



HAL
open science

Caractérisation structurale et fonctionnelle d'une nouvelle interaction entre les protéines RPAP3 et TRBP

Yoann Abel

► **To cite this version:**

Yoann Abel. Caractérisation structurale et fonctionnelle d'une nouvelle interaction entre les protéines RPAP3 et TRBP. Biologie structurale [q-bio.BM]. Université de Lorraine, 2016. Français. NNT : 2016LORR0285 . tel-01752415

HAL Id: tel-01752415

<https://hal.univ-lorraine.fr/tel-01752415>

Submitted on 2 Jul 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>



UNIVERSITÉ
DE LORRAINE



École Doctorale BioSE (Biologie-Santé-Environnement)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITÉ DE LORRAINE

Mention : « Sciences de la Vie et de la Santé »

par Yoann ABEL

**Caractérisation structurale et fonctionnelle d'une
nouvelle interaction entre les protéines RPAP3 et
TRBP**

Le 14 décembre 2016

Membres du jury :

Rapporteurs :	Dr Sébastien PFEFFER	Directeur de recherche, CNRS, Strasbourg
	Pr Nicolas LEULLIOT	Professeur, Université de Paris Descartes, Paris
Examineurs :	Dr Christine ALLMANG-CURA	Chargée de recherche, CNRS, Strasbourg
	Dr Natacha DREUMONT	Maître de conférence, Université de Lorraine, Nancy
	Pr Bruno CHARPENTIER	Professeur, Université de Lorraine, Nancy, Directeur de doctorat
	Dr Mathieu REDERSTORFF	Maître de conférence, Université de Lorraine, Nancy, Co-directeur de doctorat
Membre invité :	Dr Christiane BRANLANT	Directrice de recherche, Université de Lorraine, Nancy

**Laboratoire IMoPA: Ingénierie Moléculaire et Physiopathologie Articulaire
UMR 7365 CNRS-UL, Faculté de Médecine, Bâtiment Biopôle,
9 Avenue de la Forêt de Haye, BP 184, 54505 Vandoeuvre-lès-Nancy**

« L'urgent est fait, l'impossible est en cours, pour les miracles prévoir un délai »

Anonyme

Récapitulatif des principales protéines impliquées dans l'assemblage des snoRNP et dans la maturation des microARN

Protéines humaines et leurs homologues chez la levure impliquées dans l'assemblage des snoRNP :

Humain		<i>S. cerevisiae</i>	
Nom	Signification	Nom	Signification
Fibrillarine	-	Nop1	Nucleolar protein 1
NOP56	Nucleolar protein 56 kDa	Nop56	Nucleolar protein 56 kDa
NOP58	Nucleolar protein 58 kDa	Nop58	Nucleolar protein 58 kDa
SNU13/15.5	Small NUClear ribonucleoprotein-associated protein of 13 kDa	Snu13	Small NUClear ribonucleoprotein-associated protein of 13 kDa
HSP90	Heat Shock Protein of 90 kDa	Hsp82	Heat Shock Protein of 82 kDa
RPAP3	RNA PolymeraseII-Associated Protein 3	Tah1	TPR repeat-containing protein Associated with Hsp82
PIH1D1	PIH1 domain-containing protein 1	Pih1 Nop17	Protein Interaction with Hsp82 Nucleolar protein of 17 kDa
TIP48	48 kDa TATA box-binding protein-Interacting Protein	Rvb2	RuvB-like protein 2
TIP49	49 kDa TATA box-binding protein-Interacting Protein	Rvb1	RuvB-like protein 1
NUFIP	Nuclear fragile X mental retardation-interacting protein 1	Rsa1	Ribosome assembly protein 4
ZNHIT3 TRIP3	Zinc finger HIT domain-containing protein 3 Thyroid Receptor-Interacting Protein 3	Hit1	High temperature Growth 1
BCD1	Box C/D snoRNA protein 1	Bcd1	Box C/D snoRNA protein 1
Dyskérine	-	Cbf5	Centromere Binding Factor 5
NOP10	Nucleolar protein of 10 kDa	Nop10	Nucleolar protein of 10 kDa
GAR1	Glycine and Arginine Rich protein	Gar1	Glycine and Arginine Rich protein
NHP2	Non-Histone Protein 2	Nhp2	Non-Histone Protein 2
SHQ1	Small nucleolar RNAs of the box H/ACA family quantitative accumulation protein 1	Shq1	Small nucleolar RNAs of the box H/ACA family quantitative accumulation protein 1
NAF1	Nuclear Association Factor 1	Naf1	Nuclear Association Factor 1

Protéines humaines impliquées dans la maturation des microARN :

Humain	
Nom	Signification
Drosha	-
DGCR8	DiGeorge syndrome Critical Region 8
XPO5	exportine-5
Dicer	-
TRBP/TARBP2	Trans-Activation-Responsive RNA-Binding Protein
PACT	Protein ACTivator of PKR
AGO	Argonaute
GW182 TNRC6A	Glycine-Tryptophan protein of 182 kDa TriNucleotide Repeat-Containing gene 6A protein

Abréviations

%:	pourcentage
¹³ C :	noyau de carbone 13
¹⁵ N :	noyau d'azote 15
¹ H :	Proton
3-AT :	3 Amino-Triazol
³² P :	Phosphore 32
Abs :	Absorbance
ADN :	Acide DésoxyriboNucléique
ADNc :	ADN complémentaire
ADP :	Adénosine Di-phosphate
ARN :	Acide RiboNucléique
ARNm :	ARN messenger
ARNr :	ARN ribosomique
ARNt :	ARN de transfert
ATP :	Adénosine Tri-phosphate
CAB :	Cajal Bodies Box
Da/kDa :	Dalton/kiloDalton
DMEM :	Dulbecco's Modified Eagle's Medium
DNase :	Désoxynucléase
dNTP :	Désoxyribonucléotide triphosphate
DTT :	DiThioThréitol
<i>E. coli</i> :	<i>Escherichia coli</i>
EDTA :	Éthylène Diamine Tétra-Acétique
GST :	Glutathione S-Transférase
h/min/s :	heure/minute/seconde
HEK :	Human Epithelium Kidney
Hela :	cellules d'Henrietta Lacks
IPTG:	Isopropyl β-D-1-thiogalactopyranoside
K/°C :	Kelvin/degrés Celcius
lncARN :	long ARN non-codant
mCi :	milliCurie
Mhz :	mégahertz
mmol :	millimole
ng/μg/mg/g :	nanogramme/microgramme/milligramme/gramme

NLS :	Nuclear Localization Signal
nm/ μ m/cm :	nanomètre/micromètre/centimètre
nM/ μ M/M :	nanomolaire/micromolaire/molaire
NoLS :	Nucleolar Localization Signal
PBS :	Phosphate Buffered Saline
PCR :	Polymease Chain Reaction
PEI :	PolyEtherImide
PEG :	PolyÉthylène Glycol
ppm :	partie par million
RMN :	Résonance Magnétique Nucléaire
RMSD :	Root-Mean-Square Deviation
RNase :	Ribonucléase
rNTP :	Ribonucléotide triphosphate
<i>S. cerevisiae</i> :	<i>Saccharomyces cerevisiae</i>
SAM :	S-Adénosyl Méthionine
scaARN :	Small Cajal body-specific RNAs
SDS :	Sodium Dodecyl Sulfate
SDS-PAGE :	SDS-PolyAcrylamide Gel Electrophoresis
snARN :	small nuclear RNA
snoARN :	small nucleolar RNA
SVF :	Sérum de Veau Fœtal
TBE :	Tris-Borate-EDTA
TE :	Tris-EDTA
TPR :	TetratricoPeptide Repeat
U :	Unité
uDO :	unité de Densité Optique
UTR :	UnTranslated Region
v/v :	volume à volume
WT :	Wild Type
YPD :	Yeast extract-Peptone-Dextrose
μ l/ml/l :	microlitre/millilitre/litre

Sommaire

Introduction

Avant-propos.....	1
Chapitre 1 : Les snoARN.....	4
I. Les snoRNP à boîtes C/D.....	5
A. Les snoARN à boîtes C/D.....	5
1. Structure des snoARN à boîtes C/D.....	5
2. Maturation des snoARN à boîtes C/D.....	8
B. Les protéines des snoRNP à boîtes C/D.....	10
1. La protéine Nop1/Fibrillarine.....	10
2. Les protéines Nop56/NOP56 et Nop58/NOP58.....	13
3. La protéine Snu13/SNU13.....	15
C. Biogenèse des snoRNP à boîtes C/D.....	16
1. Assemblage des protéines cœur.....	16
2. La protéine chaperonne HSP90.....	19
a. Généralités.....	19
b. Données structurales.....	21
c. Cycle catalytique.....	24
d. Rôle dans l'assemblage des snoRNP à boîtes C/D.....	27
3. Le complexe hR2TP, un cofacteur de la protéine chaperonne HSP90.....	27
a. Découverte du complexe R2TP/hR2TP.....	27
b. Les protéines TIP48/TIP49.....	29
i. Données Structurales.....	29
ii. Données fonctionnelles.....	33
c. La protéine Pih1 (Nop17)/PIH1D1.....	35
i. Données Fonctionnelles.....	35
ii. Données structurales.....	38
d. La protéine Tah1/RPAP3.....	41
i. Données fonctionnelles.....	41
ii. Données structurales.....	43
e. Rôle du complexe (h)R2TP dans l'assemblage des snoRNP.....	45
4. Le complexe NUFIP/ZNHIT3 (Rsa1/Hit1).....	52
a. La protéine Rsa1/NUFIP.....	52
b. La protéine Hit1/ZNHIT3 (TRIP3).....	54
c. Données structurales.....	55
5. La protéine Bcd1/BCD1.....	57
a. Données fonctionnelles.....	57
b. Données structurales.....	60
6. Vers un modèle d'assemblage des snoRNP à boîtes C/D ?.....	62
II. Les snoRNP à boîtes H/ACA.....	67
A. Les snoARN à boîtes H/ACA.....	67
1. Structure des snoARN à boîtes H/ACA.....	67
2. Maturation.....	69
B. Les protéines cœur des snoRNP à boîtes H/ACA.....	72
1. La protéine Cbf5/Dyskérine.....	72
2. La protéine Gar1/GAR1.....	74
3. La protéine Nop10/NOP10.....	78
4. La protéine Nhp2/NHP2.....	80
C. Les facteurs d'assemblage des snoRNP à boîtes H/ACA.....	81
1. La protéine Shq1/SHQ1.....	81
2. La protéine Naf1/NAF1.....	85

D. Assemblage des snoRNP à boîtes H/ACA.....	86
Chapitre 2 : Les microARN.....	92
I. Maturation des microARN.....	92
A. Organisation génétique des microARN.....	92
B. Maturation canonique.....	93
1. Maturation du pri-microARN par le microprocesseur.....	93
2. Export du pré-microARN dans le cytoplasme.....	98
3. Maturation du pré-microARN.....	100
a. Dicer, une RNase de type III de classe 3.....	100
b. Les protéines TRBP et PACT.....	108
4. Formation de la microRNP mature.....	111
a. Le RLC.....	111
i. Généralités.....	111
ii. Les protéines Argonautes.....	111
iii. Rôle de la protéine HSP90 dans le recrutement du microARN.....	113
iv. Séparation des brins guide et passager.....	113
b. Le complexe RISC.....	114
C. Les autres voies de maturation.....	114
1. Maturation d'un pré-microARN par AGO2.....	115
2. Les miRtrons et simtrons.....	115
3. pré-microARN portant une coiffe m ⁷ G.....	118
4. pré-microARN dérivés d'ARNt.....	118
5. Les isomiRs.....	118
II. Fonctions canoniques et non canoniques des microARN.....	122
A. Fonctions canoniques des microARN.....	122
1. Interactions avec l'ARN messenger cible.....	122
2. Rôle du RISC au sein du duplex microARN:ARNm.....	124
a. Inhibition et activation de la traduction de l'ARNm.....	124
b. Dégradation de l'ARNm par l'intermédiaire du RISC.....	124
c. Rôle de la protéine GW182.....	125
B. Fonctions non-canoniques des microARN et protéines associées.....	126
1. Des microRNP sont présentes et actives dans les noyaux de cellules humaines.....	126
2. Rôles des microARN nucléaires et des protéines associées.....	128
Chapitre 3 : Les petits ARN dérivés de snoARN.....	130
I. Des snoARN particuliers.....	130
A. snoARN et trafic intracellulaire du cholestérol.....	130
B. snoARN et édition/épissage.....	131
C. Les « sno-lncRNA ».....	131
1. Les snoARN SNORD115 et SNORD116.....	131
2. Les long ARNnc liés à des snoARN.....	132
II. Les snoARN, précurseurs possibles de microARN ?.....	133
A. Mise en évidence.....	133
B. Des fonctions diverses pour les sdARN.....	134
1. sdARN et épissage alternatif.....	134
2. Des sdARN comme biomarqueurs de cancers ?.....	135
C. Un processus de maturation inconnu.....	136

Objectifs

Résultats et Discussion

Chapitre 1 : Confirmation et caractérisation de l'interaction entre RPAP3 et TRBP.....	140
I. Présentation des travaux préliminaires au doctorat.....	140
A. Identification d'une nouvelle interaction.....	140
B. L'interaction entre TRBP et RPAP3 est directe.....	143
II. Vérification de l'interaction <i>in vivo</i>	143
III. Caractérisation de l'interaction entre RPAP3 et TRBP.....	151
A. Détermination des domaines fonctionnels putatifs de RPAP3.....	151
B. Détermination des domaines fonctionnels putatifs de TRBP.....	152
C. Détermination des domaines impliqués dans l'interaction entre TRBP et RPAP3.....	152
1. La technique de co-expression et de co-purification chez <i>E. coli</i>	152
a. Utilisation du robot TECAN™.....	152
b. Réalisation des différentes constructions et création de vecteurs.....	155
c. Résultats du crible de co-expression.....	156
2. Les domaines TPR n°1 de RPAP3 et DRBM3 de TRBP forment un complexe stable <i>in vitro</i> et <i>in vivo</i>	160
a. Confirmation de l'existence d'un complexe stable <i>in vitro</i>	160
a. Confirmation de l'existence d'un complexe stable <i>in vitro</i>	160
b. Choix des bornes pour l'étude structurale du complexe entre le premier domaine TPR de RPAP3 et le domaine DRBM3 de TRBP.....	168
c. Mise en évidence de l'existence d'un complexe stable <i>in vivo</i> par un test d'interaction double-hydride.....	170
D. Conclusion.....	173
IV. Études structurales du complexe TRBP DRBM3/RPAP3 TPR n°1.....	176
A. Identification des résidus de RPAP3 et de TRBP impliqués dans la formation du complexe protéique.....	176
1. Utilisation des données RMN pour identifier des acides aminés de RPAP3 candidats pour leur implication dans l'interaction avec TRBP.....	176
2. Utilisation des données de cristallographie pour identifier des résidus de TRBP candidats pour leur implication dans l'interaction avec RPAP3.....	182
3. Confirmation par co-expression de l'impact de différents variants de RPAP3 et de TRBP sur la formation du complexe protéique.....	187
B. Résolution de la structure tridimensionnelle du complexe TRBP 262-366/RPAP3 133-255.....	198
V. Conclusion.....	200
Chapitre 2 : Étude de la fonction du complexe RPAP3-TRBP.....	203
I. Analyse des partenaires de TRBP et de RPAP3.....	203
A. Le domaine 133-255 de RPAP3 interagit toujours avec HSP90.....	203
B. Absence de formation d'un complexe RPAP3-TRBP-Dicer.....	203
C. RPAP3 n'interagit pas avec PACT.....	210
D. Conclusion.....	218
II. Caractérisation des variants de TRBP.....	223
III. Rôle dans la génération d'isomiR ?.....	228
A. Production et purification du pré-microARN 30E.....	228
B. Production et purification des protéines TRBP, Dicer et RPAP3 entière.....	230
1. Production et purification de la protéine TRBP.....	230
2. Production et purification de la protéine Dicer.....	231
C. Expérience de production d'isomiR <i>in vitro</i>	232
IV. Rôle dans la biogenèse d'ARNnc ?.....	233

Conclusions et Perspectives

I. TRBP et RPAP3 forment un complexe stable <i>in vitro</i> et <i>in vivo</i>	237
II. L'étude structurale du complexe TRBP/RPAP3 a fourni de précieuses informations fonctionnelles.....	238
A. L'interaction entre TRBP et RPAP3 est fortement spécifique.....	238
B. TRBP interagit-elle avec la protéine RPAP3 libre ou avec le complexe hR2TP ?.....	239
C. Résolution de la structure 3D de l'interface d'interaction entre RPAP3 et TRBP.....	240
III. Mise en place d'outils moléculaires pour l'étude fonctionnelle du complexe.....	242

Matériel et Méthodes

I. Matériel.....	246
A. Les souches de cellules.....	246
1. Les souches de bactéries.....	246
2. Les souches de levures.....	246
3. Les lignées de cellules humaines.....	246
4. Les lignées de cellules d'insectes.....	247
B. Les oligonucléotides.....	247
C. Les plasmides.....	254
1. Pour le clonage Gateway®.....	254
2. Pour le double-hybride chez la levure <i>S. cerevisiae</i>	254
3. Pour la surproduction de protéines recombinantes chez <i>E. coli</i>	255
4. Pour la production de protéines recombinantes en baculovirus.....	255
5. Pour les expériences d'immunoprécipitation.....	256
6. Pour la transcription d'un ARN.....	256
D. Les milieux de cultures.....	256
1. Cultures de bactéries.....	256
2. Cultures de levures.....	257
3. Culture de cellules humaines.....	257
4. Culture de cellules d'insectes.....	257
E. Les tampons et solutions d'usage courant.....	257
F. Les anticorps.....	258
II. Méthodes.....	260
A. Méthodes relatives aux acides nucléiques.....	260
1. Clonage.....	260
a. PCR.....	260
b. Clonage par ligature.....	260
c. Mutagenèse dirigée.....	261
d. Clonage par recombinaison Gateway®	261
i. Recombinaison avec le vecteur pDONR207 :.....	261
ii. Recombinaison avec les vecteurs de destination :.....	261
e. Clonage par recombinaison par la technique d'In-Fusion™ de Clontech.....	262
f. Transformation de bactérie <i>E. coli</i> DH5α ou BL21.....	262
i. Obtention de bactérie <i>E. coli</i> DH5α ou BL21 super-compétentes.....	262
ii. Transformation de bactérie <i>E. coli</i> DH5α par choc thermique.....	262
2. Obtention du pre-microARN 30E.....	263
a. Préparation de la matrice ADN.....	263
b. Transcription <i>in vitro</i>	263
c. Purification du pre-microARN 30E.....	263
d. Déphosphorylation et marquage d'un ARN.....	263
3. Transfection de cellules HEK par un siARN.....	264
a. Préparation d'un siARN à partir de deux oligoribonucléotides de synthèses.....	264
b. Transfection de cellules HEK au phosphate de calcium.....	264

4. Transcription inverse et PCR quantitative.....	264
a. Extraction des ARN totaux de cellule H.E.K.....	264
b. Élimination de l'ADN génomique.....	265
c. Transcription inverse des ARNm.....	265
d. PCR quantitative.....	265
5. Préparation des ARN pour puce à ARN non-codant.....	265
6. Coupure du pré-microARN 30E par Dicer en présence de TRBP et/ou de RPAP3.....	265
B. Méthodes relatives aux protéines.....	266
1. Méthodes relatives à la levure <i>Saccharomyces cerevisiae</i>	266
a. Préparation de levures <i>Saccharomyces cerevisiae</i> compétentes.....	266
b. Transformation de la levure <i>Saccharomyces cerevisiae</i>	266
c. Double-hybride	266
2. Transfection de cellule HEK par un plasmide.....	267
3. Préparation d'extraits protéiques à partir de cellule Hela/HEK.....	267
4. Immunoprécipitation.....	267
5. Western-blot.....	268
6. Immunofluorescence.....	268
7. Duolink™	269
8. Production et purification de protéines recombinante chez <i>E. coli</i>	270
a. Transformation de bactéries <i>E. coli</i> BL21 par plusieurs vecteurs d'expression.....	270
b. Production de protéines recombinantes chez <i>E. coli</i> BL21.....	270
i. Pour les analyses par RMN homo- et hétéronucléaire.....	270
ii. Pour les cribles de co-expression chez <i>E. coli</i> BL21.....	270
iii. Pour les expériences de co-expression chez <i>E. coli</i>	271
iv. Pour les expériences <i>in vitro</i>	271
c. Purification de protéines recombinantes chez <i>E. coli</i> BL21.....	271
i. Purification par chromatographie d'affinité sur résine TALON™.....	271
ii. Purification par chromatographie d'exclusion stérique.....	272
iii. Purification des cribles de co-expressions à l'aide du robot TECAN™ par chromatographie d'affinité.....	273
iv. Purification par chromatographie d'affinité à haute pression.....	273
9. Production et purification de protéines recombinantes par baculovirus.....	274
a. Obtention d'un virus recombinant P0.....	274
b. Amplification du virus recombinant en flasque T175.....	274
c. Amplification du virus en plaque 6 puits.....	274
d. Production de protéines recombinantes par infection de cellules Sf9.....	275
e. Purification des protéines recombinantes.....	275
10. Expériences RMN.....	276
C. Méthodes relatives à la culture de cellules humaines.....	276
1. Décongélation de cellules HeLa/HEK.....	276
2. Amplification de cellules HeLa/HEK.....	276
3. Congélation de cellules HeLa/HEK.....	276
D. Méthodes relatives à la culture de cellule d'insectes.....	277
1. Amplification de cellules Sf9.....	277
2. Culture liquide de cellules Sf9.....	277

Bibliographie

Annexes

Introduction

Avant-propos

Il existe dans toutes cellules deux types d'ARN, les ARN messagers, qui sont traduits en protéines, et les ARN non-codants (ARNnc), dont l'importance fonctionnelle a pendant longtemps été sous-estimée. L'étude des ARNnc a connu ces dernières années un regain d'intérêt très important en raison du nombre important de processus biologiques dans lesquels ils sont engagés (pour revue Clerget *et al.*, 2015, voir Annexe) (**Figure 1**). Les ARNnc sont ainsi notamment impliqués dans la maturation du pré-ARNm (i.e. snARN, Weinberg et Penman, 1968 ; pour revue Butcher et Brow, 2005), et dans la traduction des ARNm (i.e. ARNt et ARNr, Hoagland *et al.*, 1958 ; Brenner *et al.*, 1961 ; Palade, 1955), dans la maturation d'autres ARN (i.e. snoARN, pour revue Maxwell et Fournier, 1995), dans la sécrétion des protéines (i.e. 7SL, Walter et Blobel, 1982) mais également dans le développement cellulaire (i.e. lncARN ; pour revue Mercer *et al.*, 2009) ou encore dans le contrôle de l'expression génique (i.e. microARN ; Lee *et al.*, 1993). Enfin, depuis une dizaine d'années, les progrès réalisés notamment en matière de séquençage à haut-débit ont entraîné l'identification de nombreux ARNnc tels que les tiARN (« transcription initiation RNA »), les PASR (« promoter-associated short RNA », TASR (« termini-associated short RNAs »), les splicing-ARN (« Splice-site associated smRNAs »), dont les fonctions ne sont toutefois pas encore toujours clairement établies (Taft *et al.*, 2009 ; Kapranov *et al.*, 2007 ; Taft *et al.*, 2010).

Ces différents ARNnc sont le plus souvent intégrés au sein de complexes protéiques, formant alors des particules ribonucléoprotéiques (RNP), assemblées selon des processus de maturation complexes et souvent propres à chaque ARN. De façon surprenante, de récentes études ont toutefois montré que certains ARNnc pouvaient servir de précurseurs à d'autres ARNnc. Ainsi, la caractérisation des ARNnc associés à certaines protéines comme les protéines de la famille Argonaute (l'une des protéines associées aux microARN matures, voir Chapitre 2) a montré l'existence de petits ARN dérivés de snoARN, ou sdRNA (« snoRNA-derived RNA »). Ainsi, bien que les snoARN aient été identifiés il y a de nombreuses années et que leurs fonctions semblaient avoir été identifiées depuis, cette découverte a entraîné un regain d'intérêt pour l'étude de leur maturation, et plus particulièrement sur leur possibilité de générer de petits ARN de type microARN. Il semble aujourd'hui d'autant plus intéressant de réaliser ces études qu'il apparaît que ces petits ARN dérivés et/ou leurs précurseurs seraient impliqués dans de nombreuses pathologies, comme le cancer de la prostate (Dong *et al.*, 2008), du foie (Kawaji *et al.*, 2008), du poumon (Liao *et al.*, 2010), du cerveau (Dong *et al.*, 2009) ou encore certaines leucémies (Valleron *et al.*, 2012).

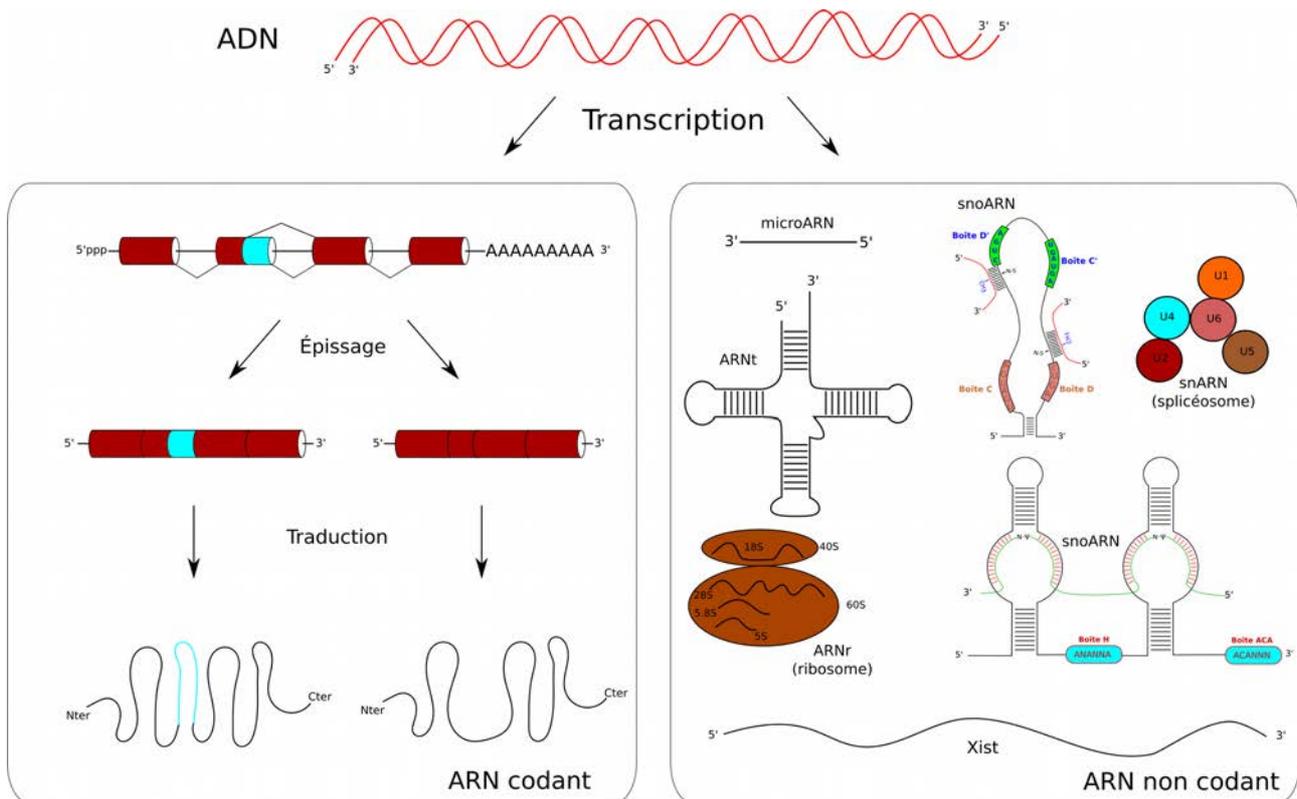


Figure 1 : ARN codants et ARN non-codants : un bilan non exhaustifs

Chaque cellule contient deux types d'ARN, les ARNm qui sont obtenus suite à la maturation des pré-ARNm et qui sont traduits en protéines, et les ARNnc qui, bien qu'ils ne soient pas traduits en protéines, sont impliqués dans la plupart des processus biologiques. Ainsi, les ARNr participent à la formation du ribosome, dont la fonction est de traduire les ARNm en protéines (Palade, 1955). Les petits ARN nucléolaires, ou snoARN, sont eux nécessaires à la maturation des ARNr (pour revue Maxwell et Fournier, 1995). Les ARNt portent au niveau de leur extrémité 5' un acide aminé et permettent ainsi de faire le lien entre le code génétique et le code protéique (Hoagland *et al.*, 1958, Brenner *et al.*, 1961). Enfin, les snARN sont impliqués dans le processus de maturation du pré-ARNm en ARNm (épissage) via le splicéosome (Weinberg et Penman, 1968 ; pour revue Butcher et Brow, 2005). De très longs ARN, comme Xist, ont également été identifiés et sont généralement impliqués dans le développement cellulaire (pour revue Mercer *et al.*, 2009). Enfin, récemment, de petits ARN, appelés microARN, ont été identifiés comme étant des régulateurs majeurs de l'expression génique (Lee *et al.*, 1993).

Mon travail de doctorat fait suite à un travail préliminaire qui concernait l'étude de la maturation des sdRNA (voir la partie **Objectifs**). Mon introduction se concentrera donc plus spécifiquement sur cette nouvelle famille d'ARNnc et sur les 2 familles d'ARNnc liées, les microARN et les snoARN, ainsi que sur les différentes protéines impliquées dans leur biogenèse.

Chapitre 1 : Les snoARN

Les premiers snoARN furent identifiés par des expériences de fractionnement nucléaire (Reddy *et al.*, 1981 ; Kiss et Filipowicz, 1993) qui ont montré l'existence de petits ARNnc présents spécifiquement dans le nucléole et impliqués dans la maturation des ARNr. Comme la grande majorité des ARN, les snoARN sont stabilisés au sein de complexes, contenant à la fois l'ARN et des protéines spécifiques, appelés particules RiboNucléoProtéines (RNP). Des expériences supplémentaires d'immunoprécipitation dirigées contre certaines protéines constituant ces snoRNP ont permis de classer les snoARN en 2 groupes, les snoARN associés à la protéine Fibrillarine (Tyc et Steitz, 1989) et ceux associés à la protéine Gar1 (Girard *et al.*, 1992).

Les snoRNP du premier groupe correspondent aux snoRNP à boîtes C/D, en référence à la présence de 2 séquences hautement conservées au niveau du snoARN, les motifs C et D. Les snoRNP du second groupe correspondent aux snoRNP à boîtes H/ACA, caractérisées par la présence de 2 autres séquences conservées, les motifs H et ACA.

Les snoRNP à boîtes C/D sont historiquement décrits comme impliqués dans la maturation des ARNr, soit dans le clivage du pré-ARNr (Hughes et Ares, 1991), soit dans la catalyse de 2'-O-méthylation de riboses de certains nucléotides. Les snoRNP à boîtes H/ACA sont eux principalement impliqués dans la conversion d'uridines en pseudouridines au niveau d'ARNr, à l'exception notable de la Télomérase, une scaRNP à boîtes H/ACA impliquée dans le maintien de la longueur des télomères (Nugent et Lundblad, 1998 ; Mitchell *et al.*, 1999). Les snoARN ne réalisent pas eux-mêmes ces modifications mais agissent en tant que guides en s'hybridant de manière spécifique aux ARN cibles. Les modifications sont réalisées par des enzymes présentes au sein des snoRNP. Ainsi la Fibrillarine est responsable des réactions de 2'-O-méthylation et la Dyskérine des réactions de pseudouridylation.

Les scaARN ne sont pas à proprement parler des snoARN puisqu'ils ne sont pas localisés dans le nucléole mais dans les corps de Cajal à cause d'une séquence particulière appelée motif CAB (Richard *et al.*, 2003). Ils possèdent néanmoins les mêmes types de séquences conservées que les snoARN et interviennent dans la maturation de snARN (impliqués dans le processus d'épissage) en catalysant les mêmes types de modifications.

Enfin, il est intéressant de noter que malgré les progrès réalisés en matière de séquençage et en bio-informatique, les cibles d'environ 20 % des snoARN connus à ce jour n'ont toujours pas été identifiées (Jorjani *et al.*, 2016). Ces snoARN sont alors appelés des snoARN orphelins.

I. Les snoRNP à boîtes C/D

A. Les snoARN à boîtes C/D

1. Structure des snoARN à boîtes C/D

Les snoARN à boîtes C/D sont des petits ARN ayant une taille comprise entre 50 et 200 nucléotides et caractérisés par la présence de 2 séquences hautement conservées. Le motif (ou boîte) C est situé au niveau de l'extrémité 5' du snoARN et a pour séquence consensus 5'-RUGAUGA-3' (ou R est une purine). La boîte D est située au niveau de l'extrémité 3' du snoARN et a pour séquence consensus 5'-CUGA-3'. La plupart des snoARN contiennent également deux autres séquences moins conservées, les boîtes C' et D' qui sont présentes au centre du snoARN. L'ARN cible vient s'hybrider au niveau de l'élément antisens. Cet élément, d'une longueur de 10 à 21 nucléotides et complémentaire à l'ARN cible, se situe directement en amont des boîtes D et D'. Toutefois il a été montré que pour certains snoARN, l'ARNr cible s'hybridait à la fois au niveau de l'élément antisens mais aussi à distance de ces derniers, ce qui augmentait d'environ 5 fois l'efficacité de méthylation du ribose du nucléotide cible (van Nues *et al.*, 2011). De manière remarquable, le ribose du nucléotide modifié est toujours celui du 5ème nucléotide apparié au snoARN en amont de la boîte D et D' (**Figure 2**).

Les boîtes C et D s'associent pour former un motif hélice-boucle-hélice, le motif C/D, qui se structure en *Kink-Turn* ou *K-Turn* (Klein *et al.*, 2001). Ce motif comporte 2 hélices I et II séparées par des appariements non canoniques **G•A** et **A•G** au niveau des boîtes C et D et par une paire **U•U** au début de l'hélice II (Marmier-Gourrier *et al.*, 2003). Ces appariements sont responsables de la présence de 3 nucléotides non appariés (**Figure 3**). Ce repliement a la particularité de laisser une uridine libre, rejetée en dehors du motif et donc potentiellement libre d'établir des interactions avec des protéines qui reconnaissent ce motif singulier, les protéines de la famille L7Ae (Watkins *et al.*, 2002). Enfin, ce motif peut être également stabilisé par des ions métalliques (Goody *et al.*, 2004).

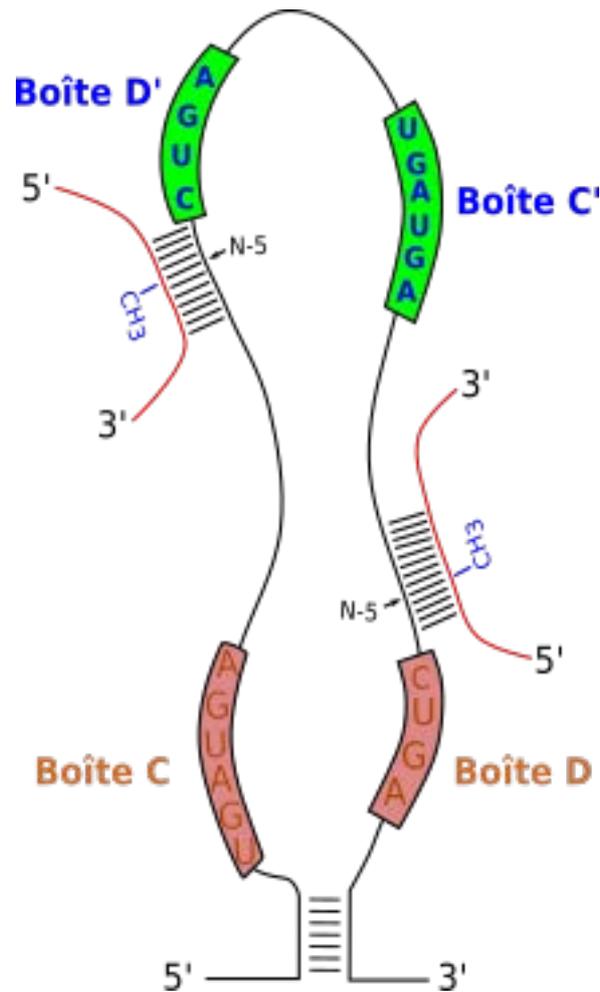


Figure 2: snoARN à boîtes C/D.

Les snoARN à boîtes C/D sont caractérisés par la présence de deux courtes séquences très conservées : la boîte C au niveau de l'extrémité 5' et la boîte D au niveau de l'extrémité 3'. La plupart d'entre-eux comportent également des boîtes C' et D', de séquences moins conservées.

L'ARNr cible vient s'hybrider sur le snoARN à boîtes C/D au niveau d'une séquence nommée élément antisens, qui est situé en amont des boîtes D et D'. Le résidu qui est 2'-O-méthylé au niveau de son ribose se trouve toujours en position 5 en amont des boîtes D et D' (résidu marqué N-5 sur la figure). Les boîtes C/D et C'/D' forment, en se rapprochant, un motif tridimensionnel particulier nommé *Kink-Turn* (*K-Turn*) reconnu par des protéines de la famille L7Ae.

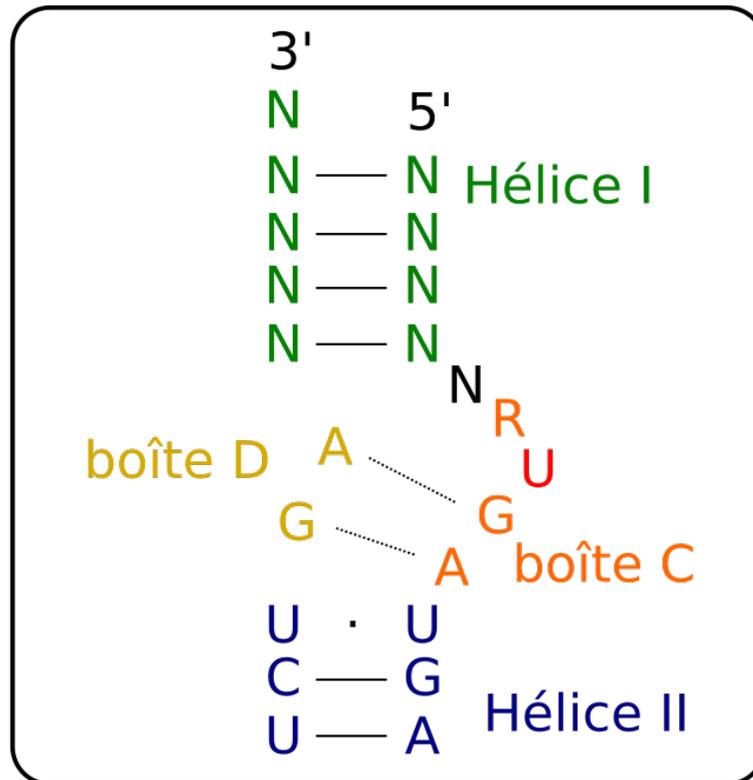


Figure 3: Structure secondaire consensus du motif *Kink-Turn* des boîtes C/D (d'après Marmier-Gourrier *et al.*, 2003).

Les boîtes C/D (et C'/D') s'associent pour former un motif en hélice-boucle-hélice qui se structure en *Kink-Turn* ou *K-Turn* (Klein *et al.*, 2001). Ce motif comporte 2 hélices I et II séparées par 3 nucléotides non appariés faisant suite à la présence d'appariements non canoniques G·A et A·G entre les boîtes C et D et d'une paire U·U au début de l'hélice II. Parmi les 3 nucléotides non appariés, l'uridine (ici en rouge) est projeté en dehors du motif et va être reconnu par des protéines de la famille L7Ae dont la protéine Snu13/SNU13 (Watkins *et al.*, 2002)

R : Purine.

N : n'importe quel nucléotide.

2. Maturation des snoARN à boîtes C/D

Chez les vertébrés, la plupart des gènes codant des snoARN sont localisés au sein d'introns de gènes codant des protéines impliquées dans la biogenèse des ribosomes ou dans la traduction des ARNm. La maturation de ces snoARN est donc dépendante du processus d'épissage. Plus précisément, il semble que la position du snoARN au sein de l'intron qui va être épissé soit critique. Ainsi, il a été montré que la plupart des snoARN introniques se trouvaient à une distance comprise entre 70 et 80 nucléotides en amont de l'extrémité 3' de l'intron hôte et que l'altération de cette distance impactait la libération du snoARN de l'intron (Hirose et Steitz, 2001).

De plus, *Hirose et al.* ont également montré en 2006 qu'il existait une distance cruciale d'environ 50 nucléotides entre la région comprenant la séquence des snoARN et le point de branchement (Hirose *et al.*, 2006). Ainsi, ils ont montré que lors de la première réaction de trans-estérification, le facteur d'épissage IBP160 est recruté par le complexe C1 du splicéosome, qui, si un snoARN est présent à la bonne distance, stabilise le repliement du snoARN et/ou facilite le recrutement de certaines protéines des snoRNP, notamment la protéine SNU13 qui se fixe au niveau du *K-Turn* (Watkins *et al.*, 2000). Une fois la deuxième réaction de transestérification effectuée et l'intron relargué sous forme de lasso, la liaison 2'-5' phosphodiester est hydrolysée par l'enzyme de débranchement DBR1 (**Figure 4**). Chez la levure *S. cerevisiae*, il a été montré que les extrémités 5' et 3' des snoARN introniques étaient ensuite maturées par des exonucléases (Villa *et al.*, 1998) et par l'exosome (Allmang *et al.*, 1999).

Il a toutefois été montré que des snoARN introniques présents à plus longue distance pouvaient toujours être maturés, mais de manière moins efficace et a priori via un mécanisme ne nécessitant pas d'interaction entre le snoARN, des protéines du snoARN et des protéines de la machinerie d'épissage (Hirose *et al.*, 2003). Des travaux récents semblent également indiquer l'existence de deux types de snoARN à boîtes C/D selon leur localisation au sein de l'intron hôte. Ainsi les snoARN distant d'environ 75 nucléotides de l'extrémité 3' de l'intron posséderaient une fois maturés des extrémités « longues », c'est-à-dire une boîte C située à 5-6 nucléotides de l'extrémité 5', contre 4 à 5 nucléotides pour les formes « courtes », et une boîte D située à 4-5 nucléotides de l'extrémité 3', contre 2 à 3 nucléotides pour les formes « courtes » (Deschamps-Francoeur *et al.*, 2014).

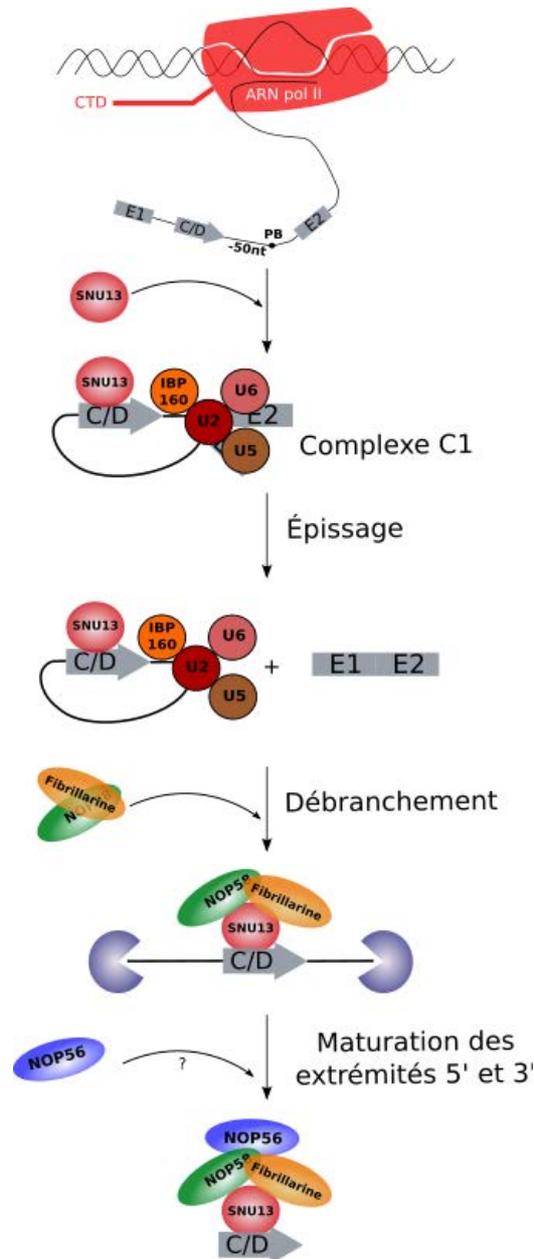


Figure 4 : Processus de maturation d'un snoARN à boîtes C/D intronique (d'après Kiss *et al.*, 2006).

La majorité des snoARN à boîtes C/D est produite à partir d'introns et est excisée de manière dépendante de l'épissage des pré-ARNm. Lors de la première réaction de transestérification, le facteur d'épissage IBP160 est recruté par le complexe C1 du spliceosome. Ce facteur d'épissage stabilise le repliement du snoARN et/ou facilite le recrutement de certaines protéines cœurs des snoRNP, notamment la protéine SNU13. Une fois la deuxième réaction de transestérification effectuée et l'intron relargué sous forme de lasso, la liaison 2'-5' phosphodiester est hydrolysée par l'enzyme de débranchement DBR1. Les extrémités 5' et 3' du pré-snoARN sont ensuite maturées respectivement par des exonucléases (Villa *et al.*, 1998) et par l'exosome (Allmang *et al.*, 1999).

B. Les protéines des snoRNP à boîtes C/D

Les snoRNP à boîtes C/D (**Figure 5**) ont tout d'abord été isolées par des expériences d'immunoprécipitation réalisées contre la protéine Fibrillarine (Tyc et Steitz, 1989). Des expériences de synthétiques létaux réalisées chez la levure *S. cerevisiae* avec des mutants de la protéine Nop1, l'homologue de la protéine Fibrillarine, ont permis de mettre en évidence 2 nouvelles protéines associées aux snoARN à boîtes C/D, les protéines Nop56 et Nop58 (Gautier *et al.*, 1997 ; Wu *et al.*, 1998 ; Lafontaine et Tollervey, 1999, 2000). Ces deux protéines ont pour homologues respectivement les protéines NOP56 et NOP58 chez l'homme. Enfin, des expériences de purification de la snoRNP U3, une snoRNP à boîtes C/D impliquée dans les clivages précoces du pré-ARNr, ont permis de révéler une dernière protéine associée à l'ensemble des snoRNP à boîtes C/D, la protéine Snu13 (chez la levure *S. cerevisiae*) / SNU13 (chez l'homme) (Watkins *et al.*, 2000).

1. La protéine Nop1/Fibrillarine

La Fibrillarine (Nop1 chez la levure *S. cerevisiae*) est l'enzyme qui est responsable, au sein des snoRNP à boîtes C/D, du transfert d'un groupement méthyle fourni par la *S*-Adénosyl-Méthionine sur le groupement hydroxyle présent sur le carbone 2' du ribose cible (**Figure 6**). C'est une protéine essentielle, d'une taille d'environ 38 kDa, initialement identifiée comme l'un des composants des nucléoles (Ochs *et al.*, 1985).

La Fibrillarine peut être divisée en 2 domaines, un domaine N-terminal et un domaine portant l'activité méthyltransférase (MTase). Le domaine N-terminal comporte un domaine GAZ et un domaine « linker » entre le domaine N-Terminal et le domaine MTase. Le domaine GAZ est impliqué dans des interactions avec d'autres protéines, comme le facteur d'épissage p32 (Yanagida *et al.*, 2004), les méthyltransférases PRMT1/PRMT5 ou encore la protéine SMN (Jones *et al.*, 2001 ; Pellizzoni *et al.*, 2001). Il comporte un signal de localisation nucléaire ou NLS (Levitskiï *et al.*, 2004). De plus, le domaine GAZ peut être méthylé au niveau de certains résidus arginine (Lischwe *et al.*, 1985), probablement via l'action de PRMT1 et/ou de PRMT5. De manière intéressante, ce domaine est absent chez l'homologue d'Archaea de la Fibrillarine.

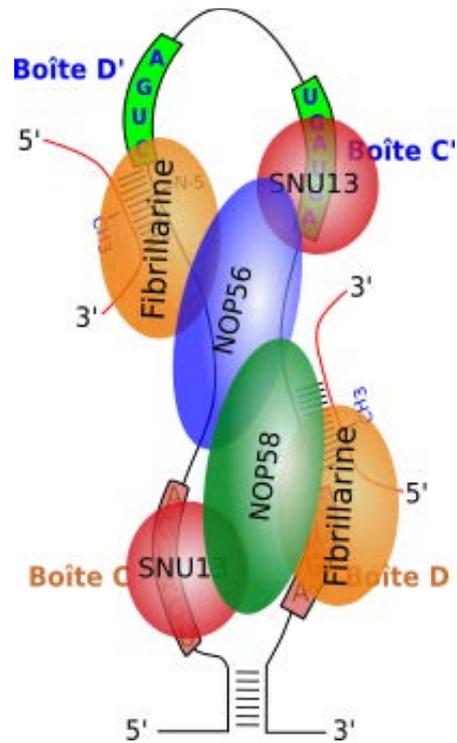


Figure 5: snoRNP à boîtes C/D.

Les snoRNP à boîtes C/D sont composées d'un snoARN à boîtes C/D et de 4 protéines cœur. Ces protéines sont la Fibrillarine, responsable de la 2'-O-méthylation, les protéines NOP56 et NOP58 et de la protéine SNU13. Cette dernière est une protéine appartenant à la famille L7Ae et va reconnaître les motifs en *K-Turn* présents au niveau des boîtes C/D et C'/D' du snoARN.

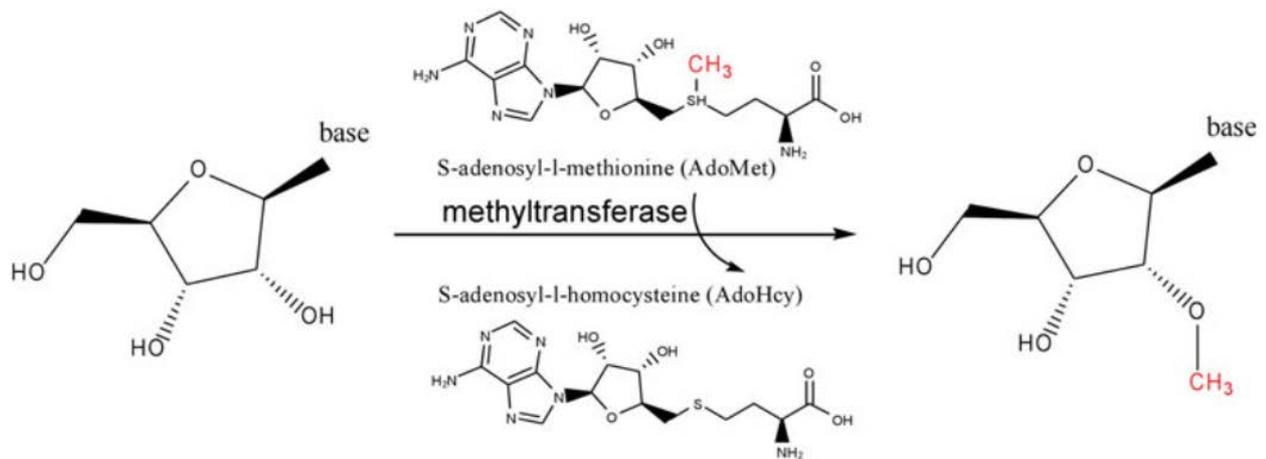


Figure 6: Réaction de transfert du groupement méthyle de la SAM au ribose.

La réaction de 2'-*O*-méthylation réalisée par la Fibrillarine consiste en un transfert de groupement méthyle (ici en rouge) fournit par la *S*-Adénosyl-Méthionine (SAM) sur le groupement hydroxyle du carbone 2' du ribose de l'ARN cible (pour revue Liang et Li, 2010).

Le domaine MTase adopte un repliement dit de Rossmann, caractéristique des domaines d'interaction protéines-acides nucléiques. Ce repliement est constitué d'un feuillet β , composé de 7 brins β parallèles, flanqué par 7 hélices α qui portent la triade catalytique K138/D231/K260 (**Figure 7**). Ce domaine peut être divisé en deux régions, un domaine R d'environ 90 résidus et un domaine riche en hélice α . Le domaine R est caractérisé par la présence d'un motif de fixation à l'ARN (GCVYAVCF) (Aris et Blobel, 1991), mais il a été montré que ce motif n'était pas suffisant pour fixer l'ARN substrat. En effet Rakitina *et al* ont montré en 2011 qu'il existait 2 autres sites de fixation pour l'ARN, l'un au niveau du domaine R et l'autre au niveau du domaine riche en hélices α . Ces domaines fonctionneraient avec une certaine synergie mais indépendamment l'un de l'autre. Il est alors possible que la protéine agisse non pas sur un mais deux substrats. Le domaine riche en hélices α semble interagir avec la protéine NOP56, une autre protéine cœur des snoRNP à boîtes C/D (McKeegan *et al.*, 2007 ; Lechertier *et al.*, 2009).

Enfin, indépendamment de son rôle au sein des snoRNP, il a récemment été montré chez la levure *S. cerevisiae*, chez l'homme et chez les plantes que la Fibrillarine était également responsable de la méthylation de l'histone H2A au niveau des promoteurs des gènes codant l'ARNr (Tessarz *et al.*, 2014 ; Loza-Muller *et al.*, 2015).

2. Les protéines Nop56/NOP56 et Nop58/NOP58

Comme indiqué précédemment, les protéines nucléolaires (Nop) Nop56 et Nop58 ont été identifiées suites à des expériences de synthétiques létaux réalisées chez la levure *S. cerevisiae* avec des mutants de la protéine Nop1. Ces deux protéines, qui possèdent une très forte homologie (45 % d'identité), ont respectivement pour homologue chez l'homme les protéines NOP56 et NOP58, dont elles partagent également une très forte homologie (52 % d'identité entre les protéines Nop56 et NOP56) (Gautier *et al.*, 1997). Bien que très proches l'une de l'autre, il a été montré que la protéine Nop56 n'est pas indispensable pour la stabilité des snoRNP à boîtes C/D contrairement à la protéine Nop58, *in vivo* (Lafontaine et Tollervey, 1999). De même, des expériences de complémentations croisées ont montré que ces deux protéines n'étaient pas interchangeables (Gautier *et al.*, 1997).

Les protéines NOP56 et NOP58 possèdent un domaine C-terminal très riche en résidus basiques (notamment des lysines) et en acides glutamiques. Ce domaine a été montré comme ayant à la fois un rôle de signal de localisation nucléaire (NLS) et nucléolaire (NoLS) (Pradet-Balade *et al.*, 2011).

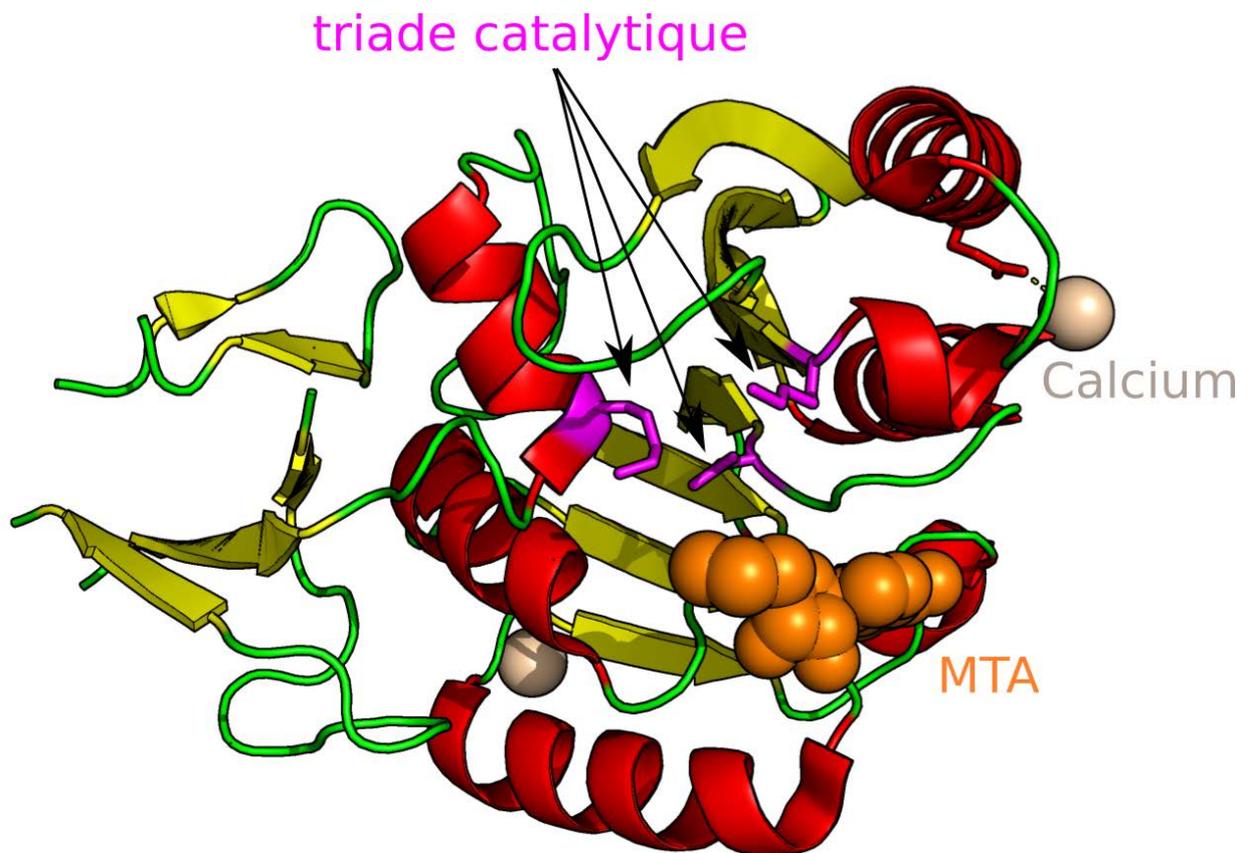


Figure 7: Structure tridimensionnelle de la protéine Fibrillarine humaine résolue par diffraction des rayons X. (code PDB : 2IPX)

La structure révèle la présence de MTA, ou 5'-MéthylThioAdenosine, au niveau du site actif. Le MTA est un analogue non fonctionnel de la *S*-Adénosyl-Méthionine, le co-facteur naturel de la Fibrillarine.

Les 3 résidus de la triade catalytique (K138, D231, K260) sont représentés en violet. La structure du domaine GAR de la protéine humaine n'a pas été résolue.

Ce domaine est également connu pour interagir avec l'enzyme TGS1 qui est responsable de l'hyperméthylation de la coiffe m⁷G en m^{2,2,7}G. Cette coiffe est présente à l'extrémité 5' des précurseurs des snoARN à boîtes C/D transcrits par l'ARN polymérase II (notamment les snoARN U3, U8 et U13) (Mouaikel *et al.*, 2002).

Les protéines NOP56 et NOP58 possèdent également un domaine conservé NOP, un domaine « coiled-coil » et un domaine N-terminal. Le domaine NOP (Watkins *et al.*, 2002) est constitué d'hélices α et est connu pour être un domaine d'interaction protéine-ARN ou protéine-protéine (Liu *et al.*, 2007). Le domaine « coiled-coil » permet la formation d'un hétérodimère NOP56/NOP58 (Aittaleb *et al.*, 2003 ; Lin *et al.*, 2011 ; Bizarro *et al.*, 2014). Enfin le domaine N-terminal est impliqué dans l'interaction avec la protéine Fibrillarine (Aittaleb *et al.*, 2003).

De manière intéressante, il a été montré chez des patients souffrant d'une forme particulière d'ataxie cérébro-spinale, la SCA36, l'augmentation de la répétition d'une séquence hexa-nucléotidique [GGCCUG] au sein d'un intron de l'ARNm codant la protéine NOP56 (Kobayashi *et al.*, 2011). Cette répétition est reconnue par le facteur d'épissage SRSF2, ce qui aurait notamment pour conséquence la séquestration de ce facteur, altérant le fonctionnement normal de certaines cellules, notamment les motoneurones.

3. La protéine Snu13/SNU13

La protéine SNU13 (ou 15.5/15.5K/15.5kD, ou Snu13 chez la levure *S. cerevisiae*) a été initialement identifiée comme un composant important d'une autre particule RNP, la snRNP U4, qui est une composante importante du spliceosome (Nottrott *et al.*, 1999). Elle partage avec d'autres protéines un domaine particulier de fixation à l'ARN. Ces protéines forment ainsi la famille des protéines L7Ae (Watkins *et al.*, 2000 ; Boulon *et al.*, 2008). Chez les eucaryotes, cette famille inclut, en plus de la protéine Snu13/SNU13, la protéine NHP2, qui est impliquée dans la stabilité des sno/scaRNP à boîtes H/ACA (Henras *et al.*, 1998), et la protéine SBP2, qui est impliquée dans l'incorporation de sélénocystéine via la reconnaissance d'une structure secondaire particulière au niveau de la région 3' UTR de certains ARNm, appelée l'élément SECIS (« SelenoCysteine Insertion Sequence ») (Allmang *et al.*, 2002). Contrairement aux autres protéines cœur des snoARN à boîtes C/D, la protéine SNU13 a la particularité, comme les autres protéines L7Ae, de se fixer directement à l'ARN au niveau du motif en *K-Turn* (voir la partie **I. A. 1** de ce chapitre).

La première structure résolue fut celle de la protéine SNU13 en association avec le motif en *K-Turn* du snARN U4 (Vidovic *et al.*, 2000). Cette structure (**Figure 8**) montre que la protéine SNU13 adopte une structure globulaire avec une alternance d'hélices α et de brins β , formant une structure dite de « sandwich α - β - α » (Vidovic *et al.*, 2000). L'interaction avec le motif en *K-Turn* de l'ARN se fait grâce à la présence d'une poche formée par les résidus E61 et K86 dans laquelle l'uridine protubérante peut se placer et être stabilisée par différentes liaisons hydrogènes présentes entre l'uridine et ces deux résidus (Vidovic *et al.*, 2000). L'uridine est également stabilisée par l'intermédiaire de liaisons hydrogènes entre la chaîne principale des acides aminés A39 et I100 et son groupement phosphate et grâce à des contacts hydrophobes entre les chaînes latérales des isoleucines 65 et 100 et l'uracile (Vidovic *et al.*, 2000). La présence de ces multiples contacts explique pourquoi la substitution de cette uridine abolit complètement l'interaction avec la protéine SNU13. En revanche contrairement à l'uridine, il a été montré en utilisant le snoARN U14 comme modèle que si la présence des 2 hélices I et II était indispensable à la formation du complexe entre la protéine SNU13 et le snoARN U14, la séquence en elle-même n'avait pas autant d'importance (Watkins *et al.*, 2002). Enfin, des expériences de GST pull-down n'ont pas montré d'interaction entre la protéine SNU13 et les autres protéines cœurs des snoRNP à boîtes C/D (McKeegan *et al.*, 2007), ce qui suggère que l'assemblage de ces protéines cœurs sur l'ARN nécessite l'intervention d'autres facteurs protéiques.

C. Biogenèse des snoRNP à boîtes C/D

1. Assemblage des protéines cœur

Comme décrit précédemment, parmi les différentes protéines cœur des snoRNP à boîtes C/D, seule la protéine Snu13/SNU13 interagit directement avec l'ARN via le motif en *K-Turn*. Cette première interaction semble donc être nécessaire pour pouvoir recruter les 3 autres protéines cœur, probablement via des interactions avec la protéine Snu13/SNU13. Toutefois, il a été montré que la protéine SNU13 pouvait s'associer indifféremment avec les *K-Turn* présents sur les snoARN à boîtes C/D, le snARN U4 ou encore le snoARN U3, dont la structure diffère des snoARN à boîtes C/D canonique (Watkins *et al.*, 2000 ; Cléry *et al.*, 2007). Ainsi la protéine Snu13/SNU13 semble donc capable, par la reconnaissance d'un motif extrêmement similaire, d'initier l'assemblage de trois types de particules RNP différentes, ce qui suggère l'existence d'éléments de différenciation supplémentaires.

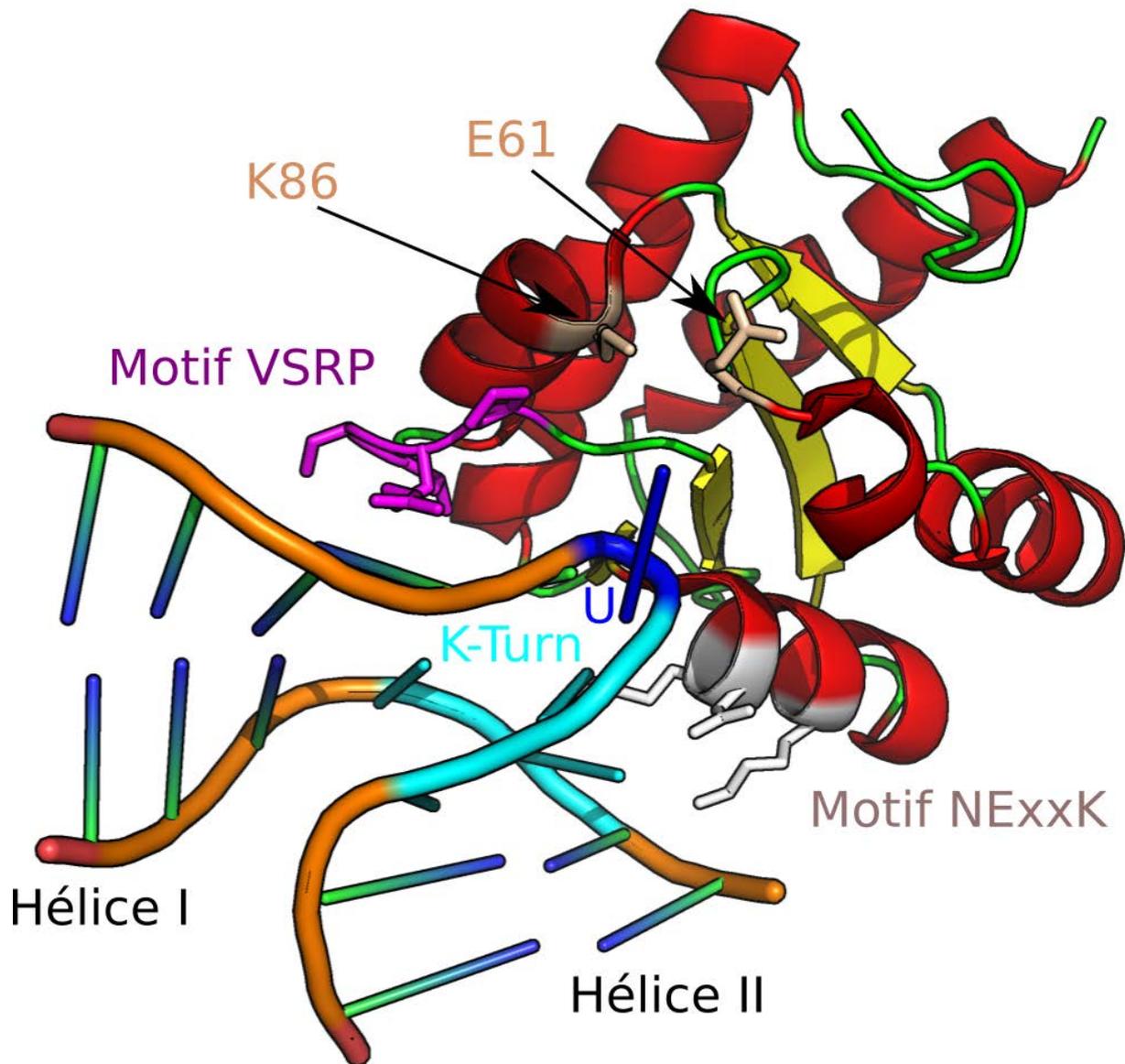


Figure 8: Structure tridimensionnelle de la protéine SNU13 humaine résolue par diffraction des rayons X en complexe avec le motif en *K-Turn* du snARN U4 (code PDB : 1E7K ; Vidovic *et al.*, 2000).

La présence d'appariements non canoniques entre 2 paires de G·A/A·G entre les deux hélices entraîne la formation d'une structure en *K-Turn* au niveau de l'ARN (ici en bleu clair). Au sein de cette structure, un uracile (ici en bleu) est projeté en dehors du motif et est reconnu par la protéine SNU13.

Le résidu uridine est stabilisé par des interactions avec la protéine SNU13, notamment au niveau des acides aminés E61 et K86 (en brun). Les acides aminés V95, S96, R97 et P98 du motif VSRP (en violet) participent à la stabilisation de l'hélice I du *K-turn*. L'hélice II est stabilisée par des interactions avec les résidus N40, E41 et K44 du motif NExxK (en gris).

Watkins *et al.* ont notamment montré que la protéine SNU13 pouvait se fixer sur le *K-Turn* des snoARN à boîtes C/D quelle que soit la séquence de l'hélice II. Ceci n'est toutefois pas le cas pour les 3 autres protéines cœur (Watkins *et al.*, 2000). Dans cette étude, les auteurs ont également montré que cette séquence était importante pour la localisation nucléolaire des snoRNP. De plus, il a été montré que plusieurs acides aminés de surface de la protéine SNU13 étaient importants pour le recrutement et/ou la stabilisation de certaines protéines au sein des différentes particules matures. C'est notamment le cas pour la protéine 55 kDa au sein des snoRNP U3, de la protéine hPRP31 et du complexe CYPH/hPRP4/hPRP3 au sein de la snRNP U4 et enfin de la protéine NOP56 pour les snoRNP à boîtes C/D (Schultz *et al.*, 2006). Ainsi, il semble que suivant la séquence de cette hélice, des protéines différentes puissent être recrutées au sein des futures snoRNP/snRNP, et être stabilisées via des interactions avec la protéine SNU13.

De plus, comme décrit précédemment, les protéines NOP56 et NOP58 peuvent interagir ensemble par l'intermédiaire d'un domaine « coiled-coil » et des interactions entre la protéine Fibrillarine et la protéine NOP56 ont été montrées au niveau du domaine riche en hélices α de la Fibrillarine.

Une première hypothèse pour l'assemblage des protéines cœurs sur le snoARN à boîtes C/D serait donc l'existence d'un mécanisme qui, couplé à la transcription/épissage du snoARN, permettrait le recrutement de la protéine SNU13 au niveau du *K-Turn* du motif C/D et/ou du motif C'/D' par l'intermédiaire de la protéine IBP160. Par la suite, deux pré-complexes Fibrillarine-NOP56 et Fibrillarine-NOP58 seraient recrutés sur chacune des protéines SNU13 fixées sur l'ARN par l'intermédiaire de contacts à la fois au niveau du snoARN et de la protéine SNU13 pour former une particule mature (Bizarro *et al.*, 2014). Toutefois, cette première hypothèse semble contredite par le fait qu'aucune interaction n'ait pu être montrée entre la protéine SNU13 et les autres protéines cœur, ce qui explique probablement pourquoi il n'a jamais été possible chez les vertébrés de reconstituer une snoRNP à boîtes C/D stable et fonctionnelle à partir des 4 protéines cœur et d'un snoARN produits *in vitro* (pour revue Verheggen *et al.*, 2015). Ainsi, il semble que le mécanisme d'assemblage pour obtenir des snoRNP à boîtes C/D stables et fonctionnelles chez les vertébrés soit un peu plus complexe.

En effet, il a notamment été montré que la maturation de ces particules nécessitait un passage dans les Corps de Cajal (Narayanan *et al.*, 1999) ce qui implique l'existence de transporteurs spécifiques

et la nécessité d'une maturation particulière au sein de ces corps nucléaires. Il a ainsi été montré pour les snoARN transcrits par l'ARN polymérase II que le transporteur PHAX était impliqué dans le transport du snoRNP immature dans les Corps de Cajal. En effet, ce transporteur cible la coiffe m⁷G des snoRNP immatures par l'intermédiaire du complexe de fixation de la coiffe (CBC), pour transporter la particule dans les Corps de Cajal où la coiffe sera hyperméthylée en m^{2,2,7}G par l'enzyme TGS1. Ensuite, les snoRNP sont transportées dans le nucléole via le transporteur CRM1 (pour revue Verheggen et Bertrand, 2012). Pour les snoARN introniques, qui ne possèdent donc pas de coiffe, il a été montré que l'isoforme longue de la protéine TGS1, Tgs1 LF, interagissait avec les protéines NOP56 et NOP58 au niveau de leur domaine de localisation nucléolaire, ce qui empêcherait la particule d'aller dans le nucléole. La fixation du transporteur CRM1 de manière Ran-GTP dépendante sur la protéine Tgs1 LF libérerait alors ce signal et permettrait à la particule de migrer dans le nucléole (pour revue Verheggen et Bertrand, 2012).

De plus, de nombreuses expériences réalisées au cours de ces 15 dernières années ont mis en évidence l'existence d'un certain nombre de nouvelles protéines qui interagissent avec l'ARN et les protéines cœur des snoRNP à boîtes C/D et qui seraient nécessaires à la stabilité *in vivo* de ces particules chez les vertébrés. L'assemblage des snoRNP à boîtes C/D demanderait alors non seulement la présence des 4 protéines cœur et d'un snoARN, mais également de divers facteurs d'assemblage permettant le recrutement et/ou la stabilisation des protéines cœur sur le snoARN, ainsi que le transport de la snoRNP dans les Corps de Cajal puis dans le nucléole. Il a ainsi été montré que les protéines HSP90, NUFIP, BCD1, ZNHIT3 ainsi que le complexe hR2TP seraient impliqués à divers niveaux dans l'assemblage des snoRNP à boîtes C/D. Ces facteurs ne sont toutefois pas retrouvés au sein des particules matures.

2. La protéine chaperonne HSP90

a. Généralités

La protéine HSP90 (pour « Heat Shock Protein 90 »), ou Hsp82 chez la levure *S. cerevisiae*, est une protéine chaperonne essentielle, impliquée dans l'activation et la stabilisation de protéines spécifiques, appelées protéines clientes. Initialement identifiée comme impliquée dans les mécanismes mis en place par les cellules pour survivre à un choc thermique, il a par la suite été montré que la protéine HSP90 était exprimée de manière constitutive et était impliquée dans le maintien de l'intégrité de nombreuses protéines (pour revue Taipale *et al.*, 2010). Chez l'homme, il

existe deux isoformes α et β , aussi appelées HSP90AA et HSP90AB, et respectivement Hsp82 et Hsc82 chez la levure *S. cerevisiae* (Carretero-Paulet *et al.*, 2013). Les protéines HSP90 sont principalement localisées dans le cytoplasme, néanmoins il existe des isoformes spécifiques au niveau de la mitochondrie (protéine TRAP) ou du réticulum endoplasmique (protéine HSP90B) (Altieri *et al.*, 2012 ; Marzec *et al.*, 2012).

De manière générale, la protéine HSP90 va détecter au niveau de ses protéines clientes la présence de résidus hydrophobes en surface, c'est-à-dire exposés au solvant, et interagir avec afin de permettre à la protéine de se structurer correctement et ainsi éviter des phénomènes d'agrégations. Si la protéine cliente est trop déstructurée, la protéine HSP90 va recruter le protéasome qui va alors dégrader la protéine cliente (Theodoraki et Caplan, 2012).

En revanche, contrairement aux autres protéines chaperonnes, HSP90 n'interagit pas avec toutes les protéines présentant un problème de repliement, mais seulement avec un certain type de protéines et ce, uniquement si elles sont proches de leur état natif (Zhao et Houry, 2005). Ainsi, il a été montré que parmi ses protéines clientes, était surtout retrouvé des protéines kinases et des facteurs de transcription, et en particulier des récepteurs aux hormones stéroïdiennes (Joab *et al.*, 1984), aux glucocorticoïdes (Rafestin-Oblin *et al.*, 1989), à la progestérone (Catelli *et al.*, 1985) et aux œstrogènes (Redeuilh *et al.*, 1987). Toutefois, il a également été montré que la protéine HSP90 était impliquée dans l'assemblage de nombreux complexes protéiques, dont l'ARN polymérase II (Boulon *et al.*, 2010), le complexe mTOR (Delgoffe *et al.*, 2009) ou encore les snoRNP (Boulon *et al.*, 2008).

La faculté d'HSP90 d'interagir avec un si large éventail de protéines vient du fait que dans la plupart des cas, cette interaction est réalisée par l'intermédiaire de co-facteurs ou co-chaperonnes, qui agiront comme un adaptateur entre la protéine cliente et la protéine HSP90 (pour revue Röhl *et al.*, 2013).

b. Données structurales

La protéine HSP90 contient 3 domaines, un domaine N-terminal, un domaine médian et un domaine C-terminal (**Figure 9**). Le domaine N-terminal est impliqué dans la fixation de l'ATP, dont l'énergie fournie lors de son hydrolyse en ADP participerait au bon repliement de la protéine cliente. L'implication de l'ATP dans le fonctionnement d'HSP90 a été confirmée suite à la résolution par diffraction des rayons X de la structure de la protéine Hsp82 de levure *S. cerevisiae* (Prodromou *et al.*, 1997). En effet, la structure de cette protéine montre l'existence d'un domaine très conservé au niveau de l'extrémité N-terminale, capable de fixer l'ATP/ADP. À partir des informations fournies par cette structure, des mutations ont été introduites dans ce domaine, au niveau de résidus potentiellement impliqués dans la fixation et l'hydrolyse de l'ATP. Ces mutations eurent un effet dramatique sur la viabilité des levures (Panaretou *et al.*, 1998 ; Obermann *et al.*, 1998), démontrant ainsi l'importance de l'ATP dans le fonctionnement de la protéine HSP90. De même, la structure du domaine N-terminal de la protéine humaine montre là aussi l'existence d'un domaine de fixation et d'hydrolyse de l'ATP/ADP (**Figure 10**). Chez les eucaryotes, ce domaine est relié au domaine médian par l'intermédiaire d'un petit domaine de type « linker », riche en glycine, qui est également impliqué dans la fixation de l'ATP.

Le domaine médian est impliqué dans la majorité des interactions avec les protéines clientes d'HSP90. Il a par ailleurs été montré que les principales différences concernant les protéines clientes prises en charge par les protéines HSP90AA ou HSP90AB étaient dues à des résidus présents au niveau de ce domaine (Synoradzki et Bieganowski, 2015). La structure de ce domaine n'a pas encore été déterminée chez l'homme, contrairement à son homologue chez la levure *S. cerevisiae* (Meyer *et al.*, 2003) dont l'homologie avec HSP90AA est d'environ 60 %, et dont la structure de la protéine entière a été obtenue (Ali *et al.*, 2006). Enfin le domaine C-terminal se compose essentiellement d'un domaine qui permet à la protéine HSP90 de se dimériser (Minami *et al.*, 1994). Cette dimérisation semble être essentielle à l'activité de la protéine chaperonne. En effet des expériences sur des mutants d'HSP90 ont montré un fort ralentissement de l'activité catalytique de la protéine lorsque le domaine C-terminal était absent (Harris *et al.*, 2004). Ce domaine est également responsable du recrutement de facteurs protéiques qui vont être capables de moduler l'activité d'HSP90.

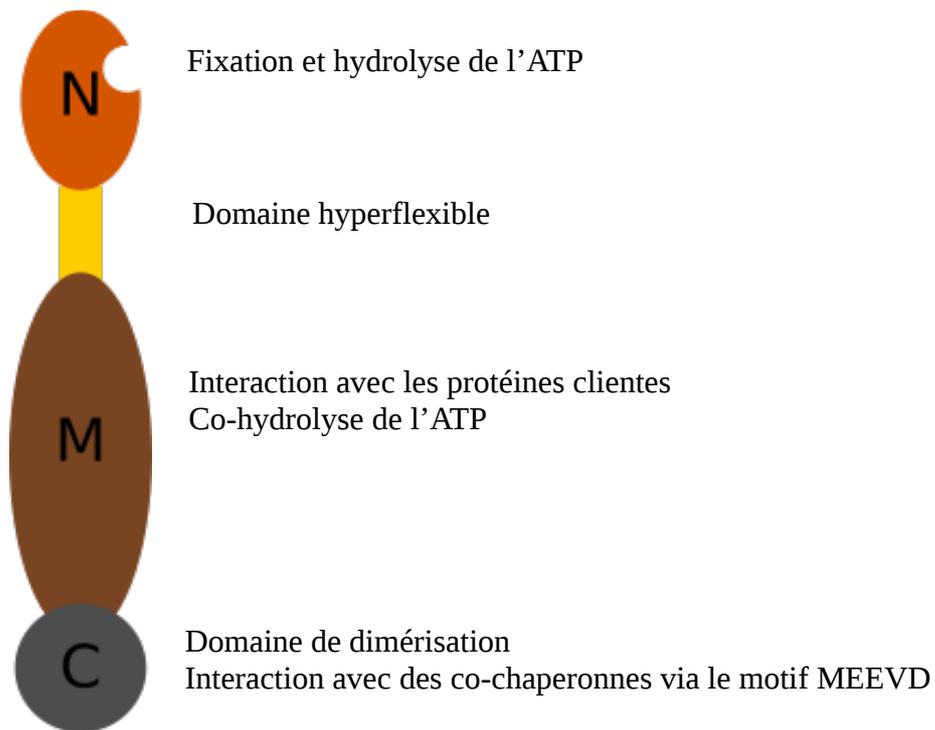


Figure 9: Schéma de l'organisation de la protéine HSP90.

La protéine HSP90 est constituée de 3 domaines permettant notamment l'interaction avec les protéines clientes (domaine M), la fixation de l'ATP (domaine N) et l'interaction avec des co-chaperonnes (domaine C)

N : domaine N-terminal ; M : domaine Median ; C : domaine C-terminal

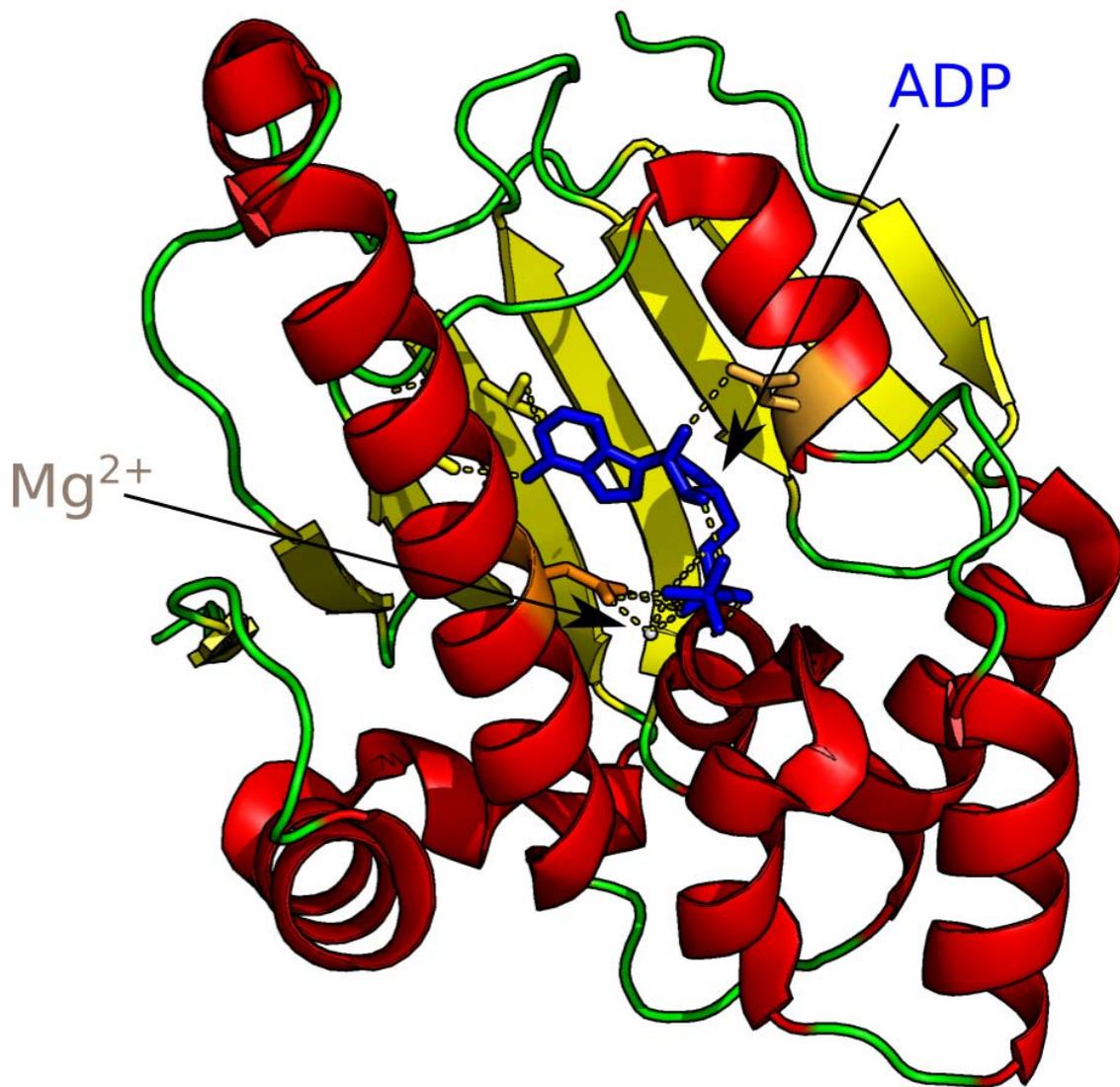


Figure 10: Structure 3D du domaine N-terminal de la protéine HSP90 humaine (code PDB : 2XK2).

La résolution de la structure tridimensionnelle de la protéine HSP90 humaine montre la présence d'une molécule d'ADP (ici en bleu) en interaction avec des résidus spécifiques de la protéine. Un ion magnésium (ici en gris) est également présent et stabilisé par des résidus de la protéine.

Ces protéines, ou co-chaperonnes, vont reconnaître l'extrémité MEEVD de la protéine HSP90 (IEEVD dans le cas de la protéine HSP70) via un domaine particulier, appelé domaine TPR (« Tetratricopeptide Repeat »). Ces domaines sont formés d'une répétition (entre 3 et 16) d'un motif de 34 acides aminés repliés en hélice-boucle-hélice, appelé motif TPR (Young *et al.*, 1998). L'interaction entre le ou les domaines TPR de la co-chaperonne et le motif MEEVD de la protéine HSP90 se fait via 5 résidus très conservés qui forment un motif nommé « dicarboxylate clamp » (**Figure 11**), une poche (ou pince) dans laquelle l'extrémité MEEVD d'HSP90 va s'insérer (Russell *et al.*, 1999 ; Scheufler *et al.*, 2000).

c. Cycle catalytique

De manière intéressante, il a été montré que le domaine N-terminal seul ne présentait aucune activité catalytique pour l'ATP. De plus, comme indiqué précédemment, la protéine HSP90 fonctionne sous la forme d'un dimère via son domaine C-terminal qui permet également de recruter les co-facteurs d'HSP90. Enfin, il semble que l'hydrolyse de l'ATP soit en partie permise grâce à des résidus présents au sein du domaine médian (Meyer *et al.*, 2003). Ainsi, l'ATP se fixerait au niveau du domaine N-terminal puis serait « présenté » au domaine médian grâce au domaine « linker » très flexible présent entre les deux domaines avant d'être hydrolysé.

L'ensemble de ces données a conduit à un modèle de fonctionnement, dans lequel le dimère d'HSP90 forme une pince moléculaire qui évolue entre deux états « ouvert » ou « fermé » suite à la fixation d'ATP. Au cours de ce cycle, le dimère d'HSP90 interagit avec une protéine cliente au niveau des domaines médians via une association avec une ou plusieurs protéines adaptatrices. Ce cycle peut aussi être régulé par l'intermédiaire de l'interaction entre le domaine TPR d'une protéine co-chaperonne et l'extrémité MEEVD d'HSP90 (**Figure 12**). Certaines imprécisions demeurent toutefois concernant ce cycle d'activité. Ainsi, il a aussi été montré que l'interface créée par l'interaction entre les deux domaines C-terminaux de chaque monomère n'était pas figée dans le temps mais dynamique (Ratzke *et al.*, 2010). De même, il semble que le domaine C-terminal puisse fixer l'ATP mais uniquement quand le domaine N-terminal est lui-même déjà occupé avec de l'ATP (Marcu *et al.*, 2000 ; Söti *et al.*, 2002).

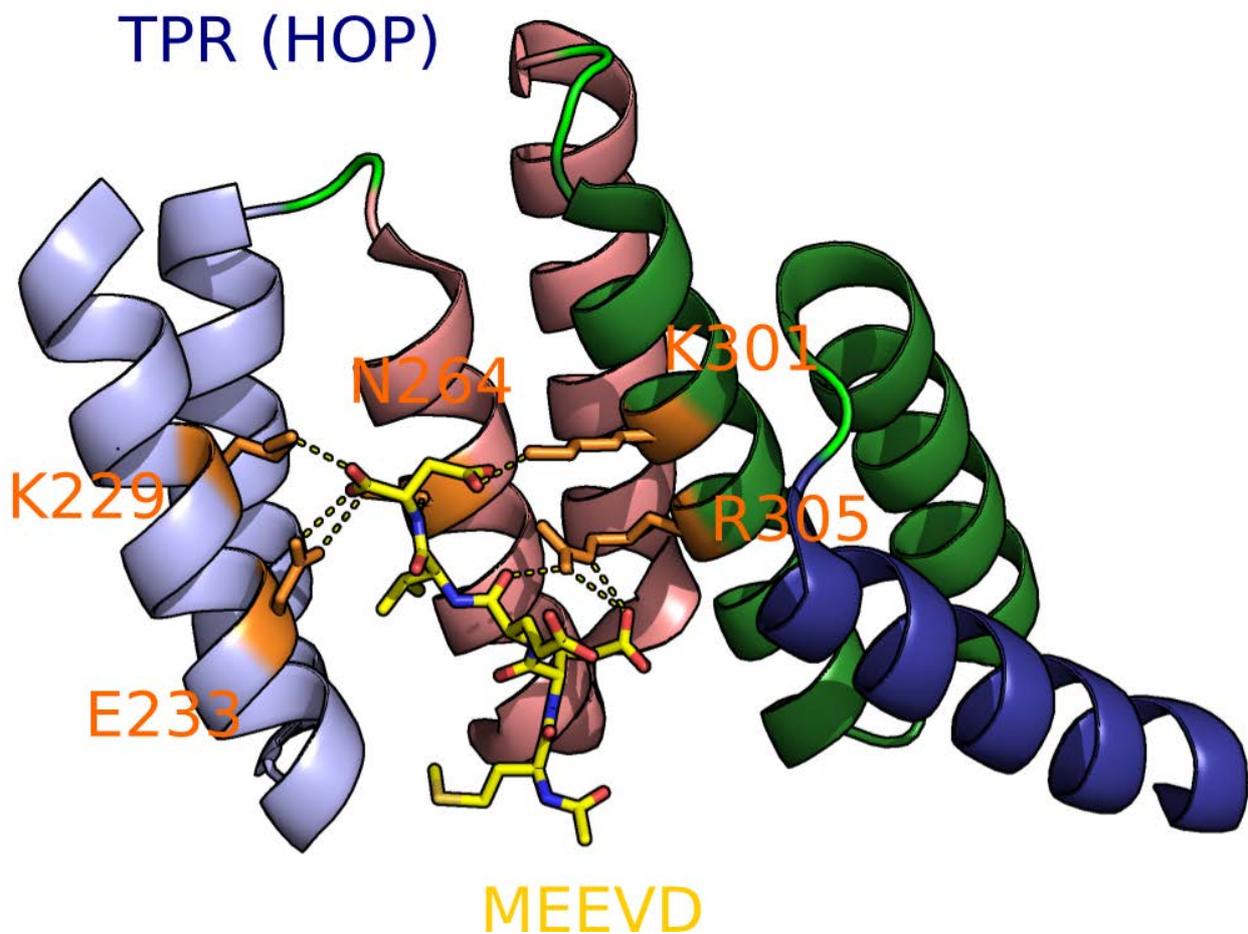


Figure 11: Motif MEEVD de la protéine HSP90 en complexe avec l'un des domaines TPR d'une protéine cochaperone humaine (HOP) (code PDB : 1ELR).

Les résidus, ici en orange, du second domaine TPR d'une des protéines co-chaperonne d'HSP90, la protéine HOP (« HSP70-HSP90 Organizing Protein »), forment une « dicarboxylate clamp » qui interagit par des liaisons électrostatiques avec l'extrémité MEEVD (en jaune) de la protéine HSP90 (Scheufler *et al.*, 2000). Le nom de ce motif vient du fait qu'il interagit avec le peptide MEEVD principalement par des liaisons hydrogènes avec deux acides carboxyliques, l'extrémité C-terminale du peptide et la chaîne latérale de l'acide aspartique.

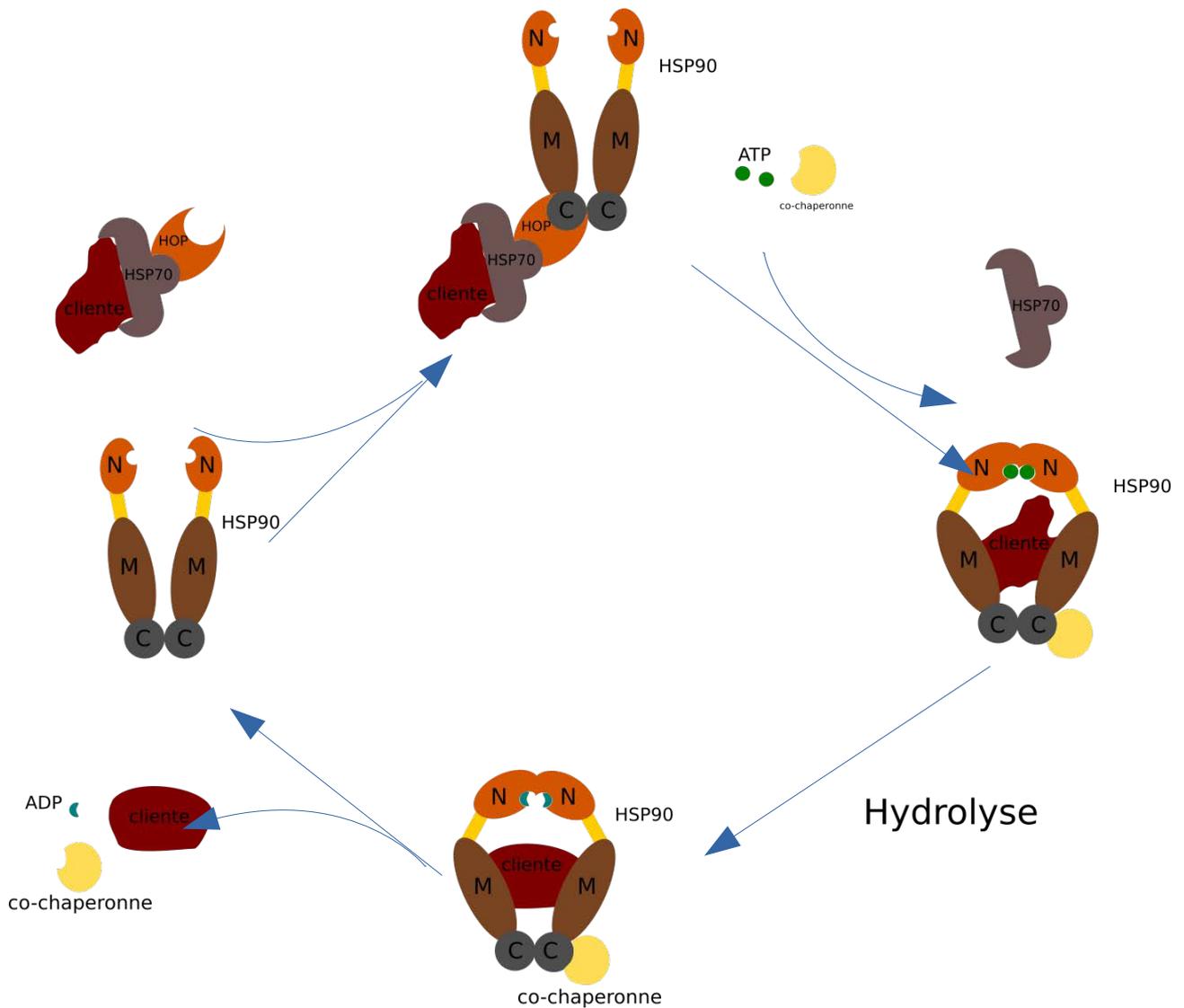


Figure 12: Cycle d'activité de la chaperone HSP90.

La prise en charge d'une protéine cliente par HSP90 se fait via une protéine adaptatrice, comme HSP70, qui reconnaît la protéine cliente. Cette protéine est ensuite chargée sur le domaine médian de HSP90 par l'intermédiaire de l'action d'une co-chaperonne comme HOP via une interaction entre un domaine TPR et le motif MEEVD de protéine HSP90. La fixation d'une molécule d'ATP au niveau des domaines N-terminaux entraîne un changement de conformation du dimère HSP90 qui prend une conformation « fermée ». L'hydrolyse de l'ATP en ADP fournit l'énergie nécessaire au bon repliement de la protéine cliente qui est ensuite relarguée (adapté de Li *et al.*, 2011).

d. Rôle dans l'assemblage des snoRNP à boîtes C/D

La geldanamycine est un inhibiteur compétitif de la fixation de l'ATP sur le domaine N-terminal d'HSP90 (**Figure 13**) (Miyata, 2005). En utilisant cet inhibiteur sur des cellules HEK293, notre équipe et celle d'E. Bertrand à Montpellier ont montré que la protéine HSP90 était impliquée dans la stabilité des particules ARN-protéine contenant des protéines de la famille L7Ae. Ainsi, l'inhibition de la protéine HSP90 par la geldanamycine entraîne une diminution de l'accumulation des snoARN à boîtes C/D, H/ACA et du snARN U4. De plus, il a été montré que cette inhibition avait un impact sur le taux des protéines SNU13, NHP2, SBP2 et NOP58 (Boulon *et al.*, 2008). Ensembles, ces données indiquent un rôle possible de la protéine HSP90 dans la formation de certains complexes RNP, dont les snoRNP à boîtes C/D, probablement en contrôlant le bon repliement des protéines cœur sur l'ARN lors de l'assemblage de la particule. Toutefois, l'assemblage de ces particules a lieu au sein du noyau, or la protéine HSP90 n'y est présente que dans des situations particulières comme lors d'un stress thermique. La protéine HSP90 agit donc probablement en amont de l'assemblage de la particule, au niveau du cytoplasme.

3. Le complexe hR2TP, un cofacteur de la protéine chaperonne HSP90

a. Découverte du complexe R2TP/hR2TP

L'analyse à haute résolution des partenaires de la protéine Hsp82 chez la levure *S. cerevisiae* a révélé l'existence de 2 nouveaux cofacteurs, les protéines Tah1 et Pih1 (Zhao *et al.*, 2005). Une analyse des partenaires de ces 2 co-chaperonnes a montré qu'elles interagissaient toutes les deux avec les protéines Rvb1 et Rvb2 (Zhao *et al.*, 2005). Ces 4 protéines forment alors un complexe dont l'un des rôles serait de réguler l'activité de la protéine chaperonne Hsp82. Ce complexe a été nommé en référence aux protéines qui le composent, à savoir les protéines Rvb1, Rvb2, Tah1 et Pih1 : soit le complexe R2TP (Zhao *et al.*, 2005). De manière remarquable, certaines des protéines constituant le complexe R2TP avaient par le passé déjà été identifiées comme impliquées dans la biogenèse des snoRNP à boîtes C/D. Ainsi, il avait été montré chez la levure *S. cerevisiae* que la protéine Rvb2 était nécessaire à la production de snoARN à boîtes C/D et H/ACA. En effet, des études génétiques chez la levure *S. cerevisiae* ont montré qu'une déplétion de Rvb2 entraînait un problème de localisation cellulaire de la protéine Nop1, l'homologue chez la levure de la protéine humaine Fibrillarine (King *et al.*, 2001). De plus, des expériences de double-hybride chez la levure *S. cerevisiae* ont montré l'existence d'une interaction entre la protéine cœur des snoRNP à boîtes C/D, la protéine Nop58, et la protéine Pih1, dont la déplétion affecte la maturation du pré-ARNr (Gonzales *et al.*, 2005).

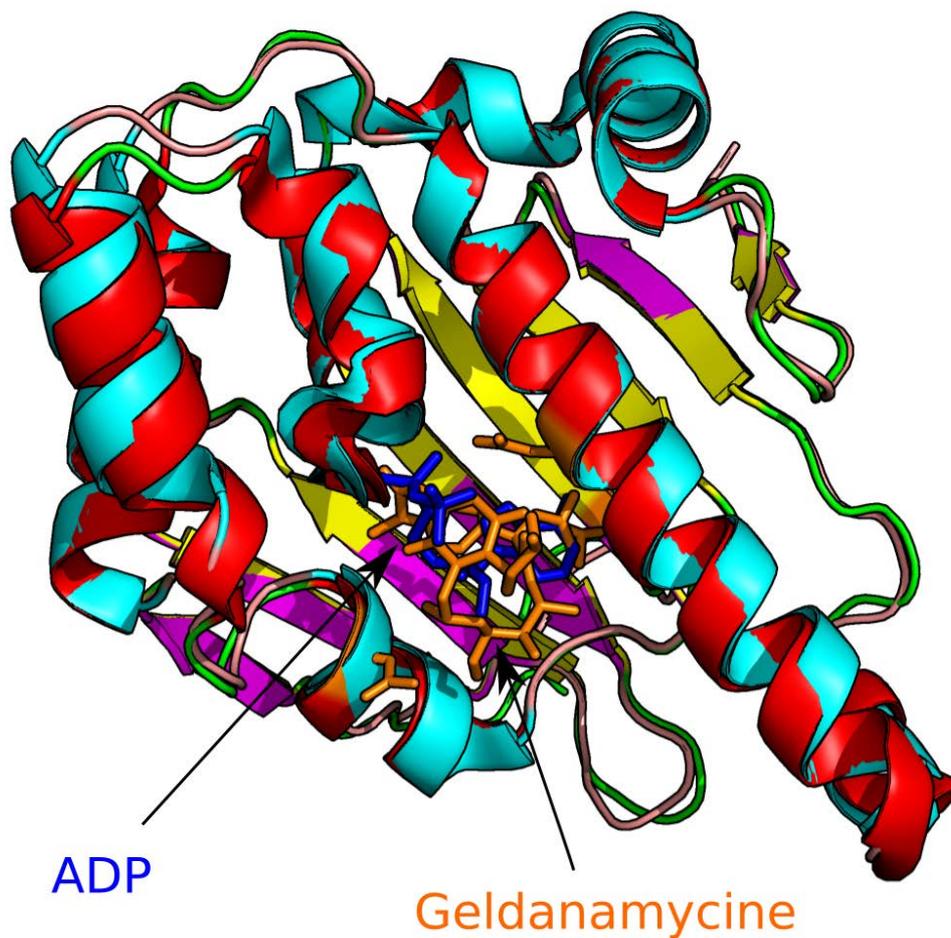


Figure 13: Superposition des structure du domaine N-terminal d'HSP90 en complexe avec l'ADP (en bleu) et de la Geldanamycine (en orange).

La superposition a été réalisée sur *Pymol* à partir des structures résolues du domaine N-terminal d'HSP90 en complexe avec l'ADP (code PDB : 2xk2), ici en bleu, et du même domaine en complexe avec la Geldanamycine (code PDB : 1OSF), ici en orange. L'alignement montre clairement que ces deux molécules partagent le même site de fixation au niveau d'HSP90.

Chez l'homme, des analyses des partenaires de la protéine HSP90 humaine ont mis en évidence l'existence d'un complexe dont la composition était proche du R2TP de levure (Te *et al.*, 2007). Ce complexe est composé des protéines TIP48, TIP49 et PIH1D1, respectivement les homologues humaines des protéines Rvb2, Rvb1 et Pih1. Une analyse des bases de données d'interaction protéines-protéines chez la drosophile a mis en évidence l'existence d'interactions entre les protéines Hsp82 et Pih1 de drosophile et une protéine appelée Spaghetti (Giot *et al.*, 2003). Un homologue de la protéine Spaghetti a été trouvé chez l'homme, la protéine FLJ21908, aussi appelée RPAP3 pour « RNA Polymerase II-Associated Protein » (Boulon *et al.*, 2008).

Les protéines TIP48, TIP49, PIH1D1 et RPAP3 forment ainsi le complexe hR2TP, homologue humain du complexe R2TP de levure. Des expériences de double-hybride et des tests d'interaction *in vitro* réalisées par notre laboratoire et celui d'E. Bertrand ont montré un large réseau d'interaction entre les protéines du hR2TP et des protéines cœur des snoRNP à boîtes C/D. Ainsi, des interactions directes ont été mises en évidence au sein du complexe hR2TP entre les protéines RPAP3 et PIH1D1, et entre les protéines TIP48 et TIP49. D'autres interactions ont également été montrées entre la protéine RPAP3 et les protéines TIP48 et HSP90 et entre les protéines PIH1D1 et NOP58. De plus, des expériences de sédimentation différentielles sur gradient de glycérol d'extraits cytoplasmiques de cellules HeLa ont montré l'existence de macro-complexes composés des protéines NOP58, Fibrillarine, RPAP3, TIP48, TIP49 et HSP90. Ensembles, ces données suggèrent l'implication du complexe hR2TP, via les interactions entre HSP90, RPAP3 et PIH1D1, dans la modulation de l'activité chaperonne d'HSP90, nécessaire au remodelage des protéines cœur lors de l'assemblage des snoRNP (Boulon *et al.*, 2008).

b. Les protéines TIP48/TIP49

i. Données Structurales

Les protéines TIP48 et TIP49 (« TATA-binding protein Interacting Protein »), également appelées respectivement RuvBL2 et RuvBL1 ou encore Reptine et Pontine (et Rvb2 et Rvb1 chez la levure *S. cerevisiae*), ont depuis longtemps été identifiées comme nécessaires à la viabilité cellulaire. Ainsi, elles seraient impliquées dans de nombreux processus biologiques fondamentaux comme la régulation de la transcription, les réponses aux dommages de l'ADN, le développement cellulaire, l'apoptose, le remodelage des histones ou encore la maturation des snoRNP (pour revue Jha et Dutta, 2009). Les protéines TIP48 et TIP49, qui partagent 43 % d'identité et 63 % de similarité

(Kanemaki *et al.*, 1999 ; Gorynia *et al.*, 2011), sont des ATPases appartenant à la superfamille des AAA+ (« ATPase Associated with diverse cellular Activities »), initialement identifiées comme des ADN hélicases.

La structure de la protéine TIP49 humaine (**Figure 14**) a été résolue et a permis de montrer que cette protéine était constituée de trois domaines DI, DII et DIII (Matias *et al.*, 2006). Le domaine DI contient deux motifs très conservés dans les ATPases, le motif Walker A (GPPGTGKT), impliqué dans la fixation de l'ATP, et le motif Walker B (DEVH), impliqué dans l'hydrolyse de l'ATP. Ce domaine contient également un motif senseur I ainsi qu'une arginine très conservée qui forme un motif appelé « arginine-finger » (R357). Cette arginine joue un rôle important pour l'hydrolyse de l'ATP au sein de la structure quaternaire.

Le domaine DIII contient le motif senseur II et forme avec le domaine DI le cœur protéique AAA+, nécessaire et suffisant à l'activité ATPase de la protéine (Silva-Martin *et al.*, 2016). Le domaine DII est lui situé au sein du domaine DI entre les motifs Walker A et Walker B. Il adopte un repliement de type OB (« Oligonucleotide-Binding ») qui est rattaché au cœur AAA+ par un linker très flexible. De manière intéressante, le domaine DII est absent dans les protéines homologues bactériennes (Putnam *et al.*, 2001). Ce domaine possède une structure proche de celle d'un domaine de la protéine RPA, une protéine liant l'ADN et qui est impliquée dans la réplication chez les eucaryotes. De plus, ce domaine présente à sa surface des résidus principalement basiques (Matias *et al.*, 2006). Enfin, des expériences ont montré que ce domaine pouvait lier aussi bien l'ADN que l'ARN. Il a donc été proposé qu'il pouvait servir d'intermédiaire entre le cœur AAA+ et un substrat ADN/ARN (Matias *et al.*, 2006).

Chacune des protéines semble pouvoir former un homohéxamère (**Figure 15**) qui montre la présence d'une molécule d'ATP entre chaque monomère (Petukhov *et al.*, 2012). Toutefois, il a aussi été montré que les protéines TIP48 et TIP49 pouvaient interagir ensemble (Kanemaki *et al.*, 1999). La résolution de la structure d'un complexe formé entre ces deux protéines, sans le domaine DII trop flexible, a montré que TIP49 et TIP48 s'organisaient cette fois en un hétérohéxamère, alternant un monomère de chaque protéine (Gorynia *et al.*, 2011).

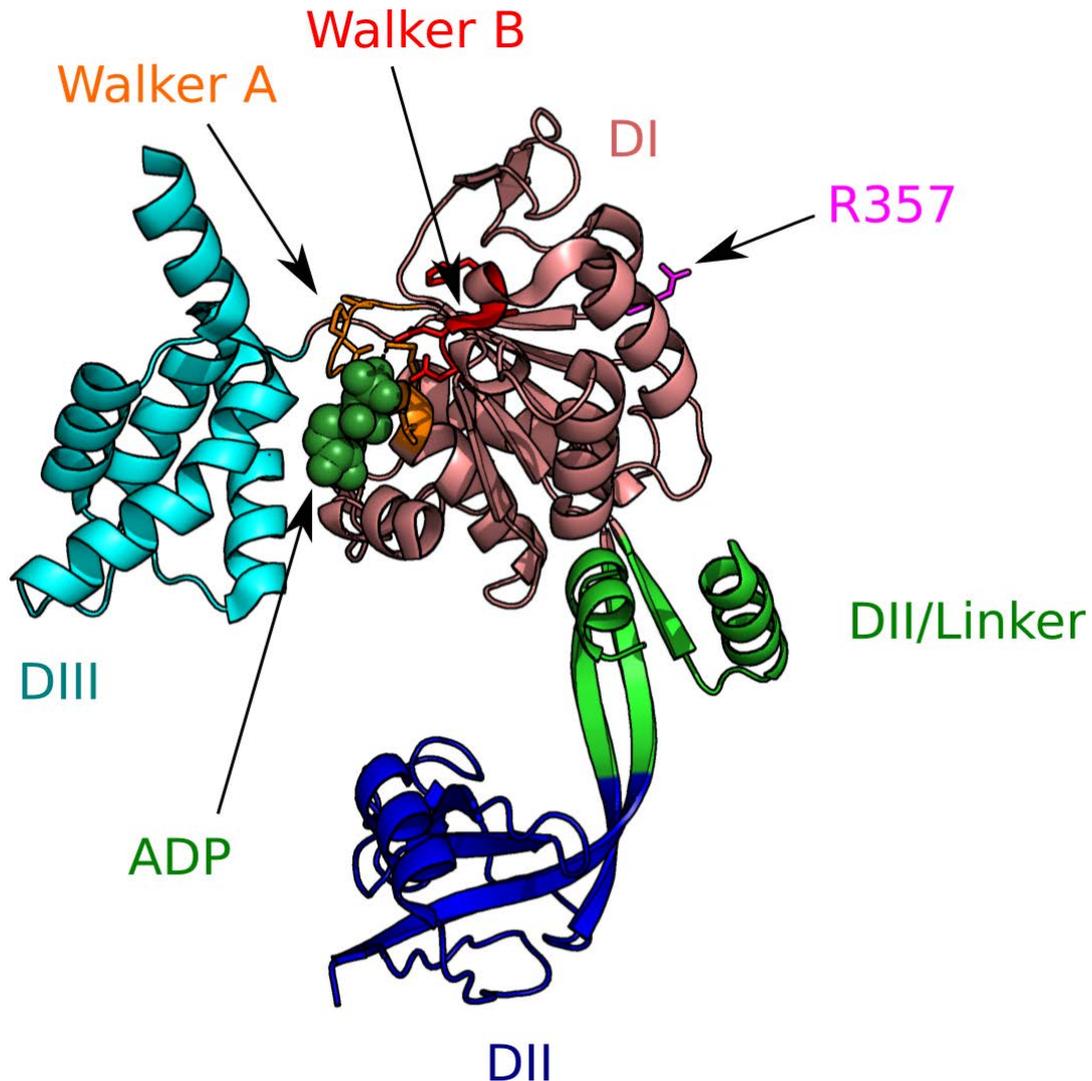


Figure 14: Structure 3D de la protéine TIP49 résolue par diffraction aux rayons X (code PDB : 2C9O).

Le monomère TIP49 est constitué de 3 domaines. Le cœur AAA+ est formé par les domaine DI (ici en rouge) et DIII (ici en cyan). Le domaine DI fixe l'ATP (ici en orange) et participe à son hydrolyse avec le domaine DIII. Le domaine DII (ici en bleu) est rejeté sur le côté du cœur protéique afin de permettre les interactions avec l'ARN/ADN, voire certaines protéines. Le domaine DII est relié au cœur de la protéine par un linker (ici en vert) très flexible.

DI : Domaine I ; DII : Domaine II ; DIII: Domaine III

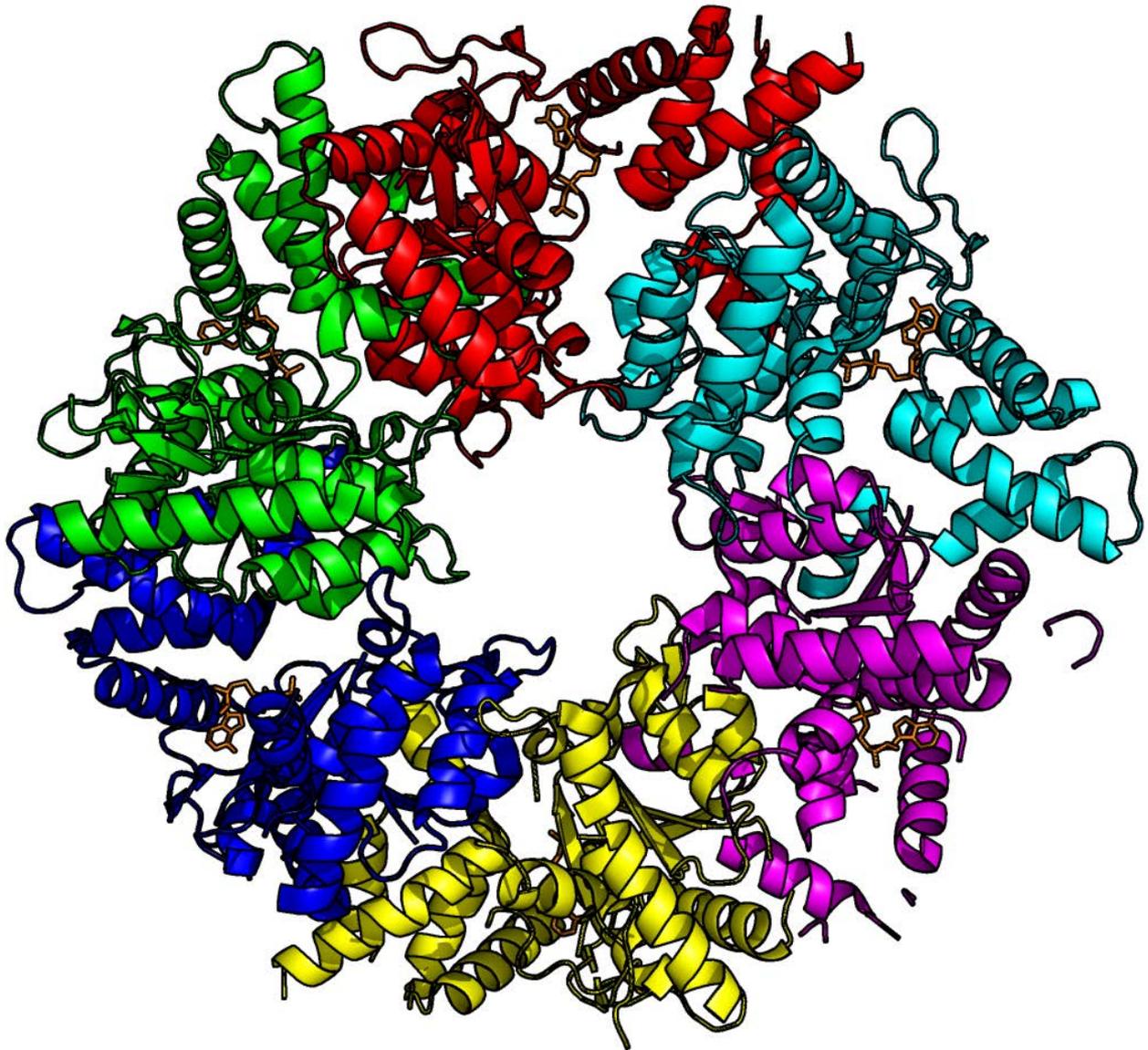


Figure 15: Structure 3D de l'homohéxamère TIP48 humain (code PDB : 3UK6).

Le domaine DII de la protéine TIP48 a été retiré pour faciliter la cristallisation. La résolution de cette structure montre un homohéxamère (chaque monomère est d'une couleur différente) formé par des interactions entre les domaines AAA+ de chaque monomère (Petukhov *et al.*, 2012). L'ADP est ici en orange, au sein de chaque monomère, bloqué par le monomère suivant.

Enfin, la résolution récente de la structure par diffraction des rayons X des homologues de ces deux protéines chez *Chaetomium thermophilum*, un organisme thermophile, combinée à l'étude de données SAXS (Diffusion des rayons X aux petits angles) et de cryo-électro-microscopie, montrent que deux hétérohexamères pourraient interagir par l'intermédiaire du domaine DII pour former un dodécamère (**Figure 16**) (Silva-Martin *et al.*, 2016).

ii. Données fonctionnelles

La protéine TIP49 a été identifiée pour la première fois lors d'un crible d'interaction avec la protéine TBP (« TATA-Binding Protein »), une sous-unité du facteur de transcription de l'ARN polymérase II TFIID, dans des extraits nucléaires rénaux de rat (Kanemaki *et al.*, 1997). La protéine TIP48, qui peut former un complexe avec la protéine TIP49, a été identifiée peu après chez la levure *S. cerevisiae* (Rvb2) et dans les cellules de mammifère (Kanemaki *et al.*, 1999). Du fait de la présence des motifs Walker A et B caractéristiques des ATPases, et d'une importante homologie avec l'hélicase RuvB bactérienne, les auteurs les ont tout d'abord identifiées comme des hélicases ATP-dépendantes. L'activité ATPase a été confirmée peu après, et semble être stimulée par la présence d'ADN simple-brin (Makino *et al.*, 1999), malgré des premiers résultats qui semblaient indiquer le contraire (Qiu *et al.*, 1998).

Toutefois, ce rôle d'hélicase reste controversé dans la littérature, même s'il semble admis que la fixation et l'hydrolyse de l'ATP soient responsables d'un changement de conformation de l'héxa/dodécamère, qui serait ensuite transmis au substrat (protéine, ADN ou ARN ; pour revue Hanson et Whiteheart, 2005). De plus, diverses études sur ces hélicases ont montré qu'elles étaient retrouvées dans de nombreux complexes protéiques, comme le complexe d'acétylation des histones TIP60 (Ikura *et al.*, 2000) ou encore les complexes de remodelage de la chromatine INO80 (Shen *et al.*, 2000) et SWR-C (Krogan *et al.*, 2003). Les protéines TIP48 et TIP49 sont également retrouvées en interaction avec la protéine TERT, un composant majeur de la télomérase (Venteicher *et al.*, 2008), et avec un certain nombre de facteurs de transcription comme c-Myc (Dugan *et al.*, 2002), E2F (Tarangelo *et al.*, 2015), la β -Caténine (Feng *et al.*, 2003) ou encore TBP (Kanemaki *et al.*, 1997). L'ensemble de ces données suggère ainsi plutôt un rôle au niveau du remodelage des protéines qu'une fonction hélicase.

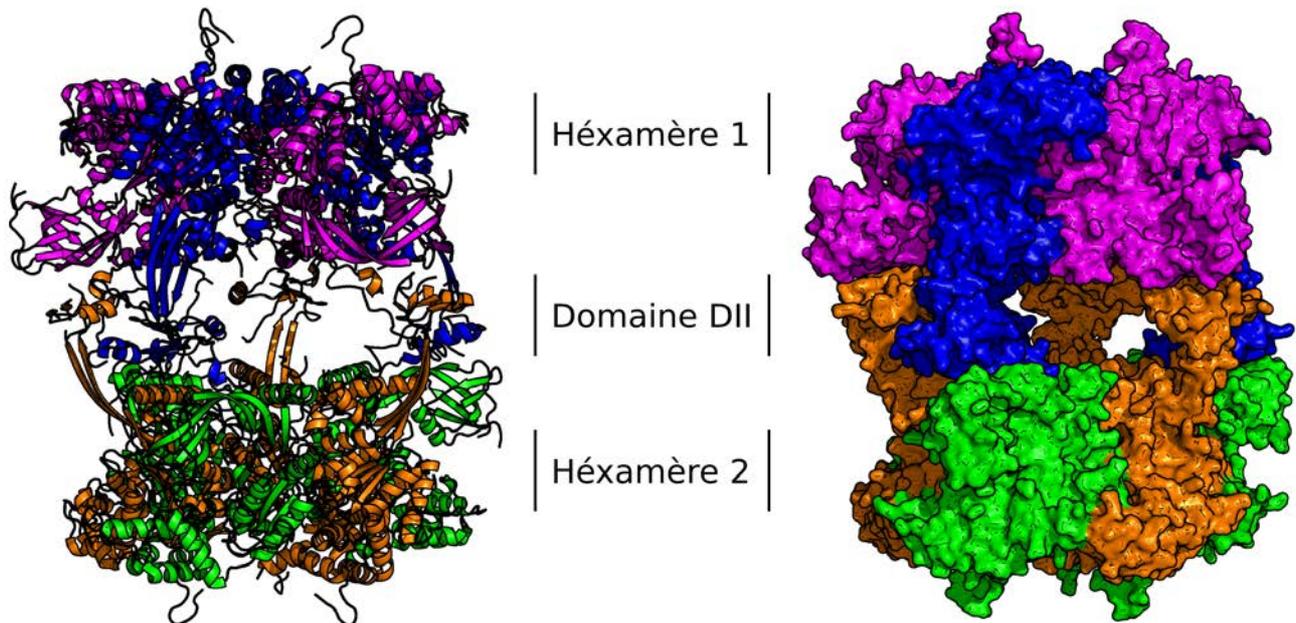


Figure 16: Reconstitution de la structure 3D du dodécamère TIP48-TIP49.

Cette reconstitution a été faite à partir de la structure de l'héxamère TIP48-TIP49 de *Chaetomium thermophilum* (code PDB : 5FM7) avec le logiciel *Pymol*.

D'après les données de cryo-électro-microscopie, l'interface entre les deux hétérohéxamères se ferait par l'intermédiaire du domaine DII (Silva-Martin *et al.*, 2016).

À gauche : Vue des structures secondaires de la reconstitution du dodécamère.

À droite : La même reconstitution mais avec une vue de la surface du complexe.

Chaque couleur correspond au monomère TIP49 (en bleu et en orange) ou TIP48 (en rose ou en vert).

Enfin, comme décrit précédemment, les protéines TIP48 et TIP49 font également partie du complexe hR2TP, un macro-complexe contenant également les protéines RPAP3 et PIH1D1. Au sein de ce complexe, il a été montré que les protéines TIP48 et TIP49 étaient impliquées dans la voie de signalisation des PIKK (Izumi *et al.*, 2010), dans l'apoptose, via des interactions avec la protéine MONAD (Saeki *et al.*, 2006 ; Ni *et al.*, 2009), dans l'assemblage de l'ARN polymérase II (Boulon *et al.*, 2010) et dans la biogenèse des snoRNP à boîtes C/D (King *et al.*, 2001 ; Boulon *et al.*, 2008) (**Figure 17**).

c. La protéine Pih1 (Nop17)/PIH1D1

i. Données Fonctionnelles

L'ORF (« Open Reading Frame ») YHR034C de levure code une protéine nommée Nop17 (« Nucleolar protein 17 ») qui fut initialement identifiée par des expériences de double-hybride chez la levure *S. cerevisiae* comme partenaire des protéines Nop53, une protéine nécessaire à la biogenèse des ARNr 25S et 5,8S (Granato *et al.*, 2005), et Nop58, l'une des protéines cœur des snoRNP à boîtes C/D (Gonzales *et al.*, 2005). L'étude de cette protéine a révélé que sa délétion chez la levure n'était pas létale mais entraînait un phénotype thermosensible. (Gonzales *et al.*, 2005). Ainsi, en l'absence de Nop17, la croissance des levures est ralentie à 37 °C. De plus, des analyses à haut débit ont révélé la présence des protéines Nop17 et Nop56 au sein des mêmes complexes pré-ribosomiques (Krogan *et al.*, 2004). L'ensemble ces données suggère que la protéine Nop17 serait impliquée dans la biogenèse des ARNr. Toutefois, les travaux de Gonzales *et al.* semblent indiquer que cette protéine interviendrait plutôt au niveau de l'assemblage des snoRNP à boîtes C/D. En effet, les auteurs ont montré par des analyses de suivi de fluorescence de protéines fusionnées à une étiquette GFP qu'en l'absence de Nop17, les protéines cœur des snoRNP à boîtes C/D (Nop1, Nop58, Nop56, Snu13) ne s'accumulaient plus dans le nucléole mais étaient retrouvées de manière aberrante dans le cytoplasme et le noyau. Enfin bien que la protéine Nop17 n'interagisse pas directement avec les snoARN, sa déplétion affecte l'interaction entre Nop58 et les snoARN U3 ou U14 (Gonzales *et al.*, 2005).

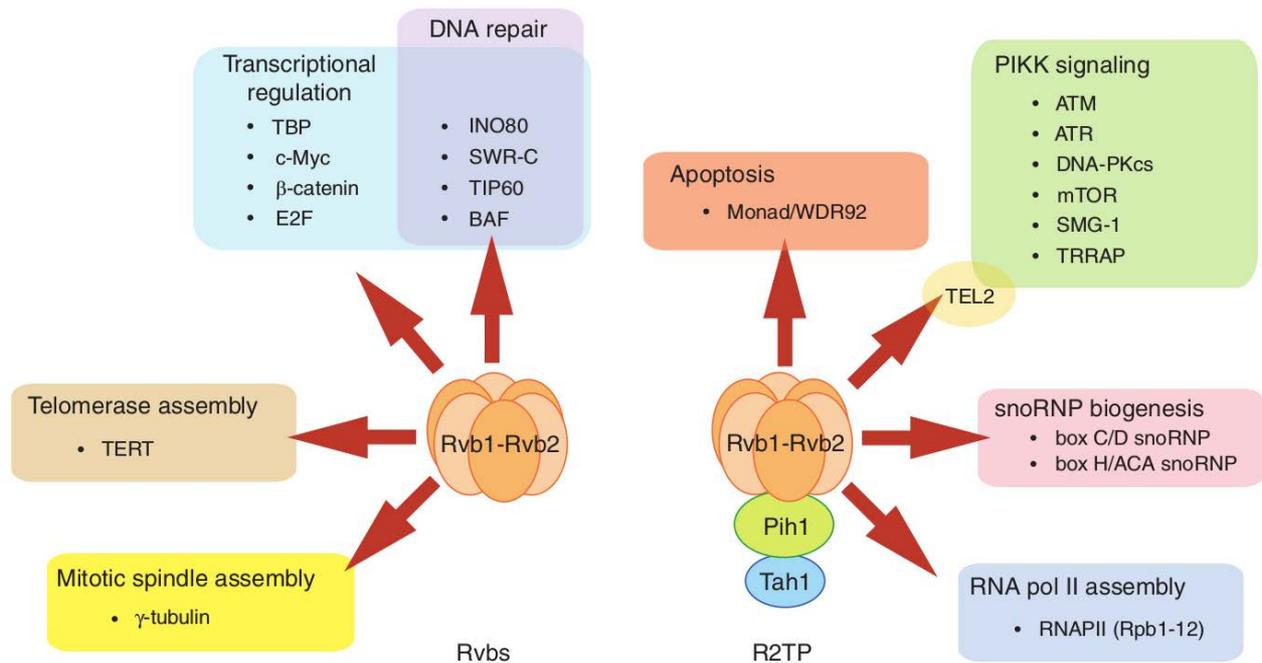


Figure 17: Fonctions des protéines TIP49/Rvb1 et TIP48/Rvb2 (Kakihara et Saeki, 2014).

Les protéines TIP48/TIP49 ont été montrées comme impliquées dans la réparation de l'ADN, l'assemblage de la télomérase ou encore la régulation de la transcription par l'intermédiaire d'interactions avec de nombreux facteurs de transcription.

Au sein du complexe R2TP qui contient également les protéines Pih1/PIH1D1 et Tah1/RPAP3, les protéines TIP48 et TIP49 interviennent dans la voie de signalisation des PIKK, dans l'apoptose, dans l'assemblage de l'ARN polymérase II et dans l'assemblage des snoRNP.

Parallèlement à ces travaux, une équipe identifia, toujours chez la levure *S. cerevisiae*, un nouveau partenaire de la protéine chaperonne Hsp82. Désignée initialement par le nom de son ORF, YHR034C, la protéine fut renommée Pih1, pour « Protein interacting with Hsp82 » (Zhao *et al.*, 2005). L'analyse de l'interaction entre Pih1/Nop17 et Hsp82 fut le point de départ de la mise en évidence de l'existence du complexe R2TP et de son rôle probable dans l'assemblage des snoRNP à boîtes C/D (Zhao *et al.*, 2005 ; Zhao *et al.*, 2008 ; Boulon *et al.*, 2008).

Au sein de ce complexe, des interactions ont été observées entre la protéine Pih1 et les protéines Tah1, une co-chaperonne d'Hsp82 (Zhao *et al.*, 2005, Back *et al.*, 2013), et Rvb1/2. Plus précisément, il a été montré *in vitro* et *in vivo* que Pih1 était une protéine instable requérant la présence du complexe chaperonne/co-chaperonne Hsp82/Tah1 pour ne pas s'agréger (Zhao *et al.*, 2008). De plus, d'autres auteurs ont montré qu'une fois le complexe Tah1/Pih1 formé, celui-ci inactivait l'activité ATPase d'Hsp82 (Eckert *et al.*, 2010). Toutefois, des travaux récemment réalisés dans notre laboratoire suggèrent que, chez la levure en phase stationnaire de croissance, la présence de la protéine Tah1 est suffisante pour la stabilisation de la protéine Pih1, indépendamment de la présence d'Hsp82 (Quinternet *et al.*, 2015).

L'ensemble de ces données suggère l'implication de Pih1 dans l'assemblage des snoRNP à boîtes C/D, où la protéine Pih1 serait stabilisée grâce à son interaction avec Tah1 et, hors phase stationnaire de croissance, en présence également d'Hsp82. Une fois la protéine Pih1 stabilisée, la protéine Hsp82 inactive libérerait le complexe Tah1/Pih1, stable, qui pourrait ensuite interagir avec les protéines Rvb1 et Rvb2 pour former le complexe R2TP, qui à son tour interagirait avec Nop58 par l'intermédiaire de Pih1.

Chez l'homme, une analyse des bases de données du génome a montré l'existence d'un homologue de la protéine Pih1 de levure. Cette protéine, initialement identifiée sous le code FLJ20643 fut nommée PIH1D1 (« Protein Interacting with HSP90 Domain containing protein 1 ») (Zhao *et al.*, 2005 ; Te *et al.*, 2007 ; Horejsí *et al.*, 2010). Comme chez la levure, la protéine PIH1D1 s'associe avec HSP90, RPAP3, TIP48, TIP49, NOP58 et NOP56 (Te *et al.*, 2007 ; Boulon *et al.*, 2008 ; Yoshida *et al.*, 2013). De plus, cette protéine semble pouvoir reconnaître spécifiquement des protéines phosphorylées grâce son domaine N-terminal. C'est notamment le cas pour la protéine TEL2, une protéine essentielle à la stabilité des complexes mTOR ou SMG1 (Horejsí *et al.*,

2010 ; Kamano *et al.*, 2013 ; Horejsí *et al.*, 2014). Ainsi, il est possible que cela soit aussi le cas concernant l'interaction entre PIH1D1 et le domaine NOP de NOP58 (Quinternet *et al.*, 2015).

La protéine PIH1D1 est également impliquée dans la régulation de l'apoptose induite par la doxorubicine, une molécule utilisée pour le traitement de certains cancers. En effet, des expériences d'ARN interférence dirigées contre l'ARNm de PIH1D1 ont montré une augmentation de la mort cellulaire par apoptose des cellules traitées avec la doxorubicine (Inoue *et al.*, 2010). La protéine PIH1D1 semble également impliquée dans l'expression des gènes codant les ARNr. Plus précisément, la protéine PIH1D1 interagit avec l'histone H4 au niveau des promoteurs des gènes codant les ARNr, puis recrute le complexe de remodelage de la chromatine SWI/SNF pour initier leur transcription par l'ARN polymérase I (Zhai *et al.*, 2012).

ii. Données structurales

Les protéines PIH1D1 humaine et de levure contiennent deux domaines conservés, un domaine N-terminal PIH1D1-N et un domaine C-terminal PIH1D1-C. Le domaine PIH1D1-N est capable de reconnaître des phosphopeptides/protéines, comme la protéine TEL2 suite à sa phosphorylation par la protéine CK2 (Horejsí *et al.*, 2010). La structure du domaine PIH1D1-N de la protéine humaine libre (**Figure 18**) ou en interaction avec un phosphopeptide TEL2 (**Figure 19**) a été résolue par diffraction des rayons X. Le domaine N-terminal est organisé en une alternance de feuillets β et d'hélices α , se finissant par deux hélices α . Cette organisation est assez rare parmi les structures connues capables d'interagir avec des protéines phosphorylées. La structure en complexe avec les 11 résidus du phosphopeptide TEL2 est sensiblement la même, à l'exception de l'hélice comprise entre les deux feuillets qui n'est plus visible dans le cristal, probablement en raison d'une perte de structure secondaire. De manière intéressante, TEL2 comporte 4 résidus très conservés DSDD, dont la sérine est phosphorylée (pS) par la protéine CK2. Seul le motif DpSDD est capable d'interagir efficacement avec PIH1D1, par l'intermédiaire de liaisons hydrogènes avec les lysines 57 et 64 de PIH1D1, qui sont très conservées (Horejsí *et al.*, 2014).

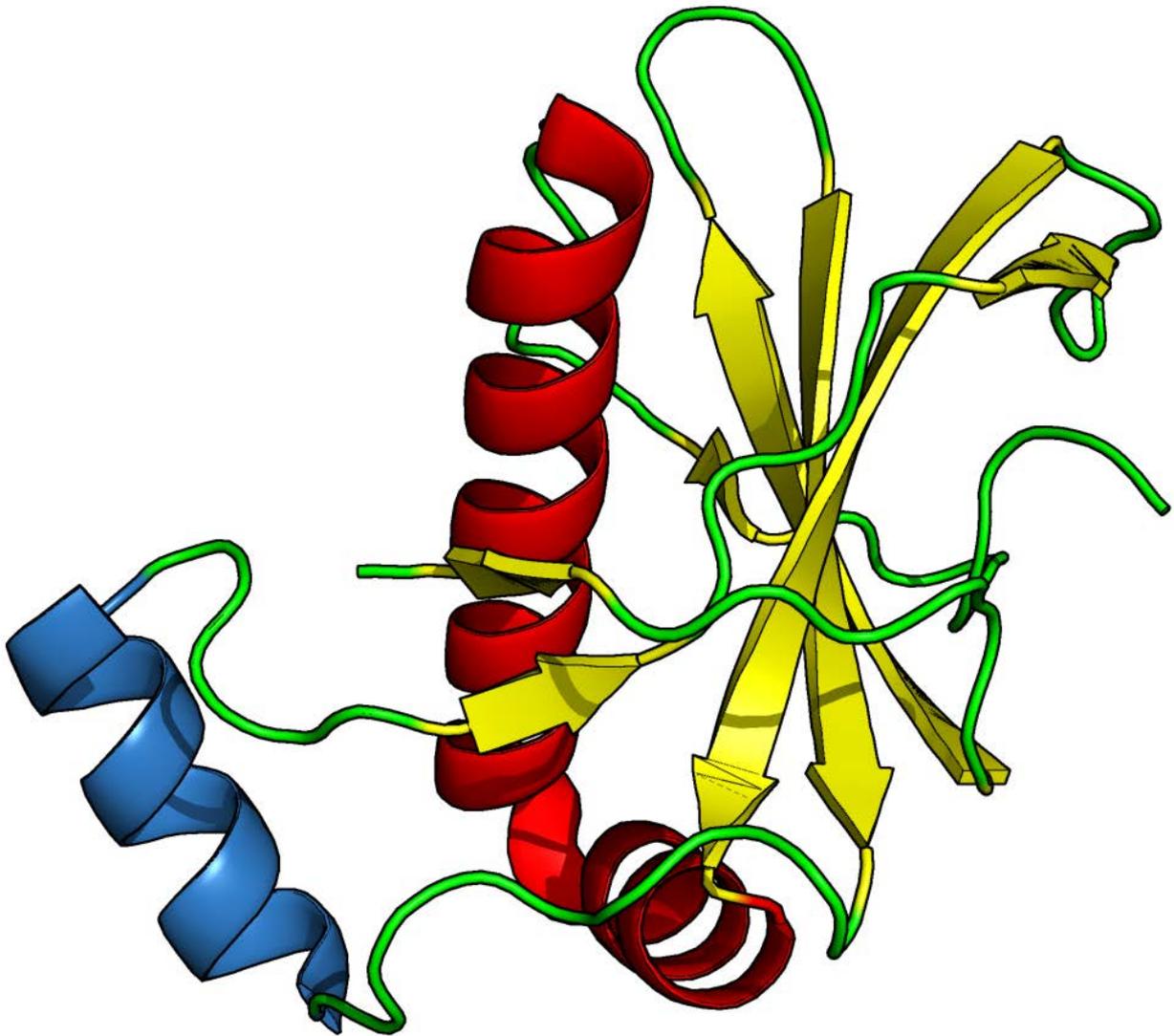


Figure 18: Structure 3D résolue par diffraction des rayons X du domaine N-terminal de la protéine humaine PIHD1 (code PDB : 4PSF)

Le domaine N-terminal de PIHD1 (PIHD1-N) forme un motif de type feuillet β - hélice α - feuillet β - hélices α . Lors de l'interaction de ce domaine avec TEL2, la structure du domaine n'est pas modifiée à l'exception d'une hélice (ici en bleu) qui perd sa structure secondaire (Horejsí *et al.*, 2014).

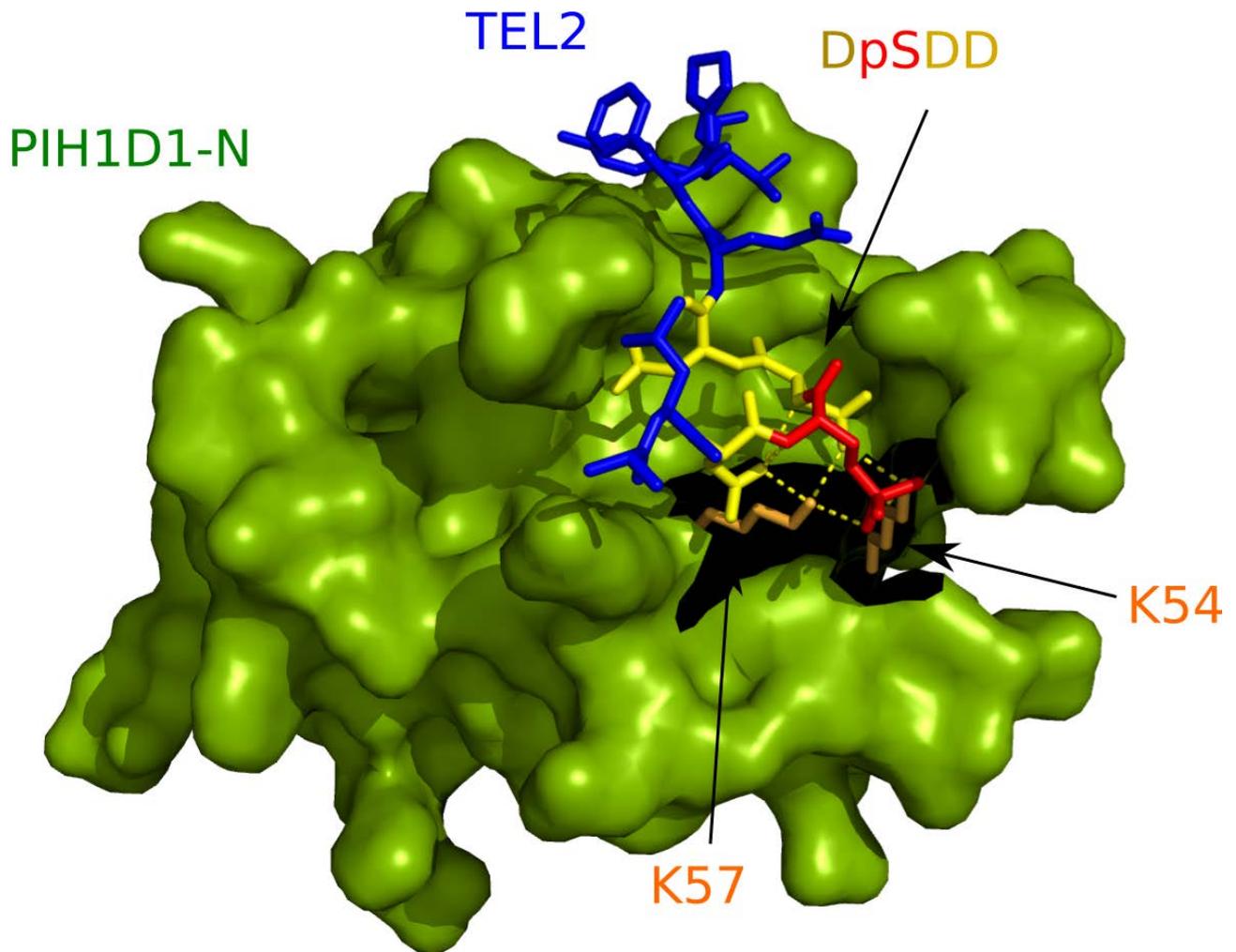


Figure 19: Vue de la surface de la structure 3D résolue par diffraction des rayons X du domaine N-terminal de la protéine humaine PIH1D1 en interaction avec un phosphopeptide TEL2 (code PDB : 4PSI).

Le domaine N-terminal de PIH1D1 interagit avec le phosphopeptide TEL2 par l'intermédiaire de liaisons hydrogènes entre les lysines K54 et K57 de PIH1D1 et le motif DpSDD de TEL2 (Horejsí *et al.*, 2014).

pS : Sérine phosphorylée.

Le domaine PIH1-C de la protéine de levure et humaine interagit respectivement avec les protéines Tah1 (Paci *et al.*, 2012 ; Quinternet *et al.*, 2015) et RPAP3 (Pal *et al.*, 2014). La structure de l'interaction entre les protéines de levure a été résolue par RMN dans notre laboratoire (Quinternet *et al.*, 2015) et montre que le domaine C-terminal de Pih1 est composé de 7 brins β . De manière remarquable, le domaine C-terminal de Tah1 ne se structure qu'une fois qu'il interagit avec le domaine PIH1D1-C pour former un brin β qui va alors s'intégrer dans le feuillet β de Pih1 (**Figure 20**).

Enfin, chez la levure, la séquence retrouvée entre les domaines N et C-terminaux contient également deux régions intrinsèquement désordonnées : les régions IDR1 et IDR2. Ces motifs sont connus pour intervenir dans des interactions protéines-protéines (Mészáros *et al.*, 2011) et il a été montré qu'elles étaient suffisantes pour permettre l'interaction entre la protéine Pih1 et les protéines Rvb1 et Rvb2 (Paci *et al.*, 2012).

d. La protéine Tah1/RPAP3

La protéine Tah1 a été identifiée comme l'un des partenaires de la protéine Hsp82 puis comme membre du complexe R2TP. Son homologue humaine correspond à la protéine RPAP3, elle aussi impliquée dans le complexe hR2TP. Toutefois, ces deux protéines sont de tailles très différentes (633 acides aminés pour RPAP3 contre 111 pour Tah1), bien qu'elle possède un domaine commun. Ceci suggère que les protéines RPAP3 et Tah1 partagent des fonctions communes mais que la protéine RPAP3 est probablement impliquée dans d'autres mécanismes chez l'homme, notamment grâce à ses domaines additionnels.

i. Données fonctionnelles

La protéine Tah1, pour « TPR-containing associated with Hsp82 » est une co-chaperonne d'Hsp82. Elle a également été décrite comme un partenaire de la protéine Pih1 et forme avec les protéines Rvb1 et Rvb2 le complexe R2TP (voir paragraphes précédents). Chez l'homme, la protéine RPAP3 est un homologue fonctionnel de la protéine Tah1. En effet, il a été montré qu'elle interagissait également avec la protéine HSP90, dont elle régule probablement l'activité comme son homologue chez la levure. RPAP3 interagit également avec les protéines PIH1D1, TIP48 et TIP49 au sein du complexe hR2TP.

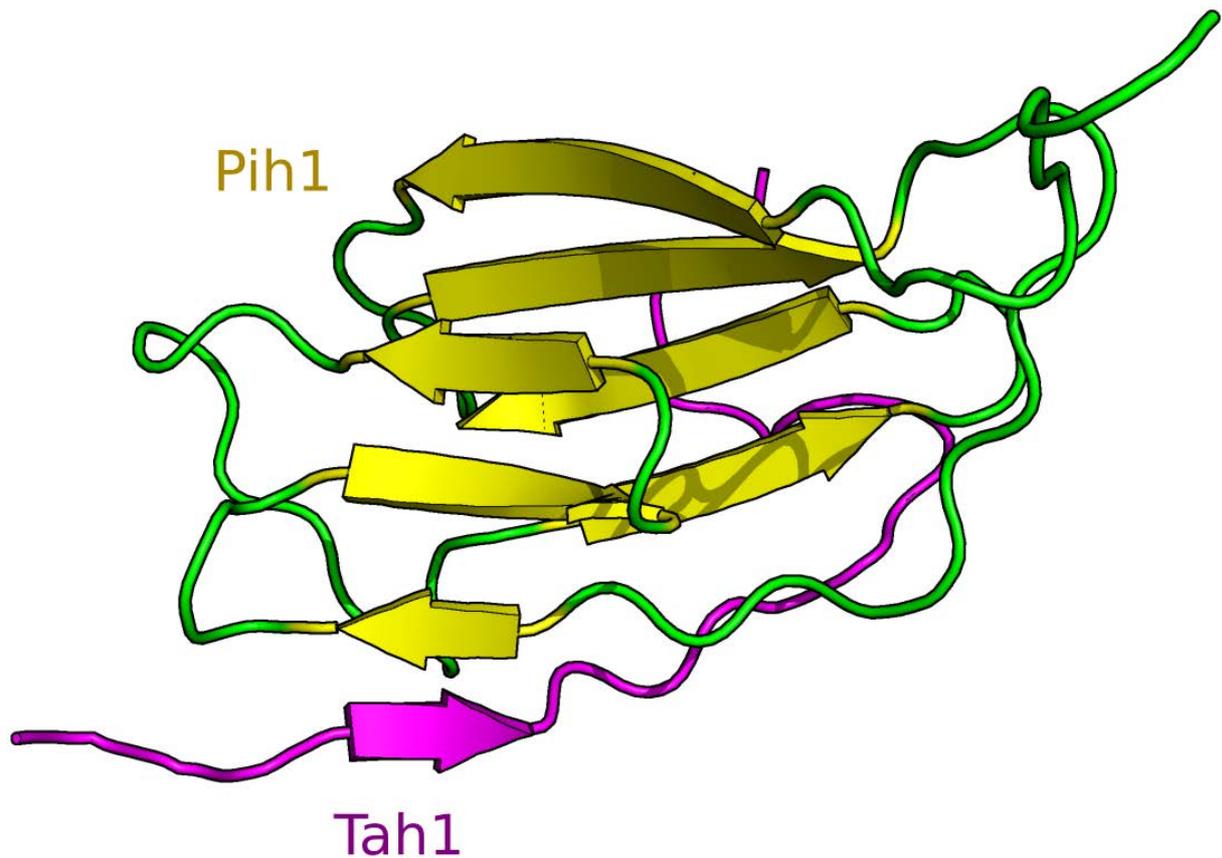


Figure 20: Structure RMN de l'interface d'interaction entre le domaine C-terminal de Tah1 (93-11, en violet) et le domaine N-terminal de Pih1 (24-111, en jaune) (Code PDB : 2MNJ).

Bien que le domaine C-terminal de Tah1 ne présente aucune structure secondaire lorsque la protéine est libre, l'interaction avec le domaine N-terminal de Pih1 entraîne la formation d'un brin β (en violet) qui va interagir et s'intégrer dans le feuillet β (en jaune) déjà présent au niveau du domaine de Pih1 (Quinternet *et al.*, 2015).

Toutefois, contrairement à la protéine Tah1, la protéine RPPA3 possède deux sites de fixation pour HSP90. Il est donc possible que RPAP3 interagisse directement avec les 2 monomères pour former un complexe tripartite HSP90-HSP90-RPAP3. Il n'est pas non plus exclu que l'un des deux sites interagisse avec le peptide IEEVD de la protéine HSP70, qui est l'un des partenaires essentiels d'HSP90. De plus, les domaines TPR sont connus pour être des domaines d'interaction protéines-protéines. Il est donc possible d'imaginer une configuration où l'un des monomères HSP90 serait en interaction avec l'un des deux TPR, tandis que le second domaine TPR serait en interaction avec une protéine partenaire de RPAP3 (par exemple une protéine cliente d'HSP90).

Initialement, la protéine RPAP3 avait été isolée lors d'un crible d'interaction avec l'ARN polymérase II, d'où son nom de « RNA Polymerase-Associated Protein 3 » (Jeronimo *et al.*, 2007). Ainsi, il a été montré que cette protéine formait un complexe avec les protéines RPAP1, RPAP2 et RPAP4/ABX1, toutes partenaires de l'ARN polymérase II. De plus, la protéine RPAP3 fut également identifiée lors d'expériences d'immunoprécipitation comme un partenaire de la protéine MONAD, une protéine comportant un domaine WD40 et impliquée dans l'induction de la mort cellulaire par apoptose induite par le TNF α ou la cycloheximide (Itsuki *et al.*, 2008). Bien que des interactions aient été retrouvées avec les protéines PIH1D1 et TIP48/Reptin (Jeronimo *et al.*, 2007 ; Ni *et al.*, 2009), l'implication du complexe hR2TP dans ces différents mécanismes n'est pas encore parfaitement comprise.

De façon remarquable, il existe trois isoformes de la protéine RPAP3 (**Figure 21**). Une isoforme longue (I), une seconde plus courte de 34 acides aminés, issue de l'exclusion de l'exon 12 (II) et enfin une dernière (III) dont le domaine N-terminal est absent (Ota *et al.*, 2004). Seules les deux premières isoformes ont été isolées et caractérisées *in vivo*. Il a ainsi été montré que la protéine PIH1D1 interagissait *in vivo* uniquement avec l'isoforme I, suggérant que seule cette isoforme se trouverait au sein du complexe hR2TP, tandis que l'isoforme II serait associée avec la protéine MONAD (Yoshida *et al.*, 2013).

ii. Données structurales

La protéine Tah1 de levure, dont la structure a été résolue par notre laboratoire (Back *et al.*, 2013), comporte 2 domaines. Un domaine N-terminal comprenant un domaine d'interaction protéine-protéine TPR et un domaine C-terminal non structuré.

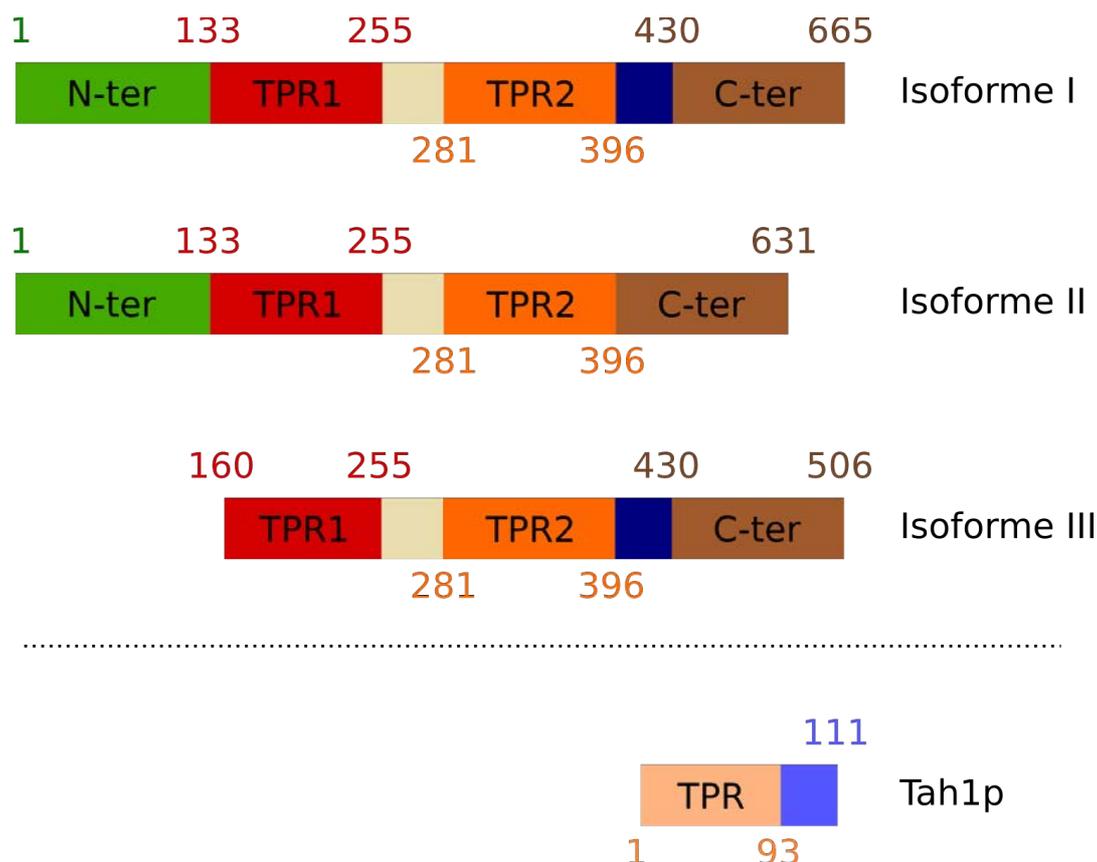


Figure 21: Organisation des domaines des trois isoformes de RPAP3, comparées à la protéine Tah1 de levure.

La protéine RPAP3 existe sous 3 isoformes. L'isoforme I est la plus longue et possède un domaine N-terminal (en vert), un premier domaine TPR (en rouge), un second domaine TPR (en orange), un domaine d'interaction avec PIH1D1 (en bleu) et un domaine C-terminal (en brun).

L'isoforme II ne possède pas le domaine d'interaction avec PIH1D1 suite à l'exclusion de l'exon 12 lors d'un événement d'épissage alternatif.

L'isoforme III qui n'a pour le moment été isolée que dans les banques de données ne possède plus le domaine N-terminal ni une partie du premier domaine TPR. En revanche, elle possède toujours le domaine d'interaction avec la protéine PIH1D1 et peut donc en théorie interagir avec PIH1D1.

Le domaine N-terminal est actuellement supposé être un domaine de dimérisation. Le domaine C-terminal serait quant à lui un domaine d'interaction protéine-protéine, tout comme les domaines TPR qui peuvent en plus interagir avec le motif M/IEEVD de la protéine HSP90/HSP70.

L'homologue chez la levure de RPAP3, la protéine Tah1, est beaucoup plus petite et ne contient que deux domaines. Un domaine TPR incomplet (2,5 motifs TPR contre 3 classiquement) et un domaine C-terminal. Le domaine C-terminal interagit avec Pih1 semble n'avoir qu'une homologie fonctionnelle avec le domaine d'interaction de RPAP3 avec PIH1D1.

Le domaine TPR est constitué de motifs TPR correspondant à une succession d'hélices α . Ces domaines TPR sont des domaines d'interaction protéine-protéine et sont retrouvés au sein des protéines qui modulent l'activité de la protéine Hsp82. Contrairement aux autres domaines TPR connus, la résolution de la structure de Tah1 a révélé une configuration particulière pour ce domaine TPR. En effet, alors que la plupart des domaines TPR sont constitués de 3 motifs TPR, la protéine Tah1 ne contient que 2,5 motifs TPR (**Figure 21 et 22**). Toutefois, comme d'autres protéines contenant des domaines TPR, la protéine Tah1 peut interagir avec le motif MEEVD de la protéine chaperonne Hsp82, par l'intermédiaire des acides aminés du domaine TPR formant une « dicarboxylate clamp » (**Figure 23**).

Chez l'homme, la protéine RPAP3 est constituée de 4 domaines. Un domaine N-terminal, deux domaines TPR et un domaine C-terminal. Contrairement à la protéine Tah1, les deux domaines TPR ont une constitution classique avec 3 motifs TPR entiers auxquels s'ajoute une hélice supplémentaire pour le domaine TPR 1. Comme pour la protéine Tah1, chaque domaine TPR peut interagir avec le motif MEEVD de la protéine HSP90 par l'intermédiaire d'une « dicarboxylate clamp » (**Figure 24 et 25**).

La structure du domaine C-terminal a été résolue par RMN au laboratoire mais n'a pas encore été publiée. Celle du domaine N-terminal n'a quant à elle pas encore été résolue et la fonction de ces deux domaines est encore inconnue.

e. Rôle du complexe (h)R2TP dans l'assemblage des snoRNP

Grâce aux différentes interactions décrites précédemment et à divers travaux supplémentaires (McKeegan *et al.*, 2007 et 2009 ; Boulon *et al.*, 2008) le rôle du complexe (h)R2TP dans l'assemblage des snoRNP à boîtes C/D commence à être compris. Tout d'abord, comme décrit pour d'autres particules RNP, il est souvent nécessaire de remodeler les protéines d'une particule pour obtenir un complexe RNP stable et fonctionnel (McKeegan *et al.*, 2007). Concernant les snoRNP, ce rôle a tout d'abord été attribué à la protéine chaperonne Hsp82/HSP90, avec pour le complexe (h)R2TP uniquement un rôle de modulateur, de par l'interaction entre RPAP3 et HSP90 (Tah1/Hsp82). Ce modèle ne prend toutefois pas en compte l'impossibilité d'isoler des complexes stables entre le (h)R2TP et HSP90 (Zhao *et al.*, 2008).

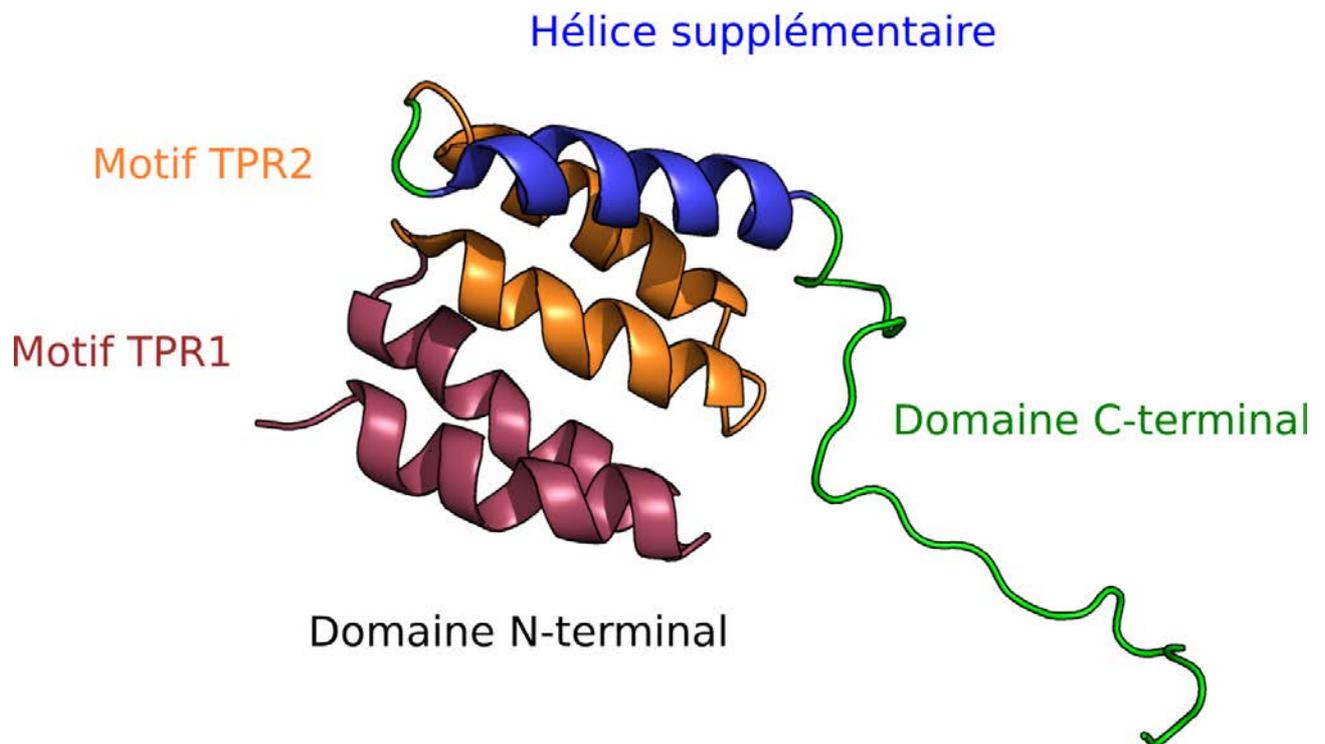


Figure 22: Structure 3D de la protéine Tah1 entière résolue par RMN (code PDB : 2LSU).

La protéine Tah1p comporte 2 domaines, un domaine TPR à son extrémité N-terminale et un domaine C-terminal.

Le domaine TPR est constitué de 2,5 motifs TPR, soit 2 fois le motif hélice-boucle-hélice plus une hélice supplémentaire (Back *et al.*, 2013).

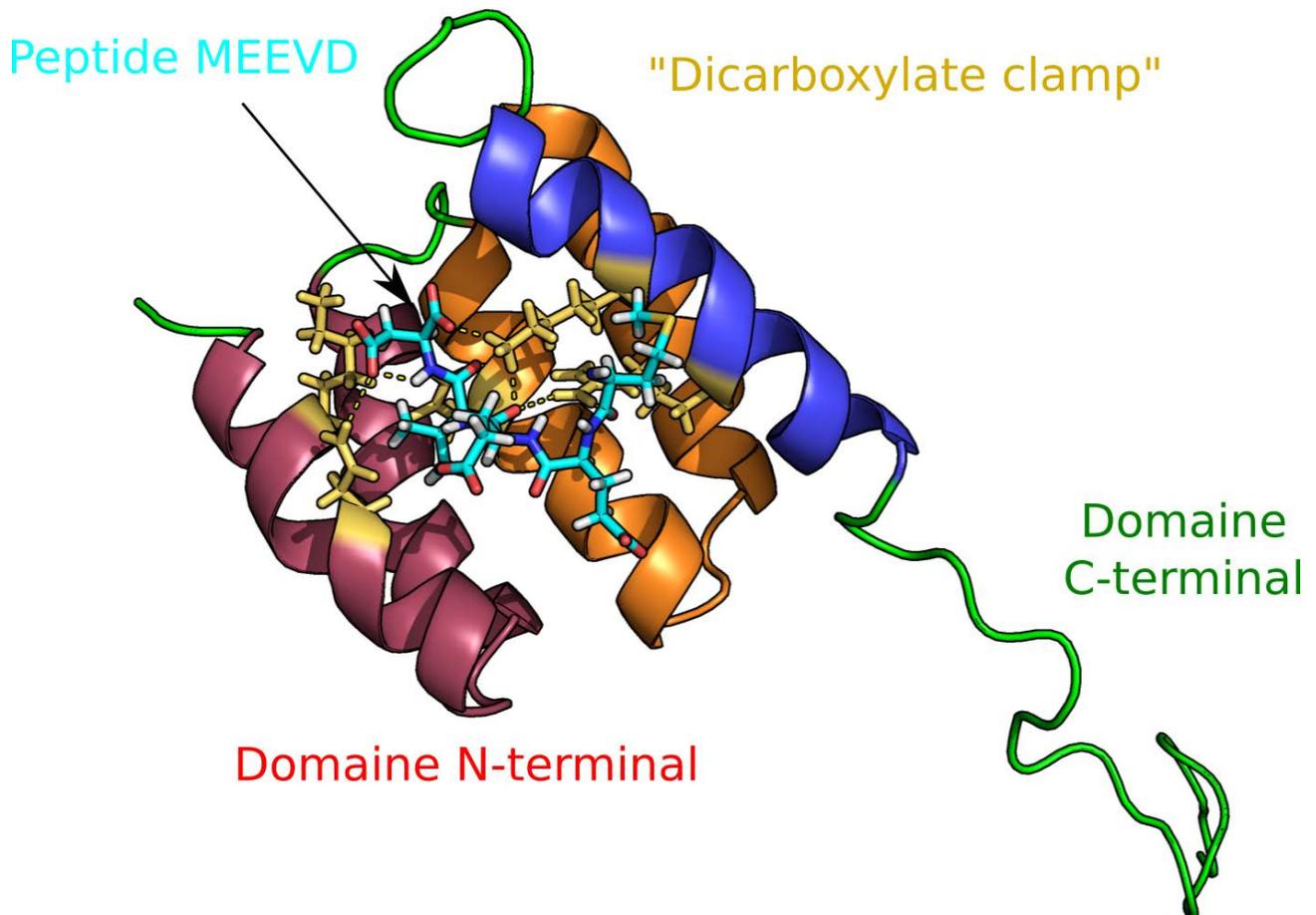


Figure 23: Structure 3D de la protéine Tah1 entière en interaction avec le peptide MEEVD d’Hsp82 résolue par RMN (code PDB : 2LSV).

Le peptide MEEVD de la protéine hsp82 interagit avec le domaine TPR de la protéine Tah1 par l’intermédiaire de résidus (K8-N12-N43-K79-R83) qui forment une pince appelée « dicarboxylate clamp ». (Back *et al.*, 2013).

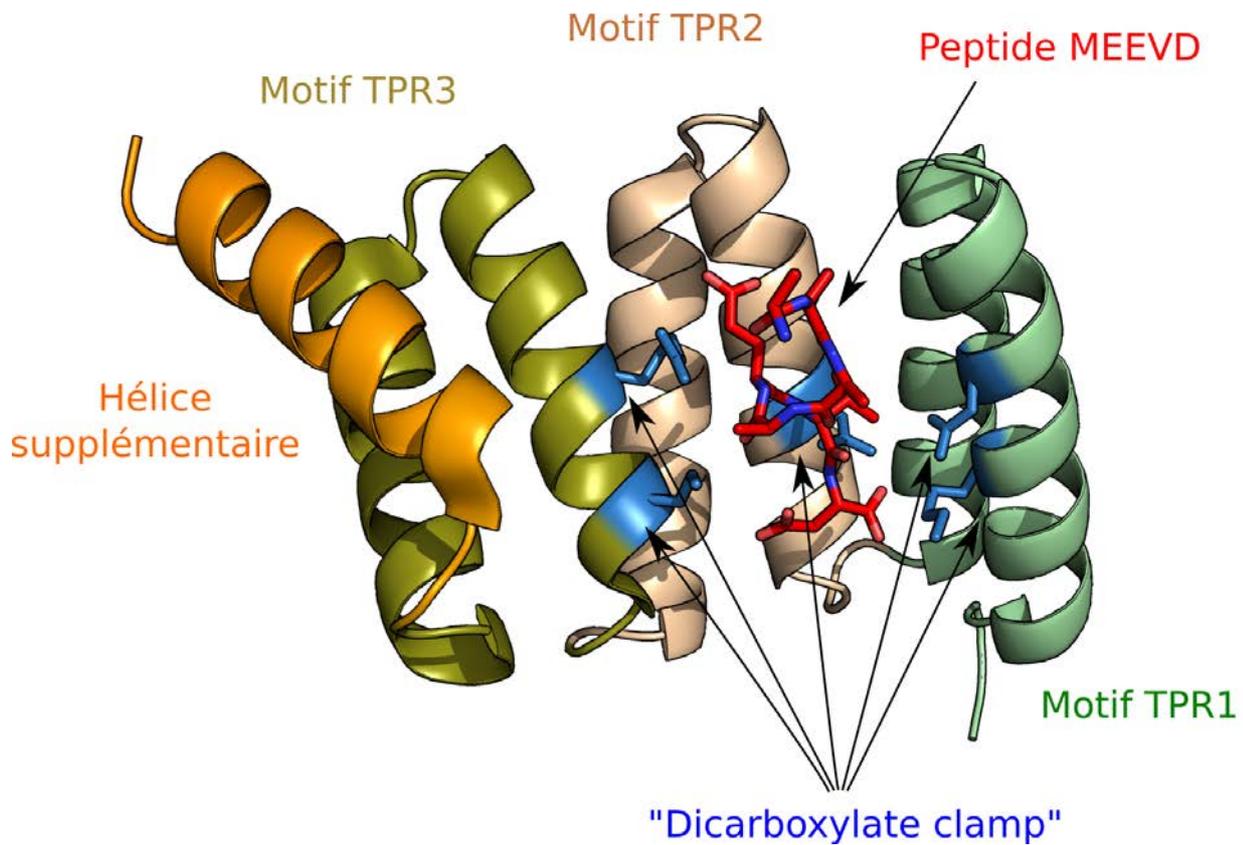


Figure 24: Structure 3D du domaine TPR n°1 de la protéine RPAP3 en interaction avec le peptide MEEVD de la protéine HSP90 résolue par diffraction des rayons X (code PDB : 4CGV).

Contrairement au domaine TPR de son homologue chez la levure Tah1, le domaine TPR de RPAP3 est constitué de 3 motifs TPR entiers et se termine par une hélice supplémentaire. Comme pour Tah1, le peptide MEEVD de la protéine HSP90 interagit avec ce domaine par l'intermédiaire d'interactions avec les résidus formant la « carboxylate clamp » (Pal *et al.*, 2014).

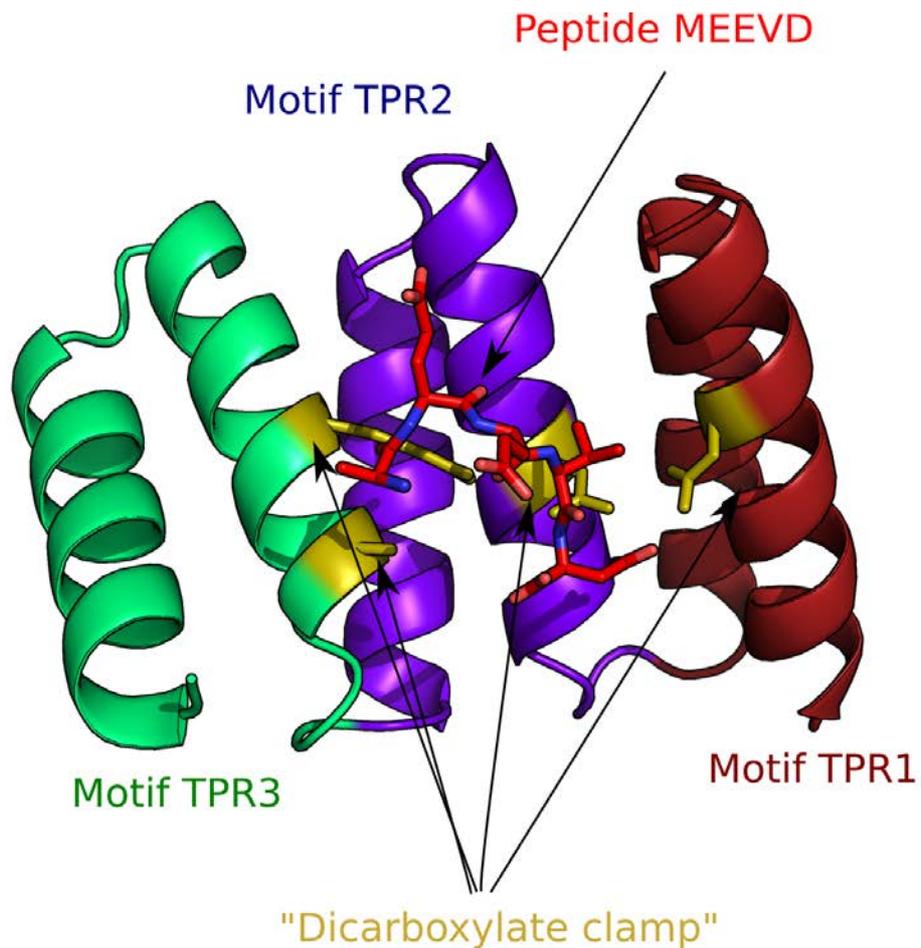


Figure 25: Structure 3D du domaine TPR n°2 de la protéine RPAP3 en interaction avec le peptide MEEVD de la protéine HSP90 résolue par diffraction des rayons X (code PDB : 4CGW).

Ce second domaine TPR présente la même structure que le premier domaine TPR à l'exception de l'hélice supplémentaire qui est absente. Comme pour Tah1 et le premier domaine TPR, le peptide MEEVD de la protéine HSP90 interagit avec ce domaine par l'intermédiaire d'interaction avec les résidus formant la « dicarboxylate clamp » (Pal *et al.*, 2014).

En outre, Hsp82/HSP90 semble absente du noyau en conditions normales de croissance cellulaire. Dans le noyau, ce rôle de remodelage semble plutôt revenir aux ATPase AAA+ TIP48-TIP49 au sein du complexe hR2TP. En effet, les protéines TIP48 et TIP49 peuvent utiliser l'ATP pour modifier la conformation d'une protéine. En outre le complexe (h)R2TP a pu être isolé à la fois dans le noyau (McKeegan *et al.*, 2009) et dans le cytoplasme (Boulon *et al.*, 2008), contrairement à la protéine Hsp82/HSP90. La localisation nucléaire du complexe (h)R2TP étant probablement due à l'interaction de Pih1/PIH1D1 avec la protéine Nop58/NOP58. En effet, les protéines Nop56/NOP56 et Nop58/NOP58 sont actuellement les seules protéines cœur des snoRNP à boîtes C/D dans lesquelles un signal de localisation nucléaire a été décrit (Pradet-Balade *et al.*, 2011).

Toutefois, le complexe (h)R2TP semble plutôt être le regroupement de 2 sous-complexes stables, le dodécamère/héxamère TIP48/TIP49 (ou Rvb2/Rvb1) d'une part et le complexe RPAP3/PIH1D1 (ou Tah1/Pih1) d'autre part, qui peuvent avoir d'autres rôles séparément. Si le premier complexe semble pouvoir s'assembler spontanément, ce n'est pas le cas du second qui nécessiterait l'intervention de la protéine chaperonne Hsp82/HSP90 pour éviter l'agrégation de la protéine PIH1D1.

Un modèle possible pour la constitution du complexe (h)R2TP pourrait donc consister tout d'abord en la stabilisation dans le cytoplasme de la protéine Pih1/PIH1D1 via son interaction avec le complexe RPAP3/HSP90 (Tah1/Hsp82) puis au recrutement du dodécamère TIP48/TIP49. Une fois constitué, le complexe interagirait, généralement par l'intermédiaire de Pih1/PIH1D1, avec différents partenaires supplémentaires (Nop56/NOP56, TEL2, SBP2, RPAP1/2/4, ...) (**Figure 26**).

Concernant l'assemblage des snoRNP à boîtes C/D d'eucaryotes, si l'interaction entre PIH1D1 et NOP58 semble importante, il a également été montré par GST pull-down que les protéines TIP48 et TIP49 interagissaient directement avec la Fibrillarine, NOP56, NOP58 (McKeegan *et al.*, 2009), PIH1D1 et SNU13 (McKeegan *et al.*, 2007 et 2009). Toutefois, il semble que le complexe (h)R2TP seul ne soit pas suffisant pour obtenir des particules stables et fonctionnelles. Des travaux supplémentaires ont ainsi mis en évidence de nouvelles interactions entre les protéines cœur des snoRNP à boîtes C/D, le complexe (h)R2TP et des facteurs d'assemblage potentiels, dont les protéines Rsa1/NUFIP, Hit1/ZNHIT3 et Bcd1/BCD1.

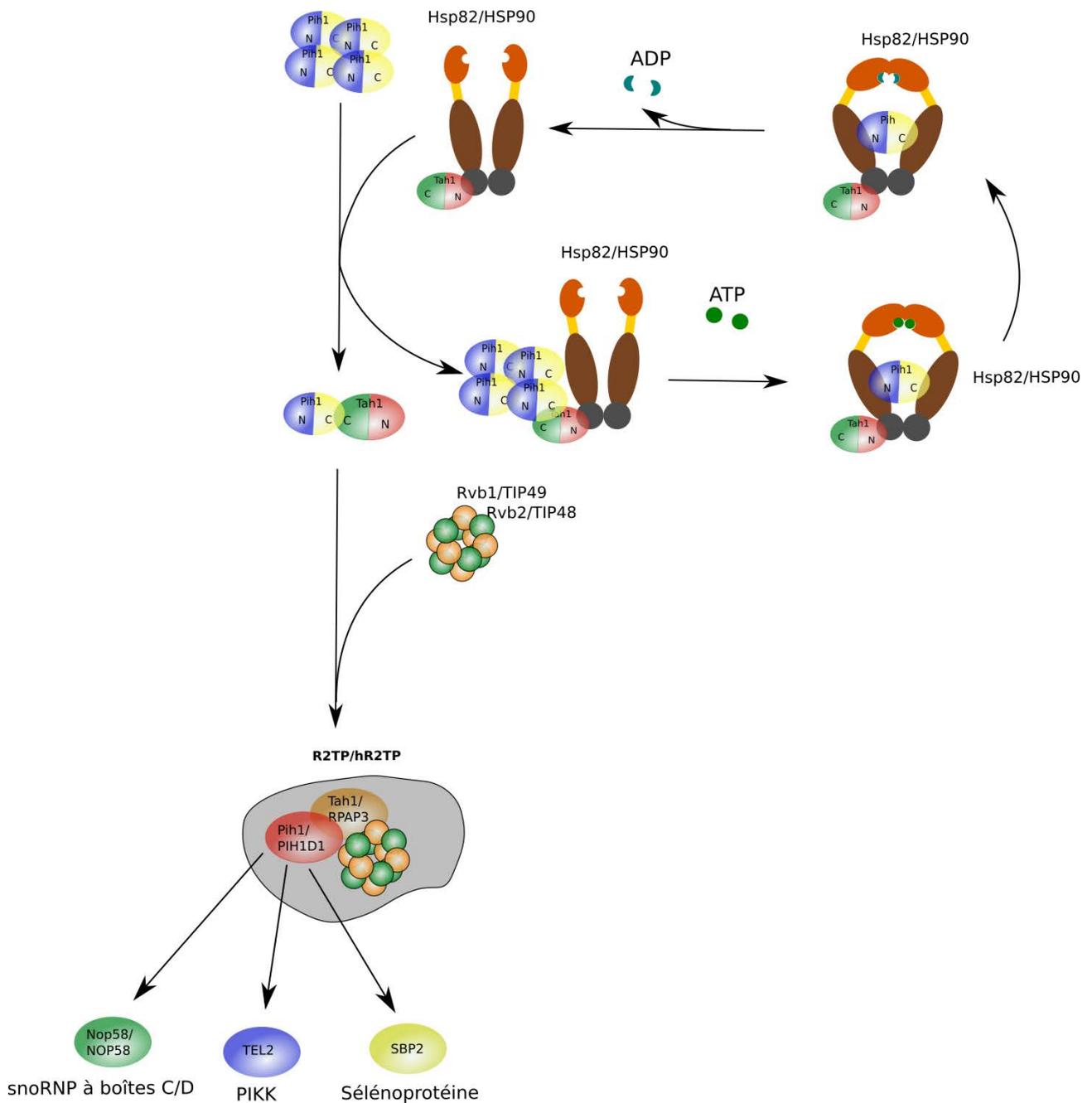


Figure 26: Modèle d'assemblage du (h)R2TP (adapté de Zhao *et al.*, 2008).

La protéine Pih1/PIH1D1 non repliée est prise en charge par la protéine chaperonne Hsp82/HSP90 par l'intermédiaire de la protéine Tah1/RPAP3. Une fois la protéine stabilisée, le complexe Tah1/Pih1 (RPAP3/PIH1D1) peut interagir avec le dodécamère Rvb1/Rvb2 (TIP49/TIP48) pour former le complexe R2TP/hR2TP. Suivant le partenaire de Pih1/PIH1D1, le complexe R2TP/hR2TP pourra intervenir dans différents mécanismes biologiques.

4. Le complexe NUFIP/ZNHIT3 (Rsa1/Hit1)

a. La protéine Rsa1/NUFIP

La protéine de levure Rsa1 (« Ribosome assembly 1 ») fut initialement identifiée lors de cribles synthétiques létaux contre la protéine Ddp6, une ARN hélicase impliquée dans la biogenèse des ribosomes (Kressler *et al.*, 1999). Des expériences supplémentaires d'immunoprécipitation, de double- et de triple-hybride ont permis ensuite de montrer que Rsa1 s'associait d'une part avec les snoARN à boîtes C/D (dont le snoARN U3) et d'autre part avec la protéine Snu13, et ont démontré l'existence d'un complexe ternaire avec les trois partenaires. Rsa1 interagit en outre avec la protéine Nop58 (Boulon *et al.*, 2008).

Chez l'homme, la protéine NUFIP (« NUClear FMRP Interacting Protein ») a été isolée suite à un crible double-hybride réalisé contre la protéine SNU13. Initialement, cette protéine était connue pour interagir avec l'isoforme nucléaire de la protéine FMRP (« Fragile X Mental Retardation Protein), une protéine impliquée dans le syndrome de l'X fragile (Bardoni *et al.*, 1999). Bien que partageant très peu d'identité (voir ci-après), les protéines NUFIP et Rsa1 sont toutes les deux nécessaires à la biogenèse des snoRNP. En effet, des expériences de siRNA dirigés contre elles ont montré une diminution significative du taux des snoARN U3 et U14 (Boulon *et al.*, 2008).

De plus, des expériences d'immunoprécipitations de la protéine NUFIP réalisées sur des extraits nucléaires de cellules HeLa ont permis d'observer son association avec des snoRNP immatures, notamment avec les pré-snoRNP U3 et U8 (McKeegan *et al.*, 2007), confirmant son implication dans la biogenèse des snoRNP à boîtes C/D. Enfin, des expériences de GST pull-down ont montré que NUFIP interagissait également avec des protéines du (h)R2TP, notamment PIH1D1 (Boulon *et al.*, 2008), TIP48 et TIP49 mais aussi avec l'ensemble des protéines cœur des snoRNP à boîtes C/D (McKeegan *et al.*, 2007).

La protéine Rsa1/NUFIP semble donc être un autre facteur impliqué dans l'assemblage des snoRNP à boîtes C/D, et permettrait, par l'intermédiaire de ses nombreux partenaires, le recrutement du complexe (h)R2TP au niveau des pré-snoRNP. Toutefois, comme pour le complexe hR2TP, la protéine NUFIP interagit également avec les protéines NHP2 et SBP2 *in vitro* et *in vivo* (Boulon *et al.*, 2008). Elle semble donc aussi intervenir dans la maturation d'autres particules RNP.

Des expériences d'immunoprécipitations sur des extraits cellulaires transfectées par différentes protéines ont depuis permis d'affiner un peu plus le rôle de la protéine NUFIP lors de l'assemblage des snoRNP à boîtes C/D (Bizarro *et al.*, 2014). Ces expériences ont été réalisées sur des extraits dont la préparation permet de séparer d'une part les complexes protéiques présents dans le cytoplasme et dans les fractions solubles du nucléoplasme (« Highly Extractable fraction », fraction HE) et d'autre part les complexes présents dans le reste du nucléoplasme (« Less Extractable fraction », fraction LE), notamment dans le nucléole (Boulon *et al.*, 2010). Ces expériences ont été couplées à du SILAC (« Stable Isotope Labeling with Amino acids in Cell culture ») et l'analyse des complexes immunoprécipités a été finalement réalisée par spectrométrie de masse.

Ces travaux ont ainsi permis d'identifier dans la fraction HE une faible association entre la protéine NUFIP et les protéines cœur Fibrillarine, SNU13 et NOP58, et une interaction plus forte avec une nouvelle protéine, ZNHIT3. Au niveau de la fraction LE, le taux des protéines SNU13 et NOP58 associées à la protéine NUFIP étaient respectivement multipliés par un facteur 3 ou 8 par rapport à la fraction HE, alors que celles de la protéine ZNHIT3 y étaient divisées par deux (Bizarro *et al.*, 2014). Ce résultat semble indiquer que la protéine ZNHIT3 doit être libérée pour permettre l'association des protéines cœur avec la protéine NUFIP au niveau du nucléole (fraction LE) (Bizarro *et al.*, 2014). De manière intéressante, NUFIP a également été retrouvée associée avec divers facteurs d'épissage (PRPF8, PRPF19, SNRPD1 et SNRNP200) au niveau de la fraction LE. La majorité des snoARN chez les mammifères étant introniques, ces associations suggèrent que NUFIP seraient, au niveau de la fraction LE, associée avec les snoARN. Ensemble, ces résultats suggèrent qu'une fois dans le nucléole, un complexe protéique composé des protéines cœur s'associerait avec le snoARN naissant et la protéine NUFIP, entraînant alors le départ de ZNHIT3 (Bizarro *et al.*, 2014).

Toutefois, alors que de précédents travaux avaient montré l'existence d'une interaction entre la protéine NUFIP et les protéines TIP48 et TIP49 (Boulon *et al.*, 2008), celle-ci n'a pas été retrouvée par ces expériences d'immunoprécipitation (Bizarro *et al.*, 2014). L'une des raisons évoquée par les auteurs serait que ces associations soient trop labiles pour être observées avec cette approche (Bizarro *et al.*, 2014). Des expériences de GST pull-down réalisées sur des extraits totaux de cellules HeLa ont permis aux auteurs de confirmer que ces associations n'étaient retrouvées que lorsque les échantillons n'étaient pas fractionnés (Bizarro *et al.*, 2014). En revanche, l'interaction

entre les protéines PIH1D1 et NUFIP n'a pas été retrouvée. Ces observations semblent indiquer que le sous-complexe RPAP3-PIH1D1 se sépare du sous-complexe TIP48-TIP49 très tôt après le recrutement du complexe hR2TP (Bizarro *et al.*, 2014).

b. La protéine Hit1/ZNHIT3 (TRIP3)

La protéine Hit1 (« High temperature Growth 1 ») a été initialement identifiée chez la levure *S. cerevisiae* lors de la recherche de facteurs protéiques dont la mutation entraîne un défaut de croissance à haute température (Kawakami *et al.*, 1992). Son implication dans l'assemblage des snoRNP à boîtes C/D a été mis en avant suite à des expériences de « TAP-Tag » réalisées contre la protéine Rsa1 qui ont montré une association entre ces deux protéines (Rothé *et al.*, 2014b). Des expériences de co-expression chez *E. coli* ont ensuite permis de montrer que cette interaction était directe et compatible avec l'interaction précédemment observée entre les protéines Rsa1 et Snu13 (Boulon *et al.*, 2008 ; Rothé *et al.*, 2014b). De plus, il a été montré que la protéine Hit1 était indispensable à la stabilité de la protéine Rsa1 (Rothé *et al.*, 2014b) et donc nécessaire à la biogenèse des snoRNP. Enfin, des expériences de synthétiques létaux ont montré l'existence d'un lien génétique entre les gènes *HIT1*, *RS1* et *PIH1* (Rothé *et al.*, 2014b).

La protéine TRIP3 (« Thyroid hormone Receptor Interacting Protein 3 »), l'homologue humaine de la protéine Hit1, a été initialement identifiée comme une partenaire du récepteur nucléaire hépatocytaire HNF-4 α . Ce récepteur est impliqué dans la régulation de l'expression de gènes liés au métabolisme du glucose et des mutations dans son gène ont été associées à une forme sévère de diabète (Iwahashi *et al.*, 2002). TRIP3 fut renommée ZNHIT3, pour « ZiNc finger HI-TYPE containing 3 », suite à l'identification de son interaction avec la protéine NUFIP, et à son homologie avec la protéine de levure Hit1 (Bizarro *et al.*, 2014 ; Rothé *et al.*, 2014b). L'interaction directe entre la protéine NUFIP et ZNHIT3 a été confirmée par GST pull-down (Bizarro *et al.*, 2014) et comme pour la protéine Hit1, il a été montré par ARN interférence qu'elle était essentielle à la stabilité de la protéine NUFIP (Rothé *et al.*, 2014b).

Bizarro *et al.* ont montré que la protéine ZNHIT3 s'associait avec la protéine NUFIP dans la fraction LE, et avec les protéines NUFIP, TIP48, TIP49, SNU13, NOP58 et hBCD1 dans la fraction HE. De manière intéressante, l'association entre ces protéines n'est pas affectée par un traitement à la RNase (Bizarro *et al.*, 2014). Il semble donc que l'ARN ne soit pas indispensable à la stabilité de

ces complexes. Toutefois, conformément aux précédents travaux qui relient la protéine SNU13 au snoARN, il est possible que le complexe qui contient les protéines NUFIP, ZNHIT3 et SNU13 contiennent aussi le snoARN (Rothé *et al.*, 2014b). Il semblerait néanmoins que la protéine ZNHIT3 ne s'associe de manière stable qu'avec des snoARN incapables de s'assembler correctement, contrairement à la protéine NUFIP (Bizarro *et al.*, 2014). L'hypothèse actuelle serait que la protéine ZNHIT3 soit rapidement relarguée une fois le snoARN, s'il est intègre, associé aux protéines cœur et aux protéines du hR2TP (Bizarro *et al.*, 2014 ; Rothé *et al.*, 2014a ; Rothé *et al.*, 2014b).

c. Données structurales

D'autres travaux ont montré que les interactions directes entre les protéines NUFIP et SNU13 et entre les protéines NUFIP et ZNHIT3 n'étaient pas mutuellement exclusives et un complexe ZNHIT3-NUFIP-SNU13 a pu être purifié *in vitro*, chez la levure (Rothé *et al.*, 2014b) et chez l'homme (Quinternet *et al.*, 2016). De plus, il a été montré chez la levure que ce complexe pouvait également se lier aux snoARN à boîtes C/D et à la protéine Nop58 (Rothé *et al.*, 2014b). Bien qu'homologues fonctionnels, les protéines NUFIP et Rsa1 ne présentent pratiquement aucune homologie de séquence mis à part un domaine de 32 résidus au niveau de l'extrémité N-terminale appelé motif PEP (hPEP chez l'Homme, yPEP chez la levure) (Boulon *et al.*, 2008). Il a été montré chez la levure et chez l'homme que cette séquence était suffisante pour permettre l'interaction avec la protéine Snu13/SNU13 (Boulon *et al.*, 2008 ; Rothé *et al.*, 2014a ; Rothé *et al.*, 2014b ; Bizzaro *et al.*, 2014 ; Quinternet *et al.*, 2016). Enfin, la protéine Rsa1 contient également au niveau de son extrémité C-terminale un domaine de localisation nucléaire.

La structure de l'interface d'interaction entre la protéine Snu13 et le domaine yPEP a été résolue par diffraction des rayons X (Bizarro *et al.*, 2014). Cette structure (**Figure 27**) montre que le domaine yPEP contient au niveau de son extrémité N-terminale une longue hélice α de 17 acides aminés reliée par une boucle à une petite hélice α de 5 résidus. Ce domaine interagit notamment avec la protéine Snu13 par le biais d'interactions électrostatiques avec les hélices 3 et 5 de Snu13, mais également par la formation d'une poche hydrophobe autour d'un résidu tryptophane du domaine yPEP par certains résidus de l'hélice 5 (Bizarro *et al.*, 2014).

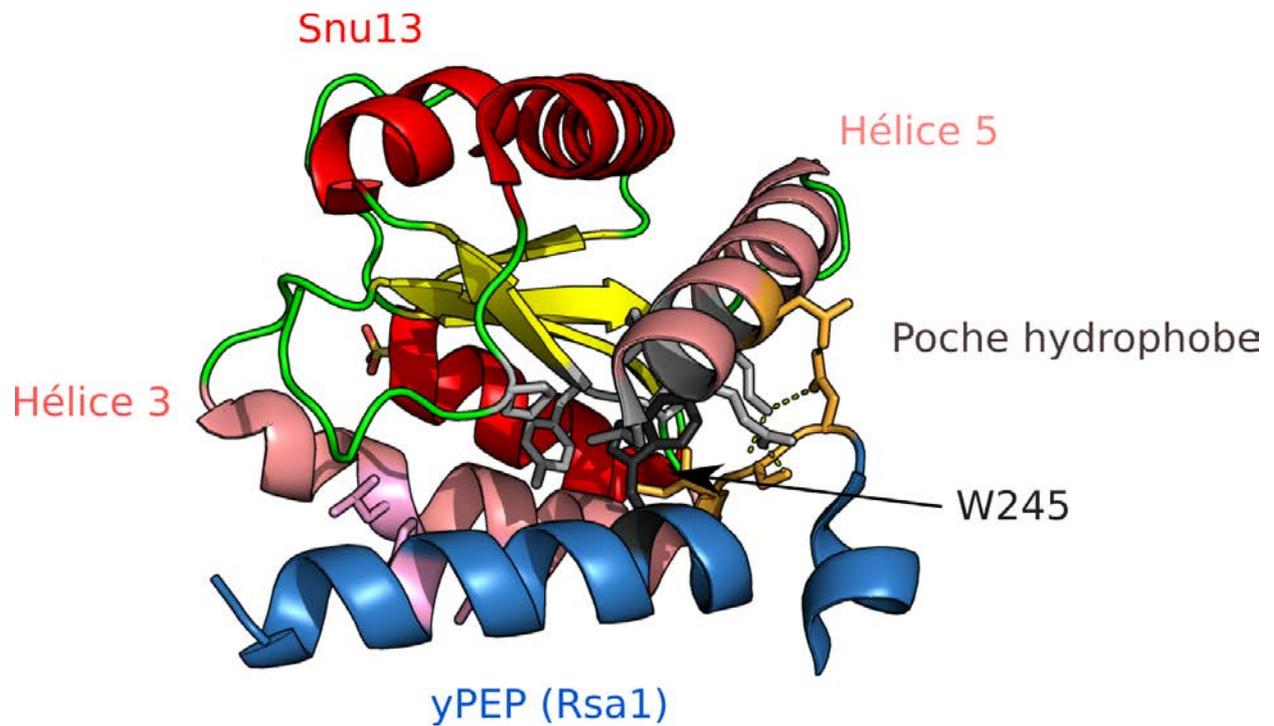


Figure 27: Structure 3D de l'interface entre la protéine de levure Snu13 (4-125) et le domaine yPEP de la protéine Rsa1 (239-265). (code PDB : 4NUT).

Le domaine yPEP de Rsa1 interagit avec la protéine Snu13 par l'intermédiaire de liaisons hydrogène avec l'hélice 5, d'interactions hydrophobes avec l'hélice 3 et par la formation d'une poche hydrophobe entre le résidu tryptophane 245 de Rsa1 et la protéine Snu13.

Bien que la structure de l'interface entre les protéines humaines NUFIP et SNU13 n'ait pas été résolue, l'analyse des séquences de ces deux protéines montre une conservation de cette poche hydrophobe (Bizarro *et al.*, 2014) et tout indique que ce type d'interaction soit conservé chez l'homme (Quinternet *et al.*, 2016).

Les protéines Rsa1 et Hit1 interagissent respectivement par l'intermédiaire de leur domaine C et N-terminaux. La structure de cette interaction a été résolue par RMN (**Figure 28**) par notre laboratoire (Rothé *et al.*, 2014b). Le domaine N-terminal de la protéine Hit1 comprend cinq hélices α , alors que le domaine C-terminal de la protéine Rsa1 est formé de 2 hélices α stabilisées par la protéine Hit1. De manière intéressante, aucune structure similaire n'existe dans les bases de données (Rothé *et al.*, 2014b).

Si les protéines NUFIP et ZNHIT3 n'ont que peu d'homologies de séquence avec leurs homologues fonctionnels respectifs chez la levure, notre laboratoire a récemment montré qu'elles interagissaient de la même manière (Quinternet *et al.*, 2016). Ainsi, la protéine NUFIP interagit par l'intermédiaire de son domaine C-terminal (462-495) avec le domaine N-terminal de ZNHIT3 (85-155). La structure de ce complexe résolue par RMN par notre laboratoire (**Figure 29**) montre que le domaine N-terminal de ZNHIT3 forme, comme Hit1, 5 hélices α qui vont stabiliser l'hélice α du domaine N-terminal de NUFIP. De légères différences ont toutefois pu être observées, indiquant que ce complexe serait moins stable chez l'homme. Ainsi, l'absence d'une seconde hélice au niveau du domaine C-terminal de la protéine NUFIP par rapport à la Rsa1 empêcherait le complexe d'être aussi stable que chez la levure (Quinternet *et al.*, 2016).

5. La protéine Bcd1/BCD1

a. Données fonctionnelles

L'analyse du taux de snoARN issus de différentes souches de levures a révélé que la répression de l'expression de la protéine Bcd1 (« Box C/D snoRNA protein ») entraînait une accumulation des snoARN à boîtes C/D, suggérant que cette protéine soit nécessaire à l'assemblage des snoRNP à boîtes C/D (Peng *et al.*, 2003 ; Hiley *et al.*, 2005). Chez l'homme, une protéine homologue, appelée BCD1, a été identifiée et des expériences d'ARN interférence ont montré une réduction significative du taux des snoARN U3, U8 et U14 ainsi qu'un changement de localisation du snoARN U3 lorsque l'ARNm de BCD1 était ciblé (McKeegan *et al.*, 2007).

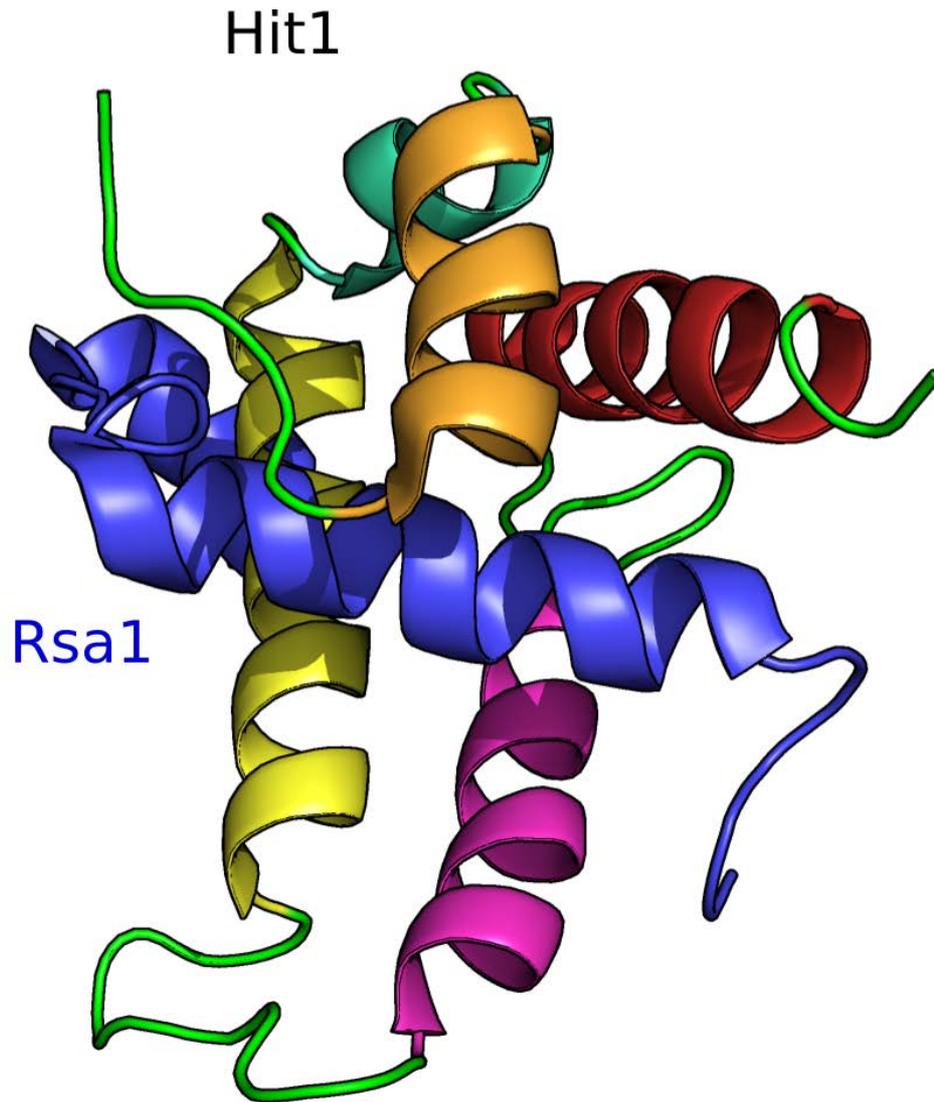


Figure 28: Structure 3D de l'interaction entre la protéine Rsa1 (317-352) et Hit1 (70-164) résolue par RMN (code PBD : 2MJF).

Les domaines N et C-terminaux des protéines Hit1 et Rsa1 forment respectivement cinq et deux hélices α . Les hélices de la protéine Rsa1 sont stabilisées par la protéine Hit1.

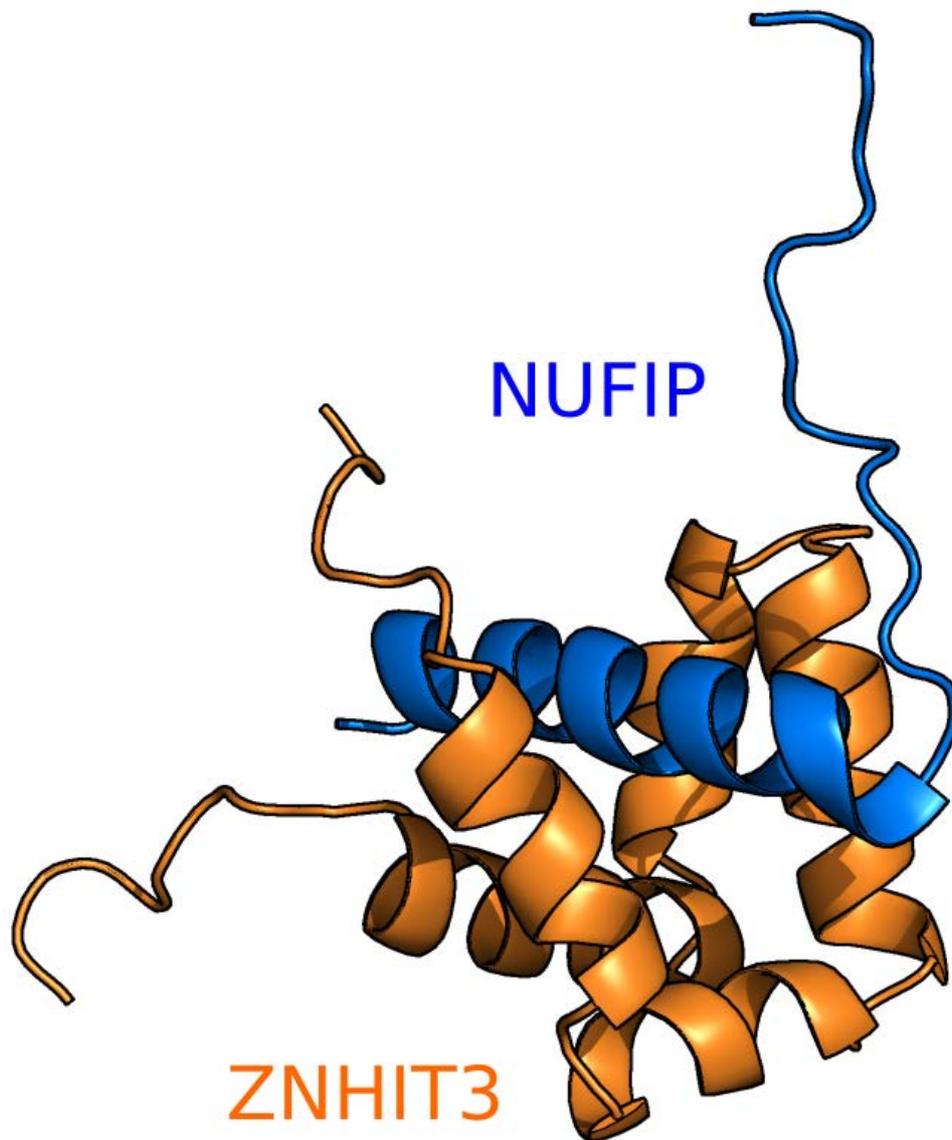


Figure 29: Structure 3D de l'interaction entre les protéines NUFIP (462-495) et ZNHIT3 (81-155) résolue par RMN dans notre laboratoire. (code PDB : 5L85).

Les domaines N et C-terminaux des protéines ZNHIT3 et NUFIP forment des hélices α (5 pour ZNHIT3 et 1 pour NUFIP). De la même manière que pour Rsa1/Hit1, l'hélice de la protéine NUFIP est stabilisée par ZNHIT3 (Quinternet *et al.*, 2016)

De plus, il a été montré que la protéine BCD1 interagissait directement avec d'autres facteurs d'assemblage, notamment les protéines NUFIP, TIP48 et TIP49, ainsi qu'avec la Fibrillarine (McKeegan *et al.*, 2007 et 2009).

Enfin Bizarro *et al.* ont montré par des immunoprécipitations réalisées sur la protéine BCD1 qu'elle s'associait avec les protéines TIP48, TIP49, NUFIP et ZNHIT3 dans la fraction HE, et avec les protéines TIP49 et SNU13 dans la fraction LE (Bizarro *et al.*, 2014). L'ensemble de ces données indique que la protéine BCD1, au même titre que les protéines NUFIP et ZNHIT3, serait un facteur majeur de la biogenèse des snoRNP à boîtes C/D.

b. Données structurales

La protéine Bcd1/BCD1 appartient à la même famille que les protéines Hit1/ZNHIT3, appelée Zf-HIT. Ces protéines sont caractérisées par la présence à leur extrémité N-terminale d'un domaine ZHD (« Zf-HIT Domain ») contenant un doigt de zinc. Au sein de cette famille, ce doigt à zinc se structure à l'aide de deux motifs CCCC et CCHC qui vont coordonner chacun un atome de zinc (**Figure 30**). Pour BCD1 ou ZNHIT3, ainsi que leurs homologues chez la levure, il a été montré que ce domaine n'était pas impliqué dans les interactions avec les protéines Rsa1/NUFIP et Snu13/SNU13 (McKeegan *et al.*, 2007).

En revanche, il a été montré pour la protéine Bcd1 de levure que ce domaine était nécessaire à la fonction de Bcd1 dans l'assemblage des snoRNP à boîtes C/D (Bragantini *et al.*, 2016), mais sa fonction spécifique n'a pas encore été identifiée. En effet, les domaines en doigt à zinc sont généralement associés à des interactions avec les acides nucléiques, toutefois il a été montré que ce n'était pas le cas pour le domaine ZHD de Bcd1, même si la protéine semble pouvoir se lier de manière non spécifique à l'ARN (Bragantini *et al.*, 2016). De manière intéressante, il a également été montré par l'utilisation de protéines chimériques que le domaine ZHD de la protéine Hit1, mais également d'autres protéines appartenant à famille Zf-HIT, pouvait remplacer le domaine ZHD de la protéine Bcd1 (Bragantini *et al.*, 2016). Ceci peut s'expliquer par la grande similitude entre les structures de ces deux domaines. En revanche, la simple présence d'un motif en doigt à zinc n'est pas suffisante pour restaurer la fonction de la protéine Bcd1.

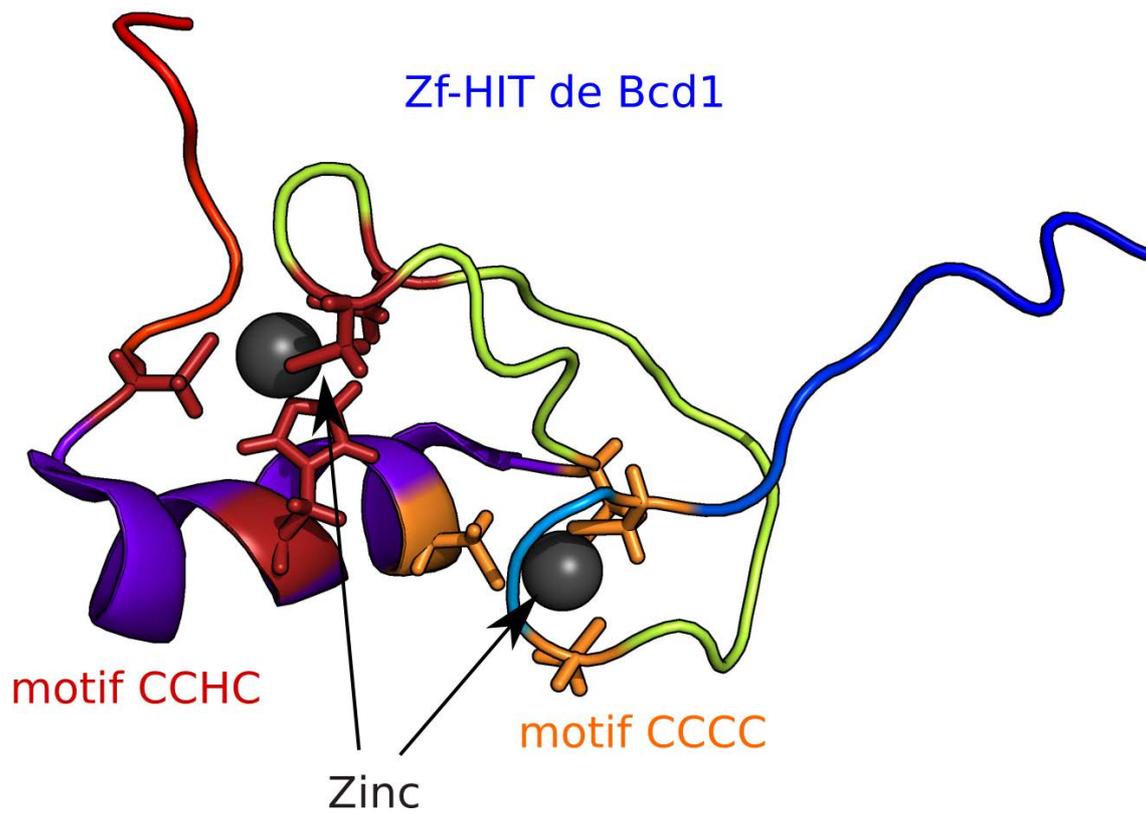


Figure 30: Structure du domaine Zf-HIT de la protéine Bcd1 résolue par RMN par notre laboratoire (code PDB : 2N94).

Le domaine en doigt de zinc (ZHD) de la protéine Bcd1 est constitué de 2 motifs entrecroisés CCCC et CCHC qui vont chacun coordonner un atome de zinc (Bragantini *et al.*, 2016).

Bien que la fonction du domaine ZHD de Bcd1/BCD1 ne soit pas encore élucidée, ce domaine semble important pour la fonction de la protéine lors de l'assemblage de snoRNP à boîtes C/D. Une des hypothèses actuelles serait que ce domaine serve d'intermédiaire entre la protéine Bcd1/BCD1 et les protéines TIP48/Rvb2 et TIP48/Rvb1 (Bragantini *et al.*, 2016).

6. Vers un modèle d'assemblage des snoRNP à boîtes C/D ?

Trente-cinq ans après leur découverte, les mécanismes nécessaires à l'assemblage des snoRNP à boîtes C/D ne sont toujours pas entièrement compris. Toutefois, il a été possible de mettre à jour un large réseau d'interactions directes entre les snoARN, les protéines cœur des snoRNP à boîtes C/D et un certain nombre de facteurs d'assemblage (**Figure 31**), tels que le complexe (h)R2TP ou le complexe NUFIP/ZNHIT3 (Rsa1/Hit1). De plus, des expériences d'immunoprécipitation, de double- et triple-hybride ont permis d'isoler plusieurs complexes intermédiaires. Ainsi, bien qu'il reste encore des zones d'ombre, l'ensemble de ces travaux nous permette aujourd'hui d'émettre des hypothèses plus précises sur le mécanisme d'assemblage séquentiel des snoRNP à boîtes C/D chez la levure (Rothé *et al.*, 2014b) et par extrapolation, chez l'homme (**Figure 32**).

Pour résumer, il a été montré que seule la protéine Snu13/SNU13 pouvait interagir directement et spécifiquement avec l'ARN au niveau du motif C/D structuré en *K-Turn* et être l'élément nucléateur de l'assemblage de la particule. Toutefois, il a également été montré que la protéine Snu13/SNU13 n'interagissait pas avec les autres protéines cœur. Il a également été montré que la formation d'une particule stable nécessitait un remodelage des protéines cœur, probablement via l'activité ATPase des protéines AAA+ Rvb1/TIP49 et Rvb2/TIP48 du complexe (h)R2TP. Au niveau du réseau d'interaction, il a été montré que la protéine Nop58/NOP58 pouvait interagir avec la protéine Pih1/PIH1D1, contrairement à la protéine Nop56/NOP56. Enfin, les deux protéines Nop56/NOP56 et Nop58/NOP58 peuvent interagir toutes les deux par l'intermédiaire de leur domaine en « coiled-coil » mais également avec la protéine Nop1/Fibrillarine.

L'existence de deux motifs de type C/D au niveau de la plupart des snoARN (motif C/D et C'/D') suggère fortement qu'une snoRNP à boîtes C/D mature contiendrait deux sites d'interaction avec l'ARN cible (l'élément anti-sens) et par conséquent 2 protéines Nop1/Fibrillarine et 2 protéines Snu13/SNU13.

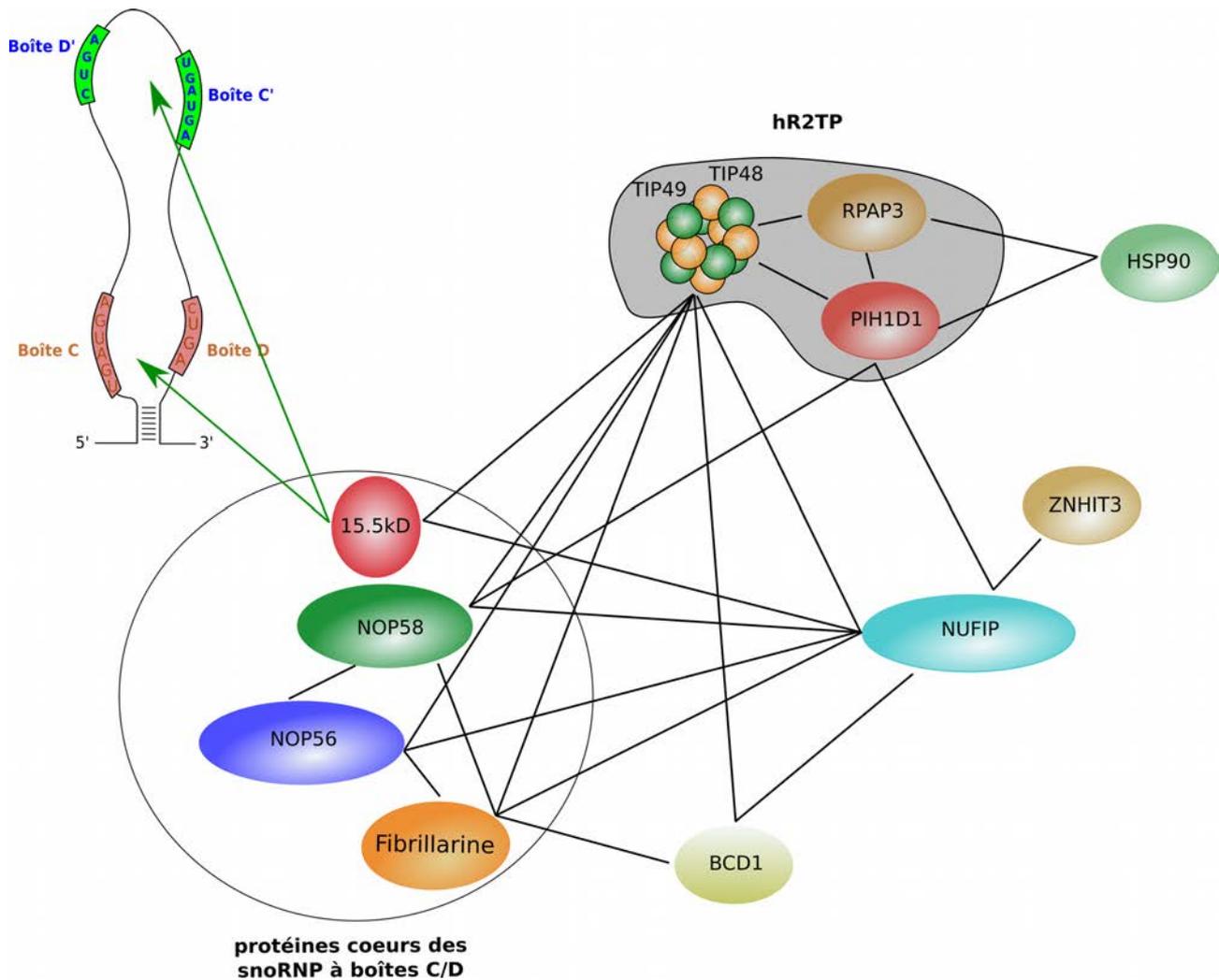


Figure 31: Résumé des interactions directes connues chez l'homme entre les protéines coeurs des snoRNP à boîtes C/D, le complexe hR2TP et les protéines NUFIP, ZNHIT3 et BCD1 (McKeegan et al., 2007 et 2009 ; Boulon et al., 2008 ; Rothé et al., 2014b ; Bizarro et al., 2014).

Grâce à des expériences de GST pull-down, de co-expression et de traduction *in vitro*, il a été possible de mettre en évidence un très large réseau d'interaction entre les protéines des snoRNP à boîtes C/D, le snoARN et de nombreux facteurs d'assemblage. La résolution des structures de certaines de ces interactions a par la suite permis d'isoler les domaines impliqués dans ces interactions.

Flèche verte : interaction protéine/ARN

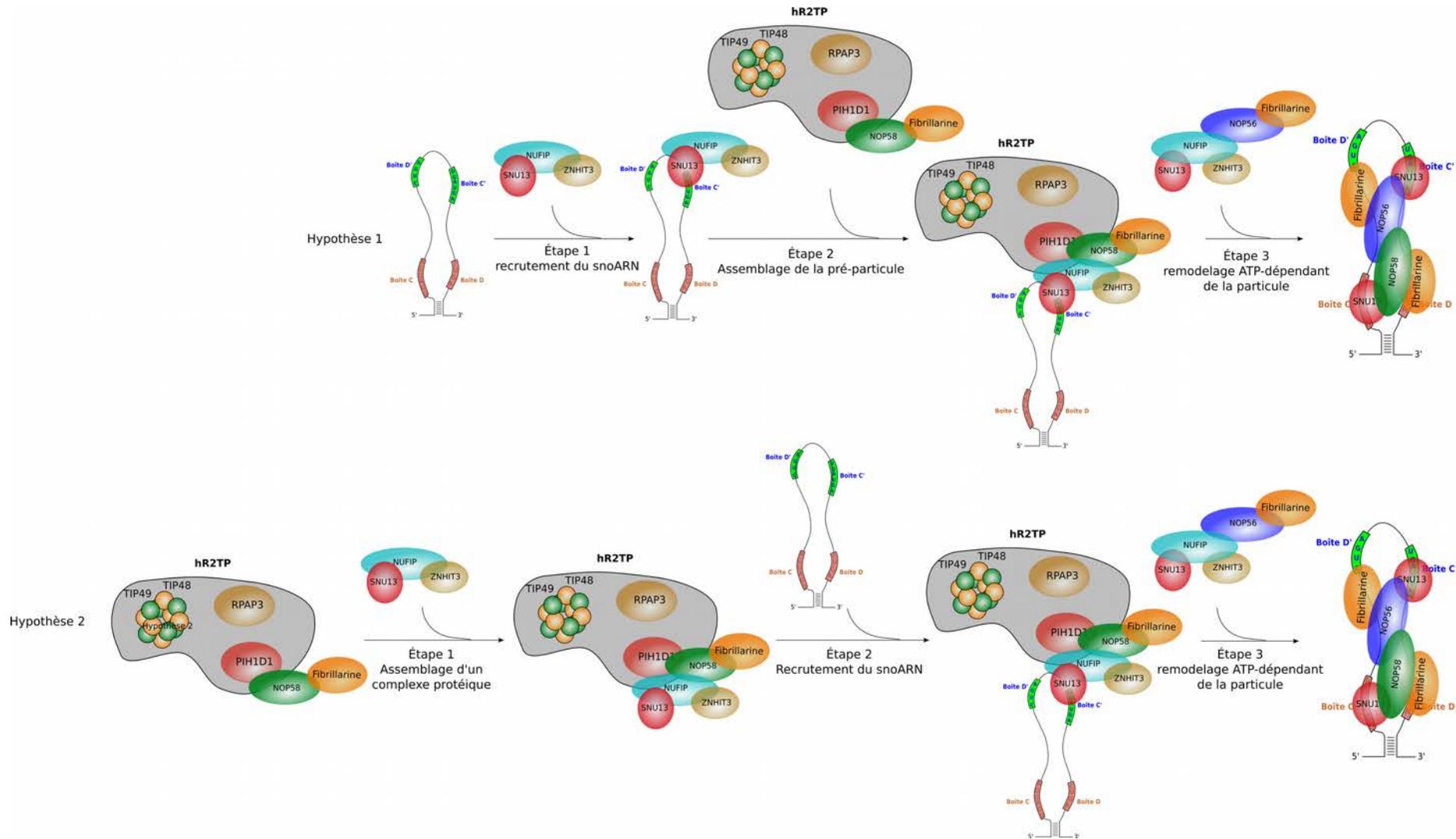


Figure 32: Assemblage des snoRNP à boîtes C/D, hypothèses actuelles (adapté de Rothé *et al.*, 2014b).

La première hypothèse suppose que l'interaction entre le motif C/D du snoARN naissant et la protéine SNU13 est l'élément qui initie l'assemblage de la particule par le recrutement d'un complexe SNU13/NUFIP/ZNHIT3 sur l'ARN. La protéine NUFIP recrute ensuite le complexe hR2TP via son interaction avec la protéine PIH1D1 ce qui permet également de recruter les protéines NOP58 et Fibrillarine. Suite au recrutement des autres protéines cœurs, l'activité ATPase des protéines TIP48 et TIP49 permet de stabiliser l'ensemble des protéines cœurs sur le snoARN.

La seconde hypothèse se différencie au niveau du recrutement du snoARN qui n'aurait lieu qu'une fois le complexe hR2TP recruté par le complexe SNU13/NUFIP/ZNHIT3.

Enfin, la découverte des protéines Rsa1/NUFIP et Hit1/ZNHIT3 a permis de mettre en évidence l'intermédiaire possible entre la protéine Snu13/SNU13, le snoARN, les autres protéines cœurs et les protéines du complexe (h)R2TP.

La formation d'un complexe SNU13-NUFIP-ZNHIT3 (Snu13/Rsa1/Hit1) permettrait ainsi le recrutement d'une part du snoARN au niveau de la protéine Snu13/SNU13 et d'autre part du complexe (h)R2TP, par l'intermédiaire d'une interaction entre la protéine Rsa1/NUFIP et la protéine Pih1/PIH1D1. Le complexe (h)R2TP recruterait alors les protéines cœur Nop58/NOP58 et Nop1/Fibrillarine par l'intermédiaire d'une part d'une interaction entre les protéines Pih1/PIH1D1 et Nop58/NOP58, et d'autre part d'une interaction entre les protéines Nop58/NOP58 et Nop1/Fibrillarine (**Figure 32**, hypothèse 1). Une seconde hypothèse est également envisagée avec le recrutement de snoARN suite au recrutement du (h)R2TP par le complexe SNU13-NUFIP-ZNHIT3 (**Figure 32**, hypothèse 2).

Le recrutement sur la pré-particule d'une seconde protéine Snu13/SNU13 au niveau du motif C'/D' associée avec une seconde enzyme Nop1/Fibrillarine et la protéine Nop56/NOP56 n'est pas encore totalement compris et plusieurs hypothèses sont possibles. Ainsi, le complexe formé des protéines Nop1/Fibrillarine-Nop56/NOP56 pourrait être recruté sur la particule naissante par l'intermédiaire d'un second complexe Snu13/SNU13-Rsa1/NUFIP-Hi1/ZNHTI3, ou directement sur la particule naissante par l'intermédiaire d'une interaction entre les protéines Nop58/NOP58 et Nop56/NOP56 via leur domaine « coiled-coil ». Et le motif C'/D' pourrait également être suffisamment stabilisé par la présence des autres protéines pour pouvoir interagir directement avec une seconde protéine Snu13/SNU13.

Enfin, la résolution de la structure 3D de sRNP à boîtes C/D, des particules d'archaea homologues aux snoRNP à boîtes C/D, a apporté des informations supplémentaires. En effet, il semblerait que les sRNP existent sous deux formes, l'une active et l'autre non (Xue *et al.*, 2010). Le passage de la forme inactive à la forme active nécessite un mouvement de rotation au niveau du site actif de la particule qui passe d'une forme ouverte à une forme fermée. Par homologie avec cette structure, il a été proposé que Rsa1/NUFIP, par l'intermédiaire de son interaction avec la protéine Snu13/SNU13, permette de bloquer la particule dans une conformation « ouverte » inactive (Rothé *et al.*, 2014a et 2014b ; Bizarro *et al.*, 2014). La particule ne deviendrait alors active qu'après le départ de la

protéine Rsa1/NUFIP, suite au remodelage des protéines sur le snoARN par l'action des protéines AAA+ Rvb1/TIP49 Rvb2/TIP48.

De manière remarquable, l'ensemble des facteurs d'assemblage et des protéines cœur des snoRNP à boîtes C/D sont bien conservés entre la levure *S. cerevisiae* et l'homme, à l'exception notable de la protéine RPAP3. En effet, bien qu'elle interagisse avec les protéines PIH1D1 et HSP90 comme son homologue Tah1, elle possède 3 domaines supplémentaires, dont 2 domaines d'interaction putatif protéine-protéine. Si son rôle au niveau de l'assemblage des snoRNP à boîtes C/D semble se limiter à la stabilisation de la protéine Pih1/PIH1D1 par l'intermédiaire du recrutement de la protéine chaperonne Hsp82/HSP90, permettant l'assemblage du complexe (h)R2TP, la présence de domaine supplémentaires chez l'homme suggère qu'elle puisse intervenir dans d'autres mécanismes biologiques.

II. Les snoRNP à boîtes H/ACA

Les snoRNP à boîtes H/ACA sont des ribonucléoprotéines constituées d'un snoARN et de protéines cœur dont l'une porte l'activité catalytique de la particule. La fonction principale des snoRNP à boîtes H/ACA est de catalyser dans le nucléole la conversion d'uridines en pseudouridines (ψ) au niveau des ARNr. Au sein des Corps de Cajal, les scaRNP à boîtes H/ACA catalysent la même réaction au niveau des snARN. La télomérase est une scaRNP à boîtes H/ACA particulière impliquée dans le maintien de la longueur des télomères. Les protéines cœur des snoRNP à boîtes H/ACA sont les protéines NHP2, GAR1, NOP10 et la Dyskérine, qui porte l'activité pseudouridine synthase.

La conversion de l'uridine en pseudouridine nécessite (i) la rupture de la liaison N-glycosidique entre le carbone 1' du ribose et la base, (ii) la rotation de la base et enfin (iii) la formation d'une nouvelle liaison entre le carbone 1' du ribose et l'uracile (**Figure 33**). La pseudouridine possède un groupement donneur d'hydrogène de plus que l'uridine, capable de former une troisième liaison hydrogène. De plus, la liaison carbone-carbone entre la base et le ribose est plus stable que la liaison N-glycosidique de l'uridine (Kierzek *et al.*, 2014). Cette modification est très fréquente au sein des ARN, au point qu'initialement la pseudouridine avait été identifiée comme la « cinquième base » (Davis et Allen, 1957).

A. Les snoARN à boîtes H/ACA

1. Structure des snoARN à boîtes H/ACA

Les snoARN à boîtes H/ACA sont des ARN nucléolaires non-codant d'environ 130 nucléotides (Lestrade et Weber, 2006). Ils sont caractérisés par la présence de deux tiges-boucles reliées par une région charnière contenant la boîte H de séquence consensus 5'-ANANNA-3' (N pouvant être n'importe quel nucléotide). La seconde tige-boucle est immédiatement suivie par une séquence 5'-ACA-3' très conservée. Cette séquence est localisée exactement à 3 nucléotides de l'extrémité 3' du snoARN. Les boîtes H et ACA sont nécessaires à la stabilité et la maturation du snoARN ainsi qu'à l'activité de la particule (Bortolin *et al.*, 1999).

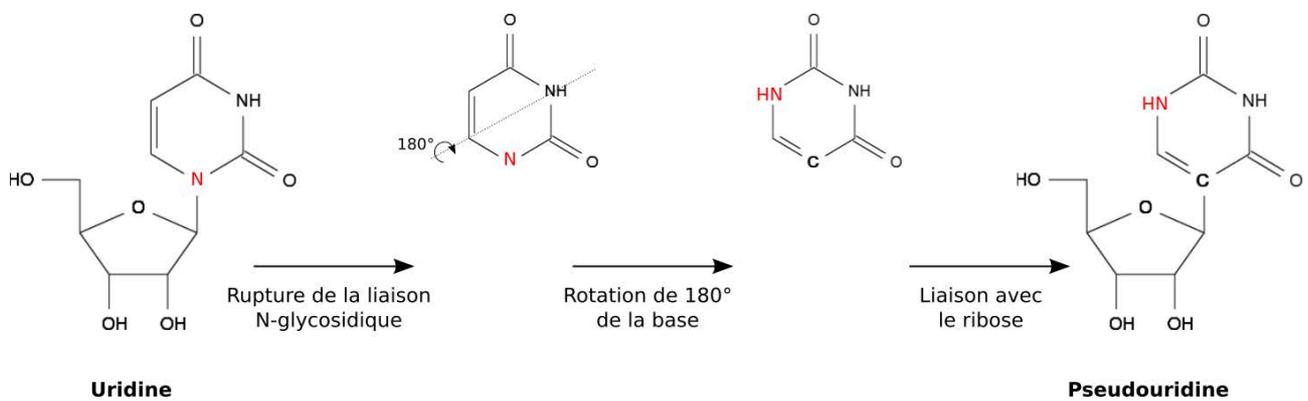


Figure 33: Processus de conversion de l'uridine en pseudouridine.

L'uracile est tout d'abord séparé du ribose suite à l'hydrolyse de la liaison N-glycosidique. Il subit ensuite une rotation de 180° avant d'être de nouveau lié au ribose par une liaison carbone-carbone. La conversion de l'uridine en pseudouridine permet notamment de libérer un nouveau groupement NH (en rouge), donneur d'hydrogène.

Chaque tige-boucle contient une boucle interne, appelée poche de pseudouridylation. Ces poches sont formées par 2 séquences simple brin capables de s'apparier avec 3 à 10 nucléotides de l'ARNr cible (Ganot *et al.*, 1997). Cet appariement conduit à l'exposition de deux nucléotides 5'-UN-3' non-appariés avec le snoARN au niveau de la poche de pseudouridylation. L'uridine non-appariée, toujours située à une distance de 14 ou 16 nucléotides des boîtes H ou ACA, sera isomérisée en pseudouridine (**Figure 34**). Contrairement aux sARN à boîtes H/ACA, leurs homologues chez les archaëa, et aux snoARN à boîtes C/D, les snoARN à boîtes H/ACA ne possèdent pas de structure en *K-Turn*. Enfin, les scaARN à boîtes H/ACA, qui ont pour cibles les snARN, possèdent une boîte additionnelle, la boîte CAB, au niveau des boucles, et qui permet leurs adressages et leurs maintiens dans les Corps de Cajal.

2. Maturation

Chez les vertébrés, mis-à-part l'ARN de la Télomérase, tous les snoARN à boîtes H/ACA sont hébergés dans les introns de pré-ARNm, notamment de protéines impliquées dans la biogenèse des ribosomes (Kiss et Filipowicz, 1995). Suite à l'épissage, le snoARN est libéré sous forme d'un lasso et comme pour les snoARN à boîtes C/D, l'enzyme de débranchement DBR1 est nécessaire à leur libération complète (Kiss et Filipowicz, 1995 ; Ooi *et al.*, 1998). En revanche, contrairement aux snoARN à boîtes C/D, il ne semble pas exister de distance optimale entre l'intron et le site 3' d'épissage (Richard *et al.*, 2006). Ainsi, contrairement aux snoRNP à boîtes C/D, l'assemblage des snoRNP à boîtes H/ACA semble être un processus co-transcriptionnel (**Figure 35**). De même, contrairement aux snoRNP à boîtes C/D, la maturation des snoRNP à boîtes H/ACA ne nécessite pas de passage par les Corps de Cajal.

Tout comme pour les snoARN à boîtes C/D, les extrémités 5' et 3' du snoARN doivent être maturées mais les enzymes responsables de cette maturation n'ont pas encore été identifiées chez les mammifères. Toutefois, il semble y avoir une différence de comportement entre les deux classes de snoARN. En effet, il a été montré *in vitro* que l'excision des 5 à 8 derniers nucléotides de l'extrémité 3' du snoARN était un processus beaucoup plus lent pour les snoARN à boîtes H/ACA que pour les snoARN à boîtes C/D, ce qui suggère une protection de cette extrémité contre la dégradation, probablement suite la fixation des protéines cœur (Kiss et Filipowicz, 1993). De plus, il semble que la maturation de l'extrémité 3' soit facilitée par l'ajout d'une queue oligo(A) par l'enzyme PAPD5.

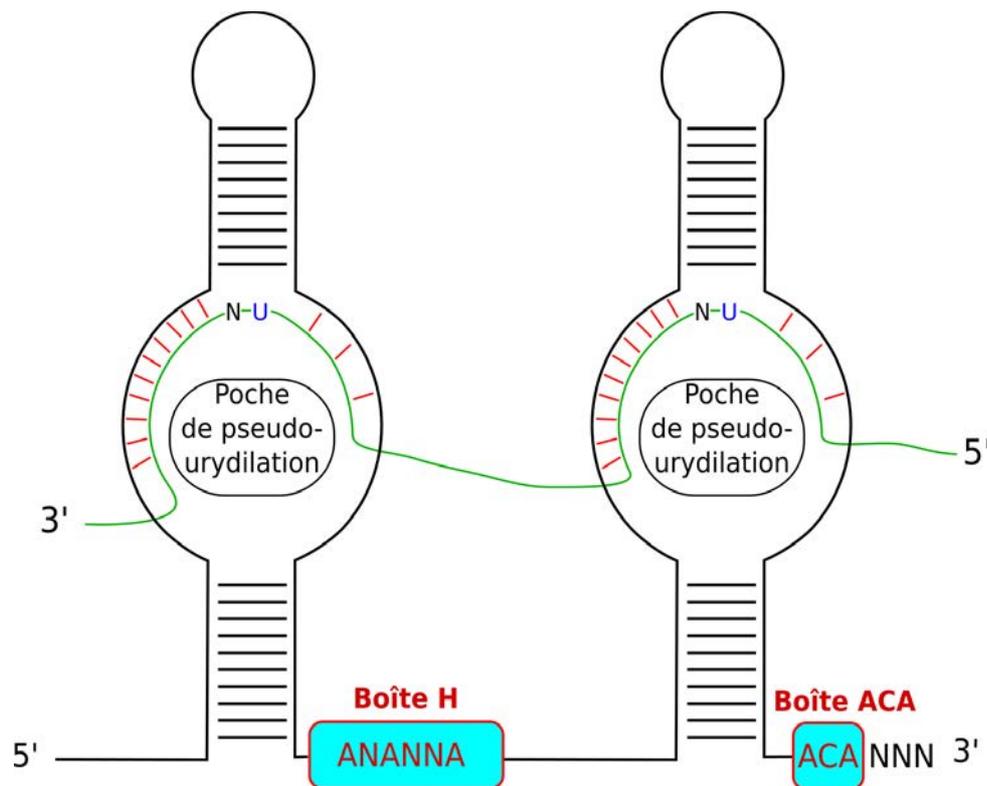


Figure 34 : snoARN à boîtes H/ACA.

Les snoARN à boîtes H/ACA sont caractérisés par la présence de deux tige-boucles connectées par une région charnière qui contient la séquence de la « boîte H ».

La seconde tige-boucle est suivie par un segment simple brin contenant une séquence de trois nucléotides appelée « boîte ACA » qui est localisée exactement à trois nucléotides en amont de l'extrémité 3' du snoARN.

L'une ou les deux tiges-boucles contiennent des boucles internes de 9 à 13 nucléotides sur chaque brin, appelées poches de pseudouridylation. L'ARNr cible (en vert) vient s'hybrider au niveau de cette région, laissant 2 nucléotides non appariés, dont une uridine (en bleu), qui sera isomérisée en pseudouridine.

N : n'importe quel nucléotide.

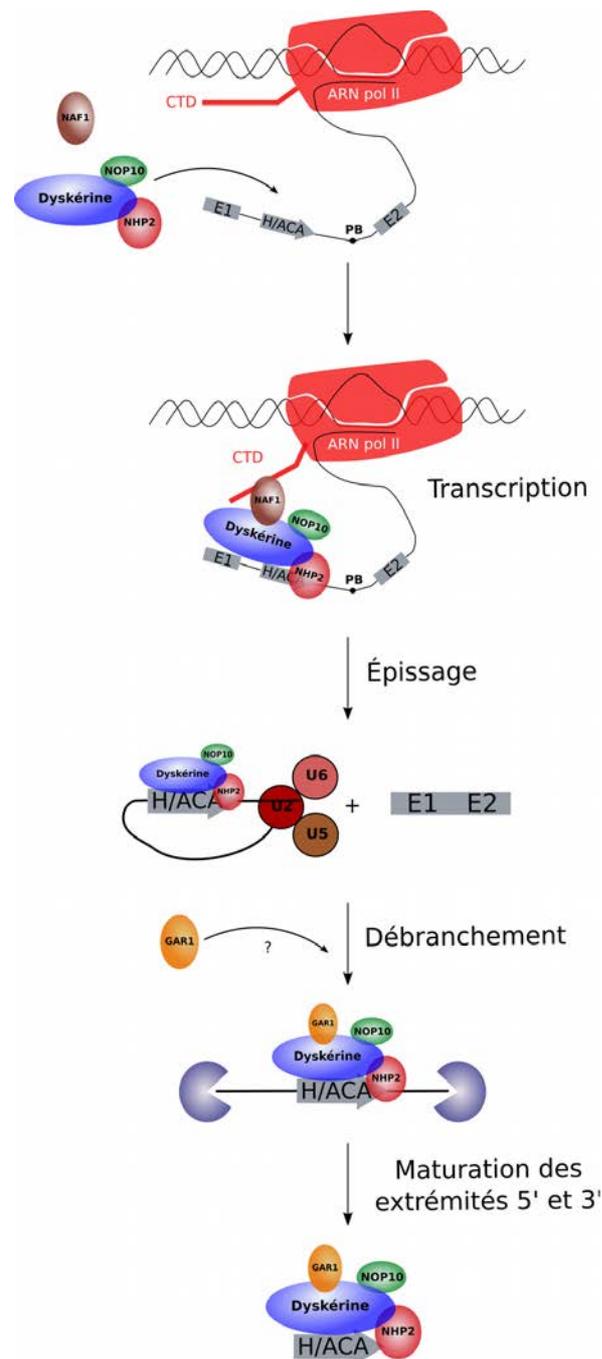


Figure 35: Processus de maturation d'un snoARN à boîtes H/ACA (Kiss *et al.*, 2006).

Contrairement à la maturation des snoARN à boîtes C/D, la maturation des snoARN à boîtes H/ACA est un processus découplé de l'épissage. Ainsi, les protéines cœur des snoRNP à boîtes H/ACA sont recrutées sur le snoARN lors de la transcription du gène hôte. Après épissage, l'intron contenant le snoARN en cours de maturation est linéarisé suite à l'action de l'enzyme de débranchement DBR1. Pour finir les extrémités 5' et 3' sont maturées par des exonucléases.

Cette séquence est ensuite reconnue par l'enzyme PARN (« PolyA RNase») et va permettre la maturation précise de l'extrémité 3' du snoARN (Berndt *et al.*, 2012). Il est possible que ce processus puisse servir de contrôle qualité pour l'assemblage de la particule. En effet, de récentes études sur l'ARN de la télomérase, une scaRNP à boîtes H/ACA particulière, ont montré qu'il existerait un équilibre entre d'une part l'ajout et la suppression de cette queue oligo(A) et l'assemblage de la particule, et d'autre part la reconnaissance puis la dégradation de l'extrémité 3' sans queue poly(A) par l'exosome (Shukla *et al.*, 2016).

B. Les protéines cœur des snoRNP à boîtes H/ACA

Comme décrit précédemment, les snoRNP à boîtes H/ACA (**Figure 36**) ont tout d'abord été isolées par des expériences d'immunoprécipitation réalisées contre la protéine de levure Gar1 (Girard *et al.*, 1992). Ces expériences ont montré que les snoARN à boîtes H/ACA étaient associés à 3 autres protéines, les protéines Nhp2, Nop10 et la protéine Cbf5, une pseudouridine synthase (Lafontaine *et al.*, 1998 ; Watkins *et al.*, 1998 ; Henras *et al.*, 1998). Chez les vertébrés, ces protéines correspondent respectivement aux protéines GAR1, NHP2, NOP10 et la Dyskérine/NP57 (respectivement chez l'homme et chez la souris).

1. La protéine Cbf5/Dyskérine

La protéine Dyskérine est une protéine d'environ 58 kDa qui a été initialement identifiée chez l'homme pour son implication dans la dyskératose congénitale liée au chromosome X ou X-DC. L'X-DC est une maladie génétique comparable à la progéria dont l'une des causes est la présence de mutations au niveau du gène *DKC1*, le gène codant la Dyskérine (Heiss *et al.*, 1998). Chez la levure *S. cerevisiae*, la protéine Cbf5 (« Centromere Binding Factor 5 ») est une protéine essentielle à la survie cellulaire, initialement décrite comme une protéine capable d'interagir *in vitro* avec les centromères et les microtubules (Jiang *et al.*, 1993). D'autres protéines homologues ont ensuite été identifiées chez la souris (NAP57), la drosophile (Mfl) et les plantes (AtNAP57), ainsi que chez les archaea (aCbf5) (Cerrudo *et al.*, 2015).

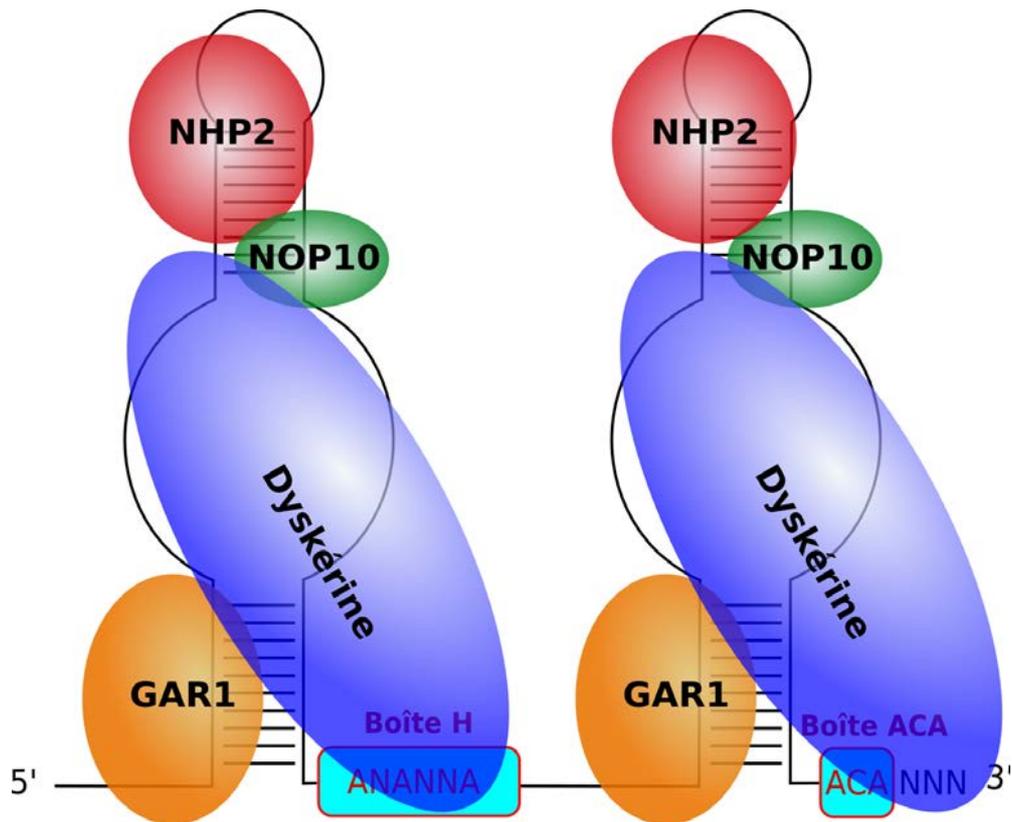


Figure 36: snoRNP à boîtes H/ACA.

Les snoRNP à boîtes H/ACA sont composées d'un snoARN à boîtes H/ACA et de 4 protéines cœur. Ces protéines sont la Dyskérine, responsable de la pseudouridylation, des protéines NOP10 et GAR1 et de la protéine NHP2, une protéine appartenant à la famille des protéines L7Ae.

Ces protéines contiennent un domaine « Dyskerin-like » (DKCLD), dont la fonction n'a pas encore été identifiée mais qui est caractéristique de cette famille de protéines. Elles contiennent également un domaine TruB_N, similaire au domaine N-terminal des pseudouridines synthases des protéines bactériennes de la famille TruB, qui contient le site actif. Enfin, elles comportent aussi un domaine PUA (« PseudoUridine synthase and Archaeosine transglycosylase ») très conservé. Ce domaine est retrouvé dans d'autres familles de protéines et est caractéristique d'un domaine de liaison à l'ARN, notamment avec les boîtes ACA (Reichow *et al.*, 2007 ; Liang *et al.*, 2009). Chez les eucaryotes, des domaines supplémentaires sont retrouvés au niveau des extrémités N- et C-terminales (domaines NTE et CTE, respectivement) (Li *et al.*, 2011). Les domaines NTE et CTE contiennent des séquences de localisation nucléaire (NLS) et nucléolaire (NoLS) (Cerrudo *et al.*, 2015). De plus, le domaine CTE contient de nombreuses répétitions KEE/D. De manière intéressante, une isoforme cytoplasmique dépourvue du domaine CTE a été isolée chez l'homme dans divers tissus et lignées cellulaires au niveau du cytoplasme (Angrisani *et al.*, 2011). Cette nouvelle isoforme suggère que la protéine Dyskérine pourrait avoir d'autres rôles biologiques.

La structure de la protéine humaine n'a pas encore été résolue contrairement à son homologue de levure (Li *et al.*, 2011), où elle l'a été au sein d'un complexe contenant également les protéines Nop10 et Gar1 (**Figure 37**). En se basant sur la forte homologie entre la protéine humaine et la protéine de levure (73 % d'identité), un modèle de la protéine humaine a néanmoins été calculé à partir des données structurales de la protéine de levure (Cerrudo *et al.*, 2015). Les résultats suggèrent que la protéine humaine adopterait une structure identique à la protéine de levure (RMSD de 0,1 Å entre le modèle humain et la structure de levure).

2. La protéine Gar1/GAR1

La protéine Gar1 (« Glycine and Arginine rich protein 1 ») de levure a été initialement identifiée lors d'un crible dont le but était d'identifier les protéines contenant un domaine riche en glycine et en arginine (ou domaine GAR), du même type que celui présent au sein de la Fibrillarine (Girard *et al.*, 1992). Des expériences d'immunoprécipitation sur des extraits totaux de cellules HeLa ont permis d'identifier la protéine humaine GAR1 comme l'homologue de la protéine Gar1 de levure (Dragon *et al.*, 2000).

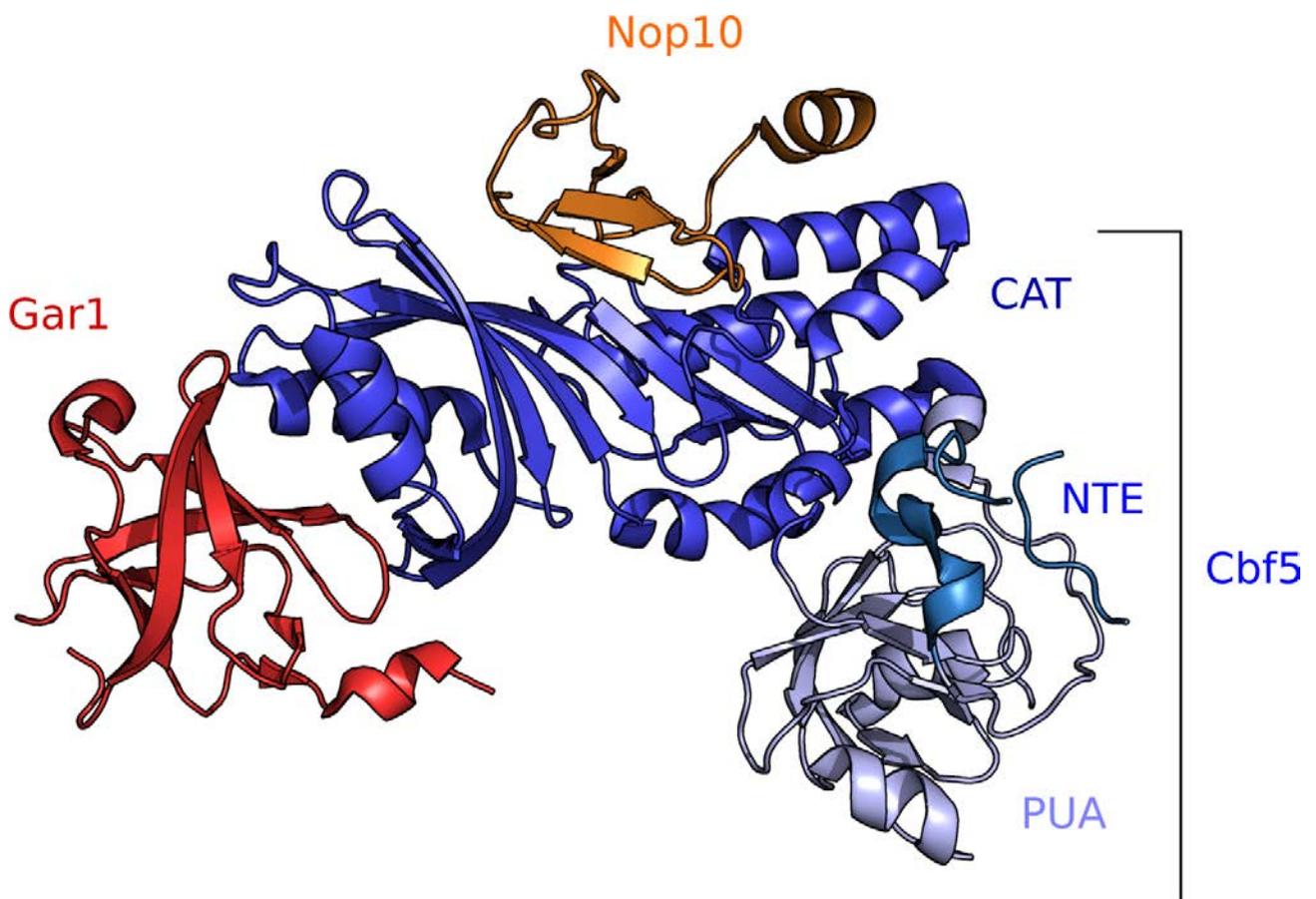


Figure 37: Structure 3D de la protéine de levure Cbf5 en complexe avec les protéines Gar1 et Nop10 résolue par diffraction des rayons X (code PDB : 3U28 ; Li *et al.*, 2011).

Toutes les protéines homologues à la protéine Cbf5 comportent un domaine catalytique (CAT) et un domaine de liaison à l'ARN (PUA). Les protéines eucaryotes possèdent en plus un domaine N-terminal (NTE) ainsi qu'un domaine C-terminal (CTE) riche en répétition KEE/D.

Le domaine CTE n'est pas visible au sein du cristal. De même, la structure du domaine NTE n'a pas été entièrement résolue.

Toutefois, il a été montré que contrairement aux autres protéines cœur des snoRNP à boîtes H/ACA, la protéine Gar1/GAR1 n'était pas requise pour la stabilité des snoARN bien qu'elle soit nécessaire à l'activité de la particule (Bousquet-Antonelli *et al.*, 1997 ; Girard *et al.*, 1992).

La protéine Gar1/GAR1 contient ainsi un domaine central, qui est nécessaire et suffisant pour la fonction de la protéine chez les archaea, flanqué par deux domaines GAR (« Glycine and Arginine Rich ») essentiels à la localisation de la protéine chez les eucaryotes (Pogacíc *et al.*, 2000), ce qui est cohérent avec le fait que, comme pour la protéine Fibrillarine, ces domaines soient absents de la protéine Gar1 d'archaea (Reichow *et al.*, 2007).

Seule la structure du domaine central de la protéine Gar1 est visible au sein du complexe Nop10-Cbf5-Gar1 (**Figure 38**). Il comporte 6 brins β ainsi que deux extensions au niveau de ses extrémités N- et C-terminales, appelées respectivement NTE et CTE. Le domaine NTE, dont la structure a été partiellement résolue, contient deux prolines (P33 et P34) très conservées chez les eucaryotes. Chez l'homme, il a été montré que la présence de mutations au niveau de ces deux résidus entraînait une agrégation de la protéine et supprimait sa localisation nucléolaire (Pogacíc *et al.*, 2000). La structure révèle que ces deux prolines se trouvent au sein d'un cœur hydrophobe et qu'elles sont donc probablement impliquées dans le maintien de l'intégrité de la structure (Li *et al.*, 2011). La protéine Gar1 interagit avec la protéine Cbf5 par l'intermédiaire de liaisons hydrogènes entre le domaine central de la protéine Gar1 et le domaine CAT de la protéine Cbf5, mais également par des interactions hydrophobes au niveau de l'hélice α de son domaine C-terminal (Li *et al.*, 2011).

Chez les archaea, il a été montré que l'interface entre ces deux protéines jouait un rôle crucial dans le recrutement et la modification de l'ARN substrat (Duan *et al.*, 2009). En effet, aCbf5 contient, au niveau de cette interface, une boucle flexible, appelée la « thumb loop », qui est impliquée dans le maintien de l'ARN cible au niveau du site actif. Cette boucle peut adopter une conformation « ouverte », qui la rapproche de la protéine aGar1, ou une conformation « fermée », qui lui permet de stabiliser l'ARN cible au niveau du site catalytique de la protéine aCbf5. La structure de cette interface chez la levure montre d'importantes différences par rapport aux protéines d'archaea (Li *et al.*, 2011). En effet, la boucle y adopte une conformation proche de la conformation « fermée », et ce même en l'absence de substrat.

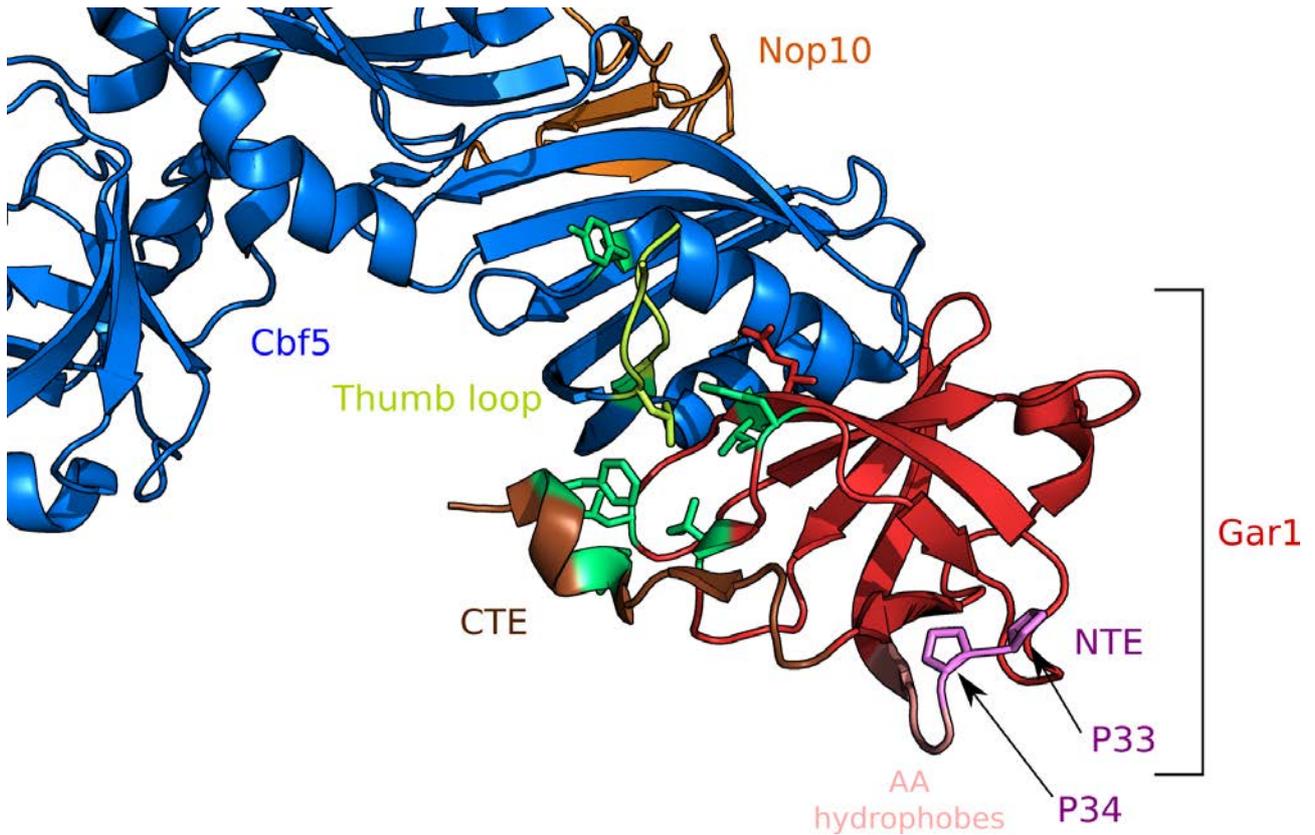


Figure 38: Structure 3D de la protéine de levure Gar1 en complexe avec les protéines Cbf5 et Nop10, résolue par diffraction des rayons X (code PDB : 3U28 ; Li *et al.*, 2011).

Seule la structure du domaine central de la protéine Gar1 a été résolue. Ce domaine interagit avec la protéine Cbf5 par l'intermédiaire de liaisons hydrogènes. Le domaine central des protéines Gar1 eucaryotes contient également deux extensions au niveau de ses extrémités N- et C-terminales, appelées respectivement NTE et CTE. La partie NTE contient deux prolines très conservées au sein d'un cœur hydrophobe, qui sont nécessaires à la stabilité de la protéine. La partie CTE participe de son côté à la régulation de l'activité de la protéine Cbf5 en contrôlant la position de la « thumb loop » (en jaune) par l'intermédiaire de résidus qui forment une poche hydrophobe (en vert).

Cette nouvelle conformation est forcée par la position d'un certain nombre de chaînes latérales d'acides aminés qui d'une part empêchent physiquement la boucle de prendre une véritable conformation « ouverte » et d'autre part stabilisent la boucle au niveau de Gar1 en formant une poche hydrophobe. Bien que cette boucle adopte une conformation très différente de son homologue chez les archaea, il a été montré qu'elle régulait toujours l'activité catalytique de la protéine Cbf5. Ainsi, des mutations réalisées dans le domaine CTE de Gar1 au niveau d'acides aminés engagés dans l'interaction avec la « thumb loop » ainsi que dans la boucle ont montré un fort impact sur l'activité enzymatique en présence de différents substrats. Le domaine CTE de Gar1 serait donc impliqué, de par son interaction avec la « thumb loop » de la protéine Cbf5, dans la régulation du renouvellement du substrat, probablement par un changement de conformation de certains résidus au niveau de la poche hydrophobe de Gar1 (Li *et al.*, 2011).

3. La protéine Nop10/NOP10

La protéine Nop10/NOP10 (« Nucleolar Protein 10 ») est une petite protéine de 10 kDa, associée à la protéine Gar1/GAR1 (Henras *et al.*, 1998 ; Pogacíc *et al.*, 2000). Au sein du complexe Nop10-Cbf5-Gar1, la protéine Nop10 adopte une structure allongée (**Figure 39**). Elle est composée d'un domaine N-terminal qui contient 2 brins β reliés à une unique hélice α C-terminale par une longue région charnière flexible (Li *et al.*, 2011). La protéine Nop10 vient se plaquer contre la protéine Cbf5 avec l'hélice de son domaine C-terminal qui se retrouve contre l'hélice 6 de la protéine Cbf5.

Le domaine N-terminal contient un motif YTLK conservé, y compris chez les archaea, impliqué dans l'interaction avec la protéine Cbf5 via des interactions hydrophobes (Reichow et Varani, 2008). Chez les archaea, ce domaine contient également 4 cystéines qui vont coordonner un ion Zn^{2+} , ce qui permet de stabiliser le motif YTLK (Reichow et Varani, 2008). Chez les eucaryotes, le domaine N-terminal de la protéine libre n'est pas structuré du fait de l'absence de ces 4 cystéines. Des études par RMN sur la protéine libre et en solution avec la protéine Cbf5 ont montré que la fixation de la protéine Cbf5 sur la protéine Nop10 via le motif YTLK induisait une structuration du domaine N-terminal identique à celle adoptée par la protéine aNop10 d'archaea (Reichow et Varani, 2008 ; Li *et al.*, 2011).

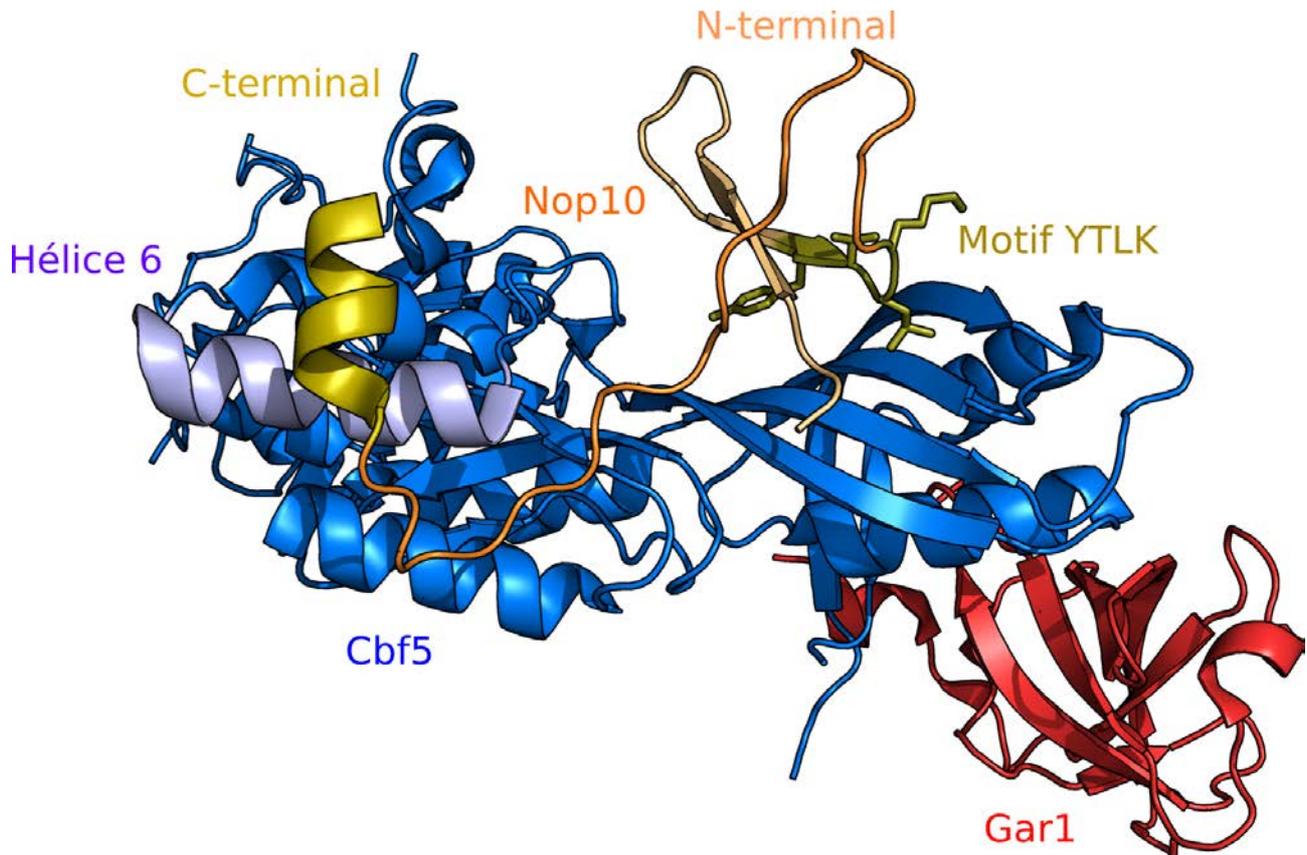


Figure 39: Structure 3D de la protéine de levure Nop10 en complexe avec les protéines Cbf5 et Gar1 résolue par diffraction des rayons X (code PDB : 3U28 ; Li *et al.*, 2011).

La protéine Nop10 est formée par un domaine N-terminal contenant 2 brins β reliés à une unique hélice α C-terminale par une longue région charnière. La protéine Nop10 vient se plaquer contre la protéine Cbf5 avec notamment l'hélice du domaine C-terminal (ici incomplète) qui se retrouve contre l'hélice 6 de la protéine Cbf5 en formant un angle droit. L'interaction entre la protéine Cbf5 et le motif conservé YTLK du domaine N-terminal de la protéine Nop10 induit une structuration de ce domaine sous la forme d'un feuillet β .

La protéine Nop10/NOP10 est une protéine essentielle pour la stabilité et la maturation des snoRNP à boîtes H/ACA. Ainsi, la résolution de la structure des sRNP à boîtes H/ACA d'archaea (Liang *et al.*, 2009) a montré que la protéine aNop10 était impliquée dans de nombreuses interactions intermoléculaires au sein de la particule et qu'elle servirait ainsi d'adaptateur entre les différents composants de la sRNP. Les acides aminés impliqués dans ces interactions sont retrouvés au sein de protéine Nop10, ce qui suggère que ce rôle est conservé au sein des snoRNP eucaryotes (Reichow et Varani, 2008). De plus, des études ont montré que l'interaction entre les protéines Cbf5 et Nhp2 dépendait de la protéine Nop10 (Wang et Meier, 2004), et des études de pontages covalents ont montré l'existence de contacts entre le snoARN et la protéine Nop10 (Dragon *et al.*, 2000).

4. La protéine Nhp2/NHP2

La protéine Nhp2/NHP2 (« Non-Histone Protein 2 ») a été identifiée en même temps que la protéine Nop10/NOP10 et est également essentielle à la stabilité et la maturation des snoRNP à boîtes H/ACA (Henras *et al.*, 1998 ; Pogacíc *et al.*, 2000). Elle appartient à la famille des protéines L7Ae, mais contrairement aux autres protéines de cette famille, elle est la seule à ne pas reconnaître spécifiquement les motifs en *K-Turn*. Ce motif n'est néanmoins pas présent au niveau des snoARN à boîtes H/ACA chez les eucaryotes. Des études *in vitro* ont montré par ailleurs que la protéine de levure avait tout de même la capacité d'interagir directement avec l'ARN (Henras *et al.*, 2001).

La structure de la protéine de levure a été résolue par RMN. Elle montre que la protéine Nhp2 adopte un repliement comparable aux autres protéines L7Ae avec notamment la présence d'une succession d'hélices α et de brins β qui forment un « sandwich α - β - α » (Koo *et al.*, 2011). La protéine Nhp2 adopte une structure très proche de la protéine SNU13. Par exemple, les motifs VSRP et NExxK impliqués dans l'interaction de la protéine SNU13 avec l'ARN sont relativement bien conservés au sein de la protéine Nhp2 (respectivement TKRP et KExxK). Toutefois, la protéine Nhp2 peut adopter deux conformations suite à un changement d'isomérisation *cis/trans* de la proline 83, alors que ce résidu a une conformation exclusivement *cis* dans les autres protéines L7Ae (Koo *et al.*, 2011).

Or, ce résidu, qui est localisé à l'extrémité N-terminale de l'hélice 3, se trouve dans la zone d'interaction présumée avec le snoARN et la protéine Nop10. Il a alors été montré que la stabilisation de la conformation *cis* de la proline 83 entraînait un déplacement de l'hélice 3 (**Figure 40**) et stabilisait *in vivo* l'interaction de la protéine Nhp2 avec le snoARN et la protéine Nop10. Ce changement de conformation entraîne également une modification de la position de la sérine 82 voisine dont les groupements amide et carboxyle de la chaîne principale se rapprochent de la poche d'interaction avec l'ARN. Ce résidu a alors une position comparable avec celle de l'acide glutamique E61 de la protéine SNU13. De plus, bien que chez les eucaryotes les snoARN à boîtes H/ACA ne possèdent pas de motif en *K-Turn*, des alignements de séquences ont montré dans 73 % des cas la présence d'une uridine conservée, située à une distance comparable de l'uridine protubérante du *K-Turn* des sRNP à boîtes H/ACA d'archaea. Les auteurs proposent ainsi que l'acide aminé 82, indépendamment de sa chaîne latérale, soit impliqué dans la reconnaissance de ce résidu uridine (Koo *et al.*, 2011). Ainsi, contrairement aux autres protéines de la famille L7Ae, l'interaction entre la protéine Nhp2 et l'ARN serait dépendante d'un changement de conformation probablement favorisé par l'association de la protéine Nhp2 avec les autres protéines des snoRNP à boîtes H/ACA (Koo *et al.*, 2011).

C. Les facteurs d'assemblage des snoRNP à boîtes H/ACA

Contrairement aux snoRNP à boîtes C/D, il a été possible de reconstituer *in vitro* des particules fonctionnelles d'archaea (Charpentier *et al.*, 2005) mais également de levures (Li *et al.*, 2011). Toutefois il semble que chez les eucaryotes, l'assemblage *in vivo* des snoRNP à boîtes H/ACA soit plus complexe et nécessite l'intervention de facteurs d'assemblage, en particulier les protéines NAF1/Naf1 et SHQ1/Shq1 (Dez *et al.*, 2002 ; Yang *et al.*, 2002 ; Fatica *et al.*, 2002).

1. La protéine Shq1/SHQ1

La protéine Shq1/SHQ1 (« Small nucleolar RNAs of the box H/ACA family Quantitative accumulation protein ») est une protéine conservée chez les eucaryotes, nécessaire à l'assemblage des snoRNP à boîtes H/ACA mais pas à leur activité (Grozdanov *et al.*, 2009). Ce facteur d'assemblage a été initialement identifié comme un partenaire potentiel de la RNase III Rnt1 lors de tests double-hybride (Yang *et al.*, 2002). Toutefois, il n'a pas été possible de confirmer l'interaction par GST pull-down, ce qui suggère l'association entre ces deux protéines plutôt qu'une interaction stable.

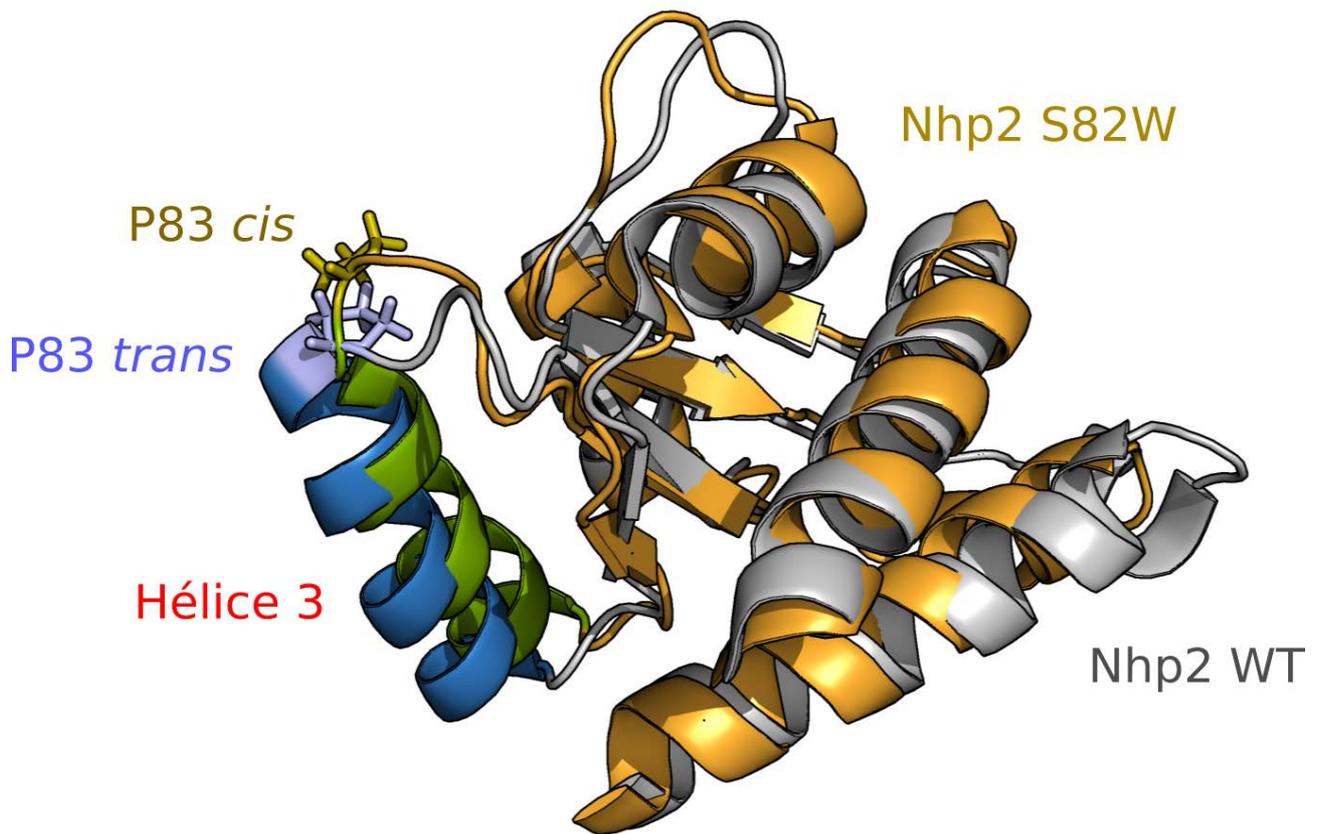


Figure 40: Comparaison des différentes structures 3D de la protéine Nhp2 induites par le changement de conformation de la proline 83 (Koo *et al.*, 2011).

Koo *et al.* ont substitué la sérine 82 par un tryptophane afin de forcer la proline 83 à adopter majoritairement une conformation *cis*. La comparaison des deux structures montre notamment un déplacement significatif de l'hélice α_3 .

Code PDB : Nhp2 WT : 2LBX - Nhp2 S82W : 2LBW

En revanche, il a été montré que la protéine SHQ1 humaine interagissait directement avec la protéine Dyskérine, indépendamment de la présence des autres protéines cœur (Grozdanov *et al.*, 2009). La déplétion de cette protéine chez les levures entraîne une diminution spécifique du taux de snoARN à boîtes H/ACA (Yang *et al.*, 2002). Enfin, des expériences d'immunofluorescence ont montré que la protéine était présente dans le noyau mais pas dans le nucléole.

La protéine Shq1/SHQ1 contient un domaine N-terminal CS et un domaine C-terminal caractéristique « SHQ1-Specific Domain » (SSD). Le domaine CS possède une homologie de structure avec certaines co-chaperonnes de la protéine HSP90 et contient 7 brins β ainsi qu'une petite hélice α (Godin *et al.*, 2009). Des expériences de complémentation chez la levure *S. cerevisiae* ont montré que séparément, les domaines CS et SSD ne pouvaient restaurer l'activité de la protéine (Machado-Pinilla *et al.*, 2012). Le domaine SSD est nécessaire pour stabiliser la protéine Cbf5/Dyskérine (Machado-Pinilla *et al.*, 2012), avec laquelle il interagit (Grozdanov *et al.*, 2009).

Contrairement à la plupart des domaines CS connus, le domaine CS de la protéine Shq1/SHQ1 n'interagit pas avec la protéine HSP90 (Godin *et al.*, 2009). En revanche, un rôle de chaperonne pour la protéine Shq1/SHQ1 a pu être démontré *in vitro* (Godin *et al.*, 2009). Enfin, il a été montré aussi que la sérine 63 du domaine CS pouvait être phosphorylée par la protéine CK1 (« Casein Kinase 1 »), connue pour réguler l'activité de nombreuses protéines impliquées dans divers processus cellulaires dont la réparation de l'ADN, le transport membranaire ou encore le transport nucléaire (Godin *et al.*, 2009).

Le domaine C-terminal SSD est responsable de l'interaction avec la protéine Cbf5/Dyskérine, au niveau du domaine PUA (Walbott *et al.*, 2011). La structure 3D de ce domaine en interaction avec la protéine Cbf5 a été résolue par diffraction aux rayons X (Walbott *et al.*, 2011). Le domaine SSD forme une structure unique appelée « maki fold » (Walbott *et al.*, 2011) composé de 17 hélices α et de 2 brins β insérés entre les hélices 16 et 17 (**Figure 41**).

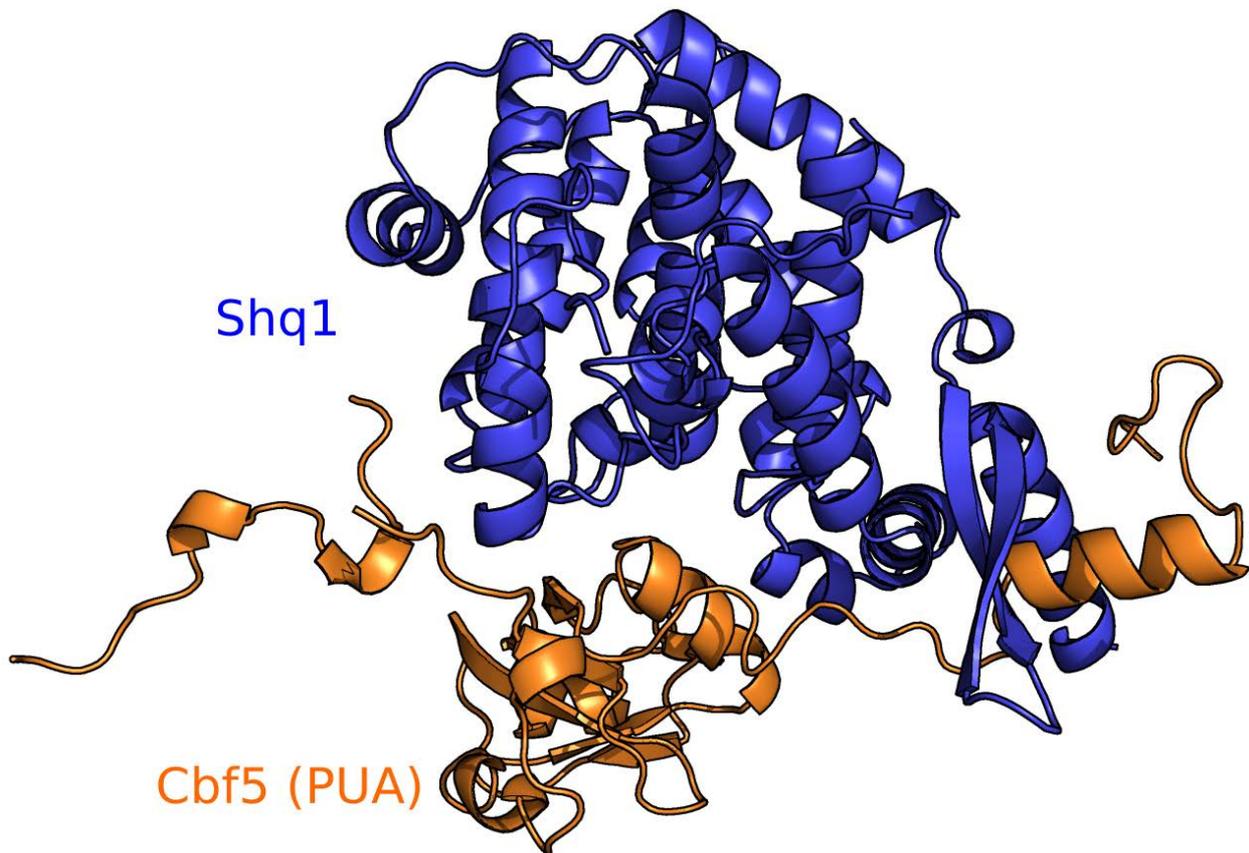


Figure 41: Structure 3D du domaine C-terminal de la protéine Shq1 en interaction avec le domaine PUA de la protéine Cbf5 résolue par diffraction des rayons X (code PDB : 3ZV0).

Le domaine C-terminal de la protéine Shq1 adopte un repliement unique appelé « maki fold », composé de 17 hélices α et de 2 brins β . L'analyse des résidus exposés en surface montre la présence conservée de résidus acides. De plus, le recrutement de la protéine Shq1 sur le domaine PUA de la protéine Cbf5 s'effectue au niveau des mêmes résidus qui sont impliqués, chez l'archaea, dans l'interaction entre le domaine PUA et la boîte ACA de l'ARN (Walbott *et al.*, 2011). Ainsi, la protéine Shq1 interagit avec la protéine Cbf5 en mimant la configuration d'un snoARN, empêchant ainsi la fixation trop précoce d'un snoARN.

Les structures de l'interface d'interaction entre les protéines Cbf5 et Shq1 et entre la protéine d'archaea aCbf5 et l'ARN ont été comparées. Les données indiquent que la fixation de la protéine Shq1 sur la protéine Cbf5 implique les mêmes régions, et empêche l'interaction du domaine PUA avec la boîte ACA du snoARN (Walbott *et al.*, 2011). De plus, la compétition entre l'ARN et la protéine Shq1 pour la fixation sur la protéine Cbf5 a pu être observée *in vitro* (Walbott *et al.*, 2011). Ces données suggèrent ainsi que le rôle de la protéine Shq1 lors de l'assemblage des snoRNP à boîtes H/ACA serait d'empêcher le recrutement prématuré du snoARN sur la protéine Cbf5.

2. La protéine Naf1/NAF1

La protéine Naf1 (« Nuclear association factor 1 ») a été initialement identifiée au cours d'un crible double-hybride exhaustif chez la levure *S. cerevisiae* comme un partenaire potentiel des protéines Cbf5, Shq1 et Nhp2 (Ito *et al.*, 2001). Tout comme pour la protéine Shq1, la déplétion de cette protéine chez la levure et chez l'homme entraîne une diminution spécifique du taux de snoARN à boîtes H/ACA (Yang *et al.*, 2002 ; Dez *et al.*, 2002 ; Hoareau-Aveilla *et al.*, 2006). Des expériences de GST pull-down et de traduction *in vitro* ont montré que la protéine Naf1 interagissait directement avec les protéines Cbf5 et Nhp2 (Fatica *et al.*, 2002). Ces interactions ont également été retrouvées chez l'homme entre la protéine NAF1 et les protéines Dyskérine et NHP2 (Hoareau-Aveilla *et al.*, 2006). De plus, des expériences d'immunofluorescence ont montré que la protéine NAF1 était localisée principalement dans le nucléoplasme, mais également dans le cytoplasme (Kittur *et al.*, 2006). En revanche, comme les autres facteurs d'assemblage, elle n'a pas été retrouvée dans le nucléole (Hoareau-Aveilla *et al.*, 2006).

Enfin, il a été montré que le recrutement de la Dyskérine par la protéine NAF1 s'effectuait au niveau des sites de transcription des gènes codant les snoARN (Hoareau-Aveilla *et al.*, 2006), où une association entre la protéine Naf1 et la queue CTD de l'ARN polymérase II a pu être observée (Fatica *et al.*, 2002). Ainsi, l'une des fonctions de la protéine Naf1/NAF1 serait de faire le lien entre l'ensemble des protéines cœur, à l'exception de la protéine Gar1/GAR1, et le snoARN en cours de transcription (Richard *et al.*, 2006). En effet, il a été montré que les protéines Cbf5 et Naf1 étaient recrutées de manière co-transcriptionnelle durant l'assemblage des snoRNP à boîtes H/ACA (Yang *et al.*, 2005). En revanche, il n'a pas été possible d'isoler de complexes contenant à la fois la protéine Naf1/NAF1 et la protéine Gar1/GAR1. En effet, contrairement aux autres protéines cœur, il semblerait que la protéine Gar1/GAR1 s'associe tardivement avec la pré-snoRNP, une fois la

protéine Naf1/NAF1 libérée (Darzacq *et al.*, 2006). La résolution de la structure 3D d'une partie de la protéine Naf1 a ainsi permis de montrer que les protéines Naf1 et Gar1 possédaient un domaine identique et étaient de fait en compétition pour leur interaction avec la protéine Cbf5/Dyskérine.

En effet, la protéine Naf1/NAF1 est composée de 4 domaines. Un domaine N-terminal riche en sérines, un domaine central, un domaine contenant un signal NLS et un domaine C-terminal riche en prolines et en glutamines. La structure du domaine central a été résolue par diffraction des rayons X (Leulliot *et al.*, 2007). Ce domaine, qui semble pouvoir former un homodimère, contient 7 brins β et une hélice α insérée entre les brins 5 et 6 (**Figure 42**). La structure de chaque monomère est très proche de celle adoptée par le domaine central de la protéine Gar1 (RMSD de 1,64 Å sur 69 résidus et des séquences identiques à 19%), à l'exception de l'hélice α qui est absente des feuilletts β de la protéine Gar1 (Leulliot *et al.*, 2007). La superposition de la structure de ce domaine à la structure de la protéine Gar1 au sein du complexe Nop10-Cbf5-Gar1 (**Figure 43**) montre clairement que les protéines Gar1 et Naf1 partagent le même site d'interaction avec la protéine Cbf5, ce qui explique pourquoi la protéine Gar1 n'est recrutée que tardivement lors de l'assemblage des snoRNP à boîtes H/ACA, suite à son échange avec la protéine Naf1 (Leulliot *et al.*, 2007 ; Li *et al.*, 2011).

Comme décrit précédemment, la protéine Gar1/GAR1 joue un rôle majeur dans la fonction de la snoRNP. En retardant le plus longtemps possible l'association de la protéine GAR1/Gar1 avec le reste de la particule, la protéine Naf1/NAF1 permettrait l'inhibition des particules immatures.

D. Assemblage des snoRNP à boîtes H/ACA

Comme pour les snoRNP à boîtes C/D, la compilation des données structurales et des réseaux d'interaction nous permet d'avoir une idée assez précise, même si elle est encore incomplète, du mécanisme d'assemblage des snoRNP à boîtes H/ACA.

Tout d'abord, il semble qu'*in vitro*, l'assemblage de la particule puisse se faire chez les archaea et chez les eucaryotes uniquement à partir des 4 protéines cœur, contrairement aux snoRNP à boîtes C/D (Charpentier *et al.*, 2005 ; Li *et al.*, 2011). Toutefois, *in vivo*, l'assemblage de la particule semble être plus complexe et nécessiter l'intervention d'au moins deux facteurs d'assemblage permettant de réguler le processus d'assemblage.

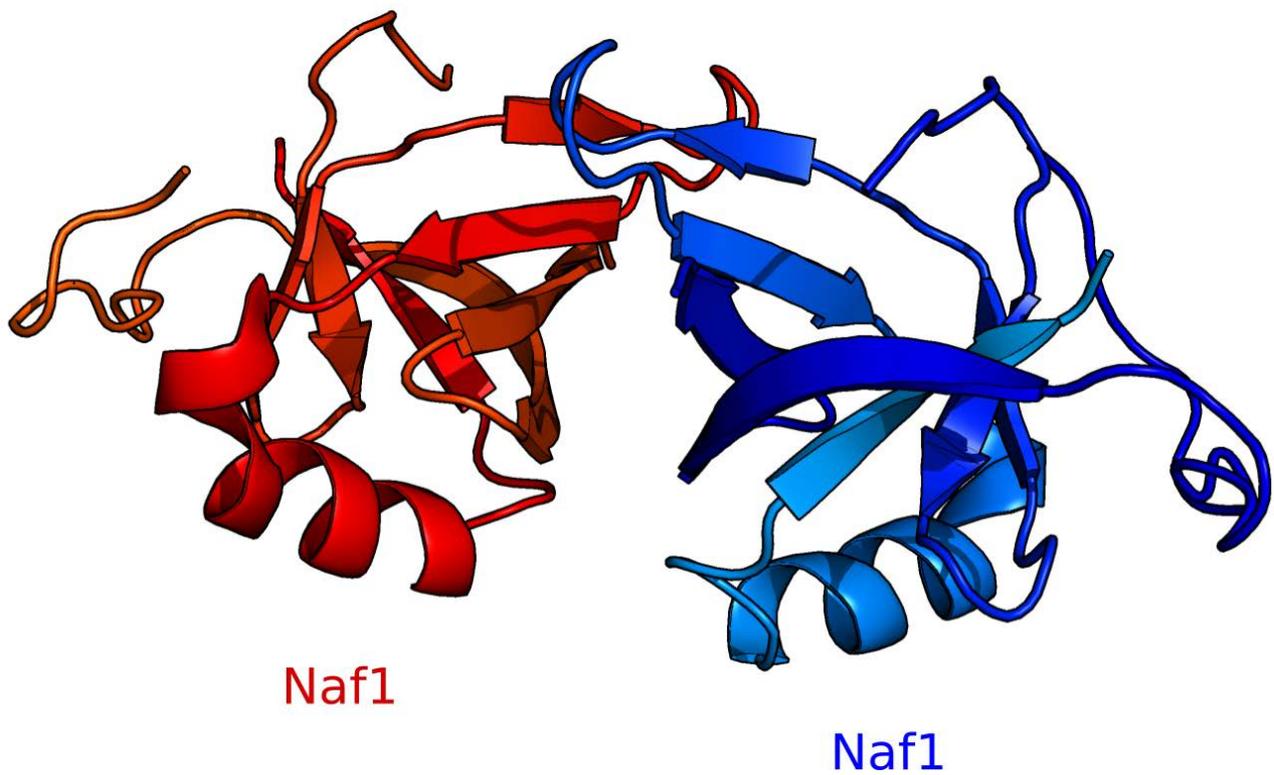


Figure 42: Structure 3D du domaine central de la protéine Naf1 résolue par diffraction des rayons X (code PDB : 2V3M).

Le domaine central de la protéine Naf1 est capable de former un homodimère. Chaque monomère est composé de 7 brins β et d'une hélice α insérée entre les brins 5 et 6. Ce domaine adopte une structure similaire au domaine central de la protéine Gar1.

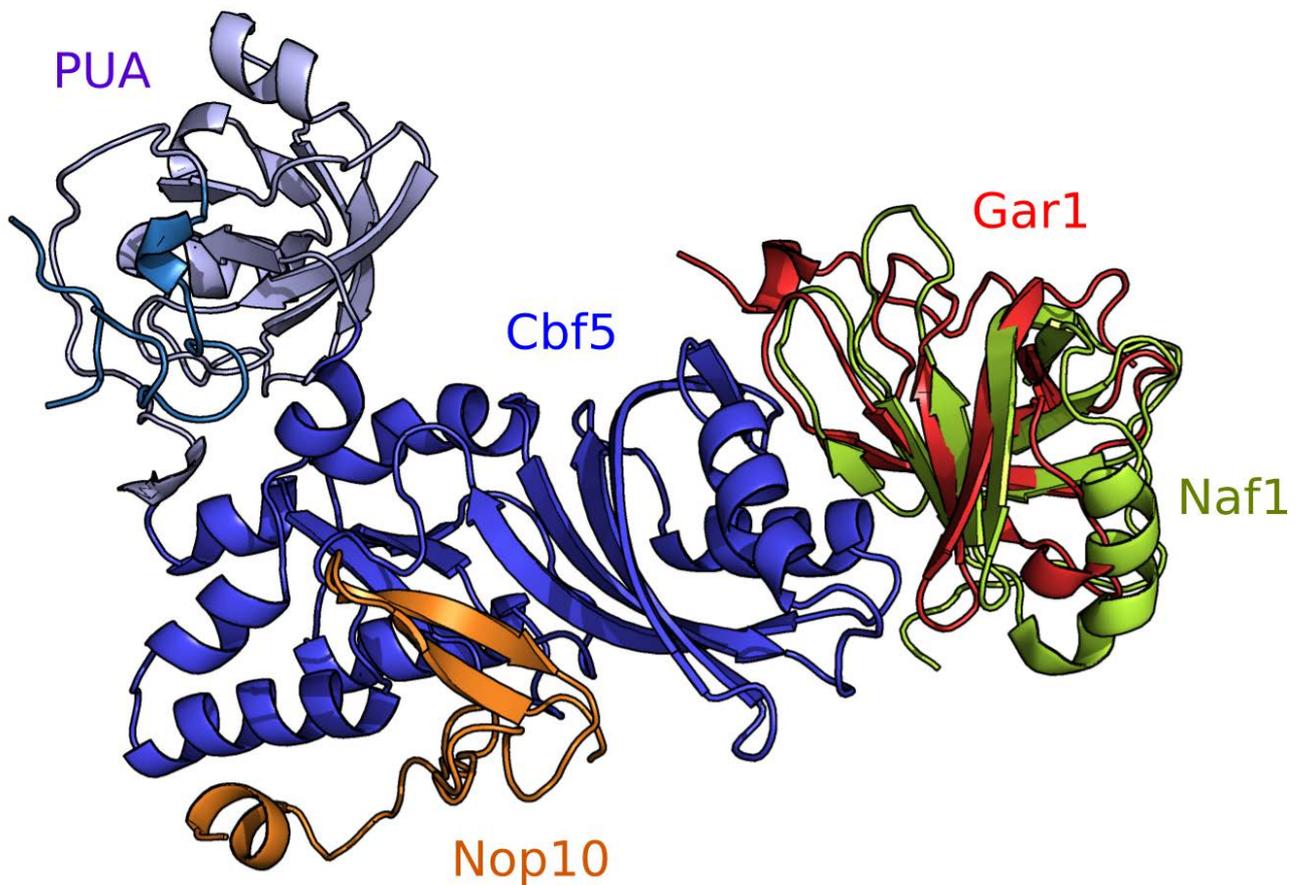


Figure 43: Superposition des structures des domaines centraux des protéines Naf1 et Gar1 au sein du complexe Nop10-Cbf5-Gar1.

La superposition a été réalisée à l'aide du logiciel *Pymol*. Elle montre une forte similarité entre les domaines centraux des protéines Naf1 et Gar1, notamment au niveau des feuillettes β . Cette similarité explique l'absence de la protéine Gar1 lorsque la protéine Naf1 est associée avec la protéine Cbf5 et la nécessité d'un remodelage protéique pour échanger la protéine Naf1 par la protéine Gar1 lors des dernières étapes de l'assemblage des snoRNP à boîtes H/ACA.

Ainsi, les données structurales suggèrent que les protéines Nop10/NOP10, Nhp2/NHP2 et Cbf5/Dyskérine formeraient un pré-complexe (Li *et al.*, 2011) nécessitant *in vivo* l'intervention du facteur d'assemblage Shq1/SHQ1 et probablement la formation de deux sous-complexes, d'une part entre les protéines Nhp2/NHP2 et Nop10/ NOP10 et d'autre part entre les protéines Cbf5/Dyskérine et Shq1/SHQ1. En effet, la protéine Nhp2/NHP2 ne peut pas s'associer à la protéine Cbf5/Dyskérine en l'absence de la protéine Nop10/NOP10 (Koo *et al.*, 2011). De même, la protéine Shq1/SHQ1 s'associe précocement avec la protéine Cbf5/Dyskérine et semble avoir plusieurs rôles. Ainsi, la résolution de la structure du complexe Shq1-Cbf5 a permis de montrer que l'interaction entre la protéine Shq1 et la protéine Cbf5 empêchait le recrutement du snoARN. De plus, la protéine Shq1/SHQ1, de part sa fonction de protéine chaperonne, pourrait également avoir pour rôle de stabilisation de la protéine Cbf5/Dyskérine *in vivo*.

Une fois le complexe SHQ1-Dyskérine-NOP10-NHP2 (ou Shq1-Cbf5-Nop10-Nhp2 chez la levure) formé (**Figure 44**, étape 1), celui-ci serait transféré du cytoplasme vers le noyau. Il est aussi possible que les deux sous-complexes SHQ1-Dyskérine et NOP10-NHP2 soient transportés indépendamment dans le noyau avant de s'associer. Une fois dans le noyau, le complexe peut être recruté par la protéine Naf1/NAF1 au niveau des sites de transcription des gènes codant des snoARN à boîtes H/ACA par l'intermédiaire de l'association de cette protéine avec l'extrémité CTD de l'ARN polymérase II (Yang *et al.*, 2005).

Toutefois, s'il est admis que les interactions des protéines Naf1/NAF1 et Gar1/GAR1 avec la protéine Cbf5/Dyskérine sont mutuellement exclusives, le mécanisme qui empêche la protéine Gar1/GAR1 d'interagir avec la protéine Cbf5/Dyskérine n'a pas encore été identifié. La protéine Naf1/NAF1 étant capable de faire la navette entre le noyau et le cytoplasme (Darzacq *et al.*, 2006), il est également possible que l'arrivée de la protéine Naf1/NAF1 sur le pré-complexe se fasse au niveau du cytoplasme.

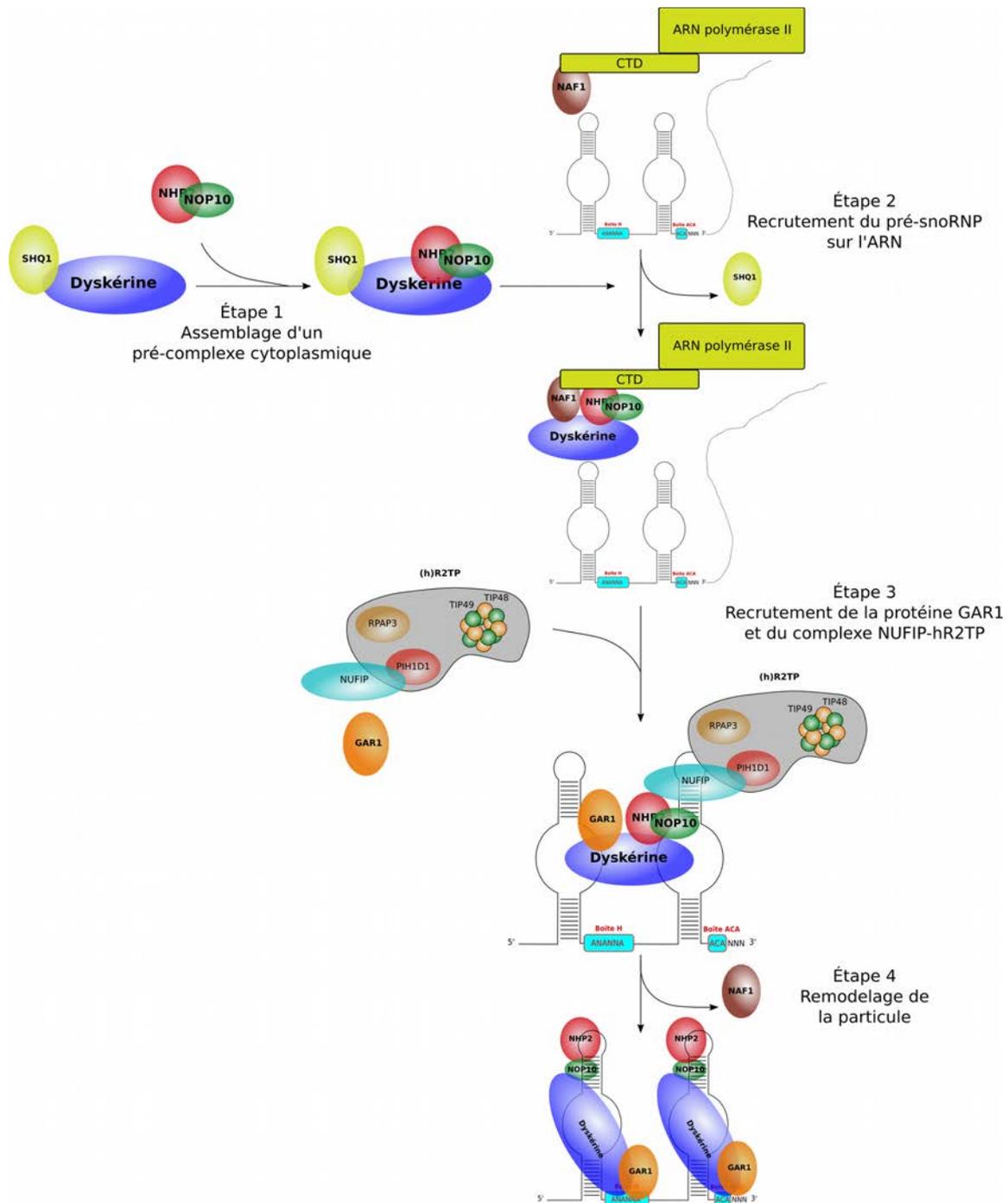


Figure 44: Hypothèses d'assemblage des snoRNP à boîtes H/ACA.

Bien qu'encore hypothétique, un certain nombre d'étapes du mécanisme d'assemblage des snoRNP à boîtes H/ACA semblent aujourd'hui comprises. Ainsi, la première étape de la biogenèse de ces particules serait la formation d'un pré-complexe cytoplasmique contenant au moins 3 des 4 protéines cœur. Le complexe serait ensuite transféré dans le noyau pour être recruté au niveau des sites de transcription des gènes codant les snoARN à boîtes H/ACA. Une fois le snoARN et le pré-complexe assemblés, la dernière protéine cœur est alors recrutée. Comme pour les snoRNP à boîtes C/D, le processus d'assemblage nécessite *in vivo* l'intervention de deux facteurs d'assemblage, Shq1/SHQ1 et Naf1/NAF1, dont la fonction principale semble être d'inhiber les particules immatures. De plus l'intervention du complexe (h)R2TP lors du remodelage de la pré-particule n'est pas non plus écartée.

Le recrutement du snoARN sur la protéine Cbf5/Dyskérine entraîne le relargage du facteur Shq1/SHQ1 qui recouvrait le domaine PUA et empêchait la reconnaissance de la boîte ACA du snoARN par la protéine Cbf5/Dyskérine (**Figure 44**, étape 2). Suite au recrutement du snoARN, la particule présente une activité catalytique, mais l'absence de la protéine Gar1/GAR1 l'empêche de catalyser efficacement les réactions (Li *et al.*, 2011). Comme décrit précédemment, le recrutement de la protéine Gar1/GAR1 sur la protéine Cbf5/Dyskérine nécessite le départ du facteur d'assemblage Naf1/NAF1 (**Figure 44**, étape 3). Le mécanisme qui permet l'échange entre ces deux protéines n'est pas encore identifié, toutefois il pourrait nécessiter un remodelage des protéines de la pré-particule et l'intervention du complexe (h)R2TP (**Figure 44**, étape 4). En effet, des interactions directes ont été mises en évidence entre la protéine Nhp2/NHP2 et la protéine Rsa1/NUFIP d'une part, entre les protéines AAA⁺ et la protéine Cbf5/Dyskérine d'autre part, ainsi qu'entre la protéine Cbf5/Dyskérine et la protéine Pih1/Pih1D1 (Boulon *et al.*, 2008 ; Machado-Pinilla *et al.*, 2012). De plus, des expériences d'ARN interférence dirigées contre les protéines TIP48 et TIP49 ainsi que l'inhibition de la protéine HSP90 par la geldanamycine ont montré une diminution du taux de snoARN à boîtes H/ACA (Machado-Pinilla *et al.*, 2012 ; Boulon *et al.*, 2008). Ainsi, il est possible que lors de l'échange entre la protéine Naf1/NAF1 et la protéine Gar1/GAR1, le complexe (h)R2TP soit recruté par l'intermédiaire de la protéine Rsa1/NUFIP afin de permettre aux protéines Rvb1/TIP49 et Rvb2/TIP48 d'interagir avec les protéines cœur pour former la particule mature (Machado-Pinilla *et al.*, 2012). De même, il semble que les protéines Rvb1/TIP49 et Rvb2/TIP48 puissent interagir avec la protéine Shq1/SHQ1. Ainsi, l'intervention du complexe (h)R2TP pour permettre le relargage de la protéine Shq1/SHQ1 et le recrutement du snoARN est également possible (Machado-Pinilla *et al.*, 2012).

Chapitre 2 : Les microARN

Les microARN sont des ARN non-codants simple-brin d'environ 22 nucléotides de long. Tout comme les snoARN, les microARN matures s'associent avec des protéines pour former une particule ribonucléoprotéique, le complexe RISC (« RNA-Induced Silencing Complex »), impliqué dans l'ARN interférence. Au sein du complexe RISC, les microARN s'apparient à un ARNm cible, entraînant l'inhibition de sa traduction (Huntzinger *et al.*, 2013), et en général, la dégradation de l'ARNm cible (Eichhorn *et al.*, 2014).

Le premier microARN, lin-4, a été initialement identifié en 1993 chez le vers nématode *C. elegans* (Lee *et al.*, 1993). D'autres microARN furent par la suite identifiés chez *C. elegans* puis dans d'autres organismes dont l'homme (Lagos-Quintana *et al.*, 2001 ; Lau *et al.*, 2001 ; Lee et Ambros, 2001). Aujourd'hui la dernière version (21) de la base de donnée miRBase, qui regroupe ces microARN (Griffiths-Jones *et al.*, 2006), compte des milliers de microARN dont un peu plus de 2800 pour l'homme.

I. Maturation des microARN

A. Organisation génétique des microARN

Les microARN sont transcrits sous la forme d'un précurseur, le pri-microARN. Chez les mammifères, l'organisation des gènes codant les pri-microARN est très variable. Ainsi, comme beaucoup d'autres ARNnc, 50% de ces gènes sont trouvés dans les introns d'autres gènes (Murchison et Hannon, 2004), les autres disposant de leur propre unité de transcription. Dans ce dernier cas, la très grande majorité sont malgré tout transcrits par l'ARN polymérase II, polyadénylés et une coiffe m⁷G est ajoutée à leur extrémité 5'. De rares pri-microARN ont été trouvés au sein d'exon de gènes codant des protéines (Rodriguez *et al.*, 2004). Enfin, dans environ 50 % des cas, les séquences des pri-microARN sont situées à une distance inférieure à 10 kb les unes des autres et sont groupées sous forme de « clusters » qui seront transcrits à partir d'un unique promoteur sous la forme d'un ARN polycistronique (pour revue Kim *et al.*, 2009).

B. Maturation canonique

La maturation canonique des microARN correspond à la première voie de biogenèse mise en évidence (**Figure 45**). Cette voie commence par la transcription d'un long précurseur primaire ou pri-microARN (Lee *et al.*, 2002). En général, chaque pri-microARN est composé d'une structure en tige-boucle d'environ 35 nucléotides de long et de deux extrémités 5' et 3' simple brin. La maturation du pri-microARN en un microARN simple-brin nécessite deux événements de clivage, l'un dans le noyau et l'autre dans le cytoplasme, et aboutit à la formation du complexe RISC. Comme pour l'assemblage des snoRNP, de nombreuses protéines vont intervenir dans ce processus sans être retrouvées dans la particule mature.

1. Maturation du pri-microARN par le microprocesseur

Après sa transcription, le pri-microARN est reconnu puis maturé en pré-microARN par le microprocesseur, un complexe protéique d'environ 650 kDa composé principalement de la protéine Drosha et de son cofacteur DGCR8 (« DiGeorge Critical Region 8 »). Au sein de ce complexe, le pri-microARN va être clivé au niveau de ses extrémités simple-brin par la ribonucléase de type III Drosha (Lee *et al.*, 2003) à environ 11 nucléotides de la jonction ARN double-brin/ARN simple-brin (Han *et al.*, 2006) et 22 nucléotides de la jonction de la boucle (Ma *et al.*, 2013) (**Figure 46**).

La protéine DGCR8 (« DiGeorge Syndrome Critical Region 8 ») a initialement été identifiée pour son lien dans une maladie génétique rare, le syndrome de DiGeorge (DGS). Les patients souffrant de ce syndrome, qui comprend notamment un déficit immunitaire sévère, une hypokaliémie, et un faciès anormal, présentent une délétion au niveau du chromosome 22 (région 22q11.2). Cette région contient de nombreux gènes, dont le gène codant DGCR8. Chez les souris, la délétion du gène orthologue, *Dgcr8*, entraîne des syndromes comparables au DGS. La protéine DGCR8 est ainsi suspectée d'être impliquée dans le développement du syndrome de DiGeorge (Shiohama *et al.*, 2003).

La protéine Drosha est une RNase de type III Mg²⁺ dépendante de 160 kDa qui a été initialement identifiée comme une protéine potentiellement impliquée dans la maturation des pré-ARN ribosomiaux (Wu *et al.*, 2000).

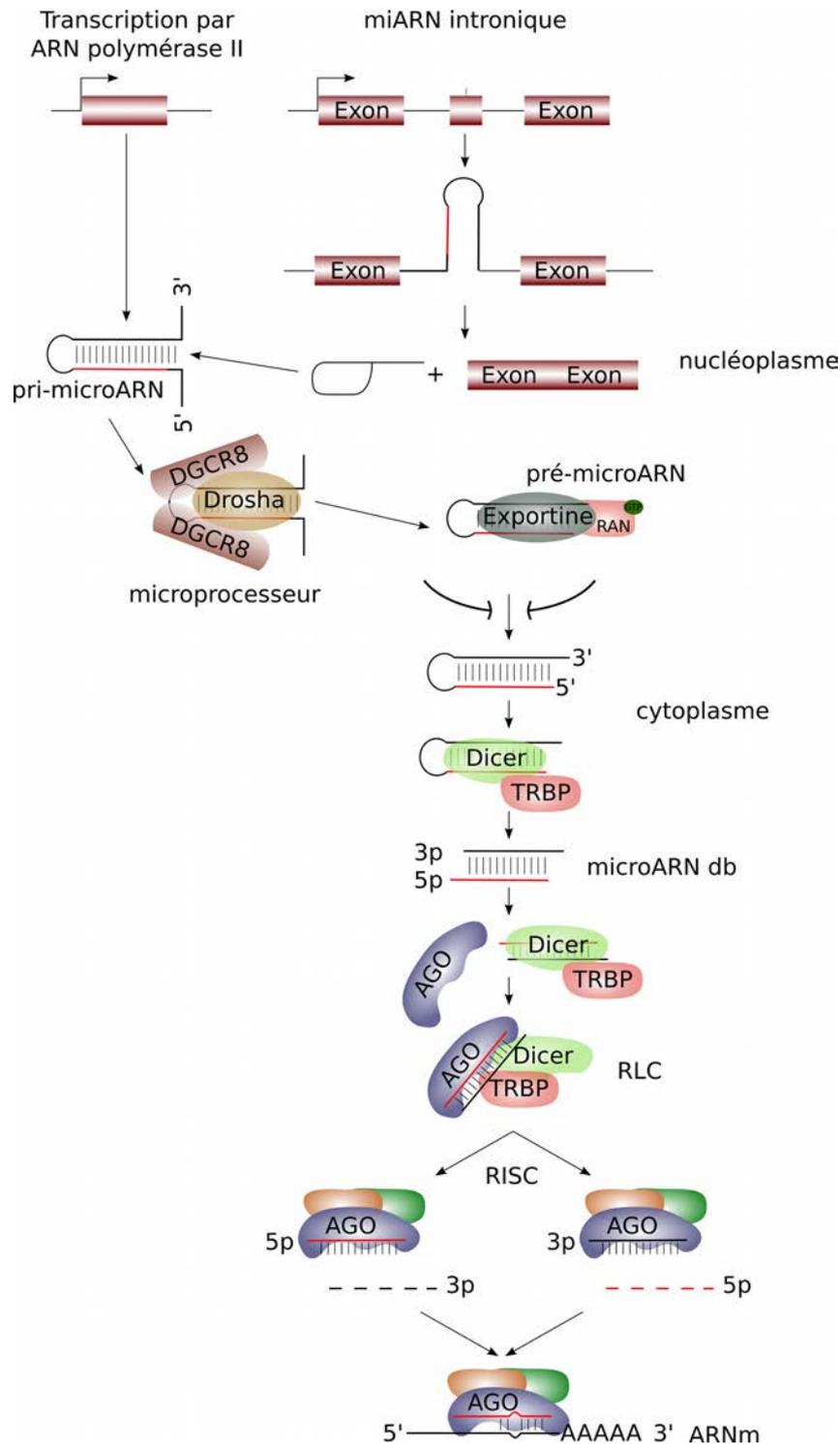


Figure 45: Maturation canonique des microARN (adapté de Winter *et al.*, 2009).

Le pri-microARN est maturé en pré-microARN par le microprocesseur notamment composé des protéines Drosha et DGCR8. Le pré-microARN est ensuite exporté dans le cytoplasme par le système RAN-GTP/Exportine-5. Une fois dans le cytoplasme, le pré-microARN est maturé en microARN double-brin par la protéine Dicer et l'un de ses cofacteurs, TRBP ou PACT. Le duplex est ensuite chargé sur une protéine Argonaute (1 à 4) puis le brin guide est intégré au sein du complexe RISC. Le brin passager est dégradé.

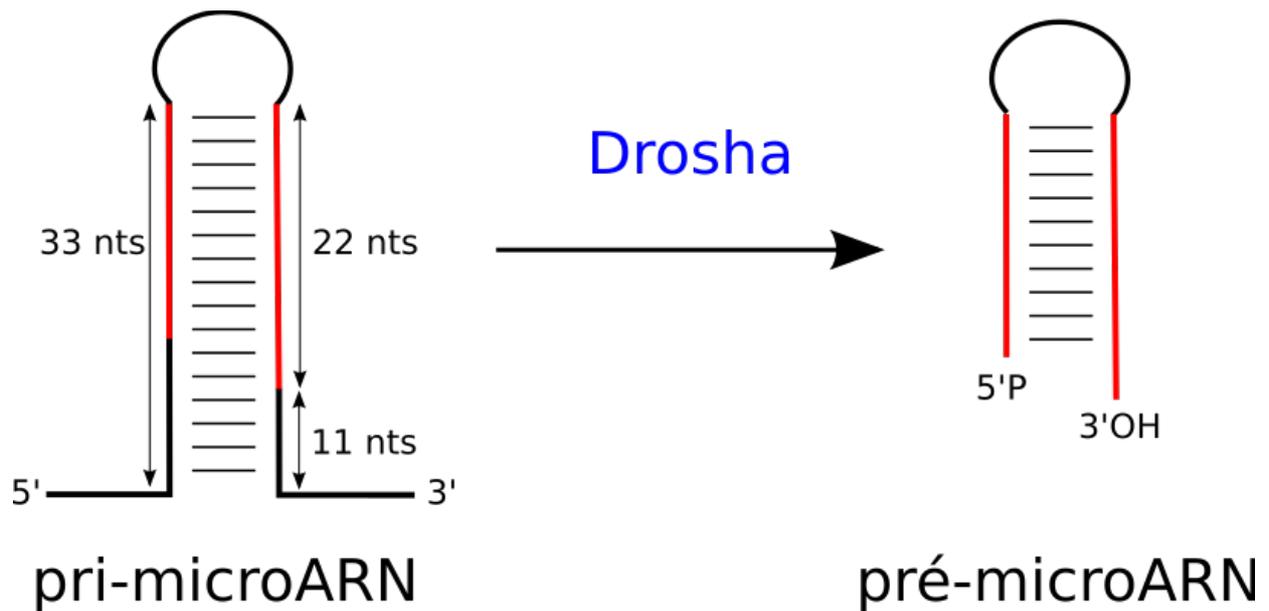


Figure 46: Schéma représentant un pri- et un pré-microARN.

Les pri-microARN sont composés d'une structure en tige-boucle de 33 nucléotides de long et de deux extrémités 5' et 3' simple brin.

Les pré-microARN sont issus maturés à partir des pri-microARN par l'enzyme Drosha. Celle-ci clive la tige du pri-microARN à 11 nucléotides de la jonction ARN double-brin/ARN simple-brin et à 22 nucléotides de la boucle. Le pré-microARN ainsi produit a une taille d'environ 70 nucléotides de long et contient 2 nucléotides non-appariés au niveau de l'extrémité 3' ainsi que des extrémités 5'-phosphate et 3'-OH.

Elle possède deux domaines RNase III (RIIIda et RIIIdb) impliqués dans le clivage du pri-microARN, un domaine C-terminal de liaison à l'ARN double-brin (dsRBD) caractéristique des RNases de type III de classe 2 (pour revue MacRae et Doudna, 2007), un domaine central (DC) indispensable pour son activité (Han *et al.*, 2004) relié au domaine RIIIda par l'intermédiaire d'un connecteur, et enfin d'une région N-terminale riche en proline (domaine PRR) et en arginine/sérine (domaine RS) (**Figure 47**). L'implication de Drosha dans la maturation canonique des microARN a été notamment démontrée par l'utilisation de siARN dirigés contre elle, entraînant une diminution du taux de microARN mature (Lee *et al.*, 2003).

Au sein du microprocesseur, la protéine Drosha est associée à 2 protéines DGCR8 (**Figure 47**). Ces interactions sont nécessaires à la reconnaissance précise du pri-microARN. En effet, bien que la protéine Drosha possède un domaine de fixation à l'ARN, il a été montré qu'elle avait une très faible affinité pour ce dernier (Nguyen *et al.*, 2015). De même, la protéine DGCR8 seule interagit avec l'ARN de manière non-spécifique (Roth *et al.*, 2013). Récemment, la structure tridimensionnelle de la protéine Drosha humaine en interaction avec une partie des deux protéines DGCR8 (**Figure 47**) a pu être résolue par diffraction des rayons X et a permis de mieux comprendre le mécanisme de reconnaissance et de maturation du pri-microARN par le microprocesseur (Kwon *et al.*, 2016). Cette structure montre que les deux protéines DGCR8 interagissent avec Drosha au niveau des domaines RIIID. De manière intéressante, l'interaction des deux protéines DGCR8 au niveau des domaines RIIID est asymétrique, et il a été montré que seule l'interaction entre DGCR8 et le domaine RIIIdb était essentielle pour la stabilité de Drosha (Kwon *et al.*, 2016). La structure montre également que les domaines RIIIda et RIIIdb forment un dimère intramoléculaire RIIID qui permet la formation du centre catalytique. Au niveau de ce centre, les domaines RIIIda et RIIIdb vont, respectivement, cliver les extrémités 3' et 5' du pri-microARN (Bernstein *et al.*, 2001). Ces domaines étant très conservés au sein des RNases de type III, Kwon *et al.* ont réalisé un modèle comprenant un pri-microARN à partir de la structure 3D d'une autre RNase de type III en complexe avec un ARN double-brin.

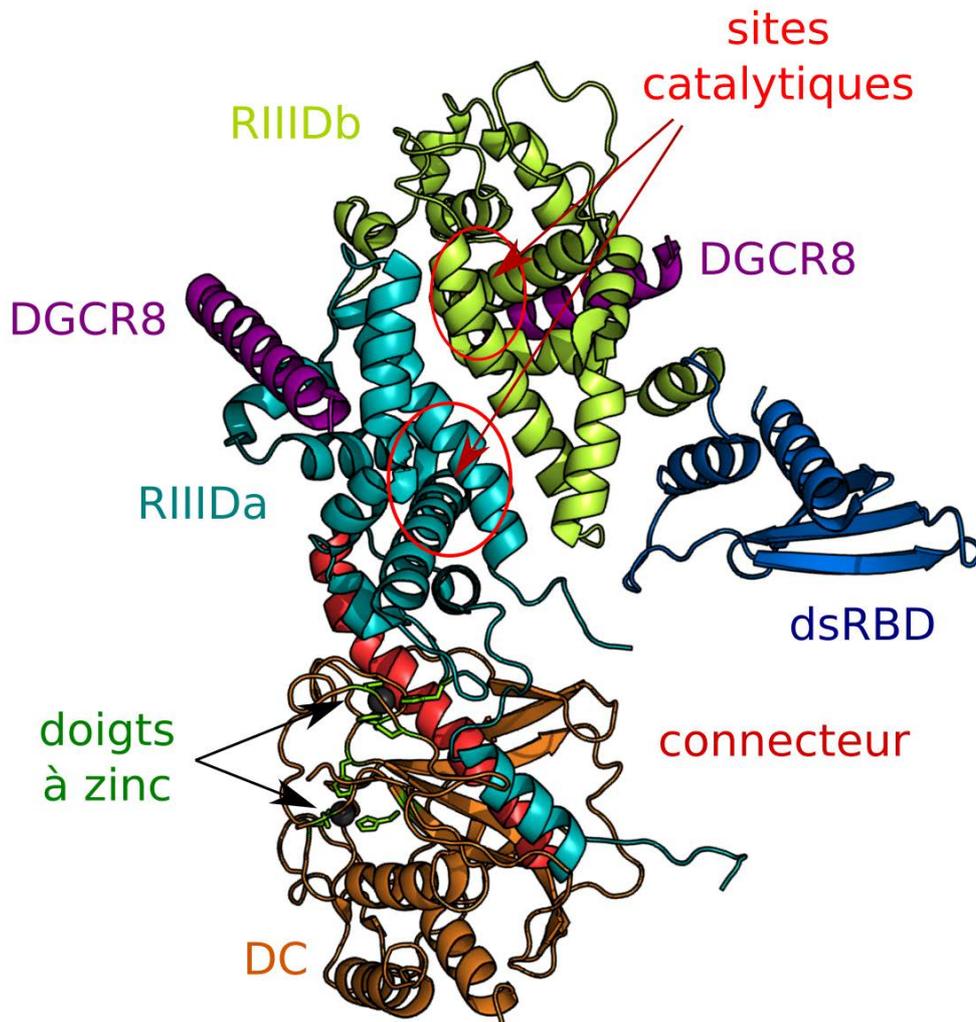


Figure 47: Structure 3D de la protéine Drosha en interaction avec le domaine C-terminal de la protéine DGCR8, résolue par diffraction des rayons X (code PDB : 5B16 ; Kwon *et al.*, 2016).

La protéine Drosha est une RNase de type III impliquée dans la maturation du pri-microARN en pré-microARN. Elle est composée de deux domaines RIIID (a et b) qui forment un dimère intramoléculaire qui permet la formation d'un centre catalytique au sein duquel les deux brins du pri-microARN seront clivés. La protéine Drosha contient également deux doigts à zinc, indispensables à la catalyse de la réaction d'hydrolyse, ainsi qu'un domaine central (DC). Enfin, un domaine d'interaction avec l'ARN double-brin est également présent.

Deux protéines DGCR8 interagissent avec Drosha de manière asymétrique afin de positionner le pri-microARN correctement au niveau du centre catalytique.

Les structures 3D des domaines PRR et RS de l'extrémité N-terminale, ainsi qu'une partie du domaine DC, n'ont pas été résolues.

Ce modèle (**Figure 48**) montre que le domaine central (DC) de Drosha se situerait plutôt au niveau des extrémités simple-brin du pri-microARN, ce qui permettrait au domaine RIIID de se positionner correctement. Le modèle, combiné à des données obtenues grâce à plusieurs variants du domaine DC, suggère que deux hélices du domaine RIIIDa, appelées MB et Bump, ainsi que 2 doigts à zinc (ZnF1 et ZnF2), seraient nécessaires pour la reconnaissance du pri-microARN et pour la stabilité de la structure de Drosha, respectivement (Kwon *et al.*, 2016). Enfin, l'orientation des domaines C-terminaux de DGCR8 (CTT) au niveau de la protéine Drosha suggère que ces deux protéines s'orienteraient parallèlement l'une par rapport à l'autre (**Figure 48**).

Bien que le complexe Drosha-DGCR8 soit suffisant pour maturer les pri-microARN en pré-microARN, il semble qu'*in vivo* le microprocesseur contienne d'autres facteurs protéiques (Gregory *et al.*, 2004). En effet, il a ainsi été montré *in vivo* que la protéine Drosha s'associait avec de nombreuses protéines, dont des hélicases et des hnRNP.

Suite à l'action du microprocesseur, le pré-microARN résultant a une taille d'environ 70 nucléotides et possède les caractéristiques d'un produit issu du clivage par une RNase de type III (Lee *et al.*, 2003), c'est-à-dire des extrémités 5'-Phosphate, 3'-OH et 2 nucléotides non appariés au niveau de l'extrémité 3' (**Figure 46**).

2. Export du pré-microARN dans le cytoplasme

Le pré-microARN est ensuite exporté dans le cytoplasme (**Figure 45**) par l'intermédiaire du système Exportine-5/RAN-GTP (« RAS-related Nuclear protein ») (Yi *et al.*, 2003). L'Exportine-5 est un transporteur nucléo-cytoplasmique qui, en présence de RAN-GTP, va reconnaître des ARN contenant une structure double-brin (Gwizdek *et al.*, 2003). L'interaction avec cet ARN est indépendant de la séquence ou de la présence de la boucle au niveau du pré-microARN. En revanche, la présence des 2 nucléotides non-appariés au niveau de l'extrémité 3' est primordiale pour l'interaction du pré-microARN avec la protéine (Zeng et Cullen, 2004). De même, il semble y avoir un contrôle au niveau de la taille de l'ARN puisque des ARN structurés en tige-boucle de moins de 16 nucléotides ne sont pas exportés dans le cytoplasme (Zeng et Cullen, 2004). Dans le noyau, l'Exportine-5 (ou XPO5) forme un complexe ternaire avec la protéine RAN, qui est liée au GTP, et le pré-microARN qu'elle protège de la dégradation par des nucléases (Zeng et Cullen, 2004).

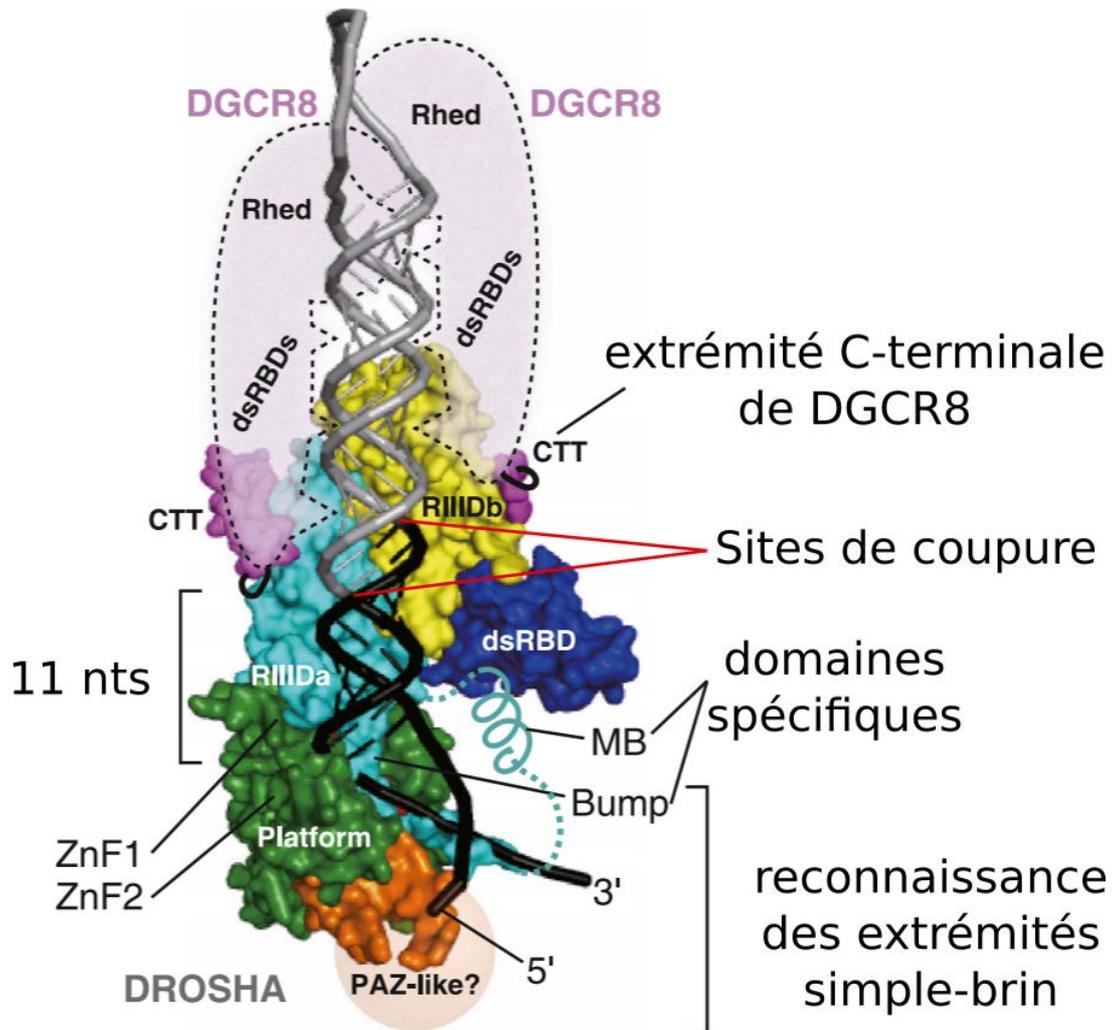


Figure 48: Modèle d'interaction du microprocesseur avec le pri-microARN (adapté de Kwon *et al.*, 2016).

Ce modèle montre que la protéine Drosha reconnaît les extrémités simple-brin du pri-microARN grâce à son domaine DC et à deux hélices spécifiques, MB et Bump, qui vont permettre de positionner le pri-microARN correctement au niveau du centre catalytique. Ce positionnement est aidé par deux protéines DGCR8 qui permettent le bon positionnement Drosha sur le pri-microARN.

Des expériences de dynamique moléculaire suggèrent que la formation de ce complexe est permise par un changement de conformation à la fois au niveau de l'Exportine-5, suite à son interaction avec la protéine RAN-GTP, et au niveau des deux nucléotides non-appariés présents à l'extrémité 3'OH du pré-microARN (Wang *et al.*, 2011). Le complexe est ensuite transloqué dans le cytoplasme, via les pores nucléaires. Puis le GTP est hydrolysé en GDP ce qui induit un changement de conformation de la protéine RAN, la dissociation du complexe et la libération du pré-microARN.

La structure du complexe ternaire a été résolue par diffraction des rayons X (Okada *et al.*, 2009). L'Exportine-5 contient 20 motifs HEAT (« Huntingtin, Elongation factor 3, protein phosphatase 2A, TOR1 ») qui forment une structure très flexible en forme de U contenant un tunnel à l'intérieur duquel vient se positionner l'ARN (**Figure 49**). Ce dernier est stabilisé grâce à la présence de résidus basiques à l'intérieur du tunnel. Les domaines HEAT 12 à 15 forment une enclave dans laquelle viennent s'insérer les deux nucléotides non-appariés de l'extrémité 3' de l'ARN (**Figure 50**). Enfin, la protéine RAN-GTP interagit avec l'Exportine-5 avec l'une des branches du U, au niveau des motifs HEAT 1 à 7 (**Figure 50**). L'export du pré-microARN dans le cytoplasme correspond à l'étape limitante du processus de maturation des microARN (Yi *et al.*, 2005), mais bien que le complexe Exportine-5/RAN-GTP semble être le principal système d'export nucléaire pour les pré-microARN, des expériences d'ARN interférence contre l'Exportine-5 n'ont montré une diminution que pour 30% des microARN matures, ce qui suggère l'existence d'un ou plusieurs autres systèmes d'export nucléaire (Kim *et al.*, 2016).

3. Maturation du pré-microARN

a. Dicer, une RNase de type III de classe 3

Une fois dans le cytoplasme le pré-microARN subit une nouvelle étape de maturation (**Figure 45**) au cours de laquelle la boucle sera clivée. Cette étape permet d'obtenir un microARN double-brin d'environ 22 nucléotides de long ayant 2 nucléotides non-appariés au niveau de chacune de ses extrémités (**Figure 51**). La maturation du pré-microARN est effectuée par un second complexe qui contient, entre autre, la protéine Dicer.

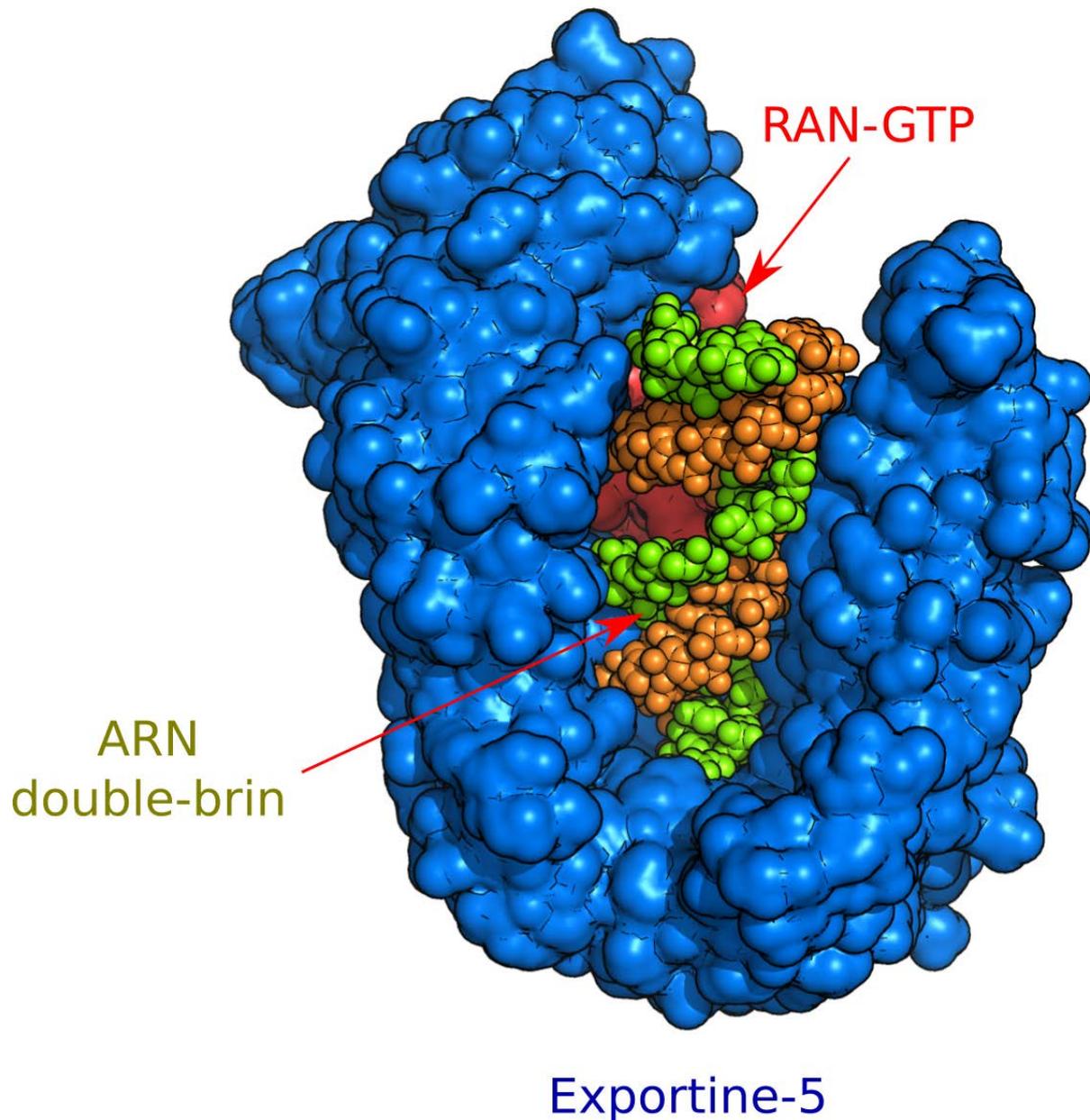


Figure 49: Surface de structure 3D du complexe ternaire RAN-GTP/Exportine-5/ARN double-brin résolue par diffraction des rayons X (code PDB : 3A6P).

L'exportine-5 est un transporteur cytoplasmique qui permet l'export du pré-microARN dans le cytoplasme. L'Exportine-5 interagit avec la protéine RAN, elle-même liée au GTP, et avec le pré-microARN. Au sein de ce complexe, le pré-microARN est totalement protégé des nucléases.

Une fois dans le cytoplasme, le GTP est hydrolysé en GDP ce qui entraîne un changement de conformation de l'Exportine-5 et la libération du pré-microARN.

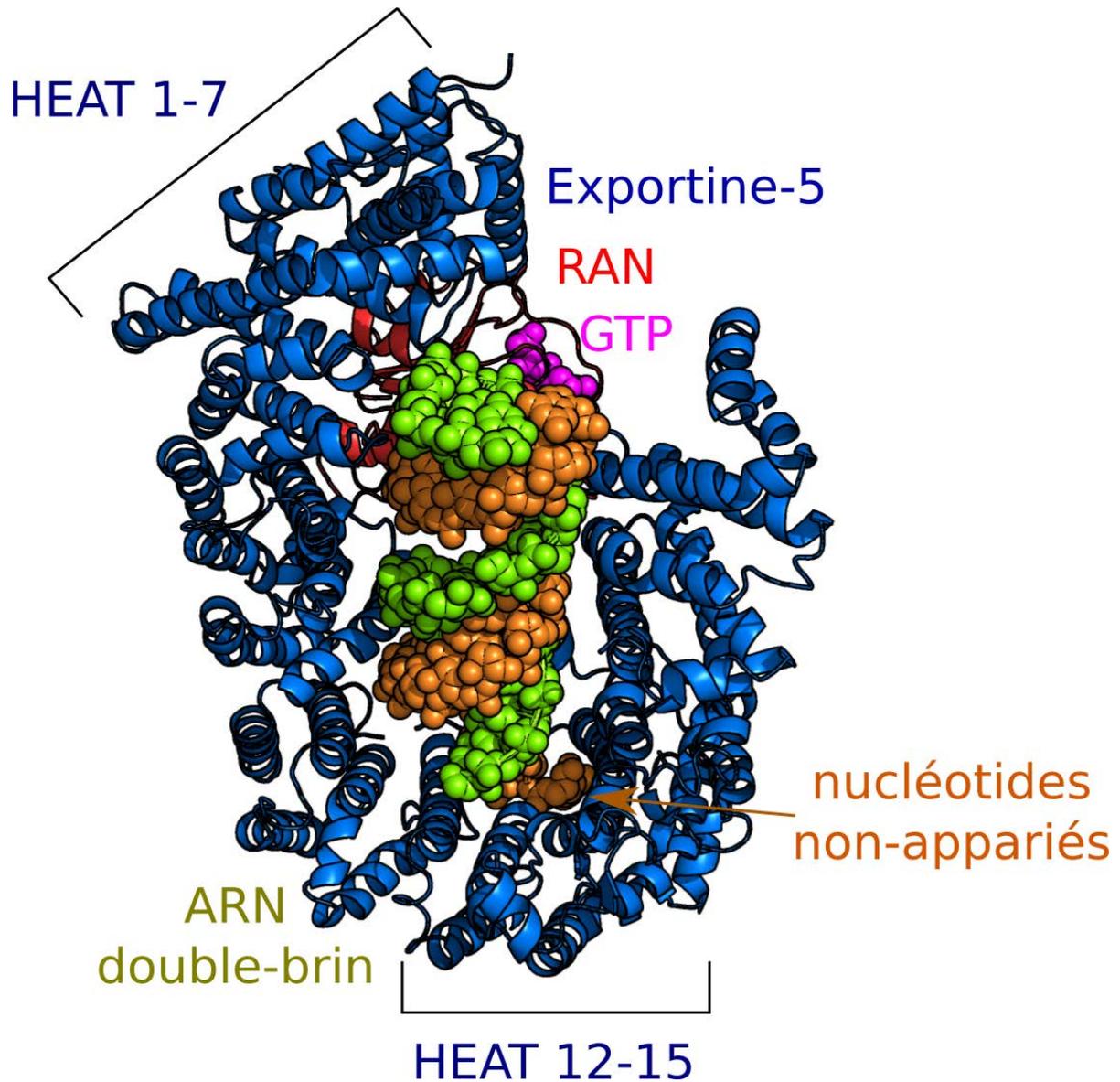


Figure 50: Structure 3D du complexe ternaire RAN-GTP/Exportine-5/ARN double-brin résolue par diffraction des rayons X (code PDB : 3A6P).

L'Exportine-5 contient 20 motifs HEAT qui forment un U au sein duquel le pré-microARN vient se positionner. Les nucléotides non-appariés au niveau de l'extrémité 3' du pré-microARN sont stabilisés via les domaines HEAT 12 à 15. La protéine RAN-GTP interagit avec l'Exportine-5 au niveau des domaines HEAT 1 à 7.

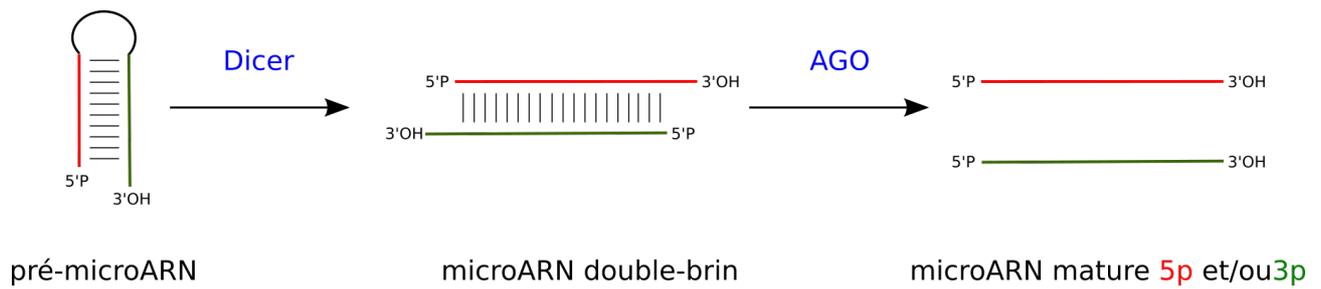


Figure 51: Schéma représentant les étapes permettant d'obtenir un microARN mature à partir du pré-microARN.

Une fois dans le cytoplasme, la boucle du pré-microARN est clivée par la protéine Dicer pour former un microARN double-brin. L'un des brins, le brin guide (5p ici) du microARN sera ensuite intégré au sein d'une protéine AGO (voir ci-après) pour former une particule RISC mature. L'autre brin, ou brin passager (3p ici), sera dégradé dans la majorité des cas.

La protéine Dicer étant une RNase de type III, le microARN double-brin aura une extrémité 3'-OH débordante avec 2 nucléotides non-appariés et une extrémité 5'-phosphate.

Dicer est une RNase de type III Mg^{2+} dépendante d'environ 220 kDa (Macrae *et al.*, 2006). Tout comme Drosha, elle possède deux domaines de type RNase III, RIIIa et RIIIb, responsables de l'hydrolyse du pré-microARN, ainsi qu'un domaine C-terminal de liaison à l'ARN double-brin (dsRBD). Elle possède en plus au niveau de son extrémité N-terminale, un domaine DexD ou HEL1, un domaine hélicase (HELICc) composé de deux sous-domaines HEL2i et HEL2, et un domaine DUF283 (« Domain of Unknow Function »). Ces domaines sont séparés des domaines RIIIa/b par un domaine PAZ (Piwi/Argonaute/Zwille) (Lau *et al.*, 2009), impliqué dans la reconnaissance du pré-microARN (**Figure 52**). De par ses nombreux domaines supplémentaires, la protéine Dicer a été classée dans la classe 3 des RNases de type III.

La structure de la protéine humaine n'a pas encore été résolue. Toutefois un mécanisme de reconnaissance du substrat et de catalyse a pu être proposé, à partir d'études biochimiques et structurales, notamment de microscopie électronique (Lau *et al.*, 2009 ; Wang *et al.*, 2009 ; Lau *et al.*, 2012). Combinées, ces données ont permis de montrer que Dicer se structurait en forme de « L » (**Figure 53**) : le domaine PAZ (divisé ici en deux sous-domaines PAZ et Platform) y serait inséré au niveau de la tête et les domaines hélicases DexD et HELICc au niveau de la base. Enfin les domaines RNase III (RIIIa/b) seraient situés au centre du L.

Tout comme pour Drosha, les domaines RNase IIIa et b s'assemblent en un dimère intramoléculaire afin de former un unique centre catalytique RIII (Zhang *et al.*, 2004) au sein duquel le domaine RNase IIIa hydrolysera le brin 3' et le domaine RNase IIIb le brin 5' du pré-microARN (Macrae *et al.*, 2006). Chez *Giardia intestinalis*, la structure de Dicer (Macrae *et al.*, 2006) révèle que les domaines PAZ et RIII, qui sont séparés par un domaine connecteur, se trouvent à environ 65 Å l'un de l'autre, ce qui correspond à la longueur du microARN produit (environ 26 nucléotides de long chez cet organisme). De plus, la délétion du domaine PAZ de la protéine Dicer chez *G. intertinalis* entraîne la perte de cette spécificité de longueur au niveau du produit (MacRae *et al.*, 2007). Chez les mammifères, des expériences de microscopie électronique ont montré que le domaine PAZ était plus proche du centre catalytique, ce qui expliquerait la différence de longueur des ARN après coupure (Lau *et al.*, 2012).



Figure 52: Organisation des domaines de la protéine Dicer humaine (Lau *et al.*, 2009).

La protéine Dicer contient de nombreux domaines, dont des domaines hélicase DEXD (ou HEL1) et HELICc (composé des sous-domaines HEL2 et HEL2i), un domaine d'interaction avec l'ARN (dsRBD), deux domaines de type RNase III (RIIIa et RIIIb) qui peuvent former un dimère intramoléculaire et un domaine PAZ qui reconnaît l'extrémité 5'-phosphate libre du pré-microARN.

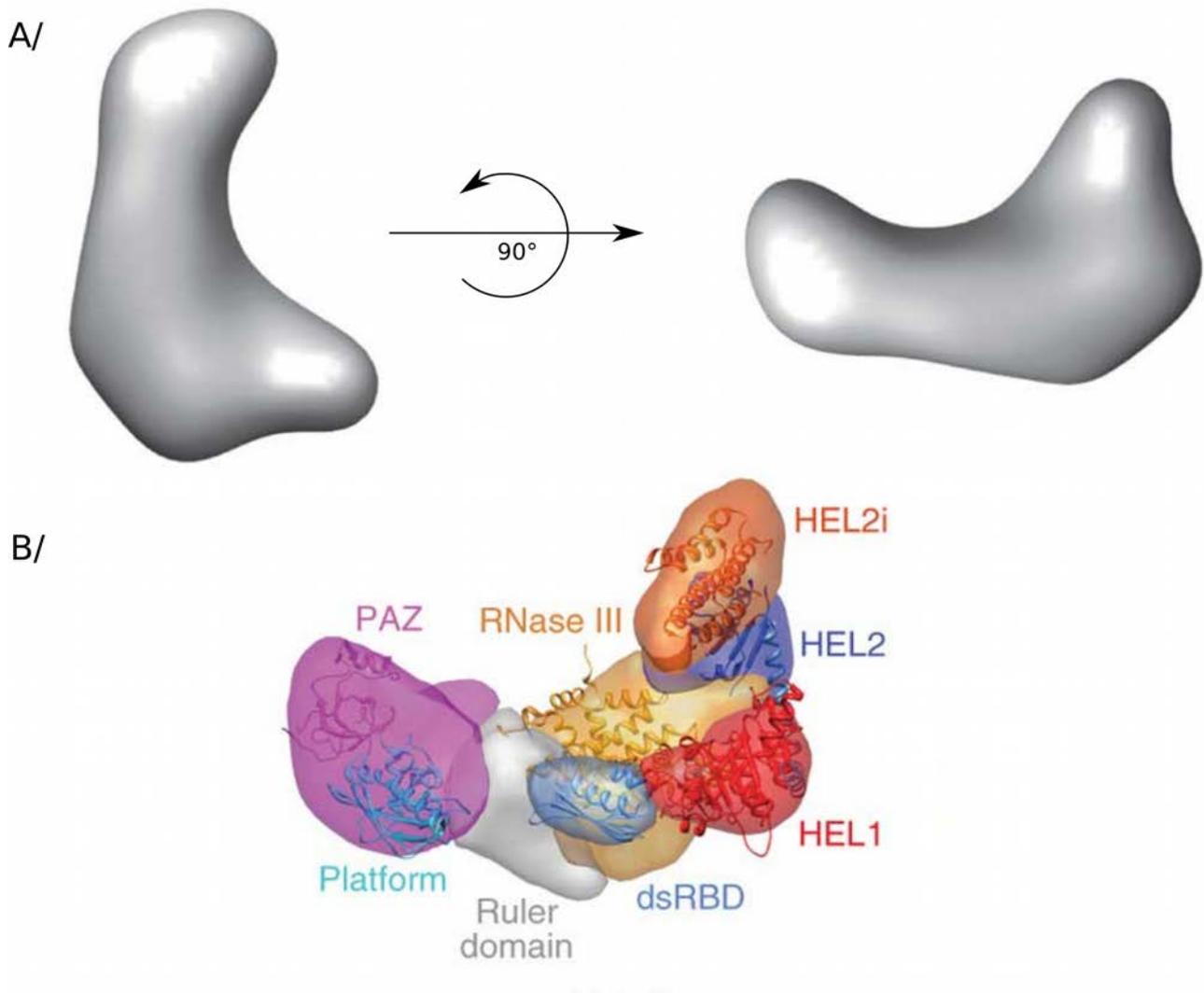


Figure 53: (A) Structure 3D de Dicer résolue à basse résolution par microscopie électronique et (B) reconstitution de la structure tertiaire de Dicer à partir des structures 3D des domaines d'homologues de Dicer résolus par diffraction des rayons X (Lau *et al.*, 2012).

La protéine Dicer forme ici un L dont la base est composée de domaines hélicase, la tête du domaine PAZ et le corps des domaines RNase. Le pré-microARN vient se positionner au niveau des domaines PAZ et hélicases.

Le clivage du pré-microARN par Dicer nécessite, chez les mammifères et la drosophile, la présence d'un groupement phosphate au niveau de l'extrémité 5' de l'ARN. Ce groupement est reconnu par le domaine PAZ au niveau d'une poche formée par des résidus basiques qui vont stabiliser l'ARN par le biais de liaisons hydrogènes (Park *et al.*, 2011). Toutefois, il semble que dans certains cas le domaine PAZ reconnaisse plutôt les deux nucléotides non-appariés au niveau de l'extrémité 3' du pré-microARN, (Gu *et al.*, 2012).

Chez les invertébrés, les domaines hélicases fixent et hydrolysent l'ATP nécessaire à la maturation des siARN à partir de longs ARN double-brin (Zamore *et al.*, 2000). Bien que ces domaines soient conservés chez les mammifères, des expériences de mutation au niveau des domaines Walker A ont montré que l'efficacité de Dicer n'est pas affectée par l'absence d'ATP (Zhang *et al.*, 2002). De plus, la délétion de ces domaines hélicases a entraîné l'augmentation de l'activité de Dicer *in vitro* et *in vivo* pour la production de siARN (Ma *et al.*, 2008 ; Flemer *et al.*, 2013 ; Kennedy *et al.*, 2015). Des données supplémentaires suggèrent que chez les mammifères les domaines hélicases auraient un autre rôle. En effet, les données de microscopie électronique montrent que la protéine Dicer aurait trois états de conformation (Taylor *et al.*, 2013) en fonction de la présence et du type de substrat (pré-microARN ou long ARN double-brin). Ainsi, en l'absence de substrat, Dicer serait dans un état de repos (« canonical state »), puis, en présence d'un substrat, alternerait entre un état *ouvert* ou *fermé*. La conformation ouverte permet l'hydrolyse des pré-microARN qui se fixent au niveau du domaine PAZ et des hélicases, le long du centre catalytique. La conformation fermée n'a elle été observée que lorsque le substrat est un ARN double-brin parfait (pré-siARN). Dans ce dernier cas, le substrat, qui interagit différemment au niveau des domaines hélicases, se retrouve plus loin des domaines RIII. Cette différence d'orientation du pré-microARN et du pré-siARN vis-à-vis du domaine catalytique de Dicer semble expliquer les différences de réactivités observées pour ces deux ARN, et suggère un rôle pour ces domaines hélicase dans le contrôle de la spécificité du substrat de Dicer.

Bien que Dicer possède un domaine de liaison à l'ARN double-brin (dsRBD), le recrutement du pré-microARN par Dicer est facilité par la formation au préalable d'un complexe entre l'ARN et l'un des cofacteurs de Dicer (Lee et Doudna, 2012), les protéines TRBP (aussi nommée TARBP2) et PACT (Chendrimada *et al.*, 2005 ; Lee *et al.*, 2006).

b. Les protéines TRBP et PACT

La protéine TRBP/TARBP2 (« Trans-activation-responsive RNA-Binding Protein 2 ») a été initialement identifiée comme un partenaire de l'ARN TAR du virus VIH (Gatignol *et al.*, 1991). Ainsi, il a été montré qu'une sur-expression de TRBP stimulait la production de virus VIH (Battisti *et al.*, 2003), d'une part via l'inhibition de la voie PKR (« Protein Kinase R ») et d'autre part en levant l'inhibition traductionnelle causée par la structure de l'ARN TAR (Benkirane *et al.*, 1997 ; Daher *et al.*, 2001 ; Dorin *et al.*, 2003). La protéine TRBP est un homodimère dont chaque monomère contient deux domaines de fixation à l'ARN double-brins DRBM (« Double-stranded RNA Binding Motif ») qui comportent un feuillet de trois brins β flanqué par deux hélices α , et un troisième domaine DRBM3, appelé domaine MeDiPAL (« Merlin, Dicer, PKR activator, PACT Liaison »). Bien que proche des autres domaines DRBM, ce domaine est un domaine d'interaction protéine-protéine. Il n'est toutefois pas nécessaire pour la dimérisation de TRBP qui s'effectue par l'intermédiaire des domaines DRBM1 et 2 (Laraki *et al.*, 2008).

Le domaine MeDiPAL interagit notamment avec la protéine PKR, une kinase à sérine/thréonine qui joue un rôle central dans la défense immunitaire innée. En effet, au cours d'une infection virale, la présence d'ARN viral peut être détectée par la protéine PKR par l'intermédiaire de ses domaines DRBM. L'interaction entre la protéine et l'ARN entraîne l'homodimérisation de la protéine et son activation par auto-phosphorylation. Une fois activée, la protéine PKR va phosphoryler le facteur d'initiation de la traduction eIF2 α , ce qui conduit à un blocage de la traduction et par conséquent à un ralentissement de l'infection virale (pour revue García *et al.*, 2006). Les protéines PKR et TRBP peuvent former un hétérodimère par l'intermédiaire de leurs domaines DRBM respectifs. Cette interaction a pour conséquence d'inhiber l'activation de la protéine PKR, probablement en bloquant son homodimérisation (Daher *et al.*, 2009 ; Gupta *et al.*, 2003).

La protéine PACT (« Protein Activator of PKR ») contient également 2 domaines DRBM et un domaine DRBM C-terminal MeDiPAL qui est impliqué dans l'interaction avec d'autres protéines. Cette homologie avec TRBP s'explique par le fait que les gènes codant ces deux protéines sont deux paralogues ayant divergé suite à la duplication d'un gène ancestral commun (Murphy *et al.*, 2008). Tout comme TRBP, la protéine PACT interagit avec la protéine PKR. Contrairement à son homologue, il a été montré que cette interaction augmentait la phosphorylation de la protéine PKR, du facteur eIF2 α , et donc inhibait l'initiation de la traduction (Patel et Sen, 1998). L'activation de la

protéine PKR nécessite toutefois la phosphorylation préalable de la protéine PACT, qui fait généralement suite à divers stress, notamment oxydants (Patel *et al.*, 2000 ; Peters *et al.*, 2006). En outre, il a été montré que l'état d'activation de PACT pouvait être modulé par la protéine TRBP avec laquelle elle interagit (Kok *et al.*, 2007 ; Daher *et al.*, 2009). Ainsi, en l'absence de stress, les protéines forment un hétérodimère via leurs trois domaines DRBM respectifs ce qui empêche l'activation de la protéine PKR par la protéine PACT. Suite à un stress, la sérine 287 de PACT est phosphorylée, ce qui conduit à la libération de PACT et à l'activation de la protéine PKR (Singh *et al.*, 2011).

Lors de la biogenèse des microARN, les protéines TRBP et PACT vont interagir, via leurs domaines DRBM 1 et 2 (sous forme de monomère) avec les pré-microARN qui seront ensuite recrutés par Dicer. L'interaction avec Dicer est réalisée par l'intermédiaire des extrémités C-terminales des domaines MeDiPAL de TRBP et PACT (aussi appelée domaine C4) et au niveau des domaines hélicase de Dicer (Daniels *et al.*, 2009). Comme indiqué précédemment, ces domaines hélicase seraient impliqués dans le contrôle de la spécificité du substrat, qui dépend ainsi du partenaire de Dicer. En effet, il a été montré *in vitro* que PACT n'était impliquée que dans la maturation des pré-microARN alors que TRBP pouvait intervenir à la fois dans la maturation des pré-microARN et des pré-siARN (Lee *et al.*, 2013), même si dans le cas de certains pré-microARN, comme le pré-miR-31, il a été montré que TRBP pouvait au contraire inhiber leur maturation (Lee et Doudna, 2012). Cette spécificité de substrat n'est pas encore très bien comprise, toutefois elle semble être liée à la fois à la manière dont l'ARN interagit avec les deux domaines DRBM N-terminaux de TRBP et PACT, à la manière dont le substrat est ensuite orienté par rapport à Dicer et également à des changements de conformation induits au niveau de Dicer suite à son interaction avec le pré-microARN ou le pré-siARN (**Figure 54**) (Lee *et al.*, 2013). En outre, la résolution de la structure tridimensionnelle de l'interaction entre le dernier domaine DRBM de TRBP et le domaine hélicase de Dicer par diffraction des rayons X a permis de montrer que les interfaces d'interaction Dicer-TRBP et Dicer-PACT étaient conservées et donc de confirmer que ces interactions étaient mutuellement exclusives au niveau de Dicer (Wilson *et al.*, 2015). Enfin, il a également été montré que suivant le partenaire de Dicer, le microARN généré n'avait pas forcément la même extrémité (voir le paragraphe I. C. 5 de ce chapitre).

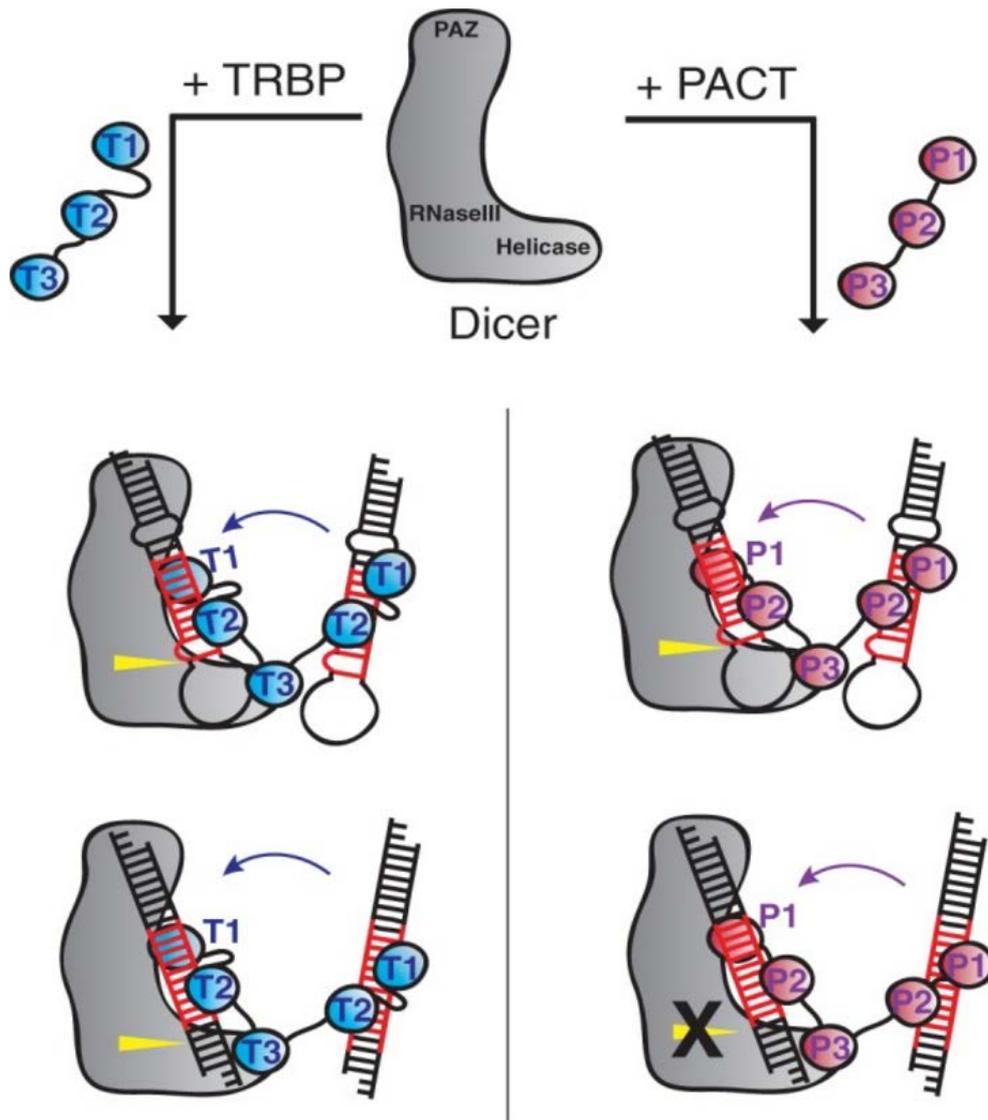


Figure 54: Rôles de TRBP et PACT dans la biogenèse des microARN et des siARN (Lee *et al.*, 2013).

La protéine Dicer a besoin d'un cofacteur afin de maturer correctement son substrat. Ce dernier interagit avec l'ARN et permet le bon positionnement de ce dernier au niveau de Dicer. Il existe deux cofacteurs connus pour Dicer, TRBP et PACT. Bien qu'ils soient homologues, il semble qu'ils ne partagent pas les mêmes fonctions. Entre autre, il a été montré que contrairement à TRBP, la protéine PACT ne participe pas à la maturation des longs ARN double-brin.

4. Formation de la microRNP mature

a. Le RLC

i. Généralités

L'obtention d'une microRNP mature nécessite l'élimination de l'un des deux brins du microARN double-brin nouvellement formé (brin passager) et le recrutement du brin restant (brin guide) au sein du complexe RISC (**Figure 45**). Ce complexe est constitué de plusieurs protéines dont une protéine cœur qui appartient à la famille des protéines Argonautes. Cette protéine est initialement recrutée avec les protéines Dicer et TRBP/PACT pour former le RLC (« RISC-Loading Complex ») (MacRae *et al.*, 2008). La structure du RLC a pu être résolue à basse résolution par microscopie électronique et montre que la protéine AGO interagit probablement avec le domaine C-terminal de Dicer (Wang *et al.*, 2009).

Il a aussi été montré que TRBP faciliterait le recrutement de la protéine AGO (Chendrimada *et al.*, 2005). Une fois la protéine AGO recrutée, le microARN double-brin est dans un premier temps transféré sur la protéine AGO (Chendrimada *et al.*, 2005) puis dans une seconde étape, les deux brins du duplex ARN sont séparés, le brin guide est sélectionné et le brin passager est dégradé (voir ci-après). Contrairement à la première étape, la sélection du brin guide est indépendante des protéines Dicer et TRBP/PACT (Betancur et Tomari, 2012 ; Suzuki *et al.*, 2015).

ii. Les protéines Argonautes

Il existe, chez les mammifères, 4 protéines AGO homologues (AGO1 à AGO4) très conservées. Elles comportent toutes un domaine N, un domaine PAZ (« PIWI/Argonaute/Zwille »), un domaine MID (« MIDdle ») et un domaine PIWI (« P-element-Induced Whlmpy testes »), ainsi que deux domaines de liaison L1 et L2, situés respectivement entre le domaine N-terminal et le domaine PAZ et entre les domaines PAZ et MID (**Figure 55**). La protéine AGO interagit avec le duplex ARN par l'intermédiaire des domaines MID et PAZ qui reconnaissent respectivement l'extrémité 5' et 3' du futur brin guide (Ma *et al.*, 2004 ; Frank *et al.*, 2010). Le choix de ce brin est réalisé en fonction des nucléotides présents aux niveaux des extrémités 5' du duplex ARN et la stabilité thermodynamique de ces extrémités (Schwarz *et al.*, 2003). Ainsi, il a été montré que le domaine MID interagissait préférentiellement avec le brin contenant un uracile ou une adénine au niveau de son extrémité 5' (Frank *et al.*, 2010). De même, il a été récemment montré que le domaine MID interagissait préférentiellement avec l'extrémité 5' la moins stable par la réalisation de liaisons ioniques avec les phosphates des nucléotides 2 à 4 du futur brin guide (Suzuki *et al.*, 2015).

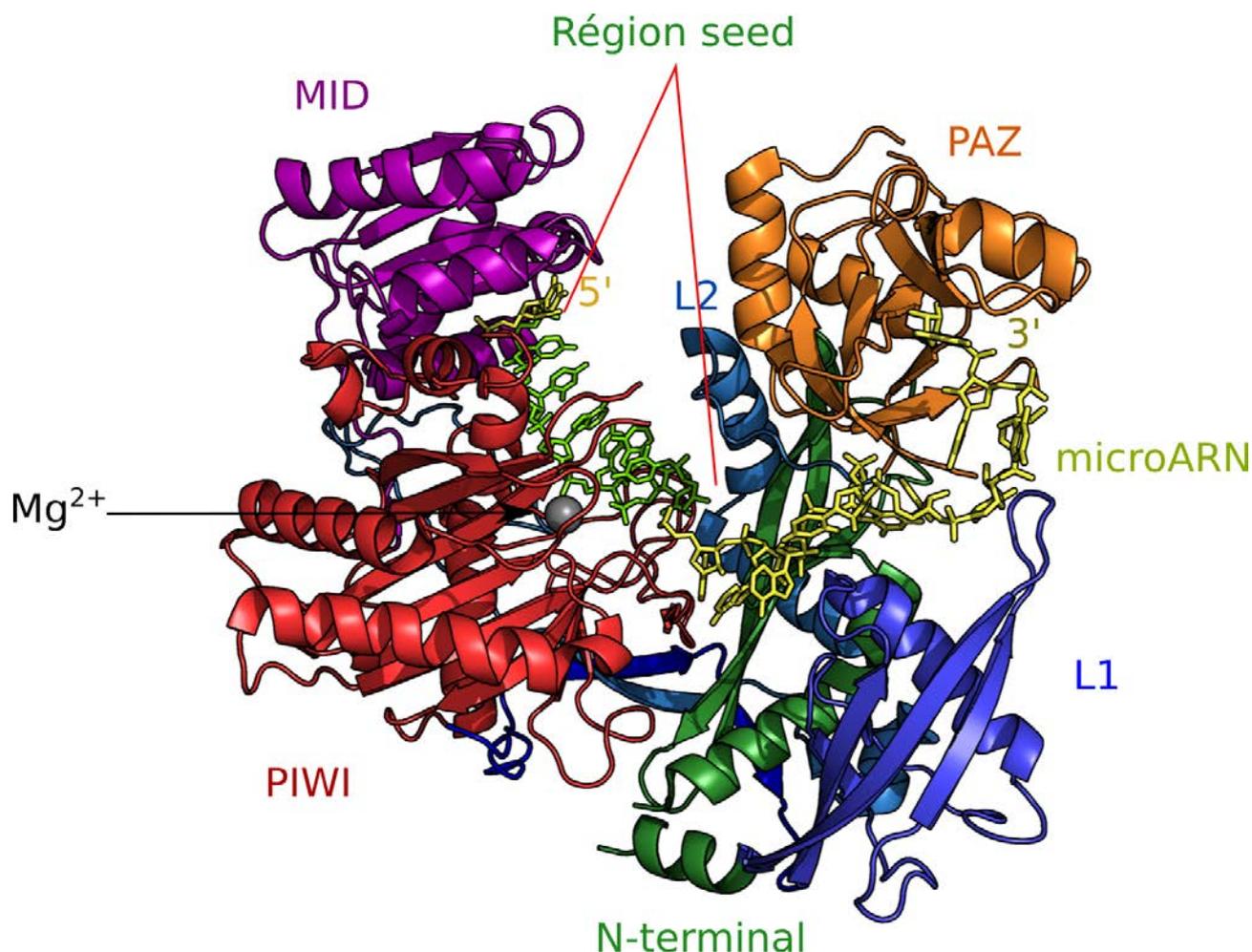


Figure 55: Structure 3D de la protéine AGO2 humaine en complexe avec un microARN de 21 nucléotides de long (code PDB : 4W5N).

Les protéines Argonautes comportent toutes un domaine N-terminal, un domaine PAZ, un domaine MID et un domaine PIWI, ainsi que deux domaines de liaison L1 et L2, situés respectivement entre le domaine N-terminal et le domaine PAZ et entre les domaines PAZ et MID. Les extrémités 5' et 3' du brin guide du duplex microARN interagissent respectivement avec les domaines MID et PAZ. Le domaine N-terminal facilite la séparation des brins guide et passager.

Le brin guide est alors positionné de telle sorte que les bases de ses nucléotides 2 à 7, ou région « seed », soient dirigées vers le solvant. Cette région est alors disponible pour interagir avec un ARN cible par l'intermédiaire d'interactions de type Watson-Crick.

Le domaine PIWI possède, uniquement chez la protéine AGO2, une activité de type RNase H. Cette activité n'est possible que si la position de l'ion Mg^{2+} est modifiée lors de la formation d'un duplex parfait entre le brin guide et l'ARN cible (ou entre le brin guide et le brin passager). L'activité RNase H est également conservée au sein de la protéine AGO3 mais les différences présentes au sein des domaines N-terminaux de ces deux protéines sont suffisantes pour inhiber cette activité au sein de la protéine AGO3.

La structure des nucléotides 8 à 11 n'a pas pu être résolue (Schirle *et al.*, 2014)

Toutefois, dans le cas où les deux brins auraient des extrémités 5' avec des stabilités trop proches, il a été montré que chaque brin pouvait former une particule RISC différente. La particule qui contient le microARN issus respectivement de l'extrémité 5' ou 3' du pré-microARN est alors noté -5p et -3p (**Figure 45**). De plus, plusieurs études ont montré que dans certains cas le brin « passager » n'était pas forcément dégradé et pouvait alors également former une particule RISC (Kuchenbauer *et al.*, 2011 ; Okamura *et al.*, 2008 ; Martin *et al.*, 2014).

iii. Rôle de la protéine HSP90 dans le recrutement du microARN

In vitro, il a été possible de reconstituer un RLC fonctionnel à partir des protéines purifiées AGO2, Dicer et TRBP (MacRae *et al.*, 2008). Toutefois, il semble qu'*in vivo* le recrutement du duplex ARN sur la protéine AGO soit dépendant de la protéine chaperonne HSP90. En effet, il a été montré d'une part que la protéine HSP90 stabilisait la protéine AGO lorsque celle-ci était libre (Johnston *et al.*, 2010), et d'autre part, qu'elle permettrait à la protéine AGO de changer de conformation tridimensionnelle pour faciliter le recrutement des microARN double-brin (Pare *et al.*, 2013).

iv. Séparation des brins guide et passager

Une fois le complexe AGO-microARN double-brin formé, le duplex ARN est déstabilisé et le brin passager dégradé. Cette étape est effectuée par le domaine PIWI des protéines AGO, toutefois, deux mécanismes ont été mis en évidence, en fonction de la protéine AGO impliquée et du type de substrat (microARN ou siARN). Les domaines PIWI des 4 protéines AGO possèdent une structure proche des domaines catalytiques des endonucléases ayant une activité RNase H, à la différence qu'ils possèdent une activité dite de « slicing », c'est-à-dire l'hydrolyse d'un brin d'ARN lorsque celui-ci se trouve au sein d'un duplex ARN:ARN (et non d'ARN:ADN). L'hydrolyse est catalysée par une tétrade catalytique DEDH qui n'est présente qu'au sein des domaines PIWI des protéines AGO2 et AGO3. Toutefois, seule la protéine AGO2 possède une activité de « slicing » fonctionnelle (Faehnle *et al.*, 2013). Cependant, des expériences réalisées avec des protéines chimériques ont montré que le remplacement du domaine N-terminal de la protéine AGO3 par le domaine N-terminal de la protéine AGO2 suffisait à restaurer l'activité de « slicing » d'AGO3 (Faehnle *et al.*, 2013).

Le mécanisme principal de séparation des deux brins des microARN correspond à un mécanisme de « débobinage » de duplex. Ainsi, le microARN double-brin n'interagit avec la protéine AGO que

par l'intermédiaire du brin guide, laissant ainsi le brin passager libre. Le duplex est alors déstabilisé notamment grâce au domaine N-terminal d'AGO puis le brin passager est éjecté du complexe (Kawamata *et al.*, 2009 ; Kwak et Tomari, 2012). Lorsque le duplex est un siARN (appariement parfait entre les brins guide et passager), la séparation des deux brins est en plus facilitée par l'hydrolyse du brin passager par le domaine PIWI de la protéine AGO2 (Leuschner *et al.*, 2006). Toutefois, bien que les protéines AGO1, 3 et 4 ne possèdent pas l'activité de « slicing », elles ont également été retrouvées associées avec des siARN. Dans ces cas, la séparation des deux brins du siARN se ferait selon le même mécanisme que pour les microARN (Park et Shin, 2015). De même, la protéine AGO2 a été retrouvée associée avec des microARN. Toutefois, contrairement au mécanisme décrit pour les siARN, le brin passager ne sera pas clivé par AGO2 mais éjecté du duplex microARN comme avec les autres protéines AGO. En effet, l'activité de « slicing » de la protéine AGO2 est inhibée par la présence de mésappariements entre le brin guide et le brin passager, au niveau de la région « seed » (nucléotides 2 et 8) et/ou au niveau de la région 3' (position 12 à 16) (Yoda *et al.*, 2010 ; Liu *et al.*, 2012).

b. Le complexe RISC

Une fois le brin passager éjecté, le complexe RISC peut s'assembler autour de la protéine AGO et du microARN. La protéine Dicer n'étant pas indispensable au recrutement du microARN sur les protéines AGO (Liu *et al.*, 2005), il n'est pas certain qu'elle fasse partie du complexe RISC. De même, la présence des protéines TRBP/PACT au sein du RISC n'a pas non plus été clairement établie, mais des expériences d'immunoprécipitation dirigées contre les protéines AGO ont permis de montrer qu'elles étaient principalement associées avec la protéine GW182 (Meister *et al.*, 2005 ; Landthaler *et al.*, 2008). La protéine GW182 est une protéine de 182 kDa possédant des répétitions GW/WG (Eystathiou *et al.*, 2002), qui interagit directement avec les protéines AGO au niveau de leur domaine PIWI (Till *et al.*, 2007 ; Schirle et MacRae, 2012). Cette protéine est essentielle pour permettre l'inhibition de la traduction de l'ARNm ciblé par le complexe RISC (Lian *et al.*, 2009) (voir le paragraphe II. A. 2 ci-après).

C. Les autres voies de maturation

De nombreux travaux conduits ces dernières années, utilisant notamment des expériences de séquençage à haut-débit, ont révélé l'existence de différentes classes de petits ARN ayant une structure et des fonctions similaires aux microARN. En revanche, ces recherches ont montré aussi

que la maturation de ces petits ARN, ou microARN non-canoniques, était indépendante de certains complexes protéiques impliqués dans la biogenèse des microARN canoniques.

1. Maturation d'un pré-microARN par AGO2

Le microARN miR-451 est à ce jour le seul microARN connu dont la maturation est réalisée par AGO2. La maturation du pri-microARN-451 par le microprocesseur produit un pré-microARN d'environ 40 nucléotides de long, trop court pour être pris en charge par Dicer. En revanche, il a été montré qu'il pouvait être maturé par la protéine AGO2 par le biais de son activité de « slicing ». Toutefois, suite à cette coupure, l'extrémité 3' du futur microARN-451 doit encore être maturée. Bien que cette partie du mécanisme ne soit pas encore clairement établit, la maturation de cette extrémité est probablement réalisée par des exonucléases 3'-5' (Cheloufi *et al.*, 2010) (**Figure 56**). Le microARN-451 a également été retrouvé en complexe avec la protéine AGO1, toutefois, il a été montré que cette protéine était incapable de maturer le pré-microARN-451 (Cheloufi *et al.*, 2010).

2. Les miRtrons et simtrons

Les Mirtrons sont des microARN dont la maturation est indépendante du microprocesseur. Ils ont issu de la maturation d'un intron (ou pré-miRtron) et non d'un pri-microARN. Le processus de maturation est ainsi dépendant à la fois de la transcription d'un gène hôte, mais également du processus d'épissage (Ruby *et al.*, 2007). Suite au débranchement de l'intron, celui-ci est directement pris en charge par la protéine XPO5 qui va l'exporter dans le cytoplasme. Une fois dans le cytoplasme, le pré-miRtron poursuit sa maturation en microARN en utilisant la même voie de biogenèse que les microARN canoniques. Toutefois, dans certains cas, les extrémités 5' et/ou 3' du pré-miRtron doi(ven)t subir une étape de maturation supplémentaire, réalisée par une exonucléase, avant qu'il ne soit exporté dans le cytoplasme (Flynt *et al.*, 2010) (**Figure 57**).

Les simtrons (« splicing-independent mirtron-like ») sont une catégorie de miRtrons dont il a été montré que leur maturation était indépendante du processus d'épissage (Havens *et al.*, 2012). Les simtrons représentent une nouvelle catégorie très intéressante de microARN. En effet, bien que Drosha soit nécessaire à leur maturation, ce n'est pas le cas des autres protéines impliquées dans la maturation des microARN canoniques.

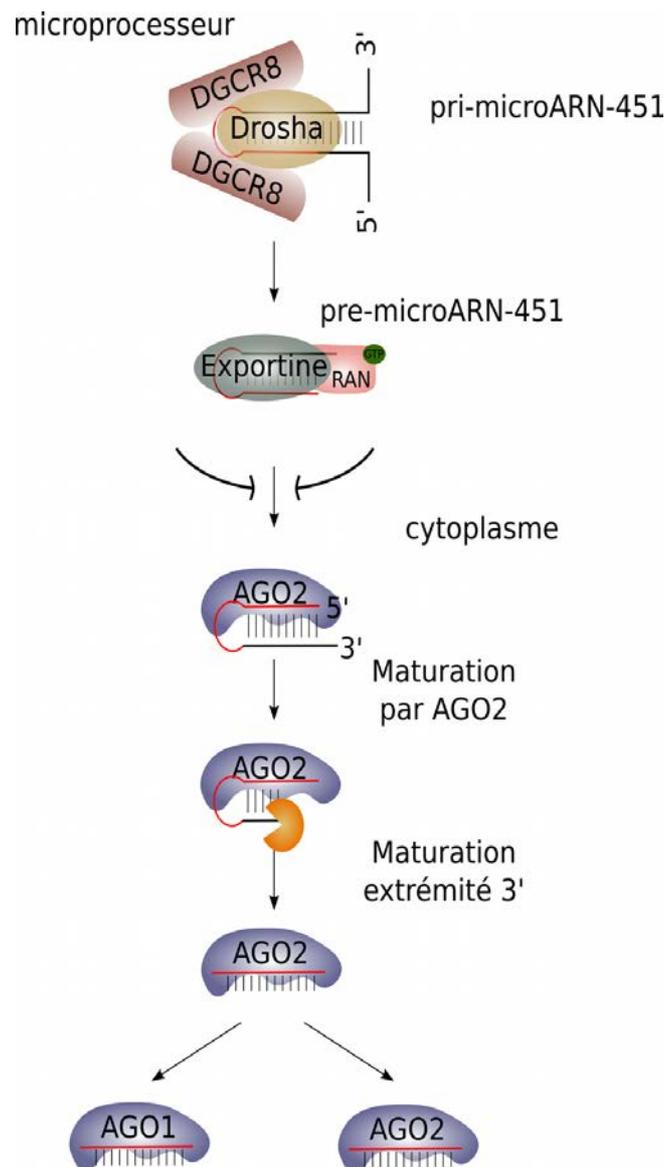


Figure 56 : Maturation non-canonique du microARN-451.

Le microARN-451 est maturé indépendamment de Dicer. Le pri-microARN est pris en charge par le microprocesseur qui forme un pré-microARN plus petit que les pré-microARN canoniques. Une fois exporté dans le cytoplasme, il ne peut être pris en charge par Dicer. La seconde coupure est effectuée par la protéine AGO2, via son activité de « slicing ». L'extrémité 3' du duplex obtenu est ensuite maturée par une ou des exonucléases qui n'ont pas encore été identifiées.

Enfin, contrairement aux microARN canoniques, la séquence du microARN-451 se situe au niveau de l'un des brins et de la boucle du pri-microARN (en rouge sur la figure).

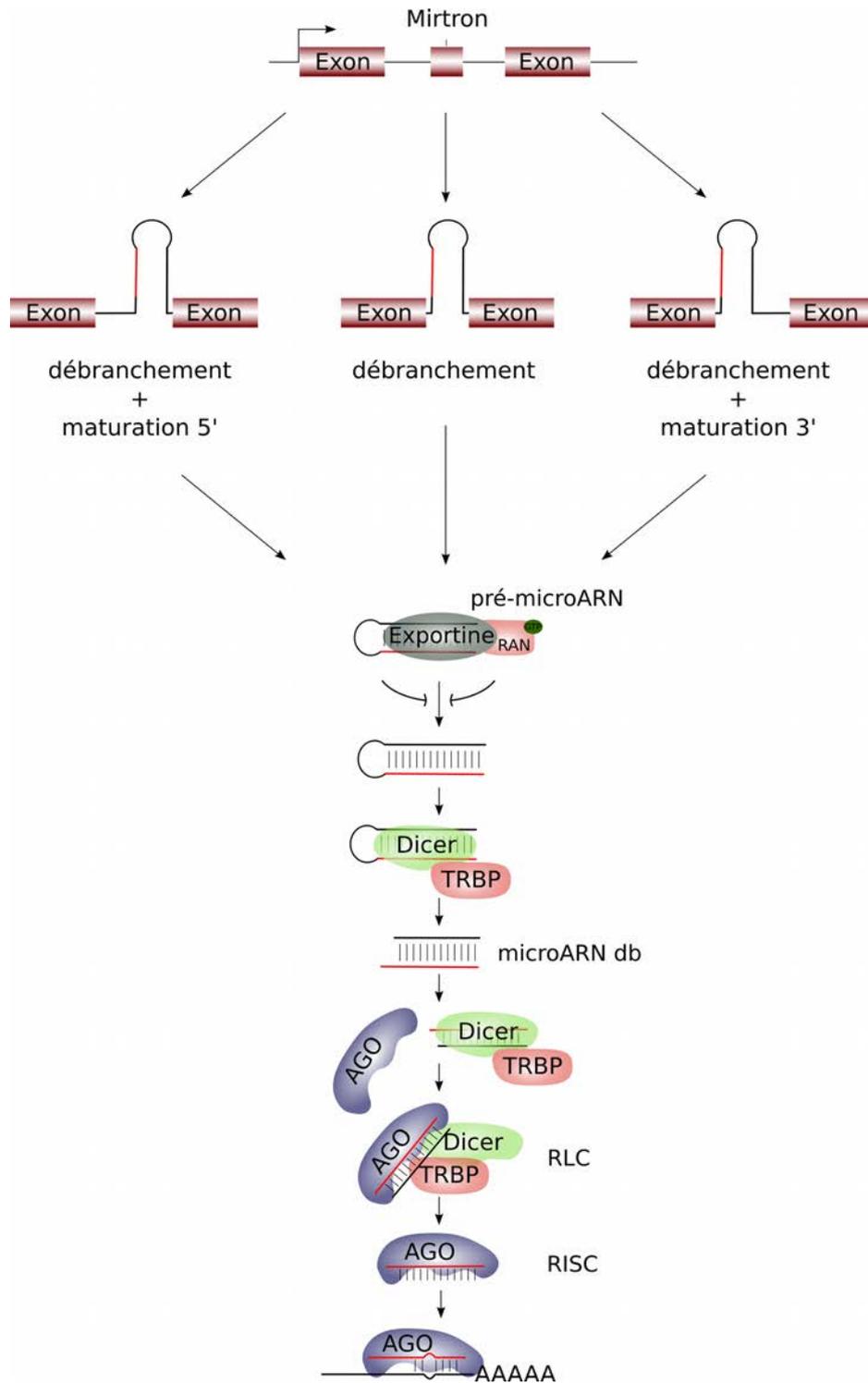


Figure 57: Maturation non-canonique des miRtrons (adapté de Westholm et Lai, 2011).

Les miRtrons sont issus d'un mécanisme indépendant du microprocesseur mais dépendant de l'épissage. Suite à l'épissage, l'intron est débranché puis exporté dans le cytoplasme où il sera ensuite mûri de manière canonique. Dans certains cas, les extrémités 5' ou 3' de l'intron doivent être mûries grâce à des exonucléases.

Ainsi, la maturation des simtrons est indépendante des protéines DGCR8, Dicer et AGO2. De plus, leur export dans le cytoplasme est également indépendant de la protéine XPO5 (Havens *et al.*, 2012) (**Figure 58**).

3. pré-microARN portant une coiffe m⁷G

Récemment, une nouvelle famille de microARN dont la maturation est indépendante du microprocesseur a été mise en évidence. Le pré-microARN est directement transcrit par l'ARN polymérase II et possède une coiffe m⁷G à son extrémité 5' (Xie *et al.*, 2013). Ainsi les extrémités 5' et 3' de ces pré-microARN sont générées respectivement par l'initiation et l'arrêt de la transcription. De plus, il a été montré qu'ils étaient exportés dans le cytoplasme via le système PHAX-CRM1 et non par l'Exportine-5. Le pré-microARN est ensuite maturé par Dicer, toutefois seul le brin 3p sera effectivement incorporé au sein du RISC. En effet, la présence de la coiffe m⁷G au niveau de l'extrémité 5' du pré-microARN augmentant la stabilité thermodynamique de cette extrémité, seul l'extrémité 5' du brin 3p sera reconnue par une protéine AGO et finalement incorporé au sein du RISC (Xie *et al.*, 2013) (**Figure 59**).

4. pré-microARN dérivés d'ARNt

L'analyse de données de séquençage à haut-débit d'ARN totaux de cellule HeLa a permis d'identifier une vingtaine de microARN issus des extrémités 5' et 3' d'ARN de transfert (Cole *et al.*, 2009). Ces microARN sont issus du clivage d'un ARNt par Dicer. Ils sont ensuite recrutés par les protéines AGO1 ou AGO2. Toutefois, ces microARN n'ont pas été retrouvés dans d'autres types cellulaires, par conséquent, leur fonction biologique n'est pas encore clairement établie. En revanche, dans le cadre de certaines infections virales, il a été montré que des microARN viraux pouvaient être obtenus suite à la maturation d'ARN de transfert par la RNase Z puis par Dicer (Bogerd *et al.*, 2010 ; Yeung *et al.*, 2009).

5. Les isomiRs

Les progrès en matière de séquençage à haut-débit ont également permis de montrer l'existence d'une certaine hétérogénéité de longueur et parfois de séquence au sein de microARN pourtant issus d'un même locus. Ainsi, l'alignement de ces séquences a montré qu'il existait de nombreux microARN dont les séquences ne différaient que d'un ou deux nucléotides.

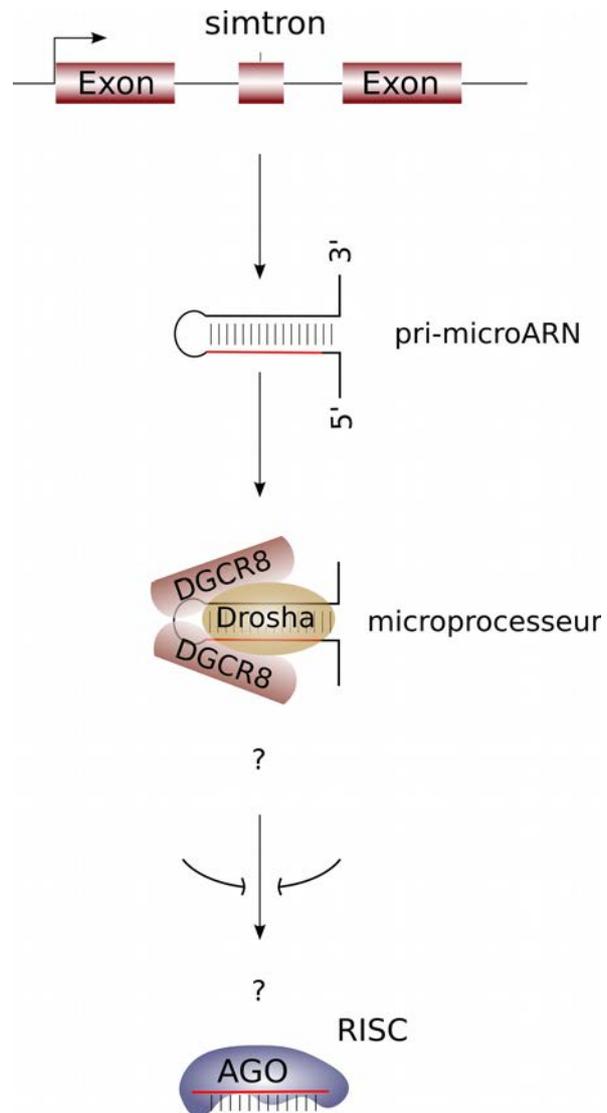


Figure 58: Maturation non-canonique des simtrons.

Peu d'information sont connues sur la maturation des simtrons. Toutefois, il a été montré que leur maturation était dépendante du microprocesseur mais indépendante des autres protéines impliquées dans la maturation canonique (Exportine-5, Dicer, AGO2)

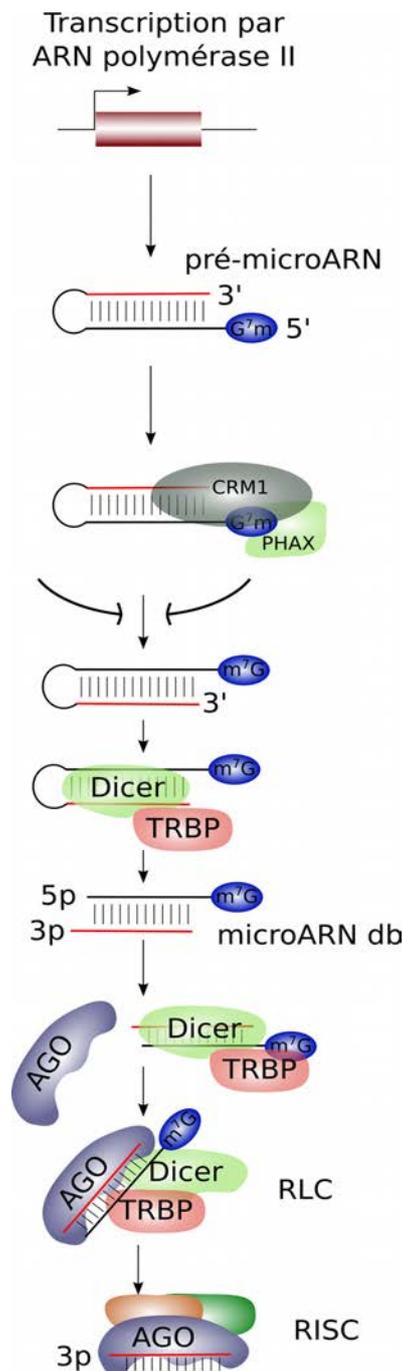


Figure 59: Maturation des microARN portant une coiffe 5' m⁷G.

Le pré-microARN est directement transcrit par l'ARN polymérase II. Ses extrémités 5' et 3' correspondent respectivement aux sites d'initiation et d'arrêt de la transcription. L'extrémité 5' porte ainsi une coiffe m⁷G qui sera reconnue par le système d'export cytoplasmique PHAX/CRM1. Une fois dans le cytoplasme, le pré-microARN est mûré de manière canonique. Seul le brin 3p sera recruté au sein du RISC (Xie *et al.*, 2013).

Ces différences se situent principalement au niveau de l'extrémité 3', parfois au niveau de l'extrémité 5', et, dans de rares cas, au sein du microARN en raison de la présence de SNPs (« Single Nucleotide Polymorphisme ») (Landgraf *et al.*, 2007 ; Lee *et al.*, 2010 ; Cloonan *et al.*, 2011 ; Ryan *et al.*, 2010 ; Chen et Rajewsky, 2006 ; Saunders *et al.*, 2007). Ces microARN sont appelés des isomiRs. Ils sont la conséquence de variations au cours de la maturation canonique. Bien que les éléments régulant la production de ces isomiRs ne sont pas encore entièrement connus, un certain nombre de travaux a montré qu'ils seraient produits suite à des imprécisions dans les « mesures » réalisées par les protéines Drosha et Dicer lors de la maturation respective des pri- et pré-microARN, générant ainsi des microARN avec des extrémités alternatives (Wu *et al.*, 2009) (Starega-Roslan *et al.*, 2010).

De récents résultats de séquençage à haut-débit, obtenus à partir de cellules de souris exprimant un variant de la protéine de Dicer incapable d'interagir avec les protéines TRBP et PACT, ont montré que dans certains cas, en l'absence de cofacteur, le microARN produit par Dicer présentait un nucléotide de moins au niveau de son extrémité 3' (Wilson *et al.*, 2015). Des expériences réalisées *in vitro* ont ensuite permis de comprendre un peu plus ce phénomène. Ainsi, dans certains cas, Dicer va générer un microARN double-brin de 24 nucléotides de long (au niveau du brin 5p) en présence de TRBP, contre 23 ou 24 nucléotides de long en son absence ou en présence de PACT (Wilson *et al.*, 2015). Toutefois, ce mécanisme ne concerne que certains microARN, et les microARN concernés ne sont affectés ni systématiquement, ni uniformément. Ainsi, dans le cas du pré-microARN-200a par exemple, la présence de TRBP engendrera 2 microARN de tailles différentes contre un seul en présence ou en absence de PACT (Lee *et al.*, 2013). Cette différence de comportement selon les pré-microARN concernés n'a pas encore été expliquée et il n'a pas encore été possible de la corrélérer avec la présence putative d'autres facteurs protéiques, des différences de stabilité thermodynamique au niveau des extrémités 5' et 3' du pré-microARN, ou encore l'existence de structures secondaires au sein de ces ARN (Wilson *et al.*, 2015).

La génération de ces isomiRs a pour principale conséquence de modifier l'identité des nucléotides présents au niveau des extrémités 5' ou 3' des microARN double-brin. Ces extrémités sont très importantes lors du recrutement du microARN double-brin par l'une des protéines AGO, qui va préférentiellement reconnaître l'extrémité la moins stable thermodynamiquement. Ainsi, pour certains isomiRs, l'incorporation du brin 3p au sein du RISC serait privilégiée au détriment du brin

5p (Wilson *et al.*, 2015). Les modifications de l'extrémité 5' peuvent également entraîner un changement ou un décalage de la région qui va interagir avec l'ARN cible (région « seed »), et donc modifier la ou les cibles du microARN (Azuma-Mukai *et al.*, 2008).

Bien que plus rares, la génération d'isomiR suite à des modifications post-transcriptionnelles au niveau du pri-microARN ou de duplex ARN a également été identifiée, telles que des délétions de nucléotides au niveau de l'extrémité 5' par des exonucléases 5'-3', l'ajout de nucléotides (A ou U notamment) au niveau de l'extrémité 3' par des nucléotidyltransférases (Sun *et al.*, 2012) ou encore la conversion d'adénosine en inosine par des enzymes de la famille ADAR (Adenosine Deaminase Acting on RNA) (Blow *et al.*, 2006).

Ces modifications peuvent avoir d'autres conséquences. Ainsi, l'ajout de résidus uridines au niveau de l'extrémité 3' du microARN est généralement un signal de dégradation alors que l'ajout d'adénosine augmente la stabilité du microARN (Shen et Goodman, 2004 ; Katoh *et al.*, 2009). La conversion d'adénosine au niveau du pri-microARN en inosine par les enzymes ADAR change le type d'interaction entre le nucléotide converti et l'ARN cible, puisque l'inosine est capable d'interagir avec les nucléotides A, C et U. De plus, la présence d'inosine peut également modifier la structure du pri-microARN. Ainsi, la conversion d'adénosine au niveau du pri-microARN en inosine peut avoir pour conséquences d'inhiber la maturation du microARN (Yang *et al.*, 2006), empêcher le recrutement du duplex ARN par les protéines AGO (Iizasa *et al.*, 2010) ou encore changer les cibles du microARN (Kawahara *et al.*, 2007).

II. Fonctions canoniques et non canoniques des microARN

A. Fonctions canoniques des microARN

1. Interactions avec l'ARN messenger cible

Classiquement, le RISC interagit avec l'ARNm cible par l'intermédiaire du microARN qui, comme pour les snoRNP, sert de guide afin de permettre le recrutement de protéines effectrices au niveau de l'ARNm cible et ainsi en bloquer la traduction. Le duplex ARNm: microARN contient généralement un appariement parfait au niveau des nucléotides 2 à 8 du microARN (région « seed ») et une boucle interne au niveau des nucléotides 9 à 12 (Lewis *et al.*, 2003 ; Doench et Sharp, 2004 ; pour revue Bartel, 2009). Toutefois de récentes études montrent que moins de 40 % des interactions au sein du duplex ARNm: microARN impliquent la formation d'un duplex parfait

au niveau de la région « seed » (Helwak *et al.*, 2013), Ainsi, il a été montré que dans certains cas, la région 3' du microARN peut également former un appariement parfait avec l'ARNm qui compenserait la présence de mésappariements au niveau de la région « seed » (pour revue Bartel, 2009). Enfin, il semble que l'accessibilité de l'ARNm (Marín et Vaníček, 2011) et la stabilité du complexe AGO-microARN-ARNm soient également des facteurs importants (Leoni et Tramontano, 2016).

Généralement, la formation du duplex se fait au niveau de l'extrémité 3'UTR (« UnTranslated Region ») de l'ARNm cible (pour revue Bartel *et al.*, 2009). Certains microARN peuvent également s'hybrider au niveau de la région codante (Duursma *et al.*, 2008 ; Forman *et al.*, 2008 ; Hafner *et al.*, 2010), voir dans au niveau de la 5'UTR. Toutefois, il a été montré que la présence de ribosome au niveau de la région codante et de la 5'UTR de l'ARNm diminuait fortement l'affinité des particules RISC pour l'ARNm dans ces régions (Gu *et al.*, 2009).

Habituellement, un même ARNm peut être ciblé simultanément par différentes particules RISC. Le nombre de particules interagissant avec l'ARNm peut être modulé par différents facteurs protéiques qui vont alors bloquer ou libérer les sites d'interaction avec les différentes particules RISC. Ainsi, certains ARNm peuvent avoir des sites de polyadénylations alternatifs qui sont utilisés en fonction de l'état physiologique de la cellule. Suivant cet état, la région 3'UTR de l'ARNm ne sera pas identique, et par conséquent l'ARNm ne sera pas ciblé par le même nombre de particules RISC (pour revue Di Giammartino *et al.*, 2011). Il a également été montré que des sites d'interaction avec l'ARNm de certaines RBP (« RNA-binding proteins ») pouvaient être en partie ou totalement confondus avec les sites d'interaction de microARN, ce qui influe sur l'accessibilité de ces sites pour les microARN (Kedde *et al.*, 2007 et 2010). De plus, l'édition post-transcriptionnelle de l'ARNm peut également moduler les interactions entre l'ARNm et les microARN (Borchert *et al.*, 2009 ; Liang et Landweber, 2007).

Enfin, il a été montré récemment que de longs ARN non-codants, ainsi que des ARNnc circulaires (circARN), pouvaient interagir avec des microARN afin de les séquestrer, régulant ainsi également leur disponibilité (Cesana *et al.*, 2011 ; Hansen *et al.*, 2013).

2. Rôle du RISC au sein du duplex microARN:ARNm

a. Inhibition et activation de la traduction de l'ARNm

L'inhibition de la traduction de l'ARNm par le RISC semble pouvoir s'effectuer à plusieurs niveaux. En effet, la traduction d'un ARNm est un processus finement régulé dont chaque étape peut potentiellement être régulé par le RISC. Ainsi, il a été montré que le RISC pouvait interférer avec la circularisation de l'ARNm en bloquant l'interaction entre la sous-unité eIF4G du facteur eIF4F, qui reconnaît la coiffe m⁷G de l'ARNm, et les protéines PABP, qui interagissent avec la queue poly(A) de ce même ARNm (Humphreys *et al.*, 2005 ; Mathonnet *et al.*, 2007). Le RISC peut également recruter le facteur eIF6, inhibant ainsi la formation du ribosome en bloquant l'association des 2 sous-unités (Chendrimada *et al.*, 2007 ; Wang *et al.*, 2008). Bien que plus rare, la traduction de l'ARNm cible peut également être bloquée par le RISC au cours de l'étape d'élongation. Ainsi, il a été montré que le RISC pouvait, lors de cette étape, faciliter la dissociation précoce des deux sous-unités du ribosome, ou encore entraîner la libération et/ou la dégradation prématurée du polypeptide naissant (Maroney *et al.*, 2006 ; Nottrott *et al.*, 2006 ; Petersen *et al.*, 2006).

Dans de rares circonstances cellulaires particulières, comme une carence en nutriment ou lors d'un stress, il semble que certains microARN puissent au contraire activer la traduction d'ARNm. Ainsi, il a été montré que plusieurs microARN interagiraient avec la 5'UTR de l'ARNm et en faciliteraient l'initiation de la traduction (Ørom *et al.*, 2008 ; Tsai *et al.*, 2009).

b. Dégradation de l'ARNm par l'intermédiaire du RISC

Les ARNm soumis au phénomène d'interférence par les RISC se concentrent alors dans le cytoplasme au sein de corps cytoplasmiques, les « P-Bodies » (« Processing-Bodies »), au sein desquels est retrouvé l'ensemble des protéines impliquées dans le maintien de l'inhibition de la traduction de ces ARNm (Eulalio *et al.*, 2007). Au sein de ces structures, l'inhibition de la traduction de l'ARNm par un ou plusieurs RISC reste en principe réversible tant que l'intégrité de l'ARNm n'est pas affectée. Ainsi, il a été montré que lors d'un stress, par exemple, le blocage de la traduction d'ARNm par les microARN pouvait être levée (Maroney *et al.*, 2006).

Bien que cela reste soumis à discussion, de récentes études ont montré que l'inhibition de l'expression des gènes par les microARN serait un processus dynamique qui commencerait par l'inhibition de la traduction de l'ARNm cible et aboutirait à la dégradation de ce dernier par l'intermédiaire du RISC (Guo *et al.*, 2010 ; Béthune *et al.*, 2012). Lors de cette seconde étape, il a

notamment été proposé que le RISC recruterait des complexes de déadénylation et/ou de décoiffage, entraînant alors la dégradation exonucléolytique de l'ARNm, dont les extrémités ne sont plus protégées (pour revue Eulalio *et al.*, 2008). De plus, il est possible que dans certains cas, l'interaction entre le microARN et l'ARNm forme un duplex parfait. Dans ces cas, il a été montré que l'ARNm subissait une coupure endonucléolytique par la protéine AGO2 si cette dernière est présente au sein du RISC concerné (Yekta *et al.*, 2004 ; Karginov *et al.*, 2010).

c. Rôle de la protéine GW182

Comme indiqué précédemment, les protéines qui composent le RISC ne sont certainement pas toutes encore clairement identifiées. Toutefois, il semble que le complexe soit constitué, *a minima*, d'une protéine AGO (1 à 4) et de la protéine GW182 (ou TNRC6). La protéine GW182 comporte un domaine N-terminal qui permet son interaction avec les protéines AGO. Elle contient des répétitions GW/WG, un domaine central UBA (« UBiquitin-Associated ») et un domaine C-terminal SD (« Silencing Domain ») qui contient notamment une région M1, un motif PAM2, une région M2 et un domaine C-terminal (Braun *et al.*, 2011). La protéine GW182 permet de faire le lien avec les protéines impliquées dans l'inhibition de la traduction et/ou la dégradation de l'ARNm cible et le complexe RISC. En effet, elle permet le recrutement des complexes de déadénylation CCR4-CAF1-NOT et PAN2-PAN3 en interagissant directement avec les protéines NOT1 et PAN3 par l'intermédiaire de son domaine SD (Braun *et al.*, 2011). Le recrutement de ces complexes de déadénylation au niveau de la 3'UTR de l'ARNm est également facilité par l'interaction du domaine SD avec la protéine PABPC1 (« Poly(A)-Binding Protein C1 ») qui reconnaît spécifiquement la queue poly(A) des ARNm.

Plusieurs modèles ont été proposés pour expliquer le rôle de ces interactions dans l'inhibition de la traduction des ARNm. Ainsi, la protéine GW182 pourrait être un compétiteur du facteur eIF4G pour l'interaction avec la protéine PABP et ainsi empêcher la circularisation de l'ARNm nécessaire à l'initiation de la traduction (Fabian *et al.*, 2009). L'interaction de la protéine GW182 avec PABP pourrait également réduire l'affinité de cette protéine pour la queue poly(A), réprimant ainsi la traduction (Fabian *et al.*, 2009). Enfin, un dernier modèle propose que cette interaction facilite le recrutement des complexes de déadénylation (Fabian *et al.*, 2009).

Le recrutement des complexes de déadénylation permet ensuite de recruter le complexe de décoiffage DCP1-DCP2. Une fois les extrémités 5' et 3' de l'ARNm déprotégées, celui-ci est rapidement dégradé par des exonucléases 5'-3' (XRN1) et 3'-5' (Exosome) (pour revue Houseley et Tollervey, 2009 et Braun *et al.*, 2013) (**Figure 60**).

B. Fonctions non-canoniques des microARN et protéines associées

1. Des microRNP sont présentes et actives dans les noyaux de cellules humaines

Plusieurs études ont suggéré assez tôt que les protéines impliquées dans l'ARN interférence seraient présentes et actives dans les noyaux de cellules de mammifères. Ainsi, dès 2005, il a été montré qu'il était possible de dégrader l'ARN nucléaire 7SK par l'utilisation de siARN spécifiques, suggérant la présence du complexe RISC dans le noyau (Robb *et al.*, 2005). Des expériences d'immunoprécipitation sur des fractions nucléaires et cytoplasmiques ainsi que des expériences d'immunofluorescence ont par la suite confirmé la présence des protéines AGO (Robb *et al.*, 2005 ; Ohrt *et al.*, 2008 ; Gagnon *et al.*, 2014), Dicer, TRBP et GW182 (Gagnon *et al.*, 2014 ; Kalantari *et al.*, 2016) dans le noyau. Ces travaux suggèrent également que le complexe RISC nucléaire n'est pas assemblé dans le noyau mais bien dans le cytoplasme puis importé, notamment par le biais de l'Importine-8 (Weinmann *et al.*, 2009 ; Wei *et al.*, 2014). De plus, Gagnon *et al.* ont montré que le recrutement du microARN sur la protéine AGO ne pouvait pas s'effectuer dans le noyau, probablement en raison de l'absence de la protéine HSP90 au sein de celui-ci.

Récemment, il a été montré que la protéine GW182 était capable de faire la navette entre le cytoplasme et le noyau via l'Importine- β . (Nishi *et al.*, 2013 ; Schraivogel *et al.*, 2015). De plus la localisation nucléaire des protéines AGO2 et GW182 semble être influencée par leur disponibilité cytoplasmique et notamment par l'existence ou non de complexes cytoplasmiques AGO2-GW182 (Schraivogel *et al.*, 2015). Un signal de localisation nucléaire non-canonique a également été identifié au sein de l'extrémité C-terminale de Dicer permettant ainsi à cette protéine d'être importée dans le noyau par les Importine-7, 8 et β (Doyle *et al.*, 2013).

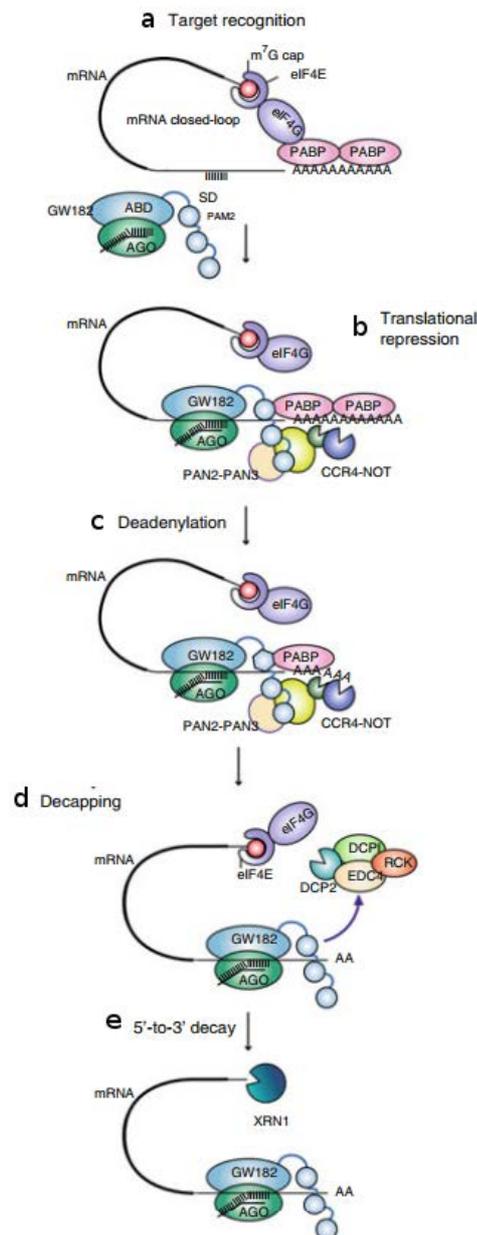


Figure 60: Inhibition de la traduction d'un ARNm par un microARN (pour revue Braun *et al.*, 2013).

(a) L'initiation de la traduction d'un ARNm nécessite la circularisation de l'ARNm grâce à l'interaction des protéines PABP présentes au niveau de l'extrémité 3' et du facteur d'initiation eIF4G qui interagit avec la coiffe m⁷G au niveau de l'extrémité 5'.

(b) L'interaction du RISC sur l'ARNm empêche l'interaction entre la protéine PABP et le facteur eIF4G grâce à l'interaction de la protéine GW182 avec PABP

(c) L'inhibition de la traduction de l'ARNm peut conduire à sa dégradation, suite au recrutement de complexes de déadénylation PAN2-PAN3 et CCR4-NOT.

(d) La coiffe de l'ARNm cible est ensuite retirée par l'intermédiaire de la protéine GW182 qui recrute le complexe DCP1-DCP2.

(e) Une fois déprotégé, l'ARNm est dégradé par des exonucléases 5'-3' (XNR1) et/ou 3'-5' (Exosome)

Enfin, des études par séquençage à haut-débit ont permis d'identifier un peu plus de 300 microARN nucléaires (J.-Y. Liao *et al.*, 2010). De manière intéressante, des travaux sur le microARN-29b, l'un des rares microARN dont la présence est fortement enrichie dans le noyau, ont montré l'existence au niveau de son extrémité 3' d'un motif hétéranucléotide, responsable de son importation dans le noyau (Hwang *et al.*, 2007).

2. Rôles des microARN nucléaires et des protéines associées

Au sein du noyau, les microARN et/ou leurs protéines associées semblent intervenir dans de nombreux mécanismes moléculaires. Ainsi, la protéine Dicer a été trouvée en interaction avec l'ARN polymérase II au niveau de certains sites de transcription. Elle empêcherait alors la formation d'ARN double-brin au niveau de ces sites, afin d'éviter l'induction de la voie de réponse à l'interféron et l'apoptose (White *et al.*, 2014). De même, la protéine AGO1 a également été trouvée associée à l'ARN polymérase II dans des cellules cancéreuses au niveau de promoteurs de différents oncogènes (Huang *et al.*, 2013). De plus, Hung *et al.* ont montré que la présence de AGO1 au niveau de ces gènes était corrélée à la présence de marques épigénétiques H3K4me3, synonymes de transcription active.

Les protéines AGO2, Dicer et Drosha seraient également impliquées dans la réparation de l'ADN (Gao *et al.*, 2014 ; Francia *et al.*, 2012). En effet, il a été montré que Drosha et Dicer étaient nécessaires à la génération de diARN (« double-strand break induced RNAs ») suite à l'apparition d'une coupure dans une double-hélice d'ARN. Ces diARN vont alors recruter la protéine RAD51, par l'intermédiaire de la protéine AGO2, afin d'activer la réparation de l'ADN par recombinaison homologue (Gao *et al.*, 2014).

La protéine AGO2 coopérerait également avec l'enzyme SETDB1, une méthyltransférase spécifique de la lysine 9 de l'histone H3, afin d'inhiber la transcription de certains gènes (Cho *et al.*, 2014). De même, il a été montré que les microARN-217, -101 et -9 diminueraient l'expression du long ARN non-codant nucléaire MALAT1 avec lequel ils interagissent directement (Leucci *et al.*, 2013 ; Wang *et al.*, 2015). Le mécanisme exact permettant cette inhibition n'a toutefois pas encore été élucidé.

Des régulations en *cis* ont également été identifiées. Ainsi, la transcription d'un ARN contenant la séquence du microARN-320 au niveau du promoteur du gène POLR3D (une sous-unité de l'ARN polymérase III) permettrait l'association au niveau de ce promoteur des protéines AGO1, EZH2 (une méthyltransférase appartenant au complexe PRC2) et d'histones H3 tri-méthylées au niveau de leur lysine 27 (H3K27me3) (Kim *et al.*, 2008). Les modifications épigénétiques H3K27me3 sont généralement retrouvées au niveau de promoteurs inactifs, ainsi le microARN-320 serait impliqué dans la répression en *cis* de la transcription du gène POLR3D. Inversement, il a été montré que les protéines AGO2 et GW182, associées au microARN-589-5p, activeraient la transcription du gène codant la protéine COX-2, une enzyme impliquée dans la régulation de l'inflammation de manière indirecte, en ciblant un ARN présent au niveau du promoteur de ce gène (Matsui *et al.*, 2013).

Les protéines AGO1 et AGO2 pourraient être également impliquées dans la modulation de l'épissage (Ameyar-Zazoua *et al.*, 2012). Ainsi, il a été montré qu'elles étaient capables d'interagir avec l'ARN polymérase II lors de la phase d'élongation de la transcription, et de faciliter le recrutement du spliceosome. En utilisant le gène CD44 comme modèle, les auteurs ont montré que le recrutement des protéines AGO1 et AGO2 permettrait l'apparition de marques épigénétiques H3K9me3 au niveau du gène transcrit, ce qui aurait pour conséquence de ralentir la progression de l'ARN polymérase, favorisant ainsi le recrutement du spliceosome et l'inclusion d'un exon alternatif. Toutefois, il n'est pas encore clairement établi si des microARN sont associés avec les protéines AGO1 et/ou AGO2 lors de ce processus.

Il semblerait aussi que le microprocesseur, ou tout du moins la protéine Drosha, puisse avoir une fonction cytoplasmique. En effet, récemment, une nouvelle isoforme de Drosha, issue d'un épissage alternatif, a été identifiée comme étant exclusivement cytoplasmique (Dai *et al.*, 2016). Bien que le rôle de cette isoforme ne soit pas encore établi, les auteurs ont montré que son expression était augmentée dans de multiples lignées tumorales.

Chapitre 3 : Les petits ARN dérivés de snoARN

I. Des snoARN particuliers

Les snoARN sont indispensables à la maturation des ARNr, puisqu'ils sont impliqués dans la catalyse de coupures endonucléolytiques au niveau du pré-ARNr (snoRNP U3, U8 et U17 par exemple), ou dans des événements de modification post-transcriptionnelle (2'-*O*-méthylation ou pseudouridylation). Toutefois, il semble qu'un certain nombre de snoARN soient impliqués dans d'autres processus biologiques.

A. snoARN et trafic intracellulaire du cholestérol

Il a récemment été montré que le snoARN à boîtes H/ACA U17, un snoARN impliqué dans la maturation du pré-ARNr (Atzorn *et al.*, 2004), serait également impliqué dans la régulation de l'élimination du cholestérol par les mitochondries (Jinn *et al.*, 2015). Le mécanisme est encore en grande partie inconnu, mais serait indépendant du rôle du snoARN U17 dans la maturation de l'ARNr. En effet, les auteurs ont montré que la délétion de l'un des deux allèles codant le snoARN U17 n'était pas suffisante pour impacter la maturation du pré-ARNr, alors qu'elle l'était pour modifier le trafic intracellulaire du cholestérol vers les mitochondries. Des expériences d'ARN interférence contre l'ARNm hôte du snoARN U17 ont permis d'identifier l'ARNm codant la protéine HUMMR (« Hypoxia-Upregulated Mitochondrial Movement Regulator »), qui est retrouvée dans les membranes des mitochondries, comme une cible potentielle du snoARN U17. En effet, non seulement l'expression de cet ARNm est inversement corrélée à celle du snoARN U17, mais il possède également des séquences complémentaires avec le snoARN. Toutefois, aucun site de pseudouridylation n'a pu être identifié au niveau de l'ARNm HUMMR, suggérant l'existence d'un autre mécanisme permettant sa régulation par le snoARN U17.

Cette équipe de chercheurs avait déjà montré l'implication du snoARN à boîtes C/D U60 dans la régulation du trafic intracellulaire du cholestérol (Brandis *et al.*, 2013). Ce snoARN avait initialement été identifié comme catalysant la 2'-*O*-méthylation du résidu G4340 de l'ARNr 28S. Toutefois, Brandis *et al.* ont montré, comme pour le snoARN U17, qu'il n'y avait pas de lien entre la méthylation de ce résidu et le contrôle du trafic intracellulaire du cholestérol, et donc que l'implication d'U60 dans ce second mécanisme n'était pas liée à son activité de méthylation. De manière intéressante, d'autres snoARN à boîtes C/D, notamment les snoARN U32a, U33 et U35a

semblent également jouer un rôle dans la protection des cellules contre le stress oxydant et le phénomène de lipotoxicité (Michel *et al.*, 2011).

B. snoARN et édition/épissage

D'autres études ont montré que certains snoARN orphelins pourraient avoir pour cible non pas des ARNr, mais des ARNm. Ainsi, le snoARN à boîtes C/D SNORD115, un snoARN orphelin, présente une complémentarité de 18 nucléotides avec un ARNm spécifique du cerveau. Cet ARN code pour le récepteur à la sérotonine 5-HT_{2c}, et possède cinq adénosines pouvant être édités et désaminés en inosine (Burns *et al.*, 1997 ; Kishore et Stamm, 2006), donnant ainsi lieu à des récepteurs moins actifs. De manière remarquable, le snoARN SNORD115 catalyserait la 2'-O-méthylation de l'un de ces cinq résidus, ce qui empêcherait sa conversion en inosine, (Vitali *et al.*, 2005). D'autres travaux ont par la suite démontré que cette régulation pouvait également impliquer le contrôle, par les snoARN, de l'épissage alternatif du pré-ARNm 5-HT_{2c}. En effet, l'hybridation du snoARN sur le pré-ARNm empêcherait l'activation d'un site inhibiteur d'épissage au niveau de l'exon V, conduisant ainsi à la production de la forme fonctionnelle du récepteur. En l'absence du snoARN, le site inhibiteur serait actif, conduisant à l'exclusion d'une partie de l'exon V (Vb), ce qui entraîne l'apparition d'un codon stop prématuré dans l'exon VI et donc à la formation d'une forme tronquée du récepteur au niveau de son domaine C-terminal (Kishore et Stamm, 2006). Depuis, d'autres études ont identifié plusieurs snoARN impliqués dans la régulation de l'épissage alternatif de différents pré-ARNm. Ainsi, une étude récente a montré que le snoARN à boîtes C/D SNORD27 pouvait s'hybrider avec différents pré-ARNm, dont celui du facteur de transcription e2F7 (Falaleeva *et al.*, 2016). Comme pour SNORD115, les auteurs ont montré que l'hybridation du snoARN masquait un site cryptique d'épissage, entraînant alors l'utilisation d'un site alternatif.

C. Les « sno-lncRNA »

1. Les snoARN SNORD115 et SNORD116

Le snoARN SNORD115 se trouve dans une région du chromosome 15 dont l'organisation est atypique. En effet, le locus 15q11-q13 est constitué de 2 gènes codant des protéines, SNURF-SNRPN et UBE3A, entre lesquels sont retrouvés plusieurs gènes de snoARN, notamment les snoARN à boîtes C/D SNORD116 et SNORD115. Ces snoARN sont organisés sous la forme de deux clusters au sein desquels chaque snoARN est répété plusieurs fois (chez l'homme, 29 copies pour SNORD116, 48 pour SNORD115). Le locus est par ailleurs soumis à l'empreinte génomique parentale et seul l'allèle paternel est exprimé, à l'exception du gène codant la protéine UBE3A qui

peut être exprimé à partir des deux allèles, sauf dans les cellules neuronales, où il n'est exprimé qu'à partir de l'allèle maternel. Enfin, l'ensemble de cette région formerait une unique unité transcriptionnelle, produisant un long ARN d'environ 470 kb (Runte *et al.*, 2001). De manière intéressante, il a été montré que la délétion d'une région d'environ 100 kb comprenant notamment le cluster des gènes SNORD116 était responsable de la maladie de Prader-Willi (Sahoo *et al.*, 2008 ; Bieth *et al.*, 2015).

2. Les long ARNnc liés à des snoARN

Tout comme le snoARN SNORD115, SNORD116 est un snoARN à boîtes C/D orphelin. Toutefois, une étude publiée en 2012 a proposé qu'il pourrait former une nouvelle catégorie d'ARNnc appelée sno-lncRNA (« snoRNA-related lncRNA » ou long ARNnc liés à des snoARN). Cette nouvelle catégorie d'ARNnc comporte à chacune de ses extrémités des séquences caractéristiques de snoARN (Yin *et al.*, 2012). Les protéines cœurs des snoRNP à boîtes C/D sont recrutées sur ces extrémités, conduisant à la formation d'une particule composée de deux snoRNP reliées par un long ARN simple-brin. La présence de ces deux snoRNP pourrait permettre de protéger l'ARN simple-brin de la dégradation par des exonucléases (Yin *et al.*, 2012). Les auteurs ont également montré que cette région simple-brin avait la capacité d'interagir avec le facteur d'épissage FOX2 pour le séquestrer. La modification de la disponibilité ce facteur a pour conséquence l'altération du profil d'épissage alternatif des pré-ARNm FOX2-dépendants (Yin *et al.*, 2012). Cinq sno-lncARN ont été identifiés au sein du cluster SNORD116, comprenant les snoARN SNORD116-6 et SNORD116-7, SNORD116-13 et SNORD116-14, SNORD116-18 et SNORD116-19, SNORD116-20 et SNORD116-21 et SNORD116-26 et SNORD116-27. Plus récemment, une étude bio-informatique a permis d'identifier 5 sno-lncARN supplémentaires présents dans d'autres régions chromosomiques, dont l'un contiendrait un snoARN à boîtes C/D et à boîtes H/ACA à ses extrémités 5' et 3' (Zhang *et al.*, 2014). La fonction de ces nouveaux sno-lncARN n'a pas encore été clairement établie.

De manière intéressante, les travaux de Deschamps-Francoeur *et al.* ont montré que le facteur d'épissage FOX2 était également impliqué dans la maturation des snoARN. En effet, lors de cette étude (Deschamps-Francoeur *et al.*, 2014), deux catégories de snoARN introniques ont été identifiées en fonction de la position des boîtes C et D par rapport aux extrémités du snoARN mature (voir paragraphe I. A. 2 du chapitre 1). Il a ainsi été montré que FOX2 stabilisait les snoARN aux extrémités « courtes » dont l'assemblage ne dépendrait pas, contrairement aux snoARN ayant des

extrémités « longues », de la protéine NOP58.

II. Les snoARN, précurseurs possibles de microARN ?

Les snoARN sont ainsi impliqués dans de nombreux processus cellulaires et leur population semble plus hétérogène qu'initialement envisagée. Les fonctions potentielles des snoARN se sont encore accrues lorsqu'en 2008 des travaux ont montré que les snoARN pouvaient être maturés en de petits ARN, formant ainsi une nouvelle catégorie d'ARNnc appelée sdARN (« sno-derived RNA »), et être dans certains cas une source de microARN.

A. Mise en évidence

Pour ces travaux, les auteurs ont réalisé des expériences d'immunoprécipitation des protéines AGO1 et AGO2 puis ont séquencé les ARN associés. De manière surprenante, de nombreux ARN identifiés étaient issus de la maturation de snoARN (Ender *et al.*, 2008). Les auteurs ont notamment montré que le scaARN à boîtes H/ACA ACA45, prédit pour guider la pseudouridylation du résidu U37 du snARN U2, pouvait former un microARN fonctionnel, associé préférentiellement à la protéine AGO1. En 2009 et 2011, deux autres équipes ont également identifié des microARN issus de snoARN, parfois appelés sno-miR (« snoRNA-microRNA »), provenant de snoARN à boîtes C/D ou à boîtes H/ACA (Scott *et al.*, 2009 ; Brameier *et al.*, 2011).

En parallèle, l'analyse par séquençage à haut-débit des petits ARN (<40 nucléotides de long) présents dans les différents compartiments cellulaires et subcellulaires (cytoplasme, noyau et nucléole) de cellules humaines a montré la présence dans le nucléole d'ARN dont la taille variait de 18 à 30 nucléotides, majoritairement issus de snoARN (Bai *et al.*, 2014b). De manière intéressante, la majorité (98%) des sdARN isolés par Bai *et al.* ont pour origine des snoARN à boîtes C/D. Ces sdARN peuvent contenir les boîtes C et/ou D mais sont dans leur grande majorité dépourvus de la région guide antisens. Les expériences réalisées par Bai *et al.* n'ont toutefois pas permis de savoir si ces petits ARN étaient intégrés au sein de particule RISC. La taille importante de certains fragments (>25 nucléotides) et leur localisation nucléolaire pourrait l'empêcher, bien que plusieurs microARN aient été localisés au sein du nucléole (Bai *et al.*, 2014a). En conclusion, il existerait donc différentes catégories de petits ARN dérivés de snoARN, dont seuls les sno-miR seraient effectivement incorporés au sein des particules RISC.

De manière intéressante, Stamm et son équipe ont rapporté que le snoARN à boîtes C/D SNORD115 pouvait former des sdARN d'un peu moins de 40 nucléotides de long (Kishore *et al.*, 2010). Ces sdARN auraient pour fonction la régulation de l'épissage alternatif du pré-ARNm du récepteur à la sérotonine 5-HT_{2c} et probablement d'autres pré-ARNm (DPM2, TAF1, RALGPS1, PBRM1 et CRHR1). Ces conclusions font toutefois toujours l'objet de débats (Bortolin-Cavaillé et Cavaillé, 2012).

La grande sensibilité des appareils de séquençage pourrait également permettre de détecter de simples produits de dégradation, et la détection de tels produits issus de snoARN n'est pas à exclure, dans certains cas, même si nombre de ces espèces dérivées de snoARN sont conservées au sein de différentes lignées cellulaires (Scott *et al.*, 2012). Enfin, en 2013, une nouvelle étude de Kishore *et al.* a conclu que bien que les sdARN existent probablement, ils ne seraient en revanche que très rarement incorporés au sein de particules RISC (Kishore *et al.*, 2013). Cette conclusion doit toutefois être nuancée. En effet, les auteurs ne se sont intéressé qu'aux ARN associés à la protéine AGO2, or les résultats des précédentes études sur les sno-miR suggéraient que ces ARN s'associaient préférentiellement avec les autres protéines AGO, notamment AGO1 (Ender *et al.*, 2008).

B. Des fonctions diverses pour les sdARN

1. sdARN et épissage alternatif

Outre les sno-miR qui sont incorporés au sein des particules RISC, les sdARN semblent intervenir dans différents mécanismes cellulaires. Ainsi, le sdARN issu du snoARN SNORD88C serait impliqué dans le contrôle de l'épissage alternatif de plusieurs pré-ARNm (Scott *et al.*, 2012 ; pour revue Abel *et al.*, 2014). Les snoARN à boîtes C/D SNORD88A, B et C sont issus de la maturation de différents introns présents au sein du gène C19orf48. Cette famille de snoARN est très conservée chez les mammifères et certains membres possèdent, en plus des boîtes canoniques C et D, une boîte M (Ono *et al.*, 2011), complémentaire à certains pré-ARNm, notamment celui codant la protéine FGFR3, un récepteur membranaire pour le FGF (« Fibroblast Growth Factor ») (Scott *et al.*, 2012). L'analyse des séquences des sdARN issus du SNORD88C a montré que certains d'entre eux contenaient la boîte M, or Scott *et al.* avaient montré que l'expression transitoire de cette séquence était suffisante pour modifier le profil d'épissage du pré-ARNm FGFR3 (Scott *et al.*, 2012). Cet effet sur le pré-ARNm est retrouvé lorsque le snoARN SNORD88C entier est exprimé,

mais pas en présence des petits ARN ne contenant pas la boîte M. Ainsi, bien que la modification du profil d'épissage puisse être obtenue à partir du snoARN entier, ces travaux suggèrent que des sdARN peuvent également jouer un rôle dans la régulation de l'épissage alternatif de certains pré-ARNm.

2. Des sdARN comme biomarqueurs de cancers ?

Ces dernières années, les recherches de nouveaux marqueurs moléculaires pour le diagnostic précoce de différents cancers ont montré des corrélations entre le taux de survie à ces cancers et le niveau d'expression de certains snoARN (pour revue Abel *et al.*, 2014). Ainsi, dans les cas de cancers bronchiques non à petites cellules (CBNPC), le niveau d'expression des snoARN à boîtes H/ACA SNORA42 et des snoARN à boîtes C/D SNORD33, 66 et 76 est significativement augmenté (Liao *et al.*, 2010). De plus, il a été montré que le taux d'expression du SNORA42 était inversement corrélé aux chances de survie pour ce type de cancer (Mei *et al.*, 2012). Inversement, l'augmentation de l'expression du snoARN SNORD71 est associée avec un pronostic favorable dans le cas des lymphomes T périphériques (Valleron *et al.*, 2012b). Le snoARN SNORD50 serait également un marqueur potentiel pour le cancer de la prostate (Dong *et al.*, 2008) et du cerveau (Dong *et al.*, 2009). Récemment, l'étude par séquençage à haut-débit des petits ARN de patients souffrant d'un cancer de la prostate a permis d'identifier environ 80 sdARN ayant une taille comprise entre 15 et 29 nucléotides de long dont le niveau d'expression est modifié par rapport aux personnes saines (Martens-Uzunova *et al.*, 2015).

Enfin, une étude récente par microarray a permis de montrer des liens entre snoARN, sdARN et le facteur de transcription p53. Cette protéine est un suppresseur de tumeur qui joue un rôle important dans les mécanismes de survie cellulaire et d'apoptose. Ainsi, une fois la protéine stabilisée suite à l'activation d'oncogènes ou à la présence de cassures au niveau de l'ADN, celle-ci va activer ou réprimer différents gènes. Lors de cette étude, les auteurs ont identifié 6 snoARN dont l'expression est fortement régulée en présence de la protéine p53 (Yu *et al.*, 2015). Ces snoARN sont tous issus de la maturation du gène hôte SNHG1, dont la transcription serait alors inhibée par la protéine p53. De plus, des expériences de HIT-CLIP contre la protéine AGO2 ont permis aux auteurs d'identifier un microARN, sno-miR-28, dont le précurseur serait l'un de ces snoARN, SNORD28. Ce microARN aurait pour cible l'ARNm codant une sous-unité du facteur d'initiation de la transcription TFIID, le facteur TAF9B, dont la fonction consiste à stabiliser la protéine p53 lors de

son activation. Les auteurs ont alors montré que l'inhibition de la traduction de l'ARNm TAF9B par ce sno-miR déstabilisait la protéine p53. Le sno-miR-28 serait ainsi un oncogène, inhibant indirectement l'expression du facteur de transcription p53 (Yu *et al.*, 2015).

C. Un processus de maturation inconnu

Les données de la littérature ne permettent pas encore de bien comprendre comment ces petits ARN sont maturés, que se soit les sdARN en général ou les sno-miR en particulier. Toutefois, quelques pistes ont pu être explorées. La maturation des snoARN et des microARN nécessite l'intervention de RNase de type III (Rnt1 pour les snoARN chez la levure *S. cerevisiae*, Drosha et Dicer pour les microARN). et la structure des snoARN est similaire à celle des pri- et pré-microARN. Ainsi, il a été montré *in vitro* que Dicer pouvait accepter certains snoARN comme substrat, par exemple le scaARN ACA45 (Ender *et al.*, 2008). Toutefois, dans le cas des sdARN non incorporés dans des particules RISC, il a été montré que leur maturation était indépendante de la voie de maturation des microARN (Bai *et al.*, 2014b), ce qui suggère l'existence de différentes catégories de petits ARN dérivés de snoARN. Des chercheurs ont alors émis l'hypothèse de l'existence d'un lien évolutif entre snoARN et microARN. Ainsi, les précurseurs de pré-microARN auraient pu initialement correspondre à des snoARN. Ces derniers auraient alors perdu peu à peu leur fonction initiale de guide pour acquérir de nouvelles propriétés fonctionnelles (Scott et Ono, 2011).

D'autres liens entre microARN et snoARN ont été observés, notamment concernant les protéines impliquées dans la maturation respective de ces deux familles d'ARNnc. Ainsi, des expériences de co-immunoprécipitation dirigées contre les protéines AGO1 et AGO2 ont mis en évidence l'existence d'une association entre ces protéines et les protéines NOP56 et Fibrillarine, respectivement (Höck *et al.*, 2007). Enfin, récemment, Cáceres et son équipe ont montré que la protéine DGCR8 était impliquée dans la dégradation des snoARN. Ainsi, il semble que DGCR8, indépendamment de Drosha, puisse être recrutée au niveau des snoARN à boîtes C/D et H/ACA matures, interagir avec la protéine RRP6, une sous-unité de l'exosome, et permettre ainsi la dégradation du snoARN (Macias *et al.*, 2015).

Bien que snoARN et microARN semblent avoir de nombreux points communs, et malgré l'existence de liens possibles entre certaines protéines cœurs des snoRNP et les protéines Argonautes, aucune étude n'a encore été entreprise pour étudier de manière exhaustive les liens

possibles entre la maturation des microARN et l'assemblage des snoRNP. Ainsi, les processus impliqués dans la maturation spécifique des sdARN, et plus particulièrement des sno-miR, à partir de snoARN précurseurs, demeurent encore largement inconnus.

Objectifs

La biogenèse des snoRNP et des microRNP comporte de nombreuses étapes de maturation faisant intervenir un très grand nombre de facteurs. La compréhension de toutes ces étapes n'est pas encore totale et plusieurs résultats récents ont amené leur lot de surprises. Ainsi, les fonctions possibles des snoRNP et des microRNP semblent être en réalité beaucoup plus variées qu'initialement anticipé. Récemment, des expériences de séquençage à haut-débit ont permis l'identification de nouveaux petits ARNnc issus de la maturation (ou de la dégradation) d'autres ARNnc, notamment de snoARN. Certains de ces petits ARN dérivés de snoARN (sdARN) sont intégrés au sein d'une particule RISC fonctionnelle, formant ainsi une nouvelle classe d'ARNnc appelée sno-miR.

Depuis leur mise en évidence, très peu d'études ont porté sur l'élucidation des mécanismes permettant la maturation des sdARN ou des sno-miR à partir de leurs snoARN précurseurs. Bien que des expériences d'immunoprécipitation aient précédemment montré l'existence de liens possibles entre les protéines Argonautes et certaines protéines cœurs des snoRNP à boîtes C/D (Höck *et al.*, 2007), aucune hypothèse fonctionnelle n'avait été avancée concernant la fonction possible de tels liens. Quant aux premiers travaux réalisés sur les sdARN issus du scaARN ACA45, leurs résultats montraient l'implication de la protéine Dicer, mais pas des protéines Drosha et DGCR8 (Ender *et al.*, 2008), suggérant un mécanisme de maturation différent des miARN canoniques. Plus récemment, deux études ont montré que DGCR8 pouvait, outre son activité dans le complexe microprocesseur, elle aussi interagir avec les snoARN et être impliquée dans leur dégradation via le recrutement de l'exosome (Macias *et al.*, 2013 et 2015). Mais aucune étude n'a été menée afin d'identifier les liens et implications respectives possibles des voies de maturation des snoRNP et des microARN dans la maturation des sdARN et des sno-miR. En particulier, aucune étude n'a proposé d'étudier l'existence possible d'un processus de coopération ou de compétition entre ces voies conduisant à la maturation des sdARN et des sno-miR.

Au cours de mon stage de master 2, nous avons réalisé un crible double-hybride chez la levure *S. cerevisiae*, en ciblant des interactions entre des facteurs protéiques impliqués dans l'assemblage des snoRNP et des facteurs de maturation des microARN. Le but de ce crible était d'identifier des connexions entre ces deux voies de biogenèse, pouvant être impliquées dans la maturation de petits ARN à partir de snoARN précurseurs. Grâce à ce crible, nous avons mis en évidence une nouvelle interaction entre la protéine RPAP3, un composant du complexe hR2TP, qui est notamment impliqué dans l'assemblage des snoRNP (Kakihara et Saeki, 2014), et la protéine TRBP, un des

cofacteurs possibles de la RNase Dicer (Chendrimada *et al.*, 2005). L'existence de cette nouvelle interaction, si elle était confirmée, pouvait alors indiquer un rôle du complexe TRBP/RPAP3 dans la maturation des sdARN. Toutefois, comme pour DGCR8, il est également possible que les protéines TRBP et RPAP3 soient impliquées respectivement dans la maturation des snoRNP et des microRNP, voire dans la maturation d'autres particules ribonucléoprotéiques.

La première partie de mon travail de doctorat a ainsi consisté à valider et confirmer l'interaction entre RPAP3 et TRBP, aussi bien *in vitro* qu'*in vivo*, puis à identifier les domaines impliqués dans cette interaction, l'objectif suivant étant d'en déterminer la structure tridimensionnelle par RMN.

À l'aide de ces différents résultats, le dernier objectif a ensuite consisté à tenter d'identifier les rôles possibles du complexe TRBP/RPAP3 dans la cellule par une étude des relations structure/fonction de ses composants. J'ai notamment utilisé les données RMN sur le complexe, obtenues en collaboration avec le Dr Marc Quinternet de la plateforme RMN de la Fédération de Recherche FR3209, pour identifier les résidus impliqués dans la formation du complexe, et en caractériser l'importance fonctionnelle par mutagenèse dirigée, en étudiant les conséquences de ces mutations sur la maturation des microARN, des snoARN, des sdARN et des sno-miR, mais également sur d'autres ARNnc putatifs.

Résultats et Discussion

Chapitre 1 : Confirmation et caractérisation de l'interaction entre RPAP3 et TRBP

I. Présentation des travaux préliminaires au doctorat

Afin d'identifier de possibles liens entre les voies de maturation des snoRNP et des microARN, nous avons réalisé au cours de mon stage de deuxième année de master un crible d'interaction entre les protéines impliquées dans ces deux voies de biogenèse, en utilisant la technique de double-hybride chez la levure *S. cerevisiae*.

A. Identification d'une nouvelle interaction

La technique du double-hybride repose sur l'expression dans la levure de deux protéines de fusion (**Figure 61**). La première protéine chimérique (protéine « appât »), dont la séquence codante est portée par le plasmide pAS2 (voir Matériel et Méthodes), est fusionnée à l'extrémité C-terminale du domaine de fixation à l'ADN du facteur de transcription Gal4 (Gal4-BD). L'ADNc de la seconde protéine (la protéine « proie »), est présent sur le plasmide pACT2, en fusion avec l'extrémité C-terminale du domaine d'activation de la transcription du même facteur de transcription Gal4 (Gal4-AD). En cas d'interaction entre les deux partenaires protéiques testés, la réunion des deux domaines fonctionnels de l'activateur transcriptionnel Gal4 permet la transcription d'un gène rapporteur placé sous le contrôle de l'opérateur Gal1, qui constitue l'élément de réponse à l'activateur Gal4. Dans notre cas, ce gène rapporteur est le gène *HIS3*, impliqué dans le métabolisme de l'histidine. L'interaction des deux protéines de fusion permet ainsi la croissance des levures sur un milieu ne contenant pas d'histidine. Un inhibiteur compétitif du produit du gène *HIS3*, le 3-Amino-Triazol (3-AT), est également utilisé avec des concentrations croissantes afin de déterminer la force et la spécificité de l'interaction.

A l'issue de ce crible, nous avons mis en évidence, notamment, une nouvelle interaction entre les protéines TRBP et RPAP3 (**Figure 62A**). En effet, les résultats montrent une croissance des levures, y compris en présence d'une forte concentration en 3-AT (40 mM), lorsqu'elles co-expriment les protéines TRBP et RPAP3. Comme nous l'avons vu au cours de l'introduction, la protéine TRBP est l'un des cofacteurs possibles de la protéine Dicer, alors que la protéine RPAP3 est l'un des composants du complexe hR2TP qui intervient notamment dans l'assemblage des snoRNP à boîtes C/D (**Figure 62B**).

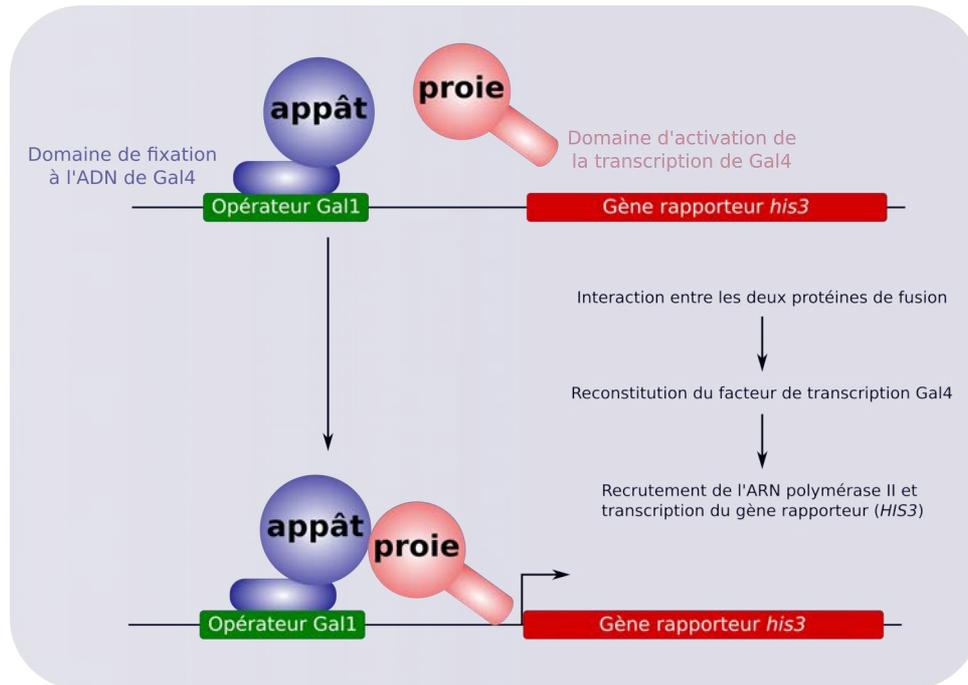


Figure 61: Principe du double-hybride chez la levure *S. cerevisiae*.

La technique de double-hybride repose sur l'expression chez la levure de deux protéines, « appât » et « proie », fusionnées au domaine de fixation à l'ADN (Gal4-BD) et d'activation de la transcription (Gal4-AD) du facteur de transcription Gal4.

En cas d'interaction entre les deux partenaires protéiques testés, le facteur de transcription Gal4 est reconstitué et permet la transcription d'un gène rapporteur placé sous le contrôle de l'opérateur Gal1.

Ici, le gène rapporteur est le gène *HIS3* qui code une enzyme impliquée dans la biosynthèse de l'histidine. L'interaction des deux protéines de fusion et donc l'expression du gène *HIS3* permet ainsi la croissance des levures sur un milieu ne contenant pas d'histidine. Un inhibiteur du produit du gène *HIS3*, le 3-AT (3-AminoTriazol) est également utilisé à différentes concentrations afin de tester la force de l'interaction.

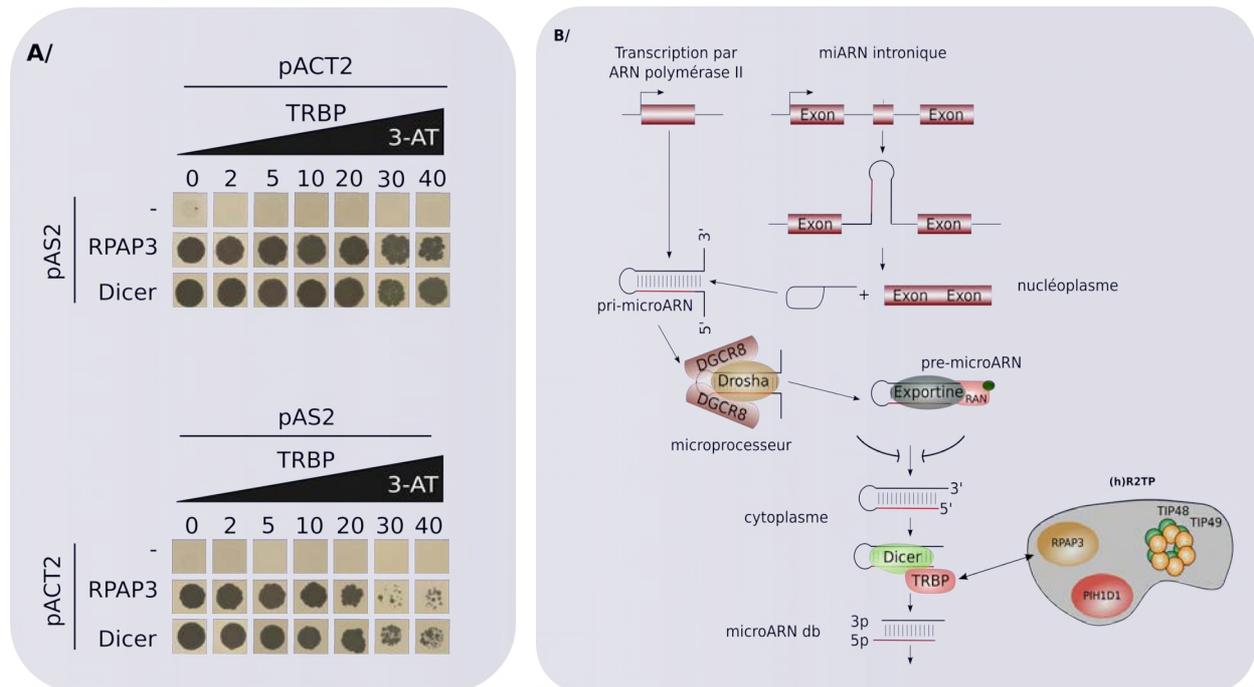


Figure 62: Mise en évidence d’une nouvelle interaction entre les protéines RPAP3 et TRBP lors d’un criblé d’interaction double-hybride.

A/ Le criblé double-hybride réalisé entre différentes protéines impliquées dans l’assemblage des snoRNP et la maturation des microARN a permis de mettre en évidence une nouvelle association entre les protéines TRBP et RPAP3. Cette association semble très stable puisque la croissance des levures est observée jusqu’à 40 mM de 3-AT, soit autant que pour l’interaction de notre témoin, entre TRBP et Dicer.

B/ Schéma indiquant le lien possible entre maturation des microARN et assemblage des snoRNP. Comme indiqué dans l’introduction, la protéine TRBP est notamment connue pour être l’un des cofacteurs possibles de la RNase III Dicer qui intervient dans la maturation des pré-microARN en microARN matures double-brin. La protéine RPAP3 est de son côté l’un des constituants du complexe (h)R2TP.

0-2-5-10-20-30-40 : différentes concentrations en 3-AT utilisée, en mM

De manière remarquable, ces résultats de double-hybride montre que l'interaction entre TRBP et RPAP3 serait aussi forte que celle entre Dicer et TRBP.

B. L'interaction entre TRBP et RPAP3 est directe

La technique de double-hybride ne permet pas de s'assurer que l'interaction entre deux partenaires soit bien directe. En effet, bien qu'étant dans un système hétérologue, il est possible qu'une protéine de la levure fasse office de médiateur de l'interaction entre les deux partenaires testés. De plus, bien que des contrôles à l'aide de vecteurs n'exprimant que les domaines Gal4-AD ou Gal4-BD aient été réalisés, le risque d'observer des faux-positifs ne peut pas être écarté. Nous avons donc voulu dans un premier temps vérifier la validité de cette interaction et déterminer son caractère direct ou indirect. Pour cela, des expériences de GST pull-down ont été réalisées. Les protéines RPAP3 et TRBP ont alors été produites et purifiées, en fusion respectivement avec une étiquette GST et 6-histidines. La protéine fusionnée à l'étiquette GST a alors été immobilisée sur des billes de glutathion-sépharose avant d'être incubée avec la seconde protéine. Un western-blot à l'aide d'anticorps dirigés contre cette seconde protéine est alors réalisé afin de valider l'interaction directe entre les deux protéines. Deux témoins négatifs sont également effectués afin de s'assurer que la seconde protéine n'interagit pas de manière non spécifique avec les billes ou avec la GST. Malgré une légère interaction non-spécifique avec les billes de sépharose, les résultats montrent clairement que la protéine GST-RPAP3 retient spécifiquement la protéine hist-TRBP, traduisant la validité de l'interaction et son caractère direct (**Figure 63**). En effet, bien qu'un très léger signal soit visible au niveau des pistes correspondant à la protéine hist-TRBP incubée uniquement avec les billes, ou avec les billes et la GST, l'ajout de protéine GST-RPAP3 permet d'obtenir un signal beaucoup plus important, indiquant une interaction directe et spécifique entre TRBP et RPAP3.

II. Vérification de l'interaction *in vivo*

Suite à notre observation inédite d'une interaction directe entre les protéines RPAP3 et TRBP, de nombreuses questions se sont posées, notamment concernant la possibilité d'observer cette interaction *in vivo*. En effet, les expériences de double-hybride combinées au « GST pull-down » précédemment réalisées ne permettent de conclure que sur la possibilité d'une interaction directe entre RPAP3 et TRBP *in vitro*. De plus, ces deux protéines sont principalement décrites pour être localisées dans des compartiments cellulaires différents.

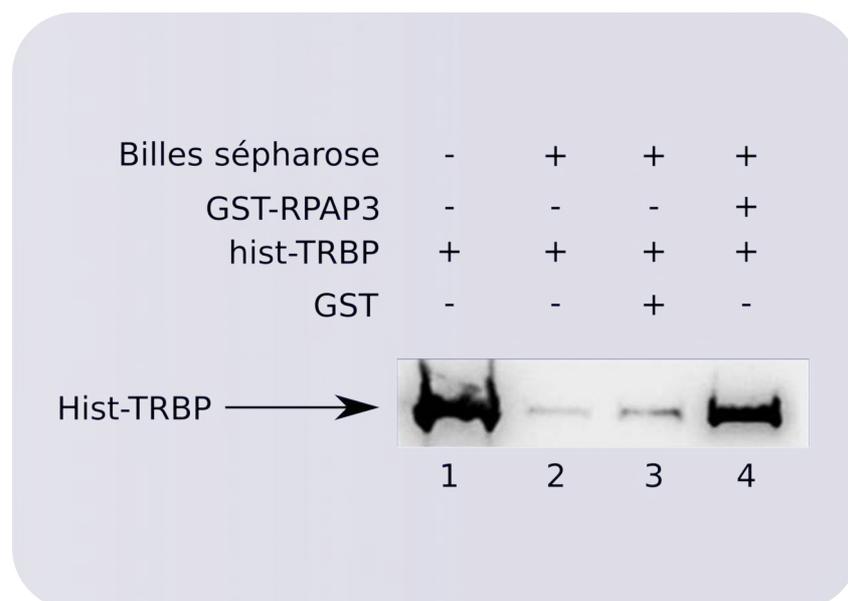


Figure 63: GST pull-down entre les protéines TRBP et RPAP3.

Les protéines TRBP et RPAP3 ont été produites fusionnées respectivement à une étiquette 6-histidines et à la GST. Les protéines purifiées ont ensuite été utilisées pour vérifier si l'interaction entre ces deux protéines est directe ou non, par « GST pull-down ». Pour cela, la protéine GST-RPAP3 a été immobilisée sur des billes de sépharose substituées avec du glutathion puis incubée avec la protéine hist-TRBP préalablement purifiée. Après plusieurs lavages, un western-blot a été réalisé à l'aide d'anticorps dirigés contre la protéine TRBP. Des essais témoins ont également été réalisés afin de vérifier que la protéine hist-TRBP n'interagissait pas de manière non spécifique avec les billes de sépharose ou avec l'étiquette GST seule.

Les résultats obtenus montrent une légère interaction non spécifique de la protéine hist-TRBP avec les billes de glutathion-sépharose (piste 2 et 3). Toutefois, en présence de GST-RPAP3 (piste 4), la protéine hist-TRBP est spécifiquement et fortement retrouvée sur les billes, traduisant l'existence d'une interaction directe et spécifique entre ces deux protéines.

La première piste nous permet de vérifier la spécificité de l'anticorps utilisé lors du western-blot pour détecter la protéine hist-TRBP.

En effet, l'assemblage des snoRNP est un mécanisme principalement nucléaire, il est fort probable que la protéine RPAP3 y soit principalement localisée. Au contraire, la maturation des pré-microARN en microARN double-brin par le complexe Dicer/TRBP est réalisée au sein du cytoplasme. Nous avons donc dans un premier temps réalisé des expériences d'immunoprécipitation sur des extraits cellulaires afin de valider cette association *in vivo*.

De nombreuses mises au point ont été nécessaires pour la réalisation des expériences d'immunoprécipitation. En effet, afin d'être au plus près des conditions cellulaires et physiologiques, nous avons dans un premier temps essayé de réaliser ces immunoprécipitations sur les protéines TRBP ou RPAP3 endogènes, à partir d'extrait totaux de cellules HEK et HeLa. Malheureusement, dans ces conditions, nous n'avons pas été capables de détecter l'association entre RPAP3 et TRBP, probablement en raison du caractère beaucoup trop labile et transitoire de l'interaction, ce qui constitue une observation fréquente de nombreuses interactions protéine-protéine connues. Pour le vérifier, nous avons effectué les mêmes expériences en utilisant cette fois des extraits totaux préalablement traité à l'EGS (Éthylène Glycol bis(Succinimidyl succinate)), afin de générer des pontages covalents entre toutes protéines distantes au maximum de 16 Å. L'EGS fut choisi en grande partie pour sa capacité à agir directement dans les cellules. Malheureusement, là encore les résultats ne furent pas concluants. En effet, pour mettre au point cette expérience, nous avons utilisé comme témoin positif l'association entre RPAP3 et PIH1D1, qui est détectable par immunoprécipitation classique. Toutefois, malgré de nombreux essais, nous n'avons pas réussi à détecter cette association sur les cellules traitées par l'EGS.

Nous avons alors entrepris de réaliser les expériences d'immunoprécipitation en présence de la protéine TRBP, sur-exprimée transitoirement en fusion avec une étiquette V5 qui sera utilisée pour immunoprécipiter la protéine (voir Matériel et Méthodes). Cette nouvelle approche nous a ainsi permis de confirmer l'association *in vivo* de la protéine RPAP3 endogène avec la protéine V5-TRBP (**Figure 64**, piste 1 et 2). En effet, bien qu'il semble que RPAP3 interagisse légèrement avec les billes (piste 1), la présence d'une bande de plus forte intensité lors de l'immunoprécipitation de la protéine V5-TRBP (piste 2), permet de confirmer l'association de ces deux protéines *in vivo*. De plus, il semble que TRBP (endogène et/ou exogène) interagisse légèrement avec les billes en l'absence d'anticorps anti-V5 (piste 1), ce qui pourrait expliquer, du moins en partie, la détection de RPAP3 lorsque la protéine V5-TRBP n'est pas spécifiquement immunoprécipitée.

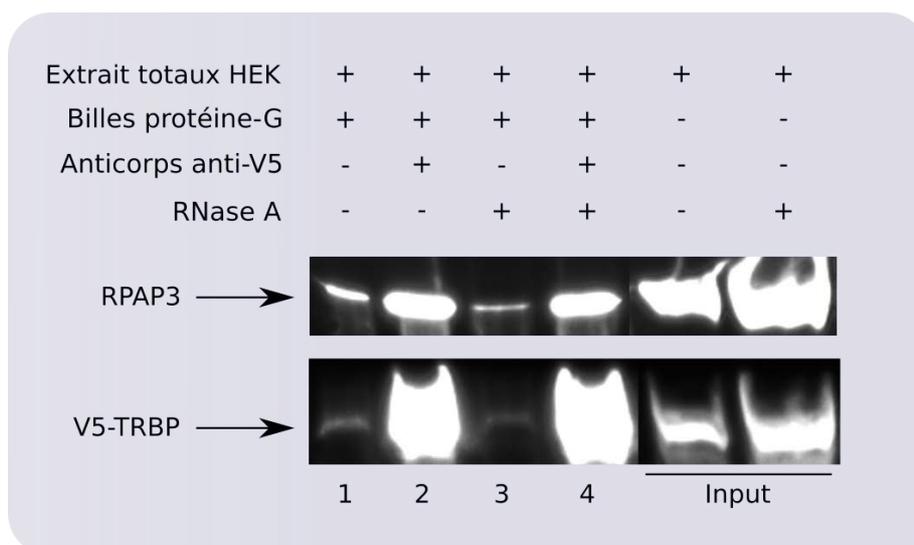


Figure 64: Les protéines RPAP3 et V5-TRBP forment un complexe *in vivo*.

Des extraits totaux de cellules HEK préalablement transfectées avec un vecteur permettant l'expression de la protéine TRBP fusionnée à une étiquette V5 ont été incubés avec des billes magnétiques couplées à la protéine G avec ou sans anticorps dirigé contre l'étiquette V5. La co-immunoprécipitation de la protéine RPAP3 endogène est vérifiée par un western blot à l'aide d'un anticorps dirigé contre RPAP3. La présence de la protéine V5-TRBP sur les billes est également vérifiée par western-blot à l'aide d'un anticorps anti-TRBP. L'expérience est réalisée avec et sans ajout de RNase A, afin de voir si l'ARN contribue à la formation du complexe TRBP/RPAP3 *in vivo*.

Les résultats montrent que la protéine RPAP3 est spécifiquement retenue lorsque la protéine V5-TRBP est immunoprécipitée (piste 3 et 4), ce qui confirme l'existence *in vivo* d'une association entre ces deux protéines, y compris en l'absence d'ARN (piste 4).

Les pistes « d'input » permettent de confirmer que les protéines RPAP3 et TRBP sont bien présentes au sein de l'extrait cellulaire utilisé.

Input : témoin d'expression interne (voir Matériel et Méthodes)

Les protéines RPAP3 et TRBP pouvant toutes deux s'associer à l'ARN, nous avons en outre voulu vérifier si, *in vivo*, l'interaction était dépendante ou non de la présence d'ARN. Pour cela, les expériences ont également été réalisées en présence de RNase A (**Figure 64**, piste 3 et 4). Les résultats obtenus, identiques en présence ou en absence de RNase A, nous ont permis de conclure que l'association entre les protéines RPAP3 et TRBP était indépendante de l'ARN.

Afin d'éliminer un biais possible dû à la rupture de la compartimentation cellulaire des protéines lors de la préparation des extraits, nous avons également réalisé des expériences de « colocalisation » *in vivo* par Duolink™ sur des cellules HeLa. Développé par la société Olink, cette technique permet de visualiser par microscopie confocale l'association *in situ* de deux protéines sous réserve qu'elles soient distantes au maximum d'environ 40 nm (**Figure 65**). Dans un premier temps, deux anticorps d'espèces différentes et spécifiques des protéines d'intérêt sont incubés avec des cellules fixées au formaldéhyde et préalablement perméabilisées. Ensuite, deux anticorps secondaires (ou sondes « Proximity Ligation Assay ») spécifiquement dirigés contre les anticorps primaires sont ajoutés. Les sondes PLA sont constituées d'anticorps dont la partie constante a été fusionnée avec une sonde oligodésoxyribonucléotide. Les deux oligodésoxyribonucléotides sont ensuite reliées par l'intermédiaire d'une ADN ligase et de deux amorces, si et seulement si elles sont assez proches l'une de l'autre, formant alors un « cercle » qui est ensuite amplifié à l'aide d'une ADN polymérase. La détection de l'association se fait grâce à une sonde nucléotidique spécifique de la région amplifiée et portant un fluorophore. Plusieurs sondes vont s'hybrider concentrant ainsi le signal émis par le fluorophore au niveau des complexes protéiques, qui sont visualisés sous forme d'un point fluorescent par microscopie confocale. La spécificité des anticorps primaire a été préalablement vérifiée par immunofluorescence, mais des témoins négatifs doivent également être réalisés afin de s'assurer de la spécificité du signal détecté. Il faut en effet vérifier que chacun des deux anticorps primaires seul ne permette pas de générer un signal en l'absence du second anticorps. De même, nous avons vérifié qu'aucun signal n'était détecté lorsque seules les sondes PLA sont utilisées. Nous avons également vérifié la spécificité de l'association entre RPAP3 et TRBP en réalisant le même type d'expérience avec la protéine GAPDH. En effet, cette protéine ne s'associe *a priori* pas avec TRBP et RPAP3, par conséquent, aucun signal ne devrait être détecté pour ces expériences de Duolink™.

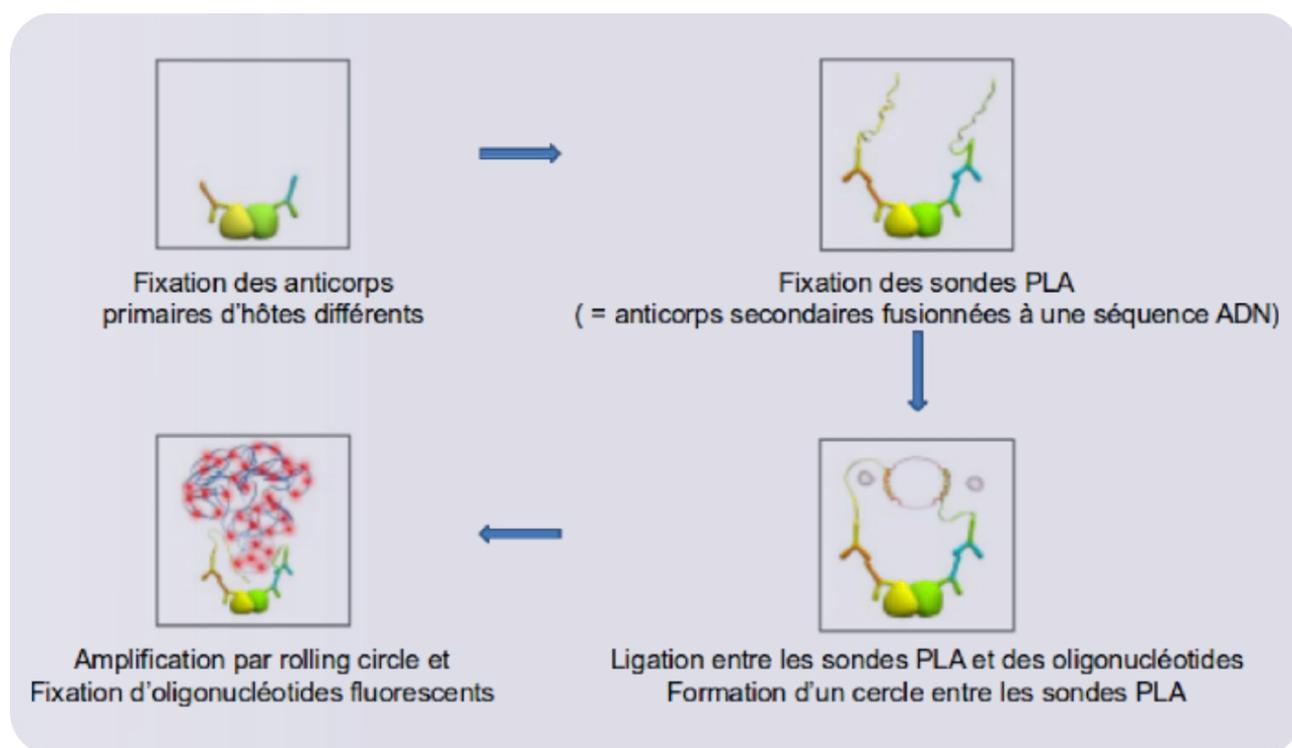


Figure 65: Principe du Duolink™.

Le Duolink™ est une technique développée par la société Olink qui permet de visualiser par microscopie confocale l'association *in cellulo* de deux protéines endogènes. Dans un premier temps, deux anticorps d'espèces différentes et spécifiques des protéines d'intérêt sont incubés avec des cellules HeLa préalablement fixées au formaldéhyde et perméabilisées.

Ensuite, deux anticorps secondaires ou sondes PLA spécifiques des espèces des anticorps primaires sont ajoutées. Ces anticorps sont fusionnés au niveau de leur partie constante avec un oligodéoxyribonucléotide. Les deux oligodéoxyribonucléotides vont ensuite être liés entre eux suite à l'action d'une ADN ligase et de deux amorces.

Si les deux anticorps secondaires sont assez proches l'un de l'autre, le produit de ligature pourra être amplifié. La détection de l'association se fait ensuite grâce à une sonde oligodéoxyribonucléotide complémentaire et spécifique de la séquence amplifiée et portant un fluorophore.

Les résultats présentés dans la **Figure 66** montrent un signal significatif uniquement pour l'association entre les protéines RPAP3 et TRBP endogènes. Ainsi, pour les témoins négatifs et pour l'association entre la GAPDH et TRBP ou RPAP3, seul un léger point fluorescent par cellule est parfois détecté. Contrairement à l'expérience entre TRBP et RPAP3, où de très nombreux points fluorescents sont détectés au niveau des cellules, confirmant ainsi l'existence d'une association au sein de ces cellules entre ces deux protéines. De plus, la réalisation conjointe de l'expérience de Duolink™ avec le marquage fluorescent du noyau des cellules (en bleu) et de leurs cytosquelettes (en rouge) permet d'obtenir la localisation subcellulaire du complexe TRBP/RPAP3. De manière intéressante, il semble que cette association soit ainsi principalement localisée dans le noyau, bien qu'observée également dans le cytoplasme. Ces différentes localisations restent toutefois cohérentes avec ce qui a pu être décrit pour les protéines RPAP3 et TRBP séparément. En effet, si les protéines RPAP3 et TRBP sont *a priori* principalement localisées, respectivement, au niveau du noyau et du cytoplasme, il a été montré d'une part que les différentes protéines du complexe hR2TP, dont RPAP3, étaient également présentes dans le cytoplasme (Boulon *et al.*, 2008) et d'autre part que TRBP était également présente dans le noyau (Gagnon *et al.*, 2014).

Ainsi, ces différentes expériences (double-hybride, GST pull-down, immunoprécipitation et Duolink™) nous ont permis d'une part d'identifier une nouvelle interaction entre les protéines RPAP3 et TRBP, qui sont respectivement impliquées dans l'assemblage de différentes RNPs, dont les snoRNP, et la maturation des microARN, et d'autre part de confirmer le caractère direct de cette interaction, ainsi que son existence *in vivo* dans des cellules HEK et HeLa. Nous avons ensuite souhaité identifier les domaines de TRBP et de RPAP3 responsables de l'interaction entre ces deux protéines par des approches biochimiques. Le but était également de mettre au point les conditions de production et de purification d'un complexe entre TRBP et RPAP3 dont la taille serait compatible avec l'utilisation d'approches par RMN en solution pour en déterminer la structure tridimensionnelle. De plus, l'identification de ces domaines au niveau des deux protéines pourrait en outre nous fournir de premières informations précieuses concernant les fonctions possibles de ce nouveau complexe. En effet, comme indiqué dans l'introduction, les protéines RPAP3 et TRBP ont déjà été décrites comme interagissant directement avec de nombreuses protéines.

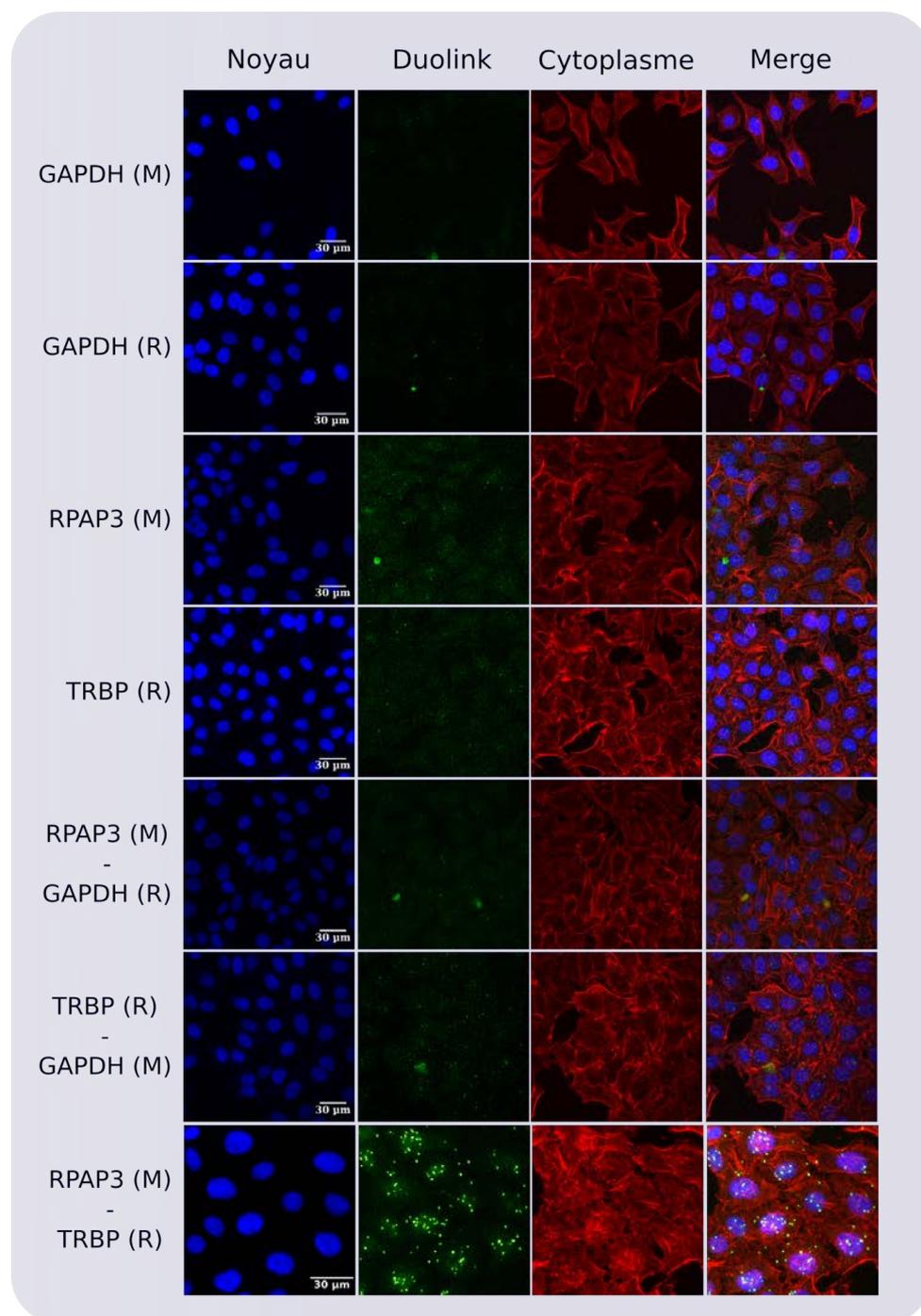


Figure 66: Les protéines RPAP3 et TRBP endogènes s'associent *in vivo*.

Expériences de Duolink™ réalisées sur les protéines TRBP, RPAP3 et GAPDH. Un point vert « brillant » correspond à la détection d'une association entre les deux protéines testées. Les expériences réalisées avec uniquement un anticorps dirigé contre les protéines RPAP3, TRBP ou GAPDH sont des témoins négatifs qui permettent de vérifier la spécificité du signal. Les noyaux et le cytosquelette des cellules ont été également marqués afin de déterminer la localisation subcellulaire du complexe TRBP/RPAP3. L'expérience RPAP3 (M)-TRBP (R) présente un nombre de points supérieurs au bruit de fond visible avec les expériences témoins, confirmant l'existence d'une association entre RPAP3 et TRBP *in vivo*. Le signal est principalement visible dans le noyau des cellules, bien que certains points soient également présents dans le cytoplasme, ce qui indique que le complexe serait présent au sein de ces compartiments cellulaires.

(M)/(R) : origine de l'espèce de l'anticorps primaire utilisé (Mouse/Rabbit).

Merge : superposition des 3 images obtenues par microscopie confocale.

Ainsi, RPAP3 interagit directement avec la protéine chaperonne HSP90 (Pal *et al.*, 2014) et les protéines du complexe hR2TP, notamment la protéine PIH1D1 (Yoshida *et al.*, 2013) et probablement les protéines TIP48/49 (Boulon *et al.*, 2008), tandis que TRBP interagit au minimum avec les protéines Dicer (Wilson *et al.*, 2015) et PACT, tout en étant également probablement capable de former des homodimères (Laraki *et al.*, 2008). Ainsi, isoler les domaines impliqués dans la formation du complexe entre TRBP et RPAP3 nous renseignerait alors sur la possibilité pour ces différentes protéines d'interagir ou non avec le complexe, fournissant ainsi une première indication sur la ou les fonctions de ce complexe protéique.

III. Caractérisation de l'interaction entre RPAP3 et TRBP

Au moment où nous avons démarré cette étude et initié la sélection des différents domaines putatifs pour ces deux protéines, les structures tridimensionnelles des deux domaines TPR de RPAP3 et du domaine DRBM3 de TRBP n'avaient encore été ni résolues ni publiées (voir Introduction). Nous avons donc utilisé des outils de prédiction de structure secondaire et les données de la littérature pour identifier et délimiter les différents domaines possibles.

A. Détermination des domaines fonctionnels putatifs de RPAP3

En l'absence de données structurales pour la protéine RPAP3, nous nous sommes basés sur les différentes données disponibles à ce moment-là dans les bases des données (notamment Uniprot) et sur les prédictions de structure secondaire réalisées alors. Nous avons ainsi, dans un premier temps, déterminé 4 domaines potentiellement fonctionnels pour RPAP3. Deux domaines N- et C-terminaux, prédits pour se structurer en hélices α , et deux domaines TPR (« Tetratricopeptide Repeat »). Cependant peu après le début de nos clonages, la résolution au laboratoire par le Dr Marc Quinternet de la structure tridimensionnelle de ces deux domaines TPR par RMN a montré que les bornes que nous avons choisies pour ces deux domaines ne comprenaient pas une dernière hélice α qui semble être importante pour la stabilité de ces domaines. En effet, les domaines TPR sont constitués de plusieurs motifs TPR eux-mêmes formés de 2 hélices α reliées par une boucle. Dans le cas de RPAP3, les deux domaines TPR étaient prédits pour contenir 3 motifs TPR, mais d'après les travaux du Dr Quinternet, une dernière hélice α est nécessaire afin de stabiliser l'ensemble du domaine TPR. De plus, nous avons initialement travaillé avec l'isoforme II de la protéine RPAP3, or, comme nous l'avons précédemment évoqué, cette isoforme correspond à d'épissage alternatif de l'isoforme I et contient une trentaine de résidus de moins juste en aval du second domaine TPR, qui

sont nécessaires pour l'interaction de RPAP3 avec PIH1D1 *in vivo*. Nous avons donc généré différentes constructions contenant ou non ces résidus, à partir d'une construction qui nous a été généreusement fournie par l'équipe d'Edouard Bertrand (IGMM, Montpellier). Nous avons ainsi sélectionné et cloné 7 nouveaux sous-domaines potentiellement fonctionnels de RPAP3 (**Figure 67**).

B. Détermination des domaines fonctionnels putatifs de TRBP

Pour la protéine TRBP, nous avons dans un premier temps isolé 3 fragments en procédant de la même façon que pour RPAP3. Ainsi, la base de donnée Uniprot répertorie 2 isoformes a et b de TRBP, qui comportent chacune 2 domaines de liaison à l'ARN et un domaine d'interaction protéine-protéine (voir Introduction) mais qui diffèrent au niveau de leur extrémité N-terminale (résidus 1 à 21 manquants dans l'isoforme b). Ces trois domaines DRBM ont donc été sélectionnés et clonés. De plus, les prédictions de structures secondaires pour le dernier domaine DRBM de TRBP (262-366) indiquent que ce domaine, qui généralement contient un feuillet β flanqué de deux hélices α , se trouverait plutôt entre les bornes 293 et 366. Un domaine plus petit étant généralement plus facile à étudier en RMN, nous avons donc également cloné un dernier fragment plus court (293-366) pour le domaine DRBM3 de TRBP (**Figure 68**).

C. Détermination des domaines impliqués dans l'interaction entre TRBP et RPAP3

Une fois les différents fragments déterminés, nous avons réalisé dans un premier temps un large crible de co-expression chez *E. coli* à l'aide d'un robot TECANTM, afin de tester un grand nombre de combinaisons et de conditions d'expression et de purification possibles. Dans un second temps, les interactions ainsi identifiées ont été confirmées *in vivo* par double-hybride chez la levure *S. cerevisiae*.

1. La technique de co-expression et de co-purification chez *E. coli*

a. Utilisation du robot TECANTM

Les retours d'expérience que nous avons au laboratoire montrent qu'*in vitro* l'interaction entre deux protéines est très souvent dépendante des conditions de salinité du tampon utilisé lors de la purification du complexe. Nous avons donc réalisé le crible de co-expression dans deux conditions, soit en présence d'une forte ou d'une faible concentration en NaCl (respectivement 400 et 50 mM).



Figure 67: Sélection de domaines potentiellement fonctionnels de RPAP3.

Ces domaines ont été sélectionnés à partir d'informations disponibles dans la base de données Uniprot mais également suite à la résolution des structures tridimensionnelles des domaines TPR par RMN par le Dr Marx Quinternet au laboratoire.

Les prédictions de structures secondaires indiquent que les domaines N- et C-terminaux seraient constitués de plusieurs hélices α . Les domaines TPR n°1 et 2 contiennent 3 motifs TPR, chacun composé de deux hélices α reliées par une boucle, auxquels s'ajoute une hélice α supplémentaire nécessaire à la stabilisation du domaine.

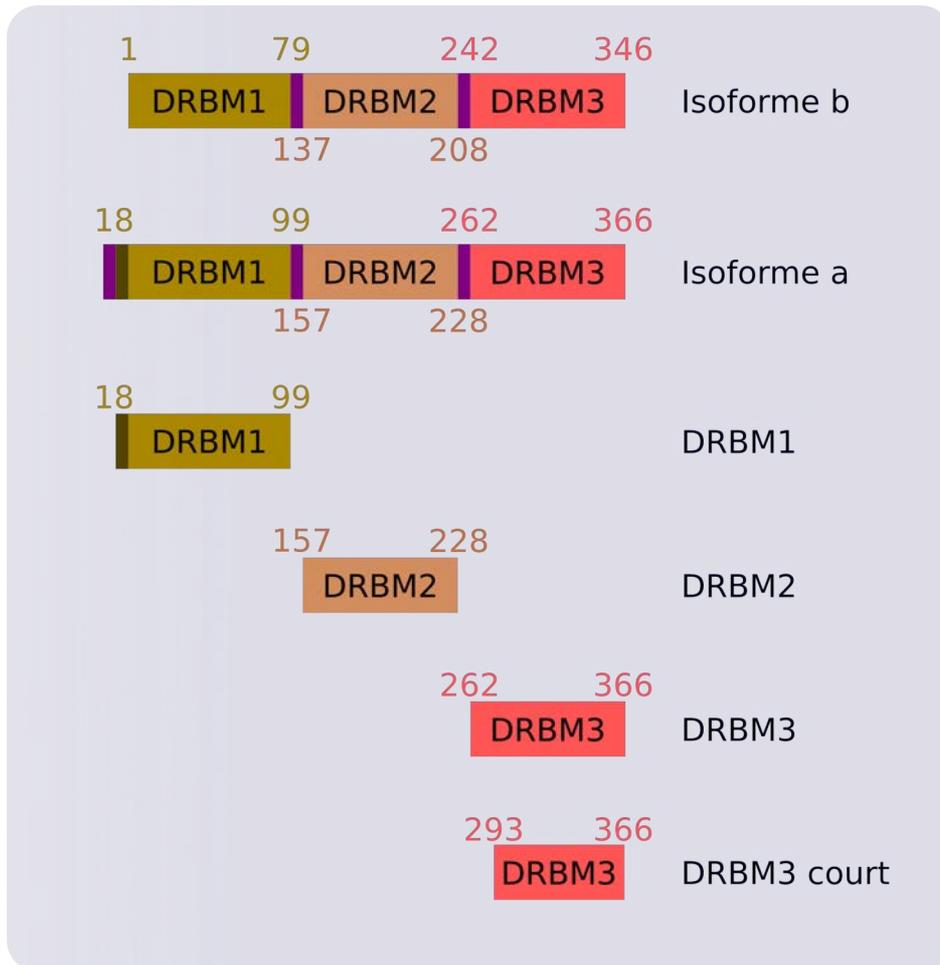


Figure 68: Sélection de domaines potentiellement fonctionnels de TRBP.

Comme pour RPAP3, les domaines putatifs de TRBP ont été sélectionnés à partir d'informations disponibles dans la base de données Uniprot et par des prédictions de structures secondaires.

La protéine TRBP est composée de trois domaines DRBM. Les deux premiers interagissent avec l'ARN alors que le dernier est connu pour être un domaine d'interaction protéines-protéines interagissant notamment avec la protéine Dicer (voir Introduction).

Différentes bornes ont été sélectionnées pour le dernier domaine DRBM3. En effet, le fragment 262-366 était annoté dans la base de données Uniprot comme nécessaire et suffisant pour l'interaction de TRBP avec Dicer, or des prédictions de structures secondaires ont montré que la structuration typique des domaines DRBM, hélice α – feuillet β – hélice α se situerait plutôt dans le fragment 293-366, qui a donc été également sélectionné.

De plus, la co-purification de complexe repose sur la production de deux protéines dont l'une est fusionnée à une étiquette 6-histidines, or la présence de cette étiquette pourrait altérer la formation du complexe, notamment en masquant les résidus impliqués dans l'interaction ou encore en générant d'importantes contraintes stériques. Nous avons donc testé ces interactions en produisant l'une ou l'autre des protéines (ou domaines) fusionnées à une étiquette 6-histidines. Enfin, il est également nécessaire de vérifier que les protéines (ou domaines) produites sans étiquette n'interagissent pas de manière non-spécifique avec la résine. Ainsi, pour chaque interaction testée, il est nécessaire de réaliser 6 tests. L'utilisation du robot TECAN™, qui nous permet de réaliser 96 expériences en une seule séquence, a alors permis un gain de temps considérable.

b. Réalisation des différentes constructions et création de vecteurs

Les ADNc codant les différentes protéines ou leurs sous-fragments ont été clonés dans les vecteurs de co-expressions pnEA3ch et pnCS permettant leur expression, respectivement avec ou sans étiquette 6-histidines au niveau de leur extrémité N-terminale. Comme nous ne disposions pas de vecteurs de co-expression compatibles avec l'approche de clonage Gateway®, le clonage des différents fragments et des protéines entières a été réalisé dans les vecteurs pnEA3ch et pnCS de manière classique (utilisation d'enzymes de restriction) par Marie-Ève Chagot. Toutefois, nous avons décidé de développer les vecteurs Gateway correspondant afin de disposer d'une gamme optimale de vecteurs compatibles. En effet, tous les autres vecteurs que nous avons utilisés pour le double-hybride, les « GST pull-down » et les immunoprécipitations sont compatibles avec le clonage Gateway® (vecteurs pAS2, pACT2, pDEST15, pDEST17 et pcDNA 3.1/nV5-DEST™). En parallèle des clonages conventionnels, nous avons donc généré la version Gateway® des vecteurs pnEA3ch, pnCS et pnYC. Ce dernier vecteur est notamment utilisé pour co-exprimer une troisième protéine lors de la purification de complexes composés de trois partenaires. Toutefois, il a tout d'abord été nécessaire de remplacer le gène de résistance au chloramphénicol du vecteur pnYC. En effet, la séquence insérée dans ces trois vecteurs pour les rendre compatibles avec le système de clonage Gateway® contient également ce gène. Nous l'avons remplacé par un gène de résistance à la kanamycine.

La construction du vecteur pnYK a été réalisée par recombinaison homologue à l'aide du kit In-Fusion™ de Clontech. Tout d'abord, le gène de résistance à la kanamycine a été amplifié par PCR à partir du vecteur pBlunt. Le vecteur pnYC a été amplifié sous forme linéaire par PCR à l'aide

d'oligodésoxyribonucléotides s'hybridant de part et d'autre du gène de résistance au chloramphénicol. Le gène de résistance à la kanamycine a alors été intégré au vecteur par recombinaison homologue (voir Matériel et Méthodes).

La construction de vecteurs de destination compatibles avec la technique de clonage Gateway® nécessite l'intégration, au niveau de la cassette de clonage du vecteur, d'une séquence comportant les deux séquences de recombinaison Gateway®, AttR1 et AttR2. Afin de rester au plus près des protocoles utilisés en routine au laboratoire, la cassette Gateway®, amplifiée à partir d'un autre vecteur Gateway® (pDEST17), a été clonée dans les vecteurs pnEA3ch, pnCS et pnYK entre les sites *NdeI* et *BamHI* utilisés classiquement. Ce clonage a nécessité la réalisation au préalable de deux mutagenèses dirigées sur le vecteur pDEST17, afin d'éliminer deux sites *BamHI* présents au sein de la cassette Gateway®. L'ajout de cette cassette Gateway® a pour principale conséquence d'entraîner la présence de 15 acides aminés supplémentaires au niveau de l'extrémité N-terminale de la protéine. Une fois ces vecteurs disponibles, nous y avons cloné les gènes codant les protéines TRBP et RPAP3 entières afin de valider l'utilisation de ces vecteurs pour les expériences de co-expression.

c. Résultats du crible de co-expression

Le crible de co-expression réalisé avec le robot TECAN™ a été effectué en collaboration avec l'équipe du Dr Xavier Manival. Bien que l'utilisation du robot nous permette de réaliser 96 expériences en une seule fois, il n'était pas envisageable de tester l'ensemble des combinaisons possibles pour déterminer les domaines de TRBP et de RPAP3 impliqués dans la formation du complexe. Nous avons donc sélectionné différents fragments et combinaisons de fragments devant nous permettre d'identifier les domaines nécessaires et suffisants pour la formation du complexe TRBP-RPAP3. Un témoin positif entre les protéines TRBP (isoforme b) et RPAP3 (isoforme II) a également été réalisé.

Les résultats des témoins négatifs (**Figure 69**) montrent qu'aucune de nos protéines ou fragments de protéines n'interagit de manière non spécifique avec les billes de nickel, à l'exception de la protéine TRBP entière (**Figure 69**, encadré rouge).

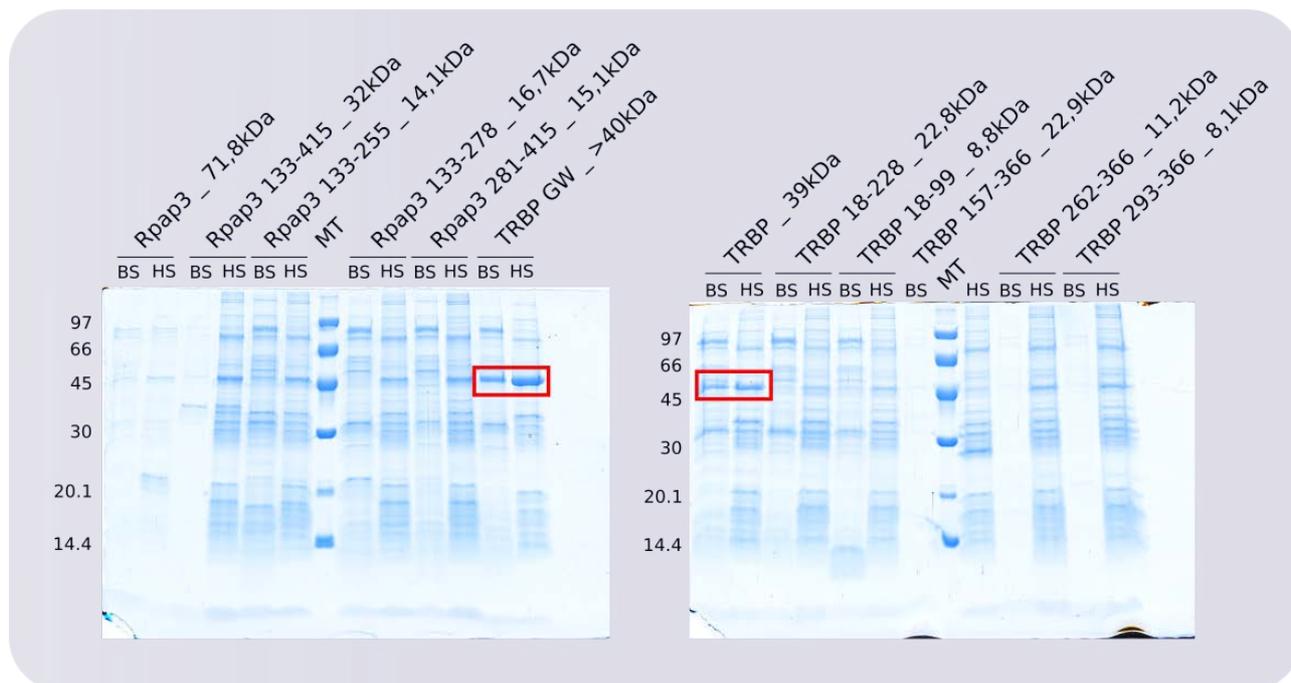


Figure 69: Témoins négatifs des tests de co-expression réalisés avec l'aide du robot TECAN™

Des bactéries *E. coli* BL21 pRAR2 ont été transformées par le vecteur pnCS codant les différents domaines sélectionnés ainsi que les protéines TRBP (isoforme b) et RPAP3 (isoforme II) entières. Après plusieurs étapes réalisées en grande partie par le robot TECAN™, les protéines sont ensuite incubées avec des billes magnétiques couplées aux ions Ni^{2+} . Ces protéines n'ayant pas d'étiquette 6-histidines, elles ne doivent théoriquement pas être retenues sur ces billes. Cette expérience permet alors de vérifier que ces protéines n'interagissent pas de manière non-spécifique avec les billes magnétiques.

Ainsi, ici aucune des protéines ou fragments testés n'interagit de manière non-spécifique avec les billes de nickel, à l'exception de la protéine TRBP entière (encadré rouge), qu'elle soit ou non produite à partir d'un vecteur Gateway®.

GW : L'ADNc codant la protéine a été insérée dans le vecteur pnCSGW.

Les tailles théoriques des différents fragments et protéines entières sont indiquées dans le nom de la piste.

HS/BS : le tampon de lyse contient respectivement 400 ou 50 mM de NaCl.

MT : marqueur de taille. La taille est indiquée en kDa.

Les autres résultats (**Figure 70**) confirment à nouveau l'interaction directe entre TRBP et RPAP3 (Gel 1, piste TRBP/RPAP3) et permettent de voir que l'interaction semble la plus stable dans le cas i) où c'est la protéine TRBP est fusionnée à l'étiquette 6-histidines et ii) pour une concentration en NaCl de 50 mM (piste BS – Bas Sel). Ces résultats sont également retrouvés avec les constructions réalisées avec les vecteurs pnEA3ch et pnCS Gateway® (gel n°2, piste TRBP GW/RPAP3 GW), validant l'utilisation de ces nouveaux vecteurs.

L'ensemble des autres résultats nous ont permis d'identifier les domaines 133-255 de RPAP3 (TPR n°1) et 293-366 de TRBP (DRBM3) comme nécessaires et suffisants pour la formation d'un complexe entre ces deux protéines (gel n°1, piste RPAP3 133-255/TRBP BS ; gel n°5, piste TRBP 293-366/RPAP3 133-278 BS et HS). Toutefois, l'interaction avec le domaine 133-255 de RPAP3 n'a été testée qu'avec la protéine TRBP entière, or un résultat positif n'est visible que dans le cas où la protéine TRBP n'est pas fusionnée à l'étiquette 6-histidines (Gel 1, piste RPAP3 133-255/TRBP BS). Ainsi, bien que le signal semble légèrement plus fort comparé au témoin négatif, il y a un risque que le résultat de l'interaction entre RPAP3 133-255 et TRBP ne soit qu'un faux positif, dû à l'interaction non spécifique de TRBP avec les billes magnétiques de nickel. Toutefois, le premier domaine TPR de RPAP3 semble bien être impliqué dans l'interaction avec TRBP puisque nous obtenons des résultats positifs avec le domaine 133-278 de RPAP3 (Gel 2, 4 et 5). Des interactions sont également retrouvées avec le domaine 262-366 de TRBP (Gel 5, piste TRBP 262-366/RPAP3 133-278). Ces premiers résultats indiquent que l'interaction entre RPAP3 et TRBP aurait lieu entre le premier domaine de RPAP3 et le domaine DRBM3 de TRBP. L'ensemble des combinaisons entre ces domaines n'ayant pu être testé avec le robot, nous avons souhaité vérifier ces résultats en réalisant une nouvelle série de co-expressions manuellement. Pour cela, le même type d'expérience a été réalisé sur des volumes de culture plus importants (200 ml contre 3 avec le robot) en utilisant la résine TALON™, substituée avec du cobalt (voir Matériel et Méthodes). Les domaines N- et C-terminaux de RPAP3 n'étant pas impliqués dans l'interaction, nous avons réalisé ces tests entre les deux domaines TPR de RPAP3 (133-255 et 281-415) et les trois domaines DRBM de TRBP (18-99, 157-228, 262-366 et 293-366).

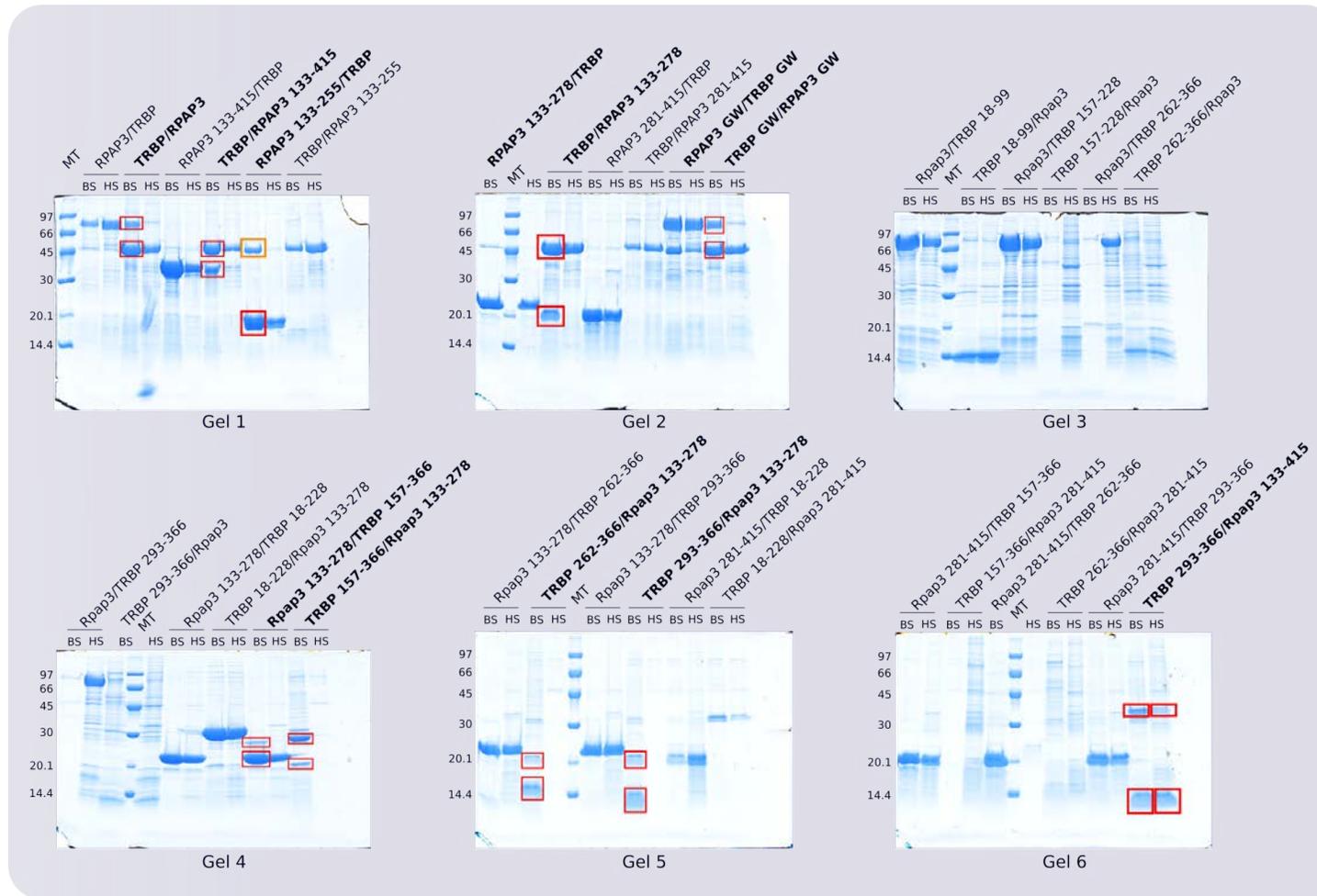


Figure 70: Résultats des tests de co-expression réalisés avec l'aide du robot TECAN™.

Des bactéries *E. coli* pRAR2 ont été co-transformées avec 2 vecteurs portant les séquences de TRBP (entière ou des domaines) et de RPAP3 (entière ou des domaines). L'une des protéines (la première indiquée dans la légende) est produite en fusion avec une étiquette 6-histidines (vecteur pnEA3ch ou pnEA3chGW) ce qui permet de la purifier sur des billes magnétiques de nickel. L'éluât a ensuite été fractionné sur un gel de protéines dénaturant afin de vérifier si la seconde protéine a été ou non co-purifiée (encadré rouge).

Les résultats montrent d'une part une interaction directe entre les protéines entières (gel 1, piste TRBP/RPAP3) et d'autre part que les domaines RPAP3 133-255, 133-278 et TRBP 262-366, 293-366 seraient suffisants pour permettre l'interaction entre ces deux protéines (gel 1, 2 et 5).

HS/BS : le tampon de lyse contient respectivement 400 ou 50 mM de NaCl.

MT : marqueur de taille. La taille est indiquée en kDa.

De plus, puisque l'interaction *in vitro* entre TRBP et RPAP3 n'est pas observée lorsque RPAP3 porte l'étiquette 6-histidines, probablement à cause de contraintes stériques, nous avons réalisé ces expériences uniquement avec la protéine TRBP, ou ses fragments, étiquetés. Enfin, les premières expériences préliminaires de production et de purification à plus grande échelle (1 litre) de ces domaines, réalisées par le Dr Quinternet, ont montré un meilleur rendement lors de l'utilisation de tampon de purification contenant une concentration en NaCl de 300 mM. Cette condition a donc été utilisée pour la suite de nos travaux.

De la même manière que pour le crible réalisé avec le robot, nous avons également réalisé des témoins négatifs (**Figure 71**). Les résultats de la co-expression sont présentés **Figure 72**, pour les domaines DRBM 1 et 2, et **Figure 73**, pour les domaines DRBM3 (262-366 et 293-366). Les résultats indiquent d'une part que les différents domaines que nous avons utilisés sont bien exprimés (Figure 71 à 73, pistes So) et solubles (Figure 71 à 73, pistes SnSo), et d'autre part que l'interaction entre RPAP3 et TRBP n'est retrouvée qu'entre les domaines 133-255 de RPAP3 (TPR n°1) et 262/293-266 de TRBP (DRBM3) (Figure 73, pistes F). Ainsi, nous confirmons qu'*in vitro*, les domaines 133-255 de RPAP3 et 293-366 de TRBP sont bien nécessaires et suffisants pour l'interaction entre TRBP et RPAP3.

Ces constructions ont alors été sélectionnées pour la suite de nos études structurales et fonctionnelles sur le complexe. Toutefois, les domaines 133-278 de RPAP3 et 262-366 de TRBP ont également été choisis. En effet, si des domaines plus courts sont généralement préférés pour des études par RMN, ils pourraient en revanche ne pas être suffisamment stables ou correctement repliés pour permettre leur étude structurale.

2. Les domaines TPR n°1 de RPAP3 et DRBM3 de TRBP forment un complexe stable *in vitro* et *in vivo*

a. Confirmation de l'existence d'un complexe stable *in vitro*

a. Confirmation de l'existence d'un complexe stable *in vitro*

Il est possible qu'un complexe visualisé par une approche de co-expression puisse se dissocier une fois purifié et élué de la résine. Afin de confirmer l'existence d'un complexe stable *in vitro* entre le TPR n°1 de RPAP3 et le domaine DRBM3 de TRBP, nous avons choisi de soumettre à une chromatographie d'exclusion stérique (voir Matériel et Méthodes) les différents complexes que nous avons préalablement sélectionnés.

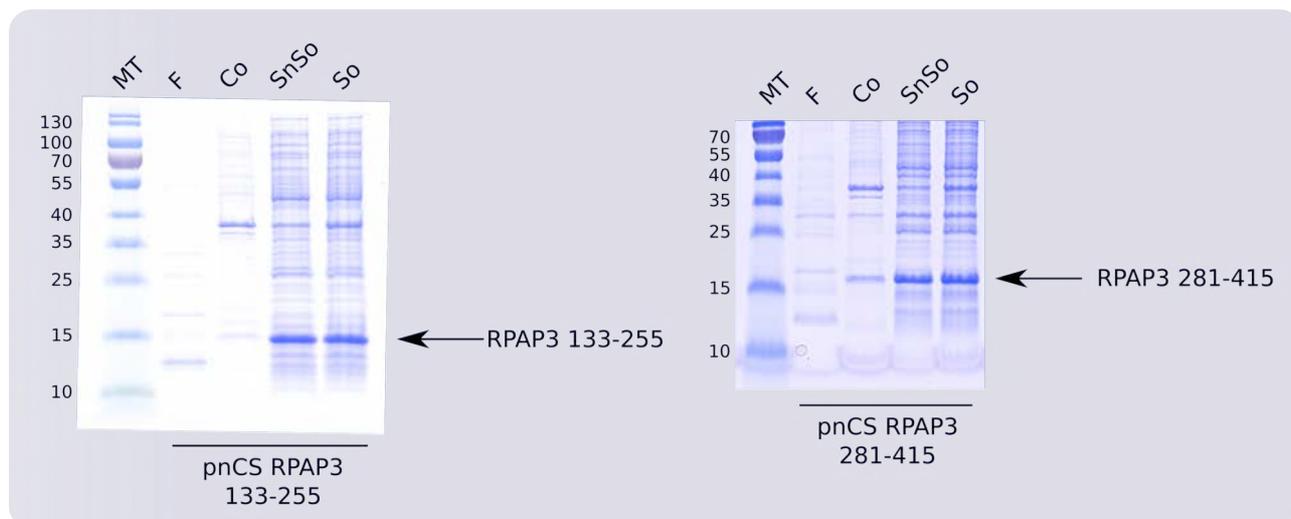


Figure 71: Témoins négatifs de co-expression.

Afin de pouvoir interpréter les co-expressions réalisées entre les domaines DRBM de TRBP et les deux domaines TPR de RPAP3, des témoins négatifs ont été réalisés afin de s'assurer que les domaines TPR n'interagissaient pas de manière non-spécifique avec la résine TALON™.

Au cours de la purification, des fractions sont prélevées à diverses étapes afin de vérifier que la protéine est exprimée (piste So), soluble (piste SnSo) ou non (piste Co). Ces fractions sont ensuite fractionnées sur un gel de protéines dénaturant. Une fraction de la résine TALON™ après purification est également fractionnée sur ce gel (piste F).

Les résultats montrent que les deux domaines TPR (133-255 et 281-415) de RPAP3 sont bien exprimés et solubles (piste SO et SnSo, respectivement), et qu'ils n'interagissent pas de manière non-spécifique avec la résine TALON™ (piste F).

MT : marqueur de taille. La taille est indiquée en kDa.

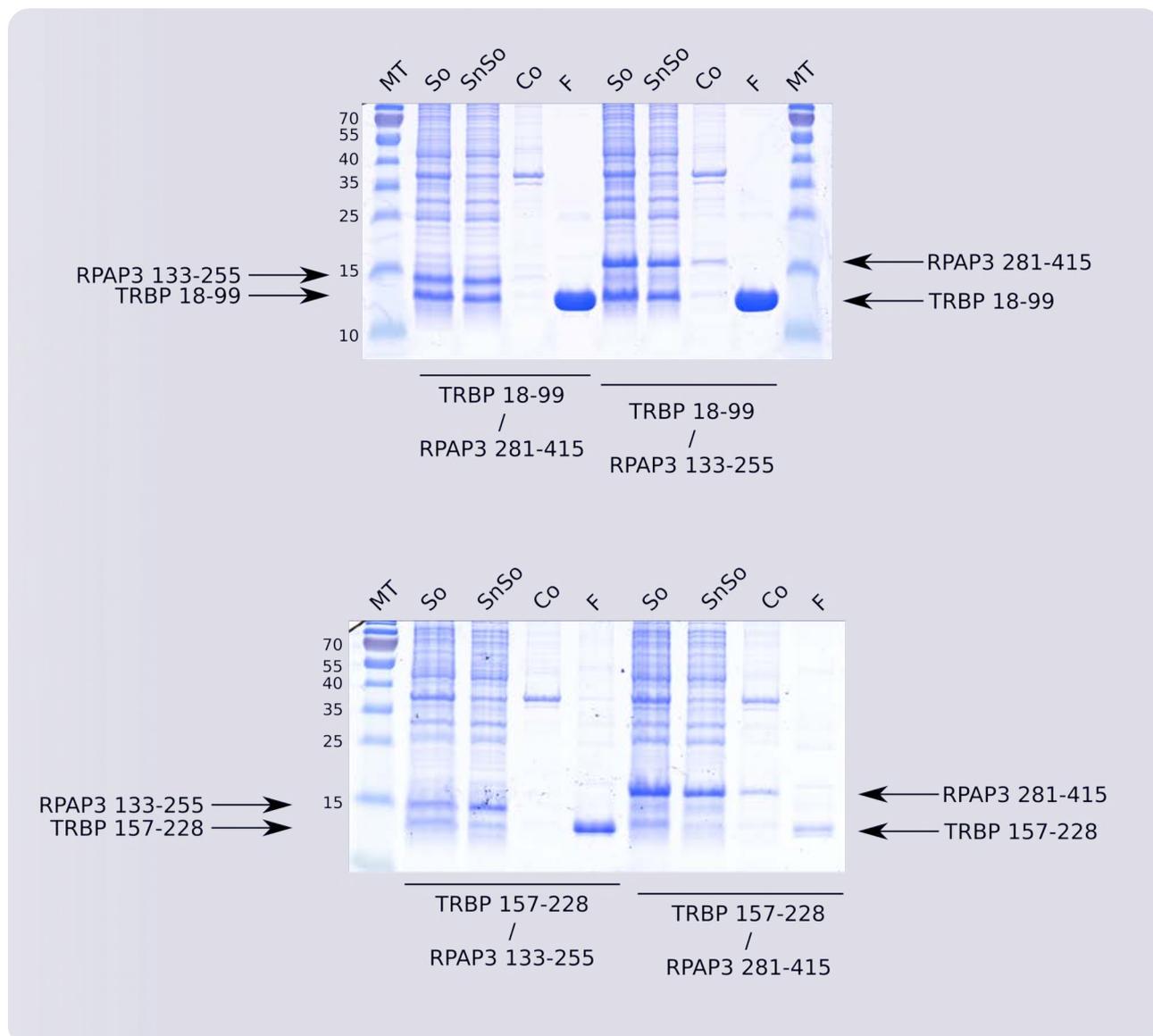


Figure 72: Co-expression entre les domaines DRBM1 et 2 de TRBP et les TPR n°1 et 2 de RPAP3.

Comme décrit Figure 71, lors de la co-purification de ces complexes, des aliquotes sont prélevées à diverse étapes afin de vérifier que les protéines soient bien exprimées (piste So), solubles (piste SnSo) ou non (piste Co). Une aliquote de la résine TALON™ après purification est également prélevée (piste F). L'ensemble de ces échantillons est ensuite fractionné sur un gel de protéines dénaturant. La présence de deux bandes de forte intensité au niveau de la piste F permet de conclure à une interaction directe entre les deux protéines testées.

Les résultats obtenus pour les domaines DRBM1 et 2 de TRBP (qui sont fusionnés à une étiquette 6-histidines) montrent qu'il n'y a pas d'interaction entre ces domaines et les domaines TPR de RPAP3. En effet, au niveau des pistes F, une seule bande, qui correspond au domaine DRBM1 ou 2 de TRBP fusionné à l'étiquette 6-histidines, est visible. Les domaines TPR de RPAP3 ne sont donc pas co-purifiés avec les domaines DRBM1 et 2 de TRBP.

MT : marqueur de taille. La taille est indiquée en kDa.

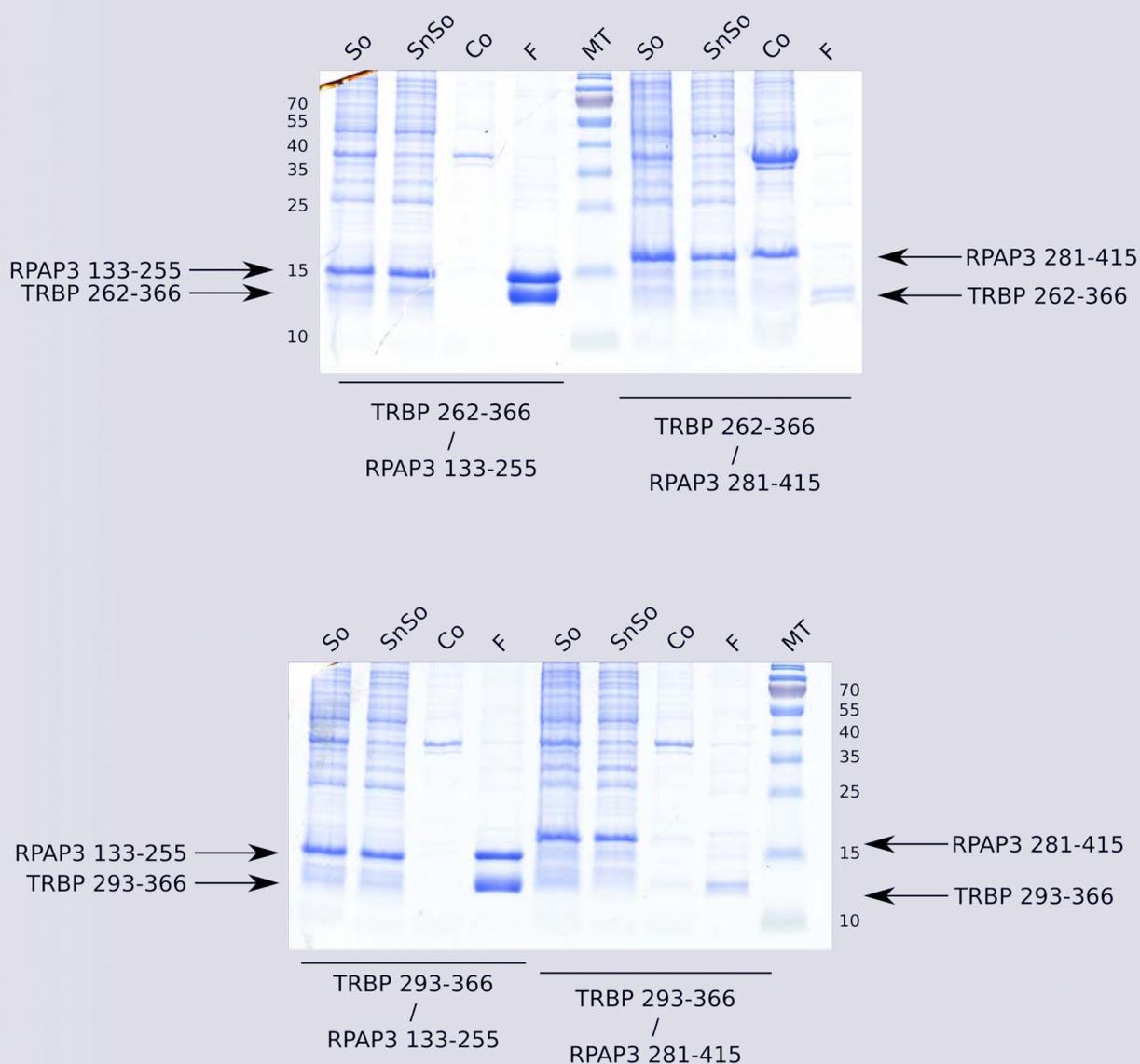


Figure 73: Co-expression entre les domaines 262-366 et 293-366 (DRBM3) de TRBP et les TPR n°1 et 2 de RPAP3.

Comme décrit précédemment, lors de la co-purification de ces complexes, des aliquotes sont prélevées à diverse étapes afin de vérifier que les protéines sont bien exprimées (piste So), solubles (piste SnSo) ou non (piste Co). Une aliquote de la résine TALON™ après purification est également prélevée (piste R). L'ensemble de ces échantillons est ensuite fractionné sur un gel de protéines dénaturant.

Les résultats obtenus pour le domaine DRBM3 (262-366 et 293-366) de TRBP (qui est fusionné à une étiquette 6-histidines) montrent clairement une interaction directe avec le TPR n°1 de RPAP3 (piste R). Dans le cas du domaine TPR n°2, seule la bande correspondant au domaine de TRBP, fusionné à une étiquette 6-histidines est visible.

MT : marqueur de taille. La taille est indiquée en kDa.

Dans un premier temps, la purification de ces différents complexes nous a permis d'identifier que les complexes TRBP 262-366/RPAP3 133-278 et TRBP 262-366/RPAP3 133-278 (**Figure 74** et **Figure 75**, respectivement) étaient les meilleurs candidats pour les approches structurales. En effet, les complexes formés avec le domaine 293-366 de TRBP ne semblent pas suffisamment stables après coupure de l'étiquette 6-histidines. Des essais ont été réalisés à l'aide d'un second protocole d'élution qui consiste à éluer le complexe à l'aide d'un tampon contenant 300 mM d'imidazole, ce protocole nous a permis de purifier le complexe mais malheureusement avec des rendements beaucoup trop faibles.

Une fois les protéines purifiées et l'étiquette 6-histidines retirée, nous avons vérifié que le complexe était toujours stable. Ainsi, après élution, le complexe a été concentré, puis injecté sur une colonne de chromatographie d'exclusion stérique (voir Matériel et Méthodes), afin de séparer les différentes molécules en fonction de leur taille. Le contenu de ces fractions collectées a ensuite été vérifié par électrophorèse sur un gel de protéines dénaturant. La réalisation d'une chromatographie d'exclusion stérique nous permet également de changer le tampon dans lequel se trouvent les complexes (tampon Hepes avec 300 mM de NaCl) par un tampon compatible avec les expériences de RMN (tampon Phosphate avec 150 mM de NaCl). La diminution de la concentration en NaCl ne semble pas altérer la stabilité du complexe d'après nos précédents résultats de co-expression.

La réalisation de ces expériences de chromatographie nous a ainsi permis de confirmer l'existence d'un complexe stable *in vitro* entre le domaine 262-366 de TRBP et le domaine 133-278 (ou 133-255) de RPAP3. Toutefois, le chromatogramme montre la présence de deux pics au niveau du volume d'élution correspondant au complexe (**Figure 76**, fractions 28 à 35), indiquant que le complexe ne serait pas pur. Un dosage au Nanodrop™ à 260 et 280 nm des fractions collectées nous a permis de vérifier qu'une partie du complexe était co-purifiée avec des acides nucléiques (**Figure 76**) et malheureusement la chromatographie d'exclusion stérique ne semble pas être assez résolutive pour séparer ces deux espèces (complexe avec et sans ARN). Or l'interaction, même non spécifique, d'acides nucléiques avec le complexe risque d'en modifier la structure tridimensionnelle. Après de multiples mises au point, nous sommes parvenus à séparer ces deux espèces en utilisant le PEI à une concentration de 0,05 % (contre 0,0125% initialement).

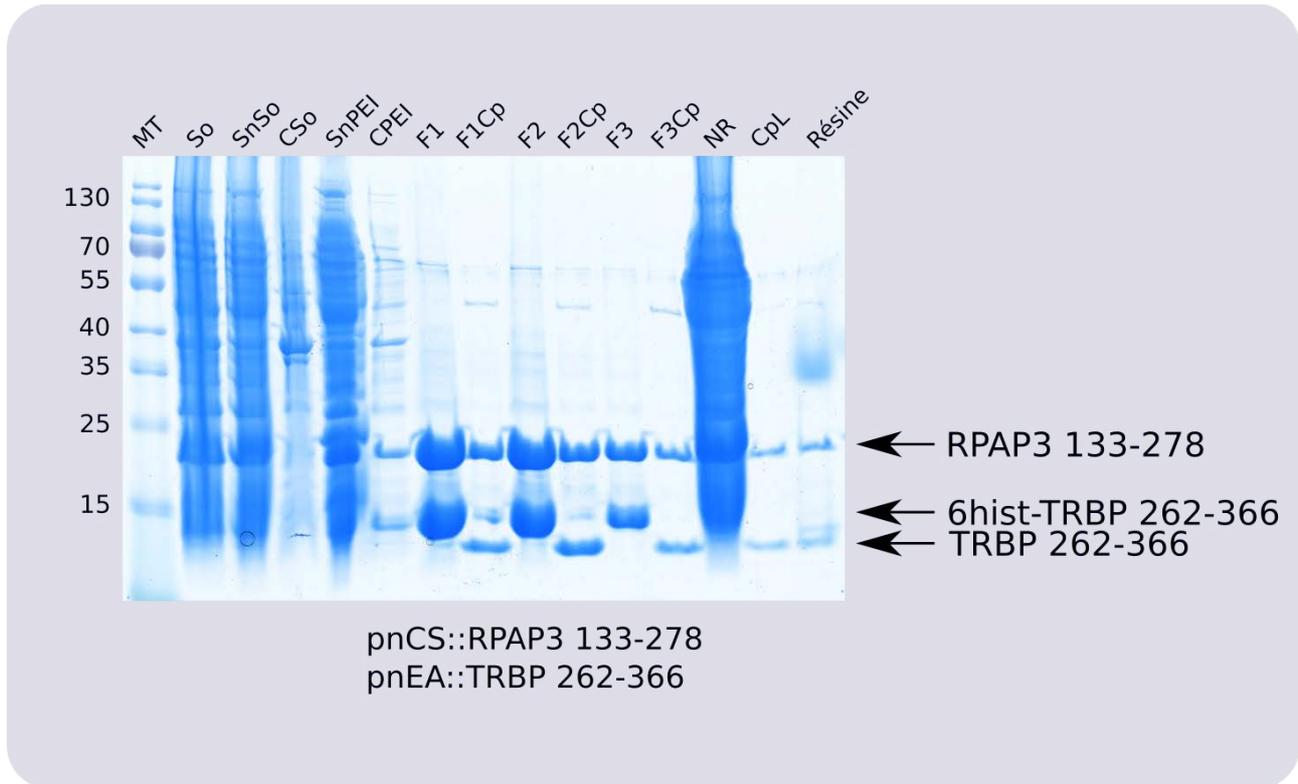


Figure 74: Production et purification du complexe TRBP 262-366/RPAP3 133-278 sur résine TALON™
Analyse sur gel dénaturant des diverses fractions collectées lors de la purification du complexe TRBP 262-366/RPAP3 133-278.

- MT : Marqueur de taille (les tailles sont indiquées en kDa).
 So : Extrait total après sonication.
 SnSo : Surnageant de sonication.
 Cso : Culot de sonication.
 SnPEI : Surnageant après traitement au PEI (0,05 %).
 CPEI : Culot après traitement au PEI.
 F1-F3 : Résine TALON™ après incubation avec le surnageant de PEI.
 F1Cp-F3Cp : Élué du complexe de la résine TALON™ après coupure de l'étiquette par la PreScission.
 NR : Surnageant après traitement au PEI après purification du complexe (Non Retenu).
 Cpl : Résines F1+F2+F3 lavées après élution.
 Résine : Résine F1+F2+F3 après élution.

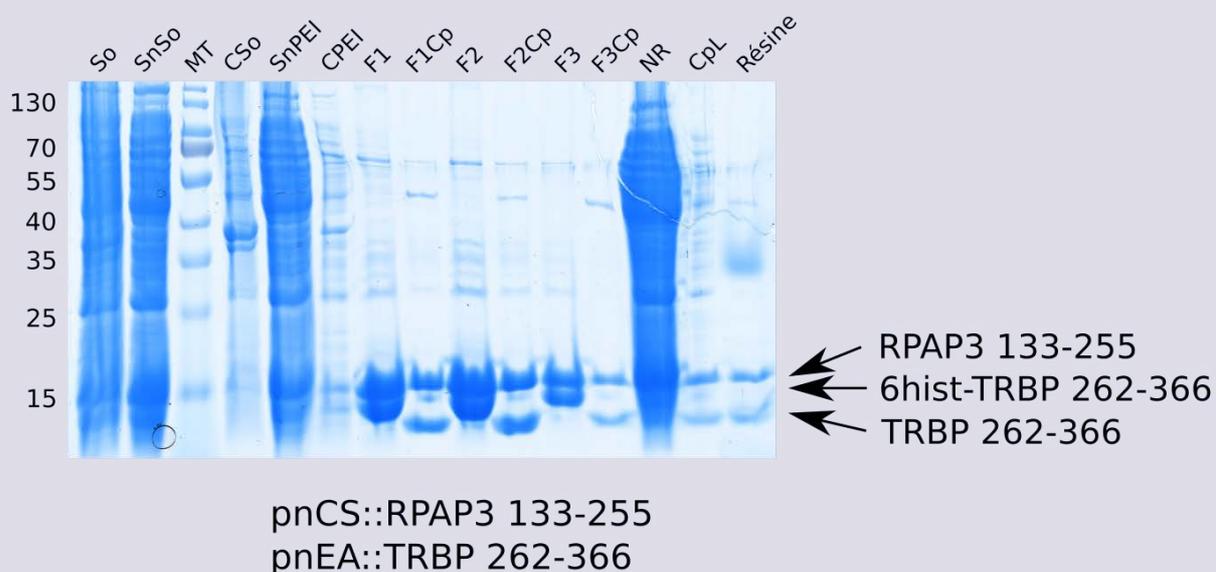


Figure 75: Purification du complexe TRBP 262-366/RPAP3 133-255 sur résine TALON™

Analyse sur gel dénaturant des diverses fractions collectées lors de la purification du complexe TRBP 262-366/RPAP3 133-255.

- MT : Marqueur de taille (les tailles sont indiquées en kDa).
 So : Extrait total après sonication.
 SnSo : Surnageant de sonication.
 Cso : Culot de sonication.
 SnPEI : Surnageant après traitement au PEI (0,05 %).
 CPEI : Culot après traitement au PEI.
 F1-F3 : Résine TALON™ après incubation avec le surnageant de PEI.
 F1Cp-F3Cp : Élué du complexe de la résine TALON™ après coupure de l'étiquette par la PreScission.
 NR : Surnageant après traitement au PEI après purification du complexe (Non Retenu).
 Cpl : Résines F1+F2+F3 lavées après élué.
 Résine : Résine F1+F2+F3 après élué.

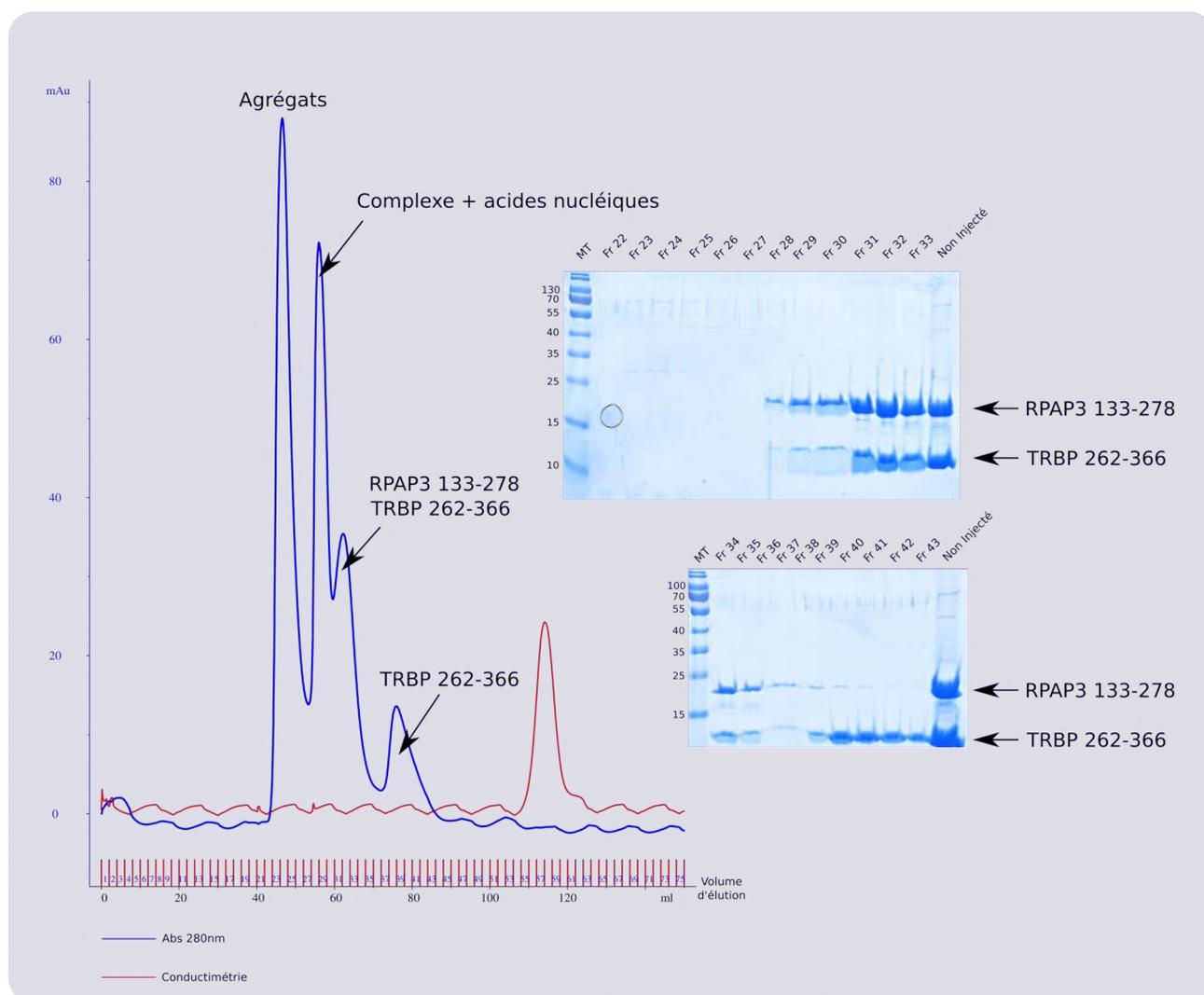


Figure 76: Purification du complexe TRBP 262-366/RPAP3 133-278 par chromatographie d'exclusion stérique.

Analyse par chromatographie d'exclusion stérique du complexe TRBP 262-366/RPAP3 133-278 précédemment purifié.

Les protéines contenues dans les fractions F1Cp, F2Cp et F3Cp (voir ci-avant) ont été combinées, concentrées, puis injectées sur une colonne s75 de 120 ml. Après élution, les différentes fractions ont été fractionnées sur un gel de protéines dénaturant.

Le chromatogramme et le gel indiquent que les fragments 262-366 de TRBP et 133-278 de RPAP3 forment un complexe stable *in vitro* qui est élué aux alentours de 60 ml. Toutefois, le pic d'élution du complexe est composé de 2 épaulements, qui correspondent au complexe purifiés avec ou sans acides nucléiques contaminants.

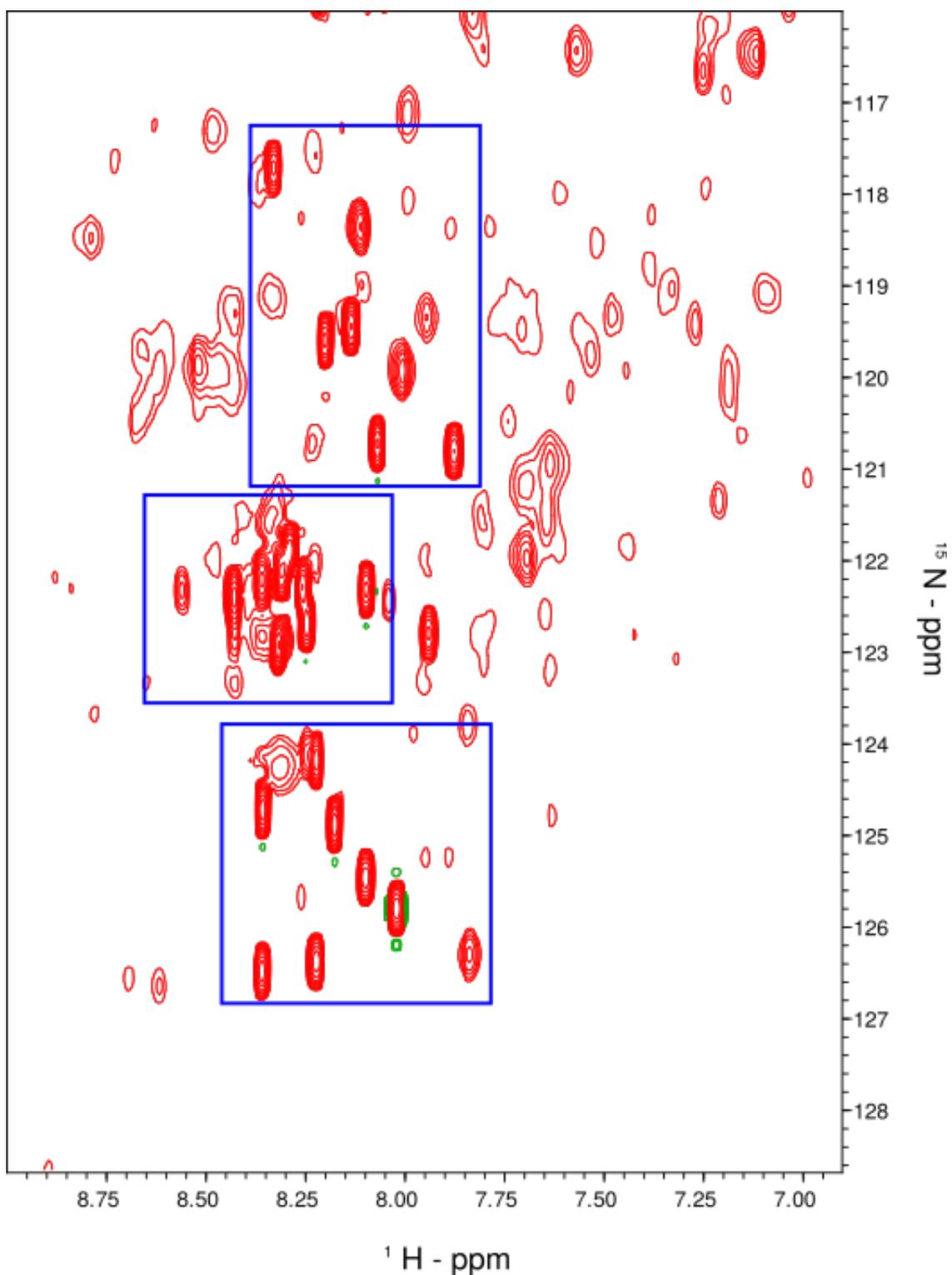
Le fragment 262-366 de TRBP étant produit en excès, une forme libre est également retrouvée (pic d'élution vers 75 ml).

La concentration en PEI ne peut pas être augmentée au-delà de ces valeurs, car elle provoquerait alors la précipitation du complexe, mais surtout entraînerait la réduction des atomes de nickel présent sur la résine TALON™, empêchant ainsi la purification.

b. Choix des bornes pour l'étude structurale du complexe entre le premier domaine TPR de RPAP3 et le domaine DRBM3 de TRBP

Les expériences de co-expression nous ont permis d'identifier différents fragments suffisants pour la formation du complexe TRBP-DRBM3/RPAP3-TRP1 : les fragments 262-366 et 293-366 de TRBP et les fragments 133-255 et 133-278 de RPAP3. Suite aux expériences de purification *in vitro*, nous avons éliminé le fragment 293-266 de TRBP dont le rendement de production était trop faible dans l'optique de la réalisation d'études structurales. Les complexes TRBP 262-366/RPAP3 133-255 et TRBP 262-366/RPAP3 133-278 ont alors été produits en présence d'une source azotée contenant de l'azote 15 (voir Matériel et Méthodes) afin de réaliser des expériences de RMN 2D HSQC (« Heteronuclear Single Quantum Coherence »). Cette expérience permet notamment de déterminer l'état de repliement des protéines étudiées. Une expérience HSQC ^{15}N - ^1H permet de corréler chaque noyau d'azote 15 (^{15}N) avec le ou les noyaux d'hydrogène auxquels il est lié de manière covalente. Il en résulte un spectre ^1H - ^{15}N où chaque « tâche », ou pic, correspond à un groupement NH (groupement amide de la chaîne principale, et les amides primaires des chaînes latérales). Grâce à cette expérience, il est ainsi possible de visualiser l'ensemble des acides aminés de la protéine, à l'exception des prolines qui contiennent un groupement amide secondaire et non primaire. Un spectre qui présente des pics peu intenses et bien dispersés est généralement synonyme d'une protéine (ou d'un complexe) correctement repliée, alors qu'un spectre présentant des pics de très forte intensité et groupés vers le centre du spectre sera plutôt caractéristique d'une protéine possédant des zones non structurées.

Le spectre HSQC ^{15}N - ^1H obtenu à 20 °C pour le complexe TRBP 262-366/RPAP3 133-278 montre clairement la présence de zones non structurées au sein de ce complexe (**Figure 77**). Celui du complexe TRBP 262-366/RPAP3 133-255 présentait de son côté quelques pics superposés, ce qui traduisait ma probable difficulté à poursuivre l'étude structurale de ce complexe par RMN. La qualité du spectre HSQC ^{15}N - ^1H pour ce second complexe a toutefois pu être améliorée en réalisant les expériences de RMN à une température de 30 °C. En revanche, même à cette température, le spectre HSQC ^{15}N - ^1H du complexe TRBP 262-366/RPAP3 133-278 montrait toujours d'importantes zones non structurées.



Suite à ces premiers résultats de RMN, le complexe TRBP 262-366/RPAP3 133-255 a été choisi pour poursuivre l'étude structurale de l'interaction entre TRBP et RPAP3 par RMN. Il a également été décidé de poursuivre ces études en réalisant les expériences de RMN à 30 °C.

c. Mise en évidence de l'existence d'un complexe stable *in vivo* par un test d'interaction double-hybride

Toutefois, avant de poursuivre ces études, nous avons vérifié que les domaines 262-366 de TRBP et 133-255 de RPAP3 formaient bien un complexe stable *in vivo*. Pour cela, nous avons réalisé une nouvelle expérience de double-hybride chez la levure *S. cerevisiae*. Lors de ces essais, nous avons en outre testé de nouveaux éléments, tels l'isoforme I de RPAP3, ainsi que le fragment 279-449 (voir **Figure 67**), qui n'avaient pas été testés jusque-là. Cette protéine et ce fragment contiennent tous deux un domaine supplémentaire qui est nécessaire pour l'interaction de RPAP3 avec PIH1D1 (Yoshida *et al.*, 2013). L'ensemble des ADNc codants les différents fragments ont été clonés dans les vecteurs Gateway[®] pAS2 et pACT2. Un témoin positif d'interaction entre Dicer et TRBP a également été ajouté.

La **Figure 78** montre que pour les expériences de double-hybride avec la protéine TRBP entière, les levures sont capables de se développer sur milieu dépourvu d'histidine uniquement lorsqu'elles sont co-transformées avec le vecteur pAS2-TRBP et les vecteurs pACT2-Dicer, pACT2-RPAP3 (isoforme I ou II), ou pACT2-RPAP3 133-255. Les levures peuvent survivre en présence d'une forte concentration en 3-AT, ce qui indique que l'association entre ces protéines (ou domaines) est très stable. La **Figure 79** montre le même type de résultats pour le domaine 262-366 de TRBP. Toutefois, les levures co-transformées avec les vecteurs pAS2-TRBP 262-366 et pACT2-RPAP3 iso II ne se développent que jusqu'à une concentration en 3-AT de 20 mM, ce qui peut indiquer que l'interaction entre ces deux protéines est moins stable que celles avec la protéine Dicer, le domaine 133-255 de RPAP3 ou l'isoforme I de RPAP3. Ainsi, la protéine TRBP s'associe *in vivo* indifféremment avec l'isoforme I ou II de RPAP3 et avec le domaine TPR n°1 mais pas avec son domaine TPR n°2, que le domaine d'interaction avec PIH1D1 soit ou non présent (**Figure 78**). De même, ces résultats montrent que le domaine 262-366 de TRBP est suffisant pour permettre l'association *in vivo* de RPAP3 (entière ou TPR n°1) avec TRBP (**Figure 79**).

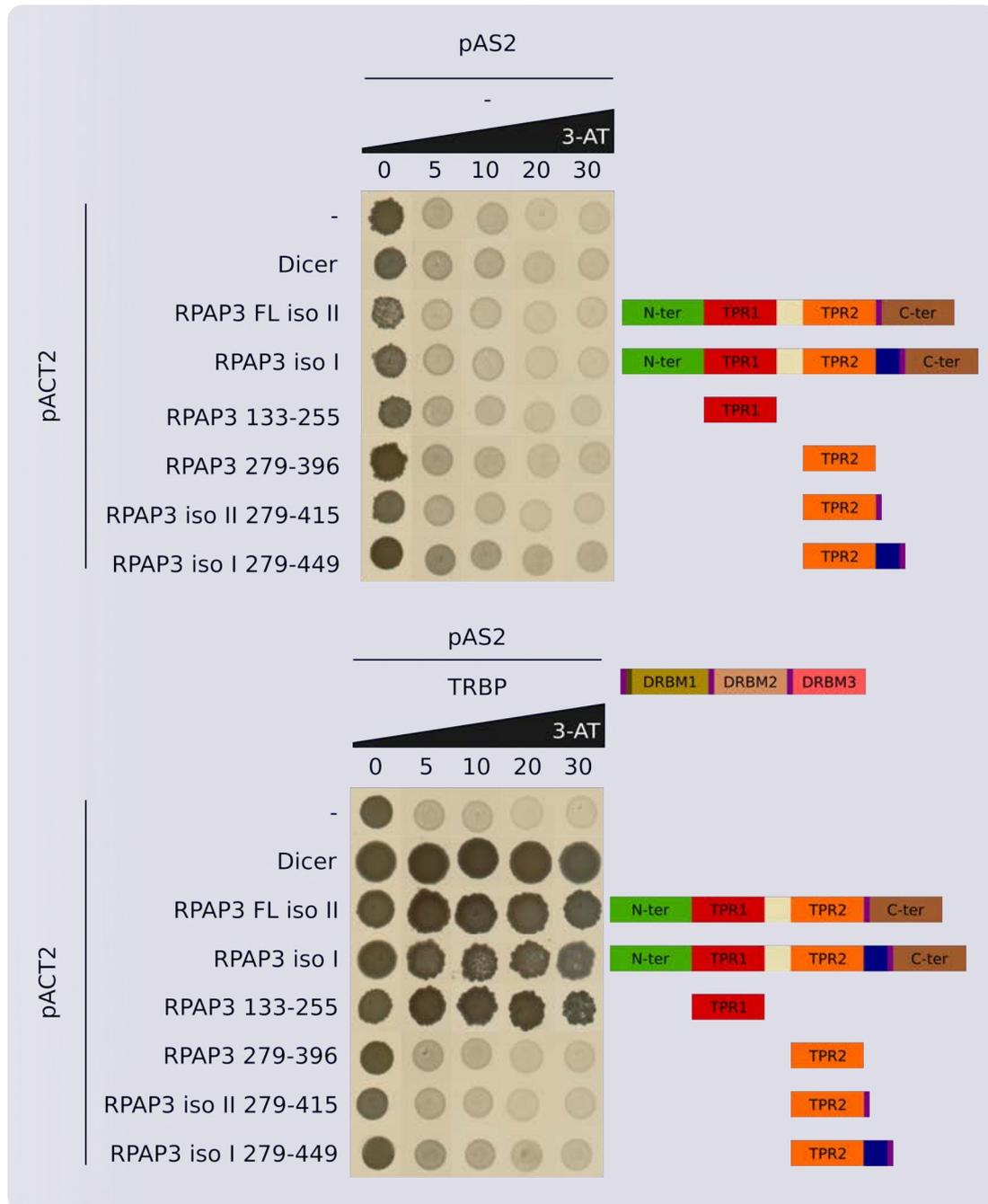


Figure 78: Le domaine 133-255 de RPAP3 est impliqué dans l'interaction de RPAP3 avec TRBP *in vivo*.

Les levures *S. cerevisiae* transformées avec le vecteur pAS2-TRBP survivent sur un milieu dépourvu en histidine lorsqu'elles sont co-transformées avec les vecteurs pACT2-Dicer, pACT2-RPAP3 (isoforme I et II) ou pACT2-RPAP3-133-255, et ce même en présence d'une forte concentration en 3-AT. Ces tests d'interaction en double-hybride montrent donc qu'*in vivo*, la protéine TRBP peut s'associer avec les protéines Dicer, RPAP3 et surtout avec le domaine TPR n°1 de RPAP3, confirmant ainsi les résultats obtenus *in vitro*.

0-5-10-20-30 : différentes concentration en 3-AT, en mM.

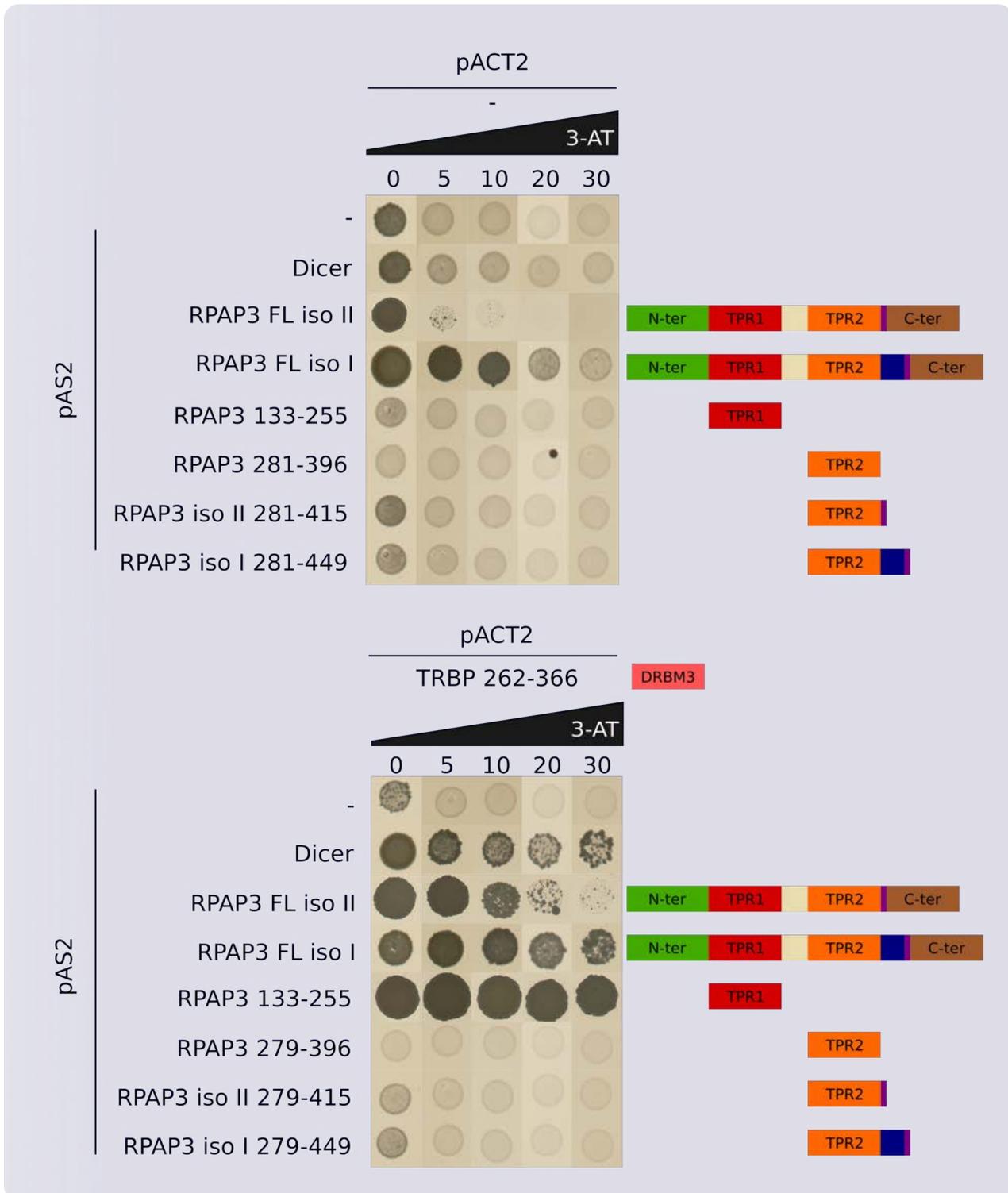


Figure 79: Les domaines 133-255 de RPAP3 et 262-366 de TRBP sont impliqués dans la formation du complexe entre ces deux protéines *in vivo*.

Les levures *S. cerevisiae* transformées avec le vecteur pACT2-TRBP-262-366 survivent sur un milieu dépourvu en histidine lorsqu'elles sont co-transformées avec les vecteurs pAS2-RPAP3 (isoforme II et I, bien que le témoin négatif de cette dernière soit positif jusqu'à 10 mM de 3-AT) ou pAS2-RPAP3-133-255, et ce même en présence d'une forte concentration en 3-AT. Ces tests en double-hybride montrent donc qu'*in vivo*, la protéine TRBP peut s'associer avec la protéine RPAP3 et surtout avec le domaine TPR n°1 de RPAP3, confirmant ainsi les résultats obtenus *in vitro*.

0-5-10-20-30 : différentes concentration en 3-AT, en mM.

D. Conclusion

L'objectif principal de cette seconde partie était d'identifier les domaines de TRBP et de RPAP3 impliqués dans la formation du complexe entre ces deux protéines. Pour cela, nous avons dans un premier temps identifié plusieurs domaines potentiellement fonctionnels, principalement à l'aide des informations disponibles dans la base de donnée Uniprot, et de prédictions de structure secondaire. Ainsi, la protéine TRBP a été divisée en trois domaines, correspondant aux trois domaines DRBM, et la protéine RPAP3 en 4 domaines, correspondant à deux domaines N- et C-terminaux et à deux domaines TPR. La résolution, au laboratoire, de la structure tridimensionnelle des domaines TPR de RPAP3 par RMN par le Dr Quinternet dans le cadre d'un autre projet de recherche, nous a permis d'affiner nos choix pour la détermination des bornes de ce domaine. De même, des prédictions de structures secondaires de TRBP ont abouti à l'utilisation d'extrémités alternatives pour le domaine DRBM3.

De manière intéressante, la résolution de la structure du domaine DRBM3 de TRBP par diffraction des rayons X par Wilson *et al.* (voir paragraphe IV. A. 2 ci-après) a confirmé, *a posteriori*, nos choix pour l'utilisation du domaine 262-366 de TRBP. En effet, ils ont montré par digestion ménagée que le domaine 258-366 de TRBP comprenait d'une part le domaine DRBM (fragment 289-363) et d'autre part un domaine N-terminal (258-289) faisant le lien entre les domaines DRBM2 et 3. Seul la structure du fragment 289-366 a pu être résolue, la densité électronique obtenue pour le domaine 258-289 étant trop faible pour être construite (Wilson *et al.*, 2015). De plus, les auteurs indiquent que le domaine 258-289 serait nécessaire à la fois pour stabiliser le domaine DRBM3 et pour le protéger de la dégradation par des protéases. Les difficultés que nous avons rencontrées lors de la production du domaine 293-366, alors que nous avons montré qu'il était nécessaire et suffisant pour permettre l'interaction de TRBP avec RPAP3, pourraient ainsi s'expliquer par l'absence de ce fragment N-terminal, ce qui déstabiliserait alors le domaine DRBM.

Grâce à l'utilisation d'un robot TECAN™, nous avons été en mesure de tester par co-expression et co-purification chez *E. coli* un grand nombre de combinaisons possibles entre tous ces différents domaines. Les résultats préliminaires indiquaient une interaction probable entre le domaine TPR n°1 de RPAP3 et le domaine DRBM3 de TRBP. L'interaction *in vitro* et *in vivo* entre les domaines 133-255 de RPAP3 (TPR n°1) et 262-366 (DRBM3) de TRBP a été par la suite confirmée par co-expression et double-hybride (**Figure 80**).

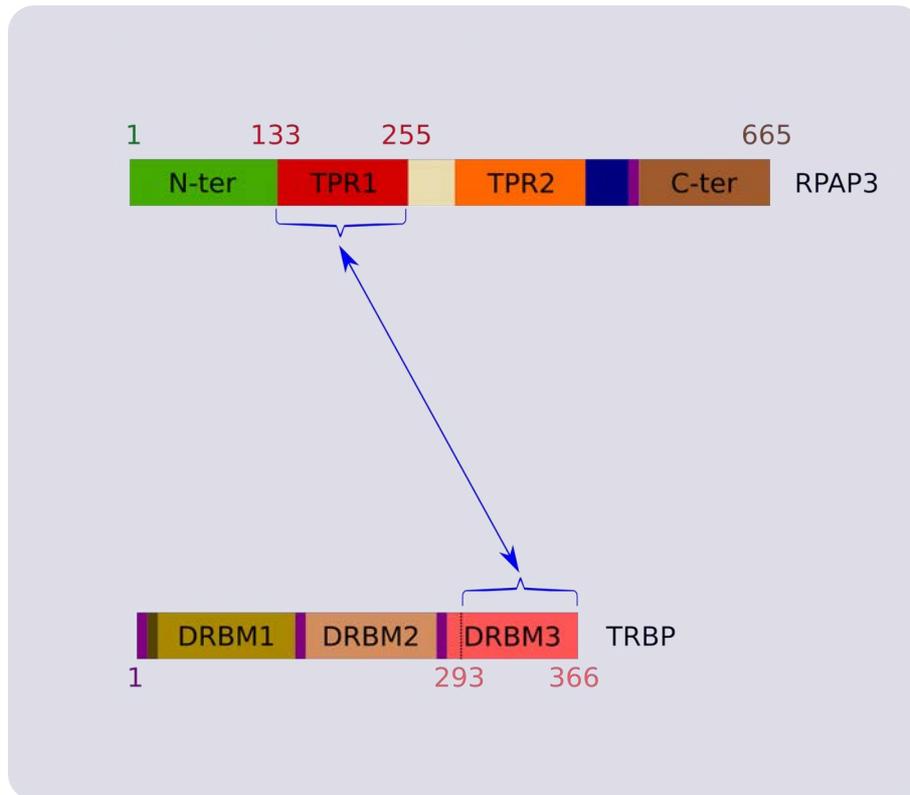


Figure 80: L'interaction directe entre RPAP3 et TRBP se fait entre les domaines DRBM3 de TRBP et le TPR n°1 de RPAP3.

L'ensemble de nos expériences nous ont permis d'identifier les domaines 133-255 de RPAP3 et 293-366 de TRBP comme nécessaires et suffisants pour permettre la formation du complexe entre ces deux protéines *in vitro* et *in vivo*.

Ces expériences nous ont également permis de mettre au point les conditions de production et de purification les plus efficaces pour obtenir un complexe TRBP DRBM3/RPAP3 TPR1 en quantité suffisante pour des études RMN 2D et 3D. Et les premières expériences RMN que nous avons alors réalisées ont permis d'éliminer certains fragments qui soit se produisaient très mal (DRBM3 293-366), soit comportaient des zones non structurées (TPR n°1 133-278).

L'objectif suivant de notre travail était de résoudre par RMN la structure tridimensionnelle du complexe TRBP 262-366/RPAP3 133-255, mais surtout, d'identifier les acides aminés impliqués dans cette interaction. En effet, la détermination ou la modélisation de l'interface d'interaction entre ces deux domaines devaient nous permettre d'identifier et d'isoler des variants afin de réaliser une analyse structure-fonction approfondie, dans le but d'obtenir des informations sur la fonction biologique du complexe TRBP/RPAP3.

Nous avons choisi d'utiliser la RMN pour résoudre cette structure pour plusieurs raisons. D'une part, bien que le complexe soit un peu grand pour la RMN (228 résidus), sa production dans un milieu contenant 50 % d'eau lourde D₂O a permis d'améliorer la définition des pics. En effet, la qualité d'un signal RMN dépend en partie du nombre de noyaux qui vont être détectés. Plus ce nombre est important, moins le signal va être visible. Dans un milieu deutéré à 50 %, une partie (en théorie 50%) des atomes d'hydrogènes vont être remplacés par du deutérium qui ne va alors plus contribuer au signal capté par la RMN, diminuant ainsi artificiellement le nombre de noyaux de la protéine. D'autre part, la structure tridimensionnelle à 30 °C du domaine 133-255 de RPAP3 libre avait déjà été résolue au laboratoire par le Dr. Quinternet, facilitant ainsi le travail d'attribution sur le complexe. Enfin, contrairement à la cristallographie, les expériences de RMN permettent d'obtenir de précieuses informations bien avant d'avoir résolu entièrement la structure tridimensionnelle.

IV. Études structurales du complexe TRBP DRBM3/RPAP3 TPR n°1

A. Identification des résidus de RPAP3 et de TRBP impliqués dans la formation du complexe protéique

1. Utilisation des données RMN pour identifier des acides aminés de RPAP3 candidats pour leur implication dans l'interaction avec TRBP

Le fait de disposer de l'attribution complète de l'un des partenaires, RPAP3, nous offrait la possibilité de pouvoir comparer rapidement les spectres RMN 2D HSQC ^{15}N - ^1H avant et après la formation du complexe entre RPAP3 et TRBP, afin d'identifier rapidement les acides aminés dont l'environnement spatial est modifié suite à la formation du complexe. En effet, le principe général de la RMN est de mesurer le temps nécessaire pour un noyau atomique particulier (^1H , ^{15}N , ^{13}C) de revenir à son état d'équilibre après avoir été excité par une impulsion électromagnétique. Ce temps de relaxation est ensuite converti en fréquence puis en déplacement chimique (mesuré en ppm) par l'utilisation d'une transformée de Fourier. Le déplacement chimique d'un noyau étant notamment dépendant de son environnement spatial, un noyau qui ne serait pas impliqué dans la formation d'un complexe verrait son déplacement chimique nettement moins affecté qu'un noyau situé au niveau de l'interface.

L'ensemble des atomes du squelette protéique du complexe a été attribué à partir des spectres RMN fournis par le Dr Quinternet réalisés sur un échantillon deutéré à 50 % et doublement marqué (^{15}N , ^{13}C) d'une part, et d'autre part des données issues de l'attribution du domaine RPAP3 133-255 libre précédemment réalisée par le Dr Quinternet. Les différentes expériences RMN, HSQC ^{15}N - ^1H , HNCA, HNCACB, HNCO, HNCOCACB et HNHA, ont été réalisées à 30 °C. Les spectres obtenus ont alors été utilisés pour attribuer les différents groupements du squelette peptidique du complexe. Ainsi, comme indiqué précédemment, les expériences HSQC ^{15}N - ^1H ont été utilisées pour attribuer les groupements NH des différents acides aminés. Les expériences HNCA permettent d'attribuer, pour chaque groupement NH, les différents carbones alpha qui y sont liés de manière covalente, ce qui correspond donc aux $\text{C}_{\alpha i}$ et $\text{C}_{\alpha i-1}$ (**Figure 81**). De la même manière, les expériences HNCACB permettent d'attribuer l'ensemble des carbones alpha et bêta liés à un groupement NH (donc $\text{C}_{\alpha i}$, $\text{C}_{\alpha i-1}$, $\text{C}_{\beta i}$ et $\text{C}_{\beta i-1}$). Les expériences HNCO et HNHA sont, comme leur nom l'indique, utilisées pour attribuer les atomes Co et H alpha. Enfin, les expériences HNCOCACB vont, comme pour les expériences HNCACB permettre l'attribution des $\text{C}_{\alpha i}$ et $\text{C}_{\beta i}$, mais, contrairement à ces dernières, uniquement pour les résidus $i-1$.

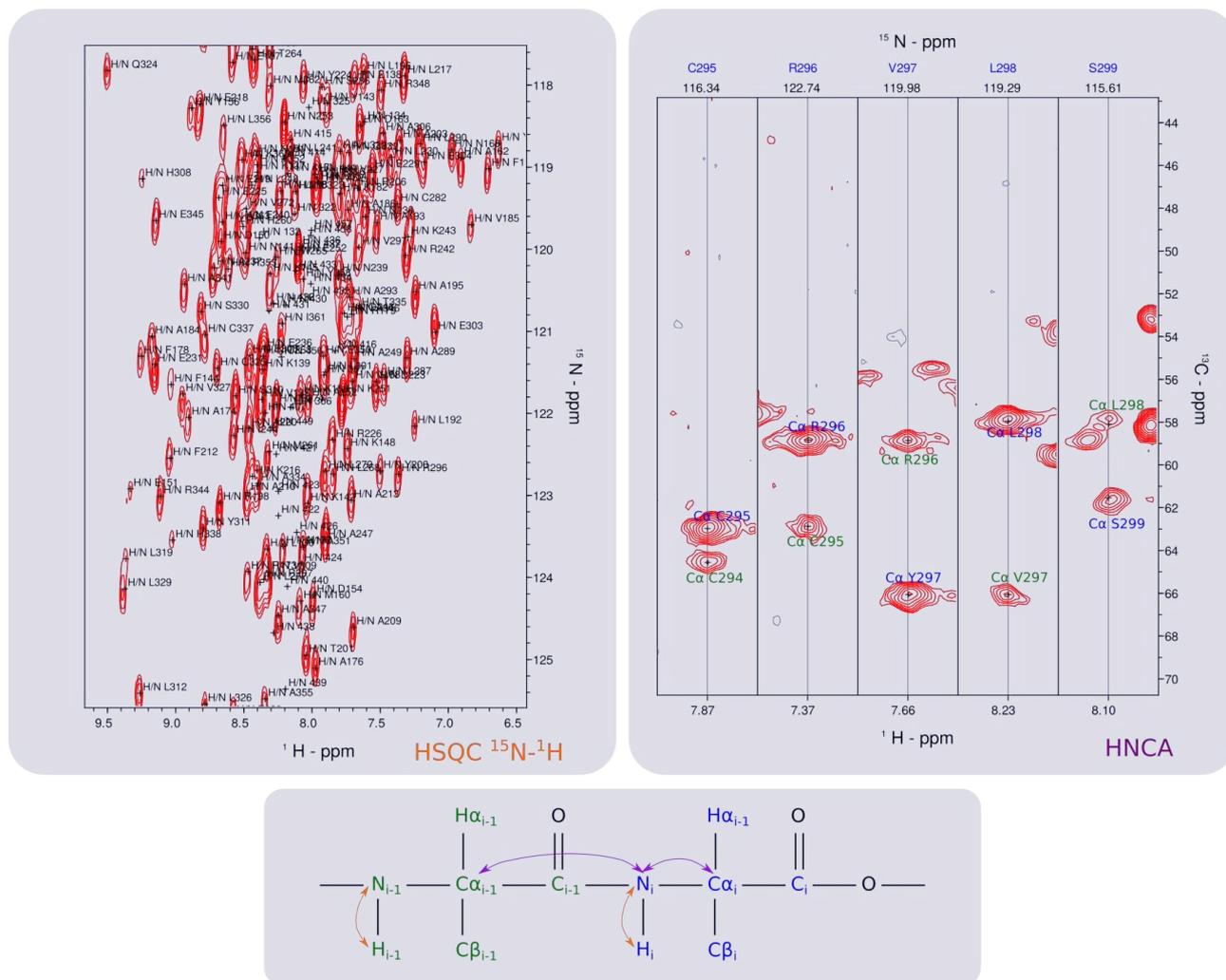


Figure 81: Exemple d'expériences RMN 2D (HSQC $^{15}\text{N}-^1\text{H}$) et 3D (HNCA).

En combinant différentes expériences de RMN 2D et 3D, il est possible de faire correspondre chaque pic d'un spectre (ici une expérience HSQC $^{15}\text{N}-^1\text{H}$ et une expérience HNCA) avec le noyau d'un atome d'un résidu i (en bleu) ou $i-1$ (en vert). Par exemple, une expérience HSQC $^{15}\text{N}-^1\text{H}$ permet notamment d'attribuer le groupement NH de la chaîne peptidique principale d'un acide aminé, alors qu'une expérience HNCA sert à attribuer le carbone α d'un résidu i et $i-1$.

Une fois l'attribution de l'ensemble de ces noyaux réalisée, les déplacements chimiques correspondant aux groupements NH du squelette peptidique du domaine 133-255 de RPAP3 libre ont été comparés aux déplacements chimiques de ces mêmes résidus lorsque ce domaine interagit avec le domaine 262-366 de TRBP (**Figure 82**). La différence de déplacement chimique ($\Delta\delta$) est obtenue à partir de la formule suivante :

$$\Delta\delta = \sqrt{(\delta H_{\text{libre}} - \delta H_{\text{complexe}})^2 + (\delta N_{\text{libre}} - \delta N_{\text{complexe}})^2 * 0,1}$$

où δ correspond au déplacement chimique des protons (H) ou des noyaux d'azotes 15 (N), pour le domaine RPAP3 133-255 libre ou en complexe avec le domaine 262-366 de TRBP. La différence de déplacement chimique des noyaux d'azote 15 est pondérée par un facteur 0,1 afin de compenser la différence d'échelle spectrale entre les dimensions des protons (entre 6 et 9 ppm) et des noyaux d'azote 15 (entre 100 et 130 ppm). Ce facteur correspond à la fréquence de résonance sur le spectromètre RMN des noyaux d'azote 15 (60 MHz) divisée par celle des protons (600 MHz). Nous obtenons ainsi un graphique qui permet de visualiser cette différence de déplacement chimique (en ppm) pour chaque acide aminé du domaine 133-255 de RPAP3 (**Figure 83**).

De manière intéressante, il semble que seules les extrémités C-terminales des motifs TPR1 et 2 du domaine 133-255 de RPAP3 soient impactées par la formation du complexe. En effet, l'interaction du domaine 262-366 de TRBP avec le domaine 133-255 de RPAP3 a pour conséquence de modifier de manière très importante le déplacement chimique des résidus 136, 160 à 164 (motif TPR1) et 189 (motif TPR2) de RPAP3 (**Figure 83**). Il y a donc de très fortes probabilités pour que ces résidus soient impliqués directement ou localisés dans une zone proche de l'interface d'interaction entre RPAP3 et TRBP. Toutefois, la modification du déplacement chimique d'un acide aminé suite à la formation d'un complexe témoigne d'un changement dans l'environnement spatial du résidu mais n'est pas synonyme de son implication directe dans la formation de ce complexe.

Nous avons donc décidé de substituer par des alanines l'ensemble des acides aminés dont le déplacement chimique est significativement modifié suite à la formation du complexe afin de tester par co-expression leur implication et leur importance dans la formation du complexe TRBP 262-366/RPAP3 133-255 par co-expression. Pour cela, nous avons calculé l'écart-type pour les valeurs obtenues pour $\Delta\delta$, puis sélectionnés les acides aminés dont le déplacement chimique était modifié d'au moins 2 fois cette valeur (**Figure 84**).

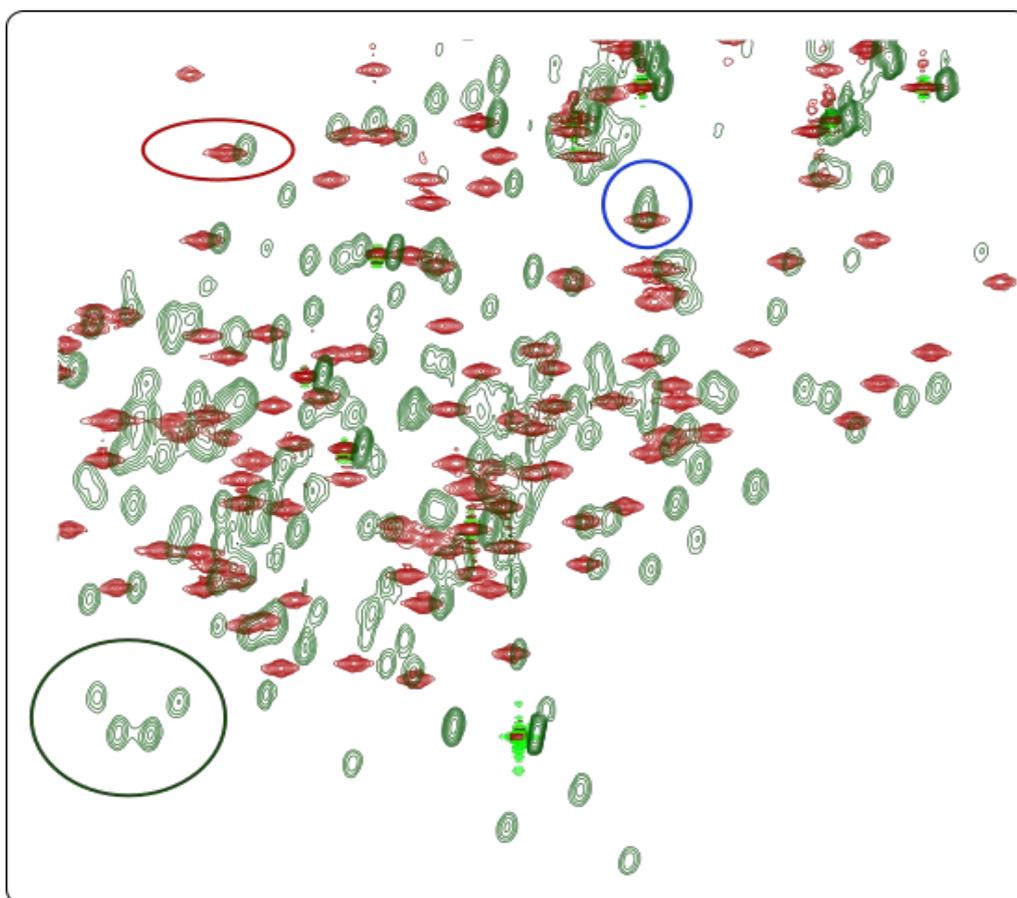


Figure 82: Impact de l'interaction de TRBP (262-366) sur RPAP3 (133-255) sur le déplacement chimique des résidus du domaine de RPAP3.

Les spectres HSQC ^{15}N - ^1H du domaine 133-255 de RPAP3 (en vert) et du complexe RPAP3 133-255 / TRBP 262-366 (en rouge) ont été superposés.

Lorsque un résidu est proche de l'interface d'interaction, son environnement spatial et donc son déplacement chimique est modifié, de manière plus ou moins importante (cercle rouge). Au contraire, si le résidu n'est pas impliqué dans l'interaction, la valeur de son déplacement chimique restera inchangée (cercle bleu).

Entouré en vert : exemple de pics retrouvés uniquement sur le spectre vert, qui correspondent donc à des acides aminés du domaine 262-366 de TRBP

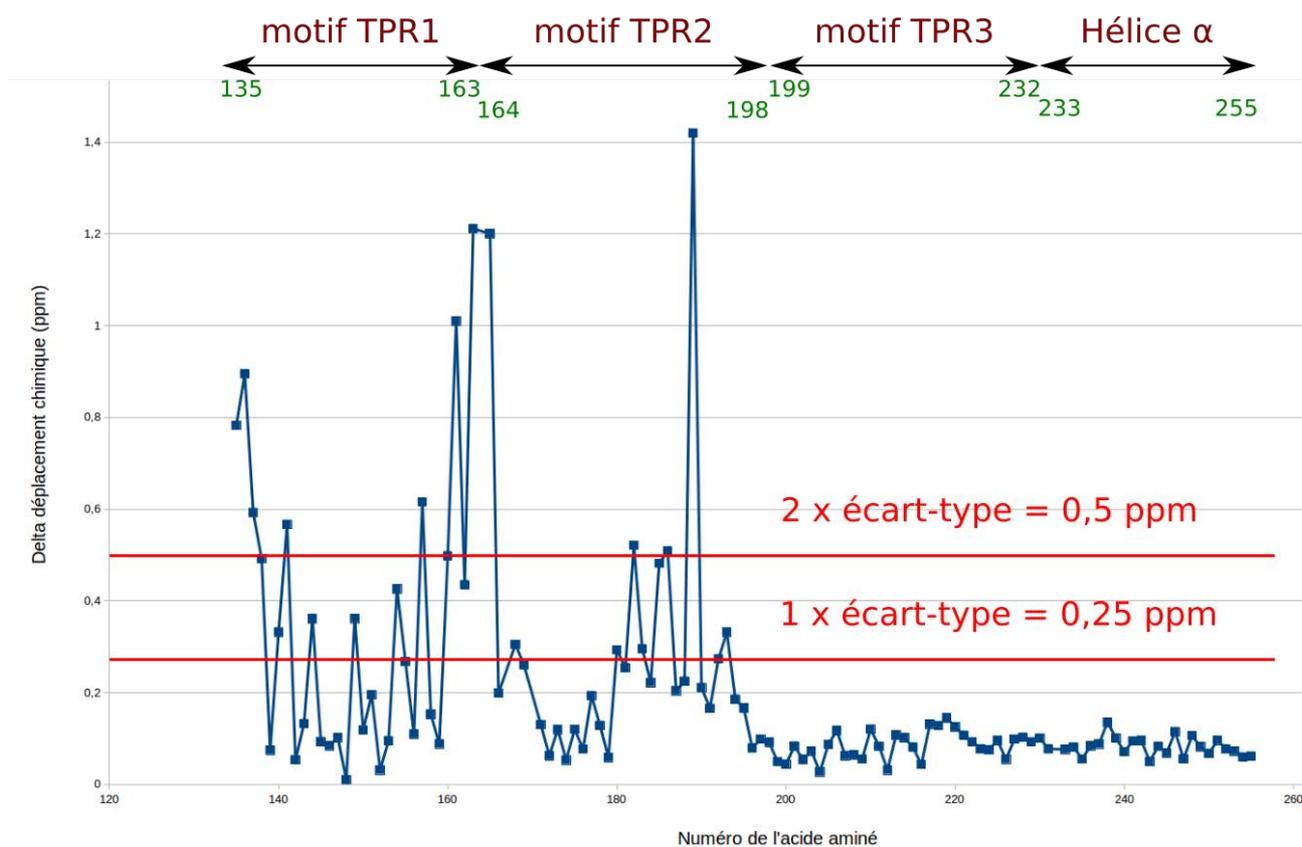


Figure 83: Résultats de l'expérience de titrage de RPAP3 133-255 par le domaine 262-366 de TRBP.

Les déplacements chimiques des protons et azotes 15 de chaque acide aminé du domaine 133-255 de RPAP3 libre et en complexe avec le domaine 262-366 de TRBP ont été comparés. Les résultats montrent que l'environnement spatial de certains résidus des motifs TPR1 et 2 du domaine TPR n°1 de RPAP3 serait modifié lors de la formation du complexe. Il est donc fortement probable que ces résidus soient proches, voire directement situés au niveau de l'interface d'interaction.

L'écart-type a été calculé sur l'ensemble des valeurs de $\Delta\delta$.

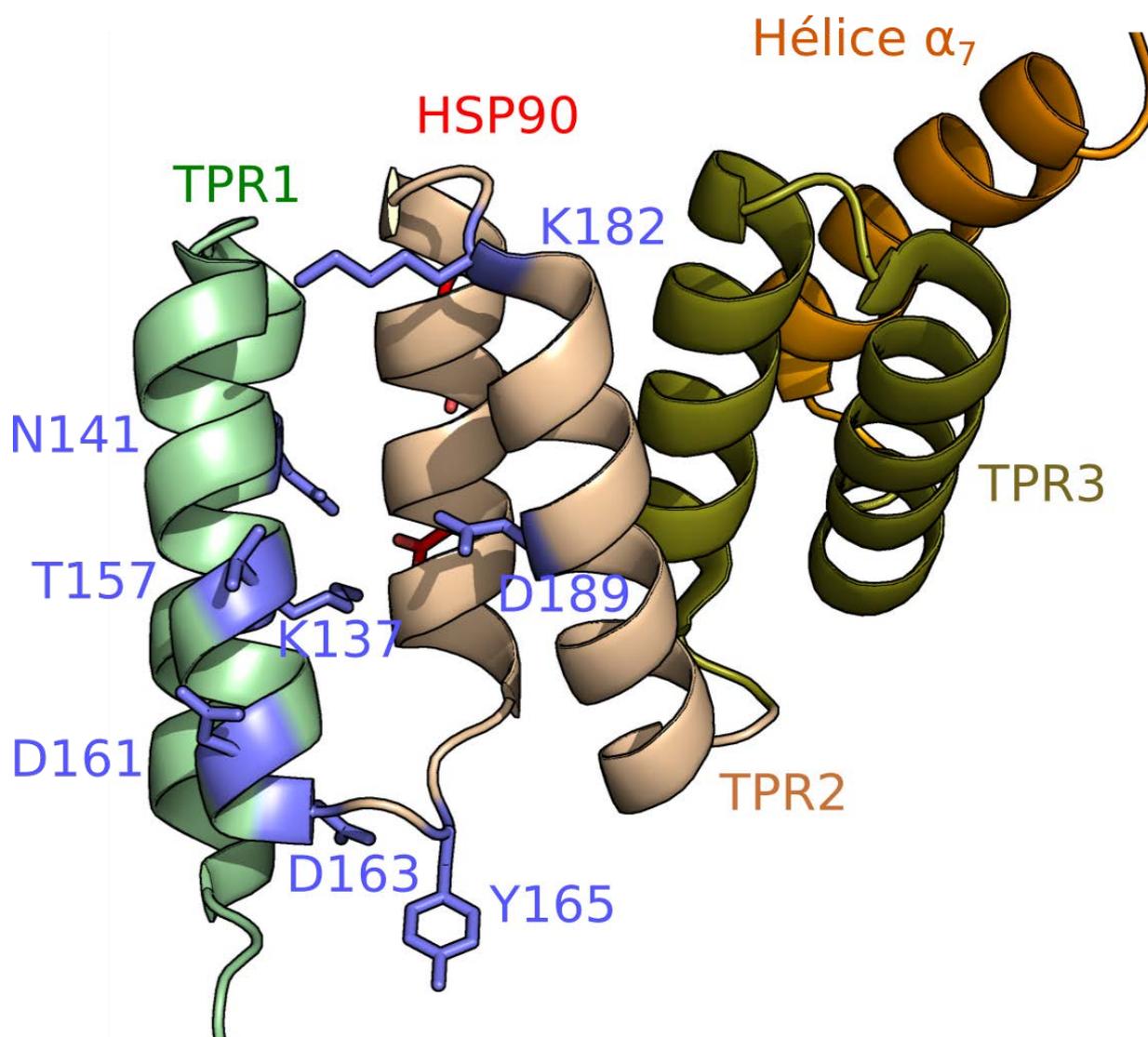


Figure 84: Résidus sélectionnés suite aux résultats de titrage par RMN (code PDB : 4CGV).

Les résultats de titrage RMN montrent que seuls les motifs TPR1 et 2 du domaine TPR n°1 de RPAP3 seraient impliqués dans l'interaction du domaine 133-255 de RPAP3 avec le domaine 262-366 de TRBP.

Nous avons alors sélectionné les résidus dont les déplacements chimiques étaient fortement modifiés suite à cette interaction (en bleu sur la figure).

Chaque acide aminé a été muté en alanine.

De manière intéressante, tous les acides aminés sélectionnés sont situés sur une même face du premier domaine TPR de RPAP3, à l'exception de la leucine 137 et de l'asparagine 141 (**Figure 84**).

Lors de ces expériences, les structures de deux domaines TPR de RPAP3, en interaction avec le peptide SRMEEVD de la chaperonne HSP90 ont été publiées (Pal *et al.*, 2014). Elles montrent qu'HSP90 interagit avec le premier domaine TPR de RPAP3 par l'intermédiaire de 5 résidus, dont la leucine 137 et l'asparagine 141, qui forment une sorte de double « pince », appelée « dicarboxylate clamp », qui va interagir avec deux acides carboxyliques d'HSP90 (COOH terminal et un acide aspartique) (voir Introduction). Afin de vérifier que le site d'interaction du premier domaine TPR de RPAP3 avec HSP90 n'est pas impliqué dans l'interaction de ce domaine avec TRBP, nous avons également muté l'asparagine 172, qui est le seul acide aminé de la « dicarboxylate clamp » à interagir directement avec le groupement carboxylate de l'extrémité C-terminale d'HSP90 et dont la mutation est suffisante pour empêcher l'interaction de RPAP3 avec HSP90 (**Figure 85**).

Enfin, nous avons également muté plusieurs autres acides aminés présents sur l'autre face du premier domaine TPR (**Figure 86**). Des double et triple mutations ont également été réalisées afin d'augmenter nos chances d'isoler des variants qui pourraient ensuite être utilisés pour étudier l'effet de l'absence de formation du complexe TRBP/RPAP3 au sein de cellules humaines. En tout nous avons sélectionné une vingtaine de résidus à muter pour le domaine 133-255 de RPAP3 (**Figure 87**)

2. Utilisation des données de cristallographie pour identifier des résidus de TRBP candidats pour leur implication dans l'interaction avec RPAP3

En l'absence de données structurales publiées sur le domaine DRBM3 de TRBP, un travail similaire a été réalisé pour essayer d'identifier des résidus de TRBP impliqués dans la formation du complexe protéique. Malheureusement, le domaine 262-366 n'est pas stable lorsqu'il n'est pas associé à un partenaire et s'agrège rapidement. À partir des données de RMN dont nous disposons (attribution de la chaîne principale du complexe), le Dr Quinternet a pu réaliser une première modélisation de la structure du squelette peptidique du domaine 262-366 de TRBP. Les données dont nous disposons n'ont toutefois pas permis de modéliser l'extrémité N-terminale du domaine (262-292). Il a toutefois été possible d'obtenir une première image de la structure globale du reste du domaine (**Figure 88**).

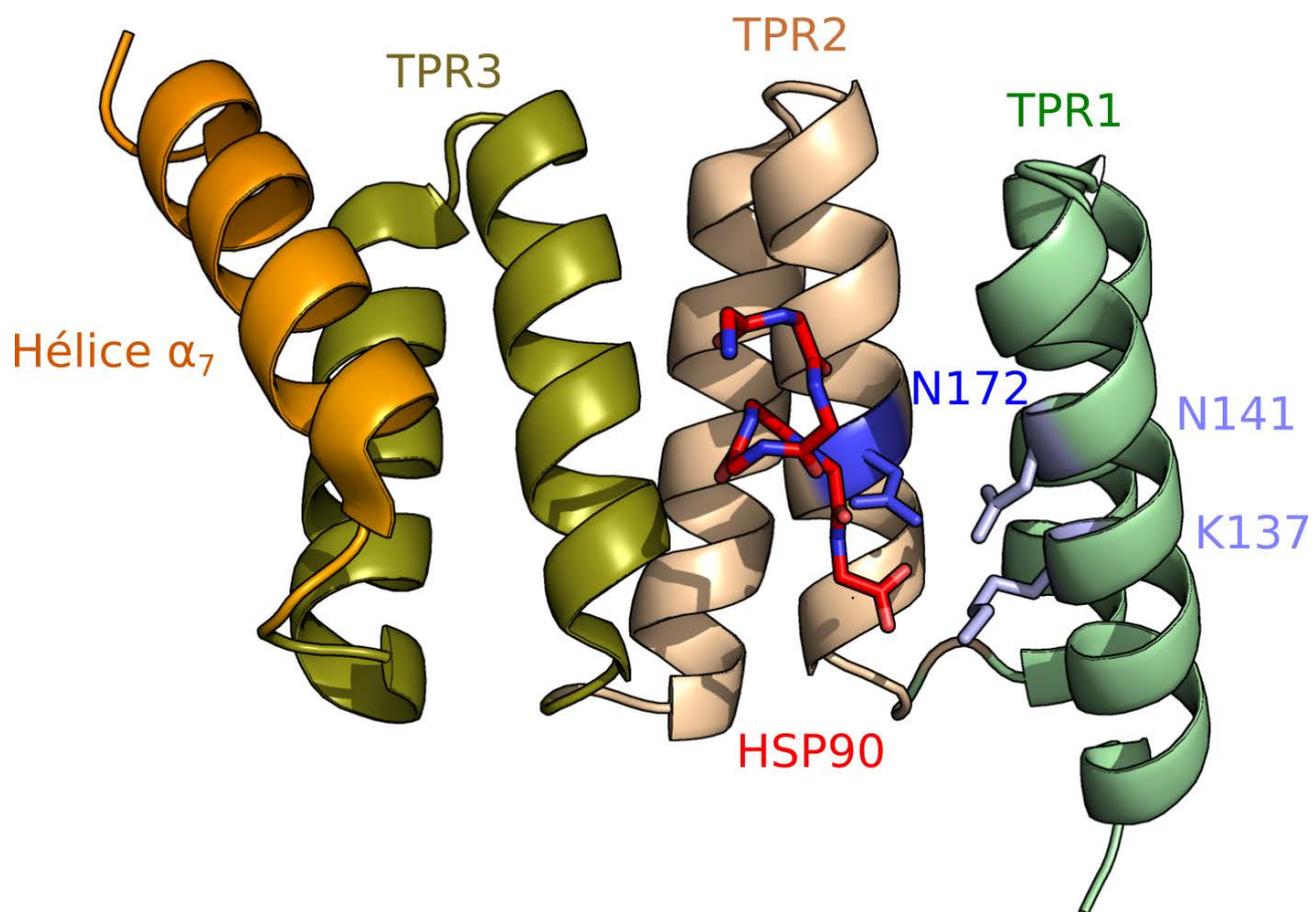


Figure 85: Résidus du domaine 133-255 de RPAP3 impliqués dans l'interaction du domaine TPR avec le peptide HSP90 (code PDB : 4CGV).

Les résidus sélectionnés suite au titrage RMN sont principalement regroupés sur la face opposée au site d'interaction du domaine TPR avec le peptide SRMEEVD d'HSP90 (en rouge), à l'exception des résidus N141 et K137. Ces résidus permettent l'interaction d'HSP90 avec le domaine TPR n°1 de RPAP3, tout comme l'asparagine 172 qui interagit directement avec l'extrémité carboxylate d'HSP90. Nous avons donc également sélectionné ce résidu (en bleu) afin de vérifier que le domaine 262-366 de TRBP n'interagissait pas avec le domaine TPR n°1 de RPAP3 par l'intermédiaire de la région interagissant avec HSP90.

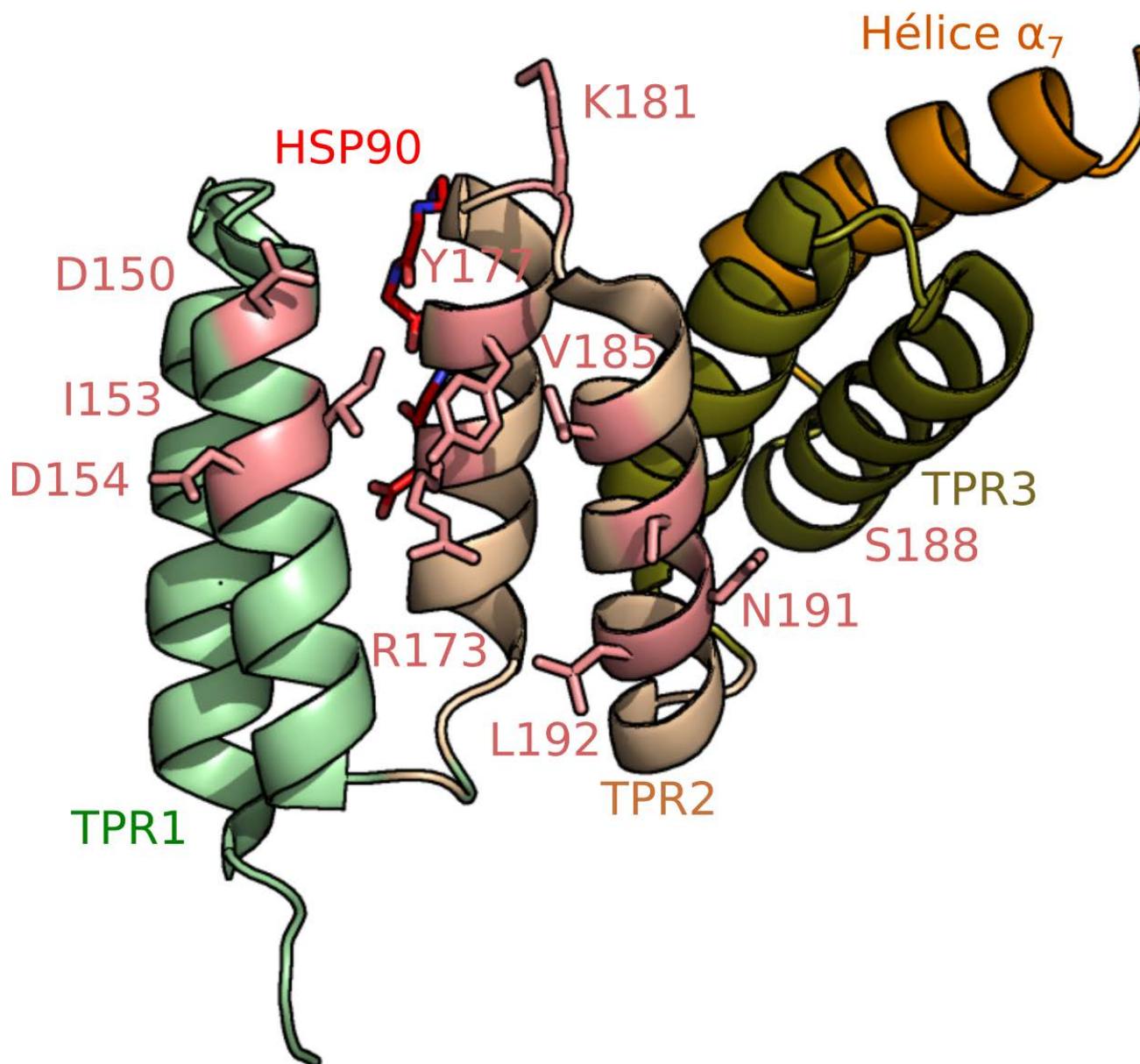


Figure 86: Sélection d'acides aminés supplémentaires suite aux résultats de titrage RMN (code PDB : 4CGV).

Les résultats de titrage RMN suggèrent que l'interface d'interaction du domaine 133-255 de RPAP3 avec le domaine 262-366 de TRBP se situerait sur la face opposée au site d'interaction avec HSP90. Afin d'augmenter nos chances d'isoler un variant intéressant, nous avons également muté plusieurs résidus présents sur cette face.

Chaque acide aminé a été substitué par une alanine.

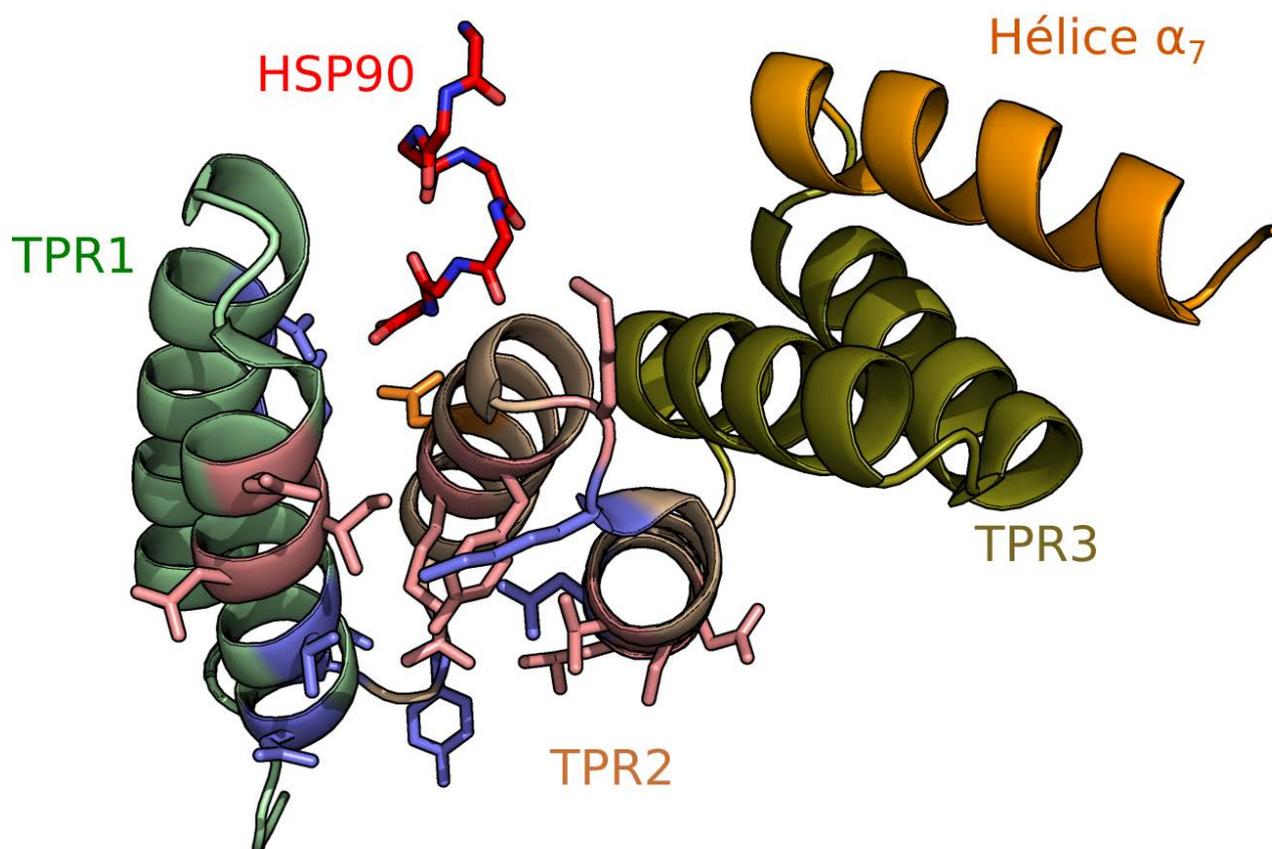


Figure 87: Bilan des acides aminés du domaine 133-255 de RPAP3 qui ont été mutés (code PDB : 4CGV).

Suite aux résultats de titrage RMN, nous avons sélectionné les acides aminés dont les déplacements chimiques étaient les plus impactés par l'interaction du domaine 133-255 de RPAP3 avec le domaine 262-366 de TRBP (en bleu), ainsi que de nombreux acides aminés présents sur la face « convexe » du domaine TPR (en rose). L'asparagine N172 (en orange), qui est impliquée dans l'interaction de RPAP3 avec HSP90, a également été sélectionnée afin de vérifier si le site d'interaction de RPAP3 avec HSP90 était impliqué ou non dans l'interaction de RPAP3 avec TRBP.

Listes des acides aminés mutés :

TPR n°1, Hélice α_1 :	K137, N141
TPR n°1, Hélice α_2 :	D150, I153, D154, T157, D161
Boucle TPR n°1 – TPR n° 2 :	D163, Y165
TPR n°2, Hélice α_1 :	N172, R173, V185, Q177
Boucle TPR n°1 Hélice α_1 - Hélice α_2 :	K181
TPR n°2, Hélice α_2 :	S188, N191, L192

Chaque acide aminé a été substitué par une alanine.

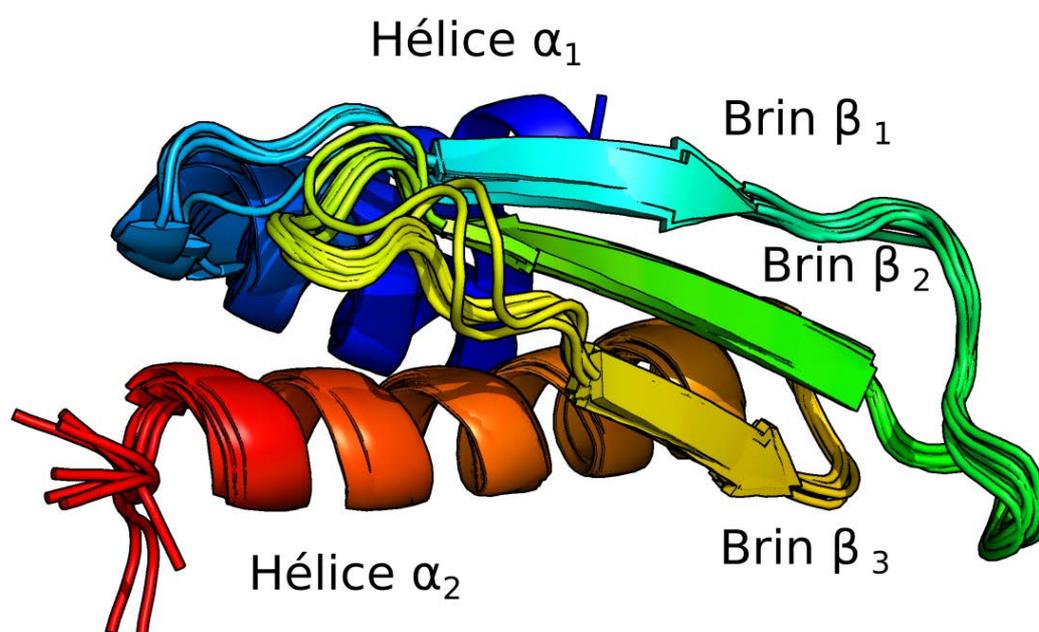


Figure 88: Résultat des 10 meilleures prédictions de structure secondaire pour le domaine 293-366 de TRBP.

Ces calculs ont été réalisés à l'aide du logiciel CS-Rosetta à partir des données expérimentales sur le complexe TRBP 262-366/RPAP3 133-255 obtenues par RMN.

Comme la plupart des domaines DRBM, le domaine DRBM3 de TRBP serait composé d'un feuillet β formé de 3 brins β anti-parallèles et flanqué de deux hélices α . Ces résultats ont été utilisés pour définir un premier lot de résidus à muter qui pourraient être impliqués dans l'interaction avec RPAP3.

Ainsi, le domaine 293-366 de TRBP contiendrait un feuillet β , composé de 3 brins β anti-parallèles, flanqué de deux hélices α . De plus, il semble que l'une des faces du feuillet β soit plaquée contre les deux hélices α , tandis que l'autre se trouve exposée au solvant.

Il est alors possible que cette seconde face puisse servir d'interface pour l'interaction avec différents partenaires protéiques, dont RPAP3. Nous avons donc, dans un premier temps, décidé de réaliser des mutations au niveau des résidus présents au niveau des trois brins β du feuillet. Les résidus dont les chaînes latérales sont, d'après nos prédictions/résultats, directement exposées au solvant, ont alors été mutés en alanine. Deux acides aminés localisés au niveau des hélices α_1 et α_2 (respectivement E300 et Q357) ont également été sélectionnés (**Figure 89**). Dans un second temps, nous avons défini d'autres acides aminés intéressants à l'aide des données structurales de TRBP tout juste publiées (Wilson *et al.*, 2015). En effet, la structure tridimensionnelle du domaine DRBM3 (289-363) de TRBP en interaction avec un domaine de Dicer (269-391), résolue par diffraction des rayons X, a été publiée par le groupe de Jennifer Doudna peu après le début de nos travaux sur l'interface TRBP/RPAP3 (Wilson *et al.*, 2015). Cette structure montre (**Figure 90**) que le domaine DRBM3 de TRBP contient effectivement un feuillet β flanqué de deux hélices α , tout comme le montrait notre modèle (**Figure 91**). Le domaine de Dicer, composé uniquement d'hélices α , venant alors interagir avec le domaine DRBM3 de TRBP au niveau du feuillet β , comme nous l'avons également envisagé. Le Dr Quinternet a alors utilisé ces nouvelles données pour identifier par bio-informatique les acides aminés de TRBP qui seraient importants pour l'interaction avec Dicer, et donc probablement pour celle avec RPAP3. Au final, une vingtaine de ces résidus ont été mutés en alanine (**Figure 92**), et tout comme pour RPAP3, nous avons également réalisé des doubles et triples mutations, au niveau des acides aminés situés dans les différents brins β . Enfin, nous avons également muté l'arginine 354 en acide glutamique dans l'espoir que la présence d'une charge négative au niveau de l'hélice α_2 empêche la formation du complexe protéique.

3. Confirmation par co-expression de l'impact de différents variants de RPAP3 et de TRBP sur la formation du complexe protéique

L'ensemble des mutations ont été réalisées sur les vecteurs pnCS-RPAP3 133-255 ou pnEA3ch-TRBP 262-366 par mutagenèse dirigée. L'effet de ces mutations sur la formation du complexe protéique a alors été vérifié par co-expression et co-purification. Les différents complexes protéiques ont été purifiés sur résine TALON™ (voir le paragraphe II.B.8.c.i de la partie Matériel et Méthodes) et une aliquote de la résine (F) a été fractionnée sur un gel de protéines dénaturant.

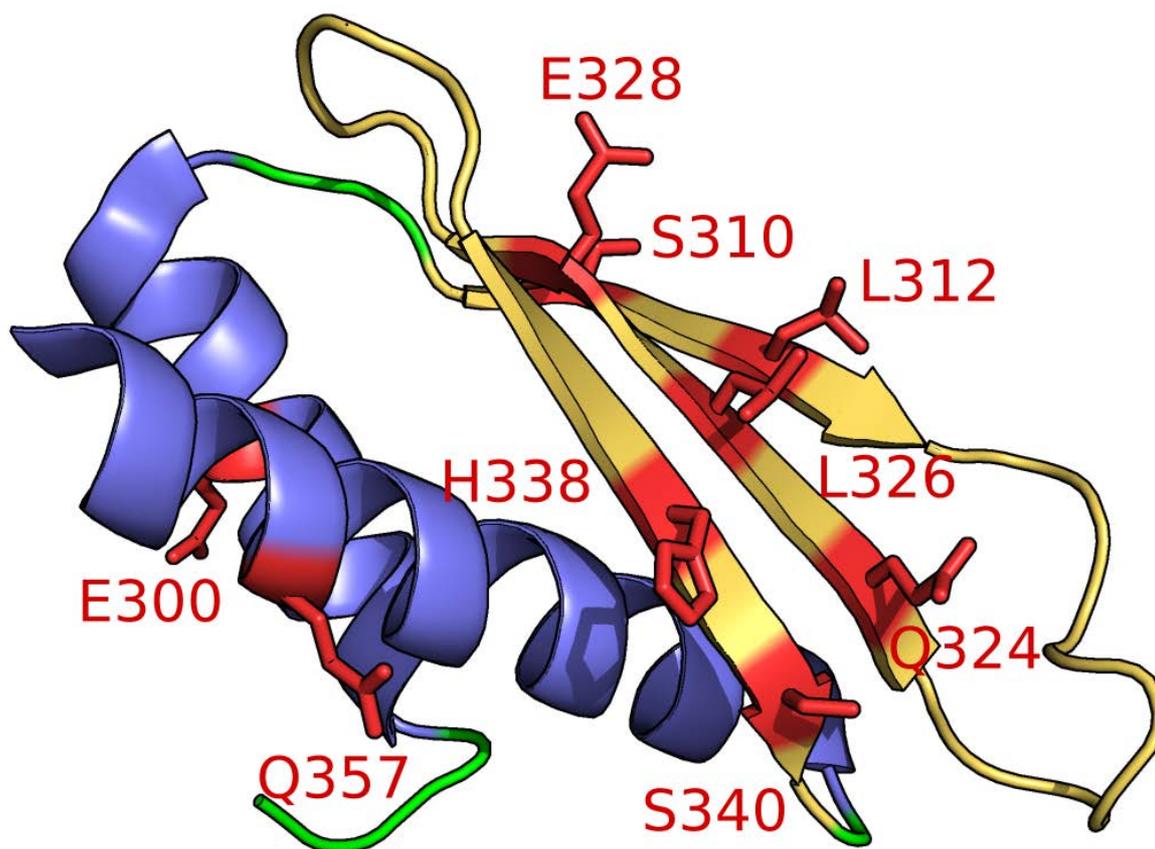
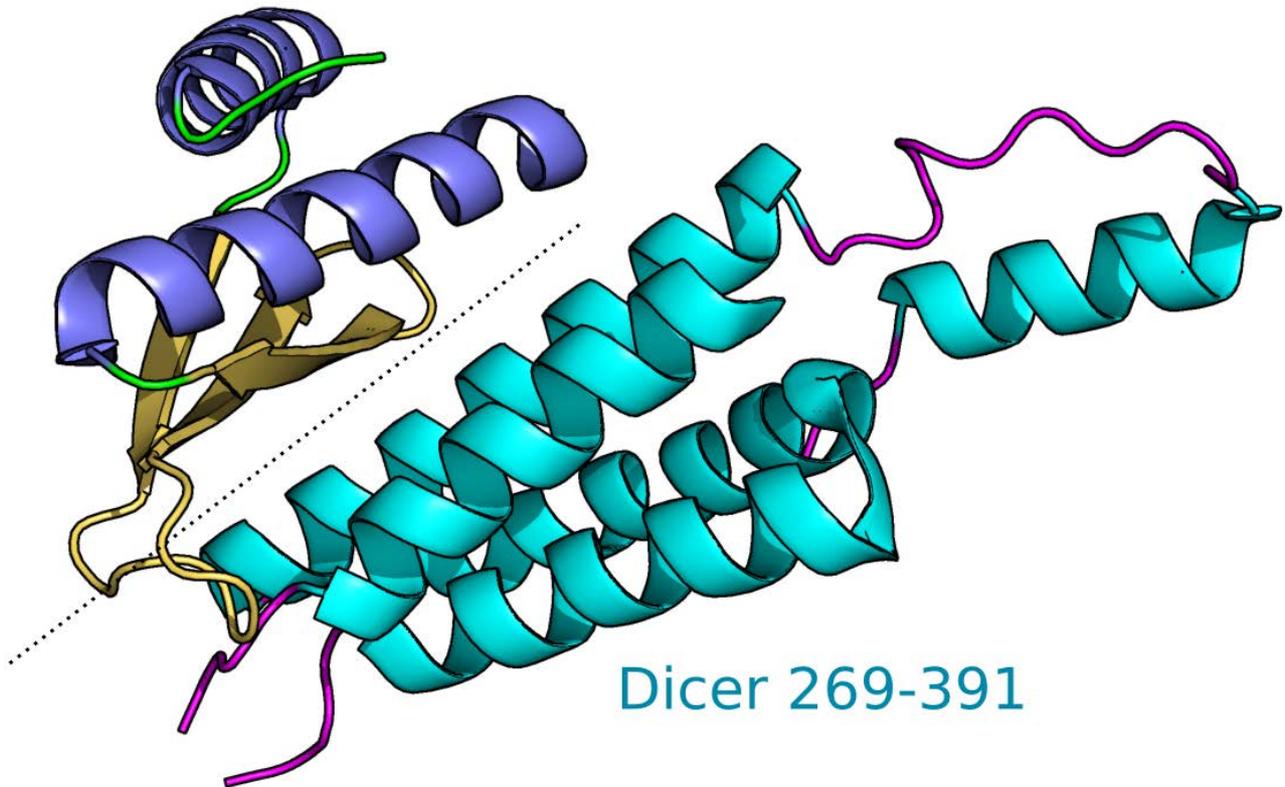


Figure 89 : Acides aminés du domaine DRBM3 de TRBP sélectionnés suite aux prédictions de structures secondaires réalisées à partir de nos données RMN.

Les résidus des brins β dont les chaînes latérales seraient, d'après les calculs effectués, exposées au solvant, ont été mutés en alanine, ainsi que les résidus E300 et Q357 présents au niveau des deux hélices.

TRBP 289-363



Dicer 269-391

Figure 90: Structure de l'interface d'interaction entre le domaine 289-363 de TRBP (DRBM3) et 269-391 de Dicer, résolue par diffraction des rayons X (code PDB : 4WYQ).

Le domaine DRBM3 de TRBP (289-363) est composé de 3 brins β anti-parallèles qui forment un feuillet β , ce dernier étant flanqué par 2 hélices α . Le domaine de Dicer, composé uniquement d'hélices α , vient interagir avec le domaine DRBM3 de TRBP principalement au niveau du feuillet β (Wilson *et al.*, 2015).

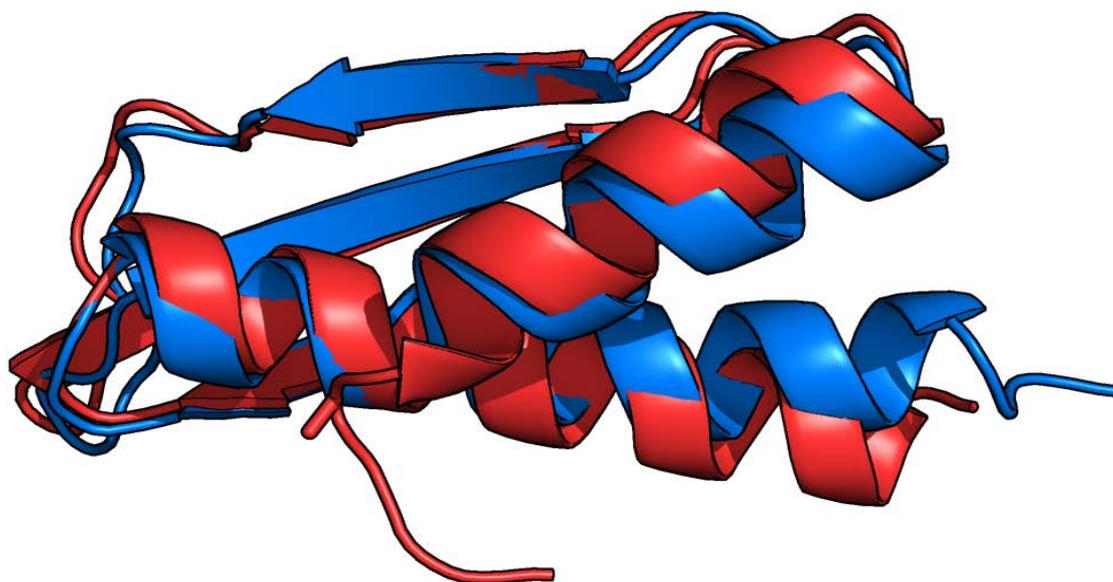


Figure 91: Alignement entre la structure du domaine 293-366 de TRBP obtenu par CS-Rosetta et la structure du domaine 289-363 résolue par diffraction des rayons X.

L'alignement, réalisé à l'aide du logiciel *Pymol*, montre un RMSD au niveau des carbones alpha de 1,371 Å.

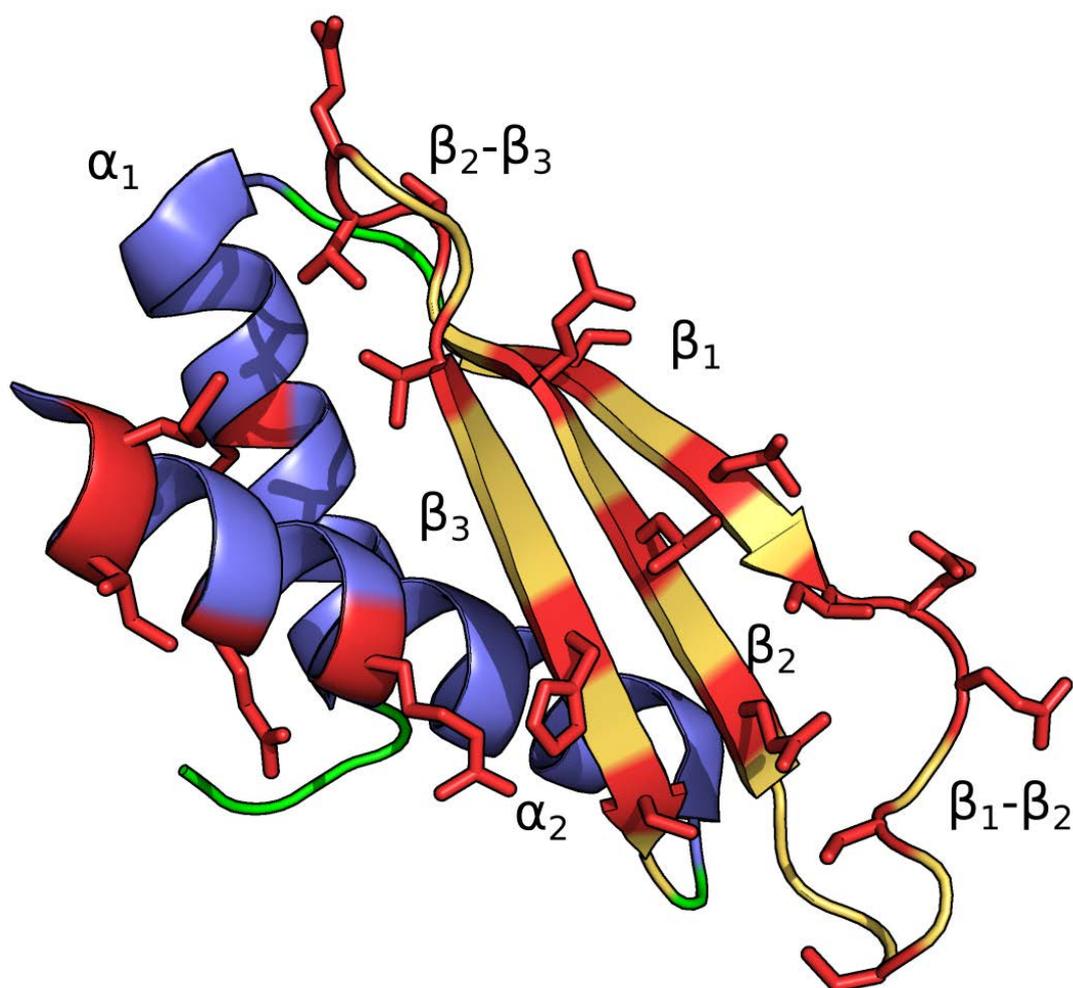


Figure 92: Bilan des acides aminés du domaine 262-366 de TRBP qui ont été mutés (code PDB : 4WYQ).

À partir des données structurales publiées pour le complexe TRBP/Dicer, nous avons sélectionné une vingtaine d'acides aminés potentiellement impliqués dans l'interaction avec le domaine 133-255 de RPAP3.

Listes des acides aminés mutés :

Hélice α_1 :	E300
Brin β_1 :	S310, L312
Boucle β_1 - β_2 :	I314, E315, E316, S318, S320
Brin β_2 :	Q324, L326, E328
Boucle β_2 - β_3 :	S330, T331, Q332
Brin β_3 :	T335, H338, S340
Hélice α_2 :	R354, Q357, I361, M362

Chaque acide aminé a été muté en alanine, à l'exception de l'arginine 354 qui a également été mutée en acide glutamique afin d'augmenter les chances d'empêcher l'interaction de TRBP avec RPAP3 par l'ajout d'une charge négative.

Les mutations pouvant entraîner d'importantes modifications au niveau du repliement intrinsèque de la protéine, qui en soi pourrait entraîner la perte de l'interaction, nous avons également déposé sur les gels les fractions correspondant aux protéines totales (So), solubles (SnSo) et insolubles (Co), afin de vérifier que tous les fragments soient toujours exprimés et solubles.

Les résultats des co-expression sont présentés **Figures 93, 94 et 95**. Pour la très grande majorité des mutations réalisées, aucun effet n'a été observé quant à la formation du complexe protéique. Cela se traduit au niveau des gels par la présence de deux bandes de forte intensité aux alentours de 15 kDa (pistes F). Ces bandes correspondent aux domaines 133-255 de RPAP3 (bande la plus haute) et 262-366 de TRBP (bande la plus basse), dont les tailles respectivement attendues sont de 13,5 kDa et 11,5 kDa. Dans certains cas, l'absence de l'une des deux protéines dans la piste F, se traduit également par son absence dans les pistes So et/ou SnSo, mais à sa présence au niveau de la piste Co, ce qui traduit le fait que la (ou les) mutation réalisée ait rendu la protéine insoluble. Cela semble notamment être le cas pour les mutations D189A, N191A, L192A (**Figure 94**) et T157A (**Figure 95**) du domaine 133-255 de RPAP3. Ces mutations ne permettent donc pas de conclure quant à l'implication de ces résidus dans l'interaction. L'ensemble de nos résultats nous a néanmoins permis d'identifier 2 résidus de RPAP3 et 5 de TRBP (**Figure 95**) dont la mutation empêche la formation du complexe, sans conséquences sur la structure des protéines mutées. Ainsi, d'après nos résultats, les résidus D161 (motif TPR1) et V185 (motif TPR2) de RPAP3 (**Figure 96**) et Q324, E328, T331 (feuillet β), Q357 et R354 (hélice α_2) de TRBP (**Figure 97**) sont impliqués dans l'interaction du domaine 262-366 de TRBP avec le domaine 133-255 de RPAP3. Ainsi, suite à la mutation de ces résidus, seule la bande qui correspond au domaine 262-366 de TRBP fusionné à l'étiquette 6-histidines est retrouvé au niveau des pistes F, alors que les deux protéines sont toujours exprimées et solubles. En effet pour ces 7 résidus, les deux bandes qui correspondent aux deux domaines sont toujours visibles au niveau des pistes So et SnSo.

De manière intéressante, pour le résidu R354, la formation du complexe protéique n'est totalement abolie que pour la mutation R354E (**Figure 95**). En effet, dans le cas de la mutation R354A, un complexe est toujours visible sur gel (**Figure 95**). Le fait que seule la mutation de l'arginine 354 en acide glutamique ait un effet suggère fortement qu'elle soit impliquée dans la formation d'une ou plusieurs interactions électrostatiques avec un ou plusieurs résidus de RPAP3, via sa chaîne latérale.

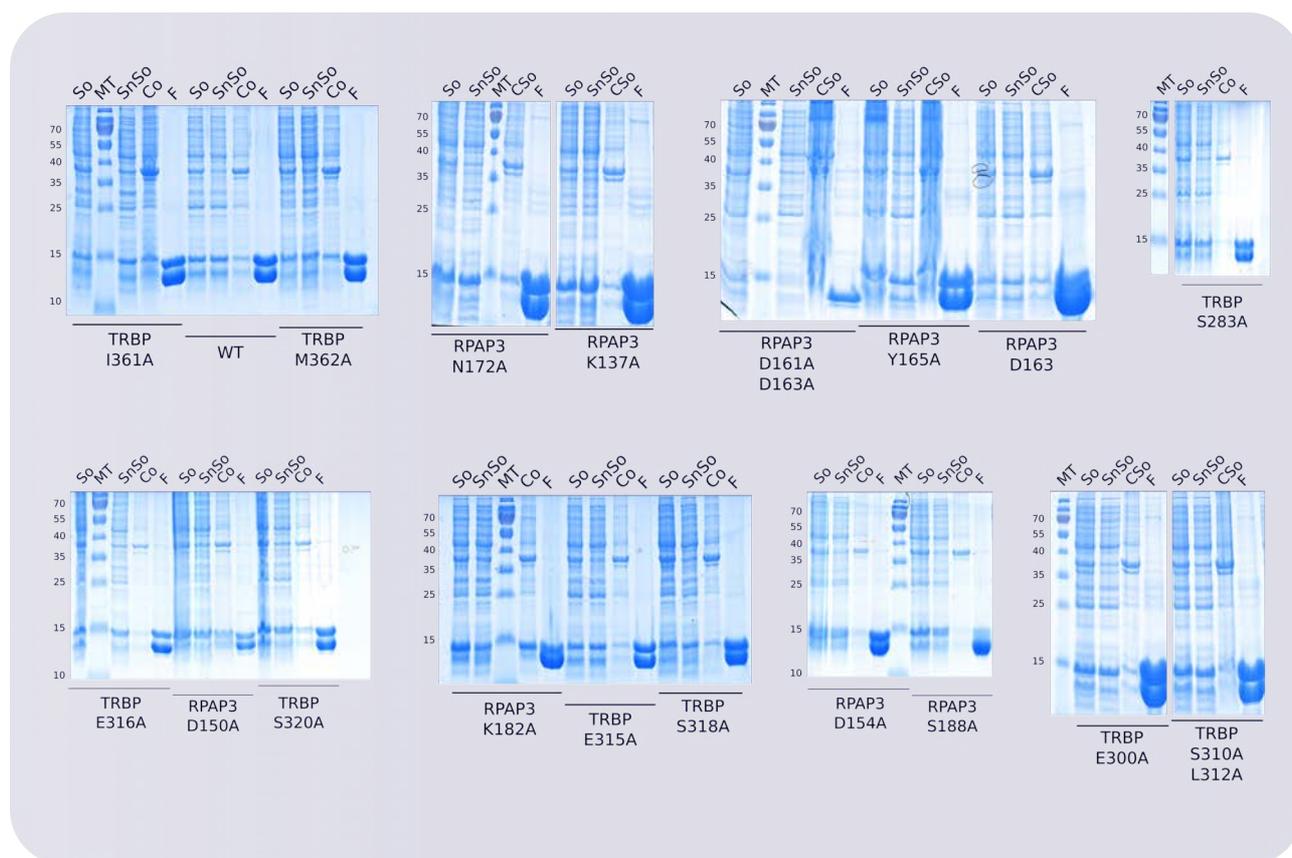


Figure 93 : Résultats (1/3) des expériences de co-expression réalisées sur les variants de RPAP3 ou de TRBP.

Comme décrit précédemment, lors de la co-purification de ces complexes, des aliquotes sont prélevées à diverse étapes afin de vérifier que les protéines soient bien exprimées (piste So), solubles (piste SnSo) ou insolubles (piste Co). Une aliquote de la résine TALON™ après purification est également prélevée (piste F). L'ensemble de ces échantillons est ensuite fractionné sur un gel de protéines dénaturant.

Les expériences de co-expression ont été réalisées avec les domaines 133-255 de RPAP3 et 262-366 de TRBP dont les séquences codantes étaient portées respectivement par les vecteurs pnCS et pnEA3ch.

Seul le nom du fragment muté est indiqué sous le gel concerné, ainsi « TRBP I361A » correspond aux résultats de co-expression entre le domaine 262-366 de TRBP dont l'isoleucine 361 a été mutée en alanine et le domaine 133-255 de RPAP3 sauvage.

WT : expérience témoin réalisée avec les vecteurs pnAE3ch-TRBP 262-366 sauvage et pnCS-RPA3 133-255 sauvage.

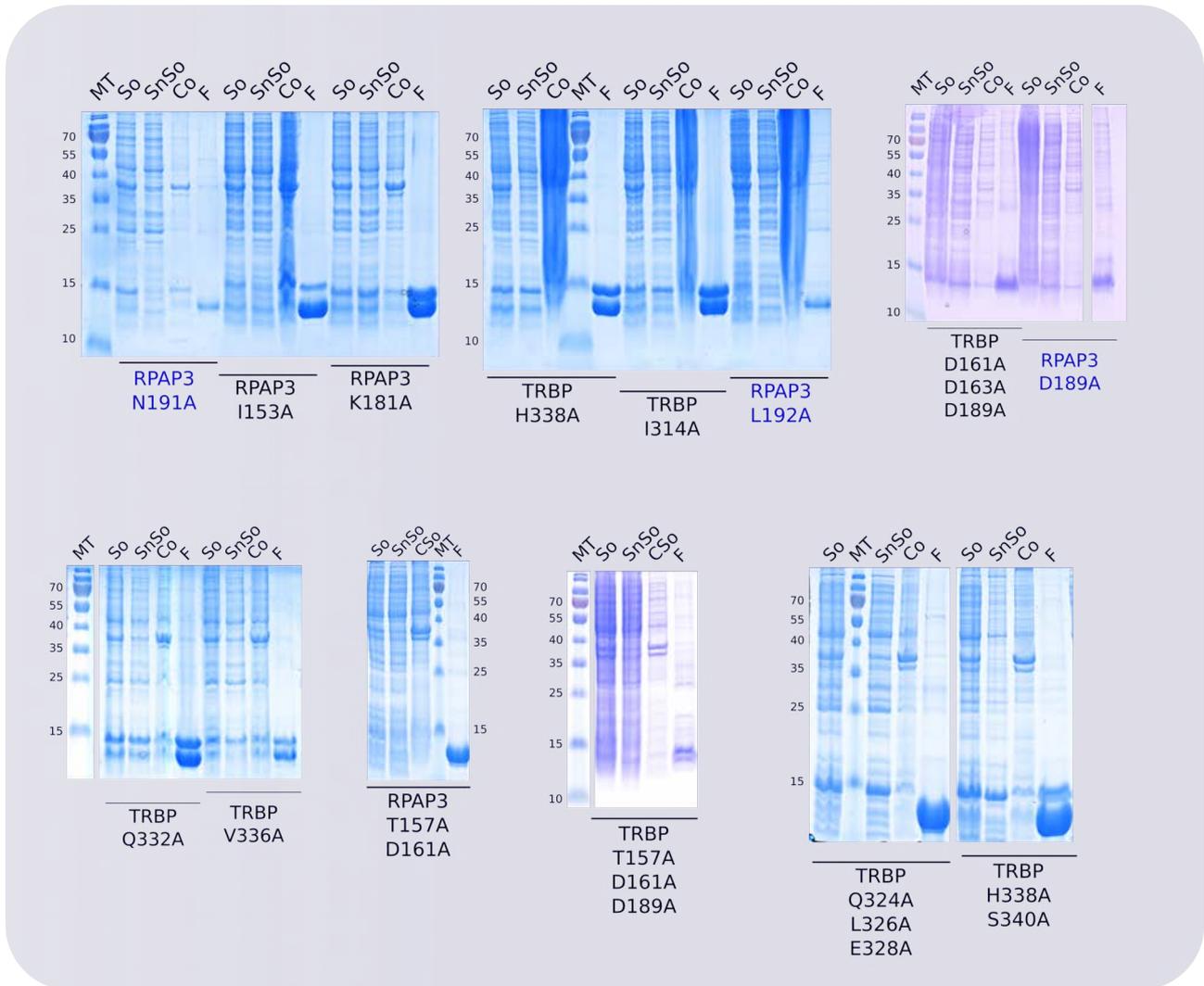


Figure 94 : Résultats (2/3) des expériences de co-expression réalisées sur les variants de RPAP3 ou de TRBP.

Comme décrit précédemment, lors de la co-purification de ces complexes, des aliquotes sont prélevées à diverse étapes afin de vérifier que les protéines soient bien exprimées (piste So), solubles (piste SnSo) ou insolubles (piste Co). Une aliquote de la résine TALON™ après purification est également prélevé (piste F). L'ensemble de ces échantillons est ensuite fractionné sur un gel de protéines dénaturant.

Les expériences de co-expression ont été réalisées avec les domaines 133-255 de RPAP3 et 262-366 de TRBP dont les séquences codantes étaient portées respectivement par les vecteurs pnCS et pnEA3ch.

En bleu, les variants simples qui semblent avoir un impact sur la production et/ou la solubilité de la protéine.

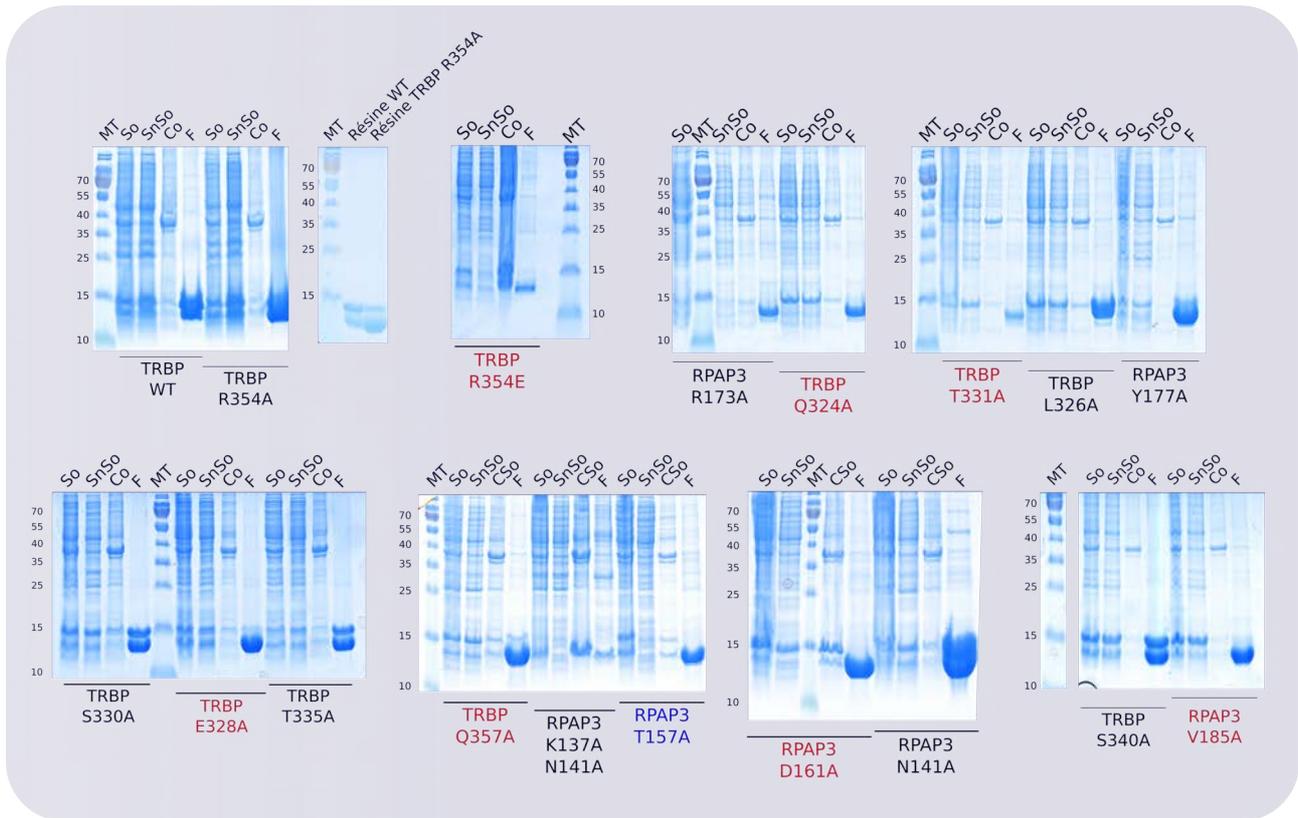


Figure 95 :Résultats (3/3) des expériences de co-expression réalisées sur les variants de RPAP3 ou de TRBP.

Comme décrit précédemment, lors de la co-purification de ces complexes, des aliquotes sont prélevés à diverse étapes afin de vérifier que les protéines soient bien exprimés (piste So), solubles (piste SnSo) ou insolubles (piste Co). Une aliquote de la résine TALON™ après purification est également prélevée (piste F). L'ensemble de ces échantillons est ensuite fractionné sur un gel de protéines dénaturant.

Les expériences de co-expression ont été réalisées avec les domaines 133-255 de RPAP3 et 262-366 de TRBP dont les séquences codantes étaient portées respectivement par les vecteurs pnCS et pnEA3ch. En bleu, les variants qui semblent avoir un impact sur la production et/ou la solubilité de la protéine.

En rouge, les variants qui empêchent la formation du complexe protéique sans avoir, *a priori*, d'effets importants sur la structure de la protéine.

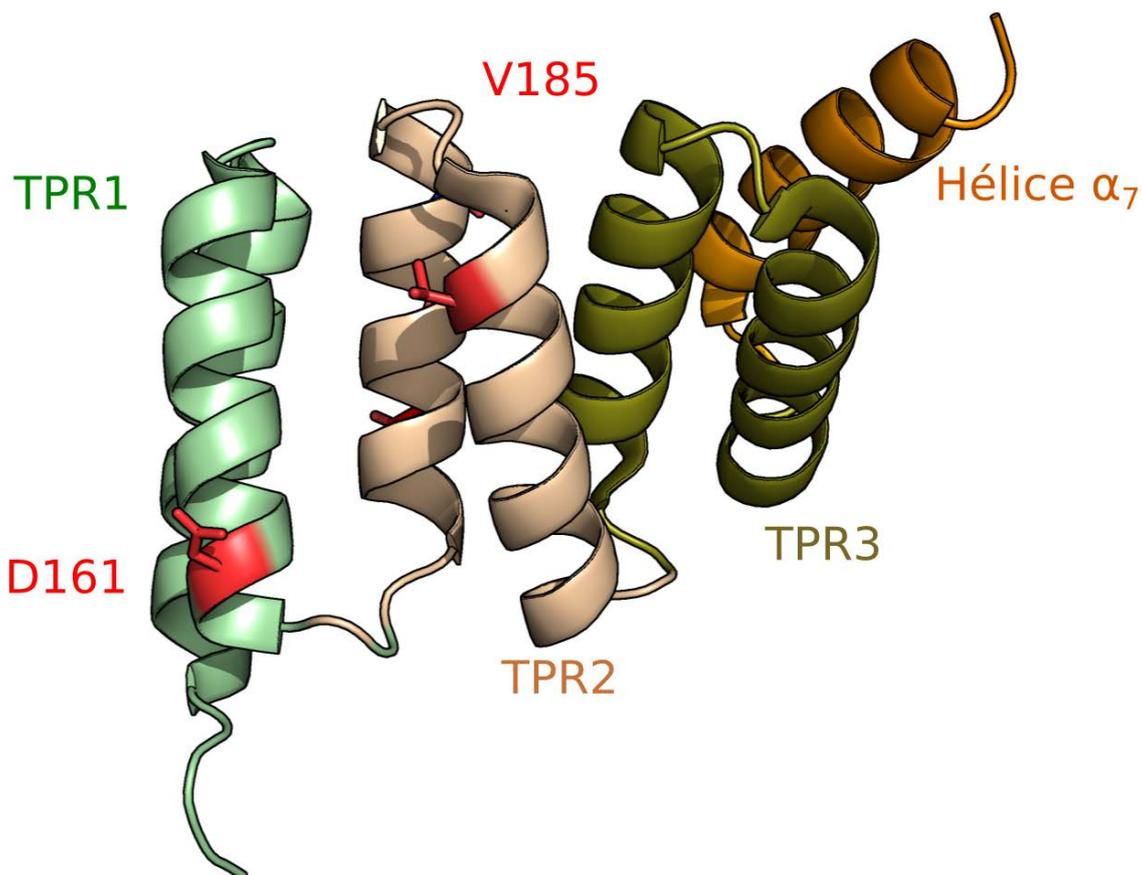


Figure 96: Bilan des acides aminés du domaine 133-255 de RPAP3 qui sont proposés, à l'issue des tests de co-expression, être impliqués dans l'interaction avec TRBP.

Nos expériences de co-expression nous ont permis d'identifier l'acide aspartique 161 (motif TPR n°1) et la valine 185 (motif TPR n°2) comme nécessaires pour l'interaction du domaine 133-255 de RPAP3 avec le domaine 262-366 de TRBP

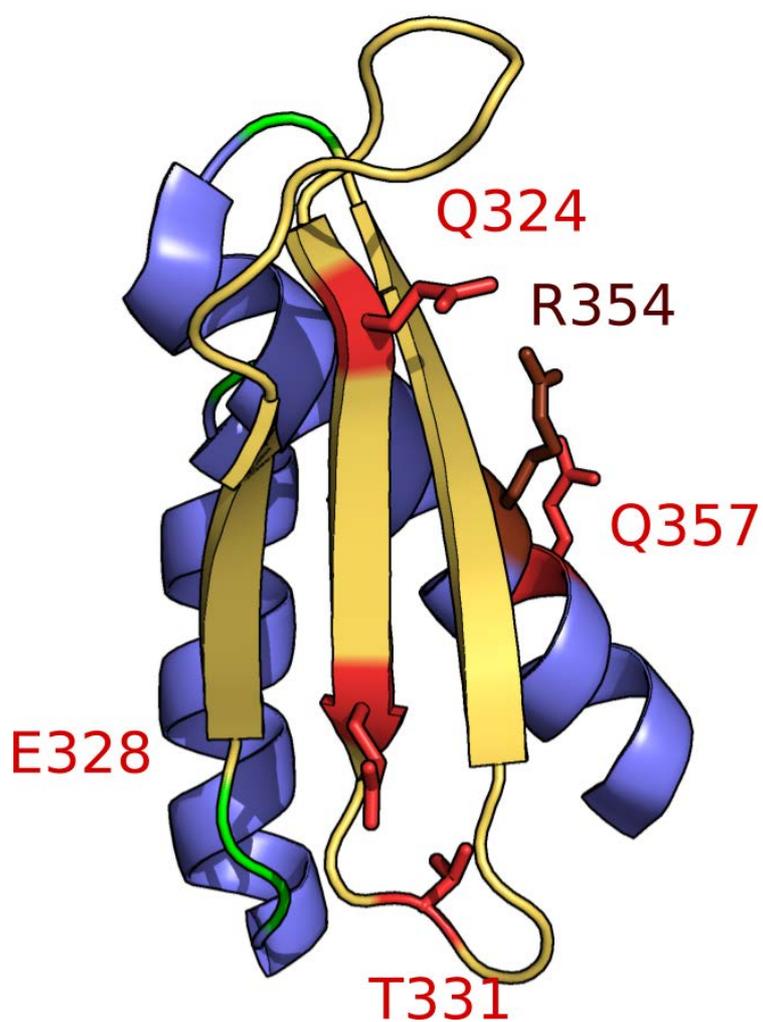


Figure 97: Bilan des acides aminés du domaine 262-366 de TRBP proposés, à l'issue des tests de co-expression, être impliqués dans l'interaction avec RPAP3.

Nos expériences de co-expression nous ont permis d'identifier l'acide glutamique 328, la thréonine 331, les glutamines 324 et 357 ainsi que l'arginine 354 comme impliqués dans l'interaction du domaine 262-366 de TRBP avec le domaine 133-255 de RPAP3.

L'absence d'effet sur la formation du complexe protéique lors de la mutation des acides aminés N172, K137 et N141 indique que le site d'interaction du domaine 133-255 de RPAP3 avec la protéine HSP90 n'est probablement pas impliqué dans l'interaction avec TRBP. De plus, nos résultats indiquent que le domaine 262-366 de TRBP interagirait avec RPAP3 par l'intermédiaire d'une part de l'hélice α_2 et d'autre part du feuillet β , soit via les mêmes éléments que pour son interaction avec Dicer. Il est donc fortement probable que les interactions de TRBP avec, d'une part RPAP3, et d'autre part Dicer, qui impliquent non seulement le même domaine de TRBP, mais également *a priori* la même surface d'interaction, ne puissent pas avoir lieu simultanément. Enfin, certains de ces résultats pourraient expliquer en partie pourquoi le domaine TPR n°2 de RPAP3 n'interagit pas avec TRBP.

En effet, ces domaines sont structurellement conservés (domaine TPR composé de 3 motifs TPR) mais nos résultats de double-hybride et de co-expression montrent clairement que le second domaine TPR de RPAP3 n'est pas impliqué dans l'interaction avec TRBP. Un alignement de séquence réalisé entre ces deux domaines TPR (**Figure 98**) montre d'importantes différences au niveau de l'hélice 4. Surtout, il semble que les résidus du TPR n°1 que nous avons précédemment identifiés comme impliqués dans l'interaction avec RPAP3 ne sont pas présents dans le second domaine TPR. Ainsi, la valine 185 du premier domaine TPR est remplacée par un acide glutamique dans le second domaine TPR, et l'acide aspartique 161, dont la mutation en alanine est suffisante pour abolir l'interaction de RPAP3 avec TRBP, correspond à une alanine au niveau du second domaine TPR. L'absence de conservation de ces deux résidus entre les deux domaines TPR pourrait donc expliquer, du moins en partie, la différence de comportement de ces domaines vis-à-vis de TRBP.

B. Résolution de la structure tridimensionnelle du complexe TRBP 262-366/RPAP3 133-255

Parallèlement à l'identification des résidus impliqués dans la formation du complexe protéique, nous avons poursuivi l'étude structurale du complexe afin d'en résoudre la structure tridimensionnelle par RMN. L'ensemble des noyaux des atomes du squelette protéique du complexe a pu être attribué, à l'exception de la glycine 339 et de la glutamine 357 de TRBP. Une fois ces attributions effectuées, il a été nécessaire de réaliser un travail similaire pour les noyaux des atomes des chaînes latérales des différents acides aminés. Toutefois, pour ce type d'attribution, la taille du complexe protéique que nous utilisons nécessite l'utilisation d'un spectromètre de plus haut champ.

Pour cela, nous avons pu obtenir du temps pour 2 semaines d'expériences sur le spectromètre RMN de 900 MHz situé à la plateforme RMN de Gif-sur-Yvette de l'ICSN. Nous avons donc produit et purifié un échantillon deutéré à 50 % et doublement marqué (^{15}N , ^{13}C) (**Figure 99**).

Malheureusement, après moins d'une journée dans le spectromètre 900 MHz, le complexe protéique a précipité. L'ensemble des expériences nécessitant au moins 1 semaine d'acquisition, il ne nous a pas été possible d'utiliser les spectres obtenus. Ce problème de stabilité avait déjà été identifié durant nos précédentes expériences sur la plateforme RMN de la fédération FR3209, mais plusieurs éléments du protocole de purification avaient été optimisés, comme l'ajout de DTT lors de la préparation de l'échantillon, qui n'ont toutefois pas permis de résoudre le problème. Pour la poursuite de cet aspect du projet, d'autres mises au point vont devoir être réalisées. De nombreux résidus très intéressants ayant été toutefois déjà identifiés, dans le cadre de mon travail de doctorat, nous avons choisi de nous concentrer pour la suite sur l'étude fonctionnelle du complexe protéique.

V. Conclusion

Les objectifs de cette partie étaient de confirmer puis de caractériser l'interaction entre les protéines RPAP3 et TRBP. Par des approches de double-hybride, d'immunoprécipitation, de Duolink™, de co-expression et de RMN, nous avons pu confirmer l'existence d'une interaction directe entre ces deux protéines *in vitro* et *in vivo* puis identifier les domaines impliqués dans la formation de ce complexe protéique. Ainsi, l'interaction entre RPAP3 et TRBP implique le premier domaine TPR de RPAP3 et le dernier domaine DRBM de TRBP. Ces domaines sont connus pour être impliqués respectivement dans l'interaction de RPAP3 avec la protéine HSP90 (Pal *et al.*, 2014) et dans celles de TRBP avec les protéines Dicer (Haase *et al.*, 2005) ou PACT (Laraki *et al.*, 2008). L'étude structurale du complexe TRBP 262-366/RPAP3 133-255 par RMN ne nous a pas permis pour le moment d'en résoudre entièrement la structure tridimensionnelle, toutefois, nos résultats préliminaires de RMN ont permis de nous fournir d'autres informations. Ainsi, nous avons réalisé, en collaboration avec le Dr Quinternet, des expériences de titrage sur le complexe. Grâce à ces expériences, nous avons identifié un certain nombre de résidus du domaine 133-255 de RPAP3 dont les déplacements chimiques sont fortement modifiés lors de la formation du complexe protéique, ce qui généralement suggère qu'ils sont soit directement impliqués dans la formation du complexe, soit proches des résidus qui le sont.

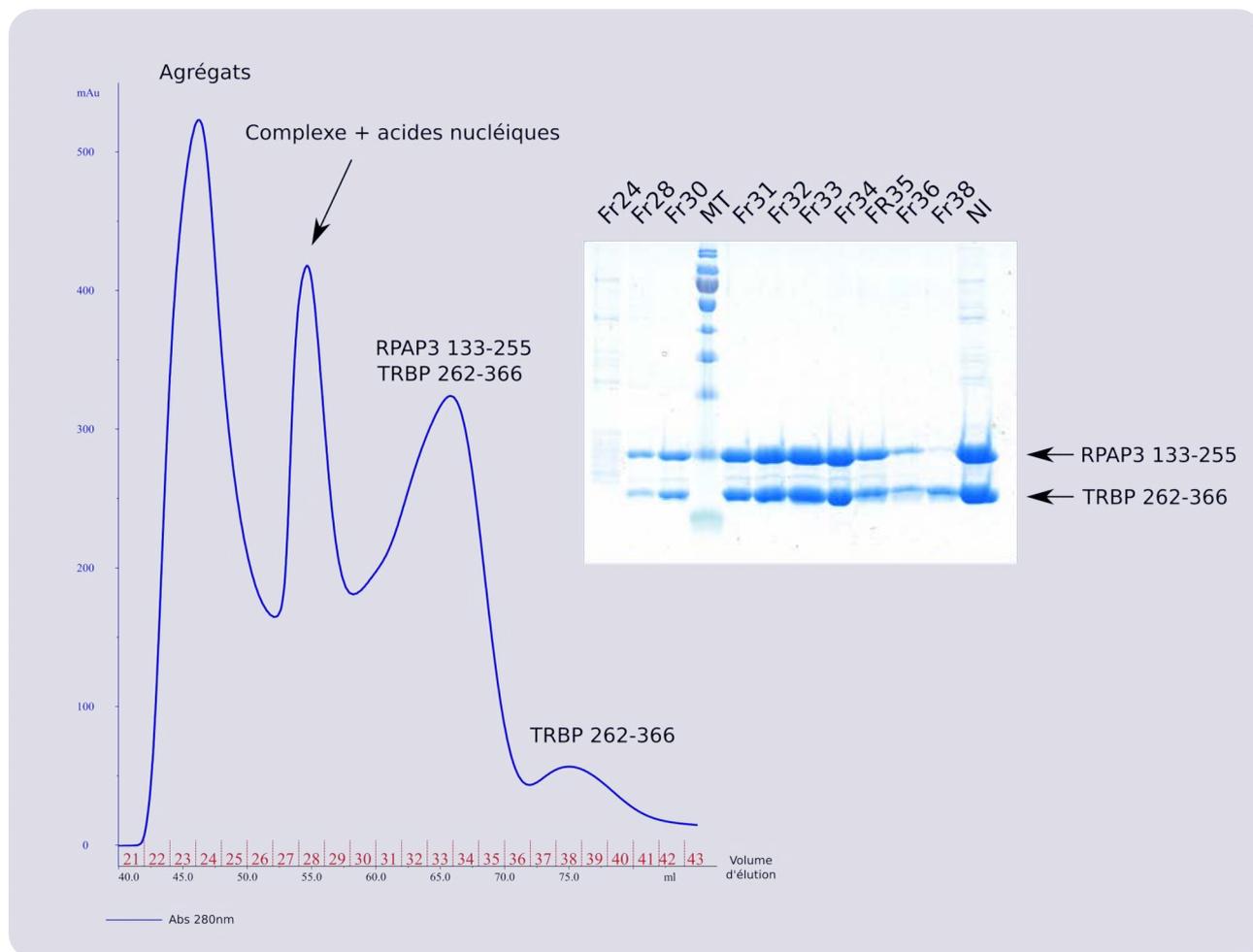


Figure 99: Purification d'un complexe TRBP 262-366/RPAP3 133-255 triplement marqué ^{15}N , ^{13}C , ^2H pour utilisation sur la plateforme RMN de Gif-sur-Yvette.

Le complexe a été produit comme précédemment décrit à partir d'une culture de deux litres de bactéries *E. coli* BL21 pRAR2. La production est effectuée dans un milieu minimum contenant 50 % d'eau lourde et en présence d'une source de carbone 13 et d'azote 15. Après purification sur une résine TALON™, le complexe est purifié par chromatographie d'exclusion stérique. Les fractions qui contiennent uniquement le complexe (32 à 35) sont ensuite combinées puis concentrées.

De nombreux acides aminés ont été sélectionnés, tant au niveau de RPAP3 que de TRBP, et nous avons ainsi réalisé une cinquantaine de mutations des deux protéines qui ont été testés par co-expression et co-purification. Sept variants, qui empêchent la formation du complexe entre TRBP et RPAP3, ont ainsi été identifiés, 2 au niveau de RPAP3 (D161A et V185A) et 5 au niveau de TRBP (Q324A, E328A, T331A, Q357A et R354E).

Ces résultats nous fournissent plusieurs informations. Ainsi, l'interface d'interaction entre RPAP3 et TRBP se situerait au niveau des hélices α_2 des motifs TPR1 et 2 du domaine TPR n°1 de RPAP3 et au niveau du feuillet β (plus précisément du brin β_2) et de l'hélice α_2 de TRBP. Ces données indiquent que le site d'interaction de ce domaine TPR avec HSP90 ne serait donc pas impliqué dans l'interaction avec TRBP. L'importance du résidu D161, et dans une moindre mesure, celle de la valine 185, dans la formation du complexe protéique nous fournit également une hypothèse pour expliquer l'absence d'interaction du second domaine TPR de RPAP3 avec TRBP, puisque ces résidus ne sont pas conservés au sein du second domaine TPR. Enfin, le fait que RPAP3 interagisse au niveau de TRBP exactement au même endroit que Dicer suggère fortement que ces deux interactions soient mutuellement exclusives. Nous avons ainsi décidé d'utiliser ces données structurales afin d'obtenir des informations concernant la ou les fonctions putatives du complexe TRBP/RPAP3.

Chapitre 2 : Étude de la fonction du complexe RPAP3-TRBP

I. Analyse des partenaires de TRBP et de RPAP3

D'après la littérature, les domaines TPR n°1 de RPAP3 et DRBM3 de TRBP seraient respectivement impliqués dans l'interaction de RPAP3 avec la protéine chaperonne HSP90 et dans l'interaction de TRBP avec les protéines Dicer ou PACT. Les mutations que nous avons effectuées sur ces domaines suggèrent par ailleurs que l'interaction de RPAP3 avec HSP90 serait toujours possible lorsque RPAP3 interagit avec TRBP, alors que Dicer ne devrait pas être capable de s'associer à TRBP si cette dernière est liée à RPAP3. Nous avons ainsi souhaité dans un premier temps confirmer *in vitro* ces hypothèses par des approches biochimiques et structurales (co-expression et RMN), ce qui devait nous aider à mieux appréhender la ou les fonctions biologiques du complexe protéique TRBP/RPAP3.

A. Le domaine 133-255 de RPAP3 interagit toujours avec HSP90

Pour confirmer l'existence d'une interaction entre le complexe TRBP 262-366/RPAP3 133-255 et la protéine HSP90, nous avons réalisé de nouvelles expériences de titrage RMN afin de déterminer une constante de dissociation (K_d) entre le peptide MEEVD d'HSP90 et le complexe TRBP 262-366/RPAP3 133-255. Il est en effet possible de déterminer cette constante en réalisant plusieurs expériences HSQC ^{15}N - ^1H successives à partir d'un complexe TRBP 262-366/RPAP3 133-255 marqué (^{15}N , ^{13}C) en présence de différentes concentrations du peptide MEEVD de la protéine HSP90 (non marquée). Ces expériences, réalisées par le Dr Quinternet, n'ont montré aucune différence significative concernant l'affinité d'HSP90 pour le domaine 133-255 de RPAP3, que la protéine TRBP soit présente ou non. Ces résultats confirment ainsi nos premières données sur le fait qu'une interaction entre RPAP3 et HSP90 est toujours possible en présence de TRBP, puisque les deux événements font intervenir des éléments distincts de RPAP3.

B. Absence de formation d'un complexe RPAP3-TRBP-Dicer

L'étude d'un complexe RPAP3/TRBP/Dicer n'est pas possible par RMN du fait de la taille de ces trois protéines, même dans le cas où seuls les domaines impliqués seraient utilisés. Nous avons donc choisi de réaliser des expériences de co-expression entre ces trois protéines. Il ne nous a toutefois pas été possible de produire la protéine Dicer sous sa forme entière (230 kDa) chez *E. coli* BL21. Nous avons donc utilisé le domaine 256-595 de Dicer, en se basant sur les informations

disponibles à ce moment-là dans la base de données Uniprot pour déterminer les bornes à utiliser de ce domaine. Ce choix a toutefois dans un premier temps été validé par la réalisation d'un test d'interaction double-hybride chez la levure *S. cerevisiae* entre ce domaine de Dicer et la protéine TRBP entière (**Figure 100**) ou son domaine 262-366 (**Figure 101**). Dans les deux cas, l'interaction *in vivo* du domaine 265-595 de Dicer avec TRBP est apparue aussi stable que l'interaction entre les deux protéines entières, confirmant ainsi que le domaine 256-595 de Dicer était suffisant pour la formation du complexe TRBP/Dicer.

Nous avons donc cloné la séquence codant le domaine 256-595 de Dicer dans le vecteur pnYKGW qui a été utilisé pour co-transformer des bactéries *E. coli* BL21 pRAR2 avec les vecteurs pnCS-RPAP3 133-255 et pnEA3ch-TRBP 262-366. Tout comme le vecteur pnCS, ce vecteur permet l'expression chez *E. coli* d'une protéine sans étiquette. Comme décrit précédemment, le complexe a été purifié sur résine TALON™, et la formation d'un complexe entre ces trois protéines a été vérifiée sur gel. Nous avons également vérifié que le domaine 256-595 de Dicer n'interagit pas de manière non-spécifique avec la résine à partir de bactéries *E. coli* BL21 pRAR2 transformées uniquement par le vecteur pnYKGW-Dicer 256-595 (**Figure 102**). Un témoin positif d'interaction entre les domaines 256-595 de Dicer et 262-366 de TRBP a également été réalisé (**Figure 102**).

Les résultats des tests de co-expression confirment bien l'interaction directe et spécifique du domaine 265-295 de Dicer avec le domaine 262-366 de TRBP (**Figure 102**). Les résultats de la co-expression entre les 3 partenaires putatifs (**Figure 103**) montrent clairement également la présence de 3 bandes dont les tailles apparentes correspondent aux tailles attendues des trois domaines, ce qui suggère qu'un complexe entre ces trois protéines existe. Pour le vérifier, nous avons ensuite soumis l'échantillon à une chromatographie d'exclusion stérique et les différentes fractions collectées ont ensuite été fractionnées sur un gel de protéines dénaturant (**Figure 104**). De manière surprenante, nous avons alors observé la présence non pas d'un, mais de deux complexes distincts, correspondant d'une part aux domaines 256-595 de Dicer et 262-366 de TRBP et d'autre part aux domaines 262-366 de TRBP et 133-255 de RPAP3. Nous avons également voulu vérifier si un résultat identique était retrouvé entre le domaine 256-595 de Dicer et les protéines TRBP et RPAP3 entières. Malheureusement, il semble que la protéine TRBP se dégrade en un fragment de taille comparable avec la taille apparente du domaine de Dicer, ce qui rend l'interprétation de ces expériences impossibles.

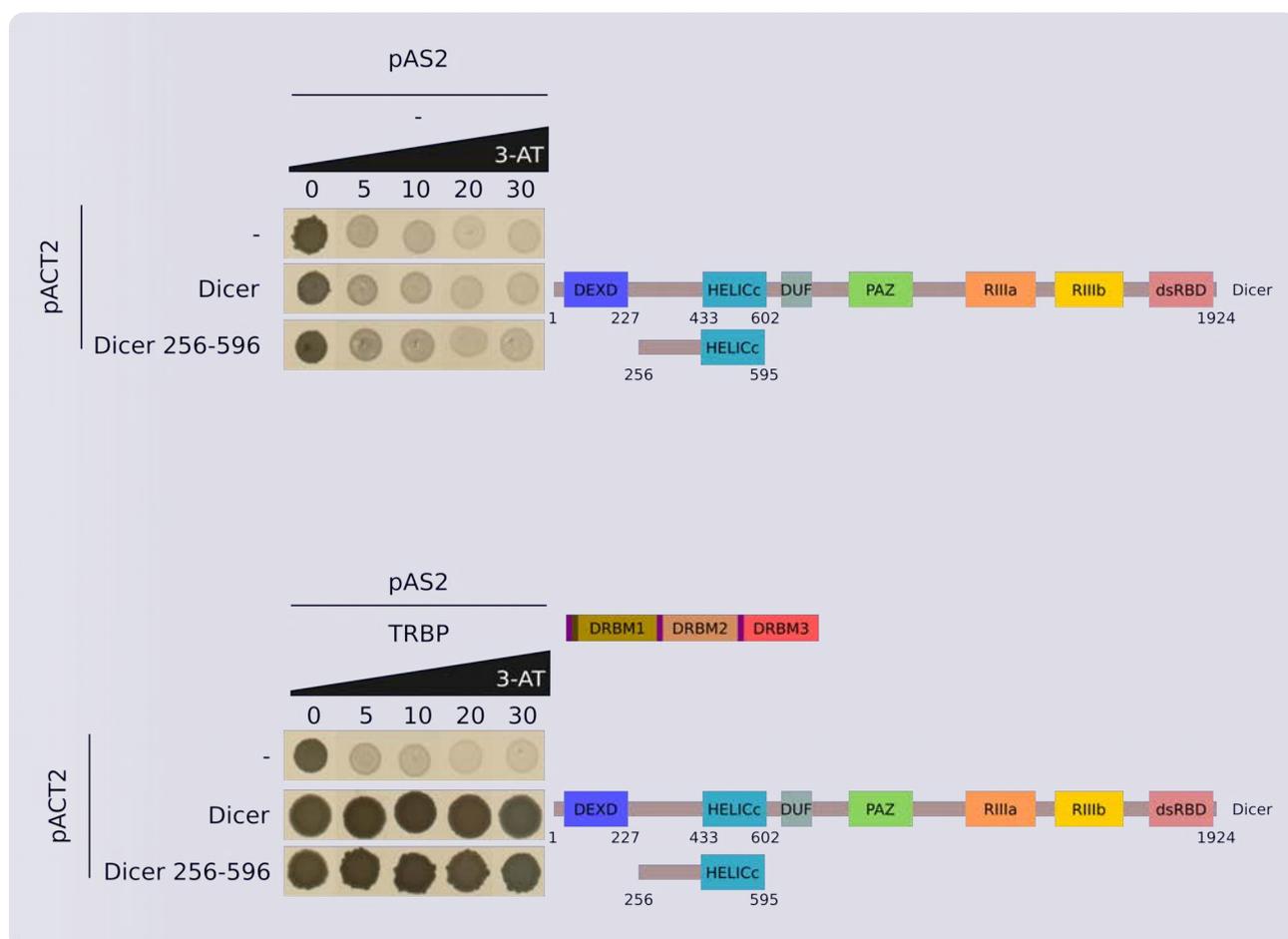


Figure 100: Le domaine 256-595 de Dicer et la protéine TRBP forment un complexe stable *in vivo*.

Un test d'interaction double-hybride chez la levure *S. cerevisiae* a été réalisé entre la protéine TRBP entière et le domaine 256-595 de Dicer. Un témoin positif a également été réalisé entre les protéines TRBP et Dicer entières. La stabilité du complexe *in vivo* est vérifiée par l'utilisation de 3-AT à différentes concentrations (de 5 à 30 mM).

Ces résultats montrent que le domaine 256-595 de Dicer est suffisant pour interagir avec la protéine TRBP *in vivo*.

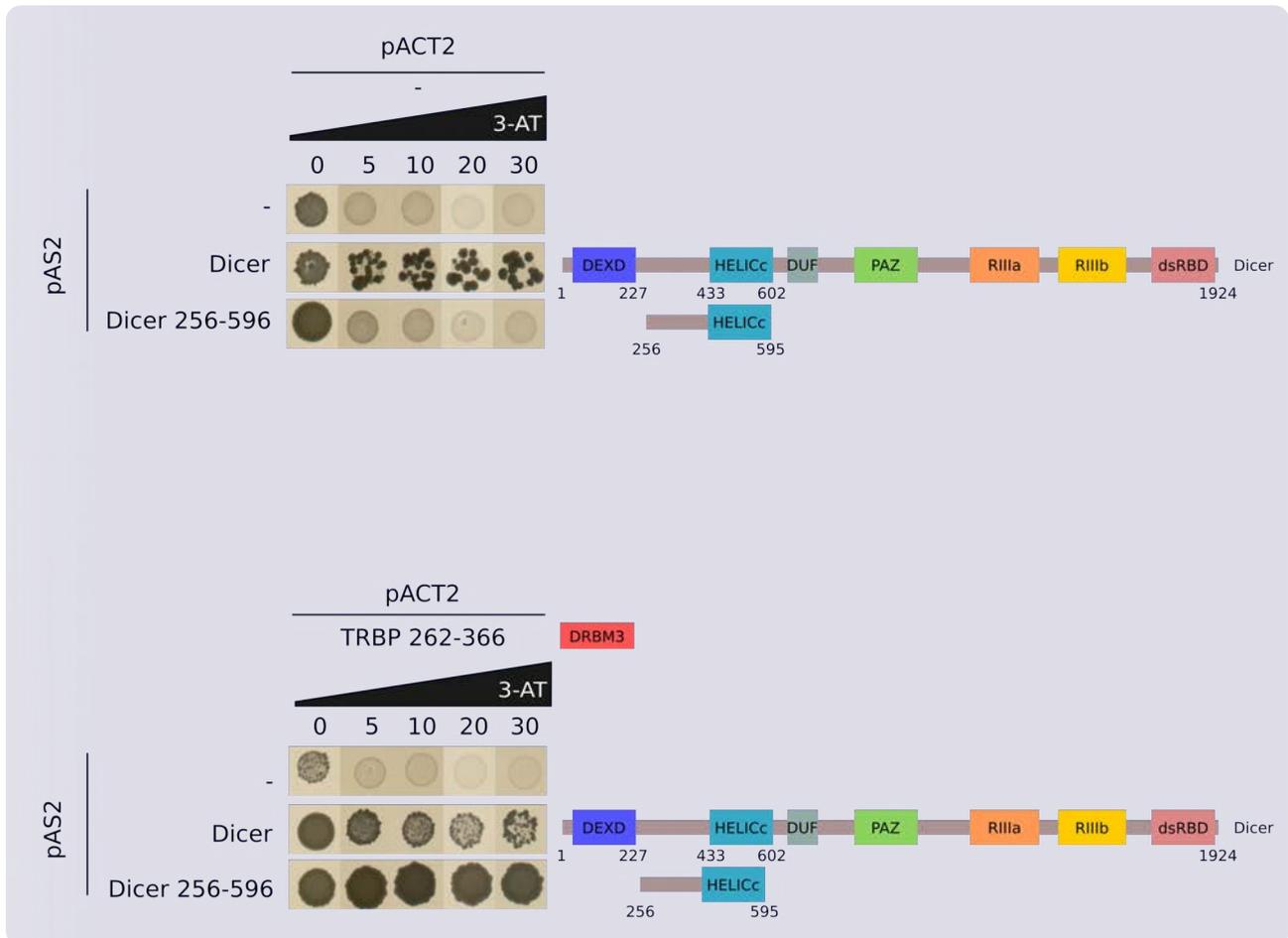


Figure 101 : Les domaines 256-595 de Dicer et 262-366 TRBP forment un complexe stable *in vivo*.

Un test d'interaction double-hybride chez la levure *S. cerevisiae* a été réalisé entre la protéine TRBP entière et le domaine 256-595 de Dicer. Un témoin positif a également été réalisé entre le domaine 262-366 de TRBP et la protéine Dicer entière. La stabilité du complexe *in vivo* est vérifiée par l'utilisation de 3-AT à différentes concentrations (de 5 à 30 mM).

Les résultats obtenus pour le témoin positif ne sont malheureusement pas exploitables puisque la protéine Dicer semble capable d'activer la transcription du gène *HIS3* lorsqu'elle est fusionnée au domaine de fixation à l'ADN de Gal4 (vecteur pAS2). Toutefois, ils confirment que le domaine 256-595 de Dicer est suffisant pour interagir avec le domaine 262-366 de TRBP *in vivo*.

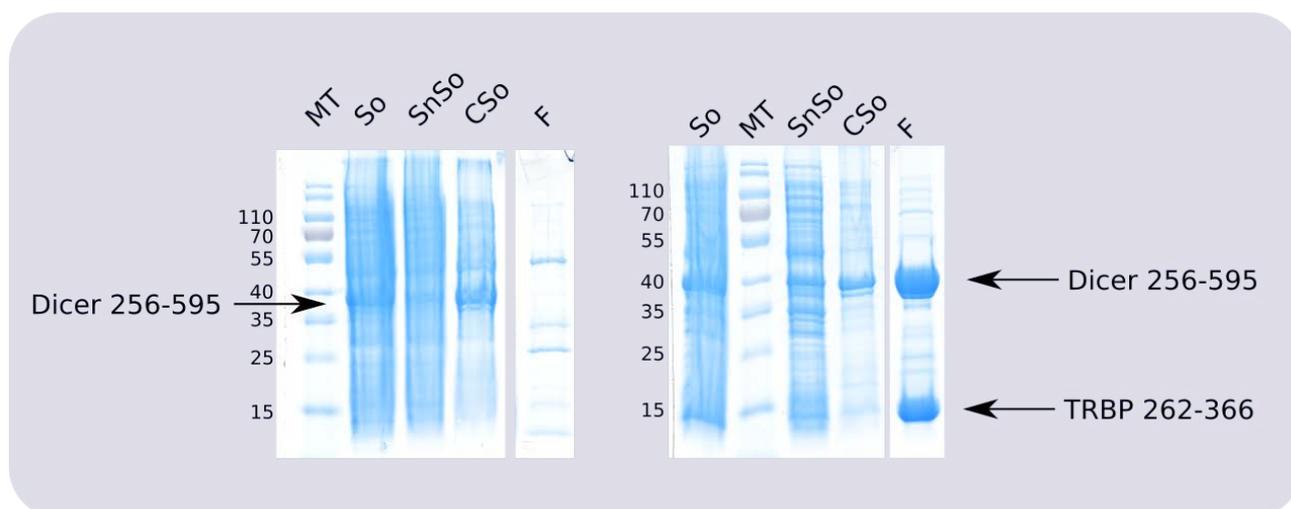


Figure 102: Témoins négatifs (à gauche) et positifs (à droite) de co-expression pour le domaine 256-595 de Dicer.

Les expériences de co-expression ont été réalisées à partir de bactéries *E. coli* BL21 pRAR2 transformées par le vecteur pnYKGW-Dicer 256-595 (témoins négatifs) ou par les vecteurs pnEA3ch-TRBP 262-366 et pnYKGW-Dicer 256-595 (témoins positifs).

La purification du complexe ou du domaine de Dicer a été réalisée sur une résine TALON™ qui permet de retenir spécifiquement la protéine TRBP 262-366 qui porte une étiquette 6-histidines.

Des aliquotes prélevées à différentes étapes de la purification sont ensuite fractionnées sur un gel de protéines dénaturant.

Ces gels montrent d'une part que le domaine 256-595 de Dicer peut être exprimé dans les bactéries *E. coli* BL21 pRAR2 (piste So) et qu'elle est majoritairement soluble (piste SnSo) et d'autre part qu'elle interagit directement et spécifiquement avec le domaine 262-366 de TRBP (piste F).

- So : Extrait total (Sonicat).
- SnSo : Surnageant de Sonication.
- Co : Culot de Sonication.
- F : Fraction fixée sur la résine.
- MT : marqueur de taille (en kDa).

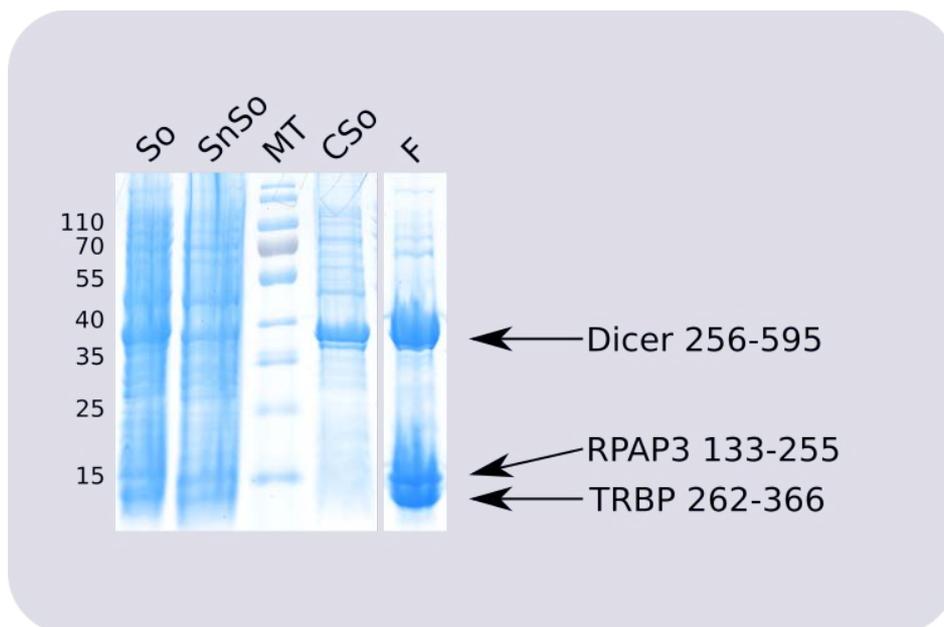


Figure 103 : Résultats de la co-expression entre les domaines 256-595 de Dicer, 133-255 de RPAP3 et 262-366 de TRBP.

Les expériences de co-expression ont été réalisées à partir de bactéries *E. coli* BL21 pRAR2 transformées par les vecteurs pnYKGW-Dicer 256-595, pnCS-RPAP3 133-255 et pnEA3ch-TRBP 262-366.

La purification du complexe a été réalisée sur une résine TALON™ qui permet de retenir spécifiquement la protéine TRBP 262-366 qui porte une étiquette 6-histidines.

Des aliquotes prélevées à différentes étapes de la purification sont ensuite fractionnées sur un gel de protéines dénaturant.

Les résultats montrent que les domaines 256-595 de Dicer et 133-255 de RPAP3 co-purifient avec le domaine 262-366 de TRBP, ce qui peut indiquer que ces trois domaines forment un complexe *in vitro*, toutefois il est nécessaire de réaliser une chromatographie d'exclusion stérique pour le confirmer.

- So : Extrait total (Sonicat).
- SnSo : Surnageant de Sonication.
- Co : Culot de Sonication.
- F : Fraction fixée sur la résine.
- MT : marqueur de taille (en kDa).

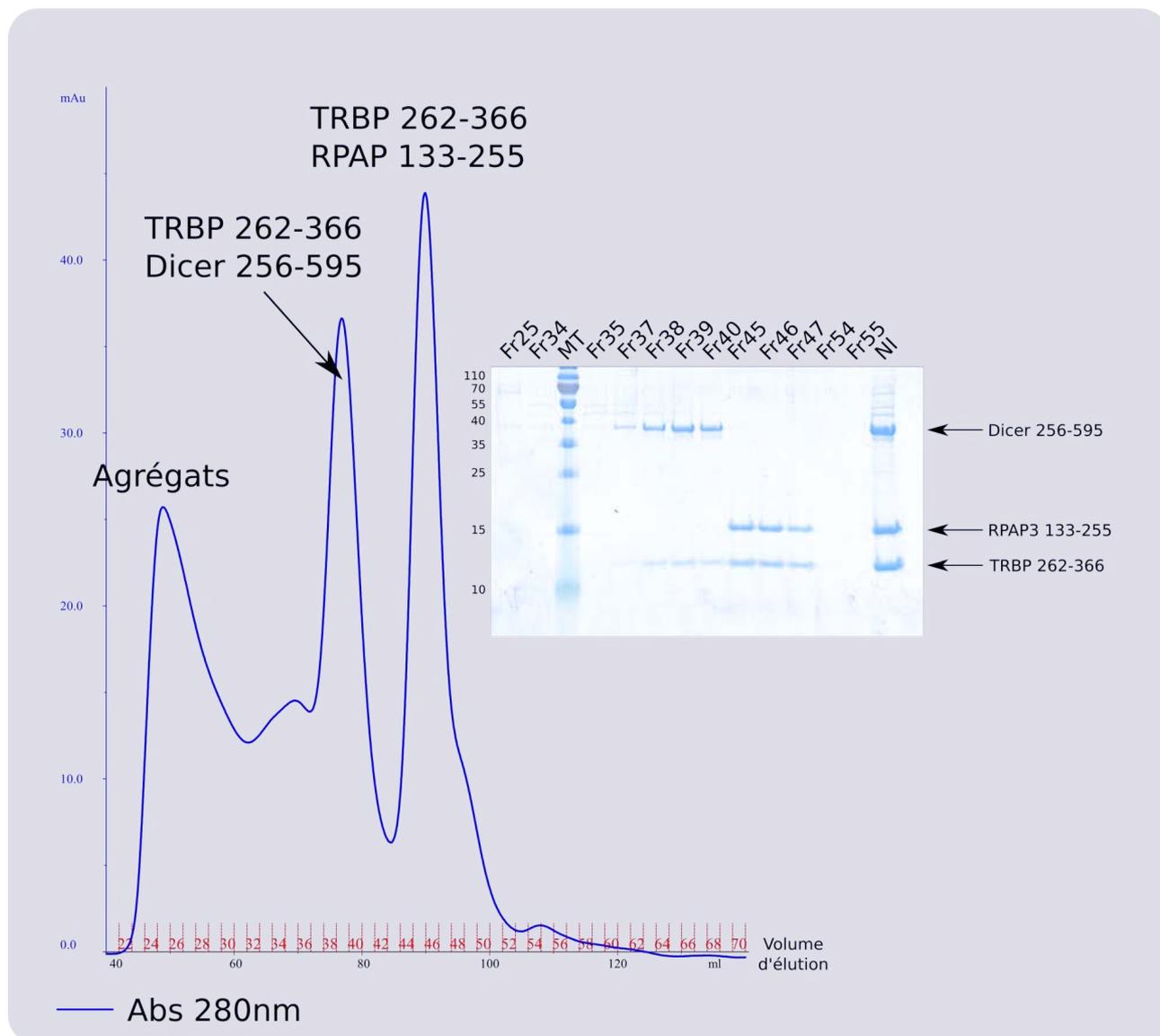


Figure 104: Les domaines 262-366 de TRBP, 133-255 de RPAP3 et 256-595 de Dicer ne forment pas de complexe à trois *in vitro*.

Après purification, le complexe est concentré puis injecté sur une colonne s200 préalablement équilibrée dans du tampon de purification (voir Matériel et Méthodes). Les complexes protéiques sont ensuite élués à une vitesse d'un millilitre par minute et des fractions de 2 ml sont collectées. Une aliquote de 10 μ l de chaque fraction est ensuite fractionnée sur un gel de protéines dénaturant.

Le chromatogramme montre que la solution injectée contient deux types de complexe, un premier entre les domaines 262-366 de TRBP et 256-595 de Dicer (volume d'éluion d'environ 76 ml) et un second entre les domaines 262-366 de TRBP et 133-255 de RPAP3 (volume d'éluion d'environ 94 ml).

Fr : Fraction

NI : Non Injecté

Toutefois, ces résultats et ceux obtenus suite aux expériences réalisées sur les différents variants de TRBP semblent valider notre hypothèse selon laquelle l'interface d'interaction avec les protéines Dicer et RPAP3 soit commune au niveau de TRBP.

C. RPAP3 n'interagit pas avec PACT

Comme décrit dans l'introduction, les protéines TRBP et PACT sont deux protéines paralogues qui interagissent notamment avec la RNase Dicer. Tout comme TRBP, la protéine PACT est composée de trois domaines DRBM et interagit avec Dicer par l'intermédiaire du troisième d'entre-eux. Lors du crible double-hybride effectué au cours de mon stage de master 2, aucune interaction n'a été visualisée entre les protéines RPAP3 et PACT, ce qui constituait une observation surprenante, puisque les domaines DRBM3 de PACT et TRBP partagent près de 53 % d'identité. De plus, un alignement de séquence réalisé entre ces deux domaines DRBM3 montre que les résidus que nous avons précédemment identifiés comme impliqués dans l'interaction avec RPAP3, à l'exception de l'arginine 354, sont conservés (**Figure 105**). Bien que TRBP et PACT interagissent toutes les deux avec Dicer, la littérature montre qu'elles ne présentent toutefois pas les mêmes fonctions : en effet, en présence de TRBP, la protéine Dicer va pouvoir cliver à la fois les pré-microARN et pré-siARN alors qu'en présence de PACT, d'une part, seuls les pré-microARN seront maturés par Dicer (Lee *et al.*, 2013), et d'autre par des isomiRs peuvent être générés (Wilson *et al.*, 2015). Nous avons donc voulu confirmer que ces deux protéines avaient bien un comportement différent vis-à-vis de RPAP3, ce qui nous pourrait nous éclairer/renseigner quant à la fonction du complexe TRBP/RPAP3.

Dans un premier temps, nous avons vérifié que les résultats de double-hybride n'étaient pas des faux-négatifs. Pour cela, nous avons réalisé de nouvelles expériences de co-expressions entre d'une part les protéines PACT et RPAP3 entières et d'autre part les domaines 133-255 de RPAP3 et le domaine de 212-313 de PACT dont l'alignement avec TRBP montre qu'il correspond au domaine 262-366 de TRBP. Les co-expressions ont été réalisées à partir de bactéries *E. coli* BL21 pRAR2 transformées par les vecteurs pnEA3ch-PACT (ou pnEA3ch-PACT 212-313) et pnCS-RPAP3 (ou pnCS RPAP3 133-255). Comme précédemment décrit, la protéine codée par le vecteur pnEA3ch est purifiée à l'aide de résine TALON™ et la présence de la seconde protéine est vérifiée sur gel de protéines dénaturant. Un témoin positif a également été réalisé entre les domaines 212-313 de PACT (pnEA3ch) et 256-595 de Dicer (pnYKGW).

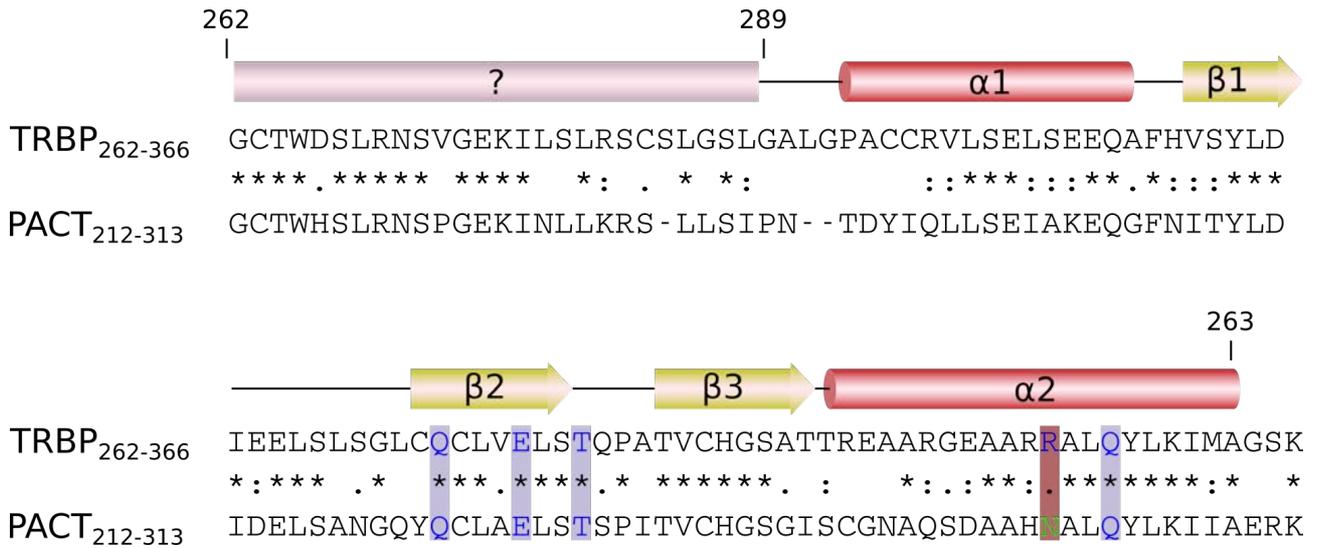


Figure 105: Alignement de séquences entre les domaines DRBM3 de TRBP (262-366) et de PACT (212-313).

L'alignement réalisé entre les troisièmes domaines DRBM des protéines TRBP et PACT montre qu'ils sont fortement conservés (53 % d'identité), notamment au niveau des brins β . De plus, les résidus impliqués dans l'interaction du domaine 262-366 de TRBP avec le domaine 133-255 de RPAP3 sont également conservés (en bleu), à l'exception notable de l'arginine 354 (en brun). Il semble ainsi difficile d'expliquer l'absence d'interaction observée entre les protéines RPAP3 et PACT sur la base de ces séquences.

Les résultats de co-expression (**Figure 106**) montrent clairement que la protéine PACT est incapable d'interagir directement avec la protéine RPAP3, que ce soit avec la protéine entière (**Figure 106A**) ou avec le domaine 133-255 (**Figure 106B**). Il est toutefois possible que ce résultat soit lié à un défaut de repliement de la protéine PACT qui ne se replie pas correctement *in vitro*. En effet, il semble qu'une quantité importante de la protéine PACT produite ne soit pas soluble (**Figure 106A et B**, piste Co). Toutefois, un résultat identique est observé concernant l'interaction entre les domaines 212-313 de PACT et 133-255 de RPAP3 (**Figure 106C**), or notre témoin positif montre clairement l'existence d'une interaction directe entre les domaines 212-313 de PACT et 256-595 de Dicer (**Figure 106C**). Ainsi, si un problème de solubilité est possible concernant les résultats obtenus avec la protéine PACT entière, ce n'est pas le cas pour ceux obtenus avec le domaine 212-313.

Ces expériences de co-expression confirment donc nos premiers résultats de double-hybride et montrent que l'interaction entre les protéines RPAP3 et TRBP est extrêmement spécifique. Nous avons ensuite essayé d'identifier les raisons qui pourraient expliquer l'absence d'interaction entre RPAP3 et PACT. Dans un premier temps, nous avons généré plusieurs variants du domaine 262-366 de TRBP en nous basant sur l'alignement de séquence réalisé avec PACT et sur nos précédents résultats. Ces mutants ont été identifiés selon différents critères. Ainsi, nous nous sommes uniquement intéressés aux résidus non conservés entre les domaines 262-366 de TRBP et 212-313 de PACT et situés soit sur le feuillet β , soit sur l'hélice α_2 . Pour chacun de ces résidus, nous nous sommes ensuite assurés, à l'aide de la structure du domaine 289-363 TRBP, que la chaîne latérale était bien exposée au solvant et non, par exemple, vers une autre partie du domaine de TRBP. L'alignement de séquence montre que les principales différences entre les domaines DRBM3 de TRBP et de PACT se situeraient au niveau des extrémités N-terminales des deux hélices α , or la structure du domaine de TRBP montre que les chaînes latérales de ces résidus sont dirigées vers le feuillet β et ne seraient donc pas accessibles. Les résidus des brins β sont également très conservés. Ainsi, il semble que les rares résidus répondant à nos critères soient principalement situés au niveau des boucles situées entre les brins β , à l'exception du résidu N354 dont nous savons qu'il est nécessaire à l'interaction avec RPAP3. Nous avons ainsi sélectionné les acides aminés L319, A334 et R354 de TRBP (**Figure 107**). Ces résidus correspondent respectivement aux acides aminés A266, I281 et N301 de PACT.

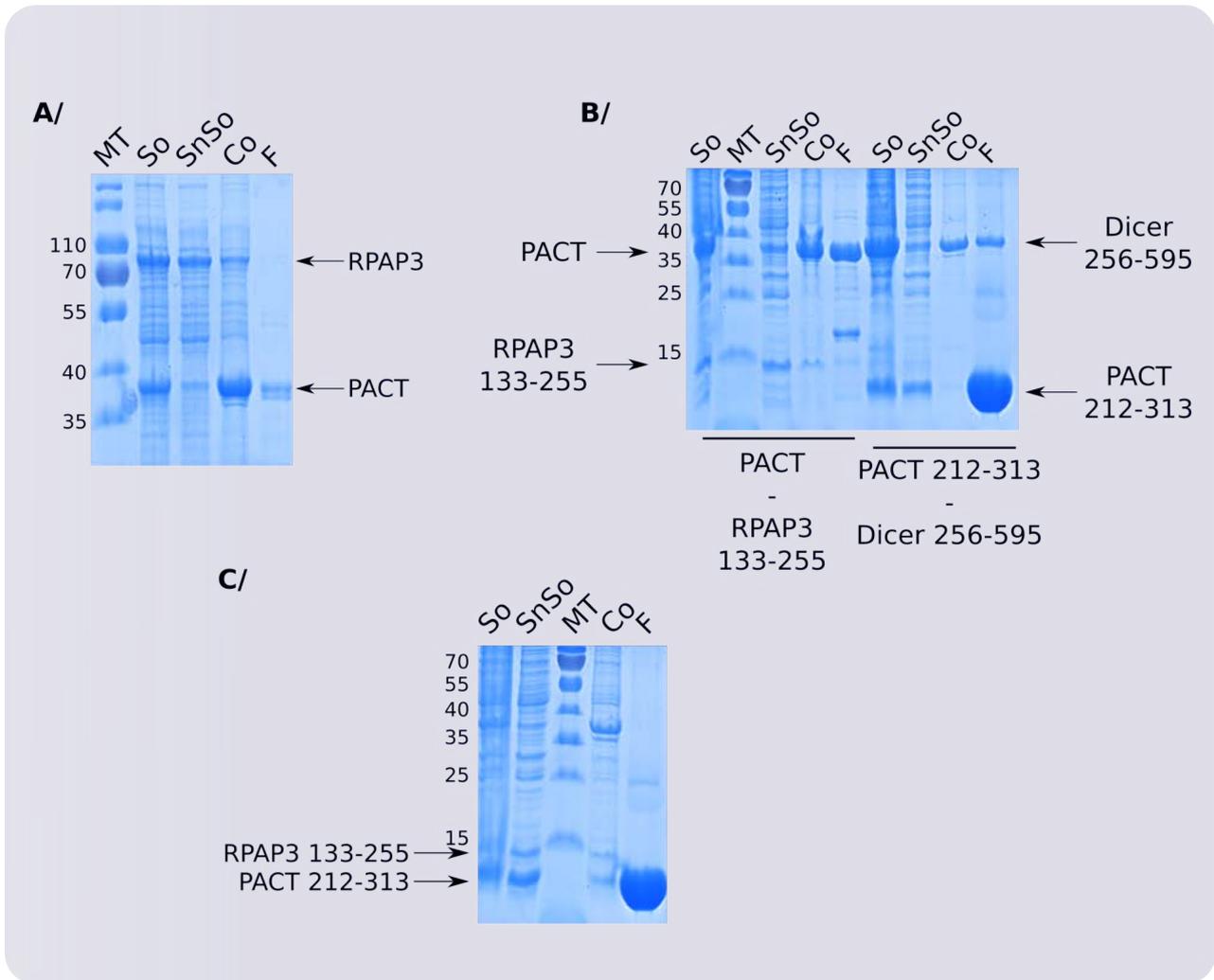


Figure 106: La protéine PACT n'interagit pas avec RPAP3.

Des expériences de co-expression ont été réalisées entre d'une part la protéine PACT (entière ou le domaine DRBM3 212-313) et la protéine RPAP3 (entière ou le domaine 133-255). Un témoin positif entre le domaine 212-313 de PACT et le domaine 256-595 de Dicer a également été réalisé.

Les co-expressions ont été réalisées dans les mêmes conditions qu'avec TRBP, en présence de 300 mM de NaCl et en étiquetant la protéine (ou le fragment) de PACT

Bien que la protéine PACT (ou le domaine 212-313) soit effectivement retenue sur la résine TALON™ (piste F des différents gels), aucune interaction n'est visible avec les autres protéines/domaines, à l'exception du domaine 256-595 de Dicer (piste F, gel B).

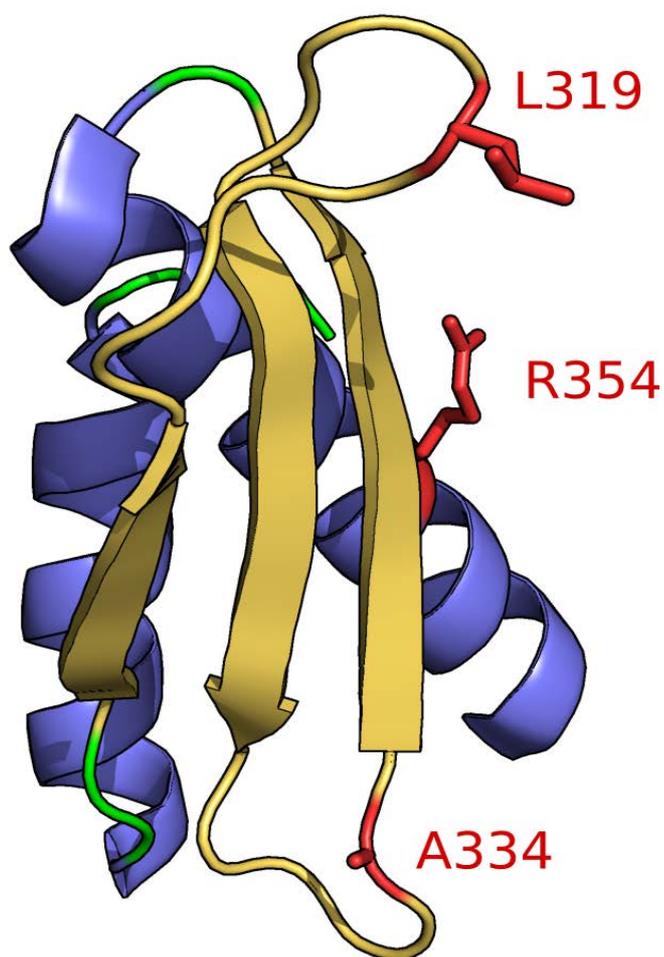


Figure 107: Variants du domaine 293-366 de TRBP sélectionnés pour étudier l'absence d'interaction entre PACT et RPAP3 (code PDB : 4WYQ).

Les résidus de TRBP les moins conservés entre les domaines DRBM3 de TRBP et de PACT et potentiellement impliqués dans des interactions, ont été sélectionnés puis mutés afin de correspondre aux acides aminés présents au niveau de PACT.

Ainsi, les acides aminés de TRBP L319, A334 et R354 ont été respectivement mutés en alanine, isoleucine et asparagine.

Nous les avons alors mutés respectivement en alanine, isoleucine et asparagine, afin de vérifier si les différences observées au niveau de ces acides aminés étaient suffisantes pour expliquer l'absence d'interaction entre PACT et RPAP3. Un triple variant L319A/A334I/R354N a également été généré. Enfin nous avons réalisé l'expérience inverse en réalisant la triple mutations A266L/I281A/N301R au niveau du domaine 212-313 de PACT, afin de vérifier si un changement au niveau de ces résidus permettrait l'interaction de ce domaine de PACT avec le domaine 133-255 de RPAP3. Les différents variants de TRBP ou de PACT ont ensuite été co-exprimés avec le domaine 133-255 de RPAP3 dans des bactéries *E. coli* BL21 pRAR2 puis purifiés sur une résine TALON™. Un témoin positif d'interaction a également été réalisé entre les domaines non mutés de TRBP et de RPAP3.

Malheureusement, aucune de ces expériences ne s'est révélée concluante. Ainsi, les différentes mutations ponctuelles réalisées n'ont pas été suffisantes pour empêcher la formation du complexe entre les domaines de TRBP et de RPAP3 (**Figure 108**), y compris la mutation de l'arginine 354 en asparagine. Les mutations compensatoires réalisées au niveau du domaine 212-313 de PACT n'ont pas non plus eu d'effet (**Figure 109**), alors que la triple mutation du domaine de TRBP semble surtout avoir pour conséquence de modifier la structure du domaine de TRBP qui n'est alors plus retenu sur la résine.

Bien que nous n'ayons pas réussi à identifier les causes de l'absence d'interaction entre PACT et RPAP3 à l'aide de ces variants, nous pouvons tout de même tirer quelques informations de ces expériences. Ainsi, il est intéressant de noter que la triple mutation du domaine de TRBP ait eu *a priori* un impact drastique sur le repliement de la protéine, alors que cela ne semble pas être le cas pour PACT. Il n'est donc pas impossible que les autres différences observées entre ces deux domaines aient plus d'importance que prévu. Ainsi, il est possible que la présence de résidus non conservés au niveau des extrémités N-terminales des deux hélices α ait pour conséquence de modifier légèrement le repliement du domaine, ce qui pourrait alors avoir une influence sur la fonction de PACT. Étant donné que le domaine 212-313 de PACT semble pouvoir être produit en quantité importante, nous avons ensuite voulu vérifier cette hypothèse en résolvant la structure tridimensionnelle de ce domaine par RMN. Un échantillon marqué ^{15}N a alors été produit et purifié puis une HSQC ^{15}N - ^1H a été réalisée par le Dr Quinternet.

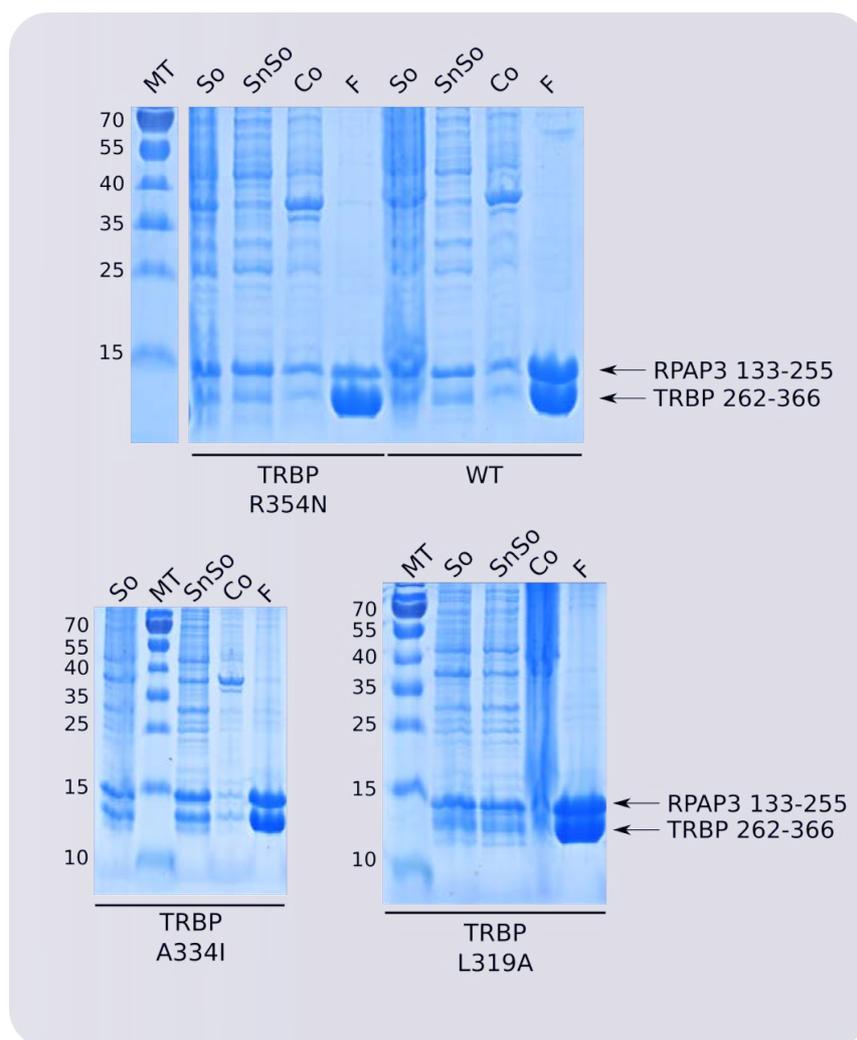


Figure 108 : Résultats des co-expressions réalisées sur les mutants L319A, A334I et R354N du domaine DRBM3 de TRBP.

Les co-expressions ont été réalisées en présence de 300 mM de NaCl et avec le domaine 262-366 de TRBP fusionné à une étiquette 6-histidines.

Quelles que soit les mutations réalisées, le complexe entre les domaines de RPAP3 et de TRBP est toujours présent (piste F), ce qui indique que ces résidus ne sont pas impliqués dans l'interaction et donc qu'ils ne sont pas un élément déterminant qui pourrait expliquer l'absence d'interaction entre PACT et RPAP3.

Toutefois, il semble que pour le mutant R354N, la stœchiométrie entre les deux protéines ne soit plus respectée, ce qui pourrait suggérer qu'en présence de ce variant, l'interaction est plus labile/instable.

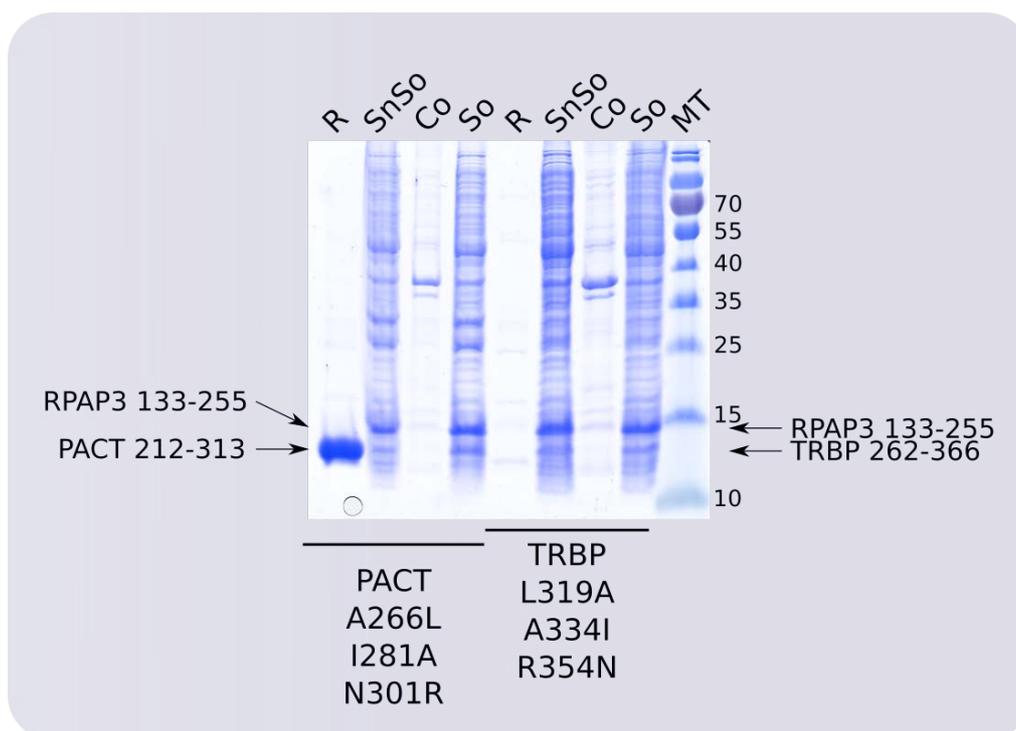


Figure 109: Résultats des co-expressions réalisées sur les triples variants L319A/A334I/R354N du domaine 262-366 de TRBP et A226L/I281A/N301R du domaine 212-313 de PACT.

Les co-expressions ont été réalisées en présence de 300 mM de NaCl et avec le domaine 262-366 de TRBP ou 212-313 de PACT fusionné à une étiquette 6-histidines.

La triple mutation réalisée au niveau de TRBP semble avoir un effet important sur son repliement, sa stabilité ou sa solubilité puisque la protéine n'est plus produite.

À l'inverse, la triple mutation compensatoire de PACT n'est toujours pas suffisante pour permettre son interaction avec le domaine 133-255 de RPAP3

De manière intéressante, le spectre HSQC a révélé la présence de près de 200 pics alors que seule une centaine était attendue, pour ce domaine d'environ 100 résidus, ce qui suggère que le domaine 212-313 de PACT pourrait former un homodimère. Un échantillon a alors été fourni pour une analyse par spectrométrie de masse native réalisée par le Dr Hortense Mazon, qui a permis de confirmer cette hypothèse. Ainsi, en solution, le domaine 212-313 de PACT est majoritairement (entre 80 et 90%) présent sous forme d'homodimère, ce qui pourrait expliquer pourquoi la protéine PACT ne forme pas de complexe avec la protéine RPAP3 contrairement à TRBP. Cette hypothèse est appuyée par plusieurs études qui ont montré que bien que les protéines TRBP et PACT puissent former des homo- et hétérodimères, elles ne le font pas par l'intermédiaire des mêmes domaines (Laraki *et al.*, 2008). Ainsi, la protéine PACT peut former des homodimères et hétérodimères (avec TRBP notamment) par l'intermédiaire de ses trois domaines DRBM, contrairement à la protéine TRBP dont l'homodimérisation se fait par l'intermédiaire de ses deux premiers domaines DRBM (**Figure 110**). Cette différence de disponibilité du dernier domaine DRBM pourrait alors expliquer l'absence d'interaction entre PACT et RPAP3.

D. Conclusion

Nos précédentes études structurales et biochimiques nous ont permis d'émettre plusieurs hypothèses concernant l'impact de la formation du complexe entre les protéines TRBP et RPAP3 sur les réseaux d'interaction respectifs de ces deux protéines, notamment avec les protéines Dicer et HSP90. Ainsi, la localisation probable de l'interface d'interaction entre les protéines TRBP et RPAP3 suggérait l'existence de contraintes stériques suffisantes pour empêcher la formation d'un complexe RPAP3/TRBP/Dicer, contrairement à l'interaction entre les protéines HSP90 et RPAP3. Grâce à la réalisation d'expériences de co-expression et de titrage RMN, nous avons pu confirmer ces 2 hypothèses. Ainsi, nous avons montré que l'interaction entre HSP90 et le TPR 133-255 était toujours possible une fois le complexe avec TRBP formé, alors que l'interaction entre Dicer et TRBP n'est pas possible lorsque TRBP est lié à RPAP3. De plus, nous avons confirmé que malgré l'existence d'une très forte conservation entre les domaines DRBM3 de TRBP et de PACT, cette dernière n'interagissait pas avec RPAP3. Bien que nous n'ayons pas identifié avec précision les éléments structuraux qui pourraient expliquer l'absence d'interaction entre PACT et RPAP3, nous avons observé la formation d'un homodimère qui pourrait l'empêcher.

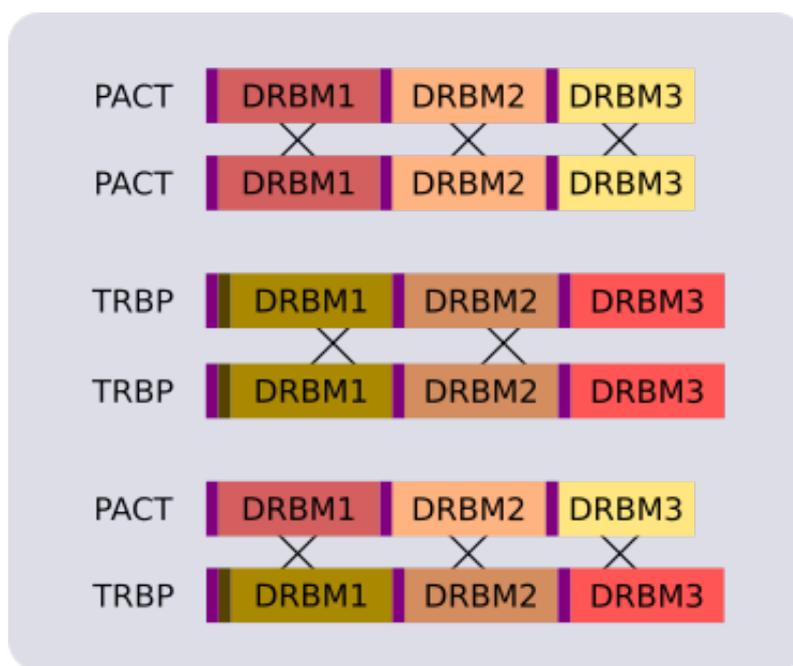


Figure 110: Bilan des interactions entre PACT et TRBP (d'après Laraki *et al.*, 2008).

La littérature montre que les protéines PACT et TRBP peuvent former des homo et hétérodimères. La protéine PACT forme un homodimère par l'intermédiaire de ses trois domaines DRBM alors que seuls les deux premiers domaines DRBM sont impliqués pour TRBP, ce qui indique que malgré l'existence d'une très forte homologie entre les domaines DRBM3 de TRBP et de PACT, il existe de légères différences fonctionnelles et structurales

De plus, la littérature indique que contrairement à TRBP, le domaine DRBM3 de PACT participe à l'homodimérisation de cette dernière. Si la protéine Dicer a semble-t-il une affinité suffisante pour PACT pour rompre l'homodimère, il semble donc que cela ne soit pas le cas de RPAP3. Nous n'avons toutefois pas été en mesure d'en identifier les raisons.

La littérature nous apprend également que la protéine RPAP3 n'interagit pas seulement avec la protéine HSP90 au niveau de son domaine TPR n°1 mais également au niveau du domaine TPR n°2 (Pal *et al.*, 2014), ainsi qu'avec la protéine PIH1D1 au niveau d'un domaine situé après le TPR n°2 (Yoshida *et al.*, 2013). Nous avons essayé de vérifier si un complexe à trois TRBP/RPAP3/PIH1D1 pouvait exister en réalisant de nouvelles expériences de co-expression, malheureusement, nous n'avons pas obtenu de résultats exploitables. En effet, non seulement la protéine PIH1D1 interagit de manière non-spécifique avec la résine TALON™, mais elle s'est également montrée instable et sensible à la dégradation. De plus, nous nous sommes rendu compte *a posteriori* que nos expériences avaient été réalisées avec l'isoforme II de RPAP3. Or, les données actuelles de la littérature indiquent que seule l'isoforme I de RPAP3 interagit avec la protéine PIH1D1. En effet, au sein de l'isoforme II, le domaine d'interaction entre RPAP3 et PIH1D1 n'est pas présent suite à un événement d'épissage alternatif (Yoshida *et al.*, 2013). Toutefois, la présence d'un domaine TPR entre les sites d'interaction de RPAP3 pour TRBP et PIH1D1 suggère fortement la possibilité de formation d'un complexe entre les protéines TRBP, RPAP3 et PIH1D1. Il serait toutefois intéressant de renouveler ces expériences avec l'isoforme I de RPAP3.

Nos travaux ont ainsi permis de mettre à jour les réseaux d'interaction des protéines RPAP3 et TRBP (**Figure 111**), ce qui nous permet d'émettre plusieurs hypothèses concernant les rôles physiologiques du complexe TRBP/RPAP3. En effet, ce n'est pas la première fois qu'une différence fonctionnelle est observée entre les protéines PACT et TRBP. Ainsi, comme décrit dans l'introduction, il a été montré que ces deux protéines avaient notamment un rôle différent lors de la génération d'isomiRs par la protéine Dicer (Wilson *et al.*, 2015). En effet, en l'absence de cofacteur ou en présence de PACT, il a été montré que la protéine Dicer pouvait, pour certains pré-microARN, générer des isomiRs, c'est-à-dire des microARN double-brins dont le site de coupure par Dicer est décalé d'un nucléotide (**Figure 112**).

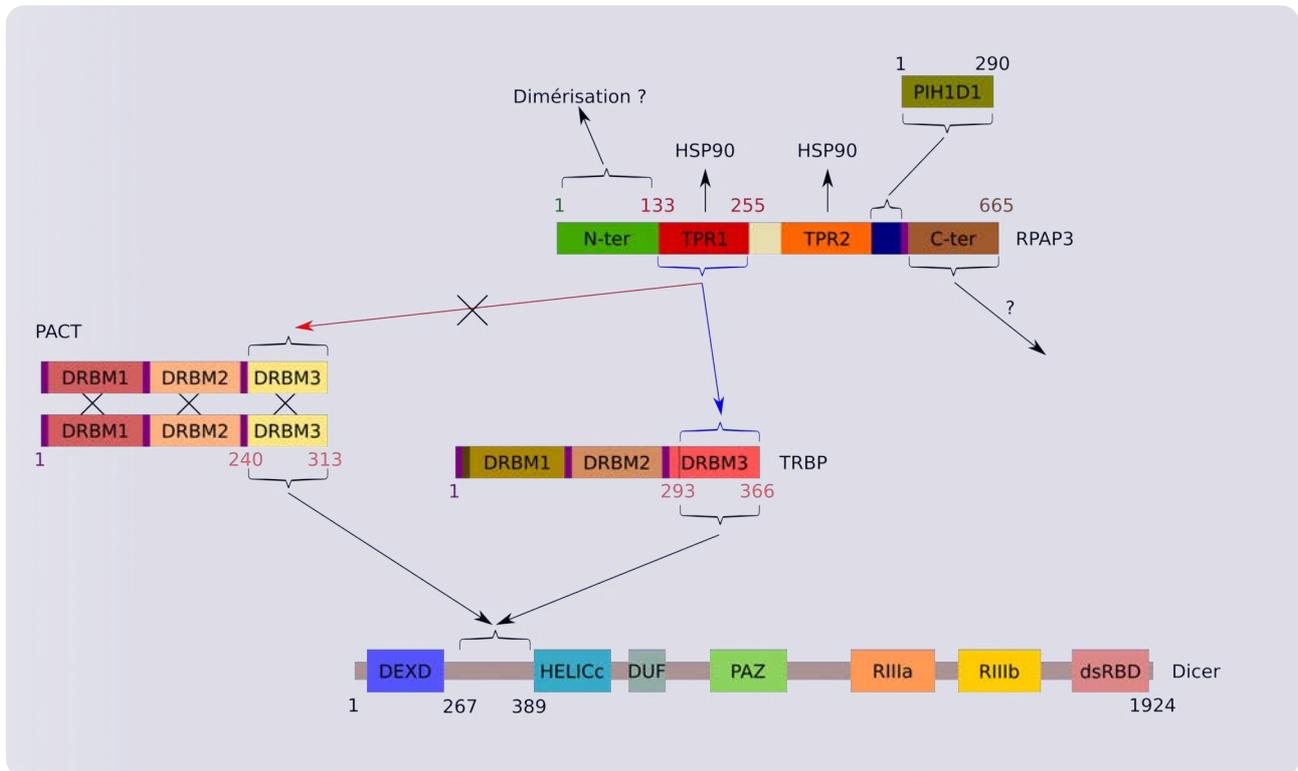


Figure 111: Récapitulatif des réseaux d'interactions des protéines RPAP3 et TRBP suite à ce travail de doctorat.

La littérature montre que la protéine RPAP3 interagit par l'intermédiaire de ces deux domaines TPR avec deux protéines HSP90 (Pal *et al.*, 2014). Ses domaines N- et C-terminaux ne sont toutefois pas encore caractérisés et au niveau de l'isoforme I, un domaine d'interaction avec la protéine PIH1D1 est présent (Pal *et al.*, 2014). Nous avons montré que le domaine TPR n°1, et uniquement lui, pouvait interagir avec le domaine DRBM3 de TRBP, mais pas avec PACT, dont la présence sous forme de dimère pourrait expliquer cette absence d'interaction. Enfin, nous avons montré que les complexes TRBP DRBM3 / RPAP3 TPR n°1 et TRBP / Dicer étaient mutuellement exclusifs.

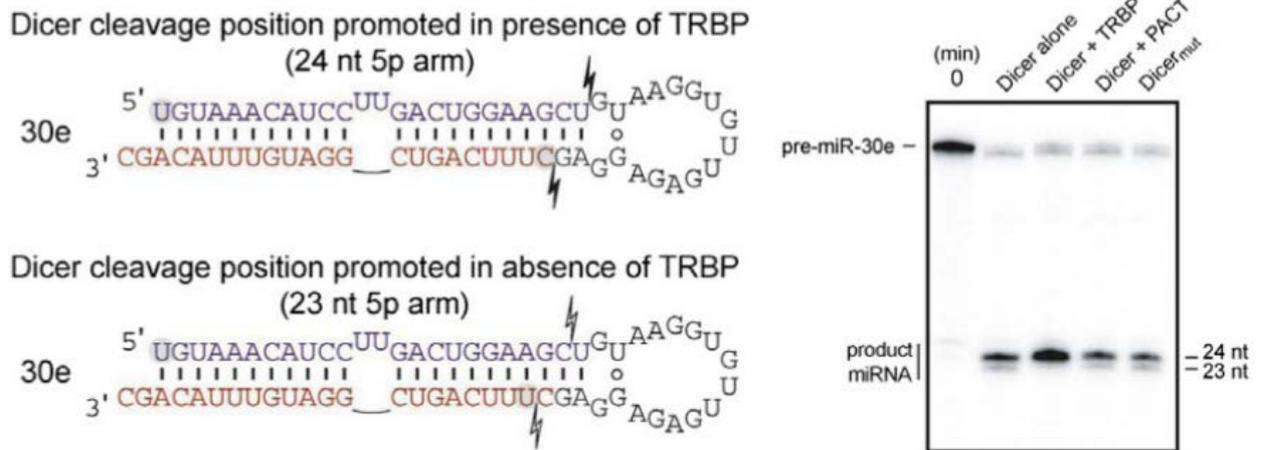


Figure 112: Résultats tirés de Wilson *et al.*, 2015 montrant la génération d'isomiRs par Dicer en l'absence de TRBP.

Wilson et al ont montré qu'en l'absence de TRBP ou en présence de PACT, la RNase Dicer pouvait, dans certains cas, générer un microARN double-brin dont le site de coupure est décalé d'un nucléotide, formant alors un isomiR.

Le changement du résidu présent au niveau des extrémités du microARN double-brin peut alors modifier la séquence reconnue par le microARN au niveau des ARNm et donc ses cibles. De plus, dans certains cas ce changement de résidu peut avoir un effet sur la sélection du brin guide et du brin passager qui est en grande partie basée sur la stabilité thermodynamique des extrémités 5' du microARN double-brin. L'interaction entre RPAP3 et TRBP étant observée à la fois dans le cytoplasme et dans le noyau, il est possible que l'une des fonctions du complexe TRBP/RPAP3 soit de réguler la disponibilité de TRBP vis-à-vis de Dicer, régulant ainsi la génération d'isomiRs par Dicer. Nous avons donc voulu tester cette hypothèse. Pour cela, nous avons tenté de reproduire les expériences de production d'isomiRs par Dicer publiées par Wilson *et al.*, en présence ou non de la protéine/RPAP3.

Cette hypothèse sur le rôle du complexe TRBP/RPAP3 n'est pas la seule possible. Afin d'obtenir une vision plus globale sur le rôle de ce complexe dans la biogenèse des ARNnc, nous avons réalisé des expériences d'ARN interférence avec des siARN dirigés contre les protéines RPAP3 ou TRBP pour analyser l'effet de la perte d'expression de l'une ou l'autre de ces protéines sur la maturation de l'ensemble des ARNnc à l'aide de puces à ARN. De plus, comme nous avons précédemment identifié certains des acides aminés de TRBP et de RPAP3 impliqués dans la formation du complexe, il pourrait être intéressant de les utiliser afin d'étudier l'effet spécifique de la perte d'interaction entre les deux protéines sur la maturation des ARNnc. Mais pour cela, nous devons nous assurer que ces mutations n'ont pas d'effet sur les autres partenaires potentiels de TRBP et de RPAP3. Des expériences de co-expression avec ces différents partenaires et d'immunoprécipitation ont ainsi été réalisées.

II. Caractérisation des variants de TRBP

Le domaine 262-366 de TRBP n'étant malheureusement pas stable sous sa forme libre, il nous est impossible d'utiliser des techniques de biophysique de type DC (« Dichroïsme Circulaire ») afin d'analyser l'effet des différentes mutations sur le repliement de ce domaine. TRBP ayant comme principal partenaire la protéine Dicer, nous avons décidé de réaliser des expériences de co-expression entre les domaines 262-366 de TRBP, sauvage ou muté, et 256-595 de Dicer. Ainsi, la visualisation d'une interaction nous indiquera non seulement que le repliement du domaine de TRBP n'a pas *a priori* été fortement modifié par la mutation, mais également que la mutation n'a pas d'effet sur l'interaction de TRBP avec Dicer. Dans un tel cas, il serait alors envisageable

d'utiliser cette construction pour tester *in vivo* l'effet de la perte d'interaction entre RPAP3 et TRBP tout en ayant l'assurance d'avoir limité *a minima* les effets que pourrait induire la mutation en dehors du complexe TRBP/RPAP3.

Deux mutations de TRBP, E328A et R354E, ont été sélectionnées sur les 5 précédemment identifiées. La première est localisée au niveau du feuillet β du domaine 262-366 de TRBP, la seconde au niveau de son hélice α_2 . Toutes les deux sont suffisantes pour empêcher l'interaction de TRBP avec RPAP3, mais nous avons voulu vérifier qu'elles n'empêchaient pas l'interaction de TRBP avec Dicer. Nous avons ainsi réalisé des co-expressions entre d'une part le domaine 262-366 de TRBP E328A et le domaine 256-595 de Dicer (**Figure 113**) et entre les domaines 262-366 de TRBP sauvage (WT) ou R354E, 133-255 de RPAP3 et 256-595 de Dicer (**Figure 114**). Les résultats montrent qu'aucune de ces mutations n'est suffisante pour empêcher l'interaction du domaine 262-366 de TRBP avec le domaine 256-595 de Dicer *in vitro*, ce qui suggère fortement que le repliement du domaine 262-366 de TRBP n'est pas affecté par l'une ou l'autre de ces mutations.

Nous avons également souhaité valider ces mutations *in vivo*. Pour cela, la protéine V5-TRBP sauvage ou portant la mutation E328A a été exprimée de manière transitoire dans des cellules HEK, puis des immunoprécipitations ont été réalisées à l'aide d'un anticorps dirigé contre l'étiquette V5. Ainsi, nous avons confirmé par western-blot que la mutation empêchait bien l'interaction entre TRBP et RPAP3 *in vivo* (**Figure 115**). En revanche, nous n'avons pas été en mesure de visualiser l'association *in vivo* de TRBP (sauvage ou mutée) avec Dicer. Cette dernière n'étant pas non plus visible au niveau du témoin d'expression interne (« input »), nous prévoyons de les renouveler prochainement. Ces expériences seront également prochainement réalisées sur le second variant sélectionné, R354E.

Ces premiers résultats sont toutefois encourageants et jusque-là maintenant nos données *in vivo* confirment nos résultats *in vitro*. Ces variants vont donc être prochainement utilisés pour étudier l'effet de l'absence d'interaction entre TRBP et RPAP3 sur la maturation des ARNnc. Pour cela, les protéines TRBP E328A ou R354E seront exprimées transitoirement dans des cellules HEK pendant 72h, puis les ARN totaux de ces cellules seront extraits et enfin séquencés et/ou analysés sur des micropuces à ARN.

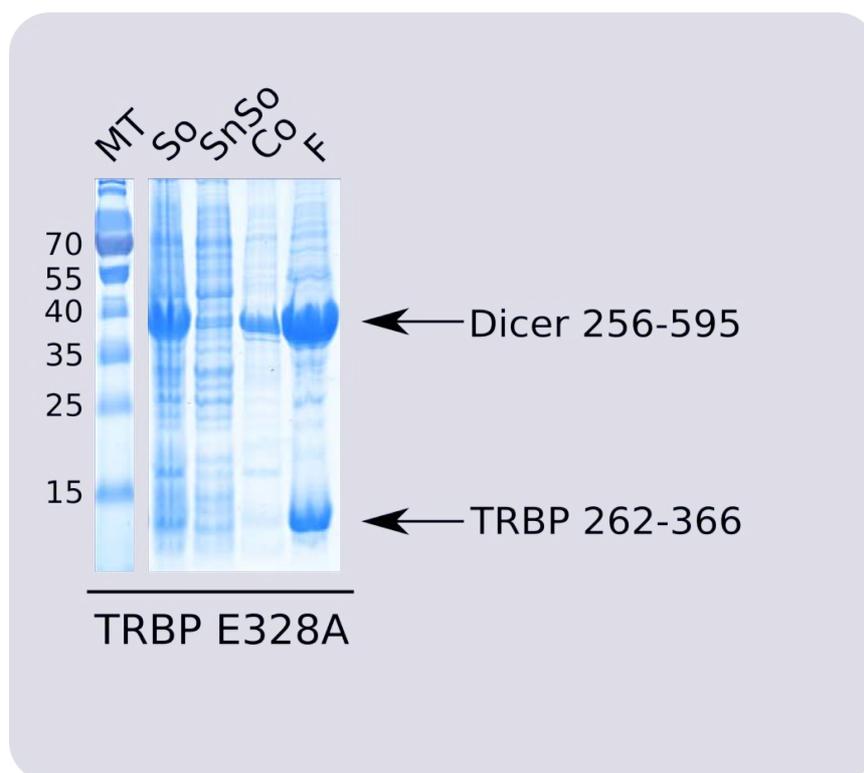


Figure 113: La mutation E328A de TRBP n'affecte pas l'interaction entre les domaines 262-366 de TRBP et 256-595 de Dicer.

La co-expression du domaine 262-366 de TRBP, dont le résidu E328 a été muté, avec le domaine 256-595 de Dicer a été effectuée de la même manière que précédemment décrit.

Bien qu'une quantité importante du domaine de Dicer ne soit pas soluble (piste Co), l'expérience permet de conclure que la mutation E328A n'empêche pas l'interaction entre ces deux domaines, ce qui suggère également que cette mutation n'a pas d'effet significatif sur le repliement du domaine 262-366 de TRBP.

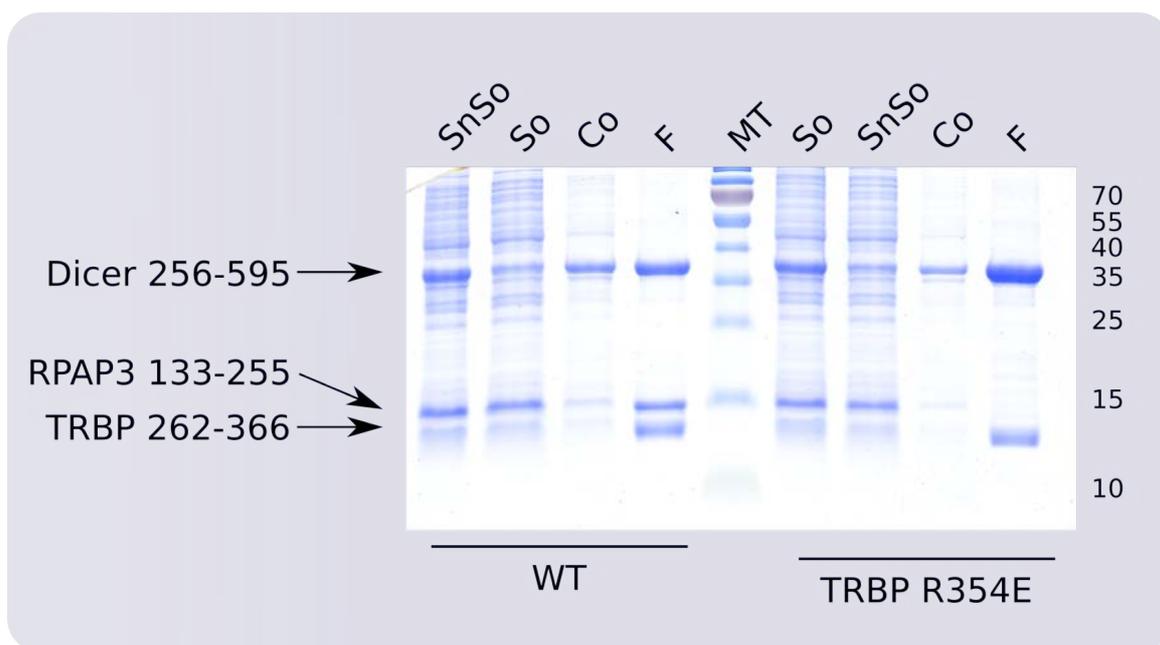


Figure 114: La mutation R354A de TRBP n'affecte que l'interaction entre les domaines de TRBP et de RPAP3.

La co-expression du domaine 262-366 de TRBP, dont le résidu R354 a été muté, avec le domaine 256-595 de Dicer et 133-255 de RPAP3 a été effectuée de la même manière que précédemment décrit.

L'expérience permet d'une part de conclure que la mutation R354A de TRBP n'empêche pas l'interaction entre TRBP et Dicer, et d'autre part confirme que l'interaction entre TRBP et RPAP3 est bien perdue en présence de cette mutation.

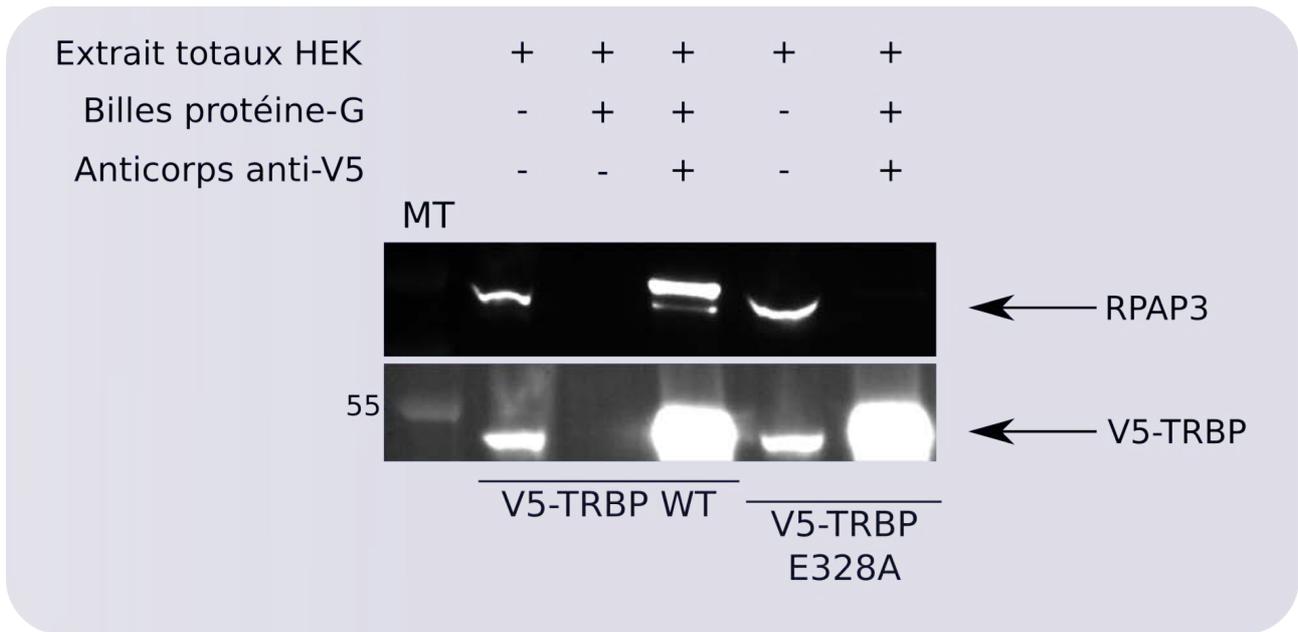


Figure 115: *In vivo*, la mutation E328A de TRBP empêche son association avec RPAP3

Après transfection, les cellules HEK sont lysées et la protéine V5-TRBP WT ou E328A est immunoprécipitée à l'aide d'un anticorps dirigé contre l'étiquette V5. Un western-blot est alors réalisé afin de vérifier d'une part que la protéine V5-TRBP a bien été immunoprécipitée (à l'aide d'un anticorps dirigé contre TRBP) et d'autre part que la protéine RPAP3 n'interagit plus avec TRBP lorsque le résidu E328 est muté.

Des contrôles sont réalisés afin de vérifier d'une part que les protéines sont bien exprimées et détectées dans les cellules HEK et d'autre part que la protéine RPAP3 n'interagit pas de manière non spécifique avec les billes magnétiques protéine-G.

Les résultats montrent qu'en présence de la mutation E328A, la protéine V5-TRBP n'interagit plus avec RPAP3.

Afin d'éliminer les effets potentiellement causés par la perte d'interaction de TRBP avec d'autres facteurs que Dicer, une expérience symétrique sera également réalisée avec la protéine RPAP3 mutée au niveau de son acide aspartique 161. Ainsi, seuls les effets retrouvés dans ces deux expériences seront analysés.

III. Rôle dans la génération d'isomiR ?

A. Production et purification du pré-microARN 30E

Comme discuté précédemment, l'une de nos hypothèses sur le rôle de l'interaction entre TRBP et RPAP3 pourrait correspondre à un effet indirect de régulation de l'activité de Dicer, via la séquestration de TRBP, pouvant conduire à la formation d'isomiR. Pour tester cette hypothèse, nous avons souhaité reproduire les expériences de Wilson *et al.* qui montrent la production *in vitro* d'un isomiR issu du pré-miR NN. Brièvement, le principe du test consiste à cliver le pré-microARN par Dicer en présence ou non de l'un de ses cofacteurs, tel que TRBP. Dans notre cas, nous avons voulu réaliser l'expérience en présence des protéines TRBP et RPAP3. Cette expérience nécessitait donc d'une part de produire et purifier le pré-microARN, mais également les protéines Dicer, TRBP et RPAP3. De plus, puisque le but de l'expérience est de détecter un décalage au niveau du site de coupure du pré-microARN, celui-ci doit être, à l'issue de sa transcription, strictement uniforme pour permettre la détection précise des isomiR. Pour cela, nous avons choisi d'une part de travailler sur le pré-microARN 30E, soit le même que Wilson *et al.*, et d'autre part de le purifier tel que décrit dans la publication, en s'appuyant sur l'utilisation de deux ribozymes. En effet, généralement, la transcription *in vitro* d'un ARN nécessite l'ajout en 5' de séquences qui vont améliorer l'efficacité de la transcription. De plus, la transcription ne s'arrête généralement pas au niveau d'un nucléotide précis. L'utilisation de deux ribozymes, l'un en 5' et l'autre en 3' du pré-microARN, doit ainsi permettre une maturation précise des extrémités 5' et 3' du pré-microARN 30E. L'ADN codant le pré-microARN a donc été cloné dans un vecteur de transcription bordé respectivement en 5' et en 3' par les séquences des ribozymes HammerHead (HH) et HDV (**Figure 116A**). Une fois transcrit, l'ARN marqué est chauffé puis refroidi plusieurs fois dans un thermocycleur, en présence d'une forte concentration en magnésium (40 mM), dans le but de permettre aux ribozymes de se replier correctement et donc de cliver l'ARN aux sites choisis au niveau des extrémités 5' et 3' du pré-microARN 30E.

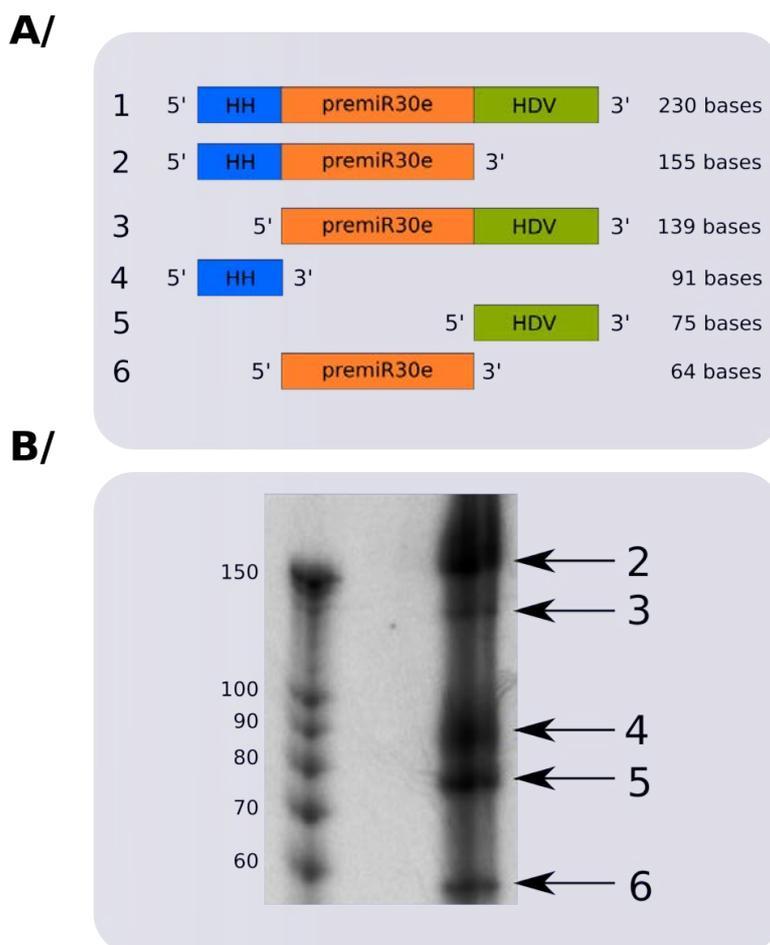


Figure 116: Production *in vitro* du pré-microARN 30E à partir d'un précurseur ARN clivé en 5' et en 3' respectivement par les ribozymes HammerHead (HH) et HDV.

A/ Représentation des différents fragments d'ARN pouvant être obtenus à partir d'un ARN transcrit contenant 2 ribozymes.

B/ Image du gel dénaturant d'acrylamide-urée 8M sur lequel a été fractionnée une partie de du pré-microARN 30E produit *in vitro*.

L'utilisation de ribozymes permet de cliver l'ARN transcrit (forme 1) au niveau de sites précisément définis, soit respectivement au niveau des extrémités 5' et 3' du pré-microARN 30E par les ribozymes HammerHead (HH) et HDV.

Ces différentes étapes de clivage n'étant pas efficaces à 100 %, 5 formes, en plus de l'ARN précurseur, peuvent être présentes en solution, en fonction du/des ribozyme(s) actif(s). Ainsi, si seul l'HDV a clivé l'ARN transcrit, les formes 2 et 5 seront présentes ; pour l'HH, se seront les formes 3 et 4 qui seront observées. Enfin, seule la forme 6 correspond au pré-microARN 30E mature, clivé à ses extrémités par les deux ribozymes.

Ces formes ont des tailles différentes et peuvent donc être isolées sur un gel dénaturant acrylamide-urée.

150-100-90-80-70-60 : Taille du marqueur de taille ARN, en bases.

Toutefois, l'efficacité de coupure des ribozymes n'est pas de 100 %. Après fractionnement sur gel d'acrylamide-urée, jusqu'à 6 bandes de taille différentes sont observées (**Figure 116A et B**), correspondant aux différents intermédiaires pouvant être générés suite aux coupures du transcrit par l'un, l'autre ou les deux ribozymes. La bande d'intérêt est alors identifiée par sa taille (forme 6 sur la Figure 116A et B).

Une fois la production du pré-microARN au point, l'expérience a été reproduite sur un transcrit non radiomarké. La bande correspondant au pré-microARN a alors été découpée, purifiée et déphosphorylée, avant d'être marquée au niveau de son extrémité 5' puis incubée avec les protéines Dicer, TRBP et/ou RPAP3.

B. Production et purification des protéines TRBP, Dicer et RPAP3 entière

Afin de tester la production d'isomiR à partir du pré-microARN 30E, il était donc également nécessaire de produire et de purifier les protéines Dicer, TRBP et RPAP3. Pour cela, nous avons utilisé deux stratégies : la production en système bactérien ou l'utilisation d'un baculovirus. En effet, de nombreux tests d'expression ont montré que seules TRBP et RPAP3 étaient exprimées de façon satisfaisante en système bactérien, et que, après les premiers tests de purification, TRBP n'était pas stable une fois purifiée. Nous avons donc opté pour une production via l'infection de cellules d'insecte Sf9 par un baculovirus recombinant pour produire les protéines Dicer et TRBP. RPAP3 a quant à elle pu être produite de manière « classique », à partir de 5 l de culture d'*E. coli* BL21, puis purifiée sur une colonne de nickel par HPLC avant une dernière étape de chromatographie d'exclusion stérique (voir « Matériel et Méthodes »).

Pour Dicer et TRBP, la production et la purification furent nettement plus complexes. En effet, la technique de production par baculovirus étant nouvelle au laboratoire, il a fallu réaliser de multiples mises au point, que ce soit au niveau de la production du baculovirus, que de la purification elle-même.

1. Production et purification de la protéine TRBP

Pour TRBP, le protocole suivant a pu être utilisé avec succès. Tout d'abord, un baculovirus recombinant a été obtenu par co-transfection de cellules d'insecte Sf9 avec un bacmide linéaire et un vecteur pAC8 portant la séquence codant TRBP fusionnée à une étiquette 6-histidines. Le virus

recombinant a ensuite été amplifié dans des boîtes T175 avant l'infection des cellules d'insecte (voir « Matériel et Méthodes »). D'autres mises au point ont ensuite été nécessaires pour purifier la protéine, notamment pour les étapes de lyse cellulaire et de fixation. En effet, pour les premiers essais de purification, nous avons utilisé le même protocole que pour une production bactérienne, avec notamment une étape de lyse cellulaire par sonication, suivie d'une purification par HPLC. Malheureusement, ce protocole s'est avéré inadapté pour les cellules Sf9, notamment l'utilisation d'ultrasons qui a eu pour conséquence la dégradation de TRBP, ainsi qu'une trop grande viscosité du lysat, incompatible avec l'HPLC. Ces points ont pu être améliorés avec succès par l'utilisation, dans un second temps, d'un homogénéisateur de Dounce, permettant une lyse plus « douce » des cellules, et d'un tampon riche en glycérol (5%) afin de protéger les protéines lors de la lyse. Enfin, TRBP a été purifiée sur de la résine TALON™ et non en HPLC.

2. Production et purification de la protéine Dicer

Une fois le protocole mis au point pour TRBP, nous l'avons utilisé pour produire Dicer, malheureusement sans succès, la protéine ne s'exprimant pas sous forme soluble. Afin d'améliorer sa solubilité, nous avons modifié le vecteur pAC8 afin d'y ajouter une étiquette MBP, comme pour le vecteur utilisé par Wilson *et al.* Pour cela, nous avons inséré par recombinaison homologue la séquence codant la protéine MBP (« Maltose Binding Protein ») au sein du vecteur pAC8, en aval de l'étiquette 6-histidines mais en amont de la séquence codant le site de coupure à la PreScission, cette séquence ayant été amplifiée par PCR à partir d'un vecteur généreusement fourni par le Dr Benjamin Chagot. L'utilisation de ce vecteur a permis de produire la protéine Dicer mais avec un rendement extrêmement faible. Notre protocole a alors été adapté une dernière fois à partir du protocole publié par De et MacRae en 2011 pour la production et la purification des protéines Dicer, TRBP et AGO2. Ce protocole diffère du nôtre sur 2 principaux points, au niveau de l'amplification du virus et au niveau de l'infection des cellules Sf9. Ainsi, le virus est amplifié non pas 1 fois mais 3 fois, en plaques 6 puits (au lieu des flasques T175). Enfin, un volume plus faible de virus est utilisé pour infecter les cellules. Combinées, ces modifications nous ont permis de produire puis de purifier Dicer. Toutefois, les quantités restent assez faibles et bien qu'elles soient suffisantes pour réaliser des tests de production d'isomiR, elles ne le sont pas pour espérer réaliser une étape de purification supplémentaire par chromatographie d'exclusion stérique.

C. Expérience de production d'isomiR *in vitro*

Une fois le pré-microARN 30E, les protéines Dicer, TRBP et RPAP3 produits et purifiés, plusieurs expériences de clivage du pré-microARN par la RNase Dicer ont été réalisées. Ainsi, un témoin positif de maturation du pré-microARN en microARN a été réalisé en incubant le pré-microARN 30E marqué avec les protéines Dicer et TRBP (ratio 1:1). Un témoin positif de génération d'isomiR a également été réalisé en incubant le pré-microARN uniquement avec Dicer (Wilson *et al.*, 2015). Pour finir, différents tests ont été réalisés en incubant avec le pré-microARN les protéines Dicer, TRBP et RPAP3 (ratio 1:1:1, 1:1:2 et 1:1:10, respectivement). L'utilisation de RPAP3 à différentes concentrations devrait permettre d'augmenter la génération d'isomiR par Dicer en séquestrant TRBP. L'ensemble des produits de ces expériences, ainsi qu'une aliquote du pré-microARN 30E servant de témoin interne ont alors été fractionnés sur un gel dénaturant acrylamide-urée puis visualisés à l'aide d'un phosphimager.

L'ensemble de nos essais se sont révélés infructueux. En effet, nous n'avons pas été capables de générer/d'observer d'isomiR. Toutefois, aucun de nos témoins n'a pu être validé, que se soit pour la maturation du pré-microARN 30E en microARN, que pour la génération d'isomiR par Dicer. Ainsi, seules des bandes de très forte intensité mais de tailles inférieures à 10 nt ont pu être visualisées sur le gel, y compris pour le témoin interne, ce qui traduit un problème de dégradation non spécifique de l'ARN. Cette dégradation pourrait s'expliquer par la présence de RNase dans les échantillons utilisés après les étapes de purification du pré-microARN, ou encore par la dégradation de l'ARN suite aux différents traitements réalisés pour le purifier et le marquer. Plusieurs tests seront nécessaires pour identifier l'origine du problème et améliorer le protocole pour réaliser à nouveau cette expérience.

Bien que ces expériences ne nous aient pas permis de tester notre hypothèse sur le contrôle de la production d'isomiR suite à la formation d'un complexe entre TRBP et RPAP3, elles nous ont permis de réaliser les premières mises au point nécessaires à la production et à la purification des protéines Dicer, TRBP et RPAP3. Il devrait être possible, en modifiant légèrement les protocoles utilisés, d'obtenir ces différentes protéines en quantité et qualité compatibles avec des études biochimiques et structurales. Notamment, il pourrait être intéressant de réaliser des expériences d'ITC (« Isothermal Titration Calorimetry ») entre ces différentes protéines afin d'obtenir les constantes de dissociation des différents complexes (Dicer/TRBP et TRBP/RPAP3), ce qui

permettrait, entre autre, de valider ou non la possibilité pour RPAP3 de déplacer l'équilibre du complexe Dicer/TRBP vers un complexe TRBP/PAP3 (dans le cas où la constante de dissociation du complexe TRBP/RPAP3 serait plus faible que celle du complexe Dicer/TRBP). Enfin, des tests de cristallisation entre les protéines TRBP et RPAP3 pourraient être envisagés. Afin d'améliorer les protocoles de purification, plusieurs pistes sont envisageables. Ainsi, il faudrait probablement ajouter des étapes de purification supplémentaires pour RPAP3. Concernant Dicer, une piste d'amélioration possible serait de la produire simultanément TRBP en infectant les cellules Sf9 par les deux baculovirus, afin de favoriser la stabilisation de Dicer par la formation d'un complexe Dicer/TRBP.

IV. Rôle dans la biogenèse d'ARNnc ?

En complément de nos expériences avec les variants E328A et R354E de TRBP et D161 de RPAP3, nous avons souhaité analyser de manière plus globale les effets liés à l'absence de TRBP ou de RPAP3 sur la maturation de l'ensemble des ARNnc. Pour cela, nous avons utilisé des siARN dirigés contre l'une ou l'autre de ces protéines puis extrait les ARN qui ont alors été analysés sur des micropuces à ARN par la plateforme du laboratoire luxembourgeois LIH dirigée par Laurent Vallar.

Pour RPAP3, nous avons utilisé avec succès un siARN commercial (On Target de ThermoFisher Scientific) comprenant un mélange de 4 siARN différents. Comme indiqué dans la partie « Matériel et Méthodes », un triplicat biologique de cellules HEK a été transfecté deux fois par ces siARN sur une période de 96 h (transfection à J0 et à J+48 h), puis les ARN totaux de ces cellules ont été extraits et purifiés sur colonne avant d'être envoyés au Luxembourg pour leur analyse. Un témoin négatif a été réalisé en transfectant le même lot de cellule HEK par un siARN dirigé contre l'ARNm codant la luciférase. Le niveau d'expression de RPAP3 après transfection a été vérifié par western blot et quantifié par rapport à l'actine, qui sert de témoin interne, et par rapport au signal obtenu pour le témoin luciférase. Les résultats indiquent que suite au traitement par les siARN, le taux d'expression de RPAP3 est inhibé d'environ 87 % pour l'ensemble du triplicat biologique, avec un écart-type de 6,4 %. Bien qu'une légère variation du taux d'expression de RPAP3 soit visible, nous avons tout de même envoyé ces ARN (5 µl à 100 ng/µl) pour analyse au LIH.

L'inhibition de l'expression de TRBP par des siARN fut plus difficile à mettre au point. En effet, nos premiers essais effectués à l'aide du même type de kit que pour RPAP3 n'ont eu aucun effet sur

l'expression de TRBP. Différentes conditions ont été testées, notamment en augmentant la quantité de siARN utilisée, sans plus de succès. Nous avons alors opté pour l'utilisation de deux oligonucléotides complémentaires synthétisés à façon ciblant l'ARNm de TRBP, mais là encore les résultats ne furent pas concluants. Ces différents échecs nous ont amené à suspecter un problème au niveau de l'anticorps utilisé pour détecter TRBP. En regardant les différentes publications utilisant le même anticorps dirigé contre TRBP que nous utilisons, nous nous sommes aperçu qu'il était bien employé pour des expériences d'immunofluorescence (Gagnon *et al.*, 2014), mais pas pour les western-blot, qui dans la plupart des cas utilisent un anticorps non-commercial. Nous avons donc vérifié l'effet de nos siARN par RT-qPCR en utilisant une gamme de dilution d'ADNc provenant de cellules HEK transfectées par le siARN luciférase pour notre quantification relative. Les résultats de qPCR ont confirmé notre hypothèse puisque nous avons observé une diminution du taux d'ARNm de TRBP d'un peu plus de 60 % pour les triplicats biologiques, avec toutefois un écart-type de 7 %. De manière intéressante, il semble que ce taux soit fortement réduit lorsque les cellules sont arrêtées aux alentours de 100 % de confluence. Le protocole de transfection a donc été modifié par rapport à RPAP3 avec une incubation de seulement 52h, mais toujours avec deux transfactions (J0 et J+46 h).

Bien que le taux d'inhibition soit plus faible que pour RPAP3 et qu'il ne reflète non pas le taux de protéine mais celui d'ARNm, les échantillons d'ARN totaux ont été analysés au LIH. Malheureusement, nous n'avons pas pu obtenir de résultats intéressants pour ces différents échantillons (siTRBP, siRPAP3 et contrôle siLuciférase). En effet, d'une part les ARN des échantillons siTRBP se sont dégradés durant le transport, et n'ont donc pas été analysés, et d'autre part, les résultats obtenus pour les siRPAP3 ont montré une forte hétérogénéité entre les différents triplicats biologiques, ne permettant d'observer aucune espèce dont l'expression puisse être considérée comme significativement altérée. Ces expériences sont actuellement reproduites pour de nouvelles analyses. Les fortes disparités observées entre les triplicats de siRPAP3 pouvant s'expliquer par des effets induits suite à une trop forte confluence des cellules, nous avons ainsi décidé d'utiliser le même protocole que pour TRBP, soit une transfection à J0, une seconde à J+46 h et un arrêt 6 h plus tard.

Conclusions et Perspectives

Longtemps sous-estimées, les fonctions des ARN non-traduits en protéines (ARNnc) se révèlent aujourd'hui aussi nombreuses et diverses qu'importantes pour le bon fonctionnement des cellules. Les différentes familles d'ARNnc sont ainsi impliquées dans l'épissage des pré-ARNm (snARN), la traduction des ARNm en protéines (ARNr, ARNt), la maturation des ARNr (snoARN), le contrôle de la traduction (microARN) ou encore dans le développement cellulaire (lncARNnc). Ces dernières années, les progrès en matière de séquençage à haut-débit ont révélé l'existence de familles d'ARNnc supplémentaires, dont les fonctions sont encore souvent inconnues ou peu documentées (tiARN, spliARN, PASR, TASR, ...). En outre plusieurs études ont montré qu'il pouvait exister des liens entre des familles d'ARNnc distinctes, par exemple au niveau de leurs voies de biogenèse. Ainsi, grâce à des approches d'immunoprécipitation et de séquençage à haut-débit, il a été montré que certains snoARN pouvaient être à la fois impliqués dans leur fonction canonique, à savoir la maturation des ARNr, et servir également de précurseurs pour la production d'autres ARN plus petits, les sdARN (pour « snoRNA-derived RNA » – ARN dérivés de snoARN) (Ender *et al.*, 2008). Parmi les sdARN, certains semblent même pouvoir être maturés en ARN de type microARN, retrouvés parfois dans la littérature sous le terme de sno-miR. Bien que les voies de maturation canonique des microRNP et des snoRNP soient aujourd'hui relativement bien comprises, il existe très peu d'informations concernant la maturation de microARN à partir de snoARN précurseurs.

Ce manque d'informations nous a conduits à réaliser un crible double-hybride chez la levure *S. cerevisiae* entre différentes protéines impliquées dans la biogenèse des snoRNP et des microRNP, dans le but de mettre en évidence des liens possibles entre ces deux voies de biogenèse. Ce crible a permis d'identifier une nouvelle interaction entre les protéines TRBP et RPAP3. Ces protéines sont impliquées respectivement dans la maturation du pré-microARN par la RNase de type III Dicer, et dans le complexe hR2TP qui intervient notamment dans l'assemblage des snoRNP. Cette nouvelle observation a ainsi soulevé de nombreuses questions, notamment sur l'implication possible de TRBP ou de RPAP3 dans de nouveaux processus de maturation que ceux pour lesquels ces protéines étaient connues, voire dans le processus de maturation des sdARN ou des sno-miR (**Figure 117**).

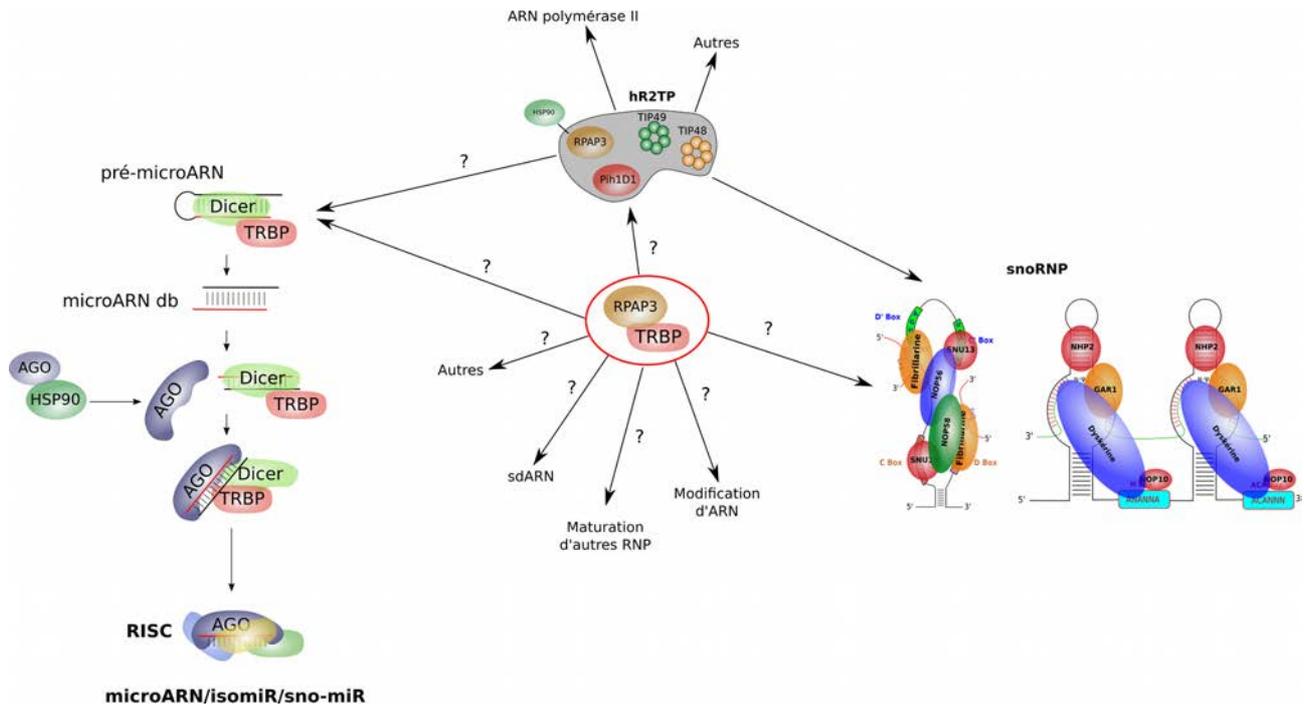


Figure 117: Rôle possible du complexe TRBP/RPAP3 dans la biogenèse des microARN, snoARN ou sno-miR.

Les travaux de mon doctorat ont permis de mettre en évidence une nouvelle interaction entre les protéines TRBP et RPAP3, impliquées respectivement dans la maturation du pré-microARN et dans le complexe hR2TP.

Potentiellement, cette interaction peut indiquer un rôle de RPAP3 et/ou du complexe hR2TP dans la maturation des microARN, voir des sno-miR ou encore indiquer un nouveau rôle pour TRBP dans l'assemblage des snoRNP voir dans la maturation et/ou la modification d'autres particules RNP putatives.

En collaboration avec l'équipe du Dr Xavier Manival et avec le Dr Marc Quinternet de la Fédération de Recherche FR3209 Bioingénierie moléculaire, cellulaire et thérapeutique, nous avons, au cours de mon doctorat, décidé d'une part de caractériser au niveau structural l'interaction entre les protéines TRBP et RPAP3, et d'autre part, commencé à mettre en place les outils nécessaires à l'étude fonctionnelle du complexe qui est toujours en cours.

I. TRBP et RPAP3 forment un complexe stable *in vitro* et *in vivo*

Grâce à des approches de biochimie mais également d'imagerie cellulaire, nous avons confirmé l'interaction directe entre TRBP et RPAP3 *in vitro* et la présence d'un complexe formé au moins de TRBP et de RPAP3. Les images obtenues suite aux expériences de Duolink™ montrent que le complexe TRBP/RPAP3 est principalement observé au niveau du noyau, mais également dans le cytoplasme. Bien qu'il puisse être intéressant de confirmer ces observations, notamment par la réalisation d'immunoprécipitations sur des extraits cytoplasmiques et nucléaires, ces données sont en accord avec la littérature. En effet, la protéine RPAP3, tout comme les autres protéines du complexe hR2TP, ont une localisation à la fois nucléaire et cytoplasmique (Boulon *et al.*, 2008). Parallèlement, des études récentes d'immunofluorescence, couplées à l'utilisation de la microscopie confocale, ont montré que les protéines impliquées dans la biogenèse des microARN étaient également présentes dans le noyau, à l'exception de la protéine chaperonne HSP90 (Gagnon *et al.*, 2014). La présence dans ces deux compartiments cellulaires d'un complexe TRBP/RPAP3 pourrait s'expliquer à la fois par la formation du complexe dans l'un des compartiments qui serait ensuite importé ou exporté dans le second, ou par la formation indépendante des complexes cytoplasmiques et nucléaires. De manière intéressante, la présence du complexe TRBP/RPAP3 dans les deux compartiments suggère en outre plusieurs fonctions putatives, selon sa localisation. Il est également possible que le complexe ne soit pas uniquement constitué de ces deux protéines et que sa composition, et donc probablement sa fonction, diffèrent suivant la localisation cellulaire du complexe. Il pourrait alors être intéressant de réaliser des fractionnements d'extraits cytoplasmiques et nucléaires sur gradient de glycérol pour déterminer la composition des complexes TRBP/RPAP3 cytoplasmiques et nucléaires, couplé à des immunoprécipitations suivies d'analyses par western-blot ou spectrométrie de masse. Alternativement, l'utilisation d'une lignée cellulaire exprimant de manière constitutive les protéines TRBP ou RPAP3 fusionnées à une étiquette nous permettrait, par immunoprécipitation puis analyse en spectrométrie de masse, d'identifier les partenaires potentiels du complexe TRBP/RPAP3.

II. L'étude structurale du complexe TRBP/RPAP3 a fourni de précieuses informations fonctionnelles

La protéine RPAP3 contient deux domaines TPR (« Tetratricopeptide Repeat ») qui sont impliqués dans son interaction avec la protéine chaperonne HSP90 (Pal *et al.*, 2014). Ces domaines TPR sont composés de 3 motifs TPR comprenant chacun deux hélices α reliées par une boucle. La protéine TRBP est quant à elle composée de trois domaines DRBM (« Double-stranded RNA-Binding Motif »), dont les deux premiers sont impliqués dans l'interaction avec l'ARN, alors que le dernier est impliqué dans des interactions protéines-protéines, notamment avec la RNase Dicer (Wilson *et al.*, 2015). Grâce à l'utilisation conjointe d'approches de biologie moléculaire et de biochimie, nous avons pu montrer que le premier domaine TPR de RPAP3 et le dernier domaine DRBM de TRBP étaient impliqués dans la formation du complexe entre ces deux protéines, soulevant alors plusieurs questions.

A. L'interaction entre TRBP et RPAP3 est fortement spécifique

Les structures du domaine DRBM3 de TRBP et du premier domaine TPR de RPAP3 (TPR n°1) sont respectivement conservées avec le domaine DRBM3 de PACT et le second domaine TPR de RPAP3 (TPR n°2). Toutefois, nos résultats indiquent d'une part que PACT n'interagit pas avec RPAP3 et d'autre part que le domaine TPR n°2 de RPAP3 n'est pas impliqué dans l'interaction avec TRBP. Les données obtenues suite à la réalisation de nombreux variants, sur la base notamment d'expériences RMN, suggèrent que de légères différences au niveau des domaines TPR n°1 et 2 pourraient permettre d'expliquer l'absence d'interaction du domaine TPR n°2 avec TRBP. En effet, nous avons montré que les substitutions de l'acide aspartique 161 ou de la valine 185 du domaine TPR n°1 étaient suffisantes pour empêcher la formation du complexe avec le domaine DRBM3 de TRBP, or l'alignement réalisé entre les deux domaines TPR de RPAP3 montre que ces résidus ne sont pas conservés au niveau du second domaine TPR.

En revanche, ces mutations ne nous ont pas permis d'expliquer l'absence d'interaction entre RPAP3 et PACT alors que les prédictions de structures secondaires pour le domaine DRBM3 de PACT montrent un repliement identique au domaine DRBM3 de TRBP. Il serait donc intéressant de réaliser des mutations plus larges afin d'identifier les éléments structuraux qui pourraient expliquer l'absence d'interaction entre PACT et RPAP3. À plus long terme, la résolution de la structure de PACT, ou de son domaine DRBM3, qui n'ont pas encore été réalisées, par cristallographie ou par RMN, permettrait de déterminer plus précisément ces éléments. Une autre explication possible

pourrait résider dans le fait que, malgré leur homologie, les protéines PACT et TRBP ne se comportent pas toujours de la même manière. Ainsi, il a été montré que TRBP et PACT pouvaient former des homodimères. Toutefois, concernant TRBP, l'homodimérisation n'implique que les deux premiers domaines DRBM, alors que pour PACT, les trois domaines DRBM vont établir des contacts entre eux lors de l'homodimérisation (Laraki *et al.*, 2008), ce qui suggère que les légères différences existant entre les deux domaines DRBM3 de TRBP et de PACT pourraient modifier significativement leurs affinités relatives. Les expériences de RMN et de spectrométrie de masse native que nous avons réalisées sur le domaine DRBM3 de PACT vont également dans ce sens. En effet, nous avons montré qu'en solution, ce domaine de PACT était présent à plus de 80 % sous la forme d'un dimère. Dans l'optique où la structure 3D du domaine DRBM3 de PACT serait résolue, il pourrait être intéressant de muter les résidus impliqués dans la formation de l'homodimère afin de vérifier si leur absence modifie le repliement de ce domaine et permet *in fine* l'interaction avec RPAP3.

B. TRBP interagit-elle avec la protéine RPAP3 libre ou avec le complexe hR2TP ?

Nous avons ensuite voulu savoir si l'interaction entre TRBP et RPAP3 pouvait bloquer l'association de l'une ou l'autre de ces protéines avec leurs partenaires habituels, notamment les protéines HSP90 et PIH1D1 pour RPAP3, et Dicer pour TRBP. L'utilisation de la RMN pour réaliser le titrage du domaine DRBM3 de TRBP en complexe avec le TPR n°1 de RPAP3 par le peptide MEEVD d'HSP90 nous a permis de montrer que l'interaction entre HSP90 et RPAP3 était toujours possible en présence de TRBP. En revanche, nous avons montré grâce à des approches biochimiques (co-expression notamment) et structurales que les protéines Dicer et RPAP3 partageaient probablement la même interface d'interaction au niveau de TRBP et qu'en conséquence les formations des complexes TRBP/Dicer et TRBP/RPAP3 étaient mutuellement exclusives. Nous n'avons pas pu confirmer l'existence d'un complexe formé des 3 protéines RPAP3, PIH1D1 et TRBP simultanément *in vitro*, toutefois, l'existence d'un tel complexe reste possible d'après nos observations et d'après la littérature (Yoshida *et al.*, 2013). En effet, les travaux de Yoshida *et al.* ont montré qu'*in vivo* seule l'isoforme I de RPAP3, qui comprend un domaine supplémentaire entre le second domaine TPR et le domaine C-terminal par rapport à l'isoforme II, pouvait interagir avec PIH1D1. Cette interaction étant suffisamment éloignée du premier TPR, il est donc fort probable qu'un complexe TRBP/RPAP3/PIH1D1 puisse exister. Enfin, récemment, l'équipe du Dr Manival a isolé le domaine de PIH1D1 impliqué dans l'interaction avec RPAP3. Ce domaine pourrait ainsi

être utilisé pour réaliser les expériences de co-expression avec les protéines TRBP et RPAP3 en lieu et place de la protéine entière qui n'est pas stable *in vitro*. La confirmation de l'existence d'un tel complexe est clairement nécessaire pour la poursuite de l'étude fonctionnelle du complexe TRBP/RPAP3. En effet, non seulement la protéine RPAP3 interagit avec PIH1D1, mais surtout, elle forme avec les protéines TIP48 et TIP49 le complexe hR2TP dont le rôle est important pour l'assemblage de diverses particules ribonucléoprotéiques, dont les snoRNP. L'immunoprécipitation du complexe TRBP/RPAP3 étant maintenant au point, il serait donc intéressant de vérifier si l'interaction *in vivo* de TRBP avec RPAP3 n'implique pas également l'ensemble du complexe hR2TP, ce qui pourrait indiquer un nouveau rôle possible pour ce complexe dans la maturation des microARN ou d'autres RNP à identifier. De plus, puisque l'interaction entre TRBP et RPAP3 n'a aucun impact sur l'interaction de cette dernière avec la chaperonne HSP90, et que HSP90 est impliquée *in vivo* dans le recrutement du microARN double-brin par le RLC (Pare *et al.*, 2013), un rôle de RPAP3 dans cette étape de recrutement du RLC pourrait aussi être envisagé, impliquant ou non le reste du complexe hR2TP. En effet, il a été montré chez les mammifères que la protéine HSP90 stabilisait la protéine AGO (Johnston *et al.*, 2010) et qu'elle était nécessaire au changement de conformation de cette dernière, nécessaire au recrutement du microARN double-brin par le complexe RLC (Pare *et al.*, 2013). La protéine RPAP3 étant une co-chaperonne d'HSP90, il est possible d'imaginer que la reconnaissance de la protéine cliente AGO par HSP90 puisse se faire par l'intermédiaire du complexe TRBP/RPAP3 qui interagirait alors d'une part avec AGO et d'autre part avec HSP90 (**Figure 118**). Il serait donc intéressant de tester cette hypothèse, notamment en étudiant la possibilité d'une interaction directe entre AGO et TRBP, voire entre AGO et RPAP3

C. Résolution de la structure 3D de l'interface d'interaction entre RPAP3 et TRBP

Bien que les expériences de RMN nous aient permis d'obtenir de nombreuses informations, notamment sur les résidus respectifs de TRBP et RPAP3 impliqués dans la formation du complexe, nous n'avons malheureusement pas réussi à résoudre par RMN la structure tridimensionnelle de l'interface d'interaction entre les domaines TPR n°1 de RPAP3 et DRBM3 de TRBP. Nous avons opté pour l'utilisation de la RMN, malgré la taille des complexes, puisque nous avons déjà accès à l'attribution complète du premier domaine TPR de RPAP3 qui a été réalisée au laboratoire. Cela nous a permis, entre autre, d'obtenir rapidement une liste de résidus potentiellement impliqués dans l'interaction de RPAP3 avec TRBP.

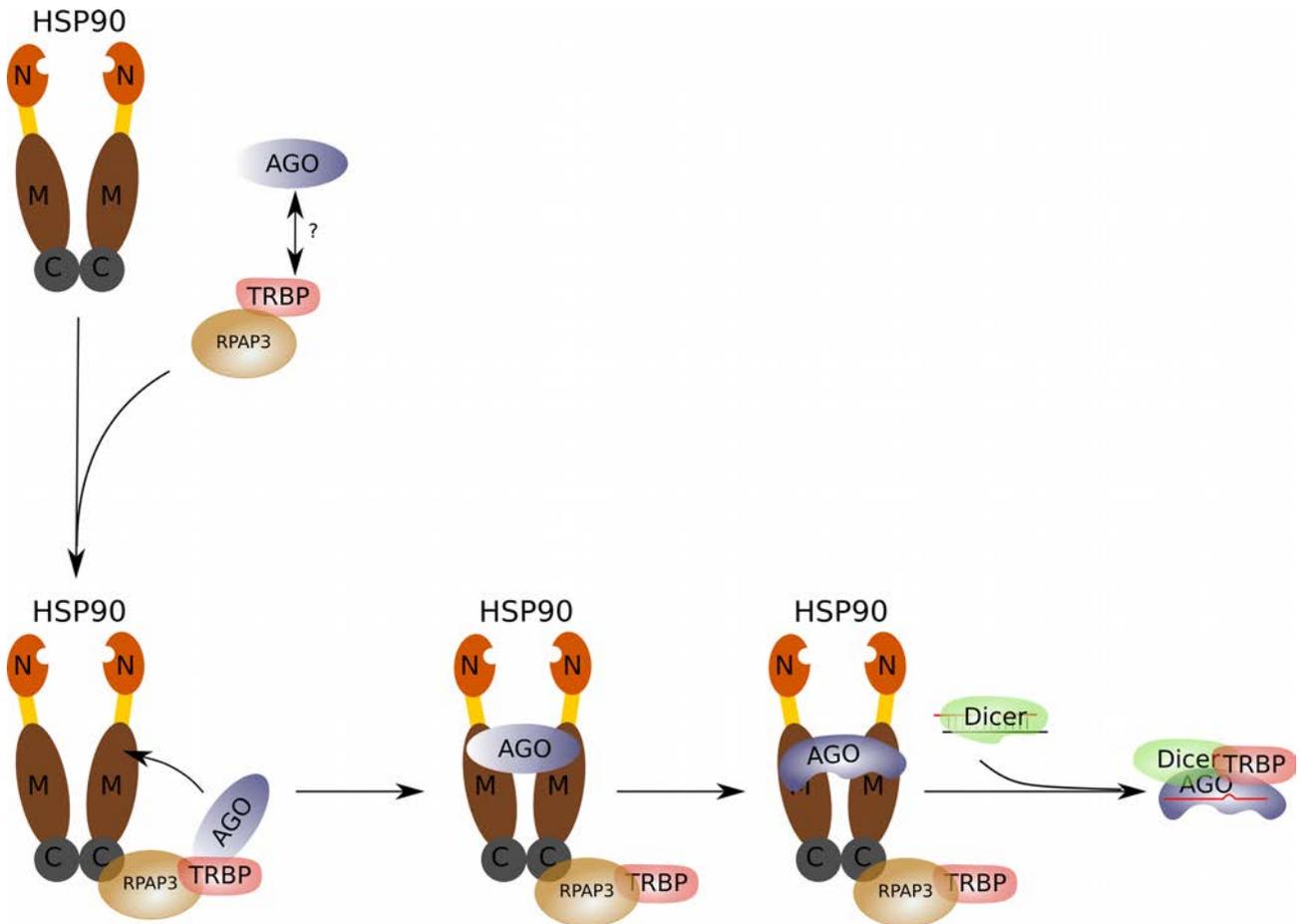


Figure 118: Implication du complexe TRBP/RPAP3 dans le recrutement du microARN par la protéine AGO ?

Il a été montré que la protéine chaperonne HSP90 était nécessaire pour stabiliser la protéine AGO et pour modifier sa structure tridimensionnelle afin de permettre le recrutement du microARN (Pare *et al.*, 2013). RPAP3 étant une co-chaperonne d'HSP90, capable d'interagir avec elle par l'intermédiaire de ses domaines TPR, il est possible que le complexe TRBP/RPAP3 soit impliqué dans la formation d'un complexe entre HSP90 et AGO2 par l'intermédiaire de contact d'une part entre HSP90 et RPAP3 et d'autre part entre TRBP et/ou RPAP3 et AGO.

Puisque cette approche ne nous a pas permis de résoudre la structure du complexe, il serait intéressant de faire rapidement des essais de cristallisation, d'une part sur les domaines 133-255 de RPAP3 et 262-366 de TRBP, mais également, étant donné les résultats des premiers tests de co-expression réalisés à l'aide du robot TECAN™, sur les protéines entières. En parallèle, il pourrait être également intéressant de réaliser des études structurales par SAXS (« Small Angle X-rays Scattering »). En effet, les structures des trois domaines DRBM de TRBP ont été résolues (PDB ID : 3LLH et 2CPN ; Wilson *et al.*, 2015), ainsi que celle des deux domaines TPR de RPAP3 (Pal *et al.*, 2014). Dans le cas où les essais de cristallisation seraient infructueux, des études par SAXS sur la protéine TRBP entière en complexe avec les deux domaines TPR permettraient peut-être d'obtenir un modèle à basse résolution de l'interface d'interaction en combinant les données SAXS et les données structurales des différents domaines de TRBP et de RPAP3.

III. Mise en place d'outils moléculaires pour l'étude fonctionnelle du complexe

Suite à nos études biochimiques et structurales, nous avons émis plusieurs hypothèses quant au rôle du complexe TRBP/RPAP3 *in vivo*. En particulier, la spécificité d'interaction de RPAP3 pour TRBP, et le caractère mutuellement exclusif de l'interaction entre TRBP et Dicer, suggèrent que RPAP3 pourrait influencer indirectement l'activité de Dicer, via la biodisponibilité de TRBP (**Figure 119**). Un tel mécanisme de régulation, s'il était avéré, pourrait avoir des conséquences fonctionnelles importantes, puisqu'il a été montré qu'en absence de TRBP, l'activité de la RNase Dicer était modifiée, entraînant dans certains cas la génération d'isomiR (Wilson *et al.*, 2015). Nous avons voulu tester cette possibilité en reproduisant les expériences réalisées par Wilson *et al.* en présence ou non de la protéine RPAP3. Nous n'avons malheureusement pas réussi à aller au bout de ces expériences, ayant été confrontés à de nombreuses difficultés pour la production du substrat par transcription d'un ribozyme ainsi que pour la production de la protéine Dicer dans les cellules d'insectes. Ce test d'activité, qui sera indispensable par la suite, devra donc faire l'objet de nombreuses optimisations. Toutefois, puisque nous arrivons à produire et purifier les protéines Dicer, TRBP et RPAP3, il serait possible de valider en partie notre hypothèse sans pour autant réaliser une expérience aussi complexe. En effet, dans le cas où nous arriverions à purifier des quantités plus importantes de ces trois protéines, il nous serait possible de réaliser des expériences d'ITC et de définir ainsi les différentes constantes d'affinités entre Dicer et TRBP et entre TRBP et RPAP3.

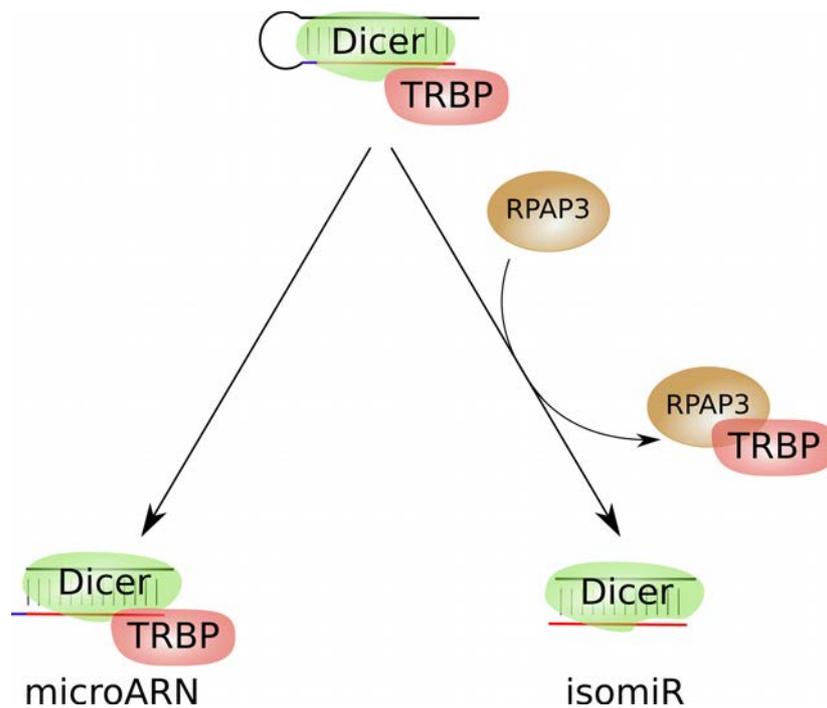


Figure 119: Implication du complexe TRBP/RPAP3 dans le contrôle de la génération d’isomiR par Dicer.

Il a été montré que l’absence de TRBP pouvait entraîner, pour certains microARN, la génération par Dicer d’isomiR. Ces isomiRs correspondent à des microARN dont le site de coupure a été décalé d’un nucléotide par rapport au site canonique (ici en bleu) (Wilson *et al.*, 2015). En fonction des affinités relatives entre les protéines Dicer, TRBP et RPAP3, il est possible que la protéine RPAP3 ait pour rôle de séquestrer TRBP afin de réguler la production d’isomiRs.

Il serait également possible de réaliser ces expériences en présence des différents variants que nous avons identifiés, afin de déterminer si une compétition existe bel et bien entre RPAP3 et Dicer pour TRBP, permettant de conforter l'une de nos hypothèses.

Nos études structurales et biochimiques nous ont permis d'identifier plusieurs acides aminés de TRBP et de RPAP3 dont la mutation abolissait l'interaction entre les deux protéines. Nous avons en outre confirmé *in vitro* que les mutants R354E et D328A de TRBP étaient toujours capables d'interagir avec le domaine 256-595 de Dicer. Ces derniers, ainsi que les mutants V185A et D161A de RPAP3 pourront alors être utilisés afin d'analyser l'effet spécifique de la perte d'interaction entre TRBP et RPAP3 sur la maturation de différents ARNnc (microARN, snoARN et sdARN). Toutefois, pour ces études, il sera nécessaire dans un premier temps de prendre en compte les effets potentiels dus à une perte d'interaction entre TRBP et ses autres partenaires putatifs. En effet, TRBP étant impliquée dans d'autres mécanismes moléculaires, il est fort probable qu'elle interagisse avec d'autres protéines que Dicer, et ce vraisemblablement par l'intermédiaire de son domaine DRBM3. Il serait alors possible que la perte d'interaction entre TRBP et l'un de ses autres partenaires soit en tout ou partie responsable des résultats obtenus suite à l'expression des variants de TRBP. Dans un second temps, il sera nécessaire de s'assurer que les différentes mutations de RPAP3 et de TRBP n'ont pas d'effets significatifs sur le repliement de ces protéines. Des études par dichroïsme circulaire devraient nous permettre de nous en assurer. Toutefois, il n'est pour le moment pas possible de les réaliser sur les protéines RPAP3 et TRBP entières, ni sur le domaine DRBM3 de TRBP. En effet, nos protocoles de purification ne sont actuellement pas suffisamment au point pour pouvoir réaliser ces expériences. En revanche, il est actuellement possible d'étudier par dichroïsme circulaire l'effet des mutations de RPAP3 sur le repliement du premier domaine TPR. L'amélioration des protocoles de production en système baculovirus (pour TRBP) et de purification (pour RPAP3) devrait nous permettre à moyen terme de confirmer l'absence d'effets des différentes mutations sur le repliement des protéines entières. En revanche, il semble exclu d'étudier le domaine DRBM3 de TRBP, qui, d'après nos expérimentations, n'est pas stable *in vitro* lorsqu'il n'interagit pas avec une autre protéine.

Ces variants seront prochainement sur-exprimés dans des cellules HEK293 et l'effet sur la biogenèse des ARNnc sera étudié par séquençage à haut-débit des ARNnc. Nous espérons que les excès respectifs de TRBP ou de RPAP3 exogènes soient suffisants pour permettre de négliger la

contribution des protéines sauvages endogènes. Toutefois, il serait beaucoup plus intéressant de tester à plus ou moins court terme ces effets dans des lignées cellulaires invalidées pour TRBP ou RPAP3. Bien qu'il soit possible d'utiliser des siARN qui ne cibleraient que les protéines endogènes, la réalisation de lignées totalement invalidées pour l'une ou l'autre de ces protéines, notamment par l'utilisation de la technique CRISPR-Cas9, fournirait un puissant outil génétique. En effet, disposer d'une telle lignée permettrait de tester l'effet de n'importe quel variant de RPAP3 ou de TRBP sur la biogenèse de ARNnc, notamment des mutants de l'interface d'interaction ou encore, par exemple, au niveau des « dicarboxylate clamp » des deux domaines TPR de RPAP3. Ces derniers permettraient alors d'étudier en outre le rôle putatif du complexe RPAP3/HSP90 dans la biogenèse des microARN.

Les différentes hypothèses sur les fonctions du complexe TRBP/RPAP3 sont basées sur les résultats expérimentaux que nous avons obtenus par des approches de biologie moléculaire, biochimiques et structurales, avec le postulat que l'interaction entre TRBP et RPAP3 puisse avoir un rôle dans la biogenèse des microARN, des snoARN, des sdARN ou des sno-miR. Toutefois, ces protéines sont également impliquées dans d'autres processus biologiques. Ainsi, TRBP peut inhiber la voie PKR (Park *et al.*, 1994 ; Daher *et al.*, 2009), et des liens entre RPAP3 et la survie cellulaire ont été montrés (Ni *et al.*, 2009 ; Yoshida *et al.*, 2013). L'identification d'autres protéines ou enzymes (RNase, enzymes de modification de l'ARN, ...), associées à TRBP et/ou à RPAP3 pourrait ainsi permettre d'identifier d'autres fonctions pour le complexe TRBP/RPAP3.

Matériel et Méthodes

I. Matériel

A. Les souches de cellules

1. Les souches de bactéries

- Pour l'amplification de plasmides :

***E. coli* DH5 α :**

supE44, Δ lacU169 (ϕ 80 lacZ Δ M15), hsdR17, recA1, gyrA96, thi-1, relA1

- Pour l'amplification de plasmides Gateway® vide :

***E. coli* DH5 α DB3.1 :**

F- gyrA462 endA1 glnV44 Δ (sr1-recA) mcrB mrr hsdS20(r_B^- , m_B^-) ara14 galK2 lacY1 proA2
rpsL20(Sm^r) xyl5 Δ leu mtl1

- Pour la production de protéines recombinantes :

***E. coli* BL21 (DE3) pRAR2 :**

FF⁻ ompT gal dcm lon hsdS_B(r_B^- m_B^-) λ (DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])
pRAR2(cm^R)

2. Les souches de levures

***S. cerevisiae* Y187 (Mat α) :**

gal4 Δ , gal80 Δ , ad2-101, his3-200, leu2-3,112, lys2-801, trp1-901, ura3-52,
URA3::GAL1UASGAL1TATA-LacZ LYS2::GAL1UASHISTATA-HIS3

***S. cerevisiae* CG1945 (Mat α) :**

ura3-52, his3-200, lys2-801, ase2-101, trp1-901, leu2-3,112, gal4-542, gal80-538, cyh-2,
LYS2::GAL1UASGAL1TATA-HIS3, URA3::GAL417mers(X3)-CyC1TATA-LacZ

3. Les lignées de cellules humaines

HeLa : Référence ATCC CCL-2

HEK : Référence ATCC CRL-1573

4. Les lignées de cellules d'insectes

Les cellules Sf9 ont été fournies par la plateforme d'expression en baculovirus de l'IGBMC de Strasbourg.

B. Les oligonucléotides

Les principaux oligonucléotides utilisés sont décrits dans le **Tableau I**.

N°	Séquence	Description
9092	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA ACTTCAGCAAATAAAGCAATCG	Clonage de RPAP3 (1-132) + séquence AttB1 (5')
9093	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CACTTTTGTGAATCTACATGAATC	Clonage de RPAP3 (1-132) + séquence AttB2 et STOP (3')
9094	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA GCTCTTGTTTTAAAAGAAAAGG	Clonage de RPAP3 (133-231) + séquence AttB1 (5')
9095	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CATTCTAGTTCTAATACTCTTTC	Clonage de RPAP3 (133-231) + séquence AttB2 et STOP (3')
9096	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA CCAAATAACTTTGAAGCAACAAATG	Clonage de RPAP3 (232-280) + séquence AttB1 (5')
9097	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CAATTCTGTTGTGCTTCAATTTG	Clonage de RPAP3 (232-280) + séquence AttB2 et STOP (3')
9098	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA AAGCAGCAGGCCATTCAGAG	Clonage de RPAP3 (281-380) + séquence AttB1 (5')
9099	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CATTCCAGAAGTAAACAGTTTC	Clonage de RPAP3 (281-380) + séquence AttB2 et STOP (3')
9100	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CACACATCAATAGTCTGTATCAAATTAC	Clonage de RPAP3 isoforme II (281-415) + séquence AttB2 et STOP (3')
9101	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA CCAGATAGCACTACTGCTGC	Clonage de RPAP3 isoforme II (417-631) + séquence AttB1 (5')
9102	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CAACCACCGTATCTTTTCTTGAG	Clonage de RPAP3 isoforme II (417-631) + séquence AttB2 et STOP (3')
9192	GCTAAGGAAGCTAAAATGATTGAACAAGATG GATTGC	Amplification du gène de résistance à la kanamycine + en 5' les séquences de recombinaison homologue avec le pnYC (5')
9193	ACTGCCTTAAAAAATCAGAAGAACTCGTCA AGAAG	Amplification du gène de résistance à la kanamycine + en 5' les séquences de recombinaison homologue avec le pnYC (3')
9194	TTTTTTTAAGGCAGTTATTGGTG	Amplification du plasmide pnYC sans amplifier le gène de résistance au chloramphénicol (5')

9195	TTTAGCTTCCTTAGCTCCTGA	Amplification du plasmide pnYC sans amplifier le gène de résistance au chloramphénicol (3')
9200	GGCTTTTAGTAAGCCGGGTCCAGATCTTTACGCCCC	Mutagenèse d'un site BamHI de la cassette gateway du pDEST17 (5')
9201	GGGGCGTAAAGATCTGGACCCGGCTTACTAA AAGCC	Mutagenèse d'un site BamHI de la cassette gateway du pDEST17 (3')
9234	AGCAGCCGATTGTCTGTTGTGCCC	Séquençage du gène de résistance Kanamycine du pnYK (Reverse)
9235	TGCTGAAGAGCTTGGCGGCGAATGGG	Séquençage du gène de résistance Kanamycine du pnYK (Forward)
9284	CTAGTTATTGCTCAGCGG	Séquençage au niveau du terminateur T7 (Reverse)
9653	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA AGTGAAGAGGAGCAAGGCTCCGGC	Clonage de TRBP isoforme a + séquence AttB1 (5')
9654	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CACTTGCTGCCTGCCATGATC	Clonage de TRBP isoforme a + séquence AttB2 et STOP (3')
9655	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA AGTATAGAGCAAATGCTGGCCGCCAAC	Clonage de TRBP a (18-99) + séquence AttB1 (5')
9656	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CAGCTCCCCCTTTGAGGTGTTTGAGG	Clonage de TRBP a (18-99) + séquence AttB2 et STOP (3')
9657	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA GAGTGCAACCCCGTT GGTGCTCTGC	Clonage de TRBP a (157-228) + séquence AttB1 (5')
9658	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CACACCGTGTGCACTCGAAGCAGC	Clonage de TRBP a (157-228) + séquence AttB2 et STOP (3')
9659	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA GGTGACCTGGGATTCTCTACG	Clonage de TRBP a (262-336) + séquence AttB1 (5')
9710	AAGGAGATATACATATGACAAGTTTGTACAAA AAAGCTG	Amplification de la cassette gateway du pDEST17 modifié + séquence en 5' nécessaire à une recombinaison avec le pnCS digéré par Nde1 (5')
9711	TTAAGCTCGAGCATATGACCACTTTGTACAAG AAAGC	Amplification de la cassette gateway du pDEST17 modifié + séquence en 5' nécessaire à une recombinaison avec le pnCS digéré par Nde1 (3')
9899	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA GCCTGCTGCCGTGTCTCAGClonage de TRBP a (262-336) + séquence AttB1 (5')	Clonage de TRBP a (293-336) + séquence AttB1 (5')
9937	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA GCAAACCTCGTTCCAGCTCGAATC	Clonage de RPAP3 isoforme II (513-631) + séquence AttB1
9938	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA CAGGCCATTTAGAGAAAGATCGGG	Clonage de RPAP3 isoforme II (281-396) + séquence AttB1
9939	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CATTCTTTTTAATTTTGAGAGTTCAG	Clonage de RPAP3 isoforme II (281-396) + séquence AttB2 et STOP (3')

9940	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CAATATGAGTTTTCTTTGGATGC	Clonage de RPAP3 isoforme II (133-255) + séquence AttB2 et STOP (3')
A393	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA ATGGATTGTGGACCATTACTGACAG	Clonage de Dicer (256-595) + séquence AttB1 et un codon START (5')
A394	GGGGACCACTTTGTACAAGAAAGCTGGGTCT CAATCAACCGACTTGGAACAC	Clonage de Dicer (256-595) + séquence AttB1 et un codon START (3')
A465	GCTCTTGTTTTAGCAGAAAAGGGCAATAAATA C	Mutagenèse RPAP3 K137A (5')
A466	GTATTTATTGCCCTTTTCTGCTAAAACAAGAGC	Mutagenèse RPAP3 K137A (3')
A467	GTTTTAAAAGAAAAGGGCGCAAATACTTCA AACAAAGG	Mutagenèse RPAP3 N141A (5')
A468	CCTTGTTTGAAGTATTTTGCGCCCTTTTCTTTT AAAAC	Mutagenèse RPAP3 N141A (3')
A469	GCAATTGACTGCTACGCAAAGGCATGGATGC CG	Mutagenèse RPAP3 T157A (5')
A470	CGGCATCCATGCCTTTTGCGTAGCAGTCAATT GC	Mutagenèse RPAP3 T157A (3')
A471	GCTACACAAAAGGCATGGCAGCCGATCCATAT AATCCC	Mutagenèse RPAP3 D161A (5')
A472	GGGATTATATGGATCGGCTGCCATGCCTTTTGT GTAGC	Mutagenèse RPAP3 D161A (3')
A473	GGCATGGATGCCGCACCATATAATCCCGTG	Mutagenèse RPAP3 D163A (5')
A474	CACGGGATTATATGGTGCGGCATCCATGCC	Mutagenèse RPAP3 D163A (3')
A475	GGCATGGATGCCGATCCAGCAAATCCCGTGTT G	Mutagenèse RPAP3 T165A (5')
A476	CAACACGGGATTTGCTGGATCGGCATCCATGC C	Mutagenèse RPAP3 T165A (3')
A477	GCTGTTGCTGAGTCTGCATGTAATTTAGCAG	Mutagenèse RPAP3 D189A (5')
A478	CTGCTAAATTACATGCAGACTCAGCAACAGC	Mutagenèse RPAP3 D189A (3')
A479	CCCGTGTTGCCAACGGCAAGAGCGTCAGC	Mutagenèse RPAP3 N172A (5')
A480	GCTGACGCTCTTGCCGTTGGCAACACGGG	Mutagenèse RPAP3 N172A (3')
A481	GTTTTAGCAGAAAAGGGCGCAAATACTTCA AACAAAGG	Mutagenèse RPAP3 N141A et K137A (5')
A482	CCTTGTTTGAAGTATTTTGCGCCCTTTTCTGCT AAAAC	Mutagenèse RPAP3 N141A et K137A (3')
A483	GCTACACAAAAGGCATGGCAGCCGCACCATAT AATCCCGTG	Mutagenèse RPAP3 D161A et D163A (5')
A484	CACGGGATTATATGGTGCGGCTGCCATGCCTTT TGTGTAGC	Mutagenèse RPAP3 D161A et D163A (3')
A485	GCAATTGACTGCTACGCAAAGGCATGGCAG CCGATCCATATAATCCC	Mutagenèse RPAP3 T157A et D161A (5')

A486	GGGATTATATGGATCGGCTGCCATGCCTTTTGC GTAGCAGTCAATTGC	Mutagenèse RPAP3 T157A et D161A (3')
A691	CGTGTCTCAGTGCCTCTCTGAGGAGC	Mutagenèse TRBP E300A (5')
A692	GCCTTTCACGTGCGCATAACGCAGATATTGAGGA GC	Mutagenèse TRBP S310A et L312A (5')
A693	GTGGACTCTGCGCATGCGCAGTGGCACTGTCC ACCCAGCC	Mutagenèse TRBP E324A, L326A et E328A (5')
A694	GCCACTGTGTGTGCAGGCGCAGCAACCACCA GGG	Mutagenèse TRBP H338A et S340A (5')
A695	CGCCGTGCCCTGGCATACTCAAGATCATGGC	Mutagenèse TRBP Q357A (5')
A696	GCTCCTCAGAGAGTGCCTGAGGACACG	Mutagenèse TRBP E300A (3')
A697	GCTCCTCAATATCTGCGTATGCGACGTGAAAG GC	Mutagenèse TRBP S310A et L312A (3')
A698	GGCTGGGTGGACAGTGCCACTGCGCATGCGC AGAGTCCAC	Mutagenèse TRBP E324A, L326A et E328A (3')
A699	CCCTGGTGGTTGCTGCGCCTGCACACACAGT GGC	Mutagenèse TRBP H338A et S340A (3')
A700	GCCATGATCTTGAGGTATGCCAGGGCACGGCG	Mutagenèse TRBP Q357A (3')
A739	GCTCTTGTTTTAGCAGAAAAGGGCGCAAATA C	Mutagenèse RPAP3 K137A et N141A (5')
A740	TATTTTGCGCCCTTTTCTGCTAAAACAAGAGC	Mutagenèse RPAP3 K137A et N141A (3')
A829	GTGGACTCTGCGCATGCCTGGTGG	Mutagenèse TRBP Q324A (5')
A830	CCACCAGGCATGCGCAGAGTCCAC	Mutagenèse TRBP Q324A (3')
A831	GTGGACTCTGCCAGTGCAGTGGAACTGTC CACCCAGCC	Mutagenèse TRBP L326A (5')
A832	GGCTGGGTGGACAGTTCCACTGCGCACTGGC AGAGTCCAC	Mutagenèse TRBP L326A (3')
A833	GTGGACTCTGCCAGTGCCTGGTGGCACTGTCC ACCCAGCC	Mutagenèse TRBP E328A (5')
A834	GGCTGGGTGGACAGTGCCACCAGGCACTGGC AGAGTCCAC	Mutagenèse TRBP E328A (3')
A835	GCCACTGTGTGTGCAGGCTCTGCAACCACCA GGG	Mutagenèse TRBP H338A (5')
A836	CCCTGGTGGTTGCAGAGCCTGCACACACAGT GGC	Mutagenèse TRBP H338A (3')
A837	GCCACTGTGTGTCATGGCGCAGCAACCACCA GGG	Mutagenèse TRBP S340A (5')
A838	CCCTGGTGGTTGCTGCGCCATGACACACAGTG GC	Mutagenèse TRBP S340A (3')
A839	CCCTCCGCAGTGCATCCCTGGGCTCCC	Mutagenèse TRBP C282A (5')
A840	GGGAGCCCAGGGATGCACTGCGGAGGG	Mutagenèse TRBP C282A (3')

A841	CCTCCGCAGTTGCGCACTGGGCTCCCTGG	Mutagenèse TRBP S283A (5')
A842	CCAGGGAGCCCAGTGCGCAACTGCGGAGG	Mutagenèse TRBP S283A (3')
A843	GGATATTGAGGAGCTGGCACTGAGTGGACTCTGC	Mutagenèse TRBP S318A (5')
A844	GCAGAGTCCACTCAGTGCCAGCTCCTCAATATCC	Mutagenèse TRBP S318A (3')
A847	CCTGGTGGAACTGGCAACCCAGCCGGCC	Mutagenèse TRBP S330A (5')
A848	GGCCGGCTGGGTTGCCAGTTCACCAGG	Mutagenèse TRBP S330A (3')
A849	GGTGGAACTGTCCGCACAGCCGGCCACTGTG	Mutagenèse TRBP T331A (5')
A850	CACAGTGGCCGGCTGTGCGGACAGTTCACC	Mutagenèse TRBP T331A (3')
A851	GGAAGTGTCCACCGCACCGGCCACTGTG	Mutagenèse TRBP Q332A (5')
A852	CACAGTGGCCGGTGCAGTGGACAGTTC	Mutagenèse TRBP Q332A (3')
A853	CCCAGCCGGCCGCAGTGTGTCATGGC	Mutagenèse TRBP T335A (5')
A854	GCCATGACACACTGCGGCCGGCTGGG	Mutagenèse TRBP T335A (3')
A855	CCAGCCGGCCACTGCATGTCATGGCTCTGC	Mutagenèse TRBP V336A (5')
A856	GCAGAGCCATGACATGCAGTGGCCGGCTGG	Mutagenèse TRBP V336A (3')
A857	GCTACCTGGATATTGCAGAGCTGAGCCTGAGTGG	Mutagenèse TRBP E315A (5')
A858	CCACTCAGGCTCAGCTCTGCAATATCCAGGTA GC	Mutagenèse TRBP E315A (3')
A859	GCTACCTGGATATTGAGGCACTGAGCCTGAGTGG	Mutagenèse TRBP E316A (5')
A860	CCACTCAGGCTCAGTGCCTCAATATCCAGGTA GC	Mutagenèse TRBP E316A (3')
A861	GGAGCTGAGCCTGGCAGGACTCTGCCAGTGC	Mutagenèse TRBP S320A (5')
A862	GCACTGGCAGAGTCCTGCCAGGCTCAGCTCC	Mutagenèse TRBP S320A (3')
A863	GGAAGTGTCCACCCAGGCAGCCACTGTGTGCATGGC	Mutagenèse TRBP P333A (5')
A864	GCCATGACACACAGTGGCTGCCTGGGTGGACAGTTC	Mutagenèse TRBP P333A (3')
A865	GCAGTACCTCAAGGCAATGGCAGGCAGC	Mutagenèse TRBP I361A (5')
A866	GCTGCCTGCCATTGCCTTGAGGTA CTGC	Mutagenèse TRBP I361A (3')
A867	GCAGTACCTCAAGATCGCAGCAGGCAGCAAG	Mutagenèse TRBP M362A (5')
A868	CTTGCTGCCTGCTGCGATCTTGAGGTA CTGC	Mutagenèse TRBP M362A (3')
A869	GAGGCTGCCCGCGCAGCCCTGCAGTACC	Mutagenèse TRBP R354A (5')
A870	GGTACTGCAGGGCTGCGCGGGCAGCCTC	Mutagenèse TRBP R354A (3')
A871	CGTCAGCTACCTGGATGCAGAGGAGCTGAGC C	Mutagenèse TRBP I314A (5')
A872	GGCTCAGCTCCTCTGCATCCAGGTAGCTGACG	Mutagenèse TRBP I314A (3')

A877	GAGGCTGCCCCGCGAGGCCCTGCAGTACC	Mutagenèse TRBP R354E (5')
A878	GGTACTGCAGGGCCTCGCGGGCAGCCTC	Mutagenèse TRBP R354E (3')
A879	GCATATTTTAGACTGGCAAATTTGCTGTTGC	Mutagenèse RPAP3 K181A (5')
A880	GCAACAGCAAATTTTCCAGTCTAAAATATGC	Mutagenèse RPAP3 K181A (3')
A881	GCATATTTTAGACTGAAAGCATTGCTGTTGC	Mutagenèse RPAP3 K182A (5')
A882	GCAACAGCAAATGCTTTCAGTCTAAAATATGC	Mutagenèse RPAP3 K182A (3')
A883	CTGAAAAAATTTGCTGCAGCTGAGTCTGATTG	Mutagenèse RPAP3 V185A (5')
A884	CAATCAGACTCAGCTGCAGCAAATTTTTTCAG	Mutagenèse RPAP3 V185A (3')
A885	GCTGTTGCTGAGGCAGATTGTAATTTAGC	Mutagenèse RPAP3 S188A (5')
A886	GCTAAATTACAATCTGCCTCAGCAACAGC	Mutagenèse RPAP3 S188A (3')
A887	CAAACAAGGAAAATATGCAGAAGCAATTGAC TGC	Mutagenèse RPAP3 D150A (5')
A888	GCAGTCAATTGCTTCTGCATATTTTCCTTGTTT G	Mutagenèse RPAP3 D150A (3')
A889	GGAAAATATGATGAAGCAGCAGACTGCTACA CAAAGG	Mutagenèse RPAP3 I153A (5')
A890	CCTTTTGTGTAGCAGTCTGCTGCTTCATCATAT TTTCC	Mutagenèse RPAP3 I153A (3')
A891	GGAAAATATGATGAAGCAATTGCATGCTACAC AAAAGG	Mutagenèse RPAP3 D154A (5')
A892	CCTTTTGTGTAGCATGCAATTGCTTCATCATAT TTTCC	Mutagenèse RPAP3 D154A (3')
A893	CGTGTTGCCAACGAACGCAGCGTCAGCATATT TTAG	Mutagenèse RPAP3 R173A (5')
A894	CTAAAATATGCTGACGCTGCGTTCGTTGGCAA CACG	Mutagenèse RPAP3 R173A (3')
A895	GCGTCAGCAGCATTTAGACTGAAAAAATTTGC	Mutagenèse RPAP3 Y177A (5')
A896	GCAAATTTTTTCAGTCTAAATGCTGCTGACGC	Mutagenèse RPAP3 Y177A (3')
A897	GCTGAGTCTGATTGTGCATTAGCAGTTGCC	Mutagenèse RPAP3 N191A (5')
A898	GGCAACTGCTAATGCACAATCAGACTCAGC	Mutagenèse RPAP3 N191A (3')
A899	GCTGAGTCTGATTGTAATGCAGCAGTTGCC	Mutagenèse RPAP3 L192A (5')
A900	GGCAACTGCTGCATTACAATCAGACTCAGC	Mutagenèse RPAP3 L192A (3')
B092	GGATCCGATTTCCCGGGCGGCCGC	Amplification de pAC8-6hist (5')
B093	CATATGGGGCCCCCTGGAACAGAACTTCC	Amplification de pAC8-6hist (3')
B094	CAGGGGCCCCATATGAAAAGCCCTGCTTTGCA AC	Clonage de Dicer dans pAC8 (5') + séquence homologue pAC8
B095	CGGGAAATCGGATCCTCAGCTATTGGGAACCT GA	Clonage de Dicer dans pAC8 (3') + séquence homologue pAC8

B121	GGGGGATCCGGATGTTTACTGATGAGTCCG TGAGGACGAAACGGTACCCGGTACCGTCTGT AAACATCCTTGACTGG	Clonage du pre-miR 30E dans le p2RZ + ribozyme Hammerhead et BamHI (5')
B122	CCCGCTAGCGCTGTAAACATCCGACTGAAAGC TCCTCTCAACACCTTACAGCTTCCAGTCAAGG ATGTTTACAGACGGTACCGGG	Clonage du pre-miR 30E dans le p2RZ + NheI (3')
B424	CGACTCACTATAGGGCCCGACTTCTTTTACGA CTCACTCTAGACACAGGATCCGGATGTTTACA CTG	Ajout d'une séquence entre le promoteur du p2RZ et le ribozyme Hammerhead (5')
B425	CAGTGTAACATCCGGATCCTGTGTCTAGAGT GAGTCGTAAAAGAAGTCGGGCCCTATAGTGA GTCG	Ajout d'une séquence entre le promoteur du p2RZ et le ribozyme Hammerhead (3')
B456	CTGGAAGTTCTGTTCCAGGG	Linéarisation du vecteur pAC8 6hist-Dicer (5')
B457	ACCCGCGGAGTGATGATG	Linéarisation du vecteur pAC8 6hist-Dicer (3')
B458	CATCACTCCGCGGGTATGAAAATCGAAGAAG GTAAAC	Clonage de MBP + séquence homologue du pAC8 (5')
B459	GAACAGAACTTCCAGGTTTTCTCGATCCCGA G	Clonage de MBP + séquence homologue du pAC8 (3')
B512	GAGGCTGCCCGCAATGCCCTGCAGTACC	Mutagenèse TRBP R354N (5')
B513	GGTACTGCAGGGCATTGCGGGCAGCCTC	Mutagenèse TRBP R354N (3')
B514	GGGGATATCCATATGTCCCAGAGCAGGCACCG CGCCGAGG	Clonage de PACT + NdeI (5')
B515	GGGGATATCCATATGGGATGTACTTGGCATTCC TTGAGG	Clonage de PACT (212-313) + NdeI (5')
B516	GGGGGATCCTTACTTTCTTTCTGCTATTATCTTT AAATACTGC	Clonage de PACT + BamHI (3')
B641	GGAAGTGTCCACCCAGCCGATCACTGTGTGTC ATGG	Mutagenèse TRBP A334I (5')
B642	CCATGACACACAGTGATCGGCTGGGTGGACA GTTCC	Mutagenèse TRBP A334I (3')
B643	GGAGCTGAGCGCAAGTGGACTCTGC	Mutagenèse TRBP L319A (5')
B644	GCAGAGTCCACTTGCGCTCAGCTCC	Mutagenèse TRBP L319A (3')
B875	GGTTGCACCTGGGATAGCTTAAGGAACAGCG TGGGCGAGAAGATCCTGTCCC	Mutagenèse de la région de TRBP ciblée par un siARN (5'). Version avec un codon rare
B876	GGGACAGGATCTTCTCGCCCACGCTGTTTCTT AAGCTATCCAGGTGCAACC	Mutagenèse de la région de TRBP ciblée par un siARN (3'). Version avec un codon rare
B877	GGTTGCACCTGGGATAGCTTACGCAACAGCGT GGGCGAGAAGATCCTGTCCC	Mutagenèse de la région de TRBP ciblée par un siARN (5'). Version sans codon rare

B878	GGGACAGGATCTTCTCGCCCACGCTGTTGCGT AAGCTATCCCAGGTGCAACC	Mutagenèse de la région de TRBP ciblée par un siARN (3'). Version sans codon rare
5852	TTTTTTTTTTTTTTTTTTTT	Transcription inverse des ARNm
8536	AACAGCCTCAAGATCATCAGC	Expérience de PCR quantitative sur la GAPDH (5')
8537	GGATGATGTTCTGGAGAGCC	Expérience de PCR quantitative sur la GAPDH (3')
B879	GGTTGCACCTGGGATTCTCTAC	Expérience de PCR quantitative sur TRBP (262-366) (5')
B963	TCTCTACGAAATTCAGTAGG	Expérience de PCR quantitative sur TRBP (266-366) (5')
B880	TCACTTGCTGCCTGCCATGATC	Expérience de PCR quantitative sur TRBP (262-366) (3')

Tableau I : Liste des oligonucléotides qui ont été utilisés au cours du doctorat.

Le numéro, correspondant à la nomenclature utilisée au laboratoire, la séquence et la fonction de chaque oligonucléotide sont précisés.

C. Les plasmides

1. Pour le clonage Gateway®

pDONR™ 207 (5,6 kpb ; Life Technologies) : Ce plasmide a été utilisé pour la première étape (recombinaison BP) de clonage par la technique Gateway® de fragments PCR d'intérêts. Le plasmide pDONR™ 207 contient une origine de répllication High copy rep, un gène de résistance à la gentamycine, le gène toxique *ccdB* suivi du gène de résistance au chloramphénicol, bordés par les séquences de recombinaison attP1 et attP2.

2. Pour le double-hybride chez la levure *S. cerevisiae*

pACT2, pGBKT7 et pAS2 (8,1 kpb, 7,3 kpb et 8,4 kpb ; Clontech) : Ces plasmides ont été utilisés pour tester les interactions entre les protéines d'intérêts par la technique du double hybride chez *S. cerevisiae*. Au laboratoire les plasmides pACT2 et pAS2 ont été modifiés afin de les rendre compatible avec la technique de clonage Gateway®. Le plasmide pACT2 permet l'expression de la protéine d'intérêt fusionnée au domaine d'activation de la transcription du facteur Gal4 (Gal4-AD) au niveau de la région N-terminale, tandis que les plasmides pGBKT7 et pAS2 permettent l'expression de la protéine d'intérêt fusionnée au domaine de fixation à l'ADN du facteur Gal4 (Gal4-BD) au niveau de la région N-terminale. L'expression de ces protéines est sous le contrôle du promoteur ADH1 et du terminateur ADH1. Les plasmides pACT2 et pAS2 contiennent une origine de répllication High copy ColE1 et High copy 2 μ permettant sa répllication chez la bactérie et la

levure respectivement. Le plasmide pGBKT7 contient une origine de réplication High copy rep et high copy 2 μ permettant sa réplication chez la bactérie et la levure respectivement. Le plasmide pACT2 contient un gène de résistance à l'ampicilline et un gène d'auxotrophie LEU2 tandis que les plasmides pGBKT7 et pAS2 contiennent un gène de résistance à la kanamycine et l'ampicilline respectivement et un gène d'auxotrophie TRP1.

3. Pour la surproduction de protéines recombinantes chez *E. coli*

pnEA3ch, pnCS, pnYK (5,8 kpb, 3,5 kpb et 4,5 kpb) : Ces plasmides ont été utilisés pour tester les interactions entre protéines d'intérêts par la technique de co-expression et pour la production et la purification de protéines. Le plasmide pnYK a été créé à partir du pnYC. Le plasmide pnEA3CH permet l'expression de protéines fusionnées à une étiquette 6-Histidines au niveau de l'extrémité N-terminale, tandis que les plasmides pnCS et pnYK permettent l'expression de protéines non étiquetées. Ces trois plasmides contiennent un gène *lacI*, un promoteur et un terminateur T7 et un site RBS (« Ribosome Binding Site »). Le plasmide pnEA3CH contient un gène de résistance à l'ampicilline un site de coupure à la protéase PreScission situé entre l'étiquette 6-Histidines et l'ORF de la protéine d'intérêt. Les plasmides pnCS et pnYK contiennent respectivement un gène de résistance à la spectinomycine et à la kanamycine. Les plasmides pnEA3Ch, pnCS et pnYK possèdent respectivement une origine de réplication pBR322, p15A et CDF.

pnEA3chGW, pnCSGW, pnYKGW (6,8 kpb, 3,5 kpb et 5,5 kpb) : Ces plasmides ont été réalisés à partir des vecteurs pnEA3ch, pnCS et pnYK respectivement, afin de les rendre compatibles avec la technique de clonage Gateway®. Dans ce cas, les plasmides contiennent alors le gène de résistance au chloramphénicol et le gène toxique *ccdB*, bordés par les séquences de recombinaison attR1 et attR2.

4. Pour la production de protéines recombinantes en baculovirus

pAC8-6hist et pAC8-6hist-MBP : Le plasmide pAC8-6hist a été fourni par le l'IGBMC. Ces deux vecteurs ont été utilisés pour produire et purifier des protéines en infectant des cellules d'insectes Sf9 par un baculovirus recombinant. Le génome de ce baculovirus provient de la recombinaison entre le vecteur pAC-6hist (ou le vecteur pAC8-6hist-MBP) et un bacmide lui aussi fourni par l'IGBMC. Le vecteur pAC8-6hist permet l'expression d'une protéine fusionnée à une étiquette 6-histidines. Le plasmide pAC8-6hist-MBP a été réalisé à partir du vecteur pAC8-6hist par insertion

d'une séquence permettant l'expression d'une protéine fusionnée à son extrémité N terminale à une étiquette 6-Histidine-MBP. Ces 2 plasmides contiennent également un site de coupure à la protéase PreScission situé entre l'étiquette 6-Histidines ou 6-Histidines-MBP et l'ORF de la protéine d'intérêt et un gène de résistance à l'ampicilline.

5. Pour les expériences d'immunoprécipitation

pcDNA 3.1™ et pcDNA 3.1/nV5-DEST™ (5,4 kpb et 7,1 kpb ; Life Technologies) : Ces plasmides ont été utilisés transférer des cellules H.E.K. Le plasmide pcDNA 3.1/nV5-DEST™ permet l'expression de protéines fusionnées à l'étiquette V5 au niveau de l'extrémité N-terminale et est compatible avec le système Gateway® et contient ainsi le gène de résistance au chloramphénicol et le gène toxique *ccdB*, bordés par les séquences de recombinaison attR1 et attR2. Les vecteurs pcDNA 3.1™ et pcDNA 3.1/nV5-DEST™ contiennent une origine de répllication pUC et SV40, le promoteur CMV, le terminateur BGH qui est suivi d'un site de polyadénylation, ainsi qu'un gène de résistance à l'ampicilline et à la néomycine.

6. Pour la transcription d'un ARN

p2RZ (2,7 kpb ; Addgene) : Ce plasmide a été utilisé pour transcrire un pré-microARN fusionné en 3' avec le ribozyme HDV. Il contient de plus un gène de résistance à l'ampicilline, un promoteur T7 et une origine de répllication ApR.

D. Les milieux de cultures

Les milieux de culture de bactéries et de levures sont autoclavés à 121 °C, sous 1 bar de pression, pendant 20 minutes. Pour la préparation des milieux solides, de l'agar est additionné à raison de 17 g/l avant le passage à l'autoclave.

1. Cultures de bactéries

Milieu LB (Luria Bertani) : Peptone tryptique de caséine 5 g/l ; extrait de levure 10 g/l ; NaCl 5 g/l. Le pH est ajusté à 7,5 avec du NaOH 10 N.

Milieu Graffinity I : LB 2X ; glucose 0,5 g/l.

Milieu Graffinity II : Hepes 20 mM pH 7 ; LB 2X ; lactose 0,6 g/l, IPTG 1 mM.

Les milieux peuvent être complétés après stérilisation avec des antibiotiques à une concentration finale de 100 µg/ml pour l'ampicilline, 30 µg/ml pour le chloramphénicol, 15 µg/ml pour la gentamycine, 30 µg/ml pour la spectinomycine et 50 µg/ml pour la kanamycine.

2. Cultures de levures

Milieu riche YPD : Extrait de levure 10 g/l ; bactopeptone 20 g/l ; galactose 2 g/l

Milieu sélectif : Base azotée de levure sans acide aminé 10 g/l ; glucose 20 g/l ; mélange d'acides aminés (20 g/l de chaque acide aminé)

3. Culture de cellules humaines

Milieu DMEM (Dulbecco's Modified Eagle's Medium) : Milieu commercial (Sigma) supplémenté avec 10 % de sérum de veau fœtal, 20 mM final de glutamine, 1000 U/ml final de pénicilline et 1000 µg/ml final de streptomycine.

4. Culture de cellules d'insectes

Milieu Grace : Milieu commercial (Life Technologie) supplémenté avec 10 % de sérum de veau fœtal.

E. Les tampons et solutions d'usage courant

TE : Tris 10 mM, pH 8 ; EDTA 1 mM.

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

PBS-Tween 0, 1 % : Na₂HPO₄ 10 mM ; KH₂PO₄ 1,8 mM, pH 7,3 ; NaCl 140 mM ; KCl 2,7 mM ; Tween-20 0,1 %.

Tampon de migration SDS-PAGE : Glycine 25 g/l ; Tris 3 g/l ; SDS 0,1 %.

Bleu de dépôt 2X pour SDS-PAGE : Tris 5 mM, pH 6,8 ; glycérol 25 % ; SDS 2 % ; bleu de bromophénol 0,01 % ; β-mercaptoéthanol 5 %.

Bleu de formamide 2X : EDTA 20 mM ; bleu de bromophénol 0,05 % ; bleu de xylène cyanol 0,05 % ; dans de la formamide désionisée.

TfbI : Acétate de potassium 30 mM ; KCl 100 mM ; CaCl₂ 10 mM ; glycérol 15 % et MnCl₂ 50 mM (ajouté extemporanément).

TfbII : MOPS-Na pH 7 10 mM ; CaCl₂ 75 mM ; KCl 10 mM et Glycérol 15 %.

Tampon de transcription 10X : Hepes 800 mM, pH 7,5 ; spermidine 10 mM ; DTT 200 mM et Triton-X100 0.1 %.

HBS 2X : Hepes 50 mM, pH 7,5 ; NaCl 0,28 M ; KCl 10 mM ; Na₂HPO₄ 1,5 mM.

Tampon d'hybridation : Tris 10 mM, pH 8 ; NaCl 50 mM ; EDTA 1 mM.

HNTG : Hepes 20 mM, pH 7,9 ; NaCl 150 mM ; glycérol 10 %, MgCl₂ 1 mM ; EGTA 1 mM ; anti-protéase 1X.

Tampon bas sel : Tris-HCl 20 mM, pH 8 ; NaCl 50 mM ; Imidazole 7 mM.

Tampon haut sel : Tris-HCl 20 mM, pH 8 ; NaCl 400 mM ; Imidazole 7 mM.

Tampon de purification : Hepes 25 mM, pH 7,5 ; NaCl 300 mM ; TCEP 0,5 mM ; Imidazole 10 mM.

Tampon d'éluion : Hepes 25 mM, pH 7,5 ; NaCl 300 mM ; TCEP 0,5 mM ; Imidazole 300 mM.

Tampon pour chromatographie d'exclusion stérique (RMN) : Na₂HPO₄,12H₂O 10 mM, pH 6.4 ; NaCl 150 mM ; TCEP 0,5 mM.

Milieu M9 : Na₂HPO₄,12H₂O 42 mM ; KH₂PO₄ 22 mM ; NaCl 8,6 mM ; Métaux Mix I 200 µl/l ; Métaux Mix II 100 µl/l ; MgSO₄ 2 mM ; ¹⁵NH₄Cl 1 g/l ; Glucose ¹³C 2 g/l ; Biotine 1 mg/l ; Acide folique 1 mg/l ; Nicotinamide 1 mg/l ; Riboflavine 0,1 mg/l ; Thiamine 5 mg/l.

Métaux Mix I : FeCl₃ 50 mM.

Métaux Mix II : CaCl₂ 20 mM ; MnCl₂ 10 mM ; ZnSO₄ 10 mM ; CoCl₂ 2 mM ; CuCl₂ 2 mM ; NiCl₂ 2 mM ; Na₂MoO₄ 2 mM ; Na₂SeO₃ 2 mM ; H₃BO₃ 2 mM.

Tampon de réaction : Hepes 40 mM, pH 7,5 ; NaCl 100 mM ; MgCl₂ 3 mM ; DTT 1 mM.

Tampon de transfection HEPES : Hepes 25 mM, pH 7 ; NaCl 140 mM ; CaCl₂ 125 mM.

Tampon de lyse : Na₂HPO₄,12H₂O 50 mM, pH 8 ; NaCl 300 mM ; Imidazole 10 mM ; Triton 0,05 % ; glycérol 5 % ; TCEP 0,5 mM ; inhibiteur de protéase EDTA-free 1X.

Tampon de lavage : Na₂HPO₄,12H₂O 50 mM, pH 8 ; NaCl 300 mM ; Imidazole 10 mM ; glycérol 5 % ; TCEP 0,5 mM.

Tampon d'éluion II : Na₂HPO₄,12H₂O 50 mM, pH 8 ; NaCl 300 mM ; Imidazole 300 mM ; Triton 0,05 % ; glycérol 5 % ; TCEP 0,5 mM.

F. Les anticorps

Les anticorps suivant ont été utilisés pour les analyses par Western-blot, les immunoprécipitations, ainsi que pour les expériences de Duolink™ et d'immunofluorescence :

- anti-RPAP3 (référence : SAB1407956 - Sigma-Aldrich) : Anticorps monoclonal produit chez la souris. Utilisé dilué au 1/250^{ème} pour les expériences de Duolink™ et au 1/100^{ème} pour les expériences d'immunofluorescence.

- anti-RPAP3 (référence : SAB1411438 - Sigma-Aldrich) : Anticorps monoclonal produit chez le lapin. Utilisé à une concentration de 1/2000^{ème} pour les expériences de western-blot.
- anti-TRBP (référence : ab72110 - Abcam) : Anticorps polyclonal produit chez le lapin. Utilisé dilué au 1/100^{ème} pour les expériences de DuolinkTM, à une concentration de 1/500^{ème} pour les expériences de Western-blot et dilué au 1/100^{ème} pour les expériences d'immunofluorescence.
- anti-TRBP (référence : ab129325 - Abcam) : Anticorps monoclonal produit chez la souris. Utilisé à une concentration de 1/500^{ème} pour les expériences de western-blot.
- anti-Dicer (référence : ab14601 - Abcam) : Anticorps monoclonal produit chez la souris. Utilisé à une concentration de 1/400^{ème} pour les expériences de western-blot.
- anti-Actine (référence : ab3280 - Abcam) : Anticorps monoclonal produit chez la souris. Utilisé à une concentration de 1/400^{ème} pour les expériences de western-blot.
- anti-GAPDH (référence : ab9485 - Abcam) : Anticorps polyclonal produit chez le lapin. Utilisé dilué au 1/750^{ème} pour les expériences de Duolink.
- anti-GAPDH (référence : ab8245 - Abcam) : Anticorps monoclonal produit chez la souris. Utilisé dilué au 1/750^{ème} pour les expériences de Duolink.
- anti-V5 (référence : GTX117997 - ThermoFischer Scientific) : Anticorps polyclonal produit chez le lapin. Utilisé à une concentration de 1/4000^{ème} pour les expériences de Western-blot.
- anti-V5 (référence : 37-7500 - ThermoFischer Scientific) : Anticorps monoclonal produit chez la souris. Utilisé pour des immunoprécipitations.
- anti-lapin (référence 31460 - ThermoFischer Scientific) : Anticorps secondaire couplé à la peroxydase et produit chez la chèvre. Utilisé à une concentration de 1/20000^{ème} pour les expériences de western-blot.
- anti-souris (référence : 515035003 - ThermoFischer Scientific) : Anticorps secondaire couplé à la peroxydase et produit chez le mouton. Utilisé à une concentration de 1/10000^{ème} pour les expériences de western-blot.
- anti-souris couplé à de l'alexa-fluor 488 (ThermoFischer Scientific) : Utilisé dilué au 1/50^{ème} pour les expériences d'immunofluorescence.
- anti-lapin couplé à de l'alexa-fluor 555 (ThermoFischer Scientific) : Utilisé dilué au 1/250^{ème} pour les expériences d'immunofluorescence.

II. Méthodes

A. Méthodes relatives aux acides nucléiques

1. Clonage

a. PCR

Les expériences de PCR ont été réalisées à l'aide d'un thermocycleur Agilent. Le programme utilisé est une dénaturation de 5 min à 95 °C suivie de 30 cycles comprenant une dénaturation à 95°C de 30s, puis d'une hybridation à la température de demi-dénaturation moins 5 °C et enfin d'une élongation à 72 °C. Après les 30 cycles, une élongation de 10 min à 72 °C est ajoutée puis les amplicons sont stockés à 20 °C dans l'appareil. Les amplifications ont été réalisées dans 50 µl à partir de 50 ng d'ADN matrice auquel a été ajouté 10 µl de tampon 5X HF (ThermoFisher Scientific), 8 µl d'un mélange des 4 dNTP à une concentration de 1,25 mM chacun, 1 µl à 10 µM de chaque oligodésoxyribonucléotide et 2 unités d'enzyme « Phusion DNA Polymerase » de ThermoFisher Scientific.

b. Clonage par ligature

Les différents ADN codant les protéines ou fragments de protéines ont été amplifiés par PCR comme décrit précédemment en utilisant des oligodésoxyribonucléotides portant des sites de restriction à leurs extrémités (voir **Tableau I**). Vingt microlitres de PCR sont digérés à 37 °C par 1 unité de chaque enzyme de restriction FastDigest (ThermoFisher Scientific) dans du tampon FastDigest (ThermoFisher Scientific) pour un volume final de 30 µl. Les enzymes de restriction sont ensuite inactivées à une température et pendant un temps spécifique de l'enzyme utilisée, en accord avec ce qui est préconisé par le fournisseur. Dix microgrammes du plasmide de clonage sont digérés par 10 unités de chaque enzyme de restriction FastDigest utilisée pour l'insert dans du tampon FastDigest à 37 °C durant 20 min dans un volume final de 100 µl. Le plasmide est ensuite extrait par traitement au phénol/chloroforme (v/v) et précipité. Pour ce faire le volume est tout d'abord ajusté à 200 µl avec de l'eau ultra-pure, puis un volume équivalent de phénol/chloroforme est ajouté et le mélange est mélangé pendant 1 min avant d'être centrifugé pendant 5 min à 10000 g. Une précipitation alcoolique est alors effectuée sur le surnageant à l'aide de 3 volumes d'éthanol 100 % et de 0,3 M d'acétate de sodium. Le mélange est ensuite incubé à -20 °C pendant 30 min avant d'être centrifugé pendant 30 min à 10000 g à 4 °C. L'éthanol 100 % est retiré et le culot d'ADN est lavé par ajout de 800 µl d'éthanol 70 %. Après une centrifugation pendant 15 min à 4 °C à 10000 g l'éthanol 70 % est totalement retiré puis le culot est finalement dissous dans 20 µl d'eau ultra-pure. La ligature est réalisée par l'incubation d'une nuit à température ambiante d'un mélange

réactionnel de 10 µl contenant 50 ng de vecteur digéré, un excès d'insert digéré de trois fois par rapport au vecteur, 5 unités de T4 DNA ligase (ThermoFisher Scientific) et du tampon de ligature « T4 DNA ligase buffer » (ThermoFisher Scientific). Cinquante microlitres de bactéries *E. coli* DH5α super-compétentes sont ensuite transformées par 2 µl de ce mélange réactionnel. Les clones positifs sont amplifiés par mini-préparation ou midi-préparation de plasmides, puis ceux-ci sont séquencés par la société GATC.

c. Mutagenèse dirigée

Le programme utilisé est le même que ci-dessus à l'exception du nombre de cycle qui n'est que de 18. Les amplifications ont été réalisées dans 50 µl à partir de 50 ng d'ADN matrice auquel a été ajoutés 10 µl de tampon 5X HF, 8 µl d'un mélange des 4 dNTP à une concentration de 1,25 mM chacun, 0,3 µl à 10 µM de chaque oligodésoxyribonucléotide portant la mutation voulue (voir **Tableau I**) et 2 unités d'enzyme « Phusion DNA Polymerase » de ThermoFischer. Vingt microlitres du mélange de PCR sont incubés dans un volume final de 50 µl avec 1 µl de *DpnI* FastDigest (ThermoFisher Scientific) et 3 µl de tampon FastDigest 10X (ThermoFisher Scientific) à 37 °C pendant 30 min. Cinquante microlitres de bactéries *E. coli* DH5α super-compétentes sont finalement transformées par 2 µl de ce mélange réactionnel. Les clones positifs sont amplifiés par mini-préparation ou midi-préparation de plasmides, puis séquencés par la société GATC.

d. Clonage par recombinaison Gateway®

i. Recombinaison avec le vecteur pDONR207 :

Les différents ADN codant pour les protéines ou fragments de protéines ont été amplifiés par PCR en utilisant des oligodésoxyribonucléotides portant des sites de recombinaison AttB1 et AttB2 (voir **Tableau I**). Deux microlitres du mélange PCR sont ensuite incubés avec 2 µl de TE, 1 µl à 100 ng/µl de pDONR207 et 2 µl du mixte BP Clonase II (ThermoFisher Scientific) pendant 2 h à toute une nuit à température ambiante. Cinquante microlitres de bactéries *E. coli* DH5α super-compétentes sont finalement transformées par 2 µl de ce mélange réactionnel. Les clones positifs sont amplifiés par mini-préparation ou midi-préparation de plasmides, puis ceux-ci sont séquencés par la société GATC.

ii. Recombinaison avec les vecteurs de destination :

Cent nanogrammes de pDONR207 recombinants sont incubés avec 2 µl de TE, 100 ng de vecteur de destination et 2 µl du mixte LR Clonase II (ThermoFisher Scientific) pendant 2 h à toute une nuit

à température ambiante. Cinquante microlitres de bactéries *E. coli* DH5 α super-compétentes sont finalement transformées par 2 μ l de ce mélange réactionnel. Les clones positifs sont amplifiés par mini-préparation ou midi-préparation de plasmides, puis ceux-ci sont séquencés par la société GATC.

e. Clonage par recombinaison par la technique d'In-Fusion™ de Clontech

L'insert est amplifié par PCR comme indiqué ci-dessus. Le vecteur de destination est linéarisé au niveau du site d'insertion par PCR. Les matrices PCR sont éliminées par un traitement à la *DpnI*. Pour ce faire, les cinquante microlitres de réaction de PCR sont incubés dans un volume final de 60 μ l avec 2 μ l de *DpnI* FastDigest et 6 μ l de tampon FastDigest 10X à 37 °C pendant 30 min. L'insert et le vecteur linéarisé sont ensuite purifiés à l'aide du kit « QIAEX II Gel Extraction » de Qiagen. La réaction de recombinaison est effectuée dans un volume final de 10 μ l avec 50 ng d'insert purifié, 150 ng de vecteur linéarisé purifié et 2 μ l de réactif « In-Fusion HD Enzyme Premix 5X » (Clontech). La réaction est incubée pendant 15 min à 50 °C puis placée dans la glace. Cinquante microlitres de bactéries *E. coli* DH5 α super-compétentes sont finalement transformées par 2 μ l de ce mélange réactionnel. Les clones positifs sont amplifiés par mini-préparation ou midi-préparation de plasmides, puis ceux-ci sont séquencés par la société GATC.

f. Transformation de bactérie *E. coli* DH5 α ou BL21

i. Obtention de bactérie *E. coli* DH5 α ou BL21 super-compétentes

Une culture de 100 ml de milieu LB estensemencée avec 10 uDO_{600 nm} d'une culture saturante d'*E. coli* (DH5 α ou BL21). Arrivée à une concentration cellulaire située entre 0,6 et 0,7 uDO_{600 nm}/ml, la culture est centrifugée 30 min à 2500 g, puis le culot est repris dans 30 ml de tampon TfbI et incubé pendant 2 h à 4 °C. Après une centrifugation de 10 min à 2500 g, le culot est dissous dans 4 ml de TfbII puis aliquoté. Les aliquotes sont finalement congelées par de l'azote liquide et conservées à -80 °C. Pour les bactéries *E. coli* BL21, les cultures sont réalisées en présence de chloramphénicol à une concentration finale de 30 μ g/ml.

ii. Transformation de bactérie *E. coli* DH5 α par choc thermique

Un à cinq microlitres de mélange à transformer sont incubés avec 50 μ l de bactéries *E. coli* DH5 α super-compétentes pendant 30 min à 4 °C. La transformation est effectuée par un choc thermique de 45 s à 42 °C puis pendant 2 min à 4 °C. Si nécessaire les bactéries sont régénérées pendant 45 min sous agitation à 37 °C après l'ajout d'un volume de 200 μ l de milieu LB avant d'être étalées sur des

boîtes LB + antibiotiques et incubées une nuit à 37 °C.

2. Obtention du pre-microARN 30E

a. Préparation de la matrice ADN

Dix microgrammes de vecteur p2RZ_C-premicroARN 30E sont digérés dans un volume final de 300 µl par 10 µg d'enzyme *XhoI* FastDigest (ThermoFischer Scientific) dans 30 µl de tampon FastDigest 10X pendant 1h à 37 °C. Le vecteur linéarisé est purifié par traitement au phénol/chloroforme (v/v) et précipité comme décrit dans le paragraphe II. A. 1. b. du chapitre Matériel et Méthodes.

b. Transcription *in vitro*

Cinq microlitres de p2RZ_C-premicroARN 30E à une concentration de 500 ng/µl sont incubés avec 5 µl de tampon de transcription 10X, 2,5 µl d'ARN polymérase T7 (produite au laboratoire), 25 mM de MgCl₂, 10 µl de rNTP à 25 mM chacun et 10 U de RNase Inhibiteur (ThermoFischer Scientific) dans un volume final de 50 µl.

c. Purification du pre-microARN 30E

Cinquante microlitres de transcrit, additionnés de 50 µl de bleu de formamide 2X sans xylène cyanol, sont fractionnés sur un gel d'acrylamide-urée 8M 12,5 %. La bande correspondant au pre-microARN 30E, visualisée par la technique d'ombrage aux UVs, est découpée sur le gel et l'ARN est élué du gel dans 800 µl de tampon d'éluion (kit RPA III de ThermoFischer Scientific) pendant toute une nuit à 4 °C sur roue. Le pre-microARN est ensuite purifié par traitement au phénol/chloroforme (v/v) et précipité comme décrit dans le paragraphe II. A. 1. b du chapitre Matériel et Méthodes.

d. Déphosphorylation et marquage d'un ARN

Un microgramme d'ARN purifié est incubé avec 1 µl d'enzyme FastAP Thermosensitive Alkaline Phosphatase 1 U/µl (ThermoFischer Scientific), 2 µl de tampon FastAP 10X (ThermoFischer Scientific), dans un volume final de 20 µl, pendant 20 min à 37 °C. La réaction est stoppée en l'incubant pendant 10 min à 75 °C. L'ARN déphosphorylé est purifié sur une colonne « Mini Quick Spin Oligo » (Roche) puis marqué. Pour cela, 200 ng d'ARN déphosphorylés sont incubés avec 1 µl de T4 polynucleotide kinase (ThermoFischer Scientific, 10 U/µl), 2 µl de tampon A 10X (ThermoFischer Scientific) et 1 µl de γ -³²P-ATP 3000 mCi/mmol 10 mCi/ml, dans un volume final

de 20 µl pendant 1 h à 37 °C. L'ARN est finalement purifié à l'aide d'une colonne « Mini Quick Spin Oligo » (Roche).

3. Transfection de cellules HEK par un siARN

a. Préparation d'un siARN à partir de deux oligoribonucléotides de synthèses

Les deux oligoribonucléotides sont incubés à concentration égale dans du tampon d'hybridation pendant 5 min à 95 °C puis laissés pendant 1 h à T °C ambiante.

b. Transfection de cellules HEK au phosphate de calcium

La transfection de cellules HEK par des siARN (OnTarget™ de Dharmacon ou duplexe d'oligoribonucléotides) a été réalisée par la méthode du phosphate de calcium. Trois millilitres de milieu DMEM sontensemencés à raison de 100000 cellules HEK par puits de 3,5 cm. Vingt-quatre heures plus tard, les cellules sont transfectées par le siRNA. Pour cela, 5 µl de siRNA à 10 µM sont incubés avec 37 µl de TE 0,1X, 5,2 µl de CaCl₂ 2 M et 42 µl d'HBS 2X, mélangés par bullage et incubés pendant 30 min à température ambiante. Le milieu est ensuite ajouté goutte-à-goutte aux cellules. Six heures plus tard, le milieu de culture est changé. Quarante-huit heures après la première transfection les cellules sont de nouveau transfectées selon le même protocole et le milieu de culture est de nouveau changé 6 heures après la seconde transfection. Vingt-quatre à 48 h après la seconde transfection le milieu de culture est retiré par aspiration et les cellules sont grattées et récupérées dans 1 ml de PBS (ThermoFischer Scientific). Après une centrifugation de 15 min à 2500 g, le PBS est retiré et les cellules sont congelées à -80 °C.

4. Transcription inverse et PCR quantitative

a. Extraction des ARN totaux de cellule H.E.K.

Les cellules sont reprises dans du Trizol (ThermoFischer Scientific) à raison de 1 ml pour 10 cm² de surface de culture. Du chloroforme est ajouté à la solution (0,2 ml/ml de Trizol) et le mélange est agité fortement avant une étape de centrifugation à 12000 g pendant 20 min à 4 °C. Après centrifugation, les ARN totaux présents dans la phase aqueuse sont prélevés et précipités avec 500 µl d'isopropanol pour 1 ml de trizol et 0,3 M d'acétate de sodium 1 h à -20 °C. Après une centrifugation à 12 000 g pendant 20 min à 4 °C, l'isopropanol est retiré et le culot d'ARN est lavé avec 800 µl d'éthanol 70 %. Après une centrifugation à 12000 g pendant 5 min à 4 °C, l'éthanol est retiré et le culot est dissous dans 10 µl d'eau ultra-pure.

b. Élimination de l'ADN génomique

Deux microgrammes d'ARN sont incubés avec 2 µl d'enzyme RQ1 (ThermoFisher Scientific) et 1 µl de tampon RQ1 10X dans un volume final de 10 µl pendant 40 min à 37 °C. La réaction est stoppée par l'ajout d'1 µl de tampon RQ1 Stop et une incubation à 65 °C pendant 10 min.

c. Transcription inverse des ARNm

Deux microgrammes d'ARN traités avec l'enzyme RQ1 sont incubés avec 1 µl d'enzyme MMLV (ThermoFischer Scientific), 2 µl à 100 µM d'oligodésoxyribonucléodes polydT (voir **Tableau I**) et 8 µl de dNTP 100 mM dans un volume final de 40 µl pendant 1 h à 42 °C.

d. PCR quantitative

Un microlitre d'ADNc (voir ci-dessus) est incubé avec 0,35 µl de chaque oligodésoxyribonucléotides (100 µM) et 5 µl de SyBr Green MasterMix 2X (ThermoFischer Scientific) dans un volume final de 10 µl. Le programme de PCR quantitative est le suivant : une dénaturation initiale pendant 10 min à 95 °C suivie de 40 cycles [dénaturation 95 °C, 15 s ; hybridation 60 °C, 30 s ; élongation 72 °C 30 s]. Ce programme est suivi d'une étape visant à déterminer une courbe de fusion permettant de vérifier qu'un seul produit de PCR est amplifié : 40 cycles [dénaturation 95 °C, 15 s ; hybridation/élongation 60 °C, 1 min ; dénaturation 95 °C avec un passage de 60 °C à 95 °C en raison d'un degré toutes les deux secondes].

5. Préparation des ARN pour puce à ARN non-codant

Après extraction des ARN totaux au Trizol (voir ci-dessus), les ARN sont purifiés à l'aide du kit « miRCURY RNA isolation » d'Exiqon. Pour cela, le volume d'ARN est ajusté à 100 µl avec de l'eau ultra-pure et ajouté à 250 µl de tampon de lyse (kit). Deux cents microlitres d'éthanol 100 % sont alors ajoutés au mélange qui est ensuite passé sur colonne. La colonne est ensuite lavée 3 fois avec 400 µl de solution de lavage (kit). Les dernières traces de la solution de lavage sont éliminées par une centrifugation de 2 min à 14000 g puis les ARN sont élués dans 35 µl de tampon d'éluion (kit). Les ARN sont ensuite concentrés par précipitation alcoolique à froid, lavés puis dissous dans 5 à 10 µl d'eau ultra-pure et conservés à -80 °C.

6. Coupure du pré-microARN 30E par Dicer en présence de TRBP et/ou de RPAP3

Dix nanomolaires de pré-microARN 30E marqué en 5' avec du ³²P sont incubés dans 20 µl de tampon de réaction à 65 °C pendant 10 min puis rapidement placés dans la glace. La réaction de

coupure par Dicer est alors réalisé en présence de 200 nM de la protéine Dicer purifiée, du complexe Dicer-TRBP ou en présence des protéines Dicer, TRBP et RPAP3 dans des concentrations spécifiées, pendant une heure à 37 °C. Les produits de coupures sont ensuite fractionnés sur un gel dénaturant acrylamide-urée 8 M de 15 % et visualisés à l'aide d'un phosphorimager.

B. Méthodes relatives aux protéines

1. Méthodes relatives à la levure *Saccharomyces cerevisiae*

a. Préparation de levures *Saccharomyces cerevisiae* compétentes

Une souche de levure estensemencée dans du 10 ml de milieu YPD et mise sous agitation pendant 48 h à 30 °C. À partir de cette culture en phase stationnaire, 50 ml de milieu YPD sontensemencés à 0,1 uDO_{600 nm}/ml et mis sous agitation à 30 °C jusqu'à obtenir une concentration cellulaire située entre 0,5 à 0,8 uDO_{600 nm}/ml. La culture est alors centrifugée pendant 10 min à 2500 g à 4 °C, puis le culot est repris et lavé dans 30 ml de tampon (Tris-HCl pH 7,5 10 mM, EDTA 1 mM, Acétate de lithium 100 mM). Une centrifugation de 10 min à 2500 g à 4 °C est effectuée et le culot est repris dans 500 µl de ce même tampon.

b. Transformation de la levure *Saccharomyces cerevisiae*

Pour chaque transformation, un volume de 50 µl de suspension cellulaire est incubé pendant 30 min à 30 °C, en présence d'1 µg d'ADN plasmidique, de 45 µg d'ADN de sperme de saumon (Invitrogen), préalablement dénaturé pendant 15 min à 96 °C, et de 300 µl de tampon (Tris-HCl pH 7,5 10 mM, EDTA 1 mM, acétate de lithium 100 mM, PEG 4000 40 %). Le mélange est ensuite incubé pendant 10 min à 42 °C. Un volume de 600 µl d'eau stérile est ajouté au mélange qui est ensuite centrifugé pendant 1 min à 6000 g, puis un volume de 800 µl est retiré. Finalement le culot est repris dans le volume restant et est étalé sur un milieu sélectif solide.

c. Double-hybride

La souche Y187 (Mat α) est transformée par le plasmide pGBKT7 ou pAS2 qui permet l'expression d'une protéine fusionnée au domaine de fixation à l'ADN de Gal4, tandis que la souche CG1945 (Mat a) est transformée par le plasmide pACT2 qui permet l'expression d'une protéine fusionnée au domaine d'activation de la transcription de Gal4. Les souches haploïdes transformées sont ensuite diluées séparément dans 50 µl d'eau stérile, avant d'être croisées par la superposition d'un volume de 8 µl de chaque dilution sur milieu solide YPD. Après 2 jours d'incubation à 30 °C, les souches diploïdes potentiellement générées sont diluées séparément dans 50 µl d'eau stérile, avant d'être

sélectionnées par le dépôt d'un volume de 8 µl de chaque dilution sur un milieu solide sélectif (Trp⁻, Leu⁻). Après quelques jours d'incubation à 30 °C, les souches diploïdes sont à nouveau diluées séparément dans 50 µl d'eau stérile, avant de tester l'interaction entre les deux protéines d'intérêt par des dépôts de 8 µl de chaque dilution sur milieu solide sélectif (Trp⁻, Leu⁻, His⁻) avec différentes concentrations en 3-AT (Sigma). Le niveau de croissance est analysé après trois à quatre jours d'incubation à 30 °C.

2. Transfection de cellule HEK par un plasmide

Dix millilitres de milieu DMEM sontensemencés avec 3 millions de cellule H.E.K. en boîte de pétri de 10 cm. Vingt-quatre heures plus tard, les cellules sont transfectées par 15 µg de plasmide. Pour chaque réaction de transfection, 15 µg d'ADN sont dilués dans 250 µl de NaCl 150 mM d'une part, et 30 µl de JetPEI (Polyplus Transfection) sont dilués dans 250 µl de NaCl 150 mM d'autre part. Ces deux dilutions sont mélangées puis incubées 1 h à température ambiante. Le mélange est ensuite ajouté goutte-à-goutte sur les cellules en culture. Après 48 à 72 h de culture, le milieu des cellules est retiré par aspiration et les cellules sont récoltées dans 1 ml de PBS (ThermoFischer Scientific) puis culottées et congelées à -80 °C.

3. Préparation d'extraits protéiques à partir de cellule Hela/HEK

Les cellules culottées sont reprises dans 100 ou 500 µl de tampon HNTG puis incubés pendant 20 min dans la glace. Après 20 min de centrifugation à 10000 g, le surnageant contenant les protéiques totales est récupéré et conservé si nécessaire à -20 °C.

4. Immunoprécipitation

Pour chaque immunoprécipitation, 100 µl de billes magnétiques couplées à la protéine G (ThermoFischer Scientific) sont lavées trois fois avec 500 µl de PBS (ThermoFischer Scientific) puis incubées pendant 2 h sur roue à température ambiante avec 10 µg d'anticorps dilués dans 500 µl de PBS. Un témoin négatif est réalisé sans anticorps. Un extrait protéique est préparé pour chaque immunoprécipitation à partir de 8 boîtes de 10 cm de cellule H.E.K. arrivées à 90 % de confluence. Pour cela, les cellules sont récoltées puis les protéines totales sont extraites en incubant le culot cellulaire avec 500 µl de tampon HNTG comme décrit précédemment. Si nécessaire, l'extraction est réalisée en présence de 15 µg de RNase A (ThermoFischer Scientific) par boîtes. Cinquante microlitres d'extrait protéique total sont conservés pour servir de témoin interne d'expression (« Input »). Après trois lavages avec 500 µl de PBS, les billes sont incubées avec

l'extrait protéique pendant 2 h sur roue à température ambiante. Les billes sont ensuite lavées 3 fois avec 500 µl de tampon PBS puis reprises dans 40 µl de bleu dénaturant 2X. Quinze microlitres de témoin d'expression interne, de témoin négatif et de l'expérience d'immunoprécipitation sont alors fractionnés sur un gel de protéines dénaturant puis analysés par western-blot.

5. Western-blot

Le transfert des protéines d'un gel dénaturant sur une membrane PVDF préalablement activée avec de l'éthanol 100 % est réalisé à l'aide d'un appareil Pierce G2 de ThermoFischer Scientific, sous un courant de 1,5 A pendant 10 min. Après rinçage dans de l'éthanol 100 %, la membrane est incubée pendant 30 min avec une solution composée de 5 % de lait dilué dans du PBS-Tween 0,1 %. La membrane est ensuite lavée 2 fois pendant 10 min avec du PBS-Tween 0,1 % puis incubée avec l'anticorps primaire dilué (voir paragraphe I. F du chapitre Matériel et Méthodes) dans une solution de lait 5 % dilué dans du PBS-Tween 0,1 % pendant 1 h. Après 3 lavages de 10 min avec du PBS-Tween 0,1 %, la membrane est incubée pendant 45 min avec l'anticorps secondaire correspondant également dilué dans une solution de lait 5 % dilué dans du PBS-Tween 0,1 %. La membrane est de nouveau lavée 3 fois 10 min avec du PBS-Tween 0,1 % puis révélée à l'aide du kit ECL prime d'Amersham. Si nécessaire les anticorps primaires et secondaires sont éliminés de la membrane à l'aide du kit « re-Blot plus middle » de Millipore et la membrane est de nouveau incubée avec une solution de lait 5 % dilué dans du PBS-Tween 0,1 %.

6. Immunofluorescence

Trois millilitres de milieu DMEM sontensemencés avec 100000 cellules HeLa. dans un puits de 3,5 cm sur une lamelle de microscope stérile. Une fois les cellules à 80-90 % de confluence, le milieu est retiré par aspiration et les lamelles sont récupérées. Après trois lavages avec du PBS (ThermoFischer Scientific), les cellules sont fixées sur la lamelle pendant 20 min à l'aide d'une solution de formaldéhyde 3 %. Après trois nouveaux lavages avec du PBS, les cellules sont perméabilisées pendant 5 min à l'aide d'une solution de triton 0,5 % dilué dans du PBS puis rincées 3 fois avec du PBS. Les cellules sont ensuite incubées pendant 30 min avec une goutte (environ 50 µl) de solution de blocage (kit Duolink™ d'Olink). Une fois les lamelles délicatement récupérées, les cellules sont lavées 3 fois avec du PBS puis incubées avec 200 µl d'anticorps dilués (voir paragraphe I. F du chapitre Matériel et Méthodes) dans une solution de dilution pour anticorps (kit Duolink™ d'Olink) pendant 1 h30 min à température ambiante. Après 3 lavages avec du PBS,

les cellules sont incubées avec 200 µl d'anticorps secondaires diluées (voir paragraphe I. F du chapitre Matériel et Méthodes) dans une solution de dilution pour anticorps (kit Duolink™ d'Olink) pendant 1 h30 min. Après 3 lavages dans du PBS, les lamelles sont montées sur des lames de microscope avec 12 µl de milieu de montage (kit Duolink™ d'Olink). Si nécessaire, les noyaux des cellules sont marqués avec du To-Pro-3 (ThermoFisher Scientific) dilué au 1/1000^{ème} directement dans le milieu de montage. L'acquisition des images a été réalisée sur un microscope SP5 de la plateforme de microscopie confocale de la Fédération de recherche FR3209.

7. Duolink™

Trois millilitres de milieu DMEM sontensemencés avec 100000 cellules HeLa. dans un puits de 3,5 cm sur une lamelle de microscope stérile. Une fois les cellules à 80-90 % de confluence, le milieu est retiré par aspiration et les lamelles sont récupérées. Après trois lavages avec du PBS (ThermoFischer Scientific), les cellules sont fixées sur la lamelle pendant 20 min à l'aide d'une solution de formaldéhyde 3 %. Après trois nouveaux lavages avec du PBS, les cellules sont perméabilisées pendant 5 min à l'aide d'une solution de triton 0,5 % dilué dans du PBS puis rincées 3 fois avec du PBS. Les cellules sont ensuite incubées pendant 30 min avec une goutte (environ 50 µl) de solution de blocage (kit Duolink™ d'Olink). Après 3 lavages avec du PBS, les cellules sont incubées avec 50 µl d'anticorps dilués (voir paragraphe I. F du chapitre Matériel et Méthodes) dans une solution de dilution pour anticorps (kit Duolink™ d'Olink) pendant 1 h à température ambiante. Après 3 nouveaux lavages avec du PBS, les cellules sont incubées avec 50 µl de sonde PLA (kit Duolink™ d'Olink) diluée au 1/5^{ème} dans du PBS pendant 1 h à 37 °C en environnement humide. Après 2 lavages de 5 min avec du PBS, les cellules sont incubées avec 50 µl de solution de ligature (kit Duolink™ d'Olink) diluée au 1/40^{ème} pendant 30 min à 37 °C en chambre humide. Les cellules sont ensuite lavées pendant 2 min avec du PBS, deux fois, puis incubées pendant 100 min à 37 °C en chambre humide avec 50 µl d'ADN polymérase (kit Duolink™ d'Olink) diluée au 1/80^{ème} dans une solution d'amplification (kit Duolink™ d'Olink) elle-même diluée au 1/5^{ème} dans de l'eau ultra-pure. Après 2 lavages de 10 min, les lamelles sont montées sur des lames de microscope avec 12 µl de milieu de montage (kit Duolink™ d'Olink). Si nécessaire, les noyaux des cellules sont marqués avec du To-Pro-3 (ThermoFisher Scientific) dilué au 1/1000^{ème} directement dans le milieu de montage. Avant montage, le cytoplasme des cellules peut être marqué. Pour cela les cellules sont incubées pendant 20 min avec une unité de phalloïdine couplée à de l'alexa-fluor 546 (ThermoFischer Scientific) puis lavées 3 fois avec du PBS. L'acquisition des images a été réalisée

sur un microscope SP5 de la plateforme de microscopie confocale de la Fédération de recherche 3209.

8. Production et purification de protéines recombinante chez *E. coli*

a. Transformation de bactéries *E. coli* BL21 par plusieurs vecteurs d'expression

Des bactéries *E. coli* BL21 sont transformées par un mélange de 2 ou 3 plasmides portant les ORF des protéines à co-exprimer. Pour cela, 30 µl de bactérie *E. coli* BL21 super-compétentes sont incubées pendant 30 min dans la glace avec 200 ng de chaque vecteur. La transformation des bactéries est effectuée par choc thermique comme décrit précédemment (voir paragraphe II. A. 1. f du chapitre Matériel et Méthodes) puis les bactéries sont régénérées pendant 1 h à 37 °C avec 200 µl de milieu LB avant d'être étalées sur une gélose LB + antibiotiques.

b. Production de protéines recombinantes chez *E. coli* BL21

i. Pour les analyses par RMN homo- et hétéronucléaire

Une préculture de bactéries *E. coli* BL21, précédemment transformées, réalisées en milieu LB contenant les antibiotiques requis est utilisée pour ensemençer une culture réalisée en milieu minimum. Pour cela, un litre de milieu M9 ¹⁵N (pour la RMN homonucléaire) ou ¹⁵N et ¹³C (pour la RMN hétéronucléaire) ne contenant pas d'antibiotique est ensemençée à partir de la préculture à une densité optique de 0,3 uDO_{600 nm}/ml. La production des protéines recombinantes est induite par l'ajout d'IPTG à une concentration finale de 0,5 mM lorsque la densité optique est d'environ 0,8 uDO_{600 nm}/ml et en plaçant la culture à 20 °C pendant une nuit. Si nécessaire, le milieu M9 utilisé pour la culture est réalisé avec 50 % d'eau deutérée.

ii. Pour les cribles de co-expression chez *E. coli* BL21

Une préculture est réalisée dans un puits de Deep-Well™. Pour cela, 3 ml de milieu LB contenant les antibiotiques nécessaires sont ensemençées par des bactéries *E. coli* BL21 précédemment transformées et mis en culture à 37°C pendant une nuit. Une culture de 1,5 ml de milieu Graffinity I est ensuite ensemençée avec la préculture à une densité optique de 0,1 uDO_{600 nm}/ml puis placée à 37 °C. Une fois la densité optique à 1,2 uDO_{600 nm}/ml, la production de protéines recombinantes est induite par l'ajout d'1,5 ml de milieu Graffinity II à la culture qui est placée à 20 °C pendant une nuit.

iii. Pour les expériences de co-expression chez *E. coli*

Une culture de 200 ml de milieu LB sans antibiotique estensemencée à partir d'une préculture en milieu riche, contenant les antibiotiques nécessaires, de bactéries *E. coli* BL21 précédemment transformées, à une densité optique de 0,1 uDO_{600 nm}/ml. La production des protéines recombinantes est induite par l'ajout d'IPTG à une concentration finale de 0,5 mM lorsque la densité optique est d'environ 0,8 uDO_{600 nm}/ml et en plaçant la culture à 20 °C pendant une nuit.

iv. Pour les expériences *in vitro*

Plusieurs cultures d'un litre de milieu LB sans antibiotique sontensemencées à partir d'une préculture réalisées en milieu LB + antibiotiques adéquats de bactéries *E. coli* BL21, précédemment transformées, à une densité optique de 0,1 uDO_{600 nm}/ml. La production des protéines recombinantes est induite par l'ajout d'IPTG à une concentration finale de 0,5 mM lorsque la densité optique est d'environ 0,8 uDO_{600 nm}/ml et en plaçant la culture à 20 °C pendant une nuit.

c. Purification de protéines recombinantes chez *E. coli* BL21

i. Purification par chromatographie d'affinité sur résine TALON™

Après une centrifugation à 4 °C de 40 min à 10000 g d'une culture d'un litre, le culot cellulaire est dilué dans 50 ml de tampon de purification (voir le paragraphe I. E du chapitre Matériel et Méthodes). La lyse des bactéries est effectuée à 4 °C par ultra-sons à l'aide de 3 cycles de sonication (Branson Sonifier 250, « output » à 5 et « duty cycle » sur 50%) de 2 min entrecouper de 2 min de pause dans la glace. Une aliquote de 10 µl du sonicat total (So) est mise de côté puis le surnageant de sonication est récupéré après une centrifugation de 40 min à 15000 g et une aliquote de 10 µl du surnageant de sonication (SnSo) est prélevée. Le culot est remis en suspension dans 10 ml de tampon de purification et une aliquote de 5 µl est prélevée (Co). Si nécessaire, les acides nucléiques présents dans le surnageant de sonication sont éliminés en ajoutant de 0,0125 % à 0,1 % de PEI. Le surnageant après traitement au PEI est alors récupéré après une centrifugation de 30 min à 15000 g à 4 °C et après avoir été filtré à l'aide d'un filtre 0,22 µm. Une aliquote de 10 µl est prélevée (SnPEI). Le culot de PEI, contenant principalement des acides nucléiques, est repris dans 10 ml de tampon de purification et une aliquote de 5 µl est prélevée (CPEI). Le surnageant de PEI est incubé avec 600 µl de résine TALON™ (Clontech), préalablement lavée et remise en suspension à 50 % dans du tampon de purification, pendant 30 min sur roue à 4 °C. Après une centrifugation de 10 min à 700 g à 4 °C, le surnageant est incubé une seconde puis une troisième fois avec de la résine TALON™. Après une centrifugation de 10 min à 700 g à 4 °C, une aliquote de

10 µl du surnageant contenant les protéines non retenues sur la résine TALON™ est prélevée (NR) puis les 3 résines sont lavées. Pour cela, elles sont remises en suspension dans 10 ml de tampon de purification puis centrifugées pendant 10 min à 700 g à 4 °C. Le lavage est répété 3 fois puis une aliquote de 10 µl de chaque fixation est prélevée (F1, F2, F3). Les résines sont ensuite remises en suspension dans 1 ml de tampon de purification et incubées pendant une nuit sur roue à 4 °C avec 15 µg de PreScission (produite au laboratoire). Après une centrifugation à 4 °C de 10 min à 700 g, le surnageant contenant principalement les protéines purifiées est récupéré ainsi qu'une aliquote de 10 µl de chaque fixation (F1Cp, F2Cp, F3Cp). Les résines sont lavées 2 fois avec 1 ml de tampon de purification sur roue à 4 °C et une aliquote de 10 µl du surnageant est prélevée après une nouvelle centrifugation (CpL). Enfin, une dernière aliquote de 10 µl correspondant à la résine après coupure et après lavage est prélevée (Résine). Les différentes aliquotes sont fractionnées sur un gel de protéines dénaturant avec du bleu dénaturant 2X. Après coloration au bleu de coomassie, les fractions F1Cp, F2Cp, F3Cp et CpL contenant les protéines purifiées sont combinées puis concentrées à l'aide d'un concentrateur Amicon™ Ultra (EMB Millipore) adapté puis conservées à -20 °C. Si trois fixations ne sont pas suffisantes pour purifier l'ensemble des protéines recombinantes, des fixations supplémentaires peuvent être réalisées. Il en est de même pour le nombre de lavage après coupure à la PreScission s'il reste des protéines recombinantes sur la résine (fraction résine).

Le même protocole est utilisé pour la purification de complexes de protéines recombinantes pour les expériences de co-expression chez *E. coli*, excepté qu'une seule fixation a été réalisée et que les protéines ne sont pas éluées.

ii. Purification par chromatographie d'exclusion stérique

La purification est réalisée sur un appareil Äkta Prime™ de chez GE Healthcare Life Sciences. Suivant la taille des protéines à purifier, une colonne s200 ou s75 (GE Healthcare Life Sciences) de 120 ml est équilibrée dans le tampon souhaité. Un volume maximum de 5 ml de protéines à purifiées est injecté dans une boucle d'un volume au moins 2 fois supérieur, puis chargé sur la colonne à raison d'1 ml/min. Des fractions de 2 ml sont alors collectées. Des aliquotes de 10 µl des fractions intéressantes sont prélevées puis fractionnées sur un gel de protéines dénaturant avec du bleu dénaturant 2X. Une aliquote correspondant à l'échantillon avant injection (NI) est également fractionnée sur le gel. Les fractions qui contiennent les protéines d'intérêts sont combinées puis concentrées à l'aide d'un concentrateur Amicon™ Ultra (EMB Millipore) adapté puis conservées à

-20 °C.

iii. Purification des cribles de co-expressions à l'aide du robot TECAN™ par chromatographie d'affinité

Le crible d'interaction protéine-protéine est effectué à partir de cultures réalisées en Deep-Weel™ 24 puits et à l'aide d'un robot TECAN™. Les tests d'interaction sont réalisés en présence de deux concentrations différentes de sel (50 et 400 mM). Les 3 ml de culture (voir le paragraphe II. B. 8. b.ii précédent) sont divisés en 2 dans 2 Deep-Weel™ 24 puits par un robot TECAN™, puis les cultures sont centrifugés pendant 20 min à 4000 g à 4 °C. Le surnageant est retiré et le culot est repris dans 1,5 ml de tampon haut sel (pour le premier deep-weel™) et bas sel (pour le second deep-weel™) par le robot. Après une sonication d'1 min (Utracell™ Bioblock Scientific, « duty cycle » 50 %), une aliquote de 10 µl correspondant à l'extrait protéique total (So) est prélevée par le robot et ajoutée à 10 µl de bleu dénaturant 2X. Après une centrifugation de 30 min à 4000 g, une aliquote de 10 µl du surnageant, correspondant aux protéines solubles (SnSo), est prélevée et ajoutée à 10 µl de bleu dénaturant 2X. Le surnageant est incubé avec 50 µl de billes magnétiques sur lesquelles sont coordonnées des atomes de nickel (Dynabeads™ His-tag, ThermoFischer Scientific) pendant 30 min dans un Deep-Weel™ 48 puits. Le Deep-well™ est ensuite placé sur un aimant par le robot et le surnageant est retiré et les billes sont lavées 3 fois avec 1,5 ml de tampon haut ou bas sel avant d'être remises en suspension dans 20 µl de bleu dénaturant 2X par le robot TECAN™. Dix microlitres des aliquotes (So et SnSo) et 10 µl des billes magnétiques sont finalement fractionnées sur un gel de protéines dénaturant.

iv. Purification par chromatographie d'affinité à haute pression

Les cultures sont centrifugées pendant 50 min à 4000 g puis les culots sont repris dans du tampon de purification à raison de 30 ml par litre de culture. La lyse des bactéries est effectuée par ajout de 400 U de Benzonase™ (Merk Millipore) et par ultra-sons (voir le paragraphe II. B. 8. c. i ci-dessus). Une aliquote de 10 µl des protéines totales (So) est prélevée puis les extraits sont centrifugés pendant 30 min à 15000 g. Une aliquote de 10 µl du surnageant et du culot, préalablement repris dans 10 ml de tampon de purification, sont prélevées puis le surnageant est filtré à l'aide d'un filtre 0,22 µm avant d'être injecté par l'Äkta Avant™ (GE Healthcare Life Sciences) sur une colonne de nickel de 5 ml préalablement équilibrée avec le tampon de purification. La colonne est ensuite lavée avec le tampon de purification puis les protéines sont éluées par le tampon d'éluion et collectées dans des fractions de 5 ml. Une aliquote de 10 µl de chaque fraction est prélevée puis l'ensemble

des aliquotes est fractionné sur un gel de protéines dénaturant avec du bleu dénaturant 2X. Après une coloration du gel au bleu de coomassie, les fractions qui contiennent les protéines d'intérêts sont combinées puis concentrées à l'aide d'un concentrateur Amicon™ Ultra (EMB Millipore) adapté et congelées à -20 °C.

9. Production et purification de protéines recombinantes par baculovirus

a. Obtention d'un virus recombinant P0

Deux millions de cellules Sf9 sontensemencées dans une flasque T25 et laissées adhérer pendant 1 h à 27 °C. Deux microgrammes de vecteur pAC8 sont co-transfectés dans les cellules Sf9 avec 500 ng de bacmide linéarisé par l'enzyme *Bsu36I*. Pour cela, 2 µg (500 ng/µl) de pAC8 sont incubés avec 500 ng de bacmide linéarisé (100 ng/µl) et 750 µl de tampon de transfection HEPES puis ajoutés goutte-à-goutte dans une solution contenant 680 µl de milieu Grace non supplémenté et 75 µl de SVF. Après une incubation de 15 min à température ambiante, le milieu de transfection est ajouté sur les cellules Sf9 dont le milieu de culture aura été préalablement retiré. Après une incubation de 4 h à 27 °C, le milieu de transfection est retiré et 5 ml de milieu Grace supplémenté avec 10 % de SVF sont ajoutés sur les cellules qui sont incubées pendant au moins 7 jours à 27 °C. Les cellules sont ensuite collectées puis centrifugées pendant 10 min à 2000 g. Le surnageant, correspondant au P0, est alors conservé à 4 °C à l'abri de la lumière entre 6 et 12 mois.

b. Amplification du virus recombinant en flasque T175

À partir d'une culture liquide de cellule Sf9, 20 millions de cellules (à une concentration minimale de 1 million de cellules/ml) sontensemencés en flasque T175 et laissés adhérer pendant 1 h à 27 °C. Le milieu de culture est retiré, tout en laissant au moins 6 ml de milieu, et 100 µl de virus P0 sont ajoutés sur les cellules. Après une incubation d'une heure à 27 °C, 19 ml de milieu Grace sont ajoutés aux cellules qui sont incubés pendant une semaine à 27 °C. Les cellules sont ensuite collectées puis centrifugées pendant 10 min à 2000 g. Le surnageant, correspondant au P1, est alors conservé à 4 °C.

c. Amplification du virus en plaque 6 puits

À partir d'une culture liquide de cellule Sf9, 1 million de cellules (à une concentration minimale de 1 million de cellule/ml) estensemencé dans un puits d'une plaque 6 puits et laissé adhérer 20 min à 27 °C. Le milieu de culture est retiré et 200 µl de virus P0, P1 ou P2 sont ajoutés sur les cellules. Après une incubation de 5 min sous hotte, 3,5 ml de milieu Grace sont ajoutés aux cellules qui sont

incubés pendant 4 jours à 27 °C. Les cellules sont ensuite collectées puis centrifugées pendant 10 min à 2000 g. Le surnageant, correspondant respectivement au P1, P2 ou P3 est alors conservé à 4 °C.

d. Production de protéines recombinantes par infection de cellules Sf9

Quatre cents millions de cellules Sf9, ayant une viabilité d'au moins 95 %, sont infectés par 19 ml de virus recombinant P1 (obtenu en flasque T175) ou par 7 ml de virus recombinant P3 dans un volume final de 200 ml. Après 72 h, les cellules sont centrifugées 10 min à 3000 g puis le culot cellulaire est conservé à -80 °C.

e. Purification des protéines recombinantes

Le culot de cellule est dilué dans 10 ml de tampon de lyse puis les cellules sont lysées par choc osmotique à l'aide d'un homogénéiseur de Dounce et une aliquote de 10 µl correspondant aux protéines totales (So) est récupérée. Après une centrifugation de 20 min à 20000 g, une aliquote de 10 µl du surnageant (SnSo), contenant les protéines solubles est prélevée. Il en est de même pour le culot (Co). Le surnageant est alors incubé pendant 40 min avec 500 µl de résine TALON™ pure (Clontech) sur roue à 4 °C. Après une centrifugation de 10 min à 700 g à 4 °C, une aliquote de 10 µl du surnageant contenant les protéines non retenues sur la résine TALON™ est prélevée (NR) puis la résine est lavée. Pour cela, elle est remise en suspension dans 10 ml de tampon de lavage puis centrifugées pendant 10 min à 700 g à 4 °C. Le lavage est répété 4 fois puis une aliquote de 10 µl de la résine est prélevée (F1). La résine est ensuite remise en suspension dans 1,5 ml de tampon d'élution II et les protéines sont éluées pendant 40 min à 4 °C sur roue puis récupérées par une centrifugation de 10 min à 700 g à 4 °C. Une aliquote de 10 µl est alors récupérée (E1). Une seconde élution est ensuite réalisée puis une aliquote de 10 µl de la résine après élution (Résine) est prélevée. Les diverses aliquotes sont ensuite fractionnées sur un gel de protéines dénaturant avec 10 µl de bleu dénaturant 2X. Les fractions contenant les protéines recombinantes sont ensuite combinées et dialysées une nuit dans 2 litres de tampon de lavage. Pour la purification de TRBP, 15 µg de PreScission (produite au laboratoire) sont ajoutés dans le boudin de dialyse. Le contenu du boudin de dialyse est ensuite récupéré puis concentrée à l'aide d'un concentrateur adapté et conservé à -20 °C.

10. Expériences RMN

L'attribution des spectres RMN a été réalisée à l'aide du logiciel CARA. Les expériences de RMN ont été réalisées par le Dr Marc Quinternet de la fédération de Recherche FR3209 sur un spectrophotomètre RMN Bruker Avance III à haut champ de 600 MHz, équipé d'une cryosonde.

C. Méthodes relatives à la culture de cellules humaines

1. Décongélation de cellules HeLa/HEK

Dans une boîte de pétri de 10 cm, un tube de cellule HeLa stock est décongelé rapidement et ajouté à 10 ml de milieu DMEM contenant 10 % de SVF, de la glutamine à une concentration finale de 20 mM et 2 antibiotiques : pénicilline 1000 U/ml finale et streptomycine 1000 µg/ml finale.

2. Amplification de cellules HeLa/HEK

Lorsque les cellules arrivent à confluence, le milieu de culture est aspiré. La boîte est rincée avec 1 ml de trypsine-EDTA 1X pour éliminer le reste de milieu. Après aspiration, 1 ml de la même solution est ajouté et les cellules sont incubées à 37 °C pendant au moins 5 min. Enfin, 10 ml de milieu de culture (voir ci-dessus) sont ajoutés pour arrêter l'action de la trypsine. Les cellules sont homogénéisées et individualisées par plusieurs aspirations et refoulements successifs. Dans une nouvelle boîte de pétri, 2 ml de cellules trypsinisées sont ajoutés à 8 ml du même milieu, puis l'ensemble est incubé pendant 2 jours à 37 °C.

3. Congélation de cellules HeLa/HEK

Comme précédemment, les cellules sont trypsinisées et remise en suspension dans 10 ml de milieu DMEM additionné de 10 % de SVF par flasque. Le contenu de chaque flasque est rassemblé et les cellules sont comptées à l'aide d'un compteur Countess™ (Initrogen). Les cellules sont ensuite centrifugées à 2000 g pendant 5 min puis le culot cellulaire est repris dans un volume de DMEM contenant 20 % de SVF et 10 % de DMSO, de façon à ce que la concentration finale soit égale à 5 millions de cellules par millilitre. Les cellules sont aliquotées par fractions de 1 ml dans des tubes Eppendorf à congélation puis placées dans une boîte de congélation. La boîte est ensuite placée à -80 °C pendant au moins 48 h, puis les tubes de cellules sont conservés dans l'azote liquide.

D. Méthodes relatives à la culture de cellule d'insectes

1. Amplification de cellules Sf9

Dans une flasque T75, 15 ml de milieu Grace supplémenté avec 10 % de SVF sontensemencés avec 1,5 millions de cellules Sf9. Après 6 jours d'incubation à 27 °C, les cellules sont récoltées puis comptées à l'aide d'un compteur Countess™ (Initrogen) et de nouveau amplifiées en flasque T75.

2. Culture liquide de cellules Sf9

Deux cents millilitres de milieu Grace supplémenté avec 10 % de SVF sontensemencés avec 100 millions de cellules Sf9 ayant une viabilité d'au moins 95 %. Arrivée à une concentration de 2 millions de cellules/ml, la culture est de nouveau diluée à 0,5 millions/ml. Seuls 3 ou 4 passages peuvent être effectués.

Bibliographie

ABEL Y., CLERGET G., BOURGUIGNON-IGEL V., SALONE V., REDERSTORFF M. [Beyond usual functions of snoRNAs]. *Medecine Sciences: M/S*, 2014, **30** : 297-302.

AITTALEB M., RASHID R., CHEN Q., PALMER J.R., DANIELS C.J., LI H. Structure and function of archaeal box C/D sRNP core proteins. *Nature Structural Biology*, 2003, **10** : 256-263.

ALI M.M.U., ROE S.M., VAUGHAN C.K., MEYER P., PANARETOU B., PIPER P.W., PRODROMOU C., PEARL L.H. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature*, 2006, **440** : 1013-1017.

ALLMANG C., CARBON P., KROL A. The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. *RNA*, 2002, **8** : 1308-1318.

ALLMANG C., KUFEL J., CHANFREAU G., MITCHELL P., PETFALSKI E., TOLLERVEY D. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *The EMBO journal*, 1999, **18** : 5399-5410.

ALTIERI D.C., STEIN G.S., LIAN J.B., LANGUINO L.R. TRAP-1, the mitochondrial Hsp90. *Biochimica Et Biophysica Acta*, 2012, **1823** : 767-773.

AMEYAR-ZAZOUA M., RACHEZ C., SOUIDI M., ROBIN P., FRITSCH L., YOUNG R., MORZOVA N., FENOUIL R., DESCOSTES N., ANDRAU J.-C., MATHIEU J., HAMICHE A., AIT-SI-ALI S., MUCHARDT C., BATSCHÉ E., HAREL-BELLAN A. Argonaute proteins couple chromatin silencing to alternative splicing. *Nature Structural & Molecular Biology*, 2012, **19** : 998-1004.

ANGRISANI A., TURANO M., PAPARO L., DI MAURO C., FURIA M. A new human dyskerin isoform with cytoplasmic localization. *Biochimica Et Biophysica Acta*, 2011, **1810** : 1361-1368.

ARIS J.P., BLOBEL G. cDNA cloning and sequencing of human fibrillarin, a conserved nucleolar protein recognized by autoimmune antisera. *Proceedings of the National Academy of Sciences of the United States of America*, 1991, **88** : 931-935.

ATZORN V., FRAGAPANE P., KISS T. U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. *Molecular and Cellular Biology*, 2004, **24** : 1769-1778.

AZUMA-MUKAI A., OGURI H., MITUYAMA T., QIAN Z.R., ASAI K., SIOMI H., SIOMI M.C. Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105** : 7964-7969.

BACK R., DOMINGUEZ C., ROTHÉ B., BOBO C., BEAUFILS C., MORÉRA S., MEYER P., CHARPENTIER B., BRANLANT C., ALLAIN F.H.-T., MANIVAL X. High-resolution structural analysis shows how Tah1 tethers Hsp90 to the R2TP complex. *Structure (London, England: 1993)*, 2013, **21** : 1834-1847.

BAI B., LIU H., LAIHO M. Small RNA expression and deep sequencing analyses of the nucleolus reveal the presence of nucleolus-associated microRNAs. *FEBS Open Bio*, 2014, **4** : 441-449.

BAI B., YEGNASUBRAMANIAN S., WHEELAN S.J., LAIHO M. RNA-Seq of the nucleolus reveals abundant SNORD44-derived small RNAs. *PLoS One*, 2014, **9** : e107519.

BARDONI B., SCHENCK A., MANDEL J.L. A novel RNA-binding nuclear protein that interacts with the fragile X mental retardation (FMR1) protein. *Human Molecular Genetics*, 1999, **8** : 2557-2566.

BARTEL D.P. MicroRNAs: target recognition and regulatory functions. *Cell*, 2009, **136** : 215-233.

BATTISTI P.-L., DAHER A., BANNWARTH S., VOORTMAN J., PEDEN K.W.C., HISCOTT J., MOULAND A.J., BENAROUS R., GATIGNOL A. Additive activity between the trans-activation response RNA-binding protein, TRBP2, and cyclin T1 on HIV type 1 expression and viral production in murine cells. *AIDS research and human retroviruses*, 2003, **19** : 767-778.

BENKIRANE M., NEUVEUT C., CHUN R.F., SMITH S.M., SAMUEL C.E., GATIGNOL A., JEANG K.T. Oncogenic potential of TAR RNA binding protein TRBP and its regulatory interaction with RNA-dependent protein kinase PKR. *The EMBO journal*, 1997, **16** : 611-624.

BERNDT H., HARNISCH C., RAMMELT C., STÖHR N., ZIRKEL A., DOHM J.C., HIMMELBAUER H., TAVANEZ J.-P., HÜTTELMAIER S., WAHLE E. Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA*, 2012, **18** : 958-972.

BERNSTEIN E., CAUDY A.A., HAMMOND S.M., HANNON G.J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 2001, **409** : 363-366.

BETANCUR J.G., TOMARI Y. Dicer is dispensable for asymmetric RISC loading in mammals. *RNA (New York, N.Y.)*, 2012, **18** : 24-30.

BÉTHUNE J., ARTUS-REVEL C.G., FILIPOWICZ W. Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Reports*, 2012, **13** : 716-723.

BIETH E., EDDIRY S., GASTON V., LORENZINI F., BUFFET A., CONTE AURIOL F., MOLINAS C., CAILLEY D., ROORYCK C., ARVEILER B., CAVAILLÉ J., SALLES J.P., TAUBER M. Highly restricted deletion of the SNORD116 region is implicated in Prader-Willi Syndrome. *European journal of human genetics: EJHG*, 2015, **23** : 252-255.

- BIZARRO J., CHARRON C., BOULON S., WESTMAN B., PRADET-BALADE B., VANDERMOERE F., CHAGOT M.-E., HALLAIS M., AHMAD Y., LEONHARDT H., LAMOND A., MANIVAL X., BRANLANT C., CHARPENTIER B., VERHEGGEN C., BERTRAND E. Proteomic and 3D structure analyses highlight the C/D box snoRNP assembly mechanism and its control. *The Journal of Cell Biology*, 2014, **207** : 463-480.
- BLOW M.J., GROCOCK R.J., VAN DONGEN S., ENRIGHT A.J., DICKS E., FUTREAL P.A., WOOSTER R., STRATTON M.R. RNA editing of human microRNAs. *Genome Biology*, 2006, **7** : R27.
- BOGERD H.P., KARNOWSKI H.W., CAI X., SHIN J., POHLERS M., CULLEN B.R. A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs. *Molecular Cell*, 2010, **37** : 135-142.
- BORCHERT G.M., GILMORE B.L., SPENGLER R.M., XING Y., LANIER W., BHATTACHARYA D., DAVIDSON B.L. Adenosine deamination in human transcripts generates novel microRNA binding sites. *Human Molecular Genetics*, 2009, **18** : 4801-4807.
- BORTOLIN M.L., GANOT P., KISS T. Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *The EMBO journal*, 1999, **18** : 457-469.
- BORTOLIN-CAVAILLÉ M.-L., CAVAILLÉ J. The SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) gene clusters at the imprinted Prader-Willi locus generate canonical box C/D snoRNAs. *Nucleic Acids Research*, 2012, **40** : 6800-6807.
- BOULON S., MARMIER-GOURRIER N., PRADET-BALADE B., WURTH L., VERHEGGEN C., JADY B.E., ROTHE B., PESCIA C., ROBERT M.-C., KISS T., BARDONI B., KROL A., BRANLANT C., ALLMANG C., BERTRAND E., CHARPENTIER B. The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *The Journal of Cell Biology*, 2008, **180** : 579-595.
- BOULON S., PRADET-BALADE B., VERHEGGEN C., MOLLE D., BOIREAU S., GEORGIEVA M., AZZAG K., ROBERT M.-C., AHMAD Y., NEEL H., LAMOND A.I., BERTRAND E. HSP90 and Its R2TP/Prefoldin-like Cochaperone Are Involved in the Cytoplasmic Assembly of RNA Polymerase II. *Molecular cell*, 2010, **39** : 912-924.
- BOUSQUET-ANTONELLI C., HENRY Y., G'ELUGNE J.P., CAIZERGUES-FERRER M., KISS T. A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs. *The EMBO journal*, 1997, **16** : 4770-4776.

BRAGANTINI B., TIOTIU D., ROTHÉ B., SALIOU J.-M., MARTY H., CIANFÉRANI S., CHARPENTIER B., QUINTERNET M., MANIVAL X. Functional and Structural Insights of the Zinc-Finger HIT protein family members Involved in Box C/D snoRNP Biogenesis. *Journal of Molecular Biology*, 2016, **428** : 2488-2506.

BRAMEIER M., HERWIG A., REINHARDT R., WALTER L., GRUBER J. Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. *Nucleic Acids Research*, 2011, **39** : 675-686.

BRANDIS K.A., GALE S., JINN S., LANGMADE S.J., DUDLEY-RUCKER N., JIANG H., SIDHU R., REN A., GOLDBERG A., SCHAFFER J.E., ORY D.S. Box C/D Small Nucleolar RNA (snoRNA) U60 Regulates Intracellular Cholesterol Trafficking. *Journal of Biological Chemistry*, 2013, **288** : 35703-35713.

BRAUN J.E., HUNTZINGER E., FAUSER M., IZAURRALDE E. GW182 Proteins Directly Recruit Cytoplasmic Deadenylation Complexes to miRNA Targets. *Molecular Cell*, 2011, **44** : 120-133.

BRAUN J.E., HUNTZINGER E., IZAURRALDE E. The role of GW182 proteins in miRNA-mediated gene silencing. *Advances in Experimental Medicine and Biology*, 2013, **768** : 147-163.

BRENNER S., JACOB F., MESELSON M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, 1961, **190** : 576-581.

BURNS C.M., CHU H., RUETER S.M., HUTCHINSON L.K., CANTON H., SANDERS-BUSH E., EMESON R.B. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature*, 1997, **387** : 303-308.

BUTCHER S.E., BROW D.A. Towards understanding the catalytic core structure of the spliceosome. *Biochemical Society Transactions*, 2005, **33** : 447-449.

CARRETERO-PAULET L., ALBERT V.A., FARES M.A. Molecular evolutionary mechanisms driving functional diversification of the HSP90A family of heat shock proteins in eukaryotes. *Molecular Biology and Evolution*, 2013, **30** : 2035-2043.

CERRUDO C.S., MENGUAL GÓMEZ D.L., GÓMEZ D.E., GHIRINGHELLI P.D. Novel insights into the evolution and structural characterization of dyskerin using comprehensive bioinformatics analysis. *Journal of Proteome Research*, 2015, **14** : 874-887.

CESANA M., CACCHIARELLI D., LEGNINI I., SANTINI T., STHANDIER O., CHINAPPI M., TRAMONTANO A., BOZZONI I. A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endogenous RNA. *Cell*, 2011, **147** : 358-369.

CHARPENTIER B., MULLER S., BRANLANT C. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Research*, 2005, **33** : 3133-3144.

CHELOUFI S., DOS SANTOS C.O., CHONG M.M.W., HANNON G.J. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature*, 2010, **465** : 584-589.

CHEN K., RAJEWSKY N. Natural selection on human microRNA binding sites inferred from SNP data. *Nature Genetics*, 2006, **38** : 1452-1456.

CHENDRIMADA T.P., FINN K.J., JI X., BAILLAT D., GREGORY R.I., LIEBHABER S.A., PASQUINELLI A.E., SHIEKHATTAR R. MicroRNA silencing through RISC recruitment of eIF6. *Nature*, 2007, **447** : 823-828.

CHENDRIMADA T.P., GREGORY R.I., KUMARASWAMY E., NORMAN J., COOCH N., NISHIKURA K., SHIEKHATTAR R. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, 2005, **436** : 740-744.

CHO S., PARK J.S., KANG Y.-K. AGO2 and SETDB1 cooperate in promoter-targeted transcriptional silencing of the androgen receptor gene. *Nucleic Acids Research*, 2014, **42** : 13545-13556.

CLERGET G., ABEL Y., REDERSTORFF M. Small non-coding RNAs: a quick look in the rearview mirror. *Methods in Molecular Biology (Clifton, N.J.)*, 2015, **1296** : 3-9.

CLÉRY A., SENTY-SÉGAULT V., LECLERC F., RAUÉ H.A., BRANLANT C. Analysis of sequence and structural features that identify the B/C motif of U3 small nucleolar RNA as the recognition site for the Snu13p-Rrp9p protein pair. *Molecular and Cellular Biology*, 2007, **27** : 1191-1206.

CLOONAN N., WANI S., XU Q., GU J., LEA K., HEATER S., BARBACIORU C., STEPTOE A.L., MARTIN H.C., NOURBAKHS E., KRISHNAN K., GARDINER B., WANG X., NONES K., STEEN J.A., MATIGIAN N.A., WOOD D.L., KASSAHN K.S., WADDELL N., SHEPHERD J., LEE C., ICHIKAWA J., MCKERNAN K., BRAMLETT K., KUERSTEN S., GRIMMOND S.M. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biology*, 2011, **12** : R126.

COLE C., SOBALA A., LU C., THATCHER S.R., BOWMAN A., BROWN J.W.S., GREEN P.J., BARTON G.J., HUTVAGNER G. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA (New York, N.Y.)*, 2009, **15** : 2147-2160.

DAHER A., LARAKI G., SINGH M., MELENDEZ-PEÑA C.E., BANNWARTH S., PETERS A.H.F.M., MEURS E.F., BRAUN R.E., PATEL R.C., GATIGNOL A. TRBP control of PACT-induced phosphorylation of protein kinase R is reversed by stress. *Molecular and Cellular Biology*, 2009, **29** : 254-265.

DAHER A., LONGUET M., DORIN D., BOIS F., SEGERAL E., BANNWARTH S., BATTISTI P.L., PURCELL D.F., BENAROUS R., VAQUERO C., MEURS E.F., GATIGNOL A. Two dimerization domains in the trans-activation response RNA-binding protein (TRBP) individually reverse the protein kinase R inhibition of HIV-1 long terminal repeat expression. *The Journal of Biological Chemistry*, 2001, **276** : 33899-33905.

DAI L., CHEN K., YOUNGREN B., KULINA J., YANG A., GUO Z., LI J., YU P., GU S. Cytoplasmic Drosha activity generated by alternative splicing. *Nucleic Acids Research*, 2016 : gkw668.

DANIELS S.M., MELENDEZ-PEÑA C.E., SCARBOROUGH R.J., DAHER A., CHRISTENSEN H.S., EL FAR M., PURCELL D.F., LAINÉ S., GATIGNOL A. Characterization of the TRBP domain required for Dicer interaction and function in RNA interference. *BMC Molecular Biology*, 2009, **10** : 38.

DARZACQ X., KITTUR N., ROY S., SHAV-TAL Y., SINGER R.H., MEIER U.T. Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *The Journal of Cell Biology*, 2006, **173** : 207-218.

DAVIS F.F., ALLEN F.W. Ribonucleic acids from yeast which contain a fifth nucleotide. *The Journal of Biological Chemistry*, 1957, **227** : 907-915.

DELGOFFE G.M., KOLE T.P., COTTER R.J., POWELL J.D. Enhanced interaction between Hsp90 and raptor regulates mTOR signaling upon T cell activation. *Molecular Immunology*, 2009, **46** : 2694-2698.

DESCHAMPS-FRANCOEUR G., GARNEAU D., DUPUIS-SANDOVAL F., ROY A., FRAPPIER M., CATALA M., COUTURE S., BARBE-MARCOUX M., ABOU-ELELA S., SCOTT M.S. Identification of discrete classes of small nucleolar RNA featuring different ends and RNA binding protein dependency. *Nucleic Acids Research*, 2014, **42** : 10073-10085.

DEZ C., NOAILLAC-DEPEYRE J., CAIZERGUES-FERRER M., HENRY Y. Naf1p, an Essential Nucleoplasmic Factor Specifically Required for Accumulation of Box H/ACA Small Nucleolar RNPs. *Molecular and Cellular Biology*, 2002, **22** : 7053-7065.

- DI GIAMMARTINO D.C., NISHIDA K., MANLEY J.L. Mechanisms and Consequences of Alternative Polyadenylation. *Molecular Cell*, 2011, **43** : 853-866.
- DOENCH J.G., SHARP P.A. Specificity of microRNA target selection in translational repression. *Genes & Development*, 2004, **18** : 504-511.
- DONG X.-Y., GUO P., BOYD J., SUN X., LI Q., ZHOU W., DONG J.-T. Implication of snoRNA U50 in human breast cancer. *Journal of Genetics and Genomics = Yi Chuan Xue Bao*, 2009, **36** : 447-454.
- DONG X.-Y., RODRIGUEZ C., GUO P., SUN X., TALBOT J.T., ZHOU W., PETROS J., LI Q., VESSELLA R.L., KIBEL A.S., STEVENS V.L., CALLE E.E., DONG J.-T. SnoRNA U50 is a candidate tumor suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. *Human molecular genetics*, 2008, **17** : 1031-1042.
- DORIN D., BONNET M.C., BANNWARTH S., GATIGNOL A., MEURS E.F., VAQUERO C. The TAR RNA-binding protein, TRBP, stimulates the expression of TAR-containing RNAs in vitro and in vivo independently of its ability to inhibit the dsRNA-dependent kinase PKR. *The Journal of Biological Chemistry*, 2003, **278** : 4440-4448.
- DOYLE M., BADERTSCHER L., JASKIEWICZ L., GÜTTINGER S., JURADO S., HUGENSCHMIDT T., KUTAY U., FILIPOWICZ W. The double-stranded RNA binding domain of human Dicer functions as a nuclear localization signal. *RNA (New York, N.Y.)*, 2013, **19** : 1238-1252.
- DRAGON F., POGACIĆ V., FILIPOWICZ W. In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs. *Molecular and Cellular Biology*, 2000, **20** : 3037-3048.
- DUAN J., LI L., LU J., WANG W., YE K. Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Molecular Cell*, 2009, **34** : 427-439.
- DUGAN K.A., WOOD M.A., COLE M.D. TIP49, but not TRRAP, modulates c-Myc and E2F1 dependent apoptosis. *Oncogene*, 2002, **21** : 5835-5843.
- DUURSMA A.M., KEDDE M., SCHRIER M., SAGE C. LE, AGAMI R. miR-148 targets human DNMT3b protein coding region. *RNA*, 2008, **14** : 872-877.
- ECKERT K., SALIOU J.-M., MONLEZUN L., VIGOUROUX A., ATMANE N., CAILLAT C., QUEVILLON-CHÉRUEL S., MADIONA K., NICAISE M., LAZEREG S., VAN DORSSELAER A., SANGLIER-CIANFÉRANI S., MEYER P., MORÉRA S. The Pih1-Tah1 cochaperone complex inhibits Hsp90 molecular chaperone ATPase activity. *The Journal of Biological Chemistry*, 2010, **285** : 31304-31312.

- EICHHORN S.W., GUO H., MCGEARY S.E., RODRIGUEZ-MIAS R.A., SHIN C., BAEK D., HSU S.-H., GHOSHAL K., VILLÉN J., BARTEL D.P. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Molecular Cell*, 2014, **56** : 104-115.
- ENDER C., KREK A., FRIEDLÄNDER M.R., BEITZINGER M., WEINMANN L., CHEN W., PFEFFER S., RAJEWSKY N., MEISTER G. A human snoRNA with microRNA-like functions. *Molecular Cell*, 2008, **32** : 519-528.
- EULALIO A., BEHM-ANSMANT I., SCHWEIZER D., IZAURRALDE E. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Molecular and Cellular Biology*, 2007, **27** : 3970-3981.
- EULALIO A., HUNTZINGER E., IZAURRALDE E. Getting to the Root of miRNA-Mediated Gene Silencing. *Cell*, 2008, **132** : 9-14.
- EYSTATHIOY T., CHAN E.K.L., TENENBAUM S.A., KEENE J.D., GRIFFITH K., FRITZLER M.J. A phosphorylated cytoplasmic autoantigen, GW182, associates with a unique population of human mRNAs within novel cytoplasmic speckles. *Molecular Biology of the Cell*, 2002, **13** : 1338-1351.
- FABIAN M.R., MATHONNET G., SUNDERMEIER T., MATHYS H., ZIPPRICH J.T., SVITKIN Y.V., RIVAS F., JINEK M., WOHLSCHEGEL J., DOUDNA J.A., CHEN C.-Y.A., SHYU A.-B., YATES J.R., HANNON G.J., FILIPOWICZ W., DUCHAINE T.F., SONENBERG N. Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Molecular Cell*, 2009, **35** : 868-880.
- FAEHNLE C.R., ELKAYAM E., HAASE A.D., HANNON G.J., JOSHUA-TOR L. The Making of a Slicer: Activation of Human Argonaute-1. *Cell reports*, 2013, **3** : 1901-1909.
- FALALEEVA M., PAGES A., MATUSZEK Z., HIDMI S., AGRANAT-TAMIR L., KOROTKOV K., NEVO Y., EYRAS E., SPERLING R., STAMM S. Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. *Proceedings of the National Academy of Sciences*, 2016, **113** : E1625-E1634.
- FATICA A., DLAKIĆ M., TOLLERVEY D. Naf1 p is a box H/ACA snoRNP assembly factor. *RNA (New York, N.Y.)*, 2002, **8** : 1502-1514.
- FENG Y., LEE N., FEARON E.R. TIP49 regulates beta-catenin-mediated neoplastic transformation and T-cell factor target gene induction via effects on chromatin remodeling. *Cancer Research*, 2003, **63** : 8726-8734.

- FLEMR M., MALIK R., FRANKE V., NEJEPINSKA J., SEDLACEK R., VLAHOVICEK K., SVOBODA P. A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. *Cell*, 2013, **155** : 807-816.
- FORMAN J.J., LEGESSE-MILLER A., COLLER H.A. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proceedings of the National Academy of Sciences*, 2008, **105** : 14879-14884.
- FRANCIA S., MICHELINI F., SAXENA A., TANG D., DE HOON M., ANELLI V., MIONE M., CARNINCI P., D'ADDA DI FAGAGNA F. Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature*, 2012, **488** : 231-235.
- FRANK F., SONENBERG N., NAGAR B. Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature*, 2010, **465** : 818-822.
- GAGNON K.T., LI L., CHU Y., JANOWSKI B.A., COREY D.R. RNAi Factors are Present and Active in Human Cell Nuclei. *Cell reports*, 2014, **6** : 211-221.
- GANOT P., BORTOLIN M.-L., KISS T. Site-Specific Pseudouridine Formation in Preribosomal RNA Is Guided by Small Nucleolar RNAs. *Cell*, 1997, **89** : 799-809.
- GAO M., WEI W., LI M.-M., WU Y.-S., BA Z., JIN K.-X., LI M.-M., LIAO Y.-Q., ADHIKARI S., CHONG Z., ZHANG T., GUO C.-X., TANG T.-S., ZHU B.-T., XU X.-Z., MAILAND N., YANG Y.-G., QI Y., RENDTLEW DANIELSEN J.M. Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. *Cell Research*, 2014, **24** : 532-541.
- GARCÍA M.A., GIL J., VENTOSO I., GUERRA S., DOMINGO E., RIVAS C., ESTEBAN M. Impact of Protein Kinase PKR in Cell Biology: from Antiviral to Antiproliferative Action. *Microbiology and Molecular Biology Reviews*, 2006, **70** : 1032-1060.
- GATIGNOL A., BUCKLER-WHITE A., BERKHOUT B., JEANG K.T. Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. *Science (New York, N.Y.)*, 1991, **251** : 1597-1600.
- GAUTIER T., BERGÈS T., TOLLERVEY D., HURT E. Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Molecular and Cellular Biology*, 1997, **17** : 7088-7098.
- GIRARD J.P., LEHTONEN H., CAIZERGUES-FERRER M., AMALRIC F., TOLLERVEY D., LAPEYRE B. GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *The EMBO journal*, 1992, **11** : 673-682.

- GONZALES F.A., ZANCHIN N.I.T., LUZ J.S., OLIVEIRA C.C. Characterization of *Saccharomyces cerevisiae* Nop17p, a novel Nop58p-interacting protein that is involved in Pre-rRNA processing. *Journal of Molecular Biology*, 2005, **346** : 437-455.
- GOODY T.A., MELCHER S.E., NORMAN D.G., LILLEY D.M.J. The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA*, 2004, **10** : 254-264.
- GORYNIA S., BANDEIRAS T.M., PINHO F.G., MCVEY C.E., VONRHEIN C., ROUND A., SVERGUN D.I., DONNER P., MATIAS P.M., CARRONDO M.A. Structural and functional insights into a dodecameric molecular machine – The RuvBL1/RuvBL2 complex. *Journal of Structural Biology*, 2011, **176** : 279-291.
- GRANATO D.C., GONZALES F.A., LUZ J.S., CASSIOLA F., MACHADO-SANTELLI G.M., OLIVEIRA C.C. Nop53p, an essential nucleolar protein that interacts with Nop17p and Nip7p, is required for pre-rRNA processing in *Saccharomyces cerevisiae*. *The FEBS journal*, 2005, **272** : 4450-4463.
- GREGORY R.I., YAN K.-P., AMUTHAN G., CHENDRIMADA T., DORATOTAJ B., COOCH N., SHIEKHATTAR R. The Microprocessor complex mediates the genesis of microRNAs. *Nature*, 2004, **432** : 235-240.
- GRIFFITHS-JONES S., GROCOCK R.J., VAN DONGEN S., BATEMAN A., ENRIGHT A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, 2006, **34** : D140-144.
- GROZDANOV P.N., ROY S., KITTUR N., MEIER U.T. SHQ1 is required prior to NAF1 for assembly of H/ACA small nucleolar and telomerase RNPs. *RNA (New York, N.Y.)*, 2009, **15** : 1188-1197.
- GU S., JIN L., ZHANG F., SARNOW P., KAY M.A. Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nature Structural & Molecular Biology*, 2009, **16** : 144-150.
- GU S., JIN L., ZHANG Y., HUANG Y., ZHANG F., VALDMANIS P.N., KAY M.A. The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. *Cell*, 2012, **151** : 900-911.
- GUPTA V., HUANG X., PATEL R.C. The carboxy-terminal, M3 motifs of PACT and TRBP have opposite effects on PKR activity. *Virology*, 2003, **315** : 283-291.
- GWIZDEK C., OSSAREH-NAZARI B., BROWNAWELL A.M., DOGLIO A., BERTRAND E., MACARA I.G., DARGEMONT C. Exportin-5 mediates nuclear export of minihelix-containing RNAs. *The Journal of Biological Chemistry*, 2003, **278** : 5505-5508.

HAFNER M., LANDTHALER M., BURGER L., KHORSHID M., HAUSSEER J., BERNINGER P., ROTHBALLER A., ASCANO M., JUNGKAMP A.-C., MUNSCHAUER M., ULRICH A., WARDLE G.S., DEWELL S., ZAVOLAN M., TUSCHL T. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 2010, **141** : 129-141.

HAN J., LEE Y., YEOM K.-H., KIM Y.-K., JIN H., KIM V.N. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development*, 2004, **18** : 3016-3027.

HAN J., LEE Y., YEOM K.-H., NAM J.-W., HEO I., RHEE J.-K., SOHN S.Y., CHO Y., ZHANG B.-T., KIM V.N. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, 2006, **125** : 887-901.

HANSEN T.B., JENSEN T.I., CLAUSEN B.H., BRAMSEN J.B., FINSEN B., DAMGAARD C.K., KJEMS J. Natural RNA circles function as efficient microRNA sponges. *Nature*, 2013, **495** : 384-388.

HANSON P.I., WHITEHEART S.W. AAA+ proteins: have engine, will work. *Nature Reviews. Molecular Cell Biology*, 2005, **6** : 519-529.

HARRIS S.F., SHIAU A.K., AGARD D.A. The crystal structure of the carboxy-terminal dimerization domain of htpG, the Escherichia coli Hsp90, reveals a potential substrate binding site. *Structure (London, England: 1993)*, 2004, **12** : 1087-1097.

HAVENS M.A., REICH A.A., DUELLI D.M., HASTINGS M.L. Biogenesis of mammalian microRNAs by a non-canonical processing pathway. *Nucleic Acids Research*, 2012, **40** : 4626-4640.

HEISS N.S., KNIGHT S.W., VULLIAMY T.J., KLAUCK S.M., WIEMANN S., MASON P.J., POUSTKA A., DOKAL I. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nature Genetics*, 1998, **19** : 32-38.

HENRAS A., DEZ C., NOAILLAC-DEPEYRE J., HENRY Y., CAIZERGUES-FERRER M. Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. *Nucleic Acids Research*, 2001, **29** : 2733-2746.

HENRAS A., HENRY Y., BOUSQUET-ANTONELLI C., NOAILLAC-DEPEYRE J., GÉLUGNE J.P., CAIZERGUES-FERRER M. Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *The EMBO journal*, 1998, **17** : 7078-7090.

HILEY S.L., BABAK T., HUGHES T.R. Global analysis of yeast RNA processing identifies new targets of RNase III and uncovers a link between tRNA 5' end processing and tRNA splicing. *Nucleic Acids Research*, 2005, **33** : 3048-3056.

- HIROSE T., IDEUE T., NAGAI M., HAGIWARA M., SHU M.-D., STEITZ J.A. A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing. *Molecular Cell*, 2006, **23** : 673-684.
- HIROSE T., SHU M.-D., STEITZ J.A. Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Molecular Cell*, 2003, **12** : 113-123.
- HIROSE T., STEITZ J.A. Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, 2001, **98** : 12914-12919.
- HOAGLAND M.B., STEPHENSON M.L., SCOTT J.F., HECHT L.I., ZAMECNIK P.C. A soluble ribonucleic acid intermediate in protein synthesis. *The Journal of Biological Chemistry*, 1958, **231** : 241-257.
- HOAREAU-AVEILLA C., BONOLI M., CAIZERGUES-FERRER M., HENRY Y. hNaf1 is required for accumulation of human box H/ACA snoRNPs, scaRNPs, and telomerase. *RNA*, 2006, **12** : 832-840.
- HÖCK J., WEINMANN L., ENDER C., RÜDEL S., KREMMER E., RAABE M., URLAUB H., MEISTER G. Proteomic and functional analysis of Argonaute-containing mRNA–protein complexes in human cells. *EMBO Reports*, 2007, **8** : 1052-1060.
- HOŘEJŠÍ Z., STACH L., FLOWER T.G., JOSHI D., FLYNN H., SKEHEL J.M., O'REILLY N.J., OGRODOWICZ R.W., SMERDON S.J., BOULTON S.J. Phosphorylation-dependent PIH1D1 interactions define substrate specificity of the R2TP cochaperone complex. *Cell Reports*, 2014, **7** : 19-26.
- HOREJŠÍ Z., TAKAI H., ADELMAN C.A., COLLIS S.J., FLYNN H., MASLEN S., SKEHEL J.M., DE LANGE T., BOULTON S.J. CK2 phospho-dependent binding of R2TP complex to TEL2 is essential for mTOR and SMG1 stability. *Molecular Cell*, 2010, **39** : 839-850.
- HOUSELEY J., TOLLERVEY D. The Many Pathways of RNA Degradation. *Cell*, 2009, **136** : 763-776.
- HUANG V., ZHENG J., QI Z., WANG J., PLACE R.F., YU J., LI H., LI L.-C. Ago1 Interacts with RNA polymerase II and binds to the promoters of actively transcribed genes in human cancer cells. *PLoS genetics*, 2013, **9** : e1003821.
- HUMPHREYS D.T., WESTMAN B.J., MARTIN D.I.K., PREISS T. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102** : 16961-16966.

HUNTZINGER E., IZAURRALDE E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews. Genetics*, 2011, **12** : 99-110.

HUNTZINGER E., KUZUOGLU-ÖZTÜRK D., BRAUN J.E., EULALIO A., WOHLBOLD L., IZAURRALDE E. The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. *Nucleic Acids Research*, 2013, **41** : 978-994.

HWANG H.-W., WENTZEL E.A., MENDELL J.T. A hexanucleotide element directs microRNA nuclear import. *Science (New York, N.Y.)*, 2007, **315** : 97-100.

IIZASA H., WULFF B.-E., ALLA N.R., MARAGKAKIS M., MEGRAW M., HATZIGEORGIOU A., IWAKIRI D., TAKADA K., WIEDMER A., SHOWE L., LIEBERMAN P., NISHIKURA K. Editing of Epstein-Barr Virus-encoded BART6 MicroRNAs Controls Their Dicer Targeting and Consequently Affects Viral Latency. *Journal of Biological Chemistry*, 2010, **285** : 33358-33370.

IKURA T., OGRYZKO V.V., GRIGORIEV M., GROISMAN R., WANG J., HORIKOSHI M., SCULLY R., QIN J., NAKATANI Y. Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell*, 2000, **102** : 463-473.

INOUE M., SAEKI M., EGUSA H., NIWA H., KAMISAKI Y. PIH1D1, a subunit of R2TP complex, inhibits doxorubicin-induced apoptosis. *Biochemical and Biophysical Research Communications*, 2010, **403** : 340-344.

ITO T., CHIBA T., OZAWA R., YOSHIDA M., HATTORI M., SAKAKI Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences of the United States of America*, 2001, **98** : 4569-4574.

ITSUKI Y., SAEKI M., NAKAHARA H., EGUSA H., IRIE Y., TERAO Y., KAWABATA S., YATANI H., KAMISAKI Y. Molecular cloning of novel Monad binding protein containing tetratricopeptide repeat domains. *FEBS letters*, 2008, **582** : 2365-2370.

IWAHASHI H., YAMAGATA K., YOSHIUCHI I., TERASAKI J., YANG Q., FUKUI K., IHARA A., ZHU Q., ASAKURA T., CAO Y., IMAGAWA A., NAMBA M., HANAFUSA T., MIYAGAWA J., MATSUZAWA Y. Thyroid Hormone Receptor Interacting Protein 3 (Trip3) Is a Novel Coactivator of Hepatocyte Nuclear Factor-4 α . *Diabetes*, 2002, **51** : 910-914.

IZUMI N., YAMASHITA A., IWAMATSU A., KURATA R., NAKAMURA H., SAARI B., HIRANO H., ANDERSON P., OHNO S. AAA+ proteins RUVBL1 and RUVBL2 coordinate PIKK activity and function in nonsense-mediated mRNA decay. *Science Signaling*, 2010, **3** : ra27.

JERONIMO C., FORGET D., BOUCHARD A., LI Q., CHUA G., POITRAS C., THÉRIEN C., BERGERON D., BOURASSA S., GREENBLATT J., CHABOT B., POIRIER G.G., HUGHES T.R., BLANCHETTE M., PRICE D.H., COULOMBE B. Systematic Analysis of the Protein Interaction Network for the Human Transcription Machinery Reveals the Identity of the 7SK Capping Enzyme. *Molecular Cell*, 2007, **27** : 262-274.

JHA S., DUTTA A. RVB1/RVB2: running rings around molecular biology. *Molecular Cell*, 2009, **34** : 521-533.

JIANG W., MIDDLETON K., YOON H.J., FOUQUET C., CARBON J. An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules. *Molecular and Cellular Biology*, 1993, **13** : 4884-4893.

JINN S., BRANDIS K.A., REN A., CHACKO A., DUDLEY-RUCKER N., GALE S.E., SIDHU R., FUJIWARA H., JIANG H., OLSEN B.N., SCHAFFER J.E., ORY D.S. snoRNA U17 Regulates Cellular Cholesterol Trafficking. *Cell Metabolism*, 2015, **21** : 855-867.

JOHNSTON M., GEOFFROY M.-C., SOBALA A., HAY R., HUTVAGNER G. HSP90 Protein Stabilizes Unloaded Argonaute Complexes and Microscopic P-bodies in Human Cells. *Molecular Biology of the Cell*, 2010, **21** : 1462-1469.

JONES K.W., GORZYNSKI K., HALES C.M., FISCHER U., BADBANCHI F., TERNS R.M., TERNS M.P. Direct Interaction of the Spinal Muscular Atrophy Disease Protein SMN with the Small Nucleolar RNA-associated Protein Fibrillarin. *Journal of Biological Chemistry*, 2001, **276** : 38645-38651.

JORJANI H., KEHR S., JEDLINSKI D.J., GUMIENNY R., HERTEL J., STADLER P.F., ZAVOLAN M., GRUBER A.R. An updated human snoRNAome. *Nucleic Acids Research*, 2016, **44** : 5068-5082.

KAKIHARA Y., SAEKI M. The R2TP chaperone complex: its involvement in snoRNP assembly and tumorigenesis. *Biomolecular Concepts*, 2014, **5** : 513-520.

KALANTARI R., HICKS J.A., LI L., GAGNON K.T., SRIDHARA V., LEMOFF A., MIRZAEI H., COREY D.R. Stable association of RNAi machinery is conserved between the cytoplasm and nucleus of human cells. *RNA (New York, N.Y.)*, 2016, **22** : 1085-1098.

KAMANO Y., SAEKI M., EGUSA H., KAKIHARA Y., HOURY W.A., YATANI H., KAMISAKI Y. PIH1D1 interacts with mTOR complex 1 and enhances ribosome RNA transcription. *FEBS letters*, 2013, **587** : 3303-3308.

KANEMAKI M., KUROKAWA Y., MATSU-URA T., MAKINO Y., MASANI A., OKAZAKI K., MORISHITA T., TAMURA T.A. TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuvB-like DNA helicase, TIP49a. *The Journal of Biological Chemistry*, 1999, **274** : 22437-22444.

KANEMAKI M., MAKINO Y., YOSHIDA T., KISHIMOTO T., KOGA A., YAMAMOTO K., YAMAMOTO M., MONCOLLIN V., EGLY J.-M., MURAMATSU M., TAMURA T. Molecular Cloning of a Rat 49-kDa TBP-Interacting Protein (TIP49) That Is Highly Homologous to the Bacterial RuvB. *Biochemical and Biophysical Research Communications*, 1997, **235** : 64-68.

KAPRANOV P., CHENG J., DIKE S., NIX D.A., DUTTAGUPTA R., WILLINGHAM A.T., STADLER P.F., HERTEL J., HACKERMÜLLER J., HOFACKER I.L., BELL I., CHEUNG E., DRENKOW J., DUMAIS E., PATEL S., HELT G., GANESH M., GHOSH S., PICCOLBONI A., SEMENTCHENKO V., TAMMANA H., GINGERAS T.R. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science (New York, N.Y.)*, 2007, **316** : 1484-1488.

KATOH T., SAKAGUCHI Y., MIYAUCHI K., SUZUKI T., KASHIWABARA S., BABA T., SUZUKI T. Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes & Development*, 2009, **23** : 433-438.

KAWAHARA Y., ZINSHTEYN B., SETHUPATHY P., IIZASA H., HATZIGEORGIOU A.G., NISHIKURA K. Redirection of Silencing Targets by Adenosine-to-Inosine Editing of miRNAs. *Science*, 2007, **315** : 1137-1140.

KAWAJI H., NAKAMURA M., TAKAHASHI Y., SANDELIN A., KATAYAMA S., FUKUDA S., DAUB C.O., KAI C., KAWAI J., YASUDA J., CARNINCI P., HAYASHIZAKI Y. Hidden layers of human small RNAs. *BMC genomics*, 2008, **9** : 157.

KAWAKAMI K., SHAFER B.K., GARFINKEL D.J., STRATHERN J.N., NAKAMURA Y. Ty Element-Induced Temperature-Sensitive Mutations of *Saccharomyces Cerevisiae*. *Genetics*, 1992, **131** : 821-832.

KAWAMATA T., SEITZ H., TOMARI Y. Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nature Structural & Molecular Biology*, 2009, **16** : 953-960.

KEDDE M., VAN KOUWENHOVE M., ZWART W., OUDE VRIELINK J.A.F., ELKON R., AGAMI R. A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. *Nature Cell Biology*, 2010, **12** : 1014-1020.

- KEDDE M., STRASSER M.J., BOLDAJIPOUR B., VRIELINK J.A.F.O., SLANCHEV K., SAGE C. LE, NAGEL R., VOORHOEVE P.M., DUIJSE J. VAN, ØROM U.A., LUND A.H., PERRAKIS A., RAZ E., AGAMI R. RNA-Binding Protein Dnd1 Inhibits MicroRNA Access to Target mRNA. *Cell*, 2007, **131** : 1273-1286.
- KENNEDY E.M., WHISNANT A.W., KORNEPATI A.V.R., MARSHALL J.B., BOGERD H.P., CULLEN B.R. Production of functional small interfering RNAs by an amino-terminal deletion mutant of human Dicer. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112** : E6945-6954.
- KIERZEK E., MALGOWSKA M., LISOWIEC J., TURNER D.H., GDANIEC Z., KIERZEK R. The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Research*, 2014, **42** : 3492-3501.
- KIM D.H., SÆTROM P., SNØVE O., ROSSI J.J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105** : 16230-16235.
- KIM Y.-K., KIM B., KIM V.N. Re-evaluation of the roles of DROSHA, Exportin 5, and DICER in microRNA biogenesis. *Proceedings of the National Academy of Sciences*, 2016, **113** : E1881-E1889.
- KING T.H., DECATUR W.A., BERTRAND E., MAXWELL E.S., FOURNIER M.J. A Well-Connected and Conserved Nucleoplasmic Helicase Is Required for Production of Box C/D and H/ACA snoRNAs and Localization of snoRNP Proteins. *Molecular and Cellular Biology*, 2001, **21** : 7731-7746.
- KISHORE S., GRUBER A.R., JEDLINSKI D.J., SYED A.P., JORJANI H., ZAVOLAN M. Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biology*, 2013, **14** : R45.
- KISHORE S., KHANNA A., ZHANG Z., HUI J., BALWIERZ P.J., STEFAN M., BEACH C., NICHOLLS R.D., ZAVOLAN M., STAMM S. The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Human Molecular Genetics*, 2010, **19** : 1153-1164.
- KISHORE S., STAMM S. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science (New York, N.Y.)*, 2006, **311** : 230-232.
- KISS T., FAYET E., JÁDY B.E., RICHARD P., WEBER M. Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harbor Symposia on Quantitative Biology*, 2006, **71** : 407-417.

KISS T., FILIPOWICZ W. Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes & Development*, 1995, **9** : 1411-1424.

KISS T., FILIPOWICZ W. Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene RCC1. *The EMBO journal*, 1993, **12** : 2913-2920.

KITTUR N., DARZACQ X., ROY S., SINGER R.H., MEIER U.T. Dynamic association and localization of human H/ACA RNP proteins. *RNA*, 2006, **12** : 2057-2062.

KLEIN D.J., SCHMEING T.M., MOORE P.B., STEITZ T.A. The kink-turn: a new RNA secondary structure motif. *The EMBO journal*, 2001, **20** : 4214-4221.

KOBAYASHI H., ABE K., MATSUURA T., IKEDA Y., HITOMI T., AKECHI Y., HABU T., LIU W., OKUDA H., KOIZUMI A. Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. *American Journal of Human Genetics*, 2011, **89** : 121-130.

KOK K.H., NG M.-H.J., CHING Y.-P., JIN D.-Y. Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *The Journal of Biological Chemistry*, 2007, **282** : 17649-17657.

KOO B.-K., PARK C.-J., FERNANDEZ C.F., CHIM N., DING Y., CHANFREAU G., FEIGON J. Structure of H/ACA RNP Protein Nhp2p Reveals Cis/Trans Isomerization of a Conserved Proline at the RNA and Nop10 Binding Interface. *Journal of Molecular Biology*, 2011, **411** : 927-942.

KRESSLER D., LINDER P., DE LA CRUZ J. Protein trans-acting factors involved in ribosome biogenesis in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 1999, **19** : 7897-7912.

KROGAN N.J., KEOGH M.-C., DATTA N., SAWA C., RYAN O.W., DING H., HAW R.A., POOTOOLAL J., TONG A., CANADIEN V., RICHARDS D.P., WU X., EMILI A., HUGHES T.R., BURATOWSKI S., GREENBLATT J.F. A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Molecular Cell*, 2003, **12** : 1565-1576.

KROGAN N.J., PENG W.-T., CAGNEY G., ROBINSON M.D., HAW R., ZHONG G., GUO X., ZHANG X., CANADIEN V., RICHARDS D.P., BEATTIE B.K., LALEV A., ZHANG W., DAVIERWALA A.P., MNAIMNEH S., STAROSTINE A., TIKUISIS A.P., GRIGULL J., DATTA N., BRAY J.E., HUGHES T.R., EMILI A., GREENBLATT J.F. High-Definition Macromolecular Composition of Yeast RNA-Processing Complexes. *Molecular Cell*, 2004, **13** : 225-239.

KUCHENBAUER F., MAH S.M., HEUSER M., MCPHERSON A., RÜSCHMANN J., ROUHI A., BERG T., BULLINGER L., ARGIROPOULOS B., MORIN R.D., LAI D., STARCZYNOWSKI D.T., KARSAN A., EAVES C.J., WATAHIKI A., WANG Y., APARICIO S.A., GANSER A., KRAUTER J., DÖHNER H., DÖHNER K., MARRA M.A., CAMARGO F.D., PALMQVIST L., BUSKE C., HUMPHRIES R.K. Comprehensive analysis of mammalian miRNA* species and their role in myeloid cells. *Blood*, 2011, **118** : 3350-3358.

KWAK P.B., TOMARI Y. The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nature Structural & Molecular Biology*, 2012, **19** : 145-151.

KWON S.C., NGUYEN T.A., CHOI Y.-G., JO M.H., HOHNG S., KIM V.N., WOO J.-S. Structure of Human DROSHA. *Cell*, 2016, **164** : 81-90.

LAFONTAINE D.L., BOUSQUET-ANTONELLI C., HENRY Y., CAIZERGUES-FERRER M., TOLLERVEY D. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes & Development*, 1998, **12** : 527-537.

LAFONTAINE D.L., TOLLERVEY D. Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA (New York, N.Y.)*, 1999, **5** : 455-467.

LAFONTAINE D.L., TOLLERVEY D. Synthesis and assembly of the box C+D small nucleolar RNPs. *Molecular and Cellular Biology*, 2000, **20** : 2650-2659.

LAGOS-QUINTANA M., RAUHUT R., LENDECKEL W., TUSCHL T. Identification of novel genes coding for small expressed RNAs. *Science (New York, N.Y.)*, 2001, **294** : 853-858.

LANDGRAF P., RUSU M., SHERIDAN R., SEWER A., IOVINO N., ARAVIN A., PFEFFER S., RICE A., KAMPHORST A.O., LANDTHALER M., LIN C., SOCCI N.D., HERMIDA L., FULCI V., CHIARETTI S., FOÀ R., SCHLIWKA J., FUCHS U., NOVOSEL A., MÜLLER R.-U., SCHERMER B., BISSELS U., INMAN J., PHAN Q., CHIEN M., WEIR D.B., CHOKSI R., DE VITA G., FREZZETTI D., TROMPETER H.-I., HORNUNG V., TENG G., HARTMANN G., PALKOVITS M., DI LAURO R., WERNET P., MACINO G., ROGLER C.E., NAGLE J.W., JU J., PAPAVALIOU F.N., BENZING T., LICHTER P., TAM W., BROWNSTEIN M.J., BOSIO A., BORKHARDT A., RUSSO J.J., SANDER C., ZAVOLAN M., TUSCHL T. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, 2007, **129** : 1401-1414.

LANDTHALER M., GAIDATZIS D., ROTHBALLER A., CHEN P.Y., SOLL S.J., DINIC L., OJO T., HAFNER M., ZAVOLAN M., TUSCHL T. Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA*, 2008, **14** : 2580-2596.

- LARAKI G., CLERZIUS G., DAHER A., MELENDEZ-PEÑA C., DANIELS S., GATIGNOL A. Interactions between the double-stranded RNA-binding proteins TRBP and PACT define the Medipal domain that mediates protein-protein interactions. *RNA biology*, 2008, **5** : 92-103.
- LAU N.C., LIM L.P., WEINSTEIN E.G., BARTEL D.P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science (New York, N.Y.)*, 2001, **294** : 858-862.
- LAU P.-W., GUILLEY K.Z., DE N., POTTER C.S., CARRAGHER B., MACRAE I.J. The molecular architecture of human Dicer. *Nature Structural & Molecular Biology*, 2012, **19** : 436-440.
- LAU P.-W., POTTER C.S., CARRAGHER B., MACRAE I.J. Structure of the human Dicer-TRBP complex by electron microscopy. *Structure (London, England: 1993)*, 2009, **17** : 1326-1332.
- LECHERTIER T., GROB A., HERNANDEZ-VERDUN D., ROUSSEL P. Fibrillarin and Nop56 interact before being co-assembled in box C/D snoRNPs. *Experimental Cell Research*, 2009, **315** : 928-942.
- LEE H.Y., DOUDNA J.A. TRBP alters human precursor microRNA processing in vitro. *RNA*, 2012, **18** : 2012-2019.
- LEE H.Y., ZHOU K., SMITH A.M., NOLAND C.L., DOUDNA J.A. Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing. *Nucleic Acids Research*, 2013, **41** : 6568-6576.
- LEE R.C., AMBROS V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science (New York, N.Y.)*, 2001, **294** : 862-864.
- LEE R.C., FEINBAUM R.L., AMBROS V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 1993, **75** : 843-854.
- LEE Y., AHN C., HAN J., CHOI H., KIM J., YIM J., LEE J., PROVOST P., RADMARK O., KIM S., KIM V.N. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 2003, **425** : 415-419.
- LEE Y., HUR I., PARK S.-Y., KIM Y.-K., SUH M.R., KIM V.N. The role of PACT in the RNA silencing pathway. *The EMBO journal*, 2006, **25** : 522-532.
- LEE Y., JEON K., LEE J.-T., KIM S., KIM V.N. MicroRNA maturation: stepwise processing and subcellular localization. *The EMBO journal*, 2002, **21** : 4663-4670.
- LEONI G., TRAMONTANO A. A structural view of microRNA–target recognition. *Nucleic Acids Research*, 2016 : gkw043.

LESTRADE L., WEBER M.J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Research*, 2006, **34** : D158-162.

LEUCCI E., PATELLA F., WAAGE J., HOLMSTRØM K., LINDOW M., PORSE B., KAUPPINEN S., LUND A.H. microRNA-9 targets the long non-coding RNA MALAT1 for degradation in the nucleus. *Scientific Reports*, 2013, **3** : 2535.

LEULLIOT N., GODIN K.S., HOAREAU-AVEILLA C., QUEVILLON-CHERUEL S., VARANI G., HENRY Y., VAN TILBEURGH H. The box H/ACA RNP assembly factor Naf1p contains a domain homologous to Gar1p mediating its interaction with Cbf5p. *Journal of Molecular Biology*, 2007, **371** : 1338-1353.

LEUSCHNER P.J.F., AMERES S.L., KUENG S., MARTINEZ J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO reports*, 2006, **7** : 314-320.

LEVITSKIĪ S.A., MUKHAR'IAMOVA K.S., VEĪKO V.P., ZATSEPINA O.V. [Identification of signal sequences determining the specific nucleolar localization of fibrillarin in HeLa cells]. *Molekuliarnaia Biologiia*, 2004, **38** : 483-492.

LEWIS B.P., SHIH I. -HUN., JONES-RHOADES M.W., BARTEL D.P., BURGE C.B. Prediction of mammalian microRNA targets. *Cell*, 2003, **115** : 787-798.

LI J., RICHTER K., BUCHNER J. Mixed Hsp90-cochaperone complexes are important for the progression of the reaction cycle. *Nature Structural & Molecular Biology*, 2011, **18** : 61-66.

LIAN S.L., LI S., ABADAL G.X., PAULEY B.A., FRITZLER M.J., CHAN E.K.L. The C-terminal half of human Ago2 binds to multiple GW-rich regions of GW182 and requires GW182 to mediate silencing. *RNA (New York, N.Y.)*, 2009, **15** : 804-813.

LIANG B., ZHOU J., KAHEN E., TERNS R.M., TERNS M.P., LI H. Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. *Nature Structural & Molecular Biology*, 2009, **16** : 740-746.

LIANG H., LANDWEBER L.F. Hypothesis: RNA editing of microRNA target sites in humans? *RNA*, 2007, **13** : 463-467.

LIAO J., YU L., MEI Y., GUARNERA M., SHEN J., LI R., LIU Z., JIANG F. Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. *Molecular Cancer*, 2010, **9** : 198.

LIAO J.-Y., MA L.-M., GUO Y.-H., ZHANG Y.-C., ZHOU H., SHAO P., CHEN Y.-Q., QU L.-H. Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. *PloS One*, 2010, **5** : e10563.

- LISCHWE M.A., OCHS R.L., REDDY R., COOK R.G., YEOMAN L.C., TAN E.M., REICHLIN M., BUSCH H. Purification and partial characterization of a nucleolar scleroderma antigen (Mr = 34,000; pI, 8.5) rich in NG,NG-dimethylarginine. *The Journal of Biological Chemistry*, 1985, **260** : 14304-14310.
- LIU J., RIVAS F.V., WOHLSCHEGEL J., YATES J.R., PARKER R., HANNON G.J. A role for the P-body component GW182 in microRNA function. *Nature Cell Biology*, 2005, **7** : 1261-1266.
- LIU S., LI P., DYBKOV O., NOTTROT S., HARTMUTH K., LÜHRMANN R., CARLOMAGNO T., WAHL M.C. Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science (New York, N.Y.)*, 2007, **316** : 115-120.
- LIU X., JIN D.-Y., MCMANUS M.T., MOURELATOS Z. Precursor microRNA-programmed silencing complex assembly pathways in mammals. *Molecular Cell*, 2012, **46** : 507-517.
- LOZA-MULLER L., RODRÍGUEZ-CORONA U., SOBOL M., RODRÍGUEZ-ZAPATA L.C., HOZAK P., CASTANO E. Fibrillar methylation of H2A in RNA polymerase I trans-active promoters in *Brassica oleracea*. *Frontiers in Plant Science*, 2015, **6**.
- MA E., MACRAE I.J., KIRSCH J.F., DOUDNA J.A. Autoinhibition of human Dicer by its internal helicase domain. *Journal of Molecular Biology*, 2008, **380** : 237-243.
- MA H., WU Y., CHOI J.-G., WU H. Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, **110** : 20687-20692.
- MA J.-B., YE K., PATEL D.J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, 2004, **429** : 318-322.
- MACHADO-PINILLA R., LIGER D., LEULLIOT N., MEIER U.T. Mechanism of the AAA+ ATPases pontin and reptin in the biogenesis of H/ACA RNPs. *RNA (New York, N.Y.)*, 2012, **18** : 1833-1845.
- MACIAS S., CORDINER R.A., GAUTIER P., PLASS M., CÁCERES J.F. DGCR8 Acts as an Adaptor for the Exosome Complex to Degrade Double-Stranded Structured RNAs. *Molecular Cell*, 2015, **60** : 873-885.
- MACRAE I.J., DOUDNA J.A. Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Current Opinion in Structural Biology*, 2007, **17** : 138-145.
- MACRAE I.J., MA E., ZHOU M., ROBINSON C.V., DOUDNA J.A. In vitro reconstitution of the human RISC-loading complex. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105** : 512-517.

MACRAE I.J., ZHOU K., LI F., REPIC A., BROOKS A.N., CANDE W.Z., ADAMS P.D., DOUDNA J.A. Structural basis for double-stranded RNA processing by Dicer. *Science (New York, N.Y.)*, 2006, **311** : 195-198.

MAKINO Y., KANEMAKI M., KUROKAWA Y., KOJI T., TAMURA T. A A rat RuvB-like protein, TIP49a, is a germ cell-enriched novel DNA helicase. *The Journal of Biological Chemistry*, 1999, **274** : 15329-15335.

MARCU M.G., CHADLI A., BOUHOUCHE I., CATELLI M., NECKERS L.M. The heat shock protein 90 antagonist novobiocin interacts with a previously unrecognized ATP-binding domain in the carboxyl terminus of the chaperone. *The Journal of Biological Chemistry*, 2000, **275** : 37181-37186.

MARÍN R.M., VANÍČEK J. Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Research*, 2011, **39** : 19-29.

MARMIER-GOURRIER N., CLÉRY A., SENTY-SÉGAULT V., CHARPENTIER B., SCHLOTTER F., LECLERC F., FOURNIER R., BRANLANT C. A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA (New York, N.Y.)*, 2003, **9** : 821-838.

MARONEY P.A., YU Y., FISHER J., NILSEN T.W. Evidence that microRNAs are associated with translating messenger RNAs in human cells. *Nature Structural & Molecular Biology*, 2006, **13** : 1102-1107.

MARTENS-UZUNOVA E.S., HOOGSTRATE Y., KALSBEK A., PIGMANS B., VREDENBREGT-VAN DEN BERG M., DITS N., NIELSEN S.J., BAKER A., VISAKORPI T., BANGMA C., JENSTER G. C/D-box snoRNA-derived RNA production is associated with malignant transformation and metastatic progression in prostate cancer. *Oncotarget*, 2015, **6** : 17430-17444.

MARTIN E.C., ELLIOTT S., RHODES L.V., ANTOON J.W., FEWELL C., ZHU Y., DRIVER J.L., JODARI-KARIMI M., TAYLOR C.W., FLEMINGTON E.K., BECKMAN B.S., COLLINS-BUROW B.M., BUROW M.E. Preferential star strand biogenesis of pre-miR-24-2 targets PKC-alpha and suppresses cell survival in MCF-7 breast cancer cells. *Molecular Carcinogenesis*, 2014, **53** : 38-48.

MARZEC M., ELETTO D., ARGON Y. GRP94: An HSP90-like protein specialized for protein folding and quality control in the endoplasmic reticulum. *Biochimica Et Biophysica Acta*, 2012, **1823** : 774-787.

MATHONNET G., FABIAN M.R., SVITKIN Y.V., PARSYAN A., HUCK L., MURATA T., BIFFO S., MERRICK W.C., DARZYNKIEWICZ E., PILLAI R.S., FILIPOWICZ W., DUCHAINE T.F., SONENBERG N. MicroRNA Inhibition of Translation Initiation in Vitro by Targeting the Cap-Binding Complex eIF4F. *Science*, 2007, **317** : 1764-1767.

MATIAS P.M., GORYNIA S., DONNER P., CARRONDO M.A. Crystal structure of the human AAA+ protein RuvBL1. *The Journal of Biological Chemistry*, 2006, **281** : 38918-38929.

MATSUI M., CHU Y., ZHANG H., GAGNON K.T., SHAIKH S., KUCHIMANCHI S., MANOHARAN M., COREY D.R., JANOWSKI B.A. Promoter RNA links transcriptional regulation of inflammatory pathway genes. *Nucleic Acids Research*, 2013, **41** : 10086-10109.

MAXWELL E.S., FOURNIER M.J. The small nucleolar RNAs. *Annual Review of Biochemistry*, 1995, **64** : 897-934.

MCKEEGAN K.S., DEBIEUX C.M., BOULON S., BERTRAND E., WATKINS N.J. A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. *Molecular and Cellular Biology*, 2007, **27** : 6782-6793.

MCKEEGAN K.S., DEBIEUX C.M., WATKINS N.J. Evidence that the AAA+ Proteins TIP48 and TIP49 Bridge Interactions between 15.5K and the Related NOP56 and NOP58 Proteins during Box C/D snoRNP Biogenesis. *Molecular and Cellular Biology*, 2009, **29** : 4971-4981.

MEI Y., LIAO J., SHEN J., YU L., LIU B., LIU L., LI R., JI L., DORSEY S., JIANG Z., KATZ R., WANG J.-Y., JIANG F. Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. *Oncogene*, 2012, **31** : 2794-2804.

MEISTER G., LANDTHALER M., PETERS L., CHEN P.Y., URLAUB H., LÜHRMANN R., TUSCHL T. Identification of novel argonaute-associated proteins. *Current biology: CB*, 2005, **15** : 2149-2155.

MERCER T.R., DINGER M.E., MATTICK J.S. Long non-coding RNAs: insights into functions. *Nature Reviews Genetics*, 2009, **10** : 155-159.

MÉSZÁROS B., SIMON I., DOSZTÁNYI Z. The expanding view of protein–protein interactions: complexes involving intrinsically disordered proteins. *Physical Biology*, 2011, **8** : 35003.

MEYER P., PRODROMOU C., HU B., VAUGHAN C., ROE S.M., PANARETOU B., PIPER P.W., PEARL L.H. Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. *Molecular Cell*, 2003, **11** : 647-658.

MICHEL C.I., HOLLEY C.L., SCRUGGS B.S., SIDHU R., BROOKHEART R.T., LISTENBERGER L.L., BEHLKE M.A., ORY D.S., SCHAFFER J.E. Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. *Cell Metabolism*, 2011, **14** : 33-44.

MINAMI Y., KIMURA Y., KAWASAKI H., SUZUKI K., YAHARA I. The carboxy-terminal region of mammalian HSP90 is required for its dimerization and function in vivo. *Molecular and Cellular Biology*, 1994, **14** : 1459-1464.

MIYATA Y. Hsp90 inhibitor geldanamycin and its derivatives as novel cancer chemotherapeutic agents. *Current Pharmaceutical Design*, 2005, **11** : 1131-1138.

MOUAIKEL J., VERHEGGEN C., BERTRAND E., TAZI J., BORDONNÉ R. Hypermethylation of the Cap Structure of Both Yeast snRNAs and snoRNAs Requires a Conserved Methyltransferase that Is Localized to the Nucleolus. *Molecular Cell*, 2002, **9** : 891-901.

MURPHY D., DANCIS B., BROWN J.R. The evolution of core proteins involved in microRNA biogenesis. *BMC Evolutionary Biology*, 2008, **8** : 92.

NARAYANAN A., SPECKMANN W., TERNS R., TERNS M.P. Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Molecular Biology of the Cell*, 1999, **10** : 2131-2147.

NGUYEN T.A., JO M.H., CHOI Y.-G., PARK J., KWON S.C., HOHNG S., KIM V.N., WOO J.-S. Functional Anatomy of the Human Microprocessor. *Cell*, 2015, **161** : 1374-1387.

NI L., SAEKI M., XU L., NAKAHARA H., SAIJO M., TANAKA K., KAMISAKI Y. RPAP3 interacts with Reptin to regulate UV-induced phosphorylation of H2AX and DNA damage. *Journal of Cellular Biochemistry*, 2009, **106** : 920-928.

NISHI K., NISHI A., NAGASAWA T., UI-TEI K. Human TNRC6A is an Argonaute-navigator protein for microRNA-mediated gene silencing in the nucleus. *RNA (New York, N.Y.)*, 2013, **19** : 17-35.

NOTTROTT S., HARTMUTH K., FABRIZIO P., URLAUB H., VIDOVIC I., FICNER R., LÜHRMANN R. Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *The EMBO journal*, 1999, **18** : 6119-6133.

NOTTROTT S., SIMARD M.J., RICHTER J.D. Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nature Structural & Molecular Biology*, 2006, **13** : 1108-1114.

VAN NUES R.W., GRANNEMAN S., KUDLA G., SLOAN K.E., CHICKEN M., TOLLERVEY D., WATKINS N.J. Box C/D snoRNP catalysed methylation is aided by additional pre-rRNA base-pairing. *The EMBO journal*, 2011, **30** : 2420-2430.

OBERMANN W.M., SONDERMANN H., RUSSO A.A., PAVLETICH N.P., HARTL F.U. In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis. *The Journal of Cell Biology*, 1998, **143** : 901-910.

OCHS R.L., LISCHWE M.A., SPOHN W.H., BUSCH H. Fibrillarlin: a new protein of the nucleolus identified by autoimmune sera. *Biology of the Cell / Under the Auspices of the European Cell Biology Organization*, 1985, **54** : 123-133.

OHRT T., MÜTZE J., STAROSKE W., WEINMANN L., HÖCK J., CRELL K., MEISTER G., SCHWILLE P. Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Research*, 2008, **36** : 6439-6449.

OKAMURA K., PHILLIPS M.D., TYLER D.M., DUAN H., CHOU Y., LAI E.C. The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nature Structural & Molecular Biology*, 2008, **15** : 354-363.

ONO M., SCOTT M.S., YAMADA K., AVOLIO F., BARTON G.J., LAMOND A.I. Identification of human miRNA precursors that resemble box C/D snoRNAs. *Nucleic Acids Research*, 2011, **39** : 3879-3891.

OOI S.L., SAMARSKY D.A., FOURNIER M.J., BOEKE J.D. Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA. *RNA (New York, N.Y.)*, 1998, **4** : 1096-1110.

ØROM U.A., NIELSEN F.C., LUND A.H. MicroRNA-10a Binds the 5'UTR of Ribosomal Protein mRNAs and Enhances Their Translation. *Molecular Cell*, 2008, **30** : 460-471.

PACI A., LIU X.H., HUANG H., LIM A., HOURLY W.A., ZHAO R. The Stability of the Small Nucleolar Ribonucleoprotein (snoRNP) Assembly Protein Pih1 in *Saccharomyces cerevisiae* Is Modulated by Its C Terminus. *Journal of Biological Chemistry*, 2012, **287** : 43205-43214.

PAL M., MORGAN M., PHELPS S.E.L., ROE S.M., PARRY-MORRIS S., DOWNS J.A., POLIER S., PEARL L.H., PRODROMOU C. Structural basis for phosphorylation-dependent recruitment of Tel2 to Hsp90 by Pih1. *Structure (London, England: 1993)*, 2014, **22** : 805-818.

PALADE G.E. A small particulate component of the cytoplasm. *The Journal of Biophysical and Biochemical Cytology*, 1955, **1** : 59-68.

PARE J.M., LAPOINTE P., HOBMAN T.C. Hsp90 cochaperones p23 and FKBP4 physically interact with hAgo2 and activate RNA interference-mediated silencing in mammalian cells. *Molecular Biology of the Cell*, 2013, **24** : 2303-2310.

PARK J.-E., HEO I., TIAN Y., SIMANSHU D.K., CHANG H., JEE D., PATEL D.J., KIM V.N. Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*, 2011, **475** : 201-205.

PARK J.H., SHIN C. Slicer-independent mechanism drives small-RNA strand separation during human RISC assembly. *Nucleic Acids Research*, 2015, **43** : 9418-9433.

PASQUINELLI A.E., REINHART B.J., SLACK F., MARTINDALE M.Q., KURODA M.I., MALLER B., HAYWARD D.C., BALL E.E., DEGNAN B., MÜLLER P., SPRING J., SRINIVASAN A., FISHMAN M., FINNERTY J., CORBO J., LEVINE M., LEAHY P., DAVIDSON E., RUVKUN G. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 2000, **408** : 86-89.

PATEL C.V., HANDY I., GOLDSMITH T., PATEL R.C. PACT, a stress-modulated cellular activator of interferon-induced double-stranded RNA-activated protein kinase, PKR. *The Journal of Biological Chemistry*, 2000, **275** : 37993-37998.

PATEL R.C., SEN G.C. PACT, a protein activator of the interferon-induced protein kinase, PKR. *The EMBO journal*, 1998, **17** : 4379-4390.

PELLIZZONI L., BACCON J., CHARROUX B., DREYFUSS G. The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. *Current biology: CB*, 2001, **11** : 1079-1088.

PENG W.T., ROBINSON M.D., MNAIMNEH S., KROGAN N.J., CAGNEY G., MORRIS Q., DAVIERWALA A.P., GRIGULL J., YANG X., ZHANG W., MITSAKAKIS N., RYAN O.W., DATTA N., JOJIC V., PAL C., CANADIEN V., RICHARDS D., BEATTIE B., WU L.F., ALTSCHULER S.J., ROWEIS S., FREY B.J., EMILI A., GREENBLATT J.F., HUGHES T.R. A panoramic view of yeast noncoding RNA processing. *Cell*, 2003, **113** : 919-933.

PETERSEN C.P., BORDELEAU M.-E., PELLETIER J., SHARP P.A. Short RNAs Repress Translation after Initiation in Mammalian Cells. *Molecular Cell*, 2006, **21** : 533-542.

PETUKHOV M., DAGKESAMANSKAJA A., BOMMER M., BARRETT T., TSANEVA I., YAKIMOV A., QUÉVAL R., SHVETSOV A., KHODORKOVSKIY M., KÄS E., GRIGORIEV M. Large-Scale Conformational Flexibility Determines the Properties of AAA+ TIP49 ATPases. *Structure*, 2012, **20** : 1321-1331.

- POGACIĆ V., DRAGON F., FILIPOWICZ W. Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. *Molecular and Cellular Biology*, 2000, **20** : 9028-9040.
- PRADET-BALADE B., GIRARD C., BOULON S., PAUL C., AZZAG K., BORDONNÉ R., BERTRAND E., VERHEGGEN C. CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *The EMBO journal*, 2011, **30** : 2205-2218.
- PUTNAM C.D., CLANCY S.B., TSURUTA H., GONZALEZ S., WETMUR J.G., TAINER J.A. Structure and mechanism of the RuvB holliday junction branch migration motor1. *Journal of Molecular Biology*, 2001, **311** : 297-310.
- QIU X.-B., LIN Y.-L., THOME K.C., PIAN P., SCHLEGEL B.P., WEREMOWICZ S., PARVIN J.D., DUTTA A. An Eukaryotic RuvB-like Protein (RUVBL1) Essential for Growth. *Journal of Biological Chemistry*, 1998, **273** : 27786-27793.
- QUINTERNET M., CHAGOT M.-E., ROTHÉ B., TIOTIU D., CHARPENTIER B., MANIVAL X. Structural Features of the Box C/D snoRNP Pre-assembly Process Are Conserved through Species. *Structure (London, England: 1993)*, 2016.
- QUINTERNET M., ROTHÉ B., BARBIER M., BOBO C., SALIOU J.-M., JACQUEMIN C., BACK R., CHAGOT M.-E., CIANFÉRANI S., MEYER P., BRANLANT C., CHARPENTIER B., MANIVAL X. Structure/Function Analysis of Protein–Protein Interactions Developed by the Yeast Pih1 Platform Protein and Its Partners in Box C/D snoRNP Assembly. *Journal of Molecular Biology*, 2015, **427** : 2816-2839.
- RAKITINA D.V., TALIANSKY M., BROWN J.W.S., KALININA N.O. Two RNA-binding sites in plant fibrillarin provide interactions with various RNA substrates. *Nucleic Acids Research*, 2011, **39** : 8869-8880.
- RATZKE C., MICKLER M., HELLENKAMP B., BUCHNER J., HUGEL T. Dynamics of heat shock protein 90 C-terminal dimerization is an important part of its conformational cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107** : 16101-16106.
- REDDY R., LI W.Y., HENNING D., CHOI Y.C., NOHGA K., BUSCH H. Characterization and subcellular localization of 7-8 S RNAs of Novikoff hepatoma. *The Journal of Biological Chemistry*, 1981, **256** : 8452-8457.
- REICHOW S.L., HAMMA T., FERRÉ-D'AMARÉ A.R., VARANI G. The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Research*, 2007, **35** : 1452-1464.

REICHOW S.L., VARANI G. Nop10 is a conserved H/ACA snoRNP molecular adaptor. *Biochemistry*, 2008, **47** : 6148-6156.

RICHARD P., DARZACQ X., BERTRAND E., JÁDY B.E., VERHEGGEN C., KISS T. A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *The EMBO journal*, 2003, **22** : 4283-4293.

RICHARD P., KISS A.M., DARZACQ X., KISS T. Cotranscriptional recognition of human intronic box H/ACA snoRNAs occurs in a splicing-independent manner. *Molecular and Cellular Biology*, 2006, **26** : 2540-2549.

ROBB G.B., BROWN K.M., KHURANA J., RANA T.M. Specific and potent RNAi in the nucleus of human cells. *Nature Structural & Molecular Biology*, 2005, **12** : 133-137.

RODRIGUEZ A., GRIFFITHS-JONES S., ASHURST J.L., BRADLEY A. Identification of mammalian microRNA host genes and transcription units. *Genome Research*, 2004, **14** : 1902-1910.

RÖHL A., ROHRBERG J., BUCHNER J. The chaperone Hsp90: changing partners for demanding clients. *Trends in Biochemical Sciences*, 2013, **38** : 253-262.

ROTH B.M., ISHIMARU D., HENNIG M. The Core Microprocessor Component DiGeorge Syndrome Critical Region 8 (DGCR8) Is a Nonspecific RNA-binding Protein. *Journal of Biological Chemistry*, 2013, **288** : 26785-26799.

ROTHÉ B., BACK R., QUINTERNET M., BIZARRO J., ROBERT M.-C., BLAUD M., ROMIER C., MANIVAL X., CHARPENTIER B., BERTRAND E., BRANLANT C. Characterization of the interaction between protein Snu13p/15.5K and the Rsa1p/NUFIP factor and demonstration of its functional importance for snoRNP assembly. *Nucleic Acids Research*, 2014, **42** : 2015-2036.

ROTHÉ B., SALIOU J.-M., QUINTERNET M., BACK R., TIOU D., JACQUEMIN C., LOEGLER C., SCHLOTTER F., PEÑA V., ECKERT K., MORÉRA S., DORSSELAER A.V., BRANLANT C., MASSENET S., SANGLIER-CIANFÉRANI S., MANIVAL X., CHARPENTIER B. Protein Hit1, a novel box C/D snoRNP assembly factor, controls cellular concentration of the scaffolding protein Rsa1 by direct interaction. *Nucleic Acids Research*, 2014, **42** : 10731-10747.

RUBY J.G., JAN C.H., BARTEL D.P. Intronic microRNA precursors that bypass Drosha processing. *Nature*, 2007, **448** : 83-86.

RUNTE M., HÜTTENHOFER A., GROSS S., KIEFMANN M., HORSTHEMKE B., BUITING K. The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Human Molecular Genetics*, 2001, **10** : 2687-2700.

- RUSSELL L.C., WHITT S.R., CHEN M.S., CHINKERS M. Identification of conserved residues required for the binding of a tetratricopeptide repeat domain to heat shock protein 90. *The Journal of Biological Chemistry*, 1999, **274** : 20060-20063.
- RYAN B.M., ROBLES A.I., HARRIS C.C. Genetic variation in microRNA networks: the implications for cancer research. *Nature Reviews Cancer*, 2010, **10** : 389-402.
- SAEKI M., IRIE Y., NI L., YOSHIDA M., ITSUKI Y., KAMISAKI Y. Monad, a WD40 repeat protein, promotes apoptosis induced by TNF- α . *Biochemical and Biophysical Research Communications*, 2006, **342** : 568-572.
- SAHOO T., DEL GAUDIO D., GERMAN J.R., SHINAWI M., PETERS S.U., PERSON R.E., GARNICA A., CHEUNG S.W., BEAUDET A.L. Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nature Genetics*, 2008, **40** : 719-721.
- SAUNDERS M.A., LIANG H., LI W.-H. Human polymorphism at microRNAs and microRNA target sites. *Proceedings of the National Academy of Sciences*, 2007, **104** : 3300-3305.
- SCHEUFLER C., BRINKER A., BOURENKOV G., PEGORARO S., MORODER L., BARTUNIK H., HARTL F.U., MOAREFI I. Structure of TPR Domain–Peptide Complexes. *Cell*, 2000, **101** : 199-210.
- SCHIRLE N.T., MACRAE I.J. The crystal structure of human Argonaute2. *Science (New York, N.Y.)*, 2012, **336** : 1037-1040.
- SCHIRLE N.T., SHEU-GRUTTADAURIA J., MACRAE I.J. Structural basis for microRNA targeting. *Science*, 2014, **346** : 608-613.
- SCHRAIVOGEL D., SCHINDLER S.G., DANNER J., KREMMER E., PFAFF J., HANNUS S., DEPPING R., MEISTER G. Importin- β facilitates nuclear import of human GW proteins and balances cytoplasmic gene silencing protein levels. *Nucleic Acids Research*, 2015, **43** : 7447-7461.
- SCHWARZ D.S., HUTVÁGNER G., DU T., XU Z., ARONIN N., ZAMORE P.D. Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, 2003, **115** : 199-208.
- SCOTT M.S., AVOLIO F., ONO M., LAMOND A.I., BARTON G.J. Human miRNA precursors with box H/ACA snoRNA features. *PLoS computational biology*, 2009, **5** : e1000507.
- SCOTT M.S., ONO M. From snoRNA to miRNA: Dual function regulatory non-coding RNAs. *Biochimie*, 2011, **93** : 1987-1992.

- SCOTT M.S., ONO M., YAMADA K., ENDO A., BARTON G.J., LAMOND A.I. Human box C/D snoRNA processing conservation across multiple cell types. *Nucleic Acids Research*, 2012, **40** : 3676-3688.
- SHEN B., GOODMAN H.M. Uridine Addition After MicroRNA-Directed Cleavage. *Science*, 2004, **306** : 997-997.
- SHEN X., MIZUGUCHI G., HAMICHE A., WU C. A chromatin remodelling complex involved in transcription and DNA processing. *Nature*, 2000, **406** : 541-544.
- SHIOHAMA A., SASAKI T., NODA S., MINOSHIMA S., SHIMIZU N. Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. *Biochemical and Biophysical Research Communications*, 2003, **304** : 184-190.
- SHUKLA S., SCHMIDT J.C., GOLDFARB K.C., CECH T.R., PARKER R. Inhibition of telomerase RNA decay rescues telomerase deficiency caused by dyskerin or PARN defects. *Nature Structural & Molecular Biology*, 2016, **23** : 286-292.
- SILVA-MARTIN N., DAUDÉN M.I., GLATT S., HOFFMANN N.A., KASTRITIS P., BORK P., BECK M., MÜLLER C.W. The Combination of X-Ray Crystallography and Cryo-Electron Microscopy Provides Insight into the Overall Architecture of the Dodecameric Rvb1/Rvb2 Complex. *PLoS One*, 2016, **11** : e0146457.
- SINGH M., CASTILLO D., PATEL C.V., PATEL R.C. Stress-induced phosphorylation of PACT reduces its interaction with TRBP and leads to PKR activation. *Biochemistry*, 2011, **50** : 4550-4560.
- SÖTI C., RÁCZ A., CSERMELY P. A Nucleotide-dependent molecular switch controls ATP binding at the C-terminal domain of Hsp90. N-terminal nucleotide binding unmasks a C-terminal binding pocket. *The Journal of Biological Chemistry*, 2002, **277** : 7066-7075.
- STAREGA-ROSLAN J., KROL J., KOSCIANSKA E., KOZLOWSKI P., SZLACHCIC W.J., SOB CZAK K., KRZYZOSIAK W.J. Structural basis of microRNA length variety. *Nucleic Acids Research*, 2010 : gkq727.
- SUN Z., MILLER R.A., PATEL R.T., CHEN J., DHIR R., WANG H., ZHANG D., GRAHAM M.J., UNTERMAN T.G., SHULMAN G.I., SZTALRYD C., BENNETT M.J., AHIMA R.S., BIRNBAUM M.J., LAZAR M.A. Hepatic Hdac3 promotes gluconeogenesis by repressing lipid synthesis and sequestration. *Nature Medicine*, 2012, **18** : 934-942.

- SUZUKI H.I., KATSURA A., YASUDA T., UENO T., MANO H., SUGIMOTO K., MIYAZONO K. Small-RNA asymmetry is directly driven by mammalian Argonautes. *Nature Structural & Molecular Biology*, 2015, **22** : 512-521.
- SYNORADZKI K., BIEGANOWSKI P. Middle domain of human Hsp90 isoforms differentially binds Aha1 in human cells and alters Hsp90 activity in yeast. *Biochimica Et Biophysica Acta*, 2015, **1853** : 445-452.
- TAFT R.J., GLAZOV E.A., CLOONAN N., SIMONS C., STEPHEN S., FAULKNER G.J., LASSMANN T., FORREST A.R.R., GRIMMOND S.M., SCHRODER K., IRVINE K., ARAKAWA T., NAKAMURA M., KUBOSAKI A., HAYASHIDA K., KAWAZU C., MURATA M., NISHIYORI H., FUKUDA S., KAWAI J., DAUB C.O., HUME D.A., SUZUKI H., ORLANDO V., CARNINCI P., HAYASHIZAKI Y., MATTICK J.S. Tiny RNAs associated with transcription start sites in animals. *Nature Genetics*, 2009, **41** : 572-578.
- TAFT R.J., SIMONS C., NAHKURI S., OEY H., KORBIE D.J., MERCER T.R., HOLST J., RITCHIE W., WONG J.J.-L., RASKO J.E.J., ROKHSAR D.S., DEGNAN B.M., MATTICK J.S. Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. *Nature Structural & Molecular Biology*, 2010, **17** : 1030-1034.
- TAIPALE M., JAROSZ D.F., LINDQUIST S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nature Reviews. Molecular Cell Biology*, 2010, **11** : 515-528.
- TARANGELO A., LO N., TENG R., KIM E., LE L., WATSON D., FURTH E.E., RAMAN P., EHMER U., VIATOUR P. Recruitment of Pontin/Reptin by E2f1 amplifies E2f transcriptional response during cancer progression. *Nature Communications*, 2015, **6** : 10028.
- TE J., JIA L., ROGERS J., MILLER A., HARTSON S.D. Novel subunits of the mammalian Hsp90 signal transduction chaperone. *Journal of Proteome Research*, 2007, **6** : 1963-1973.
- TESSARZ P., SANTOS-ROSA H., ROBSON S.C., SYLVESTERSEN K.B., NELSON C.J., NIELSEN M.L., KOUZARIDES T. Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature*, 2014, **505** : 564-568.
- THEODORAKI M.A., CAPLAN A.J. Quality control and fate determination of Hsp90 client proteins. *Biochimica Et Biophysica Acta*, 2012, **1823** : 683-688.
- TILL S., LEJEUNE E., THERMANN R., BORTFELD M., HOTHORN M., ENDERLE D., HEINRICH C., HENTZE M.W., LADURNER A.G. A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. *Nature Structural & Molecular Biology*, 2007, **14** : 897-903.

- TSAI N.-P., LIN Y.-L., WEI L.-N. MicroRNA mir-346 targets the 5'-untranslated region of receptor-interacting protein 140 (RIP140) mRNA and up-regulates its protein expression. *Biochemical Journal*, 2009, **424** : 411-418.
- TYC K., STEITZ J.A. U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *The EMBO journal*, 1989, **8** : 3113-3119.
- VALLERON W., LAPREVOTTE E., GAUTIER E.-F., QUELEN C., DEMUR C., DELABESSE E., AGIRRE X., PRÓSPER F., KISS T., BROUSSET P. Specific small nucleolar RNA expression profiles in acute leukemia. *Leukemia*, 2012, **26** : 2052-2060.
- VALLERON W., YSEBAERT L., BERQUET L., FATACCIOLI V., QUELEN C., MARTIN A., PARRENS M., LAMANT L., DE LEVAL L., GISSELBRECHT C., GAULARD P., BROUSSET P. Small nucleolar RNA expression profiling identifies potential prognostic markers in peripheral T-cell lymphoma. *Blood*, 2012, **120** : 3997-4005.
- VENTEICHER A.S., MENG Z., MASON P.J., VEENSTRA T.D., ARTANDI S.E. Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. *Cell*, 2008, **132** : 945-957.
- VERHEGGEN C., BERTRAND E. CRM1 plays a nuclear role in transporting snoRNPs to nucleoli in higher eukaryotes. *Nucleus*, 2012, **3** : 132-137.
- VERHEGGEN C., PRADET-BALADE B., BERTRAND E. SnoRNPs, ZNHIT proteins and the R2TP pathway. *Oncotarget*, 2015, **6** : 41399-41400.
- VIDOVIC I., NOTTROTT S., HARTMUTH K., LÜHRMANN R., FICNER R. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Molecular Cell*, 2000, **6** : 1331-1342.
- VILLA T., CERADINI F., PRESUTTI C., BOZZONI I. Processing of the intron-encoded U18 small nucleolar RNA in the yeast *Saccharomyces cerevisiae* relies on both exo- and endonucleolytic activities. *Molecular and Cellular Biology*, 1998, **18** : 3376-3383.
- VITALI P., BASYUK E., LE MEUR E., BERTRAND E., MUSCATELLI F., CAVAILLÉ J., HUTTENHOFER A. ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *The Journal of Cell Biology*, 2005, **169** : 745-753.
- WALBOTT H., MACHADO-PINILLA R., LIGER D., BLAUD M., RÉTY S., GROZDANOV P.N., GODIN K., VAN TILBEURGH H., VARANI G., MEIER U.T., LEULLIOT N. The H/ACA RNP assembly factor SHQ1 functions as an RNA mimic. *Genes & Development*, 2011, **25** : 2398-2408.

WALTER P., BLOBEL G. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature*, 1982, **299** : 691-698.

WANG B., YANEZ A., NOVINA C.D. MicroRNA-repressed mRNAs contain 40S but not 60S components. *Proceedings of the National Academy of Sciences*, 2008, **105** : 5343-5348.

WANG C., MEIER U.T. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *The EMBO journal*, 2004, **23** : 1857-1867.

WANG H.-W., NOLAND C., SIRIDECHADILOK B., TAYLOR D.W., MA E., FELDERER K., DOUDNA J.A., NOGALES E. Structural insights into RNA processing by the human RISC-loading complex. *Nature Structural & Molecular Biology*, 2009, **16** : 1148-1153.

WANG X., LI M., WANG Z., HAN S., TANG X., GE Y., ZHOU L., ZHOU C., YUAN Q., YANG M. Silencing of long noncoding RNA MALAT1 by miR-101 and miR-217 inhibits proliferation, migration, and invasion of esophageal squamous cell carcinoma cells. *The Journal of Biological Chemistry*, 2015, **290** : 3925-3935.

WATKINS N.J., DICKMANN A., LÜHRMANN R. 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. *Molecular and Cellular Biology*, 2002, **22** : 8342-8352.

WATKINS N.J., SÉGAULT V., CHARPENTIER B., NOTTROT S., FABRIZIO P., BACHI A., WILM M., ROSBASH M., BRANLANT C., LÜHRMANN R. A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, 2000, **103** : 457-466.

WEI Y., LI L., WANG D., ZHANG C.-Y., ZEN K. Importin 8 regulates the transport of mature microRNAs into the cell nucleus. *The Journal of Biological Chemistry*, 2014, **289** : 10270-10275.

WEINBERG R.A., PENMAN S. Small molecular weight monodisperse nuclear RNA. *Journal of Molecular Biology*, 1968, **38** : 289-304.

WEINMANN L., HÖCK J., IVACEVIC T., OHRT T., MÜTZE J., SCHWILLE P., KREMMER E., BENES V., URLAUB H., MEISTER G. Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell*, 2009, **136** : 496-507.

WESTHOLM J.O., LAI E.C. Mirtrons: microRNA biogenesis via splicing. *Biochimie*, 2011, **93** : 1897-1904.

- WHITE E., SCHLACKOW M., KAMIENIARZ-GDULA K., PROUDFOOT N.J., GULLEROVA M. Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA. *Nature Structural & Molecular Biology*, 2014, **21** : 552-559.
- WILSON R.C., TAMBE A., KIDWELL M.A., NOLAND C.L., SCHNEIDER C.P., DOUDNA J.A. Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis. *Molecular Cell*, 2015, **57** : 397-407.
- WINTER J., JUNG S., KELLER S., GREGORY R.I., DIEDERICH S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature Cell Biology*, 2009, **11** : 228-234.
- WU H., XU H., MIRAGLIA L.J., CROOKE S.T. Human RNase III is a 160-kDa protein involved in preribosomal RNA processing. *The Journal of Biological Chemistry*, 2000, **275** : 36957-36965.
- WU H., YE C., RAMIREZ D., MANJUNATH N. Alternative processing of primary microRNA transcripts by Drosha generates 5' end variation of mature microRNA. *PloS One*, 2009, **4** : e7566.
- WU P., BROCKENBROUGH J.S., METCALFE A.C., CHEN S., ARIS J.P. Nop5p is a small nucleolar ribonucleoprotein component required for pre-18 S rRNA processing in yeast. *The Journal of Biological Chemistry*, 1998, **273** : 16453-16463.
- XIE M., LI M., VILBORG A., LEE N., SHU M.-D., YARTSEVA V., ŠESTAN N., STEITZ J.A. Mammalian 5'-Capped MicroRNA Precursors that Generate a Single MicroRNA. *Cell*, 2013, **155** : 1568-1580.
- XUE S., WANG R., YANG F., TERNS R.M., TERNS M.P., ZHANG X., MAXWELL E.S., LI H. Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Molecular Cell*, 2010, **39** : 939-949.
- YANAGIDA M., HAYANO T., YAMAUCHI Y., SHINKAWA T., NATSUME T., ISOBE T., TAKAHASHI N. Human fibrillarin forms a sub-complex with splicing factor 2-associated p32, protein arginine methyltransferases, and tubulins alpha 3 and beta 1 that is independent of its association with preribosomal ribonucleoprotein complexes. *The Journal of Biological Chemistry*, 2004, **279** : 1607-1614.
- YANG P.K., HOAREAU C., FROMENT C., MONSARRAT B., HENRY Y., CHANFREAU G. Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. *Molecular and Cellular Biology*, 2005, **25** : 3295-3304.
- YANG P.K., ROTONDO G., PORRAS T., LEGRAIN P., CHANFREAU G. The Shq1p-Naf1p Complex Is Required for Box H/ACA Small Nucleolar Ribonucleoprotein Particle Biogenesis. *Journal of Biological Chemistry*, 2002, **277** : 45235-45242.

YANG W., CHENDRIMADA T.P., WANG Q., HIGUCHI M., SEEBURG P.H., SHIEKHATTAR R., NISHIKURA K. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nature Structural & Molecular Biology*, 2006, **13** : 13-21.

YEUNG M.L., BENNASSER Y., WATASHI K., LE S.-Y., HOUZET L., JEANG K.-T. Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid. *Nucleic Acids Research*, 2009, **37** : 6575-6586.

YI R., DOEHLE B.P., QIN Y., MACARA I.G., CULLEN B.R. Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *RNA (New York, N.Y.)*, 2005, **11** : 220-226.

YI R., QIN Y., MACARA I.G., CULLEN B.R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development*, 2003, **17** : 3011-3016.

YIN Q.-F., YANG L., ZHANG Y., XIANG J.-F., WU Y.-W., CARMICHAEL G.G., CHEN L.-L. Long noncoding RNAs with snoRNA ends. *Molecular Cell*, 2012, **48** : 219-230.

YODA M., KAWAMATA T., PAROO Z., YE X., IWASAKI S., LIU Q., TOMARI Y. ATP-dependent human RISC assembly pathways. *Nature Structural & Molecular Biology*, 2010, **17** : 17-23.

YOSHIDA M., SAEKI M., EGUSA H., IRIE Y., KAMANO Y., URAGUCHI S., SOTOZONO M., NIWA H., KAMISAKI Y. RPAP3 splicing variant isoform 1 interacts with PIH1D1 to compose R2TP complex for cell survival. *Biochemical and Biophysical Research Communications*, 2013, **430** : 320-324.

YOUNG J.C., OBERMANN W.M., HARTL F.U. Specific binding of tetratricopeptide repeat proteins to the C-terminal 12-kDa domain of hsp90. *The Journal of Biological Chemistry*, 1998, **273** : 18007-18010.

YU F., BRACKEN C.P., PILLMAN K.A., LAWRENCE D.M., GOODALL G.J., CALLEN D.F., NEILSEN P.M. p53 Represses the Oncogenic Sno-MiR-28 Derived from a SnoRNA. *PLoS ONE*, 2015, **10**.

ZAMORE P.D., TUSCHL T., SHARP P.A., BARTEL D.P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, 2000, **101** : 25-33.

ZENG Y., CULLEN B.R. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Research*, 2004, **32** : 4776-4785.

ZHAI N., ZHAO Z. -L., CHENG M. -B., DI Y. -W., YAN H. -X., CAO C. -Y., DAI H., ZHANG Y., SHEN Y. -F. Human PIH1 associates with histone H4 to mediate the glucose-dependent enhancement of pre-rRNA synthesis. *Journal of Molecular Cell Biology*, 2012, **4** : 231-241.

ZHANG H., KOLB F.A., BRONDANI V., BILLY E., FILIPOWICZ W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *The EMBO Journal*, 2002, **21** : 5875-5885.

ZHANG H., KOLB F.A., JASKIEWICZ L., WESTHOF E., FILIPOWICZ W. Single processing center models for human Dicer and bacterial RNase III. *Cell*, 2004, **118** : 57-68.

ZHANG X.-O., YIN Q.-F., WANG H.-B., ZHANG Y., CHEN T., ZHENG P., LU X., CHEN L.-L., YANG L. Species-specific alternative splicing leads to unique expression of sno-lncRNAs. *BMC Genomics*, 2014, **15** : 287.

ZHAO R., DAVEY M., HSU Y.-C., KAPLANEK P., TONG A., PARSONS A.B., KROGAN N., CAGNEY G., MAI D., GREENBLATT J., BOONE C., EMILI A., HOURLY W.A. Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell*, 2005, **120** : 715-727.

ZHAO R., HOURLY W.A. Hsp90: a chaperone for protein folding and gene regulation. *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire*, 2005, **83** : 703-710.

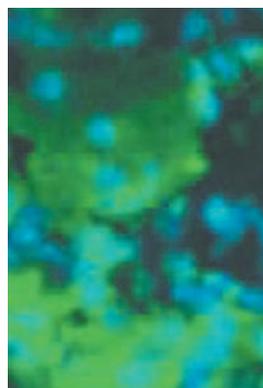
ZHAO R., KAKIHARA Y., GRIBUN A., HUEN J., YANG G., KHANNA M., COSTANZO M., BROST R.L., BOONE C., HUGHES T.R., YIP C.M., HOURLY W.A. Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. *The Journal of Cell Biology*, 2008, **180** : 563-578.

Annexes

> Les petits ARN nucléolaires (snoARN), impliqués dans la modification post-transcriptionnelle d'ARN non codants, figurent, avec les ARN de transfert et les ARN ribosomiques, parmi les premiers ARN non codants identifiés. Pourtant, plusieurs nouvelles facettes de leur éventail d'activités biologiques possibles ont été mises en évidence ces dernières années, engendrant un regain d'intérêt important pour l'étude de cette famille d'ARN. Leur maturation en ARN fonctionnels plus petits, leur implication, ainsi que celle de leurs petits ARN dérivés, dans la régulation de l'épissage alternatif, ou encore leurs liens possibles avec plusieurs pathologies, y compris le cancer, sont évoqués dans cette revue. <

Les petits ARN nucléolaires nous surprennent encore !

Yoann Abel, Guillaume Clerget,
Valérie Bourguignon-Igel, Véronique Salone,
Mathieu Rederstorff



Université de Lorraine,
CNRS, UMR 7365 IMoPA, Biopôle,
9, avenue de la forêt de Haye,
54506 Vandœuvre-lès-Nancy,
France.
mathieu.rederstorff@univ-lorraine.fr

Les cellules de tous les organismes vivants contiennent deux catégories distinctes d'acides ribonucléiques (ARN) : les ARN messagers (ARNm), qui sont traduits en protéines après leur maturation, et les ARN non codants (ARNnc). Longtemps sous-estimés, les ARNnc ont connu un regain d'intérêt considérable au cours des dernières années, en raison de leur très grand nombre et des fonctions importantes qui leur sont associées. En effet, les ARNnc participent à la plupart des processus biologiques, depuis la réplication du matériel génétique et son expression dans la cellule, jusqu'au développement de l'organisme [1]. La taille des ARNnc est très variable. Certains peuvent être constitués de milliers de bases, comme l'ARN Xist (17-kb) qui participe à l'inactivation d'un des deux chromosomes X chez les femelles des mammifères. D'autres ARNnc, dits microARN, sont d'extrême petite taille (21-23 bases) et interviennent dans l'extension de l'expression génétique (*silencing*) qui, chez l'homme, correspond à une régulation post-transcriptionnelle de l'expression des gènes [2].

Les snoARN : des molécules guide pour des enzymes de modification « aveugles »

Les petits ARN nucléolaires (snoARN) constituent une famille de plus de 300 ARN non codants, dont on pen-

sait tout connaître ou presque. En effet, la première fonction décrite pour les snoARN consiste en la modification post-transcriptionnelle d'autres ARNnc, comme les ARN ribosomiques (ARNr) ou les petits ARN nucléolaires (snARN). Ces modifications correspondent principalement à des clivages du pré-ARNr, des méthylation de nucléotides au niveau des riboses ou encore à l'isomérisation des uridines (pseudo-uridylation). La présence de motifs conservés ou boîtes, leur structure tridimensionnelle et la nature des modifications qu'ils guident, permettent de distinguer deux familles majeures de snoARN : les snoARN à boîtes C/D et les snoARN à boîtes H/ACA, guidant respectivement les réactions de méthylation ou de pseudo-uridylation (Figure 1). Une autre famille, les petits ARN spécifiques des corps de Cajal (scaARN, pour *small* Cajal ARN), qui ne sont pas à proprement parler des snoARN puisqu'ils ne sont pas localisés dans le nucléole, présentent néanmoins des boîtes C/D et/ou H/ACA et peuvent donc guider l'une ou l'autre, ou les deux types de modifications correspondantes. Localisés dans les corps de Cajal, des compartiments nucléaires non délimités par une membrane, à l'instar des nucléoles, les scaARN interviennent dans la maturation de certains petits ARN nucléolaires (snARN), tels que U1 ou U2, impliqués dans la machinerie d'épissage.

Les snoARN, comme la grande majorité des ARN non codants, ne portent pas d'activité catalytique intrinsèque et s'associent à des protéines catalytiques « aveugles », incapables de reconnaître leur cible, formant ainsi des particules ribonucléoprotéiques (RNP). Au sein de ces particules, les snoARN assurent une fonction de guide. En effet, la position à modifier dans l'ARN cible est spécifiée par la complémentarité des bases entre de courtes séquences antisens localisées dans les snoARN et les séquences de leurs ARN cibles (Figure 1).

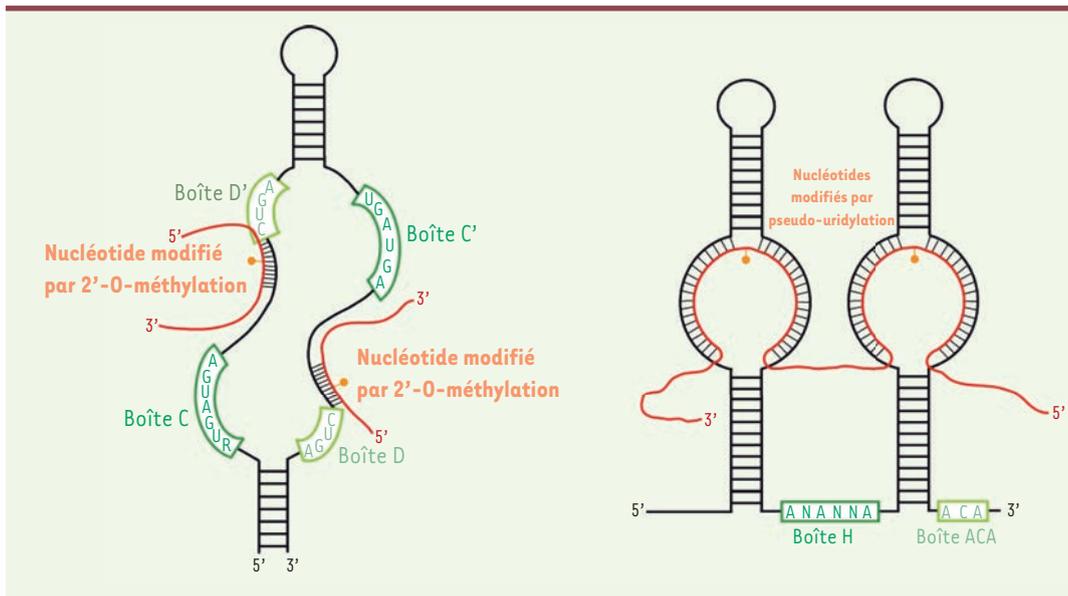


Figure 1. Représentation schématique des structures secondaires des snoARN. Les snoARN à boîtes C/D (à gauche) présentent, en amont de leurs boîtes D et D', une région antisens s'appariant à l'ARN cible (en rouge) et permettant de guider la 2'-O-méthylation du ribose du nucléotide situé cinq nucléotides en amont de la boîte (cette position est indiquée en orange)

par la fibrillarine. Les snoARN à boîtes H/ACA (à droite) présentent deux régions antisens appelées poches de pseudo-uridylation, complémentaires de l'ARN cible (en rouge). Le premier résidu non apparié dans chaque boucle (représenté en orange), une uridine, est isomérisé en pseudo-uridine par la dyskérine [42].

Parmi les snoARN, plusieurs font toutefois exception à ces règles. En effet, l'ARN U3, le snoARN à boîtes C/D le plus abondant dans la cellule, présente une particularité fonctionnelle remarquable, puisqu'il ne guide pas la méthylation d'un ARN cible, mais les clivages nucléolytiques précoces au cours de la maturation du précurseur 35S de l'ARN ribosomique, toujours par le biais d'appariements spécifiques. D'autres snoARN ne présentent, quant à eux, aucune complémentarité de séquence avec des ARN cibles potentiels classiques (ARNr, snARN) et ont ainsi été désignés snoARN orphelins. L'identification de leur fonction reste un enjeu scientifique majeur.

Les snoARN peuvent être les précurseurs d'autres ARN non codants plus petits

Les progrès en matière de séquençage à haut-débit au cours de ces dernières années ont permis de lancer des analyses à grande échelle sur les petits ARNnc humains. En 2008, une étude a montré l'existence de petits ARNnc stables, produits de maturation de snoARN précurseurs [3]. La même année, deux autres groupes ont identifié des petits ARNnc dérivés de snoARN en association avec des protéines Argonautes (Ago1 et Ago2), protéines qui s'associent aux microARN au sein des particules RISC (*RNA induced silencing complex*) impliquées dans le *silencing* des gènes [4, 5, 45]. Ainsi, une nouvelle classe de petits ARN dérivés de snoARN a été mise en évidence : les sdARN (*snoRNA-derived small RNA*, petits ARN dérivés de snoARN). Ces nouveaux petits ARN (environ 20 nucléotides pour les plus petits) peuvent être générés à partir de snoARN précurseurs à boîtes C/D ou à boîtes H/ACA [6] (Figure 2). La production et la fonction du sdARN ACA45, produit à partir du scaARN ACA45, un sous-type de snoARN à

boîtes H/ACA, ont été les premières documentées dans la littérature scientifique. En effet, le scaARN ACA45, dont la structure en tige-boucle rappelle celle des précurseurs de microARN, subit une maturation en sdARN (sdARN ACA45) de 22-23 bases, comparable à celle d'un microARN. Certains auteurs emploient même le terme de sno-miR. Ender *et al.* [4] ont montré que cette maturation était, dans ce cas, indépendante de la RNase de type III Drosha, mais nécessitait l'action de la RNase de type III Dicer, toutes deux impliquées dans la maturation classique des microARN [45]. Enfin, les auteurs ont montré, par un test rapporteur, que le sdARN ACA45 pouvait cibler l'ARNm codant pour une sous-unité protéique du médiateur CDC2L6. Depuis, d'autres sdARN ont été identifiés par des immunoprécipitations de complexes argonaute-ARN, ou prédits à l'aide de modélisations bio-informatiques. Ainsi, les protéines Ago7 (chez *Arabidopsis*) et Ago1 (chez *S. pombe*) ont également été trouvées en complexe avec de petits ARN dérivés de snoARN [6]. Plus récemment, chez l'homme, plusieurs fragments dérivés de snoARN ont aussi été observés au sein de particules RISC [7]. Une autre étude montre que les sdARN semblent ne pas être incorporés dans les complexes Ago2 de manière très efficace [8]. Il n'en demeure pas moins que certaines espèces de snoARN présentent donc à la fois les caractéristiques et les fonctions des snoARN classiques, et peuvent générer des sno-miR fonctionnels [9]. Une hypothèse qui permet d'expliquer ces observations est que les précurseurs

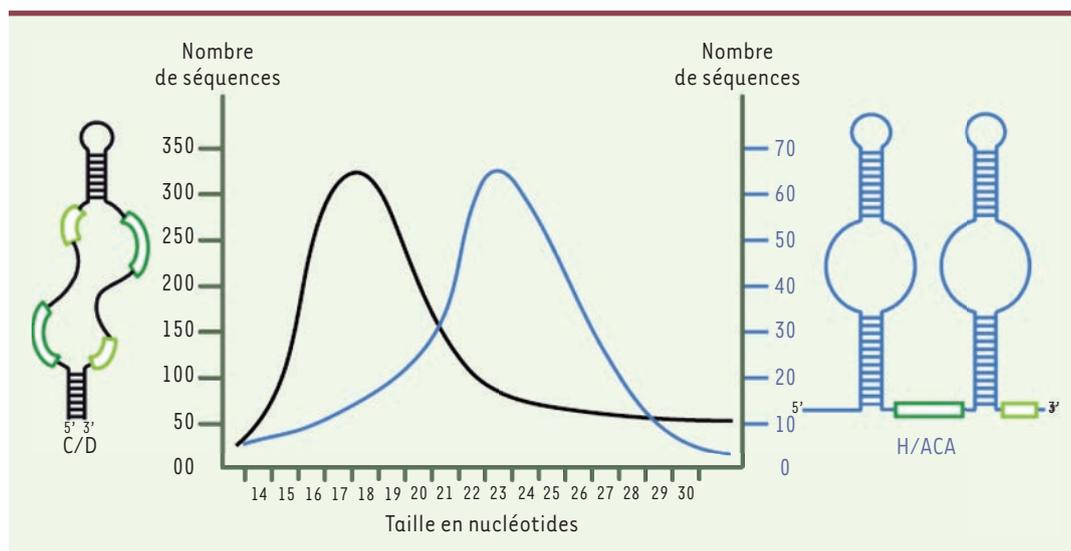


Figure 2. Caractéristiques des miARN issus de snoARN à boîtes C/D ou H/ACA chez l'homme. Les miARN dérivés de snoARN à boîtes C/D sont plus abondants que ceux dérivés de snoARN à boîtes H/ACA. Ces derniers ont toutefois une taille moyenne légèrement supérieure (21-24 contre 17-19 nucléotides, abscisse). En ordon-

née : nombre de séquences correspondant à des sno-miR pour un million de séquences localisées sur le génome (d'après [6]).

de certains microARN dériveraient de snoARN ayant perdu leur fonction initiale de guide de modification post-transcriptionnelle d'ARN, et acquis une nouvelle fonctionnalité au cours de l'évolution. Cette hypothèse est étayée par le fait que de nombreux snoARN orphelins ont été décrits comme étant les précurseurs de sdARN [6].

Ce lien surprenant entre snoARN et microARN est conforté par l'existence de plusieurs similitudes entre les processus de maturation des microARN et des snoARN. En effet, ils nécessitent tous deux l'action de nucléases (endo- et exonucléases), notamment de type RNase III (Rnt1 chez la levure et les plantes pour la maturation des snoARN polycistroniques, Drosha et Dicer pour la maturation des microARN). De plus, il a été montré que Dicer pouvait accepter certains snoARN comme substrats [4]. En outre, des interactions ont été observées entre des protéines participant à l'une ou l'autre voie de maturation. Ainsi, la fibrillarine et la protéine NOP56 (*56-kDa nucleolar protein*) ont été retrouvées associées aux protéines Argonautes Ago2 et Ago1, respectivement. Enfin, chez les plantes, on a montré que la protéine chaperonne Hsp90 (*90-kDa heat shock protein*) intervenait à la fois dans la biogenèse des snoARN et celle des microARN [10].

Les pré-ARN messagers : de nouvelles cibles pour les snoARN

Les snoARN à boîtes C/D SNORD115 et SNORD116 murins ont été identifiés en 2000 par le groupe d'Hüttenhofer [11]. Ils sont tous deux exprimés préférentiellement dans le cerveau et possèdent plusieurs caractéristiques originales. En effet, les gènes *SNORD115* et *SNORD116* sont localisés dans le locus 7C chez la souris, et 15q11-13 chez l'homme. Ces régions chromosomiques sont soumises à l'empreinte génomique parentale, et ces gènes ne sont exprimés qu'à partir de l'allèle paternel [12, 13]. De plus, de multiples copies de ces gènes sont présentes (respectivement 47 et 27 répétitions des gènes humains *SNORD115* et *SNORD116*). Chez l'homme, la délétion de l'allèle paternel de ce locus conduit au développement du syndrome de Prader-Willi (locus PW/

AS), une maladie génétique rare (prévalence de 1 sur 25 000) entraînant des troubles du développement (taille réduite, hypogonadisme) et du comportement (hypotonie, retard mental, hyperphagie due à l'absence de sensation de la satiété conduisant à une obésité morbide) [14-16]. Des études génétiques ont montré que la seule délétion de la région correspondant aux copies de *SNORD116* était suffisante pour induire un phénotype pathologique ; toutefois, la sévérité des troubles augmente avec l'étendue de la région absente. Aucune cible canonique du snoARN SNORD116 n'a pu être identifiée à ce jour. On peut donc proposer que ce dernier, tout comme d'autres snoARN orphelins, participe à des processus biologiques encore non identifiés.

Par ailleurs, le snoARN SNORD115 ne semble pas jouer un rôle essentiel dans l'apparition du syndrome Prader-Willi, puisque la seule délétion du *cluster* SNORD115 n'entraîne aucun symptôme chez les individus qui en sont porteurs. Il n'est cependant pas exclu que la perte de sa fonction, associée à celle de SNORD116, contribue à la variabilité du phénotype. Néanmoins, la compréhension de la responsabilité des snoARN dans l'apparition de la pathologie demeure partielle, et l'implication possible d'un nouveau gène, toujours au même locus, a été rapportée [17].

De manière remarquable, le snoARN orphelin SNORD115 possède une longue complémentarité de 18 nucléotides avec un pré-ARNm spécifique du cerveau codant pour le récepteur de la sérotonine 5-HT_{2c} [11]. Ce récepteur transmembranaire couplé à la protéine G est impliqué dans le contrôle de l'appétit et de l'humeur. Le pré-ARNm codant pour ce récepteur peut subir, dans le cytoplasme, différents événements d'édition

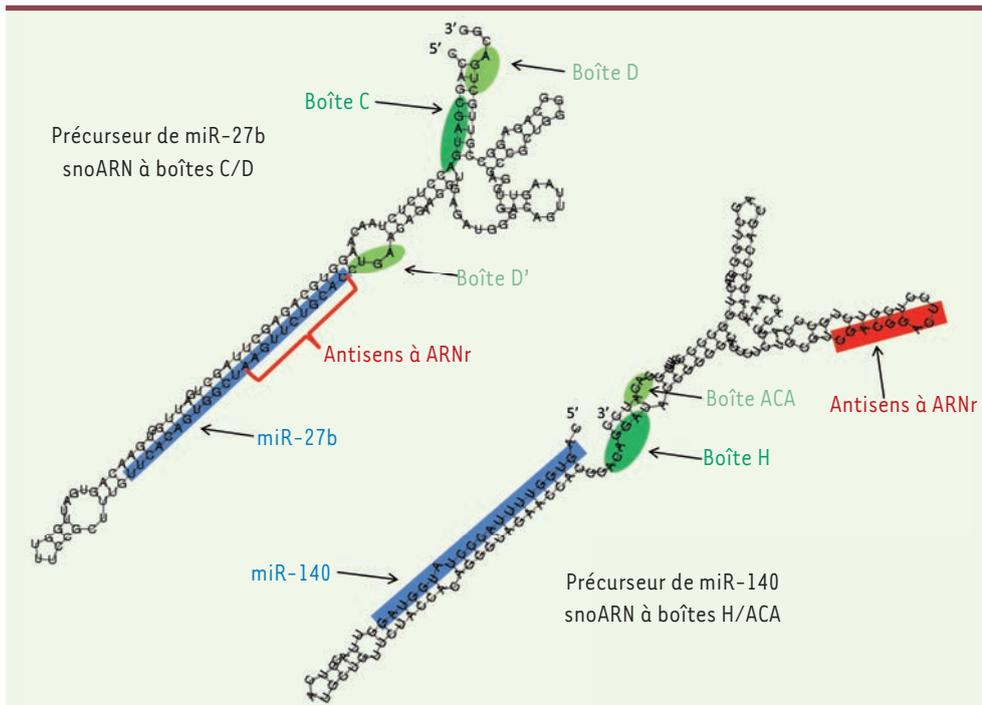


Figure 3. Les précurseurs des miARN 27b et 140 ressemblent à des snoARN. Les structures secondaires des précurseurs des miARN miR-27b et miR-140 sont représentées, avec les boîtes C/D ou H/ACA indiquées en vert et la région antisens complémentaire à une région cible dans un ARN ribosomique en rouge. Le miARN mature correspondant est représenté en bleu [22, 43, 44].

vés du snoARN SNORD115 ont été observés [19-21]. Les détails concernant leur nombre, abondance et extrémités sont l'objet de débats [19-21]; ils pourraient néanmoins être

(conversion d'adénosine en inosine *via* les enzymes ADAR [*adenosine deaminase acting on RNA*] -1 et -2) au niveau de cinq sites spécifiques (A à E) de l'exon V [13]. Cet exon V correspond à la boucle II intracellulaire du récepteur à laquelle s'associe la protéine G indispensable à la transduction du signal par la sérotonine. Au cours de la traduction en protéines, les inosines de l'ARNm sont reconnues comme des guanosines. Aussi, en fonction du taux d'édition des différents sites, plusieurs isoformes du récepteur peuvent être produites, chacune présentant une affinité différente pour la protéine G. Le snoARN SNORD115 pourrait ainsi guider la méthylation d'un nucléotide correspondant au site d'édition C. Les expérimentations *in vitro* de Hüttenhofer et son équipe vont dans ce sens, qui montrent que l'expression du snoARN SNORD115 inhibait spécifiquement la déamination de l'adénosine au site C [13]. Ainsi, en réduisant l'efficacité de l'édition au site C, le snoARN SNORD115 pourrait moduler finement le couplage du récepteur 5-HT_{2c} à la protéine G, et donc la transduction du signal par la sérotonine. Ainsi, une nouvelle classe de cibles possibles pour les snoARN, les pré-ARNm, a été mise en évidence et a engendré un regain d'intérêt important pour l'étude de cette famille de molécules. En outre, une autre étude [18] semble indiquer que le snoARN SNORD115 serait impliqué dans la régulation de l'épissage alternatif du pré-ARNm codant le récepteur 5-HT_{2c}, selon un mécanisme indépendant de l'édition. En effet, en présence du snoARN SNORD115, l'ARN mature produit contiendrait l'exon V en entier, conduisant à la forme fonctionnelle du récepteur à la sérotonine. En revanche, en son absence, un site proximal d'épissage au niveau de l'exon V serait utilisé, conduisant à l'apparition d'un codon stop prématuré dans l'exon VI. Le récepteur ainsi produit est amputé de sa partie carboxy-terminale, et son affinité pour la protéine G serait alors 10 à 100 fois moins importante. Il est intéressant de mentionner que des petits ARN déri-

impliqués dans l'épissage du pré-ARNm du récepteur de la sérotonine, ainsi que dans celui d'autres pré-ARNm cibles potentiels [20]. Des observations similaires ont été faites pour d'autres snoARN et leurs petits fragments dérivés, comme SNORD88C, impliqué dans l'épissage du pré-ARNm de FGFR3 (*fibroblast growth factor receptor 3*) [22]. Ceci indique que le snoARN SNORD115, comme de nombreux snoARN, serait le précurseur de petits ARNnc fonctionnels, dont le nombre, la nature, les voies de biogenèse ainsi que la fonction exacte devront être caractérisés plus précisément [19]. Bien que l'implication éventuelle des snoARN ou de leurs petits ARN dérivés dans les phénomènes d'épissage ou d'édition de pré-ARNm soient encore mal comprises, l'élucidation des mécanismes biologiques et l'identification des partenaires mis en jeu lors de la maturation des sno-miR pourraient permettre de mieux comprendre les mécanismes de pathologies complexes comme, par exemple, le syndrome de Prader-Willi.

snoARN et cancer ?

Le rôle des microARN dans les processus de différenciation cellulaire, mais également dans le développement de pathologies, en particulier l'évolution tumorale, dont ils peuvent servir de marqueurs, est aujourd'hui largement documenté [23, 24]. À l'instar des microARN, des snoARN sont également retrouvés parmi les ARN circulants, représentant des biomarqueurs potentiels intéressants [25, 26], et de nouveaux résultats tendent

à indiquer que plusieurs snoARN sont également impliqués dans des processus tumoraux [27-31]. Par exemple, le snoARN à boîtes C/D U50 est sous-exprimé chez certains patients atteints de cancer de la prostate, et sa surexpression inhibe la prolifération cellulaire *in vitro*. [32-34]. Un autre snoARN, à boîte H/ACA cette fois, est surexprimé dans certains types de cancers du poumon [25, 35]. Ainsi, des membres des deux familles de snoARN semblent impliqués dans la cancérogenèse, où ils peuvent se comporter comme oncogènes ou comme gènes suppresseurs de tumeurs. Bien que les mécanismes précis afférents à ces observations soient encore mal compris, une hypothèse attrayante est que l'existence de défauts de maturation des ARN ribosomiques, cibles canoniques des snoARN, et leur assemblage dans les ribosomes, puissent affecter la traduction protéique dans les cellules. Or, des défauts d'assemblage et de biogenèse des ribosomes ont été associés à l'évolution de processus tumoraux, en particulier chez des patients atteints de cancers du sein [36]. Une hypothèse alternative serait que les snoARN concernés, ou des sous-fragments de ces snoARN, puissent intervenir en dehors de la biogenèse des ribosomes, expliquant les effets observés au niveau tumoral. Les phénomènes de production de microARN à partir de snoARN précurseurs décrits plus haut pourraient en être une illustration. Les données de la littérature montrent que deux microARN issus de précurseurs ressemblant à des snoARN, les miR-27b et miR-140, ont une forte influence sur l'évolution de certaines tumeurs [37-39] (Figure 3). Ainsi, une expression élevée de miR-140 dans les ostéosarcomes et dans les tumeurs du côlon ralentit la prolifération cellulaire, et entraîne la chimiorésistance de la tumeur [40]. La surexpression de miR-27b dans des neuroblastomes ralentit également la prolifération cellulaire par une répression de l'expression du gène codant pour les récepteurs PPAR γ (*peroxisome proliferator-activated receptor γ*) [41].

Conclusion

La famille des snoARN illustre parfaitement l'importance de la compréhension des mécanismes conduisant à la maturation ou l'assemblage des ARNnc au sein de particules ribonucléoprotéiques fonctionnelles. Elle constitue un domaine d'étude aussi fascinant que primordial. Ainsi, les snoARN, ces « anciens » ARN auxquels de nouveaux rôles sont attribués, ne cessent de nous surprendre. Leur maturation en ARN fonctionnels plus petits ou la démonstration de leur implication, ainsi que celle de leurs petits ARN dérivés récemment identifiés, dans la régulation de l'édition et l'épissage alternatif, ou le lien possible de certains d'entre eux avec le syndrome de Prader-Willi ou les processus tumoraux, en sont des exemples. L'élucidation des mécanismes de maturation des snoARN en microARN pourrait ainsi permettre d'envisager, à plus long terme, des approches thérapeutiques pour ces pathologies. \diamond

SUMMARY

Beyond usual functions of snoRNA

Small nucleolar RNAs or snoRNAs, principally implicated in post-transcriptional chemical modification of other RNAs, were among the first

non-coding RNA identified, together with ribosomal and transfer RNA. Lately, snoRNA have been involved in various unexpected functions, which renewed researcher's interest for these molecules. SnoRNA processing into smaller functional RNA species (sdRNA for snoRNA-derived RNA) or into miRNA (sno-miR), snoRNA mediated regulation of messenger RNA alternative splicing or snoRNA links to human disorders, including cancers, are some of the topics developed in this review. \diamond

LIENS D'INTÉRÊT

Les auteurs déclarent n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

REMERCIEMENTS

Nous tenons à remercier le Centre National de la Recherche Scientifique (CNRS), l'Université de Lorraine (UL) et la Ligne nationale contre le cancer (comités Meurthe-et-Moselle et Moselle) pour leur soutien financier. Nous tenons également à remercier Christiane Branlant et Bruno Charpentier pour leur relecture et commentaires sur le manuscrit ainsi que l'ensemble du laboratoire pour les discussions stimulantes.

RÉFÉRENCES

1. Rederstorff M. Une approche originale de sélection de nouveaux ARN non codants. *Med Sci (Paris)* 2011 ; 27 : 343-5.
2. Cavaille J. Des microARN comme s'il en pleuvait... *Med Sci (Paris)* 2004 ; 20 : 399-401.
3. Kawaji H, Nakamura M, Takahashi Y, et al. Hidden layers of human small RNAs. *BMC Genomics* 2008 ; 9 : 157.
4. Ender C, Krek A, Friedlander MR, et al. A human snoRNA with microRNA-like functions. *Mol Cell* 2008 ; 32 : 519-28.
5. Saraiya AA, Wang CC. snoRNA, a novel precursor of microRNA in *Giardia lamblia*. *PLoS Pathog* 2008 ; 4 : e1000224.
6. Taft RJ, Glazov EA, Lassmann T, et al. Small RNAs derived from snoRNAs. *RNA* 2009 ; 15 : 1233-40.
7. Brameier M, Herwig A, Reinhardt R, et al. Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. *Nucleic Acids Res* 2011 ; 39 : 675-86.
8. Kishore S, Gruber AR, Jedlinski DJ, et al. Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biol* 2013 ; 14 : R45.
9. Scott MS, Ono M. From snoRNA to miRNA: Dual function regulatory non-coding RNAs. *Biochimie* 2011 ; 93 : 1987-92.
10. Iwasaki S, Kobayashi M, Yoda M, et al. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol Cell* 2010 ; 39 : 292-9.
11. Cavaille J, Buiting K, Kiefmann M, et al. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci USA* 2000 ; 97 : 14311-6.
12. Rederstorff M, Huttenhofer A. Small non-coding RNAs in disease development and host-pathogen interactions. *Curr Opin Mol Ther* 2010 ; 12 : 684-94.
13. Vitali P, Cavaille J. Petits ARN C/D et syndrome de Prader-Willi. *Med Sci (Paris)* 2005 ; 21 : 1017-9.
14. De Smith AJ, Purmann C, Walters RG, et al. A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum Mol Genet* 2009 ; 18 : 3257-65.
15. Duker AL, Ballif BC, Bawle EV, et al. Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur J Hum Genet* 2010 ; 18 : 1196-201.
16. Sahoo T, del Gaudio D, German JR, et al. Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet* 2008 ; 40 : 719-21.

RÉFÉRENCES

17. Schaaf CP, Gonzalez-Garay ML, Xia F, et al. Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and autism. *Nat Genet* 2013 ; 45 : 1405-8.
18. Kishore S, Stamm S. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* 2006 ; 311 : 230-2.
19. Bortolin-Cavaille ML, Cavaille J. The SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) gene clusters at the imprinted Prader-Willi locus generate canonical box C/D snoRNAs. *Nucleic Acids Res* 2012 ; 40 : 6800-7.
20. Kishore S, Khanna A, Zhang Z, et al. The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Hum Mol Genet* 2010 ; 19 : 1153-64.
21. Rederstorff M, Bernhart SH, Tanzer A, et al. RNPomics: defining the ncRNA transcriptome by cDNA library generation from ribonucleo-protein particles. *Nucleic Acids Res* 2010 ; 38 : e113.
22. Scott MS, Ono M, Yamada K, et al. Human box C/D snoRNA processing conservation across multiple cell types. *Nucleic Acids Res* 2012 ; 40 : 3676-88.
23. Kahn A. L'impérialisme des micro-ARN s'étend maintenant au cancer. *Med Sci (Paris)* 2005 ; 21 : 687-9.
24. Ladeiro Y, Zucman-Rossi J. Micro-ARN (miARN) et cancer : le cas des tumeurs hépatocellulaires. *Med Sci (Paris)* 2009 ; 25 : 467-72.
25. Liao J, Yu L, Mei Y, et al. Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. *Mol Cancer* 2010 ; 9 : 198.
26. Zhang L, Yang M, Marks P, et al. Serum non-coding RNAs as biomarkers for osteoarthritis progression after ACL injury. *Osteoarthritis Cartilage* 2012 ; 20 : 1631-7.
27. Martens-Uzunova ES, Olvedy M, Jenster G. Beyond microRNA - Novel RNAs derived from small non-coding RNA and their implication in cancer. *Cancer Lett* 2013 ; 340 : 201-11.
28. Su H, Xu T, Ganapathy S, et al. Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene* 2013 ; doi : 10.1038/onc.2013.89.
29. Teittinen KJ, Laiho A, Uusimäki A, et al. Expression of small nucleolar RNAs in leukemic cells. *Cell Oncol (Dordr)* 2013 ; 36 : 55-63.
30. Valleron W, Ysebaert L, Berquet L, et al. Small nucleolar RNA expression profiling identifies potential prognostic markers in peripheral T-cell lymphoma. *Blood* 2013 ; 120 : 3997-4005.
31. Williams GT, Farzaneh F. Are snoRNAs and snoRNA host genes new players in cancer? *Nat Rev Cancer* 2012 ; 12 : 84-8.
32. Dong XY, Guo P, Boyd J, et al. Implication of snoRNA U50 in human breast cancer. *J Genet Genomics* 2009 ; 36 : 447-54.
33. Dong XY, Rodriguez C, Guo P, et al. SnoRNA U50 is a candidate tumor-suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. *Hum Mol Genet* 2008 ; 17 : 1031-42.
34. Soeno Y, Fujita K, Kudo T, et al. Generation of a mouse model with down-regulated U50 snoRNA (SNORD50) expression and its organ-specific phenotypic modulation. *PLoS One* 2013 ; 8 : e72105.
35. Mei YP, Liao JP, Shen J, et al. Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. *Oncogene* 2011 ; 31 : 2794-804.
36. Belin S, Beghin A, Solano-Gonzalez E, et al. Dysregulation of ribosome biogenesis and translational capacity is associated with tumor progression of human breast cancer cells. *PLoS One* 2009 ; 4 : e7147.
37. Buffa FM, Camps C, Winchester L, et al. microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer Res* 2011 ; 71 : 5635-45.
38. Husted S, Sokilde R, Rask L, et al. MicroRNA expression profiles associated with development of drug resistance in Ehrlich ascites tumor cells. *Mol Pharm* 2011 ; 8 : 2055-62.
39. Susuki D, Kimura S, Naganuma S, et al. Regulation of microRNA expression by hepatocyte growth factor in human head and neck squamous cell carcinoma. *Cancer Sci* 2011 ; 102 : 2164-71.
40. Song B, Wang Y, Xi Y, et al. Mechanism of chemoresistance mediated by miR-140 in human osteosarcoma and colon cancer cells. *Oncogene* 2009 ; 28 : 4065-74.
41. Lee JJ, Drakaki A, Iliopoulos D, Struhl K. MiR-27b targets PPARgamma to inhibit growth, tumor progression and the inflammatory response in neuroblastoma cells. *Oncogene* 2011 ; 31 : 3818-25.
42. Eddy SR. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2001 ; 2 : 919-29.
43. Ono M, Scott MS, Yamada K, et al. Identification of human miRNA precursors that resemble box C/D snoRNAs. *Nucleic Acids Res* 2011 ; 39 : 3879-91.
44. Scott MS, Avolio F, Ono M, et al. Human miRNA precursors with box H/ACA snoRNA features. *PLoS Comput Biol* 2009 ; 5 : e1000507.
45. Dunoyer P. La bataille du silence. *Med Sci (Paris)* 2009 ; 25 : 505-12.

TIRÉS À PART

M. Rederstorff

Chapter 1

Small Non-Coding RNAs: A Quick Look in the Rearview Mirror

Guillaume Clerget, Yoann Abel, and Mathieu Rederstorff

Abstract

The revolution of miRNA discovery, in the early 2000s, shed a new light in the exciting field of small non-coding RNAs. Since then, and owing to outstanding breakthroughs in RNomic techniques, novel small non-coding RNA families have been regularly discovered, e.g., piRNAs, tiRNAs, and many others.

In this review, we provide a very succinct historical and functional overview on most prominent small non-coding RNA families.

Key words miRNA, snoRNA, piRNA, tRNA, rRNA, tiRNA, snRNA

1 Introduction

1.1 Dawn of Small Non-Coding RNAs

It has been more than 60 years that the first soluble cytoplasmic small non-coding RNA (sncRNA), a yeast alanine transfer RNA (tRNA), was identified and its primary structure determined [1, 2]. Together with ribosomal RNAs (rRNA), the nucleic acid moieties of the ribosomes, tRNAs play a central role in protein synthesis, as the adaptor molecules providing the correct amino acid in response to a specific codon on the mRNA [3]. Only 20 years later were the first nuclear RNAs uncovered: uridine-rich RNAs, or U-RNAs, were observed to be very abundant, nucleus or sometimes even nucleolus specific, and highly conserved [4, 5]; small nuclear RNAs (snRNA) U1, U2, U4, U5, and U6 were later shown to be involved in splicing, in association with several proteins within a ribonucleoprotein particle (RNP) called spliceosome [6]. Base pairing among snRNAs and between snRNAs and pre-messenger RNA (pre-mRNA) regions account for the proper conformation of both spliceosome and pre-mRNA, eventually leading to release of the introns and mature mRNAs [7]. Interestingly, most small nucleolar RNAs (snoRNAs) are processed from the released introns [8]. Within snoRNPs,

Author contributed equally with all other contributors.

Mathieu Rederstorff (ed.), *Small Non-Coding RNAs: Methods and Protocols*, Methods in Molecular Biology, vol. 1296, DOI 10.1007/978-1-4939-2547-6_1, © Springer Science+Business Media New York 2015

snoRNAs guide posttranscriptional RNA modifications of other ncRNAs, e.g., rRNAs and snRNAs: 2'-O-methylations for C/D box snoRNAs and pseudouridylation for H/ACA box snoRNAs. Some snoRNAs, such as U3 or U8, are instead involved in early cleavages of pre-rRNA during ribosome biogenesis in the nucleolus [9].

In the late 1990s, pioneering studies aimed at identifying novel non-coding RNAs using experimental approaches combining library generation and sequencing identified numerous novel snoRNAs in various models including human, using material that was usually discarded: small total RNA in the size range of 50–500 nucleotides [10]. At that time, why would one have looked for smaller functional RNAs? It would both have been nonsense and a waste of time. Nonsense? Actually no... antisense!

1.2 miRNAs: The Revolution!

In the early 1990s, the *lin-14* gene was shown to be involved in temporal development in the worm *C. elegans*. The *lin-4* gene product, most probably a polypeptide, was shown to negatively regulate *lin-14* expression by directly acting on the 3' untranslated region of its pre-mRNA [11]. In 1993, both Ambros and Ruvkun labs discovered that the *lin-4* gene product actually was a small non-coding RNA, of 22 nucleotides in length in its mature form, targeting the 3' UTR of the *lin-14* pre-mRNA several times, owing to imperfect antisense base pairing [12, 13]. Unfortunately enough, apparently, the *lin-4* small RNA neither had additional counterparts in worm nor was it found to be conserved in other species. However, in 2000, Pasquinelli and coworkers identified that *let-7*, another 22-nucleotide-long small temporal RNA (stRNA) involved in *C. elegans* developmental timing, was conserved in human and drosophila, anticipating the likely discovery of many other small regulatory RNAs [14]. Revolution came less than 1 year later, when Tuschl, Bartel, and Ambros labs reported, in the same issue of Science, the observation of an abundant class of novel small regulatory RNAs in worm, human, and drosophila, termed since microRNAs (miRNAs) [15–17]. Even certain viruses were next shown to encode their own miRNAs [18, 19].

It was rapidly observed that the miRNA maturation and assembly machinery were the same as the one implicated in the formation of short interfering RNAs (siRNAs), involved in the RNA interference (RNAi) pathway described in worm a couple of years earlier by the 2006 Nobel Prize awardees Andrew Fire and Craig Mello [20, 21] and even earlier in plants [22]. miRNA/siRNA maturation, from a double-stranded precursor, is now well understood [23] but still open to surprises; recently it was discovered that miRNAs could be processed from other types of ncRNAs, such as snoRNAs [24, 25].

In human, siRNAs, within the RNA-induced silencing complex (RISC), mediate perfectly complementary mRNA target cleavages

[26, 27]. On the other hand, miRNAs regulate gene expression by imperfectly base pairing to the 3' UTR of the targeted mRNA, repressing its translation, which eventually leads to the mRNA decay by different possible mechanisms [28, 29]. miRNAs can act as regulators of epigenetic modifications as well, via DNA or histone modifications, therefore controlling gene expression at the transcriptional level [30].

1.3 Small Non-Coding RNAs Return!

Another sncRNA family, the Piwi-interacting RNAs (piRNA), discovered about half a decade after miRNAs, also regulates transcription [31]. piRNAs are 27–29-nucleotide-long RNAs. They prevent transposon dissemination in germinal cell lines, owing to an original “ping-pong” mechanism [32, 33]. The smallest eukaryotic ncRNAs known to date, the transcription initiation RNAs (tiRNAs), sized 17–18 nucleotides, are generated upon stalling or backtracking of RNA polymerase II (RNAPII) near the transcription start sites (TSSs) and might be involved in transcription initiation regulation as well [34, 35]. TSSs appear to be the source of bunches of small and long non-coding RNAs (lncRNAs) [36], such as transcription start site-associated RNAs (TSSa-RNAs) and promoter-associated RNAs (PASRs) in animals [37] or promoter upstream transcripts (PROMPTs) [38] and cryptic/XrnI-sensitive unstable transcripts (CUTs/XUTs) in yeast [39–41]. While lncRNAs are generally involved in epigenetic and chromatin modifications [42–44], the functionality of short transcripts is debated. They might be involved in maintaining chromatin into a transcriptionally active state or in maintaining RNAPII available near TSSs [45].

1.4 Small Non-Coding RNA: Not the End

Most if not all the human genome appears to be actively transcribed [46–48]. Nevertheless, most transcripts rather belong to the TUF (transcripts of unknown function) rather than to the ncRNA family. Progresses in transcriptomics/RNomics enabled the identification of thousands of short and long transcripts [49–55]. Many of them were attributed a function but it is only the top of the iceberg. Whether every TUF is indeed functional is debated, and experimental efforts will be mandatory to answer this question.

There are no doubts that novel ncRNAs and novel ncRNA families will still be discovered in the (next) future.

Acknowledgment

This work was supported by the Centre National pour la Recherche Scientifique, the Université de Lorraine, the Région Lorraine and La Ligue contre le Cancer.

References

1. Holley RW, Apgar J, Doctor BP (1960) Separation of amino acid-specific "soluble"-fraction ribonucleic acids. *Ann N Y Acad Sci* 88:745–751
2. Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, Penswick JR, Zamir A (1965) Structure of a ribonucleic acid. *Science* 147(3664):1462–1465
3. Eddy SR (2001) Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2(12):919–929. doi:[10.1038/35103511](https://doi.org/10.1038/35103511), 35103511 [pii]
4. Busch H, Reddy R, Rothblum L, Choi YC (1982) SnRNAs, SnRNPs, and RNA processing. *Annu Rev Biochem* 51:617–654. doi:[10.1146/annurev.bi.51.070182.003153](https://doi.org/10.1146/annurev.bi.51.070182.003153)
5. Weinberg RA, Penman S (1968) Small molecular weight monodisperse nuclear RNA. *J Mol Biol* 38(3):289–304, doi:0022-2836(68)90387-2 [pii]
6. Lerner MR, Steitz JA (1979) Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc Natl Acad Sci U S A* 76(11):5495–5499
7. Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA (1980) Are snRNPs involved in splicing? *Nature* 283(5743):220–224
8. Matera AG, Terns RM, Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* 8(3):209–220. doi:[10.1038/nrm2124](https://doi.org/10.1038/nrm2124), nrm2124 [pii]
9. Marmier-Gourrier N, Clery A, Schlotter F, Senty-Segault V, Branlant C (2011) A second base pair interaction between U3 small nucleolar RNA and the 5'-ETS region is required for early cleavage of the yeast pre-ribosomal RNA. *Nucleic Acids Res* 39(22):9731–9745. doi:[10.1093/nar/gkr675](https://doi.org/10.1093/nar/gkr675), gkr675 [pii]
10. Huttenhofer A, Kiefmann M, Meier-Ewert S, O'Brien J, Lehrach H, Bachellerie JP, Brosius J (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J* 20(11):2943–2953. doi:[10.1093/emboj/20.11.2943](https://doi.org/10.1093/emboj/20.11.2943)
11. Wightman B, Burglin TR, Gatto J, Arasu P, Ruvkun G (1991) Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. *Genes Dev* 5(10):1813–1824
12. Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75(5):843–854, doi:0092-8674(93)90529-Y [pii]
13. Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* 75(5):855–862, doi:0092-8674(93)90530-4 [pii]
14. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Muller P, Spring J, Srinivasan A, Fishman M, Finnerty J, Corbo J, Levine M, Leahy P, Davidson E, Ruvkun G (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408(6808):86–89. doi:[10.1038/35040556](https://doi.org/10.1038/35040556)
15. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. *Science* 294(5543):853–858. doi:[10.1126/science.1064921](https://doi.org/10.1126/science.1064921), 294/5543/853 [pii]
16. Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294(5543):858–862. doi:[10.1126/science.1065062](https://doi.org/10.1126/science.1065062), 294/5543/858 [pii]
17. Lee RC, Ambros V (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294(5543):862–864. doi:[10.1126/science.1065329](https://doi.org/10.1126/science.1065329), 294/5543/862 [pii]
18. Nathans R, Chu CY, Serquina AK, Lu CC, Cao H, Rana TM (2009) Cellular microRNA and P bodies modulate host-HIV-1 interactions. *Mol Cell* 34(6):696–709. doi:[10.1016/j.molcel.2009.06.003](https://doi.org/10.1016/j.molcel.2009.06.003), S1097-2765(09)00394-3 [pii]
19. Pfeffer S, Zavolan M, Grasser FA, Chien M, Russo JJ, Ju J, John B, Enright AJ, Marks D, Sander C, Tuschl T (2004) Identification of virus-encoded microRNAs. *Science* 304(5671):734–736. doi:[10.1126/science.1096781](https://doi.org/10.1126/science.1096781), 304/5671/734 [pii]
20. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391(6669):806–811. doi:[10.1038/35888](https://doi.org/10.1038/35888)
21. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293(5531):834–838. doi:[10.1126/science.1062961](https://doi.org/10.1126/science.1062961), 1062961 [pii]
22. Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a chimeric chalcone synthase

- gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* 2(4):279–289. doi:[10.1105/tpc.2.4.279](https://doi.org/10.1105/tpc.2.4.279), 2/4/279 [pii]
23. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116(2):281–297, doi: [S0092867404000455](https://doi.org/S0092867404000455) [pii]
 24. Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G (2008) A human snoRNA with microRNA-like functions. *Mol Cell* 32(4):519–528. doi:[10.1016/j.molcel.2008.10.017](https://doi.org/10.1016/j.molcel.2008.10.017), S1097-2765(08)00733-8 [pii]
 25. Rother S, Meister G (2011) Small RNAs derived from longer non-coding RNAs. *Biochimie* 93(11):1905–1915. doi:[10.1016/j.biochi.2011.07.032](https://doi.org/10.1016/j.biochi.2011.07.032), S0300-9084(11)00290-2 [pii]
 26. Tuschl T, Zamore PD, Lehmann R, Bartel DP, Sharp PA (1999) Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev* 13(24):3191–3197
 27. Zamore PD, Tuschl T, Sharp PA, Bartel DP (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101(1):25–33. doi:[10.1016/S0092-8674\(00\)80620-0](https://doi.org/10.1016/S0092-8674(00)80620-0), S0092-8674(00)80620-0 [pii]
 28. Braun JE, Huntzinger E, Izaurralde E (2013) The role of GW182 proteins in miRNA-mediated gene silencing. *Adv Exp Med Biol* 768:147–163. doi:[10.1007/978-1-4614-5107-5_9](https://doi.org/10.1007/978-1-4614-5107-5_9)
 29. Eulalio A, Tritschler F, Izaurralde E (2009) The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. *RNA* 15(8):1433–1442. doi:[10.1261/rna.1703809](https://doi.org/10.1261/rna.1703809), rna.1703809 [pii]
 30. Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* 303(5658):672–676. doi:[10.1126/science.1093686](https://doi.org/10.1126/science.1093686), 1093686 [pii]
 31. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, Chien M, Russo JJ, Ju J, Sheridan R, Sander C, Zavolan M, Tuschl T (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442(7099):203–207. doi:[10.1038/nature04916](https://doi.org/10.1038/nature04916), nature04916 [pii]
 32. Chuma S, Pillai RS (2009) Retrotransposon silencing by piRNAs: ping-pong players mark their sub-cellular boundaries. *PLoS Genet* 5(12):e1000770. doi:[10.1371/journal.pgen.1000770](https://doi.org/10.1371/journal.pgen.1000770)
 33. Siomi MC, Sato K, Pezic D, Aravin AA (2011) PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* 12(4):246–258. doi:[10.1038/nrm3089](https://doi.org/10.1038/nrm3089), nrm3089 [pii]
 34. Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K, Irvine K, Arakawa T, Nakamura M, Kubosaki A, Hayashida K, Kawazu C, Murata M, Nishiyori H, Fukuda S, Kawai J, Daub CO, Hume DA, Suzuki H, Orlando V, Carninci P, Hayashizaki Y, Mattick JS (2009) Tiny RNAs associated with transcription start sites in animals. *Nat Genet* 41(5):572–578. doi:[10.1038/ng.312](https://doi.org/10.1038/ng.312), ng.312 [pii]
 35. Valen E, Preker R, Andersen PR, Zhao X, Chen Y, Ender C, Dueck A, Meister G, Sandelin A, Jensen TH (2011) Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nat Struct Mol Biol* 18(9):1075–1082. doi:[10.1038/nsmb.2091](https://doi.org/10.1038/nsmb.2091), nsmb.2091 [pii]
 36. Jacquier A (2009) The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat Rev Genet* 10(12):833–844. doi:[10.1038/nrg2683](https://doi.org/10.1038/nrg2683), nrg2683 [pii]
 37. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA (2008) Divergent transcription from active promoters. *Science* 322(5909):1849–1851. doi:[10.1126/science.1162253](https://doi.org/10.1126/science.1162253), 1162253 [pii]
 38. Flynn RA, Almada AE, Zamudio JR, Sharp PA (2011) Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc Natl Acad Sci U S A* 108(26):10460–10465. doi:[10.1073/pnas.1106630108](https://doi.org/10.1073/pnas.1106630108), 1106630108 [pii]
 39. Neil H, Malabat C, d’Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457(7232):1038–1042. doi:[10.1038/nature07747](https://doi.org/10.1038/nature07747), nature07747 [pii]
 40. van Dijk EL, Chen CL, d’Aubenton-Carafa Y, Gourvenec S, Kwapisz M, Roche V, Bertrand C, Silvain M, Legoix-Ne P, Loeillet S, Nicolas A, Thermes C, Morillon A (2011) XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475(7354):114–117. doi:[10.1038/nature10118](https://doi.org/10.1038/nature10118), nature10118 [pii]
 41. Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in

- yeast. *Nature* 457(7232):1033–1037. doi:[10.1038/nature07728](https://doi.org/10.1038/nature07728), nature07728 [pii]
42. Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10(3):155–159. doi:[10.1038/nrg2521](https://doi.org/10.1038/nrg2521), nrg2521 [pii]
 43. Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* 136(4):629–641. doi:[10.1016/j.cell.2009.02.006](https://doi.org/10.1016/j.cell.2009.02.006), S0092-8674(09)00142-1 [pii]
 44. Yazgan O, Krebs JE (2007) Noncoding but nonexpendable: transcriptional regulation by large noncoding RNA in eukaryotes. *Biochem Cell Biol* 85(4):484–496. doi:[10.1139/O07-061](https://doi.org/10.1139/O07-061), o07-061 [pii]
 45. Aalto AP, Pasquinelli AE (2012) Small non-coding RNAs mount a silent revolution in gene expression. *Curr Opin Cell Biol* 24(3):333–340. doi:[10.1016/j.ccb.2012.03.006](https://doi.org/10.1016/j.ccb.2012.03.006), S0955-0674(12)00037-3 [pii]
 46. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicke P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermuller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbelt J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaoz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Loytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Srinivasan M, Church D, Rosenbloom K, Kent WJ, Stone EA, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameer A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Xu M, Haidar JN, Yu Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyas E, Hallgrimsdottir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, De Jong PJ (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146):799–816. doi:[10.1038/nature05874](https://doi.org/10.1038/nature05874)
 47. Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, Rinn JL, Ponting CP, Stadler PF, Morris KV, Morillon A, Rozowsky JS, Gerstein MB, Wahlestedt C, Hayashizaki Y, Carninci P, Gingeras TR, Mattick JS (2011) The reality of pervasive transcription. *PLoS Biol* 9(7):e1000625. doi:[10.1371/journal.pbio.1000625](https://doi.org/10.1371/journal.pbio.1000625), discussion e1001102, 10-PLBI-PS-8040R3 [pii]
 48. Clark MB, Choudhary A, Smith MA, Taft RJ, Mattick JS (2013) The dark matter rises: the expanding world of regulatory RNAs. *Essays*

- Biochem 54:1–16. doi:[10.1042/bse0540001](https://doi.org/10.1042/bse0540001), bse0540001 [pii]
49. Jochl C, Rederstorff M, Hertel J, Stadler PF, Hofacker IL, Schrettl M, Haas H, Huttenhofer A (2008) Small ncRNA transcriptome analysis from *Aspergillus fumigatus* suggests a novel mechanism for regulation of protein synthesis. *Nucleic Acids Res* 36(8):2677–2689. doi:[10.1093/nar/gkn123](https://doi.org/10.1093/nar/gkn123), gkn123 [pii]
 50. Metzker ML (2010) Sequencing technologies – the next generation. *Nat Rev Genet* 11(1):31–46. doi:[10.1038/nrg2626](https://doi.org/10.1038/nrg2626), nrg2626 [pii]
 51. Rederstorff M (2012) Generation of cDNA libraries from RNP-derived regulatory non-coding RNAs. *Methods Mol Biol* 925:211–218. doi:[10.1007/978-1-62703-011-3_14](https://doi.org/10.1007/978-1-62703-011-3_14)
 52. Rederstorff M, Bernhart SH, Tanzer A, Zywicki M, Perfler K, Lukasser M, Hofacker IL, Huttenhofer A (2010) RNPomics: defining the ncRNA transcriptome by cDNA library generation from ribonucleo-protein particles. *Nucleic Acids Res* 38(10):e113. doi:[10.1093/nar/gkq057](https://doi.org/10.1093/nar/gkq057), gkq057 [pii]
 53. Rederstorff M, Huttenhofer A (2011) cDNA library generation from ribonucleoprotein particles. *Nat Protoc* 6(2):166–174. doi:[10.1038/nprot.2010.186](https://doi.org/10.1038/nprot.2010.186), nprot.2010.186 [pii]
 54. Shen M, Eyraes E, Wu J, Khanna A, Josiah S, Rederstorff M, Zhang MQ, Stamm S (2011) Direct cloning of double-stranded RNAs from RNase protection analysis reveals processing patterns of C/D box snoRNAs and provides evidence for widespread antisense transcript expression. *Nucleic Acids Res* 39(22):9720–9730. doi:[10.1093/nar/gkr684](https://doi.org/10.1093/nar/gkr684), gkr684 [pii]
 55. Skreka K, Schafferer S, Nat IR, Zywicki M, Salti A, Apostolova G, Griehl M, Rederstorff M, Dechant G, Huttenhofer A (2012) Identification of differentially expressed non-coding RNAs in embryonic stem cell neural differentiation. *Nucleic Acids Res* 40(13):6001–6015. doi:[10.1093/nar/gks311](https://doi.org/10.1093/nar/gks311), gks311 [pii]

Chapter 18

Gene Expression Knockdown by Transfection of siRNAs into Mammalian Cells

Yoann Abel and Mathieu Rederstorff

Abstract

SiRNA induced gene silencing or RNA interference is a powerful tool to knock down gene expression and perform gene function studies. In this chapter, we describe a basic method to silence gene expression by transfecting a specific synthetic siRNA into mammalian HeLa cells.

Key words RNAi, siRNA, Transfection, Knockdown

1 Introduction

The discovery of the RNA interference mechanisms opened the way to easily and efficiently knock down gene expression in mammalian cells [1]. Small interfering RNAs or siRNA are small double-stranded RNAs of ≈ 21 – 23 nucleotides long able to target and cleave complementary mRNAs (Fig. 1). SiRNAs can be synthesized and then directly transfected into the cell or produced from a short hairpin RNA precursor (shRNA). Once in the cytoplasm, the mature siRNA (guide strand) is loaded in the RNA Induced Silencing Complex (RISC). The RISC complex is a multi-protein particle. One of its constituent, the protein Argonaute 2 (Ago 2), possesses an RNase H-like domain involved in mRNA cleavage which occurs at a position of 10 or 11 nucleotides from the 5' end of the complementary siRNA [2]. SiRNA knockdown is transitory and generally not 100 % efficient, but is generally sufficient to observe the consequences of the targeted gene knockdown while transfected cells survive the treatment, even if an essential gene is targeted. In this chapter, we present a simple method using siRNAs to knock down the expression of a gene of interest in HeLa cells.

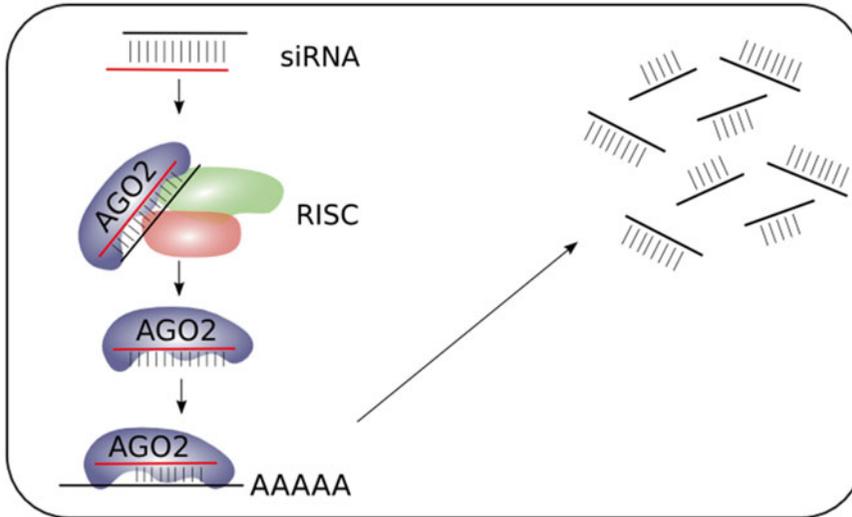


Fig. 1 siRNA induced gene silencing. After transfection, the guide strand of the double-stranded siRNA is loaded into the RISC complex, while the passenger strand is degraded. The loaded RISC targets a specific mRNA which will be cleaved by Argonaute 2

2 Materials

2.1 Mammalian Cells Culture

1. HeLa cells (ATCC) (*see Note 1*).
2. DMEM (Dulbecco's Modified Eagle Medium) cell culture medium, supplemented with 10 % of fetal bovine serum, 2.9 mg/ml of glutamine, and 10 units/ml of Penicillin-Streptomycin.
3. 6 wells plates.

2.2 siRNA Transfection

1. Chemically synthesized siRNAs (ON-Target, Dharmacon): resuspend the siRNAs in H₂O to a working concentration of 100 μ M (*see Note 2*).
2. 0.1 \times TE: 100 mM Tris-HCl, pH 8, 50 mM EDTA.
3. 2 M CaCl₂.
4. 2 \times HBS buffer: 50 mM HEPES, pH 7.5, 0.28 M NaCl, 10 mM KCl, 1.5 mM Na₂HPO₄ (*see Note 3*).

3 Methods

3.1 Mammalian Cells Culture

1. One day prior transfection, seed between 1 and 2 $\times 10^5$ HeLa cells in DMEM medium in 6 wells plates. Incubate cells at 37 $^{\circ}$ C with 5 % CO₂ overnight.
2. 1–5 h before transfection, change HeLa cells medium and add 3 ml of fresh DMEM medium.

3.2 siRNA Transfection

1. Dilute each siRNA in RNase-free H₂O to a final concentration of 10 μ M (*see Note 4*).
2. For each individual transfection, combine in this order 5 μ l of diluted siRNA, 31.8 μ l of 0.1 \times TE, 5.2 μ l of 2 M CaCl₂, and 42 μ l of 2 \times HBS. Do not vortex (*see Note 5 and 6*).
3. Mix by pipetting up and down about 15 times (*see Note 7*).
4. Incubate for 20 min at room temperature.
5. Add the transfection mixture (84 μ l) drop by drop in each well.
6. Incubate the cells at 37 °C with 5 % CO₂ for 24–96 h, but change the culture medium to remove the calcium-phosphate precipitate 12 h post-transfection.

4 Notes

1. HeLa cells are routinely used in most laboratories but numerous other cell lines can easily be transfected as well.
2. Different manufacturers offer siRNA solution, but we found the ON-TARGET siRNA (set of 4) from Dharmacon the most efficient ones in our hands.
3. Different transfection reagents and solutions are available on the market, but in our hands, calcium-phosphate transfection was the most efficient one regarding transfection of siRNAs.
4. Every siRNA should be tested to find out the best experimental conditions to get a maximal silencing efficiency. If the effect of a siRNA is too weak, you can increase its concentration or proceed with a second transfection after 24–48 h. Second transfection also enables to prolong the siRNA effect. Attention must be paid as multiple transfections and/or siRNA prolonged effects might lead to toxic effects for the cells. Generally, cells should be split (1/5) prior to second transfection.
5. A master mix can be assembled if multiple transfections of the same siRNAs have to be performed.
6. Synthetic siRNAs are fragile and should not be vortexed.
7. Pipetting also enables to inject air bubbles into the transfection mixture, which enhances formation of the fine transfection precipitate.

Acknowledgment

This work was supported by the Centre National pour la Recherche Scientifique, the Université de Lorraine, the Région Lorraine and the Ligue nationale pour la recherche contre le cancer (comités 54 et 57).

References

1. Fire A, Xu SQ, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391: 806–811
2. Sontheimer EJ (2005) Assembly and function of RNA silencing complexes. *Nat Rev Mol Cell Biol* 6:127–138

Récemment, plusieurs études ont permis de mettre en évidence un lien possible entre la maturation des snoARN et celle des microARN, deux familles d'ARNnc impliquées respectivement dans la maturation d'autres ARN, tels les ARNr, et dans la régulation de l'expression des gènes. En effet, ces études ont montré que certains microARN pouvaient être issus de la maturation d'un snoARN précurseur, formant alors une nouvelle classe d'ARNnc, les sdARN (ARN dérivés de snoARN). Le manque d'informations disponibles sur les mécanismes de maturation possibles des sdARN nous a conduit à réaliser un crible double-hybride chez la levure *S. cerevisiae* entre différentes protéines impliquées dans la biogenèse des snoARN et des microARN. Ce crible a permis d'identifier une nouvelle interaction entre les protéines TRBP, impliquée dans la maturation du pré-microARN en microARN, et RPAP3, une protéine du complexe d'assemblage hR2TP. L'observation de cette interaction soulève plusieurs questions, telles que son implication possible dans la maturation de microARN à partir de snoARN précurseurs, ou encore sur l'implication possible de TRBP et RPAP3 dans la biogenèse des snoRNP ou des microARN, respectivement. Nous avons alors entrepris de caractériser l'interaction entre TRBP et RPAP3. Par des approches de biologie moléculaire, de biochimie et de biologie structurale, nous avons d'une part confirmé l'existence du complexe *in vitro* et *in vivo* et identifié les domaines respectifs de TRBP et de RPAP3 impliqués dans l'interaction, ainsi qu'identifié plusieurs mutations intéressantes au niveau de l'interface d'interaction. L'utilisation de ces mutants devrait nous permettre d'étudier l'effet de la perte d'interaction entre TRBP et RPAP3 sur la maturation de différents ARNnc. Enfin, nos travaux ont également permis d'observer que l'interaction entre TRBP et RPAP3 n'était pas compatible avec la formation du complexe entre TRBP et la RNase Dicer. Or, il avait été montré que l'absence de TRBP au niveau de Dicer entraînait dans certains cas la génération de microARN avec des extrémités, et donc des spécificités, altérées (iso-miRs). La formation d'un complexe entre les protéines TRBP et RPAP3 pourrait donc constituer un moyen possible de réguler la disponibilité de TRBP et, indirectement, l'activité de Dicer.

Mots clés : snoRNP, microARN, sdARN, TRBP, RPAP3, Dicer

Recently, several studies have described a possible link between snoRNA and microRNA maturation, two ncRNA families involved in the maturation of other RNAs, such as rRNA, and in the regulation of gene expression, respectively. Indeed, these studies have shown that some microRNAs could be matured from a snoRNA precursor, forming a new class of ncRNAs, the sdRNAs (snoRNA-derived RNAs). The mechanism of sdRNA maturation are still poorly understood. Therefore, we have performed a two-hybrid screen in the yeast *S. cerevisiae* between different proteins involved in microRNAs and snoRNAs biogenesis. Interestingly, we observed a novel interaction between TRBP, involved in the maturation of microRNAs, and RPAP3, a member of the hR2TP complex. The observation of this interaction raises several questions, such as its possible involvement in the maturation of microRNAs from snoARN precursors, or on the possible involvement of TRBP or RPAP3 in snoRNP or microRNA biogenesis, respectively. Using various molecular biology and biochemical approaches, we undertook the functional and structural characterization of the TRBP/RPAP3 interaction. First, we confirmed the interaction both *in vitro* and *in vivo* and we identified the TRBP and RPAP3 domains involved in the interaction, as well as several interesting mutations in the binding interface. Using these mutants should allow us to study the effects of this mutations on the maturation of different ncRNAs. Additionally, we showed that the interaction between TRBP and RPAP3 and between TRBP and the RNase Dicer were mutually exclusive. Interestingly, it was shown that in the absence of TRBP, Dicer processing resulted, in some cases, in the generation of microRNAs with different ends, and thus, with altered specificity (iso-miRs). The interaction between TRBP and RPAP3 could therefore also constitute a possible way to regulate the availability of TRBP, and eventually the activity of Dicer.

Key words : snoRNP, microRNA, sdRNA, TRBP, RPAP3, Dicer