hybridization of this loop with the target RNA, a pseudouridylation pocket is formed. In this structure, an unpaired dinucleotide of the target RNA is surrounded by two short bimolecular helices; it contains the U residue converted into a Ψ residue (17,28). Genetic data in yeast, as well as biochemical analysis of purified H/ACA snoRNPs, provided strong evidence for the presence of an evolutionarily conserved RNA: Ψ -synthase in H/ACA snoRNPs (Cbf5p in yeast, dyskerin or NAP57 in

*To whom correspondence should be addressed. Tel: +33 3 83 68 43 16; Fax: +33 3 83 68 43 07; Email: bruno.charpentier@maem.uhp-nancy.fr

human) (29–31). H/ACA snoRNPs contain three other core proteins (denoted Nhp2p, Nop10p and Gar1p in yeast, and NHP2, NOP10 and GAR1 in human) (32,33). Active H/ACA snoRNPs were immunopurified from HeLa cell extracts (34). By mild denaturing treatment of yeast H/ACA snoRNPs, a complex formed by proteins Cbf5p, Gar1p and Nop10p was obtained, showing the occurrence of protein–protein interaction in the RNPs (35). Recently, assays for assembly of mammalian H/ACA snoRNPs were performed using proteins produced by *in vitro* transcription/translation (36). In accordance with previous data obtained by immunoprecipitation experiments and *in vitro* reconstitution of H/ACA snoRNP in cellular extract (37,38), the authors demonstrated that the NOP10–NAP57 interaction is a prerequisite for NHP2 binding and that the NAP57–NOP10–NHP2 core trimer specifically recognizes H/ACA snoRNAs (36). In spite of these progresses in understanding how H/ACA snoRNP components interact together, up to now, functional H/ACA snoRNPs could only be obtained by incubation of a snoRNA transcript with a cytosolic extract (34,36).

Based on the successful reconstitution of active archaeal C/D sRNPs using *in vitro* transcribed RNA and recombinant proteins, we tested the possibility to reconstitute archaeal H/ACA sRNPs from recombinant molecules. Three H/ACA sRNAs have recently been identified in the hyperthermophile archaeon *Archaeoglobus fulgidus* (39). The secondary structures of these RNAs are quite less homogeneous than those of eukaryal H/ACA snoRNAs. They are folded into one, two or three stem–loop structures, each containing a pseudouridylation pocket (39). In addition, they all exhibit a sequence able to form a terminal pseudo K-turn motif recognized by the protein L7Ae (25,40,41). Protein L7Ae is proposed to have a common ancestor with the eukaryal Snu13p/15.5 kDa and Nhp2 proteins (20,33,42). In addition, based on DNA sequence analysis, homologues of the eukaryal Cbf5, Nop10 and Gar1 proteins were expected to be present in archaea (38,42). However, their interactions with H/ACA sRNAs have not been demonstrated.

Here, we report the reconstitution of a *Pyrococcus abyssi* pseudouridylation-competent particle by incubation of an *in vitro* transcribed H/ACA sRNA with the four recombinant archaeal L7Ae, aCBF5, aNOP10 and aGAR1 proteins. RNA–protein interactions were analyzed by gel retardation assays. An optimal guided pseudouridylation activity was obtained in the presence of the four core proteins. However, particles assembled in the absence of aGAR1 can specifically modify, albeit less rapidly, an RNA substrate. As it was the case for C/D sRNPs, the development of conditions for *in vitro* reconstitution of active H/ACA sRNPs is expected to open the door for further investigation on the architecture and the mechanism of action of these particles.

MATERIALS AND METHODS

Oligonucleotides

Pairs of primers used for PCR amplification of *P.abyssi* open reading frames (ORFs) and DNA templates for T7 *in vitro* transcription of Pab91 sRNA, and for site-directed mutagenesis can be obtained upon request (see also Supplementary Material).

Recombinant protein production

The ORFs of archaeal L7Ae, aNOP10, aGAR1 and aCBF5 proteins (GenBank accession nos C75109, CAB49761, CAB49230 and CAB49444, respectively) were PCR-amplified from the genomic DNA of the *Pyrococcus abyssi* GE5 strain. They were cloned in plasmid pGEX-6P-1 (Pharmacia) to produce GST fusion protein. A multi-step PCR strategy was used to generate the (Asp-82 → Ala, GAC → GCC) site-specific mutation in the aCBF5 domain II.

The recombinant GST-L7Ae, GST-aNOP10, GST-aGAR1, GST-aCBF5 and GST-aCBF5/D82A proteins were produced in *E.coli* BL21 CodonPlus cells (Novagen) and purified under native conditions using Glutathione-Sepharose 4B (Pharmacia) as previously described (43). The fusion proteins were cleaved on beads with 80 U of PreScission protease (Pharmacia) per ml of Glutathione-Sepharose bead suspension. Cleavage was performed overnight at 4°C followed by incubation at 65°C during 15 min. The precipitated contaminant proteins were eliminated by a 20 min centrifugation at 16 000 g. For their use in the GST pull-down experiments, the GST-fusion proteins were eluted from the beads by incubation with reduced glutathione following the manufacturer recommendations, and the reduced glutathione was subsequently exchanged with the PBS buffer by loading the eluted proteins on Econo-Pac 10DG columns (Biorad). The quality of protein preparations was assessed by SDS–PAGE and Coomassie staining.

In vitro binding assays

Purified GST or GST-fusion proteins were bound to Glutathione-Sepharose beads in PBS buffer. Recombinant proteins (1 µg each) were incubated for 1 h at 25°C with 15 µl of beads in 100 µl of PBS buffer containing 500 mM NaCl. Bound proteins were washed three times with PBS/NaCl buffer and eluted by boiling in Laemmli loading buffer. After SDS–PAGE, proteins were visualized by Coomassie Blue staining.

DNA templates used for *in vitro* transcription

The Pab91 sRNA sequence was PCR-amplified from the *Pyrococcus abyssi* GE5 genomic DNA using a forward primer containing the T7 RNA polymerase promoter. The DNA templates for production of the Pab91mtKT, Pab91mtACA and Pab91mtKTmtACA variant sRNAs were produced by PCR amplification. Before use, all the PCR amplification products were cloned in the pTAdv vector (Clontech) and their sequences were checked by dideoxynucleotide sequencing.

In vitro transcription of guide and target RNAs

The WT and variant Pab91 sRNAs, and the target RNAs, were *in vitro* transcribed by T7 RNA polymerase using PCR amplification fragments. They were purified by gel electrophoresis. For production of cold RNAs, transcription reactions were carried out as previously described (44). The same conditions were used for synthesis of uniformly labeled RNAs, except that the reaction was performed with 20 µCi of [α -³²P]CTP (800 Ci/mmol) (Amersham), 4 mM each of ATP, UTP and GTP, and 0.13 mM CTP. Subsequent RNA treatments were as previously described (45).

Electrophoresis mobility shift assays (EMSA)

About 50 fmol of uniformly ^{32}P -labeled RNA were mixed with 1 μg of yeast tRNA (Roche) in 3.5 μl of buffer D (150 mM KCl, 1.5 mM MgCl_2 , 0.2 mM EDTA and 20 mM HEPES, pH 7.9). For competition experiments increasing amounts of cold competitor sRNA were added. To study the sRNP-target RNA association, unlabelled RNA-S (2.4 pmol in most experiments or more as indicated in the figures) was added at this stage. Except for K_d determinations, each recombinant proteins was used at a 200 nM concentration and the mixture (4.5 μl final volume in buffer D) was incubated for 10 min at 65 or 20°C. RNA-protein complexes were resolved at room temperature in 6% Triborate non-denaturing polyacrylamide gels, as previously described (44). The dried gels were analyzed with a phosphorimager (Typhoon 9410, Amersham Biosciences) and the amounts of radioactivity in the bands were estimated with the ImageQuant software. The percentage of RNA in each RNP was calculated from the radioactivity in each band relative to the total radioactivity in the lane.

In vitro sRNP guide pseudouridylation assay

Unlabeled sRNA (4 pmol, ~ 800 nM) and $[\alpha\text{-}^{32}\text{P}]\text{CTP}$ -labeled target RNA-S or RNA-L (150 fmol, ~ 30 nM) were mixed at room temperature in buffer D and treated as described for EMSA assays. The proteins (200 nM each) were added at room temperature; the final 4.5 μl solution was incubated at 65 or 20°C. For time course experiments, first, the reaction volume was scaled up to 50 μl , using the same RNA and protein concentrations, second, the reaction was started by addition of the proteins after incubation of the RNA solution at 65°C for 10 min. At time intervals, 4.5 μl aliquots were collected. For each aliquot, the reaction was stopped by addition of 150 μl of H_2O and 150 μl of a phenol:chloroform (1:1) mix. The extracted RNAs were ethanol-precipitated. The modified RNA targets were totally digested with 0.5 U of T2 RNase (GibcoBRL) for 12 h at 37°C. The resulting 3'-mononucleotides were chromatographed on thin-layer cellulose plates as described elsewhere (46,47). For reaction performed on the RNA-S, one dimension chromatography was sufficient to fractionate the three labeled 3'-mononucleotides (Figure S1). The radioactivity in the spots or the bands were quantified with a phosphorimager (Amersham Biosciences) using the ImageQuant software. Taking into account the total number of U residues in the target RNA, we determined the amount of Ψ formed (expressed in moles per mole of target RNA).

RESULTS

Proteins L7Ae and aCBF5 bind directly to the Pab91 sRNA

By using computer analysis, we searched for the presence of H/ACA sRNAs in the *Pyrococcus* genera and found in *P.abyssi* a small H/ACA sRNA, which we denoted as Pab91 (Muller, S., Leclerc, F., Behm-Ansmant, I., Charpentier, B. and Branlant, C., in preparation). Its counterpart in *A.fulgidus* is the recently identified Afu46 sRNA (39). We showed that sRNA Pab91 directs pseudouridylation at position 2685 in the *P.abyssi* 23S rRNA (Muller, S., Leclerc, F., Behm-Ansmant, I., Charpentier, B. and Branlant, C., in preparation). The secondary

structures of *P.abyssi* Pab91 and *A.fulgidus* Afu46 sRNAs consist of a single stem-loop structure (Figure 1A). It is closed by an L7Ae binding site, which consists of two A•G and G•A sheared pairs, linked by a 4 nt long terminal loop with a 3' terminal U residue (20,25,39). The Pab91 stem-loop structure is flanked by a short 3' sequence containing the conserved ACA triplet (Figure 1A). We chose this simple sRNA molecule as a model to develop conditions for *in vitro* reconstitution of archaeal H/ACA sRNPs. The Pab91 sRNA and its variants were synthesized by *in vitro* transcription and the four core proteins aCBF5, aNOP10, aGAR1 and L7Ae were produced as recombinant proteins in *E.coli* and purified.

Except for protein L7Ae, nothing was known on the individual RNA binding capacity of the archaeal H/ACA core proteins, thus, we first tested their direct binding on the WT Pab91 sRNA. Incubation was performed at 65°C, in the presence of 150 mM KCl and a yeast tRNA mixture in excess was used as the competitor (see Materials and Methods). The resulting complexes were analyzed by electrophoresis mobility-shift assays (EMSA) performed at room temperature (Figure 1B). The WT Pab91 sRNA was totally shifted into an RNP1 complex by incubation with protein L7Ae at a 200 nM concentration. In accordance with previous data on the *A.fulgidus* Afu46 sRNA (25), an estimated K_d of ~ 100 nM was found for complex RNP1 (Figure 1C). As expected from the present knowledge on the determinants required for L7Ae/Snu13p binding on the K-turn structures (22,25,40,41), no binding of protein L7Ae was observed after disruption of the two A•G and G•A sheared pairs in the L7Ae recognition motif (sRNA variant Pab91mtGA, Figures 1A and 2D). In contrast, the ACA triplet was not required for L7Ae binding (sRNA variant Pab91mtACA, Figures 1A and 2C). Concerning the three other proteins, only aCBF5 led to the formation of an RNP complex (RNP2 in Figure 1B). However, even at high aCBF5 protein concentration, its yield of formation was low (Figure 1C). We thus tested the specificity of this interaction by competition experiments using as the competitor, WT or variant Pab91 sRNAs (Figure 1D and E). As the Pab91mtACA sRNA, with a UGU triplet instead of the ACA triplet (i) did not compete formation of RNP2 by the WT sRNA, and (ii) was not able to form an RNP2 complex (Figure 2C), we concluded that the presence of the ACA box is required for aCBF5 binding on the Pab91 sRNA. In contrast, formation of RNP2 with the Pab91mtGA variant sRNA (Figure 2D) showed that an intact L7Ae binding site is not required for RNP2 formation. Hence, in the conditions used, only protein L7Ae and to a lesser extent protein aCBF5 interact directly with the Pab91 sRNA, in addition, aCBF5 binding depends upon the presence of the ACA motif.

Protein aCBF5 but not L7Ae favors protein aNOP10 recruitment

As binding of protein L7Ae strongly favors C/D sRNP assembly (18), we tested whether it could increase aCBF5 binding on a H/ACA sRNA. When the two proteins were incubated with the WT Pab91 sRNA, a second complex (RNP3) was detected in addition to RNP1 (Figure 2A). According to its electrophoretic mobility, it contained both L7Ae and aCBF5. As found for RNP2, RNP3 was obtained in low yield compared to RNP1. However, the yield of RNP3 formation was higher than that of

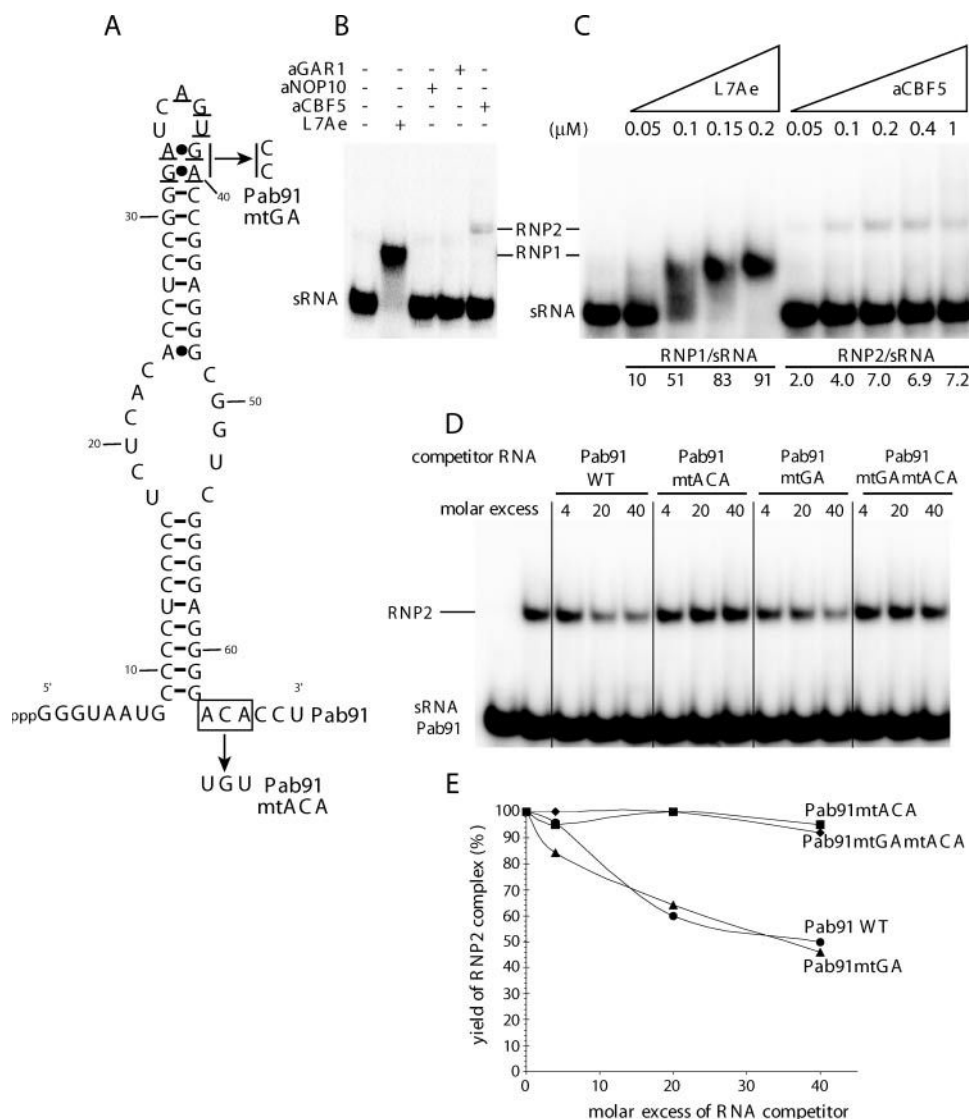


Figure 1. Both L7Ae and aCBF5 bind directly the Pab91 sRNA. (A) Secondary structure model of the Pab91 sRNA, folded according to the structure proposed for its counterpart Afu46 sRNA in *A. fulgidus* (39). The ACA motif is boxed, residues of the L7Ae binding site are underlined. The GA to CC and ACA to UGU substitutions in variants Pab91mtGA, Pab91mtACA and Pab91mtGAmtACA are shown. (B–D) Test of the individual binding of the core proteins by EMSA. *In vitro* transcribed ³²P-labeled WT Pab91 sRNA (50 fmol) (Panels B and C), or a mix of labeled WT RNA (50 fmol) and competitor unlabeled WT or variant Pab91 sRNA (0.2, 1 or 2 pmol, corresponding to 4, 20 and 40 molar excess, respectively) (Panel D) were incubated at 65°C with one of the L7Ae, aNOP10, aGAR1 or aCBF5 proteins (200 nM each in panels B and at concentrations increasing from 0.05 to 1 μM in panel C) or with protein aCBF5 (200 nM in panel D). The RNP complexes formed were resolved by electrophoresis at room temperature on native 6% polyacrylamide gels and visualized by exposure to Phosphorimager screens. Positions of the RNA/protein complexes (RNP1 and RNP2) and the free sRNA are indicated. The RNP yields (given below each lane in panel C) were expressed as the percentage of Pab91 sRNA in RNP1 or RNP2 as estimated by radioactivity measurement. The gel in panel D was exposed for a longer time as compared to the ones in panels B and C. (E) The RNP2 yields in the gel of panel D were estimated as in panel C. The values were expressed as a percentage of the RNP2 yield formed with the WT Pab91 sRNA in absence of competitor RNA (value of 100%). The values were plotted against the molar excess of cold competitor RNAs.

RNP2 (Figure 2A), the mean value for three different experiments was of 7.5 ± 1.5 for RNP2 and of 16.6 ± 1.3 for RNP3. Hence, the presence of protein L7Ae may have slightly increased aCBF5 association with the RNA. Whereas only RNP1 was formed in the presence of the two L7Ae and aNOP10 proteins (Figure 2A), an RNP4 complex was formed with the aCBF5–aNOP10 protein pair, RNP4 had a lower mobility compared to RNP2 (Figure 2A). Hence, protein aCBF5, but not L7Ae, allows the recruitment of protein aNOP10.

Incubation of the sRNA in the presence of the three L7Ae, aCBF5 and aNOP10 proteins (Figure 2B) led to the formation of a larger RNP5 complex that was expected to contain the

three proteins. However, the large amount of residual RNP1 complex reflected a partial association of proteins aNOP10 and aCBF5 with the L7Ae/sRNA complex. The absence of RNP3 complex suggested either the rapid association of protein aNOP10 after binding of protein aCBF5 or the association of a complex formed by these two proteins. Binding of protein aCBF5 was probably a limiting step in our *in vitro* reconstitution experiments. These assumptions are in agreement with our observation of similar RNA-sequence requirements for binding of aNOP10 and aCBF5: namely, the necessity of the ACA motif, but not of the two A•G and G•A sheared-pairs (Figure 2C and D). Accordingly, no RNP4 was formed when

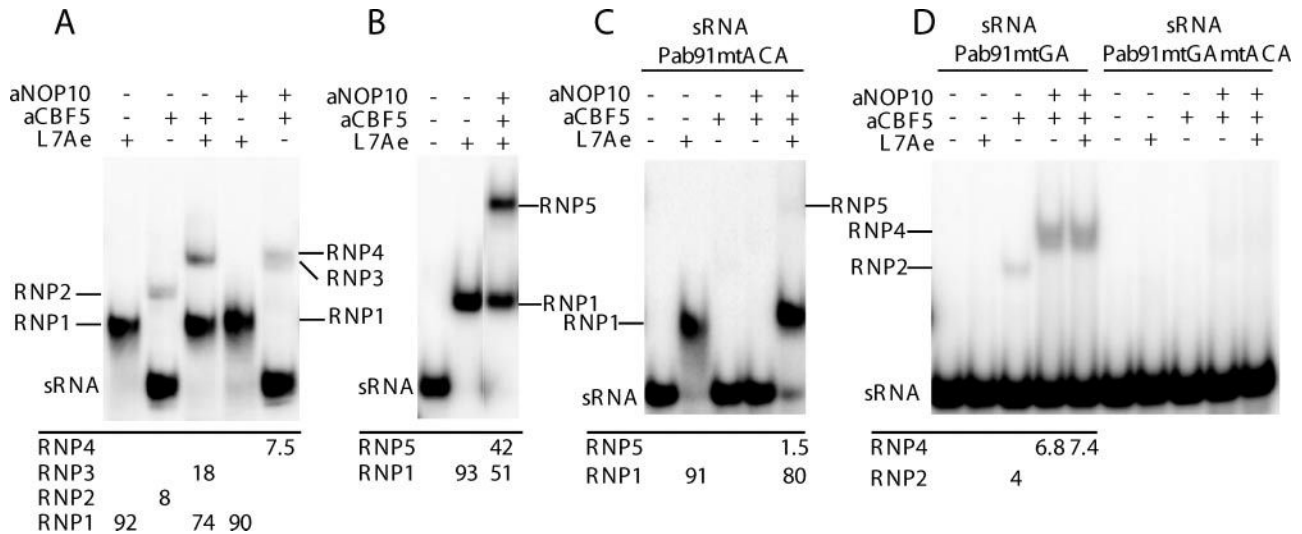


Figure 2. aCBF5 is required for aNOP10 association and association of both proteins depends upon the ACA sequence. (A and B) Identification by EMSA of the RNP complexes formed with the radiolabeled WT Pab91 sRNA (50 fmol) and various combinations of the core proteins (200 nM each). Incubation was at 65°C. The names of RNPs revealed by phosphorimager analysis are indicated. The percentage of sRNA present in each RNP as referred to the total RNA amount is indicated below each lane. (C and D) The same experiments as in panels A and B were repeated with the variant Pab91mtACA sRNA (Panel C), Pab91mtGA (Panel D) and Pab91mtGAmTACA (Panel D) sRNAs.

the Pab91mtACA sRNA variant was incubated with proteins aNOP10 and aCBF5 (Figure 2C), whereas both RNP2 and RNP4 complexes were formed with the Pab91mtGA sRNA variant (Figure 2D). Only a very low amount of RNP5 was detected upon incubation of the Pab91mtACA sRNA with the aNOP10, aCBF5 and L7Ae proteins (Figure 2C).

To test for a possible interaction between aCBF5 and aNOP10, we coated the GST-aNOP10 fusion protein on Glutathione-Sepharose beads and tested whether protein aCBF5 binds to these beads in the presence of 500 mM NaCl. As shown by SDS-PAGE analysis, protein aCBF5 but not L7Ae was retained (Figure 3). As aCBF5 was not retained in the control experiment performed with the GST alone, we concluded that aCBF5 and aNOP10 can interact together, and the interaction is stable even at high salt concentration.

Presence of the three aCBF5, aNOP10 and L7Ae proteins is required for efficient association of an RNA substrate

Next, we investigated whether the identified RNP complexes can interact with an RNA substrate. We produced a 22 nt long RNA substrate (RNA-S, Figure 4A), carrying the *P. abyssi* 23S rRNA fragment extending from positions 2673 to 2692, that is targeted by the Pab91 sRNA (Muller, S., Leclerc, F., Behm-Ansmant, I., Charpentier, B. and Branlant, C., in preparation). By using an unlabeled RNA-S target and a labeled sRNA, we first verified that, in the absence of protein and in the buffer and temperature conditions used for complex assembly, no stable interaction takes place between the radiolabeled guide sRNA and the targeted RNA (Figure 4B, lane 1). Then, we tested whether the presence of the RNA substrate was altering the electrophoretic mobility of complexes formed by incubation of the WT sRNA with various combinations of the H/ACA sRNP proteins, at 20 or 65°C.



Figure 3. The GST-aNOP10 and GST-aGAR1 fusion proteins bind aCBF5 *in vitro*. Different combinations (as indicated above each lanes) of purified recombinant proteins aCBF5 (C), aNOP10 (N), aGAR1 (G) and L7Ae (L) were incubated in buffer containing 500 mM NaCl in the presence of Glutathione-Sepharose beads coated with the GST, GST-aGAR1 or GST-aNOP10 proteins. The bound proteins were fractionated by SDS-PAGE and stained by Coomassie blue.

When proteins L7Ae and aCBF5 were incubated individually with the sRNA, addition of the RNA target did not modify significantly the yield and electrophoretic mobilities of RNP1 and RNP2 (Figure 4B). In contrast, addition of the RNA target upon incubation of the WT sRNA with the aCBF5-aNOP10 protein pair, increased the yield of RNP formation (Figure 4C, lanes 2–5). The RNP4/RNA-S complex obtained was designated as complex CI. The differences in the percentages of sRNAs in RNP4 and CI in three distinct experiments are significant as judged by a $P < 0.01$ measured by a Student's paired *t*-test. In addition, the RNP1 and RNP5 complexes detected upon incubation of the guide RNA with the three L7Ae, aCBF5 and aNOP10 proteins were almost completely

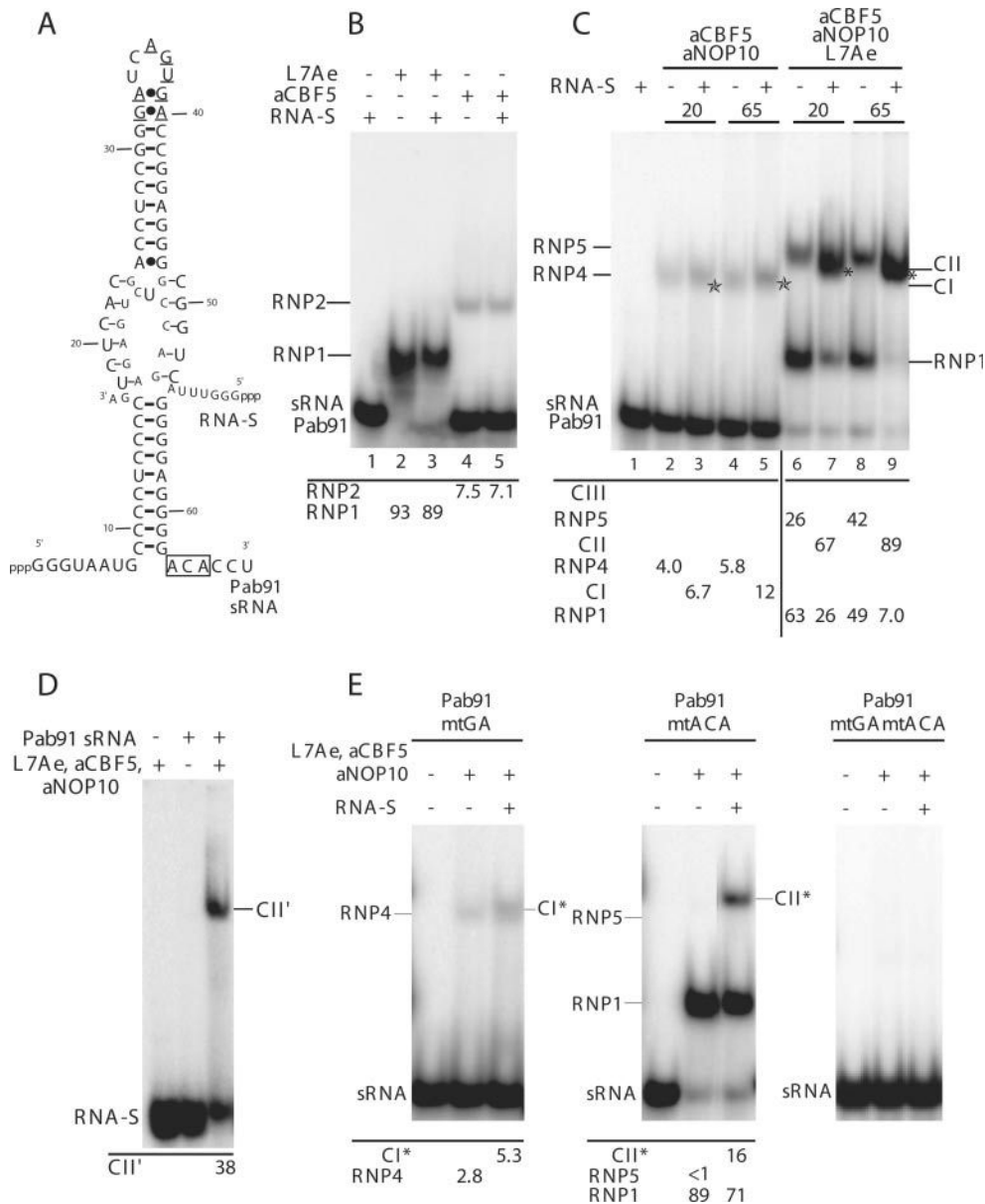


Figure 4. Presence of the target RNA favors sRNP assembly *in vitro*. (A) Model of the interaction between the Pab91 sRNA and the RNA-S target. The U residue converted into a Ψ residue is in bold. (B and C) Identification by gel shift assays of complexes formed with a radiolabeled WT Pab91 sRNA (50 fmol) and various combinations of the core proteins (200 nM each), in the presence or absence of a 50 molar excess of unlabeled target RNA-S (as indicated above the lanes). Incubation was at 65°C in panel B, and 20 or 65°C in panel C. The names of RNP complexes identified by phosphorimager analysis are indicated, as well as the percentage of Pab91 sRNA in each complex. (D) Same experiment as in lane 9 of panel C, except that RNA-S was the radiolabeled RNA (150 fmol) and that 2 pmol of unlabeled Pab91 sRNA were used. (E) The RNA-S target stabilizes sRNP complex formation with the variant Pab91mtGA and Pab91mtACA sRNAs. Similar experiments as in panel C, lanes 8 and 9, were performed with the three variant Pab91 sRNAs.

converted into a unique CII complex in the presence of the RNA substrate (Figure 4C, lanes 6–9). Complex CII had an electrophoretic mobility higher than that of RNP5. The yield of complex CII formation was slightly higher at 65°C compared to 20°C. To verify the incorporation of the RNA substrate into the CII complex, the same experiment was repeated using an unlabeled guide sRNA and a radiolabeled RNA-S target. As expected, a large part of the RNA substrate was shifted into a CII' complex, which had an electrophoretic mobility similar to that of complex CII and is expected to be identical (Figure 4D). Altogether, the data suggested an

efficient binding of the RNA substrate on the RNP complex when protein aNOP10 was present in this complex.

Presence of a uridine at the isomerization site reinforces the yield of complex CII formation

To test for the specificity of the RNA target selection by RNP5, we incubated the WT radiolabeled sRNA with the L7Ae, aCBF5 and aNOP10 proteins, in the presence of increasing amounts of the WT RNA substrate (RNA-S) or a variant RNA substrate (RNA-S-mtU). This variant RNA had a U to

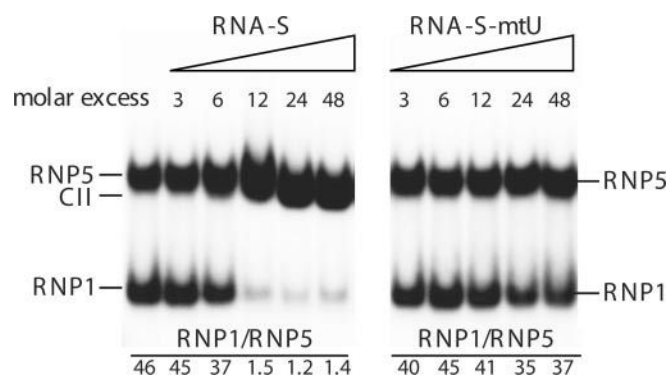


Figure 5. The targeted U residue is required for stable sRNP/RNA-S complex formation. Incubations was as in Figure 4C, lane 9, except that the WT (RNA-S) or mutated (RNA-S-mtU) target RNA was used at concentrations varying from 22 to 530 nM (molar excess from 3 to 48, as indicated above the lanes). The percentages of Pab91 sRNA in RNP5 and complex CII are given below the lanes.

C substitution at the isomerization site. Interestingly, whereas in the presence of 0.6 pmol of RNA-S (12-fold molar excess), almost all the guide sRNA was found in complex CII (Figure 5), a large part of the RNP1 was not displaced into complex CII when the same experiment was performed with RNA-S-mtU. Even at a 48-fold molar excess of RNA-S-mtU, RNP1 was still detected (Figure 5). This strongly suggests that the occurrence of a U residue at the isomerization site is important for efficient formation of a stable sRNP/RNA-S complex.

Binding of the targeted RNA requires at least one of the two sRNA conserved motifs

As RNP4 can be formed with the Pab91mtGA variant sRNA and as RNP5 was detected in trace amounts with the Pab91mtACA variant sRNA (Figure 2), we tested whether the RNA-S target can associate with these RNPs. As shown in Figure 4E, a CI-like complex (CI*), similar to the CI complex obtained with the WT sRNA, was formed with the Pab91mtGA variant and a CII-like complex (CII*) was also obtained with the Pab91mtACAv variant. The yield of CII* formation was lower (16%) compared to that of CII (89%). However, taking into account the very low level of RNP5 formation on the Pab91mtACA variant in the absence of RNA substrate, this latter strongly reinforces complex formation on the Pab91mtACA variant sRNA. The positive effect of substrate addition is not observed in the double mutant (sRNA variant Pab91mtGAmtACA, Figure 4E).

A negative effect of protein aGAR1 on RNP assembly is counteracted by the RNA target

Surprisingly, incubation of the WT Pab91 sRNA with proteins aCBF5 and aGAR1 revealed a negative effect of protein aGAR1 on RNP2 formation (Figure 6B). In addition, in the presence of aGAR1, no large RNP complex was detected when the three L7Ae, aCBF5 and aNOP10 proteins were incubated with the sRNA (Figure 6C). Even the RNP1 complex was found in lower yields in these conditions. Thus, in the incubation conditions used, protein aGAR1 had either a negative effect on protein assembly or a destabilization effect on the complex formed.

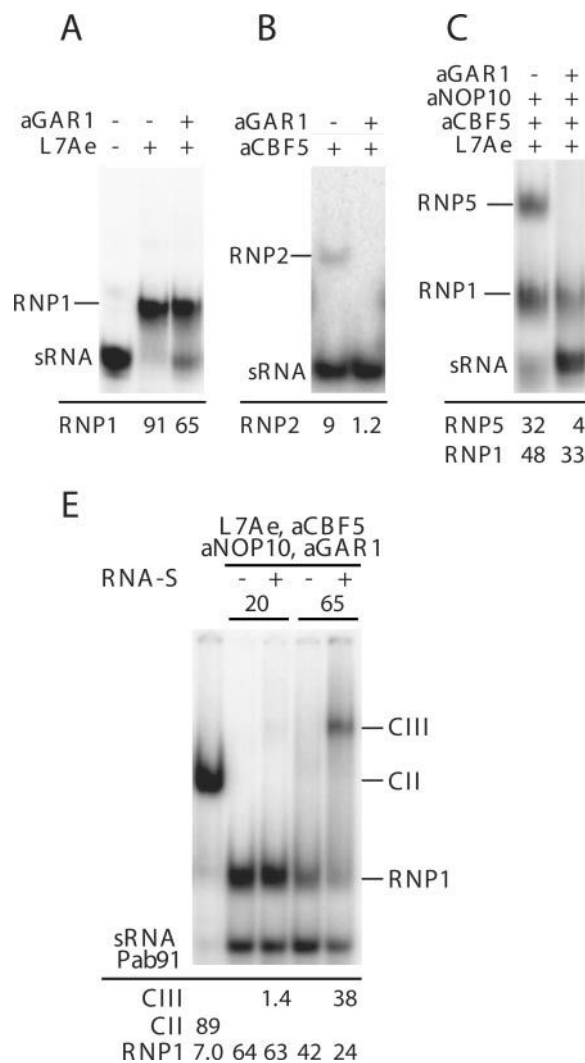


Figure 6. A decreased yield of RNP1, RNP2 and RNP5 formation in the presence of protein aGAR1 is counteracted by RNA-S. (A–C) The radiolabeled Pab91 sRNA (50 fmol) was incubated at 65°C with individual proteins or protein combinations, as indicated above the panels, in the absence or the presence of protein aGAR1. All the proteins were used at a 200 nM concentration. The percentage of Pab91 sRNA in RNP1 (Panel A), RNP2 (Panel B), and RNP5 (Panel C), is given below the lanes. (D) Same experiment as in Figure 4C, lanes 6–9, with the L7Ae, aCBF5, aNOP10 and aGAR1 protein set, and RNA-S substrate when indicated on top of the lanes.

We hypothesized that the aGAR1 negative effect on sRNP formation may be due to the formation of an aCBF5–aGAR1 interaction that destabilized the aCBF5–aNOP10 interaction. To test this hypothesis, Glutathione-Sepharose beads were coated with the GST–aGAR1 fusion protein. As shown by SDS–PAGE analysis (Figure 3), protein aCBF5, but not aNOP10 or L7Ae were retained on the beads, showing a direct interaction between aCBF5 and aGAR1. In addition, aNOP10 was not retained on the beads even when incubated together with aCBF5. Thus, the detected aCBF5–aGAR1 interaction is probably exclusive of the aCBF5–aNOP10 interaction. Interestingly, however, the negative effect of aGAR1 on RNP formation was partially counteracted in the presence of the RNA target. This is evidenced by the formation of a highly retarded complex CIII upon incubation of RNA-S with the WT

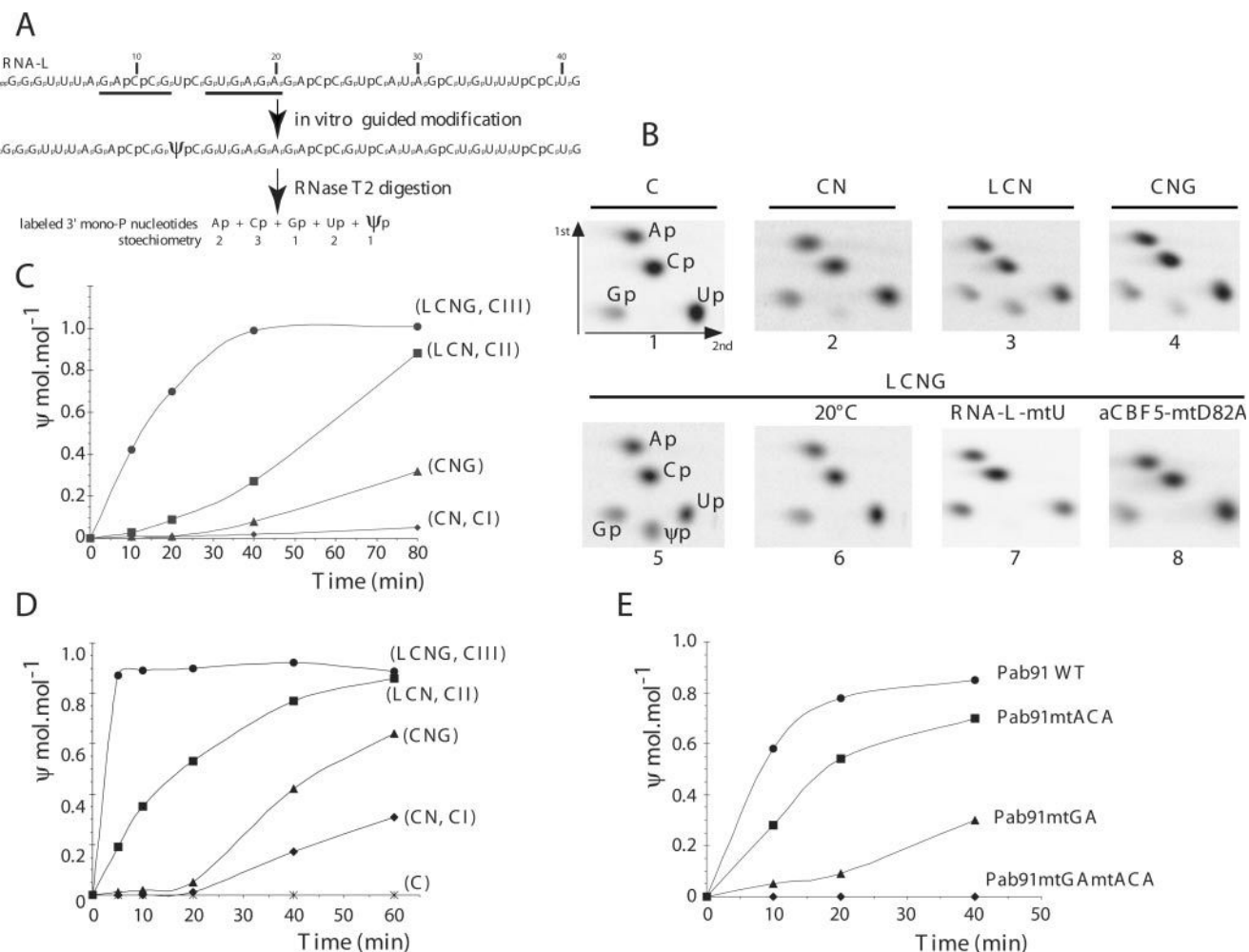


Figure 7. RNA-guided pseudouridylation of RNA target. (A) The 41 nt long RNA-L target used for analysis of RNP activity. Labeled phosphate after incorporation of [α -³²P]CTP are in bold. The sequence complementary to Pab91 sRNA is underlined. After T2 RNase digestion, residues U13p, U26p and U37p were labeled. Residue U13 is the expected pseudouridylation site (see also Figure 4A). (B) 2D-TLC analysis of Ψ formation. The RNA target (150 fmol) was mixed with 4 pmol of unlabeled WT sRNA and various combinations of the core proteins (200 nM each) and incubated for 80 min at 65°C (chromatograms 1–5 and 7–8), or 20°C (chromatogram 6). Incubation conditions are described in Materials and Methods. The following protein sets were used: aCBF5 alone (C), aCBF5 + aNOP10 (CN), L7Ae + aCBF5 + aNOP10 (LCN), aCBF5 + aNOP10 + aGAR1 (CNG), the complete set of core proteins (LCNG). In chromatogram 8, the variant aCBF5-mtD82A protein was used instead of the WT aCBF5 enzyme. After RNase T2 digestion and 2D TLC analysis, the Ap, Cp, Up, Gp and Ψ p spots indicated in (B), were identified by comparison with reference maps (47). (C–E) Time course of the *in vitro* pseudouridylation reaction using the RNA-L (Panels C and E) or RNA-S (Panel D) target. The unlabeled WT or variant Pab91 sRNA was incubated with the various protein combinations (C, CN, LCN, CNG and LCNG). The amount of Ψ residue formed was estimated by 2D TLC for RNA-L and 1D TLC for RNA-S (see Figure S1, Supplementary Material).

guide RNA and the four core proteins (Figure 6D). The CIII electrophoretic mobility was markedly lower than that of complex CII. Interestingly, complex CIII was not detected at 20°C.

The *in vitro* assembled complex CIII is active in pseudouridylation

To test for the RNA: Ψ -synthase activity of the reconstituted sRNP/RNA-S complexes, we used an approach based on the nearest-neighbor analysis (46). To get an idea of the specificity of the reaction, we used a 41 nt long RNA substrate (RNA-L) that carried a 3' extension compared to RNA-S. This RNA contained several uridine residues in addition to the one targeted by the Pab91 sRNA (Figure 7A). We verified that the CI, CII and CIII series of complexes was obtained with RNA-L

(data not shown). Then, a uniformly labeled RNA-L was produced by *in vitro* transcription in the presence of [α -³²P]CTP. After its digestion by RNase T2, three of the released 3'-monophosphate U residues were labeled, because they were followed by a C residue in the RNA molecule. One of them (U13) was the Pab91 sRNP target (Figure 6A). The labeled RNA-L was incubated 60 min with the WT Pab91 sRNA and the four core proteins in condition of complex CIII formation. Fractionation by thin layer chromatography of the nucleotides released upon T2 RNase digestion revealed a high yield of Ψ formation (Figure 7B). The experiment was repeated several times and for each experiment, the radioactivity of the fractionated 3'-monophosphate residues was measured with a phosphorimager. By taking into account the numbers of A, C, G and U residues followed by a C residue in RNA-L, we estimated the yield of U to Ψ conversion.

The mean value was found to be of 0.98 ± 0.05 mole of Ψ residue per mole of target RNA. In agreement with the idea that aCBF5 is the catalyst, no Ψ formation was detected when the aCBF5–D82A variant was used instead of the WT protein (Figure 7B). However, incubation of RNA-L with the aCBF5 protein alone was not sufficient for Ψ formation (Figure 7B). In agreement with the absence of complex CIII formation at 20°C (Figure 6D), Ψ formation only took place at an elevated temperature (Figure 7B). Pseudouridylation specifically occurred at the Pab91 targeted site, since after U to C substitution at this position no Ψ formation was detected (Figure 7B).

Some of the partially assembled H/ACA sRNPs also display specific RNA: Ψ -synthase activity

When the same test of activity was performed with the CI complex containing proteins aCBF5 and aNOP10, a very low level of Ψ formation was detected (0.08 ± 0.04 mol.mol⁻¹, mean value of three distinct experiments, Figure 7B and C). This low value was markedly increased (0.56 ± 0.08 mol.mol⁻¹), when the L7Ae protein was added to the incubation mixture (Figure 7B and C). Hence, the presence of protein L7Ae together with proteins aCBF5 and aNOP10 is sufficient for the formation of active particles. Whereas the activity of these particles was increased by addition of protein aGAR1, RNPs formed with the aCBF5, aNOP10 and aGAR1 protein combination were quite less efficient than RNPs assembled with the L7Ae, aCBF5 and aNOP10 protein set (mean value of 0.15 ± 0.07 mol.mol⁻¹, Figure 7B and C).

The short RNA-S substrate is more efficiently modified than the RNA-L substrate

We compared the kinetics of Ψ formation in both RNA-L and RNA-S (Figure 7C and D and Supplementary Material). In condition allowing complex CIII formation, one Ψ residue per mole of RNA-L was formed after 40 min of incubation. In the absence of aGAR1 (CII complex), a doubling of the incubation time was required to get the same level of modification (Figure 7C). In the presence of aGAR1 and the absence of L7Ae, only 30% of the RNA-L molecules were modified after 80 min of incubation and only trace amounts of modification were observed in the absence of both L7Ae and aGAR1 (CI complex). In contrast, when using the RNA-S substrate, a plateau of modification (90%) was reached after only 5 min of incubation in conditions of complex CIII formation. In the absence of protein aGAR1 (CII complex), the plateau was reached after 60 min of incubation. When the aCBF5–aNOP10–aGAR1 and aCBF5–aNOP10 protein combinations were used, 60 and 30% of the molecules were, respectively, modified after 60 min of incubation. Nevertheless, for both RNA substrates, a 20-min lag period was observed when protein L7Ae was not present in the reaction mixture. Hence, in the *in vitro* conditions that we used, both proteins L7Ae and aGAR1 increased the kinetics of the reaction.

Pseudouridylation by archaeal sRNP can occur in the absence of the ACA motif

As the Pab91mtGA and Pab91mtACA sRNA variants, respectively, formed CI-like and CII-like complexes with proteins L7Ae, aCBF5 and aNOP10 (Figure 4E), we tested

whether these RNAs could guide RNA-L pseudouridylation in the presence of the four core proteins. As shown in Figure 7E, whereas the double Pab91mtGAmtACA mutant had no RNA guiding capacity, the Pab91mtGA and Pab91mtACA variants both allowed RNA-L pseudouridylation. In addition, after a 40-min incubation, the pseudouridylation activity of the Pab91mtACA sRNP was ~80% compared to that of the WT sRNP. Thus, the ACA motif is not essential for H/ACA sRNA activity *in vitro*.

DISCUSSION

Here for the first time, an active H/ACA RNP is obtained by incubation of its purified components. This is also the first study on archaeal H/ACA sRNP assembly. Up to now, only the H/ACA protein L7Ae had been produced *in vitro* and was shown to bind H/ACA sRNAs (25). By this reconstitution approach, we were able to dissect the relative roles of the four core proteins in H/ACA sRNP formation and activity. Our results bring important information on H/ACA RNPs in general and on the specificity of the archaeal system, compared to the eukaryal system.

The L7Ae–aCBF5–aNOP10–sRNA complex assembly in Archaea

In Eukarya, based on its very broad RNA specificity, protein NHP2/Nhp2p (L7Ae counterpart) was proposed to get its ability to recognize snoRNAs specifically upon incorporation into a GAR1–NAP57/CBF5–NOP10–NHP2 protein complex (36). This complex is formed in the cytoplasm, and then protein–snoRNA recognition takes place in the nucleus. The absence of cell compartmentalization in Archaea may explain the need for a more specific interaction of protein L7Ae with the H/ACA sRNAs. As expected, the pseudo K-turn motif of sRNA Pab91 is required for L7Ae binding. Interestingly, the aCBF5 RNA: Ψ -synthase also has an RNA binding capacity that is dependent on the presence of the ACA motif (Figure 1). However, even at high aCBF5 concentration, only a fraction of the sRNA molecules was bound to the protein (Figure 1C). One possible explanation is that aCBF5 does not form a stable complex with the Pab91 sRNA and dissociate during the electrophoresis. Alternatively, a small number of the sRNA molecules may have the conformation required for aCBF5 binding, or the aCBF5 proteins cleaved from the GST moiety have a limited solubility. As the yield of aCBF5 association was doubled in the presence of protein L7Ae (Figure 2A), L7Ae may either favor aCBF5 recruitment by protein–protein interaction or fold the sRNA in a more favorable conformation. No direct binding of the eukaryal NAP57/CBF5 protein on H/ACA snoRNAs was described, up to now. This, in spite of the presence in both the archaeal and eukaryal protein of a PUA domain supposed to bind RNA, and also of the greater size of the eukaryal NAP57/CBF5 proteins compared to archaeal proteins (42). In contrast, the aCBF5 and NAP57/CBF5 proteins have the common property to interact with proteins aNOP10 (Figure 3) and NOP10/Nop10p (35,36), respectively. Association of aNOP10 with the sRNA is strictly dependent on the presence of aCBF5 (Figure 2A). Protein L7Ae alone does not recruit aNOP10, although association of the aCBF5–aNOP10 complex protein pair with the

sRNA is markedly reinforced in the presence of protein L7Ae (Figure 2B). Hence, as found for the C/D box sRNPs (18,22), protein L7Ae likely plays an important role in H/ACA sRNP assembly. It may either reinforce the sRNA folding or interact with the aCBF5–aNOP10 pair. In favor of the latter hypothesis, an RNP5 complex containing the three proteins was formed on the Pab91mtACA variant, which is unable to bind protein aCBF5 alone. Nevertheless, the very low yield of RNP5 formation with this variant RNA and the absence of stable L7Ae–aCBF5–aNOP10 trimer detection by GST pull-down experiments (Figure 3), is not in favor of the formation of an L7Ae–aCBF5–aNOP10 complex prior association with the sRNA, as found in Eukarya (35,36).

Interestingly, our data show that the aCBF5–aNOP10–sRNA complex and to a greater extent the L7Ae–aCBF5–aNOP10–sRNA complex are stabilized in the presence of the RNA substrate (Figure 4). This suggests a higher affinity of the aCBF5–aNOP10 heterodimer and L7Ae–aCBF5–aNOP10 heterotrimer for the RNA structure formed after base-pair interaction of the substrate RNA with the sRNA than for the free sRNA.

The L7Ae–aCBF5–aNOP10–sRNA complex is active in pseudouridylation

The small amount of RNP2 complex, formed by protein aCBF5 alone, had no detectable RNA:Ψ-synthase activity. The minimal set of proteins needed to detect activity is the aCBF5–aNOP10 pair (Figure 7), which is in agreement with the requirement of aNOP10 for efficient binding of the RNA substrate (Figure 4C). Protein aNOP10 has probably the capability to stabilize the guide RNA/target RNA interaction. The fact that the RNA substrate association strongly depends upon the presence of the targeted U residue (Figure 5) reinforces the idea of a needed RNA–protein interaction for stable association of the RNA substrate. It also indicates that the possibility to form the bipartite helices in the pseudouridylation pocket is not sufficient for a stable docking of the RNA substrate on the sRNP.

Although measurable, the activity observed with the two aCBF5 and aNOP10 proteins was low. The aCBF5–aNOP10–aGAR1 protein triplet was slightly more efficient (Figure 7). However, the L7Ae–aCBF5–aNOP10 protein triplet was quite more efficient. This reinforced efficiency may be linked to the positive effect of protein L7Ae on association of the aCBF5–aNOP10 pair. Accordingly, for both RNA-L and RNA-S substrates, a 20-min lag period without detectable RNA:Ψ-synthase activity was observed in the absence of protein L7Ae. This lag period may be needed to get an efficient aCBF5–aNOP10 recruitment in the absence of protein L7Ae. Accordingly, the same lag period is observed with the Pab91mtGA sRNA mutated in the L7Ae binding site.

aGAR1 increases the pseudouridylation efficiency of the Pab91 sRNP complexes

We are facing two puzzling observations. On the one hand, except for RNP1, the yield of RNP formation in the absence of RNA substrate is strongly reduced in the presence of protein aGAR1 (Figure 6). On the other hand, addition of protein aGAR1 increases the yield of substrate pseudouridylation at a given time point (Figure 7). Importantly also, in the presence

of the RNA substrate, the four proteins form a stable CIII complex (Figure 6E). This suggests that in the *in vitro* conditions used, the sRNP core protein complex is only stable in the presence of the RNA substrate. Furthermore, whereas aCBF5–aGAR1 and aCBF5–aNOP10 interactions were detected in our GST pull-down experiments, we did not detect a stable aCBF5–aNOP10–aGAR1 triple interaction. This may be due to the presence of the GST tag on the proteins. However, an exclusive interaction of the free aCBF5 protein with either aNOP10 or aGAR1 is in agreement with the data listed above. Triple interaction may be stable only in presence of both the sRNA and the target RNA. In light of the strong differences observed between the archaeal and eukaryal sRNP assembly, it should be pointed out that proteins L7Ae and Nhp2p/NHP2 have different behaviors. Whereas L7Ae binds strongly and specifically to the sRNA and may be the first protein bound to the sRNA, protein Nhp2p/NHP2 has poor and non specific RNA binding properties (48). It probably gains its specific RNA binding property in association with the other proteins in the aCBF5–aNOP10–NHP2 heterotrimer or in the aCBF5–aNOP10–aGAR1–NHP2 heterotetramer (36). Difference between the eukaryal and archaeal systems may essentially concern the sRNP and snoRNP assembly pathways. Observation by crosslink experiments of the location of protein GAR1 near the catalytic center is consistent with a possible interaction with the RNA substrate and with its possible involvement in the catalytic reaction. The enhancement of the yield of pseudouridylation that we observed in the presence of protein aGAR1 (Figure 7) suggests that this protein increases the catalytic efficiency of the particles: it may either favor a conformational transition leading to active sRNP particles or increase the kinetics of dissociation after the reaction. Further experiments are underway to test for these alternative explanations. Indeed, due to our utilization of an excess of RNP components compared to RNA substrate (molar excess of sRNA compared to RNA target of 26), we have no information on the turnover of the reaction. Nevertheless, this point will be important to address, since no ATP-dependent RNA helicase activity is used in the *in vitro* system, although the interaction formed by the guide Pab91 sRNA and its RNA substrate has a high free energy (–17.7 kcal/mol). Remarkably, as a C instead of a U at the pseudouridylation position had a strong negative effect on substrate assembly (Figure 5), destabilization of the target RNA may simply be due to Ψ formation.

In conclusion, we report here the successful development of an efficient *in vitro* pseudouridylation system that is based on the use of recombinant archaeal components. The advantage of this system compared to the eukaryal system is the great solubility of the archeal H/ACA core proteins. The assembly conditions developed will facilitate further studies on H/ACA sRNP structure and function. In addition, they represent a very useful tool for the *in vitro* generation of a Ψ residue at defined positions in RNAs.

Note

After we submitted this manuscript, Baker *et al.* reported in *Genes Dev.*, the reconstitution and activity of H/ACA sRNP particles by using the proteins and a modified Pf9 H/ACA guide sRNA from *Pyrococcus furiosus*.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Hannu Myllykallio (P. Forterre Lab) for his gift of *P. abyssi* genomic DNA. F. Leclerc is thanked for providing information on H/ACA sRNAs before publication. S. Massenet and Y. Motorin are thanked for their critical reading of the manuscript. The work was supported by the Centre National de la Recherche Scientifique (CNRS), the French Ministère de la Recherche et des Nouvelles Technologies (MRNT), the Actions Concertées Incitatives (ACI) Microbiologie and BCMS of the french MRNT and the Bioingénierie PRST of the Conseil Régional Lorrain. Funding to pay the Open Access publication charges for this article was provided by the Service Commun de Documentation de l' UHP-Nancy I.

Conflict of interest statement. None declared.

REFERENCES

- Decatur, W.A. and Fournier, M.J. (2003) RNA-guided nucleotide modification of ribosomal and other RNAs. *J. Biol. Chem.*, **278**, 695–698.
- Massenet, S., Mouglin, A. and Branlant, C. (1998) Posttranscriptional modifications in the U small nuclear RNAs. In Grosjean, H. and Benne, R. (eds), *The Modification and Editing of RNA*. ASM Press, New York, pp. 201–227.
- Bonnerot, C., Pintard, L. and Lutfalla, G. (2003) Functional redundancy of Spb1p and a snR52-dependent mechanism for the 2'-O-ribose methylation of a conserved rRNA position in yeast. *Mol. Cell*, **12**, 1309–1315.
- Donmez, G., Hartmuth, K. and Luhrmann, R. (2004) Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA*, **10**, 1925–1933.
- King, T.H., Liu, B., McCully, R.R. and Fournier, M.J. (2003) Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol. Cell*, **11**, 425–435.
- Lapeyre, B. and Purushothaman, S.K. (2004) Spb1p-directed formation of Gm2922 in the ribosome catalytic center occurs at a late processing stage. *Mol. Cell*, **16**, 663–669.
- Yang, C., McPheeters, D.S. and Yu, Y.T. (2005) psi 35 in the branch site recognition region of U2 snRNA is important for pre-mRNA splicing in *S. cerevisiae*. *J. Biol. Chem.*, **280**, 6655–6662.
- Ansmant, I. and Motorin, I. (2001) Identification of RNA modification enzymes using sequence homology. *Mol. Biol. (Mosk.)*, **35**, 248–267.
- Ofengand, J. (2002) Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett.*, **514**, 17–25.
- Omer, A.D., Ziesche, S., Decatur, W.A., Fournier, M.J. and Dennis, P.P. (2003) RNA-modifying machines in archaea. *Mol. Microbiol.*, **48**, 617–629.
- Bachelier, J.P., Cavaille, J. and Huttenhofer, A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
- Terns, M.P. and Terns, R.M. (2002) Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr.*, **10**, 17–39.
- Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
- Kiss-Laslo, Z., Henry, Y. and Kiss, T. (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.*, **17**, 797–807.
- Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R. and Dennis, P.P. (2000) Homologs of small nucleolar RNAs in Archaea. *Science*, **288**, 517–522.
- Ganot, P., Caizergues-Ferrer, M. and Kiss, T. (1997) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.*, **11**, 941–956.
- Ni, J., Tien, A.L. and Fournier, M.J. (1997) Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell*, **89**, 565–573.
- Omer, A.D., Ziesche, S., Ebhardt, H. and Dennis, P.P. (2002) *In vitro* reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc. Natl. Acad. Sci. USA*, **99**, 5289–5294.
- Tran, E., Brown, J. and Maxwell, E.S. (2004) Evolutionary origins of the RNA-guided nucleotide-modification complexes: from the primitive translation apparatus? *Trends Biochem. Sci.*, **29**, 343–350.
- Watkins, N.J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C. and Luhrmann, R. (2000) A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, **103**, 457–466.
- Bortolin, M.L., Bachelier, J.P. and Clouet-d'Orval, B. (2003) *In vitro* RNP assembly and methylation guide activity of an unusual box C/D RNA, cis-acting archaeal pre-tRNA(Trp). *Nucleic Acids Res.*, **31**, 6524–6535.
- Tran, E.J., Zhang, X. and Maxwell, E.S. (2003) Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J.*, **22**, 3930–3940.
- Rashid, R., Aittaleb, M., Chen, Q., Spiegel, K., Demeler, B. and Li, H. (2003) Structural requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles. *J. Mol. Biol.*, **333**, 295–306.
- Singh, S.K., Gurha, P., Tran, E.J., Maxwell, E.S. and Gupta, R. (2004) Sequential 2'-O-methylation of archaeal pre-tRNA^{Trp} nucleotides is guided by the intron-encoded but trans-acting Box C/D Ribonucleoprotein of Pre-tRNA. *J. Biol. Chem.*, **279**, 47661–47671.
- Rozhdestvensky, T.S., Tang, T.H., Tchirkova, I.V., Brosius, J., Bachelier, J.P. and Huttenhofer, A. (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.*, **31**, 869–877.
- Balakin, A.G., Smith, L. and Fournier, M.J. (1996) The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell*, **86**, 823–834.
- Bortolin, M.L., Ganot, P. and Kiss, T. (1999) Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *EMBO J.*, **18**, 457–469.
- Ganot, P., Bortolin, M.L. and Kiss, T. (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, **89**, 799–809.
- Heiss, N.S., Girod, A., Salowsky, R., Wiemann, S., Pepperkok, R. and Poustka, A. (1999) Dyskerin localizes to the nucleolus and its mislocalization is unlikely to play a role in the pathogenesis of dyskeratosis congenita. *Hum. Mol. Genet.*, **8**, 2515–2524.
- Lafontaine, D.L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M. and Tollervey, D. (1998) The box H ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.*, **12**, 527–537.
- Meier, U.T. and Blobel, G. (1994) NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria. *J. Cell Biol.*, **127**, 1505–1514.
- Watkins, N.J., Gottschalk, A., Neubauer, G., Kastner, B., Fabrizio, P., Mann, M. and Luhrmann, R. (1998) Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA*, **4**, 1549–1568.
- Henras, A., Henry, Y., Bousquet-Antonelli, C., Noaillac-Depeyre, J., Gelugne, J.P. and Caizergues-Ferrer, M. (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J.*, **17**, 7078–7090.
- Wang, C., Query, C.C. and Meier, U.T. (2002) Immunopurified small nucleolar ribonucleoprotein particles pseudouridylylate rRNA independently of their association with phosphorylated Nopp140. *Mol. Cell Biol.*, **22**, 8457–8466.
- Henras, A.K., Capeyrou, R., Henry, Y. and Caizergues-Ferrer, M. (2004) Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs. *RNA*, **10**, 1704–1712.
- Wang, C. and Meier, U.T. (2004) Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.*, **23**, 1857–1867.

37. Dragon, F., Pogacic, V. and Filipowicz, W. (2000) *In vitro* assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs. *Mol. Cell Biol.*, **20**, 3037–3048.
38. Pogacic, V., Dragon, F. and Filipowicz, W. (2000) Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. *Mol. Cell Biol.*, **20**, 9028–9040.
39. Tang, T.H., Bachellerie, J.P., Rozhdestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J. and Huttenhofer, A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA*, **99**, 7536–7541.
40. Hamma, T. and Ferre-D'Amare, A.R. (2004) Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. *Structure (Camb.)*, **12**, 893–903.
41. Charron, C., Manival, X., Clery, A., Senty-Segault, V., Charpentier, B., Marmier-Gourrier, N., Branlant, C. and Aubry, A. (2004) The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J. Mol. Biol.*, **342**, 757–773.
42. Watanabe, Y. and Gray, M.W. (2000) Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic Acids Res.*, **28**, 2342–2352.
43. Charron, C., Manival, X., Charpentier, B., Branlant, C. and Aubry, A. (2004) Purification, crystallization and preliminary X-ray diffraction data of L7Ae sRNP core protein from *Pyrococcus abyssi*. *Acta Crystallogr. D. Biol. Crystallogr.*, **60**, 122–124.
44. Marmier-Gourrier, N., Clery, A., Senty-Segault, V., Charpentier, B., Schlotter, F., Leclerc, F., Fournier, R. and Branlant, C. (2003) A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA*, **9**, 821–838.
45. Mougin, A., Gregoire, A., Banroques, J., Segault, V., Fournier, R., Brule, F., Chevrier-Miller, M. and Branlant, C. (1996) Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *RNA*, **2**, 1079–1093.
46. Motorin, Y., Keith, G., Simon, C., Foiret, D., Simos, G., Hurt, E. and Grosjean, H. (1998) The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA*, **4**, 856–869.
47. Keith, G. (1995) Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie*, **77**, 142–144.
48. Henras, A., Dez, C., Noaillac-Depeyre, J., Henry, Y. and Caizergues-Ferrer, M. (2001) Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. *Nucleic Acids Res.*, **29**, 2733–2746.

D. Résultats complémentaires

Le développement de la méthode d'assemblage et l'ensemble des données décrites dans l'article N°1, avaient été obtenus en utilisant un seul des sRNA H/ACA que nous avons mis en évidence. La question qui se posait était de savoir quel était le degré de généralité des résultats ainsi obtenus. Afin de répondre à cette question j'ai réalisé des études similaires à la fois sur un autre sRNA ne contenant qu'une seule structure tige-boucle (Pab21) et sur d'autres sRNA contenant plusieurs structures tige-boucles (Pab35, Pab105 et Pab40).

1. Mise en évidence d'une interaction directe de aNOP10 avec le sRNA H/ACA Pab21

Contrairement à ce qui a été observé pour Pab91, une étude par retard sur gel de complexes formés entre aNOP10 et le sRNA Pab21 nous a permis de mettre en évidence la formation d'une interaction directe entre ces 2 molécules (Fig. R.12). Le Kd apparent de la fixation de aNOP10 sur Pab21 estimé à partir de ces expériences de gel retard est de l'ordre de 80 nM.

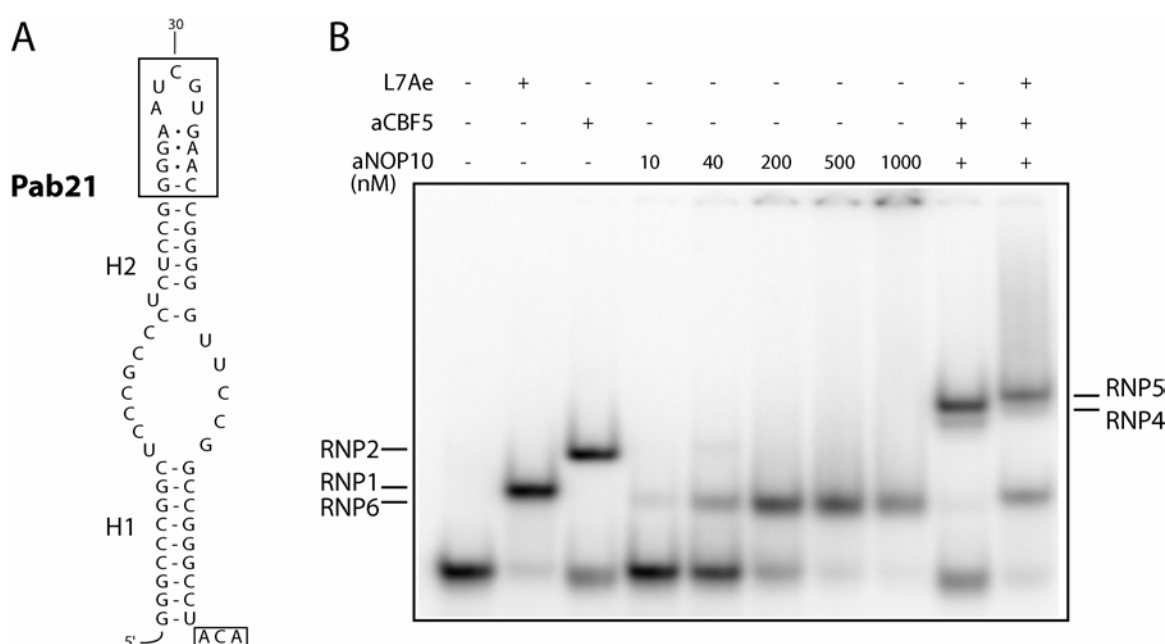


Figure R.12 : Etude par gel retard de l'assemblage de particules sRNP avec le sRNA Pab21 et démonstration de la fixation directe de aNOP10 sur le sRNA.

A. Structure secondaire proposée pour le sRNA Pab21. **B.** Résultats d'expériences de retard sur gel réalisées entre le sRNA Pab21 et les protéines L7Ae, aCBF5 et aNOP10, dans les conditions précisées dans la publication N°1. La concentration en protéines est de 200 nM sauf exceptions indiquées au dessus des pistes. La particule formée par l'interaction de aNOP10 sur le sRNA est appelée RNP6. Les noms des autres particules correspondent aux particules présentées dans la publication N°1.

2. L'affinité de aCBF5 pour le sRNA Pab21 est supérieure à celle pour le sRNA Pab91

Comme on peut le voir en comparant les résultats de la figure R.12 avec ceux de la publication N°1, l'affinité de aCBF5 pour le sRNA Pab21 est nettement supérieure à celle pour le sRNA Pab91. Or, Pab21 présente comme Pab91 une boîte ACA et une hélice H1 de 9 pb. La différence d'affinité de aCBF5 pouvait provenir de la différence de séquence des hélices H1 des deux sRNA, de l'identité des résidus de la poche interne ou de l'identité des nucléotides présents en 5' ou en 3' de la structure en tige-boucle. Pour répondre à cette question nous avons produit un ensemble de mutants des sRNA Pab91 et Pab21. Ainsi, l'hélice basale (H1) de Pab91 a été remplacée par celle de Pab21 et, inversement, l'hélice basale de Pab21 a été remplacée par celle de Pab91 (Fig. R.13). On observe que la substitution de l'hélice basale de Pab91 par celle de Pab21 améliore la fixation de aCBF5 et surtout la fixation du complexe aCBF5-aNOP10. Par contre, la substitution de l'hélice H1 de Pab21 par celle du sRNA Pab91 diminue un peu l'affinité de aCBF5 pour le sRNA Pab21. La recherche de déterminants de séquence ou de structure pouvant favoriser la fixation de aCBF5 est actuellement poursuivie par JB. Fourmann au laboratoire.

Afin d'essayer d'expliquer les résultats obtenus par l'échange des hélices des sRNA Pab21 et Pab91, des modèles de structure 3D des hélices H1 de sRNA Pab21 et Pab91 ont été construits par F. Leclerc (Fig. R.14). La séquence contenant la boîte ACA en 3' de l'hélice H1 est incluse dans ce modèle ainsi que deux résidus en 5' des hélices H1 des sRNA. Pour faciliter la modélisation, les hélices H1 ont été fermées par des boucles identiques dans les deux sRNA. On mesure une élévation moyenne par tour de paire de bases ("rise" ou "h-Rise") de 2,71Å pour l'hélice H1 de Pab21 et de 2,83Å pour celle de Pab91. L'élévation moyenne pour une hélice de forme A (ADN ou ARN) est comprise entre 2,56 et 3,03Å en fonction de la séquence. L'hélice de Pab91 est donc un peu plus étirée que l'hélice de Pab21, ce qui est sans doute le résultat d'une succession de purines. La distance qui correspond à une hauteur d'hélice d'environ 1 tour complet mesurée entre 2 résidus homologues dans chacune des hélices H1 est plus élevée pour le sRNA Pab91 (28.97Å) que pour Pab21 (26.85Å). La partie basale de l'hélice H1 de Pab91 est plus structurée et stable que celle de Pab21, ceci en raison de différences de séquence dans cette portion de l'hélice. Ainsi, une trop forte stabilité de la partie basale de l'hélice H1 associée à l'étirement de cette hélice dans le sRNA Pab91 pourrait donc être défavorable à la fixation de aCBF5 sur ce sRNA. L'affinité de aCBF5 pour d'autres sRNA identifiés et pour des sRNA pour lesquels d'autres mutations auraient été réalisées dans l'hélice H1 va être étudiée par JB. Fourmann.

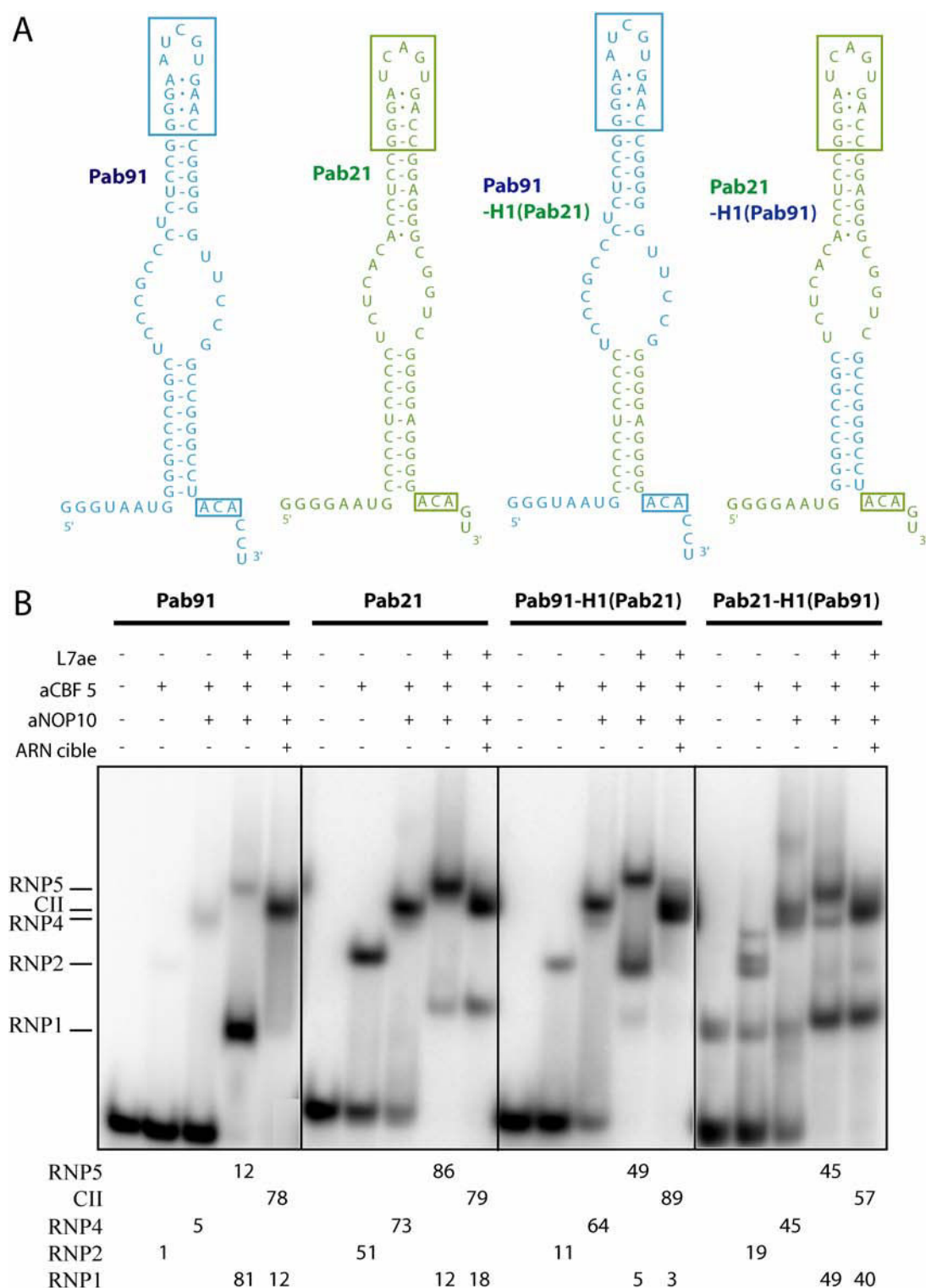


Figure R.13 : Etude par expériences de gel retard de déterminants augmentant le taux de fixation de aCBF5 sur les sRNA H/ACA. A. Structures secondaires proposées pour le sRNA Pab21, le sRNA Pab91 et les variants produits pour ces deux sRNA. **B.** Résultats de tests d'assemblage des protéines L7Ae, aCBF5, aNOP10 et de l'ARN cible sur les sRNA par des expériences de gel retard réalisées dans les mêmes conditions que celles qui sont décrites dans la publication N°1. Le pourcentage de l'ARN présent dans les complexes par rapport à la quantité d'ARN totale dans la piste est indiqué au-dessous.

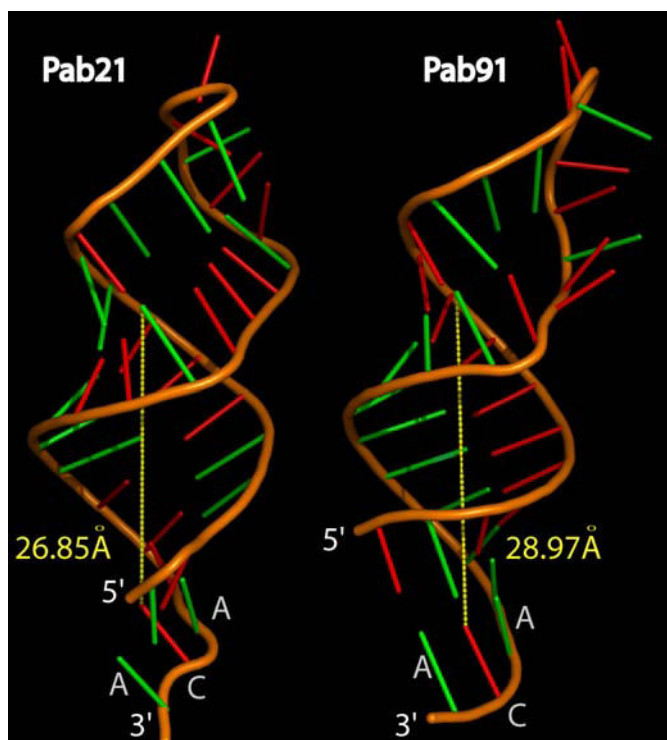


Figure R.14 : Modèles 3D des parties basales des sRNA H/ACA Pab21 et Pab91 incluant l'hélice H1 et la boîte ACA.

Les coordonnées de chacun des 2 modèles ont été obtenues après simulations par dynamique moléculaire et calcul des coordonnées moyennes (sur une trajectoire de 5 ns). Les chaînes phosphodiester sont représentées par un ruban cylindrique orange et les bases par des bâtonnets (en vert pour les purines et en rouge pour les pyrimidines). La hauteur de l'hélice basale (environ 1 tour d'hélice) des sRNA Pab21 et Pab91 est indiquée en angstroms.

3. Etude de l'assemblage des protéines sur des sRNA H/ACA à plusieurs motifs

On pouvait se demander si la fixation des 4 protéines des sRNP H/ACA était facilitée dans le cas où plusieurs motifs étaient présents dans l'ARN, ceci grâce à une fixation coopérative des protéines sur les différents motifs. Nous avons réalisé des expériences de retard sur gel en utilisant le sRNA Pab105 entier et différentes combinaisons de protéines recombinantes (Fig. R.15). Dans le cas de aCBF5, nous avons comparé les résultats obtenus pour le sRNA entier aux résultats obtenus en utilisant séparément chacune des structures tige-boucles. Dans le cas de l'ARN entier deux bandes retardées apparaissent en présence de aCBF5, la plus basse étant de plus forte intensité. Ainsi la bande de plus forte mobilité électrophorétique correspondrait à la fixation d'une seule molécule d'aCBF5 sur l'ARN et la bande de plus faible intensité et de plus faible mobilité électrophorétique correspondrait à la fixation de deux molécules de aCBF5 sur l'ARN. Ces données suggéraient l'existence d'un site ayant une plus faible affinité que l'autre, ce que nous avons confirmé en réalisant les expériences de retard sur gel avec chacune des 2 structures tige-boucles. Comme le montre la figure R.15, la structure tige-boucle 5' a une meilleure affinité que la structure tige-boucle 3'. Il est intéressant de noter que l'ajout de aNOP10 dans les cas du sRNA entier favorise la fixation de l'hétéro-duplex aCBF5-aNOP10 sur chacune des structures tige-boucles puisque dans ce cas la bande largement majoritaire correspond à celle de faible mobilité électrophorétique. On peut donc penser que aNOP10 qui, comme l'ont montré X. Manival et C. Charron au laboratoire (Manival *et al.*, 2006) peut dimériser, favorise la fixation du couple aCBF5-aNOP10 sur la structure tige-boucle 3'. Une

explication possible de la plus faible affinité de la structure tige-boucle 3' pour aCBF5 peut être la présence d'un trinucléotide AUA au lieu de ACA.

Il est intéressant de remarquer que les parties 5' et 3' indépendamment ne présentent pas une affinité pour aCBF5 à la hauteur de ce qui est observé pour le sRNA entier. Cette étude qui doit être approfondie par JB. Fourmann pourrait suggérer que l'assemblage des 4 protéines est facilité sur des sRNA comportant plusieurs motifs.

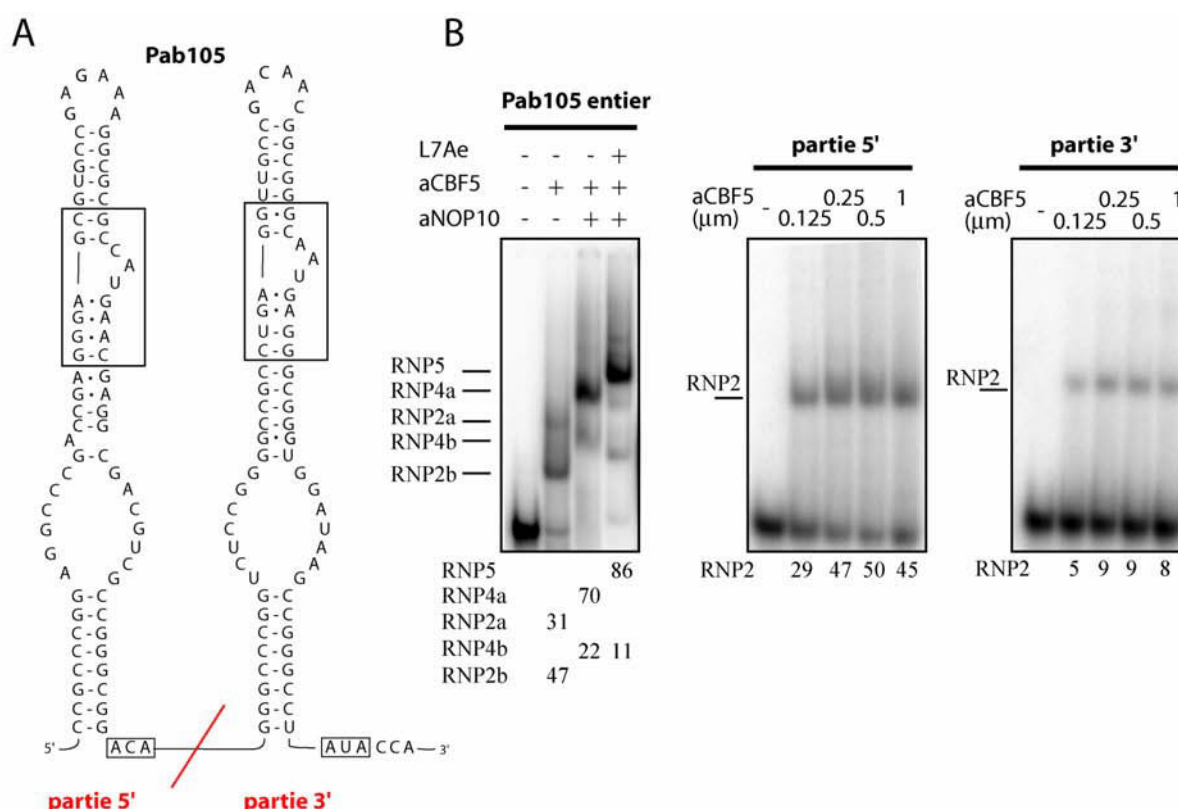


Figure R.15 : Etude par expériences de retard sur gel de l'assemblage de particules sRNP avec le sRNA Pab105 entier ou avec les parties 5' et 3'.

A. Structure secondaire proposée pour le sRNA Pab105. Le trait rouge représente la séparation entre les deux parties 5' et 3'. **B.** Résultats d'expériences retard sur gel réalisées dans les conditions décrites dans la publication N°1 avec des protéines à une concentration de 200 nM dans le premier panneau et des concentrations en protéine indiquées en haut des pistes dans les 2 autres panneaux. Le pourcentage de l'ARN présent dans les différents complexes par rapport à la quantité d'ARN totale dans la piste est indiqué au-dessous. Les noms des complexes correspondent aux noms présentés dans publication N°1 à la différence qu'un indice (a ou b) est utilisé pour distinguer deux formes correspondant à la fixation de 1 ou 2 molécules de aCBF5 ou aCBF5-aNOP10 par molécule de sRNA.

4. Cas particulier du sRNA Pab35 comportant une structure tige-boucle suppl  mentaire en 5'

Le r  le potentiel de la tige-boucle particuli  re qui est pr  sente en 5' du sRNA Pab35 est encore inconnu. Les exp  riences de Northern Blot que nous avons r  alis  es sur les ARN totaux de *P. abyssi* avec la sonde sp  cifique du sRNA Pab35 sugg  rent, d'apr  s la taille de l'unique ARN d  tect  , qu'une petite structure tige-boucle est effectivement pr  sente en 5' des deux structure tige-boucles contenant des boucles de pseudouridylaton.

De mani  re surprenante l'  l  ment 5' terminal du sRNA Pab35 comporte 2 bo  tes ACA, l'une dans la s  quence en simple-brin en 5', l'autre dans la boucle terminale. Cette observation soulevait la question d'un r  le   ventuel de cette structure tige-boucle dans la fixation de aCBF5    l'ARN. Pour r  pondre    cette question, nous avons compar   l'affinit   de aCBF5 pour l'ARN Pab35 entier et pour une forme tronqu  e o   la partie 5' comportant la structure tige-boucle suppl  mentaire   tait   limin  e (Fig. R.16). Nos r  sultats de retard sur gel montrent que la pr  sence de cette structure tige-boucle renforce la fixation de aCBF5    l'ARN. Il est    noter que les deux types de particules form   par association avec les 4 prot  ines sont actives *in vitro*.

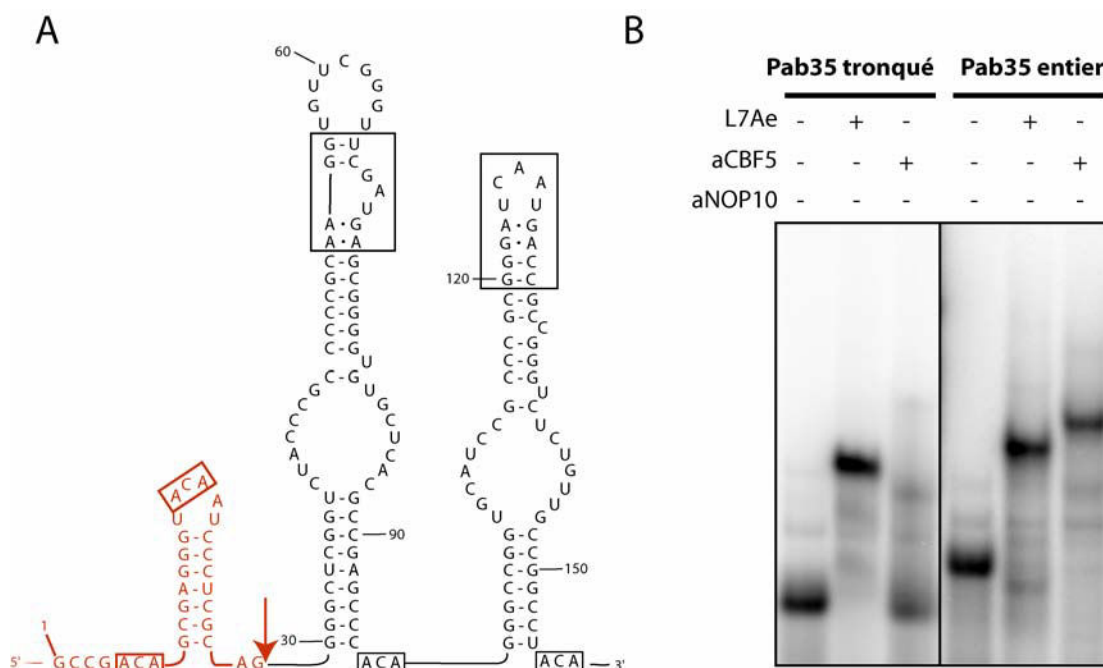


Figure R.16 : Etude par retard sur gel de l'assemblage de particules sRNP Pab35 comportant ou non une tige-boucle particuli  re en 5'.

A. Structure secondaire propos  e pour le sRNA Pab35. La partie rouge repr  sente la tige-boucle suppl  mentaire pr  sente en 5'. **B.** R  sultats d'exp  riences de retard sur gel r  alis  es dans les conditions pr  cis  es dans la publication N  1.

E. Discussion complémentaire & Perspectives

1. Trois protéines peuvent se fixer directement sur les sRNA H/ACA d'archaea : L7Ae, aCBF5 et aNOP10

- L7Ae

Comme observé par l'équipe d'A. Huttenhofer, nous observons bien que L7Ae se lie aux sRNA H/ACA, ceci par reconnaissance de motifs en K-turn ou du motif en K-loop présent dans la partie apicale des structures tige-boucles guidant la pseudouridylation. JB. Fourmann a observé que l'affinité de L7Ae pour les motifs en K-turn est supérieure à celle pour les motifs de type K-loop (4 nM contre 100 nM). Néanmoins, on observe autant de motifs de type K-loop que de motifs en K-turn dans les sRNA H/ACA de *P. abyssi*.

-aCBF5

L'affinité des sRNA H/ACA pour cette protéine est très variable suivant l'ARN. Elle peut être très faible, comme c'est le cas pour le sRNA Pab91, ou forte, comme c'est le cas pour le sRNA Pab21. En plus de l'identité du trinucléotide terminal, nous avons montré que la séquence de l'hélice H1 est également importante. Mais cet élément n'est pas encore suffisant à lui seul pour expliquer les différences marquées d'affinité que nous avons observées suivant les ARN. A ce propos, il faut noter qu'en parallèle à nos travaux de reconstitution d'une particule sRNP H/ACA active, l'équipe de Terns a aussi réussi la reconstitution (Baker *et al.*, 2005), ceci en utilisant un variant du sRNA Pf9 de *P. furiosus*. Ces auteurs avaient observé une forte affinité de ce sRNA pour aCBF5 et de ce fait ont essayé de définir par mutagenèse dirigée les déterminants de la forte affinité de cet ARN pour aCBF5. Ils avaient ainsi montré l'importance des résidus de la poche interne du sRNA pour la fixation de aCBF5.

Par ailleurs, un phénomène de coopérativité de fixation de 2 molécules d'aCBF5 sur des sRNA renfermant deux structures tige-boucle n'est pas exclu. Nous n'avons pas clairement distingué d'effet de coopérativité de fixation de aCBF5 sur le sRNA Pab105. Cependant, on remarque que la présence d'un deuxième motif améliore légèrement la fixation de aCBF5 par rapport à la somme des fixations de aCBF5 sur les motifs 5' et 3' isolés. La petite structure en tige-boucle supplémentaire située en 5' du sRNA Pab35 est précédée en amont par un trinucléotide ACA et contient dans sa boucle un autre trinucléotide ACA. Nos données suggèrent que cette tige-boucle supplémentaire pourrait renforcer la fixation de aCBF5 sur le sRNA Pab35. Il faut noter que la séquence de cette tige-boucle supplémentaire située en 5' est très conservée chez les *Pyrococcus*, ce qui renforce l'idée d'un rôle fonctionnel. Des mutations de chacun des trinucléotides ACA de cette région seront réalisées pour tester leur importance dans la fixation de aCBF5.

-aNOP10

Nos données révèlent une forte affinité de aNOP10 pour le sRNA Pab21. JB. Fourmann a aussi observé une interaction possible entre aNOP10 et le sRNA Pab105, bien que moins nette

que celle que nous avons observée pour Pab21. Par contre, aucune interaction directe n'a été observée entre aNOP10 et les autres sRNA H/ACA de *P. abyssi* caractérisés. Le seul point commun entre la structure tige-boucle unique de Pab21 et l'une des deux tige-boucles du sRNA Pab105, celle située en 5', est l'existence d'un résidu protubérant dans la partie basale de l'hélice H2. JB. Fourmann a montré que l'addition d'un résidu protubérant dans l'hélice H2 du sRNA Pab91 ne génère pas une affinité décelable pour aNOP10. Il restera à tester l'effet sur la fixation de aNOP10, de la délétion de ce résidu protubérant dans l'hélice H2 de Pab21. Une interaction de aNOP10 avec la partie basale de l'hélice H2 serait en accord avec les données structurales montrant trois résidus de aNOP10 interagir avec les groupement ribose et phosphate de 2 résidus de la partie 5' de l'hélice H2 du sRNA H/ACA (Fig. III.24) (Li et Ye, 2006).

On pourrait imaginer que l'affinité de aNOP10 pour le sRNA supplée une faible affinité de aCBF5 pour l'ARN, mais de manière surprenante, à la fois aCBF5 et aNOP10 ont une affinité élevée pour l'ARN.

2. aNOP10 joue un rôle primordial pour l'activité

Bien que aCBF5 présente les motifs caractéristiques des ARN:Ψ-synthases elle n'est pas active sur l'ARNr cible en absence de aNOP10. En effet, aNOP10 est requise pour qu'il y ait fixation du substrat. Les études structurales réalisées au laboratoire par X. Manival et C. Charron sur le complexe aCBF5-aNOP10 ont montré qu'il s'agissait d'un complexe très stable. Il est donc probable qu'*in cellulo* la fixation de aCBF5 à l'ARN guide se fasse sous la forme d'un complexe aCBF5-aNOP10. Les travaux de mutagenèse dirigée de aNOP10 et de aCBF5 réalisés au laboratoire après la mise au point des expériences de reconstitution ont permis de montrer que le domaine PUA de aCBF5 est requis pour la reconnaissance du sRNA en particulier du trinuécléotide ACA (Manival *et al.*, 2006). La protéine aNOP10 se replie en 2 domaines : le domaine N-ter forme une hélice qui interagit avec aCBF5 mais ne conduit pas à la formation d'une particule active, alors que le domaine C-ter forme un doigt à Zn qui ne se lie pas seul à aCBF5 mais peut le faire en présence d'ARN substrat, ce qui conduit à une activité réduite de moitié par rapport à l'utilisation de aNOP10 entière (Manival *et al.*, 2006). Ces données suggèrent que l'interaction entre aNOP10 et aCBF5 induit des transitions structurales dans aCBF5 favorisant son action sur l'ARN cible en présence de sRNA H/ACA.

3. L7Ae et aGAR1 augmentent l'efficacité de la réaction catalytique

La protéine aGAR1 interagit avec aCBF5 et la structure 3D de l'hétéro-trimère aCBF5-aNOP10-aGAR1 a été établie (Rashid *et al.*, 2006) (Fig.III.22). Aucun complexe entre L7Ae et les autres protéines des sRNP n'avait été isolé. D'après la structure 3D établie pour une sRNP H/ACA entière (Li et Ye, 2006), L7Ae interagit uniquement avec aNOP10. Ces interactions ne reposeraient que sur 7 et 6 résidus de L7Ae et aNOP10 respectivement et ne seraient pas assez stables pour qu'un complexe soit formé en absence de sRNA. On sait que les protéines de la famille L7Ae/L30 favorisent la structuration de l'ARN lors de leur fixation, en particulier la formation de la structure en K-turn (Marmier-Gourrier *et al.*, 2003 ; Clery *et al.*, 2007 ; Youssef

et al., 2007). On peut donc imaginer que L7Ae, en se fixant sur le sRNA H/ACA, favorise la formation des structures tige-boucles. Cette action structurale est sans doute tr s importante pour les souches poussant   des temp ratures tr s  lev es. En accord avec cette hypoth se, le taux de complexes CI et CII obtenu en pr sence du sRNA Pab91 et de son ARN cible est nettement plus  lev  lorsque L7Ae est pr sente en plus du complexe aNOP10-aCBF5 (Fig. 4C de la publication N 1). Il est probable  galement que des interactions L7Ae-aNOP10 expliquent l'augmentation marqu e de la cin tique de la r action en pr sence de L7Ae (Fig. 7 de la publication N 1). Une  tude r cente de l' quipe de M. Terns renforce l'id e d'un r le de L7Ae sur la structuration de l'ARN (Youssef *et al.*, 2007). Elle montre en effet, que la fixation de L7Ae sur le sRNA H/ACA n'induit pas seulement des changements conformationnels au niveau du K-turn, mais aussi au niveau de l'h lice H2, de la boucle apicale et de la poche de pseudouridylation. Plus pr cis ment, la fixation de L7Ae sur le K-turn favoriserait la d limitation de la poche de pseudouridylation en stabilisant la formation de l'h lice H2, ce qui favoriserait la fixation du substrat et de aCBF5 (Fig. R.17) (Youssef *et al.*, 2007).

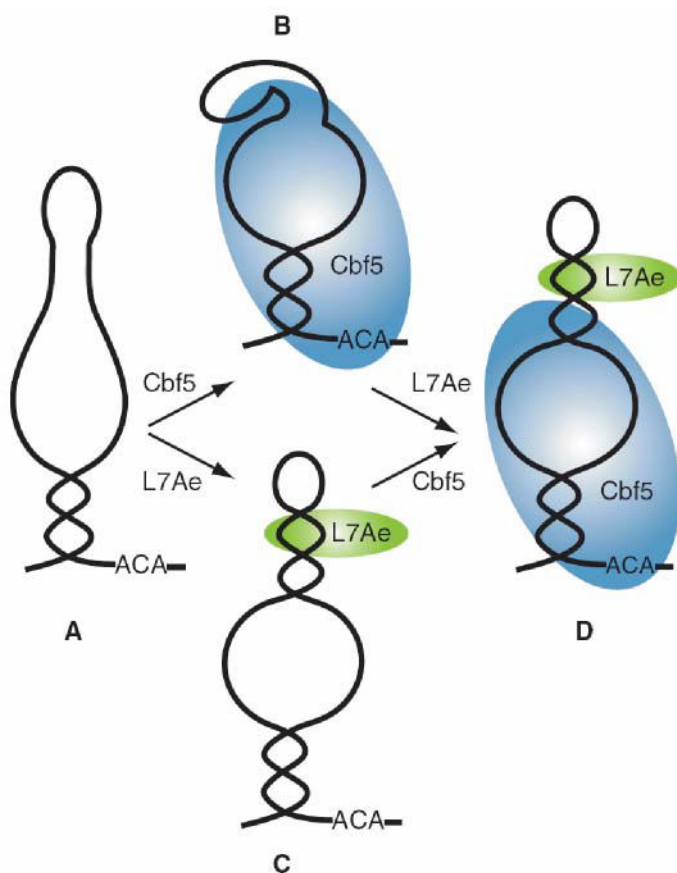


Figure R.17 : Mod le sch matique repr sentant les modifications conformationnelles d'un sRNA H/ACA induites par la fixation de L7Ae (d'apr s Youssef *et al.*, 2007).

Les donn es provenant d'empreintes sugg rent que l'h lice H2 n'est pas repl e de fa on stable en absence de L7Ae (A et B). La prot ine aCBF5 peut interagir avec la bo te ACA, l'h lice H1, la poche de pseudouridylation, la base de l'h lice H2 ainsi qu'avec, en absence de L7Ae, la partie 5' du motif en K-turn (B). La fixation de L7Ae induit le repliement correct de l'h lice H2 et permet donc la d finition de la poche de pseudouridylation (C et D). L7Ae est donc en comp tition avec aCBF5 pour sa fixation   la partie 5' du motif en K-turn.

L'importance des interactions mises en place entre L7Ae et aNOP10 est démontrée par les travaux récents de JB. Fourmann et B. Charpentier. Les données montrent en effet, que la mutation de résidus de L7Ae ou de aNOP10 impliqués dans l'interaction entre ces deux protéines conduit à une perte de l'effet positif de L7Ae sur la cinétique de la pseudouridylation.

Le rôle de aGAR1 est encore discuté. Il est envisageable que aGAR1, dont la présence accélère significativement la vitesse de la réaction, soit impliqué dans le turnover de la réaction. Comme nous l'avons expliqué dans l'introduction, aGAR1 interagit fortement avec la région $\beta 7$ de aCBF5. Or, cette boucle pourrait interagir avec l'hélice intermoléculaire formée entre l'élément antisens 3' du sRNA et la séquence en 5' de l'uridine cible dans l'ARNr (Fig III.28 ; III.23) (Li et Ye, 2006 ; Rashid *et al.*, 2006). Si la boucle $\beta 7$ de aCBF5 positionnée dans une conformation dite « fermée » stabilise l'appariement entre le sRNA et sa cible, aGAR1 pourrait, en interagissant avec cette boucle, induire une conformation « ouverte » et ainsi favoriser la dissociation de l'ARN cible et du sRNA. Ce système pourrait ainsi contrôler l'état d'avancement de la réaction. Ainsi, lorsque aucun ARN cible n'est apparié dans la poche interne du sRNA H/ACA, la fixation de aGAR1 sur aCBF5 pourrait provoquer la dissociation de aCBF5/aNOP10 du sRNA expliquant les résultats présentés dans la figure 6 de la publication N°1. Lorsqu'un ARN substrat est apparié dans la poche de pseudouridylation, aGAR1 pourrait favoriser son appariement ou provoquer la dissociation de l'ARN cible et du sRNA comme nous l'avons déjà évoqué ci-dessus. Dans ce dernier cas, il est également possible que la modification du résidu U en Ψ intervienne dans la libération de l'ARN cible puisque l'identité du résidu à cette position influence l'affinité de l'ARN substrat comme le montre le changement du résidu U en C dans la figure de la publication N°1. Par ailleurs, des données récentes de JB. Fourmann et B. Charpentier montrent une faible affinité d'un substrat pseudouridylé pour la sRNP. Enfin, un autre rôle possible de aGAR1 pourrait être lié à son interaction avec aCBF5 qui pourrait modifier la conformation de certains résidus du site actif.

Ces différentes hypothèses doivent être vérifiées.

4. Développement d'une méthode permettant de rechercher des sRNA à boîtes H/ACA dans tous les génomes d'Archaea

A. Principe

Sur la base des données obtenues à partir des sRNA H/ACA que nous avons caractérisés et des données publiées par d'autres laboratoires, il était alors possible d'envisager le développement d'une méthode de recherche informatique dédiée aux gènes de sRNA H/ACA au sein des séquences génomiques d'archaea. Cette méthode a été conçue de façon à ne pas avoir à extraire les séquences interORF, comme nous le proposons pour la recherche de gènes de tout type d'ARNnc, en particulier des ARNnc aux caractéristiques non connues. J'ai donc entrepris ce nouveau développement informatique sous la responsabilité de F. Leclerc. Afin de réaliser une recherche spécifique des sRNA H/ACA, nous avons choisi de commencer par sélectionner au sein des génomes, des séquences pouvant correspondre à des ARNnc ayant les caractéristiques de sRNA H/ACA en utilisant le programme ERPIN. Le programme ERPIN permet l'identification de motifs d'ARN en construisant un profile de structure secondaire d'un type d'ARNnc à partir d'un alignement de séquences (Gautheret *et al.*, 2001). Les séquences des sRNA H/ACA identifiés ont ainsi été alignées selon les contraintes de structuration secondaire. Cette première étape nous a conduit à la description de sRNA H/ACA candidats. Dans un second temps, des cibles possibles des sRNA H/ACA potentiels, sélectionnés à partir de la première étape, sont recherchées en utilisant le logiciel RNAMOT comme décrit précédemment (Gautheret *et al.*, 1993).

Cette approche, ainsi que son application à la recherche de sRNA dans les génomes de *M. jannaschii* et *T. kodakarensis* ont été publiées dans un chapitre du volume 425 de Methods in Enzymology consacré à la recherche et à l'étude de la génération des modifications post-transcriptionnelles dans les ARN.

Chapitre d'ouvrage N°2 :

A Dedicated Computational Approach for the Identification of Archaeal H/ACA sRNAs.
Sébastien MULLER, Bruno CHARPENTIER, Christiane BRANLANT and Fabrice LECLERC
Methods in Enzymology, 2007 volume 425:21-53.

CHAPITRE D'OUVRAGE N°2

A DEDICATED COMPUTATIONAL APPROACH FOR THE IDENTIFICATION OF ARCHAEOAL H/ACA sRNAs

Sébastien Muller, Bruno Charpentier, Christiane Branlant, and Fabrice Leclerc

Contents

1. Introduction	356
2. Method	358
2.1. Search for H/ACA-like motifs	359
2.2. Search for targets of the H/ACA-like motifs	374
2.3. Phylogenetic and experimental validation of the results	382
3. Conclusions	383
References	384

Abstract

Whereas dedicated computational approaches have been developed for the search of C/D sRNAs and snoRNAs, as yet no dedicated computational approach has been developed for the search of archaeal H/ACA sRNAs. Here we describe a computational approach allowing a fast and selective identification of H/ACA sRNAs in archaeal genomes. It is easy to use, even for biologists having no special expertise in computational biology. This approach is a stepwise knowledge-based approach, combining the search for common structural features of H/ACA motifs and the search for their putative target sequences. The first step is based on the ERPIN software. It depends on the establishment of a secondary structure-based “profile.” We explain how this profile is built and how to use ERPIN to optimize the search for H/ACA motifs. Several examples of applications are given to illustrate how powerful the method is, its limits, and how the results can be evaluated. Then, the possible target rRNA sequences corresponding to the identified H/ACA motifs are searched by use of a descriptor-based method (RNAMOT). The principles and the practical aspects of this method are also explained, and several examples are given here as well to help users in the interpretation of the results.

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, Nancy Université, Faculté des Sciences et Techniques, Vandoeuvre-les-Nancy, France

Methods in Enzymology, Volume 425
ISSN 0076-6879, DOI: 10.1016/S0076-6879(07)25015-3

© 2007 Elsevier Inc.
All rights reserved.

1. INTRODUCTION

Pseudouridine (Ψ) is one of the most abundant posttranscriptionally modified nucleotides found in tRNAs, rRNAs, and UsnRNAs (Rozenski *et al.*, 1999). Uridine to pseudouridine conversion can be catalyzed by a single protein with RNA recognition capacity and RNA/ Ψ -synthase activity. It can also be catalyzed by H/ACA RiboNucleoProtein complexes (H/ACA sRNPs) containing a guide RNA with H/ACA boxes, called snoRNAs and scaRNAs in Eukarya and sRNAs in Archaea. These RNAs are associated with a set of proteins, Nhp2p, Nop10p, Cbf5, and Gar1p in yeast, NHP2, NOP10, Dyskerin (or NAP57), and GAR1 in human (Henras *et al.*, 1998; Khanna *et al.*, 2006; Lafontaine *et al.*, 1998; Meier and Blobel, 1994; Wang and Meier, 2004; Watkins *et al.*, 1998), and their homologs L7Ae, aNOP10, aCBF5, and aGAR1 in Archaea (Baker *et al.*, 2005; Charpentier *et al.*, 2005; Rozhdestvensky *et al.*, 2003). Proteins aCBF5 and Cbf5/Dyskerin belong to the TruB family of RNA/ Ψ -synthases (Hamma *et al.*, 2005; Lafontaine *et al.*, 1998; Manival *et al.*, 2006; Meier and Blobel, 1994; Rashid *et al.*, 2006; Wang and Meier, 2004).

The guiding properties of H/ACA RNAs are based on the formation of a complex structure: two sequences from the RNA substrate, that are separated by a 5'-UN-3' dinucleotide, base pair with the two strands of an internal loop of the guide RNA (pseudouridylation pocket). The U residue in the 5'-UN-3' dinucleotide is the target site of the reaction (Balakin *et al.*, 1996; Ganot *et al.*, 1997a,b; Ni *et al.*, 1997).

Several studies combining computational and experimental approaches have been dedicated to the identification of eukaryal H/ACA snoRNAs in human, mouse, *D. melanogaster*, *S. cerevisiae*, and *S. pombe* (Huang *et al.*, 2004, 2005; Kiss *et al.*, 2004; Li *et al.*, 2005; Schattner *et al.*, 2004, 2006; Torchet *et al.*, 2005; Yuan *et al.*, 2003). Most of these RNAs were found to have a two stem-loop structure. Most generally, each of the two stem loops contains a pseudouridylation pocket. These stem loops are linked by a single-stranded sequence containing the ANANNA H box and the 3' stem loop is flanked by a single-stranded element containing an ANA trinucleotide. Whereas the global architecture of eukaryal snoRNAs is highly conserved, the sizes and base compositions of their stem-loop structures are highly variable.

Knowledge regarding archaeal H/ACA sRNAs is more limited, because they have only been characterized in *Archaeoglobus fulgidus* (Tang *et al.*, 2002a), *Pyrococcus* species (Baker *et al.*, 2005; Charpentier *et al.*, 2005; Rozhdestvensky *et al.*, 2003), and *Methanocaldococcus jannaschii* (Thebault *et al.*, 2006). Identified Ψ residues in archaeal rRNAs are also very limited: 4, 6, and 3 Ψ residues were detected in the 23S rRNA of *Halobacterium*

halobium, *Sulfolobus acidocaldarius*, and *Haloarcula marismortui*, respectively (Del Campo *et al.*, 2005; Massenet *et al.*, 1999; Ofengand and Bakin, 1997). However, the available data reveal some specific features of archaeal H/ACA sRNAs compared with their eukaryal counterparts. Indeed, although the target U residue is identified by the same base-pairing rules between the guide RNA and the targeted sequence, the global architecture of H/ACA sRNAs is more variable. One, two, or three contiguous stem-loop structures containing a pseudouridylation pocket can be present. They may even contain an additional stem-loop structure without a pseudouridylation pocket (see Pf6, Rozhdestvensky *et al.*, 2003). In addition, compared with their eukaryal counterparts, they are structurally more constrained, with a strong conservation of the relative positions, sizes, and base compositions of their structural elements. In each stem-loop structure, the stems delineating the pseudouridylation pocket are highly enriched in G-C and C-G pairs. Each stem loop also displays a K-turn motif or a K-loop motif (Charpentier *et al.*, 2005; Charron *et al.*, 2004; Hamma *et al.*, 2005; Klein *et al.*, 2001; Li and Ye, 2006; Rozhdestvensky *et al.*, 2003; Vidovic *et al.*, 2000) in its apical part. K-turn and K-loop motifs are characterized by the presence of an A•G and G•A sheared pair flanked by a U residue. Each stem-loop structure is tailed by an ANA box that is more frequently an ACA triplet. By inspection of the secondary structure of the identified H/ACA sRNAs, we observed that the K-turn or K-loop motif on the one hand and the ACA box on the other hand are located at a well-defined distance from the targeted U residue in the pseudouridylation pocket.

Therefore, although not unique, the characteristic features of the H/ACA sRNA stem-loop structures, flanked by their associated ANA box, are specific enough to represent a signature that can be used for H/ACA sRNA gene identification by computational search in sequence databases. For simplification, we will designate these modular elements as H/ACA motifs. On the basis of the peculiar structure formed by the H/ACA sRNA and the RNA substrate, the finding by computational analysis of putative target sequences can also be used as a complementary source of information to discriminate true H/ACA sRNA genes from false-positive DNA sequences.

Taking advantage of the specificity of archaeal H/ACA sRNAs, we developed a knowledge-based approach that combines the consecutive searches for sequences coding for H/ACA motifs in complete sequences of archaeal genomes and their complementary RNA targets in rRNA sequences. The identification of H/ACA motifs is performed by use of a profile-based approach, taking advantage of the present knowledge on H/ACA sRNAs. The knowledge used includes data from comparative sRNA sequence analysis (Baker *et al.*, 2005; Charpentier *et al.*, 2005; Rozhdestvensky *et al.*, 2003; Tang *et al.*, 2002a; Thebault *et al.*, 2006), *in vitro* reconstitution of active H/ACA sRNPs by use of various WT and mutated H/ACA sRNAs and various RNA substrates (Charpentier *et al.*, 2005; S. Muller *et al.*, unpublished

results) and the recent determination of the 3D structure of a complete H/ACA sRNP particle (Li and Ye, 2006). Then, the identification of possible RNA targets associated with the identified H/ACA motifs is performed by use of a descriptor-based approach, taking into account the rules of complementarities that were also inferred from the data from *in vitro* reconstitution (S. Muller *et al.*, unpublished data). Because it is a knowledge-based approach, each unknown H/ACA motif associated with an identified target can, after an appropriate validation, contribute to increase and refine our knowledge on the H/ACA structural features and their target recognition. Thus, the general performance of the approach may be improved.

This chapter is dedicated to biologists who have no special expertise in computational methods. Our goal is to teach how to use the proposed strategy to get an H/ACA sRNA gene identification that will be as extensive as possible with a minimized number of false-positive results. The use of a computational approach for the prediction of H/ACA sRNA genes in Archaea and the prediction of their target sites are as important as *in vitro* reconstitution assays that are available to validate the results (see Chapter 16 by Charpentier *et al.*). The combined use of computational predictions and *in vitro* tests should increase considerably our knowledge on archaeal sRNAs in the near future.

2. METHOD

Our strategy for the detection of H/ACA sRNAs follows a stepwise and iterative procedure in which the first step is the search for H/ACA-like motifs through archaeal genomes, and the second step is the determination of the associated RNA target(s) in rRNAs (Fig. 15.1). In the first step (Fig. 15.1, step 1), H/ACA-like motifs are detected by use of the profile-based ERPIN program (Gautheret and Lambert, 2001). This program has been applied to the search of a wide range of RNA motifs (Lambert *et al.*, 2002, 2004; Legendre *et al.*, 2005). Once H/ACA-like motifs are identified, their putative target(s) in rRNAs are searched (Fig. 15.1, step 2) by use of the descriptor-based RNAMOT program. This program has also been extensively applied to the identification of several kinds of RNA motifs, and the data obtained have been experimentally validated (Bourdeau *et al.*, 1999; Laferriere *et al.*, 1994; Lescure *et al.*, 1999, 2002). In the present strategy, the use of RNAMOT is atypical, because this program is applied to the search of RNA motifs formed by partial base pairing of the internal loops of the H/ACA motifs with ribosomal RNAs. Because of the peculiar application of both ERPIN and RNAMOT in the present strategy, we had to create an appropriate H/ACA profile for ERPIN and several RNA target descriptors for RNAMOT. We will describe the rationale used for their generation

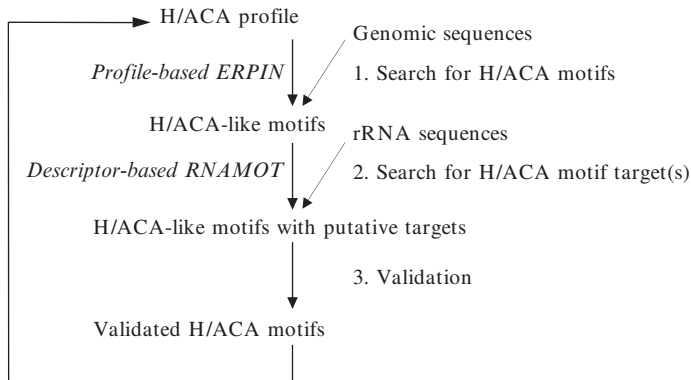


Figure 15.1 The multistep strategy proposed for the identification of H/ACA motifs. A description of the three steps leading to the identification of new H/ACA sRNA candidates: first an ERPIN-based step is used for the search of H/ACA-like motifs, then, a descriptor-based step is used for the search of their targets. Each of them includes tests of the validity of the results and a final global validation step is performed, taking into account the present knowledge on H/ACA sRNAs and rRNA pseudouridylation. The sequences of the new validated H/ACA motifs can be integrated in the H/ACA profile, and the search can be repeated on the same genome or run on other genomes.

and how to use them to achieve maximal efficiency and specificity of the delivered data.

By this procedure, H/ACA-like motifs that exhibit at least one putative RNA target are identified. They are then subjected to an evaluation step (Fig. 15.1, step 3). This evaluation step is based on: (1) current knowledge on Ψ positions in archaeal rRNA, (2) comparison with already identified archaeal H/ACA sRNAs and, (3) experimental data obtained from structure–function analysis of H/ACA sRNA by use of the *in vitro* reconstituted system. Obviously, the final proof of the validity of the prediction requires experimental tests of the proposed guiding property by use of the *in vitro* assembly procedure (see Chapter 16 by Charpentier *et al.*) and the identification of a Ψ residue at the targeted position in the rRNA by the CMCT approach (see Chapter 2). When the demonstration is completed, the newly identified H/ACA motifs can be used to enrich the ERPIN profile.

2.1. Search for H/ACA-like motifs

The search for H/ACA-like motifs requires the establishment of a profile, as implemented in the ERPIN program. This H/ACA profile is based on our present understanding of the common structural features of archaeal H/ACA motifs. Once the profile is built, the search can be started. The basic ERPIN command, parameters, and options and the present stage of their optimization

are explained by use of various examples of searches of H/ACA motifs in archaeal genomes. Through the proposed interpretations of the data obtained and by giving some tips, we intend to help users speed up their search, analyze and evaluate the results, and improve the initial profile.

2.1.1. Requirements

2.1.1.1. Hardware and software The ERPIN program used in this approach runs and has been tested under most UNIX platforms (for more details about compatible operating systems, see references). The ERPIN program is available as a stand-alone version that can be obtained from D. Gautheret (Gautheret and Lambert, 2001) and as a web server (<http://tagc.univ-mrs.fr/erpin/>) (Lambert *et al.*, 2004). The results presented here were obtained with the most recent release of ERPIN (version 5.5).

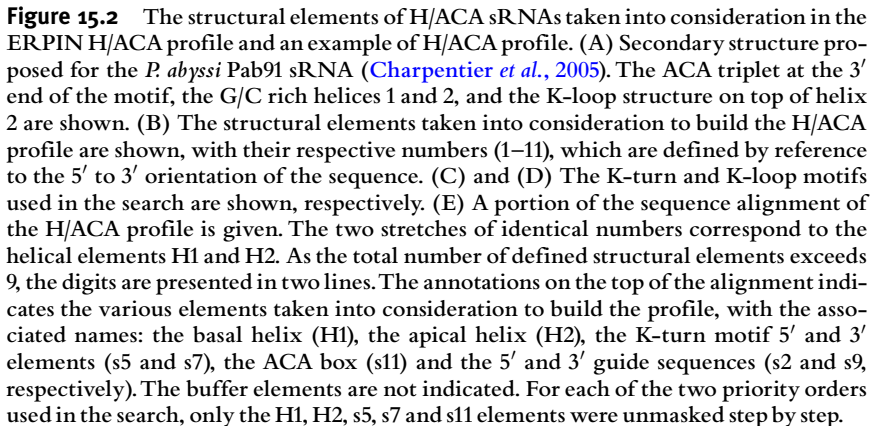
2.1.1.2. Sequence data The sequence format used for ERPIN applications is FASTA. Already published archaeal genomic sequences can be downloaded from the NCBI ftp site: <ftp://ftp.ncbi.nih.gov/genomes/Bacteria>. The sequences of the 26 already identified H/ACA sRNAs sequences were used to establish the H/ACA profile (Baker *et al.*, 2005; Charpentier *et al.*, 2005; Rozhdestvensky *et al.*, 2003; Tang *et al.*, 2002a). This profile is available at <http://tagc.univ-mrs.fr/asterix/erpin/>.

2.1.2. Establishment of the ERPIN profile

The H/ACA profile is established on the basis of a sequence alignment of all the H/ACA motifs of the already identified H/ACA sRNAs (Baker *et al.*, 2005; Charpentier *et al.*, 2005; Rozhdestvensky *et al.*, 2003; Tang *et al.*, 2002a). It is based on and includes secondary structure information, and its correct establishment is critical for the sensitivity and specificity of the ERPIN search.

To build this profile, the sequences aligned are considered as a succession of single-stranded (sx) and double-stranded (Hx) elements, x is a number defined according to the position of the element in the H/ACA sequence (5' to 3'). Each nucleotide in a defined structural element carries the number attributed to this element (Fig. 15.2B). These numbers correspond to the first lines of the profile (Fig. 15.2E). Segments known to be single-stranded, like the ACA triplet or the two guide sequences, or segments that are single-stranded or double-stranded, depending on the considered H/ACA motif, are each defined by a unique stretch of a given number. In contrast, obligatory helices are identified by the presence of two stretches of residues that are all assigned to the same number and have identical lengths. When two stretches with an identical numbering are defined in the profile, the RNA sequences that do not contain the corresponding base-pair interaction are not selected in the search.

Constitution of the profile is complicated by the fact that, despite a strong conservation, H/ACA motifs show some structural variations. For



instance, the sizes of helices 1 and 2 slightly vary from one motif to the other, and a limited number of bulge residues or mismatches can be present, in particular in helix 2. An extreme case of variation is exemplified in the second motif of the *P. furiosus* Pf7 sRNA, where the 5' strand of the pseudouridylation pocket is separated from helix 1 by an additional stem-loop structure (see Pf7, Rozhdestvensky *et al.*, 2003). Finally, the presence of either an apical K-loop or an apical K-turn motif with a terminal stem loop increases the diversity of the H/ACA motifs. For an optimized prediction of H/ACA motifs, we had to take this variability into consideration. This was of high importance, because ERPIN does not select sequences that contain bulge residues and internal loops in the double-stranded regions identified by the H/ACA profile. To overcome this program limitation, two base-paired elements, H1 and H2, were defined, corresponding to the estimated minimal regular succession of base pairs present in helices 1 and 2 (7 bps for H1 and 5 bps for H2, respectively). Furthermore, to accommodate the size variations of helices 1 and 2, segments that can be either single stranded or double stranded were included in the profile ("buffer" segments). A maximal size was defined for each "buffer" segment, by taking into account constant distances in H/ACA motifs (distances between the ACA trinucleotide and the pseudouridylation pocket and between the K-turn/K-loop motif and this pocket, respectively), as well as the overall lengths of known H/ACA motifs.

To be able to select RNA with either a K-loop or a K-turn motif, we defined these elements in the H/ACA profile by their common structural features: namely, two A•G and G•A sheared pairs with sequence constraints at the position 5' to the GA sequence in the 5' strand and at the two successive positions 5' to the GA sequence in its 3' strand (Fig. 15.2). Therefore, the two possible motifs are defined by the presence of a BGA (where B is any nucleotide except A) and RUGA elements that are separated by a loop. The maximal size of this loop was fixed to 37nts.

The 41 H/ACA motifs of the 26 identified H/ACA sRNAs were manually aligned, taking into account the following structural elements (Fig. 15.2B and E): a basal helical element H1 (No. 1), the 5' and 3' strands of the pseudouridylation pocket (Nos. 2 and 9, maximum lengths 11 and 12nts, respectively), an upper helical element H2 (No. 3), one K-turn or K-loop motif defined by a BGA element in the 5' strand, a RUGA element in the 3' strand and an apical loop (Nos. 5, 7, and 6, respectively), the ANA triplet (No. 11) and 3 "buffer" elements (Nos. 4, 8, and 10) that can be single or double stranded and are used for flexibility. Only three of the conserved structural elements have a sequence imposed by the presence of conserved residues in all the aligned RNA sequences of the profile, namely, the 5' and 3' elements of the K-turn/K-loop motif and the ANA box (Nos. 5, 7, and 11, respectively). The "buffer" elements numbered 4, 8, and 10 can be considered as gaps, which can be filled or not. The maximal size of elements 4, 8, and 10 were fixed to 3, 4, and 2nts, respectively.

The preceding description of the rationale for generation of the ERPIN profile will help the user of this approach to introduce new H/ACA sequences in the profile and will also facilitate the interpretation of the data obtained by use of this profile.

2.1.3. Practical procedure

The ERPIN searches based on the defined profile are performed by use of a local version of the program or the ERPIN web server version accessible at <http://tagc.univ-mrs.fr/asterix/erpin/>. Note that the current version of the ERPIN server does not allow the modifications of the profile and optional parameters. However, this server access provides a 2D structure representation of the selected putative H/ACA motif.

First, one needs to specify the genome sequences and the H/ACA profile to be used. To this end, the “compulsory” commands <profile> and <genome> allow the user to enter the names and locations of the ERPIN profile and sequence database, respectively. By default, the search is performed on both strands of the genome. Then, various options are available to finely tune the ERPIN search.

When activating the “no mask” option, the genomic screening will involve the simultaneous search of all the structural elements defined in the profile. We do not recommend this possibility; because of the great number of elements defined in the H/ACA profile, the computer search will be very slow. To increase the performance of the search (both speed and efficiency) screening for the presence of some selected structural elements can be done with a priority order. This order is defined by the use of successive masks. The first mask restricts the search to one or more selected element(s) (the unmasked elements, command “unmask”). The masked elements will influence the search by their delimited sizes. The sequences of the unmasked elements of the H/ACA motifs, which are aligned in the profile, are compared with the inspected genomic sequences. Then, the remaining selected masked elements are unmasked step by step (“-add” option). The first step is fast, because only a few elements are searched through the entire genome. In the second step, the added elements are only searched within the sequence portions of the genome identified in the first step and so on in the following steps.

In this priority order strategy, a cutoff value can be specified to determine how stringent the search is at each of the steps defined by the successive masks (i.e., how similar the elements have to be in comparison with those included in the profile to be identified as hits). Different cutoff values can be defined for the different steps of the selection, depending on the relative importance of the elements specified in the mask. The definition of the masks, their order of use, and the cutoff values used at each step have an impact on the results.

To explain how these cutoff values are defined, we have to introduce the notion of lod-scores and scores. By comparison of one residue (single-stranded

element) or a pair of residues (double-stranded elements) in a given element of one aligned H/ACA motif in the profile with the corresponding residues, or pairs of residues, in the other aligned motifs, ERPIN can establish a score of similarity (lod-score) for individual residues or pairs of residues. On the basis of the sum of the lod-scores of all the residues or pairs of residues in a given element, ERPIN defines a score for this element. By extension, when a mask is used, a score can be established for the overall unmasked regions of each of the H/ACA motifs aligned in the profile. The cutoff value for the step of the search where this mask is used will be defined by reference to the score values established for all the motifs aligned in the profile.

The cutoff values can be defined as absolute values or as percentages. We will use percentages. They indicate the minimal percentage of sequences in the profile that are captured as hits on the basis of their calculated scores. When a 100% or higher percentage is used as cutoff (100% is the default cutoff value if no specification is given), all the sequences aligned in the profile will be captured as hits. Lower percentages indicate that not all the sequences in the profile are selected, thus, the search in the analyzed genome will be highly stringent. On the other hand, a higher cutoff value (>100%) indicates that sequences with some divergence relative to the aligned motifs of the profile can be selected. Note that the possibility of defining cutoff values >100% is available only for releases of ERPIN >5.3.

The two highly discriminating structural elements in the H/ACA motifs are the helical elements H1 and H2 and the K-turn/K-loop structure, respectively. Searches can be done by giving prevalence to one or the other of these two elements. In the following applications, we illustrate the relative efficiencies of two ordered series of three successive masks. In order 1, preference is given to the selection of the helical elements H1 and H2. Indeed, the presence of these two elements H1 (No. 1) and H2 (No. 3) is first tested, then the 5' and 3' strands of the K-turn/K-loop (Nos. 5 and 7) are searched, and finally, the presence of a putative ACA motif (No. 11) is investigated. When the second series of masks is used, designated as order 2, the first two steps are inverted, so that priority is given to the two strands of the K-turn/K-loop in the search. In these two procedures, the elements 2, 4, 6, 8, 9, and 10, namely the buffer elements, the apical loop and the two strands of the internal loop, are always masked.

2.1.4. Illustrative tests performed on archaeal genomes whose H/ACA sRNAs have been characterized

We will first present various tests run on genomes for which H/ACA sRNAs were characterized (Baker *et al.*, 2005; Charpentier *et al.*, 2005; Rozhdestvensky *et al.*, 2003; Tang *et al.*, 2002a; Thebault *et al.*, 2006). These tests illustrate the influence of the priority order and the cutoff values on the detection of H/ACA-like motifs. They also show how the ERPIN

parameters are optimized, taking into consideration the present state of the profile. In the blind tests presented in Table 15.1 (columns denoted blind tests), no prior knowledge on the H/ACA sRNAs identified in the analyzed species or in the genus, in the case of the *Pyrococcus* species, was included. To this end, the *A. fulgidus* genome was subjected to an ERPIN search by use of a profile in which the five identified *A. fulgidus* H/ACA motifs were eliminated. Similarly, the *M. jannaschii* genome was subjected to an ERPIN search by use of a profile in which the six identified *M. jannaschii* H/ACA motifs were eliminated. Finally, for the searches performed on each of the *Pyrococcus* genomes, the 30 motifs identified in *P. abyssi*, *P. furiosus*, and *P. horikoshii* were eliminated from the profile. In parallel, we evaluated the effect of the presence in the H/ACA profile of motifs belonging to three species of the same genus (*Pyrococcus*) on searches performed on the *A. fulgidus* and *M. jannaschii* genomes. To this end, we performed blind tests on the *A. fulgidus* and *M. jannaschii* genomes by use of an H/ACA profile containing motifs from only one of the *Pyrococcus* species (*P. abyssi*) (Table 15.1B). For each of these blind tests, the two priority orders (1 and 2) were applied, as well as three different cutoff values (95%, 100%, and 110%). The putative H/ACA motifs selected in the assay are next compared with the already identified H/ACA motifs. “Positive results” in Table 15.1 correspond to motifs already identified, whereas the other ones are denoted “likely false-positive results.”

As illustrated in Table 15.1A, priority order 1 always allows the identification of the larger number of validated H/ACA motifs. The detection of 60–90% of the known H/ACA motifs with a cutoff value of 110% demonstrates the performance of the approach. The number of likely false-positive results increases with this cutoff value. However, as will be explained later, the inspection of the motifs and their location in the genome allows an efficient discrimination of likely false-positive results.

Some interesting observations can be made by inspection of Table 15.A and B: the presence of motifs from three species of the *Pyrococcus* genus in the profile did not bias the search. On the contrary, it has a marked positive effect on the selection of true H/ACA motifs in both *A. fulgidus* and *M. jannaschii*. The presence of all the H/ACA motifs from *Pyrococcus* is even required for an efficient selection of the H/ACA motifs from *A. fulgidus*.

For training in data interpretation and to demonstrate that secondary structure analysis of the candidate motifs helps discriminate the false-positive results, we provide the potential secondary structures of the candidate motifs selected for *M. jannaschii* in the blind tests presented in Table 15.1A (Fig. 15.3), and we comment on these results. As evidenced in Fig. 15.3 (panel B1), all of the four selected motifs corresponding to validated H/ACA motifs (denoted FW1, FW3, RC1, and RC3 in the search compilation) have an ACA triplet and a canonical K-turn structure, whereas the two likely false-positive sequences (FW2 and RC2 motifs, respectively) contain a CCA and an ACG motif instead of the ACA triplet (panel B2). Neither of them

Table 15.1A Tests of the effects of priority order, cutoff values and the use of a second step of selection in ERPIN searches

	Number of H/ACA sRNA motifs already identified	Cutoff	Blind Tests				Second Step Tests			
			Order (1)		Order (2)		Order (1)		Order (2)	
			Number of positive results	Number of likely false-positive results	Number of positive results	Number of likely false-positive results	Number of positive results	Number of likely false-positive results	Number of positive results	Number of likely false-positive results
<i>A. fulgidus</i>	5 ⁽¹⁾	95%	1	0	1	0	1	0	1	0
		100%	2	1	2	1	2	0	2	0
		110%	3	3	3	3	4	3	3	2
<i>P. abyssi</i>	10 ^(2, 3, 4)	95%	4	0	2	0	9	1	5	0
		100%	4	0	2	0	9	1	5	0
		110%	9	1	4	1	9	2	5	2
<i>P. furiosus</i>	10 ^(2, 3, 4)	95%	5	0	3	0	9	0	7	0
		100%	5	0	3	0	9	0	7	0
		110%	7	1	4	0	9	1	7	0
<i>P. horikoshii</i>	10 ^(2, 3, 4)	95%	3	0	2	0	9	0	6	0
		100%	3	0	2	0	9	0	6	0
		110%	6	2	4	0	9	1	6	0
<i>M. jannaschii</i>	6 ⁽⁵⁾	95%	3	1	3	0	3	1	3	0
		100%	3	1	3	0	3	1	3	0
		110%	4	2	4	1	4	1	5	0

¹ Tang *et al.*, 2002.² Rozhdestvensky *et al.*, 2003.³ Charpentier *et al.*, 2005.⁴ Baker *et al.*, 2005.⁵ Thébault *et al.*, 2006.

Table 15.1B Tests of the effects of priority order, cutoff values and the use of a second step of selection in ERPIN searches

	Number of H/ACA sRNA motifs already identified	Cutoff	Blind Tests			
			Order (1)		Order (2)	
			Number of positive results	Number of likely false-positive results	Number of positive results	Number of likely false-positive results
<i>A. fulgidus</i>	5 ⁽¹⁾	95%	1	0	1	0
		100%	1	0	1	0
		110%	1	1	1	1
<i>M. jannaschii</i>	6 ⁽⁵⁾	95%	3	1	3	0
		100%	3	1	3	0
		110%	3	1	3	1

(A) Blind tests performed on each of the archaeal genomes whose H/ACA sRNAs have been studied (*A. fulgidus*, *P. abyssi*, *P. furiosus*, *P. horikoshii* and *M. jannaschii*). In the first step of these blind tests, for each of the studied genomes, the known H/ACA motifs of this species or of species of the same genus were removed for the H/ACA profile used for the ERPIN search. Searches were run with the two orders of priority and with three different cutoff values (95, 100 and 110%). In a given test, identical cutoff values were used for each of the masks. The name of the studied genome is given in the first column. The number of H/ACA motifs already identified in this species is indicated in the second column. The numbers of true H/ACA motifs (positives results) and likely false positive H/ACA motifs found in the ERPIN search are given for each assay. Results obtained with each of the priority orders are shown in parallel. In the second step of these blind tests, the true H/ACA motifs that were found in the first step were included in the H/ACA profiles. The tests use the same parameters as in the first step.

(B) Similar blind tests were performed on the *A. fulgidus* and *M. jannaschii* genomes with a profile lacking the *P. furiosus* and *P. horikoshii* H/ACA motifs.

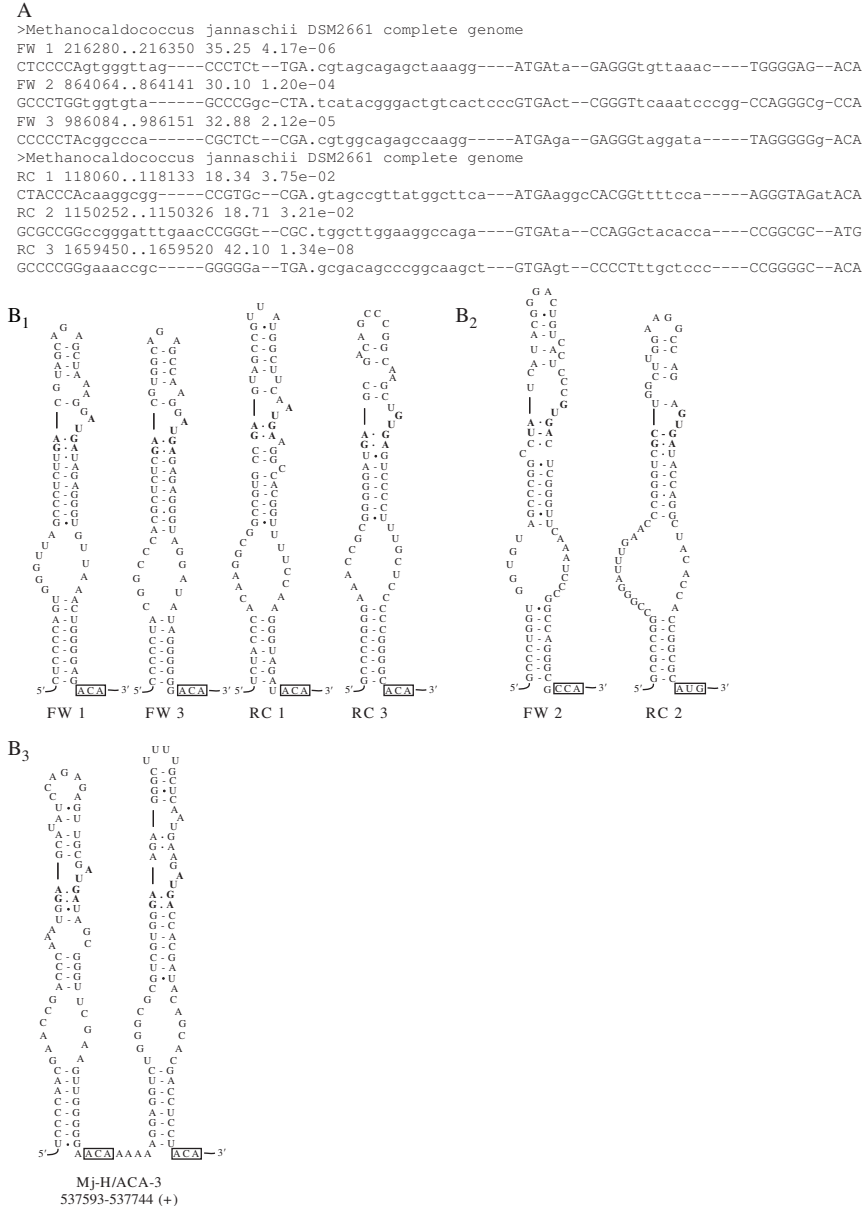


Figure 15.3 Results of an ERPIN search run on the *M. jannaschii* genomic sequence and the secondary structures proposed for the identified candidates. (A) The results of the ERPIN program, as displayed in output, are shown. The ERPIN program was run on the *M. jannaschii* genome sequence, using as a profile an alignment of the known H/ACA motifs, except for the ones identified in *M. jannaschii* (Thebault *et al.*, 2006) (see Table 15.4). The names of the candidates are defined by reference to the DNA strand screened and the order of appearance. ERPIN gives the positions relative to the forward

contain the canonical A•G and G•A sheared pairs tandem. In addition, they correspond to tRNA sequences, tRNA^{Asp}(GUC) and tRNA^{Gly}(UCC), respectively. Therefore, these likely false-positive H/ACA-like motifs can be unambiguously discarded from the ERPIN results. Note that most tRNA sequences can be folded into a stem-loop structure, explaining their detection with the profile that we used. The undetected motifs correspond to two motifs present in a unique sRNA (Mj-H/ACA-3, [Thebault et al., 2006](#)). The 5' motif has probably been discarded because of the internal loop present in helix 2. The second one probably differs by the sequences of helices 1 and 2 compared with the H/ACA motifs of the alignment.

Interestingly, when the four validated H/ACA motifs found in the blind test (FW1, FW3, RC1, and RC3) are included in the profile, the 3'-terminal motif of RNA Mj-H/ACA-3 is found in the search by use of the priority order 2 and a cutoff value of 110% ([Table 15.1A](#)). Hence, the use of both priority orders may be interesting in some specific cases.

More generally, a stepwise investigation including a second run after the inclusion in the profile of the validated H/ACA motifs selected in a first run, strongly improves the number of selected motifs and decreases the number of likely false-positive results ([Table 15.1A](#)). When the priority order 1 is used, only one of the previously identified H/ACA motifs is not found in the three *Pyrococcus* species, the one proposed to guide U to Ψ conversion at position 2575 in the *P. furiosus* LSU rRNA (see Pf3, stem I, [Rozhdestvensky et al., 2003](#)).

2.1.5. Example of the search of H/ACA-like motifs in a yet unexplored genome

The genome from *Thermococcus kodakarensis*, a species belonging to the same order as species of the *Pyrococcus* genus, was used in this teaching example. The profile used contained the 41 known archaeal H/ACA motifs. The search was first performed with priority order 1, and the cutoff value defined for each step of the selection was 110%. Hence, the following ERPIN command was used:

```
erpin<profile><genome>-1,11-umask 1 3-add 5 7-add 11-cutoff 110%110%110%
```

“1,11” indicates the beginning and the end of the chain of structural elements considered in the profile. In this case, all the elements of the profile are considered.

strand for each selected H/ACA-like motif. (B) A secondary structure is proposed for each of the candidates. The motifs in Panel B1 were retained after inspection of their structures, whereas those in panel B2 were discarded. Panel B3 represents the proposed secondary structures of the two known H/ACA motifs that were not detected in the first step of the blind search.

The search time was short: less than 3 min of CPU time on a 3.2-GHz P IV. Sixteen candidate motifs were identified. Six of them were encoded by the plus strand and 10 by the minus strand. At the last step of the selection, ERPIN computes an e-value reflecting the statistical significance of the selected H/ACA-like motifs. This e-value represents the probability of encountering this motif at random (Lambert *et al.*, 2005). Hence, it depends on the scores calculated for all the structural elements and on the size of the database. The output file displays the candidate sequences that satisfy the cutoff values defined at each selection step and the e-values of these sequences (Fig. 15.4). Only candidates with a negative e-value deserve further analysis.

In the search performed on the *Thermococcus kodakarensis* genome, nine candidates had negative e-values between 1e-06 and 4e-02, which was an indication for a good fit with the H/ACA motifs in the profile. In contrast, seven candidates had e-values between 1e-01 and 4, suggesting that they may correspond to false-positive results. The possible secondary structure of each candidate is shown in Fig. 15.4 (panels B1 and B2). The nine candidates with satisfying e-values displayed all the expected structural elements (panel B1). Only the FW3 candidate has an AUA motif instead of the ACA motif. Despite its high e-value, the FW1 motif has a canonical K-loop, an ACA motif, and helices 1 and 2 with correct lengths. The high e-value observed may be due to the presence of two G·U pairs in helix 1. Furthermore, motifs FW1, FW2 and FW3 may belong to the same H/ACA sRNA,

A

```
>Thermococcus-kodakarensis
FW 1 462059..462114 21.90 7.56e-01
TGCCCCTCgcgaggga-----CCCCG--GGGgaga-----ATGAac--CGGGGggcgatgc---GGGGGCA--ACA
FW 2 462115..462195 33.44 1.32e-04
GCCCGGcctcagcaggtccCGCG--GGAgggccttcgcgcgtcccgagcagc--ATGAc--CGCGGgaaccagc---GCCGGGC--ACA
FW 3 462201..462272 27.43 3.49e-02
GGCCCGCcgagggttag-----CCCCGc--CAAgggtggcggtacgccttcg-----ATGAgg--GCGGGagttaccg---GCGGGCC--ATA
FW 4 899083..899151 32.54 3.74e-04
GCCCCCGggaaccgc-----GGGGGg--TGAgggcctcggcgagcc-----GTGAtt--CCCTTcgctccc---CGGGGGC--ACA
FW 5 1545183..1545239 35.55 7.44e-06
GCCCTCCcctctcacac-----CCCCG--GGAgcg-----GTGAac--CGGGGggcggtcg---GGAGGGC--ACA
FW 6 1769611..1769691 19.95 1.64e+00
CGCCCGGgtggtgta-----GCCCGgc--CCAtcatcagggactgtcactccc-----GTGAac--CGGGTtcaaatccggcCGGGGCGccATA
>Thermococcus-kodakarensis
RC 1 47849..47907 35.19 1.28e-05
GGCCTGGcgtccgcctt-----CCCCGg--GGAaac-----GTGAac--CGGGGgttcctg---CCAGGCCt--ACA
RC 2 465056..465126 34.46 3.57e-05
GGGCCCGcgtctccgga-----CCGCT--GGAgggggctctctgcccgtca-----ATGAg--AGCGGgggacagc---CGGGCCC--ACA
RC 3 465127..465197 33.20 1.77e-04
CCCGCCGgtggcccggt-----GTCCC--GGAgggggccgagggccaca-----GTGAa--GGGATgagacgatgc---CGGGCGG--ACA
RC 4 865615..865690 19.53 1.91e+00
GGGAGGAtcatgac-----TCCGT--CAActccagcctgacgaccttctaagtATGAttc--TCGGGttaaagc---TGCTCCC--AGA
RC 5 986951..987023 19.50 1.93e+00
AGCCACGctggaggttcag--CCAGGctCGGAaaagagttc-----GTGAagacCGGGGaaatgcccg---CGAGGCTttACA
RC 6 1107274..1107331 34.45 3.64e-05
GGGCCGGgagcatccac-----CGCGG--GGAgga-----GTGAac--CGCGGgcctctgtac---CCGGCCC--ACA
RC 7 1107332..1107400 30.26 3.64e-03
GGGCTCGgtacaaccgc---CTCCG--CAAggtatcggtgttc-----GTGAg--CGGAGcgtgctcacgc--CGAGCCC--ACA
RC 8 1371594..1371659 18.26 2.94e+00
GTCTCTTacctgaaatg---GGCGT--CAAggtgttcatt-----GTGAg--AAAGTtagaagcct---GGAGGATg--AAA
RC 9 1404046..1404129 21.71 8.22e-01
CGGCGCtattcctgcgc---TGCGGa--TGAtacctgaggtatagctgagagaccgGTGAggc--CAACGgtaagacc---GGCCCGCa--ATA
RC 10 1460837..1460911 17.72 3.52e+00
TCGAGGgtctgaaagt-----CCAGAA--TGAacggaagacagt-----ATGAcagtCTTGttaaagtttaaaCCCTCGGcgAGA
```

Figure 15.4 (continued)

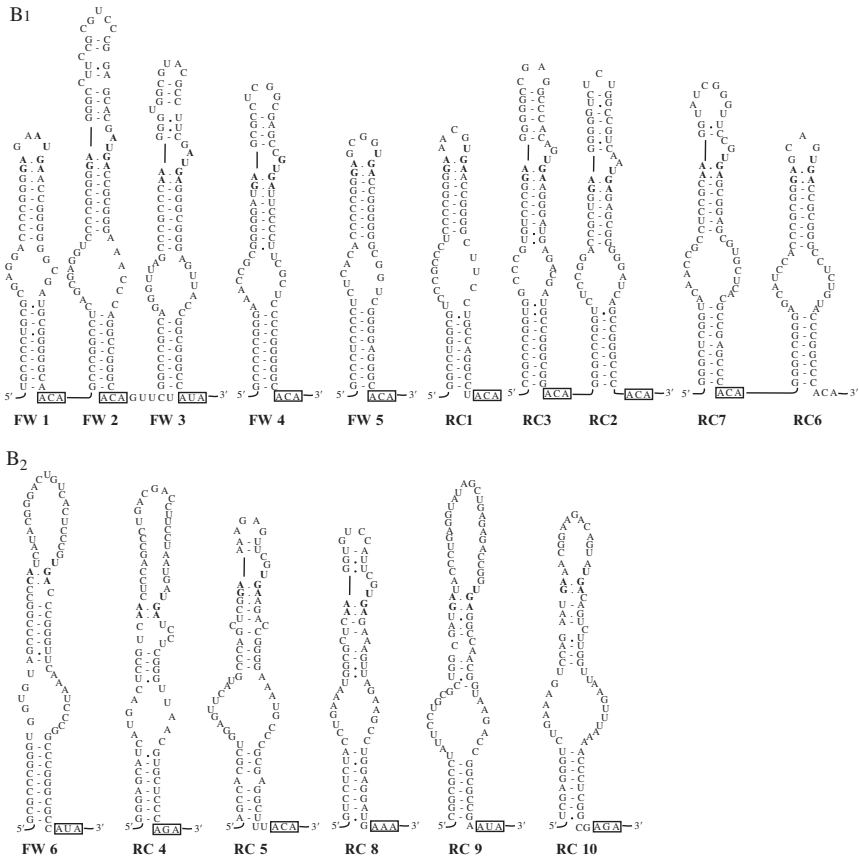


Figure 15.4 Results of an ERPIN search run on a yet-unstudied genome, *Thermococcus kodakarensis*, and proposed secondary structures of the identified candidates. Same legend as in Fig. 15.3. Panel (A) also indicates the absolute score and the e-value of the candidate. Panel (B₁) displays the candidates retained after inspection of their structures, and panel (B₂) displays the discarded candidates.

because their coding sequences are adjacent in the genome. Two pairs of motifs with a negative e-value are also encoded by adjacent sequences (RC3 and RC2 on the one hand, and RC7 and RC6 on the other hand). Therefore, they likely belong to two H/ACA sRNAs, each containing two H/ACA motifs. In addition, the coding sequences of the 10 motifs represented in panel B₁ are all located between ORFs.

In contrast, the six candidate motifs with an e-value between 1e-01 and 4 (panel B₂) can be discarded for several reasons: (1) except for motif RC5, they do not contain an ACA triplet; (2) most of them contain bulge residues or an internal loop in helix 1 or in helix 2; (3) the large terminal loops of motifs FW6, RC4, RC9, and RC10 are not structured. Finally, the

candidate motif FW 6 corresponds to a tRNA^{Asp}(GUC), and the coding sequences of the five other motifs are located within ORFs. On the basis of our present knowledge, sRNA genes may partially overlap ORFs but are not included in ORFs. A strong argument for the validation of the 10 H/ACA motifs in *T. kodakarensis* (Fig. 15.4, panel B1) is their homology with the 10 H/ACA sRNAs found in the *Pyrococcus* species. Hence, we predict that the 10 motifs detected in *T. kodakarensis* are the homologs of the 10 experimentally demonstrated *Pyrococcus* H/ACA motifs.

Interestingly, no new putative motifs with a satisfying negative e-value were found when a second screening was performed after insertion of the 10 validated motifs in the profile by use of the same parameters as in the first screening step (order 1 and cutoff values of 110%). Only the number of motifs with a high e-value increased, suggesting that all of the true H/ACA motifs were detected in the first search. Therefore, the 41 known H/ACA motifs aligned in the profile seem to provide enough information for an exhaustive selection of the *T. kodakarensis* motifs in a single run.

We next tested whether introduction of the probable *T. kodakarensis* H/ACA motifs in the profile would increase the selection of the *Pyrococcus* species known H/ACA motifs in blind tests (Table 15.2). This is, indeed, the case. When a blind ERPIN search is performed on the *P. abyssi* genome with a profile including the H/ACA motifs from *A. fulgidus*, *M. jannaschii*, *P. furiosus*, and *P. horikoshii*, nine of the known H/ACA motifs are detected with the priority order 1 and a 95% cutoff value (Table 15.2). No false-positive result is found with a 95% cutoff value. When the *T. kodakarensis* motifs are also added in the alignment, the 10 *P. abyssi* motifs are found when a cutoff value of 95% is used. However, the number of false-positive results increases in this case (Table 15.2).

2.1.6. Recommendations

On the basis of the data presented in Tables 15.1 and 15.2, the cutoff values used should be defined, taking into account the number and origin of the H/ACA motifs aligned in the profile. If they belong to species that are phylogenetically closely related to the studied species, small cutoff values can be used. They are expected to allow a selection of the true motifs with a minimum of noise. When a phylogenetically distant species is studied, we recommend the use of a cutoff value of 110% for each step to select true motifs that may exhibit divergences relative to the canonical structure. Hence, the closer the species, the smaller the cutoff values can be.

As in most cases, use of priority order 1 gives better results; we recommend the use of this order and only exceptionally, when needed, order 2.

As illustrated by the blind tests in Table 15.1, to complete the identification of H/ACA motifs in a new species, it may be worthwhile performing a second screening, after inclusion in the profile of the H/ACA-like motifs that were identified in the first step and contain all the needed characteristic structural features.

Table 15.2 Tests of the effect of the integration of the *T. kodakarensis* H/ACA motifs identified in this work, on blind tests performed on the *P. abyssi* genome

			Blind Tests performed with the profile A ^a				Bind Tests performed with the profile B ^a			
			Order (1)		Order (2)		Order (1)		Order (2)	
			Number of positive results	Number of likely false-positive results	Number of positive results	Number of likely false-positive results	Number of positive results	Number of likely false-positive results	Number of positive results	Number of likely false-positive results
<i>P. abyssi</i>	10 ^(2,3,4)	95%	9	0	7	0	10	1	7	1
		100%	10	1	8	0	10	2	8	2
		110%	10	1	8	1	10	4	8	5

^a The profile A contains the H/ACA motifs of *A. fulgidus*, *M. jannaschii*, *P. horikoshii* and *P. furiosus*. The profile B corresponds to the profile A with, in addition, the *T. kodakarensis* H/ACA motifs.

The H/ACA motifs detected in *T. kodakarensis* are included in the H/ACA profile available at <http://tagc.univ-mrs.fr/asterix/erpin/>. We are currently analyzing the H/ACA motifs of all completely sequenced archaeal genomes, and an updated H/ACA profile including not experimentally verified motifs will soon be available at the same URL.

Concerning the interpretation of the data, we recommend the selection of H/ACA-like motifs with negative e-values for the subsequent analysis (screening for the target sequence). Motifs with e-values between 1 and $1e-02$ can be maintained if they contain almost all the needed structural elements; a default in one of the structural elements may be accepted at this stage of the selection, because screening for target sequences will also help in discriminating true candidates.

Note that one strong discriminatory feature of ERPIN results is the overlap of candidate coding sequences with sequences coding for tRNAs, rRNAs, or proteins. Note also that, as observed in *T. kodakarensis*, when the apical loop closing the K-turn motif is long enough, an important criterion for the validation of the H/ACA motif is the possibility of forming a stem-loop structure in this apical region.

Finally, the newly identified sRNAs have to be experimentally confirmed with the approaches described in the third section of this chapter.

2.2. Search for targets of the H/ACA-like motifs

The identification of a putative RNA target for a given H/ACA motif is initially based on the search for a dinucleotide containing a uridine residue at the 5' position. This dinucleotide should be flanked by two sequences complementary to the two strands of the expected pseudouridylation pocket. To this end, one has to first establish the most likely secondary structure of the H/ACA motif, and we will comment in the following on the difficulties that may be encountered at this step. Once the 2D structure is established, a series of RNA descriptors, describing the various possible base-paired structures that can be formed by the target RNA and the two guiding sequences, have to be settled (Fig. 15.5). The different steps in target identification are explained in the following.

2.2.1. Requirements

2.2.1.1. Hardware and software The RNAMEOT program (Gautheret *et al.*, 1993) version 2.1 can be run on most UNIX platforms. The source and a tutorial are available at <http://pages-perso.esil.univ-mrs.fr/~dgaut/download/>. Any equivalent simple descriptor-based software such as RNAMotif (Macke *et al.*, 2001) can also be used. We selected the RNAMEOT program because it is simple and easy to use.

2.2.1.2. Sequence data The sequence format used for RNAMEOT is FASTA. The ribosomal RNA sequences of each of the studied archaeal species were extracted from their genomic sequence subsets, including all the noncoding RNAs (file with extension .frn obtained at <ftp://ftp.ncbi.nih.gov/genomes/Bacteria>).

2.2.2. Procedure to build the RNAMEOT descriptors and run the program

2.2.2.1. Construction of the descriptors On the basis of our present knowledge of the archaeal H/ACA system, there is no requirement concerning the identity of the residue at the 3' position in the single-stranded dinucleotide (Fig. 15.5). Altogether, the two base-paired regions most generally contain at least nine base pairs, and one wobble pair can be included in these nine pairs.

In Fig. 15.5, we illustrate the various descriptors that have to be built to look for the possible targets of the H/ACA motif Pab91 of *P. abyssi*. A roughly equivalent number of base pairs in the two helical regions is expected to be the most frequent situation (4 and 5 base pairs or 5 and 4 base pairs). However, one cannot exclude the possibility of having highly asymmetrical base-pair interactions (1 + 8, 2 + 7, 3 + 6, 6 + 3, 7 + 2, and 8 + 1).

We will explain how one of the descriptors needed to look for one of the base-paired configurations is built; all of the other descriptors are built in the same way. The descriptor used as an example in Fig. 15.5 (panel C) is devoted to the search of a target sequence forming base-pair interactions including 4 and 5 residues of the 3' and 5' strands of the pseudouridylation pocket, respectively (descriptor designated as 4 + 5). The descriptor file includes two lines: the first one identifies the searched sequence, s1 in this example. The second line defines the sequence requirements.

```
s1
s1 11 : 11 YUYYUNRGGYU
```

The two numbers separated by a colon correspond to the minimum and maximum nucleotide lengths of the searched target sequence; Y and R correspond to pyrimidine and purine residues, respectively.

If no positive result is obtained when the first series of descriptors corresponding to a total of nine base pairs is used, it is recommended to look for possible targets able to form only eight or seven base pairs. The use of this low number of base pairs may allow the selection of interactions that include mismatches. Indeed, such interactions cannot be selected directly by the RNAMEOT descriptors described previously.

In targets that include mismatches, an extension of the possible base-pair interactions should be tested by inspection of the guide and target

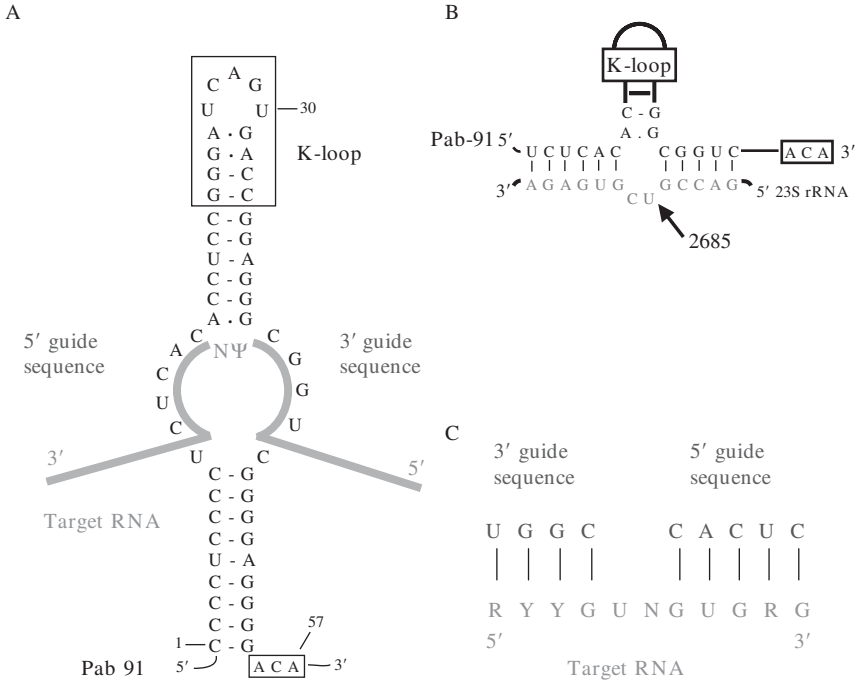


Figure 15.5 The strategy used for the identification of the targets of the H/ACA-like motifs. (A) The secondary structure of the *P. abyssi* Pab91 sRNA is shown with a schematic representation of the target RNA base pair interactions formed with the 5' and 3' guide sequences of the H/ACA motif (Charpentier *et al.*, 2005). The ΨN dinucleotide obtained after U to Ψ conversion is shown. (B) The known interaction formed between the *P. abyssi* Pab91 sRNA and its known target sequence in *P. abyssi* 23S rRNA. The targeted U2685 residue is indicated (Charpentier *et al.*, 2005). (C) As an example, we show the structure that would be used to define a 5 + 4 descriptor for the selection of target sequences of RNA Pab91. Once the 5' and 3' guide sequences are specified, the possible complementary sequences are written, taking into account possible wobble G·U base-pair interactions (Y = pyrimidines, R = purines) and the chosen numbers of base pairs to be formed with the 5' and the 3' guide sequence, respectively. The rRNA sequences are then screened with a descriptor of these possible complementary sequences by use of RNAMEOT.

sequences. Even when a putative interaction including nine base pairs is found, we recommend looking for possible extensions of the base-pair interactions by inspection of the flanking sequences.

2.2.2.2. How to start the search? The program RNAMEOT is started using the following command line:

```
rnamot -s -s <rRNA sequence> -d <descriptor> -o <output> -t
```

The name of the executable is “*nmot*.” The first “-s” specifies the mode of use (-s for search, by opposition to -a for alignment); the second “-s” is used to provide the name and location of the file containing the rRNA sequences that will be screened, and “-d” the name and location of the descriptor file. Finally, the “-o” option specifies the file in which the results are written. The “-t” option activates the display of the search stage and the number of positive results obtained. A search has to be executed for each of the descriptors. However, whatever descriptor is used, the search time does not exceed a few seconds.

2.2.3. Example of application: search for the target sequences of the *M. jannaschii* FW1 motif

In the *M. jannaschii* motif FW1, helices 1 and 2 contain nine and six successive Watson–Crick base pairs, respectively (Fig. 15.6, panel A). In addition, one wobble G•U pair can be present at the extremity of helix 2. The formation or the absence of this wobble pair modifies the pseudouridylation pocket and, therefore, the putative target sequences. Hence, two series of descriptors were built as described earlier. In the first series, the G•U pair at the extremity of helix 2 was considered to be formed. In the second series, it was considered to be opened. No putative target sequence was detected for this second series of descriptors. In contrast, two putative target sequences were detected with the first series, suggesting the formation of the G•U pair in the guide sRNA. One of the target sequences (Fig. 15.6, panel B, hit sequence 1) was obtained by use of a descriptor designed for the search of sequences forming three and six base pairs with the 3' and 5' guiding sequences, respectively (3 + 6 descriptor). The second target sequence (panel B, hit sequence 2) was selected with a 4 + 5 descriptor. Hit sequence 1 belongs to 16S rRNA, and residue U1261 would be the target. Hit sequence 2 belongs to 23S rRNA, and residue U2015 would be the target.

Inspection of both rRNA flanking sequences reveals the possible extension by 2 (G–C and G•U) and 1 (G–C) base pairs of the interactions that can be formed with the 5' guide sequences for hit sequences 1 and 2, respectively. Therefore, the base-pair interaction formed for hit sequence 1 involves two stretches of three and eight base pairs. However, each of them contains two wobble pairs. Thus far, such a high number of G•U pairs has never been described for the previously studied H/ACA sRNAs and snoRNAs. Therefore, hit sequence 1 may correspond to a false-positive result. In contrast, hit sequence 2 forms two stretches of four and six uninterrupted Watson–Crick base pairs. A strong argument in favor of the validity of hit sequence 2 is the observation of H/ACA motifs proposed to target similar positions in the 23S rRNAs of *P. furiosus* (H/ACA motif Pf7) and *A. fulgidus* (H/ACA motif Afu4) (Rozhdestvensky *et al.*, 2003) (Fig. 15.6, panels C and D).

Figure 15.6 Target sequences found for the *M. jannaschii* FW1 motif by an RNAMEOT search. (A) Secondary structure proposed for the *M. jannaschii* FW 1 candidate. The compensatory mutations and single base substitutions, insertions, or deletions found in the *A. fulgidus* Afu4 (*A.f*) and *P. furiosus* Pf7 (*P.f*) motifs compared with the *M. jannaschii* FW1 motif are indicated. Δ indicates one missing nucleotide. (B) The possible base-pair interactions between the *M. jannaschii* FW1 motif and the two putative targeted sequences identified for this motif by the RNAMEOT search (hit sequences 1 and 2) are shown. The distances between the basal extremity of helix 2 and either the K-loop or the ACA box are indicated. The target sequences proposed for the third H/ACA motif of the *P. furiosus* Pf7 sRNA (C) and the third H/ACA motif of the *A. fulgidus* Afu4 sRNA (Rozhdestvensky *et al.*, 2003) (D), and the interactions that they can form with the guide RNA are represented.

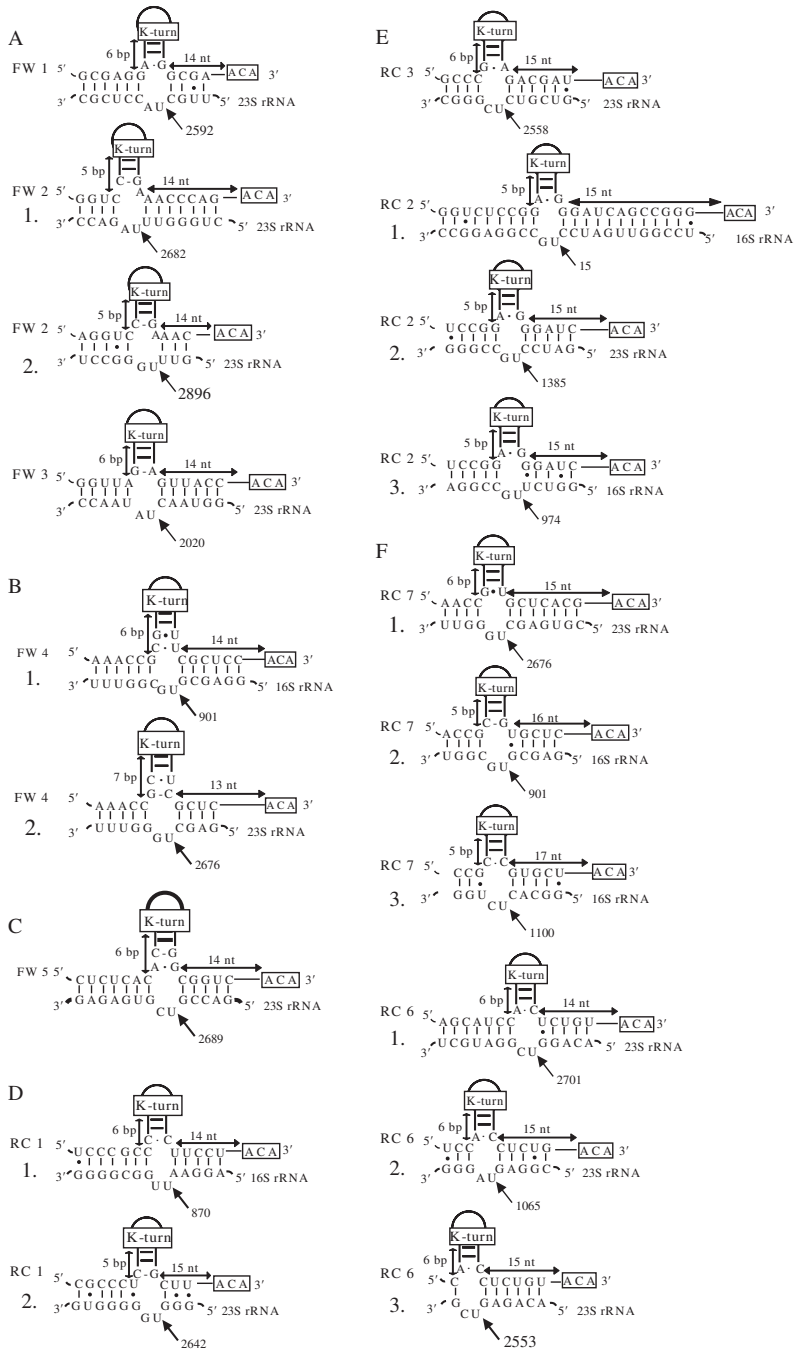
2.2.4. Application to the search of target sequences for the 10 H/ACA-like motifs found in *T. kodakarensis*

may be opened so that its two nucleotides can be included in the guide sequences (Fig. 15.7, panel A). At least two alternative lengths of helix 2 are possible for almost all the selected motifs of *T. kodakarensis* (Fig. 15.7).

When the various series of descriptors built for each of the motifs are used, no putative target sequence was found for the six motifs that we discarded after inspection of their primary and secondary structures. In contrast, up to 3 putative target sequences were found for each of the 10 motifs that we retained after this inspection. In Fig. 15.7, we present the corresponding base-pair interactions by order of stability for each of the motifs. The order of presentation also takes into account the fact that some of the H/ACA motifs are expected to belong to the same sRNA (FW1, FW2, and FW3; RC2 and RC3; RC6 and RC7).

Most generally, when two or three possible target sequences were detected for a given motif, they differ by their stabilities and/or by the distance between helix 2 and the ACA trinucleotide. For instance, among the two hit sequences found for each of the FW2 and RC1 motifs, the hit sequences numbered 2 form shorter base-pair interactions with the guide sequences compared with the hit sequences numbered 1. In cases in which three putative target sequences were found (motifs RC2, RC6 and RC7), most frequently, one of them forms significantly more stable interactions than the two other ones. Therefore, in each case, we prefer the hit sequence forming the more stable interaction (hits numbered 1). However, peculiar results were obtained for the RC2 motif: both hits 1 and 2 form stable interactions (9 Watson–Crick base pairs and a G•U wobble pair for hit sequence 2 and 19 Watson–Crick pairs and two G•U pairs for hit sequence 1). By use of hit sequence 2, the K-turn structure and the ACA triplet are located at correct distances from the basal extremity of helix 2 (Fig. 15.7E). Thus, despite the lower stability of the interaction formed by hit sequence 2 compared with hit sequence 1, we have no criterion to eliminate this hit sequence. Furthermore, it should be noted that the interactions that can be formed with hit sequence 1 are unusually long. Formation of these interactions would result in the disruption of a large part of helix 1. Because helix 1 is expected to be important for the folding and the activity of the H/ACA sRNP complex, the possible interaction between the hit sequence 1 and motif RC2 may have another function than RNA pseudouridylation. Interestingly, this hit sequence is located 3 nts downstream from the 16S rRNA 5' extremity. We may imagine an RNA chaperone role of motif RC2, in addition to its possible role in pseudouridylation.

Note that no Watson–Crick base pair in helix 1 is opened in any of the interactions formed with the other H/ACA motifs and hit sequences. Only the G•U pair at the extremity of helix 1 in motif RC1 may be formed or opened, depending on whether a G•U pair is included or not in the interaction established between hit sequence 1 and the 5' strand of the guide sequence (Fig. 15.7D).



On the basis of the data presented in Fig. 15.7, the selected hit sequences can form from 4 up to 7 base pairs with the 5' and 3' guide sequences. Interestingly, for most of the more favorable hit sequences, a noncanonical pair is present at the basal extremity of helix 2, which may be required to ensure enough flexibility in the cruciform structure.

2.2.5. Tips for searching and analyzing H/ACA targets

As evidenced by the preceding examples, a major difficulty in the prediction of target sequences comes from the necessity of establishing the correct secondary structure of the putative H/ACA-like motif and to define the basal extremity of helix 2. Because of the irregularity and size variations of helix 2, several possible conformations can sometimes be proposed for a given motif. In contrast, helix 1, which consists in a regular stretch of Watson–Crick pairs, is generally easier to predict.

In addition, even when the correct secondary structure has been established, the size variability of the base-pair interactions formed with the 5' and 3' guide sequences increases the number of RNAMOT searches that have to be run for a given motif.

As illustrated by the *T. kodakarensis* example given previously, we recommend the initial use of a series of descriptors corresponding to interactions including 9 base pairs. A reduction of this number of base pairs in the first round of the selection would unnecessarily increase the number of false-positive results and limit the chance of obtaining positive results. Hits with a lower level of complementarity can be searched in a second step, if no hit is obtained with the 9 base pair descriptors. Restriction of the starting descriptors to interactions only including Watson–Crick pairs would avoid the selection of interactions containing one wobble G•U pair such as the one found for the *T. kodakarensis* motif FW1.

For interpretation of the results note that: (1) a non-Watson–Crick base pair is often present at the basal extremity of helix 2, (2) the distances between the basal extremity of helix 2 and the K-turn/K-loop sequence on the one hand, and the ACA triplet on the other hand, are most often 5–6 bps and 14–15nts, respectively, (3) the base-pair interactions with both guide sequences include most frequently 4–7 residues; however, more asymmetrical base-pair interactions cannot be excluded.

Figure 15.7 Putative target sequences identified for the 10 H/ACA motifs of *T. kodakarensis* by an RNAMOT search. Panels A–F display the base-pair interactions that can be formed between the putative target sequences identified by use of RNAMOT and each of the 10 H/ACA motifs identified for *T. kodakarensis*. Each panel corresponds to a given H/ACA sRNA. For each motif the possible base-pair interactions are represented according to their stability.

As illustrated by the *T. kodakarensis* example used in this chapter, the search of putative target sequences in rRNAs by RNAMEOT can confirm the discrimination of true H/ACA motifs made by inspection of the H/ACA-like motifs obtained in the ERPIN search. However, if an H/ACA motif bears all the characteristic features of archaeal H/ACA motifs and if no putative target sequence is found in rRNAs, its target sequence may be contained in another kind of RNA. In this case, the search for possible target sequences in other RNAs will be useful.

2.3. Phylogenetic and experimental validation of the results

2.3.1. Phylogenetic validation

When new H/ACA motifs and their putative targets have been selected by the proposed approach, several comparisons with known data on archaeal H/ACA motifs may be helpful for their validation. This was illustrated in the interpretation of the RNAMEOT hit sequences found for the FW1 motif of *M. jannaschii*, where three homologous H/ACA motifs from *P. furiosus*, *A. fulgidus*, and *M. jannaschii*, that differ by a limited number of base substitutions, insertions or deletions (Fig. 15.6), guide modifications at similar positions in 23S rRNA. Therefore, as a first step in the validation of the data, we recommend the comparison of the putative H/ACA motifs obtained for a new archaeal genome with the already identified H/ACA motifs. However, note that phylogenetic comparisons are limited by the rapid evolution of sRNAs.

Although pseudouridylation sites in archaeal rRNAs have only been identified in a limited number of species and are not strongly conserved from one species to the other (Del Campo *et al.*, 2005; Ofengand, 2002), a comparison of the putative target sites detected by RNAMEOT with known pseudouridylation sites in archaeal rRNAs may also be useful, especially when Ψ residues have been experimentally identified in rRNAs from phylogenetically related species.

We will again use the *M. jannaschii* FW1 motif to illustrate this point. One of the two putative target sites found by the RNAMEOT (U2015 in 23S rRNA) was experimentally found to be pseudouridylated in the 23S rRNA of archaeal species (Ofengand and Bakin, 1997; and for review, Ofengand, 2002; corrected in Del Campo *et al.*, 2005). The corresponding positions in the yeast and human large subunit rRNAs are also pseudouridylated, and the H/ACA snoRNAs snR191 in yeast and hU19 in human are, respectively, proposed to guide the modifications (Badis *et al.*, 2003; Bortolin and Kiss, 1998).

On the contrary, no counterpart of the second putative target residue (U1261) of motif FW1 was found to be pseudouridylated, neither in archaea nor in eukarya. Therefore, taken together, the FW1 motif of *M. jannaschii* can reasonably be considered as guiding U to Ψ conversion at position U2015 in 23S rRNA.

2.3.2. Experimental validations

The situation described previously for the FW1 motif of *M. jannaschii* is a highly favorable one in terms of validation. It is not always possible to find homologous H/ACA motifs in other species and/or the presence of a Ψ residue at the targeted position in another archaeal species. When this is not the case, direct experimental proof is particularly important to validate the data.

A first simple and important experiment to do is a Northern blot or primer extension analysis on total RNA extracted from the studied archaea. This will allow verification of the presence of the proposed H/ACA sRNA. Radiolabeled primers complementary to the H/ACA sRNAs candidates can be used for this purpose. In addition, these kinds of experiments will give information on the length of the sRNAs because, as explained before, archaeal sRNAs have variable lengths. A protocol for primer extension analysis is proposed in Chapter 2 in this volume by Motorine *et al.*

Another validation approach involves localizing the pseudouridylation positions in rRNAs by use of the CMCT-RT approach as described by Motorine *et al.* in Chapter 2. This is not an easy technique to handle. However, a detailed protocol and tips are given by Motorine *et al.*

Finally, the complete verification of the proposed target sequences can be obtained by *in vitro* reconstitution of an active H/ACA sRNP by use of *in vitro* transcribed guide RNA and target sequence and recombinant aCBF5, aNOP10, L7Ae, and aGAR1 proteins, as described in Chapter 16 by Charpentier *et al.*, in this volume.



3. CONCLUSIONS

The knowledge-based approach described herein, which combines the search for H/ACA motifs and their respective target(s), is an efficient approach as illustrated in the numerous examples given in this chapter. Its efficiency will be further improved, on enrichment of the H/ACA profile by inclusion of new validated motifs. One does not need much expertise in computing to use the ERPIN and RNAMOT softwares, which are easy to use. Moreover, a web server version of ERPIN is available for the user to become familiar with the method. No data other than the structural or functional features of H/ACA sRNAs, which are described in this chapter, is necessary.

Compared with the results recently obtained by application of the MilPat tool to the identification of H/ACA-like motifs in the genomes from *M. jannaschii* and three *Pyrococcus* species (Thebault *et al.*, 2006), the approach proposed in this chapter is more directed and, therefore, gives a very limited number of false-positive results. By the use of the H/ACA profile

enriched by the H/ACA motifs of *T. kodakarensis* identified in this work, we detected, in the blind test presented in Table 15.2, the 10 H/ACA motifs present in *P. abyssi*, *P. furiosus*, and *P. horikoshii*. When the MilPat approach is used, among the 89 to 148 candidates H/ACA motifs found for the different *Pyrococcus* species studied, only 6 to 7 of the known motifs were detected, depending on the species. We think that after inclusion of the H/ACA motifs from a limited number of archaeal species belonging to different archaeal orders, which is currently being done by our team, we will be able to propose soon (at site <http://tagc.univ-mrs.fr/asterix/erpin/>), a highly powerful tool for the search of H/ACA sRNAs in any archaeal species. The strength of our approach is the coupling of the search of H/ACA-like motifs to the search of their target sequences. As evidenced in the given examples, both steps in this strategy participate in the selection of the true H/ACA motifs among the identified hit motifs.

We would like to point out that the strategy proposed in this chapter is well suited to the search of archaeal H/ACA motifs because of their structural specificities (presence of a K-turn or a K loop and of G-C rich helices). This strategy would be much less efficient if applied to the search of snoRNA coding sequences. However, a specific computational approach dedicated to eukaryal H/ACA snoRNAs has been developed (Schattner *et al.*, 2006).

REFERENCES

- Badis, G., Fromont-Racine, M., and Jacquier, A. (2003). A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *RNA* **9**, 771–779.
- Baker, D. L., Youssef, O. A., Chastkofsky, M. I., Dy, D. A., Terns, R. M., and Terns, M. P. (2005). RNA-guided RNA modification: Functional organization of the archaeal H/ACA RNP. *Genes Dev.* **19**, 1238–1248.
- Balakin, A. G., Smith, L., and Fournier, M. J. (1996). The RNA world of the nucleolus: Two major families of small RNAs defined by different box elements with related functions. *Cell* **86**, 823–834.
- Bortolin, M. L., and Kiss, T. (1998). Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. *RNA* **4**, 445–454.
- Bourdeau, V., Ferbeyre, G., Pageau, M., Paquin, B., and Cedergren, R. (1999). The distribution of RNA motifs in natural sequences. *Nucleic Acids Res.* **27**, 4457–4467.
- Charpentier, B., Muller, S., and Branlant, C. (2005). Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res.* **33**, 3133–3144.
- Charron, C., Manival, X., Clery, A., Senty-Segault, V., Charpentier, B., Marmier-Gourrier, N., Branlant, C., and Aubry, A. (2004). The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J. Mol. Biol.* **342**, 757–773.

- Del Campo, M., Recinos, C., Yanez, G., Pomerantz, S. C., Guymon, R., Crain, P. F., McCloskey, J. A., and Ofengand, J. (2005). Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of *Haloarcula marismortui* and *Deinococcus radiodurans*. *RNA* **11**, 210–219.
- Ganot, P., Bortolin, M. L., and Kiss, T. (1997a). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**, 799–809.
- Ganot, P., Caizergues-Ferrer, M., and Kiss, T. (1997b). The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.* **11**, 941–956.
- Gautheret, D., and Lambert, A. (2001). Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J. Mol. Biol.* **313**, 1003–1011.
- Gautheret, D., Major, F., and Cedergren, R. (1993). Modeling the three-dimensional structure of RNA using discrete nucleotide conformational sets. *J. Mol. Biol.* **229**, 1049–1064.
- Hamma, T., Reichow, S. L., Varani, G., and Ferre-D'Amare, A. R. (2005). The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat. Struct. Mol. Biol.* **12**, 1101–1107.
- Henras, A., Henry, Y., Bousquet-Antonelli, C., Noaillac-Depeyre, J., Gelugne, J. P., and Caizergues-Ferrer, M. (1998). Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J.* **17**, 7078–7090.
- Huang, Z. P., Zhou, H., He, H. L., Chen, C. L., Liang, D., and Qu, L. H. (2005). Genome-wide analyses of two families of snoRNA genes from *Drosophila melanogaster*, demonstrating the extensive utilization of introns for coding of snoRNAs. *RNA* **11**, 1303–1316.
- Huang, Z. P., Zhou, H., Liang, D., and Qu, L. H. (2004). Different expression strategy: Multiple intronic gene clusters of box H/ACA snoRNA in *Drosophila melanogaster*. *J. Mol. Biol.* **341**, 669–683.
- Khanna, M., Wu, H., Johansson, C., Caizergues-Ferrer, M., and Feigon, J. (2006). Structural study of the H/ACA snoRNP components Nop10p and the 3' hairpin of U65 snoRNA. *RNA* **12**, 40–52.
- Kiss, A. M., Jady, B. E., Bertrand, E., and Kiss, T. (2004). Human box H/ACA pseudouridylation guide RNA machinery. *Mol. Cell Biol.* **24**, 5797–5807.
- Klein, D. J., Schmeing, T. M., Moore, P. B., and Steitz, T. A. (2001). The kink-turn: A new RNA secondary structure motif. *EMBO J.* **20**, 4214–4221.
- Laferriere, A., Gautheret, D., and Cedergren, R. (1994). An RNA pattern matching program with enhanced performance and portability. *Comput. Appl. Biosci.* **10**, 211–212.
- Lafontaine, D. L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.* **12**, 527–537.
- Lambert, A., Legendre, M., Fontaine, J. F., and Gautheret, D. (2005). Computing expectation values for RNA motifs using discrete convolutions. *BMC Bioinformatics* **6**, 118.
- Lambert, A., Fontaine, J. F., Legendre, M., Leclerc, F., Permal, E., Major, F., Putzer, H., Delfour, O., Michot, B., and Gautheret, D. (2004). The ERPIN server: An interface to profile-based RNA motif identification. *Nucleic Acids Res.* **32**, W160–W165.
- Lambert, A., Lescure, A., and Gautheret, D. (2002). A survey of metazoan selenocysteine insertion sequences. *Biochimie* **84**, 953–959.
- Legendre, M., Lambert, A., and Gautheret, D. (2005). Profile-based detection of micro-RNA precursors in animal genomes. *Bioinformatics* **21**, 841–845.
- Lescure, A., Gautheret, D., Carbon, P., and Krol, A. (1999). Novel selenoproteins identified *in silico* and *in vivo* by using a conserved RNA structural motif. *J. Biol. Chem.* **274**, 38147–38154.
- Lescure, A., Gautheret, D., and Krol, A. (2002). Novel selenoproteins identified from genomic sequence data. *Methods Enzymol.* **347**, 57–70.

- Li, L., and Ye, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443**, 302–307.
- Li, S. G., Zhou, H., Luo, Y. P., Zhang, P., and Qu, L. H. (2005). Identification and functional analysis of 20 Box H/ACA small nucleolar RNAs (snoRNAs) from *Schizosaccharomyces pombe*. *J. Biol. Chem.* **280**, 16446–16455.
- Macke, T. J., Ecker, D. J., Gutell, R. R., Gautheret, D., Case, D. A., and Sampath, R. (2001). RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res.* **29**, 4724–4735.
- Manival, X., Charron, C., Fourmann, J. B., Godard, F., Charpentier, B., and Branlant, C. (2006). Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5–aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res.* **34**, 826–839.
- Massenet, S., Motorin, Y., Lafontaine, D. L., Hurt, E. C., Grosjean, H., and Branlant, C. (1999). Pseudouridine mapping in the *Saccharomyces cerevisiae* spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol. Cell Biol.* **19**, 2142–2154.
- Meier, U. T., and Blobel, G. (1994). NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria. *J. Cell Biol.* **127**, 1505–1514.
- Ni, J., Tien, A. L., and Fournier, M. J. (1997). Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**, 565–573.
- Ofengand, J. (2002). Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett.* **514**, 17–25.
- Ofengand, J., and Bakin, A. (1997). Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaeobacteria, mitochondria and chloroplasts. *J. Mol. Biol.* **266**, 246–268.
- Rashid, R., Liang, B., Baker, D. L., Youssef, O. A., He, Y., Phipps, K., Terns, R. M., Terns, M. P., and Li, H. (2006). Crystal structure of a Cbf5–Nop10–Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol. Cell* **21**, 249–260.
- Rozenski, J., Crain, P. F., and McCloskey, J. A. (1999). The RNA Modification Database: 1999 update. *Nucleic Acids Res.* **27**, 196–197.
- Rozhdestvensky, T. S., Tang, T. H., Tchirkova, I. V., Brosius, J., Bachellerie, J. P., and Huttenhofer, A. (2003). Binding of L7Ae protein to the K-turn of archaeal snoRNAs: A shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.* **31**, 869–877.
- Schattner, P., Barberan-Soler, S., and Lowe, T. M. (2006). A computational screen for mammalian pseudouridylation guide H/ACA RNAs. *RNA* **12**, 15–25.
- Schattner, P., Decatur, W. A., Davis, C. A., Ares, M., Jr., Fournier, M. J., and Lowe, T. M. (2004). Genome-wide searching for pseudouridylation guide snoRNAs: Analysis of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res.* **32**, 4281–4296.
- Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., Elge, T., Brosius, J., and Huttenhofer, A. (2002a). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. USA* **99**, 7536–7541.
- Thebault, P., de Givry, S., Schiex, T., and Gaspin, C. (2006). Searching RNA motifs and their intermolecular contacts with constraint networks. *Bioinformatics* **22**, 2074–2080.
- Torchet, C., Badis, G., Devaux, F., Costanzo, G., Werner, M., and Jacquier, A. (2005). The complete set of H/ACA snoRNAs that guide rRNA pseudouridylations in *Saccharomyces cerevisiae*. *RNA* **11**, 928–938.
- Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R., and Ficner, R. (2000). Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell* **6**, 1331–1342.

- Wang, C., and Meier, U. T. (2004). Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.* **23**, 1857–1867.
- Watkins, N. J., Gottschalk, A., Neubauer, G., Kastner, B., Fabrizio, P., Mann, M., and Luhrmann, R. (1998). Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA* **4**, 1549–1568.
- Yuan, G., Klambt, C., Bachellerie, J. P., Brosius, J., and Huttenhofer, A. (2003). RNomics in *Drosophila melanogaster*: Identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.* **31**, 2495–2507.

B. Discussion

L'application de cette approche aux séquences génomiques des trois *Pyrococcus* a d'abord permis l'identification d'un 6^{ième} candidat, Pab160 (Fig. R.18). Le candidat Pab160 comporte un seul motif et présente les mêmes caractéristiques structurales que les 5 autres candidats. Celui-ci n'avait pu être détecté par l'utilisation de RNAMOT au sein des séquences IRC sélectionnées. En effet, sa séquence chevauche de plus de 15 nt une ORF adjacente. En raison de la trop faible spécificité de la recherche de motifs H/ACA par l'approche basée sur RNAMOT, nous ne pouvions appliquer ce programme à des séquences génomiques entières. Au contraire, l'application de ERPIN révèle un faible taux de faux positifs et permet une recherche de nouveaux candidats dans les séquences génomiques entières et non plus seulement dans les séquences IRC sélectionnées. Il est à noter que la seule alternative de recherche de sRNA H/ACA d'archaea, basée sur l'utilisation du programme MilPat, conduit à la détection d'un grand nombre de faux positifs (près d'une centaine) (Thebault *et al.*, 2006).

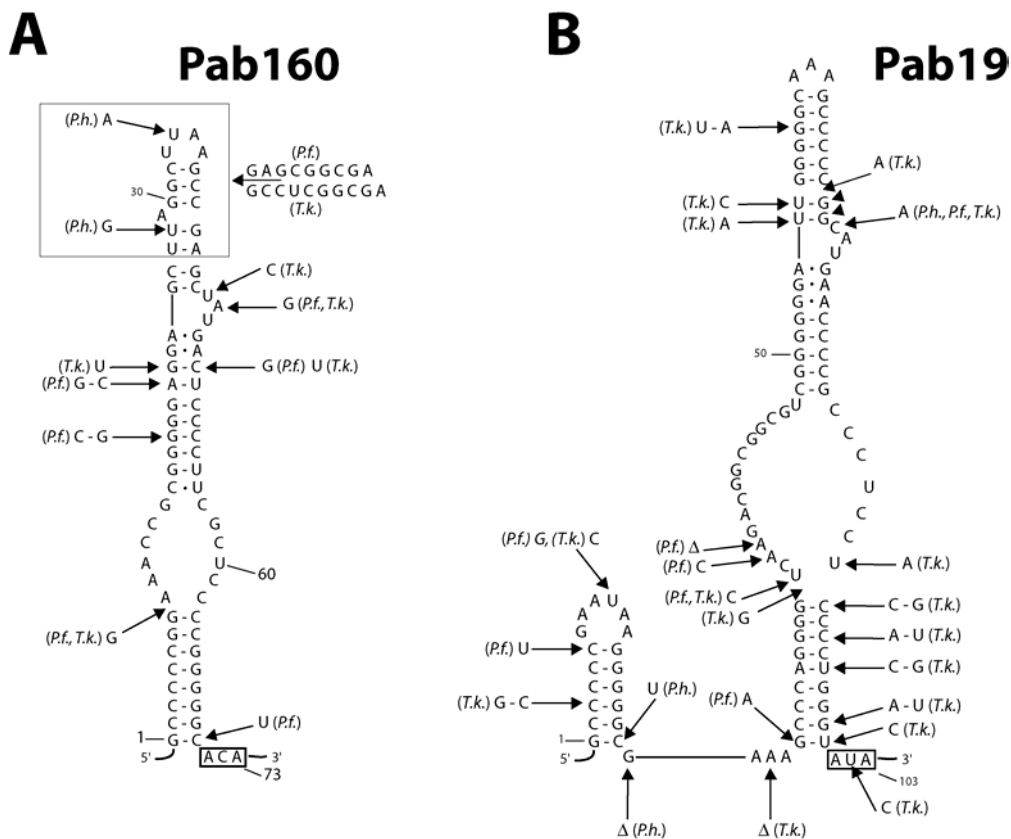


Figure R.18 : Structures secondaires proposées pour les 2 sRNA H/ACA supplémentaires identifiés chez trois espèces de *Pyrococcus* et chez *T. kodakarensis*. Les séquences et les structures secondaires proposées par MFOLD sont présentées pour Pab160 (A) et Pab19 (B). Les boîtes ACA et motifs K-turn sont encadrés. Les différences et les mutations compensatoires entre les homologues provenant des espèces de *P. abyssi*, *P. horikoshii* (*P.h.*), *P. furiosus* (*P.f.*) et *T. kodakarensis* (*T.k.*) sont indiquées.

Plus tardivement, l'étude des caractéristiques des sRNA H/ACA d'archaea nous a suggéré une certaine variabilité de la taille de la séquence de l'élément antisens en 5' de la poche interne du sRNA. Nous avons donc construit un alignement de séquences autorisant une plus grande taille de cette séquence. Nous avons ainsi identifié un 7^{ième} sRNA H/ACA : Pab19. Le candidat Pab19 comporte un seul motif qui présente, à l'exception de la taille de la séquence de l'élément antisens 5', des caractéristiques similaires aux autres motifs H/ACA (Fig. R.18). Il est à noter que son gène est localisé 30 nt en amont de la région 5'ETS de l'ARNr 16S. Cette séquence ETS forme, par repliement sur une autre région, les motifs Bulge-Helix-Bulge requis pour la maturation des ARNr (Tang *et al.*, 2002b). Ainsi, il est envisageable que la maturation de ce sRNA soit liée à la maturation des ARNr.

Le sRNA Pab160 est l'homologue de Pf9 de *P. furiosus* (Rozhdestvensky *et al.*, 2003) et de Tko4 (FW4) de *T. kodakarensis* (Muller *et al.*, 2007a) et Pab19 est l'homologue de Afu190 (Tang *et al.*, 2002a) et nous avons également identifié un homologue chez *T. kodakarensis* que nous avons appelé Tko19.

5. Etude de la contribution des sRNA H/ACA à la pseudouridylation des ARNr de P. abyssi

A ce stade, des progrès importants avaient été faits dans le domaine de la caractérisation des sRNA H/ACA potentiels de plusieurs espèces d'archaea. Par contre, les informations concernant l'existence, le nombre et la localisation des résidus Ψ dans les ARNr d'archaea étaient très limitées. De plus, les quelques études expérimentales qui avaient été réalisées avaient été faites sur des espèces autres que celles où des sRNA H/ACA avaient été recherchés par bioinformatique. Il s'agissait des ARNr 23S de *S. acidocaldarius* poussant à 80°C (Massenet *et al.*, 1999) et deux espèces halophiles, *H. halobium* (Ofengand et Bakin, 1997) et *H. marismortui* (Del Campo *et al.*, 2005 ; Kirpekar *et al.*, 2005) poussant à 42°C et 50°C. Un nombre limité de résidus Ψ , 6, 4 et 3 respectivement, avait été détecté dans ces espèces. Une analyse expérimentale approfondie des résidus Ψ dans les ARNr d'une espèce hyperthermophile d'archaea devait permettre de savoir si, comme proposé pour les 2'-O-méthylations, le nombre de résidus Ψ dans les ARNr augmentait en fonction de la température de croissance (Dennis *et al.*, 2001). Par ailleurs, la réalisation, en parallèle, dans une même espèce d'une recherche exhaustive des sRNA H/ACA et des résidus présents dans les ARNr devait permettre d'assigner chacun de ces sRNA H/ACA à une cible dans l'ARNr. Ceci devait apporter des informations importantes sur la spécificité de reconnaissance des séquences ARNr ciblés par les sRNA H/ACA. Enfin, ce type d'analyse devait permettre de savoir si la formation de tous les résidus Ψ des ARNr d'archaea est catalysée par le système reposant sur un ARN guide.

En couplant les méthodes bioinformatiques que nous avons développées, la méthode de RT-CMCT utilisée pour l'identification des résidus Ψ dans les ARN que nous maîtrisons au laboratoire (chapitre d'ouvrage N°1) et avec la méthode de reconstitution de sRNP H/ACA actives *in vitro* que nous avons développée, nous disposons de tous les outils nécessaires pour réaliser les différents volets de recherche cités ci-dessus.

Ainsi, nous avons choisi de travailler sur des espèces du genre *Pyrococcus* (*P. abyssi*, *P. furiosus* et *P. horikoshii*) qui sont des hyperthermophiles se développant de façon optimale à 100°C. Dans un premier temps, par une combinaison d'approches bioinformatiques, nous avons identifié 7 sRNA H/ACA potentiels dans ces espèces et, toujours par emploi d'approches bioinformatiques, nous avons recherché leurs cibles potentielles dans les ARNr. Plusieurs cibles potentielles ont été prédites pour la plupart des motifs H/ACA de ces sRNA. En utilisant la méthode RT-CMCT, nous avons vérifié si des résidus Ψ étaient présents aux positions prédites dans les ARNr de *P. abyssi*. Nous avons aussi testé la présence de résidus Ψ à toutes les positions dans les ARNr 16S et 23S où des modifications avaient été détectées dans d'autres espèces d'archaea. Nous avons de ce fait analysé environ 25% de la séquence des ARNr de *P.*

abyssi. Ceci nous a permis de commencer à faire un lien entre résidus Ψ présents dans les ARNr et les motifs H/ACA des différents sRNA H/ACA que nous avons mis en évidence. La relation position modifiée/motif H/ACA n'était pas univoque. En effet, plusieurs cibles étaient attribuées à plusieurs motifs. Afin d'avoir une réponse définitive sur le lien entre les résidus Ψ détectés et les motifs H/ACA identifiés, nous avons alors réalisé la reconstitution de particules H/ACA à partir de chacun des sRNA H/ACA produits par transcription *in vitro* en présence des protéines aCBF5, aNOP10, aGAR1 et L7Ae produites sous forme recombinantes chez *E. coli*. Puis nous avons testé l'activité *in vitro* de ces sRNP sur des petits fragments d'ARNr correspondant à leur cibles potentielles. Nous avons ainsi pu attribuer 14 des 16 résidus Ψ identifiés expérimentalement à 11 motifs H/ACA présents dans les 7 sRNA H/ACA identifiés. La détection de résidus Ψ à 3 positions des ARNr a été rendue impossible par la présence de pauses de la transcriptase inverse. Donc, au total, nous estimons que les ARNr de *P. abyssi* contiennent au moins 17 résidus Ψ , ce qui correspond au nombre le plus élevé observé à ce jour chez les archaea. Ceci s'explique sans doute par la température de croissance élevée de *P. abyssi*. Ces modifications sont localisées dans des régions fonctionnellement très importantes des ARNr (sites A et P et centre présentant l'activité peptidyl transférase). Huit des modifications sont localisées dans le domaine V de l'ARNr 23S, dans la partie participant à l'activité peptidyl transférase. Ces 8 résidus Ψ participent à la formation des hélices et pourraient donc avoir un rôle stabilisateur de la structure. Certains des résidus Ψ du domaine IV et du domaine V de l'ARNr 23S sont des résidus très conservés chez les eucaryotes. Il est intéressant de noter que deux des résidus Ψ expérimentalement détectés dans l'ARNr 23S ne sont attribués à aucun sRNA H/ACA et doivent donc être synthétisés par des systèmes enzymatiques ne nécessitant pas un sRNA H/ACA. Nous avons montré que la formation de l'un des deux résidus Ψ orphelins peut être catalysée par le complexe aCBF5/aNOP10/aGAR1, comme c'est le cas pour le résidu Ψ 55 des ARNt. La formation de l'ensemble des résidus Ψ des ARNr d'archaea dépend donc à la fois d'ARN: Ψ -synthases nécessitant un ARN guide et d'ARN: Ψ -synthases n'en nécessitant pas.

Enfin, l'examen des duplex sRNA H/ACA-ARNr substrats fonctionnels nous a permis de définir des règles de spécificité de reconnaissance des ARNr cibles par les sRNA H/ACA. Nous avons mis en évidence des contraintes structurales au niveau de la distance entre la structure en K-turn et le résidu U cible et entre la boîte ACA et le résidu U cible. De plus, nous avons montré que la règle de la présence de 2 résidus 5'-UN-3' entre les 2 appariements n'est pas stricte. La présence de 3 résidus 5'-UNN-3' ne modifie pas l'activité. Enfin, les données montrent que la structure de l'appariement formé avec la séquence guide 3' est nettement plus importante que celle de l'appariement formé avec la séquence guide 5'.

L'ensemble de ces données a été rédigé sous la forme d'un article sur le point d'être soumis au journal Nucleic Acids Research, dont je suis premier auteur.

Publication N°2 :

Combined *in silico* and experimental analyses of the *P. abyssi* H/ACA sRNAs and their target sites reveal the implication of RNA guided and non guided RNA:Ψ-synthase activities for Ψ residues formation in ribosomal RNAs

Sébastien Muller, Fabrice Leclerc, Isabelle Behm-Ansmant, Jean-Baptiste Fourmann, Fujihiko Matsunaga, Bruno Charpentier and Christiane Branlant.

Préparé pour soumission à **Nucleic Acids Res.**

PUBLICATION N°2

Combined *in silico* and experimental analyses of the *P. abyssi* H/ACA sRNAs and their target sites reveal the implication of RNA guided and non guided RNA:Ψ-synthase activities for Ψ residues formation in ribosomal RNAs

Sébastien Muller¹, Fabrice Leclerc¹, Isabelle Behm-Ansmant¹, Jean-Baptiste Fourmann¹, Fujihiko Matsunaga^{2,3}, Bruno Charpentier¹ and Christiane Branlant^{1*}.

1. Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP, Faculté des Sciences et Techniques, Université Henri Poincaré Nancy I, P.O. box 239, 54506 Vandoeuvre-lès-Nancy, France

2. Institut de Génétique et Microbiologie, URA CNRS 1354, Université Paris Sud, 91405 Orsay, France

3. Department of Genetic Resources Technology, Faculty of Agriculture, Kyushu University, Fukuoka 812-8581, Japan

* corresponding author

SUMMARY

Here, in the view to increase our knowledge on the RNA pseudouridylation catalysts in archaea, we performed an extended computational search of H/ACA sRNAs in three *Pyrococcus* species (*P. abyssi*, *P. furiosus* and *P. horikoshii*). Then, we predicted their putative target sites in rRNAs by computational analysis and we investigated the presence of the predicted Ψ residues in the *P. abyssi* rRNAs by using the RT-CMCT approach. Finally, the complete demonstration of the link between the detected Ψ residues in rRNAs and the identified H/ACA sRNAs was obtained by testing the *in vitro* activity of reconstituted H/ACA sRNPs. The data obtained demonstrate the presence of a large number of Ψ residues in the pyrococcal rRNAs, at least 17 residues (4 in 16S rRNA and 13 in 23S rRNA). Most of these residues are located in rRNA segments of high functional importance (central pseudoknot of 16S rRNA, A and P sites, PTC center). Some of them are highly conserved in eukarya suggesting an ancient origin of these conserved modifications. The data obtained also demonstrate that the formation of two of the identified Ψ residues in 23S rRNA is not catalyzed by the H/ACA sRNAs. Surprisingly, we showed that one of them can be formed by the aCBF5/aNOP10/aGAR1 complex in the absence of guide sRNA. Finally, our results bring important lights on the rules for H/ACA sRNA specificity and activity: we found that the length of helix H2 (distance between the K-turn and the targeted U residue) is of 5 or 6 bps, the distance between the conserved ANA motif and the targeted U residue is of 14 or 15 nts. In contrast to these rigid constraints, we found that the stability of the base-pair interaction formed by the 3' guide sequence and the substrate rRNA is highly important, but not that formed with the 5' guide sequence. In addition, the rule of the presence of a UN single-stranded dinucleotide between the two intermolecular interactions in archaea is not strict, since we showed the functionality of an interaction with a single-stranded UNN trinucleotide.

INTRODUCTION

Conversion of uridine into pseudouridine (Ψ) residues is one of the most abundant post-transcriptional modifications of tRNAs, rRNAs and UsnRNAs 1. Compared to U residues, Ψ residues have the advantage to form an additional hydrogen bond at the N1–H position. In addition, the carbon-carbon link between the ribose and the base in Ψ residues limits the flexibility of the ribose backbone, which favors and stabilizes base-pair interactions (2, 3, 4, for review 5). A role of Ψ residues in stabilization of tRNA 3D structures has been well documented (6), and the functional importance of Ψ residues in U2 snRNA for the activity of splicing complexes was demonstrated (7, 8, 9, 10). Compared to bacterial rRNAs, the eukaryal rRNAs contain a large number of Ψ residues and in both domains of life Ψ residues are concentrated in areas of the rRNAs expected to play important functional roles, in particular domains IV and V, which are directly involved in the peptidyl transferase activity (11, 12, 13). Taken individually, pseudouridylations in rRNAs are not essential for cell growth. However, complete abolition of Ψ formation in rRNAs is deleterious for ribosome assembly and activity (14, 12, 15, 16, 17, 18, 19, 20). Recent data suggest a possible involvement of Ψ residues in rRNAs in : i) subunit interaction (14, 20), ii) the translocation step (17, 20) and iii) the folding of 23S rRNA in an active form at the peptidyl transferase centre (PTC) (15). The strong increase of pseudouridylation sites in eukarya compared to bacteria is explained by the use of different catalysts: stand-alone enzymes carrying both the RNA recognition capacity and the catalytic activity are used in bacteria (21, 22, 23), whereas U to Ψ conversions are catalysed by H/ACA snoRNPs in eukarya (for review, 24). The H/ACA snoRNPs contain 4 proteins and an H/ACA snoRNA that defines the rRNA targeted residues. Recent data revealed a similar RNA-guided system in archaea (25, 26, 27, 28). The U residue to be modified is defined by formation of base-pair interactions between the H/ACA sRNA and the rRNA (for review, 24, 29, 27, 28).

Most of the eukaryal H/ACA snoRNAs contain two characteristic stem-loop structures, with an internal loop (pseudouridylation pocket) complementary to two nucleotide stretches that border the targeted U residue. We will denote these stem-loop structures H/ACA SLS. In this intermolecular interaction which is formed between the H/ACA SLS and the rRNA target, the targeted U residue and its 5' nucleotide are single-stranded. The archaeal sRNAs counterparts of the snoRNAs have more diverse architectures. They are composed of one, two or three H/ACA SLS (25, 26, 28, 27, 30). In contrast to the eukaryal snoRNAs, the sRNAs have a K-turn or a K-loop motif in each H/ACA SLS, which binds protein L7Ae. Each of the H/ACA SLS is followed by an ANA trinucleotide, which is most frequently an ACA trinucleotide. The H/ACA SLS followed by the ANA trinucleotide will be denoted an H/ACA motif. As successful reconstitution of active H/ACA sRNPs were achieved, using an *in vitro* transcribed H/ACA sRNA and the recombinant archaeal H/ACA sRNP proteins (28, 27), strong progresses were recently made in the understanding of the H/ACA sRNP structure and function. Like the eukaryal H/ACA snoRNPs, the archaeal H/ACA sRNPs contain 4 proteins: aCBF5 is the RNA:Ψ-synthase, aNOP10 is required for aCBF5 activity, L7Ae binds the K-turn or K-loop motif, and aGAR1 may facilitate the H/ACA sRNP turnover (31, 32, 33, 27, 28, 34, 26, 35). Both L7Ae and aGAR1 strongly reinforces the sRNP activity (27). The 3D structure of protein complexes and of a complex formed by an H/ACA sRNA and the 4 proteins was solved by X-ray crystallography at high resolution (36). The ANA sequence 3' to the H/ACA SLS is needed for binding of aCBF5 to the guide RNA. In the crystal structure, aCBF5 also interacts with the 5' strand of the pseudouridylation pocket, helix H1 and helix H2 of the H/ACA SLS. However, up to now, no 3D structure was obtained for an H/ACA sRNP bound to its RNA substrate. Therefore, little is known on the structural requirements imposed by formation of the RNA target-H/ACA sRNA duplex and its impact on the sRNP activity. The NMR structure established for a complex formed between an H/ACA SLS and a small target RNA revealed the capability of the two RNAs to interact together at the high concentration used for NMR analysis

(37, 38). In this structure, the heterologous helix formed by hybridization of the RNA target with the 3' guide element of the sRNA is stacked on helix H1, while the heterologous helix formed by the hybridization of the RNA target with the 5' guide element of the sRNA is stacked on helix H2. In contrast to this extended knowledge on the catalyst, at present, little is known on the presence of Ψ residues in archaeal rRNAs. The presence of Ψ residues has only been investigated in *Halobacterium halobium* (39), *Haloarcula marismortui* (19, 18), *Sulfolobus acidocaldarius* (40) and *Archaeoglobus fulgidus* (25). Unfortunately, the search for putative H/ACA sRNA genes by computational analysis was made for other species: *Pyrococcus furiosus* (41, 26, 28), *Methanococcus jannaschii* (42), *Thermococcus kodakarensis* (30) and *Sulfolobus solfataricus* (43). Experimental RNomics approach was only applied to *A. fulgidus* and 3 H/ACA sRNAs were detected by this method (25). For a better definition of the rules that govern the H/ACA sRNP specificities and efficiencies, it was of high importance to make for one given species: i) the identification of all putative H/ACA sRNAs, ii) an experimental analysis of the presence of Ψ residues at the predicted sites in rRNAs, and then, iii) to try to assign the detected Ψ residues to the detected H/ACA sRNAs.

To this end, here we combined different approaches in order to define experimentally the targets sites of all the putative H/ACA sRNAs predicted by computational analysis for a hyperthermophile archaeal species from the *Pyrococcus* order, *P. abyssi*. First we developed and used various computational approaches to identify all the putative H/ACA sRNAs of this species and their putative target sites in rRNAs. Secondly we applied the CMCT approach in order to test the presence of Ψ residues at the predicted sites in 16S and 23S rRNAs, and, finally, we attributed each of the identified Ψ residues to a given H/ACA sRNA by testing the activity of reconstituted H/ACA sRNPs on small rRNA fragments expected to contain a targeted U residue. Based on the obtained data, we brought new lights on the rules involved in the H/ACA sRNA-rRNA recognition and on structural constraints of the H/ACA sRNAs.

MATERIALS AND METHODS

Computational method

Genomes sequences and annotations

The sequences and annotations of the 11 archaeal genomes available at the beginning of the study, *Aeropyrum pernix K1* (44), *Archaeoglobus fulgidus DSM4304* (45), *Halobacterium sp.NRC1* (46), *Methanococcus jannashii DSM2661* (47), *Methanobacterium thermoautotrophicum deltaH* (48), *Pyrococcus abyssi GE5* (Heilig and Genoscope, 1999, unpublished), *Pyrococcus horikoshii OT3* (49), *Pyrococcus furiosus DSM3638* (50), *Sulfolobus solfataricus P2* (51), *Thermoplasma acidophilum DSM1728* (52), and *Thermoplasma volcanium GSSI* (53) were downloaded in Fasta and Genbank formats from NCBI ftp site : <ftp://ftp.ncbi.nih.gov/genomes/Bacteria>.

Listing Inter-Coding-Regions (ICR)

The positions of the coding regions which include ORF, rRNA, tRNA, RNaseP and 7S RNA were listed for the 11 archaeal genomes available at the beginning of the study. Identified C/D box sRNA positions were consciously not included. ICR sequences were then extracted from the genomes sequences using ExtractICR, a script written in awk and perl languages that automatically sorts out the ICR sequences. ExtractICR reads in the table of ORF positions and genomes sequences and exports ICR sequences longer than 15 nucleotides. The exported ICR sequences also include the 15 nucleotides of the downstream and upstream flanking ORF (for convenience see results). ICR sequences were afterwards formatted and assembled, using Readseq (<ftp://iubio.bio.indiana.edu/molbio/readseq>) and Formatdb (NCBI toolkit), in 4 databases. One of the databases lists all the 11 archaeal genomes ICR sequences while the three others are specific to each *Pyrococcus* species.

Search for sequence similarity between ICR sequences

The databases sequences were compared to each other using Multiblast, a script written in awk and perl languages that enables an iterative comparison of sequences in databases via Blast (54). Defaults parameters were set for Blastn. Significant similarities were assumed to be shown with a value of $e < 0.01$. Three kinds of information were obtained. The first comparisons were run on each database independently in order to remove all the repeated elements. Sequences with a size greater than 50 nucleotides, which presented a significant sequence similarity with other sequences from the same genome several times were thought to be repeat elements hence unlikely to harbour any H/ACA sRNA genes. Then a second comparison was done between the *Pyrococcus* databases and archaeal database. We considered that a sequence with a size greater than 30 nucleotides present once in a *Pyrococcus* species but similar to several sequences present in another archaeal genome ICR sequences was not usable for comparative genomics because likely to correspond to some repeat elements or low complexity sequences in other archaeal genomes. The last comparison was run on the *Pyrococcus* species database against the other two *Pyrococcus* databases. Only ICR sequences from one *Pyrococcus* presenting significant sequence similarity, with a size greater than 18 nucleotides, with one ICR sequence from each of the two other *Pyrococcus* were selected. ICR which showed similarities only at the beginning or at the end of their sequences, that means in the part of the sequence belonging to adjacent ORFs, were not selected. Groups of three ICR sequences showing similarities were then aligned using Clustal-W (55), a global alignment program. Only ICR, called conserved ICR, presenting similarities extended to 30 nucleotides displaying more than 70% sequence similarity were selected.

Filtering out likely ORFs

Conserved ICR sequences were screened for some ORF. GeneMarkS (56) is an ORF finder which calculates the probability for sequences to be ORF. GeneMarkS was not initially used to

annotate *Pyrococcus* genomes. GeneMarkS was run on the three whole *Pyrococcus* genomes, and ICR sequences that might be ORF were filtered out.

Filtering out C/D box sRNA genes

The remaining conserved ICR sequences were afterwards screened for some C/D box sRNA genes. RNAMOT (57) is a pattern descriptor that enables such a search. RNAMOT was run on the conserved ICR sequences and the results were corroborated by some recent works (58, 59). The ICR sequences containing C/D box sRNA which, alone, explained the similarities between the three species, were filtered out.

Searching for H/ACA box sRNA genes and their targets

RNAMOT (57) was used for an initial search. As it only allows a search with low constraints, RNAMOT was only run on the conserved ICR sequences. We looked for a hairpin comprising two helices called H1 and H2 helices flanking a 5-11 nts long internal loop. This hairpin has to display an 8-35 nts long apical loop and has to be tailed by an ACA box (ANA). For a search with higher constraints, we ran the ERPIN software (60) on the entire *Pyrococcus* genomes. ERPIN builds helix and single strand lod-score profiles from sequences alignment and screens genomes sequences for occurrences of these profiles. An alignment of H/ACA box sRNAs based on the previous determinants was given as an input, and the output results were added to enrich and modulate this alignment. For each candidate, at least one target, *i.e.*, a uridine residue flanked by two sequences complementary to the bipartite guide sequences of the internal loop, was predicted using RNAMOT (30).

Experimental method

***P. abyssi* strain GE 5 cultures**

P. abyssi strain GE 5 cells were grown as described (61) at 95°C in Vent Sulfothermophiles Medium (20 g/l NaCl, 0.25 g/l KCl, 0.05 g/l NaBr, 0.5 g/l SrCl₂-6H₂O, 0.08 g/l boric acid, 3 g/l PIPES, 1 g/l yeast extracts, 4 g/l peptone, 0.1% Resazurine, 20% MgSO₄, 5% CaCl₂, 5% KH₂PO₄, pH=6.8).

Total RNA isolation

P. abyssi cells growth was stopped at the end of the exponential phase. The cells were centrifuged and then washed with washing buffer (1M sorbitol, 25 mM Hepes, pH=7.0). The cells were frozen and stocked at -80°C. Total RNA isolation was performed using the method developed by Chomczynski and Sacchi (1986) (62) which proved to be efficient with *Pyrococcus* cells (63). About 10¹⁰ cells were dissolved by adding 4 ml of solution D (4 M guanidinium thiocyanate, 25 mM sodium citrate pH 7.0, 0.5% sarcosyl, 0.1 M β mercaptoethanol). A phenol/chloroform extraction was performed afterwards. Total RNAs were then precipitated using sodium acetate 0.3 M in ethanol, and quantified with a Genequant (Pharmacia).

Northern Blotting analysis

10 µg of *P. abyssi* total RNA and 5' end-labelled size standards 100 bp DNA ladders were fractionated on 6% denaturing polyacrylamide gels (8 M urea, 0.5x TBE buffer) and electrotransferred to a nylon membrane (Hybond-N+, Amersham). The RNA was immobilized on the membrane by UV crosslink. Pre-hybridization was carried out at 58°C in the pre-hybridization buffer (SSPE [52.6 g/l NaCl, 8.3 g/l Na₂HPO₄-2H₂O, 2.2 g/l EDTA, pH=7.4], Denhardt's [1 g/l Ficoll, 1 g/l polyvinylpyrrolidone, 1 g/l BSA], 0.5% SDS) for 1 h. Oligonucleotides complementary to RNA sequences (Table 1) were 5' end-labelled with [γ -

³²P]ATP and T4 polynucleotide kinase (PNK) for 1h at 37°C. Hybridization was carried out at 58°C for 16 h by adding 100 ng of labelled oligonucleotide. Membranes were washed four times in SSPE at 42°C for 5 min and were visualized on a Typhoon 9410 (Amersham Biosciences).

Mapping of pseudouridine (Ψ) residues in the *P. abyssi* rRNA

The *N*-cyclohexyl-*N'*-(2-morpholinoethyl)-carbodiimide metho-*p*-toluolsulfonate (CMCT) modification protocol was adapted from Bakin and Ofengand (1993) (64), with the previously described modifications (65). The modification by CMCT was performed with 5 µg of *P. abyssi* total RNA. Positions of CMCT modifications were identified by primer extension analysis, using the AMV RT (QBiogene, USA). Primers were 5' end-labelled using [γ-³²P]ATP (3000 Ci/mmol) and T4 PNK. RNA sequencing was done with 20 µg of *P. abyssi* total RNA.

Recombinant protein production

The ORFs of archaeal L7Ae, aNOP10, aGAR1, and aCBF5 proteins, (GenBank accession nos. C75109, CAB49761, CAB49230, and CAB49444, respectively) were PCR amplified from the genomic DNA of the *Pyrococcus abyssi* GE5 strain. They were cloned in plasmid pGEX-6P-1 (Pharmacia) to produce GST fusion protein. The recombinant GST-L7Ae, GST-aNOP10, GST-aGAR1, GST-aCBF5 proteins were produced in *E. coli* BL21 CodonPlus cells (Novagen) and purified under native conditions using Glutathione-Sepharose 4B (Pharmacia) as previously described (27). The quality of the protein preparations was assessed by SDS-PAGE and Coomassie staining.

DNA templates used for *in vitro* transcription

The Pab21, Pab35, Pab40, Pab91, Pab105 and Pab160 sRNA sequences were PCR amplified from the *Pyrococcus abyssi* GE5 genomic DNA using a forward primer containing the T7 RNA

polymerase promoter. Before use, all the PCR amplification products were cloned in the pTAdv vector (Clontech) and their sequences were checked by dideoxynucleotide sequencing.

***In vitro* transcription of guide and target RNAs**

The 7 H/ACA sRNAs, and the target RNAs, were transcribed *in vitro* by T7 RNA polymerase using PCR amplification fragments. They were purified by gel electrophoresis. For production of cold RNAs, transcription reactions were carried out as previously described (66). The same conditions were used for the synthesis of uniformly labeled RNAs, except that the reactions were performed with 20 μ Ci of [α -³²P] NTP to be incorporated (800 Ci/mmol) (Amersham), 0.13 mM of the same NTP and 4 mM each of the three other NTP. Subsequent RNA treatments were described as previously (67).

***In vitro* sRNP guide pseudouridylation assay**

Unlabeled sRNA (4 pmoles) and [α -³²P] NTP-labeled targets (150 fmoles) were mixed at room temperature in buffer D (150 mM KCl; 1.5 mM MgCl₂; 0.2 mM EDTA; 20 mM HEPES, pH 7.9) and each recombinant protein was added at a 200 nM concentration. The mixture (4.5 μ l final volume in buffer D) was incubated at 65°C for 80 min. The reaction was stopped by the addition of 150 μ l of H₂O and 150 μ l of a phenol:chloroform (1:1) mix. The extracted RNAs were ethanol-precipitated. The modified RNA targets were totally digested with 0.5 unit of T2 RNase (GibcoBRL) or P1 RNase (Boehringer) at 37°C for 12 h. The resulting 3'-mononucleotides were chromatographed on thin-layer cellulose plates following two different procedures (N1-N2 or N1-R2) as described elsewhere (68, 69). For the reactions performed with P1 RNase digestions, one or two dimension chromatography was performed to fractionate the labeled 5'-mononucleotides. The radioactivity in the spots or the bands were quantified with a phosphorimager (Typhoon 9410, Amersham Biosciences) using the ImageQuant software. When digested by RNaseT2 we took into account the total number of residues that precede the [α -³²P]

NTP in the target RNA, and thereby assessed the amount of Ψ formed (expressed in mol per mol of U in the target RNA). When digested by RNaseP1 we took only into account the total number of U residues in the target RNA.

RESULTS

Identification of 7 putative H/ACA sRNAs in *P. abyssi*

When we started this study, only a limited number of archaeal H/ACA sRNAs were identified. Therefore, to look for H/ACA sRNA genes in *P. abyssi*, we first used a comparative analysis of the genomic sequences linking the ORFs and known RNA genes in three *Pyrococcus* species (*P. abyssi*, *P. furiosus* and *P. horikoshii*). The strategy used is described in Figure 1 and in Material and Methods. The idea was that DNA sequences corresponding to non-coding RNAs of functional importance should be highly conserved in the three genomes as compared to their bordering sequences within the ORF linking regions. By using this approach and taking into account the characteristic features of previously identified H/ACA sRNAs in *A. fulgidus* (25), we detected five putative H/ACA sRNA genes, namely sRNAs Pab21, Pab35, Pab105, Pab40 and Pab91 (Fig. 2). The sRNAs Pab21, Pab105, Pab35 and Pab40 are the counterparts of the sRNAs Pf1, Pf3, Pf6 and Pf7 found in *P. furiosus*, respectively (26, 28). The sRNAs Pab21, Pab105, Pab35, Pab40 and Pab91 are the counterparts of the sRNAs Tko8 (RC1), Tko9 (RC2-3), Tko10 (RC6-7), Tko1 (FW1-3) and Tko6 (FW5) found in *T. kodakarensis*, respectively (30). The expression of the five putative H/ACA sRNAs detected was verified by Northern blot analysis (Fig. 3) and a secondary structure model was built for each of them (Fig. 2). Note that base substitutions in *P. furiosus*, *P. horikoshii* or *T. kodakarensis* are either located in single-stranded regions or compensate each other to preserve the 2D structure. Therefore, we used the 2D structures established for each of the *P. abyssi* H/ACA SLS expected to contain a pseudouridylation pocket in order to build an RNA profile to be used in a computational search

for H/ACA sRNA genes by the ERPIN software 60. As recently described (30), this computational search can be done on the complete *P. abyssi* genomic sequence. An additional putative H/ACA sRNA was found by this procedure, the sRNA Pab160 (Fig. 2), which is the counterpart of sRNAs Pf9 and Tko4 (FW4), found in *P. furiosus* (28) and in *T. kodakarensis*, respectively (30). Therefore, by using as the constraints, a minimum length of helix H1 fixed to 7 bps and a minimum length of helix H2 fixed to 5 bps, and a length of each guiding sequences comprised between 6 and 12 nts, we found six putative H/ACA sRNAs in *P. abyssi* as previously found for *T. kodakarensis* (30). However, we reasoned that by using this procedure some H/ACA sRNAs may have escape detection because of a slightly different conformation of their H/ACA SLS. In particular, we tested whether some H/ACA sRNAs may have extended pseudouridylation pockets, as is the case for the internal loop of the sRNA Pab40 (Fig. 2). To this end, in the ERPIN search we extended the limits of the sizes of the 5' and 3' guide sequences (up to 19 nts). Using these less stringent conditions in the ERPIN search, we found a seventh putative H/ACA sRNA in *P. abyssi*, the sRNA Pab19. This sRNA can be folded into a small 5' stem-loop structure followed by a putative H/ACA SLS containing a K-turn structure, and a large putative pseudouridylation pocket. It is followed by an AUA 3' terminal trinucleotide. The lengths of stems H1 and H2 are canonical, 7 and 5 nts respectively. However, the length of the 5' guide sequence is longer than usual (15 nts) (Fig. 2). We found counterparts of the putative Pab19 sRNA in *P. furiosus*, *P. horikoshii* and *T. kodakarensis* (Fig. 2). Therefore, taken together, our computational and northern blot analyses suggested the presence of 7 H/ACA sRNAs in *P. abyssi*.

To predict the possible target sites in the *P. abyssi* rRNAs of the 7 putative H/ACA sRNAs identified, we used the method based on the utilization of the RNAMOT software that we described recently (30). All the predicted target sites are given in Table I and the corresponding base-pairs interactions are shown in Figure 4. One, up to 4, distinct target sites were predicted for each putative pseudouridylation pockets. Taken together, fifteen target sites were predicted

in 23S rRNA, and 8 in 16S rRNA. Among the overall 23 predicted sites, 8 were previously predicted for the *P. furiosus* H/ACA sRNAs (26). Note that the intermolecular interaction proposed for the Pab19 sRNA leaves three single-stranded residues 5'-UGC-3' instead of the usual 5'-UN-3' dinucleotide between the two intermolecular interactions (Fig. 4G).

Experimental search for Ψ residues at the predicted target sites in rRNAs

We used the CMCT-RT approach (65) in order to test whether Ψ residues are indeed present at the positions predicted to be targeted by the identified H/ACA sRNAs. To this end, total RNA was prepared from *P. abyssi* cellular extracts and treated with CMCT with or without alkaline treatment. Positions of the alkaline-resistant Ψ residues were detected by primer extension analyses using a large series of oligonucleotide primers. The 16S and 23S rRNA segments that were probed by these primer extension analyses represent 20% and 27% of the 16S and 23S rRNAs, respectively and in particular most of the domains IV and V sequences in 23S rRNA. These two domains are known to contain a large number of post-transcriptional modifications and are located at the peptidyl transferase center (for review, 12). By this approach, we identified 3 and 11 Ψ residues in 16S and 23S rRNAs, respectively, 8 of them are located in domain V of 23S rRNA.

The experimental analysis confirmed Ψ formation at 12 of the predicted positions (positions 27, 891, 922 in 16S rRNA, and positions 1932, 2549, 2554, 2588, 2672, 2685, 2697, 2794 and 2930 in 23S rRNA) (Table 1 and Fig. 5). In addition, the absence of Ψ formation was clearly demonstrated at 6 of the predicted positions (positions 892, 995 and 1122 in 16S rRNA and 278, 2250 and 2552) (Table 1 and Fig. 5). For three of the predicted positions, the reverse transcriptase analysis was obscured by the presence of an RT pause at the level of the stop expected after CMCT modification (positions 1017 in 16S rRNA, 1377 and 2016 in 23S rRNA). Therefore, we could not determine whether U to Ψ conversion occurred at these positions.

Interestingly, for one of these positions (2016 in 23S rRNA) formation of a Ψ residue was detected at the corresponding position in two archaeal species: *H. halobium* (39) and *H. marismortui* (18, 19), and in nearly all the organisms which were studied up to now (13). Therefore, modification at position 2016 in 23S rRNA may also exist in the *P. abyssi* 23S rRNA and could be directed by the Pab40-3 H/ACA motif (Fig. 4). No information was available in other species corresponding to the positions 1017 in 16S rRNA and 1377 in 23S rRNA.

Interestingly, in the course of this analysis, Ψ residues were found to be formed at two positions that were not predicted to be target sites of the identified H/ACA sRNAs (positions 2585 and 2603 in 23S rRNA). Even by relaxing the constraints in the ERPIN search for putative H/ACA sRNA genes, we could not detect H/ACA sRNAs corresponding to these positions. We concluded that they may be formed by catalysts that do not involve H/ACA sRNAs.

The 7 H/ACA sRNPs are active at 15 target sites *in vitro*

To test experimentally the relationships that we established between the identified Ψ residues in the *P. abyssi* rRNAs and the identified H/ACA sRNAs, we used the sRNP *in vitro* reconstitution approach developed in the laboratory (27). To this end, each of the 7 identified H/ACA sRNAs was transcribed. The four H/ACA sRNP proteins aCBF5, aNOP10, aGAR1 and L7Ae from *P. abyssi* were produced as recombinant proteins in *E. coli* and purified. Assembly of individual proteins and different combinations of these proteins on the H/ACA sRNAs was tested by electrophoresis-mobility shift assay (EMSA). Incubation was performed as previously described at 65°C (27, 66). sRNPs could be assembled for each H/ACA sRNA tested. Interestingly, we observed a higher affinity of aCBF5 for the Pab21 sRNA compared to the previously studied Pab91 sRNA. In addition, protein aNOP10 was also able to bind to this RNA (data not shown). Such binding was not observed for any other of the 5 H/ACA sRNA tested.

Then, we measured the activities of the reconstituted sRNPs on their predicted target sites in rRNAs. To this end, we used the adaptation of the nearest-neighbour method that we developed for H/ACA sRNPs (27, 66). For each of the 23 predicted target sites in rRNAs, we produced a 20-nt long RNA substrates with the targeted U residue located at the center of the molecule. When T2 RNase digestion was used, these RNA substrates were *in vitro* transcribed in the presence of [α -³²P]ATP, CTP, GTP or UTP depending on the identity of the residue located 5' to the targeted U residue. When P1 RNase digestion was used, transcription was done in the presence of [α -³²P]UTP. For each RNA substrate incubation was performed at 65°C for 80 min, with the four core proteins (LCNG) in the presence or the absence of the H/ACA sRNA. The conditions used were previously described (27). Then, RNAs were extracted, digested by RNase T2 or RNase P1 and fractionated by thin layer chromatography (using one of the two procedures described in Material and Methods) (Fig. 6).

Only 15 of the 23 predicted substrates were found to be modified by the reconstituted sRNPs and rates of U to Ψ conversion were ranging from 30 to 100% (Fig. 6 and Table 1). The data obtained confirmed that each of the 11 H/ACA motifs is capable to guide pseudouridylation. Interestingly, the Pab19 H/ACA motif which has a large internal loop and forms an intermolecular interaction with three single-stranded residues is active. Deletion of the additional single stranded C residue, which is located 3' to the 5'-UN-3' dinucleotide (mutant S-C1019 Δ), Fig. 4G and 6D), showed that sRNA Pab19 can target the WT and mutated RNA substrates at very similar rates. Therefore, we concluded that requirement of two single-stranded residues located between the two intermolecular interactions is not a strict rule in archaeal H/ACA sRNAs. Interestingly also, among the two possible conformations of the second H/ACA motif in sRNA Pab40, only the conformation with a small stem-loop structure in the pseudouridylation pocket is active *in vitro*. However, the yield of U to Ψ conversion is lower for this motif as compared to that found for the other motifs, probably because of a reduced kinetics of assembly of the active sRNP.

Interestingly, some of the H/ACA motifs can guide modification at 2 or even 3 positions. For instance, the H/ACA motif 1 in sRNA Pab35 can guide U to Ψ conversion at two positions in 23S rRNAs (2672 and 2930). The H/ACA motif 2 in this sRNA can also guide modification at two positions in 23S rRNA (2549 and 2697). Therefore, altogether, the 2 H/ACA motifs of sRNA Pab35 can guide modification at 4 positions in 23S rRNA. While the H/ACA motif 1 in sRNA Pab105 guides only one modification at position 2554 in 23S rRNA, the second H/ACA motif of this sRNA can guide modification at two positions in 23S rRNA (2794 and 1377) and at one position in 16S rRNA (27). Therefore, altogether the sRNA Pab105 can also guide modification at 4 positions in rRNAs. In contrast, we found that all the H/ACA sRNAs that contain a single motif (Pab21, Pab91, Pab160 and Pab19) guide a unique position in rRNAs. This is also the case for each of the H/ACA motifs present in the sRNA Pab40. Altogether, the 11 H/ACA motifs of the 7 *P. abyssi* sRNAs can guide U to Ψ conversion at 15 sites in the *P. abyssi* rRNAs (Table 1).

Only 12 of these sites were clearly demonstrated to be modified *in vivo* by the CMCT approach (Fig. 5 and Table 1). As mentioned above, the experimental detection failed at the three other positions because of reverse transcriptase pauses, which were masking the expected CMCT stops. Based on the strong *in vitro* activity found at these three rRNA positions, namely 1017 in 16S rRNA and 1377 and 2016 in 23S rRNA, we concluded that these three positions are probably modified *in vivo*. As mentioned above, this is as much likely for position 2016 in 23S rRNA as this position is almost universally modified. In spite of the prediction that modifications at positions 922 in 16S rRNA and 2672 in 23S rRNA, can be guided by two distinct H/ACA motifs, our experimental data revealed that this is not the case. Therefore, modification at one given position is likely guided by only one H/ACA motif.

Constraints on the H/ACA sRNA structure and H/ACA sRNA target interaction

Based on the data of our reconstitution experiments, we could try to define rules concerning both the H/ACA sRNA structure and the base-pair interactions between the H/ACA sRNA and the RNA substrate. In several of the predicted interactions, that turned to be non functional, the length of helix H2, corresponding to the distance between the pseudouridylation pocket and the K-turn structure, was too short (4bps) or too long (7bps). The most frequent length for helix H2 was of 6 bps (14 of the 15 active interactions). Some of these non-functional interactions also had too long or too short distances between the ACA trinucleotide and the targeted U residue (17 or 13 nts). The most frequent distance found in the active interactions were of 15 or 14 nts (9 and 4 of the active interactions, respectively). A strict conservation of the distance between, on the one hand the K-turn stem and, on the other hand the ACA trinucleotide, with the targeted U residue can be explained by steric hindrances between proteins of the sRNPs and the needed interactions between these proteins as discussed below. Another important information obtained from our data is the strong importance of the base-pair interaction that is established with the 3'-guide sequence of the pseudouridylation pocket compared to that established with the 5'-guide sequence. This is exemplified by the efficient modification at positions 2672 and 2549 in 23S rRNA, which are guided by the H/ACA motifs 1 and 2 of sRNA Pab35, respectively (Fig. 4). Indeed this strong efficiency contrasts with the poor base-pair interaction formed between the 5' strand of the pseudouridylation pocket and the rRNA substrate. Furthermore, on the contrary, in spite of a canonical length of helix H1 (6 bps) and a canonical distance between the ACA trinucleotide and the target U residue (15 nts), the motif 2 of sRNA Pab105 did not guide modification at position 995 in 16S rRNA, probably because of the low stability of the interaction formed between the 3'-guide sequence of the sRNA and the rRNA substrate (4 bps including two G.U pairs). Therefore, we concluded that the specificity of the H/ACA sRNA guiding activity is ensured by structural rules, which concern the length of helix H1, the distance of the ACA triplet with the targeted residue and the stability of the base-pair interaction between

the 3' guide sequence and the rRNA substrate. In contrast, unexpectedly, the presence of the 5'-UN-3' single-stranded dinucleotide between the two intermolecular interactions is not a strict rule, since a 5'-UNN-3' trinucleotide can lead to a functional interaction.

One of the two orphan Ψ residues in 23S rRNA can be formed without guide sRNA

For two of the Ψ residues that we detected experimentally in 23S rRNA positions 2585 and 2603 respectively, we did not find putative guide H/ACA sRNAs. Interestingly, we observed that the sequence containing the modified residue at position 2603 can be folded into a stem-loop structure showing some similarity with the T Ψ C stem-loop structure of tRNA (Fig. S1, Supplementary Data). As this tRNA stem-loop structure was found to be modified by aCBF5 in the absence of guide sRNA (70, 71), we tested whether the aCBF5/aNOP10/aGAR1 protein complex can catalyse U to Ψ conversion at position 2603 in the 23S rRNA by using a 20-nt long RNA substrate corresponding to the 23S rRNA portion from position 2595 to 2611. Interestingly, a significant yield of U to Ψ conversion (50%) was detected in this RNA substrate. According to the well established 2D structure of domain V in 23S rRNA (72, 73), this segment does not undergo the stem-loop structure shown in figure S1 in the entire 23S rRNA. However, the aCBF5/aNOP10/aGAR1 complex may modify position 2603 at a stage during 23S rRNA synthesis, when this conformation is formed. When similar experiments were performed with a 23S rRNA segment containing the second orphan Ψ residue (position 2585), we failed to detect Ψ formation. Therefore, the catalyst for pseudouridylation at this position remains to be identified.

DISCUSSION

A high number of Ψ residues in *P. abyssi* rRNAs

By using the CMCT method, we could map 14 Ψ residues in the *P. abyssi* rRNAs. In addition, 3 of the H/ACA motifs that we identified can act *in vitro* at 3 positions in rRNAs where RT pauses obscured the identification of Ψ residues. As, we analysed only 20% and 27% of the 16S and 23S rRNA sequences by the RT-CMCT approach, we cannot exclude the presence of a few additional Ψ residues in the non studied areas of these rRNAs, that may be catalysed by RNA: Ψ -synthases without the use of guide sRNA. Hence, taken all these information into account, we conclude that the 16S and 23S rRNAs from *P. abyssi* contain at least 4 and 13 Ψ residues, respectively. Up to now, this is the highest number of Ψ residues detected in archaeal rRNAs. Indeed, in the previous studies on the 23S rRNA from archaeal species, only 3, 4, 6 and 5 Ψ residues were found in the *H. marismortui*, *H. halobium*, *S. acidocaldarius* and *A. fulgidus* species, respectively (18, 19, 39, 40, 25). Interestingly, the two halophiles species, *H. marismortui* and *H. halobium* that contain the smallest number of Ψ residues are growing at 50 and 42°C, respectively. *S. acidocaldarius* and *A. fulgidus* grow at 80°C and 83°C, respectively, whereas *P. abyssi* where we detected the largest number of Ψ residues is growing at 98°C. Therefore, as already proposed for archaeal 2'-O-methylations (74) there may be a link between the number of Ψ residues present in archaeal rRNAs and the growth temperatures of the organisms. However, a confirmation of this statement requires further analysis of the number of Ψ residues in rRNAs of a larger number of archaeal species growing at different temperatures.

8 Ψ residues are present in the 23S rRNA region located at the PTC

Interestingly, 2 and 8 among the 13 Ψ residues detected in the *P. abyssi* 23S rRNA are located in the functional domains IV and V, respectively. One of the residues detected in domain IV (position 2016) corresponds to one of the two highly conserved Ψ residues found in the terminal

loop of stem-loop structure 69 (SLS69). SLS69 is known to contact both 16S rRNA and the tRNA bound at the acceptor site (A site) (20). In addition, SLS69, that is located at the subunit interface, was proposed to play a role in tRNA translocation (75, 76, for review, 12). Interestingly, in *S. cerevisiae*, formation of the two Ψ residues in the terminal loop of SLS69 was found to confer a growth advantage (16).

The 8 Ψ residues, which are concentrated in the part of domain V considered to be directly involved in the peptidyl-transferase catalytic center (PTC), are all involved in base-pair interactions (Fig. 7). Six residues were detected in the corresponding region of *S. cerevisiae* (77, 39). Three of them are conserved in *P. abyssi* (positions 2549, 2588 and 2603) (Fig. 7). None of these 3 conserved Ψ residues was found to play an essential role in yeast (15). However, due to their involvement in the RNA 2D structure the 8 Ψ residues found in the 23S rRNA PTC region may be very important in stabilizing the conformation of this 23S rRNA region at the high growth temperature of *P. abyssi*. Another reason for a high concentration of Ψ residues in this RNA region may be a cooperative effect of Ψ residues in stabilizing the RNA conformation, as previously proposed (13).

Nevertheless, only the absence of Ψ 2919 formation in the *S. cerevisiae* 25S rRNA region, which is located at the PTC, was found to have a strong effect on translation. This functional importance is probably related to the interaction taking place between this 23S rRNA region and the tRNA bound to the A site (78). Pseudouridylation at this position was, up to now, observed in Eucarya, not in Bacteria. We detected no Ψ residue at this position (2647) in the *P. abyssi* 23S rRNA, in spite of a U to Ψ conversion proposed at this position in the *A. fulgidus* 23S rRNA. The proposed guide H/ACA sRNA for this position in this species was the Afu4 sRNA. Like its counterpart Pab40 in *P. abyssi*, sRNA Afu4 contains 3 H/ACA motifs. Whereas the proposed target sites for motifs 1 and 3 of the Pab40 and Afu4 sRNAs are in perfect agreement, there is a difference for motif 2. It was proposed to guide modification at position 2601 in the A.

fulgidus 23S rRNA (corresponding to position 2647 in the *P. abyssi* 23S rRNA), whereas we demonstrated that motif 2 of Pab40 is active at position 1932 in 23S rRNA. The difference of specificity of motif 2 from these two species may be simply due to a point mutation (G to U substitution at position 85 as referred to the Pab40 numbering). This base substitution modifies the secondary structure of this H/ACA motif and therefore its guiding specificity. In addition, we should also mention, that the target site that was proposed for motif 2 of sRNA Afu4 is probably not the best target. Indeed, the distance between the ACA triplet and the targeted U residues is of 16 bps and we found that this is not a favourable distance. Another target site, which fits the rules that we established in this study can be proposed for this motif 2. It corresponds to position 2632 in the *A. fulgidus* 23S rRNAs, corresponding to position 2678 in the *P. abyssi* 23S rRNA. The U residue at this position is not converted into a Ψ residue in *P. abyssi*, but this is the case in *H. marismortui* (18, 19), in *H. halobium* (39) and in some eukaryote species (13). Altogether, these observations illustrate how a single point mutation can dramatically modify the specificity of an H/ACA sRNA.

We detected the formation of one Ψ residue in domain II. This domain plays an important role in binding of elongation factors and the presence of some conserved Ψ residues in domain II was already found in Bacteria and Eukarya. Finally we detected two residues in domain VI. Pseudouridine formation in this domain is specific to Eukarya. Noticeably, the pseudouridylation site at position 2930 in the *P. abyssi* 23S rRNA is conserved in mouse and human LSU rRNA (Fig. 7).

The 4 Ψ residues detected in 16S rRNA are located close to the A and P sites. The location of Ψ residues in the bacterial and eukaryal SSU rRNAs is rather variable from one species to the other. No Ψ residue was detected in the 16S rRNA from *H. volcanii*, whereas 5 Ψ residues were found in *S. solfataricus* (79). Here, we detected 4 residues in the *P. abyssi* 16S rRNA. Interestingly, two of them (positions 27 and 891) are located within or very close to the essential central pseudoknot of the 16S rRNA (80). They may therefore be important for its formation or

its stability. In addition, this central pseudoknot structure is located in close vicinity of the P site (81). Interestingly, the third Ψ residue detected at position 922 also belongs to a 16S rRNA segment located at the P-site. Finally residue U1017 belongs to a segment that was shown to be involved in site A formation, therefore, the 4 Ψ residues that we detected in the *P. abyssi* 16S rRNA are located in a region close or at the A and P sites.

Area of pseudouridylation but not sites of pseudouridylation are conserved in archaea

Although we observed a strong conservation of H/ACA sRNAs and their pseudouridylation sites in the three *Pyrococcus* species tested, as well as in the phylogenetically related *T. kodakarensis* species (30), conservation is poor with species of other genus. For instance, only one of the three Ψ positions found in the 23 rRNAs of *H. halobium* and in *H. marismortui* is common with the 17 Ψ positions that we found in *P. abyssi*, and only one of the 6 Ψ residues found in the *S. acidocaldarius* 23S rRNA is conserved in *P. abyssi* (Fig. 7). To reinforce this statement, we verified experimentally that in *P. abyssi* rRNAs, there was no additional Ψ formation, which may be independent of the detected H/ACA sRNAs and may occur at the positions where Ψ residues were detected in *H. halobium* and *H. marismortui* and in *S. acidocaldarius* (data not shown). Concerning *A. fulgidus* that belongs to a genus which is phylogenetically more closely related to the *Pyrococcus* genus than the *Sulfolobus*, *Halobacterium* and *Haloarcula* genus, 3 of the 6 Ψ positions detected in the *A. fulgidus* 23S rRNA are conserved in *P. abyssi* as well as one modified position in 16S rRNA. We verified experimentally the absence of modification in *P. abyssi* rRNAs at the 3 additional positions found in *A. fulgidus*. In addition, by using our computational approaches we verified that no other H/ACA sRNAs than the 3 described previously (25) are present in *A. fulgidus*. Therefore, the number of H/ACA sRNAs varies strongly from one archaeal genus to the other (from 3 in *A. fulgidus* to 7 in *P. abyssi*).

Interestingly, comparison of the pseudouridylation sites in *P. abyssi* rRNAs and in eukaryal rRNAs reveals the conservation of several sites, namely positions 2016 in domain IV, 2549,

2588 and 2603 in domain V, in the LSU rRNA (Fig. 7). The highly conserved Ψ residue at position 2698 is not found but one Ψ residue is present at its 5' position (2697). Therefore, the pseudouridylation sites conserved in eukaryotes are likely very ancient modification sites that may have appear early in evolution. Surprisingly, we showed that one of these conserved Ψ residue in domain V is not guided by an H/ACA sRNA in *P. abyssi*. Therefore, Ψ residues in archaeal rRNAs may be generated by both RNA guided and non guided RNA: Ψ -synthases.

The aCBF5/aNOP10/aGAR1 complex may act on rRNA without need for a guide sRNA

Whereas none of the 23 20-nt long rRNA substrates that we used as substrate were modified in the absence of the H/ACA sRNA (Fig. 6), when we used the 20-nt long RNA substrate containing residue U2603, half of this residue was converted into a Ψ residue, after a 80 min incubation, in the absence of guide RNA. Therefore we conclude that some stem-loop structures in 23S rRNA can be modified by the aCBF5/aNOP10/aGAR1 complex in the absence of guide sRNA as found for tRNA (70, 71). In spite of a shorter terminal loop, the stem-loop structure that can be formed by 23S rRNA around position 2603, shows some similarities with the T Ψ C stem-loop structure of tRNAs (Fig. S1). It will be interesting to produce variants of this stem-loop structure to increase our knowledge on the specificity of the aCBF5-aNOP10-aGAR1 complex. Although the stem-loop structure formed by the small substrate is not the one adopted by this 23S rRNA region in the entire molecule, we cannot exclude the possibility that it is formed at some stage during 23S rRNA synthesis or 50S subunit assembly. Interestingly, in yeast the corresponding modification (position 2876) is guided by the sR34 snoRNA (82, 83, 84). As Pus10, acts like the aCBF5/aNOP10/aGAR1 at position 55 in tRNAs (70), this enzyme may also be involved in the generation of residue Ψ 2603 in *P. abyssi* 23S rRNA. There is still one remaining orphan pseudouridylated position in the *P. abyssi* 23S rRNA (position 2585). Only two RNA: Ψ -synthases are present in archaea in addition to aCBF5 and Pus10, one

homolog of TruA and one homolog of TruD (Pus7) (85). In all kingdom of life, where it is present, Pus7 is known to modify U residue within a conserved sequence RSUNΨAR (R= purine, S= G/C, N= any nucleotide) (86). Residues 2603 and 2585 are not located in sequences that fit the consensus sequence of the Pus7 modification sites. Therefore, it would be interesting to test the homolog of TruA for modification at position 2585.

Interestingly, in agreement with the absence in *P. abyssi* of homologues of the *E. coli* RluE, RluB and RluC, which are responsible for rRNA pseudouridylation at positions corresponding to positions 2549, 2697 and 2672 in the *P. abyssi* 23S rRNA, one single H/ACA sRNA in *P. abyssi*, sRNA Pab35, guides U to Ψ conversion at these 3 positions.

Structural determinants for H/ACA sRNA specificity

Our data shed light on structural determinants of the sRNA and sRNA-rRNA interaction that will be useful for the identification of new H/ACA sRNAs and their target sites. The distance between the ACA box and the uridine residue to be modified is invariably either 14 or 15-nt long in contrast to the 16-nt length described for the eukaryal system (87). The distance between the K-turn and the pseudouridylation site is invariably of 5 or 6 bps and helix H2 does not contain more than one bulged nucleotide. We can try to explain these constraints. Indeed, on the basis of the present knowledge on sRNP structure and the sRNA-rRNA 3D structure, NMR data showed that the intermolecular helices P1S and P2S formed with the 3' and 5' guide sequences, respectively, can be coaxially stacked on helices H1 and H2 of the snoRNA (37, 38). Two discontinuous pseudo-helices can be formed with the substrate folded into a U-shape, so that the targeted U residue is protruding at the centre of the RNA structure (37). One has to consider how these RNA duplexes interact with the protein moiety of the sRNP. Based on the established sRNP 3D structure, the H2-P2S pseudo-helix likely interacts with protein L7Ae bound to the K-turn or K-loop (26, 28, 27) and to a lesser extent with aNop10 and aCBF5 through helix H2 (36). The X-ray structure analyses showed that several interactions are established between

proteins aCBF5 and aNOP10 (88, 89, 90, 36) and also between proteins L7Ae and aNop10 (36). Finally, the H1-P1S pseudo-helix is expected to interact with protein aCBF5 bound to the ACA triplet (28, 27, 36). Thus, it is highly conceivable that variations in the lengths of those two pseudo-helices disturb the positioning of aCBF5 relative to the targeted U residue and/or prevent some protein-protein interaction in the core protein structure, which are required for the activity (27, 71).

Finally, little is known on the requirements concerning substrate-sRNA base-pair interaction. We observed a greater importance of the interaction formed between the 3' guide sequence and the substrate (P1S), compared to that formed between the 5' guide sequence and the substrate (PS2). These observations are consistent with the closer contact of aCBF5 with the 3' guide sequence than with the 5' guide sequence that is observed in the 3D structure established for an sRNP (36). Remarkably, the length of the required base-pair interaction in archaea seems to be shorter than that needed in the eukaryal system (at least 10 bps long). This released requirement in archaea likely explains why a given H/ACA motif in archaea can guide Ψ formation at up to three different positions in rRNAs.

REFERENCES

1. Rozenski, J., Crain, P.F. and McCloskey, J.A. (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res*, **27**, 196-197.
2. Davis, D.R., Veltri, C.A. and Nielsen, L. (1998) An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNA^{Lys}, tRNA^{His} and tRNA^{Tyr}. *J Biomol Struct Dyn*, **15**, 1121-1132.
3. Arnez, J.G. and Steitz, T.A. (1994) Crystal structure of unmodified tRNA^{Gln} complexed with glutamyl- tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry*, **33**, 7560-7567.
4. Auffinger, P. and Westhof, E. (1998) In Grosjean, H. and Benne, R. (eds.), *The modification and editing of RNA*. ASM Press, Washington, DC.
5. Charette, M. and Gray, M.W. (2000) Pseudouridine in RNA: what, where, how, and why. *IUBMB Life*, **49**, 341-351.
6. Auffinger, P. and Westhof, E. (2001) An extended structural signature for the tRNA anticodon loop. *Rna*, **7**, 334-341.
7. Yu, Y.T., Shu, M.D. and Steitz, J.A. (1998) Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J*, **17**, 5783-5795.
8. Newby, M.I. and Greenbaum, N.L. (2002) Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat Struct Biol*, **9**, 958-965.
9. Segault, V., Will, C.L., Sproat, B.S. and Luhrmann, R. (1995) In vitro reconstitution of mammalian U2 and U5 snRNPs active in splicing: Sm proteins are functionally interchangeable and are essential for the formation of functional U2 and U5 snRNPs. *Embo J*, **14**, 4010-4021.
10. Yang, C., McPheeters, D.S. and Yu, Y.T. (2005) Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in *Saccharomyces cerevisiae*. *J Biol Chem*, **280**, 6655-6662.
11. Veldman, G.M., Klootwijk, J., de Regt, V.C., Planta, R.J., Branlant, C., Krol, A. and Ebel, J.P. (1981) The primary and secondary structure of yeast 26S rRNA. *Nucleic Acids Res*, **9**, 6935-6952.
12. Decatur, W.A. and Fournier, M.J. (2002) rRNA modifications and ribosome function. *Trends Biochem Sci*, **27**, 344-351.
13. Ofengand, J. (2002) Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett*, **514**, 17-25.
14. Meroueh, M., Grohar, P.J., Qiu, J., SantaLucia, J., Jr., Scaringe, S.A. and Chow, C.S. (2000) Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of *Escherichia coli* 23S rRNA. *Nucleic Acids Res*, **28**, 2075-2083.
15. King, T.H., Liu, B., McCully, R.R. and Fournier, M.J. (2003) Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell*, **11**, 425-435.
16. Badis, G., Fromont-Racine, M. and Jacquier, A. (2003) A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *Rna*, **9**, 771-779.
17. Sumita, M., Desaulniers, J.P., Chang, Y.C., Chui, H.M., Clos, L., 2nd and Chow, C.S. (2005) Effects of nucleotide substitution and modification on the stability and structure of helix 69 from 28S rRNA. *Rna*, **11**, 1420-1429.
18. Del Campo, M., Recinos, C., Yanez, G., Pomerantz, S.C., Guymon, R., Crain, P.F., McCloskey, J.A. and Ofengand, J. (2005) Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of *Haloarcula marismortui* and *Deinococcus radiodurans*. *Rna*, **11**, 210-219.
19. Kirpekar, F., Hansen, L.H., Rasmussen, A., Poehlsgaard, J. and Vester, B. (2005) The archaeon *Haloarcula marismortui* has few modifications in the central parts of its 23S ribosomal RNA. *J Mol Biol*, **348**, 563-573.
20. Mengel-Jorgensen, J., Jensen, S.S., Rasmussen, A., Poehlsgaard, J., Iversen, J.J. and Kirpekar, F. (2006) Modifications in *Thermus thermophilus* 23 S ribosomal RNA are centered in regions of RNA-RNA contact. *J Biol Chem*, **281**, 22108-22117.

21. Hoang, C. and Ferre-D'Amare, A.R. (2001) Cocystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell*, **107**, 929-939.
22. Hoang, C., Chen, J., Vizthum, C.A., Kandel, J.M., Hamilton, C.S., Mueller, E.G. and Ferre-D'Amare, A.R. (2006) Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure. *Mol Cell*, **24**, 535-545.
23. Hamma, T. and Ferre-D'Amare, A.R. (2006) Pseudouridine synthases. *Chem Biol*, **13**, 1125-1135.
24. Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145-148.
25. Tang, T.H., Bachellerie, J.P., Rozhdestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J. and Huttenhofer, A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc Natl Acad Sci U S A*, **99**, 7536-7541.
26. Rozhdestvensky, T.S., Tang, T.H., Tchirkova, I.V., Brosius, J., Bachellerie, J.P. and Huttenhofer, A. (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res*, **31**, 869-877.
27. Charpentier, B., Muller, S. and Branlant, C. (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res*, **33**, 3133-3144.
28. Baker, D.L., Youssef, O.A., Chastkofsky, M.I., Dy, D.A., Terns, R.M. and Terns, M.P. (2005) RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev*, **19**, 1238-1248.
29. Decatur, W.A. and Fournier, M.J. (2003) RNA-guided nucleotide modification of ribosomal and other RNAs. *J Biol Chem*, **278**, 695-698.
30. Muller, S., Charpentier, B., Branlant, C. and Leclerc, F. (2007) A Dedicated Computational Approach for the Identification of Archaeal H/ACA sRNAs. *Methods Enzymol*, **425**, 355-387.
31. Meier, U.T. and Blobel, G. (1994) NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria. *J Cell Biol*, **127**, 1505-1514.
32. Lafontaine, D.L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M. and Tollervey, D. (1998) The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev*, **12**, 527-537.
33. Wang, C. and Meier, U.T. (2004) Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *Embo J*, **23**, 1857-1867.
34. Henras, A., Henry, Y., Bousquet-Antonelli, C., Noaillac-Depeyre, J., Gelugne, J.P. and Caizergues-Ferrer, M. (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J*, **17**, 7078-7090.
35. Watkins, N.J., Gottschalk, A., Neubauer, G., Kastner, B., Fabrizio, P., Mann, M. and Luhrmann, R. (1998) Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA*, **4**, 1549-1568.
36. Li, L. and Ye, K. (2006) Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature*, **443**, 302-307.
37. Wu, H. and Feigon, J. (2007) H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *Proc Natl Acad Sci U S A*, **104**, 6655-6660.
38. Jin, H., Loria, J.P. and Moore, P.B. (2007) Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Mol Cell*, **26**, 205-215.
39. Ofengand, J. and Bakin, A. (1997) Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaeobacteria, mitochondria and chloroplasts. *J Mol Biol*, **266**, 246-268.
40. Massenet, S., Ansmant, I., Motorin, Y. and Branlant, C. (1999) The first determination of pseudouridine residues in 23S ribosomal RNA from hyperthermophilic Archaea *Sulfolobus acidocaldarius*. *FEBS Lett*, **462**, 94-100.
41. Klein, R.J., Misulovin, Z. and Eddy, S.R. (2002) Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc Natl Acad Sci U S A*, **99**, 7542-7547.

42. Thebault, P., de Givry, S., Schiex, T. and Gaspin, C. (2006) Searching RNA motifs and their intermolecular contacts with constraint networks. *Bioinformatics*, **22**, 2074-2080.
43. Zago, M.A., Dennis, P.P. and Omer, A.D. (2005) The expanding world of small RNAs in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Microbiol*, **55**, 1812-1828.
44. Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A. *et al.* (1999) Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res*, **6**, 83-101, 145-152.
45. Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D. *et al.* (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*, **390**, 364-370.
46. Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N.S., Thorsson, V., Sbrogna, J. *et al.* (2000) Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci U S A*, **97**, 12176-12181.
47. Bult, C., White, O., Olsen, G., Zhou, L., Fleischmann, R., Sutton, G., Blake, J., FitzGerald, L., Clayton, R., Gocayne, J. *et al.* (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058-1073.
48. Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K. *et al.* (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol*, **179**, 7135-7155.
49. Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A. *et al.* (1998) Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res*, **5**, 55-76.
50. Robb, F.T., Maeder, D.L., Brown, J.R., DiRuggiero, J., Stump, M.D., Yeh, R.K., Weiss, R.B. and Dunn, D.M. (2001) Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. *Methods Enzymol*, **330**, 134-157.
51. She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., Chan-Weiher, C.C., Clausen, I.G., Curtis, B.A., De Moors, A. *et al.* (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A*, **98**, 7835-7840.
52. Ruepp, A., Graml, W., Santos-Martinez, M.L., Koretke, K.K., Volker, C., Mewes, H.W., Frishman, D., Stocker, S., Lupas, A.N. and Baumeister, W. (2000) The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature*, **407**, 508-513.
53. Kawashima, T., Amano, N., Koike, H., Makino, S., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T. *et al.* (2000) Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proc Natl Acad Sci U S A*, **97**, 14257-14262.
54. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.
55. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.
56. Besemer, J., Lomsadze, A. and Borodovsky, M. (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res*, **29**, 2607-2618.
57. Gautheret, D., Major, F. and Cedergren, R. (1993) Modeling the three-dimensional structure of RNA using discrete nucleotide conformational sets. *J Mol Biol*, **229**, 1049-1064.
58. Gaspin, C., Cavaille, J., Erauso, G. and Bachellerie, J.P. (2000) Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J Mol Biol*, **297**, 895-906.
59. Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R. and Dennis, P.P. (2000) Homologs of small nucleolar RNAs in Archaea. *Science*, **288**, 517-522.
60. Gautheret, D. and Lambert, A. (2001) Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J Mol Biol*, **313**, 1003-1011.

61. Charbonnier, F., Erauso, G., Barbeyron, T., Prieur, D. and Forterre, P. (1992) Evidence that a plasmid from a hyperthermophilic archaebacterium is relaxed at physiological temperatures. *J Bacteriol*, **174**, 6103-6108.
62. Chomczynski, P. and Sacchi, N. (1986) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, **162**, 156-159.
63. Nagahisa, K., Ezaki, S., Fujiwara, S., Imanaka, T. and Takagi, M. (1999) Sequence and transcriptional studies of five clustered flagellin genes from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *FEMS Microbiol Lett*, **178**, 183-190.
64. Bakin, A. and Ofengand, J. (1993) Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center - analysis by the application of a new sequencing technique. *Biochemistry*, **32**, 9754-9762.
65. Motorin, Y., Muller, S., Behm-Ansmant, I. and Branlant, C. (2007) Identification of Modified Residues in RNAs by Reverse Transcription-Based Methods. *Methods Enzymol*, **425**, 21-53.
66. Charpentier, B., Fourmann, J.B. and Branlant, C. (2007) Reconstitution of archaeal H/ACA sRNPs and test of their activity. *Methods Enzymol*, **425**, 389-405.
67. Mougin, A., Gregoire, A., Banroques, J., Segault, V., Fournier, R., Brule, F., Chevrier-Miller, M. and Branlant, C. (1996) Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *RNA*, **2**, 1079-1093.
68. Keith, G. (1995) Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie*, **77**, 142-144.
69. Motorin, Y. and Grosjean, H. (1998) In Grosjean, H. and Benne, R. (eds.), *The modification and editing of RNA*. ASM Press, Washington, DC, pp. 543-550.
70. Roovers, M., Hale, C., Tricot, C., Terns, M.P., Terns, R.M., Grosjean, H. and Droogmans, L. (2006) Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic Acids Res*, **34**, 4293-4301.
71. Muller, S., Fourmann, J.B., Loegler, C., Charpentier, B. and Branlant, C. (2007) Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:{Psi}55-synthase and RNA-guided RNA:{Psi}-synthase activities. *Nucleic Acids Res*.
72. Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R. *et al.* (1981) Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res*, **9**, 6167-6189.
73. Gutell, R.R., Larsen, N. and Woese, C.R. (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev*, **58**, 10-26.
74. Dennis, P.P., Omer, A. and Lowe, T. (2001) A guided tour: small RNA function in Archaea. *Mol Microbiol*, **40**, 509-519.
75. Harms, J., Schlutzenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F. and Yonath, A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, **107**, 679-688.
76. Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. and Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883-896.
77. Branlant, C., Krol, A., Machatt, M.A., Pouyet, J., Ebel, J.P., Edwards, K. and Kossel, H. (1981) Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. *Nucleic Acids Res.*, **9**, 4303-4324.
78. Barta, A., Kuechler, E., Branlant, C., Sri Widada, J., Krol, A. and Ebel, J.P. (1975) Photoaffinity labelling of 23 S RNA at the donor-site of the *Escherichia coli* ribosome. *FEBS Lett*, **56**, 170-174.
79. Noon, K.R., Bruenger, E. and McCloskey, J.A. (1998) Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus*. *J Bacteriol*, **180**, 2883-2888.
80. Juzumiene, D.I. and Wollenzien, P. (2001) Arrangement of the central pseudoknot region of 16S rRNA in the 30S ribosomal subunit determined by site-directed 4-thiouridine crosslinking. *RNA*, **7**, 71-84.
81. Bullard, J.M., van Waes, M.A., Bucklin, D.J., Rice, M.J. and Hill, W.E. (1998) Regions of 16S ribosomal RNA proximal to transfer RNA bound at the P-site of *Escherichia coli* ribosomes. *Biochemistry*, **37**, 1350-1356.

82. Samarsky, D.A., Balakin, A.G. and Fournier, M.J. (1995) Characterization of three new snRNAs from *Saccharomyces cerevisiae*: snR34, snR35 and snR36. *Nucleic Acids Res*, **23**, 2548-2554.
83. Ni, J., Tien, A.L. and Fournier, M.J. (1997) Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell*, **89**, 565-573.
84. Ganot, P., Bortolin, M.L. and Kiss, T. (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, **89**, 799-809.
85. Watanabe, Y. and Gray, M.W. (2000) Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaeobacteria. *Nucleic Acids Res*, **28**, 2342-2352.
86. Behm-Ansmant, I., Urban, A., Ma, X., Yu, Y.-T., Motorin, Y. and Branlant, C. (2003) The *Saccharomyces cerevisiae* U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. *RNA*, **9**, 1371-1382.
87. Ganot, P., Caizergues-Ferrer, M. and Kiss, T. (1997) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev*, **11**, 941-956.
88. Hamma, T., Reichow, S.L., Varani, G. and Ferre-D'Amare, A.R. (2005) The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat Struct Mol Biol*, **12**, 1101-1107.
89. Manival, X., Charron, C., Fourmann, J.B., Godard, F., Charpentier, B. and Branlant, C. (2006) Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res*, **34**, 826-839.
90. Rashid, R., Liang, B., Baker, D.L., Youssef, O.A., He, Y., Phipps, K., Terns, R.M., Terns, M.P. and Li, H. (2006) Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol Cell*, **21**, 249-260.

FIGURE LEGENDS

Table 1. Predicted and experimentally identified target sites of the identified H/ACA sRNAs

The predicted target positions in the 16S and 23S rRNAs are indicated for each of the H/ACA sRNA candidates. The detection by the CMCT-RT analysis of a Ψ residue at the predicted position in the 16S or 23S rRNA from *P. abyssi* is indicated by “+” in the second lane. The detection of an *in vitro* activity of the reconstituted H/ACA sRNP candidate at the predicted site is indicated by “+”, “++” or “+++” in the third lane. The number of “+” is proportional to the rate of modification detected after a 80 min incubation (25-50%, 50-80% and 80-100%, respectively). The target positions that were already predicted for the *P. furiosus* Pf1, Pf3, Pf6, Pf7 and Pf9 sRNAs (26, 28), which are the counterparts of the *P. abyssi* Pab21, Pab105, Pab35, Pab40 and Pab160 sRNAs are indicated by an asterisk.

Figure 1. The multistep process used in the computer based comparative genomics approach

Description of the 5 steps of the computer based comparative genomics approach that we used in this study: sequences and annotations were first downloaded. The Inter-Coding-Regions (ICR) sequences were then extracted from the genomic sequences and assembled in databases used for the next steps. Comparison of the ICR sequences by BLAST resulted in the identification of 120 conserved ICR. Protein-coding sequences and known C/D box sRNA genes were filtered out of these conserved sequences. Within the resulting 72 conserved ICR, 7 displayed putative genes for H/ACA sRNA candidates. Helices H1 and H2 that delimit the pseudouridylation pocket are indicated on the scheme on the bottom right. The ANA sequence at the 3' end of the RNA and the K-turn motif in the apical part of the H/ACA SLS are boxed.

Figure 2. Secondary structure of the 7 H/ACA sRNA candidates

The sequences and the secondary structure models established for the Pab21 (A), Pab35 (B), Pab40 (C), Pab91 (D), Pab105 (E) and Pab160 (F) *P. abyssi* sRNAs are shown. The ANA sequence at the 3' end of the RNA and the K-turn motif in the apical part of the H/ACA SLS are boxed. Base substitution in *P. furiosus* (*P.f.*), *P. horikoshii* (*P.h.*) and *T. kodakarensis* (*T.k.*) are shown. Two different folding of the H/ACA SLS1 in Pab21 sRNA and SLS2 in Pab40 sRNA can be proposed (a. and b., panels A and C).

Figure 3. Test by Northern blot analysis of the expression in *P. abyssi* of the H/ACA sRNA candidates

The expression of the putative H/ACA sRNAs was tested by Northern blot analysis in total RNA extracts from *P. abyssi* cells grown until the end of the exponential growth phase. The sequences of the complementary probes used to detect the expression of Pab21, Pab35, Pab40, Pab91, Pab105 and Pab160 sRNAs are given as supplementary data. Lane 1 corresponds to the assay performed with a 5' end-labeled probe specific of the tested H/ACA sRNA. A 5' end-labeled DNA ladder was loaded in lane 2. The detected transcripts are shown with an arrow on the left of each autoradiogram.

Figure 4. Predicted base-pair interaction between the various H/ACA motifs of the 7 detected H/ACA sRNAs and their putative target sites

Interaction between the rRNA segments predicted to be targets of Pab21 (A), Pab35 (B), Pab40 (C), Pab91 (D), Pab105 (E) and Pab160 (F) sRNAs and these sRNAs are represented. The expected targeted U residues are indicated by an arrow and by numbers giving their position in the 16S and 23S rRNAs. The distance between the K-turn motif and the pseudouridylation pocket is given in bps, the one between the ANA sequence and the pseudouridylation pocket is given in nts. The two possible intermolecular interactions proposed for the SLS motif of Pab21

and the SLS2 motif of Pab40 result from the two possible alternative confirmations of these H/ACA motifs (Fig. 2, panels A and C).

Figure 5. Identification of Ψ residues in the *P. abyssi* 16S (Panel A) and 23S (Panel B) rRNAs by the RT-CMCT method

Position of residues in 16S or 23S rRNAs is given on the right side of the autoradiograms. The *P. abyssi* total RNA extracted from cells grown up to the end of the exponential phase was treated with CMCT, during 2, 10 and 20 min, except for the first lane (same conditions without CMCT during 20 min). The CMCT-modified RNAs were or were not subjected to an alkaline treatment at pH 10.4 as indicated above the lanes. The appearance of an RT stop of increasing intensity in lanes 2, 3 and 4 reveals the presence of Ψ residues (indicated by horizontal arrows). Lanes U, G, C, and A correspond to the RNA sequencing ladder.

Figure 6. *In vitro* activity analysis of the 7 reconstituted sRNP candidates.

H/ACA sRNP particles were assembled by using *in vitro* transcribed H/ACA sRNA, the four recombinant proteins L7Ae (L), aCBF5 (C), aNOP10 (N) and aGAR1 (G) (LCNG) and *in vitro* transcribed RNA substrate as previously described (27). For each RNA target an assay was performed in the presence or in the absence of the guide RNA. To this end 150 fmoles of RNA targets were mixed with the core proteins L7Ae (L), aCBF5 (C), aNOP10 (N) and aGAR1 (G) (200 nM each) with or without 4 pmoles of the H/ACA sRNA. The RNA targets were transcribed using [α - ^{32} P] ATP, CTP, GTP or UTP and digested, after incubation, by RNase T2 (Panels A and B) or RNase P1 (Panels C and D). The obtained NMPs were fractionated by TLC on cellulose plates using either the N1-N2 (Panels A and C), or the N1-R2 (Panel B) or the N1 (Panel D) migration systems. Ap, Cp, Gp, Up and Ψ p or pU and p Ψ spots are indicated with their specific ratio in each of the target sequences. The Ψ spots are circled. The amount of Ψ formation was assessed and is indicated.

Figure 7. Comparison of Ψ residues location in the domain V of the *P. abyssi* 23S rRNA corresponding to the peptidyl transferase center

The secondary structure of the *P. abyssi* 23S rRNA is adapted from the one of the *Thermococcus celer* 23S rRNA (M67497) (Gutell website, www.rna.icmb.utexas.edu). Numbering of residues is that of the *P. abyssi* 23S rRNA. The Ψ residues detected in this work are indicated in red and by an arrow marked with a P, those found in *E. coli*, *B. subtilis*, *D. radiodurans*, *T. thermophilus*, *Z. mays* chloroplasts, *H. marismortui*, *H. halobium*, *A. fulgidus*, *S. acidocaldarius*, *S. cerevisiae*, *D. melanogaster*, *M. musculus* and *H. sapiens* are indicated by arrows marked by E, B, O, U, Z, L, A, F, S, Y, D M and H, respectively.

Figure S1. The aCBF5/aNOP10/aGAR1 complex can catalyse U to Ψ conversion at position 2603 in the 23S rRNA in the absence of a guide RNA

The secondary structure of the T Ψ C arm of the tRNA^{Asp} (Panel A) and the similar folding proposed for the 23S rRNA portion from position 2594 to 2611 (Panel B) are represented. The U residue converted into Ψ residue is circled in red. The short 23S rRNA substrate (150 fmoles) was incubated with the aCBF5/aNOP10/aGAR1 (200 nM each) protein complex in the absence of a guide RNA. The experiment was performed in conditions described in Figure 6. The yield of Ψ formation was assessed and is indicated. Ap, Cp, Gp, Up and Ψ p spots are indicated with their specific ratio in each of the target sequences. The Ψ spot is circled.

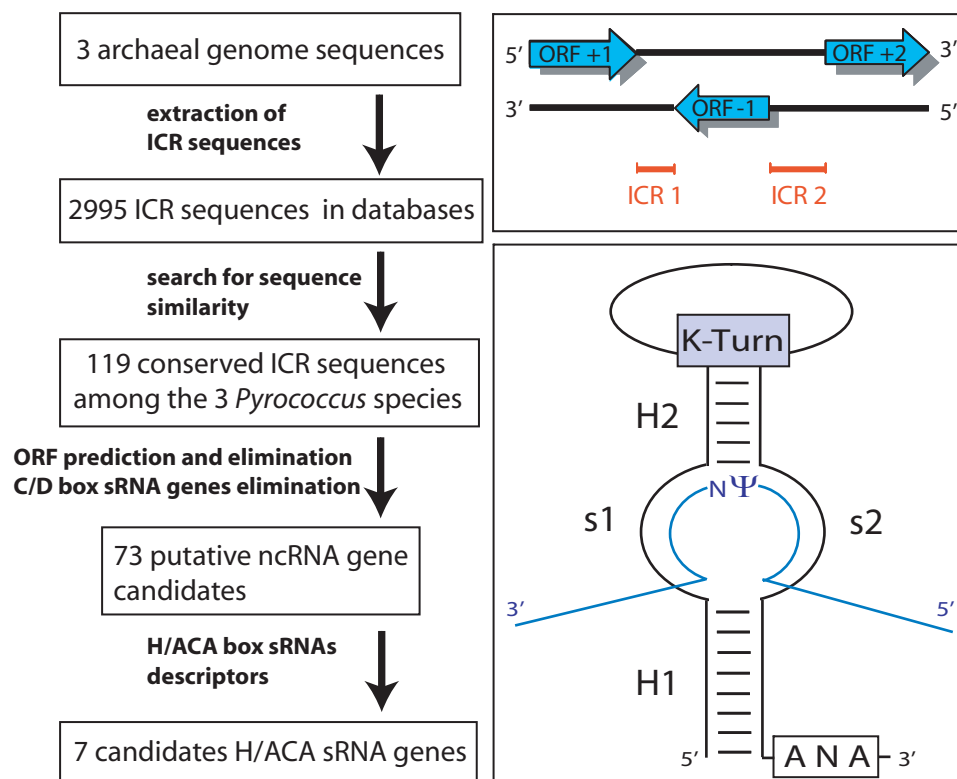


Figure 1

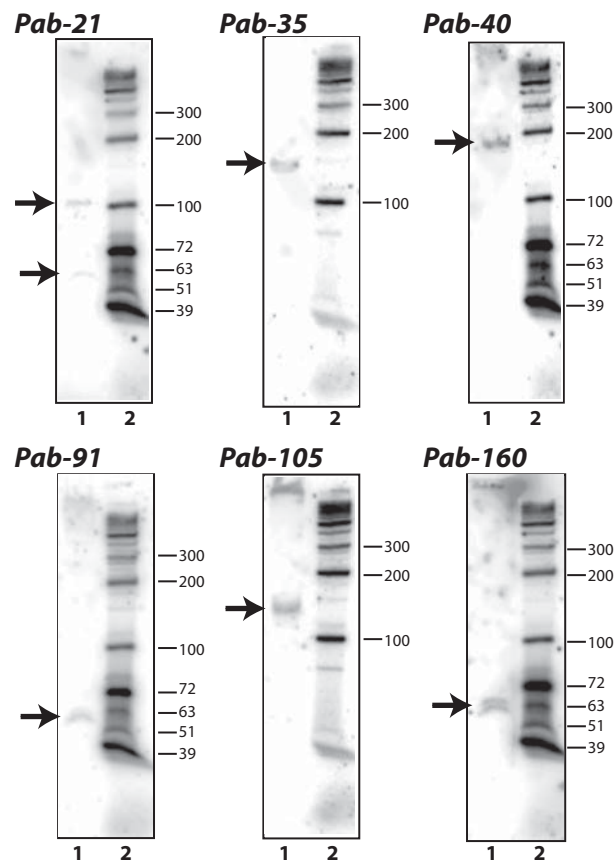


Figure 3

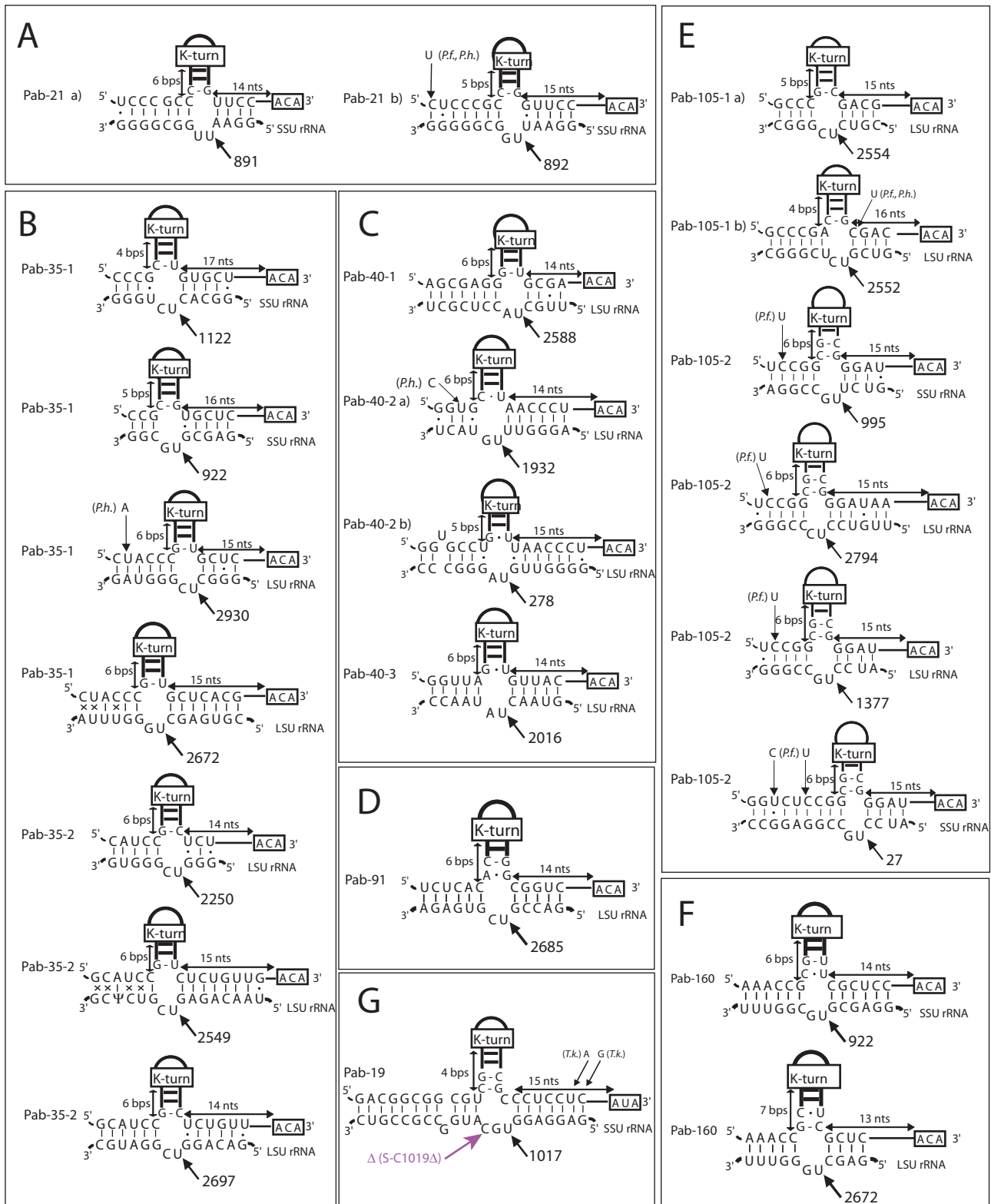


Figure 4

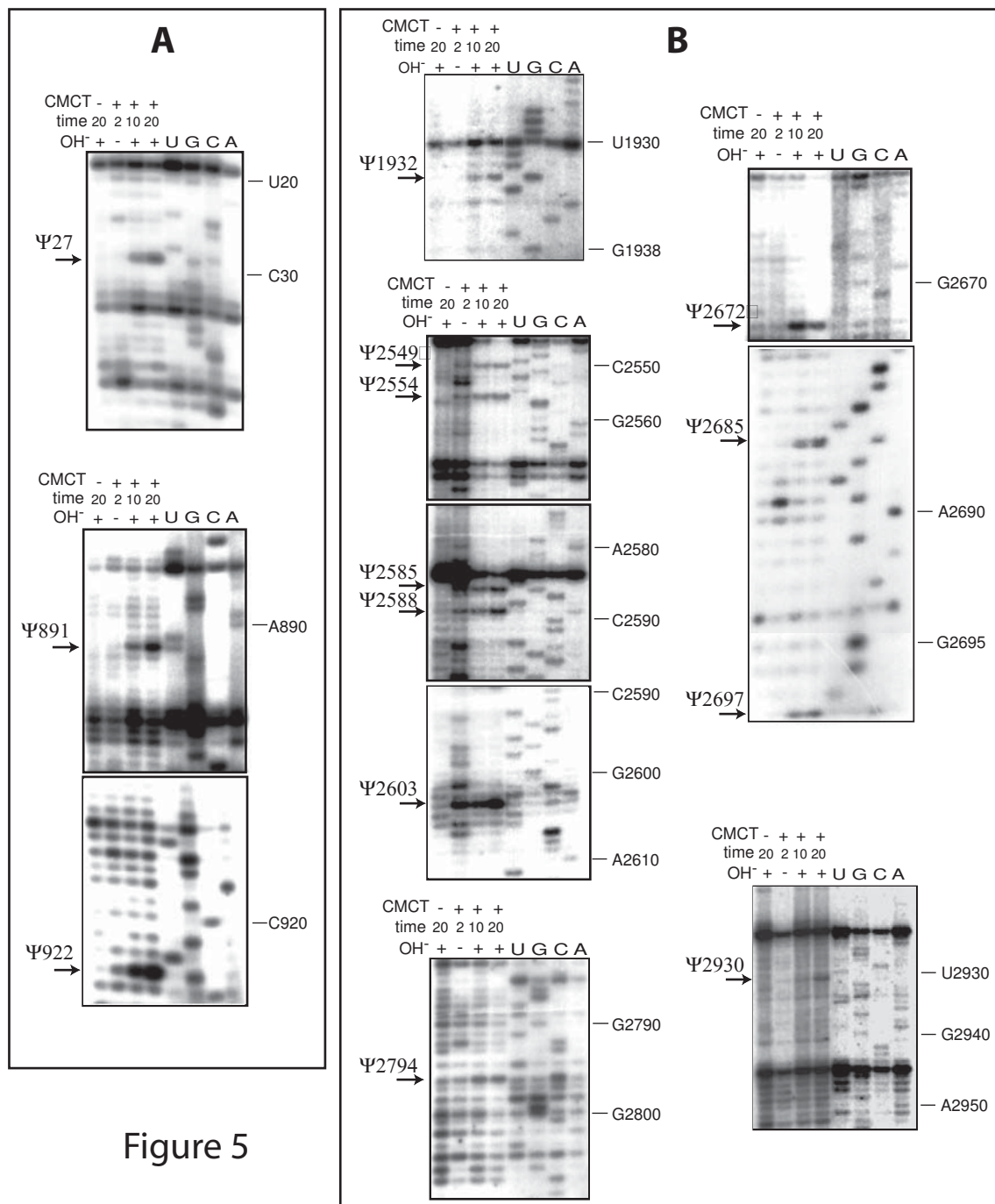


Figure 5

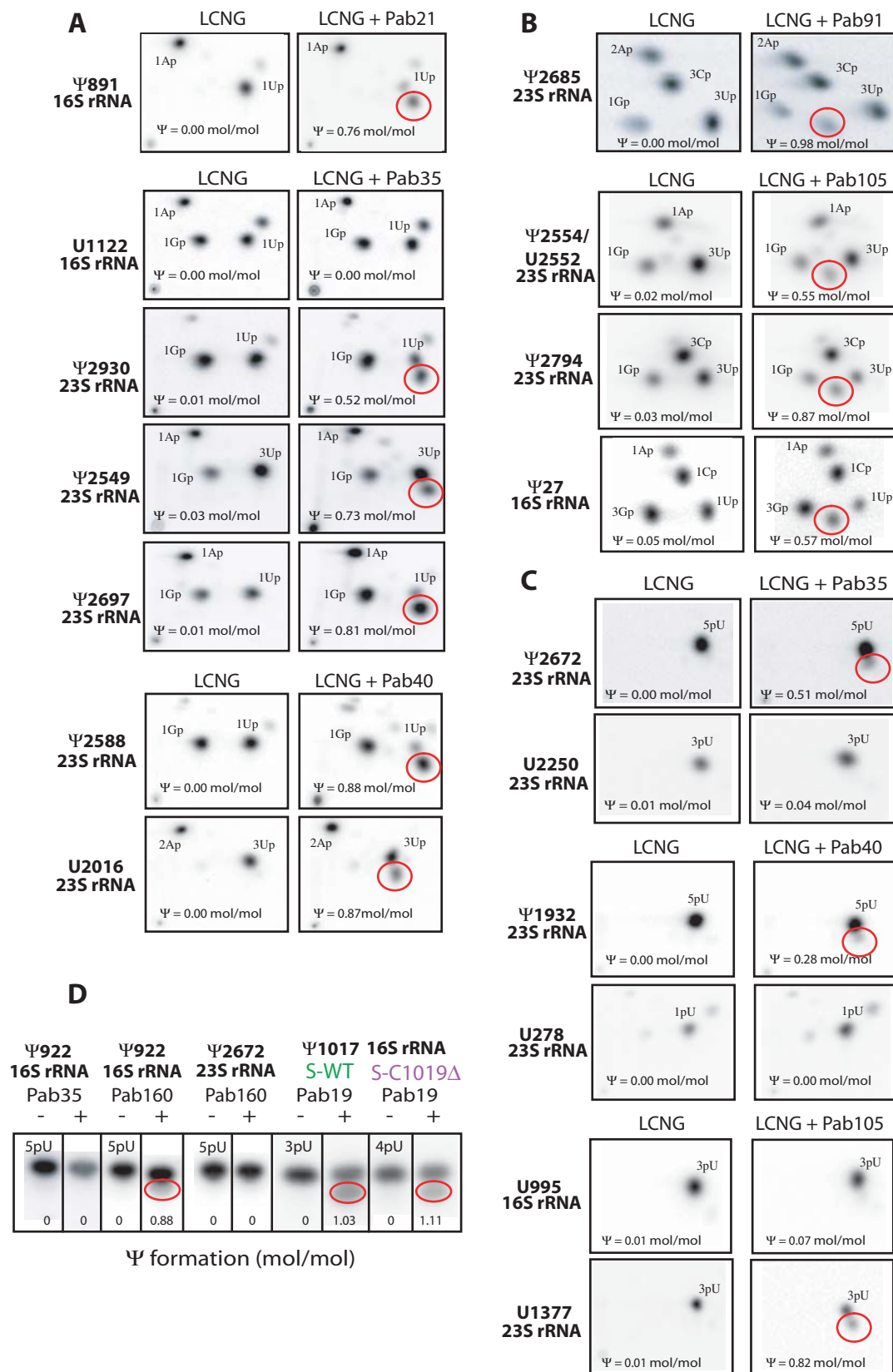


Figure 6

<i>P. abyssi</i> H/ACA	Predicted targets	RT-CMCT analysis	<i>In vitro</i> activity
Pab21	SSU 891	+	++
	SSU 892*	-	-
Pab35	motif 1 : SSU 1122	-	-
	motif 1 : SSU 922	+	-
	motif 1 : LSU 2930	+	++
	motif 1 : LSU 2672	+	++
	motif 2 : LSU 2250	-	-
	motif 2 : LSU 2549	+	++
	motif 2 : LSU 2697*	+	+++
Pab40	motif 1 : LSU 2588*	+	+++
	motif 2 : LSU 278*	-	-
	motif 2 : LSU1932	+	+
	motif 3 : LSU 2016*	-	+++
Pab91	LSU 2685	+	+++
Pab105	motif 1 : LSU 2552*	-	-
	motif 1 : LSU 2554	+	++
	motif 2 : SSU 995	-	-
	motif 2 : SSU 27*	+	++
	motif 2 : LSU 1377	-	+++
	motif 2 : LSU 2794	+	+++
Pab160	LSU 2672	+	-
	SSU 922*	+	+++
Pab19	SSU 1017	-	+++

Table 1

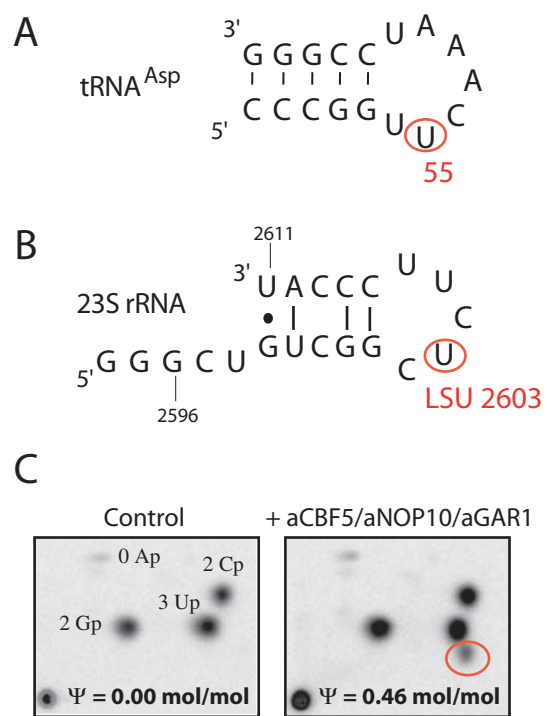


Figure S1

6. Etude comparative de l'action de aCBF5 avec et sans sRNA H/ACA

A. Contexte et démarche

En parallèle de la recherche informatique des sRNA H/ACA, j'ai réalisé une étude comparative des gènes potentiels d'ARN:Ψ-synthases chez les archaea. Le nombre d'ARN:Ψ-synthases est très limité chez ces organismes en comparaison avec la situation observée chez *E. coli* (11 enzymes) et chez *S. cerevisiae* (10 enzymes), où ces enzymes ont été largement étudiées. Lors de ces comparaisons nous avons en particulier remarqué, à l'exception de l'enzyme aCBF5 que nous avons montré faire partie des sRNP H/ACA, l'absence d'homologue de l'enzyme TruB d'*E. coli* chez les archaea (Fig. R.19). Or l'enzyme bactérienne TruB est responsable de la conversion du résidu U55 en résidu Ψ dans tous les ARNt à l'exception de l'ARNt^{Met} initiateur. Chez *S. cerevisiae*, cette conversion de U en Ψ dans la boucle terminale du bras TΨC des ARNt est catalysée par Pus4p, une enzyme de la famille TruB (Becker *et al.*, 1997 ; Gu *et al.*, 1998). Or, la conversion de U en Ψ en position 55 des ARNt qui est universelle dans le monde vivant, a aussi été mise en évidence dans des ARNt élongateurs d'archaea (Gupta, 1984 ; Rozenski *et al.*, 1999). La question était donc de savoir quelle enzyme parmi les 4 ARN:Ψ-synthases alors détectées chez les archaea, aPus7, aPus10, aTruA et aCBF5 catalyse la modification en position 55 des ARNt, sachant que la séquence conservée TΨC des ARNt est importante lors de la reconnaissance des ARN élongateurs au site A des ribosomes (Barta *et al.*, 1975 ; Sanbonmatsu *et al.*, 2005). Nous avons alors décidé de tester si aCBF5 peut être active sur les ARNt d'archaea, mais aussi de levure. C'est effectivement le cas et nous avons observé une augmentation de l'activité catalytique en présence des protéines aNOP10 et aGAR1, en particulier en absence du trinuécléotide CCA 3' terminal des ARNt. Malheureusement, deux autres équipes ont fait la même observation en parallèle et ont publié ensemble leur résultats (Roovers *et al.*, 2006). Les auteurs observaient également qu'une autre ARN:Ψ-synthase, aPus10, était capable de catalyser la conversion du résidu U55 des ARNt d'archaea en résidu Ψ. La question restait donc de savoir si seule l'une des deux enzymes agit en position 55 des ARNt d'archaea *in cellulo*. Par ailleurs, l'activation par aNOP10 et aGAR1 de la modification de la position U55 des ARNt par aCBF5 n'avait pas été étudiée de manière approfondie, de même que l'effet de mutations dans l'ARNt sur cette activité.

Enfin, grâce à cette dualité d'activité d'aCBF5 avec et sans ARN guide nous avons la possibilité de tester le rôle d'acides aminés conservés d'aCBF5 et de aNOP10 dans son activité catalytique selon qu'elle est guidée ou non guidée par un sRNA H/ACA. Cette partie du travail a été prise en charge par B. Charpentier et JB. Fourmann. Ainsi, nous avons pu publier l'ensemble de nos données sur l'action d'aCBF5 sur les ARNt d'archaea dans un article publié dans NAR où je suis premier auteur.

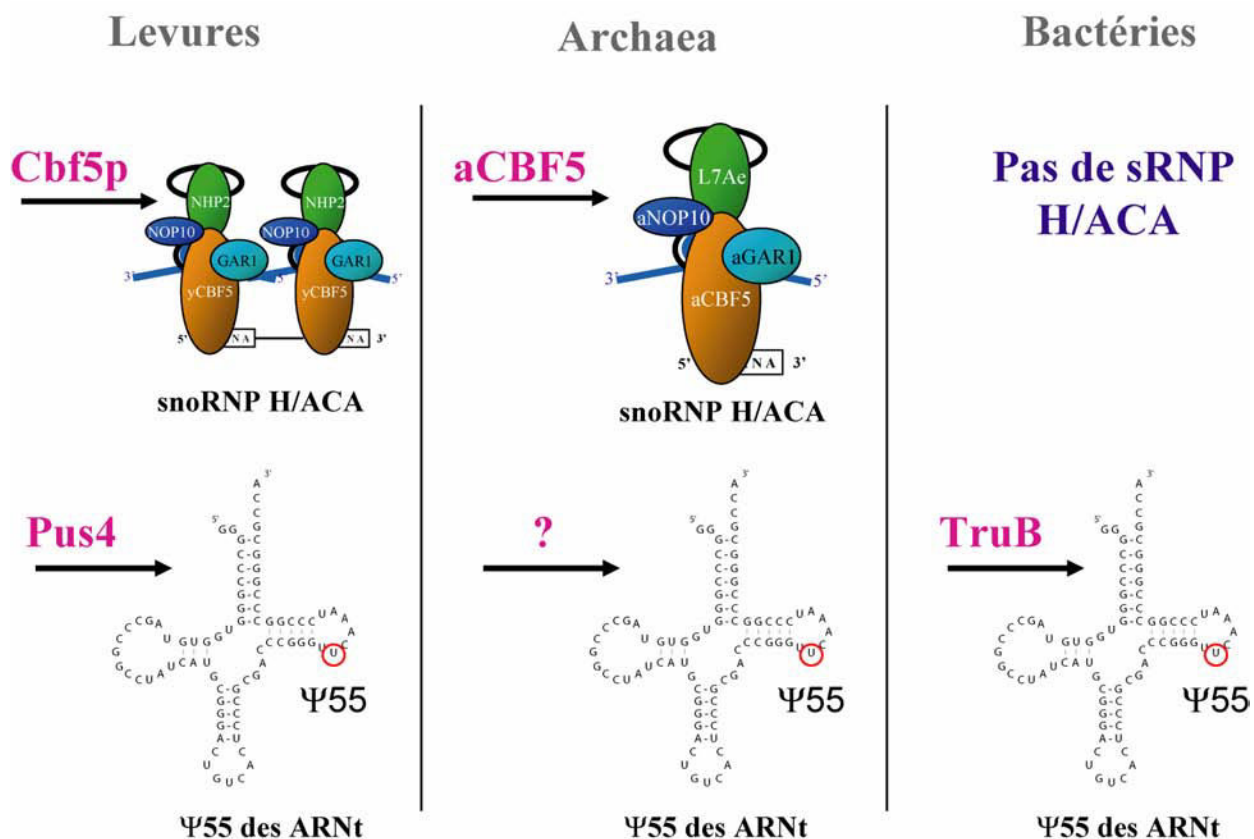


Figure R. 19 : Distribution et fonction des ARN:Ψ-synthases de la famille TruB chez les bactéries, les archaea et les levures.

Les espèces eucaryotes, représentées ici par les levures, contiennent deux ARN:Ψ-synthases de la famille TruB, Pus4 et Cbf5p. L'enzyme Pus4 est impliquée dans la pseudouridylation des ARNt en position 55, et est par conséquent l'homologue fonctionnel de l'enzyme TruB des bactéries. L'enzyme yCBF5 fait partie des snoRNP H/ACA. Chez les archaea, aCBF5 est impliquée dans les sRNP H/ACA, mais aucun homologue fonctionnel de TruB de bactéries n'a été identifié.

B. Résultats

Dans cet article nous montrons que la fixation d'aCBF5 à l'ARNt dépend du domaine PUA et de la séquence CCA en 3' de l'ARNt. Rappelons que l'équipe avait montré auparavant que ce domaine est nécessaire à la fixation du trinucletotide ACA des sRNA H/ACA (Manival *et al.*, 2006). Nous avons montré qu'en absence du motif CCA terminal de l'ARNt, aNOP10 renforce fortement la fixation de aCBF5 à l'ARNt et de ce fait son activité en position 55. Nous avons observé un niveau similaire d'activité en présence de aGAR1, cependant, le mode d'activation par aGAR1 n'a pas été déterminé dans ce cas. En utilisant à la fois l'approche basée sur une hydrolyse de l'ARNt par la RNase T2 (méthode du plus proche voisin), une analyse par RT-CMCT et en produisant un variant U55C, j'ai pu montrer que l'action de aCBF5 sur l'ARNt est restreinte à la position 55. J'ai aussi étudié l'effet de la délétion de différentes structures tige-boucles de l'ARNt sur l'efficacité catalytique de aCBF5 seule ou au sein du complexe aCBF5/aNOP10/aGAR1 en position 55 de l'ARNt. J'ai ainsi pu montrer qu'aucune activité

d'aCBF5 n'est observée en absence du bras anticodon. La présence de aNOP10 ou de aGAR1 est requise pour qu'il y ait activité. Nous en concluons donc que le bras anticodon favorise la fixation de aCBF5 sur l'ARNt. Enfin, j'ai produit un très court fragment de l'ARNt^{Asp} de *P. abyssi* correspondant au bras accepteur et au bras TΨC. L'activité de aCBF5 est faible sur ce substrat mais elle est fortement renforcée en présence d'aNOP10 et aGAR1. Ces données montraient, comme nous l'avons observé pour l'activité guidée de aCBF5 une forte importance de l'interaction entre aNOP10 et aCBF5 pour l'activité non guidée de cette enzyme. De ce fait JB. Fourmann et B. Charpentier ont réalisé une série de mutants dans aCBF5 et aNOP10 et nous avons comparé les activités guidées et non guidées des complexes aCBF5/aBOP10 formés avec ces protéines mutées. De manière intéressante, nous avons constaté que la mutation de certains résidus conservés de aNOP10 ou de aCBF5 altèrent l'activité guidée mais peu l'activité non guidée et, vice versa, la mutation de résidus conservés de aNOP10 et de aCBF5 altèrent l'activité guidée mais pas l'activité non guidée. La conservation phylogénétique de résidus impliqués dans l'activité non guidée d'aCBF5 renforce l'idée qu'elle a un rôle biologique qui peut être la modification de la position U55 des ARNt, mais aussi la modification de plusieurs positions dans les ARNr, comme la position 2603 dans l'ARNr 23S de *P. abyssi* que nous avons décrite ci-avant.

Publication N°3 :

Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Ψ55-synthase and RNA-guided RNA:Psi-synthase activities.

Sébastien Muller, Jean-Baptiste Fourmann, Christine Loegler, Bruno Charpentier and Christiane Branlant.

PUBLICATION N°3

Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Ψ⁵⁵-synthase and RNA-guided RNA:Ψ-synthase activities

Sébastien Muller, Jean-Baptiste Fourmann, Christine Loegler, Bruno Charpentier* and Christiane Branlant

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP, Nancy Université, Université Henri Poincaré-BP 239, 54506 Vandoeuvre-Lès-Nancy cedex, France

Received July 3, 2007; Revised and Accepted July 25, 2007

ABSTRACT

Protein aNOP10 has an essential scaffolding function in H/ACA sRNPs and its interaction with the pseudouridine(Ψ)-synthase aCBF5 is required for the RNA-guided RNA:Ψ-synthase activity. Recently, aCBF5 was shown to catalyze the isomerization of U55 in tRNAs without the help of a guide sRNA. Here we show that the stable anchoring of aCBF5 to tRNAs relies on its PUA domain and the tRNA CCA sequence. Nonetheless, interaction of aNOP10 with aCBF5 can counterbalance the absence of the PUA domain or the CCA sequence and more generally helps the aCBF5 tRNA:Ψ⁵⁵-synthase activity. Whereas substitution of the aNOP10 residue Y14 by an alanine disturbs this activity, it only impairs mildly the RNA-guided activity. The opposite effect was observed for the aNOP10 variant H31A. Substitution K53A or R202A in aCBF5 impairs both the tRNA:Ψ⁵⁵-synthase and the RNA-guided RNA:Ψ-synthase activities. Remarkably, the presence of aNOP10 compensates for the negative effect of these substitutions on the tRNA: Ψ⁵⁵-synthase activity. Substitution of the aCBF5 conserved residue H77 that is expected to extrude the targeted U residue in tRNA strongly affects the efficiency of U55 modification but has no major effect on the RNA-guided activity. This negative effect can also be compensated by the presence of aNOP10.

INTRODUCTION

Isomerization of uridine into pseudouridine (5-ribosyluracil, Ψ) is a prevalent post-transcriptional modification of

cellular RNAs (1) and this modified residue is frequently found in functionally important RNA regions. Consistently, an important role of Ψ residues was demonstrated for specific codon–anticodon recognition and ribosome function (2–8). Furthermore, one of the highly conserved Ψ residue in U2 snRNA was found to play a key role in pre-mRNA splicing (9,10).

Two types of catalysts can generate Ψ residues at specific sites in a pre-existing polyribonucleotide chains. One type consists in a single protein with an RNA:Ψ-synthase activity. It recognizes specifically one or more RNA segments or motifs and converts the targeted U residues within these segments into Ψ residues. A large panel of such RNA:Ψ-synthases are found in most cell types and they are up to now the unique system detected in bacteria (11). These bacterial RNA:Ψ-synthases were the first to be characterized (12–15). The search for their homologs by inspection of various genomes led to the classification of these enzymes into five families, TruA, TruB, TruD, RsuA, RluA (16–18). Based on sequence alignment and 3D structure analysis of the catalytic domains of several RNA:Ψ-synthases, five conserved structural motifs (I, II, IIa, III and IIIa) were identified and they represent the signature for RNA:Ψ-synthase activity (16,19,20). Although only a limited number of amino acids in these motifs are conserved, these motifs have highly conserved conformations in the 3D structures established for RNA:Ψ-synthases (11). The few conserved amino acids play important roles for catalysis: a universally conserved aspartate residue in motif II is the essential catalytic amino acid (17,21–24). A conserved basic residue (Arg or Lys) in motif III makes a salt bridge with the catalytic Asp residue. A conserved Tyr residue in motif IIa has a stacking interaction with the uracyl residue to be modified in the target RNA (25). Two other

*To whom correspondence should be addressed. Tel: 33 3 83 68 43 16; Fax: 33 3 83 68 43 07; Email: bruno.charpentier@maem.uhp-nancy.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© 2007 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

hydrophobic residues, a leucine in motif IIIa and an isoleucine or a valine residue in motif III, are also highly conserved, but their functional implication in the catalytic mechanism is less understood (19,25).

In the various RNA:Ψ-synthases identified up to now, the conserved catalytic domain is associated with a large variety of additional domains, which are likely involved in the specific recognition of the target RNA motif. One of these domains, the PUA domain (PseudoUridine and Archaeosine tRNA guanine transglycosylase), has largely been studied. It is found at the C-terminus of members of the TruB family of RNA:Ψ-synthases (25–30). As deduced from the crystal structure of the complex formed between the *Escherichia coli* TruB enzyme and the T stem-loop structure of tRNA (31), the PUA domain is expected to play a key role in RNA recognition by interacting with the CCA 3'-terminal sequence of tRNAs (25). This proposed interaction of TruB with a sequence common to all tRNAs fits with the general activity of TruB at position 55 of all elongator tRNAs.

The second type of catalyst for U to Ψ conversion consists in RNP complexes containing a small RNA with H/ACA motifs (the H/ACA snoRNAs and scaRNAs in Eukarya and H/ACA sRNAs in archaea) and a well-defined set of proteins (32–34). One of these proteins denoted aCBF5 in archaea, Cbf5p in yeast and CBF5/NAP57/Dyskerin in human displays the characteristic motifs I, II, IIa, III and IIIa found in RNA:Ψ-synthases and belongs to the TruB family of RNA:Ψ-synthase (16,19). In each RNP, the RNA component base pairs with one or more targeted RNA sequences and through this interaction acts as a guide to specify the uridyl residues that will be modified (32–34). Knowledge on the mechanism of the H/ACA RNPs activity is largely based on data obtained with the archaeal H/ACA RNPs that could be reconstituted using an *in vitro* transcribed RNA and recombinant proteins (35,36). In addition to the protein bearing the signature of RNA:Ψ-synthases, each H/ACA RNP contains three distinct proteins (NOP10, GAR1 and NHP2 in human; Nop10p, Gar1p and Nhp2p in yeast; and aNOP10, aGAR1 and L7Ae in archaea). The three archaeal aCBF5, aNOP10 and aGAR1 proteins form an heterotrimer in solution (35,36). The 3D structures of the aCBF5-aNOP10 heterodimer and aCBF5-aNOP10-aGAR1 heterotrimer complexes were solved by X-ray crystallography (27–29).

Very recently, the crystal structure of an entire archaeal box H/ACA sRNP was determined (30). Through reconstitution experiments, we showed that in archaea the aCBF5-aNOP10 complex is the minimal set of proteins required to get an RNA-guided RNA:Ψ-synthase activity (36). However, the two other proteins aGAR1 and L7Ae both increase the efficiency of the catalytic reaction. Formation of an H/ACA sRNA-aCBF5-aNOP10 complex depends upon aCBF5 association with the sRNA and the conserved ACA motif in the sRNA plays an essential role for this association (35,36). More precisely, the PUA domain of aCBF5 interacts with the ACA motif of H/ACA sRNAs (27,30). According to our previous reconstitution assays, formation of an sRNA-aCBF5-aNOP10 complex is required for binding

of the RNA substrate and recent structural studies on NOP10 revealed its capability to bind RNA (37). By *in vitro* reconstitution of H/ACA sRNPs with truncated aNOP10 proteins, we showed that its C-terminal half is sufficient for sRNP activity, whereas its N-terminal half reinforces the aCBF5-aNOP10 interaction (27). The L7Ae protein binds a K-turn motif present in each stem-loop structures of the archaeal H/ACA sRNAs (35,36,38,39). The direct interaction of L7Ae with aNOP10 and the sRNA stabilizes the association of the other proteins with the sRNA (30). In light of the higher rate of pseudouridylation observed in the presence of aGAR1, we proposed that protein aGAR1 might enhance the turnover of the reaction catalyzed by the sRNP (36).

A recent *in vitro* study showed that in addition to its central role in the RNA guided activity of H/ACA sRNPs, the archaeal aCBF5 enzyme has the ability to convert residue U55 of an *in vitro* transcribed tRNA into a Ψ residue (40). This activity seems to be increased in the presence of the L7Ae, aNOP10 and aGAR1 set of proteins (40). Altogether, these data revealed the ability of aCBF5 to act without guide RNA in case of a tRNA substrate. However, an *in vivo* dual role of aCBF5, as rRNA:Ψ-synthase when included in H/ACA sRNPs and as tRNA:Ψ55-synthase when acting as a free protein, is still not demonstrated since another archaeal RNA:Ψ-synthase PusX/Pus10 is also able to catalyze U to Ψ conversion at position 55 in tRNAs. Nevertheless, the capability of aCBF5 to act as a free protein and in sRNPs represented an opportunity: (i) to delineate the tRNA structures needed to get an activity of aCBF5 in the absence of guide RNA, (ii) to compare the aNOP10 and aGAR1 requirements for the guided and non guided activity of aCBF5, (iii) to compare the aCBF5 amino acid requirement for its guided and non-guided activity. Here, by site-directed mutagenesis of the *Pyrococcus abyssi* and *Saccharomyces cerevisiae* tRNAs, we show that a truncated version of tRNA including the acceptor and TΨC stem-loop structures linked by a flexible sequence is sufficient to get aCBF5 activity in the absence of a guide RNA. In the absence of the 3' terminal CCA, the aCBF5 activity becomes strictly dependent on either aNOP10 or aGAR1. Furthermore, by site-directed mutagenesis of both aNOP10 and aCBF5 we demonstrate distinct amino acid requirements for the RNA-guided and non RNA-guided activities of aCBF5. The results obtained are discussed based on the 3D structures of the archaeal H/ACA sRNP proteins and their complexes.

MATERIALS AND METHODS

Production of protein variants

The aCBF5 D82A and CΔ(PUA) protein variants carrying, respectively, an Asp to Ala substitution at position 82 and a truncation of the PUA domain, were previously described (27,36). The PCR approach was used for site-directed mutagenesis of protein aCBF5 (K53A, H77A and R202A variants) and aNOP10 (Y14A, H31A and P32A variants). The sequence of the oligonucleotides used for the PCR reactions are available upon request.

Purification of recombinant proteins

Wild-type and variant aCBF5 and aNOP10 proteins as well as the aGAR1 and TruB proteins were produced as GST fusion proteins as previously described (36). The GST moiety was removed by cleavage with the PreScission protease. The various proteins were stored at -80°C in the following buffer 150 mM NaCl, 1 mM EDTA, 1 mM DTT, glycerol 10%, 50 mM Tris-HCl pH 7.

In vitro transcription of tRNAs, H/ACA sRNA and the sRNP target RNA

Wild-type and mutated tRNA transcripts, the Pab91 H/ACA sRNA and the RNA target RNA-S (36) were produced by *in vitro* transcription of PCR products. The forward primers used for PCR amplification generated the T7 RNA polymerase promoter. DNA templates for amplifications of sequences coding for the WT yeast tRNA^{Asp} and variants of this tRNA ($\Delta\text{T-tRNA}^{\text{Asp}}$, $\Delta\text{D-tRNA}^{\text{Asp}}$) were a generous gift of C. Florentz (IBMC, Strasbourg). Additional yeast tRNA^{Asp} variants ($\Delta\text{a.c.-tRNA}^{\text{Asp}}$, $\Delta\text{acceptor-tRNA}^{\text{Asp}}$, $\Delta\text{acceptor-tRNA}^{\text{Asp-U55C}}$) were produced by site-directed mutagenesis in the course of the PCR amplification. The sequence coding for the WT *P. abyssi* tRNA^{Asp} was PCR amplified using genomic DNA from the *P. abyssi* GE5 strain. Sequences encoding the variants tRNA^{Asp}-ACA and tRNA^{Asp}- ΔCCA were obtained by using a reverse primer carrying the mutated sequence. Sequences encoding the WT semi-tRNA^{Asp} or the variant semi-tRNA^{Asp}- ΔCCA or semi-tRNA^{Asp}-U55C were obtained by hybridization of two complementary oligonucleotides. The other archaeal tRNA^{Asp} variants (tRNA^{Asp}-U55C) were produced by site-directed mutagenesis in the course of a PCR amplification. DNA templates for production of the *P. abyssi* Pab91 sRNA and its target RNA (RNA-S) were previously described (36). The RNA transcripts were purified by denaturing gel electrophoresis. Labeling was done at their 5' end for EMSA experiments or during *in vitro* transcription by incorporation of [α -³²P]CTP for the pseudouridylation assays. Conditions for transcription and labeling were previously described (41).

Electrophoresis mobility shift assays (EMSA)

EMSA were performed in the conditions previously described (36,41). Briefly, radiolabeled RNAs (50 fmol) were incubated in buffer D [150 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA and 20 mM HEPES (pH 7.9)] with proteins at a 200 nM concentration, at 65°C for 10 min. To test for the association of the Pab91 sRNP with its target RNA (complex CII), 2.5 pmol of RNA-S were added. Complexes were visualized by autoradiography after fractionation by non-denaturing polyacrylamide gel electrophoresis.

In vitro pseudouridylation assay

The tRNA: Ψ 55-synthase activities of purified aCBF5 or H/ACA-RNP complexes were measured by the nearest-neighbor approach in the conditions previously described (36,41). The RNA-guided activity of reconstituted

H/ACA sRNPs was measured as previously described by mixing ~ 4 pmol of Pab91 sRNA with ~ 150 fmol of the [α -³²P]CTP-labeled target RNA-S (36). The non-RNA-guided reaction of aCBF5 on tRNA was tested on ~ 50 fmol of [α -³²P]CTP-labeled tRNAs. For both RNA-guided and non-RNA-guided reactions, samples were pre-incubated at 65°C and the reaction was started by addition of the proteins (200 nM each). Aliquots were collected at different time points, the RNAs were extracted and then digested by RNase T2. The resulting 3'-mononucleotides were fractionated by thin-layer cellulose chromatography. The radioactivity in the spots was quantified using the ImageQuant software after exposure of a phosphorimager screen. The amount of Ψ residue formed was determined taking into account the total number of U residues in the tRNA or RNA-S molecules. Some of the values given in the text are expressed with standard errors of the mean values. All the time-course analyses were repeated at least three times and gave reproducible data. Some control experiments were performed with the *E. coli* TruB recombinant protein. The same procedure was used to measure the activity, except that the reaction was performed at 37°C .

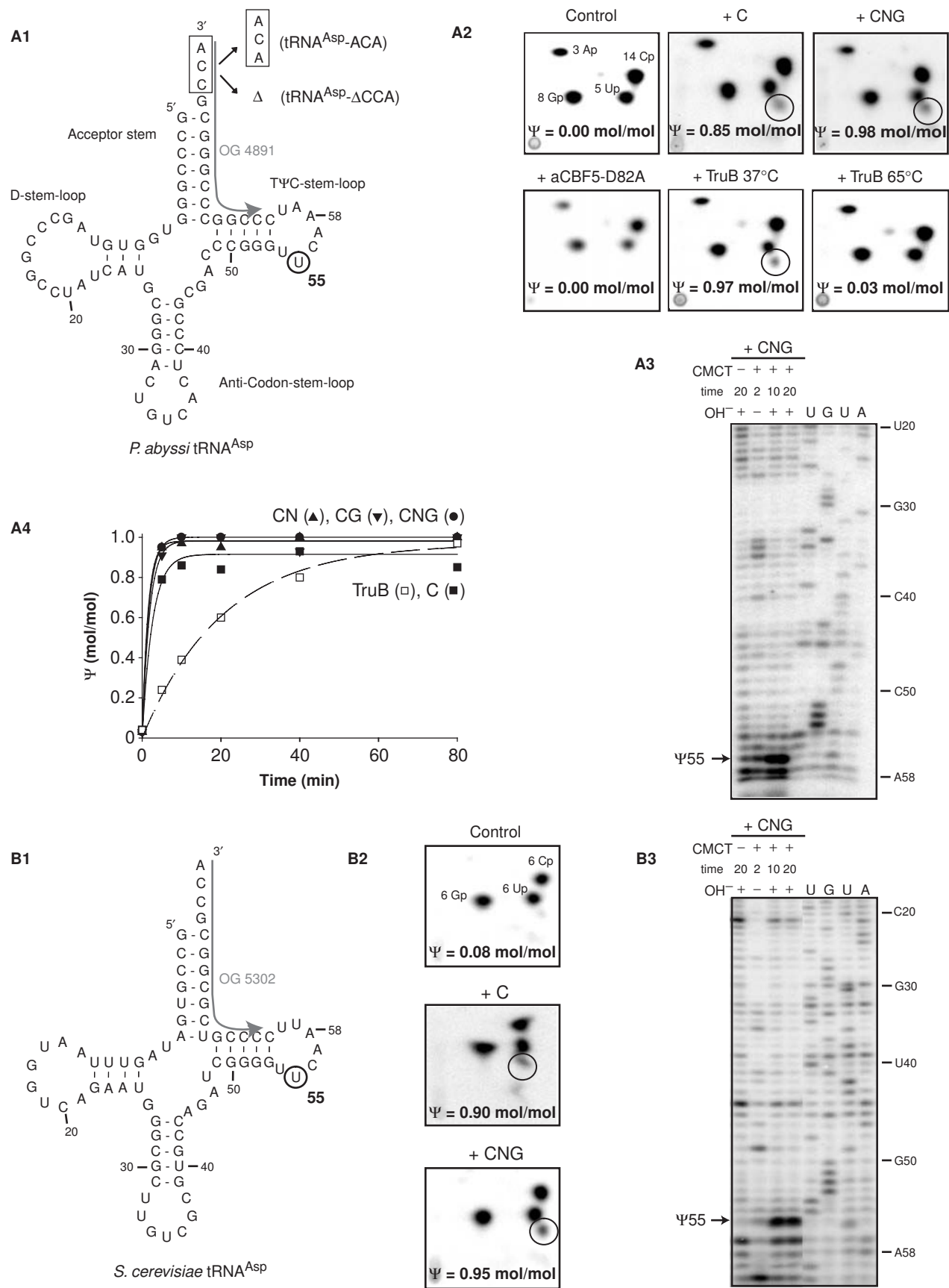
Mapping the site of pseudouridylation in tRNAs

We used the *N*-cyclohexyl-*N'*-(2-morpholinoethyl)-carbodiimide metho-*p*-toluolsulfonate (CMCT) modification protocol of (42), adapted by (43). CMCT modifications were performed with 0.5 μg of *in vitro* transcribed archaeal or yeast tRNA^{Asp}. Positions of CMCT modifications were identified by primer extension analysis, using the AMV RT (QBiogene, USA) in the conditions previously described (44) and the primer oligonucleotides OG4891 for archaeal tRNA^{Asp} and OG5302 for *S. cerevisiae* tRNA^{Asp} (Figure 1A1 and B1). Primers were 5'-end labeled using [γ -³²P]ATP (3000 Ci/mmol) and T4 polynucleotide kinase. RNA sequencing was done with 4 μg of *in vitro* transcribed RNA [for details on the technique, see (45)].

RESULTS

The kinetics of U to Ψ conversion at position 55 in tRNAs is enhanced in the presence of proteins aNOP10 and aGAR1

Previous studies revealed a U to Ψ conversion activity of the *P. furiosus* aCBF5 protein at position 55 in the *P. furiosus* tRNA^{Asp}(GUC), i.e. a tRNA: Ψ 55-synthase activity. Such an activity is enhanced in the presence of the aNOP10, aGAR1 and L7Ae proteins set (40). For a deeper understanding of these activation properties, here we quantified the relative stimulation effects of proteins aNOP10, aGAR1 and L7Ae alone and in various combinations, on the tRNA: Ψ 55-synthase activity of aCBF5. The assays were performed with the *P. abyssi* recombinant proteins aCBF5, aNOP10, aGAR1 and L7Ae, and with both *P. abyssi* tRNA^{Asp}(GUC) and *S. cerevisiae* tRNA^{Asp}(GUC) substrates (Figure 1A1 and B1). We used the nearest neighbor and the RT-CMCT approaches in order to verify that modifications occurred only at position 55 in tRNAs. For the



nearest-neighbor approach, both substrates were labeled by incorporation of [α - ^{32} P]CTP during transcription. After the pseudouridylation reactions, the tRNA transcripts were digested with RNase T2 and the products were fractionated by 2D chromatography (Figure 1A2 and B2). CMCT modifications followed by alkaline treatments were performed on cold RNA transcripts, and positions of modifications were identified by reverse transcription (Figure 1A3 and B3) using oligonucleotide primers that were complementary to the 3' terminal sequence of tRNAs (Figure 1A1 and B1). The pseudouridylation reaction was performed at 65°C, for 80 min using a 200 nM concentration of each protein. Time point experiments were performed on the *P. abyssi* tRNA^{Asp} with: (i) aCBF5 alone, (ii) aCBF5 in the presence of aNOP10 or aGAR1, or of both proteins and (iii) aCBF5 in the presence of aNOP10 and L7Ae, or the three L7Ae, aNOP10, aGAR1 proteins. Aliquots of the reaction mixtures were analyzed after 5, 10, 20, 40 and 80 min of incubation (Figures 1A4 and S1). For comparison, a time point experiment was performed at 37°C with a recombinant *E. coli* TruB enzyme (Figure 1A2).

Altogether, the results obtained demonstrated that protein aCBF5 is able to modify the *P. abyssi* and *S. cerevisiae* tRNA^{Asp} at the same high level after 80 min of incubation ($\sim 0.85 \text{ mol} \pm 0.05$ and $\sim 0.9 \text{ mol} \pm 0.05$ of Ψ per mol of tRNA, respectively). In both cases, only a slight increase of the level of modification was observed in the presence of aNOP10 and aGAR1 (Figure 1A2 and B2). However, as evidenced by the time point experiments, although the kinetics of the reaction with aCBF5 at 65°C is very fast compared to that for TruB at 37°C, it was markedly increased in the presence of aNOP10 or aGAR1. On the contrary, protein L7Ae showed no stimulatory effect (see Supplementary Data, Figure S1), which is in contrast to data obtained for the sRNA-guided activity of aCBF5 (35,36). The CMCT analysis did confirm that aCBF5 acted only at position 55 in both tRNAs. Finally, we confirmed the general capability of aCBF5 to act at position 55 in tRNAs by using other archaeal and bacterial transcripts as the substrates (data not shown).

Altogether, we concluded that aCBF5 has the ability to modify various tRNAs at position 55 and this independently of the organisms from which they originate.

The reaction is very rapid and is nearly complete after a 15-min incubation. However, the kinetics of this reaction is markedly increased in the presence of aNOP10 or aGAR1. In this case, the reaction is completed after a 7-min incubation. Furthermore, the catalytic property of aCBF5, alone or acting in the presence of aNOP10 and aGAR1 is highly specific of position 55.

Proteins aNOP10 and aGAR1 have a strong stimulatory effect on Ψ 55 formation in the absence of the aCBF5 PUA domain or of the tRNA 3' terminal CCA motif

Our previous data on the sRNA guided activity of aCBF5 revealed a strong importance of the PUA domain of aCBF5 for this activity and a negative effect of mutations in the ACA sequence of the H/ACA sRNA (27). The recent crystal structure of an H/ACA sRNP demonstrated a direct contact of the PUA domain with the ACA motif (30). Therefore, we tested whether the PUA domain was also required for the activity of aCBF5 at position 55 in tRNAs. As shown in Figure 2, no U to Ψ conversion occurred in the *P. abyssi* tRNA^{Asp} when it was incubated with an aCBF5 protein lacking the PUA domain (Δ (PUA)). Interestingly, this tRNA: Ψ 55-synthase activity was partially restored in the presence of the aNOP10 and aGAR1 proteins pair (Δ (PUA)NG). However, the kinetics was slower and the efficiency of modification after 80 min was lower. The individual aNOP10 and aGAR1 proteins had similar but very low capabilities to restore the tRNA: Ψ 55-synthase activity of the truncated aCBF5 protein (Figure 2).

As previous data had revealed the absence of aCBF5 activity on a tRNA lacking the 3' terminal CCA motif (40), we tested whether aNOP10 or aGAR1 or both proteins can compensate for the absence of this motif (Figure 3). We found in this case that individual aNOP10 or aGAR1 can restore a strong aCBF5 activity (CN and CG, respectively) at position 55 of tRNAs and this activity was reinforced to some extent when the two proteins were present (CNG). Remarkably, modification with the CNG complex was almost complete after a 10-min incubation, even in the absence of the CCA motif (Figure 3). The truncated tRNA was modified at a slower rate and lower yield, when TruB was used as the catalyst. Taken together, the data revealed a very important role of the PUA

Figure 1. The *P. abyssi* aCBF5 protein modifies position 55 in *P. abyssi* and *S. cerevisiae* tRNA^{Asp}. (A1) Secondary structure of the *P. abyssi* tRNA^{Asp} used in this study. Residue U55 is circled. The CCA sequence at the 3' end is boxed. The CCA to ACA substitution and the CCA deletion in variants tRNA^{Asp}-ACA and tRNA^{Asp}- Δ CCA, are shown. (A2) 2D Thin-layer-chromatography (TLC) analysis of Ψ formation in *P. abyssi* tRNA^{Asp}. The *P. abyssi* tRNA^{Asp} transcript was labeled by the incorporation of [α - ^{32} P]CTP during transcription. The ^{32} P-labeled tRNA substrate (50 fmol) was mixed with either the *P. abyssi* aCBF5 protein (C), or its aCBF5-D82A variant, or the aCBF5-aNOP10-aGAR1 complex (CNG), or the *E. coli* TruB enzyme (200 nM each). Incubation was performed for 80 min at 65°C for the archaeal enzymes and at 37 or 65°C for TruB. After T2 RNase digestion, the amount of Ψ residue formed was measured by 2D TLC and expressed in moles per moles of tRNA. The data shown are representative examples of at least three independent experiments. (A3) CMC/RT analysis for the detection of Ψ formation at position 55 in the *P. abyssi* tRNA^{Asp}. The unlabeled tRNA (2 μ g, 80 pmol) was incubated with proteins aCBF5, aNOP10 and aGAR1 (CNG, 200 mM each) for 80 min at 65°C. Treatment with CMCT was carried out for 2, 10 or 20 min as described in Materials and Methods section with (+) or without (–) alkaline treatment. Positions of Ψ residues were identified by primer extension analysis using oligonucleotide OG 4891 (represented by the arrow in panel A1). Lanes U, G, C and A correspond to the RNA sequencing ladder. (A4) Time course of Ψ 55 formation by aCBF5 protein in the *P. abyssi* tRNA^{Asp} and effect of the addition of proteins aNOP10 and aGAR1. The *P. abyssi* tRNA^{Asp} substrate was labeled as in panel A2 and incubated with various protein combinations: aCBF5 alone (C), aCBF5 and aNOP10 (CN), aCBF5 and aGAR1 (CG), a set of the three proteins (CNG) or TruB. For each time point, an aliquot of the reaction mixture was analyzed by 2D TLC. The amount of Ψ residue was estimated as in panel A2. (B1) Secondary structure of the *S. cerevisiae* tRNA^{Asp}. The sequence of oligonucleotide OG 5302 used as the primer for reverse transcription is indicated by an arrow. (B2) 2D-TLC analysis of Ψ formation by aCBF5 in the *S. cerevisiae* tRNA^{Asp} (same conditions as in panel A2). (B3) Detection of Ψ 55 formation in the yeast tRNA^{Asp} by the CMC/RT approach (same conditions as in panel A3).

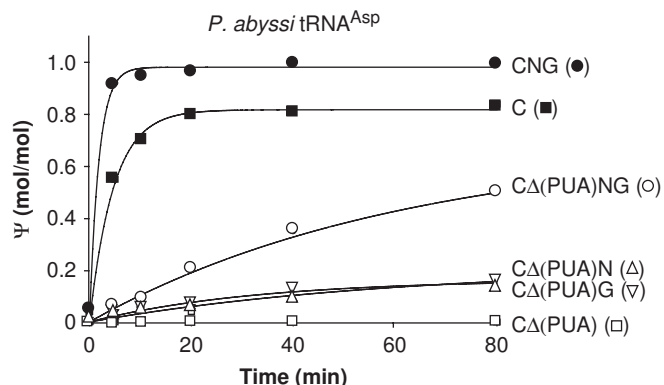


Figure 2. The PUA domain of aCBF5 is required for Ψ 55 formation. Time-course experiments of Ψ 55 formation were performed as in Figure 1 panel A4, using either the full-length aCBF5 (C) or aCBF5 truncated of its PUA domain [C Δ (PUA)], in the presence or the absence of the aNOP10 and/or aGAR1 proteins [CNG, C Δ (PUA)N, C Δ (PUA)G and C Δ (PUA)NG].

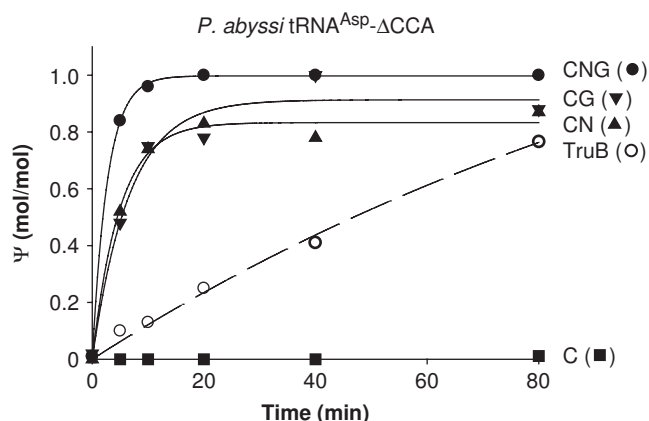


Figure 3. The aCBF5–aNOP10–aGAR1 complex can efficiently modify a tRNA lacking a 3' CCA sequence. Time-course experiments of Ψ 55 formation in the variant tRNA^{Asp}– Δ CCA by the *E. coli* TruB enzyme or various combinations of the aCBF5 (C), aNOP10 (N) and aGAR1 (G) proteins (same conditions as in Figure 2).

domain for the aCBF5 tRNA: Ψ 55-synthase activity and showed the capability of the aNOP10–aGAR1 protein pair to compensate partially for its absence. Moreover, the presence of the auxiliary proteins fully restores the aCBF5 activity in the absence of the tRNA CCA 3' terminal motif.

The aNOP10–aGAR1 protein pair can compensate for severe truncations in tRNAs

As we found that aNOP10 and aGAR1 can compensate for CCA deletion, we wondered whether they can compensate for more severe truncations in tRNAs. As aCBF5 was acting on the *S. cerevisiae* tRNA^{Asp}, and as several variants of this tRNA had already been produced (Figure 4A1), we tested the capability of aCBF5 (C) and the aCBF5–aNOP10–aGAR1 complex (CNG) to convert U55 into a Ψ residue in these truncated tRNAs (Figure 4A2). In RNAs Δ T-tRNA^{Asp}, Δ a.c.-tRNA^{Asp} and Δ D-tRNA^{Asp}, the T stem-loop,

the anticodon stem-loop, and the D stem-loop were deleted, respectively. The acceptor stem was truncated in RNA Δ acceptor-tRNA^{Asp}, but the CCA 3'-terminal motif was conserved (Figure 4A1). Obviously, no Ψ formation was detected in the Δ T-tRNA^{Asp} substrate which was lacking residue U55. Only a very low level of modification was detected in the Δ a.c.-tRNA^{Asp} when aCBF5 was used alone, while a high level of modification was detected when aNOP10 and aGAR1 were present in the incubation mixture (Figure 4A2). Similarly, no modification of the Δ acceptor-tRNA^{Asp} substrate was observed when the incubation was performed with aCBF5 alone, whereas almost a complete modification was detected for these two truncated tRNAs in the presence of aNOP10 and aGAR1 (Figure 4A2). Modification only occurred at position 55 in the truncated substrates since it disappeared after U to C mutation at this position (Δ acceptor-tRNA^{Asp}-U55C). Interestingly, the Δ D-tRNA^{Asp} was modified by aCBF5 at a high level in absence of the auxiliary proteins (Figure 4A2). As expected from previous data (46), TruB was able to modify a tRNA lacking the acceptor stem without the need for auxiliary proteins. Based on the activity of aCBF5 alone on the Δ D-tRNA^{Asp}, we concluded that aCBF5 alone, like TruB, does not need the tRNA 3D structure for activity. However, in contrast to TruB, to act alone aCBF5 needs the presence of its target T stem-loop but also the anticodon stem-loop and the acceptor stem.

Roovers *et al.* (40) had found that aCBF5 does not modify a truncated *P. furiosus* tRNA^{Asp} that contains the acceptor stem fused to the T stem-loop. We confirmed that for an equivalent *P. abyssi* substrate, no activity was restored upon addition of the aNOP10–aGAR1 protein pair (data not shown). However, as aCBF5 was active on the yeast Δ D-tRNA^{Asp} substrate, we postulated that flexibility between the acceptor stem and the T stem-loop might be an important feature for aCBF5 activity. Therefore, we produced a truncated RNA substrate (semi-tRNA^{Asp}) by truncation of the *P. abyssi* tRNA^{Asp}, which contained a UGAC single-stranded sequence acting as a flexible link between the acceptor stem and the T stem-loop (Figure 4B1). The aCBF5 protein alone had a modest but significant activity on this substrate (mean value <0.2 mol.mol⁻¹). This activity was abolished in the absence of the 3' terminal CCA sequence (Figure 4B2). However, in the presence of the aNOP10–aGAR1 protein pair both the semi-tRNA^{Asp} and semi-tRNA^{Asp}– Δ CCA lacking the 3' terminal CCA sequence were almost completely modified. As expected, TruB was active on both substrates in the absence of auxiliary proteins. Hence, we concluded that the aCBF5–aNOP10–aGAR1 heterotrimer, but not aCBF5 alone, has catalytic properties similar to that of TruB.

aNOP10 reinforces the formation of aCBF5–tRNA complexes

As the aNOP10 stimulatory effect on the aCBF5 activity seemed to consist in an increase of the kinetics of the reaction (36), and as aNOP10 was known to bind RNA (30,37), we hypothesized that aNOP10 might facilitate the

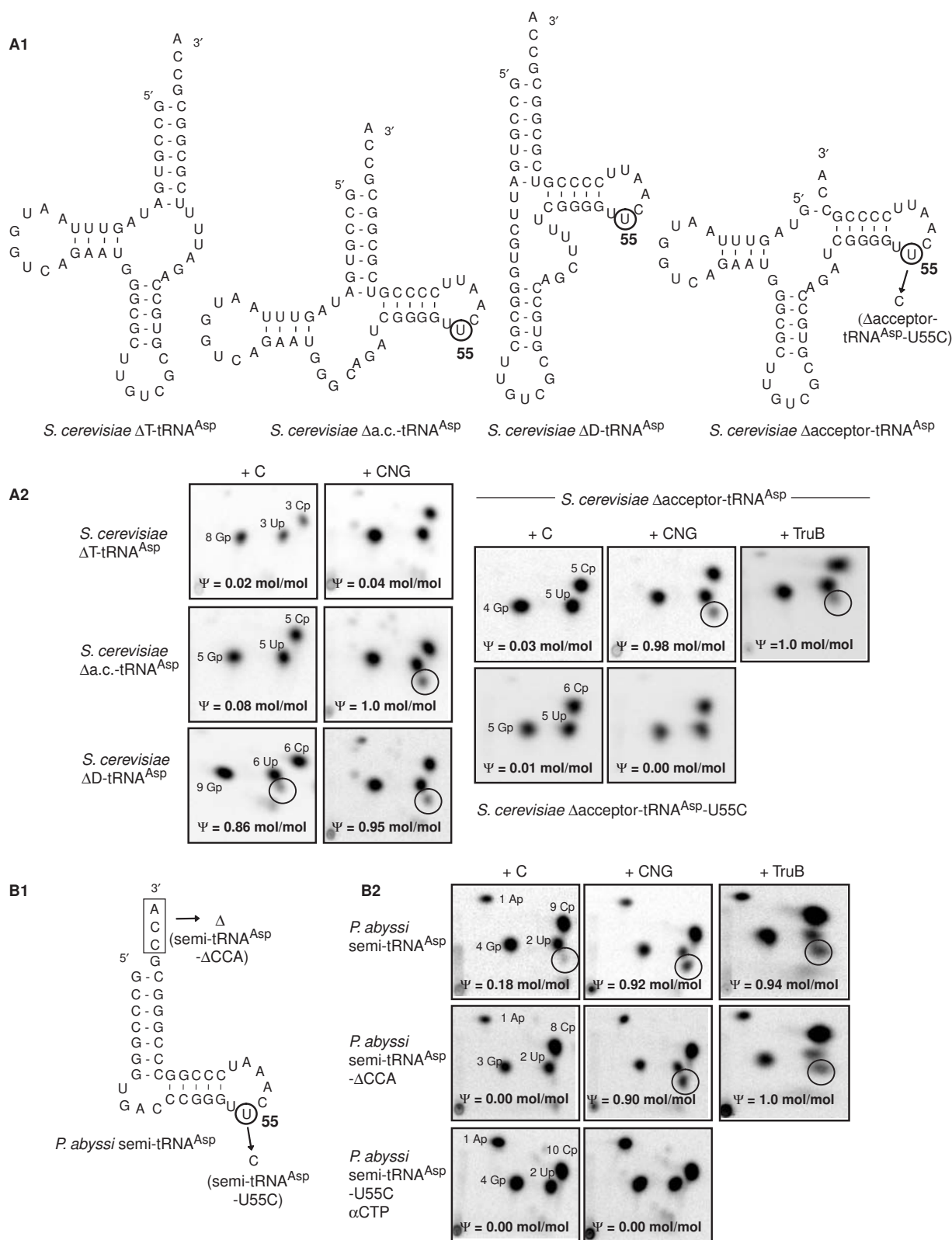


Figure 4. tRNA structural requirements for Ψ55 formation by aCBF5 or the aCBF5-aNOP10-aGAR1 complex. **(A1)** Secondary structure of the yeast truncated tRNAs^{Asp} used in this study. The DNA templates for transcription of the ΔT-tRNA^{Asp}, Δa.c.-tRNA^{Asp}, ΔD-tRNA^{Asp} and Δacceptor-tRNA^{Asp} were obtained by deletion of the T stem-loop, anticodon stem-loop, D stem-loop and acceptor stem sequences of the *S. cerevisiae* tRNA^{Asp} template sequence (Figure 1B1). **(B1)** The template sequence for the semi-tRNA^{Asp} was obtained by successive site-directed mutagenesis of the *P. abyssi* tRNA^{Asp} template sequence (Figure 1A1). The variant semi-tRNA^{Asp}-ΔCCA with a deletion of the CCA sequence and the variant semi-tRNA^{Asp}-U55C with a U to C substitution are shown. **(A2 and B2)** 2D-TLC analysis of Ψ formation by aCBF5, the aCBF5-aNOP10-aGAR1 complex or TruB, in the various truncated tRNAs. The same experiments conditions as in Figure 1A2 and B2 were used.

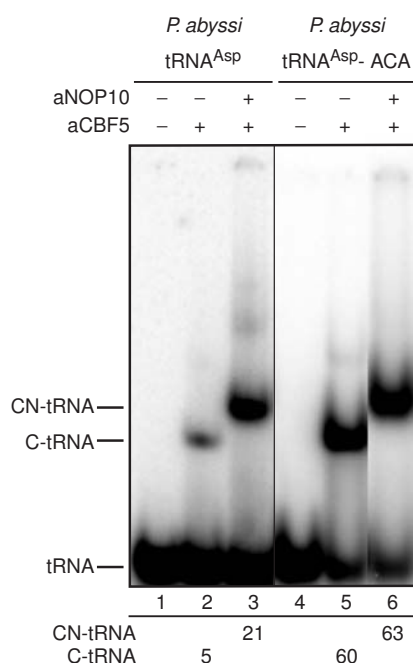


Figure 5. Analysis by electrophoresis mobility shift assays (EMSA) of the complexes formed between protein aCBF5 and archaeal tRNAs. The radiolabeled tRNA^{Asp} (lanes 1–3) or tRNA^{Asp}-ACA (lanes 4–6) (50 fmol) were incubated with aCBF5 protein (C) (200 nM), in the presence or the absence of aNOP10 (N) (200 nM) as indicated on top of each lane. Positions of the various complexes C–tRNA and CN–tRNA are indicated on the left side of the autoradiogram. The amount of the complexes was estimated as described in Materials and Methods section.

association of aCBF5 with the tRNA substrate. To test this possibility, we estimated by EMSA experiments the level of complexes formed by the association of aCBF5 with a ³²P-labeled *P. abyssi* tRNA^{Asp}. The tests were performed in the presence or absence of aNOP10 (Figure 5). The level of complex formation was increased by a factor of ~4 in the presence of aNOP10 (Figure 5, compare lane 2 with lane 3), whereas no significant increase was observed in the presence of aGAR1 (data not shown). As aCBF5 is known to interact with the ACA conserved sequence of H/ACA sRNAs, we compared the affinity of aCBF5 for the WT tRNA^{Asp} to that for a tRNA carrying an ACA triplet instead of the CCA triplet. The affinity of aCBF5 was indeed largely increased in the presence of an ACA triplet (Figure 5, compare lane 2 with lane 5) and aNOP10 did not increase the level of complex formation in this case (Figure 5, lane 6). Therefore, the stimulatory effect of aNOP10 on the kinetics of aCBF5 action on tRNAs may in part be due to its capability to facilitate the aCBF5–tRNA interaction. As previously observed for the aCBF5–RNA-guided activity, the stimulatory effect of aGAR1 on the aCBF5 activity is likely of another nature.

The conserved H31 and P32 residues in aNOP10 have different implications for the RNA-guided and non-RNA-guided activities of aCBF5

Based on the 3D structures established for the aCBF5–aNOP10 dimer and the aCBF5–aNOP10–aGAR1 trimer,

the direct interactions between some conserved residues in the central flexible linker of aNOP10 (Figure 6A) with conserved residues of aCBF5 were proposed to influence the sRNA-guided activity of aCBF5 (28,29). In particular, the aNOP10 residues Y14 and H31 were found to form hydrogen bonds with residues E199 and R201 which are present in the β12 strand of aCBF5 (Figure 6B) (29). These interactions were proposed to influence the orientation of the side-chain of residue R202 that is located on the opposite face of the β12 strand. Interestingly, in the crystal structure of the complex formed between the *E. coli* TruB RNA:Ψ-synthase and a tRNA T stem-loop, the counterpart of residue R202 was proposed to interact with the phosphate on the 5' side of the targeted uridine (31). On the other hand, in the aCBF5–aNOP10 complex, residue P32 in the aNOP10 flexible linker was found to interact by Van der Waals interaction with the conserved residue P54, which is located in motif I of aCBF5 (Figure 6C). This interaction was proposed to facilitate the formation of an hydrogen bond between residue K53 and the carbonyl oxygen of the catalytic residue D82 (28).

In light of these structural data, we used alanine substitutions to test for the relative importance of residues Y14, H31 and P32 for the RNA-guided and non-RNA-guided activities of aCBF5. Aiming at this, the stimulatory effects of each aNOP10 variant (Y14A, H31A and P32A) on the aCBF5 activity were tested on both the *P. abyssi* tRNA^{Asp} lacking the CCA sequence (Figure 6D) and the RNA-S substrate (RNA target) of the *P. abyssi* Pab91 sRNPs (Figure 6F). To this end, sRNPs were reconstituted with each of the variant aNOP10 proteins. Whereas the H31A substitution had only a limited effect on the kinetics and the yield of Ψ55 formation in the archaeal tRNA^{Asp} (Figure 6D), the Y14A substitution strongly slowed down the kinetics of this modification and the P32A substitution had a negative effect on the modification yield (~0.7 mol of Ψ per mole of tRNA, instead of ~0.9 mol of Ψ per mole of tRNA in the presence of the WT aNOP10 protein). Interestingly, the Y14A and P32A variants only had a limited effect on the sRNA-guided activity of aCBF5 (Figure 6F). The H31A mutation was in contrast strongly deleterious for this activity: only ~0.2 mol of Ψ per mole of target RNA was obtained after a 80 min incubation. Hence, mutations in aNOP10 have marked different effects on the RNA-guided and non-RNA-guided activities of aCBF5.

To try to define the reasons for the observed differences, we tested by EMSA the ability of the aNOP10 variant proteins to reinforce the aCBF5 association with the *P. abyssi* tRNA^{Asp} and with the Pab91 sRNA and its target RNA sequence (Figure 6E and G). In agreement with the time-course results, both Y14A and P32A aNOP10 variants were less efficient (~0.5-fold) than the H31A aNOP10 variant and the WT aNOP10 protein for increasing the aCBF5 association with the tRNA (Figure 6E). On the contrary, only the H31A variant had a dramatic effect on the yield of complex CII (sRNA-L7Ae–aCBF5–aNOP10–RNA target) formation, suggesting that the H31A mutation has an important negative effect on substrate recruitment in the H/ACA sRNP. Therefore, we concluded that whereas residues

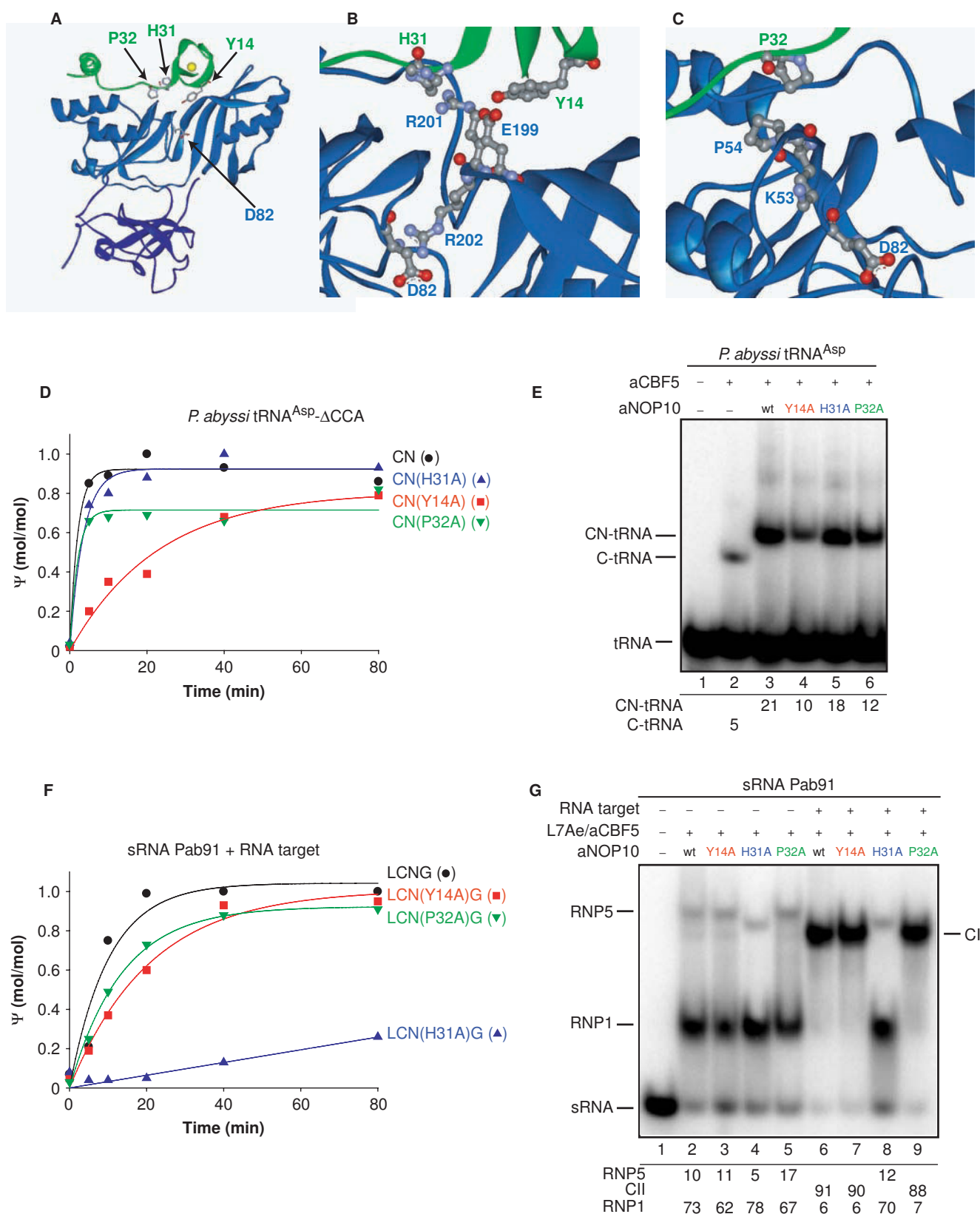


Figure 6. Role of the conserved residues in the aNOP10 protein for the non-RNA-guided and RNA-guided activities of aCBF5. (A) Ribbon representation of the heterodimer formed by interaction of aCBF5 (blue) with aNOP10 protein (green). The side chains of the aNOP10 residues P32, H31 and Y14 are represented, as well as the aCBF5 catalytic D82. (B) The residues expected to be involved in a network of interactions between aNOP10 and aCBF5 (29) are indicated on the *P. abyssi* structure. (C) Residues proposed to be involved in another network of interaction between aNOP10 and aCBF5 (28) are shown on the *P. abyssi* structure. (D) Time course of Ψ55 formation by aCBF5 (C) in the presence of the wild-type (N)

Y14 and P32 are highly important for the aCBF5 tRNA: Ψ 55-synthase activity, residue H31 plays an essential role in substrate recruitment in the sRNA-guided activity of aCBF5.

aNOP10 and aGAR1 can compensate for R202A and K53A mutations in aCBF5 for the non-guided but not for the RNA-guided activity

Next, we tested the role of the aCBF5 residues R202 and K53 whose functions in the catalytic activity were proposed to be modulated by the aNOP10 residues H31 and Y14, and by residue P32, respectively (28,29). The aCBF5 variants R202A and K53A were produced, and the effect of each substitution was tested on both RNA-guided and non-RNA-guided activities. As evidenced in Figure 7A1, both R202A and K53A substitutions in aCBF5 abolished its activity at position 55 in the *P. abyssi* tRNA^{Asp}. However, this activity was fully restored upon addition of aNOP10 and aGAR1 in the incubation mixture (Figure 7A3). Here again, as evidenced by EMSA, the defect of activity observed for CBF5 alone was at least in part due to its low affinity for the tRNA (Figure 7A2). The addition of aNOP10 enhanced the association of aCBF5 with the tRNA, which likely explains the restoration of the tRNA: Ψ 55-synthase activity. Hence, residues R202 and K53 likely play a role in tRNA association and aNOP10 can compensate for their mutations. In contrast, aNOP10 and aGAR1 were not able to compensate the negative effect of the R202A and K53A mutations in aCBF5 for the Pab91 sRNP activity (Figure 7B1), and here again, the low activity of the reconstituted sRNPs could be explained by a defect in substrate recruitment as the CII complex was formed at lower yields and displays a peculiar electrophoretic mobility (Figure 7B2, lanes 7 and 9). Hence, we concluded that residues K53 and R202 are important for tRNA binding in the non-RNA-guided system and for recruitment of the RNA substrate in the RNA-guided system. However, in this latter case, aNOP10 and aGAR1 cannot compensate for the absence of these residues. Therefore, residues R202 and K53 appear to play a more crucial role in the RNA-guided than in the non-RNA-guided system.

For both the RNA-guided and non-RNA-guided activities, the aCBF5 conserved residue H77 is dispensable in the presence of aNOP10 and aGAR1

In the *E. coli* TruB enzyme, residue H43 was proposed to be involved in the flipping of residue U55 out of the helical stack (31). The counterpart of residue H43 in aCBF5 is residue H77. As expected, the H77A mutation in aCBF5 almost completely impaired the aCBF5 activity on tRNA

(Figure 7A1) and this is also likely to be explained by the lower affinity of aCBF5 for the tRNA (Figure 7A2). However, this activity could be restored in the presence of aNOP10 and aGAR1 (Figure 7A3) and we showed that protein aNOP10 restored an efficient association of aCBF5 with the tRNA (Figure 7A2). Interestingly, the H77A mutation had a very limited effect on the Pab91 sRNP activity (Figure 7B1) consistent with an efficient recruitment of the RNA substrate observed for this variant (Figure 7B2). Hence, we concluded that residue H77 does not play an essential role for the aCBF5 RNA: Ψ -synthase activity, provided that the aNOP10 and aGAR1 proteins are present.

DISCUSSION

The present data shed light on common and specific properties of aCBF5 and TruB. They also bring a comparative analysis of the sRNA-guided and non-guided activities of aCBF5. Finally, they delineate the role of aNOP10 in the aCBF5 non-guided activity.

The major role of the PUA domain in tRNA recognition by aCBF5

Interestingly, we found that aCBF5 does not modify the Δ a.c.-tRNA^{Asp} and Δ acceptor-tRNA^{Asp} but can act on the Δ D-tRNA^{Asp}. Therefore, like for TruB, the 3D structure of tRNA is not required for recognition of the targeted U residue in tRNAs by aCBF5. However, in addition to the T stem-loop, both the anticodon stem-loop and acceptor stem contain determinants which are necessary for aCBF5 activity on tRNAs. By these extended structural requirements, aCBF5 differs from TruB, that is able to use the T stem-loop alone as a substrate (46). This different behavior may be due to variations in both the catalytic and PUA domains of aCBF5 compared to the ones of TruB. The importance of the PUA domain for the aCBF5 tRNA: Ψ 55-synthase activity is demonstrated by abolition of this activity upon PUA truncation. Previous data also revealed the strong importance of the PUA domain for H/ACA sRNP activity (27). Consistently, in the H/ACA sRNP structure, numerous protein contacts take place with the minor groove of the P1 stem and the conserved ACA motif (30). Although no 3D structure was established for an H/ACA sRNP bound to its substrate, recent NMR data (47,48) suggest that the heterologous helix P1S, which is formed by interaction of the 3' strand of the pseudouridylation pocket and the target RNA, is coaxially stacked on helix P1 of the sRNA and altogether these two helices contain a rather conserved number of base pairs (10–12).

or variants H31A (filled triangle), Y14A (filled square), P32A (filled inverted triangle) of aNOP10 protein in the tRNA- Δ CCA. Experiments were performed as in Figure 1A4. (E) Analysis by EMSA of the complexes formed with the radiolabeled tRNA^{Asp}, the aCBF5 protein with or without the WT or variant Y14A, H31A and P32A aNOP10 proteins. (F) Time-course analysis of the RNA-guided Ψ formation in an RNA target. The previously described Pab91 sRNA target (36) was radiolabeled during transcription by incorporation of [α -³²P]CTP. It was incubated at 65°C with the unlabeled *P. abyssi* Pab91 sRNA, and the three *P. abyssi* proteins L7Ae (L), aCBF5 (C) and aGAR1 (G) with the WT (N) (filled circle) or variant aNOP10 proteins (indicated as in panel D). After T2 RNase digestion, the amount of Ψ formation was estimated by 1D-TLC analysis. (G) Analysis by EMSA of the amount of the RNP5 and CII complexes (36), formed with the radio-labeled Pab91 sRNA (50 fmol), proteins L7Ae, aCBF5 and either the WT or variants aNOP10. The formation of complex CII was obtained by incubation of the guide sRNA with the protein mixture and a 50 molar excess of an unlabeled RNA target, which carries a *P. abyssi* 23S rRNA fragment complementary to the Pab91 pseudouridylation pocket.

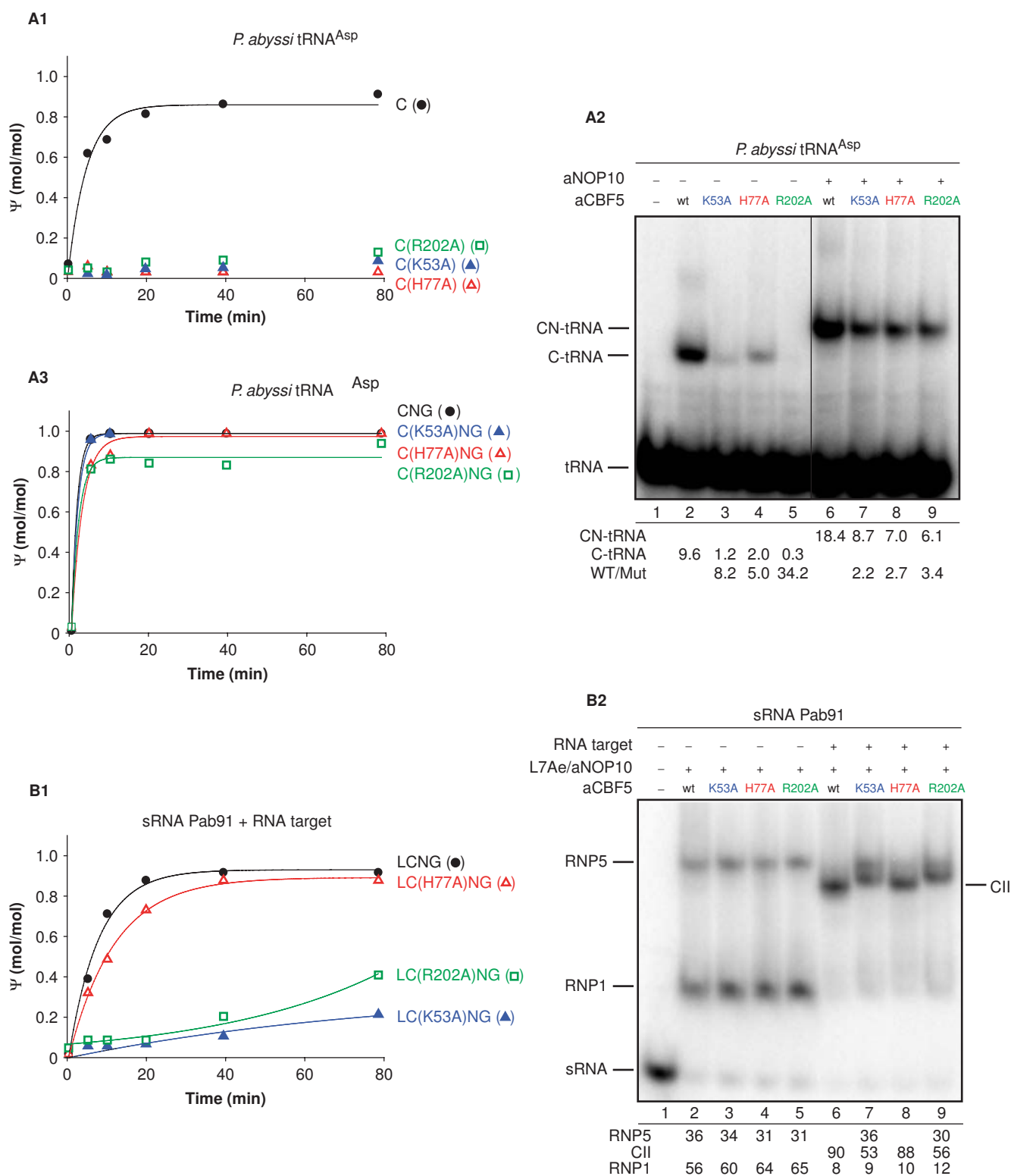


Figure 7. Effect of the substitution of residues in the aCBF5 catalytic site on its non-RNA-guided and RNA-guided activities. **(A1)** Time-course analysis of the non-RNA-guided Ψ_{55} formation in tRNA^{Asp} by the wild type (C) (filled circle), or K53A, H77A and R202A variant aCBF5 proteins (filled triangle, open triangle, open square, respectively). Analysis was performed as in figure 1. **(A2)** Analysis by EMSA of the complexes formed with the radiolabeled tRNA^{Asp}, the WT (C) or variant aCBF5 proteins in the presence or absence of aNOP10 (N). The percentages of bound tRNA in the presence of WT or mutated aCBF5 and in the absence (C-tRNA) or presence of aNOP10 (CN-tRNA) are given as well as the ratios between tRNA bound to the mutated aCBF5 proteins versus that bound to WT aCBF5. **(A3)** The same experiments as in panel A1 were performed but in the presence of the auxiliary proteins aNOP10 (N) and aGAR1 (G). **(B1)** Time-course analysis of an RNA-guided Ψ formation carried out as in Figure 6F with the wild type or variant K53A, H77A and R202A aCBF5 proteins. **(B2)** Analysis by EMSA of the RNP5 and CII complexes. The experiments were performed as in Figure 6G, except that the variant aCBF5 proteins are used.

The heterologous helix P2S, that is formed with the 5' strand of the pseudouridylation pocket, is basically stacked with the upperstem of the sRNA (P2) (47,48). The similarity between the T stem-loop structure recognized by TruB and the P2S heterologous helix formed by the guide RNA and the RNA substrate was pointed out (30,31). The tRNA acceptor stem flanked by the CCA motif would correspond to helix P1 of the sRNA and its conserved ACA motif, respectively. Based on this assumption, the finding that helix P1S, but not P2S, is stacked on the P1 stem of the sRNA can explain the requirement of a flexible link between the T and acceptor stem for aCBF5 activity. On the basis of this model of interaction between aCBF5 and tRNAs, the CCA motif of the tRNA, referred to the PUA domain, likely has the same position as the conserved ACA motif of H/ACA sRNAs (30). Therefore, like in TruB, the PUA domain of aCBF5 may interact with both the T and acceptor stems and the CCA motif, which explains the requirement of the acceptor stem and the 3' terminal CCA for activity of the free aCBF5 protein (Figures 3 and 4). It should be pointed out that although TruB is active on the T stem-loop alone, the solved 3D structures of the RNA free and RNA bound TruB revealed, that upon tRNA binding, the PUA domain in TruB moves towards the active site and also makes interactions with residues of the acceptor stem (25). In the H/ACA sRNP structure, the two conserved adenosines of the ACA motif developed both stacking interactions and hydrogen bonds with several residues of the aCBF5 PUA domain. In particular, the 5' proximal A residue inserts into an RNA-protein hybrid pocket formed between ribose moieties of the RNA minor groove and protein residues (30). Pocket formation is induced by RNA binding. Our gel-shift assays demonstrate a lower affinity of aCBF5 for a tRNA with an authentic CCA 3' terminal sequence compared to a mutated tRNA with an ACA 3' terminal sequence. Hence, an important issue would be to know how the RNA-protein interaction takes place when the proximal A residue in ACA is replaced by a C residue. The lower affinity for a CCA sequence compared to an ACA sequence may explain the needed interaction with the anticodon stem-loop for full activity of the free aCBF5 enzyme. This interaction may mimic that of aCBF5 with helix P2 in the sRNP. Interestingly, ACA to CCA substitution in a sRNA decreased substantially the amounts of aCBF5-sRNA complexes (data not shown). As aCBF5 is expected to have evolved from a pre-existing free standing enzyme, it would be interesting to investigate by 3D structure analysis of an aCBF5-tRNA complex, how the PUA domain of the pre-existing enzyme evolved in order to recognize both tRNAs and sRNAs.

It should also be pointed out that another possible explanation for the absence of activity on tRNAs of an aCBF5 without PUA domain may be the low thermostability of the truncated protein. Indeed, in contrast to the full-length aCBF5, the truncated aCBF5 cannot be enriched in the *E. coli* cellular extract by heating at 65°C. It precipitated with the *E. coli* proteins. Therefore, the PUA domain of aCBF5, which is wrapped by the

N-terminal extremity of the molecule, likely plays an important role for aCBF5 thermostability.

Proteins aNOP10 and aGAR1 but not L7Ae reinforce the activity of aCBF5 on tRNAs

Whereas the individual addition of aNOP10 or aGAR1 stimulates the kinetics of aCBF5 activity on tRNA (Figure 1), no stimulatory effect was detected in the presence of L7Ae (Figure S1). A reasonable explanation is the absence of stable interaction between aCBF5 and L7Ae, while both aNOP10 and aGAR1 directly interact with this protein, as well in the presence, as in the absence of sRNA (35,36). Indeed, a direct interaction is formed between aNOP10 and L7Ae in the H/ACA sRNP structure (30), but this interaction is not stable by itself, it depends upon the binding of L7Ae on the K-turn or K-loop structure of the H/ACA sRNA. Based on these observations, it is reasonable to propose that aCBF5 can either act alone on tRNAs or in association with aNOP10 and aGAR1 *in vivo*. *In vitro*, the stimulatory effects of both aNOP10 and aGAR1 are particularly strong for tRNAs lacking the CCA sequence (Figure 3) and each of these proteins can compensate for the absence of this sequence. Therefore, it might be that in archaeal cells, the aCBF5-aNOP10, aCBF5-aGAR1 or aCBF5-aNOP10-aGAR1 complexes would modify archaeal tRNA precursors, before activity of the CCA-terminal transferase (49).

Our gel-shift assays strongly suggest that the activation property of aNOP10 on aCBF5 results in part from its capability to increase tRNA binding to aCBF5. In the 3D structure of an H/ACA sRNP, aNOP10 only forms a few contacts with the guide sRNA (30). However, aNOP10 is required for an efficient association of the substrate RNA on the complex formed by the guide sRNA and aCBF5 (36). Similarly, the tight interaction that aNOP10 established with aCBF5 (27–30) probably favors the tRNA recruitment (Figure 5). This aNOP10 activity together with the peculiarity of the aCBF5 PUA domain may compensate for the absence of the Ins1 and Ins2 segments in the aCBF5 catalytic domain as compared to the TruB catalytic domain. The Ins1 and Ins2 segments play a crucial role for TruB activity, since Ins2 forms the floor of the T Ψ C loop binding cavity and the Ins1 a binding thumb clamping this loop in the cavity (31). The disappearance of these two segments in aCBF5 is likely required to accommodate in the active site the sRNP four-helix-bundle structure formed upon annealing of a target RNA to the sRNA pseudouridylation pocket. The lack of Ins1 and Ins2 segments in aCBF5 may explain the strong requirement for the CCA 3'-terminal motif, the acceptor stem and anticodon stem-loop for full activity of the free aCBF5 protein on tRNAs. It also likely explains the enhancer effect on activity of aNOP10, which increases the tRNA affinity. Up to now, the role of aGAR1 in the sRNA-guided system is less understood. GAR1 interacts with the RNA substrate in eukaryal snoRNP (50). However, this interaction seems to depend upon an additional domain present in the eukaryal protein as

compared to the archaeal protein. The protein aGAR1 also develops strong contact with the aCBF5 catalytic domain (29,30). Nonetheless, this may not help to reinforce the capability of aCBF5 to bind tRNA. In addition, aGAR1 was proposed to increase the turnover of the H/ACA sRNP (36). Therefore, this activity may also be needed to reinforce the aCBF5 tRNA:Ψ55-synthase activity.

Different networks of interaction are involved in the guided and non-guided activities of aCBF5

As mentioned above, the limited number of contacts between aNOP10 and the RNA in the H/ACA sRNP (30) suggests that the observed reinforcement of tRNA association in the presence of aNOP10 largely results from subtle conformational changes in the aCBF5 catalytic domain upon aNOP10–aCBF5 interaction. Superimposition of the 3D structures of the free and RNA-bound TruB enzyme (51) with the 3D structure of aCBF5 within the aCBF5–aNOP10 heterodimer shows a comparable positioning of the catalytic residues (27–30). In the absence of 3D structure of the free aCBF5 protein, we assume that in the free aCBF5 enzyme, an induced fit process has to take place upon tRNA binding for the optimization of the active site conformation. Interestingly, our data show that distinct residues of aNOP10 are involved in activation of the RNA-guided and non-guided activities of aCBF5 (Figure 6). Two conserved residues Y14 and H31 of aNOP10 were proposed to act in a common network of interactions linking aNOP10 to residue R202 in the aCBF5 catalytic site (29). Our data strongly suggest that the correct positioning of residue R202 in aCBF5 within the H/ACA sRNP involves residue H31 but not residue Y14, whereas the correct positioning of residue R202 for efficient tRNA:Ψ55-synthase activity depends upon residue Y14. Hence, conservation of these two residues in the course of evolution may reflect the conservation of the two enzymatic activities of aCBF5. Our data demonstrate a functional role in both the tRNA:Ψ55-synthase activity and the sRNA-guided activities of the conserved K residue in the aCBF5 motif I (K53 in aCBF5 and its counterpart K19 in *E. coli* TruB) (Figure 7). This residue is involved in the binding of the tRNA substrate and in the recruitment of the target RNA of the sRNP. Our data also show that the conserved P32 residue in aNOP10, which was proposed to link aNOP10 to the aCBF5 motif I (28), is essential, neither for the tRNA:Ψ55-synthase activity nor for the sRNA-guided activity. Another illustration of the difference between the aCBF5 RNA-guided and non-guided activities is the capability of aNOP10 to compensate for the absence of the lateral chains of residues K53 and R202 in the aCBF5 tRNA:Ψ55-synthase activity, but not in the sRNA-guided aCBF5 activity (Figure 7A3 and B1). Similarly, the substitution of residue K19 only has a mild effect on the TruB activity and leads to a thermosensitivity of the enzyme (52). Interestingly, the mutation of the conserved proline in motif I in the yeast Cbf5p also leads to a thermosensitivity of this enzyme (52,53). In light of these

data, the positive effect of aNOP10 on the K53A aCBF5 variant may consist in a structural stabilization of motif I in the aCBF5–aNOP10 complex, allowing the reaction to proceed at the 65°C incubation temperature.

Residue R202 is specifically conserved in members of the TruB family. Its lateral chain is expected to interact with the 5' phosphate of the targeted U residue (31). In agreement with these findings, we found that the R202A substitution affects tRNA association with aCBF5 and the binding of the target RNA on H/ACA sRNPs (Figure 7). Here again, protein aNOP10 can compensate for the R202A substitution. Another residue in the TruB catalytic center (Y76 from motif II) is expected to interact with the targeted U residue and is proposed to act as a general base during the catalytic mechanism (54). Other residues in aCBF5, like Y76 and/or Y179 which are expected to be located in the vicinity of the targeted U residue, can participate to the stabilization of the targeted residue. Interaction of aNOP10 with aCBF5 may favor their activity by subtle conformational changes in the aCBF5 active site.

Conservation of the crucial flipping mechanism in aCBF5

Upon tRNA binding on TruB, the uracil base at position 55 in the TΨC loop is extruded from the tRNA structure by a base-flipping mechanism that opens the tRNA 3D structure. Residue H43 probably plays a crucial role in this activity (31). Our observation that the mutation of residue H77, the counterpart of H43 in aCBF5, abolishes its tRNA:Ψ55-synthase activity strongly suggests that a similar mechanism is used for tRNA modification by aCBF5. Interestingly, aNOP10 can compensate for the H77A substitution. It would be highly interesting to determine by structural analysis how residue U55 is flipped out in the aCBF5 H77A variant associated with aNOP10. We also found that the RNA-guided activity is not impaired by the H77A substitution and this may be due to the fact that extrusion of the targeted U residue in H/ACA sRNPs is mainly mediated by the interaction between the guide and the target RNAs (47,48).

In spite of the absence of requirement of residue H77 for H/ACA sRNP activities, it should be noted that this residue is present in the eukaryal orthologs of aCBF5. This raises the question of a possible activity of the eukaryal Cbf5p/Dyskerin enzyme at position 55 in some eukaryal tRNAs. In connection with this observation, it should be pointed out that Cbf5 was found to co-purify with the selenocysteine tRNA in the protist *Euglena gracilis* (55).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

I. Motorin is thanked for helpful discussion. S.M. and J.-B.F. are fellows from the French Ministère de la Recherche et des Nouvelles Technologies (MNRT).

The work was supported by the Centre National de la Recherche Scientifique (CNRS), the MNRT and the PRST « Bioingénierie » of Région Lorraine. Funding to pay the Open Access publication charges for this article was provided by CNRS.

Conflict of interest statement. None declared.

REFERENCES

1. Rozenski, J., Crain, P.F. and McCloskey, J.A. (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res.*, **27**, 196–197.
2. Decatur, W.A. and Fournier, M.J. (2002) rRNA modifications and ribosome function. *Trends Biochem. Sci.*, **27**, 344–351.
3. King, T.H., Liu, B., McCully, R.R. and Fournier, M.J. (2003) Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol. Cell*, **11**, 425–435.
4. Sumita, M., Desaulniers, J.P., Chang, Y.C., Chui, H.M., Clos, L. II and Chow, C.S. (2005) Effects of nucleotide substitution and modification on the stability and structure of helix 69 from 28S rRNA. *RNA*, **11**, 1420–1429.
5. Badis, G., Fromont-Racine, M. and Jacquier, A. (2003) A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *RNA*, **9**, 771–779.
6. Meroueh, M., Grohar, P.J., Qiu, J., SantaLucia, J.Jr, Scaringe, S.A. and Chow, C.S. (2000) Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of *Escherichia coli* 23S rRNA. *Nucleic Acids Res.*, **28**, 2075–2083.
7. Kirpekar, F., Hansen, L.H., Rasmussen, A., Poehlsgaard, J. and Vester, B. (2005) The archaeon *Haloarcula marismortui* has few modifications in the central parts of its 23S ribosomal RNA. *J. Mol. Biol.*, **348**, 563–573.
8. Del Campo, M., Recinos, C., Yanez, G., Pomerantz, S.C., Guymon, R., Crain, P.F., McCloskey, J.A. and Ofengand, J. (2005) Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of *Haloarcula marismortui* and *Deinococcus radiodurans*. *RNA*, **11**, 210–219.
9. Yu, Y.T., Shu, M.D. and Steitz, J.A. (1998) Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J.*, **17**, 5783–5795.
10. Yang, C., McPheeters, D.S. and Yu, Y.T. (2005) Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **280**, 6655–6662.
11. Hamma, T. and Ferre-D'Amare, A.R. (2006) Pseudouridine synthases. *Chem. Biol.*, **13**, 1125–1135.
12. Kammen, H.O., Marvel, C.C., Hardy, L. and Penhoet, E.E. (1988) Purification, structure, and properties of *Escherichia coli* tRNA pseudouridine synthase I. *J. Biol. Chem.*, **263**, 2255–2263.
13. Nurse, K., Wrzesinski, J., Bakin, A., Lane, B.G. and Ofengand, J. (1995) Purification, cloning, and properties of the tRNA Psi 55 synthase from *Escherichia coli*. *RNA*, **1**, 102–112.
14. Wrzesinski, J., Bakin, A., Nurse, K., Lane, B.G. and Ofengand, J. (1995) Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from *Escherichia coli*. *Biochemistry*, **34**, 8904–8913.
15. Wrzesinski, J., Nurse, K., Bakin, A., Lane, B.G. and Ofengand, J. (1995) A dual-specificity pseudouridine synthase: An *Escherichia coli* synthase purified and cloned on the basis of its specificity for Ψ 746 in 23S RNA is also specific for Ψ 32 in tRNA^{Phe}. *RNA*, **1**, 437–448.
16. Koonin, E.V. (1996) Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res.*, **24**, 2411–2415.
17. Kaya, Y. and Ofengand, J. (2003) A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA*, **9**, 711–721.
18. Behm-Ansmant, I., Urban, A., Ma, X., Yu, Y.-T., Motorin, Y. and Branlant, C. (2003) The *Saccharomyces cerevisiae* U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. *RNA*, **9**, 1371–1382.
19. del Campo, M., Ofengand, J. and Malhotra, A. (2004) Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of *Escherichia coli*. *RNA*, **10**, 231–239.
20. Kaya, Y., Del Campo, M., Ofengand, J. and Malhotra, A. (2004) Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *J. Biol. Chem.*, **279**, 18107–18110.
21. Huang, L., Pookanjanatavip, M., Gu, X. and Santi, D.V. (1998) A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry*, **37**, 344–351.
22. Ramamurthy, V., Swann, S.L., Paulson, J.L., Spedaliere, C.J. and Mueller, E.G. (1999) Critical aspartic acid residues in pseudouridine synthases. *J. Biol. Chem.*, **274**, 22225–22230.
23. Conrad, J., Niu, L., Rudd, K., Lane, B.G. and Ofengand, J. (1999) 16S ribosomal RNA pseudouridine synthase RsaA of *Escherichia coli*: deletion, mutation of the conserved Asp102 residue, and sequence comparison among all other pseudouridine synthases. *RNA*, **5**, 761–763.
24. Raychaudhuri, S., Niu, L., Conrad, J., Lane, B.G. and Ofengand, J. (1999) Functional effect of deletion and mutation of the *Escherichia coli* ribosomal RNA and tRNA pseudouridine synthase RluA. *J. Biol. Chem.*, **274**, 18880–18886.
25. Ferre-D'Amare, A.R. (2003) RNA-modifying enzymes. *Curr. Opin. Struct. Biol.*, **13**, 49–55.
26. Aravind, L. and Koonin, E.V. (1999) Novel predicted RNA-binding domains associated with the translation machinery. *J. Mol. Evol.*, **48**, 291–302.
27. Manival, X., Charron, C., Fourmann, J.B., Godard, F., Charpentier, B. and Branlant, C. (2006) Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res.*, **34**, 826–839.
28. Hamma, T., Reichow, S.L., Varani, G. and Ferre-D'Amare, A.R. (2005) The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat. Struct. Mol. Biol.*, **12**, 1101–1107.
29. Rashid, R., Liang, B., Baker, D.L., Youssef, O.A., He, Y., Phipps, K., Terns, R.M., Terns, M.P. and Li, H. (2006) Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol. Cell*, **21**, 249–260.
30. Li, L. and Ye, K. (2006) Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature*, **443**, 302–307.
31. Hoang, C. and Ferre-D'Amare, A.R. (2001) Cocystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell*, **107**, 929–939.
32. Bachellerie, J.P., Cavaille, J. and Huttenhofer, A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
33. Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
34. Omer, A.D., Ziesche, S., Decatur, W.A., Fournier, M.J. and Dennis, P.P. (2003) RNA-modifying machines in archaea. *Mol. Microbiol.*, **48**, 617–629.
35. Baker, D.L., Youssef, O.A., Chastkofsky, M.I., Dy, D.A., Terns, R.M. and Terns, M.P. (2005) RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev.*, **19**, 1238–1248.
36. Charpentier, B., Muller, S. and Branlant, C. (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res.*, **33**, 3133–3144.
37. Khanna, M., Wu, H., Johansson, C., Caizergues-Ferrer, M. and Feigon, J. (2006) Structural study of the H/ACA snoRNP components Nop10p and the 3' hairpin of U65 snoRNA. *RNA*, **12**, 40–52.
38. Rozhdetsvensky, T.S., Tang, T.H., Tchirkova, I.V., Brosius, J., Bachellerie, J.P. and Huttenhofer, A. (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif

- for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.*, **31**, 869–877.
39. Hamma, T. and Ferre-D'Amare, A.R. (2004) Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. *Structure*, **12**, 893–903.
 40. Roovers, M., Hale, C., Tricot, C., Terns, M.P., Terns, R.M., Grosjean, H. and Droogmans, L. (2006) Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic Acids Res.*, **34**, 4293–4301.
 41. Charpentier, B., Fourmann, J.B. and Branlant, C. (2007) Reconstitution of archaeal H/ACA sRNPs and test of their activity. *Methods Enzymol.*, **425**, 389–405.
 42. Bakin, A. and Ofengand, J. (1993) Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center - analysis by the application of a new sequencing technique. *Biochemistry*, **32**, 9754–9762.
 43. Massenet, S., Ansmant, I., Motorin, Y. and Branlant, C. (1999) The first determination of pseudouridine residues in 23S ribosomal RNA from hyperthermophilic Archaea *Sulfolobus acidocaldarius*. *FEBS Lett.*, **462**, 94–100.
 44. Mougin, A., Gregoire, A., Banroques, J., Segault, V., Fournier, R., Brule, F., Chevrier-Miller, M. and Branlant, C. (1996) Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *RNA*, **2**, 1079–1093.
 45. Motorin, Y., Muller, S., Behm-Ansmant, I. and Branlant, C. (2007) Identification of modified residues in RNAs by reverse transcription-based methods. *Methods Enzymol.*, **425**, 21–53.
 46. Gu, X., Yu, M., Ivanetich, K.M. and Santi, D.V. (1998) Molecular recognition of tRNA by tRNA pseudouridine 55 synthase. *Biochemistry*, **37**, 339–343.
 47. Jin, H., Loria, J.P. and Moore, P.B. (2007) Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Mol. Cell*, **26**, 205–215.
 48. Wu, H. and Feigon, J. (2007) H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *Proc. Natl Acad. Sci. USA*, **104**, 6655–6660.
 49. Tomita, K., Ishitani, R., Fukai, S. and Nureki, O. (2006) Complete crystallographic analysis of the dynamics of CCA sequence addition. *Nature*, **443**, 956–960.
 50. Wang, C. and Meier, U.T. (2004) Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.*, **23**, 1857–1867.
 51. Pan, H., Agarwalla, S., Moustakas, D.T., Finer-Moore, J. and Stroud, R.M. (2003) Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proc. Natl Acad. Sci. USA*, **100**, 12648–12653.
 52. Spedaliere, C.J., Hamilton, C.S. and Mueller, E.G. (2000) Functional importance of motif I of pseudouridine synthases: mutagenesis of aligned lysine and proline residues. *Biochemistry*, **39**, 9459–9465.
 53. Zebajarian, Y., King, T., Fournier, M.J., Clarke, L. and Carbon, J. (1999) Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Mol. Cell. Biol.*, **19**, 7461–7472.
 54. Phannachet, K., Elias, Y. and Huang, R.H. (2005) Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase. *Biochemistry*, **44**, 15488–15494.
 55. Russell, A.G., Schnare, M.N. and Gray, M.W. (2004) Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis*. *RNA*, **10**, 1034–1046.

Supplementary data

Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Ψ55-synthase and RNA-guided RNA:Ψ-synthase activities

S. Muller, J.-B. Fourmann, C. Loegler, B. Charpentier, and C. Branlant

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 UHP-CNRS,

Université des Sciences et Techniques Henri Poincaré Nancy I, 54506 Vandoeuvre-Lès-

Nancy cedex, France.

Figure S1: Analysis of the influence of protein L7Ae on Ψ55 formation in *P. abyssi* tRNA^{Asp}. The tRNAs were labeled by incorporation of [α -³²PCTP] during transcription. The ³²P-labeled tRNA substrate (50 fmol) was mixed with aCBF5 (C), aCBF5-aNOP10 (CN), or the aCBF5-aNOP10-aGAR1 complex (CNG) in the absence or the presence of protein L7Ae (L) (LCN and LCNG). Incubation was performed for 80 min at 65°C. After T2 RNase digestion, the amount of Ψ residue formed was measured by 2D TLC and expressed in moles per moles of tRNA.

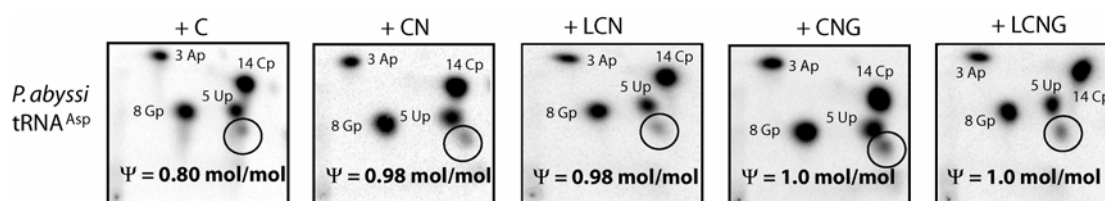


Figure S1

Muller et al.

C. Résultats complémentaires

1. Le complexe aCBF5/aNOP10/aGAR1 est capable de modifier les ARNt de différentes espèces

Nous montrons ici des résultats annoncés comme données non montrées dans l'article N°3 et qui concernent la capacité du complexe aCBF5/aNOP10/aGAR1 à modifier différents ARNt issus de différentes espèces et même un précurseur d'ARNt contenant un intron. Nous avons en effet cloné sous contrôle d'un promoteur T7 les gènes des ARNt : deux isotypes différents provenant de *P. abyssi* : les ARNt^{Tyr} et ARNt^{Asp}, un ARNt contenant un intron : l'ARNt^{Tyr} de *S. solfataricus*, et des ARNt de bactérie et de levure : l'ARNt^{Gly} de *E. coli* et l'ARNt^{Asp} de *S. cerevisiae*. Ceci nous a permis de tester l'activité de complexes aCBF5/aNOP10/aGAR1 sur les produits de transcription de ces gènes (Fig. R.20).

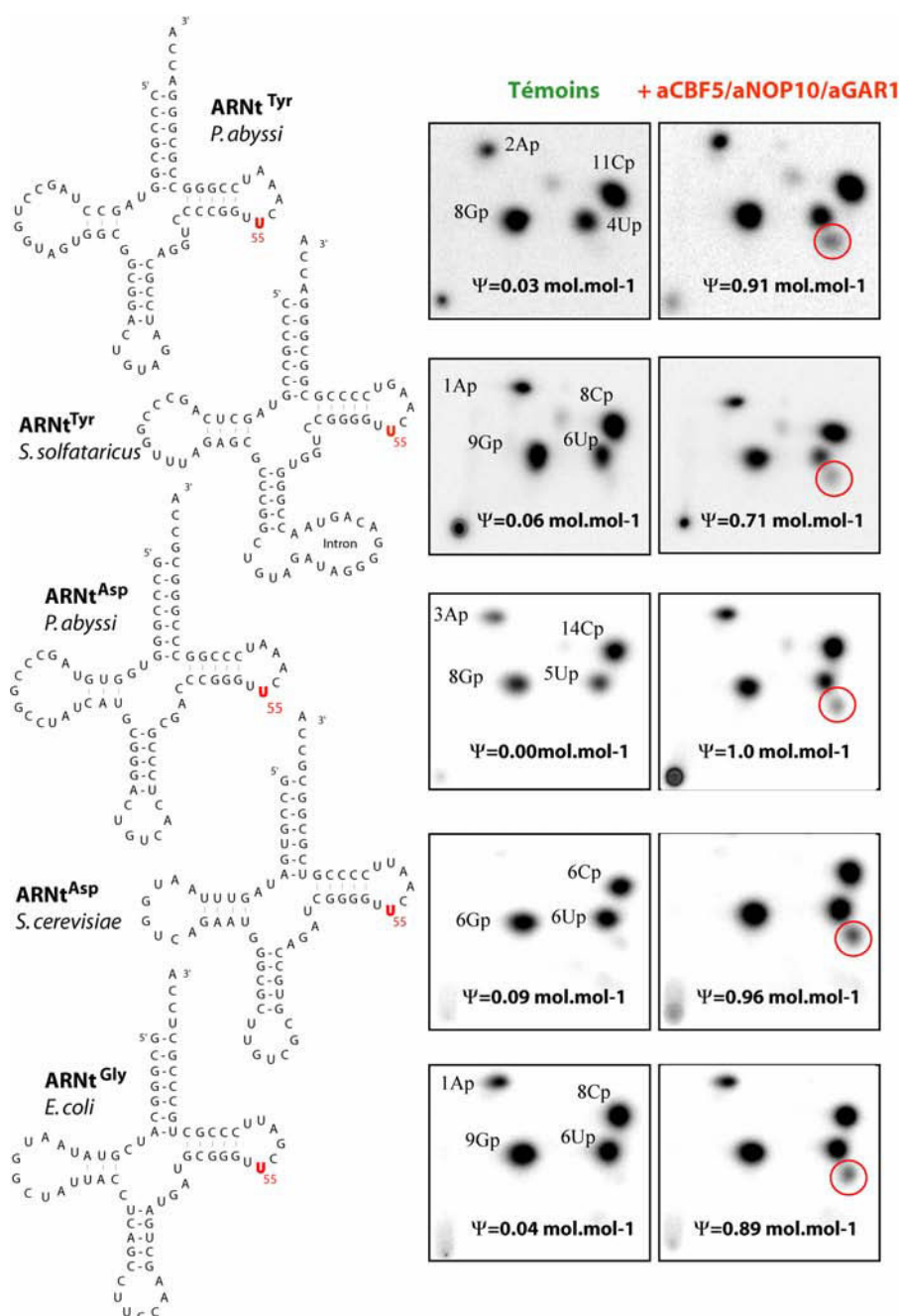


Figure R.20 : Conversion du résidu U en position 55 de plusieurs ARNt en résidu Ψ par action du trimère aCBF5/aNOP10/aGAR1.

Les tests d'activité ont été réalisés comme décrit dans la publication N°1 et dans la partie Matériel & Méthodes. Le taux de modification à la position 55 a été estimé en tenant compte du nombre de résidus A, C, G et U marqués dans chaque ARNt.

Nous montrons que le complexe des 3 protéines de *P. abyssi* est capable de modifier à un taux élevé (entre 70 et 96%) le résidu 55 des ARNt quelque soit leur espèce d'origine et ceci à 65°C. La structure 2D des ARNt d'*E. coli* et de *S. cerevisiae* à cette température est suffisamment stable pour que la structure en tige-boucle du bras TΨC soit reconnue par le complexe de protéines. Il est intéressant de noter que la présence d'un intron dans l'ARNt^{Tyr} de *S. solfataricus* affecte peu l'efficacité de conversion par rapport à celle obtenue pour les ARNt dépourvus d'introns (Fig. R.20).

2. Etude de l'efficacité relative d'association des protéines aCBF5 et aNOP10 avec les ARNt^{Asp} de *P. abyssi* et de *S. cerevisiae* et leurs variants

Dans l'article, nous avons mesuré l'action d'aCBF5 seule ou en complexe avec aNOP10 et aGAR1 sur différents variants de l'ARNt^{Asp} de *S. cerevisiae*, dont certains nous avaient été fournis par Catherine Florentz (IBMC, Strasbourg), ainsi que sur l'ARNt^{Asp} de *P. abyssi* et ses variants. Ici, nous montrons l'étude, par des expériences de retard sur gel, de l'association des protéines aCBF5, aCBF5 D82A et aNOP10 sur l'ARNt^{Asp} sauvage et sur différents variants de cet ARNt pour lesquels l'activité enzymatique du trimère aCBF5/aNOP10/aGAR1 a été testée dans la publication N°3 (Fig. R.21).

Les données des expériences de retard sur gel présentées dans la figure R.21 suggèrent fortement la fixation de 2 molécules d'aCBF5 sur la plupart des ARNt étudiés. Néanmoins, dans certains cas tels que pour le sauvage, en absence du bras TΨC ou du bras accepteur, la fixation d'une seule molécule est majoritaire. La fixation de 2 molécules est prédominante dans le cas où le bras anticodon est absent. Les données obtenues pour le variant dépourvu de bras TΨC (Δ T-tRNA^{Asp}, Fig. R.21) montrent que ce bras TΨC ne semble pas être indispensable pour qu'il y ait fixation de l'enzyme. La protéine aCBF5 semble donc avoir une reconnaissance large des ARNt. Il est intéressant de noter que l'ajout de aNOP10 limite le nombre de molécules d'aCBF5 fixées sur une molécule d'ARNt puisqu'un seul complexe est alors formé. Notons que l'absence du bras D (Δ D-tRNA^{Asp}, Fig. R.21) renforce à la fois la fixation de aCBF5 et celle du dimère aCBF5/aNOP10, ceci sans doute du fait de l'absence d'interaction de la boucle TΨC avec la boucle D. Ceci est en bon accord avec les données d'activité montrant que aCBF5 seule est active sur ce substrat et que l'activité est légèrement renforcée en présence de aNOP10 et aGAR1. En ce qui concerne l'ARNt^{Asp} de *P. abyssi*, la fixation de aCBF5 seule est faible sur l'ARNt sauvage et aNOP10 est nécessaire pour renforcer la fixation. Par contre, la conversion du CCA 3' terminal en ACA renforce fortement la fixation d'aCBF5. De manière surprenante, alors qu'aCBF5 seule se fixe au semi-tRNA^{Asp} de *P. abyssi*, la présence de aNOP10 ne permet pas la visualisation d'un complexe semi-ARNt^{Asp}-aCBF5-aNOP10 et empêche également la visualisation d'un complexe semi-ARNt^{Asp}-aCBF5 ceci en dépit de l'activité du complexe aCBF5/aNOP10/aGAR1 sur cet ARN.

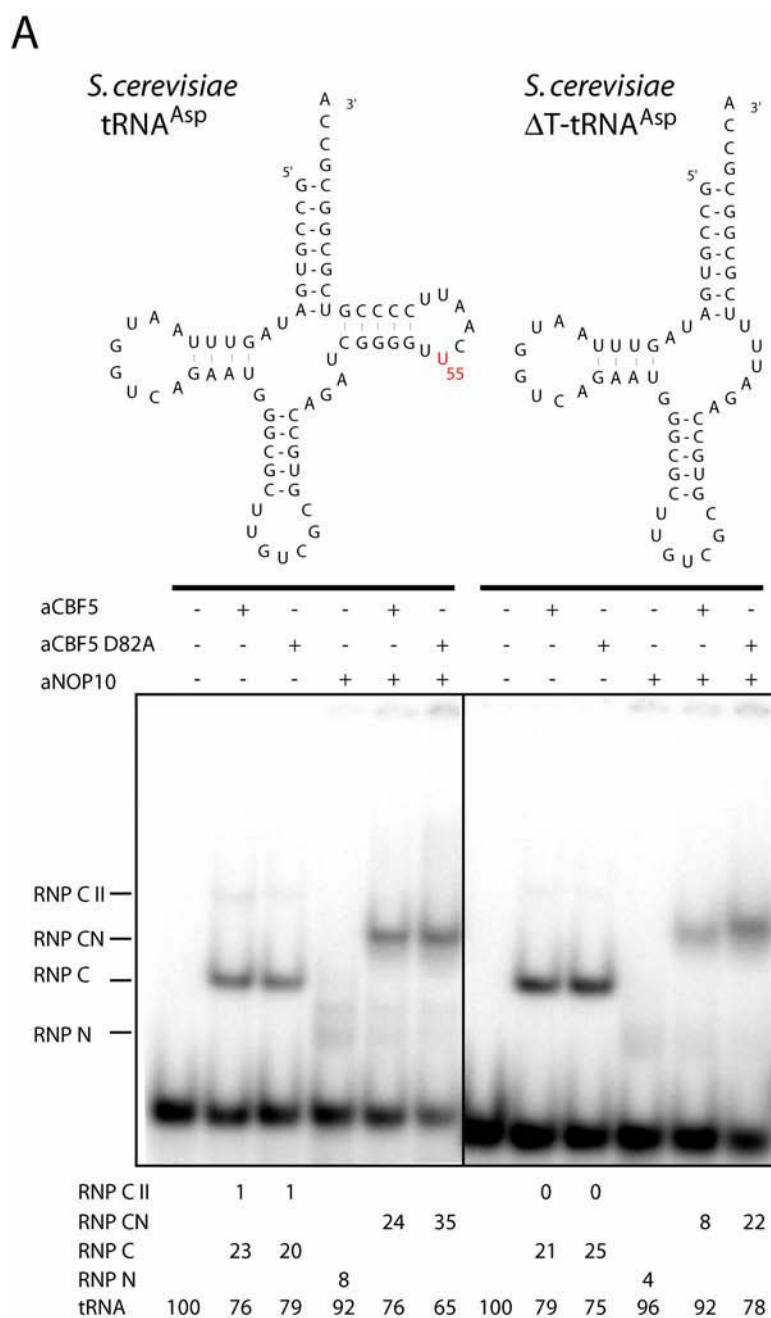


Figure R.21 : Analyse, par expériences de retard sur gel, de l'interaction de aCBF5 et aNOP10 avec l'ARNt^{Asp} et ses variants.

A. et **B.** Etude de l'ARNt^{Asp} sauvage de *S. cerevisiae* et de ses variants présentés dans la publication N°3. **C.** Etude de l'ARNt^{Asp} sauvage de *P. abyssi* et de ses variants présentés dans la publication N°3. Les complexes ont été formés avec 50 fmol d'ARNt marqué, en présence de protéines à une concentration de 200 nM dans les conditions présentées dans la partie Matériel & Méthodes. La formation du complexe a été testée par électrophorèse en gel de polyacrylamide non dénaturant à 6%. Les positions des différents complexes obtenus correspondant à la fixation de aNOP10 (RNP N), aCBF5 ou son variant D82A (RNP C et RNP CII) et celle du complexe aCBF5/aNOP10 (RNP CN) est indiquée. Le pourcentage d'ARN présent dans chaque complexe par rapport à la quantité totale d'ARN présente dans la piste est indiqué en dessous de chaque piste.

B

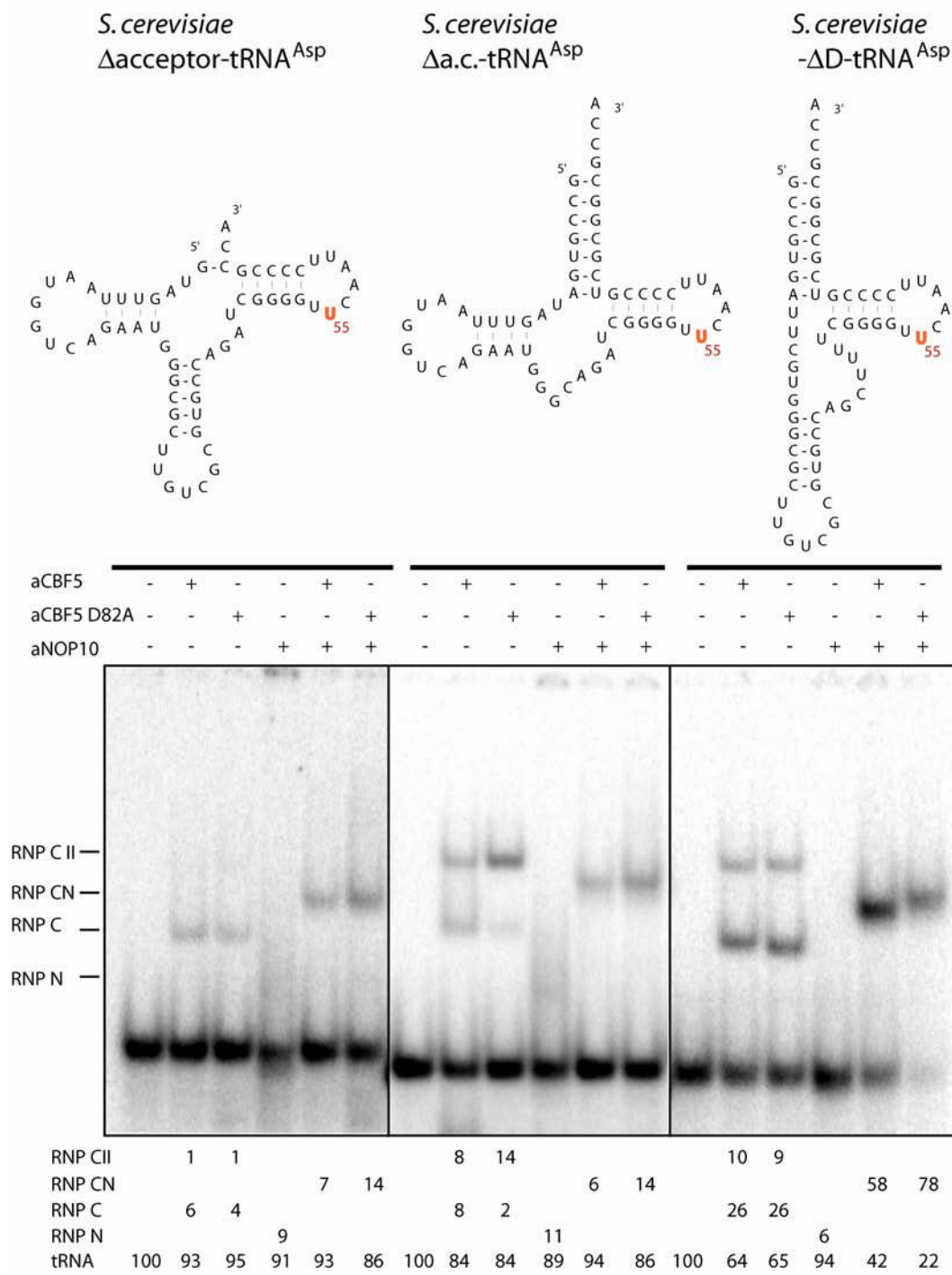


Figure R.21 (suite)

P. abyssi tRNA^{Asp}

P. abyssi tRNA^{Asp}-ACA

P. abyssi semi-tRNA^{Asp}

aCBF5

aCBF5 D82A

aNOP10

RNP C II

RNP CN

RNP C

RNP N

0 1 10 26

2 4

1

100 98 95 99 90 74

2 6 63 63 19 70 0 0

60 51 0 0

100 38 43 100 37 37 100 81 30 100 100 100

3. L'ARN : Ψ -synthase aPus10 peut agir en position 55 et 54 des ARNt

Dans l'article publié par Roovers *et al.* (2006) le taux de résidu Ψ formé par l'action de l'enzyme aPus10 sur l'ARNt^{Asp} de *P. furiosus* était de 120%, ce qui suggérait une action de aPus10 au niveau de plusieurs positions dans cet ARNt.

Nous avons produit et purifié la protéine aPus10 de *P. abyssi* et j'ai réalisé des essais de modification de l'ARNt^{Asp} de *P. abyssi*, marqué par incorporation de $[\alpha\text{-}^{32}\text{P}]\text{-UTP}$, avec l'enzyme aPus10. Après hydrolyse par la RNase P1 qui libère des nucléotides 5' monophosphates (5'NMP), j'ai en effet observé un taux de modification de 130%, contre 100% quand la même expérience est réalisée en présence du complexe aCBF5/aNOP10/aGAR1 (CNG). Ces données confirmaient une action de aPus10 à plusieurs sites dans l'ARNt^{Asp} de *P. abyssi*. Une possibilité était que cette enzyme puisse modifier le résidu U54 contigu au résidu U55. J'ai utilisé la méthode du plus proche voisin en utilisant un marquage de l'ARNt^{Asp} substrat soit par incorporation de $[\alpha\text{-}^{32}\text{P}]\text{-UTP}$, soit par incorporation de $[\alpha\text{-}^{32}\text{P}]\text{-CTP}$. Dans le premier cas, après hydrolyse par la RNase T2 qui libère des mononucléotides 3' phosphate (3'NMP), la conversion du résidu U54 en Ψ pouvait être détectée car un seul doublet UU (54-55) est présent dans l'ARNt^{Asp}. Dans le second cas, la conversion du résidu U55 en Ψ pouvait être analysée.

Comme on peut le constater dans la figure R.22, alors que aPus10 et le complexe CNG agissent chacun en position 55, seule aPus10 est active en position 54, mais ceci à un faible taux (0,10 mol/mol).

J'ai répété le même type d'expérience avec le variant tronqué de l'ARNt^{Asp} de *P. abyssi* (semi-tRNA^{Asp}). Un marquage de l'ARN en position 55 par de l' $[\alpha\text{-}^{32}\text{P}]\text{-CTP}$ confirme la formation du résidu Ψ 55 dans cet ARN tronqué, ceci à la fois par l'action de aPus10 et par l'action du complexe CNG. Par conséquent, aPus10, comme le complexe CNG, est active en absence des bras D et anticodon de l'ARNt. Un marquage par du CTP d'un variant U55C conduit à la détection d'un taux de formation de Ψ en position 54 de 0,39 mol/mol alors qu'un marquage de l'ARN non muté en position 55 conduit à la détection d'un taux très élevé de formation de Ψ en position 54 (0,81 mol/mol). Il est donc vraisemblable que l'activité en position 54 soit favorisée par la présence d'un résidu U cible en position 55.

Par contre, de façon surprenante, lorsque nous avons utilisé comme substrat l'ARNt^{Asp} de *S. cerevisiae* dépourvu du bras anticodon, nous n'avons détecté aucune activité de aPus10 ni en position 55 ni en position 54. Dans ce substrat l'interaction entre les boucles T Ψ C et D peut avoir lieu mais avait aussi lieu dans l'ARNt^{Asp} entier de *P. abyssi*. Ces expériences complémentaires montrent donc l'existence de différences de spécificité entre aPus10 et le complexe CNG.

D. Discussion complémentaire et Perspectives

Comme deux autres équipes (Roovers *et al.*, 2006) nous avons montré l'action de aCBF5 en position 55 des ARNt. Nos données montrent en plus que l'association d'aCBF5 avec aNOP10 et aGAR1 permet de pallier à l'absence du triplet CCA dans l'ARNt et permet une action efficace d'aCBF5 sur des ARNt tronqués et même sur un demi-ARNt ne renfermant que les bras accepteur et bras TΨC. Nous montrons également que la protéine aNOP10 renforce l'association d'aCBF5 avec ces substrats. Ceci peut être en accord avec les capacités d'interaction avec l'ARN décrites pour aNOP10 sur la base de données RMN (Khanna *et al.*, 2006). Il est intéressant de noter que nous avons mis en évidence une activité du complexe aCBF5/aNOP10/aGAR1 sur le demi ARNt cité ci-dessus, alors que Roovers *et al.*, (2006) qui avaient testé l'activité sur un autre ARNt tronqué n'avaient pas observé d'activité. La différence entre ces deux substrats est l'existence d'une boucle assurant une flexibilité entre le bras accepteur et l'hélice du bras TΨC. Cette observation est en accord avec le fait que dans l'architecture formée entre le sRNA H/ACA et l'ARN cible, une jonction triple existe entre les deux empilements d'hélices PS1 sur H1 et PS2 sur H2 (Jin *et al.*, 2007; Wu et Feigon, 2007). Toujours dans le domaine des considérations structurales, on peut constater de fortes similarités entre la structure formée par le bras accepteur avec la structure tige-boucle TΨC d'un ARNt et celle issue de l'association entre le sRNA H/ACA et l'ARN cible (Hoang et Ferre-D'Amare, 2001). L'hélice H1 correspondrait au bras accepteur et l'appariement intermoléculaire correspondrait au bras TΨC.

Dans cette optique, le triplet CCA 3' terminal de l'ARNt correspond à la boîte ANA. La distance qui sépare cette boîte du dernier nucléotide apparié en aval du résidu U ciblé est exactement de 14 nucléotides (Fig. R.23). Notons que le nucléotide en amont de la séquence CCA 3' terminal est le plus souvent non apparié dans les ARNt et il existe aussi parfois un nucléotide non apparié à la même position entre l'hélice H1 et la boîte ACA dans certains sRNA H/ACA. Il faut aussi noter que le résidu U54 est supposé être apparié avec le résidu A58 et que cet appariement est absent dans l'ARNt initiateur qui lui n'est pas modifié par TruB chez les bactéries et Pus4 chez la levure (Becker *et al.*, 1997; Gu *et al.*, 1998). L'absence de cet appariement pourrait peut-être aussi expliquer une absence de modification de l'ARNt initiateur d'archaea par aCBF5 ou Pus10.

La distance de 14 nt observée entre le CCA 3'terminal et la position 55 des ARNt correspond à la distance la plus fréquente entre le triplet ANA du sRNA et le résidu cible de l'ARNr. Il faut remarquer que le nucléotide en 5' du résidu U à modifier dans l'ARN cible est toujours apparié à l'élément antisens 3' au sein de la poche de pseudouridylation. L'appariement correspondant dans l'ARNt est un déterminant de l'action de TruB/Pus4 (Becker *et al.*, 1997 ; Gu *et al.*, 1998). Sa présence est sans doute nécessaire à l'action de aCBF5. Ainsi, aCBF5 sans l'aide des protéines aNOP10 ou aGAR1 serait capable, à l'instar de TruB, de reconnaître cette partie de l'ARNt.

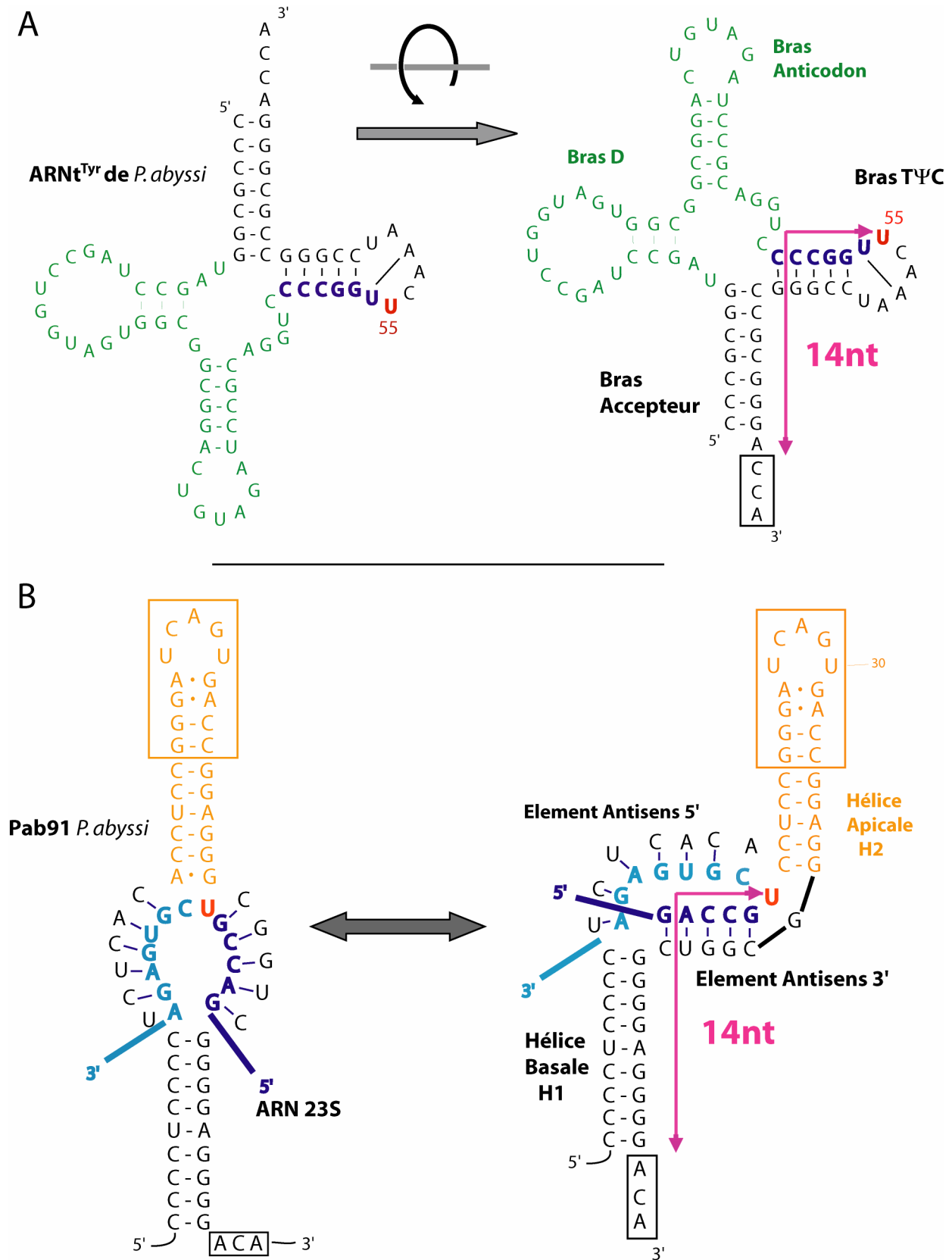


Figure R.23 : Analogie entre les structures 2D des ARNt et du duplex sRNA H/ACA-ARN cible.
 A. ARNt^{Tyr} de *P. abyssi*. B. sRNA H/ACA Pab91 de *P. abyssi*.

Aussi bien nos études sur les ARNt et leurs variants que celles sur les divers sRNA H/ACA de *P. abyssi* montrent que aCBF5 seule semble avoir une spécificité large d'interaction avec les ARN. En effet, nos expériences de retard sur gel montraient que même un ARNt tronqué de son bras TΨC se lie à aCBF5. Par ailleurs, nous avons observé la fixation de 2 molécules d'aCBF5 sur certains variants d'ARNt. Nos données montrent que aNOP10 permettrait de restreindre la spécificité de fixation de aCBF5. Il est vraisemblable que l'interaction aNOP10/aCBF5 induise des changements de conformation, qui favorisent l'interaction de aCBF5 avec ses cibles au détriment d'interactions aspécifiques. D'autre part, il n'est pas exclu que aNOP10 interagisse aussi avec l'ARN une fois associé à aCBF5. En effet, aNOP10 peut clairement compenser l'absence de trinuécléotides CCA 3' terminal dans l'ARNt et aNOP10 renforce fortement, comme nous l'avons montré par expériences de retard sur gel, le taux de fixation de aCBF5 à l'ARNt, en particulier si il est dépourvu de son CCA 3' terminal. Nous avons observé un effet activateur de aGAR1 sur l'action d'aCBF5 sur l'ARNt sans CCA 3' terminal similaire à celui d'aNOP10 et les deux effets sont cumulatifs. Dans l'état actuel des connaissances, il est difficile d'expliquer cet effet. En effet, aGAR1 ne semble pas interagir directement avec l'ARN. Par contre, l'interaction d'aGAR1 avec aCBF5 pourrait aussi induire un changement de conformation favorisant la fixation spécifique d'aCBF5 à son substrat. Ainsi, l'interaction de la boucle β7 d'aCBF5 avec aGAR1 pourrait maintenir le site actif d'aCBF5 dans une conformation ouverte favorisant à la fois l'entrée et la sortie du substrat. On peut imaginer que le fragment Ins1 de TruB absent dans aCBF5 joue un rôle équivalent à aGAR1. Le fait que des éléments présents dans TruB constituent des protéines individualisées dans le système reposant sur l'emploi d'ARN guide pourrait être lié à la fois à des contraintes structurales correspondant à la nécessité de fixer le duplex d'ARN sRNA H/ACA-rRNA, mais aussi à un besoin de renforcer la spécificité d'action d'une enzyme qui utilise un grand nombre de cofacteurs (sRNA H/ACA) pour modifier de manière spécifique un grand nombre de substrats différents. De la même manière, le passage du système intron-auto-épissable de type II au système qui fait intervenir des UsnRNA nécessite la mise en œuvre d'un très grand nombre de protéines qui jouent aussi un rôle majeur dans la spécificité d'action du système.

L'effet de aGAR1 reste inexploré. La protéine aGAR1 améliore la vitesse de réaction mais ne modifie pas le profil d'assemblage en gel retard. En effet, on ne visualise ni la formation d'un complexe de plus haut poids moléculaire, ni l'effet négatif observé dans les particules sRNP H/ACA (résultats non montrés).

Le fait que nous ayons observé que des acides aminés très conservés de aCBF5 soient importants pour le système non guidé, mais pas pour le système guidé et inversement soient important pour le système non guidé, mais pas pour le système guidé suggère que, même si aPus10 est capable de catalyser la conversion du résidu U55 en résidu Ψ, l'activité non guidée de aCBF5 a un rôle biologique et n'est pas un simple artefact d'études réalisées *in vitro*. D'un autre côté, aPus10 qui est un représentant d'une nouvelle famille d'ARN:Ψ-synthases, est présente, d'après notre analyse, dans tous les génomes d'archaea séquencés disponibles. La question du

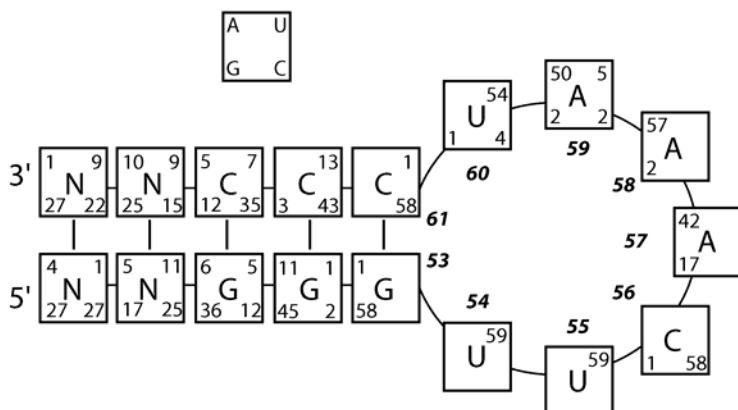
r le de aPus10 et du complexe aCBF5/aNOP10/aGAR1 reste donc ouverte. Tous les g nes d'ARNt d'archaea ne contiennent pas la s quence codant le triplet CCA 3' terminal. On peut imaginer que, dans ce cas l , le complexe aCBF5/aNOP10/aGAR1 qui a une forte activit , m me en absence de ce triplet, soit le catalyseur de la conversion U en   en position 55, ceci avant ajout du CCA terminal. On peut aussi imaginer que vu l'importance fonctionnelle de la modification en position 55 de l'ARNt, les archaea poss dent deux syst mes capables de catalyser sa formation. Les souches d'*E. coli* ne sont pas viables en absence d'expression de TruB. Chez *E. coli*, il n'existe pas d'homologue de aPus10. Les levures ne pr sentent pas non plus de repr sentant de Pus10.

Comme nous l'avons montr  dans le chapitre pr c dent, le complexe aCBF5/aNOP10/aGAR1 pourrait avoir un r le dans la modification des ARNr, puisque nous avons montr  son action possible *in vitro* en position 2603 de l'ARNr de *P. abyssi*. Les fonctions non guid es de ce complexe pourraient donc ne pas  tre limit es   l'ARNt. Par ailleurs, aPus10 pourrait, comme nous l'avons d j  d montr  pour la position 54 des ARNt, avoir une ou plusieurs autres fonctions que la modification en position 55 des ARNt. Il faut noter que la pr sence d'un r sidu   a  t  d tect e en position 54 de tous les ARNt d'archaea mais aussi dans de nombreux ARNt d'animaux qui ont  t   tudi s (Fig. R.24). Or, aucune enzyme responsable de cette modification n'avait  t  jusqu'  l  identifi e. Il est donc possible que aPus10 soit l'enzyme impliqu e. Il est int ressant de remarquer que la famille Pus10 comporte des repr sentants chez les animaux et les archaea mais est absente des autres phylums. On note donc une corr lation entre la pr sence de g nes codant Pus10 et la pr sence de r sidus   en position 54 (Fig. R.24). La faible efficacit  de modification en position 54 observ e *in vitro* pourrait  tre compens e par l'existence de prot ines favorisant l'ouverture de la structure 3D de l'ARNt et donc l'action de aPus10.

Nous avons montr  que le complexe aCBF5/aNOP10/aGAR1 agissait sur l'une des deux positions pseudouridyl es de l'ARNr 23S de *P. abyssi* (position 2603) pour lesquelles aucun sRNA H/ACA n'avait  t  identifi  mais pas sur la seconde (position 2585). Nous avons test  l'action de aPus10   chacune de ces deux positions, mais nous n'avons pas observ  d'activit . Il appara t donc que aPus10 et le complexe aCBF5/aNOP10/aGAR1 ont des sp cificit s diff rentes de reconnaissance de l'ARN.

Une information tr s importante qui aide   comprendre cette diff rence vient d' tre publi e,   savoir, la structure 3D de la prot ine PUS10 humaine (McCleverty *et al.*, 2007). Les cristaux ont  t  obtenus en absence d'ARNt substrat. L'enzyme comporte deux domaines. Un premier domaine N-terminal qui contient un doigt   Zinc ainsi qu'un domaine THUMP (THioUridine synthases, RNA M thyltransferases and Pseudouridine synthases). Un domaine catalytique en C-terminal, pr sente les motifs retrouv s dans les 5 autres familles d'ARN: -synthases, ainsi qu'un domaine « pouce » et un domaine « index ». Il est int ressant de noter que la structure 3D du domaine catalytique de TruB est celle qui, parmi les repr sentants des 5 familles d'ARN: -synthases, se superpose le mieux avec la structure 3D du domaine catalytique de PUS10.

A



B

Positions	54	55	56	57	58
Modifications					
Archaea	m1Ψ (50/59) Ψ (7/59)	Ψ (57/59)	Cm (58/59)	m1I (47/59)	m1A (2/59)
Animaux	T (57/109) Tm (13/109) Ψ (11/109)	Ψ (101/109)	-	-	m1A (100/109)
Plantes	T (31/44)	Ψ (44/44)	-	-	m1A (44/44)
Unicellulaires euc.	T (55/63)	Ψ (58/63)	-	-	m1A (49/63)
Bactéries	T (86/133)	Ψ (126/133)	-	-	m1A (8/133)

Figure R.24 : Distribution des nucléotides modifiés dans le bras TΨC des ARNt d'archaea (A) et d'animaux, de plantes, d'unicellulaires eucaryotes et de bactéries (B)

(adapté de la banque de modification des ARNt, Rozenski *et al.*, 1999).

Les nucléotides modifiés ont été recherchés sur 59, 109, 44, 63 et 133 ARNt provenant respectivement, des espèces d'archaea, d'animaux, de plantes, d'unicellulaires eucaryotes et de bactéries.

La structure 3D d'un complexe PUS10-ARNt a été modélisée. Dans ce modèle, la boucle TΨC est en interaction avec le site catalytique et le résidu U55 est en contact direct avec l'Asp catalytique, par analogie avec la structure 3D établie pour le co-cristal entre TruB et la tige-boucle TΨC (Fig. R.25) (Roovers *et al.*, 2006 ; McCleverty *et al.*, 2007). Dans ce modèle la jonction entre les domaines N- et C- ter permet d'accommoder le bras accepteur de l'ARNt. Ceci est en accord avec nos résultats qui montrent qu'un petit ARN contenant les bras accepteur et le bras TΨC est substrat de la réaction. Ainsi, le triplet CCA 3' terminal de l'ARNt interagit avec des résidus du domaine THUMP. Dans ce modèle, le bras accepteur est localisé à proximité d'une boucle désordonnée qui pourrait être stabilisée lors de l'interaction ARNt-PUS10. Il est probable qu'en absence de tige-boucle anticodon, la tige-boucle D déstabilise la fixation de la tige-boucle TΨC, ce qui pourrait expliquer nos résultats (Fig. R.22). Il faut également noter que l'analyse d'un alignement de la séquence du domaine N-ter de PUS10 humaine avec celles des domaines N-ter de ses homologues d'archaea montre que seuls les résidus du domaine THUMP et du doigt à Zn sont conservés. Ainsi, les spécificités de substrat des protéines Pus10 animales ou d'archaea pourraient être différentes. Il serait intéressant de réaliser une étude comparative de leur spécificité d'action chez les archaea et chez les vertébrés. Il est possible que chez les vertébrés, du fait de la présence d'un équivalent de Pus4, l'enzyme PUS10 n'agisse qu'en

position 54 de l'ARNt. Le positionnement du r sidu 54 au niveau de l'Asp catalytique dans le site actif est compatible avec la structure 3D obtenue (Fig. R.25).

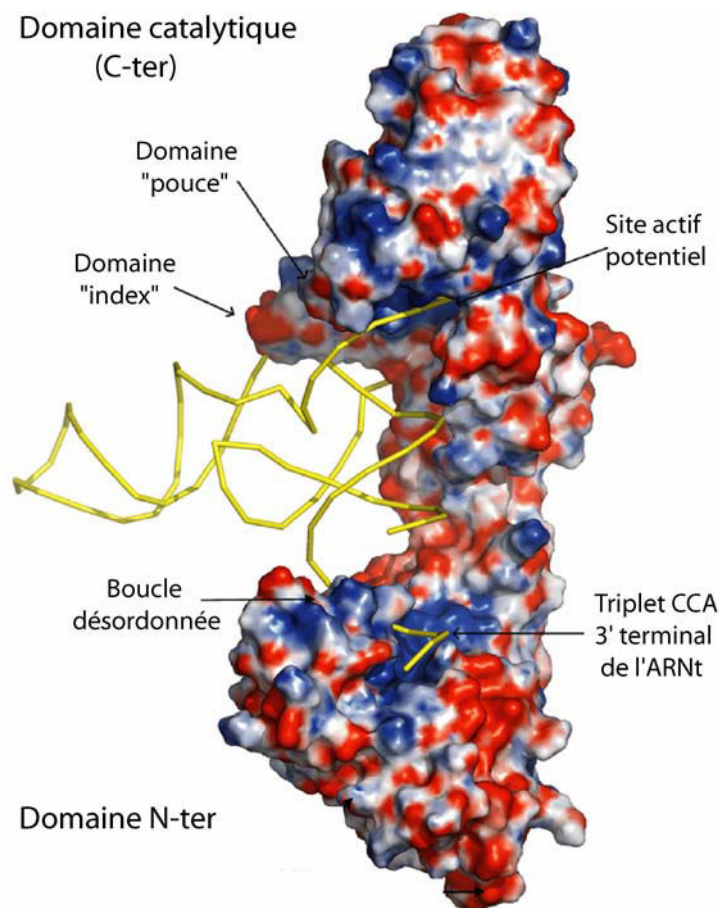


Figure R.25 : Mod lisation d'un ARNt li    la prot ine PUS10 humaine (d'apr s McCleverty *et al.*, 2007). La surface de la prot ine PUS10 est repr sent e selon le potentiel de charges  lectrostatiques. Les charges positives et n gatives sont repr sent es en bleu et en rouge respectivement. La cha ne ribose-phosphate de l'ARNt est sch matis e en jaune.

Enfin, il est   noter que A. Urban a v rifi , au laboratoire, que Cbf5p de *S. cerevisiae* n'agit pas en position 55 des ARNt et il serait int ressant de v rifier s'il en est de m me pour la dysk rine chez les vert br s.

7. Identification de sRNA H/ACA capables de cibler la position 35 dans les ARNt^{Tyr} de certaines espèces du royaume des crénotes

A. Contexte & Intérêts

Jusque là, il avait été montré que toutes les conversions de résidus U en Ψ dans les ARNt, aussi bien chez les Bactéries que chez les Eucaryotes, sont catalysées par des enzymes ne nécessitant pas d'ARN guide. L'existence possible de sRNA H/ACA pour la pseudouridylation des ARNt d'archaea restait une question ouverte. En effet, peu de gènes susceptibles de coder des ARN:Ψ-synthases avaient été mis en évidence chez les archaea : les gènes d'aCBF5 et aPus10 mentionnées précédemment et d'un homologue de TruD/Pus7 (aPus7) (Watanabe *et al.*, 2000). Par ailleurs, chez les euryotes uniquement, un homologue de TruA est également présent. L'enzyme aPus10 peut modifier, comme nous l'avons décrit, les résidus U en positions 54 et 55 des ARNt, et aCBF5 les résidus U en position 55. Il ne restait donc plus que aTruA (chez les euryotes) et aPus7 comme ARNt:Ψ-synthases potentielle. L'enzyme Pus7 était connue pour agir en position 13 de certains ARNt eucaryotes comme son homologue TruD bactérien (Kaya et Offengand, 2003). Par ailleurs, chez les eucaryotes, Pus7 est également responsable de la modification en position 35 du pré-ARNt^{Tyr} (GUA) (Ma *et al.*, 2003 ; Behm-Ansmant *et al.*, 2003) (Tableau III.2 et 3), modification qui était jusque-là considérée comme spécifique de ces organismes.

En dépit de ce faible nombre d'ARNt:Ψ-synthases potentielles chez les archaea, plusieurs résidus Ψ avaient été détectés dans l'espèce de *Haloferax volcanii* pour laquelle tous les isotypes d'ARNt ont été analysés (Fig. III.15). Par conséquent, nous pouvions nous demander si le système dépendant des sRNA H/ACA pouvait être impliqué dans la formation de ces résidus. Une telle situation était d'autant plus envisageable qu'il avait été montré que la formation d'un grand nombre de 2'-O-méthylations dans les ARNt d'archaea est catalysée par le biais de sRNA C/D.

B. Résultats

J'ai recherché, grâce à l'approche bioinformatique que nous avons développée (Muller *et al.*, 2007a) des sRNA H/ACA capables de guider des pseudouridylation dans les ARNt. Finalement, je n'ai détecté qu'un petit nombre de sRNA H/ACA présentant des complémentarités avec les ARNt. Ces sRNA H/ACA pourraient donc être dédiés à la modification des ARNt. Certains de ces sRNA H/ACA, identifiés dans l'espèce *A. pernix* de l'ordre des Desulfurococcales et dans les 4 espèces de l'ordre des Sulfolobales dont les génomes ont été séquencés, guideraient une même pseudouridylation dans les ARNt : la formation du résidu Ψ35 dans le pré-ARNt^{Tyr}(GUA) qui

contient un intron. Comme l'enzyme Pus7 catalyse la formation de ce résidu chez les eucaryotes, j'ai comparé les protéines aPus7 d'archaea avec celles des espèces pour lesquelles j'avais mis en évidence le sRNA H/ACA potentiel. J'ai ainsi détecté la présence de substitutions d'acides aminés supposés être importants dans l'activité catalytique dans les motifs II et IIIa des 5 espèces pour lesquelles j'avais mis en évidence le sRNA H/ACA dédié au pré-ARNt^{Tyr}. J'ai alors testé l'activité de la sRNP H/ACA de *S. solfataricus* avec les protéines aCBF5/aNOP10/aGAR1 de *P. abyssi* et j'ai ainsi pu vérifier qu'il guide bien la pseudouridylation en position 35 des pré-ARNt^{Tyr} de *S. solfataricus*. J'ai entrepris avec I. Motorin et A. Urban, une étude comparative de la spécificité d'action des ARN:Ψ-synthases aPus7 de *P. abyssi* et de *S. solfataricus*, ceci en produisant ces protéines sous forme recombinante dans *E. coli* et en utilisant différents ARN substrats. Nous avons ainsi pu montrer que l'enzyme aPus7 de *P. abyssi* agit comme son homologue eucaryote à la fois en position 13 et 35 des ARNt mais, à la différence de son homologue eucaryote, elle peut agir en position 35 en absence d'intron, et elle agit sans doute aussi à d'autres positions des ARNt. La génération des mutations trouvées dans les motifs II et IIIa des enzymes aPus7 de *S. solfataricus* dans l'enzyme aPus7 de *P. abyssi* abolit son activité à 55°C. Ceci est en accord avec la très faible activité de l'enzyme aPus7 de *S. solfataricus* à cette température. Néanmoins, nous avons constaté qu'à température élevée, l'enzyme aPus7 de *S. solfataricus* peut agir très faiblement en position 13 mais pas en position 35 des ARNt. Ceci explique sans doute la présence d'un sRNA H/ACA dédié à la modification de la position 35 dans ces espèces, mais l'absence de détection de sRNA H/ACA dédié à la modification de la position 13.

Nous montrons donc que le recours à des sRNA H/ACA pour la pseudouridylation des ARNt n'est qu'exceptionnelle chez les archaea, mais aussi que les sRNA sont un réservoir potentiel permettant de parer à l'apparition d'une mutation dans une ARN:Ψ-synthase dont la fonction cellulaire est certainement essentielle.

Ces données ont été rédigées sous la forme d'un article qui va être soumis à RNA et dont je suis premier auteur.

Publication N°4 :

Deficiency of the tRNA^{Tyr}:Ψ35-synthase Pus7 in Archaea of the Sulfolobales order is rescued by the presence of a specific H/ACA sRNA

Sébastien Muller, Alan Urban, Arnaud Hecker, Fabrice Leclerc, Christiane Branlant, Yuri Motorin.

Préparé pour soumission à RNA

PUBLICATION N°4

Deficiency of the tRNA^{Tyr}:Ψ35-synthase Pus7 in Archaea of the Sulfolobales order is rescued by the presence of a specific H/ACA sRNA

Sébastien Muller¹, Alan Urban¹, Arnaud Hecker², Fabrice Leclerc¹, Christiane Branlant^{*,1} and Yuri Motorin¹

¹Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP
Nancy Université, BP 239, 54506 Vandoeuvre-les-Nancy Cedex, FRANCE

²Institut de Génétique et Microbiologie, Université Paris-Sud, IFR115, UMR8621-CNRS,
91405 Orsay, FRANCE

Short running title : Pseudouridine Ψ35 formation in archaeal tRNAs

*Corresponding author :

Christiane BRANLANT

Laboratoire Maturation des ARN et Enzymologie Moléculaire

UMR 7567 CNRS-UHP Nancy I

Faculté des Sciences, BP 239

54506 Vandoeuvre-les-Nancy Cedex, FRANCE

Tel. + 33 (0).3.83.68.43.03

Fax + 33 (0).3.83.68.43.07

e-mail christiane.branlant@maem.uhp-nancy.fr

Keywords : Archaea / pseudouridine / RNA:pseudouridine-synthase / tRNA / sRNA-guide

ABSTRACT

The presence of a pseudouridine residue at position 35 (Ψ 35) is a characteristic feature of all known eukaryotic tRNA^{Tyr}(GUA). Ψ 35 formation in *S. cerevisiae* pre-tRNA^{Tyr}(GUA) is catalyzed by the RNA: Ψ -synthase Pus7, which also acts at position 13 in several tRNAs and at position 35 in U2 snRNA. In Bacteria, tRNA: Ψ -synthase TruD catalyzes the formation of pseudouridine residue at position 13 in tRNAs, but not at position 35. Here we found that the archaeal homologue (aPus7) of eukaryal Pus7 and bacterial TruD is able to modify tRNAs at position 13 and also position at 35 in both intron-containing and intron less tRNA^{Tyr}(GUA). Surprisingly, in several species of the Sulfolobales and Desulfurococcales orders, substitutions of amino acids, which are important for Pus7 activity, abolish Pus7 activity at position 35 in tRNA^{Tyr}(GUA). Despite the absence of tRNA: Ψ 35-synthase activity of Pus7 in these species, cell-free extract conserves the activity towards Ψ 35, complementing the deficiency of endogenous aPus7. In these species, we found that Ψ 35 formation in tRNA^{Tyr}(GUA) is insured by aCBF5-based machinery associated with a specific H/ACA sRNA that guides U35 modification. We demonstrated this activity by using both *in vitro* reconstitution assays and cellular extracts. This is the first demonstration that an H/ACA sRNA can guide pseudouridine formation in tRNAs.

INTRODUCTION

In all domains of life, pseudouridine (Ψ) formation is the most frequent post-transcriptional modification of RNAs. These U to Ψ conversions are catalysed by two major pathways: either by a stand-alone enzyme (specific RNA: Ψ -synthase), that can have narrow or large specificity (for review, Hamma & Ferre-D'Amare, 2006), or by a highly specific small ribonucleoprotein particle (H/ACA snoRNP or H/ACA scaRNP in eucarya, and H/ACA sRNP in Archaea) (Lafontaine *et al.*, 1998; Watkins *et al.*, 1998). H/ACA RNPs contain a small RNA that defines the targeted U residue by base-pair interaction with the RNA substrate (Charpentier *et al.*, 2007). As well in Eukarya as in Archaea, H/ACA sRNPs also contain 4 proteins: Nop10, Gar1 L7ae/Nhp2p and Cbf5/Dyskerin, which carries the activity of RNA: Ψ -synthase. Recent data showed that the aCBF5/aNOP10/aGAR1 complex in Archaea may also catalyze tRNA pseudouridylation without guide RNA (Charpentier *et al.*, 2007; Muller *et al.*, 2007a).

Pseudouridine residues are universally found in ribosomal RNA (rRNAs) and in tRNAs (Sprinzl & Vassilenko, 2005). In Eukarya, their presence was also detected in the spliceosomal UsnRNAs and some snoRNAs (for review, Massenet *et al.*, 1998)). Conversion of U into Ψ residue in RNAs modifies the local conformation of RNA molecules. Indeed, the additional free N₁-H position on the nucleobase allows the formation of an additional base-pair interactions, either in *cis* within the RNA molecule, or in *trans* with another RNA molecule or a hydrogen bond interaction with a protein. Secondly, substitution of the C-N bond between the ribose and the nucleobase by a C-C bond limits the flexibility of the ribose conformation that also favours RNA-RNA base-pair interactions (Auffinger & Westhof, 1997; Massenet *et al.*, 1998).

In all organisms, tRNA molecules can be pseudouridylated at multiple sites (7 positions in *E. coli*, at least 15 positions in *S. cerevisiae*, and even more in higher eukaryotes) (Sprinzl &

Vassilenko, 2005). Among the Ψ residues in tRNAs, the eukaryal specific Ψ 35 residue in the anticodon loop of tRNA^{Tyr}(GUA) was shown to play an important role in mRNA decoding (Tomita *et al.*, 1999). Up to now, for both Bacteria and Eukarya, pseudouridylation in tRNAs were shown to be catalysed by numerous stand-alone enzymes and most of the tRNA: Ψ -synthases in *E. coli* and in *S. cerevisiae* have been identified (for review, Ansmant & Motorin, 2001; Hamma & Ferre-D'Amare, 2006). Some of them were found to have quite narrow specificity, as is the case for the bacterial TruB, TruD and TruC enzymes. They are responsible for U to Ψ conversion at one single site (positions 55, 13 and 65 in tRNAs, respectively). Other bacterial tRNA: Ψ -synthases have larger specificities, TruA acts at several sites in tRNAs (Kammen *et al.*, 1988), while RluA modifies both tRNAs and rRNAs (Raychaudhuri *et al.*, 1999). In contrast, in Eukarya, the numerous pseudouridylations in cytoplasmic rRNAs are catalysed by H/ACA snoRNPs. Therefore, in these species, tRNA: Ψ -synthases are either dedicated only to tRNA modification or act both on tRNAs and UsnRNAs (Pus1p and Pus7p) (Massenet *et al.*, 1999; Behm-Ansmant *et al.*, 2003). Here again, some of the tRNA: Ψ -synthases dedicated to tRNAs have a very narrow specificity, as is the case for Pus4p or Pus6p, Pus8p and Pus9p (acting at position 55, 31 et 32, respectively) (Becker *et al.*, 1997; Ansmant *et al.*, 2001; Behm-Ansmant *et al.*, 2004). Other enzymes may have very large spectrum of substrates, like Pus1p which acts at 8 distinct positions in tRNAs and at position 44 in U2 snRNA (Massenet *et al.*, 1999), or Pus7p that converts U into Ψ in 3 distinct types of RNAs, in U2 snRNA at position 35, in some tRNAs at position 13 and in the pre-tRNA^{Tyr}(GUA) at position 35 (Behm-Ansmant *et al.*, 2003). Altogether, 10 distinct RNA: Ψ -synthases were characterized in *E. coli* and in *S. cerevisiae*, however, in *E. coli* five of them act on tRNAs (Kammen *et al.*, 1988; Gutgsell *et al.*, 2000; Wrzesinski *et al.*, 2000; Del Campo *et al.*, 2001; Kaya & Ofengand, 2003), while 8 tRNA-specific enzymes are present in *S. cerevisiae* (Simos *et al.*, 1996; Becker *et al.*, 1997; Motorin *et al.*, 1998;

Ansmant *et al.*, 2001; Behm-Ansmant *et al.*, 2003; Behm-Ansmant *et al.*, 2004; Hepfer *et al.*, 2005; Behm-Ansmant *et al.*, 2007).

Studies on archaeal tRNA:Ψ-synthases are far less advanced. Indeed, whereas rRNA pseudouridylation and 2'-O-methylation in Archaea were shown to be catalysed via sRNA-guides (Omer *et al.*, 2000; Dennis *et al.*, 2001), little is known on the tRNA modification systems. In addition, post-transcriptional modifications in tRNAs have only been identified in a limited number of archaeal species: all tRNA molecules were sequenced only for *Haloferax volcanii* (Gupta, 1984). However, this halophilic organism may not be good representative of the whole archaeal domain. Indeed, the diversity of the archaeal organisms is extremely wide, since Archaea includes organisms growing not only at normal salt concentration, pH, temperature and pressure, but also many species able to grow in various extreme conditions, in particular at high pressure and elevated temperatures (Chaban *et al.*, 2006). For some of these hyperthermophilic Archaea, global profile of post-transcriptional modifications in rRNA and/or tRNA was studied by the HPLC/MS approach (McCloskey *et al.*, 2001). However, the identified modified residues were not localized in individual RNA molecules. Only 2'-O-methylations in many archaeal tRNAs were localized, because they are catalyzed by sRNPs containing C/D box sRNAs (Renalier *et al.*, 2005). Therefore, studies on the tRNA modification enzymes in Archaea were essentially focused on a limited number of post-transcriptional modifications, which are specific of the archaeal domain, such as methylations (m¹I58, Gm18) and archaeosine (Gregson *et al.*, 1993; Grosjean *et al.*, 1995; Ishitani *et al.*, 2002; Roovers *et al.*, 2006; Muller *et al.*, 2007b). Recently peculiar catalysts for the universal pseudouridylation at position 55 in tRNAs were also identified (the aCBF5/aNOP10/aGAR1 complex or the newly identified PsuX/Pus10 enzyme) (Roovers *et al.*, 2006; Muller *et al.*, 2007b). One puzzling observation in Archaea, is a limited number of putative genes encoding RNA:Ψ-synthases that can be detected in the entirely sequenced

genomes using sequence homology with bacterial or eukaryotic counterparts. Only genes encoding homologs of the bacterial TruB, TruD/Pus7 and PsuX/Pus10 enzymes are found in all species. In euryarchaeota species an additional gene encoding a member of the TruA family was detected (Watanabe & Gray, 2000). The archaeal homologue of TruB is the aCBF5 enzyme involved in H/ACA sRNP formation. PsuX/Pus10 was recently found to modify position 55 in tRNAs. Therefore up to now, the activities of the TruD/Pus7 homologue and the TruA family member have never been studied.

Based on the data obtained for the *H. volcanii* tRNAs and some other archaeal species, several positions in archaeal tRNAs can be pseudouridylated. The most frequent ones are positions 13, 39, 54 and 55. Furthermore, *in vitro* tests made with a *Pyrococcus furiosus* cellular extract and *H. volcanii* tRNA transcripts confirmed the formation of Ψ residues at position 39 and 55 (Constantinesco *et al.*, 1999). As mentioned above, aCBF5 and Pus10 are both implicated in U to Ψ conversion at position 55. The enzymes acting at the other pseudouridylation sites still have to be identified. One reasonable candidate for modification at position 13 is the archaeal aPus7. In addition, as yeast Pus7p also modifies U35 in the pre-tRNA^{Tyr}(GUA) (Behm-Ansmant *et al.*, 2003), even though Ψ 35 was not found in the *H. volcanii* tRNA^{Tyr}(GUA) (Gupta, 1984), an activity of the aPus7 enzyme at position 35 in tRNA cannot be excluded. The absence of Ψ 35 in the *H. volcanii* tRNA^{Tyr}(GUA) could be due to the absence of intron in its gene. Indeed, Ψ 35 formation in both yeast and plant tRNAs^{Tyr} was found to be intron-dependent (Johnson & Abelson, 1983; Szweykowska-Kulinska & Beier, 1992). Taking into account a low number of tRNA: Ψ -synthases in most Archaea, Ψ residues in tRNAs from these organisms could be formed not only by stand-alone enzymes, but also by H/ACA sRNPs, like is the case of certain tRNA 2'-O-methylations catalysed by C/D box sRNPs.

Here, we tried to address several open questions concerning pseudouridylation of archaeal tRNAs. By using the computational method that we recently developed for the search of H/ACA sRNA motifs (Muller *et al.*, 2007a), we identified an H/ACA sRNA able to guide Ψ 35 formation in the pre-tRNA^{Tyr}(GUA) in some archaeal species of the Sulfolobales and Desulfurococcales orders. We demonstrated its activity by *in vitro* reconstitution of the corresponding H/ACA sRNP and by a series of tests using archaeal cellular extracts. In parallel, we observed that the species that express this particular H/ACA sRNA, possess Pus7 proteins bearing amino acid substitutions expected to affect their activity. Therefore, by using recombinant proteins, we compared the activities of aPus7 enzymes from *P. abyssi* and *S. solfataricus* on tRNA^{Asp}, tRNA^{Tyr}(GUA) and pre-tRNA^{Tyr}(GUA). *P. abyssi* does not express the tRNA^{Tyr}(GUA)-specific H/ACA sRNA, while it is present in *S. solfataricus*. We showed that Pab Pus7 can act at position 13 in tRNAs and at position 35 in both intron-less and intron-containing tRNA^{Tyr}(GUA). However, tRNA^{Tyr}(GUA) is not modified by enzyme originating from the species that express a H/ACA sRNA guiding U to Ψ conversion at position 35.

RESULTS

Some H/ACA sRNAs in Archaea are predicted to guide Ψ formation in tRNAs

By applying the computational method, that we recently developed for the search of H/ACA sRNA genes in archaeal genomes (Muller *et al.*, 2007a), we found putative H/ACA sequence sRNAs which may direct U to Ψ conversion in tRNAs. Interestingly, all of them were predicted to guide the conversion of the same U target: position 35 in the intron-containing pre-tRNAs^{Tyr}(GUA) (Fig. 1A and 1B). One of these unusual H/ACA sRNAs was detected in the archaea of the Desulfurococcales order, *Aeropyrum pernix* (sRNA Ape7) and four other ones were found in the four species of the Sulfolobales order, namely *Sulfolobus solfataricus*

(sRNA Sso1), *Sulfolobus tokodaii* (sRNA Sto1), *Sulfolobus acidocaldarius* (sRNA Sac1) and *Metallosphaera sedula* (sRNA Mse2). These 5 putative H/ACA RNA motifs all contain a K-turn located at the expected distance (5 bps) from the pseudouridylation pocket and the conserved ACA trinucleotide is present at the 3' -end of the stem-loop structure (Fig. 1). Interestingly, the DNA genes for the *S. solfataricus*, *S. tokodaii*, *S. acidocaldarius* H/ACA motifs are directly flanked by DNA sequences corresponding to C/D box sRNAs: the already characterized Sso-159 (from *S. solfataricus*) (Tang *et al.*, 2002) and C/D box sRNAs that we denoted Sto1 sR4 and Sac1 sR39 for *S. tokodaii* and *S. acidocaldarius*, respectively). By using the complementary DNA probe, we verified the presence of Sso1 H/ACA RNA in total RNA fraction extracted from *S. solfataricus* cells. The northern blot analysis shown in Figure 1C reveals the presence of a long form of this RNA, that likely contains both the H/ACA and the C/D motifs, and a short form corresponding to the H/ACA motif alone. Therefore, the putative Sso1 H/ACA sRNA is indeed expressed in *S. solfataricus*.

The predicted RNA-basepairing interactions between each of the putative H/ACA sRNAs and the corresponding pre-tRNA^{Tyr}(GUA) species involve both the anticodon stem-loop of tRNA and the intron sequences (Fig. 1B). These interactions all fit to the criteria that we recently established for efficient H/ACA sRNA-target RNA interaction (Muller *et al.*, 2007a). Indeed, as for the H/ACA sRNA-rRNA interaction found in *P. abyssi*, the distance between the K-turn and the internal loop is of 5 bps, the distance between the targeted U residue and the ACA motif is of 14 nts and a stable base-pair interaction is formed between the 3'-guide sequence and the target RNA (Fig. 1B). In case of the Sac1 and the Mse2 sRNAs, the predicted interaction with the 3' sequence is exceptionally long as compared to previously characterized H/ACA sRNAs that direct rRNA pseudouridylation (Muller *et al.*, in preparation). However, the interaction predicted for the 3'-guide sequence of the *A. pernix* sRNA has a lower stability (Fig. 1B). In favour of the biological significance of the predicted

H/ACA sRNA-tRNA interactions, we found that sequence variations in the Sulfolobales H/ACA 5'-guide elements correlated to sequence variations in the pre-tRNA^{Tyr}(GUA) intron, in order to preserve the possible interaction (Fig. 1B). No other possible target sites in other RNAs were found for the Mse2, Sso1, Sto1 and Sac1 H/ACA sRNAs. Only the Ape-7 motif can guide pseudourylation at position 2444 in 23S rRNA, as found previously for the *P. abyssi* Pab91 H/ACA sRNA (Fig. 1A) (Charpentier *et al.*, 2005).

Therefore we concluded that the Mse2, Sso1, Sto1 and Sac1 H/ACA sRNAs likely guide pseudourylation in the pre-tRNA^{Tyr}(GUA) and this may also be the case for the Ape-7 H/ACA sRNA.

An sRNA-guided Ψ35 formation is detected in pre-tRNA^{Tyr} *in vitro*

For an experimental confirmation of the pseudouridylation guiding capability of the identified H/ACA sRNAs at position 35 in the pre-tRNA^{Tyr}(GUA), we used the *in vitro* reconstitution assay established for archaeal H/ACA sRNPs (Charpentier *et al.*, 2007). We expected that the *P. abyssi* proteins should be active in a H/ACA sRNP reconstitution assay even if an heterologous H/ACA sRNA is used. To check this hypothesis, we used EMSA to test the association between an *in vitro* transcribed Sso1 H/ACA sRNA and various combinations of the four *P. abyssi* proteins (L7Ae (L), aCBF5 (C), aNOP10 (N) and aGAR1 (G)) produced in *E. coli* (data not shown). Association pattern obtained for Sso1 H/ACA sRNA was very similar to the pattern observed previously for the *P. abyssi* Pab91 sRNA (Charpentier *et al.*, 2005). Thus we tested the activity of the reconstituted particles on an [α -³²P]ATP-labeled *S. solfataricus* pre-tRNA^{Tyr}(GUA) transcript (2 hours incubation at 65°C in the presence of the guide RNA and the 4 LCNG proteins). Labelling by [α -³²P]ATP incorporation followed by RNase T2 digestion allowed the detection of the released Ψp by 2D thin layer chromatography (2D TLC). As a control, the assay was also performed in the absence of the

Sso1 RNA guide (Fig. 2B). In addition, in order to verify that Ψ 35 formation can occur in the Sso pre-tRNA^{Tyr} *in cellulo*, the pre-tRNA^{Tyr}(GUA) transcript was also incubated for 2 hours at 65°C in a *S. solfataricus* cellular extract. As shown in Figure 2B, Ψ formation was detected in the extract and after incubation in conditions of sRNP reconstitution, and it is strictly dependent on the presence of the Sso1 RNA. Although, several UA dinucleotides are present in the pre-tRNA^{Tyr}(GUA), only U35 was expected to be a target. To confirm this hypothesis, we used two complementary approaches. First, the unlabeled pre-tRNA^{Tyr}(GUA) was incubated in conditions for *in vitro* reconstitution of the Sso1 RNP activity and, in parallel, in the Sso in cellular extract. The two RNA samples were subjected to CMCT-RT analysis. The appearance of a reverse transcriptase stop at position 36 in the RT elongation profile confirmed Ψ formation at position 35 in both assays. Finally, a complete demonstration that modification occurred only at position 35 was obtained by repeating the reconstitution assay and the incubation in the Sso cellular extract with an [α -³²P]ATP-labeled pre-tRNA^{Tyr} containing a U35C base substitution. No formation of Ψ residue was detected in the 2D-TLC profile neither with the reconstituted H/ACA sRNPs nor with the cellular extract (Fig. 2B). Altogether, these data clearly demonstrated that the Sso1 H/ACA sRNA can guide Ψ 35 formation in the Sso pre-tRNA^{Tyr}(GUA).

The presence of an H/ACA sRNA guiding Ψ 35 formation in archaeal pre-tRNA^{Tyr} is accompanied by amino acid sequence variations in the Pus7 protein

The identification of an H/ACA sRNA directing modification at position 35 in the archaeal pre-tRNA^{Tyr}(GUA) was surprising, since genes encoding Pus7-like enzymes were detected in all archaeal genomes and, as mentioned above, Pus7 is responsible for this modification both in plants and in yeasts. Local sequence alignment of the Pus7-like proteins from various archaeal species representating most of the archaeal genera, together with the 5 Sulfolobales

and *A. pernix* species is shown on Fig. 3B. We observed a high degree of divergence in motifs IIIa and II for the four Sulfolobales and *A. pernix* Pus7 compared to the other archaeal species (Fig. 3B). This sequence divergence correlates with the presence of H/ACA sRNA guiding tRNA modification. Motif IIIa and II are highly conserved amino acid sequences in all RNA:Ψ-synthases and are located at the catalytic pocket (Kaya *et al.*, 2004). Motif II contains the catalytic Asp residue involved in U isomerisation (Huang *et al.*, 1998). As expected, most of the highly conserved residues in motifs IIIa and II (K19, F25, D77, R78, H79, according to the numbering in the *P. abyssi* Pus7 sequence) are found in the Pus7 amino acid sequences from species that do not express an H/ACA guide for position Ψ35 in pre-tRNA^{Tyr} (Fig. 3A). In contrast, in the Pus7 amino acid sequences from the 4 Sulfolobales species and *A. pernix*, the conserved basic residue in motif IIIa (K19 in *P. abyssi*) is replaced by an isoleucine residue (I27 in *S. solfataricus*) and the arginine-histidine pair in motif II (R78/H79 in *P. abyssi*) is changed into an alanine (or threonine) - asparagine (or cysteine) pair (A90/N91 in *S. solfataricus*) (Fig. 3B). Other amino acid substitutions are also observed around these conserved positions. However, the catalytic aspartic residue remains conserved (D89 in *S. solfataricus*). Altogether, these observations suggested an altered activity of the Pus7-like enzymes in *S. solfataricus*, *S. tokodaii*, *S. acidocaldarius*, *M. sedula* and *A. pernix*.

Pab Pus7 modifies tRNAs at position 13 and 35 while Sso Pus7 fails to catalyze Ψ35 formation

To check the effects of amino acid substitutions at the catalytic pocket of the *S. solfataricus* Pus7-like enzyme and more generally to test for the activity of the archaeal Pus7-like enzymes, we produced the *P. abyssi* (Pab) and the *S. solfataricus* (Sso) Pus7-like proteins as His₆-tagged fusions in *E. coli*. We used various RNA transcripts to test activities of purified recombinant proteins (Fig. 4A): two of them corresponded to *S. solfataricus* RNAs (Sso pre-

tRNA^{Tyr}(GUA) and Sso tRNA^{Asp}(GUC)) and two other ones to *P. abyssi* RNAs (Pab tRNA^{Tyr}(GUA) and Pab tRNA^{Asp}(GUC)). The two tRNA^{Asp} transcripts were used to detect the potential activity of the archaeal Pus7-like enzyme at position 13. The use of intronless tRNA^{Tyr} and a pre-tRNA^{Tyr} transcript with an intron was expected to bring information on the functional importance of an intron for modification at position 35. The *in vitro* activities of the recombinant proteins were tested by the nearest neighbour approach. As all the U residues expected to be modified were followed by an A residue, all the transcripts were labeled by incorporation of [α -³²P]ATP during transcription. Control experiments were performed with the *in vitro* transcribed tRNA^{Tyr} labeled by [α -³²P]GTP, [α -³²P]CTP or [α -³²P]UTP incorporation, as well as with an [α -³²P]GTP-labeled Sso tRNA^{Tyr}. Incubation conditions are given in Materials and Methods. The assays were first performed at 55°C, a temperature expected to preserve the intact tRNA 2D and 3D structure. As above, formation of Ψ residues was detected by 2D TLC after digestion of the RNA sample by RNase T2 (Fig. 4 B and C). Whereas no Ψ formation was detected in the control experiments using [α -³²P]GTP, CTP or UTP labeled RNAs (see panel C in Fig. 4), Ψ formation was detected for all the [α -³²P]ATP labeled substrates that we incubated with Pab Pus7p. This was a strong indication for the requirement of an A residue 3' to the targeted U residue for activity of Pab Pus7. In contrast, no Ψ formation was detected for any of the tested substrates when Sso Pus7 was used for incubation. This demonstrated the strong negative influence of the amino acid substitution found in the catalytic pocket of Sso Pus7 on the tRNA: Ψ -synthase activity of this enzyme.

We noticed that Pab Pus7 modified all four RNA transcripts but at variable levels (more than one mole of Ψ /mole tRNA for the Pab tRNA^{Tyr}, the Sso pre-tRNA^{Tyr} and the Sso tRNA^{Asp} transcripts and less than one mole of Ψ /mole tRNA for the Pab tRNA^{Asp} transcript (Fig. 4B)). To interpret these data, we had to take into account the fact that all Ψ residues formed 5' to an A residue were detected when [α -³²P]ATP labeled transcripts are used in assays. Therefore, to

test whether Pab Pus7 acts at position 13 and 35 in tRNAs like its eukaryal homologue, we generated a U13C substitution in the Pab tRNA^{Asp}(GUC) transcript and U35C substitutions in both the Pab tRNA^{Tyr}(GUA) and Sso pre-tRNA^{Tyr}(GUA) transcripts. As illustrated in Figure 4B, in spite of the presence of a two UA dinucleotides in the Pab tRNA^{Asp} molecule, U13C substitution in this tRNA completely abolished Ψ formation by Pab Pus7p. Therefore, we concluded that the Pab Pus7 enzyme, like its eukaryal counterpart, converts U into Ψ at position 13 in tRNAs. Similarly, in spite of the presence of numerous UA dinucleotides in the Pab tRNA^{Tyr} transcript (four altogether), only two moles Ψ /mole RNA were formed in the WT tRNA and only one in the tRNA mutant. Finally, whereas 5 UA dinucleotides are present in the Sso pre-tRNA^{Tyr} transcript, only 1.47 moles of Ψ were formed per RNA molecule and only half of this amount (0.71 mole Ψ /mole tRNA) was detected upon U35C substitution. Therefore, we concluded that the archaeal Pab Pus7 enzyme can modify position Ψ 35 in tRNAs, independently of the presence of an intron. In the molecule, other modifications in the U35C Pab tRNA^{Tyr} transcript may occur at position 8, 13 or 21. The 1.15 mole Ψ formed in Sso tRNA^{Asp} may correspond to U13 Ψ conversion plus a modification at one of the other UA dinucleotides. We detected only 0.77 mole of Ψ /RNA in Pab tRNA^{Tyr} transcript after incubation with the Pab cellular extract (Fig. 4). However, here again, the amount of Ψ formation was divided by a factor of 2 (0.42 mole of Ψ /mole tRNA) after U35C substitution. Taking into account the fact that efficiency of modification in the cell-free extract is lower compared to that obtained with purified recombinant enzymes, we considered that at least two uridines are likely modified upon incubation of the Pab tRNA^{Tyr} transcript in a Pab cellular extract and one of them is the residue U35.

Important roles of the conserved basic residue in motif IIIa and KR/RH pair in motif II of archaeal Pus7

The above data showed a very low or even undetectable *in vitro* activity at 55°C of the Sso Pus7 on the 4 tRNA substrates that we tested. Based on the amino acid sequence alignment of motifs IIIa et II of the archaeal Pus7 enzymes (Fig. 3), at least 3 amino acid substitutions could explain this inactivity of Sso Pus7. In order to dissociate the effect of the conserved basic residue in motif IIIa, from the effect of the substitutions in the conserved KR/RH pair of motif II, the K27I substitution found in the Sso motif IIIa was transferred into Pab Pus7 sequence by a site-directed mutagenesis (variant Pab K19I). We also transferred the R90A/H91N double substitution in the Sso motif II into Pab Pus7 sequence (variant R78A/H79N). In addition, we produced a Pab Pus7 variant enzyme mutated in both motifs (K19I + R78A/H79N). Conversely, we tested whether we could restore activity in Sso Pus7 by the individual I27K substitution in motif IIIa, the double A90R and N91H substitutions in motif II or by combination of these 3 amino acid substitutions. The variant Pab Pus7 and Sso Pus7 enzymes were produced as recombinant proteins in *E. coli* and purified using their His₆-tag sequence. As the WT Pab Pus7 modifies only position 13 in the Pab tRNA^{Asp}, this substrate was used to test the activity of the variant proteins. Each test of activity at 55°C was done in triplicate and the mean values obtained are shown in Figure 5. In accordance with the data mentioned above, about 0.8 mole of Ψ residue were formed in Pab tRNA^{Asp} by WT Pab Pus7, whereas a very low activity was detected with the WT Sso Pus7. Both the K19I and R78A/H79N substitutions in Pab Pus7 decreased the activity towards position 13 in the Pab tRNA^{Asp} transcript at 55°C (Fig. 5A). Furthermore, combination of the 3 amino acid substitutions completely abolished the activity of Pab Pus7 at this temperature. Restoration of an RH pair in motif II of Sso Pus7 had almost no effect on the Sso Pus7 activity, whereas restitution of the conserved basic residue in motif IIIa (I27K variant) allowed significant

increase of U to Ψ conversion. However, combination of all 3 substitutions had no positive effect. Therefore, we concluded that both K19 in motif IIIA and R78/H79 in motif II are important for Pab Pus7 enzyme activity at position 13 in tRNA. Some activity can be restored in the Sso Pus7 by introduction of the basic residue in motif IIIa..

As both *S. solfataricus* and *P. abyssi* grow at elevated temperatures (80°C and 100°C, respectively) we tested the activities of the WT and mutated enzymes at 80°C. Obviously, in these *in vitro* conditions, the tRNA 2D structure is expected to be considerably compromised. However, at this elevated temperature, we were able to detect some activity of the WT Sso Pus7 at position 13 in the Pab tRNA^{Asp} (Fig. 5B). The individual K19I or R78A/H79N substitutions had almost no effect on the Pab Pus7 activity at this temperature. However, their combination nearly abolished the enzymatic activity. In contrast, the I27K substitution in Sso Pus7 increased its activity at position 13 in the Pab tRNA^{Tyr} at 80°C (0.6 against 0.4 Ψ mole of Ψ / mole RNA for the variant and WT enzymes, respectively). However, here again the A78R and N79H substitutions had no positive effect on the Sso Pus7 activity. Therefore, we concluded that in any of the active site sequence context tested, and at any temperature, a basic residue in motif IIIa is important for Pus7 activity. Furthermore, at elevated temperatures the Sso Pus7 displays low, but detectable, activity at position 13. In contrast, according to the results obtained with the pre-tRNA^{Tyr} at 80°C (Fig. 6), even at this high temperature, it has no activity at position 35 in pre-tRNA, which probably explains the need for an H/ACA guided tRNA modification in *Sulfolobus*.

***P. abyssi* and *S. solfataricus* cellular extracts contain different enzymatic activities for Ψ 35 formation in tRNA^{Tyr}**

For a complete demonstration that *P. abyssi* uses Pab Pus7 for Ψ 35 modification in tRNA, whereas *S. solfataricus* uses an H/ACA sRNA-guided system to accomplish this modification, we designed the following strategy. It was based on our observation that the archaeal Pus7

enzymes require UA sequence context for U to Ψ conversion. We expected that introduction of an A36G substitution in the Pab tRNA^{Tyr} transcript should abolish U to Ψ conversion by Pab Pus7 at position 35 in this tRNA. In contrast, the same A36G base substitution should not affect modification machinery at position 35 in Sso pre-tRNA^{Tyr}, since H/ACA sRNA guide is not sensitive to the identity of the residue located 3' to the modified U (this residue as well as the target U residue are not base-paired, see Fig. 1). In order to release labeled Ψ p from position 35 after RNase T2 digestion, the variant A36G Pab tRNA^{Tyr} and A36G Sso pre-tRNA^{Tyr} transcripts were labeled by [α -³²P]GTP incorporation. As expected (Fig. 6A), in spite of the presence of 4 UG dinucleotides in the Pab tRNA^{Tyr} transcript, no Ψ formation was detected after incubation of this tRNA either with Pab Pus7 or with a Pab cellular extract. In contrast, formation of one Ψ residue was still observed upon reconstitution of the Sso1 H/ACA sRNP in the presence of the Sso A36G pre-tRNA^{Tyr} variant. In addition, the same level of U to Ψ conversion was detected upon incubation of the [α -³²P]ATP labeled WT Sso pre-tRNA^{Tyr} or the [α -³²P]GTP labeled Sso A36G pre-tRNA^{Tyr} transcripts in an Sso cellular extract (Fig. 6B).

Taken together these data brought a clear demonstration that in a Pab cellular extract, Pab Pus7 is solely responsible for Ψ 35 formation in the Pab tRNA^{Tyr}(GUA), whereas, in an Sso cellular extract the Sso H/ACA sRNP catalyses modification of the Sso pre-tRNA^{Tyr}(GUA) at the same position. Therefore, the deficiency of Pus7 activity that we found in *S. solfataricus*, is compensated by the presence of a specific H/ACA sRNA guide in this species. This is the first observation of an H/ACA sRNA guide dedicated to tRNA modification.

DISCUSSION

In this work, we performed a detailed analysis of the activity of archaeal Pus7 enzymes on tRNAs. We showed that the absence of fully active Pus7 in Sulfolobales can be compensated

by the presence of an H/ACA guide RNA that directs modification at position 35 in the pre-tRNA^{Tyr}(GUA).

The archaeal Pus7 enzymes is active at position 13 and 35 in tRNAs

The archaeal Pus7 enzymes belong to the TruD family of tRNA:Ψ-synthases. Whereas TruD in *E. coli* modifies only position 13 in tRNA^{Glu} (Kaya & Ofengand, 2003), here we show that, like its eukaryal counterpart, the archaeal Pab Pus7 enzyme can modify both positions 13 and 35 in tRNAs and maybe at some other positions (Ma *et al.*, 2003). Accordingly, 13 out of the 42 sequenced *H. volcanii* tRNAs, contain a Ψ residue at position 13 (Gupta, 1984). Except the universal Ψ55 residue, U13 is the most frequently pseudouridylated position in archaeal tRNAs. Interestingly, whereas both in plants and in yeast, the U to Ψ conversion at position 35 in tRNA^{Tyr} is strictly intron dependent (Nishikura & De Robertis, 1981; Johnson & Abelson, 1983; Szweykowska-Kulinska & Beier, 1992), here we observed the efficient formation of a Ψ35 residue in the intronless Pab tRNA^{Tyr}(GUA). Modification of U35 occurred both upon incubation with the recombinant Pab Pus7 protein and in a cellular extract. Therefore, according to these data, the point of view on Ψ35 formation in archaeal tRNA^{Tyr} should be revised. Indeed, archaeal tRNA^{Tyr} was not expected to be modified at U35 since bacterial TruD does not modify bacterial tRNAs at this position and archaeal tRNA genes were supposed to be intronless (Gupta, 1984).

However, recent data showed that about half of the tRNA^{Tyr}(GUA) genes in archaea contains an intron (Sprinzl & Vassilenko, 2005). In addition, here we demonstrate that Pab Pus7 can act on both tRNA^{Tyr}(GUA) precursor and on intronless tRNA. Therefore, tRNAs^{Tyr}(GUA) in archaea are likely pseudouridylated at position 35. Interestingly, before the present data, the only example of the formation of a Ψ35 residue in intronless tRNAs concerns the mitochondrial tRNA^{Asn}(GUU) from the starfish *Asteria amurensis* (Tomita *et al.*, 1999).

Archaeal Pus7 modifies RNA only at UA dinucleotides

A comparison of the *S. cerevisiae* Pus7p RNA substrates revealed that this enzyme modifies U residues which are comprised in a highly conserved sequence (Pu₋₄(G/c)₋₃U₋₂N₋₁ΨA₊₁Pu₊₂) (Behm-Ansmant *et al.*, 2003). A similar consensus sequence was proposed to be required for modification at position 35 of the plant tRNAs^{Tyr}(GUA) (Szweykowska-Kulinska & Beier, 1992). In addition, most of the tRNA sequences surrounding Ψ13 residues present in Bacteria, plant and animal tRNAs fit to the above consensus. Here, in agreement with this observation, we demonstrated that an A residue located 3' to the targeted U residue is strictly required for Pab Pus7 activity. In addition, at each of the 3 positions, where we could demonstrate an efficient U to Ψ conversion by Pab Pus7 (positions 13 in the Pab tRNA^{Asp}(GUC) and positions 35 in the Pab tRNA^{Tyr}(GUA) and Sso pre-tRNA^{Tyr}(GUA) respectively), the modified sequence also fit to the above consensus. Remarkably, a second UA dinucleotide which is present in the Pab tRNA^{Asp} is not modified by Pab Pus7. The absence of modification may be explained by the divergence of the surrounding sequences. A pyrimidine (U residue), instead of a Pu residue is found at position +2 and a C residue, instead of a U residue is located at position -2. In the three other RNAs tested in the present work, more than one sequence was fitting the consensus sequence and this may explain the formation of Ψ residues at multiple sites during incubation with Pab Pus7. However, it should be noticed that when Pab Pus7 was incubated at high temperature (80°C) with the various tRNA substrates that we tested, a larger yield of pseudouridylation was observed, suggesting a strong decrease of the Pab Pus7 specificity in these conditions, where the tRNA and pre-tRNA 3D and 2D structures are likely disrupted. Recent results obtained for archaeal tRNA:m5C methyltransferase *P. abyssi* Trm4 demonstrate that interaction with Archease modulate both the solubility of the protein and its substrate specificity (Auxilien *et al.*, 2007). Similar

interactions between Pab Pus7 and its cellular environnement may restrict the interaction of the modification enzyme with its RNA targets.

Peculiar Pus7 proteins are found in some crenarchaeota species

Sequence alignment showed that known Pus7 sequences from species of the Sulfolobales and the Desulforococcales orders, that both belong to the Crenarchaeota kingdom, bear several amino acid substitutions in the highly conserved motifs IIIa and II. This may be a general feature of Sulfolobales Pus7, since all the species whose genomes were entirely sequenced displays this particularity. In contrast, this is not a general property of the Desulforococcales order, since many other species, like *S. marinus*, *I. hospital* and *H. butylicus* have conventional Pus7 proteins. Both these motifs are located at the enzyme active site and may directly participate in catalysis. Mutations in these conserved motifs strongly decrease the Pus7 activity, and the observed negative effect can be explained on the basis of the 3D structures established for the closely related TruD protein (Hoang & Ferre-D'Amare, 2004) and other RNA:Ψ-synthases (Del Campo *et al.*, 2004). Indeed, the conserved lysine residue in motif IIIa (K19 in Pab Pus7) is expected to make a salt bridge with the catalytic aspartic acid residue (D77 in Pab Pus7, Fig. 3). This interaction facilitates the nucleophilic attack of the COO⁻ group on the C6 of the uridine base. In Sso Pus7, as the lysine residue in motif IIIa is replaced by an isoleucine (I27), a salt bridge cannot be formed with the catalytic D89 residue (Fig. 3). In consequence, only a poor activation of the COO⁻ group of D89 is expected. This probably explains almost complete inactivity of Sso Pus7 at 55°C which we observed for Sso Pus7. However, some activity of Sso Pus7 at position 13 in tRNAs was detected at 80°C. Thus, at high temperatures other amino acids of the catalytic site may contribute to D89 activation due to the higher flexibility of the protein backbone. Our observation that the Sso Pus7 activity is markedly increased when the conserved basic residue in motif IIIa is restored, also points out the functional importance of this residue for catalysis. In contrast, amino acid

residues in motif II (RH/KR) seem to be less important for U isomerization. Accordingly, some of the Desulfurococcales species, that do not contain Ψ 35-specific H/ACA sRNA, show mutations in the RH/KR conserved pair in motif II of Pus7. However, all these proteins conserved the basic residue of motif II, therefore these Pus7 enzymes are probably able to modify U at position 35 in tRNA^{Tyr}(GUA).

In contrast, no activity of Sso Pus7 at position 35 was detected, both at 55°C and at 80°C. This restricted specificity of Sso Pus7 towards tRNAs compared to Pab Pus7 may be explained by a difference in orientation of the tRNA substrate in substrate binding pocket. Indeed, modification of tRNA at position 13 implies the interaction with the tRNA D-stem, while U35 is located in the anticodon. Thus, some additional stabilizing RNA-protein interactions outside of the enzyme active site may compensate for the absence of proper catalytic functional groups. .

Even if Sso Pus7 activity at position 13 in tRNAs is very low, it may be sufficient to ensure a minimum level of modification at this position *in vivo*. In contrast, Sso Pus7 is probably unable to ensure sufficient modification at position 35 in tRNAs^{Tyr}, explaining why a specific H/ACA sRNA dedicated to Ψ 35 formation in tRNA^{Tyr} appeared in Sulfolobales and *A. pernix*

The high functional importance of Ψ 35 in archaeal tRNA^{Tyr} may explain the requirement for an H/ACA sRNA in Sulfolobales

The need for a guide RNA for Ψ 35 formation in Sulfolobales tRNA^{Tyr} may be related to a high functional importance of Ψ 35 in tRNA^{Tyr}. A great number of experiments were performed in order to define the function of this modified residue in tRNA^{Tyr}. First of all, residue Ψ 35 was shown to reinforce the stability of the anticodon stem-loop structure by establishment of a hydrogen bond between the N₁-H of its nucleobase and the ribose O₂' of the residue U33. This N₁-H...O₂' hydrogen bond is more stable than the weak C₅-H...O₂' bond

formed in the absence of pseudouridine (Auffinger & Westhof, 1997). This structural stabilization was proposed to influence the efficiency of tRNA^{Tyr} aminoacylation by its cognate aminoacyl-tRNA synthetase (Bare & Uhlenbeck, 1986). However, primary action seems to consist in the stabilisation of the codon-anticodon interaction. This was shown by intensive studies of suppressor tRNAs^{Tyr}. For instance, the absence of Ψ35 in SUP6 tRNA^{Tyr}(UUA) was found to decrease its *in vivo* ability to read-through the ochre stop codon UAA (Johnson & Abelson, 1983). Similarly, the presence of residue Ψ35 increased the tRNA^{Tyr} suppression activity on ochre (UAA) and amber (UAG) stop codons in the Tobacco Mosaic Virus RNA (Zerfass & Beier, 1992). One can imagine that the stabilization of codon-anticodon interactions is particularly important for organisms, growing at very high temperatures, like the hyperthermophilic Archaea. Therefore, it is reasonable to expect an important role of residue Ψ35 in the codon-anticodon interaction during mRNA decoding in these organisms.

Sulfolobales and *A. pernix* are the first organisms found to contain H/ACA guides for tRNA modification

Up to now, as well in Bacteria as in Eukarya, all tRNA posttranscriptional modifications were shown to be catalyzed by enzymes that do not require a guide RNA. In contrast, an important implication of C/D box sRNAs for tRNA 2'-O-methylation was already demonstrated in Archaea (Renalier *et al.*, 2005). However, up to now, H/ACA sRNA dedicated to tRNAs had not been discovered. The H/ACA sRNAs that we found in four Sulfolobales species and *A. pernix* are the first example of a tRNA-specific H/ACA guide RNAs. By using the computational approach that we developed for the search of H/ACA sRNA genes in archaeal genomes (Muller, 2007 #49}, we tested whether other genes for tRNA specific H/ACA sRNAs exist in the 48 available complete archaeal genome sequences. However, no other

examples of H/ACA sRNAs dedicated to tRNA modification were found. Therefore, we concluded that the H/ACA sRNA present in Sulfolobales and *A. pernix* constitute a unique example of adaptative evolution: an H/ACA sRNA that was probably initially dedicated to rRNA modification was deviated from its original target to complement the deficiency of a tRNA:Ψ³⁵-synthase. The shift from a non-guided to guided post-transcriptional machinery was already observed for 2'-*O*-methylation at position 56 in archaeal tRNAs (Renalier *et al.*, 2005). Formation of Cm56 in the tRNA TΨ-loop is catalysed by the specific tRNA: 2'-*O*-methyltransferase (aTrm56) in Archaea (Renalier *et al.*, 2005). Surprisingly, the homologous protein is missing in the crenarchaeon *Pyrobaculum aerophilum*, while the Cm56 modification is still present. The absence of aTrm56 in this organism is in fact compensated by the appearance of a C/D box sRNA dedicated to modification at this position (Renalier *et al.*, 2005). Hence, in this case, an additional C/D sRNA was used to achieve an important 2'-*O*-methylation which is normally catalyzed by a stand-alone enzyme. Thus, in order to complement the absence of an important RNA:Ψ-synthase, an H/ACA sRNA guided system was selected to insure modification of tRNAs, which are normally pseudouridylated by stand-alone enzymes. Altogether, these findings demonstrate the extreme plasticity of the RNA modification machinery in Archaea, and more generally, the adaptation potential constituted by the sRNA guides.

tRNA and rRNA modification may be tightly related in Archaea

The presence of tRNA-specific H/ACA sRNA guides in certain archaeal organisms demonstrate the tight relations existing between modification of various classes of RNA the cell. Indeed, in Archaea and eukaryotes sRNA-guided machinery is mostly dedicated to rRNA, but also to sn/snoRNA. In our study we characterized several tRNA-specific H/ACA sRNA in Sulfolobales. Interestingly, the Ape-7 sRNA, in addition to tRNA modification, has also conserved the possibility to guide Ψ formation at position 2444 in 23S rRNA (Fig. 1), U

residue in analogous position is modified by the Pab91 sRNA in *P. abyssi* (Charpentier *et al.*, 2005). The Ape-7 sRNA probably gained an additional activity at position 35 of the pre-tRNA^{Tyr}(GUA) after deleterious mutations in Pus7p. The Ape-7 sRNA with the two possible activities may correspond to a transitory situation. Note that base-pair interaction between the pre-tRNA^{Tyr} and the 5' guide sequence is strong but the interaction with the 3' guide has a lower stability. In contrast, the four other H/ACA sRNAs found in Sulfolobales do not guide anymore modification in rRNAs. They are strictly dedicated to tRNAs and interactions with both the 5' and the 3' guide sequences are extremely stable.

Remarkably, the genes encoding Sso1, Sto1 and Sac1 H/ACA sRNAs contain a segment encoding a C/D box sRNA at their 3'-extremities (Fig. 1). In this work we showed that a transcript containing both the H/ACA and the C/D sRNA is indeed expressed in *S. solfataricus*. This observation is in agreement with recent data showing that this C/D sRNA denoted Sso-159 is expressed as a 120-nt long RNA (Tang *et al.*, 2002). The Sso-159 sRNA was proposed to guide 2'-*O*-methylation at position U2621 in 23S rRNA. The C/D motifs of the Sto1 sR4 and Sac1 sR39 C/D box sRNAs, can also guide 2'-*O*-methylation at the same positions in 23S rRNA (Fig. 7). In addition, we propose that the these 3 C/D sRNAs might also guide 2'-*O*-methylation of 23S rRNA at positions 2645 and 2648 in (*S. solfataricus* numbering, Fig. 7). Altogether, this study show that sRNA sequences which have a capacity of rapid evolution represent a flexible adaptative reservoir for RNA modification in archaea and probably also in Eukarya.

MATERIALS AND METHODS

DNA constructs for the production of Pus7 proteins and their variants

The DNA fragments encoding WT Pus7 proteins from *P. abyssi* and *S. solfataricus* were obtained by a PCR amplification using genomic DNA from these two species. Oligonucleotide primers, used for amplification contained NheI and BamHI restriction sites for forward and reverse primers, respectively. Amplified DNA was digested by NheI and BamHI and inserted into pET28b plasmid digested by the same restrictases. The variants of Pus7 proteins were produced by site-directed mutagenesis (QuickChange kit, Stratagene, USA). The sequence of the oligonucleotides used for the PCR amplifications and site-directed mutagenesis are available on request.

Purification of recombinant proteins

Wild-type and variant Pus7 proteins were produced in *E. coli* as His₆-tagged proteins and purified using affinity chromatography on Ni-NTA agarose (Qiagen, France). Induction of protein production was done by 1 mM IPTG in LB media supplemented by 50 µg/ml of kanamycin. Cell pellet was resuspended in Lysis buffer (20 mM Tris-HCl, pH7.7; 300 mM NaCl; 1 mM DTT ;10 mM Imidazole) and sonicated. Cell-free extract was directly loaded onto Ni-NTA agarose. After washing step at 20 mM of imidazole in the same buffer, His₆-tagged proteins were eluted in the Lysis buffer containing 250 mM of imidazole. Purified proteins were stored at ambient temperature in the elution buffer supplemented with , 10% glycerol, or at -80°C by adjusting the glycerol concentration to 50%.

The *P. abyssi* aCBF5, aNOP10, aGAR1 and L7Ae were produced as GST fusion proteins as previously described (Charpentier *et al.*, 2005). They were stored at -80°C in the following buffer 50 mM Tris-HCl pH 7.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, glycerol 10%.

***In vitro* transcription of tRNAs, H/ACA sRNA and the sRNP target RNA**

Wild-type and mutated tRNA^{Asp} and tRNA^{Tyr} from *P. abyssi* and *S. solfataricus* and the Sso1 H/ACA sRNA from *S. solfataricus* were produced by *in vitro* transcription of linearized plasmids. The DNA templates coding for those tRNAs and sRNA were amplified by PCR using genomic DNAs of *P. abyssi* strain GE5 or *S. solfataricus* strain P2. Forward primer used for amplification bears the sequence of T7 RNA polymerase promoter. Amplified products were cloned in pUC18 vector through SmaI restriction site. Linearization was performed using BstNI restriction endonuclease.

Templates encoding the minisubstrates derived from *S. solfataricus* tRNA^{Tyr} (see Figure 6) were obtained by annealing of two complementary oligonucleotides. Additional tRNA^{Asp} and tRNA^{Tyr} variants were produced by site-directed mutagenesis (QuickChange kit, Stratagene, USA). All synthetic RNA transcripts were purified by denaturing gel electrophoresis in 8% urea-PAAG. For the pseudouridylation assays, the corresponding RNA transcripts were uniformly labeled with [α -³²P]ATP or other [α -³²P]NTPs. Conditions for transcription and labeling were described previously (Charpentier *et al.*, 2007).

Cellular extracts

About 10¹⁰ cells from *P. abyssi* or *S. solfataricus* pellets grown to exponential phase were resuspended in the extraction buffer (25 mM Tris-HCl, pH 7,5 ; 25 mM KCl; 5 mM MgCl₂ ; 10 % Glycérol ; 2 mM DTT). The cells were sonicated for 5 min followed by centrifugation at 10 000 g. The concentration of glycerol was adjusted at 20% for storage at -20°C.

***In vitro* pseudouridylation assay**

The tRNA:Ψ35-synthase activity of purified Pus7 or H/ACA sRNP complexes was measured by the nearest-neighbor approach in the conditions described previously (Charpentier *et al.*, 2005; Charpentier *et al.*, 2007). The sRNA guided activity of reconstituted H/ACA sRNPs was measured as previously described by mixing of ~ 4 pmol of Sso1 sRNA with ~150 fmol

of the [α - ^{32} P]ATP-labeled tRNA in buffer D (20 mM HEPES, pH 7.9, 150 mM KCl; 1.5 mM MgCl₂; 0.2 mM EDTA) along with aCBF5, aNOP10, aGAR1 and L7Ae proteins (Charpentier *et al.*, 2005). The non RNA-guided reaction of Pus7 on tRNA was tested on ~50 fmol of [α - ^{32} P]ATP-labeled tRNAs in buffer Psi (100 mM Tris-HCl, pH8.0; 100 mM AcNH₄; 5 mM MgCl₂; 2 mM DTT; 100 μ M EDTA). The activity assays with cell-free extracts were performed with the same amount of labeled tRNA in buffer D in the presence of a similar amount of UDO_{260nm} of *P. abyssi* and *S. solfataricus* cellular extracts. For both sRNA-guided and non RNA-guided reactions, samples were incubated at 55°C, 65°C or 80°C and the reaction was started by addition of the proteins (200 nM each). Activity assays with cellular extracts were carried out at 65°C. Then, the RNAs were extracted and digested by RNase T2. The resulting 3' phosphate-mononucleotides were fractionated by 2D thin-layer cellulose chromatography (Keith, 1995). The radioactivity in the spots was quantified using the ImageQuant software after exposure of a phosphorimager screen. The amount of Ψ residue formed was determined taking into account the total number of U residues in the tRNA.

Mapping the site of pseudouridylation in tRNAs

Pseudouridine mapping in RNAs was done using the *N*-cyclohexyl-*N'*-(2-morpholinoethyl)-carbodiimide metho-*p*-toluolsulfonate (CMCT) modification protocol and primer extension with the reverse transcriptase (Bakin & Ofengand, 1993) for review see (Motorin *et al.*, 2007). CMCT modification was performed with 0.5 μ g of *in vitro* transcribed *S. solfataricus* tRNA^{Tyr}. Positions of alkali-resistant CMCT modifications were identified by primer extension analysis, using the AMV RT (QBiogene, USA) in the conditions previously described (Motorin *et al.*, 2007) using an oligonucleotides complementary to the 14 nt at the 3' end of the tRNA. Primers were 5'-end labeled using T4 polynucleotide kinase and [γ - ^{32} P]ATP (3000 Ci/mmol). RNA sequencing was done with 4 μ g of *in vitro* transcribed RNA (for details on the CMCT/RT technique, see (Motorin *et al.*, 2007)).

ACKNOWLEDGMENTS

S. Muller and A. Urban were a pre-doctoral fellow supported by the French Ministère de la Recherche et des Nouvelles Technologies (MNRT). The work was supported by the Centre National de la Recherche Scientifique (CNRS), the MNRT and the PRST « Bioingénierie » of Région Lorraine. We thank P. Forterre (IGM, Orsay, France) for helpful discussions.

BIBLIOGRAPHY

- Ansmant I, Motorin I. 2001. [Identification of RNA modification enzymes using sequence homology]. *Molekuliarnaia biologiya* 35:248-267.
- Ansmant I, Motorin Y, Massenet S, Grosjean H, Branlant C. 2001. Identification and characterization of the tRNA:Psi 31-synthase (Pus6p) of *Saccharomyces cerevisiae*. *J Biol Chem* 276:34934-34940.
- Auffinger P, Westhof E. 1997. RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *J Mol Biol* 269:326-341.
- Auxilien S, El Khadali F, Rasmussen A, Douthwaite S, Grosjean H. 2007. Archease from *Pyrococcus abyssi* improves substrate specificity and solubility of a tRNA m5C methyltransferase. *J Biol Chem* 282:18711-18721.
- Bakin A, Ofengand J. 1993. Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center - analysis by the application of a new sequencing technique. *Biochemistry* 32:9754-9762.
- Bare LA, Uhlenbeck OC. 1986. Specific substitution into the anticodon loop of yeast tyrosine transfer RNA. *Biochemistry* 25:5825-5830.
- Becker HF, Motorin Y, Planta RJ, Grosjean H. 1997. The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs. *Nucleic Acids Res* 25:4493-4499.
- Behm-Ansmant I, Branlant C, Motorin Y. 2007. The *Saccharomyces cerevisiae* Pus2 protein encoded by YGL063w ORF is a mitochondrial tRNA:Psi27/28-synthase. *RNA* 13:1641-1647.
- Behm-Ansmant I, Grosjean H, Massenet S, Motorin Y, Branlant C. 2004. Pseudouridylation at position 32 of mitochondrial and cytoplasmic tRNAs requires two distinct enzymes in *Saccharomyces cerevisiae*. *J Biol Chem* 279:52998-53006.
- Behm-Ansmant I, Urban A, Ma X, Yu Y-T, Motorin Y, Branlant C. 2003. The *Saccharomyces cerevisiae* U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. *RNA* 9:1371-1382.
- Chaban B, Ng SY, Jarrell KF. 2006. Archaeal habitats--from the extreme to the ordinary. *Can J Microbiol* 52:73-116.
- Charpentier B, Fourmann JB, Branlant C. 2007. Reconstitution of Archaeal H/ACA sRNPs and Test of their Activity. *Methods Enzymol* 425:389-405.
- Charpentier B, Muller S, Branlant C. 2005. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res* 33:3133-3144.
- Constantinesco F, Motorin Y, Grosjean H. 1999. Transfer RNA modification enzymes from *Pyrococcus furiosus*: detection of the enzymatic activities in vitro. *Nucleic Acids Res* 27:1308-1315.
- Del Campo M, Kaya Y, Ofengand J. 2001. Identification and site of action of the remaining four putative pseudouridine synthases in *Escherichia coli*. *RNA* 7:1603-1615.
- Del Campo M, Ofengand J, Malhotra A. 2004. Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of *Escherichia coli*. *RNA* 10:231-239.
- Dennis PP, Omer A, Lowe T. 2001. A guided tour: small RNA function in Archaea. *Mol Microbiol* 40:509-519.
- Gregson JM, Crain PF, Edmonds CG, Gupta R, Hashizume T, Phillipson DW, McCloskey JA. 1993. Structure of the archaeal transfer RNA nucleoside G*-15 (2-amino-4,7-dihydro- 4-oxo-7-beta-D-ribofuranosyl-1H-pyrrolo[2,3-d]pyrimidine-5-carboximide (archaeosine)). *J Biol Chem* 268:10076-10086.
- Grosjean H, Constantinesco F, Foiret D, Benachenhou N. 1995. A novel enzymatic pathway leading to 1-methylinosine modification in *Haloferax volcanii* tRNA. *Nucleic Acids Res* 23:4312-4319.
- Gupta R. 1984. *Halobacterium volcanii* tRNAs. Identification of 41 tRNAs covering all amino acids, and the sequences of 33 class I tRNAs. *J Biol Chem* 259:9461-9471.
- Gutgsell N, Englund N, Niu L, Kaya Y, Lane BG, Ofengand J. 2000. Deletion of the *Escherichia coli* pseudouridine synthase gene *truB* blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells. *RNA* 6:1870-1881.
- Hamma T, Ferre-D'Amare AR. 2006. Pseudouridine synthases. *Chem Biol* 13:1125-1135.
- Hepfer CE, Arnold-Croop S, Fogell H, Steudel KG, Moon M, Roff A, Zaikoski S, Rickman A, Komsisky K, Harbaugh DL, Lang GI, Keil RL. 2005. DEG1, encoding the tRNA:pseudouridine synthase Pus3p, impacts HOT1-stimulated recombination in *Saccharomyces cerevisiae*. *Mol Genet Genomics* 274:528-538.
- Hoang C, Ferre-D'Amare AR. 2004. Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. *RNA* 10:1026-1033.

- Huang L, Pookanjanatavip M, Gu X, Santi DV. 1998. A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry* 37:344-351.
- Ishitani R, Nureki O, Fukai S, Kijimoto T, Nameki N, Watanabe M, Kondo H, Sekine M, Okada N, Nishimura S, Yokoyama S. 2002. Crystal structure of archaeosine tRNA-guanine transglycosylase. *J Mol Biol* 318:665-677.
- Johnson PF, Abelson J. 1983. The yeast tRNA^{Tyr} gene intron is essential for correct modification of its tRNA product. *Nature* 302:681-687.
- Kammen HO, Marvel CC, Hardy L, Penhoet EE. 1988. Purification, structure, and properties of Escherichia coli tRNA pseudouridine synthase I. *J Biol Chem* 263:2255-2263.
- Kaya Y, Del Campo M, Ofengand J, Malhotra A. 2004. Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *J Biol Chem* 279:18107-18110.
- Kaya Y, Ofengand J. 2003. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA* 9:711-721.
- Keith G. 1995. Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie* 77:142-144.
- Lafontaine DL, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D. 1998. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev* 12:527-537.
- Ma X, Zhao X, Yu YT. 2003. Pseudouridylation (Psi) of U2 snRNA in S.cerevisiae is catalyzed by an RNA-independent mechanism. *EMBO J* 22:1889-1897.
- Massenet S, Motorin Y, Lafontaine DL, Hurt EC, Grosjean H, Branlant C. 1999. Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol Cell Biol* 19:2142-2154.
- Massenet S, Mougin A, Branlant C. 1998. Posttranscriptional modifications in the U small nuclear RNAs. In: Grosjean H, Benne R, eds. *The Modification and Editing of RNA*. New York: ASM Press. pp 201-227.
- McCloskey JA, Graham DE, Zhou S, Crain PF, Ibba M, Konisky J, Soll D, Olsen GJ. 2001. Post-transcriptional modification in archaeal tRNAs: identities and phylogenetic relations of nucleotides from mesophilic and hyperthermophilic Methanococcales. *Nucleic Acids Res* 29:4699-4706.
- Motorin Y, Keith G, Simon C, Foiret D, Simos G, Hurt E, Grosjean H. 1998. The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA* 4:856-869.
- Motorin Y, Muller S, Behm-Ansmant I, Branlant C. 2007. Identification of Modified Residues in RNAs by Reverse Transcription-Based Methods. *Methods Enzymol* 425:21-53.
- Muller S, Charpentier B, Branlant C, Leclerc F. 2007a. A Dedicated Computational Approach for the Identification of Archaeal H/ACA sRNAs. *Methods Enzymol* 425:355-387.
- Muller S, Fourmann JB, Loegler C, Charpentier B, Branlant C. 2007b. Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Psi55-synthase and RNA-guided RNA:Psi-synthase activities. *Nucleic Acids Res* 35:5610-5624.
- Nishikura K, De Robertis EM. 1981. RNA processing in microinjected Xenopus oocytes. Sequential addition of base modifications in the spliced transfer RNA. *J Mol Biol* 145:405-420.
- Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP. 2000. Homologs of small nucleolar RNAs in Archaea. *Science*, 288:517-522.
- Raychaudhuri S, Niu L, Conrad J, Lane BG, Ofengand J. 1999. Functional effect of deletion and mutation of the Escherichia coli ribosomal RNA and tRNA pseudouridine synthase RluA. *J Biol Chem* 274:18880-18886.
- Renalier MH, Joseph N, Gaspin C, Thebault P, Mougin A. 2005. The Cm56 tRNA modification in archaea is catalyzed either by a specific 2'-O-methylase, or a C/D sRNP. *RNA* 11:1051-1063.
- Roovers M, Hale C, Tricot C, Terns MP, Terns RM, Grosjean H, Droogmans L. 2006. Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic Acids Res* 34:4293-4301.
- Simos G, Tekotte H, Grosjean H, Segref A, Sharma K, Tollervey D, Hurt EC. 1996. Nuclear pore proteins are involved in the biogenesis of functional tRNA. *EMBO J* 15:2270-2284.
- Sprinzl M, Vassilenko KS. 2005. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* 33:D139-140.
- Szweykowska-Kulinska Z, Beier H. 1992. Sequence and structure requirements for the biosynthesis of pseudouridine (psi 35) in plant pre-tRNA(Tyr). *EMBO J* 11:1907-1912.
- Tang TH, Bachellerie JP, Rozhdetsvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Huttenhofer A. 2002. Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proc Natl Acad Sci USA* 99:7536-7541.
- Tomita K, Ueda T, Watanabe K. 1999. The presence of pseudouridine in the anticodon alters the genetic code: a possible mechanism for assignment of the AAA lysine codon as asparagine in echinoderm mitochondria. *Nucleic Acids Res* 27:1683-1689.

- Watanabe Y, Gray MW. 2000. Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaeobacteria. *Nucleic Acids Res* 28:2342-2352.
- Watkins NJ, Gottschalk A, Neubauer G, Kastner B, Fabrizio P, Mann M, Luhrmann R. 1998. Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA*, 4:1549-1568.
- Wrzesinski J, Bakin A, Ofengand J, Lane BG. 2000. Isolation and properties of *Escherichia coli* 23S-RNA pseudouridine 1911, 1915, 1917 synthase (RluD). *IUBMB Life* 50:33-37.
- Zerfass K, Beier H. 1992. Pseudouridine in the anticodon G psi A of plant cytoplasmic tRNA(Tyr) is required for UAG and UAA suppression in the TMV-specific context. *Nucleic Acids Res* 20:5911-5918.

FIGURE LEGENDS

Figure 1 Panel A: Structures of tRNA^{Tyr}(GUA) Ψ35-specific sRNA guides from *A. pernix*, *M. sedula*, *S. solfataricus*, *S. tokodaii* and *S. acidocaldarius*. Characteristic K-turns and H/ACA sequences are boxed. C/D box sRNAs present at the 3' extremities of Sso1 (sR159), Sac1 (sR4) and Sto1 (sR39) are indicated. Panel B: Proposed base-pairing of Ape7, Mse2, Sso1, Sto1 and Sac1 with their tRNA^{Tyr}(GUA) or LSU RNA targets. An arrow indicates the site of U to Ψ conversion. The distance between the pseudouridylation pocket and K-turn and ACA-box is shown. Panel C: Northern blot analysis of Sso1 expression in *S. solfataricus*. Total RNA was separated on urea – PAAG, RNAs were transferred to Nylon membrane (Hybond N+, GE Healthcare, USA) and hybridized with Sso1-specific [³²P]-labeled oligonucleotide. Molecular weight marker is given on the left. Arrows indicate the positions of free Sso1 sRNA guide (H/ACA) and its C/D-box 3'-extended form (H/ACA+C/D).

Figure 2 Sso1 sRNA-guided modification of Ψ35 in pre-tRNA^{Tyr}(GUA). Panel A: Cloverleaf structure of unmodified tRNA transcript corresponding to *S. solfataricus* tRNA^{Tyr}(GUA) precursor, position of U35 is circled. The sequence of the intron present in the anticodon loop of this tRNA is given in small letters. Panel B: *In vitro* reconstitution tests to reveal Sso1-guided Ψ35 formation. [α -³²P]ATP-labeled tRNA^{Tyr} precursor was incubated for 90 min at 65°C in the presence of L7Ae/CBF5/NOP10/GAR1 (LCNG), the H/ACA sRNA Sso1 and the four proteins (Sso1/LCNG) and *S. solfataricus* cell-free extract (Sso extract). Additional incubation in the Sso extract was performed with [α -³²P]ATP-labeled Sso pre-tRNA^{Tyr}(GUA) with U35->C substitution. Modified transcripts were digested by RNase T2 and the resulting 3'-NMPs were analysed by 2D TLC (Keith, 1995). Position of the

pseudouridine spot is circled. The 3'-NMP composition of the RNase T2 digest as well as molar proportion of the formed Ψ residue are indicated on each panel. Panel C: CMCT-RT mapping of the pseudouridine residue formed upon *S. solfataricus* tRNA^{Tyr}(GUA) modification in the reconstituted sRNPs (Sso1/LCNG) and in *S. solfataricus* cell-free extract (Sso extract). Modified RNA was treated by CMCT, for 2, 10 and 20 min with (+) or without (-) subsequent alkaline treatment (OH-) and analyzed by primer extension. A control experiment was performed in the absence of CMCT treatment. Lanes U, G, C and A correspond to the sequencing ladders obtained with the same oligonucleotide. An arrow shows the position of the RT stop corresponding to Ψ 35.

Figure 3 Panel A: Multiple sequence alignment of Pus7 proteins from various Archaea. Only the regions of Motif IIIa and Motif II containing the catalytic D residue are shown. Positions of the conserved motifs are given for each sequence. Arrows indicate the organisms possessing the tRNA^{Tyr}-specific H/ACA sRNA. Panel B: Differences found in the active sites of Pab Pus7 and Sso Pus7. 3D modelling of the active sites of Pab Pus7 (left) and Sso Pus7 (right) was done using the X-ray structure of *M. masei* TruD (PDB 1Z2Z). Only a highly conserved amino acids are shown.

Figure 4 Enzymatic activity of WT Pab Pus7, Sso Pus7 and Pab extract on different tRNA substrates. Panel A : Cloverleaf structures of tRNA^{Tyr}(GUA) and tRNA^{Asp}(GUC) from *P. abyssi* and tRNA^{Tyr}(GUA) precursor and tRNA^{Asp}(GUC) from *S. solfataricus*. All 5'-UA-3' contexts present in these tRNAs are indicated by arrows. Potential natural sites of Ψ formation (U13 in tRNA^{Asp} and U35 in tRNA^{Tyr}) are circled. Panels B: Results of *in vitro* activity tests performed with two *P. abyssi* tRNA substrates and their U35C and U13C variants. Panel C: Results of *in vitro* activity tests performed with two *S. solfataricus* tRNA substrates and their U35C variant of tRNA^{Tyr}. All

transcripts were [α - ^{32} P]ATP-labeled and incubated for 90 min with the recombinant proteins (Pab Pus7 or Sso Pus7) at 55°C or with Pab extract at 65°C. Modified RNAs were analysed as described in the legend to Figure 2. Panel D: Control incubations performed at 55°C with Pab Pus7 and *P. abyssi* tRNA^{Tyr} labeled by [α - ^{32} P]CTP, [α - ^{32} P]UTP and [α - ^{32} P]GTP. Incubation conditions as as for Panel B and C.

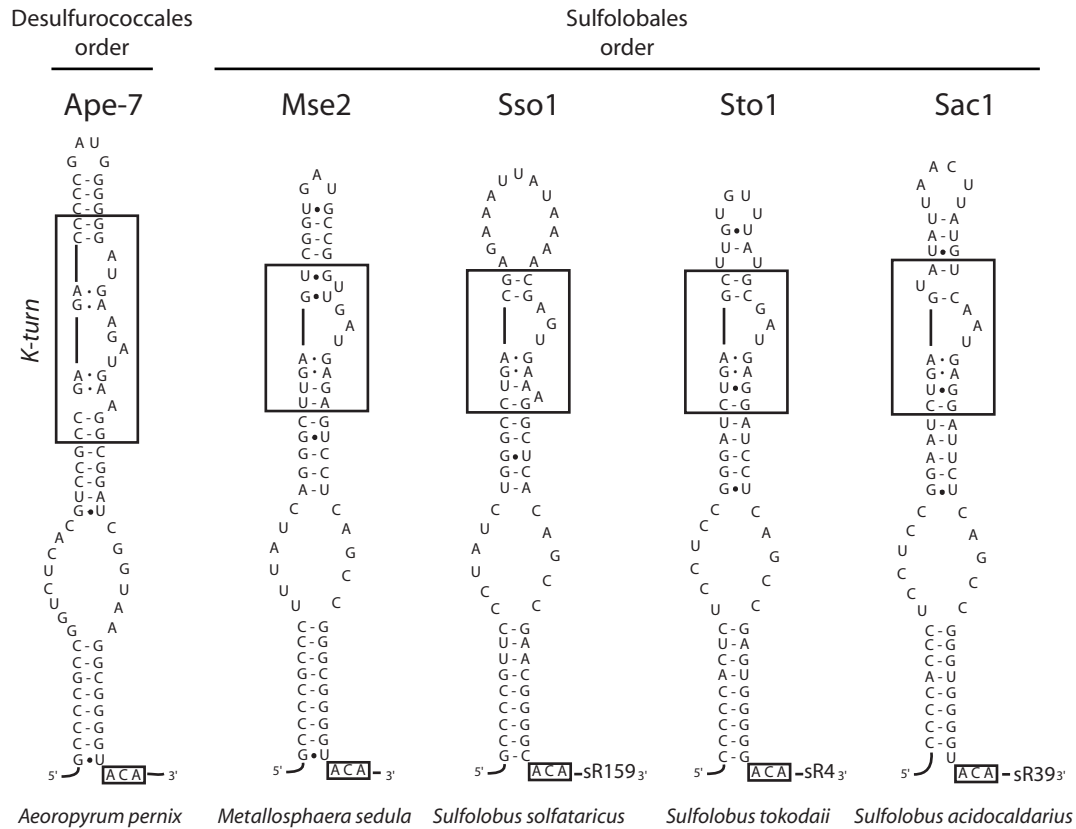
Figure 5 Enzymatic activity of Pab Pus7 and Sso Pus7 protein variants with [α - ^{32} P]ATP-labeled *P. abyssi* tRNA^{Asp}(GUC). Enzymatic tests were performed as described in the legend of Figure 4, except that point-mutated variants of Pab Pus7 and Sso Pus7 were used. Tests were done at 55°C (Panel A) and at 80°C (Panel B) and a molar ration of formed Ψ residue/tRNA is given in a histogram. Error bars were calculated using experiments done in triplicate and independent quantification of Ψ MP spot relative to four other NMP spots.

Figure 6 Pseudouridine Ψ 35-formation in WT and A36G variants of tRNA^{Tyr} from *P. abyssi* and pre-tRNA^{Tyr} from *S. solfataricus*. Incubations for 90 min with the recombinant Pab Pus7 were performed at 55°C, while the tests with *P. abyssi* (Pab), *S. solfataricus* (Sso) extracts and LCNG/Sso1 sRNP were performed at 65°C (the optimal temperature for cell-free extracts). [α - ^{32} P]ATP or [α - ^{32} P]GTP-labeled tRNA transcripts were used. Analysis of Ψ formation was done as described in the legend to Fig. 2. The 3'-NMP composition of the RNase T2 digest as well as molar proportion of the formed Ψ residue are indicated on each panel. Additional spots visible in [α - ^{32} P]ATP-labeled tRNA incubated in cellular extracts correspond to formation of methylated nucleotides naturally present in many archaeal tRNAs (m¹G37, Cm56, m¹I57 and m¹A58).

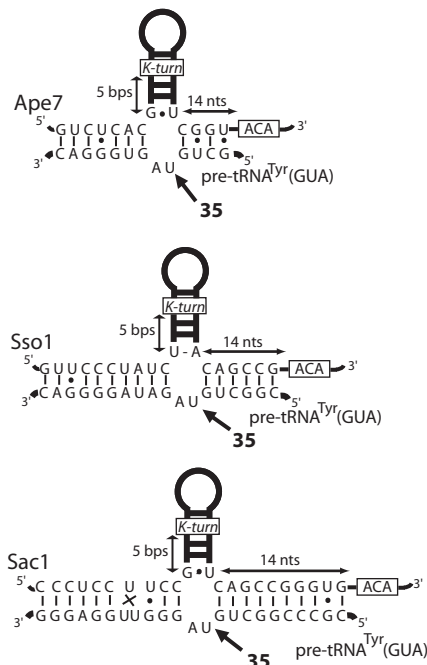
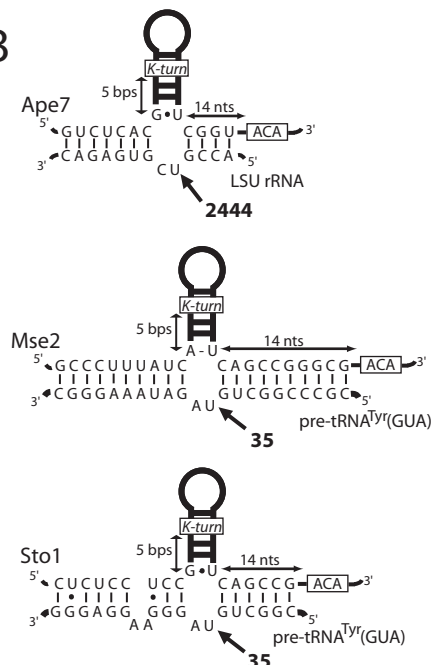
Figure 7 Panel A: Predicted secondary structure of the extended form of Sso1 sRNA containing both H/ACA and C/D motifs. K-turn, ACA-box, as well as C/D and C'/D'

boxes are indicated. Panel B: Predicted base-pairing between Sso1-sR159, Sto1-sR4 and Sac1-sR39 with the corresponding 23S rRNA targets. Two additional potential targets in 23S rRNA for sR159 are shown on the bottom.

A



B



C

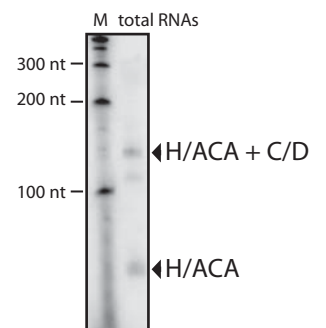


Figure 1

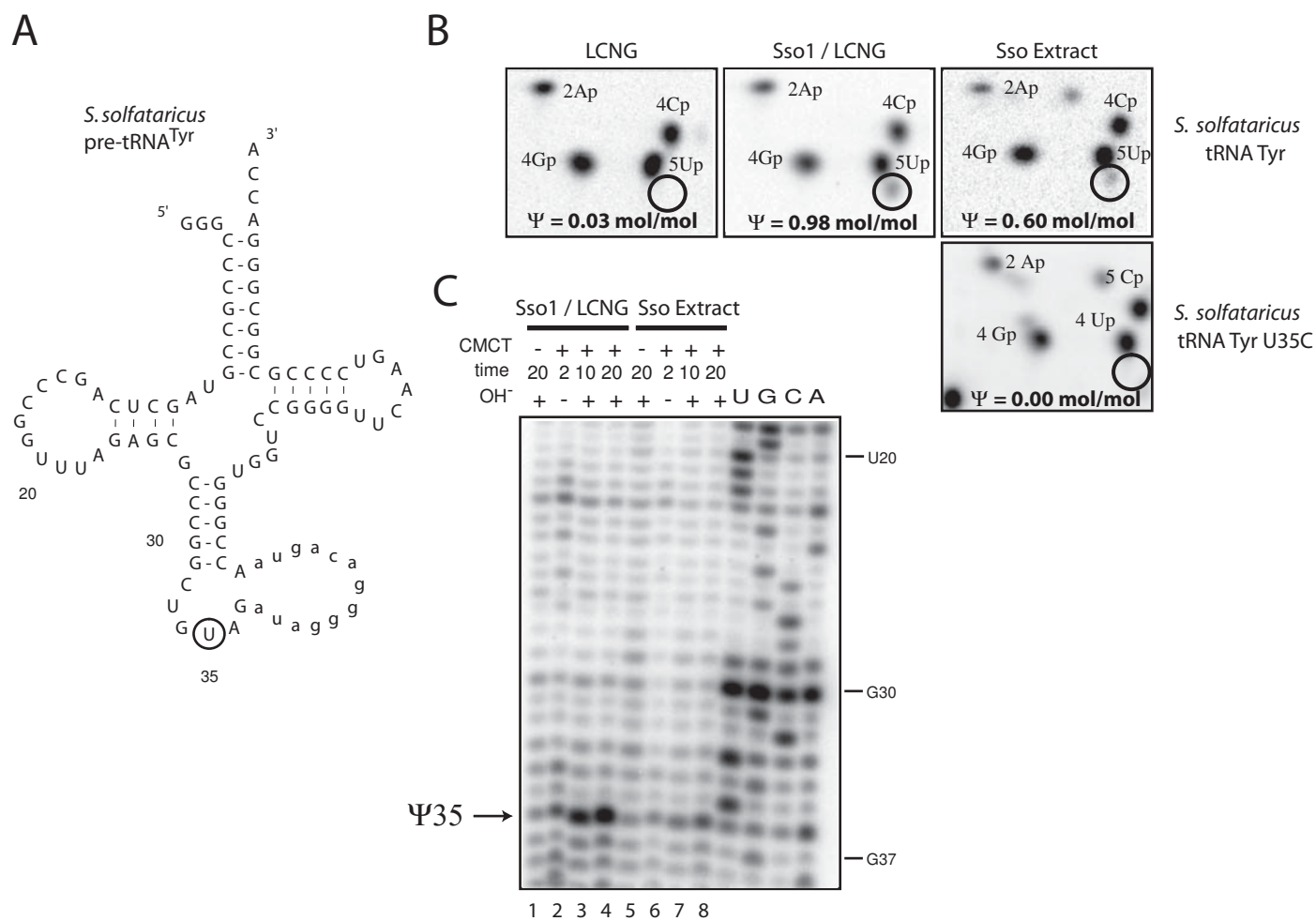


Figure 2

A

		K19	F25	Pab Pus7	D77 R78/H79	
		I27	F33	Sso Pus7	D89 A90/N91	
<i>M. acetivorans</i>	24	GLGGRL	QOEVE	FIVKEITN	78 SQKRISVAGTKD	KRALTTQKI
<i>M. mazei</i>	21	GLGGQL	QOEIE	FIVKEITN	75 SQKRISVAGTKD	KRALTTQKI
<i>H. salinarium</i>	22	GTGGRL	RDSPAD	FRVRELEA	82 SRERVRWAGTKD	RHAVTTQLF
<i>A. fulgidus</i>	16	GMGGEI	KAEPED	FVVEEIAE	70 SQKRISFAGTKD	KRALTVQYF
<i>T. acidophilum</i>	14	RPVIRI	KENPED	FTVEEVAD	68 SRKRISFAGTKD	KARKIQYF
<i>T. volcanium</i>	8	EEAIEI	KRDPED	FSVEEIID	62 SRKRVTFAGTKD	KRAVKLQYF
<i>P. furiosus</i>	16	GIGGKI	VYPED	FIVKEVIP	70 HYSEIGFAGTKD	RHAVTYQYI
<i>P. horikoshii</i>	17	GIGGKI	VYPED	FIVKEVIP	71 HYSEIGFAGTKD	RHAVTYQYI
<i>P. abyssi</i>	13	GIGGRI	KVPED	FIVREIIP	66 HYSEIGFAGTKD	RHAVTYQYI
<i>M. jannaschii</i>	51	-TGGI	IKYPE	FIVVEITP	119 QRKHFGFAGNKD	KYAVTTQRV
<i>M. thermauto.</i>	13	GTGGRI	RVHNRD	FVEVEIIP	67 DRRRMGFAGMKD	KRAVTRQWI
→ <i>S. solfataricus</i>	21	NELDGT	TERPNG	FKVIEEID	79 LHSKVNYIGIKD	ANAITSQLV
→ <i>S. tokodaii</i>	14	SPLKVF	INRPDG	FRVIEEIS	71 LHSKIHYIGIKD	TNAITEQIV
→ <i>S. acidocaldarius</i>	22	EPLQIS	INRPEG	FKVIEEIS	79 MKRKIHYLGIKD	ANAITQQII
→ <i>M. sedula</i>	20	ASIRAE	IPRPDG	FRVTEEID	81 LGEKPRYLGIKD	TNAVTEQLI
<i>P. aerophilum</i>	18	PSGGVI	KRSPED	FVVEEVLL	86 SHRDVSVGGLKD	TAVTSQII
→ <i>A. pernix</i>	25	RGPEAL	IMFPWG	FRVVESRV	77 -CRLYSYAGLKD	ACSVSWQHV
						★
						Motif II
						Motif IIIa

B

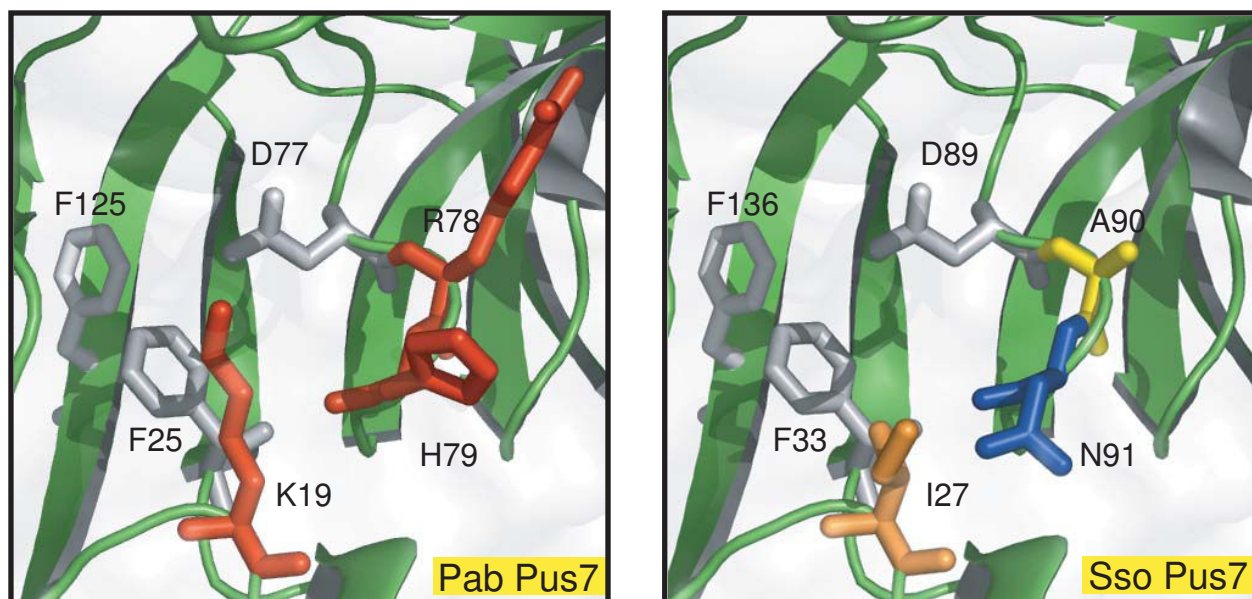


Figure 3

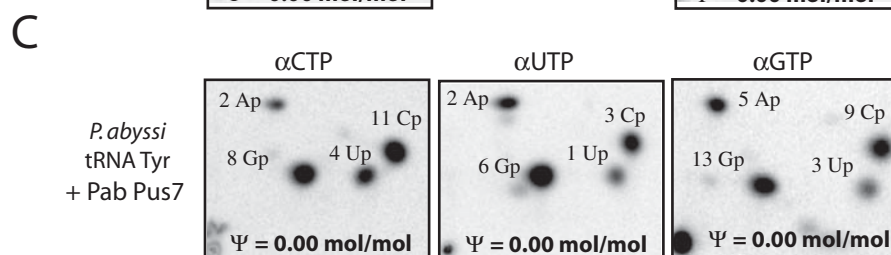
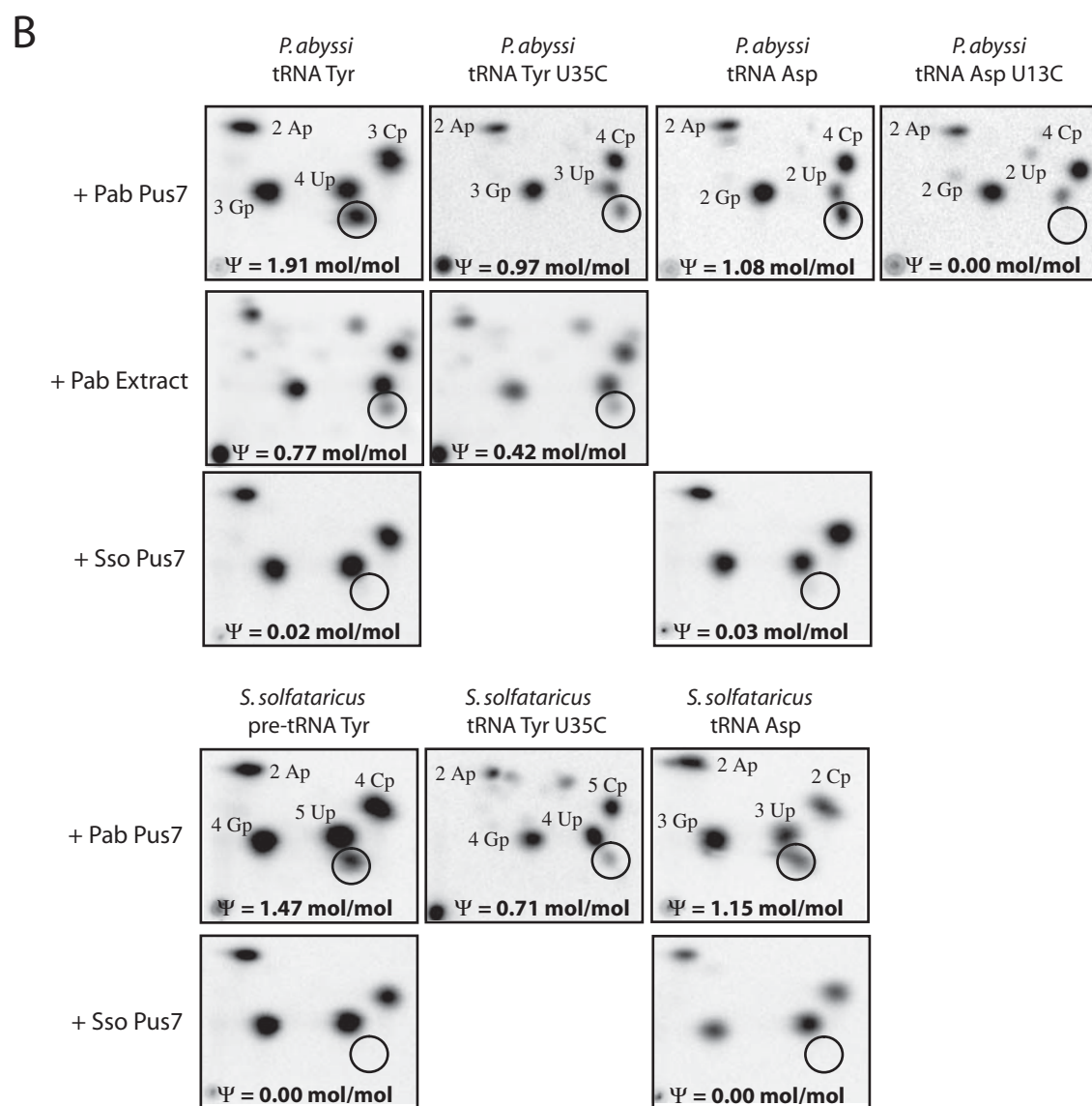
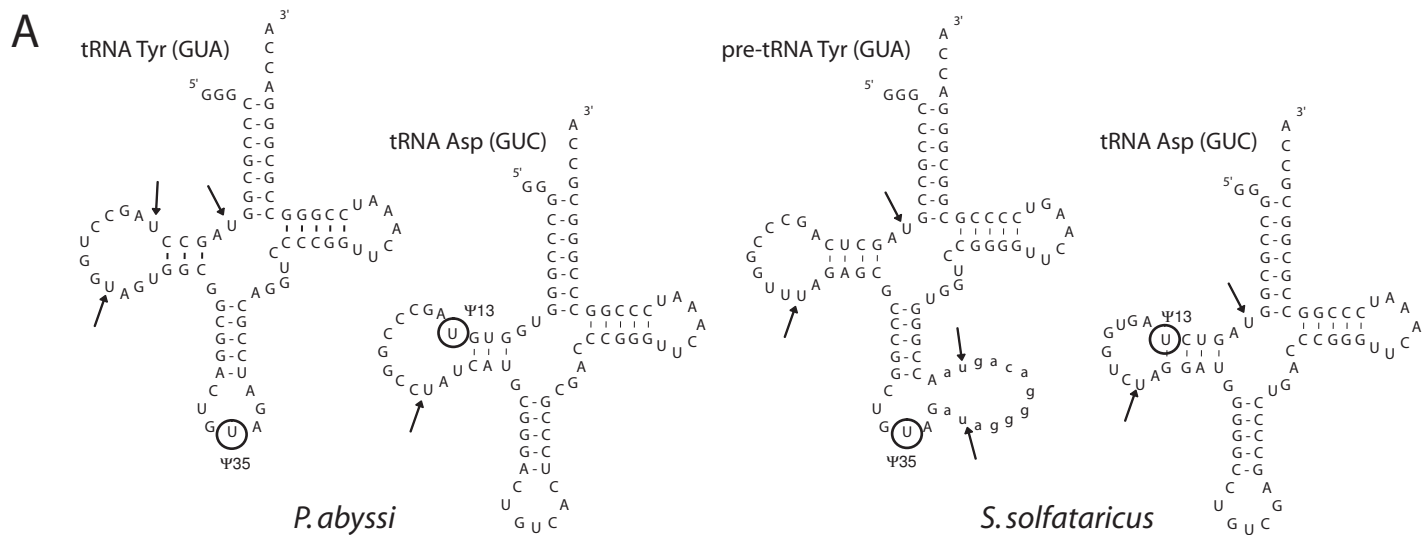
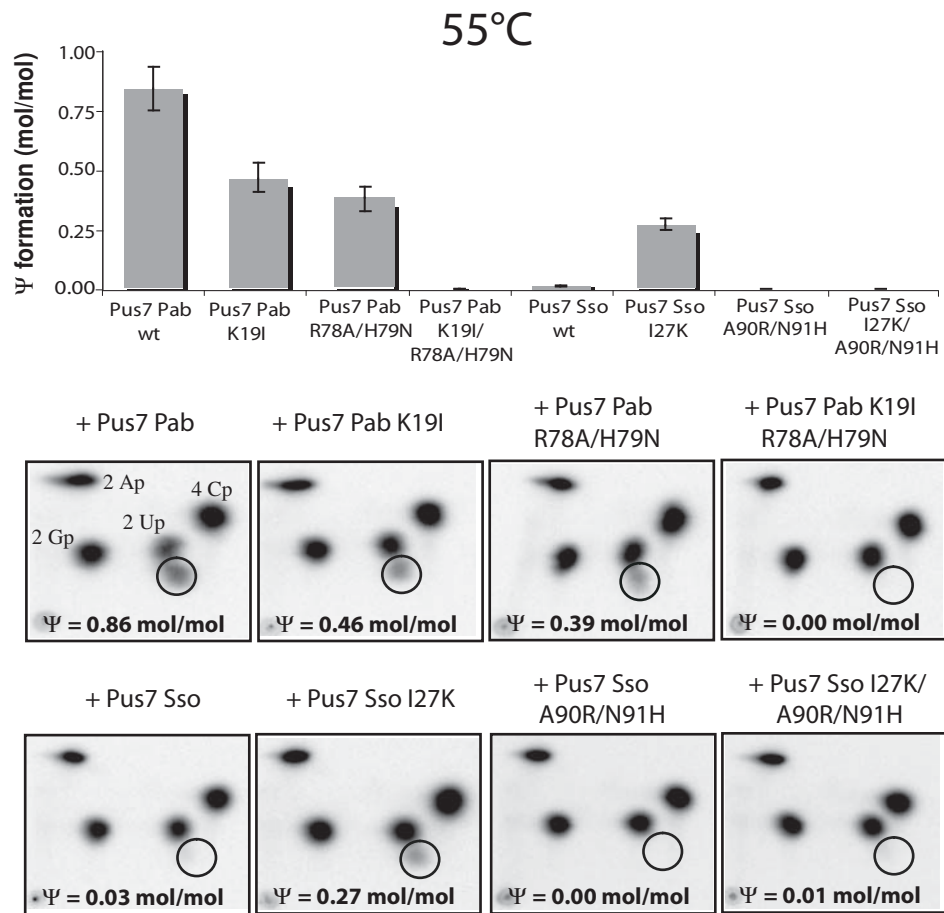


Figure 4

A



B

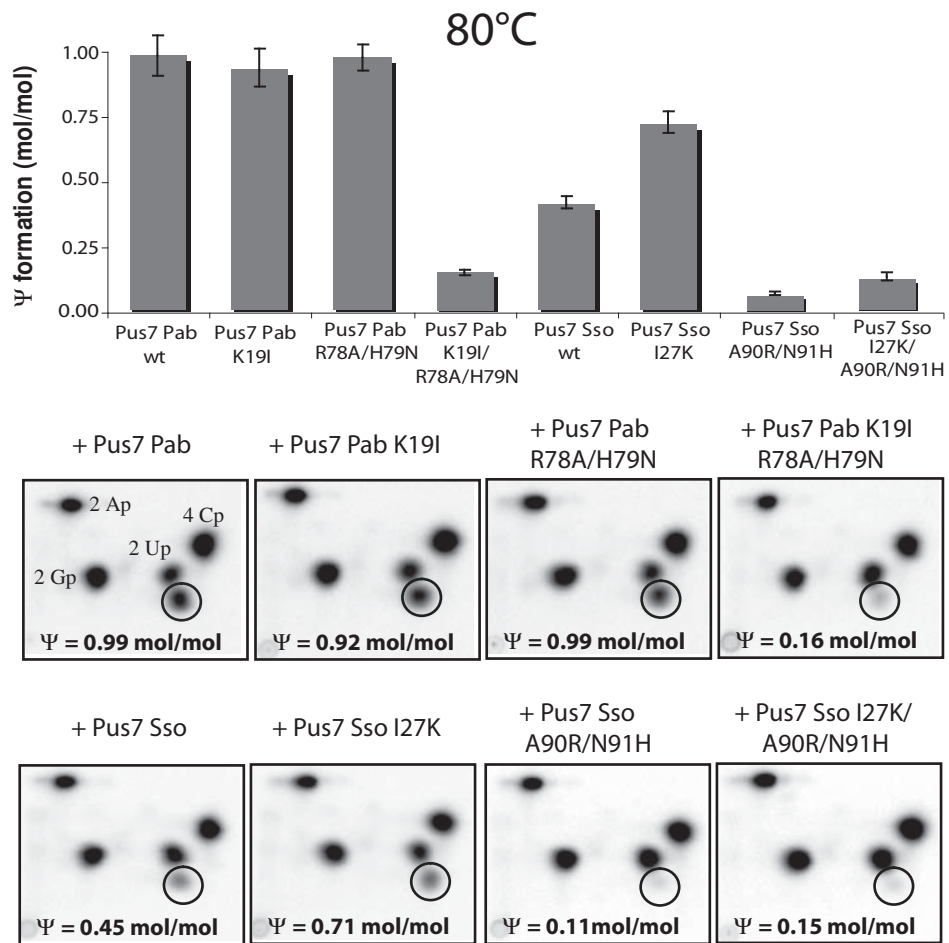


Figure 5

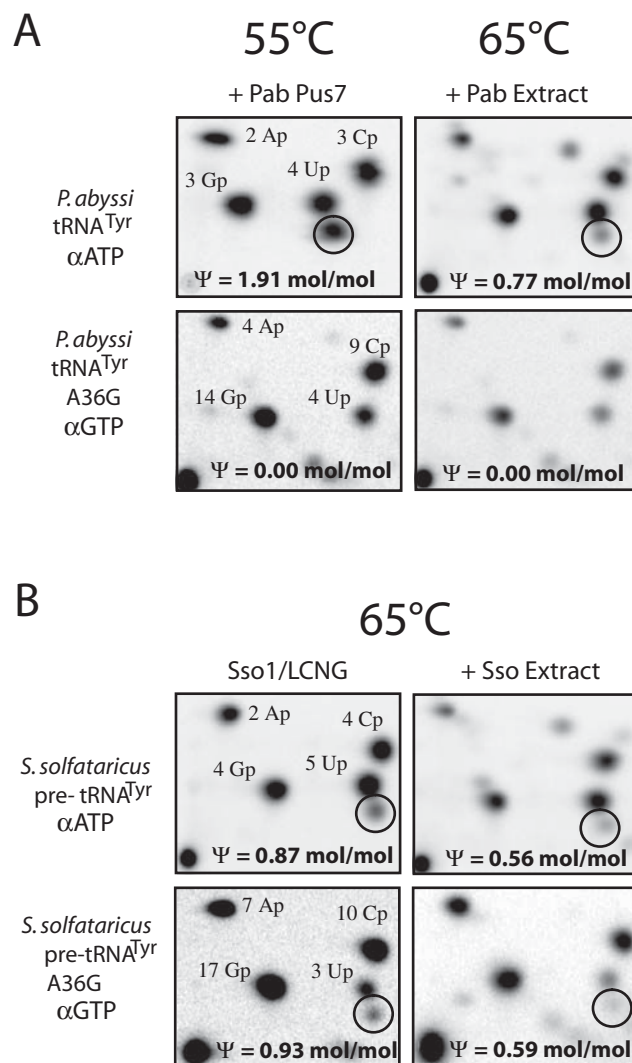
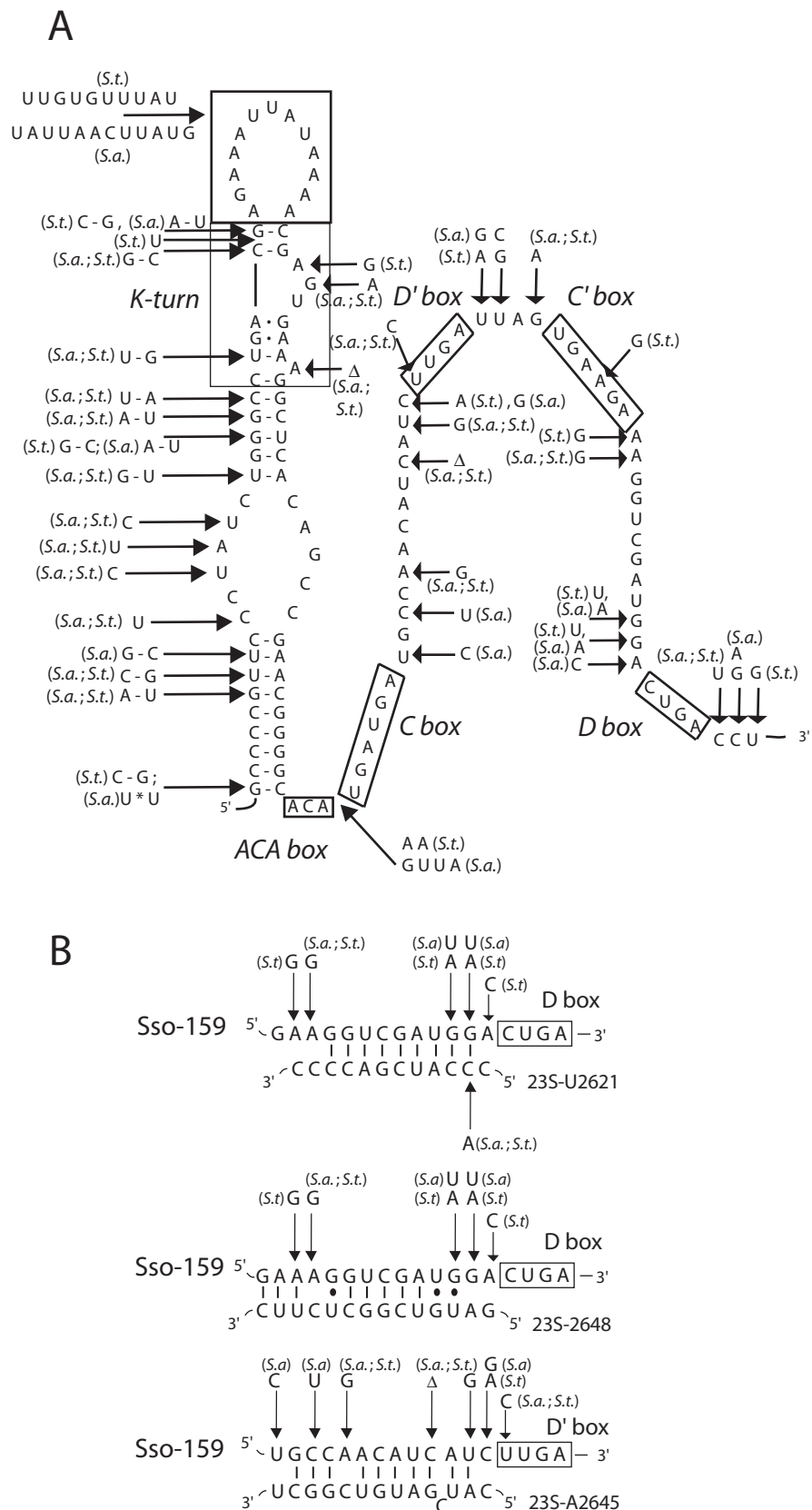


Figure 6



8. Identification de gènes potentiels de sRNA à boîtes H/ACA dans l'ensemble des séquences de génomes d'Archaea disponibles

A. Contexte & Intérêts

Jusqu'à présent, les quelques études publiées et nos études décrites précédemment avaient permis l'identification de 3, 7, 7, 7, 5, 7 et 1 gènes de sRNA H/ACA dans, respectivement les génomes de *A. fulgidus*, *P. furiosus*, *P. abyssi*, *P. horikoshii*, *M. jannaschii*, *T. kodakarensis* et *S. solfataricus* (Tang *et al.*, 2002a ; Rozhdestvensky *et al.*, 2003 ; Baker *et al.*, 2005 ; Charpentier *et al.*, 2005 ; Thebault *et al.*, 2006 ; Muller *et al.*, 2007b ; Zago *et al.*, 2005). La méthode bioinformatique que nous avons développée permettait une recherche plus spécifique et plus rapide de gènes potentiels de motifs H/ACA dans l'ensemble des génomes d'archaea séquencés disponibles. Une telle recherche devrait nous permettre de mieux définir les caractéristiques communes à tous les sRNA H/ACA d'archaea, de mieux comprendre leur modes d'appariement avec leur(s) cible(s), de savoir si leur présence est ou n'est pas uniforme dans les différentes espèces d'archaea et en particulier si leur nombre peut avoir un lien direct avec la température de croissance et enfin d'étudier l'évolution des gènes de sRNA H/ACA au sein du domaine des archaea.

B. Démarche

Nous avons appliqué l'approche présentée dans le chapitre d'ouvrage N°2 aux 46 séquences génomiques d'archaea disponibles. Nous avons comparé entre eux les résultats obtenus pour chaque espèce et nous avons essayé de mettre en place une nomenclature et de dresser un bilan de la distribution des sRNA H/ACA dans les espèces d'archaea. Certains sRNA présentant des propriétés particulières ont été analysés expérimentalement grâce à la technique de reconstitution de particules actives *in vitro* présentée dans la publication N°1.

C. Résultats et Discussion

1. Identification, dans les 46 génomes disponibles, d'un total de 162 motifs qui peuvent être regroupés en 35 familles

Au stade où en étaient les études au début de cette partie du travail 37 gènes de sRNA H/ACA d'archaea avaient été identifiés par voie expérimentale (Tang *et al.*, 2002a) ou bioinformatique (Rozhdestvensky *et al.*, 2003 ; Baker *et al.*, 2005 ; Charpentier *et al.*, 2005 ; Zago *et al.*, 2005 ; Thebault *et al.*, 2006 ; Muller *et al.*, 2007a). Ces 37 gènes correspondent à 56

motifs H/ACA. Nous avons décrit dans les chapitres précédents des applications de l'approche bioinformatique basée sur l'emploi du logiciel ERPIN que nous avons mise au point. Ici, nous décrivons les résultats obtenus par son application à l'ensemble des 46 génomes d'archaea séquencés et disponibles. Au total, nous avons ainsi identifié 114 gènes de sRNA H/ACA correspondant à 162 motifs H/ACA. Les positions dont la pseudouridylation peut être guidée par chacun de ces motifs H/ACA sont données dans le tableau R.3. L'alignement des séquences des 162 motifs H/ACA est présenté en Annexes.

Nous avons ainsi détecté l'existence de gènes de sRNA H/ACA dans l'ensemble des génomes de crénotes, à l'exception de celui de *C. symbiosium*, ainsi que dans l'ensemble des génomes d'euryotes à l'exception de 4 espèces de Methanosarcinales et une espèce de l'ordre des Thermoplasmatales (*F. acidarmanus*). Dans les espèces où des sRNA H/ACA potentiels sont trouvés le nombre de motifs H/ACA varie entre 1 et 17. Le génome de *M. kandleri* est celui qui comporte le plus de motifs H/ACA potentiels : 17 correspondant à 13 gènes de sRNA H/ACA. Il est à noter que nous avons confirmé par Northern Blot l'expression de 5 de ces 13 sRNA H/ACA à partir d'ARN extrait de cellules de cet archaea (Mka1, Mka7, Mka8, Mka15 et Mka16 ; résultats non montrés). La mise en évidence de l'expression des autres motifs H/ACA identifiés est actuellement poursuivie.

Pour chacun des motifs H/ACA identifiés dans les 46 espèces d'archaea, nous avons été capables d'identifier au moins une cible dans les ARNr ou dans les ARNt. Comme décrit dans la publication N°2, nous avons démontré que le motif Pab105-2 de *P. abyssi* est impliqué dans la modification de trois sites différents dans les ARNr de *P. abyssi*. De ce fait, nous avons pris en considération la possibilité qu'il en soit de même pour les autres motifs H/ACA mis en évidence dans cette partie du travail. En effet, pour certains des candidats détectés nous avons parfois proposé deux, voire trois cibles potentielles différentes. Le tableau R.4 indique les cibles potentielles de chacun des motifs H/ACA. La numérotation des positions au sein des ARNr 23S et 16S est celle des résidus correspondant dans l'ARNr de *P. abyssi*. En prenant en compte ainsi une des espèces d'archaea comme référence, nous pouvions comparer les cibles des sRNA H/ACA découverts dans les différentes espèces. Ceci nous a permis de regrouper les différents motifs H/ACA considérés comme homologues sur la base du fait qu'ils ciblent une, deux ou trois positions identiques. Par exemple, les boucles internes des motifs Afu4-1 (le premier motif en 5'de Afu4), Pab40-1 et Mthe1-1 présentent une séquence qui leur permet de cibler les positions 2542, 2588, et 2508 de leurs ARNr 23S respectifs. L'alignement des séquences des ARNr 23S de ces 3 espèces révèle qu'il s'agit de la même position relative, à savoir la position 2588 dans l'ARNr 23S de *P. abyssi* que nous avons utilisé comme référence. Ces trois motifs ont donc été regroupés dans la famille M1. De cette manière nous avons pu classer tous les motifs H/ACA identifiés en 35 familles appelées M1 à M35.

Ordre	Genre / esp�ce	sRNA H/ACA	Positionn� dans IRC ?	Position d�but	Position de fin	Sens
Archaeoglobales	<i>A. fulgidus</i>	Afu4-1	Y	2119145	2119214	FW
		Afu4-2	Y	2119215	2119287	FW
		Afu4-3	N	2119288	2119356	FW
		Afu46	Y	413565	413622	FW
		Afu190	Y	220968	221023	FW
Thermococcales	<i>P. abyssi</i>	Pab19	Y	204831	204911	FW
		Pab21	Y	230574	230517	RC
		Pab35-1	Y	318260	318329	FW
		Pab35-2	Y	318330	318387	FW
		Pab40-1	Y	382389	382446	FW
		Pab40-2	Y	382447	382528	FW
		Pab40-3	Y	382529	382599	FW
		Pab91	Y	1074568	1074512	RC
		Pab105-1	Y	1335797	1335724	RC
		Pab105-2	Y	1335722	1335648	RC
		Pab160	N	289063	288992	RC
	<i>P. furiosus</i>	Pf1	Y	163991	163933	RC
		Pf3-1	Y	942696	942622	RC
		Pf3-2	Y	942621	942544	RC
		Pf6-1	Y	1666344	1666412	FW
		Pf6-2	Y	1666413	1666470	FW
		Pf7-1	Y	1732713	1732772	FW
		Pf7-2	Y	1732773	1732854	FW
		Pf7-3	Y	1732855	1732926	FW
		Pf9	Y	1865160	1865095	RC
		Pfu19	Y	136711	136790	FW
		Pfu91	Y	1226103	1226159	FW
	<i>P. horikoshii</i>	Pho19	Y	190759	190839	FW
		Pho21	Y	215833	215776	RC
		Pho35-1	Y	1531443	1531511	FW
		Pho35-2	Y	1531512	1531569	FW
		Pho40-1	Y	1597422	1597481	FW
		Pho40-2	Y	1597482	1597563	FW
		Pho40-3	Y	1597564	1597634	FW
		Pho91	N	1007477	1007533	FW
		Pho105-1	Y	592081	592154	FW
		Pho105-2	Y	592155	592230	FW
		Pho160	Y	1647846	1647918	FW
	<i>T. gammatolerans</i>	Tga1-1		542215	542586	FW
		Tga1-2		542287	542356	FW
		Tga2		551052	551108	FW
		Tga3-1		945880	945949	FW
		Tga3-2		945950	946007	FW
		Tga4		1107568	1107636	FW
		Tga5-1		1396663	1396717	FW
		Tga5-2		1396718	1396798	FW
		Tga5-3		1396805	1396876	FW
		Tga19		1950591	1950511	RC
	<i>T. kodakaraensis</i>	Tko1-1	Y	462059	462114	FW
		Tko1-2	Y	462115	462195	FW
		Tko1-3	Y	462201	462272	FW
		Tko4	Y	899083	899151	FW
		Tko6	Y	1545183	1545239	FW
		Tko8	Y	47907	47849	RC
		Tko9-1	Y	465197	465127	RC
		Tko9-2	Y	465125	465056	RC
		Tko10-1	Y	1107400	1107332	RC
		Tko10-2	Y	1107331	1107274	RC
		Tko19	Y	2022646	2022727	FW

Ordre	Genre / espèce	sRNA H/ACA	Positionné dans IRC ?	Position début	Position de fin	Sens
Methanococcales	<i>M. jannashii</i>	Mja1	Y	216281	216350	FW
		Mja5	Y	986084	986151	FW
		Mja4	Y	1659519	1659450	RC
		Mja3-1	Y	537593	537665	FW
		Mja3-2	Y	537671	537744	FW
		Mja2	N	118134	118060	RC
	<i>M. maripaludis</i> S2 chrom. I	Mmar1-1	Y	226214	226132	RC
		Mmar1-2	Y	226121	226051	RC
		Mmar1-3	Y	226047	225975	RC
		Mmar2	Y	609894	609822	RC
	<i>M. vannielii</i>	Mva1-1	Y	1264869	1264798	RC
		Mva1-2	Y	1264787	1264717	RC
		Mva1-3	Y	1264714	1264640	RC
		Mva2	Y	1510036	1509979	RC
	<i>M. aeolicus</i>	Mae1-1	Y	356199	356272	FW
		Mae1-2	Y	356284	356355	FW
		Mae2	Y	558629	558559	RC
		Mae3	Y	623830	623902	FW
Methanopyrales	<i>M. kandleri</i>	Mka1	Y	259838	259910	FW
		Mka7-1	Y	1618775	1618845	FW
		Mka7-2	Y	1618856	1618929	FW
		Mka8-1	Y	576670	576592	RC
		Mka8-2	Y	576589	576516	RC
		Mka8-3	Y	576503	576421	RC
		Mka9	Y	720506	720437	RC
		Mka11	N	789429	789359	RC
		Mka15	N	1327111	1327050	RC
		Mka16	Y	1419184	1419106	RC
		Mka17	Y	879815	879895	FW
		Mka18	Y	19803	19726	RC
		Mka19-1	Y	262797	262723	RC
		Mka19-2	Y	262713	262628	RC
		Mka20	Y	1105352	1105429	FW
		Mka22	Y	1219408	1219471	FW
		Mka24	Y	1161531	1161606	FW
Thermoplasmatales	<i>T. acidophilum</i>	Tac7-1	Y	1394487	1394433	RC
		Tac7-2	Y	1394430	1394354	RC
		Tac7-3	Y	1394351	1394296	RC
	<i>T. volcanium</i>	Tvo3-1	Y	267938	267883	RC
		Tvo3-2	Y	267880	267805	RC
		Tvo3-3	Y	267801	267746	RC
	<i>P. torridus</i>	Pto1-1	Y	276027	275971	RC
		Pto1-2	Y	275969	275889	RC
		Pto1-3	Y	275887	275832	RC
Nanoarchaeales	<i>N. equitans</i>	Neq1-1	Y	339421	339494	FW
		Neq1-2	Y	339495	339566	FW
		Neq3	Y	427540	427599	FW
		Neq4	Y	432133	432188	FW

Tableau R.3 : Positions des motifs H/ACA détectés dans les génomes des espèces d'archaea analysées. Nous avons indiqué si les motifs H/ACA sont positionnés dans une séquence IRC (Y) ou chevauchent une ORF (N). Les positions de début et de fin du motif H/ACA considéré sont indiquées avec le sens du brin d'ADN considéré (FW, Forward ou brin + ; RC, Reverse Complement ou brin -). Les cases grisées correspondent à des motifs H/ACA qui ont aussi été identifiés par d'autres équipes.

Ordre	Genre / espèce	sRNA H/ACA	Positionné dans IRC ?	Position début	Position de fin	Sens
Halobacteriales	<i>H. sp1</i>	Hsp2	Y	1068407	1068478	FW
		Hsp3	Y	1068586	1068662	FW
		Hsp4	Y	1874591	1874660	FW
	<i>H. marismortui S2 Chrom. I</i>	Hmar1	Y	317126	317196	RC
		Hmar2	Y	316907	316983	RC
	<i>H. walsbyi</i>	Hwal1	Y	2216529	2216602	FW
		Hwal2	Y	2216633	2216710	FW
		Hwal3	Y	66298	66227	RC
	<i>N. pharaonis</i>	Nph1	Y	1461831	1461904	FW
		Nph2	Y	1461926	1462003	FW
		Nph3	Y	216841	216768	RC
Methanosarcinales	<i>M. thermophila</i>	Mthe1-1	Y	1596012	1595941	RC
		Mthe1-2	Y	1595939	1595864	RC
		Mthe1-3	Y	1595860	1595789	RC
Methanomicrobiales	<i>M. marisnigri</i>	Mcm1-1	Y	1003173	1003096	RC
		Mcm1-2	Y	1003091	1003015	RC
	<i>M. boonei</i>	Cme1-1	Y	826269	826194	RC
		Cme1-2	Y	829188	826109	RC
	<i>M. labreanum</i>	Mla1-1	Y	341761	341688	RC
		Mla1-2	Y	341682	341609	RC
	<i>M. hungatei</i>	Mhu1-1	Y	617845	617772	RC
		Mhu1-2	Y	617761	617683	RC
Methanobacteriales	<i>M. smithii</i>	Msm1-1	Y	515980	515907	RC
		Msm1-2	Y	515904	515830	RC
	<i>M. thermoauto.</i>	Mth1-1	Y	1607494	1607563	FW
		Mth1-2	Y	1607567	1607641	FW
	<i>M. stadtmanae</i>	Mst1	Y	415042	415117	FW
Desulfurococcales	<i>A. pernix</i>	Ape2	Y	1242202	1242274	FW
		Ape6	Y	1490689	1490613	RC
		Ape7	Y	1237087	1237162	FW
		Ape8	Y	963567	963646	FW
		Ape12	Y	1221285	1221385	FW
	<i>I. hospitalis</i>	Isp1	Y	374747	374814	FW
		Isp2	Y	204696	204626	RC
		Isp3	Y	1066758	1066829	FW
		Isp4	Y	181749	181831	FW
	<i>H. butylicus</i>	Hbu1	Y	73144	73075	RC
		Hbu2	Y	890911	890839	RC
	<i>S. marinus</i>	Sma1	Y	406641	406564	RC
		Sma2	Y	875287	875212	RC
		Sma3	Y	1398888	1398827	RC
		Sma4	Y	1116936	1117023	FW
Sulfolobales	<i>S. solfataricus</i>	Sso1	Y	2080193	2080123	RC
		Sso109	Y	595511	595586	FW
	<i>S. tokodaii</i>	Sto1	Y	2245765	2245834	FW
		Sto4	Y	419012	418942	RC
	<i>S. acidocaldarius</i>	Sac1	Y	130091	130023	RC
		Sac2	Y	458983	459058	FW
	<i>M. sedula</i>	Mse1	Y	99820	99891	FW
		Mse2	Y	1987235	1987303	FW
Thermoproteales	<i>P. aerophylum</i>	Pae1	Y	648914	648982	FW
	<i>P. arsenaticum</i>	Par1	Y	236307	236238	RC
	<i>P. calidifontis</i>	Pca1	Y	649112	649180	FW
	<i>P. islandicum</i>	Pis1	Y	1136984	1136915	RC
	<i>T. pendens</i>	Tpe1	Y	367928	367848	RC

Ordre	Genre / espèce	sRNA H/ACA	Cibles	Positions équivalentes chez <i>P. abyssi</i>	Famille
Archaeoglobales	<i>A. fulgidus</i>	Afu4	1st hp : 16S 1167 Ψ	16S 1179	M1
			1st hp : 23S 2542 Ψ	23S 2588 Ψ	
			2nd hp : 23S 2632 Ψ	23S 2678	M2
			2nd hp : 23S 2601 Ψ	23S 2647	
			3rd hp : 23S 1364	23S 1374	M3
			3rd hp : 23S 1970 Ψ	23S 2016	
		Afu46	23S 2639 Ψ	23S 2685 Ψ	M4
Thermococcales	<i>P. abyssi</i>	Afu190	16S 1004	16S 1017	M5
		Pab19	16S 1017		M5
		Pab21	16S 891 Ψ		M6
		Pab35	1st hp : 23S 2930 Ψ		M7
			1st hp : 23S 2672 Ψ		
			2nd hp : 23S 2549 Ψ		M8
			2nd hp : 23S 2697 Ψ		
		Pab40	1st hp : 23S 2588 Ψ		M1
			2nd hp : 23S 1932 Ψ		M9
			3rd hp : 23S 2016		M3
		Pab91	23S 2685 Ψ		M4
		Pab105	1st hp : 23S 2554 Ψ		M10
			2nd hp : 16S 27 Ψ		
			2nd hp : 23S 1377		M11
			2nd hp : 23S 2794 Ψ		
		Pab160	16S 922 Ψ		M12
	<i>P. furiosus</i>	Pf1	16S 879	16S 891 Ψ	M6
		Pf3	1st hp : 23S 2577	1st hp : 23S 2554 Ψ	M10
			2nd hp : 16S 15	2nd hp : 16S 27 Ψ	
			2nd hp : 23S 1400	2nd hp : 23S 1377	M11
		Pf6	2nd hp : 23S 2817	2nd hp : 23S 2794 Ψ	
			1st hp : 23S 2953	1st hp : 23S 2930 Ψ	M7
			1st hp : 23S 2695	1st hp : 23S 2672 Ψ	
			2nd hp : 23S 2572	2nd hp : 23S 2549 Ψ	M8
		Pf7	2nd hp : 23S 2720	2nd hp : 23S 2697 Ψ	
			1st hp : 23S 2611	1st hp : 23S 2588 Ψ	M1
			2nd hp : 2701	2nd hp : 2678	M2
		Pf9	3rd hp : 23S 2039	3rd hp : 23S 2016	M3
			16S 910	16S 922 Ψ	M12
		Pfu19	16S 1005	16S 1017	M5
		Pfu91	23S 2708	23S 2685 Ψ	M4
	<i>P. horikoshii</i>	Pho19	16S 1005	16S 1017	M5
		Pho21	16S 879	16S 891 Ψ	M6
		Pho35	1st hp : 23S 2925	1st hp : 23S 2930 Ψ	M7
			1st hp : 23S 2667	1st hp : 23S 2672 Ψ	
			2nd hp : 23S 2544	2nd hp : 23S 2549 Ψ	M8
			2nd hp : 23S 2692	2nd hp : 23S 2697 Ψ	
		Pho40	1st hp : 23S 2583	1st hp : 23S 2588 Ψ	M1
			2nd hp : 2673	2nd hp : 2678	M2
			3rd hp : 23S 2011	3rd hp : 23S 2016	M3
		Pho91	23S 2680	23S 2685 Ψ	M4
		Pho105	1st hp : 23S 2549	1st hp : 23S 2554 Ψ	M10
			2nd hp : 16S 15	2nd hp : 16S 27 Ψ	
			2nd hp : 23S 1373	2nd hp : 23S 1377	M11
	<i>T. gammatolerans</i>	Pho160	2nd hp : 23S 2789	2nd hp : 23S 2794 Ψ	
		Tga1	16S 910	16S 922 Ψ	M12
			1st hp : 23S 2546	23S 2554 Ψ	M10
			2nd hp : 16S 15	16S 27 Ψ	
			2nd hp : 16S 974	16S 995	M11
		Tga2	2nd hp : 23S 1372	23S 1377	
			23S 2677	23S 2685 Ψ	M4
		Tga3	1st hp : 23S 2664	23S 2672 Ψ	M7
			2nd hp : 23S 2689	23S 2697 Ψ	M8
		Tga4	16S 901	16S 922 Ψ	M12
		Tga5	1st hp : 2580	23S 2588 Ψ	M1
			2nd hp : 2670	23S 2678	M2
			3rd hp : 23S 2008	23S 2016	M3
		Tga19	16S 996	16S 1017	M5

Tableau R.4 : Les différentes positions des ARNr ou ARNt des résidus ciblées par les motifs H/ACA identifiés.

Les positions des résidus U ciblés sont indiquées avec la numérotation des ARNr 16S et 23S de l'espèce considérée. La position correspondante dans les ARNr 16S et 23S de *P. abyssi* est indiquée lorsqu'elle est identifiée. Elle est utilisée comme référence. La famille à laquelle appartient le motif H/ACA est également indiquée (M1 à M35). Les cases grisées correspondent aux motifs H/ACA identifiés ou aux prédictions de cibles réalisées par d'autres équipes. Les symboles Ψ indiqués en bleu ou en rouge correspondent respectivement aux résidus qui ont été détectés dans les ARNr de l'espèce considérée ou dans les ARNr de *P. abyssi*.

Tableau R.4
(suite)

Ordre	Genre / esp�ce	sRNA H/ACA	Cibles	Positions �quivalentes chez <i>P. abyssi</i>	Famille
Thermococcales	<i>T. kodakaraensis</i>	Tko1	1st hp: 2592	2588 Ψ	M1
			2nd hp: 2682	23S 2678 Ψ	M2
			3rd hp: 23S 2020	23S 2016	M3
		Tko4	16S 901	16S 922 Ψ	M12
		Tko6	23S 2689	23S 2685 Ψ	M4
		Tko8	16S 870	16S 891 Ψ	M6
		Tko9	1st hp: 23S 2558	23S 2554 Ψ	M10
			2nd hp: 16S 15	16S 27 Ψ	M11
			2nd hp: 16S 974	16S 995	
			2nd hp: 23S 1385	23S 1377	
		Tko10	1st hp: 23S 2676	23S 2672 Ψ	M7
			2nd hp: 23S 2701	23S 2697 Ψ	M8
		Tko19	16S 996	16S 1017	M5
Methanococcales	<i>M. jannashii</i>	Mja1	23S : 2015	23S 2016	M3
		Mja2	16S 584	16S 609	M16
		Mja3	1st hp : 23S 2595	23S 2596	M14
			2nd hp : 23S 2645	2646	M15
		Mja4	16S 899	16S 922 Ψ	M12
		Mja5	16S 507	16S 532	M13
	<i>M. maripaludis</i> S2	Mmar1	1st hp :23S 2562	23S 2596	M14
			2nd hp : 23S 2612	23S 2646	M15
			2nd hp : 23S 1048	23S 1077	
			3rd hp : 23S 1984	23S 2016	M3
	<i>M. vannielli</i>	Mva1	1st hp: 23S 2562	23S 2596	M14
			2nd hp: 23S 2612	23S 2646	M15
			3rd hp: 1984	23S 2016	M3
		Mva2	16S 571	16S 609	M16
	<i>M. aeolicus</i>	Mae1	1st hp: 23S 2446	23S 2596	M14
			2nd hp: 23S 1867	23S 2016	M3
			23S 1867	23S 2016	M3
			16S 547	16S 609	M16
Methanopyrales	<i>M. kandleri</i>	Mka1	16S 707	-	M17
		Mka7	1st hp : 16S 12	16S 24	M18
			2nd hp : 16S 730	16S 724	M19
		Mka8	1st hp : 23S 2703	23S 2596	M14
			2nd hp : 23S 2785	23S 2678	M2
			3rd hp : 23S 2111	23S 2016	M3
		Mka9	16S 8	16S 19	M20
		Mka11	23S 2718	23S 2611	M21
		Mka15	23S 2804	23S 2697 Ψ	M8
		Mka16	16S 928	16S 922 Ψ	M12
		Mka17	23S 2389	-	M22
		Mka18	23S 2825	23S 2718	M23
		Mka19	1st hp: 16S 1486	16S 1481	M24
			2nd hp: 16S 1356	16S 1350	M25
		Mka20	16S 625	16S 620	M29
		Mka22	23S 2710	23S 2603 Ψ	M30
		Mka24	tRNA Glu 25	-	M31
Thermoplasmatales	<i>T. acidophilum</i>	Tac7	1st hp: 23S 2528	23S 2603 Ψ	M30
			2nd hp : 23S 2597	23S 2672 Ψ	M7
			3rd hp : 23S 1941	23S 2016	M3
	<i>T. volcanium</i>	Tvo3	1st hp: 23S 2529	23S 2603 Ψ	M31
			2nd hp : 23S 2597	23S 2672 Ψ	M7
			3rd hp :23S 1942	23S 2016	M3
	<i>P. torridus</i>	Pto1	1st hp: 23S 2525	23S 2603 Ψ	M31
			2nd hp : 23S 2594	23S 2672 Ψ	M7
			3rd hp :23S 1937	23S 2016	M3
Nanoarchaeales	<i>N. equitans</i>	Neq1	1st hp : 23S 1930	23S 1932 Ψ	M9
			1st hp : 23S 851	23S 854	
			2nd hp : 23S 2014	23S 2016	M3
		Neq3	16S 1035	16S 1017	M5
		Neq4	16S 956	16S 938	M32

Tableau
R.4
(suite)

Ordre	Genre / espèce	sRNA H/ACA	Cibles	Positions équivalentes chez <i>P. abyssi</i>	Famille
Halobacteriales	<i>H. sp NRC1</i>	Hsp2	23S 2605 Ψ	23S 2678 Ψ	M2
		Hsp3	23S1940 Ψ	23S 2016	M3
			23S 1942 Ψ	23S 2018	
		Hsp4	23S 2590 Ψ	-	M33
	<i>H. marismortui</i>	Hmar1	23S 2619 Ψ	23S 2678	M2
		Hmar2	23S 1956 Ψ	23S 2016	M3
			23S 1958 Ψ	23S 2018	
	<i>H. walsbyi</i>	Hwal1	23S 2627	23S 2678 Ψ	M2
		Hwal2	23S 2607	23S 2016	M3
			23S 2609	23S 2018	
		Hwal3	2613	23S 2590 Hsp	M33
	<i>N. pharaonis</i>	Nph1	23S 2607	23S 2678 Ψ	M2
		Nph2	23S1942	23S 2016	M3
			23S 1944	23S 2018	
		Nph3	2593	23S 2590 Hsp	M33
Methanosarcinales	<i>M. thermophila</i>	Mthe1	1st hp: 23S 2508	23S 2588 Ψ	M1
			2nd hp: 23S 2598	23S 2678	M2
			3rd hp: 23S 1937	23S 2016	M3
Methanomicrobiales	<i>M. marisnigri</i>	Mcm1	1st hp: 23S 2220	23S 2286	M34
			1st hp: 23S 2611	23S 2678	M2
			2nd hp: 23S 1949	23S 2016	M3
			2nd hp: 23S 1951	23S 2018	
			2nd hp: 23S 2235	23S 2301	
			1st hp: 23S 2610	23S 2678	M2
	<i>M. boonei</i>	Cme1	2nd hp: 1948	23S 2016	M3
			2nd hp: 1950	23S 2018	
			1st hp: 23S 2576	23S 2647	M2
	<i>M. labreanum</i>	Mla1	2nd hp: 23S 1939	23S 2016	M3
Methanobacteriales	<i>M. smithii</i>	Msm1	1st hp: 23S 2538/2530	23S 2678	M2
			2nd hp: 23S 1873/1865	23S 2016	M3
	<i>M. thermoauto.</i>	Mth1	1st hp: 23S 2720/2726	23S 2678	M2
			2nd hp: 23S 2051/2056	23S 2016	M3
	<i>M. stadtmanae</i>	Mst1	23S 2006	23S 2016	M3
Desulfurococcales	<i>A. pernix</i>	Ape2	23S 1808	23S 2057	M26
			23S 2541	23S 2782	
		Ape6	23S 1773	23S 2022	M27
		Ape7	23S 2444	23S 2685 Ψ	M4
			Pre-tRNA ^{Tyr} 35	-	
		Ape8	23S 1685	23S 1934	M28
		Ape12	23S 2308	23S 2549 Ψ	M8
	<i>I. hospitalis</i>	Isp1	23S 2737	23S 2685 Ψ	M4
		Isp2	23S 2067	23S 2022	M27
		Isp3	23S 2061	23S 2016	M3
		Isp4	23S 2100	23S 2056	M26
	<i>H. butylicus</i>	Hbu1	23S 2779	23S 2685 Ψ	M4
		Hbu2	23S 2108	23S 2016	M3
	<i>S. marinus</i>	Sma1	23S 2857	23S 2685 Ψ	M4
		Sma2	23S 2129	23S 2022	M27
		Sma3	23S 1770	-	M35
			23S 827	750	
		Sma4	16S 1073	16S 1017	M5
Sulfolobales	<i>S. solfataricus</i>	Sso1	Pre-tRNA ^{Tyr} 35	-	M4
		Sso109	23S 2598 Ψ	23S 2549 Ψ	M8
	<i>S. tokodaii</i>	Sto1	Pre-tRNA ^{Tyr} 35	-	M4
		Sto4	23S 2596	23S 2549 Ψ	M8
	<i>S. acidocaldarius</i>	Sac1	Pre-tRNA ^{Tyr} 35	-	M4
		Sac2	23S 2561	23S 2549 Ψ	M8
	<i>M. sedula</i>	Mse1	23S 2590	23S 2549 Ψ	M8
		Mse2	Pre-tRNA ^{Tyr} 35	-	M4

Ordre	Genre / esp�ce	sRNA H/ACA	Cibles	Positions �quivalentes chez <i>P. abyssi</i>	Famille
Thermoproteales	<i>P. aerophilum</i>	Pae1	16S 1723	16S 1017	M5
	<i>P. arsenaticum</i>	Par1	16S 1722	16S 1017	M5
	<i>P. caldifontis</i>	Pca1	16S 1008	16S 1017	M5
	<i>P. islandicum</i>	Pis1	16S 1008	16S 1017	M5
	<i>T. pendens</i>	Tpe1	tRNA thr 47	-	M32
			16S 931	16S 938	

Tableau R.4 (suite)

On remarque que certaines familles comportent de nombreux repr sentants comme c'est le cas de la famille M3 ayant un membre dans tous les g nomes d'euryotes qui comportent des sRNA H/ACA potentiels, ainsi que dans quelques g nomes de cr notes (Tableau R.5). Les familles M4, M5 et M8 sont elles aussi tr s repr sent es   la fois chez certains euryotes et chez certains cr notes. D'autres familles ne comptent qu'un seul repr sentant (14 parmi les 35 familles), c'est le cas des familles correspondant aux nombreux motifs H/ACA d tect s dans *M. kandleri* et pour lesquels aucun  quivalent n'a  t  trouv  dans une autre esp ce. Cette situation pourrait s'expliquer par la position isol e de *M. kandleri* dans l'arbre phylog n tique du Domaine Archaea (Fig. I.4). Notons  galement qu'une seule esp ce (*M. aeolicus*) comporte deux repr sentants d'une m me famille (M3). Les deux motifs H/ACA appel s Mae1-2 et Mae2 ne correspondent pas   une r p tition de s quence puisque d'une part le motif Mae1-2 est associ    un autre motif et d'autre part les s quences Mae1-2 et Mae2 comportent 16 nt diff rents sur un total de 72 nt. Cette redondance de repr sentants d'une m me famille est un cas particulier. En effet, une esp ce donn e ne contient en g n ral qu'un seul membre d'une famille donn e.

2. Localisation des g nes de sRNA H/ACA dans les g nomes d'archaea

L'analyse des positions des g nes potentiels de motifs H/ACA au sein des g nomes r v le qu'ils sont, dans la plupart des cas, localis s dans les r gions IRC (Tableau R.3). Seul un faible nombre d'entre eux chevauche l g rement les ORF adjacentes. C'est le cas notamment des motifs Afu4-3, de Pab160, Pho91 ou Mja2. Il est int ressant de noter que les g nes de sRNA H/ACA d tect s sont souvent localis s entre des ORF dans des r gions qui comportent en plus des g nes d'ARNt et de sRNA C/D. Hormis leur association avec des sRNA C/D d j  d crite dans la publication N 4 et discut e ci-dessous, il ne semble pas que les sRNA H/ACA soient co-transcrits avec d'autres types d'ARNnc. Par ailleurs, on remarque souvent la pr sence, en 5' des sRNA H/ACA identifi s, de r gions riches en r sidus A/U qui pourraient constituer des promoteurs transcriptionnels ind pendants. Comme  voqu  en introduction, les donn es actuelles ne permettent pas de d finir les s quences et les positions des promoteurs d'archaea et il n'est donc pas possible de rep rer ces promoteurs en amont des g nes de sRNA H/ACA.

Euryotes

Familles de motifs H/A	Positions des cibles (Numération de P. abyssi)	Archaeoglobus fulgidus	Pyrococcus abyssi	Pyrococcus furiosus	Pyrococcus horikoshii	Thermococcus kodakarensis	Methanocaldococcus jannaschii	Methanocaldococcus marisnigellus	Methanocaldococcus aequus	Methanocaldococcus thermophilus	Methanopyrus kandleri	Halobacterium equitans	Nitrospira sp.	Halobacterium walsbyi	Halobacterium marisnigellus	Natrononas pharaonis	Methanosarcina thermophila	Methanospirillum hungatei	Methanococcus marisnigellus	Methanopropinquum	Methanoregula boonei
M1	LSU 2588	Afu4-1	Pf1-1	Pho40-1	Tko1-1	Tga5-1						Hsp2	Hwa1	Hmar1	Nph1	Mthe1-1					
M2	LSU 2678 / LSU 2647	Afu4-2	Pf1-2	Pho40-2	Tko1-2	Tga5-2										Mthe1-2	Mhu1-1			Mla1-1	Cne1-1
M3	LSU 2016/2018	Afu4-3	Pf1-3	Pho40-3	Tko1-3	Tga5-3						Hsp3	Hwa2	Hmar2	Nph2	Mthe1-3	Mhu1-2	Mem1-2	Mla1-2	Cne1-2	
M4	LSU 2685 / pre-rRNA tr 35	Afu46	Pfu91	Pho91	Tko6	Tga2															
M5	SSU 1017	Afu190	Pfu19	Pho19	Tko19	Tga19															
M6	SSU 891		Pf1	Pho21	Tko8																
M7	LSU 2672 / LSU 2930		Pf6-1	Pho35-1	Tko10-1	Tga3-1															
M8	LSU 2697 / LSU 2549		Pf6-2	Pho35-2	Tko10-2	Tga3-2															
M9	LSU 1932																				
M10	LSU 2554		Pf13-1	Pho106-1	Tko9-1	Tga1-1															
M11	SSU 27 / LSU 1377 / LSU 2794		Pf13-2	Pho106-2	Tko9-2	Tga1-2															
M12	SSU 922		Pf9	Pho160	Tko4	Tga4															
M13	SSU 532																				
M14	LSU 2596																				
M15	LSU 2646 / LSU 1077																				
M16	SSU 609																				
M17	SSU 707 Mka																				
M18	SSU 24																				
M19	SSU 724																				
M20	SSU 19																				
M21	LSU 2611																				
M22	LSU 2389 Mka																				
M23	LSU 2718																				
M24	SSU 1481																				
M25	SSU 1350																				
M26	LSU 2057 / LSU 2782																				
M27	LSU 2022																				
M28	LSU 1934																				
M29	SSU 625 Mka																				
M30	LSU 2603																				
M31	tRNA Glu 25																				
M32	SSU 938																				
M33	LSU 2590 Hsp																				
M34	LSU 2286																				
M35	LSU 1770 Sma / LSU 750																				

		Euryotes										Crénotes									
Familles de motifs H/ACA	Positions des cibles (Numération de H/ACA)	Methanospiraea										Methanobrevibacter									
		Methanospiraea					Methanobrevibacter					Methanobrevibacter					Methanobrevibacter				
		Mth1-1	Mth1-2	Mst1	Mth1-1	Msm1-1	Tac7-2	Tac7-3	Tac7-1	Tac7-2	Tac7-3	Tac7-1	Tac7-2	Tac7-3	Tac7-1	Tac7-2	Tac7-3	Tac7-1	Tac7-2	Tac7-3	Tac7-1
M1	LSU 2588																				
M2	LSU 2678 / LSU 2647																				
M3	LSU 2016/2018																				
M4	LSU 2685 / pre-rRNA tr 35																				
M5	SSU 1017																				
M6	SSU 891																				
M7	LSU 2672 / LSU 2530																				
M8	LSU 2697 / LSU 2549																				
M9	LSU 1932																				
M10	LSU 2554																				
M11	SSU 27 / LSU 1377 / LSU 2794																				
M12	SSU 922																				
M13	SSU 532																				
M14	LSU 2596																				
M15	LSU 2646 / LSU 1077																				
M16	SSU 609																				
M17	SSU 707 Mka																				
M18	SSU 24																				
M19	SSU 724																				
M20	SSU 19																				
M21	LSU 2611																				
M22	LSU 2389 Mka																				
M23	LSU 2718																				
M24	SSU 1481																				
M25	SSU 1350																				
M26	LSU 2057 / LSU 2782																				
M27	LSU 2022																				
M28	LSU 1934																				
M29	SSU 625 Mka																				
M30	LSU 2603																				
M31	rRNA Glu 25																				
M32	SSU 938																				
M33	LSU 2590 Hsp																				
M34	LSU 2286																				
M35	LSU 1770 Sma / LSU750																				

Tableau R.5 : Regroupement en 35 familles des différents motifs H/ACA mis en évidence dans les espèces dont la séquence génomique est disponible.

Une famille regroupe tous les motifs H/ACA qui ciblent un ou plusieurs résidus Ψ ayant des positions similaires dans les ARNr 16S et 23S. Un trait rouge indique la séparation entre euryotes et crénote.

La localisation du sRNA H/ACA Pab19 est intéressante. En effet, son gène est localisé une trentaine de nucléotides en amont de la région 5'ETS de l'ARNr 16S. Du fait de cette localisation particulière nous avons d'ailleurs initialement écarté ce candidat de la liste des motifs H/ACA potentiels de *P. abyssi*. En effet, nous ne savions pas si la conservation de sa séquence observée était due à la présence d'un sRNA H/ACA ou à la conservation de séquence de la région 5'ETS qui, chez les archaea, joue un rôle essentiel pour la formation de motifs Bulge-Helix-Bulge reconnus par la machinerie de maturation des ARNr. Il est également intéressant de noter que le gène du sRNA Pab21 est localisé quelques centaines de nucléotides en aval de la région 3'ETS du transcrit des ARNr. L'examen des autres génomes d'archaea montre également un lien possible entre la localisation de l'opéron des ARNr et la localisation de certains gènes de sRNA H/ACA. A l'exception de ce lien possible, nous n'avons pas détecté de lien conservé au niveau de la localisation relative dans les génomes des gènes des sRNA H/ACA et de gènes de protéines ou d'autres ARNnc. L'association de sRNA H/ACA avec des sRNA C/D du côté 3' (7 cas) ou 5' (1 seul cas) reste exceptionnelle puisqu'elle ne concerne que 8 motifs H/ACA.

3. Lien entre sRNA H/ACA et modifications des ARNr d'archaea

Nos études, ainsi que les études réalisées par d'autres équipes, ont permis de détecter expérimentalement 14, 4, 3, 5 et 6 résidus Ψ dans, respectivement, les ARNr des espèces de *P. abyssi*, *H. marismortui*, *H. halobium*, *A. fulgidus* et *S. acidocaldarius* (publication N°2 ; Kirpekar *et al.*, 2005 ; Del Campo *et al.*, 2005 ; Ofengand et Bakin, 1997 ; Tang *et al.*, 2002a ; Massenet *et al.*, 1999). Dans la publication N°2 nous avons pu définir, parmi les motifs H/ACA détectés chez *P. abyssi*, lesquels sont responsables de chacune des modifications détectées. Par ailleurs, l'équipe de A. Huttenhofer a aussi pu faire un lien entre les résidus Ψ qu'ils ont mis en évidence dans les ARNr de *A. fulgidus* et les 5 motifs H/ACA qu'ils avaient détecté dans cette espèce. Il faut néanmoins remarquer que nos travaux nous ont conduit à des corrections des propositions faites par cette équipe (publication N°2).

Ainsi, chez *P. abyssi*, nous avons attribué 12 des résidus Ψ détectés expérimentalement aux 7 sRNA H/ACA identifiés, et nous avons proposé que aCBF5 soit impliquée dans une pseudouridylation ne reposant pas sur l'utilisation d'un sRNA H/ACA (Ψ 2603 de l'ARNr 23S) (publication N°2). Il faut remarquer que les attributions décrites pour *P. abyssi* sont valables pour les deux espèces du genre *Thermococcus* dont la séquence génomique est disponible, à l'exception de l'absence d'un représentant de la famille M6 chez *T. gammatolerans*. Un point très intéressant est la présence, chez *M. kandleri* et chez trois espèces de l'ordre des Thermoplasmatales, de motifs H/ACA (M30) capables de cibler la position équivalente au résidu Ψ 2603 de l'ARNr 23S de *P. abyssi*, position non associée à un sRNA H/ACA dans cette espèce et que nous avons montrée être formée par le complexe aCBF5/aNOP10/aGAR1 en absence de guide. Comme dans la publication N°4, il pourrait donc s'agir d'une transition entre un système

guidé et un système non guidé. Ainsi, suivant les espèces considérées, la même ARN:Ψ-synthase pourrait agir avec ou sans guide. Autre point important, cette étude ne nous a pas permis d'identifier un motif H/ACA qui puisse guider la modification de la seule position qui reste orpheline (Ψ2585 dans l'ARNr 23S).

Chez *H. halobium* (également nommé *sp1* ou *salinarium*), nous avons identifié 3 sRNA H/ACA correspondant à 3 motifs. Nous avons pu assigner 3 des 4 résidus Ψ détectés expérimentalement aux motifs Hsp2, Hsp3 et Hsp4 (Tableau R.4 et 5). Comme présenté ci-après, Hsp3 pourrait également, par un appariement particulier, diriger la pseudouridylation à la 4^{ème} position. Ceci permettrait d'expliquer la formation des 4 résidus Ψ détectés expérimentalement.

Chez *H. marismortui* où nous avons seulement identifié 2 motifs H/ACA, Hmar1 et Hmar2, 3 résidus Ψ ont été détectés expérimentalement. Hmar1 appartient à la même famille que Hsp2 (M2) et Hmar2 à la même famille que Hsp3 (M3). En accord avec les observations précédentes, ces deux motifs sont capables de guider les 3 résidus Ψ détectés dans les ARNr de cette espèce. Aucun équivalent de Hsp4 attribué à la modification de la position 2590 dans l'ARNr 23S de *H. halobium* n'a été identifié chez *H. marismortui*. Ceci est en accord avec l'absence de résidu Ψ à la position équivalente dans les ARNr de *H. marismortui* (Kirpekar *et al.*, 2005 ; Del Campo *et al.*, 2005). On peut supposer d'après les motifs H/ACA que nous avons mis en évidence dans les deux autres espèces halophiles, *H. walsbyi* et *N. pharaonis*, que les ARNr de ces espèces contiennent les mêmes résidus Ψ que ceux de *H. halobium*.

Chez *S. acidocaldarius*, la situation est plus surprenante, puisque parmi les 6 résidus Ψ détectés dans l'ARNr 23S, un seul peut être attribué à un motif H/ACA, Sac2 (M8). Un autre motif (Sac1, M4) a été identifié mais il ne semble pas impliqué dans la modification des ARNr comme démontré dans la publication N°4, il est impliqué dans la modification d'un ARNt. Notons que ces deux motifs sont conservés dans les autres espèces de l'ordre des Sulfolobales, et qu'au sein de cet ordre, aucun autre motif H/ACA potentiel n'a pu être détecté. L'ordre des Sulfolobales semble se démarquer des espèces citées ci-dessus. En effet, nous ne sommes pas parvenus à détecter les motifs H/ACA qui permettraient de guider les modifications de 5 résidus Ψ (résidus Ψ2018, Ψ2024, Ψ2435, Ψ2492 et Ψ2505). Parmi ceux-ci,

- 3 sont présents dans le domaine V (Ψ2435, Ψ2492 et Ψ2505) et ne sont pas, à l'exception du résidu Ψ2435 également détecté chez *D. melanogaster*, modifiés dans les ARNr des organismes analysés, et

- 2 résidus Ψ sont localisés dans le domaine IV (Ψ2018, Ψ2024) pour lesquels nous avons identifié, dans d'autres espèces, des représentants de la famille M3 capables de cibler la position U2018 (Tableau R.5).

Ces observations soulèvent le problème de l'existence possible d'une catalyse sRNA H/ACA-indépendante de la formation de résidus Ψ dans les ARNr de certaines espèces d'archaea. Il sera important de vérifier expérimentalement cette hypothèse. Par conséquent, on ne peut exclure la co-existence de systèmes guidés et non-guidés chez certaines espèces.

Les motifs H/ACA des familles M1 à M35 que nous avons caractérisées ciblent soit l'ARNr 23S (21/35), soit l'ARNr 16S (14/35) ou encore les ARNt (2/35). Au sein des ARNr 23S, ils ciblent principalement des résidus des domaines V (12/21) et IV de l'ARNr 23S (5/21) qui sont situés au niveau du site catalytique. On remarque que certaines des familles dédiées au domaine V ciblent des résidus qui sont également pseudouridylés dans les ARNr bactériens ou eucaryotes (Fig. R.26B). De même, dans le domaine IV, les résidus Ψ qui sont conservés dans la plupart des ARNr procaryotes et eucaryotes analysés peuvent être ciblés par les familles M3 ou M31 (Fig. R.26A).

4. Caractéristiques structurales des motifs H/ACA identifiés

Il est évident que les caractéristiques des sRNA H/ACA que nous avons identifiés reposent sur les contraintes que nous avons imposées lors de leur recherche. Cependant, l'emploi du programme ERPIN et l'addition au fur et à mesure de leur détection des différents motifs H/ACA détectés dans chaque espèce nous a permis d'introduire un certain niveau de souplesse dans les contraintes imposées.

Il en ressort que tous les candidats présentent : 1) un motif en K-turn ou un motif en K-loop, 2) une boîte ANA (à l'exception de Mka15 qui présente un triplet GGA), 3) une hélice H1 riche en G/C et une hélice H2 comportant des mésappariements. Il est à noter que certains des motifs en K-turn ou K-loop que nous avons pris en considération présentent parfois certaines différences par rapport au motif consensus, à savoir, un doublet 5'-GA-3' du côté 5' où le résidu G est parfois remplacé par résidu A et, dans un cas, un résidu C. D'autre part, la partie 3' est le plus souvent 5'-NRUGA-3' (Fig. II.20). Il faut remarquer que certains motifs H/ACA présentent la particularité de contenir ce qui pourrait correspondre à deux motifs en K-turn adjacents (voir Ape7, Fig. 1 de la publication N°4).

Les motifs H/ACA identifiés mettent en évidence une grande variabilité de la taille de la partie du sRNA H/ACA située au dessus du motif en K-turn. Elle peut atteindre jusqu'à 60 nt dans les motifs identifiés mais peut être presque nulle dans le cas de la présence d'un motif en K-loop. Afin de ne pas manquer certains motifs H/ACA il est nécessaire de prendre en compte cette variabilité. Un juste milieu doit donc nécessairement être pris en compte pour une recherche efficace. Un autre point de variabilité que nous avons noté est celui de la taille de l'élément anti-sens en 5', comme c'est le cas pour le motif Pab40-2 pour lequel la taille de 24 nt s'explique par la formation d'une structure tige-boucle au sein de cet élément. Comme mentionné dans l'article N°2, nous n'avons pas, au départ, identifié le motif Pab19 de *P. abyssi* du fait de la taille importante de l'élément anti-sens en 5' (15 nt). Il est donc, là aussi, important de tenir compte de la variation possible de la taille de cet élément dans les recherches, mais là encore le relâchement des contraintes lors de la recherche peut conduire à l'obtention de nombreux faux positifs.

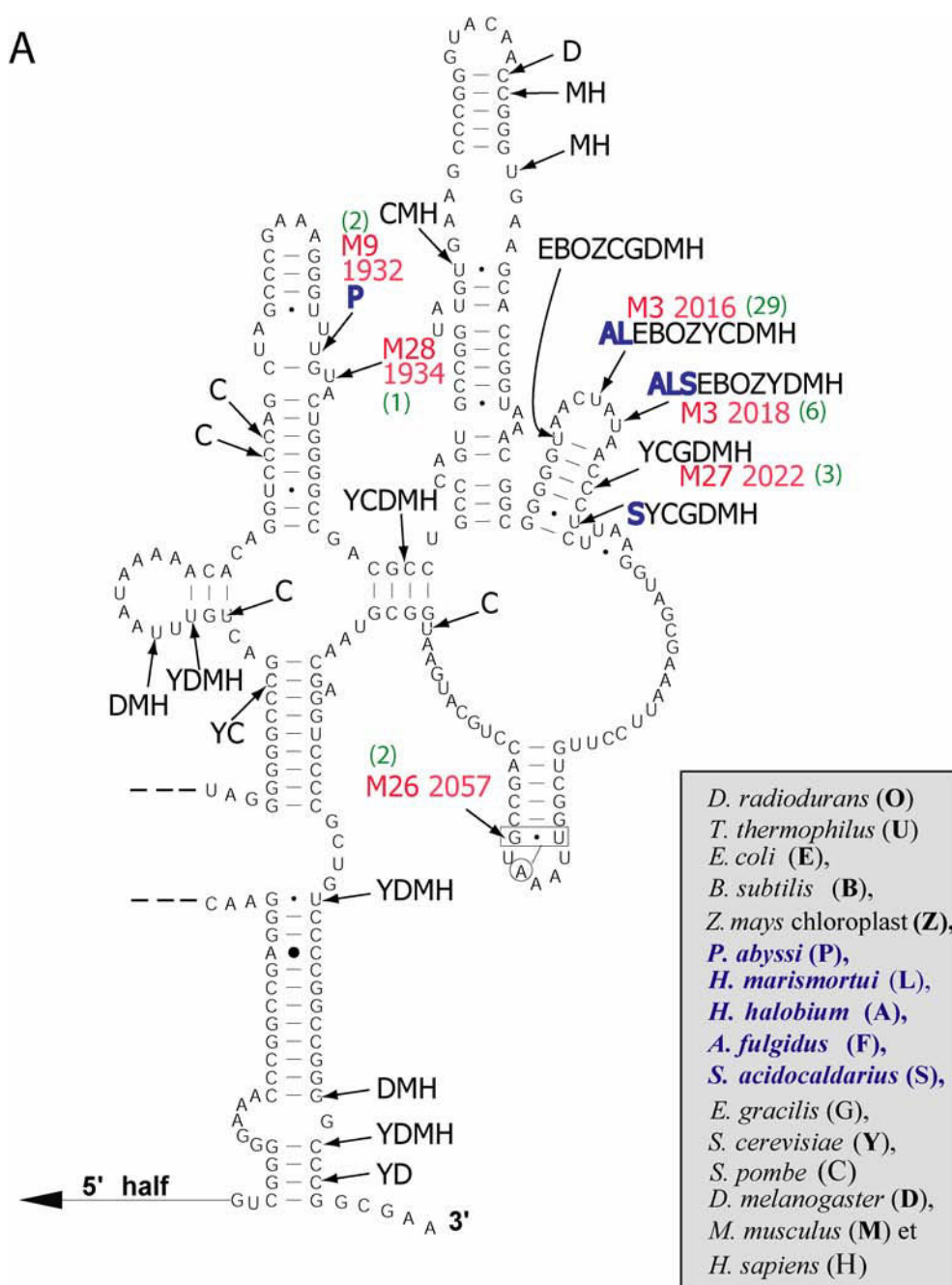


Figure R.26 : Localisation des résidus Ψ identifiés expérimentalement, ou supposés guidés par des sRNA H/ACA d'après notre étude, au sein du domaines IV (Panneau A) et du site PTC du domaine V (Panneau B) des ARNr 23S d'archaea.

La structure secondaire de l'ARNr 23S de *P. abyssi* est adaptée de la structure de l'ARNr 23S de *T. celer* (M67497). La numérotation qui y est indiquée est celle de *P. abyssi*. Les positions des résidus Ψ détectés sont indiquées par une lettre (voir l'encart). Les résidus Ψ détectés dans les espèces d'archaea sont indiqués en bleu. Les numéros des familles et les positions des résidus ciblés par celles-ci sont indiquées en rouge. Le nombre de représentant par famille est représenté en vert entre parenthèses.

5. Caractéristiques des appariements proposés entre les motifs H/ACA identifiés et leurs ARN cibles

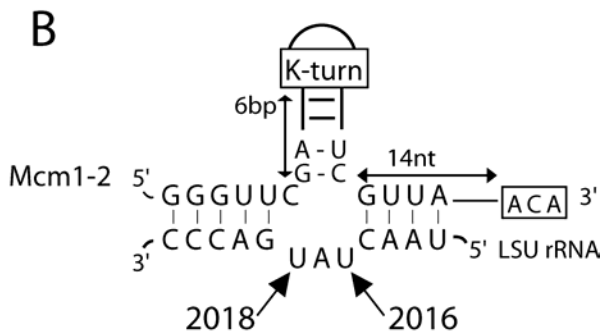
Comme évoqué dans le chapitre d'ouvrage N°2, des règles précises semblent définies lors de l'interaction du motif H/ACA avec sa cible. Celles-ci concernent principalement 1) la distance appelée d1 (entre la boîte ACA et le résidu U ciblé) comprise entre 14 et 15 nt, 2) la distance d2 (entre le motif en K-turn ou K-loop et le résidu U ciblé) comprise entre 5 et 6 pb, et 3) un appariement du sRNA avec sa cible sur au moins 8 pb avec une stabilité forte imposée pour l'appariement formé avec l'élément antisens 3' (publication N°2). L'examen de l'ensemble des appariements des motifs H/ACA des 35 familles révèle que ces règles semblent applicables dans les autres espèces d'archaea à l'exception de quelques écarts observés :

- la distance d1 peut être de 13 nt ou de 16 nt. Cependant, il semblerait qu'une distance d1 de 13 nt ne soit autorisée que lorsque l'hélice basale est formée à partir d'un mauvais empilement des paires de bases. Réciproquement, un bon empilement des paires de bases d'une longue hélice basale favoriserait une distance d1 de 16 nt.

- la distance d2 pourrait être de 7 pb lorsque l'empilement des paires de bases est optimal.

Par ailleurs, la règle de la présence de seulement 2 résidus non appariés (5'-UN-3') entre les 2 appariements réalisés avec les éléments antisens 5' et 3' n'est pas stricte. Comme nous l'avons montré pour le sRNA Pab19 de *P. abyssi* (Publication N°2), des représentants de la famille M3 pourraient également autoriser la présence de trois nucléotides non appariés. Il est intéressant de noter dans ce cas que le troisième nucléotide est un résidu U (5'-UNU-3') et que celui-ci pourrait être modifié par ce type d'appariement. En effet, les motifs H/ACA de la famille M3 pourraient guider la modification du résidu en position U2016 mais aussi du résidu en position U2018. Ces deux résidus voisins sont effectivement modifiés dans les ARNr de *H. halobium* (Bakin et Offengand, 1997) et de *H. marismortui* (Del Campo *et al.*, 2005 ; Kirpekar *et al.*, 2005) et parmi les motifs H/ACA que nous avons identifiés seuls le sRNA Hsp3 et son homologue Hmar2 présentent des séquences guides permettant de cibler ces deux positions (Fig. R.27). Ce type d'appariement a également été observé chez d'autres espèces dont *M. marisnigri* chez laquelle le motif Mcm1-2 appartenant à la famille M3 est également proposé pour la modification de ces deux résidus (Fig. R.27). La présence du troisième nucléotide pourrait être autorisée par des caractéristiques structurales différentes de la jonction entre l'hélice apicale et les éléments antisens 5' et 3' :

- dans le cas du motif Pab19 nous observons qu'un résidu C n'est ni apparié avec un résidu de l'hélice apicale, ni compris dans la séquence de l'élément antisens 3' (Fig. 4G de la publication N°2),
- dans le cas du motif Hsp3, l'hélice apicale se termine, du côté de la boucle interne du motif H/ACA, par deux paires de bases non canoniques C.U (Fig. R.27A),
- dans le cas du motif Mcm1-2 on observe qu'un résidu C n'est ni apparié avec un résidu de l'hélice apicale, ni compris dans la séquence de l'élément antisens 5' (Fig. R.27B).



Le troisième résidu non apparié est ici un U qui pourrait être modifié par le biais de cet appariement particulier. Les différences avec la séquence de Nph2 de *N. pharaonis* (*N.p.*) sont indiquées. Pour rappel, les noms *halobium*, *sp1* ou *salinarium* décrivent la même espèce

antisens pr  sents en amont des bo  tes D et bo  tes D' ciblent des positions dans les ARNr qui sont proches dans la structure, les motifs H/ACA, lorsqu'ils sont associ  s, ne ciblent pas forc  ment des r  sidus proches en s  quence dans les ARNr. Ceci est en accord avec notre observation que les motifs H/ACA d'archaea semblent constitu  s d'unit  s fonctionnelles   changeables. Les diff  rentes combinaisons que nous observons soul  vent un certain nombre de questions, en particulier, sur des   changes entre modules de sRNA H/ACA au sein des g  nomes d'archaea. On peut bien s  r envisager que des recombinaisons puissent avoir lieu entre des r  gions contenant plusieurs g  nes successifs de motifs H/ACA. Les s  quences des h  lices H1 de ces motifs H/ACA sont en effet assez conserv  es d'un motif    un autre. Par ailleurs, on peut aussi imaginer des changements rapides de la s  quence cibl  e par un motif H/ACA en un motif ciblant une autre s  quence cible par un nombre tr  s limit   de mutations ponctuelles comme nous l'avons montr   pour les sRNA Pab160 et Pab35-1 qui, bien que tr  s similaires au niveau de leur s  quence, guident des modifications    des positions diff  rentes dans les ARNr et sont de ce fait class  s dans 2 familles diff  rentes selon notre mode de classification. Par ailleurs, en ce qui concerne l'individualisation de motifs H/ACA, une possibilit   est l'existence d'insertions d'ADN entre les s  quences ADN correspondant    deux motifs H/ACA. En plus du m  canisme conduisant au caract  re modulaire des motifs H/ACA, l'observation de sRNA H/ACA    1, 2 ou 3 motifs soul  ve la question de l'importance fonctionnelle que peut avoir l'association de plusieurs motifs.

Une premi  re hypoth  se envisag  e est que cette association soit la cons  quence de la transmission des g  nes de sRNA. En effet, il est plus probable que des motifs H/ACA soient maintenus au cours de l'  volution si ces motifs sont associ  s par deux ou par trois. Si c'est le cas, la distance qui s  pare deux motifs associ  s pourrait varier au cours de l'  volution. En effet, on remarque en analysant les repr  sentants des familles M2 et M3 qui sont souvent associ  s, que deux motifs provenant d'esp  ces de l'ordre des Halobacteriales pr  sentent de telles variations : les distances qui s  parent les motifs Nph1 et Nph2 de *N. pharaonis*, Hwa1 et Hwa2 de *H. walsbyi*, Hsp2 et Hsp3 de *H. halobium*, Hmar1 et Hmar2 de *H. marismortui* sont respectivement de 21 nt, 26 nt, 108 nt et 146 nt (Tableau R.3). Il serait int  ressant de savoir si ces motifs sont transcrits ensemble ou ind  pendamment l'un de l'autre, et s'ils sont matur  s ou non. Il faut aussi rappeler que certains motifs H/ACA semblent associ  s    des sRNA    bo  tes C/D. Nos r  sultats de Northern Blot concernant le sRNA Pab21 montrent deux bandes correspondant    deux formes sugg  rant qu'un motif H/ACA peut   tre co-transcrit avec un sRNA C/D mais qu'une   tape de maturation est sans doute requise pour s  parer les motifs H/ACA et C/D (Publication N  2, Fig. 3). Nos exp  riences r  alis  es avec le sRNA Sso1 montrent   galement que la pr  sence du sRNA C/D en 3' du motif H/ACA n'alt  re ni l'assemblage de la particule sRNP, ni l'activit   de cette derni  re (r  sultats non montr  s). Finalement, une cons  quence de cette hypoth  se serait l'observation de clusters de motifs H/ACA i.e. l'association de plus de trois motifs H/ACA au sein d'un m  me transcrit ou d'un m  me sRNA H/ACA voire l'association de plusieurs sRNA C/D avec plusieurs motifs H/ACA. Or, nous n'avons encore

jamais remarqué, dans les génomes d'archaea, l'existence de gènes de sRNA comportant plus de trois motifs H/ACA, ni même une association d'un motif H/ACA avec plusieurs sRNA C/D ou l'association d'un sRNA C/D avec plusieurs motifs H/ACA.

Une seconde hypothèse reposerait sur l'importance fonctionnelle de l'association de ces motifs pour l'efficacité de l'interaction avec les protéines des sRNP H/ACA. Il est probable que, comme nous l'avons déjà discuté auparavant (chapitre 3), la présence d'un second motif H/ACA favorise la fixation de aCBF5 sur le premier motif H/ACA. Par ailleurs, il faut noter que certains sRNA présentent des structurations supplémentaires. Ainsi, le sRNA Mka8 de *M. kandleri* qui comprend trois motifs, présente deux hélices supplémentaires (Fig. R.29). La première est formée par l'appariement de la région en amont de l'hélice H1 du premier motif 5' avec la région charnière entre le premier motif et le deuxième motif. La seconde est formée par appariement de la région en 5' avec la région en aval du dernier motif en 3'. Des structures similaires ont été observées pour le candidat Mka7 issu de la même espèce, mais aucune structuration supplémentaire n'a été relevée pour les sRNA H/ACA détectés dans les autres espèces. Il reste cependant envisageable que de telles structurations aient échappé à notre analyse focalisée sur le motif H/ACA et son appariement avec l'ARN cible. Le rôle de ces structurations supplémentaires, comme le rôle de structures en tige-boucles additionnelles présentes du côté 5' et décrites précédemment pour Pab35 ou Pab19, reste inexploré. Nous pouvons toutefois remarquer que ces structures convergent toutes vers la formation d'hélices qui rejettent les trois résidus du triplet ACA en dehors de celles-ci. Ces résultats suggèrent donc un rôle potentiel dans l'interaction du sRNA avec une ou plusieurs molécules de aCBF5.

Nous avons envisagé l'évolution possible de motifs H/ACA associés aux motifs de la famille M3. Cette dernière a été choisie car représentée dans la plupart des génomes d'archaea (Tableau R.5). La figure R.28 propose au moins trois types de transitions d'une famille à une autre permettant à un motif H/ACA d'évoluer pour cibler un nouveau résidu :

- une modification dans les séquences des éléments antisens 5' et 3' permet le plus souvent à un motif H/ACA de cibler un nouveau résidu, c'est le cas par exemple de la transition entre des représentants des familles M1 et M14 (ex. : motifs Pho40-1 et Mka8-1), ou entre représentants des familles M2 et M7 (ex. : motifs Hsp2 et Pto1-2).

- une modification dans la séquence du motif en K-turn peut conduire à la formation d'un nouveau motif en K-turn qui induit une nouvelle structuration et donc la définition de nouveaux éléments antisens 5' et 3', c'est le cas de la transition entre représentants des familles M2 et M9 (ex. : motifs Pho40-2 et Pab40-2) (Fig. R.30),

- une altération de la taille de l'hélice basale ou de la taille de l'élément antisens 3' affecte la distance d1 et conduit également à la définition de nouveaux éléments antisens. C'est le cas de la transition entre représentants des familles M7 et M12 (ex.: Pab35-1 et Pab160).

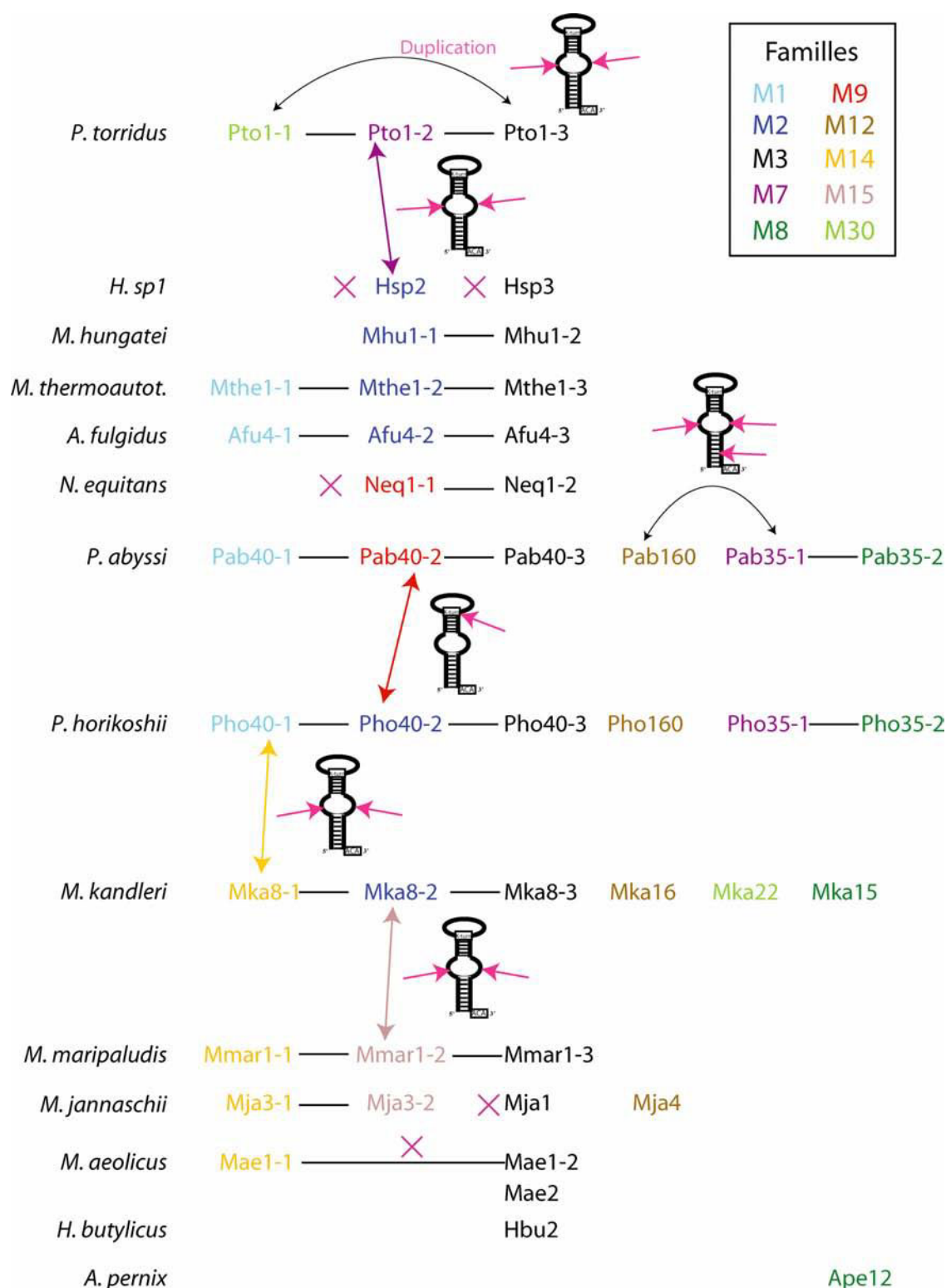


Figure R.28 : Représentation schématique de transitions d’une famille de motif H/ACA à une autre. Les familles sont indiquées par un code couleur défini dans l’encart. Les croix représentées en rose indiquent une absence d’association entre motifs H/ACA ou l’absence d’un motif H/ACA. Les flèches représentées en rose indiquent les principales différences qui existent entre deux motifs H/ACA et qui sont proposées comme étant à l’origine des transitions entre les familles de motifs H/ACA.

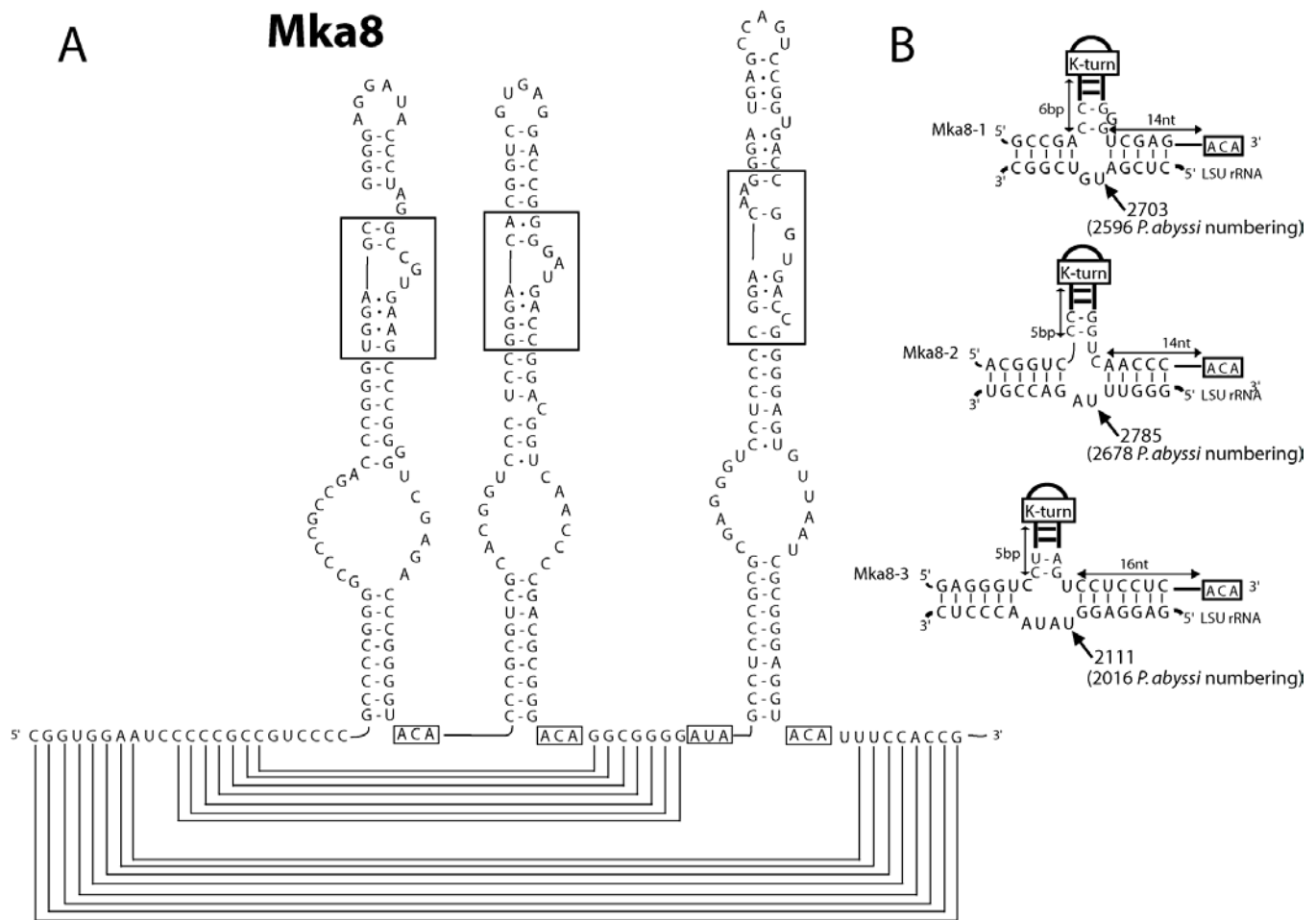


Figure R.29 : Représentation de la structure secondaire particulière du sRNA H/ACA Mka8 issu de *M. kandleri* (A). Les appariements prédits entre chacun des trois motifs de Mka8 et les ARNr 23S sont proposés en B.

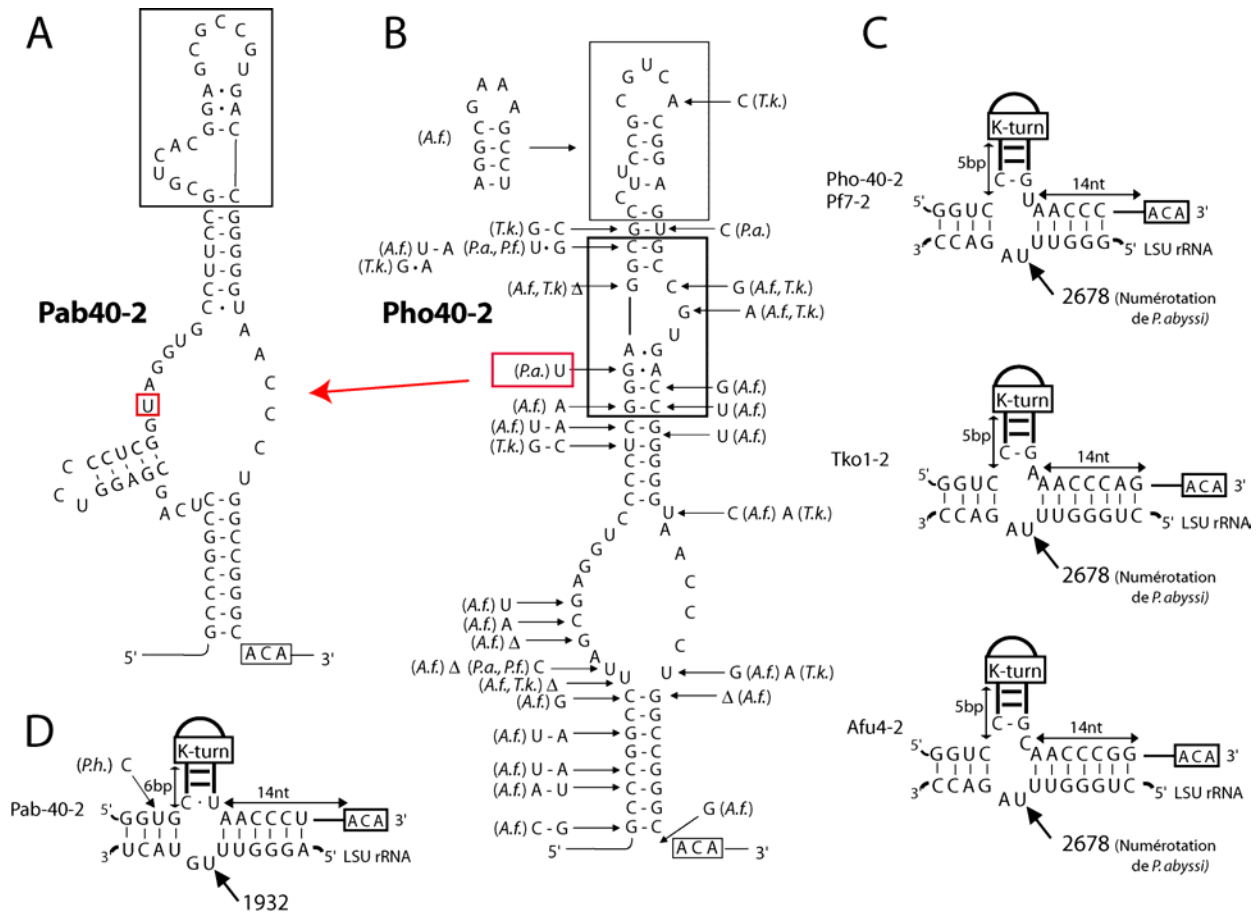


Figure R.30 : Représentation de l'influence de la formation du motif en K-turn sur la définition de la cible d'un motif H/ACA. Les structures secondaires des motifs Pab40-2 (A.) et Pho40-2 (B.) sont représentées. Les différences de nucléotides entre Pho40-2 et les équivalents identifiés chez *A. fulgidus* (A.f.), *T. kodakarensis* (T.k.) et *P. furiosus* (P.f.) sont indiquées sur la structure secondaire de Pho40-2. Les appariements prédits entre l'ARNr 23S et les motifs Pho40-2/Pf7-2, Tko1-2 et Afu4-2 sont représentés (C.). La mutation du résidu G de l'élément 5' du motif en K-turn (B.) est proposée comme étant à l'origine d'une structuration différente qui conduit le motif Pab40-2 à cibler un résidu différent de celui ciblé par les motifs Pho40-2, Pf7-2, Tko1-2 et Afu4-2 (D.).

Le programme RNA Forester (Package Vienna 1.6.4 ; Hofacker *et al.*, 1994) nous a permis de comparer entre elles les structures secondaires des motifs H/ACA selon un modèle d'évolution basé sur des travaux récents (Jiang *et al.*, 1995 ; Hoechsmann *et al.*, 2003). Cet alignement a abouti à la construction d'un arbre qui représente les relations possibles entre les motifs H/ACA (Fig. R.31). Son analyse révèle, par exemple, la proximité des motifs Pto1-1 et Pto1-3. Celle-ci suggère qu'une duplication de gène suivie d'une altération des séquences des éléments antisens, aurait permis la transition entre familles M3 et M30 représentées ici par Pto1-1 et Pto1-3 (Fig. R.28).

Cet arbre met également en évidence d'autres phénomènes de duplications récentes. Ainsi, par exemple, 4 motifs de *M. kandleri* (Mka1, Mka7-1, Mka7-2 et Mka19-1) seraient issus de duplications de gènes relativement récentes en comparaison à des événements similaires concernant les autres motifs. Par ailleurs, ces 4 motifs H/ACA sont les seuls représentants des familles auxquelles ils appartiennent.

Cette étude phylogénétique des motifs H/ACA doit encore être approfondie et permettrait, grâce à l'identification des phénomènes de duplication et de transition d'une famille à une autre, de présenter les motifs H/ACA comme des orthologues ou des paralogues.

7. Relation entre nombre de motifs H/ACA et conditions de croissance

Le tableau R.6 présente un bilan du nombre de motifs H/ACA identifiés dans les génomes d'archaea analysés. Des motifs H/ACA ont été identifiés dans tous les ordres d'archaea à l'exception des Cenarchaeales. Cet ordre ne compte pour le moment qu'un seul représentant dont le génome a pu être analysé, *C. symbiosium*. L'examen d'autres représentants est nécessaire pour conclure à une absence de motifs H/ACA dans ce phylum. Il faut cependant remarquer que ce représentant est psychrophile. Parmi les Methanosarcinales, une seule espèce présente des motifs H/ACA, il s'agit de *M. thermophila* qui est aussi le seul thermophile de cet ordre dont le génome a été analysé. En effet, les autres représentants de cet ordre sont mésophiles et ne présentent pas de motif H/ACA. Parmi les Thermoplasmatales, le génome de *F. acidarmanus* est le seul qui ne comporte pas de motif H/ACA. Les trois autres représentants de cet ordre présentent tous les trois 3 motifs H/ACA et sont thermophiles au contraire de *F. acidarmanus* qui est mésophile. Il semblerait donc qu'il y ait un lien entre la température optimale de croissance et le maintien de motifs H/ACA au sein d'un phylum.

Une corrélation entre nombre de sRNA C/D et température de croissance a été proposée (Dennis *et al.*, 2001). Il avait été remarqué que les espèces du genre *Pyrococcus* présentent environ 55 sRNA C/D, plus de 60 pour *P. aerophilum* contre 13 pour *S. solfataricus* et aucun pour *H. spl* ou *M. thermoautorophilum* (Tableau R.6). Il semble, d'une part que la recherche de sRNA C/D n'a pas été appliquée à suffisamment de génomes pour conclure à une corrélation de ce type et que, d'autre part, une telle corrélation n'est pas confirmée pour les motifs H/ACA d'archaea. En effet, il faut remarquer que les méthanogènes mésophiles de l'ordre des Methanomicrobiales ou des Methanobacteriales comportent 1 à 3 motifs H/ACA alors que des représentants de l'ordre des Thermoproteales, représenté par plusieurs hyperthermophiles, ne comportent qu'un seul motif H/ACA. De plus, le fait que les Thermoproteales ou les Desulfurococcales ne comportent pas autant de motifs que les Thermococcales ou les Methanopyrales suggère que, pris dans son ensemble, il ne semble pas exister de corrélation directe entre le nombre de motifs H/ACA détectés et la température optimale de croissance.

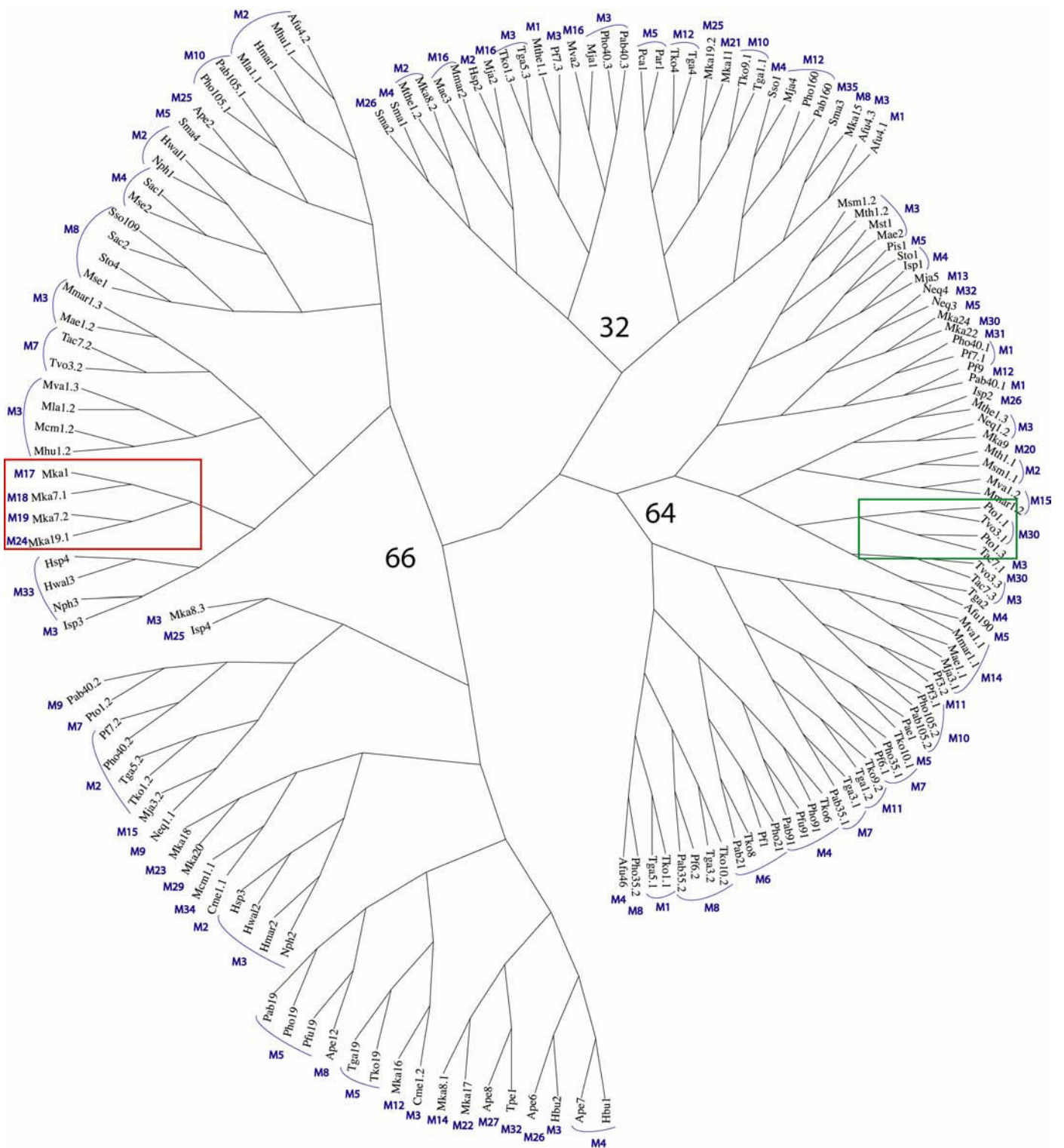


Figure R.31 : Représentation radiale de l'arbre phylogénétique établi pour les motifs H/ACA par l'utilisation du programme RNAForester (Hofacker *et al.*, 1994 ; Jiang *et al.*, 1995 ; Hoechsmann *et al.*, 2003).

Les familles auxquelles appartiennent les motifs H/ACA présentés sont représentées en bleu. Les régions encadrées en vert et en rouge correspondent aux branches comportant les motifs Pto1-1 et Pto1-3, et les motifs Mka1, Mka7-1, Mka7-2 et Mka19-1 respectivement.

Le pH ou la concentration saline du milieu peuvent être invoqués mais nous n'avons pas pu relever de corrélation de ce type. Seul le positionnement sur l'arbre phylogénétique semble donner des réponses en accord avec la situation observée dans les espèces considérées. On remarque en effet que les espèces dont le génome présente le plus de gènes de sRNA H/ACA, i.e. les espèces des ordres des Methanopyrales, Thermococcales et Methanococcales, sont celles qui branchent à proximité de la racine de l'arbre phylogénétique des archaea et qui ont donc conservé le plus de caractéristiques communes avec un ancêtre commun (Fig. I.6). Cette observation suggère que l'ancêtre commun des archaea comportait sans doute un nombre élevé de gènes de motifs H/ACA. Les gènes de ces motifs H/ACA auraient pu être perdus du fait des fortes pressions exercées pour maintenir une petite taille de génome et à la faveur de changements de conditions de croissance. Ainsi, les espèces d'archaea recensées ont évolué dans des niches écologiques différentes qui expliquent certainement pourquoi ces espèces ont été plus enclines à perdre ou à acquérir de nouveaux motifs.

Par ailleurs, il faut noter qu'il n'existe pas de lien entre le pourcentage G/C ou la taille de la séquence génomique d'une espèce d'archaea et le nombre de motifs H/ACA. Une telle corrélation aurait pu traduire un biais dans la recherche de ces motifs. En effet, l'alignement fourni à ERPIN est riche en G/C (environ 70%) et il est donc plus probable d'identifier des motifs H/ACA dans des espèces dont la séquence génomique est riche en G/C telles que *M. kandleri*, *H. sp1* ou *M. maripaludis*. De même, il est plus probable d'identifier des motifs H/ACA dans des séquences de plus grande taille. Or, nous n'observons pas de relation de ce type. Les 4 motifs H/ACA de *N. equitans* en sont une illustration : le génome de *N. equitans* est le plus petit de tous les génomes d'archaea et le pourcentage G/C de sa séquence génomique est, avec *M. jannashii* et *M. vanniellii*, parmi les plus faibles. En outre, il était surprenant de relever des motifs H/ACA dans le petit génome de cet archaea qui a une vie parasitaire. Son hôte, *I. hospitalis*, comporte également 4 motifs H/ACA. Ensemble, les deux espèces comportent donc 8 motifs dont seulement 2 appartiennent à la même famille (M3) qui est la plus représentée parmi toutes les espèces d'archaea. Il ne semble donc pas qu'il y ait eu transfert horizontal de gènes de sRNA H/ACA entre les espèces de *I. hospitalis* et *N. equitans*. De même, entre espèces également éloignées sur le plan phylogénétique mais occupant la même niche écologique telles que les thermophiles acidophiles de l'ordre des Sulfolobales et des Thermoplasmatales, nous n'observons pas de motif H/ACA de la même famille et de ce fait aucun transfert latéral de ces gènes n'est proposé.

(Page suivante) Tableau R.6 : Bilan du nombre de motifs H/ACA identifiés dans les 46 génomes d'archaea analysés. Certaines caractéristiques des espèces et des génomes considérés sont indiquées. Les valeurs les plus faibles sont indiquées en bleu, les valeurs les plus élevées sont indiquées en rouge. Les cases jaunes correspondent aux validations expérimentales réalisées par d'autres équipes, les cases vertes correspondent à nos validations expérimentales. Ces validations, à l'exception de celles que nous avons conduites chez *P. abyssi*, sont toutes partielles. Le nombre de sRNA C/D identifiés est indiqué pour certaines espèces analysées (Dennis *et al.*, 2001).

Ordre	Genre / esp�ce	T opt (�C)	Taille (Mpb)	%GC	Motifs H/ACA pr�dits	sRNA C/D pr�dits
Desulfurococcales	<i>Aeropyrum pernix</i>	90-95	1,67	56	5	23
	<i>Hyperthermus butylicus</i>	95-106	1,67	54	2	
	<i>Ignicoccus sp.</i>	90	1,29	57	4	
	<i>Staphylothermus marinus</i>	92	1,57	35	4	
Thermoproteales	<i>Thermofilum pendens</i>	88	1,78	57	1	
	<i>Pyrobaculum islandicum</i>	100	1,82	49	1	
	<i>Pyrobaculum aerophilum</i>	100	2,22	59	1	(60-70)
	<i>Pyrobaculum calidifontis</i>	90-95	2,01	57	1	
	<i>Pyrobaculum arsenaticum</i>	95	2,12	55	1	
Sulfolobales	<i>Sulfolobus solfataricus</i>	75-80	2,99	36	2	13
	<i>Sulfolobus acidocaldarius</i>	80	2,23	37	2	29
	<i>Sulfolobus tokodai</i>	80	2,69	33	2	
	<i>Metallosphaera sedula</i>	75	2,19	46	2	
Cenarchaeales	<i>Cenarchaeum symbiosum</i>	10	2,05	58	0	
Nanoarchaeales	<i>Nanoarchaeum equitans</i>	90	0,49	31	4	
Archaeoglobales	<i>Archaeoglobus fulgidus</i>	83	2,18	49	5	4
Halobacteriales	<i>Halobacterium sp.</i>	42	2,01	68	3	0
	<i>Haloarcula marismortui</i>	50	4,28	62	2	
	<i>Haloquadratum walsbyi</i>	40	3,13	48	3	
	<i>Natronomonas pharaonis</i>	45	2,59	63	3	
Methanococcales	<i>Methanocaldococcus jannashii</i>	85	1,66	31	6	8
	<i>Methanococcus maripaludis</i>	35-40	1,66	69	4	
	<i>Methanococcus aeolicus</i>	42	1,57	33	4	
	<i>Methanococcus vannieli</i>	35-40	1,72	31	4	
Methanopyrales	<i>Methanopyrus kandleri</i>	100-105	1,69	61	17	
Methanosarcinales	<i>Methanococco�ides burtonii</i>	23	2,58	41	0	
	<i>Methanosaeta thermophila</i>	50-65	1,88	54	3	
	<i>Methanosarcina acetivorans</i>	35	5,75	43	0	
	<i>Methanosarcina barkeri</i>	35	4,83	39	0	
	<i>Methanosarcina mazei</i>	37	4,09	42	0	
Methanobacteriales	<i>Methanothermobacter thermoautotrophicum</i>	62-67	1,75	50	2	0
	<i>Methanobrevibacter smithii</i>	37-40	1,85	31	2	
	<i>Methanosphaera stadtmanae</i>	37	1,77	27	1	
Methanomicrobiales	<i>Methanospirillum hungatei</i>	37	3,54	45	2	
	<i>Methanoculleus marisnigri</i>	21-25	2,48	62	2	
	<i>Methanoregula boonei</i>	37	2,54	54	2	
	<i>Methanocorpusculum labreanum</i>	37	1,8	50	2	
Thermococcales	<i>Pyrococcus abyssi</i>	96	1,77	45	11	56
	<i>Pyrococcus horikoshii</i>	98	1,74	42	11	51
	<i>Pyrococcus furiosus</i>	100	1,9	41	11	52
	<i>Thermococcus kodakaraensis</i>	95	2,08	53	11	
	<i>Thermococcus gammatolerans</i>	85-90	2,04	53	10	
Thermoplasmatales	<i>Thermoplasma acidophilum</i>	55-60	1,56	46	3	
	<i>Thermoplasma volcanium</i>	55-60	1,58	40	3	
	<i>Picrophilus torridus</i>	60	1,55	36	3	
	<i>Ferroplasma acidarmanus</i>	40	1,94	36	0	

8. Présence de sRNA H/ACA dans les éléments extra-chromosomiques ?

Puisque certains génomes d'archaea ne comportent aucun motif H/ACA, et que, pour certaines espèces, nous ne sommes pas parvenus à attribuer des motifs H/ACA à des résidus Ψ dits orphelins, il restait envisageable que les éléments extra-chromosomiques comportent des motifs H/ACA. Les séquences de certains plasmides peuvent dépasser 400 kpb et sont donc de bons candidats pour abriter des gènes de sRNA. Nous avons donc analysé 44 séquences de plasmides d'archaea ainsi qu'une dizaine de séquences de virus d'archaea, mais nous ne sommes pas parvenus à identifier de motif H/ACA.

Cependant, en utilisant ERPIN avec un alignement de sRNA C/D, nous sommes parvenus à détecter des sRNA C/D au sein de plasmides de *S. islandicus*. Nous avons également réussi à définir les résidus ciblés par ces sRNA C/D dans les séquences de l'ARNr 16S de *S. islandicus*.

Ces résultats suggèrent qu'il est probable que des plasmides ou des virus comportent des sRNA H/ACA qui ciblent les ARNr des espèces qui les portent. Tous les plasmides et les virus présents dans les cellules de *P. abyssi* ou *S. acidocaldarius* ne sont pas caractérisés ou séquencés, ce qui expliquerait peut-être pourquoi nous ne parvenons pas à identifier les motifs H/ACA impliqués dans les modifications de certains résidus Ψ détectés expérimentalement dans les ARNr de ces espèces.

CONCLUSION

Nos travaux nous ont conduit à développer plusieurs approches originales permettant :

- la recherche de gènes d'ARNnc par une approche bioinformatique de génomique comparative applicable à des espèces d'archaea suffisamment proches sur le plan phylogénétique,
- la recherche bioinformatique de gènes de sRNA H/ACA dans les séquences génomiques d'archaea ainsi que la recherche de leurs cibles dans les ARNnc,
- la mise au point des conditions d'étude de l'assemblage de particules sRNP H/ACA et la reconstitution de particules actives *in vitro*.

L'application de ces approches bioinformatiques et expérimentales a abouti à plusieurs conclusions.

Nous avons identifié 7 gènes de sRNA H/ACA correspondant à 11 motifs H/ACA dans chacune des trois espèces de *Pyrococcus* que nous avons étudiées. Nous avons détecté expérimentalement 14 résidus Ψ dans les ARNr de *P. abyssi* et il est probable, compte tenu des limites de notre méthode de détection et du fait que nous n'avons analysé que 25% des séquences des ARNr, que les ARNr de *P. abyssi* présentent au moins 17 résidus Ψ . Parmi ces 17 résidus Ψ , 15 ont pu être expérimentalement attribués à l'action des 11 motifs H/ACA identifiés. L'un des deux résidus pour lequel aucun sRNA H/ACA n'a pu être attribué pourrait être modifié par le complexe aCBF5/aNOP10/aGAR1 sans avoir recours à un sRNA H/ACA.

Nous avons démontré que aCBF5 et aNOP10 sont indispensables à l'activité des particules sRNP H/ACA, alors que L7Ae et aGAR1 améliorent la vitesse de formation de résidus Ψ . Notre étude a contribué à la définition des interactions entre sRNA H/ACA et les protéines des sRNP d'une part, et entre sRNA H/ACA et les ARN cibles d'autre part. Nous avons notamment montré que l'accès au triplet ANA en 3' du motif H/ACA est primordial pour la fixation de aCBF5 et que des structurations supplémentaires peuvent contribuer à cette fixation. Nous avons également proposé que l'appariement de l'ARN cible avec l'élément antisens 3' du sRNA H/ACA est stabilisé par la présence de aCBF5 et que des contraintes structurales, principalement les distances d1 et d2 séparant respectivement le résidu U ciblé de la boîte ANA et du motif en K-turn ou K-loop, doivent être respectées.

Nous avons suggéré que aCBF5 peut catalyser, sans avoir recours à un sRNA H/ACA, la formation du résidu Ψ 55 dans les ARNt élongateurs, et dans ce système non guidé, les protéines aNOP10 et aGAR1 amélioreraient la vitesse de formation du résidu Ψ . Nous avons ainsi démontré que deux réseaux d'interactions différents entre aNOP10 et aCBF5 semblent impliqués dans le système guidé (sRNP H/ACA) et le système non guidé.

Nous avons mis en évidence près de 90 nouveaux sRNA H/ACA dans les autres espèces d'archaea dont le génome est disponible, portant ainsi le nombre de sRNA H/ACA identifiés dans 46 espèces d'archaea à 123, correspondant à 162 motifs H/ACA. Nous avons établi une nomenclature permettant de regrouper les 162 motifs H/ACA en 35 familles fonctionnelles. Ceci

nous a permis de mettre en évidence le caractère modulaire des motifs H/ACA d'archaea : les motifs H/ACA semblent en effet pouvoir s'associer indifféremment des résidus qu'ils ciblent.

Nous avons identifié, parmi ces 35 familles, une famille de motifs H/ACA qui cible la position 35 des ARNt^{Tyr} de certains crénotes. L'étude d'un représentant de cette famille chez *S. solfataricus*, Sso1, a montré que ce sRNA H/ACA pourrait compenser une absence d'activité de l'enzyme aPus7 pourtant supposée responsable de cette modification.

Nos résultats ont donc fortement participé à la compréhension de la contribution des ARN:Ψ-synthases et des sRNP H/ACA à la modification des ARNr et ARNt. Sur la base de nos résultats, nous avons représenté dans la figure C.1 un bilan des modes d'action de 3 des 4 ARN:Ψ-synthases identifiées dans le domaine Archaea. Nous proposons que :

- l'enzyme aCBF5, associée aux protéines aNOP10 et aGAR1, pourrait être impliquée dans la modification de la position 55 de tous les ARNt et pourrait également être directement impliquée dans la modification de résidus des ARNr comme le suggèrent nos données concernant le résidu Ψ2603 de l'ARNr 23S de *P. abyssi*,
- l'enzyme aCBF5, au sein d'une particule sRNP H/ACA peut modifier de nombreux résidus des ARNr 16S et 23S, mais aussi des résidus des ARNt comme le montrent nos données sur la formation du résidu 35 de l'ARNt^{Tyr},
- l'enzyme aPus10 pourrait être impliquée à la fois dans la modification des résidus en position 54 et 55 des ARNt,
- l'enzyme aPus7 de *P. abyssi* est impliquée dans les modifications des résidus en position 13 des ARNt et 35 des ARNt^{Tyr} (GUA), l'enzyme aPus7 de *S. solfataricus* a une faible affinité en position 13 et ne semble pas avoir d'activité en position 35 d'où l'attribution de cette modification à une sRNP H/ACA (Sso1).

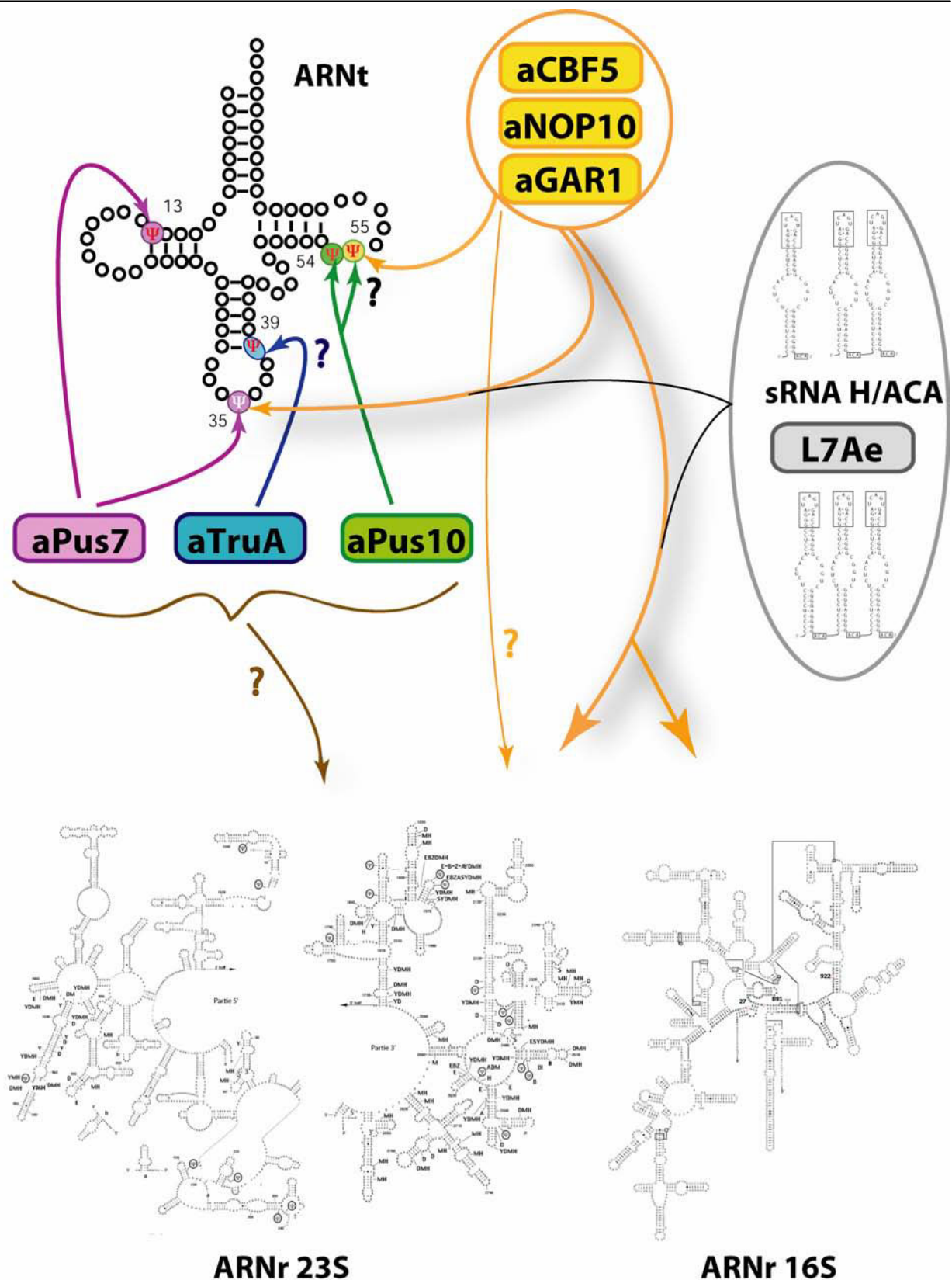


Figure C.1 : Représentation schématique de la contribution aux pseudouridylations des ARNr et ARNt proposée pour les 4 ARN:Ψ-synthases d'archaea

Les flèches ombrées correspondent au recours aux particules sRNP H/ACA. Les positions des résidus modifiés dans les ARNr ne sont pas détaillées. Les « ? » indiquent des attributions suggérées par nos résultats ou par analogie avec les bactéries chez lesquelles l'enzyme TruA est impliquée dans la modification des résidus aux positions 38 et 39 des ARNt.

REFERENCES

A

- Ackermann, H. W. (2001). Frequency of morphological phage descriptions in the year 2000. Brief review. *Arch Virol* **146**, 843-57.
- Agris, P. F. (1996). The importance of being modified: roles of modified nucleosides and Mg^{2+} in RNA structure and function. *Progress in Nucleic Acid Research* **53**, 79-129.
- Aiba, H. (2007). Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr Opin Microbiol* **10**, 134-9.
- Aittaleb, M., Rashid, R., Chen, Q., Palmer, J. R., Daniels, C. J., et Li, H. (2003). Structure and function of archaeal box C/D sRNP core proteins. *Nature Struct. Biol.* **10**, 256-263.
- Allers, T., et Mevarech, M. (2005). Archaeal genetics - the third way. *Nat Rev Genet* **6**, 58-73.
- Althoff, S., Selinger, D., et Wise, J. A. (1994). Molecular evolution of SRP cycle components: functional implications. *Nucleic Acids Res* **22**, 1933-47.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402.
- Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L., et Storz, G. (1997). A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* **90**, 43-53.
- Altuvia, S., et Wagner, E. G. (2000). Switching on and off with RNA. *Proc Natl Acad Sci U S A* **97**, 9824-6.
- Amadio, J. P., et Walsh, C. A. (2006). Brain evolution and uniqueness in the human genome. *Cell* **126**, 1033-5.
- Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T., et Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr Biol* **13**, 807-18.
- Andersen, A. A., et Panning, B. (2003). Epigenetic gene regulation by noncoding RNAs. *Curr Opin Cell Biol* **15**, 281-9.
- Ansmant, I., Massenet, S., Grosjean, H., Motorin, Y., et Branlant, C. (2000). Identification of the *Saccharomyces cerevisiae* RNA:pseudouridine synthase responsible for formation of Ψ_{2819} in 21S mitochondrial ribosomal RNA. *Nucleic Acids Res* **28**, 1941-1946.
- Ansmant, I., et Motorin, I. (2001). [Identification of RNA modification enzymes using sequence homology]. *Mol Biol (Mosk)* **35**, 248-67.
- Ansmant, I., Motorin, Y., Massenet, S., Grosjean, H., et Branlant, C. (2001). Identification and characterization of tRNA: Ψ_{31} -synthase (Pus6p) in yeast *S. cerevisiae*. *JBC* **276**, 34934-34940.
- Aoki, H., Dekany, K., Adams, S. L., et Ganoza, M. C. (1997). The gene encoding the elongation factor P protein is essential for viability and is required for protein synthesis. *J Biol Chem* **272**, 32254-9.
- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203-7.
- Aravind, L., et Koonin, E. V. (1999). DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res* **27**, 4658-70.
- Arena, F., Ciliberto, G., Ciampi, S., et Cortese, R. (1978). Purification of pseudouridylate synthetase I from *Salmonella typhimurium*. *Nucleic Acids Res* **5**, 4523-4536.
- Argaman, L., et Altuvia, S. (2000). hflA repression by OxyS RNA kissing complex formation at two sites results in a stable antisense-target RNA complex. *J.Mol.Biol.* **28**, 1101-12.
- Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E. G., Margalit, H., et Altuvia, S. (2001). Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* **11**, 941-50.
- Arn, E. A., et Abelson, J. N. (1996). The 2'-5' RNA ligase of *Escherichia coli*. Purification, cloning, and genomic disruption. *J Biol Chem* **271**, 31145-53.
- Arnez, J. G., et Steitz, T. A. (1994). Crystal structure of unmodified tRNA^{Gln} complexed with glutamyl- tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry* **33**, 7560-7567.
- Assens, C., Liautard, J. P., Sri-Widada, J., Brunel, C., et Jeanteur, P. (1982). Highly purified snRNPs retain antigenic determinants towards anti Sm antibodies. *Biochem Biophys Res Commun* **106**, 953-60.
- Atzorn, V., Fragapane, P., et Kiss, T. (2004). U17/snoR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. *Mol Cell Biol* **24**, 1769-78.
- Auffinger, P., et Westhof, E. (1998). Effects on pseudouridylation on tRNA hydration and dynamics: a theoretical approach. In "The modification and editing of RNA" (H. Grosjean et R. Benne, Eds.). ASM Press, Washington, DC.
- Auxilien, S., Crain, P. F., Trewyn, R. W., et Grosjean, H. (1996). Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. *JouRNal of Molecular Biology* **262**, 437-458.
- Avner, P., et Heard, E. (2001). X-chromosome inactivation: counting, choice and initiation. *Nat Rev Genet* **2**, 59-67.

B

- Babitzke, P., et Romeo, T. (2007). CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr Opin Microbiol* **10**, 156-63.
- Bachellerie, J. P., et Cavaille, J. (1998). Small nucleolar RNAs guide the ribose methylations of eukaryotic rRNAs. In "Modification and Editing of RNA" (H. Grosjean et R. Benne, Eds.), pp. 255-272. American Society for Microbiology, Washington DC.
- Bachellerie, J. P., Cavaille, J., et Huttenhofer, A. (2002). The expanding snoRNA world. *Biochimie* **84**, 775-90.
- Bacher, C. P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., et Heard, E. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol* **8**, 293-9.
- Badger, J. H., et Olsen, G. J. (1999). CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol* **16**, 512-24.
- Badis, G., Fromont-Racine, M., et Jacquier, A. (2003). A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *RNA* **9**, 771-9.
- Bagni, C., et Lapeyre, B. (1998). Gar1p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element. *J Biol Chem* **273**, 10868-73.
- Baker, D. L., Youssef, O. A., Chastkofsky, M. I., Dy, D. A., Terns, R. M., et Terns, M. P. (2005). RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev* **19**, 1238-48.
- Bakin, A., Kowalak, J. A., McCloskey, J. A., et Ofengand, J. (1994a). The single pseudouridine residue in *Escherichia coli* 16S RNA is located at position 516. *Nucleic Acids Res* **22**, 3681-3684.
- Bakin, A., Lane, B. G., et Ofengand, J. (1994b). Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry* **33**, 13475-13483.
- Bakin, A., et Ofengand, J. (1995). Mapping of the 13 pseudouridine residues in *Saccharomyces cerevisiae* small subunit ribosomal RNA to nucleotide resolution. *Nucleic Acids Res* **23**, 3290-3294.
- Balakin, A. G., Smith, L., et Fournier, M. J. (1996). The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86**, 823-34.
- Baliga, N. S., Goo, Y. A., Ng, W. V., Hood, L., Daniels, C. J., et DasSarma, S. (2000). Is gene expression in *Halobacterium* NRC-1 regulated by multiple TBP and TFB transcription factors? *Mol Microbiol* **36**, 1184-5.
- Baliga, N. S., Bonneau, R., Facciotti, M. T., Pan, M., Glusman, G., Deutsch, E. W., Shannon, P., Chiu, Y., Weng, R. S., Gan, R. R., et al. (2004). Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome Res* **14**, 2221-34.
- Ban, N., Nissen, P., Hansen, J., Moore, P. B., et Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905-920.
- Bapteste, E., Brochier, C., et Boucher, Y. (2005). Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* **1**, 353-63.
- Bare, L. A., et Uhlenbeck, O. C. (1986). Specific substitution into the anticodon loop of yeast tyrosine transfer RNA. *Biochemistry* **25**, 5825-30.
- Barneche, F., Gaspin, C., Guyot, R., et Echeverria, M. (2001). Identification of 66 box C/D snoRNAs in *Arabidopsis thaliana*: extensive gene duplications generated multiple isoforms predicting new ribosomal RNA 2'-O-methylation sites. *J Mol Biol* **311**, 57-73.
- Barns, S. M., Delwiche, C. F., Palmer, J. D., et Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A* **93**, 9188-93.
- Barry, E. R., et Bell, S. D. (2006). DNA replication in the archaea. *Microbiol Mol Biol Rev* **70**, 876-87.
- Barta, A., Kuechler, E., Branlant, C., Sri Widada, J., Krol, A., et Ebel, J. P. (1975). Photoaffinity labelling of 23 S RNA at the donor-site of the *Escherichia coli* ribosome. *FEBS Lett* **56**, 170-4.
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-97.
- Bartlett, M. S., Thomm, M., et Geiduschek, E. P. (2004). Topography of the euryarchaeal transcription initiation complex. *J Biol Chem* **279**, 5894-903.
- Bartlett, M. S. (2005). Determinants of transcription initiation by archaeal RNA polymerase. *Curr Opin Microbiol* **8**, 677-84.
- Bayfield, M. A., Thompson, J., et Dahlberg, A. E. (2004). The A2453-C2499 wobble base pair in *Escherichia coli* 23S ribosomal RNA is responsible for pH sensitivity of the peptidyltransferase active site conformation. *Nucleic Acids Res* **32**, 5512-8.
- Beattie, T. L., Zhou, W., Robinson, M. O., et Harrington, L. (1998). Reconstitution of human telomerase activity in vitro. *Curr Biol* **8**, 177-80.
- Becker, H. F., Motorin, Y., Planta, R. J., et Grosjean, H. (1997a). The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalysing the formation of Ψ55 in both mitochondrial and cytoplasmic tRNAs. *Nucleic Acids Research* **25**, 4493-4499.
- Becker, H. F., Motorin, Y., Sissler, M., Florentz, C., et Grosjean, H. (1997b). Major identity determinants for enzymatic formation of ribothymidine and pseudouridine in the TΨ-loop of yeast tRNAs. *JMB* **274**, 505-518.

- Behm-Ansmant, I., Urban, A., Ma, X., Yu, Y.-T., Motorin, Y., et Branlant, C. (2003). The *Saccharomyces cerevisiae* U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. *RNA* **9**, 1371-1382.
- Behm-Ansmant, I., Grosjean, H., Massenet, S., Motorin, Y., et Branlant, C. (2004). Pseudouridylation at position 32 of mitochondrial and cytoplasmic tRNAs requires two distinct enzymes in *Saccharomyces cerevisiae*. *J Biol Chem* **279**, 52998-3006.
- Behm-Ansmant, I., Massenet, S., Immel, F., Patton, J. R., Motorin, Y., et Branlant, C. (2006). A previously unidentified activity of yeast and mouse RNA:pseudouridine synthases 1 (Pus1p) on tRNAs. *RNA* **12**, 1583-93.
- Beja, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., Jovanovich, S. B., Gates, C. M., Feldman, R. A., Spudich, J. L., et al. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902-6.
- Beja, O., Spudich, E. N., Spudich, J. L., Leclerc, M., et DeLong, E. F. (2001). Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786-9.
- Belay, N., Johnson, R., Rajagopal, B. S., de Macario, E. C., et Daniels, L. (1988). Methanogenic bacteria from human dental plaque. *Appl Environ Microbiol* **54**, 600-3.
- Belay, N., Mukhopadhyay, B., Conway de Macario, E., Galask, R., et Daniels, L. (1990). Methanogenic bacteria in human vaginal samples. *J Clin Microbiol* **28**, 1666-8.
- Bell, S. D., et Jackson, S. P. (1998). Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol* **6**, 222-8.
- Bell, S. D., Magill, C. P., et Jackson, S. P. (2001). Basal and regulated transcription in Archaea. *Biochem Soc Trans* **29**, 392-5.
- Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P., et White, M. F. (2002). The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science* **296**, 148-51.
- Bell, S. D. (2005). Archaeal transcriptional regulation--variation on a bacterial theme? *Trends Microbiol* **13**, 262-5.
- Bell, S. D., et Jackson, S. P. (2001). Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol* **4**, 208-13.
- Bell, S. P., et Dutta, A. (2002). DNA replication in eukaryotic cells. *Annu Rev Biochem* **71**, 333-74.
- Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., et al. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* **37**, 766-70.
- Bernstein, E., Caudy, A. A., Hammond, S. M., et Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-6.
- Besemer, J., Lomsadze, A., et Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* **29**, 2607-18.
- Billoud, B., Kontic, M., et Viari, A. (1996). Palingol: a declarative programming language to describe nucleic acids' secondary structures and to scan sequence database. *Nucleic Acids Res* **24**, 1395-403.
- Björk, G. R. (1995). Biosynthesis and function of modified nucleosides. In "tRNA: Structure, biosynthesis and function" (D. Söll et U. RajBhandary, Eds.), pp. 165-206. ASM Press, Washington.
- Bloch, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H. W., et Stetter, K. O. (1997). *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles* **1**, 14-21.
- Blum, B., Bakalara, N., et Simpson, L. (1990). A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell* **60**, 189-98.
- Boch, J., Kempf, B., et Bremer, E. (1994). Osmoregulation in *Bacillus subtilis*: synthesis of the osmoprotectant glycine betaine from exogenously provided choline. *J Bacteriol* **176**, 5364-71.
- Boelens, R., et Gualerzi, C. O. (2002). Structure and function of bacterial initiation factors. *Curr Protein Pept Sci* **3**, 107-19.
- Bolhuis, H., Poele, E. M., et Rodriguez-Valera, F. (2004). Isolation and cultivation of Walsby's square archaeon. *Environ Microbiol* **6**, 1287-91.
- Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodriguez-Valera, F., Pfeiffer, F., et Oesterhelt, D. (2006). The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. *BMC Genomics* **7**, 169.
- Bolt, E. L. (2005). Helicases that interact with replication forks: new candidates from archaea. *Biochem Soc Trans* **33**, 1471-3.
- Bono, H., Nikaido, I., Kasukawa, T., Hayashizaki, Y., et Okazaki, Y. (2003). Comprehensive analysis of the mouse metabolome based on the transcriptome. *Genome Res* **13**, 1345-9.
- Bortolin, M. L., Ganot, P., et Kiss, T. (1999). Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *Embo J* **18**, 457-69.
- Bortolin, M. L., Bachellerie, J. P., et Clouet-d'Orval, B. (2003). *In vitro* RNP assembly and methylation guide activity of an unusual box C/D RNA, cis-acting archaeal pre-tRNA(Trp). *Nucl. Acids Res.* **31**, 6524-6535.
- Boucher, Y., et Doolittle, W. F. (2002). Something new under the sea. *Nature* **417**, 27-8.

- Bourdeau, V., Ferbeyre, G., Pageau, M., Paquin, B., et Cedergren, R. (1999). The distribution of RNA motifs in natural sequences. *Nucleic Acids Res* **27**, 4457-67.
- Bousquet-Antonelli, C., Henry, Y., Gelugne, J. P., Caizergues-Ferrer, M., et Kiss, T. (1997). A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs. *EMBO J* **16**, 4770-4776.
- Branlant, C., Krol, A., Machatt, A., et Ebel, J. P. (1981a). The secondary structure of the protein L1 binding region of ribosomal 23S RNA. Homologies with putative secondary structures of the L11 mRNA and of a region of mitochondrial 16S rRNA. *Nucleic Acids Res* **9**, 293-307.
- Branlant, C., Krol, A., Machatt, M. A., Pouyet, J., Ebel, J. P., Edwards, K., et Kossel, H. (1981b). Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. *Nucleic Acids Res* **9**, 4303-24.
- Branlant, C., Krol, A., Machatt, M. A., Pouyet, J., Ebel, J. P., Edwards, K., et Kossel, H. (1981c). Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. *Nucleic Acids Res* **9**, 4303-24.
- Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol* **10**, 102-9.
- Bredberg, K., Persson, J., Christiansson, M., Stenberg, B., et Holst, O. (2001). Anaerobic desulfurization of ground rubber with the thermophilic archaeon *Pyrococcus furiosus*--a new method for rubber recycling. *Appl Microbiol Biotechnol* **55**, 43-8.
- Brennan, R. G., et Link, T. M. (2007). Hfq structure, function and ligand binding. *Curr Opin Microbiol* **10**, 125-33.
- Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B., et Cohen, S. M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**, 25-36.
- Brennecke, J., Stark, A., Russell, R. B., et Cohen, S. M. (2005). Principles of microRNA-target recognition. *PLoS Biol* **3**, e85.
- Brochier, C., Forterre, P., et Gribaldo, S. (2004). Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the *Methanopyrus kandleri* paradox. *Genome Biol* **5**, R17.
- Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F., et Forterre, P. (2005). Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biol* **6**, R42.
- Brock, T. D., Brock, K. M., Belly, R. T., et Weiss, R. L. (1972). *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch Mikrobiol* **84**, 54-68.
- Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., et Rastan, S. (1992). The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515-26.
- Brockdorff, N. (2002). X-chromosome inactivation: closing in on proteins that bind *Xist* RNA. *Trends Genet* **18**, 352-8.
- Brosius, J., et Tiedge, H. (2004). RNomenclature. *RNA Biol* **1**, 81-3.
- Brown, J. W., Clark, G. P., Leader, D. J., Simpson, C. G., et Lowe, T. (2001). Multiple snoRNA gene clusters from *Arabidopsis*. *RNA* **7**, 1817-32.
- Brugger, K., Chen, L., Stark, M., Zibat, A., Redder, P., Ruepp, A., Awayez, M., She, Q., Garrett, R. A., et Klenk, H. P. (2006). The genome of *Hyperthermus butylicus*: a sulfur-reducing, peptide fermenting, neutrophilic Crenarchaeote growing up to 108 degrees C. *Archaea* **2**, 127-35.
- Brule, H., Holmes, W. M., Keith, G., Giege, R., et Florentz, C. (1998). Effect of a mutation in the anticodon of human mitochondrial tRNA^{Pro} on its post-transcriptional modification pattern. *Nucleic Acids Res* **26**, 537-43.
- Bult, C., White, O., Olsen, G., Zhou, L., Fleischmann, R., Sutton, G., Blake, J., FitzGerald, L., Clayton, R., Gocayne, J., et al. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058-1073.

C

- Caffarelli, E., Fatica, A., Prislei, S., De Gregorio, E., Fragapane, P., et Bozzoni, I. (1996). Processing of the intron-encoded U16 and U18 snoRNAs: the conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA. *Embo J* **15**, 1121-31.
- Cahill, N. M., Friend, K., Speckmann, W., Li, Z. H., Terns, R. M., Terns, M. P., et Steitz, J. A. (2002). Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J* **21**, 3816-3828.
- Cannio, R., Contursi, P., Rossi, M., et Bartolucci, S. (1998). An autonomously replicating transforming vector for *Sulfolobus solfataricus*. *J Bacteriol* **180**, 3237-40.
- Carbone, M. L., Solinas, M., Sora, S., et Panzeri, L. (1991). A gene tightly linked to CEN6 is important for growth of *Saccharomyces cerevisiae*. *Curr Genet* **19**, 1-8.
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-63.
- Carrington, J. C., et Ambros, V. (2003). Role of microRNAs in plant and animal development. *Science* **301**, 336-8.

- Carter, R. J., Dubchak, I., et Holbrook, S. R. (2001). A computational approach to identify genes for functional RNAs in genomic sequences. *Nucleic Acids Res* **29**, 3928-38.
- Cavaillé, J., Nicoloso, M., et Bachellerie, J.-P. (1996). Targeted ribose methylation of RNA *in vivo* directed by tailored antisense RNA guides. *Nature* **383**, 732-735.
- Cavaille, J., Buiting, K., Kieffmann, M., Lalande, M., Brannan, C. I., Horsthemke, B., Bachellerie, J. P., Brosius, J., et Huttenhofer, A. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. U S A*, **97**, 14311-14316.
- Cavaille, J., Seitz, H., Paulsen, M., Ferguson-Smith, A. C., et Bachellerie, J. P. (2002). Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum. Mol. Genet.* **11**, 1527-1538.
- Cavalier-Smith, T. (2006). Rooting the tree of life by transition analyses. *Biol Direct* **1**, 19.
- Cavicchioli, R., Siddiqui, K. S., Andrews, D., et Sowers, K. R. (2002). Low-temperature extremophiles and their applications. *Curr Opin Biotechnol* **13**, 253-61.
- Cavicchioli, R., Curmi, P. M., Saunders, N., et Thomas, T. (2003). Pathogenic archaea: do they exist? *Bioessays* **25**, 1119-28.
- Cavicchioli, R. (2006). Cold-adapted archaea. *Nat Rev Microbiol* **4**, 331-43.
- Ceci, M., Gaviraghi, C., Gorrini, C., Sala, L. A., Offenhauser, N., Marchisio, P. C., et Biffo, S. (2003). Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. *Nature* **426**, 579-84.
- Chaban, B., Ng, S. Y., et Jarrell, K. F. (2006). Archaeal habitats--from the extreme to the ordinary. *Can J Microbiol* **52**, 73-116.
- Chanfreau, G., Rotondo, G., Legrain, P., et Jacquier, A. (1998). Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *Embo J* **17**, 3726-37.
- Charette, M., et Gray, M. W. (2000). Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* **49**, 341-51.
- Charpentier, B., Muller, S., et Branlant, C. (2005). Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res* **33**, 3133-44.
- Charpentier, B., Fourmann, J. B., et Branlant, C. (2007). Reconstitution of archaeal H/ACA sRNPs and test of their activity. *Methods Enzymol* **425**, 389-405.
- Charron, C., Manival, X., Clery, A., Senty-Segault, V., Charpentier, B., Marmier-Gourrier, N., Branlant, C., et Aubry, A. (2004). The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J Mol Biol* **342**, 757-73.
- Chaudhuri, B. N., Chan, S., Perry, L. J., et Yeates, T. O. (2004). Crystal structure of the apo forms of psi 55 tRNA pseudouridine synthase from *Mycobacterium tuberculosis*: a hinge at the base of the catalytic cleft. *J Biol Chem* **279**, 24585-91.
- Chen, C. L., Perasso, R., Qu, L. H., et Amar, L. (2007). Exploration of pairing constraints identifies a 9 base-pair core within box C/D snoRNA-rRNA duplexes. *J Mol Biol* **369**, 771-83.
- Chen, J., et Patton, J. R. (1999). Cloning and characterization of a mammalian pseudouridine synthase. *RNA* **5**, 409-419.
- Chen, L., Brugger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H. P., et al. (2005). The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. *J Bacteriol* **187**, 4992-9.
- Chen, S., Lesnik, E. A., Hall, T. A., Sampath, R., Griffey, R. H., Ecker, D. J., et Blyn, L. B. (2002). A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *Biosystems* **65**, 157-77.
- Chen, Y., Sierzputowska-Gracz, H., Guenther, R., Everett, K., et Agris, P. F. (1993). 5-Methylcytidine is required for cooperative binding of Mg²⁺ and a conformational transition at the anticodon stem-loop of yeast phenylalanine tRNA. *Biochemistry* **32**, 10249-53.
- Cheng, X., et Roberts, R. J. (2001). AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res* **29**, 3784-95.
- Cherry, R., Desbruyeres, D., Heyraud, M., et Nolan, C. (1992). High levels of natural radioactivity in hydrothermal vent polychaetes. *C.R. Acad. Sci. Paris Ser. III* **315**, 21-26.
- Choquet, C. G., Patel, G. B., Beveridge, T. J., et Sprott, G. D. (1992). Formation of unilamellar liposomes from total polar lipid extracts of methanogens. *Appl Environ Microbiol* **58**, 2894-900.
- Chouari, R., Le Paslier, D., Daegelen, P., Ginestet, P., Weissenbach, J., et Sghir, A. (2005). Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. *Environ Microbiol* **7**, 1104-15.
- Christensen, S. K., et Gerdes, K. (2003). RelE toxins from bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. *Mol Microbiol* **48**, 1389-400.
- Ciampi, M. S., Arena, F., et Cortese, R. (1977). Biosynthesis of pseudouridine in the *in vitro* transcribed tRNA^{Tyr} precursor. *FEBS Lett* **77**, 75-82.
- Claesson, C., Lustig, F., Boren, T., Simonsson, C., Barciszewska, M., et Lagerkvist, U. (1995). Glycine codon discrimination and the nucleotide in position 32 of the anticodon loop. *J. Mol. Biol* **247**, 191-6.
- Cliften, P. F., Hillier, L. W., Fulton, L., Graves, T., Miner, T., Gish, W. R., Waterston, R. H., et Johnston, M. (2001). Surveying *Saccharomyces* genomes to identify functional elements by comparative DNA sequence analysis. *Genome Res* **11**, 1175-86.

- Clouet d'Orval, B., Bortolin, M. L., Gaspin, C., et Bachellerie, J. P. (2001). Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res* **29**, 4518-29.
- Cohn, W. E. (1951). Some results of the applications of ion-exchange chromatography to nucleic acid chemistry. *J Cell Physiol Suppl* **38**, 21-40.
- Condo, I., Ciammaruconi, A., Benelli, D., Ruggero, D., et Londei, P. (1999). Cis-acting signals controlling translational initiation in the thermophilic archaeon *Sulfolobus solfataricus*. *Mol Microbiol* **34**, 377-84.
- Connolly, B. A., Fogg, M. J., Shuttleworth, G., et Wilson, B. T. (2003). Uracil recognition by archaeal family B DNA polymerases. *Biochem Soc Trans* **31**, 699-702.
- Conrad, J., Sun, D., Englund, N., et Ofengand, J. (1998). The *rluC* gene of *Escherichia coli* codes for a pseudouridine synthase that is solely responsible for synthesis of pseudouridine at positions 955, 2504, and 2580 in 23 S ribosomal RNA. *J Biol Chem* **273**, 18562-18566.
- Conrad, J., Niu, L., Rudd, K., Lane, B. G., et Ofengand, J. (1999). 16S ribosomal RNA pseudouridine synthase RsuA of *Escherichia coli*: Deletion, mutation of the conserved Asp102 residue, and sequence comparison among all other pseudouridine synthases. *RNA* **5**, 761-763.
- Consortium, I. H. G. S. (2004). Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45.
- Cortese, R., Kammen, H. O., Spengler, S. J., et Ames, B. N. (1974). Biosynthesis of pseudouridine in transfer ribonucleic acid. *J Biol Chem* **249**, 1103-1108.
- Costa, F. F. (2005). Non-coding RNAs: new players in eukaryotic biology. *Gene* **357**, 83-94.
- Costa, F. (2007). Non-coding RNAs: lost in translation? *Gene* **386**, 1-10.
- Coventry, A., Kleitman, D. J., et Berger, B. (2004). MSARI: multiple sequence alignments for statistical detection of RNA secondary structure. *Proc Natl Acad Sci U S A* **101**, 12102-7.
- Coverley, D., et Laskey, R. A. (1994). Regulation of eukaryotic DNA replication. *Annu Rev Biochem* **63**, 745-76.
- Cubeddu, L., et White, M. F. (2005). DNA damage detection by an archaeal single-stranded DNA-binding protein. *J Mol Biol* **353**, 507-16.
- Cubonova, L., Sandman, K., Hallam, S. J., Delong, E. F., et Reeve, J. N. (2005). Histones in crenarchaea. *J Bacteriol* **187**, 5482-5.
- Curnow, A. W., Kung, F. L., Koch, K. A., et Garcia, G. A. (1993). tRNA-guanine transglycosylase from *Escherichia coli*: gross tRNA structural requirements for recognition. *Biochemistry* **32**, 5239-46.

D

- Darzacq, X., et Kiss, T. (2000). Processing of intron-encoded box C/D small nucleolar RNAs lacking a 5',3'-terminal stem structure. *Mol. Cell. Biol.* **20**, 4522-4531.
- Darzacq, X., Jady, B. E., Verheggen, C., Kiss, A. M., Bertrand, E., et Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O- methylation and pseudouridylation guide RNAs. *EMBO J.* **21**, 2746-2756.
- Darzacq, X., Kittur, N., Roy, S., Shav-Tal, Y., Singer, R. H., et Meier, U. T. (2006). Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J Cell Biol* **173**, 207-18.
- Das, S., Paul, S., Bag, S. K., et Dutta, C. (2006). Analysis of Nanoarchaeum equitans genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. *BMC Genomics* **7**, 186.
- Davis, D. R. (1998). Biophysical and Conformational Properties of Modified Nucleosides in RNA (Nuclear Magnetic Resonance Studies. *Eds H. Grosjean and R. Benne Modification and Editing of RNA*, 85-102.
- Davis, D. R., Veltri, C. A., et Nielsen, L. (1998). An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNALys, tRNAHis and tRNATyr. *J Biomol Struct Dyn* **15**, 1121-32.
- Davis, F. F., et Allen, F. W. (1957). Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* **227**, 907-15.
- de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* **19**, 2100-10.
- de Lange, T. (2006). Lasker Laurels for telomerase. *Cell* **126**, 1017-20.
- De Rosa, M., Morana, A., Riccio, A., Gambacorta, A., Trincone, A., et Incani, O. (1994). Lipids of the archaea: a new tool for bioelectronics. *Biosens Bioelectron* **9**, 669-675.
- Decatur, W. A., et Fournier, M. J. (2002). rRNA modifications and ribosome function. *Trends Biochem Sci* **27**, 344-51.
- Del Campo, M., Kaya, Y., et Ofengand, J. (2001). Identification and site of action of the remaining four putative pseudouridine synthases in *Escherichia coli*. *RNA* **7**, 1603-15.
- Del Campo, M., Ofengand, J., et Malhotra, A. (2004). Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of *Escherichia coli*. *RNA* **10**, 231-9.
- Del Campo, M., Recinos, C., Yanez, G., Pomerantz, S. C., Guymon, R., Crain, P. F., McCloskey, J. A., et Ofengand, J. (2005). Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of *Haloarcula marismortui* and *Deinococcus radiodurans*. *RNA* **11**, 210-9.
- Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, O., et Salzberg, S. L. (1999). Alignment of whole genomes. *Nucleic Acids Res* **27**, 2369-76.

- DeLong, E. F., Wu, K. Y., Prezelin, B. B., et Jovine, R. V. (1994). High abundance of Archaea in Antarctic marine picoplankton. *Nature* **371**, 695-7.
- DeLong, E. F., et Pace, N. R. (2001). Environmental diversity of bacteria and archaea. *Syst Biol* **50**, 470-8.
- DeLong, E. F. (2005). Microbial community genomics in the ocean. *Nat Rev Microbiol* **3**, 459-69.
- Deng, L., Starostina, N. G., Liu, Z. J., Rose, J. P., Terns, R. M., Terns, M. P., et Wang, B. C. (2004). Structure determination of fibrillarlin from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biochem Biophys Res Commun* **315**, 726-32.
- Dennis, P. P. (1997). Ancient ciphers: translation in Archaea. *Cell* **89**, 1007-10.
- Dennis, P. P., Omer, A., et Lowe, T. (2001). A guided tour: small RNA function in Archaea. *Mol Microbiol* **40**, 509-19.
- Dennis, P. P., et Omer, A. (2005). Small non-coding RNAs in Archaea. *Curr Opin Microbiol* **8**, 685-94.
- Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A., Martinez-Arias, R., Henne, A., Wiezer, A., Baumer, S., Jacobi, C., et al. (2002). The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *J Mol Microbiol Biotechnol* **4**, 453-61.
- DeRosa, M., et Gambacorta, A. (1988). The lipids of archaeobacteria. *Prog. Lipid Res.* **27**, 153-175.
- DeRosa, M., Trincone, A., Nicolaus, B., et Gambacorta, A. (1991). Archaeobacteria: lipids, membrane structures, and adaptations to environmental stresses. *di Prisco G. (Ed) Life Under Extreme Conditions, Springer-Verlag, Berlin Heidelberg*, 61-87.
- Derrick, W. B., et Horowitz, J. (1993). Probing structural differences between native and in vitro transcribed *Escherichia coli* valine transfer RNA: evidence for stable base modification-dependent conformers. *Nucleic Acids Res* **21**, 4948-53.
- Dethlefsen, L., Eckburg, P. B., Bik, E. M., et Relman, D. A. (2006). Assembly of the human intestinal microbiota. *Trends Ecol Evol* **21**, 517-23.
- Dez, C., Noaillac-Depeyre, J., Caizergues-Ferrer, M., et Henry, Y. (2002). Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs. *Mol Cell Biol* **22**, 7053-65.
- Di BeRNArdo, D., Gardner, T. S., et Collins, J. J. (2004). Robust identification of large genetic networks. *Pac Symp Biocomput*, 486-97.
- Dijk, J., et Reinhardt, R. (1986). The structure of DNA-binding proteins from eu- and archaeobacteria. *Bacterial Chromatin.*, 185-218.
- Dokudovskaya, S., Dontsova, O., Shpanchenko, O., Bogdanov, A., et Brimacombe, R. (1996). Loop IV of 5S ribosomal RNA has contacts both to domain II and to domain V of the 23S RNA. *RNA* **2**, 146-52.
- Dopson, M., Baker-Austin, C., Hind, A., Bowman, J. P., et Bond, P. L. (2004). Characterization of *Ferroplasma* isolates and *Ferroplasma acidarmanus* sp. nov., extreme acidophiles from acid mine drainage and industrial bioleaching environments. *Appl Environ Microbiol* **70**, 2079-88.
- Dore, J., Pochet, P., BeRNAlie, A., Goderel, I., Morvan, B., et Rambaud, J. (1995). Enumeration of H₂ utilizing methanogenic archaea, acetogenic and sulfate-reducing bacteria from human feces. *FEMS Microbiol. Ecol.* **17**, 279-284.
- Droogmans, L., Haumont, E., de Henau, S., et Grosjean, H. (1986). Enzymatic 2'-O-methylation of the wobble nucleoside of eukaryotic tRNA^{Phe}: specificity depends on structural elements outside the anticodon loop. *Embo J* **5**, 1105-9.
- Dsouza, M., Larsen, N., et Overbeek, R. (1997). Searching for patterns in genomic data. *Trends Genet* **13**, 497-8.
- Dubchak, I., Brudno, M., Loots, G. G., Pachter, L., Mayor, C., Rubin, E. M., et Frazer, K. A. (2000). Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res* **10**, 1304-6.

E

- Eckburg, P. B., Lepp, P. W., et Relman, D. A. (2003). Archaea and their potential role in human disease. *Infect Immun* **71**, 591-6.
- Eddy, S. R. (1999). Noncoding RNA genes. *Curr Opin Genet Dev* **9**, 695-9.
- Eddy, S. R. (2001). Non coding RNA genes. *Nature reviews genetics* **2**, 919-929.
- Eddy, S. R. (2002). Computational genomics of noncoding RNA genes. *Cell* **109**, 137-40.
- Eddy, S. R. (2005). A model of the statistical power of comparative genome sequence analysis. *PLoS Biol* **3**, e10.
- Edqvist, J., Grosjean, H., et Sträby, K. B. (1992). Identity elements for N²-dimethylation of guanosine-26 in yeast tRNAs. *Nucleic Acids Res* **20**, 6575-81.
- Edqvist, J., Blomqvist, K., et Sträby, K. B. (1994). Structural elements in yeast tRNAs required for homologous modification of guanosine-26 into dimethylguanosine-26 by the yeast *Trm1* tRNA-modifying enzyme. *Biochemistry* **33**, 9546-9551.
- Edqvist, J., Sträby, K. B., et Grosjean, H. (1993). Pleiotrophic effects of point mutations in yeast tRNA^{Asp} on the base modification pattern. *Nucleic Acids Res* **21**, 413-417.
- Egorova, K., et Antranikian, G. (2005). Industrial relevance of thermophilic Archaea. *Curr Opin Microbiol* **8**, 649-55.
- Ehresmann, C., Stiegler, P., Carbon, P., et Ebel, J. P. (1977). Recent progress in the determination of the primary sequence of the 16 S RNA of *Escherichia coli*. *FEBS Lett* **84**, 337-41.

- Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., et Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494-8.
- Elferink, M. G., de Wit, J. G., Demel, R., Driessen, A. J., et Konings, W. N. (1992). Functional reconstitution of membrane proteins in monolayer liposomes from bipolar lipids of *Sulfolobus acidocaldarius*. *J Biol Chem* **267**, 1375-81.
- Eliceiri, G. L. (1999). Small nucleolar RNAs. *Cell Mol Life Sci* **56**, 22-31.
- Erdmann, V. A., Barciszewska, M. Z., Hochberg, A., de Groot, N., et Barciszewski, J. (2001). Regulatory RNAs. *Cell Mol Life Sci* **58**, 960-77.
- Ericsson, U. B., Nordlund, P., et Hallberg, B. M. (2004). X-ray structure of tRNA pseudouridine synthase TruD reveals an inserted domain with a novel fold. *FEBS Lett* **565**, 59-64.
- Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O., et Salzberg, S. L. (2000). Prediction of transcription terminators in bacterial genomes. *J Mol Biol* **301**, 27-33.
- Erzberger, J. P., Pirruccello, M. M., et Berger, J. M. (2002). The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *Embo J* **21**, 4763-73.
- Eshete, M., Marchbank, M. T., Deutscher, S. L., Sproat, B., Leszczynska, G., Malkiewicz, A., et Agris, P. F. (2007). Specificity of phage display selected peptides for modified anticodon stem and loop domains of tRNA. *Protein J* **26**, 61-73.

F

- Falb, M., Pfeiffer, F., Palm, P., Rodewald, K., Hickmann, V., Tittor, J., et Oesterhelt, D. (2005). Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis*. *Genome Res* **15**, 1336-43.
- Fatica, A., Dlakic, M., et Tollervey, D. (2002). Naf1 p is a box H/ACA snoRNP assembly factor. *RNA* **8**, 1502-14.
- Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401-410.
- Ferre-D'Amare, A. R. (2003). RNA-modifying enzymes. *Curr Opin Struct Biol* **13**, 49-55.
- Filipowicz, W., et Pogacic, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.* **14**, 319-327.
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., et Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-11.
- Fitz-Gibbon, S. T., Ladner, H., Kim, U. J., Stetter, K. O., Simon, M. I., et Miller, J. H. (2002). Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Proc Natl Acad Sci U S A* **99**, 984-9.
- Folichon, M., Arluison, V., Pellegrini, O., Huntzinger, E., Regnier, P., et Hajnsdorf, E. (2003). The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res* **31**, 7302-10.
- Forbes, A. J., Patrie, S. M., Taylor, G. K., Kim, Y. B., Jiang, L., et Kelleher, N. L. (2004). Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS. *Proc Natl Acad Sci U S A* **101**, 2678-83.
- Forterre, P., Bouthier De La Tour, C., Philippe, H., et Duguet, M. (2000). Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria. *Trends Genet* **16**, 152-4.
- Forterre, P., Brochier, C., et Philippe, H. (2002). Evolution of the Archaea. *Theor Popul Biol* **61**, 409-22.
- Forterre, P. (2002). A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet* **18**, 236-7.
- Forterre, P. (2005). The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie* **87**, 793-803.
- Forterre, P. (2006a). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res* **117**, 5-16.
- Forterre, P. (2006b). Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc Natl Acad Sci U S A* **103**, 3669-74.
- Foster, P. G., Huang, L., Santi, D. V., et Stroud, R. M. (2000). The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I. *Nat Struct Biol* **7**, 23-7.
- Franzmann, P. D., Liu, Y., Balkwill, D. L., Aldrich, H. C., Conway de Macario, E., et Boone, D. R. (1997). *Methanogenium frigidum* sp. nov., a psychrophilic, H₂-using methanogen from Ace Lake, Antarctica. *Int J Syst Bacteriol* **47**, 1068-72.
- Franzmann, P. D., Springer, N., Ludwig, W., Conway De Macario, E., et Rohde, M. (1992). A methanogenic archaeon from Ace Lake, Antarctica: *Methanococcoides burtonii* sp. nov. *Systematic and Applied Microbiology* **15**, 573-581.
- Frazer, K. A., Elnitski, L., Church, D. M., Dubchak, I., et Hardison, R. C. (2003). Cross-species sequence comparisons: a review of methods and available resources. *Genome Res* **13**, 1-12.
- French, S. L., Santangelo, T. J., Beyer, A. L., et Reeve, J. N. (2007). Transcription and Translation are Coupled in Archaea. *Mol Biol Evol.*

- Fricke, W. F., Seedorf, H., Henne, A., Kruer, M., Liesegang, H., Hedderich, R., Gottschalk, G., et Thauer, R. K. (2006). The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H₂ for methane formation and ATP synthesis. *J Bacteriol* **188**, 642-58.
- Frith, M. C., Pheasant, M., et Mattick, J. S. (2005). The amazing complexity of the human transcriptome. *Eur J Hum Genet* **13**, 894-7.
- Fujikane, R., Komori, K., Shinagawa, H., et Ishino, Y. (2005). Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, *Pyrococcus furiosus*. *J Biol Chem* **280**, 12351-8.
- Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S., et Imanaka, T. (2005). Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res* **15**, 352-63.
- Furuno, M., Pang, K. C., Ninomiya, N., Fukuda, S., Frith, M. C., Bult, C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., et al. (2006). Clusters of inteRNally primed transcripts reveal novel long noncoding RNAs. *PLoS Genet* **2**, e37.
- Futterer, O., Angelov, A., Liesegang, H., Gottschalk, G., Schleper, C., Schepers, B., Dock, C., Antranikian, G., et Liebl, W. (2004). Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proc Natl Acad Sci U S A* **101**, 9091-6.

G

- Gabory, A., Ripoche, M. A., Yoshimizu, T., et Dandolo, L. (2006). The H19 gene: regulation and function of a non-coding RNA. *Cytogenet Genome Res* **113**, 188-93.
- Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., et al. (2002). The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res* **12**, 532-42.
- Ganot, P., Bortolin, M. L., et Kiss, T. (1997a). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, **89**, 799-809.
- Ganot, P., Caizergues-Ferrer, M., et Kiss, T. (1997b). The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev* **11**, 941-56.
- Garcia, G. A., et Goodenough-Lashua, D. M. (1998). Mechanisms of RNA-Modifying and -Editing Enzymes. In "The modification and editing of RNA" (H. Grosjean et R. Benne, Eds.), pp. 135-68. ASM Press, Washington, DC.
- Garcia, J. L., Patel, B. K., et Ollivier, B. (2000). Taxonomic, phylogenetic, and ecological diversity of methanogenic archaea. *Anaerobe* **6**, 205-26.
- Gardner, W. L., et Whitman, W. B. (1999). Expression vectors for *Methanococcus maripaludis*: overexpression of acetohydroxyacid synthase and beta-galactosidase. *Genetics* **152**, 1439-47.
- Gaspin, C., Cavaille, J., Erauso, G., et Bachellerie, J. P. (2000). Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J Mol Biol* **297**, 895-906.
- Gautheret, D., Major, F., et Cedergren, R. (1993). Modeling the three-dimensional structure of RNA using discrete nucleotide conformational sets. *J Mol Biol* **229**, 1049-64.
- Gautheret, D., et Lambert, A. (2001). Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J Mol Biol* **313**, 1003-11.
- Geiduschek, E. P., et Ouhammouch, M. (2005). Archaeal transcription and its regulators. *Mol Microbiol* **56**, 1397-407.
- Geraghty, D. S., Ding, M., Heintz, N. H., et Pederson, D. S. (2000). Premature structural changes at replication origins in a yeast minichromosome maintenance (MCM) mutant. *J Biol Chem* **275**, 18011-21.
- Gerdes, K., Gultyaev, A. P., Franch, T., Pedersen, K., et Mikkelsen, N. D. (1997). Antisense RNA-regulated programmed cell death. *Annu Rev Genet* **31**, 1-31.
- Gerdes, K., et Wagner, E. G. (2007). RNA antitoxins. *Curr Opin Microbiol* **10**, 117-24.
- Gernhardt, P., Possot, O., Foglino, M., Sibold, L., et Klein, A. (1990). Construction of an integration vector for use in the archaeobacterium *Methanococcus voltae* and expression of a eubacterial resistance gene. *Mol Gen Genet* **221**, 273-9.
- Gilbert, W. (1986). The RNA World. *Nature* **319**, 618.
- Gillet, R., et Felden, B. (2001). Emerging views on tmRNA-mediated protein tagging and ribosome rescue. *Mol Microbiol* **42**, 879-85.
- Ginger, M. R., Shore, A. N., Contreras, A., Rijnkels, M., Miller, J., Gonzalez-Rimbau, M. F., et Rosen, J. M. (2006). A noncoding RNA is a potential marker of cell fate during mammary gland development. *Proc Natl Acad Sci U S A* **103**, 5781-6.
- Ginisty, H., Sicard, H., Roger, B., et Bouvet, P. (1999). Structure and functions of nucleolin. *J Cell Sci* **112** (Pt 6), 761-72.
- Girard, A., Sachidanandam, R., Hannon, G. J., et Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199-202.

- Girard, J. P., Bagni, C., Caizergues-Ferrer, M., Amalric, F., et Lapeyre, B. (1994). Identification of a segment of the small nucleolar ribonucleoprotein- associated protein GAR1 that is sufficient for nucleolar accumulation. *J Biol Chem* **269**, 18499-506.
- Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L., et Hilario, E. (2002). Inteins: structure, function, and evolution. *Annu Rev Microbiol* **56**, 263-87.
- Goldberg, A. D., Allis, C. D., et Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell* **128**, 635-8.
- Goo, Y. A., Roach, J., Glusman, G., Baliga, N. S., Deutsch, K., Pan, M., Kennedy, S., DasSarma, S., Ng, W. V., et Hood, L. (2004). Low-pass sequencing for microbial comparative genomics. *BMC Genomics* **5**, 3.
- Gordeliy, V. I., Labahn, J., Moukhametzianov, R., Efremov, R., Granzin, J., Schlesinger, R., Buldt, G., Savopol, T., Scheidig, A. J., Klare, J. P., et al. (2002). Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* **419**, 484-7.
- Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Luhrmann, R., et Fabrizio, P. (1999). Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *Embo J* **18**, 4535-4548.
- Grabowski, B., et Kelman, Z. (2001). Autophosphorylation of archaeal Cdc6 homologues is regulated by DNA. *J Bacteriol* **183**, 5459-64.
- Green, C. J., Kammen, H. O., et Penhoet, E. E. (1982). Purification and properties of a mammalian tRNA pseudouridine synthase. *J Biol Chem* **257**, 3045-3052.
- Green, R., et Noller, H. F. (1996). In vitro complementation analysis localizes 23S rRNA posttranscriptional modifications that are required for Escherichia coli 50S ribosomal subunit assembly and function. *RNA* **2**, 1011-21.
- Green, R., et Noller, H. F. (1999). Reconstitution of functional 50S ribosomes from in vitro transcripts of Bacillus stearothermophilus 23S rRNA. *Biochemistry* **38**, 1772-9.
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., et Eddy, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Res* **31**, 439-41.
- Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., et Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33**, D121-4.
- Grigoriev, A. (1998). Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res* **26**, 2286-90.
- Grogan, D. W., Carver, G. T., et Drake, J. W. (2001). Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon Sulfolobus acidocaldarius. *Proc Natl Acad Sci U S A* **98**, 7928-33.
- Grogan, D. W. (2004). Stability and repair of DNA in hyperthermophilic Archaea. *Curr Issues Mol Biol* **6**, 137-44.
- Grosjean, H., Sprinzl, M., et Steinberg, S. (1995). Posttranscriptionally modified nucleosides in transfer RNA: Their locations and frequencies. *Biochimie* **77**, 139-141.
- Grosjean, H., Edqvist, J., Sträby, K. B., et Giegé, R. (1996). Enzymatic formation of modified nucleosides in tRNA: Dependence on tRNA architecture. *JMB* **255**, 67-85.
- Grosjean, H., Szweykowska-Kulinska, Z., Motorin, Y., Fasiolo, F., et Simos, G. (1997). Intron-dependent enzymatic formation of modified nucleosides in eukaryotic tRNAs : A review. *Biochimie* **79**, 293-302.
- Gu, X., Yu, M., Ivanetich, K. M., et Santi, D. V. (1998). Molecular recognition of tRNA by tRNA pseudouridine 55 synthase. *Biochemistry* **37**, 339-343.
- Gu, X., Liu, Y., et Santi, D. V. (1999). The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA. *Proc Natl Acad Sci U S A* **96**, 14270-5.
- Guo, R., Xue, H., et Huang, L. (2003). Ssh10b, a conserved thermophilic archaeal protein, binds RNA in vivo. *Mol Microbiol* **50**, 1605-15.
- Gupta, R. (1984). Halobacterium volcanii tRNAs. Identification of 41 tRNAs covering all amino acids, and the sequences of 33 class I tRNAs. *JouRNAI of Biological Chemistry* **259**, 9461-9471.
- Gutgsell, N. S., Del Campo, M. D., Raychaudhuri, S., et Ofengand, J. (2001). A second function for pseudouridine synthases: A point mutant of RluD unable to form pseudouridines 1911, 1915, and 1917 in Escherichia coli 23S ribosomal RNA restores normal growth to an RluD-minus strain. *RNA* **7**, 990-8.
- Guy, C. P., et Bolt, E. L. (2005). Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. *Nucleic Acids Res* **33**, 3678-90.

H

- Hallam, S. J., Putnam, N., Preston, C. M., Detter, J. C., Rokhsar, D., Richardson, P. M., et DeLong, E. F. (2004). Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**, 1457-62.
- Hallam, S. J., Konstantinidis, K. T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., Preston, C., de la Torre, J., Richardson, P. M., et DeLong, E. F. (2006). Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. *Proc Natl Acad Sci U S A* **103**, 18296-301.
- Hamilton, C. S., Greco, T. M., Vizthum, C. A., Ginter, J. M., Johnston, M. V., et Mueller, E. G. (2006). Mechanistic investigations of the pseudouridine synthase RluA using RNA containing 5-fluorouridine. *Biochemistry* **45**, 12029-38.

- Hamma, T., et Ferre-D'Amare, A. R. (2004). Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. *Structure (Camb)* **12**, 893-903.
- Hamma, T., Reichow, S. L., Varani, G., et Ferre-D'Amare, A. R. (2005). The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat Struct Mol Biol* **12**, 1101-7.
- Hamma, T., et Ferre-D'Amare, A. R. (2006). Pseudouridine synthases. *Chem Biol* **13**, 1125-35.
- Hanzelka, B. L., Darcy, T. J., et Reeve, J. N. (2001). TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIIE α . *J Bacteriol* **183**, 1813-8.
- Hardin, J. W., et Batey, R. T. (2006). The bipartite architecture of the sRNA in an archaeal box C/D complex is a primary determinant of specificity. *Nucleic Acids Res* **34**, 5039-51.
- Haring, M., Vestergaard, G., Rachel, R., Chen, L., Garrett, R. A., et Prangishvili, D. (2005). Virology: independent virus development outside a host. *Nature* **436**, 1101-2.
- Haseltine, C., Hill, T., Montalvo-Rodriguez, R., Kemper, S. K., Shand, R. F., et Blum, P. (2001). Secreted euryarchaeal microhalocins kill hyperthermophilic crenarchaea. *J Bacteriol* **183**, 287-91.
- Hastings, M. L., Milcarek, C., Martincic, K., Peterson, M. L., et Munroe, S. H. (1997). Expression of the thyroid hormone receptor gene, *erbA* α , in B lymphocytes: alterNAtive mRNA processing is independent of differentiation but correlates with antisense RNA levels. *Nucleic Acids Res* **25**, 4296-300.
- Hausner, W., Lange, U., et Musfeldt, M. (2000). Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J Biol Chem* **275**, 12393-9.
- Hausner, W., Wettach, J., Hethke, C., et Thomm, M. (1996). Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. *J Biol Chem* **271**, 30144-8.
- Heard, E. (2004). Recent advances in X-chromosome inactivation. *Curr Opin Cell Biol* **16**, 247-55.
- Hendrickson, E. L., Kaul, R., Zhou, Y., Bovee, D., Chapman, P., Chung, J., Conway de Macario, E., Dodsworth, J. A., Gillett, W., Graham, D. E., et al. (2004). Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. *J Bacteriol* **186**, 6956-69.
- Henneke, G., Flament, D., Hubscher, U., Querellou, J., et Raffin, J. P. (2005). The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. *J Mol Biol* **350**, 53-64.
- Henras, A., Henry, Y., Bousquet-Antonelli, C., Noaillac-Depeyre, J., Gelugne, J. P., et Caizergues-Ferrer, M. (1998). Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J* **17**, 7078-7090.
- Henras, A., Dez, C., Noaillac-Depeyre, J., Henry, Y., et Caizergues-Ferrer, M. (2001). Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. *Nucl. Acids Res* **29**, 2733-2746.
- Henras, A. K., Dez, C., et Henry, Y. (2004). RNA structure and function in C/D and H/ACA s(no)RNPs. *Curr Opin Struct Biol* **14**, 335-43.
- Hershberg, R., Altuvia, S., et Margalit, H. (2003). A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res* **31**, 1813-20.
- Hethke, C., Bergerat, A., Hausner, W., Forterre, P., et Thomm, M. (1999). Cell-free transcription at 95 degrees: thermostability of transcriptional components and DNA topology requirements of *Pyrococcus* transcription. *Genetics* **152**, 1325-33.
- Higuchi, M. L., Santos, M. H., Roggerio, A., Kawakami, J. T., Bezerra, H. G., et Canzian, M. (2006). A role for archaeal organisms in development of atherosclerotic vulnerable plaques and myxoid matrices. *Clinics* **61**, 473-8.
- Hirose, T., Ideue, T., Nagai, M., Hagiwara, M., Shu, M. D., et Steitz, J. A. (2006). A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing. *Mol Cell* **23**, 673-84.
- Hoang, C., et Ferre-D'Amare, A. R. (2001). Cocystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **107**, 929-39.
- Hoang, C., et Ferre-D'Amare, A. R. (2004). Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. *RNA* **10**, 1026-33.
- Hoang, C., Hamilton, C. S., Mueller, E. G., et Ferre-D'Amare, A. R. (2005). Precursor complex structure of pseudouridine synthase TruB suggests coupling of active site perturbations to an RNA-sequestering peripheral protein domain. *Protein Sci* **14**, 2201-6.
- Hoang, C., Chen, J., Vizthum, C. A., Kandel, J. M., Hamilton, C. S., Mueller, E. G., et Ferre-D'Amare, A. R. (2006). Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure. *Mol Cell* **24**, 535-45.
- Hoechsmann, M., Toeller, T., Giegerich, R., et Kurtz, S. (2003). Local Similarity of RNA Secondary Structures. *Proc. of the IEEE Bioinformatics Conference CSB* **2003**, 159-168.
- Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, S., Tacker, M., et Schuster, P. (1994). Fast Folding and Comparison of RNA Secondary Structures. *Monatsh. Chem.* **125**, 167-188.
- Holm, L., et Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**, 123-38.
- Holmes, M. L., et Dyall-Smith, M. L. (1990). A plasmid vector with a selectable marker for halophilic archaeobacteria. *J Bacteriol* **172**, 756-61.

- Holmes, M. L., Nuttall, S. D., et Dyll-Smith, M. L. (1991). Construction and use of halobacterial shuttle vectors and further studies on Haloferax DNA gyrase. *J Bacteriol* **173**, 3807-13.
- Holmes, W. M., Andraos-Selim, C., Roberts, I., et Wahab, S. Z. (1992). Structural requirements for tRNA methylation. Action of Escherichia coli tRNA(guanosine-1)methyltransferase on tRNA(1Leu) structural variants. *J Biol Chem* **267**, 13440-5.
- Huang, L., Ku, J., Pookanjanatavip, M., Gu, X., Wang, D., Greene, P. J., et Santi, D. V. (1998a). Identification of two Escherichia coli pseudouridine synthases that show multisite specificity for 23S RNA. *Biochemistry* **37**, 15951-15957.
- Huang, L., Pookanjanatavip, M., Gu, X., et Santi, D. V. (1998b). A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry* **37**, 344-351.
- Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., et Stetter, K. O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63-7.
- Huber, H., Hohn, M. J., Stetter, K. O., et Rachel, R. (2003). The phylum Nanoarchaeota: present knowledge and future perspectives of a unique form of life. *Res Microbiol* **154**, 165-71.
- Huber, R., Huber, H., et Stetter, K. O. (2000). Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. *FEMS Microbiol Rev* **24**, 615-23.
- Hur, S., et Stroud, R. M. (2007). How U38, 39, and 40 of many tRNAs become the targets for pseudouridylation by TruA. *Mol Cell* **26**, 189-203.
- Huttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J. P., et Brosius, J. (2001). RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *Embo J* **20**, 2943-53.
- Huttenhofer, A., et Schattner, P. (2006). The principles of guiding by RNA: chimeric RNA-protein enzymes. *Nat Rev Genet* **7**, 475-82.
- Huttenhofer, A., et Vogel, J. (2006). Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res* **34**, 635-46.
- Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., et Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834-8.
- Hutvagner, G., et Zamore, P. D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056-60.
- Hutvagner, G. (2005). Small RNA asymmetry in RNAi: function in RISC assembly and gene regulation. *FEBS Lett* **579**, 5850-7.

I

- Ideno, A., Furutani, M., Iwabuchi, T., Iida, T., Iba, Y., Kurosawa, Y., Sakuraba, H., Ohshima, T., Kawarabayashi, Y., et Maruyama, T. (2004). Expression of foreign proteins in Escherichia coli by fusing with an archaeal FK506 binding protein. *Appl Microbiol Biotechnol* **64**, 99-105.
- Inagaki, Y., et Ford Doolittle, W. (2000). Evolution of the eukaryotic translation termination system: origins of release factors. *Mol Biol Evol* **17**, 882-9.
- Ingalls, A. E., Shah, S. R., Hansman, R. L., Aluwihare, L. I., Santos, G. M., Druffel, E. R., et Pearson, A. (2006). Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. *Proc Natl Acad Sci U S A* **103**, 6442-7.
- Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S., et Yokoyama, S. (2003). Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell* **113**, 383-94.

J

- Jacquier, A. (1998). RNA-catalysed RNA processing. In "Eukaryotic mRNA processing" (A. R. Krainer, Ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.
- Jady, B. E., et Kiss, T. (2001). A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *Embo J* **20**, 541-51.
- Jady, B. E., Bertrand, E., et Kiss, T. (2004). Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *J Cell Biol* **164**, 647-52.
- Jain, R., Rivera, M. C., Moore, J. E., et Lake, J. A. (2002). Horizontal gene transfer in microbial genome evolution. *Theor Popul Biol* **61**, 489-95.
- Jarrell, K. F., Bayley, D. P., et Kostyukova, A. S. (1996). The archaeal flagellum: a unique motility structure. *J Bacteriol* **178**, 5057-64.
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., et Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem Sci* **23**, 403-5.
- Jeanthon, C., L'Haridon, S., Reysenbach, A. L., Corre, E., Vernet, M., Messner, P., Sleytr, U. B., et Prieur, D. (1999). Methanococcus vulcanius sp. nov., a novel hyperthermophilic methanogen isolated from East Pacific Rise, and

- identification of *Methanococcus* sp. DSM 4213T as *Methanococcus fervens* sp. nov. *Int J Syst Bacteriol* **49** Pt 2, 583-9.
- Jelinska, C., Conroy, M. J., Craven, C. J., Hounslow, A. M., Bullough, P. A., Waltho, J. P., Taylor, G. L., et White, M. F. (2005). Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. *Structure* **13**, 963-71.
- Jiang, H.-Q., Motorin, Y. A., Jin, Y.-X., et Grosjean, H. (1997). Pleiotropic effects of intron removal on base modifications pattern of yeast tRNA^{Phe}: an *in vitro* study. *Nucleic Acids Res* **25**, 2694-2701.
- Jiang, T., Wang, J. T. L., et Zhang, K. (1995). Alignment of Trees - An Alternative to Tree Edit. *Theoretical Computer Science* **143**, 137-148.
- Jiang, W., Middleton, K., Yoon, H. J., Fouquet, C., et Carbon, J. (1993). An essential yeast protein, CBF5p, binds *in vitro* to centromeres and microtubules. *Molecular and Cellular Biology* **13**, 4884-4893.
- Jin, H., Loria, J. P., et Moore, P. B. (2007). Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Mol Cell* **26**, 205-15.
- Joanny, G., Le Derout, J., Brechemier-Baey, D., Labas, V., Vinh, J., Regnier, P., et Hajnsdorf, E. (2007). Polyadenylation of a functional mRNA controls gene expression in *Escherichia coli*. *Nucleic Acids Res* **35**, 2494-502.
- Johnson, L., et Soll, D. (1970). *In vitro* biosynthesis of pseudouridine at the polynucleotide level by an enzyme extract from *Escherichia coli*. *Proc Natl Acad Sci U S A* **67**, 943-50.
- Johnson, P. F., et Abelson, J. (1983). The yeast tRNA^{Tyr} gene intron is essential for correct modification of its tRNA product. *Nature* **302**, 681-687.
- Jokela, M., Eskelinen, A., Pospiech, H., Rouvinen, J., et Syvaioja, J. E. (2004). Characterization of the 3' exonuclease subunit DPI of *Methanococcus jannaschii* replicative DNA polymerase D. *Nucleic Acids Res* **32**, 2430-40.
- Jolivet, E., L'Haridon, S., Corre, E., Forterre, P., et Prieur, D. (2003a). *Thermococcus gammatolerans* sp. nov., a hyperthermophilic archaeon from a deep-sea hydrothermal vent that resists ionizing radiation. *Int J Syst Evol Microbiol* **53**, 847-51.
- Jolivet, E., Matsunaga, F., Ishino, Y., Forterre, P., Prieur, D., et Myllykallio, H. (2003b). Physiological responses of the hyperthermophilic archaeon "*Pyrococcus abyssi*" to DNA damage caused by ionizing radiation. *J Bacteriol* **185**, 3958-61.

K

- Kachlany, S. C., Planet, P. J., Bhattacharjee, M. K., Kollia, E., DeSalle, R., Fine, D. H., et Figurski, D. H. (2000). Nonspecific adherence by *Actinobacillus actinomycetemcomitans* requires genes widespread in bacteria and archaea. *J Bacteriol* **182**, 6169-76.
- Kalayoglu, M. V., Hoerneman, B., LaVerda, D., Morrison, S. G., Morrison, R. P., et Byrne, G. I. (1999). Cellular oxidation of low-density lipoprotein by *Chlamydia pneumoniae*. *J Infect Dis* **180**, 780-90.
- Kammen, H. O., Marvel, C. C., Hardy, L., et Penhoet, E. E. (1988). Purification, structure, and properties of *Escherichia coli* tRNA pseudouridine synthase I. *J Biol Chem* **263**, 2255-2263.
- Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., et al. (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res* **14**, 331-42.
- Kandler, O., et König, H. (1998). Cell wall polymers in Archaea (Archaeobacteria). *Cell Mol Life Sci* **54**, 305-8.
- Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S., et Gingeras, T. R. (2005). Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res* **15**, 987-97.
- Karlin, D. A., Jones, R. D., Stroehlein, J. R., Mastromarino, A. J., et Potter, G. D. (1982). Breath methane excretion in patients with unresected colorectal cancer. *J Natl Cancer Inst* **69**, 573-6.
- Karlin, S. (1999). Bacterial DNA strand compositional asymmetry. *Trends Microbiol* **7**, 305-8.
- Karner, M. B., DeLong, E. F., et Karl, D. M. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507-10.
- Kashefi, K., et Lovley, D. R. (2003). Extending the upper temperature limit for life. *Science* **301**, 934.
- Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C. C., Suzuki, M., Kawai, J., et al. (2005). Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564-6.
- Kates, M. (1996). Structural analysis of phospholipids and glycolipids in extremely halophilic archaeobacteria. *J Microbiol. Meth.* **25**, 113-128.
- Kawano, M., Reynolds, A. A., Miranda-Rios, J., et Storz, G. (2005). Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res* **33**, 1040-50.
- Kawarabayashi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., et al. (1998). Complete sequence and gene organization of the genome of a hyper- thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res* **5**, 55-76.

- Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., *et al.* (1999). Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res* **6**, 83-101, 145-52.
- Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., *et al.* (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. *DNA Res* **8**, 123-40.
- Kawashima, T., Amano, N., Koike, H., Makino, S., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T., *et al.* (2000). Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proc Natl Acad Sci U S A* **97**, 14257-62.
- Kaya, Y., *et Ofengand, J.* (2003). A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA* **9**, 711-21.
- Kaya, Y., Del Campo, M., Ofengand, J., *et Malhotra, A.* (2004). Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *J Biol Chem* **279**, 18107-10.
- Keiler, K. C. (2007). Physiology of tmRNA: what gets tagged and why? *Curr Opin Microbiol* **10**, 169-75.
- Keith, G. (1995). Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie* **77**, 142-144.
- Kelman, Z., *et White, M. F.* (2005). Archaeal DNA replication and repair. *Curr Opin Microbiol* **8**, 669-76.
- Kennedy, S. P., Ng, W. V., Salzberg, S. L., Hood, L., *et DasSarma, S.* (2001). Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res* **11**, 1641-50.
- Kerr, I. D., Wadsworth, R. I., Cubeddu, L., Blankenfeldt, W., Naismith, J. H., *et White, M. F.* (2003). Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein. *Embo J* **22**, 2561-70.
- Kessler, A., Brinkman, A. B., van der Oost, J., *et Prangishvili, D.* (2004). Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon *sulfolobus*. *J Bacteriol* **186**, 7745-53.
- Khaitovich, P., Tenson, T., Kloss, P., *et Mankin, A. S.* (1999). Reconstitution of functionally active *Thermus aquaticus* large ribosomal subunits with in vitro-transcribed rRNA. *Biochemistry* **38**, 1780-8.
- Khan, M. S., *et Maden, B. E.* (1977). Nucleotide sequence relationships between vertebrate 5.8 S ribosomal RNAs. *Nucleic Acids Res* **4**, 2495-505.
- Khanna, M., Wu, H., Johansson, C., Caizergues-Ferrer, M., *et Feigon, J.* (2006). Structural study of the H/ACA snoRNP components Nop10p and the 3' hairpin of U65 snoRNA. *RNA* **12**, 40-52.
- Kim, D. F., *et Green, R.* (1999). Base-pairing between 23S rRNA and tRNA in the ribosomal A site. *Mol Cell* **4**, 859-64.
- Kimball, S. R. (1999). Eukaryotic initiation factor eIF2. *Int J Biochem Cell Biol* **31**, 25-9.
- King, T. H., Liu, B., McCully, R. R., *et Fournier, M. J.* (2003). Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell* **11**, 425-35.
- Kirpekar, F., Hansen, L. H., Rasmussen, A., Poehlsgaard, J., *et Vester, B.* (2005). The archaeon *Haloarcula marismortui* has few modifications in the central parts of its 23S ribosomal RNA. *J Mol Biol* **348**, 563-73.
- Kis-Papo, T., *et Oren, A.* (2000). Halocins: are they involved in the competition between halobacteria in saltern ponds? *Extremophiles* **4**, 35-41.
- Kiss, A. M., Jady, B. E., Bertrand, E., *et Kiss, T.* (2004). Human box H/ACA pseudouridylation guide RNA machinery. *Mol Cell Biol* **24**, 5797-807.
- Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109**, 145-8.
- Kiss-Laszlo, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M., *et Kiss, T.* (1996). Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85**, 1077-1088.
- Kiss-Laszlo, Z., Henry, Y., *et Kiss, T.* (1998). Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *Embo J* **17**, 797-807.
- Kittur, N., Darzacq, X., Roy, S., Singer, R. H., *et Meier, U. T.* (2006). Dynamic association and localization of human H/ACA RNP proteins. *RNA* **12**, 2057-62.
- Klein, D. J., Schmeing, T. M., Moore, P. B., *et Steitz, T. A.* (2001). The kink-turn: a new RNA secondary structure motif. *Embo J* **20**, 4214-21.
- Klein, R. J., Misulovin, Z., *et Eddy, S. R.* (2002). Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc Natl Acad Sci U S A* **99**, 7542-7.
- Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., *et al.* (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**, 364-70.
- Kokoeva, M. V., Storch, K. F., Klein, C., *et Oesterhelt, D.* (2002). A novel mode of sensory transduction in archaea: binding protein-mediated chemotaxis towards osmoprotectants and amino acids. *Embo J* **21**, 2312-22.
- Kolb, F. A., Engdahl, H. M., Slagter-Jager, J. G., Ehresmann, B., Ehresmann, C., Westhof, E., Wagner, E. G., *et Romby, P.* (2000). Progression of a loop-loop complex to a four-way junction is crucial for the activity of a regulatory antisense RNA. *Embo J* **19**, 5905-15.

- Kolb, F. A., Westhof, E., Ehresmann, C., Ehresmann, B., Wagner, E. G., et Romby, P. (2001). Bulged residues promote the progression of a loop-loop interaction to a stable and inhibitory antisense-target RNA complex. *Nucleic Acids Res* **29**, 3145-53.
- Kolodrubetz, D., et Burgum, A. (1991). Sequence and genetic analysis of NHP2: a moderately abundant high mobility group-like nuclear protein with an essential function in *Saccharomyces cerevisiae*. *Yeast* **7**, 79-90.
- Konings, W. N., Albers, S. V., Koning, S., et Driessen, A. J. (2002). The cell membrane plays a crucial role in survival of bacteria and archaea in extreme environments. *Antonie Van Leeuwenhoek* **81**, 61-72.
- Konneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., et Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543-6.
- Koonin, E. V. (1996). Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res* **24**, 2411-2415.
- Koonin, E. V., Makarova, K. S., et Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**, 709-42.
- Kressler, D., Linder, P., et de La Cruz, J. (1999). Protein trans-acting factors involved in ribosome biogenesis in *Saccharomyces cerevisiae*. *Mol Cell Biol* **19**, 7897-912.
- Krishnan, L., Sad, S., Patel, G. B., et Sprott, G. D. (2001). The potent adjuvant activity of archaeosomes correlates to the recruitment and activation of macrophages and dendritic cells in vivo. *J Immunol* **166**, 1885-93.
- Kulik, E. M., Sandmeier, H., Hinni, K., et Meyer, J. (2001). Identification of archaeal rDNA from subgingival dental plaque by PCR amplification and sequence analysis. *FEMS Microbiol Lett* **196**, 129-33.
- Kushner, D. J. (1964). Lysis and dissolution of cells and envelopes of an extremely halophilic bacterium. *J Bacteriol* **87**, 1147-56.
- Kyrpides, N. C., et Ouzounis, C. A. (1999). Transcription in archaea. *Proc Natl Acad Sci U S A* **96**, 8545-50.

L

- Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L., et Ellenberger, T. E. (1996). Structural basis for the excision repair of alkylation-damaged DNA. *Cell* **86**, 321-9.
- Laferrere, A., Gautheret, D., et Cedergren, R. (1994). An RNA pattern matching program with enhanced performance and portability. *Comput Appl Biosci* **10**, 211-2.
- Lafontaine, D. L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., et Tollervey, D. (1998a). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.* **12**, 527-537.
- Lafontaine, D. L. J., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., et Tollervey, D. (1998b). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev* **12**, 527-537.
- Lafontaine, D. L., et Tollervey, D. (1998). Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem. Sci.* **23**, 383-388.
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., et Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853-8.
- Lambert, A., Lescure, A., et Gautheret, D. (2002). A survey of metazoan selenocysteine insertion sequences. *Biochimie* **84**, 953-9.
- Lambert, A., Fontaine, J. F., Legendre, M., Leclerc, F., Permal, E., Major, F., Putzer, H., Delfour, O., Michot, B., et Gautheret, D. (2004). The ERPIN server: an interface to profile-based RNA motif identification. *Nucleic Acids Res* **32**, W160-5.
- Lane, B. G. (1998). Historical Perspectives on RNA Nucleoside Modifications. *Eds H. Grosjean and R. Benne Modification and Editing of RNA*, 1-20.
- Lane, B. G., Ofengand, J., et Gray, M. W. (1992). Pseudouridine in the large-subunit (23 S-like) ribosomal RNA. The site of peptidyl transfer in the ribosome? *FEBS Lett* **302**, 1-4.
- Lane, B. G., Ofengand, J., et Gray, M. W. (1995). Pseudouridine and O2'-methylated nucleosides. Significance of their selective occurrence in rRNA domains that function in ribosome- catalyzed synthesis of the peptide bonds in proteins. *Biochimie* **77**, 7-15.
- Lange, T. S., Borovjagin, A., Maxwell, E. S., et Gerbi, S. A. (1998). Conserved boxes C and D are essential nucleolar localization elements of U14 and U8 snoRNAs. *Embo J* **17**, 3176-87.
- Lange, U., et Hausner, W. (2004). Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. *Mol Microbiol* **52**, 1133-43.
- Langer, D., Hain, J., Thuriaux, P., et Zillig, W. (1995). Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A* **92**, 5768-72.
- Langworthy, T. A., Smith, P. F., et Mayberry, W. R. (1972). Lipids of *Thermoplasma acidophilum*. *J Bacteriol* **112**, 1193-200.
- Lao-Sirieix, S. H., Pellegrini, L., et Bell, S. D. (2005). The promiscuous primase. *Trends Genet* **21**, 568-72.
- Lapeyre, B., et Purushothaman, S. K. (2004). Spb1p-directed formation of Gm2922 in the ribosome catalytic center occurs at a late processing stage. *Mol Cell* **16**, 663-9.

- Lapidot, M., et Pilpel, Y. (2006). Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep* **7**, 1216-22.
- Lau, N. C., Lim, L. P., Weinstein, E. G., et Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858-62.
- Lavorgna, G., Dahary, D., Lehner, B., Sorek, R., Sanderson, C. M., et Casari, G. (2004). In search of antisense. *Trends Biochem Sci* **29**, 88-94.
- Lease, R. A., Cusick, M. E., et Belfort, M. (1998). Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc Natl Acad Sci U S A* **95**, 12456-61.
- Lecointe, F., Simos, G., Sauer, A., Hurt, E. C., Motorin, Y., et Grosjean, H. (1998). Characterization of yeast protein Deg 1 as Pseudouridine synthase (Pus 3) catalysing the formation of Ψ_{38} and Ψ_{39} in tRNA anticodon loop. *JBC* **273**, 1316-1323.
- Lecointe, F., Namy, O., Hatin, I., Simos, G., Rousset, J. P., et Grosjean, H. (2002). Lack of pseudouridine 38/39 in the anticodon arm of yeast cytoplasmic tRNA decreases in vivo recoding efficiency. *J Biol Chem* **277**, 30445-53.
- Lee, R. C., Feinbaum, R. L., et Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843-54.
- Lee, R. C., et Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862-4.
- Lee, Y., Jeon, K., Lee, J. T., Kim, S., et Kim, V. N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *Embo J* **21**, 4663-70.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-9.
- Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., et Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**, 4051-60.
- Legendre, M., Lambert, A., et Gautheret, D. (2005). Profile-based detection of microRNA precursors in animal genomes. *Bioinformatics* **21**, 841-5.
- Lenz, D. H., Miller, M. B., Zhu, J., Kulkarni, R. V., et Bassler, B. L. (2005). CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*. *Mol Microbiol* **58**, 1186-202.
- Leontis, N. B., et Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499-512.
- Lepp, P. W., Brinig, M. M., Ouverney, C. C., Palm, K., Armitage, G. C., et Relman, D. A. (2004). Methanogenic Archaea and human periodontal disease. *Proc Natl Acad Sci U S A* **101**, 6176-81.
- Lescoute, A., Leontis, N. B., Massire, C., et Westhof, E. (2005). Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res* **33**, 2395-409.
- Lescure, A., Gautheret, D., Carbon, P., et Krol, A. (1999). Novel selenoproteins identified in silico and in vivo by using a conserved RNA structural motif. *J Biol Chem* **274**, 38147-54.
- Lescure, A., Fagegaltier, D., Carbon, P., et Krol, A. (2002). Protein factors mediating selenoprotein synthesis. *Curr Protein Pept Sci* **3**, 143-151.
- Leulliot, N., Godin, K. S., Hoareau-Aveilla, C., Quevillon-Cheruel, S., Varani, G., Henry, Y., et Van Tilbeurgh, H. (2007). The Box H/ACA RNP Assembly Factor Naf1p Contains a Domain Homologous to Gar1p Mediating its Interaction with Cbf5p. *J Mol Biol*.
- Leung, A. K., et Lamond, A. I. (2002). In vivo analysis of NHPX reveals a novel nucleolar localization pathway involving a transient accumulation in splicing speckles. *J Cell Biol* **157**, 615-29.
- Li, L., et Ye, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443**, 302-7.
- Li, Y. Y., Qin, L., Guo, Z. M., Liu, L., Xu, H., Hao, P., Su, J., Shi, Y., He, W. Z., et Li, Y. X. (2006). In silico discovery of human natural antisense transcripts. *BMC Bioinformatics* **7**, 18.
- Liang, X. H., Liu, L., et Michaeli, S. (2001). Identification of the first trypanosome H/ACA RNA that guides pseudouridine formation on rRNA. *J Biol Chem* **276**, 40313-8.
- Liang, X. H., Hury, A., Hoze, E., Uliel, S., Myslyuk, I., Apatoff, A., Unger, R., et Michaeli, S. (2007). Genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Leishmania major* indicates conservation among trypanosomatids in the repertoire and in their rRNA targets. *Eukaryot Cell* **6**, 361-77.
- Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B., et Bartel, D. P. (2003a). Vertebrate microRNA genes. *Science* **299**, 1540.
- Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., Burge, C. B., et Bartel, D. P. (2003b). The microRNAs of *Caenorhabditis elegans*. *Genes Dev* **17**, 991-1008.
- Limbach, P. A., Crain, P. F., et McCloskey, J. A. (1994). Summary: the modified nucleosides of RNA. *Nucleic Acids Res* **22**, 2183-96.
- Lin, S. L., Miller, J. D., et Ying, S. Y. (2006). Intronic MicroRNA (miRNA). *J Biomed Biotechnol* **2006**, 26818.
- Littlefield, O., Korkhin, Y., et Sigler, P. B. (1999). The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci U S A* **96**, 13668-73.
- Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., et Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437-41.
- Liu, J., Gough, J., et Rost, B. (2006). Distinguishing protein-coding from non-coding RNAs through support vector machines. *PLoS Genet* **2**, e29.

- Liu, M. Y., Gui, G., Wei, B., Preston, J. F., 3rd, Oakford, L., Yuksel, U., Giedroc, D. P., et Romeo, T. (1997). The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J Biol Chem* **272**, 17502-10.
- Livny, J., Fogel, M. A., Davis, B. M., et Waldor, M. K. (2005). sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes. *Nucleic Acids Res* **33**, 4096-105.
- Livny, J., Brenic, A., Lory, S., et Waldor, M. K. (2006). Identification of 17 *Pseudomonas aeruginosa* sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. *Nucleic Acids Res* **34**, 3484-93.
- Livny, J., et Waldor, M. K. (2007). Identification of small RNAs in diverse bacterial species. *Curr Opin Microbiol* **10**, 96-101.
- Londei, P. (2005). Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol Rev* **29**, 185-200.
- Lopez, P., Philippe, H., Myllykallio, H., et Forterre, P. (1999). Identification of putative chromosomal origins of replication in Archaea. *Mol Microbiol* **32**, 883-6.
- Lorenz, C., von Pelchrzim, F., et Schroeder, R. (2006). Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat Protoc* **1**, 2204-12.
- Lowe, T. M., et Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-64.
- Lowe, T. M., et Eddy, S. R. (1999). A computational screen for methylation guide snoRNAs in yeast. *Science* **283**, 1168-71.
- Lucas, S., Toffin, L., Zivanovic, Y., Charlier, D., Moussard, H., Forterre, P., Prieur, D., et Erauso, G. (2002). Construction of a shuttle vector for, and spheroplast transformation of, the hyperthermophilic archaeon *Pyrococcus abyssi*. *Appl Environ Microbiol* **68**, 5528-36.
- Lucht, J. M., et Bremer, E. (1994). Adaptation of *Escherichia coli* to high osmolarity environments: osmoregulation of the high-affinity glycine betaine transport system proU. *FEMS Microbiol Rev* **14**, 3-20.
- Lukowiak, A. A., Narayanan, A., Li, Z. H., Terns, R. M., et Terns, M. P. (2001). The snoRNA domain of vertebrate telomerase RNA functions to localize the RNA within the nucleus. *RNA* **7**, 1833-44.
- Lundgren, M., Andersson, A., Chen, L., Nilsson, P., et BeRNander, R. (2004). Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A* **101**, 7046-51.
- Lundgren, M., et BeRNander, R. (2005). Archaeal cell cycle progress. *Curr Opin Microbiol* **8**, 662-8.
- Lustig, F., Boren, T., Claesson, C., Simonsson, C., Barciszewska, M., et Lagerkvist, U. (1993). The nucleotide in position 32 of the tRNA anticodon loop determines ability of anticodon UCC to discriminate among glycine codons. *Proc. Natl. Acad. Sci. USA* **90**, 3343-7.
- Lykke-Andersen, J., Aagaard, C., Semionenkova, M., et Garrett, R. A. (1997). Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem Sci* **22**, 326-31.

M

- Ma, X., Zhao, X., et Yu, Y. T. (2003). Pseudouridylation (Psi) of U2 snRNA in *S.cerevisiae* is catalyzed by an RNA-independent mechanism. *Embo J* **22**, 1889-97.
- Ma, X., Yang, C., Alexandrov, A., Grayhack, E. J., Behm-Ansmant, I., et Yu, Y. T. (2005). Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *Embo J* **24**, 2403-13.
- Macke, T. J., Ecker, D. J., Gutell, R. R., Gautheret, D., Case, D. A., et Sampath, R. (2001). RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res* **29**, 4724-35.
- Maden, B. E. H. (1990). The numerous modified nucleotides in eukaryotic ribosomal RNA. *Progr. Nucl. Acid Res. Mol. Biol.* **39**, 241-303.
- Maeder, D. L., Weiss, R. B., Dunn, D. M., Cherry, J. L., Gonzalez, J. M., DiRuggiero, J., et Robb, F. T. (1999). Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics* **152**, 1299-305.
- Magill, C. P., Jackson, S. P., et Bell, S. D. (2001). Identification of a conserved archaeal RNA polymerase subunit contacted by the basal transcription factor TFB. *J Biol Chem* **276**, 46693-6.
- Magot, M., Ollivier, B., et Patel, B. K. (2000). Microbiology of petroleum reservoirs. *Antonie Van Leeuwenhoek* **77**, 103-16.
- Makarova, K. S., et Koonin, E. V. (2003). Comparative genomics of Archaea: how much have we learned in six years, and what's next? *Genome Biol* **4**, 115.
- Manival, X., Charron, C., Fourmann, J. B., Godard, F., Charpentier, B., et Branlant, C. (2006). Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res* **34**, 826-39.
- Marc, F., Sandman, K., Lurz, R., et Reeve, J. N. (2002). Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. *J Biol Chem* **277**, 30879-86.

- Marinsek, N., Barry, E. R., Makarova, K. S., Dionne, I., Koonin, E. V., et Bell, S. D. (2006). GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep* **7**, 539-45.
- Marintchev, A., Kolupaeva, V. G., Pestova, T. V., et Wagner, G. (2003). Mapping the binding interface between human eukaryotic initiation factors 1A and 5B: a new interaction between old partners. *Proc Natl Acad Sci U S A* **100**, 1535-40.
- Marker, C., Zemmann, A., Terhorst, T., Kiefmann, M., Kastenmayer, J. P., Green, P., Bachellerie, J. P., Brosius, J., et Huttenhofer, A. (2002). Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr Biol* **12**, 2002-13.
- Marmier-Gourrier, N., Clery, A., Senty-Segault, V., Charpentier, B., Schlotter, F., Leclerc, F., Fournier, R., et Branlant, C. (2003). A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA*, **9**, 821-838.
- Marquet, R. (1998). Importance of modified nucleotides in replication of retroviruses, plant pararetroviruses, and retrotransposons. In "The modification and editing of RNA" (H. Grosjean et R. Benne, Eds.), pp. 517-533. ASM Press, Washington, DC.
- Maruyama, T., Suzuki, R., et Furutani, M. (2004). Archaeal peptidyl prolyl cis-trans isomerases (PPIases) update 2004. *Front Biosci* **9**, 1680-720.
- Masse, E., et Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**, 4620-5.
- Massenet, S., Mougin, A., et Branlant, C. (1998). Posttranscriptional modifications in the U small nuclear RNAs. In "The Modification and Editing of RNA" (H. Grosjean et R. Benne, Eds.), pp. 201-227. ASM Press, New York.
- Massenet, S., et Branlant, C. (1999). A limited number of pseudouridine residues in the human atac spliceosomal UsnRNAs as compared to human major spliceosomal UsnRNAs. *RNA* **5**, 1495-503.
- Massenet, S., Motorin, Y., Lafontaine, D. L. J., Hurt, E. C., Grosjean, H., et Branlant, C. (1999). Pseudouridine mapping in the *Saccharomyces cerevisiae* spliceosomal UsnRNAs reveals that pseudouridine synthase Pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Molecular and Cellular Biology* **19**, 2142-2154.
- Matera, A. G. (1999). Nuclear bodies: multifaceted subdomains of the interchromatin space. *Trends Cell Biol* **9**, 302-9.
- Matera, A. G., Terns, R. M., et Terns, M. P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* **8**, 209-20.
- Mathews, D. H., Sabina, J., Zuker, M., et Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**, 911-40.
- Mathrani, I. M., Boone, D. R., Mah, R. A., Fox, G. E., et Lau, P. P. (1988). *Methanohalophilus zhilinae* sp. nov., an alkaliphilic, halophilic, methylotrophic methanogen. *Int J Syst Bacteriol* **38**, 139-42.
- Matsunaga, F., Forterre, P., Ishino, Y., et Myllykallio, H. (2001). In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc Natl Acad Sci U S A* **98**, 11152-7.
- Matsunaga, F., Norais, C., Forterre, P., et Myllykallio, H. (2003). Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep* **4**, 154-8.
- Matte, A., Louie, G. V., Sivaraman, J., Cygler, M., et Burley, S. K. (2005). Structure of the pseudouridine synthase RsuA from *Haemophilus influenzae*. *Acta Crystallograph Sect F Struct Biol Cryst Commun* **61**, 350-4.
- Matte-Tailliez, O., Brochier, C., Forterre, P., et Philippe, H. (2002). Archaeal phylogeny based on ribosomal proteins. *Mol Biol Evol* **19**, 631-9.
- Mattick, J. S., et Makunin, I. V. (2006). Non-coding RNA. *Hum Mol Genet* **15 Spec No 1**, R17-29.
- McCleverty, C. J., Hornsby, M., Spraggon, G., et Kreusch, A. (2007). Crystal Structure of Human Pus10, A Novel Pseudouridine Synthase. *J Mol Biol*.
- Meier, U. T., et Blobel, G. (1994). NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria. *J Cell Biol* **127**, 1505-14.
- Meile, L., Abendschein, P., et Leisinger, T. (1990). Transduction in the archaebacterium *Methanobacterium thermoautotrophicum* Marburg. *J Bacteriol* **172**, 3507-8.
- Meroueh, M., Grohar, P. J., Qiu, J., SantaLucia, J., Jr., Scaringe, S. A., et Chow, C. S. (2000). Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of *Escherichia coli* 23S rRNA. *Nucleic Acids Res* **28**, 2075-83.
- Merryman, C., Moazed, D., Daubresse, G., et Noller, H. F. (1999a). Nucleotides in 23S rRNA protected by the association of 30S and 50S ribosomal subunits. *J Mol Biol* **285**, 107-13.
- Merryman, C., Moazed, D., McWhirter, J., et Noller, H. F. (1999b). Nucleotides in 16S rRNA protected by the association of 30S and 50S ribosomal subunits. *J Mol Biol* **285**, 97-105.
- Metcalf, W. W., Zhang, J. K., Apolinario, E., Sowers, K. R., et Wolfe, R. S. (1997). A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. *Proc Natl Acad Sci U S A* **94**, 2626-31.
- Miller, T. L., et Wolin, M. J. (1982). Enumeration of *Methanobrevibacter smithii* in human feces. *Arch Microbiol* **131**, 14-8.
- Mitchell, J. R., Cheng, J., et Collins, K. (1999a). A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol. Cell. Biol.* **19**, 567-576.

- Mitchell, J. R., Wood, E., et Collins, K. (1999b). A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402**, 551-5.
- Mitchell, J. R., et Collins, K. (2000). Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. *Mol Cell* **6**, 361-71.
- Miyata, T., Suzuki, H., Oyama, T., Mayanagi, K., Ishino, Y., et Morikawa, K. (2005). Open clamp structure in the clamp-loading complex visualized by electron microscopic image analysis. *Proc Natl Acad Sci U S A* **102**, 13795-800.
- Mizutani, K., Machida, Y., Unzai, S., Park, S. Y., et Tame, J. R. (2004). Crystal structures of the catalytic domains of pseudouridine synthases RluC and RluD from *Escherichia coli*. *Biochemistry* **43**, 4454-63.
- Moller, T., Franch, T., Hojrup, P., Keene, D. R., Bachinger, H. P., Brennan, R. G., et Valentin-Hansen, P. (2002). Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell* **9**, 23-30.
- Momiyama, Y., Ohmori, R., Taniguchi, H., Nakamura, H., et Ohsuzu, F. (2004). Association of *Mycoplasma pneumoniae* infection with coronary artery disease and its interaction with chlamydial infection. *Atherosclerosis* **176**, 139-44.
- Moore, J. M., Query, C. C., et Sharp, P. A. (1993). Splicing of precursors to messenger RNAs by the spliceosome. In "The RNA world" (R. Gesteland et J. A. Drenth, Eds.), pp. 303-58. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.
- Moore, T., Zhang, Y., Fenley, M. O., et Li, H. (2004). Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure (Camb)* **12**, 807-18.
- Morin, A., Auxilien, S., Senger, B., Tewari, R., et Grosjean, H. (1998). Structural requirements for enzymatic formation of threonylcarbamoyladenosine (t6A) in tRNA: an in vivo study with *Xenopus laevis* oocytes. *RNA* **4**, 24-37.
- Motorin, Y., et Grosjean, H. (1998). Chemical structures and classification of posttranscriptionally modified nucleosides in RNA. In "The modification and editing of RNA" (H. Grosjean et R. Benne, Eds.), pp. 543-50. ASM Press, Washington, DC.
- Motorin, Y., Keith, G., Simon, C., Foiret, D., Simos, G., Hurt, E., et Grosjean, H. (1998). The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA* **4**, 856-69.
- Motorin, Y. A., Bec, G., R. Tewari, et Grosjean, H. (1997). Transfer RNA recognition by the *E. coli* Δ^2 -isopentenyl-pyrophosphate:tRNA Δ^2 -isopentenyl transferase: Dependence on the anticodon arm structure. *RNA* **3**, 721-733.
- Motorin, Y., Muller, S., Behm-Ansmant, I., et Branlant, C. (2007). Identification of Modified Residues in RNAs by Reverse Transcription-Based Methods. *Methods Enzymol* **425**, 21-53.
- Mount, S. M. (1996). AT-AC introns: an ATtACk on dogma [comment]. *Science* **271**, 1690-2.
- Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., et Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* **16**, 720-8.
- Mueller, S. O., et Slany, R. K. (1995). Structural analysis of the interaction of the tRNA modifying enzymes Tgt and QueA with a substrate tRNA. *FEBS Lett* **361**, 259-64.
- Muller, B., Allmansberger, R., et Klein, A. (1985). Termination of a transcription unit comprising highly expressed genes in the archaeobacterium *Methanococcus voltae*. *Nucleic Acids Res* **13**, 6439-45.
- Muller, S., Charpentier, B., Branlant, C., et Leclerc, F. (2007a). A Dedicated Computational Approach for the Identification of Archaeal H/ACA sRNAs. *Methods Enzymol* **425**, 355-87.
- Muller, S., Fourmann, J. B., Loegler, C., Charpentier, B., et Branlant, C. (2007b). Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:{Psi}55-synthase and RNA-guided RNA:{Psi}-synthase activities. *Nucleic Acids Res*.
- Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., et Darst, S. A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285-90.
- Murakami, K. S., Masuda, S., et Darst, S. A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296**, 1280-4.
- Myllykallio, H., et Forterre, P. (2000). Mapping of a chromosome replication origin in an archaeon: response. *Trends Microbiol* **8**, 537-9.

N

- Narayanan, A., Speckmann, W., Terns, R., et Terns, M. P. (1999). Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Mol Biol Cell* **10**, 2131-47.
- Nazar, R. N., Sitz, T. O., et Busch, H. (1976). Sequence homologies in mammalian 5.8S ribosomal RNA. *Biochemistry* **15**, 505-8.
- Ng, W. V., Ciufo, S. A., Smith, T. M., Bumgarner, R. E., Baskin, D., Faust, J., Hall, B., Loretz, C., Seto, J., Slagel, J., et al. (1998). Snapshot of a large dynamic replicon in a halophilic archaeon: megaplasmid or minichromosome? *Genome Res* **8**, 1131-41.
- Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., Lasky, S. R., Baliga, N. S., Thorsson, V., Sbrogna, J., et al. (2000). Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci U S A* **97**, 12176-81.

- Ni, J., Tien, A. L., et Fournier, M. J. (1997). Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**, 565-573.
- Nicoloso, M., Qu, L. H., Michot, B., et Bachellerie, J. P. (1996). Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2'-O- ribose methylation of rRNAs. *J Mol Biol* **260**, 178-95.
- Nicolson, G. L., Gan, R., et Haier, J. (2003). Multiple co-infections (Mycoplasma, Chlamydia, human herpes virus-6) in blood of chronic fatigue syndrome patients: association with signs and symptoms. *Apmis* **111**, 557-66.
- Niewmierzyczna, A., et Clarke, S. (1999). S-Adenosylmethionine-dependent methylation in *Saccharomyces cerevisiae*. Identification of a novel protein arginine methyltransferase. *J Biol Chem* **274**, 814-24.
- Nilsen, T. W. (2003). The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* **25**, 1147-9.
- Nissen, P., Hansen, J., Ban, N., Moore, P. B., et Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920-30.
- Niu, L., et Ofengand, J. (1999). Cloning and characterization of the 23S RNA pseudouridine 2633 synthase from *Bacillus subtilis*. *Biochemistry* **38**, 629-635.
- Nolivos, S., Carpousis, A. J., et Clouet-d'Orval, B. (2005). The K-loop, a general feature of the *Pyrococcus* C/D guide RNAs, is an RNA structural motif related to the K-turn. *Nucleic Acids Res* **33**, 6507-14.
- Noon, K. R., Bruenger, E., et McCloskey, J. A. (1998). Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus*. *J Bacteriol* **180**, 2883-8.
- Normand, C., Capeyrou, R., Quevillon-Cheruel, S., Mougin, A., Henry, Y., et Caizergues-Ferrer, M. (2006). Analysis of the binding of the N-terminal conserved domain of yeast Cbf5p to a box H/ACA snoRNA. *RNA* **12**, 1868-82.
- Norris, P. R., Burton, N. P., et Foulis, N. A. (2000). Acidophiles in bioreactor mineral processing. *Extremophiles* **4**, 71-6.
- Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R., et Luhrmann, R. (1999). Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J.* **18**, 6119-6133.
- Nurse, K., Wrzesinski, J., Bakin, A., Lane, B. G., et Ofengand, J. (1995). Purification, cloning, and properties of the tRNA Psi 55 synthase from *Escherichia coli*. *RNA* **1**, 102-112.

O

- Ochsenreiter, T., Selezi, D., Quaiser, A., Bonch-Osmolovskaya, L., et Schleper, C. (2003). Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ Microbiol* **5**, 787-97.
- O'Connor, E. M., et Shand, R. F. (2002). Halocins and sulfolobocins: the emerging story of archaeal protein and peptide antibiotics. *J Ind Microbiol Biotechnol* **28**, 23-31.
- Ofengand, J., et Bakin, A. (1997). Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaeobacteria, mitochondria and chloroplasts. *J Mol Biol* **266**, 246-268.
- Ofengand, J., et Fournier, M. J. (1998). The pseudouridine residues of rRNA: number, location, biosynthesis and function. In "The modification and editing of RNA" (H. Grosjean et R. Benne, Eds.), pp. 229-54. ASM Press, Washington, DC.
- Ofengand, J., Del Campo, M., et Kaya, Y. (2001a). Mapping pseudouridines in RNA molecules. *Methods* **25**, 365-73.
- Ofengand, J., Malhotra, A., Remme, J., Gutsell, N. S., Del Campo, M., Jean-Charles, S., Peil, L., et Kaya, Y. (2001b). Pseudouridines and pseudouridine synthases of the ribosome. *Cold Spring Harb Symp Quant Biol* **66**, 147-59.
- Ofengand, J. (2002). Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett* **514**, 17-25.
- Ohman, M., et Wagner, E. G. (1989). Secondary structure analysis of the RepA mRNA leader transcript involved in control of replication of plasmid R1. *Nucleic Acids Res* **17**, 2557-79.
- Olivas, W. M., Muhlrads, D., et Parker, R. (1997). Analysis of the yeast genome: identification of new non-coding and small ORF-containing RNAs. *Nucleic Acids Res* **25**, 4619-25.
- Omer, A. D., Lowe, T. M., Russell, A. G., Ebhardt, H., Eddy, S. R., et Dennis, P. P. (2000). Homologs of small nucleolar RNAs in Archaea. *Science*, **288**, 517-522.
- Omer, A. D., Ziesche, S., Ebhardt, H., et Dennis, P. P. (2002). *In vitro* reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc. Natl Acad. Sci. U S A*, **99**, 5289-5294.
- Omer, A. D., Ziesche, S., Decatur, W. A., Fournier, M. J., et Dennis, P. P. (2003). RNA-modifying machines in archaea. *Mol. Microbiol.* **48**, 617-629.
- Oren, A. (2002). Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications. *J Ind Microbiol Biotechnol* **28**, 56-63.
- Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D., et DeLong, E. F. (2001). Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* **293**, 484-7.
- Ouhammouch, M., et Geiduschek, E. P. (2005). An expanding family of archaeal transcriptional activators. *Proc Natl Acad Sci U S A* **102**, 15423-8.

P

- Palmer, J. R., et Daniels, C. J. (1995). In vivo definition of an archaeal promoter. *J Bacteriol* **177**, 1844-9.
- Pan, H., Agarwalla, S., Moustakas, D. T., Finer-Moore, J., et Stroud, R. M. (2003). Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proc Natl Acad Sci U S A* **100**, 12648-53.
- Pan, H., Ho, J. D., Stroud, R. M., et Finer-Moore, J. (2007). The crystal structure of E. coli rRNA pseudouridine synthase RluE. *J Mol Biol* **367**, 1459-70.
- Pan, T., et Sosnick, T. (2006). RNA folding during transcription. *Annu Rev Biophys Biomol Struct* **35**, 161-75.
- Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degan, B., Muller, P., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86-9.
- Pasquinelli, A. E. (2002). MicroRNAs: deviants no longer. *Trends Genet* **18**, 171-3.
- Pasquinelli, A. E., Hunter, S., et Bracht, J. (2005). MicroRNAs: a developing story. *Curr Opin Genet Dev* **15**, 200-5.
- Pasquinelli, A. E., et Ruvkun, G. (2002). Control of developmental timing by microRNAs and their targets. *Annu Rev Cell Dev Biol* **18**, 495-513.
- Patel, G. B., Nash, J. H., Agnew, B. J., et Sprott, G. D. (1994). Natural and Electroporation-Mediated Transformation of *Methanococcus voltae* Protoplasts. *Appl Environ Microbiol* **60**, 903-907.
- Pauler, F. M., Koerner, M. V., et Barlow, D. P. (2007). Silencing by imprinted noncoding RNAs: is transcription the answer? *Trends Genet* **23**, 284-92.
- Pearson, W. R., et Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* **85**, 2444-8.
- Pellizzoni, L., Baccon, J., Charroux, B., et Dreyfuss, G. (2001). The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. *Curr Biol* **11**, 1079-88.
- Perret, V., Garcia, A., Puglisi, J., Grosjean, H., Ebel, J.-P., Florentz, C., et Giegé, R. (1990). Conformation in solution of yeast tRNA^{Asp} transcripts deprived of modified nucleotides. *Biochimie* **72**, 735-744.
- Pesole, G., Liuni, S., et D'Souza, M. (2000). PatSearch: a pattern matcher software that finds functional elements in nucleotide and protein sequences and assesses their statistical significance. *Bioinformatics* **16**, 439-50.
- Pestova, T. V., Lomakin, I. B., Lee, J. H., Choi, S. K., Dever, T. E., et Hellen, C. U. (2000). The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* **403**, 332-5.
- Pestova, T. V., et Kolupaeva, V. G. (2002). The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev* **16**, 2906-22.
- Phannachet, K., et Huang, R. H. (2004). Conformational change of pseudouridine 55 synthase upon its association with RNA substrate. *Nucleic Acids Res* **32**, 1422-9.
- Phannachet, K., Elias, Y., et Huang, R. H. (2005). Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase. *Biochemistry* **44**, 15488-94.
- Pichon, C., et Felden, B. (2005). Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. *Proc Natl Acad Sci U S A* **102**, 14249-54.
- Pickford, A. S., et Cogoni, C. (2003). RNA-mediated gene silencing. *Cell Mol Life Sci* **60**, 871-82.
- Pienkowska, J., Michalowski, D., Krzyzosiak, W. J., et Szweykowska-Kulinska, Z. (2002). Pseudouridylation of U(35) in the anticodon of *Arabidopsis thaliana* pre-tRNA(Tyr) depends on length rather than structure of an intron. *Biochim Biophys Acta* **1574**, 137-44.
- Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J. S., Bejerano, G., Baertsch, R., et al. (2006). Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* **2**, e168.
- Ponting, C. P., et Lunter, G. (2006). Evolutionary biology: human brain gene wins genome race. *Nature* **443**, 149-50.
- Portnoy, V., Evguenieva-Hackenberg, E., Klein, F., Walter, P., Lorentzen, E., Klug, G., et Schuster, G. (2005). RNA polyadenylation in Archaea: not observed in *Haloferax* while the exosome polynucleotidylates RNA in *Sulfolobus*. *EMBO Rep* **6**, 1188-93.
- Portnoy, V., et Schuster, G. (2006). RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R. *Nucleic Acids Res* **34**, 5923-31.
- Prado, A., Da Costa, M., et Madeira, V. (1988). Effect of growth temperature on the lipid composition of two strains of *Thermus*. *J. Gen. Microbiol.* **134**, 1653-1660.
- Prangishvili, D., Holz, I., Stieger, E., Nickell, S., Kristjansson, J. K., et Zillig, W. (2000). Sulfolobins, specific proteinaceous toxins produced by strains of the extremely thermophilic archaeal genus *Sulfolobus*. *J Bacteriol* **182**, 2985-8.
- Prangishvili, D., Stedman, K., et Zillig, W. (2001). Viruses of the extremely thermophilic archaeon *Sulfolobus*. *Trends Microbiol* **9**, 39-43.
- Prangishvili, D., et Garrett, R. A. (2005). Viruses of hyperthermophilic Crenarchaea. *Trends Microbiol* **13**, 535-42.

- Prangishvili, D., Forterre, P., et Garrett, R. A. (2006a). Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* **4**, 837-48.
- Prangishvili, D., Vestergaard, G., Haring, M., Aramayo, R., Basta, T., Rachel, R., et Garrett, R. A. (2006b). Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. *J Mol Biol* **359**, 1203-16.
- Prescott, E. M., et Proudfoot, N. J. (2002). Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci U S A* **99**, 8796-801.
- Preston, C. M., Wu, K. Y., Molinski, T. F., et DeLong, E. F. (1996). A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc Natl Acad Sci U S A* **93**, 6241-6.

Q

- Qu, L. H., Meng, Q., Zhou, H., et Chen, Y. Q. (2001). Identification of 10 novel snoRNA gene clusters from *Arabidopsis thaliana*. *Nucleic Acids Res* **29**, 1623-30.
- Quigley, G. J., et Rich, A. (1976). Structural domains of transfer RNA molecules. *Science* **194**, 796-806.
- Qureshi, S. A., Khoo, B., Baumann, P., et Jackson, S. P. (1995). Molecular cloning of the transcription factor TFIIB homolog from *Sulfolobus shibatae*. *Proc Natl Acad Sci U S A* **92**, 6077-81.
- Qureshi, S. A., et Jackson, S. P. (1998). Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol Cell* **1**, 389-400.

R

- Rachel, R., Wyszchony, I., Riehl, S., et Huber, H. (2002). The ultrastructure of *Ignicoccus*: evidence for a novel outer membrane and for intracellular vesicle budding in an archaeon. *Archaea* **1**, 9-18.
- Ramamurthy, V., Swann, S. L., Paulson, J. L., Spedaliere, C. J., et Mueller, E. G. (1999). Critical aspartic acid residues in pseudouridine synthases. *J Biol Chem* **274**, 22225-30.
- Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., et Claverie, J. M. (2004). The 1.2-megabase genome sequence of *Mimivirus*. *Science* **306**, 1344-50.
- Rashid, R., Aittaleb, M., Chen, Q., Spiegel, K., Demeler, B., et Li, H. (2003). Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles. *J. Mol. Biol.* **333**, 295-306.
- Rashid, R., Liang, B., Baker, D. L., Youssef, O. A., He, Y., Phipps, K., Terns, R. M., Terns, M. P., et Li, H. (2006). Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol Cell* **21**, 249-60.
- Rasmussen, A. A., Eriksen, M., Gilany, K., Udesen, C., Franch, T., Petersen, C., et Valentin-Hansen, P. (2005). Regulation of ompA mRNA stability: the role of a small regulatory RNA in growth phase-dependent control. *Mol Microbiol* **58**, 1421-9.
- Ray, W. C., Munson, R. S., Jr., et Daniels, C. J. (2001). Tricross : using dot-plots in sequence-id space to detect uncataloged intergenic features. *Bioinformatics* **17**, 1105-12.
- Raychaudhuri, S., Conrad, J., Hall, B. G., et Ofengand, J. (1998). A pseudouridine synthase required for the formation of two universally conserved pseudouridines in ribosomal RNA is essential for normal growth of *Escherichia coli*. *RNA* **4**, 1407-1417.
- Raychaudhuri, S., Niu, L., Conrad, J., Lane, B. G., et Ofengand, J. (1999). Functional effect of deletion and mutation of the *Escherichia coli* ribosomal RNA and tRNA pseudouridine synthase RluA. *JBC* **274**, 18880-18886.
- Reddy, R. (1988). Compilation of small RNA sequences. *Nucleic Acids Res* **16**, r71-85.
- Reddy, T. R., et Suryanarayana, T. (1989). Archaeobacterial histone-like proteins. Purification and characterization of helix stabilizing DNA binding proteins from the acidothermophile *Sulfolobus acidocaldarius*. *J Biol Chem* **264**, 17298-308.
- Redlak, M., Andraos-Selim, C., Giege, R., Florentz, C., et Holmes, W. M. (1997). Interaction of tRNA with tRNA (guanosine-1)methyltransferase: binding specificity determinants involve the dinucleotide G36pG37 and tertiary structure. *Biochemistry* **36**, 8699-709.
- Rees, H. C., Grant, W. D., Jones, B. E., et Heaphy, S. (2004). Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods. *Extremophiles* **8**, 63-71.
- Regalia, M., Rosenblad, M. A., et Samuelsson, T. (2002). Prediction of signal recognition particle RNA genes. *Nucleic Acids Res* **30**, 3368-77.
- Reichow, S. L., Hama, T., Ferre-D'Amare, A. R., et Varani, G. (2007). The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res* **35**, 1452-64.
- Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., et Bartel, D. P. (2002). MicroRNAs in plants. *Genes Dev* **16**, 1616-26.
- Reizer, J., Grossowicz, N., et Barenholz, Y. (1985). The effect of growth temperature on the thermotropic behavior of the membranes of a thermophilic *Bacillus*. Composition-structure-function relationships. *Biochim Biophys Acta* **815**, 268-80.

- Renalier, M. H., Joseph, N., Gaspin, C., Thebault, P., et Mougin, A. (2005). The Cm56 tRNA modification in archaea is catalyzed either by a specific 2'-O-methylase, or a C/D sRNP. *RNA* **11**, 1051-63.
- Renfrow, M. B., Naryshkin, N., Lewis, L. M., Chen, H. T., Ebright, R. H., et Scott, R. A. (2004). Transcription factor B contacts promoter DNA near the transcription start site of the archaeal transcription initiation complex. *J Biol Chem* **279**, 2825-31.
- Repoila, F., Majdalani, N., et Gottesman, S. (2003). Small non-coding RNAs, co-ordinators of adaptation processes in *Escherichia coli*: the RpoS paradigm. *Mol Microbiol* **48**, 855-61.
- Reysenbach, A. L., Liu, Y., Banta, A. B., Beveridge, T. J., Kirshtein, J. D., Schouten, S., Tivey, M. K., Von Damm, K. L., et Voytek, M. A. (2006). A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. *Nature* **442**, 444-7.
- Richard, D. J., Bell, S. D., et White, M. F. (2004). Physical and functional interaction of the archaeal single-stranded DNA-binding protein SSB with RNA polymerase. *Nucleic Acids Res* **32**, 1065-74.
- Richard, P., Darzacq, X., Bertrand, E., Jady, B. E., Verheggen, C., et Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *Embo J* **22**, 4283-93.
- Ridanpaa, M., van Eenennaam, H., Pelin, K., Chadwick, R., Johnson, C., Yuan, B., vanVenrooij, W., Pruijn, G., Salmela, R., Rockas, S., et al. (2001). Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. *Cell* **104**, 195-203.
- Rivas, E., et Eddy, S. R. (2000). Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics* **16**, 583-605.
- Rivas, E., et Eddy, S. R. (2001). Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* **2**, 8.
- Rivas, E., Klein, R. J., Jones, T. A., et Eddy, S. R. (2001). Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol* **11**, 1369-73.
- Rivera, M. C., Jain, R., Moore, J. E., et Lake, J. A. (1998). Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci U S A* **95**, 6239-44.
- Robertson, C. E., Harris, J. K., Spear, J. R., et Pace, N. R. (2005). Phylogenetic diversity and ecology of environmental Archaea. *Curr Opin Microbiol* **8**, 638-42.
- Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., BeRNander, R., et Bell, S. D. (2004). Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* **116**, 25-38.
- Robinson, N. P., et Bell, S. D. (2005). Origins of DNA replication in the three domains of life. *Febs J* **272**, 3757-66.
- Roll-Mecak, A., Shin, B. S., Dever, T. E., et Burley, S. K. (2001). Engaging the ribosome: universal IFs of translation. *Trends Biochem Sci* **26**, 705-9.
- Romier, C., Ficner, R., et Suck, D. (1998). Structural basis of base exchange by tRNA-guanine transglycosylases. In "The modification and editing of RNA" (H. Grosjean et R. Benne, Eds.), pp. 169-182. ASM Press, Washington, DC.
- Roovers, M., Hale, C., Tricot, C., Terns, M. P., Terns, R. M., Grosjean, H., et Droogmans, L. (2006). Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic Acids Res* **34**, 4293-301.
- Rosenshine, I., Tchelet, R., et Mevarech, M. (1989). The mechanism of DNA transfer in the mating system of an archaebacterium. *Science* **245**, 1387-9.
- Rother, M., et Metcalf, W. W. (2005). Genetic technologies for Archaea. *Curr Opin Microbiol* **8**, 745-51.
- Rould, M. A., Perona, J. J., et Steitz, T. A. (1991). Structural basis of anticodon loop recognition by glutamyl-tRNA synthetase. *Nature* **352**, 213-8.
- Rowlands, T., Baumann, P., et Jackson, S. P. (1994). The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. *Science* **264**, 1326-9.
- Rozenski, J., Crain, P. F., et McCloskey, J. A. (1999). The RNA Modification Database: 1999 update. *Nucleic Acids Res* **27**, 196-7.
- Rozhdestvensky, T. S., Tang, T. H., Tchirkova, I. V., Brosius, J., Bachellerie, J. P., et Huttenhofer, A. (2003). Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res* **31**, 869-77.
- Rubin, G. M. (1973). The nucleotide sequence of *Saccharomyces cerevisiae* 5.8 S ribosomal ribonucleic acid. *J Biol Chem* **248**, 3860-75.
- Ruepp, A., Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. N., et Baumeister, W. (2000). The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508-13.
- Ruggero, D., Grisendi, S., Piazza, F., Rego, E., Mari, F., Rao, P. H., Cordon-Cardo, C., et Pandolfi, P. P. (2003). Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. *Science* **299**, 259-62.
- Russell, A. G., Schnare, M. N., et Gray, M. W. (2004). Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis*. *RNA* **10**, 1034-46.
- Russell, N., et Fukunaga, N. (1990). A comparison of thermal adaptation of membrane lipids in psychrophilic and thermophilic bacteria. *FEMS Microbiol. Rev.* **75**, 171-182.

S

- Saetrom, P., Sneve, R., Kristiansen, K. I., Snove, O., Jr., Grunfeld, T., Rognes, T., et Seeberg, E. (2005). Predicting non-coding RNA genes in *Escherichia coli* with boosted genetic programming. *Nucleic Acids Res* **33**, 3263-70.
- Saito, R., et Tomita, M. (1999). Computer analyses of complete genomes suggest that some archaeobacteria employ both eukaryotic and eubacterial mechanisms in translation initiation. *Gene* **238**, 79-83.
- Samuel, B. S., Hansen, E. E., Manchester, J. K., Coutinho, P. M., Henrissat, B., Fulton, R., Latreille, P., Kim, K., Wilson, R. K., et Gordon, J. I. (2007). Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc Natl Acad Sci U S A* **104**, 10643-8.
- Samuelsson, T., et Olsson, M. (1990). Transfer RNA pseudouridine synthases in *Saccharomyces cerevisiae*. *J Biol Chem* **265**, 8782-8787.
- Sanbonmatsu, K. Y., Joseph, S., et Tung, C. S. (2005). Simulating movement of tRNA into the ribosome during decoding. *Proc Natl Acad Sci U S A* **102**, 15854-9.
- Sandman, K., et Reeve, J. N. (2005). Archaeal chromatin proteins: different structures but common function? *Curr Opin Microbiol* **8**, 656-61.
- Sandman, K., et Reeve, J. N. (2006). Archaeal histones and the origin of the histone fold. *Curr Opin Microbiol* **9**, 520-5.
- Santangelo, T. J., et Reeve, J. N. (2006). Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J Mol Biol* **355**, 196-210.
- Santangelo, T. J., Cubonova, L., James, C. L., et Reeve, J. N. (2007). TFB1 or TFB2 Is Sufficient for *Thermococcus kodakaraensis* Viability and for Basal Transcription in Vitro. *J Mol Biol* **367**, 344-57.
- Sapp, J. (2005). The prokaryote-eukaryote dichotomy: meanings and mythology. *Microbiol Mol Biol Rev* **69**, 292-305.
- Schattner, P. (2002). Searching for RNA genes using base-composition statistics. *Nucleic Acids Res* **30**, 2076-82.
- Schattner, P., Decatur, W. A., Davis, C. A., Ares, M., Jr., Fournier, M. J., et Lowe, T. M. (2004). Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res* **32**, 4281-96.
- Schattner, P., Brooks, A. N., et Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* **33**, W686-9.
- Scherer, P. A., Vollmer, G. R., Fakhouri, T., et Martensen, S. (2000). Development of a methanogenic process to degrade exhaustively the organic fraction of municipal "grey waste" under thermophilic and hyperthermophilic conditions. *Water Sci Technol* **41**, 83-91.
- Schiraldi, C., Giuliano, M., et De Rosa, M. (2002). Perspectives on biotechnological applications of archaea. *Archaea* **1**, 75-86.
- Schleper, C., Jurgens, G., et Jonuscheit, M. (2005). Genomic studies of uncultivated archaea. *Nat Rev Microbiol* **3**, 479-88.
- Schoemaker, R. J., et Gulyaev, A. P. (2006). Computer simulation of chaperone effects of Archaeal C/D box sRNA binding on rRNA folding. *Nucleic Acids Res* **34**, 2015-26.
- Schultz, A., Nottrott, S., Watkins, N. J., et Luhrmann, R. (2006). Protein-protein and protein-RNA contacts both contribute to the 15.5K-mediated assembly of the U4/U6 snRNP and the box C/D snoRNPs. *Mol Cell Biol* **26**, 5146-54.
- Schultz, D. W., et Yarus, M. (1994a). tRNA structure and ribosomal function. I. tRNA nucleotide 27-43 mutations enhance first position wobble. *J Mol Biol* **235**, 1381-94.
- Schultz, D. W., et Yarus, M. (1994b). tRNA structure and ribosomal function. II. Interaction between anticodon helix and other tRNA mutations. *J Mol Biol* **235**, 1395-405.
- Selinger, D. W., Cheung, K. J., Mei, R., Johansson, E. M., Richmond, C. S., Blattner, F. R., Lockhart, D. J., et Church, G. M. (2000). RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat Biotechnol* **18**, 1262-8.
- Senger, B., Auxilien, S., Englisch, U., Cramer, F., et Fasiolo, F. (1997). The modified wobble base inosine in yeast tRNA^{Ile} is a positive determinant for aminoacylation by isoleucyl-tRNA synthetase. *Biochemistry* **36**, 8269-75.
- Serebrov, V., Vassilenko, K., Kholod, N., Gross, H. J., et Kisselev, L. (1998). Mg²⁺ binding and structural stability of mature and in vitro synthesized unmodified *Escherichia coli* tRNA^{Phe}. *Nucleic Acids Res* **26**, 2723-8.
- Sharp, P. A. (1991). "Five easy pieces". *Science* **254**, 663.
- She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., et al. (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* **98**, 7835-40.
- Shi, H., et Moore, P. B. (2000). The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **6**, 1091-105.
- Shibata, R., Bessho, Y., Shinkai, A., Nishimoto, M., Fusatomi, E., Terada, T., Shirouzu, M., et Yokoyama, S. (2007). Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. *Biochem Biophys Res Commun* **355**, 122-8.

- Shin, J. H., Jiang, Y., Grabowski, B., Hurwitz, J., et Kelman, Z. (2003). Substrate requirements for duplex DNA translocation by the eukaryal and archaeal minichromosome maintenance helicases. *J Biol Chem* **278**, 49053-62.
- Shine, J., et Dalgarno, L. (1974). The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A* **71**, 1342-1346.
- Simos, G., Tekotte, H., Grosjean, H., Segref, A., Sharma, K., Tollervey, D., et Hurt, E. C. (1996). Nuclear pore proteins are involved in the biogenesis of functional tRNA. *EMBO Journal* **15**, 2270-2284.
- Singh, S. K., Gurha, P., Tran, E. J., Maxwell, E. S., et Gupta, R. (2004). Sequential 2'-O-methylation of archaeal pre-tRNA^{Trp} nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pre-tRNA. *J Biol Chem* **279**, 47661-71.
- Sivaraman, J., Sauve, V., Larocque, R., Stura, E. A., Schrag, J. D., Cygler, M., et Matte, A. (2002). Structure of the 16S rRNA pseudouridine synthase RsuA bound to uracil and UMP. *Nat Struct Biol* **9**, 353-8.
- Sivaraman, J., Iannuzzi, P., Cygler, M., et Matte, A. (2004). Crystal structure of the RluD pseudouridine synthase catalytic module, an enzyme that modifies 23S rRNA and is essential for normal cell growth of Escherichia coli. *J Mol Biol* **335**, 87-101.
- Sledjeski, D. D., Gupta, A., et Gottesman, S. (1996). The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in Escherichia coli. *Embo J* **15**, 3993-4000.
- Slesarev, A. I., Mezhevaya, K. V., Makarova, K. S., Polushin, N. N., Shcherbinina, O. V., Shakhova, V. V., Belova, G. I., Aravind, L., Natale, D. A., Rogozin, I. B., et al. (2002). The complete genome of hyperthermophile Methanopyrus kandleri AV19 and monophyly of archaeal methanogens. *Proc Natl Acad Sci U S A* **99**, 4644-9.
- Slupska, M. M., King, A. G., Fitz-Gibbon, S., Besemer, J., Borodovsky, M., et Miller, J. H. (2001). Leaderless transcripts of the crenarchaeal hyperthermophile Pyrobaculum aerophilum. *J Mol Biol* **309**, 347-60.
- Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., et al. (1997). Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics. *J Bacteriol* **179**, 7135-55.
- Soppa, J. (1999a). Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box. *Mol Microbiol* **31**, 1589-92.
- Soppa, J. (1999b). Transcription initiation in Archaea: facts, factors and future aspects. *Mol Microbiol* **31**, 1295-305.
- Speckmann, W. A., Terns, R. M., et Terns, M. P. (2000). The box C/D motif directs snoRNA 5'-cap hypermethylation. *Nucleic Acids Res* **28**, 4467-73.
- Spedaliere, C. J., Hamilton, C. S., et Mueller, E. G. (2000). Functional importance of motif I of pseudouridine synthases: mutagenesis of aligned lysine and proline residues. *Biochemistry* **39**, 9459-65.
- Spedaliere, C. J., et Mueller, E. G. (2004). Not all pseudouridine synthases are potently inhibited by RNA containing 5-fluorouridine. *RNA* **10**, 192-9.
- Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., et Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **26**, 148-53.
- Sprott, G. D., Meloche, M., et Richards, J. C. (1991). Proportions of diether, macrocyclic diether, and tetraether lipids in Methanococcus jannaschii grown at different temperatures. *J Bacteriol* **173**, 3907-10.
- Starostina, N. G., Marshburn, S., Johnson, L. S., Eddy, S. R., Terns, R. M., et Terns, M. P. (2004). Circular box C/D RNAs in Pyrococcus furiosus. *Proc Natl Acad Sci U S A* **101**, 14097-101.
- Steinberg, D. (1991). Antioxidants and atherosclerosis. A current assessment. *Circulation* **84**, 1420-5.
- Stetter, K. O. (2006). History of discovery of the first hyperthermophiles. *Extremophiles* **10**, 357-62.
- Stevens, S. W., et Abelson, J. (1999). Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc Natl Acad Sci U S A* **96**, 7226-31.
- Stillier, J. W., et Hall, B. D. (2002). Evolution of the RNA polymerase II C-terminal domain. *Proc Natl Acad Sci U S A* **99**, 6091-6.
- Storz, G. (2002). An expanding universe of noncoding RNAs. *Science* **296**, 1260-3.
- Storz, G., Opdyke, J. A., et Zhang, A. (2004). Controlling mRNA stability and translation with small, noncoding RNAs. *Curr Opin Microbiol* **7**, 140-4.
- Stoughton, R. B. (2005). Applications of DNA microarrays in biology. *Annu Rev Biochem* **74**, 53-82.
- Stuart, K. D., Schnauffer, A., Ernst, N. L., et Panigrahi, A. K. (2005). Complex management: RNA editing in trypanosomes. *Trends Biochem Sci* **30**, 97-105.
- Suckow, J. M., Amano, N., Ohfuku, Y., Kakinuma, J., Koike, H., et Suzuki, M. (1998). A transcription frame-based analysis of the genomic DNA sequence of a hyper-thermophilic archaeon for the identification of genes, pseudogenes and operon structures. *FEBS Lett* **426**, 86-92.
- Sumita, M., Desaulniers, J. P., Chang, Y. C., Chui, H. M., Clos, L., 2nd, et Chow, C. S. (2005). Effects of nucleotide substitution and modification on the stability and structure of helix 69 from 28S rRNA. *RNA* **11**, 1420-9.
- Sunita, S., Zhenxing, H., Swaathi, J., Cygler, M., Matte, A., et Sivaraman, J. (2006). Domain organization and crystal structure of the catalytic domain of E.coli RluF, a pseudouridine synthase that acts on 23S rRNA. *J Mol Biol* **359**, 998-1009.
- Suryadi, J., Tran, E. J., Maxwell, E. S., et Brown, B. A., 2nd. (2005). The crystal structure of the Methanocaldococcus jannaschii multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the box C/D RNAs. *Biochemistry* **44**, 9657-72.

- Svobodova, J., et Svoboda, P. (1988). Membrane fluidity in *Bacillus subtilis*. Physical change and biological adaptation. *Folia Microbiol (Praha)* **33**, 161-9.
- Szewczak, L. B., DeGregorio, S. J., Strobel, S. A., et Steitz, J. A. (2002). Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem. Biol.* **9**, 1095-1107.
- Szymanski, M., Barciszewska, M. Z., Barciszewski, J., et Erdmann, V. A. (2000). 5S ribosomal RNA database Y2K. *Nucleic Acids Res* **28**, 166-7.
- Szymanski, M., Erdmann, V. A., et Barciszewski, J. (2003). Noncoding regulatory RNAs database. *Nucleic Acids Res* **31**, 429-31.

T

- Taft, R. J., Pheasant, M., et Mattick, J. S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288-99.
- Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., Elge, T., Brosius, J., et Huttenhofer, A. (2002a). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc Natl Acad Sci U S A* **99**, 7536-41.
- Tang, T. H., Rozhdestvensky, T. S., d'Orval, B. C., Bortolin, M. L., Huber, H., Charpentier, B., Branlant, C., Bachellerie, J. P., Brosius, J., et Huttenhofer, A. (2002b). RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucl. Acids Res.* **30**, 921-930.
- Tang, T. H., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., Bachellerie, J. P., et Huttenhofer, A. (2005). Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol Microbiol* **55**, 469-81.
- Tarn, W. Y., et Steitz, J. A. (1997). Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci* **22**, 132-7.
- Terns, M. P., et Terns, R. M. (2001). Macromolecular complexes: SMN--the master assembler. *Curr Biol* **11**, R862-4.
- Terns, M. P., et Terns, R. M. (2002). Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr.* **10**, 17-39.
- Thebault, P., de Givry, S., Schiex, T., et Gaspin, C. (2006). Searching RNA motifs and their intermolecular contacts with constraint networks. *Bioinformatics* **22**, 2074-80.
- Thisted, T., Sorensen, N. S., Wagner, E. G., et Gerdes, K. (1994). Mechanism of post-segregational killing: Sok antisense RNA interacts with Hok mRNA via its 5'-end single-stranded leader and competes with the 3'-end of Hok mRNA for binding to the mok translational initiation region. *Embo J* **13**, 1960-8.
- Thomm, M., Hausner, W., et Hethke, C. (1994). Transcription factors and termination of transcription in *Methanococcus*. *Syst Appl Microbiol* **16**, 648-655.
- Thompson, D., Wong, K., Humphry-Baker, R., Wheeler, J., Kim, J., et Ranavavare, S. (1992). Tetraether bolaform amphiphiles as models of archaebacterial membrane lipids: Raman spectroscopy, ³¹P NMR, X-ray scattering, and electron microscopy. *J. Am.Chem. Soc.* **114**, 9035-9042.
- Thompson, J. D., Higgins, D. G., et Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80.
- Thore, S., Mayer, C., Sauter, C., Weeks, S., et Suck, D. (2003). Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. *J Biol Chem* **278**, 1239-47.
- Tjaden, B., Saxena, R. M., Stolyar, S., Haynor, D. R., Kolker, E., et Rosenow, C. (2002). Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res* **30**, 3732-8.
- Toledo-Arana, A., Repoila, F., et Cossart, P. (2007). Small noncoding RNAs controlling pathogenesis. *Curr Opin Microbiol* **10**, 182-8.
- Tollervey, D. (1987). A yeast small nuclear RNA is required for normal processing of pre- ribosomal RNA. *Embo J* **6**, 4169-75.
- Tollervey, D., et Kiss, T. (1997). Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell Biol.* **9**, 337-342.
- Torarinsson, E., Klenk, H. P., et Garrett, R. A. (2005). Divergent transcriptional and translational signals in Archaea. *Environ Microbiol* **7**, 47-54.
- Torchet, C., Badis, G., Devaux, F., Costanzo, G., Werner, M., et Jacquier, A. (2005). The complete set of H/ACA snoRNAs that guide rRNA pseudouridylation in *Saccharomyces cerevisiae*. *RNA* **11**, 928-38.
- Tran, E. J., Zhang, X., et Maxwell, E. S. (2003). Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J.* **22**, 3930-3940.
- Tran, E., Brown, J., et Maxwell, E. S. (2004). Evolutionary origins of the RNA-guided nucleotide-modification complexes: from the primitive translation apparatus? *Trends Biochem Sci* **29**, 343-50.
- Tran, E., Zhang, X., Lackey, L., et Maxwell, E. S. (2005). Conserved spacing between the box C/D and C'/D' RNPs of the archaeal box C/D sRNP complex is required for efficient 2'-O-methylation of target RNAs. *RNA* **11**, 285-93.

- Tremblay, A., Lamontagne, B., Catala, M., Yam, Y., Larose, S., Good, L., et Elela, S. A. (2002). A physical interaction between Gar1p and Rnt1pi is required for the nuclear import of H/ACA small nucleolar RNA-associated proteins. *Mol Cell Biol* **22**, 4792-802.
- Tsui, H. C., Arps, P. J., Connolly, D. M., et Winkler, M. E. (1991). Absence of hisT-mediated tRNA pseudouridylation results in a uracil requirement that interferes with *Escherichia coli* K-12 cell division. *J Bacteriol* **173**, 7395-7400.
- Tworowska, I., et Nikonowicz, E. P. (2006). Base pairing within the psi32,psi39-modified anticodon arm of *Escherichia coli* tRNA(Phe). *J Am Chem Soc* **128**, 15570-1.
- Tycowski, K. T., Aab, A., et Steitz, J. A. (2004). Guide RNAs with 5' caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Curr Biol* **14**, 1985-95.

U

- Uliel, S., Liang, X. H., Unger, R., et Michaeli, S. (2004). Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organisation, and unique functions. *Int J Parasitol* **34**, 445-54.
- Unniraman, S., Prakash, R., et Nagaraja, V. (2001). AlterNate paradigm for intrinsic transcription termination in eubacteria. *J Biol Chem* **276**, 41850-5.

V

- Valadkhan, S. (2005). snRNAs as the catalysts of pre-mRNA splicing. *Curr Opin Chem Biol* **9**, 603-8.
- van den Burg, B. (2003). Extremophiles as a source for novel enzymes. *Curr Opin Microbiol* **6**, 213-8.
- van Hoek, A. H., van Alen, T. A., Sprakel, V. S., Leunissen, J. A., Brigge, T., Vogels, G. D., et Hackstein, J. H. (2000). Multiple acquisition of methanogenic archaeal symbionts by anaerobic ciliates. *Mol Biol Evol* **17**, 251-8.
- Vapnik, V. N. (1998). Statistical learning theory. *John Wiley & Sons, Inc., N.Y.*
- Vassilyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassilyeva, M. N., Borukhov, S., et Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**, 712-9.
- Venema, J., et Tollervey, D. (1999). Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu Rev Genet* **33**, 261-311.
- Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I., et Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672-6.
- Vianna, M. E., Conrads, G., Gomes, B. P., et Horz, H. P. (2006). Identification and quantification of archaea involved in primary endodontic infections. *J Clin Microbiol* **44**, 1274-82.
- Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R., et Ficner, R. (2000). Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell*, **6**, 1331-1342.
- Villa, T., Ceradini, F., et Bozzoni, I. (2000). Identification of a novel element required for processing of intron-encoded box C/D small nucleolar RNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol* **20**, 1311-20.
- Vitali, P., Royo, H., Seitz, H., Bachellerie, J. P., Huttenhofer, A., et Cavaille, J. (2003). Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res* **31**, 6543-51.
- Vivona, J. B., et Kelman, Z. (2003). The diverse spectrum of sliding clamp interacting proteins. *FEBS Lett* **546**, 167-72.
- Vogel, J., Bartels, V., Tang, T. H., Churakov, G., Slagter-Jager, J. G., Huttenhofer, A., et Wagner, E. G. (2003). RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* **31**, 6435-43.
- Vogel, J., et Sharma, C. M. (2005). How to find small non-coding RNAs in bacteria. *Biol Chem* **386**, 1219-38.
- Vogels, G. D., Hoppe, W. F., et Stumm, C. K. (1980). Association of methanogenic bacteria with rumen ciliates. *Appl Environ Microbiol* **40**, 608-12.
- Vorobyova, E., Soina, V., Gorlenko, M., Zalinuva, N., Mamukelashvili, A., Gilichinsky, D., Rivkina, E., et Vishnevetskaya, T. (1997). The deep cold biosphere: facts and hypothesis. *FEMS Microbiol Rev* **20**, 277-290.
- Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P. J., et Dokal, I. (2001). The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* **413**, 432-5.
- Vulliamy, T. J., Marrone, A., Knight, S. W., Walne, A., Mason, P. J., et Dokal, I. (2006). Mutations in dyskeratosis congenita: their impact on telomere length and the diversity of clinical presentation. *Blood* **107**, 2680-5.

W

- Wagner, E. G., et Flardh, K. (2002). Antisense RNAs everywhere? *Trends Genet* **18**, 223-6.
- Wagner, E. G., et Simons, R. W. (1994). Antisense RNA control in bacteria, phages, and plasmids. *Annu Rev Microbiol* **48**, 713-42.
- Walne, A. J., Vulliamy, T., Marrone, A., Beswick, R., Kirwan, M., Masunari, Y., Al-Qurashi, F. H., Aljurf, M., et Dokal, I. (2007). Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. *Hum Mol Genet* **16**, 1619-29.

- Wang, C., Ding, C., Meraz, R. F., et Holbrook, S. R. (2006). PSOL: a positive sample only learning algorithm for finding non-coding RNA genes. *Bioinformatics* **22**, 2590-6.
- Wang, C., et Meier, U. T. (2004). Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *Embo J* **23**, 1857-67.
- Wang, H., Boisvert, D., Kim, K. K., Kim, R., et Kim, S. H. (2000). Crystal structure of a fibrillarin homologue from *Methanococcus jannaschii*, a hyperthermophile, at 1.6 Å resolution. *EMBO J.* **19**, 317-323.
- Warbrick, E. (2000). The puzzle of PCNA's many partners. *Bioessays* **22**, 997-1006.
- Washietl, S., et Hofacker, I. L. (2004). Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. *J Mol Biol* **342**, 19-30.
- Washietl, S., Hofacker, I. L., Lukasser, M., Huttenhofer, A., et Stadler, P. F. (2005a). Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol* **23**, 1383-90.
- Washietl, S., Hofacker, I. L., et Stadler, P. F. (2005b). Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A* **102**, 2454-9.
- Wassarman, K. M., Zhang, A., et Storz, G. (1999). Small RNAs in *Escherichia coli*. *Trends Microbiol* **7**, 37-45.
- Wassarman, K. M., et Storz, G. (2000). 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* **101**, 613-23.
- Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G., et Gottesman, S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* **15**, 1637-51.
- Wassarman, K. M. (2002). Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell* **109**, 141-4.
- Wassarman, K. M. (2007). 6S RNA: a small RNA regulator of transcription. *Curr Opin Microbiol* **10**, 164-8.
- Watanabe, Y., et Gray, M. W. (2000). Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaeobacteria. *Nucleic Acids Res* **28**, 2342-52.
- Watanabe, Y., Yokobori, S., Inaba, T., Yamagishi, A., Oshima, T., Kawarabayasi, Y., Kikuchi, H., et Kita, K. (2002). Introns in protein-coding genes in Archaea. *FEBS Lett* **510**, 27-30.
- Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., et al. (2003). The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci U S A* **100**, 12984-8.
- Watkins, N. J., Gottschalk, A., Neubauer, G., Kastner, B., Fabrizio, P., Mann, M., et Luhrmann, R. (1998). Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA*, **4**, 1549-1568.
- Watkins, N. J., Leverette, R. D., Xia, L., Andrews, M. T., et Maxwell, E. S. (1996). Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA* **2**, 118-33.
- Watkins, N. J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C., et Luhrmann, R. (2000). A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, **103**, 457-466.
- Watkins, N. J., Dickmanns, A., et Luhrmann, R. (2002). Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol. Cell. Biol.* **22**, 8342-8352.
- Weinstein, L. B., et Steitz, J. A. (1999). Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol* **11**, 378-84.
- Westhof, E., Dumas, P., et Moras, D. (1985). Crystallographic refinement of yeast aspartic acid transfer RNA. *J Mol Biol* **184**, 119-45.
- Westhof, E., Dumas, P., et Moras, D. (1988). Hydration of transfer RNA molecules: a crystallographic study. *Biochimie* **70**, 145-65.
- White, M. F., et Bell, S. D. (2002). Holding it together: chromatin in the Archaea. *Trends Genet* **18**, 621-6.
- Wilderman, P. J., Sowa, N. A., FitzGerald, D. J., FitzGerald, P. C., Gottesman, S., Ochsner, U. A., et Vasil, M. L. (2004). Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc Natl Acad Sci U S A* **101**, 9792-7.
- Will, C. L., et Luhrmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr Opin Cell Biol* **9**, 320-8.
- Will, C. L., et Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol* **13**, 290-301.
- Willkomm, D. K., Minnerup, J., Huttenhofer, A., et Hartmann, R. K. (2005). Experimental RNomics in *Aquifex aeolicus*: identification of small non-coding RNAs and the putative 6S RNA homolog. *Nucleic Acids Res* **33**, 1949-60.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonnrhein, C., Hartsch, T., et Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* **407**, 327-39.
- Winkler, A. A., Bobok, A., Zonneveld, B. J., Steensma, H. Y., et Hooykaas, P. J. (1998). The lysine-rich C-terminal repeats of the centromere-binding factor 5 (Cbf5) of *Kluyveromyces lactis* are not essential for function. *Yeast* **14**, 37-48.

- Wise, J. A., Tollervey, D., Maloney, D., Swerdlow, H., Dunn, E. J., et Guthrie, C. (1983). Yeast contains small nuclear RNAs encoded by single copy genes. *Cell* **35**, 743-751.
- Woese, C. R. (1968). The fundamental nature of the genetic code: prebiotic interactions between polynucleotides and polyamino acids or their derivatives. *Proc Natl Acad Sci U S A* **59**, 110-7.
- Woese, C. R., et Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**, 5088-5090.
- Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**, 221-71.
- Woese, C. R., Kandler, O., et Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**, 4576-9.
- Woese, C. R. (2004). The Archaeal Concept and the World it Lives in: A Retrospective. *Photosynth Res* **80**, 361-72.
- Wrzesinski, J., Bakin, A., Nurse, K., Lane, B. G., et Ofengand, J. (1995). Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from *Escherichia coli*. *Biochemistry* **34**, 8904-8913.
- Wu, H., et Feigon, J. (2007). H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *Proc Natl Acad Sci U S A* **104**, 6655-60.
- Wuchter, C., Abbas, B., Coolen, M. J., Herfort, L., van Bleijswijk, J., Timmers, P., Strous, M., Teira, E., Herndl, G. J., Middelburg, J. J., et al. (2006). Archaeal nitrification in the ocean. *Proc Natl Acad Sci U S A* **103**, 12317-22.

X

- Xia, L., Watkins, N. J., et Maxwell, E. S. (1997). Identification of specific nucleotide sequences and structural elements required for intronic U14 snoRNA processing. *RNA* **3**, 17-26.
- Xie, J., Zhang, M., Zhou, T., Hua, X., Tang, L., et Wu, W. (2007). Sno/scaRNAbase: a curated database for small nucleolar RNAs and cajal body-specific RNAs. *Nucleic Acids Res* **35**, D183-7.
- Xie, Y., et Reeve, J. N. (2004). Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *J Bacteriol* **186**, 3492-8.
- Xu, P., Vernooy, S. Y., Guo, M., et Hay, B. A. (2003). The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr Biol* **13**, 790-5.

Y

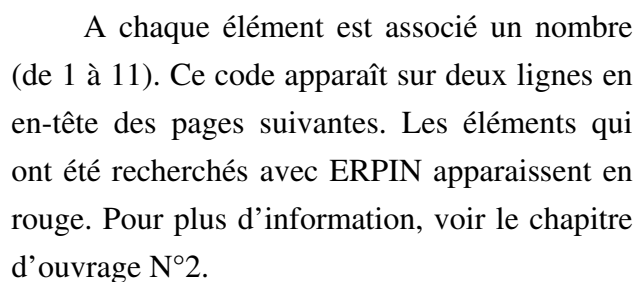
- Yang, C., McPheeters, D. S., et Yu, Y. T. (2005a). Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in *Saccharomyces cerevisiae*. *J Biol Chem* **280**, 6655-62.
- Yang, C. Y., Zhou, H., Luo, J., et Qu, L. H. (2005b). Identification of 20 snoRNA-like RNAs from the primitive eukaryote, *Giardia lamblia*. *Biochem Biophys Res Commun* **328**, 1224-31.
- Yang, P. K., Rotondo, G., Porras, T., Legrain, P., et Chanfreau, G. (2002). The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis. *J Biol Chem* **277**, 45235-42.
- Yao, Z., Weinberg, Z., et Ruzzo, W. L. (2006). CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics* **22**, 445-52.
- Yarian, C. S., Basti, M. M., Cain, R. J., Ansari, G., Guenther, R. H., Sochacka, E., Czerwinska, G., Malkiewicz, A., et Agris, P. F. (1999). Structural and functional roles of the N1- and N3-protons of psi at tRNA's position 39. *Nucleic Acids Res* **27**, 3543-9.
- Yarus, M., et Barrell, B. G. (1971). The sequence of nucleotides in tRNA Ile from *E. coli* B. *Biochem Biophys Res Commun* **43**, 729-34.
- Ye, K. (2007). H/ACA guide RNAs, proteins and complexes. *Curr Opin Struct Biol* **17**, 287-92.
- Yi, R., O'Carroll, D., Pasolli, H. A., Zhang, Z., Dietrich, F. S., Tarakhovsky, A., et Fuchs, E. (2006). Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nat Genet* **38**, 356-62.
- Youssef, O. A., Terns, R. M., et Terns, M. P. (2007). Dynamic interactions within sub-complexes of the H/ACA pseudouridylation guide RNP. *Nucleic Acids Res*.
- Yuan, G., Klamt, C., Bachellerie, J. P., Brosius, J., et Huttenhofer, A. (2003). RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res* **31**, 2495-507.
- Yue, D., Kintanar, A., et Horowitz, J. (1994). Nucleoside modifications stabilize Mg²⁺ binding in *Escherichia coli* tRNA(Val): an imino proton NMR investigation. *Biochemistry* **33**, 8905-11.

Z

- Zago, M. A., Dennis, P. P., et Omer, A. D. (2005). The expanding world of small RNAs in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Microbiol* **55**, 1812-28.
- Zebarjadian, Y., King, T., Fournier, M. J., Clarke, L., et Carbon, J. (1999). Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Mol. Cell. Biol.* **19**, 7461-7472.

- Zemann, A., op de Bekke, A., Kiefmann, M., Brosius, J., et Schmitz, J. (2006). Evolution of small nucleolar RNAs in nematodes. *Nucleic Acids Res* **34**, 2676-85.
- Zeng, Y., Yi, R., et Cullen, B. R. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci U S A* **100**, 9779-84.
- Zerfass, K., et Beier, H. (1992). Pseudouridine in the anticodon G psi A of plant cytoplasmic tRNA(Tyr) is required for UAG and UAA suppression in the TMV-specific context. *Nucleic Acids Res* **20**, 5911-8.
- Zhang, A., Wassarman, K. M., Rosenow, C., Tjaden, B. C., Storz, G., et Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol* **50**, 1111-24.
- Zhao, K., Chai, X., et Marmorstein, R. (2003). Structure of a Sir2 substrate, Alba, reveals a mechanism for deacetylation-induced enhancement of DNA binding. *J Biol Chem* **278**, 26071-7.
- Zillig, W., Tu, J., et Holz, I. (1981). Thermoproteales--a third order of thermoacidophilic archaeobacteria. *Nature* **293**, 85-6.
- Zuker, M., et Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* **9**, 133-48.
- Zuker, M., et Jacobson, A. B. (1998). Using reliability information to annotate RNA secondary structures. *RNA* **4**, 669-79.

ANNEXES



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Résumé

La conversion des résidus U en pseudouridine (Ψ) est la modification la plus fréquente dans les ARN. Cette réaction est catalysée par des ARN: Ψ -synthases qui fonctionnent soit seules, soit au sein de particules snoRNP H/ACA composées de 4 protéines et d'un snoRNA. Celui-ci assure la reconnaissance de l'ARN cible par appariement de bases. Les pseudouridylations des ARNr eucaryotes sont catalysées par des snoRNP alors que les modifications des ARNt sont catalysées par des enzymes dépourvues de guide. Aucun snoRNA H/ACA n'est décrit chez les Bactéries. Trois équivalents des snoRNA H/ACA (sRNA H/ACA) avaient été découverts chez les Archaea lorsque j'ai débuté ma thèse, mais leur contribution aux modifications des ARNr n'avait pas été démontrée. Nous avons développé une approche bioinformatique permettant l'identification exhaustive de gènes de sRNA H/ACA et de leurs cibles. Nous avons ainsi identifié 7 sRNA H/ACA chez *Pyrococcus abyssi* et détecté 14 Ψ dans les ARNr. Nous avons réussi la reconstitution de particules H/ACA actives *in vitro*, ce qui nous a permis d'étudier leur mode d'assemblage et d'assigner 12 des Ψ détectées aux 7 sRNA. L'application à d'autres espèces d'archaea de l'approche bioinformatique développée nous a permis d'identifier environ 90 sRNA, nous avons ainsi pu étudier leur évolution. De plus, nous avons expérimentalement démontré pour la première fois qu'un sRNA H/ACA est impliqué dans la modification d'un ARNt. Enfin, nous avons montré que l'enzyme aCBF5 des sRNP H/ACA peut catalyser la pseudouridylation en position 55 des ARNt sans utilisation de sRNA et nous avons identifié des résidus de aCBF5 impliqués dans les activités guidée et/ou non guidée.

Mots-Clés : archaea, *Pyrococcus abyssi*, sRNA H/ACA, pseudouridine, ARN guide, ARN: Ψ -synthase

Title : Development and application of computational methods dedicated to the identification of H/ACA box sRNAs within archaeal genomes and functional study of the corresponding RNAs and H/ACA sRNPs.

Summary

The conversion of U into Pseudouridine (Ψ) residue is the most prevalent RNA modification. This reaction is catalysed either by stand-alone RNA: Ψ -synthases or by H/ACA snoRNPs composed of 4 proteins and a snoRNA. In snoRNPs the recognition of the U residue is carried out by base-pair interactions between the snoRNA and the RNA target. SnoRNPs are involved in eukaryal rRNA modifications whereas stand-alone RNA: Ψ -synthase are involved in tRNA and bacterial rRNA modifications. Three counterparts of the H/ACA snoRNAs had been discovered in Archaea when I started my PhD work, but their contribution to rRNA modification was not demonstrated. We developed a computational approach to identify the H/ACA sRNA genes and their rRNA targets. Its application to the *Pyrococcus abyssi* genome revealed 7 candidate genes. We verified their expression and experimentally identified 14 Ψ in the *P. abyssi* rRNAs. We succeeded in the *in vitro* reconstitution of active H/ACA sRNPs and thereby studied the sRNP assembly mechanism, identified important H/ACA sRNA structural features and assigned 12 of the identified Ψ residues to the 7 sRNAs. The application of the computational approach to the other archaeal genomes revealed 90 sRNA genes. This repertoire shed new light on the function and the evolution of this ncRNA class. In addition, we experimentally demonstrated for the first time that a few H/ACA sRNAs are involved in tRNA modification. We also showed that the enzyme of the H/ACA sRNP (aCBF5) is involved in Ψ 55 formation in all archaeal tRNAs without the use of an H/ACA sRNA. We compared the guided and non-guided CBF5 activities and identified aminoacids which are important for these activities.

Key Words: archaea, *Pyrococcus abyssi*, H/ACA sRNA, pseudouridine, RNA guide, RNA: Ψ -synthase.