Evolution des gènes codant pour le petit ARN nucléolaire U3 chez les levures de la sous-famille des Saccharomycetoideae

Fabienne Brulé-Morabito

▶ **To cite this version:**

Fabienne Brulé-Morabito. Evolution des gènes codant pour le petit ARN nucléolaire U3 chez les levures de la sous-famille des Saccharomycetoideae. Biologie moléculaire. Université Henri Poincaré - Nancy 1, 1996. Français. NNT : 1996NAN10281 . tel-01747119

HAL Id: tel-01747119
https://hal.univ-lorraine.fr/tel-01747119

Submitted on 29 Mar 2018

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

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

AVERTISSEMENT

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

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

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

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

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

FACULTÉ DES SCIENCES

UFR Sciences et Techniques Biologiques
Ecole Doctorale Biologie et Santé

Thèse

S.C.D. - U.H.P.
BIBLIOTHÈQUE
NANCY 1
SCIENCES

Présentée pour l'obtention du titre de

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

Par Fabienne BRULE

Evolution des gènes codant pour le petit ARN nucléolaire U3 chez les levures de la sous-famille des Saccharomycetoideae.

Soutenance publique prévue le 13 novembre 1996 devant la commission d'examen :

Membres du Jury :

Rapporteurs:	Mme J. Marie	Directeur de Recherche INSERM, Gif sur Yvette
	Mr J.P. Bachellerie	Directeur de Recherche CNRS, Toulouse
Examinateurs :	Mme C. Finance	Professeur, Université Henri Poincaré, Nancy I
	Mme C. Branlant	Directeur de Recherche CNRS, Nancy
	Mr D. Tollervey	Responsable d'équipe à l'EMBL, Heidelberg, Allemagne

J'exprime ma gratitude à Monsieur le Professeur Guy BRANLANT pour m'avoir accueillie dans son laboratoire et permis de réaliser ce travail.

J'exprime une grande reconnaissance à Madame Christiane BRANLANT qui a dirigé ce travail. Par ces conseils continus, sa rigueur et son enthousiasme, elle m'a guidé, sans relâche, tout au long du travail expérimental et de la rédaction de ce mémoire. Je la remercie pour la formation scientifique qu'elle m'a apportée au cours de ces années passées à ses côtés.

Je suis très sensible à l'honneur que m'ont fait Madame Joëlle MARIE et Monsieur Jean-Pierre BACHELLERIE d'accepter d'être rapporteur de cette thèse, Madame le Professeur Chantal FINANCE et Monsieur David TOLLERVEY d'avoir accepté de juger ce travail.

Je les remercie vivement.

Je remercie Mesdames Evelyne MYSLINSKI et Véronique SEGAULT qui ont guidé, avec enthousiasme, mes premiers pas dans la recherche.

Je voudrais aussi remercier mes collègues du laboratoire pour leur amitié et leur aide, et plus particulièrement, Nathalie ROBAS et Annie MOUGIN pour leurs conseils, Régis FOURNIER et surtout, Anne GREGOIRE pour le "bout de chemin" que nous avons fait ensemble.

J'ai eu beaucoup de plaisir à travailler parmi vous.

Que tous mes parents et amis trouvent ici la reconnaissance et le respect que je leur exprime pour m'avoir soutenue et encouragée dans les instants difficiles, et aussi pour avoir été présents dans les moments de joie.

Je vous dédie à tous ce travail.

à Daniel

à Chillou et sa Maman

ABREVIATIONS

ADN	: acide désoxyribonucléique
ADNr	: ADN ribosomique
ARN	: acide ribonucléique
ARNm	: ARN messager
ARNr	: ARN ribosomique
ARNt	: ARN de transfert
ATCC	: "american type culture collection"
ATP	: adénosine 5' triphosphate
Ci	: Curie
CMCT	: 1-cyclohexyl-3-(2-morpholinoéthyl)-carbodiimide-métho-p-toluène sulfonate
Da, kDa	: dalton, kilodalton
dATP	: désoxyadénine triphosphate
db	: double brin
ddNTP	: didésoxyribonucléotide triphosphate
DMS	: diméthylsulfate
dNTP	: désoxyribonucléotide triphosphate
DO	: densité optique
DTT	: dithiothréitol
EDTA	: acide éthylène diamine tétra-acétique
ETS	: "external transcribed sequence"
GTP	: guanosine 5' triphosphate
h	: heure(s)
hnRNP	: "heterogeneous nuclear ribonucleoprotein particle"
ITS	: "internal transcribed sequence"
kb	: kilobase
min	: minute(s)
nt	: nucléotides
NTS	: "non transcribed sequence"
p/v	: poids/volume
pb	: paire de bases
PM	: poids moléculaire
pré-ARNm	: pré-ARN messager
pré-ARNr	: pré-ARN ribosomique
PRP	: "precursor mRNA processing"
qsp	: quantité suffisante pour
RNase	: ribonucléase
S	: svedberg
sb	: simple brin
SDS	: sodium dodécylsulfate
snoRNA	: "small nucleolar RNA"
snoRNP	: "small nucleolar ribonucleoprotein particle"
snRNA	: "small nuclear RNA"
snRNP	: "small nuclear ribonucleoprotein particle"
TBE	: Tris-borate-EDTA
TE	: Tris-EDTA
TEMED	: N, N, N', N' tétraméthyl éthylène diamine
Tris	: Tris (hydroxyméthyl) aminométhane
U	: unité enzymatique
V	: volts
v/v	: volume/volume
Ψ	: pseudouridine

Introduction

A. Le snoRNA U3 et la maturation des ARN ribosomiques (ARNr)

Chez les eucaryotes, la biosynthèse des ribosomes a lieu dans un sous-compartiment du noyau: le nucléole. Ont, à la fois, lieu au sein du nucléole: la transcription et la maturation des ARNr 18S, 28S et 5,8S, la génération des modifications post-transcriptionnelles de ces ARN et leur assemblage avec une grande partie des protéines ribosomiques. Seules quelques modifications post-transcriptionnelles ont lieu dans le cytoplasme et l'intégration de l'ARNr 5S au sein de la grande sous-unité ribosomique.

Nous décrirons plus particulièrement la situation chez *Saccharomyces cerevisiae* puisque notre travail a concerné cet organisme.

A1. Transcription et maturation des ARN ribosomiques.

A1.1 Les unités transcriptionnelles des ARNr.

Chez la levure *S. cerevisiae*, les deux grands ARNr ont une taille inférieure à celle des grands ARNr de vertébrés : 1801 au lieu de 1900 nucléotides pour l'ARNr de la petite sous-unité, d'où la dénomination 17S au lieu de 18S, et 3393 au lieu de 4700 nucléotides pour l'ARNr de la grande sous-unité, d'où la dénomination 26S au lieu de 28S.

Comme chez tous les eucaryotes, les ARNr 17S, 5,8S et 26S de levure sont co-transcrits par l'ARN polymérase I, à partir d'unités d'ADN ribosomiques répétées (Bell *et al.*, 1977; Rubstov *et al.*, 1980; Bayev *et al.*, 1981). Ces unités répétées environ 150 fois chez la levure, font chacune environ 9,1kb. L'ensemble de ces unités est localisé sur le bras long du chromosome XII (Petes, 1979). L'organisation de ces unités répétées en tandem est représentée dans la figure 1 (Bell *et al,* 1977; Raué et Planta, 1991). Chaque couple d'unités d'ADNr est séparé par 2 régions non transcrites : NTS1 et NTS2 (NTS pour Non Transcribed Spacer), encadrant un gène de l'ARNr 5S qui est, lui, transcrit par l'ARN polymérase III, hors du nucléole.

L'ARN polymérase I produit un transcrit de 6645 nt (Udem et Warner, 1972) (Fig. 1). Au sein de ce précurseur, les trois ARNr sont séparés par des régions "espaceur", appelées ITS_1 et ITS_2 (Internal Transcribed Spacer). Les régions 5' et 3' transcrites en amont et aval des séquences correspondant aux ARN 17S et 26S (Fig. 1) sont appelées 5' et 3' ETS (External Transcribed Sequence). La libération des trois ARNr à partir d'un même transcrit primaire permet d'avoir une production coordonnée de ces ARN.

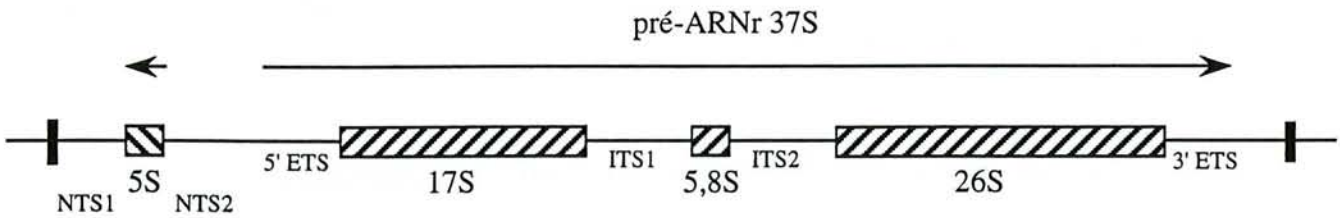


Figure 1 : Unité de transcription des ARN ribosomiques chez *Saccharomyces cerevisiae*. L'ARNr 5S est transcrit par l'ARN polymérase III sur le brin opposé par rapport à la transcription du précurseur 37S des ARNr 17S, 5,8S et 26S (Aarstad et Oyen, 1975), le précurseur 37S est transcrit par l'ARN polymérase I.

L'organisation du pré-ARNr est la même chez tous les eucaryotes avec, dans l'ordre, 17S (18S)-5,8S-26S (28S). Mais les régions ETS et ITS ont des tailles et des séquences très variables. Ainsi, chez l'Homme, le précurseur a un coefficient de sédimentation de 45S, ce qui correspond à une taille de 13000 nt (Tiollais *et al.*, 1971).

A1.2 Structure du pré-ARNr.

Les régions du pré-ARNr les plus conservées au cours de l'évolution correspondent aux ARNr. *S. cerevisiae* a été le premier eucaryote pour lequel une structure secondaire a pu être proposée, à la fois pour le 17S (Hogan *et al.*, 1984) et pour le 26S (Veldman *et al.*, 1981). Une large part des structures secondaires formées dans ces ARN sont retrouvées dans les ARN d'eucaryotes supérieurs (Michot *et al.*, 1984), ainsi que dans les ARNr bactériens (Branlant *et al.*, 1981). Ainsi, si l'on prend l'exemple du grand ARN de la grande sous-unité: cet ARN renferme 2900nt chez *E. coli*. Chacun des domaines ou sous-domaines de l'ARNr 23S bactérien (Branlant *et al.*, 1981; Noller, 1984) a un homologue dans l'ARNr 26S de levure (Branlant *et al.*, 1981; Veldman *et al.*, 1981) ou 28S de souris (Michot *et al.*, 1984). L'ARNr 26S (3393nt) renferme des séquences excédentaires par rapport à l'ARNr bactérien, qui sont situées dans les domaines variables du grand ARNr. L'ARNr 28S de souris (4700nt) renferme des régions supplémentaires de plus grande taille et de séquences différentes de celles trouvées chez *S. cerevisiae*, elles se situent dans les domaines variables du grand ARNr. Par ailleurs, l'équipe de C. Branlant a montré que l'ARN 5,8S d'eucaryote correspond à l'extrémité 5' de l'ARNr 23S bactérien (Branlant *et al.*, 1981; Veldman *et al.*, 1981). Une insertion a dû avoir lieu au cours de l'évolution dans la région codant pour la partie 5' terminale du grand ARNr, avec développement concomitant d'un processus de maturation éliminant le segment d'ARN excédentaire au niveau du précurseur, d'où la nature nouvelle de l'équivalent de l'ARNr 23S chez les eucaryotes.

Bien que les régions espaceurs du précurseur des 3 ARNr 18/17S, 5,8S et 28/26S contiennent des séquences nécessaires à la maturation de ce précurseur, leurs tailles et leurs séquences varient beaucoup suivant les espèces (Van der Sande *et al.*, 1992 ; Raué et Planta, 1995 ; Van Nues *et al.*, 1995). La comparaison des structures secondaires possibles des régions 5'ETS, ITS1, ITS2 et 3'ETS, chez différentes espèces de levures, a tout de même permis de mettre en évidence certains motifs structuraux conservés (Yeh et Lee, 1992 ; Van Nues *et al.*, 1994 ; 1995 ; Kempers-Veenstra *et al.*, 1986). Ces structures conservées entre différentes levures correspondent le plus souvent aux séquences essentielles des mécanismes de maturation du pré-ARNr (Yeh et Lee, 1990; Raué et Planta, 1995; Allmang *et al.* 1996).

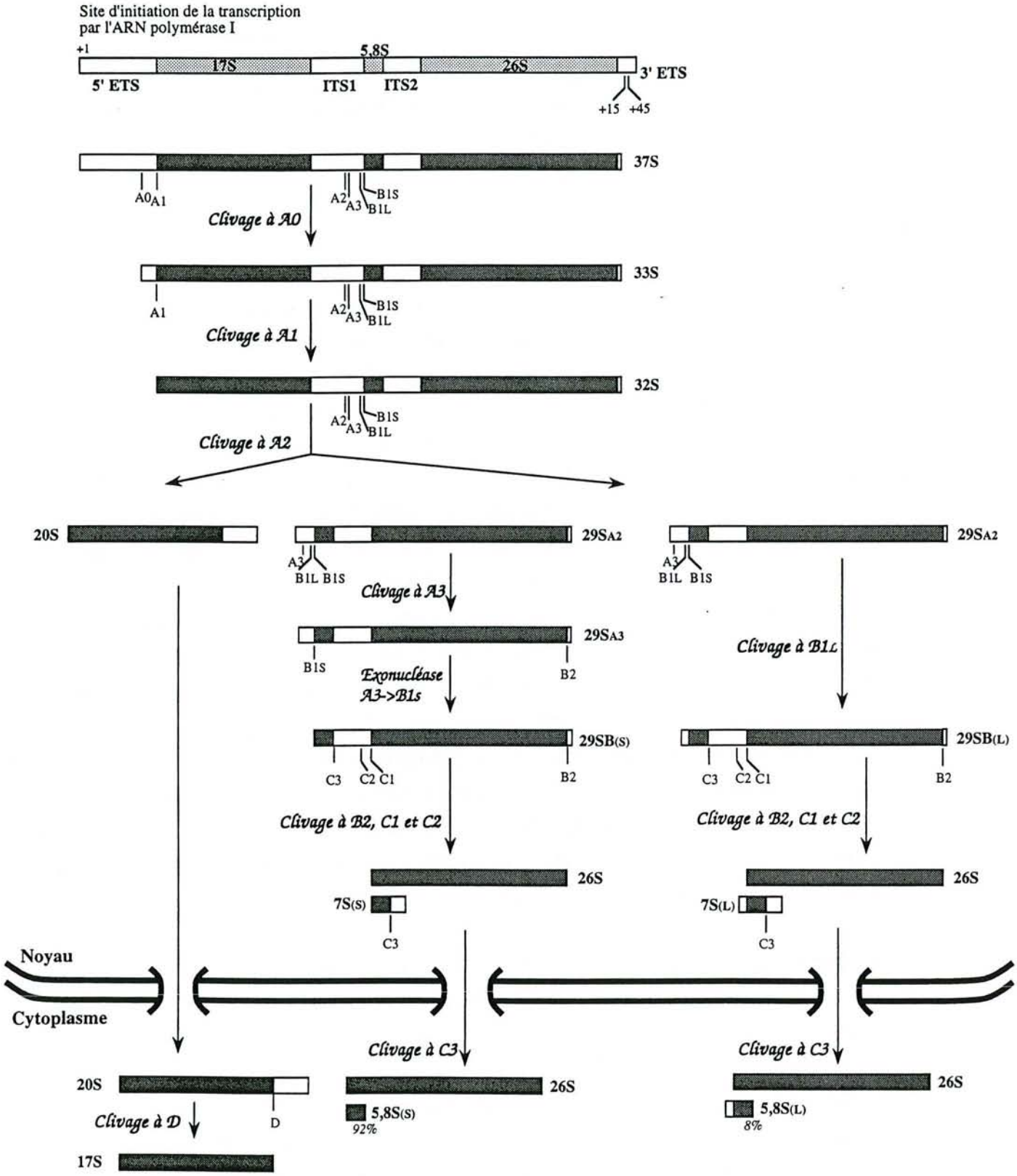


Figure 2 : Les deux voies alternatives de maturation du précurseur 37S des ARN ribosomiques chez *Saccharomyces cerevisiae* .
(Henry *et al.*, 1994; Raué et Planta, 1995)

A1.3 Les étapes de clivage du transcrit primaire des ARN ribosomiques.

Les ARNr 5,8S, 17S et 26S sont libérés à partir du long transcrit primaire par une cascade de coupures endonucléolytiques représentées dans la figure 2. La maturation du pré-ARN a été étudiée chez de nombreux organismes eucaryotes, mais c'est pour *S. cerevisiae* que les connaissances sur ce processus sont les plus approfondies. Chez *S. cerevisiae*, la première étape de maturation consiste en l'élimination de toute ou partie de la région 3' ETS (Kempers-Veenstra *et al.*, 1986), réaction tellement rapide que le premier intermédiaire détectable (le pré-ARNr 37S) est le produit de cette première étape (Fig. 2). Le premier clivage du pré-ARN 37S, qui a lieu dans la région 5' ETS au site A_0 (Hughes et Ares, 1991) libère un pré-ARNr 33S. Suit alors un clivage au site A_1 à l'extrémité 5' de l'ARN 17S conduisant au pré-ARNr 32S (Retel et Planta, 1970). La coupure au site A_2 (Fig. 2), dans la région ITS_1, sépare alors les pré-ARNr 20S et $29S_{A2}$, qui donneront respectivement lieu aux ARNr 17S et 26S+5,8S (Udem et Warner, 1972; Veldman *et al.*, 1980).

Une simple coupure endonucléolytique libère l'ARNr 17S à partir de l'intermédiaire 20S (Udem et Warner, 1972; Trapman et Planta, 1976).

Par contre, la maturation de l'intermédiaire $29S_{A2}$ est beaucoup plus complexe et peut se faire selon 2 voies selon qu'un ARNr 5,8S long (L) ou court (S) est produit (Henry *et al.*, 1994). En effet, 8% des ARNr 5,8S présentent une extension de 6 ou 7 nt à leur extrémité 5' terminale (Fig. 2) : dans la voie la plus utilisée, le pré-ARNr $29S_{A2}$ est clivé en A_3 (pré-ARNr $29S_{A3}$). Le pré-ARNr $29S_{A3}$ ainsi obtenu est alors dégradé en 5' jusqu'au site B_1 (S). L'extrémité 5' de l'ARNr 5,8S (S) est alors obtenue. Dans la deuxième voie, le pré-ARNr $29S_{A2}$ est directement clivé au site B_1 (L) générant l'extrémité 5' de l'ARNr 5,8S (L) (Henry *et al.*, 1994).

Les coupures aux sites C1, C2 et B2 (Fig. 2), sur les pré-ARNr 29SB(S) et 29SB(L) (obtenus à partir du pré-ARNr $29S_{A2}$, en suivant la première ou la deuxième voie de maturation), vont libérer l'ARNr 26S et le pré-ARNr 7S (S) ou (L). Une dernière coupure au site C3 va enfin générer les ARNr 5,8S (S) ou (L).

Deux formes d'ARNr 5,8S (longue et courte) sont aussi observées chez toutes les autres espèces eucaryotes, ce qui suggère que l'existence des 2 voies de maturation décrites ci-dessus peuvent avoir un caractère de généralité chez les eucaryotes (Henry *et al.*, 1994).

A1.4 Modifications post-transcriptionnelles du transcrit et association avec des protéines.

Très tôt après sa transcription, le pré-ARNr est associé à des protéines, un grand nombre d'entre elles, mais pas la totalité, entrant dans la composition des ribosomes (Epstein *et al.*, 1984; Raué et Planta, 1991). Par ailleurs, ce complexe ribonucléoprotéique, d'environ 90S, renferme aussi des petits ARN nucléolaires, eux-mêmes associés à des protéines.

Comme nous l'avons mentionné ci-dessus, en plus des étapes de clivage, le pré-ARNr subit, dans le nucléole, des modifications post-transcriptionnelles, essentiellement des méthylations en position 2' des riboses et la formation de pseudouridines. Ces modifications post-transcriptionnelles semblent avoir lieu très rapidement après la transcription, avant certaines étapes de maturation. La formation des 2'-O méthylations du pré-ARNr est nécessaire au déroulement normal du processus de maturation (Vaughan *et al.*, 1967). Ces modifications pourraient soit éviter la rupture de certaines liaisons phosphodiester, soit jouer un rôle dans la structuration du pré-ARNr ou des intermédiaires de maturation (Maden, 1990). Le nombre de modifications post-transcriptionnelles dans les ARNr semble avoir nettement augmenté au cours de l'évolution. Ainsi, la plupart des modifications post-transcriptionnelles présentes dans l'ARNr 23S bactérien (8 résidus pseudouridines (Ψ), 15 2'-O méthylations et 2 bases modifiées) (Branlant *et al.*, 1981; Maden, 1990; pour revue, voir Bakin et Ofengand, 1993; Ofengand *et al.*, 1995) sont retrouvées dans les ARNr 26S de levure et 28S de souris (Veldman *et al.*, 1981; Michot *et al.*, 1984), mais ces ARNr comportent un grand nombre de 2'-O méthylations et de pseudouridines non présentes dans l'ARN bactérien. Ainsi, l'ARNr 26S de *S. cerevisiae* contient 43 Ψ et 36 à 37 bases méthylées (Veldman *et al.*, 1981; Ofengand *et al.*, 1995), et l'ARNr 28S de souris comporte 57 Ψ et 63 à 65 bases méthylées (Maden, 1990). Ce sont majoritairement ces modifications supplémentaires qui ont lieu dans le nucléole, les modifications de bases présentes dans l'ARNr bactérien ayant lieu dans le cytoplasme.

A2. Les éléments impliqués dans la maturation des ARN ribosomiques.

A2.1. Stratégies d'identification des éléments intervenant dans la maturation des ARNr.

A2.11 Les éléments intervenant en cis.

Pendant longtemps, la caractérisation des éléments intervenant en cis dans le processus de maturation des ARNr a été difficile à aborder du fait du grand nombre de copies (150 à 200) des unités d'ADNr chromosomique.

Deux approches ont été développées pour résoudre ce problème.

La première approche a consisté en l'insertion de séquences étrangères dans les régions codantes d'ARNr portées par un plasmide, ces insertions étant choisies de façon à n'affecter ni la maturation, ni la fonction des ARNr. Elles permettent de distinguer les ARNr plasmidiques des ARNr produits à partir des gènes chromosomiques (Musters *et al.*, 1989). L'utilisation de ce système a permis de démontrer que des délétions, dans les régions 5' ETS ou ITS1, conduisent à un défaut de maturation de l'ARNr 17S, alors que la synthèse de l'ARNr 26S n'est pas altérée. Des délétions

dans la région ITS2 ont des conséquences opposées : un défaut de maturation de l'ARNr 26S (Van der Sande *et al*, 1992).

La seconde approche a été développée par l'équipe de Nomura (Wittekind *et al.*, 1988; Nogi *et al.*, 1991). Elle consiste à utiliser une souche de *S. cerevisiae* où l'ARN polymérase I est thermosensible et à transformer cette souche par une construction portant une unité d'ADNr placée sous le contrôle d'un promoteur Gal, reconnu par l'ARN polymérase II. A température non permissive, seule cette unité d'ADNr est exprimée. Il est donc possible d'étudier l'effet de mutations de cette unité transcriptionnelle d'ADNr. Ce système a déjà été largement utilisé par l'équipe de Tollervey pour décortiquer le mécanisme de maturation chez *S. cerevisiae*.

A2.12 Etudes des éléments intervenant en trans.

La purification des composants nucléolaires, le clonage de leurs gènes, puis l'étude de mutants de ces composants ou la recherche de suppresseurs sont les principales techniques d'étude des éléments agissant en trans dans le mécanisme de maturation des ribosomes.

C'est ainsi qu'a pu être mise en évidence l'intervention de protéines, ainsi que de petits ARN nucléolaires (snoRNA). En plus des protéines directement associées au pré-ARNr, ces approches ont aussi permis d'identifier des protéines essentielles au processus, associées aux différents snoRNA constituant ainsi des particules ribonucléoprotéiques (snoRNP).

A2.2 Les snoRNA.

A2.21. Les snoRNA de vertébrés (Tableau I)

La présence de petits ARN nucléolaires métaboliquement stables a été mise en évidence chez les vertébrés en 1976 (Zieve et Penman, 1976). Le snoRNA U3, le snRNA le plus abondant, a alors été caractérisé: son association avec les complexes renfermant les pré-ARNr suggérait déjà un rôle possible dans la maturation de cet ARN précurseur (Zieve et Penman, 1976; Busch *et al.*, 1982; Epstein *et al.*, 1984).

Deux autres snoRNA sont aussi présents à un fort nombre de copies dans le nucléole, les snoRNA U8 et U13 dont les séquences ont été établies en 1989 (Tyc et Steitz, 1989). La séquence d'un autre snoRNA, le snoRNA U14, a été publiée peu après (Liu et Maxwell, 1990).

Alors que les snoRNA U3, U8 et U13 présentent une coiffe triméthylée en 5', du même type que celle mise en évidence pour les UsnRNA spliceosomaux (^{2,2,7}mGpppAmUm) (Tyc et Steitz, 1989), le snoRNA U14 présente une extrémité 5' phosphate (Leverette *et al.*, 1992), ce qui s'explique par le fait que les gènes du snoRNA U14 sont contenus dans différents introns du gène de la protéine de choc thermique hsc70 (Liu et Maxwell, 1990; Leverette *et al.*, 1992). Le snoRNA U14 est libéré par maturation des introns (Leverette *et al.*, 1992). Cette observation inattendue, faite par Liu et Maxwell (1990), a ouvert la voie à la découverte d'un nombre considérable de snoRNA

snoRNA	Taille (nt)	Cap TMG	Boîtes conservées	Boîte ACA	Origine	Codé par un intron	ARNr associé	Anticorps	Références
U3	197 à 228	Oui Non	A', A, C', B, C et D	Non	Homme Rat Souris *Xenopus laevis* Plantes	Non	5' ETS	Anti-Fb	Suh *et al.*, 1986 Stroke et Weiner, 1985 Mazan et Bachellerie, 1988 Jeppensen *et al.*, 1988 Marshallsay *et al.*, 1990 Kiss et Solymosy, 1990
U8	136 140	Oui	C et D	Non	Homme *X.laevis*	Non	ARNr 5,8S/28S	Anti-Fb	Tyc et Steitz, 1989; Peculis et Steitz, 1993 Peculis et Steitz, 1994
U13	105	Oui	C et D	Non	Homme	Non		Anti-Fb	Tyc et Steitz, 1989
U14	86 96	Non	C et D	Non	Souris, Rat, Hamster et Homme Truite *X.laevis*	introns 5, 6, 8 du gène hsc70 introns 2,4,5,6,7,8 du gène hsc70 introns 4, 5, 7 du gène hsc70 introns 3, 4 du gène rpS13	ARNr 18S (14nt + 14 nt)	Anti-Fb	Liu et Maxwell, 1990 Shanab et Maxwell, 1991 Shanab et Maxwell, 1992 Leverette *et al.*, 1992
U15	148 (U15A) 146 (U15B)	Non	C et D	Non	Homme	intron 1 du gène rpS3 intron 5 du gène rpS3		Anti-Fb	Tycowski *et al.*, 1993 Pellizzoni *et al.*, 1994
U16	106	Non	C et D	Non	Homme *X.laevis* *X.borealis*	intron 3 du gène rpL1		Anti-Fb	Fragapane *et al.*, 1993 Prislei *et al.*, 1993 Caffarelli *et al.*, 1994
U17 (E1)	205-207 220	Non Non	- -	Oui	Homme *X.laevis*	introns 1 et 2 du gène RCC1 introns 1 à 6 du gène rpS8	5' ETS/ARNr 18S	- 	Clark *et al.*, 1990 Kiss et Filipowicz, 1993 Cecconi *et al.*, 1994
U18	67-70	Non	C et D	Non	Homme	introns 1, 2 et 4 du gène rpL1	ARNr 28S (13nt)	Anti-Fb	Prislei *et al.*, 1993
U19	200	Non	-	Oui				-	Filipowicz et Kiss, 1993
U20	80	Non	C et D	Non	Homme	intron 11 du gène nucléoline	ARNr 28S (21 nt)	Anti-Fb	Nicoloso *et al.*, 1994
U21	93	Non	C et D	Non	Homme Souris Poulet Drosophile	intron 5 du gène rpL5 -"- -"- intron du gène ARF-1	ARNr 28S (13nt)	Anti-Fb	Prislei *et al.*, 1993 Qu *et al.*, 1994 Renalier *et al.*, 1996
U22	125 125	Non	C et D	Non	Homme *X.laevis*	intron du gène UHG -"-	ARNr 18S	Anti-Fb	Tycowski *et al.*, 1994

U23	147	Non	-	Oui					J.P. Bachellerie, non publiés
U24	77	Non	C et D	Non	Homme Poulet	intron 2 du gène rpL7a intron 2 du gène rpL7a	ARNr 28S (12 nt+12 nt)	Anti-Fb	Qu *et al.*, 1995
U25- U61	60 à 98	Non	C et D	Non	Homme		ARNr 18S ou ARNr 28S (10 à 14 nt)	Anti-Fb	Kiss-Làslo *et al.*, 1996
E2	154	Non	-	Oui			ARNr 28S	-	Rimoldi *et al.*, 1993 Ruff *et al.*, 1993 Nag *et al.*, 1993
E3	135	Non	-	Oui			ARNr 18S	-	Rimoldi *et al.*, 1993 Ruff *et al.*, 1993 Nag *et al.*, 1993
7-2/MRP	260-280 260	Non	-	Non	Mammifère *A. thaliana*	Non (transcrit par l'ARN polymerase III)		Anti Th/To	Clayton, 1994 Kiss *et al.*, 1992

Tableau I : Les snoRNA identifiés chez les eucaryotes supérieurs.

Sont indiqués pour chacun des snoRNA, sa taille en nombre de nucléotides, la présence ou l'absence d'une coiffe triméthylguanosine (TMG), la présence des séquences phylogénétiquement conservées A', A, B, C, C' et D, la présence ou l'absence de la boîte ACA à 3 nucléotides de l'extrémité 3' de l'ARN (définie par Balakin *et al.*, 1996), les espèces pour lesquelles ces snoRNA ont été caractérisés, les introns dans lesquels leurs gènes sont contenus lorsque c'est le cas, l'ARN ribosomique pour lequel ils présentent une complémentarité et la longueur de cette complémentarité, les anticorps avec lesquels il a été possible de les précipiter (anti-Fb= anti-fibrillarine, anti Th/To= anticorps isolés de patients atteints de maladie auto-immune) et enfin les références majeures concernant chacun de ces snoRNA.

Les gènes rpS3, rpL1, rpS8, rpL5 et rpL7a codent, respectivement pour les protéines ribosomiques S3, L1, S8, L5 et L7a.

Le gène hsc70 code pour une protéine de choc thermique, la protéine HSC70, le gène nucléoline code pour la protéine nucléolaire du même nom, le gène RCC1 code pour la protéine RCC1 intervenant dans la régulation de la division cellulaire, le gène ARF-1 de Drosophile code pour le facteur 1 de ribosylation de l'ADP et le gène UHG ne code aucune protéine identifiée à ce jour.

chez les eucaryotes supérieurs. Ils sont peu abondants dans le nucléole et leurs gènes sont contenus dans des introns de gènes de protéines (pour revue, Maxwell et Fournier, 1995) (Tableau I). Pour beaucoup d'entre eux, il s'agit de gènes de protéines ribosomiques ou de protéines nucléolaires, telles que la nucléoline. Ceci suggère une corégulation de la production de ces protéines et la production des snoRNA contenus dans les introns des gènes de ces protéines (Maxwell et Founier, 1995).

Comme nous le verrons plus loin, ces snoRNA sont impliqués dans la maturation des ARNr, d'où une possibilité de corégulation entre biosynthèse des ARNr et des protéines ribosomiques, au travers de ces snoRNA.

Le snoRNA U17 humain semble représenter un cas particulier puisqu'il est codé par 2 introns du gène RCC1, gène qui code pour une protéine régulatrice du cycle cellulaire (Kiss et Filipowicz, 1993). Le désaccord avec l'hypothèse formulée ci-dessus n'est sans doute qu'apparent. En effet, des études récentes ont montré l'implication de la protéine RCC1 dans la maturation et l'export des ARNr (Cheng *et al.*, 1995): la protéine RCC1 semble intervenir dans l'import de protéines et ARN dans le noyau et pourrait donc intervenir dans l'import des protéines ribosomiques.

Notons également le cas des snoRNA U22 et U25 à U31 codés dans les introns du gène UGH (pour "U22 gene host") dont les exons semblent dégradés et qui ne coderaient donc pour aucune protéine; l'information apportée par ce gène ne serait contenue que dans les introns (Tycowski *et al.*, 1996; Moore, 1996).

A2.22 Les snoRNA de *S. cerevisiae* (Tableau II).

Chez *S. cerevisiae*, l'emploi d'anticorps reconnaissant spécifiquement le nucléotide ^{2,2,7}mGp à l'extrémité de la coiffe particulière portée par certains UsnRNA avait permis de mettre en évidence un très grand nombre de snoRNA chez cet organisme. Une différence importante du nombre de ces snoRNA semblait donc exister entre *S. cerevisiae* et les eucaryotes supérieurs. Avec la découverte des snoRNA codés par des introns chez les eucaryotes supérieurs, il apparaît maintenant qu'un nombre important de snoRNA existe aussi bien chez les eucaryotes supérieurs que chez *S. cerevisiae*.

Quatre snoRNA codés par des introns ont, néanmoins, été mis en évidence chez *S. cerevisiae*: U18, snR38, snR39 et U24 qui sont, respectivement, contenus dans les introns des gènes de 2 sous-unités du facteur d'élongation EF1 (ß et γ), du gène de la protéine ribosomique L8 (Maxwell et Fournier, 1995) et du gène d'une protéine appartenant à la famille des protéines fixant le GTP (Qu *et al.*, 1995).

L'existence de snoRNA codés par des introns semble donc générale chez les eucaryotes. Le fait qu'un plus grand nombre de snoRNA corresponde à des unités transcriptionnelles et non à des séquences introniques chez *S. cerevisiae* s'explique vraisemblablement par le faible nombre d'introns présents chez cet eucaryote (Kalogeropoulos, 1995).

snoRNA	Taille (nt)	cap TMG	Boîtes conservées	Boîte ACA	Essentiel à la viabilité de la cellule	Effet de la perte de l'ARN	Protéines Associées	Intermédiaires de maturation	Réferences
U3	333	+	A', A, B, C, C' et D	-	Oui	maturation du 35S; 18S faible	NOP1/SOF1	35S	Riedel *et al.*, 1986; Tollervey, 1987; Beltrame et Tollervey, 1992; Hughes *et al.*, 1987; Bally *et al.*, 1988; Beltrame *et al.*, 1994
U14	125-128	-	C et D	-	Oui	maturation du 35S; 18S faible	NOP1	27S 35S	Zagorski *et al.*, 1988; Li *et al.*, 1990; Li et Fournier, 1992; Huang *et al.*, 1992; Jarmolowski *et al.*, 1990
U18 *	102	-	C et D	-				-	Qu *et al.*, 1994
U24 *	67	-	C et D	-	Non	Absence de méthylation du 26S (C_{1436} et A_{1448})	NOP1	-	Qu *et al.*, 1995 Kiss-Làslo *et al.*, 1996
MRP	340	-	-	-	Oui		POP1/SNM1	5,8S	Lygerou *et al.*, 1994; Schmitt et Clayton, 1992, 1993, 1994; Chu *et al.*, 1994
snR3	194	+	-	+	Non		NOP1	35S	Riedel *et al.*, 1986; Tollervey, 1987; Parker *et al.*, 1988
snR4	192	+	-	+	Non		NOP1	35S	Riedel *et al.*, 1986; Tollervey, 1987; Parker *et al.*, 1988
snR5	204	+	-	+	Non		NOP1	20S	Riedel *et al.*, 1986; Tollervey, 1987; Parker *et al.*, 1988
snR8	189	+	-	+	Non		NOP1	20S	Riedel *et al.*, 1986; Tollervey, 1987; Parker *et al.*, 1988
snR9	187	+	-	+	Non		NOP1	35S 20S	Riedel *et al.*, 1986; Tollervey, 1987; Parker *et al.*, 1988
snR10	245	+	-	+	Non	maturation du 35S	NOP1/GAR1/ SSB1	35S	Tollervey, 1987; Tollervey et Guthrie, 1985; Parker *et al.*, 1988; Girard *et al.*, 1992; Clark *et al.*, 1990
snR11	259	+	-	+	Non		NOP1/SSB1		Riedel *et al.*, 1986; Clark *et al.*, 1990
snR13	124	+	C et D	-	Non				Riedel *et al.*, 1986; Balakin *et al.*, 1994

snR30	605	+	-	+	Oui	maturation du 35S; 18S faible	NOP1/GAR1	35S	Bally *et al.*, 1988; Girard *et al.*, 1992; Morrissey et Tollervey, 1993
snR31	222	+	-	+	Non		NOP1	35S	Balakin *et al.*, 1993; Dandekar et Tollervey, 1993
snR32	188	+	-	+	Non		NOP1		Balakin *et al.*, 1993
snR33	183	+	-	+	Non		NOP1		Balakin *et al.*, 1993
snR34	203	+	-	+	Non		NOP1		Samarsky *et al.*, 1995
snR35	204	+	-	+	Non		NOP1		Samarsky *et al.*, 1995
snR36	182	-	-	+	Non		NOP1		Samarsky *et al.*, 1995
snR37	386	+	-	+	Non		NOP1		Balakin *et al.*, 1996
snR38 *	95		C et D	-	Non		NOP1		Balakin *et al.*, 1996
snR39 *	89		C et D	-					Nicoloso *et al.*, 1996
snR40	97	-	C et D	-			NOP1		Balakin *et al.*, 1996
snR41	110	-	C et D	-			NOP1		Balakin *et al.*, 1996
snR42 à 46 et 49			-	+					Balakin *et al.*, 1996
snR47			C et D	-					Balakin *et al.*, 1996
snR48			C et D	-					Balakin et Ni, non publiés
snR189	192	+	-	-	Non		NOP1	20S	Thompson *et al.*, 1988; Tollervey, 1987
snR190	191-194	-	C et D	-	Non		NOP1		Zagorski *et al.*, 1988

Tableau II : Les snoRNA de la levure *Saccharomyces cerevisiae*.

TMG pour triméthylguanosine

Sont indiqués dans ce tableau, la taille des snoRNA, la présence d'une coiffe triméthylguanosine (TMG), la présence des séquences phylogénétiquement conservées A', A, B, C, C' et D, la présence de la boîte ACA située à 3 nucléotides de l'extrémité 3' des snoRNA (définie par Balakin *et al.*, 1996), l'intermédiaire de maturation du pré-ARNr avec lequel le snoRNA semble associé d'après des expériences de pontage aux UV ou d'après l'observation d'un appariement, l'immunoprécipitation avec un anticorps dirigé contre la protéine NOP1 (équivalent de la fibrillarine humaine) ou avec des anticorps dirigés contre deux autres protéines nucléolaires GAR1 et SSB1, le caractère essentiel de l'ARN d'après des expériences de disruption de son gène, l'effet produit par la perte ou l'inactivation de ce gène et enfin quelques références majeures sur chacun de ces snoRNA.

* Les snoRNA U18, U24, snR38 et snR39 présentent des séquences similaires à celles des introns suivants : l'intron du gène EF-1 ß pour U18, l'intron du gène de la sous-unité ß d'une protéine G pour U24, l'intron du gène EF 1 γ pour snR38 et l'intron du gène de la protéine ribosomique L8.

L'augmentation du nombre de snoRNA identifiés chez *S. cerevisiae* (35) a amené l'équipe de M. Fournier à proposer une classification en 2 groupes, les snoRNA qui possèdent les boîtes conservées C et D (la boîte C correspond ici à la boîte C' du snoRNA U3), et les snoRNA qui possèdent à 3 nucléotides de leur extrémité 3', une séquence ACA (l'ARN de la RNase MRP n'appartient à aucun de ces deux groupes) (Balakin *et al.*, 1996). L'importance de cette séquence ACA a été testée par une analyse mutationnelle du snoRNA snR11 de *S. cerevisiae*. Cette séquence ACA ainsi qu'une structure hélicoïdale en 3' de l'ARN est nécessaire à l'accumulation de cet ARN (Balakin *et al.*, 1996). Ce classement en deux groupes est également possible pour les snoRNA de vertébrés.

Le nucléole ayant essentiellement comme fonction la biosynthèse des ribosomes, il paraîtrait logique que les snoRNA interviennent dans ce processus. Une grande limitation dans l'étude de la fonction de ces snoRNA provenait de l'absence de système *in vitro* permettant de reproduire les étapes de maturation du pré-ARNr. Seule l'étape de clivage en A_0 dans la région 5' ETS a pu être reproduite *in vitro* sur un transcrit ARN placé en présence d'un extrait cellulaire (Miller et Sollner-Webb, 1981; Kass *et al.*, 1987). C'est en utilisant cette approche qu'une implication du snoRNA U3, dans cette étape précoce de la maturation, a pu être mise en évidence (Kass *et al.*, 1990; Mougey *et al.*, 1993).

Faute de système *in vitro* pour les autres étapes, les recherches sur le rôle possible des autres snoRNA ont consisté à

- rechercher une complémentarité avec les ARNr ou les intermédiaires,
- rechercher une cosédimentation entre snoRNA et complexes ribonucléoprotéiques renfermant les intermédiaires de maturation,
- réaliser des expériences de pontage entre snoRNA et pré-ARNr,
- utiliser les méthodes de disruption de gène chez *S. cerevisiae*.

Cette dernière approche a été très informative, car elle a permis de montrer que, chez *S. cerevisiae*, les snoRNA U3, U14 et snR30 sont essentiels (Bally *et al.*, 1988; Zagorski *et al.*, 1988; Hughes et Ares, 1991) et que l'absence de ces 3 ARN entraîne une non-production de l'ARNr 18S (Li *et al.*, 1990; Hughes et Ares, 1991; Morrissey et Tollervey, 1993). Bien que non-essentiel, le snoRNA snR10 intervient aussi dans la maturation de l'intermédiaire 35S, la disruption de son gène conduit à un phénotype thermosensible (Tollervey et Guthrie, 1985). La disruption des autres gènes de snoRNA ne conduit à aucune différence détectable de croissance d'une souche haploïde (pour revue, Maxwell et Fournier, 1995). Néanmoins, Dandekar et Tollervey (1993) ont récemment observé que la disruption du gène du snoRNA snR31 conduit à une accumulation du transcrit primaire 35S, signe d'une diminution de l'efficacité des premières étapes de maturation.

Enfin, la RNase MRP, un ribozyme constitué de l'ARN MRP et de protéines telle la protéine POP1, s'est avérée avoir, en plus de sa fonction dans la réplication des ADN mitochondriaux (Chang et Clayton, 1987; Stohl et Clayton, 1992), un rôle dans la maturation des pré-ARNr (Shuai

et Warner, 1991; Chu *et al.*, 1994). Ainsi, une mutation dans le gène codant pour l'ARN MRP ou pour la protéine POP1 provoque un phénotype thermosensible avec accumulation d'un ARN 5,8S rallongé en 5' (Lygerou *et al.*, 1994). Plus récemment, Lygerou *et al.* (1996) ont montré, par une étude *in vitro*, l'implication directe de la RNase MRP dans le clivage au site A_3 du pré-ARNr $29S_{A2}$. Les clivages aux sites A2 (situé à l'extrémité 3' de l'ARNr 17S) et A3 (situé en 5' de l'ARNr 5,8S) semblent être liés, ce qui fait proposer à Allmang *et al.* (1996) que la RNase MRP interagit avec l'ensemble des facteurs regroupés pour les coupures aux sites A1 et A2, et donc avec les snoRNA U3, U14, snR30 impliqués dans la formation de l'ARNr 17S.

Chez les vertébrés, en plus du rôle du snoRNA U3 cité plus haut (Kass *et al.*, 1990; Savino et Gerbi, 1990), un rôle a pu être proposé pour les snoRNA U14 (Maxwell et Fournier, 1995), U8 (Peculis et Steitz, 1994) et U22 (Tycowski *et al.*, 1994). Tous ces snoRNAs essentiels sont impliqués dans la maturation du pré-ARNr. La perte d'un de ces snoRNA, sauf celle du snoRNA U8 (impliqué, lui, dans la production des sous-unités 5,8 et 28 S (Peculis et Steitz, 1994)), entraîne une diminution de la production de l'ARNr 18S.

Enfin, le rôle de certains snoRNA codés par des introns, dans la formation des 2'-O méthylations des ARNr, vient d'être récemment mis en évidence aussi bien chez les vertébrés que chez *S. cerevisiae* (Kiss-Làslo *et al.*, 1996; Nicoloso *et al.*, 1996).

L'équipe de J.P. Bachellerie avait en effet observé que la plupart des snoRNA codés par des introns présentaient une complémentarité avec des séquences des ARNr 18S ou 28S renfermant des 2'-O méthylations. Par ailleurs, la 2'-O méthylation est toujours située à 5nt d'une séquence de l'ARNr complémentaire d'une boîte conservée dans les snoRNA codés par des introns (Kiss-Làslo *et al.*, 1996; Nicoloso *et al.*, 1996). La preuve directe a été apportée par disruption du gène du snoRNA U24 de *S. cerevisiae*. Cette disruption entraîne bien une absence de 2'-O méthylation à la position attendue (Kiss-Làszlo *et al.*, 1996; Tollervey, 1996).

Etant donné que les snoRNA sont présents dans le nucléole sous forme de ribonucléoprotéines et que c'est sous cette forme qu'ils interviennent dans la maturation des ARNr, avant de passer à la présentation détaillée du snoRNA U3 qui a fait l'objet de mes travaux, je ferai un bilan rapide des connaissances sur les protéines nucléolaires.

A2.3 Les protéines nucléolaires impliquées dans la maturation du pré-ARNr.

Au sein du nucléole, on peut distinguer :

1/ des protéines très abondantes qui constituent une sorte de réseau de protéines délimitant le nucléole. Les 2 protéines plus étudiées sont la nucléoline et la fibrillarine chez les vertébrés (Lapeyre *et al.*, 1987; Ochs *et al.*, 1985) avec leur homologue NSR1 et NOP1 chez *S. cerevisiae* (Jansen *et al.*, 1991; Tollervey *et al.*, 1991; Kondo et Inouye, 1992);

2/ des protéines à activité enzymatique, telles que l'ARN polymérase I, les topoisomérases;

3/ des protéines moins abondantes associées aux snoRNA;

4/ les protéines ribosomiques.

Ayant travaillé sur *S. cerevisiae*, je ferai une présentation des protéines caractérisées chez cet organisme.

Parmi les protéines intervenant dans la maturation des ARN pré-ribosomiques, on peut distinguer des protéines ayant un rôle central et général, c'est le cas de NOP1 que nous présenterons en premier, des protéines directement impliquées dans la maturation de l'ARNr 17S, des protéines impliquées dans la maturation des ARNr 5,8S et 26S et enfin, des protéines jouant plutôt un rôle dans l'assemblage des sous-unités ribosomiques. C'est dans cet ordre que nous présenterons ces différentes protéines.

A2.31 La protéine NOP1, équivalent de la fibrillarine.

En 1985, Ochs et ses collaborateurs remarquaient, dans le sérum de patients atteints d'une maladie auto-immune, la présence d'anticorps dirigés contre une protéine de 34 kDa. Cette protéine, localisée dans la région fibrillaire du nucléole, a été dénommée "fibrillarine".

La fibrillarine présente, à son extrémité N terminale, un domaine répétitif de 75 à 80 acides aminés riche en glycines et en arginines (appelé domaine GAR) (Lapeyre *et al.*, 1990). Ce domaine, présent dans de nombreuses protéines, semble constituer un signal de localisation nucléolaire (Girard *et al.*, 1992). La fibrillarine a été très conservée au cours de l'évolution (Aris et Blobel, 1991), ce qui est le signe d'un rôle important de cette protéine dans la cellule.

La protéine équivalente chez la levure, NOP1, une protéine de 38 kDa, présente 67% d'homologie avec la fibrillarine humaine (Aris et Blobel, 1991; Jansen *et al.*, 1991). NOP1 est reconnue par des anticorps anti-fibrillarine (Schimmang *et al.*, 1989). Sur la base des résultats obtenus par disruption de son gène, c'est une protéine essentielle. Une complémentation partielle peut être obtenue par la fibrillarine humaine (Jansen *et al.*, 1991). Les souches présentent alors un phénotype thermosensible. A température non permissive, les souches complémentées par la fibrillarine humaine présentent une sévère altération de la maturation du pré-ARNr, aucun ARNr 17S n'est obtenu, par ailleurs, une sous-méthylation du pré-ARNr est aussi observée (Tollervey *et al.*, 1991). Différentes mutations dans le gène *nop1* conduisent à des phénotypes différents, ce qui est en accord avec l'idée que cette protéine remplit plusieurs fonctions (Tollervey *et al.*, 1993) :

Ainsi, suivant les mutations, sont affectés:

--> la méthylation du pré-ARNr,

--> la maturation du pré-ARNr (aussi bien du 17S que du 26S),

--> l'assemblage des sous-unités du ribosome.

NOP1 n'intervient pas seule dans ces différentes étapes de la biogenèse des ribosomes (maturation, méthylation, assemblage). NOP1 est associée, de manière directe ou par interaction avec d'autres protéines, à la plupart des petits ARN nucléolaires, dont le snoRNA U3. Elle est partie intégrante des snoRNP de levure (Schimmang *et al*, 1989), et de vertébrés (Tyc et Steitz, 1989) (Tableaux I et II). Des études génétiques récentes ont montré que NOP1 agit en interaction

ou en synergie avec d'autres protéines et ont permis l'identification de ces protéines (Tollervey *et al.*, 1993).

A2.32 Les protéines intervenant dans la production de l'ARNr 17S.

¤ SOF1 intervient dans la production de l'ARN 17S.

SOF1 a été découverte en recherchant un mutant suppresseur du caractère thermosensible observé dans les souches où NOP1 avait été remplacée par la fibrillarine humaine. Jansen *et al.* (1993) ont ainsi identifié un gène codant pour une protéine nucléolaire appelée SOF1. SOF1 est une protéine de 56 kDa, la disruption de son gène a montré qu'elle était essentielle chez *S. cerevisiae*. La co-immunoprécipitation du snoRNA U3 par des anticorps anti-NOP1 et anti-SOF1 démontre l'existence d'une association physique entre le snoRNA U3, SOF1 et NOP1, 3 constituants de la snoRNP U3 (Jansen *et al.*, 1993). L'isolement de mutants thermosensibles de SOF1 a permis de montrer qu'elle intervient dans la maturation du pré-ARNr. Son inactivité entraîne, en effet, une absence de production d'ARNr 17S.

¤ SSB1 et GAR1.

La protéine SSB1, associée aux snoRNA snR10 et snR11 (Clark *et al.*, 1990), et la protéine GAR1, associée aux snoRNA snR10 et snR30 (Girard *et al.*, 1992), sont également impliquées dans la production de l'ARNr 17S.

Comme nous l'avons mentionné ci-dessus, la présence d'un ou plusieurs domaines, riches en résidus glycine et arginine, appelés domaines GAR, est une caractéristique de beaucoup de protéines nucléolaires (Girard *et al.*, 1992). C'est le cas pour la nucléoline, la fibrillarine (NOP1), SSB1 et GAR1. C'est, en effet, en utilisant une sonde ADNc correspondant à ce domaine que Girard *et al.* (1992) ont mis en évidence la protéine GAR1.

Au niveau structural, les doublets de glycine des domaines GAR doivent conférer une flexibilité à la chaîne polypeptidique. De plus, le domaine GAR, chargé positivement, doit pouvoir interagir avec les acides nucléiques. Il est possible que les domaines GAR soient impliqués dans le métabolisme des ARN en modifiant la conformation de ceux-ci. Par exemple, les domaines GAR pourraient ouvrir la structure des pré-ARNr de manière à les rendre accessibles aux autres composants tels que les protéines ribosomiques ou les snoRNP (Girard *et al.*, 1992). Alors que la protéine SSB1 n'est pas essentielle (Jong et Campbell, 1986), comme les snoRNA auxquels elle s'associe, GAR1, comme le snoRNA snR30, est essentielle: sans expression de la protéine GAR1, la maturation du pré-ARNr 37S est altérée, la synthèse d'ARNr 17S n'a pas lieu (Girard *et al.*, 1992).

A2.33 Les protéines intervenant dans la production des ARNr 26S et 5,8S.

¤ NOP77.

La recherche de mutations conférant un phénotype dit "synthétique létal", avec l'allèle mutant thermosensible nop1-5, a permis à Berges *et al.* (1994) d'identifier le gène *nop77*. Le principe de la recherche de mutants synthétiques létaux repose sur l'idée que si deux protéines interviennent dans un même phénomène, des mutations réalisées individuellement dans chacune des protéines ne sont pas forcément létales. Par contre, la présence de mutations dans chacune des deux protéines peut être létale. Si une mutation a été générée dans une protéine, conférant un phénotype thermosensible, la recherche de mutations empêchant toute croissance, puis l'identification du gène muté permettent la mise en évidence de protéines partenaires dans un même processus. La létalité est testée au niveau des spores d'une tétrade.

Le gène *nop77* est essentiel pour la viabilité des levures et code pour une protéine de 77 kDa. La disruption de ce gène entraîne un défaut de méthylation et de maturation du pré-ARNr : la synthèse des ARN issus du précurseur $29S_{A2}$, en particulier les ARNr matures 26S et 5,8S, est fortement réduite, en l'absence de protéine NOP77 fonctionnelle.

NOP77 ne semble pas faire partie intégrante des snoRNP. Néanmoins, NOP77 doit pouvoir interagir de façon transitoire avec les snoRNAs ou le pré-ARNr, puisqu'elle présente quatre motifs de reconnaissance de l'ARN (appelés RRM pour RNA Recognition Motif). D'autre part, NOP77 interagit vraisemblablement avec NOP1. On peut donc penser que NOP1 permettrait l'association entre snoRNA et pré-ARNr par interaction avec des protéines fixées à chacun de ces ARN, dont NOP77.

Il est intéressant de noter que SOF1 et NOP77, qui semblent agir de manière conjointe, d'après le test génétique utilisé pour mettre NOP77 en évidence, interviennent l'une dans la production de l'ARNr 17S, l'autre dans la production des ARNr 26 et 5,8S. Ceci montre que la maturation de ces différents ARN est hautement liée.

D'autres protéines intervenant soit dans la synthèse de l'ARNr 17S, soit dans la synthèse des ARNr 5,8 et 26S ont aussi été identifiées et nous les décrirons maintenant.

¤ DRS1 et SPB4.

Une implication des protéines DRS1 et SPB4 dans la maturation de l'intermédiaire 27S en ARNr 26S est suggérée par un défaut de cette étape de maturation lorsque des mutations sont portées par les gènes drs1 et spb4 (Sachs et Davies, 1990; Ripmaster *et al.*, 1992). Les séquences prédites en acides aminés des protéines DSR1 (Ripmaster *et al.*, 1992) et SPB4 (Sachs et Davies, 1990), à partir de leur gène, comportent les motifs caractéristiques des protéines appartenant à la famille des hélicases "DEAD box", hélicases supposées dérouler les duplex ARN/ARN. Il semble donc qu'au moins 2 hélicases soient impliquées dans cette étape de la maturation. Un autre mutant rrp1,

identifié par Andrew *et al.* en 1976, présente le même défaut de maturation de l'intermédiaire 27S. Le gène muté n'a pas encore été caractérisé pour ce mutant.

NOP77 semble intervenir dans une étape touchant à la fois la production du 5,8S et du 26S, DRS1 et SPB4 dans une étape spécifique de la production de l'intermédiaire 27S. Certaines protéines interviennent spécifiquement dans une des deux voies possibles de production des ARNr 26S et 5,8S, c'est le cas de POP1.

¤ POP1.

Lygerou *et al.*, en 1994, ont, en effet, mis en évidence l'intervention d'une protéine de 100 kDa (POP1) dans la production de l'ARNr 5,8S. Cette protéine s'est avérée être un composant de la RNase MRP.

Comme nous l'avons décrit ci-dessus, deux formes d'ARNr 5,8S (S) et (L) sont produites et ceci par deux voies différentes de maturation. La forme majeure 5,8S (S) est produite à partir de l'intermédiaire $29S_{A3}$ obtenu après coupure en A3 (Henry *et al.*, 1994). La forme 5,8S (L) ne nécessite pas ce clivage en A3.

Dans des souches de levure portant une mutation dans le gène *pop1*, la coupure en A3 n'a pas lieu, d'où une diminution du taux d'ARNr 5,8S (S) et l'apparition d'un ARNr 5,8S rallongé en 5', de 7 à 8 nucléotides. La disruption du gène *pop1* révèle un caractère essentiel de la protéine POP1 (Lygerou *et al.*, 1994). Le même phénotype est obtenu par mutation du gène *rrp2* qui code pour l'ARN constituant la RNase MRP (Chu *et al.*, 1994).

La RNase MRP serait donc directement impliquée dans le clivage en A3. L'intermédiaire obtenu est alors le substrat d'exonucléases 5'->3', XRN1 et RAT1 (Stevens *et al.*, 1991; Amberg *et al.*, 1992), qui conduisent à la forme courte de l'ARN 5,8S. Ce rôle direct de la RNase MRP a d'ailleurs été confirmé récemment (Lygerou *et al.*, 1996) par le développement d'un système d'étude *in vitro*.

A2.34 Les protéines impliquées dans l'import des composants nucléolaires.

C'est le cas de la protéine NSR1 (Nuclear Signal Recognition protein). Cette protéine qui a un rôle très général dans l'import des protéines dans le noyau est, bien sûr, importante dans l'assemblage des sous-unités ribosomiques dans le nucléole.

Dans une souche mutante nsr1, la maturation des pré-ARNr (notamment la synthèse de l'ARNr 17S) est légèrement affectée à 30°C. Elle l'est, encore plus, après un changement de température de 30 à 10°C. Du fait de sa structure proche de celle de la nucléoline (protéine nucléolaire de vertébrés (Kondo et Inouye, 1992)), Lee *et al.* (1992) ont proposé que la protéine NSR1, en plus d'un rôle dans l'import des protéines ribosomiques et nucléolaires dans le noyau, soit impliquée directement dans la formation des particules pré-ribosomiques en facilitant des interactions entre des protéines ribosomiques et le pré-ARNr dans le nucléole.

Protéine	Origine	Maturation pré-ARNr	Domaines conservés	Références
NOP1 Fibrillarine	*S.cerevisiae* Homme *X.laevis*	17S 26S et 5,8S	GAR	Ochs *et al.*, 1985; Aris et Blobel, 1991; Schimmang *et al.*, 1989; Tollervey *et al.*, 1993; Jansen *et al.*, 1991; Tollervey *et al.*, 1991; Girard *et al.*, 1993a
SOF1	*S.cerevisiae*	17S 26S (-)	GAR	Jansen *et al.*, 1993; Jansen *et al.*, 1994
SSB1	*S.cerevisiae*	17S	GAR	Jong et Campbell, 1986; Jong *et al.*, 1987; Clark *et al.*, 1990
GAR1	*S.cerevisiae* *S.pombe*	17S	GAR	Girard *et al.*, 1992; Girard *et al.*, 1993b
NOP77	*S.cerevisiae*	26S		Bergès *et al.*, 1994
SPB4 DRS1	*S.cerevisiae*	26S		Sachs et Davis, 1990; Ripmaster *et al.*, 1992
POP1	*S.cerevisiae*	26S		Lygerou *et al.*, 1994
NSR1 Nucléoline	*S.cerevisiae* Mammifères	17S 26S et 5,8S	GAR RRM	Lapeyre *et al.*, 1987; Kondo *et al.*, 1992; Kondo et Inouye, 1992;
NOP3	*S.cerevisiae*	26S		Russell et Tollervey, 1992
RAT1		5,8Ss		Amberg *et al.*, 1992
XRN1		5,8S		Stevens *et al.*, 1991
RNT1	*S.cerevisiae*	17S 26S	(équivalent RNase III de procaryotes)	Elela *et al.*, 1996

Tableau III : Les protéines participant à la maturation des ARN ribosomiques chez *Saccharomyces cerevisiae* et leurs équivalents chez les vertébrés lorsqu'ils ont été identifiés.

Sont indiqués les protéines équivalentes de vertébrés lorsqu'elles ont été caractérisées, l'ARNr dont la synthèse est affectée par une mutation dans la protéine, la présence de domaine GAR (Gly ou Arg) ou RRM (RNA Recognition Domain), les références majeures concernant ces protéines.

Enfin, la protéine NOP3 provoque un retard de la maturation du précurseur 35S, ce qui diminue la production de tous les intermédiaires, de l'ARNr 17S au 26S. NOP3, protéine à la fois nucléolaire et nucléoplasmique, est supposée jouer un rôle dans l'import des composants impliqués dans la maturation du pré-ARNr ou dans l'assemblage du ribosome dans le nucléole (Russell et Tollervey, 1992)

L'ensemble des protéines déjà étudiées (Tableau III) intervenant dans la synthèse des ribosomes, appartient donc à différentes classes :

-->SPB4, DSR1, des protéines ayant probablement une activité hélicase ARN dépendante,

--> RAT1 ou XRN1, des exonucléases,

--> NOP3, NOP77 ou NSR1, des protéines non associées à des snoRNAs (ou dont l'association n'a pas été encore démontrée), mais nécessaires à la biogenèse des ribosomes,

--> SSB1, GAR1, SOF1, NOP77 et POP1, des protéines associées à certains petits ARN nucléolaires,

--> NOP1 (équivalent de la fibrillarine) associée à de nombreux snoRNA et impliquée dans pratiquement toutes les étapes de la biogenèse des ribosomes,

--> RNT1, une endonucléase, vient d'être récemment mise en évidence (Elela *et al.*, 1996).

Parmi les protéines associées à des snoRNA, GAR1, SOF1 et SSB1 sont impliquées dans la production de l'ARNr 17S. La disruption des gènes de ces trois protéines conduit au même phénotype que la disruption des gènes des snoRNA essentiels U3, U14 et snR30, à savoir une absence de production de l'ARNr 17S, d'où l'idée d'un rôle coordonné de ces trois snoRNA et de ces trois protéines dans les étapes spécifiques à la production de l'ARNr 17S, à savoir les coupures aux sites A0, A1 et A2. L'idée actuelle est que les snoRNP renfermant les snoRNA U3, U14 et snR30 et les protéines NOP1, SOF1 et GAR1, forment un macrocomplexe "le processosome" (Fig. 3), à l'image du spliceosome qui assure l'épissage des ARN pré-messagers. Les snoRNA U3 et U14 ont en commun les boîtes conservées C et D, le snoRNA snR30 ne contient pas ces boîtes. Ces trois snoRNA sont associés à la fibrillarine, mais seul le snoRNA snR30 est associé à GAR1, le snoRNA U3 est associé à SOF1.

L'importance des boîtes C et D a été testée dans le cas de U14 chez *S. cerevisiae* (Jarmolowski *et al.*, 1990). L'intégrité de ces deux séquences est essentielle à la fonction, mais aussi la présence de l'appariement entre les extrémités 3' et 5' du snoRNA U14 qui rapproche les deux boîtes dans l'espace (Fig. 4). Il est vraisemblable que ces boîtes C et D sont importantes pour la fixation de protéines d'après les résultats obtenus sur les snoRNA introniques (Tycowski *et al.*, 1993) et sur le snoRNA U3 pour lequel nous allons faire une description détaillée de l'état des connaissances.

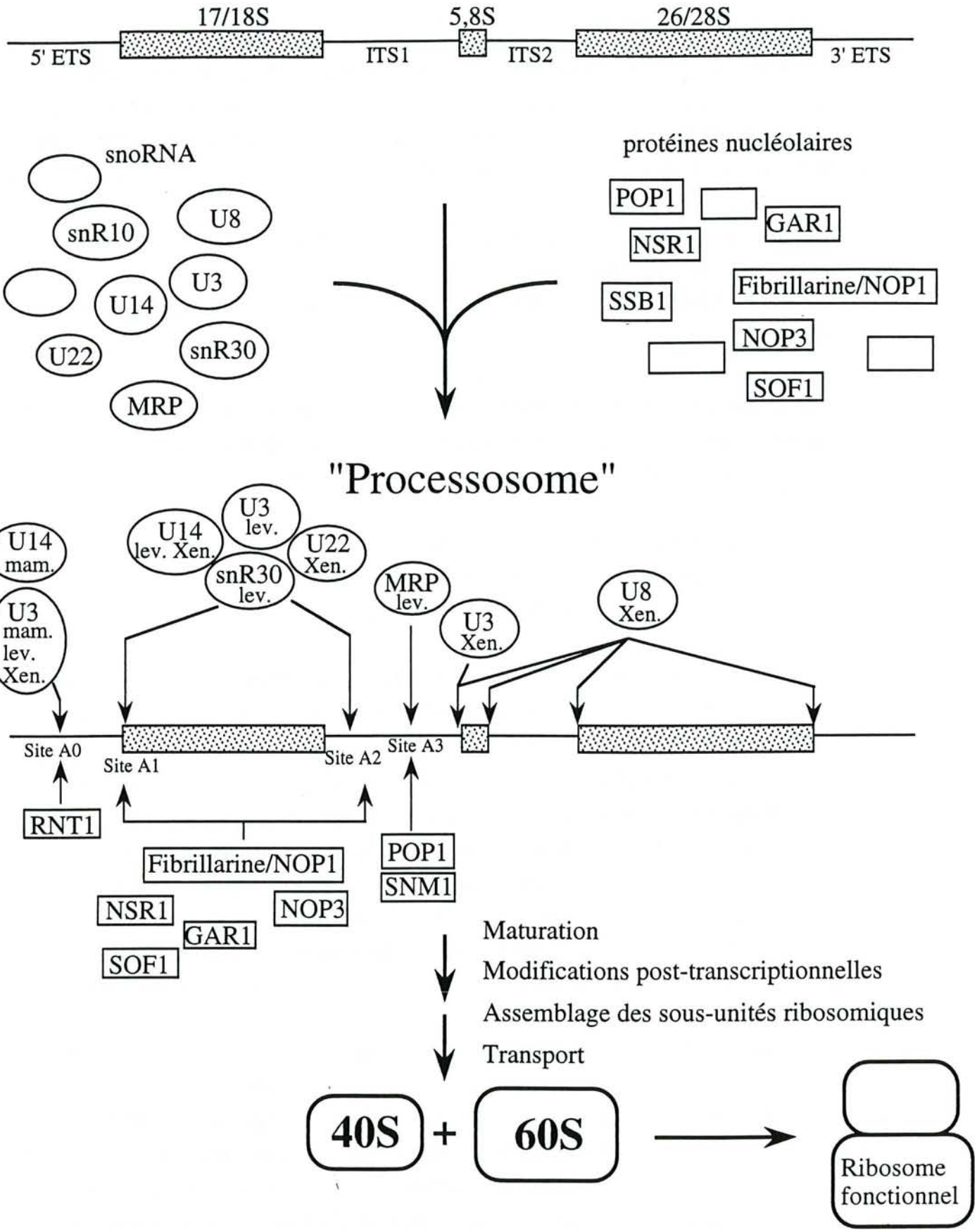


Figure 3 : Les snoRNA et les protéines nucléolaires impliqués dans les différentes étapes de maturation des ARN ribosomiques.
L'implication de ces différents composants de la machinerie de maturation des ARN ribosomiques a été mise en évidence par épuisement de la cellule en snoRNA ou en protéine. (lev. pour levure, mam. pour mammifères, Xen. pour *Xenopus laevis*)

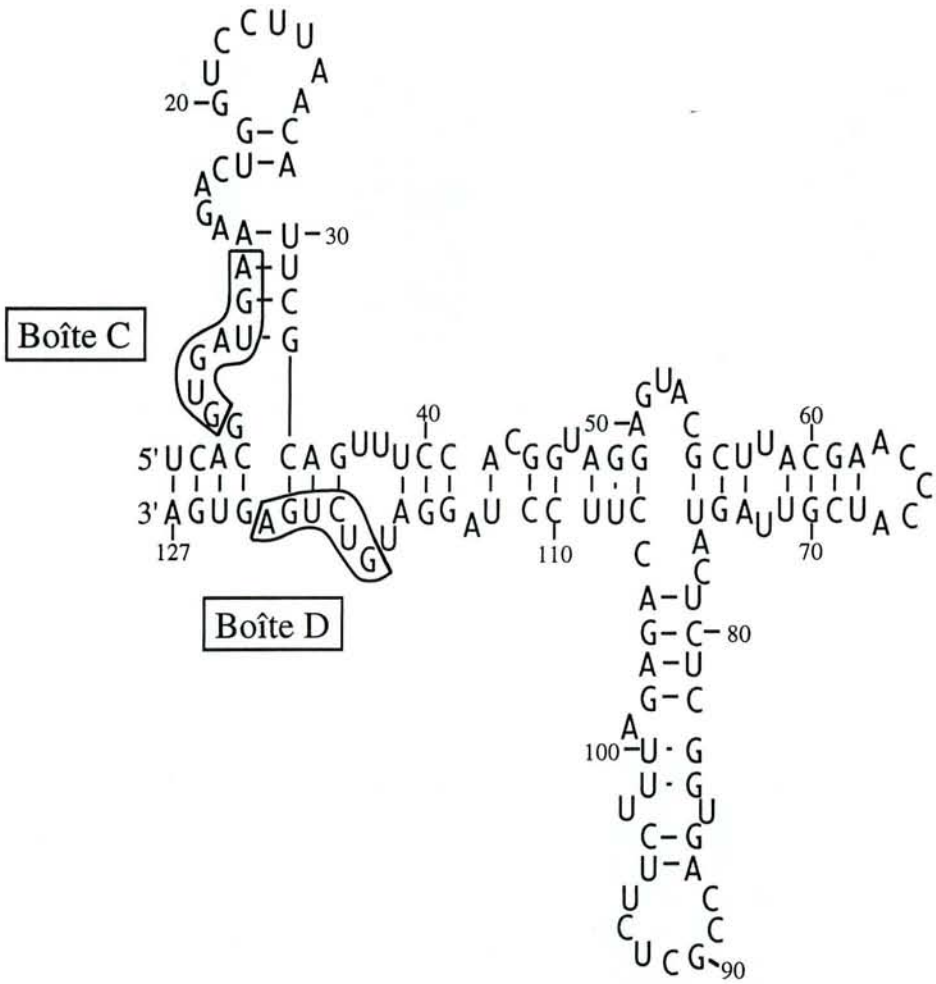


Figure 4 : Modèle de structure secondaire du snoRNA U14 de *Saccharomyces cerevisiae* proposé par Shanab et Maxwell en 1991.

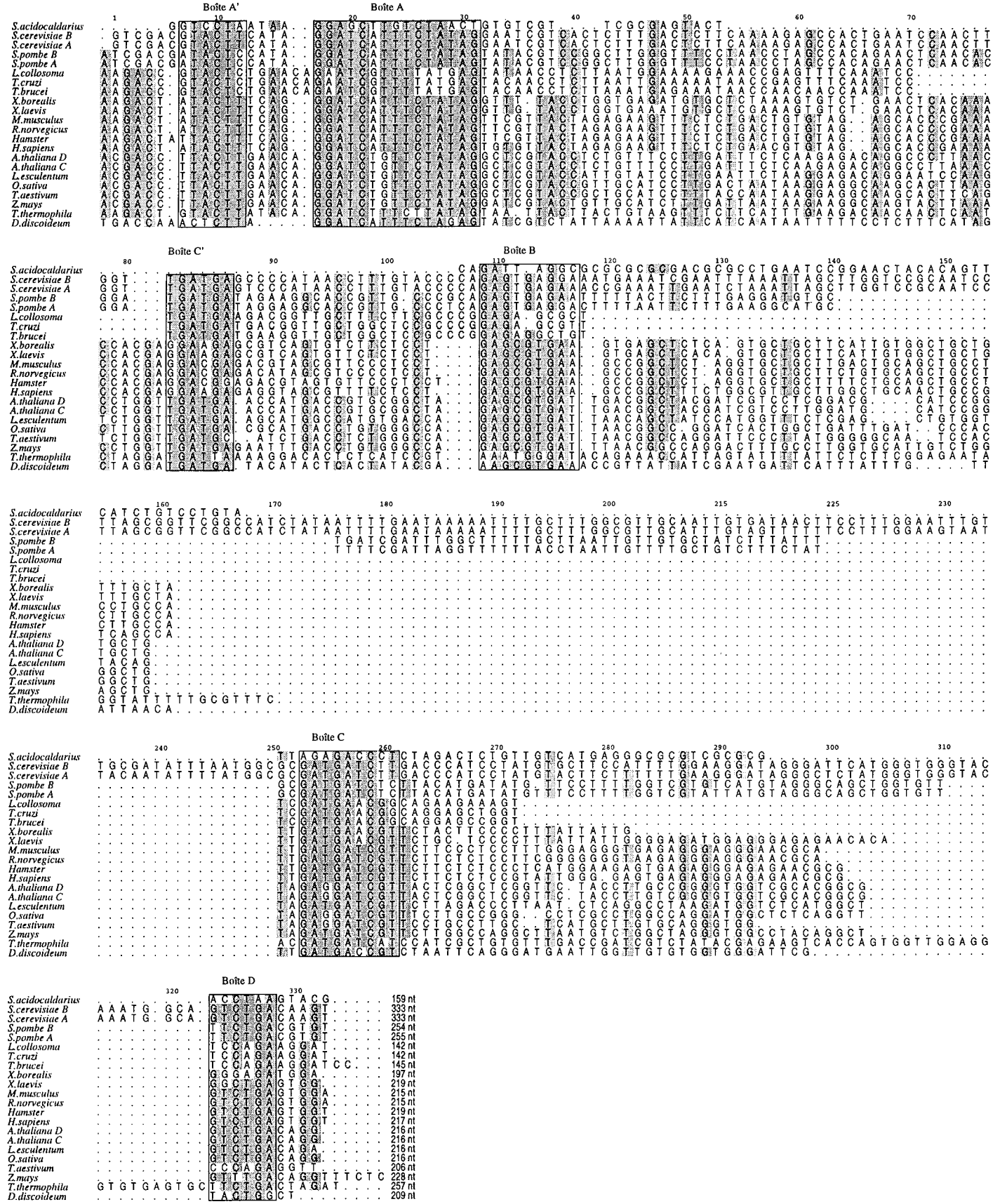


Figure 5 : Alignement des séquences des snoRNA U3 caractérisés chez différentes espèces.
Cet alignement a été réalisé avec le programme PileUp du logiciel GCG (Genetics Computer Group, Wisconsin, USA) réajusté de manière à obtenir l'alignement des boîtes conservées au cours de l'évolution A', A, C', B, C et D.
Sulfolobus acidocaldarius (Potter *et al.*, 1995), *Saccharomyces cerevisiae* (Hughes *et al.*, 1987; Myslinski *et al.*, 199 *Schizosaccharomyces pombe* (Porter *et al.*, 1988), *Leptomonas collosoma*, *Trypanosoma cruzi* et *Trypanosoma bru* (Harsthorne et Agabian, 1994), *Xenopus borealis* et *Xenopus laevis* (Jeppensen *et al.*, 1988), *Mus musculus* (Mazan Bachellerie, 1988), *Rattus norvegicus* (Stroke et Weiner, 1985), Hamster (A. Grégoire *et al.*, résultats non publié *Homo sapiens* (Suh *et al.*, 1986) *Arabidopsis thaliana* et *Tricitum aestivum* (Marshallsay *et al.*, 1990), *Lycopersic esculentum* (Kiss et Solymosy, 1990), *Oriza sativa* (Qu F., résultats non publiés) *Zea mays* (Leader *et al.*, 199· *Tetrahymena thermophila* (Orum *et al.*, 1993) et *Dictyostelium discoïdeum* (Wise et Weiner, 1980).

A3. Le snoRNA U3.

A3.1 Structure primaire des snoRNA U3.

Les premières séquences de snoRNA U3 ont été établies pour des vertébrés (Stroke et Weiner, 1985 ; Suh *et al.*, 1986 ; Mazan et Bachellerie, 1988). Depuis, des séquences de snoRNA U3 de plantes et de plusieurs eucaryotes inférieurs, tels que *Trypanosoma brucei*, *Trypanosoma cruzi*, *Tetrahymena thermophila*, *Leptomonas collosoma*, *Schizosaccharomyces pombe* (Porter *et al.*, 1988; Orum *et al.*, 1993; Harsthorne et Agabian, 1994) ont été établies.

En ce qui concerne *S. cerevisiae*, des séquences avaient été établies à partir des deux gènes A et B par Hughes *et al.* (1987). Ces séquences ont ensuite été corrigées par l'équipe de C. Branlant avant mon arrivée au laboratoire (Myslinski *et al.*, 1990), ceci du fait de l'observation de l'existence d'un intron dans chacun de ces gènes, intron qui avait été au départ considéré comme étant la séquence codante.

L'alignement des séquences établies pour les snoRNA U3 des différentes espèces (Fig. 5) met en évidence l'existence de nombreuses boîtes conservées. En plus des boîtes C (Wise et Weiner, 1980) et D (Hughes *et al.*, 1987), déjà mentionnées, et situées dans la région 3', les snoRNA U3 présentent 4 autres séquences conservées: les boîtes A, B (Wise et Weiner, 1980), A' (Myslinski *et al.*, 1990) et C' (Tycowski *et al.*, 1993). La région 5' de la molécule présente un plus grand degré de conservation que la région 3' et surtout la région centrale où de grandes insertions sont observées chez *S. cerevisiae* et *T. thermophila* (Fig. 5).

Très récemment, un ARN qui semble impliqué dans la maturation des ARN pré-ribosomiques (Durovic et Dennis, 1994), et qui comporte des séquences présentant une homologie avec les boîtes A, B et C' du snoRNA U3 (60%, 67% et 61%, respectivement) (Fig. 5), a été mis en évidence chez une archaebactérie *Sulfolobus acidocaldarius* (Potter *et al.*, 1995). D'où l'idée que l'apparition des petits ARN impliqués dans la maturation des ARN pré-ribosomiques, pourrait être antérieure à l'apparition des eucaryotes. Jusque là, il avait toujours été considéré que leur apparition était concomitante avec l'apparition d'un noyau dans les cellules. Une différence importante entre cet ARN bactérien et le snoRNA U3 de vertébré est, bien sûr, l'absence de coiffe m^3G à l'extrémité 5'. En effet, à l'exception des snoRNA U3 de plantes (Kiss *et al.*, 1985; Kiss et Solymosy, 1990; Marshallsay *et al.*, 1990; Leader *et al.*, 1994), les snoRNA U3 d'eucaryotes étudiés jusqu'à ce jour, présentent comme les UsnRNA spliceosomaux U1, U2, U4 et U5, une coiffe avec un résidu 2,2,7-triméthyl Guanosine en 5' (Filipowicz et Kiss, 1993; Maxwell et Fournier, 1995). Comme nous le verrons plus loin, la présence de cette coiffe est liée à la transcription des gènes de snoRNA U3 par l'ARN polymérase II (Mazan et Bachellerie, 1990), les snoRNA U3 de plantes étant eux, transcrits par l'ARN polymérase III (Marshallsay *et al.*, 1990). En plus de la coiffe, le snoRNA U3 de rat renferme plusieurs nucléotides modifiés, dont des pseudouridines (Reddy *et al.*, 1980; Reddy, 1988).

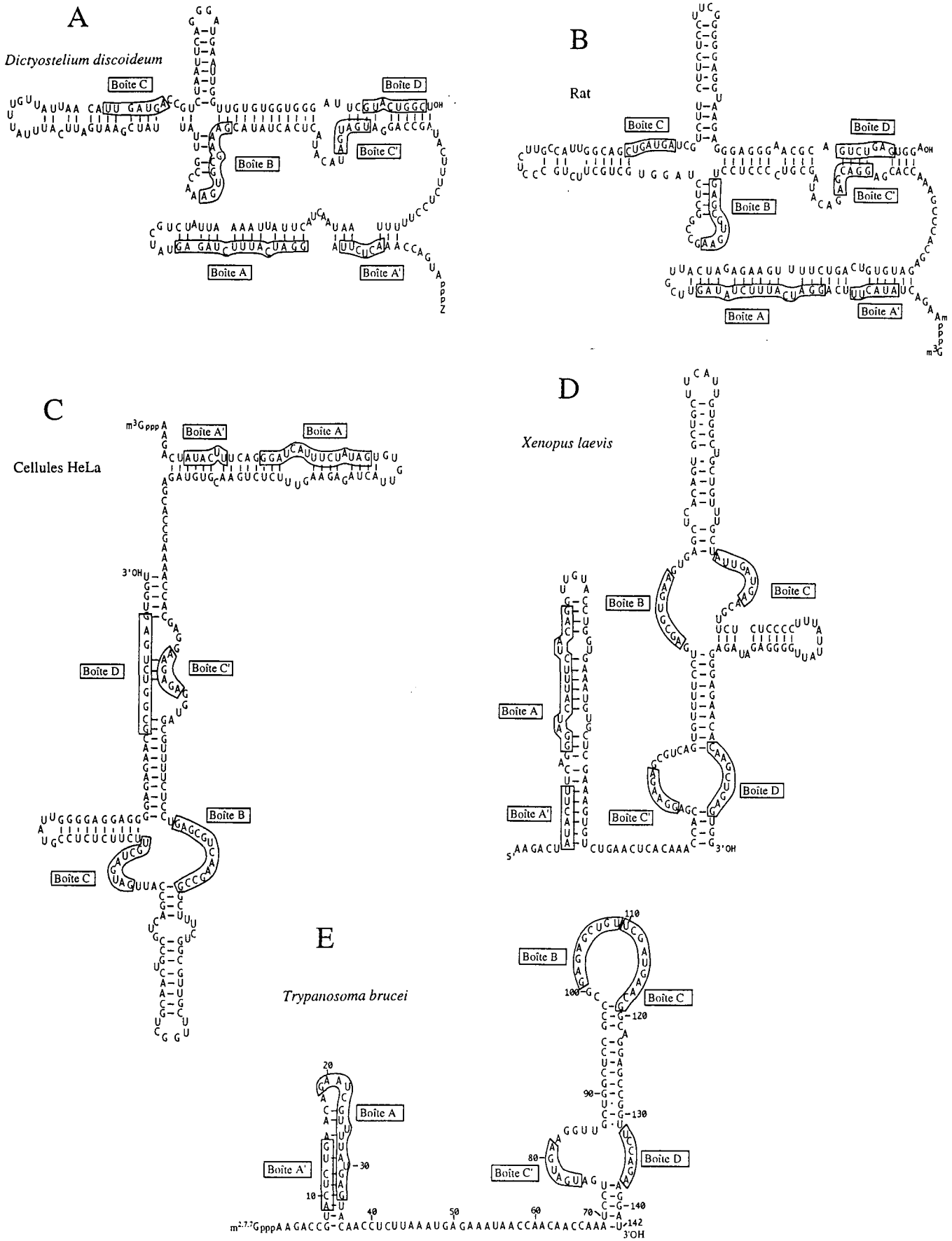


Figure 6 : Modèles de structures secondaires des snoRNA U3 de *Dictyostelium discoideum* (A), de rat (B) (Stroke et Weiner, 1985), **de cellules HeLa (C)** (Parker et Steitz, 1987), **de *Xenopus laevis* (D)** (Jeppensen *et al.*, 1988) **et de *Trypanosoma brucei*** (Harsthorne et Agabian, 1994).

Les modèles proposés pour le snoRNA de cellules HeLa (Parker et Steitz, 1987), pour le snoRNA U3 de *Xenopus laevis* (Jeppensen *et al.*, 1988) et pour le snoRNA U3 de *Trypanosoma brucei* (Harsthorne et Agabian, 1994) sont basés sur une étude expérimentale.

A3.2 Structure secondaire des snoRNA U3.

Le premier modèle de structure secondaire proposé pour le snoRNA U3 était basé sur la comparaison des deux seuls ARN U3 connus à l'époque: l'ARN de rat et celui de *Dictyostelium discoïdeum* (Stroke et Weiner, 1985) (Fig. 6). Une étude expérimentale de la structure du snoRNA U3 humain (Parker et Steitz, 1987) a ensuite permis d'affiner ce modèle (Fig. 6). Cette étude était basée sur l'emploi de sondes chimiques de la structure des ARN. Le modèle ainsi établi s'est avéré valable pour toutes les séquences publiées de snoRNA U3 de vertébrés. Une confirmation expérimentale a d'ailleurs été obtenue peu de temps après par étude du snoRNA U3 de Xénope (Jeppensen *et al.*, 1988), et plus récemment, par l'étude expérimentale du snoRNA U3 de *Trypanosoma brucei* (Hartshorne et Agabian, 1994) (Fig. 6).

Selon ce modèle (Fig. 6), l'ARN peut être décomposé en deux domaines :

=> un domaine 3' comportant une hélice et deux structures tige/boucle jointes par les boîtes B et C en simple brin. Les boîtes C' et D sont situées face à face au sein de l'hélice 4.

=> un domaine 5' formé d'une seule structure tige/boucle renfermant les boîtes A et A'. Le modèle établi pour le domaine 5' du snoRNA de vertébrés ne s'appliquait pas aux séquences établies pour le snoRNA de *Schizosaccharomyces pombe* (Porter *et al.*, 1988) et pour les snoRNA U3 de plantes (Kiss et Solymosy, 1990; Marshallsay *et al.*, 1990). En effet, d'après leur séquence, ces deux ARN semblaient présenter un domaine 5' formé de deux structures tige/boucle au lieu d'une seule comme chez les vertébrés (Fig. 7). Aucune étude expérimentale n'avait néanmoins été menée dans le cas de ces snoRNA U3.

D'après la séquence établie au laboratoire pour les variants de snoRNA U3, la situation était similaire pour *S. cerevisiae*, à savoir que le modèle établi pour la région 5' du snoRNA de vertébré ne s'appliquait pas. Par ailleurs, les snoRNA U3 de *S. cerevisiae* présentaient une longue insertion au niveau du domaine 3' (Fig. 5). La détermination expérimentale de la structure secondaire du snoRNA U3A de *S. cerevisiae* a été réalisée au laboratoire (Ségault *et al.*, 1992). Ce travail montrait que le snoRNA U3 de *S. cerevisiae* en solution comporte un domaine 5' formé de deux structures tige/boucle et un domaine 3' ayant une structure cruciforme (Fig. 8A). Mais, une étude menée par Anne Grégoire au laboratoire en collaboration avec l'équipe du Pr R. Lührmann (Marburg, Allemagne), sur la snoRNP U3, a permis de conclure qu'au sein de la snoRNP U3 de *S. cerevisiae*, le snoRNA U3 a une structure différente de l'ARN en solution, avec une structure tige/boucle 1b raccourcie dans le domaine 5' et une implication des boîtes D et C' dans une extension de l'hélice 5 (Fig. 8B).

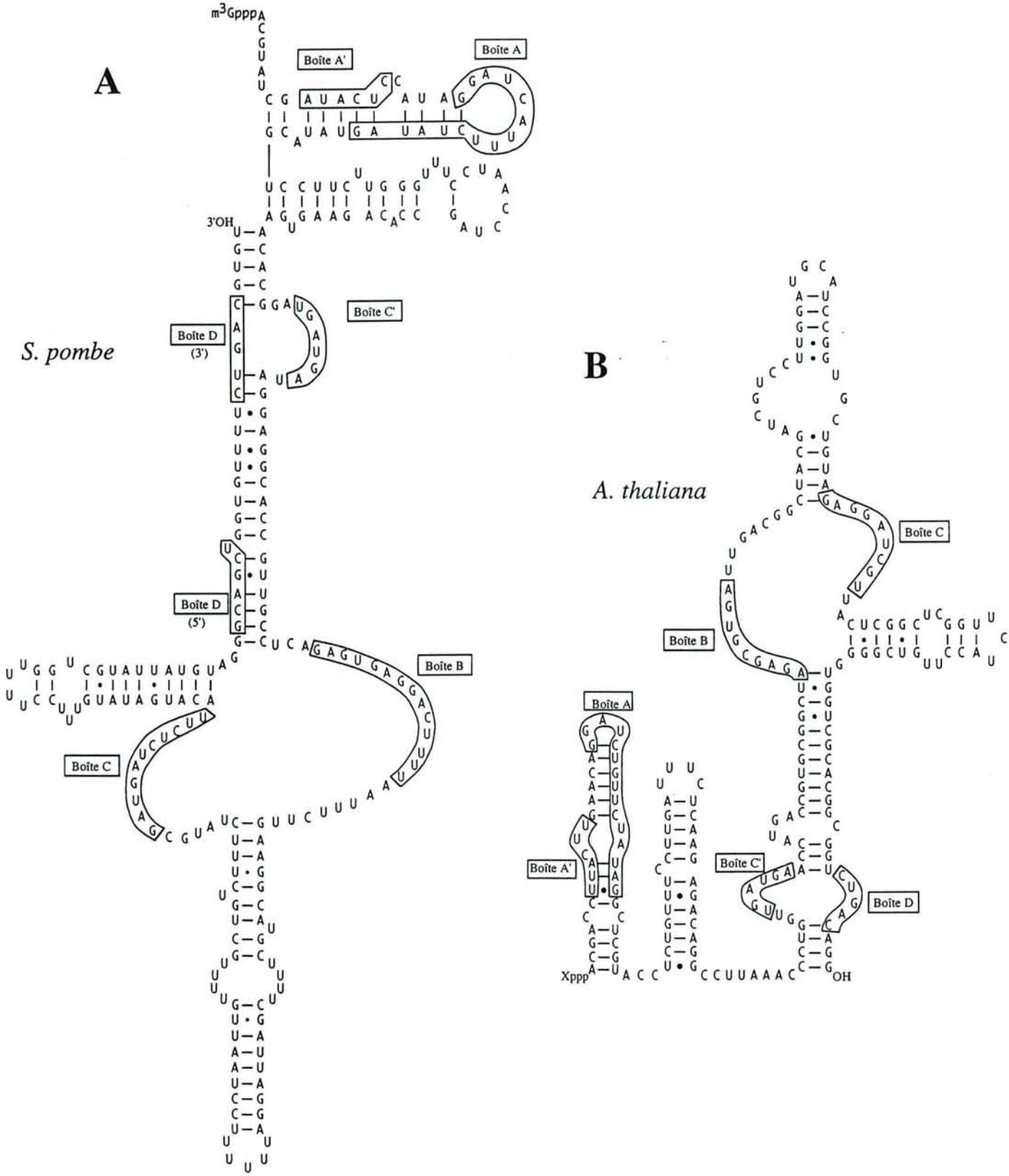


Figure 7 : Modèles de structures secondaires des snoRNA U3 de *Schizosaccharomyces pombe* (A) (Porter *et al.*, 1988) **et d'*Arabidopsis thaliana* (B)** (Marshallsay *et al.*, 1990).

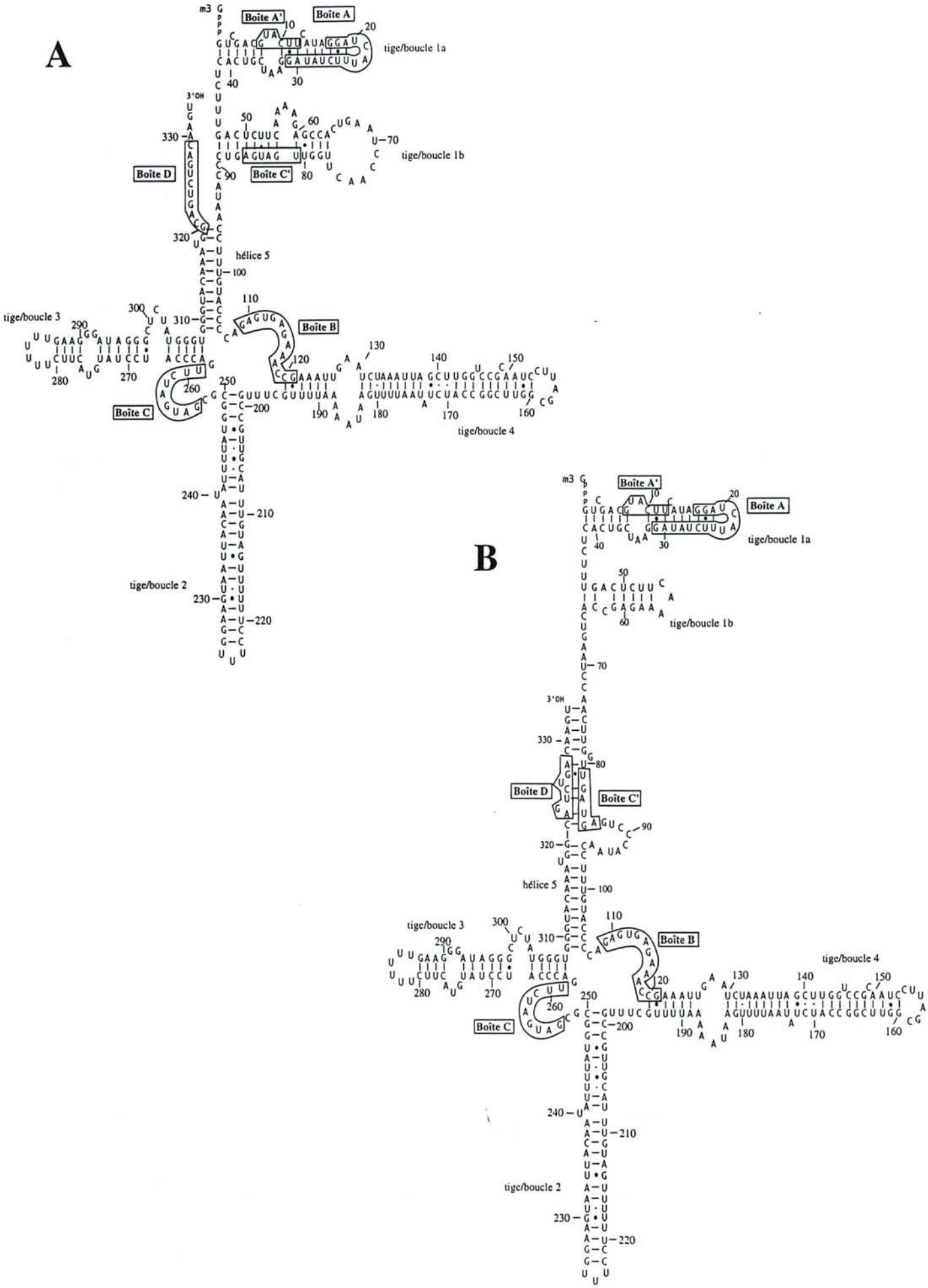


Figure 8 : Structure secondaire du snoRNA U3 de levure *S. cerevisiae*, en solution (A) (Ségault *et al.*, 1992) **ou au sein de la particule ribonucléoprotéique (B)** (A. Grégoire, résultats non publiés). Les séquences conservées phylogénétiquement sont encadrées.

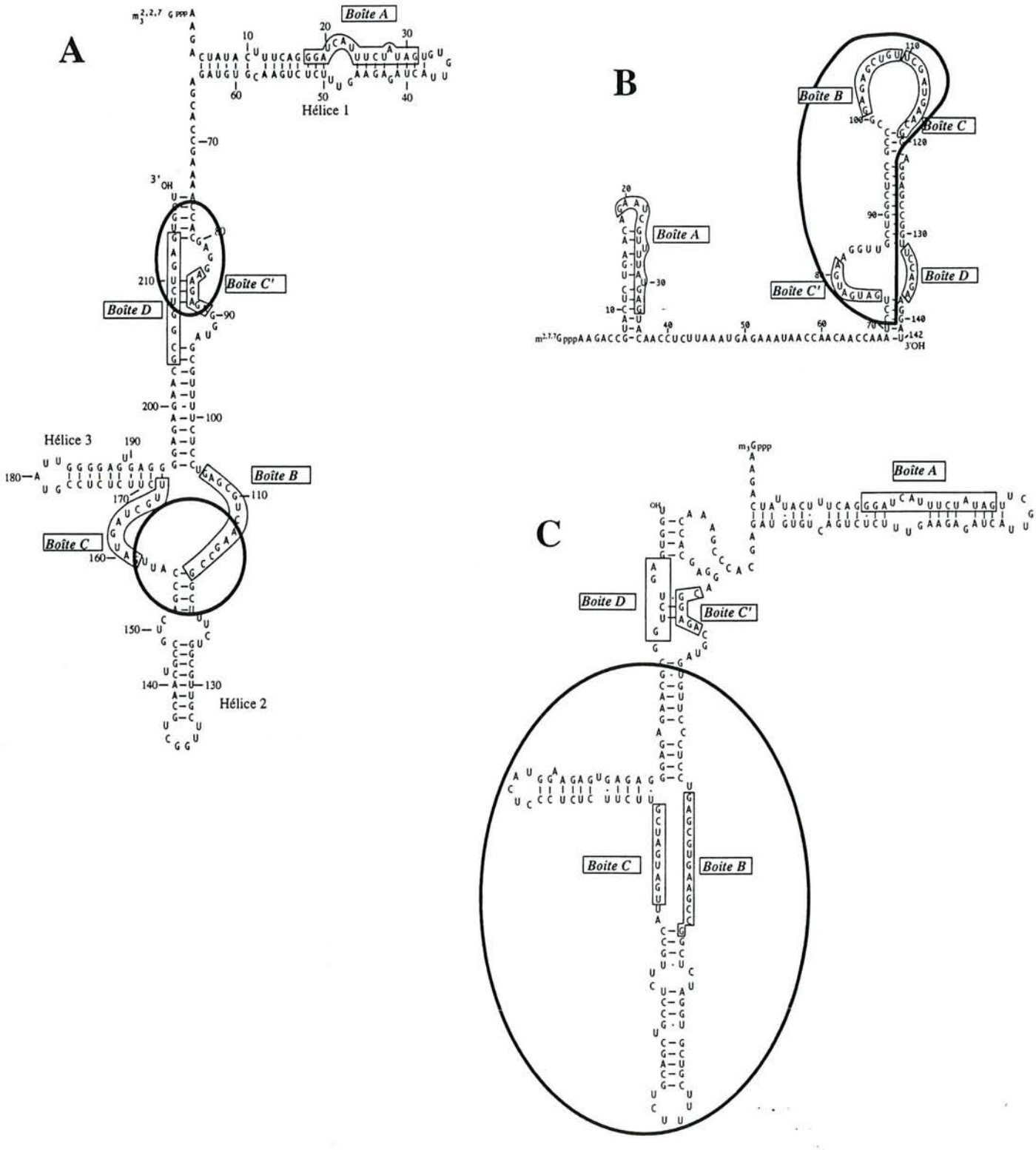


Figure 9 : Régions de fixation des protéines au sein des snoRNAs U3 de cellules HeLa (A), de *Trypanosoma brucei* (B) et de Hamster (C) (Parker et Steitz, 1987 ; Hartshorne et Agabian, 1994 ; Lübben *et al.*, 1993).

Les boîtes phylogénétiquement conservées A, B, C, C' et D sont indiquées. Les régions protégées contre l'action des RNases au sein des snoRNP de cellules HeLa (A) et de *Trypanosoma brucei* (B) sont indiquées, ainsi que la région impliquée dans la fixation de la protéine de 55kDa dans la snoRNP U3 de hamster (C).

A3.3 Les snoRNP U3.

Le complexe ribonucléoprotéique contenant le snoRNA U3 a été isolé pour la première fois en 1984 par Epstein *et al.* En analysant des extraits nucléolaires de cellules d'hépatome de rat, ces auteurs ont observé que le snoRNA U3 sédimente à un coefficient de 60S lorsqu'il est associé au pré-ARNr et de 20S lorsqu'il n'est pas associé à l'ARN pré-ribosomique. Un tel coefficient de sédimentation ne peut correspondre à un ARN libre. Donc, même lorsque le snoRNA U3 n'est pas associé au pré-ARNr, il est lié à des protéines, l'ensemble constituant une particule ribonucléoprotéique.

Jusqu'à présent, par rapport au nombre important de snoRNA U3 caractérisés, peu de snoRNP ont été étudiées.

En 1987, Parker et Steitz ont obtenu une fraction hautement purifiée en snoRNP U3 en utilisant un anticorps dirigé contre la fibrillarine. Cet anticorps était un anticorps auto-immun, produit par des patients atteints de sclérodermie (Lischwe *et al.*, 1985). L'analyse de cette fraction enrichie mettait en évidence la présence d'au moins 6 polypeptides de poids moléculaires 12,5, 13, 30, 36, 59 et 74 kDa, les 4 premières protéines étant non phosphorylées, les deux dernières étant phosphorylées. La protéine de 36 kDa correspond à la fibrillarine.

Plus récemment, Lübben *et al.* (1993) ont purifié une snoRNP U3 à partir de cellules CHO de hamster, en utilisant une chromatographie sur colonne d'affinité avec un anticorps anti-m^3G, suivie d'une chromatographie sur colonne échangeuse d'anions (monoQ). La particule ainsi purifiée ne renferme que 3 protéines de 15, 50 et 55 kDa. Ce procédé de purification qui implique un traitement à forte concentration ionique, ne permet donc pas d'obtenir une snoRNP intacte, mais seulement une particule renfermant les protéines les plus fortement associées à l'ARN. Dans ces conditions, la fibrillarine est dissociée, ce qui est en faveur d'une fixation indirecte de la fibrillarine, au travers d'interactions protéine/protéine.

Parker et Steitz (1987) avaient essayé de définir les sites de fixation des protéines sur la snoRNP U3 de cellules HeLa en comparant l'accessibilité de l'ARN déprotéinisé et de l'ARN au sein de la snoRNP, ceci en utilisant à la fois des sondes chimiques et des sondes enzymatiques. Deux zones de protection avaient ainsi été observées correspondant l'une aux boîtes conservées D et C', l'autre aux boîtes B et C (Fig. 9). Une protection de l'ensemble de la structure tige/boucle enfermant ces 4 boîtes dans l'ARN de *Trypanosoma brucei* a aussi été récemment observée (Fig. 9) (Harsthorne et Agabian, 1994).

Vu l'instabilité de la particule isolée par Lübben *et al.* (1993) à partir de cellules CHO, ces auteurs ont utilisé une autre approche pour essayer d'identifier le site de fixation de la protéine de 55 kDa associée à cette protéine. Une étude de la fixation de la protéine de 55 kDa sur des snoRNA U3 tronqués a été réalisée. Les résultats obtenus montrent que cette protéine de 55 kDa se fixe dans la région renfermant les boîtes B et C.

Comme nous l'avons déjà mentionné dans le cas de *S. cerevisiae*, seules deux protéines se fixant aux snoRNA U3 ont été identifiées, NOP1, l'équivalent de la fibrillarine et SOF1 (Jansen *et al.*, 1993). P. Fabrizio, au laboratoire du Pr R. Lührmann (Marburg, Allemagne), a pu obtenir une fraction très fortement enrichie en snoRNP U3 de *S. cerevisiae*, en utilisant une chromatographie d'affinité sur une colonne portant un anticorps dirigé contre le résidu triméthyl G de la coiffe et l'étude de l'accessibilité du snoRNA U3 au sein de cette snoRNP purifiée a été menée au laboratoire par Anne Grégoire, étude qui a permis d'observer une protection des boîtes B, C, C' et D.

L'ensemble de ces données est en faveur d'un rôle de ces 4 boîtes conservées dans la fixation des protéines. Les boîtes conservées A' et A dans le domaine 5' qui ne sont pas protégées au sein de la snoRNP, ayant plus vraisemblablement un rôle fonctionnel dans la maturation de l'ARN pré-ribosomique.

La biosynthèse des snRNP spliceosomales U1, U2 et U5 comporte une étape cytoplasmique. C'est en particulier dans le cytoplasme qu'a lieu la double méthylation en position 2 du résidu G de la coiffe en 5'. Un passage dans le cytoplasme du snoRNA U3 a été proposé chez les vertébrés (Baserga *et al.*, 1992), et il a même été proposé d'après des études de mutagenèse que cette étape nécessite la présence de l'appariement impliquant les 5 nucléotides 3' terminaux de la molécule, ainsi que la boîte D. Néanmoins, des résultats récents de Terns et Dahlberg (1994) montrent que la double méthylation de la coiffe aurait lieu dans le noyau contrairement à ce qui se passe pour les UsnRNA spliceosomaux. Pour l'instant, ces données sont donc quelques peu contradictoires.

A3.4 Mécanisme d'action du snoRNA U3.

Comme nous l'avons déjà mentionné ci-dessus, le rôle essentiel du snoRNA U3 dans l'étape de clivage en A0 a tout d'abord été mis en évidence chez les vertébrés, après la mise au point d'un système *in vitro*, où cette étape pouvait être réalisée (Miller et Sollner-Webb, 1981; Kass *et al.*, 1987).

L'élimination spécifique du snoRNA U3 de l'extrait par hydrolyse par la RNase H en présence d'oligonucléotides complémentaires au snoRNA U3 a permis de montrer le caractère essentiel de U3 pour cette étape (Savino et Gerbi, 1990).

Hughes et Ares (1991) ont construit une souche de *S. cerevisiae* où le gène B était interrompu et le promoteur du gène A placé sous le contrôle de l'opérateur Gal10. Le gène U3A pouvait ainsi être exprimé de manière conditionnelle. Lorsque l'expression du gène U3A est bloquée par croissance sur glucose, les coupures aux sites A0, A1 et A2 sont inhibées, et l'ARNr 17S n'est pas produit (Hughes et Ares, 1991).

Restait à savoir comment le snoRNA U3 intervenait dans ce processus. Dès les premières études sur le snoRNA U3 (Zieve et Penman, 1976; Busch *et al.*, 1982; Epstein *et al.*, 1984), il avait été

postulé que cet ARN intervient par appariement de bases avec l'ARN pré-ribosomique. En effet, le snoRNA U3 reste associé à l'ARN pré-ribosomique après déprotéinisation.

Le problème était de savoir où se formait l'appariement sur l'ARN pré-ribosomique, un très grand nombre de modèles d'interaction ont été proposés. Les premiers étaient basés sur l'observation de complémentarités de séquences entre le snoRNA et la région espaceur ITS 2 (Crouch *et al.*, 1983; Bachellerie *et al.*, 1983; Tague et Gerbi, 1984), par analogie avec le mécanisme d'action du snRNA U1 dans l'épissage des introns, il était proposé que le snoRNA U3 s'apparie aux extrémités de la région ITS 2, afin de permettre l'excision de cette région espaceur. La détermination de séquences des régions ITS 2 et de snoRNA U3 de différentes espèces a montré que les appariements proposés n'étaient pas conservés phylogénétiquement et n'avaient donc pas de signification biologique.

Sur la base des études de structure secondaire des snoRNA U3 de cellules HeLa (Parker et Steitz, 1987) et de Xénope (Jeppensen *et al.*, 1988), un autre modèle a alors été proposé. Selon ce modèle, la région 105-115 (la boîte B) s'apparierait, soit avec la région 3' ETS, soit à l'extrémité 3' de l'ARNr 28S. En étudiant des complexes formés *in vitro* entre des fragments d'ARN pré-ribosomiques et un extrait nucléaire, par immunoprécipitation par des anticorps dirigés contre la snoRNP U3, Parker *et al.* (1988) ont également proposé un autre type d'interaction entre le snoRNA U3 et la région 28S du pré-ARNr.

Aucune des interactions proposées ne s'est avérée conservée au cours de l'évolution et en plus, aucune n'a été confirmée par les méthodes de pontages covalents. C'est, en effet, cette approche qui a permis de définir les régions d'interaction.

En 1981, Calvet et Pederson avaient déjà utilisé le psoralène comme agent de pontage des régions d'ARN appariées pour montrer la formation d'un pontage au psoralène entre le pré-ARNr et le snoRNA U3. La région du pré-ARNr en interaction a été mieux définie, plus tard, chez le rat (Stroke et Weiner, 1989) et chez l'Homme (Maser et Calvet, 1989). Il s'agit d'un segment de la région 5' ETS proche du site de clivage A0.

Le même type d'étude a ensuite été réalisé chez la levure *S. cerevisiae* aussi bien *in vivo* qu'*in vitro*, par Beltrame et Tollervey (1992) et chez la souris par Tyc et Steitz (1992) dans des expériences *in vitro* avec des extraits cellulaires de souris. Dans les deux cas, les résultats suggéraient une interaction entre le domaine 5' du snoRNA U3 et la région 5' ETS du pré-ARNr (juste en amont du site de premier clivage A0) (Fig. 10).

Les pontages covalents obtenus au niveau du snoRNA U3 de *S. cerevisiae* sont situés dans la boîte A et dans un segment situé en aval (positions 39-49) qui n'est pas conservé au cours de l'évolution (Fig. 10).

Dans ces expériences, les nucléotides de la boîte A étaient pontés à la région +665 de l'ARN pré-ribosomique, aucune complémentarité significative n'existe entre les 2 ARN à ce niveau. Par contre, le segment 39-49 du snoRNA U3 était ponté au segment 469-479 de la région 5' ETS de

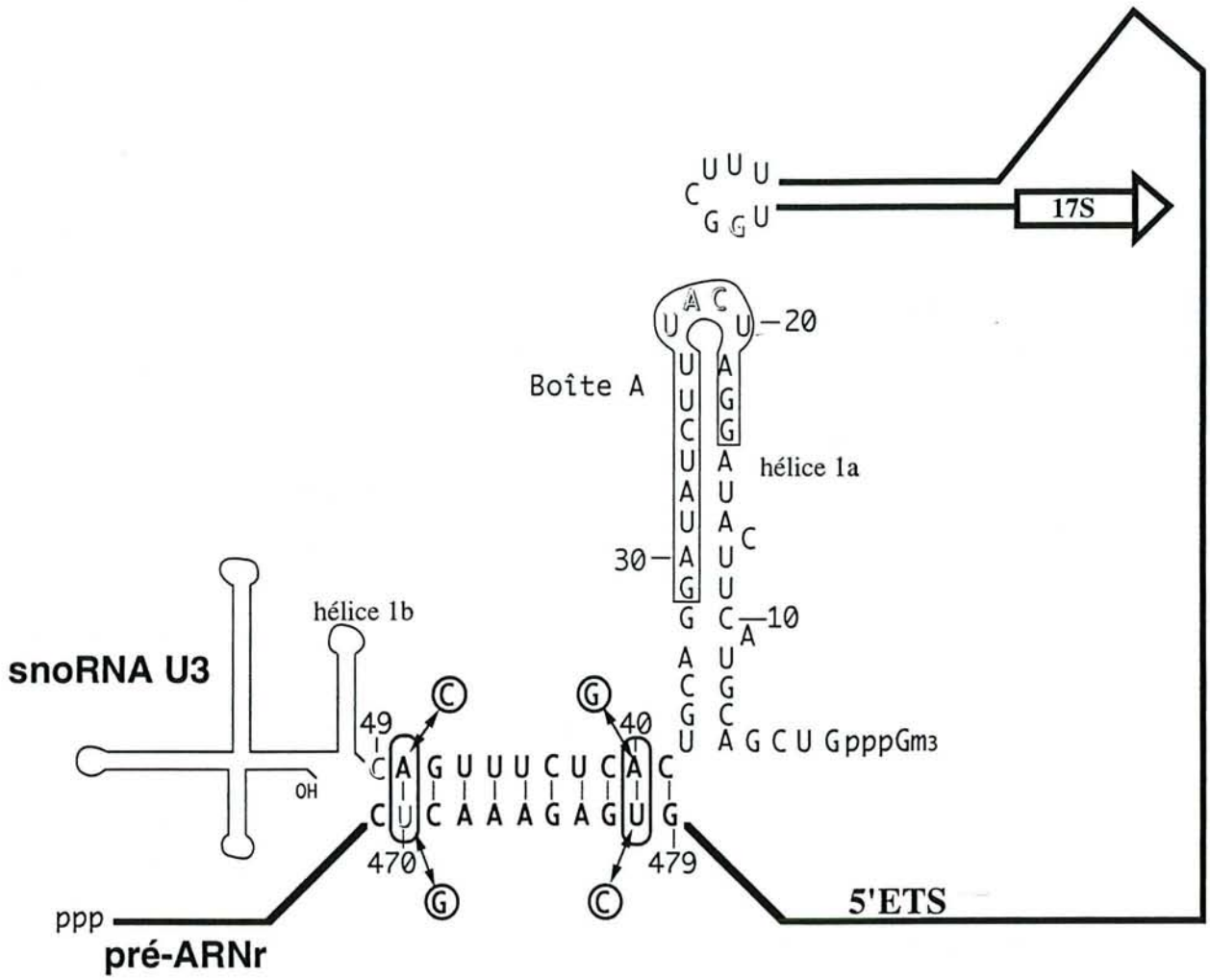


Figure 10 : Appariement entre le snoRNA U3 et la région 5'ETS du précurseur des ARN ribosomiques chez la levure *S. cerevisiae*.

L'appariement proposé par Beltrame et Tollervey (1992), d'après des expériences de pontages aux UV, *in vivo*, entre ces 2 ARN, a été confirmé par voie génétique (Beltrame et Tollervey, 1995) et par des données phylogénétiques (Brulé *et al.*, 1996).
Les nucléotides pontés sont indiqués en ombré et les mutations compensatoires observées dans les 2 séquences de *Hansenula wingei* sont indiquées par les nucléotides encerclés. La boîte A est encadrée.

l'ARN pré-ribosomique, et là, une complémentarité de 10 nucléotides est observée (Fig. 10) (Beltrame et Tollervey, 1992).

Cette région 39-49 du snoRNA U3 est en simple brin dans le modèle établi pour le snoRNA U3 en solution (Ségault *et al.*, 1992), ainsi que dans le modèle établi pour l'ARN au sein de la snoRNP par Anne Grégoire.

La délétion de 23 nucléotides du segment complémentaire au snoRNA U3 entre les positions +469 et +491 de la région 5' ETS conduit a un phénotype similaire à celui de l'inhibition de l'expression du gène du snoRNA U3 (Hughes et Ares, 1991) ou des gènes des protéines NOP1 et SOF1 (Tollervey *et al.*, 1991; Jansen *et al.*, 1993), d'où l'idée que l'appariement potentiel joue un rôle important. L'importance de cet appariement a été démontrée par deux approches, l'une génétique, l'autre phylogénétique. En effet, d'une part Beltrame *et al.* (1994) ont montré que des mutations dans la région 39-49 du snoRNA U3 ont le même effet sur la maturation du pré-ARNr que des mutations dans la région complémentaire du pré-ARNr. D'autre part, des mutations compensatoires dans la région 39-49 du snoRNA U3 compensent l'effet de mutation dans le segment complémentaire de la région 5' ETS.

C'est ce qui s'est passé au cours de l'évolution. En effet, j'ai montré au laboratoire, en collaboration avec l'équipe de D. Tollervey que chez la levure *Hansenula wingei*, deux mutations présentes dans le segment 39-49 du snoRNA U3 compensent deux mutations présentes dans le segment complémentaire de la région 5' ETS du pré-ARNr (Fig. 10). Cette étude a donné lieu à une publication dans RNA (Brulé *et al.*, 1996).

Les clivages aux sites A1 et A2 sont également inhibés par la délétion ou l'inhibition de l'expression des gènes codant pour les snoRNA U3, snR10, U14 et snR30 (Li *et al.*, 1990 ; Hughes et Ares, 1991 ; Morrissey et Tollervey, 1993) et des gènes codant pour des protéines qui leur sont associées, NOP1, GAR1, SOF1 (Tollervey *et al.*, 1991 ; Girard *et al.*, 1992 ; Jansen *et al.*, 1993). Mais le clivage en A0 ne requiert que le snoRNA U3 et les protéines constituant la snoRNP U3, telles que NOP1 ou SOF1 (Beltrame *et al.*, 1994). Le clivage en A0 serait donc spécifiquement assuré par la snoRNP U3. Cependant, la mise en évidence d'un équivalent de la RNase III procaryote chez *S. cerevisiae*, dont l'action serait notamment le clivage au site A0, laisse supposer que le snoRNA U3 jouerait plus un rôle structural que catalytique dans ce clivage, favorisant ainsi l'accès du site A0 à la RNase III (Elela *et al.*, 1996).

Des travaux très récents de J. Hughes (1996) conduisent cet auteur à proposer un nouvel appariement entre le snoRNA U3 et le pré-ARNr. Il impliquerait une séquence très conservée phylogénétiquement du snoRNA U3 (la boîte A) et 3 séquences très conservées de l'ARNr 18S, impliquées dans une structure pseudo-noeud également conservée phylogénétiquement (Neefs *et al.*, 1993; Gutell *et al.*, 1994) (Fig. 11). Le snoRNA U3 aiderait donc le pré-ARNr à adopter cette structure pseudo-noeud qui protégerait le pré-ARNr d'une dégradation rapide. Cet appariement snoRNA U3/pré-ARNr est, de plus, situé non loin du site de coupure A1. D'ailleurs, des mutations

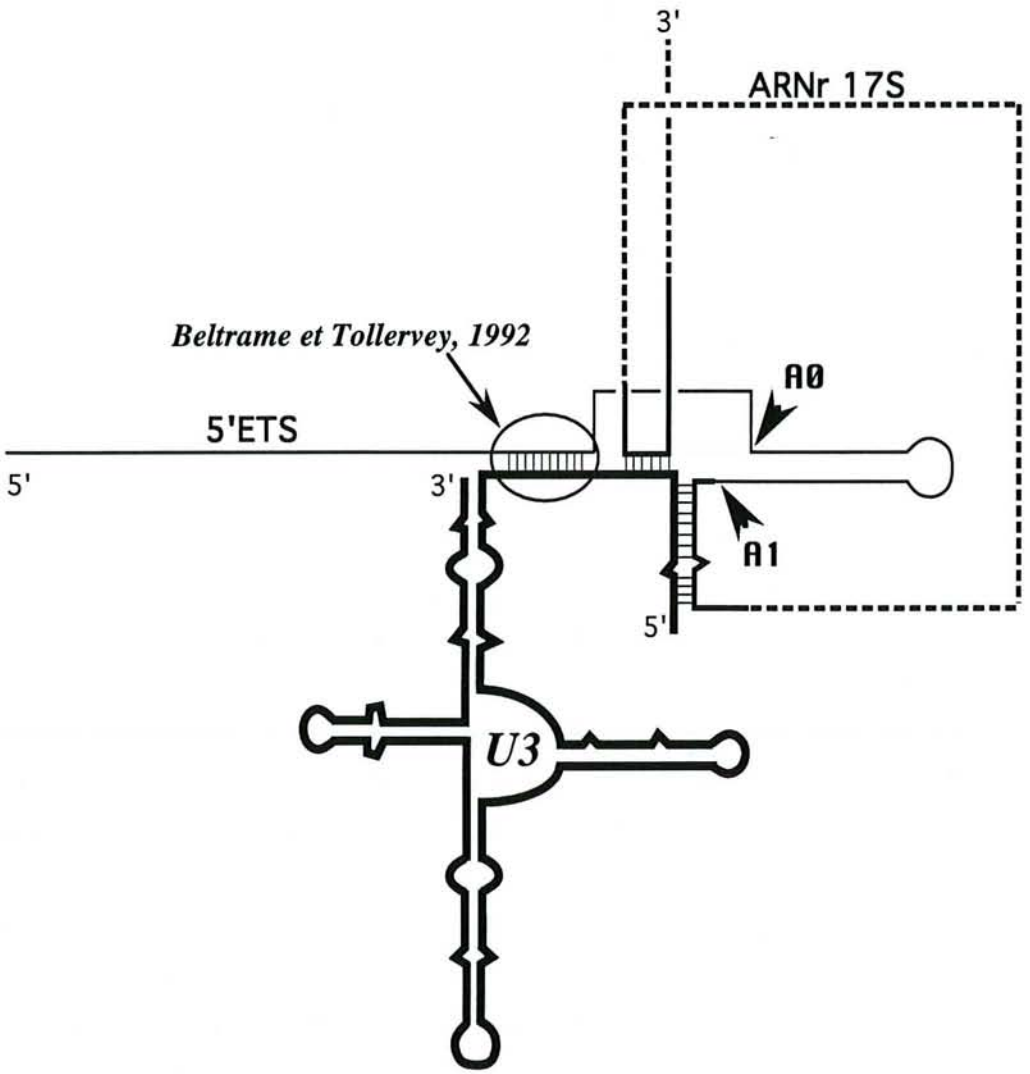


Figure 11 : Résumé des interactions snoRNA U3/pré-ARNr chez *S. cerevisiae*.
(Hughes, 1996)

dans la région concernée du snoRNA U3, dans la boîte A donc, affectent les clivages aux sites A1 et A2. Ceci n'est pas vrai en ce qui concerne le clivage au site A0, qui, lui, dépendrait donc plutôt de l'interaction snoRNA U3/pré-ARNr mise en évidence par Beltrame et Tollervey (1992) (Fig. 11).

A3.5 Les gènes des snoRNA U3.

A3.51 Transcription des gènes U3.

Comme nous l'avons déjà mentionné ci-dessus, les gènes du snoRNA U3 sont transcrits par l'ARN polymérase II, excepté chez les plantes (Marshallsay *et al.*, 1990) et les organismes inférieurs appartenant à ce phylum, tel que *Chlamydomonas reinhardti* (Solymosy *et al.*, résultats non publiés), où ils sont transcrits par l'ARN polymérase III.

Les gènes U3 de vertébrés contiennent des signaux de transcription similaires à ceux observés dans les gènes codant pour les UsnRNA (U1, U2, U4 et U5), gènes aussi transcrits par l'ARN polymérase II (Dahlberg et Lund, 1987) :

un élément proximal entre les positions -62 et -44

[GT(G/C)ACCG(T/C)GNGT(T/A)AA(G/T)GTG]

un élément renforçateur (-232, -219)

[TATG(T/C)TAAT(T/C)(A/G)TGG]

et un signal de terminaison de la transcription présent en aval de la région codante entre les positions +16 et +31

[GTTTNAAA(A/G)NNAGA].

Les gènes U3 présentent en plus un motif additionnel qui leur est spécifique. Ce motif [GATTGGCT(A/G)(T/C)] est appelé "U3 box", il est situé dix paires de bases en amont de l'élément renforçateur.

Chez les plantes, les séquences promotrices des gènes codant pour les UsnRNA spliceosomaux sont différentes de celles trouvées chez les mammifères. Les gènes U3, chez les plantes, présentent un élément distal et un élément proximal (boîte TATA "like") trouvés à la fois dans les gènes du snRNA U2 (Vankan et Filipowicz, 1989) et du snRNA U6 (Waibel et Filipowicz, 1990a, 1990b). Ces deux ARN sont, respectivement, transcrits par l'ARN polymérase II et l'ARN polymérase III. La seule différence apparente entre les deux régions promotrices des gènes codant pour ces deux UsnRNA est la distance entre l'élément "distal" et l'élément "proximal" : 32 à 36 nucléotides pour les gènes U2 (transcrits par l'ARN polymérase II) et 24 à 26 nucléotides pour les gènes U6 (transcrits par l'ARN polymérase III). L'élément "distal" et l'élément "proximal", identifiés dans les gènes codant pour l'ARN U3 de plantes, sont distants de 24 à 26 nucléotides, comme pour les gènes U6 (Kiss et Solymosy, 1990). Kiss *et al.* (1991) ont pu, à partir de ces observations démontrer que la distance entre ces deux boîtes est le paramètre qui détermine la spécificité de reconnaissance par l'ARN polymérase II ou III. Ils ont en effet converti le gène U3 de plante en un

gène transcrit par l'ARN polymérase II, simplement en changeant la distance entre l'élément "distal" et l'élément "proximal" (distance augmentée d'un tour d'hélice).

Tous les snoRNA U3 transcrits par l'ARN polymérase II ont une coiffe m^3G, alors que les snoRNA U3 transcrits par l'ARN polymérase III ont en 5' un γmonométhyl triphosphate (Shimba *et al.*, 1992) comme le snRNA U6 de vertébrés (Singh et Reddy, 1989).

A3.52 Organisation des gènes.

Chez l'Homme (Suh *et al.*, 1986 : Suh *et al.*, 1991), le rat (Stroke et Weiner, 1985) et la souris (Mazan et Bachellerie, 1988), le snoRNA U3 est codé par une famille de gènes qui comporte cinq à dix copies de gènes "fonctionnels" (dont la séquence codante correspond à celle de l'ARN ou des ARN cellulaires et qui portent des séquences promotrices) et de nombreux pseudogènes.

Peu de choses sont connues sur l'organisation des familles de gènes U3 chez les métazoaires. Chez le rat ou l'Homme, la présence de groupes de gènes n'est pas exclue, mais ils doivent être séparés par 10 à 15 kb (Stroke et Weiner, 1985 ; Yuan et Reddy, 1989). Chez la souris, l'étude poussée menée par Mazan et Bachellerie (1990) a mis en évidence une famille de quatre gènes fonctionnels, situés sur un même locus du chromosome II. Trois de ces gènes sont regroupés sur un fragment de 20 kb, dans une même orientation. Chez les eucaryotes inférieurs, tels que *Distyostelium discoïdeum* (Wise et Weiner, 1980), *Schizosaccharomyces pombe* (Porter *et al.*, 1988) ou *Saccharomyces cerevisiae* (Hughes *et al.*, 1987), les gènes du snoRNA U3 sont présents à un nombre réduit de copies (deux à cinq).

Il faut noter qu'il existe souvent deux ou trois formes (variants) de snoRNA U3, c'est le cas chez la souris (Mazan et Bachellerie, 1988), mais aussi chez la levure *S. cerevisiae* (2 variants A et B). Souvent l'une de ces deux formes est prédominante. Chez *S. cerevisiae*, dans des conditions de croissance en laboratoire, seule la forme A semble exprimée (Hughes *et al.*, 1987; Myslinski *et al.*, 1990). Par contre, lorsque le gène A est délété, le gène B prend le relais. Ce gène, qui code pour un ARN qui présente 8% de divergence par rapport au snoRNA U3A, est donc fonctionnel (Hughes *et al.*, 1987 ; Myslinski *et al.*, 1990).

Par ailleurs, les gènes U3 de levure *S. cerevisiae* présentent une particularité mise en évidence au laboratoire (Myslinski *et al.*, 1990): les deux gènes renferment un intron épissé dans un spliceosome. Avec la découverte d'un intron du même type dans le gène du snRNA U6 de *S. pombe* (Tani et Ohshima, 1989), ceci constituait les deux premières observations d'introns épissés dans un spliceosome dans des gènes codant pour des ARN métaboliquement stables. Cette observation avait remis en question l'idée que les introns épissés dans un spliceosome étaient uniquement contenus dans les gènes nucléaires de protéines. Plus tard, des introns de type intron de gène de protéine ont également été mis en évidence dans les gènes des snRNA U6 d'autres espèces de levure [*Schizosaccharomyces japonicus* var. *japonicus*, *Schizosaccharomyces japonicus* var. *versatilis*, *Schizosaccharomyces malidevorans*, *Schizosaccharomyces octosporus* (Frendewey

et al., 1990; Reich et Wise, 1990), *Rhodotorula hasegawae* et *Rhodosporium dacroydium* (Tani et Ohshima, 1991)] et dans le gène codant pour le snRNA U2 de la levure *Rhodotorula hasegawae* (Takahashi *et al.*, 1993). Au cours de ma thèse, j'ai mis en évidence ce même type d'introns dans les gènes du snoRNA U3 de différentes levures appartenant à la sous-famille des Saccharomycetoideae.

Dans l'état actuel des connaissances, les levures sont les seuls organismes où des introns épissés dans un spliceosome ont été mis en évidence dans les gènes d'ARN métaboliquement stables. Ceci est d'autant plus surprenant qu'au niveau gène nucléaire de protéine, très peu d'introns sont observés chez ces organismes. Seuls 2 à 3% des gènes de protéines chez *S. cerevisiae* présentent un intron (Woolford, 1989), de petite taille (<1 kb) localisé à l'extrémité 5' des gènes (Lamond, 1993) ou dans une région en 5' non traduite (Kalogeropoulos, 1995). Ces introns sont en général uniques à deux exceptions près [Mat α1 (Miller, 1984) et YLA8 (Mizuta *et al.*, 1992)].

Par ailleurs, ces introns ont uniquement été mis en évidence dans des gènes de petits ARN nucléaires, ce qui suggère qu'ils pourraient s'être insérés au départ dans les ARN par un processus d'épissage inverse avec retour de l'information au niveau ADN par transcription inverse. Ceci est en accord avec l'observation de la localisation de ces introns au niveau des régions fonctionnelles des UsnRNA spliceosomaux, supposées appartenir au site catalytique du spliceosome (Tani et Ohshima, 1991; Takahashi *et al.*, 1993).

Dans le cas du snoRNA U3, l'intron est aussi inséré dans une région fonctionnelle puisqu'il est situé entre les boîtes conservées A' et A du domaine fonctionnel en 5'.

Cet intron de type intron de gène nucléaire dans un gène d'ARN métaboliquement stable est éliminé par le spliceosome. C'est pourquoi dans cette introduction bibliographique, je ferai maintenant un bilan de l'état des connaissances sur la composition, l'assemblage et le mécanisme d'action du spliceosome, ceci en présentant plus en détail les connaissances chez *S. cerevisiae*, organisme qui a fait l'objet de mes travaux de thèse.

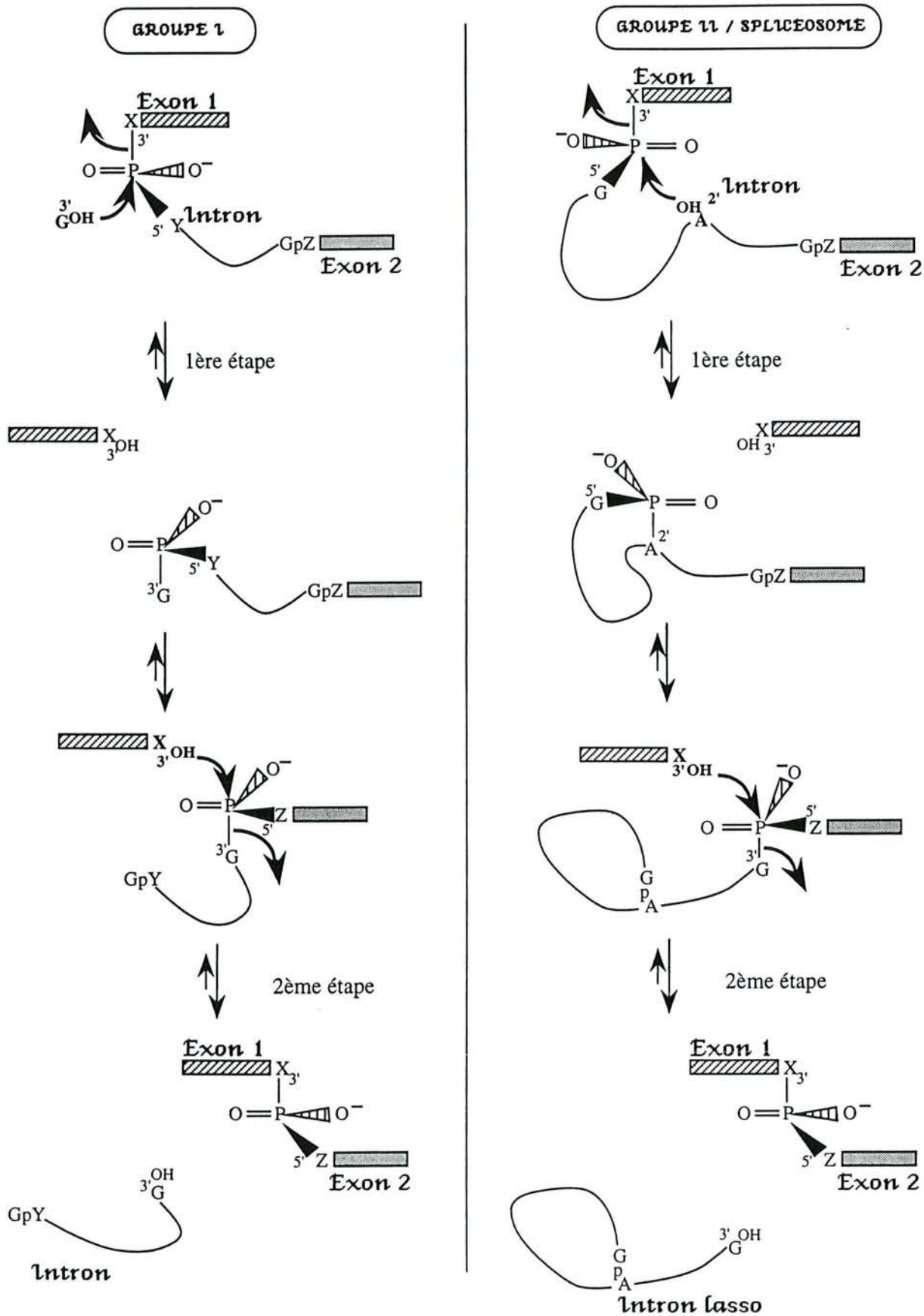


Figure 12 : Les étapes de la réaction d'épissage des introns de groupe I (à gauche), et des introns de groupe II et des ARN pré-messagers nucléaires (à droite).
(Cech 1985; 1986 ; 1987 ; Peebles *et al.* 1986; Konarska *et al.*, 1985; Grabowski *et al.*, 1985)

B. La maturation des introns dans un spliceosome

Les gènes nucléaires de protéines chez les eucaryotes, ainsi que, et comme nous l'avons vu, les gènes de UsnRNA chez les levures, mais aussi les gènes d'ARN ribosomiques nucléaires chez un nombre limité d'espèces et les gènes nucléaires d'ARNt chez les levures renferment des séquences excédentaires qui doivent être éliminées au niveau du transcrit primaire pour qu'il y ait production d'ARN mature. La découverte, en 1977, de telles séquences introniques a valu le prix Nobel à P.A. Sharp et à R.J. Roberts (Berget *et al.*, 1977; Chow *et al.*, 1977; Roberts, 1978). Cette étape de la maturation des ARN pré-messagers est appelée épissage. Les gènes de protéines nucléaires, chez les eucaryotes, peuvent contenir jusqu'à une cinquantaine d'introns, dont la taille peut varier d'une dizaine à une dizaine de milliers de nucléotides (les génomes d'organismes d'eucaryotes supérieurs renfermant beaucoup plus d'introns que ceux d'eucaryotes inférieurs). Chez les procaryotes, la présence d'introns est rare : quelques gènes d'archaebactéries (gènes codant pour des ARN de transfert ou pour le grand ARNr) (Kaine *et al.*, 1983; Kjems *et al.*, 1985) ou de cyanobactéries (gènes codant pour des ARN de transfert) (Kuhsel *et al.*, 1990; Xu *et al.*, 1990a) possèdent des introns. Des introns ont également été découverts dans des bactériophages de bactéries à Gram négatif (Chu *et al.*, 1986; Ehrenman *et al.*, 1986) et à Gram positif (Goodrich-Blair *et al.*, 1990).

On peut classer les introns en trois grandes catégories en fonction de leur mécanisme d'excision :

les introns éliminés par voie enzymatique classique. Il s'agit essentiellement des introns des gènes codant pour les ARN de transfert (Peebles *et al.*, 1983).

les introns auto-épissables, introns de groupe I et II. Ils sont largement répandus dans l'ensemble du monde vivant, dans les bactériophages de bactéries à Gram + et - (Chu *et al.*, 1986; Ehrenman *et al.*, 1986; Goodrich-Blair *et al.*, 1990), dans les génomes mitochondriaux, chloroplastiques et dans les gènes nucléaires du grand ARN ribosomique de quelques espèces. Dans les deux cas (intron de groupe I ou II), le mécanisme d'excision repose sur deux étapes de transestérification (Fig. 12).

Pour les introns de groupe I, le premier clivage au site 5' d'épissage implique une molécule extérieure, une guanosine, qui se lie à un site généré par le repliement de la séquence intronique (Cech, 1985). C'est l'hydroxyle en 3' de cette guanosine qui réalise la première attaque nucléophile. Dans le cas des introns de groupe II, la première attaque nucléophile, conduisant au clivage au site 5' d'épissage, est réalisée par l'hydroxyle en 2' d'une adénosine interne de l'intron, activée par la structure formée par l'intron. Il en résulte un intermédiaire en lasso (Peebles *et al.*, 1986; Van der Veen *et al.*, 1986). La deuxième étape de la réaction est similaire pour les 2 types d'introns. Les introns de groupe I, comme les introns de groupe II, possèdent donc toutes les informations nécessaires à leur excision. L'étude de différents introns de groupe I, d'une part, et de

informations nécessaires à leur excision. L'étude de différents introns de groupe I, d'une part, et de groupe II, d'autre part, a permis de mettre en évidence une structuration secondaire commune à chacune de ces classes d'introns (Michel *et al.*, 1982; Michel et Jacquier, 1987; Burke *et al.*, 1987; Cech, 1988). Par ailleurs, un modèle de la structure 3D du site catalytique a pu être établi pour les introns de groupe I (Michel et Westhof, 1990). De plus, très récemment, une partie du site catalytique a pu être cristallisé et le cristal étudié par diffraction aux Rayons X (E. Westhof, résultats non publiés), les résultats obtenus confirment et précisent le modèle établi par Michel et Westhof (1990).

En plus de la structuration secondaire, les introns de groupe II présentent des séquences conservées, en particulier au niveau des sites 5' et 3' d'épissage (Michel *et al.*, 1989). Le degré de conservation de séquences est moins important dans le cas des introns de groupe I. Par ailleurs, il ne s'agit pas des mêmes séquences. Les contraintes structurales sont donc très importantes pour ces deux types d'introns puisqu'ils doivent se replier pour créer un site catalytique.

Il faut noter que les introns de groupe I ou II ne possèdent pas tous une capacité d'autocatalyse, certains nécessitent l'action de protéines pour augmenter leur efficacité catalytique (pour revue, Michel et Ferat, 1995).

les introns contenus dans les ARN pré-messagers nucléaires. L'épissage des introns des ARN pré-messagers implique les 2 mêmes étapes de transestérification que l'épissage des introns de groupe II (Fig. 12), à savoir, une attaque nucléophile du phosphate en 5' de l'intron par l'hydroxyle en 2' d'un résidu adénosine de l'intron situé en amont du site d'épissage en 3' (cette étape libère le premier exon et un intermédiaire en forme de lasso), suivie d'une seconde attaque nucléophile du phosphate en 3' de l'intron par l'hydroxyle en 3' du premier exon. Cette deuxième étape permet de générer une liaison phosphodiester entre les 2 exons et de libérer l'intron (Fig. 12) (Ruskin *et al.*, 1984; Padgett *et al.*, 1984; Konarska *et al.*, 1985).

Ces 2 attaques nucléophiles sont catalysées par un large complexe ribonucléoprotéique, le spliceosome, (Grabowski *et al.*, 1985; Brody et Abelson, 1985) qui s'assemble autour de l'intron et des bordures exoniques. Ce spliceosome renferme cinq petits ARN nucléaires (U1, U2, U4, U5 et U6), constituant les composants essentiels de cette machinerie (Grabowski et Sharp, 1986), et un très grand nombre de protéines (environ 60). Les UsnRNA sont assemblés sous la forme de particules ribonucléoprotéiques, les snRNP U1, U2, U4/U6 et U5.

Il faut aussi ajouter que, chez certains protozoaires tels que le trypanosome ou chez certains métazoaires comme le nématode, l'épissage peut avoir lieu entre 2 molécules, par réunion de 2 ARN distincts. Cet épissage en trans, comme l'épissage en cis, met en jeu une structure spliceosomale renfermant des snRNP U2, U5 et U4/U6 (Agabian, 1990).

Enfin, une nouvelle classe d'introns, très peu fréquents, vient d'être mise en évidence chez la drosophile et chez les vertébrés (Hall et Padgett, 1994; Tarn *et al.*, 1995; Tarn et Steitz, 1996;

Mount, 1996). L'épissage de ces introns dits "mineurs" nécessite l'assemblage d'un autre spliceosome renfermant les snRNA U11, U12 et U5 et deux snRNA remplaçant les snRNA U4 et U6.

Nous décrirons plus particulièrement l'épissage en cis des introns majeurs contenus dans les ARN pré-messagers nucléaires, introns qui ont fait l'objet de mon travail de thèse.

Je ferai d'abord une présentation des séquences introniques et exoniques, puis des UsnRNA et UsnRNP que forment ces ARN. Ensuite, nous verrons comment ces UsnRNP s'assemblent autour de l'ARN pré-messager pour former un spliceosome, quelles sont les protéines qui interviennent dans ces processus et enfin, comment des régulations peuvent s'exercer à cette étape de l'expression des gènes.

Les progrès considérables sur la connaissance de la machinerie d'épissage ont pu être faits grâce à la mise au point de systèmes d'épissage *in vitro*, d'abord pour les vertébrés (Padgett *et al.*, 1983), puis pour la levure (Lin *et al.*, 1985). Les approches de génétique chez *S. cerevisiae* ont aussi apporté beaucoup d'informations. Jusque là, les études purement biochimiques avaient surtout été réalisées sur le système HeLa, du fait de la connaissance plus importante de UsnRNP et facteurs protéiques chez HeLa. Mais, très récemment, l'équipe de R. Lührmann a développé des méthodes de purification des UsnRNP de *S. cerevisiae* (Fabrizio *et al.*, 1994).

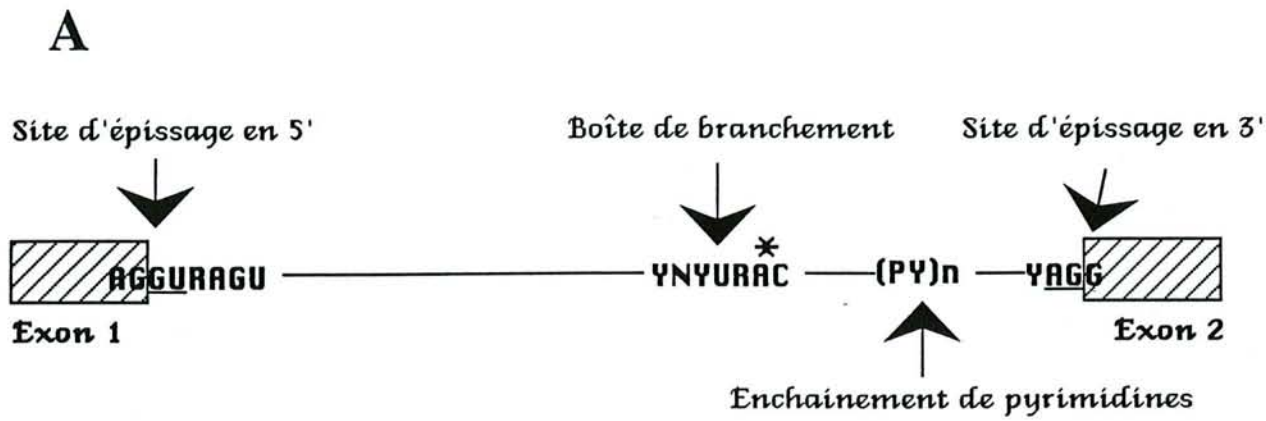
B1. Les ARN impliqués dans l'épissage.

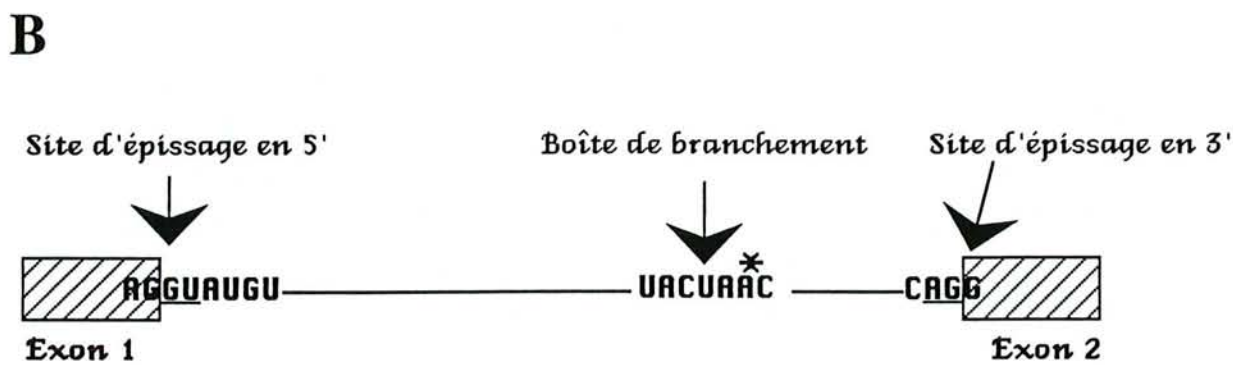
B1.1 Les introns et bordures exoniques de gènes nucléaires de protéines.

Très tôt, l'alignement des séquences introniques a permis de mettre en évidence la conservation de certaines séquences aux extrémités des introns et des bordures exoniques. Les nucléotides les plus conservés sont les dinucléotides GT et AG, localisés en 5' et 3' de tous les introns épissés par la forme majeure de spliceosome, et ceci quelle que soit l'espèce (Mount, 1982) (Fig. 13A et B). Ces 4 nucléotides sont différents chez les introns dits de classe "mineure". En effet, ces quelques introns présentent en 5' et 3' les dinucléotides AT et AC (Hall et Padgett, 1994). Ces introns sont, d'ailleurs, épissés dans un spliceosome différent de celui dans lequel sont épissés la majorité des introns (Mount, 1996) (Fig. 13C).

B1.11 Site d'épissage en 5'.

Chez les eucaryotes supérieurs, la conservation de séquence au site 5' s'étend sur 9 nucléotides (trois en amont du site de clivage et six en aval) : (A/C)AG▼GT(A/G)AGT (Green, 1986). Mais des variations importantes par rapport à la séquence consensus sont tolérées à l'exception du





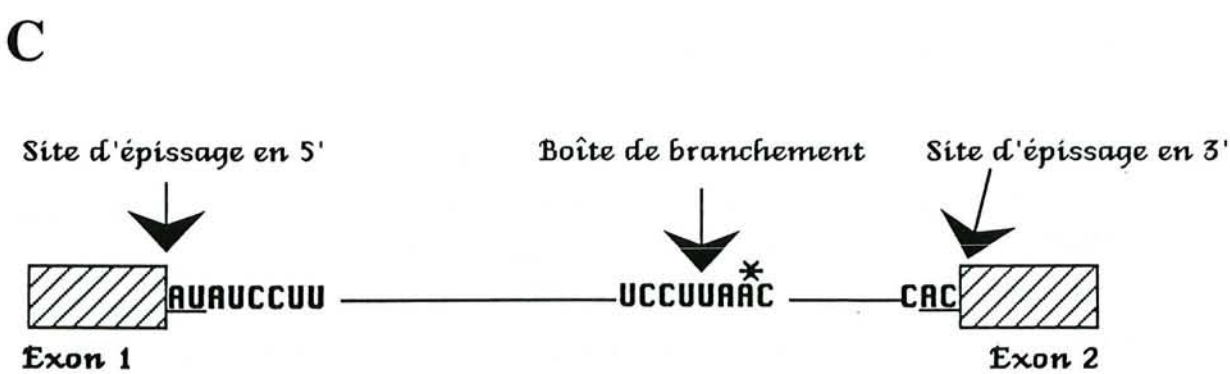


Figure 13 : Séquences introniques essentielles à la réaction d'épissage.
En A, pour les introns majeurs des ARN pré-messagers des eucaryotes supérieurs, en B pour les mêmes introns chez la levure *S. cerevisiae* (pour revue, voir Legrain et Chanfreau, 1994) et en C, pour les introns mineurs récemment mis en évidence chez les vertébrés et les insectes (Tarn et Steitz, 1996; Mount, 1996).
L'astérisque désigne le point de branchement.

dinucléotide GT toujours conservé. Des mutations dans ce dinucléotide entraînent l'activation de sites cryptiques (Krainer *et al.*, 1984).

Chez la levure *S. cerevisiae*, la conservation de séquence en 5' de l'intron est beaucoup plus stricte (GTATGT). Des divergences aux positions G_1, T_2 mais aussi G_5 ne sont pas tolérées (Newman *et al.*, 1985; Fouser et Freisen, 1986; Vijayraghavan *et al.*, 1986). Du fait de cette plus grande exigence de la machinerie d'épissage de levure vis-à-vis de la séquence 5', l'utilisation de sites cryptiques 5' n'est pas observée.

B1.12 Séquence du site de branchement.

Quel que soit l'organisme, c'est, à de rares exceptions près, une adénosine qui réalise la première attaque nucléophile et constitue le point de branchement. Elle est donc importante pour la première réaction d'épissage, mais des études récentes ont montré qu'elle est également reconnue lors de la formation du spliceosome et lors de la deuxième étape de la réaction d'épissage (Query *et al.*, 1996).

Chez les eucaryotes supérieurs, cette adénosine est située 18 à 40 nucléotides en amont du site 3' d'épissage, dans une région (PyTPu**A**Pu) dont le degré de conservation de séquence est faible (l'adénosine soulignée correspond au point de branchement) (Keller et Noon, 1984; Ruskin *et al.*, 1984; Shapiro et Senapathy, 1987).

A nouveau, chez la levure *S. cerevisiae*, la séquence de la boîte de branchement TACTA**A**C est très conservée. De ce fait, l'utilisation de sites cryptiques de branchement est possible chez les métazoaires mais pas chez *S. cerevisiae*. En général, la délétion de la boîte de branchement inhibe totalement la réaction d'épissage (Langford et Gallwitz, 1983; Pikielny *et al.*, 1983), comme les mutations du nucléotide A_6 correspondant au point de branchement (Langford *et al.*, 1984; Newman *et al.*, 1985; Fouser et Friesen, 1986; Vijayraghavan *et al.*, 1986). Seules quelques mutations ponctuelles sont tolérées telles que C_3—>T ou A_2—>C (Jacquier *et al.*, 1985; Vijayraghavan *et al.*, 1986). Jacquier *et al.* (1985) remarquaient également que, si une base mutée dans la boîte 5' ou dans la boîte de branchement n'affectait parfois pas sévèrement l'efficacité de la réaction d'épissage *in vivo*, le cumul de 2 mutations s'avérait bien plus dramatique.

L'intron des gènes U3 que nous avons étudié présente une déviation par rapport à la séquence consensus. La séquence de la boîte de branchement des introns des gènes U3A et U3B de *S. cerevisiae* est, en effet, GACTAAC au lieu de TACTAAC (Myslinski *et al.*, 1990). Ce type de dérivation est rare.

B1.13 Site d'épissage en 3'.

Les deux nucléotides à l'extrémité 3' de l'intron sont toujours AG, aussi bien chez les métazoaires que chez la levure.

Chez les métazoaires, une séquence riche en pyrimidines (70 à 93%), de longueur variable, est fréquemment présente entre les positions -15 à -5 en amont du dinucléotide AG. Cette séquence polypyrimidique est importante pour la fixation de facteurs protéiques du spliceosome et dans la première étape de la réaction d'épissage (Ruskin et Green, 1985a).

Il existe deux classes d'ARN pré-messagers chez les métazoaires, selon que le dinucléotide AG est impliqué ou non dans la première étape de la réaction d'épissage (Reed, 1989). En effet, le dinucléotide AG est nécessaire à la première réaction d'épissage lorsqu'il est précédé d'une séquence polypyrimidines courte (14nt) et il ne l'est pas lorsqu'il est précédé d'une séquence polypyrimidines longue (26nt) (Reed, 1989).

La présence d'une séquence polypyrimidines n'est pas indispensable chez la levure (Rymond et Rosbash, 1985). Il existe également 2 types d'introns chez les levures :

¤ les introns 3'S, les moins fréquents, caractérisés par une proximité entre la boîte de branchement et le site 3' d'épissage (5 à 15 nucléotides), comme les 2 introns présents dans le gène *MATa1* (Miller, 1984),

¤ les introns 3'L, les plus fréquents, pour lesquels la distance entre le site de branchement et le site 3' varie d'une vingtaine à 150 nucléotides, généralement cette distance est de l'ordre de 40 nucléotides (Parker et Patterson, 1987). Le dinucléotide AG est alors précédé d'une séquence N(U/A)APy, elle-même précédée d'une séquence riche en uridines (58 à 100%) entre les positions -7 et -13.

Les introns des 2 gènes U3 de *S. cerevisiae* que j'ai étudiés appartiennent à la première catégorie d'introns de levure : les introns 3'S, la boîte de branchement et le site AG 3' terminal étant séparés par la courte séquence $_{5'}$ACATTC(T/A)AC$_{3'}$ (Myslinski *et al.*, 1990).

Les sites 5' et 3' d'épissage et boîte de branchement sont les segments qui, nous le verrons plus loin, jouent un rôle essentiel dans la reconnaissance et la fixation des éléments constituant le spliceosome.

Les autres régions de l'intron ne présentent pas de conservation de séquence, néanmoins, ces séquences non conservées peuvent jouer un rôle dans la formation du spliceosome et dans la sélection des sites d'épissage. Une distance minimale entre les 3 séquences conservées est requise pour que l'épissage ait lieu (Cellini *et al.*, 1986a; Thompson-Jäger et Domdey, 1987). De plus, la structure secondaire, qui peut être formée par ces régions, peut intervenir dans l'efficacité d'épissage et ceci tout particulièrement chez *S. cerevisiae*. D'ailleurs, l'hypothèse de la nécessité de la formation d'une structure secondaire rapprochant la boîte 5' et la boîte de branchement pour qu'il y ait épissage des longs introns de levure avait été formulée dès 1987 (Parker et Patterson, 1987). L'équipe de M. Rosbash a, depuis, travaillé sur ce sujet : l'efficacité d'épissage du pré-ARNm RP51B dépendrait en effet de 2 séquences complémentaires situées, l'une en aval du site 5' d'épissage et l'autre en amont de la boîte de branchement (Goguel et Rosbash, 1993; Libri *et al.*,

1995). La démonstration de l'existence d'une telle structure secondaire a été réalisée au laboratoire et a donné lieu à une publication (Mougin *et al.*, 1996), à laquelle j'ai participé.

B1.2 Structure des UsnRNA U1, U2, U4, U5 et U6.

Ces cinq UsnRNA ont été observés chez tous les organismes eucaryotes dans lesquels ils ont été recherchés. Ils sont donc universels chez ces organismes.

Les premières structures primaires complètes et structures secondaires des UsnRNA ont été établies par l'équipe dans laquelle j'ai préparé ma thèse, à l'exception de celle du snRNA U6 :

—> le snRNA U1 (Branlant *et al.*, 1980; Mount et Steitz, 1981) (Fig. 14),

—> le snRNA U2 (Branlant *et al.*, 1982) (Fig. 14),

—> le snRNA U4 (Krol *et al.*, 1981a) (Fig. 14),

—> le snRNA U5 (Krol *et al.*, 1981b ; Branlant *et al.*, 1983; Krol *et al.*, 1983) (Fig. 14),

—> le snRNA U6 (Epstein *et al.*, 1980) (Fig. 15).

Depuis, de très nombreuses séquences de snRNA ont été établies (Gu et Reddy, 1996).

B1.21 Caractéristiques communes.

Les premières séquences établies avaient permis de constater que les snRNA U1, U2, U4 et U5 appartiennent à une même famille, distincte du snRNA U6 (Branlant *et al.*, 1982) :

¤ le snRNA U6 présente une extrémité 5' particulière, avec un phosphate γ-monométhylé (Singh et Reddy, 1989), alors que les autres UsnRNA possèdent une coiffe en 5' qui se distingue de celle des ARNm par la présence de 2 méthyles supplémentaires en position 2 de la guanosine (Reddy et Busch, 1983). Cette différence s'explique par le fait que le snRNA U6 est transcrit par l'ARN polymérase III alors que les snRNA U1, U2, U4 et U5 sont transcrits par l'ARN polymérase II (Kunkel *et al.*, 1986; Krol *et al.*, 1987; Reddy *et al.*, 1987; Moenne *et al.*, 1990).

¤ Les snRNA U1, U2, U4 et U5 possèdent un domaine 3' terminal A qui présente des caractéristiques communes : il est constitué d'une région simple brin, contenant la séquence PuA(U_{5-6})GPu très conservée dite séquence Sm, flanquée de deux structures tige/boucle (Branlant *et al.*, 1982) (Fig. 14). Ce domaine n'est pas présent dans le snRNA U6. Dès l'observation de ce domaine commun, C. Branlant *et al.* (1982) ont proposé qu'il permettrait de fixer les protéines communes aux différentes UsnRNP spliceosomales, appelées protéines Sm. Cette hypothèse a, ensuite, été vérifiée expérimentalement (Mattaj et De Robertis, 1985 ; Mattaj *et al.*, 1986). Il existe, en effet, 8 protéines Sm communes aux 4 UsnRNP U1, U2, U5 et U4/U6 (Brunel *et al.*, 1981;

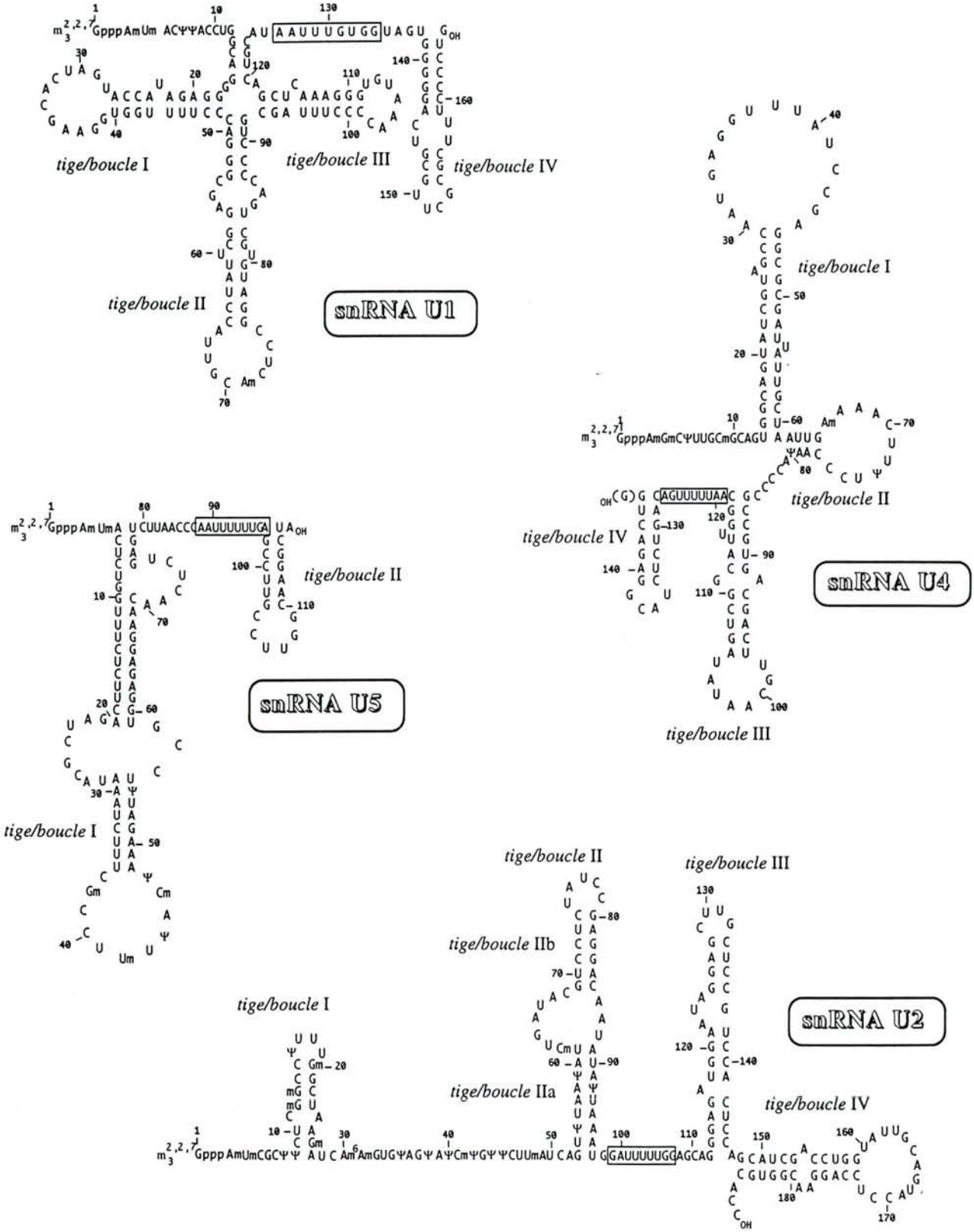


Figure 14 : Structures secondaires des snRNA spliceosomaux U1, U2, U4 et U5.
Les structures représentées sont celles des snRNA de rat établies par :
Branlant *et al.*, 1981, Mount et Steitz, 1981 pour le snRNA U1, Branlant *et al.*, 1982 pour le snRNA U2, Krol *et al.*, 1981a, Myslinski et Branlant, 1991 pour le snRNA U4 et Krol *et al.*, 1981b pour le snRNA U5.
La séquence Sm est encadrée dans les différents snRNA (Branlant *et al.*, 1982).

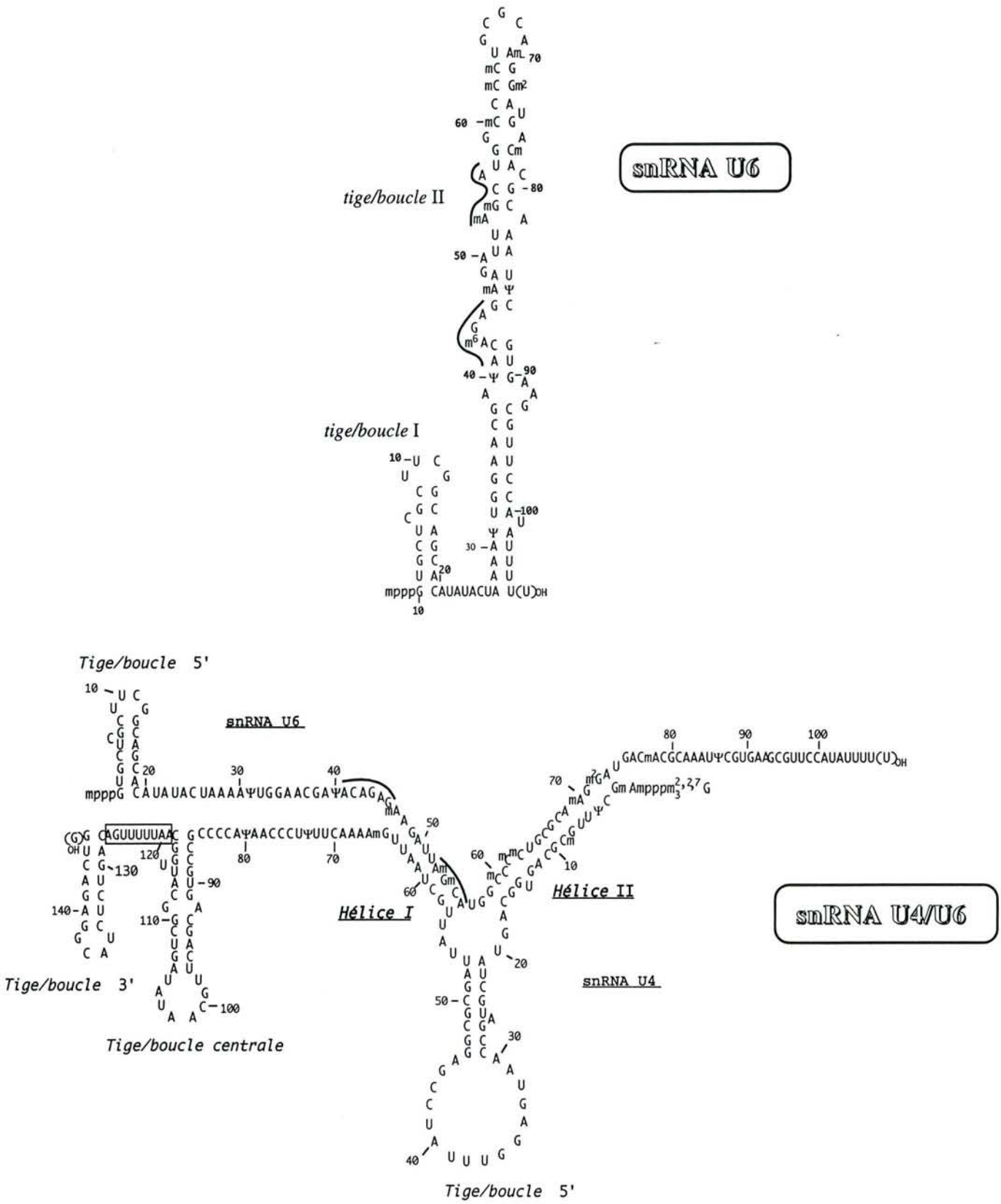


Figure 15 : Structure secondaire du snRNA U6 humain (Mougin *et al.*, résultats non publiés) **et modèle de structure secondaire du duplex d'ARN U4/U6 humain** d'après la structure proposée par Brow et Guthrie (1988) pour *S. cerevisiae*.
Le domaine Sm du snRNA U4 est encadré, les 2 séquences conservées du snRNA U6 sont surlignées.

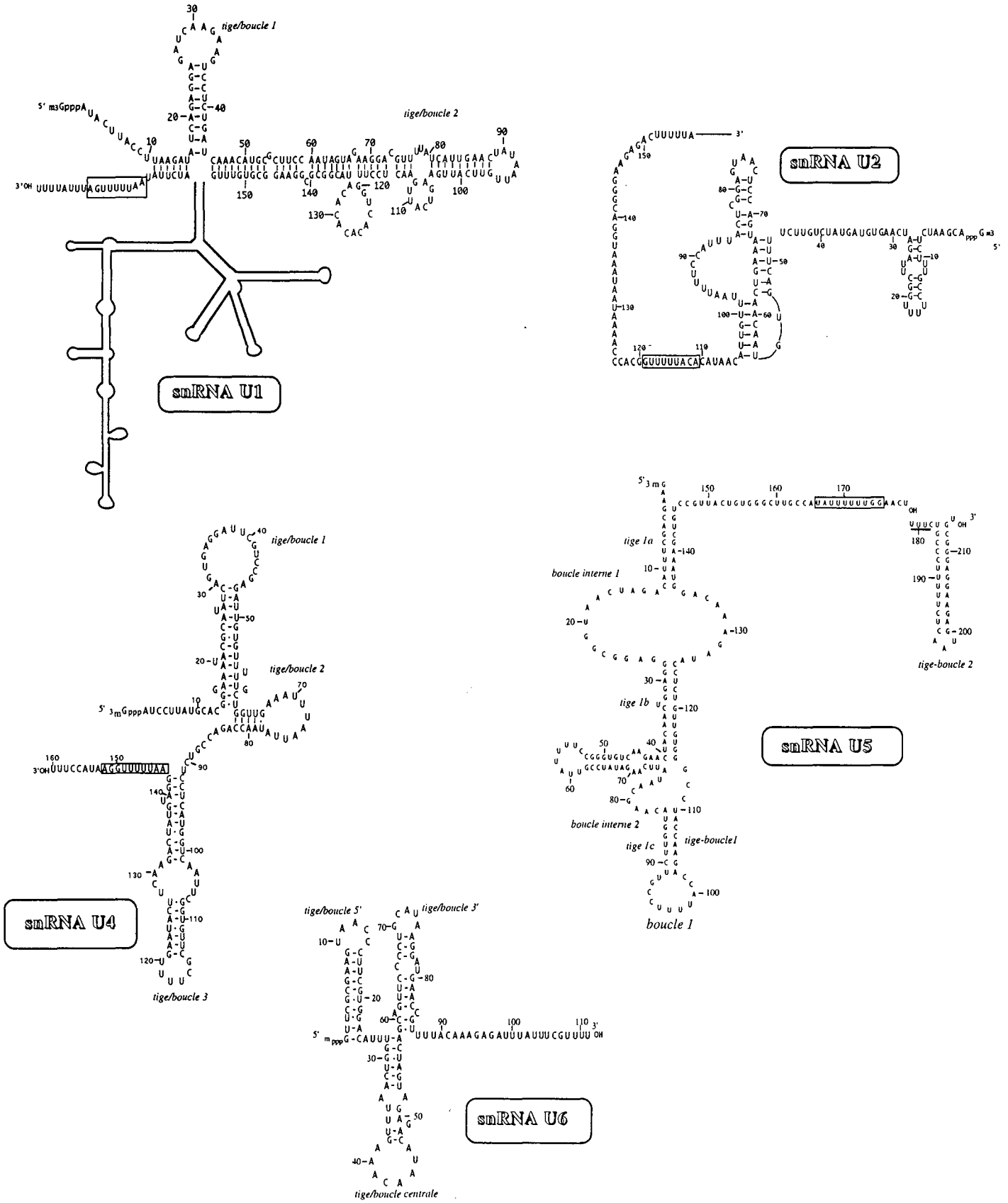


Figure 16 : Structures secondaires des snRNA U1, U2, U4, U5 et U6 de levure *S. cerevisiae*. Ces différentes structures sont extraites de Stutz *et al.* (1993) pour le snRNA U1, Igel et Ares (1988) pour le snRNA U2, Myslinski et Branlant (1991) pour le snRNA U4, Guthrie et Patterson (1988) pour le snRNA U5 et de Fortner *et al.* (1994) pour le snRNA U6. La séquence Sm est encadrée.

Liautard *et al.*, 1982; Lührmann *et al.*, 1990), les snRNA U4 et U6 étant appariés au sein d'une même particule ribonucléoprotéique (Bringmann *et al.*, 1984; Rinke *et al.*, 1985) (Fig. 15).

La comparaison des séquences de UsnRNAs d'un grand nombre d'espèces révèle l'existence de grandes variations en séquences et, pour les levures, en taille. Cependant, ce domaine A a été conservé au cours de l'évolution, ce qui est une indication du rôle important que ce domaine doit jouer.

Comme nous venons de le mentionner, les UsnRNA de l'espèce *S. cerevisiae* se distinguent très nettement de tous les autres UsnRNA caractérisés, y compris de ceux de la levure *Schizosaccharomyces pombe*, par leurs tailles inhabituelles (Fig. 16) : 569, 1174 et 214nt, respectivement, pour les snRNA U1, U2 et U5, contre 165, 188 et 117nt chez les vertébrés. Ils présentent néanmoins les caractéristiques structurales décrites ci-dessus, ainsi que des caractéristiques spécifiques pour chacun des UsnRNA, caractéristiques que nous allons décrire plus loin. Les séquences excédentaires forment des ensembles de structure tige/boucle additionnelles (Fig. 16). Dans le cas du snRNA U2, la délétion de ces séquences excédentaires n'empêche pas l'épissage (Igel et Ares, 1988). Le snRNA U4, lui, a une taille similaire chez *S. cerevisiae* et les vertébrés, bien que cet ARN ait la particularité de ne pas avoir de structure tige/boucle 3' terminale. Cette perte en séquence est compensée par une augmentation de la taille de la tige/boucle en amont de la séquence Sm (Fig. 16) (Siliciano *et al.*, 1987a; Myslinsky et Branlant, 1991). Le snRNA U6 a également une taille similaire chez *S. cerevisiae* et les vertébrés, il est le UsnRNA le plus conservé en séquence (Brow et Guthrie, 1988). Néanmoins, la structure secondaire du snRNA U6 de *S. cerevisiae* en solution est différente de celle de l'ARN de vertébrés (Brow et Vidaver, 1995; Mougin *et al.*, résultats non publiés).

Nous décrirons maintenant les séquences et motifs structuraux spécifiques de chacun des UsnRNA et conservés au cours de l'évolution.

B1.22 Motifs spécifiques à chacun des UsnRNA conservés au cours de l'évolution.

¤ Le snRNA U1.

Dix nucléotides à l'extrémité 5' du snRNA U1 ($5'U_2A_3C_4U_5U_6A_7C_8C_9U_{10}G_{11}3'$) sont strictement conservés au cours de l'évolution (pour revue, Guthrie et Patterson, 1988). L'étude réalisée au labortoire sur le snRNA U1 de *Physarum polycephalum* montre que cet ARN fait exception à la règle puisqu'en position 5, un changement de U en A est observé (Szkukalek *et al.*, 1996). Aussi bien dans l'ARN en solution (Branlant *et al.*, 1980) qu'au sein de la snRNP U1 (Rinke *et al.*, 1984), cette séquence en 5' est toujours en simple brin. Elle est complémentaire de la séquence située à la jonction entre l'exon 1 et l'intron (Lerner *et al.*, 1980; Rogers et Wall, 1980; Mount *et al.*, 1983) et est nécessaire à la réaction d'épissage (Krämer *et al.*, 1984a). Le snRNA U1 a été le premier UsnRNA dont l'implication dans la réaction d'épissage a pu être mise en évidence. En plus du site Sm et de cette séquence conservée en 5', les séquences des boucles terminales I et

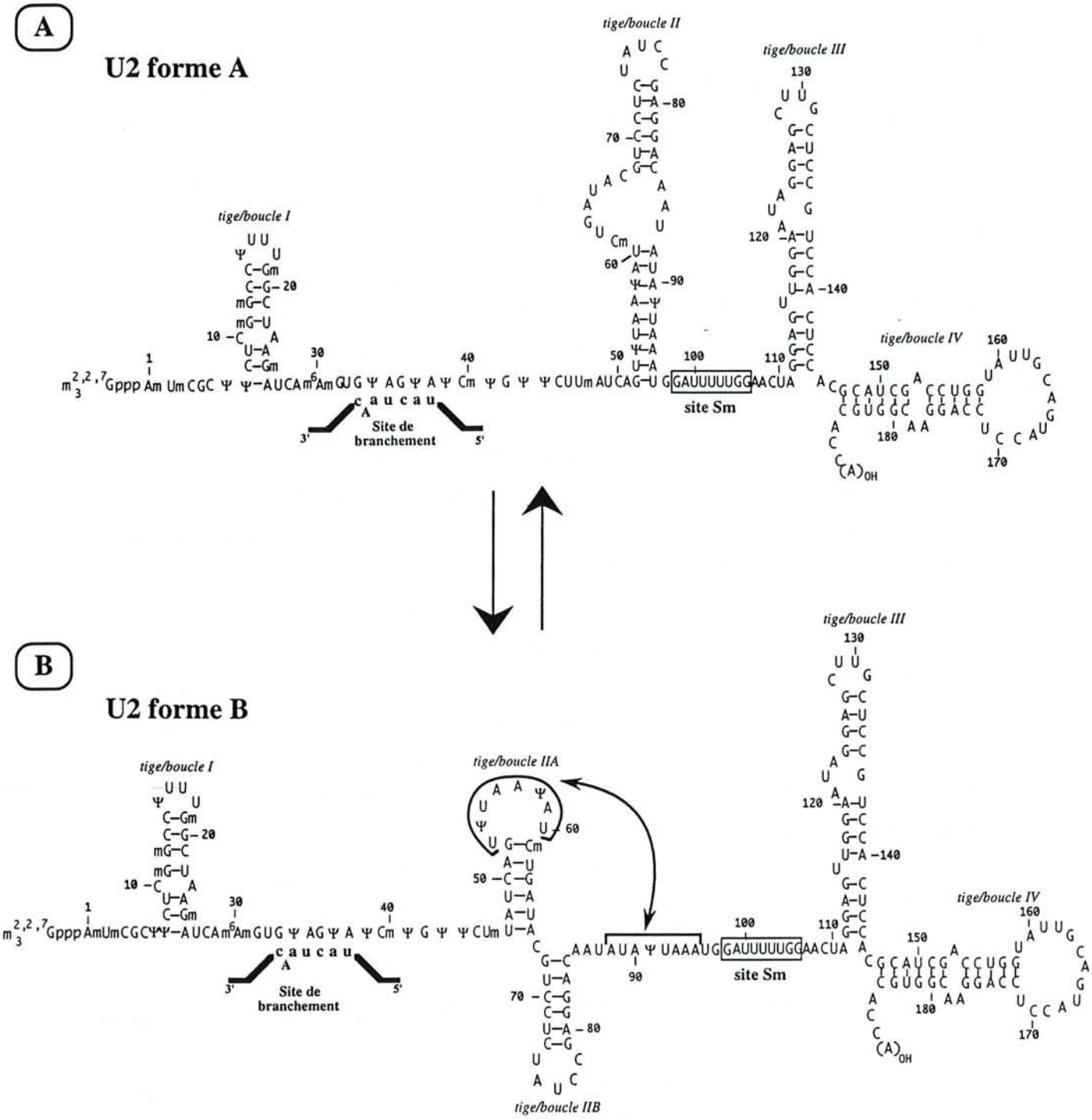


Figure 17 : Les 2 structures alternatives possibles du snRNA U2.
La forme B semble nécessaire pour la formation d'un spliceosome actif. La structure de type pseudo-nœud est représentée par la double flèche. (Branlant *et al.*, 1982; Ares et Igel, 1990; Zavanelli et Ares, 1991). La partie complémentaire de la région renfermant le site de branchement est indiquée (Guthrie et Patterson, 1988).

II, sur lesquelles se fixent, respectivement, les protéines U1-70K et U1-A, ont également été bien conservées au cours de l'évolution (Query *et al.*, 1989a; Scherly *et al.*, 1989) (Fig. 14 et 16).

¤ Le snRNA U2.

Le snRNA U2 a la particularité de pouvoir se replier selon deux structures alternatives. La possibilité de former ces 2 structures est conservée phylogénétiquement (Fig. 17), ce qui laisse supposer qu'elles ont chacune un rôle biologique. Néanmoins, des mutations bloquant le snRNA U2 de levure dans la forme A bloquent l'épissage. La forme B, comportant une structure de type pseudo-nœud (Fig. 17B), doit donc être nécessaire pour l'assemblage d'un spliceosome actif (Ares et Igel, 1990 ; Zavanelli et Ares, 1991). Des études génétiques, réalisées sur des levures *S. cerevisiae*, suggèrent que la tige/boucle IIA serait reconnue par certaines protéines indispensables à l'incorporation de la snRNP U2 dans le spliceosome (Wells et Ares, 1994). La moitié 5' terminale du snRNA U2 (nucléotides 1 à 123) est très conservée phylogénétiquement et porte de nombreuses modifications post-transcriptionnelles, ce qui suggère un rôle fonctionnel important de cette région 5' terminale du snRNA U2. Un hexanucléotide entre les nucléotides U_{32} et U_{39} est strictement conservé au cours de l'évolution. Il est complémentaire de la région renfermant le site de branchement (Guthrie et Patterson, 1988). L'importance de ce domaine phylogénétiquement conservé du snRNA U2 a été démontrée par de nombreuses approches, comme nous le verrons plus loin. En plus du site Sm et de cette région 5' terminale, la boucle terminale de la tige/boucle 3' terminale (tige/boucle IV), qui fixe la protéine B" (Scherly *et al.*, 1990), est aussi une séquence bien conservée au cours de l'évolution (pour revue, Guthrie et Patterson, 1988).

¤ Le snRNA U5.

En plus du site Sm, une seule séquence du snRNA U5 est très conservée : la boucle terminale de la tige/boucle I et l'hélice qui délimite cette boucle (Krol *et al.*, 1981b; Branlant *et al.*, 1983 ; Guthrie et Patterson, 1988) (Fig. 14 et 16). A l'exception d'une différence chez *S. cerevisiae* (Fig. 16), les 9 nucléotides de la boucle terminale I, y compris les nucléotides modifiés, sont strictement conservés (Patterson et Guthrie, 1987; Szkukalek *et al.*, 1995). Cette séquence de 9 nucléotides est accessible dans la snRNP U5 (Black et Pinto, 1989). Comme nous le verrons plus loin, ces 9 nucléotides conservés interagissent, d'après des expériences de pontage, avec les bordures exoniques. Le snRNA U5 présente, d'autre part, la particularité d'exister sous la forme de plusieurs variants différant par quelques substitutions ou délétions, ceci chez toutes les espèces étudiées. Chez l'Homme, une dizaine de variants a été recensée (Branlant *et al.*, 1983; Sontheimer et Steitz, 1992). Chez *S. cerevisiae*, deux formes fonctionnelles de snRNA U5, ayant des tailles différentes, sont transcrites à partir du même gène (Patterson et Guthrie, 1987).

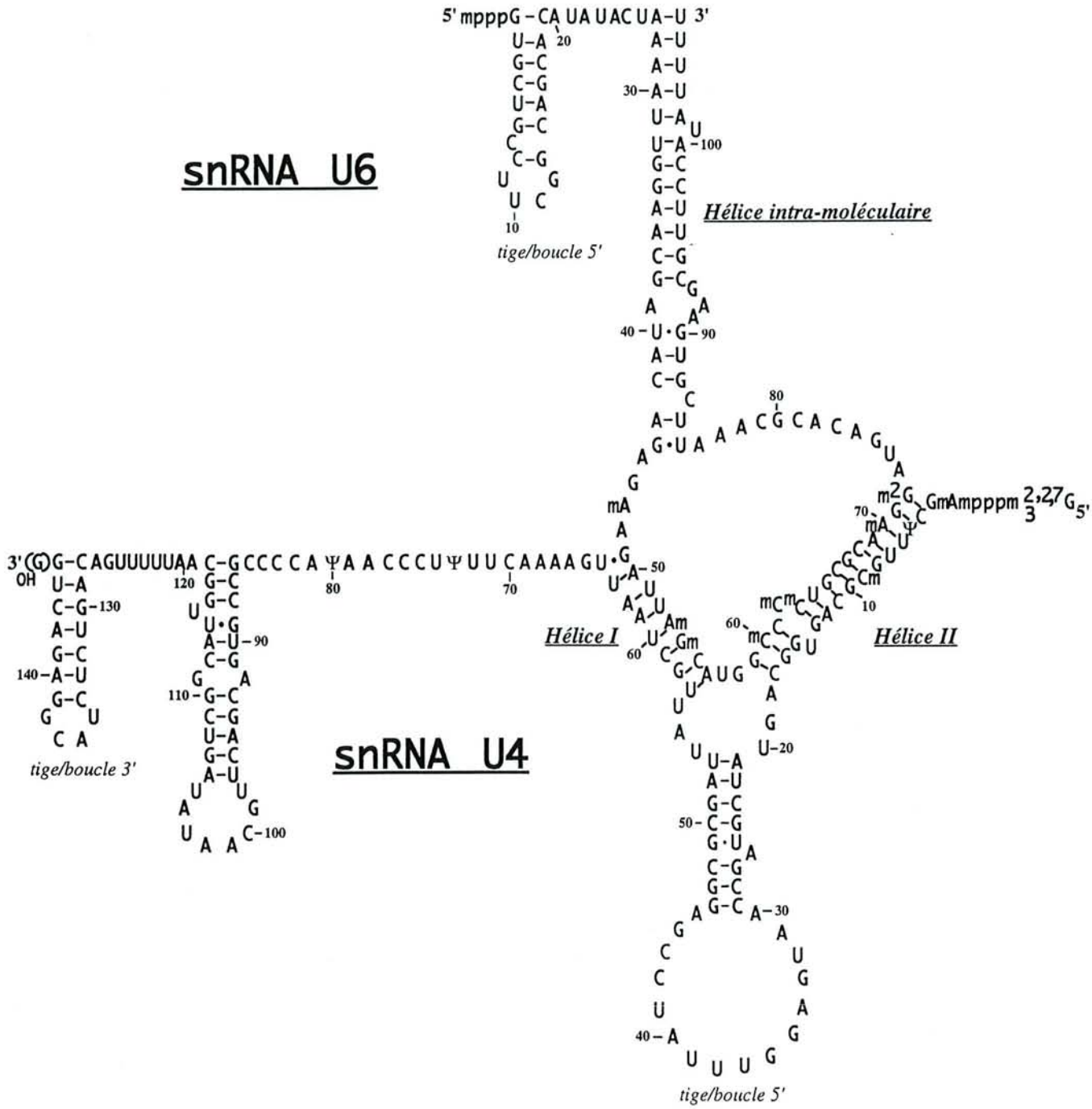


Figure 18 : Structure secondaire du duplex U4/U6 humain lorsque le snRNA U6 se replie sur lui-même (A. Mougin, résultats non publiés).

¤ Les snRNA U4 et U6.

Les snRNA U4 et U6 sont les deux snRNA spliceosomaux les plus conservés en taille (Fig. 14, 15 et 16). Le snRNA U4 présente de nombreuses variations de séquence d'une espèce à l'autre. Par contre, le snRNA U6 est très conservé, c'est le plus conservé des UsnRNA spliceosomaux (Brow et Guthrie, 1988). Son domaine central, en particulier, présente au moins 80% d'identité de séquences quelles que soient les espèces comparées (Guthrie et Patterson, 1988). Là encore, un tel degré de conservation, qui va aussi de pair avec la présence d'un grand nombre de modifications post-transcriptionnelles (Fig. 15), est le signe d'un rôle fonctionnel important de cet ARN.

Ces deux ARN peuvent exister soit sous forme libre, soit appariés au sein d'une même particule ribonucléoprotéique (Bringmann *et al.*, 1984; Rinke *et al.*, 1985).

A la suite de l'observation d'une interaction entre les snRNA U4 et U6 au sein de la snRNP U4/U6, différents modèles ont été proposés pour cette interaction. C'est la détermination des séquences des snRNA U4 et U6 de levure, puis la recherche d'une structure du duplex qui pourrait être formée avec les snRNA de toutes les espèces étudiées, qui ont permis d'aboutir à un modèle validé du point de vue phylogénétique (Brow et Guthrie, 1988) (Fig. 15). Enfin, les analyses directes de structure, analyses réalisées au laboratoire, ont permis d'affiner ces résultats (A. Mougin, résultats non publiés). Ce modèle comporte 2 hélices hétérologues dénommées I et II. A. Mougin a montré, qu'en plus, le snRNA U6 peut se replier sur lui-même au sein du duplex (Fig. 18).

En plus de la séquence Sm, seules 3 séquences du snRNA U4 présentent un certain degré de conservation, la séquence en 5' et les boucles terminales des tiges/boucles I et II qui doivent fixer des protéines (Fig. 14 et 16) (Myslinski et Branlant, 1991). Certaines séquences du snRNA U4 semblent, par ailleurs, importantes pour la formation de la tri-snRNP [U4/U6-U5], forme sous laquelle ces particules sont incorporées dans le spliceosome (Bordonné *et al.*, 1990; Xu *et al.*, 1990b; Mougin *et al.*, résultats non publiés).

Dans le domaine central très conservé en séquence du snRNA U6, deux éléments sont strictement conservés : les séquences ACAGAG, en simple brin en amont de l'hélice I du duplex U4/U6, et AGCA, impliquée dans la formation de cette hélice I (Fig. 15 et 16). Des mutations dans la séquence ACAGAG abolissent les deux étapes de transestérification *in vitro* (Fabrizio et Abelson, 1990a;1990b) et *in vivo* (Madhani *et al.*, 1990). Des mutations dans le snRNA U6, au niveau de cette même hélice I, affectent spécifiquement la deuxième étape de transestérification (Fabrizio et Abelson, 1990a; 1990b). L'effet de ces mutations n'est pas compensé par des mutations compensatoires dans le snRNA U4 (Madhani *et al.*, 1990). Ceci montre que ces séquences du snRNA U6 n'interviennent pas sous la forme du duplex U4/U6. Comme nous le verrons plus loin, ce duplex est dissocié lors de l'assemblage du spliceosome.

Désignation	poids moléculaire apparent	Présence dans les UsnRNP					
		10S	10S	17S	12S	20S	25S
	(M x 10^{-3})	**U1**	**U2**	**U2**	**U4/U6**	**U5**	**U4/U6-U5**
G	9	++	++	++	++	++	++
F	11	+	+	+	+	+	+
E	12	+	+	+	+	+	+
D1	16	+	+	+	+	+	+
D2	16,5	+	+	+	+	+	+
D3	18	+	+	+	+	+	+
B	28	+	+	+	+	+	+
B'	29	+	+	+	+	+	+
C	22	+					
A	34	+					
70K	70	+					
B"	28,5		+	+			
A'	31		+	+			
	35			++			
	50			+			
	60			+			
	66			+			
	110			+			
	120			+			
	150			+			
	160			+			
	15					+	+
	40					+	+
	52					+	+
	100					+	+
	102					+	+
	110					+	+
	116					+	+
	200					+	+
	220					+	+
	15,5						+
	20						+
	27						+
	60						+
	61				+		+
	63				+		+
	90						+

Tableau IV : Composition en protéines des UsnRNP purifiées à partir de cellules HeLa (Lührmann *et al.*, 1990; Behrens et Lührmann, 1991; Gozani *et al.*, 1994).

+: indique la présence de la protéine au sein de la particule.
La présence d'une double bande à la distance de migration de la protéine en gel de polyacrylamide/SDS est indiquée par deux signes +.
Les protéines G, F, E, D1, D2, D3, B et B' sont communes à toutes les particules.

B2. Les snRNP.

Le fractionnement en gradient de saccharose d'un extrait nucléaire a révélé que les snRNA spliceosomaux ne sont pas libres, mais associés à des protéines sous la forme de particules ribonucléoprotéiques, les UsnRNP (Guimont-Ducamp *et al.*, 1977; Gallinaro et Jacob, 1979). Les UsnRNA spliceosomaux constituent 4 snRNP distinctes, les snRNP U1, U2, U5 et U4/U6, puisque, nous l'avons déjà mentionné, les snRNA U4 et U6 sont associés au sein d'une même snRNP.

Les protéines contenues dans ces UsnRNP sont réparties en deux familles :

¤ les protéines communes ou protéines Sm, déjà mentionnées ci-dessus,

et ¤ les protéines spécifiques à chacune des UsnRNP.

Des études poussées ont été réalisées par l'équipe de R. Lürhmann sur les protéines des UsnRNP de cellules HeLa (pour revue, Lürhmann *et al.*, 1990). Les études sur les protéines des UsnRNP de *S. cerevisiae* ont débuté plus récemment, soit par voie génétique, soit par purification des UsnRNP.

B2.1 Les protéines des snRNP de cellules HeLa.

B2.11 Les protéines Sm.

Huit protéines communes aux snRNP U1, U2, U5 et U4/U6 ont été identifiées chez HeLa (Lührmann *et al.*, 1990) : B' (29 kDa), B (28 kDa), D1 (16 kDa), D2 (16,5 kDa), D3 (18 kDa), E (12 kDa), F (11 kDa) et G (9 kDa) (Tableau IV). Les protéines B et B' sont produites à partir d'un gène unique. Ces protéines forment un complexe protéique associé au site Sm. Plus récemment, une protéine commune de 69 kDa, qui semble se fixer de manière transitoire, a été identifiée (Hackl *et al.*, 1994). La fixation des protéines Sm aux snRNA dans le cytoplasme initie l'assemblage des UsnRNP. En effet, après transcription par l'ARN polymérase II, les UsnRNA U1, U2, U4 et U5 sont exportés dans le cytoplasme. Après fixation des protéines Sm, une double méthylation de la coiffe en 5' a lieu. Les pré-UsnRNP ainsi formées sont alors reconnues par des récepteurs nucléaires et réimportées dans le noyau où l'assemblage est complété par fixation des protéines spécifiques (Mattaj et De Robertis, 1985; Mattaj, 1986; Fischer *et al.*, 1993; Plessel *et al.*, 1994; Nelissen *et al.*, 1994).

D'autre part, des expériences de reconstitution des snRNP U2 et U5 *in vitro* démontrent que les protéines Sm sont interchangeables et essentielles à la fonction des snRNP U2 et U5 dans l'épissage (Ségault *et al.*, 1995). Ce système devrait permettre d'obtenir des informations structurales et fonctionnelles sur ces 2 snRNP de mammifère.

B2.12 Les protéines spécifiques à chacune des snRNP.

La composition en protéines des différentes UsnRNP purifiées de cellules HeLa est indiquée dans le Tableau IV. Cette composition, sauf dans le cas de la UsnRNP U1, varie suivant la concentration saline utilisée pour la purification des particules (pour revue, Lührmann *et al.*, 1990).

Quelle que soit la concentration saline utilisée, la snRNP U1 renferme 3 protéines spécifiques 70K, A et C (Hinterberger *et al.*, 1983; Bringmann et Lührmann, 1986). Les structures tridimensionnelles du domaine de la protéine A se fixant à la boucle terminale II du snRNA U1 et du complexe formé entre ce domaine et la boucle terminale II ont été étudiées à haute résolution à la fois par diffraction des Rayons X et par RMN (Nagai *et al.*, 1990; Hoffman *et al.*, 1991). Ce domaine de fixation de la protéine A était le premier domaine de fixation à l'ARN de type RRM mis en évidence (Query *et al.*, 1989b; Kenan *et al.*, 1991). Depuis, la présence de ce domaine a été mise en évidence dans de nombreuses protéines se liant à l'ARN, en particulier, les protéines hnRNP et les protéines SR qui interviennent aussi dans l'épissage et que nous décrirons plus loin. Ce domaine est caractérisé par la présence d'une séquence octamérique (RNP-1: RGFGFVXF) et d'une séquence hexamérique (RNP-2: LFVGNL) qui sont situées dans un ensemble de feuillets ß antiparallèles à la surface du domaine en contact avec l'ARN. Le domaine comporte en plus de cet ensemble de feuillets ß, deux hélices α.

A concentration saline élevée, une particule de 12S renfermant essentiellement 2 protéines spécifiques A' et B" est obtenue dans le cas du snRNA U2, tandis qu'à faible concentration saline, une particule de 17S renfermant 9 à 12 protéines spécifiques supplémentaires est obtenue (Behrens *et al.*, 1993a) (Tableau IV). La protéine U2B", qui présente beaucoup de similitudes avec la protéine U1A de la snRNP U1, possède aussi un domaine de fixation à l'ARN, de type RRM (Scherly *et al.*, 1990).

Pour la snRNP U5 20S, 9 protéines spécifiques sont présentes dans les particules purifiées à faible concentration saline (Bach *et al.*, 1989) (Tableau IV).

Enfin, à faible concentration en ions monovalents, les particules U4/U6 et U5 sont purifiées sous la forme d'un complexe [U4/U6-U5] 25S qui, en plus des protéines contenues dans les snRNP U4/U6 12S et U5 20S, renferme 5 protéines spécifiques (Behrens et Lührmann, 1991) (Tableau IV). C'est sous la forme de ce complexe pré-assemblé que les snRNA U4, U5 et U6 sont associés au pré-spliceosome.

B2.2 Les protéines des snRNP chez *S. cerevisiae*.

B2.21 Les protéines Sm.

Plus récemment, les gènes codant pour les homologues des protéines Sm identifiées chez HeLa ont été recherchés dans les banques de séquences des génomes de *S. cerevisiae*, en utilisant les enchaînements caractéristiques d'acides aminés observés dans toutes les protéines Sm isolées de métazoaires (Séraphin, 1995). Cette approche a permis d'identifier les gènes codant pour

l'ensemble des protéines équivalentes aux protéines Sm de métazoaires. Par ailleurs, elle a permis de mettre en évidence, chez *S. cerevisiae*, une autre série de protéines de type Sm. Le prototype de ces protéines, la protéine Uss1, mise en évidence à la fois par l'équipe de B. Séraphin (Séraphin, 1995) et par celle de J. Beggs (Cooper *et al.*, 1995), se lie au snRNA U6 de *S. cerevisiae*. B. Séraphin a montré que, non seulement cette protéine Uss1, mais l'ensemble de ces protéines de type Sm se lie au snRNA U6 de *S. cerevisiae*. Il reste maintenant à savoir s'il existe l'équivalent de ce groupe de protéines chez les vertébrés et quelle est la fonction de ces protéines.

Donc, sur le plan des protéines Sm, à part cet éventuel ensemble de protéines supplémentaires, il ne semble pas y avoir de différence majeure entre métazoaires et *S. cerevisiae*.

B2.22 Les protéines spécifiques.

Comme nous l'avons mentionné plus haut, les UsnRNA de *S. cerevisiae* se distinguent très fortement des UsnRNA de métazoaires par leur taille. On pouvait donc aussi s'attendre à des différences au niveau des protéines spécifiques de chaque UsnRNP.

Les progrès actuels des connaissances sur les protéines de *S. cerevisiae* impliquées dans l'épissage sont liés au fait que, depuis de nombreuses années, des mutants affectés dans les processus de maturation des ARN ont été isolés. Les gènes mutés de ces protéines, appelées PRP suivi d'un nombre (PRP pour Precursor mRNA Processing), ont été identifiés (pour revue, Green, 1991; Lamm et Lamond, 1993). Certains de ces gènes correspondent à des protéines appartenant aux UsnRNP (Tableau V), d'autres à des protéines associées directement au spliceosome, sans fixation intermédiaire à une snRNP.

Ainsi, les protéines PRP9, PRP11 et PRP21 (ou SPP91) sont contenues dans la snRNP U2 (Chang *et al.*, 1988; Legrain et Choulika, 1990; Arenas et Abelson, 1993; Ruby *et al.*, 1993; Wells et Ares, 1994). PRP9 intervient dans les premières étapes de la réaction d'épissage. *In vitro,* elle est nécessaire à la formation de pré-spliceosomes (Abovich *et al.*, 1990). PRP9 est l'homologue de la protéine de 60 kDa de la snRNP U2 17S purifiée à partir de cellule HeLa (Behrens *et al.*, 1993b; Krämer *et al.*, 1994). L'utilisation du système double hybride a permis de démontrer une interaction directe entre les protéines PRP9 et PRP11 avec PRP21 (Legrain et Chapon, 1993; Legrain *et al.*, 1993). Les protéines PRP11 et PRP21 sont, elles, les homologues des protéines de 66 et 120 kDa de la snRNP U2 17S humaine (Chiara *et al.*, 1994; Krämer *et al.*, 1995). Ces 3 protéines, PRP9, PRP11 et PRP21, comme leurs équivalents dans le système HeLa de 60, 66 et 120 kDa constitueraient le facteur SF3a, un facteur d'épissage intervenant dans la formation du spliceosome (Brosi *et al.*, 1993). De plus, récemment, Tang *et al.* (1996), en utilisant la technique des mutants synthétiques létaux, ont mis en évidence un équivalent levure à l'une des 2 protéines de la snRNP U2 12S (Tableau IV): la protéine U2B" dénommée YU2B".

Seules deux protéines PRP8 et PRP18 appartiennent à la snRNP U5 (Lossky *et al.*, 1987; Horowitz et Abelson, 1993). Des mutations dans PRP8 affectent la formation de la tri-snRNP

Nom de la protéine	MM (kDa)	Motif protéique	snRNP d'appartenance	Références
PRP3	56		snRNP U4/U6	Ruby et Abelson, 1991
PRP4	52	motif des sous-unités β de la protéine G	snRNP U4/U6	Banroques et Abelson, 1989 Bjorn *et al.*, 1989 Bordonné *et al.*, 1990 Dalrymple *et al.*, 1989
PRP6	104	motif Pro-X5-Tryp motif riche en leucine régions riches en Cys et His	snRNP U4/U6 snRNP U4/U6-U5	Legrain et Choulika, 1990 Galisson et Legrain, 1993
PRP8	280		snRNP U5	Jackson *et al.*, 1988 Lossky *et al.*, 1987 Whittaker *et al.*, 1990
PRP9	63	motif riche en leucine motif Cys/His	snRNP U2	Legrain et Choulika, 1990 Abovich *et al.*, 1990 Behrens *et al.*, 1993b
PRP11	30	motif Cys/His	snRNP U2	Chang *et al.*, 1988 Chiara *et al.*, 1994
PRP18	28		snRNP U5	Vijayraghavan et Abelson, 1990 Frank *et al.*, 1992 Horowitz et Abelson, 1993
PRP21	33	motif SURP	snRNP U2	Ruby et Abelson, 1991 Legrain *et al.*, 1993 Chiara *et al.*, 1994 Spikes *et al.*, 1994
PRP24	51	motif RBD	snRNP U6	Shannon et Guthrie, 1991 Strauss et Guthrie, 1991
PRP39			snRNP U1	Lockhart et Rymond, 1994
PRP40	69		snRNP U1	Kao et Siliciano, 1996
$YU_2B"$		motif RBD	snRNP U2	Tang *et al.*, 1996
SNP_1			snRNP U1	Smith et Barell, 1991
MUD_1			snRNP U1	Liao *et al.*, 1993

Tableau V : Protéines impliquées dans la maturation des ARN pré-messagers chez la levure *S. cerevisiae* et appartenant aux UsnRNP.
Les motifs structuraux mis en évidence dans la séquence protéique sont indiqués ainsi que la UsnRNP contenant la protéine (MM : masse moléculaire, RBD : RNA Binding Domain).

[U4/U6-U5] et l'assemblage du spliceosome (Brown et Beggs, 1992). PRP8, comme son homologue humain (une protéine de 220 kDa) (Anderson *et al.*, 1989), semble interagir directement avec l'ARN pré-messager pendant les deux étapes de la réaction d'épissage (Garcia-Blanco *et al.*, 1990; Whittaker et Beggs, 1991; Teigelkamp *et al.*, 1995a). Des études de pontage covalent suggèrent une interaction de PRP8 avec le site 5' d'épissage pendant la première étape de la réaction d'épissage et avec la région du site 3' d'épissage pendant la deuxième étape (Teigelkamp *et al.*, 1995b). PRP8 semble stabiliser les interactions s'établissant entre le snRNA U5 et le pré-ARNm que nous décrirons plus loin.

La protéine PRP24 a une forte affinité pour le snRNA U6 et n'interagirait que très faiblement avec la snRNP U4/U6 (Shannon et Guthrie, 1991). Il est proposé que cette protéine soit impliquée dans la déstabilisation de l'interaction U4/U6 au cours de l'assemblage du spliceosome (Ghetti *et al.*, 1995). Les protéines PRP3 et PRP4 font partie de la snRNP U4/U6 (Banroques et Abelson, 1989; Bordonné *et al.*, 1990; Ruby et Abelson, 1991). PRP 4 est retrouvée dans la tri-snRNP [U4/U6-U5], de même que les protéines PRP8 et PRP18 appartenant à la snRNP U5. La protéine PRP6 s'associe spécifiquement à la tri-snRNP et la stabilise (Galisson et Legrain, 1993).

Enfin, les protéines PRP39 et PRP40 ont récemment été mises en évidence comme 2 constituants de la snRNP U1. Ces 2 protéines ne semblent pas présenter d'homologue chez les métazoaires (Lockhart et Rymond, 1994; Kao et Siliciano, 1996), la snRNP U1 de levure serait donc plus complexe que la snRNP U1 de cellules HeLa. Deux autres protéines de cette UsnRNP ont été caractérisées par voie génétique: la protéine SNP1, présentant 30% d'homologie avec la protéine U1 70K de vertébrés (Smith et Barrell, 1991) et la protéine MUD1, qui serait l'homologue de la protéine U1A (Liao *et al.*, 1993).

Par ailleurs, les études réalisées par P. Fabrizio, dans l'équipe du Pr R. Lürhmann, qui consistent à purifier des UsnRNP de *S. cerevisiae* et à en analyser la composition en protéines, devraient permettre d'avoir très rapidement une analyse complète des protéines (Fabrizio *et al.*, 1994).

D'ores et déjà, ressort de cette étude la présence d'un beaucoup plus grand nombre de protéines dans la snRNP U1 de *S. cerevisiae* (Fabrizio *et al.*, 1994). Il est possible que, parmi les facteurs protéiques qui s'assemblent directement au pré-spliceosome en cours de formation, certains se lieraient aux UsnRNP chez *S. cerevisiae* alors que, chez les vertébrés, ils n'interviendraient que transitoirement avec les UsnRNP. Ceci pourrait être facilité par la plus grande taille des UsnRNA de levure (P. Fabrizio, résultats non publiés).

Comme nous le verrons plus loin, du fait de l'absence, chez *S. cerevisiae*, de certains facteurs protéiques essentiels chez les vertébrés, le processus d'assemblage du spliceosome, bien que procédant par les mêmes étapes (étapes que nous allons décrire maintenant), pourrait reposer sur des mécanismes légèrement différents.

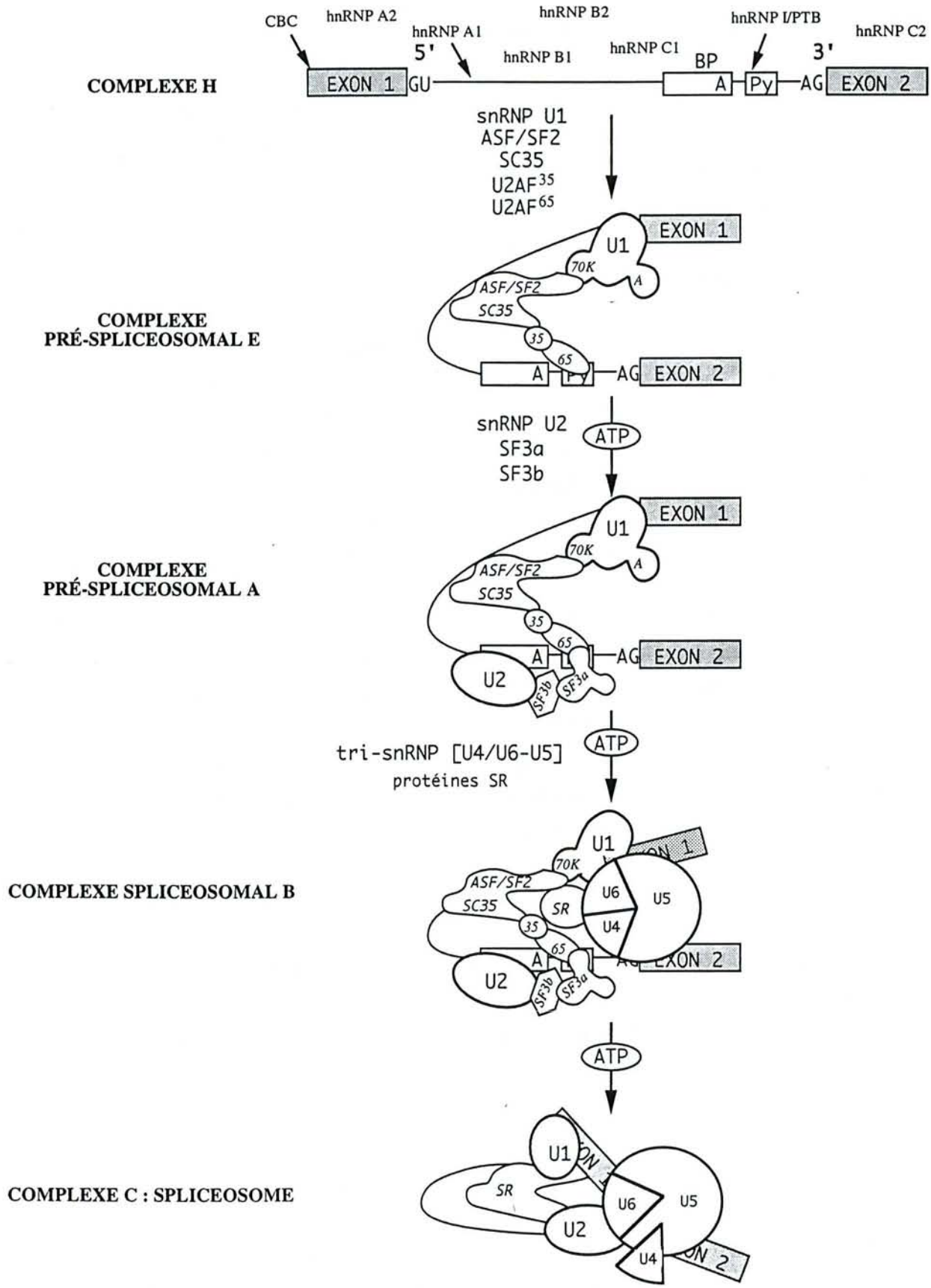


Figure 19 : Les étapes d'assemblage du spliceosome chez les eucaryotes supérieurs.
(Kramer, 1988; Konarska et Sharp, 1987; Bindereif et Green, 1987; Legrain et Chanfreau, 1994).
ASF/SF2 et SC35 sont 2 facteurs protéiques nécessaires à l'interaction de l'assemblage du spliceosome. PTB: "polypyrimidine track binding protein", U2AF: "U2 auxiliary factor" BP:"branchpoint sequence", SF: "splicing factor", SR: "protéine SR", CBC: "Cap Binding Complex".

B3. L'assemblage du spliceosome.

D'après les études réalisées *in vitro* à partir d'extraits nucléaires de cellules HeLa ou de levure, l'assemblage du spliceosome est un processus séquentiel. Les étapes mises en évidence sont identiques pour les deux organismes, même si les mécanismes fins assurant la mise en place des différents intermédiaires et la composition en protéines diffèrent quelque peu entre ces organismes. Les intermédiaires d'assemblage et de réaction ont pu être isolés en grande quantité et étudiés pour le système HeLa. Dans le cas de la levure, les études n'ont pas encore pu être menées à grande échelle. Toutefois, l'approche génétique a, de nouveau, apporté de nombreuses informations. Nous décrirons donc l'état des connaissances pour les deux systèmes.

B3.1 Assemblage du spliceosome chez les mammifères (Fig. 19).

Quatre complexes spliceosomaux (E, A, B et C), correspondant à des intermédiaires d'assemblage et des intermédiaires de la réaction d'épissage, ont été isolés (pour revue, Rymond et Rosbash, 1992; Lamm et Lamond, 1993; Moore *et al.*, 1993). Les complexes E et A sont appelés "pré-spliceosome" et les complexes B et C "spliceosome". Ces différents complexes ont d'abord été mis en évidence par électrophorèse en gel de polyacrylamide non dénaturant (Konarska et Sharp, 1986), puis purifiés par gel filtration suivie d'une colonne d'affinité, le principe étant d'utiliser un ARN pré-messager uniformément biotinylé, ce qui permet de purifier les complexes sur avidine sépharose (Reed, 1990). Ainsi, ont pu être analysés leurs contenus en UsnRNA et en protéines. Les interactions ARN/ARN (pour revues, voir Madhani et Guthrie, 1994a; Umen et Guthrie, 1995a) et ARN/protéines (Staknis et Reed, 1994a; Chiara *et al.*, 1994; Mc Millan *et al.*, 1994) au sein de ces complexes ont par ailleurs été étudiées par des expériences basées sur la formation de pontages covalents.

B3.11 Le complexe H.

Chez les eucaryotes, dès leur transcription, les ARN pré-messagers s'associent à des protéines constituant les particules hnRNP (heterogenous nuclear RiboNucleic Particle) ou complexe H (Gallinaro *et al.*, 1983). Ce complexe renferme au moins 30 protéines nucléaires associées au pré-ARNm (Bennett *et al.*, 1992a). Six d'entre elles, les protéines A1, A2, B1, B2, C1 et C2 sont particulièrement abondantes (Choi et Dreyfuss, 1984). Certaines de ces protéines, en particulier A1, pourraient jouer un rôle dans la réaction d'épissage comme nous le verrons plus loin. En effet, la protéine A1 se lie dans la région polypyrimidines de l'intron et il a été montré que la protéine A1 peut modifier le choix de sites d'épissage dans le cas ou plusieurs sites alternatifs sont en compétition (Mayeda *et al.*, 1993; Yang *et al.*, 1994; Caceres *et al.*, 1994). Certaines de ces

protéines hnRNP sont toutefois moins abondantes dans les complexes d'épissage que dans le complexe H (Bennett *et al.*, 1992a).

La protéine PTB (pour "Polypyrimidine Track Protein") (Patton *et al.*, 1991), identifiée comme une protéine se fixant spécifiquement à la région riche en pyrimidines, s'est avérée être la protéine hnRNP I (Ghetti *et al.*, 1992).

La protéine hnRNP C interagit avec l'extrémité 3' du snRNA U6 (Forne *et al.*, 1995; Mayrand *et al.*, 1996) et pourrait ainsi participer à l'association du snRNA U6 à l'ARN pré-messager.

Le complexe CBC (Cap Binding Complex) reconnait spécifiquement la structure cap des ARN pré-messagers (Izaurralde *et al.*, 1994, 1995). Ce complexe resterait associé à la structure cap pendant toute la réaction d'épissage, et interviendrait dans l'association de la snRNP U1 au site 5' d'épissage proximal (Lewis *et al.*, 1996).

B3.12 Le complexe E.

La fixation de la snRNP U1 et de certains facteurs protéiques conduisent à la formation du complexe E (Fig. 19). C'est le premier complexe de la voie d'épissage. Comment se fait la transition complexe H -> complexe E? Y a-t-il nécessairement formation d'un complexe H s'il s'agit d'épissage ou directement formation d'un complexe E? Aucune donnée n'est encore disponible sur ce point.

Le complexe E purifié par affinité contient les deux sous-unités du facteur U2AF ($U2AF^{65}$ et $U2AF^{35}$), la snRNP U1 et quelques protéines dites SAP pour "Spliceosome Associated Protein" (Bennett *et al.*, 1992b) dont la protéine PTB. L'association de la snRNP U1 se fait, en absence d'ATP, par formation d'un appariement entre l'extrémité 5' du snRNA U1 avec la région du pré-ARNm à la jonction entre exon 1 et extrémité 5' de l'intron (Mount *et al.*, 1983; Black *et al.*, 1985). Un réarrangement au sein de ce complexe permet alors vraisemblablement une interaction de la snRNP U1, et probablement directement du snRNA U1, avec les extrémités 5' et 3' de l'intron (Michaud et Reed, 1993). Différents facteurs protéiques interviennent dans la mise en place de ces interactions de la snRNP U1 : le facteur ASF/SF2 ou le facteur SC35 au niveau de l'extrémité 5' et U2AF au niveau de l'extrémité 3' de l'intron (Michaud et Reed, 1993; Wu et Maniatis, 1993). Une interaction entre ASF/SF2 et U2AF ou entre SC35 et U2AF participe sans doute au repliement de l'ARN pré-messager qui permet de rapprocher les deux extrémités de l'intron qui pourraient ainsi interagir avec la snRNP U1.

D'après des expériences *in vitro*, les facteurs ASF/SF2 et SC35 semblent interchangeables et joueraient un rôle similaire en facilitant le repliement de l'intron (Fu *et al.*, 1992). Dans le cas où plusieurs sites 5' d'épissage alternatif sont présents, le facteur ASF/SF2 favorise l'utilisation du site proximal (Caceres *et al.*, 1994).

B3.13 Le complexe A.

La fixation de la snRNP U2 conduit au complexe A, second intermédiaire d'assemblage qui a pu être isolé (Fig. 19). Dix nouvelles protéines, dont beaucoup sont des composantes de la snRNP U2, sont présentes dans le complexe A (Bennett *et al.*, 1992b; Staknis et Reed, 1994). Un certain nombre de composants de la snRNP U2 peuvent être pontés au pré-ARNm par irradiation aux UV (Staknis et Reed, 1994), ce qui suggère une interaction directe entre ces composants et le pré-ARNm. Le snRNA U2 interagit par appariement de bases avec la boîte de branchement (Zhuang *et al.*, 1989; Zhuang et Weiner, 1989; Wu et Manley, 1989). La formation de ce complexe A nécessite l'hydrolyse d'ATP (Ruskin et Green, 1985b).

B3.14 Le complexe B.

Les snRNP U4/U6 et U5 sont ensuite associées sous forme d'un pré-complexe dont nous avons déjà parlé (la tri-snRNP [U4/U6, U5]), ce qui conduit à la formation du complexe B (Fig. 19). Celui-ci contient donc les snRNP U1, U2, U4, U5 et U6 (Grabowski et Sharp, 1986; Konarska et Sharp, 1986, 1987; Michaud et Reed, 1991; Behrens et Lührmann, 1991). Ce complexe contient, en plus des protéines du complexe A, douze protéines additionnelles correspondant aux protéines de la tri-snRNP [U4/U6-U5] (Bennett *et al.*, 1992b). Une des protéines faisant partie intégrante de la snRNP U5, la protéine de 220 kDa, mentionnée ci-dessus, a pu être pontée de manière covalente par introduction de nucléotides photo-activables, à l'extrémité 3' de l'exon 1. Elle interagit vraisemblablement directement avec cette région de l'ARN dans ce complexe (Wyatt *et al.*, 1992).

Après assemblage de tous les composants nécessaires à l'épissage dans le complexe B, des transitions structurales ont lieu, qui conduisent à la formation du complexe qui porte l'activité catalytique pour la première étape de la réaction d'épissage.

Les transitions structurales conduisant au complexe C sont multiples. L'interaction U4/U6 est rompue, le snRNA U4 devenant ainsi beaucoup plus labile au sein du complexe (Lamond *et al.*, 1988; Michaud et Reed, 1991). De manière concomitante, l'interaction U4/U6 est remplacée par une interaction U2/U6, le snRNA U2 étant toujours associé à l'ARN pré-messager au niveau de la boîte de branchement. Par ailleurs, une interaction va être établie entre le snRNA U6 et l'extrémité 5' de l'intron, ce qui va déstabiliser l'interaction entre le snRNA U1 et l'ARN pré-messager (pour revue, Madhani et Guthrie, 1994a). La snRNP U5 (plus précisément le snRNA U5 et la protéine de 220K) interagit tout d'abord avec l'extrémité 3' du premier exon puis, après la première étape de la réaction d'épissage, avec l'extrémité 5' du deuxième exon d'où un rapprochement spatial des deux extrémités des exons, ce qui facilite la seconde étape de transestérification qui conduit à la ligation des deux exons (Fig. 19) (Wyatt *et al.*, 1992; Sontheimer et Steitz, 1993).

Toutes ces transitions structurales ont été proposées sur la base : 1) de la possibilité de former ces différentes interactions chez toutes les espèces, 2) des résultats d'expériences génétiques montrant la nécessité de former les différentes interactions alternatives possibles, 3) d'expériences de cinétique de formation de pontages covalents au cours d'une réaction d'épissage. Dans l'état

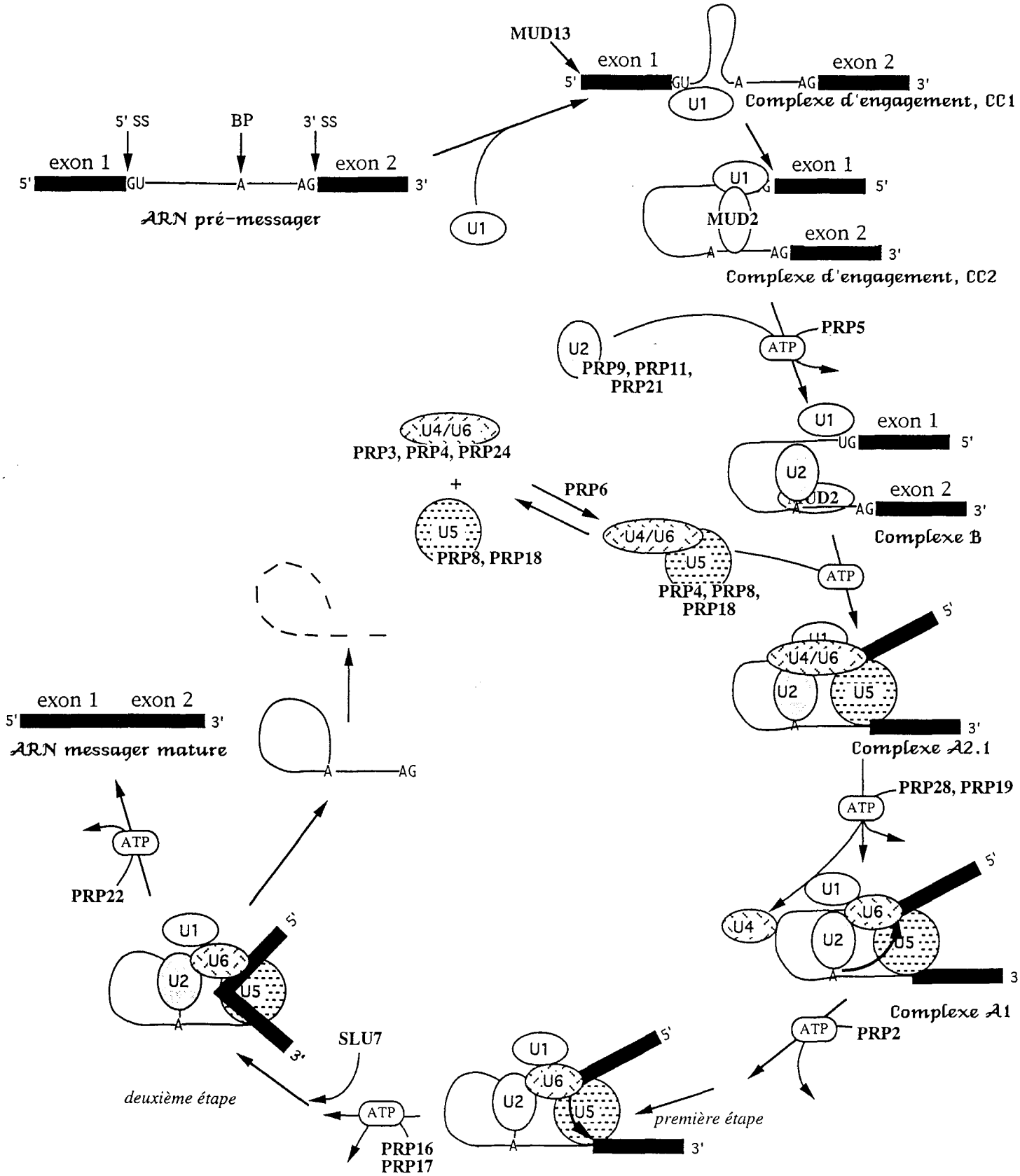


Figure 20 : Représentation schématique du cycle d'assemblage, d'action et de dissociation du spliceosome chez *S. cerevisiae*.

5' ss et 3' ss représentent les sites 5' et 3' d'épissage, respectivement. BP est la boîte de branchement. Les étapes ATP dépendantes sont indiquées par ATP encerclé. Les étapes où interviennent des protéines PRP et l'identité de celles-ci sont indiquées.

actuel des études, le type d'interaction UsnRNA /UsnRNA et UsnRNA /pré-ARNm dans le complexe B purifié selon les techniques de R. Reed et coll., n'est pas encore connu. C'est ce que A. Mougin étudie au laboratoire.

B3.15 Le complexe C.

Un complexe C correspondant à un intermédiaire bloqué après la première réaction d'épissage, peut être purifié en utilisant un ARN pré-messager muté dans la séquence AG 3' terminale de l'intron. Ce complexe C comporte quatorze nouvelles protéines, dites SAP, en plus des protéines déjà présentes dans le complexe B (Gozani *et al.*, 1994). Le complexe C renferme aussi le facteur essentiel d'épissage PSF qui doit donc intervenir dans la deuxième étape de transestérification (Lindsey *et al.*, 1995).

B3.2 Assemblage du spliceosome chez *S. cerevisiae*.

L'utilisation des nombreux mutants PRP permet, en utilisant des extraits de ces levures mutantes à température non permissive, de bloquer le processus d'assemblage du spliceosome et le processus réactionnel à leurs différentes étapes successives. L'analyse des complexes spliceosomaux, bloqués à l'étape de l'assemblage où la protéine PRP intervient, a permis, comme dans le cas des mammifères, de décortiquer ce processus (Fig. 20).

B3.21 Les complexes d'engagement "CC1 et CC2", correspondant au complexe E des vertébrés.

Les complexes d'engagement ("Commitment Complexes") ont été initialement mis en évidence *in vitro* par électrophorèse en gel de polyacrylamide non dénaturant, comme étant les premiers complexes fonctionnels de l'assemblage du spliceosome chez *S. cerevisiae* (Legrain *et al.*, 1988; Séraphin et Rosbash, 1989). La formation de ces complexes ne nécessite ni la présence d'ATP, ni la présence de la snRNP U2, par contre, ils nécessitent l'intégrité des séquences en 5' de l'intron et autour du site de branchement et la présence de la snRNP U1 (Legrain *et al.*, 1988; Séraphin et Rosbash, 1989). Le snRNA U1 interagit tout d'abord avec la séquence en 5' de l'intron, le complexe CC1 est alors formé. La transition entre le complexe CC1 et le complexe CC2 nécessite la présence de la séquence de la boîte de branchement, d'où l'idée d'une interaction supplémentaire dans le complexe CC2 entre le snRNA U1 et la boîte de branchement par l'intermédiaire d'une ou plusieurs protéines. La protéine MUD2, associée à la snRNP U1, pourrait être impliquée dans l'interaction avec la boîte de branchement (Abovich *et al.*, 1994). Cette protéine semble être l'équivalent du facteur $U2AF^{65}$ de vertébrés. Comme nous l'avons donc déjà mentionné, chez

S. cerevisiae, certains facteurs protéiques, tels que l'équivalent de U2AF, se lieraient aux snRNP au lieu de se lier directement à l'ARN pré-messager comme chez les vertébrés.

Une nouvelle protéine non-snRNP, MUD13, vient d'être identifiée dans le complexe CC1. Elle serait l'équivalent, chez *S. cerevisiae* de la protéine CBP20, une des sous-unités du complexe CBC (Cap Binding Complex, Lewis *et al.*, 1996) : elle se fixerait à l'extrémité 5' de l'ARN pré-messager au niveau du cap, et aiderait à la fixation de la snRNP U1 (Colot *et al.*, 1996).

B3.22 Le complexe pré-spliceosomal B, correspondant au complexe appelé A chez les vertébrés.

Comme chez les mammifères, ce complexe est obtenu après la fixation de la snRNP U2, étape nécessitant de l'ATP (Fig. 20).

Plusieurs protéines sont nécessaires à l'addition de la snRNP U2 : la protéine MUD2 (Abovich *et al.*, 1994), les protéines PRP5 (Dalbadie-Mc Farland et Abelson, 1990; Ruby *et al.*, 1993), PRP9 (Abovich *et al.*, 1990; Legrain *et al.*, 1993), PRP11 (Chang *et al.*, 1988) et PRP21 (Arenas et Abelson, 1993; Ruby *et al.*, 1993). Ces protéines PRP9, PRP11 et PRP21 appartiennent à la snRNP U2 (Legrain et Chapon, 1993; Arenas et Abelson, 1993; Behrens *et al.*, 1993b). Elles possèdent toutes les trois un homologue chez les mammifères, homologues qui appartiennent également à la snRNP U2 (Chiara *et al.*, 1994). Le facteur MUD2, déjà présent dans le complexe CC2, permettrait à la snRNP U2 de s'associer au complexe. Une interaction directe entre la protéine MUD2 et la protéine PRP11 a été mise en évidence (Abovich *et al.*, 1994).

B3.23 Les complexes A2-1 et A1 (ils correspondent au complexe appelé B chez les vertébrés).

Ces deux complexes correspondent respectivement à la fixation de la tri-snRNP [U4/U6-U5], puis aux changements conformationnels qui en résultent (Fig. 20). Ces deux étapes sont dépendantes de l'ATP. En plus des protéines associées à la tri-snRNP, plusieurs protéines interviennent dans la formation du complexe A1, dont la protéine PRP28, une protéine de type ARN-hélicase ATP-dépendante. Elle se lie à la tri-snRNP [U4/U6-U5] et pourrait déstabiliser le duplex U4/U6 (Strauss et Guthrie, 1991, 1994). Une autre protéine, PRP19 (Cheng *et al.*, 1993), s'associe au spliceosome lorsque l'interaction avec le snRNA U4 est déstabilisée (Tarn *et al.*, 1994), et pourrait jouer un rôle dans le remaniement qui a lieu à ce moment dans le spliceosome.

D'autres protéines PRP de type ARN-hélicase ATP-dépendante interviennent avant et après les deux étapes de transestérification. L'action de la protéine PRP2 (Cheng et Abelson, 1987; Lin *et al.*, 1987) est nécessaire pour que la première étape de transestérification ait lieu, celle de la protéine PRP16, pour que la deuxième étape ait lieu (Schwer et Guthrie, 1991). Au cours de cette deuxième étape de transestérification, interviennent les protéines PRP16, SLU4/PRP17 et SLU7 (Frank et Guthrie, 1992; Schwer et Guthrie, 1992; Jones *et al.*, 1995; Umen et Guthrie, 1995b). Ces 3

facteurs semblent interagir avec les protéines PRP18 et PRP8 de la snRNP U5. En effet, lors de la deuxième étape de la réaction d'épissage, le site 3' d'épissage est reconnu au moins en 2 étapes :

¤ La première étape, ATP-dépendante, impliquerait le facteur essentiel PRP16 aidé du facteur non-essentiel PRP17. PRP16 interagit directement avec le site 3' d'épissage après formation de l'intermédiaire lasso, hydrolyse de l'ATP, ce qui induit un changement de conformation.

Ce changement de conformation est corrélé au pontage du site 3' aux facteurs essentiels SLU7 et PRP8.

¤ La deuxième étape d'épissage, ATP-indépendante, implique, elle, les facteurs SLU7, PRP8 et PRP18, le facteur PRP18, non-essentiel, aidant sûrement SLU7 à rejoindre le site 3' d'épissage (pour revue, Umen et Guthrie, 1995b).

Quel est exactement le rôle joué par ces différents facteurs ? Interviennent-ils uniquement dans le choix du site 3' d'épissage, assurent-ils la bonne conformation des ARN impliqués dans la réaction ou interviennent-ils directement dans la deuxième réaction de transestérification? Ces 2 étapes semblent, en tous les cas, refléter deux stades de reconnaissance du bon site 3' d'épissage.

Certaines protéines PRP sont impliquées dans la dissociation du spliceosome. Ainsi, la protéine PRP22, une autre protéine de type ARN-hélicase ATP-dépendante est nécessaire pour la libération des UsnRNP après épissage (Company *et al.*, 1991).

La figure 20 représente schématiquement le processus d'assemblage d'action et de dissociation du spliceosome, avec les différentes protéines impliquées à chacune des étapes.

Beaucoup de protéines PRP identifiées qui ne font pas partie intégrante des snRNP présentent donc un motif retrouvé dans des protéines à activité ARN-hélicase ATP-dépendante. C'est le cas des protéines PRP2, PRP5, PRP16, PRP22 et PRP28 (Tableau VI). L'idée est que ces protéines, dont l'action jalonne toutes des étapes du processus, auraient un rôle essentiel dans le contrôle de la fidélité du processus. En effet, comme nous allons le voir plus loin, l'assemblage complexe et l'action du spliceosome reposent sur la reconnaissance d'un nombre très limité de séquences de l'ARN pré-messager. L'action de ces ARN-hélicases ATP-dépendantes serait nécessaire pour que le passage d'une étape à la suivante n'ait lieu que si le complexe d'assemblage ou le complexe réactionnel précédent a été correctement formé. Ces ARN-hélicases auraient un rôle de contrôle, conduisant vers une voie de dégradation tous les complexes non conformes pour le passage à l'étape suivante (Burgess et Guthrie, 1993; pour revue, Umen et Guthrie, 1995a).

Les protéines PRP de type ARN-hélicases ATP-dépendantes présentent des séquences en acides aminés caractéristiques: en plus de la séquence Asp-Glu-Ala-Asp (DEAD) ou Asp-Glu-Ala-His (DEAH), ces protéines comportent les éléments de séquence conservés dans toutes les ATPases à savoir un motif G/AXXXXGKT et DEAD ou DEAH, plus un motif QRXGRXGR (Company *et al.*, 1991). Dans le cas des protéines spliceosomales de levure, une activité ATPase a seulement été démontrée pour les protéines PRP2 et PRP16 (Schwer et Guthrie, 1991; Kim *et al.*, 1992).

Nom de la protéine	MM (kDa)	Motif protéique	Effet de mutations ou délétion	Références
PRP2	100	motif DEAH des ATPases ARN hélicases	1ère étape	Lin *et al.*, 1985; Chen et Li 1990; King et Beggs, 1990 Lee *et al.*, 1986
PRP5	52	motif DEAD des ARN hélicases	assemblage du spliceosome	Dalbadie-Mc Farland et Abelson, 1990; Ruby et Abelson, 1991
PRP6	104	motif Pro-X5-Trp motif riche en Leu régions riches en Cys et His	snRNP U4/U6 tri-snRNP U4/U6-U5	Legrain et Choulika, 1990; Galisson et Legrain, 1993
PRP7			assemblage du spliceosome	Ruby et Abelson, 1991
PRP16	120	Motif DEAH des ATPases ARN hélicases et séquences GX4GK(T/S) et SAT	2ème étape	Couto *et al.*, 1987; Burgess *et al.*, 1990; Schwer et Guthrie, 1991
PRP17	52	motif des sous-unités β de la protéine G	2ème étape	Ruby et Abelson, 1991
PRP19			1ère étape	Ruby et Abelson, 1991
PRP22	130	motif DEAH des ATPases ARN-dépendantes	dissociation du spliceosome et dégradation de l'intron	Company *et al.*, 1991
PRP25			1ère étape	Ruby et Abelson, 1991
PRP26			dissociation du spliceosome et dégradation de l'intron	Ruby et Abelson, 1991
PRP27			dissociation du spliceosome et dégradation de l'intron	Ruby et Abelson, 1991
PRP28	67	motif DEAD des ARN hélicases	1ère étape	Strauss et Guthrie, 1991
PRP29			2ème étape	Ruby et Abelson, 1991
PRP30			1ère étape	Ruby et Abelson, 1991
SPP2			1ère étape	Last *et al.*, 1987; Roy *et al.*, 1995
CUS1			1ère étape	Wells *et al.*, 1996
SLU7	44	motif "Zinc knuckle" ($CX_2CX_4HX_4C$) impliqué dans la fixation à l'ARN des rétrovirus	2ème étape	Frank et Guthrie, 1992; Jones *et al.*, 1995
MUD2	58	motif "RNA Binding Domain"	"Commitment Complex"	Abobich *et al.*, 1994
MUD13	24	motif "RNA Binding Domain"	"Commitment Complex"	Colot *et al.*, 1996

Tableau VI : Facteurs protéiques non-snRNP impliqués dans la maturation des ARN pré-messagers chez la levure *S. cerevisiae*.
Sont indiqués le nom de la protéine, sa masse moléculaire en kDa, les motifs protéiques particuliers observés dans la protéine puis l'étape de l'assemblage, de la réaction d'épissage ou de la dégradation du spliceosome où le processus est bloqué par mutation du gène de la protéine.

Chez les vertébrés, une protéine ATPase ARN-dépendante a récemment été mise en évidence dans la snRNP U5 (la protéine de 200 kDa) (Laggerbauer *et al.*, 1996).

Mais encore aucune protéine n'a été identifiée comme une ARN-hélicase. Ceci vraisemblablement parce que leur interaction avec les complexes spliceosomaux est très transitoire et qu'elles doivent être présentes en faible quantité dans les complexes purifiés.

Si peu de protéines de type ARN-hélicases ATP-dépendantes ont été caractérisées jusque là dans le système mammifère, il existe une classe de protéines pour lesquelles, au contraire, beaucoup de membres ont été isolés chez les vertébrés, ainsi que chez les plantes depuis peu (Lopato *et al.*, 1996), et très peu chez *S. cerevisiae* : les protéines SR. Ces protéines comportent, entre autre, un domaine très riche en enchaînements de séquences Ser-Arg, d'où leur nom. Ce domaine peut s'étendre jusqu'à 120 acides aminés. Des domaines SR sont trouvés dans les facteurs ASF/SF2, U2AF^{65}, U2AF^{35}, SC35 et U1-70K chez les vertébrés (Theissen *et al.*, 1986; Fu et Maniatis, 1990, 1992; Mancebo *et al.*, 1990; Ge *et al.*, 1991; Krainer *et al.*, 1991; Zamore *et al.* 1992; Zhang *et al.*, 1992). De nombreux facteurs de régulation de l'épissage comportent des domaines SR (pour revue, Fu, 1995).

Chez la levure *S. cerevisiae*, jusqu'à présent, seules 4 protéines comportant des similitudes avec les domaines caractéristiques des protéines SR ont été mises en évidence : YCL11c (Birney *et al.*, 1993a), MUD2 (Abovich *et al.*, 1994), Npl3 (Birney *et al.*, 1993b) et NOP3 (Russell et Tollervey, 1995). Il semblerait, en effet, que le nombre de protéines à domaine SR chez *S. cerevisiae* soit beaucoup plus limité. Ces protéines SR peuvent être phosphorylées (Gui *et al.*, 1994a, 1994b; Mermoud *et al.*, 1994). Une "SR protéine kinase" (SRPK1) a été mise en évidence et l'activité de cette protéine influence l'épissage. Récemment, il a été proposé que la topoisomérase I ait une activité "SR protéine kinase" (Gobert *et al.*, 1996). Les domaines SR phosphorylés pourraient faciliter l'assemblage des spliceosomes par des interactions protéine/protéine, comme c'est le cas entre les phospho-tyrosines et les domaines SH2 portés par les différentes protéines impliquées dans la formation du complexe de transduction du signal. Le nombre élevé de protéines SR chez les vertébrés pourrait permettre de palier à la faible stabilité des interactions hétérologues entre UsnRNA et ARN pré-messager chez ces espèces (Kanopka *et al.*, 1996). Les interactions protéine/protéine joueraient un rôle stabilisateur du spliceosome beaucoup plus important chez ces organismes que chez *S. cerevisiae* où la complémentarité entre UsnRNA et séquences introniques est très conservée. Chez les vertébrés, les protéines SR semblent même pouvoir compenser l'absence de certains UsnRNA, comme cela a été démontré pour le snRNA U1 (Crispino *et al.*, 1994; Tarn et Steitz, 1994). En fait, chez les vertébrés, les protéines SR essentielles à la réaction d'épissage sont impliquées dans 3 étapes lors de l'assemblage du spliceosome :

¤ Les protéines SR (dont ASF/SF2) interagissent avec la snRNP U1 pour définir le site 5' d'épissage (Kohtz *et al.*, 1994; Jamison *et al.*, 1995);

¤ Les protéines SR (dont U2AF^{65} et U2AF^{35}) interagissent pour stabiliser le complexe d'assemblage au site 3' d'épissage (Tarn et Steitz, 1995);

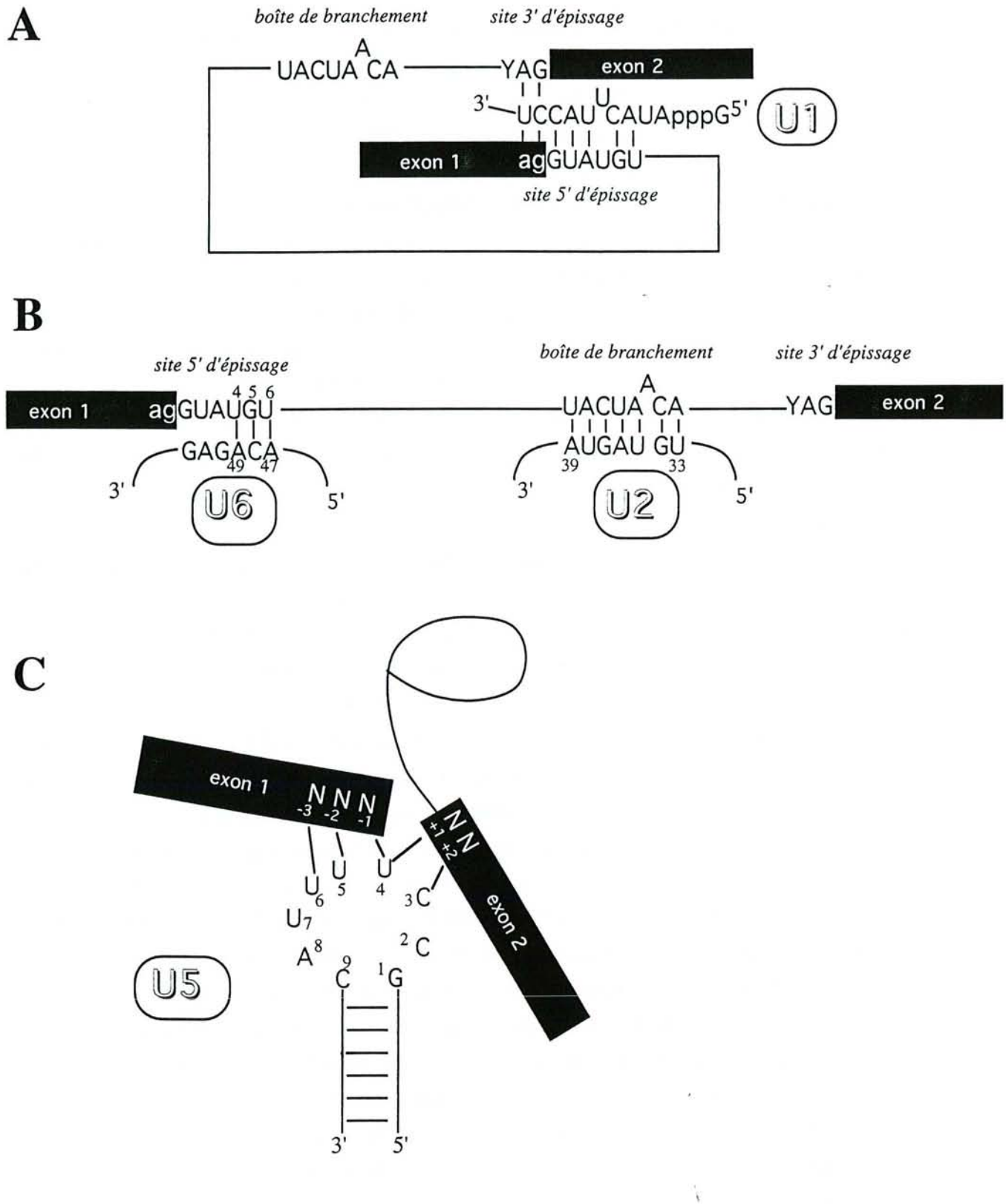


Figure 21 : Interactions snRNA/pre-ARNm s'établissant au cours de la réaction d'épissage chez *S.cerevisiae*. (Madhani et Guthrie, 1994a)
A. interaction U1/intron (est indiqué, en plus de l'interaction U1/boîte 5', l'interaction entre U1 et le site 3' d'épissage),
B. interactions U6/boîte 5' et U2/boîte de branchement,
C. interaction U5/exons, les traits entre la boucle du snRNA U5 et les exons représentent des résultats de pontages ou de mutations compensatoires.

¤ Les protéines SR, toujours par l'intermédiaire du facteur U2AF, interagissent entre elles pour former un pont entre les facteurs assemblés autour des sites 5' et 3' d'épissage (Zhang *et al.*, 1992; Zuo et Maniatis, 1996).

Les protéines SR pourraient également intervenir dans l'interaction des snRNA U2/U6 au site 3' d'épissage (Tarn et Steitz, 1995) et dans l'incorporation de la tri-snRNP [U4/U6-U5] lors de l'assemblage du spliceosome (Roscigno et Garcia-Blanco, 1995), ce qui suggère une intervention des protéines SR dans l'incorporation des snRNP U1, puis U2, et enfin U4/U6-U5 lors de l'assemblage du spliceosome.

En dépit de l'homologie des processus d'assemblage du spliceosome chez les vertébrés et chez *S. cerevisiae*, avec, en particulier, la mise en place de complexes intermédiaires et fonctionnels très similaires, des différences semblent exister dans les mécanismes assurant la mise en place de ces complexes. Les interactions UsnRNA/pré-ARNm semblent, en particulier, jouer un rôle plus important chez *S. cerevisiae* que chez les vertébrés. Nous allons maintenant décrire ces interactions et leur mise en place.

B4. Les interactions ARN-ARN mises en jeu dans l'assemblage et l'activité du spliceosome.

B4.1 Interactions entre l'ARN pré-messager et les UsnRNA.

B4.11 Entre snRNA U1 et pré-ARNm (Fig. 21A).

C'est la première interaction mise en place, elle participe à la reconnaissance des sites d'épissage. La complémentarité observée entre l'extrémité 5' du snRNA U1 et les séquences aux extrémités 5' et 3' des introns, a, dès 1980, amené à proposer une interaction directe entre ces deux ARN (Lerner *et al.*, 1980; Rogers et Wall, 1980). La formation d'un appariement entre l'extrémité 5' du snRNA U1 et la séquence 5' de l'intron fut ensuite démontrée par production de mutations compensatoires entre ces deux ARN (Zhuang et Weiner, 1986; Séraphin *et al.*, 1988; Siliciano et Guthrie; 1988; Séraphin et Rosbash, 1990). L'appariement du snRNA U1 avec l'extrémité 3' de l'intron a été démontrée par la même approche chez *S. pombe* (Reich *et al.*, 1992). Chez *S. cerevisiae*, cet appariement du snRNA U1 avec l'extrémité 3' ne semble pas indispensable (Séraphin et Kandels-Lewis, 1993), par contre, l'appariement en 5' pourrait être plus étendu (Séraphin et Kandels-Lewis, 1993; Nandabalan *et al.*, 1993).

Chez les vertébrés, cette interaction entre le snRNA U1 et le dinucléotide AG en 3' terminal serait nécessaire pour tous les introns pour lesquels ce dinucléotide est important pour la première transestérification, c'est-à-dire une large majorité des introns (Reed, 1989).

Chez les vertébrés, la stabilité de l'appariement entre le snRNA U1 et la boîte 5' est importante lorsque deux sites 5' sont en compétition (Nelson et Green, 1990). Néanmoins, la complémentarité correspondante est loin d'être le seul paramètre définissant la force d'un site donneur d'épissage (Domenjoud *et al.*, 1991). Beaucoup d'autres paramètres, dont les séquences et structures environnant la boîte 5', jouent aussi un rôle important, souvent parce qu'elles fixent des facteurs facilitant l'assemblage du spliceosome.

Alors que, chez les vertébrés, le degré de complémentarité entre snRNA U1 et région 5' est parfois très faible (Jacob et Gallinaro, 1989), chez la levure *S. cerevisiae*, à l'exception du défaut d'appariement en position 5 (Fig. 21A), une conservation absolue des complémentarités aux autres positions est observée. C'est le cas pour les extrémités 5' des introns des gènes du snRNA U3 que nous avons étudiés. Seules quelques rares déviations par rapport au consensus ont été décrites chez les levures (c'est le cas de l'intron du gène rp51b, GTACGT au lieu de GTATGT).

L'effet de mutation dans la boîte 5' complémentaire au snRNA U1 a été testé chez *S. cerevisiae*, et, comme je l'ai déjà cité, des mutations ont parfois des effets dramatiques, notamment celles des nucléotides G_1-->A, G_1-->C et T_2-->G (Newman *et al.*, 1985; Vijayraghavan *et al.*, 1986; Fouser et Friesen, 1986).

Un autre paramètre très important dans le cas de *S.cerevisiae* est l'accessibilité de la séquence en 5' de l'intron. Comme nous l'avons montré en étudiant le pré-snoRNA U3 et comme cela a été aussi montré dans le cas de l'intron de la protéine ribosomique RP51A, l'implication de la boîte 5' dans une structure secondaire de forte stabilité, structure interne au pré-ARN, bloque l'épissage *in vitro* (Goguel *et al.*, 1993; Mougin *et al*, 1996). Les effets observés *in vitro* sont renforcés par rapport aux effets observés *in vivo* (Goguel *et al.*, 1993). Il est possible que des protéines, présentes en quantité limitante dans les extraits nucléaires et ayant une activité ARN-hélicase permettant de rompre les fortes structurations internes du pré-ARN, soient plus actives *in vivo*. Néanmoins, *in vivo*, une limite de stabilité ne peut pas être dépassée non plus. Aussi, des expériences de mutagenèse dirigée ont permis de montrer que lorsque la séquence en 5' de l'intron RP51A est impliquée dans un appariement de plus de 15 pb, l'efficacité d'épissage devient très faible *in vivo* (Goguel *et al.*, 1993).

Chez les vertébrés, la limite à ne pas dépasser pour que l'épissage reste efficace est supérieure à celle observée chez *S. cerevisiae* (Solnick et Lee, 1987). La mise en place du spliceosome semble aller de pair avec un processus de dépliement de l'ARN chez ces organismes.

B4.12 L'interaction snRNA U2/pré-ARNm, la seconde interaction mise en place qui définit le point de branchement (Fig. 21B).

Là aussi, l'interaction postulée sur la base de l'observation d'une complémentarité a été démontrée par génération de mutations compensatoires entre les deux ARN (Parker *et al.*, 1987; Zhuang et Weiner, 1989; Wu et Manley, 1989). Au sein de cette interaction, l'adénosine responsable de la première attaque nucléophile forme une boucle latérale (Fig. 21B). En fait, chez

les eucaryotes supérieurs, le degré de complémentarité entre snRNA U2 et site de branchement est souvent très faible et le type d'interaction parfaite montré dans la figure 21B n'est que très rarement observé, alors que chez la levure, à de très rares exceptions près, un appariement de 6 paires de bases peut être formé entre le snRNA U2 et la boîte de branchement. Comme nous l'avons déjà mentionné, dans les introns des gènes du snRNA U3 de *S. cerevisiae* que nous avons étudiés, seulement 5 pb peuvent être formées du fait de la présence d'une séquence GACTAAC au lieu de TACTAAC au site de branchement.

Par ailleurs, des effets négatifs très marqués de l'implication de la boîte de branchement dans une structure secondaire sont observés chez *S. cerevisiae*. C'est ce que nous avons observé lors de l'étude du pré-snoRNA U3. Parmi les mutants que nous avons produits dans l'objectif de modifier la structure secondaire du pré-ARN, seul le mutant où la boîte de branchement est impliquée dans un appariement de 5 pb n'est pas épissé *in vitro* alors qu'un appariement de même stabilité au niveau de la boîte 5' permet l'épissage *in vitro*. Un fort effet, *in vitro* comme *in vivo*, de l'implication de la boîte de branchement dans une structure secondaire de 15 pb a aussi été observé dans le cas de l'intron du gène de la protéine RP51A (Goguel *et al.*, 1993).

L'effet de la structuration secondaire est plus marqué *in vivo* que l'effet de mutations ponctuelles, même au niveau du résidu A réalisant l'attaque nucléophile. En effet, sa mutation en G dans l'intron permet encore un épissage *in vivo* bien qu'à un taux plus faible (Fouser et Friesen, 1986).

Un tel effet de la structuration de la boîte de branchement n'a jamais été décrit chez les vertébrés.

Donc la formation d'un appariement fort entre le snRNA U2 et la boîte de branchement semble être un paramètre essentiel chez *S. cerevisiae*, ce qui implique une disponibilité de la boîte de branchement au sein de l'ARN pré-messager. Les interactions protéine/protéine jouent sans doute un rôle prédominant chez les vertébrés compensant le très faible appariement existant entre le snRNA U2 et la boîte de branchement.

B4.13 Les échanges d'interaction U2+U4/U6 <=> U6/U2+U4 (Fig. 22).

Comme nous l'avons déjà mentionné ci-dessus, les snRNA U4 et U6 sont associés au spliceosome sous la forme d'un duplex d'ARN présentant deux hélices hétérologues I et II et vraisemblablement un repliement interne du snRNA U6 sur lui-même (Fig. 22A) (Brow et Vidaver, 1995; Mougin *et al.*, résultats non publiés). La protéine hnRNP C, qui a une affinité pour la région 3' du snRNA U6 (Mayrand *et al.*, 1996) pourrait participer à l'ouverture de l'interaction interne du snRNA U6 (Forne *et al.*, 1995). L'existence de la formation d'un appariement de bases entre l'extrémité 3' du snRNA U6 et 5' du snRNA U2 (hélice 2) a été proposée par Hausner *et al.* (1990). La possibilité de former cet appariement est indispensable pour qu'il y ait épissage *in vitro* avec un extrait nucléaire de cellules HeLa (Datta et Weiner, 1991; Wu et Manley, 1991). Chez *S. cerevisiae*, le même type d'interaction peut être formé d'après la conservation de la complémentarité des séquences par mutations compensatoires, mais cet appariement n'est pas

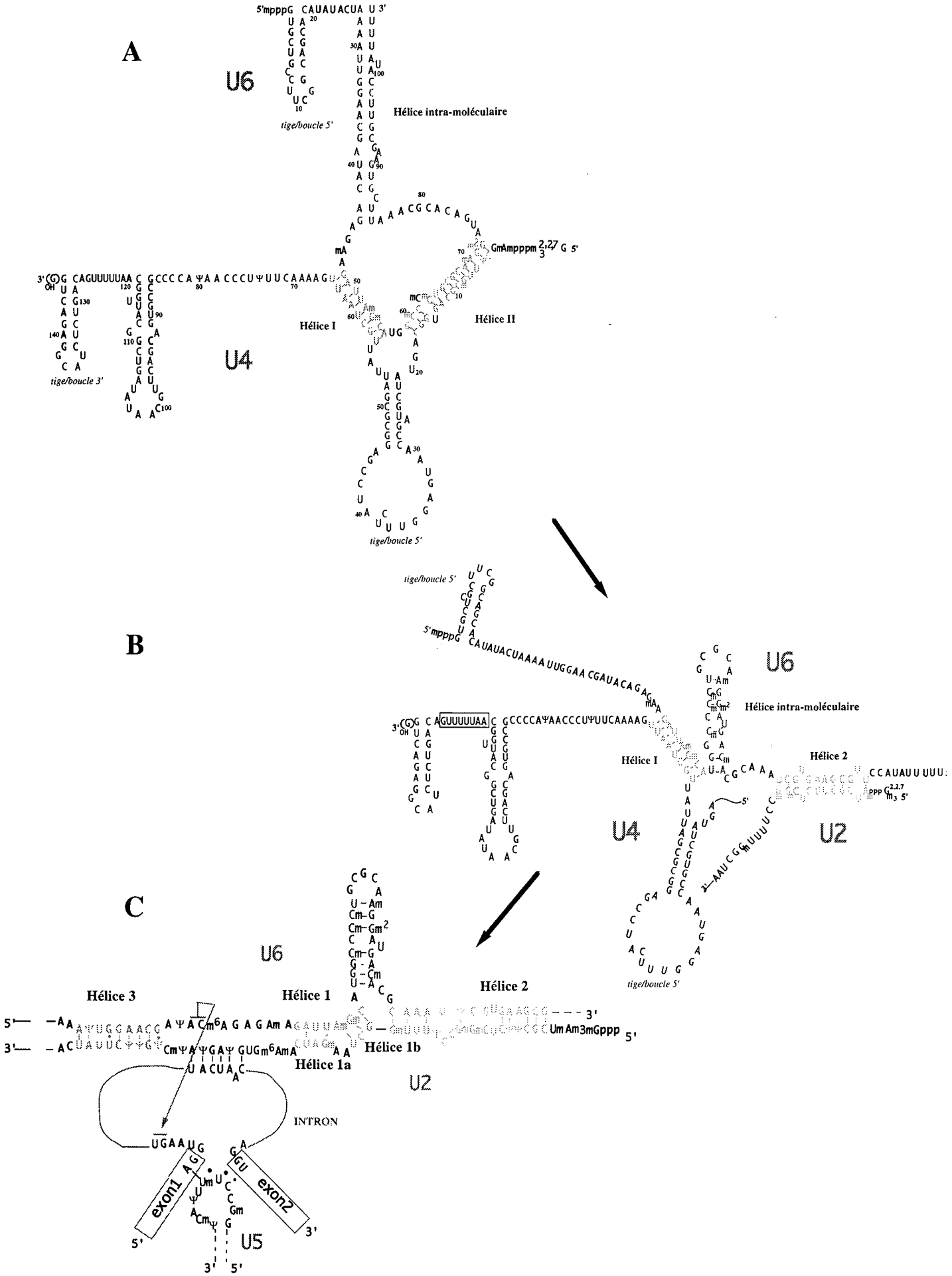


Figure 22 : Mécanisme proposé pour le passage de l'interaction U4/U6 à l'interaction U2/U6.

A. Interaction des snRNA U4 et U6 au sein de la tri-snRNP [U4/U6-U5].
B. Dissociation de l'interaction U4/U6 au profit d'une interaction U6/U2 au sein du spliceosome B.
C. Interactions ARN/ARN au coeur du spliceosome.

indispensable à la croissance cellulaire (Fabrizio *et al.*, 1989; Madhani *et al.*, 1990) et donc à l'épissage.

Une autre interaction correspondant à la formation de deux courtes hélices contiguës 1a et 1b interrompues par deux nucléotides en boucle latérale a été mise en évidence chez *S. cerevisiae* par Madhani et Guthrie (1992). Des mutations dans ces hélices 1a et 1b, qui sont très fortement conservées phylogénétiquement (Fig. 22C), affectent soit la première, soit le seconde étape de la réaction d'épissage (Fabrizio et Abelson, 1990a; 1990b; Wolff *et al.*, 1994; Yu *et al.*, 1995). La formation de l'hélice 1a est aussi nécessaire pour la formation d'un spliceosome actif dans un extrait nucléaire de cellules HeLa, la nécessité de l'hélice 1b n'a pas, elle, été démontrée (Sun et Manley, 1995).

Enfin, la possibilité de former une troisième interaction hétérologue, par formation d'une hélice 3 entre la séquence du snRNA U2 en aval de la séquence appariée à la boîte de branchement et un segment du snRNA U6 recouvrant en partie la séquence ACAGAG, a été mise en évidence plus récemment (Sun et Manley, 1995) (Fig. 22C). Son importance pour la formation du spliceosome actif a été démontrée dans des expériences *in vitro* avec un extrait de cellules HeLa. L'importance d'une telle interaction n'a pas été démontrée chez *S. cerevisiae*. Il est vraisemblable que, dans le spliceosome, seule une partie de cette hélice 3 soit formée puisque, dans sa totalité, elle implique une partie de la séquence du snRNA U2 s'appariant au site de branchement (Fig. 22C). Les études réalisées par A. Mougin au laboratoire, sur les snRNP U4/U6, tri-snRNP [U4/U6-U5], duplex d'ARN U4/U6 et snRNA U6 isolé d'extraits nucléaires de cellules HeLa ou de *S. cerevisiae*, apportent une explication possible aux différences que nous venons de rapporter entre vertébrés et *S. cerevisiae*.

En effet, A. Mougin a montré que le snRNA U6 se replie sur lui-même en une structure très stable qui peut être initiée par le repliement interne du snRNA U6 qui peut se former au sein du duplex d'ARN U4/U6 (Fig. 22A). La structuration interne du snRNA U6 de *S. cerevisiae* est différente (Brow et Vidaver, 1995) et n'est pas initiée par le repliement interne du snRNA U6 au sein du duplex d'ARN U4/U6. L'idée est que les transitions U2+U4/U6 <=> U6/U2+U4 chez *S. cerevisiae* ne doivent jamais se produire avec libération de snRNA U6 libre comme dans le cas des cellules HeLa, pour des raisons thermodynamiques. La transition U4/U6-->U2/U6 doit être initiée par formation de l'hélice hétérologue 2 entre les snRNA U2 et U6 qui évite le repliement interne du snRNA U6, avant que les hélices hétérologues I et II entre les snRNA U4 et U6 ne soient dissociées (Fig. 22B). L'hélice hétérologue 3 entre les snRNA U2 et U6 dans sa configuration partielle compatible avec l'appariement au site de branchement, pourrait peut-être aussi être mise en place en même temps, puisqu'elle est compatible avec les hélices hétérologues I et II des snRNA U4 et U6.

De la même manière, lors de la transition inverse U2/U6 vers U4/U6, les hélices hétérologues 1 et 3 entre les snRNA U2 et U6 pourraient être maintenues lors de la reformation des hélices

hétérologues I et II entre les snRNA U4 et U6, après rupture de l'hélice 1 entre les snRNA U2 et U6. Ainsi, le snRNA U6 ne serait jamais libéré sous forme libre (Fig. 22).

Cette nécessité d'éviter la formation d'un repliement interne du snRNA U6 est, sans doute, moins aiguë chez *S. cerevisiae* puisque le repliement possible dans le duplex U4/U6 ne conduit pas à un repliement interne stable du snRNA U6, d'où vraisemblablement, l'exigence moins importante du système levure vis-à-vis de la formation des hélices hétérologues 2 et 3 entre les snRNA U2 et U6.

Cet exemple montre, une fois de plus, que, tout en procédant par les mêmes types d'intermédiaires, les mécanismes de mise en place de ces intermédiaires diffèrent entre les systèmes HeLa et *S. cerevisiae*.

D'après l'importance de la séquence des hélices hétérologues 1a et 1b entre les snRNA U2 et U6, ces hélices doivent appartenir au site catalytique de même que la structure tige/boucle du snRNA U6 située entre les hélices hétérologues 1b et 2. De nombreuses mutations dans cette structure tige/boucle affectent fortement l'épissage (Wolff et Bindereif, 1993; Fortner *et al.*, 1994).

Est-ce que l'hélice hétérologue 3 n'est formée que transitoirement, est-elle présente dans les complexes actifs ? Les expériences qu'A. Mougin réalise actuellement au laboratoire sur les complexes purifiés B et C de cellules HeLa, devraient permettre d'apporter une réponse à ces questions.

B4.14 Entre snRNA U6 et pré-ARNm.

La transition U4/U6 vers U2/U6 s'accompagne vraisemblablement de la génération d'une interaction entre le snRNA U6 et l'extrémité 5' de l'intron (Fig. 21B). En effet, l'importance d'une interaction de trois paires de bases formées entre le trinucléotide ACA en 5' de la séquence conservée ACAGAG du snRNA U6 et les résidus aux positions 4, 5 et 6 en 5' de l'intron a pu être montrée à la fois chez HeLa (Wassarman et Steitz, 1992) et chez *S. cerevisiae* par des techniques de pontage (Sawa et Abelson, 1992). Les approches génétiques ont permis de montrer le caractère essentiel de cette interaction (Lesser et Guthrie, 1993; Kandel-Lewis et Séraphin, 1993).

Cette interaction est exclusive vis-à-vis de l'interaction entre le snRNA U1 et l'extrémité 5' de l'intron (Fig. 21) et vis-à-vis de la partie de l'hélice hétérologue 3 qui n'est pas non plus compatible, comme nous l'avons déjà mentionné, avec l'appariement entre le snRNA U2 et la boîte de branchement (Fig. 22C). Il est donc vraisemblable que cette hélice 3 ne se forme que partiellement dans le spliceosome actif. Par ailleurs, la mise en place de l'interaction entre le snRNA U6 et le pré-ARNm doit donc avoir lieu après déstabilisation de l'interaction entre le snRNA U1 et l'extrémité 5' de l'intron. Des travaux anciens de l'équipe de P. Sharp montraient, en effet, que le snRNA U1 n'était pas associé de manière stable aux complexes spliceosomaux réalisant les deux étapes de la réaction catalytique (Konarska et Sharp, 1987).

La snRNP U1 jouerait donc un rôle dans la reconnaissance des sites 5' et 3', et l'initiation de l'assemblage du spliceosome, mais n'interviendrait pas dans la fonction de catalyse.

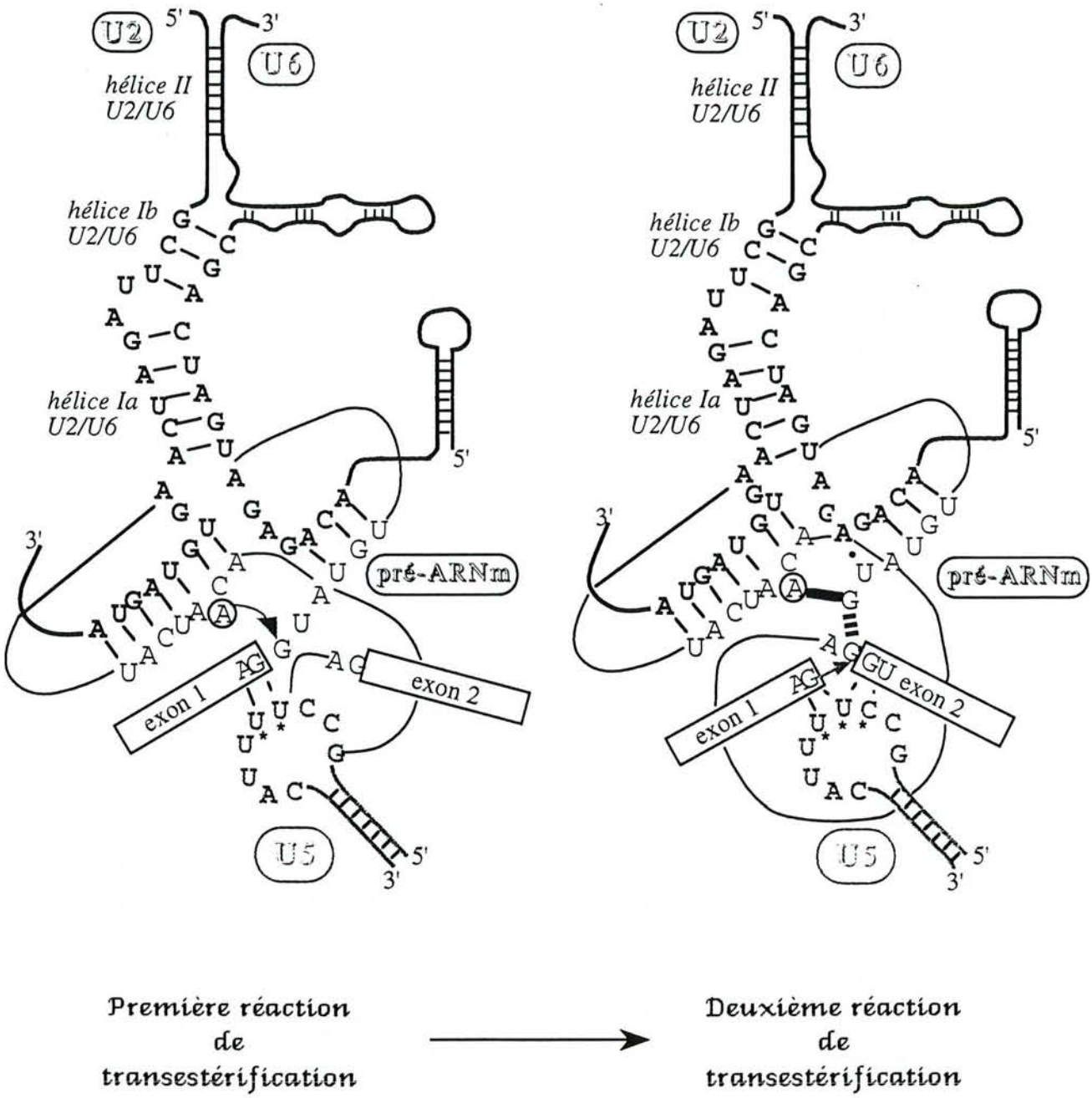


Figure 23 : Les interactions hétérologues entre ARN supposées exister au sein des complexes spliceosomaux assurant la première et la deuxième étapes de la réaction d'épissage.

Le pré-ARN messager est indiqué en noir, les ARN U2, U5 et U6 en bleu, vert et rouge, respectivement. Les astérisques indiquent les positions où la formation de liaison covalente a été mise en évidence (Sontheimer et Steitz, 1993).

Le trait épais noir indique la liaison covalente 2'-5' s'établissant entre le premier nucléotide de l'intron et le site de branchement au cours de la première réaction de transestérification. Le trait épais en pointillé entre les deux résidus G aux extrémités de l'intron indique l'interaction de type Hoogsteen proposée entre ces deux résidus (Parker et Siliciano, 1993; Chanfreau *et al.*, 1994; Deirdre *et al.*, 1995).

Ce schéma a été adapté à partir de celui de Legrain et Chanfreau (1994).

B4.15 Entre snRNA U5 et pré-ARNm.

L'ensemble des interactions entre le snRNA U5 et le pré-ARNm met en jeu la boucle de la tige/boucle I du snRNA U5 très conservée phylogénétiquement. Contrairement au reste de la molécule U5, des mutations dans cette boucle altèrent la croissance chez les levures (Frank *et al.*, 1994). Différentes approches ont permis de montrer que cette boucle terminale interagit avec les extrémités des deux exons 5' et 3' (Fig. 21C, 22C).

La mutation G->A en position 1 de l'intron inhibe la réaction d'épissage chez *S. cerevisiae* (Newman *et al.*, 1985; Fouser et Friesen, 1986; Vijayraghavan *et al.*, 1986). Mais une simple substitution dans la boucle conservée du snRNA U5 s'est révélée partiellement suppressive du phénotype, à savoir qu'un site cryptique est utilisé. Le site de coupure s'est avéré être défini par l'existence d'une complémentarité entre la boucle terminale I du snRNA U5 et la bordure exonique située en 5' de l'intron (Newman et Norman, 1991; 1992). Ce même type de conclusion a été obtenu par mutation du AG 3' terminal : des révertants mutés dans la boucle du snRNA U5 ont été obtenus. Les sites cryptiques 3' utilisés sont définis par la complémentarité de la bordure exonique située en 3' de l'intron avec la boucle terminale du snRNA U5 (Newman et Norman, 1992). Puis, des confirmations de l'existence de ces interactions ont été obtenues par des expériences de pontage basées sur l'incorporation de nucléotides photo-activables dans les séquences exoniques en bordure de l'intron. La réalisation de cinétiques de pontage a permis de montrer, qu'avant la première étape d'épissage, les nucléotides 4 et 5 de la boucle terminale du snRNA U5 interagissent avec l'extrémité 5' de l'exon 1 puis, après la première étape de la réaction, une interaction a lieu avec les extrémités des deux exons (Fig. 23) (Sontheimer et Steitz, 1993).

Toutefois, il est à noter que si la boucle I du snRNA U5 est très conservée phylogénétiquement, il n'en est pas de même pour les bornes exoniques du pré-ARNm. De ce fait, l'interaction possible avec le snRNA U5 est souvent de très faible énergie. La présence de résidus U dans cette boucle augmente les possibilités d'interactions avec la faculté de former des paires G·U ou A-U. Par ailleurs, la présence de Ψ et de résidus 2'-O-méthylés dans la boucle pourrait rigidifier la structure dans une conformation favorisant les interactions avec l'ARN pré-messager. Néanmoins, vu la faible énergie des interactions possibles, aussi bien chez les vertébrés que chez *S. cerevisiae*, une stabilisation de ces interactions par des protéines doit être indispensable. La protéine PRP8, chez *S. cerevisiae,* comme sûrement son homologue chez les vertébrés (la protéine de 220 kDa), qui appartient à la snRNP U5 et se lie à l'ARN pré-messager, doit intervenir dans cette stabilisation (Teigelkamp *et al.*, 1995b). Ces expériences étaient les premières démonstrations d'une interaction possible du snRNA U5 avec les deux bordures exoniques au cours de l'épissage.

Etant donné que le snRNA U5 interagit avec les deux bordures exoniques, on pouvait s'attendre à un effet négatif marqué de l'implication de ces deux bordures exoniques dans une structuration secondaire interne à l'ARN précurseur. D'après les études réalisées chez les plantes, l'effet d'une structuration en 3' semble moins marqué que l'effet d'une structuration en 5' (Liu *et al.*, 1995). Néanmoins, chez les levures, là encore, un cas d'occlusion d'un site 3' par son implication dans

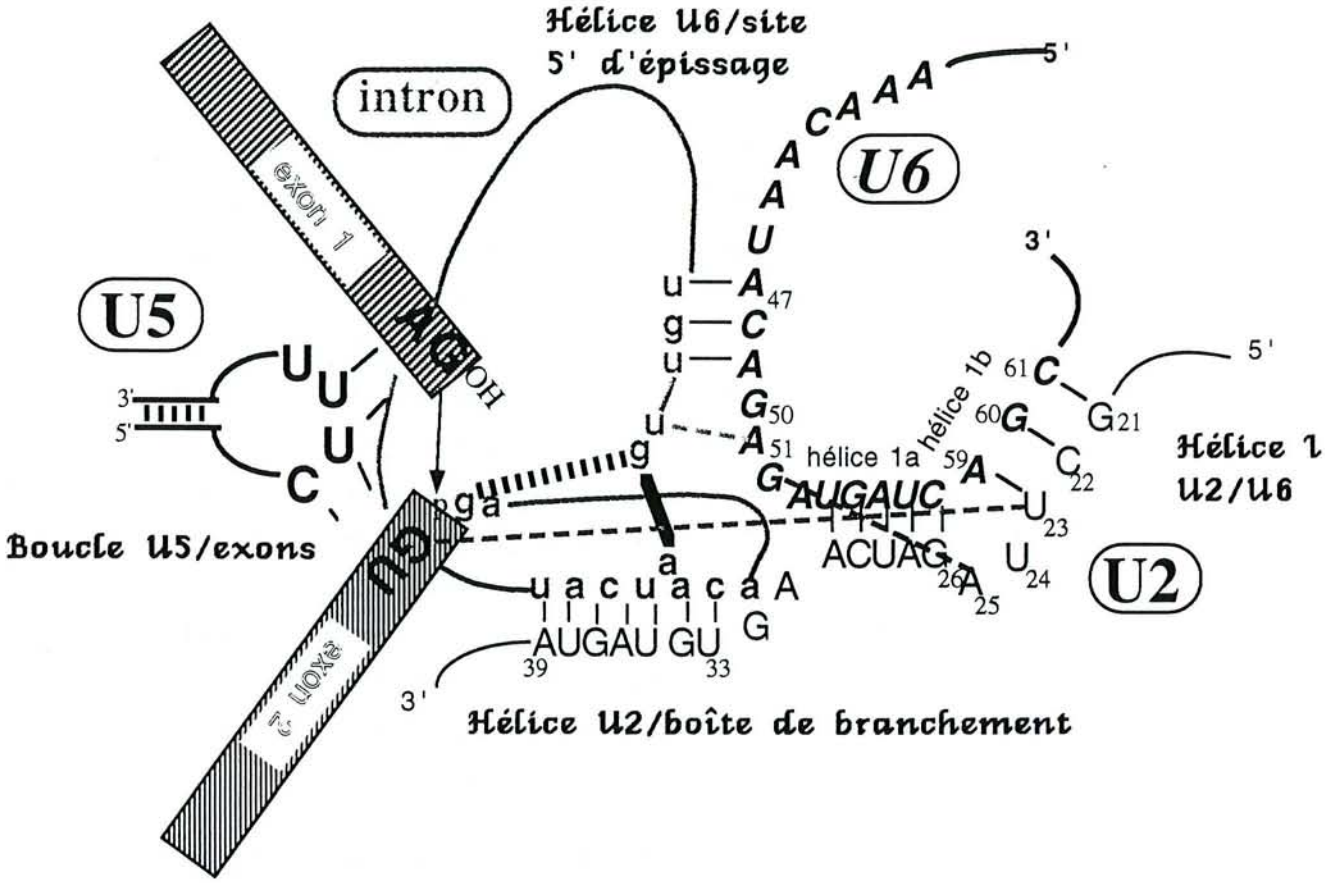


Figure 24 : Réseau d'interactions entre ARN lors de la deuxième étape de la réaction d'épissage.

La numérotation des séquences est celle utilisée chez *S. cerevisiae*. Les exons sont représentés sous forme de rectangles gris hachurés, l'intron est coloré en bleu.
Le trait épais noir représente la liaison 2'-5' A-G entre le point de branchement et l'extrémité 5' libre de l'intron (Query *et al.*, 1994).
Le trait épais pointillé représente l'interaction non canonique de type Hoogsteen mise en évidence entre les deux résidus G aux extrémités de l'intron (Parker et Siliciano, 1993; Chanfreau *et al.*, 1994; Deirdre *et al.*, 1995).
L'interaction tertiaire entre les résidus A25 et G52 des snRNA U2 et U6 est représentée en pointillé rouge.
Les pontages entre le résidu A51 du snRNA U6 et le résidu +2 de l'intron (Sontheimer et Steitz, 1993) et entre le résidu U23 et le résidu +1 de l'exon 2 (A. Newman, résultats non publiés) sont représentés en pointillé vert.
Ce schéma est adapté de Madhani et Guthrie (1994a) et de Umen et Guthrie 1995a).

une structure secondaire a été mis en évidence chez *Kluyveromyces lactis* (Deshler et Rossi, 1991). Ainsi dans l'intron du gène de l'actine, une structuration secondaire séquestre un site 3' proximal au profit d'un site 3' aval. Des mutations qui déstabilisent la structure secondaire entraînent un épissage anormal au site 3' proximal. Dans le cas du pré-snoRNA U3, un certain degré de structuration en 3' est toléré. En effet, le site 3' des introns des pré-snoRNA U3A et U3B de *S. cerevisiae* est situé à la bordure d'une boucle latérale d'une structure tige/boucle (Mougin *et al.*, 1996).

Aucune expérience ne démontre encore actuellement de manière définitive quelles sont, parmi toutes les interactions citées, celles ayant lieu à l'une ou l'autre des étapes depuis l'assemblage du complexe B (dans la nomenclature vertébrés) jusqu'à la dissociation du spliceosome, après la réaction (Fig. 22). Néanmoins, il est raisonnable de penser, d'après des études basées sur la cinétique de formation des pontages covalents et les études de l'effet de mutations dans les ARN, que :

- l'hélice formée entre le snRNA U2 et la boîte de branchement,
- les hélices hétérologues 1a et 1b formées entre les snRNA U2 et U6,
- la tige/boucle du snRNA U6 qui jouxte cette hélice hétérologue,
- l'interaction entre le snRNA U6 et la séquence en 5' de l'intron,
- et les interactions entre le snRNA U5 et, dans un premier temps, la jonction exon 1-intron, puis, dans un deuxième temps, les deux exons,

constituent, avec les protéines qui stabilisent cette architecture et en modulent sans doute la réactivité, le site catalytique des deux étapes successives de l'épissage (Fig. 23).

En plus de ces interactions, reflétant la structure secondaire des ARN au sein du spliceosome, quelques interactions impliquées dans la structure tertiaire ont été mises en évidence. La première est une interaction de type Hoogsteen entre les deux résidus G aux extrémités de l'intron (Fig. 24) (Parker et Siliciano, 1993; Chanfreau *et al.*, 1994; Deirdre *et al.*, 1995). La nécessité de former cette interaction expliquerait en partie la stricte conservation d'un résidu G à ces deux positions. Les expériences de mutagenèse, appuyées sur les prédictions d'interactions possibles entre les nucléotides, montrent qu'une interaction très similaire au niveau de la position relative des bases et des riboses peut être formée entre un résidu A et un résidu C (Parker et Siliciano, 1993; Chanfreau *et al.*, 1994). Ce sont précisément ces deux résidus qui sont trouvés aux extrémités des introns appartenant à la classe mineure d'introns récemment mise en évidence (Fig. 13C) (Hall et Padgett, 1994). Une étude par mutagenèse à saturation des snRNA U2 et U6 a récemment permis de mettre aussi en évidence des interactions tertiaires possibles entre les résidus G_{52} du snRNA U6 et A_{25} du snRNA U2 (Madhani et Guthrie, 1994b) (Fig. 24).

Il est intéressant de noter que chez les vertébrés, pour lesquels les modifications post-transcriptionnelles portées par les UsnRNA avaient été identifiées (Branlant *et al.*, 1980; Harada *et al.*, 1980; Branlant *et al.*, 1981; Krol *et al.*, 1981a; 1981b; Branlant *et al.*, 1982; 1983, pour revue,

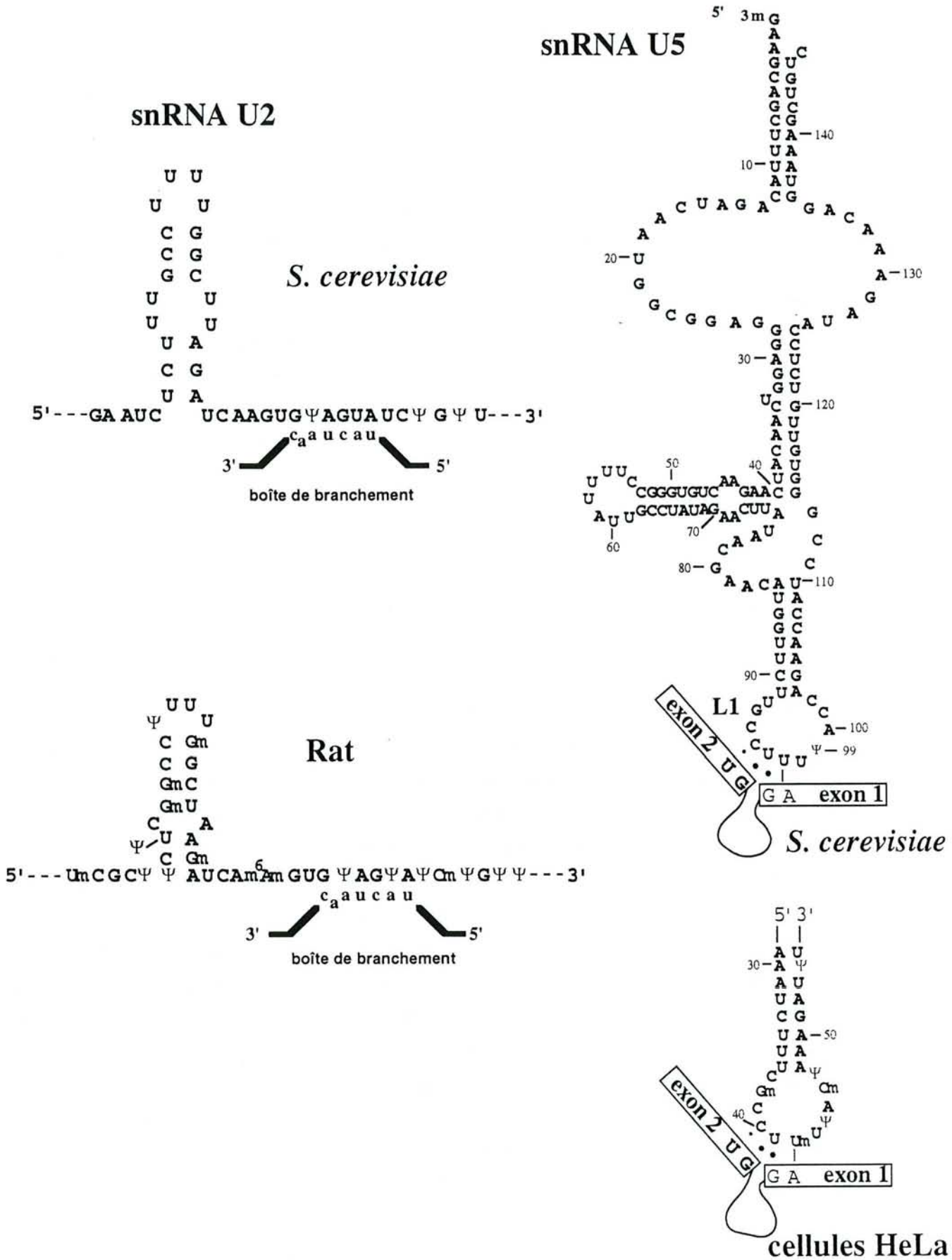


Figure 25 : Comparaison des nucléotides modifiés dans la région 5' terminale du snRNA U2 et la tige/boucle 5' terminale du snRNA U5 chez l'homme, le rat et chez *S. cerevisiae*.
L'appariement avec la région du point de branchement est indiquée pour le snRNA U2, l'interaction avec les exons pour le snRNA U5.

voir Solymosy et Pollack, 1993), environ 95% de ces modifications sont contenues dans les segments impliqués dans les interactions hétérologues (Fig. 22). Cette situation est particulièrement marquante pour les snRNA U2 et U6 et pour la boucle terminale du snRNA U5 dont les modifications post-transcriptionnelles sont fortement conservées au cours de l'évolution (Szkukalek *et al.*, 1995). On peut supposer un rôle important de ces modifications dans la mise en place des interactions hétérologues, leur stabilisation ou la fonction des ARN dans la catalyse.

Au laboratoire, S. Massenet a récemment identifié les résidus Ψ contenus dans les UsnRNA de *S. cerevisiae*. Un nombre beaucoup plus faible de résidus Ψ a été mis en évidence chez ce type d'organisme, aucun résidu Ψ n'est présent dans les snRNA U1, U4 et U6. Un seul résidu Ψ est présent dans le snRNA U5 et il s'agit d'un des deux résidus de la boucle terminale I qui sert à aligner les exons, l'autre résidu U converti en Ψ chez les vertébrés étant remplacé par un C (Fig. 25). Trois résidus Ψ sont présents dans le snRNA U2, contre 13 chez les vertébrés, un de ces résidus Ψ est contenu dans la séquence qui s'apparie à la boîte de branchement. Précisément, il forme une des deux paires de bases qui encadrent le résidu A en boucle latérale qui réalise la première attaque nucléophile. Les deux autres résidus Ψ sont situés dans la séquence juste en aval qui forme l'hélice hétérologue 3 avec le snRNA U6 et interagit avec de nombreux facteurs protéiques (Fig. 25).

L'absence de résidus Ψ dans la séquence du snRNA U1 se liant à la boîte 5' chez *S. cerevisiae* et la présence d'un seul résidu Ψ dans la séquence du snRNA U2 se liant à la boîte de branchement pourraient être liées au fait que le degré de complémentarité entre intron et UsnRNA est plus élevé chez *S. cerevisiae* et donc nécessite moins une stabilisation de l'interaction formée par la présence de résidus Ψ. On sait, en effet, que les résidus Ψ favorisent la formation de structures hélicoïdales dans les ARN (Davis et Poulter, 1991).

B4.2 Comparaison avec les interactions ARN/ARN s'établissant au cours de l'épissage des introns "mineurs" et des introns de groupe II.

Bien que des UsnRNA différents soient impliqués dans le nouveau type de spliceosome récemment mis en évidence chez les eucaryotes supérieurs, la reconnaissance du site 5' est assurée par un appariement de bases avec le snRNA U11 (Fig. 26), qui mime l'interaction entre le snRNA U1 et l'extrémité 5' des introns majeurs. Le site de branchement est reconnu par appariement de bases avec le snRNA U12 et, comme dans le cas de l'interaction avec le snRNA U2, l'adénosine responsable de la première attaque nucléophile est située en boucle latérale (Fig. 26) (pour revue, Mount, 1996). Enfin, la structure formée entre les snRNA U2 et U6 est retrouvée entre le snRNA U12 et un snRNA qui remplace le snRNA U6 (résultats non publiés de l'équipe de J. Steitz).

Non seulement ce type de structuration des ARN au sein du site catalytique semble donc très conservé entre les deux types de spliceosomes, mais, en plus, il présente de fortes homologies avec la structuration du site catalytique des introns de groupe II. Comme nous l'avons déjà mentionné,

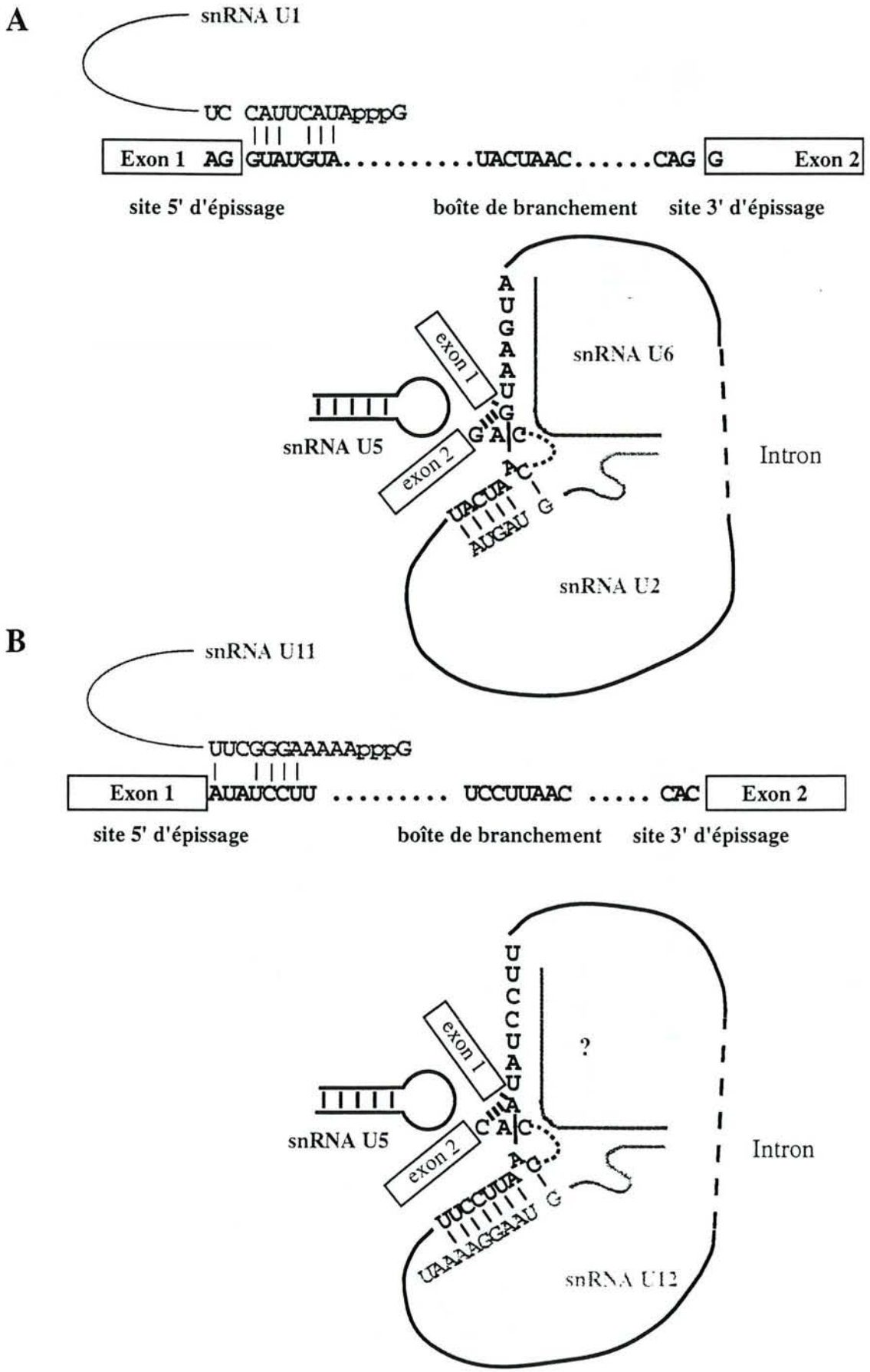


Figure 26 : Comparaison du rôle des snRNA U1 et U2 dans les spliceosomes majeurs (A) à celui de leurs homologues U11 et U12 dans la forme mineure de spliceosomes (B)(Mount, 1996).

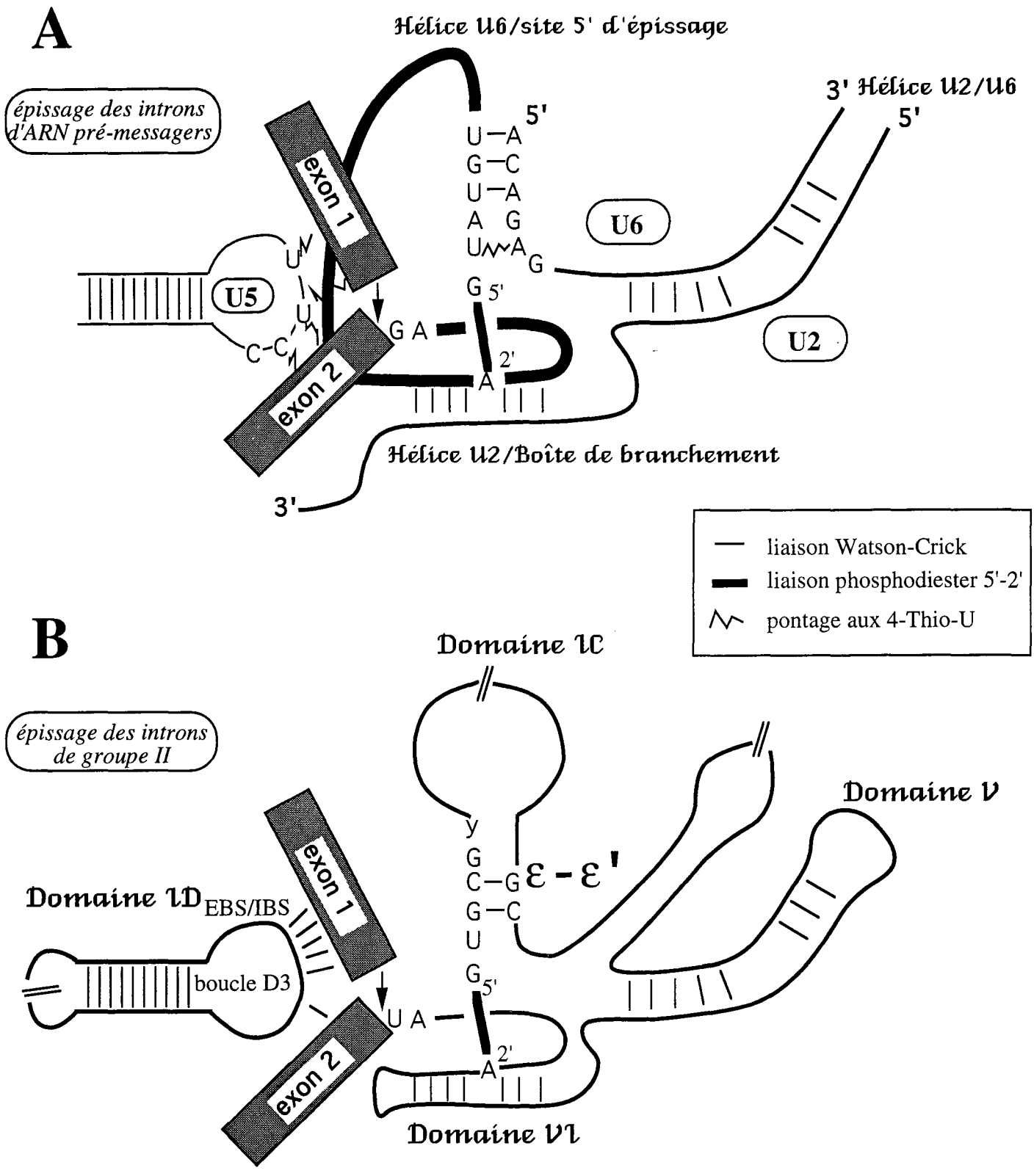


Figure 27 : Comparaison des interactions ARN/ARN mises en jeu dans l'épissage des introns des ARN pré-messagers (A) et des introns de groupe II (B) (Wise, 1993).
A. Interactions au niveau du site catalytique du spliceosome au cours de la deuxième réaction de transestérification, les séquences utilisées sont les séquences consensus chez *S. cerevisiae*,
B. Interactions au site catalytique des introns de groupe II s'établissant au cours de la deuxième réaction de transestérification. Les séquences conservées des introns de la forme IIA sont indiquées.

l'épissage des introns de groupe II et l'épissage des ARN pré-messagers impliquent les mêmes attaques nucléophiles, d'où l'idée que les introns des ARN pré-messagers pourraient avoir les introns de groupe II pour ancêtres.

En effet, une structure équivalente à celle formée par le snRNA U2 et la boîte de branchement est retrouvée dans les introns de groupe II (domaine **VI**, Fig. 27). Dans ce cas aussi, l'adénosine responsable de la première attaque nucléophile est située en boucle latérale.

Le domaine **V** des introns de groupe II, situé juste en amont du domaine VI, joue un rôle essentiel dans l'épissage de ces introns (Koch *et al.*, 1992; Chanfreau et Jacquier, 1994). Ce domaine présente certaines homologies avec la structure formée entre les snRNA U2 et U6 (Fig. 27) (Wise, 1993).

Le court appariement **ε-ε'** à longue distance entre la séquence en 5' du site de branchement et un dinucléotide du domaine 1C dans les introns de groupe II (Fig. 27) (Jacquier et Michel, 1990) peut être considéré comme l'homologue de l'interaction entre le snRNA U6 et la séquence en 5' de l'intron dans les introns épissés dans un spliceosome.

Enfin, la boucle **D3** du domaine 1D des introns de groupe II joue, pour ces introns, un rôle similaire à celui qui est proposé pour le snRNA U5 dans le cas des introns épissés dans un spliceosome (Fig. 27). En effet, la séquence EBS1 de cette boucle s'apparie à l'exon 1 lors de la première étape de la réaction et servirait à aligner les deux exons pour la deuxième étape de la réaction (Jacquier et Michel, 1987; Augustin *et al.*, 1990; Mörl et Schmelzer, 1990; Mörl *et al.*, 1992).

L'idée actuelle est qu'au cours de l'évolution des introns de groupe II, les régions codant pour certaines séquences essentielles à l'épissage de ces introns, auraient été transférées hors de la séquence codant pour l'intron et auraient constitué des gènes autonomes, les gènes des UsnRNA. Ces ARN interviendraient alors en trans, au lieu d'intervenir en cis comme les séquences dont ils auraient dérivé. Des protéines auraient alors été impliquées dans l'assemblage des différents ARN intervenant en trans, ce rôle et le nombre de ces protéines s'étant sans doute beaucoup diversifiés au cours du temps, jusqu'à l'obtention des spliceosomes actuels qui renferment environ 60 protéines.

Selon cette idée, les UsnRNA joueraient donc un rôle essentiel dans la catalyse et vraisemblablement tout particulièrement la structure formée entre les snRNA U2 et U6. Comme nous l'avons vu, néanmoins avec des protéines telles que la protéine de 220 kDa et son équivalent PRP8, certaines protéines doivent jouer un rôle très important pour la constitution de l'entité catalytique.

Avant d'en terminer avec le mécanisme de base d'assemblage et de catalyse du spliceosome, il faut tout de même noter que tout ce qui vient d'être décrit correspond à la situation où : 1) un seul intron est présent dans l'ARN pré-messager (cas presque général chez *S. cerevisiae* à deux exceptions près), 2) la séquence intronique est courte (inférieure à 550nt) (Klintz et Gallwitz, 1985), ce qui est toujours le cas chez *S. cerevisiae* (Kalogeropoulos, 1995).

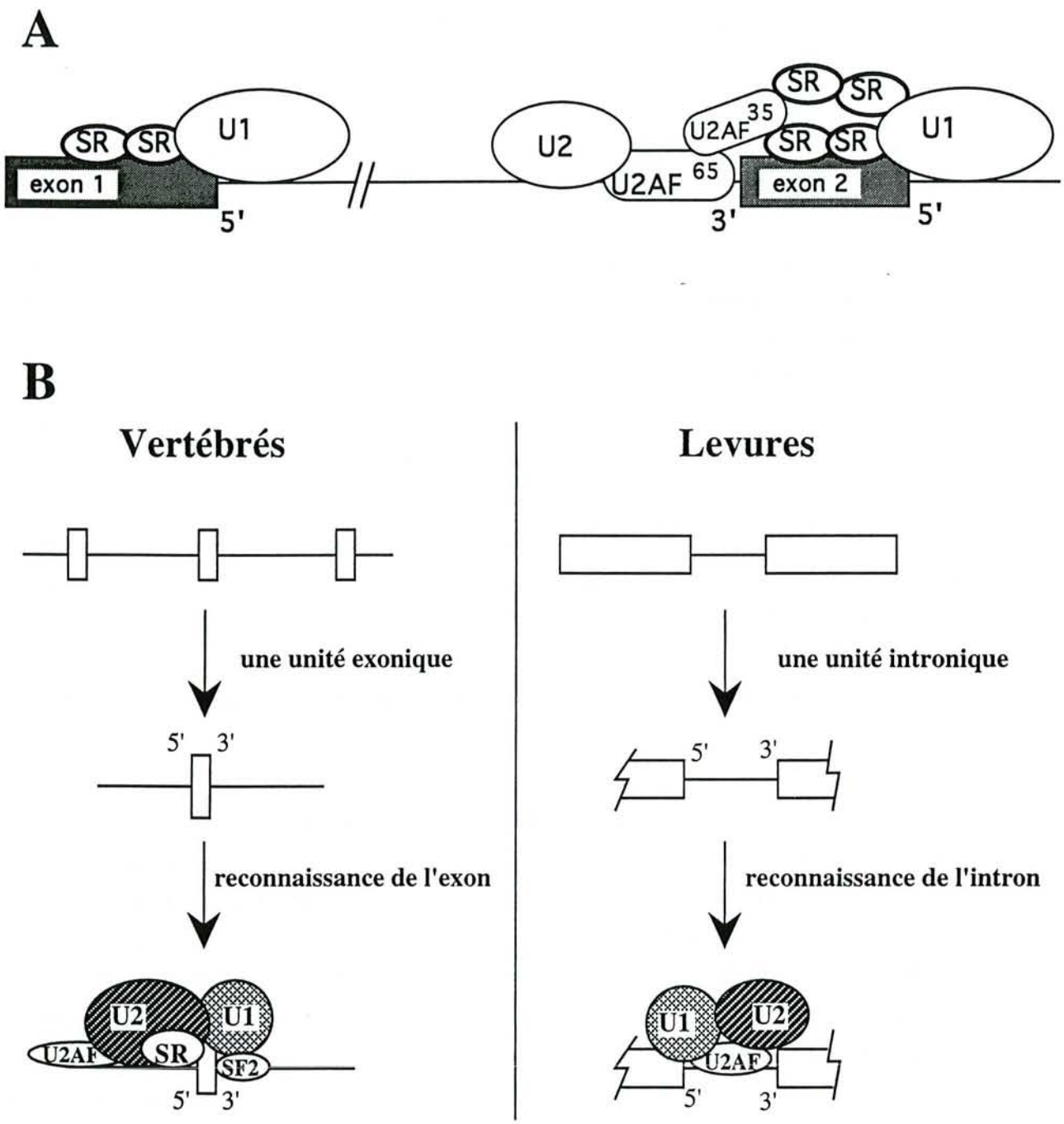


Figure 28 : Comparaison des reconnaissances des bornes introniques et exoniques chez les vertébrés et les levures (Berget, 1995).

A. Chez les vertébrés, un complexe entre les snRNP U1 et U2 se ferait au niveau de l'exon par l'intermédaire de protéines SR, ensuite, le spliceosome serait assemblé par rapprochement des extrémités de l'intron.

B. Formation du complexe "précoce" décrit en A via la "définition d'exon" pour les vertébrés (à gauche) et via la "définition d'intron" pour les eucaryotes inférieurs (à droite).

Chez les vertébrés, les exons sont généralement courts et les introns longs, parfois jusqu'à 10^4 à 10^5 bases (Hawkins, 1988). Berget et coll. ont proposé que, dans ce cas, la reconnaissance par la machinerie d'épissage ne se fait pas au niveau des introns, mais au niveau des exons (Berget, 1995) (Fig. 28). La mutation de la séquence en 5' d'un intron n'entraîne pas l'absence d'épissage de cet intron, mais l'exclusion de l'exon qui le précède dans l'ARN pré-messager (Krawczak *et al.*, 1992). Dans le cas du premier exon, la reconnaissance se ferait entre la coiffe terminale et la jonction en 5' et pour le dernier exon, entre la jonction 3' et le site de polyadénylation (Ohno *et al.*, 1987; Niwa et Berget, 1991; Izaurralde *et al.*, 1994; Lewis *et al.*, 1996; pour revue, Black, 1995).

Chez les insectes et les levures où les introns sont courts et les exons plus longs, la reconnaissance se ferait au niveau de l'intron, une mutation de la séquence en 5' entraînant l'absence d'épissage de l'intron (Fig. 28) (Talerico et Berget, 1994; Berget, 1995).

Il existe donc à ce niveau une différence importante entre *S. cerevisiae* et les vertébrés. Chez les vertébrés, où les sites d'épissage sont peu conservés, les pré-ARNm, très longs, contiennent de nombreuses séquences qui pourraient correspondre à des sites d'épissage. Ces cellules ont donc besoin d'un mécanisme particulier leur permettant de distinguer les "vrais" des "faux" sites d'épissage.

B5. Les mécanismes de régulation de l'épissage des ARN pré-messagers.

Chez les organismes où la différenciation cellulaire a lieu, la régulation de cette étape de l'expression est très complexe. En effet, suivant les types cellulaires, différentes possibilités d'épissage alternatif des ARN pré-messagers se présentent. Ces régulations reposent principalement sur l'intervention de facteurs protéiques exprimés à des taux différents suivant les types cellulaires. Certains facteurs vont être activateurs de l'utilisation de certains sites, d'autres au contraire inhibiteurs. Nous avons déjà cité le cas général du facteur ASF/SF2 qui permet d'utiliser le site 5' proximal, et de la protéine hnRNP A1 qui conduit à une utilisation du site distal. Ces protéines, qui ont des effets antagonistes, peuvent intervenir dans la sélection des sites d'épissage de la plupart des ARN pré-messagers présentant une multiplicité de sites 5'. Par contre, certains facteurs agissent spécifiquement sur le choix de sites au niveau d'un ARN pré-messager donné, car leur intervention nécessite la présence d'un site spécifique. C'est le cas des protéines Sx1, Tra1 et Tra2 qui régulent l'expression de protéines définissant le sexe chez la Drosophile au niveau des étapes d'épissage (Tian et Maniatis, 1992; Valcarcel *et al.*, 1993).

La structure secondaire des ARN pré-messagers peut aussi intervenir dans la régulation de l'utilisation des sites. Mais, là aussi, les différences entre types cellulaires dépendront de l'expression de certains facteurs protéiques qui vont soit stabiliser, soit déstabiliser ces structures secondaires.

Quelques exemples d'occlusion de sites par la formation de structures secondaires ont été mis en évidence chez les vertébrés, surtout dans le cas des ARN viraux. Ainsi, dans le cas de l'adénovirus 2, comme nous l'avons déjà mentionné, la formation d'une structure tige/boucle incluant un site 5' peut diminuer l'efficacité de ce site (Domenjoud *et al.*, 1991). La formation d'une structure tige/boucle à l'intérieur de l'intron permet, par ailleurs, l'utilisation d'un site 3' d'épissage anormalement éloigné du point de branchement (Chebli *et al.*, 1989).

Enfin, dans le cas du gène de la ß-tropomyosine de poulet, la formation d'une structure tige/boucle emprisonne un site 5' d'épissage, ce qui semble empêcher la reconnaissance de l'intron correspondant. Ce site 5' n'étant pas reconnu, c'est le site 5' de l'intron suivant qui le sera. Il en résulte une exclusion de l'exon 6B de l'ARNm de la ß tropomyosine de poulet (Libri *et al.*, 1991; Clouet d'Orval *et al.*, 1991). Libri *et al.* (1991) ont montré, *in vivo*, l'importance de la structure secondaire pour l'épissage alternatif de la ß-tropomyosine de poulet. La structure secondaire qui, impliquée dans l'épissage alternatif de ce pré-ARNm influence non seulement la fixation de la snRNP U1 sur le site 5' d'épissage (site impliqué dans une structure tige/boucle), mais également la fixation des snRNP U2 et U4/U6-U5 (Sirand-Pugnet *et al.*, 1995). La structure secondaire particulière de ce pré-ARNm interfèrerait-elle avec la communication qui s'établit entre sites 5' et 3' d'épissage d'où une déstabilisation des complexes se formant aux extrémités de l'intron ? Les résultats de Watakabe *et al.* (1989) semblent indiquer également l'intervention d'une structure secondaire dans la régulation de l'épissage alternatif du pré-ARNm de l'IgM de souris *in vitro*. Dans le cas d'épissage alternatif, les produits d'épissage étant différents en fonction des tissus cellulaires, du stade de développement, il faudrait qu'*in vivo*, ces structures secondaires puissent se former uniquement dans certains tissus par l'intermédiaire des facteurs protéiques stabilisant ou non ces structures d'où l'obtention d'un profil différent d'épissage en fonction du tissu cellulaire. Rien n'est actuellement connu à ce sujet. S'il est sûr qu'il existe une relation entre structure et efficacité d'épissage *in vitro*, cela l'est moins en ce qui concerne l'efficacité d'épissage *in vivo* (Solnick et Lee, 1987). Eperon *et al.* en 1988, ont mis en évidence que seule la formation de petites structures tige/boucle peut affecter le choix des sites d'épissage *in vivo*, ils en concluent donc que la vitesse de transcription peut influencer la formation de ces structures et donc indirectement l'épissage alternatif.

La situation semble différente chez les plantes où il a clairement été démontré que la formation de structures secondaires peut considérablement affecter l'efficacité d'épissage *in vivo* (Goodall et Filipowicz, 1991), l'utilisation de protoplastes permettant cette étude *in vivo*. L'introduction de tiges/boucles de 18 ou 24 paires de bases inhibe fortement l'efficacité d'épissage *in vivo* si cette structure séquestre le site 5' d'épissage ou si elle est placée au milieu de petits introns. Etonnamment, les structures tige/boucle séquestrant le site 3' n'empêchent pas l'utilisation de ce dernier : il semblerait que la fixation des snRNA U1 et U2 permettrait au spliceosome de déstabiliser la structure présente dans la partie 3' de l'intron (Liu *et al.*, 1995).

Chez les levures, dont les gènes renferment en général un seul intron, les seules régulations mises en évidence jusqu'à ce jour, mettent en jeu la formation de structures secondaires et la stabilisation de ces structures secondaires par des protéines.

Nous avons déjà cité le cas de l'intron du gène de l'actine de *Kluyveromyces lactis* où l'épissage correct nécessite la présence d'une structure secondaire (Deshler et Rossi, 1991). Un autre exemple, qui a été très étudié, est le rétrocontrôle de la protéine ribosomique RPL32 au niveau de l'étape d'épissage aussi bien chez *Kluyveromyces lactis* que chez *S. cerevisiae* (Eng et Warner, 1991). Dans ce cas, la séquence en 5' de l'intron est incluse dans une structure secondaire impliquant l'exon 1. La fixation de la protéine RPL32 stabilise cette structure. Cette structuration n'empêche pas l'association du snRNA U1, mais bloque les étapes ultérieures, en particulier, sans doute, la formation de l'interaction entre le snRNA U6 et l'extrémité 5' de l'intron. Il est possible que la structure formée (Li *et al.*, 1995) stabilise la snRNP U1 et bloque l'évolution des complexes pré-spliceosomaux (Vilardell et Warner, 1994).

D'autre part, l'équipe de M. Rosbash a constaté que l'efficacité d'épissage du pré-ARNm RP51B dépendrait de deux séquences complémentaires situées, l'une en aval du site 5' d'épissage et l'autre en amont de la boîte de branchement (Goguel et Rosbash, 1993; Libri *et al.*, 1995). Ils postulent donc qu'un appariement de ces deux séquences complémentaires serait indispensable à l'épissage des longs introns de levure, cet appariement permettant de rapprocher les deux sites d'épissage.

De plus, s'il a été vérifié que, chez *S. cerevisiae*, les structures secondaires s'établissant au sein de l'intron peuvent réguler *in vitro* et *in vivo* la réaction d'épissage, il semblait important de vérifier que l'effet d'une structure sur l'efficacité d'épissage *in vitro* est amoindri *in vivo*, et s'il est amoindri, dans quelles proportions. Dans ce but, Goguel *et al.* (1993) ont inséré par mutagenèse dirigée des séries de petites séquences dans un intron de façon à générer des structures tige/boucle plus ou moins grandes emprisonnant soit le site 5' d'épissage, soit la boîte de branchement. Étonnamment, l'inhibition de l'épissage *in vivo*, provoquée par l'emprisonnement des sites 5' ou des boîtes de branchement, est moins prononcée, mais très proche de l'inhibition obtenue *in vitro*. *In vivo*, six nucléotides appariés au site 5' peuvent déjà diminuer l'efficacité d'épissage, quinze nucléotides appariés bloquent complètement la réaction d'épissage et ceci dès la première étape de l'assemblage du spliceosome en empêchant le snRNA U1 de se fixer au site 5' (Goguel *et al.*, 1993).

D'une manière générale, comme nous l'avons montré tout au long de cette partie bilan concernant l'épissage, l'épissage dans un spliceosome est plus dépendant de la structure primaire des boîtes fonctionnelles de l'intron et de la structure secondaire du pré-ARNm chez *S. cerevisiae* que chez les vertébrés.

Champignons	Division
Eumycota	Sous-division
Ascomycotina	Classe
Hemiascomycetes	Sous-classe
Endomycetales (ou Saccharomycetales)	Ordre
Saccharomycetaceae	Famille
Sacchraomycetoideae	Sous-famille
Saccharomyces	Genre
Saccharomyces cerevisiae	Espèce

Tableau VII : Classification taxonomique de la levure *Saccharomyces cerevisiae* (selon Kreger van Rij, 1987).

C. Classification des levures.

Etant donné que nous avons travaillé sur différentes espèces de levures et que notre étude peut apporter des informations en matière de taxonomie, nous ferons une brève description de l'état actuel de la classification des levures.

La définition des levures est très vague : "Les levures sont des champignons chez lesquels la forme unicellulaire est prédominante" (Larpent et Larpent-Gourgaud, 1985.).

C1. Les familles, sous-familles et genres dans la classification actuelle.

Les levures sont retrouvées dans divers biotopes, terrestres, aquatiques, elles peuvent être associées à des plantes, des insectes. Elles sont également utilisées en industrie, en boulangerie et en brasserie. *S. cerevisiae* est la levure la plus étudiée, mais aussi la plus utilisée en industrie boulangère, de nombreuses levures du genre *Saccharomyces* sont utilisées en brasserie. Des levures sont responsables de mycoses, certaines sont pathogènes, telles que *Candida albicans*.

La définition des espèces et genres de levures est principalement basée sur des critères morphologiques, physiologiques, biochimiques et parfois moléculaires. La classification de la levure *S. cerevisiae* est présentée dans le Tableau VII.

La classe des Ascomycètes est séparée en deux sous-classes : Hemiascomycetideae et Ascomycetideae. La sous-classe Hemiascomycetideae est subdivisée en trois ordres : Endomycetales, Protomycetales et Taphrinales. L'ordre Endomycetales est lui-même divisé en deux familles : la famille Saccharomycetaceae et la famille Spermophthoraceae. La famille Saccharomycetaceae est divisée en quatre sous-familles : la sous-famille Schizosaccharomycetoideae qui comprend un seul genre, la sous-famille Saccharomycetoideae qui comprend 26 genres, la sous-famille Lipomycetoideae qui ne comprend qu'un seul genre et enfin la sous-famille Nadsonioideae qui comprend 4 genres (Figure 29).

Bien que classées dans le même ordre, une distance génétique importante sépare les levures bourgeonnantes des levures scissibles (Kurtzman, 1993). Cette distance importante ressort clairement des études basées sur la comparaison de séquences partielles d'ARN ribosomiques 17S et 25S (Kurtzman et Robnett, 1991). L'arbre phylogénétique établi à partir de ces données montre que le genre *Schizosaccharomyces* est très éloigné des trois autres genres *Saccharomyces*, *Debaryomyces* et *Schwanniomyces*. Les levures scissibles appartiennent au genre *Schizosaccharomyces*, et les levures bourgeonnantes appartiennent aux trois autres genres.

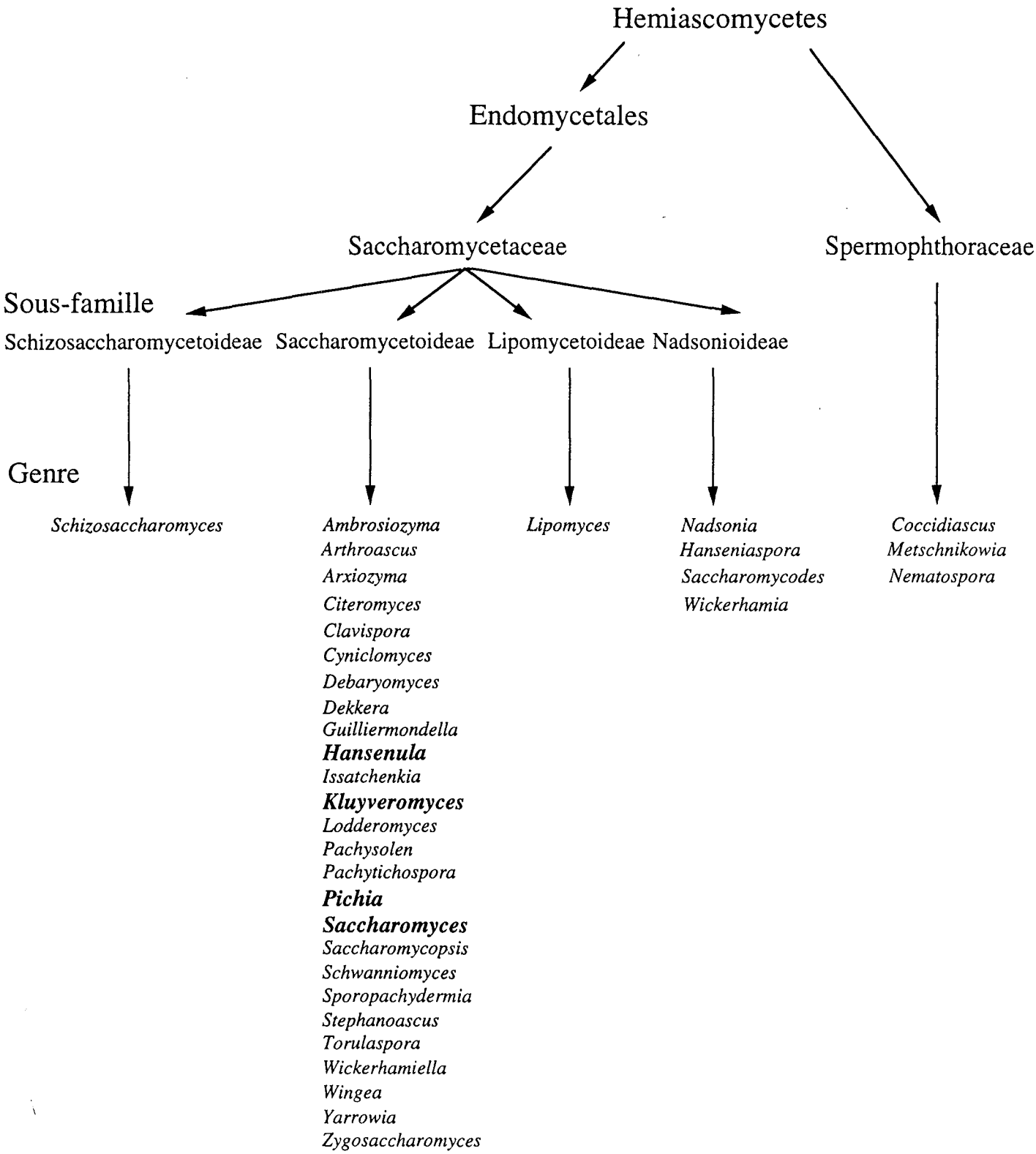


Figure 29 : Relations phylogénétiques entre les levures selon Kreger van Rij, 1987.

En plus des levures ascomycètes, il existe des champignons ascomycètes, *Aspergillus nidulans* est par exemple un champignon ascomycète filamenteux. Un arbre phylogénétique a été établi par Smith en 1989 entre des levures ascomycètes et des champignons ascomycètes sur la base de la comparaison des gènes codant pour une enzyme, la glycéraldéhyde-3-phosphate déshydrogénase (GAPDH). L'arbre phylogénétique déduit de cette étude montre que les levures ascomycètes *S. cerevisiae* et *Zygosaccharomyces rouxii* ont évolué séparément des champignons ascomycètes filamenteux avec lesquels elles sont cependant classées. Toujours d'après cette étude, les levures apparentées à *S. cerevisiae* dériveraient d'un ancêtre commun apparu très tôt dans l'évolution des eucaryotes, alors que les champignons filamenteux ascomycètes et également les basidiomycètes auraient divergé plus tard à partir d'un ancêtre commun.

C2. Les changements récents dans la classification des espèces.

Les méthodes anciennement utilisées pour établir les critères de description d'une levure, ont été décrites par Wickerman en 1951, Van der Walt en 1970, plus récemment des méthodes moléculaires ont été introduites (Van der Walt et Yarrow, 1984).

L'ensemble de ces données a conduit à reclasser certaines espèces dans un genre différent de celui où elles avaient été classées. Nous citerons d'abord les critères utilisés.

C2.1 Le mode de reproduction.

Le type de reproduction est le premier facteur étudié. Deux types de reproduction peuvent être utilisés par les levures, la reproduction végétative ou la reproduction sexuée.

La reproduction végétative peut se faire selon deux modes, la production par bourgeonnement, typique des levures du genre *Saccharomyces* et la formation de parois au milieu de la cellule, suivie d'une scission, ce type de reproduction est typique des levures du genre *Schizosaccharomyces*.

La reproduction sexuée est caractéristique des levures ascosporogènes. Le mode de formation des asques, la forme des asques, le nombre de spores par asque, la couleur des spores, leur possible relarguage de l'asque et leur mode de germination sont des critères typiques utilisables pour la classification.

La variation des formes des ascospores de levures est un bon critère de distinction des espèces, les spores pouvant être rondes, ovales, réniformes ou en forme d'aiguille. La forme des spores a été étudiée par microscopie électronique (Kurtzman *et al.*, 1975).

C2.2 Les critères physiologiques et biochimiques.

Les principaux critères physiologiques utilisés dans la classification des levures sont la fermentation et l'assimilation de sources carbonées. La description standard des espèces inclut la capacité à fermenter certains glucides pendant une période donnée. Dans une description standard, une étude de la fermentation de 18 à 30 composés carbonés est réalisée.

Les composés carbonés utilisés dans la majorité des descriptions standard sont : le galactose, le glucose, le maltose, le cellobiose, le tréhalose, le lactose, le raffinose, l'amidon soluble, le D-xylose, le L-arabinose, le D-ribose, le L-rhamnose, le ribitol, le D-mannitol, l'acide succinique, l'acide citrique et l'inositol (Kreger van Rij, 1987). D'autres composés sont parfois nécessaires pour différencier les espèces appartenant à un même genre. Il est à noter que des impuretés dans les composants testés et des conditions de tests mal adaptées peuvent conduire à des résultats aberrants qui peuvent entraîner une mauvaise classification.

L'assimilation des composés nitrés est testée de la même manière que celles des composés carbonés. L'assimilation du nitrate de potassium apporte des informations pour la classification.

La croissance à 37°C ou à d'autres températures, en particulier, la température maximale de croissance sert également de critère de classification (Vidal-Leiria *et al.*, 1979).

Avec l'augmentation du nombre d'espèces, l'ensemble des éléments clés de classification devient de plus en plus complexe. Un ensemble de caractères clés peut conduire, tout d'abord à la détermination du genre, puis de l'espèce, ou directement à l'espèce sans passer par le genre. Dans ce cas, en général, seuls sont considérés les critères physiologiques (Kreger van Rij, 1987).

Un exemple de la difficulté à établir une classification à partir de ces données est illustrée ci-dessous. Le genre *Kluyveromyces* a été défini comme un nouveau genre d'Endomycétales par Van der Walt en 1956. En 1970, cet auteur a transféré vers ce genre quatre levures précédemment classées dans le genre *Saccharomyces* par Lodder et Kreger van Rij en 1952, *S. fragilis*, *S. lactis*, *S. marxianus* et *S. veronae*, ceci principalement sur la base de la fragilité des asques et sur l'aspect réniforme des spores. En 1972, Campbell a réalisé une analyse numérique prenant en compte 15 caractères morphologiques et 43 caractères physiologiques, ceci pour 41 espèces couramment reconnues comme appartenant au genre *Saccharomyces* et 18 espèces du genre *Kluyveromyces*. Sur la base des données ainsi obtenues, *K. lactis* et *K. marxianus* ont été reclassées dans le genre *Saccharomyces* (*S. lactis* et *S. marxianus*), et l'espèce *K. fragilis* a été assimilée à *S. marxianus*. En 1984, les résultats de Van der Walt ont amené cet auteur à contester la nouvelle classification qui avait été proposée par Campbell (1972), *S. marxianus* a alors été redéfinie comme *K. marxianus* var. *marxianus* et *S. lactis* comme *K. marxianus* var. *lactis*.

Ces changements successifs de la classification des levures des genres *Saccharomyces* et *Kluyveromyces* illustrent le manque de fiabilité des critères physiologiques et morphologiques et la nécessité de faire appel à des critères moléculaires.

C2.3 Les critères moléculaires.

Les critères moléculaires ont tout d'abord reposé sur la composition en bases de l'ADN et plus particulièrement le pourcentage en G+C, sur des expériences d'hybridation entre ADN hétérologues, sur des cinétiques de réassociation d'ADN hétérologues, et enfin, sur les séquences partielles ou complètes de gènes d'ARN ribosomiques, mais aussi d'autres gènes.

En ce qui concerne les espèces que nous avons étudiées, il est intéressant de noter que l'approche moléculaire a conduit à une révision du genre d'appartenance de certaines des espèces.

Ainsi en 1970, Bicknell et Douglas ont réalisé des hybridations ADN/ADN et ADN/ARNr entre espèces du genre *Saccharomyces*. Ce genre comptait alors les espèces *S. cerevisiae*, *S. fragilis* et *S. lactis*. L'étude était basée sur la formation de duplex d'ADN en présence d'ADN compétiteur. Cette étude a permis de proposer la classification de *S. fragilis* et *S. lactis* dans le genre *Klyuveromyces*, résultats en accord avec ceux de Van der Walt en 1970, ensuite, comme nous l'avons vu, ces résultats ont été dénoncés par Campbell en 1972, puis réactualisés en 1984 par Van der Walt et Yarrow.

La combinaison de paramètres physiologiques et de l'étude de la composition en bases de l'ADN, analysée par cinétique de réassociation pour une étude des genres *Pichia* et *Hansenula*, a conduit Kurtzman (1984) à reclasser certaines espèces du genre *Hansenula* dans le genre *Pichia*. Ces deux genres avaient été séparés sur la base de leur capacité à assimiler le nitrate comme seule source d'azote (Wickerman, 1951). Par ailleurs, en 1979, une étude menée par Fuson *et al.* basée sur la détermination du pourcentage en G+C de l'ADN, par des études de cinétique de réassociation de l'ADN, a permis de définir une nouvelle espèce dans le genre *Hansenula*, l'espèce *Hansenula alnii*, qui était précédemment classifiée dans l'espèce *Hansenula wingei*.

Dans le cadre du développement de nouveaux critères moléculaires, nous soulignerons l'intérêt que représentent les séquences des unités transcriptionnelles renfermant les gènes des ARN ribosomiques, ainsi que les séquences des gènes des UsnRNA, séquences auxquelles nous nous sommes intéressées au cours de cette étude.

Des comparaisons de séquences d'ARN ribosomiques de levures ont été réalisées. Dès 1984, ainsi, Chen *et al.* ont aligné les séquences de l'ARN ribosomique 5S de 19 champignons ascomycètes, dont trois levures, *S. cerevisiae*, *S. carlsbergensis* et *K. lactis*. Les séquences de cet ARN sont identiques pour ces trois levures. L'ARN ribosomique 5S est relativement petit, 120nt. Les gènes codant pour cet ARN sont en grand nombre de copies dans la cellule. Cette multiplicité de gènes empêche une divergence trop rapide des séquences entre espèces relativement proches.

L'utilisation de séquences partielles des grands ARN ribosomiques 17S et 25S s'est avérée plus informative. En 1993, Kurtzman a réalisé une étude sur des levures phylogénétiquement beaucoup plus éloignées que celles étudiées précédemment par Chen *et al.* (1984), étude qui a permis de

confirmer un éloignement phylogénétique important entre les levures appartenant aux genres *Saccharomyces* et *Schizosaccharomyces*.

D'une manière générale, les unités d'ADN ribosomiques de différentes levures ont été étudiées de plusieurs manières. Des cartes de restriction de ces unités ont été établies pour les levures *S. cerevisiae*, *Saccharomyces carlsbergensis*, *Schizosaccharomyces pombe*, *Yarrowia lypolytica*, *Kluyveromyces lactis*, *Hansenula wingei* et *Candida utilis* (Garber *et al.*, 1988).

L'équipe de R.J. Planta (Verbeet *et al.*, 1984a) a étudié la formation d'hétéroduplex entre ces unités d'ADNr des levures *K. lactis* et *H. wingei* et *S. carlsbergensis*. Cette étude a permis de montrer une absence d'homologie significative entre les séquences transcrites ne correspondant pas aux ARN ribosomiques matures (Verbeet *et al.*, 1984a).

La même équipe a réalisé des études moléculaires plus poussées. Elle a étudié les sites d'initiation de la transcription des opérons ADNr chez les levures *K. lactis*, *H. wingei*, et les a comparé avec ceux des levures *S. carlsbergensis*, *S. cerevisiae* et *Saccharomyces rosei* (Verbeet *et al.*, 1984b). La même séquence est présente au niveau des positions +1 à +7 du site d'initiation. Mais de fortes divergences apparaissent en amont de la position -9 dans la région NTS (Non Transcribed Spacer, qui fait partie de l'unité d'ADNr). Dans la région 5'ETS, en aval de la position +23, les séquences sont conservées entre les levures *K. lactis*, *S. carlsbergensis* et *S. rosei*, mais la région 5'ETS de *H. wingei* présentait une divergence de séquence significative par rapport aux trois autres levures étudiées.

L'alignement des séquences d'ADN ribosomiques peut être combiné à la comparaison des structures secondaires. En effet, des séquences en ADN très variables peuvent permettre à l'ARN d'adopter une structure secondaire identique. Un modèle de structure secondaire complet de l'ARNr 25S avait été établi en collaboration entre l'équipe de C. Branlant et celle de R.J. Planta (Veldman *et al.*, 1981). Une étude récente faite sur une région très variable de l'ARNr 25S (hélices 17.1 et 17.2) a été réalisée chez des levures très différentes les unes des autres (Blanz *et al.*, 1989). Malgré des séquences très variables, les auteurs ont retrouvé une certaine conservation de la structure secondaire.

Les snRNA, comme nous l'avons vu, sont des ARN relativement conservés au cours de l'évolution, en particulier, le snRNA U6 (Brow et Guthrie, 1988). La séquence des UsnRNA spliceosomaux de *S. cerevisiae* a été établie (pour revue, Guthrie et Patterson, 1988). Alors que le snRNA U6 de *S. cerevisiae* est conservé en taille et en séquence par rapport aux snRNA U6 d'autres eucaryotes, la taille des snRNA U1 et U2 de *S. cerevisiae* est, respectivement, 3,5 et 6 fois, plus importante que celle des snRNA U1 et U2 de métazoaires (570nt pour U1 et 1175nt pour U2 chez *S. cerevisiae*, contre 164nt pour U1 et 187nt pour U2 chez l'homme). Les tailles des snRNA de *S. pombe* sont semblables à celles observées chez les métazoaires (Porter *et al.*, 1988; 1990; Dandekar *et al.*, 1989; Dandekar et Tollervey, 1989).

Les tailles des snRNA de 24 levures dites bourgeonnantes ont été déterminées par électrophorèse en gel de polyacrylamide, suivie d'une hybridation avec des oligonucléotides complémentaires de courtes régions conservées des snRNA (Roiha *et al.*, 1989). Cette étude permet de séparer en deux groupes ces différentes levures, le premier groupe présentant des snRNA U1 et U2 de grande taille proche de celle des snRNA U1 et U2 de *S. cerevisiae*. Le second groupe présente des snRNA U1 et U2 dont la taille est proche de celles des snRNA correspondant de métazoaires. Le premier groupe comprend des levures des genres *Saccharomyces*, *Kluyveromyces*, *Saccharomycopsis* (levures de la sous-famille des Saccharomycetoideae) et *Saccharomycodes* (levures de la sous-famille des Nadsonioideae), le second groupe comprend des levures des genres *Pichia*, *Clavispora*, *Ambrosiozyma*, *Pachysolen*, *Yarrowia* (levures de la sous-famille des Saccharomycetoideae), *Hanseniaspora*, *Nadsonia*, *Wickerhamia* (levures de la sous-famille des Nadsonioideae).

Le snRNA U6, le plus conservé en taille et en séquence au cours de l'évolution est très conservé dans les 24 levures étudiées (Roiha *et al.*, 1989).

La séquence des snRNA U1 de *Kluyveromyces lactis* et *Saccharomyces bayanus* ont été étudiées (Kretzner *et al.*, 1987).

Au niveau snoRNA nucléolaire, aucune information n'était disponible pour les espèces autres que *Saccharomyces cerevisiae* et *Schizosaccharomyces pombe*.

Matériels et Méthodes

D. Matériels.

D1. Souches de levure.

Saccharomyces cerevisiae FL 100 ATCC 28383 (Lacroute, 1968),
Saccharomyces cerevisiae GY5 ATCC 25657 (Hughes *et al.*, 1987),
Saccharomyces cerevisiae JH84 (Hughes et Ares, 1991).

Souches fournies par Mme le Docteur Petitdemange, IUT du Montet, Nancy, France :
Saccharomyces bayanus STV107
Saccharomyces florentinus STV 91,

Souche fournie par le Docteur Moll, Tépral, Champigneulles, France :
Souche de brasserie identifiée comme appartenant à l'espèce *Saccharomyces uvarum.*

Souche fournie par le Professeur Bonaly, Faculté de Pharmacie, Nancy, France :
Kluyveromyces marxianus var. *fragilis* CBS1555

Souches fournies par le Docteur Fukuhara, Institut Curie, Orsay, France :
Kluyveromyces delphensis CBS 2170
Kluyveromyces marxianus var. *lactis* CBS 2360
Kluyveromyces marxianus var. *bulgaricus* CBS 397
Pichia anomala CBS 5759
Pichia guilliermondii CBS 2030
Pichia pastoris CBS 704
Pichia salictaria CBS 5456

Souche fournie par le Docteur Tollervey, EMBL, Heidelberg, Allemagne :
Hansenula wingei ATCC 14356

Toutes les souches de levures sont conservées -80°C dans un milieu YPG (ou YPD) contenant 20% de glycérol. Les souches de levure peuvent être étalées sur boite de Pétri contenant du milieu YPG gélosé, et conservées à 4°C pendant 1 mois.

D2. Souches bactériennes.

Escherichia coli TG1 : *supE*, *hsdΔ*5, *thi*, Δ(*lac-pro*AB), F ' [*traD*36, *proA*$^+$*B*$^+$, *lacI*^{q}, *lacZ*ΔM15] (Gibson, 1984).

* Souches utilisées pour la mutagenèse dirigée :

Escherichia coli HB2151 : *ara*, *thi*, Δ(*lac-pro*AB), F ' [*proA*$^+$*B*$^+$, *lacI*^{q}, *lacZ*ΔM15] (Carter *et al.*, 1985),

Escherichia coli HB2154 : *ara*, *thi*, *mutL* :: *Tn*10, Δ(*lac-pro*AB), F ' [*proA*$^+$*B*$^+$, *lacI*^{q}, *lacZ*ΔM15] (Carter *et al.*, 1985).

Cette dernière souche possède les mêmes caractéristiques que la précédente, mais elle est mutée dans le système de réparation et ne possède pas d'ARNt suppresseur (Carter *et al.*, 1985).

D3. Milieux de culture.

Les cultures de levures sont réalisées en milieu non sélectif YPGlu (bactopeptone 10 g/l, extrait de levure 10 g/l, glucose 20 g/l). La souche de levure *S. cerevisiae* JH84 est cultivée sur milieu non sélectif YPGal (bactopeptone 10 g/l, extrait de levure 10 g/l, galactose 20 g/l).

Les levures transformées ont été cultivées en milieu non sélectif YPGlu et en milieu sélectif minimum YNBGal [YNB pour Yeast Nitrogen Base (Base azotée de levure)] sans acides aminés 6,7 g/l; Galactose 20 g/l; de l'histidine et/ou de l'adénine sont ajoutées à une concentration finale de 20 mg/l chacune). Les milieux sont stérilisés en autoclave durant 15 minutes, à 121°C et à 1 bar.

Le milieu solide est obtenu en ajoutant du bactoagar à 20g/l.

Les cultures de bactéries ont été réalisées en milieu non sélectif LB (bactopeptone 10g/l, extrait de levure 5 g/l, NaCl 5 g/l pH 7,2), stérilisé 45 min à 0,5 bar. Le milieu solide est obtenu en ajoutant du bactoagar à 20 g/l.

Si nécessaire de l'ampicilline, à une concentration finale de 100 µg/ml, est ajoutée au milieu de culture LB préalablement stérilisé, puis refroidi à 45°C.

D4. Plasmides, Bactériophages et Phagemides.

D4.1 Plasmides.

pBR322 (Bolivar *et al.*, 1977)

Ce plasmide de 4,3 kb construit à partir du plasmide ColE1, porte les gènes de résistance à l'ampicilline et à la tétracycline.

pUC18 et pUC19 (Vieira et Messing, 1982)

Ces plasmides de 2,7 kb dérivent du plasmide pBR322, dont ils possèdent la séquence de l'origine de réplication et le gène de résistance à l'ampicilline. Ils présentent l'avantage de contenir une cassette composée d'un fragment du gène de la β-galactosidase d'*E.coli* qui permet l'α-complémentation de souches ayant la caractéristique génotypique Δlac. Dans ce fragment du gène de la β-galactosidase, une cassette de clonage a été insérée. Elle posséde un grand nombre de sites uniques de coupure par des enzymes de restriction. Ces deux plasmides diffèrent par l'orientation de cette cassette de clonage par rapport au promoteur *lac*.

pMOSBlue-T (kit Amersham RPN 1719) (Fig. 30A)

Ce plasmide est utilisé pour cloner des fragments amplifiés par la méthode PCR sans traitement préalable par des enzymes de restriction ou des enzymes qui permettent d'obtenir des fragments d'ADN à bout franc. Les fragments d'ADN provenant d'amplification avec l'ADN polymérase *Taq* possèdent, en effet, à leur extrémité 3' terminale un résidu A protubérant. Cette particularité rend le clonage direct des fragments difficile. C'est pourquoi des systèmes de clonage basés sur la production d'un ADN avec un résidu T sortant ont été développés.

Dans le cas du plasmide pMOSBlueT, une hydrolyse avec la nucléase *Eco*R V est réalisée. Le site de reconnaissance de cette nucléase est localisé au milieu de la cassette de clonage du plasmide. Après linéarisation, l'ADN plasmidique est placé en présence de ddTTP et de terminal transférase (Holton et Graham, 1991). L'utilisation de ddTTP permet l'addition d'un seul résidu T lors de la réaction.

pASZ11 (Stotz et Linder, 1990) (Fig. 30B)

Ce plasmide est un plasmide navette *E. coli/S. cerevisiae* centromérique, construit à partir du plasmide pUC19 par insertion de 4 éléments de séquences lui conférant les particularités suivantes :

* une séquence ARS [pour Autonomous Replicative Sequence (Séquence de Réplication Autonome)] issue de l'ADN génomique de *S. cerevisiae* assurant une réplication du plasmide synchronisée avec celle des chromosomes. Une telle origine de réplication confère une plus grande stabilité au plasmide que l'origine de réplication du plasmide 2μ conduisant à des plasmides multicopies,

* la séquence *CEN*VI, séquence centromérique du chromosome VI de *S. cerevisiae*. Cette séquence assure une répartition équilibrée des copies du plasmide entre les deux cellules filles au cours de la mitose et son maintien en un faible nombre de copies (1 à 2 copies par cellule),

* la séquence *TRP*1 marqueur de sélection permettant la synthèse du tryptophane dans les souches *trp*1,

* la séquence *ADE*2 marqueur de sélection permettant la synthèse de l'adénine dans les souches *ade*2.

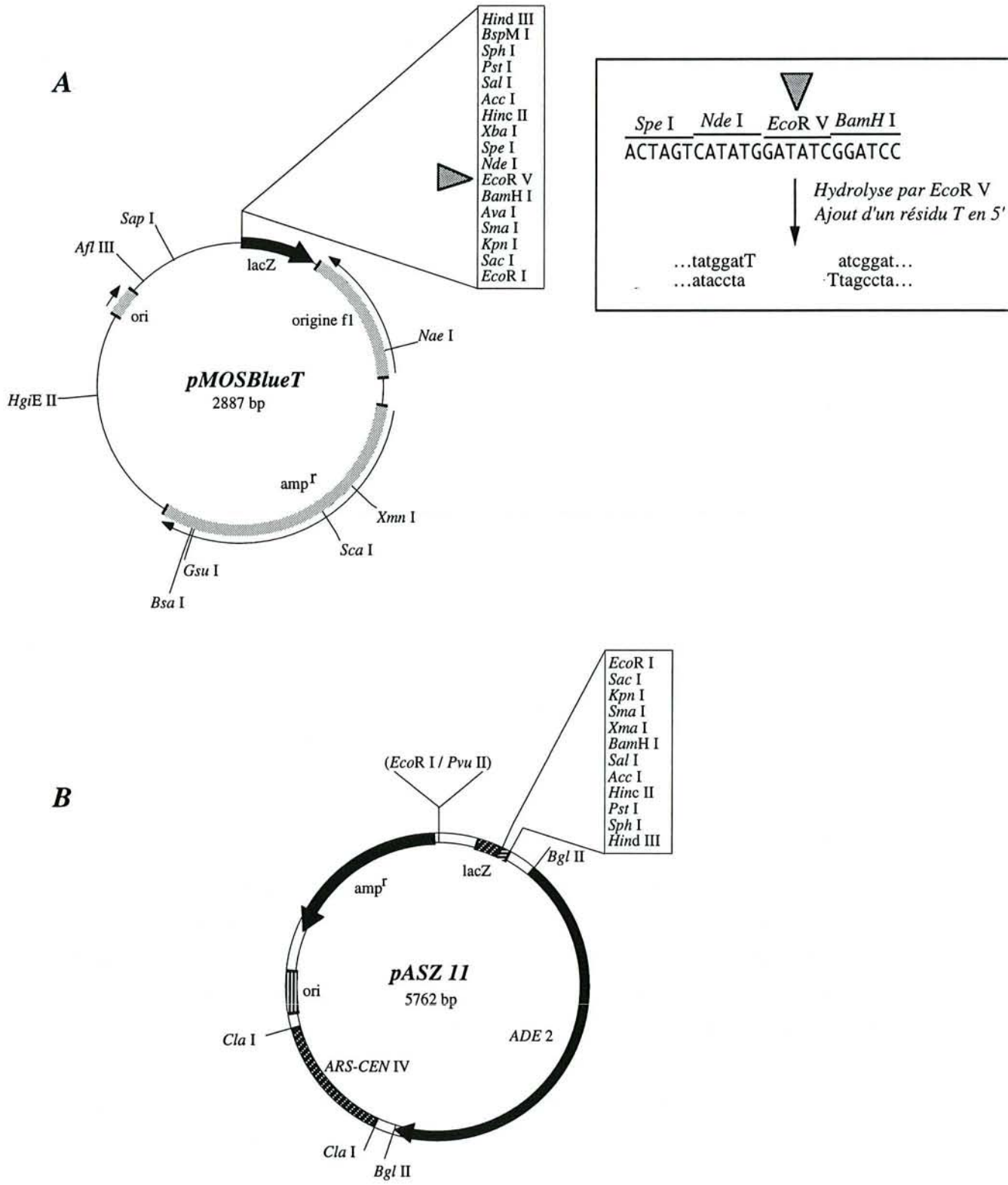


Figure 30 : A. Carte de restriction du plasmide pMOSBlueT (kit RPN1719 Amersham).
Le triangle gris indique le site EcoR V.
L'encart montre, en lettres minuscules, les extrémités du plasmide pMOSBlue-T commercialisé par Amersham.
Les sites de restriction de la casette de clonage sont indiqués.

B. Carte de restriction du plasmide pASZ11 (Stotz et Linder, 1991).
Les sites de restriction de la casette de clonage sont indiqués.

pFL1 (Chevalier *et al.*, 1980) (Fig. 31A)

Ce plasmide est un plasmide navette *E. coli /S. cerevisiae* de 7661 pb obtenu par insertion de l'origine de réplication du plasmide 2μ de *S. cerevisiae* dans le plasmide pBR322 Ce plasmide peut se répliquer à la fois chez *E. coli* et chez *S. cerevisiae*. Il contient le gène *ura3* servant de marqueur de sélection.

pFL1::snR17A (Myslinski *et al.*, 1990) (Fig. 31B)

Ce plasmide contenant le gène du snoRNA U3A avait été sélectionné à partir d'une banque génomique de la levure *S. cerevisiae*, fournie par F. Lacroute (IBMC, Strasbourg, France). Le fragment contenu dans ce plasmide comporte 8 kb.

D4.2 Bactériophages.

M13mp9, M13mp18, M13mp19 (Messing *et al.*, 1977)

Ils ont été obtenus par modification du phage M13 (Messing *et al.*, 1977). Ces phages renferment la même cassette que les plasmides pUC permettant l'insertion d'une séquence d'ADN dans le fragment du gène de la β-galactosidase. Les phages M13mp9, M13mp18 et M13mp19 diffèrent par la nature ou l'orientation de la cassette contenant les sites uniques de restriction. De plus, le phage M13mp9 contient deux mutations de type ambre dans les gènes C1 et C3.

M13mp9T7::snR17A (Ségault *et al.*, 1992)

Ce bactériophage a été obtenu en deux étapes. Tout d'abord, j'ai inséré la séquence reconnue par l'ARN polymérase du phage T7 par mutagenèse dirigée entre les sites de reconnaissance des nucléases *Sal* I et *Sma* I au niveau de la cassette de clonage du phage M13mp9. Puis, le fragment *Sal* I-*Hpa* I du plasmide recombinant pFL1::snR17A (Myslinski *et al.*, 1990), débutant à l'extrémité 5' de la région codante du gène du snoRNA U3A et se terminant à son extrémité 3' a été inséré dans le phage M13mp9T7 entre les sites *Sal* I et *Sma* I (Ségault *et al.*, 1992).

D4.3 Phagemides.

Bluescript KS + ou - et Bluescript SK + ou - (Short *et al.*, 1988) (Stratagene Inc; CA)

Ces plasmides de 2,96 kb possèdent de nombreux points communs avec les plasmides de type pUC :

* le gène de résistance à l'ampicilline,
* l'origine de réplication du plasmide pBR322,
* le fragment β-galactosidase d'*E.coli* permettant l'α-complémentation,

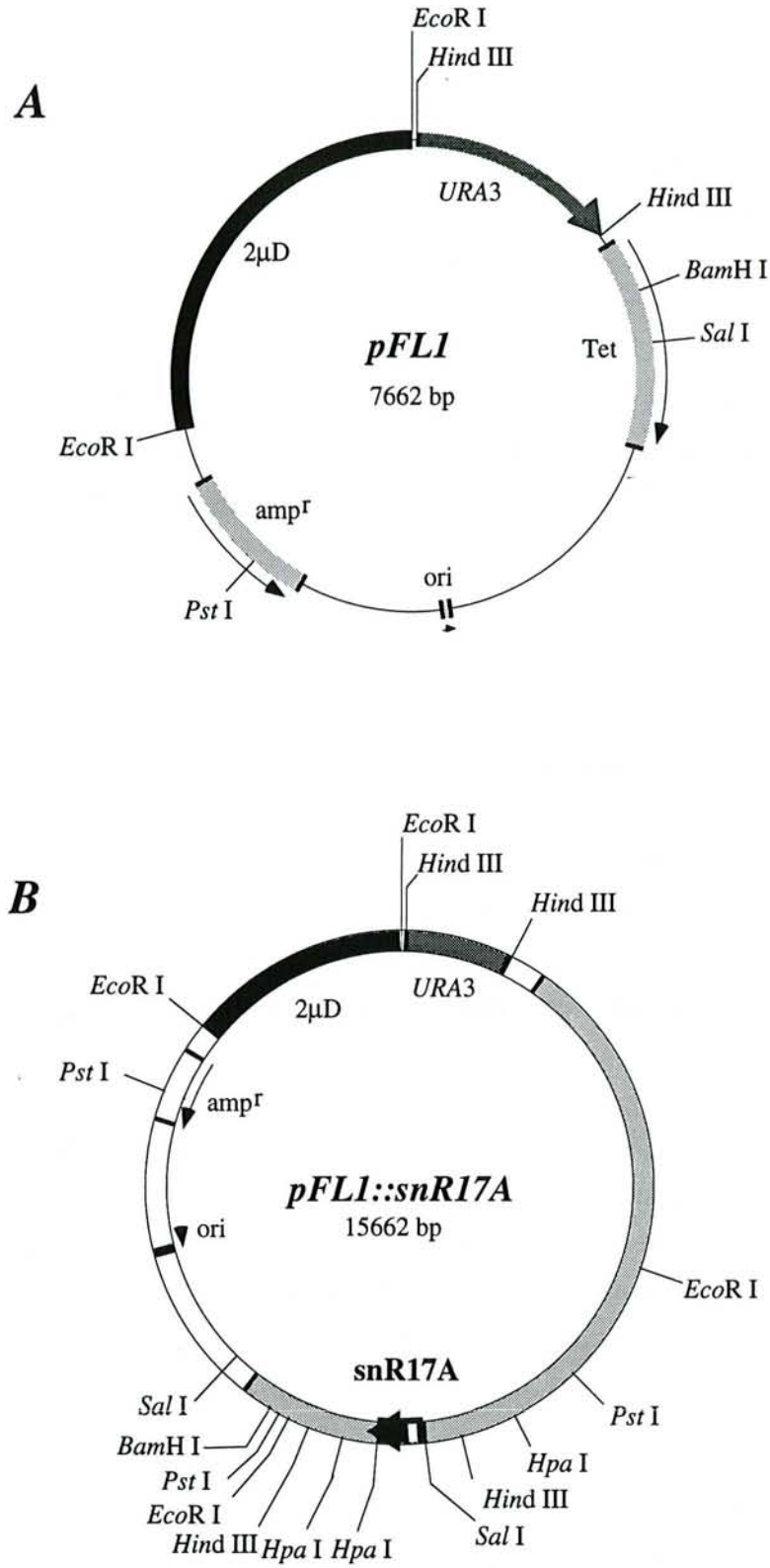


Figure 31 : A. Carte de restriction du plasmide pFL1 (Chevallier *et al.*, 1980).

B. Carte de restriction du plasmide recombinant pFL1::snR17A (Myslinski *et al.*, 1990).

Le fragment en gris clair est issu de la banque d'ADN de *S. cerevisiae.* La région codant pour le snoRNA U3A est indiquée (en noir, les exons, en balnc, l'intron).

* une cassette de clonage présentant plus de sites uniques de restriction que la cassette de clonage des plasmides pUC18 et pUC19 (les bluescripts SK et KS diffèrent par l'orientation de cette cassette de clonage).

Ces plasmides contiennent aussi :

* des promoteurs de transcription reconnus par les ARN polymérases des phages T3 et T7 d'*E.coli*,

* une séquence dérivée du phage M13 permettant la synthèse et l'encapsidation du simple brin (sb) en présence d'un phage "helper" M13KO7 (Vieira et Messing, 1987).

D5. Tampons et Solutions d'usage courant.

TE : EDTA 1 mM, Tris HCl 10mM, pH 8,0

TBE 1x : acide borique 89mM, EDTA 2mM, Tris borate 89mM, pH 8,3

Solution d'acrylamide dénaturante : Acrylamide/Bisacrylamide (23,75/1,25) 25% (poids/vol); TBE 1x; Urée 8M.

Tampon de dépôt agarose : EDTA 125mM, SDS 2%, glycérol 25%, Bleu de bromophénol 0,12%

Tampon de dépôt bleu urée: EDTA 1mM, urée 7M, Bleu de Bromophénol 0,025%, Bleu de Xylène Cyanol 0,025%, citrate de sodium 20mM pH 5.

Tampon de dépôt bleu/formamide : EDTA 20mM, Bleu de Bromophénol 0,05%, Bleu de Xylène Cyanol 0,05% dans de la formamide désionisée.

E. Méthodes.

E1. Méthodes relatives à l'ADN.

Les techniques classiques de préparation d'ADN plasmidique, d'hydrolyse par des enzymes de restriction, de séparation par électrophorèse, d'hybridation par la méthode de Southern, de transformation de cellules bactériennes sont celles décrites dans le manuel de laboratoire de Sambrook *et al.* (1989).

E1.1 Préparation d'ADN génomique de levure.

Deux méthodes de préparation de l'ADN génomique de levure ont été utilisées, une méthode de préparation à grande échelle nécessitant la formation de sphéroplastes par digestion de la paroi cellulaire par la cytohélicase avant l'extraction de l'ADN. La seconde méthode est beaucoup plus rapide, l'extraction de l'ADN est directe.

E1.11 Technique basée sur la formation de sphéroplastes de levures (Campbell et Duffus, 1988).

Formation des sphéroplastes.

Une préculture de levure est réalisée en milieu YPG à partir d'une souche de levure conservée à -80°C, avec une incubation sous agitation à 30°C pendant 12h. Cette préculture est ensuite utilisée pour ensemencer au 1/100ème, 100 ml de milieu, la culture est réalisée dans les mêmes conditions que la préculture. Après centrifugation 10 min à 4 000 tours/min à 4°C (Beckman J6), les cellules sont lavées trois fois dans 20 ml d'eau distillée. Un gramme de cellule est remis en suspension dans 4 ml de tampon (EDTA 25mM, DTT 1mM) et incubé 30 min à 30°C. Les cellules sont à nouveau centrifugées, puis reprises dans 4 ml de tampon isotonique (KCl 1,4 M, acide citrique 50mM, Na_2HPO_4 150mM, pH 5,8), contenant 25 mg de cytohélicase (IBF Biotechnics), l'incubation est réalisée à 37°C.

La formation des sphéroplastes est vérifiée par mesure de la densité optique à 800 nm (DO_{800nm}) de 10 µl de suspension cellulaire diluée dans 990 µl d'eau d'un côté, et de 10 µl de suspension cellulaire diluée dans 990 µl de sorbitol 1M. Dilués dans l'eau, les sphéroplastes éclatent, la solution de sorbitol 1 M étant hypertonique, elle permet aux cellules dont la paroi cellulaire est fragilisée de conserver leur intégrité. Lorsqu'un rapport de 2 entre les deux densités optiques mesurées est atteint, les sphéroplastes peuvent être utilisés pour préparer de l'ADN génomique de bonne qualité.

Les sphéroplastes ainsi formés sont centrifugés 5 min à 1 000 tours/min à 4°C et lavés dans une solution de sorbitol 1M.

Extraction de l'ADN génomique des sphéroplastes de levures (Cryer *et al.*, 1975).

Le culot de sphéroplastes est repris dans 7,5 ml de tampon TE qui est ensuite additionné de 1,5 ml de tampon (EDTA 250mM, SDS 2,5% (p/v), Tris-HCl 500mM pH 8,5), le tout est incubé à 65°C pendant 30 min, puis 2 ml d'acétate de potassium 5M sont ajoutés, l'incubation est poursuivie à 0°C pendant 1h.

Les débris cellulaires sont éliminés par centrifugation 15 min à 20 000 tours/min (Beckman J21) à 4°C. La précipitation des acides nucléiques est réalisée par ajout de 30 ml d'isopropanol. Le précipité est lavé à l'éthanol à 70%, ensuite repris dans 7,5 ml de tampon TE additionné de RNase A à la concentration finale de 50 mg/ml, et incubé 30 min à 37°C. L'ADN est à nouveau précipité avec un volume égal d'isopropanol, centrifugé 10 min à 10 000 tours/min à 4°C puis repris dans 1 ml de tampon TE.

E1.12 Préparation rapide d'ADN génomique de levure

3 ml de milieu YPG sont ensemencés avec une colonie de levure, et incubés à 30°C sous agitation jusqu'en phase stationnaire. Les cellules sont ensuite centrifugées, remises en suspension dans 500 µl d'eau, et à nouveau centrifugées. Le culot cellulaire est repris dans 200 µl d'une solution Triton (Triton X-100 2% (p/v), SDS 1% (p/v), NaCl 100mM, EDTA 1mM, Tris-HCl 10mM pH8), à cette solution sont ajoutés 200 µl d'un mélange phénol/chloroforme/alcool isoamylique (25/24/1) (v/v/v), et 300 mg de billes de verre (Ø = 0,45-0,50 mm). Le tout est vigoureusement agité pendant 3 min. 200 µl de tampon TE sont ensuite ajoutés, avant centrifugation pendant 5 min à 13 000 tours/min. Les acides nucléiques sont précipités deux fois à l'éthanol. Le précipité est dissous dans 50 µl de tampon TE, 5 µg de RNase A sont ajoutés, l'ensemble est incubé 30 min à 37°C. Les protéines sont éliminées par une extraction phénolique, suivie d'une extraction au chloroforme, l'ADN est à nouveau précipité à l'éthanol, puis dissous dans 100 µl de tampon TE et conservé à 4°C.

E1.2 Amplification d'ADN (Innis *et al.*, 1988).

E1.21 Principe de la méthode.

L'amplification d'ADN par polymérisation en chaîne ou PCR (pour Polymerase Chain Reaction) est utilisée pour isoler un fragment d'ADN dont, en général, les séquences bordantes sont connues. Deux oligonucléotides correspondant aux séquences de part et d'autre du segment à amplifier sont utilisés contre amorce d'une ADN polymérase. Chaque oligonucléotide est complémentaire à un brin d'ADN. Cette technique permet de multiplier exponentiellement le nombre de copies de fragment d'ADN, en utilisant une ADN polymérase thermostable extraite de la bactérie *Thermus aquaticus* (ADN polymérase *Taq*) ou *Thermus thermophilus* (ADN polymérase *Tth*).

Le principe de la réaction est le suivant, l'ADN double brin est dénaturé par chauffage à 94°C, puis refroidissement à 55°C, ce qui permet l'hybridation des deux oligonucléotides sur leur

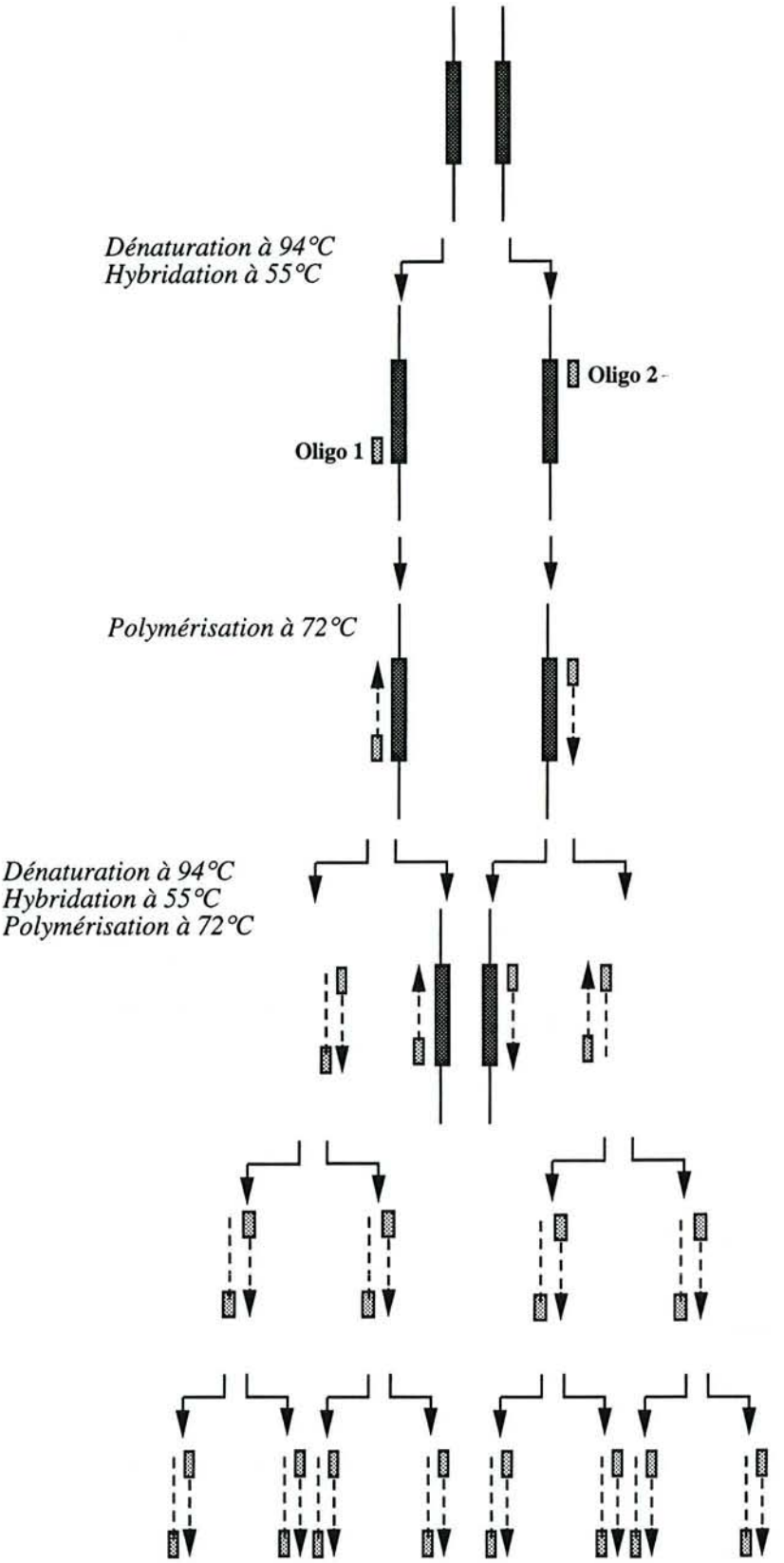


Figure 32 : Schéma du principe d'amplification d'ADN par la technique PCR.

A chaque cycle d'amplification, le nombre de fragments d'ADN double-brin est doublé, si aucun composant du mélange réactionnel n'est limitant.

ADN matrice

Amorce oligonucléotidique

ADN nouvellement synthétisé

séquence cible. La troisième étape de la réaction est l'élongation par l'ADN polymérase *Taq* des deux amorces en présence des 4 dNTP, l'élongation est réalisée à 72°C. Le nombre de copies d'ADN est doublé à chaque cycle de dénaturation, hybridation et élongation (Fig. 32).

E1.22 Protocole d'amplification.

Composition du mélange réactionnel :

1 µg d'ADN génomique
10 pmol de chaque oligonucléotide amorce
1,25mM de chaque dNTP
5 µl de tampon 10x (KCl 500mM, $MgCl_2$ 20mM, DTT 10mM, Tris-HCl 200mM pH 8,4)
2 U d'ADN polymérase *Taq* (Gibco-BRL)
H_2O qsp 50 µl.

L'amplification est réalisée dans l'appareil ThermoCycler (de Cetus Perkin-Elmer), avec le programme suivant :

1er cycle	: 3 min à 94°C
30 cycles	: 1 min à 55°C
	2 min à 72°C
	1 min à 94°C
Dernier cycle	: 10 min à 72°C
	10 min à 4°C.

Le dernier cycle de 10 min à 72°C sert à terminer l'élongation de toutes les chaînes en cours de synthèse.

E1.3 Marquage des sondes d'ADN.

Plusieurs techniques ont été employées :

* marquage à l'extrémité 5',
* marquage par la technique de translation de coupure,
* marquage par PCR.

E1.31 Marquage de l'extrémité 5'.

Cette technique consiste à ajouter un P^{32} en 5' d'un oligonucléotide de synthèse. Ce groupement est transféré de la position γ de l'ATP[γ-^{32}P] à l'extrémité 5' de l'oligonucléotide, par la polynucléotide kinase du bactériophage T4 :

50 ng d'oligonucléotide (soit 7 pmoles pour un oligonucléotide de 20 mer)
30 µCi d'ATP[γ-^{32}P] (soit 1110 Bq)

1 U de polynucléotide kinase du bactériophage T4

1 µl tampon 10x (Tris-HCl 500mM, DTT 50mM, $MgCl_2$ 100mM, spermidine 1mM, EDTA 1mM)

H_2O qsp 10 µl.

Le mélange est placé à 37°C pendant 30 min. L'oligonucléotide est ensuite purifié sur gel de polyacrylamide dénaturant à 15%, urée 8M afin d'éliminer l'excès d'ATP[γ-^{32}P].

Cette méthode est également utilisée pour phosphoryler en 5' les oligonucléotides utilisés dans les expériences de mutagenèse dirigée, la présence d'un phosphate à l'extrémité 5' de l'oligonucléotide étant nécessaire à la ligation de l'oligonucléotide avec le fragment d'ADN nouvellement synthétisé. La réaction se fait comme ci-dessus, mais en remplaçant l'ATP radioactif par 10mM d'ATP froid.

E1.32 Marquage par translation de coupure.

Ce type de marquage peut être réalisé sur un ADN double-brin circulaire ou linéaire et produit des sondes plus radioactives que celles obtenues par la technique précédente, du fait de l'incorporation de plusieurs nucléotides marqués. Le principe est l'incorporation par l'ADN polymérase I d'*E.coli* d'un nucléotide marqué après avoir statistiquement coupé un des brins de l'ADN double-brin par la DNase I. Les fragments d'ADN ainsi marqués sont dénaturés par la chaleur et utilisés dans des expériences d'hybridation d'ADN par la méthode de Southern ou d'ARN par la méthode de Northern.

Nous avons utilisé le kit commercialisé par Boehringer (Kit 976 776)

0,1 µg ADN

20 µCi dATP[α-^{32}P] (740 Bq) (3000 Ci/mmol)

1,2 mM de chacun des autres dNTP

2 µl d'un mélange contenant de l'ADN polymérase I et de la DNase I

(le nombre d'unités n'est pas donné par le fournisseur)

2 µl tampon 10x (la composition n'est pas donnée par le fournisseur).

Le mélange est placé 35 min à 15°C et la réaction est stoppée par chauffage à 65°C pendant 10 min. L'excès de nucléotides marqués non incorporés est éliminé par deux précipitations successives à l'éthanol.

E1.33 Marquage par la méthode PCR.

Des fragments d'ADN peuvent être marqués radioactivement au cours de l'amplification par incorporation d'un nucléotide marqué, tel que le dATP[α-^{32}P], ou par utilisation d'un oligonucléotide amorce marqué à son extrémité 5'. Le protocole d'amplification est celui décrit ci-dessus.

E1.4 Séquençage d'ADN (Sanger *et al.*, 1977; Tabor et Richardson, 1987).

E1.41 Principe du séquençage d'ADN.

L'ADN peut être séquencé soit sous forme simple-brin, soit sous forme double-brin. La méthode de séquençage a été décrite par Sanger *et al.* (1977) pour le séquençage d'ADN simple-brin en utilisant le fragment de Klenow de l'ADN polymérase I d'*E. coli*. Cette méthode a été adaptée par Tabor et Richardson (1987) pour séquencer l'ADN double-brin ou simple-brin avec l'ADN polymérase du bactériophage T7. Un oligonucléotide de synthèse est hybridé sur le fragment d'ADN à séquencer, il sert d'amorce à l'enzyme. Dans les deux cas, les fragments synthétisés par l'enzyme sont marqués radioactivement par l'incorporation de dATP[α-^{32}P] lors de l'étape d'élongation, ce qui permet la visualisation de ces fragments.

L'élongation des fragments est stoppée par incorporation de didésoxyribonucléotides (ddNTP), la quantité des désoxyribonucléotides (dNTP) par rapport aux ddNTP est calculée en fonction de la longueur et de la composition en bases du fragment à séquencer, afin qu'il y ait un arrêt statistique des chaînes.

E1.42 Méthode de séquençage d'ADN.

Le séquençage d'ADN db nécessite une étape préalable de dénaturation qui n'est pas nécessaire lors du séquençage d'ADN simple-brin. La réaction de séquençage en elle-même est identique dans les deux cas.

Une étape préalable de dénaturation de l'ADN double-brin circulaire à séquencer est nécessaire. 3 µg de l'ADN dissous dans 10 µl de tampon dénaturant (NaOH 0,4N) sont incubés 5 min à température ambiante. Après addition de 8 µl d'acétate d'ammonium 5M pH 7,5, l'ADN est précipé à l'éthanol, puis repris dans 5 µl H_2O MilliQ.

L'hybridation entre l'ADN à séquencer et l'amorce oligonucléotide est réalisée dans le mélange suivant :

5 µl d'ADN db dénaturé ou 1 µg d'ADN simple-brin
2 µl de tampon 5x ($MgCl_2$ 50mM, NaCl 250mM, Tris-HCl 200mM pH 7,5)
5 à 10 ng d'oligonucléotide (0,71 à 1,4 pmoles d'une amorce de 20 nucléotides)
H_2O qsp 10 µl.

Après incubation 5 min à 80°C, le mélange est lentement ramené à température ambiante avant d'ajouter :

5 µCi dATP[α-^{32}P] (185 Bq) (800 Ci/mmol; 1 mCi/ml)
1 µl DTT 100mM

2U de l'ADN polymérase du bactériophage T7

2 µl mélange d'extension 1x (mélange d'extension 5x : dCTP 7,5µM, dGTP 7,5µM, dTTP 7,5µM)

Après une incubation de 5 min à température ambiante, 3,5 µl de mélange sont transférés dans un microtube contenant 2,5 µl du mélange ddATP / dATP ou 2,5 µl de ddCTP / dCTP ou 2,5 µl de ddGTP / dGTP ou 2,5 µl de ddTTP / dTTP (ddNTP / dNTP : 20 µM / 200 µM). Le tout est placé à 37 °C pendant 5 min, la réaction est arrêtée en ajoutant 4 µl de bleu de formamide.

Les échantillons sont alors chauffés à 100°C pendant 3 min et fractionnés sur gel de polyacrylamide dénaturant (polyacrylamide 6%, urée 8M). Après migration, le gel est trempé dans une solution d'acide acétique 10% (p/v) afin de fixer l'ADN. Les fragments d'ADN sont ensuite révélés par autoradiographie.

E1.5 Mutagenèse dirigée par la méthode de Krämer (Krämer *et al.*, 1984).

E1.51 Descritif de la méthode.

Cette technique est basée sur la formation d'un hétéroduplex entre un brin du phage M13mp19 double-brin coupé par les enzymes de restriction *Hin*d III et *Eco*R I (de telle façon que la cassette de clonage soit éliminée) et le brin opposé du phage M13mp9 contenant la séquence à muter. La séquence à muter se trouve ainsi sous forme d'une fenêtre simple-brin sur laquelle l'oligonucléotide complémentaire, servant à réaliser la mutation, peut être hybridé. Le fragment Klenow de l'ADN polymérase I d'*E.coli* et l'ADN ligase du bactériophage T4 permettent ensuite de recréer un phage double-brin complet. La sélection du phage portant la séquence mutée est réalisée par transformation d'une souche d'*E.coli* HB2154 qui ne possède pas d'ARNt suppresseur et qui est mutée dans son système général de réparation d'ADN, en présence d'un tapis cellulaire d'*E.coli* HB2151. Le phage M13mp9 portant sur deux gènes essentiels des codons ambres, ce simple brin ne peut donc donner lieu à des virus dans la souche d'*E.coli* HB2154. L'emploi de la souche d'*E.coli* HB2151, qui elle n'est pas mutée dans son système de réparation, permet de conserver la séquence correcte du phage simple-brin M13mp19 portant le gène muté, qui seul donne naissance à des phages infectieux (Fig. 33).

E1.52 Protocole.

Mélange M1 :

1 µg M13mp19 db coupé par *Hin*d III et *Eco*R I

2,5 µg M13mp9 sb portant le fragment d'ADN à muter

4 µl tampon Mut 5x (KCl 937,5mM, Tris-HCl 62,5mM pH 7,5)

H_2O qsp 20 µl

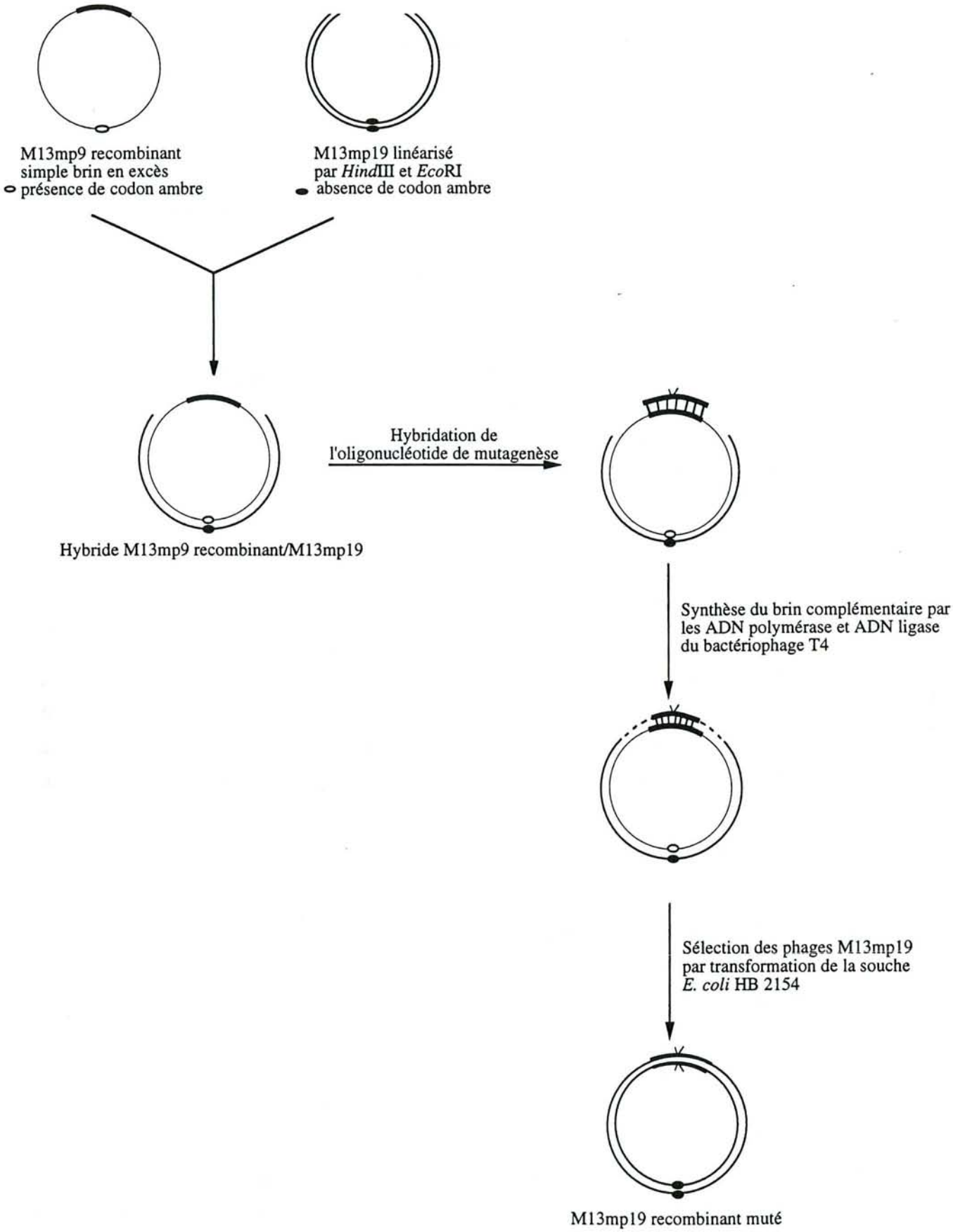


Figure 33 : Principe de la mutagenèse dirigée selon la méthode de Kramer (Kramer *et al.*, 1984).

Après incubation à 100°C pendant 3 min, puis à 65°C pendant 5 min, le mélange est ramené lentement à température ambiante.

La présence de l'hétéroduplex est vérifiée sur gel d'agarose à 0,8% (p/v).

Le mélange M2 est réalisé en ajoutant au mélange M1, l'oligonucléotide de synthèse préalablement phosphorylé en 5' par la polynucléotide kinase du bactériophage T4. Le rapport molaire entre le phage M13mp9 simple-brin et l'oligonucléotide est de 1/20ème. Le tout est incubé 5 min à 55°C.

La dernière étape consiste en une incubation d'une nuit à 15°C du mélange suivant :

mélange M2 (mélange M1 + oligonucléotide phosphorylé)
1 µl tampon ($MgCl_2$ 100mM, DTT 10mM, Tris-HCl 200mM pH 8,5)
1 µl de chaque dNTP 10mM
1 µl ATP 10mM
2,5 U du fragment de Klenow
1,5 U d'ADN ligase du bactériophage T4
1 µl DTT 100mM
H_2O qsp 40 µl

Le lendemain, l'incubation est prolongée 2 h à température ambiante en présence d'une unité supplémentaire du fragment de Klenow et de l'ADN ligase du bactériophage T4, 100 µl de cellules compétentes de la souche d'*E.coli* HB2154 sont alors transformés par 1 µl de ce mélange final, en présence de 100 µl de cellules d'*E.coli* HB2151 en phase stationnaire. Les ADN des phages contenus dans une dizaine de plages de ralentissement de croissance sont préparés par la technique de minipréparation de phages, et la présence de la mutation est recherchée par séquençage.

E1.6 Transformations de levures.

E1.61 Préparation des levures compétentes.

Un volume de 100 ml de milieu YPGlu est ensemencé au 1/1000ème à partir d'une préculture de *S. cerevisiae* en phase stationnaire de croissance. Lorsque la DO à 600nm de la culture atteint une valeur de 0,5, les cellules sont centrifugées 5 minutes à 4 000 tours/min (Beckman, J6). Après lavage avec 50 ml de Tris HCl 10mM pH 7,5, les cellules sont recentrifugées dans les mêmes conditions. Le culot est repris dans 25 ml de tampon LiT (acétate de lithium 100 mM, Tris HCl 10 mM pH7,5), et additionné de 0,25 ml de DTT 1 M. Les cellules sont ensuite incubées 40 minutes à température ambiante sous une faible agitation avant d'être centrifugées à nouveau. Le culot est alors repris dans 750 µl de tampon LiT, additionné de 7,5 µl de DTT 1 M. Les cellules sont alors compétentes et prêtes à être transformées.

E1.62 Transformation des levures compétentes.

A 100 µl de cellules compétentes sont ajoutés 50 µl de tampon LiT, 5 µl d'ADN de sperme de saumon à 10 mg/ml (traité par sonication, puis dénaturé 10 minutes à 100°C), environ 1 µg d'ADN plasmidique et 300 µl d'une solution de PEG8000 à 50% dans le tampon LiT. Après 10 minutes d'incubation à température ambiante, 50 µl de DMSO sont ajoutés. Les cellules sont alors incubées 15 minutes à 42°C, puis centrifugées 10 secondes dans une microcentrifugeuse. Les cellules sont lavées dans de l'eau stérile, et le culot est repris dans 100 µl d'eau et les cellules sont étalées sur un milieu sélectif, où seules les cellules transformées pourront se développer.

E1.63 Test de la fonctionnalité de gènes de snoRNA U3.

Ce test consiste à utiliser un crible génétique dont le principe est le suivant : la souche *S. cerevisiae* JH84 (Hughes et Ares, 1991) a été utilisée comme souche test. Cette méthode a été développée au laboratoire par R. Fournier. Dans la souche test, le gène codant pour le snoRNA U3B est interrompu par insertion du gène LEU2, le gène codant pour le snoRNA U3A est placé sous le contrôle de l'opérateur Gal10. Sur un milieu contenant du galactose comme source de carbone, le gène codant pour le snoRNA U3A est exprimé, la souche présente une croissance normale. Par contre, sur milieu contenant du glucose comme source de carbone, l'expression du gène codant pour le snoRNA U3A est réprimée, aucun snoRNA U3 n'est produit, la croissance de la souche JH84 est arrêtée. Pour pallier à ce défaut de croissance en milieu glucose, un gène codant pour le snoRNA U3 peut être apporté en trans par transformation par un plasmide portant ce gène. Pour cela, R. Fournier a introduit dans le plasmide navette centromérique pASZ11, un fragment *Hin*d III de 1,5 kb du chromosome de *S. cerevisiae* contenant le gène du snoRNA U3A et ses régions flanquantes (pASZ11::U3A). La souche *S. cerevisiae* JH84 a été transformée par ce plasmide recombinant. Sur milieu glucose, la souche recombinante présente une croissance normale.

Des échanges de régions codantes du gène du snoRNA U3A avec les régions codantes de gènes hétérologues ou des régions codantes portant des mutations ont ensuite été réalisés dans le plasmide recombinant, et la fonctionnalité de ses gènes hétérologues et des gènes mutés a été testée *in vivo*, selon le protocole suivant.

¤ Cultures de la souche non transformée.

La souche *S. cerevisiae* JH84 est d'abord étalée sur milieu solide sélectif YNB/Gal/His/Ade, puis cultivée en milieu YPGal sous agitation à 30°C.

¤ Cultures des souches transformées : répression de l'expression du gène du snoRNA U3A par le glucose.

La souche *S. cerevisiae* JH84 transformée par le plasmide pASZ11 ou un de ses dérivés est cultivée pendant 24 heures dans 20 ml de milieu YPGal. Après centrifugation, les cellules sont

lavées deux fois avec du milieu YPGlu. Les cellules sont collectées par une centrifugation de 2 minutes à 4 000 tours/min (Beckman, J6). Une fraction des cellules est reprise dans 20 ml de milieu YPGlu et sert à ensemencer 100 ml de milieu YPGlu préchauffé à une DO à 600nm de 0,5 unité. La culture est placée sous agitation à 30°C. Le reste des cellules est lavé avec du sorbitol 1M froid, et est congelé à -80°C. Cette fraction constituera le prévèvement au temps T_0.

Après 5 heures d'incubation, la DO à 600 nm de la culture est mesurée, et une fraction de la culture est diluée dans du milieu YPGlu préchauffé de façon à obtenir 100 ml de suspension cellulaire à une DO à 600 nm de 0,5. Le reste des cellules est récolté par centrifugation, puis traité comme ci-dessus avec du sorbitol 1M froid. Cette fraction constituera le prévèvement au temps T_5. Les 100 ml de culture dilués sont incubés dans les mêmes conditions que précédemment. Après 10, puis encore après 24 heures d'incubation, la même opération de dilution et collecte de cellules est réalisée (temps T_{10} et T_{24}). Enfin, après 36 heures d'incubation, la DO à 600 nm est mesurée, et la totalité des cellules est récoltée (T_{36}).

Les mesures de DO à 600 nm réalisées avant chaque dilution permettent d'établir une courbe de croissance de la souche transformée. Les ARN totaux des cellules récoltées à chaque étape de dilution sont extraits, le contenu en ARN est analysé par la méthode d'hybridation de Northern.

E2. Méthodes relatives à l'ARN.

E2.1 Préparation et fractionnement des ARN de levure en fonction de leur constante de sédimentation.

E2.11 Préculture et culture.

Une préculture est obtenue en ensemençant 10 ml de milieu YPGlu avec une colonie de levure. Le tout est incubé, sous agitation, 10h à 30°C, 500 ml de milieu YPGlu sont ensemencés au 1/100ème avec la préculture, puis incubés à 30°C, sous agitation, jusqu'à saturation. La culture est centrifugée pendant 10 min à 4 000 tours/min. Le culot cellulaire est lavé dans 10 ml de tampon d'extraction (EDTA 10mM, SDS 1%, AcNa 50 mM pH 5,3), la suspension est à nouveau centrifugée 10 min à 4 000 tours/min.

E2.12 Extraction des ARN.

Après centrifugation, les cellules sont reprises dans 10 ml de tampon d'extraction (EDTA 10mM, SDS 1%, AcNa 50 mM pH 5,3). Les ARN sont extraits en présence de 10 ml de phénol, préalablement équilibré avec le tampon (EDTA 10mM, AcNa 50mM pH 5,3) pendant 4 min à 65°C. Le tout est refroidi brutalement dans un mélange glace-éthanol, puis incubé 10 min dans ce mélange et centrifugé pendant 10 min à 4 000 tours/min. La phase aqueuse est prélevée puis réextraite avec 3ml de phénol, puis à nouveau avec 4 ml d'un mélange phénol-chloroforme (vol/vol). Les ARN

sont précipités en présence d'AcNa pH5,3 et de 2,5 ml d'éthanol pendant 1h à -20°C. Après une centrifugation de 15 min à 4 000 tours/min, le précipité est lavé à l'éthanol 70%, séché et repris dans 250 µl d'eau.

E2.13 Fractionnement des ARN en fonction de leur constante de sédimentation.

15 à 20 mg d'ARN totaux sont déposés à la surface d'un gradient continu de saccharose 10-30%, établi dans le tampon (NaCl 100mM, Tris-HCl 10 mM pH7,5). La centrifugation est effectuée à 27 000 tours/min pendant 24h à 4°C (ultracentrifugeuse Beckman L8M, rotor SW 28).

Les fractions du gradient sont récoltées à partir du fond du tube à l'aide d'une pompe péristaltique. La DO à 280 nm est lue directement grâce à un spectrophotomètre UV couplé à un enregistreur et à un collecteur de fractions (fractions de 0,5 ml). Les fractions contenant les ARN de constante de sédimentation 4S à 8S sont rassemblées, la concentration saline est ajustée à 0,3 M en acétate de sodium et 3 volumes d'éthanol absolu sont ajoutés. Les ARN sont précipités pendant 1h à -20°C, puis centrifugés.

Cette méthode permet de préparer une grande quantité d'ARN dont la constante de sédimentation est comprise entre 4 et 8S.

E2.2 Préparation rapide d'ARN totaux de levures.

Une culture de 5 ml en milieu YPGlu est réalisée, à partir d'une colonie de levure, sous agitation, pendant 10h à 30°C. Les cellules sont récoltées par centrifugation 30 secondes à 13 000 tours/min, lavées avec 300 µl d'H_2O. Après centrifugation, les cellules sont reprises dans 300 µl de tampon d'extraction (SDS 1% (p/v), EDTA 10 mM, acétate de sodium 50 mM pH 5), 300 µl de phénol saturé (acétate de sodium 50 mM pH 5, EDTA 10 mM) et 200 µg de billes de verre (Ø= 0,45-0,50mm) sont ajoutés. L'ensemble est vigoureusement agité pendant 2 min, puis placé 2 min dans la glace, à nouveau agité 2 min, et enfin incubé 5 min à 65°C. Un refroidissement brutal est réalisé en plaçant le microtube dans un mélange glace/éthanol jusqu'à formation de cristaux. La phase aqueuse du mélange prélevée après centrifugation pendant 10 min à 13 000 tours/min, est extraite au chloroforme, puis précipitée à l'éthanol. Les ARN sont repris, après lavage à l'éthanol à 70%, dans 100 µl d'H_2O.

Cette méthode permet d'obtenir environ 500 µg d'ARN. Elle est utilisée pour extraire les ARN des cellules récoltées à chaque étape de dilution lors des tests de la fonctionnalité de gènes de snoRNA U3 décrits ci-dessus.

E2.3 Hybridation de l'ARN par la technique d'hybridation de Northern.

Les ARN totaux sont fractionnés par électrophorèse en gel d'acrylamide dénaturant 15%, urée 8M, TBE 1x, suivi d'un transfert sur membrane et hybridation des ARN avec les oligodésoxynucléotides sondes marqués au ^{32}P.

E2.31 Transfert des ARN sur membrane de nylon.

Le transfert d'ARN sur membrane de nylon chargée positivement (Biodyne, Pall) peut être réalisé par deux méthodes différentes, le transfert passif ou l'électrotransfert.

Transfert passif.

Le transfert est réalisé passivement avec du tampon SSC 20x (NaCl 3M; citrate de sodium 0,3M). La membrane est directement en contact avec le gel d'électrophorèse entre des feuilles de papier Whatman 3MM. Les ARN sont ensuite fixés sur la membrane par exposition aux rayonnements UV (312 nm) pendant 4 min.

Electrotransfert.

Après électrophorèse, la membrane de nylon chargée positivement (Biodyne, Pall) et préalablement trempée dans du tampon TBE 0,5X est placée sur le gel. L'ensemble est inséré entre 2x3 feuilles de papier Whatman 3MM imbibées de TBE 0,5X. L'électrotransfert a lieu dans un appareil Transblot, Biorad, pendant 30 min à 3 mA/cm^2. Les ARN sont ensuite fixés sur la membrane par exposition aux rayonnements UV (312 nm) pendant 4 min.

E2.32 Préhybridation et hybridation.

La membrane est préhybridée dans un sac de plastique scellé contenant 5 à 10 ml de tampon de préhybridation :

7,5 ml de tampon SSC 20x (NaCl 3M; citrate de sodium 0,3M)
1,25 ml de Denhardt's 100x (Ficoll 2% ; Polyvinylpyrrolidone 2% ; BSA 2%)
0,5 ml de SDS 10%
0,31 ml d'EDTA 0,2M
H_20 qsp 25 ml

La préhybridation a lieu pendant 2 heures à 37°C.

Le tampon d'hybridation est le même que le tampon de préhybridation. Il contient en plus la sonde marquée au ^{32}P (environ 3000 cpm/ml). L'hybridation est réalisée à 37°C pendant au moins 2 heures sous agitation.

E2.33 Lavage de la membrane.

La membrane de nylon est lavée 3 fois dans 20 ml de tampon SSC 6X, 0,1% SDS, pendant 15 min à 42°C. Elle est ensuite placée en autoradiographie ou en fluorographie à -80°C pendant 4 à 12 heures au contact d'un film Fuji X-Ray.

E2.4 Transcription *in vitro*.

E2.41 Principe.

L'ADN est transcrit à partir du promoteur de l'ARN polymérase du phage T7. Afin de limiter la transcription à la région codante nous intéressant, l'ADN servant de matrice est préalablement hydrolysé par une enzyme de restriction, au niveau d'un site localisé juste en aval de la séquence du gène à transcrire.

E2.42 Préparation du fragment d'ADN à transcrire.

Après hydrolyse par une enzyme de restriction appropriée, l'ADN matrice est soumis à deux extractions, d'abord par un mélange phénol-chloroforme, puis par un mélange chloroforme-alcool isoamylique. L'ADN est ensuite précipité à l'éthanol en présence d'acétate de sodium 0,3M. Le précipité est repris dans de l'eau et la concentration d'ADN est ajustée à 500 ng/µl.

E2.43 Transcription *in vitro*.

La transcription est réalisée dans le mélange suivant :

2 µg d'ADN linéarisé

10 µl de tampon de transcription 5x (DTT 5 mM; Spermidine 1 mM; Triton X-100 0,01%; PEG 8000 80 mg/ml; $MgCl_2$ 20mM, Tris-HCl 40 mM pH 8,1)

10 µl de rNTP (12,5mM chacun)

0,5 µl de RNAsine 25U/µl (Amersham)

0,5 µl d'ARN polymérase du phage T7 60U/µl (Pharmacia) ou SP6 20U/µl (Boehringer)

H_2O qsp 30 µl.

La RNAsine est un inhibiteur de RNases extrait de placenta.

La solution est incubée 1h à 37°C.

Puis 1 µl de DNase I (5U/µl) est ajouté en présence de 30 µl de tampon DNase 10x (NaCl 50mM, DTT 1mM, BSA 0,1 mg/ml, glycérol 50% (vol/vol), Tris-HCl 20mM pH 7,6), le volume total étant ajusté à 300 µl avec de l'eau.

La solution est incubée 30 min à 37°C, extraite avec 300 µl de phénol, puis 300 µl de chloroforme. L'ARN est précipité à l'éthanol en présence d'acétate de sodium à une concentration finale de 0,3M. Le précipité est repris dans 100 µl d'H_20.

E2.44 Purification des transcrits.

Les transcrits sont purifiés sur gel de polyacrylamide à 10% dénaturant. Après immersion du gel pendant 10 min dans une solution de bromure d'éthydium et rinçage à l'eau, la bande de transcrit est repérée sous UV et découpée. L'élution est réalisée pendant 12h à 4°C dans 150 µl de tampon d'élution des ARN (Acétate d'Ammonium 500mM, Acétate de Magnésium 10mM, SDS 0,1%, EDTA 0,1mM pH 4,8). Le liquide est ensuite prélevé et l'ARN est précipité à l'éthanol.

E2.5 Séquencage de l'ARN par la transcriptase inverse (Qu *et al.*, 1983).

E2.51 Principe.

Un oligonucléotide marqué au ^{32}P à son extrémité 5', est hybridé à une région donnée de l'ARN. L'extrémité 3' de cet oligonucléotide sert d'amorce à la transcriptase inverse qui synthétise, en présence des 4 dNTP, un brin d'ADN complémentaire à l'ARN. Des ddNTP sont ajoutés au milieu réactionnel en concentration telle que la synthèse du brin d'ADN est arrêtée statistiquement. Un témoin d'élongation est effectué en absence de ddNTP.

E2.52 Méthodologie.

L'hybridation entre l'ARN et l'oligodéoxynucléotide amorce est réalisée dans les conditions suivantes :

5 ng d'oligonucléotide marqué au ^{32}P
200 ng d'ARN transcrit *in vitro* ou 20 à 100 µg d'ARN totaux
10 µl de tampon RTB 5x ($MgCl_2$ 30 mM, KCl 200mM, Tris-HCl 250 mM pH 8,3)
H_2O qsp 50 µl

Ce mélange est incubé 10 min à 65°C, puis plongé dans la glace.

Les quatre réactions d'élongation sont réalisées (U, G, C et A) à partir de l'hybride ARN/oligodéoxynucléotide dans les milieux réactionnels décrits dans le Tableau VIII. Les différentes solutions utilisées sont indiquées dans le Tableau IX.

Les réactions sont effectuées à 45°C pendant 30 min, puis sont stoppées par addition de 100 µl d'acétate de sodium 0,3M et 500 µl d'éthanol absolu. Après centrifugation à 13 000 tours/min pendant 20 min, le précipité est lavé avec 200 µl d'éthanol absolu puis séché.

	T	G	C	A
hybride ARN:ADN	10 µl	10 µl	10 µl	10 µl
dNTP 2,5 mM				
T_{10}	1,5 µl			
G_{10}		1,5 µl		
C_{10}			1,5 µl	
A_{10}				1,5 µl
ddNTP 0,1 mM	2 µl	2 µl	2 µl	2 µl
tampon RTB 5x	1 µl	1 µl	1 µl	1 µl
Transcriptase inverse 1U/µl	1 µl	1 µl	1 µl	1 µl

Tableau VIII : Composition des milieux réactionnels utilisés pour le séquençage d'ARN par la transcriptase inverse.

	A_{10}	C_{10}	G_{10}	T_{10}
dATP 10mM	2,5 µl	10 µl	10 µl	10 µl
dCTP 10mM	10 µl	2,5 µl	10 µl	10 µl
dGTP 10mM	10 µl	10 µl	2,5 µl	10 µl
dTTP 10mM	10 µl	10 µl	10 µl	2,5 µl
H_2O qsp 100 µl	67,5 µl	67,5 µl	67,5 µl	67,5 µl

Tableau IX : Composition des solutions utilisées pour le séquençage d'ARN par la transcriptase inverse.

Sondes chimiques

- DMS (Diméthylsulfate) N3-C, N1-A, N7-G et N7-A
- CMCT N3-U et N1-G
  (N-Cyclohexyl-N'-[2-(N-méthylmorpholino)-éthyl] carbodiimid-4-toluosulfonate)

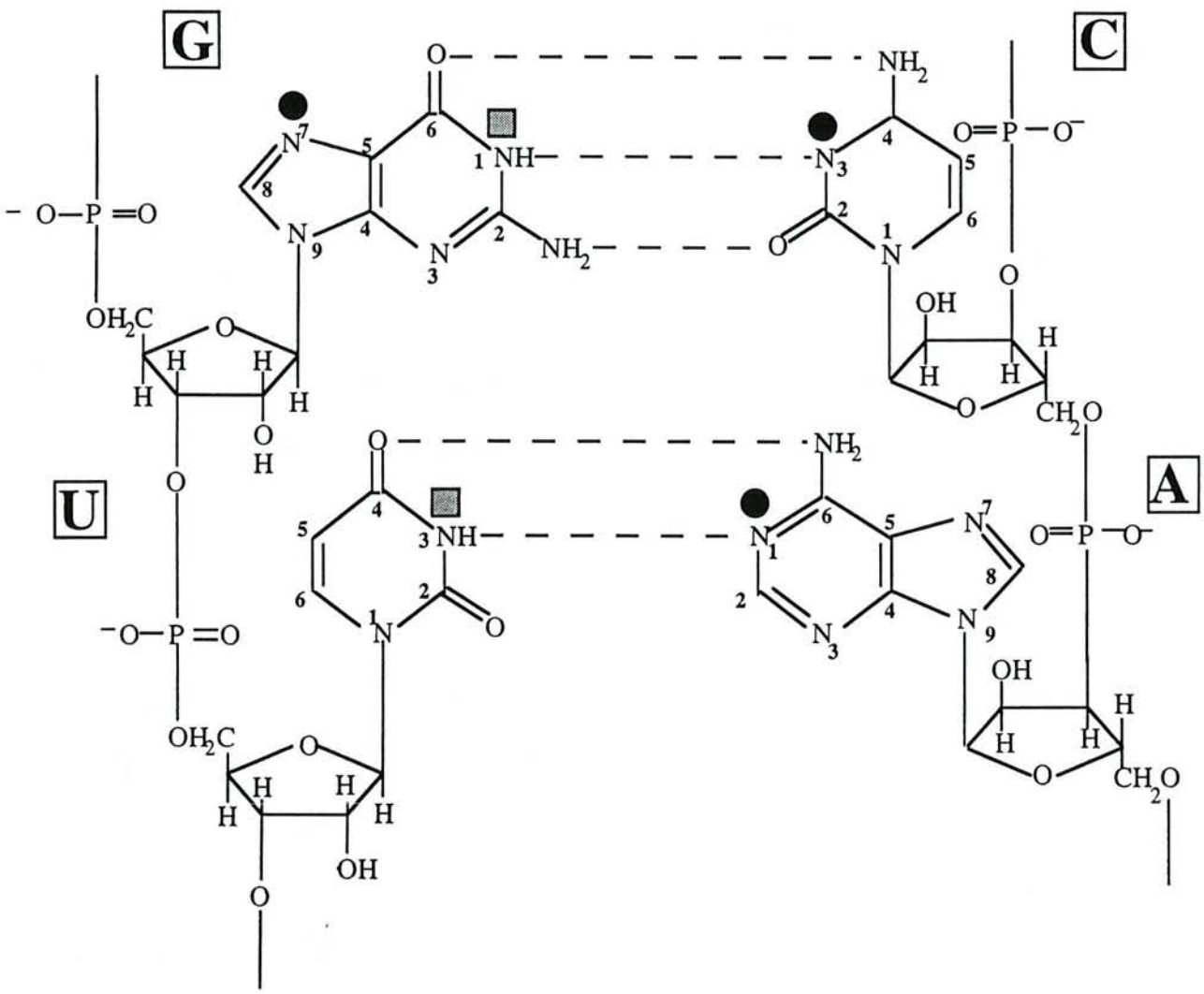


Figure 34 : Localisation des sites de modification des bases par les sondes chimiques.

▩ : site de modification par la CMCT,

● : site de modification par le DMS.

Les précipités sont repris dans 2 µl bleu de formamide. Après chauffage, 2 min à 100°C, les ADNc obtenus sont fractionnés par électrophorèse en gel de polyacrylamide 7% dénaturant. La migration est réalisée à 45V/cm, jusqu'à ce que le bleu de bromophénol ait atteint le bas du gel.

Le gel est fixé à l'acide acétique pendant 10 min, puis transféré sur du papier Whatman 3MM, recouvert d'un film plastique et séché sous vide. Le gel est ensuite placé en fluorographie à -80°C au contact d'un film Fuji X-Ray pendant 12 à 24 heures.

E2.6 Analyse de la structure secondaire des ARN (Qu et *al*, 1983; Ehresmann et *al*, 1987).

E2.61 Principe.

Des sondes chimiques ou enzymatiques sont utilisées pour identifier les régions en simple-brin ou en double brin dans l'ARN. Ces sondes sont soit des agents chimiques, soit des ribonucléases.

Deux enzymes sont utilisées :

- la nucléase S1 (Pharmacia) spécifique des régions simple-brin.
- la RNase V1 (Vassilenko et Babkina, 1965), spécifique des régions structurées, préparée à partir de venin de Cobra *Naja oxiana* qui nous avait été donnée par le Dr S.Vassilenko.

Deux sondes chimiques permettent d'identifier les nucléotides non impliqués dans une interaction de type Watson et Crick (Fig. 34).

- le diméthylsulfate (DMS, Aldrich) qui modifie les positions N7-G, N1-A et plus lentement les positions N7-A et N3-C.
- la 1-cyclohexyl-3-(2-morpholinoéthyl)-carbodiimide-métho-p-toluène sulfonate (CMCT, Merk), qui modifie les positions N3-U et plus lentement les positions N1-G.

Les hydrolyses enzymatiques et les modifications chimiques sont réalisées en présence d'ARNt dans des conditions permettant d'avoir en moyenne une coupure ou une modification par molécule.

Les sites de coupure par les enzymes et les bases modifiées sont mises en évidence par arrêt de la synthèse d'ADNc par la transcriptase inverse à partir d'une amorce oligonucléotide.

L'oligodéoxynucléotide amorce est marqué en 5' au ^{32}P. Les ADNc synthétisés sont séparés par électrophorèse en gel de polyacrylamide à 7%, urée 8M, dans un tampon TBE 1x avec, en parallèle, avec les produits d'un séquençage de l'ARN réalisé avec la transcriptase inverse et le même oligonucléotide amorce. La position des bases modifiées et des coupures enzymatiques est identifiée en référence au séquençage.

E2.62 Méthodologies appliquées aux transcrits produits *in vitro*.

Modifications chimiques.

La quantité de transcrit obtenu par transcription *in vitro* étant variable, elle est tout d'abord estimée par séquençage du transcrit à la transcriptase inverse. Si la quantité s'avère suffisante, avant de réaliser les modifications chimiques, les ARN sont dissous dans 100 µl du tampon dans lequel seront réalisées les réactions, et préincubés 10 min à 20°C. Les modifications chimiques sont réalisées en conditions natives ou semi-dénaturantes. Les modifications sont effectuées sur 2 µl de transcrit auxquels sont ajoutés 5 µg d'un mélange d'ARNt de levure (Boehringer).

¤ En condition non-dénaturante, les modifications au DMS sont réalisées dans le tampon Cacodylate de Sodium 100mM, KCl 50mM, $MgCl_2$ 100mM, pH7,5 et en conditions semi-dénaturantes, dans le tampon Cacodylate de Sodium 100 mM, EDTA 2mM. Deux concentrations en DMS ont été utilisées : soit 0,75µl de DMS pur, soit 0,75µl d'une solution de DMS 1/1 (V/V) (DMS/EtOH) par essai. Les 0,75 µl sont ajoutés aux 100 µl de tampon dans lesquels l'ARN a été préincubé. Après une incubation de 15 min, la réaction est stoppée par ajout de 10 µl d'acétate de sodium et précipitation à l'éthanol. Le précipité obtenu est lavé à l'éthanol à 70%, puis dissous dans 14 µl d'H_2O. 7 µl sont utilisés pour l'analyse par extension d'amorce avec la transcriptase inverse.

¤ En condition non-dénaturante, les modifications par la CMCT sont réalisées dans un tampon Borate de Sodium 100mM, KCl 50mM, $MgCl_2$ 100mM, pH8 et en conditions semi-dénaturantes, dans le tampon Borate de Sodium 100 mM, EDTA 2mM. Plusieurs quantités de CMCT ont été utilisées : 12,5 µl d'une solution de CMCT à 84 mg/ml ou 168 mg/ml.
Ces 12,5 µl sont ajoutés aux 100 µl de tampon dans lesquels l'ARN a été préincubé. Après 30 min d'incubation, la réaction est stoppée comme décrit ci-dessus pour les modifications au DMS. Le précipité est également lavé et repris dans 14 µl d'H_2O.

Hydrolyses enzymatiques.

¤ Les digestions par la nucléase S1 sont effectuées à 20°C dans 10 µl de tampon Acétate de Sodium 25mM, KCl 50mM, $MgCl_2$ 10mM, $ZnCl_2$ 1mM, pH4,5. Avant l'ajout de la nucléase S1, l'ARN (2 µl de transcrit + 5 µg ARNt) est préincubé dans ce tampon pendant 10 min à 20°C. Deux réactions ont été réalisées en parallèle : 2 unités de nucléase S1 sont ajoutées dans chaque tube ; dans le premier, la réaction est stoppée après 12 min d'incubation et dans le deuxième après 25 min d'incubation. Les deux réactions sont stoppées par ajout de 2 µl d'EDTA 100mM suivi d'une extraction au phénol. Après précipitation à l'éthanol et lavage à l'éthanol à 70%, les précipités d'ARN sont repris dans 14 µl d'H_2O et analysés par la transcriptase inverse (7 µl par réaction).

¤ Les digestions par la nucléase V1 sont effectuées à 0°C dans 10 µl de tampon Tris-HCl 10mM, KCl 350mM, $MgCl_2$ 10mM, pH7,5. Avant l'ajout de la nucléase V1, l'ARN (2 µl de transcrit + 5

µg ARNt) est préincubé dans ce tampon pendant 10 min à 0°C. Un rapport enzyme/ARN de 1,2 unités/µg a été utilisé. Des incubations de 2 et 7 min, parfois 10 min, ont été réalisées en parallèle. Les réactions sont arrêtées par ajout de 5 µl d'EDTA à 100mM suivi d'une extraction au phénol. Les ARN hydrolysés pendant 2 min et 7 min d'incubation sont mélangés. Après précipitation à l'éthanol et lavage à l'éthanol à 70%, les précipités d'ARN sont repris dans 14 µl d'H_2O et analysés par la transcriptase inverse (7 µl par réaction).

E3. Utilisation du logiciel GCG.

E3.1 Calcul de la stabilité des hélices d'ARN.

La stabilité des hélices d'ARN est calculée avec le programme MFold du logiciel GCG (Genetic Computer Group, Wisconsin) version Unix 8.1 (1995), programme adapté de celui publié par Jeager *et al.*, 1989. La stabilité des structures prédites par ce logiciel est calculée pour un ARN dissous en présence de NaCl 1 M à 37°C.

E3.2 Alignement des séquences.

Les alignements de séquences d'ADN ou d'ARN ont été réalisés avec le programme PileUp du logiciel GCG version Unix 8.1 (1995), programme basé sur l'algorithme développé par Feng et Doolittle (1987) Les alignements ont été ajustés avec le programme LineUp du même logiciel et la présentation des alignements (zones grisées et encadrements) a été réalisée à l'aide du logiciel Alscript (Barton, 1993).

E3.3 Calcul des pourcentages d'homologies de séquences.

Les pourcentages d'homologie entre deux séquences ont été calculés avec le programme Gap du logiciel GCG, c e programme utilise la méthode d'alignement de Needleman et Wunsch (1970).

Résultats

F. Objectifs

Lorsque je suis arrivée au laboratoire, l'équipe venait de découvrir que les deux gènes de snoRNA U3 chez *Saccharomyces cerevisiae*, contenaient un intron présentant les caractéristiques des introns épissés dans un spliceosome (Myslinski *et al.*, 1990) (Fig. 35). Comme nous l'avons mentionné dans l'introduction, il s'agissait, avec la mise en évidence d'un intron du même type dans le gène du snRNA U6 de *Schizosaccharomyces pombe* (Tani et Ohshima, 1989), des deux premières observations d'introns susceptibles d'être épissés au sein d'un spliceosome, dans des gènes codant pour des ARN jouant un rôle fonctionnel. Ce qui soulevait beaucoup de questions.

Tout d'abord, les pré-snoRNA U3 étaient-ils bien épissés par le même type de spliceosome que les ARN pré-messagers ? En effet, les deux introns mis en évidence dans les deux gènes de snoRNA U3 de *S. cerevisiae* contenaient bien un intron portant les séquences 5', 3' et de branchement retrouvées dans les introns des ARN pré-messagers. Néanmoins, la séquence de la boîte de branchement différait à une position par rapport à celle des introns de gènes de protéines (GACTAAC au lieu de TACTAAC).

Dans un premier temps, j'ai donc réalisé un certain nombre de constructions permettant de produire un pré-snoRNA U3A ou un pré-snoRNA U3B *in vitro* par transcription par l'ARN polymérase du phage T7, ceci, afin d'utiliser ces transcrits dans des expériences d'épissage *in vitro* qui ont été réalisées par J. Banroques.

J'avais réalisé 2 constructions différentes pour le pré-snoRNA U3A, les transcrits produits différaient par une extension en 5' de l'exon 1. La forte différence d'efficacité d'épissage observée pour ces transcrits a amené l'équipe à réaliser une étude approfondie de la relation structure du pré-snoRNA U3A/efficacité d'épissage. Par ailleurs, les constructions que j'avais réalisées, aussi bien pour le pré-snoRNA U3A que U3B, ont permis d'établir un modèle de structure secondaire pour ces deux ARN.

Une seconde question importante se posait lorsque je suis arrivée au laboratoire. Du fait que les gènes de snoRNA caractérisés jusque là, en particulier, les deux gènes codant pour le snoRNA U3 de *S. pombe* (Porter *et al.*, 1988; Selinger *et al.*, 1992) ne contenaient pas d'intron, il était important de savoir si les introns mis en évidence dans les deux gènes U3 de *S. cerevisiae* (Myslinski *et al.*, 1990) représentaient un cas d'espèce, ou étaient une propriété générale des levures appartenant à la sous-famille des Saccharomycetoideae. En effet, comme nous l'avons vu dans l'introduction bibliographique, alors que *S. cerevisiae* appartient à cette sous-famille, *S. pombe* appartient à la sous-famille des Schizosaccharomycetoideae. Afin de répondre à cette question, j'ai rassemblé une large collection de levures appartenant à quatre genres

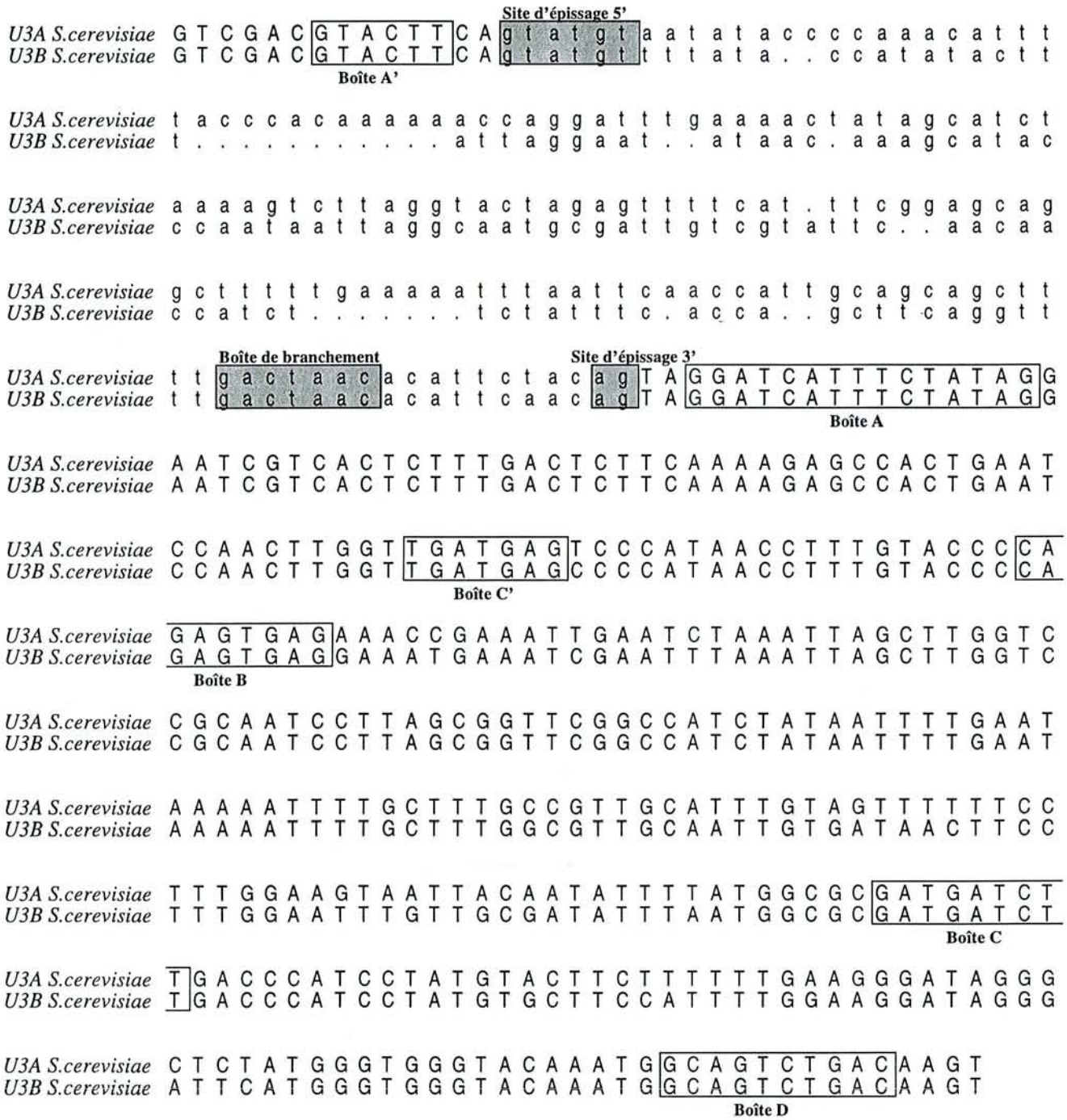


Figure 35 : Alignement des séquences des gènes codant pour les snoRNA U3A et U3B de *Saccharomyces cerevisiae* (Myslinski *et al.*, 1990).

Les régions codantes sont indiquées en lettres majuscules et les séquences introniques en lettres minuscules.
Les boîtes A', A, C', B, C et D phylogénétiquement conservées dans tous les snoRNA U3, sont encadrées. Les signaux de reconnaissance des introns épissés dans un spliceosome (boîte 5', boîte de branchement et dinucléotide AG en 3') sont indiquées en gris.

distincts de la sous-famille des Saccharomycetoideae : *Saccharomyces*, *Kluyveromyces*, *Pichia* et *Hansenula*. J'ai développé une méthode basée sur l'amplification par la technique PCR permettant d'étudier les introns des gènes de snoRNA U3 et aussi, les gènes de snoRNA U3. Comme nous le verrons, la présence d'introns chez *S. cerevisiae* ne s'est pas avérée être un cas d'espèce. Ces études nous ont permis de mettre en évidence une évolution très complexe des gènes de snoRNA U3 chez les Saccharomycetoideae.

Par ailleurs, nous avons, sur la base des données acquises par la comparaison des différents snoRNA U3 que j'avais mis en évidence chez les levures appartenant à la sous-famille des Saccharomycetoideae, développé des travaux visant à définir les régions essentielles à la fonction du snoRNA U3 chez les levures et nous avons apporté un argument phylogénétique fort en faveur d'un appariement avec la région 5' ETS du pré-ARNr.

Mes travaux de thèse étant présentés sous forme de publications publiées, sous presse ou préparées pour soumission, pour faciliter la lecture de ce manuscrit, nous commencerons par un résumé détaillé expliquant la démarche scientifique et les résultats obtenus. Les articles seront présentés sous la forme de deux sous-chapitres. Le premier concerne l'étude de la relation entre la structure secondaire des pré-snoRNA U3A et U3B de *Saccharomyces cerevisiae* et leur efficacité d'épissage *in vitro*. Cette étude correspond à un travail auquel j'ai participé, sans être le maître d'œuvre principal. Le second sous-chapitre correspond à l'étude comparative des gènes codant pour le snoRNA U3 de différentes espèces de levures. J'ai initié et développé cette étude au laboratoire avec, pour certains aspects plus génétiques ou plus structuraux, la collaboration de R. Fournier et A. Grégoire.

Ces deux sous-chapitres seront suivis d'une discussion sur les perpectives de recherche ouvertes par l'ensemble des résultats obtenus.

Enfin au cours de ma thèse, j'ai été amenée à aider l'équipe de l'unité INSERM U308 dirigée par J.P. Nicolas à développer des méthodes de biologie moléculaire. Un article a été publié dans le cadre de cette collaboration. Il est présenté en annexe.

G. Résumé détaillé des travaux réalisés et résultats obtenus

G1. Etudes sur la relation entre la structure secondaire des pré-snoRNA U3A et U3B et leur efficacité d'épissage *in vitro*.

Comme nous l'avons mentionné ci-dessus, notre premier objectif était de savoir si le pré-snoRNA U3A était bien épissé par le même type de spliceosomes que les ARN pré-messagers, et aussi de tester l'efficacité d'épissage de ce pré-snoRNA U3A. Nous avons choisi de réaliser cette étude *in vitro*, en collaboration avec J. Banroques, qui disposait d'extraits nucléaires de mutants thermosensibles de *S. cerevisiae* affectés, à température non permissive, dans le processus d'épissage des ARN pré-messagers.

Cette étude est décrite dans l'article n° 1 :

"The secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency." A. Mougin, A. Grégoire, J. Banroques, V. Ségault, R. Fournier, F. Brulé, M. Chevrier-Miller et C. Branlant. RNA, sous presse.

Pour répondre à la question posée, j'ai réalisé une première construction qui permettait de produire par transcription *in vitro* par l'ARN polymérase du phage T7, un transcrit correspondant au pré-snoRNA U3A. Cette construction conduisait à un pré-snoRNA U3A où l'exon 1 présentait une extension de 52nt en 5' par rapport à l'ARN naturel. Ce transcrit a été dénommé pU3G66 (p pour précurseur du snoRNA, G du fait de la séquence GACTAAC de la boîte de branchement et 66 parce que l'exon 1 comporte 66nt) (Fig. 36). J. Banroques a utilisé cette construction pour réaliser des tests d'épissage *in vitro*. Le transcrit, ainsi produit, était épissé *in vitro*, avec une efficacité extrêmement élevée, la meilleure efficacité décrite jusqu'à ce jour, pour un épissage *in vitro* réalisé avec un extrait nucléaire de levure (Fig. 37). Par ailleurs, en employant des extraits nucléaires de souches de *S. cerevisiae* thermosensibles, affectées dans certains des gènes indispensables à la formation d'un spliceosome actif pour la maturation des ARN pré-messagers, J. Banroques a pu, grâce à la construction que j'avais réalisée, prouver que le pré-snoRNA U3 est bien épissé par le même type de spliceosome que les ARN messagers (Fig. 37).

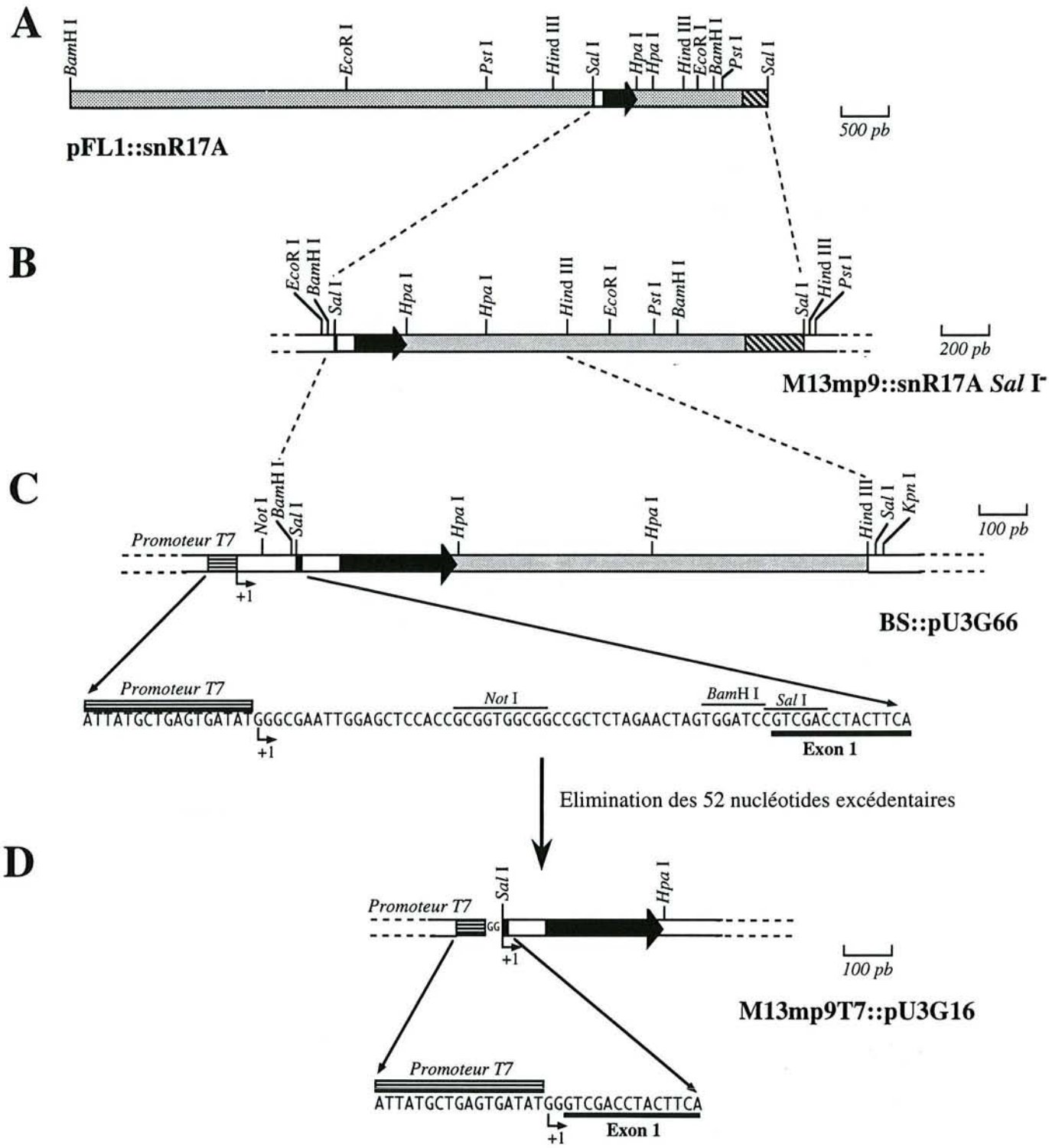


Figure 36 : Constructions génétiques réalisées pour étudier la structure secondaire du pré-snoRNA U3A et l'efficacité d'épissage des transcrits produits.

A. Carte de restriction du fragment chromosomique de 8kb inséré dans le plasmide pFL1::snR17A (Myslinski *et al.*, 1990).

B. Carte de restriction du fragment *Sal* I provenant du plasmide pFL1::snR17A et inséré dans le phage M13mp9::snR17A *Sal* I.

C. Carte de restriction du fragment *Bam*H I/*Hin*d III provenant du phage recombinant M13mp9::snR17A*Sal* I, inséré dans le phagemide pBluescript (BS::pU3G66).
La séquence du promoteur de l'ARN polymérase du phage T7 est indiquée, ainsi que celle des 52nt excédentaires entre le site +1 d'initiation de la transcription et le premier nucléotide de l'exon 1 du gène codant pour le snoRNA U3A. La séquence de l'exon 1 est également indiquée.

D. Représentation schématique du phage M13mp9T7::pU3G16.
La séquence du promoteur de l'ARN polymérase du phage T7 est indiquée, ainsi que celle du premier exon du gène U3A.

fragment de 8kb inséré dans le plasmide pFL1::snR17A

région appartenant au plasmide pFL1

région correspondant au promoteur du phage T7

gène *SNR17A* codant pour le snoRNA U3A de *Saccharomyces cerevisiae*, avec en noir les exons et en blanc l'intron.

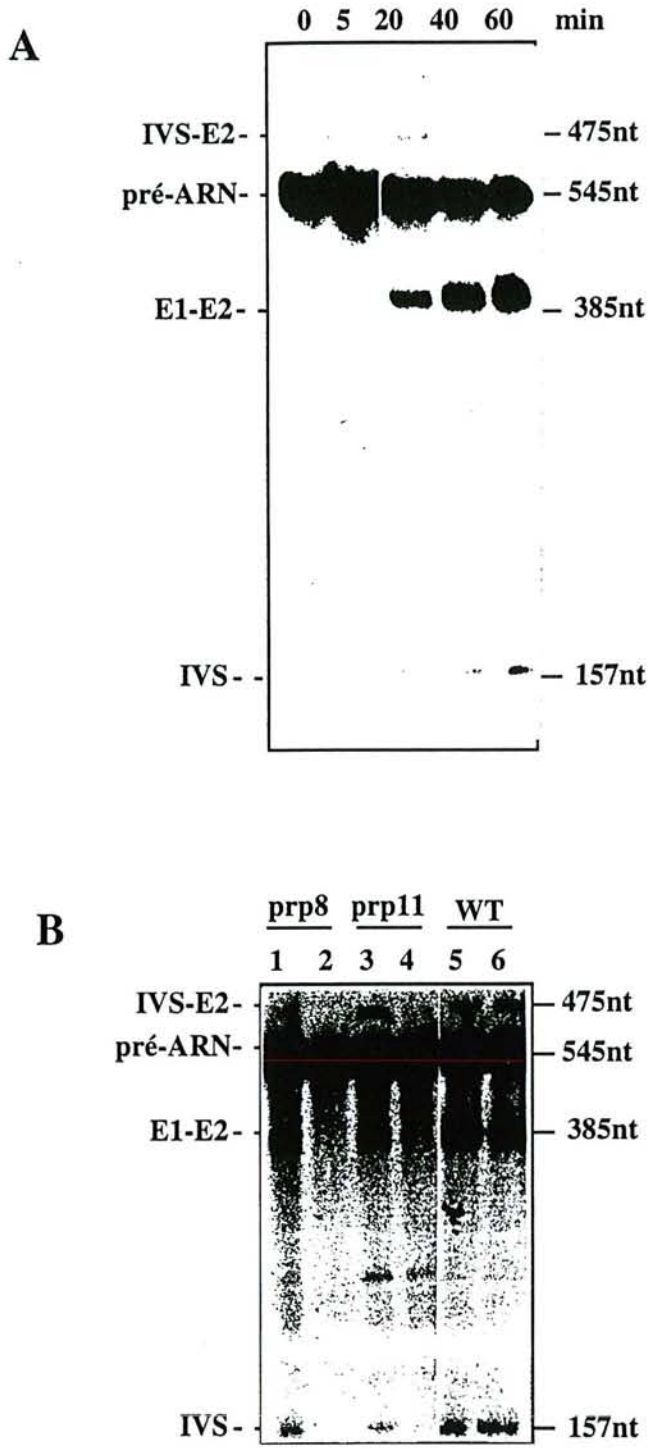


Figure 37 : A. Cinétique d'épissage *in vitro* du transcrit pU3G66.

Le transcrit pU3G66 a été incubé dans les conditions décrites par Newman *et al.* (1985). Des prélèvements ont été faits 0, 5, 20, 40 et 60 minutes à partir du début de l'incubation. Les produits de la réaction ont été fractionnés par électrophorèse en gel de polyacrylamide 4%.

Les positions du précurseur (pré-ARN), du premier intermédiare (IVS-E2), du produit mature (E1-E2) et de l'intron sous forme de lasso (IVS) sont indiquées.

B. Epissage *in vitro* du transcrit pU3G66 avec différents extraits de levure.

Le transcrit pU3G66 a été incubé dans des conditions permettant l'épissage avec un extrait de la souche sauvage EJ101 (WT) et des extraits de 2 mutants thermosensibles AHrna8 (prp8) et SPJ11.4 (prp11). Les expériences correspondant aux colonnes 1, 3 et 5 avaient été réalisées sans pré-incubation des extraits; dans les expériences correspondant aux colonnes 2, 4 et 6 les extraits avaient été incubés à température non-permissive (30 min à 30°C) avant épissage.

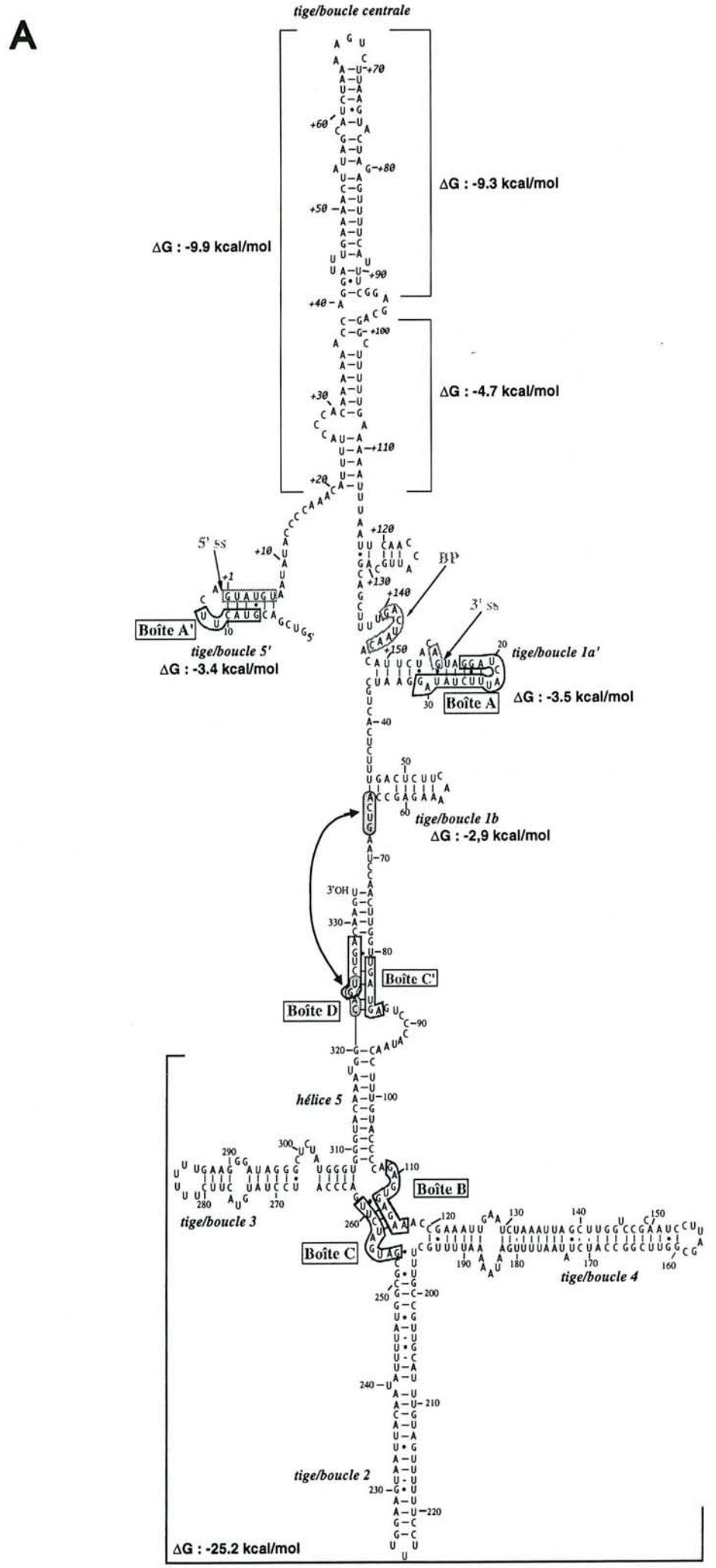
Les positions du précurseur (pré-ARN), du premier intermédiare (IVS-E2), du produit mature (E1-E2) et de l'intron sous forme de lasso (IVS) sont indiquées.

Etant donné que le transcrit pU3G66 renfermait 52nt de plus que le précurseur naturel, nous avons ensuite délété ces 52nt supplémentaires par mutagenèse dirigée. Les 52nt ne pouvaient pas être éliminés en totalité, du fait des contraintes imposées par l'ARN polymérase du phage T7. L'ARN produit par la nouvelle construction renfermait 2 résidus G supplémentaires en 5' par rapport à l'ARN naturel (Fig. 36). Ce nouveau transcrit a été dénommé pU3G16, car la taille de l'exon 1 est de 16nt. Son épissage *in vitro* s'est avéré très peu efficace. V. Ségault, A. Grégoire, A. Mougin et R. Fournier ont alors entrepris une étude de la structure secondaire des transcrits produits par les deux constructions, mais aussi, de différents mutants de ces transcrits. L'étude de la structure secondaire de ces différents ARN et de leur efficacité d'épissage a permis, d'une part, d'établir un modèle de structure secondaire du pré-snoRNA U3A (Fig. 38), d'autre part, d'établir un certain nombre de relations entre la structure de ce pré-snoRNA et son efficacité d'épissage *in vitro*. Ce qui a permis de proposer une explication à la différence de taux d'épissage observée entre les deux transcrits pU3G66 et pU3G16.

Au sein du transcrit pU3G16 ne comportant que 2 résidus G supplémentaires, ces 2 résidus et le G normalement présent en 5' de l'ARN naturel forment trois paires de bases G-C avec une série de résidus C localisée à proximité de l'extrémité 5' terminale de l'intron (Fig. 39). De plus, les études structurales menées par A. Mougin, A. Grégoire et R. Fournier ont permis de montrer que de manière inattendue au sein du pré-snoRNA U3, la séquence conservée en 5' de l'intron, qui se lie au snRNA U1 lors de l'assemblage du spliceosome, est appariée avec un segment de l'exon 1, ce qui conduit à la présence d'une structure tige/boucle 5' terminale dans le pré-snoRNA U3 (Fig. 38).

Dans le transcrit long (pU3G66), les deux résidus G 5' terminaux résultant de la séquence des sites d'initiation de la transcription par l'ARN polymérase du phage T7, sont impliqués dans une structure formée par l'extension de 52nt (Fig. 39). En d'autres termes, cette extension s'auto-structure, sans interférer avec la structure du pré-snoRNA U3. Dans ces conditions, l'épissage est parfaitement efficace *in vitro*, en dépit d'une boîte de branchement non canonique et d'un appariement de la boîte 5'. Par contre, dans le transcrit court, les 3 résidus G en 5' renforcent très fortement la stabilité de la structure tige/boucle en 5' (Fig. 39). Ce qui diminue très fortement l'efficacité d'épissage *in vitro* en présence d'une boîte de branchement non canonique (Fig. 40).

A. Grégoire, R. Fournier et J. Banroques ont montré qu'en convertissant la boîte de branchement non canonique en une boîte de branchement canonique, l'efficacité d'épissage *in vitro* du transcrit court pouvait être très fortement augmentée. Ces observations démontraient l'importance de la structure secondaire des pré-ARN de *S. cerevisiae* pour leur efficacité d'épissage. Nous démontrions aussi une interrelation qui existe entre séquence de la boîte de branchement et disponibilité de la boîte 5' dans l'assemblage du spliceosome de levure. En effet, J. Banroques a pu montrer que la faible


A
tige/boucle centrale
ΔG : -9.3 kcal/mol
ΔG : -9.9 kcal/mol
ΔG : -4.7 kcal/mol
5' ss
3' ss
BP
Boîte A'
tige/boucle 5'
ΔG : -3.4 kcal/mol
tige/boucle 1a'
Boîte A
ΔG : -3.5 kcal/mol
tige/boucle 1b
ΔG : -2,9 kcal/mol
3'OH
Boîte C'
Boîte D
hélice 5
Boîte B
Boîte C
tige/boucle 3
tige/boucle 4
tige/boucle 2
ΔG : -25.2 kcal/mol

B

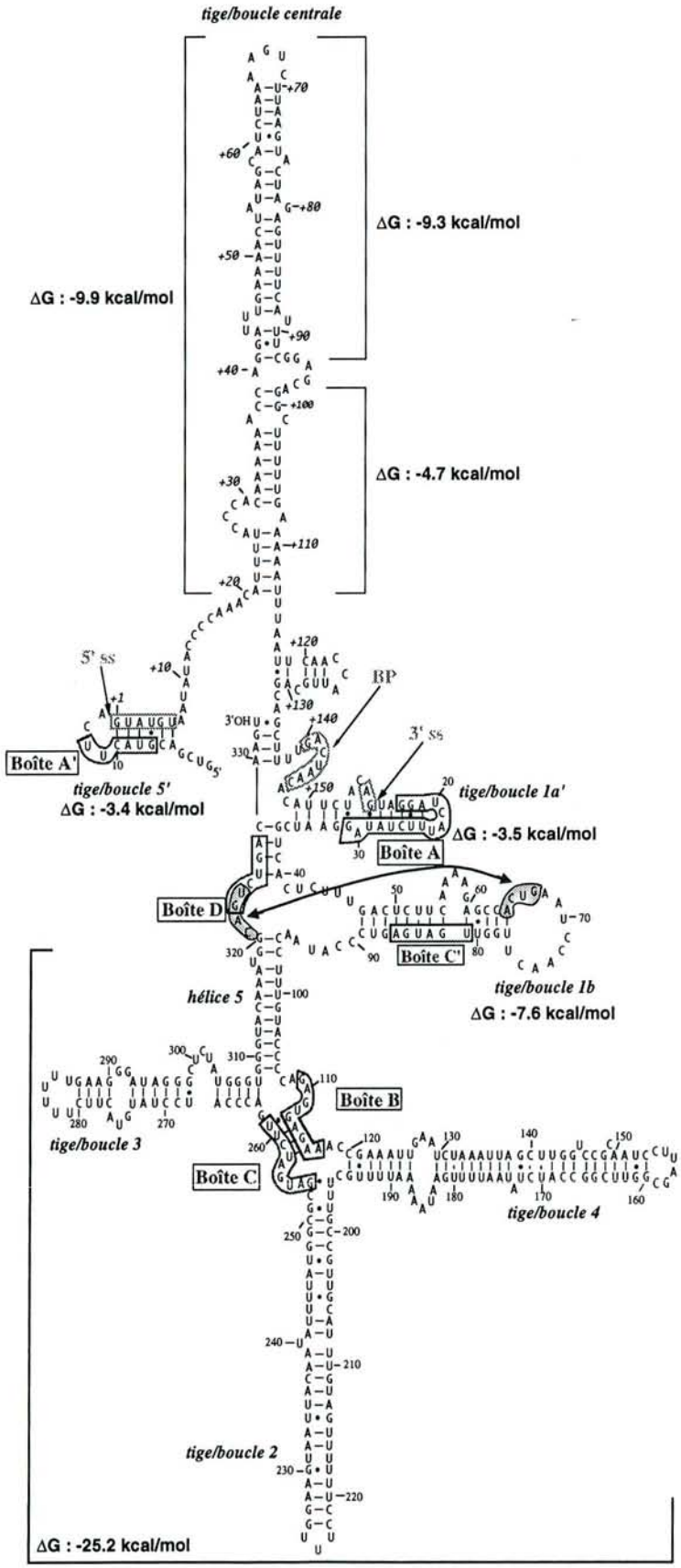
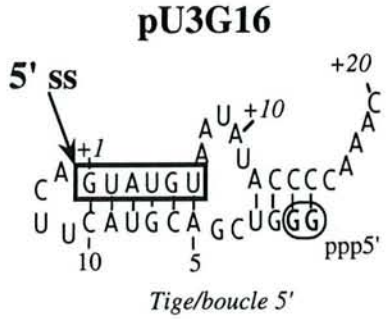


Figure 38 : Modèle de structure secondaire du pré-snoRNA U3A de *S. cerevisiae*.

Les boîtes phylogénétiquement conservées du snoRNA U3 sont indiquées, ainsi que les séquences fonctionnelles de l'intron, site 5' (5' ss), site 3' (3' ss) et boîte de branchement (BP).
L'interaction possible de type pseudo-nœud est visualisée par les 2 régions en grisé, reliées par une flèche.
La structure du domaine 3' est celle correspondant le mieux au résultat de l'analyse structurale en solution.
Les deux modèles de structuration du domaine 5' du snoRNA U3A au sein de la snoRNP (A) et en solution (B)sont représentées.
La stabilité des hélices a été calculée avec le programme MFold du logiciel GCG.

A



B

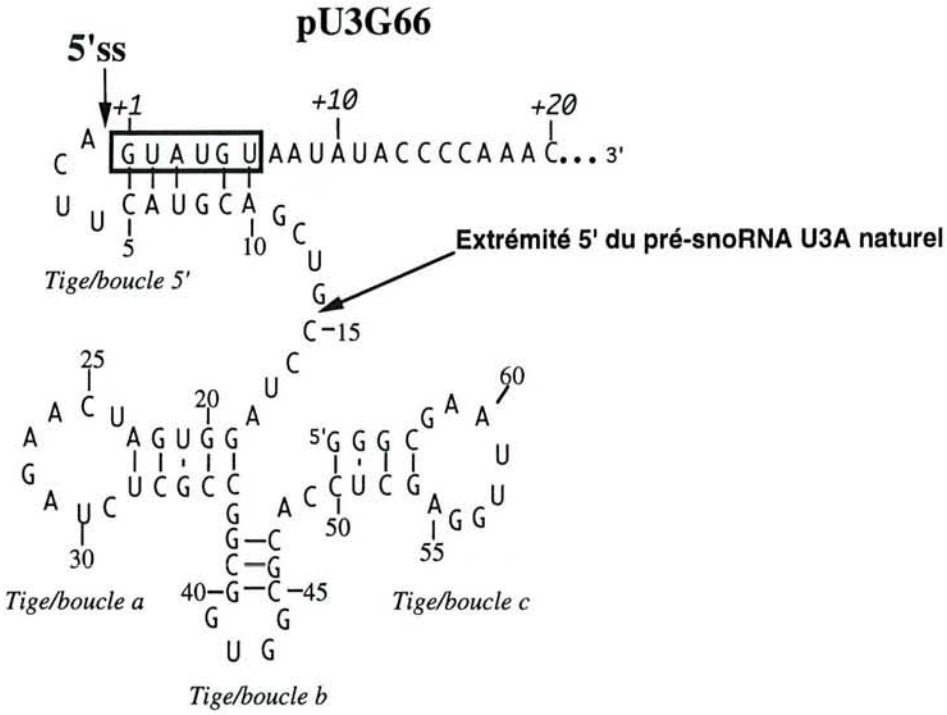


Figure 39 : Comparaison des structures secondaires des extrémités 5' des transcrits pU3G16 (A) et pU3G66 (B).

Le transcrit pU3G66 présente une extension de 52nt par rapport au transcrit pU3G16.
Les deux résidus G supplémentaires présents dans le transcrit court sont représentés en gras et sont encerclés.

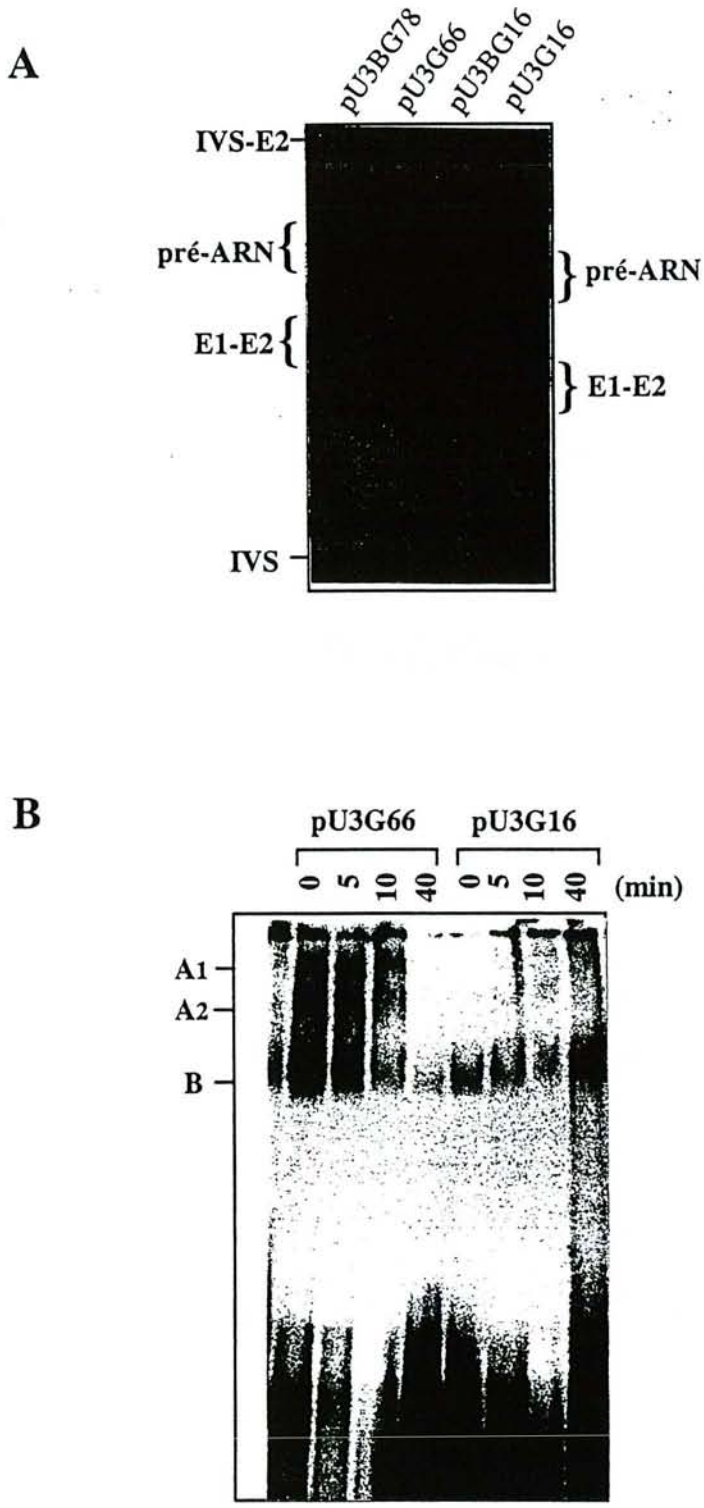


Figure 40 : A. Epissage *in vitro* de différents transcrits correspondant à des pré-ARN U3A et U3B chimériques.

Les positions du précurseur (pré-ARN), du premier intermédiaire (IVS-E2), du produit mature (E1-E2) et de l'intron sous forme de lasso (IVS) sont indiquées.

B. Cinétiques d'assemblage du spliceosome sur les transcrits pU3G66 et pU3G16 (Mougin *et al.*, 1996).

Les assemblages des complexes spliceosomaux ont été réalisés en utilisant des extraits nucléaires de la souche EJ101.

Les positions des complexes A1, A2 et B (Cheng et Abelson, 1987) sont indiquées.

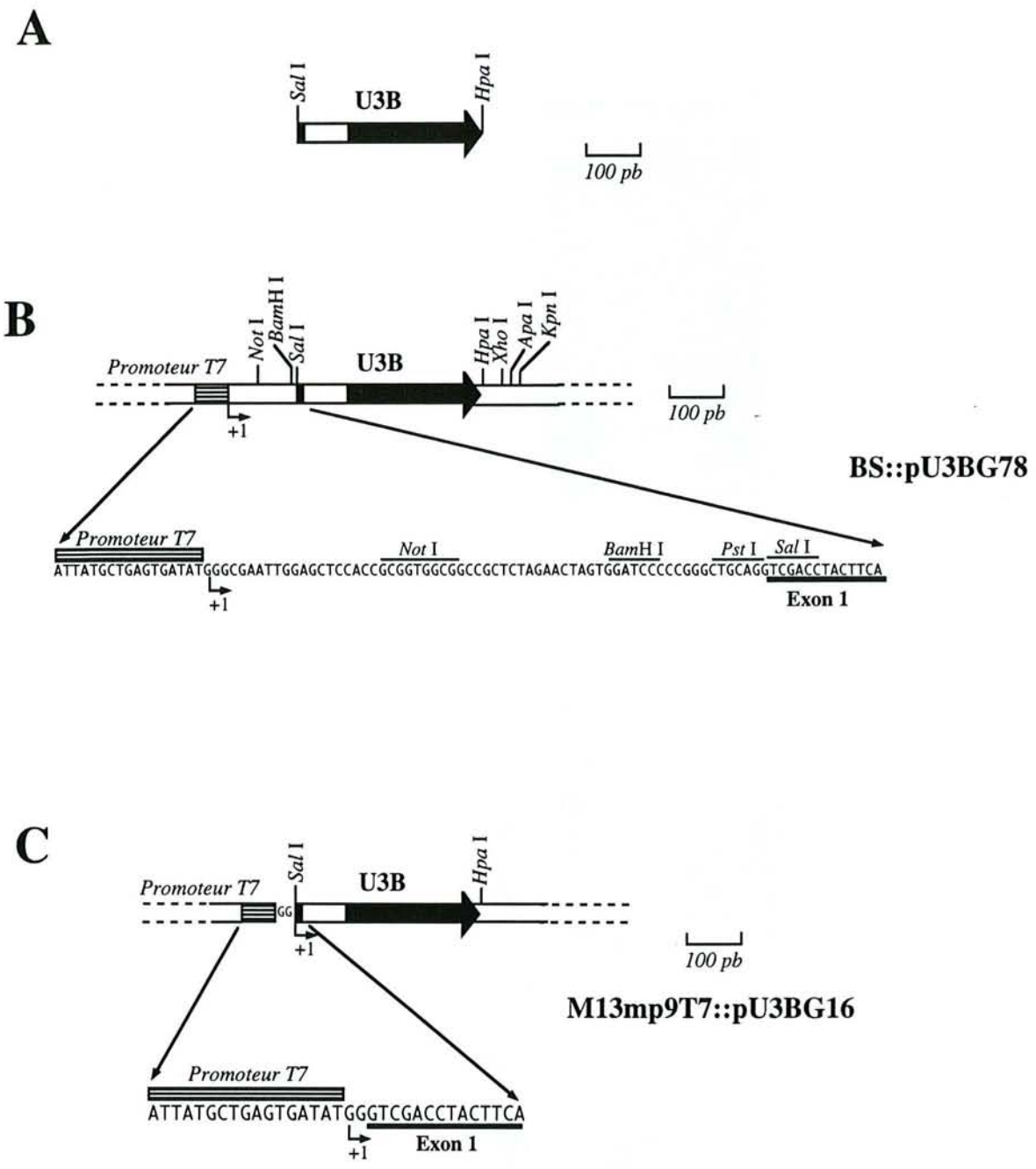


Figure 41: Constructions génétiques réalisées pour étudier la structure secondaire du pré-snoRNA U3B et son efficacité d'épissage *in vitro*.

A. Carte de restriction du fragment d'ADN obtenu par amplification et correspondant à la région codante du gène du snoRNA U3

B. Représentation schématique de la région du plasmide BS::pU3BG78 renfermant la région codante du gène du snoRNA U3B sous contrôle d'un promoteur de l'ARN polymérase du phage T7.
La séquence du promoteur de l'ARN polymérase du phage T7 est indiquée, ainsi que celle des 64nt excédentaires entre le site +1 d'initiation de la transcription et le premier nucléotide de l'exon 1 du gène codant pour le snoRNA U3B. La séquence de l'exon 1 est également représentée.

C. Représentation schématique de la région du bactériophage M13mp9T7::pU3BG16, contenant la région codante du gène du snoRNA U3B sous contrôle d'un promoteur de l'ARN polymérase du phage T7.
La séquence du promoteur de l'ARN polymérase du phage T7 est indiquée, ainsi que celle du premier exon.

région correspondant au promoteur du phage T7

U3B gène du snoRNA U3B de *Saccharomyces cerevisiae*, avec en noir les exons et en blanc l'intron.

efficacité d'épissage observée pour le transcrit pU3G16 était liée à une très faible formation de complexes d'épissage (Fig. 40).

Il était intéressant de tester si la situation était similaire pour le pré-snoRNA U3B. Pour cela, j'ai réalisé deux constructions génétiques, où la région codante du gène U3B était placée sous contrôle du promoteur de l'ARN polymérase du phage T7, l'une courte conduisant à un ARN ne comportant que 2nt supplémentaires par rapport à l'ARN naturel, l'autre conduisant à un ARN ayant une extension en 5', comme pour le transcrit pU3G66 (Fig. 41). Une différence marquée d'efficacité d'épissage a aussi été observée pour les deux types de transcrits (Fig. 40). A. Grégoire a utilisé les constructions que j'avais réalisées pour étudier la structure secondaire du pré-snoRNA U3B. Un article présentant l'analyse de cette structure est en préparation. Une comparaison des structures des pré-snoRNA U3A et U3B est décrite dans l'article 2 :

"Secondary structure conservation of the U3 small nucleolar RNA introns in *Saccharomyces*" F. Brulé, A. Grégoire, V. Ségault, A. Mougin et C. Branlant. C. R. Acad. Sci. Paris Sciences de la Vie/Life sciences, 318, 1197-1206.

Il ressort de la comparaison décrite dans cet article (Fig. 42), qu'en dépit d'une forte divergence en taille et en séquence, les introns des pré-snoRNA U3A et U3B se replient de la même manière, avec la présence d'une longue structure tige/boucle centrale, d'une boîte 5' impliquée dans une structure tige/boucle, d'une boîte de branchement en simple-brin et d'une boîte 3' dans une boucle latérale d'une structure tige/boucle. La présence d'une structure tige/boucle en 5' et d'une boîte de branchement non canonique dans le pré-snoRNA U3B, explique que nous ayons trouvé pour ce pré-ARN, comme pour le pré-snoRNA U3A, un effet sur l'épissage *in vitro* de l'addition de nucléotides renforçant la structure tige/boucle 5' terminale.

La conservation de ces différentes caractéristiques structurales dans les deux pré-ARN, avec des introns ayant, respectivement 157 et 130 nt, et présentant un taux de similitude extrêmement faible (seules les boîtes 5' et de branchement et la séquence entre la boîte de branchement et le site 3' accepteur sont conservées) (Fig. 35), suggère que la préservation de cette structuration reflète des contraintes liées à la nécessité d'une forte efficacité d'épissage des pré-snoRNA U3. En effet, le snoRNA U3 est le plus abondant de tous les snRNA chez la levure, il doit être produit à un taux élevé lorsque les levures sont en phase exponentielle de croissance, d'où la nécessité d'une forte efficacité d'épissage. Aucune accumulation du précurseur n'est observée avec un gène U3A ou U3B sauvage, quelles que soient les conditions de culture.

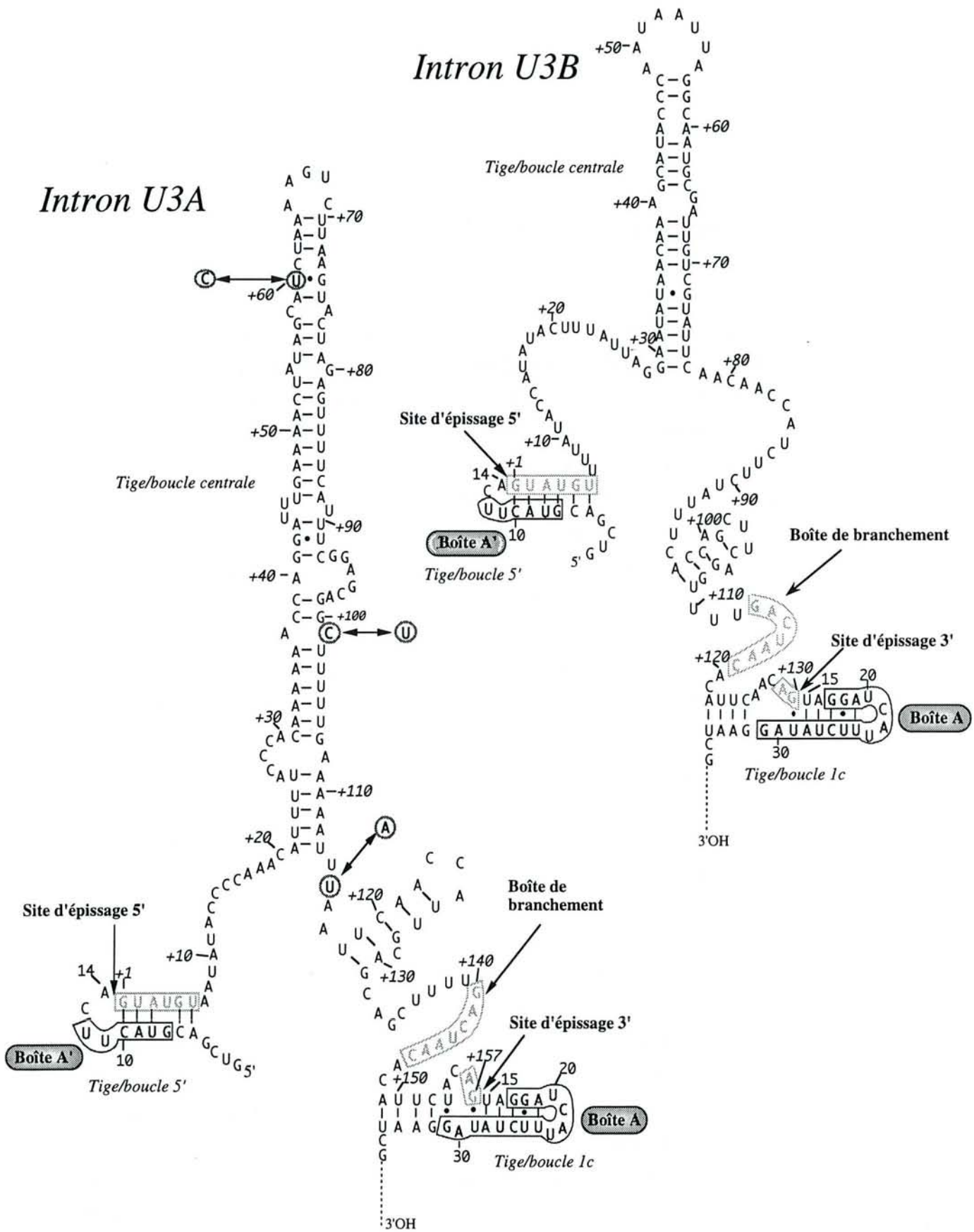


Figure 42 : Modèle de structure secondaire de l'intron des pré-snoRNA U3A, U3A* et U3B.

Les substitutions de bases entre les allèles U3A et U3A* sont encerclées en vert.
Les boîtes phylogénétiquement conservées A et A' sont encadrées en rouge, les boîtes de reconnaissance des introns (boîtes 5', de branchement et 3') sont indiquées en bleu.
Les nucléotides des exons sont numérotés selon la numérotation dans l'ARN mature (Ségault*et al.*, 1992), et les nucléotides de l'intron de *+1* à *+157* pour les introns U3A et U3A*, et *+130* pour l'intron U3B.

Il avait été proposé qu'en plus de la présence des boîtes 5', 3' et de branchement, les grands introns de levure nécessitent la présence d'un appariement rapprochant ces boîtes fonctionnelles (Parker et Patterson, 1987). Les résultats décrits dans les articles 1 et 2 en apportent une confirmation et montrent que cet appariement appartient à une longue structure tige/boucle dans le cas des pré-snoRNA U3. A. Grégoire a montré au laboratoire qu'il en était de même pour le pré-ARN messager de l'actine.

Les résultats de Hughes *et al.* (1987), et de Myslinski *et al.* (1990) au laboratoire, montraient que seul le gène du pré-snoRNA U3A est actif dans des conditions de fermentation classiques. Hughes et Ares (1991) ont montré, par ailleurs, que lorsque le gène U3A est interrompu, le gène U3B peut prendre le relais, sans qu'il y ait de problème de croissance. Ceci est en bon accord avec notre observation d'une forte similitude de la structure secondaire des deux pré-snoRNA U3 et d'une efficacité d'épissage *in vitro* similaire de ces deux précurseurs. L'absence d'expression du gène U3B *in vivo,* en présence d'un gène U3A actif, doit être liée à une inhibition transcriptionnelle.

G2. Etude de l'évolution des gènes des snoRNA U3 chez les levures.

Comme nous l'avons mentionné ci-dessus, il était important de savoir si les introns mis en évidence dans les deux gènes U3 de *Saccharomyces cerevisiae* (Myslinski *et al.*, 1990) représentaient un cas d'espèce, ou étaient une propriété générale des levures appartenant à la sous-famille des Saccharomycetoideae. Afin de répondre à cette question, j'ai rassemblé une large collection de levures appartenant à quatre genres distincts de la sous-famille des Saccharomycetoideae : *Saccharomyces*, *Kluyveromyces*, *Pichia* et *Hansenula* (Fig. 43 et 44). J'ai mis au point des conditions de culture permettant à la fois d'étudier l'ADN génomique et le contenu en ARN de ces levures, et j'ai réalisé, au départ avec V. Ségault, puis seule, une étude sur les gènes des snoRNA U3 de ces levures. Les résultats obtenus sont décrits dans les articles n°2 et 3, déjà publiés et dans l'article n°4 préparé pour soumission à J. Mol. Biol. :

2 : "Secondary structure conservation of the U3 small nucleolar RNA introns in *Saccharomyces*" F. Brulé, A. Grégoire, V. Ségault, A. Mougin et C. Branlant. C. R. Acad. Sci. Paris Sciences de la Vie/Life sciences, 318, 1197-1206,

3 : "The yeast *Hansenula wingei* U3 snoRNA gene contains an intron and its coding sequence co-evolved with the 5' ETS region of the pre-ribosomal RNA" F. Brulé, J. Venema, V. Ségault, D. Tollervey et C. Branlant. RNA, 2, 183-197,

4 : "A high diversity of U3 snoRNA genes in yeasts. U3 snoRNA function accomodates to great variations of the 3' domain" F. Brulé, R. Fournier, V. Ségault, A. Grégoire, A. Mougin et C. Branlant.

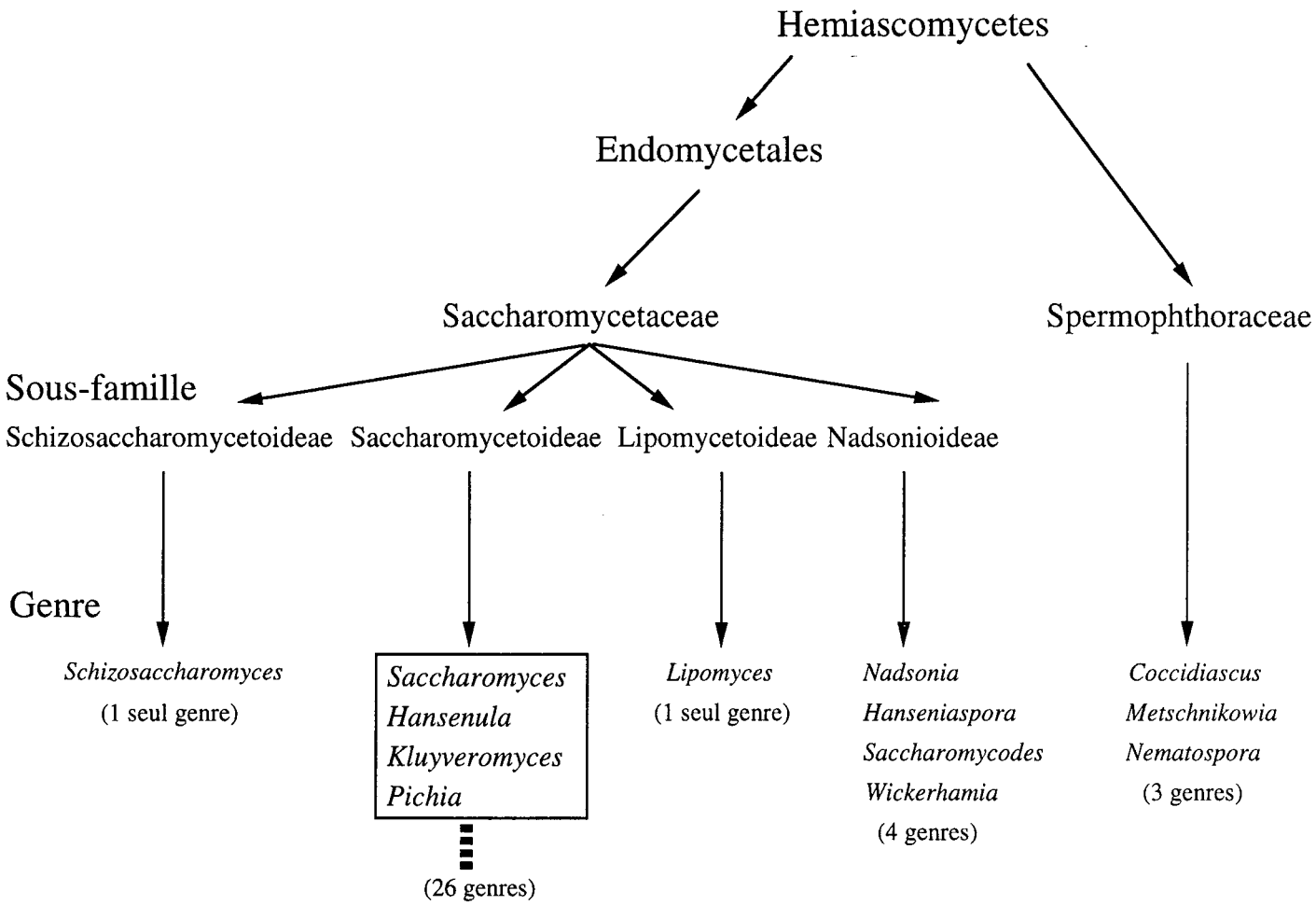


Figure 43 : Relations phylogénétiques entre les levures selon Kreger van Rij, 1987.

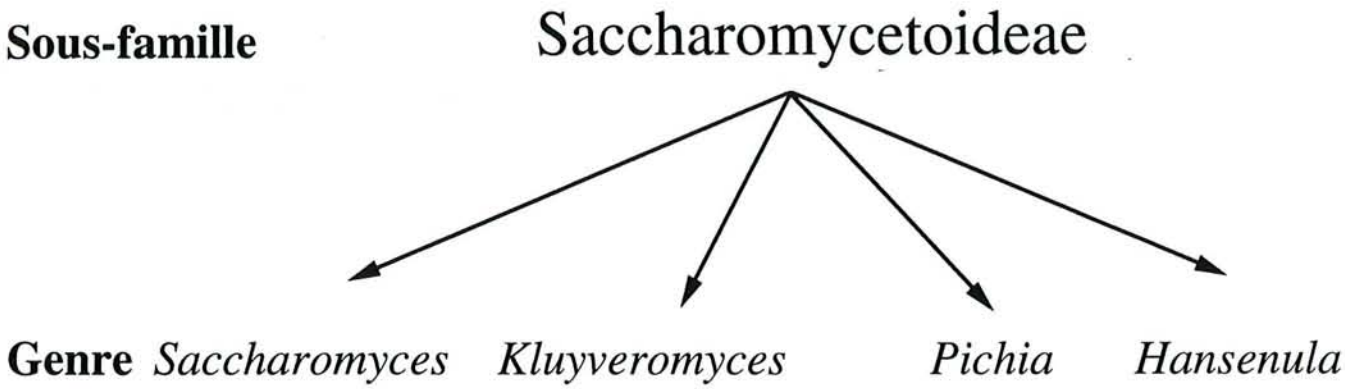


Figure 44 : Liste des espèces et souches de levures de la sous-famille des Saccharomycetoideae, étudiées dans ce travail.

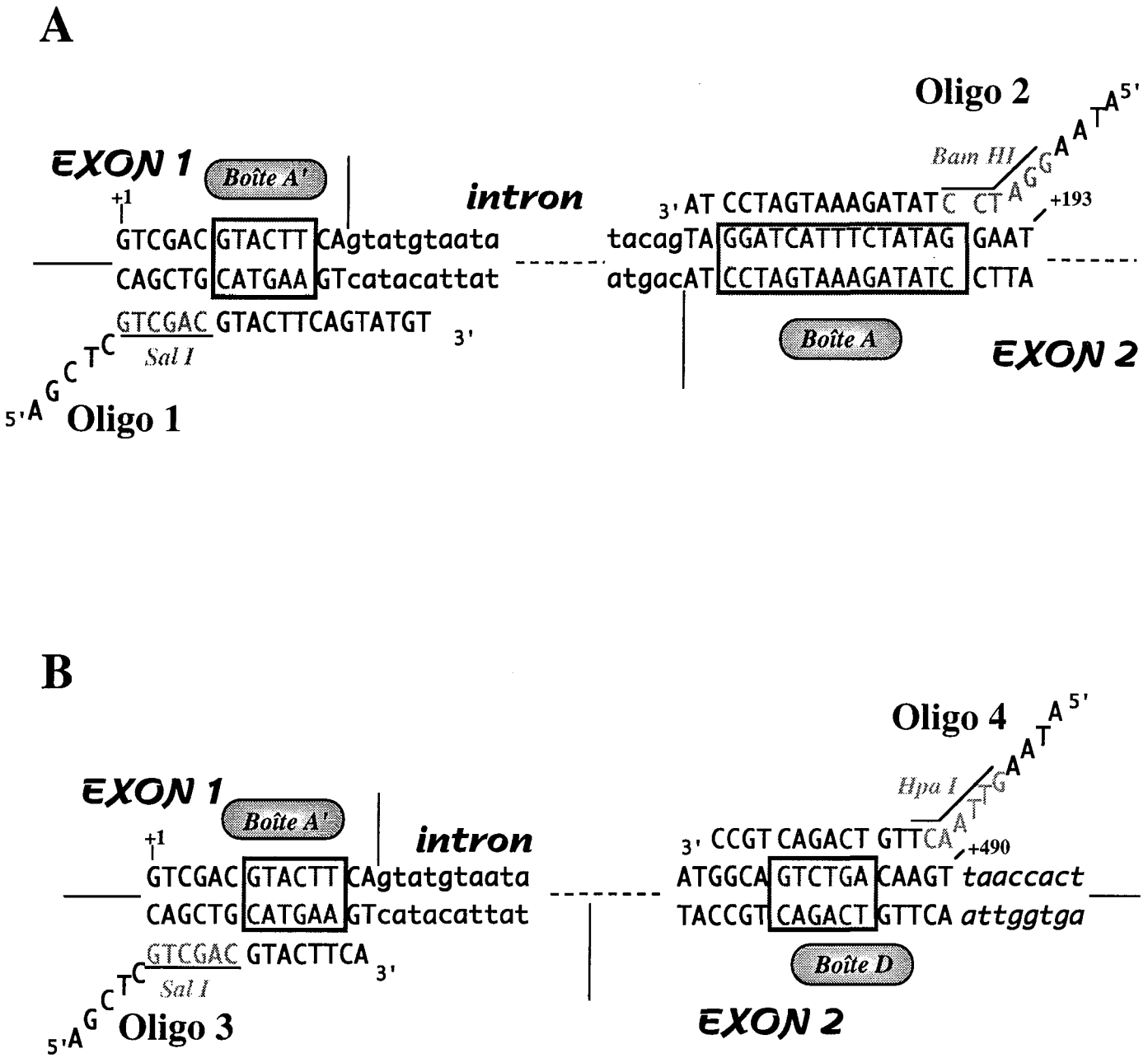


Figure 45 : Oligonucléotides utilisés pour réaliser les amplifications par PCR à partir de l'ADN génomique de levure.

A. Les oligonucléotides 1 et 2, correspondant respectivement aux boîtes phylogénétiquement conservées A' et A, ont été utilisés pour les amplifications par PCR des introns des gènes U3.

B. Les oligonucléotides 3 et 4, correspondant respectivement aux boîtes phylogénétiquement conservées A' et D, sont utilisés pour les amplifications des gènes U3

Les exons 1 et 2 sont en lettres majuscules, et l'intron en lettres minuscules.
La région non codante 3' est indiquée en lettres minuscules italiques.
Les boîtes phylogénétiquement conservées A', A et D sont encadrées en rouge.
Les sites de restriction portés par les oligonucléotides amorces utilisés pour la PCR sont en bleu

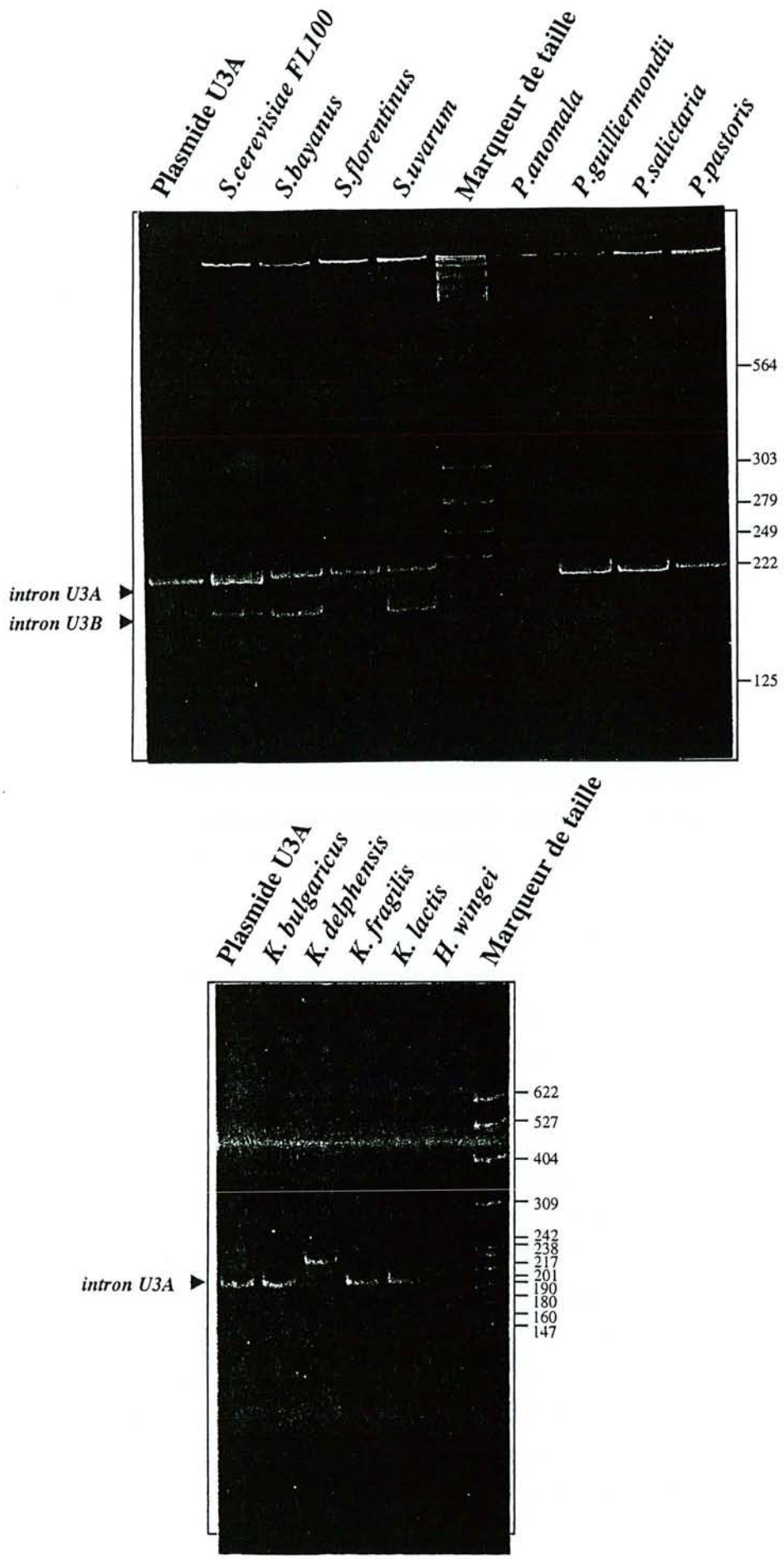


Figure 46 : Fractionnement en gel de polyacrylamide non dénaturant à 10% des produits d'amplification des introns des gènes codant pour le snoRNA U3, à partir de l'ADN génomique de différentes levures.

Le nom de chaque levure est indiqué en haut des gels.
Le plasmide pFL::snR17A (Plasmide U3A) a été utilisé comme témoin de la réaction d'amplification.
Les tailles (en paires de bases) des fragments d'ADN utilisés comme marqueurs de taille sont indiquées à droite des gels.
Le positionnement des introns U3A et U3B de *S. cerevisiae* est indiqué sur la gauche du gel.

Dans un premier temps, j'ai développé une méthode basée sur la technique PCR, permettant de détecter la présence d'introns dans les gènes du snoRNA U3, localisés à la même position dans la région codante que les introns des gènes U3A et U3B de *S. cerevisiae*. Pour cela, nous avons mis à profit la présence de boîtes phylogénétiquement conservées dans le snoRNA U3, situées de part et d'autre de l'intron dans les gènes U3A et U3B de *S. cerevisiae*. La boîte A' est située en 5' et la boîte A en 3' (Fig. 35 et 45).

Dans toutes les levures étudiées, cette approche nous a permis de détecter des introns, supposés être des introns de gènes de snoRNA U3 (Fig. 46). Un produit d'amplification de même taille que celui obtenu avec le gène U3A de *S. cerevisiae* était observé dans 12 des 14 souches étudiées. L'analyse de la séquence de ce produit d'amplification pour chacune des 12 espèces, après clonage du produit amplifié dans l'ADN du phage M13mp9 et séquençage de 5 à 10 phages recombinants pour chaque fragment amplifié, a montré qu'il s'agissait bien d'un intron et qu'il était identique à celui du gène du snoRNA U3A de *S. cerevisiae*. Nous avons seulement détecté chez certaines espèces de *Saccharomyces*, un second allèle de l'intron U3A (U3A*) différant par quelques substitutions de bases. Ce haut degré de conservation était complètement inattendu pour un intron, surtout sur la base du haut degré de divergence des introns des gènes U3A et U3B de *S. cerevisiae*.

Dans deux des espèces étudiées, *Hansenula wingei* et *Kluyveromyces delphensis*, aucun produit d'amplification correspondant à l'intron du gène U3A n'était obtenu, mais un produit de taille différente était amplifié.

Pour les espèces appartenant au genre *Saccharomyces*, un second produit d'amplification était toujours obtenu, ayant soit la taille du produit amplifié à partir du gène U3B de *S. cerevisiae* (chez *S. bayanus* et *S. uvarum*), soit une taille inférieure (chez *S. florentinus*). Enfin, dans le cas de la levure *S. uvarum*, une levure polyploïde utilisée en brasserie, nous avons observé une plus grande multiplicité des produits de l'amplification. L'analyse de la séquence des produits amplifiés, pour *S. bayanus* et *S. uvarum* ayant la même taille que le produit issu d'amplification du gène du snoRNA U3B de *S. cerevisiae*, montrait une séquence identique à celle de l'intron U3B de *S. cerevisiae*. Donc ce deuxième intron uniquement présent chez les levures du genre *Saccharomyces*, a une séquence très conservée au sein de ce genre. Il n'est cependant pas présent dans toutes les espèces, puisque chez *S. florentinus*, il est remplacé par un intron de taille plus faible (76nt) (Fig. 47), qui a une boîte de branchement canonique au lieu de la boîte GACTAAC des introns des gènes U3A et U3B.

L'analyse de la séquence des deux produits d'amplification supplémentaires chez *S. uvarum* a montré qu'il s'agissait de deux introns (U3C et U3D) qui semblent dérivés

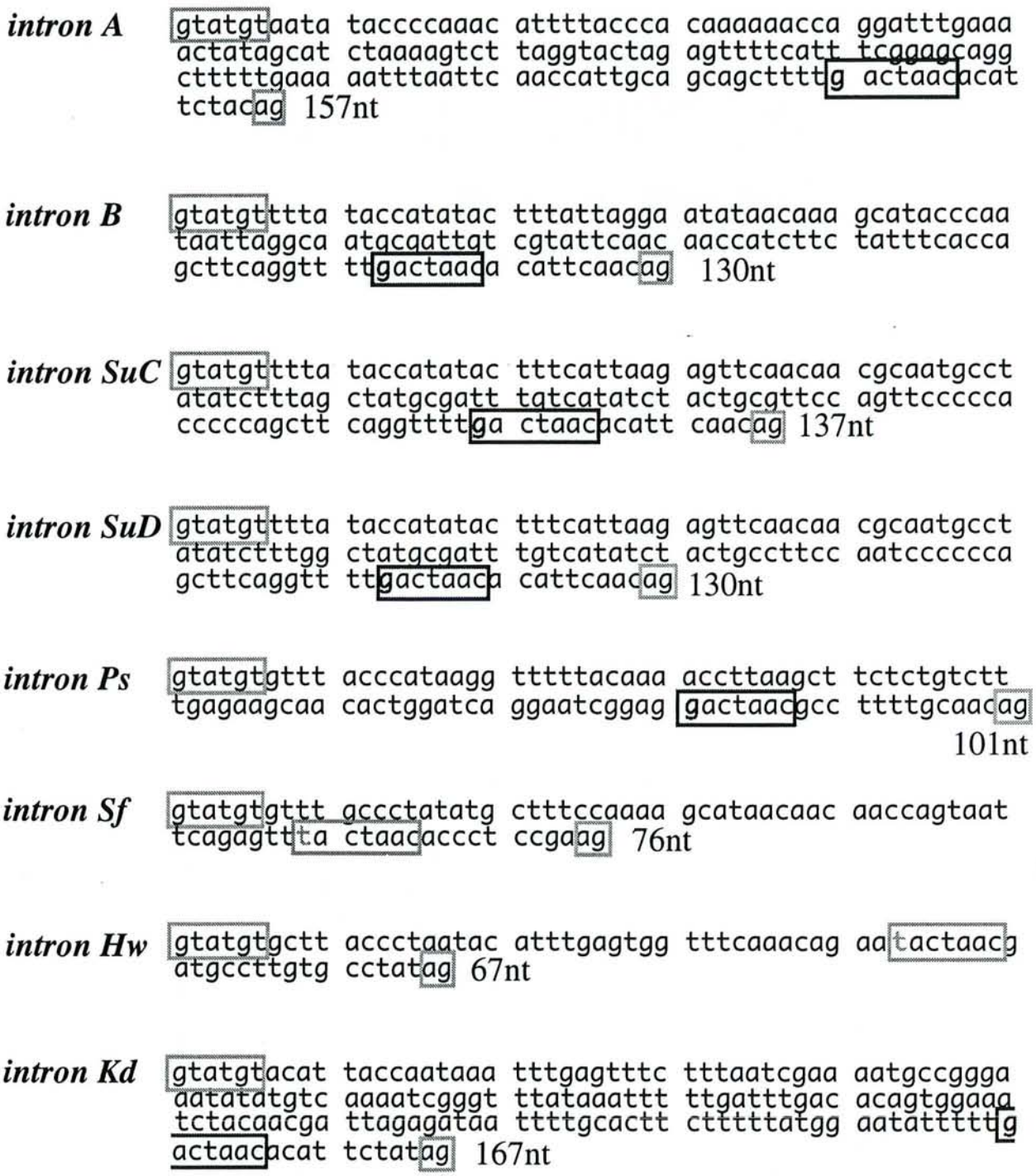


Figure 47 : Séquence des introns des gènes de snoRNA U3, mis en évidence chez différentes levures.

Intron A : *Saccharomyces cerevisiae* FL100,
Intron B : *Saccharomyces cerevisiae* FL100,
Intron SuC : *Saccharomyces uvarum,*
Intron SuD : *Saccharomyces uvarum,*
Intron Ps : *Pichia salictaria,*
Intron Sf : *Saccharomyces florentinus,*
Intron Hw : *Hansenula wingei,*
Intron Kd : *Kluyveromyces delphensis.*

Les boîtes 5' et 3' sont encadrées en vert.
Les boîtes de branchement sont en rouge ou en bleu selon qu'elles correspondent à la séquence GACTAAC (non canonique) ou TACTAAC (canonique).

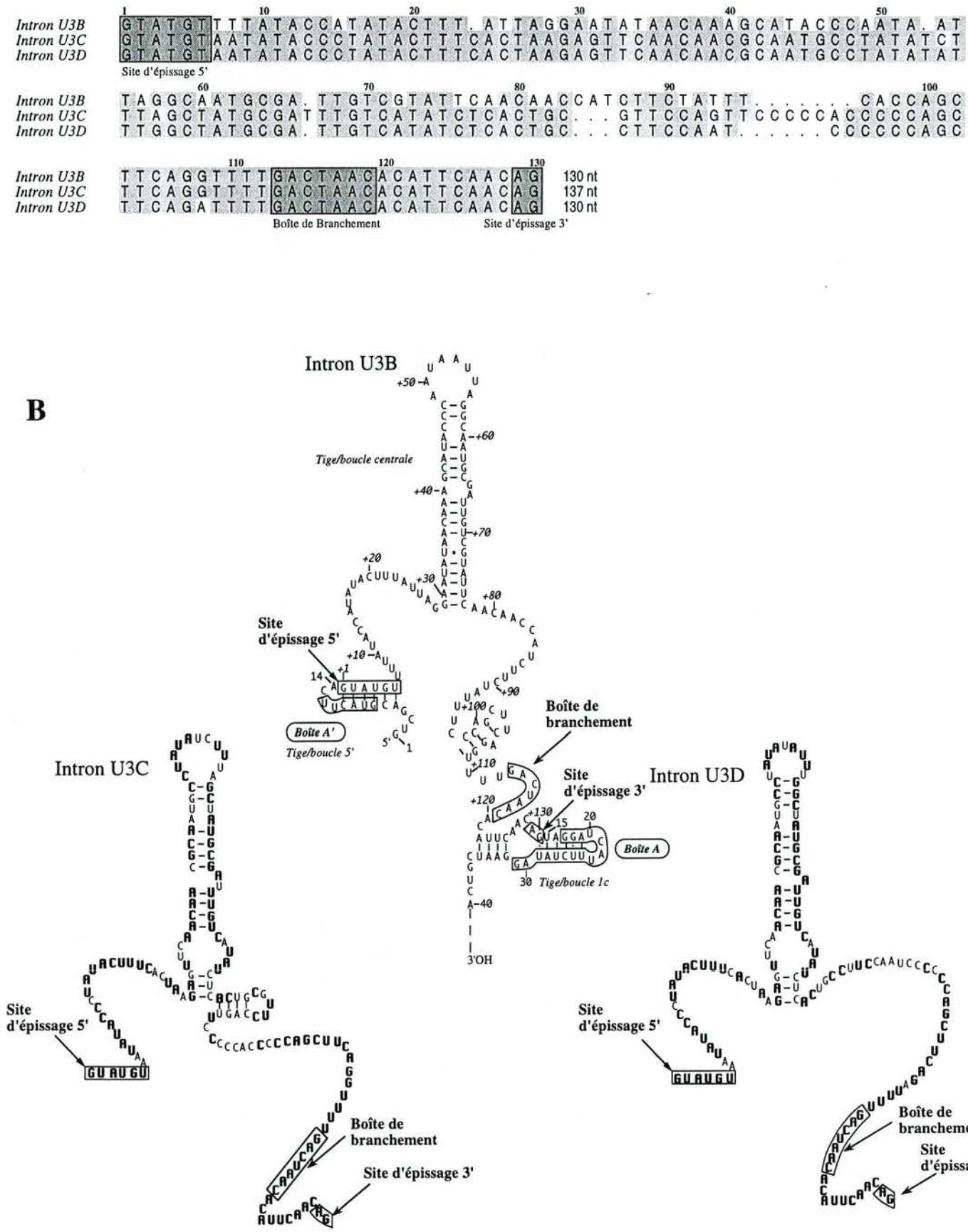


Figure 48 : Les introns U3B, U3C et U3D de *Saccharomyces uvarum*.

A. Alignement des séquences des introns U3B, U3C et U3D mis en évidence chez *S. uvarum*.
Les nucléotides conservés dans les trois séquences sont entourés de gris clair.
Les signaux d'épissage (boîtes 5', de branchement et 3') sont représentés en gris foncé.

B. Modèle de structure secondaire des introns U3B, U3C et U3D de *S. uvarum*.
Les nucléotides conservés entre les trois introns sont indiqués en gras.
Les signaux d'épissage (boîtes 5', de branchement et 3') sont encadrés.

de l'intron du gène U3B d'après leur homologie de séquence avec cet intron (Fig. 48). La complexité du génome de cette levure a peut-être conduit à ne conserver la séquence que d'un seul des allèles du gène U3B, deux autres allèles ayant dérivé en séquence. Néanmoins, il est intéressant de constater que malgré cette dérive en séquence, la possibilité de replier l'intron, selon le modèle de structure établi pour les introns des gènes U3A et U3B, est conservée pour les introns U3C et U3D. En particulier, la possibilité de former la longue structure tige/boucle centrale est préservée par des mutations compensatoires (Fig. 48).

Cette observation suggère que ces introns appartiennent à des gènes fonctionnels et renforce l'idée d'un rôle très important de cette tige/boucle centrale. Un autre argument en faveur du rôle de cette tige/boucle centrale est apporté par la comparaison des deux allèles d'introns de gènes U3A chez *S. cerevisiae* et *S. bayanus*. Les trois substitutions de bases observées sont telles qu'aucune altération de la structure tige/boucle centrale n'est observée (Fig. 42). La présence de ces deux allèles des gènes U3A chez certaines espèces du genre *Saccharomyces*, n'a pas été retrouvée chez les levures des genres *Pichia* et *Kluyveromyces*. Ceci doit résulter du fait que les levures du genre *Saccharomyces* sont diploïdes, alors que celles des genres *Pichia* et *Kluyveromyces* sont haploïdes. *S. cerevisiae* FL100 et *S. cerevisiae* GY5 sont des souches homozygotes construites en laboratoire, chacune d'elle renferme un seul des deux types d'allèles de gènes U3A. Au contraire, *S. bayanus* doit être hétérozygote puisqu'elle renferme les deux formes alléliques. Les espèces des genres *Pichia* et *Kluyveromyces* contiennent un intron U3A similaire à celui présent chez *S. cerevisiae* FL100. On peut donc supposer que le second allèle du gène U3A, observé chez les levures du genre *Saccharomyces*, est apparu après la division entre les trois genres.

L'ensemble de cette étude sur les espèces du genre *Saccharomyces* est décrite dans les articles 2 et 4.

Bien que ni l'allèle U3A* de l'intron U3A, ni l'intron U3B n'ait été détecté dans les espèces étudiées des genres *Pichia*, *Kluyveromyces* et *Hansenula*, dans le cas de l'espèce *P. salictaria*, nous avons obtenu un second produit d'amplification en plus du produit d'amplification de l'intron U3A. Il correspond à un intron de plus petite taille (101nt). Cet intron ne présente aucune homologie marquante avec l'intron des gènes U3B du genre *Saccharomyces* et il contient une boîte de branchement non canonique (Fig. 47). Les deux produits d'amplification uniques obtenus pour *Kluyveromyces delphensis* et *Hansenula wingei* ont eux aussi des tailles et des séquences très différentes. Celui de *K. delphensis* correspond à un intron ayant une boîte de branchement non canonique, alors que celui de *H. wingei* renferme un intron ayant une boîte de branchement canonique (Fig. 47).

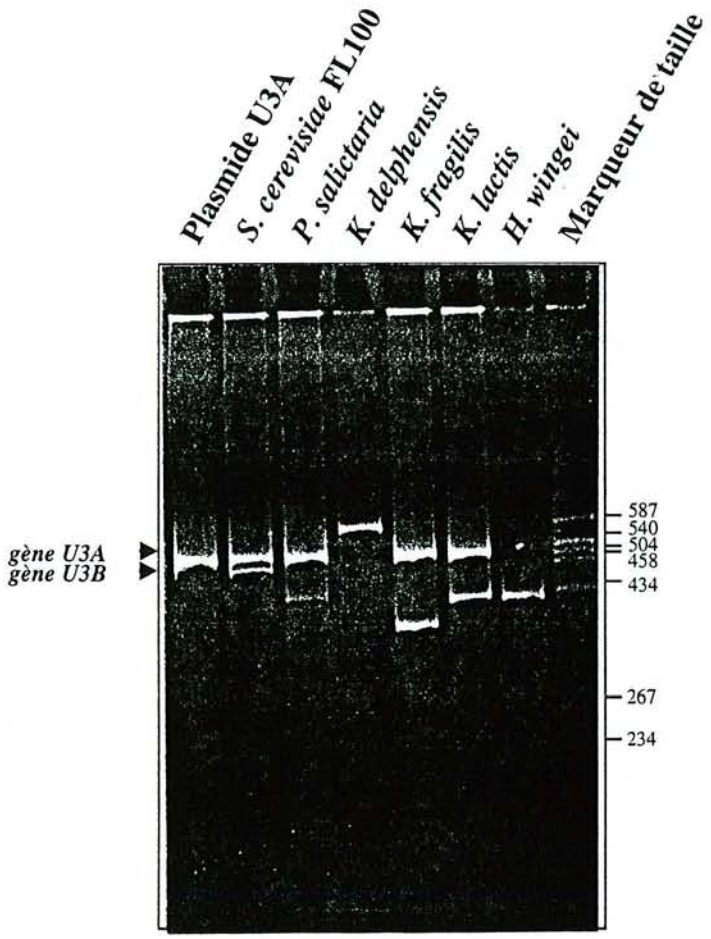


Figure 49 : Fractionnement en gel de polyacrylamide non dénaturant à 10% des produits d'amplification des gènes codant pour le snoRNA U3, à partir de l'ADN génomique de différentes levures.

Le nom de chaque levure est indiqué en haut des gels.
Le plasmide pFL::snR17A (Plasmide U3A) a été utilisé comme témoin de la réaction d'amplification.
Les tailles (en paires de bases) des fragments d'ADN utilisés comme marqueurs de taille sont indiquées à droite des gels.
Le positionnement des gènes U3A et U3B de *S. cerevisiae* est indiqué sur la gauche du gel.

Cette étude mettait donc en évidence, d'une part une conservation surprenante de l'intron du gène U3A chez les levures de la sous-famille des Saccharomycetoideae, et d'autre part une grande variété de types d'introns supposés appartenir à des gènes de snoRNA U3 chez les levures de cette sous-famille. Il restait à vérifier que ces introns appartenaient bien à des gènes de snoRNA U3.

Nous avons alors utilisé la même stratégie basée sur la méthode PCR pour amplifier les gènes de snoRNA U3 dans leur totalité. Les deux oligonucléotides amorces utilisés correspondaient aux séquences conservées aux deux extrémités de la région codante, les boîtes A' et D, respectivement (Fig. 45B). Nous avons ainsi pu montrer que la majorité des introns mis en évidence était bien contenue dans des gènes de snoRNA U3 (Fig. 49), à l'exception de l'intron court de *S. florentinus*. Aucun produit d'amplification correspondant à cet intron n'a été obtenu. Seul un gène U3A a été amplifié dans cette espèce. Il est possible que l'intron court obtenu par amplification de la région intronique à partir des boîtes A' et A, corresponde à un gène de snoRNA U3 tronqué, ou à un produit de recombinaison entre un gène de snoRNA U3 et un autre gène.

Pour l'espèce *Pichia salictaria*, où un intron U3A et un intron de séquence différente U3Ps avaient été amplifiés, nous avons bien obtenu comme attendu, l'amplification de deux gènes de snoRNA U3, un gène U3A et un gène U3Ps renfermant le second intron (Fig. 49). Pour les espèces *Hansenula wingei* et *Kluyveromyces delphensis* où un seul intron différent de l'intron U3A avait été amplifié, un seul gène de snoRNA U3 contenant cet intron a été mis en évidence. Par contre, pour les espèces *Kluyveromyces marxianus* var. *fragilis* et *Kluyveromyces marxianus* var. *lactis*, où seul l'intron U3A avait été mis en évidence, nous avons obtenu deux produits distincts d'amplification, l'un correspondant au gène U3A, l'autre à un gène de snoRNA U3 ne renfermant pas d'intron. Ces deux gènes, bien qu'issus d'espèces appartenant au même genre, ont des séquences très différentes. Au contraire, tous les gènes U3A amplifiés chez les espèces du genre *Pichia* ou du genre *Kluyveromyces* ont des séquences identiques.

L'ensemble de ces résultats est décrit dans l'article 4.

Nous retrouvions, donc, par analyse des gènes dans leur totalité, la même situation contrastée que lors de l'analyse des introns, à savoir, une conservation complète du gène U3A, aussi bien au niveau intronique qu'exonique, à l'exception de quelques mutations observées dans le second allèle présent chez certaines espèces du genre *Saccharomyces*. Ce degré inattendu de conservation contraste avec la variété des autres gènes de snoRNA U3, mis en évidence chez les différentes levures étudiées. De plus, notre étude montre

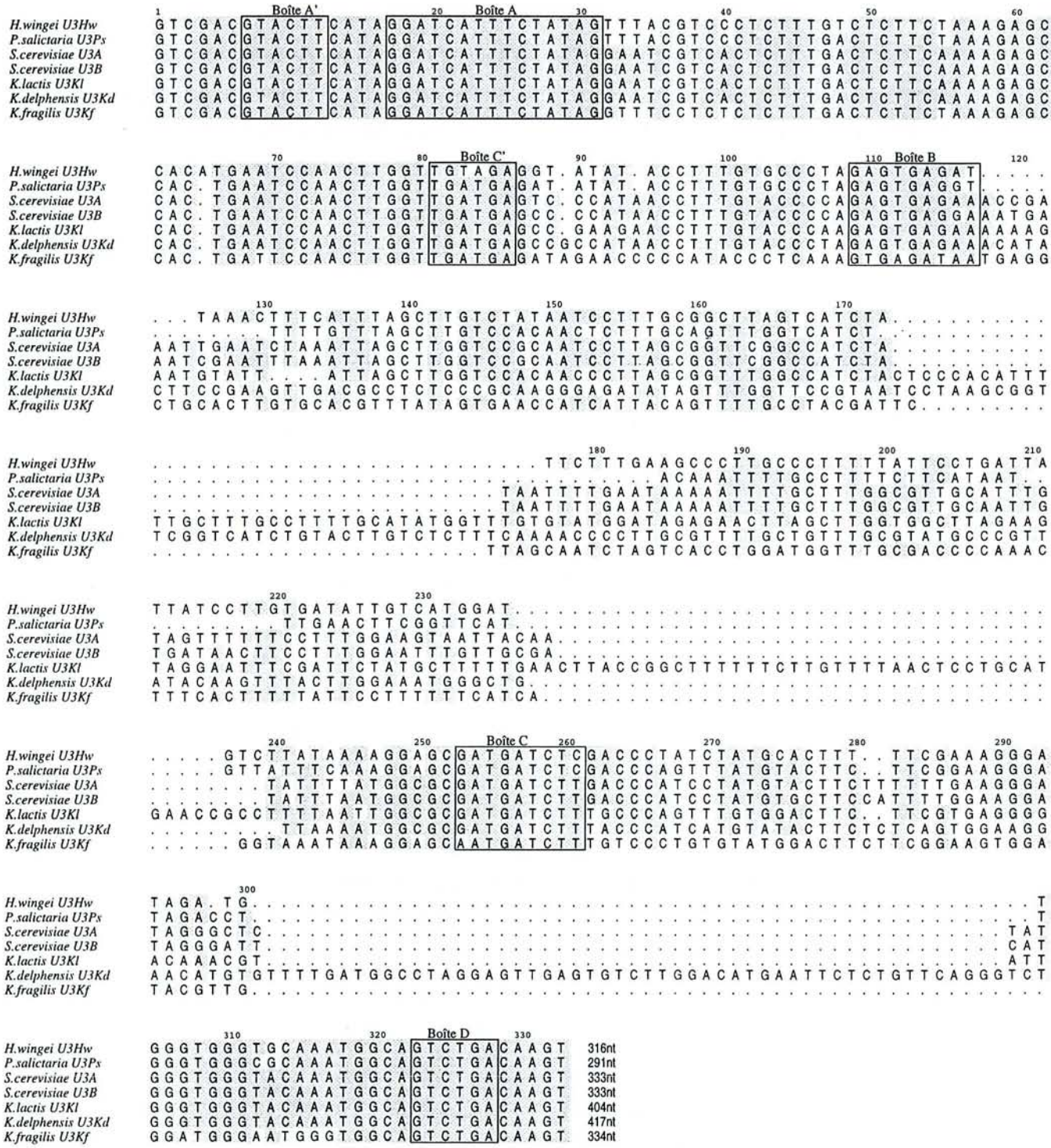


Figure 50 : Alignement des régions codantes des snoRNA U3 de levures caractérisés lors de cette étude.

Les régions codantes des snoRNA U3 de levures ont été alignées avec le programme PileUp du logiciel GCG (Genetics Computer Group, Wisconsin, USA). L'alignement a été réajusté de manière à superposer les boîtes phylogénétiquement conservées A', A, C', B, C et D.
Les nucléotides conservés dans au moins quatre séquences sont indiqués en gris.
Les boîtes A', A, C', B, C et D sont encadrées.

que des gènes U3 avec ou sans intron, peuvent être observés dans une même espèce. Par ailleurs, tous les introns mis en évidence sont toujours localisés à la même position de la région codante, même dans les gènes où, à la fois, les séquences introniques et les séquences exoniques sont différentes.

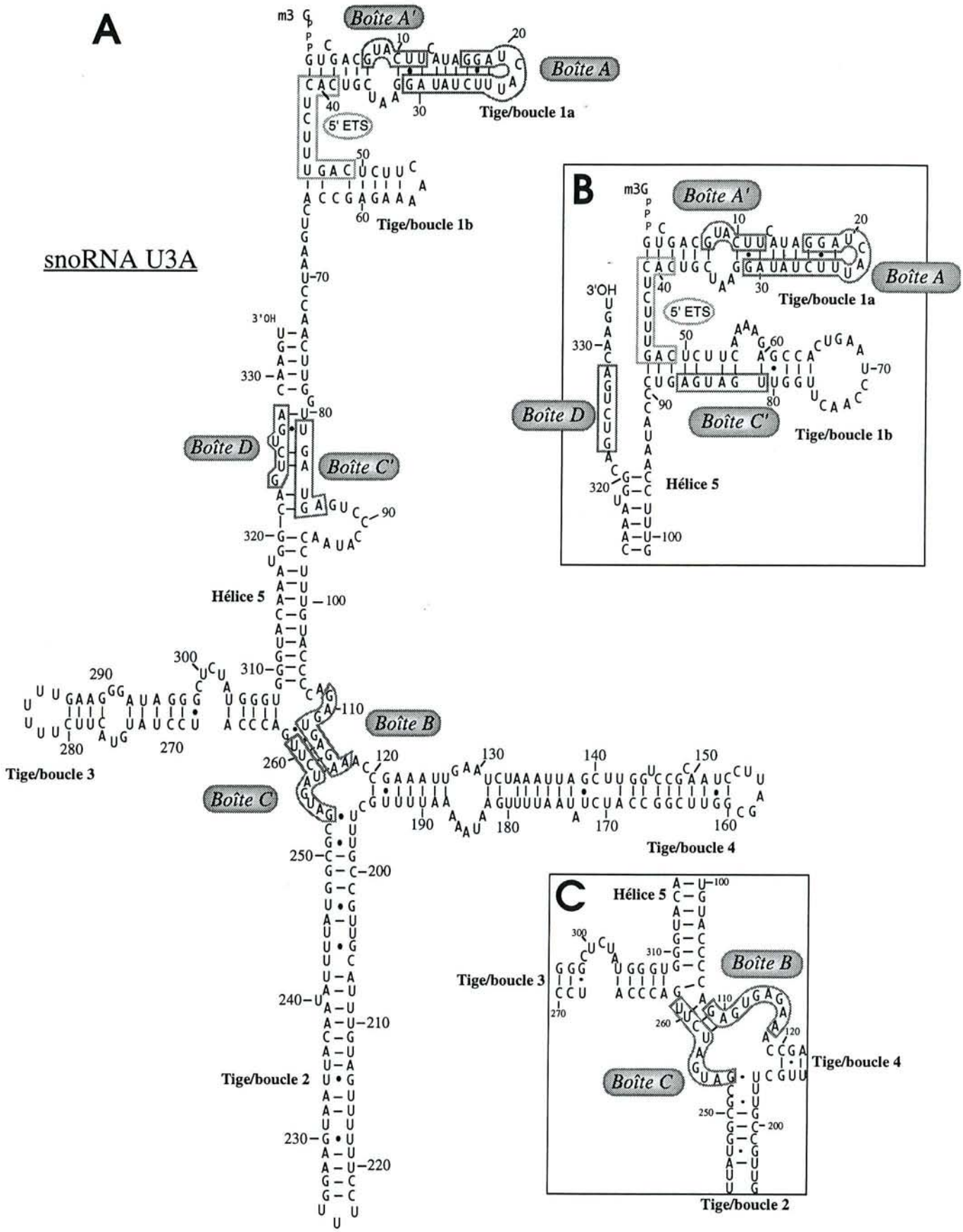
Il est intéressant de noter que l'intron est inséré dans la région 5' du snoRNA U3 supposée jouer un rôle fonctionnel important. Sur la base d'une étude génétique, J. Hughes (1996) a récemment proposé que la boîte A, qui jouxte l'intron, joue un rôle important pour la formation d'une structure pseudo-nœud dans l'ARN 17S. La localisation préférentielle des introns dans les régions fonctionnelles des ARN semble une caractéristique générale. C'est le cas pour les ARN de transfert (les introns sont localisés dans la région de l'anticodon), les ARN ribosomiques (Branlant *et al.*, 1981), les UsnRNA spliceosomaux (Tani et Ohshima, 1989, 1991; Takahashi *et al.*, 1993), et ici, pour le snoRNA U3. Cette situation est sans doute liée au fait qu'aucune insertion n'est tolérée dans ces régions fonctionnelles et que les seules insertions non létales dans ces régions sont celles éliminées par un processus d'épissage, qu'il s'agisse d'un auto-épissage ou d'un épissage catalysé en trans.

Un alignement des régions codantes des différents gènes caractérisés (Fig. 50) mettait en évidence une grande variabilité, en taille et en séquence du domaine 3'. Nous avons recherché si néanmoins, les ARN codés par ces différents gènes pouvaient se replier selon le modèle de structure secondaire proposé pour le snoRNA U3A de *S. cerevisiae*.

Un premier modèle de structure secondaire du snoRNA U3A avait été établi au laboratoire sur la base d'une analyse de l'ARN en solution (Ségault *et al.*, 1992). Cette étude avait mis en évidence l'existence de deux domaines, un domaine 5' formé de deux hélices de stabilité très moyenne et un domaine 3' très compact ayant une structure cruciforme (Fig. 51). L'étude faite récemment par A. Grégoire et A. Mougin sur la snoRNP U3 purifiée par l'équipe de R. Lührmann, étude qu'A. Méreau a complétée par une analyse de la snoRNP U3 *in vivo* par emploi du diméthylsulfate (DMS), suggérait une réorganisation d'une partie de la structure du snoRNA U3A en présence des protéines. Ce qui a conduit à proposer un second modèle de structure secondaire pour le snoRNA U3A de *S. cerevisiae*, modèle aussi compatible avec la séquence du snoRNA U3B de cette espèce (Fig. 51).

La différence entre les deux structures consiste en une extension de l'hélice 5 dans la snoRNP, extension qui inclut les boîtes phylogénétiquement conservées C' et D. Cette extension implique un remaniement de la structure tige/boucle 1b, qui est de ce fait plus courte.

Nous avons constaté que l'ensemble des ARN codés par les différents gènes de snoRNA U3 que nous avons caractérisés, peut se replier selon le modèle établi pour le


A
snoRNA U3A
m3 G
Boîte A'
Boîte A
Tige/boucle 1a
5' ETS
Tige/boucle 1b
3'OH
Boîte D
Boîte C'
Hélice 5
Tige/boucle 3
Boîte B
Boîte C
Tige/boucle 4
Tige/boucle 2
B
m3G
Boîte A'
Boîte A
Tige/boucle 1a
3'OH
5' ETS
Boîte D
Boîte C'
Tige/boucle 1b
Hélice 5
C
Hélice 5
Tige/boucle 3
Boîte B
Boîte C
Tige/boucle 4
Tige/boucle 2

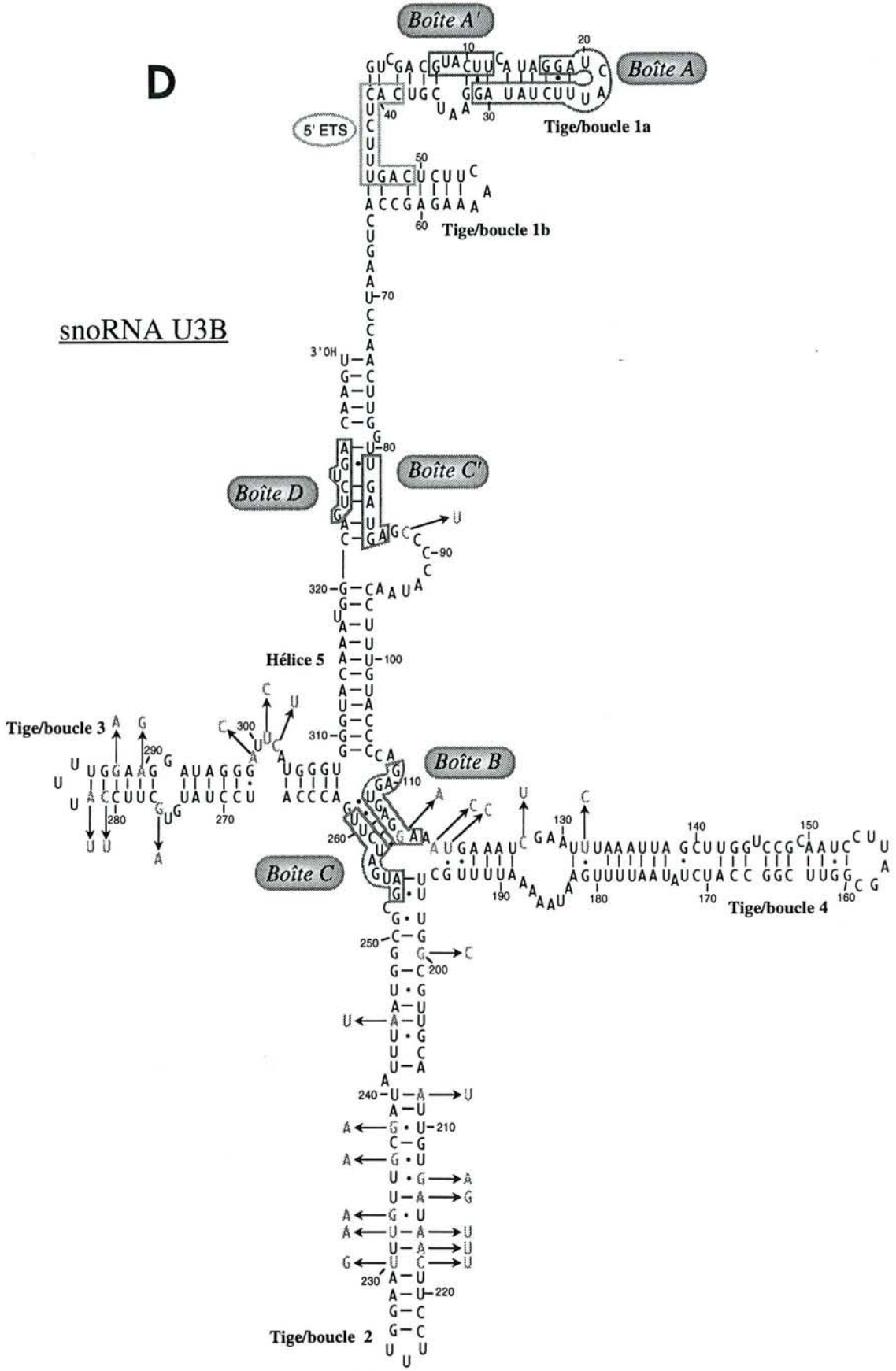


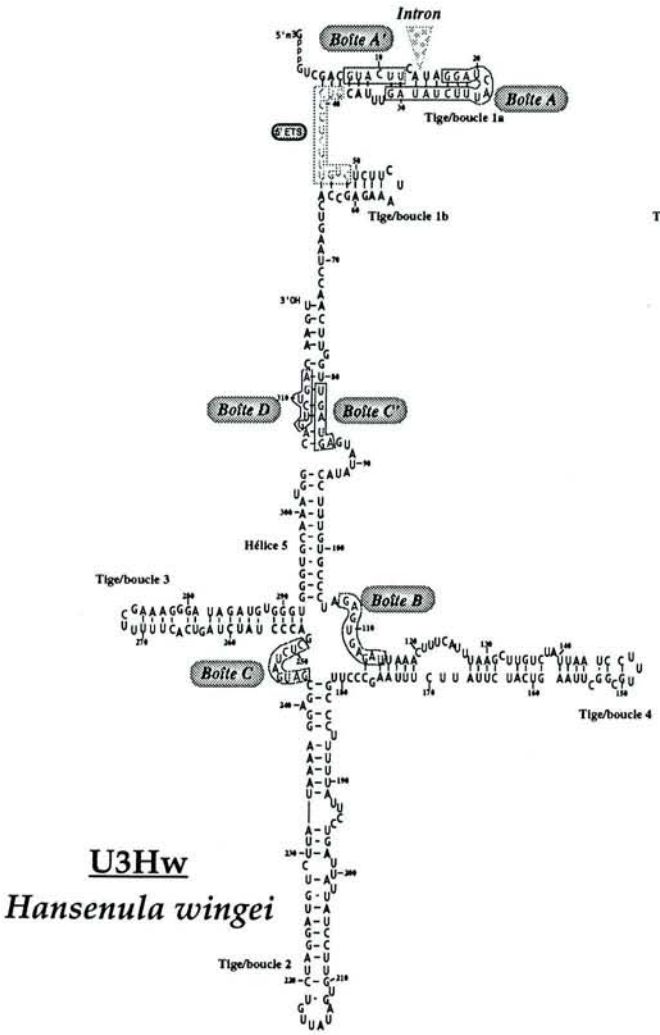
Figure 51 : Modèles de structure secondaire des snoRNA U3A et U3B de *S. cerevisiae*.

A. Ce modèle de structure secondaire est celui établi à partir de l'étude du snoRNA U3A au sein de la snoRNP U3A (A. Grégoire, résultats non publiés).

B. Structuration possible du domaine 5' du snoRNA U3A en solution (Ségault *et al.*, 1992).

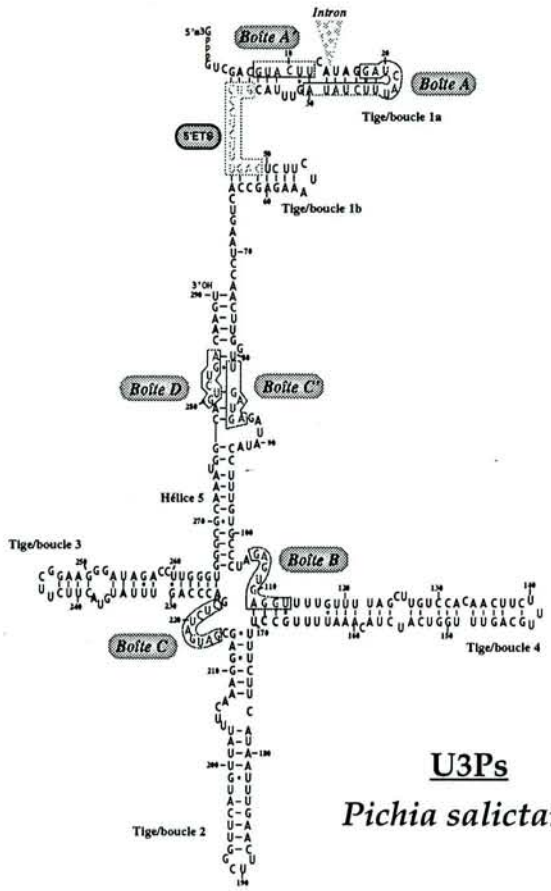
C. Interaction alternative entre les segments contenant les boîtes B et C.

D. La séquence complète est celle du snoRNA U3B, les positions où des substitutions sont observées dans le snoRNA U3A et la nature des substitutions sont indiquées en vert.



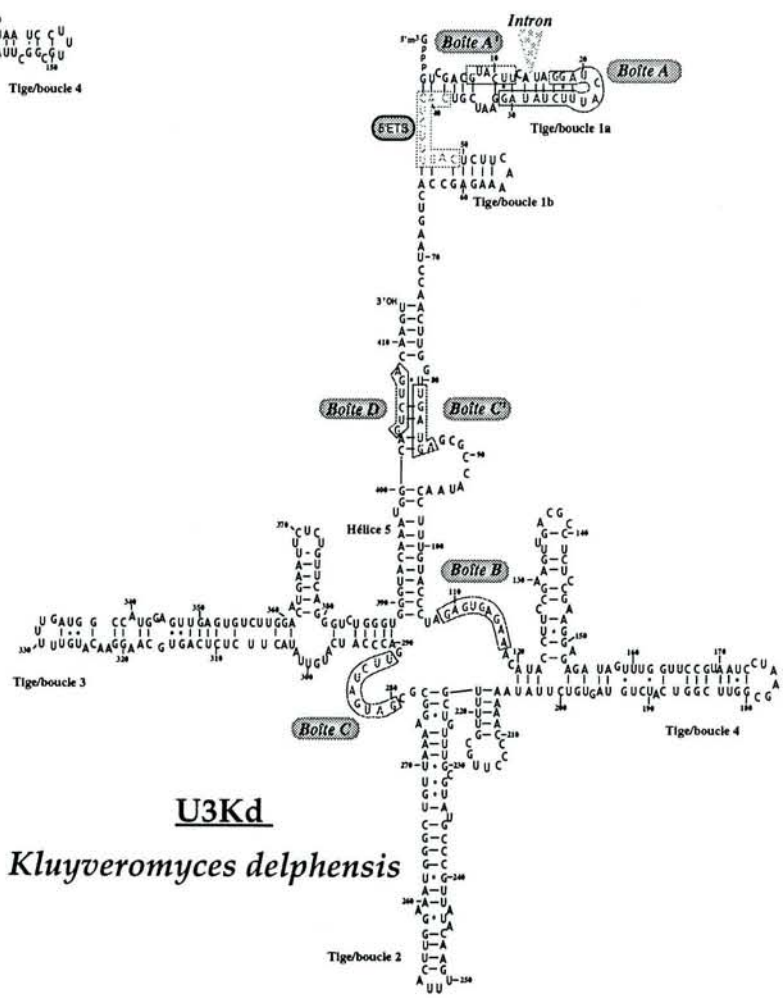
U3Hw

Hansenula wingei



U3Ps

Pichia salictaria



U3Kd

Kluyveromyces delphensis

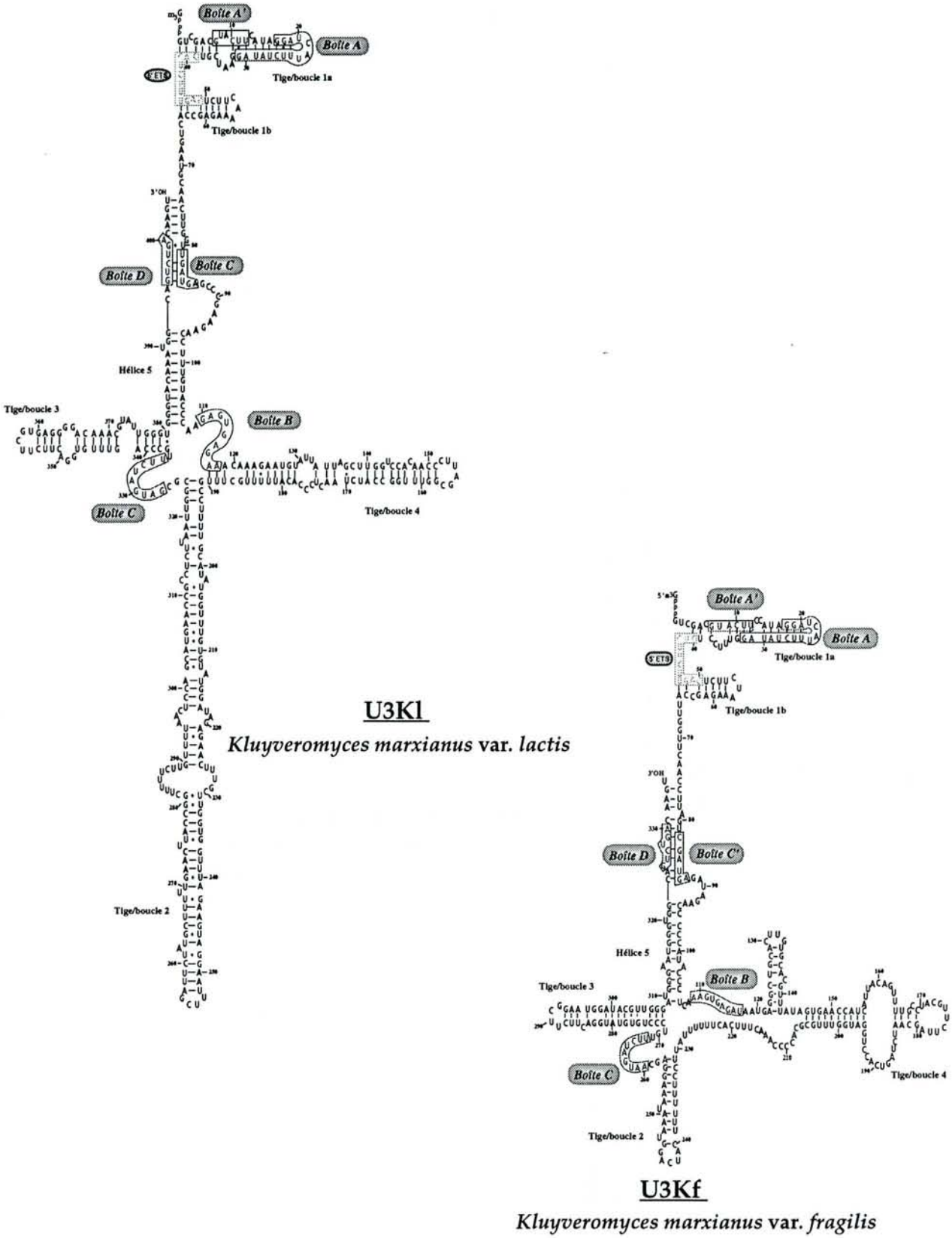


Figure 52 : Modèles de structures secondaires des snoRNA U3 extraits des levures *Hansenula wingei*, *Pichia salictaria*, *Kluyveromyces delphensis*, *Kluyveromyces marxianus* var. *fragilis* et *Kluyveromyces marxianus* var. *lactis*.

Les boîtes A', A, C', B, C et D phylogénétiquement conservées sont encadrées en rouge.
La séquence complémentaire à la région 5' ETS du pré-ARNr est encadrée en bleu.
La présence d'un intron dans les gènes codant pour ces ARN est indiquée par un triangle vert.

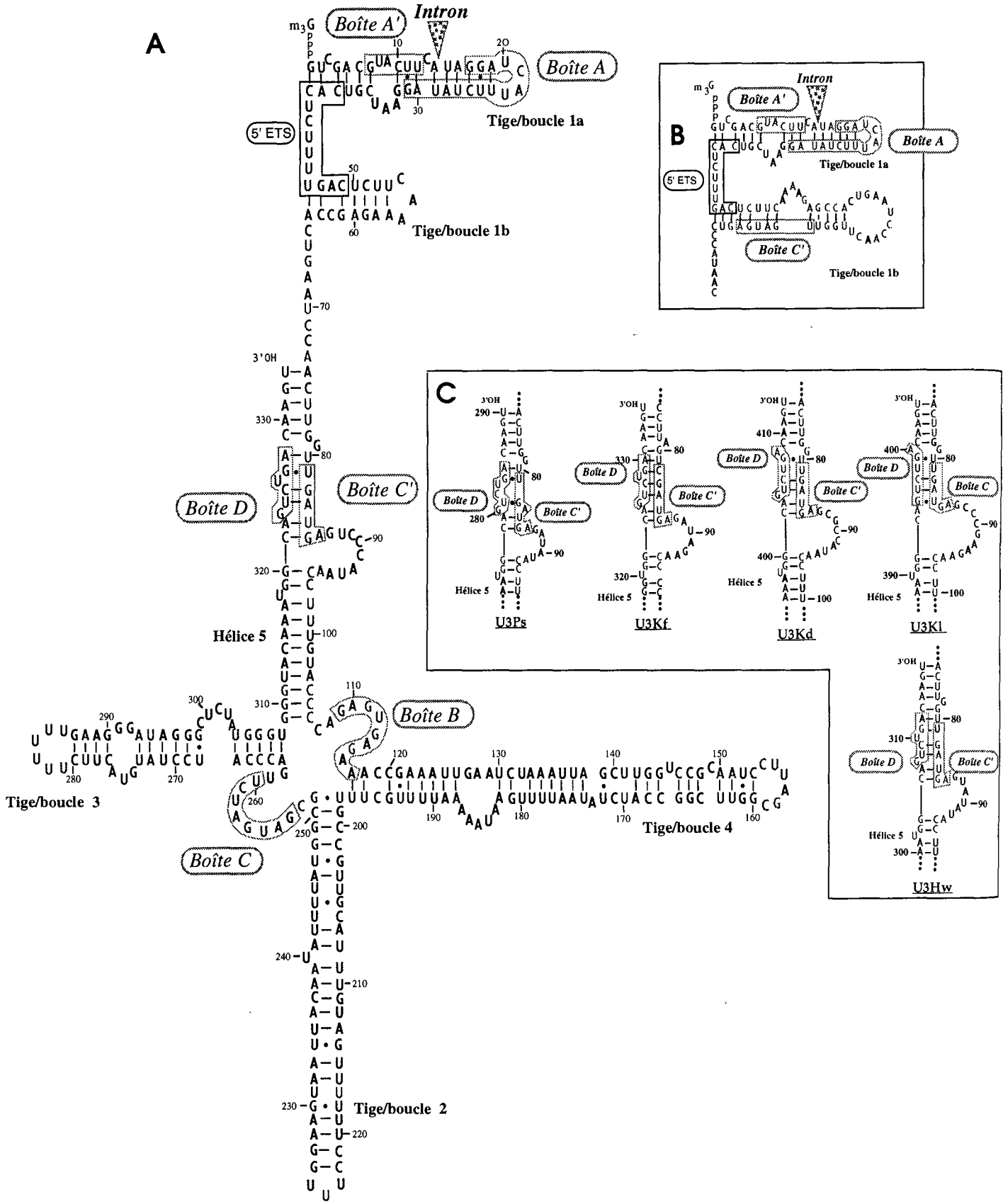


Figure 53 : Modèle de structure secondaire du snoRNA U3A de *Saccharomyces cerevisiae* sur lequel sont indiqués en rouge les nucléotides conservés entre les différents snoRNA U3 isolés des levures *Pichia salictaria* (U3Ps), *Kluyveromyces lactis* (U3Kl), *Kluyveromyces delphensis* (U3Kd), *Kluyveromyces fragilis* (U3Kf), *Hansenula wingei* (U3Hw)et *S. cerevisiae*.

A. Le modèle de structure secondaire représenté en totalité a été établi pour le snoRNA U3A de*S. cerevisiae* au sein de la snoRNP U3A (A. Grégoire, résultats non publiés).

B. Structure secondaire proposée pour le domaine 5' du snoRNA U3A de *S. cerevisiae*, après étude de cet ARN en solution (Ségault *et al.*, 1992).

C. Variations de la boucle latérale en aval de la boîte C', dans des différents snoRNA U3 de levures que j'ai caractérisés.

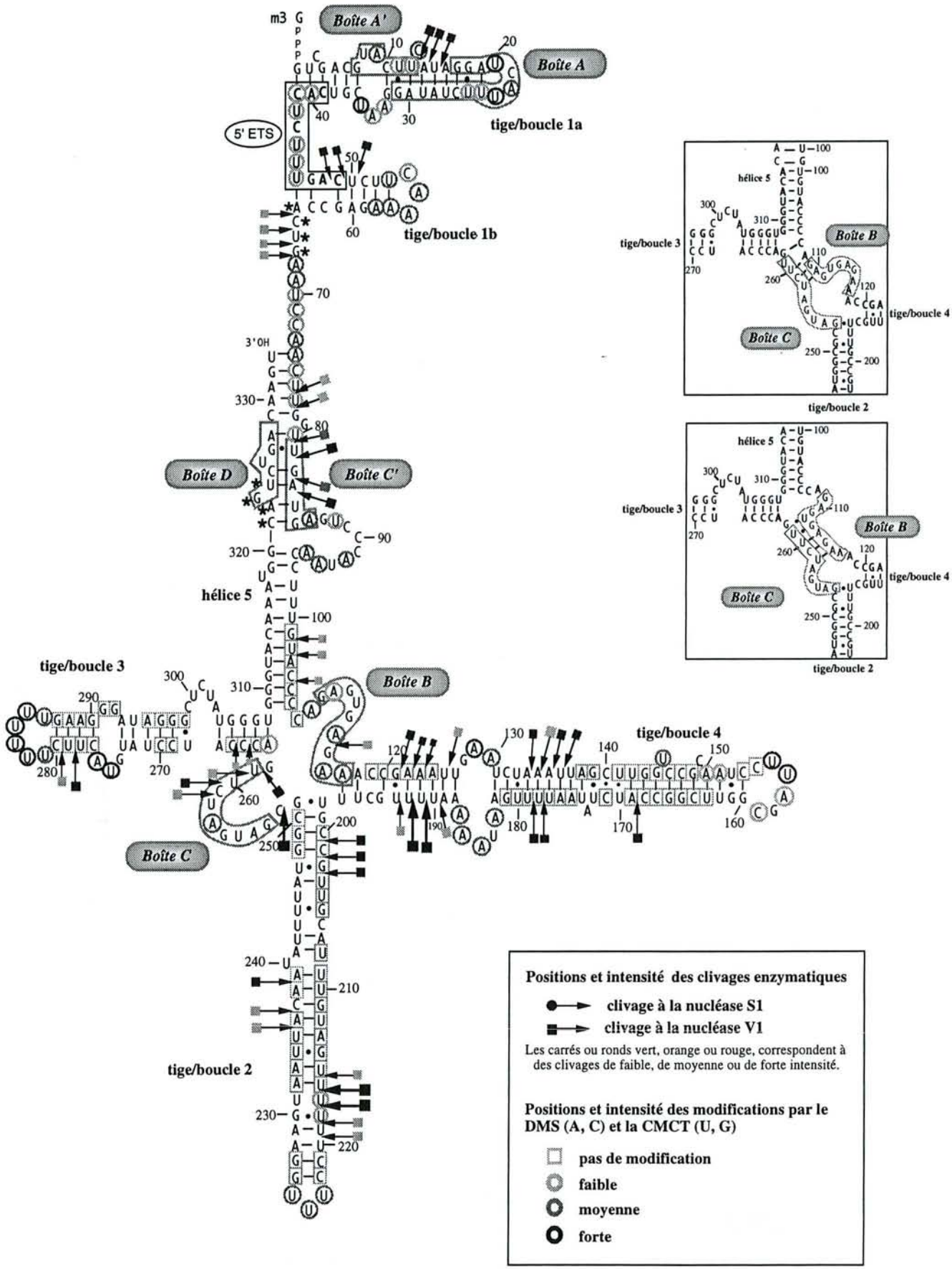


Figure 54 : Structure secondaire et accessibilité du snoRNA U3A de *S. cerevisiae* au sein de la snoRNP U3 (A. Grégoire, résultats non publiés).

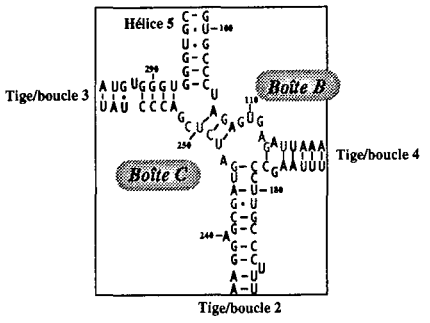
Les modifications chimiques et coupures enzymatiques sont représentées par les sigles indiqués dans l'encart en bas à droite.

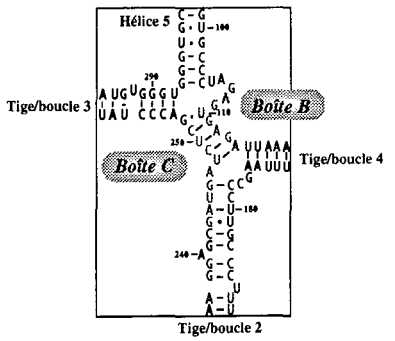
Les nucléotides qui pourraient former une structure de type pseudo-noeud sont indiqués par une astérisque.

Les 2 encarts, en haut à droite, présentent les interactions possibles entre les segments contenant les boîtes B et C.

U3Hw

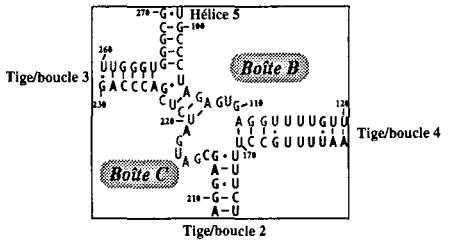
Hansenula wingei

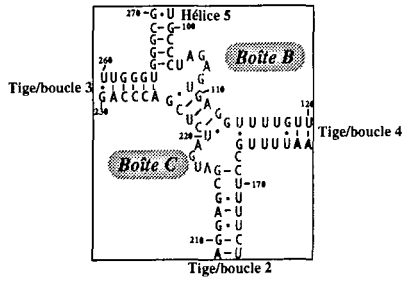




U3Ps

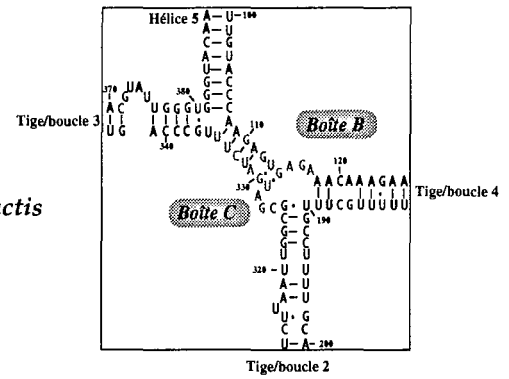
Pichia salictaria

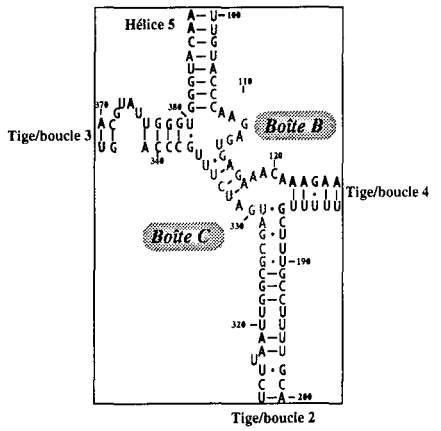




U3Kl

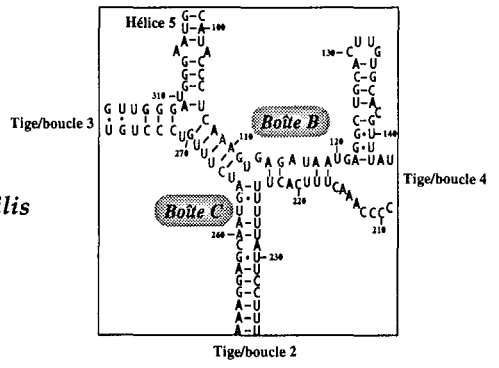
Kluyveromyces marxianus var. *lactis*

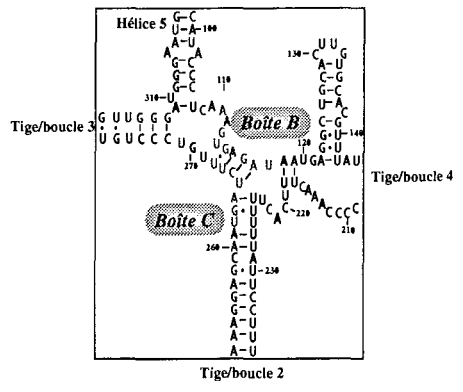




U3Kf

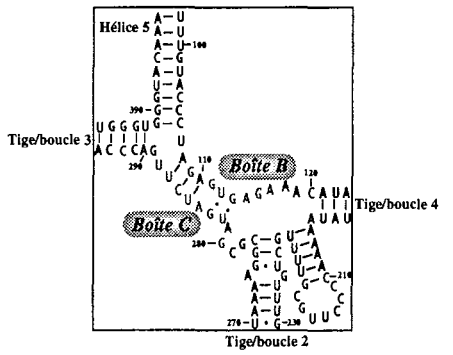
Kluyveromyces marxianus var. *fragilis*





U3Kd

Kluyveromyces delphensis



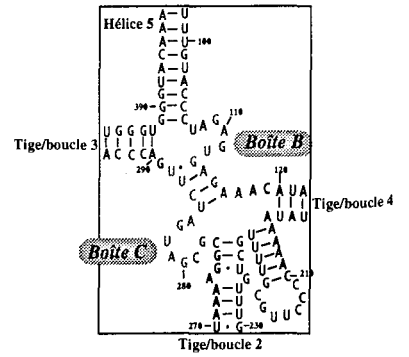


Figure 55 : Conservation des deux types d'interaction possibles entre les segments renfermant les boîtes B et C dans les snoRNA U3 caractérisés chez différentes levures (articles n°3 et 4) .

snoRNA U3A au sein de la snoRNP (Fig. 52). En particulier, au niveau de l'extension de l'hélice 5, il est intéressant de constater que la région très variable en séquence correspond à une boucle latérale (Fig. 53). La structure tige/boucle 1b, telle qu'elle était proposée sur la base de l'étude de l'ARN en solution, n'a pas un degré de stabilité conservé dans les ARN codés par les différents gènes mis en évidence, alors que la nouvelle structure 1b proposée a une stabilité similaire dans tous ces ARN.

Nous apportions donc un argument phylogénétique en faveur de la nouvelle structure proposée. Selon cette structure, le snoRNA U3 est composé d'un très court domaine 5' très conservé en séquence et formé de deux structures tige/boucle et d'un grand domaine 3' très structuré, toujours de type cruciforme, mais avec un très grand degré de variabilité en taille et en séquence. Certains des ARN codés par les gènes caractérisés, présentent même des structures tige/boucle latérales au sein des hélices 2 et 4 (Fig. 52). Une constante dans ce domaine est la présence des boîtes phylogénétiquement conservées C' et D, qui semblent interagir entre elles par un appariement de faible stabilité, ainsi que la présence des boîtes phylogénétiquement conservées B et C qui relient les hélices de la structure cruciforme.

V. Ségault (Ségault *et al.*, 1992) avait observé des coupures par la RNase de venin de cobra (RNase V1), lors de l'étude du snoRNA U3A en solution, dans le segment comportant la boîte C. Ceci a été confirmé par A. Grégoire lors de l'étude de la snoRNP U3A, une coupure au niveau de la boîte B étant aussi observée dans ce cas. Une des séquences clivées en 3' de la boîte C, UCU (positions 258 à 260) est complémentaire de deux séquences AGA présentes dans la boîte B (Fig. 54). L'une d'elles, entre les positions 114 et 116 est coupée par la RNase V1. Un appariement entre ces deux séquences pouvait donc expliquer les coupures observées. L'examen des différents snoRNA U3 que j'ai caractérisés montre que les deux possibilités d'appariement entre les régions renfermant les boîtes B et C sont conservées dans tous ces ARN. Ce qui soulève le problème de l'existence possible d'appariements alternatifs au centre de la structure cruciforme (Fig. 55).

Une autre coupure forte par la RNase V1 a été observée, aussi bien dans l'ARN en solution que dans la snoRNP, entre les positions 251 et 252 en 5' de la boîte C (Fig. 54). Nous avons observé, en comparant les différents ARN mis en évidence, qu'un prolongement de l'hélice 4 comportant une paire de bases G·U ou G-C et une paire de pyrimidines pouvait être proposé pour tous ces ARN, sauf le snoRNA U3Kd de *Kluyveromyces delphensis*. Ce prolongement de l'hélice 4 qui pourrait être stabilisé par un empilement des hélices 2 et 4, pourrait peut-être expliquer la forte coupure par la RNase V1 entre les positions 251 et 252.

Les structures montrées sur la figure 52, sont donc loin de représenter toute la complexité de la structure du snoRNA U3 de levure. Sur la base de l'apparition de

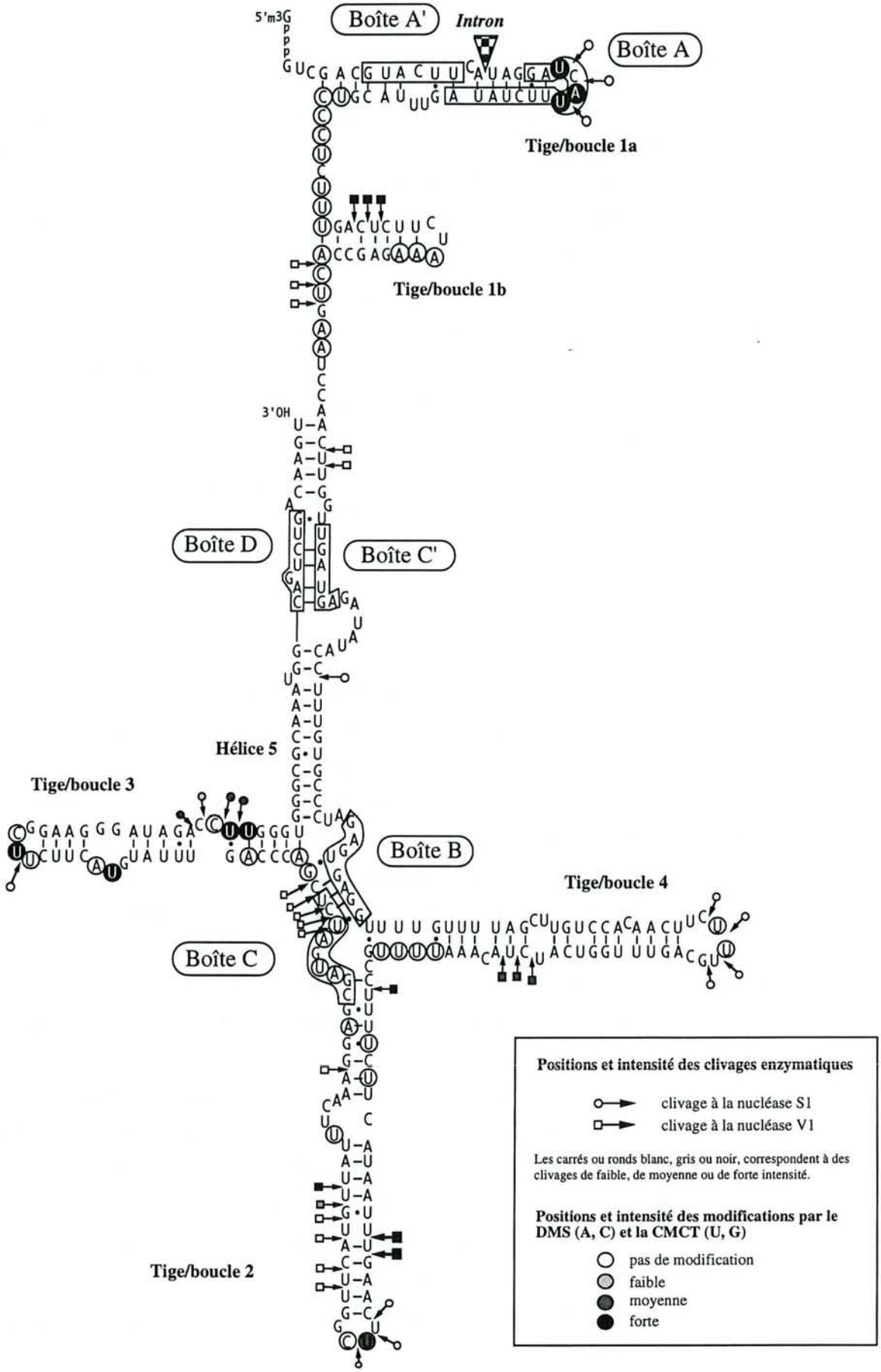


Figure 56 : Modèle de structure secondaire du snoRNA U3Ps de *Pichia salictaria*.

Ce modèle a été établi sur la base d'une étude expérimentale basée sur l'emploi de sondes chimiques et enzymatiques de la structure des ARN. Les résultats obtenus avec ces sondes sont schématisés selon le code indiqué ci-dessus.
Une des interactions possibles entre les segments de la boîte B et la boîte C est représentée.

coupures par la RNase V1 au sein de la snoRNP U3A dans le segment ACUG (64 à 67) en simple-brin dans le modèle présenté dans la figure 54, la formation d'une structure de type pseudo-nœud entre cette séquence et la séquence CAGU (321 à 324) en 5' de la boîte D a aussi été proposée. La possibilité de former cette structure de type pseudo-nœud est conservée dans les différents ARN caractérisés, puisque les deux séquences sont conservées (Fig. 52).

J'ai étudié expérimentalement la structure du snoRNA U3Ps de *Pichia salictaria* et les résultats obtenus (Fig. 56) sont en bon accord avec le modèle proposé sur la base de la comparaison avec le snoRNA U3A. L'analyse expérimentale de la structure des autres snoRNA U3 caractérisés est en cours.

D'ores et déjà sur la base des modèles établis, il ressort clairement que le domaine 5' supposé être le domaine fonctionnel interagissant directement avec l'ARN pré-ribosomique est très conservé, alors que le domaine 3' renfermant les boîtes phylogénétiquement conservées supposées interagir avec les protéines est beaucoup plus variable.

L'étude de la snoRNP U3A de *S. cerevisiae* réalisée par A. Grégoire a confirmé cette notion de domaine 3' structural servant à fixer les protéines. En effet, dans la snoRNP U3A, seules les boîtes B, C, C' et D sont protégées par les protéines, l'ensemble du domaine 5' restant accessible à l'action des nucléases et des agents chimiques, en particulier, les boîtes A' et A.

Sur la base d'une étude par formation de pontage covalent entre séquences d'ARN appariés, par un dérivé du psoralène, M. Beltrame et D. Tollervey (1992) avaient proposé une interaction entre la séquence de 10nt $_{5'}$CACUCUUUGAC$_{3'}$ (positions 39 à 49) du snoRNA U3A et la région 5' ETS du pré-ARN ribosomique. Le segment 39 à 49 du snoRNA U3 est en grande partie sous forme simple-brin dans le modèle de structure secondaire proposé (Fig. 51). Il est conservé en séquence dans les snoRNA U3 caractérisés, à l'exception de deux substitutions de bases dans le snoRNA U3Hw de *Hansenula wingei* et d'une substitution de bases dans le snoRNA U3Ps de *Pichia salictaria* (Fig. 50).

Par séquençage par la transcriptase inverse des snoRNA U3 contenus dans un extrait total d'ARN de *H. wingei* et de *P. salictaria*, j'ai pu montrer que ces deux ARN étaient les seuls exprimés dans chacune de ces espèces. Pour *H. wingei*, c'était attendu du fait que nous avions détecté un seul gène, pour *P. salictaria*, le résultat est plus inattendu puisqu'un gène U3A est présent. Les deux mutations trouvées dans *H. wingei* posaient le problème de la fonctionnalité du snoRNA U3Hw. Nous avons alors séquencé la région 5' ETS du pré-ARNr de *H. wingei* et montré que les deux mutations présentes dans le snoRNA U3Hw étaient compensées par deux substitutions dans le segment

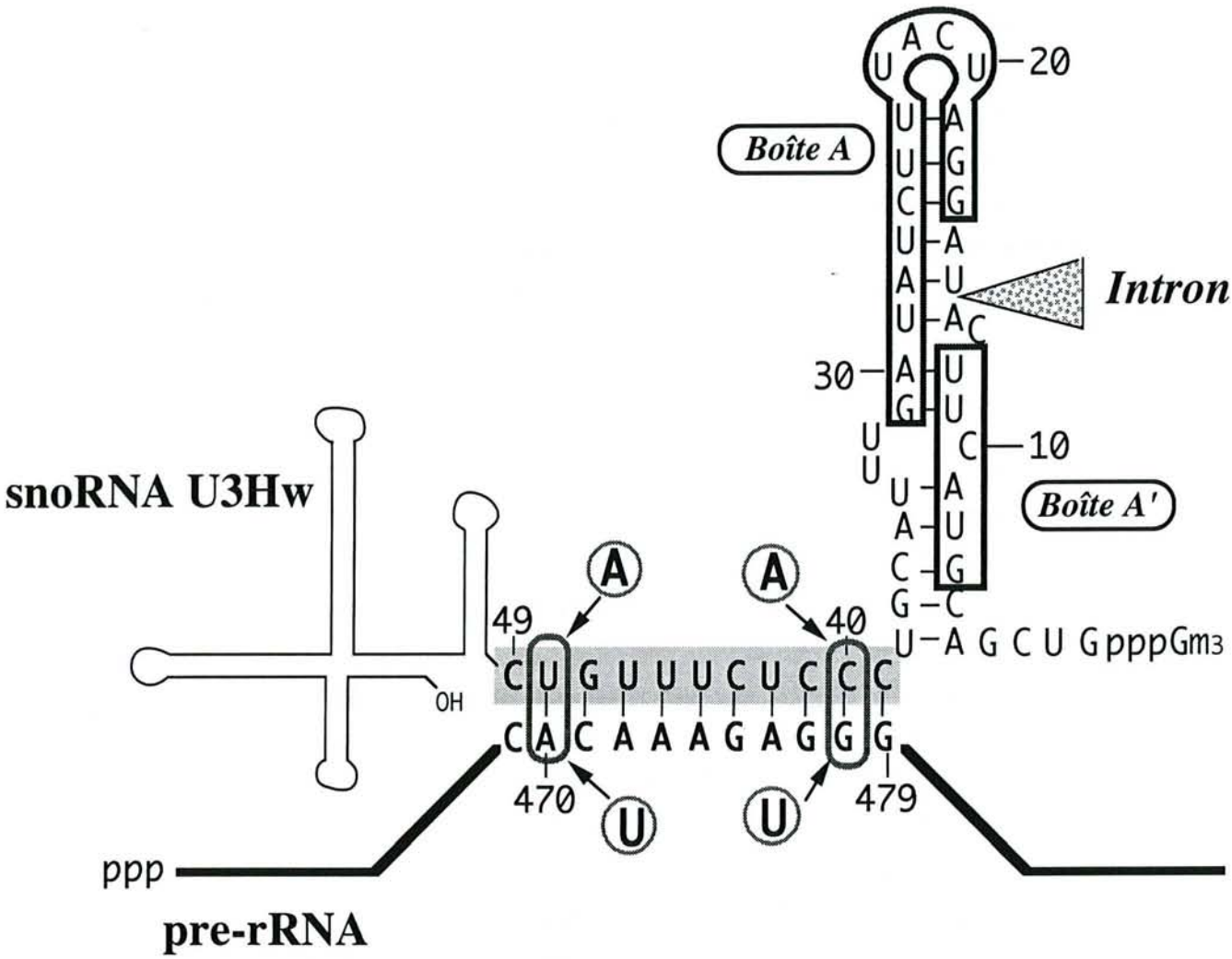


Figure 57 : Conservation de la complémentarité entre le snoRNA U3 et l'ARN pré-ribosomique chez la levure *Hansenula wingei* par mutations compensatoires (article n°3).

La région +39-+49 du snoRNA U3, en simple-brin, complémentaire du segment +470-+479 de la région 5' ETS du pré-ARNr, est représentée en jaune.
Les mutations compensatoires entre les deux segments de ces ARN, chez *S. cerevisiae,* sont indiquées par les nucléotides encerclés en vert.
Les boîtes A' et A phylogénétiquement conservées sont encadrées en rouge.
La localisation de l'intron est indiquée par un triangle bleu.

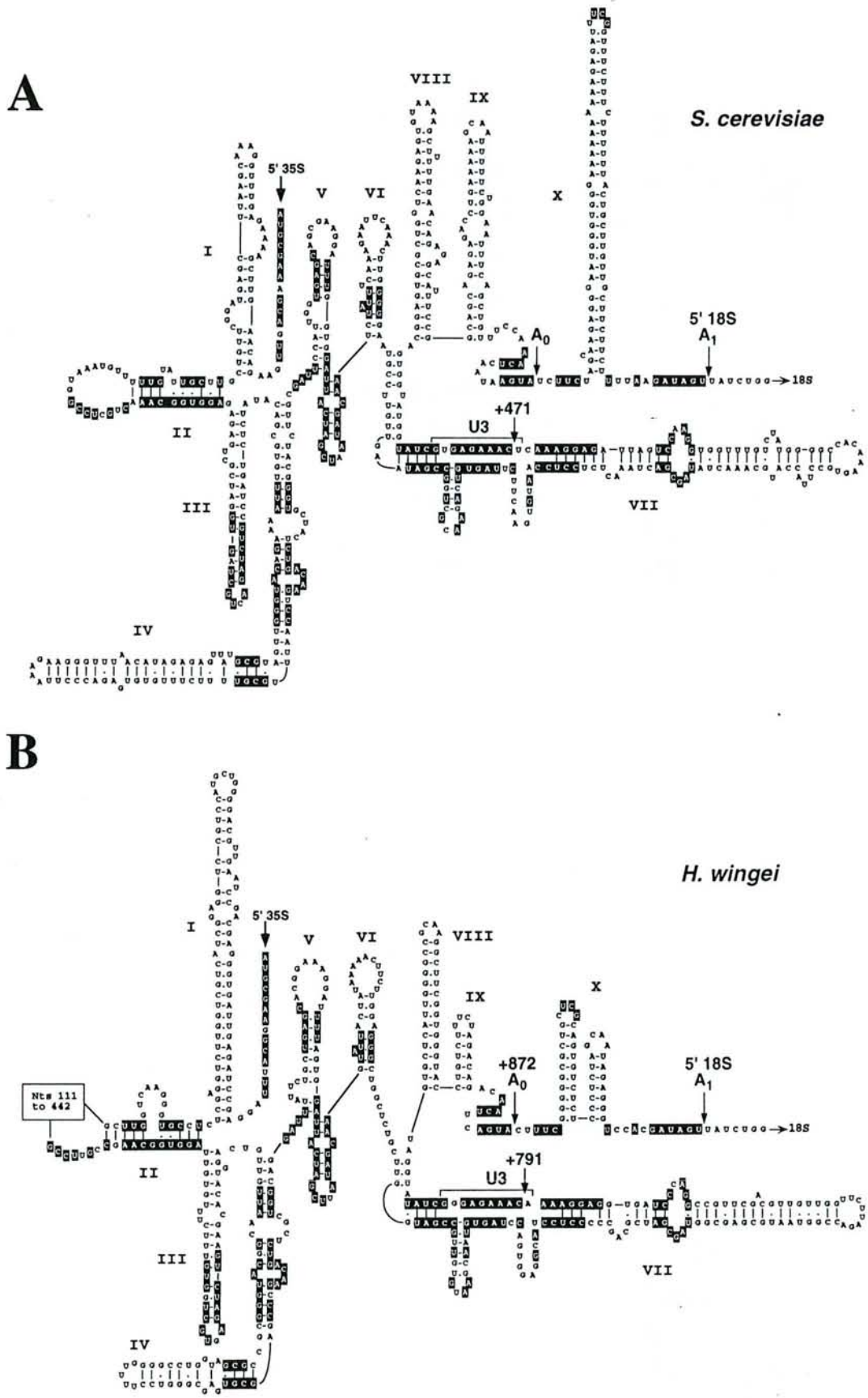


Figure 58 : Modèle de structure secondaire de la région 5' ETS du pré-ARNr de *Saccharomyces cerevisiae* (A) et de *Hansenula wingei* (B).

La numérotation des structures tige/boucle a été décrite par Klootwijk et Planta (1989) et Yeh et Lee (1992).
Les paires de bases conservées entre les deux structures sont indiquées en noir.
L'extrémité 5' de l'ARNr 18S (correspondant au site A1 de maturation) est indiquée par une flèche.
Le premier site de clivage du précurseur des ARN ribosomiques (le site A0) est également indiqué par une flèche.
Le segment complémentaire au snoRNA U3 est indiqué par un trait surmonté de "U3".

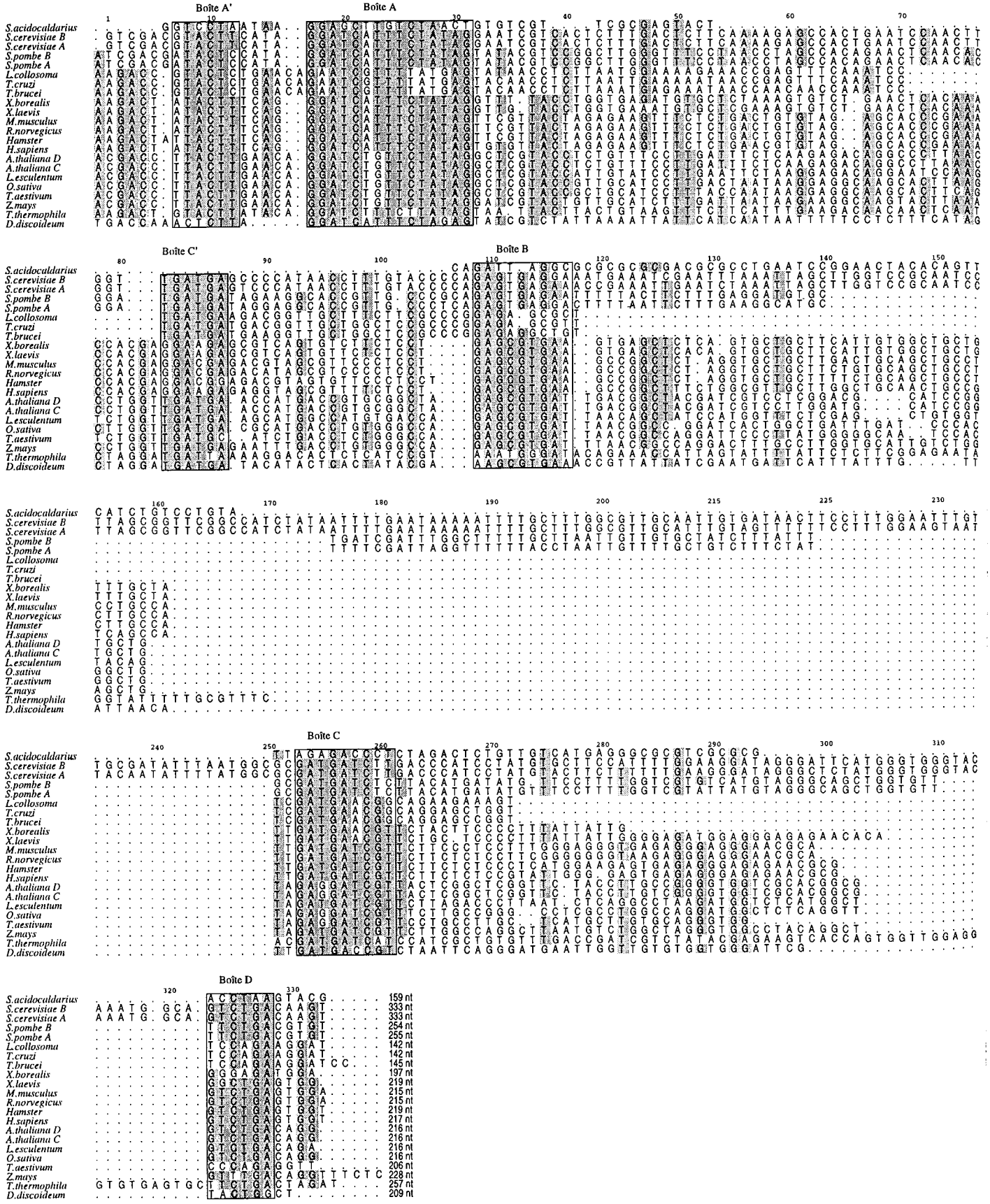


Figure 59 : Alignement des séquences des snoRNA U3 caractérisés chez différentes espèces.
Cet alignement a été réalisé avec le programme PileUp du logiciel GCG (Genetics Computer Group, Wisconsin, USA) et réajusté de manière à obtenir l'alignement des boîtes conservées au cours de l'évolution A', A, C', B, C et D.
Sulfolobus acidocaldarius (Potter *et al.*, 1995), *Saccharomyces cerevisiae* (Hughes *et al.*, 1987; Myslinski *et al.*, 1990), *Schizosaccharomyces pombe* (Porter *et al.*, 1988), *Leptomonas collosoma*, *Trypanosoma cruzi* et *Trypanosoma brucei* (Harsthorne et Agabian, 1994), *Xenopus borealis* et *Xenopus laevis* (Jeppensen *et al.*, 1988), *Mus musculus* (Mazan et Bachellerie, 1988), *Rattus norvegicus* (Stroke et Weiner, 1985), Hamster (A. Grégoire *et al.*, résultats non publiés), *Homo sapiens* (Suh *et al.*, 1986) *Arabidopsis thaliana* et *Tricitum aestivum* (Marshallsay *et al.*, 1990), *Lycopersicon esculentum* (Kiss et Solymosy, 1990), *Oriza sativa* (Qu F., résultats non publiés) *Zea mays* (Leader *et al.*, 1994), *Tetrahymena thermophila* (Orum *et al.*, 1993) et *Dictyostelium discoïdeum* (Wise et Weiner, 1980).

complémentaire de la région 5' ETS (Fig. 57). Cette observation était un argument phylogénétique fort en faveur de l'importance fonctionnelle de l'appariement proposé entre le snoRNA U3 de levure et la région 5' ETS. L'importance de cet appariement a aussi été démontrée, en parallèle, par voie génétique par création de mutations compensatoires entre les deux ARN (Beltrame et Tollervey, 1995).

Sur la base de la séquence établie pour la région 5' ETS du pré-ARNr de *H. wingei*, un modèle de structure secondaire fondé sur celui établi pour *S. cerevisiae* a été proposé. Comme on peut le voir dans la figure 58, un très faible degré de conservation de séquence est observé entre les régions 5' ETS de ces deux espèces de levures. La région complémentaire au snoRNA U3 est une des rares longues séquences conservées. L'ensemble du travail sur *H. wingei* est décrit dans l'article n° 3.

Il sera intéressant d'établir la séquence de la région 5' ETS du pré-ARNr de *P. salictaria*, afin de voir si une mutation compensatoire est aussi observée dans la région complémentaire au snoRNA U3.

Il a aussi été proposé que le snoRNA U3 de vertébrés interagisse par appariement de bases avec la région 5' ETS du pré-ARNr, mais la séquence du snoRNA U3 impliquée dans cet appariement postulé n'a pas été clairement identifiée. Dans l'état actuel des connaissances, la démonstration d'un rôle fonctionnel de l'appariement entre la séquence 39 à 49 du snoRNA U3 de levure et la région 5'ETS, démonstration à laquelle nous avons contribué, est le seul argument direct d'un mode d'intervention du snoRNA U3 dans la maturation de l'ARN pré-ribosomique. Comme nous l'avons déjà mentionné, J. Hughes (1996) a récemment proposé un rôle de la boîte A dans la formation d'une structure pseudo-nœud dans l'ARN 17S. Mais cette hypothèse doit encore être confirmée par des preuves expérimentales. R. Fournier a montré au laboratoire que la délétion d'une partie de la boîte A entraîne une non fonctionnalité du snoRNA U3.

Bien que le domaine 5' du snoRNA U3 soit, comme nous l'avons observé, très conservé chez les levures de la sous-famille des Saccharomycetoideae, ce domaine 5' est très variable en taille et en séquence si l'on compare les snoRNA U3 de levures, de protozaires, de plantes et de vertébrés (Fig. 59). Seules les boîtes A' et A sont très conservées. Cette situation est peu fréquente pour un domaine fonctionnel. Les régions des ARNr impliqués dans le site catalytique du ribosome sont très conservées (pour revue, Garrett et Rodriguez-Fonseca, 1995). Cette situation particulière du domaine 5' du snoRNA U3 s'explique sans doute par la grande variabilité de la région 5' ETS des pré-ARNr, comme nous avons pu le montrer en comparant la région 5' ETS de *H. wingei* et *S. cerevisiae*. Le degré de variabilité de cette région 5' ETS est encore plus grand entre les grands phyla champignons, plantes, animaux (Michot et Bachellerie, 1991). Une interaction de la boîte A avec l'ARNr 17S (18S) qui est lui-même très conservé, peut expliquer la grande conservation de cette boîte.

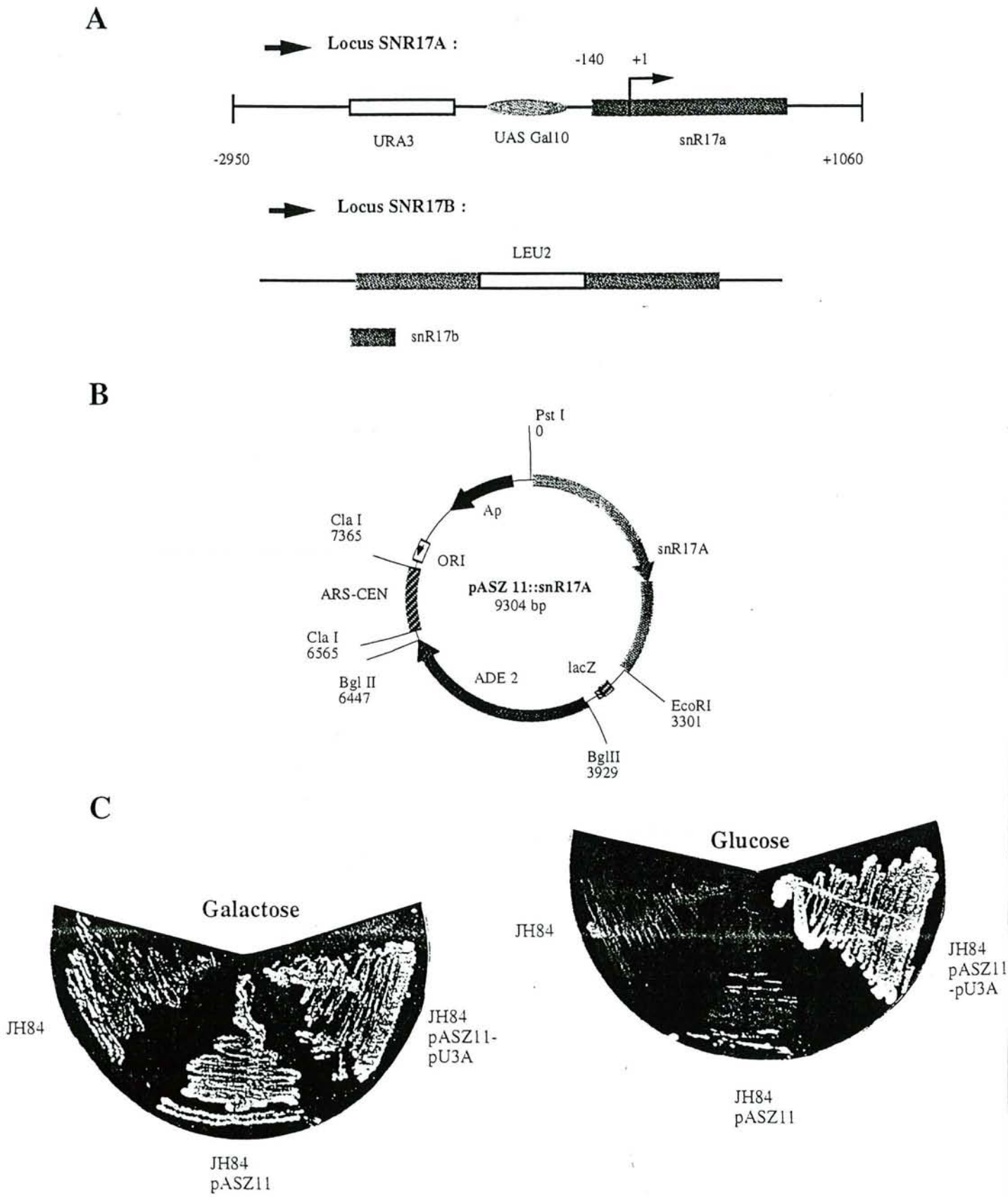


Figure 60 : Crible génétique utilisé pour tester la fonctionnalité des gènes de snoRNA U3 chez *Saccharomyces cerevisiae*.

A. Les modifications réalisées par Hughes et Ares (1991) dans les gènes chromosomiques des snoRNA U3A et U3B de la souche *S. cerevisiae*JH84, Le génotype de la souche JH84 est : Mat α snr17a.Gald:URA3 snr17b::LEU2 his3 ade2 can1 (Hughes et Ares, 1991).

B. Le vecteur navette recombinant pAZ11::snR17A, construit par R. Fournier à partir du vecteur centromérique pASZ11 (Stotz et Linder, 1990).

C. Exemple de croissance sur boîte de Pétri de la souche JH84 non transformée (JH84), transformée par le plasmide pASZ11 (JH84 pASZ11) et transformée par le plasmide pASZ11-pU3A (JH84 pASZ11-pU3A)

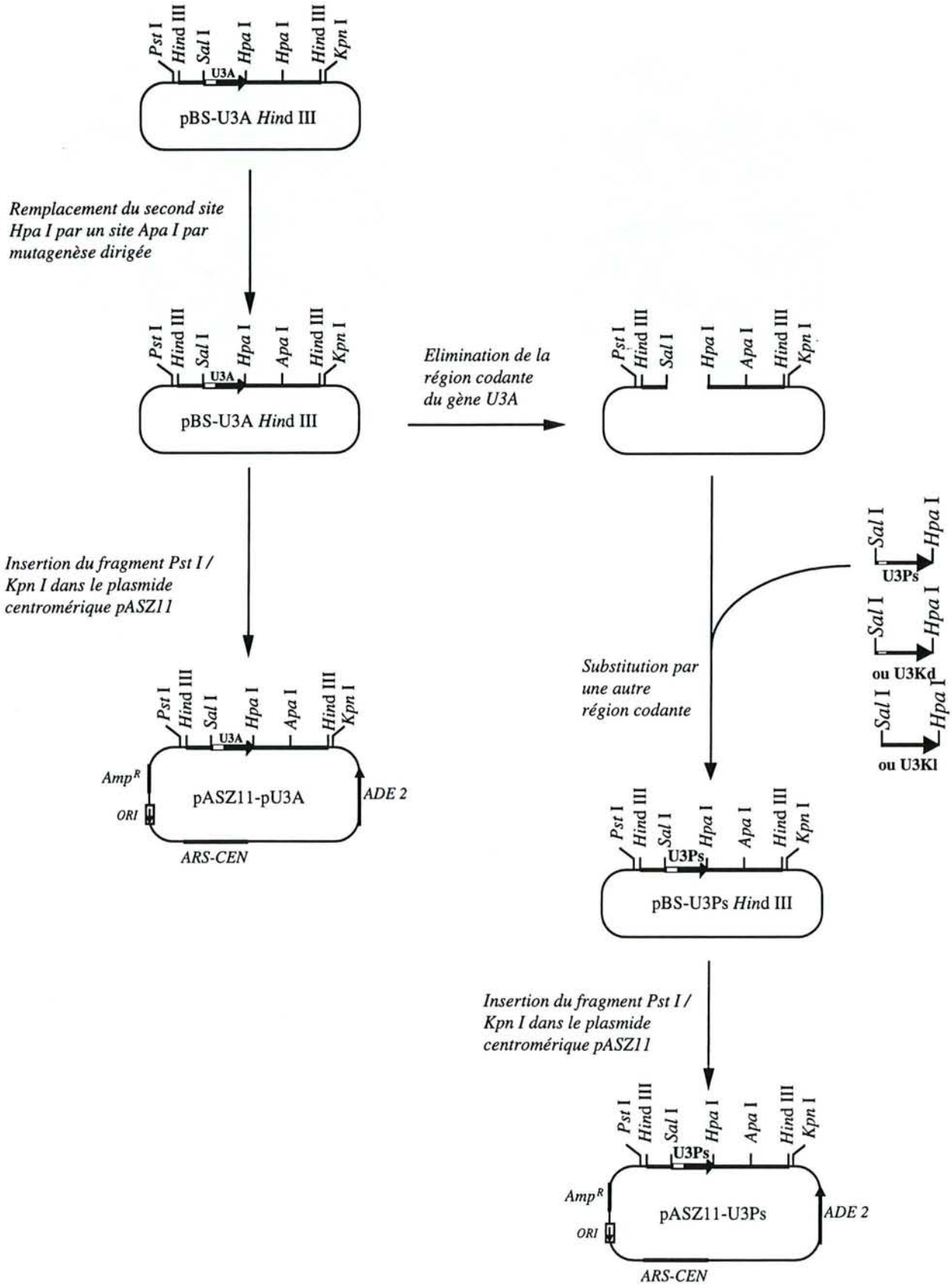


Figure 61: Représentation schématique des constructions génétiques réalisées en vue d'étudier la fonctionnalité chez *Saccharomyces cerevisiae* des snoRNA U3Ps, U3Kd et U3Kl.

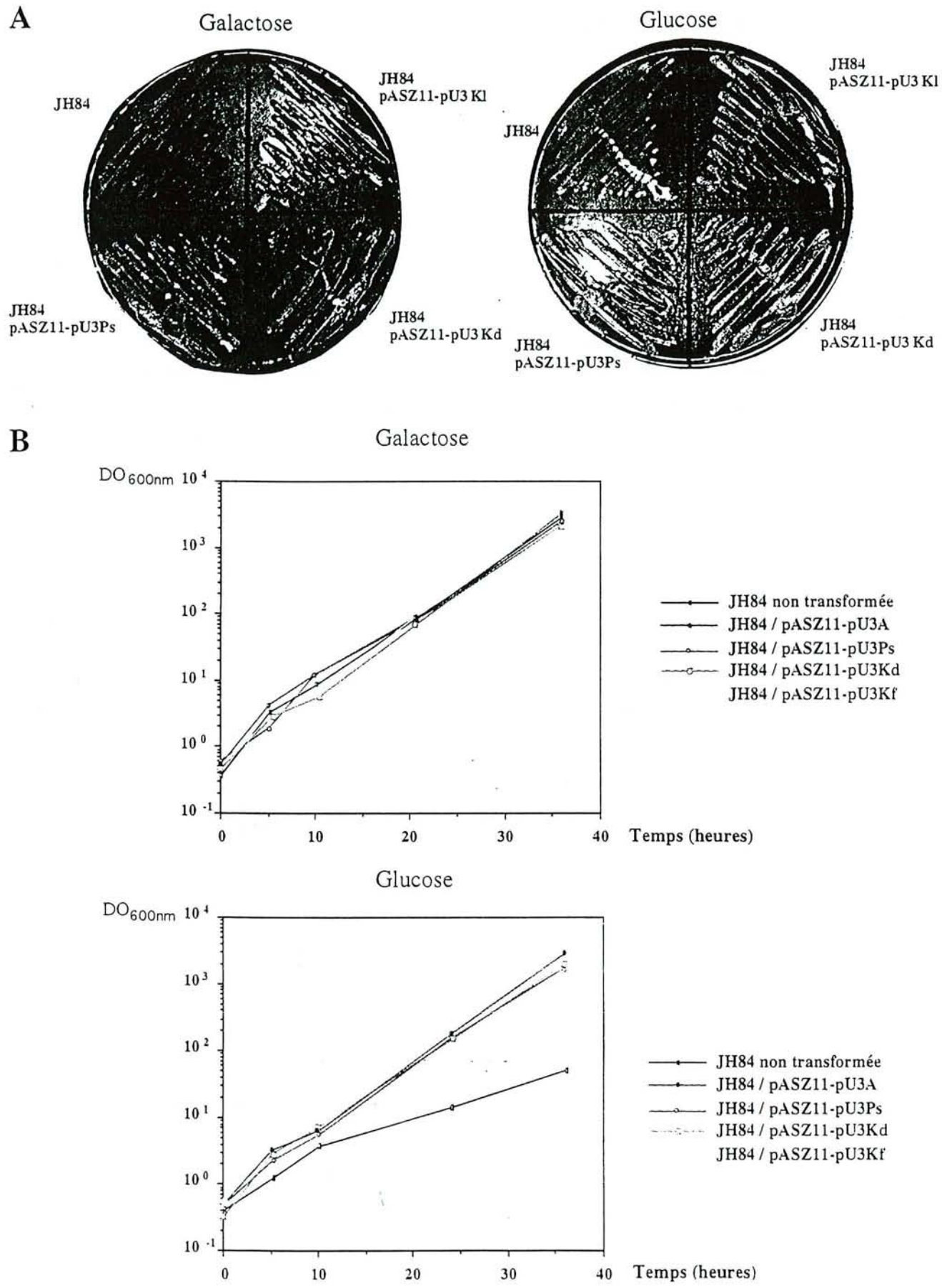


Figure 62 : Etude de la croissance de la souche *S. cerevisiae* JH84 transformée par différents plasmides pASZ11 recombinants.

A. Photographie des boîtes de Pétri réalisées en milieu YPGal ou YPGlc, sur lesquelles ont été étalées les levures JH8 non transformée, JH84 transformée par le plasmide pASZ11-pU3Ps, JH84 transformée par le plasmide pASZ11-pU3Kd, JH84 transformée par le plasmide pASZ11-pU3Kf.

B. Courbe de croissance des souches *S. cerevisiae* JH84 non transformée, transformée par le plasmide pASZ11-pU ou par le plasmide pASZ11-pU3Ps, ou par le plasmide pASZ11-pU3Kd ou le plasmide pASZ11-pU3Kf.

Au niveau du domaine 3', la variabilité est encore plus grande, puisque nous avons observé des différences marquées entre snoRNA d'un même genre de levure. Les différences de structure que nous avons observées entre les snoRNA U3 codés par les différents gènes que nous avons caractérisés, nous ont amenés à soulever la question de la fonctionnalité de ces différents ARN chez *S. cerevisiae*. Pour cela, nous avons utilisé le crible génétique que R. Fournier a développé au laboratoire. Le principe est le suivant : la souche *S. cerevisiae* JH 84, construite par J. Hughes et M. Ares (1991), a été utilisée comme souche test. Dans cette souche, le gène du snoRNA U3B a été interrompu par insertion du gène *LEU2*, et le gène du snoRNA U3A a été placé sous le contrôle de l'opérateur Gal10 (Fig. 60). Ainsi, lorsque cette souche est cultivée sur galactose, le gène du snoRNA U3A est exprimé. Par contre, lorsqu'elle est cultivée sur glucose le gène du snoRNA U3A est réprimé et inactif. Le gène U3B étant lui aussi inactif du fait de l'insertion du gène *LEU2*, aucun snoRNA U3 n'est donc produit. Cette souche ne peut pas se multiplier sur glucose, sauf si elle est transformée par un plasmide apportant en trans, un gène actif de snoRNA U3. C'est ce qu'a fait R. Fournier en clonant un fragment *Hin*d III/*Hin*d III de 1,5 Kb du chromosome de *S. cerevisiae* contenant le gène snoRNA U3A et ses régions flanquantes dans le plasmide navette centromérique pASZ11. Ce plasmide recombinant a été appelé pASZ11::pU3A. Dans ces conditions, la souche pousse bien sur glucose (Fig. 60).

Afin de tester la fonctionnalité des snoRNA U3Kd, U3Kf et U3Ps codés respectivement par les gènes U3 de *K. delphensis*, *K. marxianus* var. *fragilis* et *P. salictaria*, j'ai remplacé la région codante du gène U3A au sein du plasmide pASZ11::pU3A, par la région codante des gènes U3Ps, U3Kd et U3Kl que j'avais amplifiés par PCR (Fig. 61). Ceci a impliqué plusieurs constructions successives du fait de la présence de sites multiples de restriction dans le plasmide pASZ11. Nous avons ensuite testé la fonctionnalité des gènes ainsi obtenus. Comme nous pouvons le constater dans la figure 62, la souche *S. cerevisiae* JH84 transformée par les différents plasmides recombinants ne présente aucun problème de croissance sur glucose. Donc les snoRNA U3Ps, U3Kd et U3Kl, en dépit des grandes différences de structure de leur domaine 3', par rapport au domaine 3' des snoRNA U3A et U3B de *S. cerevisiae* sont fonctionnels chez cet organisme. Il semblait donc, que chez les levures, seules la présence des boîtes B et C au centre d'une structure à quatre branches plus ou moins régulière et celle des boîtes C' et D étaient importantes pour que le snoRNA U3 soit fonctionnel.

La comparaison des séquences de snoRNA U3 de différentes espèces (Fig. 59) montrait que l'extension d'environ 100nt présente dans les snoRNA U3 de levures de la sous-famille des Saccharomycetoideae, correspondait à la structure tige/boucle 2, et à une partie de la structure tige/boucle 4. La question se posait de savoir si ces structures

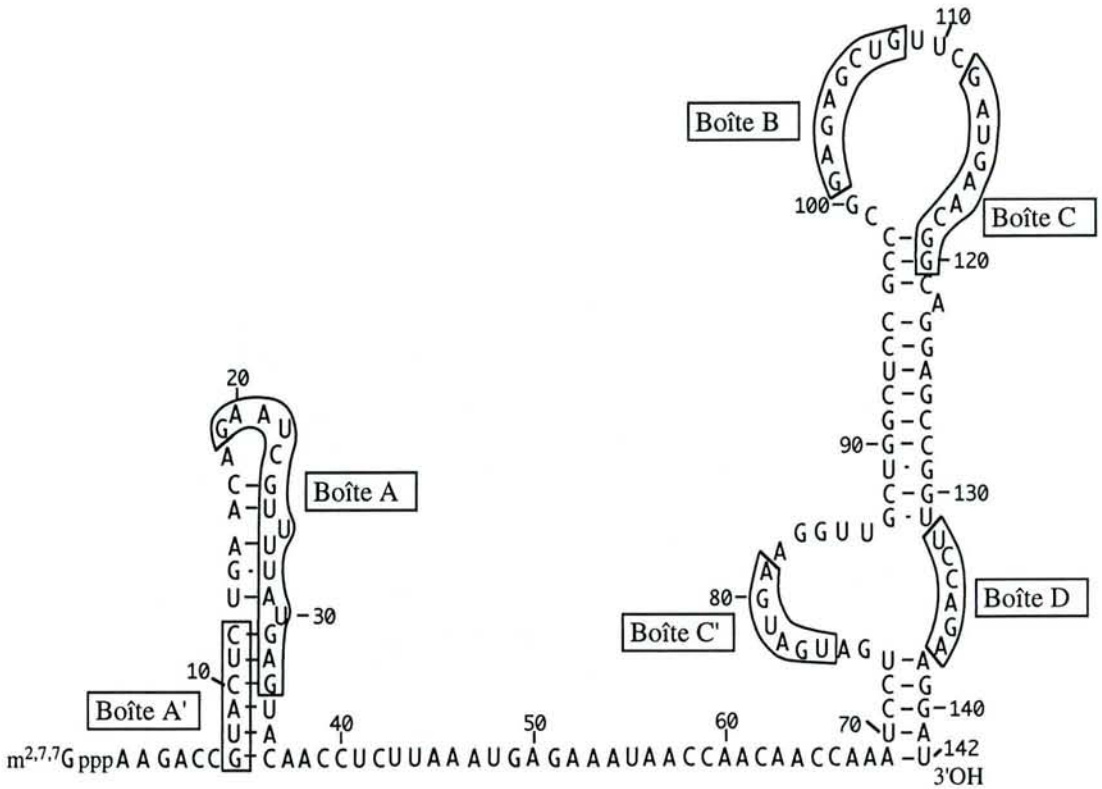


Figure 63 : Modèle de structure secondaire du snoRNA de *Trypanosoma brucei* (Harsthorne et Agabian, 1994).

Ce modèle a été proposé sur la base d'une étude expérimentale (Harsthorne et Agabian, 1994). Les boîtes A', A, C', B, C et D phylogénétiquement conservées sont encadrées.

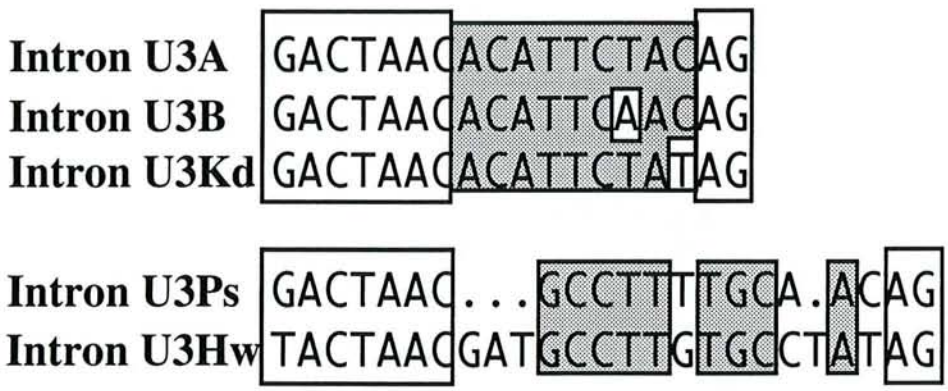


Figure 64 : Alignement des extrémités 3' des introns des gènes de snoRNA U3 de levures.

La boîte de branchement et le dinucléotide AG en 3' sont encadrés.
Les nucléotides conservés dans chaque groupe d'introns sont indiqués en gris
Le premier groupe d'introns est constitué des deux introns A et B mis en évidence dans les gènes U3A et U3B codant pour le snoRNA U3 de *Saccharomyces cerevisiae*, et de l'intron du gène unique codant pour cet ARN chez *Kluyveromyces delphensis*.
Le second groupe d'introns rassemble l'intron du deuxième gène mis en évidence chez *Pichia salictaria* et l'intron du gène unique trouvé chez *Hansenula wingei*.

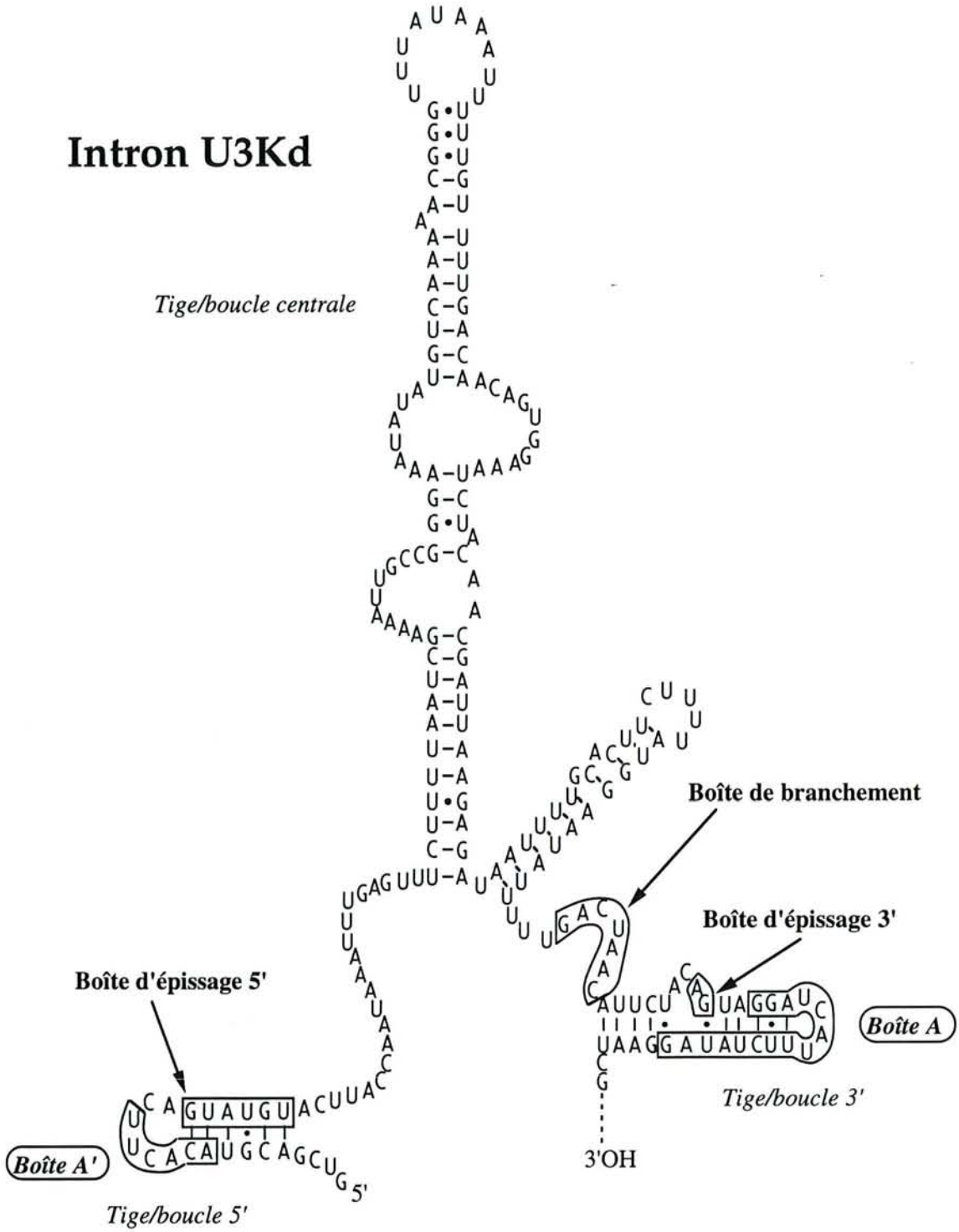


Figure 65 : Modèle de structure secondaire de l'intron du pré-snoRNA U3Kd de *Kluyveromyces delphensis.*

Les boîtes A' et A phylogénétiquement conservées sont encadrées, ainsi que les boîtes 5', 3' et de branchement de l'intron.
Ce modèle a été établi sur la base des modèles de la structure secondaire proposée pour les introns des pré-snoRNA U3A et U3B (Mougin *et al.*, 1996).

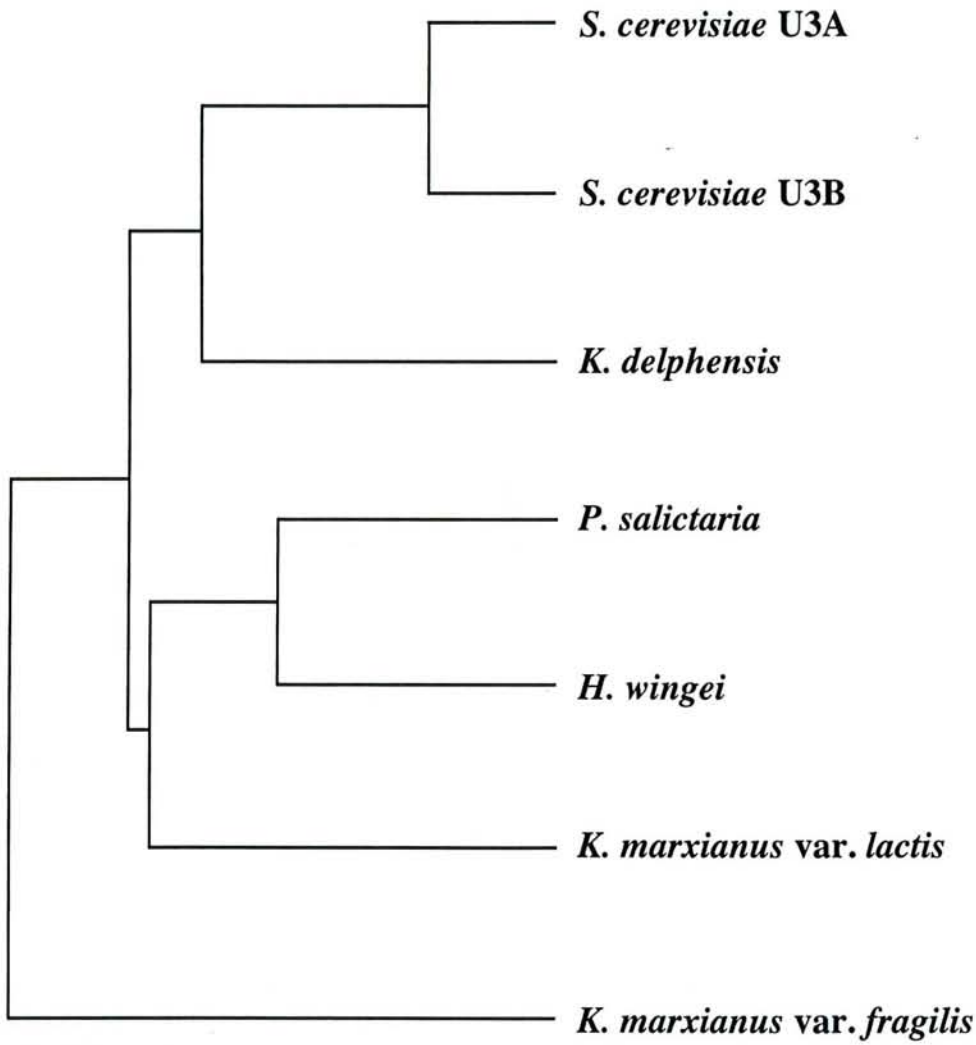


Figure 66 : Dendrogramme déduit à partir de la comparaison des régions codantes des snoRNA U3 caractérisés.

Ce dendrogramme a été réalisé avec le programme PileUp du logiel GCG.

tige/boucle étaient importantes pour la fonction du snoRNA U3 de levure. R. Fournier a délété chacune de ces structures individuellement au sein du snoRNA U3A et montré, toujours par le même crible génétique, que chacun des ARN tronqués était fonctionnel. Donc, aucune de ces deux structures tige/boucle prises individuellement, n'est indispensable. Néanmoins, le fait que j'aie retrouvé ces deux structures, même si elles n'ont pas tout à fait la même architecture, dans tous les snoRNA U3 que j'ai caractérisés, suggère que leur présence apporte un avantage sélectif, probablement du fait d'une augmentation de la stabilité et donc de la durée de vie du snoRNA U3 *in vivo*. Le domaine 3' du snoRNA de levure est, en effet, extrêmement stable. Ainsi, R. Fournier a pu montrer que des snoRNA non fonctionnels, tronqués dans leur domaine 5', sont néanmoins accumulés dans les cellules sous forme d'une particule ribonucléoprotéique. Il sera important de tester l'effet de la délétion simultanée des deux structures tige/boucle 2 et 4, puisqu'alors le snoRNA produit aurait, comme dans le cas de *Trypanosoma brucei* (Fig. 63), les boîtes B et C contiguës au sein d'une grande boucle.

Notre étude sur les levures de la sous-famille des Saccharomycetoideae avait donc mis en évidence une grande variété de gènes de snoRNA U3 et tous ceux testés s'étaient avérés fonctionnels. La question se posait de savoir quelle était la filiation de ces différents gènes. Pour répondre à cette question, j'ai, d'une part, comparé la région codante de ces différents gènes (Fig. 50), d'autre part, les régions introniques. Dans ce dernier cas, du fait de la forte divergence en taille et en séquence des introns, j'ai seulement comparé la région la plus conservée allant de la boîte de branchement au site 3' accepteur (Fig. 64).

De la comparaison des régions codantes, il ressort clairement que le gène unique mis en évidence chez *K. delphensis* est issu d'un ancêtre commun aux gènes U3A et U3B. Ce qui est aussi confirmé par la comparaison des régions 3' terminales des introns. L'intron de ce gène U3Kd peut d'ailleurs se replier selon une structure très voisine de celle établie pour les introns des gènes U3A et U3B avec une longue structure tige/boucle centrale, des structures tige/boucle en 5' et en 3' impliquant la boîte 5' et le site 3' accepteur, respectivement (Fig. 65). La structure tige/boucle centrale est flanquée en 3' d'une structure tige/boucle plus longue et plus stable que celle mise en évidence pour les introns U3A et U3B, et une structure tige/boucle peut aussi se former en 5' de la longue tige/boucle centrale. La structure secondaire proposée sur la base des complémentarités internes de l'intron est en cours de vérification expérimentale.

La comparaison des séquences codantes (présentée sous forme de dendrogramme, Fig. 66) et aussi des régions 3' terminales des introns (Fig. 64) montre une proximité phylogénétique entre le gène unique U3Hw mis en évidence chez *H. wingei* et le gène

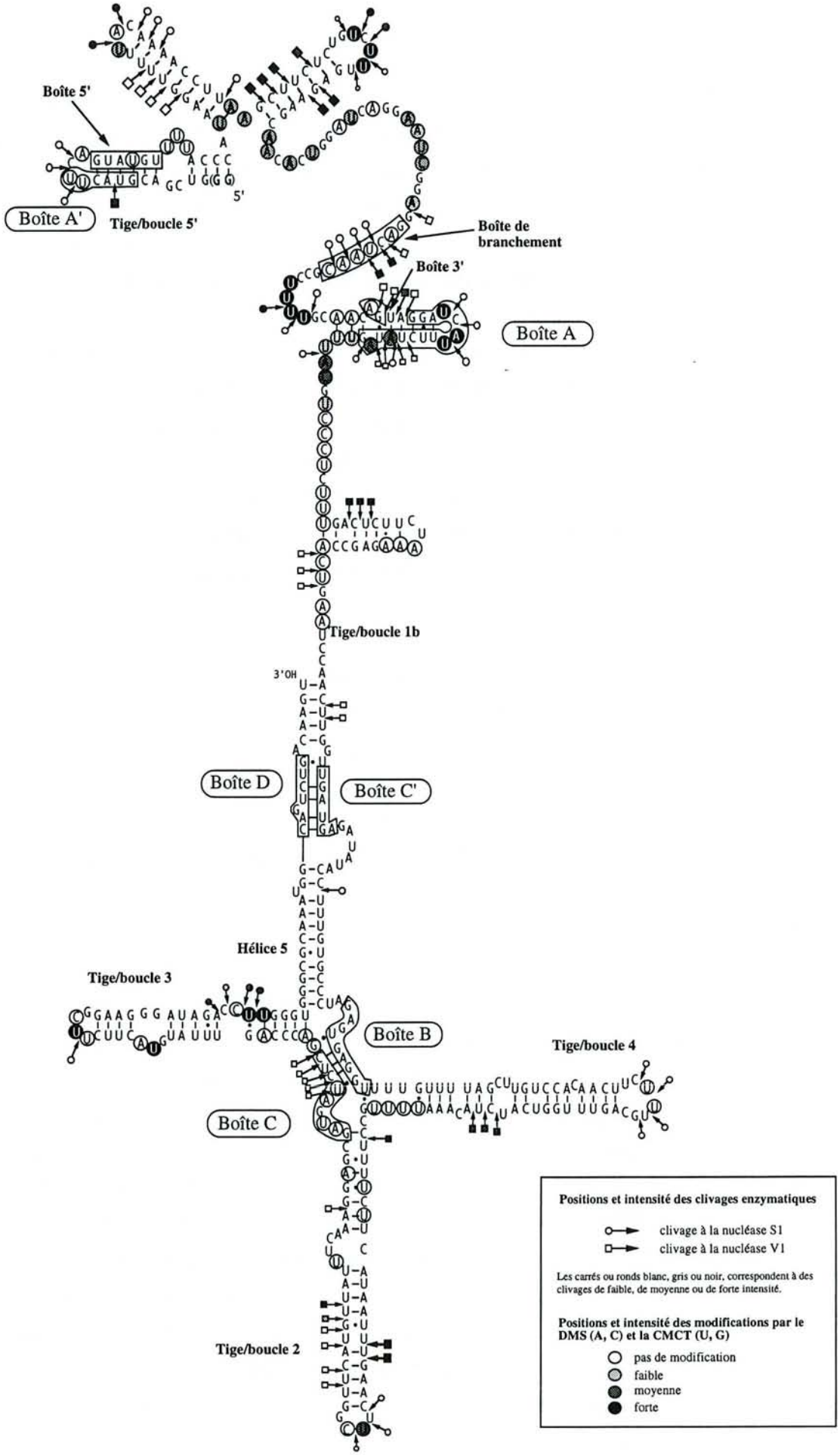


Figure 67 : Modèle de structure secondaire du pré-snoRNA U3Ps de *Pichia salictaria*.

Ce modèle a été établi sur la base d'une étude expérimentale basée sur l'emploi de sondes chimiques et enzymatiques de la structure des ARN. Les résultats obtenus avec ces sondes sont schématisés selon le code indiqué ci-dessus.

U3Ps, présent en plus du gène U3A, chez *P. salictaria*. Comme nous l'avons déjà signalé, c'est ce second gène qui est exprimé de manière prépondérante dans cette souche. Cette proximité entre *P. salictaria* et *H. wingei* est en bon accord avec le fait que dans les premières classifications basées sur des caractéristiques physiologiques, certaines espèces du genre *Pichia* avaient été classées dans le genre *Hansenula* (Kurtzman, 1984). Nos résultats confirment que les espèces des genres *Pichia* et *Hansenula* sont phylogénétiquement proches.

J'ai réalisé une étude expérimentale de la structure du pré-snoRNA U3Ps (Fig. 67). L'intron présente aussi une structuration importante, la conservation des séquences en 5' et en 3' permet toujours de former le même type de structure tige/boucle aux extrémités 5' et 3'. A la place d'une longue structure tige/boucle centrale, j'ai observé la présence de deux structures tige/boucle contiguës de taille similaire. Une petite structure tige/boucle centrale peut être proposée pour l'intron très court du gène U3Hw de *H. wingei*.

Enfin, bien que très divergents, les deux gènes sans intron de *K. marxianus* var. *fragilis* et *K. marxianus* var. *lactis* semblent phylogénétiquement plus proches de la famille des gènes U3 constituée par les gènes U3Ps et U3Hw, que de la famille constituée des gènes U3A, U3B et U3Kd.

La grande homologie des gènes U3A et U3B de *S. cerevisiae* et l'absence de gène U3B dans d'autres levures suggèrent que le gène U3B est d'origine assez récente, il est probablement issu d'une duplication du gène U3A, postérieure à la division entre les genres *Saccharomyces*, *Pichia*, *Kluyveromyces* et *Hansenula*. Si l'on tient compte du fait que le gène U3Kd, qui est unique chez *K. delphensis*, dérive d'un gène de type U3A, toutes les levures étudiées, sauf *H. wingei*, renferment un gène U3A ou de type U3A. On peut imaginer que du fait de mutations dans la région 5' ETS nécessitant la présence d'un snoRNA U3 présentant les deux mutations compensatoires, le gène U3A ait disparu dans cette espèce. Il est intéressant de noter que le gène U3Ps de *P. salictaria* qui présente une plus forte similitude avec le gène U3Hw, code pour un snoRNA U3 qui présente dans la séquence complémentaire à la région 5' ETS, une des deux substitutions présentes dans le snoRNA U3Hw de *H. wingei*.

Il est donc possible que très tôt deux familles de snoRNA U3 aient été générées chez les levures correspondant à deux formes alléliques de la région 5' ETS du pré-ARNr. L'ancêtre commun de ces deux familles de gènes de snoRNA U3, avait probablement déjà un intron, puisque dans chacune des deux familles, l'intron est localisé exactement à la même position.

La question se pose maintenant de savoir quelle est l'origine des gènes U3 sans intron mis en évidence chez les *Kluyveromyces*. Résultent-ils de la perte de l'intron porté par le

gène ancestral ? C'est ce que l'on pourrait supposer du fait de la similitude observée entre le gène U3Kl et les gènes U3Ps et U3Hw. Néanmoins, cette similitude est plus faible dans le cas du gène U3Kf, si bien que l'existence d'une troisième famille de gènes sans intron n'est pas exclue. Une étude sur un plus grand nombre d'espèces devra être réalisée pour répondre à cette question.

L'ensemble de cette étude comparative est présenté dans l'article n°4.

Publication n:°1 :

Étude de la structure secondaire du pré-ARN U3A de *S. cerevisiae* et de son rôle dans l'épissage.

A. Mougin, A. Grégoire, J. Banroques, V. Ségault, R. Fournier, F. Brulé, M. Chevrier-Miller, et C. Branlant

Sous presse dans RNA

The secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency.

Annie Mougin[1], Anne Grégoire[1], Josette Banroques[2], Véronique Ségault[1], Régis Fournier[1], Fabienne Brulé[1], Marianne Chevrier-Miller[2*] and Christiane Branlant[1§].

1. Laboratoire d'Enzymologie et de Génie Génétique, URA CNRS 457
Université de Nancy I, Faculté des Sciences
Boulevard des Aiguillettes, BP 239
54506 Vandoeuvre-lès-Nancy cedex, FRANCE.

2. Centre de Génétique Moléculaire, CNRS
Avenue de la Terrasse
91198 Gif sur Yvette, France.

2* Laboratoire de Génétique Moléculaire, Ecole Normale Supérieure
46, rue d'Ulm
75230 Paris cedex 05, France.

1§. To whom correspondance should be sent:
Laboratoire d'Enzymologie et de Génie Génétique, URA CNRS 457
Faculté des Sciences, Université de Nancy I
Boulevard des Aiguillettes, BP 239
54506 Vandoeuvre-lès-Nancy cedex, FRANCE.
Telephone: (33) 83.91.20.91.
Fax: (33) 83.91.20.93
Email: cbranlant@legg.u-nancy.fr

Running title: *Saccharomyces cerevisiae* pre-U3A snoRNA structure

Abstract

The *S. cerevisiae* U3 snoRNA genes contain long spliceosomal introns with noncanonical branch site sequences. By using chemical and enzymatic methods to probe the RNA secondary structure and site-directed mutagenesis, we established the complete secondary structure of the U3A snoRNA precursor. This is the first determination of the complete secondary structure of an RNA spliced in a spliceosome. The peculiar cruciform structure of the U3A snoRNA 3' terminal region is formed in the precursor RNA and the conserved Boxes B and C are accessible for binding the U3 snoRNP proteins. The intron forms a highly folded structure with a long central stem-loop structure that brings the 5' box and the branch site together. This is in agreement with the idea that secondary structure interactions are necessary for efficient splicing of long introns in yeast. The 3' splice site is in a bulged loop and the branch site sequence is single-stranded. Surprisingly, the 5' splice site is involved in a 6 base-pair interaction. We used *in vitro* splicing experiments to show that, despite a noncanonical branch site sequence and a base-paired 5' splice site, transcripts that mimic the authentic pre-U3A snoRNA are spliced very efficiently *in vitro*. Sequestering the 5' splice site in a more stable structure had a negative effect on splicing, which was partially compensated by converting the branch site sequence into a canonical sequence. Analysis of spliceosomal complex formation, revealed a cumulative negative effect of a base-pair interaction at the 5' splice site and of a deviation to the consensus sequence at the branch site, on the efficiency of spliceosome formation *in vitro*.

Key words: U3 snoRNA; *S. cerevisiae*; splicing; RNA secondary structure

Introduction

In vertebrates, most of the nuclear genes coding for proteins contain several introns that can be up to 100 kb in length (Hawkins, 1988). The situation is very different in yeast. Introns spliced in a spliceosome are present in less than 5% of the *Saccharomyces cerevisiae* nuclear genes (Kalogeropoulos, 1995). Except for the MATα1 and the YL8A genes (Miller, 1984; Mizuta et al., 1992), the *S. cerevisiae* nuclear genes contain no more than a single intron. In addition to protein genes, a few yeast nuclear genes that code for metabolically stable RNAs contain introns that are spliced in a spliceosome. Such introns were first observed in the spliceosomal U6 snRNA gene from *Schizosaccharomyces pombe* (Tani & Ohshima, 1989) and the nucleolar U3 snoRNA genes from *S. cerevisiae* (Myslinski et al., 1990). Later, these introns were found in the U6 snRNA genes of other *Schizosaccharomyces* species, *Rhodotorula hasegawae* and *Rhodosporidium dacryoidum* (Tani & Ohshima, 1991) and in the U2 snRNA genes from *R. hasegawae* (Takahashi et al., 1993). A growing number of small nucleolar RNAs (snoRNAs) is characterized in *S. cerevisiae* (for review, see Maxwell & Fournier, 1995). However, only the U3A and U3B snoRNA genes were found to contain an intron, and the introns of these two genes share the peculiarity of having a noncanonical sequence at the branch site (GACTAACp instead of TACTAACp) (Myslinski et al., 1990).

U3 snoRNA plays a crucial role in pre-ribosomal RNA maturation. In eukaryotes, the 18S, 5.8S and 25/28S ribosomal RNAs (rRNAs) are transcribed by RNA polymerase I as a single precursor molecule. A complex series of processing reactions are needed to eliminate the 5' and 3' external transcribed spacers (5' ETS and 3' ETS, respectively) and the internal transcribed spacers 1 and 2 (Maxwell & Fournier, 1995; Eichler & Craig, 1994; Raué & Planta, 1995). The primary event, A_0 cleavage within the 5' ETS spacer (Craig et al., 1987; Kass et al., 1987; Hughes & Ares, 1991), was successfully reproduced in extracts of cultured mouse cells (Craig et al., 1987) and in *Xenopus laevis* (Mougey et al., 1993), and it was found to be impaired upon U3 snoRNA depletion (Kass et al., 1990; Mougey et al., 1993). In yeast, when U3 snoRNAs are depleted the normal pre-rRNA intermediates in the pathway leading to the synthesis of the 18S rRNA are missing which results in an under-accumulation of mature 18S rRNA (Hughes & Ares, 1991).

We previously studied the secondary structure of the mature U3A snoRNA from *S. cerevisiae* (Ségault et al., 1992). It was interesting to compare this structure with that of the precursor RNA containing the intron. Up to now, only the secondary structures of limited portions of eukaryotic pre-mRNAs have been studied (Teare & Wollenzien, 1989; Teare & Wollenzien, 1990; Clouet d'Orval et al., 1991; Li et al., 1995). However, several reports suggest an important role of RNA secondary structure in the splicing efficiency of yeast pre-mRNAs. Computer analysis suggested the presence of a long central stem-loop structure in several *S. cerevisiae* pre-mRNA introns (cyh2, S10, RP51A and actin; Parker & Patterson, 1987). Moreover, the genetic deletion experiments of Pikielny & Rosbash (1985) and of Newman (1987) suggested that a base-pair interaction between segments located downstream from the 5' splice site element and upstream of the branch point sequence is required for

optimal splicing efficiency. This was experimentally confirmed by mutational analyses of the introns in the cyh2 ribosomal protein gene (Newman, 1987), and in the RP51B ribosomal protein gene (Goguel & Rosbash, 1993). Libri et al. (1995) recently proposed that base-pair interactions are generally needed for efficient splicing of long introns in yeast, based on randomisation selection experiments. Finally, two other examples showing the importance of pre-mRNA secondary structure on splicing efficiency in yeast are the occlusion of a 3' splice site by the formation of a stable secondary structure in the *Kluyveromyces lactis* actin gene (Deshler & Rossi, 1991) and the auto-control of the *S. cerevisiae* L32 gene by the L32 protein binding on a stem-loop structure formed by the intron and the first exon (Eng & Warner, 1991). However, all these studies suffer from the absence of information about the entire secondary structure of the intron.

In this paper, we describe the complete secondary structure of the *S. cerevisiae* pre-U3A snoRNA. This precursor RNA is highly folded; it has a long central stem-loop structure in the intron and an unexpected 5' stem-loop structure that contains the 5' splice site. We also tested the importance of this unexpected 5' stem-loop structure and the importance of the noncanonical branch site sequence on splicing efficiency. For this purpose, we produced variant pre-U3A snoRNAs by site-directed mutagenesis, and we studied their secondary structure and *in vitro* splicing efficiency. The possible implications of these two peculiar features of the pre-U3A snoRNA is discussed.

Results

1. Strategy for secondary structure analysis.

The intron of the *S. cerevisiae* U3A snoRNA gene is located in the region coding for the 5' terminal stem-loop structure 1a (Myslinsky et al., 1990; Ségault et al., 1992) (Inset in Fig. 1A) that is supposed to play an essential role in pre-ribosomal RNA maturation (Beltrame & Tollervey, 1992 and 1995). Exon 2 contains the sequence that forms the stem-loop structure 1b and the 3' terminal cruciform structure in the mature RNA (Inset in Fig. 1A). All the sequences expected to be required for binding of the U3 snoRNP proteins are contained in exon 2 (Parker & Steitz, 1987; Lübben et al., 1993; Hartshorne & Agabian, 1994 and our unpublished results on the yeast U3 snoRNP).

To determine the secondary structure of the pre-U3A snoRNA, we produced an *in vitro* transcript that contained a minimal number of extraneous nucleotides as compared to the authentic pre-U3A snoRNA: two additional G residues at the 5' end, to enhance transcription by T7 RNA polymerase, and one additional U residue at the 3' end, which was generated by HpaI digestion of the DNA template. This transcript contained a GACUAACp sequence at the branch site and a 16 nt-long exon 1, and it was designated pU3G16.

The secondary structure was probed with dimethylsulfate (DMS), 1-cyclohexyl-3-(2-morpholino ethyl)-carbodiimide-metho-p-toluene sulfonate (CMCT), and S1 and V1 RNases, under the conditions given in Materials and Methods. Positions of modification or cleavage

were identified by reverse transcription using primers 1, 2 and 3 (Fig. 1A and 2A). As will be described below, based on the properties of these agents (Ségault et al., 1992), two possible secondary structures of the pre-U3A snoRNA were deduced. Therefore, we constructed pre-U3A snoRNA variants by site-directed mutagenesis to test these two possibilities. These variant U3A pre-snoRNA transcripts were subjected to RNA secondary structure analysis and the results allowed us to propose a complete secondary structure model for the pre-U3A snoRNA. For reasons of clarity, the positions of residues within exon sequences will be numbered as in the mature U3A snoRNA (Ségault et al., 1992) and within the intron in italic numbers from *+1* to *+157*.

2. The U3 snoRNA 3' terminal cruciform structure and stem-loop structure 1b are found in the precursor RNA.

For a comparative analysis between precursor and mature RNAs, the pU3G16 transcript and an *in vitro* U3 transcript produced as previously described (Ségault et al., 1992), were subjected in parallel to chemical and enzymatic probing. The results obtained, which are summarized in Figure 1A, confirmed the presence of the 3' terminal cruciform structure and stem-loop structure 1b in the precursor RNA. Except for an increased intensity of the V1 RNase cleavages in stem-loop 3, the intensities of chemical modifications and enzymatic digestions in the 3' terminal cruciform structure were very similar for both the mature and precursor RNAs (Fig. 1A). Hence, the secondary structure and the tertiary structure of the 3' domain were almost identical in the two RNAs. In contrast, the accessibility in stem-loop structure 1b differed at several positions between the precursor and the mature RNA. V1 RNase cleavages were reinforced in the stem and both V1 and S1 cleaved in the terminal loop (Fig. 1A and B). One explanation for the additional V1 cleavages may be that a pseudo-knot structure formed between the 5' terminal ACUGp sequence of the terminal loop and the CAGUp sequence from position 321 to 324 (Fig. 1A).

3. A highly folded intron sequence.

The enzymatically- and chemically-modified intron structure was analyzed by reverse transcription with primers 2 and 3 (Fig. 2) and the results are summarized in Figure 3. Chemical and enzymatic probing experiments clearly demonstrated that the 3' splice site was located at the 3' end of a bulged loop in a stem-loop structure designated 1a'. The top part of stem-loop structure 1a' was identical to that of stem-loop structure 1a in the mature RNA. The branch site sequence ($G_{+140}ACUAAC_{+146}$) was located in a large bulged loop. A long central stem-loop structure was present at the center of the intron from position C_{+31} to G_{+107} (Fig. 2B, 2C and 3). The base of the stem (segment from position A_{+21} to U_{+25} base-paired with the segment from position A_{+109} to U_{+113}) had a lower stability, as reflected by mild chemical modifications in both nondenaturing and semi-denaturing conditions (Fig. 2B). For the 5' terminal part of the intron, two alternative conformations could fit the experimental results; in the A conformation (Fig. 3A), exon 1 is base-paired with the 5' terminal sequence of the intron, whereas in the B conformation (Fig. 3B), it is base-paired with exon 2.

4. A base-pair interaction between exon 1 and the intron 5' extremity.

Two approaches were used to test whether the V1 RNase cleavages in exon 1 resulted from a base-pair interaction with exon 2 or with the intron. In the first approach, a transcript missing the 9 nt sequence of exon 2 that is complementary to exon 1, was produced by site-directed mutagenesis (transcript pU3G16Δ9, Fig. 3C). In the second approach, a transcript missing the exon 2 sequence downstream from position 17, was produced by Sau3AI digestion of the DNA template prior to transcription (transcript pU3G16Sau3AI, Fig. 3C). As illustrated in Fig. 4A, the patterns of chemical modifications and enzymatic digestions in exon 1 were identical for the transcripts pU3G16, pU3G16Δ9 and pU3G16Sau3AI. Consequently, these three transcripts adopt the A conformation. This suggested that, in the authentic precursor RNA, the 5' terminal sequence of the intron, which is known to base-pair with U1 snRNA in the course of spliceosome assembly (Lerner et al., 1980; Mount et al., 1983; Zhuang & Weiner, 1986), is base-paired with exon 1. However, we had to verify that the 5' terminal structure found in the transcripts was not an artefact due to the presence of the two additional G residues at the 5' end. Indeed, the resulting run of three G residues at the 5' end could interact with the run of four C residues between positions C_{+13} and C_{+16} within the intron, and this might favour formation of a 5' stem-loop structure. Due to the constraints of T7 RNA polymerase, the two additional G residues at the 5' end of the transcripts could not be omitted. Thus, to check whether the 5' terminal stem-loop structure was formed in the absence of a base-pair interaction between the 5' terminal G run and the [C_{+13}-C_{+16}] segment, we deleted the C run between positions *+13* and *+16* by site-directed mutagenesis (transcript pU3G16Δ4C, Fig. 3D). The 5' terminal stem-loop structure in transcript pU3G16Δ4C was still evident in the probing experiments (Fig. 4B). Hence, we concluded that this 5' terminal structure is likely to be present in the authentic pre-U3A snoRNA.

In summary, our probing and site-directed mutagenesis experiments demonstrated the presence of a 5' terminal stem-loop structure, a central stem-loop structure and stem-loop structure 1a' in the pre-U3 snoRNA (Fig. 5). A few additional interactions of low stability were also suggested by the results of the probing experiments (Fig. 5). In particular, a base-pair interaction may occur between the GCUUp sequence upstream of the branch site (position G_{+134} to U_{+137}) and the 3' extremity of exon 2 (Fig. 5). However, due to the presence of an additional U residue at the 3' end of the transcript, this interaction might be reinforced in the T7 transcript as compared to the authentic RNA (Fig. 3A).

5. *In vitro* splicing efficiency depends upon the 5' splice site accessibility and the branch site sequence.

As U3A snoRNA is produced in large amounts in yeast, a high efficiency of splicing of the pre-U3A snoRNA is expected *in vivo*. It was, thus, surprising to find both a noncanonical branch site sequence, GACUAACp, (Myslinski et al., 1990) and a base-pair interaction between exon 1 and the intron 5' extremity in the pre-U3A snoRNA. To obtain more

information on this observation, *in vitro* splicing experiments were made with the pU3G16 transcript and a series of variant pU3 transcripts.

Transcript pU3G16 was inefficiently spliced *in vitro* and this is shown by the very low ratio of mature RNA to precursor RNA (M/P: 0.04) (Fig. 6A). In contrast, a transcript with a 52nt-long extension at the 5' end of exon 1, that arose from the BS(-) phagemid (transcript pU3G66, Fig. 6A) was spliced very efficiently (M/P: 1.37) (Fig. 6A). The pU3G66 transcript was spliced more efficiently than an actin transcript, which is considered to be one of the more efficiently spliced yeast pre-mRNA *in vitro* (Lin et al., 1985). This was particularly surprising because of the GACUAACp sequence at the branch site.

Using *S. cerevisiae* thermosensitive mutants with defect in the pre-mRNA splicing machinery (SPJ11.4 and AH rna8; Jackson et al., 1988 and Chang et al., 1988), we verified that introns in the pre-U3A snoRNA are spliced by the same machinery as introns in pre-messenger RNAs (not shown).

When the branch site sequence was converted to a canonical one in transcript pU3G16 (transcript designated as pU3U16, Fig. 6A), splicing was increased by a factor of about 15 (Fig. 6A). When the same mutation was performed in transcript pU3G66 (transcript designated as pU3U66) splicing efficiency was only increased by a factor of 1.7 (Fig. 6B). Why transcript pU3G66 was spliced very efficiently with a GACUAACp sequence, whereas transcript pU3G16 needed a UACUAACp branch site sequence for efficient splicing ? One possible explanation was the short length of exon 1 in transcript pU3G16, since 16 nt represent the shortest size for a natural yeast exon 1 spliced in a spliceosome (Duchêne et al., 1988). To test for this possibility, a series of chimeric transcripts, with 5' terminal extensions of different length and sequence, were produced as described in Materials and Methods (Fig. 6A). Transcript pU3G74 has a 60 nt extension, with a sequence different from that of transcript pU3G66. Transcripts pU3G66Δ1 and pU3G66Δ2 differ from transcript pU3G66 by 21 and 29 nt-long deletions within the 52 nt extension, respectively. As shown in Figure 6A, no splicing was observed for the pU3G74 transcript and a low level of splicing was obtained for the pU3G66Δ1 and pU3G66Δ2 transcripts (M/P ratios of 0.09 and 0.25, respectively). Hence, splicing efficiency was not correlated to the length of exon 1.

Another possible explanation for the poor splicing efficiency of transcript pU3G16 was the sequestering of the 5' splice site in the stable stem-loop structure formed at the 5' end of this transcript. In agreement with this hypothesis, a secondary structure analysis of the pU3G66, pU3G66Δ1 and pU3G66Δ2 transcripts (Fig. 7) showed a reinforced 5' stem-loop structure in the transcripts that were inefficiently spliced *in vitro* (pU3G66Δ1 and pU3G66Δ2). In contrast, the 52 nt extension of the efficiently spliced pU3G66 transcript folded on itself independently of the pre-U3A snoRNA sequence (Fig. 7) and this reduced the 5' stem-loop structure stability. Due to the presence of reverse transcriptase stops in the CCCCp sequence (between positions A_{+12} *to* A_{+17}) in control experiments, the single-stranded or double-stranded state of this sequence in transcript pU3G66 could not be determined.

As we showed that, in transcript pU3G16Δ4C, the 5' stem-loop structure has a lower stability, if the above reasoning was correct, transcript pU3G16Δ4C was expected to be spliced

more efficiently than transcript pU3G16. As shown in Figure 6B, this was indeed the case and the splicing efficiency of transcript pU3G16Δ4C is 62% of that measured for transcript pU3G66 (for comparison, the efficiency for pU3U16, in the same experiment, was 42% of that for transcript pU3G66).

Finally, a last confirmation of the strong negative effect of a reinforced 5' stem-loop structure on pre-U3 snoRNA splicing efficiency was the absence of splicing when the irregular stem of the 5' stem-loop structure was converted into a perfect 12 bp stem. In this case, splicing was abolished both in the presence of a GACUAACp branch site sequence (transcript pU3G16ds5') and a canonical branch site sequence (transcript pU3U16ds5') (Fig. 6B).
As evidenced by a kinetic study of spliceosome assembly with the pU3G66, pU3G16, pU3G16Δ4C and pU3G16ds5' transcripts (Fig. 8), an increased sequestration of the 5' splice site in the 5' stem-loop structure, in the presence of a noncanonical branch site sequence affects an early step of spliceosome assembly. After incubation for 10 or 40 minutes, the rates of B, A1 and A2 complexes formation was strongly reduced for transcript pU3G16, as compared to transcripts pU3G66 and pU3G16Δ4C. The A1 and A2 complexes were even not seen in case of transcript pU3G16ds5'.

Discussion

We established a complete secondary structure model for the pre-U3A snoRNA of *S. cerevisiae*. This is the first complete secondary structure determination of a pre-RNA spliced in a spliceosome. The *S. cerevisiae* pre-U3A snoRNA displays two specific structural features that are not expected for an efficiently spliced transcript : a noncanonical branch site sequence and a base-paired 5' box. We used site-directed mutagenesis and *in vitro* splicing experiments to study the influence of these two peculiar features on splicing efficiency. Altogether, the results described in this paper have specific implication for the understanding of U3 snoRNA biogenesis, but also for the understanding of the effect of pre-RNA secondary structure on *in vitro* splicing efficiency by the yeast splicing machinery.

The 3' domains of the mature and the precursor RNA have similar structures.

Most of the secondary structure motifs found in the *S. cerevisiae* U3A snoRNA are found in the precursor RNA. Only the basis of stem-loop structure 1a is not formed in the precursor RNA. As a consequence, depending on the relative kinetics of transcription versus spliceosome assembly, splicing of the pre-U3A snoRNA *in vivo* may take place in the presence of a highly folded exon 2. In addition, as exon 2 contains the sequences expected to bind the U3 snoRNP proteins (Parker & Steitz, 1987; Lübben et al., 1993; Hartshorne & Agabian, 1994 and our unpublished results on the yeast U3 snoRNP), one could not exclude the possibility that splicing of the pre-U3A snoRNA *in vivo* takes place on a pre-U3A snoRNP complex containing some of the U3A snoRNP proteins.

An intron central stem-loop structure in U3 snoRNA precursors.

Several studies have suggested that formation of a base-pair interaction between sequences downstream from the 5' splice site and upstream of the branch site sequence is required for efficient splicing of long introns in *S. cerevisiae* (Parker & Patterson, 1987; Pikielny & Rosbash, 1985; Newman, 1987; Goguel & Rosbash, 1993 and Libri et al., 1995). The secondary structure that we have determined for the pre-U3A snoRNA demonstrates the postulated interaction is formed in this precursor RNA. However, in the pre-U3A snoRNA, this interaction has a low stability as determined by a free energy calculation (Fig. 5) and by the mild accessibility of the base-paired segments to chemical reagents both in nondenaturing and semi-denaturing conditions (Fig. 2 and 3). The upper part of the central stem-loop structure of the pre-U3A snoRNA, which is more stable (-9.3 kcal/mol, Fig. 5), is probably required to ensure the proper interactions between the sequences downstream from the 5' splice site and upstream of the branch site sequence.

An argument in favour of an important biological role of the central stem-loop structure in the pre-U3A snoRNA intron is the possibility to form a similar stem-loop structure with a completely different sequence in the intron of the second U3 snoRNA gene of *S. cerevisiae* (U3B gene) (Brulé et al., 1995). In addition, the *S. cerevisiae* FL100 and *S. cerevisiae* GY5 U3A snoRNA genes differ at 3 positions in the intron, and two of these mutations reinforce the proposed central stem-loop structure in *S. cerevisiae* GY5 (Brulé et al., 1995 and Fig. 5).

The pre-U3A snoRNA is very efficiently spliced despite a noncanonical branch site sequence and a base-paired 5' splice site.

While an artificial 6 bp interaction that involves the 5' splice site in the RP51A pre-mRNA had a negative effect on *in vitro* splicing efficiency (Goguel et al., 1993), the pU3G16Δ4C transcript and particularly the pU3G66 transcript, that bear the naturally occurring 6 bp interaction at the 5' splice site, were spliced very efficiently *in vitro*. This is particularly surprising because these two transcripts also have noncanonical branch site sequences.

Interestingly, the 6 bp interaction at the 5' splice site of the pU3G66 transcript has the same calculated free energy (-3.4 kcal/mol) as the interaction between U1 snRNA and the pre-U3A snoRNA 5' box. When the stability of the 5' stem-loop structure was increased by -2.7 kcal/mol, upon formation of three additional G-C base-pairs (transcript pU3G16), the *in vitro* splicing efficiency was reduced by 98% relative to transcript pU3G66. However, converting the branch site into the canonical sequence partially relieved the splicing efficiency to 42% of pU3G66.

Effects of mutations in the branch site sequence on *in vitro* and *in vivo* splicing efficiency in yeast were studied previously (Jacquier et al., 1985; Vijayraghavan et al., 1986), but to our knowledge the effect of a mutation at the first position was not tested. Our results show that, at least *in vitro*, the effect of a mutation at this position strongly depends on the pre-RNA conformation.

Based on the unexpectedly high *in vitro* splicing efficiency of the pU3G66 transcript, the noncanonical branch site sequence and the base-paired 5' box in the pre-U3A snoRNA are

probably offset by an optimization of the other structural parameters that influence splicing efficiency, such as the long central stem-loop structure. Interestingly, the branch site sequence is located in a large bulged loop that is closed by an helix of low energy; this helix is probably easily destabilized upon spliceosome assembly. The 3' splice site is located at the extremity of a bulged loop in the stem-loop structure 1a'. Recent investigations on plants (Liu et al., 1995) and HeLa cells (Smith et al., 1993) showed that sequestrating the 3' splice site has a milder effect on splicing efficiency than sequestrating the 5' splice site. This may also be the case for yeast.

The noncanonical branch site and the base-paired 5' splice site in the *S. cerevisiae* pre-U3A snoRNA may provide a control mechanism for U3A snoRNA production at the splicing step. Indeed, our results show that the noncanonical branch site sequence places pre-U3A snoRNA splicing under the dependence of the availability of the 5' splice site sequence. One may imagine the existence of a negative regulator protein that would bind to the 5' terminal stem-loop structure and reinforce its stability. As a consequence, splicing efficiency would be strongly reduced. A very similar mechanism explains the retro-control of the *S. cerevisiae* protein RPL 32 at the splicing level (Eng & Warner, 1991; Vilardell & Warner, 1994 and Li et al., 1995). An argument in favour of the biological importance of a noncanonical branch site sequence and of the possibility to form a 5' stem-loop structure in the pre-U3A snoRNA is the observation of these two features in the pre-U3B snoRNA, in spite of a very high degree of nucleotide sequence divergence of the intron (Brulé et al., 1995).

Altogether, our results reinforced the idea of a strong influence of precursor RNA secondary structure on the efficiency of intron excision by the yeast splicing machinery.

Materials and Methods

Strains and growth conditions.

The *E. coli* strain TG1 (Gibson, 1984) was used for plasmid DNA preparation and the *E. coli* strains HB 2151 and HB 2154 (Carter et al., 1985) for propagation of the M13mp19 bacteriophage. Growth was achieved at 37°C in the Luria Broth Medium with 100 mg/ml of ampicillin added when necessary. For nuclear extract preparation, the *S. cerevisiae* strain EJ101 (Lin et al., 1985) and the thermosensitive strains SPJ11.4 and AH rna8 (Lustig et al., 1986) were grown according to Lin et al. (1985).

Plasmids and bacteriophages.

The pVS1::snR17A plasmid, a pUC18 derivative with an uninterrupted *S. cerevisiae* U3A snoRNA-coding sequence under the control of a T7 promoter (Ségault et al., 1992), was used to produce U3A snoRNA *in vitro* (U3 transcript). The U3A snoRNA gene (exons 1 and 2, and the intron), which was previously cloned under the control of a T7 promoter in the recombinant bacteriophage M13mp9::T7-snR17A (Ségault et al., 1992), was excised with HindIII nuclease and inserted into plasmid pUC18 that was cleaved with the same enzyme. The resulting plasmid

was used to produce pre-U3A snoRNA *in vitro* (pU3G16 transcripts). The M13mp9::T7-snR17A recombinant phage was also used as a template for site-directed mutagenesis. The BS(-) (Short et al., 1988) and SK(+) (Short et al., 1988) plasmids were used for the construction of chimeric U3A snoRNA genes. The EcoRI-SalI restriction fragment, that contains the *S. cerevisiae* actin intron bordered by portions of exon 1 and exon 2 (Lin et al., 1985), was placed under the control of a T7 promoter by inserting it in the pGEM9Zf(-) plasmid cut with the EcoRI and SalI enzymes.

Construction of chimeric U3A snoRNA gene.

Chimeric U3A snoRNA genes, with an extended exon 1 (Fig. 6A), were constructed as follows. The 2.2 kb SalI fragment of plasmid pFL1::snR17A (Myslinski et al., 1990) was inserted in the two possible orientations at the SalI restriction site of bacteriophage M13mp9. The two recombinant bacteriophages generated were denoted M13mp9::U3ASalI$^+$ (U3A gene transcription is oriented in the HindIII to EcoRI direction as referred to M13mp9 restriction sites) and M13mp9::U3ASalI$^-$ (opposite orientation of the U3A gene). To produce pre-U3A snoRNA with a 60 nt-long extension at the 5' end of exon 1, the 0.81 kb HindIII-HpaI fragment from M13mp9::U3ASalI$^+$, which contained the U3A snoRNA gene, was inserted into plasmid SK(+) cleaved with HindIII and SmaI nucleases. The transcript produced from the resulting construct has a 74 nt-long exon 1 and was denoted pU3G74. To produce a pre-U3A snoRNA with a 52 nt-long 5' terminal extension of completely different sequence, the 1.05 kb BamHI-HindIII fragment from M13mp9::U3ASalI$^-$ containing the U3A snoRNA gene, was inserted into the BS(-) plasmid cleaved with the same enzymes. The transcript produced by the resulting construct had a 66 nt-long exon 1 and was denoted pU3G66. Then, truncations were generated within the additional 52 nt segment using a unique NotI restriction site, located in the middle of the 52 nt segment, and Bal31 nuclease. Conditions for Bal31 nuclease were as follows: 5 µg of DNA were digested with 0.2 units of Bal31 nuclease, for 1 or 2 min in 10 µl of 0.1 M EDTA, 1 M NaCl, 0.1 M $CaCl_2$, 0.2 M $MgCl_2$, 0.2 M Tris-HCl pH 8, buffer. After phenol/chloroform extraction and ethanol precipitation, blunt ends were generated with 5 Units of the Klenow fragment in 10 µl of 10 mM $MgCl_2$, 50 mM Tris-HCl pH 7.6 buffer in the presence of 0.05 mM of each dNTP for 30 min at 37°C. Circularization of the blunt-ended DNA was achieved with T4 DNA ligase under standard conditions (Sambrook et al., 1989). The circularized plasmids were used to transform the *E. coli* TG1 strain. Plasmids were prepared from a series of colonies isolated on LB/ampicillin medium. The extent of deletions was determined by sequence analysis using the dideoxynucleotide method (Sanger et al., 1977). Two plasmids with a 21 nt and 29 nt-deletion extension were selected and used to produce chimeric pre-U3A snoRNA; the transcripts produced from these constructs were denoted pU3G66Δ1 and pU3G66Δ2, respectively.

Production of variant U3A snoRNA genes.

Site-directed mutageneses were achieved with the method of Kramer et al. (1984), or with the Amersham commercial kit that uses the recombinant phage M13mp9::T7snR17A as a

template. The oligonucleotides used in these experiments are listed in Table 1. The variant U3A snoRNA genes were inserted into a pUC18 plasmid after cleavage by HindIII nuclease, as described above for the wild-type gene. Oligonucleotide 5 was used to delete the 9 nt-long exon 2 sequence from position 31 to 39, and oligonucleotide 6 to delete the intron sequence between position *+13* to *+16*. Oligonucleotide 3 was used to convert the branch site sequence into a canonical sequence, starting from the pU3G16 or the pU3G66 gene. Finally, oligonucleotide 7 was used to produce a variant pre-U3A snoRNA with a reinforced 5' stem-loop structure (Fig. 6B); this mutation was made both on the wild-type gene and on the variant gene with a UACUAAC$_p$ sequence at the branch site. The name of the *in vitro* transcripts produced from these mutated genes are given in Table I.

In vitro transcription.

The HpaI restriction site located at the 3' end of the U3A coding sequence (Ségault et al., 1992) was used to linearize all DNA templates except the template of transcript pU3G74, which was linearized with the BamHI nuclease. To produce a truncated pU3G16Sau3AI transcript, the Sau3AI nuclease was used for template linearization. The pGEM::Actin plasmid was used to produce actin pre-mRNA transcript after linearization with the EcoRI nuclease. The digested DNAs were phenol extracted, ethanol precipitated and dissolved in water.

Synthesis of cold transcripts for RNA secondary structure-probing experiments were carried out in 125 µl of the following buffer: 10 mM NaCl, 10 mM $MgCl_2$, 10 mM DTT, 40 mM Tris-HCl, pH 8, containing 5 µg of linearized plasmid, 62.5 nmol of each ribonucleoside triphosphate, 125 units of RNase Guard™ (Pharmacia), 140 units of T7 RNA polymerase (Amersham) . After 2 hours of incubation at 37°C, nucleic acids were phenol extracted and ethanol precipitated. Template DNA was digested with 5 units of RNase-free DNase I (Boehringer) in 250 µl of 5 mM $MgCl_2$, 100 mM sodium acetate, pH 5 buffer, for 30 min at 37°C. After phenol extraction and ethanol precipitation, the RNA (about 10 µg) was dissolved in 120 µl of sterile water; 2 µl of this solution was used for each chemical reaction or enzymatic digestion.

Synthesis of labelled transcripts for *in vitro* splicing experiments were carried out with 0.5 µg of cleaved DNA template in 5 µl of 6 mM $MgCl_2$, 4 mM spermidine, 10 mM DTT, 40 mM Tris-HCl pH 8, in the presence of 0.5 mM ATP, CTP and GTP, 0.2 mM UTP, 80 units of T7 RNA polymerase, 10 µCi of [α^{-32}P]UTP (800 Ci/mmole) and 10 units of RNasin. Incubation was at 37°C for 2 hours. Transcription reactions were stopped by adding 5 µl formamide loading buffer. Labelled transcripts were purified by electrophoresis on 6% polyacrylamide gel and were eluted at 4°C in 0.5 M sodium acetate, 1 mM EDTA pH 8, in the presence of 2.5% phenol, for 16 hours. The RNA from the aqueous phase was ethanol precipitated and dissolved in 50 µl of H_2O for *in vitro* splicing experiments.

Preparation of splicing extract and *in vitro* splicing assay.

Yeast whole-cell extracts were prepared according to Lin et al. (1985), using either the wild-type strain (EJ101) or the thermosensitive strains (SPJ11.4 and AH rna8). Heat

inactivation of the extracts was performed at 30°C for 30 min, as previously described (Lustig et al., 1986). Splicing assays were carried out in a buffer containing 3 mM $MgCl_2$, 60 mM KPO_4 pH 7 with 1 mM spermidine, 2 mM ATP, 3% PEG, 0.4 nM uncapped transcript and 40% (V/V) of splicing extract for 1 hour at 16°C, as previously described by Newman et al., 1985. The M/P ratio was evaluated by densitometry of the autoradiogram with the Gel Doc 1000 UV Fluorescent System (Biorad). Calculation were made with the Biorad software taking into account the lengths of the mature and precursor RNAs.

Non denaturing gel electrophoresis of splicing complexes.

Complex assembly analysis was performed as previously described by Cheng & Abelson (1987). RNA transcripts were incubated under the splicing conditions described above. 5 µl aliquots were recovered after 0, 5, 10 and 40 min of incubation. Heparin at a concentration of 2 µg/µl of whole cell extract was added to these aliquots as described by Cheng & Abelson (1987), and the material was fractionated on a 4% nondenaturing polyacrylamide gel (acrylamide/bisacrylamide, 80:1) made up in 1 mM EDTA, 20 mM Tris-HCl pH 8 buffer. Gels were run at 25 V/cm for 4, 5 hours at 4°C, and were dried prior to autoradiography.

Chemical modifications.

Chemical modifications were performed either under nondenaturing conditions or under semi-denaturing conditions (Ehresmann et al., 1987; Ségault et al., 1992). Each modification was carried out on about 200 ng of *in vitro* synthesized RNA in the presence of 5 µg of a commercial yeast tRNA mixture (Boehringer). Prior to the chemical reaction, the RNA was dissolved in 100 µl of the appropriate buffer and pre-incubated 10 min at 20°C (temperature used for the modification reactions). Dimethylsulfate (DMS) modifications under nondenaturing conditions were made in 50 mM KCl, 10 mM $MgCl_2$, 100 mM sodium cacodylate, pH 7.5 buffer and under semi-denaturing conditions in 2 mM EDTA, 100 mM sodium cacodylate pH 7.5 buffer, with either 0.75 µl of pure DMS or 0.75 µl of a DMS solution (1/1 (v/v) DMS/EtOH) per assay. The incubation was for 15 min, and the reaction was stopped by ethanol precipitation. The RNA was washed with 70% ethanol and dissolved in 14 µl of H_2O. 7 µl of this solution were used for a reverse transcriptase elongation assay. 1-cyclohexyl-3-(2-morpholino ethyl)-carbodiimide-metho-p-toluene sulfonate (CMCT) modifications under nondenaturing conditions were made in 50 mM KCl, 10 mM $MgCl_2$, 100 mM sodium borate, pH 8 buffer, and under semi-denaturing conditions, in the same buffer except that 2 mM EDTA were used instead of 10 mM $MgCl_2$ and 50 mM KCl, with 1.05 or 2.1 mg of CMCT per 100 µl assay. Incubation was for 30 min. At the end of the reaction, the same procedure was used as for DMS modification.

Enzymatic cleavages.

As for chemical modifications, the enzymatic digestions were performed on T7 RNA transcripts mixed with 5 µg of commercial tRNA. V1 RNase was prepared from *Naja oxiana* venom (Vassilenko & Babkina, 1965). RNA was pre-incubated 10 min at 0°C in 10 µl of

reaction buffer (350 mM KCl, 10 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.5). Incubation was for 2, 7 or 10 min with an enzyme/RNA ratio of 1.2 unit/µg. The reaction was stopped by adding 5 µl of 100 mM EDTA, followed by phenol extraction. RNAs digested for 2 and 7 min were mixed. After ethanol precipitation, the RNA was washed with 70% ethanol, and subjected to primer extension analysis. S1 nuclease was from Pharmacia. RNA was pre-incubated 10 min at 20°C in 10 µl of reaction buffer (50 mM KCl, 10 mM $MgCl_2$, 1 mM $ZnCl_2$, 25 mM Na acetate, pH 4.5). Incubation was for 12 or 25 min at 20°C with 2 units of enzyme. The reaction was stopped by adding 2 µl of 100 mM EDTA, extracted with phenol, and the RNA was prepared for primer extension analysis as above.

Primer extension analysis.

Chemical modifications and enzymatic cleavages in the pre-U3A snoRNA transcripts were analyzed with reverse transcriptase. The oligonucleotides primers used were complementary to the pre-U3A snRNA sequences from position 319 to 333 (primer 1), 92 to 106 (primer 2), *+129* to *+150* (primer 3) and *+1* to *+19* (primer 4). Primer oligonucleotides were 5' end labelled with [γ-^{32}P] ATP (Amersham) 3000 mCurie/mMole. They were annealed to 7 µl aliquots of the RNA samples described above, for 10 min at 65°C in 40 mM KCl, 6 mM $MgCl_2$, 50 mM Tris-HCl, pH 8.3 buffer, and the solution was cooled down to room temperature. Reverse transcription was performed with 1 unit of Rous sarcoma virus 2 reverse transcriptase (Amersham) or 1 unit of Avian myeloblastosis virus reverse transcriptase (Life Science), for 30 min at 45°C, in the presence of 125 µM of each dNTP. To prepare sequencing ladders of unmodified RNA, dideoxynucleotide: deoxynucleotide mixtures (in a 1/2 ratio) were used. Reverse transcripts were fractionated by electrophoresis on a 7% polyacrylamide sequencing gel (Sambrook et al., 1989).

Estimation of RNA secondary structure stability.

Estimation of stem-loop structure stability was made with the Mfold Program of the GCG software Version 8.1 Unix (1995), that is based on the thermodynamic values proposed by Jeager et al. (1989) for an RNA in solution in 1 M NaCl buffer at 37°C.

Acknowledgements

J. Bayeul is thanked for her excellent technical assistance and Kyle Tanner for checking/reading through the manuscript. A. Grégoire, V. Ségault, R. Fournier and F. Brulé were fellows of the French Ministère de la Recherche et de l'Enseignement Supérieur. This work was supported by the Centre National de la Recherche Scientifique and the French Ministère de la Recherche et de l'Enseignement Supérieur.

References

Beltrame M, Tollervey D. 1992. Identification and functional analysis of two U3 binding sites on yeast pre- ribosomal RNA. EMBO J. 11: 1531-1542.

Beltrame M, Tollervey D. 1995. Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. EMBO J. 14: 4350-4356.

Brulé F, Grégoire A, Ségault V, Mougin A, Branlant C. 1995. Secondary structure conservation of the U3 small nucleolar RNA introns in *Saccharomyces*. C. R. Acad. Sci. Paris, Life Sciences. 318:1197-1206.

Brulé F, Venema J, Ségault V, Tollervey D, Branlant C. 1996. The yeast *Hansenula wingei* U3 snoRNA gene contains an intron and its coding sequence co-evolved with the 5' ETS region of the pre-ribosomal RNA. RNA. (In Press).

Carter P, Bedouelle H, Winter G. 1985. Improved oligonucleotide site-directed mutagenesis using M13 vectors. Nucl. Acids Res. 13: 4431-4443.

Chang TH, Clark MW, Lustig AJ, Cusick ME, Abelson J. 1988. RNA11 protein is associated with the yeast spliceosome and is localized in the periphery of the cell nucleus. Mol Cell Biol. 8: 2379-93.

Cheng SC, Abelson J. 1987. Spliceosome assembly in yeast. Genes Dev. 1: 1014-1027.

Clouet d'Orval B, d'Aubenton Carafa Y, Sirand-Pugnet P, Gallego M, Brody E, Marie J. 1991. RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. Science. 252: 1823-1828.

Craig N, Kass S, Sollner-Webb B. 1987. Nucleotide sequence determining the first cleavage site in the processing of mouse precursor rRNA. Proc Natl Acad Sci USA. 84: 629-633.

Deshler JO, Rossi JJ. 1991. Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron. Genes Dev. 5: 1252-1263.

Duchêne M, Low A, Schweizer A, Domdey H. 1988. Molecular consequences of truncations of the first exon for *in vitro* splicing of yeast actin pre-mRNA. Nucl. Acids Res. 16: 7233-7239.

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. Nucl. Acids Res. 15: 9109-9128.

Eichler DC, Craig N. 1994. Processing of eukaryotic ribosomal RNA. Prog Nucl. Acid Res Mol Biol. 49: 197-239.

Eng FJ, Warner JR. 1991. Structural basis for the regulation of splicing of a yeast messenger RNA. Cell. 65: 797-804.

Gibson TJ. 1984. Studies on the Epstein Barr virus genome. PhD thesis, Cambridge University, England.

Goguel V, Rosbash M. 1993. Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. Cell. 72: 893-901.

Goguel V, Wang Y, Rosbash M. 1993. Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing. Mol Cell Biol. 13: 6841-6848.

Hall SL, Padgett RA. 1996. Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns. Science. 271: 1716-1718.

Hartshorne T, Agabian N. 1994. A common core structure for U3 small nucleolar RNAs. Nucl. Acids Res. 22: 3354-3364.

Hawkins JD. 1988. A survey on intron and exon lengths. Nucl. Acids Res. 16: 9893-9908.

Hughes JM, Ares M. 1991. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. EMBO J. 10: 4231-4239.

Jackson SP, Lossky M, Beggs JD. 1988. Cloning of the RNA8 gene of *Saccharomyces cerevisiae*, detection of the RNA8 protein, and demonstration that it is essential for nuclear pre-mRNA splicing. Mol Cell Biol. 8: 1067-1075.

Jacquier A, Rodriguez JR, Rosbash M. 1985. A quantitative analysis of the effects of 5' junction and TACTAAC box mutants and mutant combinations on yeast mRNA splicing. Cell. 43: 423-430.

Jeager JA, Turner DH, Zucker M. 1989. Improved predictions of secondary structures for RNA. Proc. Natl. Acad. Sci. USA. 86: 7706-7710.

Kalogeropoulos A. 1995. Automatic intron detection in nuclear DNA sequences of *Saccharomyces cerevisiae*. Yeast 11: 555-565.

Kass S, Craig N, Sollner-Webb B. 1987. Primary processing of mammalian rRNA involves two adjacent cleavages and is not species specific. Mol Cell Biol. 7: 2891-2898.

Kass S, Tyc K, Steitz JA, Sollner-Webb B. 1990. The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. Cell. 60: 897-908.

Kramer B, Kramer W, Fritz HJ. 1984. Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*. Cell. 38: 879-887.

Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. 1980. Are snRNP involved in splicing? Nature. 283: 220-224.

Li H, Dalal S, Kohler J, Vilardell J, White SA. 1995. Characterization of the pre-mRNA binding site for yeast ribosomal protein L32: the importance of a purine-rich internal loop. J.Mol.Biol. 250: 447-459.

Libri D, Stutz F, Mc Carthy T, Rosbash M. 1995. RNA structural patterns and splicing: Molecular basis for an RNA-based enhancer. RNA. 1: 425-436.

Lin RJ, Newman AJ, Cheng SC, Abelson J. 1985. Yeast mRNA splicing *in vitro*. J Biol Chem. 260: 14780-14792.

Liu HX, Goodall GJ, Kole R, Filipowicz W. 1995. Effects of secondary structure on pre-mRNA splicing: hairpins sequestering the 5' but not the 3' splice site inhibit intron processing in *Nicotiana plumbaginifolia*. EMBO J. 14: 377-388.

Lübben B, Marshallsay C, Rottmann N, Lührmann R. 1993. Isolation of U3 snoRNP from CHO cells: a novel 55 kDa protein binds to the central part of U3 snoRNA. Nucl. Acids Res. 21: 5377-5385.

Lustig AJ, Lin RJ, Abelson J. 1986. The yeast RNA gene products are essential for mRNA splicing *in vitro*. Cell. 47: 953-63.

Maxwell ES, Fournier MJ. 1995. The small nucleolar RNA. Annu. Rev. Biochem. 67: 6523-6527.

Miller AM. 1984. The yeast MATa1 gene contains two introns. EMBO J. 3: 1061-1065.

Mizuta K, Hashimoto T, Otaka E. 1992. Yeast ribosomal proteins: XIII. *Saccharomyces cerevisiae* YL8A gene, interrupted with two introns, encodes a homologue of mammalian L7. Nucl. Acids Res. 20: 1011-1016.

Mougey EB, Pape LK, Sollner-Webb B. 1993. A U3 small nuclear ribonucleoprotein-requiring processing event in the 5' external transcribed spacer of *Xenopus* precursor rRNA. Mol Cell Biol. 13: 5990-5998.

Mount SM, Petterson I, Hinterberger M, Karmas A, Steitz JA. 1983. The U1 small nuclear RNA protein complex selectively binds a 5' splice site *in vitro*. Cell. 33: 509-518.

Myslinski E, Ségault V, Branlant C. 1990. An intron in the genes for U3 small nucleolar RNAs of the yeast *Saccharomyces cerevisiae*. Science. 247: 1213-1216.

Newman AJ, Lin RJ, Cheng SC, Abelson J. 1985. Molecular consequences of specific intron mutations on yeast mRNA splicing in vivo and *in vitro*. Cell. 42: 335-344

Newman A. 1987. Specific accessory sequences in *Saccharomyces cerevisiae* introns control assembly of pre-mRNAs into spliceosomes. EMBO J. 6: 3833-3839.

Parker KA, Steitz JA. 1987. Structural analysis of the human U3 ribonucleoprotein particle reveal a conserved sequence available for base pairing with pre-rRNA. Mol Cell Biol. 7: 2899-913.

Parker R, Patterson B. 1987. Architecture of fungal introns: Implications for spliceosome assembly. New York: Inouye, M. & Dubock, B. eds.

Pikielny CW, Rosbash M. 1985. mRNA splicing efficiency in yeast and the contribution of nonconserved sequences. Cell. 41: 119-126.

Raué HA, Planta RJ. 1995. The pathway to maturity: Processing of ribosomal RNA in *Saccharomyces cerevisiae*. Gene expression. (In press).

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular Cloning. A laboratory manual. New York: Cold Spring Harbor Laboratory Press eds.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74: 5463-5467.

Ségault V, Mougin A, Grégoire A, Banroques J, Branlant C. 1992. An experimental study of *Saccharomyces cerevisiae* U3 snRNA conformation in solution. Nucl. Acids Res. 20: 3443-3451.

Short JM, Fernandez JA, Sorge JA, Huse WD. 1988. Lambda ZAP: A bacteriophage Lambda expression vector with in vivo excision properties. Nucl. Acids Res. 16: 7583.

Smith CW, Chu TT, Nadal-Guinard B. 1993. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. Mol Cell Biol. 13: 4939-4952.

Takahashi Y, Urushiyama S, Tani T, Ohshima Y. 1993. An mRNA-type intron is present in the *Rhodotorula hasegawae* U2 small nuclear RNA gene. Mol Cell Biol. 13: 5613-5619.

Tani T, Ohshima Y. 1989. The gene for the U6 small nuclear RNA in fission yeast has an intron. Nature. 337: 87-90.

Tani T, Ohshima Y. 1991. mRNA-type introns in U6 small nuclear RNA genes: implications for the catalysis in pre-mRNA splicing. Genes Dev. 5: 1022-1031.

Tarn WY, Steitz JA. 1996. A novel spliceosome containing U11, U12 and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. Cell. 84: 801-811.

Teare J, Wollenzien P. 1989. Structures of human and rabbit ß-globin precursor messenger RNAs in solution. Biochemistry. 28: 6208-6219.

Teare J, Wollenzien P. 1990. The structure of a pre-mRNA molecule in solution determined with a site directed cross-linking reagent. Nucl. Acids Res. 18: 855-864.

Vassilenko SK, Babkina GT. 1965. Biokymiya. 30: 705-716.

Vijayraghavan U, Parker R, Tamm J, Iimura Y, Rossi J, Abelson J, Guthrie C. 1986. Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. EMBO J. 5: 1683-95.

Vilardell J, Warner JR. 1994. Regulation of splicing at an intermediate step in the formation of the spliceosome. Genes Dev. 8: 211-220.

Zhuang Y, Weiner AM. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell. 46: 827-835.

Zucker M. 1989. On finding all suboptimal foldings of an RNA molecule. Science. 244: 48-52.

N°	Sequence	Target sequence	Generated mutation	Resulting transcripts
3	5' ATGTGTTAGTAAAAAGCTGCTG 3'	(+*150*)-(+*129*)	G+*140*→ T	pU3U16 and pU3U66
5	5' GTCAAAGAG↓TTATAGAAATGAT 3'	(19)-(49)	Δ (G31 to C39)	pU3G16Δ9
6	5' GGGTAAAATGTTT↓TATATTACATACTG 3'	(13)-(+*29*)	Δ (C+*13* to C+*16*)	pU3G16Δ4C
7	5' TGTTTGGGGTCGACATACTGAA 3'	(11)-(+*21*)	(A+*7*) to (U+*11*)→CG	pU3G16ds5' and pU3U16ds5'

Table 1: Oligonucleotides used for site-directed mutagenesis.

The oligonucleotide sequences complementary to the snoRNA gene are underlined and vertical arrows indicate deletions. The exon-targeted sequences are numbered as in the mature U3A snoRNA (Ségault et al., 1992) and nucleotide positions in the intron are numbered by italic numbers from +*1* to +*157*. The nomenclature for transcripts produced from the mutated genes is as follow: pU3 for the pre-U3A snoRNA, G or U for GACUAACp or UACUAACp branch site sequence followed by the size of exon 1 and the identification of the mutation (*ie* Δ4C for deletion of 4 C residues).

Legends to figures

Figure 1: Secondary structure analysis of exon 2.

Transcript pU3G16 was subjected to chemical (DMS and CMCT) and enzymatic (V1 and S1 nucleases) probing of the RNA secondary structure, under nondenaturing conditions. Positions of chemical modifications and enzymatic cleavages were identified by primer extension analysis with primers 1 and 2 (inset in Panel A). The stem-loop structure nomenclature is according to Ségault et al., 1992. The location of the 5 phylogenetically conserved boxes A, B, C, C' and D (for review, see Maxwell & Fournier, 1995) and the calculated free energies of the stem-loop 1b and of the cruciform structure are shown in the inset of Panel A. The results of probing experiments are schematically represented on the proposed exon 2 structure in Panel A. Nucleotides modified by DMS or CMCT are circled. The colours of the circles indicate the intensity of modification: green, orange and red for low, medium and strong, respectively. Nucleotides in blue squares were never modified by chemical reagents. Arrows linked to a circle and arrows linked to a square indicate S1 RNase and V1 RNase cleavages, respectively. The colours of squares and circles indicate the yield of cleavage with the same rules as for chemical modifications. Differences of enzymatic cleavages in the precursor RNA as compared to the mature RNA are represented as follows: the Δ indicates decreased cleavage intensity, the + indicates increased cleavage intensity and the N indicates a new cleavage. In Panel B, the accessibility to S1 and V1 nucleases of pU3G16 transcript from position 45 to 85 is analyzed by primer extension with primer 2. Lane c, control lane without enzyme; lane s, short incubation time (2 or 7 min for V1 and 12 min for S1); lane l, long incubation time (10 min for V1 and 25 min for S1). Lanes A, C, G and U show the sequencing ladder of the respective nucleotide. Positions of loop and helix are indicated on the left of the Panel.

Figure 2: Chemical and enzymatic probing of the intron and exon 1 secondary structure.

The pU3G16 transcript was digested with S1 and V1 nucleases (same nomenclature of the Lanes as in Figure 1) and modified with DMS or CMCT chemical reagents. For chemical modifications: c represents control experiment in the absence of chemical reagent; n and d, chemical modifications made under nondenaturing and semi-denaturing conditions, respectively. For DMS, n1 and n2 refer to DMS/RNA ratios of 0.072 µl/µg and 0.144 µl/µg, respectively. In semi-denaturing conditions, the DMS/RNA ratio was of 0.144 µl/µg. Reverse transcription analysis were made with primer 2 (Panel B) and primer 3 (Panel C). The positions of primers 2 and 3 on the pre-RNA are schematically represented in Panel A, E1 and E2 indicate the position of exon 1 and 2, respectively.

Figure 3: The two possible conformations of the intron and exon 1 as deduced from RNA structure-probing experiments.

The structure of the intron and exon 1 in the A conformation is shown in Panel A, differences in the B conformation are shown in Panel B. Positions of enzymatic cleavages and chemical

modifications are indicated with the same nomenclature as in Figure 1A. The 5' and 3' splice sites are indicated as 5' ss and 3' ss, respectively. Panels C and D show differences between the pU3G16 transcript and the pU3G16Δ9, pU3G16Sau3AI transcripts and the pU3GΔ4C transcript, three variants that were made to distinguish between the two possible conformations.

Figure 4: Chemical and enzymatic probing of exon 1 structure in the pU3G16Δ9 and pU3G16Δ4C transcripts.

Transcripts pU3G16Δ9 (Fig. 3C) and pU3G16Δ4C (Fig. 3D) were subjected to enzymatic digestions and chemical modifications under nondenaturing conditions. Reverse transcriptase analysis with primer 3 (Fig. 2A) is shown in the upper part of Panels A and B, respectively. Results of probing experiments are schematically represented on the proposed structures at the bottom of each Panel. The nomenclature is that used in Figure 1A, except that the green, orange and grey colors are replaced by white, grey and black, respectively. Boxed nucleotides were clearly shown to be unmodified by chemical reagents. The 5' splice site is indicated.

Figure 5: A proposed secondary structure model for the pre-U3A snoRNA.

The sequence is that of the pre-U3A snoRNA from *S. cerevisiae* FL100 (Myslinski et al., 1990). The differences found in the pre-U3A* snoRNA from *S. cerevisiae* GY5 (Brulé et al., 1995) are indicated by circled nucleotides. The phylogenetically conserved Boxes A, B, C, C' and D, the 5' and 3' splice sites, and branch site are boxed. Helix and stem-loop structure numbering is derived from that used in the mature U3A snoRNA. The two nucleotide sequences that may be involved in a pseudoknot structure are shadowed. For the intron central stem-loop structure, the free energies calculated for the entire structure (ΔG on the left), the upper part of the structure and for the base-paired segments at the bottom (ΔGs on the right) are given.

Figure 6: Comparative analysis of the *in vitro* splicing efficiency of the chimeric and variant pre-U3A snoRNA.

Panel A: the effect of the length and sequence of an extension at the 5' end of exon 1 is tested. The sequences of the extensions are shown at the bottom and the products of the *in vitro* splicing reaction are shown above. The products were fractionated by electrophoresis on a 6% polyacrylamide gel under denaturing conditions. An actin transcript was included for comparison. Positions of precursor RNA (pre-RNA), first intermediate (IVS-E2), mature RNA (E1-E2) and intron (IVS) are as indicated. The mature/precursor RNA ratio (M/P) and the values obtained after normalization relative to the M/P ratio of transcript pU3G66 taken as 100% (M/Pnor.%) are given. Panel B : the effect of a GACUAACp or UACUAACp branch site sequence in chimeric and variant pre-U3A snoRNA transcripts is tested. The mutation in the pU3G16ds5' and pU3U16ds5' transcripts, which convert the irregular stem of the 5' stem-loop structure into a perfect 12 bp helix, is shown as well as the four cytosine deletion in transcript pU3G16Δ4C. The splicing products were analyzed as in Panel A.

Figure 7: Secondary-structure analysis of the 5' terminal extensions.
The four transcripts: pU3G66, pU3G66Δ1, pU3G66Δ2 and pU3G16ds5' were subjected to chemical and enzymatic probing of the RNA structure and the products reverse transcribed. Primer extension analysis of cleavages and modifications in transcript pU3G66, using primer 4 is shown in Panel A (same nomenclature as in Figures 1 and 2). In Panel B, the results obtained for the 5' terminal region of the four transcripts are schematically represented on the proposed secondary structures. The symbols are as in Figure 4.

Figure 8: A kinetic study of spliceosomal complex assembly on variant and chimeric pre-U3A snoRNAs.
Spliceosomal complex assembly was made with nuclear extracts of strain EJ101 under the same conditions as for the splicing assay. 5 µl aliquots were taken after the indicated times, in minutes of incubation, heparin was added and the spliceosomal complexes were fractionated on 4% nondenaturing gel. The positions of complexes A_1, A_2 and B (Cheng & Abelson, 1987) are as indicated.

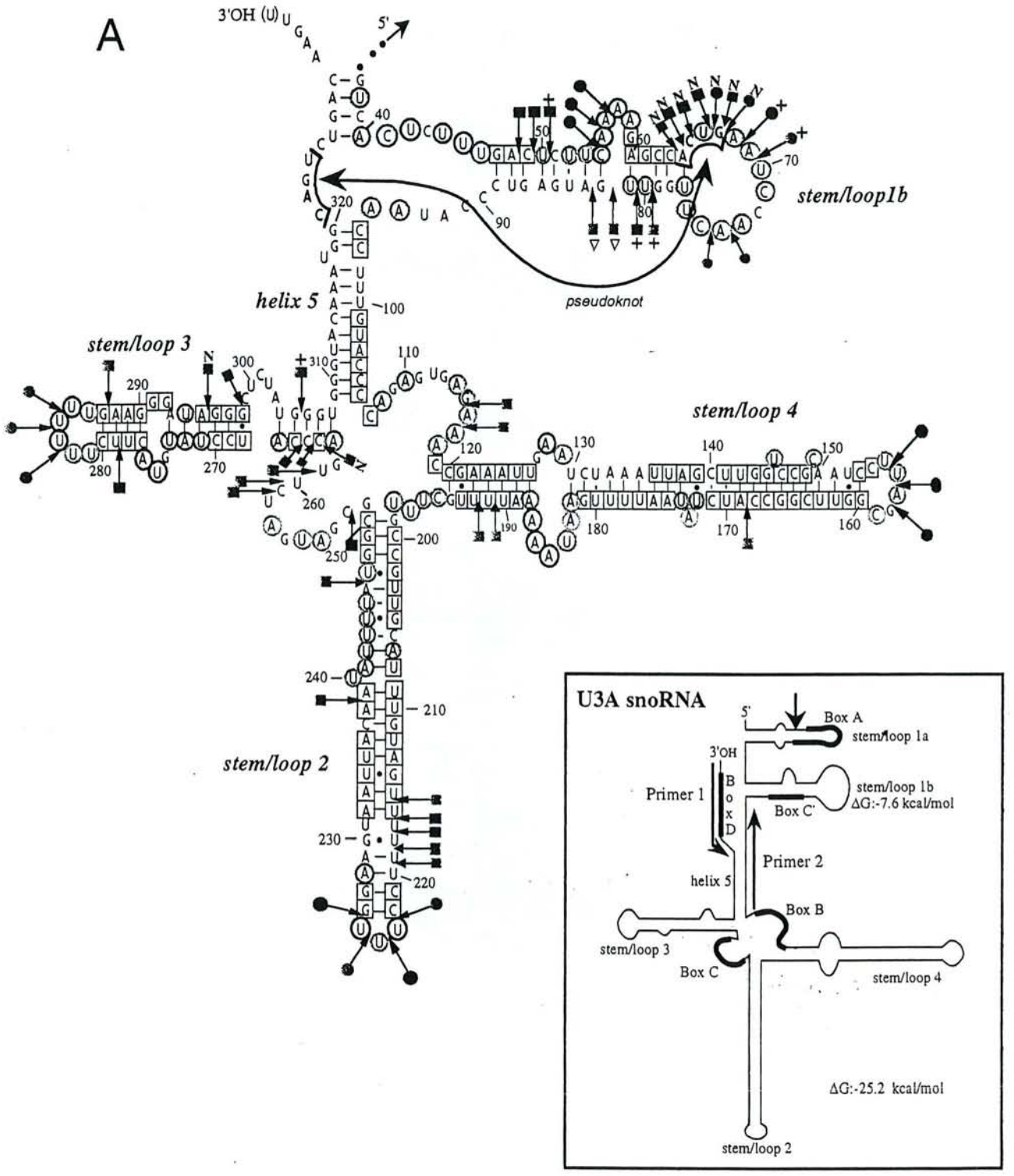

A
3'OH (U)
5'
stem/loop1b
pseudoknot
helix 5
stem/loop 3
stem/loop 4
stem/loop 2
U3A snoRNA
5'
Box A
stem/loop 1a
3'OH
Primer 1
Box D
stem/loop 1b
ΔG:-7.6 kcal/mol
Box C'
Primer 2
helix 5
Box B
stem/loop 3
Box C
stem/loop 4
ΔG:-25.2 kcal/mol
stem/loop 2

Figure 14

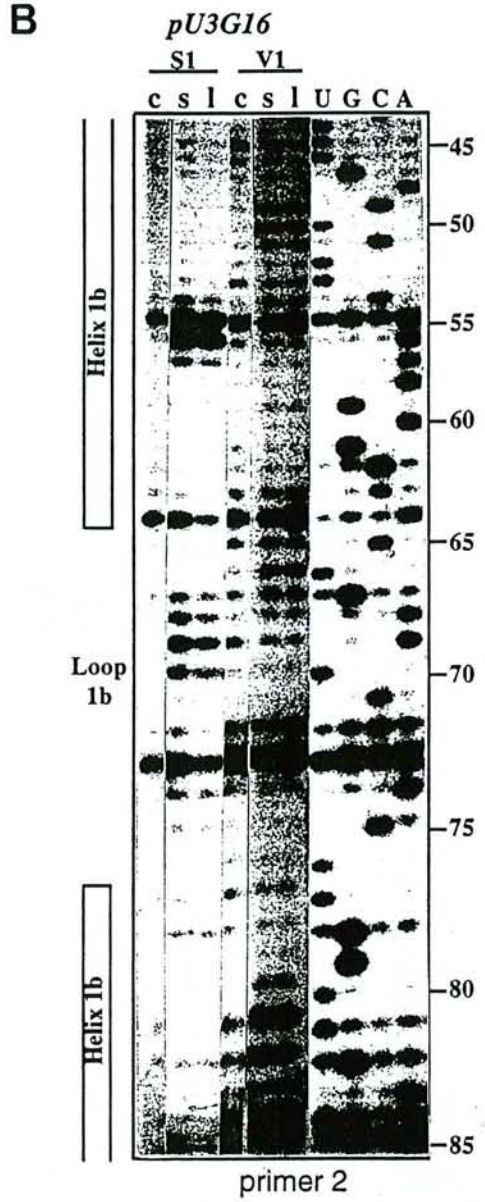
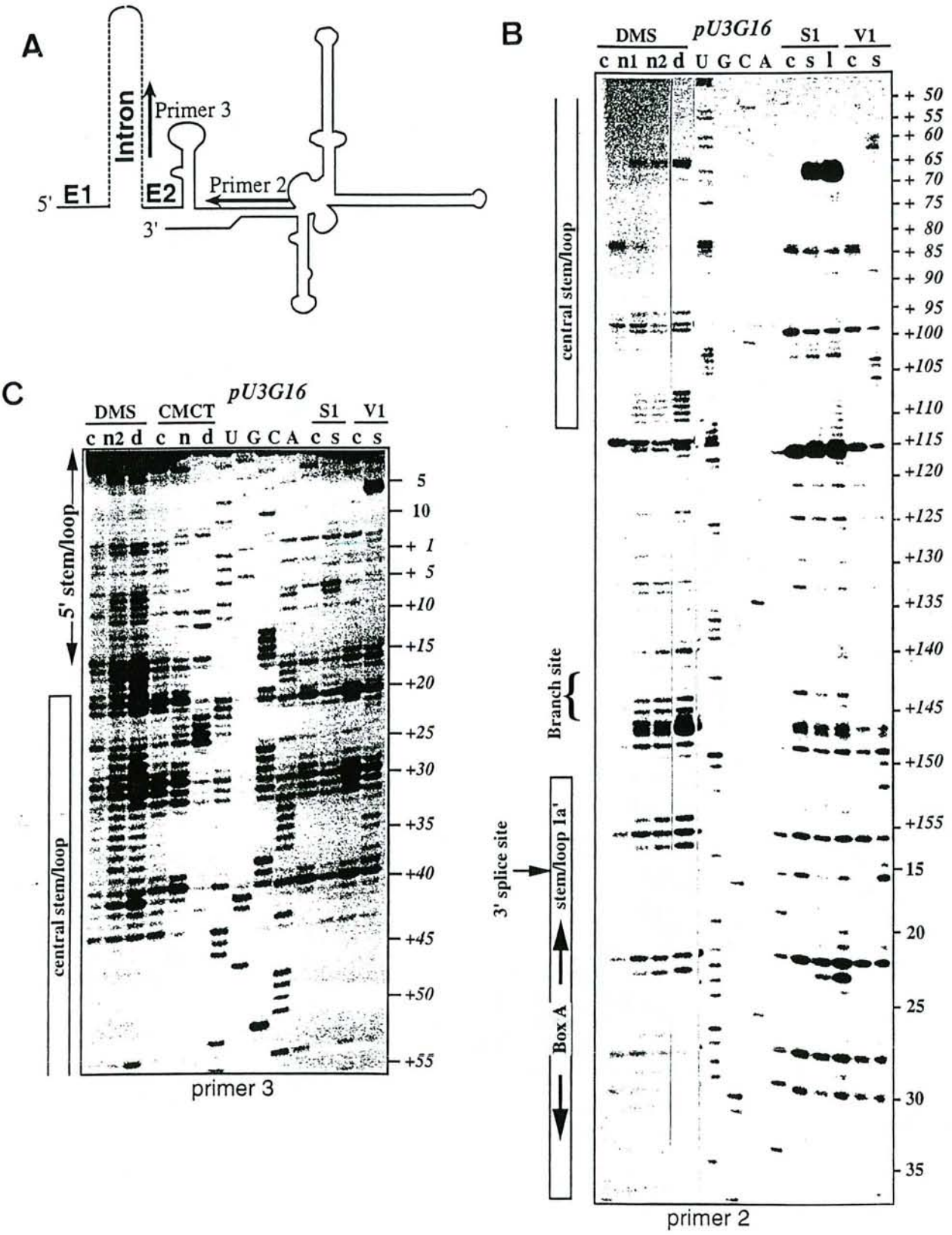
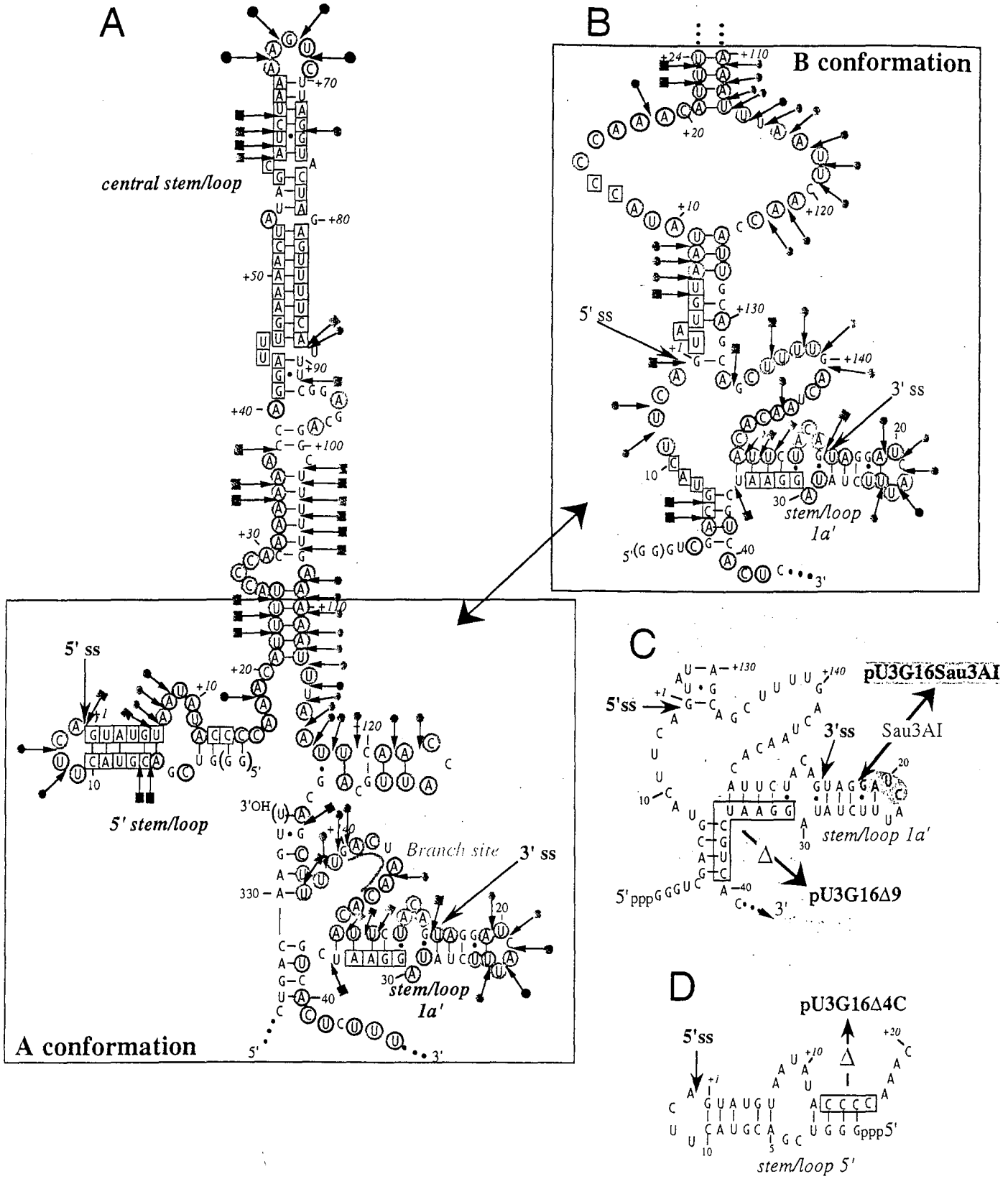
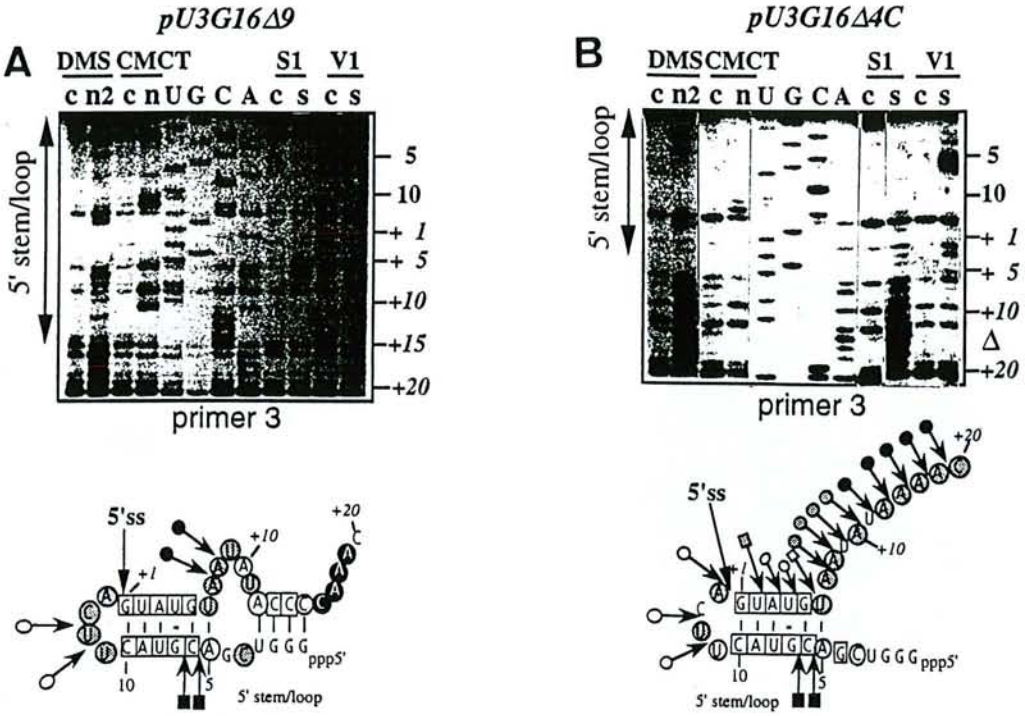
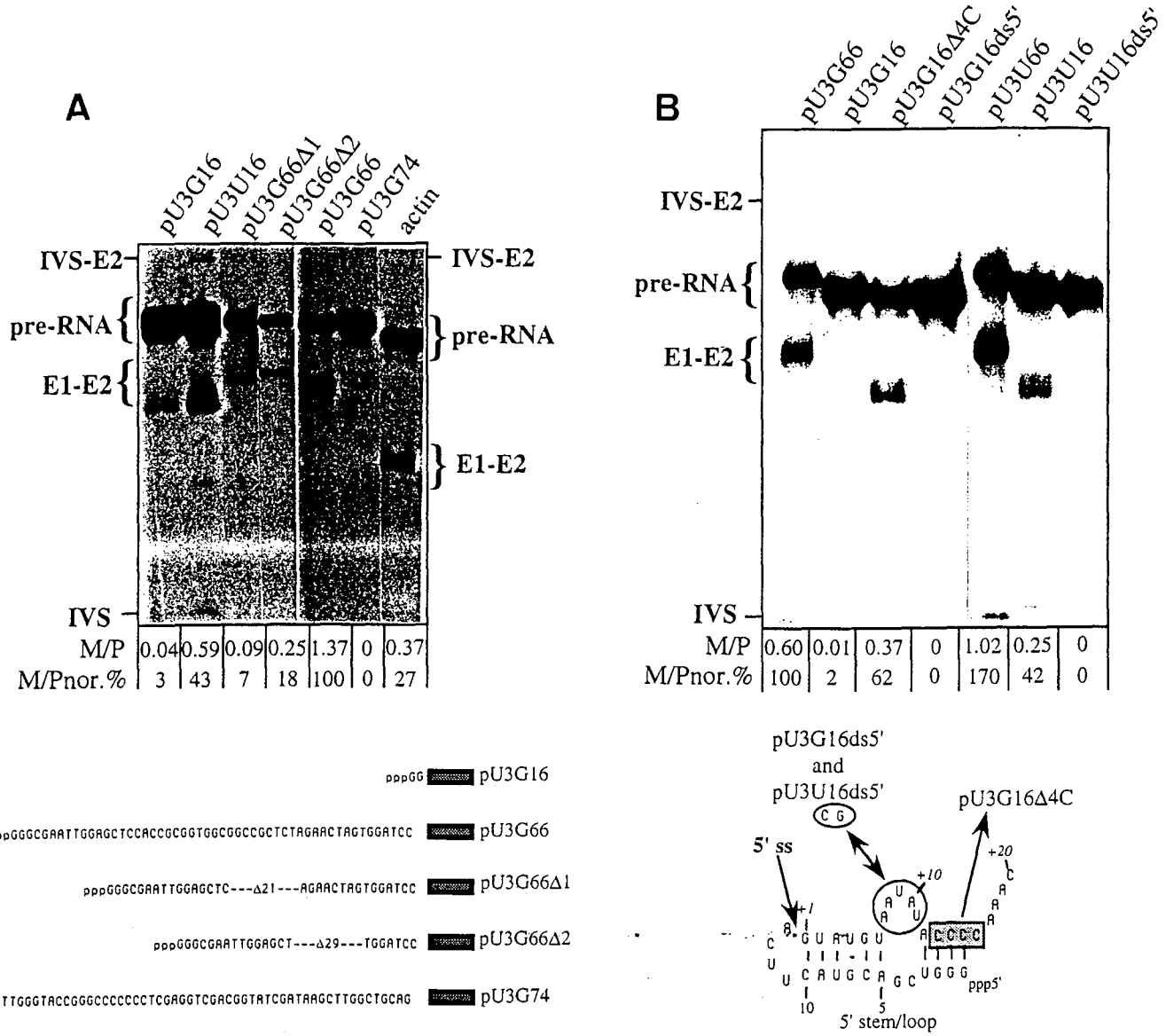

B
pU3G16
S1
V1
c s l c s l U G C A
Helix 1b
Loop 1b
Helix 1b
45
50
55
60
65
70
75
80
85
primer 2

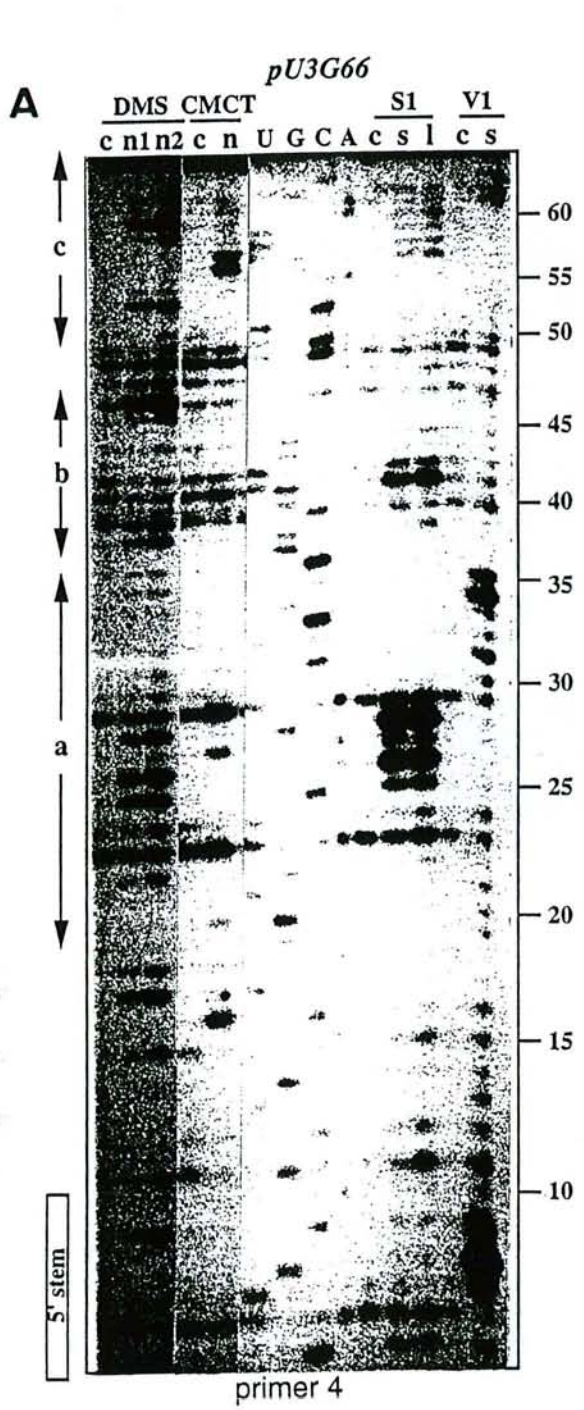
Figure 13

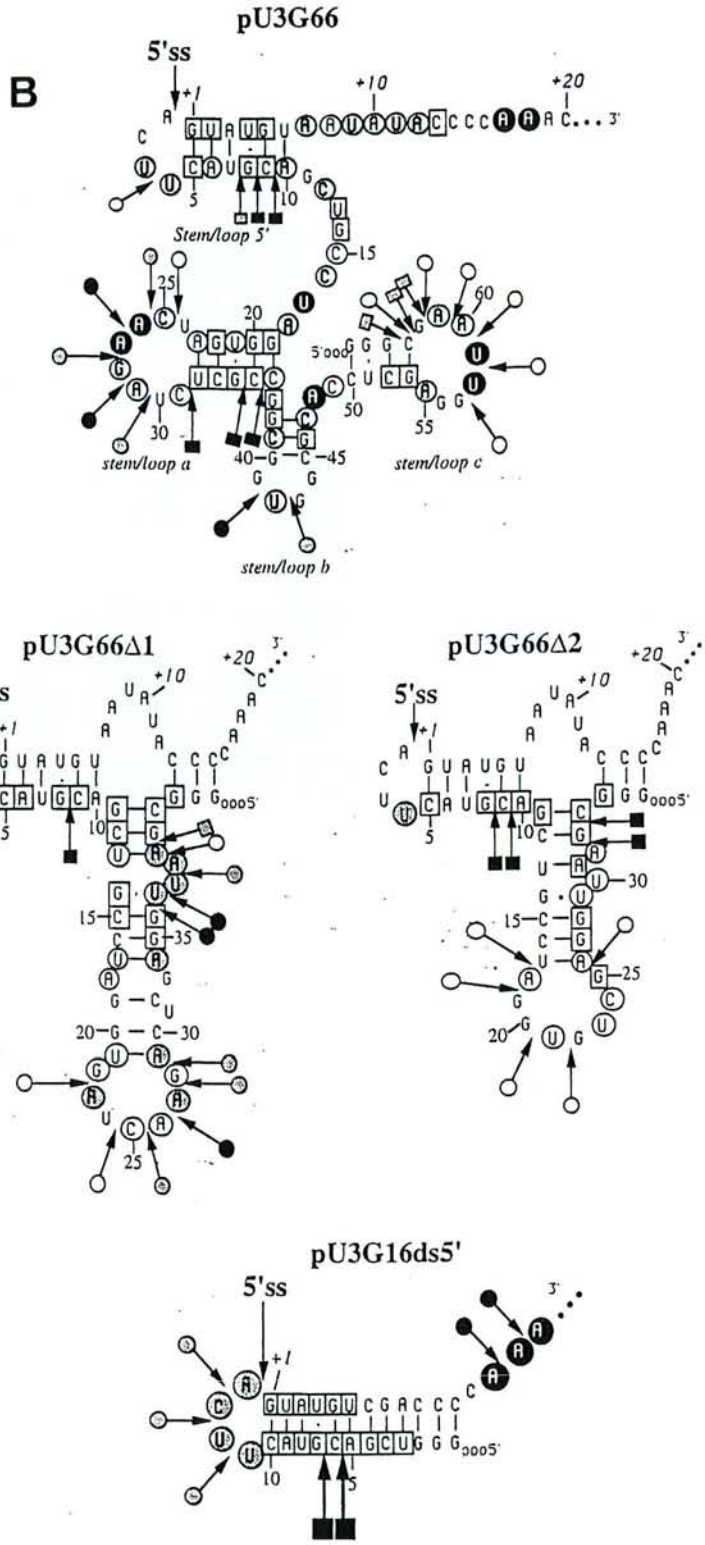

A
Intron
Primer 3
Primer 2
5'
E1
E2
3'
B
DMS
pU3G16
S1
V1
c n1 n2 d U G C A c s l c s
central stem/loop
Branch site
3' splice site
stem/loop 1a'
Box A
primer 2
C
pU3G16
DMS
CMCT
S1
V1
c n2 d c n d U G C A c s c s
5' stem/loop
central stem/loop
primer 3


A
central stem/loop
5' ss
5' stem/loop
3'OH
Branch site
3' ss
stem/loop 1a'
A conformation
B
B conformation
5' ss
3' ss
stem/loop 1a'
C
pU3G16Sau3AI
5'ss
3'ss
Sau3AI
stem/loop 1a'
pU3G16Δ9
D
pU3G16Δ4C
5'ss
stem/loop 5'


pU3G16Δ9
A
DMS CMCT
S1
V1
c n2 c n U G C A c s c s
5' stem/loop
5
10
+ 1
+ 5
+10
+15
+20
primer 3
5'ss
+1
+10
+20
10
5
5' stem/loop
pU3G16Δ4C
B
DMS CMCT
S1
V1
c n2 c n U G C A c s c s
5' stem/loop
5
10
+ 1
+ 5
+10
Δ
+20
primer 3
5'ss
+10
+20
10
5
5' stem/loop


A
pU3G16
pU3U16
pU3G66Δ1
pU3G66Δ2
pU3G66
pU3G74
actin
IVS-E2
pre-RNA
E1-E2
IVS
M/P 0.04 0.59 0.09 0.25 1.37 0 0.37
M/Pnor.% 3 43 7 18 100 0 27
B
pU3G66
pU3G16
pU3G16Δ4C
pU3G16ds5'
pU3U66
pU3U16
pU3U16ds5'
IVS-E2
pre-RNA
E1-E2
IVS
M/P 0.60 0.01 0.37 0 1.02 0.25 0
M/Pnor.% 100 2 62 0 170 42 0
pppGG pU3G16
pppGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCC pU3G66
pppGGGCGAATTGGAGCTC---Δ21---AGAACTAGTGGATCC pU3G66Δ1
pppGGGCGAATTGGAGCT---Δ29---TGGATCC pU3G66Δ2
pppGGGCGAATTGGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGGCTGCAG pU3G74
pU3G16ds5'
and
pU3U16ds5'
pU3G16Δ4C
5' ss
+1
+10
+20
ppp5'
10
5
5' stem/loop


A
pU3G66
DMS CMCT
S1
V1
c n1 n2 c n U G C A c s l c s
c
b
a
60
55
50
45
40
35
30
25
20
15
10
5' stem
primer 4


B
pU3G66
5'ss
Stem/loop 5'
stem/loop a
stem/loop b
stem/loop c
pU3G66Δ1
pU3G66Δ2
pU3G16ds5'

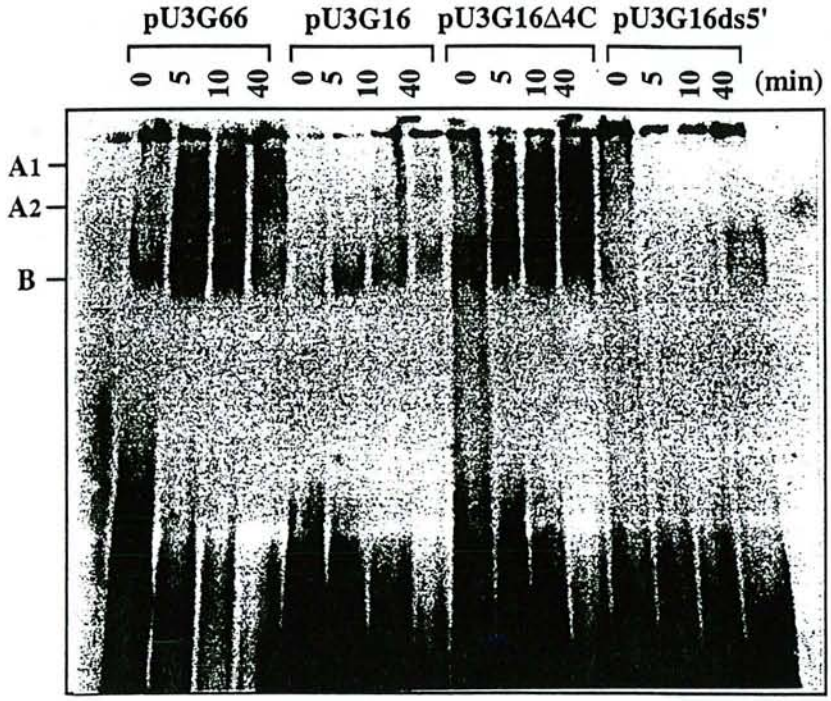

pU3G66
pU3G16
pU3G16Δ4C
pU3G16ds5'
0 5 10 40
0 5 10 40
0 5 10 40
0 5 10 40
(min)
A1
A2
B

Figure 2

Publication n:°2 :

Conservation de la structure secondaire des introns du petit ARN nucléolaire U3 chez les *Saccharomyces*

F. Brulé, A. Grégoire, V. Ségault, A. Mougin et C. Branlant

Comptes Rendus de l'Académie des Sciences de Paris (1995) vol. 318, 1197-1206

C. R. Acad. Sci. Paris, Sciences de la vie/Life sciences, 1995; 318: 1197-206
Biologie et génétique moléculaires/*Molecular biology and genetics*

Secondary structure conservation of the U3 small nucleolar RNA introns in *Saccharomyces*

Conservation de la structure secondaire des introns du petit ARN nucléolaire U3 chez les Saccharomyces

FABIENNE BRULÉ, ANNE GRÉGOIRE, VÉRONIQUE SÉGAULT, ANNIE MOUGIN, CHRISTIANE BRANLANT

Laboratoire d'enzymologie et de génie génétique, URA CNRS 457, Faculté des sciences, Université de Nancy I, boulevard des Aiguillettes, BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France.

RÉSUMÉ

En utilisant une stratégie basée sur l'amplification d'ADN génomique par la méthode PCR, nous avons montré que la présence, dans les gènes U3A et U3B codant pour le petit ARN nucléolaire U3, d'introns ayant les caractéristiques des introns éliminés dans un « épisséosome » n'est pas limitée à la levure *Saccharomyces cerevisiae*. Il s'agit, sans doute, d'une propriété ancienne des levures du genre *Saccharomyces*. Nous avons détecté les gènes U3A et U3B chez *Saccharomyces bayanus* et dans une souche de levure de brasserie et la séquence des introns U3A et U3B est très conservée. La levure industrielle renferme, en plus, 2 gènes supplémentaires de type U3B, dont les séquences introniques présentent de nombreuses mutations. Ces mutations préservent, néanmoins, la structuration secondaire de l'intron, en particulier la tige/boucle centrale. D'où l'idée d'un rôle important de cette structure tige/boucle centrale dans la formation de l'« épisséosome » et l'efficacité d'épissage. ▲

Mots clés : levure, Saccharomyces, *petit ARN nucléolaire U3, intron, épissage.*

ABSTRACT

Using a strategy based on PCR amplification of DNA and sequence analysis, we showed that the presence of introns with the characteristic features of introns spliced in a spliceosome, in the U3A and U3B snoRNA genes that code for the U3 small nucleolar RNA, is not a property restricted to Saccharomyces cerevisiae. *It is probably an ancient property of yeasts from the genus* Saccharomyces. *We detected the U3A and U3B snoRNA genes in* Saccharomyces bayanus *and in a lager brewing yeast strain. The U3A and U3B intronic sequences are highly conserved. Two additional "U3B-like" snoRNA genes were detected in the lager brewing yeast. Their intronic sequences show several differences, when compared to the U3B intronic sequence. However, despite the numerous mutations, the intron secondary structure is conserved, especially, the central structure. This strongly suggests an important role of this central stem/loop structure for spliceosome assembly and efficient splicing.* ▲

Key words: yeast, *Saccharomyces*, small nucleolar RNA U3, intron, splicing.

VERSION ABRÉGÉE

Nous avions fait, il y a quelques années, l'observation étonnante de la présence d'un intron, présentant les séquences caractéristiques des introns épissés dans un « épisséosome », dans chacun des 2 gènes (U3A et U3B) qui codent pour le petit ARN nucléolaire U3 de la levure *Saccharomyces cerevisiae*. Jusqu'à cette découverte et celle concomitante de la présence d'un intron du même type dans le gène du petit ARN nucléolaire U6, les introns épissés dans un « épisséosome » avaient toujours été considérés comme spécifiques des gènes nucléaires codant pour des protéines. Étant donné, qu'aucun intron n'a été décelé dans les gènes codant pour le petit ARN nucléolaire U3 d'autres espèces, il était intéressant de savoir si la propriété observée était générale dans le genre *Saccharomyces* et, dans le cas d'une réponse positive, d'étudier l'évolution des séquences introniques au sein de ce genre.

Note présentée par François Gros.
Note remise le 7 septembre 1995, acceptée après révision le 6 novembre 1995.

Corresponding author: C. Branlant.

Pour cela, nous avons défini une stratégie basée sur l'amplification d'ADN génomique par la méthode PCR. Les amplifications portaient soit sur l'ensemble du gène U3, soit uniquement sur la séquence intronique. Nous avons réalisé ces expériences d'amplification sur l'ADN génomique des souches: *S. cerevisiae* FL100, *S. cerevisiae* GY5, *Saccharomyces bayanus* STV107, et d'une souche de levure utilisée en brasserie. L'amplification des gènes entiers démontre la présence, dans toutes les espèces étudiées, de gènes ayant des tailles identiques à celles des gènes U3A et U3B de *S. cerevisiae*. Ce résultat est confirmé par l'amplification de fragments d'ADN correspondant aux introns U3A et U3B dans les 3 espèces.
Les produits d'amplification correspondant aux introns ont été clonés dans le phage M13mp9. Pour chaque fragment, un grand nombre de phages recombinants ont été séquencés, ce qui a permis de mettre en évidence chez *S. cerevisiae* et *S. bayanus* 2 séquences alléliques de l'intron U3A (U3A et U3A*), qui diffèrent par 3 mutations ponctuelles. Trois introns de type U3B (U3B, U3C et U3D) ont été mis en évidence dans la souche de brasserie.
Nous avons examiné quel était l'effet des différences de séquences entre les introns U3A* et U3A, d'une part, et U3C, U3D et U3B, d'autre part, sur la structure secondaire des pré-ARN U3A et U3B. Les différences observées ne modifient pas la structure secondaire des introns, en particulier la structure tige/boucle centrale. La conservation de cette structure suggère fortement qu'elle joue un rôle important dans le processus de maturation de l'intron, vraisemblablement au niveau de l'étape d'assemblage de l'« épisséosome ». ▲

For a long time after their discovery, the introns processed in a spliceosome were considered as a specificity of the nuclear genes that code for proteins. This assumption turned to be wrong upon finding of introns with the characteristic features of introns spliced in a spliceosome, in the gene coding for the U6 small nuclear RNA (U6 snRNA) of *Schizosaccharomyces pombe* [1] and in the 2 genes coding for the U3 small nucleolar RNA (U3 snoRNA) of *Saccharomyces cerevisiae* [2]. More recently, similar introns were detected in the U6 snRNA genes of other yeast species, 5 other *Schizosaccharomyces* species, *Rhodotorula hasegawae* and *Rhodosporidium dacryoidum* [3] and in the U2 snRNA gene of the yeast *Rhodotorula hasegawae* [4]. The corresponding genes from vertebrates, insects and plants have no intron [5]. Thus, the presence of introns spliced in a spliceosome in genes coding for metabolically stable RNAs seems to be restricted to the yeasts.
In addition, such introns were, up to now, exclusively found in RNA genes that code for small nuclear RNAs, either the UsnRNAs involved in splicing, or the UsnoRNAs involved in pre-ribosomal RNA maturation. For a given UsnRNA gene, the presence of an intron is not a generality in yeasts. For instance, whereas, the *S. pombe* U6 snRNA gene contains an intron [1], its *S. cerevisiae* counterpart has no intron [6]. The 2 *S. cerevisiae* U3 snoRNA genes both contain an intron [2], in contrast, the unique *S. pombe* U3 gene has no intron [7]. Finally, when introns are found in the same UsnRNA gene of 2 different yeast species, they are often located at different positions of the coding sequence. This is the case for the introns found in the U6 snRNA genes. A single intron was found in 6 different *Schizosaccharomyces* species and is located at the same position. However, the intron found in the *R. dacryoidum* gene is located at a different position and the 4 introns that were detected in the *R. hasegawae* gene [3], are located at 4 other positions. Hence, the presence of introns in UsnRNA genes and their location seems to be species or genus specific. Altogether these observations are consistent with the hypothesis that such introns were generated, recently, by accidental insertion in the UsnRNA genes of a given species or the ancestor of a given genus. The alternative hypothesis of an ancestral origin of these introns and their disappearance in higher eukaryotes, fits less with the observed sporadic distribution of the UsnRNA introns in yeasts. To get more information into this problem, it will be necessary to perform systematic study of UsnRNA intron evolution in the yeasts.
The 2 introns that we previously discovered in the *S. cerevisiae* genes coding for the U3A and U3B snoRNAs, respectively, are located at the same position of the coding sequence. But, they display strong length and sequence differences [2]. It was interesting to know whether introns were also present in the U3 snoRNA genes of other yeast species from the genus *Saccharomyces* and, if so, to determine their nucleotide sequences.
In this paper, we describe a PCR DNA amplification method for the search of introns in U3 snoRNA yeast genes with its application to the *S. cerevisiae* FL100 [8] and GY5 [9] strains, the *S. bayanus* STV107 strain and a lager brewing yeast strain. U3A and U3B snoRNA genes with introns were detected in these 3 yeast strains and the intronic nucleotide sequences were determined. In addition to the introns contained in the U3A and U3B snoRNA genes, 2 additional "U3B-like" introns were found in the lager brewing yeast strain. The results obtained are discussed in terms of intron evolution, taking into account, knowledge on *Saccharomyces* strains polyploidy and our recent data on the U3A and U3B pre-snoRNA secondary structure [10].

Materials and methods

Strains and culture conditions

The yeast strains used in this study were: *Saccharomyces cerevisiae* FL100 (ATCC 28383) [8] and GY5 (ATCC 25657) [9], *Saccharomyces bayanus* (STV107), that belongs to the french collection *Sciences et techniques du vin* and a lager brewing yeast strain considered as a *Saccharomyces uvarum* derivative [11]. This strain was a generous gift of Dr M. Moll (Tépral, Champigneulles, France). For genomic DNA preparation, the 4 strains were grown for 12 h in 10 ml of standard YPD medium (1% yeast extract, 1% bactopeptone and 2% glucose), at 30° C.
The *Escherichia coli* strain TG1 [12] was used as the host strain for production of recombinant DNA. *E. coli* TG1 was grown at 37° C, in the Luria Broth medium [13].

Preparation of yeast genomic DNA

The method used is an adaptation of that described by Cryer *et al.* [14]. A 10 ml stationary culture was centrifuged for 3 min at 13,000 rpm. The cells were ground with 0.45-0.5 mm diameter glass beads in the presence of 0.4 ml of Triton Buffer (1% Triton X-100, 1% sodium dodecyl sulfate (SDS), 100 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl pH 8.0) and 0.4 ml of a phenol/chloroform/isoamylic alcohol mixture (25/24/1). After alcohol precipitation, the DNA was incubated for 30 min at 37° C, in the presence of RNase A (Boehringer) at a 20 µg/ml concentration, in TE buffer (0.1 mM EDTA, 10 mM Tris-HCl pH 8.0). Another extraction with the phenol/chloroform/isoamyl alcohol mixture was performed. The genomic DNA was ethanol precipitated, dissolved in TE buffer and stored at 4° C.

Polymerase chain reaction (PCR)

An oligonucleotide pair 1 was designed to amplify the intronic sequence of the U3 snoRNA genes. It consists of the primer A: 5' AGCTCgaattcTACTTCAGTATGT 3', that contains the nucleotide sequence of the *S. cerevisiae* U3A snoRNA from positions 8 to 14 (exon 1)(underlined), followed by the hexanucleotide sequence at the 5' end of the intron (underlined) [2], and a sequence that generates an *Eco*R I restriction site (lower cases), and the primer B: 5' ATAAggatccTATAGAAATGATCCTA 3', complementary to the phylogenetically conserved Box A of the U3A snoRNA (underlined) [2] and contains a sequence that generates a *Bam*H I restriction site (lower cases). The oligonucleotide pair 2 was designed to amplify the entire U3 pre-snoRNA coding sequence. It consists of the primer C: 5' AGCTCGTCGACGTACTTCA 3', that displays the sequence at the 5' end of the *S. cerevisiae* U3A snoRNA from position 1 to 14 [2] and the primer D: 5' ATAAGTTTTAACTTGTCAGACTGCC 3', complementary to the 3' terminal sequence of the *S. cerevisiae* U3A snoRNA from position 475 to 490 [2].

Polymerase chain reactions were performed on 1 µg of yeast genomic DNA or on 2 ng of plasmid DNA, in 50 µl of 50 mM KCl, 2 mM $MgCl_2$, 0.2% Tween-1 (Sigma), 20 mM Tris-HCl pH 8.3, containing 0.2 mM of each deoxynucleotide triphosphate, 10 pmol of each primer and 2 units of Taq DNA polymerase (Gibco-BRL). A Perkin Elmer Cetus thermal cycler was used with the following protocol: 30 cycles of denaturation at 94° C for 1 min, annealing at 55° C for 1 min and polymerization at 72° C for 2 min, followed by a 10 min at 72° C after the final cycle. PCR products were analyzed by electrophoresis on a 6% non-denaturing acrylamide gel in TBE buffer [15].

Plasmids and recombinant DNA technology

DNA fragments, corresponding to the amplified U3 gene intronic sequence, were gel purified as described above, and eluted from the gel by diffusion in the elution buffer [15]. After phenol extraction and ethanol precipitation, they were digested with the *Eco*R I and *Bam*H I nucleases and ligated into the M13mp9 phage DNA [16], cleaved with the same enzymes. The *E. coli* TG1 strain was transformed with the ligation mixture [15]. Several recombinant M13mp9 single-stranded DNAs were prepared for each ligation. Inserts were sequenced as described by Tabor and Richardson [17].

Nucleotide sequence alignments and phylogenetic tree

Sequence alignments were performed with the PILEUP program of GCG software Version 8.1 Unix (1995). Shematic representation of sequence similarities were made with the Alscript program [18]. The phylogenetic tree was established with the "FastDNAm1" program [19, 20]. The GenBank database accession numbers of the sequences established in this paper are as follows: *S. bayanus* STV107 U3A intron (X81559), U3A* intron (X81558), U3B intron (X81557); *S. uvarum* lager brewing yeast U3A intron (X91039), U3B intron (X91038), U3C intron (X91035) and U3D intron (X91036).

Results

The strategy used

As, in spite of their length and sequence differences [2], the introns of the U3A and U3B snoRNA genes of *S. cerevisiae* are located at identical positions in the coding sequence, we hypothesized that introns in the U3 snoRNA genes of other yeast species from the genus *Saccharomyces* may also be located at the same position. The introns of the *S. cerevisiae* U3A and U3B snoRNA genes are inserted between 2 phylogenetically highly conserved sequences of the U3 snoRNA, namely, Boxes A' and A [2, 21]. We took advantage of this location to look for introns in other yeast species with the polymerase chain reaction (PCR) method for DNA amplification. An oligonucleotide pair 1 (primers A and B) was designed for this purpose. Primer A contains the sequence of Box A', followed by the 6 nucleotide sequence that is conserved at the 5' end of the yeast introns, which are spliced in a spliceosome [2]. Primer B is complementary to the Box A sequence. Primers A and B, also contained sequences that generated an *Eco*R I and a *Bam*H I restriction site, respectively. These 2 sites were used to clone the amplified DNA fragment in the M13mp9 phage DNA, in order to perform sequence analysis. We could not discard the possibility of sequence heterogeneity of the amplified DNA fragments that would result from the presence of introns with equal lengths, but different sequences, in a given yeast species. To test for this possibility, the replicative forms of 10 to 30 distinct M13mpl9 recombinant phages were prepared and sequenced for each DNA ligation experiment.

As a control that there was no gene without intron or gene with an intron located at a different position, DNA amplification experiments were also performed with the oligonucleotide pair 2 (primers C and D). Primer C contains the sequence of the exon 1 of the U3A and U3B snoRNA genes, but, no intronic sequence. Primer D is complementary to the highly phylogenetically conserved Box D at the 3' end of the U3 snoRNA [9].

Introns A and B are detected in all the strains studied

PCR amplification experiments were performed with the genomic DNA from the *S. cerevisiae* FL100 and GY5 strains, the *S. bayanus* STV107 strain and a lager brewing

yeast strain, that was previously identified as a *S. uvarum* derivative [11]. As a control, amplification was also made with the pFL1::snR17A plasmid containing a cloned *S. cerevisiae* U3A gene [2]. The amplified DNA fragments were fractionated by electrophoresis on a 10% polyacrylamide gel in parallel with DNA size markers. *Figure 1* illustrates the results obtained for *S. cerevisiae* FL100, *S. bayanus* STV107 and the lager brewing strain. As expected, amplified DNA fragments of 202 and 175 bp, corresponding to the U3A and U3B introns, respectively, were obtained for *S. cerevisiae* FL100. Amplified fragments with similar sizes were also obtained for *S. bayanus (Fig. 1)* and *S. cerevisiae* GY5 (not shown). For the lager brewing yeast, in addition to the DNA fragments corresponding to the U3A and U3B introns, a DNA fragment (C) with an electrophoretic mobility slightly inferior to that of the U3B intronic fragment was detected. Bands corresponding to DNA of higher molecular weights were also present in low yields. As this yield varied from one experiment to the other, they were considered to be artefacts of the amplification procedure.

These amplification experiments demonstrated the presence of introns in the U3 snoRNA genes of all the studied species. In addition, the introns found had sizes similar to those present in the *S. cerevisiae* FL100 U3A and U3B snoRNA genes, except for the additional band in the lager brewing yeast. Sequence analysis was required to determine whether the U3A and U3B introns had conserved sequences in all the studied species and to identify the additional fragment of the lager brewing yeast.

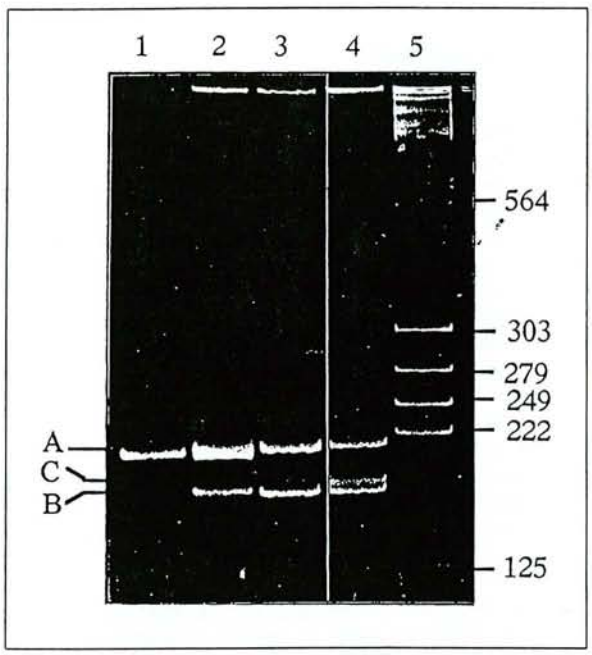


Figure 1. ***PCR amplification of the intronic DNA from the yeast U3 snoRNA genes.*** *Genomic DNAs from* S. cerevisiae *FL100 (lane 2),* S. bayanus *(lane 3) and the lager brewing yeast (lane 4) were prepared as described in* Materials and methods. *Amplifications were performed with the oligonucleotide pair 1, under the conditions described in* Materials and methods. *As a control, DNA amplification was also performed with 2 ng of the pFL1::snR17A [2] plasmid DNA (lane 1). A mixture of the pBR322* Ava *II digestion products and the λ phage DNA* Hind *III digestion products was used as size marker (lane 5). The sizes of the resulting DNA fragments are indicated on the right side. The amplified DNA fragments A and B correspond to the U3A and U3B introns, respectively. An additional amplified DNA fragment, denoted C is detected in the lager brewing yeast. The lengths of the fragments A and B correspond to the expected sizes 202 and 175 nt, respectively, as calculated from the 564 and 125 nt fragments of the λ digest. The pBR322* AvaII *digestion products revealed to be bad size markers. Their mobility is probably altered by sequence dependent DNA distorsions.*

High conservation of the intron A sequence in *Saccharomyces*

The number of M13mp9 recombinant phages that were subjected to sequence analysis, for each of the amplified DNA fragments is indicated in *Table I.* As expected, the M13mp9 recombinant phages generated by insertion of the amplified DNA fragment, that corresponded to the *S. cerevisiae* FL100 U3A intron, were found to have homogeneous sequences *(Table I)*, and this sequence was identical to that previously found in the cloned *S. cerevisiae* FL100 U3A snoRNA gene [2]. The M13mp9 recombinant clones corresponding to the *S. cerevisiae* GY5 U3A intron were also homogeneous *(Table I)*. However, 3 base substitutions were observed, as compared to the *S. cerevisiae* FL100 *(Fig. 2)*. This result is in accord with previous sequencing data on the *S. cerevisiae* GY5 [9] and FL100 [2] U3A snoRNA genes. Indeed, the gene sequences previously established [2, 9] were found to display the 3 differences, presently, observed in the intron. One additional difference was detected in the coding region. Thus, our results confirm that these differences are true strain differences and the sequence homogeneity observed for each species,

Table I
Compilation of the sequence data on the amplified DNA fragments corresponding to the U3A introns

Yeast strain	Total number of sequenced clones	Intron type	Number of clones
S. cerevisiae FL100	11	A	11
		A*	0
S. cerevisiae GY5	10	A	0
		A*	10
S. bayanus STV107	21	A	5
		A*	16
Lager brewing yeast	14	A	14
		A*	0

The amplified DNA fragments were ligated into M13mp9 phage DNA. For each strain, the total number of recombinant phages subjected to sequence analysis are indicated, as well as, the number of clones that display the U3A and the U3A* sequence, respectively.

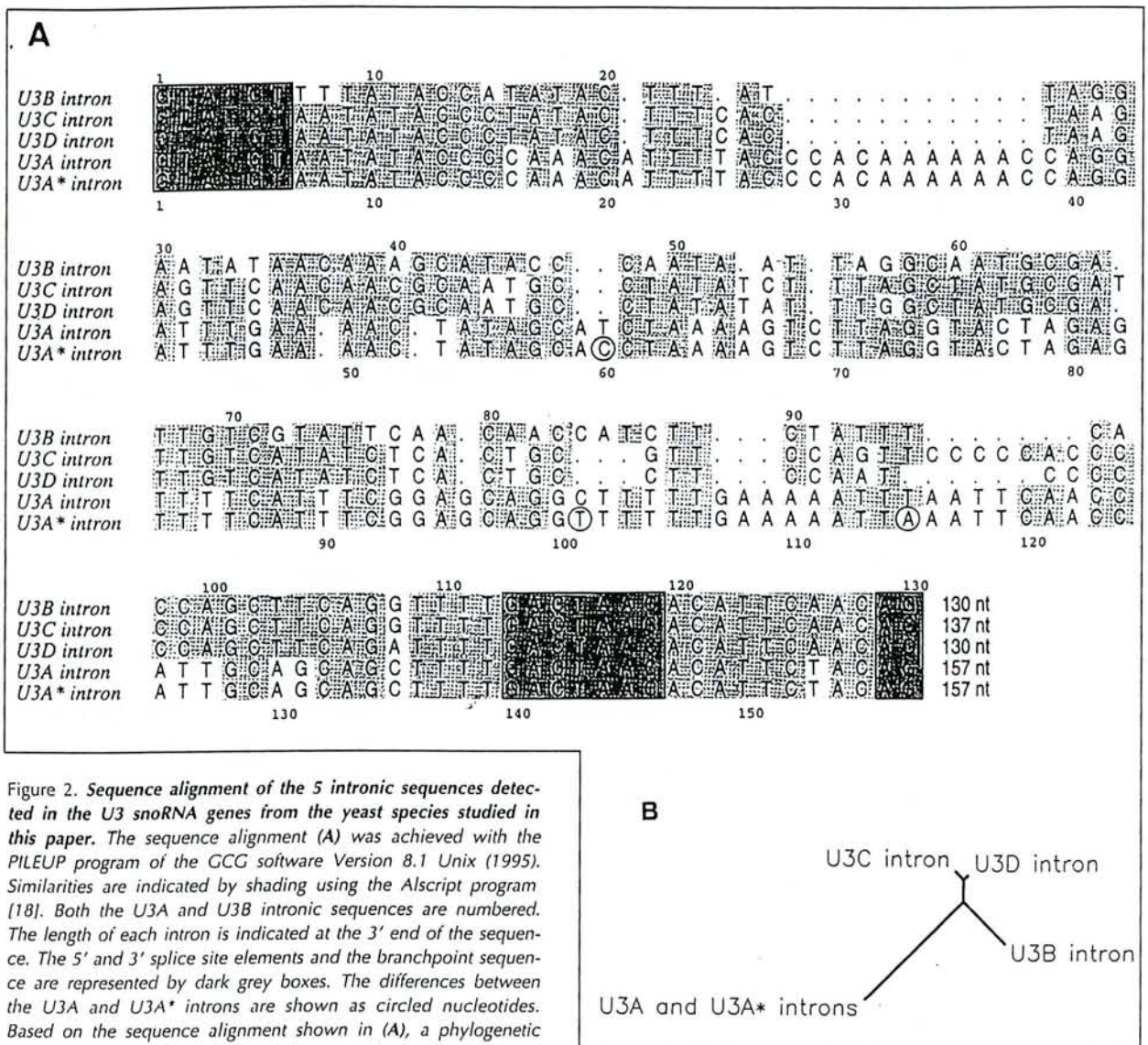


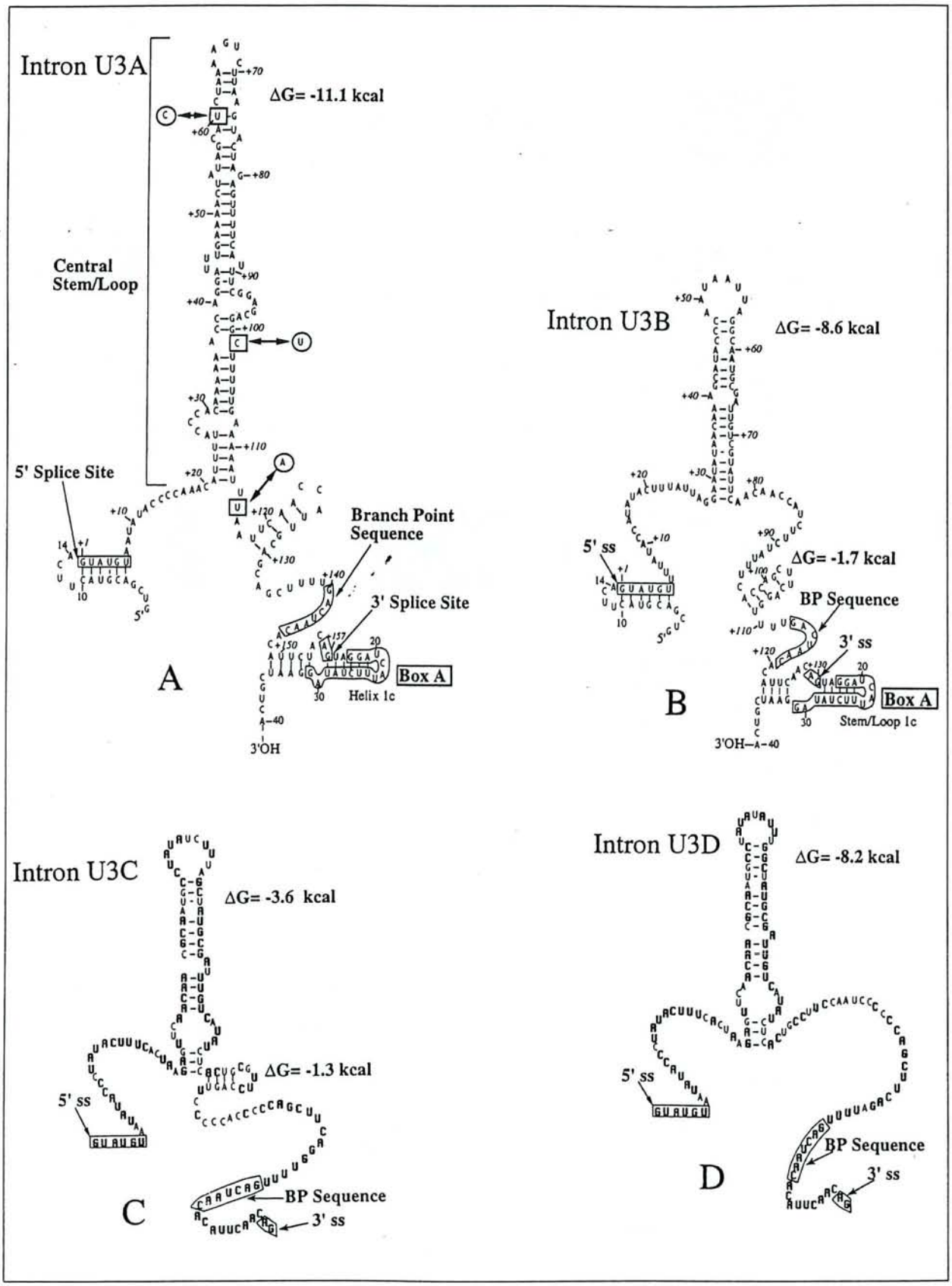
Figure 2. ***Sequence alignment of the 5 intronic sequences detected in the U3 snoRNA genes from the yeast species studied in this paper.*** *The sequence alignment* ***(A)*** *was achieved with the PILEUP program of the GCG software Version 8.1 Unix (1995). Similarities are indicated by shading using the Alscript program [18]. Both the U3A and U3B intronic sequences are numbered. The length of each intron is indicated at the 3' end of the sequence. The 5' and 3' splice site elements and the branchpoint sequence are represented by dark grey boxes. The differences between the U3A and U3A* introns are shown as circled nucleotides. Based on the sequence alignment shown in* ***(A)****, a phylogenetic tree* ***(B)*** *was established for the 5 introns with the "FastDNAm1" program [19, 20].*

shows that the 2 allelic U3A snoRNA genes in *S cerevisiae* FL100 have the U3A sequence, whereas, those in *S. cerevisiae* GY5 have the sequence, that we will now denoted U3A*. In contrast, the U3A and U3A* intronic sequences were both found upon sequencing the recombinant phage DNAs that corresponded to the amplified U3A DNA fragment from the *S. bayanus* STV107 U3A intron *(Table I)*. This suggests the presence of 2 distinct U3A snoRNA allelic genes in this strain. In the lager brewing strain, only the U3A sequence was detected.

We have recently established the secondary structure of the *S. cerevisiae* FL100 U3A pre-snoRNA using chemical and enzymatic probes and primer extension analysis [10], as previously described [22]. As can be seen in *Figure 3*, 2 of the 3 differences between the U3A and U3A* intronic sequences are located in the long central stem/loop structure, the third one is located in a single-stranded segment. The 2 differences in the stem are such that base-complementarity is conserved and even reinforced in the U3A* intron: a U·G pair in U3A is replaced by a C-G pair in U3A* and an A C pair in U3A is changed for an A-U pair in U3A*.

Two additional "U3B-like" introns in the lager brewing yeast

The amplified DNA fragment of the lager brewing yeast, that was expected to correspond to the U3B intron, contained 2 distinct amplified intronic sequences: a U3B intronic sequence identical to that of *S. cerevisiae* and another intronic sequence, that was denoted intron D *(Fig. 2A)*. The sequence established for the additional amplified DNA


Intron U3A
ΔG= -11.1 kcal
Central
Stem/Loop
5' Splice Site
Branch Point
Sequence
3' Splice Site
Box A
Helix 1c
3'OH
A
Intron U3B
ΔG= -8.6 kcal
5' ss
ΔG= -1.7 kcal
BP Sequence
3' ss
Box A
Stem/Loop 1c
3'OH
B
Intron U3C
ΔG= -3.6 kcal
ΔG= -1.3 kcal
5' ss
BP Sequence
3' ss
C
Intron U3D
ΔG= -8.2 kcal
5' ss
BP Sequence
3' ss
D

fragment C of this strain *(Fig. 1)*, also displayed the characteristic features of introns spliced in a spliceosome (presence of 5' and 3' splice site elements and branchpoint sequence, see *Fig. 2A*). The corresponding intron was denoted intron C. The 4 distinct intronic sequences found in the lager brewing yeast are aligned in *Figure 2A*. The sequences of their 5' and 3' splice site elements and branchpoint regions are conserved. This means that as found for the *S. cerevisiae* U3A and U3B introns, the U3C and U3D introns have a GACTAAC branchpoint sequence, instead of the TACTAAC sequence, found in the introns of the yeast protein genes that are spliced in a spliceosome [23].

The U3C and U3D introns are highly homologous (93% of similarity) and their sequences are much more similar to that of the U3B intron (73.5% of similarity) than to that of the U3A intron (61.5% of similarity). Only the 15 nucleotides at the 5' end of U3C and U3D introns are more similar to the U3A intron sequence, than to the U3B intron sequence. Except for this 5' end, that may have been gained by a DNA recombination event between intronic DNAs, the U3B, U3C and U3D introns were clearly derived from a common ancestor. In the filiation, U3C and U3D introns probably appeared more recently *(Fig. 2B)*. The actual divergence between the U3B introns and the U3C and U3D introns, probably, arose from punctual substitutions, short deletions, insertions and short duplications. For instance, the additional CCCCA sequence in the U3C intron, as compared to the U3D intron, corresponds to a tandemly duplicated sequence.

We recently made an experimental determination of the U3B pre-snoRNA secondary structure using chemical and enzymatic probes for RNA secondary structure determination [10]. As shown in *Figure 3*, differences between the U3B, U3C and U3D introns mainly concern the central stem/loop structure and the segment that joins this structure to the branchpoint sequence. Base-compensations between mutations preserved the possibility to form the central stem/loop structure in the 3 introns. As compared to the U3B intron, the stability of the upper part of the stem of the U3D intron is reinforced by the presence of an additional G-C pair. However, the stability of the lower part of the stem is slightly decreased by the replacement of successive A-U and U-A pairs, by a G-C pair. Altogether, these differences compensate each other and the overall stability of the central stem/loop structure of the U3B and U3D introns are very similar, as calculated with the "MFold" program of the GCG software *(Fig. 3)*. The stability of the U3C intron is inferior due to a G to A substitution at position 57, as referred to U3B intron numbering *(Fig. 3C)*. As the result of a A to C substitution at position 97, as referred to U3B intron numbering, the short stem/loop structure upstream of the branch point sequence in the U3B intron, cannot be formed in the U3C and U3D introns. However, the sequence of the U3C intron allows the formation of another short stem/loop structure, immediately downstream of the central stem/loop structure *(Fig. 3C)*. These secondary structure considerations clearly demonstrated that the U3C and U3D introns can be considered as U3B intron derivatives and, thus, that they probably belong to "U3B-like" snoRNA genes.

Only U3A, U3B and "U3B-like" snoRNA genes are present in the *Saccharomyces* strain tested

Thus, altogether, our results so far strongly suggested the presence of U3A and U3B snoRNA genes in the various strains studied, plus additional "U3B-like" snoRNA genes in the lager brewing yeast. This had to be confirmed by amplification of the entire U3 snoRNA genes. To test for this possibility DNA amplifications were made with the oligonucleotide pair 2, as primers.

DNA amplifications were achieved on genomic DNAs from *S. cerevisiae* FL100, *S. bayanus* STV107 and the lager brewing yeast strain. The amplified DNA fragments were fractionated by electrophoresis on a 6% polyacrylamide gel next to DNA size markers *(Fig. 4)*. For the 3 strains, only 2 bands with the mobility expected for the U3A and U3B amplified genes were obtained. This excluded the possibility of genes without intron or with introns of different sizes. The size differences between the U3B snoRNA gene and the 2 expected U3C and U3D snoRNA genes were not enough to get individual bands. Thus, the finding of only 2 bands for the lager yeast strain is consistent with our proposal of U3B and "U3B-like" snoRNA genes in this strain.

Discussion

In contrast to the haploid strains of the *Kluyveromyces* and *Pichia* genera, yeasts from *Saccharomyces* genera are diploid or have higher degree of polyploidy [24]. In particular, the yeast lager brewing strains usually display a high degree of polyploidy [24]. A large number of yeast lager brewing strains are closely related to the "Carlsberg lager strain #1" isolated by Hansen in 1883 [25]. Although classified as *S. cerevisiae*, these strains differ in some ways

Figure 3. ***Secondary structures models for the U3A, U3B, U3C and U3D introns.*** *The secondary structures established for the 5' halves of the* S. cerevisiae *FL100 pre-U3A and pre-U3B snoRNAs [10] are shown in* ***(A and B)****, respectively. The 5' and 3' splice site (5'ss and 3'ss) elements and branchpoint (BP) sequence of the intron, as well as, the 2 phylogenetically conserved Boxes A' and A of U3 snoRNA are indicated [2], the stem/loop structure nomenclature is that previously used [10, 22]. Nucleotide numbering is according to Grégoire* et al. *[10] (from 1 to 14 for the first exon, +1 to +157* ***(A)*** *or +1 to +130* ***(B)*** *for the introns, from 15 to 40 for the 5' end of exon 2). In* ***(A)****, the differences found in the pre-U3A* snoRNA from* S. cerevisiae *GY5 (this paper and [9]) are indicated by circled letters. Models of secondary structure were built for the U3C* ***(C)*** *and U3D* ***(D)*** *introns. Nucleotide sequence conservation in the U3C and U3D introns, as compared to the U3B intron, are indicated by bold letters. Free energies (indicated by ΔG) of the intron central and lateral stem/loop structures have been calculated with the "Mfold" program of the GCG software Version 8.1 Unix (1995). The folding parameters used in this program are described in [47, 48].*

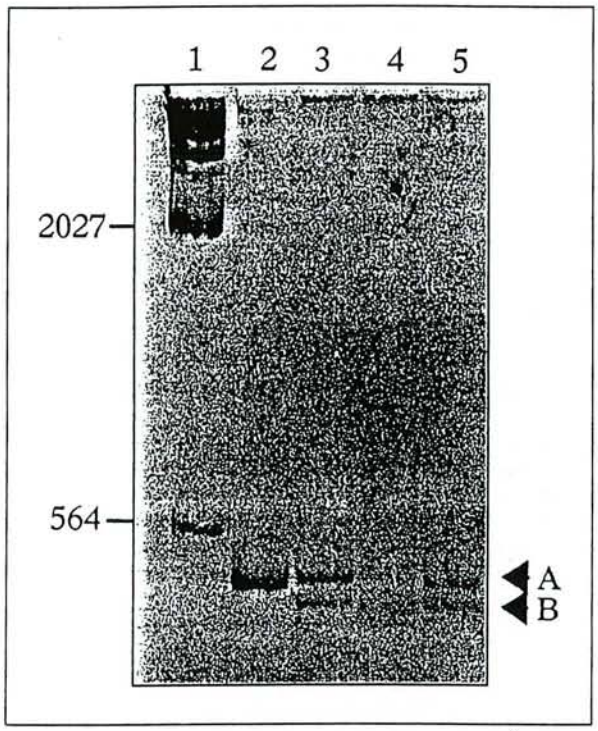


Figure 4. ***Polyacrylamide gel electrophoresis of the amplified U3 snoRNA gene DNA.*** *Genomic DNAs from* S. cerevisiae *FL100 (lane 3),* S. bayanus *STV107 (lane 4) and* S. uvarum *(lane 5) were subjected to PCR DNA amplification using the oligonucleotide pair 2. As a control, amplification was also done with the pFL1::snR17A plasmid (lane 2). The amplified DNA fragments were fractionated by electrophoresis on a 6% polyacrylamide gel. The λ phage DNA* Hind *III digestion products were used as size markers. The lengths of the fragments are indicated on the left side. The bands corresponding to the U3A and U3B genes are indicated by a triangle.*

from *S. cerevisiae* and its was show that *Saccharomyces carlsbergensis* is alloploid, containing parts of at least 2 diverged genomes, one related to *S. cerevisiae*, the other one that may be related to *Saccharomyces monacensis*. Based on its physiological and metabolic properties, the industrial strain used in this study was previously identified as a *S. uvarum* derivative and was shown to be tetraploid [24]. The diploidy or tetraploidy of the studied strains, probably, explained some of our observations. For instance, as liquid pre-cultures of *S. bayanus* STV107 were inoculated with a single colony isolated on an agar plates, we can exclude the possibility, that our observation of 2 distinct U3A and U3A* intronic sequences in *S. bayanus* STV107, resulted from the coexistence of 2 variant strains in the yeast cells that were used for genomic DNA preparation. The presence of these 2 sequences should correspond either to the existence of 2 different U3A and U3A* snoRNA genes in *S. bayanus* STV107, or more likely, to 2 different alleles of the same gene. The *S. cerevisiae* FL100 and GY5 strains that display the U3A and U3A* intronic sequence, respectively, are laboratory homozygote diploid strains. The *S. cerevisiae* wild-type heterozygote strains used to generate these 2 homozygote strains, probably, had like *S. bayanus* STV107, 2 different allelic U3A snoRNA genes (U3A and U3A*). Surprisingly, the lager brewing yeast strain studied, which has a complex U3B snoRNA genes content, displayed a unique U3A intronic sequence. Similarly, as liquid cultures of this strain were also made with a single colony isolated on an agar plate, the presence of U3B and "U3B-like" snoRNA genes in our analysis cannot be attributed to the coexistence of different variant strains in the cultures. One likely explanation is that this tetraploid strain is alloploid like *S. carslbergensis* and one genome has 2 allelic U3A intronic sequences and 2 U3B intronic sequences, whereas, the other genome has 2 allelic U3A intronic sequences and 2 different "U3B-like" intronic sequences (U3C and U3D introns). Study of U3 intronic sequences in other *Saccharomyces* species may help to identify the origin of the different genomes fused in this industrial strain, especially, to verify that one of them corresponds to *S. uvarum*, in accord with physiologic and metabolic data. Another alternative to identify these genomes, would be to determine the nucleotide sequence of other genes from this strain, for example, the *Met2* gene alleles, as Hansen and Kielland-Brandt [26] did for *S. carlsbergensis*. Interestingly, whereas, the *S. bayanus* sequence coding for the MET2 protein displays only 82.1% of similarity with the corresponding sequence of *S. cerevisiae* [27], the U3A and U3B introns of *S. cerevisiae* have the same sequence as their counterparts in *S. bayanus*. This suggests a high degree of nucleotide conservation of the U3A and U3B intronic sequences, as compared to other *Saccharomyces* genomic sequences. This is in contrast to the strong divergence observed between the U3A and U3B intronic and coding sequences and this suggests that the U3A and U3B snoRNA genes were generated by a DNA duplication event, in a very ancient ancestor of the actual *Saccharomyces*. Hence, the presence of introns in the U3 snoRNA genes is probably an extremely ancient property of the genus *Saccharomyces*.

Interestingly, the U3B snoRNA gene seems to be highly conserved in *Saccharomyces*, but, its expression is undetectable, when *S. cerevisiae* FL100 or GY5 is grown in laboratory conditions [9, 28]. Such a conservation probably reflects a functional role of this gene in the natural environment of wild-type *Saccharomyces*. Yet, the *S. cerevisiae* U3B snoRNA gene was found to be active, as it was expressed and allowed cell growth after deletion of the U3A snoRNA gene [28]. It can thus be assumed, that in spite of nucleotide sequence differences, the U3B pre-snoRNA should be spliced efficiently and the U3B snoRNA functional.

From a comparative study of the intron from yeast nuclear genes coding for proteins [29] and genetic experiments [30-32], it was suggested that efficient processing of the yeast introns spliced in a spliceosome requires the formation of a base-pair interaction between sequences located downstream from the 5' splice site element and upstream of the branch point sequence, respectively. Such base pairing was proposed to be necessary in long introns, in order to limit the distance between the 5' splice site, the branch-point sequence and the 3' splice site [29]. We recently provided experimental evidence for the presence of such base-pairing in both the U3A and U3B pre-snoRNAs [10]. The central part of the U3A and U3B introns, in spite of

their length and sequence differences, both form a long central stem/loop structure. Interestingly, the 3 differences observed between the U3A and U3A* introns preserve formation of this central stem/loop structure. Similarly, the numerous mutations found in the "U3B-like" introns, also preserve formation of the central stem/loop structure. Altogether, this strongly suggests an important role of this central structure in spliceosome assembly or splicing efficiency. This also suggests that the 2 "U3B-like" snoRNA genes in the lager brewing yeast may be functional.

In both "U3B-like" introns (U3C and U3D), the possibility to form the small stem/loop structure that is located upstream of the branch point sequence in the U3B intron, is impaired by sequence differences [10]. However, another small stem/loop structure can be formed immediately downstream from the central stem in the U3C intron, so that the linear distance between the 5' splice site element and the branch point sequence is well conserved (27 nt against 24 in U3B). In the U3D intron, no stem/loop structure can be formed, but, the distance between the central stem/loop and the branchpoint sequence is only of 32 nt *(Fig. 3D)*. Thus, formation of small stem/loop structures between the central stem/loop and the branchpoint sequence may modulate the distance between the 5' splice site element and the branchpoint sequence in yeast nuclear introns.

In spite of the conservation of the basic mechanism [33], the splicing machinery of yeast and vertebrates display some differences, mainly, the requirement of SR proteins for vertebrates introns [34] and higher constraints at the level of the primary and secondary structures in yeasts [29, 32]. Indeed, the nucleotide sequences of the 5' splice site element and branchpoint region are much more conserved in yeasts than in vertebrates [33, 35]. Up to now, no secondary structure requirement was described for constitutive splicing of vertebrates introns. Secondary structure elements were only found to be involved in splicing regulation or alternative splicing [32, 36]. These differences might be due to the different modes of recognition of the 5' and 3' splice sites in yeast and vertebrates [37]. Berget *et al.* [37, 38] proposed that the exons, which are short in vertebrates may represent the recognition element. In contrast, the recognition element would be the introns in yeasts and drosophila [39-41]. These organisms have short introns. In addition, the SR proteins may compensate for the absence of define secondary structure elements in the vertebrates introns, since, they were recently shown to be able to compensate for the absence of U1 snRNA [42]. A current hypothesis is that the ancestor of the introns spliced in a spliceosome was an autocatalytic group II intron [43, 44]. According to this hypothesis [45, 46], part of the RNA sequences contained in the intron would have been moved from *cis* to *trans*, thus, generating the U snRNA genes. The UsnRNA that act in *cis* would replace some of the intronic sequence acting in *trans* in autocatalytic introns. In addition, proteins would have been selected in order to reinforce the interaction between the UsnRNAs and the pre-mRNA. The advantage of such an evolution would be to limit the constraints on the pre-RNA primary and secondary structures, especially, at the level of introns. Based on this hypothesis, the yeast introns that are subjected to higher structural constraints, as shown by our results, would be more closely related to the ancestor autocatalytic intron, than the vertebrate introns. ▼

Acknowledgements: we are grateful to E. Petitdemange (Université de Nancy I, France), M. Moll (Tépral, Champigneulles, France) and J. Hughes (Liverpool, England) for providing us with the *S. bayanus* STV107 strain, the lager brewing strain and the *S. cerevisiae* GY5 strain, respectively. We also thanked M. Moll and F. Boutley (Tépral) for helpful discussion, G. Mulliert-Carlin for help in nucleotide sequence alignment and J. Bayeul for excellent technical assistance. F. Brulé, A. Grégoire and V. Ségault were fellows of the French ministère de la Recherche et de l'Enseignement supérieur. This work was supported by the Centre national de la recherche scientifique and the ministère de la Recherche et de l'Enseignement supérieur.

References

1. Tani T., Ohshima Y. 1989. The gene for the U6 small nuclear RNA in fission yeast has an intron. *Nature* 337: 87-90.

2. Myslinski E., Ségault V., Branlant C. 1990. An intron in the genes for U3 small nucleolar RNAs of the yeast *Saccharomyces cerevisiae*. *Science* 247: 1213-6.

3. Tani T., Ohshima Y. 1991. mRNA-type introns in U6 small nuclear RNA genes: implications for the catalysis in pre-mRNA splicing. *Genes Dev.* 5: 1022-31.

4. Takahashi Y., Urushiyama S., Tani T., Ohshima Y. 1993. An mRNA-type intron is present in the *Rhodotorula hasegawae* U2 small nuclear RNA gene. *Mol. Cell. Biol.* 13: 5613-9.

5. Solymosy F., Pollak T. 1993. Uridylate-rich small nuclear RNAs (U snRNAs), their genes and pseudogenes, and UsnRNPs in plants: structure and function. A comparative approach. *Critical Rev. Plant Sci.* 12: 275-369.

6. Brow D.A., Guthrie C. 1988. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* 334: 213-8.

7. Porter G.L., Brennwald P.J., Holm K.A., Wise J.A. 1988. The sequence of U3 from *Schizosaccharomyces pombe* suggests structural divergence of this snRNA between metazoans and unicellular eukaryotes. *Nucleic Acids Res.*16: 10131-52.

8. Lacroute F. 1968. Regulation of pyrimidine biosynthesis in *Saccharomyces cerevisiae*. *J. Bacteriol.* 95: 824-32.

9. Hughes J.M.X., Konings D.A., Cesareni G. 1987. The yeast homologue of U3 snRNA. *EMBO J.* 6: 2145-55.

10. Grégoire A., Mougin A., Ségault V., Fournier R., Brulé F., Branlant C. 1995. The secondary structure of the yeast *Saccharomyces cerevisiae* U3 pre-snoRNAs (in press).

11. Aigle M., Erbs D., Moll M. 1984. Some molecular structures in the genome of lager brewing yeasts. *Am. Soc. Brew. Chem.* 42: 1-7.

12. Gibson T.J. 1984. *Studies on the Epstein-Barr virus genome*. PhD Thesis. Cambridge University, UK.

13. Atlas R.M. 1993. *Handbook of microbiological media*. L.C.P. Ed. London: CRC Press.

14. Cryer D.R., Eccleshall R., Marmur J. 1975. *Methods in cell biology*. Prescott D.M., eds. New York: Academic Press, 39-44.

15. Sambrook J., Fritsch E.F., Maniatis T. 1989. *Molecular cloning. A laboratory manual.* New York: Cold Spring Harbor Laboratory Press.

16. Yanisch-Perron C., Vieira J., Messing J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-19.

17. Tabor S., Richardson C.C. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* 84: 4767-71.

18. Barton G.J. 1993. Alscript: a tool to format multiple sequence alignments. *Protein Eng.* 6: 37-40.

19. Olsen G.J., Matsuda H., Hagstrom R., Overbeek R. 1994. FastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput. Appl. Biosci.* 10: 41-8.

20. Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17: 368-76.

21. Wise J.A., Weiner A.M. 1980. *Dictyostelium* small nuclear RNA D2 is homologous to rat nucleolar RNA U3 and is encoded by a dispersed multigene family. *Cell* 22: 109-18.

22. Ségault V., Mougin A., Grégoire A., Banroques J., Branlant C. 1992. An experimental study of *Saccharomyces cerevisiae* U3 snRNA in solution. *Nucleic Acids Res.* 20: 3443-51.

23. Pikielny C.W., Teem J.L., Rosbash M. 1983. Evidence for the biochemical role of an internal sequence in yeast nuclear mRNA introns: implications for U1 RNA and metazoan mRNA splicing. *Cell* 34: 395-403.

24. Aigle M., Erbs D., Mol M. 1983. Determination of brewing yeast ploidy by DNA measurement. *J. Inst. Brew.* 89: 72-4.

25. Hansen E.C. 1883. Recherches sur la physiologie et la morphologie des ferments alcooliques. V. Méthodes pour obtenir des cultures pures de *Saccharomyces* et de micro-organismes analogues. *Compt. Rend. Trav. Lab. Carlsberg* 2: 92-105.

26. Hansen J., Kielland-Brandt M.C. 1994. *Saccharomyces carlsbergensis* contains two functional *Met2* alleles similar to homologues from *S. cerevisiae* and *S. uvarum*. *Gene* 140: 33-40.

27. Langin T., Faugeron G., Goyon C., Nicolas A., Rossignol J.L. 1986. The *Met2* gene of *Saccharomyces cerevisiae*: molecular cloning and nucleotide sequence. *Gene* 49: 283-93.

28. Hughes J.M.X., Ares M. 1991. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J.* 10: 4231-9.

29. Parker R., Patterson B. 1987. Architecture of fungal introns: implications for spliceosome assembly. In: Inouye M., Dubock B., eds. *Molecular biology of RNA. New perspectives.* New York, 133-48

30. Pikielny C.W., Rosbash M. 1985. mRNA splicing efficiency in yeast and the contribution of nonconserved sequences. *Cell* 41: 119-26.

31. Newman A. 1987. Specific accessory sequences in *Saccharomyces cerevisiae* introns control assembly of pre-mRNAs into spliceosomes. *EMBO J.* 6: 3833-9.

32. Goguel V., Rosbash M. 1993. Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell* 72: 893-901.

33. Moore M.J., Query C.C., Sharp P.A. 1993. Splicing of precursors to mesenger RNAs by the spliceosome. In: Gesteland R., Athuis J., eds. *The RNA world.* New York: Cold Spring Harbor Laboratory Press, 303-57.

34. Horowitz D., Krainer A.R. 1994. Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. *Trends Genet.* 10: 100-6.

35. Green M.R. 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu. Rev. Cell. Biol.* 7: 559-99.

36. Goguel V., Wang Y., Rosbash M. 1993. Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing. *Mol. Cell. Biol.* 13: 6841-8.

37. Berget S.M. 1995. Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270: 2411-4.

38. Robberson B.L., Cote G.J., Berget S.M. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* 10: 84-94.

39. Talerico M., Berget S.M. 1994. Intron definition in splicing of small *Drosophila* introns. *Mol. Cell. Biol.* 14: 3434-45.

40. Mount S.M., Burks C., Hertz G., Stormo G.D., White O., Fields C. 1992. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. *Nucleic Acids Res.* 20: 4255-62.

41. Zhang G., Taneja K.L., Singer R.H., Green M.R. 1994. Localization of pre-mRNA splicing in mammalian nuclei. *Nature* 372: 809-12.

42. Tarn W.Y., Steitz J.A. 1994. SR proteins can compensate for the loss of U1 snRNP functions *in vitro*. *Genes Dev.* 8: 2704-17.

43. Madhani H.D., Guthrie C. 1992. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 71: 803-17.

44. Steitz T.A., Steitz J.A. 1993. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* 90: 6498-502.

45. Wise J.A. 1993. Guides to the heart of the spliceosome. *Science* 262: 1978-9.

46. Weiner A.M. 1993. mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? *Cell* 72: 161-4.

47. Zucker M. 1989. On finding all suboptimal foldings of an RNA molecule. *Science* 244: 48-52.

48. Jeager J.A., Turner D.H., Zucker M. 1989. Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci. USA* 86: 7706-10.

Publication n:°3 :

Le gène du snoRNA U3 de la levure *Hansenula wingei* contient un intron et sa séquence codante a co-évolué avec la région 5' ETS du pré-ARN ribosomique.

F. Brulé, J. Venema, V. Ségault, D. Tollervey et C. Branlant

RNA (1996), vol. 2 183-197

RNA (1996), 2:183–197. Cambridge University Press. Printed in the USA.
Copyright © 1996 RNA Society.

The yeast *Hansenula wingei* U3 snoRNA gene contains an intron and its coding sequence co-evolved with the 5′ ETS region of the pre-ribosomal RNA

FABIENNE BRULÉ,[1] JAAP VENEMA,[2] VÉRONIQUE SÉGAULT,[1] DAVID TOLLERVEY,[2] and CHRISTIANE BRANLANT[1]

[1] Laboratoire d'Enzymologie et de Génie Génétique, URA CNRS 457, Université de Nancy I, Bd des Aiguillettes, BP 239, F-54506 Vandœuvre-lès-Nancy cédex, France
[2] European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

ABSTRACT

The 5′ external transcribed spacer (ETS) region of the pre-rRNA in *Saccharomyces cerevisiae* contains a sequence with 10 bp of perfect complementarity to the U3 snoRNA. Base pairing between these sequences has been shown to be required for 18S rRNA synthesis, although interaction over the full 10 bp of complementarity is not required. We have identified the homologous sequence in the 5′ ETS from the evolutionarily distant yeast *Hansenula wingei*; unexpectedly, this shows two sequence changes in the region predicted to base pair to U3. By PCR amplification and direct RNA sequencing, a single type of U3 snoRNA coding sequence was identified in *H. wingei*. As in the *S. cerevisiae* U3 snoRNA genes, it is interrupted by an intron with features characteristic of introns spliced in a spliceosome. Consequently, this unusual property is not restricted to the yeast genus *Saccharomyces*. The introns of the *H. wingei* and *S. cerevisiae* U3 genes show strong differences in length and sequence, but are located at the same position in the U3 sequence, immediately upstream of the phylogenetically conserved Box A region. The 3′ domains of the *H. wingei* and *S. cerevisiae* U3 snoRNAs diverge strongly in primary sequence, but have very similar predicted secondary structures. The 5′ domains, expected to play a direct role in pre-ribosomal RNA maturation, are more conserved. The sequence predicted to base pair to the pre-rRNA contains two nucleotide substitutions in *H. wingei* that restore 10 bp of perfect complementarity to the 5′ ETS. This is a strong phylogenetic evidence for the importance of the U3/pre-rRNA interaction.

Keywords: *H. wingei*; intron; pre-ribosomal RNA; spliceosome; U3 snoRNA

INTRODUCTION

In eukaryotes, the small subunit ribosomal RNA, 18S rRNA, and two large subunits rRNAs, 5.8S and 25/28S rRNAs, are transcribed by RNA polymerase I as a single precursor molecule (Tiollais et al., 1971). In this pre-ribosomal RNA (pre-rRNA), the rRNA coding regions are flanked and separated by noncoding regions called spacers. A complex series of processing reactions are needed to eliminate the 5′ and 3′ external transcribed spacers (5′ ETS and 3′ ETS, respectively) and the internal transcribed spacers 1 and 2 (ITS1 and 2, respectively) (for reviews, see Eichler & Craig, 1994; Raué & Planta, 1995; Venema & Tollervey, 1995).

These processing steps take place in the nucleolus, a compartment of eukaryotic nuclei devoted to the transcription, processing, and assembly of rRNA into ribosomal subunits. A growing number of small nucleolar RNAs (snoRNAs) are being characterized (for review, Filipowicz & Kiss, 1993; Fournier & Maxwell, 1993; Maxwell & Fournier, 1995). Among 20 snoRNAs characterized to date in yeast, only 5 have been shown to be required for normal pre-rRNA processing (U3, U14, snR10, snR30, and RNase MRP) (Tollervey, 1987; Li et al., 1990; Hughes & Ares, 1991; Morrissey & Tollervey, 1993; Lygerou et al., 1994).

Following its discovery (Zieve & Penman, 1976), the U3 snoRNA was found to be associated with the 28S/

Reprint requests to: C. Branlant, Laboratoire d'Enzymologie et de Génie Génétique, URA CNRS 457, Université de Nancy I, Bd des Aiguillettes, BP 239, F-54506 Vandœuvre-lès-Nancy cédex, France; e-mail: cbranlant@legg.u-nancy.fr.

35S nucleolar RNA in a deproteinized mammalian nucleolar extracts (Calvet & Pederson, 1981) and with preribosomal particles >60S in cell extracts (Epstein et al., 1984). It was therefore proposed to base pair with the pre-rRNA and to function in some aspect of the pre-rRNA processing. However, despite considerable effort, the site of the postulated base pairing could not be identified at that time (Bachellerie et al., 1983; Crouch et al., 1983; Tague & Gerbi, 1984). More recently, crosslinking experiments with psoralen derivatives revealed that U3 interacts with the 5′ ETS region of the pre-rRNA in the vicinity of the first pre-rRNA cleavage site in human (Maser & Calvet, 1989), rat (Stroke & Weiner, 1989), mouse (Tyc & Steitz, 1992), and yeast cells (Beltrame & Tollervey, 1992). In *Saccharomyces cerevisiae*, sequences crosslinked in the U3 snoRNA and 5′ ETS display a perfect complementarity of 10 nt (Beltrame & Tollervey, 1992). The U3 snoRNA sequence involved (nt 39–49) is predominantly single-stranded in the secondary structure model established previously for the *S. cerevisiae* U3A snoRNA (Ségault et al., 1992).

Evidence for the involvement of the U3 snoRNA in pre-rRNA processing has been obtained both in vertebrates and yeast. The earliest pre-rRNA processing reactions take place in the 5′ ETS region and these have been reproduced successfully in extracts of cultured cells from mouse (Craig et al., 1987) and *Xenopus laevis* (Mougey et al., 1993b). This processing event is abolished in extracts depleted of U3 snoRNA by oligonucleotide-directed RNase H cleavage or immunoprecipitation (Kass et al., 1990; Mougey et al., 1993b). In yeast, the genetic depletion of the U3 snoRNA or its associated proteins (Nop1p and Sof1p) inhibits pre-rRNA processing in the pathway leading to the synthesis of the 18S rRNA, resulting in an under-accumulation of mature 18S rRNA (Hughes & Ares, 1991; Tollervey et al., 1991; Jansen et al., 1993). Substitution of the U3-complementary sequence in the 5′ ETS region has the same effect as depletion of U3 snoRNA (Beltrame et al., 1994; Beltrame & Tollervey, 1995). Processing can be restored by mutations in U3 that restore base pairing, demonstrating a requirement for direct base pairing between U3 and the pre-rRNA (Beltrame & Tollervey, 1995). However, a mutation that only partially restores the base pairing between U3 and the pre-rRNA can also restore processing, demonstrating that the full 10 bp are not required for a productive interaction.

Although the 5′ ETS regions of both yeast and vertebrate pre-rRNAs have been shown to bind U3 and undergo U3-dependent processing, the relationship between the sites and interactions reported in yeast and vertebrates is unclear. The low conservation of transcribed spacer regions of the pre-rRNA means that clear alignments of spacer sequences from evolutionarily distant organisms is not generally possible. We therefore wished to compare the 5′ ETS from *S. cerevisiae* with that of an organism in which we expected to be able to derive a clear alignment, at least for functionally important sites, but that would still show substantial divergence. Previous analyses of ITS1 and ITS2 (Van Nues et al., 1994, 1995) have shown the spacer regions in *Hansenula wingei* to have a useful level of divergence for such studies, and it was therefore of interest to examine the sequence of the 5′ ETS from this organism.

When the U3 genes of *S. cerevisiae* were cloned and sequenced, they were found, unexpectedly, to each contain an intron with the characteristic features of introns spliced in a spliceosome (Myslinski et al., 1990). In the Schizosaccharomycetoidae subfamily (fission yeasts), the U3 genes from *Schizosaccharomyces pombe* lack introns (Porter et al., 1988) and introns are not present in the U3 genes of plants or metazoans. No other U3 genes have been sequenced from the Saccharomycetoidae subfamily (budding yeasts), and it was of interest to check whether introns are present in the U3 genes of other Saccharomycetoidae or whether this property was restricted to *S. cerevisiae*. *H. wingei* belongs to the subfamily Saccharomycetoidae, but not to the genus *Saccharomyces* (Kreger-Van-Rij, 1973) and was therefore also an interesting example to test for the generality of the presence of intron in U3 genes of the subfamily Saccharomycetoidae.

We have analyzed the sequences of the 5′ ETS and the U3 snoRNA from *H. wingei* and report that the single U3 gene in this yeast does indeed contain an intron and the snoRNA has the potential to form 10 perfect base pairs with the pre-rRNA.

RESULTS

Analysis of the 5′ ETS from *H. wingei*

The rDNA repeat of *H. wingei* (strain ACT 14356) was cloned and mapped previously by restriction digestion (Verbeet, 1982; Verbeet et al., 1984). A 1-kb *Xba* I fragment was subcloned into pUC19 and sequenced by primer walking. This region contains the entire 5′ ETS, except for a few 3′-terminal nucleotides, and 119 nt of 5′ flanking sequence. The remaining sequence was determined by sequencing on a 3-kb *Bam*H I–*Eco*R I fragment subcloned into pUC19, using a primer to conserved 18S rRNA sequence. In addition, the 3′ region of the ETS was sequenced from three other clones obtained by PCR amplification (see the Materials and methods). The 5′ end of the ETS was identified by primer extension. The *H. wingei* ETS region is 928 nt in length; substantially longer than the ETS from *S. cerevisiae* (699 nt). The overall homology between the ETS sequences is low, but, despite this and the length difference, much of the ETS can be unambiguously aligned when the *H. wingei* sequence is drawn onto the secondary structure model that was proposed for *S. ce-*

revisiae (Yeh & Lee, 1992) on the basis of chemical probing of in vitro transcribed RNA (Fig. 1). Nucleotides that are identical in the primary sequences and that occupy the same positions in the predicted structure are shown in reverse-contrast. The 5′ proximal, single-stranded regions of both sequences are clearly homologous; 3′ to this region, a stem-loop structure (domain I) can be drawn for both sequences, but this region has no significant primary sequence homology. The extra nucleotides in the 5′ ETS of *H. wingei* appear to be accommodated in a large (331 nt) insertion into the sequence corresponding to the loop of domain II in the *S. cerevisiae* ETS. The sequences 3′ to position +85 (the loop of domain II in *S. cerevisiae*) and +443 in *H. wingei* are largely colinear, with only minor differences in the lengths of stems in domains III, IV, V, and VI. In each of these proposed structures, regions of identity between the *S. cerevisiae* and *H. wingei* sequences are observed that lie on both sides of the stems. Some of them are involved in highly stable helices, as is the case in domains III and IV, that are highly supported by base compensatory mutations. In contrast, conserved sequences in domains V and VI are found in helices predicted previously with very low stability. The terminal region of stem-loop VII is well-supported phylogenetically and several base compensatory mutations are observed. The structure predicted previously for the base of this stem-loop structure is less stable and is shorter in *H. wingei*. Stem-loops designated VIII and IX are predicted to form in both ETS sequences, but this region shows no significant primary sequence homology. Domain X consists of a long, stable stem-loop structure in *S. cerevisiae*. No stem of comparable stability can be formed in *H. wingei*, and from computer predictions it is very likely that this region folds into two much shorter stems. A small 3-bp stem-loop structure was proposed immediately 5′ to the *S. cerevisiae* A_0 cleavage site. Its counterpart in *H. wingei* would have a 2-nt terminal loop, which is thermodynamically unlikely, and these nucleotides are, therefore, not drawn base paired. In the 3′ region, most of the conserved nucleotide stretches are single-stranded and are located either in the region containing the U3-complementary sequence or in the vicinity of the processing sites.

The 3′ region of the *S. cerevisiae* ETS contains two processing sites that can be detected by primer extension. Cleavage at site A_0, located between nt +609 and +610, requires the U3 snoRNA (Hughes & Ares, 1991) and the snoRNA-associated protein Nop1p (Beltrame et al., 1994). Cleavage at site A_0 is inhibited in strains genetically depleted of Nop1p by growth of a GAL::nop1 strain (Tollervey et al., 1991) on glucose-containing medium (Fig. 2, lane 2) or carrying a temperature sensitive (ts) nop1-5 mutation (Fig. 2, lane 3), which blocks many steps in pre-rRNA processing (Tollervey et al., 1993). A second primer extension stop is seen in wild-type pre-rRNA at position +471, corresponding to cleavage between nt +471/2 (Fig. 2, lane 1). This site lies within the region that base pairs with the U3 snoRNA. Cleavage at +471 is also inhibited by depletion of Nop1p (Fig. 2, lane 2), but pre-rRNA cleaved at +471 appears to accumulate in the nop1-5 strain (Fig. 2, lane 3). Neither the stop at site A_0 nor at +470 is detected by primer extension on RNA transcribed in vitro (Fig. 2, lane 4), indicating that they are not due to secondary structure or other features intrinsic to the RNA sequence. To exclude the possibility that the stops were due to some element of secondary structure that only forms in vivo, RNA from the wild-type, GAL::nop1, and nop1-5 strains was subjected to stringent denaturation (heating to 95 °C for 5 min in the presence of 5 M urea). This did not affect the intensity of the primer extension stops at A_0 or +471 (data not shown). We conclude that the primer extension stop at +471/2 is likely to represent a site of pre-rRNA processing that, at least in some pre-rRNA molecules, is cleaved prior to cleavage at sites A_0 and A_1.

Primer extension was also performed on RNA extracted from *H. wingei* (Fig. 2, lane 5) and on RNA transcribed in vitro from the cloned *H. wingei* ETS sequence (Fig. 2, lane 6). Two stops detected by primer extension on the *H. wingei* RNA, at +872 and +791 (marked with arrows in Fig. 2), indicative of cleavages between nt +871/2 and +790/1, align precisely with sites A_0 and +470 in the *S. cerevisiae* pre-rRNA. An additional strong primer extension stop is observed at position +771. This stop falls within the conserved bulged loop sequence in domain VII, 5′ to the U3-complementary region (see Fig. 1), but no stop is observed in the homologous sequence in *S. cerevisiae* and the significance of this stop is not clear. Other primer extension stops are also detected on RNA transcribed in vitro, and are therefore assumed to be due to secondary structure in the RNA. All of the primer extension stops observed on RNA synthesized in vivo in *H. wingei* appear as doublets, with the minor, upper stop 5 nt longer than the major stop. We assume this to be due to heterogeneity among the rDNA repeats in *H. wingei*. Heterogeneity between different repeats in the same organism has been reported in other species, including *S. cerevisiae* (Petes et al., 1978). Four different clones sequenced from *H. wingei* had the same sequence in the 3′ region of the 5′ ETS and we have not investigated this further.

The positions of the primer extension stops in *S. cerevisiae* and *H. wingei* are shown by arrows in Figure 1A and B. The nucleotides at the site of A_0 cleavage are different, AU in *S. cerevisiae* and AC in *H. wingei*. From the sequence and secondary structure comparison in Figure 1, the sequence in the *H. wingei* ETS that is homologous to the U3 snoRNA binding site in the *S. cerevisiae* ETS appears to be clear. This identification is also supported by the identical positions of the primer extension stops within each sequence. It is, however,

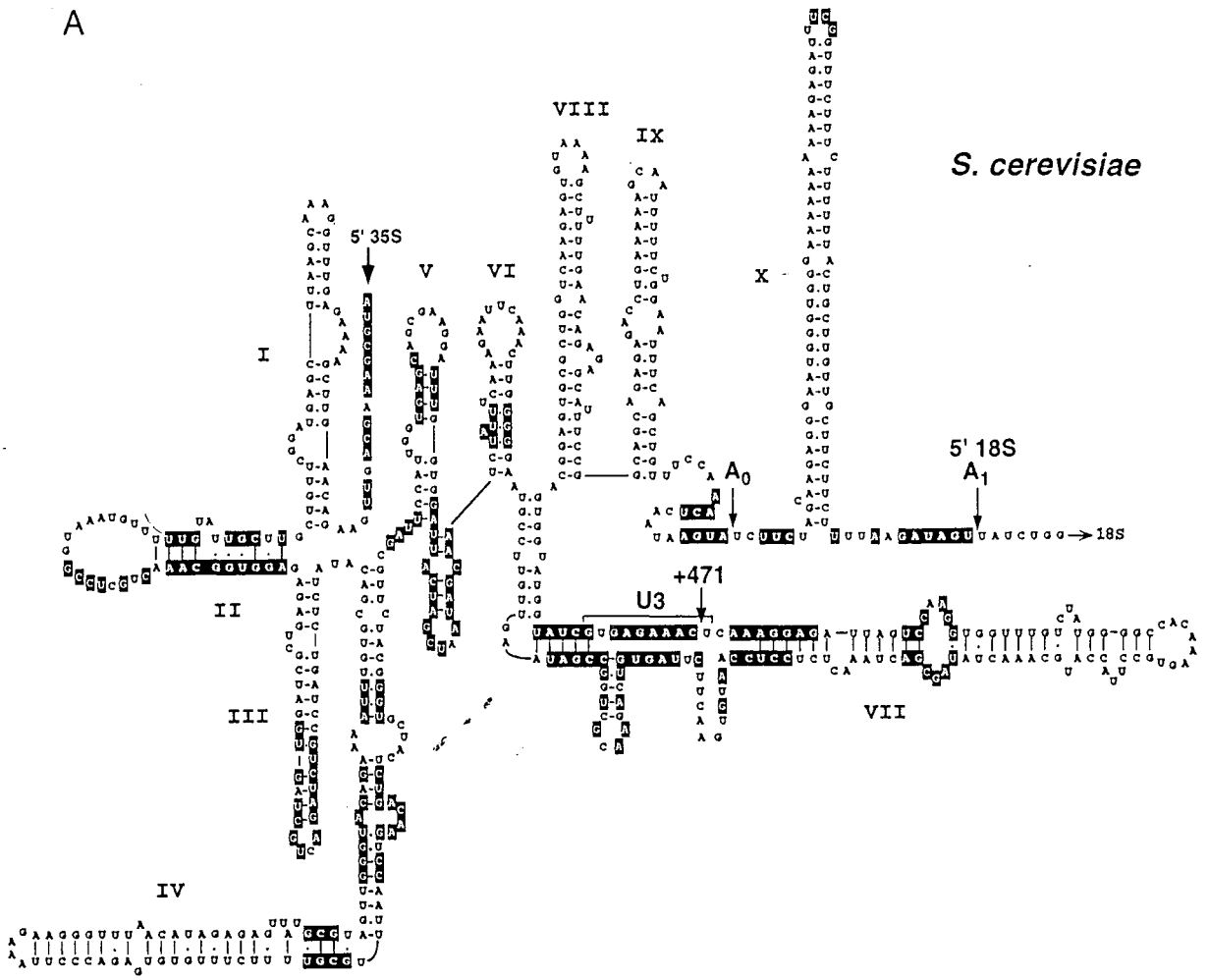


FIGURE 1. Predicted secondary structures for the 5' ETS from *H. wingei* and *S. cerevisiae*. Secondary structure shown for *S. cerevisiae* is modified from Yeh and Lee (1992). *H. wingei* ETS sequence is drawn in a similar structure; regions in which identical nucleotides are located in the same predicted structure are shown in counter-contrast. Processing sites are indicated by arrows and sequences complementary to the cognate U3 snoRNAs are overlined.

striking that there are two substitutions within the putative U3 binding sites. This suggested that the sequence of the U3 snoRNA from *H. wingei* might contain compensatory mutations restoring base complementary to this region.

An intron in the single U3 snoRNA gene sequence detected in *H. wingei*

We took advantage of the high phylogenetic conservation of the U3 snoRNA 5'- and 3'-terminal sequences (Myslinski et al., 1990), to amplify the *H. wingei* U3 snoRNA genes by the PCR method.

Two pairs of oligonucleotides, 1 and 2, were designed for this purpose. Oligonucleotide A, which is complementary to the phylogenetically highly conserved Box D (Hughes et al., 1987), is common to the two pairs. Couple 1 consists of oligonucleotide A and oligonucleotide B, which contains the sequence found at the 5' end of both the U3A and U3B snoRNA from *S. cerevisiae* (Box A') (Myslinski et al., 1990). Because this 5'-terminal sequence is less conserved in evolution than the 3'-terminal sequence, couple 2 was also used. Couple 2 consists of oligonucleotide A and the oligonucleotide C, which contains the sequence of the first exon of the *S. cerevisiae* U3A and U3B genes, followed by the 5'-terminal sequence conserved in all *S. cerevisiae* introns that are spliced in a spliceosome (Pikielny et al., 1983). Couple 2 was expected to allow amplification of U3 genes that carry an intron located at position 14 of the coding sequence, as found for the *S. cerevisiae* U3A and U3B genes (Myslinski et al., 1990).

The *H. wingei* genomic DNA was prepared as described in the Materials and methods, and the results of DNA amplification are shown in Figure 3. As a control, amplifications were also performed on the genomic DNA from the *S. cerevisiae* strain FL100. Two amplified fragments with the expected size were ob-

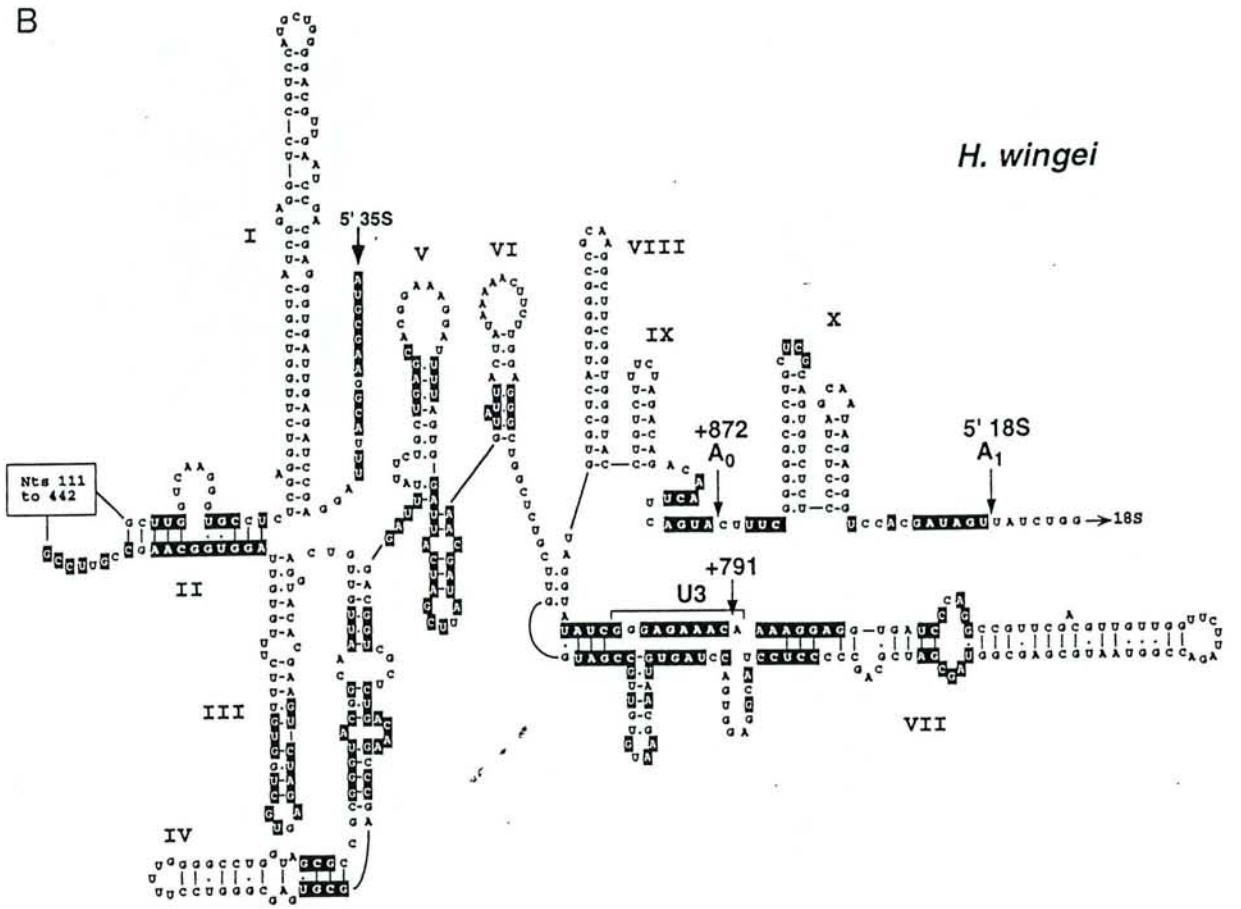


FIGURE 1B.

tained for *S. cerevisiae*. In contrast, a unique fragment with a smaller size (about 380 nt according to the size markers used in this experiment) was obtained for *H. wingei*. The same result was obtained with the oligonucleotide couples 1 and 2 (not shown). This suggested the presence of a unique U3 gene in *H. wingei*.

This was tested by nucleotide sequence analysis of the amplified DNA fragment and by Southern blot experiments. The oligonucleotide primers A and B used for amplification generated a *Sal* I and an *Hpa* I restriction site, respectively. These restriction sites were used to clone the amplified fragment in the pBluescript SK(+) plasmid. A recombinant pBS-U3Hw plasmid was purified and the inserted DNA fragment was sequenced. The amplified DNA fragment was found to have a homogeneous sequence (Fig. 5), which reinforced the idea of a unique U3 gene in *H. wingei*. The sequence of the exon 1 and of the 3′ terminus of the *H. wingei* U3 snoRNA could not be determined from the amplified DNA, because the primers used for amplification covered these sequences. The sequence of the first exon was therefore determined directly by sequencing the RNA using reverse transcriptase with oligonucleotides A, G, and H as primers (Fig. 6). A unique sequence identical to that determined for the amplified DNA fragment was displayed on the sequencing gel. Southern blot experiments were performed on the *H. wingei* genomic DNA and, as a control, on the *S. cerevisiae* genomic DNA. Digestions were performed with the *Eco*R I, *Hin*d III, and *Pst* I nucleases. The *H. wingei* DNA amplification product generated with oligonucleotide couple 1 was labeled by nick translation and used as the probe. As shown in Figure 4, in the case of *S. cerevisiae*, for each digestion, two fragments hybridized with the *H. wingei* U3 probe. In contrast, a unique hybridization fragment was found for all *H. wingei* digestions. According to the size markers, the *S. cerevisiae* hybridization fragments had the size expected for the U3A and U3B snoRNA genes and, when the DNA amplification product generated with the oligonucleotide couple 1 and the cloned *S. cerevisiae* U3A snoRNA gene (Ségault et al., 1992) was used as the probe, a similar hybridization pattern was obtained (Fig. 4). Hence, both the U3A and U3B snoRNA genes hybridized with the *H. wingei* U3 probe. Because the U3A and U3B snoRNA gene sequences are very

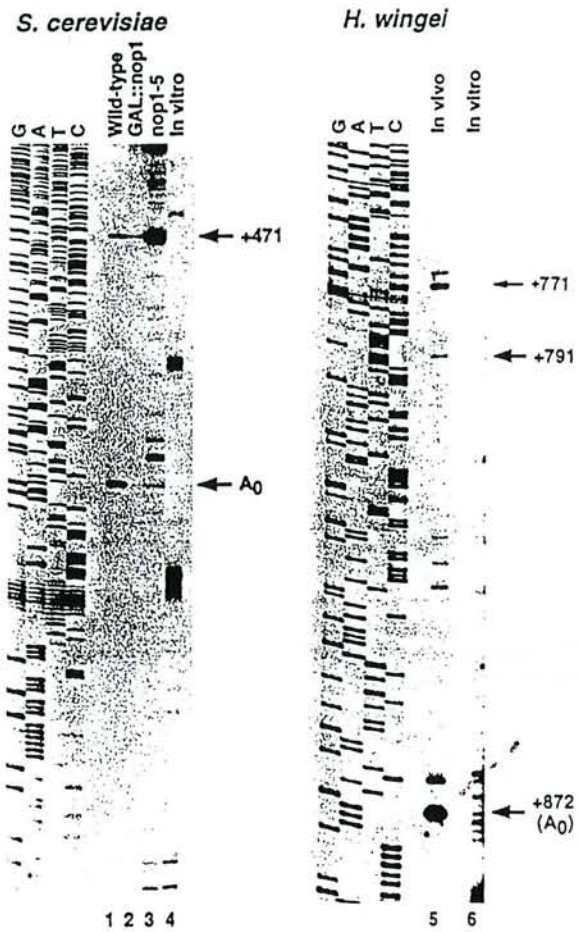


FIGURE 2. Primer extension analysis of the ETS of *S. cerevisiae* and *H. wingei*. Lane 1, wild-type strain of *S. cerevisiae*; lane 2, GAL::nop1 strain of *S. cerevisiae*; lane 3, nop1-5 strain of *S. cerevisiae*; lane 4, pre-rRNA from *S. cerevisiae* transcribed in vitro; lane 5, RNA extracted from *H. wingei*; lane 6, pre-rRNA from *H. wingei* transcribed in vitro. RNA was extracted from the GAL::nop1 strain after growth on glucose medium for 24 h and from the nop1-5 strain after growth at nonpermissive temperature (37 °C) for 6 h. RNA was transcribed in vitro from a clone with the full-length *S. cerevisiae* pre-rRNA sequence or from a clone containing the *Eco*R I-*Bam*H I region of the *H. wingei* rDNA. The primer used hybridizes to a conserved sequence within the mature 18S rRNA close to site A_1.

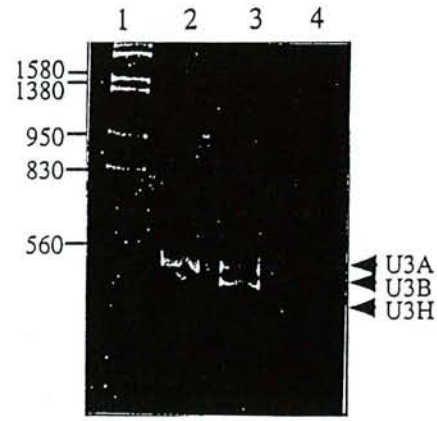


FIGURE 3. PAGE of the amplified U3 genes from *H. wingei* and *S. cerevisiae*. PCR amplification of genomic DNAs of *H. wingei* (lane 4) and *S. cerevisiae* (lane 3) and plasmid pFL1::snR17A containing the *S. cerevisiae* U3A snoRNA gene (Myslinski et al., 1990) (lane 2) was performed with the oligonucleotide couple 1. λ DNA digested with *Hind* III-*Eco*R I was used as a size marker (lane 1). Sizes of the resulting DNA fragments are indicated to the left. Amplified DNA fragments corresponding to the *S. cerevisiae* U3A and U3B genes and the *H. wingei* U3 gene (Hw) are indicated to the right.

different from the *H. wingei* U3 snoRNA gene sequence, we assumed that a second, different U3 gene present in the *H. wingei* genome would also have hybridized with the probe in the hybridization conditions used. Thus, although it remains formally possible that tandem identical U3 genes are located in the unique *Eco*R I, *Hind* III, *Pst* I genomic fragments that hybridized with the Hw U3 probe, taking into consideration the results of PCR amplification and Southern blot analysis, we concluded for the presence of a unique type of U3 snoRNA coding sequence in *H. wingei*.

The sequence of the *H. wingei* gene is aligned with the *S. cerevisiae* U3A and U3B genes in Figure 5. The 5′-terminal sequence is identical to that of the *S. cerevisiae* U3A and U3B snoRNAs and the sequences of Boxes A′, A, B, and C are all highly conserved in the three genes. From the DNA sequence, the *H. wingei* U3 snoRNA is 315 nt in length, shorter than the *S. cerevisiae* U3A and U3B snoRNAs (333 nt). The *H. wingei* gene contains an intron that is 67 nt in length, also shorter than the introns of the *S. cerevisiae* U3A and U3B genes (157 nt and 130 nt, respectively) (Myslinski et al., 1990). The intron in *H. wingei* is located at the same position within the U3 sequence as the introns in the U3A and U3B genes. The *H. wingei* intron has the 5′ and 3′ splice site sequences and intron branch site sequence characteristic of introns spliced in a spliceosome. In addition to the conserved 5′-terminal hexanucleotide, some sequence similarity is observed in the 24-nt 5′-terminal segment. Most of the sequences that are present between the 24-nt 5′-terminal segment and the branch site sequence in the U3A and U3B introns are absent in the U3Hw intron. The U3Hw intron displays a canonical sequence at the branch site (TAC TAAC), instead of the GACTAAC sequence in the U3A and U3B introns; the distance between the branch site and the 3′ splice site is 16 nt, compared to 9 nt in the *S. cerevisiae* U3 genes.

Compensatory mutations between the 5′ ETS and the U3 snoRNA complementary regions in *H. wingei*

Based on the secondary structure model established previously for the *S. cerevisiae* U3A snoRNA (Ségault

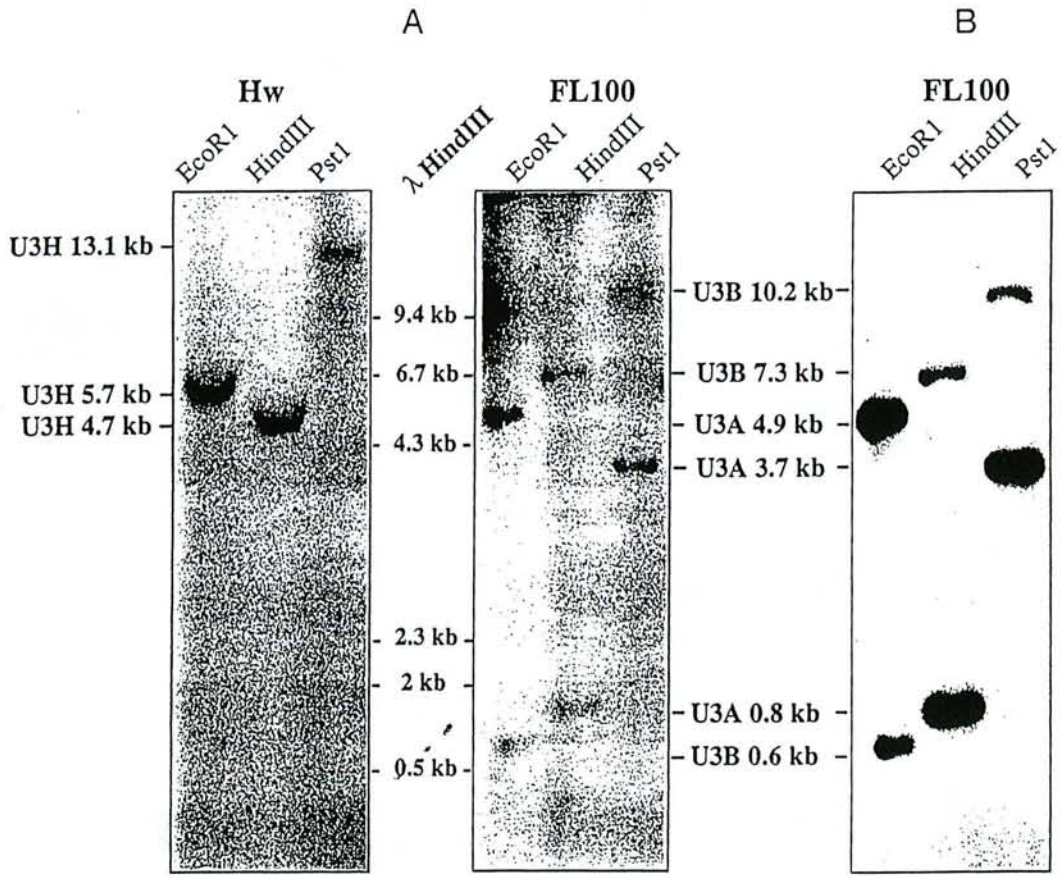


FIGURE 4. Southern blot analysis of U3 snoRNA genes in *H. wingei*. Genomic DNAs from *H. wingei* (Hw) and *S. cerevisiae* FL100 (FL100) were prepared (see the Materials and methods) and digested with the *Eco*R I, *Hind* III, and *Pst* I nucleases. Resulting fragments were fractionated by electrophoresis on a 0.8% agarose gel. The DNA fragment amplified from *H. wingei* genomic DNA using the oligonucleotide couple 1 was labeled by nick translation and used as a probe (**A**). Positions of the λ *Hind* III digestion fragments, which were used as size markers, are represented in the middle. The size of the fragments that hybridized with the probe were estimated from the size markers are are indicated on the left (*H. wingei*) and right (*S. cerevisiae*) sides of the panels. For *S. cerevisiae*, fragments corresponding to U3A or U3B genes were distinguished by reference to the restriction maps established for the U3A and U3B snoRNA genes (Hughes et al., 1987; Myslinski et al., 1990) and by hybridization with the U3A snoRNA gene amplification product (**B**) (a more intense hybridization signal is obtained for the fragments containing the U3A snoRNA gene).

et al., 1992), we built a secondary structure model for the *H. wingei* U3 snoRNA (Fig. 7). In spite of strong differences of primary structure in the 3′-terminal region, the same secondary structure can be proposed. Most of the mutations compensate each other to preserve the predicted 3′-terminal cruciform structure.

The 5′-terminal domain is more conserved at the level of the primary structure and can be folded in the same way in the two RNAs. The *H. wingei* U3 snoRNA segment between positions 38 and 50 displays two mutations as compared to *S. cerevisiae* U3A and U3B, and these two mutations compensate those detected in the 5′ ETS region of *H. wingei*. Thus, in the base paired region between the U3 snoRNA and the 5′ ETS segment, two A-U pairs in *S. cerevisiae* are replaced by a U-A and a C-G pair in *H. wingei* (Fig. 7). This is strong evidence for the biological significance of the base pair interaction between U3 and the pre-rRNA.

DISCUSSION

In contrast to the situation observed for *S. cerevisiae*, analysis at the genomic and RNA level reveals a unique U3 snoRNA sequence in *H. wingei*. As found in *S. cerevisiae*, the *H. wingei* U3 gene contains an intron with the characteristic features of introns spliced in a spliceosome. In *S. cerevisiae*, the introns of the U3 genes display some unusual features, as compared to introns contained in protein genes: a GACUAAC sequence at the branch site instead of a UACUAAC sequence and a very short distance between the branch site and the 3′ splice site (9 nt). In contrast, the intron in the *H. wingei* U3 gene has a normal UACUAAC branch site sequence located 16 nt from the 3′ splice site. Thus, the unusual characteristics of the *S. cerevisiae* U3 snoRNA introns are not general features of introns contained in U3 genes.

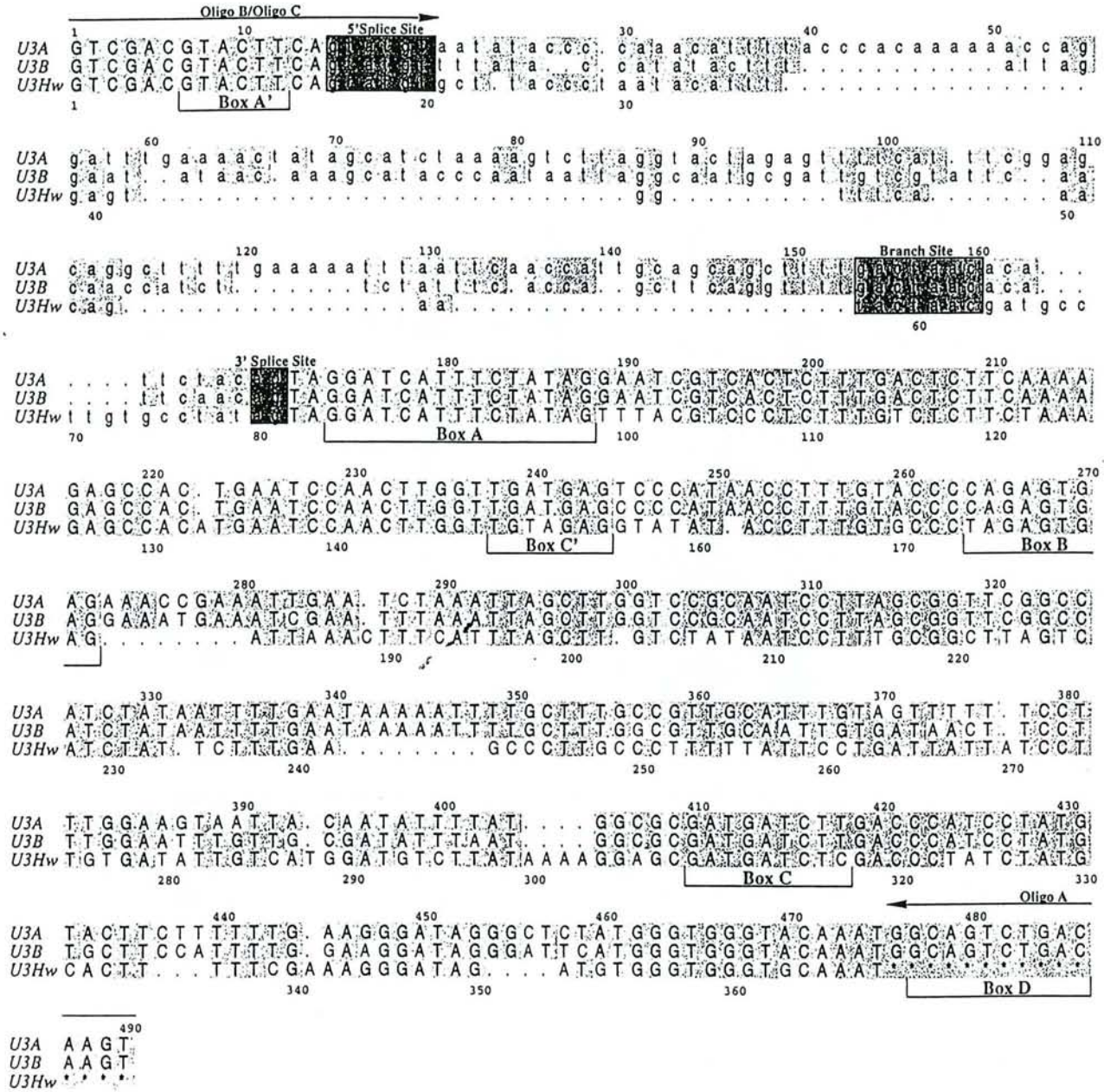


FIGURE 5. Alignment of the *H. wingei* and *S. cerevisiae* U3 gene sequences. Alignment was achieved with the PILEUP program of the GCG software. Identical nucleotides are indicated by the grey color using the Alscript program. The 5' and 3' splice sites and the intron branch sites are represented by dark grey boxes. Phylogenetically conserved Boxes, A', A, B, C, and D, are indicated by horizontal lines below the sequences. Sequences used to design the oligonucleotide primers for DNA amplification are indicated by horizontal arrows above the sequences. *S. cerevisiae* U3A and U3B sequences are from Myslinski et al. (1990). The EMBL Nucleotide Sequence Database accession number for the *H. wingei* U3 snoRNA gene is X91005.

In spite of the low level of sequence homology between the introns of the *S. cerevisiae* and *H. wingei* genes, it is notable that the *H. wingei* intron is located exactly at the same position of the coding sequence as the *S. cerevisiae* introns, that is to say, upstream of Box A. This is an interesting location, because Box A, which has a phylogenetically highly conserved sequence, is believed to play an essential role in U3 snoRNA function. Box A can be crosslinked to the 5' ETS region by psoralen derivatives in vivo (Maser & Calvet, 1989; Stroke & Weiner, 1989; Beltrame & Tollervey, 1992; Tyc & Steitz, 1992). Among the phylogenetically highly conserved regions of U3 snoRNA, only Box A is accessible in the U3 snoRNP (Parker & Steitz,

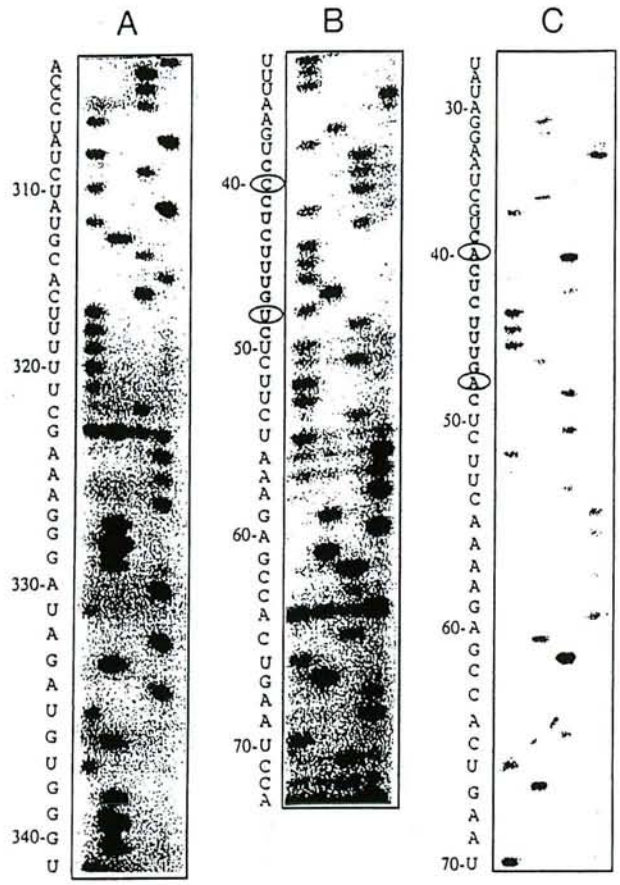


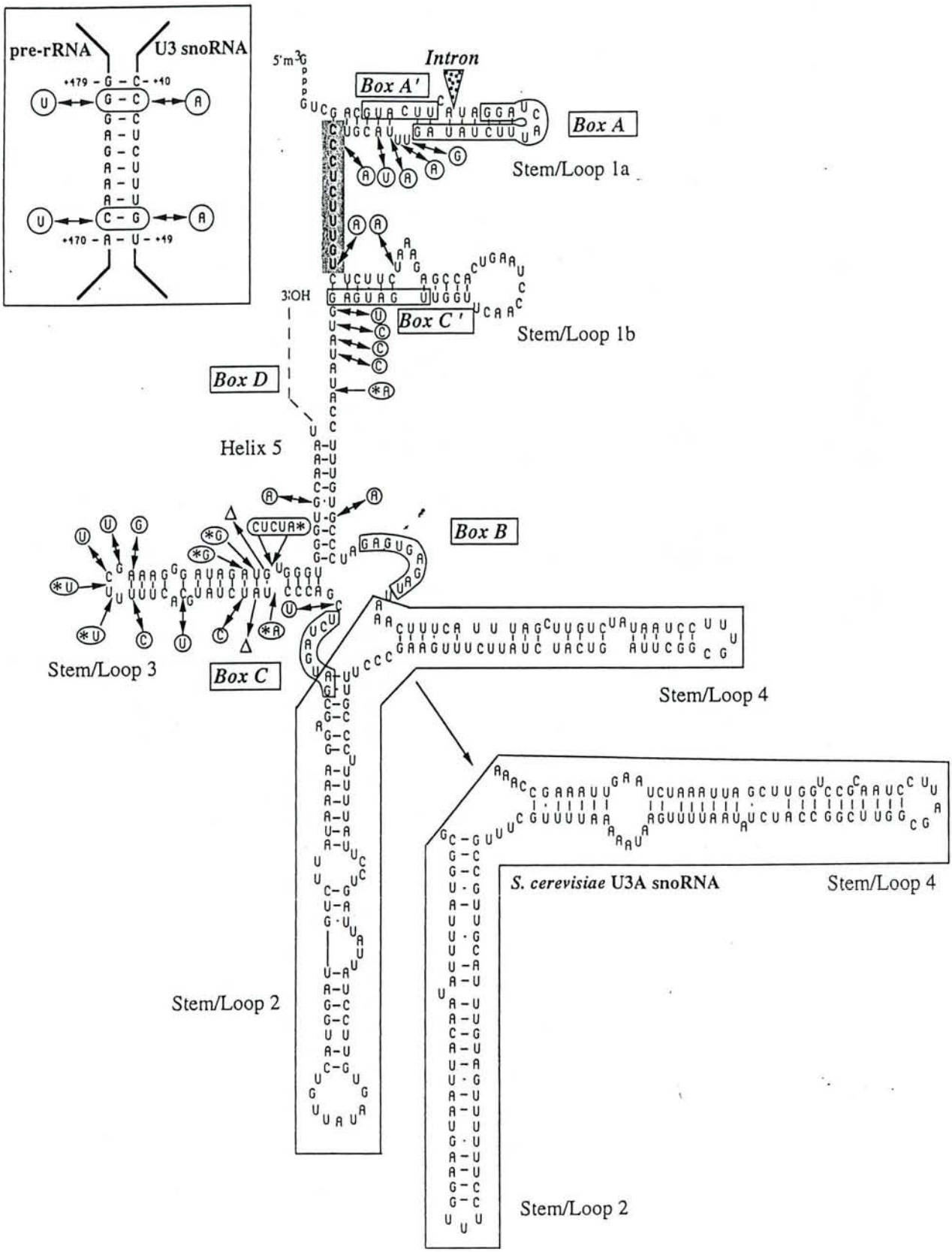
FIGURE 6. Direct sequence analysis of the *H. wingei* U3 snoRNA. The 4S-8S RNA mixture extracted from *H. wingei* or *S. cerevisiae* cells was used as template for U3 snoRNA sequence analysis using reverse transcriptase. Oligonucleotides A (**A**) and H (**B**) were used as primers for sequence analysis of the *H. wingei* U3 snoRNA. For comparison, the sequence analysis of the *S. cerevisiae* U3 snoRNA using oligonucleotide G as the primer is also shown (**C**). Nucleotide sequences are indicated to the left and are numbered starting from the 5′ end of the molecule. In B and C, the sequence differences between *H. wingei* and *S. cerevisiae* in the segment proposed to base pair with the 5′ ETS region (nt 39–49) are indicated by circled nucleotides.

1987; our unpublished results on the *S. cerevisiae* U3 snoRNP). This suggests that, whereas the other conserved sequences are probably involved mainly in protein binding, Box A is likely to make direct contact with the pre-rRNA.

Only a limited number of introns spliced in a spliceosome have been found in genes encoding metabolically stable RNAs. They were essentially found in genes coding for the spliceosomal U-snRNAs (U2 and U6) of yeasts (Tani & Ohshima, 1989, 1991; Frendewey et al., 1990; Takahashi et al., 1993) and are located in regions encoding RNA segments that are essential for the function of the U-snRNA (for review, Madhani & Guthrie, 1994). The location of introns in functional areas of metabolically stable RNAs is not restricted to introns spliced in a spliceosome. The enzymatically spliced introns from tRNA genes are located in the sequence coding for the anticodon loop (for review, see Peebles et al., 1983) and the autocatalytic introns found in genes coding for the large ribosomal rRNA are located in domain IV or V, that lie in the catalytic center of the ribosome (Branlant et al., 1981; Brimacombe et al., 1993). One reasonable explanation for this peculiar location of introns in RNA genes is that insertion in genomic DNA coding for essential RNA region can only be maintained in the progeny if the additional RNA sequence can be eliminated from the RNA transcript by a maturation process. Such location is also consistent with an extended view of the model proposed originally for the spliceosomal U6 snRNA intron (Brow & Guthrie, 1989); the insertion events may occur in highly exposed functional regions of RNA by reverse splicing events, with the subsequent return of the information to the DNA by reverse transcription.

The phylogenetically highly conserved Boxes of U3 snoRNA (A′, A, B, and C) have identical sequences in *S. cerevisiae* and *H. wingei*. Box D is probably either identical or very similar in the two yeast species, because the *H. wingei* sequence hybridizes efficiently to the oligonucleotide primer A, which has the Box D sequence found in *S. cerevisiae*. In addition to the sequences of Boxes A′, A, B, and C, the nucleotide sequences of the helix-loop structure 1b and the helix 1c are remarkably conserved in the two yeast species. In contrast, in regions corresponding to stem-loop structures 2, 3, and 4 display a high level of sequence divergence. However, in spite of this sequence divergence, stem-loop structures similar to those predicted for the *S. cerevisiae* U3A snoRNA can be proposed for the *H. wingei* U3 snoRNA. The high sequence divergence of these stems is consistent with the idea that they essentially play structural roles in the U3 snoRNP. Stem-loop structure 2 is absent from mammalian U3 snoRNAs (Ségault et al., 1992), whereas stem-loop 4 is extended in yeast. The presence of a cruciform structure, together with the conserved sequences of Boxes B and C and helix 1c, may constitute the binding site for the yeast U3 snoRNP proteins.

The transcribed spacer regions of eukaryotic pre-rRNA are, in general, very poorly conserved in evolution. It is therefore necessary to compare sequences from relatively close organisms in order to obtain clear, useful sequence alignments. Previous analyses of the ITS regions have shown that the divergence between *H. wingei* and *S. cerevisiae* is suitable for such comparisons (Van Nues et al., 1994, 1995). As was observed for ITS1 and ITS2, the 5′ ETS sequences from these organisms show low overall homology, with local regions of conservation that, generally, presumably represent functionally important sites. Overall, the comparison between the ETS sequences reveals a pattern of conservation of short localized sequences within stems, in internal loops, and in single-stranded sequences. This


pre-rRNA
U3 snoRNA
Intron
Box A'
Box A
Stem/Loop 1a
Box C'
Stem/Loop 1b
Box D
Helix 5
Box B
Box C
Stem/Loop 3
Stem/Loop 4
S. cerevisiae U3A snoRNA
Stem/Loop 4
Stem/Loop 2
Stem/Loop 2

suggests that many of these conserved elements may come together to form a highly ordered core structure. As is the case for ITS1 and ITS2 (Van Nues et al., 1994, 1995), conservation of 5′ ETS sequences is also high around sites of pre-rRNA processing.

Processing site A_0 can be identified readily in the *H. wingei* ETS. A strong primer extension stop is detected in the *H. wingei* pre-rRNA that aligns precisely with processing site A_0 in the *S. cerevisiae* sequence. However, although the sequences surrounding A_0 are conserved, the actual site of processing is different. Consistent with this, substitution of 6 nt across site A_0 in *S. cerevisiae* does not prevent accurate processing (Venema et al., 1995). The evolutionarily conserved flanking sequences are therefore likely to be important for specifying the site of A_0 processing. Similarly, the 5′ flanking sequence at site A_1 is also conserved (the 3′ flanking sequence, the mature 18S rRNA, is completely conserved in this region). These conserved nucleotides have been shown to be important for determining the position of A_1 cleavage (Venema et al., 1995). Truncation of the stem lying between A_0 and A_1 in *S. cerevisiae* (domain X in Figure 1) (Yeh & Lee, 1992) does not inhibit pre-rRNA processing (Venema et al., 1995). This extended stem structure in *S. cerevisiae* is replaced by two much shorter stems in *H. wingei*, consistent with the idea that the exact structure of this region is not important for processing.

Two portions of domain VII form stable helical structures (terminal and lateral stem/loop structures) that are strongly supported by base compensatory mutations. The parts of domain VII that are not involved in these stable helices have, in contrast, highly conserved sequences. Compared to the whole 5′ ETS region, the loosely structured areas of domain VII show a high degree of primary sequence conservation. Within this region, a long stretch of conserved sequence includes the U3 snoRNA binding site identified in *S. cerevisiae*. Only three differences are found in a 21-nt long segment and the two changes located in the U3 binding site are compensated for by two mutations in the *H. wingei* U3 snoRNA. The functional homology between the sequences at +470–+479 in *S. cerevisiae* and +790–+799 in *H. wingei* is supported by the primary sequence conservation, their similar locations in the overall secondary structure of the 5′ ETS region, and the presence of primer extension stops, assumed to the represent pre-rRNA processing sites, at identical positions within the conserved sequences (between nt +471/2 in *S. cerevisiae* and +791/2 in *H. wingei*).

In both *S. cerevisiae* and *H. wingei*, this primer extension stop is detected in RNA extracted from cells, but not in RNA transcribed in vitro. In a strain of *S. cerevisiae* depleted of Nop1p, the level of the stop declines, whereas pre-rRNA species that include this region are accumulated. In contrast, pre-rRNA species showing the stop at +471 accumulate in a nop1-5 mutant strain. Moreover, the stop cannot be removed by stringent denaturation (incubation at 95 °C in 5 M urea) from RNA extracted from *S. cerevisiae*. We cannot formally exclude the possibility that this primer extension stop is due to a site of covalent RNA modification, but no modification in any pre-rRNA spacer region has been reported and this, therefore, seems unlikely. Pre-rRNA species corresponding to cleavage at +471 have not been detected by northern hybridization; however, the pre-rRNA species cleaved at A_0 and A_3 are also not detected readily in wild-type cells. We conclude that it is likely that at least some pre-rRNA molecules are cleaved at +471/2 prior to cleavage at sites A_0 and A_1. The fact we see the same primer extension stop in *H. wingei* suggests that this is of functional importance. Cleavage sites have also been identified in the ETS sequences of vertebrate pre-rRNAs, and it is possible that the cleavages at +471 and +791 in the *S. cerevisiae* and *H. wingei* ETS sequences are homologous to these ETS cleavages (see Beltrame et al., 1994, for further discussion). The role played by the putative cleavage at +471/2 in subsequent processing reactions is, however, unclear at present.

The *H. wingei* U3 snoRNA has the potential to form 10 perfect base pairs with the ETS sequence around +790. Compared to the U3/ETS base pairing in *S. cerevisiae*, two compensatory mutations are seen in the base paired ETS/U3 sequences of *H. wingei*, providing very strong evidence for the evolutionary conservation of this interaction. In *S. cerevisiae*, compensatory mutations constructed between the ETS sequences and

FIGURE 7. A secondary structure model for the *H. wingei* U3 snoRNA. A model has been built on the basis of the structure established previously for *S. cerevisiae* U3A (Ségault et al., 1992). The nomenclature of stem-loop structures is that used previously (Ségault et al., 1992). The complete nucleotide sequence is that of *H. wingei*. Nucleotide substitutions compared to the *S. cerevisiae* U3A snoRNA are indicated by circled nucleotides. Δ indicates nucleotide deletion. Circled sequences marked with an asterisk correspond to insertions. For technical reasons, the large number of differences in the stem-loop structures 2 and 4 could not be represented. The corresponding region of the *S. cerevisiae* U3A snoRNA is represented in an inset below the *H. wingei* structure. The 15-nt sequence at the 3′ end of the *H. wingei* U3 snoRNA that could not be sequenced using our approaches is indicated by a discontinuous line. The position of the intron is shown by an arrow. Phylogenetically conserved Boxes, A′, A, B, and C, are indicated. The segment predicted to base pair with the 5′ ETS region of the pre-rRNA is colored in grey. The proposed base pair interaction between the *H. wingei* U3 snoRNA and the 5′ ETS region is shown in an inset on the left side of the figure. Differences in *S. cerevisiae* are indicated by circled nucleotides.

the U3 snoRNA have been used to demonstrate that direct base pairing to this sequence is required for U3 function (Beltrame & Tollervey, 1995). These analyses also demonstrated, however, that interaction over less than the full 10 bp, which could potentially occur, is sufficient to support pre-rRNA processing. This work shows that the full 10 bp of the interaction have been conserved in evolution and are therefore presumably all important for optimal function. The region beyond the U3-complementary sequences at +470/+790 is strikingly well conserved, making it very likely that other *trans*-acting factors also bind to the pre-rRNA in this vicinity. In *Xenopus*, a site of U3-dependent processing has been identified in the 5′ ETS (Kass et al., 1990; Mougey et al., 1993b). This site has also been shown to be the site of formation of the large complex that is visualized as the "terminal ball" in electron micrographs of nascent pre-rRNA molecules (Mougey et al., 1993a). Such terminal balls have been observed in many eukaryotes, including *S. cerevisiae* (Saffer & Miller, 1986). We envisage that the evolutionarily conserved sequences around the U3 binding site are important for the assembly of this complex.

MATERIALS AND METHODS

Strains and culture conditions

The *S. cerevisiae* FL100 (ATCC 28383) (Lacroute, 1968) and *H. wingei* (ATCC 14356) strains were grown in the standard YPD medium [1% (w/v) yeast extract, 1% (w/v) bactopeptone, and 2% (w/v) glucose] at 30 °C. For genomic DNA preparation, 10-mL cultures were grown for 12 h, whereas for RNA preparation, 500-mL cultures were grown for 10 h.

The *Escherichia coli* strain TG1 (Gibson, 1984) was used as the host strain plasmid transformation. *E. coli* was grown at 37 °C, in the Luria Broth medium complemented with 100 mg/L of ampicillin when necessary.

Preparation of yeast genomic DNA

The method used is an adaptation of that proposed by Cryer et al. (1975). A 10-mL stationary culture was centrifuged for 3 min at 13,000 rpm. The cells were ground with 0.45–0.5-mm diameter glass beads in the presence of 0.4 mL of Triton buffer [1% (v/v) Triton X-100, 1% (w/v) SDS, 100 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0] and 0.4 mL of a phenol/chloroform/isoamyl alcohol mixture (25/24/1). After alcohol precipitation, the DNA was incubated at 37 °C in the presence of RNase A (Boehringer) at a 20 μg/mL concentration, in TE buffer (0.1 mM EDTA, 10 mM Tris-HCl, pH 8.0). Another extraction with the phenol/chloroform/isoamyl alcohol mixture was performed. Genomic DNA was ethanol precipitated, dissolved in TE buffer, and stored at 4 °C.

PCR

Two pairs of oligonucleotide primers have been used for amplification of the U3 snoRNA gene. Couple 1 consists of oligonucleotide A, 5′-ATAAGTTAACTTGTCAGACTGCC-3′, complementary to the 3′-terminal sequence of the *S. cerevisiae* U3A snoRNA from position 475 to 490, and oligonucleotide B, 5′-AGCTCGTCGACGTACTTCA-3′, which contains the sequence at the 5′ end of the *S. cerevisiae* U3A snoRNA from position 1 to 14 (Myslinski et al., 1990). Couple 2 consists of oligonucleotide A and oligonucleotide C, 5′-AGCT CGTCGACGTACTTCAGTATGT-3′, which contains the sequence of the exon 1 of the *S. cerevisiae* U3A pre-snoRNA followed by the hexanucleotide at the 5′ end of the intron (Myslinski et al., 1990). A third oligonucleotide couple, 3, was used for amplification of the 5′ ETS region of the pre-rRNA. It consists of oligonucleotide D, 5′-GAAGGATTTGGTGG ATTACTAG-3′, which contains the sequence of the *S. cerevisiae* 5′ ETS region between positions 153 and 174 (Verbeet et al., 1984) and oligonucleotide E, 5′-CTACTGGCAGGAT CAACCAG-3′, which is complementary to the 20 nt at the 5′ end of the *S. cerevisiae* 18S rRNA (Bayev et al., 1980; Skryabin et al., 1984).

PCR was achieved on 1 μg of yeast genomic DNA, in 50 μL of 50 mM KCl, 2 mM $MgCl_2$, 0.2% (v/v) Tween-1 (Sigma), 20 mM Tris-HCl, pH 8.3, containing 0.2 mM of each deoxynucleotide triphosphate, 10 pmol of each primer, and 2 units of Taq DNA polymerase (Gibco-BRL). The reactions were performed in a Perkin Elmer Cetus thermal cycler with the following protocol: 30 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min, and polymerization at 72 °C for 2 min, followed by a 10-min incubation at 72 °C after the final cycle. PCR products were analyzed by electrophoresis on a 6% nondenaturing acrylamide gel in the TBE buffer (Sambrook et al., 1989).

Southern blot analysis

The genomic DNAs of *S. cerevisiae* and *H. wingei* were digested with the *Eco*R I, *Hind* III, and *Pst* I nucleases. The digestion products were fractionated by electrophoresis on a 0.8% agarose gel made up in the TBE buffer and transferred onto a nitrocellulose filter (Sambrook et al., 1989). The DNA fragment amplified from the *H. wingei* genomic DNA using the oligonucleotide couple 1 was labeled by nick translation with the Nick Translation Kit (Boehringer) and [α-32P]dATP (400 Ci/mmol) (Amersham) and used as a probe. The amplified DNA fragment generated with the same oligonucleotide couple 1 and the cloned *S. cerevisiae* U3A snoRNA (plasmid pVS1::snR17A) (Ségault et al., 1992) was also used as a probe, after nick translation. Hybridizations were performed for 12 h at 60 °C in phosphate buffer [0.5 M $NaHPO_4$, pH 7.2, 7% (w/v) SDS, 0.1% (w/v) BSA, 0.1 mM EDTA] and 300 ng of the labeled probe. After hybridization, the filters were washed twice for 30 min, in 6× SSC buffer (0.9 M NaCl, 0.09 M sodium citrate) containing 0.2% (w/v) SDS at 37 °C. Autoradiography was conducted at −80 °C with an intensifying screen.

Plasmids and recombinant DNA technology

U3 PCR products were gel purified as described above. Amplified DNA fragments were extracted from the gel by dialysis in elution buffer (Sambrook et al., 1989), followed by phenol extraction. After digestion with *Sal* I and *Hpa* I nucle-

ases, they were ligated into the pBluescript SK(+) vector (Stratagene) digested with the *Sal* I and *Sma* I nucleases. *E. coli* TG1 strain was transformed with the ligation mixture (Sambrook et al., 1989). The 5′ ETS PCR products were gel purified, using the same protocol, and inserted into the pMOS-Blue-T vector (Amersham) using the PCR Cloning Kit (Amersham).

DNA sequence determination and alignment

DNA nucleotide sequences were determined on recombinant phagemid DNA using the method of Tabor and Richardson (1987).

U3 sequence alignments were performed with the PILEUP program of GCG software, version 8.1 Unix (1995). Schematic representation of sequence similarities were made with the Alscript program (Barton, 1993). ETS alignments were made with Bestfit and RNA folding predictions were made using Mfold from GCG software, version 8.0-OpenVMS (1994).

EMBL Nucleotide Sequence Database accession numbers of the sequences established in this paper are as follows: U3Hw snoRNA gene, X91005; 5′ ETS region of *H. wingei*, X87403.

Preparation of yeast small RNAs

The method used is an adaptation of that described previously (Myslinski et al., 1990). Cells recovered after a 10-h culture were centrifuged, washed in 10 mM EDTA, 50 mM sodium acetate, pH 5.3 buffer, and ground with glass beads in 10 mM EDTA, 0.1% (w/v) SDS, 50 mM sodium acetate, pH 5.3, buffer, in the presence of an equal volume of phenol saturated with 50 mM sodium acetate, pH 5.3. The extracted RNA was ethanol precipitated and fractionated by centrifugation on a 15–30% sucrose gradient made up in 100 mM NaCl, 10 mM Tris-HCl, pH 7.6, buffer. Centrifugation was for 24 h in a SW28 rotor at 27,000 rpm and 4 °C. RNA with a sedimentation coefficient between 4S and 8S was ethanol precipitated and dissolved in water.

RNA sequence analysis

RNA nucleotide sequences were determined with reverse transcriptase using an adaptation of the method developed by Hu Qu et al. (1983). Two oligonucleotides complementary to the U3A snoRNA between positions 91 and 105 (G) and between positions 319 and 333 (A) were used for the *S. cerevisiae* U3A snoRNA, and two oligonucleotides complementary to the sequence of the U3Hw snoRNA between positions 157 and 172 (H) and between positions 369 and 383 (A) were used for the U3Hw snoRNA. The oligonucleotide primers were labeled with [γ-32P] ATP (3,000 Ci/mmol) (Amersham) and the T4 polynucleotide kinase (Sambrook et al., 1989). Primer annealing was achieved in 6 μL of buffer in the presence of 50 μg of RNA mixture. RNA and labeled primers were incubated for 10 min at 65 °C in a buffer of 40 mM KCl, 6 mM $MgCl_2$, 50 mM Tris-HCl, pH 8.3, and returned at room temperature. Reverse transcription was performed with 1 unit of Rous Sarcoma virus 2 reverse transcriptase (Amersham) for 30 min at 45 °C, in the presence of 0.12 mM of each dNTP, and with a dideoxynucleotide/deoxynucleotide ratio of 1/2. Reverse transcripts were fractionated on a 6% polyacrylamide sequencing gel (Sambrook et al., 1989).

Primer extension analysis

Primer extension was conducted as described previously (Beltrame & Tollervey, 1992) using a primer (CATGGCTT AATCTTTGAGAC) complementary to nt +34–+54 within the 18S rRNA sequence; this region is completely conserved between *S. cerevisiae* and *H. wingei*. The in vitro transcript of *S. cerevisiae* pre-rRNA was prepared from plasmid pT7rDNA (J. Morrissey & D. Tollervey, unpubl.). To construct this plasmid, the 5′ ETS was amplified by PCR using oligonucleotides that inserted a *Kpn* I site immediately 5′ to the transcription initiation site and a *Bam*H I site immediately 3′ to the *Hin*d III site within the 5′ ETS. This fragment was subcloned into pBS KS+ and verified by sequencing. The large *Hin*d III fragment, which contains the entire remaining pre-rRNA coding sequence, was then inserted into the *Hin*d III site. When linearized with *Bam*H I and transcribed with T7 RNA polymerase, this plasmid gives rise to a transcript containing the entire pre-rRNA sequence, together with 284 nt of the 3′ nontranscribed spacer. The in vitro transcript of *H. wingei* pre-rRNA was prepared from a plasmid containing the *Bam*H I–*Eco*R I fragment of the rDNA (Verbeet, 1982; Verbeet et al., 1984). When linearized with *Eco*R I, this generates a transcript that includes 475 nt of 5′ nontranscribed spacer and 1,599 nt of the 17S rRNA sequence, in addition to the entire 5′ ETS sequence.

ACKNOWLEDGMENTS

We thank H.A. Raué for support during the sequencing of the *H. wingei* ETS, A. Mougin and A. Grégoire for advice on RNA sequencing methods, and G. Beck for her participation in the Southern blotting experiments. F.B. was a fellow of the French Ministère de la Recherche et de l'Enseignement Supérieur. J.V. was the recipient of an EMBO and EU postdoctoral Fellowship. This work was funded by the Centre National de la Recherche Scientifique and the French Ministère de la Recherche et de l'Enseignement Supérieur.

Received October 4, 1995; returned for revision October 16, 1995; revised manuscript received January 15, 1996

REFERENCES

Bachellerie JP, Michot B, Raynal F. 1983. Recognition signals for mouse pre-rRNA processing. A potential role for U3 nucleolar RNA. *Mol Biol Rep* 9:79–86.

Barton GJ. 1993. ALSCRIPT: A tool to format multiple sequence alignments. *Protein Eng* 6:37–40.

Bayev AA, Georgiev OI, Hadjiolov AA, Kermekchiev MB, Nikolaev N, Skryabin KG, Zakharyev VM. 1980. The structure of the yeast ribosomal RNA genes. 2. The nucleotide sequence of the initiation site for ribosomal RNA transcription. *Nucleic Acids Res* 8: 4919–4926.

Beltrame M, Henry Y, Tollervey D. 1994. Mutational analysis of an essential binding site for the U3 snoRNA in the 5′ external transcribed spacer of yeast pre-rRNA. *Nucleic Acids Res* 22:4057–4065.

Beltrame M, Tollervey D. 1992. Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J* 11:1531–42.

Beltrame M, Tollervey D. 1995. Base-pairing between U3 and the pre-

ribosomal RNA is required for 18S rRNA synthesis. *EMBO J.* Forthcoming.
Branlant C, Krol A, Machatt MA, Pouyet J, Ebel JP, Edwards K, Kossel H. 1981. Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. *Nucleic Acids Res 9*:4303-4324.
Brimacombe R, Mitchell P, Osswald M, Stade K, Bochkariov D. 1993. Clustering of modified nucleotides at the functional center of bacterial ribosomal RNA. *FASEB J 7*:161-167.
Brow DA, Guthrie C. 1989. Splicing a spliceosomal RNA. *Nature 337*:14-15.
Calvet JP, Pederson T. 1981. Base-pairing interactions between small nuclear RNAs and nuclear RNA precursors as revealed by psoralen cross-linking in vivo. *Cell 26*:363-370.
Craig N, Kass S, Sollner-Webb B. 1987. Nucleotide sequence determining the first cleavage site in the processing of mouse precursor rRNA. *Proc Natl Acad Sci USA 84*:629-633.
Crouch RJ, Kanaya S, Earl PL. 1983. A model for the involvement of the small nucleolar RNA U3 in processing eukaryotic ribosomal RNA. *Mol Biol Rep 9*:75-78.
Cryer DR, Eccleshall R, Marmur J. 1975. *Methods in cell biology*. New York: Prescott DM.
Eichler DC, Craig N. 1994. Processing of eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol 49*:197-239.
Epstein P, Reddy R, Busch H. 1984. Multiple states of U3 RNA in Novikoff hepatoma nucleoli. *Biochemistry 23*:5421-5425.
Filipowicz W, Kiss T. 1993. Structure and function of nucleolar snRNPs. *Mol Biol Rep 18*:149-156.
Fournier MJ, Maxwell ES. 1993. The nucleolar snRNAs: Catching up with the spliceosomal snRNAs. *Trends Biochem Sci 18*:131-135.
Frendewey D, Barta I, Gillespie M, Potashkin J. 1990. *Schizosaccharomyces* U6 genes have a sequence within their introns that matches the B box consensus of tRNA internal promoters. *Nucleic Acids Res 18*:2025-2032.
Gibson TJ. 1984. Studies on the Epstein-Barr virus [thesis]. Cambridge, UK: Cambridge University.
Hughes JM, Ares M. 1991. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast preribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J 10*:4231-4239.
Hughes JM, Konings DA, Cesareni G. 1987. The yeast homologue of U3 snRNA. *EMBO J 6*:2145-2155.
Hu Qu L, Michot B, Bachellerie JP. 1983. Improved method for structure probing large RNAs: A rapid "heterologous" sequencing approach is coupled to the direct mapping of nuclease accessible sites. Application to the 5' terminal domain of eukaryotic 28S rRNA. *Nucleic Acids Res 11*:5903-5920.
Kass S, Tyc K, Steitz JA, Sollner-Webb B. 1990. The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. *Cell 60*:897-908.
Kreger-Van-Rij NJW. 1973. Endomycetales, Basidiomycetous yeasts, and related fungi. In: Ainsworth GC, Sparrow FK, Sussman AS, eds. *The fungi*. New York: Academic Press. pp 11-32.
Lacroute F. 1968. Regulation of pyrimidine biosynthesis in *Saccharomyces cerevisiae*. *J Bacteriol 95*:824-832.
Li HD, Zagorski J, Fournier MJ. 1990. Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in *Saccharomyces cerevisiae*. *Mol Cell Biol 10*:1145-1152.
Lygerou Z, Mitchell P, Petfalski E, Seraphin B, Tollervey D. 1994. The POP1 gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins. *Genes & Dev 8*:1423-1433.
Madhani HD, Guthrie C. 1994. Dynamic RNA-RNA interactions in the spliceosome. *Annu Rev Genet 28*:1-26.
Maser RL, Calvet JP. 1989. U3 small nuclear RNA can be psoralen-cross-linked in vivo to the 5' external transcribed spacer of preribosomal RNA. *Proc Natl Acad Sci USA 86*:6523-6527.
Maxwell ES, Fournier MJ. 1995. The small nucleolar RNA. *Annu Rev Biochem 67*:6523-6527.
Mougey EB, O'Reilly M, Osheim Y, Miller OLJ, Beyer A, Sollner-Webb B. 1993a. The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. *Genes & Dev 7*:1609-1619.
Mougey EB, Pape LK, Sollner-Webb B. 1993b. A U3 small nuclear ribonucleoprotein-requiring processing event in the 5' external transcribed spacer of *Xenopus* precursor rRNA. *Mol Cell Biol 13*:5990-5998.
Myslinski E, Ségault V, Branlant C. 1990. An intron in the genes for U3 small nucleolar RNAs of the yeast *Saccharomyces cerevisiae*. *Science 247*:1213-1216.
Parker KA, Steitz JA. 1987. Structural analysis of the human U3 ribonucleoprotein particle reveal a conserved sequence available for base pairing with pre-rRNA. *Mol Cell Biol 7*:2899-2913.
Peebles CL, Gegenheimer P, Abelson J. 1983. Precise excision of intervening sequences from precursor tRNAs by a membrane-associated yeast endonuclease. *Cell 44*:525-536.
Petes TD, Hereford LM, Skryabin KG. 1978. Characterization of two types of yeast ribosomal DNA gene. *J Bacteriol 134*:295-305.
Pikielny CW, Teem JL, Rosbash M. 1983. Evidence for the biochemical role of an internal sequence in yeast nuclear mRNA introns: Implications for U1 RNA and metazoan mRNA splicing. *Cell 34*:395-403.
Porter GL, Brennwald PJ, Holm KA, Wise JA. 1988. The sequence of U3 from *Schizosaccharomyces pombe* suggests structural divergence of this snRNA between metazoans and unicellular eukaryotes. *Nucleic Acids Res 16*:10131-10152.
Raué HA, Planta RJ. 1995. The pathway to maturity: Processing of ribosomal RNA in *Saccharomyces cerevisiae*. *Gene Expr*. Forthcoming.
Saffer LD, Miller OLJ. 1986. Electron microscopic study of *Saccharomyces cerevisiae* rDNA chromatin replication. Mol *Cell Biol 6*:1148-1157.
Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
Ségault V, Mougin A, Grégoire A, Banroques J, Branlant C. 1992. An experimental study of *Saccharomyces cerevisiae* U3 snRNA conformation in solution. *Nucleic Acids Res 20*:3443-3451.
Skryabin KG, Eldarov MA, Larionov VL, Bayev AA, Klootwijk J, De Regt VC, Veldman GM, Planta RJ, Georgiev OI, Hadjiolov AA. 1984. Structure and function of the nontranscribed spacer regions of yeast rDNA. *Nucleic Acids Res 12*:2955-2968.
Stroke IL, Weiner AM. 1989. The 5' end of U3 snRNA can be cross-linked in vivo to the external transcribed spacer of rat ribosomal RNA precursors. *J Mol Biol 210*:497-512.
Tabor S, Richardson CC. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc Natl Acad Sci USA 84*:4767-4771.
Tague BW, Gerbi SA. 1984. Processing of the large rRNA precursor: Two proposed categories of RNA-RNA interactions in eukaryotes. *J Mol Evol 20*:362-367.
Takahashi Y, Urushiyama S, Tani T, Ohshima Y. 1993. An mRNA-type intron is present in the *Rhodotorula hasegawae* U2 small nuclear RNA gene. *Mol Cell Biol 13*:5613-5619.
Tani T, Ohshima Y. 1989. The gene for the U6 small nuclear RNA in fission yeast has an intron. *Nature 337*:87-90.
Tani T, Ohshima Y. 1991. mRNA-type introns in U6 small nuclear RNA genes: Implications for the catalysis in pre-mRNA splicing. *Genes & Dev 5*:1022-1031.
Tiollais P, Galibert F, Boiron M. 1971. Evidence for the existence of several molecular species in the "45S fraction" of mammalian ribosomal precursor RNA. *Proc Natl Acad Sci USA 68*:1117-1120.
Tollervey D. 1987. A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J 6*:4169-4175.
Tollervey D, Lehtonen H, Carmo-Fonseca M, Hurt EC. 1991. The small nucleolar RNP protein NOP1 (fibrillarin) is required for pre-rRNA processing in yeast. *EMBO J 10*:573-583.
Tollervey D, Lehtonen H, Jansen R, Kern H, Hurt EC. 1993. Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell 72*:443-457.
Tyc K, Steitz JA. 1992. A new interaction between the mouse 5' external transcribed spacer of pre-rRNA and U3 snRNA detected by psoralen crosslinking. *Nucleic Acids Res 20*:5375-5382.
Van Nues RW, Rientjes JMJ, Morré SA, Mollee E, Planta RJ, Venema J, Raué HA. 1995. Evolutionarily conserved structural elements are critical for processing of internal transcribed spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA. *J Mol Biol 250*:24-36.
Van Nues RW, Rientjes JMJ, Van der Sande CAFM, Zerp SF, Sluiter

C, Venema J, Planta RJ, Raué HA. 1994. Separate structural elements within internal transcribed spacer 1 of *Saccharomyces cerevisiae* precursor ribosomal RNA direct the formation of 17S and 26S rRNA. *Nucleic Acids Res* 22:912-929.

Venema J, Henry Y, Tollervey D. 1995. Two distinct recognition signals define the site of endonucleolytic cleavage at the 5' end of yeast 18S rRNA. *EMBO J.* Forthcoming.

Venema J, Tollervey D. 1995. Processing of pre-ribosomal RNA in *Saccharomyces cerevisiae*. *Yeast*. Forthcoming.

Verbeet GM. 1982. Regulatory and structural sequences of the yeast ribosomal RNA operon [thesis]. Amsterdam: Vrije Universiteit.

Verbeet MP, Van Heerikhuizen H, Klootwijk J, Fontijn RD, Planta RJ. 1984. Evolution of yeast ribosomal DNA: Molecular cloning of the rDNA units of *Kluyveromyces lactis* and *Hansenula wingei* and their comparison with the rDNA units of other Saccharomycetoideae. *Mol Gen Genet 195*:116-125.

Yeh LCC, Lee JC. 1992. Structure analysis of the 5' external transcribed spacer of the precursor ribosomal RNA from *Saccharomyces cerevisiae*. *J Mol Biol 228*:827-839.

Zieve G, Penman S. 1976. Small RNA species of the HeLa cell: Metabolism and subcellular localization. *Cell 8*:19-31.

Publication n:°4 :

Une grande diversité des gènes du snoRNA U3 chez les levures. La fonction du snoRNA U3 s'accommode des grandes variations du domaine 3'

F. Brulé, R. Fournier,V. Ségault, A. Grégoire, A. Mougin et C. Branlant

soumis à Journal of Molecular Biology

A high diversity of U3 snoRNA genes in yeasts. U3 snoRNA function accommodates great variations of the 3' domain.

F. Brulé*, R. Fournier*, V. Ségault, A. Grégoire, A. Mougin and C. Branlant+.

Laboratoire d'Enzymologie et de Génie Génétique - URA CNRS 457

Faculté des Sciences

Bd des Aiguillettes - BP239

54506 Vandœuvre-lès-Nancy cedex, FRANCE

+ to whom correspondence should be sent.

Laboratoire d'Enzymologie et de Génie Génétique - URA CNRS 457

Faculté des Sciences

Bd des Aiguillettes - BP239

54506 Vandœuvre-lès-Nancy cedex, FRANCE

Tel (33) 83.91.20.91 - Fax (33) 83.91.20.93

E-mail : cbranlant@legg.u-nancy.fr

* The two authors contributed equally to this work.

Introduction

In eukaryotes, U3 snoRNA plays a crucial role in pre-ribosomal RNA maturation. The eukaryotic 18S, 5.8S and 25/28S ribosomal RNAs (rRNAs) are transcribed by RNA polymerase I as a single precursor molecule. A complex series of processing reactions are needed to eliminate the 5' and 3' external transcribed spacers (5' ETS and 3' ETS, respectively) and the internal transcribed spacers 1 and 2 (for review, Eicler & Craig, 1994; Raué & Planta, 1995). The primary event, A_0 cleavage within the 5' ETS spacer, was successfully reproduced in extracts of cultured mouse cells (Craig *et al.*, 1987) and in *Xenopus laevis* (Mougey *et al.*, 1993), and it was found to be impaired upon U3 snoRNA depletion (Kass *et al.*, 1990; Mougey *et al.*, 1993). In yeast, when U3 snoRNAs are depleted, the normal pre-rRNA intermediates in the pathway leading to the synthesis of the 17S rRNA are missing, which results in an under-accumulation of mature 17S rRNA (Hughes & Ares, 1991). A "U3 snoRNA-like" was recently found in the archaebacteria *Sulfolobus acidocaldarius* and was proposed to be required for 16S rRNA maturation (Potter *et al.*, 1995).

U3 snoRNAs are characterized by the presence of six highly phylogenetically conserved boxes, namely A', A, C', B, C and D (Wise & Weiner, 1980; Hughes *et al.*, 1987; Myslinski *et al.*, 1990; Tycowski *et al.*, 1994). Boxes D and C' are the counterparts of the Boxes D and C found in several UsnoRNAs (for review, Fournier & Maxwell, 1993). The four other boxes are unique to U3 snoRNA. Within the nucleolus, U3 snoRNA is found associated with at least 6 proteins and Boxes B, C, C' and D were proposed to be involved in recognition and binding of the U3 snoRNP proteins (Parker & Steitz, 1987; Lübben *et al.*, 1993; Harsthorne & Agabian, 1994). The two *Saccharomyces cerevisiae* U3A and U3B snoRNAs are longer (333 nt), than all other U3 snoRNAs studied (between 142 and 279 nt) (for review, Gu & Reddy, 1996). The additional sequence fold into an additional stem-loop structure, so that the 3'-terminal domain of *S. cerevisiae* U3 snoRNA has a cruciform structure (Ségault *et al.*, 1992), instead of the 3-helix structure found in the vertebrates U3 snoRNAs (Parker & Steitz, 1987; Jeppensen *et al.*, 1988). Yeast and vertebrate U3 snoRNAs also differ in the 5' domain, that forms a single stem-loop structure in vertebrate (Parker & Steitz, 1987; Jeppensen *et al.*, 1988) and a two stem-loop structure in yeast and plant U3 snoRNAs (Porter *et al.*, 1988; Marshallsay *et al.*, 1990; Ségault *et al.*, 1992). The U3 snoRNA 5'-terminal domain is supposed to be the functional domain. A 10 base-pair complementarity between the U3 snoRNA segment from positions 39 to 48 and the pre-rRNA 5' ETS region of *S. cerevisiae* is conserved by base-compensatory mutations in *Hansenula wingei* (Brulé *et al.*, 1996) and it was found to be required for production of the mature 17S rRNA in *S. cerevisiae* (Beltrame et Tollervey, 1995). Based on a phylogenetic comparison and on the results of a mutational analysis, Box A, was recently proposed to facilitate folding of a conserved pseudo-knot structure in the small subunit rRNA (Hughes, 1996).

We previously made the unexpected observation of the presence of an intron spliced in a spliceosome in the two genes that code for U3 snoRNA in *S. cerevisiae* (U3A and U3B genes) (Myslinski *et al.*, 1990). Introns from nuclear protein genes in yeasts are characterized by a high conservation of the 5' and branch site sequences (GTATGT and TACTAAC, respectively) (for review, Will *et al.*, 1995). Whereas, a canonical sequence is found at the 5'-end of the introns from the *S. cerevisiae* U3A and U3B snoRNA genes (U3A and U3B introns), an unusual GACTAAC sequence is present at the branch site (Myslinski *et al.*, 1990). Secondary structure analysis of the *S. cerevisiae* pre-U3A snoRNA revealed a highly folded structure, with a long central stem-loop structure and 5' and 3' stem-loop structures that include the 5' and 3' functional sequences, respectively (Mougin *et al.*, 1996).

Up to now, the presence of spliceosomal introns in genes that code for stable RNAs, seems to be restricted to yeasts and, more precisely, to yeast UsnRNA or UsnoRNA genes. Such introns were found in the U6 snRNA of *Schizosaccharomyces pombe* (Tani & Ohshima, 1989) and other *Schizosaccharomyces* species, *Rhodotorula hasegawae* and *Rhodosporium dacroïdeum* (Frendewey *et al.*, 1990; Reich & Wise, 1990; Tani & Ohshima, 1991) and in the U2 snRNA genes from *Rhodotorula hasegawae* (Takahashi *et al.*, 1993). This is surprising as introns spliced in a spliceosome are quite less frequent in yeasts than in other eukaryotes (Woolford, 1989). It should also be pointed out that, in spite of the large number of small nuclear RNAs characterized in *S. cerevisiae* (for review, Tollervey, 1987; Maxwell & Fournier, 1995), only the U3 snoRNA genes were found to contain an intron. Finally, the presence of introns in U3 snoRNA genes is not general in yeasts, as the *S. pombe* U3 snoRNA genes have no introns (Selinger *et al.*, 1992). *S. pombe* and *S. cerevisiae* belong to two different yeast subfamily, the Schizosacchraomycetoideae and Sacchromycetoideae subfamily, respectively (Kreger van Rij, 1987).

The presence of introns in U3 snoRNA genes may be a general property of the Saccharomycetoideae subfamily. Indeed, we previously found U3A and U3B snoRNA genes with introns in two other Saccharomyces species, namely *S. bayanus* and *S. uvarum* (Brulé *et al.*, 1995), and two additional U3 snoRNA genes with an intron, that are likely derived from the U3B snoRNA genes, were also detected in *S. uvarum*. Another member of the Saccharomycetoideae subfamily, *Hansenula wingei*, from the *Hansenula* genus, has a unique U3 snoRNA gene and it also contains an intron (Brulé *et al.*, 1996). However, this intron is short, it has a canonical TACTAAC sequence at the branch site and this gene codes for a U3 snoRNA different from the *S. cerevisiae* U3A and U3B snoRNAs (Brulé *et al.*, 1996). Altogether, this suggested a complex evolution of the yeast U3 snoRNA genes.

In the present study, to test whether the presence of intron in U3 snoRNA genes is a characteristic feature of the Saccharomycetoideae subfamily, we extended the analysis of U3 snoRNA genes to two other genus from this subfamily, *Pichia* and *Kluyveromyces*. The U3 snoRNA genes and/or the U3 snoRNA introns were cloned and sequenced by the PCR method. Surprisingly, in addition to the U3A snoRNA gene that was completely conserved in sequence (intron and exon sequences), we detected a series of very different U3 snoRNA genes. Both the

intron and exon sequences were different, and in *Pichia* and *Kluyveromyces* species U3 snoRNA genes without intron were detected. This raised the question of the functionality of these genes and we showed by a genetic test that, in spite of great variations of the 3'-terminal domain, the encoded U3 snoRNAs are functional in *S. cerevisiae*. This prompted us to test the importance of the 3'-terminal cruciform domain of the yeast U3 snoRNA for U3 snoRNA function and we showed that U3A snoRNA severely truncated in the 3'-terminal domain can function in *S. cerevisiae*.

The results obtained are discussed in terms of U3 snoRNA function, pre-snoRNA splicing and U3 snoRNA gene evolution.

Results

The U3A snoRNA intron is frequently encountered in the Saccharomycetoideae sub-family

Oligonucleotide O-1 with the 14-nt sequence of the U3A and U3B snoRNA 5'-terminus and oligonucleotide O-2 complementary to the phylogenetically conserved box A of U3 snoRNAs were used to amplify genomic DNAs in order to look for the presence of U3A and U3B snoRNA introns. A series of yeast species from the yeast Saccharomycetoideae sub-family were subjected to the analysis : a species of the *Saccharomyces* genus, *S. florentinus*, four species of the *Pichia* genus, *P. anomala*, *P. guilliermondii*, *P. pastoris* and *P. salictaria*, and five members of the *Kluyveromyces* genus, *K. marxianus* var. *bulgaricus*, *K. marxianus* var. *fragilis* and *K. marxianus* var. *lactis* and *K.delphensis* and another *Kluyveromyces* strain supposed to belong to the species *Kluyveromyces lactis*, but which has to be reclassified according to the results obtained. *S. cerevisiae* FL100 was used as a control.

As shown in Figure 1, amplification products with the same size as the amplification fragment of the *S. cerevisiae* U3A snoRNA gene were obtained for all the species tested, except *K. delphensis*. Sequence analysis of the amplified DNA fragments, after cloning in phage M13mp9, confirmed that all the species tested, except *K. delphensis*, contained an intron with a sequence identical to that of the intron from the *S. cerevisiae* U3A snoRNA gene. These introns were all designed as U3A introns. A slightly longer amplification product was obtained for *K. delphensis*. According to its nucleotide sequence, it corresponded to a 167-nt long intron (denoted U3Kd intron) (Fig. 2). As found for the *S. cerevisiae* U3A and U3B genes, the U3F intron has a GACTAAC sequence at the branch site, and it is located 9 nt upstream of the 3' terminal AG dinucleotide. However, the rest of the nucleotide sequence of the U3Kd intron is very different from that of the U3A and U3B introns.

In addition to the amplification product of the U3A intron, a smaller amplified DNA fragment was found for *S. florentinus* and for *P. salictaria*. According to their nucleotide sequences, both of them corresponded to intron sequences. The 101-nt long intron of *P. salictaria* (denoted intron U3Ps), had a GACTAAC sequence at the branch site, that is located 12 nt upstream of

the 3' terminal AG dinucleotide (Fig. 2). Its nucleotide sequence is also very different from that of the U3A and U3B introns. The smaller, 76-nt long intron, of *S. florentinus* (U3Sf intron) had a canonical TACTAAC sequence at the branch site, that is located 9 nt upstream the 3' terminal AG dinucleotide. We had to verify that the introns, that we denoted U3Sf, U3Kd and U3Ps were contained in U3 snoRNA coding sequences and this was the next step of our study.

A large diversity of U3 snoRNA genes in the Saccharomycetoideae sub-family

In order to verify that the U3Sf, U3Kd and U3Ps introns were contained in U3 snoRNA genes, the genomic regions of *S. florentinus*, *P. salictaria* and *K. delphensis* that contain U3 snoRNA genes were amplified with the oligonucleotide pair O-1 and O-3. Oligonucleotide O-3 is complementary to the phylogenetically conserved box D, located at the 3'end of U3 snoRNAs (Wise & Weiner, 1980). As a control, the experiment was extended to all the other strains for which intron of U3 snoRNA genes were studied as described above. As expected from the results of intron-amplification experiments, an amplification product corresponding to the U3A snoRNA gene was found for all the species tested, except *K. delphensis* (Fig. 3), and all the amplified U3A snoRNA genes had identical sequences. Hence, the intron and exon sequences of the U3A snoRNA genes are strictly conserved in the *Saccharomyces*, *Pichia* and *Kluyveromyces* genera.

In addition to the amplification product of the U3A snoRNA gene, a second amplification product that corresponded to a U3 snoRNA gene with the U3Ps intron, characterized above (U3Ps snoRNA gene), was found for *P. salictaria*. A unique U3 snoRNA gene was amplified for *K. delphensis* and it contained the U3Kd intron described above (U3Kd snoRNA gene) (Fig. 2 and 3). However, no amplification product containing the U3Sf intron, described above, was detected for *S. florentinus*. This intron may belong to a truncated U3 snoRNA pseudogene. Unexpectedly, two amplification products were detected for *P. anomala, P. guillermondii, P. pastoris, K. marxianus* var. *lactis* and *K. marxianus* var. *fragilis* and the K. lactis strain to be identified (Fig. 3). The second amplification product was sequenced for *K. marxianus* var. *fragilis* and the *K. lactis* strain and was found to correspond to U3 snoRNA gene without intron ($U3Kl_{403}$ and U3Kf snoRNA genes, respectively) (Fig. 3).

Alignment of the U3 snoRNA coding sequences, from all the amplified U3 snoRNA genes (Fig. 4), revealed a strong conservation of the 5' terminal region including Box C'. Box B and its upstream 15 nucleotide sequence were also highly conserved, as well as Box C and the 15 nt sequence upstream of Box D (Fig. 4A). Due to the method used, the 15 nt at the two termini of the coding regions, that served as binding sites for primer oligonucleotides, could not be determined. Long insertions were detected in the $U3Kl_{403}$ and U3Kd coding regions (Fig. 4A). None of them contained sequences that define spliceosomal introns, namely the 5' and 3' boxes and the branch site sequence. Hence, all the introns, that we found in U3 snoRNA genes of the yeast species studied, were located at the same position within the coding sequence (between positions 14 and 15 upstream of Box A) (Fig. 4A).

The various genes detected in *Pichia* and *Kluyveromyces* are expressed

The question was to know whether the various U3 snoRNA genes characterized in addition to the U3A snoRNA were expressed. To answer to this question, we used two approaches: direct sequence analysis of U3 snoRNA in total RNA extract from the various species with oligonucleotide O-3 as the primer and reverse transcriptase (Fig. 5). On the other hand, U3 snoRNAs from total RNA extracts were studied by Northern blot analysis using oligonucleotide O-3 and O-2 as the probes (Fig. 5). The results obtained show that, in spite of its high degree of conservation, in the growth conditions, the U3A snoRNA gene was expressed in none of the Pichia and Kluyveromyces species studied : the U3Ps snoRNA was expressed in *P. salictaria* and the U3 snoRNA genes without intron in *K. marxianus* var. *fragilis* and *K. marxianus* var. *lactis*. As expected, the single U3Kd gene is expressed in *K. delphensis*.

Hence, it was interesting to look for the possible secondary structure of the U3 snoRNAs encoded by these genes.

We previously studied the secondary structure of the *S. cerevisiae* U3A snoRNA in solution (Ségault *et al.*, 1992). The 3' domain of the U3A snoRNA was found to fold into a cruciform structure and the 5' domain to form two stem-loop structures, denoted 1a and 1b. However, our recent study of the purified *S. cerevisiae* U3A snoRNP (A. Grégoire *et al.*, unpublished data) suggests that in the presence of the U3 snoRNP proteins, the helix 5 of the cruciform structure may be extended up to the 3' end of the molecule (Fig. 6). In this extended helix 5, boxes C' and D would form an irregular stem. And the stem-loop structure 1b would be replaced by a smaller stem-loop structure 1b'. The present results bring argument in favour of this new secondary structure model of the yeast U3 snoRNA. Indeed, for all the U3 snoRNA coding sequences, that we established including that of the *H. wingei* U3Hw snoRNA (Brulé *et al.*, 1996), a more or less irregular stem-loop structure 1a can be formed (Fig. 6), but, the stability of the stem-loop structure 1b is not conserved, especially for the U3Kl403 and U3Kd snoRNA coding sequences. In contrast, the possibility to form the extended helix 5 and the stem-loop structure 1b' is conserved for all the coding sequences. The highly variable sequence downstream from box C' constitutes a bulged loop (Fig. 6).

Variations of the nucleotide sequence are greater in the 3'-terminal domain. However, a cruciform structure similar to that formed in the *S. cerevisiae* U3A snoRNA can be proposed for the U3 snoRNA encoded by the U3Ps and U3Kl403 genes. In this latter, as a consequence of additional sequences, the stem-loop structure 2 is highly extended. Cruciform structures can be formed with the U3Kl403 and U3Kd coding sequences, but, they are very irregular with lateral stem-loop structures in the helixes 2 and 4 (Fig. 6).

Based on these results, the expressed gene from the various U3 snoRNA genes should be functional in their respective species.

U3 snoRNAs with very different 3' domains are functional in *S. cerevisiae*

Based on the above results, one could ask the question whether each yeast genus need a specific U3 snoRNA for growth or if the U3 snoRNA expressed in a given yeast genus or a given yeast

species can function in another yeast genus or in another yeast species. To test for this possibility, we looked for the functionality in *S. cerevisiae* of some of the U3 snoRNA that are expressed in *Pichia* or *Kluyveromyces*. For this purpose, we used a genetic test. This test is based on the utilization of the *S. cerevisiae* strain JH84 built by Hughes and Ares (1991). In this strain, the U3B snoRNA gene has been disrupted and the U3A snoRNA gene is placed under the control of the Gal upstream activation sequence (UASgal10) (Guarente *et al.*, 1982). To grow on glucose this strain has to be complemented by a plasmid carrying an active U3 snoRNA gene. To complement the *S. cerevisiae* strain JH84 with an active U3 snoRNA genes, we inserted the 3.3 kb *Eco*RI-*PstI* fragment of the *S. cerevisiae* genome, that contains the U3A snoRNA gene in the centromeric plasmid pASZ11 (Stotz & Linder, 1990). The cloned fragment includes the exon and intron sequences, and the transcription initiation and termination signals (Myslinski *et al.*, 1990). As can be seen in Figure 7, *S. cerevisiae* JH84, transformed with the resulting pASZ11::pU3A plasmid, grew at the same rate on glucose as on galactose. Then, the *S. cerevisiae* pre-U3A snoRNA coding sequence from plasmid pASZ11::U3A was replaced by the U3Kf snoRNA coding sequence without intron. As shown in Figure 7, the U3Kf snoRNA coding sequence allowed growth on glucose and the growth rate was similar to that with the *S. cerevisiae* pre-U3A snoRNA coding sequence. Hence, this gene without intron that codes for a very different U3 snoRNA is functional in *S. cerevisiae*. The fact that *S. cerevisiae* does not need to have an intron in the U3 snoRNA gene for growth, is also demonstrated by the complementation of the JH84 strain by a snoRNA U3A gene with a deleted intron sequence (Fig. 7 and the recent results of J. Hughes (1996)).

Using the same approach, we also checked whether the pre-U3Ps snoRNA and pre-U3Kd snoRNA (intron + exons) were functional in *S. cerevisiae*. Growth of the recombinant JH84 strains on glucose (Fig. 7) showed that both U3Ps and U3Kd snoRNAs are functional in *S. cerevisiae*. Hence, U3 snoRNAs with highly divergent 3'-terminal domain are active in *S. cerevisiae*. This raised the question of the importance of this 3' terminal domain for U3 snoRNA function. For more information into this problem, we mutated the *S. cerevisiae* U3A snoRNA 3'-terminal domain. On the one hand, we deleted the coding sequence of the stem-loop structure 2, on this other hand, we deleted that for the stem-loop structure 4. These deletions made as previously described (Ségault *et al.*, 1992) were designed in order to preserve boxes B and C (Fig. 7). The variant coding sequences generated by site-directed mutagenesis (pU3AΔ2 and pU3AΔ4) were substituted to the wild-type coding sequence in plasmid pASZ11-pU3A. As shown in Figure 7, they both ensured growth of the JH84 strain on glucose. Hence, the stem-loop structures 2 and 4 are not needed for U3 snoRNA function.

Discussion

U3 snoRNA genes with an intron are found in all members of the yeast Saccharomycetoideae subfamily.

Taking into account our previous studies (Brulé *et al.*, 1995, 1996) and the present study that altogether encompass 14 yeast species from 4 different genera *Saccharomyces*, *Kluyveromyces*, *Pichia* and *Hansenula*, it appears that at least one U3 snoRNA gene with an intron is present in each of the studied species. This raise the possibility of a biological importance of this intron.

However in several *Pichia* and *Kluyveromyces* species, the U3 snoRNA gene without intron is expressed and the present results and the results of J. Hughes (1996) show that the presence of an intron in U3 snoRNA gene is not needed for growth of *S. cerevisiae* as the JH84 *S. cerevisiae* strain can be complemented with U3 snoRNA genes without intron. Thus, either the introns of the Saccharomycetoideae U3 snoRNAs are a sort of selfish DNA, inserted accidently in the course of evolution, and that provides no advantage to the cell. Or on the contrary, there is some selective advantage for yeasts that carry U3 snoRNA genes with an intron, and this advantage cannot be detected by a simple study of growth rate in laboratory conditions. In connection to this problem, it should be pointed out that several snoRNAs encoded by introns have been found recently, as well in vertebrates, as in yeast (for review, Bachellerie *et al.*, 1995) and that yeast introns coding for snoRNA were not found to be essential (for review, Maxwell & Fournier, 1995). An argument against the hypothesis that introns of yeast U3 snoRNA genes may code snoRNAs, is the absence in the U3 snoRNA introns of the characteristic features generally found in introns encoding snoRNA, namely a sequence complementarity with the 17S or 26S rRNA and the conserved boxes C and D (Bachellerie *et al.*, 1995; Kiss-Làslo *et al.*, 1996). Hence, the reason for the high degree of conservation of the U3A intron in Saccharomycetoideae still has to be elucidated.

A strong conservation of the U3A snoRNA gene together with a wide variety of other U3 snoRNA genes in Saccharomycetoideae.

Among the 14 Saccharomycetoideae species, studied up to now (Brulé *et al.*, 1995, 1996 and the present study), only two, namely *H. wingei* and *K. delphensis*, were found to be devoided of U3A snoRNA gene. The entire pre-U3A snoRNA sequence was established for 5 of the 12 species : *S. cerevisiae* FL100 (Myslinski *et al.*, 1990), GY5 (Hughes *et al.*, 1987), *S. florentinus*, *P. salictaria*, *K. fragilis* and *K. lactis* (this study). All these sequences were found to be identical, except the *S. cerevisiae* GY5 pre-U3A snoRNA sequence that showed 4 base substitutions, 3 in the intron and 1 in exon 2 (Hughes *et al.*, 1987; Myslinski *et al.*, 1990; Brulé *et al.*, 1996). One allele of this variant U3A snoRNA gene is present in *Saccharomyces bayanus* together with the usual U3A snoRNA gene. In spite of this high degree of conservation the U3A snoRNA was not found to be expressed in the Pichia and Kluyveromyces species that we studied.

In contrast, with this unexpectedly high degree of conservation of the U3A snoRNA gene, the additional U3 snoRNA gene that we sequenced for *P. salictaria*, *K. fragilis* and *K. lactis*, as well as, the unique *K. delphensis* and in *H. wingei* snoRNA genes are very different from the *S. cerevisiae* U3A and U3B snoRNA genes. Both the intron and the exon sequences differ strongly.

For a more clear understanding of the relationships between all these U3 snoRNA genes, we looked for the degree of similarity of their coding sequence (Table 2). The pair of sequences with the highest degree of homology was found to be the U3A and U3B snoRNA sequence (91% of similarity). This suggests that the U3B snoRNA gene may be the results of a duplication of the U3A snoRNA gene after *Saccharomyces* genes separation. Interestingly, another pair shows a high percentage of similarity, the U3Ps and U3Hw coding sequences (83% of similarity). Hence, the unique U3Hw snoRNA gene, that we found previously in *H. wingei*, and the second U3Ps snoRNA gene that we detected in *P. salictaria*, are likely to have a common ancestor. Interestingly, we previously found, that the *H. wingei* U3 snoRNA displays two differences in the 10 nt sequence from position 39 to 48, that base-pairs with the 5'-ETS region of the pre-rRNA, and these two differences were found to be compensate by two mutations, in the 5'-ETS complementary region (Brulé *et al.*, 1996). Only one of the two mutations in the segment that base-pairs with the 5'-ETS region is found in the U3Ps snoRNA (A -> C). As the U3Ps snoRNA is expressed in this species. It will be interesting to know whether a base mutation in the *P. salictaria* 5'-ETS region compensates for the A_{40} -> C mutation in the U3Ps snoRNA. The unique U3Kd snoRNA found in *K. delphensis* is more similar to the U3A and U3B snoRNAs than to the other U3 snoRNAs (79 and 75% of similarity, respectively). Hence, these 3 snoRNAs may have a common ancestor. The U3Kl snoRNA from *K. lactis*, that is encoded by a gene without intron displays more similarity with the U3Ps and U3Hw snoRNAs, than with the other snoRNAs (79 and 72%, respectively). Finally, the U3Kf snoRNA from *K. fragilis*, that is also encoded by a gene without intron is the more divergent of all the identified RNAs. However, its best score of similarity is with the U3Ps and U3Kd snoRNAs (70 and 67%, respectively).

Based on these comparison there seems to be two families of U3 snoRNAs in Saccharomycetoideae, the U3A, U3B and U3Kd genes, on the one hand, and the U3Ps, U3Hw, U3Kl and U3Kf genes, on the other hand (Fig. 4B).

The intron filiation fits with the established snoRNA filiation

Due to their rapid divergence, it is difficult to establish a filiation of intron sequences. However, if only the more conserved 3'-terminal sequences, starting at the 5'-end of the branch site sequence are compared, some relationship can be established (Fig. 8). For instance, except for one C to T substitution, the 3'-terminal sequence of the U3Kd intron is identical to that of the U3A intron (Fig. 9), which is in agreement with the filiation found at the level of the exon sequences. In addition, the long U3Kd intron can be folded in a structure that is very similar to that we established for the U3A (Mougin *et al.*, 1996) and U3B (Brulé *et al.*, 1995 and A.

Grégoire *et al.*, unpublished results) introns. This structure includes a long central stem-loop structure flanked by a smaller stem-loop structure. The 5' sequence, the AG 3'-terminal dinucleotide and the branch site sequence have exactly the same conformation, as in the pre-U3A and pre-U3B snoRNAs (Fig. 8).

The snoRNA genes from the second family have no intron or short introns and the 3'-terminal regions of the U3Hw and U3Ps introns displays significant sequence homologies (Fig. 2). Although one has a consensus TACTAAC sequence at the branch site, the other a GACTAAC sequence. Our experimental analysis of the pre-U3Ps snoRNA secondary structure (Fig. 8) showed the presence of two stem-loop structures. As proposed for the U3Kd intron, the 5' sequence, the 3'-terminal AG dinucleotide and the branch site sequence, have the same conformation as in the U3A and U3B introns. Indeed, formation of a 5' stem-loop structure is imposed in all the identified pre-U3 snoRNAs, including the pre-U3Hw snoRNA, by a conserved complementarity of exon 1 and the intron 5'-terminal sequence. The presence of the 3'-terminal AG dinucleotide at the border of a bulge loop in all pre-U3 snoRNAs is also largely due to the sequence conservation of Box A. However, the possibility to form the stem at the basis of the stem-loop structure 1a', in all pre-U3 snoRNA is due to base compensatory mutations. Furthermore, in all U3 snoRNA introns, the branch site sequence is free of base-pairing and according to our experimental study of the relationship between pre-U3A snoRNA secondary structure and *in vitro* splicing efficiency, this single-stranded state of the branch site sequence is probably the results of a selection pressure for splicing efficiency.

A short functional 5' domain and a large structural 3' domain in yeast U3 snoRNA

The present data bring phylogenetic arguments in favour of a new secondary structure model for yeast U3 snoRNAs. In this model, the 5' domain is reduced in size (64 nt), and large portions of the domain have been proposed to base-pair with the pre-rRNA (box A and the segment from positions 39 to 49). The 3' cruciform domain is increased in size by an extension of helix 5 containing boxes C' and D. The new conformation of boxes C' and D is more similar to that of boxes C' and D in vertebrates U3 snoRNAs (Parker & Steitz, 1987; Jeppensen *et al.*, 1988) and also of the counterparts of boxes C' and D in intron encoded snoRNAs (Tycowski *et al.*, 1993). Our deletion experiments of entire stem/loop structures from the 3' domain and our observation that yeast U3 snoRNAs with completely different stem/loop 2, 3 and 4 are functional, suggest that stem/loop structures 2, 3 and 4 essentially, have a structural role and that none of them are essential. This reinforced the idea that the main function of the U3 snoRNA 3'-terminal domain is to bind the U3 snoRNP proteins, and that the binding sites of this proteins is constituted by the evolutionarely conserved sequences C', B, C and D.

Materials and Methods

Strains and growth conditions

The yeast strains listed in Table 1 were used in this study.

For genomic DNA or total RNA preparation, the yeast strains were grown for 12 hours in 10 ml of standard YPD medium (1% yeast extract, 1% bactopeptone and 2% glucose), at 30°C.

The *S. cerevisiae* strain JH84 (Mata snr17a.Gald:*URA*3 snr17b::*LEU*2 his3 ade2 can1) (Hughes & Ares, 1991), generously provided by J. Hughes, was used to test for the functionality of U3 snoRNA genes. It was grown on YPG medium (1% yeast extract, 1% bactopeptone and 2% galactose) or on YPD medium.

The *Escherichia coli* strain TG1 (Gibson, 1984) was used as the host strain for production of recombinant DNA. *E. coli* TG1 was grown at 37°C, in the Luria Broth medium, with 100 µg/ml of ampicillin added when necessary.

Yeast genomic DNA preparation and amplification

The method used is an adaptation of that described by Cryer *et al.* (1975). A 10 ml stationary culture was centrifuged for 3 min at 13000 rpm. The cells were ground with 0.45-0.5 µm diameter glass beads in the presence of 0.4 ml of Triton Buffer [1% Triton X-100, 1% sodium dodecyl sulfate (SDS), 100 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl pH 8.0] and 0.4 ml of a phenol/chloroform/isoamylic alcohol mixture (25/24/1). After alcohol precipitation, the DNA was incubated for 30 min at 37°C, in the presence of RNase A (Boehringer) at a 20 µg/ml concentration, in TE buffer (0.1 mM EDTA, 10 mM Tris-HCl pH 8.0). Another extraction with the phenol/chloroform/isoamyl alcohol mixture was performed. The genomic DNA was ethanol precipitated, dissolved in TE buffer and stored at 4°C.

An oligonucleotide pair (O-1, O-2) was designed to amplify the intron sequence of U3 snoRNA genes. Primer O-1: 5'AGCTCgtatgtGTACTTCA3' has the same sequence as the 14 nt at the 5' end of the *S. cerevisiae* U3A and U3B snoRNAs (Myslinski *et al.*, 1990), the sequence in lower cases correspond to a *Sal* I restriction site.

Primer O-2: 5'ATAAggatccTATAGAAATGATCCTA3', is complementary to the phylogenetically conserved Box A of the U3A snoRNA (underlined) (Myslinski *et al.*, 1990) and contains a sequence that generates a *Bam*H I restriction site (lower cases). The oligonucleotide pair (O-1, O-3) was used to amplify the entire U3 pre-snoRNA coding sequences. Primer O-3: 5' ATAAGTTTTaacgttTCAGACTGCC 3', contains a 15-nt sequence complementary to the 3'-end of the *S. cerevisiae* U3A and U3B snoRNAs (Myslinski *et al.*, 1990), and a sequence that generates an *Hpa* I restriction site (lower cases).

Polymerase Chain Reactions were performed on 1 µg of yeast genomic DNA in 50 µl of 50 mM KCl, 2 mM $MgCl_2$, 0.2% Tween-1 (Sigma), 20 mM Tris-HCl pH 8.3, containing 0.2 mM of each deoxynucleotide triphosphate, 10 pmol of each primer and 2 units of Taq DNA polymerase (Gibco-BRL). A Perkin Elmer Cetus thermal cycler was used with the following protocol : 30 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min and

polymerization at 72°C for 2 min, followed by a 10 min incubation at 72°C after the final cycle. PCR products were analyzed by electrophoresis on a 6% non-denaturing acrylamide gel in 1XTBE buffer (Sambrook *et al.*, 1989). For production of the U3 snoRNA or pre-U3 snoRNA by *in vitro* transcription, DNA amplifications were performed with a primer oligonucleotide that generated a T7 RNA polymerase transcription initiation site.

The oligonucleotide O-4 : 5'AAATTtaatacgactcactataGTCGACGTACTTCATAGGAT3', complementary to the U3 snoRNA sequence from the 5' end to position 1 to 20 (underlined) was used for DNA amplification in the view to produce mature U3 snoRNA by *in vitro* transcription and the primer oligonucleotide O-5:

5'AATTtaatacgactcactataGTCGACGTACTTCA*GTATGT*3' complementary to the pre-U3 snoRNA sequence from the 5' end to position *+6* of intron was used. The T7 RNA polymerase promoter is represented in lower cases in these two oligonucleotides.

Cloning and sequencing of the amplified DNA fragments

Amplified DNA fragments corresponding U3 snoRNA introns were purified by electrophoresis on 8% agarose gel in TBE buffer (Sambrook *et al.*, 1989). After elution, the amplified DNA fragment were phenol extracted, and ethanol precipitated, and digested with the *Sal* I and *Bam*H I nucleases. The digestion product was ligated into the phage M13mp9 DNA (Yanish-Perron *et al.*, 1985) cleaved with the same enzymes. The *E. coli* TG1 strain was transformed with the ligation mixture (Sambrook *et al.*, 1989). Several recombinant M13mp9 single-stranded DNAs were prepared for each ligation, and their inserts sequenced as described by Tabor and Richardson (1987). The amplified fragment corresponding to entire U3 snoRNA genes were purified in the same way and directly cloned in plasmid pMOSBlue-T using the commercial kit of Amersham (RPN 1719).

Nucleotide sequence alignments

Sequence alignments were perfomed with the PILEUP program of GCG software Version 8.1 Unix (1995). Shematic representation of sequence similarities were made with the Alscript program (Barton, 1993).

The GenBank database accession numbers of the sequences established in this paper are as follows: Z78423, Z78424, Z78425, Z78426, Z78427, Z78428, Z78429, Z78429, Z78430, Z78431, Z78432, Z78433, Z78434, Z78435, Z78436, Z78437, Z78438.

Construction of plasmid pASZ11::pU3A and its derivatives

The 3.3 Kb *Eco*R I-*Pst* I fragment from plasmid pFL1::snR17A (Myslinski *et al.*, 1990) that contains the *S. cerevisiae* U3A snoRNA gene including the transcription start and termination signals was inserted into the centromeric plasmid pASZ11 (Stotz and Linder, 1990). The recombinant plasmid was designated as pASZ11::pU3A.

The following strategy was used for substitution of the *S. cerevisaie* pre-U3A snoRNA coding sequence by the pre-U3Ps, pre-U3Kd ou U3Kf coding sequence of other genes. The 1.5 kb

Hind III fragment that contains the U3A snoRNA gene was subcloned in pBluescript. In the recombinant plasmid obtained, the *Hpa* I site located 250 nucleotides downstream of the pre-U3A snoRNA coding sequence was eliminated by conversion into a *Apa* I site by site-directed mutagenesis.Then, the *Sal* I *Hpa* I from this plasmid was substituted by the fragment generated by the digestion with the *Sal* I and *Hpa* I nuclease of the amplification products corresponding to the U3Ps, U3Kd and U3Kf snoRNA genes. Finally, the 1.5 kb *Kpn* I-*Pst* I fragment from recombinant pBluescript plasmid generated in this way was cloned in pASZ11 plasmid.

The following strategy was used to produce pASZ11::U3A derivatives carrying mutated pre-U3A snoRNA coding sequences. Site-directed mutageneses of the pre-U3A snoRNA coding sequence were achieved on the recombinant phage M13mp9T7::U3A, that contains the 0.95 kb *Sal* I-*Bam*H I fragment from plasmid pFL1::U3A (Myslinski *et al.*, 1990), were made with the method of Kramer *et al.* (1984).

Oligonucleotide O-6 : 5'TTCCTATAGAAATGATCCTATGAAGTACGTCGACCTATAG3' complementary to the pre-U3A snoRNA sequences between positions -7 to 14 and 15 to 33 was used to delete the intron coding sequence. Oligonucleotide O-7:

5'GGGTCAAGATCATCGCGCAAAGCAAAATTTTTATTC3' complementary to the pre-U3A snoRNA sequences between positions 181 to 199 and 250 to 266 was used to delete the stem-loop structure 2 sequence and oligonucleotide O-8:

5'CTACAAATGCAACGGCGCCGGTTTCTCACTCTGGGG3' complementary to the pre-U3A snoRNA sequence between positions 104 to 122 and 199 to 214 to delete the stem-loop structure 4 coding sequence. After site-directed mutagenesis, the 0.95 *Sal* I-*Eco*R I fragment of plasmid pASZ11::U3A containing the wild-type coding sequence was subsituted by the 0.95 Kb *Sal* I-*Eco*R I fragment from the mutated M13mp9T7::U3A phage DNAs containing the variant coding sequence. The resulting plasmids were designated as pASZ11::U3AΔ2 for deletion of stem-loop 2 and pASZ11::U3AΔ4 for the deletion of stem-loop structure 4.

Direct sequence analysis of yeast U3 snoRNAs

The method used for preparation of a fraction of yeast RNA enriched in U3 snoRNA is an adaptation of that previously described (Myslinski *et al.*, 1990). Cells recovered after a 10h culture were centrifuged, washed in 10 mM EDTA, 50 mM sodium acetate pH 5.3 buffer and ground with glass beads in 10 mM EDTA, 0.1% (w/v) SDS, 50 mM sodium acetate pH 5.3 buffer, in the presence of an equal volume of phenol saturated with 50 mM sodium acetate pH 5.3. The extracted RNA was ethanol precipitated and fractionated by centrifugation on a 15%-30% sucrose gradient made up in 100 mM NaCl, 10 mM Tris-HCl pH 7.6 buffer. Centrifugation was for 24 hours in a SW28 rotor at 27 000 rpm and 4°C. RNAs with a sedimentation coefficient between 4S and 8S were ethanol precipitated, and dissolved in water, and used for direct sequence analysis as previously described (Qu *et al.*, 1983).

Northern blot analysis

Total RNA was extracted from yeast cells essentially as described by Domdey *et al* (1984). Cells were slowly cooled on ice 300 µl of ice cold 50 mM AcONa pH 5.0, 10 mM EDTA, 1% SDS, 300 µl of acid-washed glass-beads and 300 µl of phenol equilibrated with 50 mM AcONa pH 5.0, 10 mM EDTA, were added. The microfuge tube was vortexed twice for 30 seconds. The mixture was rapidly chilled on an ice-ethanol bath until apparition of phenol crystals. After centrifugation, the aqueous phase was treated once with phenol, once with a phenol-chloroform mixture and RNA was ethanol precipitated in the presence of 300 mM AcONa pH 5.0. After centrifugation, the RNA pellet was washed, dried and dissolved in water. Concentration was estimated by measuring the absorption at 260nm.

About 10 µg of total RNA prepared as described above was subjected to a 6% palyacrylamide (38:2) gelelectrophoresis in presence of 8M urea and 0.5X TBE buffer. After 2 hours electrophoresis at 15 V/cm, RNA was electroblotted (Biorad Trans Blot SD SemiDry Transfer Cell) on Biodyne B transfer membrane (0.45 µm, Pall) in 0.5X TBE, for 30 minutes at 3 mA/cm^2 and fixed onto the membrane by Ultra-Violet cross-linking for 5 minutes.

Pre-hybridization was at 37°C, in 6X SSC, 0.1% SDS, 5X Denhardt's reagent, 5mM EDTA, 10 µg/ml of sonicated and denatured calf thymus DNA, for 3 hours. Hybridization with the ^{32}P-5'-labelled oligonucleotide probe (4 ng/ml)was for 16 hours at 37°C, in the same buffer as for the hybridization, filters were first rinsed at room temperature in 6X SSC, 0.1% SDS and washed twice for 30 minutes in the same buffer pre-warmed at 37°C. Dried filters were autoradiographied with X-Ray films.

***In vitro* transcription.**

In vitro transcription on lineraized plasmid DNA were carried out in 125 µl of 10 mM NaCl, 10 mM $MgCl_2$, 10 mM DTT, 40 mM Tris-HCl, pH 8, containing 5 µg of linearized plasmid, 62.5 nmol of each ribonucleoside triphosphate, 125 units of RNase Guard^{TM} (Pharmacia), 140 units of T7 RNA polymerase (Amersham). When the DNA used for *in vitro* transcription was prepared by the PCR method, transcription were carried out in 20 µl of 1 mM NaCl, 6 mM $MgCl_2$, 5 mM DTT, 40 mM Tris-HCl pH 8.4, containing 250 ng of amplified fragment, 1.25 nmol of each ribonucleoside triphosphate, 20 units of RNase Guard^{TM} (Pharmacia), 140 units of T7 RNA polymerase (Amersham). In both cases, after 2 hours of incubation at 37°C, nucleic acids were phenol extracted and ethanol precipitated. Template DNA was digested with 5 units of RNase-free DNase I (Boehringer) in 250 µl of 5 mM $MgCl_2$, 100 mM sodium acetate, pH 5 buffer, for 30 min at 37°C. After phenol extraction and ethanol precipitation, the RNA (about 10 mg) was dissolved in 120 µl of sterile water; 2 µl of this solution was used for each chemical reaction or enzymatic digestion.

Chemical modifications.

Chemical modifications were performed either under nondenaturing conditions or under semi-denaturing conditions (Ehresmann *et al.*, 1987; Ségault *et al.*, 1992). Each modification was

carried out on about 200 ng of *in vitro* synthesized RNA in the presence of 5 µg of a commercial yeast tRNA mixture (Boehringer). Prior to the chemical reaction, the RNA was dissolved in 100 µl of the appropriate buffer and pre-incubated 10 min at 20°C (temperature used for the modification reactions). Dimethylsulfate (DMS) modifications under nondenaturing conditions were made in 50 mM KCl, 10 mM $MgCl_2$, 100 mM sodium cacodylate, pH 7.5 buffer and under semi-denaturing conditions in 2 mM EDTA, 100 mM sodium cacodylate pH 7.5 buffer, with either 0.75 µl of pure DMS or 0.75 µl of a DMS solution (1/1 (v/v) DMS/EtOH) per assay. The incubation was for 15 min, and the reaction was stopped by ethanol precipitation. The RNA was washed with 70% ethanol and dissolved in 14 µl of H_2O. 7 µl of this solution were used for a reverse transcriptase elongation assay. 1-cyclohexyl-3-(2-morpholino ethyl)-carbodiimide-metho-p-toluene sulfonate (CMCT) modifications under nondenaturing conditions were made in 50 mM KCl, 10 mM $MgCl_2$, 100 mM sodium borate, pH 8 buffer, and under semi-denaturing conditions, in the same buffer except that 2 mM EDTA were used instead of 10 mM $MgCl_2$ and 50 mM KCl, with 1.05 or 2.1 mg of CMCT per 100 µl assay. Incubation was for 30 min. At the end of the reaction, the same procedure was used as for DMS modification.

Enzymatic cleavages

As for chemical modifications, the enzymatic digestions were performed on T7 RNA transcripts mixed with 5 µg of commercial tRNA. For V1 RNase (Pharmacia).digestion, RNA was pre-incubated 10 min at 0°C in 10 ml of reaction buffer (350 mM KCl, 10 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.5). Incubation was for 2, 7 or 10 min with an enzyme/RNA ratio of 1.2 unit/mg. The reaction was stopped by adding 5 µl of 100 mM EDTA, followed by phenol extraction. RNAs digested for 2 and 7 min were mixed. After ethanol precipitation, the RNA was washed with 70% ethanol, dried and dissolved in 7µl water and stored at -80°C.For digestion with S1 nuclease (Pharmacia), RNA was pre-incubated 10 min at 20°C in 10 µl of reaction buffer (50 mM KCl, 10 mM $MgCl_2$, 1 mM $ZnCl_2$, 25 mM Na acetate, pH 4.5). Incubation was for 12 or 25 min at 20°C with 2 units of enzyme. The reaction was stopped by adding 2 ml of 100 mM EDTA, extracted with phenol, and the RNA was prepared for primer extension analysis as above.

Primer extension analysis

Positions of enzymatic cleavages and chemical modifications in the U3 snoRNA transcripts were analyzed by primer extension with reverse transcriptase as previously described (Qu *et al.*, 1983). Oligonucleotides were used as primer after 5'-end labelling. It was annealed to 7 ml aliquots of the RNA samples described above, for 10 min at 65°C in 40 mM KCl, 6 mM $MgCl_2$, 50 mM Tris-HCl, pH 8.3 buffer, and the solution was cooled down to room temperature. Reverse transcription was performed with 1 unit of Avian myeloblastosis virus reverse transcriptase (Life Science), for 30 min at 45°C, in the presence of 125 mM of each dNTP. To prepare sequencing ladders of unmodified RNA, dideoxynucleotide:

deoxynucleotide mixtures (in a 1/2 ratio) were used. Reverse transcripts were fractionated by electrophoresis on a 7% polyacrylamide sequencing gel (Sambrook *et al.*, 1989).

Test of U3 snoRNA functionality

Strain JH84 was transformed with plasmid pASZ11 and plasmid pASZ11::U3A or one of its derivative, using the LiCl transformation method (Ito *et al.*, 1983). Selection of transformed strain was made on minimum medium (YNB 0,67%, Galactose 2%, Histidine 20mg/l). Recombinant JH84 yeast strains were grown on YPG medium (1% yeast extract, 1% bactopeptone and 2% galactose). Repression of the U3A snoRNA gene was achieved as described by Hughes and Ares (1991) : freshly grown galactose culture were either washed once in glucose medium and used to inoculate fresh pre-warmed glucose medium, or diluted directly into fresh pre-warmed galactose medium as controls, to give an optical density of about 0.5 at the zero time-point. The cultures were subsequently diluted in order after 5h, 10h, 15h and 36h of culture to maintain exponential growth. The OD_{600nm} measured before each dilution was used to determine the growth rate. An aliquot of cells was collected at each dilution for analysis of U3 snoRNA content in toatl RNA extract, by Northern blot analysis.

Legends to tables

Table I : The yeast strains used in this study

ATCC : American Type Cellule Collection
CBS : Centraalbureau voor Schimmelcultures
STV : Sciences et Techniques du vin

Table II : Similarities of the U3 snoRNA genes coding sequences

The percentages of similarity were deduced from the alignment shown in Figure 4 using the Gap software.

Legends to figures

Figure 1 : PCR amplification of introns from yeast U3 snoRNA genes

The genomic DNAs were prepared as described in Materials and Methods. The intron DNA sequences were amplified by the PCR method with the oligonucleotide pair O-1, O-2. The amplified fragment were fractionated by electrophoresis on a 10% polyacrylamide gel in TBE buffer and visualised by ethidium bromide. A control amplification, experiment was made on plasmid pFL1::snR17A. A mixture of the pBR322 Ava II digestion products and of the λ phage DNA Hind III digestion products was used as size marker. The origin of the genomic DNA is indicated above each Lane.

Figure 2 : The nucleotide sequence of the introns from U3 snoRNA genes characterized in this study

In panel A, the nucleotide sequences established for the unique intron found in *K. delphensis* (intron U3Kd) the two introns found in *P. salictaria* (U3A and U3Ps) and in the two introns found in *S. florentinus* (U3A and U3Sf) are shown. The intron from the *S. cerevisiae* U3B snoRNA gene and of the *H. wingei* U3Hw snoRNA gene are shown for comparison. The 5' sequence, the branch site sequence and the 3' terminal AG dinucleotide are boxed.

In panel B, the similarity of the 3' terminal sequences of introns U3A, U3B and U3Kd is shown, as well as, the similarity of the 3' terminal sequences of introns U3Ps and U3Hw.

Figure 3 : PCR amplification of yeast U3 snoRNA genes.

The genomic DNAs were prepared as described in Materials and Methods. The U3 snoRNA genes were amplified by the PCR method with the oligonucleotide pair O-1, O-3. The amplified DNA fragments were fractionated by electrophoresis on a 10% polyacrylamide gel in TBE buffer and visualised by ethidium bromide. A control amplification, experiment was made on plasmid pFL1::snR17A. The origin of the genomic DNA is indicated above each Lane. Plasmid pBR322 digested with the *Hae* III nuclease was used as size marker.

Figure 4 : Comparative analysis of the coding sequences from the amplified U3 snoRNA genes.

The coding sequences established for the *K. delphensis*, *K. marxianus* var. *fragilis*, *K. marxianus* var. *fragilis*, *P. salictaria* U3 snoRNA genes characterized in the present study are aligned with the coding sequence of the *S. cerevisiae* U3A and U3B snoRNA genes (Myslinski *et al.*, 1990) and of the *H. wingei* U3Hw snoRNA gene (Brulé *et al.*, 1996). The alignment was made with the PileUp program. Positions were at least 4 nucleotides were identical in the 7 sequences aligned are surrounded by grey color. The phylogenetically conserved boxes A', A, C', B, C and D (Wise and Weiner, 1980; Hughes *et al.*, 1987; Myslinski *et al.*, 1990; Tycowski *et al.*, 1994) are boxed. Based on the alignment in Panel A, a dendrogram was established (Panel B).

Figure 5A : Direct sequence analysis of the U3 snoRNA from *P. salictaria* and *K. delphensis* (panel A)

Total RNA from *P. salictaria* and *K. delphensis* was prepared as described in Materials and Methods and direct sequence analysis was made with reverse transcriptase and oligonucleotide O-3 as the primer.

Figure 5B : Northern blot analysis of U3 snoRNA in *S. cerevisiae*, *K. marxianus* var. *bulgaricus*, *K. marxianus* var. *fragilis*, *K. marxianus* var. *lactis* and *K. delphensis* (panel B)

Total RNA was prepared as described in Materials and Methods. The 5' end labeled oligonucleotide O-3 was used as the probe.

Figure 6 : Proposed secondary structure models for the yeast U3 snoRNAs

Panel A shows the nucleotide sequence conservation in the various U3 snoRNA coding sequences that we characterized. The secondary structure model used is a revised version of the *S. cerevisiae* U3A snoRNA secondary structure based on a study of the *S. cerevisiae* U3 snoRNP (A. Grégoire *et al.*, unpublished results). The differences in the previously published model are shown in the insert. Nucleotides which are conserved in all the yeast U3 snoRNA, are represented by black characters, the variable ones are represented in grey. The phylogenetically conserved boxes A', A, C', B, C and D are boxed, as well as the 10 nt sequence complementary to the 5' ETS region of the pre-ribosomal RNA (Beltrame and Tollervey, 1992; Brulé *et al.*, 1996).

Models of secondary structures for the *P. salictaria* U3Ps snoRNA, *K. delphensis* U3Kd snoRNA, *K. marxianus* var. *lactis* U3Kl snoRNA and *K. marxianus* var. *fragilis* U3Kf snoRNA adapted from that proposed for the *S. cerevisiae* U3A snoRNA are shown in Panel B. Secondary structure of the 3' domain of *P. salictaria* U3Ps snoRNA was studied experimentally on the pre-U3Ps snoRNA and the results of chemical modifications and enzymatic digestions are shown in Panel B. Circled nucleotides were modified by DMS or CMCT. The intensity of the modification is indicated by the color of the circle, white for low, grey for medium, black for high. Phosphodiester bound cleaved by S1 nuclease are shown by arrows surmounted of a circle and those cleaved by VI RNase by arrows surmounted of a square, and the color of the circles or the squares indicate the intensity of the cleavage with the same code as above. The phylogenetically conserved boxes A', A, C', B, C and D, and the 10 nt sequence complementary to the 5' ETS region are boxed. The position of the intron is shown by a black triangle.

Figure 7 : Test of the functionality of heterologous U3 snoRNA genes and variant U3A snoRNA genes in *S. cerevisiae*

The *S. cerevisiae* JH84 strain (Hughes and Ares, 1991) was used as the host strain. Plasmid pASZ11::U3A contains the wild-type pre-U3A snoRNA coding sequence. This coding sequence was replaced by the coding sequence of the U3Kd, U3Kf and U3Ps genes in plasmid pASZ11::U3Kd, pASZ11::U3Kf and pASZ11::U3Ps, respectively, and by truncated pre-U3A snoRNA sequence in plasmids pASZ11::U3AΔIVS (intron deleted), pASZ11::U3AΔ4, pASZ11::U3AΔ2. The truncation in U3AΔ2 and U3AΔ4 snoRNA are schematically represented in Panel C.

Growth on glucose solid medium of the JH84 untransformed strain and of the JH84 strain transformed with the various plasmids is shown in Panel A. For the same series of strains, the evolution of the OD600nm of cultures made in glucose liquid medium, with serial dilutions as described in Materials and Methods, is illustrated in Panel B.

Figure 8 : Secondary structure model for the introns of the U3Ps, U3Kd and U3Hw snoRNA genes

The secondary structure of the *P. salictaria* U3Ps intron was studied experimentally. The results of enzymatic cleavages and chemical modifications are schematically represented with the same rules as in Figure 5. The 5' sequence, branch site sequence and 3' terminal AG dinucleotide are boxed.

Bibliography

Bachellerie, J. P., Michot, B., Nicoloso, M., Balakin, A., Ni, J. & Fournier, M. J. (1995). Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. *Trends Biochem. Sci.* **20**, 261-264.

Barton, G. J. (1993). ALSCRIPT a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37-40.

Beltrame, M. & Tollervey, D. (1992). Identification and functional analysis of two U3 binding sites on yeast pre- ribosomal RNA. *EMBO J* **11**, 1531-1542.

Beltrame, M. & Tollervey, D. (1995). Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. *EMBO J* **14**, 4350-4356.

Brulé, F., Grégoire, A., Ségault, V., Mougin, A. & Branlant, C. (1995). Secondary structure conservation of the U3 small nucleolar RNA introns in *Saccharomyces. C. R. Acad. Sci. Paris, Life Sciences.* **318**, 1197-1206.

Brulé, F., Venema, J., Ségault, V., Tollervey, D. & Branlant, C. (1996). The yeast *Hansenula wingei* U3 snoRNA gene contains an intron and its coding sequence co-evolved with the 5' ETS region of the pre-ribosomal RNA. *RNA* **2**, 183-197.

Craig, N., Kass, S. & Sollner-Webb, B. (1987). Nucleotide sequence determining the first cleavage site in the processing of mouse precursor rRNA. *Proc Natl Acad Sci U S A* **84**, 629-633.

Domdey, H., Apostol, B., Lin, R. J., Newman, A., Brody, E. & Abelson, J. (1984). Lariat structures are *in vivo* intermediates in yeast pre-mRNA splicing. *Cell* **39**, 611-21.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P. & Ehresmann, B. (1987). Probing the structure of RNAs in solution. *Nucleic Acids Res* **15**, 9109-9128.

Eichler, D. C. & Craig, N. (1994). Processing of eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol* **49**, 197-239.

Fournier, M. J. & Maxwell, E. S. (1993). The nucleolar snRNAs: catching up with the spliceosomal snRNAs. *TIBS* **18**, 131-135.

Frendewey, D., Barta, I., Gillespie, M. & Potashkin, J. (1990). Schizosaccharomyces U6 genes have a sequence within their introns that matches the B box consensus of tRNA internal promoters. *Nucleic Acids Res* **18**, 2025-2032.

Gibson, T. J. (1984). Studies on the Epstein Barr virus genome. PhD thesis, Cambridge University, England.

Gu, J. & Reddy, R. (1996). Small RNA database. *24* **24**, 73-75.

Guarente, L., Yocum, R. R. & Gofford, P. (1982). A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. *Proc. Natl. Acad. Sci. USA.* **79**, 7410-7414.

Hartshorne, T. & Agabian, N. (1994). A common core structure for U3 small nucleolar RNAs. *Nucleic Acids Res* **22**, 3354-3364.

Hughes, J. M. & Ares, M. (1991). Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J* **10**, 4231-4239.

Hughes, J. M. X. (1996). Functional base-pairing interaction between highly conserved elements of U3 small nucleolar RNA and the small ribosomal subunit. *1996* **259**, 645-654.

Hughes, J. M., Konings, D. A. & Cesareni, G. (1987). The yeast homologue of U3 snRNA. *EMBO J* **6**, 2145-2155.

Ito, H., Fukuda, Y., Murata, K. & Kimura, A. (1983). Transformation of intact yeast cells treated with alcali cations. *J. Bacteriol.* **153**, 163-168.

Jeppesen, C., Stebbins-Boaz, B. & Gerbi, S. A. (1988). Nucleotide sequence determination and secondary structure of *Xenopus* U3 snRNA. *Nucleic Acids Res* **16**, 2127-2148.

Kass, S., Tyc, K., Steitz, J. A. & Sollner-Webb, B. (1990). The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. *Cell* **60**, 897-908.

Kiss-Làszlo, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M. & Kiss, T. (1996). Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar. RNAs. *85* **85**, 1077-1088.

Kramer, B., Kramer, W. & Fritz, H.J. (1984). Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*. *Cell* **38**, 879-887.

Kreger van Rij, N.J.W. (1987) In *The yeasts*. Second edition. Rose, A.H. & Harrison, J.S. eds. Academic Press, London.

Lübben, B., Marshallsay, C., Rottmann, N. & Luhrmann, R. (1993). Isolation of U3 snoRNP from CHO cells: a novel 55 kDa protein binds to the central part of U3 snoRNA. *Nucleic Acids Res* **21**, 5377-5385.

Marshallsay, C., Kiss, T. & Filipowicz, W. (1990). Amplification of plant U3 and U6 snRNA gene sequences using primers specific for an upstream promoter element and conserved intragenic regions. *Nucleic Acids Res.***18**, 3459-3466.

Maxwell, E. S. & Fournier, M. J. (1995). The small nucleolar RNA. *Annu. Rev. Biochem.* **67**, 6523-6527.

Mougey, E. B., Pape, L. K. & Sollner-Webb, B. (1993). A U3 small nuclear ribonucleoprotein-requiring processing event in the 5' external transcribed spacer of Xenopus precursor rRNA. *Mol Cell Biol* **13**, 5990-5998.

Mougin, A., Grégoire, A., Banroques, J., Ségault, V., Fournier, R., Brulé, F., Chevrier-Miller, M. & Branlant, C. (1996). The secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *RNA* (in press).

Myslinski, E., Ségault, V. & Branlant, C. (1990). An intron in the genes for U3 small nucleolar RNAs of the yeast Saccharomyces cerevisiae. *Science* **247**, 1213-1216.

Parker, K. A. & Steitz, J. A. (1987). Structural analysis of the human U3 ribonucleoprotein particle reveal a conserved sequence available for base pairing with pre-rRNA. *Mol Cell Biol* **7**, 2899-913.

Porter, G. L., Brennwald, P. J., Holm, K. A. & Wise, J. A. (1988). The sequence of U3 from *Schizosaccharomyces pombe* suggests structural divergence of this snRNA between metazoans and unicellular eukaryotes. *Nucleic Acids Res* **16**, 10131-10152.

Potter, S., Durovic, P. & Dennis, P. P. (1995). Ribosomal RNA precursor processing by a eukaryotic U3 small nucleolar RNA-like molecule in an archaeon. *Science* **268**, 1056-1060.

Qu, L. H., Michot, B. & Bachellerie, J. P. (1983). Improved methods for structure probing in large RNAs : a rapid "heterologous" sequencing approach is coupled is coupled to the direct mapping of nuclease accessible sites. Application to the 5' terminal domain of eukaryotic 28S rRNA. *Nucleic Acids Res.* **11**, 5903-5920.

Raué, H. A. & Planta, R. J. (1995). The pathway to maturity: Processing of ribosomal RNA in *Saccharomyces cerevisiae. Gene expression* **5**, 71-77.

Reich, C. & Wise, J. A. (1990). Evolutionary origin of the U6 small nuclear RNA intron. *Mol Cell Biol* **10**, 5548-5552.

Sambrook J., Fritsch E.F. et Maniatis T. (1989). In *Molecular Cloning. A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, New York.

Ségault, V., Mougin, A., Grégoire, A., Banroques, J. & Branlant, C. (1992). An experimental study of Saccharomyces cerevisiae U3 snRNA conformation in solution. *Nucleic Acids Res* **20**, 3443-3451.

Selinger, D. A., Porter, G. L., Brennwald, P. J. & Wise, J. A. (1992). The two similarly expressed genes encoding U3 snRNA in Schizosaccharomyces pombe lack introns. *Mol Biol Evol* **9**, 297-308.

Stotz, A. & Linder, P. (1990). The ADE2 gene from *Saccharomyces cerevisiae*: sequence and new vectors. *Gene.* **95**, 91-98.

Tabor, S. & Richardson, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84**, 4767-4771.

Takahashi, Y., Urushiyama, S., Tani, T. & Ohshima, Y. (1993). An mRNA-type intron is present in the Rhodotorula hasegawae U2 small nuclear RNA gene. *Mol Cell Biol* **13**, 5613-5619.

Tani, T. & Ohshima, Y. (1989). The gene for the U6 small nuclear RNA in fission yeast has an intron. *Nature* **337**, 87-90.

Tani, T. & Ohshima, Y. (1991). mRNA-type introns in U6 small nuclear RNA genes: implications for the catalysis in pre-mRNA splicing. *Genes Dev* **5**, 1022-1031.

Tollervey, D. (1987). High level of complexity of small nuclear RNAs in fungi and plants. *J Mol Biol* **196**, 355-61.

Tycowski, K. T., Shu, M. D. & Steitz, J. A. (1993). A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. *Genes Dev.* **7**, 1176-1190.

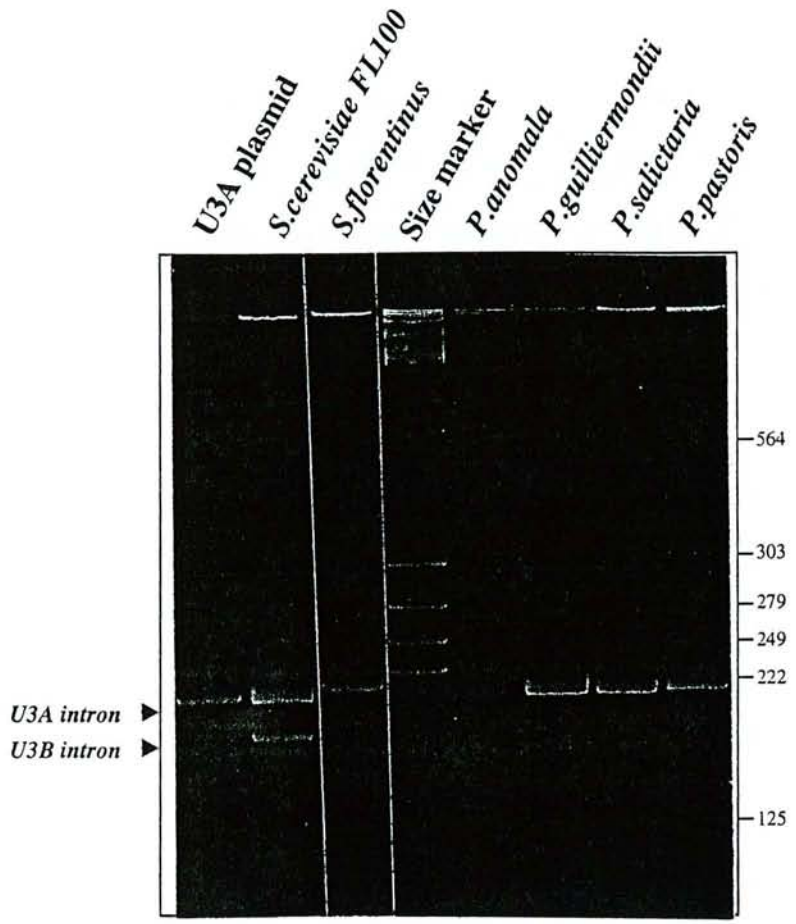
Tycowski, K. T., Shu, M. D. & Steitz, J. A. (1994). Requirement for intron-encoded U22 small nucleolar RNA in 18S ribosomal RNA maturation. *Science* **266**, 1558-1561.

Will, C. L., Fabrizio, P. & Lührmann, R. (1995). Nuclear pre-mRNA splicing. *Nucleic Acids Mol. Biol.* **9**, 342-372. F. Eckstein & D. M.J. Lilley. Springer-Verlag Berlin Heidelberg.

Wise, J. A. & Weiner, A. M. (1980). Dictyostelium small nuclear RNA D2 is homologous to rat nucleolar RNA U3 and is encoded by a dispersed multigene family. *Cell* **22**, 109-118.

Woolford, J. L. (1989). Nuclear pre-mRNA splicing in yeast. *Yeast* **5**, 439-457.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.


U3A plasmid
S.cerevisiae FL100
S.florentinus
Size marker
P.anomala
P.guilliermondii
P.salictaria
P.pastoris
564
303
279
249
222
125
U3A intron
U3B intron

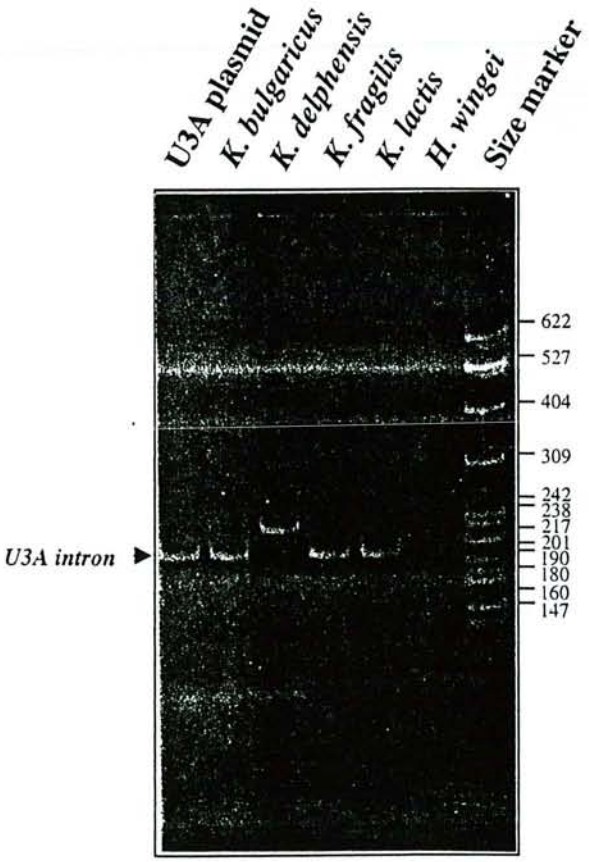

U3A plasmid
K. bulgaricus
K. delphensis
K. fragilis
K. lactis
H. wingei
Size marker
622
527
404
309
242
238
217
201
190
180
160
147
U3A intron

Figure 1

A

Intron U3A GTATGTAATA TACCCCAAAC ATTTTACCCA CAAAAAACCA GGATTTGAAA ACTATAGCAT
CTAAAAGTCT TAGGTACTAG AGTTTTCATT TCGGAGCAGG CTTTTTGAAA AATTTAATTC
AACCATTGCA GCAGCTTTTG ACTAACACAT TCTACAG **157 nt**

Intron U3B GTATGTTTTA TACCATATAC TTTATTAGGA ATATAACAAA GCATACCCAA TAATTAGGCA
ATGCGATTGT CGTATTCAAC AACCATCTTC TATTTCACCA GCTTCAGGTT TTGACTAACA
CATTCAACAG **130 nt**

Intron U3Kd GTATGTACAT TACCAATAAA TTTGAGTTTC TTTAATCGAA AATGCCGGGA AATATATGTC
AAAATCGGGT TTATAAATTT TTGATTTGAC ACAGTGGAAA TCTACAACGA TTAGAGATAA
TTTTGCACTT CTTTTTATGG AATATTTTTG ACTAACACAT TCTATAG **167 nt**

Intron U3Ps GTATGTGTTT ACCCATAAGG TTTTTACAAA ACCTTAAGCT TCTCTGTCTT TGAGAAGCAA
CACTGGATCA GGAATCGGAG GACTAACGCC TTTTGCAACA G **101 nt**

Intron U3Hw GTATGTGCTTACCCTAATAC ATTTGAGTGG TTTCAAACAG AATACTAACG ATGCCTTGTG
CCTATAG **67 nt**

Intron U3Sf GTATGTGTTT GCCCTATATG CTTTCCAAAA GCATAACAAC AACCAGTAAT TCAGAGTTTA
CTAACACCCT CCGAAG **76 nt**

B

Intron U3A GACTAACACATTCTACAG
Intron U3B GACTAACACATTCAACAG
Intron U3Kd GACTAACACATTCTATAG

Intron U3Ps GACTAAC...GCCT TTGCA.ACAG
Intron U3Hw TACTAACGATGCCT GTGCCTATAG

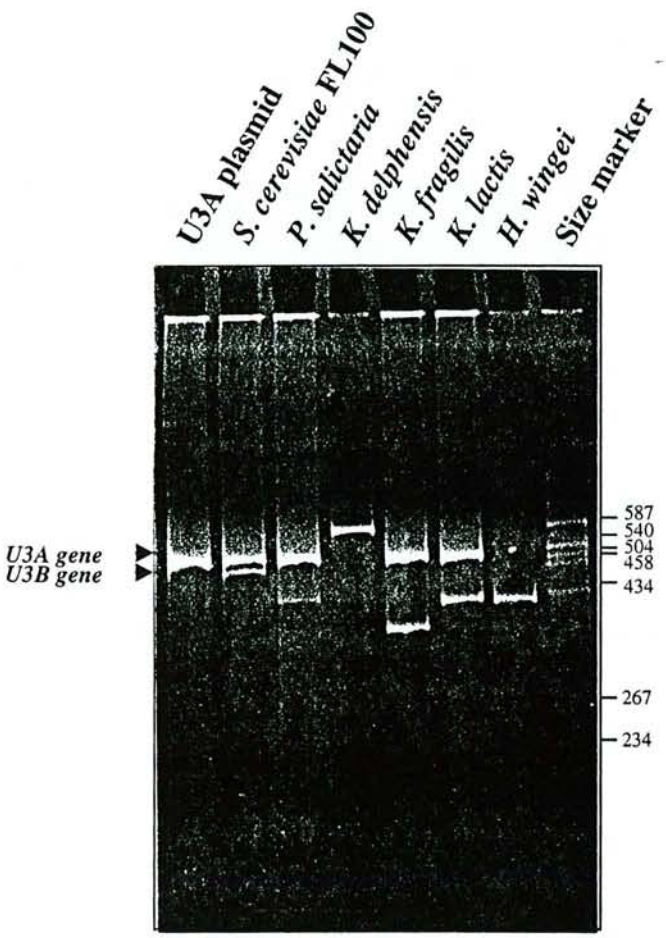
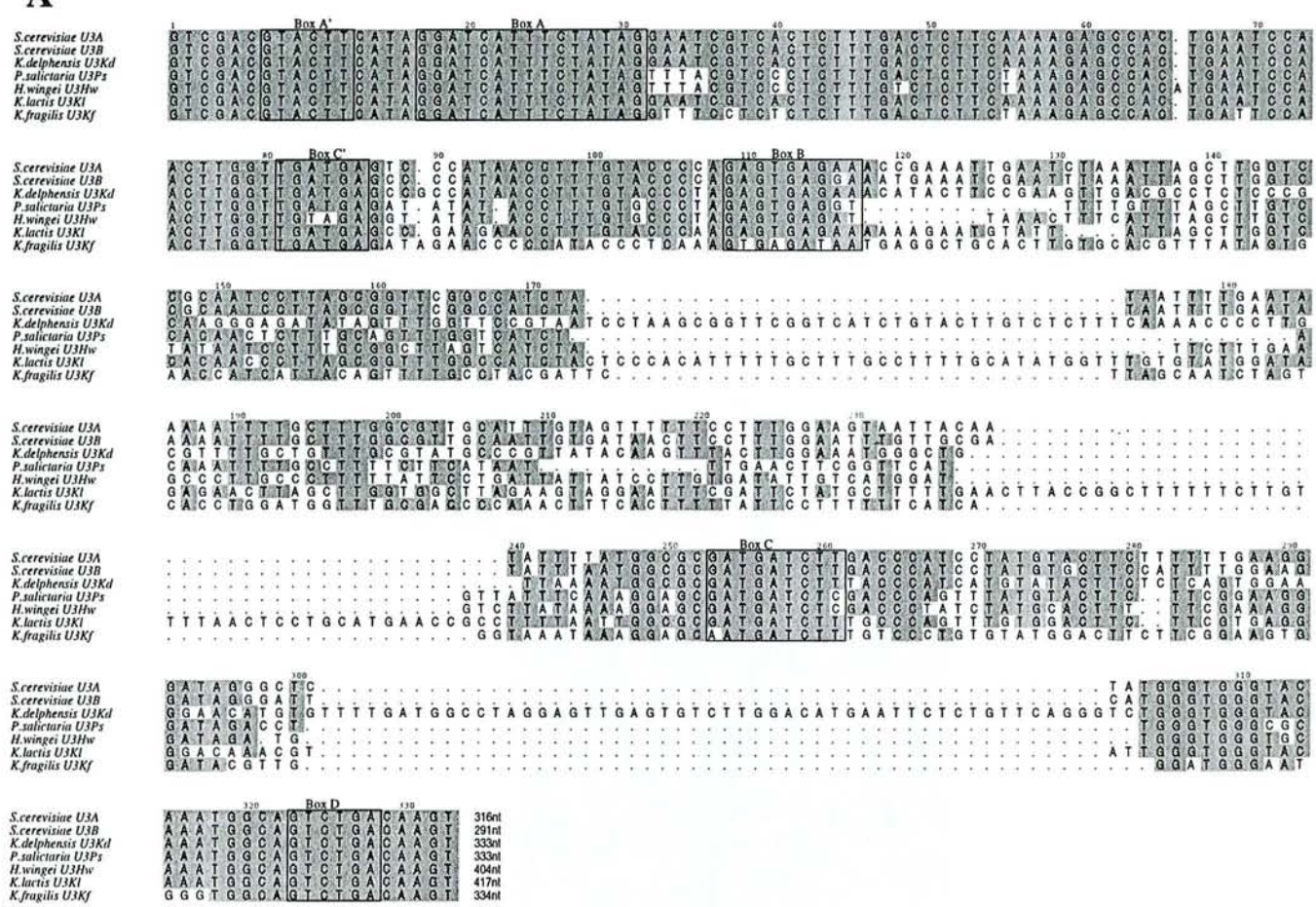
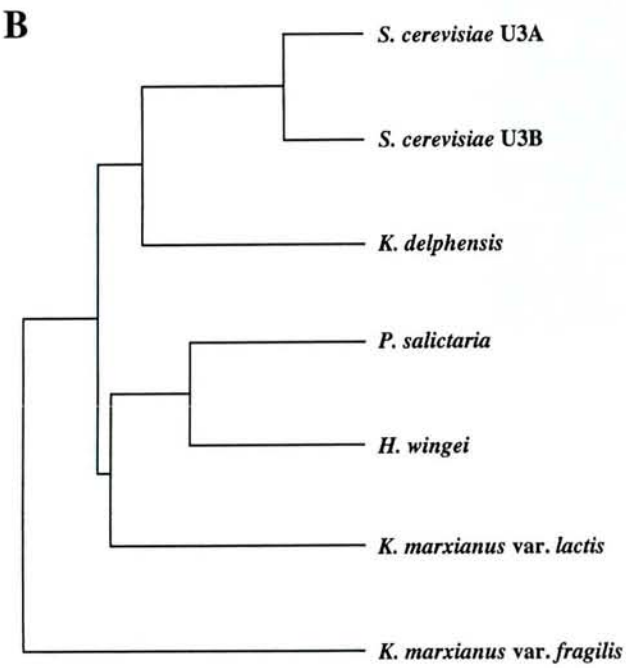
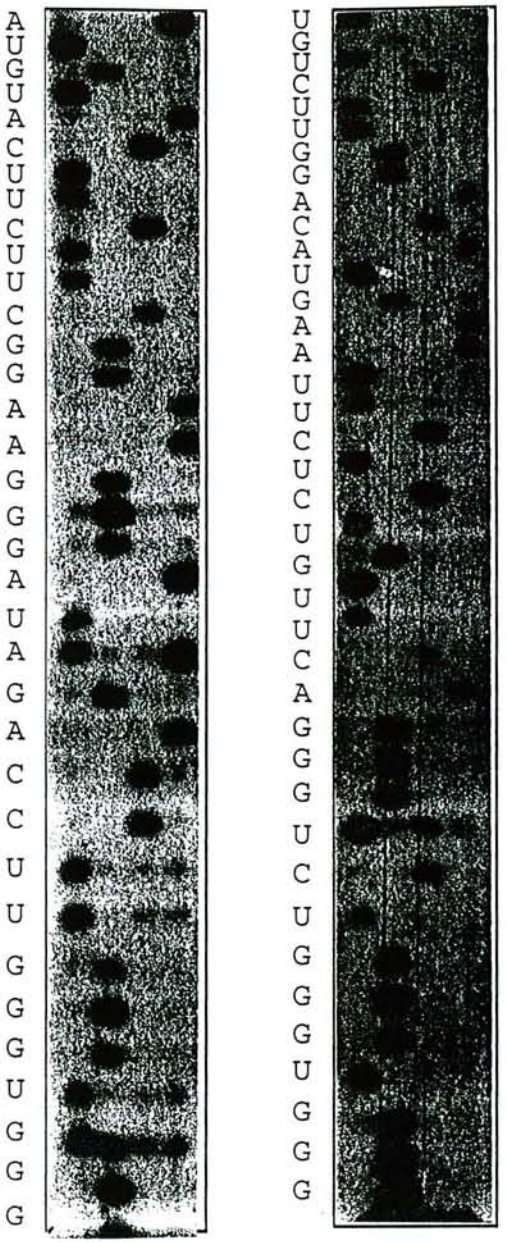

U3A plasmid
S. cerevisiae FL100
P. salictaria
K. delphensis
K. fragilis
K. lactis
H. wingei
Size marker
U3A gene
U3B gene
587
540
504
458
434
267
234

Figure 3


A
Box A'
Box A
Box C'
Box B
Box C
Box D
S.cerevisiae U3A
S.cerevisiae U3B
K.delphensis U3Kd
P.salictaria U3Ps
H.wingei U3Hw
K.lactis U3Kl
K.fragilis U3Kf
316nt
291nt
333nt
333nt
404nt
417nt
334nt


B
S. cerevisiae U3A
S. cerevisiae U3B
K. delphensis
P. salictaria
H. wingei
K. marxianus var. lactis
K. marxianus var. fragilis

A



P. salictaria *K. delphensis*

B

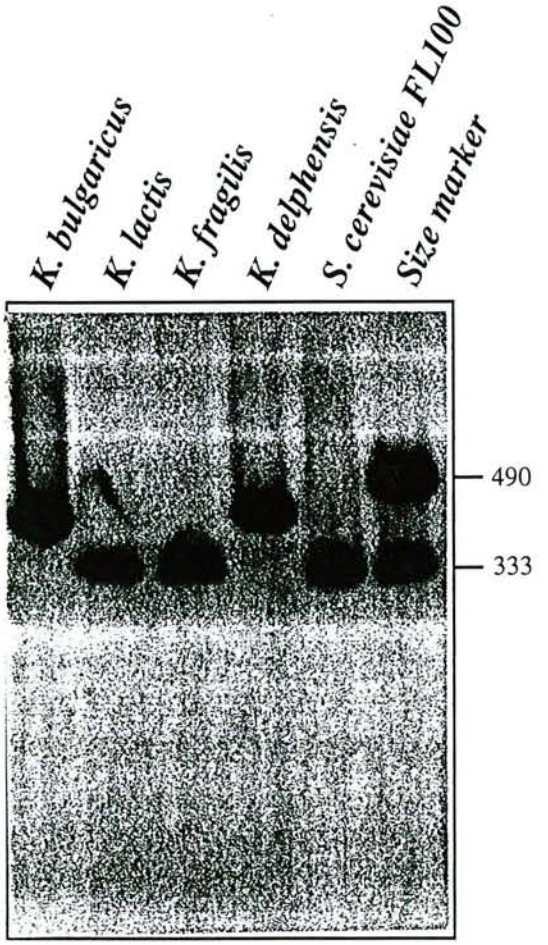
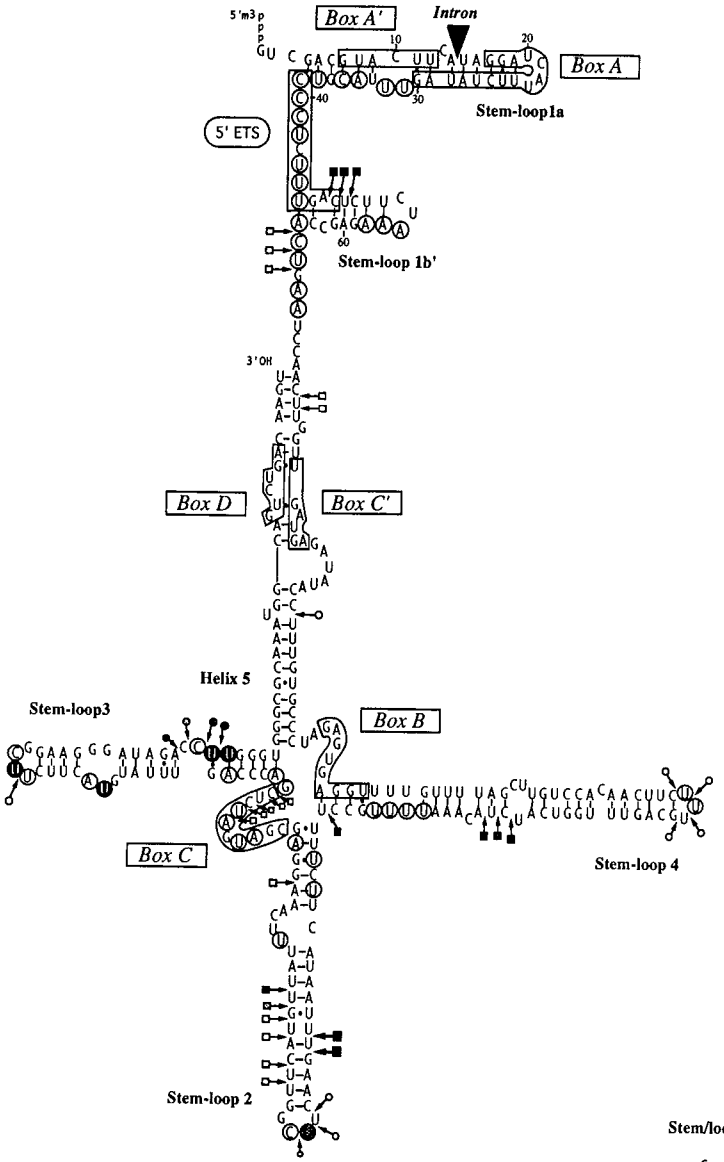
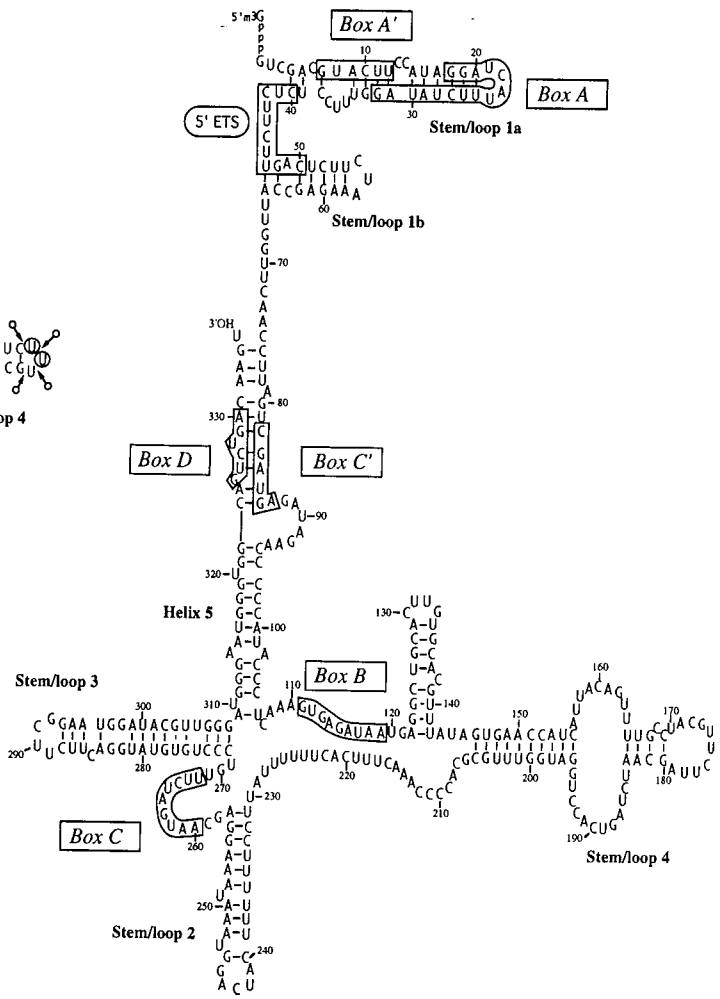


Figure 5



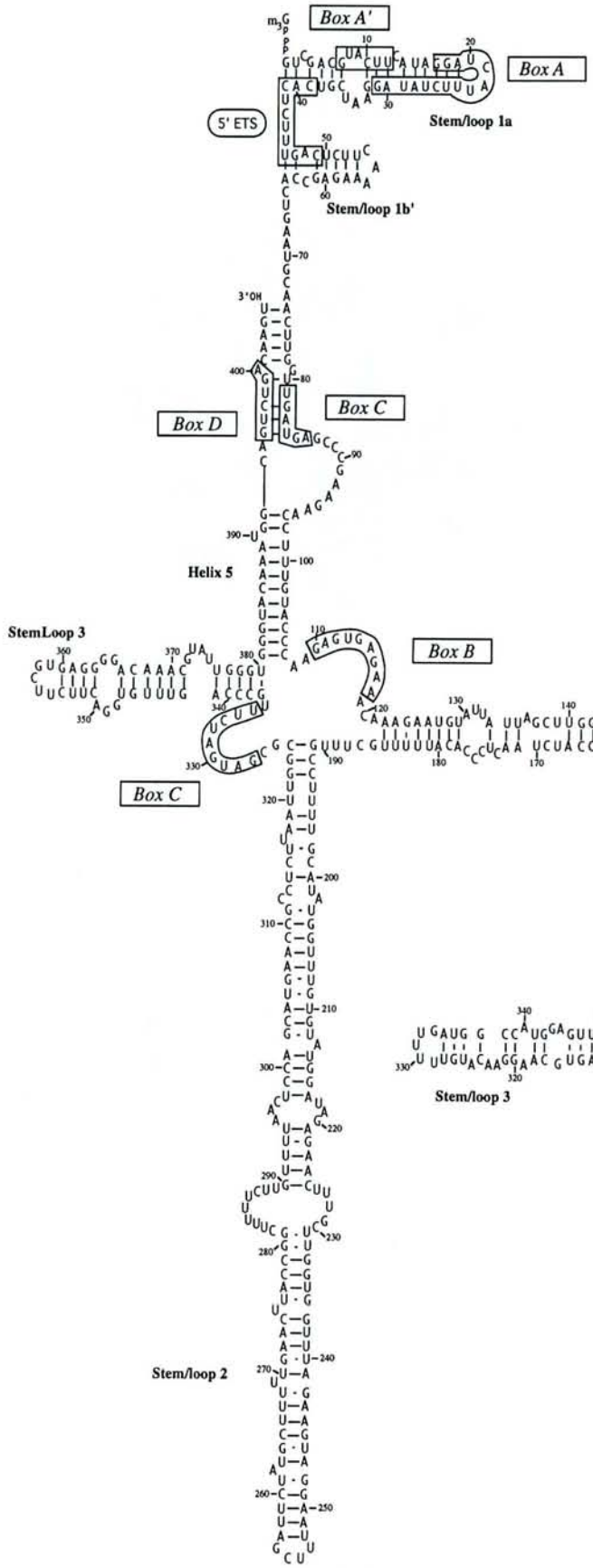
U3Ps

Pichia salictaria



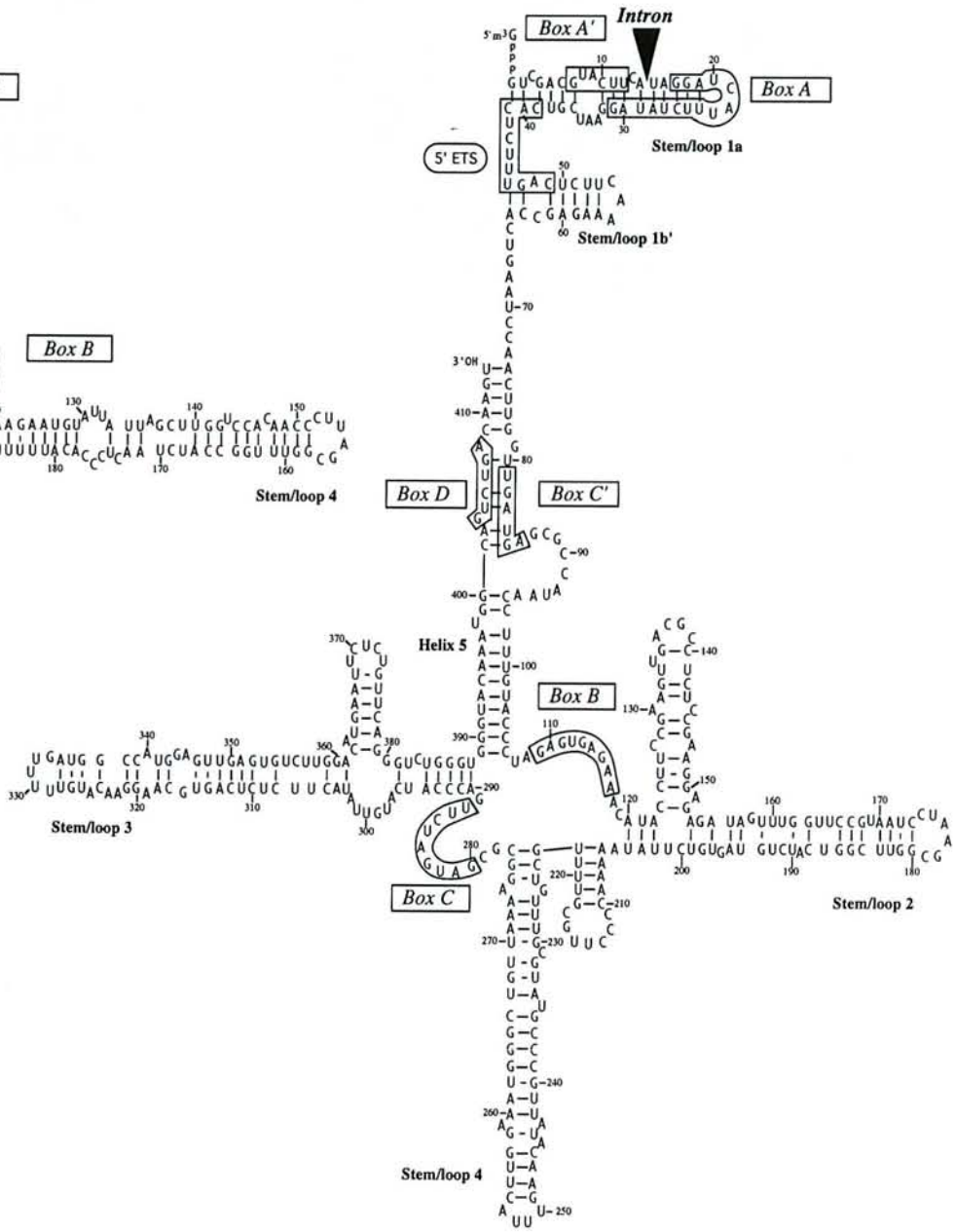
U3JKf

Kluyveromyces marxianus var. *fragilis*



U3Kl

Kluyveromyces marxianus **var.** *lactis*



U3Kd

Kluyveromyces delphensis

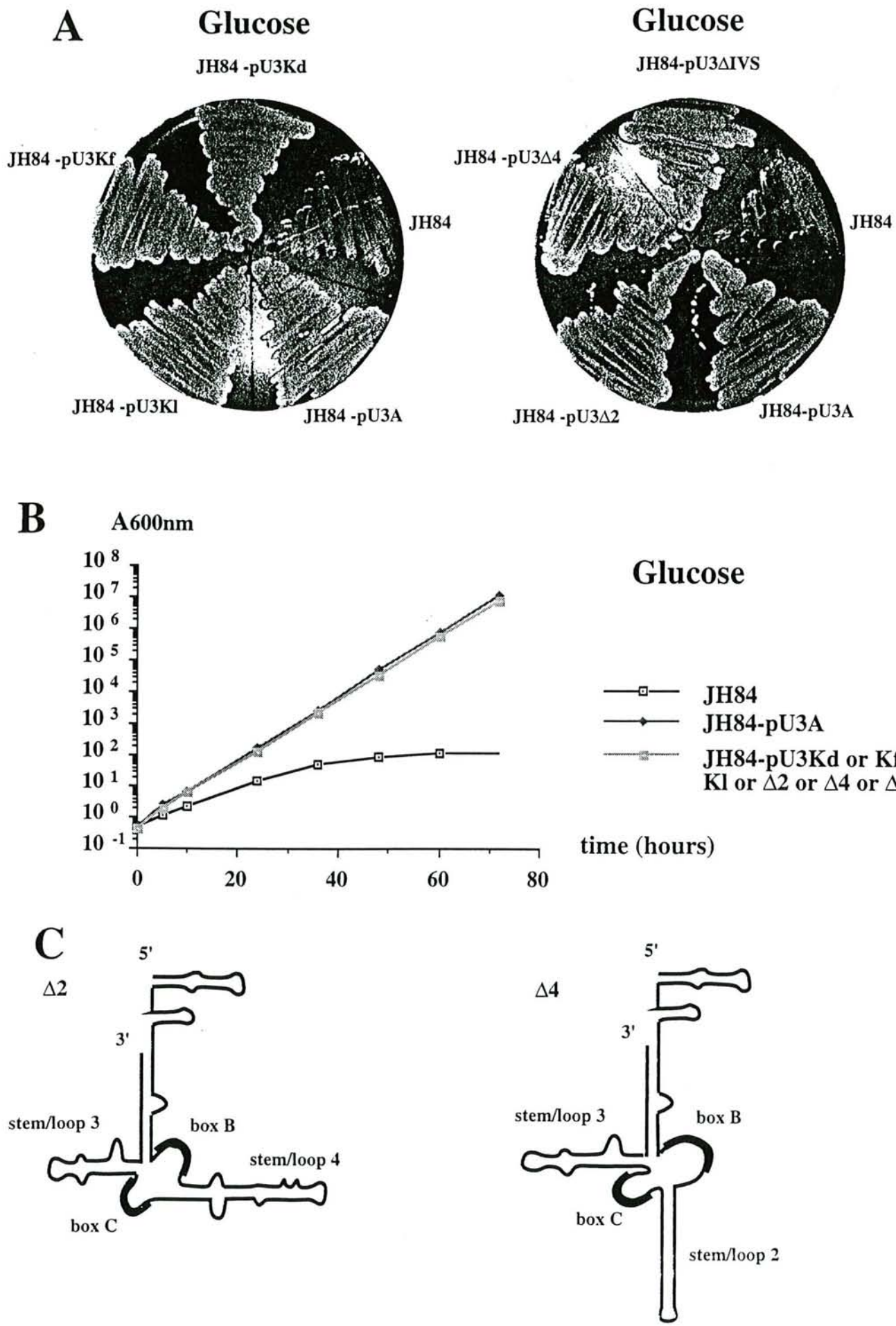
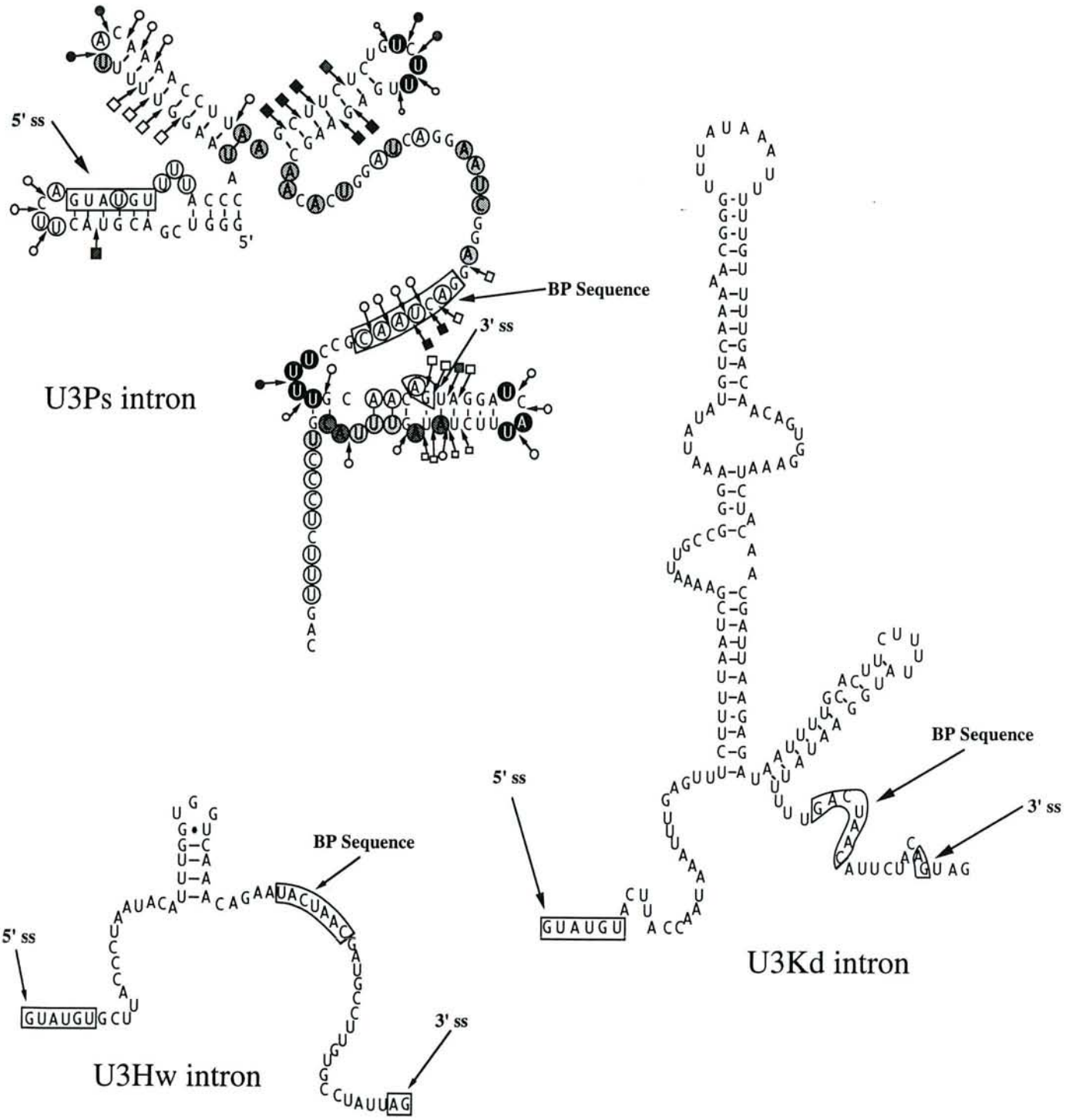

A
Glucose
JH84 -pU3Kd
JH84 -pU3Kf
JH84
JH84 -pU3Kl
JH84 -pU3A
Glucose
JH84-pU3ΔIVS
JH84 -pU3Δ4
JH84
JH84 -pU3Δ2
JH84-pU3A
B
A600nm
10 8
10 7
10 6
10 5
10 4
10 3
10 2
10 1
10 0
10 -1
0
20
40
60
80
Glucose
JH84
JH84-pU3A
JH84-pU3Kd or Kf or Kl or Δ2 or Δ4 or ΔIVS
time (hours)
C
Δ2
5'
3'
stem/loop 3
box B
stem/loop 4
box C
Δ4
5'
3'
stem/loop 3
box B
box C
stem/loop 2

Figure 7


5' ss
BP Sequence
3' ss
U3Ps intron
5' ss
BP Sequence
3' ss
U3Kd intron
5' ss
BP Sequence
3' ss
U3Hw intron

Discussion
Perspectives

H. Discussion - Perspectives

Notre étude a permis de mettre en évidence cinq nouveaux snoRNA U3 de levures appartenant aux genres *Kluyveromyces*, *Pichia* et *Hansenula* de la sous-famille des Saccharomycetoideae. Sur la base de la séquence de ces ARN et d'une étude faite par A. Grégoire sur la snoRNP U3A de *S. cerevisiae*, une révision de la structure secondaire proposée antérieurement pour le snoRNA U3 de *S. cerevisiae* est présentée.

Les différents snoRNA U3 de levure présentent un domaine 5' de petite taille, très conservé chez toutes les espèces et un domaine 3' très variable en taille et en séquence. Ce domaine 3' a cependant, une architecture cruciforme comparable chez toutes les espèces étudiées, avec au centre, les boîtes phylogénétiquement conservées B et C. Ces boîtes et les boîtes C' et D sont supposées constituer les sites de fixation des protéines de la particule ribonucléoprotéique qui se forme autour du snoRNA U3. Nous avons démontré la fonctionnalité de ces domaines 3' très variables chez *S. cerevisiae*. La présence de structures tige/boucle latérales supplémentaires, l'extension des structures tige/boucle ou encore la variabilité de la séquence des structures tige/boucle de la partie cruciforme du domaine 3' du snoRNA U3, n'affecte pas la fonctionnalité de cet ARN chez *S. cerevisiae*. La délétion de l'une ou l'autre des structures tige/boucle 2 et 4 du snoRNA U3A de *S. cerevisiae* n'affecte pas non plus la croissance. Aux vues de ces résultats, nous pouvons supposer que seulement la séquence et le positionnement des boîtes B, C, C' et D sont importants pour la fonctionnalité du domaine 3'.

Afin d'avoir une estimation plus précise de l'importance de chacun des éléments de la structure cruciforme en 3', il sera intéressant de compléter les expériences de délétions en éliminant la structure tige/boucle 3, puis les structures tige/boucles 2 à 2 ceci tout en conservant les boîtes B et C. Il serait aussi intéressant de tester l'effet d'un allongement ou d'un raccourcissement de la partie de l'hélice 5 joignant les boîtes B et C aux boîtes C' et D. Aucune variation de taille n'a été observée dans les différents ARN que j'ai mis en évidence, seule la séquence varie. On peut supposer que cette hélice est soumise à une contrainte en taille, du fait de la nécessité de préserver la distance entre le couple de boîtes B et C et le couple de boîtes C' et D, ceci afin de préserver les sites de fixation des protéines.

En utilisant le crible génétique mis au point par R. Fournier, il sera aussi intéressant de tester si l'un ou l'autre des deux appariements possibles entre la boîte B et la boîte C, pour lesquels nous avons observé une conservation dans tous les snoRNA de levure caractérisés, a une importance biologique. Pour cela, des mutations compensatoires devront être réalisées dans les deux séquences. Bien sûr, une réponse ne pourra être obtenue que si la fonction du snoRNA U3 n'impose pas une stricte conservation de la séquence des boîtes B et C.

Notre analyse comparative des différents snoRNA U3 caractérisés met aussi en évidence dans le domaine 3', la conservation de deux blocs de 3 paires G-C, l'un à la base de l'hélice 3, l'autre à la

base de l'hélice 5. On peut supposer que cette conservation est liée à des contraintes imposées par la nécessité d'avoir une structure stable. Il est d'ailleurs possible d'imaginer que ces deux hélices soient empilées et forment ensemble une longue structure hélicoïdale. Un empilement des hélices 2 et 4 est aussi possible, de telle sorte que le domaine 3' aurait en fait une structure en forme de H.

Vu son haut degré de structuration, le domaine 3', boîtes C' et D non incluses, nous paraît être un bon modèle d'étude de la structure tridimensionnelle d'ARN à quatre branches par diffraction des RX sur des cristaux. Aucune structure de ce type n'a encore été étudiée par cette approche. Le haut degré de structuration de ce domaine devrait favoriser sa cristallisation.

Le crible génétique mis au point par R. Fournier permettra aussi d'étudier le rôle des boîtes conservées dans la fixation des protéines et le mécanisme d'action du domaine 5'. En effet, nos travaux et ceux de M. Beltrame et D. Tollervey (1992) ont montré l'importance de la séquence entre les positions 39 et 49 qui s'apparie à la région 5' ETS. Mais le snoRNA U3 pourrait être multifonctionnel au niveau de la maturation des ARN pré-ribosomiques. Ainsi, une hypothèse de son implication en tant que chaperon du repliement de l'ARNr 17S a été formulée (Hughes, 1996). Cette hypothèse doit néanmoins être démontrée expérimentalement.

Par ailleurs, des travaux récents de Elela *et al.* (1996) soulèvent le problème du rôle précis de la snoRNP U3 dans la coupure au site A0. En effet, une nucléase ayant une activité similaire à celle de la RNase III bactérienne vient d'être mise en évidence chez *S. cerevisiae*. La destruction de son gène entraîne une accumulation de précurseurs non maturés des ARN ribosomiques. Un ARN modèle correspondant à la partie en double-brin de la région 5' ETS, qui contient le site A0, est clivé au site A0 par la pseudo-RNAse III purifiée de levure. De même, un ARN modèle comportant une structure en double-brin, comportant le site de maturation situé 21 nt en aval de l'extrémité 3' de l'ARNr 26S, est aussi clivé à ce site par l'enzyme purifiée, tout ceci en absence de snoRNA U3. Comment concilier le fait que le snoRNA U3 est indispensable *in vivo* (Hughes et Ares, 1991) et dans un extrait cellulaire (Tyc et Steitz, 1989) pour que la coupure en A0 ait lieu et l'activité observée pour cette nucléase ? Une possibilité est que l'enzyme à activité RNase III soit fixée sur la snoRNP U3 qui la véhiculerait au niveau des sites à cliver. Vue la forte activité vis-à-vis de tout ARN double-brin que semble avoir cette enzyme (Elela *et al.*, 1996), on peut aussi supposer que son activité soit modulée par une autre protéine répresseur et qu'un système d'activation mettant en jeu la snoRNP U3 soit nécessaire pour révéler son activité *in vivo*, vis-à-vis de sites bien définis.

La mise en évidence d'une activité RNase III chez *S. cerevisiae* d'un côté et d'un ARN qui ressemble au snoRNA U3 chez une archaebactérie, *Sulfolobus acidocaldarius*, ARN qui semble être nécessaire à une des étapes de maturation de l'ARNr 16S (Durovic et Dennis, 1994; Potter *et al.*, 1995), suggèrent que le système de maturation mettant en jeu un ARN de type snoRNA U3 est en fait très ancien.

Bien que différents modèles d'interaction entre le snoRNA U3 de vertébrés et la région 5' ETS aient été proposés (Stroke et Weiner, 1989; Maser et Calvet, 1989), aucune démonstration claire

d'une interaction précise entre ces deux ARN, telle que celle que l'équipe de D. Tollervey (Beltrame et Tollervey, 1995) et notre équipe (Brulé *et al.*, 1996) avons mis en évidence chez *S. cerevisiae*, n'a été jusqu'à présent mise en évidence chez les vertébrés. D'ailleurs, la structure secondaire du domaine 5' du snoRNA U3 de vertébrés est différente de celle que nous proposons pour le snoRNA U3 des levures. Le segment du snoRNA U3 de vertébrés, correspondant au segment simple-brin du snoRNA U3 de levure qui s'apparie avec la région 5' ETS, est impliqué dans une longue structure tige/boucle.

Le domaine 5' du snoRNA U3 de plantes comporte deux structures tige/boucle comme le snoRNA U3 de levure (Marshallsay *et al.*, 1990). Par analogie avec ce que nous avons observé chez les levures, il serait intéressant de tester si un appariement peut se former entre la séquence du snoRNA U3 de plantes en simple-brin liant les deux structures tige/boucle et la région 5' ETS de l'ARN pré-ribosomique de plantes.

Au cours de nos comparaisons, en plus du snoRNA U3 de *Hansenula wingei*, qui comportait deux mutations dans le segment complémentaire de la région 5' ETS, une de ces deux mutations est retrouvée dans le snoRNA U3 exprimé chez *Pichia salictaria*. Il sera intéressant de rechercher l'existence d'une mutation compensatoire dans la séquence de la région 5' ETS chez cette levure.

Notre étude comparative de snoRNA U3 de levures et l'observation d'un grand degré de variation du domaine 3', suggèrent que l'on devrait pouvoir produire un snoRNA U3 actif tout en ayant inséré une séquence excédentaire dans l'une des deux structures tige/boucle 2 ou 4, à condition que cette séquence s'auto-structure. Ceci nous ouvre des possibilités intéressantes pour essayer d'étudier la composition en protéines de la snoRNP U3 de levure. L'idée est d'insérer le fragment d'ARN reconnu par la protéine de l'enveloppe du phage R17 dans une de ces structures tige/boucle. Un gène portant cette insertion pourra être inséré dans le plasmide pASZ11. Les snoRNP U3 produites *in vivo* pourront ainsi être purifiées par la méthode triple-hybride qui consiste à fixer la composante snRNP étudiée sur une protéine de fusion : protéine de la coque du phage R17/glutathion transférase. La protéine de fusion peut elle-même être retenue sur glutathion sépharose. Des mutations pourront ensuite être générées dans la région codante du snoRNA U3, afin de tester l'effet de ces mutations sur la composition de la snoRNP U3.

En plus des informations apportées sur la structure et le mécanisme d'action possible du snoRNA U3, notre étude comparative des gènes du snoRNA U3 chez différentes espèces de levures apporte des informations sur l'évolution des introns portés par ces gènes. Nous avons observé la présence de gènes de snoRNA U3 sans intron chez les espèces du genre *Kluyveromyces* et ces gènes sans intron sont exprimés. Nous avons aussi montré qu'une souche de *S. cerevisiae* dont les gènes de snoRNA U3 chromosomiques sont inactivés peut être complémentée par un gène de snoRNA U3 sans intron. Donc, bien qu'au moins un gène de snoRNA U3 avec un intron soit présent dans chacune des levures que nous avons étudiées, la présence d'un tel intron n'a pas un

caractère indispensable pour la croissance de la cellule. Un point qui nous paraît très étonnant, et pour lequel nous ne sommes pas en mesure de donner une explication dans l'état actuel des connaissances, est la très forte conservation du gène du snoRNA U3, exons et intron compris. Vu le nombre croissant de petits ARN codés par des introns mis en évidence au cours des dernières années (pour revue, Bachellerie *et al.*, 1995), nous nous sommes demandé si l'intron du gène du snoRNA U3A ne pourrait pas correspondre à un petit ARN stable. Plusieurs arguments sont en défaveur de cette hypothèse. Cet intron ne présente pas les boîtes C et D observées dans un grand nombre de petits ARN nucléolaires codés par des introns. Mais surtout, selon l'étude réalisée par R. Fournier, par la méthode d'hybridation de Northern, l'intron ne semble pas être stable *in vivo*. Le précurseur ne peut pas en lui-même avoir un rôle particulier, il ne s'accumule pas dans la cellule. Le gène sans intron s'exprime apparemment au même taux que le gène avec intron, ce qui exclut la possibilité que des éléments activateurs de la transcription soient présents dans l'intron, comme c'est parfois le cas chez les vertébrés. Le problème du fort degré de conservation de séquence de cet intron reste donc entier. Il en est de même pour la conservation de la région codante du snoRNA U3A, en comparaison de la forte divergence de séquences du second gène de snoRNA U3, d'un genre de levure à un autre, voire même d'une espèce de levure à une autre au sein des genres *Kluyveromyces* et *Pichia*.

Le fait que l'intron soit situé à la même position dans la région codante suggère que l'existence de cet intron est ancienne. En ce sens, il est possible que l'apparition de gène sans intron ait été générée par transcription inverse d'un snoRNA mature suivie de l'insertion de l'ADNc dans le noyau. La comparaison de la séquence des régions flanquantes des gènes avec et sans intron dans une même espèce devrait donner des informations sur ce point.

Dans le double objectif de compléter cette étude de l'évolution des gènes de snoRNA U3 chez les levures et en vue de développer des méthodes rapides d'identification des espèces de levures, il sera intéressant d'étudier des espèces appartenant aux autres genres de la sous-famille des Saccharomycetoideae. Nous avons en effet étudié des espèces appartenant à quatre genres, or cette sous-famille en comporte 26 (Kreger van Rij, 1987). Il sera aussi intéressant d'étendre l'étude à d'autres sous-familles de levures. Par contre, comme nous l'avons mentionné dans l'introduction, la classification des espèces de levures fait encore l'objet de controverses et les méthodes biochimiques d'identification sont souvent laborieuses. En effet, l'analyse du contenu en gènes de snoRNA U3 par la méthode PCR est une méthode simple à mettre en œuvre, les produits d'amplification obtenus peuvent être spécifiques d'un genre ou d'une espèce donnés.

Notre étude apporte aussi des arguments forts confirmant le rôle très important de la structure secondaire du pré-ARN pour l'efficacité d'épissage des introns chez *S. cerevisiae*.

Les constructions que j'ai réalisées permettant de produire les snoRNA U3A et U3B *in vitro*, ont permis à notre équipe d'établir la structure secondaire de ces deux pré-ARN et d'observer, en

collaboration avec J. Banroques, une très forte différence d'efficacité d'épissage entre deux types de transcrits différant par le degré d'accessibilité de la boîte en 5' de l'intron. Les séquences que j'ai établies pour le gène allélique du snoRNA U3A*, et pour les introns U3C et U3D mis en évidence dans la levure polyploïde *Saccharomyces uvarum* apportent, par l'observation de mutations compensatoires, des arguments phylogénétiques en faveur de la validité des modèles de structure secondaire établis pour les introns U3A et U3B et démontrent l'importance de la présence d'une structure tige/boucle centrale pour ces introns de *S. cerevisiae*. Notre observation d'une structure tige/boucle centrale dans l'intron du pré-snoRNA U3Kd de *Kluyveromyces delphensis* montre que la nécessité de cette structure tige/boucle centrale pour un épissage efficace des introns longs, est sans doute générale chez tous les genres de cette sous-famille.

D'une manière générale, nous constatons que dans toutes les espèces que nous avons étudiées, une conservation complète de la conformation des trois séquences fonctionnelles de l'intron du pré-snoRNA U3 : une boîte 5' toujours appariée à l'exon 1, ce qui était très inattendu pour des pré-ARN qui doivent être épissés efficacement, un dinucléotide AG terminal en boucle latérale, et une boîte de branchement toujours accessible en simple-brin.

Les travaux de mutagenèse dirigée de A. Grégoire et R. Fournier ont montré l'importance de la nature simple-brin de la boîte de branchement pour l'efficacité d'épissage du snoRNA U3A de *S. cerevisiae*.

Il est aussi intéressant de noter que dans tous les modèles de pré-snoRNA U3 que nous avons pu établir, en dépit des variations de taille et de séquences de l'intron, la longueur de la région en simple-brin reliant la structure tige/boucle 5' terminale et la structure tige/boucle renfermant le AG 3' terminal est toujours très voisine, environ 30 nt. Il s'agit sans doute d'une distance optimale pour l'assemblage des complexes spliceosomaux. Il sera intéressant d'étudier l'effet sur l'efficacité d'épissage d'une extension ou d'un raccourcissement de cette région en simple-brin.

La grande conservation observée de la conformation globale des introns de pré-snoRNA U3 est une illustration de l'importance que cette structure secondaire doit avoir pour l'efficacité d'épissage.

Le modèle de structure secondaire établi pour le pré-snoRNA U3A va maintenant servir de base à l'équipe pour l'étude de la structure des complexes spliceosomaux formés avec le pré-snoRNA U3A. Il sera, en particulier, intéressant de déterminer à quelle étape les structures tige/boucle 5' et 3' terminales de l'intron sont ouvertes, afin de permettre la ligation des exons. D'une manière générale, il sera aussi intéressant de voir comment la structure établie *in vitro* évolue en présence des protéines présentes dans un extrait nucléaire.

Publication annexe :
L'ARN messager de l'haptocorrine comme marqueur tumoral

S. Frémont, A. Mougin, F. Boisson, F. Brulé, I. Sponne, J.L. Guéant,

C. Branlant et J.P. Nicolas

Journal of Endocrinology (1994), vol. 2 269-276

Advances in Thomas Addison's Diseases
Eds Bhatt, James, Besser, Bottazzo & Keen, Vol 2, pp 269-276
Journal of Endocrinology Ltd, Bristol (1994)

Haptocorrin messenger RNA as a tumoral marker

S Frémont, A Mougin*, F Boisson, F Brulé*, I Sponne, J L Guéant, C Branlant* and J P Nicolas

Laboratoire de Biochimie Nutritionnelle, INSERM U308, Faculté de Médecine, BP 184, 54505 Vandoeuvre les Nancy Cedex, France; *Laboratoire d'Enzymologie et de Génie Génétique, CNRS URA 457, Faculté des Sciences, BP 249, 54505 Vandoeuvre les Nancy Cedex, France

Introduction

Vitamin B_{12} (cobalamin, Cbl) is transported in the blood by two proteins. One is haptocorrin (HC) otherwise known as transcobalamin I and III (TC I and TC III), the only difference between them being their sialic acid composition. HC is found in plasma, granulocytes, and various body fluids (milk, saliva, tears and bile). Its function is still unclear. HC synthesis could occur at an early stage of granulocytes maturation (7). The other protein is transcobalamin II (TC II), which is involved in tissular exchanges. HC and TC II can be either saturated by Cbl or unsaturated (6).

Different authors have studied the blood levels of HC in liver diseases. In 1974, Burger *et al* (1) observed an increased concentration of serum HC, associated with a high serum Cbl level in malignant hepatoma (MH). Carmel noticed the same increase in patients with hepatic metastasis (3). This HC increase is observed even without any rise in alpha foeto-protein in the case of MH (10). An important increase in TC I level could precede clinical and radiological evidence of a recurring fibrolamellar hepatoma (17). This rise could be induced by synthesis of TC I by tumor cells (11). An increase in TC I and TC III can also be observed in viral hepatitis (12).

Our group recently compared the blood variations of vitamin B12 binding proteins values in malignant hepatoma and in other chronic liver diseases (hepatic metastasis, cirrhosis, chronic viral hepatitis, hepatic steatosis, hepatic cholangiocarcinoma). We showed a significant increase in total HC (saturated and unsaturated) in MH ($p < 0.005$). On the contrary, we did not note any significant TC II variation in MH (5).

The question was to know if the increased blood level of HC, in malignant hepatoma, corresponded to a greater synthesis or to a defective catabolism.

In aiming to answer this question and to know at which stage of the granulocytes maturation HC is synthesized, we developed a reverse transcription polymerase chain reaction (RT-PCR) assay to study HC mRNA in different samples.

Chronic granulocytic leukemia (CGL) is a myeloproliferative syndrome characterised by a proliferation of the myeloid series cells. The myeloid cells present in the blood, particularly myelocytes and metamyelocytes, result from this hyperactivity

(19). Thus the method was first developed with blood samples from patients suffering from CGL and then applied to hepatic biopsies.

Material and methods

Samples

Blood samples: These were obtained from the Centre de Transfusion Sanguine (Dr Schooneman) and from the Service de Médecine A (Dr. Guerci) of the Centre Hospitalier Régional Universitaire in Nancy. Peripheral blood leukocytes were isolated from blood of healthy donors by dextran gradient protocol (4). We used apheresis blood treatment cells from patients with CGL.

Tissues: Hepatic biopsies were obtained from malignant hepatoma patients who were operated on. After the hepatic tumor was removed and used for anatomopathological identification we received two samples: one from the tumor and one from surrounding healthy hepatic tissue.

RNA preparation

Samples of granulocytes or hepatic biopsies conserved at −80°C, were thawed and homogenized in 4 M guanidine isothiocyanate (14). The mixture was centrifuged through a bed of 5.7 M cesium chloride at 125 000 g for 21 hours.

The pellet was resuspended in Tris EDTA buffer and extracted once with phenol: chloroform mix and precipitated in ice cold ethanol at −20°C. The RNA extracted was quantified by spectrophotometry at 260 nm and kept at −80°C in Tris EDTA buffer. After centrifugation, the dry pellet was redissolved in Tris EDTA buffer.

Oligonucleotides used for amplification

Recently, Johnston *et al* (7) reported the cloning of a full-length cDNA for transcobalamin I. They then cloned the genomic sequence for TCI and reported mapping of the exons (9).

Four oligonucleotides, chosen within the TCI-cDNA were synthesized on a Biosearch DNA synthesizer: A (5′ primer) and B (3′ probe) are included in the second exon. C (3′ primer) and D (3′ primer) are included in the third exon (Fig. 1).

A: 5′-GAAGAAAACTACATCCG-3′
B: 5′-CAGGGTTTGGATCTGGATTCCAA-3′
C: 5′-CGAGCTTACATCTGACAATCTGCT-3′
D: 5′-CATACTCCCAAAGCCAGT-3′

The expected products synthesized by PCR are A/D: 227 bp and A/C: 189 bp.

Amplification method

RT-PCR: The RNA extracted was denaturated 10′ at 65°C with 5′ primer, and reverse transcription was achieved 45′ at 42°C by RAV_2 reverse transcriptase (Amersham, England). The resulted cDNA was amplified by PCR. It was mixed with 5′ (A) and 3′ (D) primers, deoxynucleotides in excess, PCR buffer, diethylpyrocarbonate-treated water and *Thermus thermophilus* DNA polymerase (Chelzyme, England).

The mixture, overlaid with mineral oil, was heated at 94°C for three minutes before 30 cycles of annealing (at 55°C for 30 seconds), extension (at 72°C for one minute) and denaturing (at 94°C for 30 seconds) in a thermal cycler (PREM III, LEP, England).

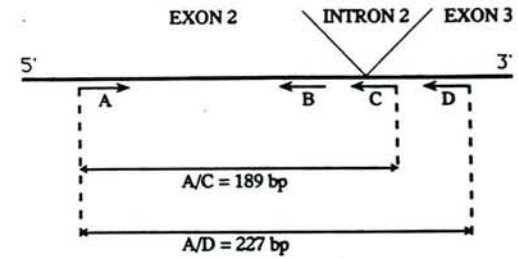


Fig. 1 (a) Diagram showing the location of the oligonucleotide primers used for RT-PCR amplification of the HC mRNA and the expected amplified products (Johnston *et al* 1989);

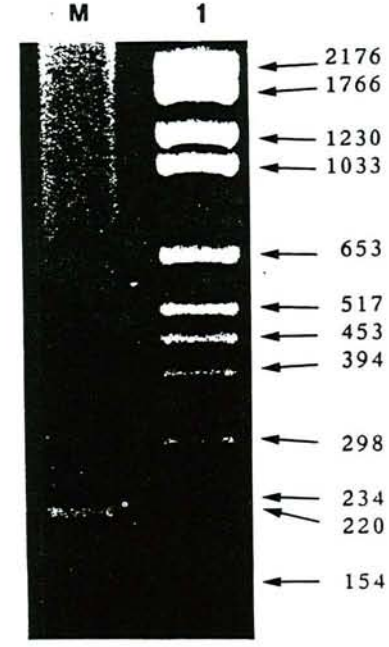


Fig. 1 (b) Fractionation on 2% agarose gel of the RT-PCR products obtained using CGL-RNA template and oligonucleotides A and D, as primers. Lane 1: amplification product. M: size marker (numbers to the right denote molecular weight in Kb);

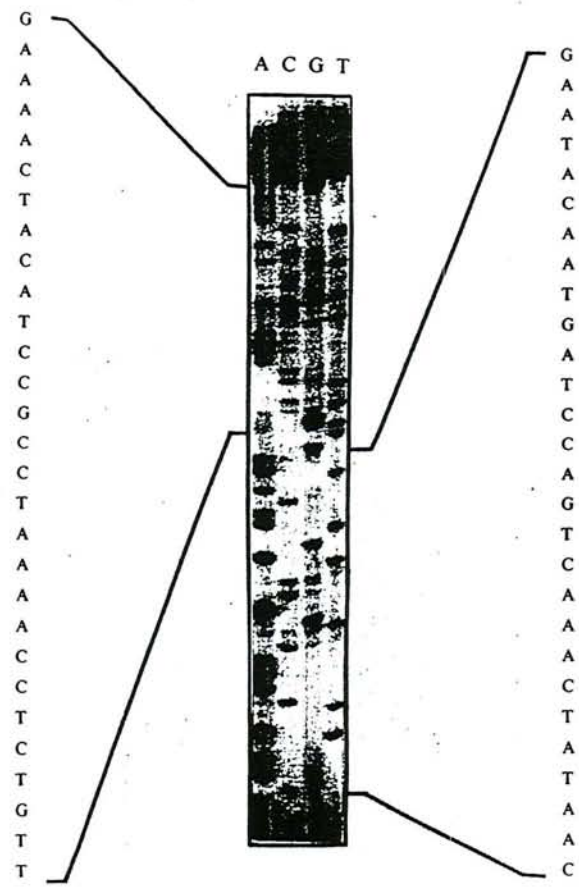


Fig. 1 (c) Nucleotide sequence analysis of the RT-PCR product presented in Figure 1 (b).

The RT-PCR products were then electrophoresed in 2% (w/v) agarose ethidium bromide-stained minigel. A mixture of Bgl 1-digested pBR 328 and Hinf 1-digested pBR 328 was used as a size marker (Boehringer, Mannheim, Germany).

Nested PCR: After excision of the DNA band from the gel, elution was performed using a silica containing membrane (Glass MAX Spin Cartridge, GIBCO BRL, USA).

The eluted fragment was reamplified by PCR with 20 cycles in the same conditions as described above, but using nested oligonucleotide C as 3′ primer.

Southern blot analysis

The PCR products fractionated on the agarose gel were blotted on to a nylon filter (Hybond N, Amersham, England) and hybridized with a specific probe by the Southern method (17). We used primer B, which is located between the two external primers, as a probe (Fig. 1). It was labelled with [γ-^{32}P] ATP (Amersham, England) by the 5′-end labelling method, using T_4 polynucleotide kinase. The labelled probe was purified on a 8% polyacrylamide gel electrophoresis.

The nylon filter was first prehybridized in a solution containing 5 x SSC, 5 x Denhardt's reagent, 0.5% (w/v) SDS, 100 mg/ml denaturated fragmented salmon sperm DNA and 50% formamide at 42°C for one hour. The prehybridization solution was completely removed and replaced by the hybridization solution (same solution without Denhardt's reagent).

The labelled probe was added (specific activity: 4.10^6 cpm/µg)) and the incubation performed at 42°C for 12 hours. The membrane was washed twice with 2 x SSPE and 0.1% SDS solution at room temperature, dried and autoradiographed.

Sequencing

DNA sequencing was performed using the dideoxy chain termination method developed by Sanger (T_7 Sequencing kit, Pharmacia, Uppsala, Sweden), using oligonucleotide B as primer (15).

Results and discussion

As it is a well known fact that HC is synthesized by myeloid series cells, we isolated RNA from granulopoietic progenitors which are present in the blood of patients suffering from CGL. After RT-PCR using oligonucleotides A and D as primers, we obtained a single fragment with the expected size for HC mRNA amplification (Fig. 1). Due to the presence of intron 2, the amplification of an eventual contaminating genomic DNA with the same oligonucleotides would have lead to a 1.8 kb amplified fragment (9). The fact that the amplified fragment does correspond to HC mRNA amplification was confirmed by Southern blot analysis, using ^{32}P-labelled oligonucleotide B as a probe (data not shown), and by nucleotide sequencing, using primer B (Fig. 1 (c)).

Our results demonstrate the presence of HC mRNA in the granulocytes or their precursors.

On the other hand, when healthy blood donors' granulocytes were used for the same kind of experiment, an HC-specific amplified fragment was observed only once among seven samples tested. One explanation could be the presence of granulopoietic progenitors in the blood of that donor, as a result of a recent infection which had stimulated the proliferation of myeloid series cells.

The presence of HC in mature granulocytes has been previously shown by several authors (2, 16. 19), but Johnston *et al* (7) have shown that its synthesis is co-ordinately

regulated with the appearance of secondary granules. More recently, these authors have shown that HC mRNA is detected at a reduced level in the bone marrow of patients with neutrophil-specific granule deficiency, as compared with normal bone marrow (8).

Our results showing the absence of HC synthesis in mature granulocytes of normal peripheral blood suggest that, like lactoferrin (13), HC is synthesized in developing neutrophils during a narrow window in the maturation process. Recently, Johnston *et al* (8) hypothesized that the expression of the genes encoding these two proteins requires the presence of a common transcription factor.

In the second part of the study, we used the developed PCR protocol to compare malignant hepatoma and healthy hepatic biopsy.

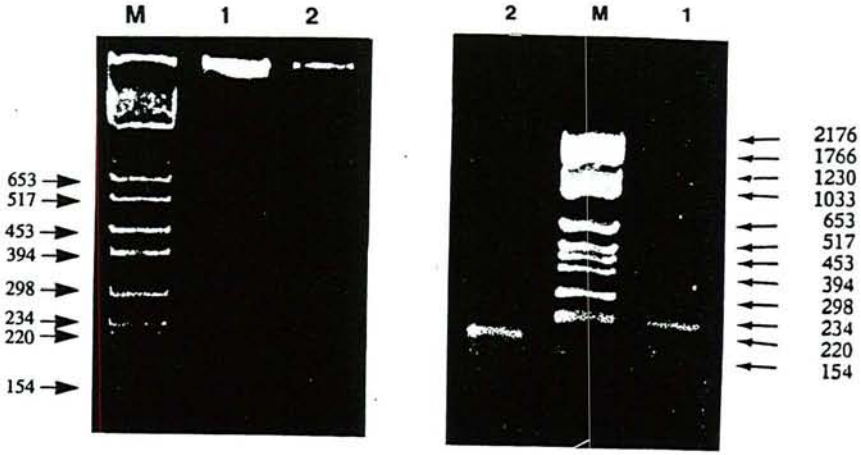


Fig. 2 (a, left) Fractionation on 10% polyacrylamide gel of the RT-PCR product, obtained using oligonucleotides A and D as primers, and malignant hepatoma RNA (lane 2) and healthy hepatic biopsy - RNA (lane 1), as template; (b, right) Nested PCR. Lane 1: RT-PCR product using oligonucleotides A and D, and malignant hepatoma RNA. Lane 2: Second PCR product obtained by amplification with oligonucleotide A and nested oligonucleotide C of previous RT-PCR product. M: Size marker.

Using MH RNA as a template, we observed a band corresponding to the expected 227bp fragment (Fig. 2 (a) lane 2). When RNA was extracted from the healthy tissue surrounding the tumor no band was detected (lane 1). Similar results were obtained with two different patients. The identity of the amplified fragment was confirmed by Southern blot hybridization (data not shown) and by reamplification of the 227 bp eluted fragment using oligonucleotides A and C as primers (Fig. 2 (b)). The expected shorter fragment (189 bp) was obtained by this nested PCR (Fig. 2 (b)).

Our results suggest that HC synthesis does not occur within a healthy liver, but within a malignant hepatoma. This could explain the significant seric HC level

observed in this pathology. We have yet to ensure that the mRNA translation is well performed. This study, using anti-HC antibodies, is currently in progress.

In conclusion, HC mRNA is observed in cells from the granulopoietic progenitors present in CGL samples but not in mature granulocytes. In this case, HC synthesis is probably related to the presence of immature cells.

With regard to hepatic tissues, HC-specific mRNA is detected only in hepatoma and not in healthy hepatic biopsies, suggesting that this protein is expressed in tumoral cells which have lost part of their differentiate properties.

The synthesis of HC in these two malignant diseases led us to propose HC as a potential marker.

References

1 Burger R L, Waxman S, Gilbert H S & Allen R H (1974) Isolation of a novel vitamin B_{12} binding protein associated with hepatocellular carcinoma. *Journal of Clinical Investigation* **53** 13

2 Carmel R & Coltman C A (1971) Non leukemic elevation of serum vitamin B12 and B12 binding capacity levels resembling that in chronic myelogenous leukemia. *Journal of Laboratory and Clinical Medicine* **78** 289-297

3 Carmel R (1975) Extreme elevation of serum transcobalamin I in patients with metastatic cancer. *The New England Journal of Medicine* **292** 282-284

4 Cartier T, Marchand J C & Garreau H (1969) La séparation des cellules sanguines. *Annales de Biologie Clinique* **27** 133-157

5 Frémont S, Champigneulle B, Gérard P, Felden F, Lambert D, Guéant J L & Nicolas J P (1991) Blood transcobalamin levels in malignant hepatoma. *Tumor Biology* **12** 353-359

6 Guéant J L & Nicolas J P (1990) Cobalamin binding glycoproteins in the digestive tract. *Cobalamin and Related Binding Proteins in Clinical Nutrition*. Eds J L Guéant & J P Nicolas. Paris: Elsevier

7 Johnston J, Bollekens J, Allen R H & Berliner N (1989) Structure of the cDNA encoding transcobalamin I, a neutrophil granule protein. *The Journal of Biological Chemistry* **264** 15754-15757

8 Johnston J, Boxer L & Berliner N (1992) Correlation of messenger RNA levels with protein defects in specific granule deficiency. *Blood* **80** 2088-2091

9 Johnston J, Yang-Feng T & Berliner N (1992) Genomic structure and mapping of the chromosomal gene for transcobalamin I (TCN I): comparison to human intrinsic factor. *Genomics* **12** 459-464

10 Kane S P, Murray Lyon I M, Paradinas F J, Johnson P J, Williams R, Orr A H & Kohn J (1978) Vitamin B12 binding protein as a tumor marker for hepatocellular carcinoma. *Gut* **19** 1105-1109

11 Kudo H, Ohshio G, Ogawa K, Kim Y C, Wakatsuki Y, Nakashima Y, Yamabe H & Inada M (1988) Distribution of vitamin B12 R-binder in carcinomas of the digestive tract. *Journal of Clinical Pathology* **41** 320-323

12 Osifo B O A, Aydola A, Parmentier Y, Gerard P & Nicolas J P (1988) Correlation between serum enzymes and serum unsaturated vitamin B12 binding proteins in primary liver carcinoma. *Enzyme* **39** 161-166

13 Rado T A, Bollekens J, Laurent G, Parler L & Benz E (1984) Lactoferrin biosynthesis during granulocytopoiesis. *Blood* **64** 1103-1109
14 Sambrook J, Fritsch E F & Maniatis T (1989) *Molecular Cloning. Laboratory Manual Second Edition*. Cold Spring Harbor: Cold Spring Harbor Laboratory
15 Sanger F, Nicklen S & Coulson A R (1977) DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences of the USA* **74** 5463-5467
16 Simons K & Weber T (1966) The vitamin B12 binding protein in human leukocytes. *Biochimica et Biophysica Acta* **117** 201-208
17 Southern E M (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98** 503-517
18 Wheeler K, Pritchard J, Luck W & Rossister M (1986) Transcobalamin I as a marker for fibroamellar hepatoma. *Medical and Pediatric Oncology* **14** 227-229
19 Zittoun J, Marquet J & Zittoun R (1975) The intracellular content of the three transcobalamins at various stages of normal and leukaemic myeloid cell development. *British Journal of Haematology* **31** 299-310

Advances in Thomas Addison's Diseases
Eds Bhatt, James, Besser, Bottazzo & Keen, Vol 2, pp 277-285
Journal of Endocrinology Ltd, Bristol (1994)

Immunodetection, molecular expression and receptor binding of transcobalamin II in CaCo2 and HT 29 cells

J L Guéant, H Schohn, R Bougrine, C Masson, I Gastin, J P Nicolas, B Leheup*, G Grignon*, I Chantret**, M Rousset** and E Nexø***
Laboratoire de Biochimie Nutritionnelle, INSERM U 308, Université de Nancy, France; *Laboratoire d'Histologie-Embryologie et Microscopie Electronique, Université de Nancy, France; **INSERM U178, Hôpital P Brousse, Villejuif, France; ***Department of Clinical Chemistry, KH University Hospital of Aarhus, Denmark

Introduction

Assimilation of cobalamin (Cbl, vitamin B_{12}) required its intraluminal binding to gastric intrinsic factor (IF) and the subsequent receptor-mediated endocytosis of the IF-Cbl complex within the ileal mucosa. Transcytosis of cobalamin in the ileum involves a transfer of cobalamin to transcobalamin II (TCII). Exprimental data have suggested that TCII is synthesized by the ileal enterocytes (9). Recently, our group has reported that differentiated HT 29 cell, derived from colon carcinoma cells, secretes TCII (10). Dix *et al* (2) and Ramanujam *et al* (7) have shown that the CaCo2 cell line, also derived from colon carcinoma, secretes a cobalamin-binding protein, related as TCII, in the culture medium. The IF-receptor is present in the apical brush border membrane of HT 29 cell and mediates the endocytosis of Cbl-IF within the cell (3). The presence of IF-receptor in a colon carcinoma cell line is not surprising if one considers that this cell line resembles foetal enterocyte and that IF-receptor is present in the foetal colon as well as in the foetal small bowel (11). The exact role of transcobalamin II in the ileal intracellular Cbl trafficking and in the transcytosis remains to be determined. The main role of TCII is to deliver cobalamin from the plasma to the target cells via a receptor-mediated endocytosis (5, 13). This receptor-mediated endocytosis of TCII-Cbl has been described for various cells such as fibroblasts, placental and endothelial cells (5) and more recently germ cells (1).

The purpose of the present paper was first to characterize the secretion of TCII from CaCo2 cells cultivated in minicell chambers; second, to study the mRNA expression of TCII in the adult intestine and colon and in CaCo2 and HT 29 cells; and third, to investigate the presence of a TCII-receptor in the brush border membranes of HT 29 cells.

Bibliographie

Aarstad K. et Oyen T.B. 1975. On the distribution of 5S RNA cistrons on the genome of *Saccharomyces cerevisiae*. FEBS Lett., 51: 227-231.

Abovich N., Legrain P. et Rosbash M. 1990. The yeast PRP6 gene encodes a U4/U6 small nuclear ribonucleoprotein particle (snRNP) protein, and the PRP9 gene encodes a protein required for U2 snRNP binding. Mol. Cell. Biol., 10: 6417-6425.

Abovich N., Liao X.C. et Rosbash M. 1994. The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. Genes Dev., 8: 843-854.

Agabian N. 1990. Trans splicing of nuclear pre-mRNAs. Cell, 61: 1157-1160.

Allmang C., Henry Y., Wood H., Morrissey J.P., Petfalski E. et Tollervey D. 1996. Recognition of cleavage site A2 in the yeast pre-rRNA. RNA, 2: 51-62.

Amberg D.C., Goldstein A.L. et Cole C.N. 1992. Isolation and characterization of RAT1: an essential gene of *Saccharomyces cerevisiae* required for the efficient nucleocytoplasmic trafficking of mRNA. Genes Dev., 6: 1173-1189.

Anderson G.J., Bach M., Lührmann R. et Beggs J.D. 1989. Conservation between yeast and man of a protein associated with U5 small nuclear ribonucleoprotein. Nature, 342: 819-821.

Andrew C., Hopper A.K. et Hall B.D. 1976. A yeast mutant defective in the processing of 27S r-RNA precursor. Mol. Gen. Genet., 144: 29-37.

Arenas J.E. et Abelson J.N. 1993. The *Saccharomyces cerevisiae* PRP21 gene product is an integral component of the prespliceosome. Proc. Natl. Acad. Sci. USA., 90: 6771-6775.

Ares M. et Igel A.H. 1990. Lethal and temperature-sensitive mutations and their suppressors identify an essential structural element in U2 small nuclear RNA. Genes Dev., 4: 2132-2145.

Aris J.P. et Blobel G. 1991. cDNA cloning and sequencing of human fibrillarin, a conserved nucleolar protein recognized by autoimmune antisera. Proc. Natl. Acad. Sci. USA, 88: 931-935.

Augustin S., Muller M.W. et Schweyen R.J. 1990. Reverse self-splicing of group II intron RNAs *in vitro*. Nature, 343: 383-386.

Bach M., Winkelmann G. et Lührmann R. 1989. 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. Proc. Natl. Acad. Sci. USA, 86: 6038-6042.

Bachellerie J.P., Michot B. et Raynal F. 1983. Recognition signals for mouse pre-rRNA processing. A potential role for U3 nucleolar RNA. Mol. Biol. Rep., 9: 79-86.

Bachellerie J.P., Michot B. Nicoloso M., Balakin A., Ni J. et Fournier M.J. 1995. Antisense snoRNAs : a family of nucleolar with long complementarities to rRNA. Trends Biochem., 20: 261-264.

Bakin A. et Ofengand J. 1993. Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. Biochemistry, 32: 9754-9762.

Balakin A.G., Lempicki R.A., Huang G.M. et Fournier M.J. 1994. *Saccharomyces cerevisiae* U14 small nuclear RNA has little secondary structure and appears to be produced by post-transcriptional processing. J. Biol. Chem., 269: 739-746.

Balakin A.G., Schneider G.S., Corbett M.S., Ni J. et Fournier M.J. 1993. SnR31, snR32, and snR33: three novel, non-essential snRNAs from *Saccharomyces cerevisiae*. Nucleic Acids Res., 21: 5391-5397.

Balakin A.G., Smith L. et Fournier M.J. 1996. The RNA world of the nucleolus: Two major families os small RNAs defined by different elements with related functions. Cell (sous presse).

Bally M., Hughes J. et Cesareni G. 1988. SnR30: a new, essential small nuclear RNA from *Saccharomyces cerevisiae*. Nucleic Acids Res., 16: 5291-5303.

Banroques J. et Abelson J.N. 1989. PRP4: a protein of the yeast U4/U6 small nuclear ribonucleoprotein particle. Mol. Cell. Biol., 9: 3710-3719.

Barton G.J. 1993. ALSCRIPT a tool to format multiple sequence alignments. Protein Eng., 6: 37-40.

Baserga S.J., Gilmore-Hebert M. et Yang X.W. 1992. Distinct molecular signals for nuclear import of the nucleolar snRNA, U3. Genes Dev., 6: 1120-1130.

Bayev A., Georgiev O.I., Hadjiolov A.A., Nikolaev N., Skryabin K.G. et Zakharyev V.M. 1981. The structure of the yeast ribosomal RNA genes. 3. Precise mapping of the 18 S and 25 S rRNA genes and structure of the adjacent regions. Nucleic Acids Res., 9: 789-799.

Behrens S.E. et Lührmann R. 1991. Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. Genes Dev., 5: 1439-1452.

Behrens S.E., Galisson F., Legrain P. et Lührmann R. 1993b. Evidence that the 60-kDa protein of 17S U2 small nuclear ribonucleoprotein is immunologically and functionally related to the yeast PRP9 splicing factor and is required for the efficient formation of prespliceosomes. Proc. Natl. Acad. Sci. USA, 90: 8229-8233.

Behrens S.E., Tyc K., Kastner B., Reichelt J. et Lührmann R. 1993a. Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. Mol. Cell. Biol., 13: 307-319.

Bell G.I., DeGennaro L.J., Gelfand D.H., Bishop R.J., Valenzuela P. et Rutter W.J. 1977. Ribosomal RNA genes of *Saccharomyces cerevisiae*. I. Physical map of the repeating unit and location of the regions coding for 5S, 5.8S, 18S and 25S ribosomal RNAs. J. Biol. Chem., 252: 8118-8125.

Beltrame M. et Tollervey D. 1992. Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. EMBO J., 11: 1531-42.

Beltrame M. et Tollervey D. 1995. Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. EMBO J., 14: 4350-4356.

Beltrame M., Henry Y. et Tollervey D. 1994. Mutational analysis of an essential binding site for the U3 snoRNA in the 5' external transcribed spacer of yeast pre-rRNA. Nucleic Acids Res., 22: 4057-4065.

Bennett J.L., Jeong-Yu S. et Clayton D.A. 1992c. Characterization of a *Xenopus laevis* ribonucleoprotein endoribonuclease. Isolation of the RNA component and its expression during development. J. Biol. Chem., 267: 21765-21772.

Bennett M., Michaud S., Kingston J. et Reed R. 1992b. Protein components specifically associated with prespliceosome and spliceosome complexes. Genes Dev., 6: 1986-2000.

Berges T., Petfalski E., Tollervey D. et Hurt E.C. 1994. Synthetic lethality with fibrillarin identifies NOP77p, a nucleolar protein required for pre-rRNA processing and modification. EMBO J. 13: 3136-3148.

Berget S.M. 1995. Exon recognition in vertebrate splicing. J. Biol. Chem., 270: 2411-2414.

Berget S.M., Moore C. et Sharp P.A. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. USA, 74: 3171-3175.

Bicknell J.N. et Douglas H.C. 1970. Nucleic acid homologies among species of *Saccharomyces*. J. Bacteriol., 101: 505-512.

Bindereif A. et Green M.R. 1987. An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. EMBO J., 6: 2415-2424.

Birnboim H.C. et Doly J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res., 7: 1513-1523.

Birney E., Kumar S. et Krainer A.R. 1993a. A putative homology of U2AF65 in *S. cerevisiae*. Nucleic Acids Res., 21: 1333.

Birney E., Kumar S. et Krainer A.R. 1993b. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res., 21: 5803-5816.

Bjorn S.P., Soltyk A., Beggs J.D. et Friesen J.D. 1989. PRP4 (RNA4) from *Saccharomyces cerevisiae*: its gene product is associated with the U4/U6 small nuclear ribonucleoprotein particle. Mol. Cell. Biol., 9: 3698-3709.

Black D.L. 1995. Finding splice sites within a wilderness of RNA. RNA, 1: 763-771.

Black D.L. et Pinto A.L. 1989. U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. Mol. Cell. Biol., 9: 3350-3359.

Black D.L., Chabot B. et Steitz J.A. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. Cell, 42: 737-750.

Blanz P.A., Unseld M.G. et Rauh I. 1989. Group-specific differences in the secondary structure of the 28S ribosomal RNA of yeasts. Yeast, 5 Spec No: S399-404.

Bolivar F., Rodriguez R.L., Betlach M.C. et Boyer H.W. 1977. Construction and characterization of new cloning vehicles. I. Ampicillin- resistant derivatives of the plasmid pMB9. Gene, 2: 75-93.

Bordonné R., Banroques J., Abelson J. et Guthrie C. 1990. Domains of yeast U4 spliceosomal RNA required for PRP4 protein binding, snRNP- snRNP interactions, and pre-mRNA splicing *in vivo*. Genes Dev., 4: 1185-96.

Branlant C., Krol A., Ebel J.P., Gallinaro H., Lazar E. et Jacob M. 1981. The conformation of chicken, rat and human U1A RNAs in solution. Nucleic Acids Res., 9: 841-58.

Branlant C., Krol A., Ebel J.P., Lazar E., Gallinaro H., Jacob M., Sri-Widada J. et Jeanteur P. 1980. Nucleotide sequences of nuclear U1A RNAs from chicken, rat and man. Nucleic Acids Res., 8: 4143-4154.

Branlant C., Krol A., Ebel J.P., Lazar E., Haendler B. et Jacob M. 1982. U2 RNA shares a structural domain with U1, U4, and U5 RNAs. EMBO J., 1: 1259-65.

Branlant C., Krol A., Lazar E., Haendler B., Jacob M., Galego-Dias L. et Pousada C. 1983. High evolutionary conservation of the secondary structure and of certain nucleotide sequences of U5 RNA. Nucleic Acids Res., 11: 8359-67.

Bringmann P. et Lührmann R. 1986. Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. EMBO J., 5: 3509-3516.

Bringmann P., Appel B., Rinke J., Reuter R., Theissen H. et Lührmann R. 1984. Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. EMBO J., 3: 1357-1363.

Brody E. et Abelson J. 1985. The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. Science, 228: 963-7.

Brosi R., Groning K., Behrens S.E., Lührmann R. et Kramer A. 1993. Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. Science, 262: 102-105.

Brow D.A. et Guthrie C. 1988. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. Nature, 334: 213-8.

Brow D.A. et Vidaver R.M. 1995. An element in human U6 RNA destabilizes the U4/U6 spliceosomal RNA complex. RNA, 1: 122-131.

Brown J.D. et Beggs J.D. 1992. Roles of PRP8 protein in the assembly of splicing complexes. EMBO J., 11: 3721-3729.

Brulé F., Grégoire A., Ségault V., Mougin A. et Branlant C. 1995. Secondary structure conservation of the U3 small nucleolar RNA introns in *Saccharomyces*. C. R. Acad. Sci. Paris, Sciences de la Vie/Life Sciences, 318: 1197-1206.

Brulé F., Venema J., Ségault V., Tollervey D. et Branlant C. 1996. The yeast *Hansenula wingei* U3 snoRNA gene contains an intron and its coding sequence co-evolved with the 5' ETS region of the pre-ribosomal RNA. RNA, 2: 183-197.

Brunel C., Widada J.S., Lelay M.N., Jeanteur P. et Liautard J.P. 1981. Purification and characterization of a simple ribonucleoprotein particle containing small nucleoplasmic RNAs (snRNP) as a subset of RNP containing heterogenous nuclear RNA (hnRNP) from HeLa cells. Nucleic Acids Res., 9: 815-830.

Burgess S., Couto J.R. et Guthrie C. 1990. A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. Cell, 60: 705-717.

Burgess S.M. et Guthrie C. 1993. A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. Cell, 73: 1377-1391.

Burke J.M., Belfort M., Cech T.R., Davies R.W., Schweyen R.J., Shub D.A., Szostak J.W. et Tabak H.F. 1987. Structural conventions for group I introns. Nucleic Acids Res., 15: 7217-7221.

Busch H., Reddy R., Rothblum L. et Choi Y.C. 1982. SnRNAs, SnRNPs, and RNA processing. Annu. Rev. Biochem., 51: 617-654.

Caceres J.F., Stamm S., Helfman D.M. et Krainer A.R. 1994. Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. Science, 265: 1706-1709.

Caffarelli E., Arese M., Santoro B., Fragapane P. et Bozzoni I. 1994. In vitro study of processing of the intron-encoded U16 small nucleolar RNA in Xenopus laevis. Mol. Cell. Biol., 14: 2966-2974.

Calvet J.P. et Pederson T. 1981. Base-pairing interactions between small nuclear RNAs and nuclear RNA precursors as revealed by psoralen cross-linking *in vivo*. Cell, 26: 363-370.

Campbell et Duffus 1988. The yeast pratical approach. IRL Press, IRL Press,

Campbell I. 1972. Numerical analysis of the genera *Saccharomyces* and *Kluyveromyces*. J. Gen. Microbiol., 73: 279-301.

Carter P., Bedouelle H. et Winter G. 1985. Improved oligonucleotide site-directed mutagenesis using M13 vectors. Nucleic Acids Res., 13: 4431-4443.

Cecconi F., Mariottini P., Loreni F., Pierandrei-Amaldi P., Campioni N. et Amaldi F. 1994. U17XS8, a small nucleolar RNA with a 12 nt complementarity to 18S rRNA and coded by a sequence repeated in the six introns of *Xenopus laevis* ribosomal protein S8 gene. Nucleic Acids Res., 22: 732-741.

Cech T.R. 1985. Self-splicing RNA: implications for evolution. Int. Rev. Cytol., 93: 3-22.

Cech T.R. 1986. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell, 44: 207-210.

Cech T.R. 1987. The chemistry of self-splicing RNA and RNA enzymes. Science, 236: 1532-1539.

Cech T.R. 1988. Conserved sequences and structures of group I introns: building an active site for RNA catalysis: a review. Gene, 73: 259-271.

Cellini A., Felder E. et Rossi J.J. 1986a. Yeast pre-messenger RNA splicing efficiency depends on critical spacing requirements between the branch point and 3' splice site. EMBO J., 5: 1023-1030.

Chanfreau G. et Jacquier A. 1994. Catalytic site components common to both splicing steps of a group II intron. Science, 266: 1383-1387.

Chanfreau G., Legrain P., Dujon B. et Jacquier A. 1994. Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3' splice site selection in *S. cerevisiae*. Nucleic Acids Res., 22: 1981-1987.

Chang D.D. et Clayton D.A. 1987. A novel endoribonuclease cleaves at a priming site of mouse mitochondrial DNA replication. EMBO J., 6: 409-417.

Chang T.H., Clark M.W., Lustig A.J., Cusick M.E. et Abelson J. 1988. RNA11 protein is associated with the yeast spliceosome and is localized in the periphery of the cell nucleus. Mol. Cell. Biol., 8: 2379-93.

Chebli K., Gattoni R., Schmitt P., Hildwein G. et Stevenin J. 1989. The 216-nucleotide intron of the E1A pre-mRNA contains a hairpin structure that permits utilization of unusually distant branch acceptors. Mol. Cell. Biol., 9: 4852-4861.

Chen J.H. et Lin R.J. 1990. The yeast PRP2 protein, a putative RNA-dependent ATPase, shares extensive sequence homology with two other pre-mRNA splicing factors. Nucleic Acids Res., 18: 6447.

Chen M.W., Anne J., Volckaert G., Huysmans E., Vandenberghe A. et De Wachter R. 1984. The nucleotide sequences of the 5 S rRNAs of seven molds and a yeast and their use in studying ascomycete phylogeny. Nucleic Acids Res., 12: 4881-4892.

Chen X., Chamorro M., Lee S.I., Shen L.X., Hines J.V., Tinoco I. et Varmus H.E. 1995. Structural and functional studies of retroviral RNA pseudoknots involved in ribosomal frameshifting: nucleotides at the junction of the two stems are important for efficient ribosomal frameshifting. EMBO J., 14: 842-852.

Cheng S.C. et Abelson J. 1987. Spliceosome assembly in yeast. Genes Dev., 1: 1014-1027.

Cheng S.C., Tarn W.Y., Tsao T.Y. et Abelson J. 1993. PRP19: a novel spliceosomal component. Mol. Cell. Biol., 13: 1876-1882.

Cheng Y., Dahlberg J.E. et Lund E. 1995. Diverse effects of the guanine nucleotide exchange factor RCC1 on RNA transport. Science, 267: 1807-1810.

Chevallier M.R., Bloch J.C. et Lacroute F. 1980. Transcriptional and translational expression of a chimeric bacterial-yeast plasmid in yeasts. Gene, 11: 11-19.

Chiara M.D., Champion-Arnaud P., Buvoli M., Nadal-Ginard B. et Reed R. 1994. Specific protein-protein interactions between the essential mammalian spliceosome-associated proteins SAP 61 and SAP 114. Proc. Natl. Acad. Sci. USA, 91: 6403-6407.

Choi Y.D. et Dreyfuss G. 1984. Isolation of the heterogeneous nuclear RNA-ribonucleoprotein complex (hnRNP): a unique supramolecular assembly. Proc Natl Acad Sci USA, 81: 7471-7475.

Chow L.T., Roberts J.M., Lewis J.B. et Broker T.R. 1977. A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. Cell, 11: 819-836.

Chu F.K., Maley G.F., West D.K., Belfort M. et Maley F. 1986. Characterization of the intron in the phage T4 thymidylate synthase gene and evidence for its self-excision from the primary transcript. Cell, 45: 157-166.

Chu S., Archer R.H., Zengel J.M. et Lindahl L. 1994. The RNA of RNase MRP is required for normal processing of ribosomal RNA. Proc Natl Acad Sci USA, 91: 659-663.

Clark M.W., Yip M.L., Campbell J. et Abelson J. 1990. SSB-1 of the yeast *Saccharomyces cerevisiae* is a nucleolar-specific, silver- binding protein that is associated with the snR10 and snR11 small nuclear RNAs. J. Cell. Biol., 111: 1741-51.

Clayton D.A. 1994. A nuclear function for RNase MRP. Proc. Natl. Acad. Sci. USA, 9192: 4615-4617.

Clouet d'Orval B., d'Aubenton Carafa Y., Sirand-Pugnet P., Gallego M., Brody E. et Marie J. 1991. RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. Science, 252: 1823-1828.

Colot H.V., Stutz F. et Rosbash M. 1996. The yeast splicing factor Mud13 is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. Genes Dev., 10: 1699-1708.

Company M., Arenas J. et Abelson J. 1991. Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. Nature, 349: 487-93.

Cooper M., Johnston L.H. et Beggs J.D. 1995. Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. EMBO J., 14: 2066-2075.

Couto J.R., Tamm J., Parker R. et Guthrie C. 1987. A trans-acting suppressor restores splicing of a yeast intron with a branch point mutation. Genes Dev., 1: 445-55.

Crispino J.D., Blencowe B.J. et Sharp P.A. 1994. Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. Science, 265: 1866-1869.

Crouch R.J., Kanaya S. et Earl P.L. 1983. A model for the involvement of the small nucleolar RNA (U3) in processing eukaryotic ribosomal RNA. Mol. Biol. Rep., 9: 75-78.

Cryer D.R., Eccleshall R. et Marmur J. 1975. Methods in Cell Biology. Prescott, D.M., Academic Press, Prescott, D.M., New York,

Dahlberg J.E. et Lund E. 1987. Structure and expression of U-snRNA genes. Mol. Biol. Rep., 12: 139-143.

Dalbadie-McFarland G. et Abelson J. 1990. PRP5: a helicase-like protein required for mRNA splicing in yeast. Proc. Natl. Acad. Sci. USA, 87: 4236-40.

Dalrymple M.A., Petersen-Bjorn S., Friesen J.D. et Beggs J.D. 1989. The product of the PRP4 gene of *S. cerevisiae* shows homology to beta subunits of G proteins. Cell, 58: 811-812.

Dandekar T. et Tollervey D. 1989. Cloning of *Schizosaccharomyces pombe* genes encoding the U1, U2, U3 and U4 snRNAs. Gene, 81: 227-235.

Dandekar T. et Tollervey D. 1993. Identification and functional analysis of a novel yeast small nucleolar RNA. Nucleic Acids Res., 21: 5386-90.

Dandekar T., Ribes V. et Tollervey D. 1989. *Schizosaccharomyces pombe* U4 small nuclear RNA closely resembles vertebrate U4 and is required for growth. J. Mol. Biol., 208: 371-9.

Das G., Henning D. et Reddy R. 1987. Structure, organization, and transcription of Drosophila U6 small nuclear RNA genes. J. Biol. Chem., 262: 1187-1193.

Datta B. et Weiner A.M. 1991. Genetic evidence for base pairing between U2 and U6 snRNA in mammalian mRNA splicing. Nature, 352: 821-824.

Davis D.R. et Poulter C.D. 1991. 1H-15N NMR studies of *Escherichia coli* tRNA(Phe) from hisT mutants: a structural role for pseudouridine. Biochemistry, 30: 4223-4231.

Deirdre A., Scadden J. et Smith C.W. 1995. Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. EMBO J., 14: 3236-3246.

Deshler J.O. et Rossi J.J. 1991. Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron. Genes Dev., 5: 1252-1263.

Domenjoud L., Gallinaro H., Kister L., Meyer S. et Jacob M. 1991. Identification of a specific exon sequence that is a major determinant in the selection between a natural and a cryptic 5' splice site. Mol. Cell. Biol., 11: 4581-4590.

Durovic P. et Dennis P.P. 1994. Separate pathways for excision and processing of 16S and 23S rRNA from the primary rRNA operon transcript from the hyperthermophilic archaebacterium Sulfolobus acidocaldarius: similarities to eukaryotic rRNA processing. Mol. Microbiol., 13: 229-242.

Ehrenman K., Pedersen-Lane J., West D., Herman R., Maley F. et Belfort M. 1986. Processing of phage T4 td-encoded RNA is analogous to the eukaryotic group I splicing pathway. Proc. Natl. Acad. Sci. USA, 83: 5875-5879.

Ehresmann C., Baudin F., Mougel M., Romby P., Ebel J.P. et Ehresmann B. 1987. Probing the structure of RNAs in solution. Nucleic Acids Res., 15: 9109-9128.

Elela S.A., Igel H. et Ares M. 1996. RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. EMBO J., 85: 115-124.

Eng F.J. et Warner J.R. 1991. Structural basis for the regulation of splicing of a yeast messenger RNA. Cell, 65: 797-804.

Eperon I.C., Ireland D.C., Smith R.A., Mayeda A. et Krainer A.R. 1993. Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. EMBO J., 12: 3607-3617.

Eperon L.P., Graham I.R., Griffiths A.D. et Eperon I.C. 1988. Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? Cell, 54: 393-401.

Epstein P., Reddy R. et Busch H. 1984. Multiple states of U3 RNA in Novikoff hepatoma nucleoli. Biochemistry, 23: 5421-5425.

Epstein P., Reddy R., Henning D. et Busch H. 1980. The nucleotide sequence of nuclear U6 (4.7 S) RNA. J. Biol. Chem., 255: 8901-8906.

Fabrizio P. et Abelson J. 1990a. Point mutations in yeast U6 snRNA can specifically block the first or second step of pre-mRNA splicing *in vitro*. Mol. Biol. Rep., 14: 135-140.`

Fabrizio P. et Abelson J. 1990b. Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing. Science, 250: 404-409.

Fabrizio P., Esser S., Kastner B. et Lührmann R. 1994. Isolation of *S. cerevisiae* snRNPs: comparison of U1 and U4/U6.U5 to their human counterparts. Science, 2645: 261-265.

Fabrizio P., McPheeters D.S. et Abelson J. 1989. *In vitro* assembly of yeast U6 snRNP: a functional assay. Genes Dev., 3: 2137-2150.

Feng D.F. et Doolittle R.F. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol., 25: 351-360.

Filipowicz W. et Kiss T. 1993. Structure and function of nucleolar snRNPs. Mol. Biol. Rep., 18: 149-156.

Fischer U., Sumpter V., Sekine M., Satoh T. et Lührmann R. 1993. Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. EMBO J., 12: 573-583.

Forne T., Rossi F., Labourier E., Antoine E., Cathala G., Brunel C. et Tazi J. 1995. Disruption of base-paired U4.U6 small nuclear RNAs induced by mammalian heterogeneous nuclear ribonucleoprotein C protein. J. Biol. Chem., 270: 16476-16481.

Fortner D.M., Troy R.G. et Brow D.A. 1994. A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. Genes Dev., 8: 221-233.

Fouser L.A. et Friesen J.D. 1986. Mutations in a yeast intron demonstrate the importance of specific conserved nucleotides for the two stages of nuclear mRNA splicing. Cell, 45: 81-93.

Fragapane P., Prislei S., Michienzi A., Caffarelli E. et Bozzoni I. 1993. A novel small nucleolar RNA (U16) is encoded inside a ribosomal protein intron and originates by processing of the pre-mRNA. EMBO J., 12: 2921-2928.

Frank D. et Guthrie C. 1992. An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. Genes Dev., 6: 2112-2124.

Frank D., Patterson B. et Guthrie C. 1992. Synthetic lethal mutations suggest interactions between U5 small nuclear RNA and four proteins required for the second step of splicing. Mol. Cell. Biol., 12: 5197-205.

Frank D.N., Roiha H. et Guthrie C. 1994. Architecture of the U5 small nuclear RNA. Mol. Cell. Biol., 14: 2180-90.

Frendewey D., Barta I., Gillespie M. et Potashkin J. 1990. Schizosaccharomyces U6 genes have a sequence within their introns that matches the B box consensus of tRNA internal promoters. Nucleic Acids Res., 18: 2025-2032.

Fu X.D. 1995. The superfamily of arginine/serine-rich splicing factors. RNA, 1: 663-680.

Fu X.D. et Maniatis T. 1990. Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. Nature, 343: 437-441.

Fu X.D. et Maniatis T. 1992. The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. Proc. Natl. Acad. Sci. USA, 89: 1725-1729.

Fu X.D., Mayeda A., Maniatis T. et Krainer A.R. 1992. General splicing factors SF2 and SC35 have equivalent activities *in vitro*, and both affect alternative 5' and 3' splice site selection. Proc. Natl. Acad. Sci. USA, 89: 11224-11228.

Fuson G.B., Price C.W. et Phaff H.J. 1979. Deoxyribonucleic acid sequence relatedness among some membres of the yeast genus *Hansenula*. Int. J. Syst. Bacteriol., 29: 64-69.

Galisson F. et Legrain P. 1993. The biochemical defects of prp4-1 and prp6-1 yeast splicing mutants reveal that the PRP6 protein is required for the accumulation of the [U4/U6.U5] tri-snRNP. Nucleic Acids Res., 21: 1555-1562.

Gallinaro H. et Jacob M. 1979. An evaluation of small nuclear RNA in hnRNP. FEBS Lett., 104: 176-182.

Gallinaro H., Puvion E., Kister L. et Jacob M. 1983. Nuclear matrix and hnRNP share a common structural constituent associated with premessenger RNA. EMBO J., 2: 953-960.

Garber R.C., Turgeon B.G., Selker E.U. et Yoder O.C. 1988. Organization of ribosomal RNA genes in the fungus *Cochliobolus heterostrophus*. Curr. Genet., 14: 573-582.

Garcia-Blanco M.A., Anderson G.J., Beggs J. et Sharp P.A. 1990. A mammalian protein of 22OkDa binds pre-mRNAs in the spliceosome: a potential homologue of the yeast PRP8 protein. Proc. Natl. Acad. Sci. USA. 87: 3082-3086.

Garrett R.A. et Rodriguez-Fonseca C. 1995. Ribosomal RNA: Structure, Evolution, Processing and Function. R. A. &. D. Zimmerman A. eds, Cleveland, Ohio.

Ge H., Zuo P. et Manley J.L. 1991. Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. Cell, 66: 373-382.

Ghetti A., Company M. et Abelson J. 1995. Specificity of Prp24 binding to RNA: a role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. RNA, 1: 132-145.

Ghetti A., Pinol-Roma S., Michael W.M., Morandi C. et Dreyfuss G. 1992. hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. Nucleic Acids Res., 20: 3671-3678.

Gibson T.J. 1984. Studies on the Epstein Barr virus genome. PhD Thesis. Université de Cambridge, Angleterre.

Girard J.P., Caizergues-Ferrer M. et Lapeyre B. 1993b. The SpGAR1 gene of Schizosaccharomyces pombe encodes the functional homologue of the snoRNP protein GAR1 of *Saccharomyces cerevisiae*. Nucleic Acids Res., 21: 2149-2155.

Girard J.P., Feliu J., Caizergues-Ferrer M. et Lapeyre B. 1993a. Study of multiple fibrillarin mRNAs reveals that 3' end formation in Schizosaccharomyces pombe is sensitive to cold shock. Nucleic Acids Res. 21: 1881-1887.

Girard J.P., Lehtonen H., Caizergues-Ferrer M., Amalric F., Tollervey D. et Lapeyre B. 1992. GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. EMBO J., 11: 673-82.

Gobert C., Bracco L., Rossi F., Olivier M., Tazi J., Lavelle F., Larsen A.K. et Riou J.F. 1996. Modulation of DNA topoisomerase I activity by p53. Biochemistry, 35: 5778-5786.

Goguel V. et Rosbash M. 1993. Splice site choice and splicing efficiency are positively influenced by pre- mRNA intramolecular base pairing in yeast. Cell, 72: 893-901.

Goguel V., Wang Y. et Rosbash M. 1993. Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing. Mol. Cell. Biol., 13: 6841-6848.

Goodall G.J. et Filipowicz W. 1991. Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. EMBO J., 10: 2635-2644.

Goodrich-Blair H., Scarlato V., Gott J.M., Xu M.Q. et Shub D.A. 1990. A self-splicing group I intron in the DNA polymerase gene of *Bacillus subtilis* bacteriophage SPO1. Cell, 63: 417-424.

Gozani O., Patton J.G. et Reed R. 1994. A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. EMBO J., 13: 3356-3367.

Grabowski P.J. et Sharp P.A. 1986. Affinity chromatography of splicing complexes: U2, U5, and U4 + U6 small nuclear ribonucleoprotein particles in the spliceosome. Science, 233: 1294-1299.

Grabowski P.J., Seiler S.R. et Sharp P.A. 1985. A multicomponent complex is involved in the splicing of messenger RNA precursors. Cell, 42: 345-353.

Green M.R. 1986. Pre-mRNA splicing. Annu. Rev. Genet., 20: 671-708.

Green M.R. 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu. Rev. Cell. Biol., 7: 559-599.

Gu J. et Reddy R. 1996. Small RNA database. Nucleic Acids Res., 24: 73-75.

Gui J.F., Lane W.S. et Fu X.D. 1994a. A serine kinase regulates intracellular localization of splicing factors in the cell cycle. Nature, 369: 678-682.

Gui J.F., Tronchere H., Chandler S.D. et Fu X.D. 1994b. Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. Proc. Natl. Acad. Sci. USA, 91: 10824-10828.

Guimont-Ducamp C., Sri-Widada J. et Jeanteur P. 1977. Occurrence of small molecular weight RNAs in Hela nuclear ribonucleoprotein particles containing HnRNA. Biochimie, 59: 755-758.

Gutell R.R., Larsen N. et Woese C.R. 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol. Rev., 58: 10-26.

Guthrie C. et Patterson B. 1988. Spliceosomal snRNAs. Annu. Rev. Genet., 22: 387-419.

Hackl W., Fischer U. et Lührmann R. 1994. A 69-kD protein that associates reversibly with the Sm core domain of several spliceosomal snRNP species. J. Cell. Biol., 1249: 261-272.

Hall S.L. et Padgett R.A. 1994. Conserved sequences in a class of rare eukaryotic nuclear introns with non- consensus splice sites. J. Mol. Biol., 239: 357-365.

Hall S.L. et Padgett R.A. 1996. Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. Science, 271: 1716-1718.

Harada F., Kato N. et Nishimura S. 1980. The nucleotide sequence of nuclear 4.8S RNA of mouse cells. Biochem. Biophys. Res. Comm., 95: 1332-1340.

Hartshorne T. et Agabian N. 1994. A common core structure for U3 small nucleolar RNAs. Nucleic Acids Res., 22: 3354-3364.

Hausner T.P., Giglio L.M. et Weiner A.M. 1990. Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. Genes Dev., 4: 2146-2156.

Hawkins J.D. 1988. A survey on intron and exon lengths. Nucleic Acids Res., 16: 9893-9908.

Hellung-Larsen P., Kulamowicz I. et Frederiksen S. 1980. Synthesis of low molecular weight RNA components in cells with a temperature- sensitive polymerase II. Biochim. Biophys. Acta., 609: 201-204.

Henry Y., Wood H., Morrissey J.P., Petfalski E., Kearsey S. et Tollervey D. 1994. The 5' end of yeast 5.8S rRNA is generated by exonucleases from an upstream cleavage site. EMBO J., 13: 2452-2463.

Hernandez N. et Keller W. 1983. Splicing of *in vitro* synthesized messenger RNA precursors in HeLa cell extracts. Cell, 35: 89-99.

Hinterberger M., Pettersson I. et Steitz J.A. 1983. Isolation of small nuclear ribonucleoproteins containing U1, U2, U4, U5, and U6 RNAs. J. Biol. Chem., 258: 2604-13.

Hoffman D.W., Query C.C., Golden B.L., White S.W. et Keene J.D. 1991. RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. Proc Natl Acad Sci USA, 88: 2495-2499.

Hogan J.J., Gutell R.R. et Noller H.F. 1984. Probing the conformation of 26S rRNA in yeast 60S ribosomal subunits with kethoxal. Biochemistry, 23: 3330-3335.

Horowitz D.S. et Abelson J. 1993. A U5 small nuclear ribonucleoprotein particle protein involved only in the second step of pre-mRNA splicing in *Saccharomyces cerevisiae*. Mol. Cell. Biol., 13: 2959-2970.

Huang G.M., Jarmolowski A., Struck J.C. et Fournier M.J. 1992. Accumulation of U14 small nuclear RNA in *Saccharomyces cerevisiae* requires box C, box D, and a 5', 3' terminal stem. Mol. Cell. Biol., 12: 4456-4463.

Hughes J.M. 1996. Functional base-pairing interaction between highly conserved elements of U3 small nucleolar RNA and the small ribosomal subunit. J. Mol. Biol., 259: 645-654.

Hughes J.M. et Ares M. 1991. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. EMBO J., 10: 4231-4239.

Hughes J.M., Konings D.A. et Cesareni G. 1987. The yeast homologue of U3 snRNA. EMBO J., 6: 2145-2155.

Igel A.H. et Ares M. 1988. Internal sequences that distinguish yeast from metazoan U2 snRNA are unnecessary for pre-mRNA splicing. Nature, 334: 450-453.

Innis M.A., Myambo K.B., Gelfand D.H. et Brow M.A. 1988. DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. Proc. Natl. Acad. Sci. USA, 85: 9436-9440.

Izaurralde E. et Mattaj I.W. 1995. RNA export. Cell, 81: 153-159.

Izaurralde E., Lewis J., Gamberi C., Jarmolowski A., McGuigan C. et Mattaj I.W. 1995. A cap-binding protein complex mediating U snRNA export. Nature, 376: 709-712.

Izaurralde E., Lewis J., McGuigan C., Jankowska M., Darzynkiewicz E. et Mattaj I.W. 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell, 78: 657-668.

Jackson S.P., Lossky M. et Beggs J.D. 1988. Cloning of the RNA8 gene of *Saccharomyces cerevisiae*, detection of the RNA8 protein, and demonstration that it is essential for nuclear pre-mRNA splicing. Mol. Cell. Biol., 8: 1067-1075.

Jacob M. et Gallinaro H. 1989. The 5' splice site: phylogenetic evolution and variable geometry of association with U1RNA. Nucleic Acids Res., 17: 2159-2180.

Jacquier A. et Michel F. 1987. Multiple exon-binding sites in class II self-splicing introns. Cell, 50: 17-29.

Jacquier A. et Michel F. 1990. Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. J. Mol. Biol., 213: 437-447.

Jacquier A., Rodriguez J.R. et Rosbash M. 1985. A quantitative analysis of the effects of 5' junction and TACTAAC box mutants and mutant combinations on yeast mRNA splicing. Cell, 43: 423-430.

Jamison S.F., Pasman Z., Wang J., Will C., Lührmann R., Manley J.L. et Garcia-Blanco M.A. 1995. U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements. Nucleic Acids Res., 23: 3260-3267.

Jansen R., Tollervey D. et Hurt E.C. 1993. A U3 snoRNP protein with homology to splicing factor PRP4 and G beta domains is required for ribosomal RNA processing. EMBO J., 12: 2549-2558.

Jansen R.P., Hurt E.C., Kern H., Lehtonen H., Carmo-Fonseca M., Lapeyre B. et Tollervey D. 1991. Evolutionary conservation of the human nucleolar protein fibrillarin and its functional expression in yeast. J. Cell. Biol., 113: 715-729.

Jarmolowski A., Zagorski J., Li H.V. et Fournier M.J. 1990. Identification of essential elements in U14 RNA of *Saccharomyces cerevisiae*. EMBO J., 9: 4503-4509.

Jeager J.A., Turner D.H. et Zucker M. 1989. Improved predictions of secondary structures for RNA. Proc. Natl. Acad. Sci. USA, 86: 7706-7710.

Jeppesen C., Stebbins-Boaz B. et Gerbi S.A. 1988. Nucleotide sequence determination and secondary structure of Xenopus U3 snRNA. Nucleic Acids Res., 16: 2127-2148.

Jones M.H., Frank D.N. et Guthrie C. 1995. Characterization and functional ordering of Slu7p and Prp17p during the second step of pre-mRNA splicing in yeast. Proc Natl Acad Sci USA, 92: 9687-9691.

Jong A.Y. et Campbell J.L. 1986. Isolation of the gene encoding yeast single-stranded nucleic acid binding protein 1. Proc Natl Acad Sci USA, 83: 877-881.

Jong A.Y., Clark M.W., Gilbert M., Oehm A. et Campbell J.L. 1987. *Saccharomyces cerevisiae* SSB1 protein and its relationship to nucleolar RNA- binding proteins. Mol. Cell. Biol., 7: 2947-2955.

Kaine B.P., Gupta R. et Woese C.R. 1983. Putative introns in tRNA genes of prokaryotes. Proc. Natl. Acad. Sci. USA, 80: 3309-3312.

Kalogeropoulos A. 1995. Automatic intron detection in nuclear DNA sequences of *Saccharomyces cerevisiae*. Yeast, 11: 555-565.

Kandels-Lewis S. et Séraphin B. 1993. Role of U6 snRNA in 5' splice site selection. Science, 262: 2035-2039.

Kanopka A., Muhlemann O. et Akusjarvi G. 1996. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. Nature, 381: 535-538.

Kao H.Y. et Siliciano P.G. 1996. Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. Mol. Cell. Biol., 16: 960-967.

Kass S. et Sollner-Webb B. 1990. The first pre-rRNA-processing event occurs in a large complex: analysis by gel retardation, sedimentation, and UV cross-linking. Mol. Cell. Biol., 10: 4920-4931.

Kass S., Craig N. et Sollner-Webb B. 1987. Primary processing of mammalian rRNA involves two adjacent cleavages and is not species specific. Mol. Cell. Biol., 7: 2891-2898.

Kass S., Tyc K., Steitz J.A. et Sollner-Webb B. 1990. The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. Cell, 60: 897-908.

Keller E.B. et Noon W.A. 1984. Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. Proc. Natl. Acad. Sci. USA, 81: 7417-7420.

Kempers-Veenstra A.E., Oliemans J., Offenberg H., Dekker A.F., Piper P.W., Planta R.J. et Klootwijk J. 1986. 3'-End formation of transcripts from the yeast rRNA operon. EMBO J., 5: 2703-2710.

Kenan D.J., Query C.C. et Keene J.D. 1991. RNA recognition: towards identifying determinants of specificity. Trends Biochem Sci, 16: 214-220.

Kim S.H., Smith J., Claude A. et Lin R.J. 1992. The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NTPase. EMBO J., 11: 2319-2326.

King D.S. et Beggs J.D. 1990. Interactions of PRP2 protein with pre-mRNA splicing complexes in *Saccharomyces cerevisiae*. Nucleic Acids Res., 18: 6559-6564.

Kiss T. et Filipowicz W. 1993. Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene RCC1. EMBO J., 12: 2913-2920.

Kiss T. et Solymosy F. 1990. Molecular analysis of a U3 RNA gene locus in tomato: transcription signals, the coding region, expression in transgenic tobacco plants and tandemly repeated pseudogenes. Nucleic Acids Res., 18: 1941-1949.

Kiss T., Marshallsay C. et Filipowicz W. 1991. Alteration of the RNA polymerase specificity of U3 snRNA genes during evolution and *in vitro*. Cell, 65: 517-526.

Kiss T., Marshallsay C. et Filipowicz W. 1992. 7-2/MRP RNAs in plant and mammalian cells: association with higher order structures in the nucleolus. EMBO J., 11: 3737-3746.

Kiss T., Toth M. et Solymosy F. 1985. Plant small nuclear RNAs. Nucleolar U3 snRNA is present in plants: partial characterization. Eur. J. Biochem., 152: 259-266.

Kiss-Laszlo Z., Henry Y., Bachellerie J.P., Caizergues-Ferrer M. et Kiss T. 1996. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. EMBO J., 85: 1077-1088.

Kjems J., Olesen S.O. et Garrett R.A. 1985. Comparison of eubacterial and eukaryotic 5S RNA structures: a chemical modification study. Biochemistry, 24: 241-250.

Klinz F.J. et Gallwitz D. 1985. Size and position of intervening sequences are critical for the splicing efficiency of pre-mRNA in the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res., 13: 3791-3804.

Koch J.L., Boulanger S.C., Dib-Hajj S.D., Hebbar S.K. et Perlman P.S. 1992. Group II introns deleted for multiple substructures retain self-splicing activity. Mol. Cell. Biol., 12: 1950-1958.

Kohtz J.D., Jamison S.F., Will C.L., Zuo P., Lührmann R., Garcia-Blanco M.A. et Manley J.L. 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature, 368: 119-124.

Konarska M.M. et Sharp P.A. 1986. Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. Cell, 46: 845-855.

Konarska M.M. et Sharp P.A. 1987. Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. Cell, 49: 763-774.

Konarska M.M., Grabowski P.J., Padgett R.A. et Sharp P.A. 1985. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature, 313: 552-557.

Kondo K. et Inouye M. 1992. Yeast NSR1 protein that has structural similarity to mammalian nucleolin is involved in pre-rRNA processing. J. Biol. Chem., 267: 16252-16258.

Kondo K., Kowalski L.R. et Inouye M. 1992. Cold shock induction of yeast NSR1 protein and its role in pre-rRNA processing. J. Biol. Chem., 267: 16259-16265.

Krainer A.R., Maniatis T., Ruskin B. et Green M.R. 1984. Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced *in vitro*. Cell, 36: 993-1005.

Krainer A.R., Mayeda A., Kozak D. et Binns G. 1991. Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. Cell, 66: 383-394.

Kramer B., Kramer W. et Fritz H.J. 1984b. Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*. Cell, 38: 879-887.

Krämer A. 1988. Presplicing complex formation requires two proteins and U2 snRNP. Genes Dev., 2: 1155-1167.

Krämer A., Keller W., Appel B. et Lührmann R. 1984a. The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. Cell, 38: 299-307.

Krämer A., Legrain P., Mulhauser F., Groning K., Brosi R. et Bilbe G. 1994. Splicing factor SF3a60 is the mammalian homologue of PRP9 of *S.cerevisiae*: the conserved zinc finger-like motif is functionally exchangeable *in vivo*. Nucleic Acids Res., 22: 5223-5228.

Krämer A., Mulhauser F., Wersig C., Groning K. et Bilbe G. 1995. Mammalian splicing factor SF3a120 represents a new member of the SURP family of proteins and is homologous to the essential splicing factor PRP21p of *Saccharomyces cerevisiae*. RNA, 1: 260-272.

Krawczak M., Reiss J. et Cooper D.N. 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum. Genet., 90: 41-54.

Kreger van Rij N.J.W. 1987. The yeasts, 2ème édition. Rose A.H. et Harrison J.S. eds. pp 5-61.

Kretzner L., Rymond B.C. et Rosbash M. 1987. *S. cerevisiae* U1 RNA is large and has limited primary sequence homology to metazoan U1 snRNA. Cell, 50: 593-602.

Krol A., Branlant C., Lazar E., Gallinaro H. et Jacob M. 1981a. Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs. Nucleic Acids Res., 9: 2699-716.

Krol A., Carbon P., Ebel J.P. et Appel B. 1987. *Xenopus tropicalis* U6 snRNA genes transcribed by Pol III contain the upstream promoter elements used by Pol II dependent U snRNA genes. Nucleic Acids Res., 15: 2463-2478.

Krol A., Ebel J.P., Rinke J. et Lührmann R. 1983. U1, U2 and U5 small nuclear RNAs are found in plants cells. Complete nucleotide sequence of the U5 RNA family from pea nuclei. Nucleic Acids Res., 11: 8583-8594.

Krol A., Gallinaro H., Lazar E., Jacob M. et Branlant C. 1981b. The nuclear 5S RNAs from chicken, rat and man. U5 RNAs are encoded by multiple genes. Nucleic Acids Res., 9: 769-787.

Kuhsel M.G., Strickland R. et Palmer J.D. 1990. An ancient group I intron shared by eubacteria and chloroplasts. Science, 250: 1570-1573.

Kunkel G.R., Maser R.L., Calvet J.P. et Pederson T. 1986. U6 small nuclear RNA is transcribed by RNA polymerase III. Proc. Natl. Acad. Sci. USA, 83: 8575-8579.

Kurtzman C.P. 1984. Synonymy of the yeast genera *Hansenula* and *Pichia* demonstrated through comparisons of deoxyribonucleic acid relatedness. Antonie Van Leeuwenhoek, 50: 209-217.

Kurtzman C.P. 1993. Systematics of the ascomycetous yeasts assessed from ribosomal RNA sequence divergence. Antonie Van Leeuwenhoek, 63: 165-174.

Kurtzman C.P. et Robnett C.J. 1991. Phylogenetic relationships among species of *Saccharomyces*, *Schizosaccharomyces*, *Debaryomyces* and *Schwanniomyces* determined from partial ribosomal RNA sequences. Yeast, 7: 61-72.

Kurtzman C.P., Smiley M.J. et Baker F.L. 1975. Scanning electron microscopy of ascospores of *Debaryomyces* and *Saccharomyces*. Mycopathol. Mycol. Appl., 55: 29-34.

Lacroute F. 1968. Regulation of pyrimidine biosynthesis in *Saccharomyces cerevisiae*. J. Bacteriol., 95: 824-832.

Laggerbauer B., Lauber J. et Lührmann R. 1996. Identification of an RNA-dependent ATPase activity in mammalian U5 snRNPs. Nucleic Acids Res., 24: 868-875.

Lamm G.M. et Lamond A.I. 1993. Non-snRNP protein splicing factors. Biochim. Biophys. Acta, 1173: 247-265.

Lamond A.I. 1993. The spliceosome. BioEssays, 15: 595-603.

Lamond A.I., Konarska M.M., Grabowski P.J. et Sharp P.A. 1988. Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. Proc. Natl. Acad. Sci. USA, 85: 411-415.

Langford C.J. et Gallwitz D. 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. Cell, 33: 519-527.

Langford C.J., Klinz F.J., Donath C. et Gallwitz D. 1984. Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. Cell, 36: 645-653.

Lapeyre B., Bourbon H. et Amalric F. 1987. Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. Proc Natl Acad Sci USA, 84: 1472-1476.

Lapeyre B., Mariottini P., Mathieu C., Ferrer P., Amaldi F., Amalric F. et Caizergues-Ferrer M. 1990. Molecular cloning of Xenopus fibrillarin, a conserved U3 small nuclear ribonucleoprotein recognized by antisera from humans with autoimmune disease. Mol. Cell. Biol., 10: 430-434.

Larpent J.P. et Larpent-Gourgaud M. 1985. Eléments de Microbiologie. Hermann ed. Editeurs des Sciences et des Arts.

Last R.L., Maddock J.R. et Woolford J.L. 1987. Evidence for related functions of the RNA genes of *Saccharomyces cerevisiae*. Genetics, 117: 619-631.

Leader D.J., Connelly S., Filipowicz W. et Brown J.W. 1994. Characterisation and expression of a maize U3 snRNA gene. Biochim. Biophys. Acta, 1219: 145-147.

Lee M.G., Lane D.P. et Beggs J.D. 1986. Identification of the RNA2 protein of *Saccharomyces cerevisiae*. Yeast, 2: 59-67.

Lee W.C., Zabetakis D. et Melese T. 1992. NSR1 is required for pre-rRNA processing and for the proper maintenance of steady-state levels of ribosomal subunits. Mol. Cell. Biol., 12: 3865-3871.

Legrain P. et Chanfreau G. 1994. Pre-mRNA splicing : from intron recognition to catalysis. Bull. Inst. Pasteur, 92: 153-179.

Legrain P. et Chapon C. 1993. Interaction between PRP11 and SPP91 yeast splicing factors and characterization of a PRP9-PRP11-SPP91 complex. Science, 262: 108-110.

Legrain P. et Choulika A. 1990. The molecular characterization of PRP6 and PRP9 yeast genes reveals a new cysteine/histidine motif common to several splicing factors. EMBO J., 9: 2775-2781.

Legrain P., Chapon C. et Galisson F. 1993. Interactions between PRP9 and SPP91 splicing factors identify a protein complex required in prespliceosome assembly. Genes Dev., 7: 1390-1399.

Legrain P., Séraphin B. et Rosbash M. 1988. Early commitment of yeast pre-mRNA to the spliceosome pathway. Mol. Cell. Biol., 8: 3755-3760.

Lerner M.R., Boyle J.A., Mount S.M., Wolin S.L. et Steitz J.A. 1980. Are snRNPs involved in splicing? Nature, 283: 220-224.

Lesser C.F. et Guthrie C. 1993. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. Science, 262: 1982-1988.

Leverette R.D., Andrews M.T. et Maxwell E.S. 1992. Mouse U14 snRNA is a processed intron of the cognate hsc70 heat shock pre- messenger RNA. Cell, 71: 1215-1221.

Lewis J.D., Izaurralde E., Jarmolowski A., McGuigan C. et Mattaj I.W. 1996. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. Genes Dev., 10: 1683-1698.

Li H., Dalal S., Kohler J., Vilardell J. et White S.A. 1995. Characterization of the pre-mRNA binding site for yeast ribosomal protein L32: the importance of a purine-rich internal loop. J. Mol. Biol., 250: 447-459.

Li H.D. et Fournier M.J. 1992. U14 function in *Saccharomyces cerevisiae* can be provided by large deletion variants of yeast U14 and hybrid mouse-yeast U14 RNAs. EMBO J., 11: 683-689.

Li H.D., Zagorski J. et Fournier M.J. 1990. Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in *Saccharomyces cerevisiae*. Mol. Cell. Biol., 10: 1145-1152.

Liao X.C., Tang J. et Rosbash M. 1993. An enhancer screen identifies a gene that encodes the yeast U1 snRNP A protein: implications for snRNP protein function in pre-mRNA splicing. Genes Dev., 7: 419-428.

Liautard J.P., Sri-Widada J., Brunel C. et Jeanteur P. 1982. Structural organization of ribonucleoproteins containing small nuclear RNAs from HeLa cells. Proteins interact closely with a similar structural domain of U1, U2, U4 and U5 small nuclear RNAs. J. Mol. Biol., 162: 623-643.

Libri D., Piseri A. et Fiszman M.Y. 1991. Tissue-specific splicing *in vivo* of the beta-tropomyosin gene: dependence on an RNA secondary structure. Science, 252: 1842-1845.

Libri D., Stutz F., McCarthy T. et Rosbash M. 1995. RNA structural patterns and splicing: molecular basis for an RNA-based enhancer. RNA, 1: 425-436.

Lin R.J., Lustig A.J. et Abelson J. 1987. Splicing of yeast nuclear pre-mRNA *in vitro* requires a functional 40S spliceosome and several extrinsic factors. Genes Dev., 1: 7-18.

Lin R.J., Newman A.J., Cheng S.C. et Abelson J. 1985. Yeast mRNA splicing *in vitro*. J. Biol. Chem., 260: 14780-14792.

Lindsey L.A., Crow A.J. et Garcia-Blanco M.A. 1995. A mammalian activity required for the second step of pre-messenger RNA splicing. J. Biol. Chem., 270: 13415-13421.

Lischwe M.A., Ochs R.L., Reddy R., Cook R.G., Yeoman L.C., Tan E.M., Reichlin M. et Busch H. 1985. Purification and partial characterization of a nucleolar scleroderma antigen (Mr = 34,000; pI, 8.5) rich in NG,NG-dimethylarginine. J. Biol. Chem., 260: 14304-14310.

Liu H.X., Goodall G.J., Kole R. et Filipowicz W. 1995. Effects of secondary structure on pre-mRNA splicing: hairpins sequestering the 5' but not the 3' splice site inhibit intron processing in *Nicotiana plumbaginifolia*. EMBO J., 14: 377-388.

Liu J. et Maxwell E.S. 1990. Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene. Nucleic Acids Res., 18: 6565-6571.

Lockhart S.R. et Rymond B.C. 1994. Commitment of yeast pre-mRNA to the splicing pathway requires a novel U1 small nuclear ribonucleoprotein polypeptide, Prp39p. Mol. Cell. Biol., 14: 3623-3633.

Lodder J. et Kreger van Rij N.J.W. 1952. The yeast, a taxonomic study. North-Holland Publishing Company, Amsterdam.

Lopato S., Mayeda A., Krainer A.R. et Barta A. 1996. Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors. Proc. Natl. Acad. Sci. USA, 93: 3074-3079.

Lossky M., Anderson G.J., Jackson S.P. et Beggs J. 1987. Identification of a yeast snRNP protein and detection of snRNP-snRNP interactions. Cell, 51: 1019-1026.

Lübben B., Marshallsay C., Rottmann N. et Lührmann R. 1993. Isolation of U3 snoRNP from CHO cells: a novel 55 kDa protein binds to the central part of U3 snoRNA. Nucleic Acids Res., 21: 5377-5385.

Lührmann R. 1990. Functions of U-snRNPs. Mol. Biol. Rep., 14: 183-192.

Lührmann R., Kastner B. et Bach M. 1990. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. Biochim. Biophys. Acta, 1087: 265-292.

Lygerou Z., Allmang C., Tollervey D. et Séraphin B. 1996. Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP *in vitro*. Science, 272: 268-270.

Lygerou Z., Mitchell P., Petfalski E., Séraphin B. et Tollervey D. 1994. The POP1 gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins. Genes Dev., 8: 1423-1433.

MacMillan A.M., Query C.C., Allerson C.R., Chen S., Verdine G.L. et Sharp P.A. 1994. Dynamic association of proteins with the pre-mRNA branch region. Genes Dev., 8: 3008-3020.

Maden B.E.H. 1990. The numerous modified nucleotides in eukaryotic ribosomal RNA. Prog. Nucleic Acids Res. Mol. Biol., 39: 241-303.

Madhani H.D. et Guthrie C. 1992. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell, 71: 803-817.

Madhani H.D. et Guthrie C. 1994a. Dynamic RNA-RNA interactions in the spliceosome. Annu. Rev. Genet., 28: 1-26.

Madhani H.D. et Guthrie C. 1994b. Randomization-selection analysis of snRNAs *in vivo*: evidence for a tertiary interaction in the spliceosome. Genes Dev., 8: 1071-1086.

Madhani H.D., Bordonne R. et Guthrie C. 1990. Multiple roles for U6 snRNA in the splicing pathway. Genes Dev., 4: 2264-2277.

Mancebo R., Lo P.C. et Mount S.M. 1990. Structure and expression of the *Drosophila melanogaster* gene for the U1 small nuclear ribonucleoprotein particle 70K protein. Mol. Cell. Biol., 10: 2492-2502.

Marshallsay C., Kiss T. et Filipowicz W. 1990. Amplification of plant U3 and U6 snRNA gene sequences using primers specific for an upstream promoter element and conserved intragenic regions. Nucleic Acids Res., 18: 3459-3466.

Maser R.L. et Calvet J.P. 1989. U3 small nuclear RNA can be psoralen-cross-linked *in vivo* to the 5' external transcribed spacer of pre-ribosomal-RNA. Proc. Natl. Acad. Sci. USA, 86: 6523-6527.

Mattaj I.W. 1986. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. Cell, 46: 905-911.

Mattaj I.W. et De Robertis E.M. 1985. Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. Cell, 40: 111-118.

Mattaj I.W., Habets W.J. et Van Venrooij W.J. 1986. Monospecific antibodies reveal details of U2 snRNP structure and interaction between U1 and U2 snRNPs. EMBO J., 5: 997-1002.

Maxwell E.S. et Fournier M.J. 1995. The small nucleolar RNA. Annu. Rev. Biochem., 67: 6523-6527.

Mayeda A., Helfman D.M. et Krainer A.R. 1993. Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. Mol. Cell. Biol., 13: 2993-3001.

Mayrand S.H., Fung P.A. et Pederson T. 1996. A discrete 3' region of U6 small nuclear RNA modulates the phosphorylation cycle of the C1 heterogeneous nuclear ribonucleoprotein particle protein. Mol. Cell. Biol., 16: 1241-1246.

Mazan S. et Bachellerie J.P. 1988. Structure and organization of mouse U3B RNA functional genes. J. Biol. Chem., 263: 19461-19467.

Mazan S. et Bachellerie J.P. 1990. Organization of the gene family encoding mouse U3B RNA: role of gene conversions in its concerted evolution. Gene, 94: 263-272.

Mermoud J.E., Cohen P.T. et Lamond A.I. 1994. Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. EMBO J., 13: 5679-5688.

Messing J., Gronenborn B., Muller-Hill B. et Hans Hopschneider P. 1977. Filamentous coliphage M13 as a cloning vehicle: insertion of a HindII fragment of the lac regulatory region in M13 replicative form *in vitro*. Proc. Natl. Acad. Sci. USA, 74: 3642-3646.

Michaud S. et Reed R. 1991. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev., 5: 2534-2546.

Michaud S. et Reed R. 1993. A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. Genes Dev., 7: 1008-1020.

Michel F. et Ferat J.L. 1995. Structure and activities of group II introns. Annu. Rev. Biochem., 64: 435-461.

Michel F. et Jacquier A. 1987. Long-range intron-exon and intron-intron pairings involved in self-splicing of class II catalytic introns. Cold Spring Harbor Symp. Quant. Biol., 52: 201-212.

Michel F. et Westhof E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J. Mol. Biol., 216: 585-610.

Michel F., Jacquier A. et Dujon B. 1982. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochimie, 64: 867-881.

Michel F., Umesono K. et Ozeki H. 1989. Comparative and functional anatomy of group II catalytic introns--a review. Gene, 82: 5-30.

Michot B. et Bachellerie J.P. 1991. Secondary structure of the 5' external transcribed spacer of vertebrate pre- rRNA. Presence of phylogenetically conserved features. Eur J Biochem, 195: 601-609.

Michot B., Hassouna N. et Bachellerie J.P. 1984. Secondary structure of mouse 28S rRNA and general model for the folding of the large rRNA in eukaryotes. Nucleic Acids Res., 12: 4259-4279.

Miller A.M. 1984. The yeast MATa1 gene contains two introns. EMBO J., 3: 1061-1065.

Miller K.G. et Sollner-Webb B. 1981. Transcription of mouse rRNA genes by RNA polymerase I: *in vitro* and *in vivo* initiation and processing sites. Cell, 27: 165-174.

Mizuta K., Hashimoto T. et Otaka E. 1992. Yeast ribosomal proteins: XIII. *Saccharomyces cerevisiae* YL8A gene, interrupted with two introns, encodes a homolog of mammalian L7. Nucleic Acids Res., 20: 1011-1016.

Moenne A., Camier S., Anderson G., Margottin F., Beggs J. et Sentenac A. 1990. The U6 gene of *Saccharomyces cerevisiae* is transcribed by RNA polymerase C (III) *in vivo* and *in vitro*. EMBO J., 9: 271-277.

Moore M.J. 1996. When the junk isn't junk. Nature, 379: 402-403.

Moore M.J., Querry C.C. et Sharp P.A. 1993. Splicing of precursors to messenger RNAs by the spliceosome. In The RNA World: 303-357.

Mörl M. et Schmelzer C. 1990. Integration of group II intron bI1 into a foreign RNA by reversal of the self- splicing reaction *in vitro*. Cell, 60: 629-636.

Mörl M., Niemer I. et Schmelzer C. 1992. New reactions catalyzed by a group II intron ribozyme with RNA and DNA substrates. Cell, 70: 803-810.

Morrissey J.P. et Tollervey D. 1993. Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. Mol. Cell. Biol. 13: 2469-2477.

Mougey E.B., Pape L.K. et Sollner-Webb B. 1993. A U3 small nuclear ribonucleoprotein-requiring processing event in the 5' external transcribed spacer of Xenopus precursor rRNA. Mol. Cell. Biol., 13: 5990-5998.

Mougin A., Grégoire A., Banroques J., Ségault V., Fournier R., Brulé F., Chevrier-Miller M. et Branlant C. 1996. The secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. RNA, (sous presse).

Mount S.M. 1982. A catalogue of splice junction sequences. Nucleic Acids Res., 10: 459-472.

Mount S.M. 1983. RNA processing. Sequences that signal where to splice. Nature, 304: 309-310.

Mount S.M. 1996. AT-AC introns: an ATtACk on dogma. Science, 271: 1690-1692.

Mount S.M. et Steitz J.A. 1981. Sequence of U1 RNA from *Drosophila melanogaster*: implications for U1 secondary structure and possible involvement in splicing. Nucleic Acids Res., 9: 6351-6368.

Mount S.M., Pettersson I., Hinterberger M., Karmas A. et Steitz J.A. 1983. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site *in vitro*. Cell, 33: 509-518.

Musters W., Venema J., van der Linden G., van Heerikhuizen H., Klootwijk J. et Planta R.J. 1989. A system for the analysis of yeast ribosomal DNA mutations. Mol. Cell. Biol., 9: 551-9.

Myslinski E. et Branlant C. 1991. A phylogenetic study of U4 snRNA reveals the existence of an evolutionarily conserved secondary structure corresponding to 'free' U4 snRNA. Biochimie, 73: 17-28.

Myslinski E., Segault V. et Branlant C. 1990. An intron in the genes for U3 small nucleolar RNAs of the yeast *Saccharomyces cerevisiae*. Science, 247: 1213-1216.

Nag M.K., Thai T.T., Ruff E.A., Selvamurugan N., Kunnimalaiyaan M. et Eliceiri G.L. 1993. Genes for E1, E2, and E3 small nucleolar RNAs. Proc. Natl. Acad. Sci. USA, 90: 9001-9005.

Nagai K., Oubridge C., Jessen T.H., Li J. et Evans P.R. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. Nature, 348: 515-520.

Nandabalan K., Price L. et Roeder G.S. 1993. Mutations in U1 snRNA bypass the requirement for a cell type-specific RNA splicing factor. Cell, 73: 407-415.

Needleman S.B. et Wunsch C.D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol., 48: 443-453.

Neefs J.M., Van de Peer Y., De Rijk P., Chapelle S. et De Wachter R. 1993. Compilation of small ribosomal subunit RNA structures. Nucleic Acids Res., 21: 3025-3049.

Nelissen R.L., Will C.L., Van Venrooij W.J. et Lührmann R. 1994. The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. EMBO J., 13: 4113-4125.

Nelson K.K. et Green M.R. 1990. Mechanism for cryptic splice site activation during pre-mRNA splicing. Proc. Natl. Acad. Sci. USA, 87: 6253-6257.

Newman A. et Norman C. 1991. Mutations in yeast U5 snRNA alter the specificity of 5' splice-site cleavage. Cell, 65: 115-123.

Newman A.J. et Norman C. 1992. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. Cell, 68: 743-754.

Newman A.J., Lin R.J., Cheng S.C. et Abelson J. 1985. Molecular consequences of specific intron mutations on yeast mRNA splicing *in vivo* and *in vitro*. Cell, 42: 335-344.

Nicoloso M., Caizergues-Ferrer M., Michot B., Azum M.C. et Bachellerie J.P. 1994. U20, a novel small nucleolar RNA, is encoded in an intron of the nucleolin gene in mammals. Mol. Cell. Biol., 14: 5766-5776.

Nicoloso M., Qu L.H., Michot B. et Bachellerie J.P. 1996. Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2'-O-ribose methylation of rRNAs. J. Mol. Biol., 260: 178-195.

Niwa M. et Berget S.M. 1991. Polyadenylation precedes splicing *in vitro*. Gene Expres., 1: 5-14.

Nogi Y., Vu L. et Nomura M. 1991. An approach for isolation of mutants defective in 35S ribosomal RNA synthesis in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA, 88: 7026-7030.

Noller H.F. 1984. Structure of ribosomal RNA. Annu. Rev. Biochem., 53: 119-162.

Ochs R.L., Lischwe M.A., Spohn W.H. et Busch H. 1985. Fibrillarin: a new protein of the nucleolus identified by autoimmune sera. Biol Cell, 54: 123-133.

Ofengand J., Bakin A., Wrzesinski J., Nurse K. et Lane B.G. 1995. The pseudouridine residues of ribosomal RNA. Biochem. Cell. Biol., 73: 915-924.

Ohno M., Sakamoto H. et Shimura Y. 1987. Preferential excision of the 5' proximal intron from mRNA precursors with two introns as mediated by the cap structure. Proc. Natl. Acad. Sci. USA, 84: 5187-5191.

Orum H., Nielsen H. et Engberg J. 1993. Sequence and proposed secondary structure of the Tetrahymena thermophila U3- snRNA. Nucleic Acids Res., 21: 2511.

Padgett R.A., Hardy S.F. et Sharp P.A. 1983. Splicing of adenovirus RNA in a cell-free transcription system. Proc. Natl. Acad. Sci. USA, 80: 5230-5234.

Padgett R.A., Konarska M.M., Grabowski P.J., Hardy S.F. et Sharp P.A. 1984. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. Science, 225: 898-903.

Parker K.A. et Steitz J.A. 1987. Structural analysis of the human U3 ribonucleoprotein particle reveal a conserved sequence available for base pairing with pre-rRNA. Mol. Cell. Biol., 7: 2899-913.

Parker K.A., Bruzik J.P. et Steitz J.A. 1988. An *in vitro* interaction between the human U3 snRNP and 28S rRNA sequences near the alpha-sarcin site. Nucleic Acids Res., 16: 10493-509.

Parker R. et Patterson B. 1987. Architecture of fungal introns: Implications for spliceosome assembly. New York. Inouye M. and Dubock B. eds.

Parker R. et Siliciano P.G. 1993. Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. Nature, 361: 660-662.

Parker R., Siliciano P.G. et Guthrie C. 1987. Recognition of the TACTAAC box during mRNA splicing in yeast

involves base pairing to the U2-like snRNA. Cell, 49:229-239.

Patterson B. et Guthrie C. 1987. An essential yeast snRNA with a U5-like domain is required for splicing *in vivo*. Cell, 49: 613-624.

Patton J.G., Mayer S.A., Tempst P. et Nadal-Ginard B. 1991. Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. Genes Dev., 5: 1237-1251.

Peculis B.A. et Steitz J.A. 1993. Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. Cell, 73: 1233-1245.

Peculis B.A. et Steitz J.A. 1994. Sequence and structural elements critical for U8 snRNP function in *Xenopus* oocytes are evolutionarily conserved. Genes Dev., 8: 2241-2255.

Peebles C.L., Gegenheimer P. et Abelson J. 1983. Precise excision of intervening sequences from precursor tRNAs by a membrane-associated yeast endonuclease. Cell, 32: 525-536.

Peebles C.L., Perlman P.S., Mecklenburg K.L., Petrillo M.L., Tabor J.H., Jarrell K.A. et Cheng H.L. 1986. A self-splicing RNA excises an intron lariat. Cell, 44: 213-223.

Pellizzoni L., Crosio C., Campioni N., Loreni F. et Pierandrei-Amaldi P. 1994. Different forms of U15 snoRNA are encoded in the introns of the ribosomal protein S1 gene of Xenopus laevis. Nucleic Acids Res., 22: 4607-4613.

Petes T.D. 1979. Yeast ribosomal DNA genes are located on chromosome XII. Proc. Natl. Acad. Sci. USA, 76: 410-414.

Pikielny C.W., Teem J.L. et Rosbash M. 1983. Evidence for the biochemical role of an internal sequence in yeast nuclear mRNA introns: implications for U1 RNA and metazoan mRNA splicing. Cell, 34: 395-403.

Plessel G., Fischer U. et Lührmann R. 1994. m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) *in vitro*: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. Mol. Cell. Biol., 14: 4160-4172.

Porter G.L., Brennwald P.J., Holm K.A. et Wise J.A. 1988. The sequence of U3 from *Schizosaccharomyces pombe* suggests structural divergence of this snRNA between metazoans and unicellular eukaryotes [published erratum appears in Nucleic Acids Res 1991 Jun 25;19(12):3484]. Nucleic Acids Res., 16: 10131-10152.

Potter S., Durovic P. et Dennis P.P. 1995. Ribosomal RNA precursor processing by a eukaryotic U3 small nucleolar RNA-like molecule in an archaeon. Science, 268: 1056-1060.

Prislei S., Michienzi A., Presutti C., Fragapane P. et Bozzoni I. 1993. Two different snoRNAs are encoded in introns of amphibian and human L1 ribosomal protein genes. Nucleic Acids Res., 21: 5824-5830.

Qu L.H., Henry Y., Nicoloso M., Michot B., Azum M.C., Renalier M.H., Caizergues-Ferrer M. et Bachellerie J.P. 1995. U24, a novel intron-encoded small nucleolar RNA with two 12 nt long, phylogenetically conserved complementarities to 28S rRNA. Nucleic Acids Res., 23: 2669-2676.

Qu L.H., Michot B. et Bachellerie J.P. 1983. Improved methods for structure probing in large RNAs : a rapid "heterologous" sequencing approach is coupled is coupled to the direct mapping of nuclease accessible sites. Application to the 5' terminal domain of eukaryotic 28S rRNA. Nucleic Acids Res., 11: 5903-5920.

Qu L.H., Nicoloso M., Michot B., Azum M.C., Caizergues-Ferrer M., Renalier M.H. et Bachellerie J.P. 1994. U21, a novel small nucleolar RNA with a 13 nt. complementarity to 28S rRNA, is encoded in an intron of ribosomal protein L5 gene in chicken and mammals. Nucleic Acids Res., 22: 4073-4081.

Query C.C., Bentley R.C. et Keene J.D. 1989a. A specific 31-nucleotide domain of U1 RNA directly interacts with the 70K small nuclear ribonucleoprotein component. Mol. Cell. Biol., 9: 4872-4881.

Query C.C., Bentley R.C. et Keene J.D. 1989b. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. Cell, 57: 89-101.

Query C.C., Moore M.J. et Sharp P.A. 1994. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev., 8: 587-597.

Query C.C., Strobel S.A. et Sharp P.A. 1996. Three recognition events at the branch-site adenine. EMBO J., 15: 1392-1402.

Raué H.A. et Planta R.J. 1991. Ribosome biogenesis in yeast. Prog. Nucleic Acid Res. Mol. Biol., 41: 89-129.

Raué H.A. et Planta R.J. 1995. The pathway to maturity: Processing of ribosomal RNA in *Saccharomyces cerevisiae*. Gene expression., 5: 71-77.

Reddy R. 1988. Compilation of small RNA sequences. Nucleic Acids Res, 16 Suppl: r71-85.

Reddy R. et Busch H. 1983. Small nuclear RNAs and RNA processing. Prog. Nucleic Acid Res. Mol. Biol., 30: 127-162.

Reddy R., Henning D. et Busch H. 1980. Substitutions, insertions, and deletions in two highly conserved U3 RNA species. J. Biol. Chem., 255: 7029-7033.

Reddy R., Henning D., Das G., Harless M. et Wright D. 1987. The capped U6 small nuclear RNA is transcribed by RNA polymerase III. J. Biol. Chem., 262: 75-81.

Reed R. 1989. The organization of 3' splice-site sequences in mammalian introns. Genes Dev., 3: 2113-2123.

Reed R. 1990. Protein composition of mammalian spliceosomes assembled in vitro. Proc. Natl. Acad. Sci. USA, 87: 8031-8035.

Reich C. et Wise J.A. 1990. Evolutionary origin of the U6 small nuclear RNA intron. Mol. Cell. Biol., 10: 5548-5552.

Reich C.I., VanHoy R.W., Porter G.L. et Wise J.A. 1992. Mutations at the 3' splice site can be suppressed by compensatory base changes in U1 snRNA in fission yeast. Cell, 69: 1159-1169.

Renalier M.H., Nicoloso M., Qu L.H. et Bachellerie J.P. 1996. SnoRNA U21 is also intron-encoded in Drosophila melanogaster but in a different host-gene as compared to warm-blooded vertebrates. FEBS Lett., 379: 212-216.

Retel J. et Planta R.J. 1970. On the mechanism of the biosynthesis of ribosomal RNA in yeast. Biochim. Biophys. Acta, 224: 458-469.

Riedel N., Wise J.A., Swerdlow H., Mak A. et Guthrie C. 1986. Small nuclear RNAs from *Saccharomyces cerevisiae*: unexpected diversity in abundance, size, and molecular complexity. Proc. Natl. Acad. Sci. USA, 83: 8097-8101.

Rimoldi O.J., Raghu B., Nag M.K. et Eliceiri G.L. 1993. Three new small nucleolar RNAs that are psoralen cross-linked *in vivo* to unique regions of pre-rRNA. Mol. Cell. Biol., 13: 4382-4390.

Rinke J., Appel B., Blocker H., Frank R. et Lührmann R. 1984. The 5'-terminal sequence of U1 RNA complementary to the consensus 5' splice site of hnRNA is single-stranded in intact U1 snRNP particles. Nucleic Acids Res., 12: 4111-4126.

Rinke J., Appel B., Digweed M. et Lührmann R. 1985. Localization of a base-paired interaction between small nuclear RNAs U4 and U6 in intact U4/U6 ribonucleoprotein particles by psoralen cross-linking. J. Mol. Biol., 185: 721-731.

Ripmaster T.L., Vaughn G.P. et Woolford J.L. 1992. A putative ATP-dependent RNA helicase involved in *Saccharomyces cerevisiae* ribosome assembly. Proc. Natl. Acad. Sci. USA, 89: 11131-11135.

Roberts R.J. 1978. Intervening sequences excised *in vitro*. Nature, 274: 530

Rogers J. et Wall R. 1980. A mechanism for RNA splicing. Proc. Natl. Acad. Sci. USA, 77: 1877-1879.

Roiha H., Shuster E.O., Brow D.A. et Guthrie C. 1989. Small nuclear RNAs from budding yeasts: phylogenetic comparisons reveal extensive size variation. Gene, 82: 137-144.

Roscigno R.F. et Garcia-Blanco M.A. 1995. SR proteins escort the U4/U6-U5 tri-snRNP to the spliceosome. RNA, 1: 692-706.

Roy J., Zheng B., Rymond B.C. et Woolford J.L. 1995. Structurally related but functionally distinct yeast Sm D core small nuclear ribonucleoprotein particle proteins. Mol. Cell. Biol., 15: 445-455.

Rubtsov P.M., Musakhanov M.M., Zakharyev V.M., Krayev A.S., Skryabin K.G. et Bayev A.A. 1980. The structure of the yeast ribosomal RNA genes. I. The complete nucleotide sequence of the 18S ribosomal RNA gene from *Saccharomyces cerevisiae*. Nucleic Acids Res., 8: 5779-5794.

Ruby S.W. et Abelson J. 1991. Pre-mRNA splicing in yeast. Trends Genet., 7: 79-85.

Ruby S.W., Chang T.H. et Abelson J. 1993. Four yeast spliceosomal proteins (PRP5, PRP9, PRP11, and PRP21) interact to promote U2 snRNP binding to pre-mRNA. Genes Dev., 7: 1909-1925.

Ruff E.A., Rimoldi O.J., Raghu B. et Eliceiri G.L. 1993. Three small nucleolar RNAs of unique nucleotide sequences. Proc Natl Acad Sci USA, 90: 635-638.

Ruskin B. et Green M.R. 1985a. Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing. Nature, 317: 732-734.

Ruskin B. et Green M.R. 1985b. Specific and stable intron-factor interactions are established early during *in vitro* pre-mRNA splicing. Cell, 43: 131-142.

Ruskin B., Krainer A.R., Maniatis T. et Green M.R. 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. Cell, 38: 317-331.

Ruskin B., Zamore P.D. et Green M.R. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell, 52: 207-219.

Russell I.D. et Tollervey D. 1992. NOP3 is an essential yeast protein which is required for pre-rRNA processing. J. Cell. Biol., 119: 737-747.

Russell I.D. et Tollervey D. 1995. Yeast Nop3p has structural and functional similarities to mammalian pre-mRNA binding proteins. Eur. J. Cell. Biol., 66: 293-301.

Rymond B.C. et Rosbash M. 1985. Cleavage of 5' splice site and lariat formation are independent of 3' splice site in yeast mRNA splicing. Nature, 317: 735-737.

Rymond B.C. et Rosbash M. 1992. Yeast pre-mRNA splicing. Jones E.W. eds, New York,

Sachs A.B. et Davis R.W. 1990. Translation initiation and ribosomal biogenesis: involvement of a putative rRNA helicase and RPL46. Science, 247: 1077-1079.

Samarsky D.A., Balakin A.G. et Fournier M.J. 1995. Characterization of three new snRNAs from *Saccharomyces cerevisiae*: snR34, snR35 and snR36. Nucleic Acids Res., 23: 2548-2554.

Sambrook J., Fritsch E.F. et Maniatis T. 1989. Molecular Cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, New York,

Sanger F., Nicklen S. et Coulson A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA, 74: 5463-5467.

Savino R. et Gerbi S.A. 1990. In vivo disruption of Xenopus U3 snRNA affects ribosomal RNA processing. EMBO J., 9: 2299-2308.

Sawa H. et Abelson J. 1992. Evidence for a base-pairing interaction between U6 small nuclear RNA and 5' splice site during the splicing reaction in yeast. Proc. Natl. Acad. Sci. USA, 89: 11269-11273.

Scherly D., Boelens W., Dathan N.A., van Venrooij W.J. et Mattaj I.W. 1990. Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B" and their cognate RNAs. Nature, 345: 502-506.

Scherly D., Boelens W., van Venrooij W.J., Dathan N.A., Hamm J. et Mattaj I.W. 1989. Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. EMBO J., 8: 4163-4170.

Schimmang T., Tollervey D., Kern H., Frank R. et Hurt E.C. 1989. A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. EMBO J., 8: 4015-24.

Schmitt M.E. et Clayton D.A. 1992. Yeast site-specific ribonucleoprotein endoribonuclease MRP contains an RNA component homologous to mammalian RNase MRP RNA and essential for cell viability. Genes Dev., 6: 1975-1985.

Schmitt M.E. et Clayton D.A. 1993. Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in *Saccharomyces cerevisiae*. Mol. Cell. Biol., 13: 7935-7941.

Schmitt M.E. et Clayton D.A. 1994. Characterization of a unique protein component of yeast RNase MRP: an RNA-binding protein with a zinc-cluster domain. Genes Dev., 8: 2617-2628.

Schwer B. et Guthrie C. 1991. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. Nature, 349: 494-499.

Schwer B. et Guthrie C. 1992. A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. EMBO J., 11: 5033-9.

Ségault V., Mougin A., Grégoire A., Banroques J. et Branlant C. 1992. An experimental study of *Saccharomyces cerevisiae* U3 snRNA conformation in solution. Nucleic Acids Res., 20: 3443-3451.

Ségault V., Will C.L., Sproat B.S. et Lührmann R. 1995. *In vitro* reconstitution of mammalian U2 and U5 snRNPs active in splicing: Sm proteins are functionally interchangeable and are essential for the formation of functional U2 and U5 snRNPs. EMBO J., 14: 4010-4021.

Selinger D.A., Porter G.L., Brennwald P.J. et Wise J.A. 1992. The two similarly expressed genes encoding U3 snRNA in *Schizosaccharomes pombe* lack introns. Mol. Biol. Evol., 9: 297-308.

Séraphin B. 1995. Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. EMBO J., 14: 2089-2098.

Séraphin B. et Kandels-Lewis S. 1993. 3' splice site recognition in *S. cerevisiae* does not require base pairing with U1 snRNA. Cell, 73: 803-812.

Séraphin B. et Rosbash M. 1989. Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell, 59: 349-358.

Séraphin B. et Rosbash M. 1990. Exon mutations uncouple 5' splice site selection from U1 snRNA pairing. Cell, 63: 619-629.

Séraphin B., Kretzner L. et Rosbash M. 1988. A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. EMBO J., 7: 2533-2538.

Shanab G.M. et Maxwell E.S. 1991. Proposed secondary structure of eukaryotic U14 snRNA. Nucleic Acids Res., 19: 4891-4894.

Shanab G.M. et Maxwell E.S. 1992. Determination of the nucleotide sequences in mouse U14 small nuclear RNA and 18S ribosomal RNA responsible for *in vitro* intermolecular base-pairing. Eur. J. Biochem., 206: 391-400.

Shannon K.W. et Guthrie C. 1991. Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA- binding motifs. Genes Dev., 5: 773-785.

Shapiro M.B. et Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res., 15: 7155-7174.

Shimba S., Buckley B., Reddy R., Kiss T. et Filipowicz W. 1992. Cap structure of U3 small nucleolar RNA in animal and plant cells is different. gamma-Monomethyl phosphate cap structure in plant RNA. J. Biol. Chem., 267: 13772-13777.

Short J.M., Fernandez J.A., Sorge J.A. et Huse W.D. 1988. Lambda ZAP : A bacteriophage Lambda expression vector with in vivo eexcision properties. Nucl. Acids Res., 16: 7583.

Shuai K. et Warner J.R. 1991. A temperature sensitive mutant of *Saccharomyces cerevisiae* defective in pre- rRNA processing. Nucleic Acids Res., 19: 5059-5064.

Siliciano P.G. et Guthrie C. 1988. 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. Genes Dev., 2: 1258-1267.

Siliciano P.G., Brow D.A., Roiha H. et Guthrie C. 1987a. An essential snRNA from *S. cerevisiae* has properties predicted for U4, including interaction with a U6-like snRNA. Cell, 50: 585-592.

Singh R. et Reddy R. 1989. Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. Proc. Natl. Acad. Sci. USA, 86: 8280-8283.

Sirand-Pugnet P., Durosay P., Clouet d'Orval B.C., Brody E. et Marie J. 1995. beta-Tropomyosin pre-mRNA folding around a muscle-specific exon interferes with several steps of spliceosome assembly. J. Mol. Biol., 251: 591-602.

Smith T.L. 1989. Disparate evolution of yeasts and filamentous fungi indicated by phylogenetic analysis of glyceraldehyde-3-phosphate dehydrogenase genes. Proc Natl Acad Sci U S A, 86: 7063-7066.

Smith V. et Barrell B.G. 1991. Cloning of a yeast U1 snRNP 70K protein homologue: functional conservation of an RNA-binding domain between humans and yeast. EMBO J., 10: 2627-2634.

Solnick D. et Lee S.I. 1987. Amount of RNA secondary structure required to induce an alternative splice. Mol. Cell. Biol., 7: 3194-3198.

Solymosy F. et Pollack T. 1993. Uridylate-rich small nuclear RNAs (UsnRNAs), their genes and pseudogenes, and UsnRNPs in plants: structure and function. A comparative approach. Critical Rev. Plant Sci., 12: 275-369.

Sontheimer E.J. et Steitz J.A. 1992. Three novel functional variants of human U5 small nuclear RNA. Mol. Cell. Biol., 12: 734-746.

Sontheimer E.J. et Steitz J.A. 1993. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science, 262: 1989-1996.

Spikes D.A., Kramer J., Bingham P.M. et Van Doren K. 1994. SWAP pre-mRNA splicing regulators are a novel, ancient protein family sharing a highly conserved sequence motif with the prp21 family of constitutive splicing proteins. Nucleic Acids Res., 22: 4510-4519.

Staknis D. et Reed R. 1994. Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. Mol. Cell. Biol., 14: 2994-3005.

Stevens A., Hsu C.L., Isham K.R. et Larimer F.W. 1991. Fragments of the internal transcribed spacer 1 of pre-rRNA accumulate in *Saccharomyces cerevisiae* lacking 5'----3' exoribonuclease 1. J. Bacteriol., 173: 7024-7028.

Stohl L.L. et Clayton D.A. 1992. *Saccharomyces cerevisiae* contains an RNase MRP that cleaves at a conserved mitochondrial RNA sequence implicated in replication priming. Mol. Cell. Biol., 12: 2561-2569.

Stotz A. et Linder P. 1990. The ADE2 gene from *Saccharomyces cerevisiae*: sequence and new vectors. Gene, 95: 91-98.

Strauss E.J. et Guthrie C. 1991. A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. Genes Dev., 5: 629-641.

Strauss E.J. et Guthrie C. 1994. PRP28, a 'DEAD-box' protein, is required for the first step of mRNA splicing *in vitro*. Nucleic Acids Res., 22: 3187-3193.

Stroke I.L. et Weiner A.M. 1985. Genes and pseudogenes for rat U3A and U3B small nuclear RNA. J. Mol. Biol., 184: 183-193.

Stroke I.L. et Weiner A.M. 1989. The 5' end of U3 snRNA can be crosslinked *in vivo* to the external transcribed spacer of rat ribosomal RNA precursors. J. Mol. Biol., 210: 497-512.

Stutz F., Liao X.C. et Rosbash M. 1993. U1 small nuclear ribonucleoprotein particle-protein interactions are revealed in *Saccharomyces cerevisiae* by *in vivo* competition assays. Mol. Cell. Biol., 13: 2126-2133.

Suh D., Busch H. et Reddy R. 1986. Isolation and characterization of a human U3 small nucleolar RNA gene. Biochem. Biophys. Res. Commun., 137: 1133-1140.

Suh D., Wright D. et Reddy R. 1991. Sequences more than 500 base pairs upstream of the human U3 small nuclear RNA gene stimulate the synthesis of U3 RNA in frog oocytes. Biochemistry, 30: 5438-5443.

Sun J.S. et Manley J.L. 1995. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. Genes Dev., 9: 843-854.

Szkukalek A., Mougin A., Grégoire A., Solymosy F. et Branlant C. 1996. A unique U5->A substitution in the Physarum polycephalum U1 snRNA: Evidence at the RNA and gene level. Biochimie, 78 (sous-presse).

Szkukalek A., Myslinski E., Mougin A., Lührmann R. et Branlant C. 1995. Phylogenetic conservation of modified nucleotides in the terminal loop 1 of the spliceosomal U5 snRNA. Biochimie, 77: 16-21.

Tabor S. et Richardson C.C. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA, 84: 4767-4771.

Tague B.W. et Gerbi S.A. 1984. Processing of the large rRNA precursor: two proposed categories of RNA-RNA interactions in eukaryotes. J. Mol. Evol., 20: 362-367.

Takahashi Y., Urushiyama S., Tani T. et Ohshima Y. 1993. An mRNA-type intron is present in the *Rhodotorula hasegawae* U2 small nuclear RNA gene. Mol. Cell. Biol., 13: 5613-5619.

Talerico M. et Berget S.M. 1994. Intron definition in splicing of small Drosophila introns. Mol. Cell. Biol., 14: 3434-3445.

Tang J., Abovich N. et Rosbash M. 1996. Identification and characterization of a yeast gene encoding the U2 small nuclear ribonucleoprotein particle B" protein. Mol. Cell. Biol., 16: 2787-2795.

Tani T. et Ohshima Y. 1989. The gene for the U6 small nuclear RNA in fission yeast has an intron. Nature, 337: 87-90.

Tani T. et Ohshima Y. 1991. mRNA-type introns in U6 small nuclear RNA genes: implications for the catalysis in pre-mRNA splicing. Genes Dev., 5: 1022-1031.

Tarn W.Y. et Steitz J.A. 1994. SR proteins can compensate for the loss of U1 snRNP functions *in vitro*. Genes Dev., 8: 2704-2717.

Tarn W.Y. et Steitz J.A. 1995. Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps. Proc. Natl. Acad. Sci. USA, 92: 2504-2508.

Tarn W.Y. et Steitz J.A. 1996. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. Cell, 84: 801-811.

Tarn W.Y., Hsu C.H., Huang K.T., Chen H.R., Kao H.Y., Lee K.R. et Cheng S.C. 1994. Functional association of essential splicing factor(s) with PRP19 in a protein complex. EMBO J., 13: 2421-2431.

Tarn W.Y., Yario T.A. et Steitz J.A. 1995. U12 snRNA in vertebrates: evolutionary conservation of 5' sequences implicated in splicing of pre-mRNAs containing a minor class of introns. RNA, 1: 644-656.

Teigelkamp S., Newman A.J. et Beggs J.D. 1995b. Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. EMBO J., 14: 2602-2612.

Teigelkamp S., Whittaker E. et Beggs J.D. 1995a. Interaction of the yeast splicing factor PRP8 with substrate RNA during both steps of splicing. Nucleic Acids Res., 23: 320-326.

Terns M.P. et Dahlberg J.E. 1994. Retention and 5' cap trimethylation of U3 snRNA in the nucleus. Science, 2641: 959-961.

Theissen H., Etzerodt M., Reuter R., Schneider C., Lottspeich F., Argos P., Lührmann R. et Philipson L. 1986. Cloning of the human cDNA for the U1 RNA-associated 70K protein. EMBO J., 5: 3209-3217.

Thompson J.R., Zagorski J., Woolford J.L. et Fournier M.J. 1988. Sequence and genetic analysis of a dispensible 189 nucleotide snRNA from *Saccharomyces cerevisiae*. Nucleic Acids Res., 16: 5587-5601.

Thompson-Jäger S. et Domdey H. 1987. Yeast pre-mRNA splicing requires a minimum distance between the 5' splice site and the internal branch acceptor site. Mol. Cell. Biol., 7: 4010-4016.

Tian M. et Maniatis T. 1992. Positive control of pre-mRNA splicing *in vitro*. Science, 256: 237-240.

Tiollais P., Galibert F. et Boiron M. 1971. Evidence for the existence of several molecular species in the "45S fraction" of mammalian ribosomal precursor RNA. Proc. Natl. Acad. Sci. USA, 68: 1117-1120.

Tollervey D. 1987. A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. EMBO J., 6: 4169-75.

Tollervey D. 1996. Small nucleolar RNAs guide ribosomal RNA methylation. Science, 273: 1056-1057.

Tollervey D. et Guthrie C. 1985. Deletion of a yeast small nuclear RNA gene impairs growth. EMBO J., 4: 3873-3878.

Tollervey D., Lehtonen H., Carmo-Fonseca M. et Hurt E.C. 1991. The small nucleolar RNP protein NOP1 (fibrillarin) is required for pre-rRNA processing in yeast. EMBO J., 10: 573-583.

Tollervey D., Lehtonen H., Jansen R., Kern H. et Hurt E.C. 1993. Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre- rRNA processing, pre-rRNA methylation, and ribosome assembly. Cell, 72: 443-57.

Trapman J. et Planta R.J. 1976. Maturation of ribosomes in yeast. I Kinetic analysis by labelling of high molecular weight rRNA species. Biochim. Biophys. Acta, 442: 265-274.

Tyc K. et Steitz J.A. 1989. U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. EMBO J., 8: 3113-3119.

Tyc K. et Steitz J.A. 1992. A new interaction between the mouse 5' external transcribed spacer of pre-rRNA and U3 snRNA detected by psoralen crosslinking. Nucleic Acids Res., 20: 5375-5382.

Tycowski K.T., Shu M.D. et Steitz J.A. 1993. A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. Genes Dev., 7: 1176-1190.

Tycowski K.T., Shu M.D. et Steitz J.A. 1994. Requirement for intron-encoded U22 small nucleolar RNA in 18S ribosomal RNA maturation. Science, 266: 1558-1561.

Tycowski K.T., Shu M.D. et Steitz J.A. 1996. A mammalian gene with introns instead of exons generating stable RNA products. Science, 379: 464-466.

Udem S.A. et Warner J.R. 1972. Ribosomal RNA synthesis in *Saccharomyces cerevisiae*. J. Mol. Biol. 65: 227-242.

Umen J.G. et Guthrie C. 1995a. The second catalytic step of pre-mRNA splicing. RNA, 1: 869-885.

Umen J.G. et Guthrie C. 1995b. Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. RNA, 1: 584-597.

Valcarcel J., Singh R., Zamore P.D. et Green M.R. 1993. The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. Nature, 362: 171-175.

Van der Sande C.A., Kwa M., Van Nues R.W., Van Heerikhuizen H., Raué H.A. et Planta R.J. 1992. Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA. J. Mol. Biol., 223: 899-910.

Van der Veen R., Arnberg A.C., Van der Horst G., Bonen L., Tabak H.F. et Grivell L.A. 1986. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. Cell, 44: 225-234.

Van der Walt J.P. 1956. Antonie van Leeuwenhoek, 22: 190.

Van der Walt J.P. 1970. The yeast, a taxonomic study. Lodder J. ed. pp 34-113.

Van der Walt J.P. et Yarrow D. 1984. The yeast, a taxonomic study. Kreger van Rij N.J.W. ed. 45-104.

Van Nues R.W., Rientjes J.M., Morre S.A., Mollee E., Planta R.J., Venema J. et Raué H.A. 1995. Evolutionarily conserved structural elements are critical for processing of Internal Transcribed Spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA. J. Mol. Biol., 250: 24-36.

Van Nues R.W., Rientjes J.M., Van der Sande C.A., Zerp S.F., Sluiter C., Venema J., Planta R.J. et Raué H.A.

1994. Separate structural elements within internal transcribed spacer 1 of *Saccharomyces cerevisiae* precursor ribosomal RNA direct the formation of 17S and 26S rRNA. Nucleic Acids Res., 22: 912-919.

Vankan P. et Filipowicz W. 1989. A U-snRNA gene-specific upstream element and a -30 'TATA box' are required for transcription of the U2 snRNA gene of *Arabidopsis thaliana*. EMBO J., 8: 3875-3882.

Vassilenko S.K. et Babkina G.T. 1965. Biokymia, 30: 705-712.

Vaughan M.H., Warner J.R. et Darnell J.E. 1967. Ribosomal precursor particles in the HeLa cell nucleus. J. Mol. Biol., 25: 235-251.

Veldman G.M., Brand R.C., Klootwijk J. et Planta R. 1980. Some characteristics of processing sites in ribosomal precursor RNA of yeast. Nucleic Acids Res., 8: 2907-2920.

Veldman G.M., Klootwijk J., Van Heerikhuizen H. et Planta R.J. 1981. The nucleotide sequence of the intergenic region between the 5.8S and 26S rRNA genes of the yeast ribosomal RNA operon. Possible implications for the interaction between 5.8S and 26S rRNA and the processing of the primary transcript. Nucleic Acids Res., 9: 4847-4862.

Verbeet M.P., Klootwijk J., Van Heerikhuizen H., Fontijn R.D., Vreugdenhil E. et Planta R.J. 1984b. A conserved sequence element is present around the transcription initiation site for RNA polymerase A in Saccharomycetoideae. Nucleic Acids Res, 12: 1137-1148.

Verbeet M.P., Van Heerikhuizen H., Klootwijk J., Fontijn R.D. et Planta R.J. 1984a. Evolution of yeast ribosomal DNA: molecular cloning of the rDNA units of *Kluyveromyces lactis* and *Hansenula wingei* and their comparison with the rDNA units of other Saccharomycetoideae. Mol Gen Genet, 195: 116-125.

Vidal-Leira R. 1979. Mycologia, 71: 494.

Vieira J. et Messing J. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene, 19: 259-268.

Vijayraghavan U. et Abelson J. 1990. PRP18, a protein required for the second reaction in pre-mRNA splicing. Mol. Cell. Biol., 10: 324-332.

Vijayraghavan U., Parker R., Tamm J., Iimura Y., Rossi J., Abelson J. et Guthrie C. 1986. Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. EMBO J., 5: 1683-95.

Vilardell J. et Warner J.R. 1994. Regulation of splicing at an intermediate step in the formation of the spliceosome. Genes Dev., 8: 211-220.

Waibel F. et Filipowicz W. 1990a. RNA-polymerase specificity of transcription of Arabidopsis U snRNA genes determined by promoter element spacing. Nature, 346: 199-202.

Waibel F. et Filipowicz W. 1990b. U6 snRNA genes of Arabidopsis are transcribed by RNA polymerase III but contain the same two upstream promoter elements as RNA polymerase II-transcribed U-snRNA genes. Nucleic Acids Res., 18: 3451-3458.

Wassarman D.A. et Steitz J.A. 1992. Interactions of small nuclear RNA's with precursor messenger RNA during *in vitro* splicing. Science, 257: 1918-1925.

Watakabe A., Inoue K., Sakamoto H. et Shimura Y. 1989. A secondary structure at the 3' splice site affects the *in vitro* splicing reaction of mouse immunoglobulin mu chain pre-mRNAs. Nucleic Acids Res., 17: 8159-8169.

Wells S.E. et Ares M. 1994. Interactions between highly conserved U2 small nuclear RNA structures and Prp5p, Prp9p, Prp11p, and Prp21p proteins are required to ensure integrity of the U2 small nuclear ribonucleoprotein in *Saccharomyces cerevisiae*. Mol. Cell. Biol., 14: 6337-6349.

Wells S.E., Neville M., Haynes M., Wang J., Igel H. et Ares M. 1996. CUS1, a suppressor of cold-sensitive U2 snRNA mutations, is a novel yeast splicing factor homologous to human SAP 145. Genes Dev., 10: 220-232.

Whittaker E. et Beggs J.D. 1991. The yeast PRP8 protein interacts directly with pre-mRNA. Nucleic Acids Res., 19: 5483-5489.

Whittaker E., Lossky M. et Beggs J.D. 1990. Affinity purification of spliceosomes reveals that the precursor RNA processing protein PRP8, a protein in the U5 small nuclear ribonucleoprotein particle, is a component of yeast spliceosomes. Proc. Natl. Acad. Sci. USA, 87: 2216-2219.

Wickerham L.J. 1951. Taxonomy of yeasts. Technical Bulletin No. 1029, United States Department of Agriculture, Washington, D.C.

Wise J.A. 1993. Guides to the heart of the spliceosome. Science, 262: 1978-1979.

Wise J.A. et Weiner A.M. 1980. Dictyostelium small nuclear RNA D2 is homologous to rat nucleolar RNA U3 and is encoded by a dispersed multigene family. Cell, 22: 109-118.

Wittekind M., Dodd J., Vu L., Kolb J.M., Buhler J.M., Sentenac A. et Nomura M. 1988. Isolation and characterization of temperature-sensitive mutations in RPA190, the gene encoding the largest subunit of RNA polymerase I from *Saccharomyces cerevisiae*. Mol. Cell. Biol., 8: 3997-4008.

Wolff T. et Bindereif A. 1993. Conformational changes of U6 RNA during the spliceosome cycle: an intramolecular helix is essential both for initiating the U4-U6 interaction and for the first step of slicing. Genes Dev., 7: 1377-1389.

Wolff T., Menssen R., Hammel J. et Bindereif A. 1994. Splicing function of mammalian U6 small nuclear RNA: conserved positions in central domain and helix I are essential during the first and second step of pre-mRNA splicing. Proc. Natl. Acad. Sci. USA, 91: 903-907.

Woolford J.L. 1989. Nuclear pre-mRNA splicing in yeast. Yeast, 5: 439-457.

Wu J. et Manley J.L. 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev., 3: 1553-1561.

Wu J.A. et Manley J.L. 1991. Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA. Nature, 352: 818-821.

Wu J.Y. et Maniatis T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell, 75: 1061-1070.

Wyatt J.R., Sontheimer E.J. et Steitz J.A. 1992. Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. Genes Dev., 6: 2542-2553.

Xu M.Q., Kathe S.D., Goodrich-Blair H., Nierzwicki-Bauer S.A. et Shub D.A. 1990a. Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. Science, 250: 1566-1570.

Xu Y., Petersen-Bjorn S. et Friesen J.D. 1990b. The PRP4 (RNA4) protein of *Saccharomyces cerevisiae* is associated with the 5' portion of the U4 small nuclear RNA. Mol. Cell. Biol., 10: 1217-1225.

Yang X., Bani M.R., Lu S.J., Rowan S., Ben-David Y. et Chabot B. 1994. The A1 and A1B proteins of heterogeneous nuclear ribonucleoparticles modulate 5' splice site selection *in vivo*. Proc. Natl. Acad. Sci. USA, 91: 6924-6928.

Yeh L.C. et Lee J.C. 1990. Structural analysis of the internal transcribed spacer 2 of the precursor ribosomal RNA from *Saccharomyces cerevisiae*. J. Mol. Biol., 211: 699-712.

Yeh L.C. et Lee J.C. 1992. Structure analysis of the 5' external transcribed spacer of the precursor ribosomal RNA from *Saccharomyces cerevisiae*. J. Mol. Biol., 228: 827-839.

Yu Y.T., Maroney P.A., Darzynkiwicz E. et Nilsen T.W. 1995. U6 snRNA function in nuclear pre-mRNA splicing: a phosphorothioate interference analysis of the U6 phosphate backbone. RNA, 1: 46-54.

Yuan Y. et Reddy R. 1989. Genes for human U3 small nucleolar RNA contain highly conserved flanking sequences. Biochim. Biophys. Acta, 1008: 14-22.

Zagorski J., Tollervey D. et Fournier M.J. 1988. Characterization of an SNR gene locus in *Saccharomyces cerevisiae* that specifies both dispensible and essential small nuclear RNAs. Mol. Cell. Biol., 8: 3282-3290.

Zamore P.D., Patton J.G. et Green M.R. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. Nature, 355: 609-614.

Zavanelli M.I. et Ares M. 1991. Efficient association of U2 snRNPs with pre-mRNA requires an essential U2 RNA structural element. Genes Dev., 5: 2521-2533.

Zhang M., Zamore P.D., Carmo-Fonseca M., Lamond A.I. et Green M.R. 1992. Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. Proc. Natl. Acad. Sci. USA, 89: 8769-8773.

Zhuang Y. et Weiner A.M. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell, 46: 827-835.

Zhuang Y. et Weiner A.M. 1989. A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev., 3: 1545-1552.

Zhuang Y.A., Goldstein A.M. et Weiner A.M. 1989. UACUAAC is the preferred branch site for mammalian mRNA splicing. Proc. Natl. Acad. Sci. USA, 86: 2752-2756.

Zieve G. et Penman S. 1976. Small RNA species of the HeLa cell: metabolism and subcellular localization. Cell, 8: 19-31.

Zuo P. et Maniatis T. 1996. The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes Dev., 10: 1356-1368.

Nom : **BRULÉ**

Prénom : **Fabienne**

DOCTORAT de l'UNIVERSITE HENRI POINCARE, NANCY-I

en BIOLOGIE MOLECULAIRE

VU, APPROUVÉ ET PERMIS D'IMPRIMER

Nancy, le 20 NOV 1996 UHP 122

Le Président de l'Université

UNIVERSITÉ HENRI POINCARÉ, NANCY I
LE PRESIDENT

J.P. FINANCE

C-F9

UNIVERSITÉ HENRI POINCARÉ · NANCY 1
24-30, rue Lionnois B.P.3069 54013 Nancy cedex
Tél : 83 32 81 81 Fax : 83 32 95 90
Adresse électronique : _____ @presn1.u-nancy.fr

Lorsque j'ai débuté ce travail, il venait d'être démontré que les gènes codant pour les snoRNA U3A et U3B de la levure *Saccharomyces cerevisiae* avaient la particularité de contenir un intron (Myslinski *et al.*, 1990, Science, 247, 1213-1216). Nous avons d'abord réalisé des constructions génétiques qui ont permis de montrer que les introns présents dans les gènes U3A et U3B étaient épissés dans le même type de spliceosome que les introns présents dans les ARN pré-messagers de *S. cerevisiae*. Ces constructions ont aussi permis d'étudier la structure secondaire des pré-snoRNA U3A et U3B. Les structures établies montraient qu'en dépit de grandes variations en taille et en séquence de leurs introns, ces deux ARN précurseurs ont la même architecture globale, avec une longue structure tige/boucle centrale, un site 5' d'épissage appartenant à une structure tige/boucle impliquant le premier exon, une boîte de branchement en simple-brin et un site 3' d'épissage dans la boucle latérale d'une structure tige/boucle.

De tous les gènes de snoRNA U3 qui étaient connus, seuls les gènes de *S. cerevisiae* contenaient un intron. Par amplification d'ADN par la méthode PCR, nous avons recherché si la présence d'intron dans les gènes de snoRNA U3 était une caractéristique des levures appartenant à 4 genres de la sous-famille des Saccharomycetoideae, *Saccharomyces*, *Kluyveromyces*, *Pichia* et *Hansenula*. Chez chacune des 13 espèces que nous avons étudiées, au moins un gène de snoRNA U3 avec un intron a été observé. De plus, l'intron du gène du snoRNA U3A était présent chez 11 des espèces étudiées, et le degré de conservation observé pour la séquence de cet intron était tout à fait inattendu.

En amplifiant les gènes de snoRNA U3 dans leur totalité, nous avons montré que cette conservation de séquence s'étend aussi aux régions exoniques. En plus du gène du snoRNA U3A très conservé chez la plupart des levures, nous avons mis en évidence la présence presque systématique d'un second gène qui ne porte pas toujours un intron et dont la séquence est très variable suivant les genres et même les espèces. Ce gène est souvent le seul exprimé en dépit de la forte conservation du gène codant le snoRNA U3A. La comparaison de l'ensemble des gènes de snoRNA U3 mis en évidence chez les différentes levures étudiées a apporté beaucoup de renseignements sur cet ARN, sa maturation et ses gènes.

1) Les gènes de snoRNA U3 de levures se classent en deux familles, l'une très conservée, à laquelle appartiennent les gènes des snoRNA U3A, U3B de *S. cerevisiae* et U3Kd de *Kluyveromyces delphensis*, l'autre dont les séquences ont beaucoup divergé dans les espèces et qui contient des gènes mis en évidence dans les genres *Kluyveromyces*, *Pichia* et *Hansenula*.

2) L'étude nous a apporté une confirmation phylogénétique des modèles de structure secondaire établis pour les pré-snoRNA U3A et U3B : des mutations compensatoires conservent la structure secondaire de l'intron U3A dans un allèle du pré-snoRNA U3A caractérisé chez *Saccharomyces bayanus* et la structure secondaire de l'intron U3B dans deux variants de cet intron mis en évidence chez *Saccharomyces uvarum*. Par ailleurs, les pré-snoRNA U3Kd de *Kluyveromyces delphensis* et U3Ps de *Pichia salictaria* peuvent se replier selon le modèle proposé pour les pré-snoRNA U3A et U3B.

3) Sur la base des résultats obtenus et d'une étude de la snoRNP U3A de *S. cerevisiae*, une révision du modèle de structure secondaire du snoRNA U3 de levure a été proposée. Cet ARN présente un domaine 5' très conservé chez toutes les levures et un domaine 3' variable en taille et en séquence. Ce domaine 3' peut, néanmoins, toujours être replié selon une structure cruciforme, avec au centre les boîtes phylogénétiquement conservées B et C et à l'extrémité d'une des hélices, les boîtes conservées C' et D, ces 4 boîtes interagissant avec les protéines de la snoRNP U3. Ce domaine 3' servirait donc essentiellement à fixer les protéines. Le domaine 5' qui renferme la boîte A et une séquence de 10nt complémentaire de la région 5' ETS du pré-ARN ribosomique aurait un rôle direct dans la maturation du pré-ARN ribosomique.

4) Nous avons pu montrer que, malgré les grandes variations en taille et en séquence de leur domaine 3', les snoRNA U3 mis en évidence chez *Kluyveromyces delphensis*, *Kluyveromyces lactis* et *Pichia salictaria*, sont fonctionnels chez *S. cerevisiae*.

5) Nous avons montré que chez *Hansenula wingei*, deux mutations compensatoires conservent la possibilité de former l'appariement de 10pb entre le snoRNA U3 et la région 5' ETS, ce qui est un argument phylogénétique fort de l'importance biologique de cet appariement.

Mots clés: levures, snoRNA U3, maturation des pré-ARNr, intron, épissage.