AVERTISSEMENT

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

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

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

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

LIENS

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

UNIVERSITÉ DE LORRAINE

UFR de Sciences et Techniques Biologiques
Ecole Doctorale BioSE (Biologie-Santé-Environnement)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE l'UNIVERSITE DE LORRAINE

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

Régis BACK

Etudes structurales des interactions protéines-protéines et ARN-protéines impliquées dans l'assemblage des snoRNP à boîtes C/D.

Le 30 Août 2012

Membres du jury :

Rapporteurs :	**M. Sébastien FRIBOURG**	**Chargé de Recherche, INSERM, Pessac**
	M. Nicolas LEULLIOT	**Professeur, Paris Descartes, Paris**
Examinateurs :	**M. Denis LAFONTAINE**	**Professeur, Université Libre de Bruxelles, Belgique**
	M. Bruno CHARPENTIER	**Professeur, Université de Lorraine, Nancy**
	M^{me} Christiane BRANLANT	**Directeur de Recherche, CNRS, Nancy** **Directeur de thèse**
	M. Xavier MANIVAL	**Charge de Recherche, CNRS, Nancy** **Co-directeur de thèse**
Membres invités:	**M. Christophe ROMIER**	**Chargé de Recherche, CNRS, Illkirch**
	M. Marc GRAILLE	**Chargé de Recherche, CNRS, Palaiseau**

Laboratoire AREMS : ARN, RNP, structure-fonction-maturation, Enzymologie moléculaire et Structurale
UMR 7214 CNRS-UL, Faculté de Médecine, Bâtiment Biopôle, 9 Avenue de la Forêt de Haye,
BP 184, 54505 Vandoeuvre-Lès-Nancy

LISTE DES ABREVIATIONS

Liste des abbréviations

3-AT	: 3-aminotriazol
A	: Adénine
A600nm	: Absorbance à la longueur d'onde de 600nm
AAA+	: ATPase associated with diverse cellular processes
ADN	: Acide désoxyribonucléique
ADN db	: Acide désoxyribonucléique double brin
ADNc	: Acide désoxyribonucléique complémentaire
ARN	: Acide ribonucléique
ARNm	: Acide ribonucléique messager
ARNnc	: ARN non codant
ARNr	: Acide ribonucléique ribosomique
ARNt	: Acide ribonucléique de transfert
ASF	: A site finger
ATP	: Adénosine 5' triphosphate
Bcd1	: Box C/D snoRNA 1
C	: Cytosine
CAB	: Cajal bodies box
CB	: Cajal Bodies
CBC	: Cap Binding Complex
Cbf5	: Centromere binding factor 5
CTD	: Carboxy Terminal Domain
Da, KDa	: Dalton, kilodalton
DC	: Dyskeratose Congénitale
DFC	: Dense Fibrillar Center
DNase	: Désoxyribonucléase
dNTP	: Désoxyribonucléotide triphosphate
DTT	: dithiothréitol
EDTA	: éthylène diamine tétraacétate
ETS	: External Transcribed Spacer
FC	: Fibrillar Center
G	: Guanine
GAR	: Glycine and Arginine Rich
GC	: Granular Components
GST	: Glutathione S Transférase
HEPES	: acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
IMAC	: Immobilized Metal Affinity Chromatography
IP	: Immuno-précipitation
IPTG	: isopropyl β-D-thiogalactoside
ITS	: Intenal Transcibed Spacer
Kan	: Kanamycine
Kb	: Kilobase
Lsm	: Sm Like
mN	: Nucléotide méthylé sur sa base
MRP	: Mitochondrial RNA Processing
NB	: Nucleolar bodies

NES : Nuclear Export Signal
Nhp2 : Non-Histone Protein 2
NLS : Nuclear Localization signal
Nm : Nucléotide 2'-O-méthylé
NMD : Non sens Mediated Decay
NoLS : Nucleolar Localization Signal
Nop : Nucleolar protein
nt : nucléotide
NTP : Ribonucléotide triphosphate
NTS : Non Transcribed Spacer
NUFIP : Nuclear FMRP Interacting Protein
ORF : Open Reading Frame
PAGE : polyacrylamide gel electrophoresis
pb : paire de base
PBS : Phosphate Buffer Saline
PCR : Polymerase Chain Reaction
PEG : Polyéthylène glycol
pGal : promoteur galactose dépendant
Pih1 : Protein interacting with Hsp90-1
PPIase : Peptidyl-Propyl cis/trans Isomerase
pré-ARNm : précurseur de l'ARN messager
pré-ARNr : précurseur de l'ARN ribosomique
PTC : Peptidyl Transferase center
PUA : PseudoUridine synthase and Archaeosine transglycosylase
R : Purine (adénine ou guanine)
R2TP : Rvb1p/Rvb2p/Tah1p/Pih1p
RBD : RNA Binding Domain
RMN : Résonance Magnétique Nucléaire
RNase : Ribonucléase
RNP : Ribonucléoparticle
Rrp9p : Ribosomal RNA Processing 9
Rsa1 : Ribosome assembly 1
Rvb1/2 : RuVB-like
S : Svedberg
SAM : S-Adenosyl-L-Methionine
sb : simple brin
scaRNA : small Cajal Bodies RNA
SDS : sodium dodecylsulfate
SECIS : Selenocystein Insertion Sequence
SMA : Spinal Muscular Atrophy
SMN : Survival of MotoNeuron
snoRNA : small nucleolar RNA
snoRNP : small nucleolar Ribonucleoparticle
snRNA : small nuclear RNA
snRNP : small nuclear RNP
Snu13 : small Nuclear ribonucleoprotein associated
SR : Serine/Arginine Rich
sRNA : small RNA
sRNP : small RNP
SSU : Small Subunit

T	: Thymine
Tah1	: TPR containing protein asociated with Hsp90
TBE	: Tris-Borate-EDTA
TBP	: TATA box Binding Protein
TE	: Tris-EDTA
TIP48/49	: TBP Interacting Protein 48/49
TPR	: Tetratricopeptid Repeat
t-UTP	: U Three Protein Complex required for transcription
UE	: Unité Enzymatique
UsnRNA	: Uridine rich small nuclear ribonucleic acid
UsnRNP	: Uridine rich small nuclear ribonucleoprotein particle
Utp	: U-three protein
UTR	: Untranslated region
Y	: Pyrimidine (cytosine, uracile ou thymine)
Y2H/Y3H	: Yeast two-hybrid/ Yeast three-hybrid
Ψ	: Pseudouridine

Les abréviations internationales utilisées pour désigner les acides aminés (nomenclature à trois lettres ou à une lettre) ne sont pas incluse dans cette liste.

N.B. **Nomenclature des protéines :**

Concernant les noms de protéines, nous avons respecté la nomenclature communément utilisée dans chaque organisme. Pour les protéines de levure seule la première lettre est en majuscule et le nom est suivi d'un « p » pour *« protein »* (*e.g.* Rsa1p). Pour les protéines humaines l'intégralité du nom apparaît généralement en lettre capitales (*e.g*. NUFIP). La séparation de deux noms de protéines par le symbole « / » fait simultanément référence à une protéine de levure et à son homologue humaine (*e.g*. Rsa1p/NUFIP). Pour les protéines du règne archae, seule la première lettre est en majuscule et le nom est précédé d'un « a » pour *« archae »* (*e.g*. aNop1). Enfin, le nom des ORF de levure correspond au nom de la protéine en majuscule et en italique (*e.g. RSA1*).

TABLE DES MATIERES

Table des matières

Préambule.....1

Introduction

1 Maturation des ARN ribosomiques 4
1.1 Les évènements de maturation des ARNr 18S, 5.8S et 25S chez la levure..... 4
1.1.1 La maturation du pré-ARNr 35S 6
1.1.2 La maturation des préARNr 20S et 27SA2 9
1.2 Les modifications des résidus 10
1.2.1 La 2'-O-méthylation 11
1.2.2 La pseudouridylation 11
1.2.3 Effet des modifications 11
1.2.3.1 La pseudouridylation 13
1.2.3.2 La 2'-O-méthylation 15
1.2.4 Evolution phylogénétique des modifications des résidus : nombre et position . 16
1.3 Les résidus modifiés au sein du ribosome 20
1.3.1 Fonction du ribosome 20
1.3.1.1 L'initiation 21
1.3.1.2 L'élongation 24
1.3.1.3 La terminaison 27
1.3.2 Structure du ribosome 28
1.3.3 Positionnement des résidus modifiés au sein du ribosome 30
1.3.3.1 Rôle des modifications post-transcriptionnelles au niveau de la petite sous-unité du ribosome 30
1.3.3.2 Rôle des modifications post-transcriptionnelles au niveau de la grande sous-unité du ribosome 31
1.3.3.2.1 Le centre PTC 34
1.3.3.2.2 L'hélice 69 36
1.3.3.2.3 « A-site finger » 38
1.4 Mise en place des modifications par les sRNP/snoRNP 38
2 Les RNP à boîtes H/ACA 40

2.1 Structure des ARN à boîtes H/ACA 40
2.2 Structure des protéines composant les RNP à boîtes H/ACA 43
2.2.1 L'enzyme aCbf5/Cbf5p 43
2.2.2 aNop10/Nop10p 47
2.2.3 aGar1/Gar1p 48
2.2.4 L7Ae/Nhp2p 49
2.3 Structure des particules à boîtes H/ACA 53
2.4 La biogenèse des RNP à boîtes H/ACA 59
2.4.1 Rôle de Naf1 et Shq1 60
2.4.2 Rôle possible du complexe S.M.N. 66
2.4.3 Rôle de la chaperonne Hsp90 et du complexe R2TP 66
3 Les RNP à boîtes C/D 67
3.1 Structure des ARN à boîtes C/D 68
3.2 Le motif ARN en *« K-turn »* 70
3.3 Structure des protéines composant les RNP à boîtes C/D 72
3.3.1 L'enzyme aNop1/Nop1p 72
3.3.2 Les protéines aNop5/Nop56p/Nop58p 74
3.3.3 La protéine L7Ae/Snu13p 78
3.4 Structure des particules à boîtes C/D 81
3.5 La snoRNP U3 89
3.5.1 Le snoRNA U3 89
3.5.2 Les protéines de la snoRNP U3 91
3.5.3 Structure de la snoRNP U3 93
3.5.4 La biogenèse de la snoRNP U3 94
3.6 La biogenèse des RNP à boîtes C/D : état des connaissances au début de ma thèse 97
3.6.1 La chaperonne Hsp90 97
3.6.2 Le complexe R2TP 101
3.6.2.1 Les protéines Rvb1 et Rvb2 101
3.6.2.2 La protéine Pih1p/NOP17 104
3.6.2.3 La protéine Tah1p/SPAGH 104
3.6.3 La protéine Rsa1p/NUFIP 105

Objectifs de la thèse

Objectifs de la thèse.....107

Résultats et discussion

CHAPITRE 1

La co-expression chez Escherichia coli : un outil de biologie moléculaire et structural

1 Présentation des travaux 110
1.1 Mise en place au laboratoire 112
1.1.1 Les vecteurs pnEA-tH et pnEA-3CH 112
1.1.2 Les vecteurs pnCS et pnYC 114
1.1.3 Clonage au sein des vecteurs 114
1.1.4 La concaténation 116
1.1.5 Le principe 116
1.2 Les séquences codantes sélectionnées 119
2 Résultats et discussion 119
2.1 Résultats obtenus lors d'une première série de criblage par co-expression 119
2.1.1 Première série d'expériences 121
2.1.2 Deuxième série d'expériences 131
2.1.3 Troisième série d'expériences 137
2.2 Bilan de l'analyse par spectrométrie de masse 137
2.3 Purification et analyse du complexe His-Snu13p/Rsa1p (231-381) 145
2.3.1 Production et purification préparative 145
2.3.2 Analyse par spectrométrie de masse *(travail réalisé par J.M. Saliou)* 146
2.3.2.1 Données obtenues en conditions dénaturantes 148
2.3.2.2 Données obtenues en conditions non dénaturantes 151
2.3.3 Bilan de l'analyse du complexe His-Snu13p/Rsa1p (231-381) 154
2.4 Purification préparative et analyse du complexe His-Rsa1p (231-381)/Hit1p 154
2.5 Production de la protéine Snu13p en utilisant le vecteur pnEA-3CH 155

2.5.1 Essais de production et de purification de la protéine Snu13p 157
2.5.2 Analyse par spectrométrie de masse 157
2.6 Le complexe Snu13p/Rsa1p/Hit1p 160
2.7 Dernières séries d'expériences de co-expression 163
2.7.1 Première série d'essais du second ensemble de co-expression 164
2.7.2 Deuxième série d'essais du second ensemble de co-expression 166
2.7.3 Troisième série d'essais du second ensemble de co-expression 170
2.7.4 Quatrième série d'essais du second ensemble de co-expression..... 170
3 Conclusion et Perspectives 177
3.1 Bilan..... 177
3.2 Vers l'étude d'entité multimériques 177

CHAPITRE 2

La machinerie d'assemblage des snoRNP à boîtes C/D : rôle de la protéine Hit1p

1 Etude structurale sur le complexe Snu13p/Rsa1p/Hit1p 183
1.1 Production du complexe Snu13p/Rsa1p (231-375)/Hit1p..... 183
1.2 Protéolyse ménagée du complexe Snu13p/Rsa1p/Hit1p 185
1.2.1 Choix de l'enzyme et de sa concentration..... 187
1.2.2 Choix du temps d'incubation 189
1.2.3 Fractionnement et analyse par spectrométrie de masse des produits d'hydrolyse ménagée ……...189
1.3 La production du fragment de Hit1p (44-164) 194
1.3.1 Analyse par DLS du fragment Hit1p (44-164)..... 194
1.3.2 Analyse par S.L.S. 196
1.3.3 Analyse par R.M.N. 196
1.4 Le complexe Rsa1p (317-375)/Hit1p (44-164) 199
1.4.1 Analyse par DLS 199
1.4.2 Analyse par S.L.S. 199
1.4.3 Tests de cristallisation 202
1.4.4 Etude par R.M.N. 202
1.5 La protéolyse ménagée en présence du snoRNA U3 205
2 Etude fonctionnelle sur les complexes Snu13p/Rsa1p/Hit1p et Rsa1p/Hit1p 206

2.1 Présentation des travaux décrits dans le manuscrit n°1 et de ma participation aux travaux 206
2.2 Manuscrit n°1 208
3 Conclusion et perspectives 232
3.1 Bilan 232
3.2 Vers un modèle structurale multiprotéique 232

CHAPITRE 3

L'interface Rsa1p/Snu13p, caractérisation fonctionnelle et structurale

1 Présentation des travaux 233
Manuscrit n°2 234
2 Données complémentaires sur le complexe Snu13p/Rsa1p 278
3 Conclusions et perspectives 278
3.1 Bilan 278
3.2 La structure tridimensionnelle du complexe Snu13p/Rsa1p : élément de compréhension de la biogenèse des snoRNP à boîtes C/D 279

CHAPITRE 4

Structure tridimensionnelle de la protéine Tah1p et de son complexe avec l'extrémité C-terminale de la chaperonne Hsp90 : les déterminants de l'interaction

1 Présentation des travaux 282
2 Résultats et discussion 283
2.1 Manuscrit n°3 283
2.2 Compléments concernant les analyses du complexe Tah1p-Hsp90 décrites dans le manuscrit n°3 328
2.3 Travaux complémentaires réalisés sur la protéine Pih1p non décrits dans le manuscrit 328
2.3.1 Travaux réalisés sur l'extrémité N-terminale de Pih1p 328
2.3.2 Travaux réalisés sur l'extrémité C-terminale de Pih1p 330
3 Conclusion et perspectives 335

3.1 Bilan 335
3.2 Etude structurale du complexe R2TP : la protéine Pih1p 337

Conclusions & Perspectives

Etude structurale de la machinerie de biogenèse des snoRNP à boîtes C/D chez la levure Saccharomyces cerevisae

1 La co-expression, un outil d'analyse pour la biologie structurale 339
2 La protéolyse ménagée, définition des domaines 341
3 La R.M.N. 341
4 Vers une étude structurale par SAXS 342
5 Vers l'incorporation des ARN dans l'étude des complexes 343

Matériel & Méthodes

1 Matériel 345
1.1 Les souches bactériennes 345
1.1.1 Pour l'amplification et l'extraction d'ADN plasmidique 345
1.1.2 Pour la surexpression de protéines recombinantes 345
1.2 Les plasmides bactériens 345
1.2.1 Pour le clonage 345
1.2.2 Pour la production de protéines 346
1.2.3 Pour la synthèse d'ARN *in vitro* 346
1.3 Les milieux de culture 346
1.4 Les tampons et solutions d'usage courant 347
2 Méthodes 348
2.1 Techniques de génie génétique 348
2.1.1 Clonage 348
2.1.1.1 Clonage direct dans les vecteurs d'expressions bactérien 349
2.1.1.2 Sous clonage par le pCR®-Blunt 349
2.1.2 Mutagenèse dirigée par PCR 350
2.2 Ingénierie des protéines 351
2.2.1 Surproduction de protéines chez *E. coli* 351

2.2.1.1 Pour les cribles de co-expression (échelle analytique) 351
2.2.1.2 Pour des cribles de cristallogenèse ou d'analyse par R.M.N. homonucléaire (échelle préparative) 351
2.2.1.3 Pour des analyses par R.M.N. hétéronucléaire (échelle préparative) 352
2.2.2 Purification de protéines 352
2.2.2.1 Lors des cribles par co-expression 352
2.2.2.2 Lors des préparations pour des études structurales 353
2.2.3 Protéolyse ménagée 353
2.2.4 Analyse par D.L.S. 354
2.2.5 Analyse par S.L.S. 354
2.3 Synthèse *in vitro* des ARN 354

Références Bibliographiques

Références Bibliographiques……..……356

PREAMBULE

Préambule

La traduction est une activité clef de toutes cellules vivantes car elle permet le décodage de l'information génétique portée par les ARNm (Acides RiboNucléique messagers) et la synthèse des liaisons peptidiques, à la base de la formation des protéines. La traduction est réalisée au sein d'une machinerie ribonucléoprotéique : le ribosome. Les ribosomes sont composés de protéines mais également d'ARNr (ARN ribosomiques) qui ont la particularité de composer le site actif. Chez les eucaryotes, cette machinerie comporte : *i*) une grande sous-unité 60S (50S chez *E. coli*), composée de 3 ARNr (25-28S suivant les espèces, 5,8S et 5S) (23S et 5S chez *E. coli*) et de 46 protéines ribosomiques (34 chez *E. coli*), qui catalyse la formation de la liaison peptidique. *ii*) une petite sous-unité 40S (30S chez *E. coli*), composée de l'ARNr 18S (16S chez *E. coli*) et de 33 protéines ribosomiques (21 chez *E. coli*), au sein de laquelle s'effectue le décodage des ARNm. Le ribosome 80S a une masse de 3,3 MDa chez la levure. Son volume est environ 40% supérieur à celui de son homologue bactérien (70S).

Les ribosomes étant les garants de la fidélité de la traduction de l'information génétique, leur biogenèse est un processus complexe et très contrôlé impliquant de nombreuses étapes. Parmi celles-ci, les clivages endo- et exo-nucléolytiques permettent de produire les ARNr matures à partir d'un grand ARN précurseur produit par l'ARN Polymérase I (le pré-ARNr 35S). Ce pré-ARNr subit aussi l'apposition de modifications de bases et de riboses (Figure 1). Ces étapes de clivages et de modifications du pré-ARNr se déroulent au sein de macro-complexes nucléolaires qui mettent en jeu des particules (les snoRNP ou *« small nucleolar ribonucleoprotein particle »*) contenant un petit ARN non codant (le snoRNA) et des protéines dont certaines d'entre elles portent une activité catalytique (Figure 1). Il existe deux grandes classes de snoRNP qui se distinguent par des séquences et motifs structuraux conservés dans le snoRNA qu'elles contiennent, et aussi par leur composition en protéines constitutives : il s'agit des snoRNP à boîtes C/D et des snoRNP à boîtes H/ACA. Ces deux familles de snoRNP participent donc à la maturation nucléolytique des ARNr mais sont également des machineries responsables des modifications des résidus des ARNr.

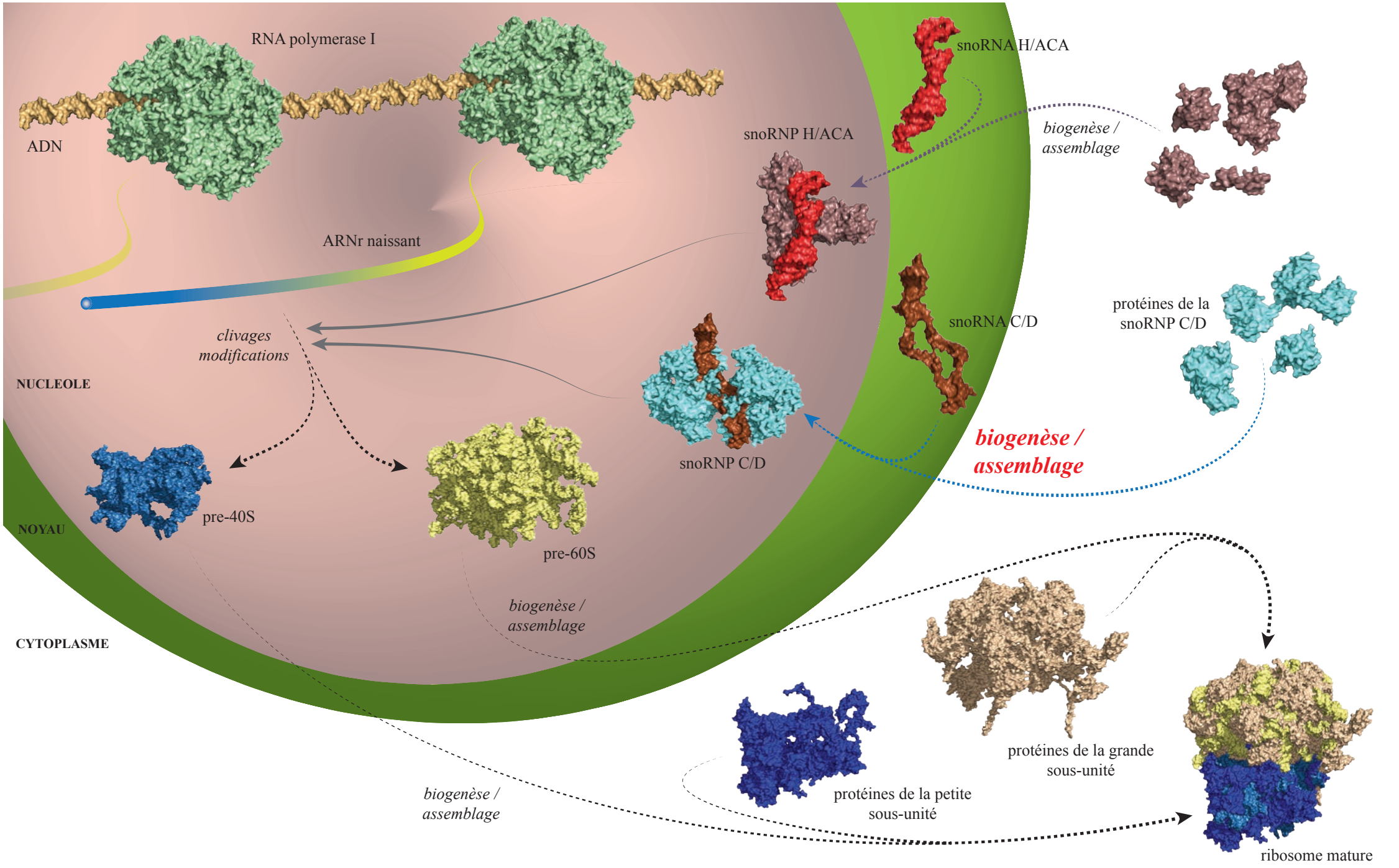


Figure 1. **Représentation schématique du contexte biologique.** L'ensemble des protéines, des ARN et l'ADN sont représentés en surface moléculaire sauf l'ARNr naissant symbolisé par un tube. Les évènements de maturation, de biogenèse et d'assemblage sont symbolisés par des flèches en pointillées. Les compartiments cellulaires sont mentionnés. Les évènements qui ont fait l'objet des études au cours de la thèse sont indiqués en rouge. Pour des raisons de clarté, l'échelle entre les différentes molécules n'est pas respectée.

Depuis de nombreuses années, le laboratoire s'intéresse aux mécanismes de biogenèse des ribosomes et, en particulier, à l'architecture et au rôle de ces snoRNP. Dès 2005, des travaux réalisés au laboratoire sur les archaea avaient clairement montré qu'il existait des particules sRNP équivalentes aux snoRNP et qu'elles peuvent être reconstituées *in vitro* en utilisant un ARN transcrit *in vitro* et des protéines recombinantes. Par contre, ce type de reconstitution n'avait pas pu être réalisée pour les snoRNP eucaryotes. Peu de temps avant que j'intègre le laboratoire, l'équipe avait trouvé l'origine de cette différence grâce à un travail de collaboration avec l'équipe d'E. Bertrand (IGM, Montpellier). Les deux équipes avaient mis en évidence l'existence de facteurs cellulaires impliqués dans la biogenèse des snoRNP *in cellulo*.

A mon arrivée au laboratoire j'ai entrepris des études de relations structures-fonctions sur ces facteurs d'assemblage et sur leurs interactions avec les protéines des snoRNP, ceci en étroite collaboration avec Benjamin Rothé (Figure 1). J'ai également consacré une grande partie de mon travail de thèse à la détermination de structures tridimensionnelles de certains facteurs mais aussi à la mise au point des conditions de leur expression et de leur purification afin d'obtenir des échantillons compatibles avec des études structurales.

Afin de mieux expliciter les résultats obtenus au cours de ma thèse, je présenterai dans la première partie introductive de mon manuscrit la maturation des ARNr comportant les clivages et les modifications des résidus. Je ferai ensuite des rappels synthétiques sur les étapes de la traduction ainsi que la structure du ribosome, dans le but de situer les résidus modifiés au sein de cette organisation. Je détaillerai enfin la mise en place de ces modifications par la machinerie des sRNP/snoRNP et ferai un bilan des données, notamment structurales sur ces particules. Pour finir, je donnerai des précisions sur les connaissances qui étaient disponibles au début de ma thèse sur les facteurs protéiques qui venaient d'être découverts comme étant impliqués dans l'assemblage des snoRNP. Ceci afin d'expliciter mes objectifs et mes stratégies de recherche qui visaient à établir les relations structures-fonctions impliquées dans les mécanismes d'assemblage des snoRNP eucaryotes, qui sont elles-mêmes requises pour la maturation des transcrits des ARN ribosomiques, composants essentiels des ribosomes.

La partie introductive de ce manuscrit sera plus particulièrement focalisée sur la levure *Saccharomyces cerevisiae* qui a fait l'objet de mes études. Quand cela sera nécessaire, nous présenterons néanmoins des données complémentaires sur d'autres organismes.

INTRODUCTION

Introduction

1 Maturation des ARN ribosomiques

1.1 Les évènements de maturation des ARNr 18S, 5.8S et 25S chez la levure

Chez les Eucaryotes, la synthèse des ribosomes requière à elle seule les 3 ARN polymérases et mobilise environ 70% de l'activité transcriptionnelle (Warner, 1999). L'ARN polymérase II synthétise les pré-ARNm codant les protéines ribosomiques et les facteurs intervenant en trans dans la biogenèse des ribosomes. L'ARN polymérase III produit le précurseur de l'ARNr 5S, tandis que l'ARN polymérase I transcrit un grand ARN 35S chez la levure (45S chez l'homme), précurseur des ARNr 5.8S, 18S et 25S (28S chez l'homme). La biogenèse des ribosomes fait ainsi appel à plus de 200 protéines non ribosomiques et à une centaine de snoRNP, qui comme nous l'avons déjà évoqué dans le préambule, jouent un rôle majeur dans les étapes de clivage du précurseur et la modification des résidus.

La biogenèse des ribosomes est un processus complexe, ordonné dans le temps et dans l'espace cellulaire. L'événement déclencheur de ce mécanisme est la synthèse du précurseur 35S des ARNr 25S, 18S et 5,8S. Il s'assemble avec des protéines ribosomiques et de nombreux facteurs nucléolaires pour former dans un premier temps un macrocomplexes 90S qui catalyse les étapes précoces de maturation du pré-ARNr permettant de libérer la pré-particule 40S. Un second macrocomplexe est alors formé autour de l'intermédiaire précurseur des ARNr 18S et 5,8S et il donnera lieu à la particule pré-60S. Ces différentes étapes ont lieu dans le nucléole, un sous domaine du noyau des cellules eucaryotes au sein duquel l'ARNr est soumis à une série ordonnée de clivages endo et exo-nucléolytiques et des modifications enzymatiques spécifiques de nucléotides (Figure 2).

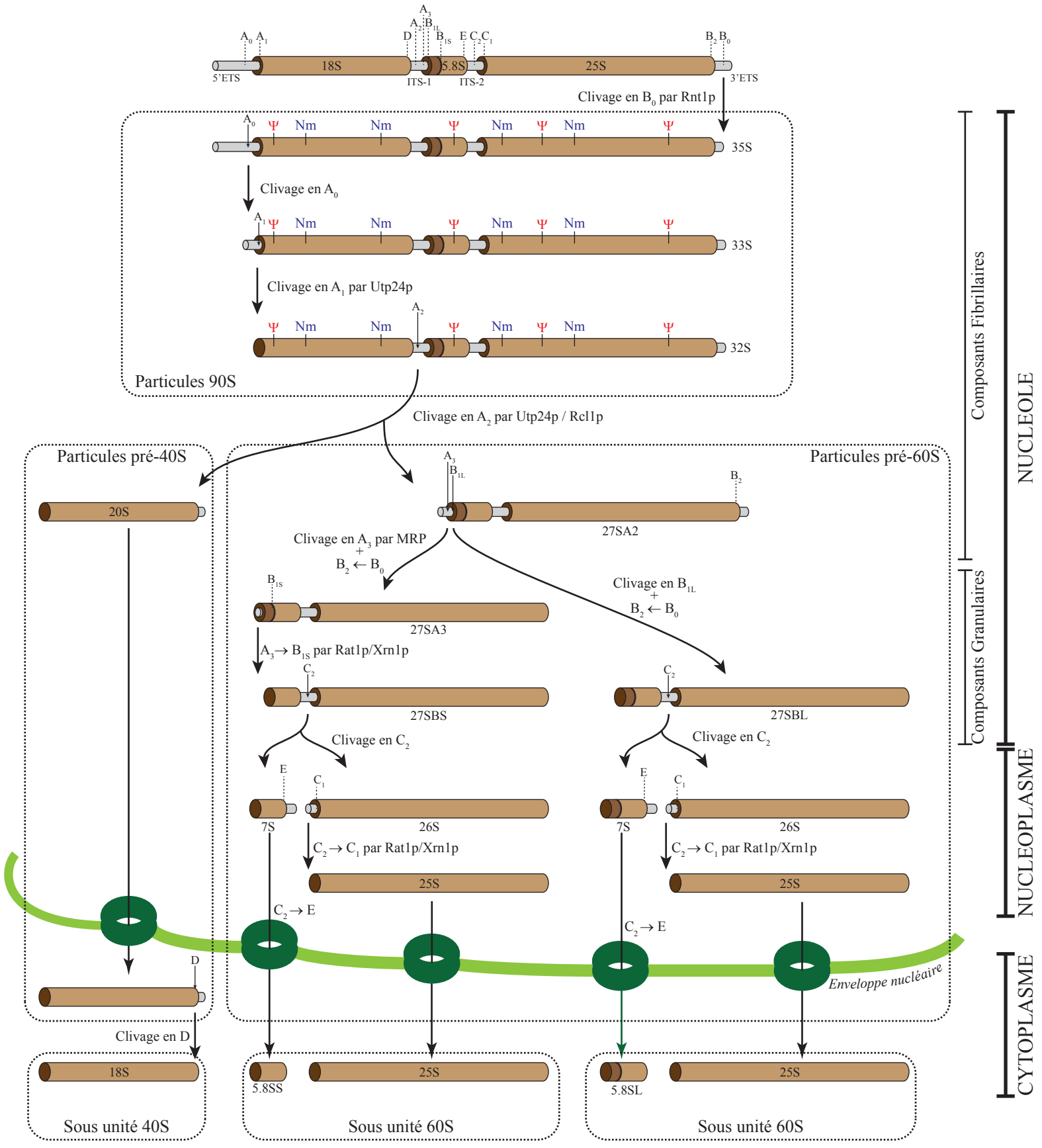


Figure 2. **Schéma simplifié de la maturation des ARNr chez *S. cerevisiae*.** Les clivages exonucléolytiques sont symbolisés par les symboles ← (3' vers 5') et → (5' vers 3'). Les modifications de résidus : pseudouridylations et 2'-O-méthylations sont représentées par les symboles Ψ et Nm, respectivement, seulement sur les ARNr 35S, 33S et 32S pour des raisons de clartés. Adapté de (Henras *et al*., 2008).

1.1.1 La maturation du pré-ARNr 35S

L'ARN polymérase I transcrit de façon continue le pré-ARN ribosomique jusqu'à la région 3'ETS (*External Transcribed Spacer*). De façon co-transcriptionnelle, la RNAse III Rnt1p va alors cliver l'ARN au niveau du site B_0 contenu dans la région 3'ETS afin de libérer la particule pré-ribosomale initiale 90S contenant le pré-ARNr 35S (Henras *et al.*, 2004; Henras *et al.*, 2008; Kufel *et al.*, 1999) (Figure 2). L'extrémité 5'-ETS du pré-ARNr 35S est éliminée par deux clivages endonucléolytiques successifs. D'abord au site A_0 donnant naissance à l'intermédiaire 33S (Hughes & Ares, 1991) puis au site A_1 de manière concomitante avec le clivage au site A_2 situé dans la séquence ITS1 (*« Internal Transcribed Spacer 1 »*). Il avait été proposé que Utp24p était l'enzyme impliquée dans ces clivages (Bleichert *et al.*, 2006) mais récemment il a été montré que l'enzyme Rcl1p serait responsable du clivage au site A_2 (Horn *et al.*, 2011). Le clivage au site A_1 générant l'intermédiaire 32S permet de délimiter l'extrémité 5' du futur ARNr 18S (Beltrame *et al.*, 1994). C'est le clivage au site A_2 suivi d'un clivage au site A_3 qui divisera le processus de maturation en deux voies distinctes : celle du pré-ARNr 20S, qui contient l'ARNr 18S, et celle du pré-ARNr 27SA2 qui renferme les ARNr 5.8S et 25S (De Jonge *et al.*, 1977; Henry *et al.*, 1994; Lindahl *et al.*, 1994) (Figure 2).

Il a également été observé chez la levure que, dans des conditions de croissance très favorables, les coupures aux sites A_0, A_1 et A_2 et donc la libération de la particule pré-40S se réalisent de façon co-transcriptionnelle (Osheim *et al.*, 2004). La particule pré-60S est libérée dans un second temps par le clivage endonucléolytique au site B_0 (Henras *et al.*, 2004).

La libération à partir du précurseur 35S, du pré-ARNr 20S qui contient l'ARNr mature 18S qui est le composant nucléique de la future petite sous-unité du ribosome ; nécessite donc la mise en place des clivages endonucléolytiques aux sites A_0, A_1 et A_2. Ces évènements requièrent l'intervention de nombreux facteurs, parmi lesquels les snoRNP à boîtes C/D U3 (Dutca *et al.*, 2011; Hughes & Ares, 1991; Hughes *et al.*, 1987; Kass *et al.*, 1990; Marmier-Gourrier *et al.*, 2011; Méreau *et al.*, 1997) et U14 (Huang *et al.*, 1992; Li & Fournier, 1992; Li *et al.*, 1990; Zagorski *et al.*, 1988) ainsi que les snoRNP à boîtes H/ACA snR10 (Tollervey, 1987) et snR30/U17 (Morrissey & Tollervey, 1993). Parmi ces snoRNP, seule la snoRNP snR10 est spécifique de la levure. D'autre part, les snoRNP U14 et snR10 ont une double fonctionnalité puisqu'elles sont également impliquées dans des réactions de modification du pré-ARNr (Lestrade & Weber, 2006; Piekna-Przybylska *et al.*, 2007).

Le rôle joué par ces snoRNP dans la maturation nucléolytique du pré-ARNr a fait l'objet de nombreux travaux, notamment la snoRNP U3 au laboratoire. On sait que l'appariement du snoRNA U3 avec les régions 5'-ETS et 18S est essentiel pour les clivages précoces aux sites A_0, A_1 et A_2 (Dutca *et al.*, 2011; Hughes, 1996; Hughes & Ares, 1991; Kass *et al.*, 1990; Marmier-Gourrier *et al.*, 2011; Méreau *et al.*, 1997), tout comme les protéines constituants la snoRNP U3 (Baserga *et al.*, 1997; Cléry *et al.*, 2007; Dunbar *et al.*, 1997; Lee & Baserga, 1997; Lee & Baserga, 1999; Wehner *et al.*, 2002). Le snoRNA U3, essentiel à la viabilité cellulaire est associé de manière stable à cinq protéines (Watkins *et al.*, 2000). Il peut en plus s'associer à un grand nombre d'autres protéines (28 facteurs protéiques) pour former le *« SSU (Small SubUnit) processome »* (Dragon *et al.*, 2002). Le complexe SSU-processome est à l'origine de la maturation nucléolytique de l'ARNr 18S et de l'assemblage de la petite sous-unité 40S. Sa formation à l'extrémité 5' du pré-ARNr naissant constitue une étape clé de la synthèse des ribosomes. La snoRNP U3 est essentielle pour la mise en place du SSU (Osheim *et al.*, 2004). Associée aux snoRNP U14, snR10 et snR30, la snoRNP U3 fait donc partie de la machinerie initiant la maturation précoce de l'ARNr 18S. Parmi ces RNP, seule la snR10 n'est pas essentielle à la viabilité cellulaire chez la levure.

La snoRNP U3 est conservée chez l'ensemble des eucaryotes. Elle est essentielle à la viabilité cellulaire et contient le snoRNA le plus abondant de la cellule (Epstein *et al.*, 1984). Bien qu'il appartienne à la famille des snoRNA à boîtes C/D, il se distingue par des caractéristiques structurales et fonctionnelles spécifiques. En effet aucune activité de 2'-O-méthylation de la particule U3 n'a été décrite à ce jour. Elle intervient dans la formation des complexes pré-ribosomiques et dans les clivages précoces de l'ARNr aux sites A_0, A_1 et A_2, comme nous venons de le voir. Sa structuration est elle aussi particulière. Le snoRNA U3 se compose de deux domaines : un court domaine 5' structuré en une ou deux tiges-boucles à l'état libre (Fournier *et al.*, 1998; Ségault *et al.*, 1992) et un long domaine 3' possédant les caractéristiques des snoRNA à boîtes C/D. Ces deux domaines sont séparés par une région charnière appelée *« 3'-hinge »*.

Le domaine 5' ou domaine fonctionnel, nécessaire à la viabilité cellulaire contient les boîtes conservées A et A' qui s'apparient au pré-ARNr par la formation d'hélices intermoléculaires (Dutca *et al.*, 2011; Marmier-Gourrier *et al.*, 2011; Méreau *et al.*, 1997; Myslinski *et al.*, 1990). La snoRNP U3 jouerait donc un rôle de chaperonne en s'associant de façon spécifique avec certaines séquences du pré-ARNr (Figure 3). Ainsi, le domaine 5' du snoRNA U3, notamment les boîtes A et A' s'associent avec la séquence 18S du pré-ARNr

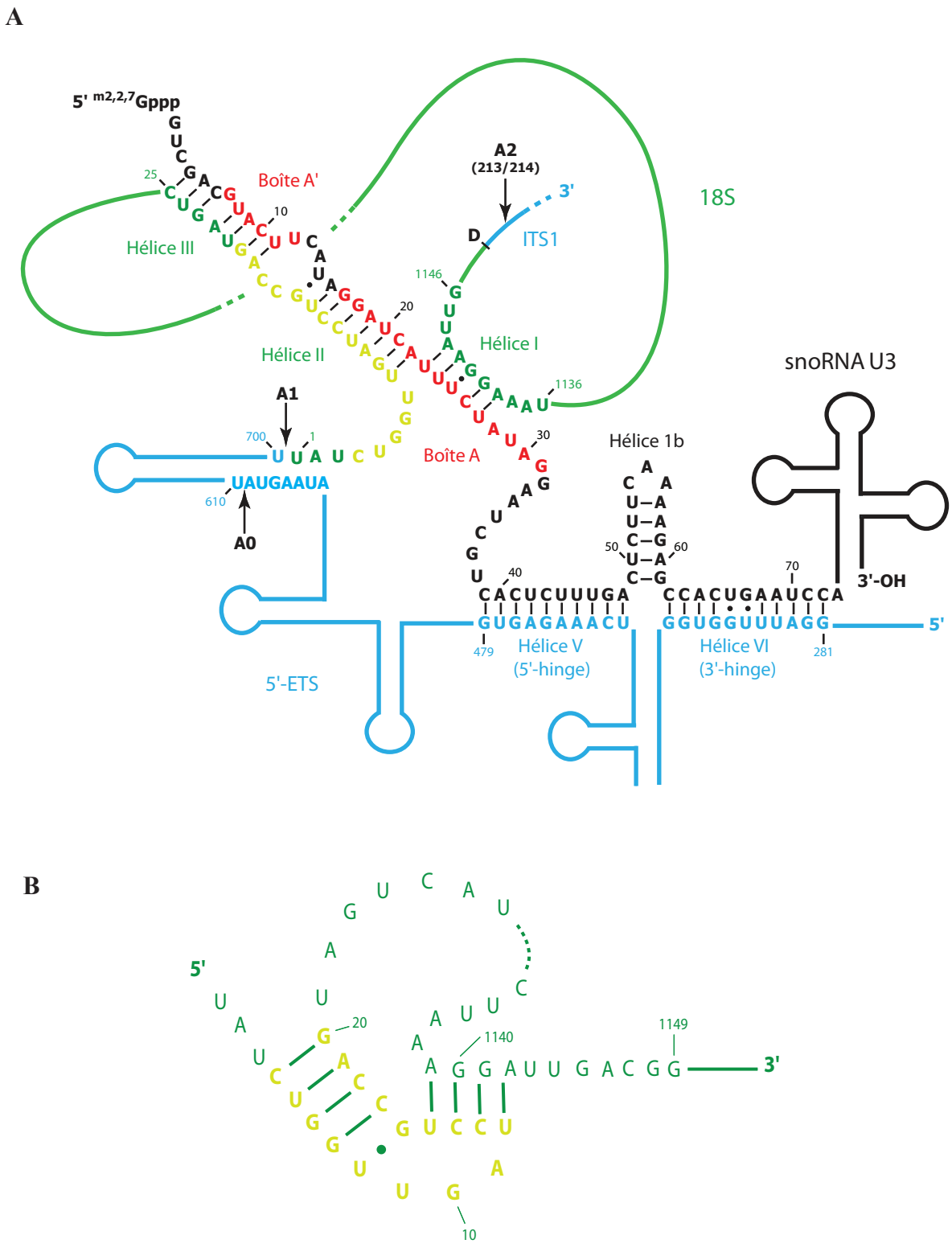


Figure 3. **Représentation schématique de l'appariement entre le snoRNA U3 et le pré-ARNr 35S chez la levure *S.cerevisiae***. (A) Association du domaine 5' du snoRNA U3 avec le pré-ARNr. L'ARNr 18S est représenté en vert et la séquence impliquée dans le pseudo noeud central de la petite sous-unité en jaune, les séquences 5'-ETS et ITS1 en bleu. Le snoRNA U3 est représenté en noir et les boîtes A et A' en rouge. Les sites de coupures A_0, A_1 et A_2 sont indiqués par une flèche. (B) Pseudo noeud central de la petite sous unité mature formé par l'ARN 18S. Adapté de (Marmier-Gourrier *et al.*, 2011).

pour former respectivement les hélices I, II et III (Borovjagin & Gerbi, 2000; Méreau *et al.*, 1997; Myslinski *et al.*, 1990), tandis que la séquence *« 3'-hinge »* de la région charnière s'apparie avec l'élément 5'-ETS du pré-ARNr pour former l'hélice V (ou *« 5'-hinge »*) (Beltrame & Tollervey, 1995; Borovjagin & Gerbi, 2000; Brulé *et al.*, 1996).

Les segments du pré-ARNr engagés dans la formation des hélices I, II et III avec le snoRNA U3 sont les segments de l'ARNr 18S qui forment ensuite le pseudo-nœud central de la petite sous-unité ribosomique mature (Figures 3). Cet élément structural est universellement conservé au sein des ribosomes (Gutell *et al.*, 1994). C'est pourquoi, J.M. Hughes et notre équipe ont proposé que la snoRNP U3 participerait au repliement séquentiel de l'ARNr par le biais de ces hélices retardant ainsi la formation du pseudo-nœud central (Hughes, 1996; Méreau *et al.*, 1997).

Initialement, une hélice IV avait été prédite entre la région charnière du snoRNA U3 et l'ARNr 18S chez la levure (Méreau *et al.*, 1997). Mais des données acquises plus récemment ont montré que la région charnière du snoRNA U3 s'associait avec la région 5'-ETS du pré-ARNr pour former une hélice VI conservée (ou 3'-hinge) (Borovjagin & Gerbi, 2000; Dutca *et al.*, 2011; Marmier-Gourrier *et al.*, 2011).

La formation de ces hélices intermoléculaires est donc nécessaire pour la production de l'ARNr 18S. A l'exception de l'hélice III qui n'est requise que pour les coupures aux sites A_1 et A_2, les appariements entre le snoRNA U3 et le pré-ARNr sont nécessaires pour les clivages endonucléolytiques A_0, A_1 et A_2.

Nous reviendrons plus en détail sur le domaine 3' du snoRNA U3 dans le paragraphe 3.5.

1.1.2 La maturation des préARNr 20S et 27SA2

Alors que la maturation du pré-ARNr 20S se poursuivra dans le cytoplasme par la coupure au site D par l'enzyme Nob1p qui génèrera l'extrémité 3' mature de l'ARNr 18S (Fatica *et al.*, 2003; Lamanna & Karbstein, 2009), le pré-ARNr 27SA2 est quant à lui maturé dans le nucléole par deux voies alternatives (Figure 2).

- La voie de maturation majoritaire (qui concerne 85% de la population des pré-ARNr 27SA2) débute par un clivage au site A_3 de la région l'ITS1 résiduelle par la RNAse MRP, ce qui donne naissance à l'intermédiaire 27SA3 (Henry *et al.*, 1994; Lygerou *et*

al., 1996; Reilly & Schmitt, 1995). Cette étape est rapidement suivie par une dégradation 5'-3' du site A_3 jusqu'au site B_{1S} par les exonucléases 5'-3' Rat1p et/ou Xrn1p, donnant l'intermédiaire $27SB_S$ (Xue *et al.*, 2000).

- La voie minoritaire (les 15% de la forme 27SA2 restant), relativement peu décrite, consiste en un clivage direct au site B_{1L} sans qu'il y ait hydrolyse au site A_3 et donne lieu à l'intermédiaire $27SB_L$.

La formation des intermédiaire $27SB_S$ et $27SB_L$ s'accompagne également de l'élimination de la séquence 3'-ETS résiduelle par une dégradation exonucléolytique B_0-B_2 permettant de définir l'extrémité 3' mature du futur ARNr 25S (Veldman *et al.*, 1980).

Les deux populations d'intermédiaires 27SB suivent alors une voie de maturation identique impliquant un clivage au site C2 et une dégradation 5'-3' par les exonucléases Rat1p ou Xrnt1p du site C2 jusqu'au site C1 qui constitue l'extrémité 5' mature de l'ARNr 25S (Geerlings *et al.*, 2000). L'extrémité 3' mature des ARN 5.8S et $5.8S_L$ (la forme minoritaire qui est plus longue de 6 nucléotides en 5') est définie par une digestion exonucléolytique 3'-5' du site C2 jusqu'au site E par l'exosome, des facteurs Rex1p et Rex2p, ainsi que l'endonucléase Ngl2p (Allmang *et al.*, 1999; Faber *et al.*, 2002; Mitchell *et al.*, 1996; van Hoof *et al.*, 2000a).

1.2 Les modifications des résidus

Les pré-ARNr subissent en plus des clivages nucléotidiques, de nombreuses modifications de résidus. Il existe deux grands types majoritaires de modifications : la 2'-O-méthylation et la pseudouridylation. De façon plus minoritaire chez les eucaryotes, il existe un troisième type de modification : la méthylation à différentes positions des bases des nucléotides.

Deux grandes classes de snoRNP catalysent les modifications du pré-ARNr au sein du complexe 90S : les snoRNP à boîtes C/D qui possèdent l'activité de 2'-O-méthylation et les snoRNP à boîtes H/ACA qui possèdent l'activité de pseudouridylation (Figure 2). Les snoRNP de ces deux familles contiennent un ARN guide, le snoRNA, qui permet de cibler spécifiquement les résidus à modifier. Nous reviendrons plus en détails sur la composition, la structure et la biogenèse de ces particules dans les paragraphes 2 et 3.

1.2.1 La 2'-O-méthylation

Cette modification consiste en l'ajout d'un groupement méthyl (CH3) sur l'oxygène lié au carbone en position 2' du ribose (Figure 4). Comme précisé auparavant, cette modification fait appel à la classe des snoRNP à boîtes C/D qui contiennent une 2'-O-méthylase : la protéine Nop1/Fibrillarine.

1.2.2 La pseudouridylation

La pseudouridylation est un processus plus complexe car il s'agit d'une isomérisation de la base uridine en pseudouridine, ceci par rupture de la liaison N1-glycosidique, suivie d'une rotation de la base de 180° autour de l'axe N3-C6 et de la formation d'une nouvelle liaison covalente à la position C5 (Figure 4). La pseudouridine se distingue du résidu uridine car elle possède 2 groupements protons imino (NH) (au lieu d'un seul) chacun étant potentiellement donneur de proton permettant donc d'établir une liaison hydrogène supplémentaire. La pseudouridylation est catalysée par une seconde famille de snoRNP : les snoRNP à boîtes H/ACA, qui contiennent une ARN : pseudouridine synthase : Cbf5/CBF5 (également appelée Dyskérine chez l'homme).

1.2.3 Effet des modifications

Les modifications des ARNr catalysées par les snoRNP sont retrouvées chez les archées et chez les eucaryotes. Il a été constaté très tôt que les modifications les plus conservées se situent dans les régions de séquences conservées ayant des fonctions importantes au sein des ribosomes (Bakin *et al.*, 1994; Branlant *et al.*, 1981; Brimacombe *et al.*, 1993; Lane *et al.*, 1995; Smith *et al.*, 1992; Veldman *et al.*, 1981) comme les régions A, P et E qui interagissent avec les ARNm et ARNt, le centre peptidyl-transférase (PTC), le tunnel de sortie du polypeptide, le centre de décodage ou encore les surfaces d'interactions de la grande et petite sous-unité. Ceci impliquerait donc que ces modifications ont une importance fonctionnelle au niveau de la structure et de la fonction de traduction du ribosome (Baudin-Baillieu *et al.*, 2009; Branlant *et al.*, 1981; Decatur & Fournier, 2002; Veldman *et al.*, 1981).

Comme déjà mentionné, certaines snoRNP ont un rôle à la fois dans la maturation nucléolytique des ARNr (vu précédemment) et dans la catalyse de modifications de l'ARNr. Par exemple, la snoRNP à boîtes H/ACA snR10 est responsable de la formation chez la levure de la Ψ2919 (numérotée parfois Ψ2923 (Baudin-Baillieu *et al.*, 2009) équivalente à la position 2554 chez *Thermus thermophilus* dans la boucle A du centre PTC. L'absence de ce résidu Ψ

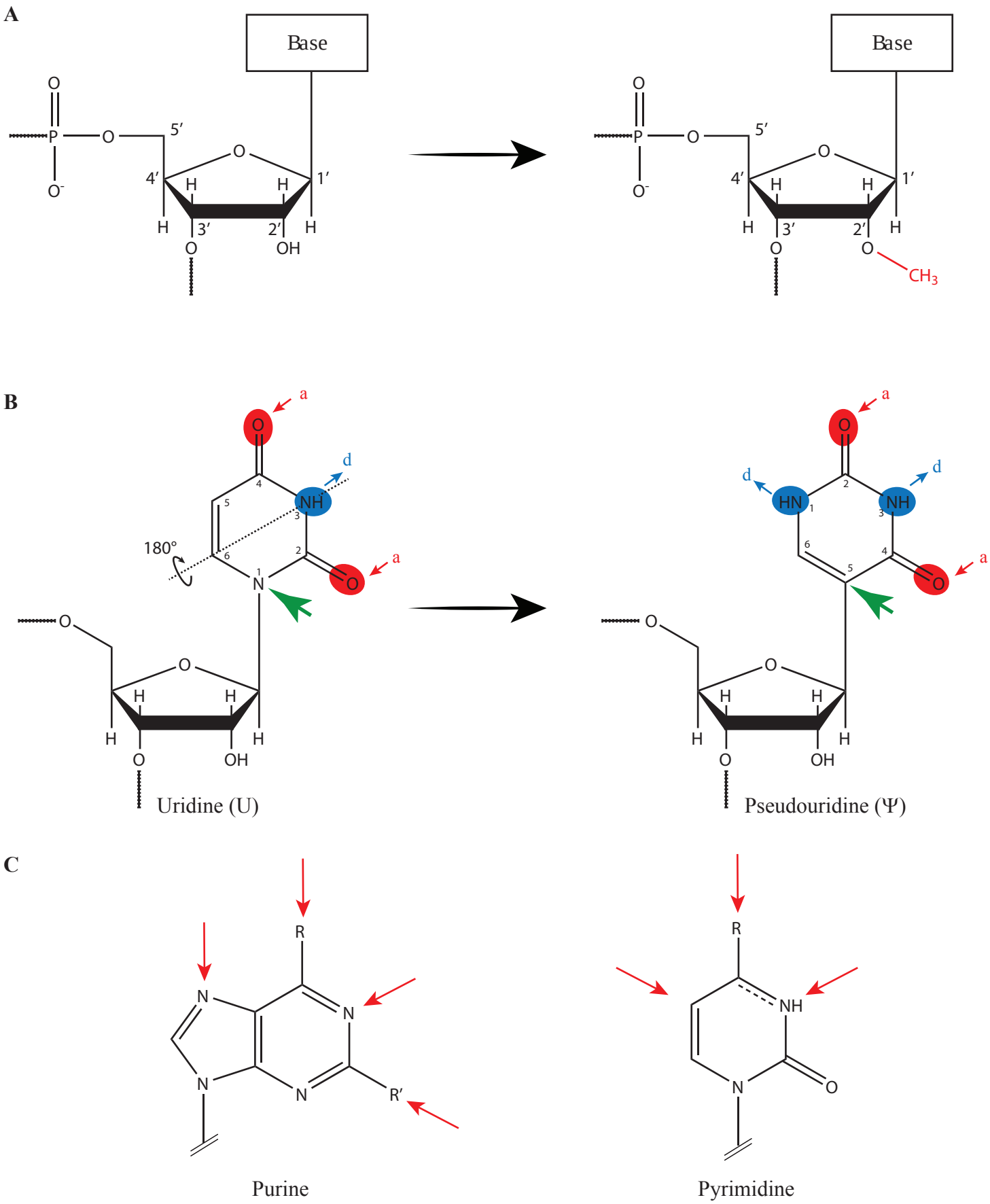


Figure 4. **Modifications post-transcriptionnelles les plus fréquentes dans les ARNr eucaryotes.** (A) Ajout d'un groupement méthyl en position 2' du ribose. (B) La conversion des résidus uridines en pseudouridines (pseudouridylation). Les groupements donneurs et accepteurs de liaisons hydrogène sont respectivement symbolisés en rouge et bleu. Les liaisons hydrogène possibles dans les sens accepteur (a) ou donneur (d) sont indiquées par des flèches avec le même code couleur. (C) Principaux sites de méthylation des bases purines et pyrimidines. Adapté de (Charette & Gray, 2000; Decatur & Fournier, 2002).

affecte la traduction mais n'a pas d'effet sur la synthèse de la sous-unité 40S (King *et al.*, 2003). Par ailleurs, le snR10 est lui-même nécessaire à la maturation nucléolytique de l'ARNr 18S (Tollervey, 1987) et donc à la production de la petite sous-unité 40S du ribosome (King *et al.*, 2003).

D'un point de vue structural, il a été montré par RMN en solution que les résidus Ψ et Nm ont le potentiel de stabiliser le repliement des ARN (Cabello-Villegas & Nikonowicz, 2005; Davis, 1995; Rife & Moore, 1998).

1.2.3.1 La pseudouridylation

La pseudouridylation est la modification la plus répandue dans les ARN, elle est présente de façon ubiquitaire dans les ARNt et fréquemment retrouvée dans les ARNr et les petits ARN nucléaires (Krol *et al.*, 1981; Limbach *et al.*, 1994; Maden, 1990; Massenet *et al.*, 1998; Steinberg *et al.*, 1993) au point d'être appelée le « cinquième nucléoside ». Les pseudouridines jouent un rôle important dans les ARNr de par leur positionnement dans des zones fonctionnellement importantes. Par exemple, dans l'ARNr 26S de levure, l'équipe de J. Ofengand a montré que 3,5% des résidus uridines sont convertis en pseudouridines, lesquelles se retrouvent comme chez *E.coli* au sein du centre PTC (Bakin *et al.*, 1994; Bakin & Ofengand, 1993; Veldman *et al.*, 1981). Les Ψ ont un effet subtil mais néanmoins significatif, qui tend à rigidifier le squelette sucre-phosphate et à augmenter l'empilement des bases. De plus, la pseudouridylation permet d'obtenir un site donneur de liaison hydrogène supplémentaire, ce qui stabilise les interactions ARN-ARN et ARN-protéine, pour revue (Charette & Gray, 2000). Il a également été proposé, à partir d'études RMN montrant un échange lent du proton N1-H de la Ψ avec le solvant (Figure 4), qu'il peut s'établir un pont aqueux intramoléculaire de manière stable (Davis & Poulter, 1991). Cette observation ayant été faite dans le contexte d'un $ARNt^{Phe}$ d'*E.coli*, la protection du proton N1-H peut également être expliquée par son environnement au sein de la structure tridimensionnelle. Le proton N1-H formerait une liaison hydrogène avec une molécule d'eau coordonnée avec les groupements oxygènes du résidu nΨ et n-1 de la chaîne phosphodiester de l'ARN (Figure 5). Une confirmation de l'existence de cette liaison a été obtenue par une analyse radiocristallographique d'un $ARNt^{Gln}$ complexé avec sa Gln-tRNA synthétase (Arnez & Steitz, 1994) et par une étude par RMN de l'effet structural de la présence d'un résidu Ψ dans un tétramère d'ARN AAΨA en solution (Davis, 1995). Les résultats obtenus confirment la coordination d'une molécule d'eau par le proton N1-H du résidu Ψ. Cette faculté serait

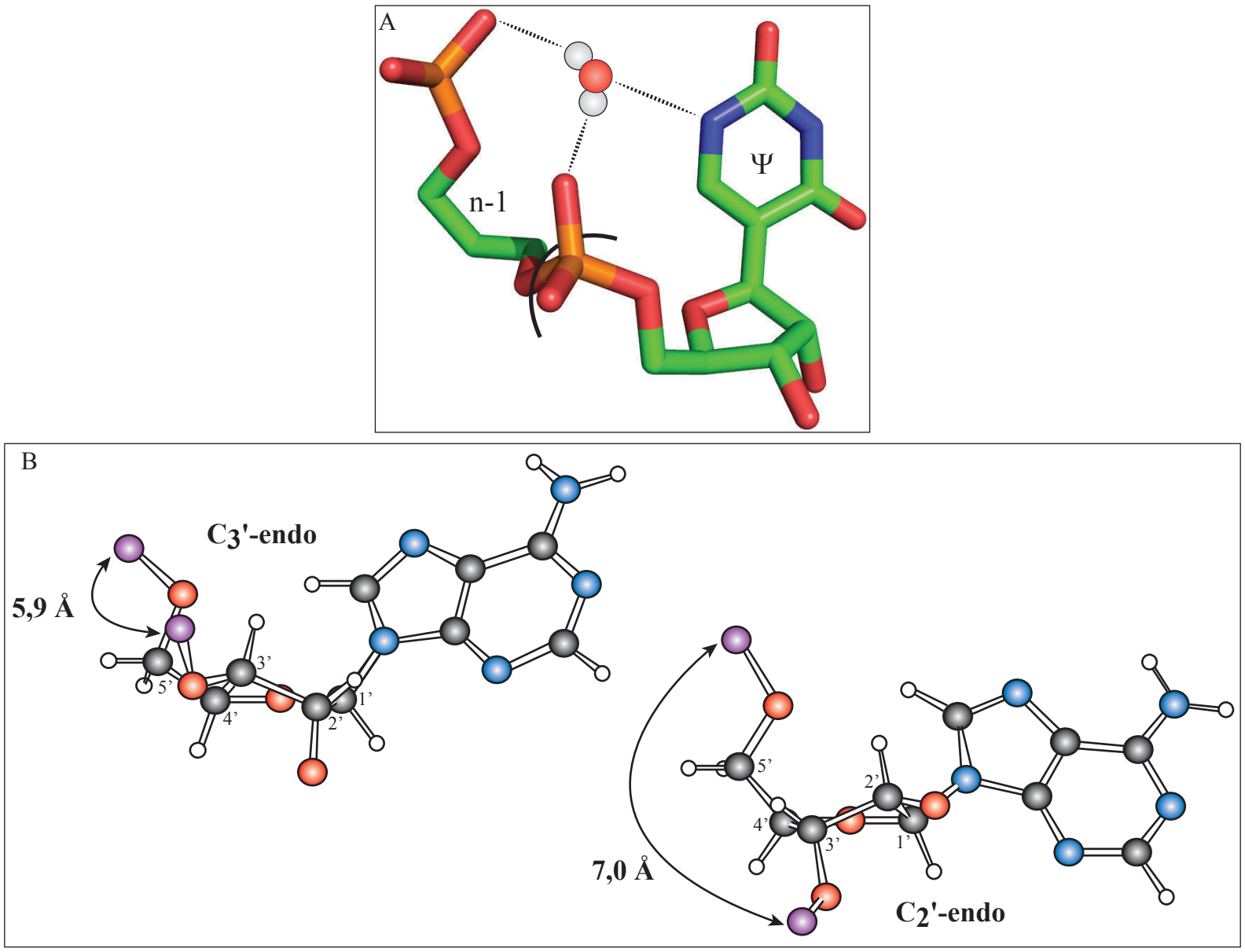


Figure 5. **Représentation schématique des conséquences structurales des pseudouridylations et 2'-O-methylations des ARN.** (A) Liaisons hydrogène pouvant être formées entre une molécule d'eau (représentée en sphères) et les oxygènes du squelette phosphodiester et le proton N1-H de la pseudouridine. (B) Comparaison entre la conformation C2' et C3' endo d'un résidu Adénosine. La distance entre les atomes de phosphates (en violet) est indiquée (Saenger, 1984). (C) Représentation en mode bâton de résidus 3'-5' diphosphate 2'-O-méthylés (purine (Adénosine) et pyrimidine (Uridine)). Adapté de (Charette & Gray, 2000; Kawai *et al.*, 1992; Prusiner *et al.*, 1974).

intrinsèque à ce résidu et indépendante de la séquence ou du contexte simple ou double brins de l'ARN. Ceci expliquerait le gain local de stabilité structurale observé au sein des ARN possédant une Ψ. Il a également observé que le résidu Ψ favorise la conformation 3'endo du sucre par rapport au résidu uridine. Ceci restreint la base du nucléotide dans une conformation *anti* et stabilise les résidus voisins, en particulier ceux précédant le site modifié, dans une conformation favorisant l'empilement des bases (Charette & Gray, 2000; Davis, 1995; Davis *et al.*, 1998).

On peut également rajouter que lors de la conversion d'un résidu Uridine en résidu Ψ on passe d'une interaction très faible C5-H$\cdots$O$_{eau}$ pour un résidu U à une plus forte interaction hydrogène N1-H$\cdots$O$_{eau}$ pour un résidu Ψ (Auffinger & Westhof, 1998; Charette & Gray, 2000) ce qui a également pour effet de réduire la flexibilité conformationnelle de la base et du squelette phosphodiester à proximité. Il a donc été proposé que les résidus Ψ améliorent l'activité du ribosome en modulant le repliement et les interactions des ARNr. La pseudouridylation diminue le caractère hydrophobe du résidu et peut donc être vue comme une stratégie antagoniste à la méthylation pour un ajustement fin de la stabilité et de l'hydrophobicité des structures locales d'ARN (Davis, 1995).

1.2.3.2 La 2'-O-méthylation

La méthylation augmente le caractère hydrophobe du nucléotide, protège le nucléotide des attaques nucléolytiques en empêchant la déprotonation de la fonction alcool en position 2' du ribose. Elle stabilise les hélices par l'augmentation de l'empilement des bases, augmente la thermostabilité des ARN et peut renforcer les interactions inter ou intramoléculaires (Kawai *et al.*, 1992). Cette modification stabilise aussi la conformation C3' endo d'un point de vue thermodynamique (- 0,8 kcal.mol^{-1}) or cette conformation est favorable à la formation des hélices d'acides nucléiques. Cet effet, intrinsèque au résidu, est dû à la répulsion stérique qui s'établirait entre le groupement carbonyl en position 2 (pour les bases pyrimidine), le groupement méthyl en 2'-O et le groupement phosphate en 3', si le résidu était dans une conformation C2' endo (Figure 5). On peut observer les mêmes effets conformationnels lors d'une 2-thiolation des résidus (Kawai *et al.*, 1992). Ces effets ont été observés seulement sur de petits ARN modèles et on connaît encore peu l'effet structural de ces modifications sur les ARNr individuels ou sur le ribosome lui-même.

Il semble qu'à un niveau individuel, la plupart des modifications apportent de petits avantages non essentiels mais que l'effet cumulé de ces modifications apporte un avantage

non négligeable à la cellule. Cette vision est appuyée par la nature diverse des machineries de modifications responsables de la synthèse des résidus Ψ et Nm (la machinerie guidée des snoRNP par rapport aux enzymes purement protéiques). L'évolution parallèle mais schématiquement différente pour créer le même type de modification, suggère que ces modifications doivent être bénéfiques. La modification chimique des nucléotides pourrait ainsi constituer un moyen d'affiner la structure des ARNr de sorte à optimiser les différents aspects fonctionnels des ARNr (Decatur & Fournier, 2002).

1.2.4 Evolution phylogénétique des modifications des résidus : nombre et position

Les premières analyses structurales des ARN ribosomiques ont été réalisées par modifications chimiques, digestions par des RNAses, études d'hybridation d'ADNc mais également sur la base de conservations de séquences. Au cours des années 1970 à 1985, les structures des ARNr de la petite et la grande sous-unité du ribosome ont fait l'objet de nombreuses études pour aboutir à la détermination de leurs séquences et de leurs structures secondaires (Figures 6 et 7). Les premières données structurales sur l'ARNr de la petite sous-unité ont été obtenues dès le début des années 1980 (Noller & Woese, 1981; Woese *et al.*, 1980), tout comme pour la grande sous-unité (Branlant *et al.*, 1981; Fox & Woese, 1975; Noller *et al.*, 1981; Veldman *et al.*, 1981). L'ensemble de ces informations a fait l'objet d'une banque de données consultable à l'adresse http://www.rna.icmb.utexas.edu/ (Cannone *et al.*, 2002).

L'ARNr de la petite sous-unité a été décomposée en quatre domaines avec une structure en pseudo nœud centrale (Figure 6). On retrouve le domaine 5', le domaine Central correspondant à la Plateforme, le domaine 3' majeur localisé dans la tête et le bec et enfin le domaine 3' mineur composé de : *i*) l'hélice 44 intervenant dans le centre de décodage et dans l'interface entre les deux sous-unités, *ii*) la séquence anti Shine-Delgarno.

Six domaines ont été définis dans l'ARNr de la grande sous-unité dont le domaine V, qui forme le PTC et participe à la formation des sites A et P (Figure 7).

En comparant le nombre et la position des modifications des ARNr au sein des organismes *Escherichia coli* (35 modifications), *Saccharomyces cerevisiae* (109 modifications) et *Homo sapiens* (plus de 206 modifications), nous nous sommes très vite rendu compte de la conservation des modifications à certaines positions (Tableau 1) mais

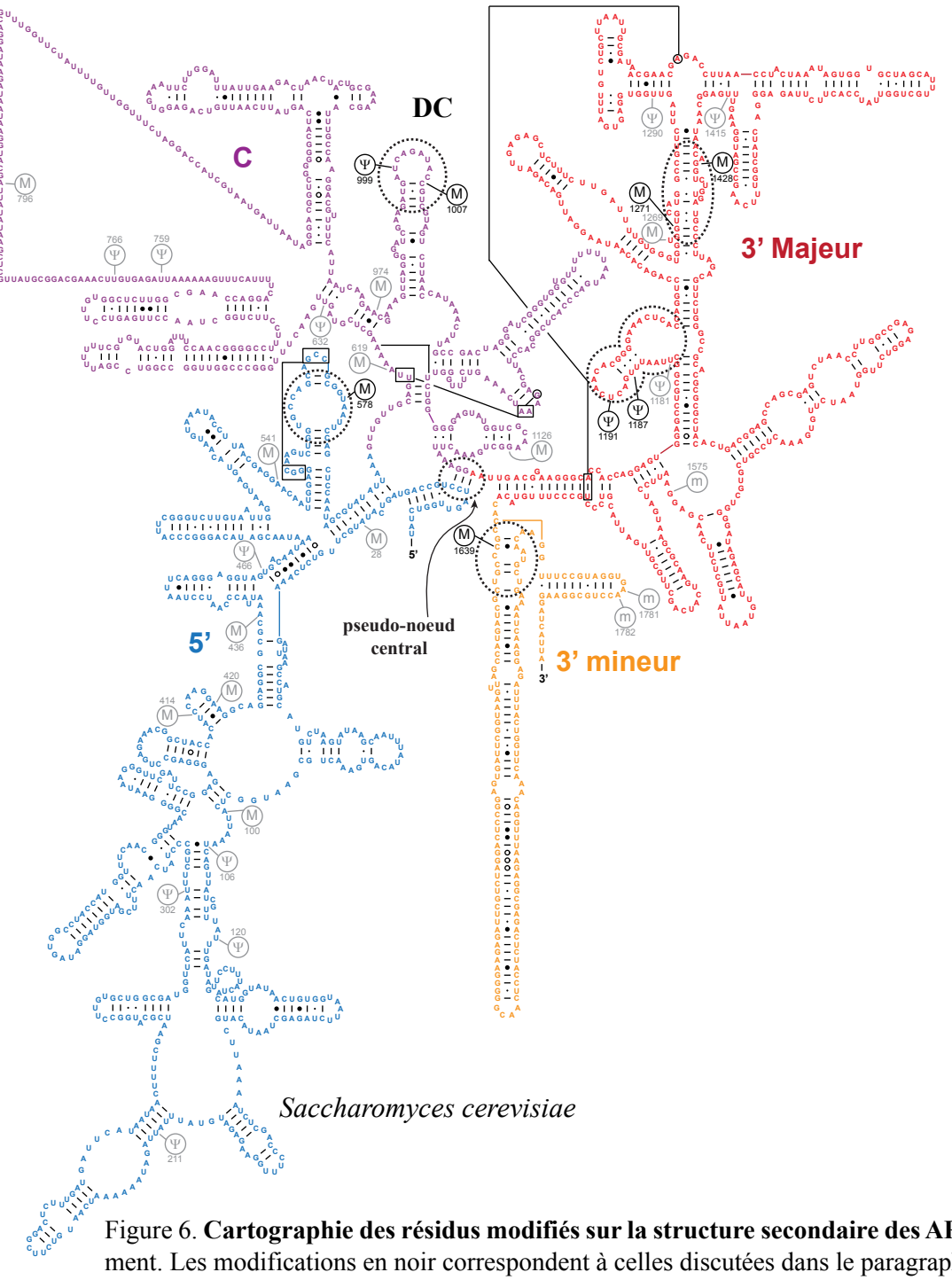


Figure 6. **Cartographie des résidus modifiés sur la structure secondaire des ARNr 18S et 16S de la petite sous unité du ribosome** de *S. cerevisiae* et *T. thermophilus* respectivement. Les modifications en noir correspondent à celles discutées dans le paragraphe 1.3.3.1. Leurs équivalents par homologie ont été placés sur l'ARN 16S de *T. thermophilus* dont la structure tridimensionnelle est utilisée dans la figure 10. Lorsque la position modifiée chez la bactérie n'existe pas, seul le numéro du résidu est indiqué. Les résidus formant le centre de décodage (DC) sont entourés en pointillés sur l'ARNr de levure. Les domaines structuraux de la plateforme, de la tête et du bec qui sont délimités en pointillés sur l'ARNr bactérien reprennent les mêmes couleurs que pour la figure 10. Les différents domaines de la petite sous unité du ribosome sont mentionnés. D'après (Baudin-Baillieu *et al.*, 2009), http://people.biochem.umass.edu/sfournier/fournierlab/snornadb/ et (Cannone *et al.*, 2002), http://www.rna.ccbb.utexas.edu/.

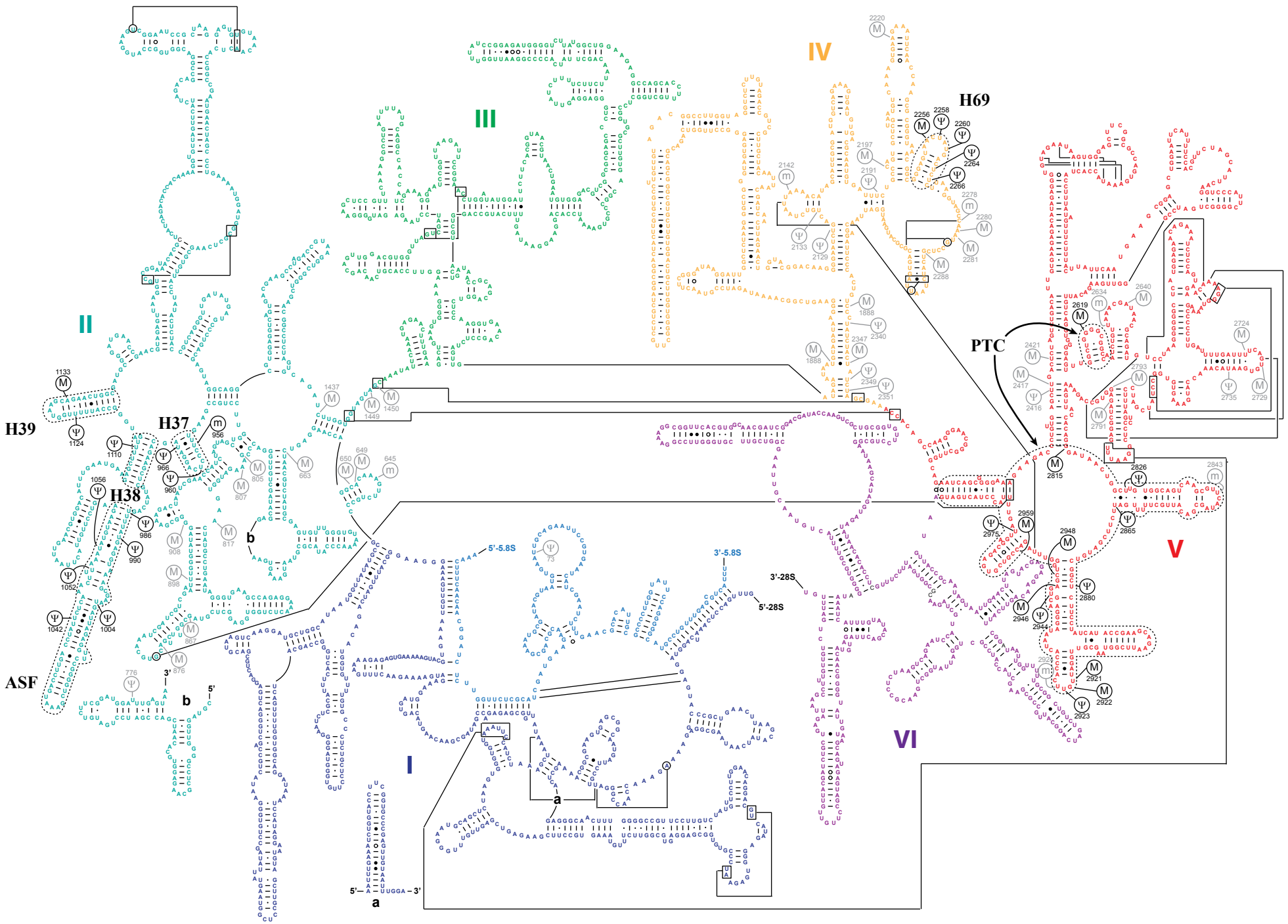


Figure 7. **Cartographie des résidus modifiés sur la structure secondaire des ARNr 5.8S et 25S de la grande sous unité du ribosome de *S. cerevisiae.*** Les modifications en noir correspondent à celles discutées dans le chapitre 1.3.3.2. Les résidus formant le centre de transfert peptidique (PTC), l'hélice 69 (H69) et le A-site finger (ASF) sont entourés en pointillés. Les chiffres romains correspondent aux domaines de la grande sous-unité du ribosome. D'après (Baudin-Baillieu *et al.*, 2009), http://people.biochem.umass.edu/sfournier/fournierlab/snornadb/ et (Cannone *et al.*, 2002), http://www.rna.ccbb.utexas.edu/.

Petite sous unité	*S.cerevisiae*		*T.thermophilus*	*E.coli*
Centre de décodage	snRP	Position modifiée		
	77	Um578	U515	U531
	31	Ψ999	U772	U788
	79	Cm1007	C780	C796
	36	Ψ1187	C940	C962
	35	$m^1acp^3Ψ1191$	$m^2{}_2G944$	m^2G966
	40	Gm1271	C1032	C1051
	56	Gm1428	G1177	A1197
	70	Cm1639	$m^4Cm1381$	$m^4Cm1402$

Grande sous unité	*S.cerevisiae*		*T.thermophilus*	*E.coli*
Centre PTC	snRP	Position modifiée		
	67	Gm2619	Gm2262	Gm2251
	38	Gm2815	G2457	G2446
	34	Ψ2826	U2468	Ψ2457
	46	Ψ2865	C2507	Cm2498
	34	Ψ2880	U2522	U2511
	52	Um2921	Um2563	Um2552
	Spb1p	Gm2922	G2564	G2553
	10	Ψ2923	U2565	U2554
	37	Ψ2944	C2586	C2575
	71	Am2946	A2588	A2577
	69	Cm2948	C2590	C2579
	73	Cm2959	A2601	A2590
	42	Ψ2975	C2617	C2606
L'hélice 69	63	Am2256	A1934	A1913
	191	Ψ2258	mU1936	$m^3Ψ1915$
	191	Ψ2260	Ψ1938	Ψ1917
	3	Ψ2264	G1942	G1921
	84	Ψ2266	U1944	U1923
A-site finger	?	m^5U956	A871	U824
	8	Ψ960	U875	U828
	43	Ψ966	C881	G834
	8	Y986	U898	C851
	49	Ψ990	G902	G855
	5	Ψ1004	U915	U868
	33	Ψ1042	U951	A905
	81	Ψ1052	C961	C915
	44	Ψ1056	G965	U919
	82	Ψ1110	G984	G939
	5	Ψ1124	A998	G953
	61	Am1133	U1008	U963

Tableau 1. **Récapitulatif des positions des résidus modifiés dans les régions fonctionnelles du ribosome de *S. cerevisiae* et leurs équivalents chez *T. thermophilus* et *E. coli***. Les snoRNP de *S. cerevisiae* catalysant ces modifications sont indiquées. D'après (Piekna-Przybylska *et al.*, 2007), http://people.biochem.umass.edu/sfournier/fournierlab/snornadb/.

aussi de l'augmentation du nombre de modifications lorsque l'on passe d'un système de synthèse purement protéique (bactéries) au système guidé eucaryote (Branlant *et al.*, 1981; Decatur & Fournier, 2002; Ofengand & Bakin, 1997; Ofengand *et al.*, 1995; Veldman *et al.*, 1981).

De plus, lorsque l'on replace les modifications de résidus sur la structure des ARNr, on observe que les régions riches en modifications sont celles impliquées dans les évènements spécifiques de la traduction ou encore dans l'interaction des deux sous-unités (Branlant *et al.*, 1981; Veldman *et al.*, 1981) (Figures 6 et 7). Il apparait aussi que ces modifications sont absentes des domaines où se fixent des protéines ribosomiques : la surface externe et la périphérie de la région de l'interface. L'absence de modifications dans ces régions indique que la majorité des interactions ARN-protéines n'est sans doute pas affectée directement par ces modifications (Decatur & Fournier, 2002). Il existe donc une forte corrélation entre les sites des modifications post-transcriptionelles et les régions fonctionnellement importantes (65 des 109 modifications soit 60% chez la levure, et 33 des 35 modifications chez *E. coli* soit 94 % (Decatur & Fournier, 2002)). C'est le cas par exemple de la région des 3 sites A, P et E pour l'interaction entre les ARNt et les ARNm.

1.3 Les résidus modifiés au sein du ribosome

Comme nous venons de le mentionner, les résidus modifiés se situent principalement dans les régions fonctionnelles du ribosome, c'est-à-dire les sites où se déroulent les étapes nécessaires à la traduction des ARNm. Afin de mieux appréhender le rôle de chacune de ces régions, il est important de rappeler le mécanisme de la traduction des ARNm

1.3.1 Fonction du ribosome

Le ribosome est un acteur essentiel pour la biogenèse des composants cellulaires et représente l'une des machineries les plus consommatrices d'énergie. Grâce aux ARNt, il permet le décodage de l'information génétique portée par les ARNm et catalyse la formation des liaisons peptidiques. La traduction peut être divisée en 3 étapes principales : l'initiation (qui est l'étape limitante), l'élongation et la terminaison (pour revue (Schmeing & Ramakrishnan, 2009)). Bien que les étapes d'élongation et de terminaison soient très similaires entre les bactéries et les eucaryotes, l'étape d'initiation repose sur des mécanismes différents, l'initiation bactérienne étant actuellement la mieux connue structuralement.

1.3.1.1 L'initiation

L'initiation bactérienne a la particularité de requérir une formylation de l'ARNt initiateur (fMet-ARNt^{fMet}). La reconnaissance du codon initiateur de l'ARNm par le ribosome est facilitée chez les bactéries par l'appariement de l'extrémité 3' de l'ARNr 16S à la séquence Shine-Delgarno de l'ARNm (également appelé *Ribosome Binding Site* ou RBS). Cette séquence précède le codon initiateur. Trois facteurs interviennent lors de cette étape, ce sont les facteurs d'initiations IF-1, IF-2 et IF-3. Parmi eux, seul IF-2 possède une activité GTPase fournissant l'énergie nécessaire aux différents changements conformationels qui se produisent lors de l'association des deux sous-unités, lors du transfert de l'ARNt-fMet vers le site PTC, ou encore lors de la rotation de la sous-unité 30S (Figure 8).

Chez les Eucaryotes, le mécanisme de base de l'initiation (Figure 9) fait appel à 9 facteurs d'initiation. L'ARNt initiateur (Met-ARNt$_i$) s'associe au facteur d'initiation eIF2 couplé à une molécule de GTP pour former un complexe ternaire (étape 2). Ce complexe peut alors s'associer à la petite sous-unité 40S du ribosome eucaryote afin de positionner l'ARNt initiateur au niveau du site P. La présence des facteurs eIF1, eIF1A et eIF3 sur la sous-unité 40S modifie sa conformation et facilite son association avec le complexe ternaire pour former le complexe 43S de pré-initiation (PIC) (étape 3). La coiffe m^{7}G de l'ARNm est liée au complexe CBC (*Cap Binding Complex*) lors de la synthèse de l'ARNm dans le noyau. Elle est ensuite reconnue par le complexe eIF4F dans le cytoplasme (étape 4) qui déplace le complexe CBC. Ce complexe eIF4F est composé des facteurs eIF4A, eIF4G et eIF4E, eIF4G reconnaissant directement la coiffe. Il recrute le complexe de pré-initiation 43S au niveau de la région 5'-UTR de l'ARNm (étape 5). Ce dernier est composé de la sous-unité 40S, des facteurs eIF1, eIF1A et eIF3 et du complexe ternaire. Il se déplace alors en direction de l'extrémité 3' de l'ARNm jusqu'au codon initiateur, c'est l'étape de « *scanning* » (étape 6) (Gross *et al.*, 2003). Chez les eucaryotes, l'ARNm adopte une forme cyclique grâce à des interactions entre eIF4G et les PABP (*Poly A Binding Protein*) fixées sur la queue poly A (von der Haar *et al.*, 2004). Lors de la reconnaissance du codon initiateur, eIF2 hydrolyse le GTP dans une réaction favorisée par eIF5 (étape 7). La reconnaissance du codon de l'ARNm par la boucle anticodon de l'ARNt$_i$-Met au sein du complexe 43S déclenche le relargage d'eIF1 de son site de fixation et la dissociation du pyrophosphate (Pi) associé au facteur eIF2, conduisant ainsi au complexe d'initiation 48S. Le complexe eIF2-GDP ainsi formé est associé à la sous-unité 40S de manière instable et est facilement éliminée. Grâce à une seconde réaction dépendante du GTP, le facteur eIF5B favorise l'association de la grande sous-unité

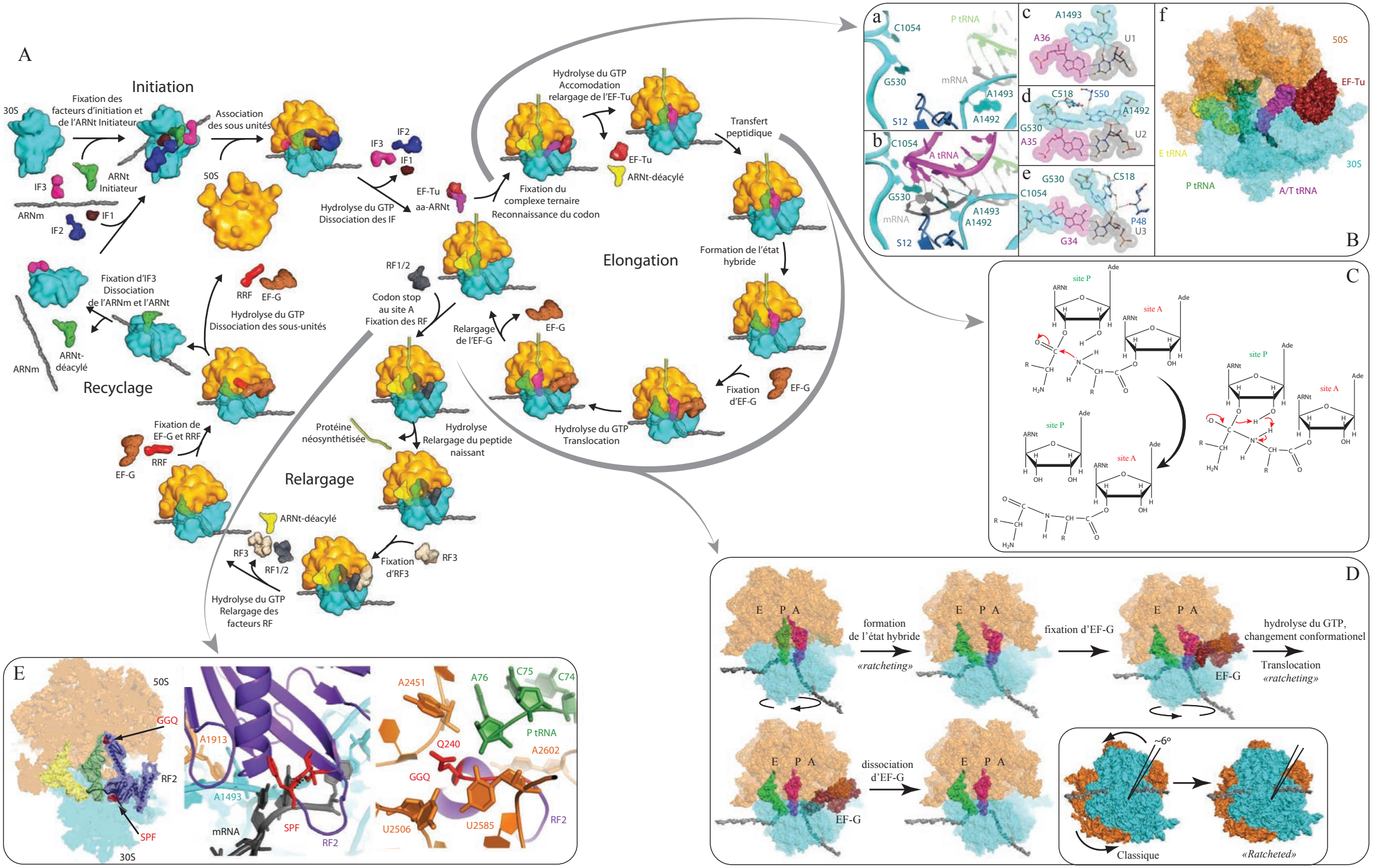


Figure 8. **Schéma simplifié de la traduction bactérienne.** Les représentations en surface moléculaire des différentes sous-unités du ribosome, des ARNt et des différents facteurs sont basées sur les structures 3D établies. Le décodage du codon par la boucle anti-codon de l'ARNt, la formation de la liaison peptidique, le mouvement de cliquet (ou *« ratcheting »*) au cours de l'élongation et l'association du facteur de relargage RF2 au site A du ribosome feront l'objet d'une description plus détaillée dans le paragraphe 1.3.1 du manuscrit. D'après (Schmeing & Ramakrishnan, 2009).

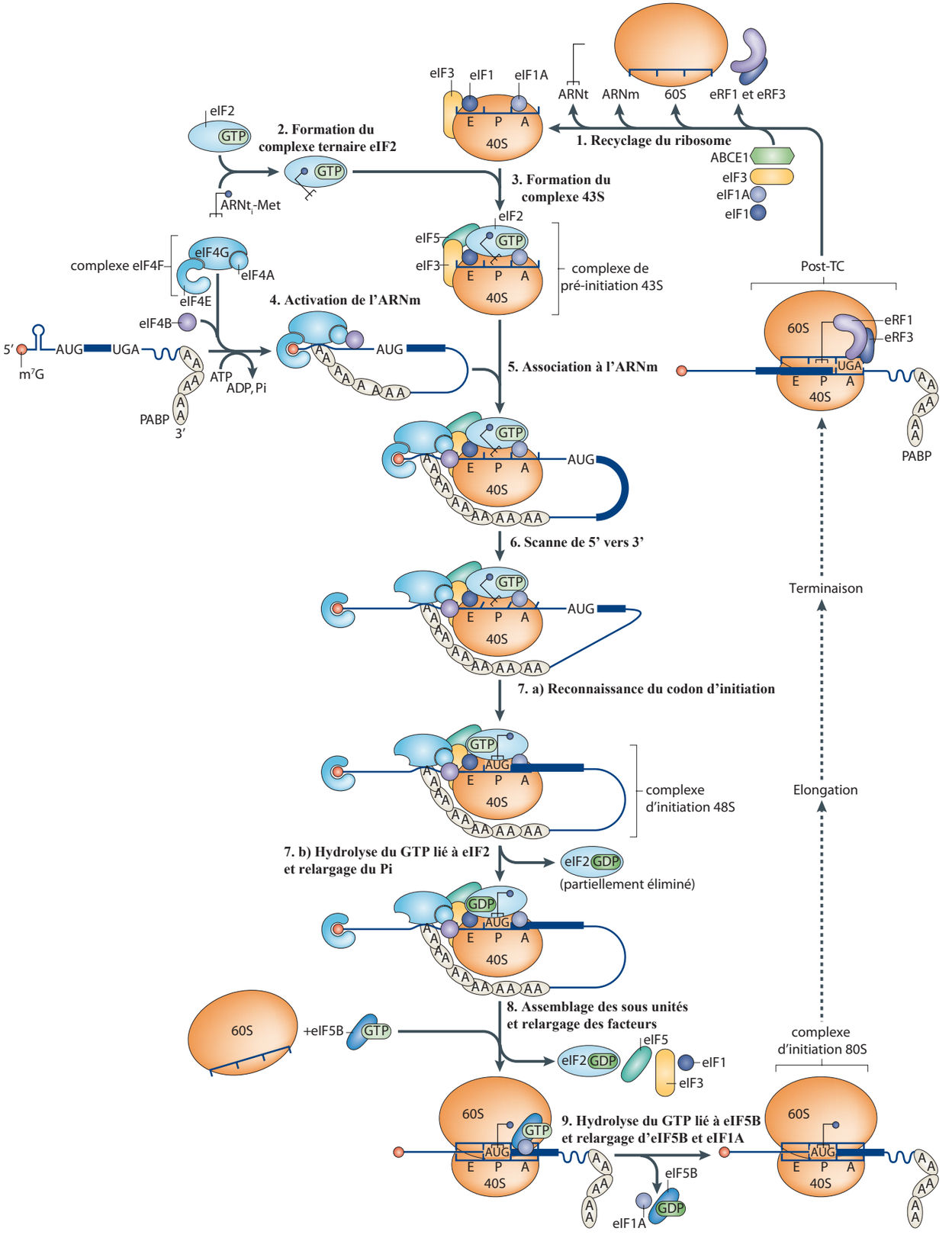


Figure 9. **Modèle d'initiation de la traduction chez les Eucaryotes.** La voie classique d'initiation de la traduction chez les eucaryotes peut être divisée en 8 étapes (2 à 9) qui conduisent à la formation d'un complexe ribosomique d'initiation de la traduction 80S au sein duquel l'ARNt^{Met} initiateur est apparié au codon initiateur fixé au site P du ribosome et qui est compétant pour démarrer les étapes d'élongation de la traduction. Adapté de (Jackson *et al.*, 2010).

ribosomique 60S sur le complexe d'initiation 48S (étape 8). L'hydrolyse du GTP par le facteur eIF5B suivie de l'association des sous-unités permet la dissociation des facteurs eIF5B et eIF1A du complexe d'initiation 80S, laissant l'ARNt$_i$-Met hybridé au codon initiateur dans le site P et l'extrémité 3' de l'ARNm libre (étape 9). Le ribosome est alors prêt à entrer dans la phase d'élongation de la synthèse protéique (pour revue (Jackson *et al.*, 2010; Lorsch & Dever, 2010)).

Bien que la majorité des ARNm utilisent le mécanisme d'initiation par « *scanning* », l'initiation de la traduction de certains ARNm fait appel au système IRES (« *Internal Ribosome Entry Sites* » ou sites d'entrée interne du ribosome). Ce système, indépendant de la coiffe, fait appel à des éléments de structure de l'ARNm pour recruter les facteurs d'initiation de la traduction et démarrer la traduction de manière interne. Les IRES sont en particulier utilisés par certains virus afin de détourner à leur profit la machinerie cellulaire de traduction ou pour permettre la traduction de plusieurs cistrons à partir d'un seul transcrit (Figure 10) (Jackson *et al.*, 2010). Des IRES sont aussi présents dans des ARNm cellulaires, en particulier des ARNm dont le taux d'expression varie dans les cellules tumorales (pour revue, (Stoneley & Willis, 2004)).

1.3.1.2 L'élongation

Le cycle de l'élongation consiste en l'addition d'un acide aminé supplémentaire à l'extrémité C-terminale de la chaîne polypeptidique en cours de synthèse. Au début d'un cycle d'élongation, le ribosome contient dans son site P un ARNt lié à l'acide aminé initiateur et son site A est vide. L'acide aminé suivant est apporté sous la forme d'un complexe ternaire eEF-1α/GTP/ARNt aminoacylé chez les eucaryotes (EF-Tu/GTP/ARNt aminoacylé chez les procaryotes). La boucle anti-codon du nouvel ARNt s'apparie alors avec le codon de l'ARNm (Figure 8B) et une nouvelle liaison peptidique est formée entre l'acide aminé du nouvel ARNt et l'acide aminé initiateur ou, au cycle suivant, la chaîne polypeptidique naissante (Figure 8C). Le facteur eEF-2 des eucaryotes (EF-G chez les procaryotes) catalyse ensuite un mouvement de translocation des ARNt et de l'ARNm au sein du ribosome afin de déplacer l'ARNt du site P au site E et le « nouvel » ARNt portant la chaîne polypeptique du site A vers le site P. Le site A est alors disponible pour accueillir un nouvel ARNt aminoacylé et pour initier un nouveau cycle d'élongation (Figure 8D).

L'étape de décodage (Figure 8, panneau B) permet la sélection fidèle de l'ARNt aminoacylé, dictée par l'ARNm, au site A du ribosome. La fidélité de la traduction ne repose

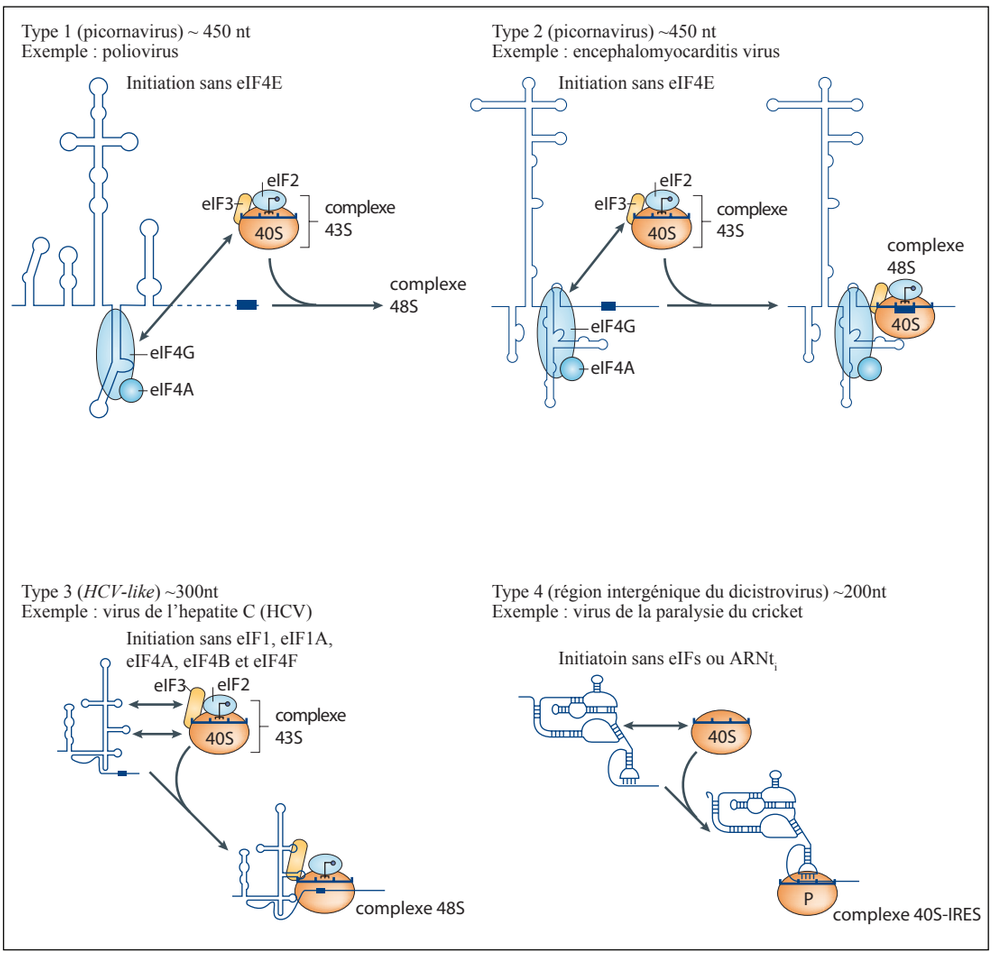


Figure 10. **Initiation de la traduction par le système IRES.** Les sites internes d'entrée du ribosome (IRES) sont des éléments structuraux qui médient le recrutement du ribosome indépendamment de la coiffe en 5' des ARNm sur des sites situés au sein de l'ARNm. Les IRES reposent sur des interactions directes avec les facteurs d'initiation eucaryotes et/ou la sous-unité 40S du ribosome. Les IRES de type 1 et 2 s'associent spécifiquement avec le domaine p50 du facteur eIF4G et cette interaction est facilitée par eIF4A. Le complexe eIF4G-A recrute alors le complexe 43S sur l'élément IRES en absence d'eIF4E. Les IRES de type 3 s'associent au facteur eIF3 et la sous-unité 40S composant le complexe 43S indépendamment des facteurs eIF4F, eIF4B, eIF1 et eIF1A. Quant aux IRES de type 4, ils sont capables de s'associer à la sous unité 40S du ribosome en absence des facteurs d'initiation et de l'ARNt^{Met} initiateur car le site P est occupé par un domaine de l'IRES qui mime l'interaction entre le codon et l'anticodon. Ainsi l'initiation médiée par les IRES est insensible aux mécanismes de régulation de la traduction tels que la phosphorylation d'eIF2 (pour le type 4) et/ou la séquestration du facteur eIF4E. Adapté de (Jackson *et al*., 2010).

pas uniquement sur la différence d'énergie entre un appariement correct et un appariement non correct entre le codon et l'anticodon (Ogle & Ramakrishnan, 2005; Xia *et al.*, 1998). Les résidus A1492, A1493 et G530 chez *E.coli* (équivalent aux résidus A1466, A1467 et G514 chez *T. thermophilus* et A1755, A1756 et G577 chez *S.cerevisiae*) interviennent dans la reconnaissance codon-anticodon au site A du ribosome en établissant des interactions avec les deux résidus aux premières positions du codon, ceci au travers du sillon mineur de la mini hélice formée par l'appariement du codon de l'ARNm et de l'anticodon de l'ARNt (Noller, 2006; Schmeing & Ramakrishnan, 2009). La mise en place de ces interactions constitue une étape de discrimination des mauvais anticodons. Cette étroite surveillance des appariements par le ribosome ne s'établie pas au niveau de la troisième position du codon, appelée *« wobble position »* qui participe à la dégénérescence du code génétique. L'énergie d'interaction de ces associations supplémentaires ne serait pas dans un premier temps utilisée pour augmenter l'affinité relative entre l'ARNt hybridé et l'ARNt quasi-hybridé mais plutôt pour induire une fermeture d'un domaine de la petite sous-unité 30S (Ogle *et al.*, 2002) qui conduit probablement à l'accélération du décodage (Gromadski & Rodnina, 2004).

Dans la forme apo du ribosome, les résidus A1492 et A1493 sont pris dans un empilement de bases de l'hélice h44 (Figure 8B panneau a). Lorsqu'un ARNt s'hybride à l'ARNm au site A du ribosome, la conformation des résidus A1492, A1493 et G530 est modifiée pour permettre une interaction avec le sillon mineur de la mini hélice ARNm-ARNt (Ogle *et al.*, 2001) (Figure 8B, panneau b). Ces résidus du ribosome contrôlent la géométrie des paires de bases à la première (Figure 8B, panneau c) et à la seconde position de la mini hélice (Figure 8B, panneau d). Les résidus S50 et P48 de la protéine ribosomique S12 interagissent également avec la seconde (Figure 8B, panneau d) et la troisième position (Figure 8B, panneau e) respectivement. Lors du chargement de l'ARNt aminoacylé au site A du ribosome par le facteur EF-Tu, conduisant au complexe ternaire, la tige-boucle de l'anticodon est fortement pliée (Schuette *et al.*, 2009; Villa *et al.*, 2009) (Figure 8B, panneau f).

La formation de la liaison peptidique est l'étape clef de la synthèse peptidique. Il est proposé qu'elle repose sur une attaque nucléophile du groupement amine de la chaîne principale de l'acide aminé porté par l'ARNt au site A du ribosome sur le carbone ester du dernier résidu de la chaîne polypeptidique en cours de synthèse portée par l'ARNt peptidylé (Figure 8C). Cette réaction a lieu au niveau du domaine V de l'ARNr 23S (28S chez la levure) de la grande sous-unité du ribosome qui fixe les extrémités $C_{74}C_{75}A_{76}$ des ARNt

aminoacylé et peptidylé : le domaine PTC (Figure 7). Il est intéressant de noter que ce site actif est essentiellement composé d'ARN qui doit donc jouer un rôle majeur dans la catalyse.

Après la formation de la liaison peptidique, la chaîne peptidique naissante est transférée à l'ARNt au site A du ribosome laissant un ARNt déacétylé au site P. Avant de procéder à un second cycle d'élongation, les ARNt et l'ARNm doivent donc se déplacer au sein du ribosome, c'est ce que l'on appelle la translocation (Figure 8D). Au cours de cette étape, L'ARNm se déplace précisément de trois bases, soit un codon. Les ARNt quant à eux passent du site A au site P et du site P au site E. Ces transferts se produisent au cours d'un mouvement de cliquet de la petite sous-unité 30S du ribosome également appelé *« ratcheting »*. Après le transfert peptidique, les ARNt se déplacent spontanément des sites A et P vers les sites P et E respectivement au sein du ribosome dans un état *« ratcheted »* dans lequel vient se fixer le facteur EF-G. Après l'hydrolyse du GTP par EF-G la petite sous-unité effectue un mouvement de cliquet inverse et le facteur EF-G se dissocie (Figure 8D). Un mouvement de cliquet implique une rotation d'environ 6° de la petite sous-unité du ribosome (Figure 8D).

1.3.1.3 La terminaison

L'élongation se poursuit jusqu'à ce qu'un codon stop de l'ARNm se retrouve dans le site A du ribosome, signalant la fin de la traduction et engageant alors le processus de terminaison de la traduction. Les facteurs de relargage (RF) de classe I reconnaissent alors le codon stop et coupent la chaine polypeptidique naissante liée au dernier ARNt fixé au site P, ce qui libère la protéine néo-synthétisée (Figure 8). Chez les bactéries, il existe deux facteurs de relargage de classe I : (*i*) RF1 qui reconnait les codons UAA et UAG, (*ii*) RF2 qui reconnaît les codons UAA et UGA. Chez les eucaryotes, il existe un facteur unique, baptisé eRF1, qui reconnait les 3 codons stop. Les tri-peptide PXT du facteur bactérien RF1 et SPF du facteur RF2 leurs confèrent leur spécificités pour les codons UAG et UGA respectivement (Ito *et al.*, 2000), alors que le motif GGQ universellement conservé est impliqué dans l'hydrolyse du peptide par les facteurs de relargage (Figure 8E) (Frolova *et al.*, 1999; Song *et al.*, 2000). La fixation du facteur de relargage induit un changement de conformation des résidus conservés du centre de décodage G530, A1492 et A1493 et l'établissement de nouvelles interactions (Korostelev *et al.*, 2008; Laurberg *et al.*, 2008; Weixlbaumer *et al.*, 2008). Cette conformation est différente de celle observée au cours du décodage par les ARNt, où elle contrôlait la géométrie de l'appariement des bases (Ogle *et al.*, 2001), conformation incompatible avec l'association du facteur de relargage (Youngman *et al.*,

2007). Par exemple, le résidu A1493 vient former un nouveau contact entre les deux sous-unités en s'empilant sur le résidu A1913 de l'ARNr 23S (Figure 8E). Le motif GGQ se positionne dans le centre PTC dans une conformation mettant en jeu ses deux résidus glycines (Korostelev *et al.*, 2008; Laurberg *et al.*, 2008; Weixlbaumer *et al.*, 2008) (Figure 8E). La glutamine du motif GGQ est un résidu conservé et le groupement amino de sa chaîne latérale porte une méthylation dont la perte provoque une diminution de l'efficacité du relargage du peptide (Dincbas-Renqvist *et al.*, 2000). Il a été proposé que la chaîne latérale de cette glutamine coordonne directement une molécule d'eau pour une attaque nucléophile permettant le relargage du poly-peptide du ribosome (Weixlbaumer *et al.*, 2008).

Les facteurs de relargage de classe II (i.e. RF3 chez les bactéries et eRF-3 chez les eucaryotes) interviennent dans un second temps. Ils accélèrent en effet la dissociation des facteurs de classe I du ribosome après la libération du peptide.

Le recyclage des ribosomes est réalisé grâce à l'intervention d'un facteur de recyclage. Chez les bactéries, le facteur de recyclage RRF (*Ribosome Recycling Factor)* agit en coopération avec le facteur EF-G (Hirashima & Kaji, 1973). L'hydrolyse du GTP provoque alors la dissociation du ribosome en deux entités : la sous-unité 50S et un complexe 30S-ARNm-ARNt déacétylé. La dissociation de ce dernier sous complexe requiert l'intervention du facteur IF3 (eIF3 chez les eucaryotes). On se retrouve donc dans la configuration initiale permettant d'initier un nouveau cycle de traduction (Figure 8).

1.3.2 Structure du ribosome

La structure tridimensionnelle globale des ribosomes est présentée dans la figure 11. La petite sous-unité s'associe avec la grande sous-unité pour former l'entité fonctionnelle. L'interface entre les 2 sous-unités est majoritairement formée par les ARN. Les premières structures à basse résolution des ribosomes ont été obtenues par cryomicroscopie électronique (Frank *et al.*, 1995; Verschoor *et al.*, 1998). Sur la base de l'enveloppe de cette machinerie, les différent domaines ont été décrit comme la Plateforme, le bec ou encore la tête mentionnés dans le paragraphe 1.2.4.

Récemment l'équipe de Yusupov a réussi une première en déterminant les structures atomiques des ribosomes de *S. cerevisiae* et *T. thermophilus* à 3Å de résolution (Ben-Shem *et al.*, 2011; Jenner *et al.*, 2010). Ceci est d'autant plus remarquable que la structure de *T. thermophilus* contient un ARNm ainsi qu'un ARNt à chacun des sites A, P et E du ribosome.

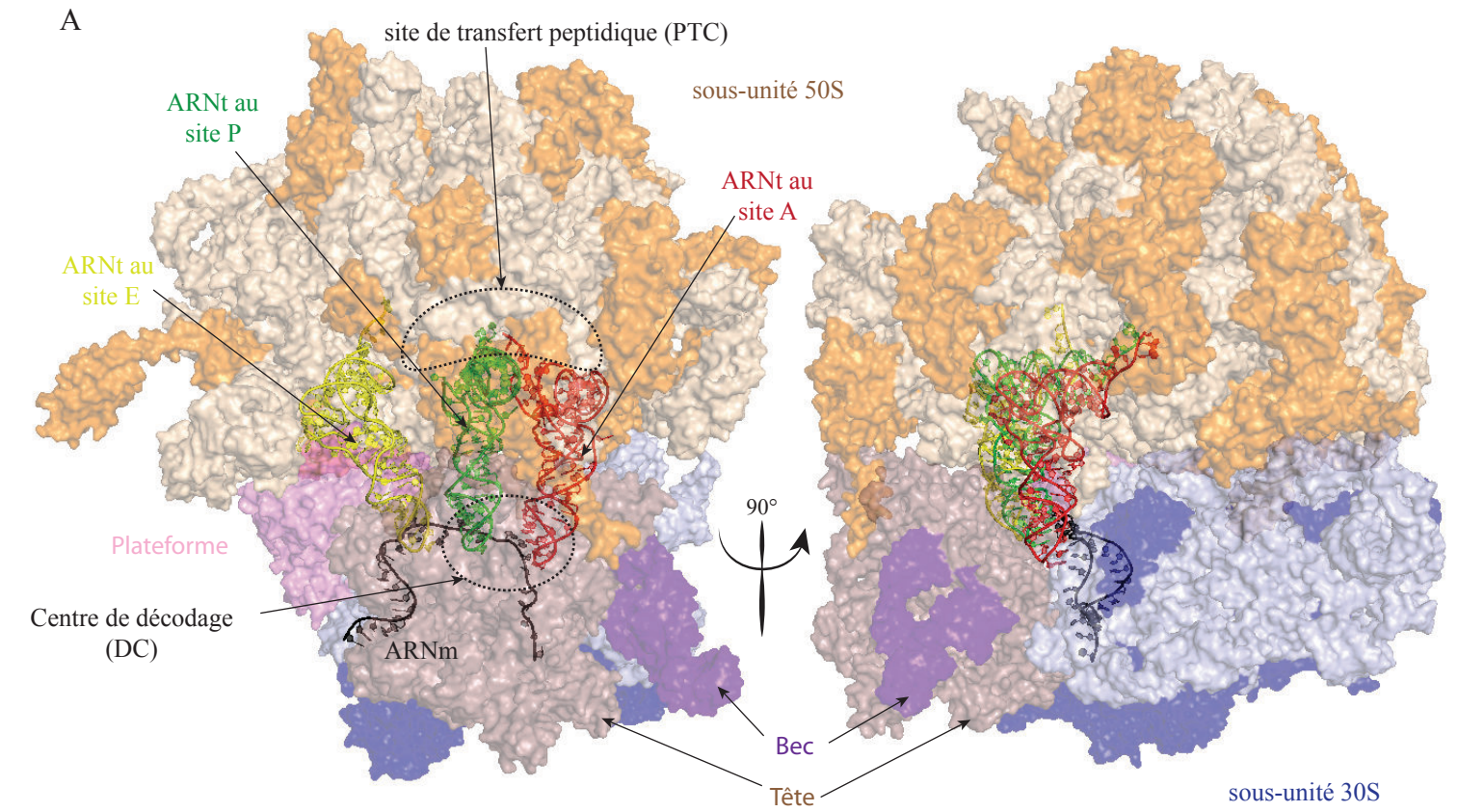


Figure 11. **Structure cristallographique à 3,1 Å du ribosome 70S de *Thermus thermophilus.*** L'ensemble des ARNr et des protéines ribosomiques des sous unités 50S et 30S sont représentés en (A) surface moléculaire et (B) mode ribbon. Les protéines sont représentées en orange et en bleu tandis que les ARNr sont représentés en orange clair et en bleu clair. La tête, le bec et la plateforme de la sous unité 30S sont respectivement indiqués en brun, violet et rose sur le panneau A. Les ARNt qui viennent se positionner séquentiellement aux sites A, P et E au cours de l'élongation, ainsi que l'ARNm sont représentés en mode ribbon en rouge, vert, jaune et noir respectivement. D'après (Jenner *et al.*, 2010), code PDB 3I8F et 3I8G.

Auparavant, l'équipe de Ramakrishnan avait également déterminé la structure du ribosome de *T. thermophilus* à 3,5 Å de résolution contenant un ARNm et un ARNt à chacun des sites A, P et E (Voorhees *et al.*, 2009). A la différence de celle obtenue en 2010, cette dernière possède des ARNt aminoacylés aux sites A et P. La forte homologie entre les ARNr de levure et d'archaea ainsi que la présence des ARNt au sein du ribosome dans les structures cristallines de *T. thermophilus* justifient l'utilisation de ces dernières dans les figures illustrant les positions des résidus modifiés au sein du ribosome.

L'ARNm se positionne dans la crevasse entre le corps et la tête de la petite sous-unité, où les codons s'apparient avec les anticodons des ARNt au sein du centre de décodage. Les ARNt se positionnent dans 3 sites distincts au sein de la grande sous-unité (Figure 11), ceci de façon séquentielle et ordonnée au cours de la traduction, comme nous venons de le voir. Il s'agit du site A, où l'ARNt aminoacylé « entre » dans le ribosome, du site P qui contient l'ARNt lié à la chaîne polypeptidique naissante et le site E ou *Exit* contenant l'ARNt déacylé provenant du site P après formation de la liaison peptidique et avant sa sortie du ribosome.

Au sein de la grande sous-unité, les extrémités 3' des ARNt présents aux sites A et P sont situées au sein du centre portant l'activité peptidyl transférase (*Peptidyl Transferase Center* ou PTC), alors que l'extrémité 3' de l'ARNt du site E est à environ 50 Å du PTC (Figure 11). C'est au niveau du PTC que le nouvel acide aminé est additionné à l'extrémité C-terminale du polypeptide en cours de synthèse.

1.3.3 Positionnement des résidus modifiés au sein du ribosome

Il existe une prédominance du type de modification en fonction du site fonctionnel ciblé : les pseudouridines (Ψ) sont principalement concentrées sur 2 zones : la tête et le bec, ainsi que le bas du corps de la petite sous-unité du ribosome (Figures 6 et 11). Les 2'-O-méthylation (Nm) chez *E. coli* et la levure sont concentrées à l'interface des deux sous-unités (Figures 6,7 et 11).

1.3.3.1 Rôle des modifications post-transcriptionnelles au niveau de la petite sous-unité du ribosome

Le centre de décodage inclue la zone de la reconnaissance codon-anticodon des sites A et P (Figure 11). Ce domaine fonctionnel du ribosome comprend 8 modifications guidées par des snoRNA chez la levure *S. cerevisiae* : 5 Nm et 3 Ψ aux positions 578, 999, 1007, 1187, 1191, 1271, 1428 et 1639 catalysées par les snoRNP 77, 31, 79, 36, 35, 40, 56 et 70

respectivement (Tableau 1, figures 6 et 12). La position 1191 est particulière car elle subit d'abord une pseudouridylation catalysée par la snoRNP 35, puis un groupement méthyl est ajouté en position 1 par l'enzyme Nep1p et enfin un groupement 3-amino-3-carboxypropyle y est greffé dans le cytoplasme pour conduire à la formation du résidu hyper-modifié $m^1acp^3\Psi$1191 (Brand *et al.*, 1978; Liang *et al.*, 2009b; Wurm *et al.*, 2010). Les résidus Um578 et Gm1271 se situent au niveau de la zone d'entrée de l'ARNt au site A tandis que les résidus Ψ1187 et Gm1428 se situent autour de l'ARNt au site A (site « *aroundA* » ou aA). Les résidus $m^1acp^3\Psi$1191 et Cm1639 sont localisés au site P et les nucléotides Ψ999 et Cm1007 au niveau du site E. L'inhibition de la synthèse de ces modifications de façon individuelle ou cumulative provoque différents effets, comme une altération de la terminaison au codon stop et de la maintenance du cadre de lecture, avec un effet plus prononcé pour les résidus du site P et aA que ceux du site A (Baudin-Baillieu *et al.*, 2009). Elle provoque également un défaut de croissance, un défaut de traduction et un déséquilibre du ratio 40S/60S (Liang *et al.*, 2009b). Il est intéressant de noter que le résidu Cm1639 est localisé à proximité de la 3^{ième} base du codon de l'ARNm au site P, qui constitue une position cruciale pour la fidélité de la traduction. La modification de cette position est conservée d'*E.coli* (m^4Cm1402) à l'homme (Cm1703), suggérant un apport bénéfique de cette modification.

En plus de ces modifications guidées par des snoRNA, les résidus A1781 et A1782 de l'ARN 18S subissent une méthylation par la diméthylase Dim1p chez *S.cerevisiae*. Ces résidus sont les équivalents des résidus 1492 et 1493 chez *T.thermophilus* ou 1518 et 1519 chez *E.coli* (modifié par l'enzyme Ksga, (O'Farrell *et al.*, 2006; Xu *et al.*, 2008) (http://people.biochem.umass.edu/fournierlab/3dmodmap/eqthermssu.php), qui se situent à proximité du centre de décodage et à l'interface des 2 sous-unités du ribosome.

1.3.3.2 Rôle des modifications post-transcriptionnelles au niveau de la grande sous-unité du ribosome

Au niveau de la grande sous-unité, presque toutes les modifications chez les bactéries et la levure sont contenues dans les domaines II, IV et V (Branlant *et al.*, 1981; Veldman *et al.*, 1981) (Figures 7 et 13). Le domaine V qui est entouré des domaines II et IV dans l'espace se situe au centre de l'interface des sous-unités et englobe le PTC (Ban *et al.*, 2000; Polacek & Mankin, 2005). Le domaine V de la levure contient plus de modifications que son homologue bactérien (Veldman *et al.*, 1981). Ces modifications de la grande sous-unité de levure se situent à proximité de son centre et forment une coquille autour des sites A et P des

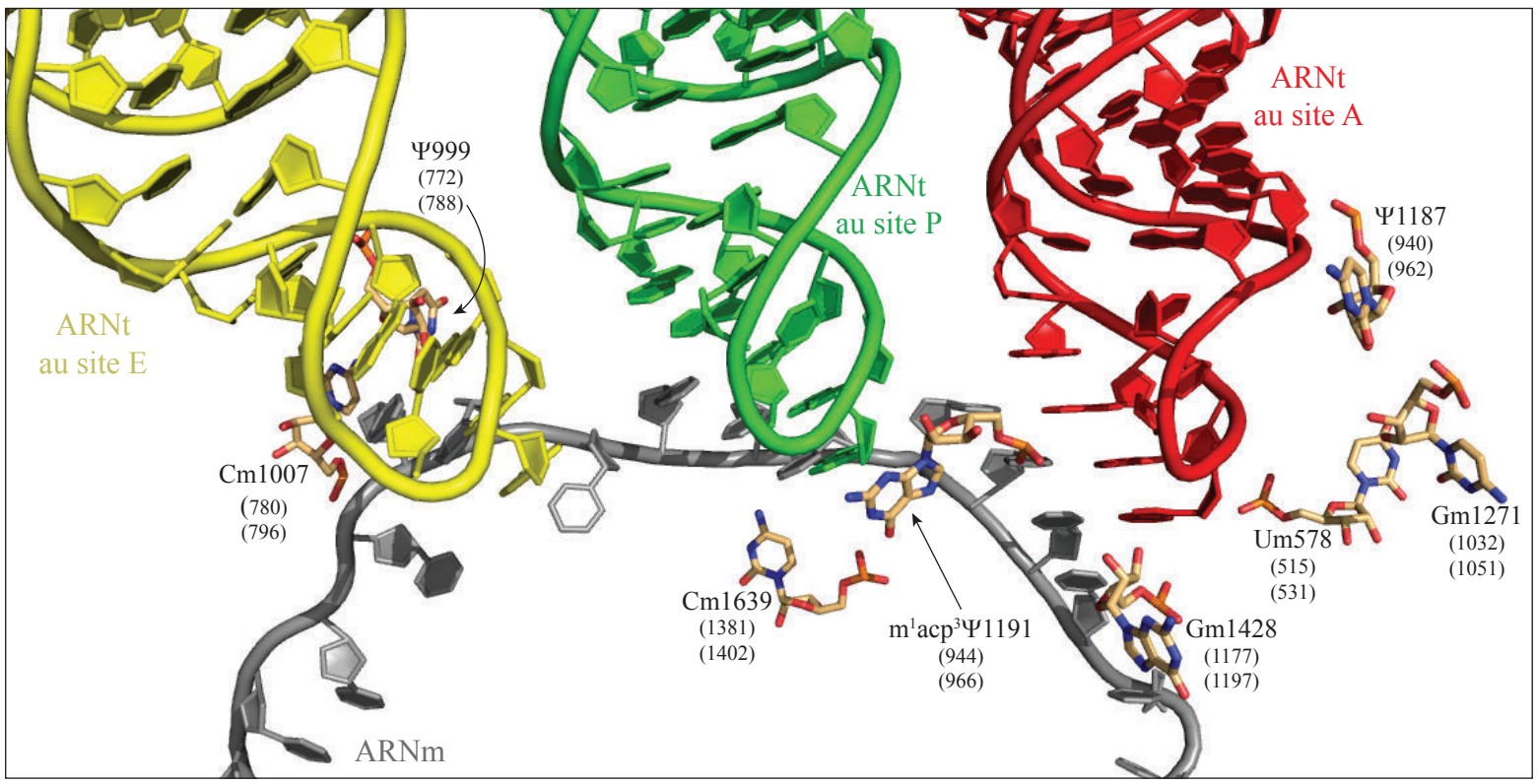


Figure 12. **Les résidus modifiés du centre de décodage de la petite sous unité du ribosome.** Les résidus modifiés sont représentés sous forme bâton. D'après (Jenner *et al.*, 2010), code PDB 3I8F et 3I8G. La structure cristallographique correspond au 70S de *Thermus thermophilus*, la numérotation des résidus modifiés correspond à celle de *Saccaromyces cerevisiae* avec entre parenthèse leur équivalent chez *T. thermophilus* et *E. coli*.

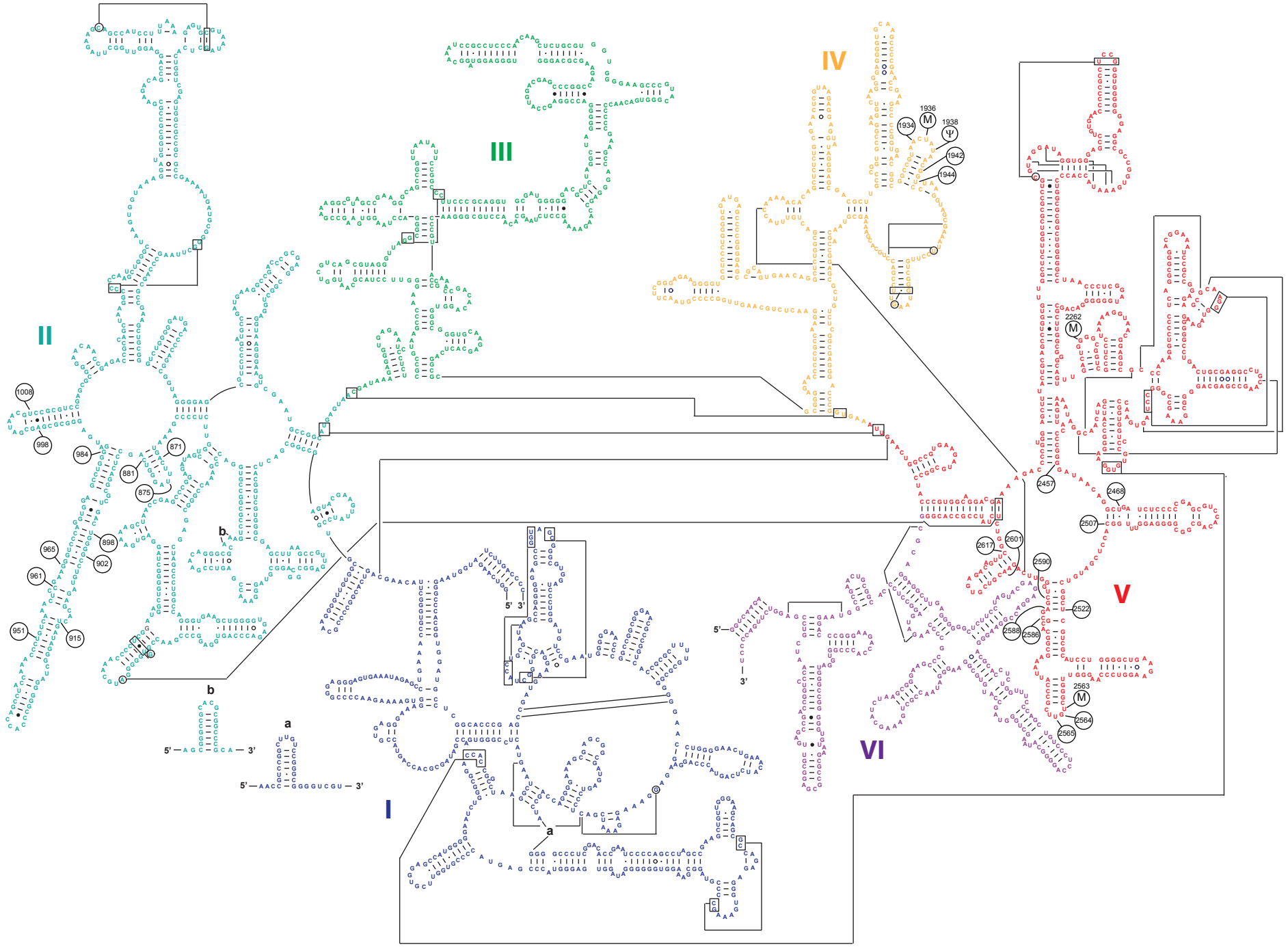


Figure 13. **Cartographie des positions équivalentes aux résidus modifiés de la grande sous-unité** de *S. cerevisiae* sur la structure secondaire de l'ARNr 23S de la grande sous-unité du ribosome de *T. thermophilus*. Les positions en noir correspondent à celles discutées dans le paragraphe 1.3.3.2. Les chiffres romains correspondent aux domaines de la grande sous-unité du ribosome. D'après (Baudin-Baillieu *et al.*, 2009), http://people.biochem.umass.edu/sfournier/fournierlab/snornadb/ et (Cannone *et al.*, 2002), http://www.rna.ccbb.utexas.edu/.

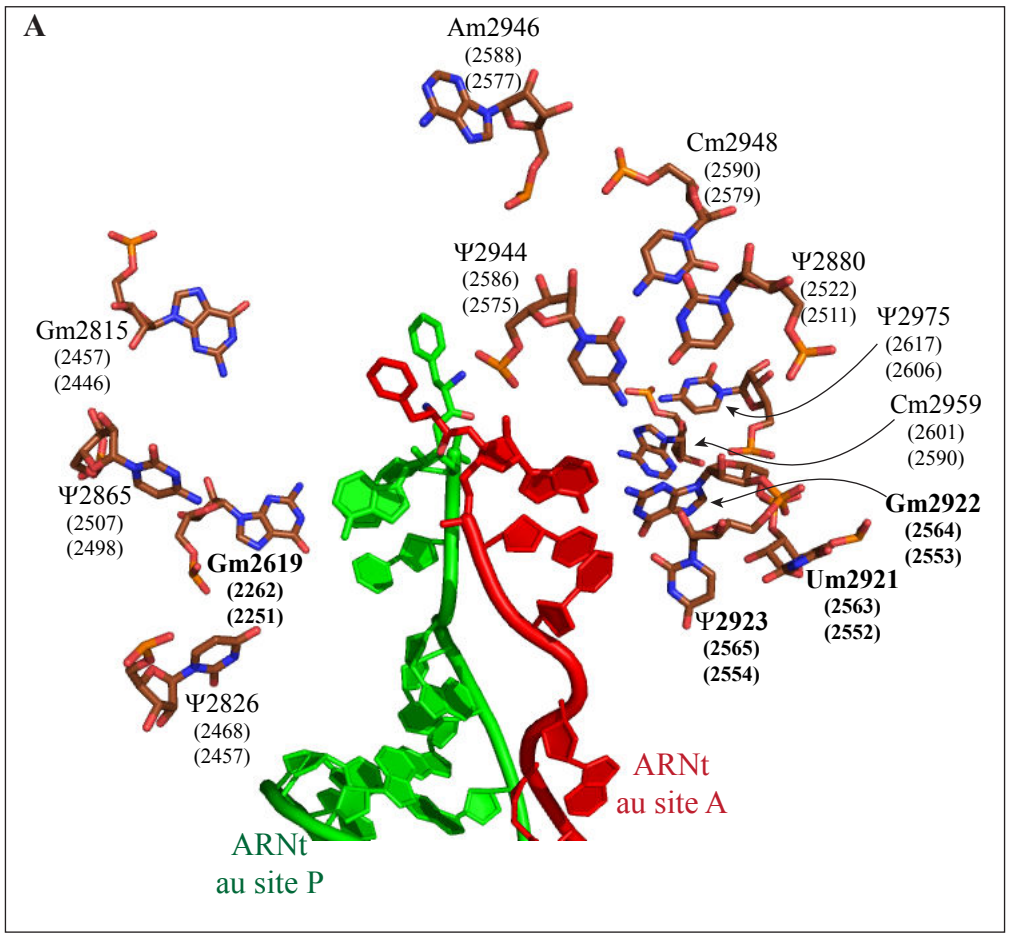
ARNt. Cette région englobe aussi l'entrée et le premier tiers du tunnel de sortie du polypeptide (Decatur & Fournier, 2002). Au sein de la grande sous-unité, il existe donc une très forte concentration en nucléotides modifiés au niveau des sites A, P et E, ainsi que dans les régions situées à proximité du centre réactionnel, du tunnel de sortie du polypeptide et des zones d'interaction établissant le pont entre les deux sous-unités.

1.3.3.2.1 Le centre PTC

Le PTC est le cœur catalytique du ribosome car il catalyse les deux principales réactions chimiques de la synthèse peptidique : la formation de la liaison peptidique et le relargage du peptide. Les structures cristallographiques confirment que le ribosome est une enzyme ARN ou ribozyme. Le PTC est riche en résidus 2'-O-méthylés et en pseudouridines (Figure 7). Le centre PTC contient six résidus Ψ, sept 2'-O-méthylations (Nm) et une méthylation de base (mN). La formation de ces résidus Ψ est guidée par 5 snoRNA à boîtes H/ACA : snR10 (à la position 2923), snR34 (aux positions 2826 et 2880), snR37 (à la position 2944), snR42 (à la position 2975) et snR46 (à la position 2865) (Tableau 1). Les 2'-O-méthylations formées aux positions 2619, 2815, 2921, 2922, 2946, 2946, 2948 et 2959 impliquent respectivement les snoRNA snR67, snR38, snR52, l'enzyme Spb1p, et les snoRNA snR71, snR69, et snR73 (Figure 14). La perte des résidus Ψ au sein du centre PTC inhibe l'activité de traduction du ribosome, notamment la perte de Ψ2919 (alias Ψ 2923 incorporé par la snR10) de la boucle A du centre PTC (Figure 14). L'inhibition de la formation de cette modification perturbe aussi la formation des polysomes (King *et al.*, 2003).

Il est intéressant de noter que le snoRNA snR10 influence aussi la production de l'ARNr 18S (Tollervey, 1987). L'effet de croissance observé dans une souche invalidée pour le gène codant le snoRNA snR10 (Δ*SNR10*) reflète donc à la fois l'effet sur la traduction et sur la synthèse de l'ARNr 18S. L'équipe de Fournier (King *et al.*, 2003) a également montré que la structuration de la grande sous-unité était perturbée en l'absence des 6 Ψ du centre PTC. Il est intéressant de noter que la 2'-O-méthylation du résidu U2921 est conservée entre la bactérie et l'homme.

Le résidu Gm2922 chez *S.cerevisiae* (l'équivalent des résidus G2564 chez *T.thermophilus* et G2553 chez *E.coli*) forme une paire de base Watson-Crick avec le nucléotide C75 de l'ARNt au site A, ce qui a pour effet de stabiliser l'extrémité 3' de l'ARNt (le motif $C^{74}C^{75}A^{76}$ qui porte l'acide aminé) (Voorhees *et al.*, 2009) (Figure 14). De façon symétrique, le nucléotide C75 de l'ARNt au site P établit une paire de base Watson-Crick



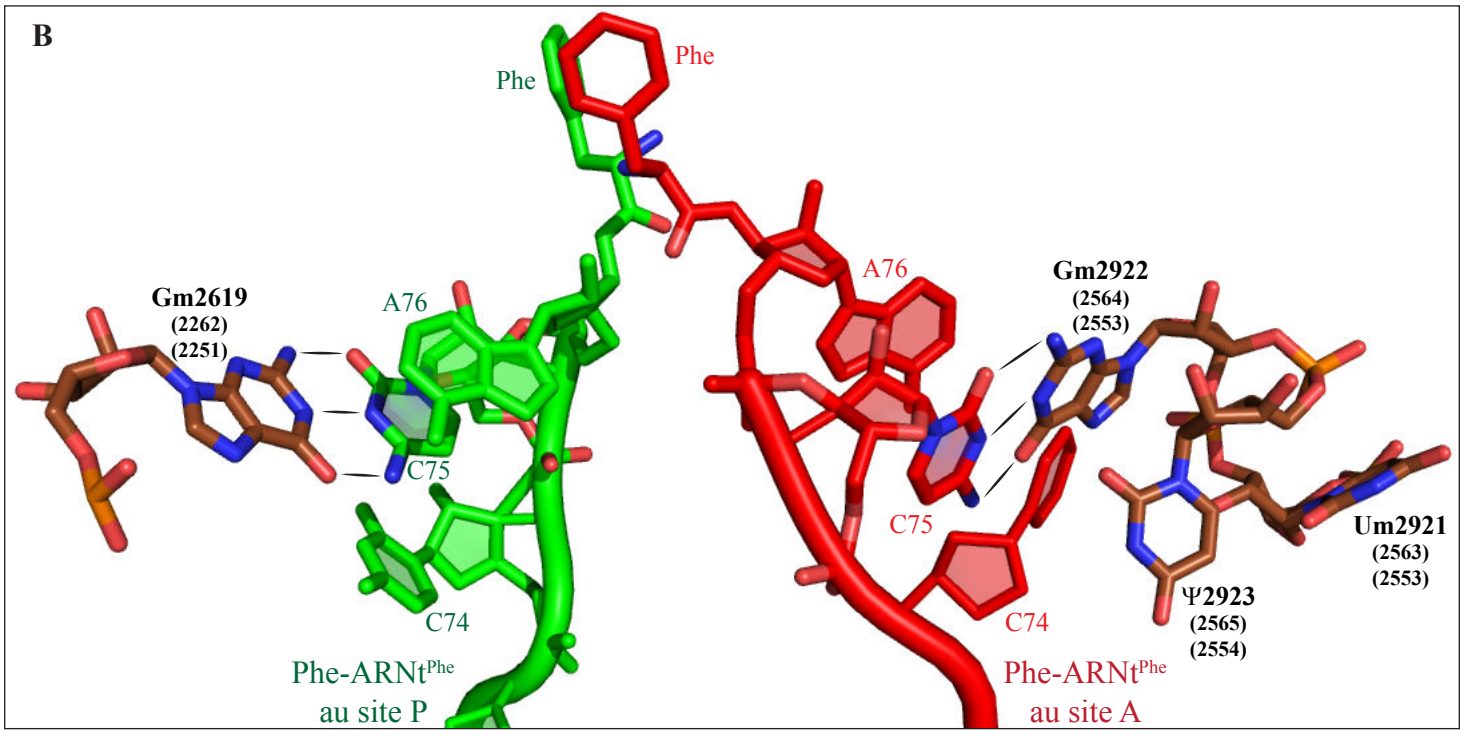
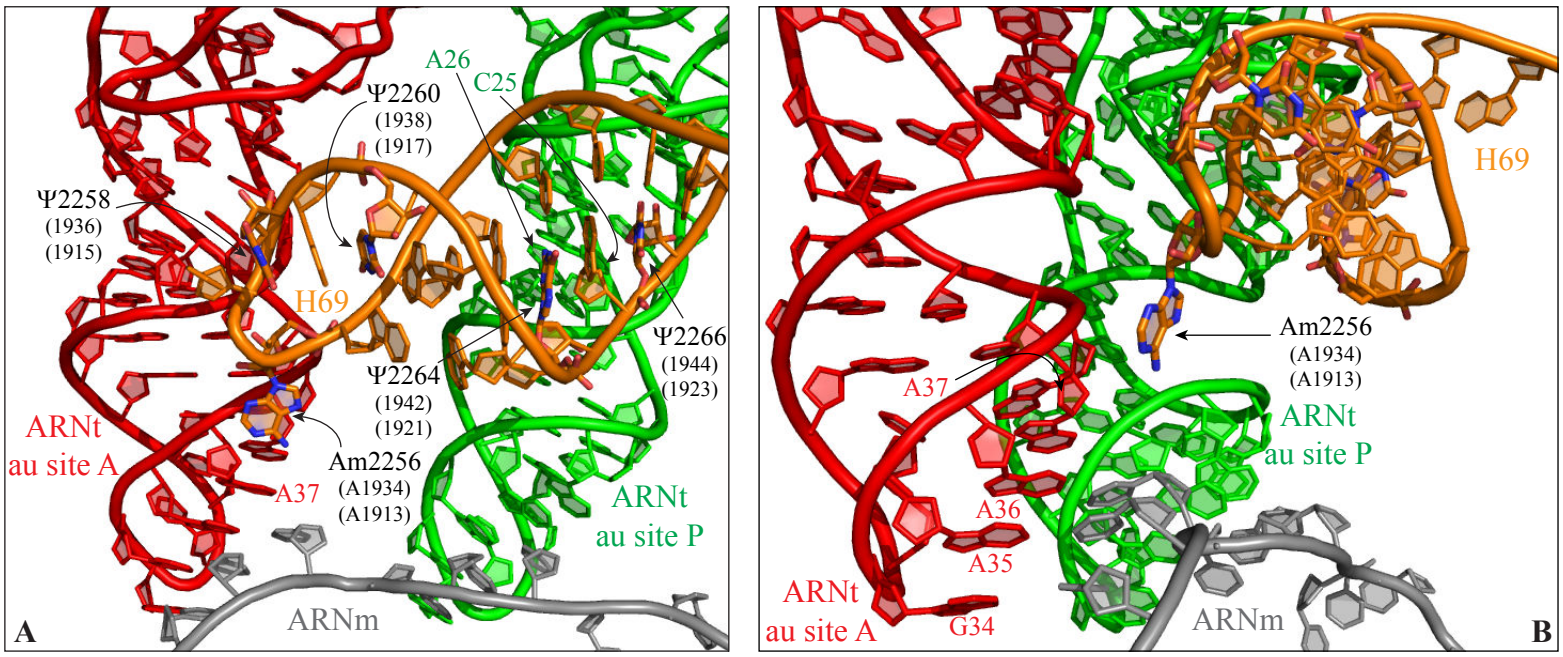


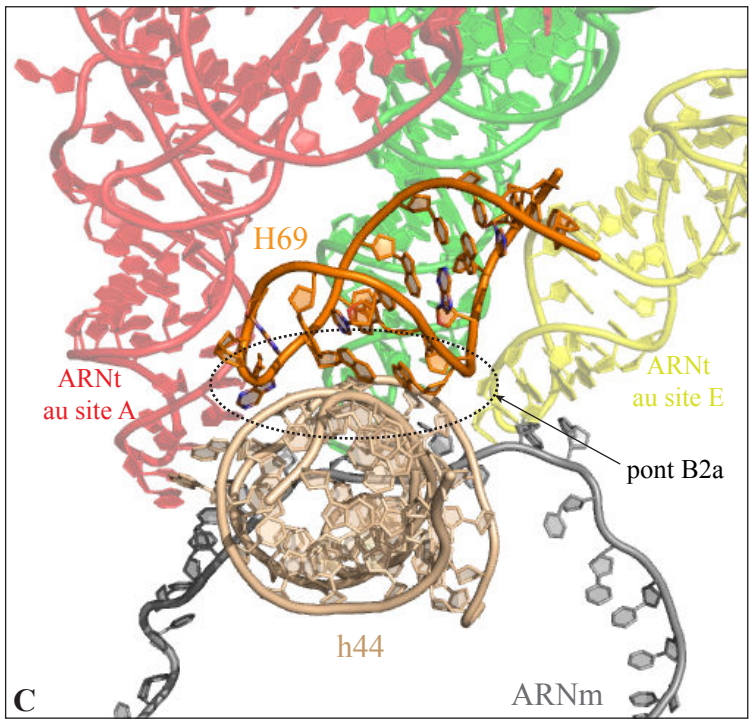
Figure 14. **Vues du centre peptidyl transférase au stade du pré-transfert peptidique.** (A) L'ensemble des résidus modifiés (Ψ ou Nm) du site PTC sont représentés en mode bâton. (B) Interaction des résidus modifiés Gm2619 et Gm2922 de l'ARNr 25S de la grande sous unité avec les nucléotides C75 des ARNt aminoacylés (Phe-ARNt^{Phe}) aux sites P et A, respectivement. D'après (Voorhees *et al.*, 2009), code PDB 2WDL et 2WDK. La structure cristallographique correspond au 70S de *Thermus thermophilus*, la numérotation des résidus modifiés correspond à celle de *Saccaromyces cerevisiae* avec entre parenthèse leurs équivalents chez *T. thermophilus* et *E. coli*. Les traits indiquent les liaisons hydrogène.

avec le résidu Gm2619 (Gm2262 chez *T.thermophilus* et Gm2251 chez *E.coli*). Un examen de la région du site actif montre, que chez *E. coli* ou chez la levure, aucune modification chimique ne se situe à moins de 10 Å du carbone tétraédrique, ce qui exclut une participation directe dans la réaction de formation de la liaison peptidique (Decatur & Fournier, 2002). Mais il est tout de même possible que les modifications environnantes puissent influencer la réaction de manière indirecte.

1.3.3.2.2 L'hélice 69

En 2007, l'équipe de M. Fournier a étudié l'effet des modifications de l'hélice 69 (H69) du domaine IV de la grande sous-unité du ribosome. Ce domaine est universellement riche en nucléotides modifiés car il possède trois Ψ chez *E. coli*, quatre Ψ et une Nm chez la levure (Figure 7), ainsi que cinq Ψ et une Nm chez l'homme (Liang *et al.*, 2007b). La position de plusieurs de ces modifications est conservée (Figure 15). Les données de cristallographie montrent que H69 établit des interactions avec l'hélice 44 (h44) de la petite sous-unité pour former le pont inter sous-unité B2a (Yusupov *et al.*, 2001) (Figure 15C et D). Dans le contexte du ribosome complet, H69 interagit également avec la tige D des ARNt aux sites A et P, suggérant un rôle dans la translocation des ARNt et la fidélité de la traduction (Gao *et al.*, 2003; Moazed & Noller, 1989; Yusupov *et al.*, 2001) (Figure 15). En effet, l'implication de l'hélice 69 dans la fidélité de traduction chez *E.coli* a été montrée par des mutations réalisées dans la région de la boucle (U1915A) ou par le blocage de la formation des modifications Ψ1911, Ψ1915 et Ψ1917, qui provoquent, entre autre, un décalage du cadre de lecture +1, la translecture du codon stop et un défaut de la terminaison de la traduction (Ejby *et al.*, 2007; Hirabayashi *et al.*, 2006). Chez la levure, le domaine H69 de 19 nucléotides porte 5 modifications (Am2256, Ψ2258 (Ψ1915 chez *E.coli*), Ψ2260 (Ψ1917 chez *E.coli*), Ψ2264 et Ψ2266) (Figure 7). Les modifications aux positions 2258 et 2260 sont conservées entre la bactérie et l'homme. La perte de ces modifications provoque un défaut de croissance, une augmentation de la sensibilité à certains antibiotiques, un défaut de traduction, une modification de la structure du ribosome et une diminution du taux d'ARNr dans les cellules (Liang *et al.*, 2007b). De façon intéressante, elle a également pour effet d'augmenter la fidélité de la traduction et la résistance aux antibiotiques aminoglycosides (G418 et kanamycine A) connus pour se lier aux hélices h44 et H69 chez *S. cerevisiae*. Ceci suggère donc une corrélation entre la fidélité de traduction et des changements structuraux du ribosome (Baudin-Baillieu *et al.*, 2009). Il est intéressant de noter que la position du résidu Ψ2266, capable de contacter les nucléotides 25 et 26 de l'ARNt au site P (Yusupov *et al.*,





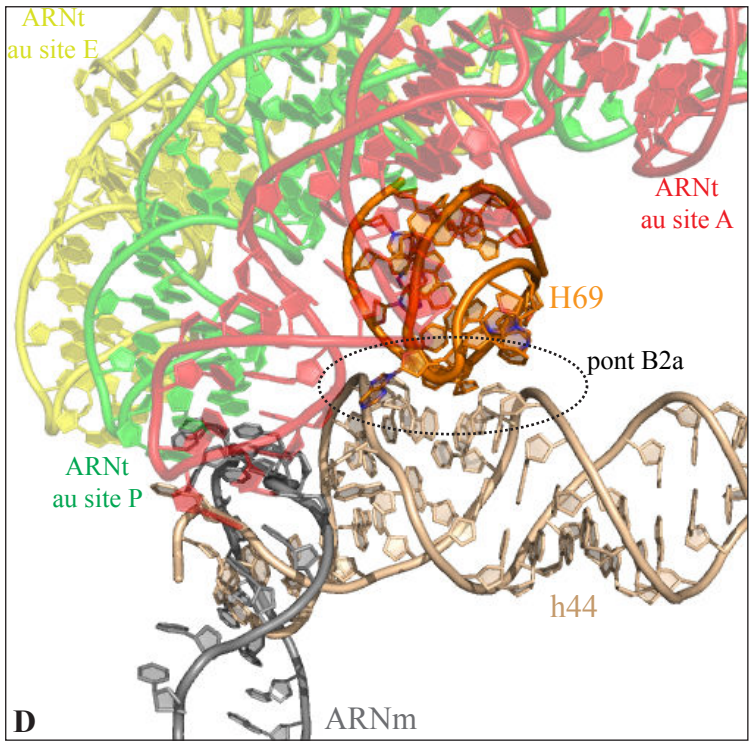


Figure 15. **Position des résidus modifiés de l'hélice 69 de la grande sous-unité 70S dans la structure cristallographique du ribosome de *Thermus thermophilus***. (A) et (B) Proximité des résidus modifiés des ARNt aux sites A et P, notamment le résidu à la position 2256 proche du résidu hyper modifié A37 en position 3' de la boucle anticodon. Pont inter sous-unité B2a qui s'établit entre les hélices 69 de la grande sous-unité et 44 de la petite sous-unité du ribosome vue de face (C) et de côté (D). D'après (Jenner *et al.*, 2010), code PDB 3I8F et 3I8G. La numérotation des résidus modifiés correspond à celle de *Saccaromyces cerevisiae* avec entre parenthèse leur équivalent chez *T. thermophilus* et *E. coli*.

2001), exerce une influence stabilisatrice sur l'interaction entre l'ARNt et le ribosome. Au sein du site A, le résidu Am2256 est situé à proximité de la base 37 hyper-modifiée en 3' de la boucle anti codon de l'ARNt, connue pour jouer un rôle crucial dans le décodage par l'ARNt (Baudin-Baillieu *et al.*, 2009).

1.3.3.2.3 « A-site finger »

Le doigt du site A (*« A-site finger »* ou ASF), conservé de la bactérie à l'Homme, est formé par l'hélice 38 appartenant au domaine II de l'ARNr 25S (Figure 7). L'ASF participe au pont inter sous-unité B1a par sa boucle terminale, établit des contacts avec les ARNt au site A et avec l'ARNr 5S, mais également avec la protéine S15 chez *S. cerevisiae* (S13 chez *E.coli*). Chez la levure, l'hélice 38 contient sept residus Ψ et il en existe également trois autres dans des petites hélices la bordant (H37 et H39) (Figures 7 et 16). Il est intéressant de noter l'absence de modifications dans le *« A-site finger »* bactérien alors que celles-ci sont conservées chez les mammifères tout comme chez les levures. Parmi les positions les plus étudiées, l'inhibition de la formation des modifications Ψ960, 986, 1004, 1042, 1052 et 1124 (équivalente aux positions 828, 851, 868, 905, 915 et 953 chez *E.coli*) n'a pas d'effet marqué sur la fidélité de la traduction (Baudin-Baillieu *et al.*, 2009), alors que ces mutations provoquent un défaut de croissance, un défaut de traduction, une sensibilité accrue aux antibiotiques ciblant le ribosome et un profil de polysome anormal (Piekna-Przybylska *et al.*, 2008).

La carte des sites des modifications sur les ARNr et l'identification des snoRNA guides correspondant a été réalisé chez *Saccharomyces cerevisiae* et sont disponible à l'adresse suivante : http://people.biochem.umass.edu/sfournier/fournierlab/snornadb/ (Piekna-Przybylska *et al.*, 2007).

1.4 Mise en place des modifications par les sRNP/snoRNP

Chez les eucaryotes, les snoRNP impliquées dans les clivages endonucléolytiques et les modifications post- ou co-transcriptionnelles du pré-ARNr 35S sont composées de 4 protéines constitutives, dont une enzyme (la 2'-O-méthylase ou la pseudouridyl-synthase) et d'un ARN guide : le snoRNA. Ces différents constituants s'assemblent sous la forme d'une particule unique. Le snoRNA a pour rôle de sélectionner l'ARNr substrat en s'appariant avec celui-ci, ce qui permet le ciblage précis du résidu à modifier. La composition précise et la biogenèse de ces particules fera l'objet d'une description détaillée dans les chapitres suivants.

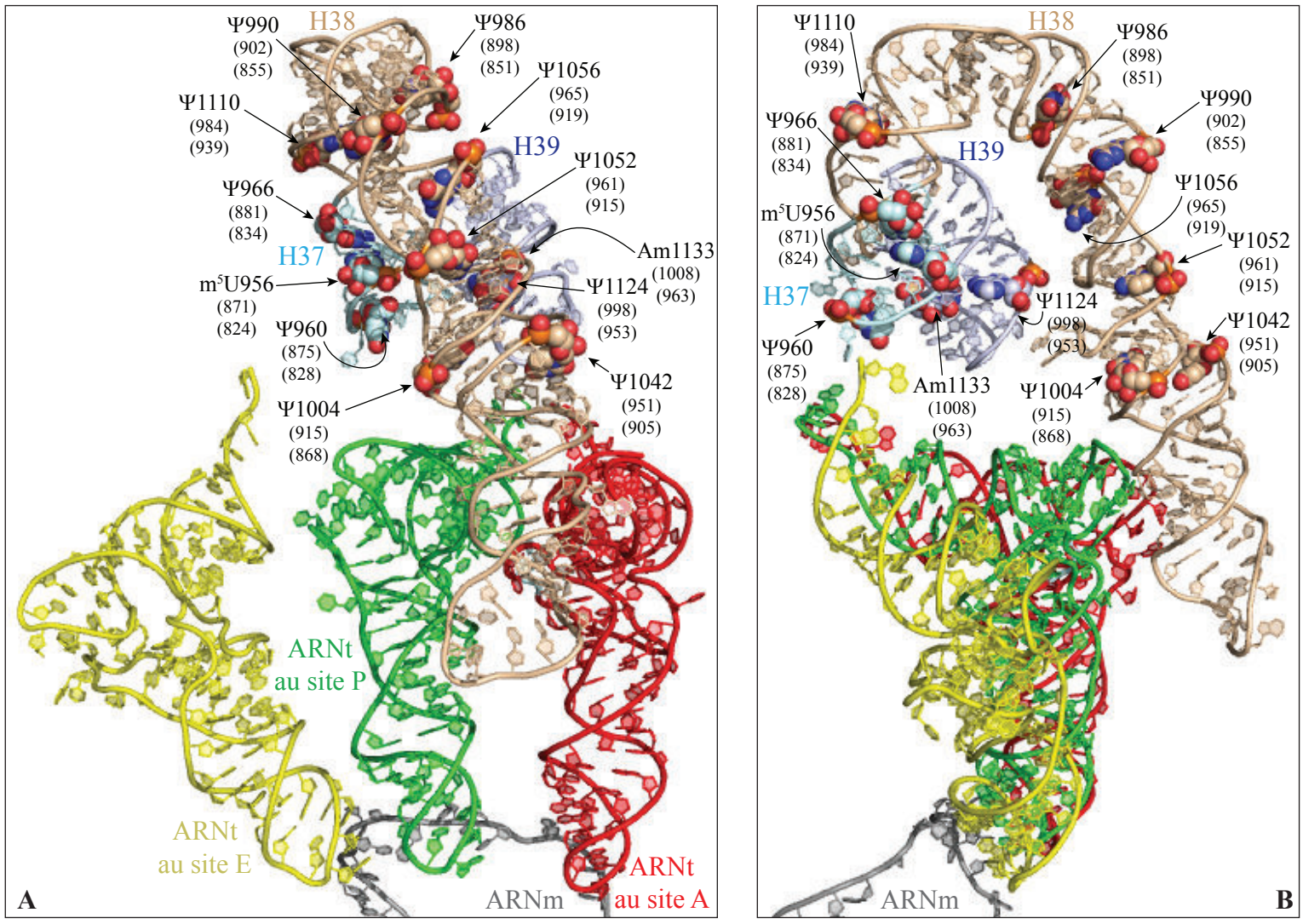


Figure 16. **Les résidus modifiés du *«A-site finger»*** et des hélices adjacentes H37 et H39 au sein de la structure cristallographique de la sous-unité 70S de *Thermus thermophilus*. (A) Vue de face et (B) de côté des hélices H37, H38 et H39. Pour des raisons de clarté, les résidus modifiés sont représentés en sphère. D'après (Jenner *et al.*, 2010), code PDB 3I8F et 3I8G. La numérotation des résidus modifiés correspond à celle de *S. cerevisiae* avec entre parenthèse leur équivalent chez *T. thermophilus* et *E. coli*.

Il est intéressant de noter que l'absence des enzymes comme Nop1p (2'-O-méthylase) (Tollervey *et al.*, 1991) ou Cbf5p (pseudouridyl-synthase) (Lafontaine *et al.*, 1998), ainsi que d'autres protéines constitutives comme Gar1p (Bousquet-Antonelli *et al.*, 1997; Girard *et al.*, 1992), ou encore de certains snoRNA snR30 (Morrissey & Tollervey, 1993), U3 (Hughes & Ares, 1991) ou U14 (Li *et al.*, 1990) inhibe la maturation des ARNr.

La production de l'ARNr 18S est principalement affectée et se matérialise par l'accumulation d'un intermédiaire aberrant, l'ARN 23S, qui résulte d'un défaut de coupure en A_0 et A_1. Par exemple, certaines mutations du gène codant la méthyltransférase Nop1p provoquent une inhibition de synthèse des ARNr 18S et 25S à la température non permissive (Tollervey *et al.*, 1993).

2 Les RNP à boîtes H/ACA

Cette machinerie RNP se compose d'un ARN guide, le s/snoRNA mais aussi de plusieurs composants protéiques dénommés respectivement chez les archaea, la levure et l'homme : L7Ae/Nhp2p/NHP2, essentiel à la stabilité de l'ARN et à l'activité optimum de la particule (Duan *et al.*, 2009; Henras *et al.*, 1998; Li *et al.*, 2011c; Watkins *et al.*, 1998a), aGar1/Gar1p/GAR1, aNop10/Nop10p/NOP10 et l'enzyme aCbf5/Cbf5p/CBF5 qui porte l'activité ARN : pseudouridine synthase (Tableau 2).

2.1 Structure des ARN à boîtes H/ACA

Responsables de la formation des Ψ (Ganot *et al.*, 1997a; Ni *et al.*, 1997) et impliquées dans la maturation nucléolytique du pré-ARNr comme nous venons de le voir, les snoRNP à boîtes H/ACA sont définies par un composant ARN unique, le snoRNA qui tire son nom de la présence de séquences conservées: la boîte « H » ou *« Hinge »* de séquence 5'-ANANNA-3' (Ganot *et al.*, 1997b) et la boîte ACA, formée par le tri-nucléotide 5'-ACA-3' (Balakin *et al.*, 1996) (Figure 17). Cet ARN adopte une structure secondaire particulière qui consiste en la succession de deux structures tiges-boucles séparées par une séquence en simple brin contenant la boîte H. La boîte ACA se situe à exactement 3 nucléotides en amont de l'extrémité 3' du snoRNA mature (Figure 17). Les deux structures tiges boucles présentent une boucle interne dont chaque brin comporte de 9 à 13 nucléotides constituant la poche de pseudouridylation. Ces séquences en simple brin constituent la séquence guide et sont spécifiques de chaque snoRNA. Elles ont pour fonction de cibler un ARN substrat précis.

A

	Homme	Levure	Archae
Protéines consitutives des particules H/ACA	CBF5	Cbf5p	aCbf5
	NOP10	Nop10p	aNop10
	GAR1	Gar1p	aGar1
	NHP2	Nhp2p	L7Ae
ARN guide	snoRNA	snoRNA	sRNA

B

	Homme	Levure	Archae
Protéines consitutives des particules C/D	Fibrillarine	Nop1p	aNop10
	NOP56	Nop56p	aNop5
	NOP58	Nop58p	
	15,5K	Snu13p	L7Ae
ARN guide	snoRNA	snoRNA	sRNA

Tableau 2. **Composition des RNP à boîtes H/ACA et C/D.** Chaque composant formant la particule (A) H/ACA et (B) C/D active est indiqué avec son homologue humain, de levure et d'archae.

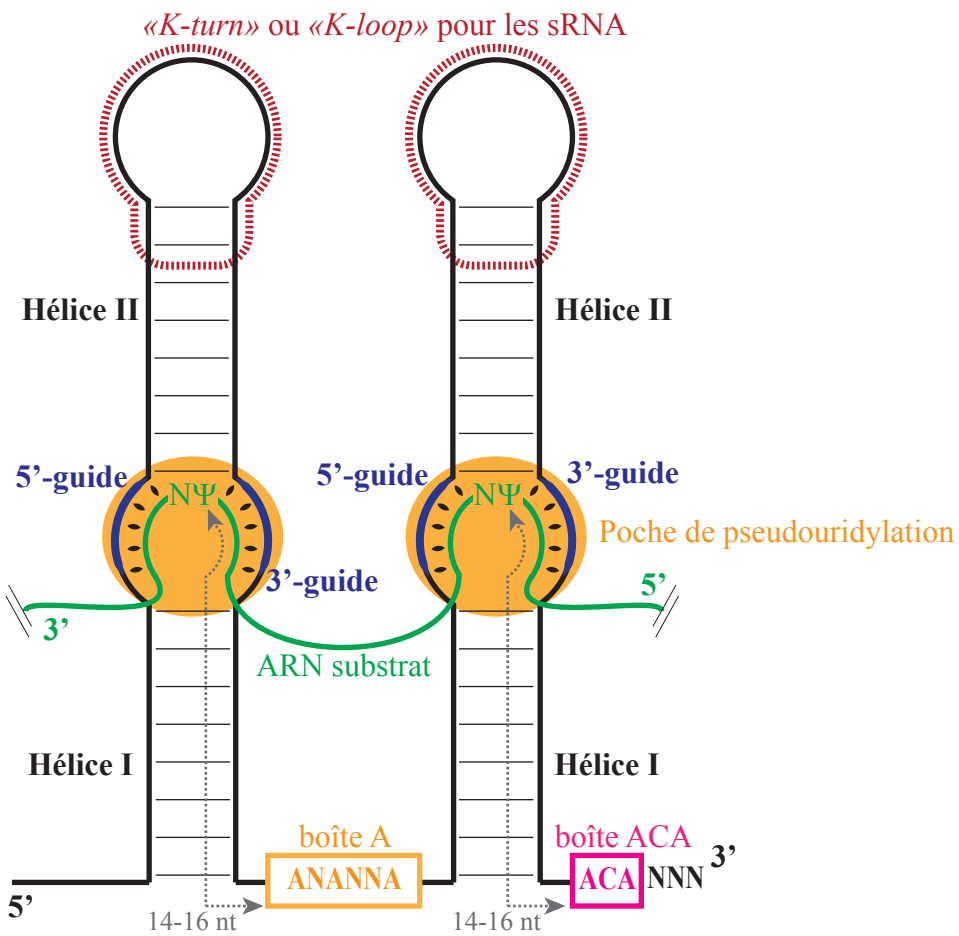


Figure 17. **Représentation schématique des s/snoRNA à boîtes H/ACA**. Au sein des sRNA d'archae, les extrémités apicales des tiges portent un motif en *« K-turn »* ou *« K-loop »*. La distance de 14-16 nucléotides entre le résidu Ψ de la poche de pseudouridylation et les boîtes A et ACA est conservée.

L'association de l'ARN guide avec son ARN substrat conduit à la formation de deux hélices intermoléculaires dans l'ARN substrat et à un dinucléotide 5'-UN-3' non apparié au sommet de la poche de l'ARN substrat (Ganot *et al.*, 1997a; Ni *et al.*, 1997) (Figure 17). C'est le résidu uridine en 5' de ce di-nucléotide qui est converti en pseudouridine. Ce résidu est toujours localisé 14 à 16 nucléotides en amont des boîtes H ou ACA. Ces éléments conservés sont nécessaires à la stabilité du snoRNA, à sa maturation, ainsi qu'à l'activité de pseudouridylation de la particule (Balakin *et al.*, 1996; Bortolin *et al.*, 1999; Ganot *et al.*, 1997b). (Pour revue (Bachellerie *et al.*, 2002; Reichow *et al.*, 2007))

A la différence des eucaryotes, les sRNA H/ACA peuvent être constitués de 1 à 3 structures en tige-boucle (Muller *et al.*, 2008). Ils se distinguent aussi par l'absence de la boîte H et par une boîte ACA de type 5'-ANA-3' située directement en aval de chaque tige-boucle. Structuralement, ils diffèrent par la présence d'un motif ARN particulier, un motif en *« K-turn »* ou *« K-loop »* (Rozhdestvensky *et al.*, 2003). Il est retrouvé à l'extrémité apicale de l'hélice II des sRNA, à une distance de 5 à 6 paires de bases de la poche de pseudouridylation (Muller *et al.*, 2008). Ce domaine fera l'objet d'une description plus détaillée au chapitre 3.2. Des travaux réalisés au laboratoire par S. Muller et J.B. Fourmann ont permis de mettre en évidence que la réaction de pseudouridylation peut avoir lieu même en présence d'un trinucléotide 5'-UNN-3' non apparié au sein de la poche de pseudouridylation (Muller *et al.*, 2008).

2.2 Structure des protéines composant les RNP à boîtes H/ACA

Ces RNP et ses constituants ont fait l'objet de nombreuses études structurales. La majorité des informations structurales ont été obtenues sur les protéines d'archaea. Ceci s'explique notamment par des protéines d'archaea plus stables et aussi plus petites que leurs homologues eucaryotes qui contiennent souvent des extensions composées de répétitions de résidus. Nous présenterons donc les données fonctionnelles obtenues principalement chez la levure, en nous appuyant sur les données structurales obtenues chez les archaea.

2.2.1 L'enzyme aCbf5/Cbf5p

Chez la levure, elle a été initialement identifiée comme interagissant avec les centromères, d'où son nom *Centromere Binding Factor 5*. Chez l'homme elle porte le nom de CBF5, NAP17 ou encore Dyskérine, en référence à son lien avec la dyskératose congénitale. Cbf5p est une protéine essentielle à la viabilité cellulaire (Jiang *et al.*, 1993; Watkins *et al.*,

1998a). Sa mutation conduit à un défaut d'activité de la snoRNP chez la levure (Zebarjadian *et al.*, 1999) et son rôle enzymatique a été confirmé *in vitro* grâce à l'utilisation de protéines d'archaea recombinantes (Charpentier *et al.*, 2005). Des expériences de déplétion réalisées chez la levure ont montré que la protéine Cbf5p est requise pour la fonction des particules H/ACA, mais également pour la stabilité des snoRNA (Lafontaine *et al.*, 1998). Tout comme les protéines des snoRNP à boîtes C/D Nop56p et Nop58p, qui feront l'objet d'une description au chapitre 3.3.2 ; Cbf5p possède une extrémité C-terminale riche en motifs KKE/D, qui interagissent avec l'hyperméthylase Tgs1p (Mouaikel *et al.*, 2002) (Figure 18). De façon intéressante, ces motifs sont absents chez les archaea et ne sont pas requis pour la fonctionnalité de l'enzyme chez les eucaryotes (Watkins *et al.*, 1998a; Winkler *et al.*, 1998).

Les données cristallographiques du complexe entre aCbf5 et aNop10 d'archaea (Manival *et al.*, 2006) ont confirmé l'homologie de séquence observée entre l'enzyme du système ribonucléoprotéique et le système purement protéique d'*E.coli*, l'ARNt Ψ synthase TruB, qui catalyse l'isomérisation de l'uridine 55 de la boucle T des ARNt en pseudouridine (Hamma *et al.*, 2005; Manival *et al.*, 2006) (Figures 18 et 19). Tout comme TruB, Cbf5 est composée de deux domaines : le domaine cœur catalytique, commun à toutes les ARN: Ψ synthases (Hamma & Ferré-D'Amaré, 2006) et le domaine PUA (« *PseudoUridine synthase and Archaeaosine transglycosylase* ») situé à l'extrémité C-terminale de la protéine (Figures 18 et 19).

Ce dernier est un motif de liaison à l'ARN retrouvé dans de nombreuses enzymes modifiant les nucléotides (Aravind & Koonin, 1999) et au sein de l'enzyme Cbf5 et ses homologues. Il assure la reconnaissance du motif conservé 3'-ACA-5' (3'-ANA-5' chez les archaea) des sno/s RNA à boîtes H/ACA (Hamma *et al.*, 2005; Li & Ye, 2006; Rashid *et al.*, 2006; Zhou *et al.*, 2011). Ce domaine est aussi impliqué dans la maladie génétique humaine « *dyskeratosis congenita* » (DKC) car plusieurs mutations identifiées chez certains patients sont concentrées dans le domaine PUA de la dyskérine (Heiss *et al.*, 1998). Le domaine PUA adopte un repliement mixte α/β (Figure 19). La comparaison des domaines PUA des protéines TruB d'*E.coli* et aCbf5 révèle la présence d'une séquence N-terminale additionnelle d'une trentaine de résidus (Figure 18) qui se positionne autour du domaine PUA (Figure 19). Cette séquence additionnelle est caractéristique des protéines aCbf5 d'archaea mais est également présente dans les protéines CBF5 eucaryotes. La délétion de ce segment N-terminal induit une forte instabilité de la protéine qui a conduit notre équipe à proposer qu'il était important pour le repliement et/ou la stabilité de l'enzyme (Manival *et al.*, 2006).

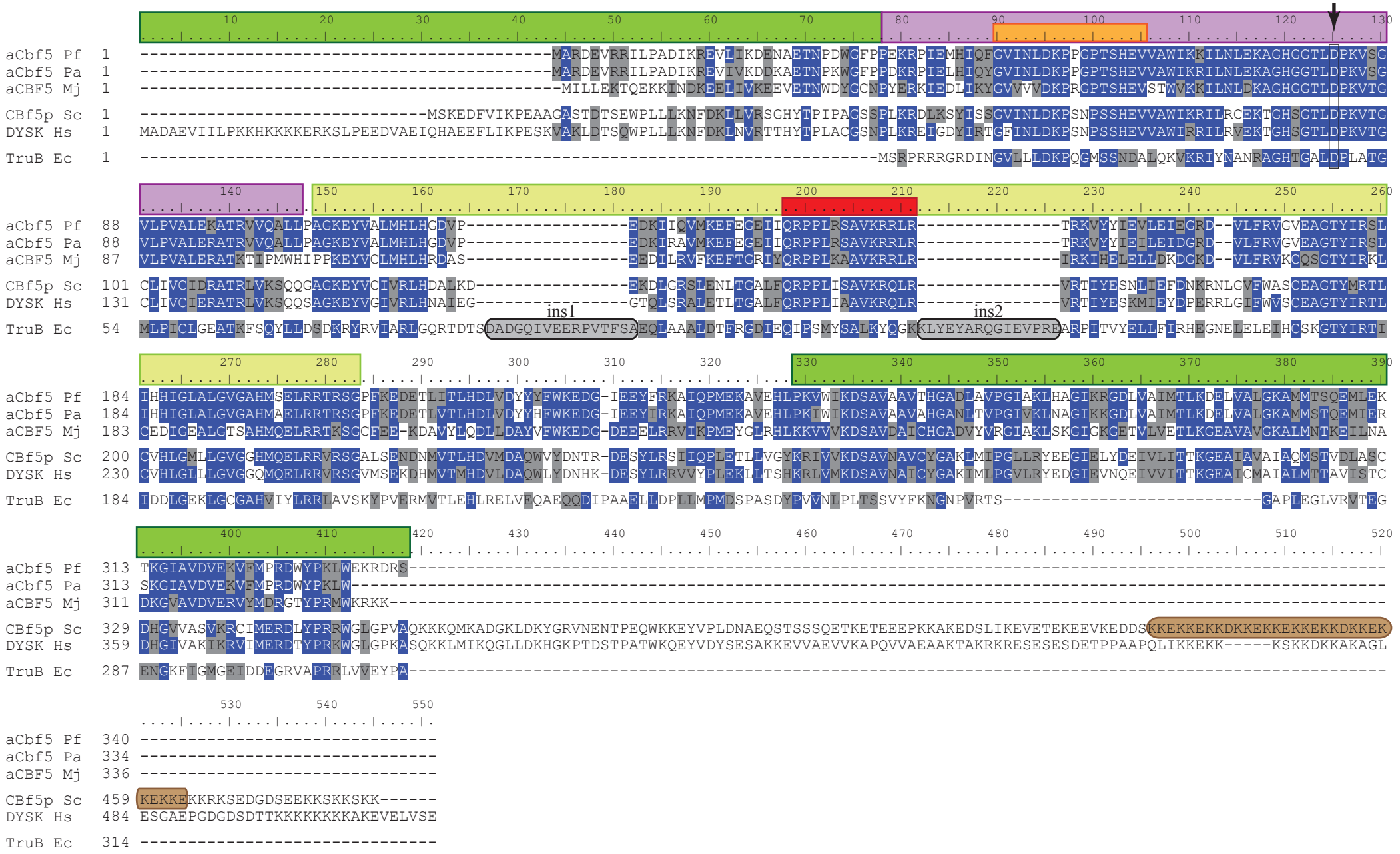


Figure 18. **Alignement des protéines Cbf5 de différentes espèces d'archaea et de leur homologue humain et de levure**. La séquence de la protéine aCbf5 de *Pyroccocus furiosus* (Pf) est comparée à ses homologues de *Pyroccocus abyssi* (Pa), *Methanococcus jannaschii* (Mj), *Saccaromyces cerevisiae* (Sc), *Homo sapiens* (Hs) et *Escherichia coli* (Ec). Le domaine PUA est symbolisé par le rectangle vert et la répétition des 10 motifs KKE/D de Cbf5p par le rectangle brun. Les séquences additionnelles spécifiques de la machinerie d'*E. coli* «ins1» et «ins2» sont indiquées par un rectangle gris. La *«thumb loop»* spécifique des protéines d'Archae et des eucaryotes remplaçant structuralement la séquence «ins2» de TruB est symbolisée par un rectangle rouge. Le motif I est indiqué par un rectangle orange. L'acide aspartique catalytique conservé est repéré par une flèche noire. Les sous domaines catalytiques D1 et D2 sont représentés par un rectangle violet et jaune, respectivement. L'alignement a été réalisé par le logiciel ClustalW2 (Larkin *et al*., 2007) selon une matrice BLOSUM. Les résidus conservés et identiques dans plus de 50% des séquences sont représentés en blanc sur fond bleu et noir sur fond gris respectivement.

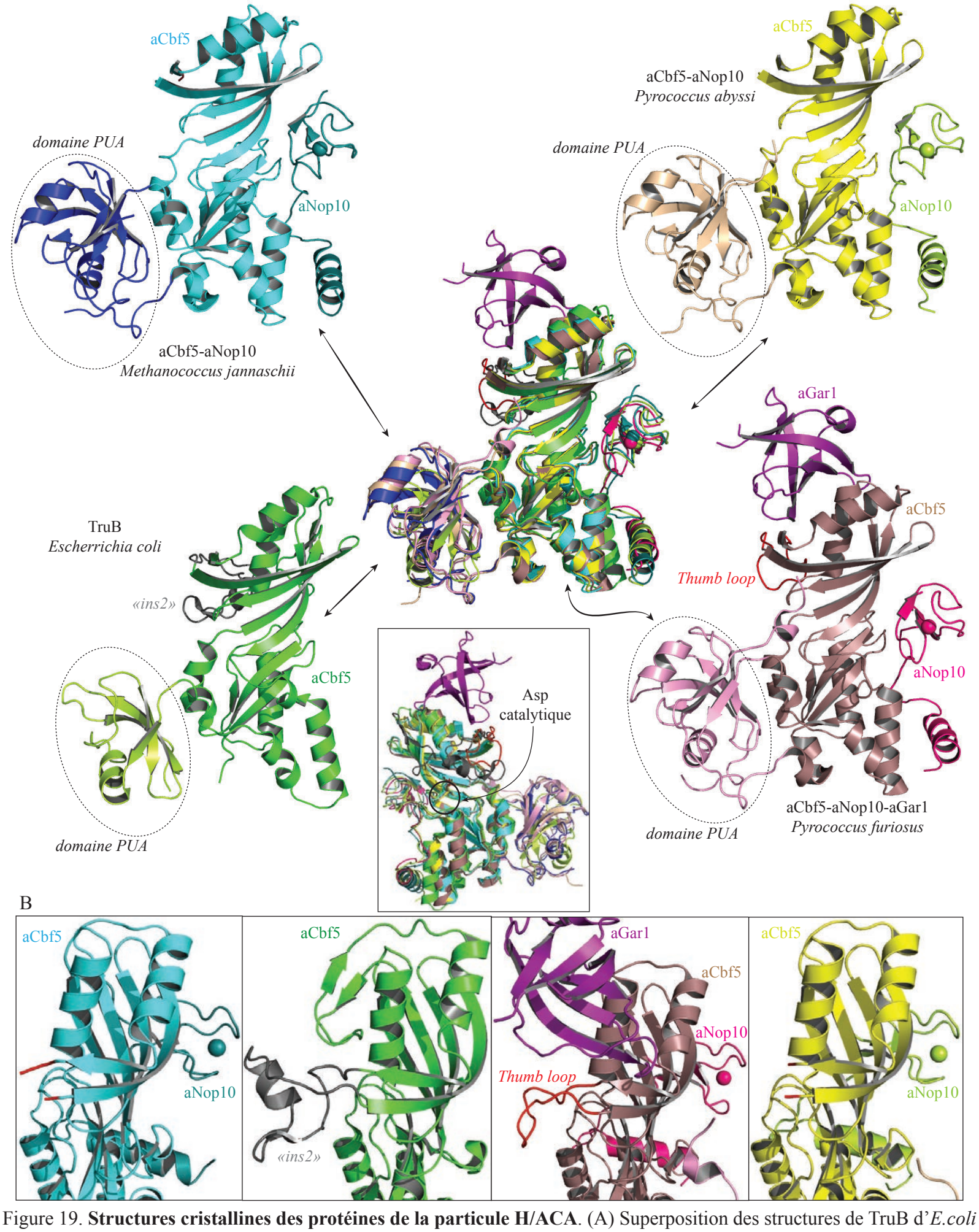


Figure 19. **Structures cristallines des protéines de la particule H/ACA**. (A) Superposition des structures de TruB d'*E.coli* (code PDB : 1K8W), des complexes aCbf5-aNop10 de *Methanococcus jannaschii* (code PDB : 2APO), aCbf5-aNop10 de *Pyrococcus abyssi* (code PDB : 2AUS) et aCbf5-aNop10-aGar1 de *Pyrococcus furiosus* (code PDB : 2EY4). L'acide aspartique catalytique est repéré par un cercle dans l'encart. (B) Vues centrées sur la zone de la *«Thumb loop»* et le l'«ins2» des protéines aCbf5 et TruB, respectivement, en utilisant le même code couleur qu'en (A). Les protéines aNop10 et aGar1 sont indiquées lorsqu'elles sont présentes. L'atome de zinc chélaté par aNop10 est représenté par une sphère. Le domaine PUA de l'enzyme est entouré en pointillé.

En plus de son rôle dans la réaction de pseudouridylation, le domaine catalytique d'aCbf5 d'archaea est impliqué dans des interactions avec les protéines aNop10 et aGar1 (Li & Ye, 2006; Manival *et al.*, 2006; Rashid *et al.*, 2006). Le domaine catalytique de l'enzyme Cbf5 est formé d'un cœur de feuillets β entouré d'hélices α formant la crevasse catalytique (Figure 19). Comme pour TruB (Ramamurthy *et al.*, 1999) un acide aspartique (D82 chez aCbf5, D48 chez TruB et D125 dans la Dyskérine humaine), conservé dans toutes les ARN : Ψ-synthases (Gu *et al.*, 1999; Huang *et al.*, 1998) (Figure 18), est essentiel à l'activité de l'enzyme (Charpentier *et al.*, 2005; Zebarjadian *et al.*, 1999). Récemment, une forme cytoplasmique de la dyskérine a été mise en évidence. Sa production résulte d'évènements d'épissage alternatif, mais sa fonction précise, qui est vraisemblablement indépendante des snoRNP, n'a pas encore été décrite (Angrisani *et al.*, 2011).

L'enzyme TruB d'*E. coli* possède deux segments polypeptidiques caractéristiques qui sont directement impliqués dans la reconnaissance de la boucle TΨC de ses ARNt substrats (Hoang & Ferré-D'Amaré, 2001). Ces segments baptisés « ins1 » et « ins2 » sont absents dans les protéines aCbf5 et leurs homologues humains (Figure 18) et l'un des deux (« ins2) est remplacé par une courte boucle dont la flexibilité rend impossible la détermination de son repliement en absence de la protéine aGar1 (Hamma *et al.*, 2005) (Figure 19). Comme nous le verrons plus tard (paragraphes 2.2.3 et 2.3), cette boucle située entre les brins β6 et β7 est baptisée *« thumb loop »* et joue un rôle dans l'activité de la particule H/ACA (Li & Ye, 2006).

Les protéines recombinantes aCbf5 d'archaea et Cbf5p de levure sont capables de se lier directement à l'ARN guide (Baker *et al.*, 2005; Charpentier *et al.*, 2005; Normand *et al.*, 2006). Mais ces études ont montrés que cette interaction est insuffisante pour guider la formation spécifique de Ψ sur les ARN ribosomiques cibles. Par contre, aCbf5 d'archae et Pus10, une autre Ψ synthase d'archaea, sont capables de catalyser sans ARN guide la formation du résidu Ψ55 des ARNt (Muller *et al.*, 2007; Roovers *et al.*, 2006) et cette activité chez Cbf5 est améliorée en présence de certaines des protéines cœur de la particule.

2.2.2 aNop10/Nop10p

La protéine Nop10 est une petite protéine (environ 60 acides aminés) essentielle à l'activité de la particule (Charpentier *et al.*, 2005) et la stabilité des snoRNA à boîtes H/ACA eucaryotes (Henras *et al.*, 1998). La présence de Cbf5 est nécessaire pour la fixation de Nop10 sur l'ARN guide et l'hétéro-complexe Cbf5-Nop10 peut se former en absence d'ARN (Charpentier *et al.*, 2005; Manival *et al.*, 2006). *In vitro*, Cbf5 complexée à Nop10 et un ARN

guide constituent l'architecture minimale ayant une activité ARN : Ψ synthase sur un ARN substrat (Charpentier *et al.*, 2005). Toutefois, la présence des 4 protéines cœurs est nécessaire pour assurer une activité optimale à la particule H/ACA. Nop10 a une structure très flexible qui comporte un motif à 3 brins β formant un « doigt à zinc » de type *β ribbon* à son extrémité N-terminale et une hélice α C-terminale, reliée à la partie N-terminale par un long segment flexible (Hamma *et al.*, 2005; Manival *et al.*, 2006; Rashid *et al.*, 2006) (Figure 19). La structure doigt à zinc comporte des cystéines arrangées selon le motif CXXC …/… CXXC qui chélatent l'atome métallique. Bien que ce motif N-terminal soit impliqué dans l'interaction avec aCbf5, la mutation des 4 cystéines n'altère ni la formation de la particule, ni son activité (Manival *et al.*, 2006). La protéine eucaryote a les mêmes propriétés structurales et dynamiques mais le doigt à zinc N-terminal est remplacé par un petit feuillet à 2 brins β (Khanna *et al.*, 2006; Reichow & Varani, 2008). La protéine aNop10 se fixe sur une large surface du domaine catalytique de aCbf5 (Hamma *et al.*, 2005; Manival *et al.*, 2006) (Figure 19). Lors de sa fixation à aCbf5, aNop10 se rigidifie en établissant de nombreux contacts avec cette dernière (environ 46% de la surface accessible de la protéine est masquée lors de l'interaction). Néanmoins elle conserve la même structure secondaire que celle observée à l'état libre (Hamma *et al.*, 2005). La région charnière entre les domaines N- et C-terminaux est la plus conservée et établit de nombreux contacts avec une zone d'aCbf5 conservée dans les ARN : Ψ synthases (le motif I) (Koonin, 1996), qui est important pour la stabilité de ces enzymes (Spedaliere *et al.*, 2000). Cette région charnière de Nop10 participe à la poche basique constituée par le site catalytique. Enfin, comme nous le verrons au paragraphe 2.3, au sein de la sRNP, aNop10 établit également des contacts avec le sRNA et la protéine L7Ae (Li & Ye, 2006).

2.2.3 aGar1/Gar1p

La protéine Gar1p est également essentielle à l'activité de la particule, mais c'est la seule à ne pas être requise pour la stabilité du snoRNA (Bousquet-Antonelli *et al.*, 1997; Girard *et al.*, 1992). Cette protéine contient deux domaines présentant des répétitions de séquences glycine / arginine qui sont localisés à chacune de ses extrémités : les domaines GAR (« G*lycin-Arginin-Rich* »). La protéine aGar1 d'archaea ne possède pas ces domaines terminaux présents chez ses homologues eucaryotes mais le domaine fonctionnel est très conservé entre archaea et eucaryotes. Les domaines GAR des protéines nucléolaires interagissent avec les ARN et participent à la localisation nucléolaire de nombreuses protéines comme la Nucléoline ou la Fibrillarine (Ginisty *et al.*, 1999). En présence de la protéine

aNop10, aGAR1 est capable de s'assembler avec aCbf5 pour former un complexe tri-partite, ceci de manière indépendante à la fixation à l'ARN. Tout comme aNop10, aGar1 ne peut pas se lier à l'ARN en absence de CBF5.

La détermination de la structure tridimensionnelle de l'hétérotrimère aCbf5-aNop10-aGar1 a permis de révéler la structuration en tonneau β composé de 6 brins de aGar1 d'archaea (Rashid *et al.*, 2006) (Figure 19). Chez les archaea, malgré son homologie structurale avec des protéines interagissant avec des ARN comme les facteurs d'initiation (IF2) et d'élongation (EF-G) de la traduction bactérienne au sein de la particule, aGar1 n'établit que des liaisons protéine-protéine avec aCbf5. De plus aGar1 se situe à environ 20 Å du site catalytique de l'enzyme. L'interaction observée entre l'ARN et Gar1 eucaryote (Bagni & Lapeyre, 1998; Wang & Meier, 2004) est donc vraisemblablement due aux domaines GAR spécifiques des homologues eucaryotes ou à une association temporaire avec l'ARN substrat au cours de son chargement ou de son relargage de la particule H/ACA (Meier, 2006; Rashid *et al.*, 2006; Yu, 2006). L'idée actuelle est que la protéine aGar1 intervient dans le recyclage (fixation-relargage) de l'ARN substrat. En effet, les protéines Cbf5 sont dépourvues d'un segment peptidique présent dans TruB qui est impliqué dans ce processus. Il reconnait la boucle TΨC des ARNt (mentionné dans le paragraphe 2.2.1, figure 19). L'absence de ce segment étend le site de fixation à l'ARN jusqu'à la tige-boucle β7/β10 (également appelée *« thumb loop »*) de l'enzyme. Cette boucle présente dans aCbf5 interagit directement avec la protéine aGar1. Il a été proposé que cette interaction intervienne dans le recyclage du substrat (Charpentier *et al.*, 2005; Li & Ye, 2006; Rashid *et al.*, 2006).

2.2.4 L7Ae/Nhp2p

La protéine eucaryote Nhp2p/NHP2 et son homologue d'archaea L7Ae font partie des protéines de la famille L7Ae/L30. Cette famille se définit par le partage d'un domaine commun de fixation à l'ARN, appelé L7Ae/L30-RBD (« *L7Ae/L30 RNA Binding Domain* »), qui est responsable de l'interaction avec des motifs structuraux particulier d'ARN : les motifs en *« K-turn »* ou *« K-loop »* (Allmang *et al.*, 2002; Cléry *et al.*, 2007; Marmier-Gourrier *et al.*, 2003; Nolivos *et al.*, 2005; Nottrott *et al.*, 1999; Vidovic *et al.*, 2000; Watkins *et al.*, 2000) déjà évoqués au chapitre 2.1, feront l'objet d'une description détaillée au chapitre 3.2. De façon étonnante, la protéine Nhp2p/NHP2 est le seul membre de la famille L7Ae/L30 qui ne reconnaît pas spécifiquement ces motifs. Sa spécificité est dégénérée (Henras *et al.*, 2001) malgré une légère préférence pour les structures « tige-boucle » imparfaites. Cette

caractéristique est cohérente avec l'absence de ce type de motifs au sein des snoRNA à boîtes H/ACA eucaryotes. Les propriétés de fixation à l'ARN de Nhp2p/NHP2 peuvent s'expliquer par le fait que le résidu acide (l'acide glutamique 61 ou aspartique 59 dans les protéines homologues 15.5K humaine et L7Ae d'archaea respectivement) qui stabilise le résidu uridine protubérant dans les motifs *« K-turn »* n'est pas présent dans cette protéine (Figure 20).

Récemment, la structure tridimensionnelle de la protéine Nhp2p de *S. cerevisiae* a été établie par RMN (Koo *et al.*, 2011) (Figure 21). Nhp2p présente un repliement similaire aux structures connues des protéines de la famille, en particulier la superposition de 3 couches de motifs hélice α - brin β - hélice α. Cependant, Nhp2p peut adopter deux conformations distinctes en solution du fait de l'isomérisation *cis*/*trans* d'une proline conservée, la proline 83, située à l'extrémité N-terminale de l'hélice α3. Dans Nhp2p en solution, la proline 83 adopte majoritairement une conformation *trans* tandis que dans les structures des autres protéines de la famille L7Ae, cette proline adopte une conformation *cis*. L'équipe de J. Feigon a montré que l'isomérisation *cis*/*trans* de P83 influence le positionnement de l'hélice α3 par rapport à l'hélice α2 mais aussi le brin β2 (Figure 21). Ce mauvais positionnement limite la formation des interactions mises en place entre les motifs en *« K-turn »* et les protéines de la famille L7Ae. L'équipe de J.Feigon a montré que la seule mutation S82W dans Nhp2p suffit pour générer une protéine ayant les capacités de reconnaissance des motifs en *« K-turn »* Au sein de la RNP d'archaea, la proline 60 de L7Ae, équivalente à la proline 83 de Nhp2p, établit des contacts directs avec la protéine aNop10, tandis que le résidu Asp59 ou Glu61/S82 chez l'homme/levure établit des liaisons hydrogène avec le résidu uracile éjecté des motifs *« K-turn »* ou *« K-loop »* (Figure 20). La mutation S82W dans Nhp2p, permet au résidu acide voisin de la proline 83 d'adopter une conformation identique à celle observée par le résidu Asp59 dans L7Ae ce qui renforce l'affinité pour les motifs en *« K-turn »*. La conformation de la proline conservée des protéines de la famille jouerait donc un rôle majeur dans le positionnement de résidus et de la chaine principale au sein du site de fixation des nucléotides. En ce qui concerne Nhp2p, malgré l'absence de véritable *« K-turn »* dans les snoRNA H/ACA, la sérine 82 pourrait être impliquée dans la reconnaissance d'un résidu U, retrouvé de façon préférentielle à une distance de 10 ou 11 paires de bases de la poche de pseudouridylation. Elle jouerait donc un rôle équivalent au résidu aspartique 59 de L7Ae. L'interaction de Nhp2p avec Nop10p au sein de la snoRNP pourrait stabiliser la conformation *cis* de la proline 83 et permettrait l'établissement de contacts optimaux entre Nhp2p et le snoRNA (Koo *et al.*, 2011).

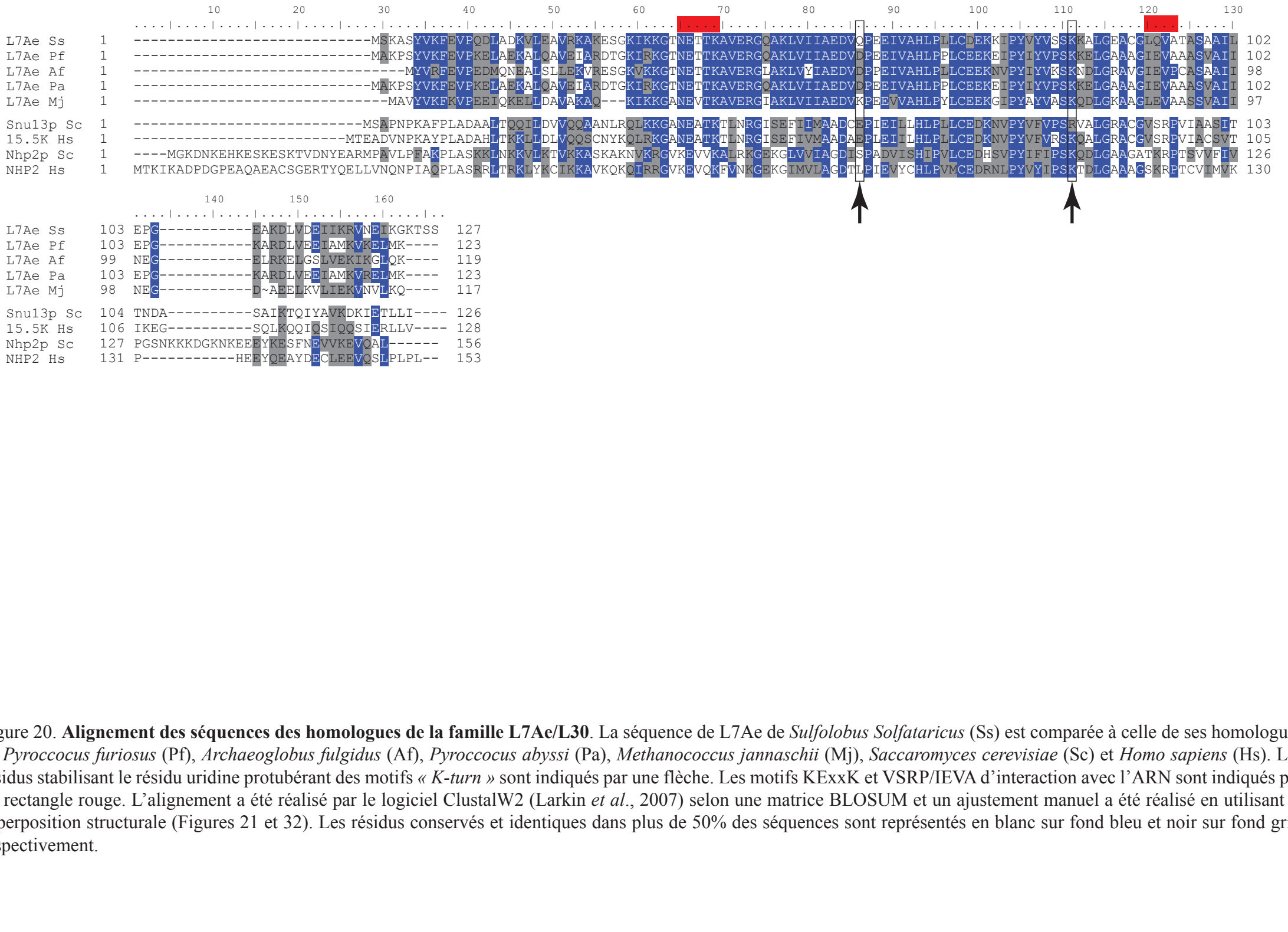


Figure 20. **Alignement des séquences des homologues de la famille L7Ae/L30**. La séquence de L7Ae de *Sulfolobus Solfataricus* (Ss) est comparée à celle de ses homologues de *Pyroccocus furiosus* (Pf), *Archaeoglobus fulgidus* (Af), *Pyroccocus abyssi* (Pa), *Methanococcus jannaschii* (Mj), *Saccaromyces cerevisiae* (Sc) et *Homo sapiens* (Hs). Les résidus stabilisant le résidu uridine protubérant des motifs *« K-turn »* sont indiqués par une flèche. Les motifs KExxK et VSRP/IEVA d'interaction avec l'ARN sont indiqués par un rectangle rouge. L'alignement a été réalisé par le logiciel ClustalW2 (Larkin *et al*., 2007) selon une matrice BLOSUM et un ajustement manuel a été réalisé en utilisant la superposition structurale (Figures 21 et 32). Les résidus conservés et identiques dans plus de 50% des séquences sont représentés en blanc sur fond bleu et noir sur fond gris, respectivement.

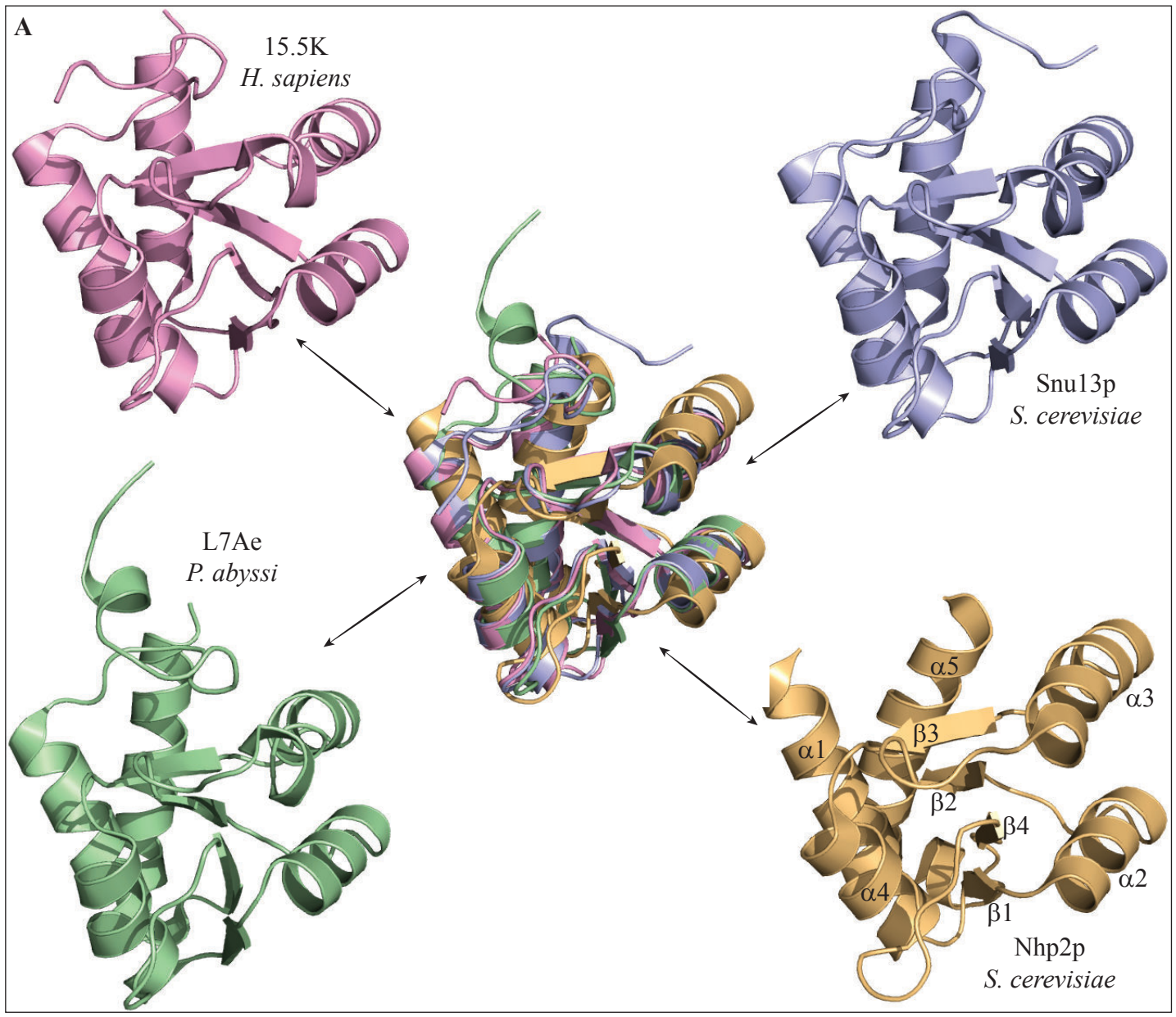


Figure 21. **Structures tridimensionnelles des protéines de la famille L7Ae/L30.** (A) Superposition des structures de 15.5K humaine (code PDB : 1E7K), Snu13p de *S.cerevisiae* (code PDB : 2ALE), L7Ae de *P.abyssi* (code PDB : 1PXW) et Nhp2p de *S.cerevisiae* (code PDB : 2LBX et 2LBW). (B) Superposition des structures de la protéine Nhp2p de *S.cerevisiae* WT (orange) et mutante S82W (bleu). Le carbone alpha de la proline 83 est représenté par une sphère noire. Les numéros d'hélices alpha et des brins beta sont indiqués.

2.3 Structure des particules à boîtes H/ACA

Au cours de ces dix dernière années, plusieurs structures tridimensionnelles de particules sRNP de *Pyrococcus furiosus* ont été résolues (Duan *et al.*, 2009; Li & Ye, 2006; Liang *et al.*, 2007a; Liang *et al.*, 2009a; Zhou *et al.*, 2011). La première correspondait à un complexe formé entre une seule tige-boucle d'ARN guide H/ACA adoptant une conformation étendue en double hélice et les protéines aCbf5, aNop10 et L7Ae (Li & Ye, 2006). Cette structure montrait que bien que la protéine aGar1 n'établit pas d'interaction avec l'ARN, elle est capable de séquestrer la boucle « pouce » (ou *« thumb loop »*) d'aCbf5 délimitée par les feuillets β7 et β10. La modélisation d'un substrat lié à la structure établie renforçait l'idée que aGar1 régule le turnover du substrat en contrôlant la position de la « *thumb loop* » (Li & Ye, 2006). Cependant, cette structure n'apportait pas d'informations sur le mode de recrutement de l'ARN substrat sur la RNP. La structure du complexe aCbf5-aNop10-aGar1 associé à l'ARN guide et à l'ARN substrat, qui a été obtenue ultérieurement, n'a pas non plus apporté de réponse claire à cette question. En effet, en absence de la protéine L7Ae, l'ARN substrat n'était pas correctement positionné dans le site actif de la particule (Liang *et al.*, 2007a). Plus récemment, l'équipe de K. Ye a apporté des éléments de réponse à cette question en déterminant la structure 3D d'une sRNP à boîtes H/ACA associée à un ARN substrat positionné dans le site actif (Duan *et al.*, 2009) (Figure 22). Les auteurs ont obtenu deux types de complexes, un premier où l'ensemble des partenaires protéiques et ARN sont présents et un second, où la protéine aGar1 est absente. Dans ce dernier cas, l'ARN à boîtes H/ACA associé aux protéines aCbf5, aNop10 et L7Ae adopte une conformation similaire à celle observée dans la structure du complexe sans substrat (Li & Ye, 2006) (Figure 22A droite et centre). Au niveau de la structure sans aGar1, l'hélice supérieure (P2) du guide interagit avec L7Ae, aNop10 et avec le sous domaine catalytique D1 de aCbf5, tandis que l'hélice inférieure (P1) est ancrée au domaine PUA de Cbf5 par l'extrémité de la boîte ACA (Figure 22B). L'ARN substrat, associé à la poche de pseudouridylation à proximité du site actif, adopte une structure en forme « d'arc de cercle » à la surface des protéines (Figure 22B). Le substrat est maintenu dans cette position par des contacts supplémentaires, en particulier, la « *thumb loop* » d'aCbf5 semble jouer un rôle clé dans l'accessibilité de la poche de pseudouridylation pour le substrat (Duan *et al.*, 2009; Liang *et al.*, 2007a; Liang *et al.*, 2009a). La structure de la particule entière sRNP H/ACA en complexe avec le substrat ARN montre que la protéine aGar1 interagit avec le sous domaine catalytique D2 de la protéine aCbf5, en adoptant un positionnement similaire à celui observé dans la structure du complexe sans substrat (Figure

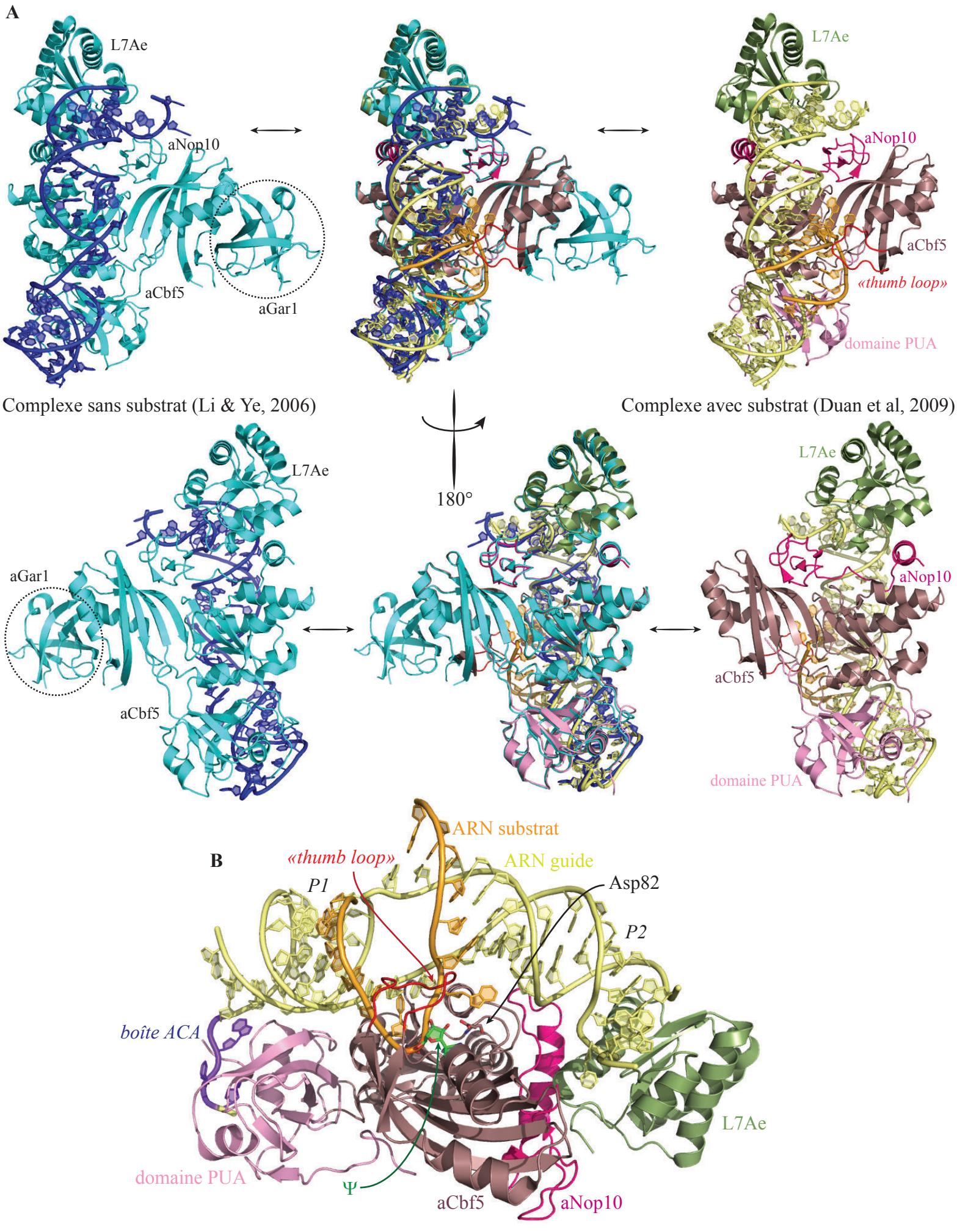


Figure 22. **Structures tridimensionnelle d'une partie d'une sRNP à boîtes H/ACA de *P. furiosus***. (A) Comparaison entre la sRNP sans l'ARN substrat (les protéines sont bleu clair et l'ARN guide bleu foncé) (Li & Ye, 2006), code PDB : 2HVY et la sRNP avec l'ARN substrat mais sans la protéine aGar1 (le nom des protéines est indiqué) (Duan *et al.*, 2009), code PDB : 3HAX. (B) L'ARN substrat adopte une structure en forme de « d'arc de cercle » qui est aligné à la surface des protéines. Le nom des protéines et des ARN sont indiqués.

23). La protéine aGar1 n'établit aucun contact avec l'ARN guide ou avec l'ARN substrat. Le reste du complexe, incluant les hélices formées entre l'ARN guide et l'ARN substrat, ainsi que la « *thumb loop »,* adopte la même structuration qu'en absence de aGar1. Néanmoins, en absence de substrat, la « *thumb loop* », stabilisée par une interaction avec la protéine aGar1, est dans une configuration ouverte favorisant l'accès à un ARN cible. Ensuite, en présence du substrat, la boucle se referme sur la poche de pseudouridylation, ce qui participerait au maintien de l'ARN cible dans le site actif de l'enzyme (Duan *et al.*, 2009; Li & Ye, 2006). Comme déjà observé auparavant, les extrémités du sRNA guide sont stabilisées d'une part grâce à l'interaction de la boîte ACA avec le domaine PUA de CBF5 et d'autre part grâce à l'association de la protéine L7Ae avec le motif ARN en *« K-loop »* (Figure 23B et C). A ce niveau, on observe l'interaction de l'acide aspartique 59 de la protéine L7Ae avec le résidu uracile projeté vers le solvant au niveau du motif *« K-turn »* de l'ARN guide (mentionné au chapitre 2.2.4). Il a été montré que des changements de conformation de l'ARN guide en présence de la protéine L7Ae favorisent également son interaction avec le substrat (Liang *et al.*, 2007a). Des résultats encore non publiés obtenus au laboratoire par J.B. Fourmann, M. Blaud et B. Charpentier, indiquent que ce rôle de la protéine L7Ae reposerait sur des contraintes induites par son interaction au sein de la particule avec la protéine aNop10.

Récemment, la structure cristallographique d'un complexe partiel Nop10p-Cbf5p-Gar1p de *S. cerevisiae,* dans lequel le domaine KKE/D de Cbf5p et les domaines GAR de la protéine Gar1p ont été retirés, a été obtenue à partir de protéines recombinantes co-exprimées chez *E. coli* (Li *et al.*, 2011c) (Figure 24). Les auteurs ont tout d'abord démontré que la RNP composée de Cbf5p, Gar1p, Nop10p, Nhp2p et de l'ARN guide snR5 est active *in vitro*. La comparaison de l'activité de ce système avec la protéine Cbf5p dépourvue du domaine KKE/D a permis de démontrer que ce dernier n'est pas requis pour la réaction. Par contre, l'incorporation de la protéine Gar1p de pleine taille (mais cependant partiellement dégradée) augmente significativement le taux de pseudouridylation comme c'est le cas pour la particule d'archaea (Charpentier *et al.*, 2005).

Enfin des expériences de pontages chimiques covalents couplées à une analyse par spectrométrie de masse ont permis aux auteurs de confirmer que la position de la protéine Nhp2p au sein de la RNP est similaire à celle occupée par la protéine L7Ae dans la particule H/ACA d'archaea (Li *et al.*, 2011c). De façon étonnante, la particule H/ACA en absence de protéine Nhp2p possède toujours une forte activité de pseudouridylation, même si celle-ci est inférieure à celle de la particule complète. Ceci contraste avec le rôle essentiel que joue son

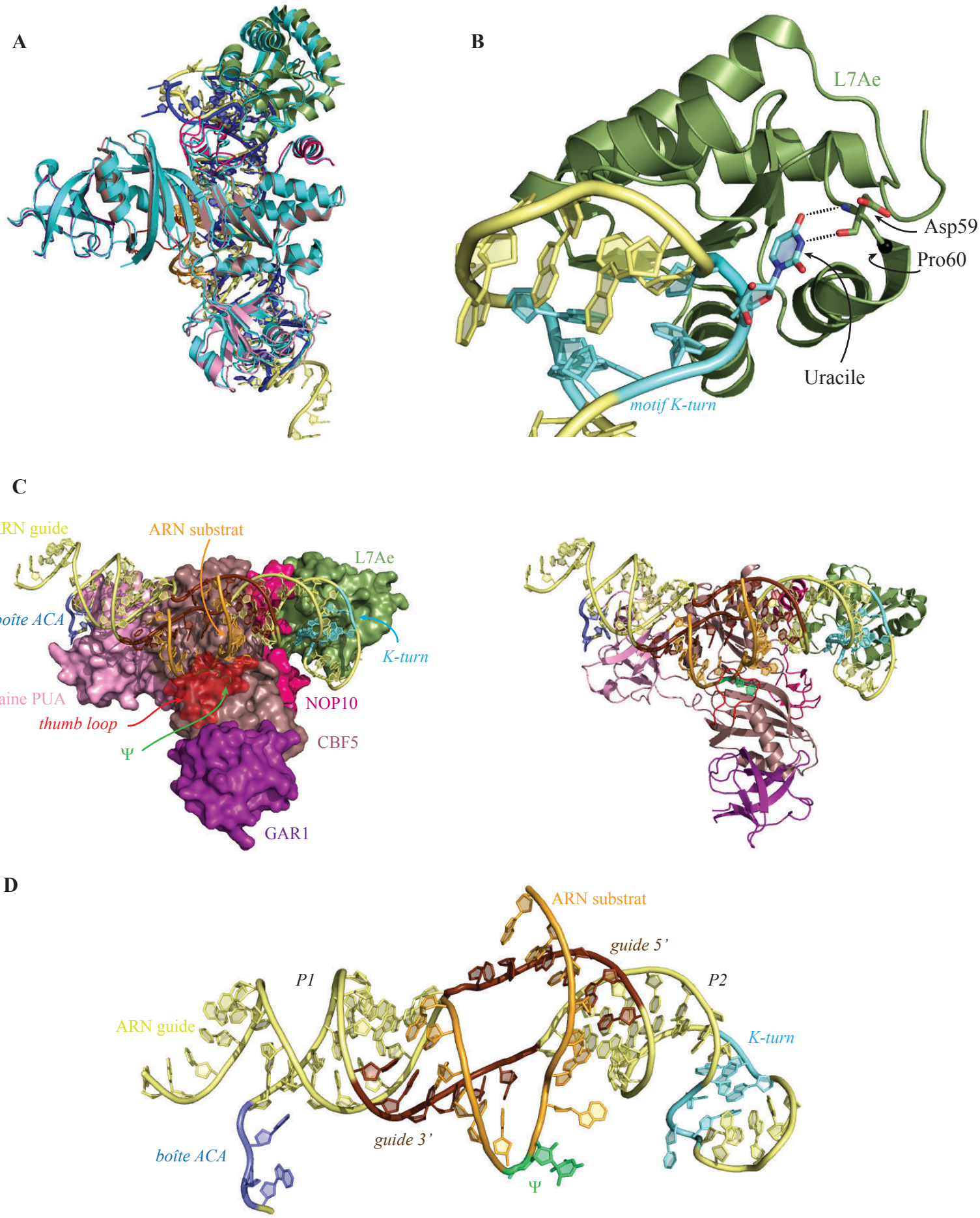


Figure 23. **Structure tridimensionnelle d'une sRNP à boîtes H/ACA entière et active de *P. furiosus***. (A) Comparaison entre la sRNP sans l'ARN substrat (les protéines sont représentées en bleu clair et l'ARN guide en bleu foncé) (Li & Ye, 2006), code PDB : 2HVY. La sRNP avec l'ARN substrat est représentée avec le même code couleur qu'en (C) (Duan *et al*., 2009), code PDB : 3HAY. (B) Détails de l'interaction entre le résidu uracile projeté du motif en *« K-turn »* et l'acide aspartique 59 de la protéine L7Ae. Le carbone alpha de la proline 60 est représenté par une sphère noire. (C) Vue d'ensemble de la sRNP H/ACA de *P. furiosus* en surface moléculaire et cartoon. (D) Détails des ARN guide H/ACA et substrat de la sRNP.

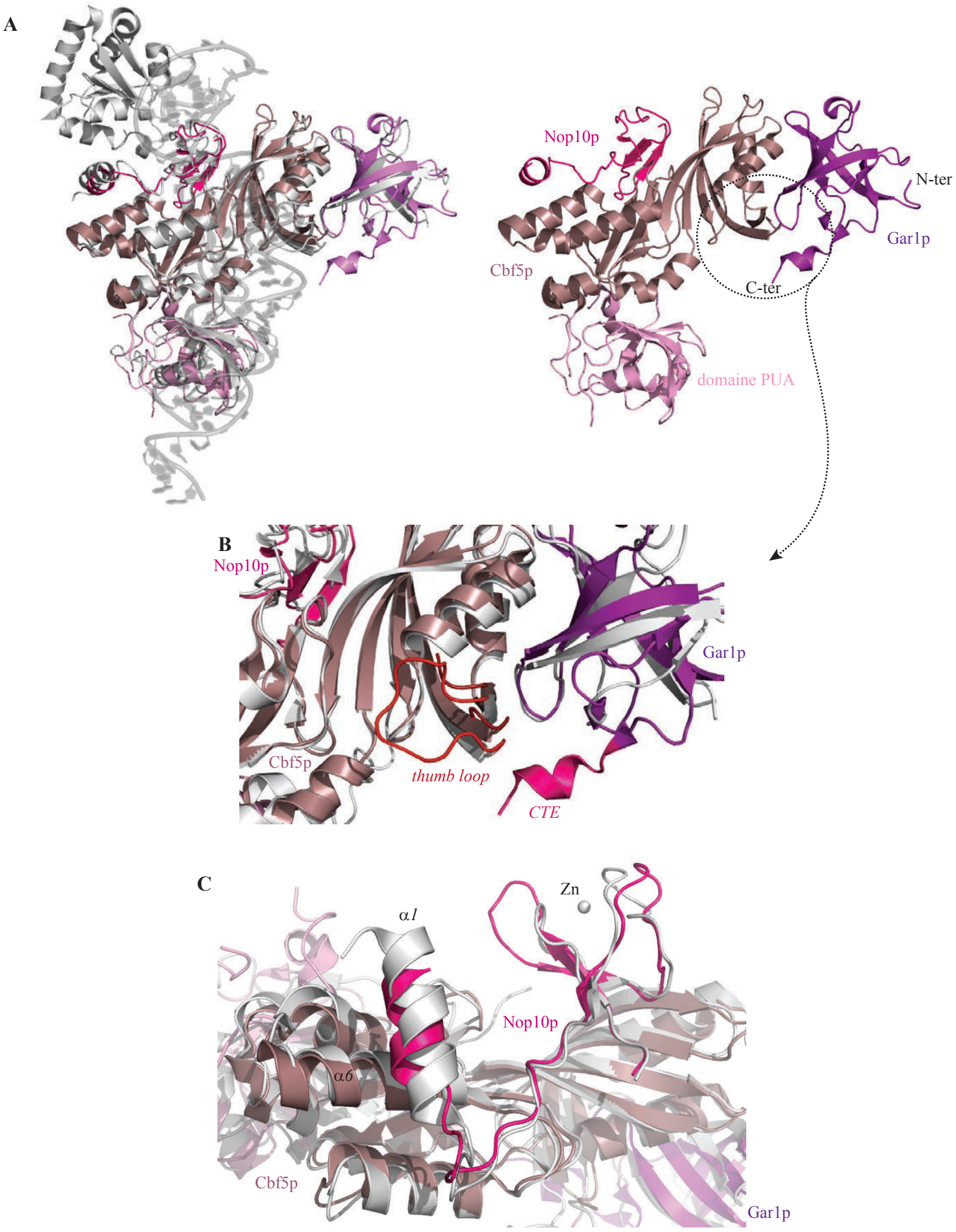


Figure 24. **Structures tridimensionnelles de la sRNP à boîtes H/ACA de *P. furiosus* et du complexe Cbf5p-Nop10p-Gar1p de *S.cerevisiae***. (A) A gauche, superposition entre la sRNP de *P.furiosus* (les protéines sont grises claires et l'ARN en transparence grise) (Duan *et al.*, 2009), code PDB : 3HAY et le complexe Cbf5p-Nop10p-Gar1p de *S. cerevisiae* (Li *et al.*, 2011), code PDB 3U28. A droite le complexe protéique tripartite de levure. (B) Vue de l'interaction entre le domaine CTE de la protéine Gar1p et la *«thumb loop»* de Cbf5p. (C) Superposition de la vue de la protéine Nop10p et de son interaction avec l'hélice α6 de Cbf5p (en couleur) avec l'équivalent chez les Archaea (en gris).

homologue d'archaea L7Ae. Nhp2p et L7Ae semblent donc présenter des différences fonctionnelles au sein de leurs RNP H/ACA respectives.

Au sein de la structure cristalline établie pour le complexe de levure, la conformation de chaque sous-unité, l'organisation globale et les interactions protéine-protéine sont similaires à ce qui a été précédemment observé chez les archaea (Hamma *et al.*, 2005; Li & Ye, 2006; Manival *et al.*, 2006; Rashid *et al.*, 2006) (Figure 24). Nop10p et Gar1p s'associent de part et d'autre du domaine catalytique de l'enzyme Cbf5p. Mise à part une légère rotation du domaine PUA d'environ 20° par rapport au domaine catalytique, les variations de positionnement des différents domaines sont mineures (Li *et al.*, 2011c).

La structure du complexe révèle aussi des caractéristiques spécifiques aux trois protéines de levure :

- La protéine Cbf5p eucaryote se distingue de son homologue archaea, par des extensions N-terminale (NTE) et C-terminale (CTE), ainsi que par la présence d'une longue queue de répétition KKE/D C-terminale (Figure 17). Ces extensions ont un intérêt tout particulier puisque 15 mutations impliquées dans la Dyskératose Congénitale sont retrouvées dans ces régions. Le domaine CTE de Cbf5p n'a pas été cristallisé, ceci résulte probablement de sa dégradation ou d'un désordre intrinsèque. Le NTE est partiellement structuré et forme une couche supplémentaire qui recouvre le tonneau β du domaine PUA sur la face opposée à la surface d'interaction avec l'ARN chez l'archaea.
- Chez la levure, Nop10p adopte une structure étendue composée d'un domaine N-terminal en feuillet, une région charnière et une hélice α C-terminale et contacte largement le domaine catalytique de Cbf5p (Figure 24). Le domaine N-terminal de *P. furiosus* est un domaine à doigt de zinc classique avec 4 résidus cystéines capables de lier un atome de zinc, tandis que le domaine homologue chez la levure ne possède pas ces cystéines et ne lie pas d'atome de zinc. Malgré cette différence, ce domaine garde le même repliement et la même interface avec Cbf5p (Figure 24C). En absence de fixation du zinc, l'association à Cbf5p est donc nécessaire au maintien de la structuration de ce domaine. Au niveau du domaine C-terminal, l'hélice α1 d'aNop10 d'archaea interagit avec l'hélice α6 d'aCbf5. Cette association est différente chez la levure car les deux hélices s'éloignent l'une de l'autre.

Wang et Meier en 2004 ont montré que Nhp2p est capable de s'associer à Cbf5p-Nop10p pour former un complexe stable même en absence d'ARN (Wang & Meier, 2004) et peut donc se lier à Nop10p d'une façon différente de L7Ae d'archaea qui elle n'est pas capable de former un complexe stable avec aCbf5-aNop10 (Baker *et al.*, 2005; Charpentier *et al.*, 2005). Les changements conformationels de la région C-terminale de Nop10p de levure génèrent donc probablement une nouvelle interface de fixation pour la protéine Nhp2p.

- Enfin, en ce qui concerne la protéine Gar1p, tout comme son homologue archaea, son domaine cœur est composé d'un tonneau β à 6 brins mais elle possède un domaine NTE et un domaine CTE supplémentaire. Le domaine NTE comporte 2 prolines très conservées chez les eucaryotes dont l'absence provoque l'agrégation de la protéine, la perte de sa localisation nucléolaire et la disparition des « *Cajal bodies* » (CB) (Pogacić *et al.*, 2000). Le domaine CTE contient une courte hélice α1 qui établit des interactions hydrophobes avec la *« thumb loop »* de Cbf5 (Figure 24B) et qui pourrait être importante pour le turnover réactionnel (Li *et al.*, 2011c).

2.4 La biogenèse des RNP à boîtes H/ACA

Les ARN à boîtes H/ACA représentent une classe de petits ARN non codants abondants et conservés. Comme nous venons de l'évoquer, ces ARN agissent au sein de complexes ribonucléoprotéiques, les snoRNP à boîtes H/ACA. Ils agissent sur les pré-ARNr en guidant l'une des modifications post-trancriptionnelles les plus abondantes : la conversion des résidus uridines en pseudouridines (Charette & Gray, 2000). Mais leurs substrats ne se limitent pas seulement aux ARNr, ces complexes sont capables de cibler également des petits ARN nucléaires impliqués dans l'épissage des ARNm : les UsnRNA. Certaines des snoRNP H/ACA en plus de guider la conversion de U en Ψ sont également impliquées dans la maturation nucléolytique des pré-ARNr. C'est le cas des snoRNA snR30 et snR10 (Bally *et al.*, 1988; Collins, 2006; Morrissey & Tollervey, 1993; Tollervey, 1987). De plus, chez les vertébrés, le complexe de la télomérase contient un ARN non codant, qui porte dans sa région 3' une structure en « tige-boucle » à boîtes H/ACA qui s'associe aux quatre protéines des snoRNP H/ACA (Antal *et al.*, 2002) (Dez *et al.*, 2001; Dragon *et al.*, 2000; Mitchell *et al.*, 1999). Ainsi, les particules snoRNP H/ACA sont essentielles dans 3 processus cellulaires fondamentaux : la synthèse des protéines, l'épissage des ARNm et la maintenance de l'intégrité du génome. Chez tous les eucaryotes, les snoRNA à boîtes H/ACA s'accumulent dans le nucléole, tandis que les ARN H/ACA médiant la pseudouridylation des snRNA

splieceosomaux autres que le UsnRNA U6, les scaRNA (« *Cajal body-specific RNAs* ») résident dans les CB des cellules animales (Darzacq *et al.*, 2002). Ces scaRNA possèdent l'ensemble des caractéristiques de la famille des snoRNA H/ACA, comme la présence des boîtes H et ACA, la poche de pseudouridylation mais aussi un signal de localisation vers les CB, la boîte CAB (pour Cajal body), qui correspond à une séquence consensus ugAG située dans la boucle apical des structures « tige-boucle » 5' et 3' de l'ARN (Richard *et al.*, 2003). Le snoRNA U6 qui transite dans le nucléole est modifié par les snoRNP.

Depuis dix ans des particules RNP d'archaea fonctionnelles ont pu être reconstituées *in vitro*, tout comme récemment des RNP de levure (Li *et al.*, 2011c). Cependant ces dernières ont nécessité une co-expression et l'utilisation de protéines tronquées, ce qui suggère que la biogenèse des RNP H/ACA d'eucaryotes est un processus complexe qui requiert de multiples facteurs d'assemblage (Boulon *et al.*, 2008; Kiss *et al.*, 2006; Matera *et al.*, 2007; Terns & Terns, 2006). Et comme nous allons le voir cela se vérifie dès la transcription de l'ARN guide à boîtes H/ACA.

Chez les archaea et les eucaryotes unicellulaires, les gènes des ARN H/ACA sont fréquemment organisés en unité de transcription mono- ou polycistronique indépendantes, alors que chez les vertébrés, on observe une intégration dans les parties introniques des unités codant les ARN H/ACA. Ainsi, tous les gènes des ARN H/ACA de vertébrés, à l'exception de l'ARN de la télomérase, sont hébergés dans des introns de pré-ARNm. Chez la levure, l'assemblage des RNP H/ACA débute de façon co-transcriptionelle sur l'ARN H/ACA naissant (Ballarino *et al.*, 2005; Yang *et al.*, 2005). De façon similaire, chez les mammifères, le recrutement des protéines sur les ARN à boîtes H/ACA introniques s'effectue sur le pré-ARNm naissant, de façon concomitante avec la sélection des sites d'épissages (Darzacq *et al.*, 2006; Richard *et al.*, 2006) (Figure 25).

2.4.1 Rôle de Naf1 et Shq1

Les facteurs d'assemblage précoces des particules H/ACA eucaryotes, Naf1 (*« Nuclear Assembly Factor 1 »*) et Shq1 (*« Small nucleolar RNAs of the box H/ACA family Quantitative accumulation 1 »*) sont les plus caractérisés à l'heure actuelle (Darzacq *et al.*, 2006; Dez *et al.*, 2002; Fatica *et al.*, 2002; Yang *et al.*, 2002). Ces protéines s'accumulent dans le nucléoplasme mais sont exclues du nucléole et des CB, où les snoRNP et scaRNP H/ACA fonctionnelles résident (Darzacq *et al.*, 2006; Dez *et al.*, 2002; Grozdanov *et al.*, 2009; Kittur *et al.*, 2006; Yang *et al.*, 2002) (Figure 25). Les protéines Naf1 et Shq1 sont

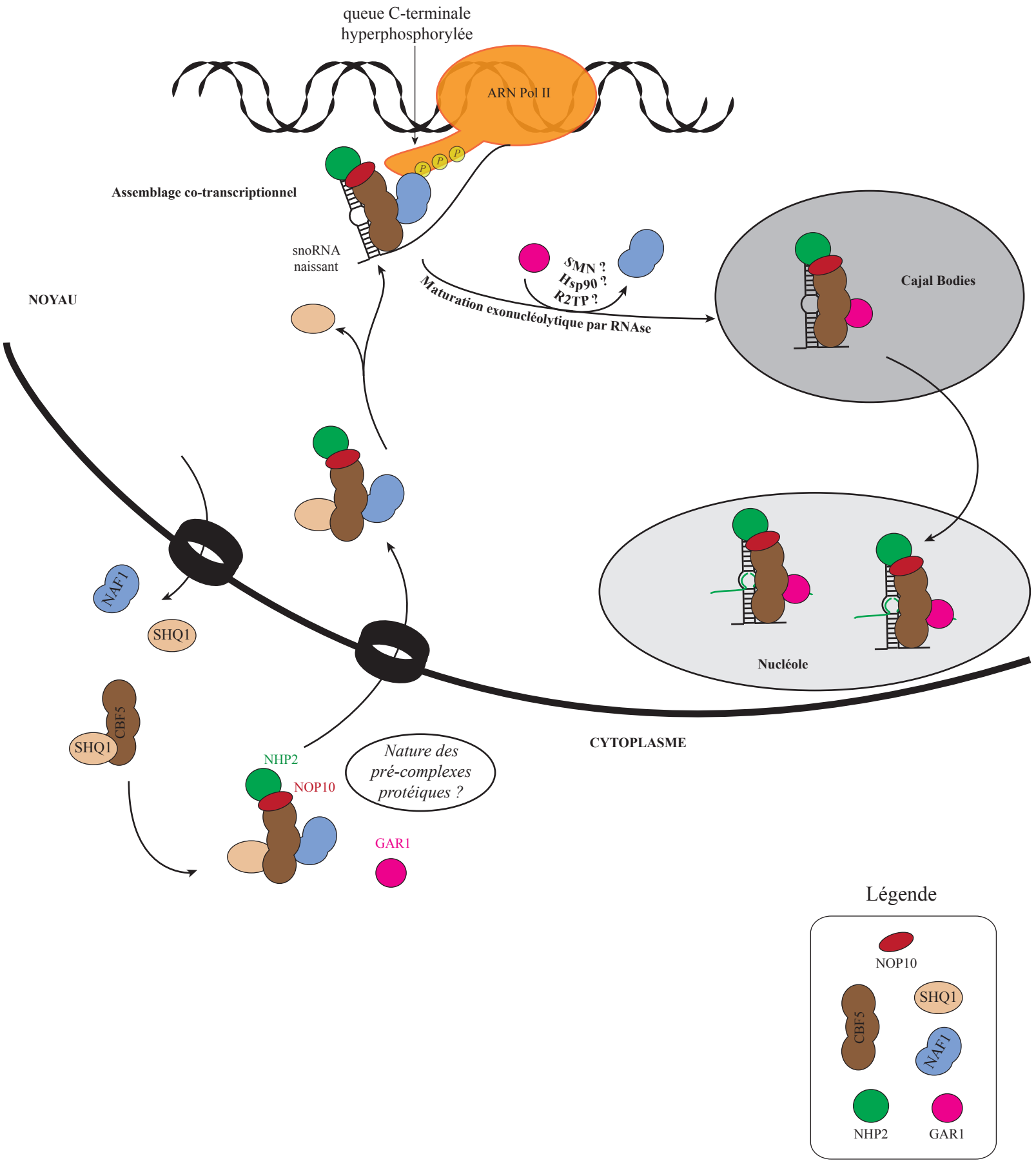


Figure 25. **Représentation schématique du modèle possible d'assemblage des snoRNP à boîtes H/ACA chez les eucaryotes**. Ce mécanisme semblerait être conservé entre l'homme et la levure mais l'intervention éventuelle du complexe S.M.N. constitue une spécificité de mammifères. D'après (Boulon et al., 2008).

toutes deux capables de s'associer à la protéine Cbf5, probablement pour stabiliser cette protéine qui est intrinsèquement instable, mais aussi pour faciliter son transport et sa localisation dans le noyau (Darzacq *et al.*, 2006; Dez *et al.*, 2002; Fatica *et al.*, 2002; Godin *et al.*, 2009). Initialement il était proposé que Shq1 ne soit pas capable de s'associer à la protéine Cbf5p/Dyskérine libre, alors que Naf1 l'était, puisqu'il avait été retrouvé associé à la pré-RNP H/ACA contenant les protéines Dyskérine-Nop10-NHP2 (Grozdanov *et al.*, 2009) (Figure 25). Mais plus récemment, il a été montré *in vitro* que la protéine Shq1p de levure est capable de se lier à la protéine Cbf5p, indépendamment de la présence des autres protéines s'associant à Cbf5p telles que Nop10p, Nhp2p, Gar1p et même Naf1p (Li *et al.*, 2011b). Il est maintenant proposé que Shq1 empêche la protéine Cbf5p/Dyskérine nouvellement synthétisée d'interagir de manière aspécifiques avec des ARN jusqu'à son incorporation au sein des pré-snoRNP (Godin *et al.*, 2009; Grozdanov *et al.*, 2009; Li *et al.*, 2011b; Walbott *et al.*, 2011). Shq1p accompagnerait Cbf5p dans le nucléoplasme en association avec Nop10p, Nhp2p et Naf1p, ceci jusqu'au lieu de transcription des snoRNA (Figure 25). Naf1 est capable d'interagir avec le domaine C-terminal phosphorylé de l'ARN polymérase II et des facteurs de transcription spécifiques de l'ARN Polymérase II (Fatica *et al.*, 2002; Yang *et al.*, 2005). L'idée est que Naf1p permette l'adressage du complexe Cbf5p-Nop10p-Nhp2p sur l'ARN H/ACA en cours de synthèse par le biais de son association avec l'ARN Polymérase II. De façon intéressante, il a été montré que les protéines Naf1, Cbf5, Nop10 et Nhp2 s'accumulent de façon spécifique au niveau des gènes des ARN H/ACA transcriptionnellement actif chez la levure et chez les mammifères (Ballarino *et al.*, 2005; Darzacq *et al.*, 2006; Yang *et al.*, 2005). Au site de transcription se forme alors une particule pré-RNP H/ACA inactive composée de l'ARN H/ACA précurseur associé au complexe Naf1-Cbf5/Dyskérine-Nop10-Nhp2 mais dans lequel la protéine Gar1 est absente (Figure 25).

La structure tridimensionnelle du facteur d'assemblage Naf1p de levure montre un domaine cœur, formé par un tonneau β à 6 brins, homologue à celui de Gar1 d'archaea et de levure. Ce domaine est responsable de l'association avec la protéine Cbf5p (Leulliot *et al.*, 2007) (Figure 26). Il est proposé que lorsque Naf1p remplace la protéine Gar1p au sein de la RNP, il ne bloque pas la réaction de pseudouridylation mais son turnover (Li *et al.*, 2011c). Bien que structuralement homologue comme nous le montre la figure 26, Naf1p ne possède pas le domaine CTE interagissant avec la « *thumb loop* » de Cbf5p et ne peut donc pas agir comme Gar1p dans l'activité de la particule, ce qui est en adéquation avec son rôle comme facteur d'assemblage (Li *et al.*, 2011c). De plus, le domaine CTE de Gar1p permettrait de

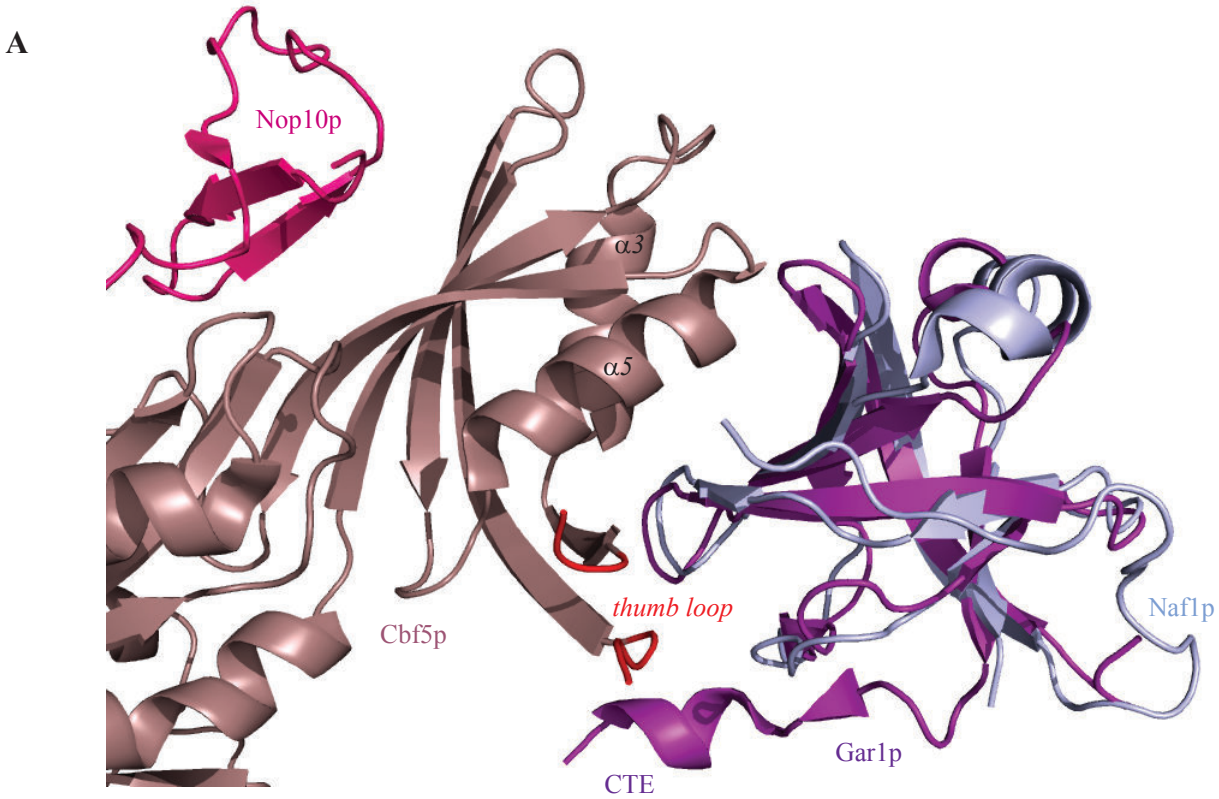


Figure 26. **Superposition des structures tridimensionnelles de Gar1p dans le complexe Cbf5p-Nop10p-Gar1p et du facteur d'assemblage Naf1p de *S.cerevisiae***. Vue de face (A) et de derrière (B) sur la partie de Cbf5p en interaction avec Gar1p. Pour le complexe Cbf5p-Nop10p-Gar1p (Li et al., 2011b) (code PDB 3U28) nous avons utilisé le même code couleur que dans la figure 24. La protéine Naf1p (Leulliot et al., 2007) (code PDB : 2V3M) représentée en bleu clair a été superposée à la protéine Gar1p du complexe.

chasser le facteur Naf1p de la protéine Cbf5 durant la maturation de la particule H/ACA (Li *et al.*, 2011c). En effet, par l'intermédiaire de leur domaine commun, Naf1p et Gar1p s'associent à la protéine Cbf5p de façon mutuellement exclusive (Darzacq *et al.*, 2006; Wang & Meier, 2004). La dernière étape de l'assemblage des snoRNP H/ACA pourrait donc être l'échange entre le facteur d'assemblage Naf1p et la protéine Gar1p. Ceci pourrait ainsi permettre de contrôler la transition d'une forme inactive de la pré-RNP à une particule H/ACA mature et fonctionnelle (Darzacq *et al.*, 2006). Des données non publiées d'A. Huttin et S. Massenet au laboratoire suggèrent que le complexe S.M.N. pourrait jouer un rôle dans cette étape.

Contrairement à la protéine Naf1p, Shq1p n'est pas retrouvée au niveau des sites de transcription des snoRNA à boîtes H/ACA (Grozdanov *et al.*, 2009; Yang *et al.*, 2005). De façon étonnante, bien qu'elle réalise la navette entre le cytoplasme et le noyau, Shq1p ne possède pas de signal de localisation nucléaire propre et sa translocation vers le noyau pourrait dépendre de son association avec Cbf5p, qui possède son propre NLS (*« Nuclar localization signal »*) (Grozdanov *et al.*, 2009). Le facteur Shq1p contient deux domaines : le domaine N-terminal nommé domaine CS (« *CHORD-containing proteins and SGT1 »*) ou « *HSP20-like domain »,* et un domaine C-terminal possédant une forte homologie avec les autres protéines Shq1, appelé domaine Shq1 (ou SSD pour « *Shq1-specific domain »*) (Figure 27). Le domaine CS est requis pour la fonction de Shq1p, néanmoins, en dépit de sa présence dans de nombreuses co-chaperonnes d'Hsp90 (e.g. Sgt1, Sba1, p23, Pih1p), aucune interaction *in vitro* n'a été détectée entre le domaine CS de Shq1p et Hsp90 de levure (Singh *et al.*, 2009). De plus, même si sa présence au sein de la protéine Shq1p améliore son association avec Cbf5p, ce domaine n'est pas suffisant pour permettre l'interaction (Godin *et al.*, 2009). Le domaine C-terminal, ou SSD, est quant à lui capable de s'associer seul à la protéine Cbf5p. En 2011, deux équipes ont résolu simultanément les structures tridimensionnelles des complexes formés entre le domaine SSD de Shq1p et Cbf5p (Walbott *et al.*, 2011) et entre Shq1p_SSD-Cbf5p-Nop10p-Gar1p (Li *et al.*, 2011b) chez *S. cerevisiae* (Figure 27). L'interface d'interaction entre le facteur d'assemblage Shq1p et la protéine Cbf5p implique le domaine PUA de l'enzyme mais aussi son extension CTE, spécifique des eucaryotes (Li *et al.*, 2011b; Walbott *et al.*, 2011). Comme mentionné auparavant (chapitre 3.1.2), cette extension C-terminale de Cbf5p n'apparait pas dans la structure cristalline du complexe Cbf5p-Nop10p-Gar1p de *S.cerevisiae* (Figure 24A). Cette absence résulte vraisemblablement de sa flexibilité et/ou d'une forte tendance à la dégradation (Li *et al.*, 2011c), néanmoins au sein du complexe avec le facteur d'assemblage cette extension adopte

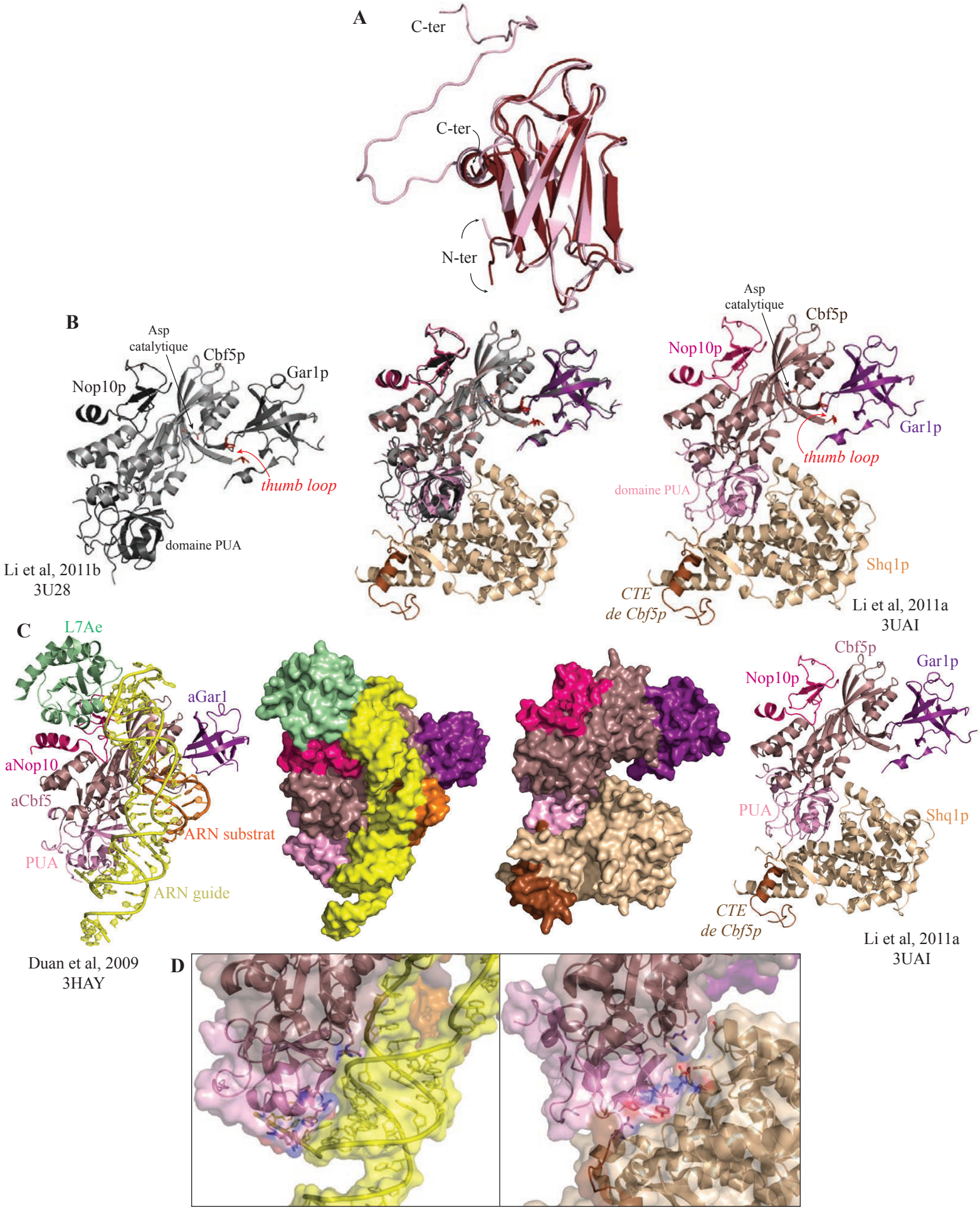


Figure 27. **Structure tridimensionnelle du facteur d'assemblage Shq1p de *S.cerevisiae***. (A) Superposition des structures du domaine N-terminal de la protéine Shq1p ou domaine CS obtenue par RMN en rose (Godin *et al*., 2009) (code PDB : 2K8Q) et par cristallographie en brun (Singh *et al*., 2009) (code PDB : 3EUD). (B) Comparaison des complexes Cbf5-Nop10p-Gar1p en gris (Li *et al*., 2011b) (code PDB : 3U28) et Cbf5p-Nop10p-Gar1p-Shq1p chez *S. cerevisiae* (Li *et al*., 2011a) (code PDB : 3UAI). (C) Comparaison entre les complexes aCbf5-aNop10-aGar1-L7Ae-ARN chez *P. furiosus* (Duan *et al*., 2009) (code PDB 3HAY) et Cbf5p-Nop10p-Gar1p-Shq1p chez *S. cerevisiae* (Li *et al*., 2011a) (code PDB : 3UAI) en mode ribbon et surface moléculaire. (D) Comparaison de la zone d'interaction entre la protéine aCbf5 et l'ARN guide (à gauche) et entre la protéine Cbf5p et la protéine Shq1p (à droite). Le code couleur est le même que (C).

une structure étendue sur la surface de Shq1p. Ce domaine CTE est composé d'une hélice α (α3') bordée de deux boucles. La boucle reliant le domaine PUA à l'hélice α3' contacte directement Shq1p. De plus, l'interaction de Shq1p sur Cbf5p est incompatible avec une association supplémentaire de Cbf5p avec le snoRNA. En effet Shq1p se fixe au niveau du domaine PUA de Cbf5p par le biais de contacts semblables à ceux retrouvés entre Cbf5p et l'ARN guide (Figure 27). En effet, il y existe une superposition parfaite entre les résidus de Cbf5p impliqués dans l'association avec Shq1p et ceux de la protéine aCbf5 d'archaea impliqués dans l'interaction avec l'ARN. Les liaisons électrostatiques avec les groupements phosphates de la chaine principale de l'ARN guide sont remplacées par des liaisons électrostatiques avec des résidus acides glutamiques et aspartiques de Shq1p, et les liaisons hydrogène sur les sucres sont remplacées par des liaisons hydrogène sur les chaînes principales et latérales de la protéine (Li *et al.*, 2011b; Walbott *et al.*, 2011). Ces résultats illustrent de façon remarquable la dépendance des certaines protéines eucaryotes vis-à-vis des facteurs d'assemblages, puisque la nature instable de Cbf5p peut être compensée par sa prise en charge par la protéine Shq1p. De plus, comme cela a été observé entre Naf1p et Gar1p, Shq1p prévient l'association non spécifique de Cbf5 avec les ARN (Godin *et al.*, 2009; Grozdanov *et al.*, 2009; Li *et al.*, 2011b; Walbott *et al.*, 2011).

2.4.2 Rôle possible du complexe S.M.N.

Il a été montré que le complexe de survie des motoneurones (S.M.N.) qui intervient dans l'assemblage de nombreuses classes de RNP comme les UsnRNP, s'associe spécifiquement à la protéine humaine GAR1, laissant suggérer un rôle de SMN dans la biogenèse des RNP à boîtes H/ACA (Jones *et al.*, 2001; Pellizzoni *et al.*, 2001; Whitehead *et al.*, 2002) (Figure 25). De plus, chez les vertébrés, les RNP H/ACA en cours d'assemblage transitent, grâce à l'adaptateur d'export des snRNA, PHAX, par les CB dans lesquels le complexe SMN est localisé (Boulon *et al.*, 2004; Pradet-Balade *et al.*, 2011). Une hypothèse envisagée est donc que l'incorporation de la protéine Gar1 soit médiée par SMN au sein des CB (Figure 25), pour revue (Matera *et al.*, 2007). Au laboratoire, A. Huttin et S. Massenet ont obtenu des arguments expérimentaux forts en faveur de cette hypothèse (communication personnelle).

2.4.3 Rôle de la chaperonne Hsp90 et du complexe R2TP

En 2008, notre équipe et celle de W. Houry ont montrés que la chaperonne ubiquitaire Hsp90 (« *Heat Shock Protein 90* ») intervenait dans l'assemblage des RNP d'eucaryotes

contenant les protéines de la famille L7Ae en particulier les snoRNP H/ACA (Nhp2p / NHP2), les snoRNP C/D (Snu13p/15.5K) et les mRNP SECIS impliquées dans la traduction des sélénoprotéines (Sbp2) (Boulon *et al.*, 2008; Zhao *et al.*, 2008). Dans la biogenèse de ces RNP, la chaperonne agirait de concert avec le complexe R2TP composé des protéines Rvb1p (Pontin ou TIP49), Rvb2p (Reptin ou TIP48), Tah1p (SPAGH) et Pih1p (NOP17) (Boulon *et al.*, 2008; Zhao *et al.*, 2005; Zhao *et al.*, 2008). Les protéines Rvb1p et Rvb2p sont des membres de la famille des AAA+ ATPases et sont connues pour intervenir également dans les fonctions d'activation ou de répression de la transcription, de transformation cellulaire, de réponse aux dommages à l'ADN et de réplication de l'ADN (Jha & Dutta, 2009). Au sein du complexe R2TP, Rvb1p et Rvb2p sont associées avec l'hétérodimère Tah1p-Pih1p. Hsp90 interagit directement avec les protéines Tah1p et Pih1p *in vitro* et cette association est également présente *in vivo* lorsque les cellules sont en phase stationnaire de croissance (Zhao *et al.*, 2008). Il a notamment été proposé que la chaperonne contrôle la stabilité de Pih1p et que son association avec cette dernière serait renforcée par la protéine Tah1 (Zhao *et al.*, 2008). Les rôles moléculaires de Hsp90 et du R2TP dans la biogenèse des snoRNP H/ACA restaient toutefois à définir. En collaboration avec l'équipe d'E. Bertrand, notre équipe a notamment montré que la protéine Rsa1p/NUFIP joue le rôle d'adaptateur entre le complexe R2TP et les pré-snoRNP. En effet, ce dernier est capable d'interagir spécifiquement avec Pih1/NOP17 et avec les membres de la famille L7Ae : Snu13p/15.5K, NHP2 et SBP2 (Boulon *et al.*, 2008). Le rôle de ces différents facteurs sera abordé plus en détail dans la partie suivante concernant la biogenèse des snoRNP à boîtes C/D sur laquelle mon travail de thèse a été plus particulièrement focalisé (cf. paragraphe 3.6).

3 Les RNP à boîtes C/D

Comme nous l'avons déjà évoqué, cette grande classe de particule RNP est responsable des 2'-O-méthylations des pré-ARNr mais aussi des UsnRNA spliceosomaux et vraisemblablement d'ARNm. Certains de ses membres comme la snoRNP U3 sont également impliqués dans la maturation nucléolytique des ARNr. Tout comme les RNP H/ACA, ces particules se composent d'un ARN guide, le s/snoRNA et d'un jeu de protéines constitutives (Tableau 2).

3.1 Structure des ARN à boîtes C/D

Les RNP à boîtes C/D sont définies par la présence de deux motifs consensus : la boîte C de séquence 5'-RUGAUGA-3' (ou R est un résidu purine) qui est localisée à proximité de l'extrémité 5' de l'ARN et la boîte D de séquence 5'-CUGA-3' localisée à proximité de l'extrémité 3' (Bachellerie *et al.*, 2002) (Figure 28). En plus, la plupart des snoRNA à boîtes C/D possèdent un second couple de boîtes C' et D' très similaire au couple C/D mais qui peut avoir dans certains cas une séquence dégénérée. Il est localisé dans la région centrale de l'ARN. Ces deux couples de boîtes sont conservés chez les archaea et les eucaryotes (Omer *et al.*, 2000). Chaque couple forme un motif structural particulier : le motif en *« K-turn »*. Ainsi les boîtes C et D s'apparient et leur appariement est prolongé par une complémentarité des deux extrémités de l'ARN (Watkins *et al.*, 2000). Les boîtes C' et D' forme une structure en *« K-turn »* chez les eucaryotes (Watkins *et al.*, 2000) et parfois en *« K-loop »* chez les archaea (Nolivos *et al.*, 2005). Le site de modification est défini par la formation d'une double hélice parfaite formée par hybridation de l'ARN substrat avec les 10 à 20 résidus situés en amont des boîtes D et/ou D' (Cavaillé *et al.*, 1996; Kiss-László *et al.*, 1996). Le résidu ciblé pour la 2'-O-méthylation est localisé de façon invariante 5 nucléotides en amont du motif conservé 5'-CUGA-3' conformément à la règle dite du « N plus cinq » (Kiss-László *et al.*, 1996; Kiss-László *et al.*, 1998). Les boîtes C' et D' sont généralement séparées par 3 à 10 nucléotides non appariés ou par une séquence de plus grande taille formant une tige-boucle interne (Kiss-László *et al.*, 1998). Chez les archaea, les sRNA à boîtes C/D sont plus petits mais présentent un couple de boîtes C'/D' plus conservé (Gaspin *et al.*, 2000; Omer *et al.*, 2000). Il est intéressant de noter que chez *Pyrococcus abyssi*, les boîtes C' et D' sont généralement séparées par une courte séquence de 3 ou 4 nucléotides, conduisant à la formation d'un motif de type *« K-loop »*, motif également retrouvé au sein des sRNA à boîtes H/ACA (Nolivos *et al.*, 2005). Chez les eucaryotes, la distance entre les motifs C/D et C'/D' varie en fonction des ARN d'une à plusieurs dizaines de résidus. En revanche, chez beaucoup d'archaea, on observe une longueur moyennes de 12 ou 13 nucléotides (Tran *et al.*, 2005). Comme nous l'aborderons par la suite, cette conservation de distance entre les deux couples de boîtes présente une grande importance structurale et fonctionnelle.

En plus de leur rôle dans le mécanisme d'action des snoRNP, les boîtes C et D sont importantes pour la biogenèse des snoRNP, la stabilité du snoRNA (Caffarelli *et al.*, 1996; Huang *et al.*, 1992; Li & Fournier, 1992), la triméthylation de leur coiffe (Speckmann *et al.*, 2000), la localisation nucléolaire (Lange *et al.*, 1998; Narayanan *et al.*, 1999; Samarsky *et al.*,

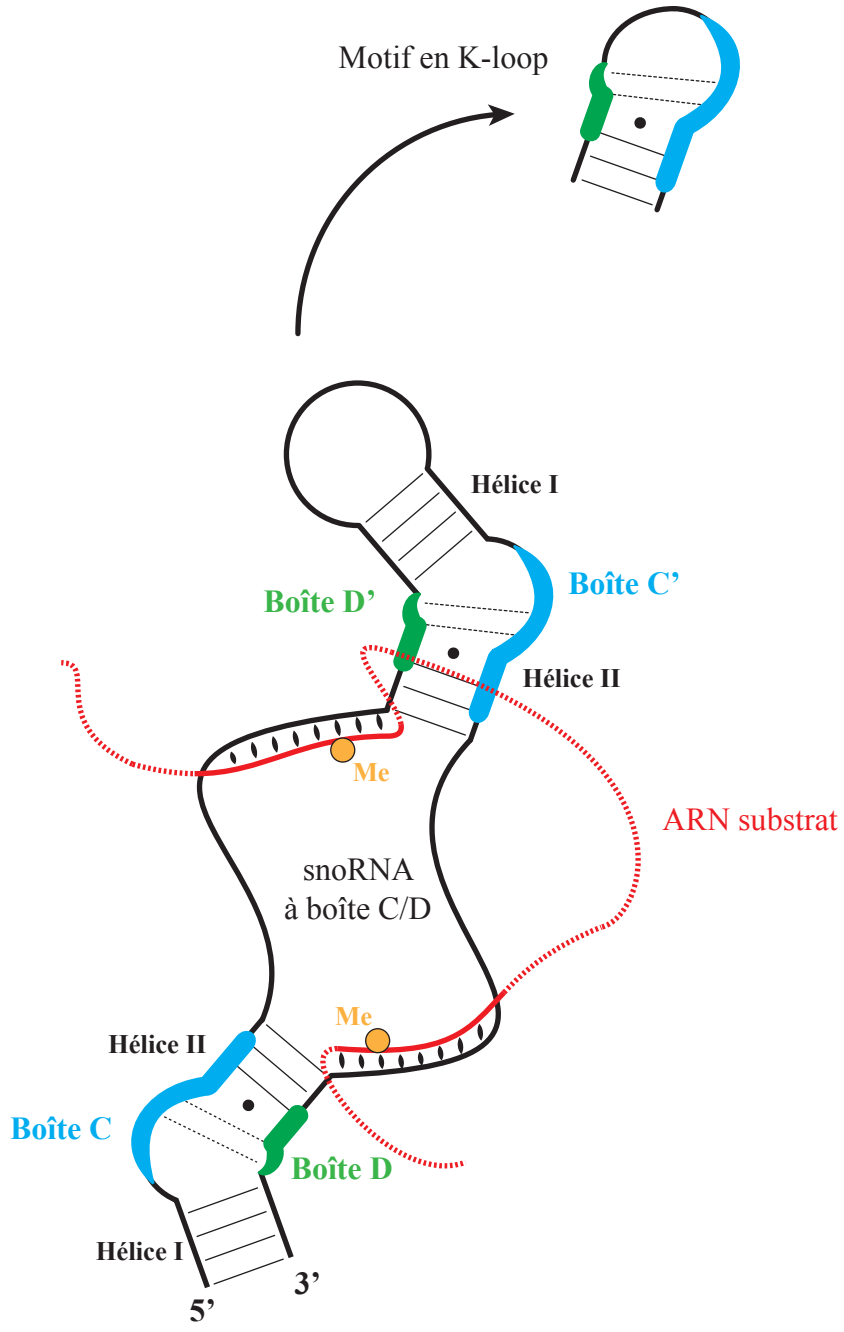


Figure 28. **Représentation schématique des s/snoRNA à boîtes C/D**. Les boîtes C, D et C', D' s'associent pour former les motifs structuraux en *« K-turn »* ou *« K-loop »* quand l'hélice I est remplacée par une courte séquence. L'hybridation de l'ARN substrat avec l'ARN guide a lieu en amont des boîtes D et D' et permet de cibler spécifiquement le ribose du résidu à 2'-O-méthylé. Le résidu cible est situé à cinq paires de bases en amont des boîtes D et D'.

1998; Speckmann *et al.*, 1999) et l'assemblage des particules (Caffarelli *et al.*, 1996; Cavaillé & Bachellerie, 1996; Watkins *et al.*, 1996; Xia *et al.*, 1997).

3.2 Le motif ARN en *« K-turn »*

Le motif en « *K-turn* » des snoRNA reconnu par les protéines de la famille L7Ae a été découvert en 2000 par notre équipe en collaboration avec l'équipe de Lührmann lors de l'étude structurale de la Snu13p. En parallèle, plusieurs motifs en « *K-turn* » ont été mis en évidence dans les ARN (Klein *et al.*, 2001). Notre équipe avec celle de Lührmann avaient alors montré que le motif en « *K-turn* » est requis pour fixer Snu13p de levure et son homologue humain la 15,5K (Vidovic *et al.*, 2000; Watkins *et al.*, 2000). Notre équipe a alors proposé que Snu13p soit l'élément nucléateur de l'assemblage des snoRNP eucaryotes (Marmier-Gourrier *et al.*, 2003).

En parallèle, la reconstitution *in vitro* des sRNP C/D d'archaea a montré que la protéine L7Ae homologue de Snu13p est requise pour initier l'assemblage des RNP sur les motifs C/D et C'/D' des sRNA C/D (Kuhn *et al.*, 2002; Omer *et al.*, 2002). D'une manière générale la famille des protéines L7Ae partagent la propriété de se lier à des motifs en *« K-turn »* ou apparenté. Le motif en « *K-turn* » se replie selon une structure en hélice-boucle interne-hélice interne d'une quinzaine de nucléotides environ (Figure 29). La boucle interne est asymétrique et se compose de 3 résidus non appariés de séquence 5'-NRU-3' côté 3' (aux positions 1, 2 et 3). Cette boucle est flanquée de l'hélice I, comportant au moins deux paires de bases G-C et de hélice II comportant deux paires de bases non Watson-Crick A•G et G•A ou « *sheared pairs* » du côté de la boucle interne. Les hélices I et II sont rapprochées dans l'espace car la chaîne phosphodiester de la boucle interne forme un angle d'environ 60° (Figure 29A). Ce coude est à l'origine du nom du motif « *kink-turn* » ou « *K-turn* ». Les 2 premiers résidus non appariés (en position 1 et 2) participent respectivement à la formation d'empilements de bases avec les hélices I et II, tandis que le troisième résidu non apparié qui est toujours une uridine dans le cas des motifs reconnus par les protéines de la famille L7Ae est projeté à l'extérieur du motif vers le solvant. Ce résidu vient se loger dans une poche formée par la protéine. Cette conformation particulière du résidu U est stabilisée par l'empilement des paires de bases G•A et A•G avec les résidus de l'hélice II. Ces résidus sont requis pour la fixation de L7Ae et Snu13p/15.5K *in vitro* et pour la fonction de la snoRNP *in vivo* (Vidovic *et al.*, 2000; Watkins *et al.*, 2000).

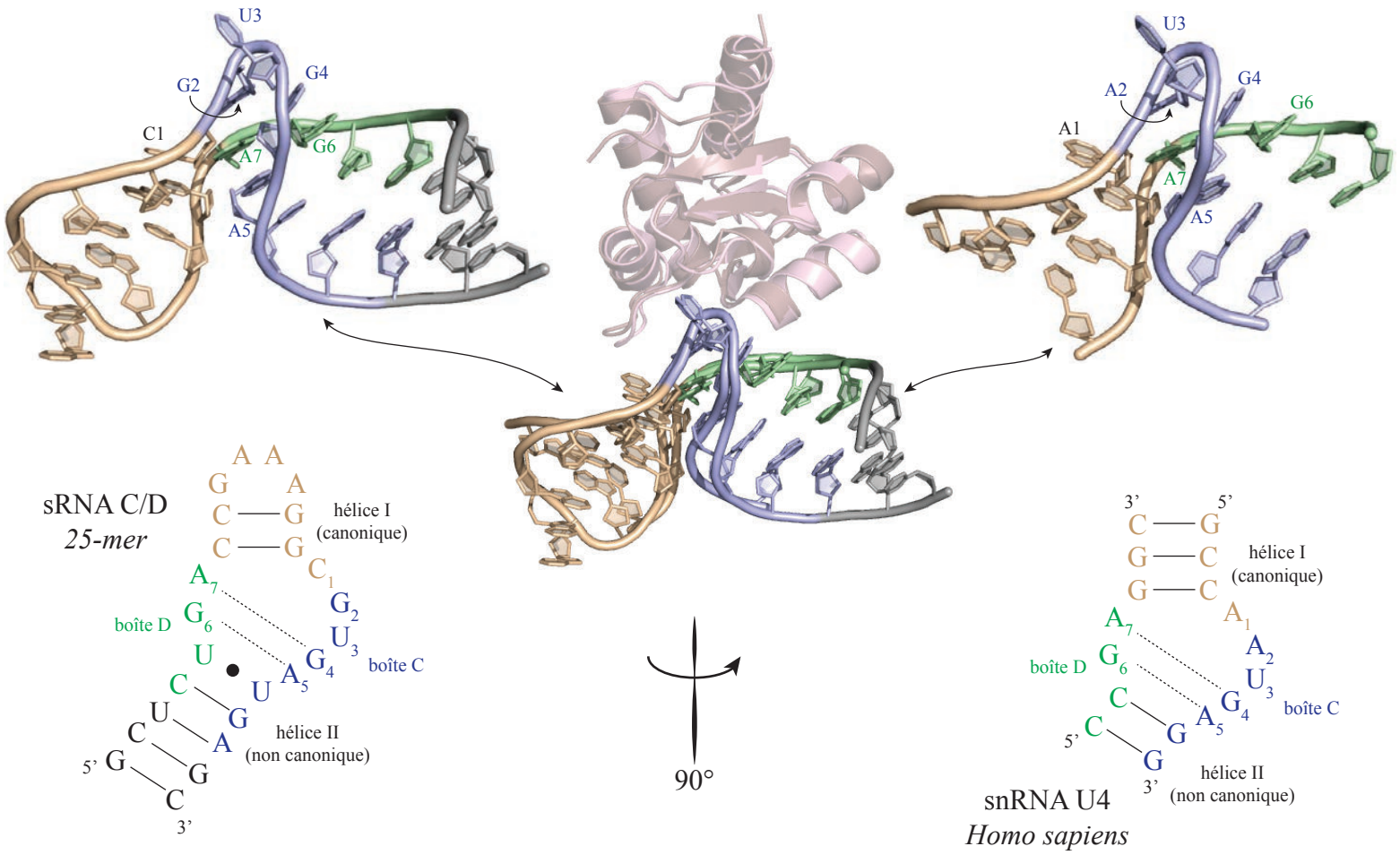


Figure 29. **Structure tridimensionnelle du motif en *« K-turn ».*** (A) Vue de face et (B) après une rotation de 90° des complexes L7Ae-sRNA C/D chez *Archaeoglobus fulgidus* à gauche (Moore *et al.*, 2004), code PDB : 1RLG et 15.5K-snRNA U4 chez *Homo sapiens* à droite (Vidovic *et al.*, 2000), code PDB 1E7K. Superposition des deux structures tridimensionnelle au centre. Les protéines L7Ae et 15.5K sont mises en transparence. La séquence, les boîtes C (bleu) et D (vert) ainsi que les hélices des ARN sont représentées.

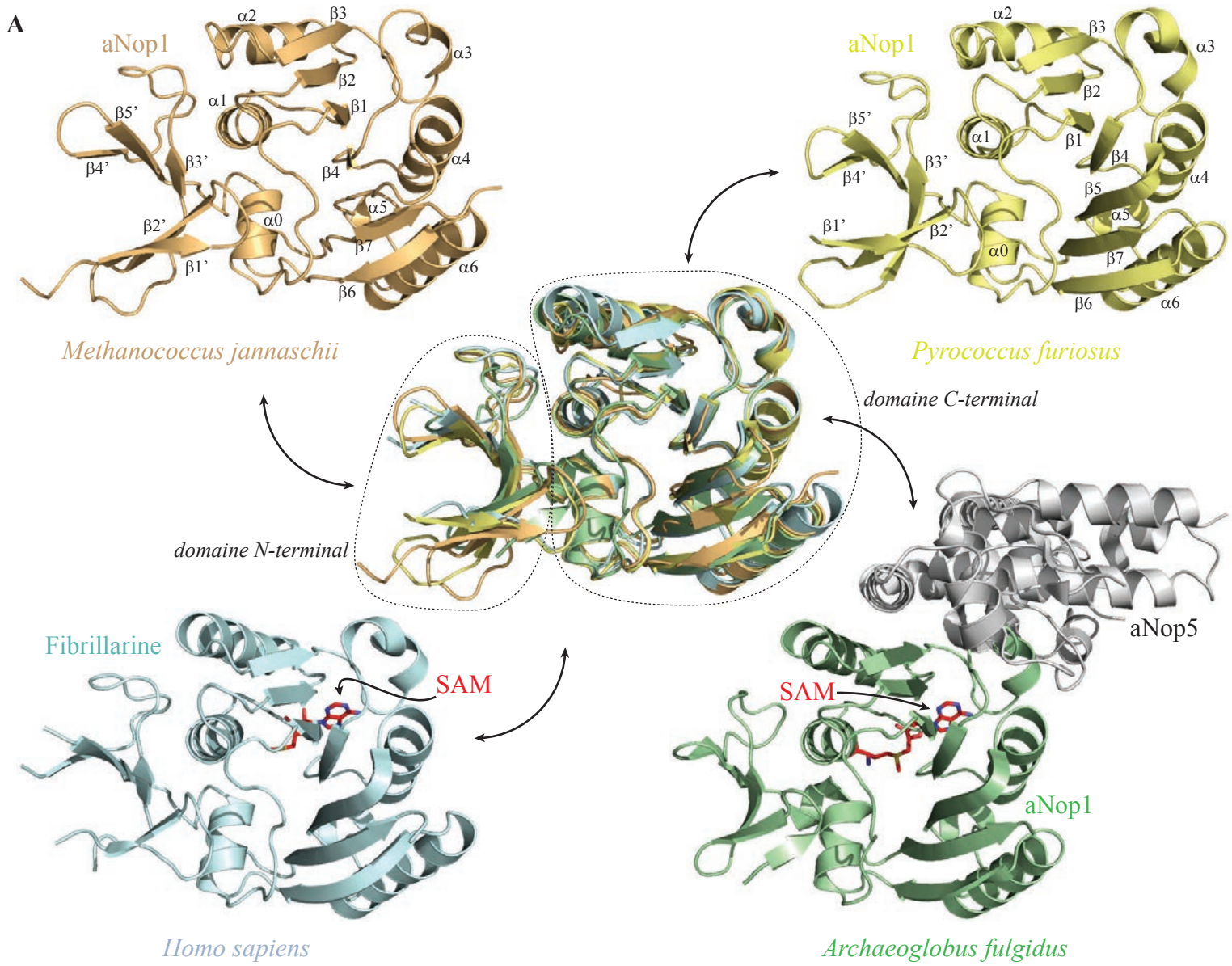
Au laboratoire N. Marmier-Gourrier a montré qu'alors que L7Ae peut reconnaître à la fois les « *K-turn* » et les « *K-loop* », Snu13p ne reconnaît que les motifs en « *K-turn* » (Marmier-Gourrier *et al.*, 2003).

3.3 Structure des protéines composant les RNP à boîtes C/D

Tout comme les RNP H/ACA, ces particules ont fait l'objet de nombreuses études ces vingt dernières années. Bien que de nombreuses données fonctionnelles soient connues chez les eucaryotes, les premières données structurales ont été obtenues pour les protéines d'archae. Ceci pour les mêmes raisons qu'évoquées précédemment : les protéines eucaryotes présentent des domaines additionnels constitués de répétitions de résidus provoquant une forte instabilité de ces protéines.

3.3.1 L'enzyme aNop1/Nop1p

La Fibrillarine, également appelée Nop1p chez la levure et aNop1 chez les archaea, est responsable du transfert du groupement méthyle sur le ribose du nucléotide substrat (Aittaleb *et al.*, 2003; Wang *et al.*, 2000). Cette protéine est essentielle à la viabilité cellulaire et à la production des ARNr chez les archaea et les eucaryotes (Amiri, 1994; Jansen *et al.*, 1991; Tollervey *et al.*, 1991; Tollervey *et al.*, 1993). Une étude réalisée chez *S. cerevisiae* a montré que son absence inhibe la méthylation des pré-ARNr et affecte tout particulièrement la production de l'ARNr 18S en inhibant les clivages précoces aux sites A_0, A_1 et A_2 (Tollervey *et al.*, 1991). Elle conduit également à une diminution du taux de nombreux snoRNA à l'exception du snoRNA U3 (Lafontaine & Tollervey, 2000; Tollervey *et al.*, 1991). Sa déplétion chez l'homme est associée à une diminution spécifique du taux de snoRNA C/D (Watkins *et al.*, 2004). La structure 3D de aNop1 libre de l'archae *Methanococcus jannaschii* (Figure 30A) (Wang *et al.*, 2000) met en évidence une forte homologie structurale avec les enzymes connues pour catalyser le transfert d'un groupement méthyl à partir du cofacteur AdoMet (S-AdenosylMethionine ou SAM) sur une variété de substrat. Depuis, la structure de la protéine NOP1 humaine a également été résolue respectivement avec la protéine aNop5 et libre chez deux autres archaea : *Archaeaglobus fulgidus* (Aittaleb *et al.*, 2003) et *Pyroccocus furiosus* (Deng *et al.*, 2004). La structure 3D de la protéine NOP1 humaine à l'état isolé a aussi été établie (Wu *et al.*). La protéine eucaryote possède un domaine GAR à son extrémité N-terminal qui n'est pas présent chez les archaea. Toutefois, les structures 3D des domaines catalytiques de aNop1 et de son homologue chez l'homme sont très similaires (Figure 30A). Leur domaine catalytique localisé en C-terminal est similaire à ceux d'autres



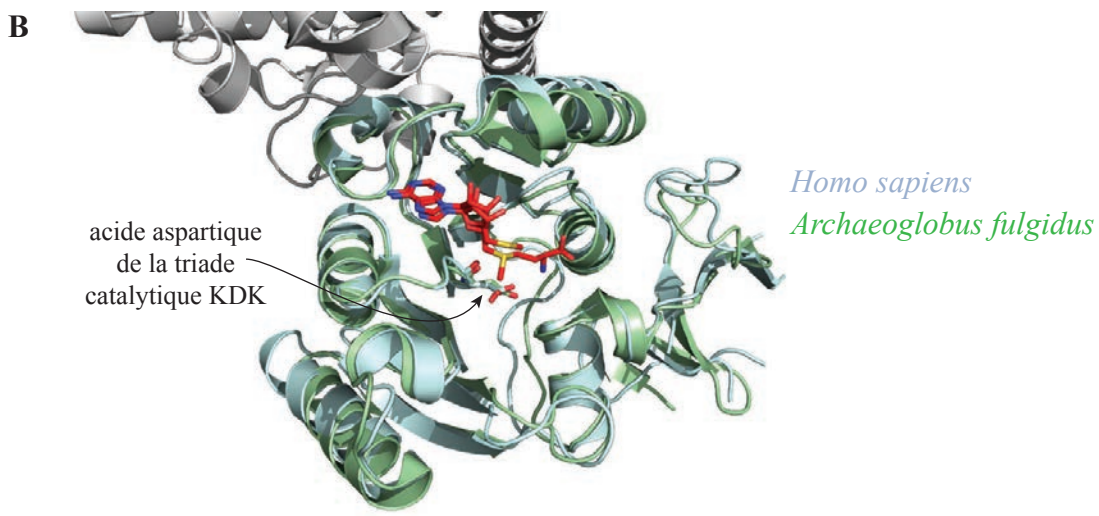


Figure 30. **Structures tridimensionnelles des protéines aNop1/Fibrillarine.** (A) Superposition des structures de aNop1/Fibrillarine de *Methanococcus jannaschii* (Wang *et al.*, 2000), code PDB : 1FBN, *Pyrococcus furiosus* (Deng *et al*., 2004), code PDB : 1PRY, *Archaeoglobus fulgidus* en présence de la protéine aNop5 en grisé (Aittaleb *et al*., 2003), code PDB : 1NT2 et *Homo sapiens* (Wu et al., non publié), code PDB : 2IPX. (B) Vue du site catalytique de l'enzyme aNop1/Fibrillarine d'*A. fulgidus* et d'*H. sapiens* avec le cofacteur SAM représenté en tige-bâton rouge ainsi que l'acide aspartique catalytique D237/D133.

méthyltransférase AdoMet-dépendantes. Ces domaines comportent un cœur de feuillet β à sept brins flanqué de part et d'autre par trois hélices α (Figure 30A). Le domaine N-terminal de aNOP1 et de son homologue humaine NOP1 adopte une structure composée de cinq brins β. Ce domaine possède une certaine variabilité structurale chez les archaea en accord avec la variabilité de leurs séquences. Des variations conformationnelles sont aussi observées entre les différentes structures de aNop1 et NOP1 au niveau du site de fixation du cofacteur SAM. Ces variations pourraient être générées par la fixation du cofacteur ou l'interaction avec les protéines aNop5/Nop56_58 (Aittaleb *et al.*, 2003). L'activité de aNOP1/NOP1 semble reposer sur une triade catalytique K-D-K conservée et localisée dans la poche de fixation du substrat (Feder *et al.*, 2003) (Figure 30B). L'acide aspartique strictement conservé de ce motif est important pour l'activité de la sRNP d'archae *in vitro* et a été proposé comme responsable de la déprotonation du groupement 2'-OH nucléophile du ribose et/ou de la stabilisation du cofacteur SAM (Aittaleb *et al.*, 2004). *In vitro*, la protéine NOP1 seule n'interagit pas, ou très peu, avec les snoRNA à boîtes C/D (Fatica *et al.*, 2000; Omer *et al.*, 2002), tandis que des expériences de pontage aux UV sur des snoRNP ont montré des contacts directs entre l'enzyme et l'ARN au sein de la RNP (Cahill *et al.*, 2002; Granneman *et al.*, 2009; Watkins *et al.*, 1998b). De façon intéressante, aNop1 n'exerce son activité qu'au sein d'une RNP à la différence de la pseudouridine synthase aCBF5 des sRNP à boîtes H/ACA. A l'état isolé, aucune activité n'a été détectée ni pour aNop1 ni pour la NOP1 humaine (Galardi *et al.*, 2002; Omer *et al.*, 2002).

3.3.2 Les protéines aNop5/Nop56p/Nop58p

Les protéines Nop56p et Nop58p font aussi partie intégrante des particules à boîtes C/D. Elles ont été identifiées chez la levure *S. cerevisiae* par une recherche de gènes mutants synthétiques létaux avec des gènes mutés de la protéine Nop1p (Gautier *et al.*, 1997). Elles partagent une forte homologie de séquence (45% d'identité) et il a été proposé qu'elles soient issues d'un évènement de duplication de gènes qui n'aurait eu lieu que chez les eucaryotes (Gautier *et al.*, 1997). Les auteurs ont également montré que ces deux protéines ne peuvent pas se substituer l'une à l'autre, suggérant ainsi l'acquisition de fonctions distinctes. En accord avec cette idée, D. Lafontaine et D. Tollervey ont montré que la déplétion de ces protéines conduit à des phénotypes différents. L'absence de Nop58p est associée à une forte diminution du taux de snoRNA à boîtes C/D et à un défaut de production de l'ARNr 18S par inhibition des clivages aux sites A_0, A_1 et A_2 (Lafontaine & Tollervey, 1999). En ce qui concerne Nop56p, son absence à un effet moins prononcé sur la production de l'ARNr 18S et

n'affecte pas le taux des snoRNA (Lafontaine & Tollervey, 2000). Les auteurs ont également montré que la délétion de Nop58p ou Nop56p induit une diminution du taux des 2'-O-méthylation des ARNr, cependant, cet effet est bien moins prononcé que celui observé pour la délétion du gène de l'enzyme Fibrillarine (Lafontaine & Tollervey, 1999; Lafontaine & Tollervey, 2000).

Ces deux protéines eucaryotes comportent une répétition KKE/D qui n'est pas présente dans les protéines d'archaea, un domaine NOP au niveau de leur domaine C-terminal qui est aussi présent dans la protéine Prp31p de la snRNP U4, un domaine « *coiled-coil* » central « coiffé » d'un domaine « *Tip* » et enfin un domaine N-terminal (Figure 31A). Le domaine KKE/D n'est pas requis pour la fonction et la localisation des protéines Nop56p et Nop58p chez la levure (Gautier *et al.*, 1997). Il semble néanmoins contenir des déterminants nécessaires à l'interaction avec l'hyperméthylase Tgs1p (Mouaikel *et al.*, 2002) et semble jouer un rôle de signal de localisation nucléaire (NLS) et nucléolaire (NoLS) dans les protéines humaines (Pradet-Balade *et al.*, 2011). Comme cela a été évoqué dans le chapitre 2.1.1, ces motifs sont retrouvés dans la pseudouridyne synthase Cbf5p des snoRNA à boîtes H/ACA. Chez les archaea, les deux protéines Nop58p/NOP58 et Nop56p/NOP56 n'ont qu'un seul homologue : la protéine aNop5. Chez ces organismes, la formation du complexe L7Ae-sRNA permet le recrutement de la protéine aNop5 qui à son tour facilite l'association de la Fibrillarine pour former une sRNP active *in vitro* (Omer *et al.*, 2002). En 2003, l'équipe de H. Li a déterminé la structure tridimensionnelle du complexe aNop5-Fibrillarine chez *A. fulgidus* (Aittaleb *et al.*, 2003) (Figure 31). Dans ce complexe, le domaine N-terminal de aNop5 adopte une topologie α/β irrégulière et interagit par l'intermédiaire de son squelette peptidique avec celui de la Fibrillarine. Le domaine N-terminal de aNOP5 est connecté par le biais d'un motif « *coiled-coil* » au domaine C-terminal très conservé de cette protéine qui est composé d'un cœur de sept hélices α. Le motif « *coiled-coil* » correspond à deux hélices super-enroulées, il possède la faculté de dimériser et favorise la formation d'un complexe quaternaire contenant deux hétérodimères aNop5-aNop1. Basé sur la forte homologie de séquence de aNop5, Nop58 et Nop56, il a été proposé que le domaine C-terminal de Nop58 et Nop56 adopte le même repliement que celui de aNop5 (Aittaleb *et al.*, 2003). Le domaine C-terminal d'aNop5 interagit avec les résidus d'aNop1 directement impliqués dans la reconnaissance du cofacteur AdoMet (SAM). Ainsi, le résidu Arg169 d'aNop5 forme une interaction forte cation-π avec la Tyr89 d'aNop1. Le cycle phénol de cette tyrosine stabilise la partie adénosine du cofacteur SAM par interaction π-π (Figure 31B). Ceci suggère que aNop5



Figure 31. **Structures tridimensionnelles des complexes aNop5-Fibrillarine chez les archaea.** (A) Organisation de la protéine aNop5 de *Archaeoglobus fulgidus* (Aittaleb *et al.*, 2003), code PDB : 1NT2. Pour plus de clarté, laNop1 est représentée en transparence. (B) Vue du site catalytique d'aNop1 d'*A. fulgidus* avec le cofacteur SAM représenté en rouge et le réseau d'interactions stabilisant le cofacteur avec utilisation des couleurs des protéines du panneau A. (C) Comparaison des structures tridimensionnelles des complexes aNop5-aNop1 obtenues pour *A. fulgidus* (Aittaleb *et al.*, 2003), code PDB : 1NT2 et *P. furiosus* (Oruganti *et al.*, 2007), code PDB : 2NNW. Au centre la superposition des deux structures.

peut stabiliser l'association aNop1-SAM (Aittaleb *et al.*, 2003). La mutation de cette arginine dans aNop5 perturbe de façon importante la catalyse (Aittaleb *et al.*, 2004). Il y a donc un lien très fort entre la aNop1 et aNop5 permettant l'activité optimale de la particule, sans doute en facilitant la fixation du cofacteur dans le site catalytique. Le domaine C-terminal d'aNop5 présente le plus grand degré de conservation de séquence de cette protéine. Néanmoins, au sein de la structure cristalline qui a été établie, ce domaine n'est impliqué ni dans des interactions protéines-protéines, ni dans des interactions protéine-SAM. La distribution des potentiels électrostatiques révèle une zone chargée positivement localisée au niveau de ce domaine C-terminal. Aussi, les auteurs suggèrent que ce domaine C-terminal est impliqué dans l'interaction avec le sRNA par le biais des charges négatives de son squelette phosphodiester. La mutation d'un résidu hautement conservé du domaine C-terminal, l'Arg224, est suffisante pour perturber l'interaction entre le sRNA et le complexe aNop5-Fibrillarine (Aittaleb *et al.*, 2003). L'équipe de J.P. Bachellerie a montré par des expériences de protections que aNop5 établit des contacts directs avec les boîtes C/D et C'/D' des sRNA (Bortolin *et al.*, 2003). Chez les archaea, la protéine aNop5 joue donc un rôle central dans la particule en établissant des contacts avec le sRNA, en recrutant l'enzyme et en modulant son activité, mais aussi en s'auto-dimérisant afin d'établir un contact physique entre les particules assemblées sur les deux couples de boîtes du sRNA, ce qui permet de former une architecture symétrique (Aittaleb *et al.*, 2003) (Lin *et al.*, 2011). Cette hypothèse est confortée par des données de mutagenèse, puisque l'activité de transfert du groupement méthyl et réduite dans le cas de mutants de aNop5 ayant perdu la faculté de dimériser et s'associant donc au sRNA sous la forme d'un seul hétérodimère aNop5-aNop1 (Rashid *et al.*, 2003; Zhang *et al.*, 2006). Tout ceci peut expliquer pourquoi aNop1 n'est pas active seule sur un ARN.

Plus récemment, l'équipe de Hong Li a déterminé la structure tridimensionnelle de l'hétérodimère aNop5-aNop1 de *P. furiosus* (Oruganti *et al.*, 2007) (Figure 31C). Les résultats obtenus montrent également une organisation symétrique du complexe sous la forme aNop1-aNop5-aNop5-aNop1 mais dans laquelle aNop5 adopte une autre conformation qui modifie considérablement la position du centre méthyl-transférase d'aNop5 au sein du complexe. Malgré la présence d'une insertion de trois hélices α au sein du domaine N-terminal, l'interface entre aNop5 et aNop1 et similaire à celle observée chez *A. fulgidus*, suggérant une conservation du mode d'association au sein de tous les complexes aNop5-aNop1 (Oruganti *et al.*, 2007). De même, malgré une insertion de 56 acides aminés chez *P. furiosus* au sein du domaine « *coiled-coil* » médiant la dimérisation (Figure 31C), ce domaine adopte le même

repliement que chez *A. fulgidus* et le même mode d'auto-association. La position de l'enzyme et du cofacteur SAM observée dans le complexe de *P. furiosus* ne semble pas compatible avec une activité simultanée sur les deux sites actifs de la RNP. De plus, la conformation du cofacteur SAM dans la structure cristalline, en particulier son carbone ε qui porte le groupement méthyl transférable est partiellement bloqué par aNop5 ce qui empêcherait son transfert sur un duplex d'ARN (Oruganti *et al.*, 2007). Il faut donc imaginer que des transitions structurales ont lieu pour permettre la méthylation (Figure 31).

3.3.3 La protéine L7Ae/Snu13p

Le motif en *« K-turn »* est reconnu par L7Ae et ses orthologues par l'intermédiaire d'un motif conservé formant une architecture en sandwich α-β-α composé d'un cœur de 4 brins β entouré par un ensemble de 6 hélices α (Figure 32). Les résidus de l'hélice α2, du brin β1 et des boucles β1-α2 et α4-β4 établissent de nombreuses interactions avec le motif ARN *« K-turn »*. Le résidu U en position 3, projeté hors de la boucle interne de l'ARN, vient se loger dans une poche hydrophobe de la protéine et établit des liaisons hydrogènes avec un résidu acide (E61 chez l'homme) et un résidu basique (K86 chez l'homme) (Figure 32B). Ainsi, il existe une forte dépendance entre le type de repliement du motif *« K-turn »* et le mode d'interaction de l'ARN avec les protéines de la famille L7Ae.

Comme mentionné dans le paragraphe 2.2.4, la protéine Nhp2 ne possède pas cette capacité de reconnaître spécifiquement les motifs en *« K-turn »* des ARN. L'analyse de séquence de cette protéine des snoRNP H/ACA fait apparaitre une variation des résidus clés localisés au niveau du site de liaison à l'ARN (Figure 20), ce qui pourrait expliquer l'absence de reconnaissance spécifiques des *« K-turn »* par cette protéine (Hamma & Ferré-D'Amaré, 2004).

Les protéines eucaryotes 15.5K/Snu13p ont une forte spécificité pour les motifs en K-turn canoniques. En revanche, la protéine d'archae L7Ae est capable de reconnaître également les motifs en *« K-loop »*, retrouvés au sein des sRNA à boîtes C/D et H/ACA et dont l'hélice I est remplacée par une boucle terminale apicale. La comparaison structurale entre L7Ae et 15.5K/Snu13p attribue cette différence aux résidus clés à l'intérieur de la boucle α4-β4 (Charron *et al.*, 2004; Cléry *et al.*, 2007; Hamma & Ferré-D'Amaré, 2004; Marmier-Gourrier *et al.*, 2003; Oruganti *et al.*, 2005) (Figure 20). La superposition des structures de Snu13p et de L7Ae a également permis de mettre en évidence une différence structurale au niveau des boucles α2-β2 et β4-α5 conduisant à subdiviser en deux groupes les protéines de la famille

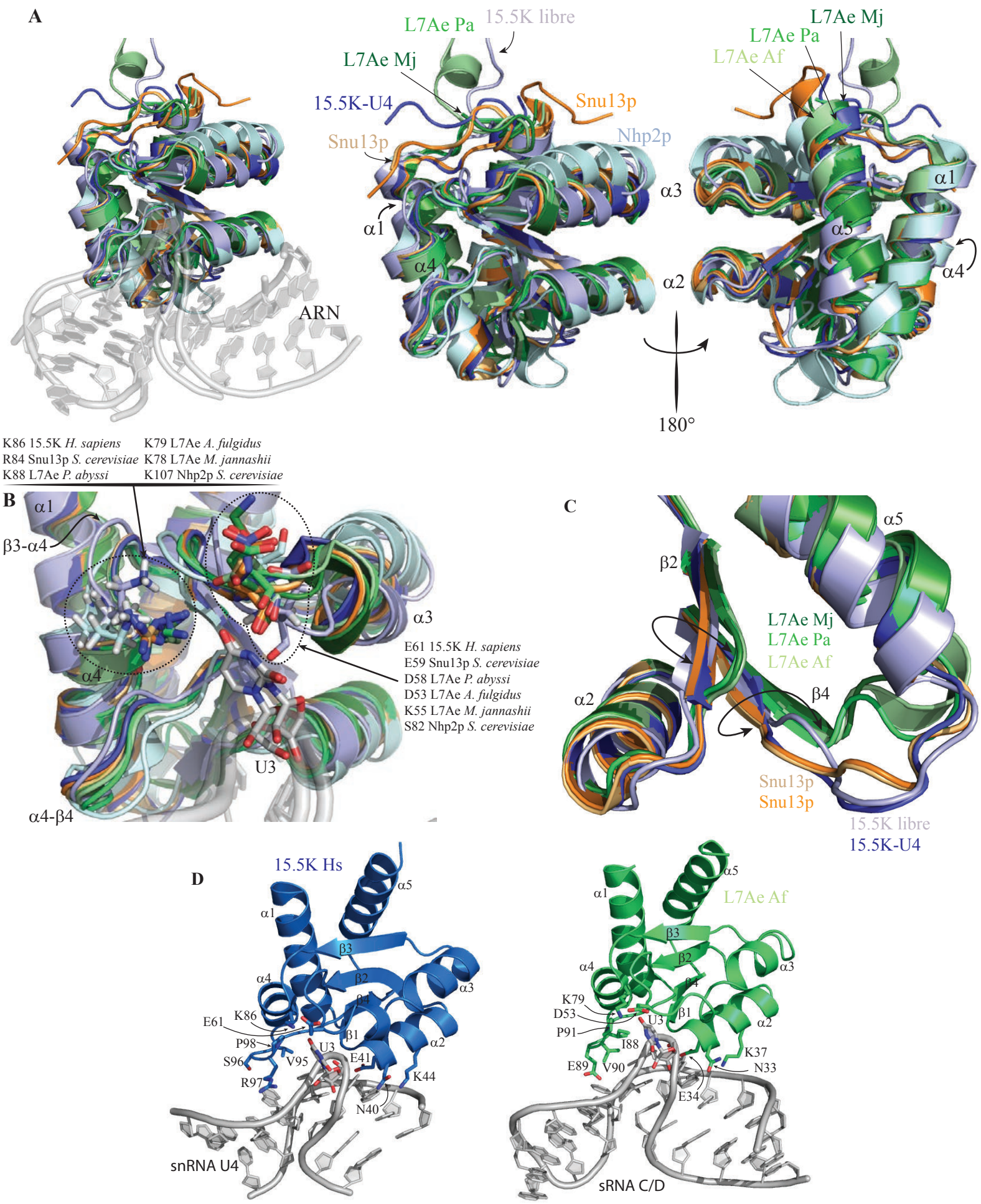


Figure 32. **Structures tridimensionnelles des protéines de la famille L7Ae**. (A) Superposition des structures de 15.5K humaine libre (code PDB : 2JNB) et en complexe avec le snRNA U4 (code PDB : 1E7K), Snu13p de *S. cerevisiae* libre (code PDB : 2ALE et 1ZWZ), L7Ae de *P. abyssi* libre (code PDB : 1PXW), de *A. fulgidus* en complexe avec le sRNA C/D (code PDB : 1RLG), de *M. jannaschii* libre (code PDB : 1RA4) et Nhp2p de *S. cerevisiae* libre (code PDB : 2LBX). (B) Résidus des protéines impliqués dans l'interaction avec le nucléotide protubérant du motif *« K-turn »*. (C) Vue de la différence structurale entre les protéines eucaryotes et d'archae repérée par les flèches. (D) Vue détaillée des résidus des motifs VSRP/IEVP et NExxK des protéines 15.5K d'*H. sapiens* en bleu et L7Ae de *A. fulgidus* en vert en interaction avec le snRNA U4 et sRNA C/D respectivement.

L7Ae, d'une part les protéines L7Ae d'archaea et leurs homologues eucaryotes d'autre part (Oruganti *et al.*, 2005) (Figure 32C). Bien que ces régions ne soient pas directement impliquées dans l'interaction avec l'ARN, les auteurs proposent que cette différence influence indirectement l'association de chaque protéine à l'ARN en modulant ses propriétés d'interaction. Chez les eucaryotes, la boucle β4-α5 possède une insertion de 2 acides aminés qui maintiennent l'extrémité C-terminale du brin β4 et qui conduit à la formation de liaisons hydrogènes supplémentaires entre les chaînes principales des brins β2 et β4. Ainsi, le brin β2 des protéines Snu13p et 15.5K est plus long du côté N-terminal que chez leurs homologues d'archaea (Figure 32). De plus, la stabilisation inter-brins β contribue à stabiliser l'extrémité C-terminale de l'hélice α2 dans les protéines eucaryotes ce qui limiterait ses mouvements lors de l'association avec l'ARN et donc réduirait la spécificité d'interaction des protéines eucaryotes avec l'ARN. Ne possédant pas ce réseaux de stabilisation, l'hélice α2 de L7Ae possèderait une flexibilité structurale plus importante et donc une plus large spécificité de substrat (Moore *et al.*, 2004; Oruganti *et al.*, 2005) d'où leur capacité à reconnaître les motifs en « *K-loop* ».

Des études physico-chimique par FRET (Goody *et al.*, 2004) et par dichroïsme circulaire (Suryadi *et al.*, 2005), ainsi que des expériences d'empreinte (Marmier-Gourrier *et al.*, 2003) ont permis de montrer que l'ARN libre n'est que partiellement sous forme de *« K-turn »*. La structure adopte une conformation *« kink »* stable que par le biais de son association avec la protéine ou en présence d'une concentration importante en cations divalents comme le magnésium.

Contrairement à l'ARN, les protéines L7Ae et 15.5K/Snu13p adoptent la même conformation globale qu'elles soient à l'état libre (Charron *et al.*, 2004; Oruganti *et al.*, 2005; Suryadi *et al.*, 2005) ou liées à l'ARN (Hamma & Ferré-D'Amaré, 2004; Moore *et al.*, 2004; Nottrott *et al.*, 1999). Ces observations suggèrent que L7Ae, 15.5K et Snu13p adoptent un repliement fixe capable de sélectionner la conformation approprié de l'ARN en solution. Par conséquent, le complexe initiateur des sno/s RNP C/D formé par L7Ae, 15.5K ou Snu13p est probablement important pour stabiliser le sno/s RNA dans une conformation favorable au recrutement des autres protéines « cœur ».

3.4 Structure des particules à boîtes C/D

Contrairement aux particules H/ACA, aucune donnée structurale de snoRNP eucaryotes même partielle n'est disponible. A l'heure actuelle, seule la structure de particules d'archae est disponible ce qui apporte néanmoins des informations précieuses transposables au système eucaryote du fait de la forte homologie de séquences.

L'assemblage des particules C/D en extrait cellulaire nécessite l'association préalable de la protéine Snu13p/15.5K sur le motif ARN en *« K-turn »* partiellement formé par les boîtes C et D (Watkins *et al.*, 2002). C'est donc la fixation de la protéine de la famille L7Ae qui est l'élément nucléateur de la formation de la particule (Marmier-Gourrier et al., 2003). Chez les archaea, ce système est conservé puisque les protéines aNop5 et aNop1/Fibrillarine ne peuvent être intégrées à la particule qu'en présence de la protéine L7Ae (Omer *et al.*, 2002). La dépendance vis-à-vis de L7Ae est moins forte pour l'assemblage des particules H/ACA d'archaea, où la protéine aCbf5 seule, peut initier l'assemblage (Baker *et al.*, 2005; Charpentier *et al.*, 2005). Elle peut même dans certains cas se suffire à elle-même pour modifier un ARNt (Roovers *et al.*, 2006). Ainsi, nous pourrions imaginer que les snoRNA à boîtes H/ACA eucaryotes aient perdu leur motif en *« K-turn »* au cours de l'évolution. Chez les eucaryotes, Nop58p et Nop1p sont recrutées indépendamment sur le snoRNA, alors que l'association plus tardive de Nop56p sur l'ARN nécessite la présence de Nop1p (Lafontaine & Tollervey, 2000). En dépit de leur homologie de séquence, les protéines Nop56p et Nop58p ont donc évolué vers un mode d'assemblage distinct.

Les RNP à boîtes C/D ont la particularité de posséder deux couples de boîtes sur le même ARN : le couple C/D et C'/D'. Chez les archaea, plusieurs études ont mis en évidence la présence d'un jeu de protéine L7Ae-aNop5-aNop1 par couple de boîtes (Omer *et al.*, 2002; Tran *et al.*, 2003). De plus, les réactions de 2'-O-méthylation, catalysées au niveau des motifs C/D et C'/D' sont les plus efficaces lorsque deux trimères sont retrouvées au sein d'une même particule (Rashid *et al.*, 2003; Tran *et al.*, 2003). A partir de ces données, l'équipe de Hong Li a proposé un modèle d'assemblage symétrique chez les archaea (Aittaleb *et al.*, 2003). Dans ce modèle, un jeu complet des trois protéines L7Ae-aNop5-Fibrillarine est associé par couple de boîte et ses deux « RNP » interagissent physiquement par le biais de la dimérisation des domaines « *coiled-coil* » des protéines aNop5. Ce mode d'assemblage serait induit par des contraintes structurales au niveau de l'ARN, notamment la distance qui sépare les deux couples de boîtes. Ceci pourrait ainsi expliquer la conservation d'une douzaine de nucléotides

entre les deux motifs (Tran *et al.*, 2005) et le fait que modifier la longueur de ce domaine en altère la fonction.

De façon remarquable, ces trois dernières années, plusieurs équipes ont réussi à obtenir des données structurales sur les s/snoRNP à boîtes C/D. Le premier modèle a été obtenu par K. Ye en 2009 chez l'archae *Sulfolobus solfataricus* (Ye *et al.*, 2009). Il se compose d'un ARN guide ne comportant qu'un seul couple de boîtes C/D, de la protéine aNop5 sans sa partie C-terminale (motifs KKE/D), aNop1 et la protéine L7Ae (Figure 33). La structure cristallographique obtenue ne correspond malheureusement pas à la conformation active, l'enzyme contenant le cofacteur SAM étant située à l'opposé du site de fixation de l'ARN guide. Le domaine N-terminal de aNop5 adopte le même repliement et la même surface d'interaction avec aNop1 que celui de aNop5 de *P. furiosus* observé par une autre équipe (Oruganti *et al.*, 2007), mais ces deux structures divergent par le positionnement spatial du domaine N-terminal de aNop5 et aNop1 par rapport au domaine « *coiled-coil* ». L'interaction observée entre la protéine L7Ae, le domaine C-terminal (ou domaine NOP) de aNop5 et l'ARN guide est en revanche très similaire à celle observée dans la structure du complexe de la snRNA U4 humaine 15.5K-PRP31-U4 5'SL (Liu *et al.*, 2007) comme le montre la figure 33B.

Quelques mois plus tard, une étude par microscopie électronique a permis d'obtenir un modèle tridimensionnel d'une sRNP de *Methanocaldococcus jannaschii* catalytiquement active et assemblée à partir de protéines recombinantes et du sRNA sR8 transcrit *in vitro* (Bleichert *et al.*, 2009). Les premiers résultats étaient surprenants car ils indiquaient que la particule obtenue avait une masse bien supérieure à celle attendue pour un complexe à deux jeux de protéines aNop5-aNop1-L7Ae. Le placement des structures cristallines au sein de l'enveloppe déterminée par microscopie électronique a tout d'abord révélé qu'il pouvait y avoir non pas un, mais deux hétérotétramères aNop5-aNop1. Par la suite une analyse d'acides aminés quantitative montrait une stœchiométrie de 1 :1 :1 entre les protéines L7Ae, aNop5 et aNop1. Quatre molécules de L7Ae pouvaient donc être incluses dans la particule. Ainsi, l'enveloppe observée en microscopie électronique de la sRNP contenait 4 copies de chaque protéine, suggérant la présence de deux molécules de sRNA avec un jeu de boîtes C/D et C'/D' chacune dans la particule. Le placement de ces deux molécules de sRNA au sein de l'enveloppe a montré que ces ARN s'associaient au complexe aNop5-Fibrillarine (initialement positionné dans l'enveloppe) dans une direction différente car orthogonale à celle du domaine « *coiled-coil* » (Bleichert *et al.*, 2009). Les auteurs ont ainsi proposé une

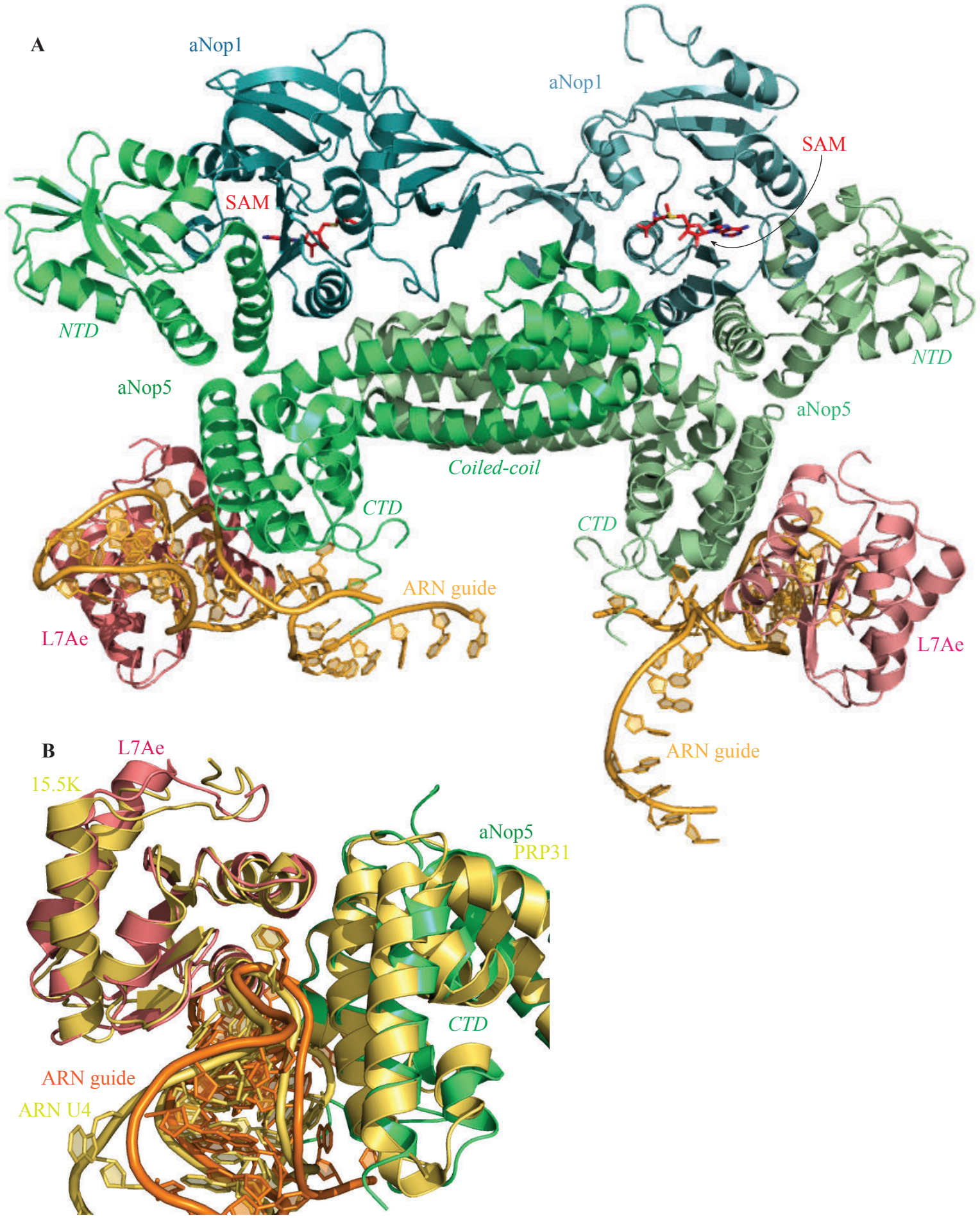
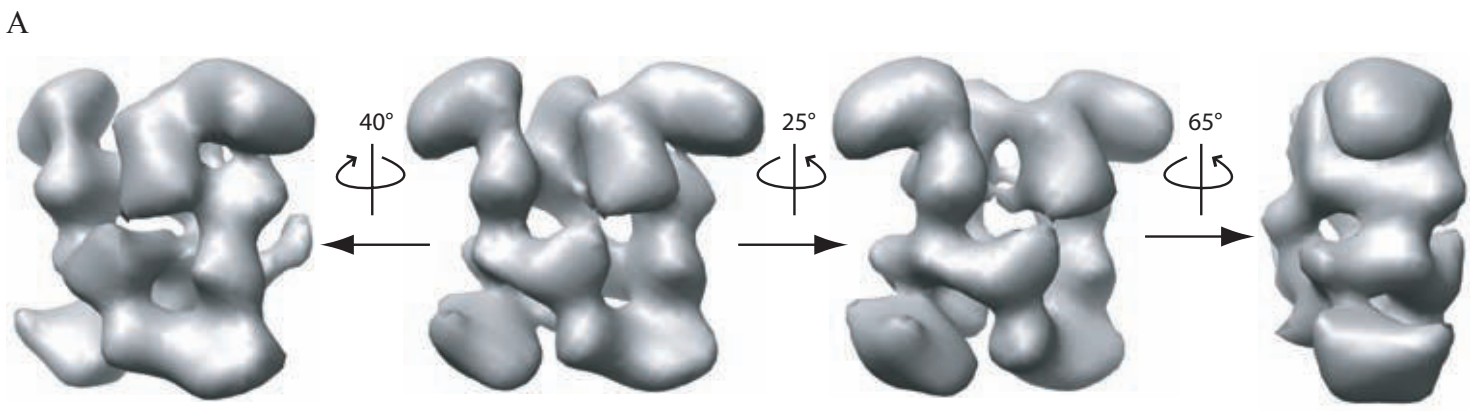


Figure 33. **Structure tridimensionnelle du complexe formé par un ARN guide en interaction avec le tétramère aNop5-aNop1-L7Ae de *Sulfolobus sulfataricus***. (A) Organisation de la particule C/D de *Sulfolobus sulfataricus* en forme de papillon (Ye *et al*., 2009), code PDB : 3ID5. Les deux complexes s'associent par la dimérisation de la protéine aNop5 via son domaine *«coiled-coil»*. (B) Superposition du module protéique de liaison à l'ARN avec celui du complexe 15.5K-PRP31-5'SL du UsnRNAU4 (Liu *et al*., 2007), code PDB : 2OZB.

organisation différente du modèle conventionnel : le modèle de la di-sRNP, contenant deux molécules de sRNA et quatre copies de chaque protéine cœur (Figure 34). Au sein de ce modèle, les protéines aNop5 ne dimérisent pas au sein d'une même RNP mais avec une protéine aNop5 assemblée sur la seconde molécule d'ARN. On comprend mal comment dans cette organisation une variation de la longueur des domaines coiled-coil puisse altérer la fonction de la particule.

En 2010, une structure cristallographique d'une demi-sRNP de *Pyrococcus furiosus* en association avec son substrat a apporté des arguments en faveur du modèle de la di-sRNP (Xue *et al.*, 2010) (Figure 35). La structuration globale du complexe et l'orientation des domaines de chaque composants de la RNP sont similaires à ce qui a été observé pour le complexe isolé aNop5-Fibrillarine de *P. furiosus* (Oruganti *et al.*, 2007) (Figure 35B), mais diffèrent des complexes de *A. fulgidus* (Aittaleb *et al.*, 2003) (Figure 32) et *S. sulfataricus* (Ye *et al.*, 2009) (Figure 33). La différence principale repose sur la position d'aNop1, qui résulte de la rotation du domaine N-terminal de la protéine aNop5 autour d'un peptide charnière. Toutefois, le substrat ARN reste relativement loin de la sous-unité catalytique d'aNop1 la plus proche. Bien que cette observation puisse résulter du positionnement incorrect du substrat lors de la cristallisation, les auteurs proposent un mécanisme de modification particulier (Xue *et al.*, 2010) compatible avec le modèle de la di-sRNP. De façon surprenante, ce mécanisme fait appel au relargage de la protéine L7Ae de la particule et à un large mouvement de l'enzyme pour que la modification puisse avoir lieu. De plus, la modification au niveau d'un motif C/D n'est pas réalisée par la Fibrillarine associée à ce motif, mais par la Fibrillarine associée au second motif du même sRNA guide (Xue *et al.*, 2010) (Figure 35).

Enfin, en 2011, l'équipe de K. Ye a déterminé la structure tridimensionnelle d'une particule C/D composée des protéines aNop5, aNop1 et L7Ae de *Sulfolobus solfataricus,* mais également d'une seule molécule d'ARN guide complète (80 nucléotides) portant les deux jeux de couples de boîtes associés à deux ARN substrats de 10 nucléotides chacun (Lin *et al.*, 2011) (Figure 36). C'est sur ce point précis que ce travail se distingue des résultats antérieurs, puisque jusqu'à présent, les structures cristallines des sRNP C/D ont été obtenues avec un « demi » ARN guide qui ne contenait qu'un seul jeu de boîtes avec des extrémités libres en association ou non avec son ARN substrat (Figures 33 et 35) (Xue *et al.*, 2010; Ye *et al.*, 2009). Or ce type de configuration, ne permettait pas de savoir si un ARN à deux couples de boîtes C/D s'associerait à un seul dimère de protéines aNop5 ou si cet ARN s'associerait à un dimère d'aNop5 (comme dans le modèle de la di-sRNP). Dans la structure obtenue par Ye et



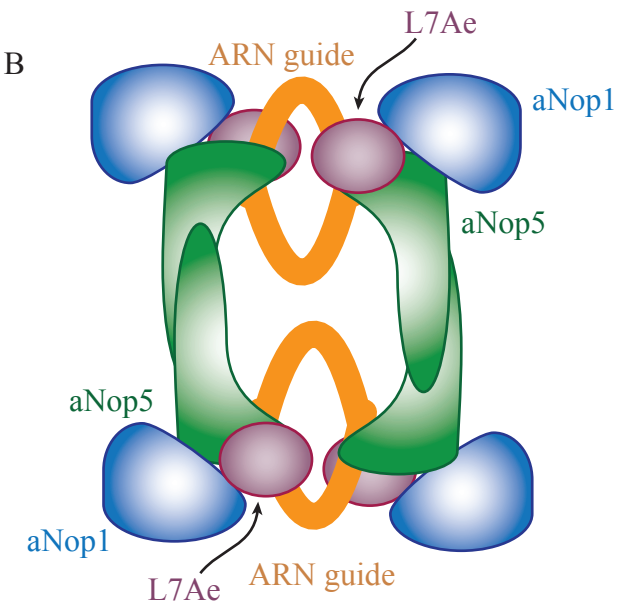


Figure 34. **Reconstitution 3D de la di-sRNP C/D de l'archae *Methanocaldococcus jannaschii* par microscopie électronique.** (A) Carte d'isodensité des volumes tridimensionnels reconstruits. (B) Modèle proposé pour la di-sRNP C/D. Le code couleur est le même que celui de la figure 33. D'après Bleichert *et al*., 2009.

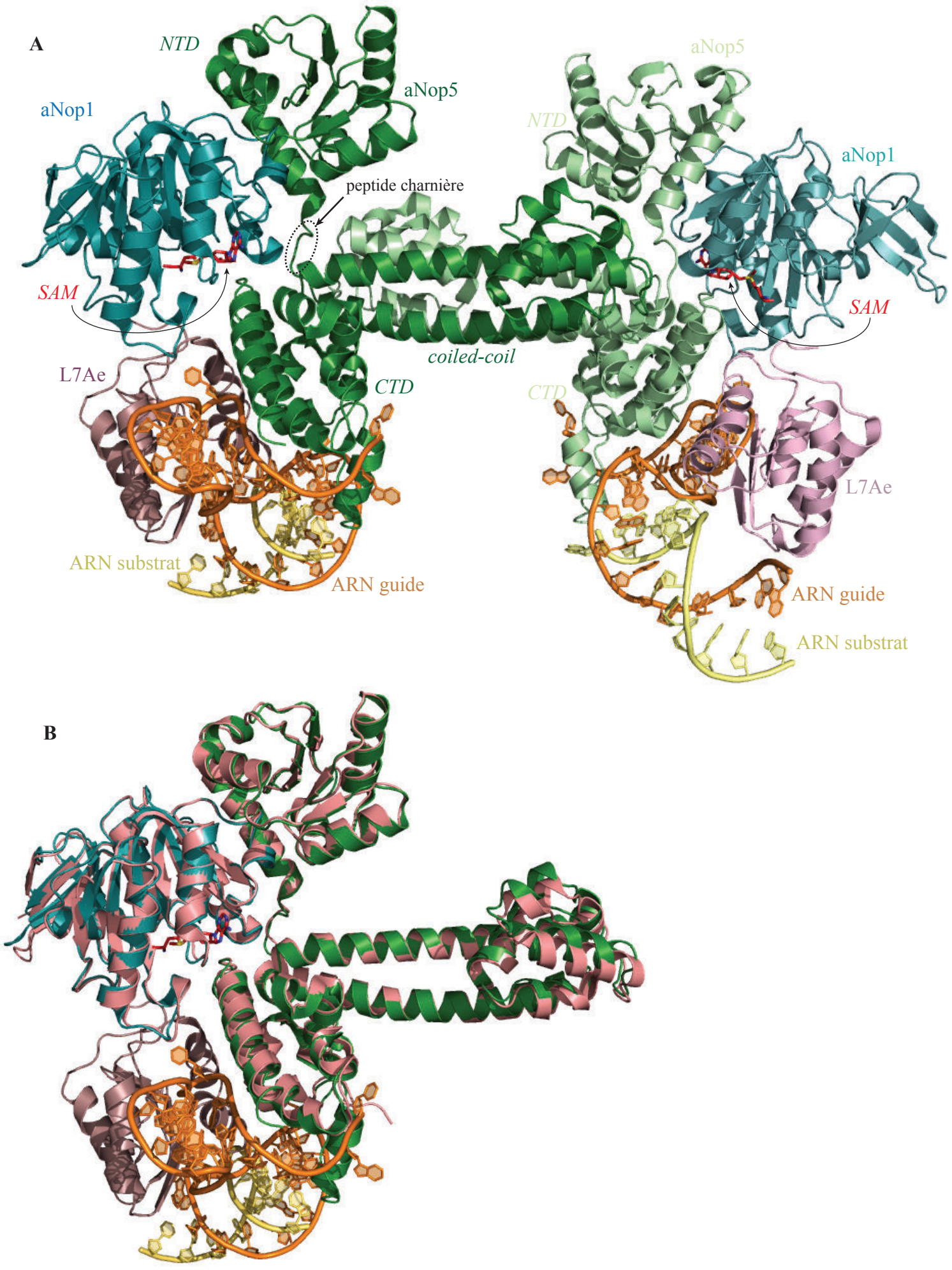


Figure 35. **Structure tridimensionnelle d'un complexe ARN guide-ARN substrat en interaction avec les protéines aNop5-aNop1-L7Ae de *Pyroccocus furiosus***. (A) Organisation de la particule C/D de *P. furiosus* (Xue *et al*., 2010), code PDB : 3NMU. Les deux complexes s'associent par la dimérisation de la protéine aNop5 via son domaine *« coiled-coil »*. (B) Comparaison avec le complexe aNop5-aNop1 de *P. furiosus* en rose (Oruganti *et al*., 2007), code PDB : 2NNW.

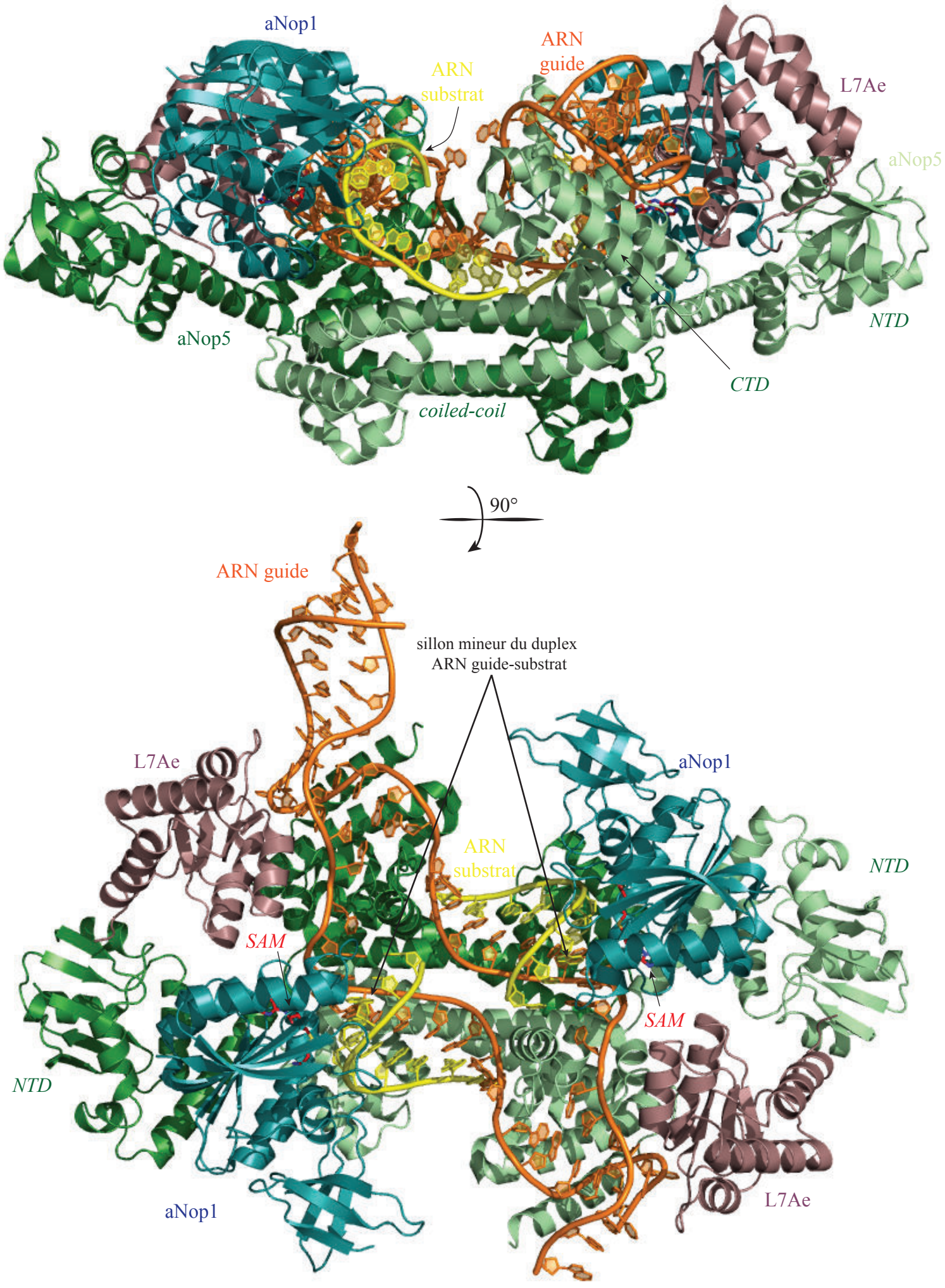


Figure 36. **Structure tridimensionnelle d'une particule sRNP à boîte C/D complète chez *Sulfolobus solfataricus***. Les protéines, leurs domaines et les ARN guides et substrat sont mentionnés. La particule présente une parfaite symétrie (Lin *et al*., 2011), code PDB : 3PLA.

collaborateurs, l'ARN guide s'assemble avec un dimère de la protéine aNop5 et avec deux copies des protéines L7Ae et aNop1, comme cela était proposé dans le modèle classique de la mono RNP (Aittaleb *et al.*, 2003; Ye *et al.*, 2009). L'organisation de la particule est basée sur le dimère de la protéine aNop5 qui sert de plateforme pour accueillir les autres composants de la particule (Figure 36). Le domaine NOP de aNop5 et la protéine L7Ae constituent le module de fixation de l'ARN et emprisonnent le motif en *« K-turn »* formé par les boîtes C/D de l'ARN guide. De cette façon, les deux motifs en *« K-turn »* sont ancrés à la face opposée de la plateforme « *coiled-coil* », plaçant la séquence guide au centre du complexe (Figure 36). Comme vu précédemment, le domaine N-terminal de la protéine aNop5 forme avec la aNop1 le module catalytique (Lin *et al.*, 2011). L'enzyme s'associe à l'hélice du duplex ARN guide-substrat par le sillon mineur de telle manière que le cinquième nucléotide de l'hélice formée par le guide et le substrat soit positionné dans le site actif. Cette conformation active de la particule permet donc la 2'-O-méthylation de deux molécules d'ARN substrats associés à chacun des motifs de l'ARN guide. De plus, il ne s'établit aucun contact entre les deux duplex ARN guide-substrat suggérant une activité enzymatique indépendante des deux motifs (Lin *et al.*, 2011).

Récemment, une étude chez la levure *S. cerevisiae* a montré un couplage spatial et fonctionnel entre les boîtes C/D et C'/D' au sein des RNP eucaryotes, comme cela avait été démontré chez les archaea (Qu *et al.*, 2011; Tran *et al.*, 2005; Tran *et al.*, 2003). Ceci démontre ainsi que cette caractéristique est conservée au cours de l'évolution des RNP à boîtes C/D. Cette étude apporte donc des données supplémentaires d'un assemblage symétrique pour les particules eucaryotes. Les auteurs ont ainsi montré qu'en dépit du caractère indépendant de l'activité de méthylation de la particule C/D, la fonctionnalité de la particule C'/D', souvent formée autour d'un motif ARN de séquence dégénérée chez les eucaryotes, requiert la présence du motif C/D voisin. Contrairement à ce qui avait été observé chez le Xénope où l'absence de détection (par *« cross-link »*) de la protéine 15.5K sur le motif C'/D' avait conduit à l'élaboration d'un modèle asymétrique (Cahill *et al.*, 2002) , un jeu complet de protéines semble associé à chaque motif. Ainsi, la faible affinité de la protéine 15.5K/Snu13p pour les motifs en « *K-loop* » (motif C'/D') observée *in vitro* (Charron *et al.*, 2004; Cléry *et al.*, 2007; Marmier-Gourrier *et al.*, 2003; Tran *et al.*, 2003; Watkins *et al.*, 2002; Watkins *et al.*, 2000) et *in vivo* par « *cross-link* » (Szewczak *et al.*, 2002) semble compensée *in vivo* par l'établissement de contacts protéines-protéines au sein de la particule entière. L'association de Nop56p avec Nop58p par leurs domaines « *coiled-coil* » doit en

particulier avoir une forte contribution. Les auteurs ont également montré qu'une distance inter-motif de 12 à 36 nucléotides est nécessaire à l'activité du motif C'/D', ce qui est en accord avec les distances observées chez la plupart des snoRNA C/D de levure. Contrairement au système archae, dans lequel cette distance est contrainte à une douzaine de résidus (Tran *et al.*, 2005), la présence de séquence inter-motif de grande taille dans les snoRNA eucaryotes est compensée par la formation de structures en tiges boucles qui rapprochent les deux motifs dans l'espace. Les auteurs proposent donc, comme pour les archaea, l'établissement d'un modèle symétrique de mono RNP, avec l'association à chaque couple de boîtes, d'un jeu de protéines Snu13p-Nop1p-Nop56p ou Nop58p pontés entre eux par la dimérisation des domaines « *coiled-coil »* des protéines Nop56p et Nop58p.

3.5 La snoRNP U3

3.5.1 Le snoRNA U3

Comme mentionné dans le paragraphe 1.1.1, le snoRNA U3 est composé de deux domaines, un domaine 5' court (décrit au paragraphe 1.1.1) et un long domaine 3' possédant les caractéristiques des snoRNA à boîtes C/D (Ségault *et al.*, 1992). Ces deux domaines sont séparés par une région charnière appelée *« 3'-hinge »*.

Le domaine 3' est très structuré et comporte deux motifs structurés en *« K-turn »* (Watkins *et al.*, 2000). Les couples de boîtes C'/D et B/C qui servent au recrutement des protéines de la snoRNP (Méreau *et al.*, 1997). Ce domaine est nécessaire pour la stabilité du snoRNA. Les deux couples de boîtes se structurent en motif *« K-turn »* pour recruter *i)* les protéines des particules C/D au niveau du couple C'/D à savoir Snu13p, Nop1, Nop56 et Nop58 (Marmier-Gourrier *et al.*, 2003; Watkins *et al.*, 2000) et *ii)* les protéines Snu13p/15.5 et Rrp9p/h55K, protéine propre à la snoRNP U3, au niveau du couple B/C (Clery, 2007). La taille de ce domaine est très variable suivant les espèces. Cela se traduit par l'existence de structures en tige-boucles plus ou moins longues entourant le couple de boîtes B/C. Ainsi chez l'homme, il existe deux motifs en tige-boucle qui encadrent le couple B/C (Parker & Steitz, 1987), tandis que chez la levure *S. cerevisiae*, ce motif se situe au centre d'une structure cruciforme comportant trois tige-boucles (Ségault *et al.*, 1992) (Figure 37). Le motif C'/D est essentiel pour l'accumulation du snoRNA U3, pour son assemblage avec les protéines de la snoRNP, pour la maturation de son extrémité 3', pour l'hyperméthylation de sa coiffe, pour sa localisation nucléolaire ainsi que pour sa fonction (Speckmann *et al.*, 2000). Les boîtes B et

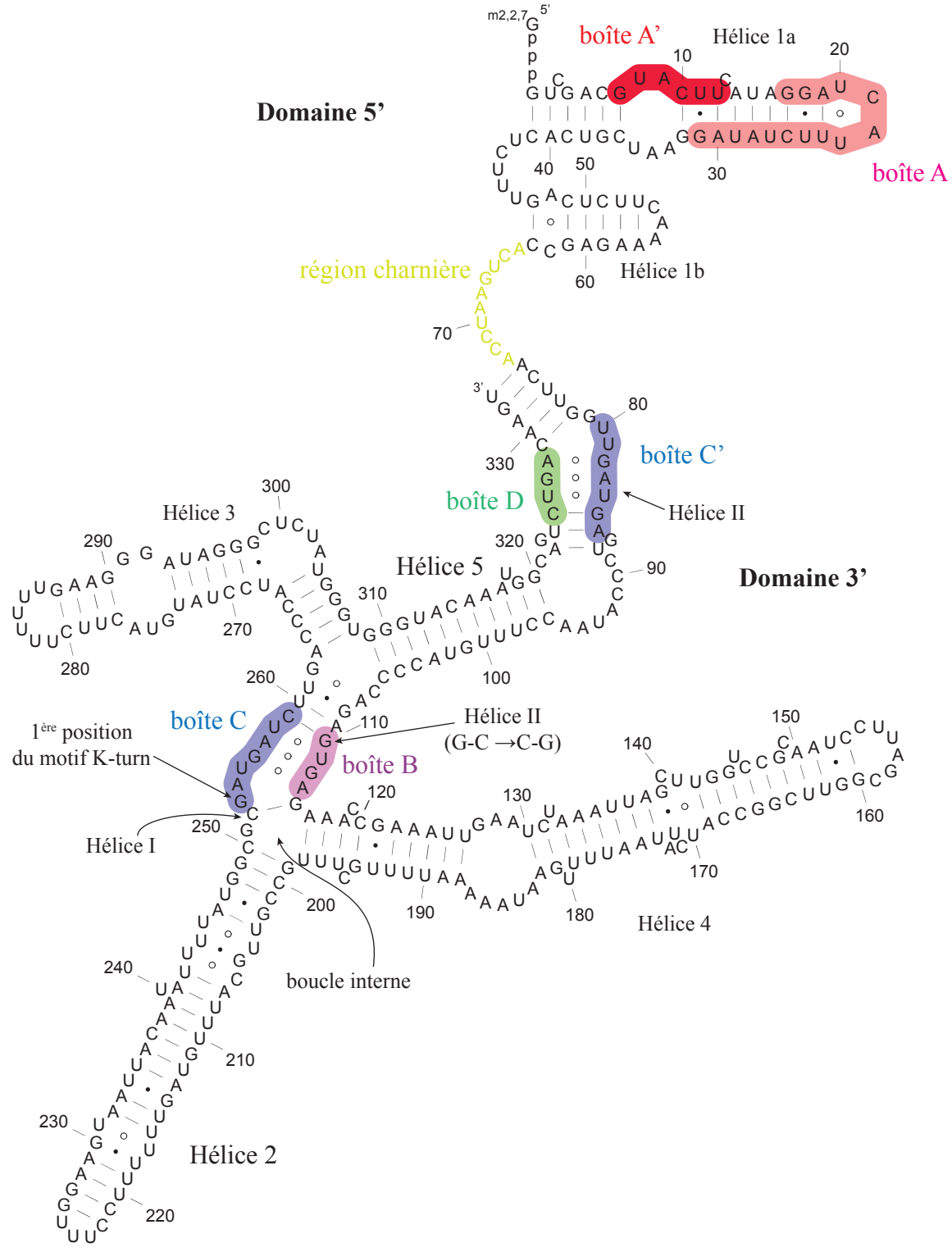


Figure 37. **Structure secondaire du snoRNA U3 de *Saccharomyces cerevisiae***. La structure a été déterminée par modificaion chimique et enzymatique en solution. D'après (Marmier-Gourrier *et al.*, 2003; Ségault *et al*., 1992).

C, quant à elles ne sont pas nécessaires pour l'accumulation du snoRNA U3, ni pour sa localisation nucléolaire, mais sont essentielles pour sa fonction (Cléry *et al.*, 2007; Samarsky & Fournier, 1998).

D'une espèce à l'autre, la taille du snoRNA U3 varie de façon importante : 333 nucléotides chez la levure contre 213 chez l'homme. Chez les levures et les vertébrés, le snoRNA U3 est transcrit par l'ARN polymérase II et porte une coiffe 2,2,7-triméthylguanosine (ou coiffe tri-méthylée) à son extrémité 5' (Reddy & Busch, 1983). Il est intéressant de noter que chez les vertébrés, le snoRNA U3 subit une réaction de pseudouridylation sur le résidu uracile en position 8 de la boîte A' (Reddy *et al.*, 1980) (Figure 37). En revanche, aucune pseudouridine n'a à ce jour été identifiée sur le snoRNA U3 de levure *S. cerevisiae* (Massenet *et al.*, 1998).

3.5.2 Les protéines de la snoRNP U3

La composition des protéines formant la snoRNP U3 a été établie à partir de la purification de cette particule chez l'humain, puis chez la levure (Lübben *et al.*, 1993; Watkins *et al.*, 2000). Comme toutes les particules C/D, la snoRNP U3 renferme des protéines Snu13p/15.5K, Nop1p/Fibrillarine, Nop56p/NOP56 et Nop58p/NOP58. Mais ces travaux ont également montrés la présence d'une protéine spécifique de cette particule : la protéine Rrp9p/U3-55K (Cléry *et al.*, 2007; Lübben *et al.*, 1993; Venema *et al.*, 2000; Watkins *et al.*, 2000) (Figure 38). Bien qu'elle soit nécessaire pour la viabilité cellulaire et pour la fonction de la snoRNP U3 (Granneman *et al.*, 2004; Lukowiak *et al.*, 2000; Venema *et al.*, 2000), l'absence de la protéine Rrp9p chez la levure affecte peu l'accumulation du snoRNA U3 alors que la protéine U3-55K est essentielle pour la stabilité de l'ARN chez l'homme (Knox *et al.*, 2011; Venema *et al.*, 2000). Structuralement, la protéine Rrp9p/U3-55K semble posséder 7 domaines WD40 qui pourraient se structurer en un large domaine en *« helix-propeler »* (Granneman *et al.*, 2002; Pluk *et al.*, 1998). Tout comme les protéines des snoRNP C/D ou la protéine PRP31, le recrutement de la protéine Rrp9p/U3-55K nécessite la présence de la protéine Snu13p/15.5K sur le motif B/C du snoRNA (Cléry *et al.*, 2007; Granneman *et al.*, 2002). *In vitro*, il a été montré que les motifs WD40 présents à l'extrémité C-terminale de la protéine U3-55K sont nécessaires et suffisants pour l'association de la protéine sur le complexe U3 B/C-15.5K (Granneman *et al.*, 2002). Au niveau des déterminants ARN, une boucle interne, en amont de l'hélice I du motif en *« K-turn »* B/C est essentielle pour la fixation de la protéine U3-55K (Granneman *et al.*, 2002) (Figure 37). Au laboratoire, A. Cléry

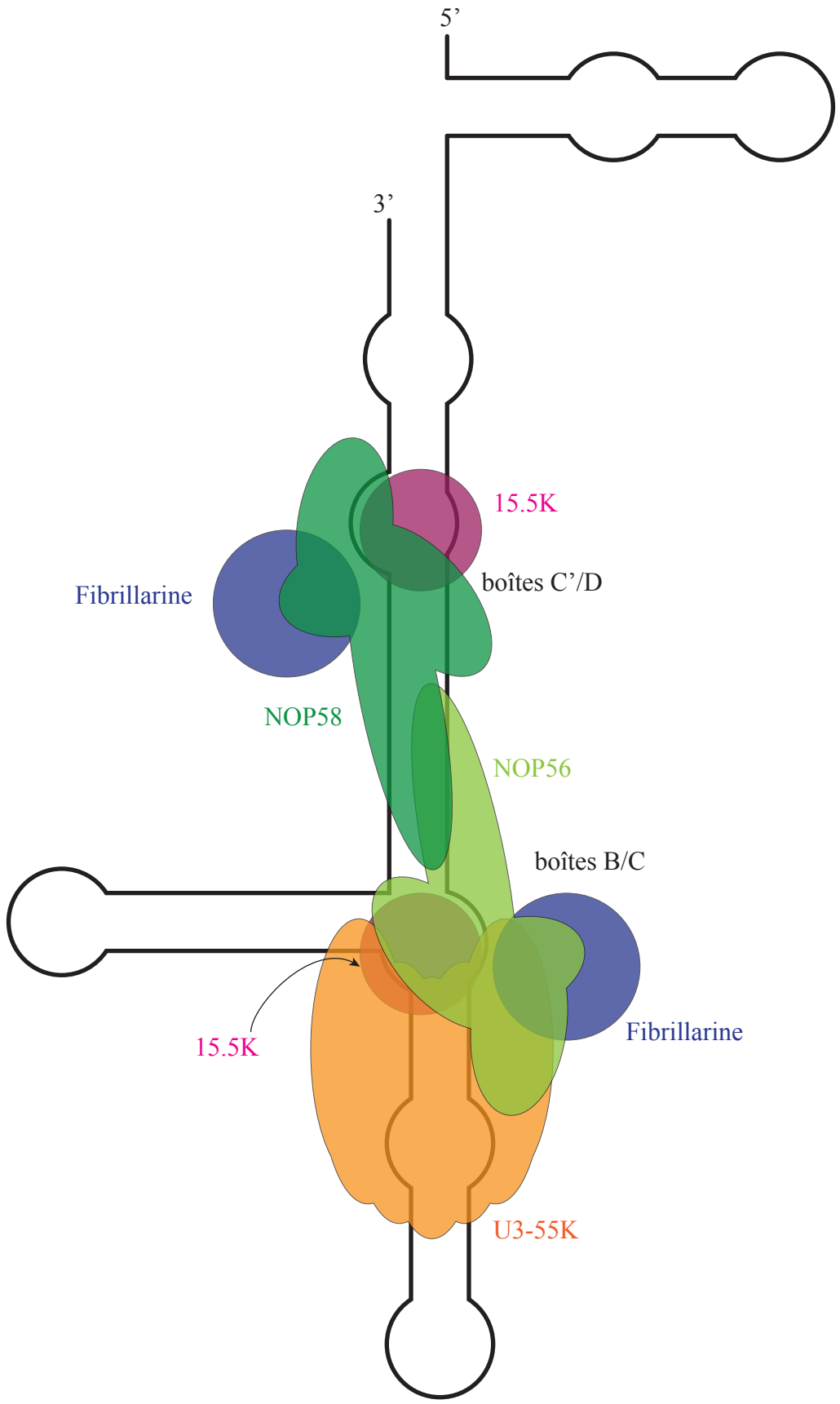


Figure 38. **Modèle d’assemblage de la snoRNP U3 humaine**. D’après (Knox *et al*., 2011).

a pu mettre en évidence trois paramètres importants pour la fixation de la protéine de levure Rrp9p sur l'ARN U3 : la présence d'une paire de base G-C inversée dans l'hélice II du *« K-turn »* (position 111-259 dans la figure 37), une hélice I minimum réduite à une seule paire de bases G-C et la présence d'un résidu G à la première position du *K-turn* (Cléry *et al.*, 2007) (position 253 dans la figure 37). Plus récemment, deux autres sites de contact de la protéine Rrp9p ont été détectés sur le snoRNA U3 : les hélices 2 et 4 (Granneman *et al.*, 2009) (Figure 37).

3.5.3 Structure de la snoRNP U3

A l'heure actuelle, aucune donnée structurale n'est disponible pour la snoRNP U3. Néanmoins, sur la base des connaissances sur les sRNP d'archae, les données obtenues par l'équipe de N.J. Watkins et notre équipe ont permis d'établir un modèle de la particule (Figure 38) et la structure 3D de la partie C'/D a été modélisée par C. Charron au laboratoire. Une coopérativité de fixation des protéines avec les motifs C'/D et B/C pourrait exister. En effet, l'incorporation de la protéine U3-55K au niveau du motif B/C compenserait la présence d'une boîte C' sous optimale pour la fixation de la protéine 15.5K aussi bien chez vertébrés (Knox *et al.*, 2011) que chez la levure (N. Rolland, données non publiées du laboratoire). L'étude décrite par Knox *et al.* a permis de mettre en évidence une interaction directe entre les protéines U3-55K et les protéines NOP56 et NOP58 au sein de la particule. Chez les levures, des données récentes ont montrées que la protéine Nop58p s'associe à l'hélice II du motif C'/D et que les protéines Nop56p et Nop1p interagissent avec la base de l'hélice 3 située à proximité du motif B/C (Granneman *et al.*, 2009).

Le modèle obtenu est un modèle pseudo-symétrique où le motif C'/D s'associerait avec les protéines 15.5K, NOP58, NOP1 et en partie NOP56 qui interagirait avec U3-55K fixée sur le motif B/C avec 15.5K et peut être une autre molécule de NOP1. Dans ce modèle, comme dans les modèles des sRNP à boîtes C/D d'archae, les protéines NOP58 et NOP56 s'auto-dimériseraient par l'intermédiaire de leur domaine « *coiled-coil* ». Des travaux menés au laboratoire sur le snoRNA U3 de levure ont permis de montrer que la longueur entre les motif C'/D et B/C est un paramètre déterminant pour la stabilité de l'ARN *in vivo* (N. Marmier-Gourrier & N. Rolland, communication personnelle). Dans le modèle proposé, cette contrainte de distance pourrait correspondre à l'espacement nécessaire pour la dimérisation des protéines Nop56p et Nop58p, associées respectivement au niveau des motifs B/C et C'/D.

3.5.4 La biogenèse de la snoRNP U3

Il existe deux modes de production des snoRNA C/D : la production des snoRNA codés par des introns dépend de l'épissage du pré-ARNm dans lequel ils sont intégrés et les snoRNA produits sous forme d'unités transcriptionnelles individuelles (mono ou polycistroniques) qui dépendent seulement d'étapes de clivage. Ainsi, plusieurs voies de maturation sont possibles. Chez la levure *S. cerevisiae*, du fait du faible nombre d'introns dans les gènes de protéines et de leurs tailles réduites, la majorité des snoRNA sont produits sous la forme d'unités transcriptionnelles monocistroniques (par exemple U3) ou polycistroniques (comme U14 et snR190) (Piekna-Przybylska *et al.*, 2007). A l'inverse, chez les vertébrés, à quelques exceptions près, telles que les snoRNA comme les U3, U8 ou U13 qui sont produits à partir d'unités transcriptionnelles indépendantes, la majorité des snoRNA sont situés dans les introns de pré-ARNm (par exemple U14) (Lestrade & Weber, 2006). Il est important de noter que U3 est codé par environ une centaine d'unités transcriptionnelles indépendantes chez l'homme. Beaucoup de gènes contenant des séquences de snoRNA codent des protéines impliquées dans la biogenèse des ribosomes, suggérant une co-régulation de la production des snoRNA et de ces protéines. Dans ce dernier cas, la libération du snoRNA implique une série de clivages endo et exonucléolytiques qui peuvent se dérouler selon deux principaux mécanismes : une voie majoritaire impliquant le débranchement préalable de la structure en lasso (Ooi *et al.*, 1998; Petfalski *et al.*, 1998) et une voie minoritaire nécessitant un clivage endonucléolytique indépendant de l'épissage pour les snoRNA H/ACA et l'intervention des protéines splicéosomales pour les snoRNP C/D (Ooi *et al.*, 1998; Petfalski *et al.*, 1998). Dans le cas des transcrits polycistroniques, le transcrit primaire subit des clivages endonucléolytiques par l'endoribonucléase Rnt1p qui libère les unités individuelles (Chanfreau *et al.*, 1998). Les extrémités de ces unités ainsi libérées, tout comme celles des snoRNA introniques, subissent une hydrolyse par les exonucléases Rat1p et Xrn1p du côté 5' (Petfalski *et al.*, 1998; Qu *et al.*, 1999) et par l'exosome du côté 3' (Allmang *et al.*, 1999; van Hoof *et al.*, 2000b). A la différence des snoRNA introniques, les snoRNA cistroniques acquièrent une coiffe m7G à leur extrémité 5' de manière co-transcriptionnelle qui subit ensuite une hyperméthylation par l'enzyme Tgs1p/TGS1 (par exemple U3) (Mouaikel *et al.*, 2002). Les snoRNA introniques pourraient aussi en acquérir une mais de manière post-transcriptionnelle.

Ainsi la biogenèse de la snoRNP U3 nécessite plusieurs étapes distinctes, telles que l'assemblage des protéines cœur de la snoRNP, la maturation nucléolytique du snoRNA,

l'hyperméthylation de la coiffe. Chez les levures des genres *Saccharomyces*, *Kluyveromyces* et *Pichia*, le gène codant le snoRNA U3 présente la particularité de contenir un intron épissé par la machinerie spliceosomale (Brulé *et al.*, 1996; Fournier *et al.*, 1998; Myslinski *et al.*, 1990). Au début des années 2000, l'équipe de D. Tollervey s'est attachée à caractériser le processus de maturation nucléolytique du snoRNA U3 chez la levure *S. cerevisiae* (Kufel *et al.*, 2000; Kufel *et al.*, 2003) (Figure 39). Le snoRNA est transcrit sous la forme d'un grand précurseur qui est progressivement hydrolysé à partir de son extrémité 3'. La forme la plus longue observée correspond à une forme non épissée du snoRNA, porteuse d'une longue extension à son extrémité 3'. Dans un second temps, il y aurait épissage par le splicéosome et clivage endonucléolytique par l'enzyme Rnt1p aux positions +58 et +21 (par rapport à l'extrémité 3' mature). Alors que les sous-produits autre que le pré-ARN U3 seraient dégradés par l'exonucléase 5'→3' Rat1p, le pré-ARN U3 serait hydrolysé rapidement jusqu'aux positions +18 et +12. Cette dégradation pourrait être catalysée par les exonucléases de la famille Rex1-3p et/ou par l'exosome. Les deux nouveaux intermédiaires (+12 et +18) sont stabilisés par les facteurs Lhp1p/La et les protéines Lsm, qui sont recrutés de façon synergique au niveau de séquences riches en uridines situées en amont des sites de coupures.

D'après les auteurs, toutes ces formes ne seraient pas associées aux protéines Nop56p, Nop58p et Nop1p et ne seraient pas encore triméthylées. Il a été proposé que le recrutement de ces trois protéines sur le motif C'/D déplace les protéines Lhp1p et Lsm, permettant ainsi la poursuite de la maturation de l'extrémité 3'. Il a été démontré que le motif C/D (C'/D pour le snoRNA U3) est nécessaire et suffisant pour induire l'hyperméthylation de la coiffe (Speckmann *et al.*, 2000). L'enzyme responsable de l'hyperméthylation de la coiffe, l'enzyme Tgs1p, serait recrutée au niveau du motif C/D (ou C'/D) par le domaine C-terminal basique (KKE/D) de la protéine Nop58p (Mouaikel *et al.*, 2002). La coiffe triméthylée protège l'ARN contre d'éventuelles dégradations exonucléolytiques en 5' et permettrait l'adressage du snoRNA U3 dans le nucléole. Dans un contexte sauvage, la séquence intronique est excisée à partir d'un long précurseur U3 dont l'extrémité 3' n'a pas été déterminée. Des travaux au laboratoire menés par B. Rothé au cours de sa thèse ont montré que, contrairement à ce qui était publié par Kufel et *al.*, chez *S. cerevisiae,* l'épissage du pré-snoRNA U3 s'effectue très tôt, avant même l'intervention de Rnt1p (en +58) et que l'intron n'est pas détecté sur les autres espèces (B. Rothé, travaux de thèse). La maturation de l'extrémité 3' est réalisée par une hydrolyse exonucléolytique 3'→5'. D'une manière générale, le motif C/D est requis pour la terminaison de la transcription et la formation de l'extrémité 3' mature. Les protéines qui

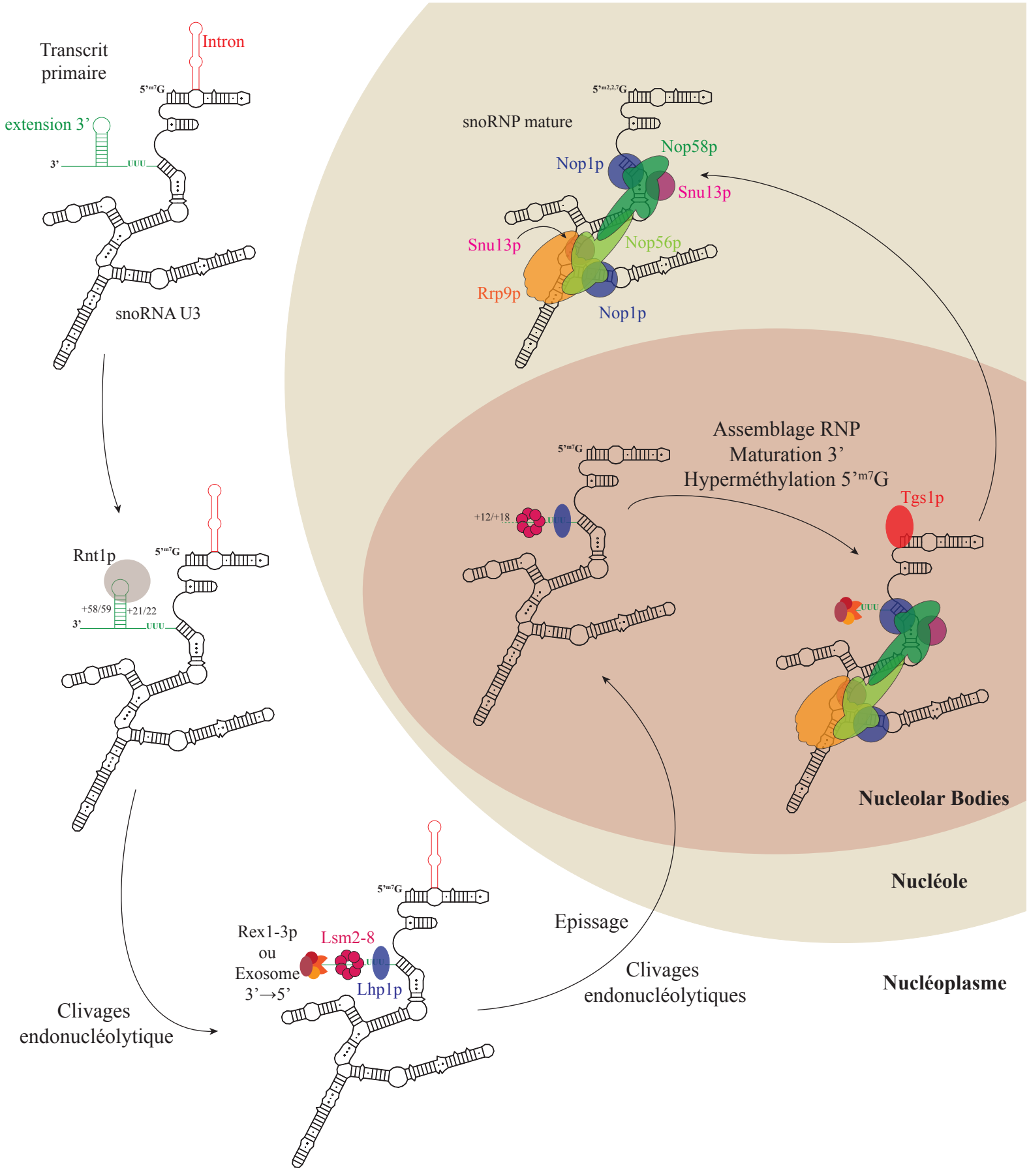


Figure 39. **Modèle de biogenèse de la snoRNP U3 chez *S. cerevisiae.*** D'après (Kufel *et al*., 2000; Verheggen *et al*., 2002) et les données de B. Rothé au laboratoire.

s'associent aux précurseurs du snoRNA stabilisent ainsi l'ARN et protègent son extrémité 3' contre la dégradation. Par la suite, des expériences de microscopie électronique ont permis de mettre en évidence la localisation nucléolaire au niveau des NB (« *Nucleolar bodies* ») des intermédiaires +12 et +18 du snoRNA U3 (Verheggen *et al.*, 2002) (Figure 39). Cette biogenèse apparaît donc comme une succession d'étapes hautement ordonnées, à la fois temporellement et spatialement.

Même si elle est moins connue que le chez la levure, la biogenèse de la snoRNP U3 chez les vertébrés comporte de nombreuses similitudes. Des expériences d'immunoprécipitation ont démontré que seule la protéine 15.5K est associée aux précurseurs du snoRNA, suggérant qu'en absence des protéines NOP58 et NOP1, ces derniers doivent être stabilisés par des facteurs supplémentaires, comme cela a été montré pour les intermédiaires +12 et +18 chez la levure. De plus, chez le rat, les précurseurs du snoRNA sont essentiellement détectés au niveau des sites de transcription et des corps de Cajal (CB), tandis que l'ARN mature est localisé dans les CB et le nucléole. L'équipe d'E. Bertrand a mis en évidence que l'enzyme TGS1 et les protéines 15.5K, NOP56 et NOP58 s'accumulent dans les CB (Verheggen *et al.*, 2002) tout comme aNOP1 (Gerbi & Borovjagin, 1997). Ceci suggère donc que la formation de la coiffe ^{m2,2,7}G, la maturation nucléolytique terminale de l'extrémité 3' et l'assemblage des protéines cœur pourrait se dérouler au sein de ce sous-compartiment nucléaire, comme cela semble être le cas chez la levure dans les NB.

3.6 La biogenèse des RNP à boîtes C/D : état des connaissances au début de ma thèse

En plus des facteurs intervenant dans la maturation du snoRNA, différents facteurs requis pour la biogenèse de la snoRNP C/D ont été identifiés avant mon arrivée au laboratoire. Ces facteurs semblaient participer aux évènements d'assemblage des particules. C'était notamment le cas des hélicases Rvb1p/TIP49, Rvb2p/TIP48, des facteurs Pih1p/NOP17, Tah1p/SPAGH, Bcd1p/BCD1 et de la protéine Rsa1p/NUFIP (Zhao, 2005, Boulon, 2008 Watkins, 2004).

3.6.1 La chaperonne Hsp90

Les protéines Hsp90 sont des chaperonnes moléculaires abondantes et très conservées qui participent à de nombreux processus cellulaires fondamentaux comme le contrôle du cycle cellulaire, la survie cellulaire, les processus de signalisation et les réponses au stress

(Borkovich *et al.*, 1989; Caplan *et al.*, 2007; Zhao *et al.*, 2005), pour revue (Richter *et al.*, 2010; Wandinger *et al.*, 2008; Young *et al.*, 2001). Sa fonction s'inscrit dans le maintien de l'homéostasie cellulaire, pour revue (Dezwaan & Freeman, 2008; Taipale *et al.*, 2010). Un bilan des protéines s'associant à la chaperonne Hsp90 est disponible aux adresses internet suivantes :

- http://www.picard.ch/downloads/Hsp90interactors.pdf
- http://www.picard.ch/downloads/Hsp90facts.pdf
- http://www.picard.ch/Hsp90Int/ (Echeverria *et al.*, 2011)

Les chaperonnes Hsp90 ont également pour fonction de maintenir de nombreuses activités cellulaires essentielles en dépit des mutations ou des variations environnementales. Il a même été montré que les chaperonnes Hsp90 jouaient un rôle important dans le développement tumoral (Trepel *et al.*, 2010).

Dans son activité de chaperonne du repliement des protéines, Hsp90 s'associe avec d'autres chaperonnes comme Hsp40 et Hsp70 et des facteurs protéiques appelés co-chaperons qui réguleraient le cycle d'activité de ces chaperonnes en inhibant ou activant leur activité ATPase. Ces co-chaperonnes serviraient à cibler les protéines substrats (Chen & Smith, 1998; Eckert *et al.*, 2010; Frydman & Hohfeld, 1997; Panaretou *et al.*, 2002; Prodromou *et al.*, 1999; Richter *et al.*, 2004; Roe *et al.*, 2004; Siligardi *et al.*, 2002).

Hsp90 forme un homodimère flexible où chaque molécule est composée de 3 domaines principaux : i) le domaine N-terminal qui contient le site de fixation à l'ATP (Prodromou *et al.*, 1997), suivit d'une région portant un grand nombre de résidus chargés et qui a une longueur variable, ii) un domaine central essentiel pour l'hydrolyse de l'ATP contenant les sites d'interaction pour les protéines clientes et de certaines co-chaperonnes (Meyer *et al.*, 2003) et iii) un domaine C-terminal responsable de la dimérisation de la chaperonne et du recrutement des co-chaperonne à domaines TPR (*« TetratricoPeptide Repeat »*) via son motif NH_2-MEEVD-COOH (Scheufler *et al.*, 2000).

Structuralement, un motif TPR consiste en une séquence de trente-quatre acides aminés souvent dégénérée formant deux hélices α antiparallèles reliées par un coude. Ces motifs TPR s'associent pour former une super-hélice dont le sillon sert de site d'interaction pour la chaperonne notamment certains acides aminés conservés au sein des motifs TPR forme la pince à carboxylate ou *« carboxylate clamp »* (Das *et al.*, 1998; Prasad *et al.*, 2010;

Scheufler *et al.*, 2000), pour revue (D'Andrea & Regan, 2003; Zeytuni & Zarivach, 2012). Nous reviendrons plus en détails sur les domaines TPR dans la partie Résultats & Discussion de ce manuscrit.

Parmi les co-chaperonnes à domaine TPR, la protéine HOP *« (Hsp70 and Hsp90 organizing protein »* (Sti1p chez la levure) est la plus connue et la mieux décrite (Chang & Lindquist, 1994; Johnson *et al.*, 1998; Scheufler *et al.*, 2000). Elle est composée de plusieurs motifs TPR, qui s'assemblent en trois domaines TPR ce qui lui permet de s'associer simultanément aux chaperonnes Hsp90 et Hsp70 par leurs motifs terminaux NH_2-EEVD-COOH. Cette double interaction permet ainsi de former le complexe Hsp90-HOP/Sti1p-Hsp70, qui constitue l'entité minimale pour le fonctionnement de la chaperonne.

Au cours de son cycle d'activité, la chaperonne Hsp90 adopte différentes conformations (Figure 40). Dans son état « apo », Hsp90 est principalement dans une conformation ouverte (ou *« V-shaped »*). Cet état est stabilisé par la présence de co-chaperonne comme HOP/Sti1p ou p50/Cdc37p qui inhibe l'activité ATPasique de la chaperonne et facilite la prise en charge de la protéine cliente notamment par la présence concomitante de la chaperonne Hsp70 (Chen & Smith, 1998; Hessling *et al.*, 2009; Johnson *et al.*, 1998; Prodromou *et al.*, 1999; Richter *et al.*, 2003; Roe *et al.*, 2004; Wegele *et al.*, 2003; Wegele *et al.*, 2006). La fixation de l'ATP favorise le passage à une conformation fermée où les domaines N-terminaux dimérisent et s'associent aux domaines centraux. Cette conformation est également stabilisée par des co-chaperonnes comme p23/Sba1p (Ali *et al.*, 2006). Après l'hydrolyse de l'ATP, les domaines N-terminaux se dissocient, l'ADP et la protéine cliente correctement repliée sont relargués et la chaperonne retourne à sa conformation initiale, certaines co-chaperonnes comme AHA1/Aha1p favorise l'hydrolyse de l'ATP (Hessling *et al.*, 2009; Meyer *et al.*, 2004). Au cours de ce cycle, plusieurs PPIases (*« Peptidyl-Prolyl isomerase »*) interviennent. Ces dernières font également partie de la famille des protéines à domaines TPR, ce domaine leur permettrait notamment de déplacer la protéine HOP/Sti1p de son site de fixation (Li *et al.*, 2011a). Il est intéressant de noter qu'en plus de cette régulation par les co-chaperonne, l'activité d'Hsp90 est également sous le contrôle de modifications post-traductionnelles, pour revue (Mollapour & Neckers, 2012).

Comme dans beaucoup d'autres organismes, il existe chez la levure *S. cerevisiae* deux isoformes de la chaperonne : la protéine Hsc82p exprimée de façon constitutive, et la protéine Hsp82p qui est exprimée, en conditions de stress thermique ou en phase stationnaire de

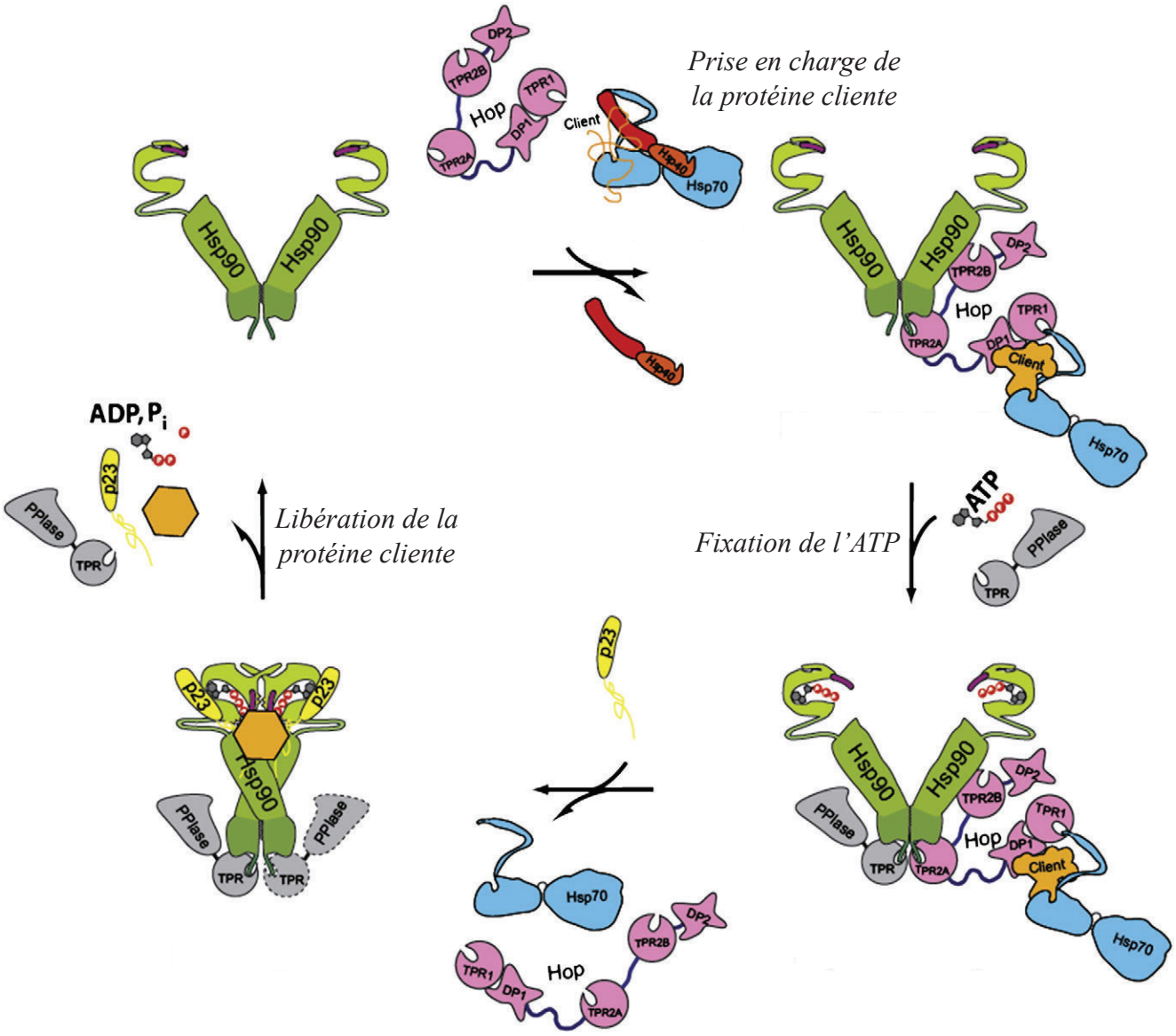


Figure 40. **Représentation schématique du cycle d'activité de la chaperonne Hsp90.** La protéine cliente est transférée de Hsp70 à Hsp90 grâce à la protéine adaptatrice HOP/Sti1p. Hsp90 adopte une conformation fermée (active pour l'hydrolyse de l'ATP) après la fixation de l'ATP. La co-chaperonne p23/Sba1p stabilise la conformation fermée de la chaperonne Hsp90 ce qui affaiblit l'association de HOP/Sti1p et facilite son élimination du complexe. Après l'hydrolyse de l'ATP, p23/Sba1p et la protéine cliente correctement repliée sont relargées de la chaperonne Hsp90. D'après (Li *et al*., 2012).

croissance (Borkovich *et al.*, 1989; Donalies & Stahl, 2001). Ces isoformes ont une localisation cellulaire principalement cytoplasmique mais semblent pouvoir faire l'objet de passages transitoires dans le noyau en phase exponentielle de croissance (Huh *et al.*, 2003; Tapia & Morano, 2010).

Le journal « Biochimica et Biophysica Acta » a récemment consacré un numéro entier à la chaperonne Hsp90 (Li *et al.*, 2012; Makhnevych & Houry, 2012; Mollapour & Neckers, 2012; Prodromou, 2012).

3.6.2 Le complexe R2TP

En 2005, l'équipe de W.A. Houry, dans un criblage à haut débit visant à définir de nouveaux partenaires de la chaperonne Hsp90 de levure, a mis en évidence l'existence d'un complexe formé de 4 protéines : les deux hélicases Rvb1p et Rvb2p associées aux protéines Tah1p et Pih1p. Ce complexe a été nommé R2TP pour « Rvb1p-Rvb2p-Tah1p-Pih1p » (Zhao *et al.*, 2005), pour revue (Kakihara & Houry, 2012).

3.6.2.1 Les protéines Rvb1 et Rvb2

En raison de leur implication dans de nombreux processus cellulaires, ces deux protéines homologues possèdent de nombreuses dénominations différentes. Ainsi TIP49 est également appelée Rvb1p chez la levure ou encore TIP49a, RUVBL1, p55, TAP54α ou Pontin52 chez les vertébrés tandis que la protéine TIP48 porte le nom de Rvb2p chez la levure, mais aussi TIP49b, RUVBL2, p50, TAP54β ou Reptin42 chez les vertébrés (Watkins *et al.*, 2002). Ces protéines font partie de la famille des ATPases AAA+ (*« ATPases associated with diverse cellular activities »*). L'association des protéines Rvb1p/TIP49 et Rvb2p/TIP48 avec un snoRNA à boîtes C/D a été mise en évidence pour la première fois chez la souris (Newman *et al.*, 2000; Watkins *et al.*, 1998b), puis chez l'homme (Watkins *et al.*, 2002). Il a été montré que la déplétion des protéines TIP49 et TIP48 affecte le taux de snoRNA à boîtes C/D chez l'homme (Watkins *et al.*, 2004). Chez la levure, l'implication des protéines Rvb1p et Rvb2p dans la biogenèse des snoRNP C/D et même H/ACA a également été démontrée (Boulon *et al.*, 2008; King *et al.*, 2001; Zhao *et al.*, 2008).

Structuralement, la protéine humaine TIP49 ou RuvBL1 a fait l'objet d'une étude cristallographique qui a permis la détermination de sa structure 3D (Matias *et al.*, 2006). La protéine s'assemble en un hexamère constitué de monomères de protéine RuvBL1 liés à une molécule d'ADP (Figure 41). Chaque monomère est constitué de trois domaines dont le

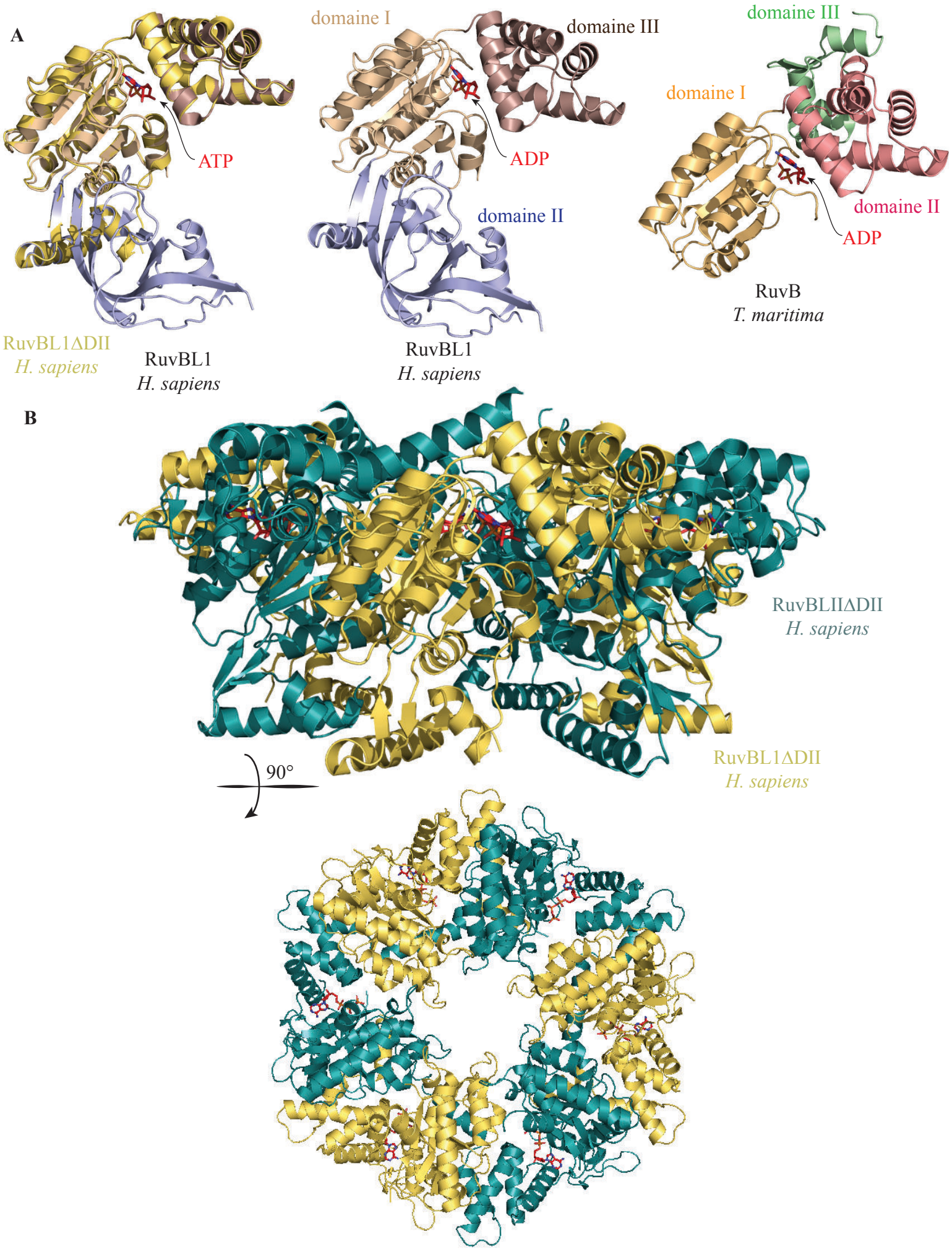


Figure 41. **Structure tridimensionnelle des protéines humaines RuvBL**. (A) Comparaison entre la protéine humaine RuvBL1 (Matias *et al*., 2006), code PDB : 2C9O, la protéine RuvB de *Thermotoga maritima* (Putnam *et al*., 2001), code PDB : 1IN7 et la protéine humaine RuvBL1DDII dans le complexe RuvBL hexamérique (Gorynia *et al*., 2011), code PDB : 2XSZ. (B) Structure hexamérique du complexe RuvBLI+RuvBLII humains. Les différents domaines des protéines et la molécule d'ATP sont indiqués.

premier et le troisième sont impliqués dans la fixation de l'ATP et son hydrolyse. Le second domaine, propre aux AAA+ protéines et absent dans l'homologue bactérienne RuvB (Putnam *et al.*, 2001), constitue un nouveau domaine de liaison à l'ADN/ARN. Les protéines AAA+ forment des structures circulaires hexamèriques et contiennent un motif conservé pour la fixation et l'hydrolyse de l'ATP : les domaines Walker A et B. Plus récemment, la structure du complexe humain RuvBL a été déterminée (Gorynia *et al.*, 2011) (Figure 41). Ce complexe consiste en un dodécamère formé par deux anneaux hétérohexamériques d'une alternance de monomères des protéines RuvBL1 et RuvBL2 liées à une molécule d'ADP/ATP. Ce complexe a été produit à partir de protéines dans lesquelles le domaine II avait été tronqué. Jusque-là, seules des données de microscopie électronique avaient permis d'appréhender la structure de ce complexe chez l'homme (Puri *et al.*, 2007) et chez la levure (Gribun *et al.*, 2008; Torreira *et al.*, 2008). Cependant, selon les études, l'état d'oligomérisation des protéines Rvb1p/TIP49 et Rvb2p/TIP48 était variable. La structure cristalline du complexe a apporté un élément de réponse (Gorynia *et al.*, 2011) en montrant clairement la possibilité de former un dodécamère constitué de deux hétérohexamères associés chacun par leur face supérieure. Chaque héxamère est composé d'une succession de monomère de RuvBL1 ΔDII et RuvBL2 ΔDII. Pour confirmer que l'état d'oligomérisation du complexe n'était pas dû aux conditions cristallines, les auteurs ont également montré la formation du dodécamère en solution par la technique de SAXS. D'autres études avaient montré que l'état de dodécamère était induit par la présence des étiquettes 6-His fusionnées à l'extrémité N-terminale des protéines de levure Rvb1p et Rvb2p (Cheung *et al.*, 2010). Les auteurs ont alors montré que l'assemblage dodécamérique du complexe humain n'est pas induit par la présence de cette étiquette 6-His. En effet, ils ont utilisés des protéines co-exprimées, fusionnées à des étiquettes différentes (6-His-RuvBL1 ΔDII et FLAG-RuvBL2 ΔDII) et ont analysé les profils d'élution par filtration après protéolyse de l'étiquette 6-His. De plus, les données en solution obtenues par SAXS, l'absence de densité électronique pour ces étiquettes dans le cristal et leur position attendue à l'extérieur des anneaux confirment leur absence d'effet sur l'état d'oligomérisation du complexe (Gorynia *et al.*, 2011).

En 2008, notre laboratoire en collaboration avec celui d'Edouard Bertrand à Montpellier a mis en évidence un lien direct entre le complexe R2TP et les protéines constitutives des snoRNP (Boulon *et al.*, 2008). De plus, ce travail a permis d'observer une diminution des snoRNA U3, U4 et l'ARN de la télomérase lorsque l'activité ATPasique de la chaperonne humaine est inhibée. Cette inhibition provoque également un défaut

d'accumulation des protéines 15.5K, NHP2, NOP58 et SBP2 dans les cellules (Boulon *et al.*, 2008). La machinerie Hsp90-R2TP semble donc être impliquée dans la biogenèse des snoRNP.

3.6.2.2 La protéine Pih1p/NOP17

La protéine Pih1p (*« Protein interacting with Hsp90-1 »*) ou Nop17p (*« Nucleolar protein 17 »*) avait été identifiée lors de la recherche de partenaires de la protéine Nop58p par l'approche double hybride (Gonzales *et al.*, 2005). Cette interaction repose sur la région 217-512 de la protéine Nop58p, qui contient le domaine NOP et une portion du domaine *« coiled-coil »*. Cette interaction a été confirmée *in vitro* par co-immunoprécipitation mais aucune association entre la protéine Pih1p et les autres composants de la particule C/D n'avaient pu être détectés. La délétion de cette protéine non essentielle confère à la cellule un défaut de croissance avec un phénotype accentué à 37°C. Ce phénotype est associé à une diminution du nombre de polysomes et à une réduction de la quantité de ribosomes. En effet, dans une souche de levure invalidée pour le gène *PIH1*, les auteurs ont observé une diminution du taux des ARNr 18S et 25S et une accumulation des précurseurs 35S et 23S. La protéine Pih1p possède une localisation nucléolaire et son absence conduit à une localisation diffuse des protéines Nop1p, Nop56p, Nop58p et Snu13p dans le noyau et dans le cytoplasme, indiquant que Pih1p est nécessaire pour la rétention nucléolaire ou l'adressage nucléolaire des protéines des snoRNP à boîtes C/D. La protéine Pih1p ne s'associe pas directement aux snoRNA à boîtes C/D, ni aux snoRNA à boîtes H/ACA, mais son absence conduit à une augmentation de l'association de la protéine Nop58p avec les snoRNA U3 et U14 sans modifier le taux de ces snoRNA. Les auteurs proposaient donc que l'interaction entre Pih1p et Nop58p régulerait le processus d'assemblage des particules C/D et agirait comme un facteur régulateur nucléolaire (Gonzales *et al.*, 2005).

3.6.2.3 La protéine Tah1p/SPAGH

Initialement, la protéine Tah1p avait été décrite en 2001 par l'équipe de Y. Sakaki comme un partenaire de la protéine Pih1p détecté par double hybride (Ito *et al.*, 2001). En 2005, lors d'un crible haut débit, l'équipe de W.A. Houry a identifié la protéine Tah1p comme un cofacteur de la chaperonne Hsp90 de levure, médiant l'interaction de la protéine Pih1p avec Hsp90 (Zhao *et al.*, 2005). Cette protéine de 111 résidus est caractérisée par la présence d'un domaine TPR (*« tetratricopeptide repeat »*) à son extrémité N-terminale. Elle tire son nom de la présence de ce domaine et de sa faculté à s'associer avec la chaperonne : *« TPR-*

containing protein associated with Hsp90 ». La protéine Tah1p a une localisation nucléaire et cytoplasmique (Huh *et al.*, 2003). Des expériences de « GST pull-down » et de double hybride ont permis de montrer que l'interaction entre les protéines Pih1p et Hsp90 est stabilisée par la protéine Tah1p, et que le domaine TPR permettrait l'interaction avec la chaperonne. Ainsi la protéine Tah1p stabiliserait la protéine Pih1p qui serait en quantité limitante dans la cellule et qui présente une instabilité en phase stationnaire de croissance lorsque Tah1p est absente (Zhao *et al.*, 2008). Le domaine TPR de la protéine Tah1p serait responsable de son association avec la chaperonne selon un mode similaire à celui observé pour la co-chaperonne HOP et son domaine TPR2A (Millson *et al.*, 2008; Scheufler *et al.*, 2000). Ainsi la protéine Tah1p serait spécifique de la chaperonne Hsp90 en discriminant les résidus spécifiques de l'extension C-terminale de la chaperonne, notamment sa méthionine 705 (Millson *et al.*, 2008). Plus récemment, il a été montré que le complexe Pih1p-Tah1p agirait comme un complexe co-chaperon qui inhiberait l'activité ATPase de la chaperonne Hsp90 (Eckert *et al.*, 2010).

La protéine Tah1p fera l'objet d'une description plus détaillée dans la partie Résultats et Discussion de ce manuscrit.

3.6.3 La protéine Rsa1p/NUFIP

La protéine de levure Rsa1p (*« Ribosome assembly 1 »*) a été mise en évidence par un crible synthétique létal chez *S. cerevisiae* comme étant impliquée dans la biogenèse des ribosomes (Kressler *et al.*, 1999). La délétion du gène *RSA1* confère un phénotype thermosensible, une diminution de la quantité de sous-unités ribosomiques 60S, ainsi qu'une accumulation du précurseur pré-ARN 23S. Il est intéressant de noter que la protéine Rsa1p possède une localisation nucléaire et contient un NLS potentiel au niveau de son extrémité C-terminale (Kressler *et al.*, 1999). Elle est en revanche exclue du nucléole. Au laboratoire, N. Marmier-Gourrier et B. Charpentier ont utilisé l'approche de triple hybride afin de rechercher de nouveaux partenaires protéiques du snoRNA U3 (Marmier-Gourrier *et al.*, 2003). En utilisant le motif B/C du snoRNA U3 comme appât, ils ont pu identifier la protéine Rsa1p comme partenaire potentiel. Des données antérieures précisaient que la protéine Rsa1p n'était pas retrouvée au sein de la particule mature (Watkins *et al.*, 2000) et qu'elle ne s'accumulait pas dans le nucléole (Kressler *et al.*, 1999). La protéine Rsa1p pourrait donc interagir de façon transitoire avec la snoRNP U3.

En parallèle, l'équipe d'E. Bertrand à Montpellier travaillait à la recherche de nouveaux partenaires de la protéine 15.5K par l'approche double hybride. Ils ont ainsi pu mettre en évidence un nouveau facteur d'intérêt : la protéine NUFIP (*« Nuclear FMRP Interacting Protein »*). Cette protéine possède un domaine à doigt de zinc de type C2H2 et un NLS bi-partite (Bardoni *et al.*, 1999). Comme Rsa1p, elle interagit par double hybride avec les protéines LSM4 et TGS1 (Boulon, communication personnelle, B. Rothé, travaux de thèse). NUFIP est faiblement détectée au niveau du cytoplasme et du nucléole et présente une localisation principalement nucléoplasmique (Bardoni *et al.*, 1999).

Ces travaux ont conduit à l'élaboration d'une collaboration entre nos deux équipes et à la publication de travaux portant sur ces protéines (Boulon *et al.*, 2008). Ces travaux ont permis de montrer que les protéines Rsa1p et NUFIP s'associent aux snoRNA par l'intermédiaire d'une interaction avec les protéines Snu13p et 15.5K de la famille L7Ae. De plus, cette interaction repose sur un domaine conservé de 37 acides aminés, nommé yPEP chez la levure et hPEP chez l'homme. Il est intéressant de noter que ce domaine correspond à la seule séquence homologue entre ces deux protéines. Cette interaction permet de recruter les protéines Rsa1p et NUFIP sur les snoRNA C/D, comme en témoigne l'existence d'un complexe ternaire Rsa1p/NUFIP-Snu13p/15.5K-snoRNA. Les deux équipes ont également proposé que Rsa1p ou NUFIP soient requises pour la biogenèse des snoRNA car en absence de ces protéines, une diminution du taux de certains snoRNA C/D est observée, notamment de la snoRNP U3 humaine.

OBJECTIFS DE LA THESE

Objectifs de la thèse

Comme nous venons de le voir, les modifications chimiques des riboses et des bases (2'-O-méthylation et pseudouridylation) des ARN ribosomiques ainsi que leurs clivages endo-nucléolytiques sont des évènements conservés chez les archaea et chez l'homme. Ils se déroulent au sein du nucléole où a lieu la biogenèse des ribosomes. Ces étapes de maturation des ARNr mettent en œuvre des particules ribonucléoprotéiques : les s/snoRNP. Ces particules sont composées d'un ARN guide, le s/snoRNA, ainsi que d'un jeu de protéines constitutives. Il existe deux grandes classes de RNP : les s/snoRNP à boîtes C/D et les s/snoRNP à boîtes H/ACA. Chez les eucaryotes, il existe beaucoup d'autres RNP comme par exemple les snRNP U1, U2, U4, U5 et U6, le complexe de la télomérase, et des mRNP particulières contenant un élement SECIS (*« SElenoCysteine Incorporation Sequence »*) chez les vertébrés. Les snoRNP C/D de la levure *Saccharomyces cerevisiae* contiennent les protéines Snu13p, Nop56p, Nop58p et Nop1p/Fibrillarine. Parmi ces protéines, la protéine Snu13p fait partie de la famille des protéines L7Ae qui se caractérise par sa capacité à se lier à un motif structural ARN particulier : le motif en *« K-turn »*. La fixation de cette protéine est l'évènement nucléateur de l'assemblage des particules. Chez les Archaea, les différents constituants des RNP semblent contenir dans leur structure l'ensemble des informations nécessaires et suffisantes pour assurer leur assemblage en particules actives. Nous avons vu que des sRNP d'Archaea à boîtes C/D et H/ACA actives *in vitro* ont pu être reconstituées à partir de protéines recombinantes et de nombreuses structures tridimensionnelles de sRNP ont été déterminées, alors que chez les eucaryotes, les approches de reconstitution sont infructueuses. Cette différence s'explique par la complexité des cellules eucaryotes avec une compartimentalisation de l'espace cellulaire et l'existence de machineries moléculaires contrôlant l'efficacité, la spécificité et la processivité de ce processus de biogenèse.

A mon arrivée au laboratoire, le groupe de B. Charpentier avait mis en évidence un nouveau partenaire de la protéine Snu13p dans le cadre d'une étude de la de la snoRNP U3 chez *S. cerevisiae*, la protéine Rsa1p. En parallèle à ce travail, l'équipe d'E. Bertrand avait montré l'association de la protéine NUFIP avec la protéine 15.5K, l'orthologue de la protéine Snu13p. Sur la base de l'existence d'une courte séquence commune aux protéines Rsa1p et NUFIP, que nous avons dénommée PEP, il a été proposé que cet élément de séquence

constitue un site d'interaction avec la protéine Snu13p/15.5K. Lors de mon stage de Master au sein du laboratoire, j'ai vérifié cette hypothèse en collaboration avec B. Rothé. En début de thèse j'ai identifié les déterminants de l'interaction entre les protéines Snu13p et Rsa1p. A ce moment-là, certains facteurs, comme les hélicases Rvb1p/TIP49 et Rvb2p/TIP48 ou les protéines Pih1p et Bcd1p avaient été identifiés par d'autres auteurs, comme pouvant jouer un rôle dans le processus d'assemblage et le contrôle de la qualité des snoRNP à boîtes C/D. Afin d'apporter des réponses sur les mécanismes d'interaction entre les protéines constituantes des snoRNP et leurs facteurs d'assemblage et de décrypter les mécanismes d'assemblage, nous nous sommes fixés comme objectif de caractériser structuralement les constituants et complexes formés par les facteurs d'assemblage et les protéines constituant les snoRNP à boîtes C/D chez *S. cerevisiae*.

- Pour cela, nous devions d'abord définir le réseau d'interaction qui peut s'établir entre les différentes protéines impliquées. Sur la base de données antérieures et de résultats obtenus au laboratoire, nous avons mis en place un système de co-expression hétérologue chez *Escherichia coli*. Ce système, implanté sur une plateforme robotisée TECAN, nous a permis de réaliser un crible d'interaction à haut débit entre les différentes protéines d'intérêt et, pour la majorité des nouveaux facteurs étudiés, d'identifier leurs domaines d'interaction. Cette démarche a été particulièrement utilisée pour analyser l'association entre les protéines Rsa1p et Snu13p. Par mutagenèse dirigée nous avons ensuite pu identifier les acides aminés impliqués dans l'interaction. Un modèle d'association entre les 2 partenaires a ainsi pu être proposé.

- En plus de décrire les interactions qui s'établissent entre les protéines, nous avons également cherché à purifier des complexes protéiques possédant des caractéristiques de solubilité, d'homogénéité et de pureté compatibles avec des études structurales. Pour cela, nous avons utilisé l'approche de co-expression que nous avons couplée avec une stratégie de protéolyse ménagée afin de ne conserver au sein des complexes, que les domaines structurés des différentes protéines ; en particulier ceux impliqués dans les interactions. L'optimisation des conditions de purification et des tampons a été réalisée par une analyse DLS (*« Dynamic Light Scattering »*), nous permettant également de définir la taille et l'homogénéité des échantillons obtenus.

- Afin d'obtenir des informations structurales sur les protéines et les complexes caractérisés, une double approche par Résonance Magnétique Nucléaire et radiocristallographie a été utilisée. En effet, le laboratoire AREMS dispose d'un accès privilégié à un spectromètre R.M.N. Bruker DRX 600 MHz équipé d'une cryosonde TCI, ainsi qu'à un diffractomètre Bruker et à deux robots de cristallisation. L'appartenance de l'équipe au réseau européen EURASNET nous a aussi permis de collaborer avec l'ETH de Zurich qui a des compétences techniques et un équipement exceptionnel dans le domaine de la R.M.N.

L'ensemble de ce travail de thèse a été réalisé en étroite collaboration avec Benjamin Rothé, doctorant sous la direction de B. Charpentier et dont l'objectif était d'obtenir des données fonctionnelles sur la machinerie dédiée à l'assemblage des snoRNP à boîtes C/D. Cette collaboration nous a permis de valider nos données structurales par des approches *in vivo*. De notre côté nous leurs avons fait bénéficier de la production de complexes et protéines purifiés. Ce travail de thèse a été mené à travers différentes collaborations : avec l'équipe de Christophe Romier (IGBMC, Strasbourg) pour l'approche par co-expression, le laboratoire de Frédéric Allain (ETH, Zurich) pour la partie structurale par RMN, le groupe de Vladimir Pena (Max Planck Institute, Göttingen) pour la protéolyse ménagée, le laboratoire d'Alain Van Dorssealer (IPHC, Strasbourg) pour les analyses en spectrométrie de masse et les équipes d'Edouard Bertrand (IGM, Montpellier) et de Solange Moréra (LEBS, Gif-sur-Yvette) qui partageaient des contrats ANR avec nous sur cette thématique.

RESULTATS

&

DISCUSSION

CHAPITRE 1

La co-expression chez Escherichia coli : un outil de biologie moléculaire et structural

1 Présentation des travaux

L'étude structurale de la biogenèse des snoRNP à boîtes C/D de *S. cerevisiae* implique l'obtention des protéines cœur de la particule, c'est-à-dire les protéines Snu13p, Nop1p, Nop56p, Nop58p et Rrp9p pour la snoRNP U3, ainsi que les facteurs d'assemblage comme Rsa1p, Hit1p ,yBcd1p, Tah1p, Pih1p, Rvb1p et Rvb2p sous une forme soluble. Les quelques protéines qui sont produites sous forme soluble chez *E. coli* ont déjà fait l'objet d'études structurales. Lorsque j'ai débuté ce travail, la plupart des protéines qui restaient à étudier posaient des problèmes de solubilité et de stabilité qui rendaient difficile leur étude structurale. Dans ce domaine, la définition de domaines protéiques qui pourraient être solubles sur la base de prédiction de solubilité et/ou de structures secondaires n'apporte pas toujours de solutions efficaces.

Comme nous avons pu le voir dans la partie introductive, il existe un réseau d'interaction entre les protéines impliquées dans la biogenèse des snoRNP à boîtes C/D et aussi entre ces dernières et les protéines cœurs de la particule. Ainsi, un criblage à haut débit par double hybride (Boulon *et al.*, 2008; Ito *et al.*, 2001; Zhao *et al.*, 2005) avait permis de commencer à définir ce réseau d'interactions (Figure 42). Cependant les données obtenues ne permettaient pas d'affirmer que toutes les interactions observées étaient directes. Afin de répondre à cette question mais aussi pour avoir une vision plus exhaustive du réseau d'interactions, ainsi que pour apporter des solutions au problème d'insolubilité de certaines des protéines étudiées, nous avons mis en place une approche de co-expression des différents partenaires protéiques chez *Escherichia coli*.

Ce type d'approche avait été développé par l'équipe de C. Romier à l'IGBMC de Strasbourg (Diebold *et al.*, 2011). Elle repose sur la capacité de certaines protéines à former des complexes stables et solubles avec leurs partenaires au sein du système d'expression hétérologue *E. coli*. Le complexe protéique ainsi formé est une nouvelle entité avec des propriétés physico-chimiques différentes de celle de ses composants. De plus, certains

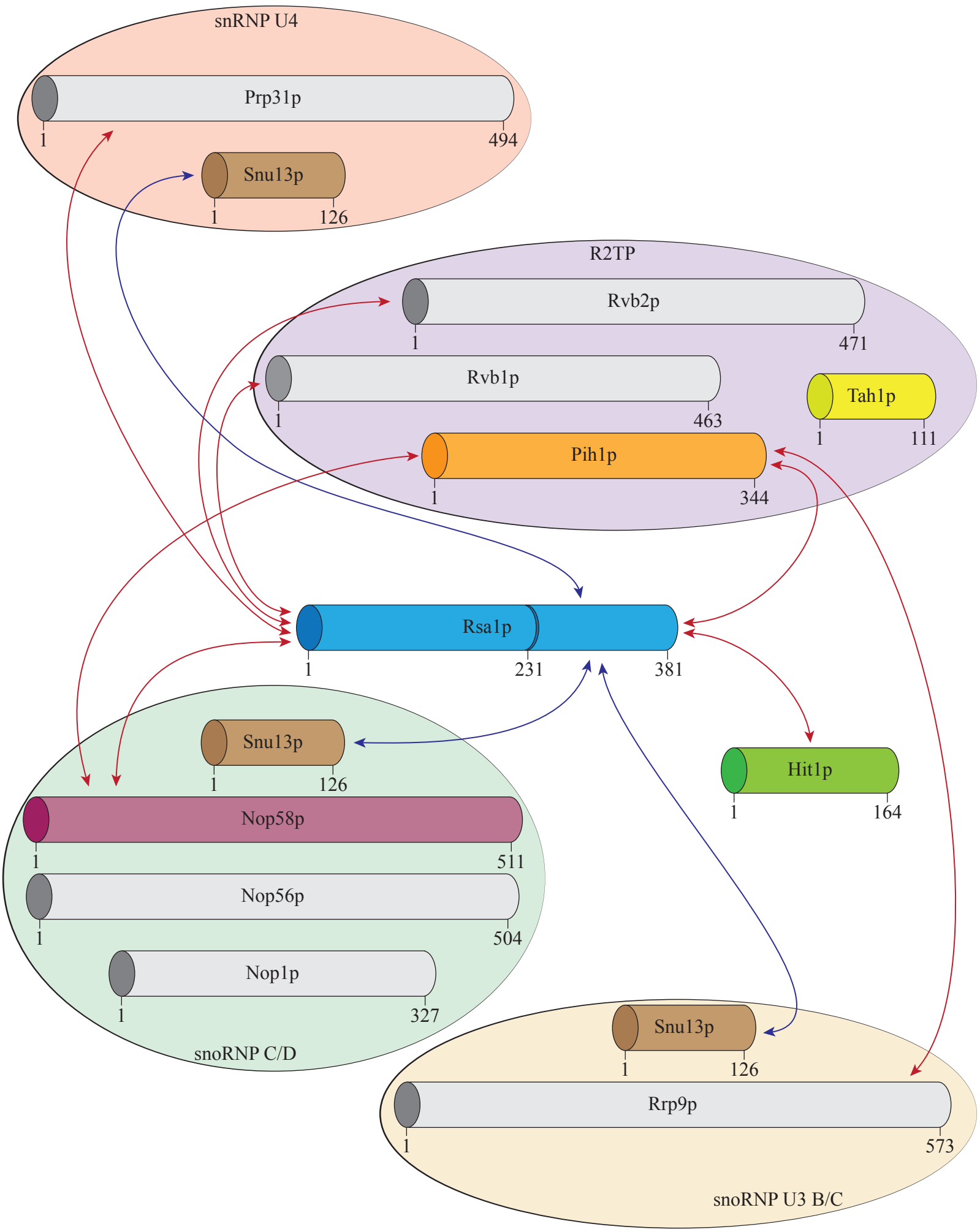


Figure 42. **Bilan des interactions entre les protéines constitutives des RNP et les facteurs d'assemblage connues au début de la thèse.** Les protéines sont représentées sous forme de cylindre dont la longueur est proportionnelle à leur taille en acides aminées. Les bornes des différents domaines des protéines sont indiquées par les numéros des résidus. Les protéines en couleurs correspondent aux protéines qui ont fait l'objet d'une étude plus appronfondie lors de la thèse. Les interactions détectées par double hybrides sont indiquées par des flèches rouges et celles connues pour être directes par des flèches bleue. D'après Boulon *et al*., 2008; Gonzales *et al*., 2005; Ito *et al*., 2001.

domaines protéiques peuvent se structurer lors de l'association avec leurs partenaires, par ce que l'on nomme le repliement induit ou *« induced folding »*.

La co-expression de protéines de levure chez *E. coli* permet d'obtenir une information sur les possibilités de formation d'interactions directes entre partenaires. Elle peut aussi renseigner sur la stabilité et la spécificité de l'interaction et permet aussi la définition de domaines minimaux d'interaction (Fribourg *et al.*, 2001).

1.1 Mise en place au laboratoire

Pour chaque candidat, que ce soit la protéine entière (FL ou *« Full Lenght »*) ou seulement un domaine de cette protéine, la première étape consistait à insérer sa séquence codante au sein de vecteur d'expression compatibles avec *E. coli*. Ces vecteurs sont de deux types selon qu'ils permettent d'exprimer la protéine de façon native ou fusionnée à une étiquette permettant ainsi de faciliter sa purification. Pour se faire, chaque plasmide possède la même casette de clonage (ou MCS), ce qui permet un sous-clonage rapide dans l'ensemble des vecteurs disponibles.

1.1.1 Les vecteurs pnEA-tH et pnEA-3CH

Initialement, les premières constructions ont été établies dans le vecteur pnEA-tH (Figure 43). Ce vecteur fait partie de la série des vecteurs pET-MCN (pET Multi-Cloning and expressioN) (Diebold *et al.*, 2011). Il s'agit de vecteurs qui ne possèdent qu'un seul promoteur pour l'ensemble des séquences codantes qui y sont insérées. On obtient donc une unité transcriptionnelle polycistronique. Le vecteur pnEA-tH dérive du vecteur commercial pET 15b, qui porte une cassette de résistance à l'ampicilline, un promoteur T7 et une origine de réplication de pBR322. Il tire son nom de la série à laquelle il appartient (pET-MCN = « n »), du vecteur dont il dérive (pET = « E »), de la casette de résistance qu'il porte (ampicilline = « A »), de la position de l'étiquette (N-terminal = « - »), du site de la protéase permettant la libération de l'étiquette (Thrombine = « t ») et de l'étiquette de fusion (poly-histidine = « H »). Ainsi le plasmide (= « p ») pnEA-tH est un vecteur permettant d'exprimer la protéine d'intérêt avec une étiquette poly-histidine (6xHis) à son extrémité N-terminale avec un site d'hydrolyse à la Thrombine. Ce vecteur est inductible par l'IPTG et la sélection des clones transformés se fait grâce à l'ampicilline.

Par la suite, au laboratoire, nous avons substitué le site de clivage par la Thrombine du vecteur pnEA-tH par un site de clivage par la protéase PreScission (ou 3C) qui est issue du

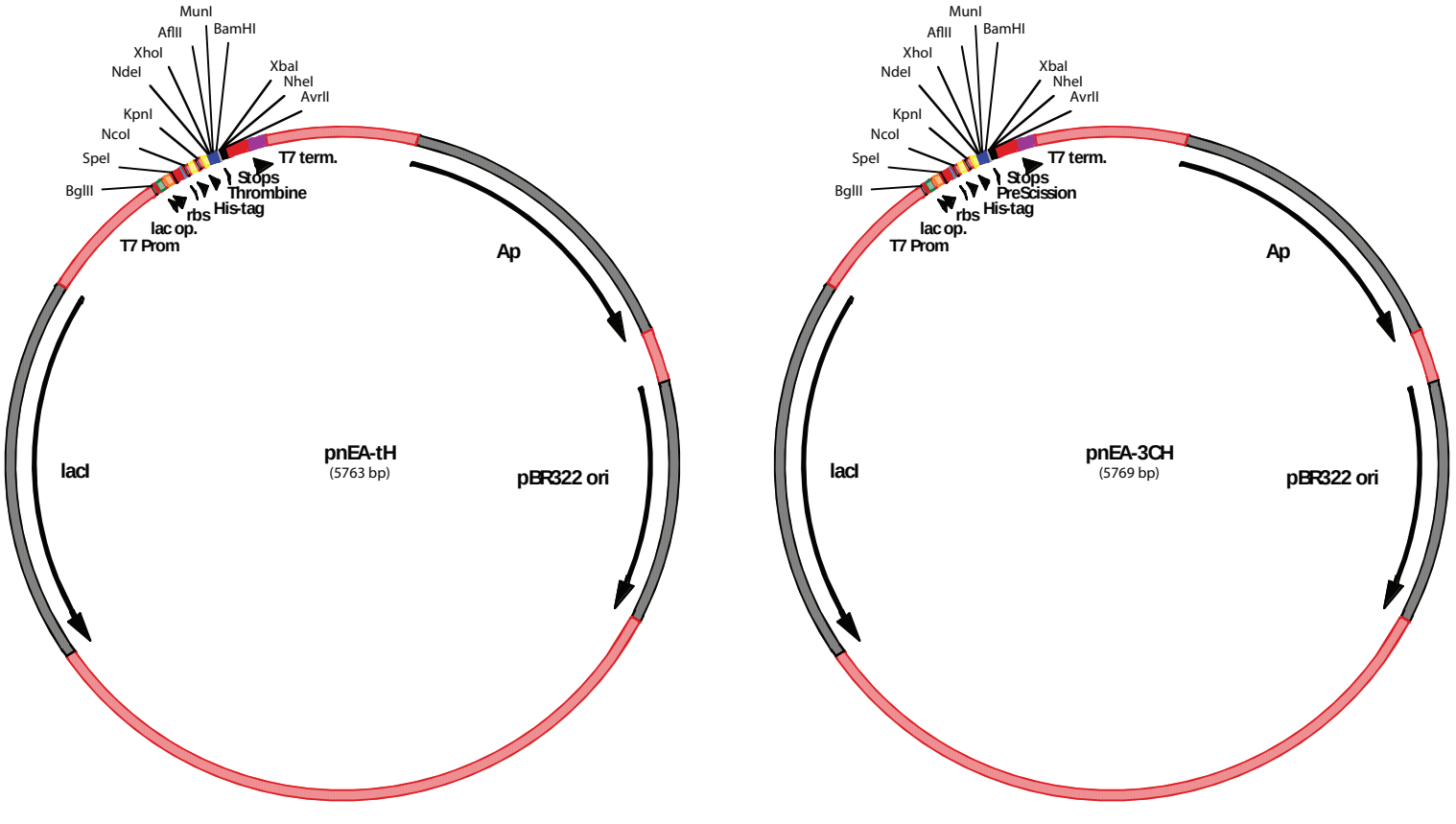


Figure 43. **Représentation schématique des plasmides pnEA-tH, pnEA-3CH, pnCS et pnYC.** D'après Diebold *et al*, 2011.

rhinovirus B humain. La protéase PreScission possède une spécificité pour son site, plus importante, que celle de la Thrombine et sa température d'activité optimum est de 4°C, ce qui est un avantage dans la manipulation de protéines eucaryotes, connues pour leur sensibilité à la dégradation et aux chocs thermiques. Nous avons ainsi créé le vecteur pnEA-3CH (Figure 43) en accord avec C. Romier qui avait construit le vecteur pnEA-tH initial.

1.1.2 Les vecteurs pnCS et pnYC

Les deux vecteurs pnCS et pnYC font également partie de la série pET-MCN mais permettent l'expression des protéines d'intérêt sans étiquette. Ils tirent également leur nom du vecteur duquel ils dérivent (pCDF-DuET = « C » et pACYC11b = « Y ») et de la casette de résistance qu'ils portent (spectinomycine = « S » et chloramphénicol = « C »). Leur origine de réplication leur est propre, CDF pour le vecteur pnCS et p15a pour le vecteur pnYC. Ces origines de réplication étant compatibles les unes avec les autres, les plasmides pnEA peuvent être co-transformés avec les plasmides pnCS et pnYC au sein de la même cellule d'*E. coli* de manière stable. Pour finir, ils possèdent également un promoteur T7 inductible par l'IPTG (Figure 43).

1.1.3 Clonage au sein des vecteurs

Le clonage des séquences codantes d'intérêt a été réalisé par l'utilisation des sites de restriction *Nde* I et *BamH* I (Figure 44). En effet, le site *Nde* I, 5'-CATATG-3' est le premier site de restriction de la cassette de clonage. Dans les vecteurs permettant la fusion de l'étiquette 6xHis, la séquence de ce site permet d'incorporer un doublet d'acides aminés NH_2-HisMet-COOH en amont de la séquence de la protéine produite. La méthionine ainsi introduite remplace celle à l'extrémité de l'ORF (*« Open reading frame »*) cloné. La position de ce site de restriction dans les vecteurs pnCS ou pnYC qui ne conduisent pas à l'ajout de l'étiquette 6xHis permet à la séquence 5'-ATG-3' du site *Nde* I de faire office de codon initiateur pour la traduction. Ainsi avec ces plasmides, la protéine est exprimée sans acide aminé supplémentaire. Le nombre d'acides aminés ajoutés et aussi limité dans le cas des vecteurs pnEA utilisé pour la fusion (Figure 44). Lorsqu'un site *BamH* I est présent au sein de la séquence codante d'intérêt, nous avons utilisé à sa place le site *Bgl* II (Figure 44), en choisissant en conséquence les oligodéoxynucléotides nécessaire à l'amplification par PCR (*« Polymerase Chain Reaction »*). En effet, les sites *BamH* I et *Bgl* II sont compatibles et permettent de conserver le clivage par *BamH* I du vecteur d'expression. Il a été nécessaire de

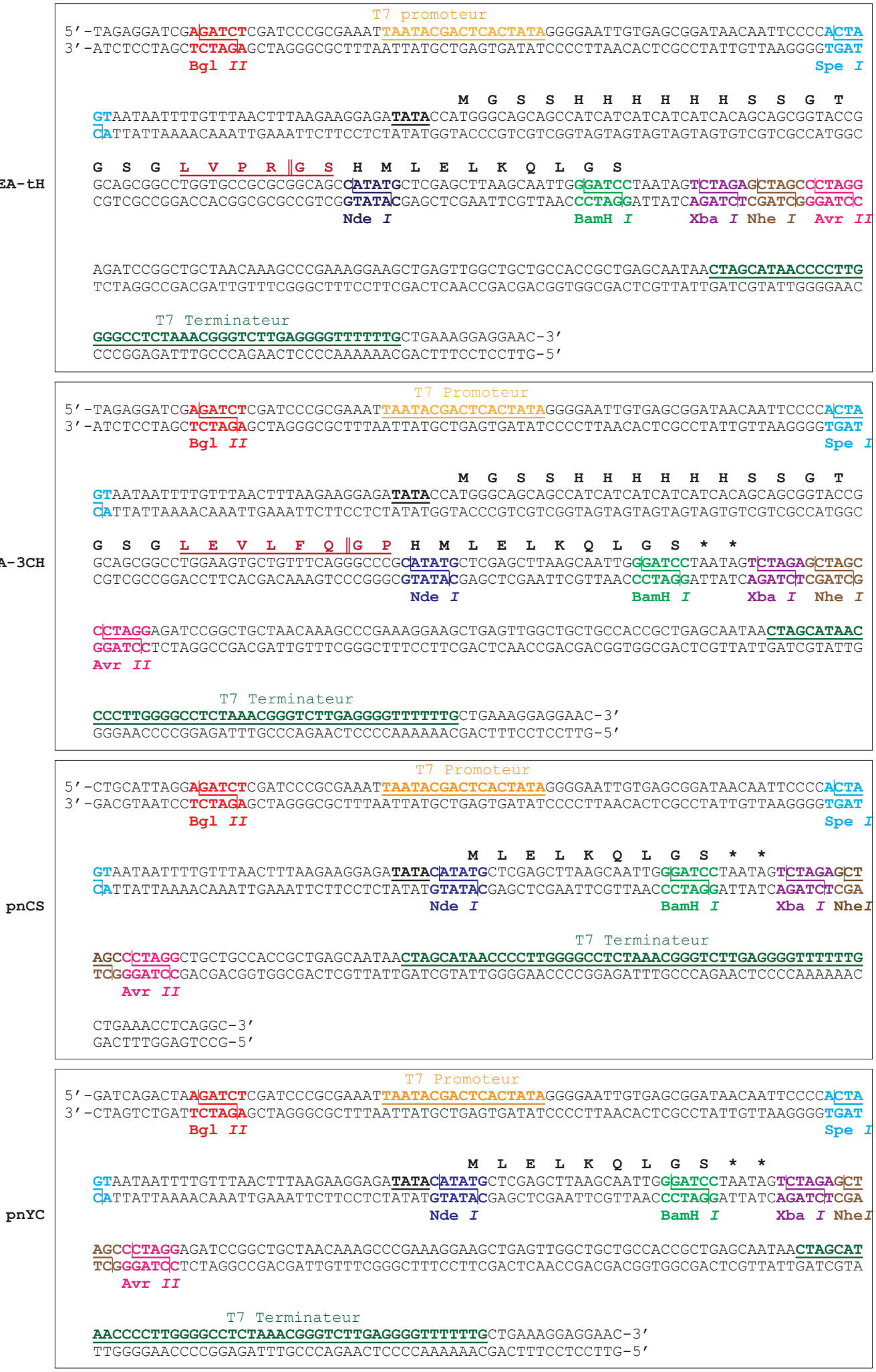


Figure 44. **Unité transcriptionnelle des vecteurs d'expression pnEA-tH, pnEA-3CH, pnCS et pnYC.** Les sites de restrictions et les séquences «T7 Promoteur» et «T7 Terminateur» sont symbolisés par des couleurs distinctes. Le site de la protéase est également représenté et la position du clivage est symbolisé par «‖». D'après Diebold *et al*, 2011.

réaliser des mutations ponctuelles silencieuses de la séquence codante pour éliminer la présence des sites *Nde* I et/ou la concomitance des sites *BamH* I et *Bgl* II.

1.1.4 La concaténation

Les vecteurs ont été conçus de manière à pouvoir rapidement, par un système de « copier-coller moléculaire » utilisant des sites de restrictions, associer plusieurs séquences codantes au sein d'un seul et même vecteur, quelques soit le vecteur (Figures 44 et 45). Dans notre cas, l'ensemble des gènes d'intérêt ont ainsi été insérés dans une unité transcriptionnelle de type polycistronique. Nous avons mis à profit cette possibilité lors d'une identification d'une interaction entre deux partenaires (ou plus). Le fait de réunir les séquences codantes de plusieurs protéines dans un seul plasmide simplifie l'étape de transformation bactérienne et diminue le nombre d'antibiotiques qui peut limiter la croissance bactérienne s'il est élevé. Cette approche permet de contourner le problème que pose le nombre limité de vecteurs compatibles dans *E. coli*, le nombre de partenaires serait ainsi limité (à trois dans notre cas) si l'on introduisait seulement une séquence codante par plasmide. Par cette méthode, il est aussi très facile de transposer l'ensemble de l'unité transcriptionnelle d'un plasmide à un autre (par exemple de pnCS à pnYC), ce qui permet de changer la résistance aux antibiotiques sans repasser par les premières étapes de clonage.

1.1.5 Le principe

L'étiquette poly-histidines a la capacité de fixer les cations divalents tels que le Nickel ou le Cobalt. Cette caractéristique est mise en œuvre lors de la chromatographie d'affinité sur métal immobilisé encore appelée IMAC (*« Immobilized Metal Affinity Chromatography »*) (Figure 46). Si la protéine porteuse de l'étiquette poly-histidines est contenue dans la fraction soluble du lysat cellulaire, elle est retenue sur les billes de sépharose sur lesquelles des ions Nickel ou Cobalt sont gréffés. Si la protéine porteuse de l'étiquette 6xHis est associée à une ou plusieurs protéines avec lesquelles elle a été co-exprimée, le complexe formé est retenu sur la résine (Figure 46). Afin de limiter les problèmes d'encombrements stériques ou les associations aspécifiques qui empêcheraient la formation du complexe, il est nécessaire de répéter les tests en fusionnant successivement l'étiquette 6xHis à chacun des partenaires du complexe. Il est également possible de faire varier la position de l'étiquette (côté N-terminal ou C-terminal de la protéine) et même de changer la nature de l'étiquette. Pour notre part nous avons seulement utilisé l'étiquette 6xHis mais nous avons systématiquement réalisé les expériences en testant l'effet de la position de l'étiquette 6xHis à chacun des partenaires

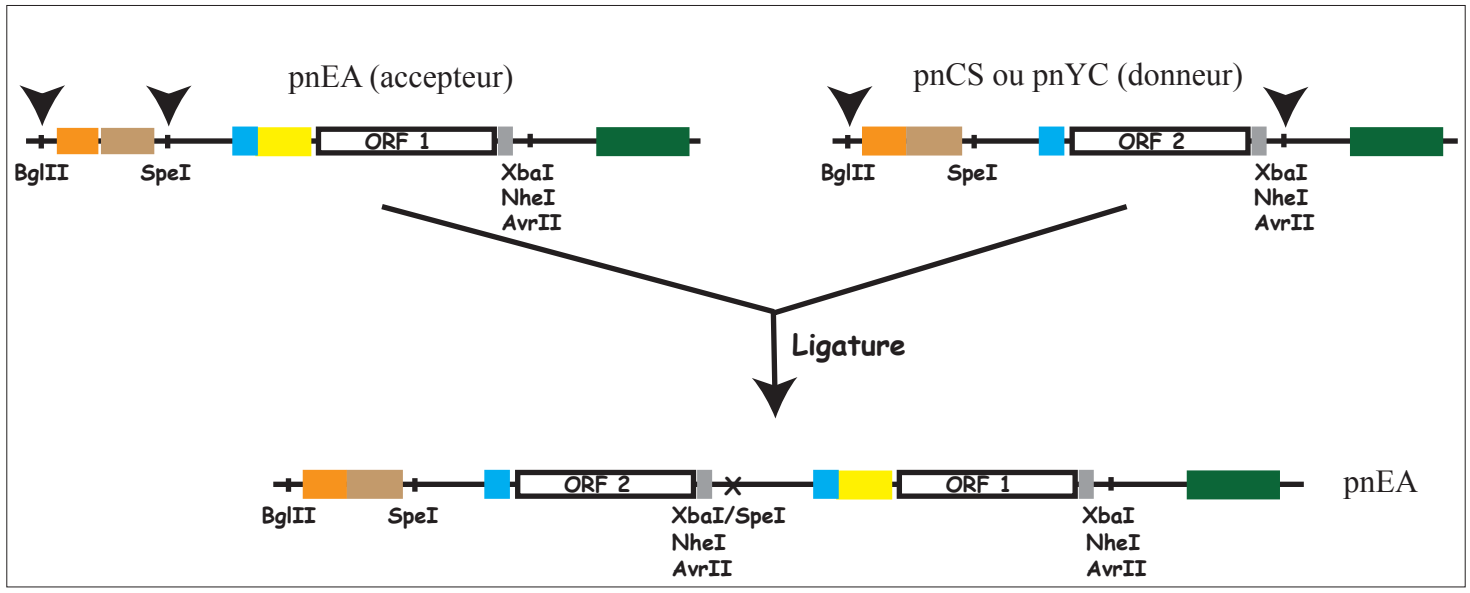


Figure 45. **Stratégie que nous avons utilisé pour la co-expression de protéines chez *E. coli***. Chaque plasmide (pnEA, pnCS et pnYC) peut être soit donneur soit accepteur, ce qui permet une grande souplesse dans le choix du plasmide accepteur. Par cette méthode, il est aussi très facile de déplacer l'ensemble des séquences codantes d'un plasmide à un autre, ce qui permet ainsi le changement rapide de résitance à un antibiotique. D'après Diebold *et al*, 2011.

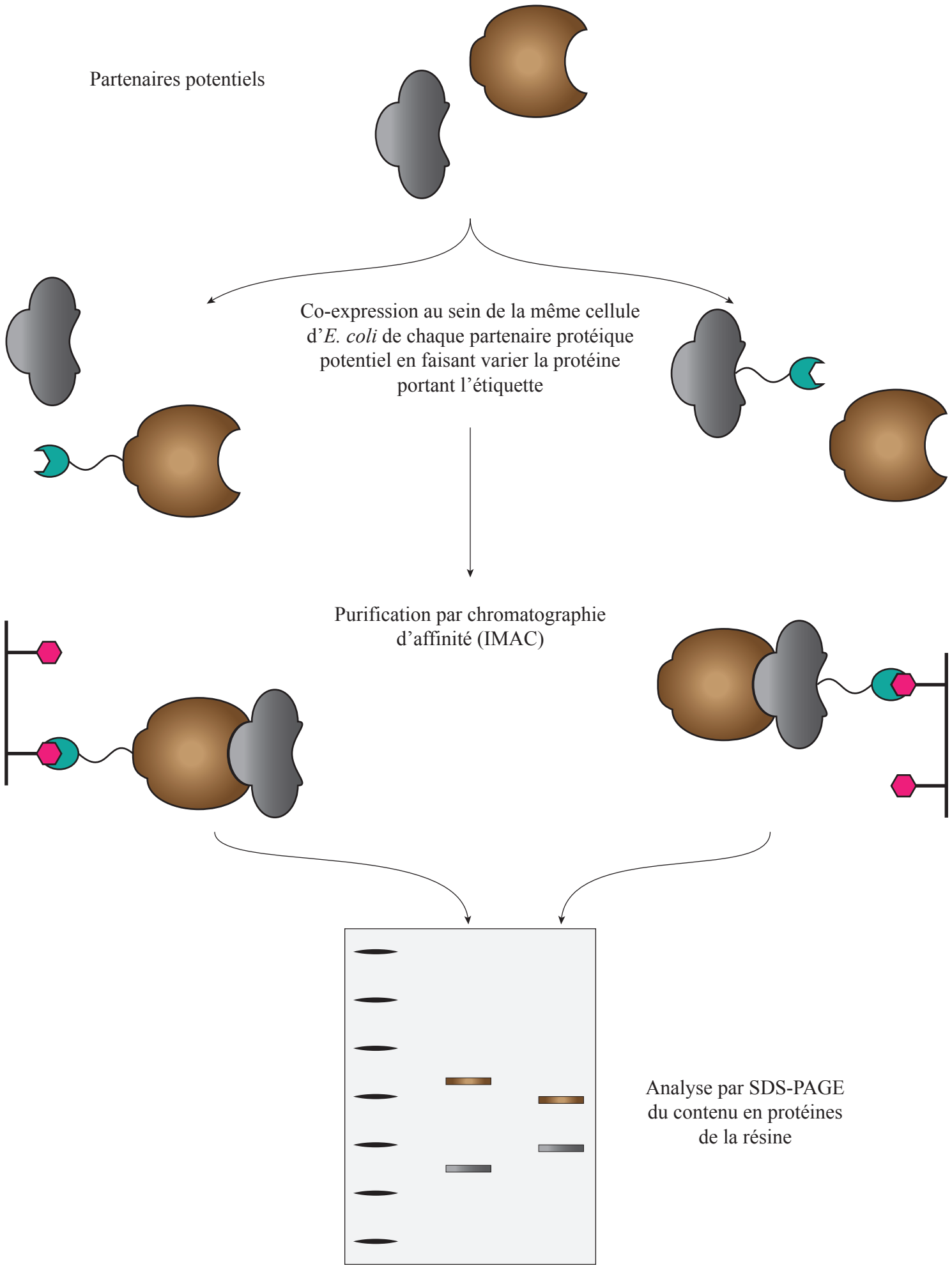


Figure 46. **Schéma du principe de l'analyse d'interactions à grande échelle par la méthode de co-expression chez *E. coli***. Au laboratoire, nous avons transposé ce mode opératoire sur une plateforme robotisée TECAN.

potentiels. De plus, chaque crible d'interaction a été réalisé dans deux tampons de concentration saline différente (50 mM et 400 mM NaCl), soit quatre conditions différentes testées pour une interaction potentielle donnée.

Lorsqu'une interaction entre deux partenaires est détectée et validée, il est ensuite possible de concaténer les séquences codantes des deux protéines au sein d'un même vecteur (au choix) et de recommencer un cycle de criblage avec d'autres protéines (Figure 47).

Une grande partie des essais d'interaction a été réalisée en collaboration avec C. Romier à l'IGBMC de Strasbourg. Cette collaboration m'a permis d'acquérir la maîtrise de la technologie et d'utiliser la plateforme robotisée TECAN permettant le criblage rapide de nombreuses interactions, avant que nous acquérions un robot à Nancy.

1.2 Les séquences codantes sélectionnées

Comme précisé auparavant, sur la base des résultats obtenus au laboratoire par l'approche du double hybride et aussi des résultats des équipes de Y. Sakaki et W.A. Houry (Ito *et al.*, 2001; Zhao *et al.*, 2005), nous avons sélectionné les couples de partenaires potentiels étudiés par co-expression. Afin d'obtenir des informations sur les domaines impliqués dans l'association des protéines, mais aussi pour éliminer certaines régions potentiellement instables, différents domaines ont été définis pour certaines des protéines sur la base de plusieurs paramètres : *i)* des prédiction de structure secondaire (Dosztanyi *et al.*, 2005a; Dosztanyi *et al.*, 2005b), *ii)* les homologies par rapport à des séquences de structures connues et, *iii)* les conservations de séquences à travers les espèces.

2 Résultats et discussion

Afin d'avoir un témoin positif de la méthode, validant notre approche expérimentale, nous avons utilisé un complexe connu et bien caractérisé *in vitro*, le complexe Tah1p-Pih1p (Eckert *et al.*, 2010; Zhao *et al.*, 2005; Zhao *et al.*, 2008).

2.1 Résultats obtenus lors d'une première série de criblage par co-expression

Dans un premier temps, trois grandes séries de co-expressions (Figures 48 à 58) ont été réalisées avec les plasmides pnCS et pnEA-tH, que nous avions à notre disposition à

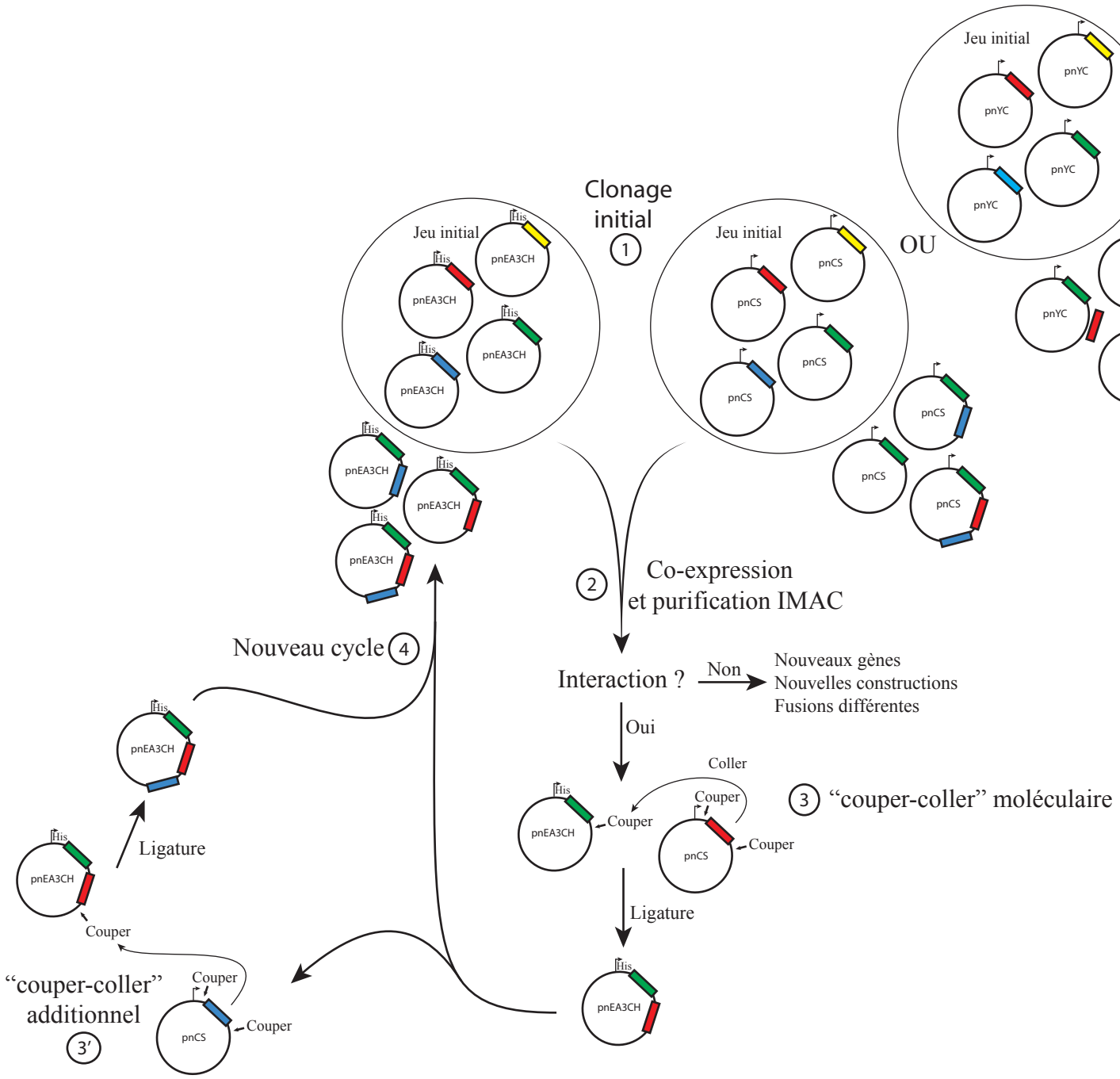


Figure 47. **Schéma du principe de la stratégie de co-expression chez *E. coli* basé sur l'utilisation des vecteurs pnEA-3CH, pnCS et pnYC et sur la concaténation**. La première étape consistait à cloner l'ensemble des séquences codantes des protéines ou fragments de protéines cibles dans les vecteurs de co-expression. La co-expression et l'analyse par gel SDS-PAGE des protéines retenues sur la colonne IMAC permettait de sélectionner les partenaires capablent d'interagir et donc de former un complexe (étape 2). Dans ce cas l'ensemble des séquences codantes peut être fusionné au sein d'un même plasmide par concaténation (étape 3). A partir de cette nouvelle construction, un nouveau cycle peut être entrepris (étape 4). D'après Diebold *et al*, 2011.

Strasbourg. Ce n'est que par la suite que nous avons modifié le plasmide pnEA-tH en convertissant sa séquence codant un site de clivage par la Thrombine en une séquence codant un site de clivage par la PreScission protéase. Cette première campagne nous a permis tout d'abord de valider le protocole expérimental par l'obtention du complexe hétérodimérique Pih1p-Tah1p (Figure 51, pistes 39, 40 et 74, 81).

Pour tester la validité de nos interprétations, la nature des protéines retenues sur les billes TALON^{TM} et visualisées sur les gels de polyacrylamides dénaturants a été identifiée par spectrométrie de masse en conditions dénaturantes. Pour cela, les bandes correspondantes ont étés excisées, les protéines contenues dans ces bandes hydrolysées par la trypsine et les produits d'hydrolyses analysés par spectrométrie de masse en conditions dénaturantes au laboratoire LSMBO à Strasbourg par J.M. Saliou dans l'équipe de S. Sanglier. Cette technique permet d'identifier la séquence d'une série de peptides issue de la protéolyse de la protéine présente dans la bande de gel et de valider ou non la présence sur la résine des deux partenaires recherchés.

Les bandes analysées lors des expériences de co-expression sont reprises dans les tableaux des figures 49 à 58 où est mentionné la protéine majoritairement détectée. Les bandes correspondantes sont indiquées par un astérisque sur les photographies des différents gels de polyacrylamides. L'ensemble des données obtenues lors de l'analyse par spectrométrie de masse en conditions dénaturantes est présenté dans les tableaux 3 à 10.

2.1.1 Première série d'expériences

Un premier constat est que d'une manière générale, nous observions plus de protéines retenues sur la résine dans la condition de haute salinité (NaCl 400 mM), qu'à basse concentration saline (NaCl 50 mM) (Figures 48 à 51 et tableaux 3 à 5).

Par cette approche, nous avons pu confirmer l'association directe entre Snu13p, une des protéines cœur des RNP à boîtes C/D, et le facteur d'assemblage qui venait d'être mis récemment en évidence au laboratoire, la protéine Rsa1p (Boulon *et al.*, 2008) (Pistes 9). La formation de ce complexe semblait sensible à l'identité de la protéine portant l'étiquette 6xHis. L'interaction était plus forte lorsque Snu13p portait l'étiquette 6xHis à son extrémité N-terminale (Piste 9 et 20). Cette expérience de co-expression, nous avait aussi permis de définir plus précisément le domaine de la protéine Rsa1p responsable de son association avec la protéine Snu13p (Pistes 10, 11 et 21, 22). En effet, les deux fragments C-terminaux de la

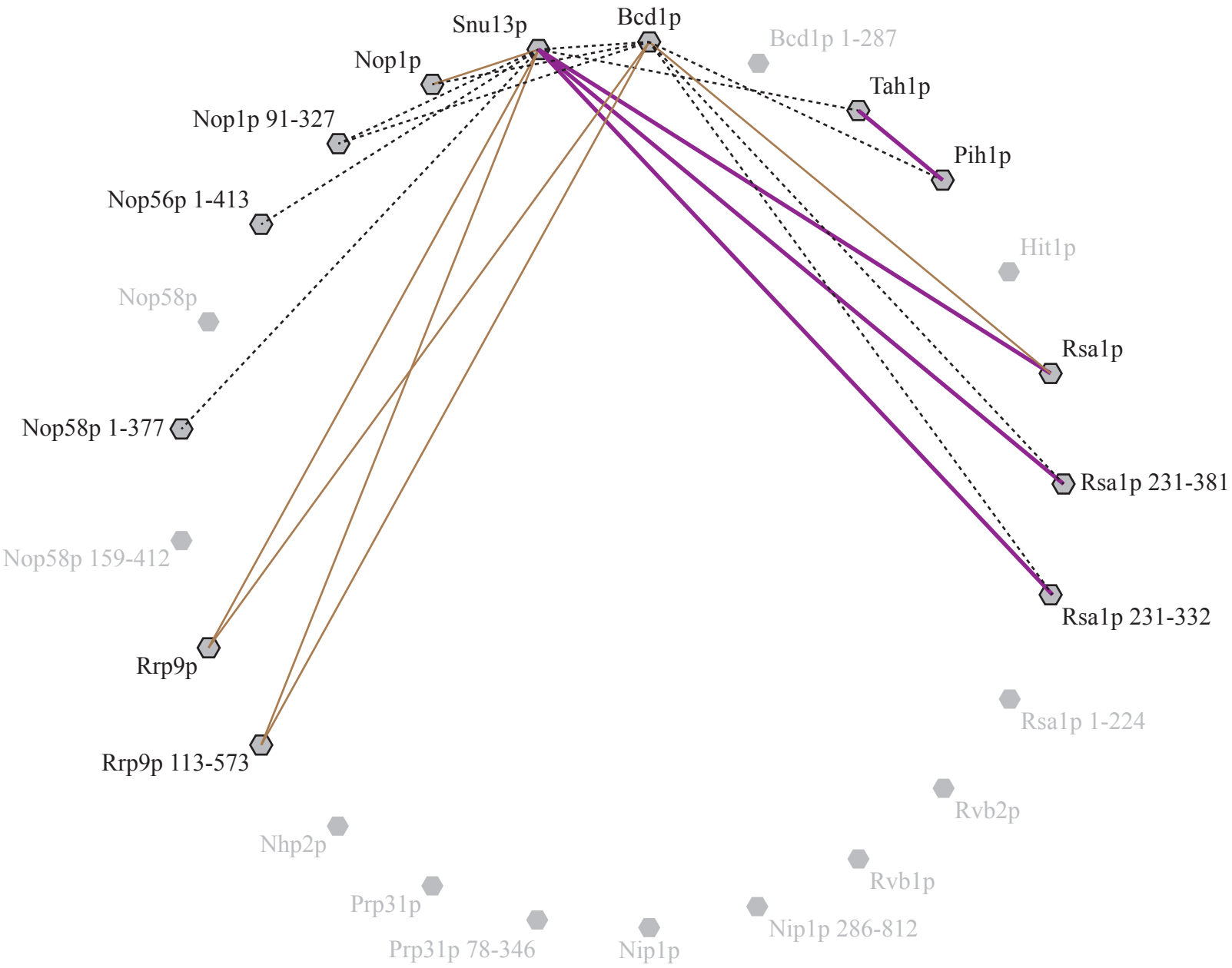
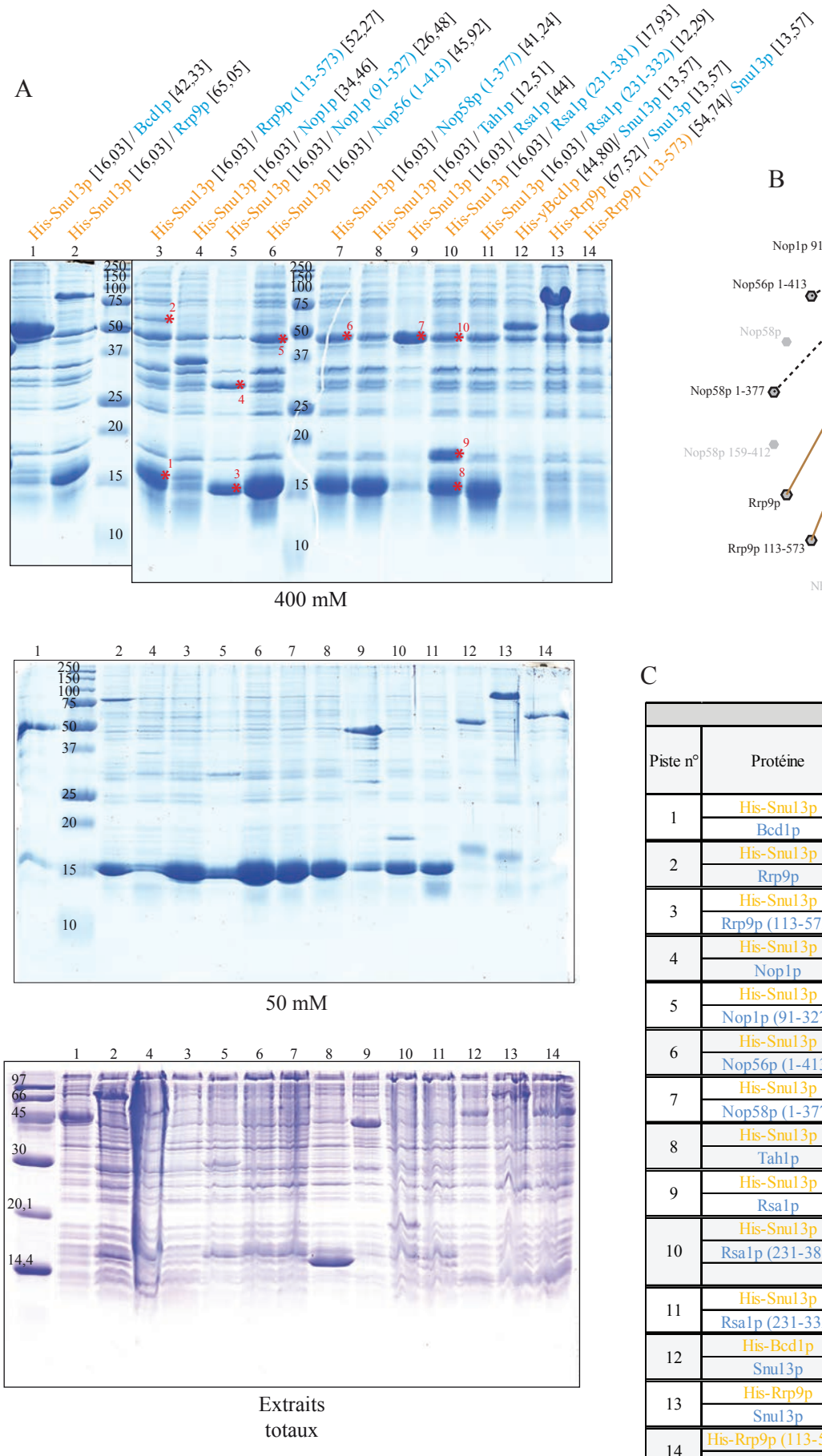
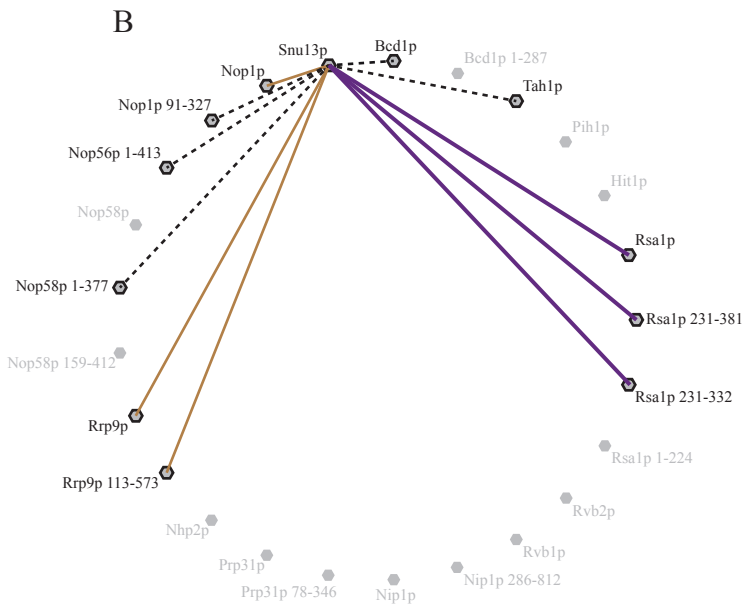


Figure 48. **Bilan des interactions testées dans la première série d'expériences de co-expression.** L'ensemble des protéines des expériences de co-expression sont indiquées. Les traits en pointillés symbolisent les interactions testées, les traits en brun les interactions nécessitant une vérification et les traits épais en violet celles qui ont été validées.

A



B



C

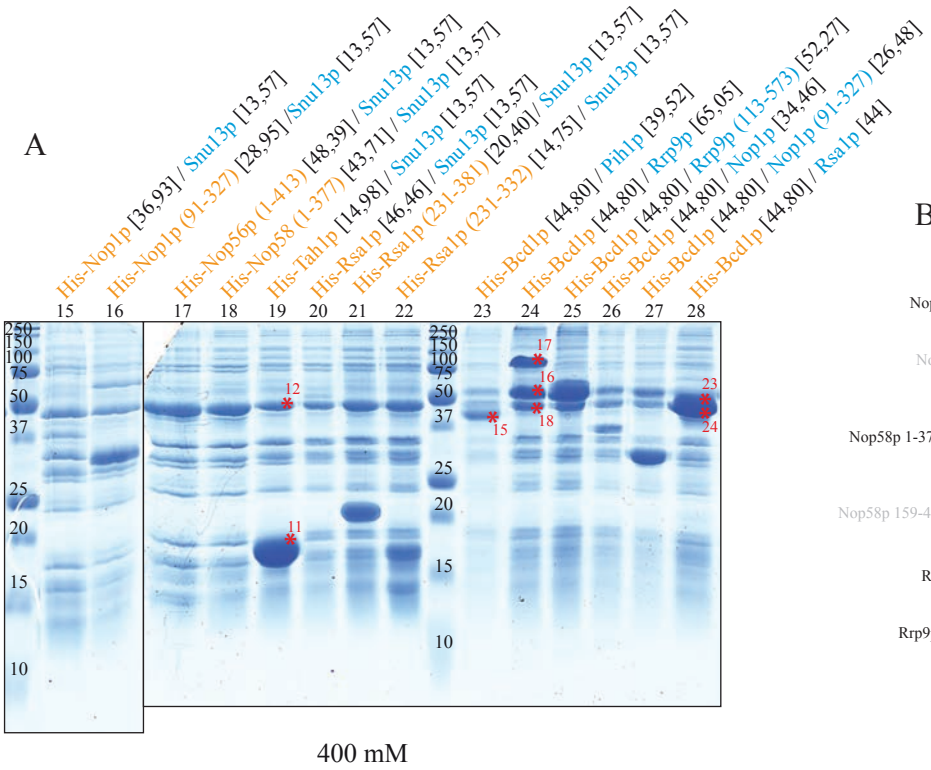
SDS-PAGE						Spectrométrie de masse	
Piste n°	Protéine	P.M. (kDa)	Expression	IMAC 50 mM	IMAC 400 mM	Bande n°	Protéine majoritairement détectée
1	His-Snu13p	16,03		Oui			
	Bcd1p	42,33	Oui	Oui	Oui		
2	His-Snu13p	16,03	Oui	Oui	Oui		
	Rrp9p	65,05	Oui	Oui	Oui		
3	His-Snu13p	16,03		Oui	Oui	1	Snu13p
	Rrp9p (113-573)	52,27			Oui	2	Rrp9p
4	His-Snu13p	16,03		Oui			
	Nop1p	34,46			Oui		
5	His-Snu13p	16,03	Oui	Oui	Oui	3	Snu13p
	Nop1p (91-327)	26,48	Oui	Oui	Oui	4	Nop1p
6	His-Snu13p	16,03	Oui	Oui	Oui		
	Nop56p (1-413)	45,92				5	EFTu
7	His-Snu13p	16,03	Oui	Oui	Oui		
	Nop58p (1-377)	41,24				6	EFTu
8	His-Snu13p	16,03	Oui	Oui	Oui		
	Tah1p	12,51	Oui				
9	His-Snu13p	16,03		Oui			
	Rsa1p	44	Oui	Oui	Oui	7	Rsa1p
10	His-Snu13p	16,03	Oui	Oui	Oui	8	Snu13p
	Rsa1p (231-381)	17,93	Oui	Oui	Oui	9	Rsa1p
						10	EFTu
11	His-Snu13p	16,03	Oui	Oui	Oui		
	Rsa1p (231-332)	12,29		Oui	Oui		
12	His-Bcd1p	44,8	Oui	Oui	Oui		
	Snu13p	13,57	Oui				
13	His-Rrp9p	67,52	Oui	Oui	Oui		
	Snu13p	13,57					
14	His-Rrp9p (113-573)	54,74	Oui	Oui	Oui		
	Snu13p	13,57					

Figure 49. **Analyse par SDS-PAGE à 15 % des interactions testées lors de la première série d'expériences de co-expression.** (A) Les numéros des pistes sont indiqués au dessus des gels. Les bandes analysées par spectrométrie de masse sont marquées par une "*" et leur numéro est indiqué. Les extraits totaux correspondent aux sonicats avant centrifugation. Les concentrations en sels (50 ou 400 mM) des tampons de purification sont indiquées sous les gels. La souche d'*E. coli* utilisée est la souche BL21 (DE3) pRAR2. (B) Schéma représentant l'ensemble des couples de protéines utilisés. Les interactions détectées par co-expression qui ont été validées par la suite sont représentées en violet et celles nécessitant une vérifiaction en brun. (C) Tableau synthétisant les résultats obtenus par le fractionnement par SDS-PAGE et analyse en spéctrométrie de masse en conditions dénaturantes. Les protéines exprimées grâce aux vecteurs pnEA-tH et pnCS sont respectivement indiquées en orange et en bleu.

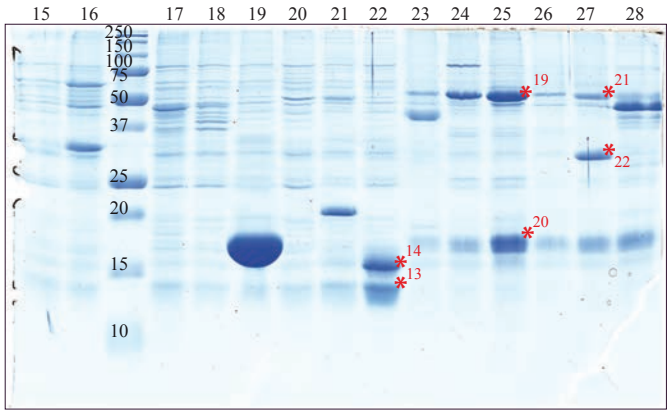
Bande n°	AC	Protéine	Taxonomie	Masse (Da)	pI	Couverture	Nombre de peptides
1	Snu13p_YEAST	13 kDa ribonucleoprotein-associated protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	13560	7.68	55%	77
	RL22_ECOLI	50S ribosomal protein L22 - E. coli (strain K12)	*E. coli*	12219	10.23	70%	37
	RL17_ECOLI	50S ribosomal protein L17 - E. coli (strain K12)	*E. coli*	14356	11.05	49%	27
	RS11_ECOLI	30S ribosomal protein S11 - E. coli (strain K12)	*E. coli*	13836	11.33	56%	15
2	Rrp9p_YEAST	Ribosomal RNA-processing protein 9 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	65014	5.55	51%	86
	RHLE_ECOLI	Putative ATP-dependent RNA helicase rhlE - E. coli (strain K12)	*E. coli*	49959	10.06	59%	33
	AGAL_ECOLI	Alpha-galactosidase - E. coli (strain K12)	*E. coli*	50624	5.52	43%	21
	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	53%	20
3	Snu13p_YEAST	13 kDa ribonucleoprotein-associated protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	13560	7.68	69%	84
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	25%	19
	RL17_ECOLI	50S ribosomal protein L17 - E. coli (strain K12)	*E. coli*	14356	11.05	37%	20
	RL22_ECOLI	50S ribosomal protein L22 - E. coli (strain K12)	*E. coli*	12219	10.23	56%	9
4	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	34444	10.24	52%	195
	RS2_ECOLI	30S ribosomal protein S2 - E. coli (strain K12)	*E. coli*	26727	6.61	65%	39
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	25%	13
	Snu13p_YEAST	13 kDa ribonucleoprotein-associated protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	13560	7.68	56%	11
5	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	84%	152
	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin OS=S. cerevisiae GN=NOP1 PE=1 SV=1	*S. cerevisiae*	34444	10.24	22%	10
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	25%	8
	Nop56p_YEAST	Ribosome biosynthesis protein SIK1 OS=S. cerevisiae GN=SIK1 PE=1 SV=1	*S. cerevisiae*	56829	8.96	13%	6
6	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	79%	151
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	25%	6
	ISCS_ECOLI	Cysteine desulfurase OS=E. coli (strain K12) GN=iscS PE=1 SV=1	*E. coli*	45061	5.94	13%	6
	Nop58p_YEAST	Nucleolar protein NOP58 OS=S. cerevisiae GN=NOP58 PE=1 SV=1	*S. cerevisiae*	56921	8.96	7%	3
7	Rsa1p_YEAST	Ribosome assembly 1 protein OS=S. cerevisiae GN=RSA1 PE=1 SV=1	*S. cerevisiae*	43971	5.77	59%	82
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	40%	21
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	25%	7
8	Snu13p_YEAST	13 kDa ribonucleoprotein-associated protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	13560	7.68	55%	83
	RL22_ECOLI	50S ribosomal protein L22 - E. coli (strain K12)	*E. coli*	12219	10.23	73%	32
	RL17_ECOLI	50S ribosomal protein L17 - E. coli (strain K12)	*E. coli*	14356	11.05	51%	26
	RS6_ECOLI	30S ribosomal protein S6	*E. coli*	15694	4.93	48%	15
9	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	27%	145
	RL13_ECOLI	50S ribosomal protein L13 - E. coli (strain K12)	*E. coli*	16009	9.91	74%	41
	RS7_ECOLI	30S ribosomal protein S7 - E. coli (strain K12)	*E. coli*	20007	10.36	45%	12
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	25%	10
10	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	69%	74
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	25%	12
	Rsa1p_YEAST	Ribosome assembly 1 protein OS=S. cerevisiae GN=RSA1 PE=1 SV=1	*S. cerevisiae*	43971	5.77	18%	7

Tableau 3. Résultats bruts de l'analyse par spectrométrie de masse en conditions dénaturantes des bandes de protéines purifiées par electrophorèse lors de la première série d'essais de co-expression. La protéine majoritaire est surlignée en vert ou gris selon qu'il s'agisse respectivement d'une protéine de levure ou d'*E. coli*.

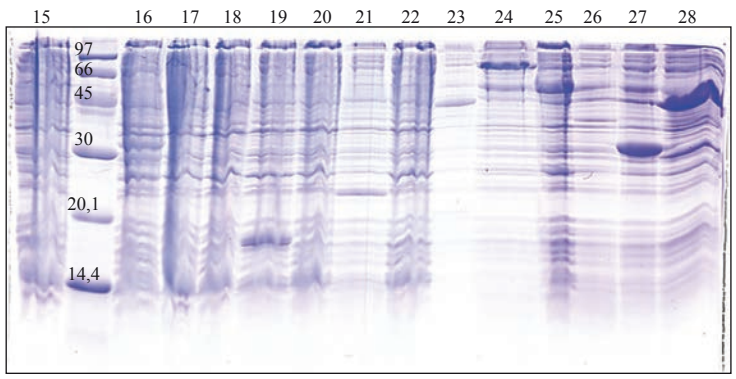
A



400 mM

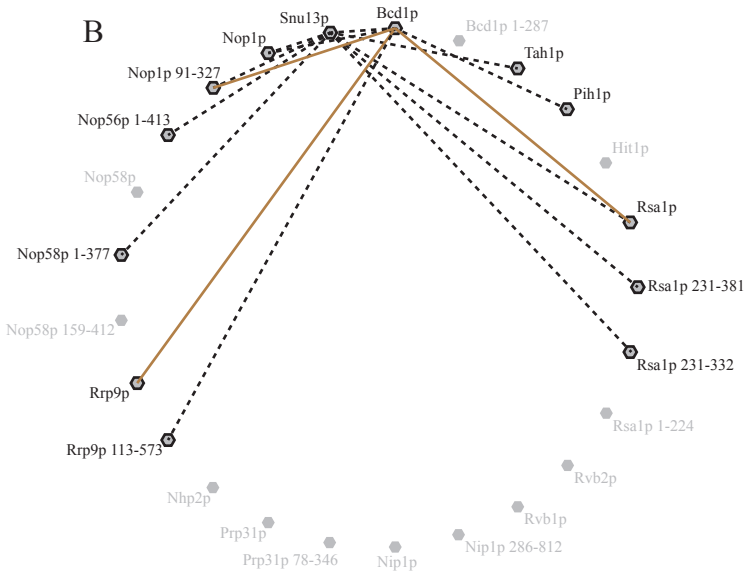


50 mM



Extraits totaux

B



C

SDS-PAGE						Spectrométrie de masse	
Piste n°	Protéine	P.M. (kDa)	Expression	IMAC 50 mM	IMAC 400 mM	Bande n°	Protéine majoritairement détectée
15	His-Nop1p	36,93			Oui		
	Snu13p	13,57					
16	His-Nop1p (91-327)	28,95	Oui	Oui	Oui		
	Snu13p	13,57					
17	His-Nop56p (1-413)	48,39					
	Snu13p	13,57					
18	His-Nop58p (1-377)	43,71					
	Snu13p	13,57					
19	His-Tah1p	14,98	Oui	Oui	Oui	11	Tah1p
	Snu13p	13,57	Oui				
						12	EFTu
20	His-Rsa1p	46,46					
	Snu13p	13,57	Oui				
21	His-Rsa1p (231-381)	20,4	Oui	Oui	Oui		
	Snu13p	13,57					
22	His-Rsa1p (231-332)	14,75	Oui	Oui	Oui	13	Rsa1p
	Snu13p	13,57	Oui	Oui	Oui	14	Snu13p
23	His-Bcd1p	44,8	Oui	Oui	Oui		
	Pih1p	39,52	Oui	Oui	Oui	15	Pih1p
24	His-Bcd1p	44,8	Oui	Oui	Oui	16	Bcd1p
	Rrp9p	65,05	Oui	Oui	Oui	17	Rrp9p
						18	EFTu
25	His-Bcd1p	44,8	Oui	Oui	Oui	19	Bcd1p
	Rrp9p (113-573)	52,27					
						20	Bcd1p
26	His-Bcd1p	44,8			Oui		
	Nop1p	34,46			Oui		
27	His-Bcd1p	44,8	Oui	Oui	Oui	21	Bcd1p
	Nop1p (91-327)	26,48	Oui	Oui	Oui	22	Nop1p
28	His-Bcd1p	44,8	Oui	Oui		23	Bcd1p
	Rsa1p	44	Oui	Oui	Oui	24	Rsa1p

Figure 50. **Analyse par SDS-PAGE à 15 % des interactions testées lors de la première série d'expériences de co-expression.** (A) Les numéros des pistes sont indiqués au dessus des gels. Les bandes analysées par spectrométrie de masse sont marquées par une "*" et leur numéro est indiqué. Les extraits totaux correspondent aux sonicats avant centrifugation. Les concentrations en sels (50 ou 400 mM) des tampons de purification sont indiquées sous les gels. La souche d'*E. coli* utilisée est la souche BL21 (DE3) pRAR2. (B) Schéma représentant l'ensemble des couples de protéines utilisés. Les interactions détectées par co-expression qui ont été validées par la suite sont représentées en violet et celles nécessitant une vérifiaction en brun. (C) Tableau synthétisant les résultats obtenus par le fractionnement par SDS-PAGE et analyse en spectrométrie de masse en conditions dénaturantes. Les protéines exprimées grâce aux vecteurs pnEA-tH et pnCS sont respectivement indiquées en orange et en bleu.

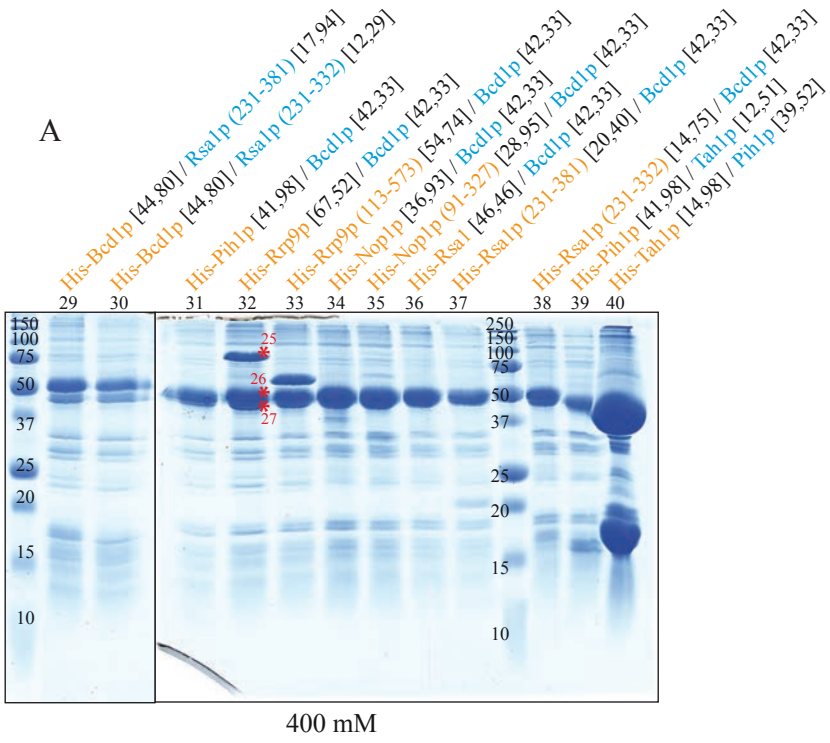
Bande n°	AC	Protéine	Taxonomie	Masse (Da)	pI	Couverture	Nombre de peptides
11	Tah1p_YEAST	TPR repeat-containing protein associated with Hsp90 OS=S. cerevisiae GN=TAH1 PE=1 SV=1	*S. cerevisiae*	12503	6.08	69%	30
	RS5_ECOLI	30S ribosomal protein S5 OS=E. coli (strain K12) GN=rpsE PE=1 SV=2	*E. coli*	17592	10.11	55%	9
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	25%	9
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	12%	7
12	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	72%	136
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	25%	11
13	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	11%	83
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	25%	15
	Snu13p_YEAST	13 kDa ribonucleoprotein-associated protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	13560	7.68	56%	12
	K2C1_HUMAN	Keratin, type II cytoskeletal 1 - H. sapiens (Human)	*H. sapiens*	65978	8.16	12%	7
14	Snu13p_YEAST	13 kDa ribonucleoprotein-associated protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	13560	7.68	74%	91
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	25%	17
	RS15_ECOLI	30S ribosomal protein S15 - E. coli (strain K12)	*E. coli*	10263	10.40	67%	10
	RS10_ECOLI	30S ribosomal protein S10 - E. coli (strain K12)	*E. coli*	11728	9.68	38%	9
15	Pih1p_YEAST	Protein interacting with Hsp90 1 OS=S. cerevisiae GN=PIH1 PE=1 SV=1	*S. cerevisiae*	39494	5.63	62%	60
	LACI_ECOLI	Lactose operon repressor OS=E. coli (strain K12) GN=lacI PE=1 SV=3	*E. coli*	38566	6.39	25%	10
	FTSZ_ECOLI	Cell division protein ftsZ OS=E. coli (strain K12) GN=ftsZ PE=1 SV=1	*E. coli*	40299	4.65	25%	8
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	16%	6
16	Bcd1p_YEAST	Zinc finger protein YHR040W OS=S. cerevisiae GN=YHR040W PE=1 SV=1	*S. cerevisiae*	42307	5.76	50%	48
	RHLE_ECOLI	Putative ATP-dependent RNA helicase rhlE OS=E. coli (strain K12) GN=rhlE PE=1 SV=3	*E. coli*	49959	10.06	30%	12
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	21%	8
	Rrp9p_YEAST	Ribosomal RNA-processing protein 9 OS=S. cerevisiae GN=RRP9 PE=1 SV=1	*S. cerevisiae*	65014	5.55	10%	6
17	Rrp9p_YEAST	Ribosomal RNA-processing protein 9 OS=S. cerevisiae GN=RRP9 PE=1 SV=1	*S. cerevisiae*	65014	5.55	45%	76
	ARNA_ECOLI	Bifunctional polymyxin resistance protein arnA OS=E. coli (strain K12) GN=arnA PE=1 SV=1	*E. coli*	74242	6.39	34%	23
	GLMS_ECOLI	Glucosamine--fructose-6-phosphate aminotransferase	*E. coli*	66853	5.56	16%	11
	Bcd1p_YEAST	Zinc finger protein YHR040W OS=S. cerevisiae GN=YHR040W PE=1 SV=1	*S. cerevisiae*	42307	5.76	20%	8
18	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	63%	80
	Bcd1p_YEAST	Zinc finger protein YHR040W OS=S. cerevisiae GN=YHR040W PE=1 SV=1	*S. cerevisiae*	42307	5.76	21%	9
	ISCS_ECOLI	Cysteine desulfurase OS=E. coli (strain K12) GN=iscS PE=1 SV=1	*E. coli*	45061	5.94	19%	9
	Rrp9p_YEAST	Ribosomal RNA-processing protein 9 OS=S. cerevisiae GN=RRP9 PE=1 SV=1	*S. cerevisiae*	65014	5.55	13%	7
19	Bcd1p_YEAST	Zinc finger protein YHR040W OS=S. cerevisiae GN=YHR040W PE=1 SV=1	*S. cerevisiae*	42307	5.76	52%	66
	RHO_ECOLI	Transcription termination factor rho OS=E. coli (strain K12) GN=rho PE=1 SV=1	*E. coli*	46975	6.75	25%	11
	AMPA_ECOLI	Cytosol aminopeptidase OS=E. coli (strain K12) GN=pepA PE=1 SV=1	*E. coli*	54845	6.82	18%	8
	Rrp9p_YEAST	Ribosomal RNA-processing protein 9 OS=S. cerevisiae GN=RRP9 PE=1 SV=1	*S. cerevisiae*	65014	5.55	11%	7
20	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	26%	16
	Bcd1p_YEAST	Zinc finger protein YHR040W OS=S. cerevisiae GN=YHR040W PE=1 SV=1	*S. cerevisiae*	42307	5.76	13%	14
	K2C1_HUMAN	Keratin, type II cytoskeletal 1 OS=H. sapiens GN=KRT1 PE=1 SV=5	*H. sapiens*	65978	8.16	16%	10
	SKP_ECOLI	Chaperone protein skp OS=E. coli (strain K12) GN=skp PE=1 SV=1	*E. coli*	17677	9.69	22%	4
21	Bcd1p_YEAST	Zinc finger protein YHR040W OS=S. cerevisiae GN=YHR040W PE=1 SV=1	*S. cerevisiae*	42307	5.76	48%	46
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	26%	15
	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin OS=S. cerevisiae GN=NOP1 PE=1 SV=1	*S. cerevisiae*	34444	10.24	40%	13
	RHO_ECOLI	Transcription termination factor rho OS=E. coli (strain K12) GN=rho PE=1 SV=1	*E. coli*	46975	6.75	33%	13

Tableau 4a. Résultats bruts de l'analyse par spectrométrie de masse en conditions dénaturantes des bandes de protéines purifiées par electrophorèse lors de la première série d'essais de co-expression. La protéine majoritaire est surlignée en vert ou gris selon qu'il s'agisse respectivement d'une protéine de levure ou d'*E. coli*.

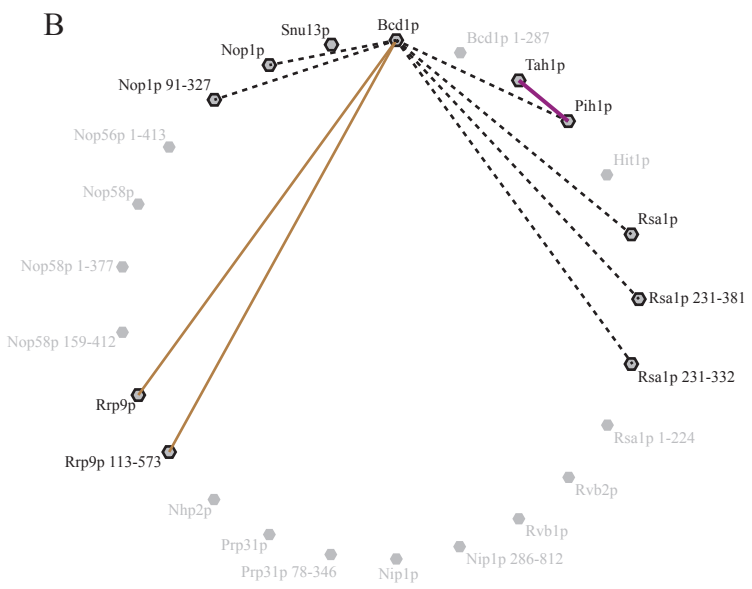
Bande n°	AC	Protéine	Taxonomie	Masse (Da)	pI	Couverture	Nombre de peptides
22	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin OS=S. cerevisiae GN=NOP1 PE=1 SV=1	*S. cerevisiae*	34444	10.24	55%	141
	RS2_ECOLI	30S ribosomal protein S2 OS=E. coli (strain K12) GN=rpsB PE=1 SV=2	*E. coli*	26727	6.61	43%	10
	YRAL_ECOLI	UPF0011 protein yraL OS=E. coli (strain K12) GN=yraL PE=1 SV=1	*E. coli*	31329	5.83	19%	4
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	16%	4
23	Bcd1p_YEAST	Zinc finger protein YHR040W - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	42307	5.76	53%	95
	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	50%	27
	HSLU_ECOLI	ATP-dependent hsl protease ATP-binding subunit hslU - E. coli (strain K12)	*E. coli*	49563	5.24	29%	15
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	25%	14
24	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	56%	131
	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	67%	88
	K2C1_HUMAN	Keratin, type II cytoskeletal 1 - H. sapiens (Human)	*H. sapiens*	65978	8.16	28%	24
	K1C9_HUMAN	Keratin, type I cytoskeletal 9 - H. sapiens (Human)	*H. sapiens*	62092	5.19	24%	14

Tableau 4b. Résultats bruts de l'analyse par spectrométrie de masse en conditions dénaturantes des bandes de protéines purifiées par electrophorèse lors de la première série d'essais de co-expression. La protéine majoritaire est surlignée en vert ou gris selon qu'il s'agisse respectivement d'une protéine de levure ou d'*E. coli*.

A



B



C

SDS-PAGE						Spectrométrie de masse	
Piste n°	Protéine	P.M. (kDa)	Expression	IMAC 50 mM	IMAC 400 mM	Bande n°	Protéine majoritairement détectée
29	His-Bcd1p	44,8	Oui	Oui	Oui		
	Rsa1p (231-381)	17,94	Oui				
30	His-Bcd1p	44,8	Oui	Oui	Oui		
	Rsa1p (231-332)	12,29					
31	His-Pih1p	41,98					
	Bcd1p	42,33	Oui	Oui	Oui		
32	His-Rrp9p	67,52	Oui	Oui	Oui	25	Rrp9p
	Bcd1p	42,33	Oui	Oui	Oui	26	Bcd1p
						27	EFTu
33	His-Rrp9p (113-573)	54,74	Oui	Oui	Oui		
	Bcd1p	42,33	Oui	Oui	Oui		
34	His-Nop1p	36,93				28	Bcd1p
	Bcd1p	42,33	Oui	Oui	Oui	29	Bcd1p
35	His-Nop1p (91-327)	28,95	Oui				
	Bcd1p	42,33		Oui	Oui		
36	His-Rsa1p	43,46					
	Bcd1p	42,33	Oui	Oui	Oui		
37	His-Rsa1p (231-381)	20,4					
	Bcd1p	42,33	Oui	Oui	Oui		
38	His-Rsa1p (231-332)	14,75					
	Bcd1p	42,33	Oui	Oui	Oui		
39	His-Pih1p	41,98	Oui	Oui	Oui		
	Tah1p	12,51	Oui	Oui	Oui		
40	His-Tah1p	14,98	Oui	Oui	Oui		
	Pih1p	39,52	Oui	Oui	Oui		

Figure 51. **Analyse par SDS-PAGE à 15 % des interactions testées lors de la première série d'expériences de co-expression.** (A) Les numéros des pistes sont indiqués au dessus des gels. Les bandes analysées par spectrométrie de masse sont marquées par une "*" et leur numéro est indiqué. Les extraits totaux correspondent aux sonicats avant centrifugation. Les concentrations en sels (50 ou 400 mM) des tampons de purification sont indiquées sous les gels. La souche d'*E. coli* utilisée est la souche BL21 (DE3) pRAR2. (B) Schéma représentant l'ensemble des couples de protéines utilisés. Les interactions détectées par co-expression qui ont été validées par la suite sont représentées en violet et celles nécessitant une vérifiaction en brun. (C) Tableau synthétisant les résultats obtenus par le fractionnement par SDS-PAGE et analyse en spéctrométrie de masse en conditions dénaturantes. Les protéines exprimées grâce aux vecteurs pnEA-tH et pnCS sont respectivement indiquées en orange et en bleu.

Bande n°	AC	Protéine	Taxonomie	Masse (Da)	pI	Couverture	Nombre de peptides
25	Rrp9p_YEAST	Ribosomal RNA-processing protein 9 OS=S. cerevisiae GN=RRP9 PE=1 SV=1	*S. cerevisiae*	65014	5.55	43%	81
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	17%	5
	Bcd1p_YEAST	Zinc finger protein YHR040W OS=S. cerevisiae GN=YHR040W PE=1 SV=1	*S. cerevisiae*	42307	5.76	8%	4
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	3
26	Bcd1p_YEAST	Zinc finger protein YHR040W OS=S. cerevisiae GN=YHR040W PE=1 SV=1	*S. cerevisiae*	42307	5.76	44%	57
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	25%	11
	Rrp9p_YEAST	Ribosomal RNA-processing protein 9 OS=S. cerevisiae GN=RRP9 PE=1 SV=1	*S. cerevisiae*	65014	5.55	7%	5
	Rsa1p_YEAST	Ribosome assembly 1 protein OS=S. cerevisiae GN=RSA1 PE=1 SV=1	*S. cerevisiae*	43971	5.77	11%	3
27	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	65%	68
	Bcd1p_YEAST	Zinc finger protein YHR040W OS=S. cerevisiae GN=YHR040W PE=1 SV=1	*S. cerevisiae*	42307	5.76	39%	25
	ISCS_ECOLI	Cysteine desulfurase OS=E. coli (strain K12) GN=iscS PE=1 SV=1	*E. coli*	45061	5.94	29%	11
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	6
28	Bcd1p_YEAST	Zinc finger protein YHR040W - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	42307	5.76	48%	132
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	38%	24
	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	34444	10.24	26%	6
	HOLB_ECOLI	DNA polymerase III subunit delta' - E. coli (strain K12)	*E. coli*	36913	6.56	31%	6
29	Bcd1p_YEAST	Zinc finger protein YHR040W - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	42307	5.76	70%	181
	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	54%	25
	RHO_ECOLI	Transcription termination factor rho - E. coli (strain K12)	*E. coli*	46975	6.75	23%	9
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	20%	8

Tableau 5. Résultats bruts de l'analyse par spectrométrie de masse en conditions dénaturantes des bandes de protéines purifiées par electrophorèse lors de la première série d'essais de co-expression. La protéine majoritaire est surlignée en vert ou gris selon qu'il s'agisse respectivement d'une protéine de levure ou d'*E. coli*.

protéine Rsa1p, le fragment 231-381 déjà mis en évidence au laboratoire (Boulon *et al.*, 2008) et le fragment 231-332 se sont avérés capables de s'associer à Snu13p dans les expériences de co-expression. Dans le contexte du complexe Snu13p/Rsa1p (231-381), l'interaction était aussi plus forte lorsque Snu13p portait l'étiquette (Pistes 10 et 21), tandis que pour le complexe Snu13p-Rsa1p (231-332) une interaction stable était détectée quelle que soit la protéine portant l'étiquette (Pistes 11 et 22). Nous observions aussi un effet du sel sur ce complexe avec une stabilisation lorsque la concentration saline était de 400 mM. L'analyse par spectrométrie de masse en conditions dénaturantes (bandes 7, 8, 9, 10, 13 et 14) avait confirmé la présence des protéines de levure Snu13p et Rsa1p. On avait également détecté la présence d'un contaminant protéique, le facteur EFTu d'*E. coli* (bande 10) qui migre sur gel de polyacrylamide au même niveau que la protéine Rsa1p.

Un second complexe semblait également se dégager de la première série d'expériences de co-expression. Il s'agissait du complexe Bcd1p/Rrp9p (entière ou le fragment 113-573) (Pistes 24, 25 et 32, 33) avec une plus grande stabilité de l'interaction à une concentration saline élevée (400 mM). L'analyse par spectrométrie de masse en conditions dénaturantes (bandes 16, 17, 18, 19, 20, 25, 26 et 27) avait confirmé la présence des protéines de levure Bcd1p et Rrp9p. On avait également détecté la présence d'un contaminant protéique, le facteur EFTu d'*E. coli* (bandes 18 et 27) qui migre sur gel de polyacrylamide à proximité de la protéine Bcd1p.

Nous observions aussi la présence de deux bandes dans les pistes correspondant au complexe Snu13p/Rrp9p (entière ou le fragment 113-573) (Pistes 2, 3 et 13, 14), surtout à une concentration saline élevée et lorsque la protéine Snu13p portait l'étiquette 6xHis. L'analyse des bandes 1 et 2 par spectrométrie de masse en conditions dénaturantes avait confirmé la nature des protéines de levure Snu13p et Rrp9p.

Il semblait s'établir un complexe entre les protéines Rsa1p et Bcd1p lorsque celles-ci étaient entières et que l'étiquette 6xHis était portée par Bcd1p avec de nouveau, une plus forte stabilité à une concentration de 400 mM en sel (Piste 28). La nature des protéines a également été confirmée par spectrométrie de masse en conditions dénaturantes (bandes 23 et 24)

L'analyse de la piste 27 laissait suggérer l'existence d'un complexe entre les protéines Bcd1p et Nop1p (91-327). Ce complexe se formerait uniquement dans le contexte où c'est la protéine Bcd1p qui porte l'étiquette 6xHis et où le fragment 91-327 de Nop1p est utilisé

comme le confirmait l'analyse en spectrométrie de masse en conditions dénaturantes (bandes 21 et 22).

Enfin la présence de deux bandes majoritaires dans la piste 5 lorsque la concentration en sel était de 400 mM laissait penser que les protéines Snu13p et Nop1p peuvent former un complexe. L'analyse des bandes 3 et 4 par spectrométrie de masse en conditions dénaturantes avait confirmé la présence des protéines Snu13p et Nop1p.

2.1.2 Deuxième série d'expériences

Dans la deuxième série d'expériences (Figures 52 à 54), un complexe entre les protéines Rsa1p et Hit1p était clairement mis en évidence (Pistes 65, 66, 67 et 68, 69, 70). A l'inverse du complexe Rsa1p/Snu13p, le complexe Rsa1p/Hit1p se formait de façon plus stable lorsque la protéine Rsa1p portait l'étiquette 6xHis mais toujours à concentration saline élevée (Piste 68, 69 et 70). De façon intéressante, l'association entre les protéines Rsa1p et Hit1p s'établissait avec la protéine Rsa1p entière (Piste 68) mais aussi avec ses deux fragments C-terminaux (231-381 et 231-332) (Pistes 69 et 70). L'interaction entre les protéines Rsa1p et Hit1p avait déjà été mise en évidence auparavant par un crible double hybride (Ito *et al.*, 2001), mais pour la première fois nos données montraient que cette association est directe et que c'est le domaine 231-332 de la protéine Rsa1p qui est impliqué dans cette interaction.

Il faut également noter la présence dans la piste 41 de deux bandes d'intensité équivalente correspondant respectivement aux poids moléculaires estimés des composants du complexe Rsa1p/Rrp9p. Le taux de rétention sur la résine était relativement faible, mais la nature des protéines Rsa1p et Rrp9p avait toutefois pu être validé grâce à l'analyse par spectrométrie de masse en conditions dénaturantes (bandes 30 et 31).

Enfin, l'examen des pistes 60, 62 et 63 suggérait la formation possible d'un complexe entre la protéine Rsa1p (entière ou fragment 231-381) et Nop58p (1-377). Ce complexe semblait être plus stable en condition saline élevée. La masse moléculaire des deux protéines était très proche dans le cas de Rsa1p entière ce qui rendait difficile la distinction des deux bandes. Par ailleurs de fortes variations de l'intensité des bandes étaient observées d'une piste à l'autre. Les analyses par spectrométrie de masse en conditions dénaturantes des bandes 39, 40, 42, 43, 44 et 45 n'avaient indiqué que la présence majoritaire de la protéine Rsa1p (bandes 39, 40, 42 et44) et du contaminant EFTu (bandes 43 et 45). Aussi nous ne pouvions pas conclure avec certitude sur la formation de ces complexes chez *E. coli*.

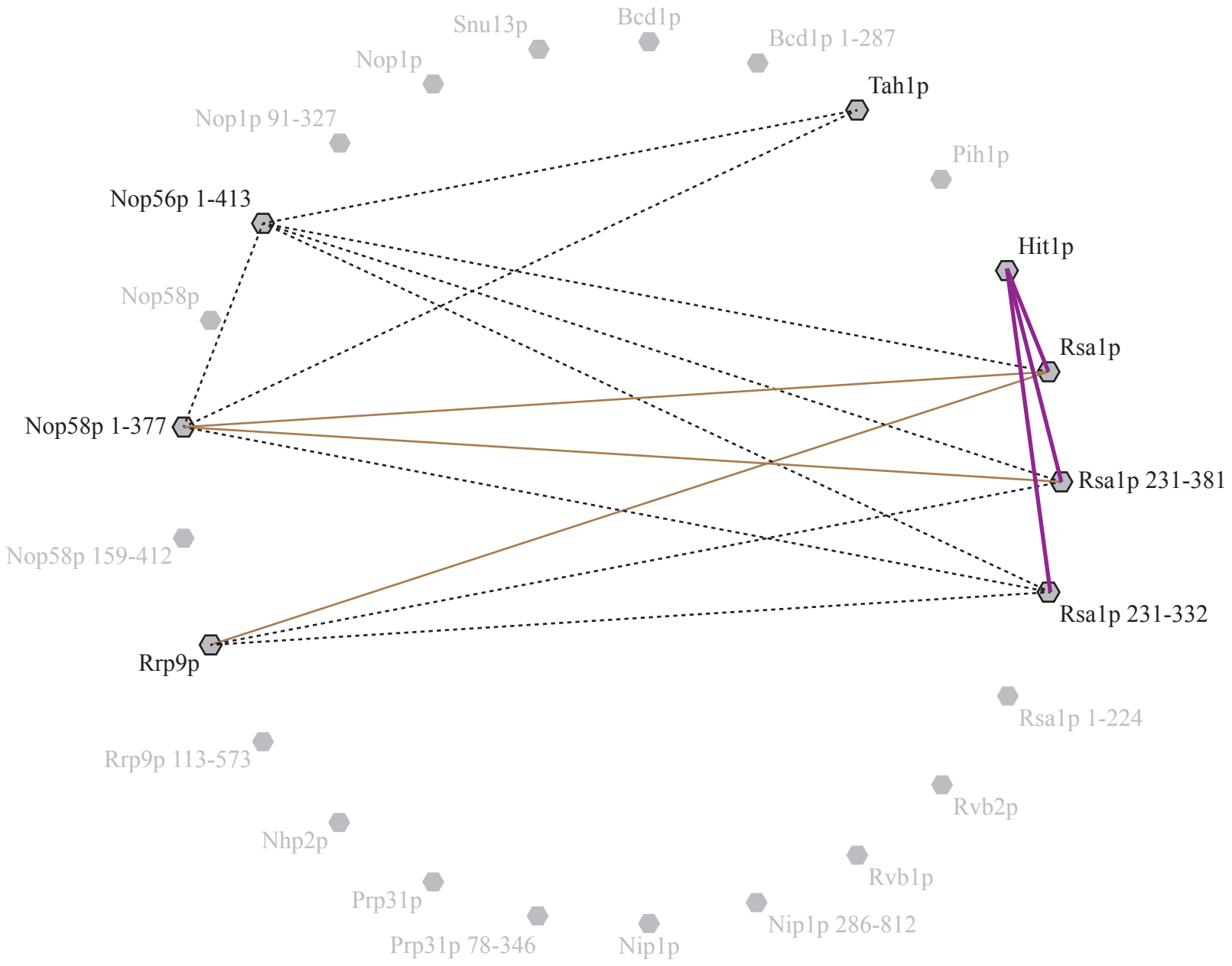
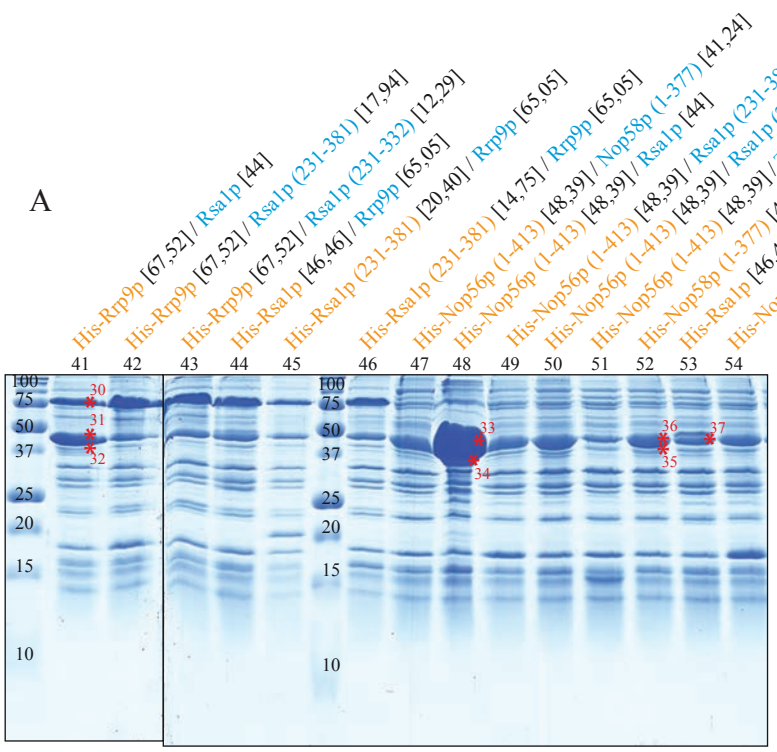
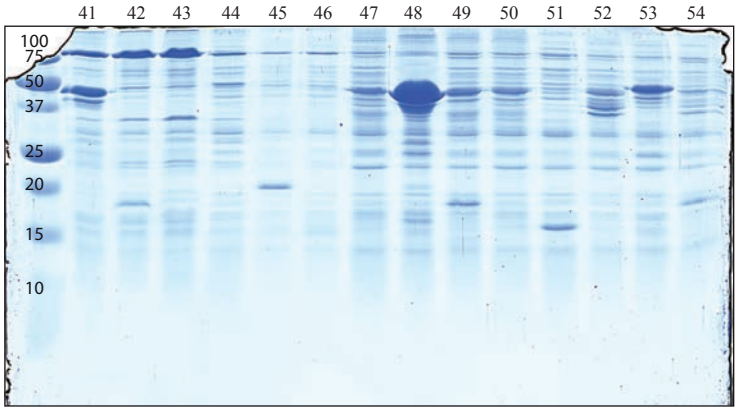


Figure 52. **Bilan des interactions testées dans la deuxième série d'expériences de co-expression.** L'ensemble des protéines des expériences de co-expression sont indiquées. Les traits en pointillés symbolisent les interactions testées, les traits en brun les interactions nécessitant une vérification et les traits épais en violet celles qui ont été validées.

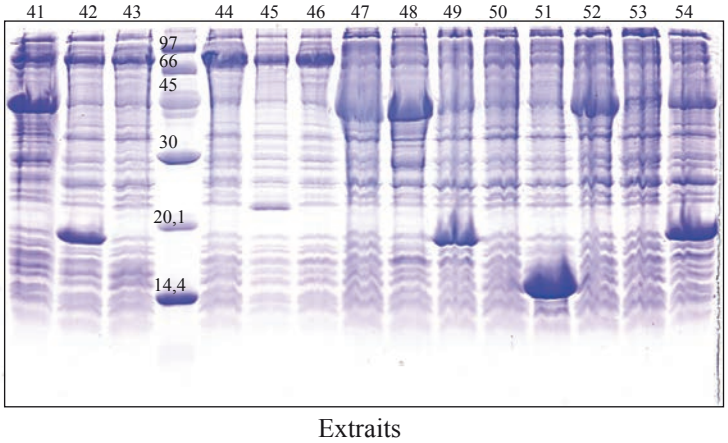
A



400 mM

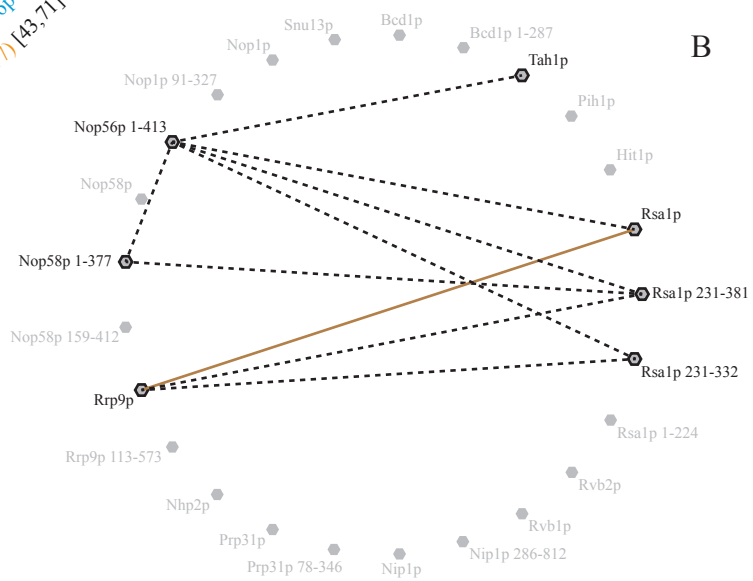


50 mM



Extraits totaux

B



C

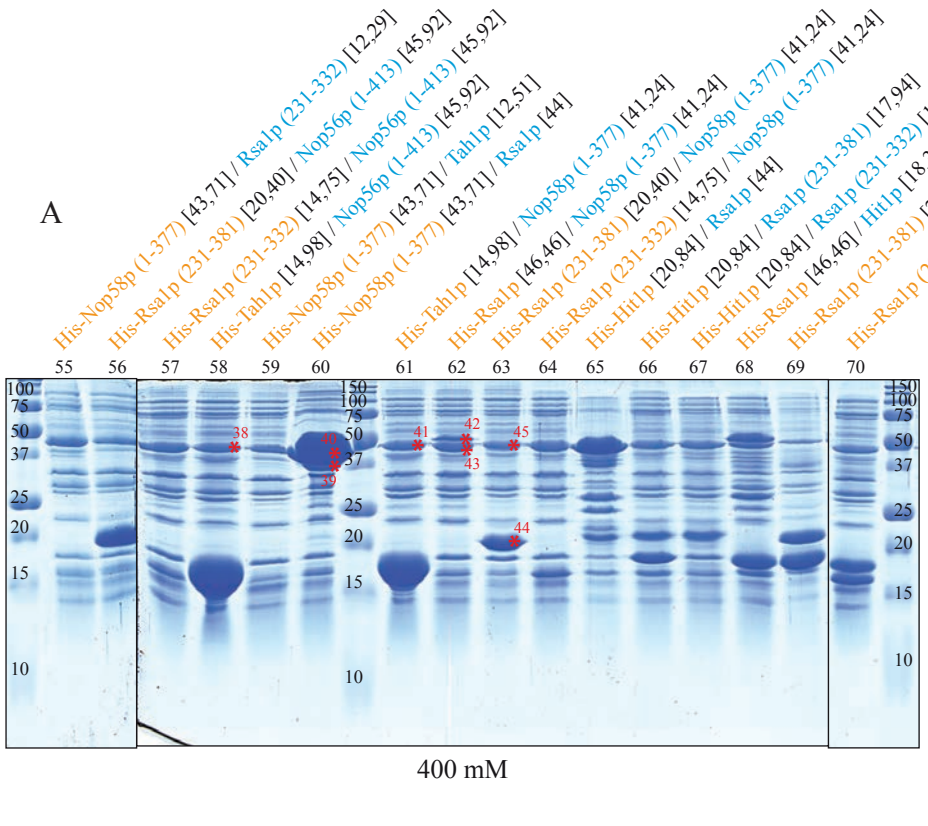
SDS-PAGE						Spectrométrie de masse	
Piste n°	Protéine	P.M. (kDa)	Expression	IMAC 50 mM	IMAC 400 mM	Bande n°	Protéine majoritairement détectée
41	His-Rrp9p	67,52	Oui	Oui	Oui	30	Rrp9p
	Rsa1p	44	Oui	Oui	Oui	31	Rsa1p
						32	EFTu
42	His-Rrp9p	67,52	Oui	Oui	Oui		
	Rsa1p (231-381)	17,94	Oui	Oui	Oui		
43	His-Rrp9p	67,52	Oui	Oui	Oui		
	Rsa1p (231-332)	12,29					
44	His-Rsa1p	46,46			Oui		
	Rrp9p	65,05	Oui	Oui	Oui		
45	His-Rsa1p (231-381)	20,4	Oui	Oui	Oui		
	Rrp9p	65,05	Oui	Oui	Oui		
46	His-Rsa1p (231-332)	14,75					
	Rrp9p	65,05	Oui	Oui	Oui		
47	His-Nop56p (1-413)	48,39					
	Nop58p (1-377)	41,24					
48	His-Nop56p (1-413)	48,39				33	Rsa1p
	Rsa1p	44	Oui	Oui	Oui	34	Rsa1p
49	His-Nop56p (1-413)	48,39	Oui	Oui	Oui		
	Rsa1p (231-381)	17,94	Oui	Oui			
50	His-Nop56p (1-413)	48,39	Oui	Oui	Oui		
	Rsa1p (231-332)	12,29					
51	His-Nop56p (1-413)	48,39	Oui		Oui		
	Tah1p	12,51	Oui	Oui			
52	His-Nop58p (1-377)	43,71	Oui			35	Nop58
	Nop56p (1-413)	45,92				36	EFTu
53	His-Rsa1p	46,46				37	EFTu
	Nop56p (1-413)	45,92					
54	His-Nop58p (1-377)	43,71	Oui				
	Rsa1p (231-381)	17,94	Oui				

Figure 53. **Analyse par SDS-PAGE à 15 % des interactions testées lors de la deuxième série d'expériences de co-expression.** (A) Les numéros des pistes sont indiqués au dessus des gels. Les bandes analysées par spectrométrie de masse sont marquées par une "*" et leur numéro est indiqué. Les extraits totaux correspondent aux sonicats avant centrifugation. Les concentrations en sels (50 ou 400 mM) des tampons de purification sont indiquées sous les gels. La souche d'*E. coli* utilisée est la souche BL21 (DE3) pRAR2. (B) Schéma représentant l'ensemble des couples de protéines utilisés. Les interactions détectées par co-expression qui ont été validées par la suite sont représentées en violet et celles nécessitant une vérifiaction en brun. (C) Tableau synthétisant les résultats obtenus par le fractionnement par SDS-PAGE et analyse en spéctrométrie de masse en conditions dénaturantes. Les protéines exprimées grâce aux vecteurs pnEA-tH et pnCS sont respectivement indiquées en orange et en bleu.

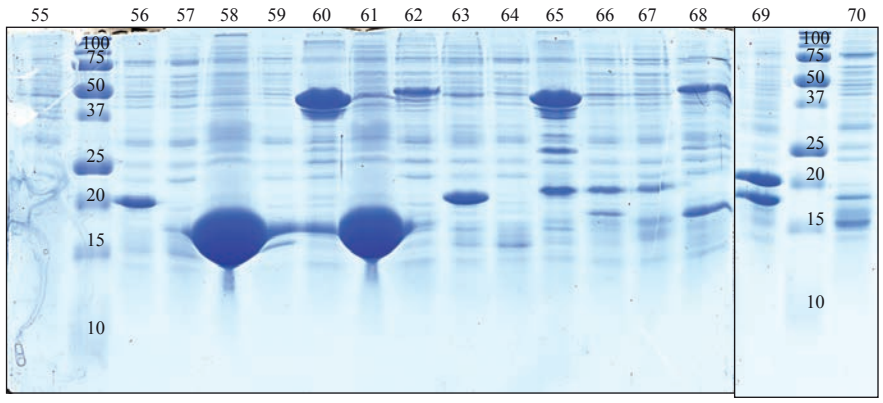
Bande n°	AC	Protéine	Taxonomie	Masse (Da)	pI	Couverture	Nombre de peptides
30	Rrp9p_YEAST	Ribosomal RNA-processing protein 9 OS=S. cerevisiae GN=RRP9 PE=1 SV=1	*S. cerevisiae*	65014	5.55	48%	79
	ARNA_ECOLI	Bifunctional polymyxin resistance protein arnA OS=E. coli (strain K12) GN=arnA PE=1 SV=1	*E. coli*	74242	6.39	21%	14
	Rsa1p_YEAST	Ribosome assembly 1 protein OS=S. cerevisiae GN=RSA1 PE=1 SV=1	*S. cerevisiae*	43971	5.77	26%	7
	SYT_ECOLI	Threonyl-tRNA synthetase OS=E. coli (strain K12) GN=thrS PE=1 SV=1	*E. coli*	73967	5.80	10%	7
31	Rsa1p_YEAST	Ribosome assembly 1 protein OS=S. cerevisiae GN=RSA1 PE=1 SV=1	*S. cerevisiae*	43971	5.77	68%	86
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	60%	29
	ENO_ECOLI	Enolase OS=E. coli (strain K12) GN=eno PE=1 SV=2	*E. coli*	45626	5.32	17%	6
	Bcd1p_YEAST	Zinc finger protein YHR040W OS=S. cerevisiae GN=YHR040W PE=1 SV=1	*S. cerevisiae*	42307	5.76	18%	6
32	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	69%	97
	Rsa1p_YEAST	Ribosome assembly 1 protein OS=S. cerevisiae GN=RSA1 PE=1 SV=1	*S. cerevisiae*	43971	5.77	56%	33
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	4
	CISY_ECOLI	Citrate synthase OS=E. coli (strain K12) GN=gltA PE=1 SV=1	*E. coli*	47984	6.21	6%	3
33	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	67%	259
	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	42%	17
	Nop56p_YEAST	Ribosome biosynthesis protein SIK1 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	56829	8.96	29%	15
	HSLU_ECOLI	ATP-dependent hsl protease ATP-binding subunit hslU - E. coli (strain K12)	*E. coli*	49563	5.24	22%	8
34	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	70%	195
	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	76%	106
	Nop56p_YEAST	Ribosome biosynthesis protein SIK1 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	56829	8.96	56%	62
	CISY_ECOLI	Citrate synthase - E. coli (strain K12)	*E. coli*	47984	6.21	13%	3
35	Nop58p_YEAST	Nucleolar protein NOP58 OS=S. cerevisiae GN=NOP58 PE=1 SV=1	*S. cerevisiae*	56921	8.96	43%	66
	PHEA_ECOLI	P-protein OS=E. coli (strain K12) GN=pheA PE=1 SV=1	*E. coli*	43084	6.21	19%	9
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	16%	8
	LACI_ECOLI	Lactose operon repressor OS=E. coli (strain K12) GN=lacI PE=1 SV=3	*E. coli*	38566	6.39	17%	7
36	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	68%	78
	Nop58p_YEAST	Nucleolar protein NOP58 OS=S. cerevisiae GN=NOP58 PE=1 SV=1	*S. cerevisiae*	56921	8.96	43%	39
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	4
	ACRA_ECOLI	Acriflavine resistance protein A OS=E. coli (strain K12) GN=acrA PE=1 SV=1	*E. coli*	42171	7.68	6%	2
37	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	67%	111
	ISCS_ECOLI	Cysteine desulfurase OS=E. coli (strain K12) GN=iscS PE=1 SV=1	*E. coli*	45061	5.94	21%	9
	Nop56p_YEAST	Ribosome biosynthesis protein SIK1 OS=S. cerevisiae GN=SIK1 PE=1 SV=1	*S. cerevisiae*	56829	8.96	19%	9
	RLMG_ECOLI	Ribosomal RNA large subunit methyltransferase G OS=E. coli (strain K12) GN=rlmG PE=1 SV=2	*E. coli*	42305	6.32	14%	5

Tableau 6. Résultats bruts de l'analyse par spectrométrie de masse en conditions dénaturantes des bandes de protéines purifiées par electrophorèse lors de la deuxième série d'essais de co-expression. La protéine majoritaire est surlignée en vert ou gris selon qu'il s'agisse respectivement d'une protéine de levure ou d'*E. coli*.

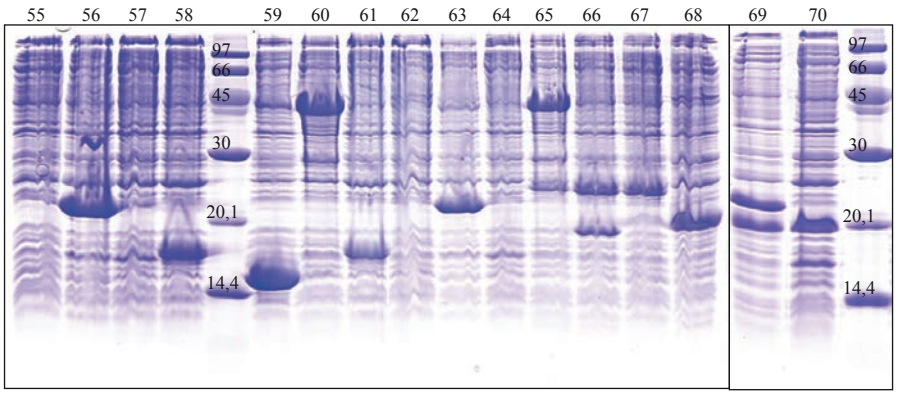
A



400 mM

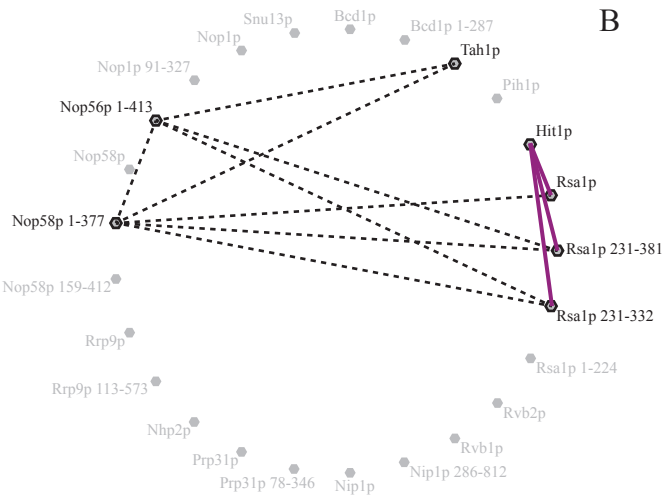


50 mM



Extraits totaux

B



C

SDS-PAGE						Spectrométrie de masse	
Piste n°	Protéine	P.M. (kDa)	Expression	IMAC 50 mM	IMAC 400 mM	Bande n°	Protéine majoritairement détectée
55	His-Nop58p (1-377)	43,71					
	Rsa1p (231-332)	12,29					
56	His-Rsa1p (231-381)	20,4	Oui	Oui	Oui		
	Nop56p (1-413)	45,92					
57	His-Rsa1p (231-332)	14,75					
	Nop56p (1-413)	45,92					
58	His-Tah1p	14,98	Oui	Oui	Oui		
	Nop56p (1-413)	45,92				38	EFTu
59	His-Nop58p (1-377)	43,71					
	Tah1p	12,51	Oui				
60	His-Nop58p (1-377)	43,71				39	Rsa1p
	Rsa1p	44	Oui	Oui	Oui	40	Rsa1p
61	His-Tah1p	14,98	Oui	Oui	Oui		
	Nop58p (1-377)	41,24				41	EFTu
62	His-Rsa1p	46,46	Oui	Oui	Oui	42	Rsa1p
	Nop58p (1-377)	41,24				43	EFTu
63	His-Rsa1p (231-381)	20,4	Oui	Oui	Oui	44	Rsa1p
	Nop58p (1-377)	41,24				45	EFTu
64	His-Rsa1p (231-332)	14,75					
	Nop58p (1-377)	41,24					
65	His-Hit1p	20,84	Oui	Oui	Oui		
	Rsa1p	44	Oui	Oui	Oui		
66	His-Hit1p	20,84	Oui	Oui	Oui		
	Rsa1p (231-381)	17,94	Oui	Oui	Oui		
67	His-Hit1p	20,84	Oui	Oui	Oui		
	Rsa1p (231-332)	12,29					
68	His-Rsa1p	46,46		Oui	Oui		
	Hit1p	18,38	Oui	Oui	Oui		
69	His-Rsa1p (231-381)	20,4	Oui	Oui	Oui		
	Hit1p	18,38	Oui	Oui	Oui		
70	His-Rsa1p (231-332)	14,75	Oui	Oui	Oui		
	Hit1p	18,38	Oui	Oui	Oui		

Figure 54. **Analyse par SDS-PAGE à 15 % des interactions testées lors de la deuxième série d'expériences de co-expression.** (A) Les numéros des pistes sont indiqués au dessus des gels. Les bandes analysées par spectrométrie de masse sont marquées par une "*" et leur numéro est indiqué. Les extraits totaux correspondent aux sonicats avant centrifugation. Les concentrations en sels (50 ou 400 mM) des tampons de purification sont indiquées sous les gels. La souche d'*E. coli* utilisée est la souche BL21 (DE3) pRAR2. (B) Schéma représentant l'ensemble des couples de protéines utilisés. Les interactions détectées par co-expression qui ont été validées par la suite sont représentées en violet et celles nécessitant une vérifiaction en brun. (C) Tableau synthétisant les résultats obtenus par le fractionnement par SDS-PAGE et analyse en spéctrométrie de masse en conditions dénaturantes. Les protéines exprimées grâce aux vecteurs pnEA-tH et pnCS sont respectivement indiquées en orange et en bleu.

Bande n°	AC	Protéine	Taxonomie	Masse (Da)	pI	Couverture	Nombre de peptides
38	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	64%	128
	ISCS_ECOLI	Cysteine desulfurase OS=E. coli (strain K12) GN=iscS PE=1 SV=1	*E. coli*	45061	5.94	19%	7
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	5
	Nop56p_YEAST	Ribosome biosynthesis protein SIK1 OS=S. cerevisiae GN=SIK1 PE=1 SV=1	*S. cerevisiae*	56829	8.96	4%	2
39	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	63%	185
	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	79%	108
	Nop58p_YEAST	Nucleolar protein NOP58 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	56921	8.96	41%	29
	LACI_ECOLI	Lactose operon repressor - E. coli (strain K12)	*E. coli*	38566	6.39	42%	15
40	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	65%	263
	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	35%	15
	MURE_ECOLI	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase - E. coli (strain K12)	*E. coli*	53310	5.42	13%	5
	Nop58p_YEAST	Nucleolar protein NOP58 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	56921	8.96	5%	4
41	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	73%	123
	Nop58p_YEAST	Nucleolar protein NOP58 OS=S. cerevisiae GN=NOP58 PE=1 SV=1	*S. cerevisiae*	56921	8.96	16%	6
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	3
	Tah1p_YEAST	TPR repeat-containing protein associated with Hsp90 OS=S. cerevisiae GN=TAH1 PE=1 SV=1	*S. cerevisiae*	12503	6.08	18%	2
42	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	62%	123
	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	80%	101
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	25%	18
	ENO_ECOLI	Enolase - E. coli (strain K12)	*E. coli*	45626	5.32	34%	11
43	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	80%	255
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	25%	13
	Nop58p_YEAST	Nucleolar protein NOP58 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	56921	8.96	28%	12
	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	33%	12
44	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	28%	170
	RL5_ECOLI	50S ribosomal protein L5 - E. coli (strain K12)	*E. coli*	20289	9.49	62%	43
	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	17%	12
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	20%	6
45	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	81%	191
	Nop58p_YEAST	Nucleolar protein NOP58 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	56921	8.96	38%	26
	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	24%	14
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	25%	12

Tableau 7. Résultats bruts de l'analyse par spectrométrie de masse en conditions dénaturantes des bandes de protéines purifiées par electrophorèse lors de la deuxième série d'essais de co-expression. La protéine majoritaire est surlignée en vert ou gris selon qu'il s'agisse respectivement d'une protéine de levure ou d'*E. coli*.

2.1.3 Troisième série d'expériences

Enfin, dans la troisième série, parmi les pistes 75, 76, 77 et 82, 83, 84 correspondant au complexe Rsa1p-Pih1p, seule la construction His-Rsa1p-Pih1p (Pistes 82, 83 et 84) semblait montrer la formation d'un complexe, que ce soit avec la protéine Rsa1p entière (Piste 82) ou avec les fragments 231-381 (Piste 83) et 231-332 (Piste 231-332). Cette observation fût confortée par les analyses par spectrométrie de masse en conditions dénaturantes qui ont montré la présence des protéines Rsa1p et Pih1p dans les bandes 48 et49.

La piste 78 où avait été déposé le couple des protéines Rrp9p et Pih1p laissait supposer la formation d'un complexe entre ces deux protéines. La présence sur gel de ces deux protéines avait également été validée par les analyses en spectrométrie de masse en conditions dénaturantes (bandes 46 et 47).

Il était également possible de visualiser la présence des protéines Nhp2p et Rsa1p au niveau de la piste 85. Les analyses en spectrométrie de masse en conditions dénaturantes des bandes 50 et 51 avaient confirmé la présence majoritaire de ces deux protéines.

Une association entre les protéines Rsa1p et Nop1p (91-327) semblait aussi exister (Pistes 103, 104, 105 et 108, 109, 110) en particulier quand la protéine Rsa1p portait l'étiquette 6xHis et que la concentration saline était faible (Pistes 108 et 109). Il semblait d'après les données que la formation du complexe soit plus efficace avec la protéine Rsa1p entière ou son fragment 231-381 qu'avec le fragment 231-332 (Pistes 108, 109 et 110). Les analyses par spectrométrie de masse en conditions dénaturantes avaient révélé la présence majoritaire des protéines Rsa1p (bandes 60 et 63) et Nop1p (bandes 61 et 64).

Nous avons aussi observé la présence de deux bandes dans les pistes 101 et 106 correspondant au complexe Nop56p (1-413)/Nop1p (91-327). L'analyse par spectrométrie de masse avait confirmé la présence majoritaire de ces deux protéines dans les bandes 58 et 59.

2.2 Bilan de l'analyse par spectrométrie de masse

La présence d'une bande contaminante qui a pu être identifiée comme étant la protéine EFTu d'*E. coli* ayant un poids moléculaire d'environ 44 kDa dans la quasi-totalité des pistes, notamment en conditions de haute salinité, limitait nos capacités d'interprétation des résultats. De plus, l'inconvénient de cette approche et que la rétention sur les billes de Sépharose de

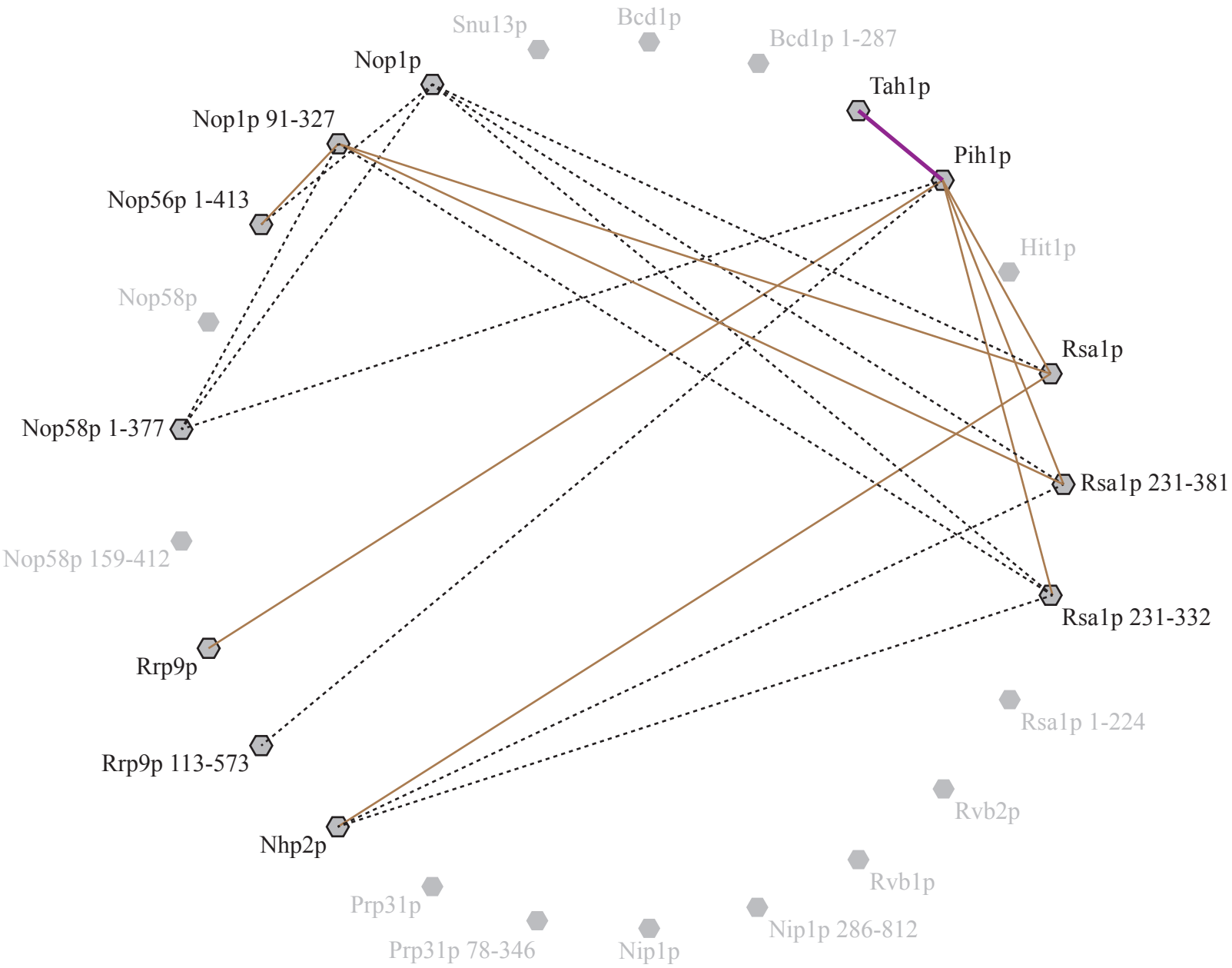
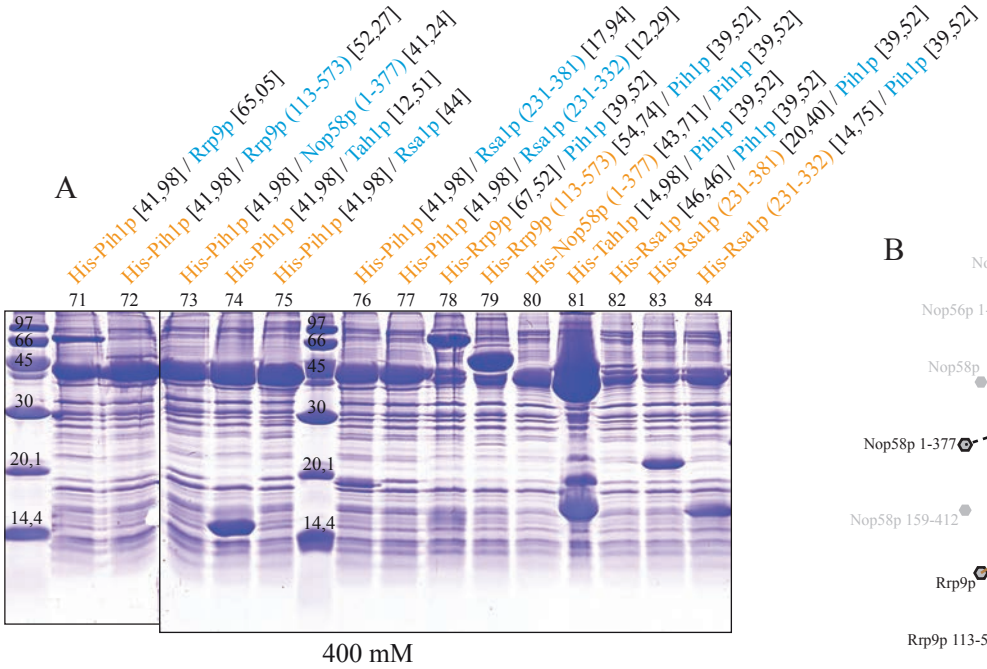
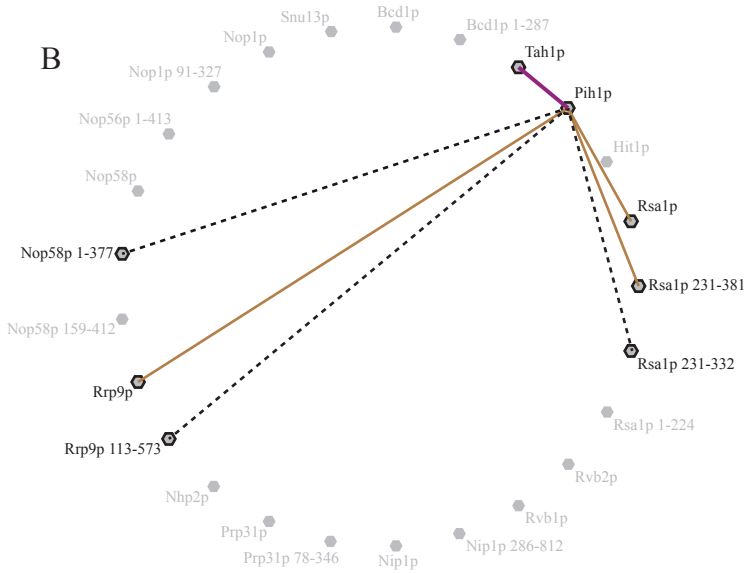


Figure 55. **Bilan des interactions testées dans la troisième série d'expériences de co-expression.** L'ensemble des protéines des expériences de co-expression sont indiquées. Les traits en pointillés symbolisent les interactions testées, les traits en brun les interactions nécessitant une vérification et les traits épais en violet celles qui ont été validées.

A



B



C

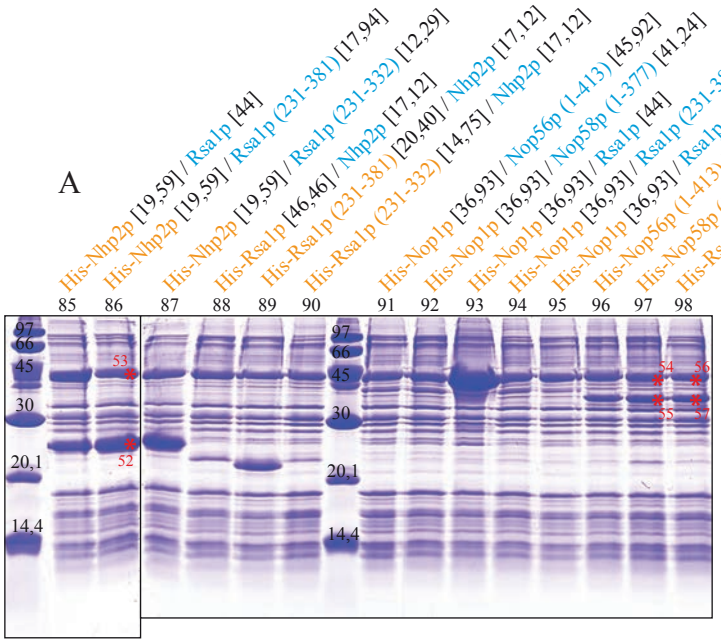
SDS-PAGE						Spectrométrie de masse	
Piste n°	Protéine	P.M. (kDa)	Expression	IMAC 50 mM	IMAC 400 mM	Bande n°	Protéine majoritairement détectée
71	His-Pih1p	41,98	Oui	Oui	Oui		
	Rrp9p	65,05	Oui	Oui	Oui		
72	His-Pih1p	41,98	Oui	Oui	Oui		
	Rrp9p (113-573)	52,27	Oui	Oui	Oui		
73	His-Pih1p	41,98	Oui	Oui	Oui		
	Nop58p (1-377)	41,24					
74	His-Pih1p	41,98	Oui	Oui	Oui		
	Tah1p	12,51	Oui	Oui	Oui		
75	His-Pih1p	41,98	Oui	Oui	Oui		
	Rsa1p	44					
76	His-Pih1p	41,98	Oui	Oui	Oui		
	Rsa1p (231-381)	17,94	Oui	Oui	Oui		
77	His-Pih1p	41,98	Oui	Oui	Oui		
	Rsa1p (231-332)	12,29					
78	His-Rrp9p	67,52	Oui	Oui	Oui	46	Rrp9p
	Pih1p	39,52	Oui	Oui	Oui	47	Pih1p
79	His-Rrp9p (113-573)	54,74	Oui	Oui	Oui		
	Pih1p	39,52	Oui		Oui		
80	His-Nop58p (1-377)	43,71	Oui				
	Pih1p	39,52	Oui	Oui	Oui		
81	His-Tah1p	14,98	Oui	Oui	Oui		
	Pih1p	39,52	Oui	Oui	Oui		
82	His-Rsa1p	46,46	Oui	Oui	Oui		
	Pih1p	39,52	Oui	Oui	Oui		
83	His-Rsa1p (231-381)	20,4	Oui	Oui	Oui	48	Rsa1p
	Pih1p	39,52	Oui	Oui	Oui	49	Pih1p
84	His-Rsa1p (231-332)	14,75	Oui	Oui	Oui		
	Pih1p	39,52	Oui	Oui	Oui		

Figure 56. **Analyse par SDS-PAGE à 15 % des interactions testées lors de la troisième série d'expériences de co-expression.** (A) Les numéros des pistes sont indiqués au dessus des gels. Les bandes analysées par spectrométrie de masse sont marquées par une "*" et leur numéro est indiqué. Les extraits totaux correspondent aux sonicats avant centrifugation. Les concentrations en sels (50 ou 400 mM) des tampons de purification sont indiquées sous les gels. La souche d'*E. coli* utilisée est la souche BL21 (DE3) pRAR2. (B) Schéma représentant l'ensemble des couples de protéines utilisés. Les interactions détectées par co-expression qui ont été validées par la suite sont représentées en violet et celles nécessitant une vérifiaction en brun. (C) Tableau synthétisant les résultats obtenus par le fractionnement par SDS-PAGE et analyse en spéctrométrie de masse en conditions dénaturantes. Les protéines exprimées grâce aux vecteurs pnEA-tH et pnCS sont respectivement indiquées en orange et en bleu.

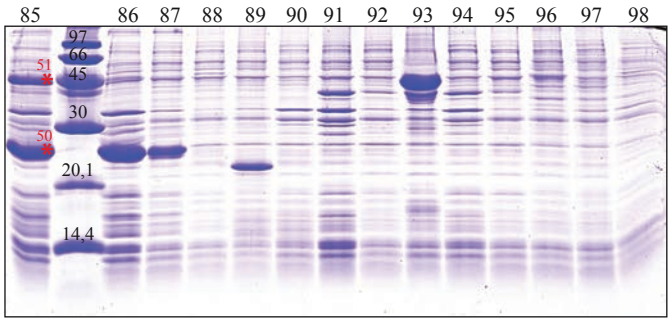
Bande n°	AC	Protéine	Taxonomie	Masse (Da)	pI	Couverture	Nombre de peptides
46	Rrp9p_YEAST	Ribosomal RNA-processing protein 9 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	65014	5.55	60%	277
	ARNA_ECOLI	Bifunctional polymyxin resistance protein arnA - E. coli (strain K12)	*E. coli*	74242	6.39	29%	17
	DHSA_ECOLI	Succinate dehydrogenase flavoprotein subunit - E. coli (strain K12)	*E. coli*	64381	5.85	16%	9
	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	21%	5
47	Pih1p_YEAST	Protein interacting with Hsp90 1 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	39494	5.63	72%	111
	Rrp9p_YEAST	Ribosomal RNA-processing protein 9 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	65014	5.55	28%	22
	ISCS_ECOLI	Cysteine desulfurase - E. coli (strain K12)	*E. coli*	45061	5.94	39%	17
	YCBX_ECOLI	Uncharacterized protein ycbX - E. coli (strain K12)	*E. coli*	40619	7.57	53%	18
48	Rsa1p_YEAST	Ribosome assembly 1 protein - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	43971	5.77	27%	158
	Pih1p_YEAST	Protein interacting with Hsp90 1 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	39494	5.63	53%	15
	RL6_ECOLI	50S ribosomal protein L6 - E. coli (strain K12)	*E. coli*	18892	9.71	50%	13
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	38%	13
49	Pih1p_YEAST	Protein interacting with Hsp90 1 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	39494	5.63	76%	169
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	25%	18
	ISCS_ECOLI	Cysteine desulfurase - E. coli (strain K12)	*E. coli*	45061	5.94	29%	11
	LACI_ECOLI	Lactose operon repressor - E. coli (strain K12)	*E. coli*	38566	6.39	38%	10

Tableau 8. Résultats bruts de l'analyse par spectrométrie de masse en conditions dénaturantes des bandes de protéines purifiées par electrophorèse lors de la troisième série d'essais de co-expression. La protéine majoritaire est surlignée en vert ou gris selon qu'il s'agisse respectivement d'une protéine de levure ou d'*E. coli*.

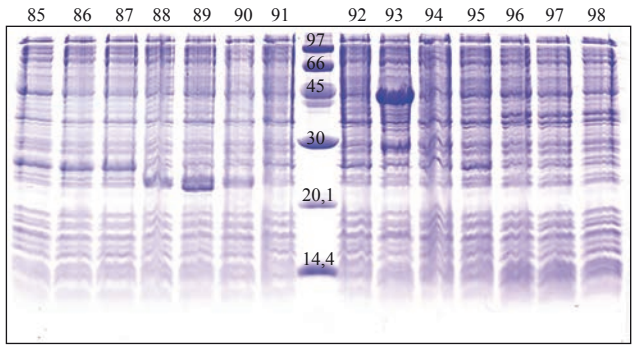
A



400 mM

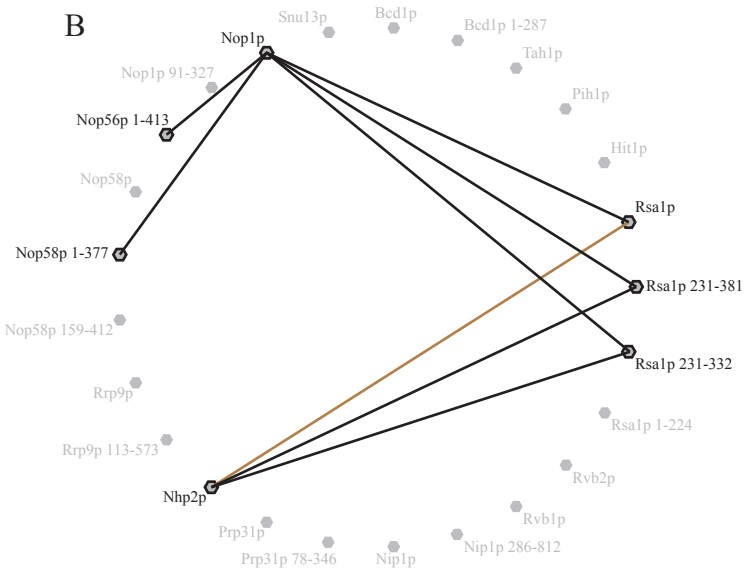


50 mM



Extraits totaux

B



C

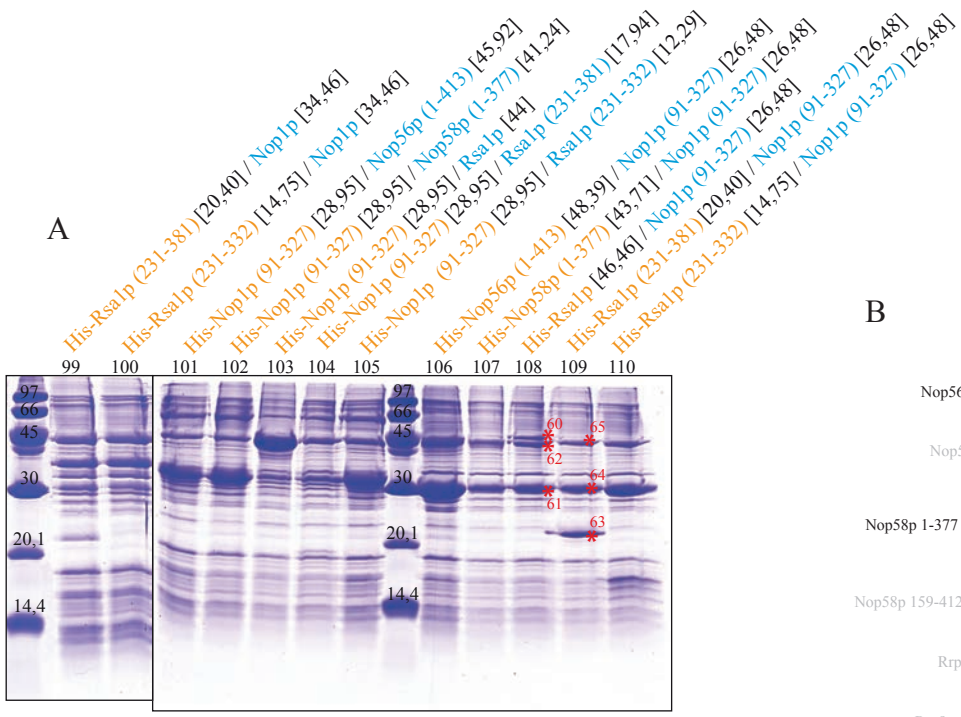
SDS-PAGE						Spectrométrie de masse	
Piste n°	Protéine	P.M. (kDa)	Expression	IMAC 50 mM	IMAC 400 mM	Bande n°	Protéine majoritairement détectée
85	His-Nhp2p	19,59	Oui	Oui	Oui	50	Nhp2p
	Rsa1p	44	Oui	Oui	Oui	51	Rsa1p
86	His-Nhp2p	19,59	Oui	Oui	Oui	52	Nhp2p
	Rsa1p (231-381)	17,94					
						53	EFTu
87	His-Nhp2p	19,59	Oui	Oui	Oui		
	Rsa1p (231-332)	12,29					
88	His-Rsa1p	46,46	Oui				
	Nhp2p	17,12	Oui				
89	His-Rsa1p (231-381)	20,4	Oui				
	Nhp2p	17,12	Oui	Oui	Oui		
90	His-Rsa1p (231-332)	14,75					
	Nhp2p	17,12	Oui				
91	His-Nop1p	36,93					
	Nop56p (1-413)	45,92					
92	His-Nop1p	36,93					
	Nop58p (1-377)	41,24					
93	His-Nop1p	36,93					
	Rsa1p	44	Oui	Oui	Oui		
94	His-Nop1p	36,93					
	Rsa1p (231-381)	17,94					
95	His-Nop1p	36,93					
	Rsa1p (231-332)	12,29					
96	His-Nop56 (1-413)	48,39					
	Nop1p	34,46	Oui	Oui	Oui		
97	HisNop58p (1-377)	43,71				54	EFTu
	Nop1p	34,46	Oui	Oui	Oui	55	Nop1p
98	His-Rsa1p	46,46				56	EFTu
	Nop1p	34,46	Oui	Oui	Oui	57	Nop1p

Figure 57. **Analyse par SDS-PAGE à 15 % des interactions testées lors de la troisième série d'expériences de co-expression.** (A) Les numéros des pistes sont indiqués au dessus des gels. Les bandes analysées par spectrométrie de masse sont marquées par une "*" et leur numéro est indiqué. Les extraits totaux correspondent aux sonicats avant centrifugation. Les concentrations en sels (50 ou 400 mM) des tampons de purification sont indiquées sous les gels. La souche d'*E. coli* utilisée est la souche BL21 (DE3) pRAR2. (B) Schéma représentant l'ensemble des couples de protéines utilisés. Les interactions détectées par co-expression qui ont été validées par la suite sont représentées en violet et celles nécessitant une vérifiaction en brun. (C) Tableau synthétisant les résultats obtenus par le fractionnement par SDS-PAGE et analyse en spéctrométrie de masse en conditions dénaturantes. Les protéines exprimées grâce aux vecteurs pnEA-tH et pnCS sont respectivement indiquées en orange et en bleu.

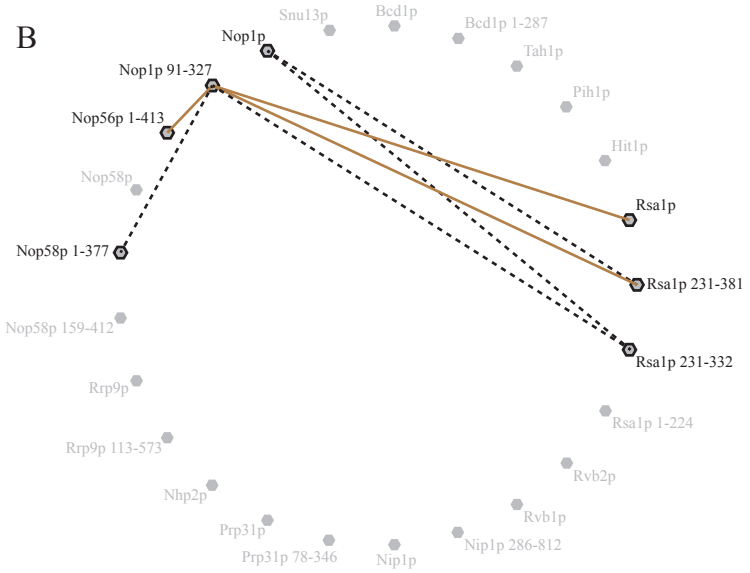
Bande n°	AC	Protéine	Taxonomie	Masse (Da)	pI	Couverture	Nombre de peptides
50	Nhp2p_YEAST	H/ACA ribonucleoprotein complex subunit 2 OS=S. cerevisiae GN=NHP2 PE=1 SV=2	*S. cerevisiae*	17111	9.73	44%	51
	CRP_ECOLI	Catabolite gene activator OS=E. coli (strain K12) GN=crp PE=1 SV=1	*E. coli*	23625	8.37	50%	13
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	4
51	Rsa1p_YEAST	Ribosome assembly 1 protein OS=S. cerevisiae GN=RSA1 PE=1 SV=1	*S. cerevisiae*	43971	5.77	45%	35
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	25%	11
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	5
	Rrp9p_YEAST	Ribosomal RNA-processing protein 9 OS=S. cerevisiae GN=RRP9 PE=1 SV=1	*S. cerevisiae*	65014	5.55	5%	4
52	Nhp2p_YEAST	H/ACA ribonucleoprotein complex subunit 2 OS=S. cerevisiae GN=NHP2 PE=1 SV=2	*S. cerevisiae*	17111	9.73	44%	49
	CRP_ECOLI	Catabolite gene activator OS=E. coli (strain K12) GN=crp PE=1 SV=1	*E. coli*	23625	8.37	52%	12
	RL4_ECOLI	50S ribosomal protein L4 OS=E. coli (strain K12) GN=rplD PE=1 SV=1	*E. coli*	22073	9.72	35%	10
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	6
53	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	58%	69
	Nhp2p_YEAST	H/ACA ribonucleoprotein complex subunit 2 OS=S. cerevisiae GN=NHP2 PE=1 SV=2	*S. cerevisiae*	17111	9.73	35%	6
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	5
54	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	55%	89
	Nop58p_YEAST	Nucleolar protein NOP58 OS=S. cerevisiae GN=NOP58 PE=1 SV=1	*S. cerevisiae*	56921	8.96	15%	7
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	5
	ACRA_ECOLI	Acriflavine resistance protein A OS=E. coli (strain K12) GN=acrA PE=1 SV=1	*E. coli*	42171	7.68	9%	3
55	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin OS=S. cerevisiae GN=NOP1 PE=1 SV=1	*S. cerevisiae*	34444	10.24	46%	59
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	13%	7
	CYSB_ECOLI	HTH-type transcriptional regulator cysB OS=E. coli (strain K12) GN=cysB PE=3 SV=1	*E. coli*	36128	6.87	15%	5
	RL2_ECOLI	50S ribosomal protein L2 OS=E. coli (strain K12) GN=rplB PE=1 SV=2	*E. coli*	29842	10.93	14%	4
56	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	67%	81
	ISCS_ECOLI	Cysteine desulfurase OS=E. coli (strain K12) GN=iscS PE=1 SV=1	*E. coli*	45061	5.94	14%	7
	K1C10_HUMAN	Keratin, type I cytoskeletal 10 OS=H. sapiens GN=KRT10 PE=1 SV=4	*H. sapiens*	59475	5.13	4%	4
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	3
57	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin OS=S. cerevisiae GN=NOP1 PE=1 SV=1	*S. cerevisiae*	34444	10.24	45%	49
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	16%	9
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	4
	CYSB_ECOLI	HTH-type transcriptional regulator cysB OS=E. coli (strain K12) GN=cysB PE=3 SV=1	*E. coli*	36128	6.87	10%	4

Tableau 9. Résultats bruts de l'analyse par spectrométrie de masse en conditions dénaturantes des bandes de protéines purifiées par electrophorèse lors de la troisième série d'essais de co-expression. La protéine majoritaire est surlignée en vert ou gris selon qu'il s'agisse respectivement d'une protéine de levure ou d'*E. coli*.

A



B



C

SDS-PAGE						Spectrométrie de masse	
Piste n°	Protéine	P.M. (kDa)	Expression	IMAC 50 mM	IMAC 400 mM	Bande n°	Protéine majoritairement détectée
99	His-Rsa1p (231-381)	20,4		Oui			
	Nop1p	34,46	Oui		Oui		
100	His-Rsa1p (231-332)	14,75					
	Nop1p	34,46	Oui	Oui	Oui		
101	His-Nop1p (91-327)	28,95	Oui	Oui	Oui		
	Nop56p (1-413)	45,92					
102	His-Nop1p (91-327)	28,95	Oui	Oui	Oui		
	Nop58p (1-377)	41,24					
103	His-Nop1p (91-327)	28,95	Oui	Oui	Oui		
	Rsa1p	44	Oui	Oui	Oui		
104	His-Nop1p (91-327)	28,95	Oui	Oui	Oui		
	Rsa1p (231-381)	17,94	Oui	Oui	Oui		
105	His-Nop1p (91-327)	28,95	Oui	Oui	Oui		
	Rsa1p (231-332)	12,29					
106	His-Nop56p (1-413)	48,39	Oui	Oui	Oui	58	Nop56p
	Nop1p (91-327)	26,48	Oui	Oui	Oui	59	Nop1p
107	his-Nop58p (1-377)	43,71	Oui				
	Nop1p (91-327)	26,48	Oui	Oui	Oui		
108	His-Rsa1p	46,46	Oui	Oui	Oui	60	Rsa1p
	Nop1p (91-327)	26,48	Oui	Oui	Oui	61	Nop1p
						62	EFTu
109	His-Rsa1p (231-381)	20,4	Oui	Oui	Oui	63	Rsa1p
	Nop1p (91-327)	26,48	Oui	Oui	Oui	64	Nop1p
						65	EFTu
110	His-Rsa1p (231-332)	14,75					
	Nop1p (91-327)	26,48	Oui	Oui	Oui		

Figure 58. **Analyse par SDS-PAGE à 15 % des interactions testées lors de la troisième série d'expériences de co-expression.** (A) Les numéros des pistes sont indiqués au dessus des gels. Les bandes analysées par spectrométrie de masse sont marquées par une "*" et leur numéro est indiqué. Les extraits totaux correspondent aux sonicats avant centrifugation. Les concentrations en sels (50 ou 400 mM) des tampons de purification sont indiquées sous les gels. La souche d'*E. coli* utilisée est la souche BL21 (DE3) pRAR2. (B) Schéma représentant l'ensemble des couples de protéines utilisés. Les interactions détectées par co-expression qui ont été validées par la suite sont représentées en violet et celles nécessitant une vérifiaction en brun. (C) Tableau synthétisant les résultats obtenus par le fractionnement par SDS-PAGE et analyse en spéctrométrie de masse en conditions dénaturantes. Les protéines exprimées grâce aux vecteurs pnEA-tH et pnCS sont respectivement indiquées en orange et en bleu.

Bande n°	AC	Protéine	Taxonomie	Masse (Da)	pI	Couverture	Nombre de peptides
58	Nop56p_YEAST	Ribosome biosynthesis protein SIK1 - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	56829	8.96	66%	132
	K1C10_HUMAN	Keratin, type I cytoskeletal 10 - H. sapiens (Human)	*H. sapiens*	59475	5.13	51%	52
	K2C1_HUMAN	Keratin, type II cytoskeletal 1 - H. sapiens (Human)	*H. sapiens*	65978	8.16	40%	44
	EFTU_ECOLI	Elongation factor Tu - E. coli (strain K12)	*E. coli*	43286	5.30	58%	23
59	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin - S. cerevisiae (Baker's yeast)	*S. cerevisiae*	34444	10.24	63%	254
	RS2_ECOLI	30S ribosomal protein S2 - E. coli (strain K12)	*E. coli*	26727	6.61	65%	24
	YRAL_ECOLI	UPF0011 protein yraL - E. coli (strain K12)	*E. coli*	31329	5.83	41%	7
	TRYP_PIG	Trypsin precursor - Sus scrofa (Pig)	*Sus scrofa*	24394	7.00	20%	8
60	Rsa1p_YEAST	Ribosome assembly 1 protein OS=S. cerevisiae GN=RSA1 PE=1 SV=1	*S. cerevisiae*	43971	5.77	27%	15
	K1C10_HUMAN	Keratin, type I cytoskeletal 10 OS=H. sapiens GN=KRT10 PE=1 SV=4	*H. sapiens*	59475	5.13	10%	7
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	12%	5
	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin OS=S. cerevisiae GN=NOP1 PE=1 SV=1	*S. cerevisiae*	34444	10.24	11%	5
61	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin OS=S. cerevisiae GN=NOP1 PE=1 SV=1	*S. cerevisiae*	34444	10.24	32%	41
	RS2_ECOLI	30S ribosomal protein S2 OS=E. coli (strain K12) GN=rpsB PE=1 SV=2	*E. coli*	26727	6.61	38%	21
	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	8%	3
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	3
62	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	47%	52
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	5
	ISCS_ECOLI	Cysteine desulfurase OS=E. coli (strain K12) GN=iscS PE=1 SV=1	*E. coli*	45061	5.94	7%	3
63	Rsa1p_YEAST	Ribosome assembly 1 protein OS=S. cerevisiae GN=RSA1 PE=1 SV=1	*S. cerevisiae*	43971	5.77	22%	18
	RL5_ECOLI	50S ribosomal protein L5 OS=E. coli (strain K12) GN=rplE PE=1 SV=2	*E. coli*	20289	9.49	29%	11
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	5
	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin OS=S. cerevisiae GN=NOP1 PE=1 SV=1	*S. cerevisiae*	34444	10.24	3%	4
64	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin OS=S. cerevisiae GN=NOP1 PE=1 SV=1	*S. cerevisiae*	34444	10.24	34%	32
	RS2_ECOLI	30S ribosomal protein S2 OS=E. coli (strain K12) GN=rpsB PE=1 SV=2	*E. coli*	26727	6.61	38%	20
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	7%	3
	YRAL_ECOLI	UPF0011 protein yraL OS=E. coli (strain K12) GN=yraL PE=1 SV=1	*E. coli*	31329	5.83	12%	3
65	EFTU_ECOLI	Elongation factor Tu OS=E. coli (strain K12) GN=tufA PE=1 SV=2	*E. coli*	43286	5.30	82%	109
	Nop1p_YEAST	rRNA 2'-O-methyltransferase fibrillarin OS=S. cerevisiae GN=NOP1 PE=1 SV=1	*S. cerevisiae*	34444	10.24	29%	11
	TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	*Sus scrofa*	24394	7.00	26%	10
	ISCS_ECOLI	Cysteine desulfurase OS=E. coli (strain K12) GN=iscS PE=1 SV=1	*E. coli*	45061	5.94	17%	6

Tableau 10. Résultats bruts de l'analyse par spectrométrie de masse en conditions dénaturantes des bandes de protéines purifiées par electrophorèse lors de la troisième série d'essais de co-expression. La protéine majoritaire est surlignée en vert ou gris selon qu'il s'agisse respectivement d'une protéine de levure ou d'*E. coli*.

certaines protéines pouvait être aspécifique en particulier dans le cas de protéines peu solubles qui pouvaient s'agréger sur la colonne en même temps que la protéine étiquetée.

La grande majorité des bandes analysées contenaient les protéines de levure attendues ce qui confirmait bien leur co-expression et que cette co-expression permettait parfois une solubilisation des protéines chez *E. coli*. Néanmoins cela ne nous permettait pas d'affirmer avec certitude qu'il y avait bien eu formation de complexe dans la mesure où comme nous l'avons dit plus haut, certaines protéines pouvaient être retenues du fait de leur agrégation sur les billes.

Par spectrométrie de masse nous avons validé la présence des protéines Snu13p et Rsa1p dans les pistes 10 et 22 de la première série de co-expression. Ce qui renforçait l'idée de la formation d'un complexe entre ces 2 protéines, notamment avec les fragments C-terminaux de la protéine Rsa1p 231-381 et 231-332.

Cette donnée a été mise à profit dans un essai de purification d'un complexe His-Snu13p/Rsa1p (231-381) produit à partir de 0,5 litre de culture afin de pousser l'analyse de ce complexe jusqu'à une étape de filtration sur gel suivie d'une analyse par spectrométrie de masse de l'échantillon liquide.

2.3 Purification et analyse du complexe His-Snu13p/Rsa1p (231-381)

Dans un premier temps, nous nous sommes focalisés sur le complexe Snu13p/Rsa1p pour lequel de nombreux arguments prouvaient son existence. On supposait aussi qu'il jouait un rôle majeur dans l'initiation de l'assemblage de toutes les particules contenant la protéine Snu13p, à savoir les snoRNP à boîtes C/D, les scaRNP C/D et la snRNP U4 (Boulon *et al.*, 2008).

2.3.1 Production et purification préparative

Notre objectif étant de réaliser des analyses structurales de ce complexe, nous avons entrepris la mise au point de conditions de production et purification à grande échelle. Pour cela, le complexe His-Snu13p/Rsa1p (231-381) a été produit chez *E. coli* dans la souche *E. coli* BL21 DE3 transformée par le plasmide pRAR2 (voir Matériel & Méthodes). Le milieu de culture utilisé etait le milieu Graffinity (Matériel & Méthodes) que nous avions utilisé pour le criblage des interactions par co-expression. Les résultats de la première série d'expériences de co-expression montraient une stabilité du complexe supérieure dans les conditions de

concentration saline élevée (400 mM), la purification préparative du complexe a donc été réalisée dans ces mêmes conditions. Les étapes et les résultats de la purification du complexe sont représentés dans la figure 59. Malgré une faible surproduction des partenaires protéiques (piste sonicat), nous avons observé deux bandes intenses correspondant bien à des protéines faisant le poids moléculaire attendu après fixation sur la résine TALON^{TM} et électrophorèse des protéines retenues. Nous avons observé une forte rétention aspécifique sur la résine, probablement dû à la faible surexpression de His-Snu13p et Rsa1p (231-381), qui de ce fait entraient faiblement en compétition avec les autres protéines de l'extrait cellulaire lors de la fixation sur les billes de Sépharose. L'élution a été réalisée par palier avec des concentrations croissantes d'imidazole (10 mM, 100 mM et 200 mM), sans réaliser d'hydrolyse par la Thrombine. L'analyse des fractions d'élution montrait que le complexe était élué en présence de 100 mM d'imidazole (10%). Cette étape de purification ne nous a cependant pas permis d'éliminer le contaminant d'environ 30 kDa présent lors de la fixation sur la résine. Ce contaminant a été identifié par spectrométrie de masse comme la « *Peptidyl-prolyl cis-trans isomerase* ». Cette protéine possède une séquence riche en résidu histidine qui explique sa capacité à se lier à la résine lors de la purification. Nous observions également des bandes secondaires minoritaires sous la bande correspondant à His-Snu13p (fractions 31 et 32) mais aussi la présence de deux bandes très proches pouvant correspondre au fragment de Rsa1p (231-381) (fraction 38). Il semblait donc que le fragment 231-381 de la protéine Rsa1p subissait une dégradation protéolytique au cours de la purification. Une étape supplémentaire de purification par exclusion de taille a permis de séparer le complexe His-Snu13p/Rsa1p (231-381) de l'excèdent de His-Snu13p libre. Cet excédent provenait du fait que c'était la protéine Snu13p qui portait l'étiquette poly-histidines. Afin d'identifier précisément les protéines purifiées et les fragments de dégradation, des analyses par spectrométrie de masse en conditions dénaturantes et non dénaturantes ont été réalisées sur les fractions 25 et 28 issues de la filtration sur gel (Figure 59). Afin de vérifier la reproductibilité de nos résultats, une autre fraction, la fraction 18, d'une seconde purification du complexe His-Snu13p/Rsa1p (231-381) a également été analysée en même temps. Le profil de fractionnement sur gel des protéines de la fraction 18 de la seconde expérience était le même que celui de la fraction 25.

2.3.2 Analyse par spectrométrie de masse *(travail réalisé par J.M. Saliou)*

Les fractions 18, 25 et 28 contenant la protéine His-Snu13p et des sous-fragments de la protéine Rsa1p (231-381) ont été concentré sur Vivaspin 500 (3 kDa) pour atteindre une concentration d'environ 40 µM. Les échantillons ont ensuite été dessalé par gel filtration sur

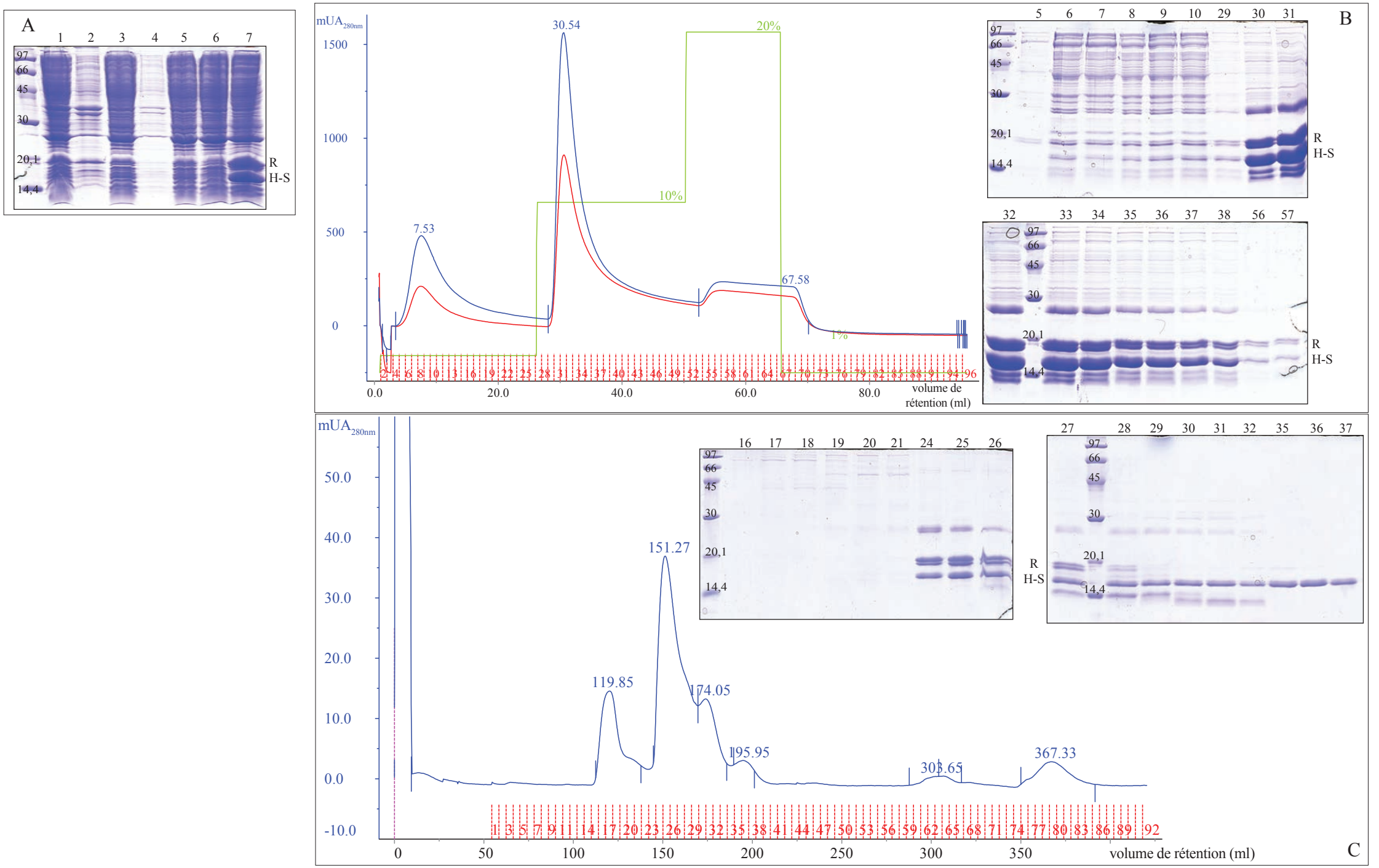


Figure 59. **Mise au point et mise en oeuvre des étapes de la purification du complexe His-Snu13p/Rsa1p (231-381).** (A) Analyse par gel SDS-PAGE des protéines obtenues lors des étapes de purification : piste 1 sonicat, piste 2 culot de sonication, piste 3 surnageant de sonication, piste 4 culot du traitement P.E.I., piste 5 surnageant du traitement P.E.I., piste 6 non retenue sur la résine TALON™, piste 7 résine TALON™. (B) Chromatographie d'affinité IMAC sur résine TALON™ : chromatogramme d'absorption à 280 nm (bleu) et 260 nm (rouge) et analyse des fractions par SDS-PAGE. (C) Chromatographie d'exclusion de taille sur Superdex 75: chromatogramme et analyse des fractions par SDS-PAGE. Les protéines Rsa1p (231-381) et His-Snu13p sont symbolisées par le «R» et «H-S» respectivement.

colonne Zeba (Pierce) contre de l'acétate d'ammonium. Après dessalage, les échantillons ont été dilués dans un solvant Acétonitrile/eau/acide formique (ACN/H2O/HCOOH) (50/50/1) pour réaliser les analyses en conditions dénaturantes ou dans de l'acétate d'ammonium 400 mM, pH 7,8 pour réaliser les analyses en conditions non dénaturantes. Ces analyses ont été réalisées avec le spectromètre de masse LCT (Waters) qui est muni d'un analyseur à temps de vol (TOF) et qui est couplé au robot d'injection Triversa Nanomate (Advion). L'ensemble des résultats a été obtenu en mode d'ionisation positif.

2.3.2.1 Données obtenues en conditions dénaturantes

Les analyses des fractions 25 et 28 ont révélé la présence de 2 espèces identiques dans les 2 fractions (Figure 60):

- une espèce A (15903.3 Da) majoritaire dont la masse moléculaire mesurée correspondait à la masse moléculaire théorique de la protéine His-Snu13p sans la Met $_{ini}$ (15903.3 Da),

- une espèce X (13793 Da) que nous n'avons pas pu identifier car sa masse moléculaire mesurée ne correspondait pas à la masse moléculaire de la protéine Rsa1p (231-381) sans la Met $_{ini}$ (17937.3 Da)

L'analyse par déconvolution des deux spectres (Figure 61) a révélé la présence d'espèces minoritaires différentes dans les 2 fractions : comme attendu, de plus hautes masses que His-Snu13p dans le cas de la fraction 25 (espèces B' et B'') et de plus faibles masses que His-Snu13p dans le cas de la fraction 28 (espèces B''' et B'''').

Les masses mesurées pour les espèces suggéraient quelles correspondaient toutes à des fragments de la protéine Rsa1p (231-381) (les séquences perdues sont en rouge) :

- la forme B' (17291 Da) pouvait correspondre au fragment 231-381 sans la méthionine N-terminale et ayant perdu ses cinq résidus C-terminaux dont la masse théorique est de 17290.5 Da :

MGTSIALITDEDVKKWREERKKMWLLKISNNKQKHMQEMGIKEDELKSQPSIFKESR
KEKQFIQSIQNQVQRGNPKIDLNLKLIQREFANENSQLLDFIRELGDVGLLEYELSQQE
KDVLFGSSEDNNKNHYKPNYKNRKPNLSRANFTRNK.

- la forme B'' (17176 Da) pouvait correspondre à la séquence suivante dont la masse théorique est de 17176.4 Da :

MGTSIALITDEDVKKWREERKKMWLLKISNNKQKHMQEMGIKEDELKSQPSIFKESR
KEKQFIQSIQNQVQRGNPKIDLNLKLIQREFANENSQLLDFIRELGDVGLLEYELSQQE
KDVLFGSSEDNNKNHYKPNYKNRKPNLSRANFTRNK

A

Séquence de Rsa1p (231-381):
MGTSIALITDEDVKKWREERKKMWLLKISNNKQKHMQEMGIKEDELKSQPSIFKESRKEKQFIQSIQNQVQRGNPKIDLN
LKLIQREFANENSQLLDFIRELGDVGLLEYELSQQEKDVLFGSSEDNNKNHYKPNYKNRKPNLSRANFTRNK
Masse moléculaire théorique: 17937.3 Da sans Met_{ini}

Séquence de His-Snu13p:
MGSSHHHHHHSSGTGSGLVPRGSHMSAPNPKAFPLADAALTQQILDVVQQAANLRQLKKGANEATKTLNRGISEFIIMAA
DCEPIEILLHLPLLCEDKNVPYVFVPSRVALGRACGVSRPVIAASITTNDASAIKTQIYAVKDKIETLLI
Masse moléculaire théorique: 15903.3 Da sans Met_{ini}

B

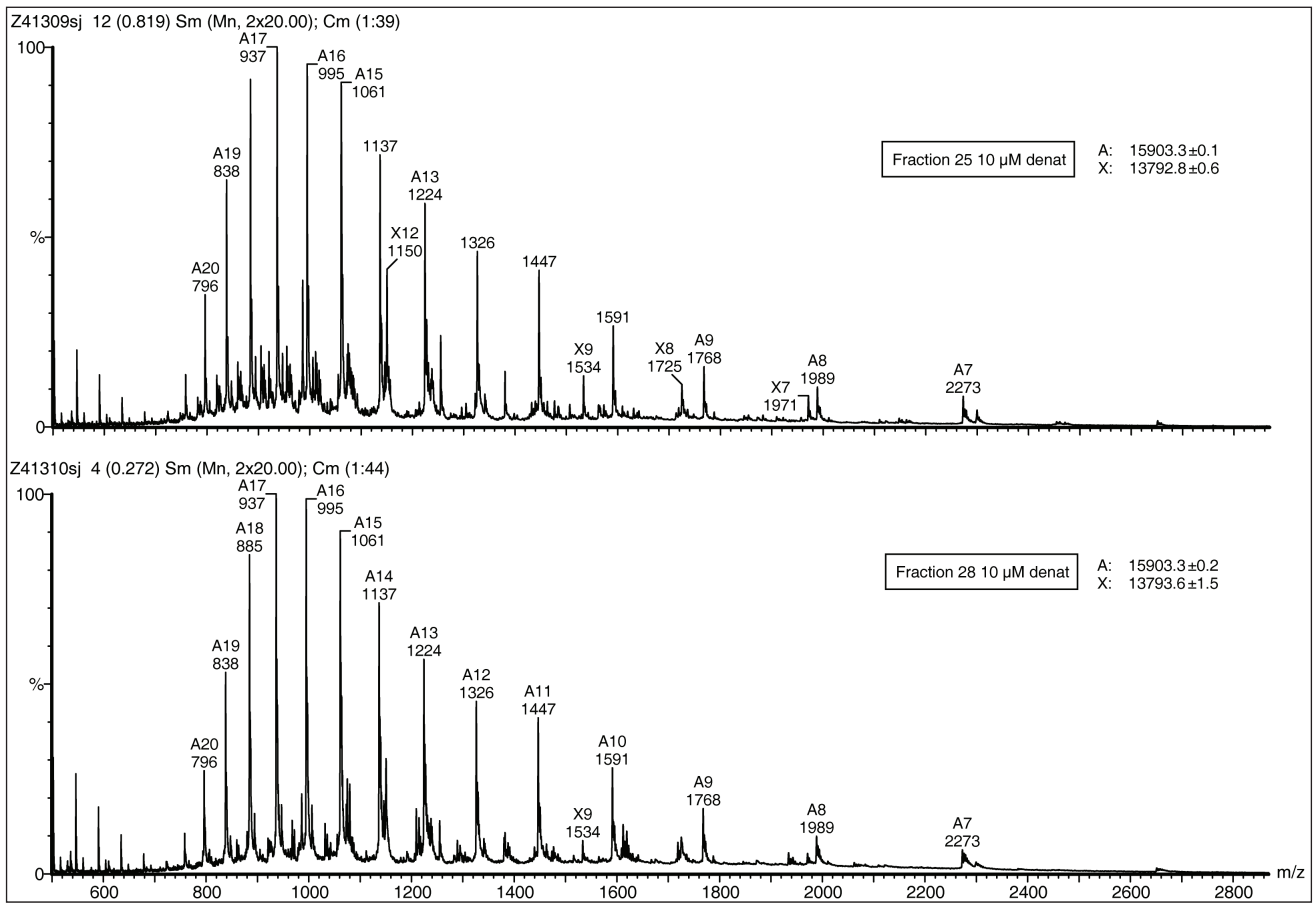
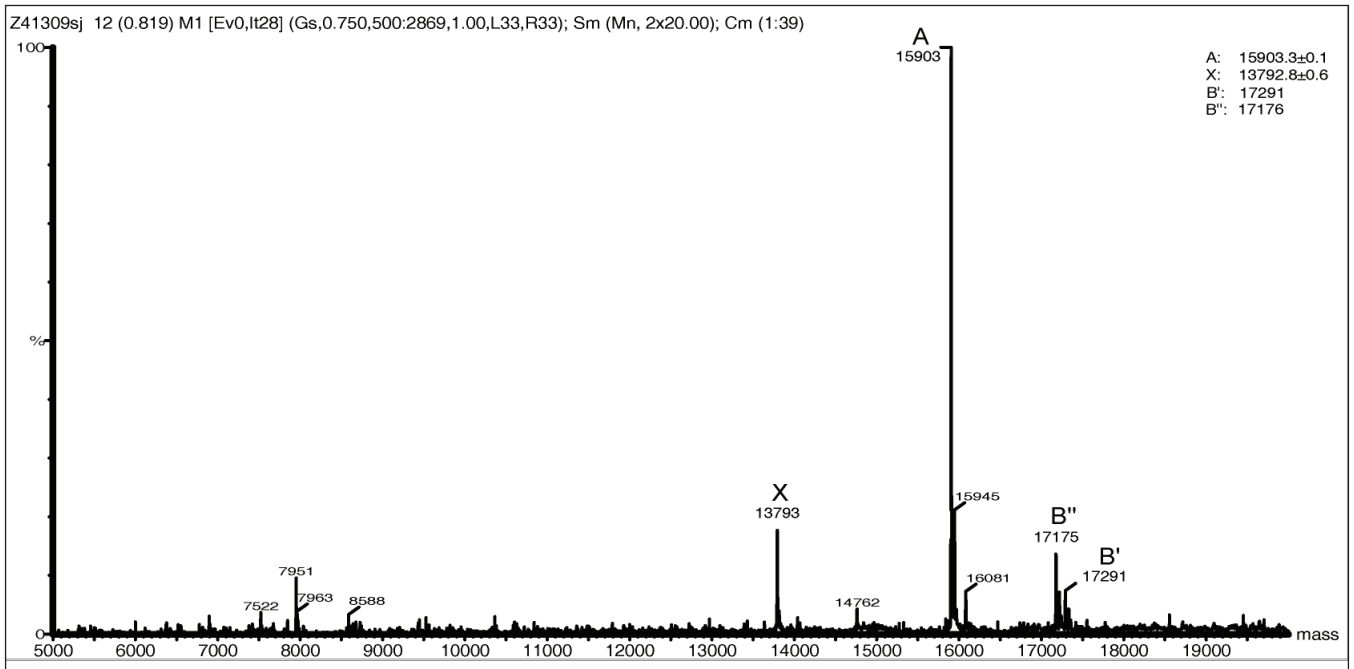


Figure 60. **Analyse par spectrométrie de masse dénaturante des protéines His-Snu13p et Rsa1p (231-381).** (A) Séquences et masses moléculaires des protéines His-Snu13p et Rsa1p (231-3818). (B) Spectres d'analyse des fractions 25 et 28 de la filtration sur gel (Figure 58) en conditions dénaturantes. Injection des échantillons à 10 µM dans un tampon ACN/H2O/HCOOH (50/50/1).

A



B

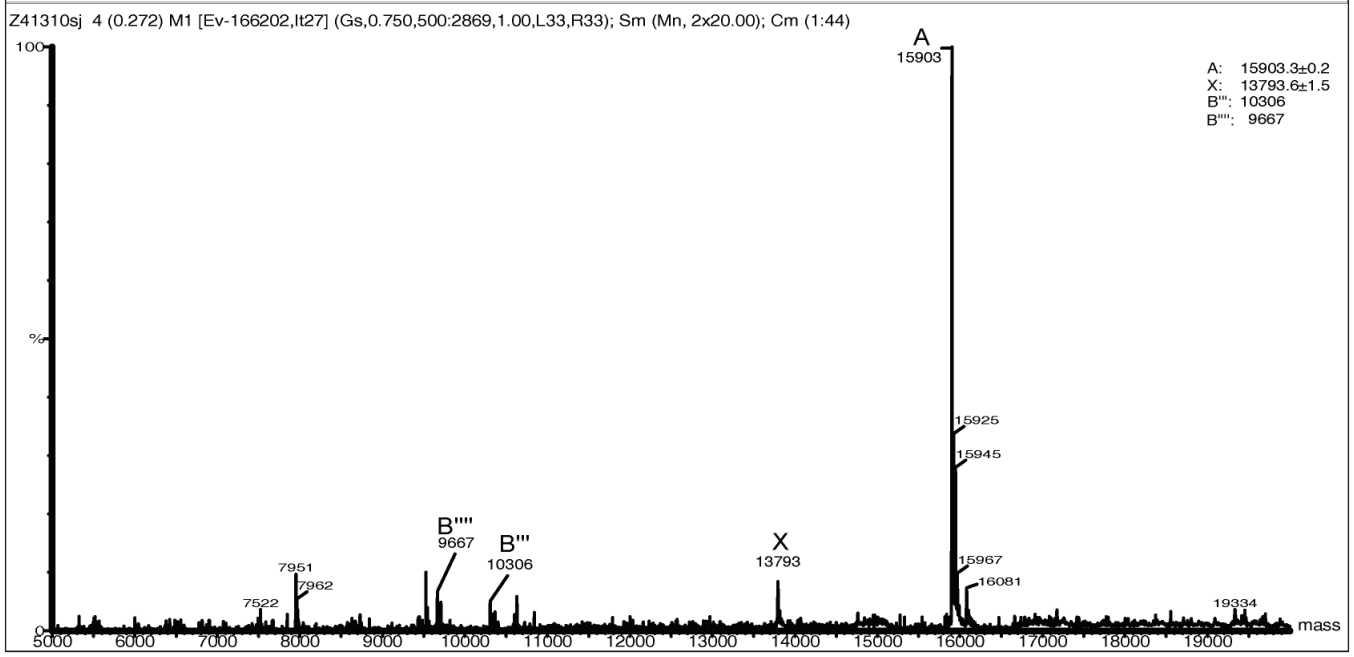


Figure 61. **Spectres des analyses par spectrométrie de masse déconvolués des fractions 25** (A) **et 28** (B) en conditions dénaturantes. Injection des échantillons à 10 µM dans un tamon ACN/H2O/HCOOH (50/50/1).

- la forme B''' (10306 Da) pouvait correspondre à la séquence suivante dont la masse théorique est de 10306.9 Da :

MGTSIALITDEDVKKWREERKKMWLLKISNNKQKHMQEMGIKEDELKSQPSIFKESR
KEKQFIQSIQNQVQRGNPKIDLNLKLIQREFANENSQLLDFIRELGDVGLLEYELSQQE
KDVLFGSSEDNNKNHYKPNYKNRKPNLSRANFTRNK

- la forme B'''' (9667 Da) pourrait correspondre à la séquence suivante dont la masse théorique est de 9667.1 Da :

MGTSIALITDEDVKKWREERKKMWLLKISNNKQKHMQEMGIKEDELKSQPSIFKESR
KEKQFIQSIQNQVQRGNPKIDLNLKLIQREFANENSQLLDFIRELGDVGLLEYELSQQE
KDVLFGSSEDNNKNHYKPNYKNRKPNLSRANFTRNK

Les analyses de la fraction 18 ont révélé la présence du fragment de la protéine Rsa1p (231-381) qui était attendu (espèce B dont la masse est de 17937.3 Da sans Met_{ini}), ainsi que la protéine His-Snu13p (espèce A) et des sous-fragments de Rsa1p (231-381) (espèce B' et B'') (Figure 62) ce qui était en accord avec le profil obtenu sur gel (Figure 59).

2.3.2.2 Données obtenues en conditions non dénaturantes

L'analyse en spectrométrie de masse réalisée sur la fraction 25 mettait en évidence la présence des fragments B et B' de Rsa1p sous forme libre et liée àSnu13p avec en plus 2 formes nouvelles de Snu13p elles aussi sous forme libre et liée à Rsa1p. (Figure 63):

- A' (15967.5 Da) dont la masse moléculaire mesurée correspondait à une masse moléculaire légèrement supérieure à celle de la protéine His-Snu13p sans la Met_{ini} (+ 63,5 Da non attribué)

- A'' (16081.9 Da) dont la masse moléculaire mesurée correspondait à la masse moléculaire de la forme gluconoylée de la protéine His-Snu13p sans la Met_{ini} (15903.3 + 178 = 16081.3 Da) (Geoghegan *et al.*, 1999).

Il était ainsi possible d'observer des complexes formés entre 2 espèces de His-Snu13p (A et A') et 2 espèces de Rsa1p (231-381) (B et B') :

- AB (33842.1 Da) dont la masse moléculaire théorique est de 33840,6 Da,

- A'B (33902.6 Da) dont la masse moléculaire théorique est de 33904,1 Da,

- AB' (33197.0 Da) dont la masse moléculaire théorique est de 33193,8 Da,

- A'B' (33254.6 Da) dont la masse moléculaire théorique est de 33257,3 Da.

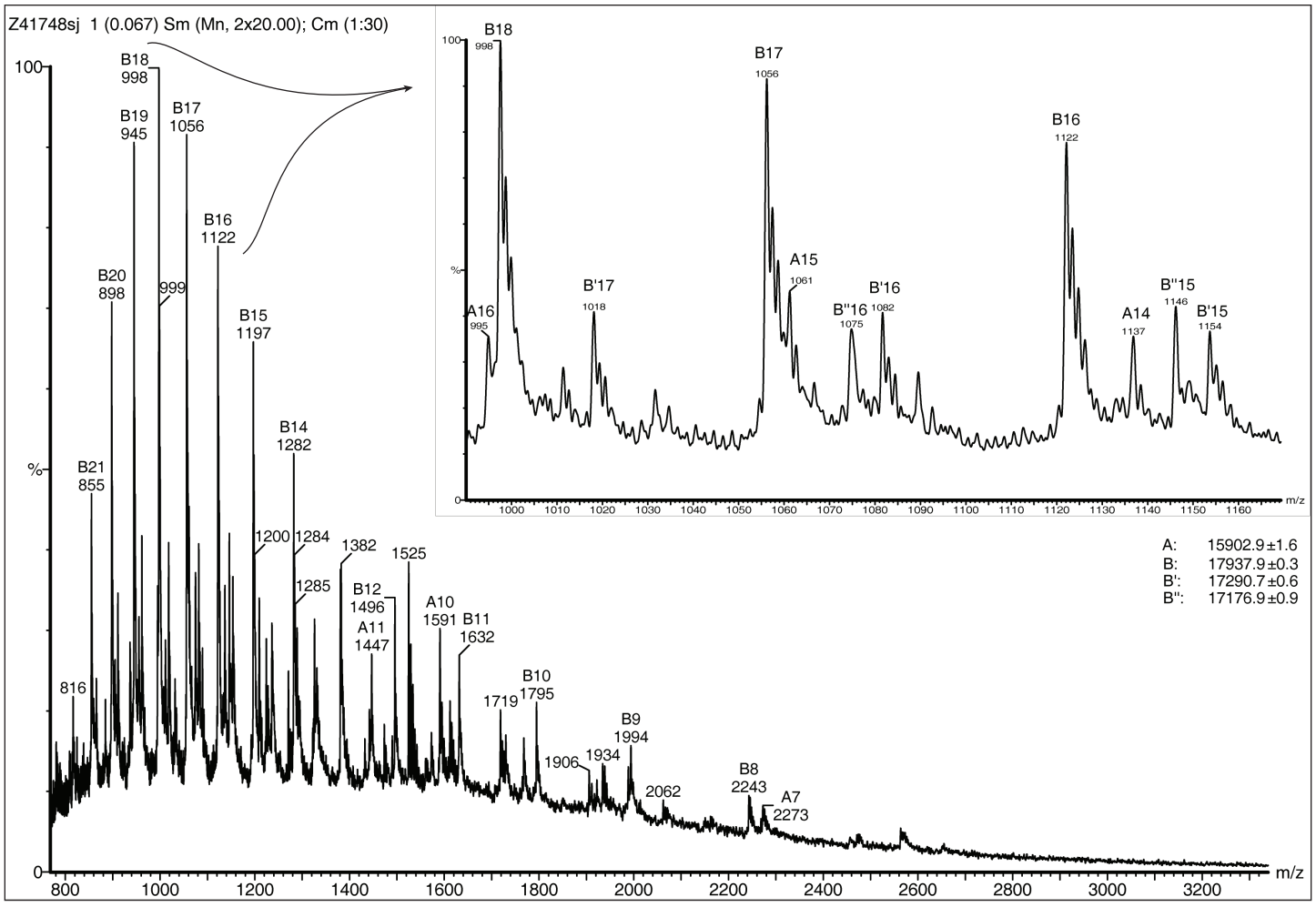


Figure 62. **Spectre des analyses par spectrométrie de masse de la fraction 18 en conditions dénaturantes et agrandissement de trois séries de pics.** Injection de l'échantillon à 10 µM dans un tampon ACN/H2O/HCOOH (50/50/1).

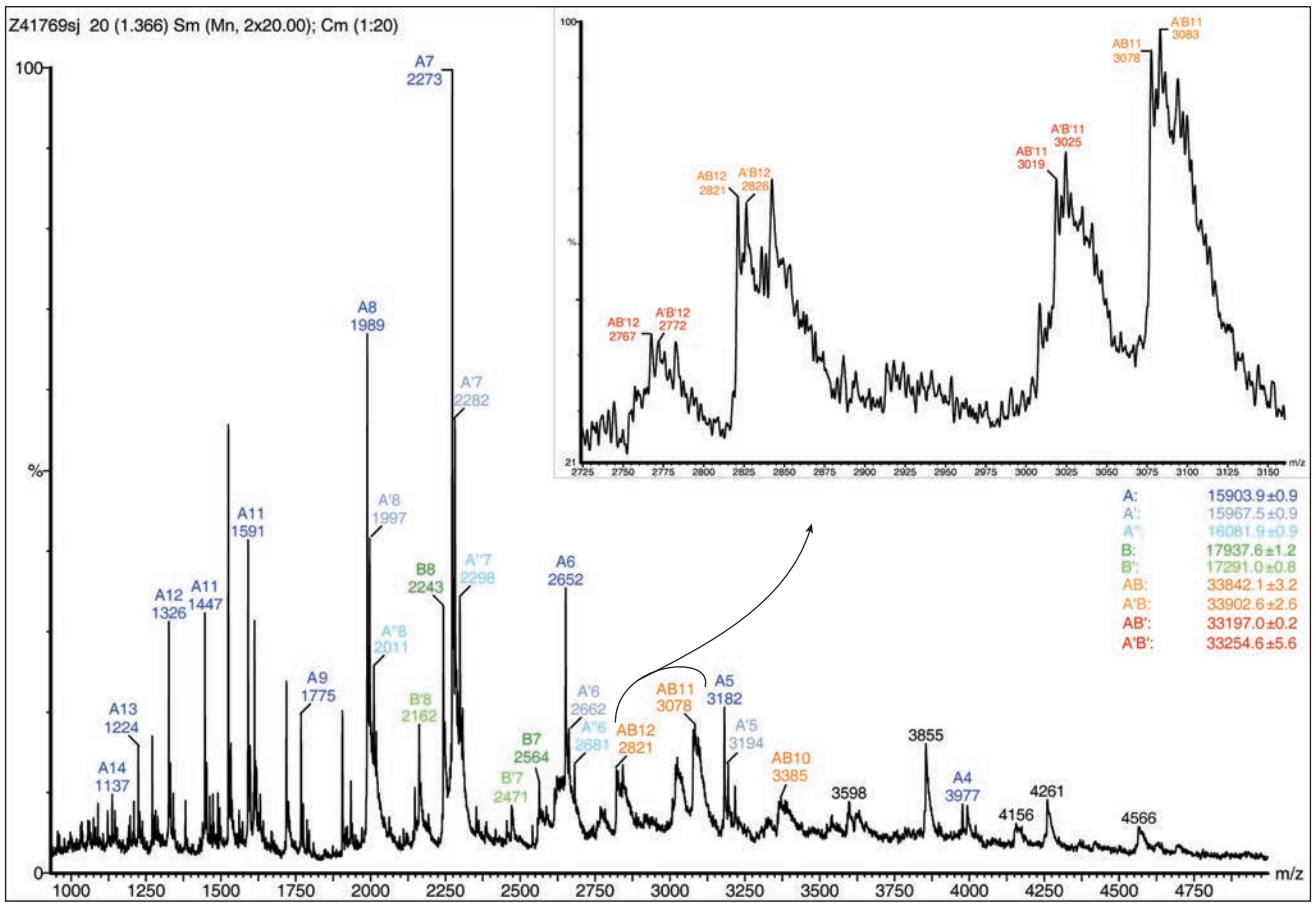


Figure 63. **Spectre de l'analyse par spectrométrie de masse de la fraction 25 en conditions non dénaturantes et agrandissement de la partie du spectre correspondant aux complexes.** Injection de l'échantillon à 10 µM dans un tampon AcNH4 300mM pH 7,5).

2.3.3 Bilan de l'analyse du complexe His-Snu13p/Rsa1p (231-381)

Les analyses des fractions d'élution de la résine TALON™ contenant His-Snu13p et Rsa1p (231-381) ont permis de confirmer la présence des deux protéines, avec en plus pour Rsa1p (231-381) la présence distincte de quatre formes tronqués en C-terminal. Les analyses en conditions non dénaturantes ont révélé que dans les conditions de purifications décrites, deux hétérodimères étaient présents en faible taux formés des protéines His-Snu13p/Rsa1p (231-381) et His-Snu13p/Rsa1p (231-376) où les résidus FTRNK étaient tronqués.

A partir de ces analyses, les séquences codantes des fragments les plus court de Rsa1p pour lesquels une interaction était détectée avec Snu13p ont donc été clonés : Rsa1p (231-375) (la forme B'') et Rsa1p (231-311) (forme B''''). Nous avons postulé que ces fragments issus de l'hydrolyse durant la production et la purification des protéines devraient lorsqu'ils sont produits directement être plus stable que le fragment (231-381) dont ils étaient issus. Par la suite, nous avons remplacé le fragment Rsa1p (231-381) par le fragment Rsa1p (231-375) dans les tests de co-expression. Ces fragments semblaient en effet conserver leurs propriétés d'association à la protéine Snu13p, ce qui suggérait que le site d'interaction de Snu13p se situait dans la région N-terminale du fragment Rsa1p (231-381). L'approche de co-expression couplée à une analyse des complexes par spectrométrie de masse en conditions dénaturantes et non dénaturantes nous avait donc permis d'affiner les connaissances sur les domaines d'interaction de Rsa1p et Snu13p.

2.4 Purification préparative et analyse du complexe His-Rsa1p (231-381)/Hit1p

Un second complexe détecté lors des co-expressions semblait intéressant à étudier : le complexe Rsa1p/Hit1p, car jusque-là, très peu d'informations étaient disponibles sur le rôle possible de hit1p dans l'assemblage des snoRNP.

Nous avons détecté ce complexe dans la deuxième série de co-expression (pistes 65, 66, 67 et 68, 69, 70) (paragraphe 2.1.2). Il semblait plus stable à une concentration saline élévée (400 mM) et lorsque l'étiquette 6xHis était portée par la protéine Rsa1p (entière ou son fragment 231-381). Nous avons donc décidé de produire le complexe His-Rsa1p (231-381)/Hit1p dans la souche BL21 DE3 d'*E. coli* contenant le plasmide pRAR2. D'après nos essais d'optimisation des conditions de purification, il est apparu préférable d'utiliser un tampon HEPES à la place du tampon Tris, ceci en maintenant le pH à 7,5 et en ajoutant du β-

mercapthoéthanol comme réducteur qui est compatible avec l'utilisation d'une résine d'affinité TALON^{TM}. Nous avons réalisé la lyse cellulaire et la fixation sur la résine en présence de 10 mM d'Imidazole, ce qui limitait la fixation aspécifique de protéines sur la résine. Du fait que le complexe semblait plus stable à forte concentration saline, celle-ci a été maintenue à 400 mM tout au long de la purification. Les étapes et les résultats de la purification du complexe sont présentés dans la figure 64. Nous observions une surexpression efficace des protéines étudiées mais aussi la présence d'un contaminant ayant une masse moléculaire d'environ 30 kDa était présent avant la fixation sur résine. Malgré la perte d'une partie du complexe dans la fraction non soluble après la lyse cellulaire (piste culot de sonication), une grande quantité de protéines était retenue sur la résine TALON^{TM}. La présence d'Imidazole à une concentration de 10 mM a empêché la rétention du contaminant protéique présent dans le surnageant de sonication sur la résine. L'élution du complexe a alors été réalisée par paliers successifs d'imidazole allant de 50, 100 et 400 mM (Figure 64). L'analyse des fractions d'élution par gel SDS-PAGE montrait que le complexe commence à être élué à une concentration de 50 mM en imidazole (5%) néanmoins du complexe était encore élué aux concentrations plus élevées. Ce comportement, mais aussi la présence de complexe sur la résine après ces étapes d'élution suggérait une tendance du complexe à établir des contacts non spécifiques avec la matrice Sépharose de la résine ou à l'agrégation. L'étape de chromatographie d'exclusion de taille a permis d'obtenir un complexe His-Rsa1p (231-381)/Hit1p ayant un degré élevé de pureté, qui était compatible avec une étude structurale par cristallographie. En effet, à l'issu de la purification, nous avons obtenus un échantillon de 300 µl à une concentration de 13,36 mg/ml.

2.5 Production de la protéine Snu13p en utilisant le vecteur pnEA-3CH

Les plasmides que nous avions utilisés pour les tests de co-expression portaient une s équence codant pour une étiquette 6xHis succédée d'une séquence générant un site de clivage par la Thrombine afin de cliver la séquence 6xHis. La présence de l'étiquette 6xHis au sein des complexes pouvait être gênante pour la réalisation d'études structurales par radiocristallographie ou par Résonance Magnétique Nucléaire (RMN). Un clivage de l'étiquette 6xHis par la Thrombine n'était pas adapté aux études sur les protéines eucaryotes car :

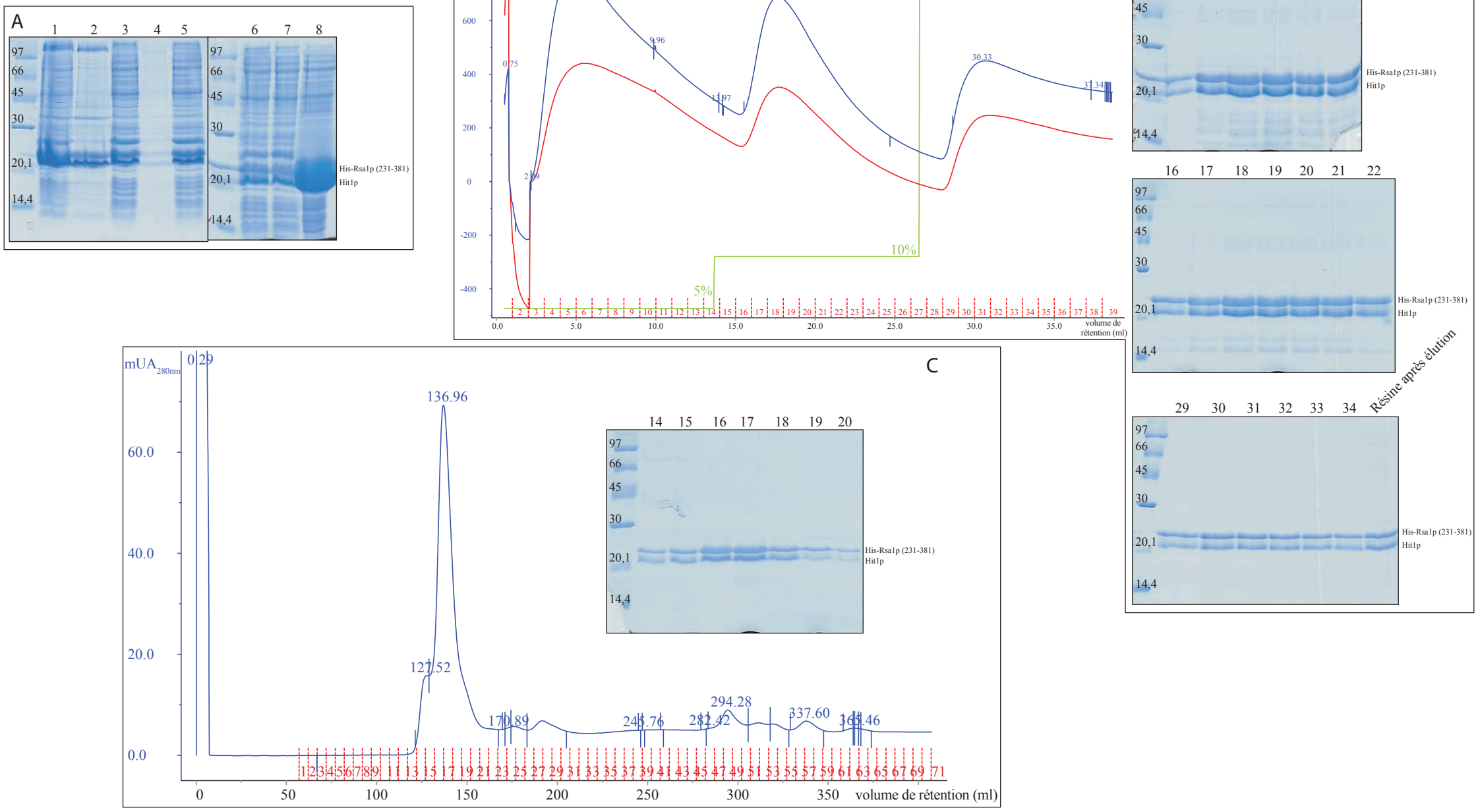


Figure 64. **Mise au point et mise en oeuvre des étapes de la purification du complexe His-Rsa1p (231-381)/Hit1p.** (A) Analyse par gel SDS-PAGE des protéines obtenues lors des étapes de purification : piste 1 sonicat, piste 2 culot de sonication, piste 3 surnageant de sonication, piste 4 culot du traitement P.E.I., pistes 5 et 6 surnageant du traitement P.E.I., piste 7 non retenue sur la résine TALON™, piste 8 résine TALON™. (B) Chromatographie d'affinité IMAC sur résine TALON™ : chromatogramme d'absorption à 280 nm (bleu) et 260 nm (rouge) et analyse des fractions par SDS-PAGE. (D) Chromatographie d'exclusion de taille sur Superdex 75: chromatogramme et analyse des fractions par SDS-PAGE.

- sa température optimale d'activité se situe autour de 45°C, or les protéines eucaryotes que nous étudions sont peu stable à cette température,
- par ailleurs, la spécificité de la Thrombine est faible par rapport à celles d'autres protéases comme la TEV ou la PreScission. Nous avions d'ailleurs constaté que l'utilisation de la Thrombine pour cliver la protéine 6xHis-Bcd1p conduisait à la protéolyse presque totale de cette protéine.

Aussi nous avons réalisé une mutagenèse dirigée du plasmide pnEA-tH afin de remplacer le site de clivage par la Thrombine (NH_2-LVPR↓GS-COOH) par un site de clivage par la PreScission protéase (NH_2-LEVLFQ↓GP-COOH). La PreScission protéase ou 3C a le double avantage en plus d'une activité optimum à 4°C de présenter une spécificité de substrat élevée.

2.5.1 Essais de production et de purification de la protéine Snu13p

Nos études de co-expression montraient que nous formions plus facilement des complexes avec Snu13p lorsque celle-ci portait l'étiquette 6xHis. Cependant, pour étudier le rôle de Snu13p dans les expériences in vitro, nous avions besoin de produire Snu13p sans étiquette. Aussi nous avons réalisé des essais afin de produire, purifier et hydrolyser la protéine Snu13p par la PreScission protéase en utilisant le plasmide pnEA-3CH (Figure 65). Comme attendu après coupure par la PreScission protéase, une bande correspondant à une protéine d'environ 14 kDa était obtenue dans le surnageant.

2.5.2 Analyse par spectrométrie de masse

Afin de s'assurer de l'efficacité et de la précision de la protéolyse par la PreScission, les protéines contenues dans le surnageant ont été analysées par spectrométrie de masse en conditions dénaturantes et non dénaturantes.

L'analyse en condition dénaturante de Snu13p révélait la présence d'une espèce A (13858.2±0.4 Da) dont la masse moléculaire mesurée correspondait à la masse moléculaire de la protéine Snu13p (13860.2 Da) (Figure 66). L'analyse de Snu13p en conditions non dénaturantes révèlait la présence de la même espèce A (13859.3±0.1 Da) correspondant à la protéine Snu13p (13860.2 Da) (Figure 66C) ce qui montrait que Snu13p ne multimérisait pas et que la protéine était correctement clivée et que son degré de pureté est excellent. Nous pouvions ainsi produire de grandes quantité de protéines Snu13p à un haut degré de pureté.

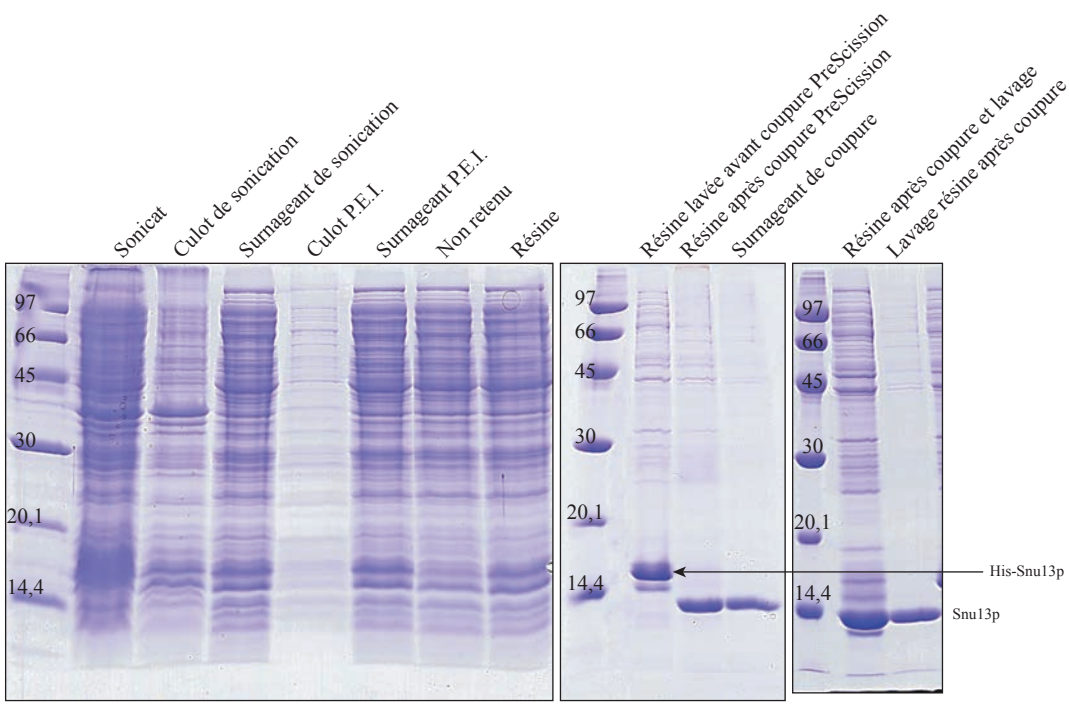
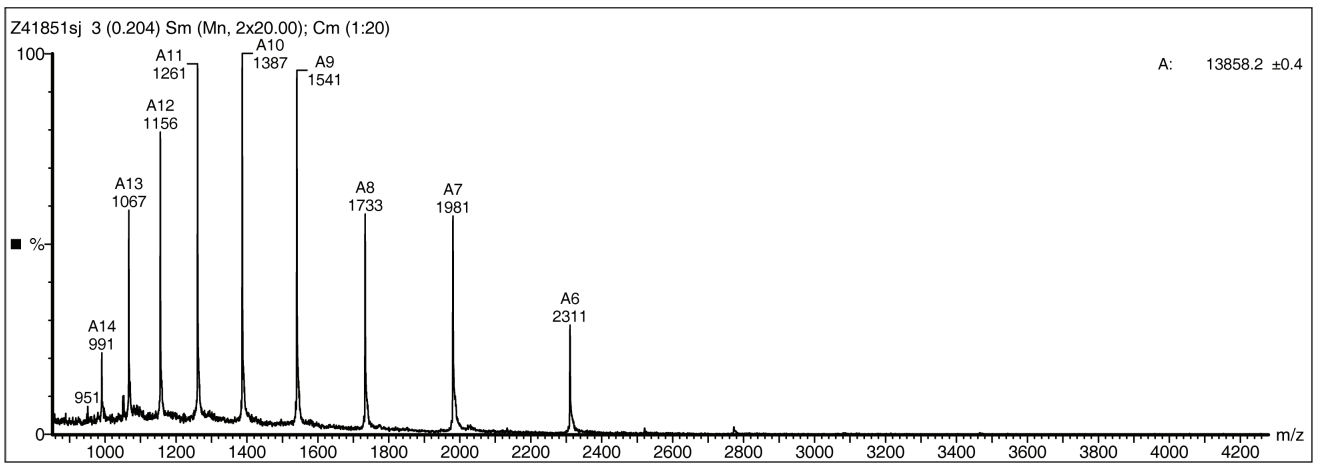


Figure 65. **Suivi des expériences de purification de la protéine Snu13p réalisées sur résine TALON™ avec élution de Snu13p libre par clivage de la protéine His-Snu13p par la PreScission protéase directement sur la résine.** Analyse par SDS-PAGE du matériel obtenu aux différentes étapes de la purification.

A

Séquence de Snu13p:
GPHMSAPNPKAFPLADAALTQQILDVVQQAANLRQLKKGANEATKTLNRGISEFIIMAADCEPIEILLHLPLLCEDKN
VPYVFVPSRVALGRACGVSRPVIAASITTNDASAIKTQIYAVKDKIETLLI
Masse moléculaire : 13860.2 Da

B



C

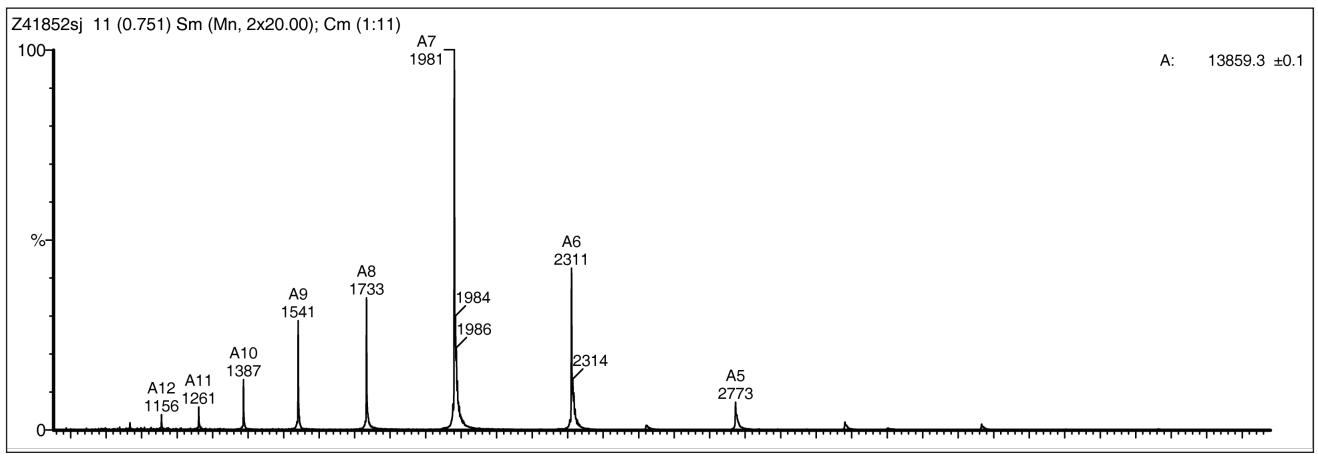


Figure 66. **Analyse par spéctrométrie de masse de la protéine Snu13p.** (A) Séquence et masse moléculaire attendue de l'échantillon Snu13p analysé. (B) Spectre de l'analyse par spectrométrie de masse de Snu13p après hydrolyse par la Préscission protéase en conditions dénaturantes. Injection de Snu13p à 5 µM dans un tampon ACN/H2O/HCOOH (50/50/1). (C) Spectre de l'analyse par spectrométrie de masse de Snu13p après hydrolyse par la Prescission protéase en conditions non dénaturantes. Injection de Snu13p à 5 µM dans un tampon ACN/H2O/HCOOH (50/50/1).

2.6 Le complexe Snu13p/Rsa1p/Hit1p

Nos données de co-expression montraient que Snu13p pouvait former un complexe avec le fragment 231-381 de Rsa1p et que ce même fragment de Rsa1p pouvait former un complexe avec Hit1p. Nous nous sommes donc demandé si le fragment Rsa1p (231-381) pouvait donc servir de linker entre Snu13p et Hit1p où au contraire si Snu13p et Hit1p pouvaient entrer en compétition pour la fixation à Rsa1p. Des données obtenues par l'approche SILAC par E. Bertrand sur les homologues humains de Snu13p, Rsa1p et Hit1p (15.5K, NUFIP et TRIP3) (communication personnelle) suggéraient la formation d'un hétérotrimère 15.5K/NUFIP/TRIP3. Nous avons alors testé la possibilité de produire le complexe Snu13p/Rsa1p/Hit1p de levure. La figure 67 présente la première purification préparative du complexe de levure His-Snu13p/Rsa1p (231-381)/Hit1p par co-expression chez *E. coli*.

Nous avions une très bonne sur-expression des trois protéines de levure ainsi qu'une bonne fixation de ces trois protéines sur la résine TALON™ (piste résine du gel SDS-PAGE, figure 67B). La fixation des trois protéines était un argument fort en faveur de la formation du complexe hétérotrimérique. Bien que l'élution par l'imidazole n'était que partielle (piste résine après élution, figure 67B), nous retrouvions bien les trois partenaires dans l'éluât avec visiblement une stœchiométrie de 1/1/1 suggérant que le complexe était stable dans nos conditions d'élution. Ce qui était confirmé par la présence d'un pic bien défini contenant les trois protéines dans le chromatogramme obtenu lors de la chromatographie d'exclusion de taille (Figure 67C). Les trois protéines partenaires étaient là encore présent à une stœchiométrie de 1/1/1 ce qui confirmait la formation d'un complexe stable entre les protéines His-Snu13p/Rsa1p (231-381)/Hit1p.

Nous avions dans cette expérience élué le complexe par ajout d'imidazole (500 mM). Nous avons ensuite réalisé des tests d'hydrolyse du complexe avec la preScission protéase en incubant l'éluât avec l'enzyme pendant une nuit à 4°C. L'analyse des produits d'hydrolyse (Figure 68A) a révélé que ni la protéine Rsa1p (231-381), ni la protéine Hit1p n'étaient ciblées par la PreScission protéase et que du fait de la diminution de sa masse moléculaire apparente, la protéine His-Snu13p avait perdu l'étiquette 6xHis. L'apparition sur le gel SDS-PAGE d'une bande correspondant à une protéine d'environ 45 kDa dans les pistes où les produits d'hydrolyse ont été fractionné correspond à la PreScission protéase qui était produite au laboratoire sous forme fusionnée à la Glutathione S-transferase (GST).

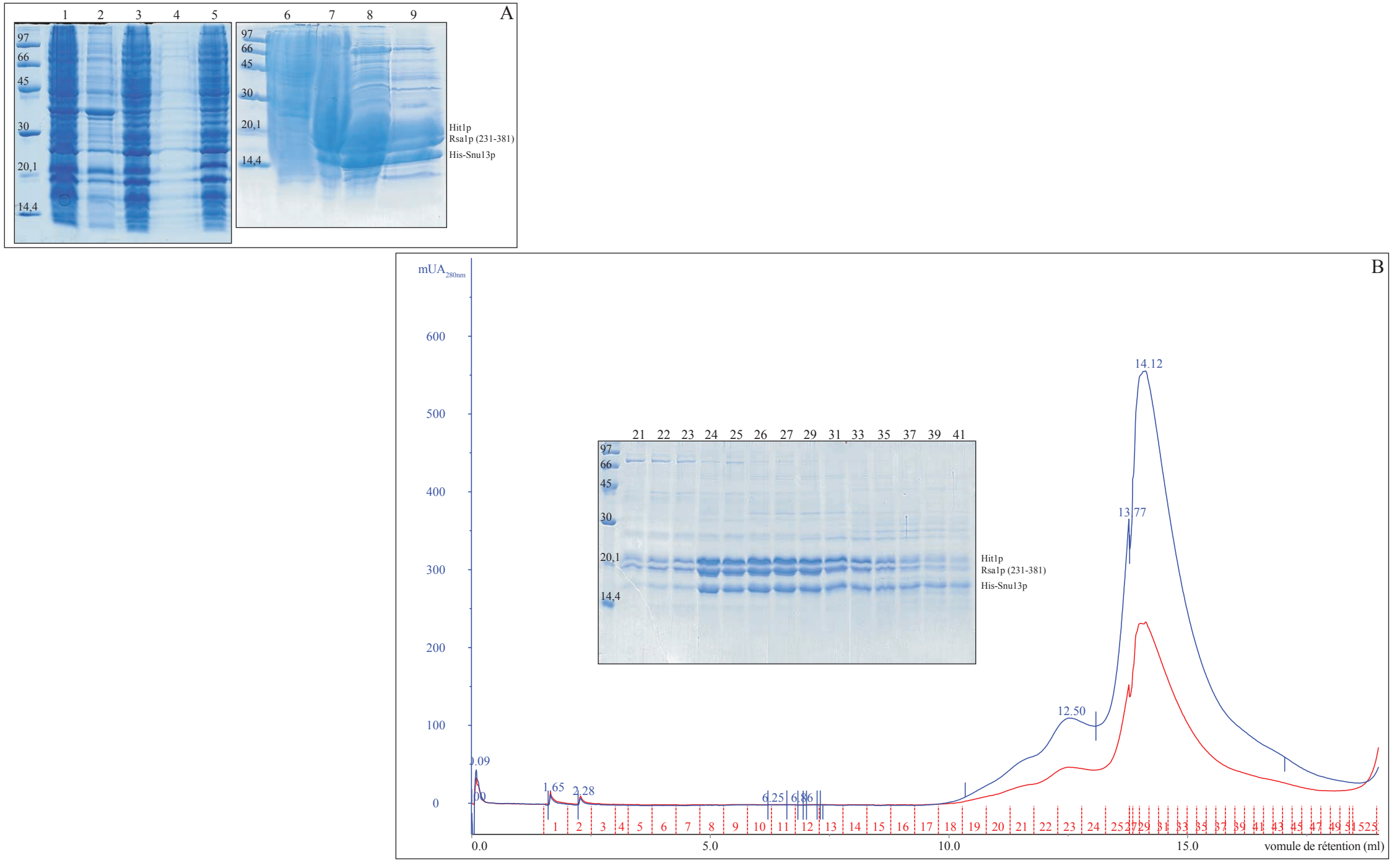


Figure 67. **Mise au point et mise en oeuvre des étapes de la purification du complexe His-Snu13p/Rsa1p (231-381)/Hit1p.** (A) Analyse par gel SDS-PAGE des protéines obtenues lors des étapes de purification : piste 1 sonicat, piste 2 culot de sonication, piste 3 surnageant de sonication, piste 4 culot du traitement P.E.I., piste 5 surnageant du traitement P.E.I., piste 6 non retenue sur la résine TALON™, piste 7 résine TALON™, piste 8 éluât, piste 9 résine après élution. (B) Chromatographie d'exclusion de taille sur Superdex 200: chromatogramme et analyse des fractions par SDS-PAGE.

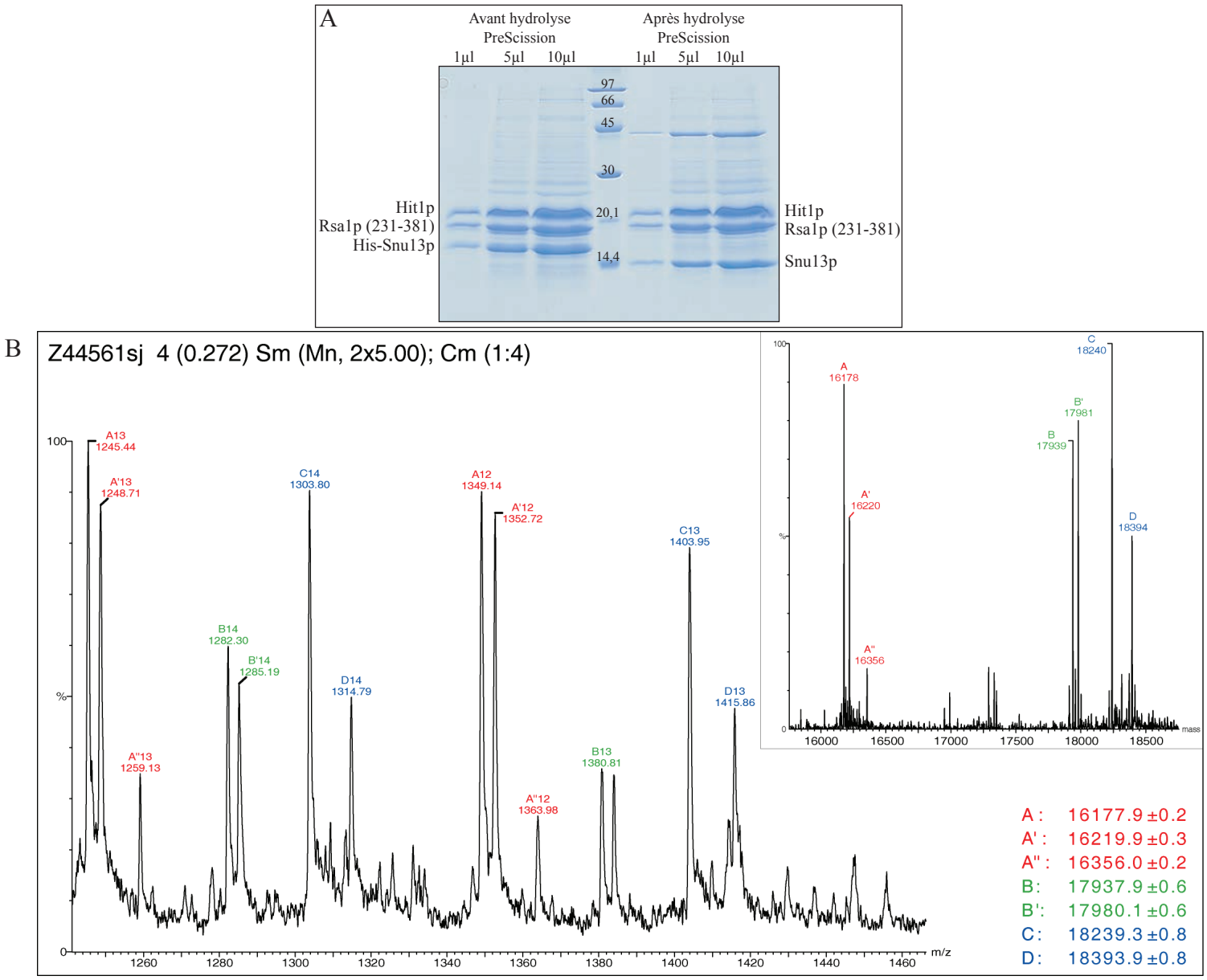


Figure 68. **Analyse du complexe His-Snu13p/Rsa1p (231-381)/Hit1p.** (A) test en solution de la protéolyse du complexe ternaire par la PreScission protéase. (B) Zoom du spectre de l'analyse par spectrométrie de masse en conditions dénaturantes du complexe en tampon ACN/H2O/HCOOH (50/50/1). Le spectre en insert correspond au résultat de la déconvolution. (C) Spectre de l'analyse par spectrométrie de masse en conditions non dénaturantes du complexe en tampon AcNH4 300mM pH 6,5. Les analyses de l'échantillon révèlent la présence des mêmes espèces que celles obtenues en conditions dénaturantes, ainsi que d'une espèce X (52572.5±5.4 Da) dont la masse moléculaire mesurée est très voisine de celle calculée pour le complexe formé par les espèces A (His-Snu13p), B (Rsa1p (231-381) et D (Hit1p) (52509,7 Da).

Une analyse par spectrométrie de masse en conditions dénaturantes et non dénaturantes a confirmé la présence des trois partenaires protéiques ainsi que des espèces dérivantes (Figure 68B) :

- A (16177,9±0.2 Da) dont la masse moléculaire mesurée correspondait à la masse moléculaire attendue pour la protéine His-Snu13p sans la Met_{ini} (16177,6 Da),
- A' (16219,9±0.3 Da) correspondant à l'espèce A avec un adduit de 42 Da,
- A'' (16356,0±0.2 Da) correspondant à la forme gluconoylée de l'espèce A (+178 Da),
- B (17937,9±0.6 Da) dont la masse moléculaire mesurée correspondait à la masse moléculaire attendu pour la protéine Rsa1p (231-381) sans la Met_{ini} (17937,3 Da),
- B' (17980,1±0.6 Da) correspondant à l'espèce D + un adduit de 42 Da,
- C et D (18239,3±0.8 et 18393,9±0.8 Da) dont les masses moléculaires mesurées étaient voisines de celle attendue pour la protéine Hit1p (18376,1 Da), mais dont les écarts de masse n'ont pas été expliqués.

Cette analyse a également montré que l'hydrolyse de la protéine Snu13p (spectres non présentés) avait lieu au site attendu et la formation du complexe dans un ratio stœchiométrique 1/1/1 (Figure 62 C).

Ces résultats ont été intégré dans un article en préparation dont je suis troisième auteur (Manuscrit n°1).

2.7 Dernières séries d'expériences de co-expression

Le premier ensemble d'expériences de co-expression que nous avons décrit apportait de nombreuses informations qui confirmaient les données précédentes obtenus au laboratoire (Boulon *et al.*, 2008) tout en précisant la nature de certaines des interactions qui s'établissant entre les différentes protéines constitutives des particules et les facteurs d'assemblage. Ces premiers résultats nous avaient aussi permis de produire les complexes Snu13p/Rsa1p et Hit1p/Rsa1p et de mettre en évidence pour la première fois la formation *in vitro* du complexe Snu13p/Rsa1p/Hit1p.

Sur la base de ces résultats, dans un second ensemble d'expériences de co-expression, nous avons cherché à inclure des partenaires supplémentaires dans les essais, en particulier, de tester la co-expression de plus de deux protéines. Malheureusement, ces essais ne se sont pas avérés aussi efficace que nous le pensions vraisemblablement parce qu'en présence de plusieurs plasmides, les différentes protéines ne semblaient pas exprimées au même taux. Par

ailleurs, faute de temps, les données obtenues restent préliminaires et méritent d'être revérifiées et approfondies par d'autres types d'expériences. Nous les décrivons tout de même car elles sont une base pour la poursuite des travaux.

Ce second ensemble d'expériences de co-expression à haut débit a également été réalisé à l'IGBMC (Strasbourg) en collaboration avec C. Romier. Nous avions pris en compte les résultats obtenus par les analyses de spectrométrie de masse et remplacé le plasmide pnEA-tH par le pnEA-3CH permettant d'introduire un site de coupure à la PreScission protéase. Nous avions les différentes ORF étudiées dans le vecteur pnYC afin de tester des complexes contenant plus de deux partenaires. Dans ce cas, du fait de contraintes liées aux résistances aux antibiotiques des plasmides pnYC et pRAR2, nous avons utilisé la souche d'*E. coli* BL21 (DE3). Les résultats de ce second ensemble d'expériences de co-expression qui a été réalisé en quatre séries d'expériences sont présentés dans les figures 69 à 78.

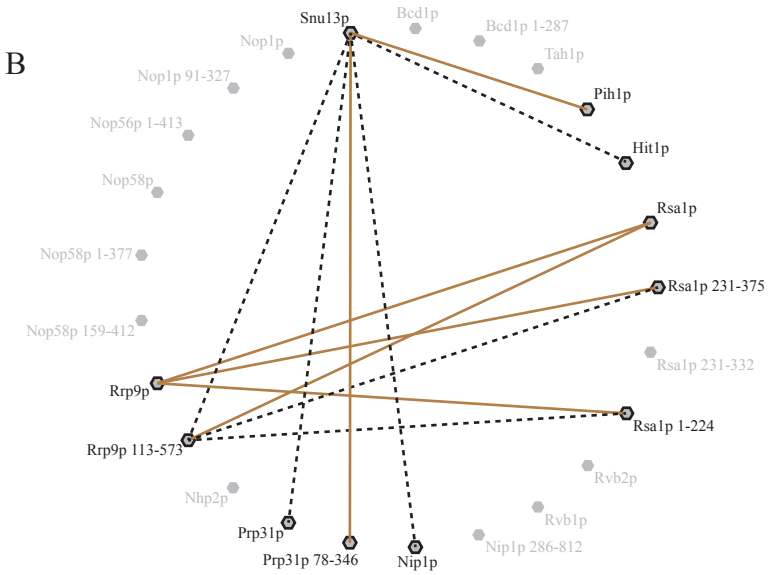
2.7.1 Première série d'essais du second ensemble de co-expression

Dans cette quatrième série d'essais nous avons cherché à savoir si au sein du complexe Snu13p/Rsa1p/Hit1p, un contact direct pouvait s'établir entre les protéines Snu13p et Hit1p (Pistes 111 et 112 de la figure 69). Dans nos conditions d'analyses, les protéines Snu13p et Hit1p ne semblent pas capables de former un hétérodimère. Nous en concluons donc qu'au sein de l'hétérotrimère, la protéine Rsa1p et plus précisément son fragment C-terminal doit faire le lien entre les deux autres partenaires.

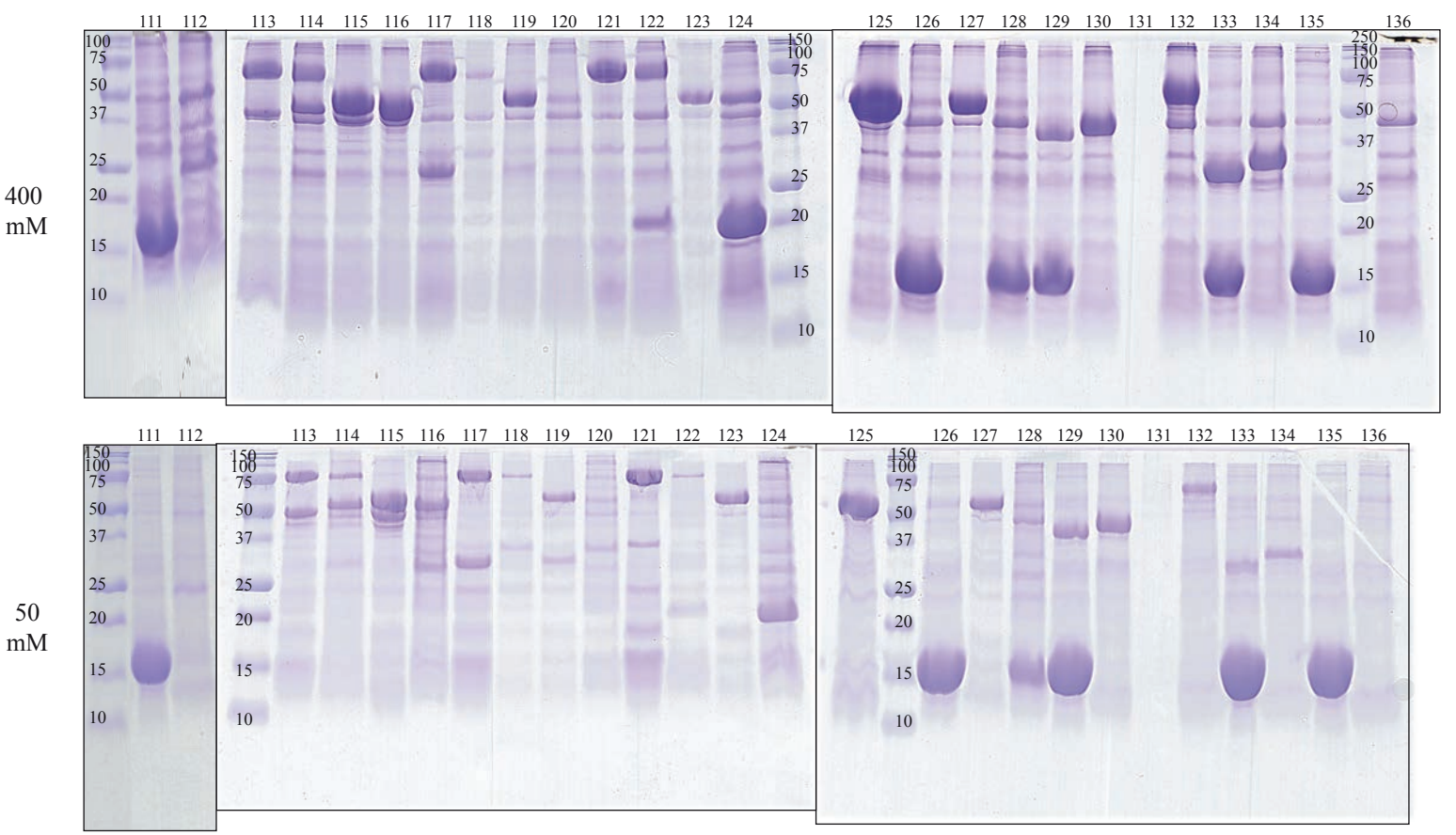
Nous avons également cherché à mieux définir le domaine de Rsa1p qui est responsable de l'association que nous avions observé dans les séries d'expériences précédentes avec la protéine Rrp9p (chapitre Résultats 2.1.2, piste 41). En collaboration avec N. Rolland au laboratoire, un domaine N-terminal de la protéine Rsa1p (résidus 1-224) a été défini et son interaction avec la protéine Rrp9p a été testée en utilisant les séquences codant différents fragments de ces deux protéines que nous avions à disposition (Figure 69, pistes 113 à 124). Dans la piste 117 nous observons des bandes de protéines qui peuvent correspondre à Rrp9p et au fragment 1-224 de Rsa1p, ceci quelle que soit la concentration en sel.

Cette quatrième série d'essais de co-expression révélait aussi la présence d'un complexe formé entre les protéines Snu13p et Pih1p (Figure 64, piste 129) et entre les protéines Snu13p et Prp31p (Figure 69, piste 133) dans les conditions utilisées. Alors que le

B



A



C

SDS-PAGE				
Piste n°	Protéine	P.M. (kDa)	IMAC 50 mM	IMAC 400 mM
111	His-Snu13p	16,03	Oui	Oui
	Hit1p	18,38		
112	His-Hit1p	20,84		
	Snu13p	13,57		
113	His-Rrp9p	67,52	Oui	Oui
	Rsa1p	44	Oui	Oui
114	His-Rsa1p	46,46	Oui	Oui
	Rrp9p	65,05	Oui	Oui
115	His-Rrp9p (113-573)	54,73	Oui	Oui
	Rsa1p	44	Oui	Oui
116	His-Rsa1p	46,46	Oui	Oui
	Rrp9p (113-573)	52,27	Oui	Oui
117	His-Rrp9p	67,52	Oui	Oui
	Rsa1p (1-224)	25,44	Oui	Oui
118	His-Rsa1p (1-224)	28,18	Oui	
	Rrp9p	65,05	Oui	
119	His-Rrp9p (113-573)	54,74	Oui	Oui
	Rsa1p (1-224)	25,44	Oui	
120	His-Rsa1p (1-224)	28,18		
	Rrp9p (113-573)	52,27		
121	His-Rrp9p	67,52	Oui	Oui
	Rsa1p (231-375)	17,93		
122	His-Rsa1p (231-375)	17,93		Oui
	Rrp9p	65,05		Oui
123	His-Rrp9p (113-573)	54,74	Oui	Oui
	Rsa1p (231-375)	17,93		
124	His-Rsa1p (231-375)	17,93	Oui	Oui
	Rrp9p (113-573)	52,27		Oui
125	His-Rrp9p (113-573)	54,74	Oui	Oui
	Snu13p	13,57		
126	His-Snu13p	16,03	Oui	Oui
	Rrp9p (113-573)	52,27		Oui
127	His-Rrp9p (113-573)	54,74	Oui	Oui
	Snu13p mut VSVPV	13,57		
128	His-Snu13p mut VSVPV	16,03	Oui	Oui
	Rrp9p (113-573)	52,27		Oui
129	His-Snu13p	16,03	Oui	Oui
	Pih1p	39,52	Oui	Oui
130	His-Pih1p	41,98	Oui	Oui
	Snu13p	13,57		
131	His-Snu13p	16,03		
	Prp31p	56,3		
132	His-Prp31p	59,04	Oui	Oui
	Snu13p	13,57		
133	His-Snu13p	16,03	Oui	Oui
	Prp31p (78-346)	30,71		Oui
134	His-Prp31p (78-346)	33,58	Oui	Oui
	Snu13p	13,57		
135	His-Snu13p	16,03	Oui	Oui
	Nip1p	93,2		
136	His-Nip1p	95,94		
	Snu13p	13,57		

Figure 69. **Analyse par SDS-PAGE à 15 % des interactions testées lors de la quatrième série d'expériences de co-expression.** (A) Les numéros des pistes sont indiqués au dessus des gels. Les concentrations en sels (50 ou 400 mM) des tampons de purification sont indiquées à gauche des gels. La souche d'*E. coli* utilisée est la souche BL21 (DE3) pRAR2. (B) Schéma représentant l'ensemble des couples de protéines utilisés. Les interactions détectées par co-expression qui ont été validées par la suite sont représentées en violet et celles nécessitant une vérifiaction en brun. (C) Tableau synthétisant les résultats obtenus par le fractionnement par SDS-PAGE . Les protéines exprimées grâce aux vecteurs pnEA-3CH et pnCS sont respectivement indiquées en rouge et en bleu.

complexe Snu13p-Pih1p était insensible aux conditions salines, le complexe Snu13p-Prp31p était plus stable en forte concentration en sel. Comme déjà observé auparavant avec les complexes contenant la protéine Snu13p, les deux complexes n'était détectés que lorsque l'étiquette 6xHis était portée par Snu13p (Figure 69, pistes 129,130 et 133,134)

2.7.2 Deuxième série d'essais du second ensemble de co-expression

Dans cette cinquième série d'expérience de co-expression, l'utilisation du vecteur pnYC nous a permis de réaliser des essais de co-expression à trois partenaires sans concaténation de l'ensemble des ORF. Cependant, dans cette série d'expression, nous n'avons pas obtenu de résultats très clairs. Nous avons néanmoins utilisé la concaténation des ORF de Pih1p et His-Tah1p dans le vecteur pnEA-3CH (Figure 72, pistes 179 à 184). Même si dans les pistes 179, 180 et 181 nous observons bien trois partenaires protéiques, la forte différence d'intensité entre les bandes correspondant à Pih1p et His-Tah1p d'un côté et Nop58p (159-412), Rsa1p et Rsa1p (1-224) de l'autre laisse suggérer une fixation aspécifique de ces deux dernières protéines sur les billes TALON^{TM}.

Lors cette cinquième série d'expérience, l'utilisation de la concaténation des ORF Rsa1p (231-375) et His-Snu13p nous a permis de vérifier l'expression de ces deux protéines à partir de ce système (Figure 72, piste 159, extraits totaux) même si le pourcentage du gel SDS n'était pas adapté à la visualisation individuelle de chaque partenaires de l'hétérotrimère.

Il est important de noter qu'au cours de cette série d'expériences nous avons bien reproduit les résultats obtenus lors de la première série d'expérience (paragraphe 2.1.1) en ce qui concerne le complexe Bcd1p-Rrp9p (Figure 72, pistes 149 à 152). Ce complexe était plus stable à forte concentration saline et lorsque l'étiquette 6xHis était portée par Rrp9p (entière ou son fragment 113-573) (pistes 149 et 151).

Le profil de fractionnement des protéines dans les pistes 139, 142, 163, et 173 est similaire, notamment en faible concentration saline. Or, le seul point commun de ces quatre dépôts est la présence simultanée des protéines Rsa1p et Snu13p portant l'étiquette 6xHis. Nous pensons donc que la bande protéique observée dans ces quatre pistes, qui correspond à une protéine légèrement supérieure à 25 kDa, correspond à un fragment de la protéine Rsa1p qui doit subir une protéolyse au cours de la purification. Ce fragment possèderait néanmoins la capacité à se lier à la protéine Snu13p.

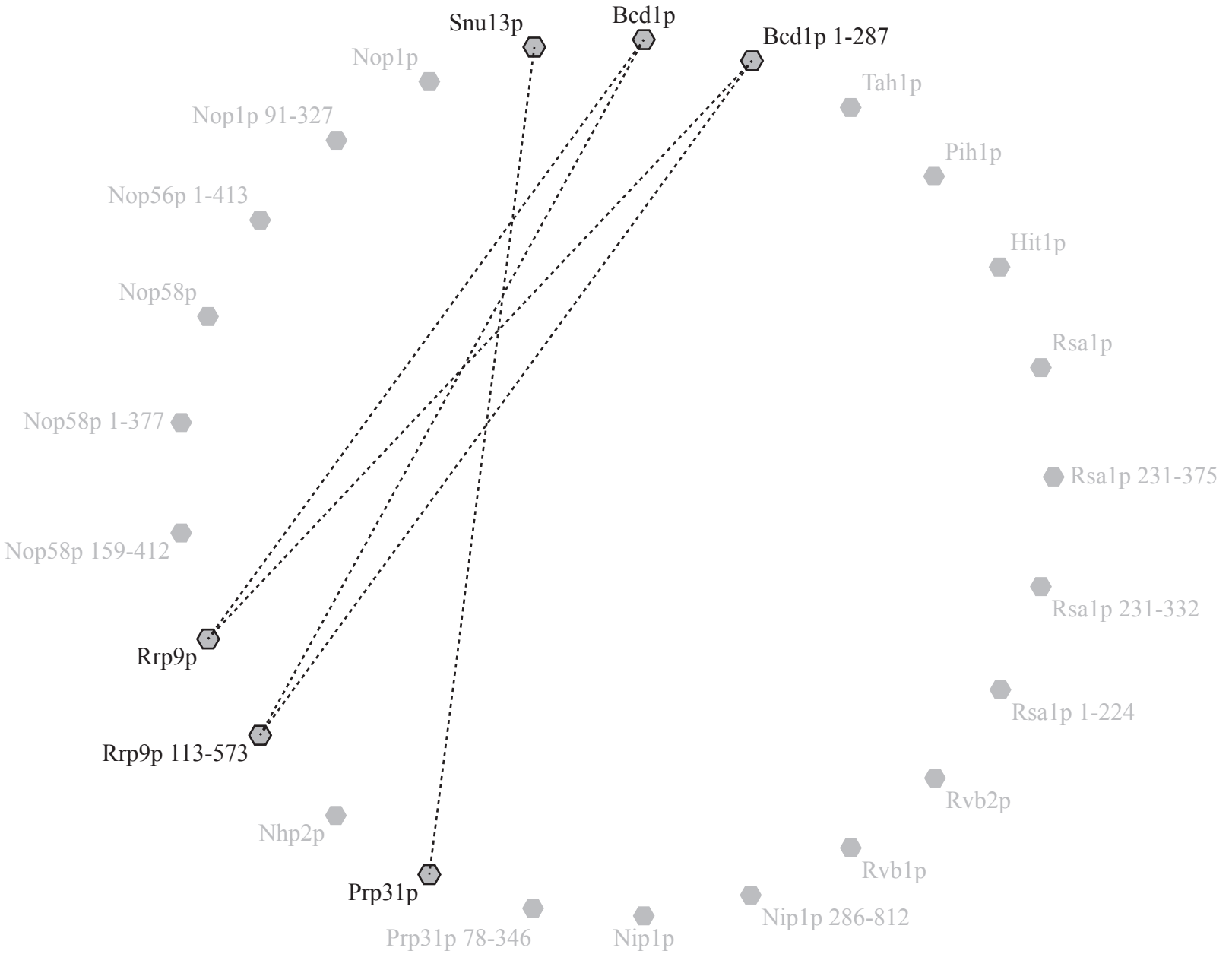


Figure 70. **L'ensemble des interactions à deux partenaire testée dans la cinquième série de co-expression.** Les traits symbolisent les interactions testées.

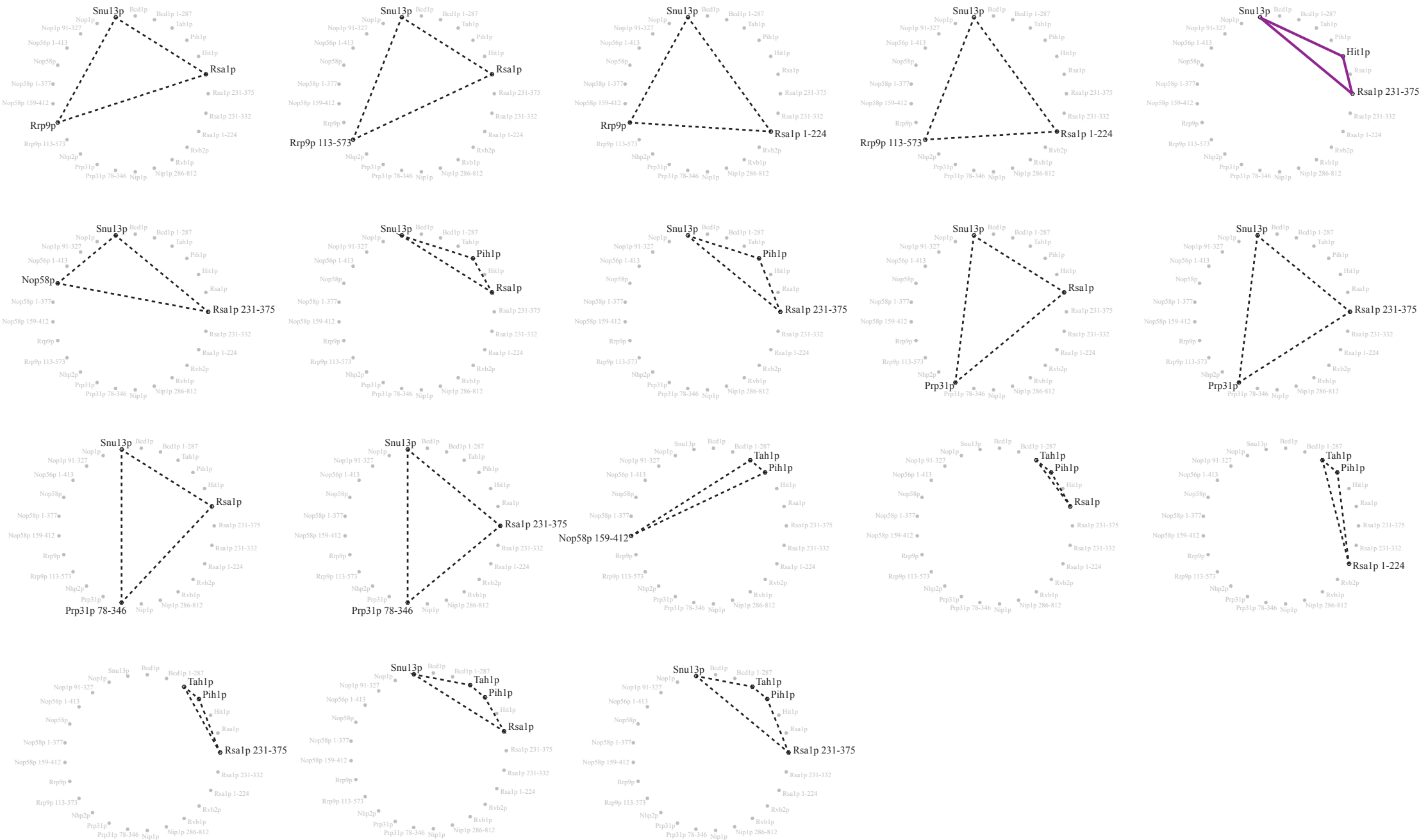


Figure 71. **L'ensemble des interactions à trois partenaires testées dans la cinquième série de co-expression.** Les traits relient l'ensemble des partenaires d'un complexe testé.

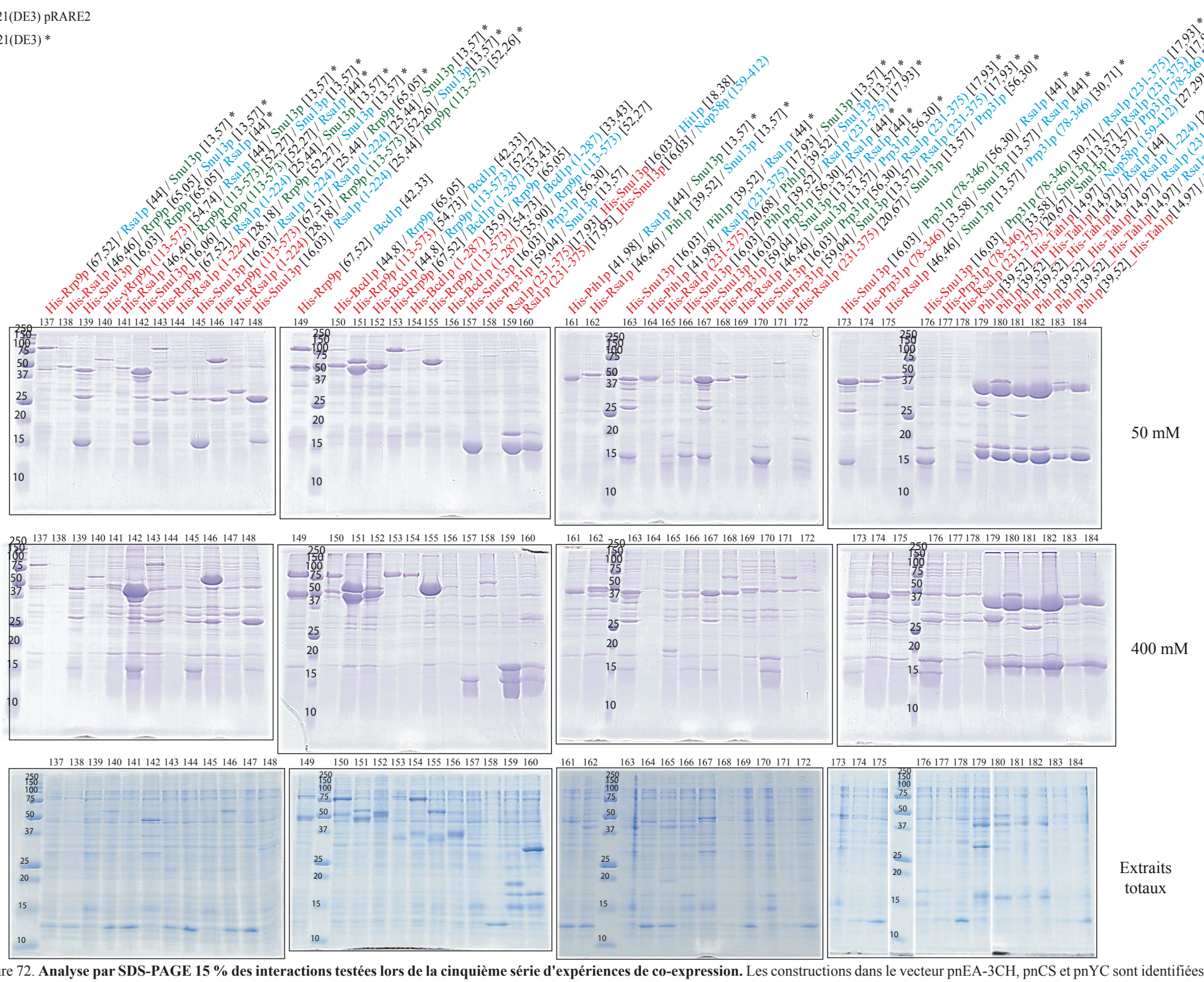


Figure 72. **Analyse par SDS-PAGE 15 % des interactions testées lors de la cinquième série d'expériences de co-expression.** Les constructions dans le vecteur pnEA-3CH, pnCS et pnYC sont identifiées en rouge, bleu et vert respectivement. Les extraits totaux correspondent aux sonicats avant centrifugation. Les concentrations en sels (50 et 400 mM) des tampons de purification sont indiquées à droite des gels. Les souches d'*E. coli* utilisées sont indiquées.

2.7.3 Troisième série d'essais du second ensemble de co-expression

Pour cette sixième série d'expérience, nous avons utilisés les plasmides pCDF::Rvb1p et pKHS::Rvb1p que nous a fournis Edouard Bertrand notre collaborateur de l'IGM à Montpellier. Ces plasmides nous ont permis d'intégrer les deux hélicases Rvb1p et Rvb2p dans les tests de co-expression (les constructions avec la protéine Rvb2p ont été réalisées au laboratoire). Les données obtenues suggèrent l'existence d'une interaction entre la protéine Prp31p, un constituant de la snRNP U4 et la protéine Rvb1p (Figure 75, pistes 221, 222, 223 et 224).

A forte concentration en sel, les deux protéines Hit1p et Prp31p sont détectées dans les pistes 217 et 219 (Figure 75) et le taux des deux protéines semble équivalent. Hit1p est présente aussi bien lorsque Prp31p ou son fragment 78-346 fusionnés à l'étiquette 6xHis sont co-exprimées.

2.7.4 Quatrième série d'essais du second ensemble de co-expression

Dans cette septième série d'expérience, les deux protéines Nop58p (159-412) et Pih1p (Figure 78, pistes 257 et 258) sont clairement visibles avec un effet de la concentration saline différente lorsque c'est soit Nop58p (159-412) soit Pih1p qui porte l'étiquette 6xHis.

Afin d'aller plus loin dans l'étude des interactions, nous avons utilisé dans cette dernière série d'essai de co-expression des ORF mutées de Snu13p et Rsa1p, ceci afin d'étudier les complexes Snu13p/Rsa1p et Snu13p/Rsa1p (231-381). Nous disposions en effet d'une ORF Snu13p codant le mutant $E_{72}D_{73}K_{74}$/AAA et d'une ORF codant le mutant $R_{249}A$ de Rsa1p. Nous avons donc introduit ces ORF mutés dans les plasmides de co-expression et testé l'effet de ces mutations sur l'association. Par cette approche, nous avons pu montrer très clairement que les mutations $E_{72}D_{73}K_{74}$/AAA dans Snu13p ou $R_{249}A$ dans Rsa1p (entière ou son fragment 231-381) abolissent l'association de ces deux protéines (Figure 78, pistes 235 à 238).

L'utilisation d'autres mutants de cette interface analysée par co-expression sera décrite plus loin dans un article en préparation (Manuscrit n°2).

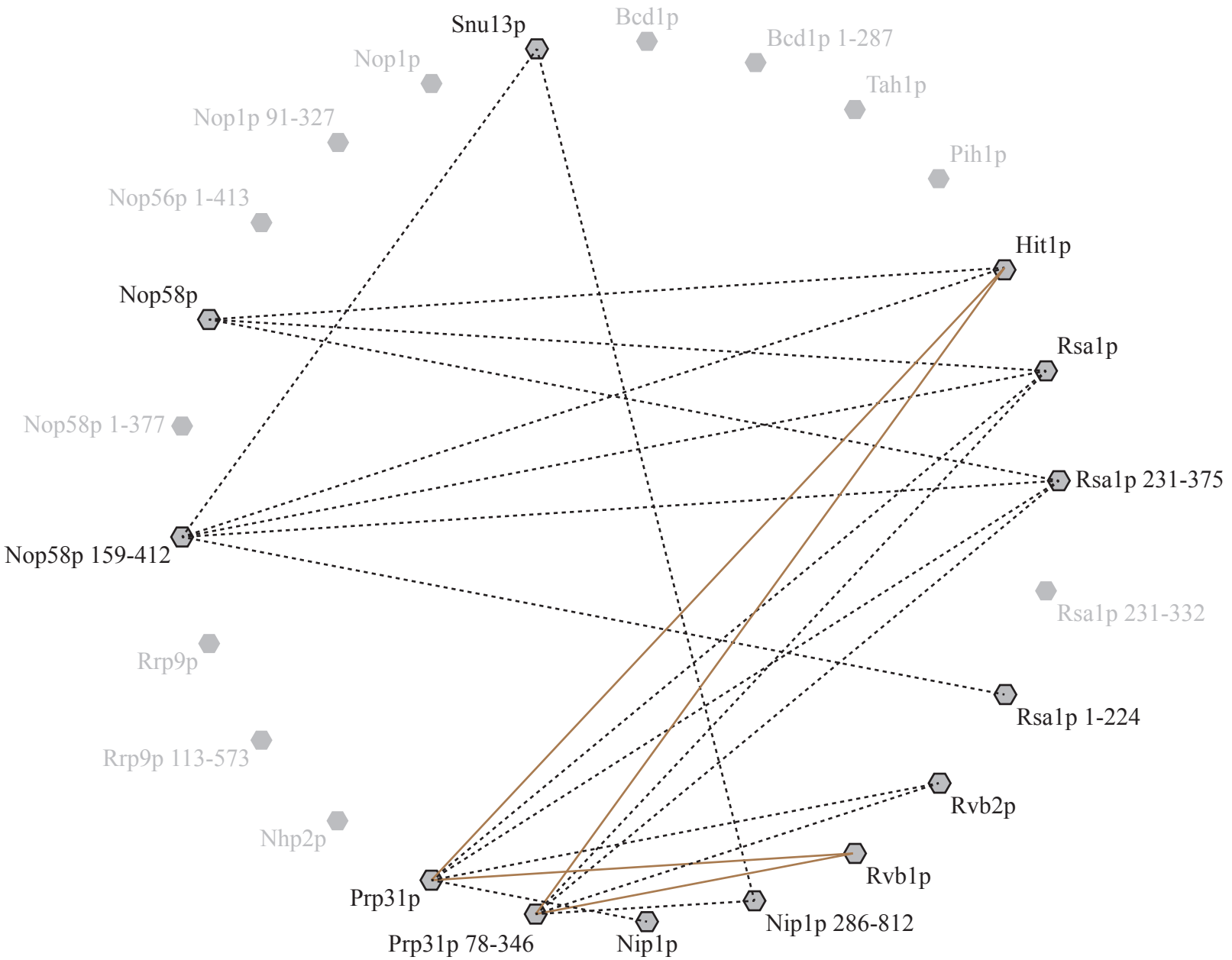


Figure 73. **L'ensemble des interactions à deux partenaire testée dans la cinquième série de co-expression.** Les traits symbolisent les interactions testées. Les intéractions qui sembleraient s'établir mais devront être revérifiées par d'autres approches sont représentées en brun

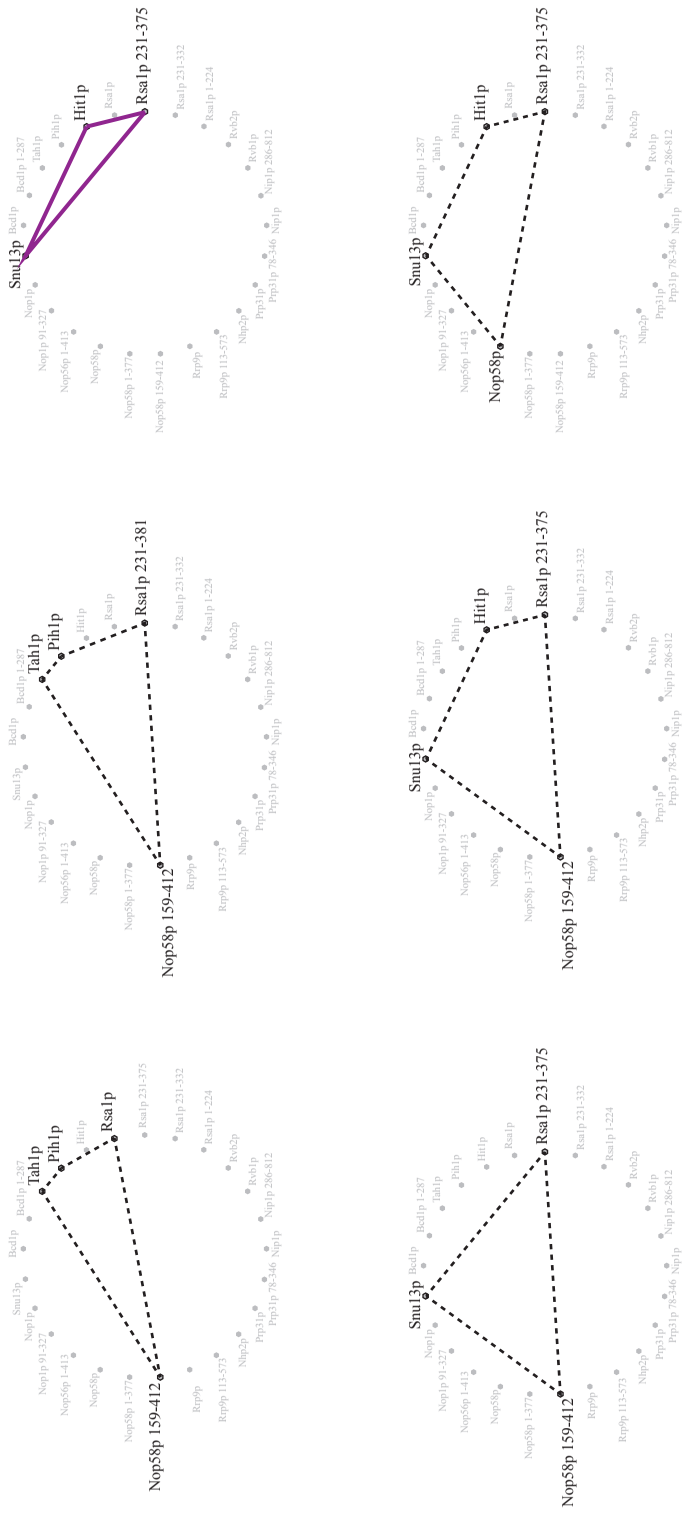


Figure 74. **Bilan des interactions à trois partenaires testées dans la sixième série d'expériences co-expression.** Les traits relient l'ensemble des partenaires d'un complexe testé.

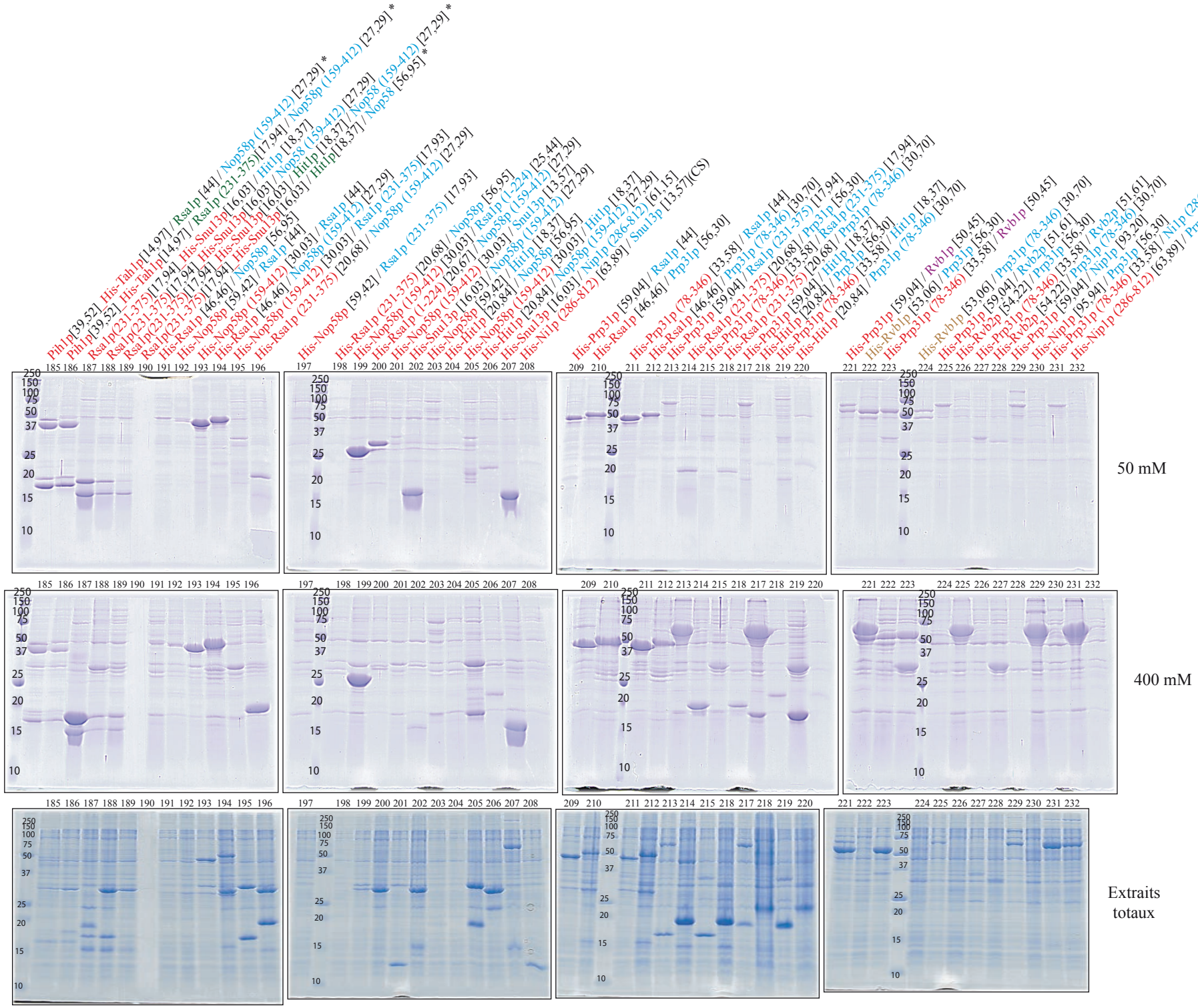


Figure 75. **Analyse par SDS-PAGE 15 % des interactions testées lors de la sixième série d'expériences co-expression.** Les constructions dans le vecteur pnEA-3CH, pnCS et pnYC sont identifiées en rouge, bleu et vert respectivement. Les extraits totaux correspondent aux sonicats avant centrifugation. Les concentrations en sels (50 et 400 mM) des tampons de purification sont indiquées à droite des gels. Les souches d'*E. coli* utilisées snot indiquées en haut à gauche.

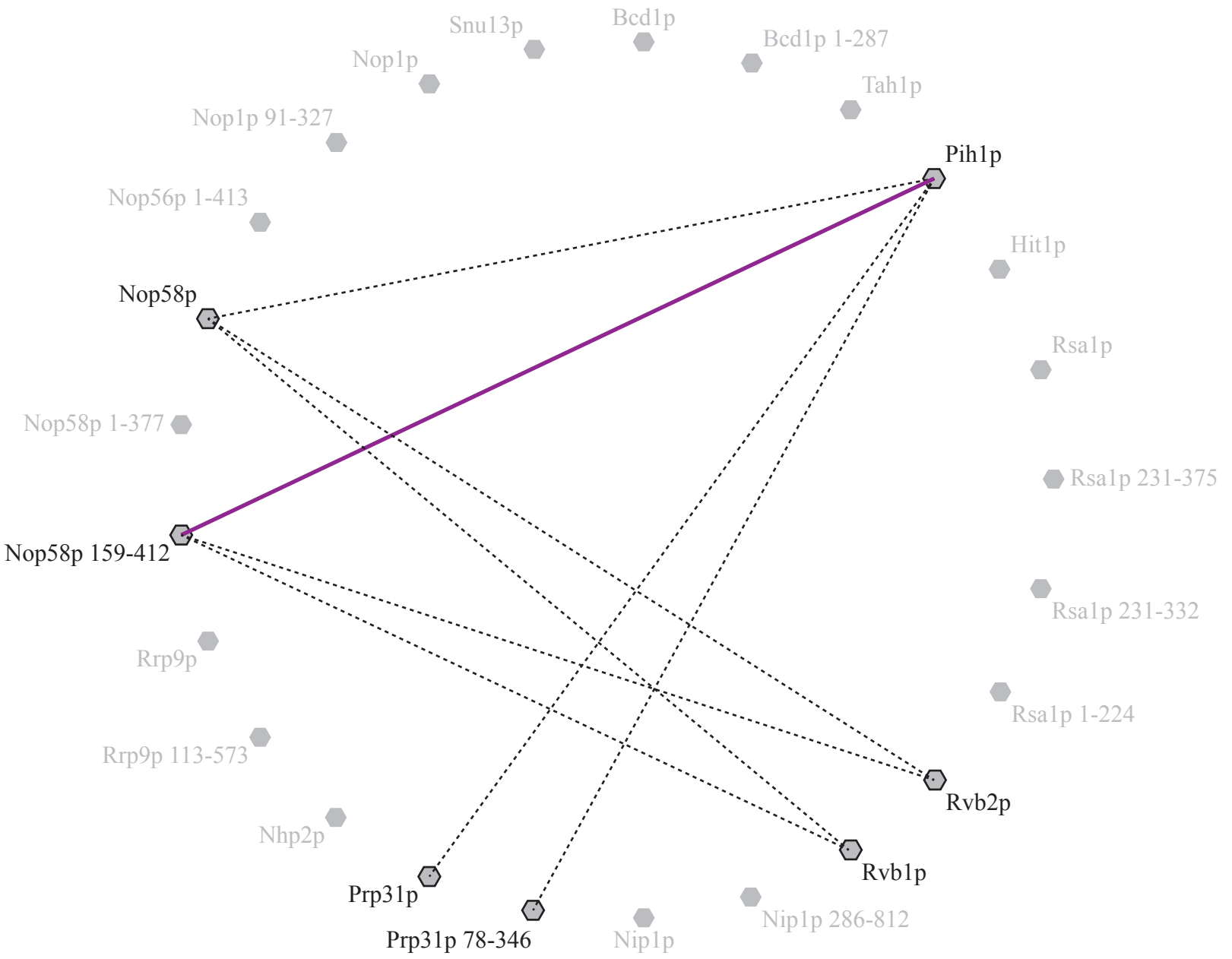


Figure 76. **L'ensemble des interactions à deux partenaire testée dans la septième série de co-expression.** Les traits symbolisent les interactions testées. Les interactions qui ont été validées par la suite par d'autres approches sont symbolysées en violet.

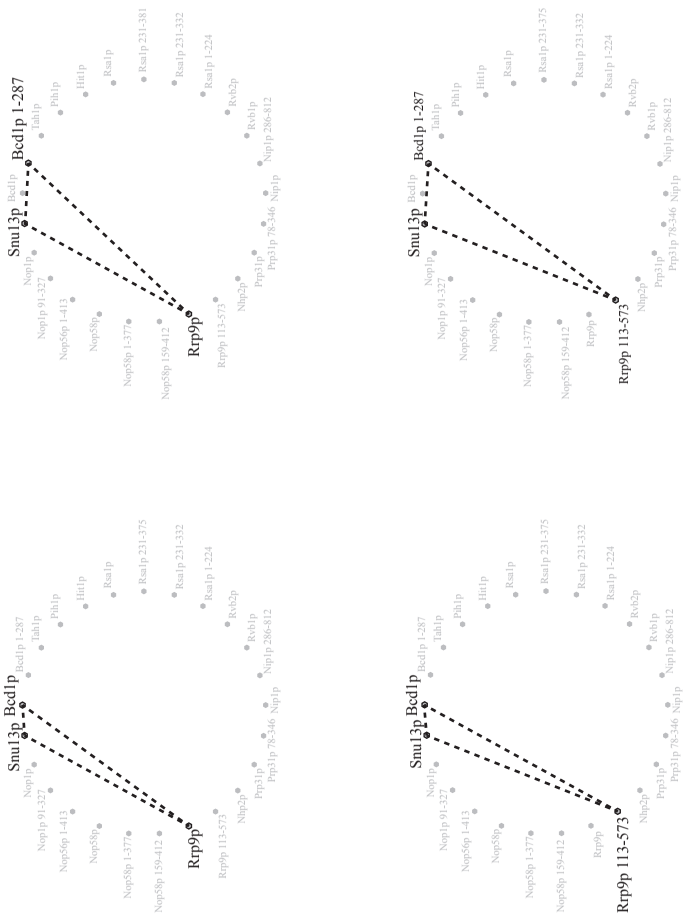


Figure 77. **Bilan des interactions à trois partenaires testées dans la septième série d'expériences co-expression.** Les traits relient l'ensemble des partenaires d'un complexe testé.

BL21(DE3) pRARE2
BL21(DE3)*
pCDF (E.Bertrand)

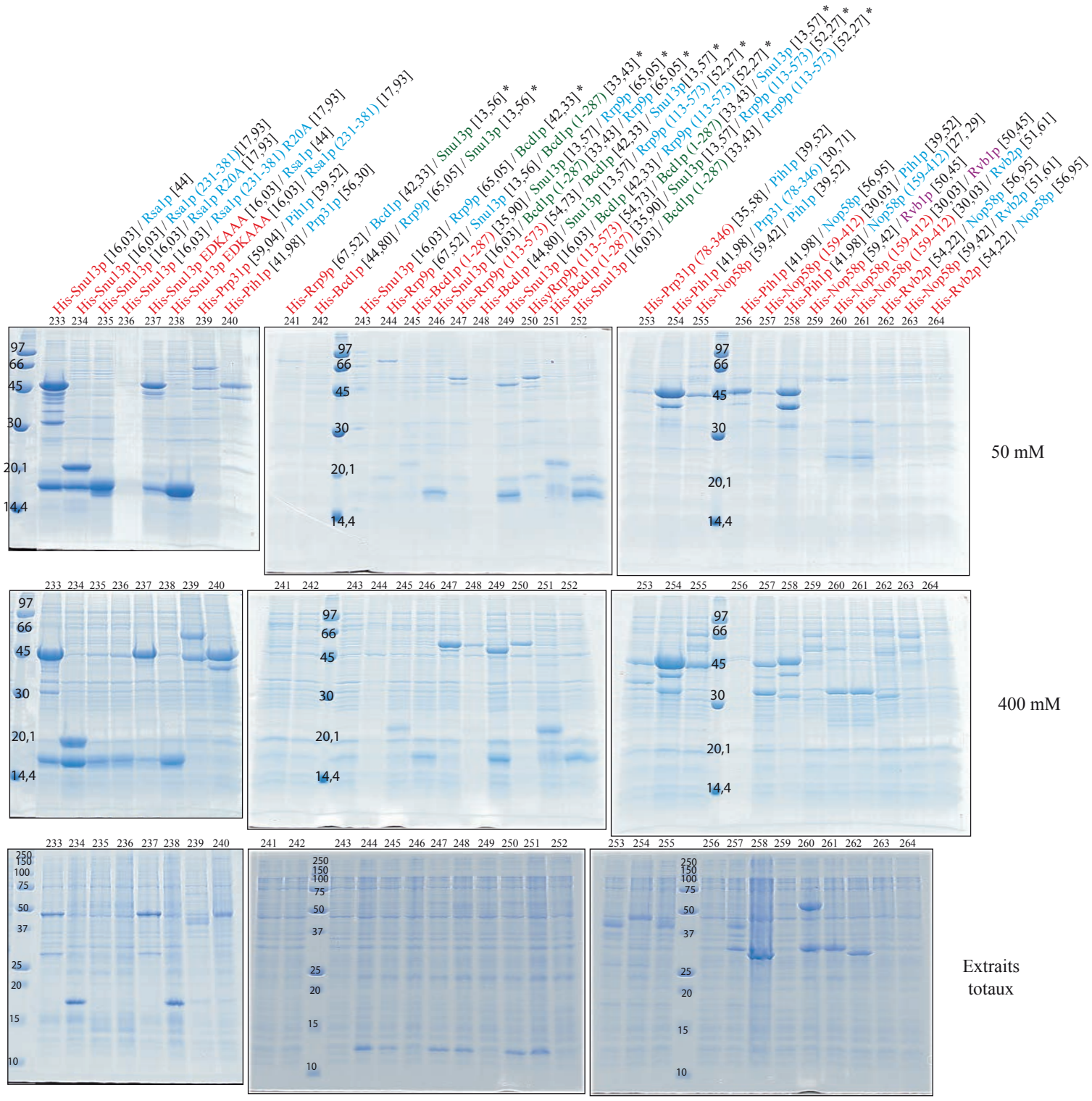


Figure 78. **Analyse par SDS-PAGE 15 % des interactions testées lors de la septième série d'expériences co-expression.** Les constructions dans le vecteur pnEA-3CH, pnCS et pnYC sont identifiées en rouge, bleu et vert respectivement. Les extraits totaux correspondent aux sonicats avant centrifugation. Les concentrations en sels (50 et 400 mM) des tampons de purification sont indiquées à droite des gels. Les souches d'*E. coli* utilisées sont indiquées en haut à gauche.

3 Conclusion et Perspectives

3.1 Bilan

Un bilan des 188 essais d'interaction à deux partenaires dans les expériences de co-expression est présenté dans la figure 79. Cette approche par co-expression nous a permis de valider un grand nombre des interactions mises en évidence par double hybride, co-immuno-précipitation ou expérience de GST Pull down. Par ailleurs, elle nous a permis de détecter d'autres interactions potentielles à confirmer par d'autres approches (Figure 80) et pour les interactions déjà existantes, de préciser les domaines d'interaction (Figure 81).

Grâce à cette approche, nous avons également pu définir des conditions permettant de produire de façon soluble des protéines, ou fragment de protéines qui, exprimés individuellement, ne l'étaient pas. C'est le cas par exemple pour la protéine Rsa1p et son fragment 231-381. Les protéines ainsi produites ont pu être utilisées dans des expériences de retard sur gel afin d'analyser leur capacité à lier les ARN comme nous le décrirons plus loin dans le manuscrit n°1.

Certains des complexes mis en évidence ont également fait l'objet d'études structurales, en particulier le complexe Snu13p/Rsa1p/Hit1p. Comme nous le verrons dans le chapitre 2 des résultats, la co-expression des trois protéines nous a permis d'obtenir ce complexe en très grande quantité et à un degré de pureté élevé. L'expérience acquise au cours de ces travaux nous permets de conclure que couplée à une analyse DLS (*« Dynamic Light Scattering »*) et à la protéolyse ménagée, l'approche de co-expression est un outil efficace pour l'obtention de complexes macromoléculaires de qualité en vue de réaliser des études structurales.

3.2 Vers l'étude d'entité multimériques

Au sein de processus biologiques cellulaires, une protéine n'est et n'agit jamais seule. De plus les études structurales effectuées jusqu'à maintenant tendent à suivre cette vision. En effet, sur l'ensemble des protéines d'une cellule, peu d'entre elles possèdent les caractéristiques d'expression, de solubilité et de purification adaptée à une étude structurale à l'état individuel. Les critères de pureté et de quantité nécessaire à la cristallographie des rayons X, la RMN ou encore le SAXS (*« Small-angle X-ray scattering »*) et la microscopie

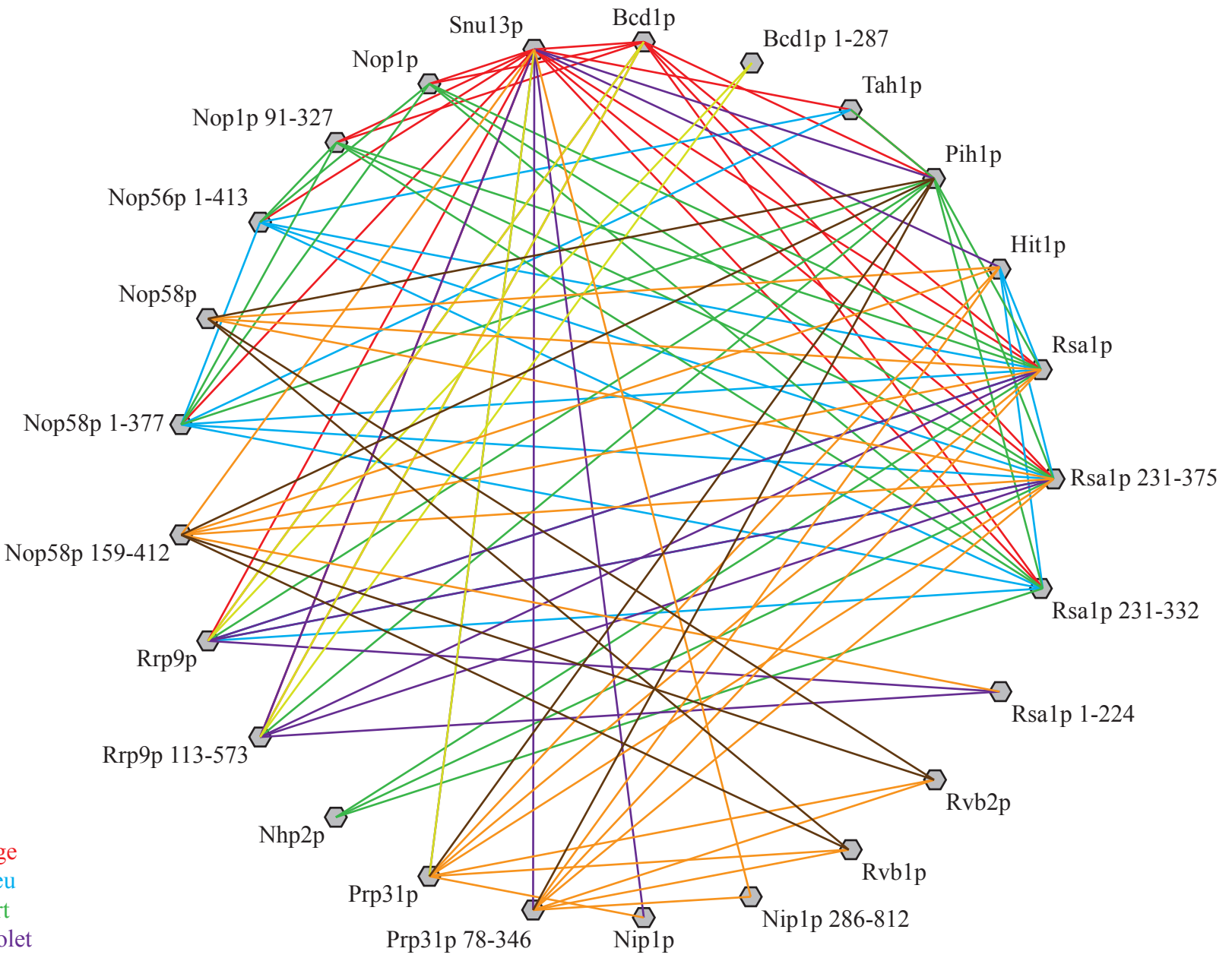


Figure 79. **Bilan de l'ensemble des essais d'interactions à deux partenaires dans les expériences de co-expression à haut débit.** Le code couleur pour chaque série d'expériences de co-expression est indiqué.

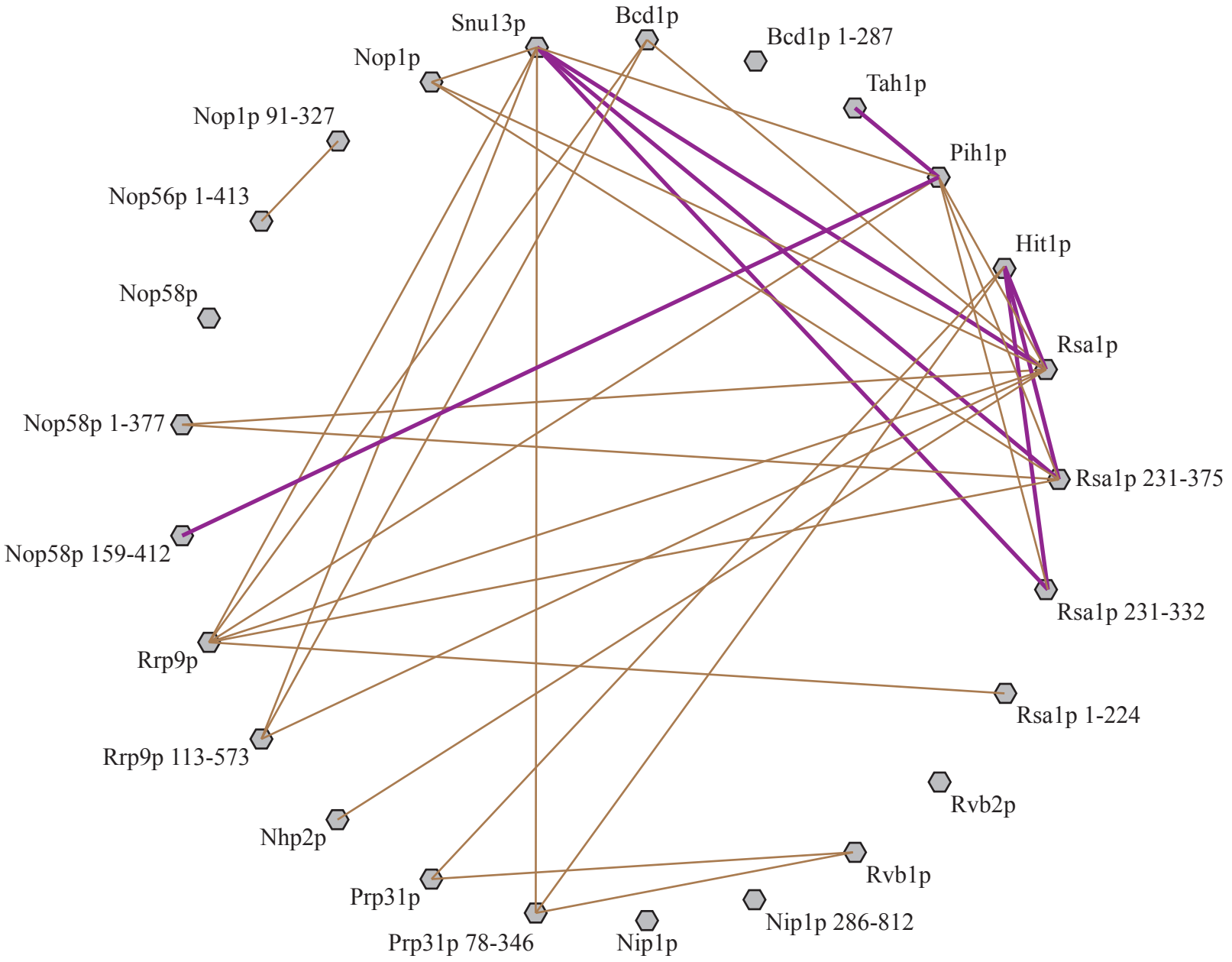


Figure 80. **Bilan des interactions à deux partenaires observées au cours de l'étude par co-expression.** Les interactions observées qui nécessites une validation par d'autres approches sont représentées en brun et celles déjà validées par d'autres approches par des traits épais violet.

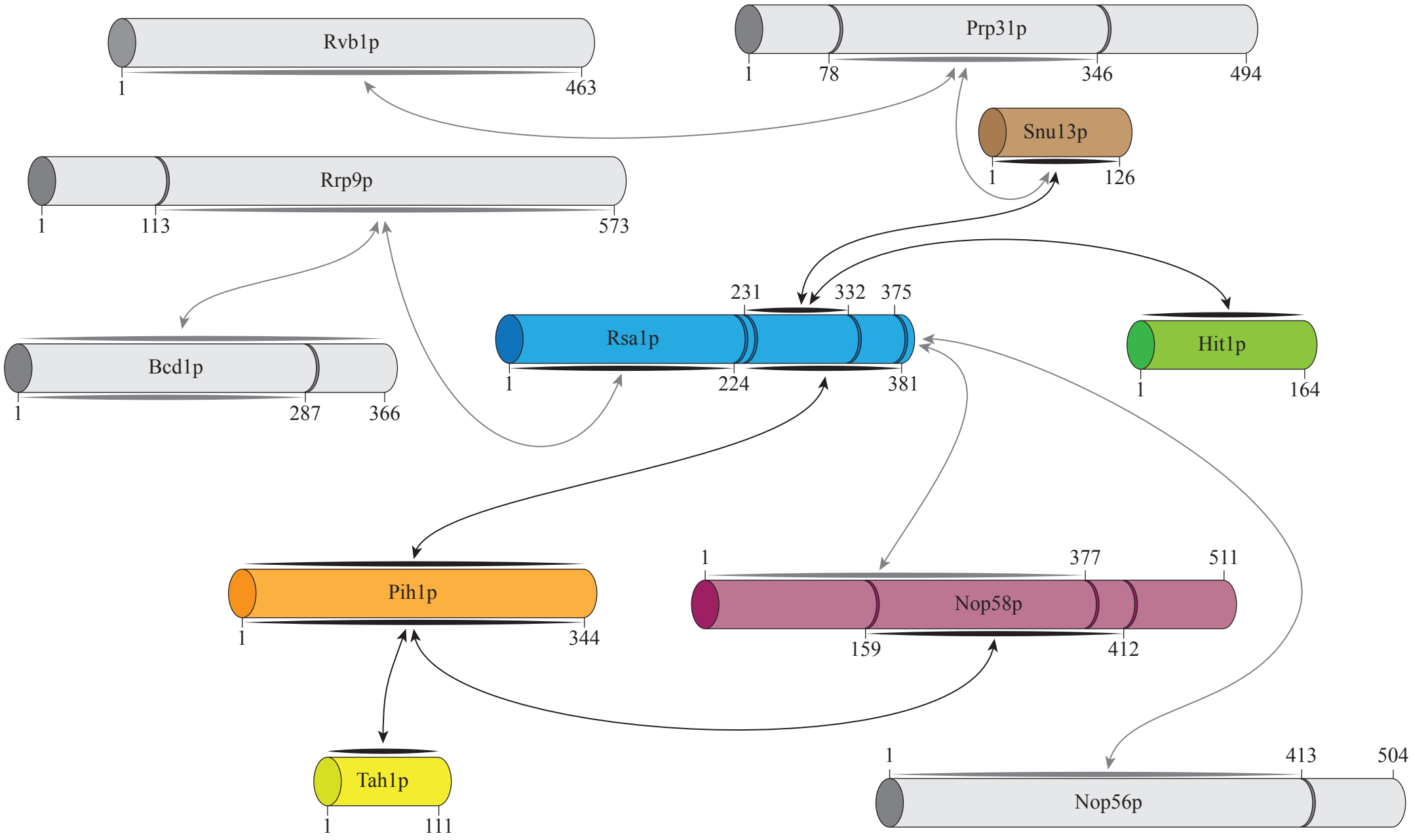


Figure 81. **Bilan des interactions mises en évidences par les séries d'expériences de co-expression.** Les protéines sont représentées sous forme de cylindre dont la longueur est proportionnelle à leur taille en acides aminées. Les bornes des différents domaines des protéines sont indiquées par les numéros des résidus. Les protéines en couleurs correspondent aux protéines qui ont fait l'objet d'une étude plus appronfondie lors de la thèse.

électronique amènent bien souvent à n'étudier qu'une partie des protéines d'intérêts ou même d'y renoncer.

La co-expression offre une alternative à ce problème en proposant d'exprimer des partenaires d'interactions sous la forme d'une entité moléculaire unique lors de la purification. Ainsi, des régions de protéines présentant une flexibilité élevée ou des caractères hydrophobes importants pourront être stabilisées par la présence de leur partenaire. La co-expression permet également de respecter la stœchiométrie des complexes évitant ainsi des assemblages fortuits dans le mélange de protéines concentrées lors des dernières étapes de préparation des échantillons. Enfin, l'expression d'une seule protéine porteuse de l'étiquette et de ces partenaires de façon native conduit à une purification relativement simplifiée et plus rapide. La chromatographie d'affinité couplée à une étape d'élution par protéolyse dirigée permet après une chromatographie d'exclusion de taille l'obtention d'un échantillon de grande pureté tout en restant dans des conditions de purification non dénaturante, conditions imposées pour le maintien du complexe.

Malgré cette avancée technique remarquable, il arrive que des complexes observés par des procédés d'extraction *in vivo* comme les immunoprécipitation (IP) ou les IP-SILAC (*« stable isotope labeling of amino acids in cell culture »*) ne peuvent pas être obtenus, même par co-expression. L'absence d'un partenaire, autre que protéique, peut être à l'origine de l'absence de la formation du complexe. Il ne faut pas oublier que les protéines que nous étudions forment une structure ribonucléoprotéique : les RNP à boîtes C/D. La composante nucléique y est donc très importante. Une des possibilités de l'absence de formation de certains complexes pourrait être l'absence de cette composante nucléique. La synthèse *in vitro* des ARN reste à l'heure actuelle un défi pour les études structurales à cause des contraintes d'homogénéité et de pureté imposées. Des systèmes d'expression dans le système hétérologue *E. coli* ont été développés ces dernières années (Ponchon *et al.*, 2009; Ponchon & Dardel, 2011) afin de palier à ces problèmes. Ces systèmes font appel à l'utilisation d'un ARN stable dans la cellule bactérienne : les ARNt. Le principe est d'insérer l'ARN cible au sein de la boucle anticodon d'un ARNt *« scaffold »*, utilisant ainsi les caractéristiques de stabilité et d'expression de cet ARNt au sein de la cellule. L'insertion d'un aptamère capable de se lier à une autre molécule comme le Séphadex permettra de purifier la molécule chimérique comme l'on ferait avec une protéine porteuse d'une étiquette d'affinité. L'une des difficultés de cette approche reste l'élution ou la séparation de l'ARN d'intérêt de son ARNt *« scaffold »*. En effet, il n'existe pas ou peu d'enzyme capable d'hydrolyser une liaison phosphodiester à un

site spécifique comme le font les protéases pour les liaisons peptidiques. L'utilisation de la RNAse H apporte une solution à ce problème mais son utilisation et sa mise en œuvre reste délicate et très coûteuse. En effet, elle nécessite l'utilisation de longs oligodésoxynucléotides qui s'hybrideront à une séquence précise de l'ARN pour former un hétéroduplexe, le substrat de la RNAse H. L'obtention de tels oligodésoxynucléotides en grande quantité et à tel niveau de pureté ainsi que leur recyclage reste onéreux et complexe. De plus, la précision de la position et la reproductibilité de l'hydrolyse de la liaison phosphodiester de l'ARN cible par la RNAse H ne sont pas assurées.

Le passage par un ARNt *« scaffold »* est une réponse au problème de la rapide dégradation des ARN « non protégés » chez *E. coli*. Une autre solution serait de mimer un ARNm avec sa coiffe mais là encore la mise en œuvre d'une telle stratégie resterait fastidieuse. L'utilisation des protéines partenaires de l'ARN d'intérêt dans le repliement et la protection de celui-ci pourrait apporter une solution. La co-expression au sein d'une même cellule des partenaires protéiques et nucléiques permettrait de former des complexes qui ne s'établiraient pas en absence d'ARN mais aussi de protéger cet ARN par la formation d'une particule. Nous parlerions alors de co-expression ARN-protéines. Même si la mise en place d'un tel système reste un challenge, les bénéfices dans l'étude structurale de tels complexes ribonucléoprotéiques seraient immenses.

CHAPITRE 2

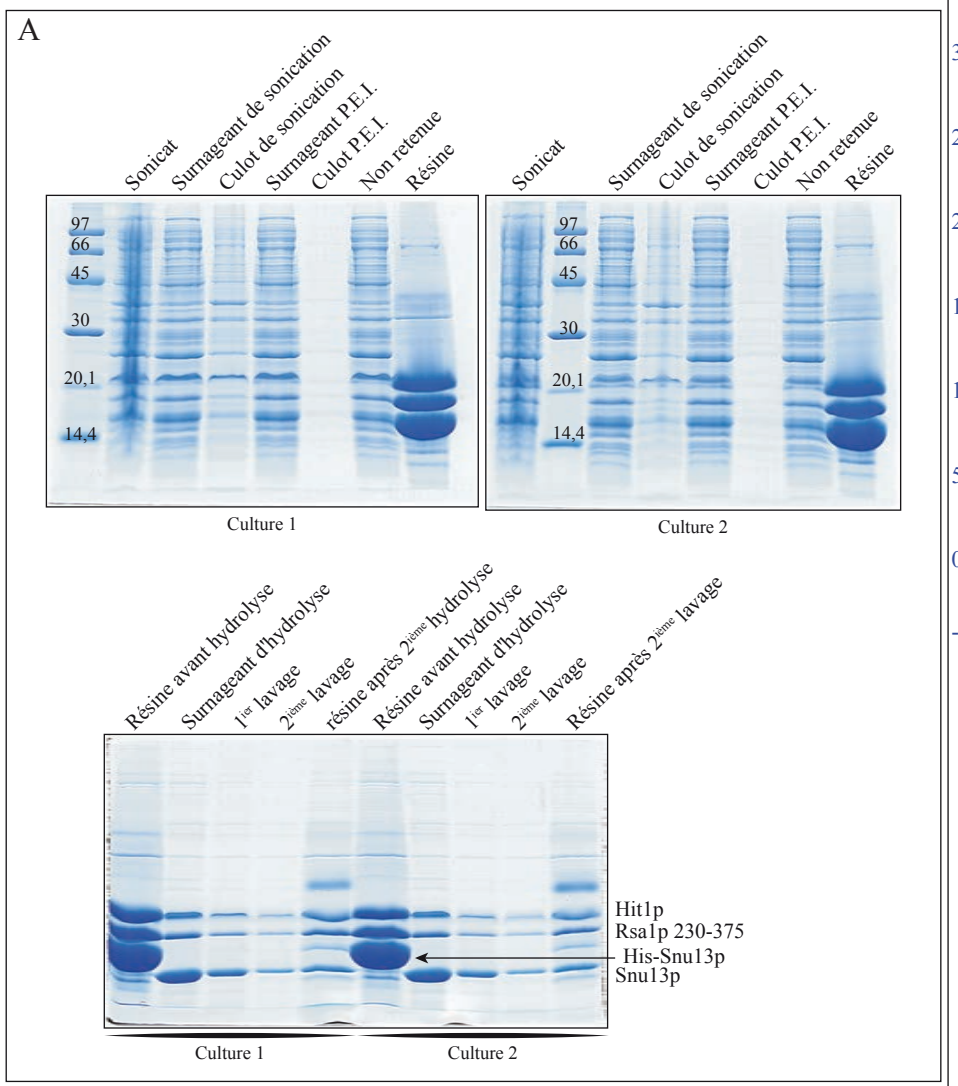
La machinerie d'assemblage des snoRNP à boîtes C/D : rôle de la protéine Hit1p

1 Etude structurale sur le complexe Snu13p/Rsa1p/Hit1p

1.1 Production du complexe Snu13p/Rsa1p (231-375)/Hit1p

Le résultat de la purification du complexe Snu13p/Rsa1p (231-381)/Hit1p (Résultats & Discussions, Chapitre 1, paragraphe 2.6) étant très concluant, nous avons entrepris sa production et sa purification à plus grande échelle afin de réaliser des cribles de cristallogenèse. Nous avons conservé les deux étapes de purification, d'abord sur la résine TALON™ retenant les protéines contenant un enchaînement d'histidines et ensuite par chromatographie d'exclusion de taille. Lors de la fixation sur la résine TALON™, nous avons toujours utilisé une faible quantité de résine afin d'accroitre la compétition entre la protéine portant l'étiquette et les contaminants protéiques d'*E. coli*. Ceci nous a permis d'obtenir une fixation plus spécifique et de limiter les contaminants protéiques à éliminer lors des étapes suivantes. La figure 82 représente les résultats obtenus lors des différentes étapes de purification à grande échelle du complexe hétérotrimérique. L'analyse en spectrométrie de masse réalisée sur le complexe His-Snu13p/Rsa1p (231-381) montrait une dégradation de l'extrémité C-terminale de la protéine Rsa1p (Résultats, Chapitre 1, paragraphe 2.3.2). La région codant le sous-fragment sans la partie C-terminale a été clonée pour produire le domaine correspondant de Rsa1p (231-375). Dans le profil de fractionnement des protéines lors de la chromatographie d'exclusion de taille, nous détections très fréquemment un pic correspondant à la protéine Snu13p non complexé (Figure 82), ceci du fait que Snu13p qui portait l'étiquette 6xHis, était présente en plus grande quantité que Rsa1p après élution du complexe fixé sur la résine TALON™.

Afin de caractériser les propriétés du complexe en vue de réaliser une étude par cristallographie, nous avons entrepris une analyse par une approche de D.L.S. ou *« Dynamic Light Scattering »* (diffusion dynamique de la lumière) du complexe hétérotrimérique dans différents tampons ceci en faisant varier le pH mais aussi la concentration saline et la température. Cette technique permet d'obtenir des informations sur le rayon hydrodynamique



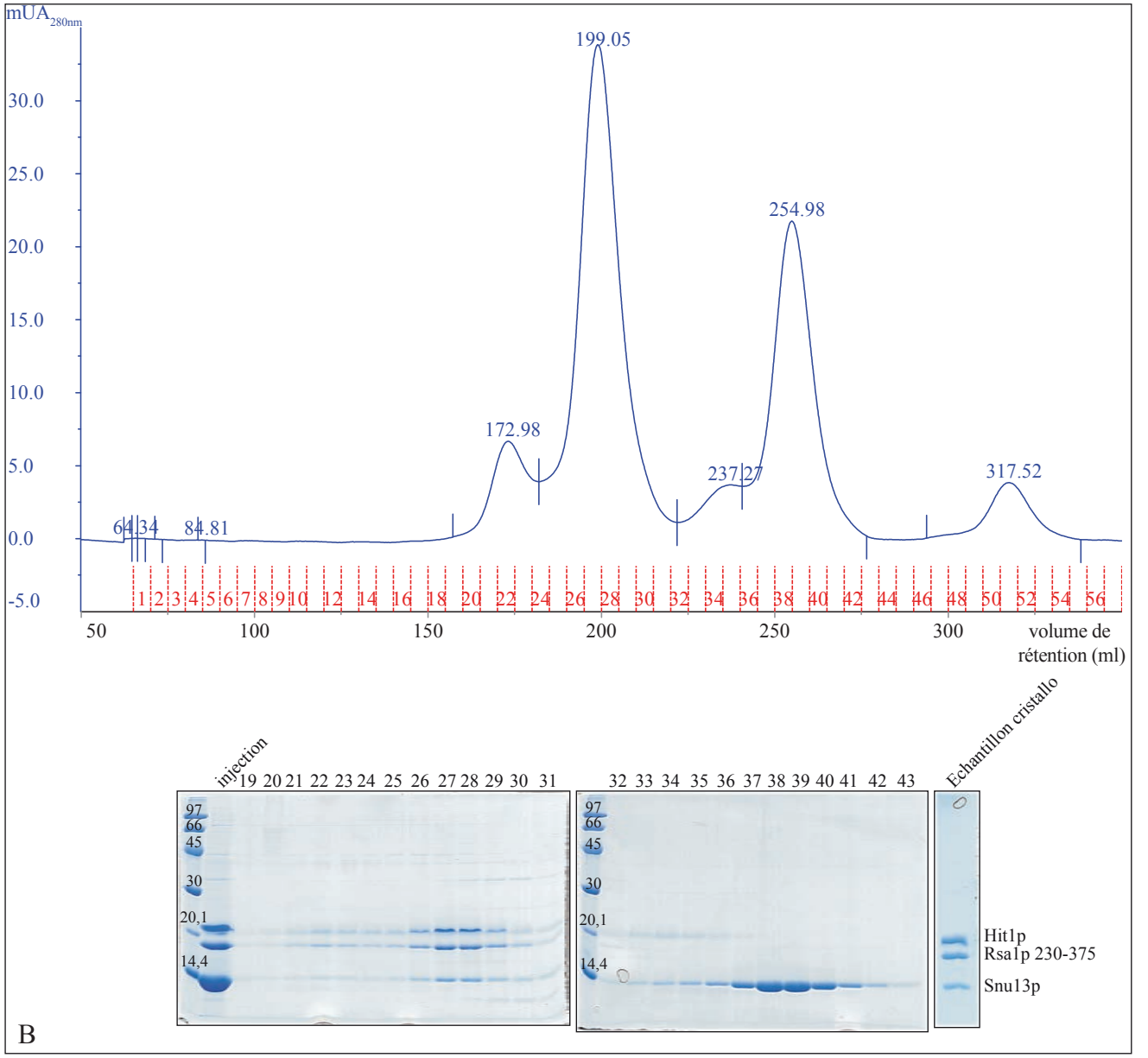


Figure 82. **Purification préparative du complexe His-Snu13p/Rsa1p (231-375)/Hit1p.** (A) Analyse par SDS-PAGE des protéines obtenues lors dès premières étapes de purification. (C) Chromatographie d'exclusion de taille (Superdex 200) du complexe et analyse par SDS-PAGE des fractions correspondant aux pics. La dernière piste correspond à l'échantillon final (13,81 mg/ml) qui a fait l'objet de tests de cristallisation.

(*« Z-Average »*) des molécules de l'échantillon mais aussi d'évaluer leur hétérogénéité (PDI ou *« PolyDispersity Indice »*). L'approche de diffusion dynamique de la lumière repose sur la variation de l'intensité de la lumière diffusée par le mouvement Brownien des particules en solution. Ces variations sont mesurées aux petits angles et au cours du temps. Le faible volume de l'échantillon (12 µl) et la possibilité de l'étudier sur une large gamme de concentration et à différentes températures font du D.L.S. une technique biophysique de choix pour étudier les échantillons destinés à des études structurales. La figure 83 présente les résultats de l'analyse du complexe Snu13p/Rsa1p (231-332)/Hit1p à 12 $mg.ml^{-1}$ (soit 0,3 mM) en fonction du pH, de la concentration saline et de la température. Il apparait clairement que l'homogénéité de ce complexe est optimale à des pH de 6,5 et 7,5. Deux tampons semblaient intéressants à utiliser notamment du fait des faibles variations observées en fonction de la température : Na PIPES pH 6,5 ; 150 mM NaCl et Na HEPES pH 7,5 ; 300 mM NaCl. Ce sont ces deux tampons que nous avons retenus pour la suite du travail.

A l'issue des différentes purifications réalisées, nous avons pu obtenir des échantillons d'une grande pureté avec des concentrations allant de 10 mg/ml à 100 mg/ml. Des tests de cristallisation ont alors été entrepris sur ces échantillons. Nous disposons au laboratoire de 12 kits commerciaux de cristallogenèse, chacun composés de 96 solutions différentes. Ainsi 1152 conditions de cristallogenèse ont été testées. Ceci a été réalisé par l'utilisation d'un robot MOSQUITO® qui permet de préparer des boîtes de cristallogenèse de 96 puits en gouttes assises. Néanmoins, malgré tous nos efforts de combinaisons de domaines protéiques et d'optimisation de conditions de tampon, aucun cristal n'a pu être obtenu, même sous la forme dite « d'oursins ». La présence de la protéine Snu13p, dont la structure a été résolue par cristallographie présente donc des caractéristiques favorables à la cristallogenèse, mais n'a pas été suffisante pour générer une nucléation de la formation de cristaux. La nature flexible des deux autres partenaires pourrait donc être un frein à la cristallisation de ce complexe.

1.2 Protéolyse ménagée du complexe Snu13p/Rsa1p/Hit1p

Afin de palier à ces problèmes de flexibilité de nos complexes, nous avons mis en place, avec l'aide de V. Pena (Max Planck Institute, Göttingen), une stratégie de protéolyse ménagée. Cette approche présente les avantages d'un faible coût et d'une mise en œuvre rapide. Elle repose sur l'accessibilité aux protéases des régions non structurées ou exposées au solvant. En faisant varier le type de protéase, on peut faire varier les sites d'hydrolyse. La

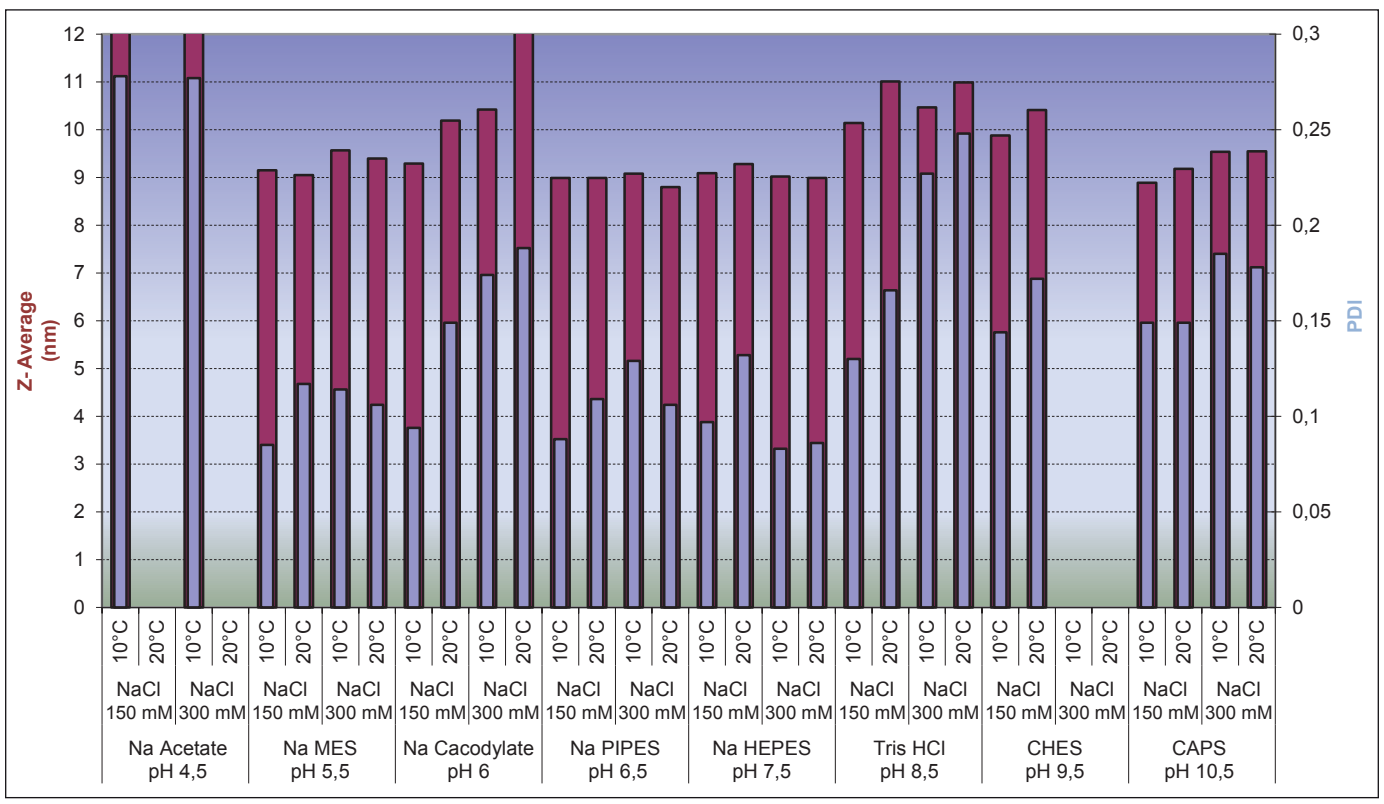


Figure 83. **Analyse par diffusion dynamique de la lumière du complexe Snu13p/Rsa1p (231-332)/Hit1p.** Les échantillons ont été utilisés à une concentration de 12 mg/ml et ont été dialysés contre des tampons dont le pH variait entre 4,5 et 10,5 et à deux concentrations salines différentes (300 et 150 mM NaCl), puis analysés à 10°C et 20°C. L'absence de résultats reflète une précipitation de l'échantillon.

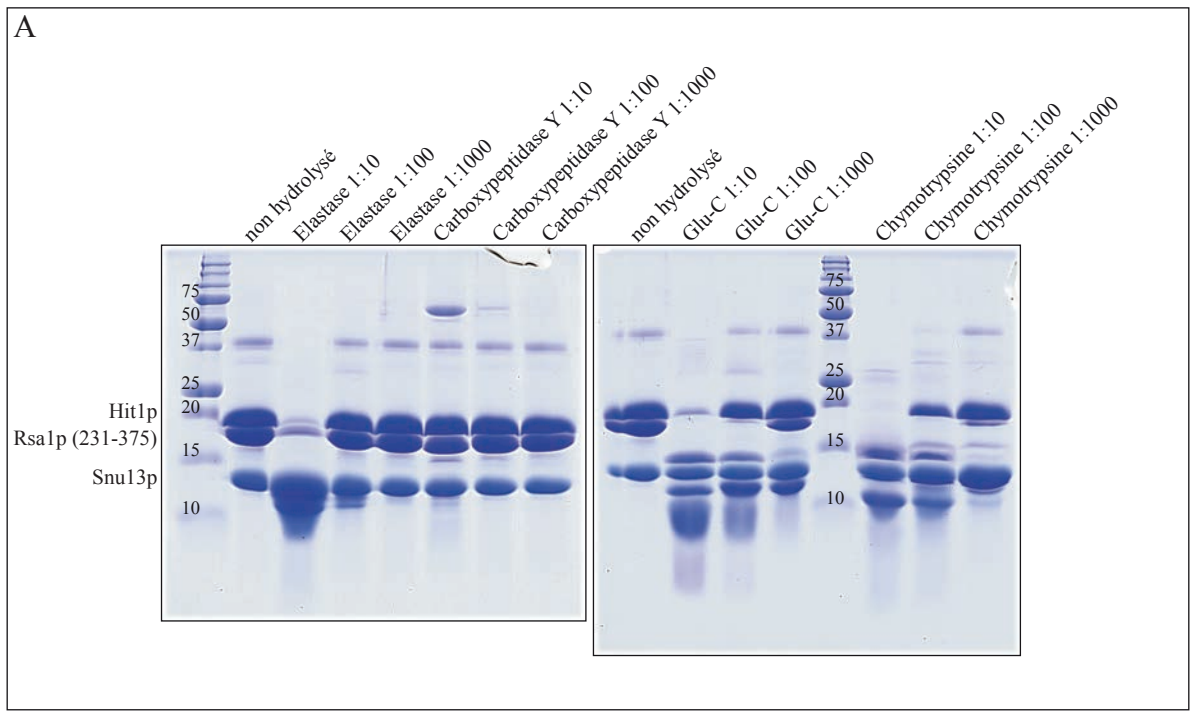
première étape fut le choix de l'enzyme, ainsi que de sa concentration, afin de ne pas avoir une hydrolyse totale de l'échantillon.

1.2.1 Choix de l'enzyme et de sa concentration

Pour mettre au point des conditions de protéolyse ménagée du complexe Snu13p/Rsa1p (231-375)/Hit1p, nous permettant d'éliminer les segments très flexibles, nous avons sélectionné quatre enzymes :

- l'Elastase, qui clive la chaîne peptidique à l'extrémité carbonyle des acides aminés à petite chaîne latérale hydrophobe comme la glycine, la valine, la leucine, l'isoleucine et tout particulièrement l'alanine.
- la Carbopeptidase Y, qui dégrade la chaîne peptidique à partir de l'extrémité C-terminale sans grande spécificité d'acides aminés.
- l'enzyme Glu-C, une endopeptidase, qui hydrolyse la chaîne peptidique au niveau du carboxyl des acides glutamiques et aspartiques
- la Chymotrypsine, qui hydrolyse la chaîne peptidique à l'extrémité C-terminale des résidus phénylalanine, tyrosine et tryptophane (et dans une moindre mesure, des résidus leucine, méthionine et histidine) à condition que ces résidus ne soient pas suivis d'une méthionine ou d'une proline.

Nous avons réalisé les essais avec trois concentrations distinctes de chaque enzyme comme décrit dans le chapitre Matériel et Méthodes. Dans cet essai, le temps d'incubation est fixé à 30 minutes et l'hydrolyse a été réalisée à température ambiante. Les résultats obtenus sont présentés dans la figure 84A. Dans les cas de l'Elastase et la Carboxypeptidase Y nous n'avons observé qu'une très légère hydrolyse du complexe à l'exception de la plus forte concentration en Elastase (1:10). Par contre, nous avons pu voir une dégradation partielle du complexe avec les enzymes Glu-C et Chymotrypsine et le profil de dégradation évoluait en fonction des concentrations utilisées. Nous pouvions notamment voir la dégradation partielle dans un premier temps de Rsa1p (231-375), puis celle de Hit1p. Cette observation était en accord avec les données sur la flexibilité attendue de ces protéines. La protéine Snu13p qui est la seule protéine constituante des snoRNP C/D ne semblait pas être affectée par la présence de protéases, confortant l'idée qu'il s'agit d'une protéine nucléatrice stable. Ce qui est en accord avec les structures 3D établies pour Snu13p et pour ses homologues 15.5K et L7Ae ou encore Nhp2p. L'hydrolyse par l'enzyme Glu-C à la concentration 1:100 conduisait à un clivage quasi-total de la protéine Rsa1p (231-375) avec l'apparition de deux bandes de forte intensité,



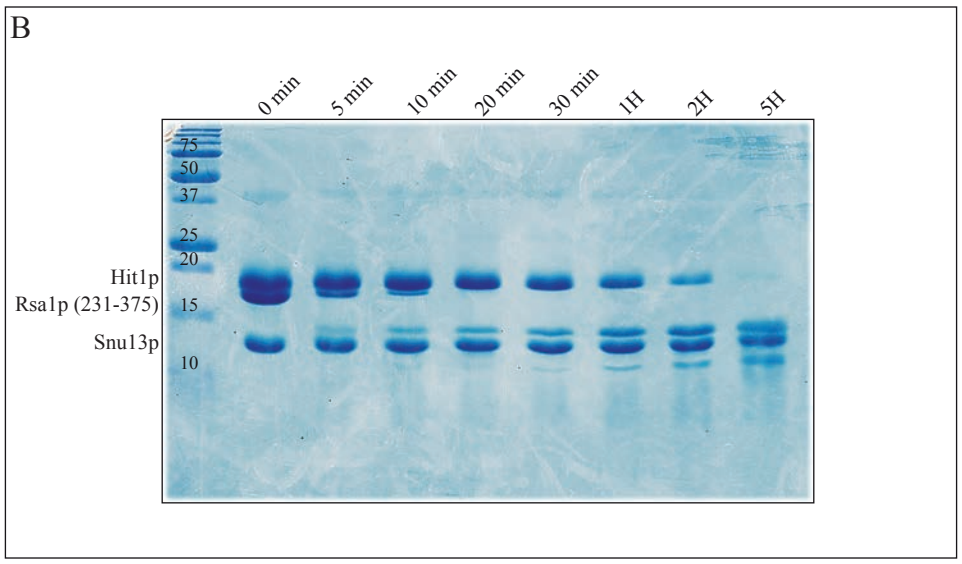


Figure 84. **Analyse des produits de la protéolyse ménagée du complexe Snu13p/Rsa1p (231-375)/Hit1p par SDS-PAGE à 15%.** (A) Choix de l'enzyme et de sa concentration. (B) Hydrolyse dans la condition Glu-C 1:100 pendant 0, 5, 10, 20, 30, 60, 120 et 300 minutes d'incubation.

ayant respectivement des poids moléculaires supérieur et inférieur à celui de la protéine Snu13p (Figure 84A). Entre les bandes correspondant à la protéine Hit1p et celle d'un poids moléculaire supérieur à la protéine Snu13p, nous n'observions aucune autre bande de protéine alors que c'était le cas pour la Chymotrypsine même à la plus faible concentration (1:1000). C'est pour cela que la condition Glu-C 1:100 a été retenue afin de réaliser une étude cinétique de cette protéolyse du complexe.

1.2.2 Choix du temps d'incubation

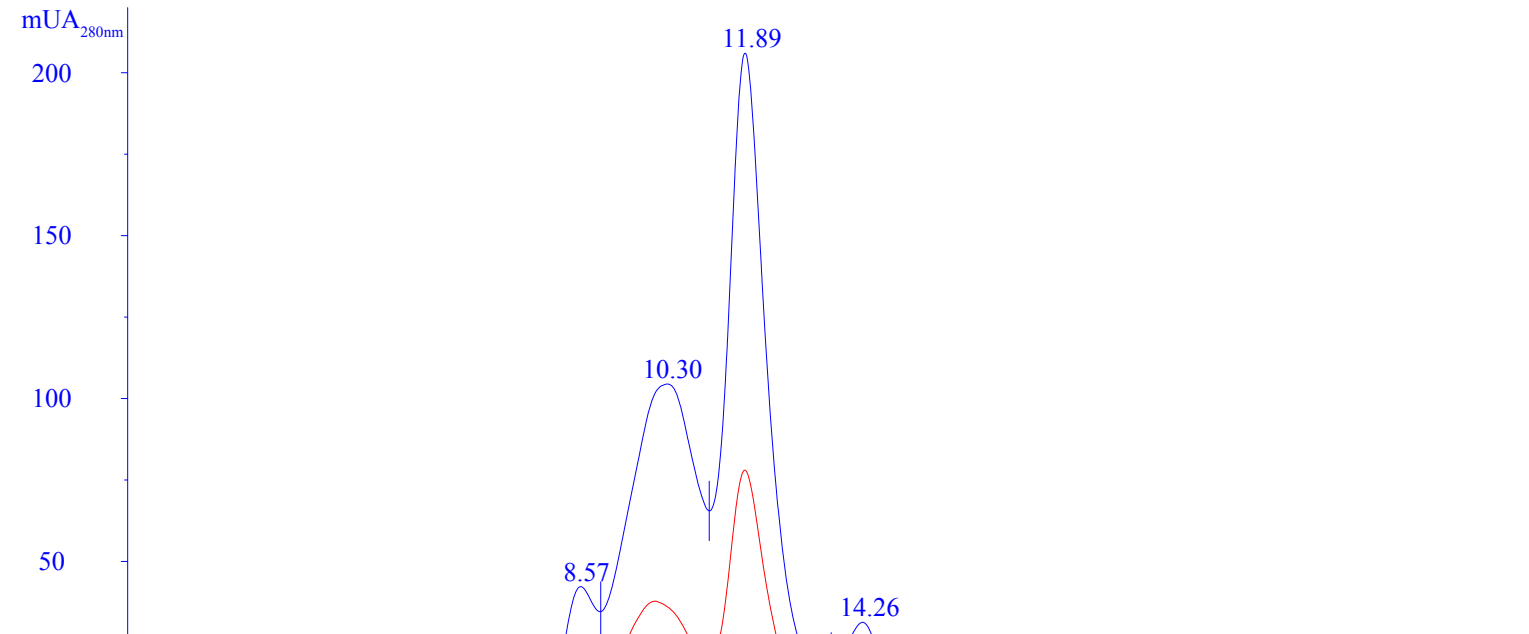
Comme nous venons de le voir, les conditions d'hydrolyse par la protéase Glu-C à la concentration 1:100 semblait conduire à la production de fragments bien définis et stables. Afin de déterminer le temps optimal d'incubation du complexe avec l'enzyme, nous avons étudié la cinétique de la réaction en prélevant des aliquots après 5, 10, 20, 30, 60, 120 et 300 minutes d'incubation et nous avons analysé les produits d'hydrolyse par SDS-PAGE (Figure 84B). Comme observé dans le panneau A, la protéine Rsa1p (231-375) est la première à être clivée dès 5 minutes de réaction. Une protéolyse de 20 minutes à température ambiante semblait être le meilleur compromis entre une dégradation totale de la protéine Rsa1p (231-375) et le maintien des protéines Snu13p et Hit1p.

1.2.3 Fractionnement et analyse par spectrométrie de masse des produits d'hydrolyse ménagée

Afin d'essayer d'identifier les produits de protéolyse obtenus, nous avons réalisé une réaction d'hydrolyse dans les conditions sélectionnées précédemment (Glu-C 1 :100) pendant 20 minutes à température ambiante, ceci en utilisant 1,5 mg de complexe Snu13p/Rsa1p (231-375)/Hit1p. Les produits de la réaction ont été fractionnés par chromatographie d'exclusion de taille (Figure 85), afin de séparer les sous-complexes présents. A l'issu de cette filtration sur gel (Superdex 75), les protéines contenues dans les fractions 33, 41 et 46 ont été analysées par spectrométrie de masse.

L'analyse de la fraction 33 a révélé la présence de plusieurs espèces dont les masses moléculaires sont données dans la figure 86. Afin d'identifier les espèces A à D, les masses moléculaires mesurées ont été comparées aux masses moléculaires théoriques des produits potentiels de digestion de chacune des trois protéines du complexe (Figure 86). Cette analyse a permis de conclure que :

- l'espèce A correspond au fragment 317-375 de Rsa1p.



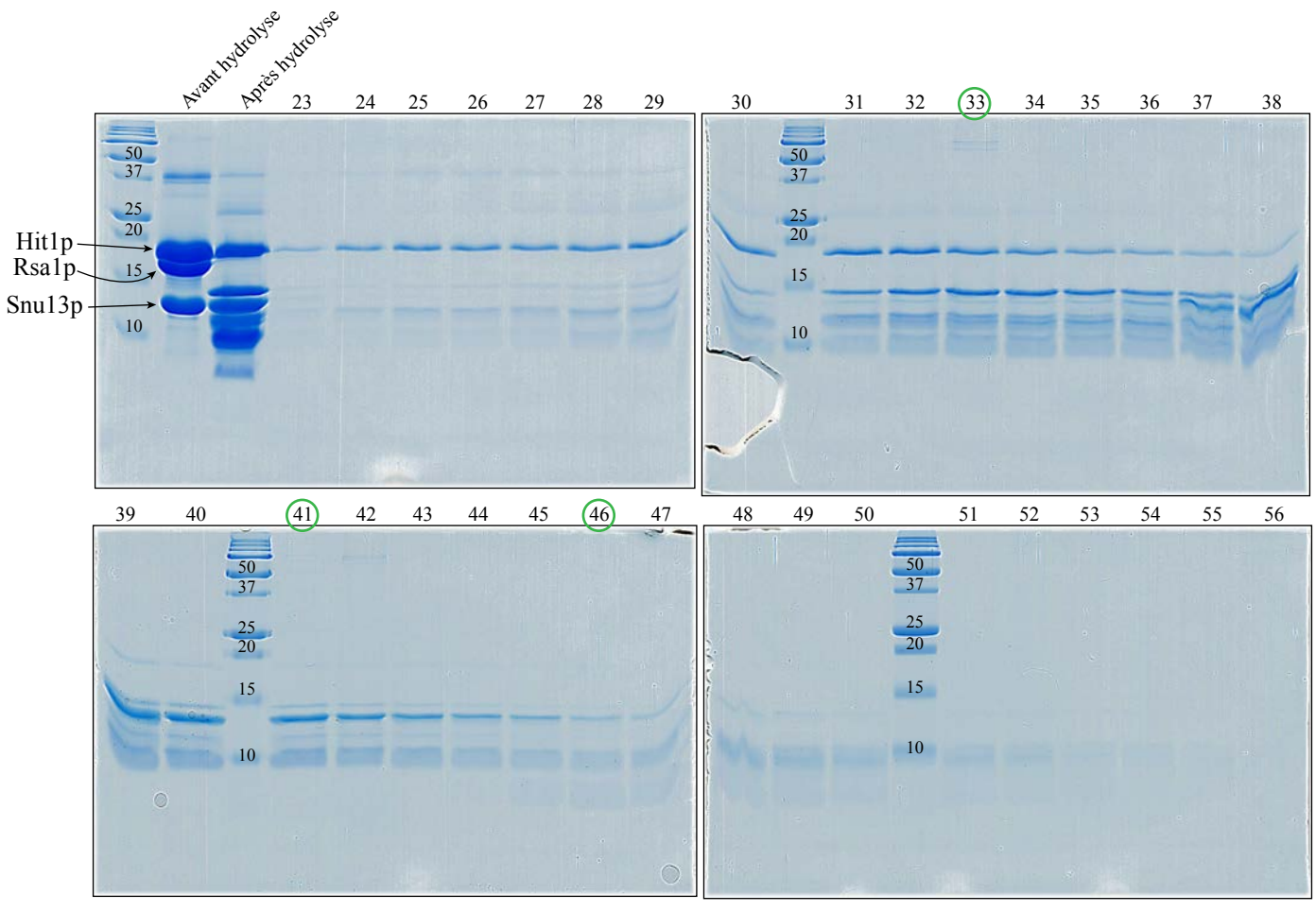
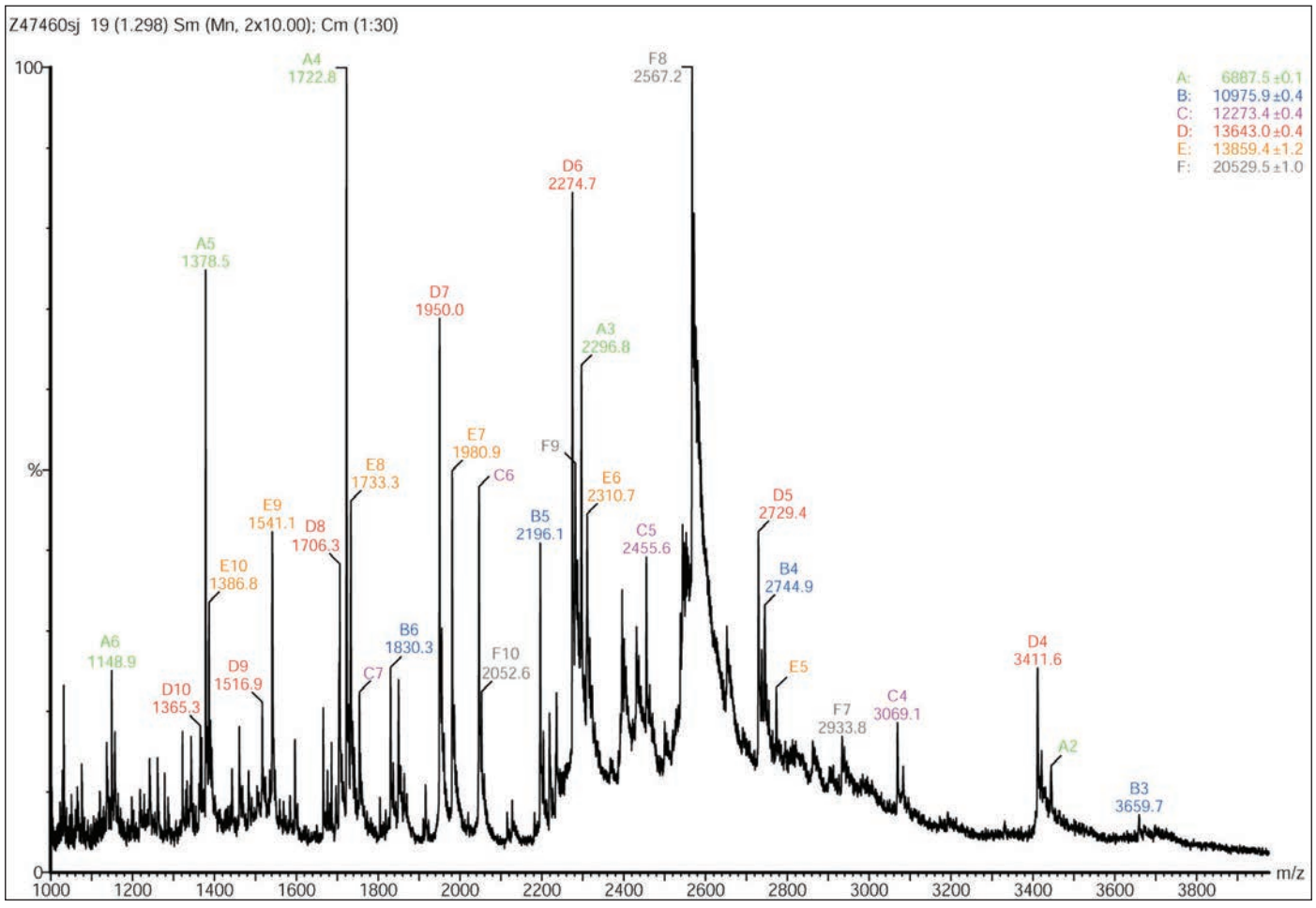


Figure 85. **Analyse de la protéolyse ménagée du complexe tripartite.** (A) Fractionnement par chromatographie d'exclusion de taille (Superdex 75) des produits de protéolyse ménagée du complexe Snu13p/Rsa1p (231-375)/Hit1p par l'enzyme Glu-C (1:100, 20 minutes d'incubation). Les courbes d'absorbance à 280nm et 260 nm sont représentées respectivement en bleu et en rouge avec la même échelle. (B) le contenu en protéines de chaque fraction a été analysé par SDS-PAGE. Les fractions 33, 41 et 46 (indiqués par un cercle vert) ont été analysées par spectrométrie de masse.

A



B

Masse mesurée	Séquences retenues	Séquence possible								
		Rsa1p			Hit1p			Snu13p		
		Bornes	Masse	Séquence	Bornes	Masse	Séquence	Bornes	Masse	Séquence
6887,5 (A)	Rsa1p (317-375)	277-335	6886,79	(K)SQPSIFKESRKEKQFIQSIQNQVQRGNPKIDLNLKLIQREFANENSQLLDFIRELGDVG(L)	11-72	6835,94	(I)CRGVDGKYKCPKCGVRYCSLKCYKDAAKHVHKESEQPRAGTEANVEVVNNDKIINSSLAMNK(T)	11-73	6886,03	(P)LADAALTQQILDVVQQAANLRQLKKGANEATKTLNRGISEFIIMAADCEPIEILLHLPLLCED(K)
		317-375	6887,56	(E)FANENSQLLDFIRELGDVGLLEYELSQQEKDVLFGSSEDNNKNHYKPNYKNRKPNLSRA(-)				9-71	6886,12	(A)FPLADAALTQQILDVVQQAANLRQLKKGANEATKTLNRGISEFIIMAADCEPIEILLHLPLLC(E)
10975,8 (B)	??	256-348	10973,5	(L)KISNNKQKHMQEMGIKEDELKSQPSIFKESRKEKQFIQSIQNQVQRGNPKIDLNLKLIQREFANENSQLLDFIRELGDVGLLEYELSQQEKDV(L)	29-126	10975,4	(C)SLKCYKDAAKHVHKESEQPRAGTEANVEVVNNDKIINSSLAMNKTLKTKAFDDIYQNSAELQELLKYNTVKFHLAKVYRILSSTVNDGSSGKMNSDLQ(K)			
					69-164	10976,6	(L)AMNKTLKTKAFDDIYQNSAELQELLKYNTVKFHLAKVYRILSSTVNDGSSGKMNSDLQKELAVNYLNTLRYGGIHYNEAIEEFCQILLDKLNAVKK(-)			
12273,4 (C)	??	241-342	12273,2	(E)DVKKWREERKKMWLLKISNNKQKHMQEMGIKEDELKSQPSIFKESRKEKQFIQSIQNQVQRGNPKIDLNLKLIQREFANENSQLLDFIRELGDVGLLEYELS(Q)	6-114	12273,2	(A)VKCGICRGVDGKYKCPKCGVRYCSLKCYKDAAKHVHKESEQPRAGTEANVEVVNNDKIINSSLAMNKTLKTKAFDDIYQNSAELQELLKYNTVKFHLAKVYRILSSTVN(D)	8-121	12273,4	(K)AFPLADAALTQQILDVVQQAANLRQLKKGANEATKTLNRGISEFIIMAADCEPIEILLHLPLLCEDKNVPYVFVPSRVALGRACGVSRPVIAASITTNDASAIKTQIYAVKDKI(E)
13643 (D)	Hit1p (44-164)	235-348	13643,6	(I)ALITDEDVKKWREERKKMWLLKISNNKQKHMQEMGIKEDELKSQPSIFKESRKEKQFIQSIQNQVQRGNPKIDLNLKLIQREFANENSQLLDFIRELGDVGLLEYELSQQEKDV(L)	44-164	13643,5	(E)SEQPRAGTEANVEVVNNDKIINSSLAMNKTLKTKAFDDIYQNSAELQELLKYNTVKFHLAKVYRILSSTVNDGSSGKMNSDLQKELAVNYLNTLRYGGIHYNEAIEEFCQILLDKLNAVKK(-)			

Figure 86. **Analyse par spectrométrie de masse des protéines contenues dans la fraction 33 du fractionnement des produits d'hydrolyse ménagée du complexe Snu13p/Rsa1p (231-375)/Hit1p par l'enzyme Glu-C** (1:100, 20 min). (A) Spectre de l'analyse par spectrométrie de masse en conditions non dénaturantes de la fraction 33 en tampon AcNH4 300mM pH 7,5. (B) Masses moléculaires théoriques des produits de digestion possible de chacune des trois protéines du complexe. Les séquences correspondant à un produit issue de 2, 1 et aucune coupures par la Glu-C sont notées respectivement en verte, orange et rouge. Pour chaque masse mesurée, il est proposé une séquence d'un produit d'hydrolyse lorsque cela était possible.

- les espèces B et C ne pouvaient pas clairement être identifiées car les séquences possibles ne correspondent pas à des stricts produits d'hydrolyse par Glu-C.
- l'espèce D correspondait au fragment C-terminal 44-164 de la protéine Hit1p.
- l'espèce E (13859.4±1.2 Da) correspondait à la protéine Snu13p non hydrolysée (13860.2 Da),
- l'espèce F (20529.5±1.0 Da) correspondait au complexe formé par les espèces A et D (6887,5 + 13643,0 = 20530,5).

L'analyse de la fraction 41 révélait la présence de deux espèces dont les masses sont données dans la figure 87. En accord avec le résultat du gel SDS PAGE (Figure 85), la protéine majoritaire dans cette fraction est l'espèce B (13860.7±0.3 Da) correspondant à la protéine Snu13p non digérée (13860.2 Da). L'espèce A minoritaire n'a pu être identifiée en raison du grand nombre de séquences possibles pour cette petite masse et du fait qu'aucune des séquences potentielles ne possèdent deux extrémités qui pouvait être issues de coupures par Glu-C.

L'analyse de la fraction 46 n'a pas permis de détecter d'espèce par manque de signal.

L'analyse de la fraction 33 nous a donc permis d'identifier deux produits majeurs d'hydrolyse par Glu-C, le fragment 317-375 de Rsa1p et le fragment C-terminal 44-164 de la protéine Hit1p. Aucun complexe hétérotrimérique n'a été observé dans les analyses par spectrométrie de masse en condition non dénaturante bien que la protéine Snu13p ait été présente avec les fragments de Rsa1p et Hit1p. Nous en avons conclu que le complexe Rsa1p (317-375)/Hit1p (44-164) n'était sans doute pas capable de s'associer à la protéine Snu13p. Ce résultat était en accord avec notre absence de détection d'un complexe entre Snu13p et Hit1p lors des expériences de co-expression (quatrième série, pistes 111 et 112) et que le site de fixation de la protéine Snu13p sur Rsa1p se situe au niveau du domaine 231-266 ((Boulon *et al.*, 2008) et manuscrit n° 2).

Cette étude apportait une information importante : la formation d'un complexe stable entre les fragments (317-375) de Rsa1p et (44-164) de Hit1p ce qui n'était pas connu. Nous affinions ainsi la définition des domaines d'interaction de ces deux protéines. De plus, ce complexe étant le produit d'une protéolyse, on pouvait supposer qu'il présentait peu de chaînes flexibles et peu structurés. Nous ne pouvons pas à ce stade éliminer l'hypothèse que la présence de Rsa1p (231-374), Hit1p (1-43) et/ou Snu13p soient nécessaire à la formation du complexe stable entre Rsa1p (317-375)/Hit1p (44-164).

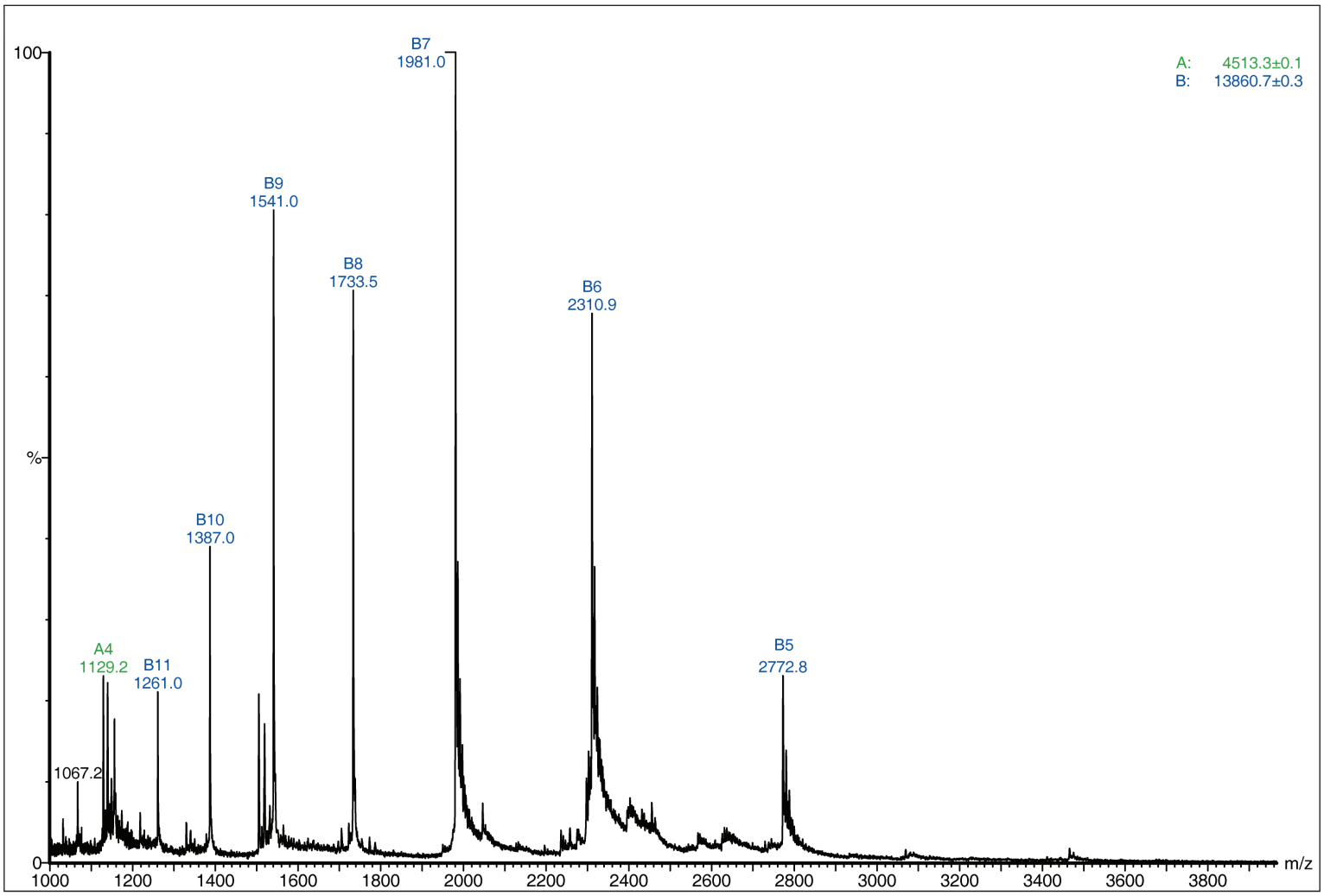


Figure 87. **Analyse par spectrométrie de masse de la fraction 41 obtenue par chromatographie d'exclusion de taille après protéolyse ménagée du complexe Snu13p/Rsa1p (231-375)/Hit1p par l'enzyme Glu-C** (1:100, 20 min). Spectre de l'analyse par spectrométrie de masse en conditions non dénaturantes dans un tampn $AcNH_4$ 300mM, pH 7,5).

Il était intéressant de noter que le domaine 1-43 de la protéine Hit1p, excisé lors de la protéolyse ménagée du complexe Snu13p/Rsa1p (231-375)/Hit1p correspondait exactement au domaine à doigt à Zinc bien caractérisé de cette protéine qui a donné lieu à la famille du doigt à Zinc « zf-Hit ».

Néanmoins, sur la base de ces résultats, nous avons procédé au clonage des régions codant ces deux fragments dans les vecteurs de co-expression, afin de produire le complexe Rsa1p (317-375)/Hit1p (44-164) en grande quantité pour une étude structurale.

1.3 La production du fragment de Hit1p (44-164)

Dans un premier temps, nous avons exprimé et purifié individuellement chaque nouveau fragment défini sur la base des résultats de spectrométrie de masse. Le fragment Rsa1p (317-375) ne faisait que 59 résidus, ce qui rendait sa purification difficile. Par contre, le fragment Hit1p (44-164) plus long était bien exprimé chez *E. coli* et facilement purifiable. Nous avons donc réalisé une purification du domaine 44-164 de la protéine Hit1p et obtenu un échantillon de haute pureté à une concentration de 8,45 mg/ml. Cet échantillon a alors été dialysé dans le tampon phosphate R.M.N. à différentes concentrations salines (200, 150, 100, 50 et 0 mM NaCl), puis analysé par D.L.S. afin de déterminer les conditions les plus favorables pour son analyse par R.M.N.

1.3.1 Analyse par DLS du fragment Hit1p (44-164)

L'approche par D.L.S. nous permettait d'obtenir des informations sur l'homogénéité et d'approcher le diamètre moyen du fragment Hit1p (44-164) analysé en fonction des conditions salines. Afin de contrôler si le changement de tampon (et/ou salinité) n'avait pas eu d'effet déstructurant sur la conformation de la protéine, nous avions également analysé un échantillon dans le tampon de la chromatographie d'exclusion de taille, dernière étape de la purification. Il s'agissait de l'échantillon dans le tampon Tris-HCl 10 mM, pH 8 ; NaCl 150 mM. Les autres échantillons correspondaient au tampon R.M.N. soit un tampon phosphate 10 mM, pH 6,6 avec différentes concentrations salines pour une même concentration de protéine (8,45 mg/ml). Les résultats obtenus sont présentés dans la figure 88. Nous constations que le sel n'influence que très peu l'homogénéité de l'échantillon comme en témoigne la faible variation de l'indice PDI. Par contre nous avons aussi constaté qu'après congélation à -80°C, ces mêmes échantillons présentaient une valeur d'indice PDI doublée indiquant que ces échantillons supportaient assez mal la congélation. Nous constations par ailleurs que l'indice

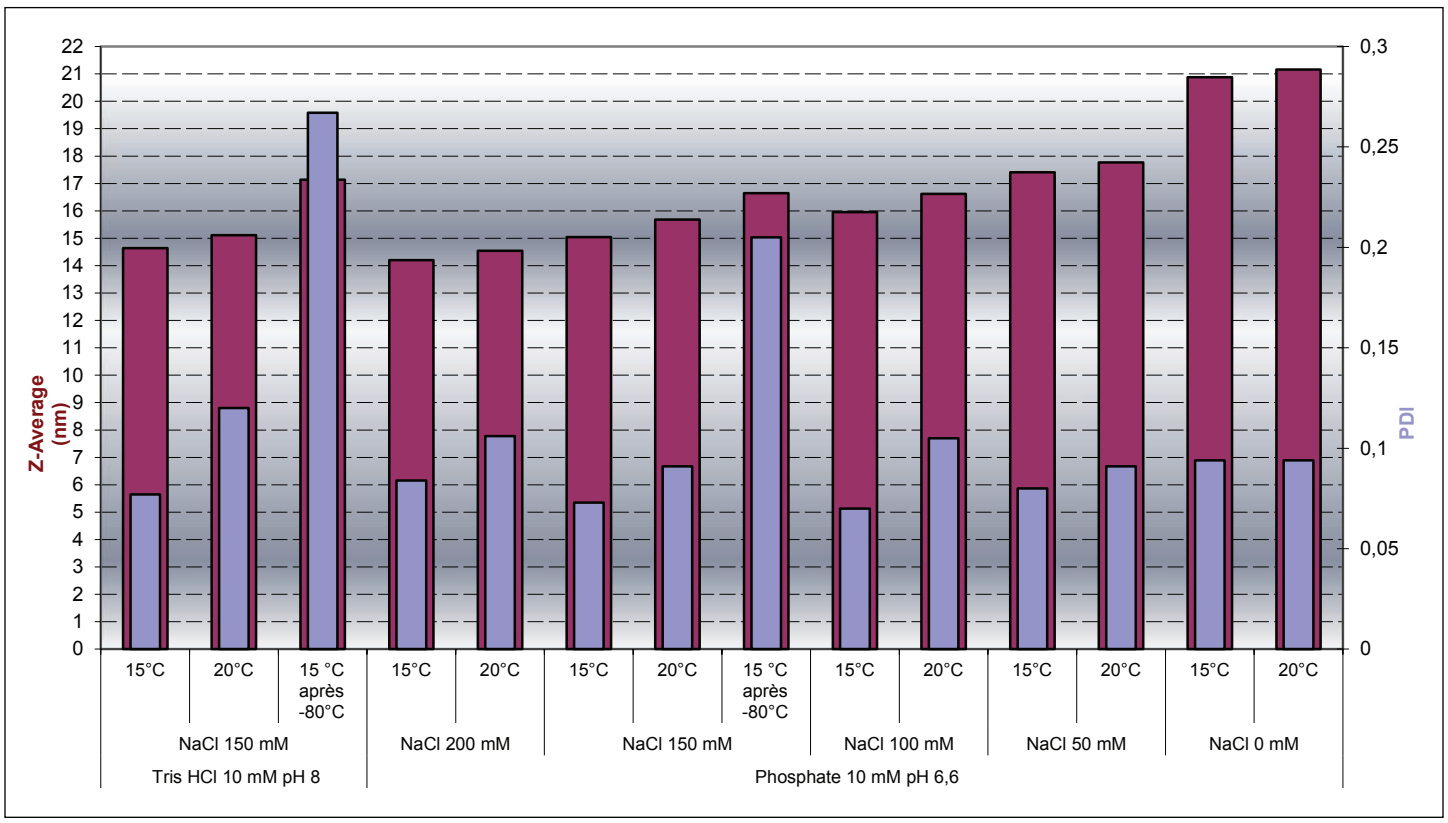


Figure 88. **Analyse par diffusion dynamique de la lumière de la protéine Hit1p (44-164).** La concentration des protéines des échantillons utilisés était de 8,45 mg/ml. Ils avaient été dialysés contre des tampons Phosphate 10 mM, pH 6,6 contenant différentes concentrations en sel (200, 150, 100, 50 et 0 mM NaCl) et ont été analysés à deux températures (15°C et 20°C). Deux échantillons ont été congelés à l'azote liquide, stockés à -80°C, puis décongélés pour tester l'effet de la congélation sur la conformation de la protéine (échantillon "15°C après -80°C").

du rayon hydrodynamique de la protéine était élevé pour une protéine de cette taille (environ 15 nm). Cela suggérait une multimérisation de la protéine, une agrégation ou qu'elle soit complètement ou partiellement dépliée. Pour choisir entre ces hypothèses, nous avons réalisé une analyse par S.L.S.

1.3.2 Analyse par S.L.S.

L'approche S.L.S. (*« Static Light Scattering »*) ou diffusion statique de la lumière est basée sur la diffusion de la lumière par les molécules. Elle permet de calculer la masse moléculaire moyenne d'une macromolécule. Dans notre cas, l'analyse de la protéine Hit1p (44-164) a clairement montré une multimérisation de la protéine avec un profil de chromatographie d'exclusion de taille, d'index de réfraction et de R.A.L.S. (*« Right Angle Light Scattering »*) multimodale (Figure 89). De plus, même si la présence de pics multiples perturbait le calcul de la masse moléculaire, la valeur très élevée que nous avions obtenue (plus de 300 000 Da) montrait aussi que la protéine multimérisait ou s'agrégeait.

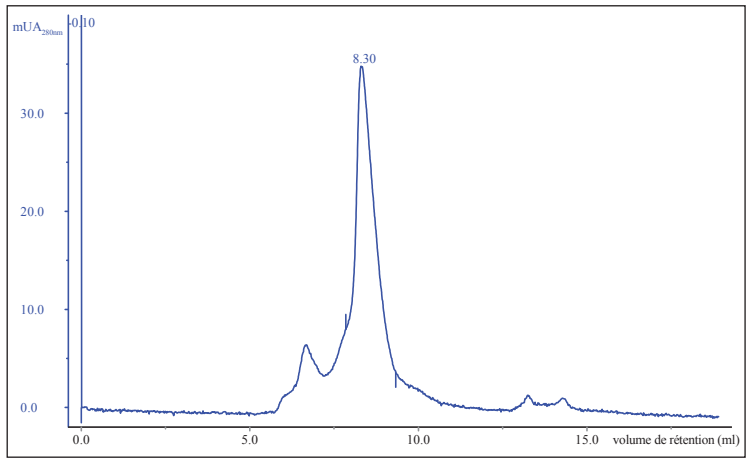
1.3.3 Analyse par R.M.N.

En parallèle de ces études, un échantillon du fragment Hit1p (44-164) a été analysé par R.M.N.

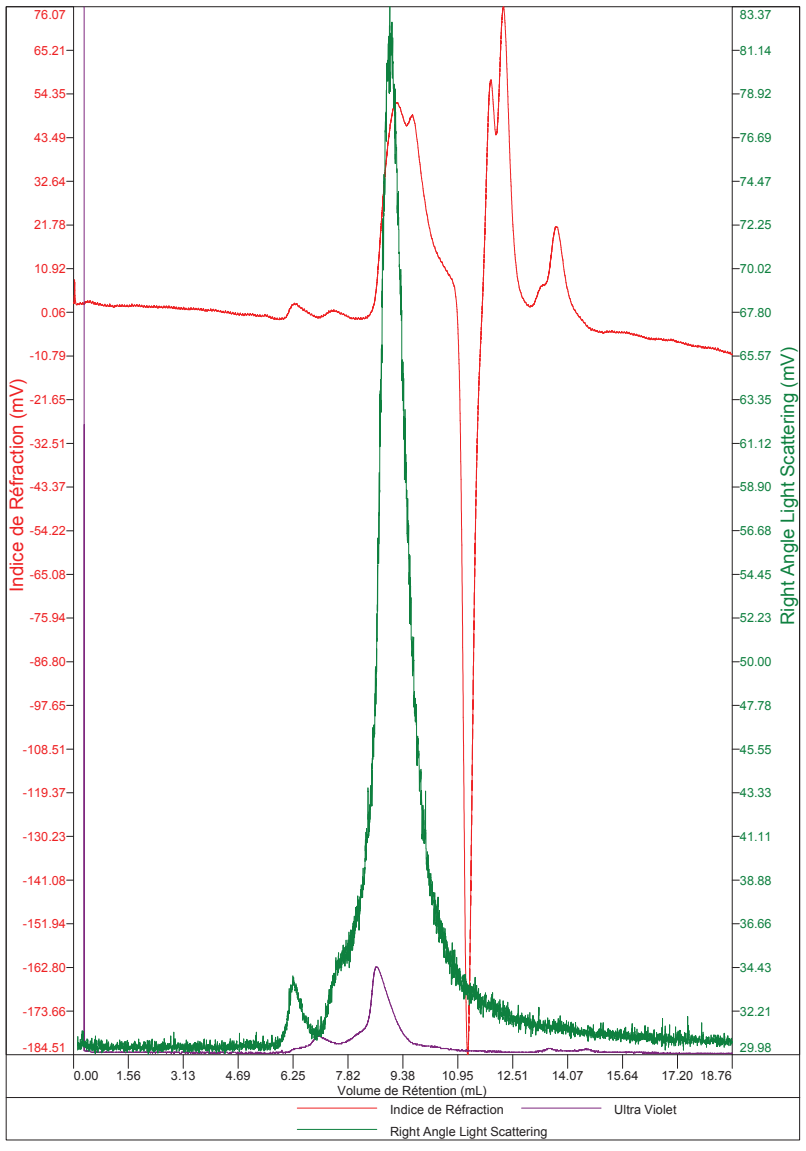
Nous avons réalisé une expérience 2D HSQC (*« Heteronuclear single quantum coherence »*) ^{1}H-^{15}N. Ce type d'expérience permet de corréler chaque azote avec le ou les hydrogènes qui lui sont liés covalemment. On peut ainsi établir une empreinte des groupements NH (i.e. amide de la chaîne principale et amine primaire des chaînes latérales) d'une protéine, à l'exception des prolines (dont l'azote ne porte aucun hydrogène) et du résidu N-terminal (l'hydrogène de la fonction amine étant labile et s'échangeant avec l'eau). Schématiquement, un spectre HSQC ^{1}H-^{15}N fournit une tâche de corrélation ou « pic » par résidu et reflète l'état de repliement d'une protéine. Une protéine structurée est caractérisée par des taches de corrélation bien dispersées.

Le spectre HSQC ^{15}N-^{1}H obtenu pour le fragment Hit1p (44-164) montrait clairement l'agrégation de la protéine ou tout du moins sa multimérisation (Figure 90). En effet, le nombre de tâches de corrélation visibles devrait être de 123 pour une protéine de 125 résidus (présence de deux prolines) et leur dispersion spectrale se situait dans une gamme allant de 7,7 à 8,5 ppm ce qui est caractéristique d'une conformation étendue.

A



B



C

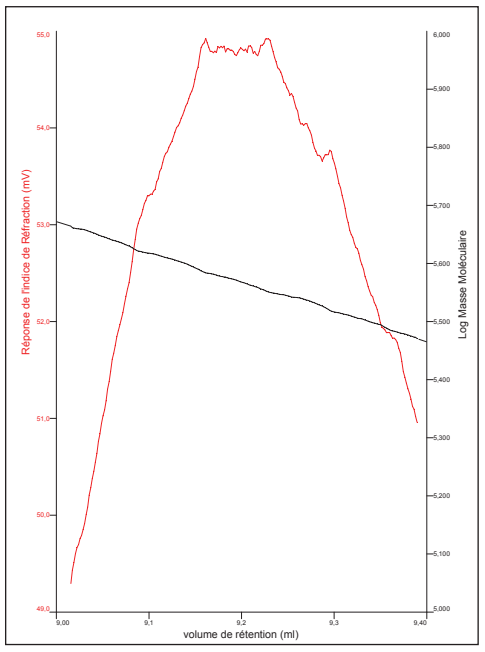


Figure 89. **Analyse par diffusion statique de la lumière de la protéine Hit1p (44-164).** La concentration en protéine de l'échantillon était de 100 µM dans un tampon Hepes 10 mM, pH 7,5 ; NaCl 150 mM. L'analyse à été réalisée à température ambiante. (A) Profil de la chromatographie d'exclusion de taille (SHODEX KW 8025) précédent l'analyse par S.L.S. (B) Superposition des profils d'absorbance des ultra-violets (profil SHODEX) en violet, de l'indice de réfraction en rouge et du R.A.L.S. en vert en fonction du volume de rétention. (C) Superposition de l'indice de réfraction et du logarithme de la masse moléculaire en fonction du volume de rétention.

A

Hit1p (44-164) :

```
   44     50        60        70        80        90       100       110       120
GPHMSEQPRAGTEANVEVVNNDKIINSSLAMNKTLKTKAFDDIYQNSAELQELLKYNTVKFHLAKVYRILSSTVNDGSSGK

     130       140       150       160
MNSDLQKELAVNYLNTLRYGGIHYNEAIEEFCQILLDKLNAVKK
```

B

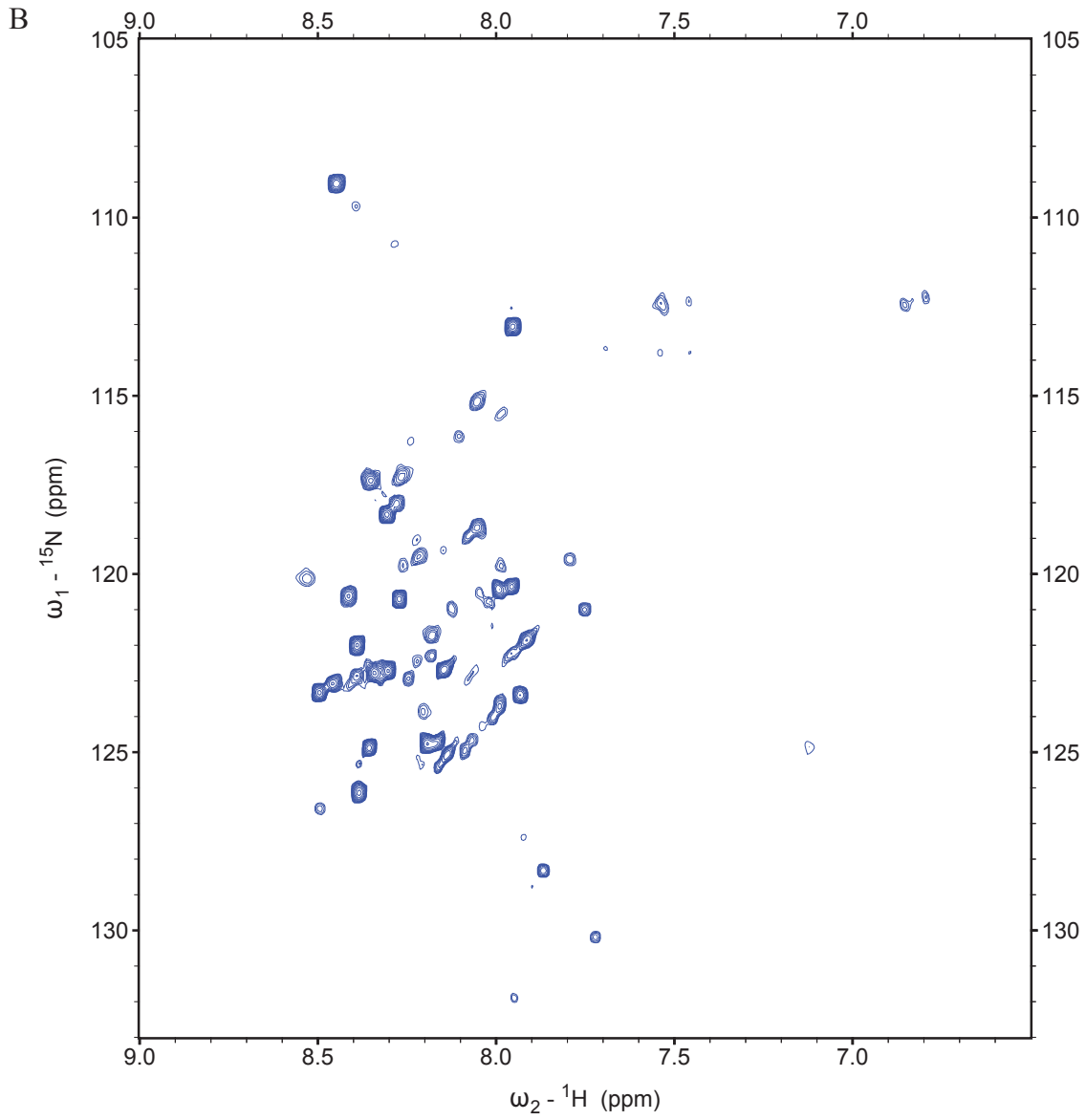


Figure 90. **Etude par R.M.N. de la protéine Hit1p (44-164).** (A) Séquence du fragment Hit1p (44-164) étudié. Les acides aminés en rouge sont rajoutés par la construction réalisée dans le vecteur pnEA-3CH. (B) Spectre ^{15}N-^{1}H TROSY (*"Transverse Relaxation Optimized Spectroscopy"*)-HSQC de la protéine Hit1p (44-164) à 288°K. La protéine a été utilisée à une concentration de 9,8 mg/ml (0,7 mM) dans du tampon phosphate 10 mM, pH 6,5; NaCl 150 mM.

Les analyses par D.L.S., S.L.S. et R.M.N. de la protéine Hit1p (44-164) produite seule montraient clairement qu'elle s'agrégeait dans les conditions utilisées même si elle gardait ses capacités de fixation à la résine TALON™ et de clivage par la PreScission ainsi que de solubilité. Ceci peut donc confirmer des résultats d'expériences de double hybride réalisées par B. Rothé (communication personnelle) qui montraient que la protéine Hit1p est capable d'interagir avec elle-même.

1.4 Le complexe Rsa1p (317-375)/Hit1p (44-164)

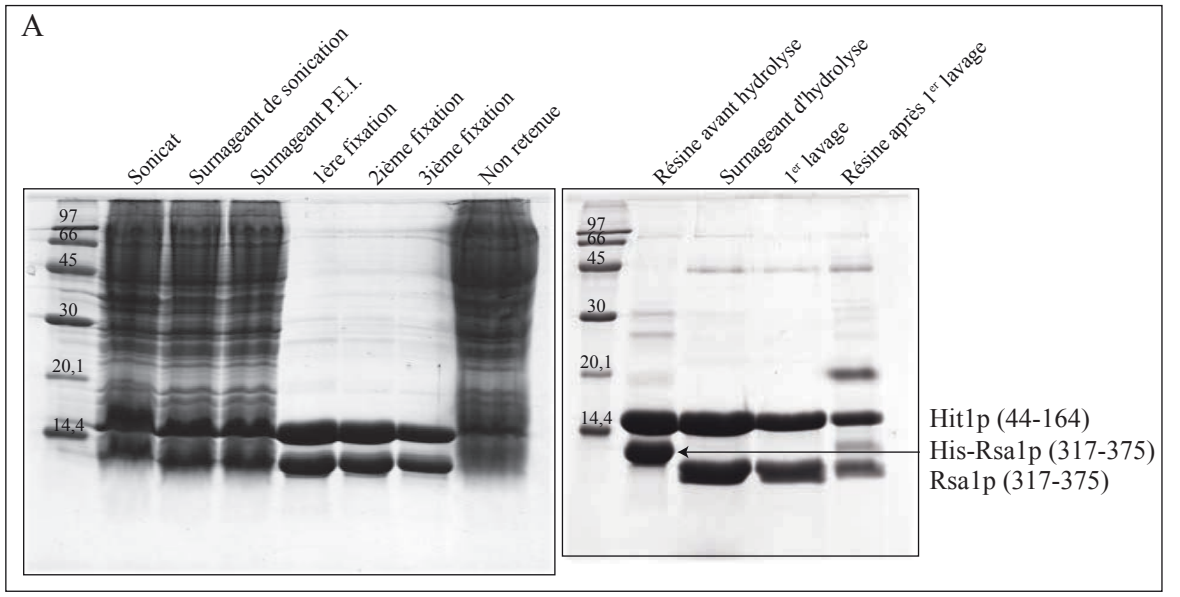
Aussi bien Rsa1p (317-375) que Hit1p (44-164) posaient donc des problèmes de purification et d'agrégation lorsqu'elles sont produites individuellement. La co-expression des protéines Rsa1p, Hit1p et Snu13p suivie d'une protéolyse s'étant avérée être un moyen de produire le complexe Rsa1p (317-375)/Hit1p (44-164) sous forme soluble, nous avons testé si la co-expression de ces deux fragments de protéines permettait de produire le complexe qu'ils forment sous forme soluble. Ce test s'est avéré positif, la co-expression de deux fragments suivie du fractionnement par chromatographie d'exclusion de taille du matériel après hydrolyse par la PreScission protéase directement sur la résine TALON™ nous a permis de montrer que les domaines 317-375 de Rsa1p et 44-164 de Hit1p s'associent de façon stable et stœchiométrique en absence de la protéine Snu13p et des domaines 231-316 de Rsa1p et 1-43 de Hit1p (Figure 91). Le doigt à Zinc de la protéine Hit1p ne joue donc pas un rôle dans cette interaction.

1.4.1 Analyse par DLS

Afin d'identifier la concentration saline optimum pour étudier le complexe Rsa1p (317-375)/Hit1p (44-164) par diffraction des rayons X, nous avons réalisé une étude par D.L.S. Les résultats obtenus (Figure 92) révèlent un rayon hydrodynamique du complexe entre 6 et 7 nm, ce qui semble en adéquation avec le complexe de 21 kDa formé par Rsa1p (317-375) et Hit1p (44-164) contrairement à celui que nous avions observé lorsque Hit1p était exprimé seul (paragraphe 1.3.1). De plus en présence de 100 mM NaCl, l'hétérogénéité du complexe était inférieure à 10% (Figure 92), ce qui est favorable pour une cristallisation du complexe.

1.4.2 Analyse par S.L.S.

Malgré de très bonnes données de D.L.S., la faible taille des protéines, notamment celle de Rsa1p (317-375) ne permettait pas de déterminer avec certitude la stœchiométrie du



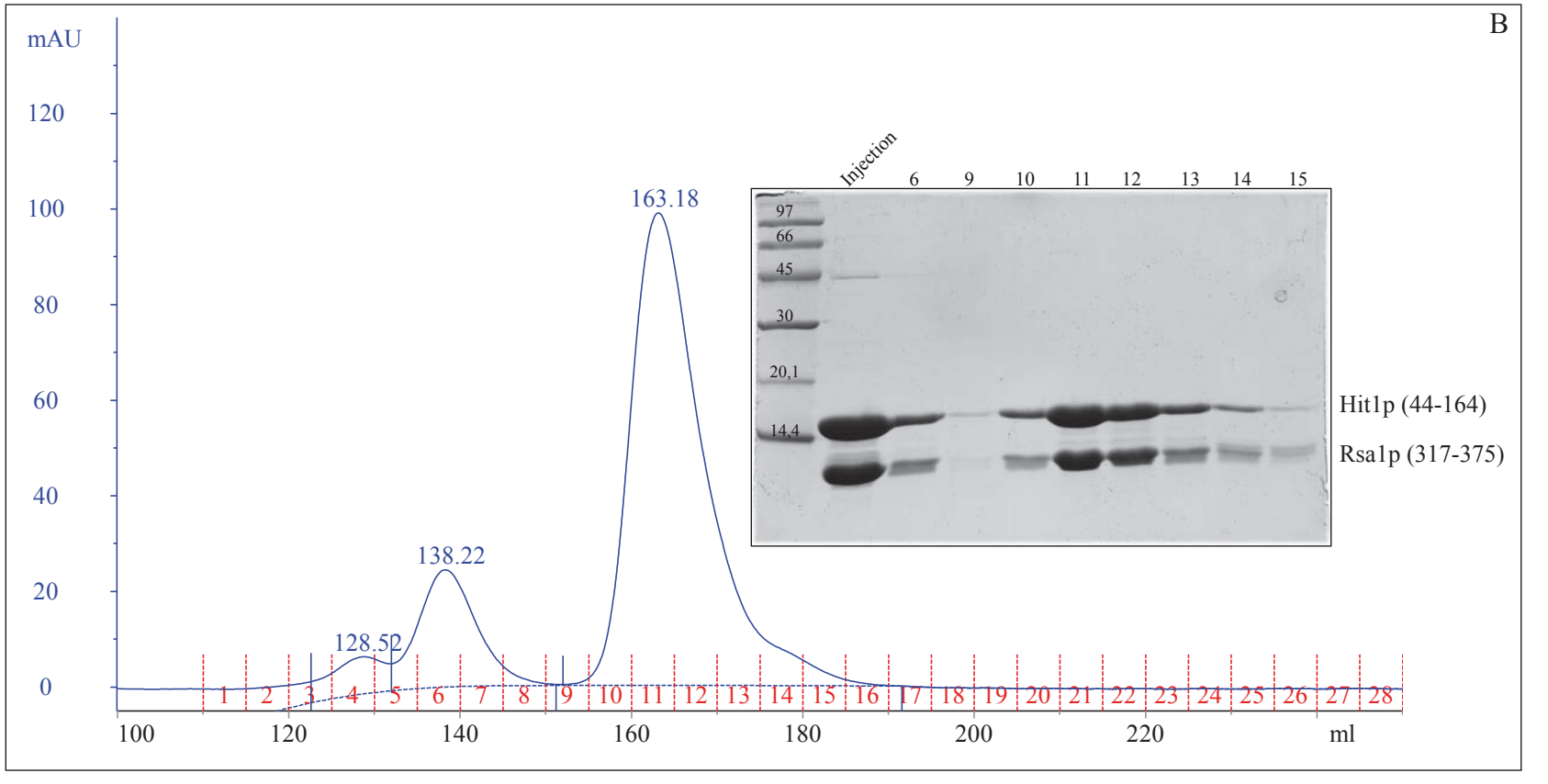


Figure 91. **Purification préparative du complexe His-Rsa1p (317-375)/Hit1p (44-164) doublement marqué ^{15}N ^{1}H.** (A) Analyse par SDS-PAGE des protéines contenues dans les fractions des premières étapes de purification. (B) Fractionnement par chromatographie d'exclusion de taille du complexe après hydrolyse par la PreScission protéase et analyse par SDS-PAGE des fractions correspondantes aux différents pics.

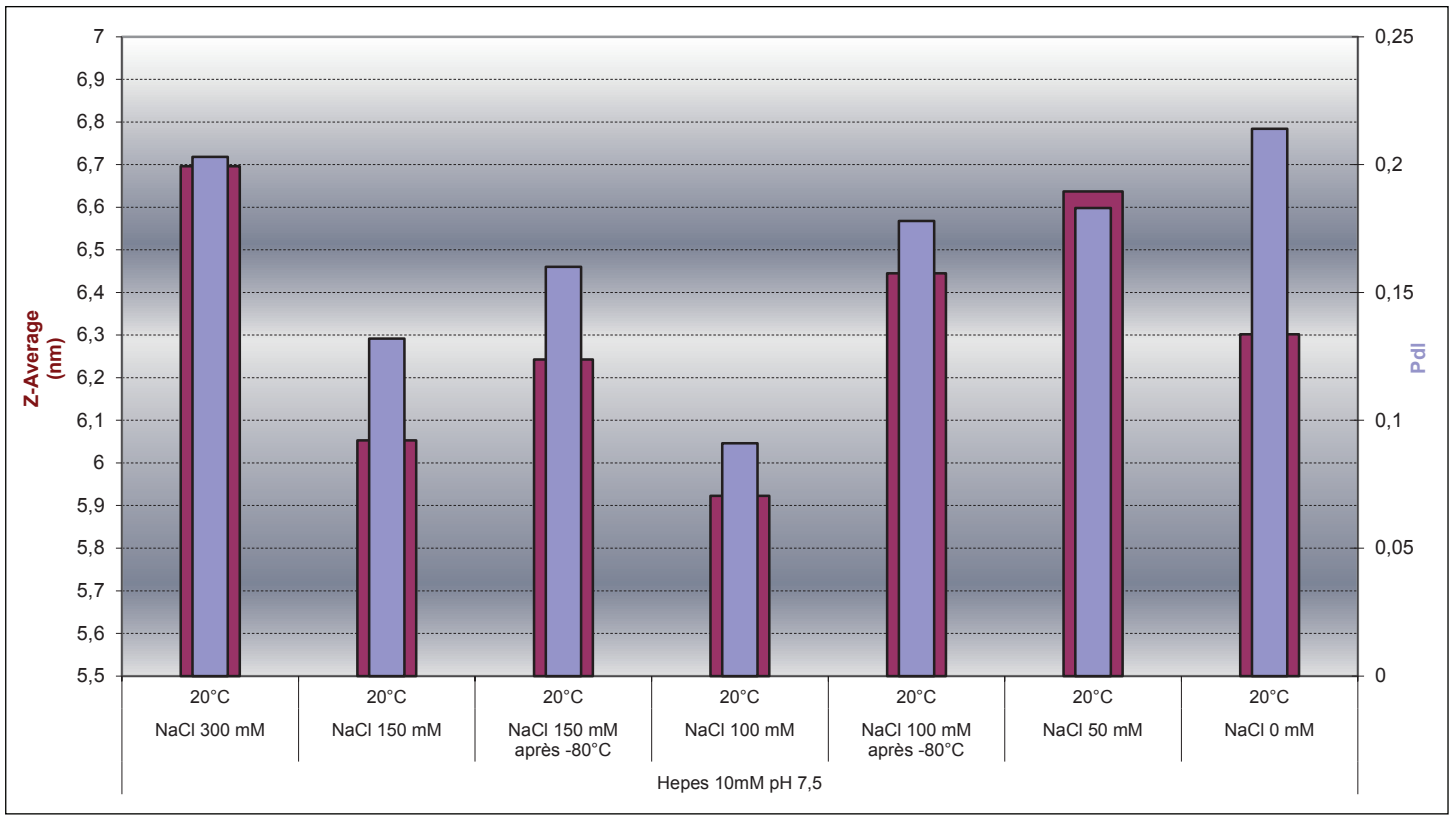


Figure 92. **Analyse par diffusion dynamique de la lumière du complexe Hit1p (44-164)/Rsa1p (317-375).** La concentration du complexe était de 12 mg/ml et avait été dialysés contre un tampon Hepes 10 mM, pH 7,5 à différentes concentrations salines (300, 150, 100, 50 et 0 mM NaCl) avant son analyse à 20°C. Deux échantillons avaient été congélés dans l'azote liquide, stockés à -80°C, puis décongélés pour analyser l'effet de la congélation sur la conformation des protéines (échantillon "après -80°C").

complexe. Pour la déterminer, nous avons procédé à une analyse S.L.S. du complexe (Figure 93). Un seul pic était détecté après la chromatogrphie d'exclusion de taille (Figure 93). Ce qui confirmait l'homogénéité de la population. De plus, les très bonnes données de réfraction et de R.A.L.S. permettaient d'estimer avec précision, la masse moléculaire moyenne du complexe à 19 kDa ce qui correspondait bien à un complexe Rsa1p (317-375)/Hit1p (44-164) avec une stœchiométrie de 1:1.

Sur la base de l'ensemble de ces analyses, nous avions décidé de débuter l'analyse du complexe Rsa1p (317-375)/Hit1p (44-164) en tentant à la fois une analyse par radiocristallographie et une analyse par R.M.N..

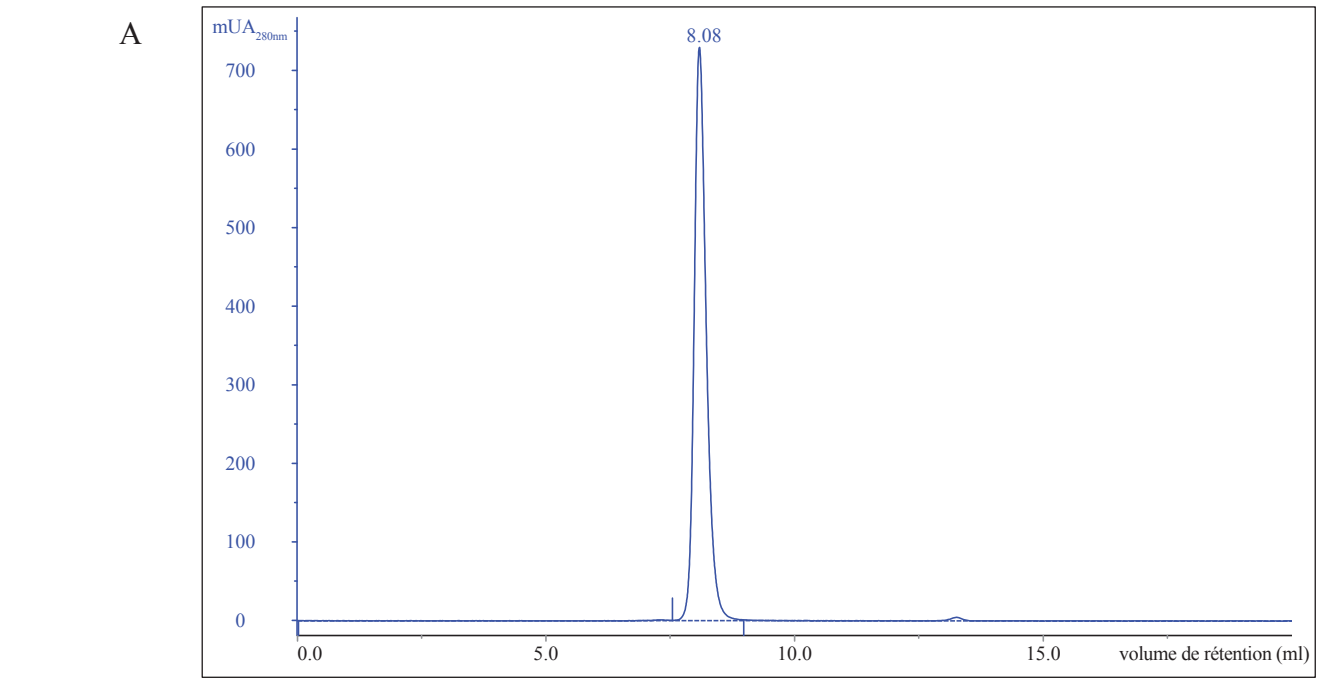
1.4.3 Tests de cristallisation

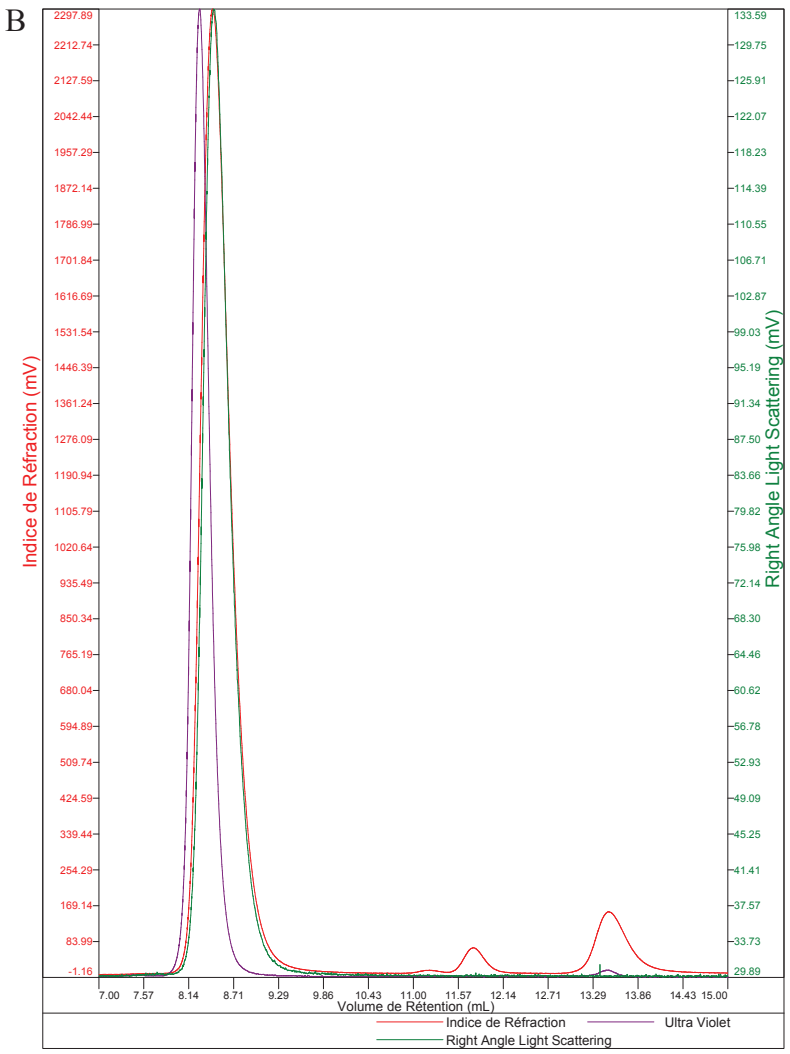
L'utilisation du robot MOSQUITO® nous avait permis de tester plus de milles conditions de cristallisation par échantillon. Nous avions utilisé des échantillons à 20, 60 et 185 mg/ml soit respectivement à des concentrations de 1, 3 et 9 mM pour réaliser les tests de cristallisation. Mais malgré tous nos efforts (tests de température, absence de sel dans l'échantillon) aucun cristal de protéine n'a pu être obtenu à ce jour.

1.4.4 Etude par R.M.N.

La masse moléculaire calculée du complexe étant de 19 kDa (185 résidus), une étude par R.M.N. était possible. Pour cela, nous avons donc réalisé une production et une purification du complexe Rsa1p (317-375)/Hit1p (44-164) doublement marqué $^{15}N/^{13}C$. Nous avons ainsi obtenu un échantillon de haute pureté à 0,97 mM. La production était donc bonne malgré l'utilisation d'un milieu minimum pour le marquage isotopique ^{15}N et ^{13}C. Afin de tester l'influence des sels sur le repliement des protéines, des spectres ^{15}N-^{1}H TROSY-HSQC ont étés enregistrés en présence de 150 mM et sans NaCl (Figure 94).

Les spectres obtenus en présence et en absence de NaCl étaient superposable indiquant que le complexe était peu sensible aux variations de concentrations en sel. Nous avons donc réalisé l'ensemble des expériences de R.M.N. à une concentration saline de 150 mM, concentration minimale pour réaliser la chromatographie d'exclusion de taille, ce qui permettait d'éviter une dialyse avant l'analyse par R.M.N. A son arrivé au laboratoire, Marc Quinternet (IR, CNRS) a repris l'étude par R.M.N. de ce complexe car le spectre ^{15}N-^{1}H HSQC que nous avons obtenu semblait satisfaisant.





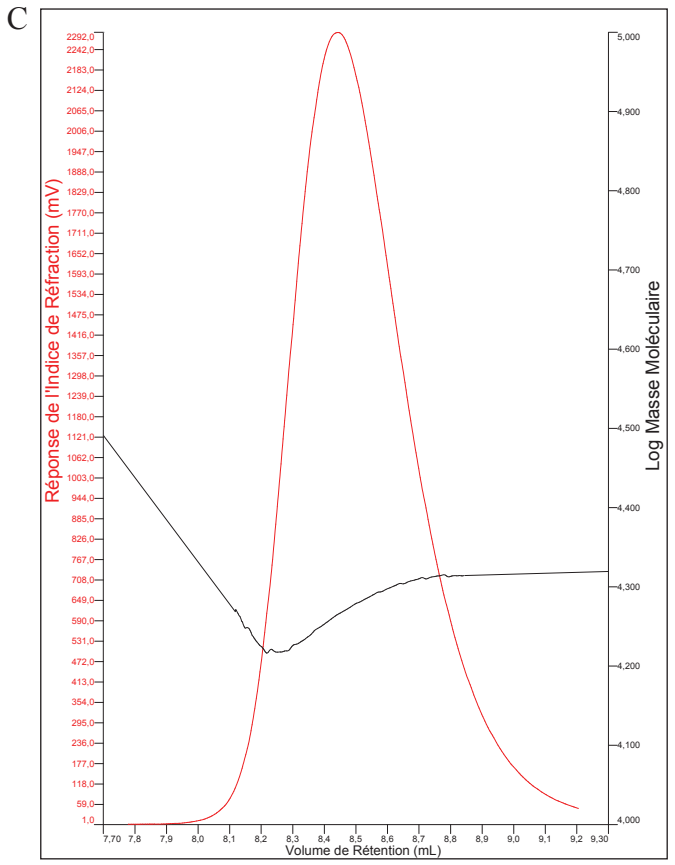


Figure 93. A**nalyse par diffusion statique de la lumière du complexe Hit1p (44-164)/Rsa1p (317-375).** Le complexe était à une concentration de 261 µM dans le tampon Hepes 10 mM, pH 7,5 ; NaCl 150 mM et analysé à température ambiante. (A) Profil de la chromatographie d'exclusion de taille (SHODEX KW 8025) précédent l'analyse par S.L.S. (B) Superposition des profils d'absorbance aux ultra-violets (profil SHODEX) en violet, de l'indice de réfraction en rouge et du R.A.L.S. en vert en fonction du volume de rétention. (C) Superposition de l'indice de réfraction et du logarithme de la masse moléculaire en fonction du volume de rétention.

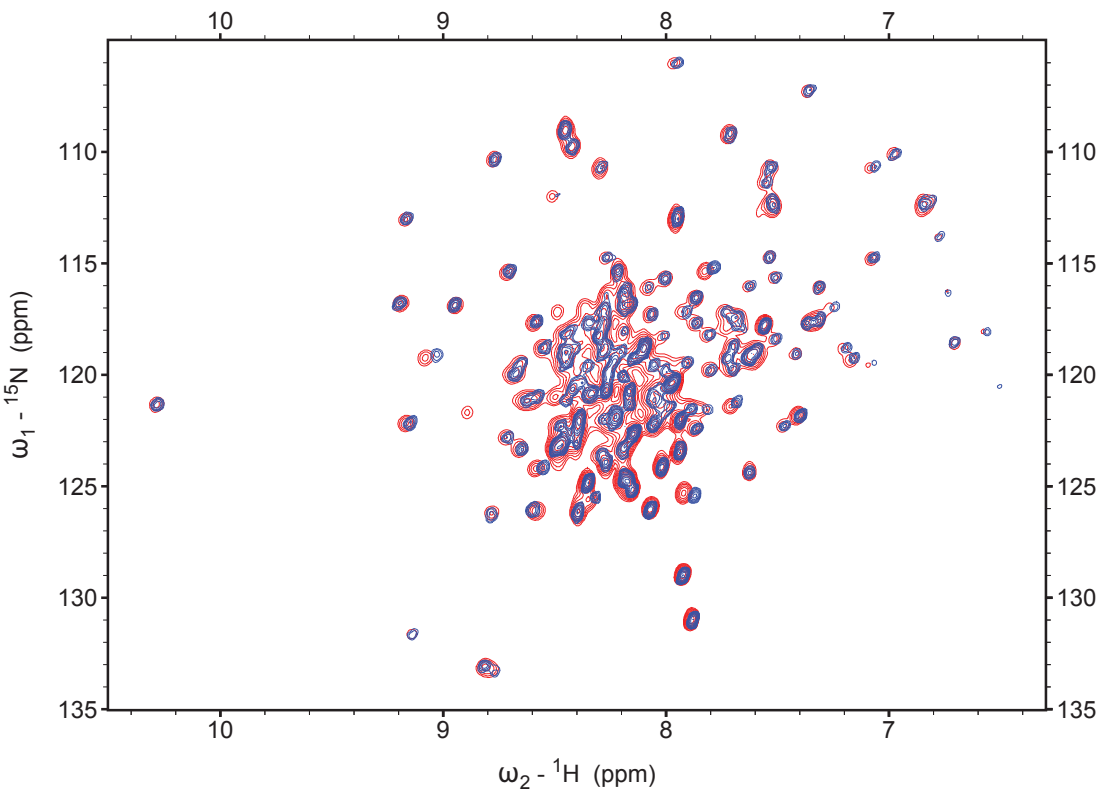


Figure 94. **Analyse par R.M.N. du complexe Rsa1p (317-375)/Hit1p (44-164)** à 299K dans un tampon phosphate 10 mM, pH 6,8. Superposition des spectres ^{15}N-^{1}H TROSY-HSQC du complexe à 150 mM NaCl (rouge) et 0 mM NaCl (bleu).

En fait, la poursuite des expériences R.M.N. sur le complexe, a permis rapidement de mettre en évidence la présence de régions flexibles donnant lieu à des pics très intenses et souvent superposés les uns sur les autres, ce qui rendait l'attribution difficile. Afin d'obtenir un gain de résolution et de sensibilité au niveau des signaux R.M.N., des expériences ont étés réalisées sur un spectromètre à 800 MHz et un échantillon triplement marqué $^{15}N/^{13}C/^{2}H$ deutéré à 50% a été produit. Sur la base de l'attribution partielle réalisée, une analyse des déplacements chimiques des résidus par le logiciel TALOS+ (Shen *et al.*, 2009) et une prédiction de la structure secondaire par le logiciel PsiPred (Jones, 1999), un modèle de structuration secondaire du complexe a pu être proposé. Ceci a confirmé le fait que certaines régions du complexe possèdent un degré important de flexibilité. En particulier les régions 44-69 de la protéine Hit1p et 353-375 de la protéine Rsa1p. Il est intéressant de noter que ces résultats recoupaient les informations obtenus par l'analyse en spectrométrie de masse des produits de protéolyse ménagée du complexe Snu13p/Rsa1p (231-375)/Hit1p (Résultats & Discussions, Chapitre 2, paragraphe 1.2.3, figure 85). En effet, lors de cette analyse nous avions détecté une hydrolyse probable de la protéine Rsa1p au niveau du résidu 348 et la production du fragment 69-164 de la protéine Hit1p avait été proposée pour expliquer l'espèce B observée (Figure 86B).

Sur la base de ces résultats, il a été décidé de continuer l'étude par R.M.N. en produisant le complexe Rsa1p (317-352)/Hit1p (70-164) qui s'est aussi avéré soluble et dont le poids moléculaire était tout particulièrement adapté à une étude structurale par R.M.N. L'ensemble des expériences nécessaires à l'attribution de la chaîne principale, des chaînes latérales mais aussi à l'établissement des contraintes NOEs ont été enregistrées et M. Quinternet réalise actuellement son attribution. Les bases moléculaires de l'interaction mise en place entre les protéines Rsa1p et Hit1p, impliqués dans la biogenèse des snoRNP à boîtes C/D, devraient donc être prochainement identifiées.

1.5 La protéolyse ménagée en présence du snoRNA U3

En utilisant le complexe Snu13p/Rsa1p (231-375)/Hit1p que j'avais produit, B. Rothé avait pu montrer l'association de ce complexe au snoRNA U3. Notre équipe avait déjà montré que par sa capacité de fixation aux motifs en *« K-turn »*, la protéine Snu13p se lie aux couples de boîtes B/C et C/D du snoRNA U3 (Marmier-Gourrier *et al.*, 2003). J'avais montré que l'interaction de Snu13p à l'ARN n'abolissait pas son interaction à Rsa1p comme nous le verrons plus loin. La fixation du complexe Snu13p/Rsa1p (231-375)/Hit1p au snoRNA U3

montrait que l'association de Snu13p à l'ARN n'empêchait pas non plus l'association de Hit1p à Rsa1p dans le complexe à trois partenaires. Néanmoins il était intéressant de savoir si la présence de l'ARN pouvait changer la conformation du complexe. Pour cela j'avais utilisé l'approche de protéolyse ménagée. Nous avions réutilisé les mêmes conditions d'hydrolyse par les protéases qu'en absence d'ARN.

Les résultats obtenus sont présentés dans la figure 95. Le complexe Snu13p/Rsa1p (231-375)/Hit1p a été incubé avec différentes protéases à différentes concentrations (comme précisé au paragraphe 1.2) en absence et en présence de l'ARN ceci avec un ratio ARN/complexe de 1:1. L'ARN utilisé correspondait au snoRNA U3 Δ2,3,4 qui est fonctionnel *in cellulo* (Cléry *et al.*, 2007; Marmier-Gourrier *et al.*, 2003). Nous retrouvions des profils d'hydrolyse très similaires à ceux obtenus en absence d'ARN. Néanmoins un point intéressant était à noter : après hydrolyse par Glu-C (1:100 et 1:1000) nous observions moins de clivage de Rsa1p qu'en absence d'ARN. Donc au sein du complexe hétérotrimérique, Rsa1p doit établir un ensemble de contacts sans doute non spécifiques avec l'ARN expliquant sa protection partielle. D'ailleurs, B. Rothé a observé en complément une protection de l'ARN par la protéine Rsa1p (231-375) lors d'expériences d'empreintes.

Il pourrait donc être intéressant de poursuivre l'étude structurale du complexe Snu13p/Rsa1p (231-375)/Hit1p en y ajoutant une composante ARN.

2 Etude fonctionnelle sur les complexes Snu13p/Rsa1p/Hit1p et Rsa1p/Hit1p

2.1 Présentation des travaux décrits dans le manuscrit n°1 et de ma participation aux travaux

Les complexes que nous avons décrits dans le premier chapitre des résultats et la première partie de ce chapitre, Snu13p/Rsa1p (qui fera l'objet d'une description plus approfondie dans le chapitre 3), Rsa1p/Hit1p et Snu13p/Rsa1p/Hit1p ont été utilisés par B. Rothé pour réaliser des études fonctionnelles. Il faut noter que jusque-là, Hit1p n'avait été décrite que comme interagissant avec Rsa1p dans un crible de double hybride à haut débit (Ito *et al.*, 2001), sa fonction dans l'assemblage des snoRNP C/D était inconnue.

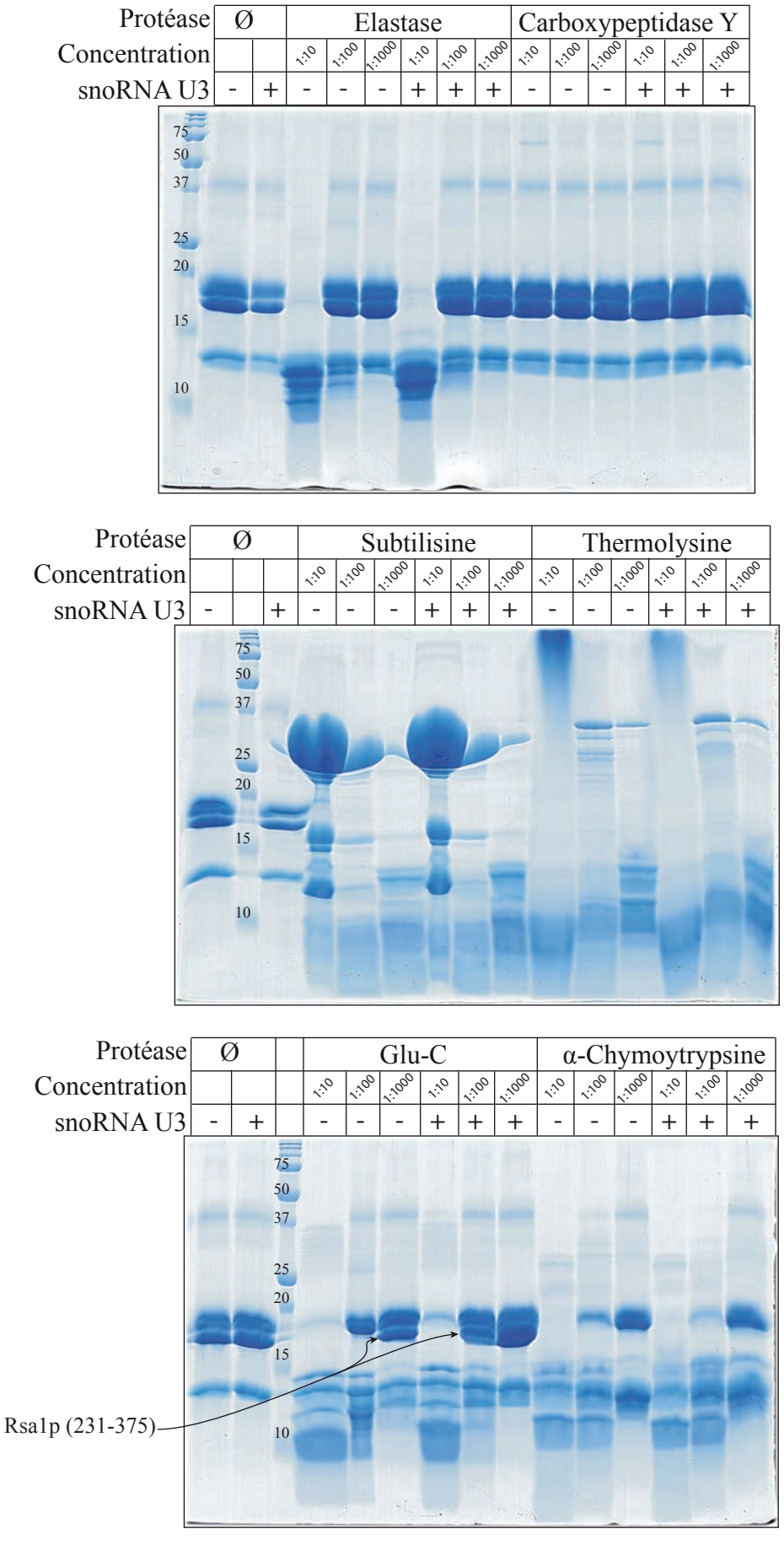


Figure 95. **Protéolyse ménagée du complexe Snu13p/Rsa1p (231-375)/Hit1p en présence et en absence du snoRNA U3.** Les produits d'hydrolyse ont été fractionnés par SDS-PAGE à 17,5%. Les complexes ARN/protéines ont été formés à un ratio ARN/complexe de 1:1.

Les données que nous venons de décrire suggéraient que son action passait par l'interaction que nous avons détecté avec Rsa1p. B. Rothé a commencé par réaliser des tests de co-purification de protéines avec Rsa1p à partir d'un extrait de levure exprimant la protéine Rsa1p-TAP, cela a permis de montrer par spectrométrie de masse que les protéines Rsa1p et Hit1p étaient associées de manière stable *in vivo* et que la protéine Hit1p pourrait jouer un rôle de co-facteur de la protéine Rsa1p. Les travaux de B. Rothé montraient en effet que la protéine Hit1p est nécessaire pour maintenir le taux de Rsa1p dans la cellule et que Hit1p s'associe avec les précurseurs du snoRNA U3 sans pour autant être détecté avec la forme mature de cet ARN. Par la recherche de mutants de levure conduisant à une létalité, B. Rothé a montré que les délétions des gènes de RSA1 ou HIT1 en présence d'une délétion du gène PIH1 conduisait toute les deux à un phénotype létal. Ce qui renforce l'idée d'un lien fonctionnel entre ces deux protéines. Une possibilité était qu'elles participaient à une même fonction essentielle de la cellule au sein d'un même complexe. Grâce aux conditions que nous avions développées pour la production chez *E. coli* des protéines Snu13p, Rsa1p, Hit1p et Nop58p et de leurs complexes et à partir de l'utilisation de protéines recombinantes obtenue à partir d'un extrait d'*E. coli* S30, B. Rothé a pu montrer la formation de complexes entre Nop58p et les complexes que nous avions purifiés Snu13p/Rsa1p (231-375), Rsa1p (231-375)/Hit1p et enfin Snu13p/Rsa1p (231-375)/Hit1p. Ce qui était en accord avec les données de co-expression (Chapitre 1 Résultats & Discussion, paragraphe 2.1.2, pistes 60, 62 et 63).

La caractérisation des propriétés d'interaction de Hit1p avec Rsa1p, les domaines d'interaction des deux protéines que j'avais mis en évidence ainsi que la formation du complexe tripartite Snu13p/Rsa1p (231-375)/Hit1p que j'ai identifié sont décrits dans le manuscrit suivant dont je suis troisième auteur. Les propriétés fonctionnelles de ces protéines et leurs complexes ont été étudiés par B. Rothé et J.M. Saliou a réalisé les analyses par spectrométrie de masse.

2.2 Manuscrit n°1

Functional link between proteins Rsa1 and Hit1 during box C/D snoRNPs biogenesis in yeast

Benjamin Rothé, Jean-Michel Saliou, Régis Back Xavier Manival, Alain Van Dorssealer, Sarah Sanglier-Cianferani, Christiane Branlant and Bruno Charpentier

Functional link between proteins Rsa1 and Hit1 during box C/D snoRNPs biogenesis in yeast

Benjamin Rothé[1], Jean-Michel Saliou[2], Régis Back[1], Xavier Manival[1], Alain Van Dorssealer[2], Sarah Sanglier-Cianferani[2], Christiane Branlant[1] and Bruno Charpentier[1]

AREMS[1] (ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale), Nancy Université, Faculté de Médecine, 9, avenue de la forêt de Haye, BP 184, 54500 Vandoeuvre-Lès-Nancy, Cédex, France. LMBSO[2] (Laboratoire de Spectrometrie de Masse BioOrganique, Strasbourg, Universite, IPHC, 25 rue Becquerel, 67087 Strasbourg, France.

§ Address correspondence to: Bruno Charpentier, +33 383 68 55 08 , Fax: +33 383 68 55 09, E-mail bruno.charpentier@maem.uhp-nancy.fr

ABSTRACT

Biogenesis of box C/D small nucleolar ribonucleoprotein particles (snoRNPs) relies on conserved *trans*-acting factors, which are proposed to facilitate the assembly of the core proteins Snu13p, Nop58p, Nop56p, and Nop1p/Fibrillarin on box C/D small nucleolar RNAs (snoRNAs). In the yeast assembly machinery, protein Rsa1 acts as a platform for direct interaction with the RNA-binding core protein Snu13, and with protein Pih1, which forms the R2TP complex by association with the Hsp82-adaptor Tah1p and the two helicases Rvb1p, Rvb2p. Here, we identified by a proteomic-based approach protein Hit1 as a novel Rsa1p-interacting partner, and we show it is required to maintain steady-state levels of Rsa1p and box C/D snoRNAs. Moreover, we found presence of Hit1p with the other known assembly factors on precursors of U3 box C/D snoRNA, and observed that synthetic lethality arises upon combination of null mutants for *RSA1* and *PIH1* genes, or *HIT1* and *PIH1* genes, but not for *RSA1* and *HIT1* genes. Moreover, core protein Nop58 is shown here to directly interact *in vitro* with purified Rsa1p-Hit1p heterodimer or Snu13p-Rsa1p-Hit1p heterotrimer. Together, our observations conduct to classify Hit1p as a new C/D snoRNP assembly factor functionally linked to Rsa1p for binding to C/D snoRNP core proteins.

INTRODUCTION

Ribonucleoprotein particles (RNPs) participate in various fundamental functions for cell growth and development. In Eukarya, snoRNPs (small nucleolar RNPs) play essential role during processing of ribosomal RNAs (32). These particles are divided into two classes, each mediating one type of chemical modification on RNA: box C/D snoRNPs catalyze ribose 2'-O-methylation, and box H/ACA snoRNPs catalyze conversion of uridine to pseudouridine. Furthermore, some snoRNPs like C/D snoRNPs U3 and U8 in mammals (16), and H/ACA snoRNPs U17 in mammals and its yeast homolog snR30 (2, 36) are not involved in those chemical modifications but in RNA endonucleolytic cleavages during pre-rRNA processing.

Each box C/D snoRNPs arises by the association of a box C/D snoRNA (small nucleolar RNA) with a common set of four core proteins called Snu13p, Nop1p, Nop56p, and Nop58p in yeast *Saccharomyces cerevisiae* (respectively, 15.5K, Fibrillarine, NOP56, and NOP58 in human) (28, 29, 51). Snu13p and 15.5K belong to the L7Ae family of RNA binding proteins whose members are components of several essential RNPs. Snu13p and 15.5K are present in C/D snoRNPs (51), in the RNP assembled on the snoRNA U3-specific boxes B and C (8), and in the U4 small nuclear ribonucleoprotein particle (U4 snRNP) involved in pre-mRNA splicing (37, 51). Nhp2 is the L7Ae protein present in H/ACA snoRNPs (18, 49) and in the H/ACA RNP of the vertebrate telomerase complex. Cytoplasmic SECIS mRNPs, which are involved in selenocystein incorporation in vertebrates proteins, contain the L7Ae protein SBP2 (1, 7, 9, 25, 51). The L7Ae family of proteins share a common RNA binding domain that specifically recognizes

the structural RNA motifs Kink-turn (K-turn) and K-loop (23, 48). Initial binding of L7Ae proteins on these RNA motifs nucleates assembly of RNP by facilitating specific recruitment of the other core proteins (8, 46, 50).

Biogenesis of nuclear L7Ae RNPs is a complex process starting from transcription of RNA precursors and their maturation, to synthesis and nuclear import of RNP core proteins, which are orderly assembled on RNA during RNA processing, and finished by the delivery of mature RNPs to functional sites (32, 53). For snRNPs the process includes assembly steps in the cytoplasm for review, (39). Several studies highlighted the involvement of *trans*-acting factors during RNPs assembly. In the case of snRNPs formation, the SMN (Survival Motor Neuron) multiprotein complex controls cytoplamic assembly steps of the particles. Indeed, SMN complex specifically recognizes snRNA sequences, and mediates the recruitment and the binding of a set of Sm proteins to form the Sm-core RNPs (24, 40). The Naf1 protein facilitates the H/ACA snoRNP co-transcriptional assembly by interacting with the Cbf5 core protein and with the RNA polymerase II CTD (10, 11, 14, 52). Recently, we and others showed that assembly of the L7Ae RNPs requires a set of conserved factors (3, 33). These factors will be written here as a couple formed by the yeast protein and human protein (yeast/human). This machinery includes Rsa1p/NUFIP, the Hsp82/Hsp90 co-chaperone R2TP complex composed of the AAA+ hexameric helicases Rvb1p/TIP49 and Rvb2p/TIP48, Tah1p/hSPAGH, and Pih1p/hPIH1(NOP17). The Hsp82/Hsp90 chaperon interacts with the R2TP complex through binding to the adaptor Tah1 (3, 33, 54).

Rsa1p and NUFIP are believed to play a central role in the RNP assembly process by interacting simultaneously with Pih1p/hPIH1, Rvb1p/hTIP49 and Rvb2p/hTIP48, and with the L7Ae family members Snu13p/15.5K, Nhp2p/hNHP2, and SBP2 (3). In particular, we demonstrated that a conserved stretch of 32 amino acids designated yPEP in Rsa1p and hPEP in NUFIP is dedicated to the interaction with proteins Snu13p/15.5K and Nhp2p/hNHP2 (3), and residues of the protein-protein interface were identified (Rothé, Back *et al*., unpublished data). Moreover, based on NMR data for yPEP and molecular modeling, we proposed a 3D model of the interaction in which Rsa1p/NUFIP binds Snu13p/15.5K opposite to the RNA binding region suggesting that the assembly factor acts as an external scaffold during RNP assembly.

In yeast, the other box C/D core protein Nop58 has been shown to be contacted by protein Pih1 (15). In a Δ*PIH1* mutant lacking Pih1p, the C/D core proteins Snu13, Nop1, Nop56, and Nop58 accumulate into the cytosol and the nucleus (15), and the amounts of co-immunoprecipitated core proteins are reduced (55). Furthermore, Pih1p is closely associated with the low molecular weight Tah1 protein (13, 54, 55), which contains a TPR domain and acts as co-factor of the Hsp82 chaperon (54) (Back *et al*., unpublished data).

This machinery dedicated to RNP assembly is based on a complex interaction network including activities of helicases and chaperons. Data obtained in human indicate that Hsp90, in association with the R2TP co-chaperon complex, participates to the RNP assembly process by controlling stability of the RNP core proteins (15.5K, NOP58, NHP2, SBP2) (3, 55). Furthermore, it has been proposed that NUFIP and the TIP48/TIP49 helicases facilitate interactions between C/D snoRNP components in a mechanism regulated by the ATPase activity of TIP48/TIP49 (3, 33, 34). Nevertheless, the precise mechanism by which the machinery assembles RNPs remains poorly understood. While protein-protein interactions were mapped (3, 13, 55), the *in vivo* existence of a complex formed by association of the full set of the identified factors is not settled. Moreover, data on interactions between R2TP and intermediates of the snoRNP biogenesis are lacking. Furthermore, the list of assembly factors could still be incomplete

and additional factors may participate in the RNP assembly.

In this work, we set out to characterize complexes containing TAP-tagged Rsa1 and Pih1 proteins. Proteomic analysis of immunoprecipitated complexes indicates that Rsa1p is not stably associated with R2TP, but co-purifies with the Hit1 protein. Here we show that Hit1p is physically and functionally linked to Rsa1p, and that it represents an additional factor of the C/D snoRNP assembly machinery. Moreover, we purified protein complexes containing the core protein Snu13 associated with the Rsa1p-Hit1p heterodimer. We show that this heterotrimer interacts with the other core box C/D snoRNP protein Nop58.

MATERIAL & METHODS

Plasmid constructions. Plasmids were manipulated by standard techniques. Constructs used for co-expression studies in *E. coli* were obtained by cloning PCR-amplified fragments between the *Nde*I and *Bam*HI sites of plasmids (pnEA-3CH and pnCS) (12). The primers used for PCR amplification introduced a *Nde*I site at the 5' end and a *Bam*HI site at the 3' end of the fragment (Rothé, Back *et al*., unpublished data). The bacterial expression vector pGEX-6P-1::*SNU13* encoding the GST-Snu13p fusion protein was already described (31). The recombinant pACT2 plasmids expressing Gal4-AD-Rsa1p, Gal4-AD-$Rsa1_{Cter}$ and used for yeast two hybrid (Y2H) assays were already described (3). Recombinant plasmids pAS2 and pACT2 expressing respectively Gal4-BD-Hit1p and Gal4-AD-HA-Hit1p were obtained by the same procedure. Plasmid expressing GFP-Rsa1p was provided by E. Bertrand (IGM, Montpellier, France).

Production of recombinant proteins. The recombinant pGEX-6P-1::*SNU13* plasmid was used to produce Snu13p fused to the glutathione S-transferase (GST) (31) in the *E. coli* CodonPlus strain (Novagen©). The recombinant GST-Snu13 protein was purified from cellular extract under native conditions, using Glutathione-Sepharose 4B as recommended by the manufacturer (GE Healthcare©). The Snu13 protein was recovered after overnight cleavage on the beads using PreScission protease (GE Healthcare©). The cleaved recombinant Snu13 protein had two additional amino acid residues at its N-terminus (G and P), compared to the native protein.

Yeast strains and measurements of RNA levels. Wild-type, knock-out, and TAP-tagged strains used in this study were provided by Euroscarf ®. Standard techniques were used for growth and handling of *S. cerevisiae* strains. Total RNAs were extracted from exponentially growing cells (OD_{600} ~0.8). Northern-blot analyses were carried out according to standard procedures on 5 µg of total RNAs using as probes 5' ^{32}P-radiolabeled oligodesoxyribo nucleotides complementary to RNAs of interest.

Synthetic lethal screen by diploid strains sporulation and dissection. Diploid strains were obtained by cross-breeding on YPD plates of simple knock-out haploid strains By4741 (MATa) and By4742 (MATα) from Euroscarf®. Diploid cells were resuspended in 2 mL of liquid sporulation medium (1% potassium acetate; 0.05% zinc acetate). Sporulation cultures were incubated on a roller wheel for 5 days at 25°C, then for 3 days at 30°C. Cells were finally dissected with a micromanipulator (Singer®) and separated on YPD plates. Strain genotypes were controlled by standard techniques including genomic DNA extraction and PCR analysis.

Purification of multiprotein complexes by protein A selection affinity. *S. cerevisiae* strains SC4186 (TAP-*RSA1*), SC3760 (TAP-*HIT1*), and SC3165 (TAP-*PIH1*) expressing TAP-tagged proteins, as well as the untagged strain BY4741, were grown at 30°C in YPD to OD_{600}~0.8-1. Protein A purifications were performed in breaking buffer (100 mM NaCl, 8% glycerol, 0.1 mM DTT, 100mM

Tris-HCl, pH8) from cells representing 2500 units of OD_{600} of the culture (43). Total cellular extracts were incubated for 12 hours at 4°C with IgG-Sepharose beads (Sigma-Aldrich®). Beads were then washed in Mobicol "classic" columns (MoBiTec®) with washing buffer (100 mM NaCl, 0.1% IGEPAL, 0.1mM DTT, 20 mM Tris-HCl, pH 7.4). After TEV cleavage, eluted proteins were separated by SDS–PAGE and stained with Coomassie Blue.

Mass Spectrometry-based Proteomic analysis. In gel digestion of SDS-PAGE gel slices was performed as previously described (35). The resulting peptide extracts were analyzed by nanoLC-MS/MS on a nanoACQUITY Ultra-Performance-LC (UPLC, Waters, Milford, UK) coupled to a SYNAPT HDMS G1 hybrid quadrupole-time-of-flight mass spectrometer (Waters, Milford, UK) equipped with a Z-spray ion source and a lock mass system. Peptide mixtures were loaded on a Symmetry C18 trap precolumn (180 µm inner diameter × 20 mm, particle size 5 µm; Waters, Milford, UK) and peptides were separated on a ACQUITY UPLC® BEH130 C18 column, 75 µm x 200 mm, 1.7 µm particle size (Waters, Milford, UK). The solvent system consisted of 0.1 % formic acid in water (solvent A) and 0.1 % formic acid in acetonitrile (solvent B). Trapping was performed during 3 min at 5 µL/min with 99 % of solvent A and 1 % of solvent B. Elution was performed at a flow rate of 400 nL/min, using 1-50 % gradient (solvent B) over 35 min at 45°C followed by 65 % (solvent B) over 5 min. Parameters of the mass spectrometer were set as follows: the capillary voltage was set to 3500 V and the cone voltage to 35 V. Mass calibration of the TOF was achieved using [Glu1]-fibrinopeptide B (GFP) over the m/z range 50-2000 in the positive ion mode. Online correction of this calibration was achieved using lock-mass on product ions derived from the [Glu1]-fibrinopeptide B (GFP). The ion $(M+2H)^{2+}$ at m/z 785.8426 is used to calibrate MS data and the fragment ion $(M+H)^{+}$ at m/z 684.3469 is used to calibrate MS/MS data during the analysis. For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The 4 most abundant peptides, preferably doubly and triply charged ions, were selected on each MS spectrum for further isolation and CID fragmentation. MS/MS fragmentation was performed in the collision cell using argon as collision gas. The complete system was fully controlled by MassLynx 4.1 (SCN 566, Waters, Milford, UK). Raw data collected during nanoLC-MS/MS analyses were processed and converted with ProteinLynx Browser 2.3 (23, Waters, Milford, UK) into *.pkl peak list format. The MS/MS data were submitted to Mascot search engine (Matrix Science, London, UK) installed in its 2.2.04 version on a local server. Searches were performed against a composite target-decoy database including both forward (target) and reversed (decoy) SwissProt protein sequences (21 January 2010, 5 1028424 total entries). Searches were performed with a tolerance on mass measurements of 0.2 Da for precursor and 25 ppm for fragment ions. Carbamidomethylation of cystein residues, oxidation of methionine residues and acetylation of protein N-term residues were searched as variable modifications. Up to one missed cleavage was allowed. Scaffold V3_00_01 (Proteome Science, Portland, USA) was used for identification validation and false positive rate estimation for protein identification. For each sample, peptides were filtered out according to the cutoff set for proteins hits with 2 or more peptides (Mascot Ion Score >20 and Mascot Ion Score - Mascot Identity Score >0) and no false positive protein was identified. Further data treatment were performed with Microsoft Excel.

Noncovalent Mass Spectrometry. Before noncovalent mass spectrometry experiments, protein buffer was exchanged against 500 mM or 1 M ammonium acetate buffer pH 7.5 using microcentrifuge gel-filtration columns (Zeba 0.5 mL, Thermo Scientific, Rockford,

IL). Protein concentrations were determined spectro photometrically. Mass spectrometry experiments were performed on a hybrid quadrupole/ion mobility separator/time-of-flight instrument (Synapt HDMS G2, Waters, Manchester, UK)) equipped with an automated chip-based nanoelectrospray source (Triversa Nanomate, Advion Biosciences, Ithaca, NY) operating in the positive ion mode. External calibration was performed with the multiply charged ions produced by 2 µM horse heart myoglobin diluted in 1:1 (v/v) water/acetonitrile acidified with 1 % (v/v) formic acid. Prior to noncovalent analysis, each sample was analyzed in classical denaturing conditions in order to asses its purity and homogeneity. Mass measurements under denaturing conditions were carried out by diluting samples to 2 µM in water/acetonitrile/formic acid (50:50:1). Accelerating voltage (Vc) was set to 20 or 40 V while pressure in the interface region (Pi) was 1.1 mbar. Molecular weights of 17599.0 ± 0.1 Da, and 18241.9 ± 0.3 Da were measured for Rsa1p$_{230-375}$-Hit1p expression and were respectively in agreement with theoretical masses calculated from the amino acid sequence (GPHM-Rsa1p(230-375) and Hit1p(2-164)). Molecular weights of 17176.8 ± 0.1 Da, 18243.4 ± 0.6 Da and 13860.4 ± 0.2 Da were measured for Snu13p-Rsa1p$_{230-375}$-Hit1p expression and were respectively in agreement with theoretical masses calculated from the amino acid sequence (GPH-Snu13p(1-129), Rsa1p(230-375), and Hit1p(2-164)). Experiments under nondenaturing conditions (noncovalent conditions) were realized in either 500 mM (for Rsa1p$_{230-375}$-Hit1p) or 1 M (for Snu13p-Rsa1p$_{230-375}$-Hit1p) ammonium acetate buffer at pH 7.5. Optimal interface parameters (Vc, Pi) were used in order to combine intact noncovalent complexes detection to efficient ion desolvation and ion transmission in the mass spectrometer (45). Data analysis was performed with MassLynx 3.5 (Waters, Manchester, UK).

RNA co-IP assays. Co-immunoprecipitation assays were realized as previously described in (3). Briefly, yeast cells expressing the TAP-tagged proteins were grown to exponential phase before centrifugation. Cells were lysed by bead-beating in breaking buffer (100 mM NaCl, 8% glycerol, 0.1 mM DTT, 100mM Tris-HCl, pH8) and lysates were added to IgG-Sepharose beads (Sigma-Aldrich®) and incubated 2 h at 4°C. Beads were washed four times in washing buffer (100 mM NaCl, 0.1% IGEPAL, 0.1mM DTT, 20 mM Tris-HCl, pH 7.4). RNAs were extracted with phenol-chloroform and analyzed by RT-PCR. cDNAs were generated using oligonucleotides OG-Rev (5'-CAGTGCAGGGT CCGAGGTATTCTGAAAACCAAACCTT TGG-3') containing the underlined sequence exogenous to the yeast genome, and used as templates for PCR amplification. PCR was then performed by using the 5' oligonucleotide OG-5'PCRa 5'-TCTGTGTCGACGTACTTCA-3' complementary to exon 1 of *snR17A* and *snR17B* (from nucleotides 1 to 14) or the OG-5'PCRb 5'-AGTCTTAGGTACTAGAGTTT-3' complement -tary to *snR17A* intron (from nucleotides 66 to 85) and the 3' oligonucleotide OG-3'PCR 5'-CA GTGCAGGGTCCGAGGTATT-3' complement-tary to the yeast exogenous sequence.

Western-blot. Crude cell extracts were prepared from pellets of 7.10^8 yeast cells resuspended in 300 µL of buffer (150 mM KCl, 5 mM MgCl2, 0,05% TRITON-X100, 20 mM Tris-HCl, pH 7,5), and lysed by bead beating. The lysates were centrifuged two times at 10,000 g for 5 min. Concentration of proteins in the supernatants obtained form the various samples were estimated by Bradford assays. Equivalent amounts of protein were fractionated on 12,5 % SDS-PAGE and western blots were performed according to standard procedures using rabbit commercial anti-GFP, anti-HA, anti-PMA1 (provided by B. André, LPMC,

IBMM-ULB, Bruxelles, Belgium), or anti-Tah1p-Pih1p antibodies (provided by S. Moreira LEBS, Gif-sur-Yvette, France).

Formation of complexes between proteins co-expressed *in E. coli.* Co-expression tests in *E. coli* were carried out using the procedure described previously (44) (Rothé, Back *et al*., unpublished data). The recombinant plasmids pnEA-3CH expressing His_6Snu13p, His_6Rsa1p, His_6Rsa1p$_{230-375}$, or His_6Tah1p, and the recombinant plasmid pnCS expressing Hit1p, Rsa1p$_{230-375}$, or Pih1p were simultaneously transformed into pRare2 *E. coli* host cells (Novagen). Growth was performed in 500 mL of 2X concentrated LB medium at 37°C to OD_{600} ~0.8. Then, the temperature was shifted to 20°C. At this point, co-expression of the proteins was induced by addition of isopropyl-β-D-thiogalactopyranoside (Euromedex) at a 1 mM final concentration. Cells were further grown overnight at 20°C. They were collected by low speed centrifugation, resuspended in buffer A (20 mM HEPES, pH 7.5; 400 mM NaCl, 5 mM β-mercaptoethanol), and lysed by sonication. The soluble fraction recovered by high speed centrifugation was mixed with 100 µl of Talon resin (Clontech). The supernatant was removed after 1 h incubation, and the resin washed extensively with buffer A. Proteins were eluted from the column using buffer A containing 500 mM imidazole. Eluted proteins were analyzed by SDS-PAGE.

Electrophoresis Mobility Shift Assays (EMSA). Yeast RNAs U14, variant of U3 snoRNA U3Δ2,3,4 (31), and $tRNA^{His}$ were synthesized by T7 RNA polymerase by *in vitro* transcription of PCR fragments in the conditions previously described (3, 6). RNA transcripts were radiolabeled at their 5' ends by T4 polynucleotide kinase and [γ-^{32}P]ATP. 500 nM of various proteins or complexes were mixed and incubated for 20 min at 4°C with 5 fmol of 5'-end ^{32}P-radiolabeled RNA in buffer D (20 mM HEPES, pH 7.9; 150 mM KCl; 1.5 mM $MgCl_2$; 0.2 mM EDTA; 10 % glycerol). The RNA-protein complexes formed were visualized by native gel electrophoresis.

Yeast two hybrid assays. For yeast two-hybrid assays (Y2H), appropriate pACTII and pAS2 plasmids were used to transformed haploid test cells (strains Y190 and Y187, respectively), which were then crossed. Diploids were selected on media suitable for double selection (Leu-, Trp-) and then plated on media suitable for triple selection (Leu-, Trp-, His-). Various concentrations of 3-Amino-1,2,4-triazol (3AT) (Sigma®) were used to evaluate the strength of the interactions. Growth was assessed after three days of incubation at 30°C.

His-tagged protein pull-down. Complex formation was tested in a volume of 20 µL of buffer D (20 mM HEPES, pH 7.9; 150 mM KCl; 1.5 mM $MgCl_2$; 0.2 mM EDTA; 10 % glycerol) in the presence of recombinant proteins at a concentration of 2.5 µM and 4 µL of ^{35}S radiolabeled Nop58p, obtained by using the kit *E. coli* T7 S30 Extract System for Circular DNA (Promega). After 20 min at 30°C, 180 µL of buffer (10 mM Tris-HCl, pH 8, 150 mM NaCl, 0.1 % Igepal, 5 mM imidazole) and 20 µL of cobalt-sepharose beads (GE Healthcare) were added. The mixture was incubated on a wheel at 4°C for 90 min. Finaly, the beads were washed three times binding buffer, and retention of radiolabeled protein was analyzed by gel electrophoresis SDS-PAGE 7.5 %.

RESULTS

Rsa1p is not stably associated with the R2TP complex

Aiming at characterizing the variety of complexes that are formed in the cell during C/D snoRNP biogenesis, we attempted to identify proteins that co-purify with the RNPs assembly factors Rsa1p and Pih1p. We used a procedure based on a single-step immunoglobulin G (IgG) affinity purification of TAP-tagged Rsa1 and Pih1 proteins followed by protein identification with proteomic analysis. Yeast strains expressing one of the proteins tagged at their

C-terminus with the tandem affinity purification (TAP) sequence were grown. Cell extracts were prepared from cells in log-phase, during which biosynthesis of snoRNPs is very active. After capture on IgG-coupled beads and cleavage with the TEV protease, the proteins co-purified with TAP-Rsa1p or TAP-Pih1p were fractionated by SDS-PAGE (Fig. S1). Gel slices were cut systematically each 2 mm, and protein contents were identified by mass spectrometry-based proteomic analysis (Supplementary data Table S1).

Proteins of interest related to RNP assembly and identified with a significant number of peptides are listed in Table 1. As expected, proteins constitutive of the whole R2TP complex mainly co-purify with TAP-Pih1p, as Rvb1p, Rvb2p, and Tah1p are identified with a high number of peptides (16 for Rvb1p, 17 for Rvb2p, and 4 for Tah1p). In parallel, C/D core proteins Snu13p, Nop56p, Nop58p, and Nop1p were also identified in association with Pih1p but with fewer peptides. Surprisingly, Rsa1p does not co-purify with any of the R2TP components as no peptides of Rsa1p were found with TAP-Pih1p. In the TAP-Rsa1p experiment, C/D core proteins were also detected with a low number of peptides. No peptide from Pih1p and from any R2TP subunits were found, which is consistent with results obtained in the TAP-Pih1p experiment.

Altogether these observations first highlight the transient feature or low stability of the interactions between these assembly factors and core proteins. In addition, the low steady-state level of the fully assembled R2TP-Rsa1p complexes can result from a transient association of Rsa1p with R2TP, or alternatively the complex is destabilized during purification due to weak interactions.

Co-purification of Hit1p with Rsa1p

As proteomic analysis identified Hit1p with a high number of peptides (7) among the proteins that co-purify with TAP-Rsa1p, we next performed the IgG affinity purification of TAP-Hit1p. Rsa1p was found to be strongly associated with Hit1p as 10 Rsa1p peptides are detected in the eluted TAP-Hit1p samples (Table 1). As in TAP-Rsa1p experiment, only fewer peptides from the C/D snoRNP core proteins Nop1p and Snu13p that co-purify with TAP-Hit1p were observed, and no peptide from the R2TP complex was detected (Table 1). These data suggest the *in vivo* existence of a stable complex containing Rsa1p and Hit1p, which is in agreement with previously published two hybrid screen results (20).

Rsa1p, Hit1p, and Pih1p are related to the stability of C/D snoRNAs

The involvement of Hit1p in snoRNP biogenesis is attested by our phenotypic analyses of the *HIT1* gene null mutant (Δ*HIT*). Indeed, this mutant shows a phenotype similar to those observed for null mutants of genes *RSA1* and *PIH1*, which respectively encode proteins Rsa1p and Pih1p involved in the assembly of the C/D snoRNPs (Fig. 1). First we showed that the disruption of genes *RSA1*, *HIT1,* or *PIH1* is correlated with a slow growth phenotype (Fig. 1A). Indeed, in standard growth conditions the Δ*RSA1*, Δ*HIT1*, and Δ*PIH1* null mutant strains have doubling times respectively ~58 %, ~56 %, and ~31 % lower than the WT cells. Northern blot analyses indicate that these slow growth phenotypes are associated with lower levels of C/D snoRNAs whereas the levels of H/ACA snoRNAs and UsnRNAs remain unchanged (Fig. 1B). The stability of some C/D snoRNAs, i.e. U14, snR190, snR70, is decreased to the same extent in the three mutants. The levels of the C/D snoRNAs snR45, snR63, and snR13 are also reduced in the Δ*PIH1* mutant, but these RNAs are barely detected in the Δ*RSA1* and Δ*HIT1* mutants. As the amount of U3 snoRNA was only found slightly decreased in the various strains, we analyzed a truncated version of the U3 snoRNA (U3Δ2,3,4) that accumulates precursors of U3 RNPs when the encoding gene is transfected in yeast cells (3, 27). We already reported the higher accumulation of U3Δ2,3,4 precursors in cells lacking Rsa1p (Fig. 1C) or Pih1p (3). Such accumulation is

also observed in Δ*HIT1* cells (Fig. 1C), indicating that Hit1p is also related to the processing of U3.

Given that snoRNPs participate in ribosome biogenesis, we further analyzed the phenotypes of the Δ*RSA1*, Δ*PIH1*, and Δ*HIT1* null mutant strains, and detected by Northern blot that these strains accumulate the 35S ribosomal RNA precursor and several intermediates of its processing pathway (Fig. 1D): intermediate 32S which results from a lower efficiency of cleavage at site A3, and intermediates 23S and 20S of the 18S maturation pathway accumulated due to a lower efficiency of cleavages at sites A1 and A3.

Altogether, Δ*HIT1* and Δ*RSA1* strains display similar marked phenotypes strongly suggesting that both proteins participate in a common step of C/D snoRNP biogenesis. Interestingly, under the tested conditions, absence of Tah1p does not strongly affect the cell growth (Fig. 1A), the snoRNA stability (Fig. 1B), nor the rRNA processing (Fig. 1D).

Like other assembly factors, Hit1p is associated *in vivo* with precursors of U3 snoRNA

We next tested the association of Rsa1p, Hit1p, and the components of the R2TP complex, with intermediates formed in the U3 snoRNA processing pathway. In yeast, RNA polymerase II produces a precursor of U3 (pre-U3) containing one intron and an extended 3'-end (27). As described in Fig. 2A, unspliced pre-U3 and spliced pre-U3 could be detected by RT-PCR using a reverse-primer hybridizing in the 3'-extension. We used such detection system to test for the presence of U3 precursors among the RNAs that are co-purified with the TAP-tagged assembly factors. First, we observed that Nop1p, Nop56p, and Nop58p are stably associated with unspliced and spliced pre-U3 (Fig. 2B, lanes 2-4), demonstrating that the core proteins are recruited early during U3 snoRNP biogenesis. In addition, Rsa1p, Hit1p, Pih1p, Rvb1p, and Rvb2p, which are not present in the mature U3 particle, are also associated with the two U3 precursors (Fig. 2C, lanes 2-6). We concluded that Hit1p is present with Rsa1p and with the R2TP assembly machinery on the pre-RNPs. Nevertheless, Tah1p was not found associated with pre-U3s (Fig. 2D, lane 2). In accordance with the absence of apparent phenotype for the Δ*TAH1* mutant strain (Fig. 1), this result seemed to minimize the role of Tah1p in snoRNP biogenesis. However, as shown by W. Houry laboratory (55), the functional importance of Tah1p appears during the stationary growth phase, during which it stabilizes Pih1p (Fig. 2E). Thus, we performed TAP purification from cells harvested in the stationary growth phase (Fig. 2D). Our results clearly show the implication of Tah1p in U3 biogenesis since pre-U3 species are selectively co-precipitated with Tah1p in an extract prepared from cells in stationary phase (Fig. 2D, lane 4). Altogether, these data indicate that Tah1p could be selectively involved with the Rsa1p-Hit1p-R2TP assembly machinery under stress conditions.

RSA1* and *HIT1*exhibit synthetic lethality with *PIH1

To further increase our knowledge of the functional link existing between the assembly factors, we tested synthetic lethality between pairs of null alleles in the non-essential genes *RSA1*, *HIT1*, *PIH1*, and *TAH1* (Fig. 3). We crossed single knock-out haploid strains to obtain heterozygous diploid strains. Tetrads of haploid cells were produced by meiosis during sporulation, dissected, and growth of the haploid cells were tested on YPD media plates. The Δ*RSA1*-Δ*HIT1* double knock-out haploid strain did not reveal any cumulative effect in the phenotypes of the single knock-out strains, as illustrated by a comparable growth defect for the Δ*RSA1*-Δ*HIT1*, Δ*RSA1*, and Δ*HIT1* strains (Fig 4A). This result supports the observation that the lack of Rsa1p or Hit1p leads to the same phenotypic effects (Fig. 1), and that these proteins are closely linked in the same function. This function is strongly connected to Pih1p as a lethal

phenotype is observed when cells combine null mutations of *RSA1* and *PIH1* genes, or null mutations of *HIT1* and *PIH1* genes.

The stability of Rsa1p relies on presence of Hit1p

The strong functional connection that we have established between Rsa1p and Hit1p is based on a control of the stability of Rsa1p by Hit1p. We evaluated the levels of the fusion protein GFP-Rsa1p when the encoding gene is expressed in the Δ*RSA1* strain and in the double knock-out Δ*RSA1*-Δ*HIT1* strain (Fig. 4B). The fusion protein is functional as ectopic expression in the Δ*RSA1* strains restores a wild type growth phenotype (data not shown). Western-blot analysis with anti-GFP antibodies indicates that the level of GFP-Rsa1p is drastically reduced in extracts prepared from cells lacking Hit1p (Fig. 4B, compare lane 2 with lane 3). The role of Hit1p for the *in vivo* stability of Rsa1p is confirmed by restoration of Rsa1p amounts upon ectopic expression of Hit1p in strain Δ*RSA1*-Δ*HIT1* (Fig. 4B, lane 4). On the contrary, absence of Rsa1p has no strong effect on the level of Hit1p fused to the epitope derived from the Influenza protein hemagglutinin (HA) (Fig. 4B, lanes 5 to 7).

A COOH-terminal fragment of Rsa1p directly interacts with Snu13p and Hit1p

In a previous work, we showed that ectopic expression in the Δ*RSA1* strain of the Rsa1p COOH-terminal fragment, extending from Pro230 to Lys381 and named $Rsa1p_{230\text{-}381}$ allows full recovery of the wild type growth phenotype (3). We also demonstrated that this fragment interacts in a yeast two hybrid (Y2H) assay with Pih1p (Boulon et al., 2008), and that it carries the conserved PEP domain (residues 230 to 263), which constitutes the Snu13p-binding domain of Rsa1p (3) (Rothé, Back *et al*., unpublished data). Here we show that $Rsa1p_{230\text{-}381}$ interacts also with Hit1p in a Y2H assay (Fig. 5A).

To better characterize the interaction network between Snu13p, Rsa1p, Pih1p, and Hit1p, we co-expressed His_6-tagged Snu13p or Rsa1p with one or two of their potential partners in *E. coli*. Co-expressed proteins were co-purified from bacterial extracts by Immobilized Metal Ion Affinity Chromatography, fractionated by SDS-PAGE and analysed by mass spectrometry. The mass spectrometric analysis revealed that the last 6 amino acids of $Rsa1p_{230\text{-}381}$ are often degraded (data not shown). This observation led us to clone and work with a new fragment deleted of these residues and named $Rsa1p_{230\text{-}375}$. Proteomic analysis of the detected bands identified Hit1p as associated with both His_6Rsa1p and $His_6Rsa1p_{230\text{-}375}$ (Fig. 5B). These data confirm the Y2H interactions, and reflect direct physical association between Rsa1p and Hit1p. In addition, we observed that Rsa1p and Hit1p also co-purify with His_6Snu13p (Fig. 5C), indicating formation of a complex involving these 3 partners in *E. coli*.

To further characterize binary ($Rsa1p_{230\text{-}375}$ and Hit1p) or ternary (Snu13p, $Rsa1p_{230\text{-}375}$, and Hit1p) complexes and to unambiguously asses their binding stoichiometries, His_6-tagged complexes were cleaved by TEV protease, purified and analyzed by noncovalent mass spectrometry. Electrospray ionization mass spectra obtained in nondenaturing conditions of $Rsa1p_{230\text{-}375}$-Hit1p confirmed the presence of a 1/1 complex with a measured mass of 35 906 ± 1 Da (expected mass of the complex 35 905 Da) in addition to free proteins $Rsa1p_{230\text{-}375}$ (17 600 ± 1 Da ; expected mass: 17599 Da) and Hit1p (18 305 ± 1 Da in agreement with Hit1p sequence after deletion of the N-terminus methionine and presence of 1 Zn^{2+} ion ; 18 245 + 63 = 18308 Da) (Fig. 5C). Similarly, noncovalent mass spectrometry analysis of the complex involving Snu13p, $Rsa1p_{230\text{-}375,}$ and Hit1p confirmed the presence of a 1/1/1 complex with a measured mass of 49 342 ± 2 Da (expected mass of the complex 49 344 Da) in addition to free Snu13p (13 861 ± 1 Da ; expected mass: 13 860 Da) and the 1/1

heterodimer Rsa1p$_{230-375}$-Hit1p (35 483 ± 5 Da ; expected mass : 35 484 Da) (Fig. 5D).
Finally, we observed by EMSA that incubation of the purified protein complexes with radiolabeled snoRNAs containing the K-turn RNA motif leads to formation of a single high molecular weigh complex displaying a slower mobility than the Snu13p-Rsa1p-RNA complex (Fig. 5D). Hence, presence of Hit1p does not impair the capacity of the Snu13p-Rsa1p heterodimer to interact with RNA.

Core protein Nop58p binds heterodimers Rsa1p-Hit1p and Tah1p-Pih1p

As presented in table 1, only small amounts of C/D core proteins were found in association with assembly factors, suggesting that the latter are transiently associated to pre-RNPs in normal growth conditions. It was previously established that Pih1p interacts with Nop58p (15), indicating that the interaction between this assembly factor and this core protein is intrinsically stable. In our precedent work, we showed a Y2H interaction between Rsa1p and the Nop58 core protein (3). Here, by using the purified recombinant His$_6$-tagged complexes in pull-down assays with the ^{35}S radiolabelled Nop58p produced in *E. coli* S30 extract, we observed that Nop58 co-purifies as expected with His$_6$Tah1p-Pih1p (Fig. 6, lane 2), and also with His$_6$Snu13p-Rsa1p$_{230-375}$ (Fig. 6, lane 3), His$_6$Rsa1p$_{230-375}$-Hit1p (Fig. 6, lane 4), and His$_6$Snu13p-Rsa1p$_{230-375}$-Hit1p (Fig. 6, lane 5), but not in absence of the His$_6$-tagged complexes (Fig. 6, lane 1). These *in vitro* binding assays confirm the direct association of Rsa1p and Nop58p. Hence, Nop58p associates directly not only with Pih1p but also with Rsa1p in absence of RNA. The presence of Tah1p bound to Pih1p, or Hit1p bound to Rsa1 does not impair the interactions Rsa1p-Nop58 or Pih1p-Nop58.

DISCUSSION

The function of Rsa1p relies on Hit1p

In this work, we have identified Hit1p (High Temperature growth, YJR055W) as a novel protein which function is related to the assembly of C/D snoRNPs. At first, we found that Hit1p and the C/D snoRNP assembly factor Rsa1p co-purify of yeast extracts (Table 1), confirming the interaction between the two proteins in a global yeast two hybrid screen (20). Here, we show that Hit1p has features in common with Rsa1p and the other R2TP assembly factors: it is associated *in vivo* with U3 snoRNA precursors, and is required for accumulation of the C/D snoRNAs. Our data indicate that the presence of Hit1p on the pre-snoRNPs relies on a direct interaction with Rsa1p. Moreover, we show that at least one of the major function of Hit1p is to stabilize Rsa1p in yeast (Fig. 6). The loss of such activity in the single-gene knock-out mutant strain Δ*HIT1* gives rise to a phenotype similar to the Δ*RSA1* strain, and deletion of the two *HIT1* and *RSA1* genes in a same strain does not lead to cumulative effects of the mutations (Fig. 1). As Hit1p was not found to interact with other characterized C/D snoRNPs assembly factors, nor with any of their core proteins, we propose that Hit1p intervenes during the assembly of C/D snoRNP through its stability effect on Rsa1p.

The specificity of Hit1p for Rsa1p relies on a direct association. Indeed, we demonstrate that the two proteins form an heterodimer when they are co-expressed as recombinant proteins in *E. coli.* The production of Rsa1p is not efficient in *E. coli* as most of the proteins accumulate in the insoluble fraction. The solubility is markedly enhanced when Rsa1p is co-expressed with its interacting proteins Snu13 (Rothé, Back *et al*, unpublished data) or Hit1 (Fig. 5B and C). This is reminiscent of the feature of the Pih1p C/D snoRNP assembly factor which is efficiently produced and purified as a recombinant protein only in the presence of its interaction partner Tah1p (13, 55). Ectopic expression of various fragments of Rsa1p in strain Δ*RSA1* indicated that the last 152 amino acids of the COOH-terminal end (Rsa1p$_{230-381}$) represent the minimal fragment able to restore normal growth ((3), and data not shown). All the physical

interactions that we have characterized in this work were obtained with the C-terminal part of Rsa1p which constitutes a platform for several non-exclusive interactions. Indeed, we show that in presence of Hit1p, the C/D snoRNAs binding protein Snu13p can still interact with the PEP motif located at the N-terminal end of Rsa1p$_{230-375}$, and that the formed complex Snu13p-Rsa1p$_{230-375}$-Hit1p can interact with Nop58, one of the core box C/D snoRNP-associated proteins.

Features of the Hit1 protein

The Hit1 protein contains a zinc finger domain, which represents the prototype of the HIT-zinc finger domain (Zf-HIT) family (22). The 3D solution structure of the Zf-HIT domain present in human protein ZNHIT2 (protein FON) classified it as a new member of the treble clef domain (17). The function of the domain remains obscure and we have no evidence whether it is required for the association of Hit1p with Rsa1p. Interestingly, the Zf-HIT domain is found in nuclear proteins involved in gene regulation and chromatin remodelling, and is also present in protein Bcd1 (Box CD 1 protein), which controls the box C/D snoRNAs stability in yeast (19, 41). The human homologue of Bcd1p was shown to be involved in the C/D RNP assembly process by interacting with NUFIP, the functional homologue of Rsa1p (33). Conservation of a similar domain in the two assembly factors Hit1p and Bcd1p raises the question of whether the two domains have similar or related functions.

TRIP3 (Thyroid Hormone Receptor Interacting protein 3), the human homologue of Hit1p, is involved in multiple hormone-dependent regulation processes, e.g. the regulation of retinoic acids receptor in neuroblastoma differentiation (30), the regulation of the HNF-4α (Hepatocyte Nuclear Factor 4α) transcription activity involved in glucose metabolism (21), and the regulation of PPARγ-mediated adipocyte differentiation (26). The precise role played by TRIP3 in these mechanisms is currently unclear. Work performed by the laboratory of E. Bertrand shows that TRIP3 participate in C/D snoRNPs biogenesis (personal communication). Furthermore, a recent study revealed a decrease of TRIP3 expression in renal cell carcinoma, and proposed TRIP3 as a new tumor suppressor gene (4). Thus, this property could rely on its capacity to control cell growth through its implication in various essential processes like nuclear-receptor dependant transcription and/or by promoting snoRNP biogenesis and ribosome synthesis.

Rsa1p-Hit1p represents a module for box C/D snoRNP assembly

Based on our data of the co-purification assays showing no accumulation of complexes composed of Rsa1p, Hit1p, Rvb1p, Rvb2p, Tah1p, and Pih1p (Table 1), we propose that in yeast, these assembly factors mainly exist as two sub-complexes. One module would be composed of the heterodimer Rsa1p-Hit1p, and a second one would correspond to the R2TP complex (Rvb1p-Rvb2p-Tah1p-Pih1p). Nevertheless, as we showed that Rsa1p and Pih1p interact in Y2H assays (3), such interaction may promote transient association between the two modules during snoRNPs assembly. The importance of Rsa1p and Pih1p for the stability of C/D snoRNA *in vivo* is well established (3, 55). In addition, lethality of the Δ*RSA1*Δ*PIH1* mutant (Fig. 3) indicates that Rsa1p and Pih1 represent the cornerstones of the assembly machinery.

Although Nop58p was shown to interact with Pih1p, our Y2H assays (3) and our *in vitro* binding assay (Fig. 6) show an association of Nop58 with the purified Rsa1p$_{230-375}$-Hit1p and the Snu13p-Rsa1p$_{230-375}$-Hit1p complex. Thus, at some stage of the C/D snoRNP biogenesis, the stable association of Nop58p with Pih1p could be destabilized in favour of an interaction with a pre-RNP composed of Snu13p-Rsa1p-Hit1p. Proteins of the two modules are present on pre-U3 snoRNPs (Fig. 2). Hence at this stage, the two modules can transiently cooperate for RNP formation: Rsa1p can

facilitate recruitment of the Nop proteins on the RNA-Snu13p pre-RNP (3, 33, 34), and R2TP can operate a conformation remodelling required for the progression of the RNP assembly by bringing helicase activity of the Rvb1p-Rvb2p proteins (3, 34).

The precise events occurring during early steps of snoRNP assembly are not known. The initial interaction established between the box C/D RNA motif and proteins Snu13p/15,5K in Eukaryotes or L7Ae in Archaea, is required for subsequent assembly of the other proteins (5, 38, 42, 47, 50). An unresolved question is whether the K-turn is initially bound by the free form of Snu13p and next by the Rsa1p-Hit1p pair or directly by a Snu13p-Rsa1p-Hit1p complex. This last scenario is possible because such heterotrimeric complex can be purified (Fig. 5C) and because association of Rsa1p and Hit1p with Snu13p does not impair the RNA binding capacity of Snu13p (Fig. 5D). In addition, given that Snu13p does not possess nuclear localization signal (NLS), one possibility would be that Rsa1p or Hit1p serves as protein cargo for the import of Snu13p in the nucleus. Interestingly, a NLS signal is present in the C-terminal region of Rsa1p. Hence, Snu13p-Rsa1p-Hit1p may represent a module for nuclear import and RNA binding.

FIGURE LEGENDS

Table 1: Identification of Hit1p as an interacting partner of Rsa1p. List of identified proteins by proteomic analysis after tandem affinity experiments of TAP-Rsa1p, TAP-Hit1p, TAP-Pih1p, and an untagged negative control (No TAP).

Figure 1: Phenotypic analysis of the single null mutant strains.
(A) Growth of the strains By4741 WT, and of its mutants Δ*RSA1*, Δ*HIT1*, Δ*PIH*, and Δ*TAH1* was monitored overtime by measuring the optical density of the cultures at 600 nm (OD_{600}). The doubling times were calculated (G), and the growth defect of the mutant strains were expressed as a percentage of the doubling time of the wild-type strain designated. (B) Rsa1p, Hit1p and Pih1p are required for accumulation of C/D snoRNAs. Northern blot analysis of snoRNAs steady-state levels in the null mutant strains. Extracted RNAs were separated by 7 % acrylamide/urea gels and transferred on Zeta-Probe membrane (Biorad®). Various RNAs (C/D snoRNAs, H/ACA snoRNAs and snRNAs) were detected by antisense oligonucleotide probes.
(C) Analysis by Northern blot of the processing of a modified U3Δ2,3,4 gene expressed in strains lacking Rsa1p (Δ*RSA1*) and in strain lacking Hit1p (Δ*HIT1*). This U3 gene accumulates pre-U3Δ2,3,4 precursors.
(D) Rsa1p, Hit1p, and Pih1p are required for pre-rRNA processing. Analysis of pre-rRNA processing by Northern blot. Cells were harvested in log phase (OD_{600} = 0.8-1) and total RNAs were extracted. Equal amounts of RNA (5 µg) were fractionated on 1.2 % glyoxal/DMSO agarose gels and transferred to a Zeta-Probe membrane (Biorad®). The membranes were consecutively hybridized with radio-labeled antisense oligonucleotides probes indicated on the scheme showing various intermediates of the pre-rRNA processing pathway (on the right).

Figure 2: The core proteins and assembly factors are associated with U3 precursors.
(A) Scheme of the pre-U3 precursor with its intron and extended 3' end. Positions of the cleavages occurring in this extension are indicated. Synthesis of cDNAs were primed by oligonucleotide Rev (broken arrow) interacting partially with the U3 3' end extension. Regions complementary to the primers PCRa, PCRb, and 3'PCR used for the PCR amplifications are indicated by dotted arrows.
(B, C, D) Extracts prepared from various TAP-tagged strains in log phase: TAP-*NOP1*, TAP-*NOP56* and TAP-*NOP58* (B), TAP-*RSA1*, TAP-*HIT1*, TAP-*PIH1*, TAP-*RVB1* and TAP-*RVB2* (C), TAP-*TAH1* both in log- and stationary phase (D), were purified on IgG sepharose beads and analyzed by RT-PCR with primers specific for the unspliced (unspl.) and spliced (spl.)

U3 precursors carrying the full size of the 3' extension.
An untagged strain was used as negative control. Inputs correspond to signals obtained from 10 % of cellular extracts used for IPs.
(E) Tah1p is required for accumulation of Pih1p in stationary phase. Analysis by Western-blot of Pih1p and Tah1p levels in WT, Δ*PIH1* and Δ*TAH1* strains.

Figure 3: Strains lacking both Rsa1p and Pih1p, or Hit1p and Pih1p are not viable. Diploid strains generated from crossbreeding between single null mutants Δ*RSA1*, Δ*PIH1*, or Δ*HIT1* of haploids strains By4741 (MATa) and By4742 (MATα) were placed under starvation for meiosis and spore formation. The obtained tetrads were dissected with a micromanipulator and haploid spores were deposited on YPD plates and incubated 1 week at 30 °C.

Figure 4: Hit1p is required for accumulation of Rsa1p
(A) Growth defect of the single and double null mutants for the *RSA1* and *HIT1* genes was analyzed as in Fig. 1A.
(B) Western blot analysis of equal amounts of total cell lysates from the single and double null mutants transformed with a HA-Hit1p or GFP-Rsa1p expression vector. The tagged proteins were detected by blotting the membrane with α-HA and α-GFP antibodies. Protein PMA1 used as a loading control was detected using specific α-PMA1 antibodies.

Figure 5: Rsa1p and Hit1p form a stable heterodimer.
(A) Hit1p interacts in Y2H assay with Rsa1p. Rsa1p and the COOH-terminal fragment ($Rsa1p_{230-381}$; amino acids 230 to 381) fused to the Gal4 activation domain interact with Hit1p fused to the Gal4 DNA binding domain. The interaction allows transcription activation of the *TRP1* reporter gene. Diploids carrying the plasmids expressing the two protein partners were grown on medium lacking Leu, His, and Trp.
(B, C) Co-expression in *E. coli* and co-purification assays. His_6-tagged Rsa1p and $Rsa1p_{230-375}$ proteins were co-expressed with Hit1p (B). His_6-tagged protein Snu13 was co-expressed with $Rsa1p_{230-375}$ and Hit1p (C). His-pull down assays from the crude cellular extracts were performed by IMAC. The proteins retained on cobalt resin were resolved by SDS-PAGE and stained with Coomassie blue (panels on the left). The ladder (in kDa) is indicated on the right. The retained complexes were eluted by site-specific cleavage of the His_6-tag with PreScission protease and the protein analyzed by mass spectrometry.
(D, E) Noncovalent mass spectrometry detection of the 1/1 $Rsa1p_{230-375}$-Hit1p and 1/1/1 Snu13p-$Rsa1p_{230-375}$-Hit1p complexes. NanoESI mass spectra were obtained under nondenaturing conditions for $Rsa1p_{230-375}$-Hit1p (10 µM in 500 mM ammonium acetate pH 7.5) (Vc = 140 V; Pi = 6 mbar) (D) and for Snu13p-$Rsa1p_{230-375}$-Hit1p (10 µM in 1M ammonium acetate pH 7.5) (Vc = 80 V; Pi = 6 mbar) (E).
(F) The trimeric complex His_6Snu13p-$Rsa1p_{230-375}$-Hit1p interacts *in vitro* with RNAs containing a K-turn RNA motif. Formation of the proteins-RNA complexes was assessed EMSA. Radiolabeled U3Δ2.3.4 and U14 RNAs containing K-turn RNA motif bound by Snu13p were incubated with various combinations of recombinant proteins and purified complexes obtained by co-expression in *E. coli*: His_6Snu13p, His_6Snu13p-$Rsa1p_{230-375}$, or the trimeric complex His_6Snu13p-$Rsa1_{230-375}$-Hit1p. The formed complexes were resolved by electrophoresis on 6 % non denaturing polyacrylamide gels.

Figure 6: Protein Nop58 interacts with purified recombinant complexes His_6Tah1p-Pih1p and His_6Rsa1p-Hit1p in His pull-down experiments. Nop58p was translated *in vitro* in bacterial S30 lysate in presence of ^3H methionine and assayed for binding to purified His_6Tah1p-Pih1p, $His_6Rsa1p_{230-375}$-Hit1p, and His_6Snu13p-$Rsa1p_{230-375}$-Hit1p.

Input represents 10% of total labeled protein used in interaction assay.

REFERENCES

1. **Allmang, C., P. Carbon, and A. Krol.** 2002. The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. RNA **8:**1308-18.
2. **Atzorn, V., P. Fragapane, and T. Kiss.** 2004. U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. Mol Cell Biol **24:**1769-78.
3. **Boulon, S., N. Marmier-Gourrier, B. Pradet-Balade, L. Wurth, C. Verheggen, B. E. Jady, B. Rothe, C. Pescia, M. C. Robert, T. Kiss, B. Bardoni, A. Krol, C. Branlant, C. Allmang, E. Bertrand, and B. Charpentier.** 2008. The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. J Cell Biol **180:**579-95.
4. **Brito, G. C., A. A. Fachel, A. L. Vettore, G. M. Vignal, E. R. Gimba, F. S. Campos, M. A. Barcinski, S. Verjovski-Almeida, and E. M. Reis.** 2008. Identification of protein-coding and intronic noncoding RNAs down-regulated in clear cell renal carcinoma. Mol Carcinog **47:**757-67.
5. **Cahill, N. M., K. Friend, W. Speckmann, Z. H. Li, R. M. Terns, M. P. Terns, and J. A. Steitz.** 2002. Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. EMBO J **21:**3816-28.
6. **Charpentier, B., S. Muller, and C. Branlant.** 2005. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. Nucleic Acids Res **33:**3133-44.
7. **Clery, A., V. Bourguignon-Igel, C. Allmang, A. Krol, and C. Branlant.** 2007. An improved definition of the RNA-binding specificity of SECIS-binding protein 2, an essential component of the selenocysteine incorporation machinery. Nucleic Acids Res **35:**1868-84.
8. **Clery, A., V. Senty-Segault, F. Leclerc, H. A. Raue, and C. Branlant.** 2007. Analysis of sequence and structural features that identify the B/C motif of U3 small nucleolar RNA as the recognition site for the Snu13p-Rrp9p protein pair. Mol Cell Biol **27:**1191-206.
9. **Copeland, P. R., J. E. Fletcher, B. A. Carlson, D. L. Hatfield, and D. M. Driscoll.** 2000. A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. EMBO J **19:**306-14.
10. **Darzacq, X., N. Kittur, S. Roy, Y. Shav-Tal, R. H. Singer, and U. T. Meier.** 2006. Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. J Cell Biol **173:**207-18.
11. **Dez, C., J. Noaillac-Depeyre, M. Caizergues-Ferrer, and Y. Henry.** 2002. Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs. Mol Cell Biol **22:**7053-65.
12. **Diebold, M. L., S. Fribourg, M. Koch, T. Metzger, and C. Romier.** 2011. Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. J Struct Biol **175:**178-88.
13. **Eckert, K., J. M. Saliou, L. Monlezun, A. Vigouroux, N. Atmane, C. Caillat, S. Quevillon-Cheruel, K. Madiona, M. Nicaise, S. Lazereg, A. Van Dorsselaer, S. Sanglier-Cianferani, P. Meyer, and S. Morera.** 2010. The Pih1-Tah1 cochaperone complex inhibits Hsp90 molecular chaperone ATPase activity. J Biol Chem **285:**31304-12.
14. **Fatica, A., M. Dlakic, and D. Tollervey.** 2002. Naf1 p is a box H/ACA snoRNP assembly factor. RNA **8:**1502-14.
15. **Gonzales, F. A., N. I. Zanchin, J. S. Luz, and C. C. Oliveira.** 2005. Characterization of Saccharomyces cerevisiae Nop17p, a novel Nop58p-interacting protein that is involved in Pre-rRNA processing. J Mol Biol **346:**437-55.
16. **Granneman, S., J. Vogelzangs, R. Luhrmann, W. J. van Venrooij, G. J. Pruijn, and N. J. Watkins.** 2004. Role of pre-rRNA base pairing and 80S complex formation in subnucleolar localization of the U3 snoRNP. Mol Cell Biol **24:**8600-10.
17. **He, F., T. Umehara, K. Tsuda, M. Inoue, T. Kigawa, T. Matsuda, T. Yabuki, M. Aoki, E. Seki, T. Terada, M. Shirouzu, A. Tanaka, S. Sugano, Y. Muto, and S. Yokoyama.** 2007. Solution structure of the zinc finger HIT domain in protein FON. Protein Sci **16:**1577-87.
18. **Henras, A., Y. Henry, C. Bousquet-Antonelli, J. Noaillac-Depeyre, J. P. Gelugne, and M. Caizergues-Ferrer.** 1998. Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. EMBO J **17:**7078-90.
19. **Hiley, S. L., T. Babak, and T. R. Hughes.** 2005. Global analysis of yeast RNA processing identifies new targets of RNase III and uncovers a link between tRNA 5' end

processing and tRNA splicing. Nucleic Acids Res **33:**3048-56.

20. **Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki.** 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A **98:**4569-74.
21. **Iwahashi, H., K. Yamagata, I. Yoshiuchi, J. Terasaki, Q. Yang, K. Fukui, A. Ihara, Q. Zhu, T. Asakura, Y. Cao, A. Imagawa, M. Namba, T. Hanafusa, J. Miyagawa, and Y. Matsuzawa.** 2002. Thyroid hormone receptor interacting protein 3 (trip3) is a novel coactivator of hepatocyte nuclear factor-4alpha. Diabetes **51:**910-4.
22. **Kawakami, K., B. K. Shafer, D. J. Garfinkel, J. N. Strathern, and Y. Nakamura.** 1992. Ty element-induced temperature-sensitive mutations of Saccharomyces cerevisiae. Genetics **131:**821-32.
23. **Klein, D. J., T. M. Schmeing, P. B. Moore, and T. A. Steitz.** 2001. The kink-turn: a new RNA secondary structure motif. EMBO J **20:**4214-21.
24. **Kolb, S. J., D. J. Battle, and G. Dreyfuss.** 2007. Molecular functions of the SMN complex. J Child Neurol **22:**990-4.
25. **Koonin, E. V., P. Bork, and C. Sander.** 1994. A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? Nucleic Acids Res **22:**2166-7.
26. **Koppen, A., R. Houtman, D. Pijnenburg, E. H. Jeninga, R. Ruijtenbeek, and E. Kalkhoven.** 2009. Nuclear receptor-coregulator interaction profiling identifies TRIP3 as a novel peroxisome proliferator-activated receptor gamma cofactor. Mol Cell Proteomics **8:**2212-26.
27. **Kufel, J., C. Allmang, G. Chanfreau, E. Petfalski, D. L. Lafontaine, and D. Tollervey.** 2000. Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding. Mol Cell Biol **20:**5415-24.
28. **Lafontaine, D. L., and D. Tollervey.** 1999. Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. RNA **5:**455-67.
29. **Lafontaine, D. L., and D. Tollervey.** 2000. Synthesis and assembly of the box C+D small nucleolar RNPs. Mol Cell Biol **20:**2650-9.
30. **Lovat, P. E., M. Annicchiarico-Petruzzelli, M. Corazzari, M. G. Dobson, A. J. Malcolm, A. D. Pearson, G. Melino, and C. P. Redfern.** 1999. Differential effects of retinoic acid isomers on the expression of nuclear receptor co-regulators in neuroblastoma. FEBS Lett **445:**415-9.
31. **Marmier-Gourrier, N., A. Clery, V. Senty-Segault, B. Charpentier, F. Schlotter, F. Leclerc, R. Fournier, and C. Branlant.** 2003. A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. RNA **9:**821-38.
32. **Matera, A. G., R. M. Terns, and M. P. Terns.** 2007. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol **8:**209-20.
33. **McKeegan, K. S., C. M. Debieux, S. Boulon, E. Bertrand, and N. J. Watkins.** 2007. A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. Mol Cell Biol **27:**6782-93.
34. **McKeegan, K. S., C. M. Debieux, and N. J. Watkins.** 2009. Evidence that the AAA+ proteins TIP48 and TIP49 bridge interactions between 15.5K and the related NOP56 and NOP58 proteins during box C/D snoRNP biogenesis. Mol Cell Biol **29:**4971-81.
35. **Miguet, L., G. Bechade, L. Fornecker, E. Zink, C. Felden, C. Gervais, R. Herbrecht, A. Van Dorsselaer, L. Mauvieux, and S. Sanglier-Cianferani.** 2009. Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis. J Proteome Res **8:**3346-54.
36. **Morrissey, J. P., and D. Tollervey.** 1993. Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. Mol Cell Biol **13:**2469-77.
37. **Nottrott, S., K. Hartmuth, P. Fabrizio, H. Urlaub, I. Vidovic, R. Ficner, and R. Luhrmann.** 1999. Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. EMBO J **18:**6119-33.
38. **Omer, A. D., S. Ziesche, H. Ebhardt, and P. P. Dennis.** 2002. In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. Proc Natl Acad Sci U S A **99:**5289-94.
39. **Patel, S. B., and M. Bellini.** 2008. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. Nucleic Acids Res **36:**6482-93.
40. **Pellizzoni, L.** 2007. Chaperoning ribonucleoprotein biogenesis in health and disease. EMBO Rep **8:**340-5.
41. **Peng, W. T., M. D. Robinson, S. Mnaimneh, N. J. Krogan, G. Cagney, Q. Morris, A. P. Davierwala, J. Grigull, X. Yang, W. Zhang, N. Mitsakakis, O. W. Ryan, N. Datta, V. Jojic, C. Pal, V.**

Canadien, D. Richards, B. Beattie, L. F. Wu, S. J. Altschuler, S. Roweis, B. J. Frey, A. Emili, J. F. Greenblatt, and T. R. Hughes. 2003. A panoramic view of yeast noncoding RNA processing. Cell **113:**919-33.

42. **Rashid, R., M. Aittaleb, Q. Chen, K. Spiegel, B. Demeler, and H. Li.** 2003. Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles. J Mol Biol **333:**295-306.
43. **Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin.** 1999. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol **17:**1030-2.
44. **Romier, C., M. Ben Jelloul, S. Albeck, G. Buchwald, D. Busso, P. H. Celie, E. Christodoulou, V. De Marco, S. van Gerwen, P. Knipscheer, J. H. Lebbink, V. Notenboom, A. Poterszman, N. Rochel, S. X. Cohen, T. Unger, J. L. Sussman, D. Moras, T. K. Sixma, and A. Perrakis.** 2006. Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. Acta Crystallogr D Biol Crystallogr **62:**1232-42.
45. **Sanglier, S., C. Atmanene, G. Chevreux, and A. V. Dorsselaer.** 2008. Nondenaturing mass spectrometry to study noncovalent protein/protein and protein/ligand complexes: technical aspects and application to the determination of binding stoichiometries. Methods Mol Biol **484:**217-43.
46. **Schultz, A., S. Nottrott, N. J. Watkins, and R. Luhrmann.** 2006. Protein-protein and protein-RNA contacts both contribute to the 15.5K-mediated assembly of the U4/U6 snRNP and the box C/D snoRNPs. Mol Cell Biol **26:**5146-54.
47. **Tran, E. J., X. Zhang, and E. S. Maxwell.** 2003. Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. EMBO J **22:**3930-40.
48. **Vidovic, I., S. Nottrott, K. Hartmuth, R. Luhrmann, and R. Ficner.** 2000. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. Mol Cell **6:**1331-42.
49. **Wang, C., and U. T. Meier.** 2004. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. EMBO J **23:**1857-67.
50. **Watkins, N. J., A. Dickmanns, and R. Luhrmann.** 2002. Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. Mol Cell Biol **22:**8342-52.
51. **Watkins, N. J., V. Segault, B. Charpentier, S. Nottrott, P. Fabrizio, A. Bachi, M. Wilm, M. Rosbash, C. Branlant, and R. Luhrmann.** 2000. A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. Cell **103:**457-66.
52. **Yang, P. K., G. Rotondo, T. Porras, P. Legrain, and G. Chanfreau.** 2002. The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis. J Biol Chem **277:**45235-42.
53. **Yong, J., L. Wan, and G. Dreyfuss.** 2004. Why do cells need an assembly machine for RNA-protein complexes? Trends Cell Biol **14:**226-32.
54. **Zhao, R., and W. A. Houry.** 2005. Hsp90: a chaperone for protein folding and gene regulation. Biochem Cell Biol **83:**703-10.
55. **Zhao, R., Y. Kakihara, A. Gribun, J. Huen, G. Yang, M. Khanna, M. Costanzo, R. L. Brost, C. Boone, T. R. Hughes, C. M. Yip, and W. A. Houry.** 2008. Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. J Cell Biol **180:**563-78.

FUNDING

This work was supported by the Agence Nationale de la Recherche [ANR-06-BLAN-0208], the ARC, the french Ministère de l'Enseignement Supérieur et de la Recherche [doctoral fellowships to B.R. and R.B.]; the Centre National de la Recherche Scientifique ; the Pôle de Recherche Scientifique et Technologique "Bioingénierie" of Région lorraine, the University of Strasbourg (UdS), the Fondation pour la Recherche Médicale and the Region Alsace. J.M.S. acknowledges the Sidaction foundation financial support.

	Protein name	Accession number	Molecular weight (Da)	No TAP		TAP-*PIH1*		TAP-*RSA1*		TAP-*HIT1*	
				Unique pep (% cover)	Spectral Count	Unique pep (% cover)	Spectral Count	Unique pep (% cover)	Spectral Count	Unique pep (% cover)	Spectral Count
	Rsa1p	Q08932	43 981	-	-	-	-	10 (35,4)	81	10 (32,0)	77
	Hit1p	P46973	18 358	-	-	2 (20,7)	2	7 (54,9)	31	7 (54,9)	49
Box C/D core protein	Snu13p	P39990	13 551	-	-	2 (36,5)	6	2 (36,5)	9	2 (36,5)	8
	Nop1p	P15646	34 195	2 (10,4)	4	4 (21,4)	5	5 (24,5)	9	6 (27,8)	9
	Nop56p	Q12460	56 849	-	-	-	-	5 (20,4)	7	-	-
	Nop58p	A6ZPE5	57 176	-	-	2 (6,8)	3	-	-	-	-
R2TP protein	Pih1p	P38768	39 503	-	-	13 (51,7)	58	-	-	-	-
	Tah1p	P25638	12 493	-	-	4 (69,4)	33	-	-	-	-
	Rvb1p	Q03940	50 436	-	-	16 (47,7)	25	-	-	-	-
	Rvb2p	Q12464	51 595	-	-	17 (46,9)	33	-	-	-	-

Table 1

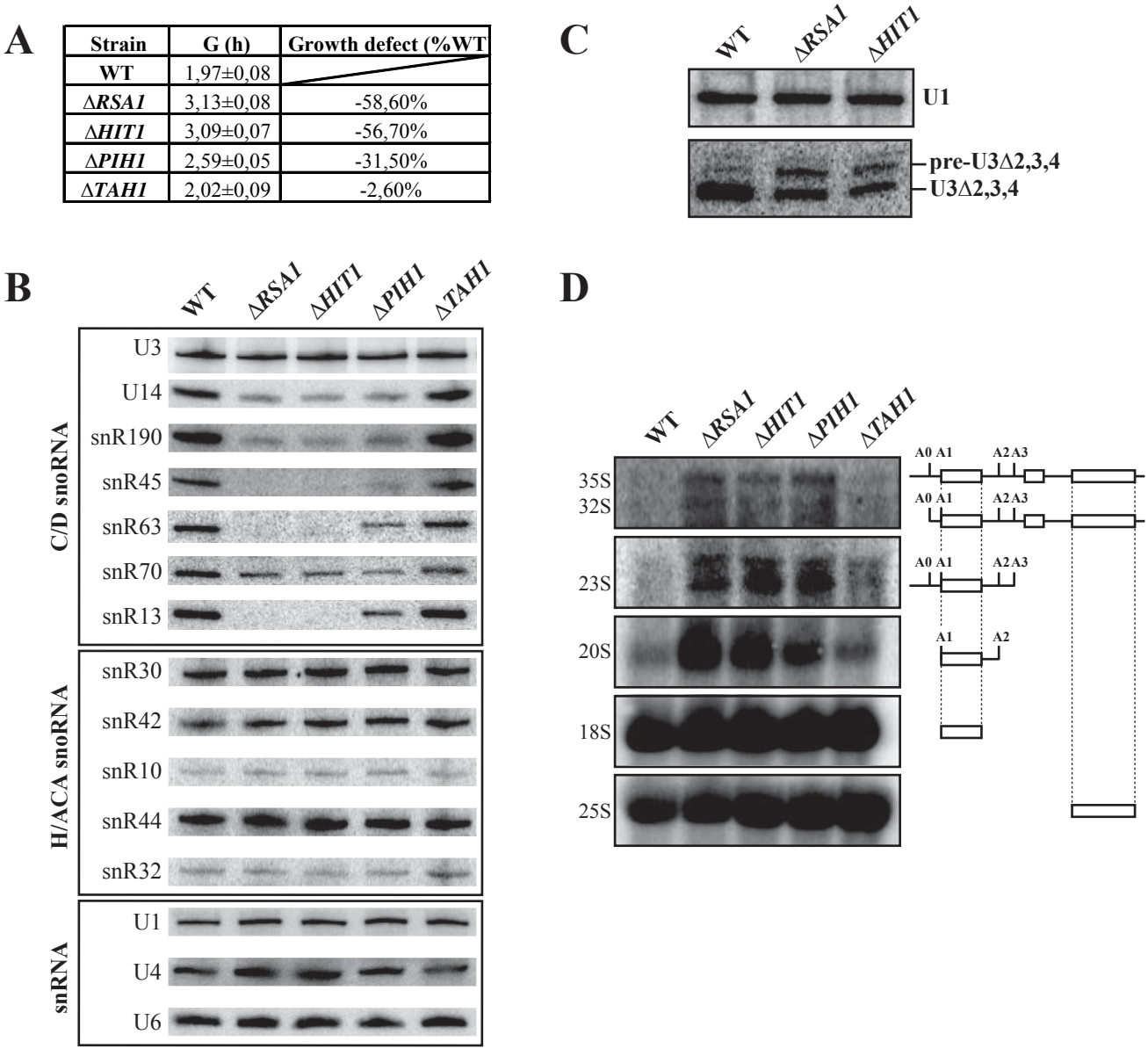


Figure 1

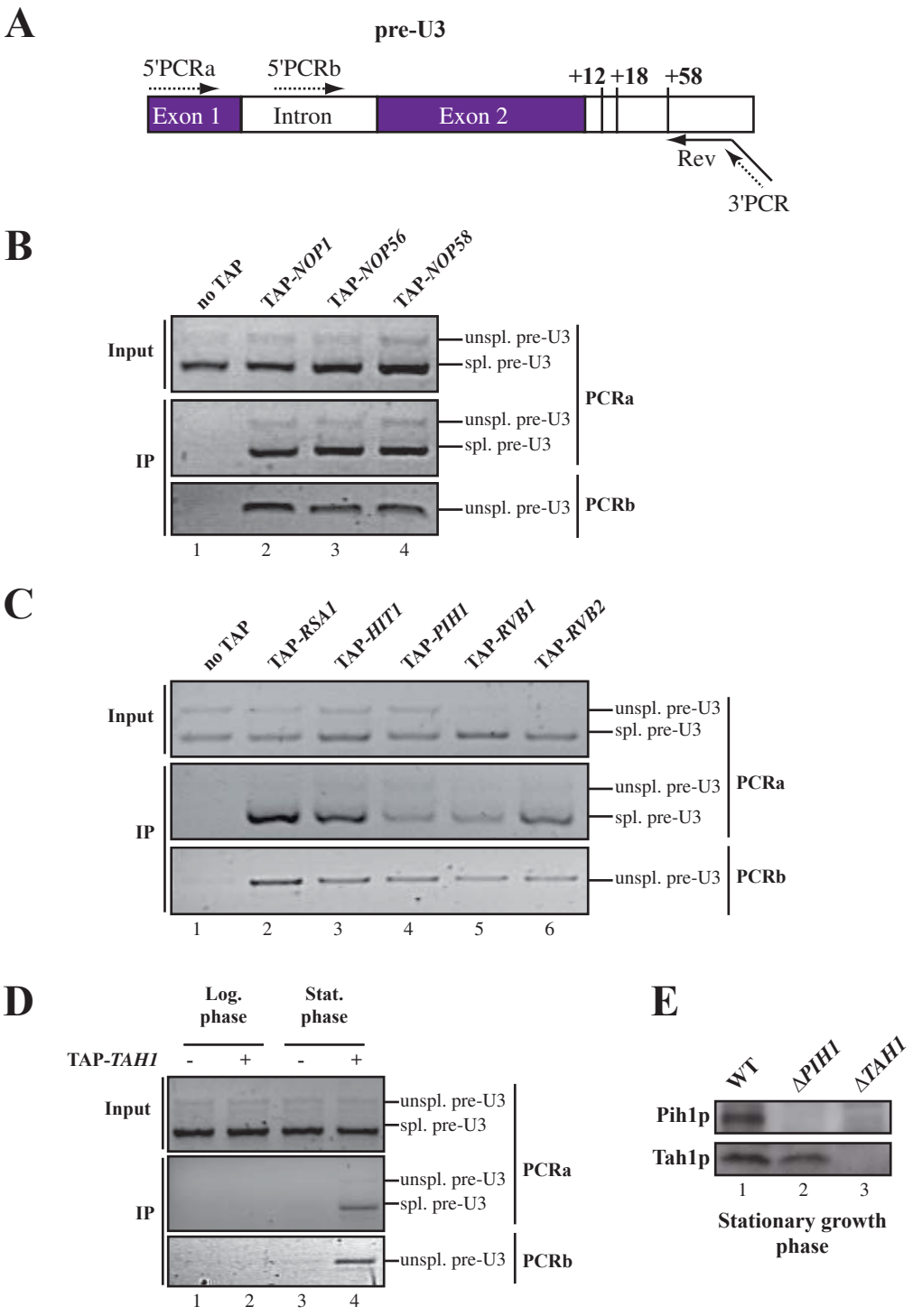

A
pre-U3
5'PCRa
5'PCRb
+12 +18 +58
Exon 1
Intron
Exon 2
Rev
3'PCR
B
no TAP
TAP-NOP1
TAP-NOP56
TAP-NOP58
Input
IP
unspl. pre-U3
spl. pre-U3
PCRa
PCRb
1 2 3 4
C
no TAP
TAP-RSA1
TAP-HIT1
TAP-PIH1
TAP-RVB1
TAP-RVB2
Input
IP
unspl. pre-U3
spl. pre-U3
PCRa
PCRb
1 2 3 4 5 6
D
Log. phase
Stat. phase
TAP-TAH1
- + - +
Input
IP
unspl. pre-U3
spl. pre-U3
PCRa
PCRb
1 2 3 4
E
WT
ΔPIH1
ΔTAH1
Pih1p
Tah1p
1 2 3
Stationary growth phase

Figure 2

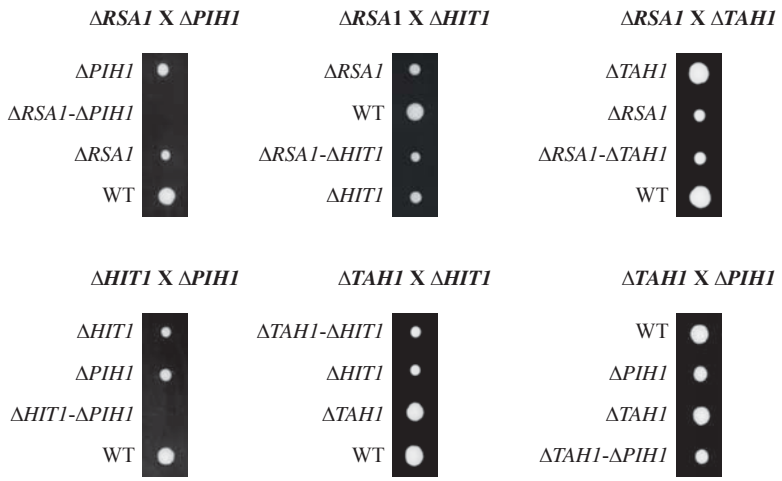


Figure 3

A

Strain	G (h)	Growth defect (%WT)
WT	1,97±0,08	
Δ*RSA1*	3,13±0,08	58,60%
Δ*HIT1*	3,09±0,07	56,70%
Δ*RSA1*-Δ*HIT1*	3,21±0,08	62,70%

B

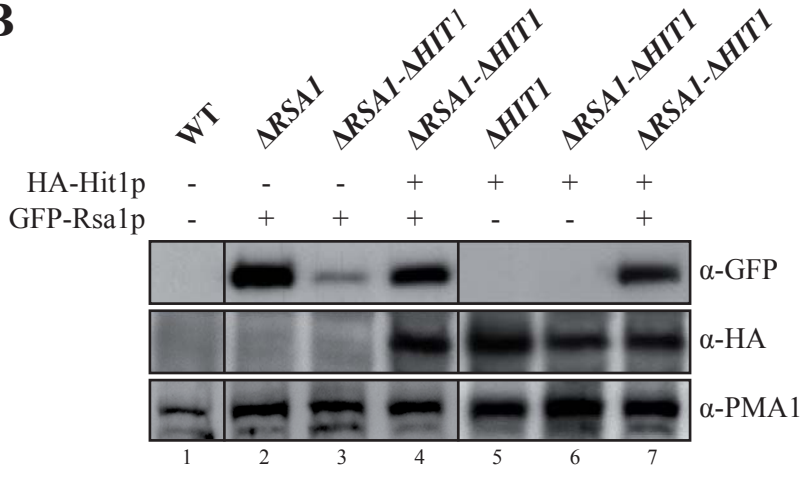


Figure 4

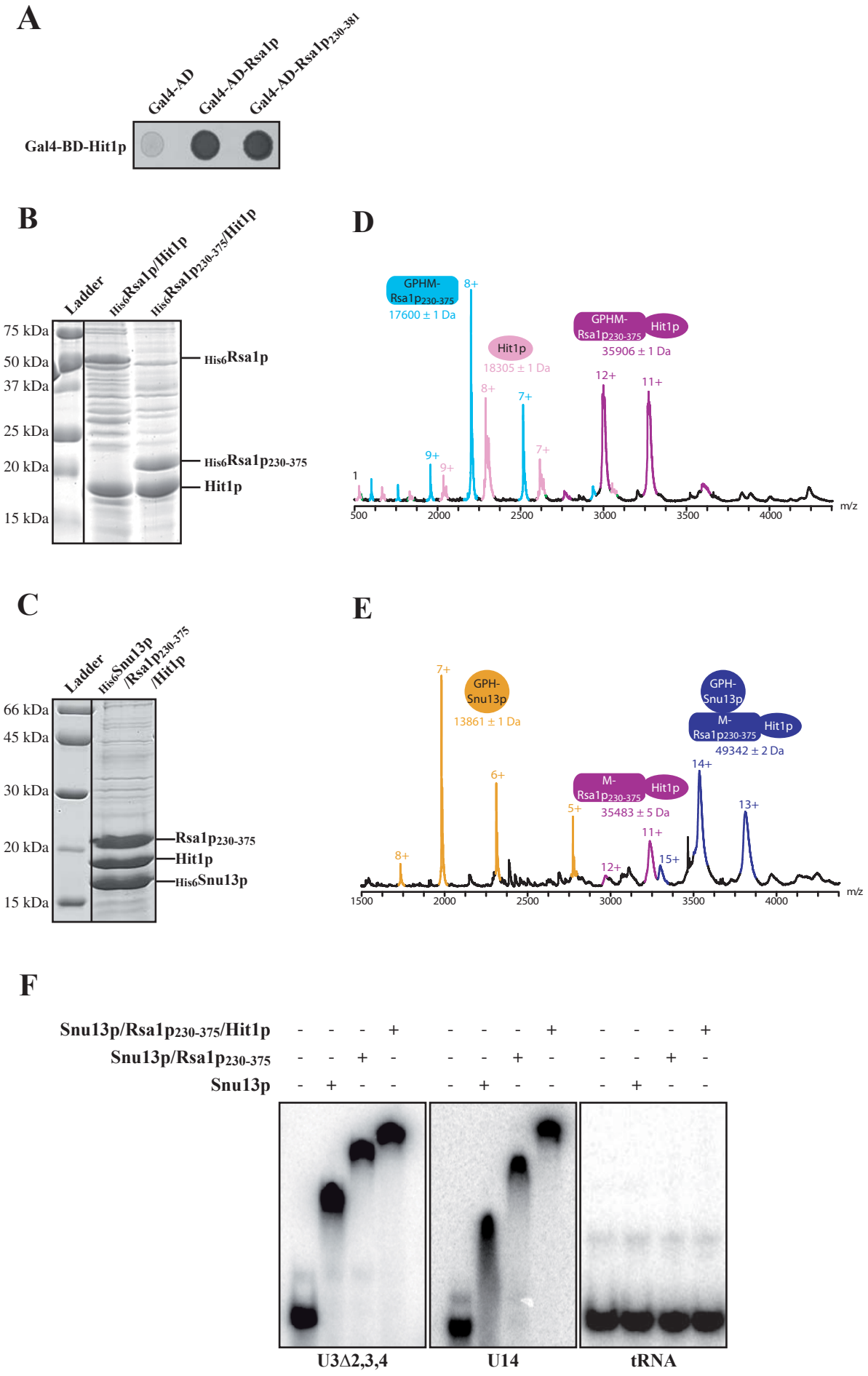

A
Gal4-AD
Gal4-AD-Rsa1p
Gal4-AD-Rsa1p230-381
Gal4-BD-Hit1p
B
Ladder
His6Rsa1p/Hit1p
His6Rsa1p230-375/Hit1p
75 kDa
50 kDa
37 kDa
25 kDa
20 kDa
15 kDa
His6Rsa1p
His6Rsa1p230-375
Hit1p
C
Ladder
His6Snu13p /Rsa1p230-375 /Hit1p
66 kDa
45 kDa
30 kDa
20 kDa
15 kDa
Rsa1p230-375
Hit1p
His6Snu13p
D
GPHM-Rsa1p230-375
17600 ± 1 Da
Hit1p
18305 ± 1 Da
GPHM-Rsa1p230-375 Hit1p
35906 ± 1 Da
m/z
E
GPH-Snu13p
13861 ± 1 Da
GPH-Snu13p M-Rsa1p230-375 Hit1p
49342 ± 2 Da
M-Rsa1p230-375 Hit1p
35483 ± 5 Da
m/z
F
Snu13p/Rsa1p230-375/Hit1p
Snu13p/Rsa1p230-375
Snu13p
U3Δ2,3,4
U14
tRNA

Figure 5

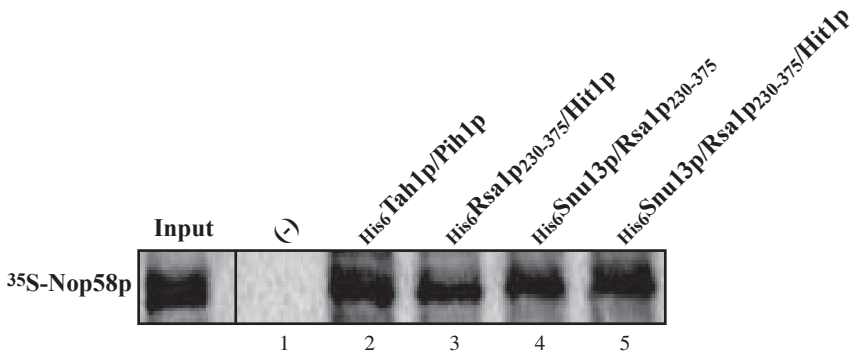

Input
(-)
His6Tah1p/Pih1p
His6Rsa1p230-375/Hit1p
His6Snu13p/Rsa1p230-375
His6Snu13p/Rsa1p230-375/Hit1p
35S-Nop58p
1
2
3
4
5

Figure 6

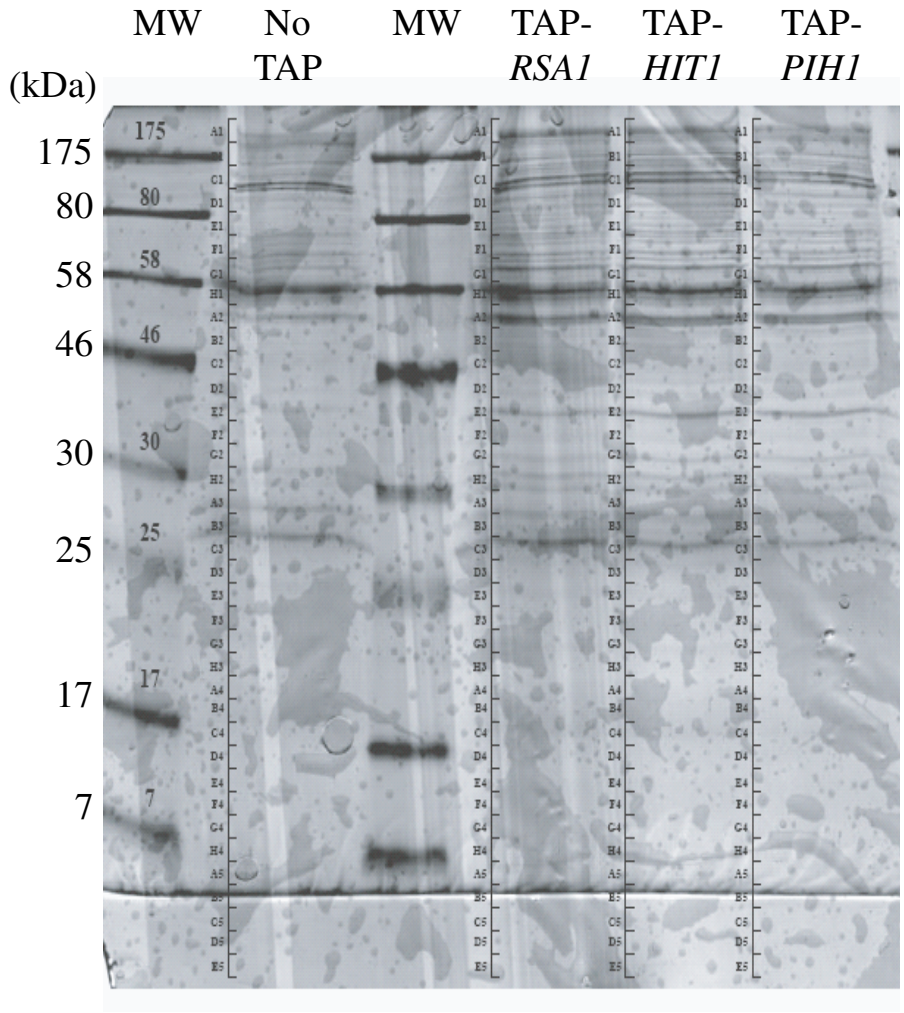


Figure S1: SDS-PAGE analysis of proteins contained in the extracts prepared from tandem affinity experiments of TAP-Rsa1p, TAP-Hit1p, TAP-Pih1p, and an untagged negative control (No TAP). Proteins were separated by 12,5% SDS-PAGE and stained with colloidal Coomassie Blue. Protein markers were fractionated on the left of the gel and their molecular weights in kDa are given.

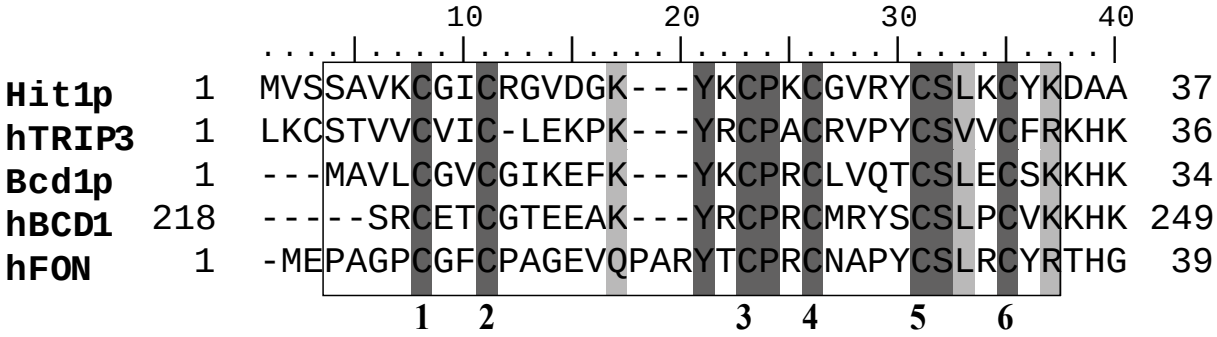


Figure S2: Zinc finger HIT domain sequence. Multiple amino acid sequence alignments were constructed using CLUSTALW. Identical residues are highlighted in dark grey and similar residues in light grey. Zf-HIT domain is framed and zinc-finger forming cysteines are numbered below the alignment.

3 Conclusion et perspectives

3.1 Bilan

L'ensemble de nos résultats et ceux obtenus par B. Rothé nous ont permis d'identifier Hit1p comme un nouveau facteur intervenant dans l'assemblage des snoRNP à boîtes C/D. Nous avons montré que Hit1p s'associait directement avec Rsa1p, quelle est requise pour assurer sa stabilité *in vivo* et qu'elle contribue ainsi de manière importante à la biogenèse des snoRNP. Comme c'est le cas pour les autres facteurs, son implication dans les processus d'assemblage pourrait être conservée dans l'ensemble des eucaryotes, puisque son homologue chez l'homme, la protéine TRIP3, interagit avec NUFIP l'orthologue de Rsa1p (E. Bertrand, communication personnelle).

La protéine Hit1p, exprimée seule chez E. coli s'agrège ou multimérise alors que co-exprimée avec les domaines (231-381) et (317-375) de Rsa1p, la protéine Hit1p entière ou ses fragments (44-164) et (70-164) sont fortement stabilisés. Ce qui montre que réciproquement l'association de Rsa1p à Hit1p participe à la stabilisation de cette dernière. B. Rothé a montré que réciproquement l'expression de Hit1p dans la levure stabilisait Rsa1p. Il est donc important de comprendre structuralement comment cette stabilisation réciproque des deux protéines s'établit grâce à leur interaction. Ceci d'autant plus que l'absence de chacune de ces protéines provoque un défaut de croissance des levures et de la biogenèse des snoRNP C/D.

3.2 Vers un modèle structurale multiprotéique

La protéine Rsa1p semble être une plateforme d'assemblage pour les constituants des snoRNP C/D et des facteurs d'assemblage. Quel est alors le rôle de Hit1p au sein du processus de biogenèse des snoRNP ? L'étude structurale du complexe Snu13p/Rsa1p/Hit1p devrait apporter des informations sur cette question. Nous allons rapidement obtenir la structure 3D du complexe formé entre les fragments Rsa1p (317-352) et Hit1p (70-164). Les études sur l'interaction établie entre Snu13p et Rsa1p décrites dans le chapitre suivant seront aussi utiles pour essayer ensuite d'avoir une vision plus globale des complexes impliqués dans l'assemblage des snoRNP à boîtes C/D. Ainsi, en accumulant des informations sur des parties du complexe Snu13p/Rsa1p/Hit1p nous espérons à terme pouvoir modéliser une structure du complexe entier.

CHAPITRE 3

L'interface Rsa1p/Snu13p, caractérisation fonctionnelle et structurale

1 Présentation des travaux

Comme nous l'avons décrit dans la partie bibliographique, la protéine Rsa1p joue un rôle dans l'assemblage des snoRNP C/D (Boulon *et al.*, 2008). Ces RNP ont la particularité de toutes contenir une protéine appartenant à la famille L7Ae. La protéine Rsa1p est capable de s'associer à ces différentes protéines de la famille L7Ae, dont la protéine Snu13p des snoRNP C/D et de la snRNP U4 de levure. De la même manière l'homologue humain de Rsa1p, la protéine NUFIP se lie aux protéines 15.5K humaine homologue à Snu13p et NHP2 des snoRNP H/ACA (Boulon *et al.*, 2008). Les travaux antérieurs de notre équipe et de celle d'E. Bertrand ont montré que l'interaction de Rsa1p et NUFIP avec les protéines de la famille L7Ae dépend d'un domaine commun de ces deux protéines appelé PEP. Chez la levure, ce domaine correspond au fragment 231-266 de la protéine Rsa1p

Le travail que nous avons réalisé en collaboration avec l'équipe d'E. Bertrand (IGM, Montpellier) a consisté à caractériser les acides aminés jouant un rôle déterminant dans les interactions Rsa1p/Snu13p et NUFIP/15.5K. Par une approche de mutagenèse dirigée, couplé à des approches de biochimie et biologie structurale en particulier la RMN, nous avons pu définir les résidus impliqués dans l'interface Rsa1p/Snu13p et proposer un modèle tridimensionnel du complexe Snu13p/PEP. Selon ce modèle, un réseau de liaisons électrostatiques s'établit entre d'une part les résidus R246 et R249 de la protéine Rsa1p et d'autre part les résidus E72 et D73 de l'hélice α3 de la protéine Snu13p. Des arguments forts en faveur de l'existence de ce réseau sont apportés par les résultats de nos études par dichroïsme circulaire, par expérience de retard sur gel et de double hybride, mais aussi nos travaux de co-expression initiés comme décrit dans le chapitre 1, paragraphe 2.7.4. De façon intéressante, la mutation ponctuelle de quatre résidus impliqués dans l'interaction Rsa1p/Snu13p selon notre modèle conduit à une diminution significative du taux des snoRNA à boîtes C/D et est associée à un défaut de croissance des cellules de levures.

L'ensemble de ces données constitue un manuscrit en préparation pour soumission dans *Nucleic Acid Research* et dont je suis co-premier auteur. Des données complémentaires sur le complexe Snu13p/Rsa1p seront également présentées à la suite de l'article.

Manuscrit n°2

Structural and functional analysis of the interaction between the small nucleolar RNP protein Snu13p/15.5K and the Rsa1p/NUFIP assembly factor demonstrate its direct role in snoRNP assembly

Régis Back*, Benjamin Rothé*, Marc Quinternet, Marie-Cécile Robert, Magali Blaud, Christophe Romier, Xavier Manival, Bruno Charpentier, Edouard Bertrand and Christiane Branlant

En préparation pour Nucleic Acid Research

*Contribution équivalente

Structural and functional analysis of the interaction between the small nucleolar RNP protein Snu13p/15.5K and the Rsa1p/NUFIP assembly factor demonstrate its direct role in snoRNP assembly

Benjamin Rothé[1], Régis Back[1], Marc Quinternet[2], Marie-Cécile Robert[3], Magali Blaud[1], Christophe Romier[4], Xavier Manival[1*], Bruno Charpentier[1*], Edouard Bertrand[3*], and Christiane Branlant[1]

[1]AREMS (ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale), Université de Lorraine, Unité Mixte de Recherche 7214 CNRS-UL, Faculté de Médecine, Bâtiment Biopôle, BP 184, Vandœuvre-lès-Nancy Cédex, France.
[2]FR CNRS-3209 (Ingénierie Moléculaire et Thérapeutique), CNRS, Université de Lorraine, Faculté de Médecine, Bâtiment Biopôle, BP 184, 54505 Vandœuvre-lès-Nancy Cédex, France.
[3]IGMM (Institute of Molecular Genetics of Montpellier), Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5535, Montpellier Cédex 5, France.
[4]IGBMC (Institut de Génétique et Biologie Moléculaire et Cellulaire), Département de Biologie et Génomique Structurales, Université de Strasbourg, CNRS, INSERM, 1 Rue Laurent Fries, BP 10142, 67404 Illkirch Cédex, France.

*To whom correspondence should be addressed. Tel: +33 3 83 68 55 08; Fax: +33 3 83 68 55 09; Email: bruno.charpentier@maem.uhp-nancy.fr.
Correspondence may also be addressed to Xavier Manival. Tel: +33 3 83 68 55 87; Fax: +33 3 83 68 43 07; E-mail: xavier.manival@maem.uhp-nancy.fr.
Correspondence may also be addressed to Edouard Bertrand. Tel: +33 4 67 61 36 47; Fax: +33 4 67 04 02 31; E-mail: edouard.bertrand@igmm.cnrs.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Present address : Magali Blaud, Laboratoire de Cristallographie et RMN Biologiques, Université Paris Descartes, CNRS-UMR8015, Paris Cédex, France.

ABSTRACT

Proper assembly of non-coding RNAs into RNPs is essential to their function. Protein NUFIP and its yeast ortholog Rsa1p are involved in assembly of box C/D small nucleolar ribonucleoprotein particles (snoRNPs) and the U4 small nuclear ribonucleoprotein particle (snRNP). These RNPs share a common L7Ae-like protein (15.5K in human, Snu13p in yeast). By interaction with 15.5K/Snu13p, NUFIP/Rsa1p is proposed to scaffold immature particles and to recruit Hsp90 through a specific co-chaperone, the R2TP complex. However, the precise mode of interaction of 15.5K/Snu13p with NUFIP/Rsa1p is not known. Here, we used yeast two hybrid assays, GST-pull down experiments, and co-expression in *Escherichia coli* to identify key residues in this interaction. By NMR structure determination, docking, and molecular dynamics simulation, we build a structural model of the Snup13p-Rsa1p interaction. In this model, amino acids R_{249} and R_{246} from Rsa1p and E_{72} and D_{73} from Snu13p form three ion pairs and the resulting interaction is stabilized by hosting of the W_{245} side chain of Rsa1p in a hydrophobic cavity located at the Snu13p surface. By mutation of the interacting residues in yeast, we demonstrated the functional importance of the predicted interaction for cell growth, snoRNA accumulation *in vivo*, and pre-snoRNP formation *in vitro*.

INTRODUCTION

During the last decades, numerous ribonucleoprotein particles (RNPs) containing small non-coding RNAs and proteins have been discovered in eukaryotes. Several of them share the common property to contain a primary RNA binding protein which belongs to the L7Ae family of protein ("L7Ae-like" proteins) (1-3). They will be denoted here L7Ae RNPs. Members of this

L7Ae-RNP family play several essential roles in eukaryotic cells: *i*) small nucleolar RNPs (box C/D and box H/ACA snoRNPs) are required for generation of mature ribosomal RNAs starting from pre-ribosomal RNA transcripts (maturation and post-transcriptional modifications) (4,5); *ii*) the spliceosomal U4 small nuclear RNP (U4 snRNA) is an essential component of the pre-mRNA splicing machinery (6) ; *iii*) small Cajal body-specific RNPs (box C/D and H/ACA scaRNPs) catalyze the post-transcriptional modifications of spliceosomal UsnRNPs (7); *iv*) the vertebrate telomerase holoenzyme is needed for telomere synthesis (8,9); and *v*) specific RNA protein complexes which are assembled on the 3' UTR regions of selenoprotein mRNAs (SECIS mRNPs) are required for selenocysteine incorporation (1,10). In their active forms, these various L7Ae RNPs consist of an RNA molecule containing a recognition motif for the "L7Ae-like" protein and a set of specific proteins. The L7Ae RNPs can be divided in different classes depending on the identity of the "L7Ae-like" protein they contain: vertebrate's protein 15.5K (Snu13p in yeast) is a component of U4 snRNP, box C/D snoRNPs and box C/D scaRNPs (3,11,12); vertebrate's protein NHP2 (Nhp2p in yeast) is found in H/ACA snoRNPs, H/ACA scaRNPs, and the telomerase RNP (13,14); finally, vertebrate's protein SBP2 is specific for selenocysteine mRNPs (1,10).

The 3D structures established for human 15.5K and yeast Snu13p proteins are very similar (15,16) showing a high conservation through-out evolution, and these two proteins recognize very similar K-turn motifs characterized by the presence of two helices, I and II separated by a 3-nt long bulge-loop (3,16). One main feature of the K-turn 3D structure is the sharp angle formed by the backbone of the bulge loop (16,17). It is generated by stacking of residues at positions 1 and 2 in the bulge on helices I and II, respectively. As a consequence, the U residue at position 3 in the loop, which is highly conserved in K-turn motifs recognized by proteins of the L7Ae family, is protruding outside of the motif and plays a key role in the RNA-protein interaction (16). The two non-canonical sheared G:A base pairs at the extremity of helix II are also involved in this interaction (3,16,18,19). Binding of 15.5K/Snu13p allows further recruitment of the other specific core proteins: PRP31 for U4 snRNP (20), U3-55K (Rrp9p in yeast) on the U3 B/C motif (18,21), and NOP56, NOP58, and Fibrillarin (Nop1p in yeast) on box C/D snoRNAs as well as box C/D scaRNAs (22,23). The variable elements in K-turn motifs, *i.e.*: length of helix I, sequence of helix II are key determinants for the selective recognition of these specific proteins (18,23,24). The recent determination of the crystal structure of the human PRP31 fragment encompassing amino acids 78 to 346 ($PRP31_{78-333}$) in complex with 15.5K and a fragment of U4 snRNA (25), as well as the crystal structure of archaeal box C/D RNPs (26-28) have provided some clues to understand how diversity is generated on a common theme: namely, how very similar architecture formed by a K-turn motif bound to an "L7Ae-like" protein can interact specifically with different sets of proteins. Although, it is well admitted that the 15.5K/Snu13p primary binding proteins play a key role in initiation of the L7Ae RNP biogenesis, one key remaining question is to know how these RNP complexes are assembled *in vivo* with a high degree of specificity. Indeed, whereas, assembly of the archaeal counterparts of snoRNPs seems to be mainly based on the capacity of their components to interact together (29-32), recent studies revealed an unexpected complexity of the RNP biogenesis pathways in (2,33-36). Eukaryotic RNAs are most frequently produced in a precursor form that is

subsequently processed and assembled with proteins. This is most generally achieved through ordered steps occurring in specific sub-cellular compartments and requiring the assistance of dedicated assembly factors. Up to now, only the biogenesis of vertebrate's spliceosomal UsnRNPs was completely deciphered (37): it requires the export of the pre-UsnRNA to the cytoplasm for 3'- and 5'-end processing and assembly with the Sm proteins (38). This step is mediated by the Survival of Motor Neurons (SMN) complex which binds to the Sm proteins, holds them together from their outside surface to form an open ring, and present them to the RNA (37). The SMN complex thus works as an external scaffold to promote RNP formation. Then, the SMN complex escorts the pre-UsnRNP to the Cajal bodies where RNA maturation and protein assembly are completed. The mature UsnRNPs are released in the nucleoplasm for assembly into splicing complexes or storage into nuclear speckles (34,37-39).

Studies on H/ACA and, especially, box C/D snoRNP assembly mechanisms are far less advanced. However, the present knowledge indicates marked differences as compared to UsnRNP assembly. For instance, the newly translated H/ACA snoRNP proteins are imported into nuclei as preformed complexes (40,41) and they interact with nascent precursor H/ACA RNAs (40,42). Interestingly, some assembly factors like NAF1/Naf1p, functions by mimicking and replacing a core protein in the immature particle, rather than as an external scaffold (43). The Sh1p assembly factor mimicks RNA binding, to prevent association of the H/ACA proteins with non-specific RNAs (44,45). In addition, as mentioned above, one fascinating question is to understand how by recognition of very similar K-turn motifs, and with the assistance of RNP assembly factors Snu13p/15.5k specifically initiates the assembly of three different kinds of RNPs (C/D box snoRNPs, the U4 snRNP and the B/C protein complex of the U3 snoRNP) (3,18,20-23). Therefore, snoRNP assembly represents a new attractive paradigm to understand RNP biogenesis in eukaryotes.

In addition to the "L7Ae-like" protein, one cellular factor, the vertebrate's NUFIP protein (Rsa1p in yeast), plays a key role in assembly of snoRNPs and more generally in assembly of all RNPs of the L7Ae family (2,35,46). This platform protein facilitates the formation of a macro-complex containing both the core RNP proteins and the assembly factors. More precisely, NUFIP/Rsa1p is proposed (*i*) to hold together core proteins in the immature particles, and (*ii*) to recruit the Hsp90 chaperone through its interactions with an Hsp90 co-chaperone complex, named the R2TP complex (2,36). The human R2TP complex is composed of four proteins: PIH1/NOP17 (Pih1p/ Nop17p in yeast), hSpagh (also named RPAP3, Tah1p in yeast), TIP48 (also known as Reptin or p50 in vertebrates, and Rvb2p in yeast) and TIP49 (also known as Pontin or p55 in vertebrates, and Rv1bp in yeast) (2). Protein PIH1/NOP17 interacts with protein hSpagh within the R2TP complex (47). By its interaction with NUFIP/Rsa1p, PIH1 allows the tethering of the R2TP complex on the pre-RNP (2), while by its interaction with HSP90 (Hsp82 in yeast), protein hSpagh/Tah1p allows the recruitment of the chaperone on the pre-RNP. The two other proteins of the R2TP complex, TIP48 and TIP49 are AAA^+ ATPases. They interact together and with protein PIH1 (2,36,47). The implication of Hsp90 in the biogenesis of L7Ae RNPs was shown by the disappearance of newly-synthesized U4, box C/D and telomerase RNAs, as well as of 15.5K, NOP58 and NHP2 proteins, upon inhibition of Hsp90 activity, suggesting that these proteins are clients of the chaperone (2,36). Therefore, as compared to UsnRNP assembly, snoRNP assembly provides a very

different paradigm for RNP assembly and is an excellent model to decipher mechanisms ensuring specificity of RNP assembly.

At the present stage of the knowledge, association of NUFIP/Rsa1p to the L7Ae protein appears to be a key determinant of the assembly process. Although NUFIP and Rsa1p show a poor degree of similarity, we previously identified a 37 amino acid sequence which is highly conserved in the two proteins (hPEP and yPEP motifs), and showed that the hPEP and yPEP motifs are essential for association of NUFIP and Rsa1p with proteins 15.5K and Snu13p, respectively (2). Here, to improve our knowledge on assembly of RNPs containing protein 15.5K/Snu13p, we used yeast molecular genetics, biochemical assays, cell imaging, NMR spectroscopy, and molecular modeling experiments to: (*i*) identify the interacting amino acids in the NUFIP-15.5k and Rsa1p-Snu13p interactions, (*ii*) establish the solution 3D structure of a fragment of the PEP domain containing the interacting amino acids and a 3D structure model of its interaction with Snu13p, (iii) determine the respective functional importance of the Rsa1p and Snu13p interacting amino acids. *In vivo* substitution of the Snu13p and Rsa1p interacting amino acids in yeast demonstrated their requirement for cell growth and the stability of box C/D snoRNAs. *In cellulo* mutation of their counterparts in 15.5K abolished 15.5K-NUFIP co-localization in Hela cell nuclei.

MATERIAL AND METHODS

Recombinant plasmids used in this study

Plasmids were manipulated by standard techniques. Mutations were generated using the QuickChange mutagenesis system according to the manufacturer's instructions (Stratagene©) and appropriate primer pairs.

The bacterial expression vector pGEX-6P-1::*SNU13* was already described (19). To express the C-terminal 152 amino acids fragment of Rsa1p in *E. coli*, a *Bgl*II-*Xho*I restriction fragment was released by digestion of plasmid pACTII::*Rsa1p*$_{230-381}$ (previously designated as pACTII::N3C1) (2), and cloned into plasmid pET30c (Novagen©). A 111 bp fragment containing the Rsa1p coding sequence of amino acids 230 to 266 was obtained by PCR amplification using two primers that generated a *Nhe*I site at one end and an *Xho*I site at the other end. This fragment was inserted between the *Nhe*I and *Xho*I sites of plasmid pET28a (Novagen©) to obtain the *E. coli* expression vector pET28a::*RSA1*$_{230-266}$.

For protein co-expression in *E. coli*, DNA fragments encoding the full length Rsa1p, the Rsa1p$_{230-381}$ fragment or protein Pih1p were amplified by standard PCR procedures, using primers that generated *Nde*I and *Bam*HI restriction sites, and they were inserted in the pnCS vector. On the other hand, the Snu13p and Tah1p ORFs were inserted in the pnEA-3CH vector in order to produce His-Tagged proteins (48-50).

For expression of WT or mutated Snu13p or Rsa1p protein in yeast, the *S. cerevisiae SNU13* and *RSA1* ORFs were PCR amplified using primers that generated a *Bam*H1 site at the ORF 5' end and an *Xho*I site at its 3' end, and were cloned between the *Bam*H1 and *Sal*I sites of a centromeric version of the pG1 (*TRP1*) expression vector (51,52). The resulting plasmids were designated as pG1::*SNU13* and pG1::*RSA1*.

The recombinant plasmids used for Y2H assays were already described (2). Briefly, the ORFs were cloned by the Gateway system (invitrogen™) in plasmid pACTII downstream from the *GAL4* transactivation domain coding sequence, or in plasmid pAS2 downstream from the *GAL4* DNA binding domain coding sequence.

The plasmids coding for mRFP-LacI-NUFIP and GFP-15.5K were generated with the Gateway technology with destination vectors previously described (53).

Production of recombinant proteins

The recombinant pGEX-6P-1::*SNU13* plasmid was used to produce Snu13p fused to the glutathione S-transferase (GST) (19) in the *E. coli* CodonPlus strain (Novagen©). The recombinant GST-Snu13p protein was purified from cellular extract under native conditions, using Glutathione Sepharose 4B as recommended by the manufacturer (GE Healthcare©). The GST-Snu13p proteins were recovered after overnight cleavage on the beads using the PreScission protease (GE Healthcare©). The cleaved recombinant Snu13 proteins had the two additional amino acids G and P at their N-terminus compared to the native protein. The recombinant pET30c::*RSA1*$_{230-381}$ and pET28a::*RSA1*$_{230-266}$ expression vectors were used to produce the Rsa1p fragments encoding amino acids 230 to 381, and 230 to 266, respectively, with an His6 N-terminal tag in the *E. coli* CodonPlus strain (Novagen©). Note that the peptide produced by plasmid pET28a::*Rsa1p*$_{230-263}$ contains at its N-terminus the His6-tag sequence and a 15 amino acid sequence resulting from the cloning into the pET28a plasmid. The recombinant $_{His6}$Rsa1p$_{230-266}$ and $_{His6}$Rsa1p$_{230-381}$ proteins were purified under native and denaturing/refolding conditions, respectively, using a 5 mL His-Trap™ HP column (GE Healthcare©). Proteins were eluted with 0.5 M imidazole at pH 8. When necessary, the His_6-tag was removed by cleavage with the Thrombin protease. After Thrombin cleavages, or imidazole elution, the obtained recombinant proteins were dialyzed against buffer D (20 mM HEPES-KOH, pH 7.9; 150 mM KCl; 1.5 mM $MgCl_2$; 0.2 mM EDTA; 10% glycerol) and stored at -80°C.

Test of protein-protein interactions by their co-expression in *E. coli*

Co-expression tests in *E. coli* were carried out using a standard procedure described previously (49). The two recombinant plasmids pET15b and pACYC18b-11b expressing, respectively, $_{His6}$Snu13p and full-length Rsa1p or its Rsa1p$_{230-381}$ fragment were used simultaneously to transform pRare2 *E. coli* host cells (Novagen). Growth was performed in 500 mL of 2X concentrated LB medium at 37°C to an absorbance at 600 nm of 0.8. Then, the temperature was switched to 20°C. At this point, co-expression of the proteins was induced by addition of isopropyl-β-D-thiogalactopyranoside (Euromedex) at a 1 mM final concentration. Cells were further grown overnight at 20°C. They were collected by low speed centrifugation, resuspended in buffer A (20 mM HEPES-NaOH, pH 7.5; 400 mM NaCl; 5 mM β-mercaptoethanol), and lysed by sonication. The soluble fraction recovered by high speed centrifugation was mixed with 100 µl of Talon resin (Clontech). After 1 hour of incubation, the supernatant was removed and the resin washed extensively with buffer A. Proteins were eluted from the column using buffer A containing 500 mM imidazole. Eluted proteins were analyzed by SDS-PAGE.

Electrophoresis Mobility Shift Assays (EMSA)

Yeast U14 box C/D snoRNA was synthesized by *in vitro* transcription with T7 RNA polymerase of a PCR fragment carrying the *S. cerevisiae* snoRNA sequences in the conditions previously described (19). After transcription, RNAs were 5'-end labelled with T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP and purified by gel electrophoresis. For EMSA, WT or variant Snu13p were mixed with $_{His6}$Rsap1p$_{230-266}$ or $_{His6}$Rsa1p$_{230-381}$ protein at a final concentration of 0.5 µM, and incubated for 20 min at 4°C with 5 fmol of ^{32}P-radiolabeled U14 RNA in buffer D (20 mM HEPES-KOH, pH 7.9; 150 mM KCl; 1.5 mM $MgCl_2$; 0.2 mM EDTA; 10% glycerol). The RNA-protein complexes

formed were resolved by native gel electrophoresis in conditions described previously (30).

Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy analyses were carried out using a Jobin-Yvon Mark VI circular dichrograph. Quartz spare split-compartment cuvettes with 0.44 cm path length per compartment were used for binding measurements. Use of these cuvettes allows accurate assessment of spectral differences between proteins in a bound and a free state or specific modifications in the RNA spectrum caused by its interaction with proteins. The relevant protein solution was placed in one compartment of the cuvette and the target protein solution in the other compartment. CD spectra were recorded before and after mixing the cuvette contents and blanks were run for each interaction and subtracted from the raw data. At least two spectra were collected before and after mixing, an average spectrum was established in order to increase the signal-to-noise ratio. Spectra were recorded from 200 to 260 nm to test protein-protein interaction with a scan speed of 0.2 $nm.s^{-1}$. Proteins were adjusted at a 4 µM final concentration in 10 mM Tris-HCl buffer, pH 7.8; 150 mM NaCl. The results are presented as normalized $\Delta\varepsilon$ values. For protein-protein interactions we used amino acid mean residue mass of 110 Da. Taking into account the residue concentration and the optical path-length of the cuvette, the precision of measurements was of $\delta(\Delta\varepsilon)$ = +/- 0,03 $dm^3.mol^{-1}.cm^{-1}$.

NMR Spectroscopy and structure calculation

For solubility reasons, a synthetic peptide of Rsa1p ($Rsa1p_{238--252}$) was used for NMR analyses (GeneCust®). The NMR sample of $Rsa1p_{238-252}$ contained 3 mM of peptide in 10 mM Phosphate buffer, pH 6.4; 95% H_2O; 5% D_2O. NMR data were acquired at 20°C on a Bruker Avance III spectrometer equipped with a Z axis gradients TCI cryoprobe. For proton resonance assignment, TOCSY spectra using 2 mixing times (30 and 60 ms) and NOESY spectra using 3 mixing times (75, 150 and 300 ms) were recorded. Assignment of heteronuclei was achieved using ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HSQC-TOCSY (with a mixing time of 60 ms) experiments in natural abundance. All spectra were processed with Topspin 3.0 (Bruker BioSpin) and analyzed with CARA (54).

Distance restraints were derived from NOESY spectra and torsion angle restraints for the ϕ angle were derived from measurements of the ^{3}J H^{N}-H^{α} coupling constant on 1D proton spectra. 200 structures of $Rsa1p_{238-252}$ were calculated with CYANA 2.1 (55). The 50 best structures were used for explicit water-refinement in CNS 1.2 (56). Finally, the 10 structures with the lowest energies were selected as the most representative structures of $Rsa1p_{238-252}$.

Docking and molecular modeling

The best NMR structure of $Rsa1p_{238-252}$ was docked on the Snu13p X-ray 3D structure (PDB code 2ALE) using the Haddock2.1/CNS1.2 protocol (57-59). In order to generate the required Ambiguous Interaction Restraints file, residues for which point mutations abolish formation of a stable complex between Snu13p and Rsa1p have been considered as active whereas residues with less significant effect or solvent accessible residues in the neighborhood of active residues have been considered as passive. The solvent accessibility was checked with MOLMOL (60). The entire $Rsa1p_{238-252}$ and the segment that encompasses the passive and active residues of Snu13p have been defined as semi-flexible. Moreover, active residues of the peptide and the segment that encompasses the passive and active residues of Snu13p have been defined as fully flexible. 1000

structures have been calculated during the rigid body energy minimization step (it0) and the top hundred best structures were used for the semi-flexible simulated annealing (it1), and the subsequent flexible explicit H_2O refinement steps (it1-water). The top ten structures with the best haddock score were selected as possible models for interaction.

Then, the structure with the best haddock score was placed in a water box of 80Å^3 containing 16257 water molecules and one chlorine ion for ensuring the electrostatic neutrality of the system. All Arg, Lys, Asp, and Glu side chains were considered ionic. The resulting system, consisting in a total number of 51,121 atoms, was energy refined. The MD program NAMD (61) was employed in combination with the CHARMM27 force field (62) in order to describe the whole molecular system. Coulomb forces were evaluated with the particle-mesh Ewald method. Chemical bonds between hydrogen and heavy atoms were constrained to their equilibrium value by means of the SHAKE algorithm. The refinement procedure was based on several energy minimization steps carried out by the conjugate gradient method: first, the backbone of the protein complex was fixed such that only the protein side chains and water molecule movements were variable; next, the whole system was relaxed. Finally, a 5-ns MD simulation was carried out in the isobaric-isothermal ensemble, maintaining the pressure and the temperature at 1.0 atm and 300.0 K, respectively, by means of Langevin dynamics and the Langevin piston approach. The equations of motion were integrated with a 2-fs time step, using the r-RESPA algorithm to update short- and long-range contributions at different frequencies.

Yeast two-hybrid assays

For yeast two-hybrid assays (Y2H), appropriate pACTII and pAS2 plasmids were used to transform haploid cells (strain CG929 or Y190, and strain Y187, respectively), which were then crossed. Diploids were selected on media suitable for double selection (Leu^-, Trp^-) and then plated on media suitable for triple selection (Leu^-, Trp^-, His^-). 3-Amino-1,2,4-triazol (3AT) was used to evaluate the strength of the interactions. Growth was assessed after three days of incubation at 30°C.

Tests of the functionality of variant Snu13p and Rsa1p proteins in yeast

Standard *S. cerevisiae* growth and handling techniques were employed. The *S. cerevisiae* strain YPH499-*GAL::SNU13* (*Mat*a *trp*1-Δ63; *his*3-Δ200; *ura*3–52; *lys*2–801; a*de*2–101; *leu*2-Δ1; *HIS3::GAL1::SNU13*) (3) was used to test whether Snu13p variants can complement depletion of the essential endogenous *SNU13* gene. To this end, strain YPH499-*GAL::SNU13* was transformed with recombinant pG1::*SNU13* plasmids encoding WT or mutated Snu13p. Cells were grown overnight at 30°C in a medium containing galactose as the carbon source (YPG). They were collected by centrifugation, washed, and diluted in YPD medium containing glucose as the carbon source. Cells were grown at 30°C and constantly maintained in exponential phase by dilution with pre-warmed YPD. The doubling time was monitored by measuring the absorbance at 600 nm (OD600).

The *S. cerevisiae* Y190-Δ*RSA1* (*YPL193W::kanMX2*) strain, used in Y2H assays, was obtained by homologous recombination. A DNA fragment carrying the kanamycin cassette flanked by sequences amplified from the *RSA1* gene was used to transform the haploid Y190 strain. Insertion of the kanamycin cassette at the *RSA1* locus was then verified by PCR amplifications. The *RSA1* knock-out strain obtained was transformed with plasmid pG1::*RSA1* encoding WT or variant Rsa1p, and grown in YPD medium.

Evaluation of snoRNA levels in yeast

Total RNAs were extracted from exponentially growing yeast cultures. Northern blot analyses were performed in standard conditions using 5 µg of total RNAs and specific 5'-end ^{32}P-radiolabeled antisens probes:
OG-U1 (5'-GACCAAGGAGTTTGCATCAATGA-3');
OG-snR63 (5'-TTATGTTGGCCACTCATCAC-3');
OG-snR190 (5'-CGAGGAAAGAAGAGACACCATTATC-3');
OG-snR45 (5'-CCTCAGATCGCTCCGAGA-3');
U14 (5'-TCACTCAGACATCCTAG-3');
OG-snR13 (5'-AGCTTGAGTTTTTCCACACC-3');
OG-U18 (5'-TGTGATAGTCAACCACATCTC-3').

His-pull down assays of *in vitro* assembled RNPs

100 fmol of 5'-end ^{32}P-radiolabeled U14 RNA prepared according to (19) were incubated with various sets of recombinant proteins during 30 min in buffer D at 30°C, 1 µM of protein Snu13p and of the $_{His6}$Tah1p-Pih1p complex were used, while the Rsa1p$_{230-381}$ fragment was used at a 0.25, 0.5 µM, 1 µM, or 2 µM concentration. The RNP complexes formed were incubated with Nickel Sepharose beads (GE Healthcare©) for 15 min at room temperature in 200 µL of binding buffer (10 mM Tris-HCl, pH 8.0; 150 mM NaCl; 0.1% IGEPAL). The beads were washed three times for 10 min with 1 mL of binding buffer. RNAs were extracted by phenol-chloroform treatments, ethanol precipitated, and fractionated on a 7% denaturing polyacrylamide gel.

Subcellular localization assay of WT and mutated 15.5K proteins in HeLa cells

The co-recruitment assay was done as previously described (53).

RESULTS

Although NUFIP and Rsa1p are expected to have similar functions in snoRNP assembly, their sequence alignment revealed a low level of sequence similarity, except for the 37 amino acid long PEP domain (Figure 1A). Presence of the PEP domain was previously found to be required to get positive NUFIP-15.5K and Rsa1p-Snu13p interactions in 2-H assays, and when using the same approach, the hPEP and yPEP domains alone were, sufficient to get a positive signal with the 15.5K and Snu13p proteins, respectively (2). Moreover, His6-yPEP was previously found to associate with Snu13p bound to U14 snoRNA (2). Altogether, this had emphasized a major role played by the PEP domain for interaction of the two platform proteins, with their respective 15.5K and Snu13p partners. Thus, we used various biochemical, biophysical and genetics methods to identify the 15.5K and Snu13p amino acids which interact with hPEP and yPEP, respectively, and also to determine which amino acids of the hPEP and yPEP motifs are crucial for these interactions. Based on the high sequence similarity of the 15.5K and Snu13p proteins (Figure 1B), and of the hPEP and yPEP motifs, we postulated very similar modes of interaction of these two proteins with their PEP partner motifs. Therefore, information on these interactions was gained by performing some of the experiments on yeast and other ones in human cells. As in this study, we used several Rsa1p sub-fragments (Figure 1A) each of them are identified as Rsa1px-y, where x and y represents the positions in the Rsa1p amino acid sequence of the amino acids located at their extremities. To homogenize the nomenclatures, the yeast and human PEP domains will be denoted here Rsa1p$_{230-266}$ and NUFIP$_{223-260}$, respectively.

15.5K-NUFIP association depends upon the 15.5K helix α3

To identify the NUFIP interacting domain of 15.5K, the surface residues in 15.5K were mutagenized into alanines. A series of ten 15.5K mutants was generated, and the resulting proteins were tested for association with NUFIP and the $NUFIP_{223-260}$ segment (hPEP), by the use of yeast two-hybrid (Y2H) assays. Mutations in 15.5K which had a negative effect on its interaction with the $NUFIP_{223-260}$ fragment also had a negative effect on the interaction with full-length NUFIP. However, due its lower stability the $NUFIP_{223-260}$-15.5K interaction was more drastically affected by mutations in 15.5K than the full-length NUFIP-15.K interaction (Figure 1C). Nine of the tested variants still interacted with NUFIP and $NUFIP_{223-260}$, while one (variant D) had lost these interactions. Variant D carried 3 amino acid substitutions ($E_{74}D_{75}K_{76}$/AAA) in a highly conserved sequence of the C-terminal part of helix α3 (Figure 1B). Next, by using the same approach, we tested the importance of each residue in the EDK sequence. The E_{74}/A mutation turned-out to be sufficient to abolish the Y2H interaction with both NUFIP and $NUFIP_{223-260}$, while the D_{75}/A and K_{76}/A mutations had no adverse effects (Figure 1C). Noticeably, among proteins able to recognize K-turn RNA motifs, the presence of an E residue at a position equivalent to that of the 15.5K E_{74} residue is a characteristic feature which is unique to proteins of the L7Ae family, namely, Snu13p, NHP2/Nhp2p and SBP2 (Figure 1B).

We also tested the impact on the 15.5K-NUFIP interaction of mutations in other parts of the 15.5K helix α3. Residues E_{64}, I_{65}, H_{68}, and L_{71} were of particular interest, since in the U4 snRNP, E_{64}, I_{65}, and H_{68} were found to establish contacts with the PRP31 Nop domain (25), and L_{71} is spatially close to E_{74}, but not conserved in NHP2, where it is replaced by a valine (Figure 1B). Variants E_{64}/A, I_{65}/A, H_{68}/A, and L_{71}/A still interacted with NUFIP, although the interaction was weaker for the L_{71}/A variant (Figure 1C and data not shown). When tested with $NUFIP_{223-260}$, the E_{64}/A and I_{65}/A 15.5K variants interacted with this small motif, while H_{68}/A and L_{71}/A had lost the interaction (Figure 1C), reflecting the generally weaker interactions obtained with the $NUFIP_{223-260}$ fragment compared to full-length NUFIP. Altogether, two main conclusions were drawn from these data: i) residue E_{74} of the 15.5K helix α3 is essential for binding to NUFIP and, ii) residues H_{68} and L_{71} likely have some indirect of direct modulation effect on this interaction.

The 15.5K $E_{74}D_{75}K_{76}$/AAA mutant has an altered intra-cellular localization and does not bind to NUFIP *in cellulo*

Having shown the functional importance of residue E_{74} in the $E_{74}D_{75}K_{76}$ motif of 15.5K for formation of the 15.5K-NUFIP interaction in a heterologous system, we then used a spatial recruitment assay in HeLa cells in order to verify that the 15.5K EDK motif is needed for the 15.5K-NUFIP interaction in an authentic *in cellulo* context (Figure 2). The approach used was based on the LacI/LacO system (40,63), that circumvents artifacts occurring during co-immunoprecipitation assays, in particular, the loss of weak interactions or on the contrary the re-association of proteins in the test tube. More precisely, a HeLa cell line containing a tandem array of LacO sites integrated in its genome was transfected with the mRFP1-LacI-NUFIP expression vector. In the U2OS parental cell line devoid of LacO sites, the fusion protein mRFP1-LacI-NUFIP had a diffuse localization in the nucleoplasm, which was similar to that of wild-type NUFIP (Figure 2A). When the GFP-15.5K fusion was co-expressed with mRFP1-LacI-NUFIP in U2OS cells, it localized in the nucleoli, Cajal bodies, and nucleoplasm as previously described, and no particular co-localization was observed with

mRFP1-LacI-NUFIP. In contrast, in the cell line containing the LacO array, GFP-15.5K co-localized with the bright spots formed by the binding of mRFP1-LacI-NUFIP to the LacO sites. This indicated that, as expected, NUFIP interacts with 15.5K *in vivo*. We then tested the 15.5K $E_{74}D_{75}K_{76}$/AAA variant in the same assay (Figure 2B). In parental U2OS cells, the localization of the mutant protein differed from its wild-type counterpart. Consistent with a defect in its incorporation into U4 and box C/D RNPs, the mutant 15.5K accumulated at a lower level in nucleoli compared to WT 15.5K and had a more diffuse localization throughout the nucleoplasm and the cytoplasm. Moreover, in cells containing mRFP1-Laci-NUFIP bound to the LacO array, GFP-15.5K $E_{74}D_{75}K_{76}$/AAA was unable to accumulate at the LacO sites, indicating that it was defective for NUFIP binding *in vivo*. Therefore, the data demonstrated the importance of the extremity of helix α3 of 15.5K for *in cellulo* binding of protein 15.5K to NUFIP.

Then, for further characterization of the 15.5K/Snu13p-NUFIP/Rsa1p interaction, we turned to *S. cerevisiae* because of our capability to produce Snu13p and Rsa1p as soluble recombinant proteins in *E. coli* and, the possibility to use yeast genetics to investigate the functional importance of the Snu13p-Rsa1p interaction in snoRNP assembly.

Amino acid(s) in the Snu13p $E_{74}D_{75}K_{76}$ motif is/are directly involved in the Snu13p-Rsa1p interaction

We hypothesized that, like for 15.5K, some residues in the Snu13p $E_{72}D_{73}K_{74}$ motif located in the C-terminal part of helix α3 (Figure 1 and 3A) are important for binding to Rsa1p. To test this hypothesis, we used biochemical and biophysical approaches, having in mind, to get for the first time a clear demonstration of formation of a stoichiometric complex between the purified Snu13p and Rsa1p proteins. To this end, we used co-expression assays in *E.coli*. The His6-tagged WT Snu13p or its His6-tagged $E_{72}D_{73}K_{74}$/AAA variant were co-expressed with either full-length Rsa1p or its C-terminal $Rsa1p_{230-381}$ fragment and their capacity of interaction were investigated by co-immobilization on Metal Ion Affinity Chromatography (IMAC) (Figure 3C). Fractionation of the IMAC purified proteins by SDS-PAGE revealed the formation of a His6-Snu13p-Rsa1p complex. Due to its low level of structuration, Rsa1p is highly sensitive to proteolysis in *E. coli*, which is evidenced by the appearance of some protein bands below the Rsa1p band in the SDS-PAGE fractionation (Figure 3C lane 2). However, taking into account these Rsa1p digestion products, we could estimate that interaction of Rsa1p and its sub-fragments with Snu13p occurred with an apparent 1:1 molar stoichiometry (Figure 3C lane 2). Snu13p also co-purified with $Rsa1p_{230-381}$, but not with the Rsa1p N-terminal domain (data not shown). No co-purification of Rsa1p was detected with the His-tagged $E_{72}D_{73}K_{74}$/AAA Snu13p variant (Figure 3C, lane 4), which revealed the importance of residue(s) within the EDK motif for the Snu13p-Rsa1p interaction. Accordingly, when recombinant GST-Snu13p or its $E_{72}D_{73}K_{74}$/AAA variant were immobilized on Glutathione Sepharose beads and incubated with the recombinant His6-$Rsa1p_{230-266}$ protein, GST-Snu13p retained significant amounts of His6-$Rsa1p_{230-266}$ protein (Supplementary data Figure S1, lane 6), while GST alone or the $E_{72}D_{73}K_{74}$/AAA Snu13p variant did not (Supplementary data Figure S1, lane 8).

At this stage of the study, two explanations could be proposed for these data: either mutation of the $E_{72}D_{73}K_{74}$ motif altered the overall Snu13p conformation and consequently abrogated the interaction with

Rsa1p, or the Snu13p $E_{72}D_{73}K_{74}$ motif is directly involved in the interaction.

By providing information on the α-helix content of proteins, far UV Circular Dichroïsm (CD) spectroscopy can reveal conformational changes occurring upon generation of mutations in proteins or upon their interaction with a partner. As Snu13p predominantly folds into α-helices (15), this approach was perfectly adapted to test whether the $E_{72}D_{73}K_{74}$/AAA mutation modified the Snu13p secondary structure. The CD spectra of WT Snu13p and its $E_{72}D_{73}K_{74}$/AAA variant recorded from 205 to 260 nm showed similar negative peaks at 208 and 222 nm, which was an indication for similar levels of α-helices (data not shown ? a montrer ?). Therefore, we concluded that the $E_{72}D_{73}K_{74}$/AAA mutation probably has limited effect on the protein structure, which favored the hypothesis of a direct physical interaction of residues in the $E_{72}D_{73}K_{74}$ motif with Rsa1p. Comparison of CD spectra collected before and after mixing the two proteins reinforced this hypothesis. For this purpose, we used a cuvette with two compartments: one was filled with Snu13p and the other one with $Rsa1p_{230-266}$ at a protein to peptide molar ratio of 1, and a CD spectrum was recorded from 205 to 260 nm before and after mixing the contents of the two compartments. When using WT Snu13p the intensity of the signal at 222 nm was modified after mixing (Figure 4), which was indicative of the formation of a direct interaction between the two partners accompanied by some structural modifications (Figure 4A, left panel). In contrast, no variation at 222 nm was detected upon mixing, when the Snu13p $E_{72}D_{73}K_{74}$/AAA variant was used (Figure 4A, right panel), reflecting its inability to interact physically with $Rsa1p_{230-266}$. Therefore, by utilization of biochemical and bio-physical approaches, we demonstrated: i) the formation of a Snu13p-Rsa1p physical interaction at a 1:1 molar ratio, ii) the direct involvement of some amino acids in the $E_{72}D_{73}K_{74}$ motif in this interaction. In addition, we developed a CD based assay to test for the Snu13p-Rsa1p interaction.

Residue E72 in the Snu13p EDK motif plays a key role in the Snu13p-Rsa1p interaction

Then, to delineate which of the three EDK amino acids was/were important for the Snu13p-Rsa1p interaction, we came back to the Y2H approach. As expected from previous results (2), WT Snu13p interacted with both full-length Rsa1p and its C-terminal half ($Rsa1p_{230-381}$). No interaction was detected with either the Snu13p $E_{72}D_{73}K_{74}$/AAA variant or the Snu13p E_{72}/A variant (Figure 3B), while the D73/A and K74/A mutations had no marked deleterious effects on the interaction (not shown). Therefore, as found for protein 15.5K, The E_{72} residue in the Snu13p EDK motif is needed for interaction with Rsa1p. Furthermore, by using the U14 snoRNA-Snu13p complex, we showed that this residue is required for interaction of both free and snoRNA bound Snu13p with Rsa1p. Indeed, as previously observed (2), $Rsa1p_{230-381}$ and $Rsa1p_{230-266}$ super-shifted the complex formed by Snu13p and an *in vitro* transcribed U14 box C/D snoRNA (Figure 3D, lanes 3 and 4), but, did not shift this complex, when the Snu13p $E_{72}D_{73}K_{74}$/AAA or the E_{72}/A variants were used.

As these experiments demonstrated the major role of residue E_{74} of the Snu13p EDK motif for interaction with Rsa1p, we decided to determine which amino-acid(s) in Rsa1p can be the partner(s) of the Snu13p E_{72} residue.

Rsa1p residues W_{245} and R_{249} are required for binding to Snu13p

Alignment of the human $NUFIP_{223-260}$ and yeast $Rsa1p_{230-266}$ sequences revealed the presence of a strictly conserved five amino

acids stretch (WREER, positions 245 to 249 in Rsa1p) (Supplementary data Figure S2). To assess whether this conserved $W_{245}R_{246}E_{247}E_{248}R_{249}$ sequence is involved in the Rsa1p-Snu13p interaction, we generated individual substitutions by alanines of almost each of its residues in the $Rsa1p_{230-266}$ protein (Figure 5A). The effects on the interaction with Snu13p was assessed, using as reporter signal, the specific modification of the CD signal at 220 nm observed upon formation of the Snu13p-Rsa1p complex (Figure 4A). While the decrease in the CD signal at 222 nm observed for association of the E247/A variant with Snu13p was similar to that found for WT $Rsa1p_{230-266}$, no decrease of this signal was detected when mixing the $Rsa1p_{230-266}$ W_{245}/A, or R_{246}/A, or R_{249}/A variants with Snu13p (Figure 4B). These data strongly suggested the direct involvement of residues W_{245}, R_{246}, and R_{249} in the Snu13p-$Rsa1p_{230-266}$ interaction. Accordingly, when we tested the ability of $Rsa1p_{230-266}$ variants to form trimeric complexes with Snu13p and the U14 snoRNA by gel-shift assays, only the $Rsa1p_{230-266}$ E_{247}/A variant super-shifted the Snu13p-U14 RNA complex (Figure 5B). Interestingly, when we introduced the amino acid substitutions in the WREER motif of $Rsa1p_{230-381}$ and tested the ability of the variant proteins to interact with Snu13p by Y2H assays, only the W_{245}/A and R_{249}/A substitutions impaired the Y2H interaction (Figure 5C). The R_{246}/A substitution had no visible effects. Similarly, when we co-expressed the WT or variant Rsa1p or $Rsa1p_{230-381}$ proteins with $_{His6}$Snu13p in *E. coli*, and co-purified $_{His6}$Snu13p with its associated proteins by IMAC, the R_{249}/A substitution in both full-length Rsa1p and $Rsa1p_{230-381}$ abolished the interaction with Snu13p, whereas the R_{246}/A variants retained some binding activity (Figure 5D).

The lower effect of the R_{246}/A substitution in our Y2H and co-expression assays compared to super-shift assays and CD measurements is likely explained by the fact that the former assays were made with $Rsa1p_{230-381}$ while the latter assays were performed with $Rsa1p_{230-266}$. As found for the 15.5K-NUFIP interaction, the Snu13p-$Rsa1p_{230-266}$ interaction is most probably less stable than the Snu13p-$Rsa1p_{230-381}$ interaction and, therefore more strongly destabilized by the R_{246}/A amino acid substitution. Altogether, the data identified two residues in Rsa1p (W_{245} and R_{249}) which are crucial for the interaction with Snu13p. They also indicated a possible modulating role of residue R_{246}. As the important residue in Snu13p was a glutamic acid (E_{74}), it sounded to find a crucial role of a lysine (R_{249}) in the partner protein, because of the possibility to form ionic interactions. In contrast, it was not obvious to imagine how residue W_{245} in Rsa1p is playing an essential role. Therefore, to go one step further in the understanding of the Snu13p-Rsa1p interaction, we then used structural and modeling approaches.

$Rsa1p_{238-252}$ folds into an alpha helix

The solution structure of Snu13p had been solved by X-ray crystallography (15). Therefore, to generate a 3D structure model of the Snu13p-Rsa1p interaction by *in silico* docking experiments, we had to solve the 3D structure of an Rsa1p fragment containing the amino acids required for interaction with Snu13p. We first attempted to establish the 3D structure of the entire $Rsa1p_{230-266}$ fragment, but encountered strong solubility problems at the concentration used for NMR studies. Therefore, we determined the 3D structure of the peptide TDEDVKKWREERKKM spanning residues 238 to 252 (named $Rsa1p_{238-252}$). The solubility of this peptide was suitable to obtain well-resolved ^{1}H-NMR spectra, and it contained the three residues (W_{245}, R_{246}, and R_{249}) found to be important for the interaction with Snu13p. When using a C^{13}, N^{15} labelled peptide all nuclei can be assigned from NMR spectra, except for

carbonyl, carboxyl and guanidium groups and tertiary amine of lysine side chains. By applying this approach to Rsa1p$_{238-252}$, only the NMR frequencies of the backbone amide groups of residues T_{238} and D_{239} and of the C^{δ} and C^{ε} of residue K_{244} remained undetermined, meaning that we unambiguously assigned the rest of the ^{1}H, ^{13}C and ^{15}N nuclei. A schematic view of the NMR distances restraints and structural statistics for the calculation are displayed in figure S3 and table S1 (supplementary data). As shown by the superimposition of the 10 structures having the lowest energies (Figure 6A) and their low RMSD values, Rsa1p$_{238-252}$ is folded into a α-helix. Interestingly, the side chains of residues W_{245}, R_{246}, and R_{249} are positioned on the same face of the helix, in a configuration favourable for the establishment of interactions with Snu13p amino acids (Figure 6B). Conversely, the side chain of the residue E_{247}, which is not required for Snu13p binding (Figure 4B), is positioned on the opposite side of the helix (Figure 6B).

The 3D structure model of the Snu13p-Rsa1p$_{238-252}$ interaction explains the crucial roles of E_{72} in Snu13p and W_{245} and R_{249} in Rsa1p

Our docking experiments with Haddock2.1/CNS1.3 (57,58) were based on: *i*) the established crystal structure of Snu13p (PDB accession number: 2ALE, (15)), *ii*) our solution structure of Rsa1p$_{238-252}$, *iii*) our finding of a 1:1 molar stoichiometry of the Snu13p-Rsa1p complex, and *iv*) our knowledge on the amino acids important for the Snu13p-Rsa1p interaction: E_{72} in Snu13p, and W_{245} and R_{249} in Rsa1p$_{238-252}$ were defined as active residues for docking calculation. Results of the docking procedure are presented in Table 1.

According to energies and RMSD values calculated on the ten models with the lowest haddock scores, Rsa1p$_{238-252}$ adopts a preferential positioning on Snu13p in which the C-terminal part of the peptide is close to the EDK motif of the protein, whereas the N-terminal part accommodates into the groove formed by helix α3 and the N-terminal part of Snu13p (Figure 6C). More interestingly, as shown on the best model provided by Haddock, the carboxylate group of residue E_{72} in Snu13p, which plays a key role in the interaction, is able to form two electrostatic bonds with Rsa1p$_{239-252}$ basic residues (one with a guanidinium group of residue R_{246}, the other one with a guanidinium group of residue R_{249}) (figure 6D). In contrast, the side chain of residue D_{73} in Snu13p, which was found to play a less important role in the interaction, is predicted to form only one salt bridge with the guanidinium group of residue R_{246} in Rsa1p$_{239-252}$ (Figure 6D).This salt bridge is also the only one formed by its partner R_{246} residue (Figure 6D). Therefore, according to our 3D model, the two acidic residues (E_{72} and D_{73}) of the Snu13p EDK motif, and the two basic residues (R_{249} and R_{246}) of the Rsa1p PEP motif are all implicated in a network of ionic interactions. However, according to our experimental data, only residues E_{72} in Snu13p and R_{249} in Rsa1p, which are both involved in formation of two ion pairs, play a critical role for the interaction. Finally, the side chain of residue W_{245} in Rsa1p$_{238-252}$ points into a large pocket formed by the hydrophobic side chains of the Snu13p residues K_7, F_9, L_{65}, L_{69}, and E_{72} (Figure 6E). Hence, by developing hydrophobic interactions with residues of this pocket the side chain of residue W_{245} should strongly participate to the stability of the interaction. Altogether, this can explain the essential role of the Rsa1p residue W_{245} in formation of the interaction.

The stability of the proposed model in an aqueous medium has been tested through a 5 ns Molecular Dynamic simulation as described in Material and Methods. Analysis

of the trajectory calculated with NAMD is available as supplementary material (supplementary data Table S2). Briefly, with a mean RMSD value of 1.73 ± 0.27 Å, it shows first that the overall 3D structure of the model does not undergo drastic changes over the simulation. Secondly, that the network of ionic interactions still exists during simulation. Finally, the location of the side chain of W245 in the Snu13p pocket described above is stable over time.

To assess the stabilizing effect generated by binding of the Rsa1p W_{245} residue in the Snu13p hydrophobic pocket, we substituted one of the hydrophobic residues of this pocket, L69, by an alanine residue. The choice of L_{69} was dictated by its close vicinity to the Rsa1p W_{245} residue (supplementary data Table S2). The effect of the mutation on the Snu13p-Rsa1p$_{230-266}$ interaction was tested by EMSA experiments. As illustrated in figure 7, the L_{69}A mutation abolished the binding of Rsa1p$_{230-266}$ to the Snu13p-U14 RNA complex, which demonstrated experimentally the stabilization effect of the Snu13p residue L_{69} on the Snu13p- Rsa1p$_{230-266}$ interaction. These data were in good agreement with the finding by Y2H assay of a role of the NUFIP residue L_{71} (the counterpart of L_{69} in Snu13p) (Figure 1C) in the NUFIP-15.5K interaction.

The amino acids essential for the Rsa1p-Snu13p interaction are important for cell growth

Deletion of the *RSA1* gene in *S. cerevisiae* leads to a growth defect, especially at high temperature (64). Hence, we decided to test whether abolition of the Snu13p-Rsa1p interaction also generates some growth phenotype. To this end, we used a yeast strain containing the chromosomal *SNU13* ORF placed under the control of the conditional *GAL1* promoter. These cells were transformed with a centromeric pG1 plasmid allowing constitutive expression of the WT or mutant Snu13p protein and cell growth was monitored in liquid media (Figure 8A). In the presence of glucose in the medium (absence of expression of the genomic *SNU13* gene), a doubling time of 3.26 hours was observed for the ectopic expression of WT Snu13p from the pG1 plasmid. This doubling time was increased to 3.56 and 3.58 hours, respectively, when the $E_{72}D_{73}K_{74}$/AAA and E_{72}A Snu13p variants were expressed. Although, the increase was not strong, it was reproducibly observed, and supports the idea of a functional importance of the identified Snu13p-Rsa1p interaction, and in particular, of the Snu13p E72 residue, playing an essential role in this interaction. In contrast, the L_{69}/A mutations had no strong effect on cell growth, probably because other amino acids of the Snu13p hydrophobic pocket can still interact with the W245 Rsa1p residue. To complete the demonstration of a functional role of the Snu13-Rsa1p interaction, WT and variant Rsa1p proteins were expressed from a centromeric plasmid in an *RSA1* KO strain, and growth rates were measured (Figure 8B). The KO strain had a doubling time ~44% slower compared to WT cells. Ectopic expression of Rsa1p restored the wild-type growth rates (2.6 hours). In contrast, only a partial restoration of cell growth was obtained for the Rsa1p R_{249}/A and W_{245}/A variants that did not bind to Snu13p (doubling time of 3.24 and 3.06 hours, respectively). The R_{246}/A substitution in Rsa1p, which did not abolish binding to Snu13p, had a more limited effect on cell growth. Therefore, substitutions of the three amino acids which play a key role in the Snu13p-Rsa1p interaction all have a negative effect on yeast cell growth showing their functional importance *in vivo*.

The Rsa1p residue R249 essential for Snu13p-Rsa1p interaction is important for box C/D snoRNA stability

As defects in snoRNA assembly are known to lead to partial degradation of snoRNAs (2,3), we guessed that if the Snu13p-Rsa1p interaction plays an important role for snoRNP assembly, its abolition *in vivo* should lead to a decreased level of cellular snoRNAs. To test this hypothesis, we used Northern blot analyses to evaluate the steady state levels of a selected number of box C/D snoRNAs in the *RSA1*$^-$ *S. cerevisiae* strain Y190 carrying a *RSA1* gene disruption, its isogenic strain with an active *RSA1* gene, and in the *RSA1*$^-$ strain complemented by ectopic expression of WT or R_{249}/A Rsa1p protein. The selected snoRNAs included U14 which is involved in both pre-rRNA 2'-O methylation and its early endo-nucleolytic cleavage steps leading to 18S rRNA (65,66), and snoRNAs (U18, snR63, snR190, snR45, snR13) which are only involved in 2'-O methylations of the pre-rRNA (66). The results obtained confirmed that disruption of the *RSA1* gene is associated to a drastic decrease of the amounts of most of the tested snoRNAs (snR63, snR190, snR45, snR13, and U18 box C/D snoRNAs) (Figure 8C). Ectopic expression of Rsa1p WT restored normal snoRNA amounts, while the expression of the variant R_{249}/A Rsa1p protein did not. Among the tested RNAs, only U14 snoRNA showed no significant decrease in the absence of Rsa1p expression or upon expression of the Rsa1p R_{249}/A variant. Due to its essential role in18S rRNA production, U14 may have a life span higher than other box C/D snoRNAs tested, or the assembly of the U14 snoRNP is less dependent upon the Snu13p-Rsa1p interaction. In contrast, the decreased levels of snoRNAs snR63, snR190, snR45, snR13, and U18 are likely indicative of a decreased efficiency of their assembly into snoRNPs. Therefore, these data strongly support the idea of a functional importance of the identified Snu13p-Rsa1p interaction for box C/D snoRNP biogenesis and as a consequence for yeast cell growth.

Residue E72 in Snu13p is required for recruitment of the Pih1-Tah1p dimer on a snoRNA in *in vitro* assays

We previously showed that $Rsa1p_{230-381}$ interacts with Pih1p, an integral component of the R2TP complex (2). Therefore, we proposed that through this interaction with Pih1p and its interaction with Snu13p, Rsa1p tether the R2TP complex onto Snu13p bound to the snoRNA. Our identification of a Snu13p residue required for the Snu13p-Rsa1p interaction (E_{72}) opened the possibility to test experimentally the above statement. For this purpose, instead of using protein Pih1 alone, which has a strong tendency to aggregate when produced alone in *E. coli*, we used a $_{His6}$Tah1p-Pih1p complex obtained by co-expression in *E.coli*. This allowed us to purify the stable $_{His6}$Tah1p-Pih1p complex on Ni-Sepharose beads (67), and to use the bound complex to test its interaction with other components of the snoRNP assembly machinery. The bound Tah1p-Pih1p complex was incubated with P^{32} labelled U14 snoRNA, in the presence or the absence of on the one hand, the WT recombinant $Rsa1p_{230-381}$ C-terminal domain and, on the other hand, the WT recombinant Snu13p protein or its E74/A variant. Retention of U14 snoRNA on the beads was detected by autoradiography after electrophoresis on a denaturing gel. As shown in Figure 9A, very little amounts of RNA was co-purified with the $_{His6}$Tah1p-Pih1p in the absence or the presence of Snu13p alone, showing a weak capacity of the $_{His6}$Tah1p-Pih1p to form a stable complex with U14 snoRNA alone or the Snu13p-U14 complex. In contrast, in the presence of the $Rsa1p_{230-381}$ C-terminal domain, retention of the snoRNA was substantially improved, indicating that this domain of Rsa1p can mediate the recruitment of the Tah1p-Pih1p complex on the Snu13p-U14 complex. Accordingly, when using the E_{72}/A Snu13p variant which

does not interact with Rsa1p no U14 retention was detected on the beads coated with the $_{His6}$Tah1p-Pih1p complex (Panel B). Thus, the Snu13p-Rsa1p interaction, that we characterized here, is required for binding of the R2TP complex to pre-snoRNP complexes. This observation likely explains the decreased amounts of snoRNAs that we detected in yeast strains with an impaired Snu13p-Rsa1p interaction.

DISCUSSION

Several proteins have been proposed to participate in snoRNP biogenesis. However, the mechanism of action of only a few of them has been clearly established. NUFIP and its yeast homolog Rsa1p were identified as key factors in the biogenesis of RNPs containing proteins of the L7Ae family. They were found to interact with proteins 15.5K and Snu13p, respectively (2,35). Based on the scaffolding properties of proteins NUFIP and Rsa1p and their capability to interact with the R2TP complex, the 15.5K-NUFIP and Snu13p-Rsa1p interactions were proposed to facilitate the recruitment of the other snoRNP core proteins and of Hsp90 on nascent RNPs (2,35). Accordingly in the absence of NUFIP and Rsa1p, a marked decrease of the box C/D snoRNP levels had been observed (2). Here, by combination of structural, molecular and cellular analyses, we were able to define the Snu13p-Rsa1p interaction at the atomic level and to demonstrate for the first time its high functional importance for box C/D snoRNP assembly *in vivo*. Our data also bring insights on how NUFIP and Rsa1p can discriminate L7Ae members within the larger L30 protein super-family.

Structural analysis identifies electrostatic and Van der Waals forces in the Snu13p-Rsa1p interaction

By combining data from Y2H assays, site-directed mutagenesis, co-expression assays, NMR analysis and computer modeling, we built a comprehensive 3D structure model of the Snu13p-Rsa1p interaction. According to this model, two basic residues in Rsa1p and two acidic residues in Snu13p form a network of ionic interactions. However, only one of the Snu13p acidic residues (E_{72}) and one of the Rsa1p basic residues (R_{249}) are indispensable for the interaction. Similarly, in Y2H assays, only the 15.5K E_{74} residue in protein 15.5K is essential for interaction of 15.5K with full length NUFIP. Inspection of our 3D structure model brings some insight to this observation: interaction of E_{72} in Snu13p with one guanidinium group of each R_{249} and R_{246} residues in Rsa1p is likely more stable than the interaction of D_{73} in Snu13p with R_{249} in Rsa1p. Moreover, formation of the D_{73}(Snu13p)-R_{249}(Rsa1p) ion pair might depend upon a favorable orientation of the R_{249} side chain, which may be imposed by formation of the E_{72}(Snu13p)-R_{249}(Rsa1p) interaction. The lower functional importance of the D_{73}(Snu13p)-R_{249}(Rsa1p) interaction as compared to the E_{72}(Snu13p)-R_{246}+R_{249}(Rsa1p) interaction is evidenced by the similar growth rate defects of cells expressing the E_{72}A Snu13p variant and cells expressing the or $E_{72}D_{73}K_{74}$/AAA Snu13p variant. Similarly, interaction of R_{249} in Rsa1p with both D_{73} and E_{72} in Snu13p seems to be sufficient to ensure association of Snu13p with Rsa1p. In contrast, the R_{246}(Rsa1p)-E_{72}(Snu13p) interaction does not allow the binding of Snu13p to Rsa1p when the R_{249} residue in Rsa1p is mutated into an alanine. Accordingly, a low functional importance of the Rsa1p R_{246} residue is evidenced by the absence of growth phenotype of yeast cells expressing the Rsa1p R_{246}A variant.

In our 3D structure model, the heterocyclic indol of residue W_{245} is located in a hydrophobic pocket formed by the aliphatic side chains of residues K_7, F_9, L_{65}, L_{69}, and E_{72} of Snu13p. The strong importance of the

W_{245} Rsa1p residue for cell growth strengthens the importance of the corresponding Van der Waals interactions. We could reasonably assume that positioning of residue W_{245} in the Snu13p pocket anchors the PEP domain of Rsa1p at the Snu13p surface. Importantly, the Rsa1p residue W_{245} is conserved in NUFIP (W_{239}), and the Snu13p hydrophobic building blocks of Snu13p hydrophobic pocket are conserved in human 15.5K protein (Figure 1). This strongly suggests that Van der Waals interactions similar to those established by the yeast Rsa1p W_{245} residue are formed at the 15.5K-NUFIP interface. Accordingly, mutation of residue L_{71} in 15.5K had a negative effect on the PEP-15.5K. Although, in our *in vitro* assays, the L_{69}/A mutation in Snu13p also had a negative effect on the Snu13p-PEP interaction, residue L_{69} is not essential for yeast cell growth. It is reasonable to think that, the four other hydrophobic amino acids in the Snu13p pocket compensate for the L_{69}/A mutation. Another function of residue W_{245} in Rsa1p might be to orient the side chain of the Snu13p E_{72} residue in a conformation favoring its electrostatic interaction with residue R_{249} in Rsa1p. Therefore, we propose that the Snu13p-Rsa1p and 15.5K-NUFIP interactions are stabilized by interdependent ionic and Van der Waals interactions, requiring the presence of three residues which are conserved from yeast to human: E_{72} in Snu13p and its E_{74} homologue in 15.5K, R_{249} in Rsa1p and its R_{243} counterpart in NUFIP, and W_{245} in Psa1p and its W_{239} counterpart in NUFIP (Figure 1).

NUFIP and Rsa1p discriminate L7Ae members through helix α3

Prior to the present study, very little was known about amino acids required for discrimination of "L7Ae-like" proteins by the Rsa1p and NUFIP platform proteins. One previous study proposed the involvement of the 15.5K E_{124} and R_{125} residues in NUFIP recruitment (22,46). The present data invalidate this conclusion, because we show that the $E_{124}R_{125}$/SS mutant has the capability to interact efficiently with NUFIP and the small $NUFIP_{223-260}$ domain in Y2H assays (Mut I in Figure 1C). (22,46). The absence of conservation of the E_{124} and R_{125} 15.5K residues in protein Snu13p also makes it very unlikely that they participate to an interaction expected to be conserved from yeast to human. Thus, the effect previously observed was likely to be an indirect effect. In contrast, our data are in better agreement with previous results showing that the $E_{74}D_{75}K_{76}N_{77}$/RKMT mutation abrogates detection of the 15.5K-NUFIP association in a cellular extract (46). The large diversity of amino acid substitutions generated in this variant may have alter the protein structure, however, one of the four substitutions targeted the 15.5K residue (E_{74}) that we identified as the key player in the 15.5K-NUFIP interaction. In the present study, to preserve the protein structure the 15.5K and Snu13p EDK motifs were substituted by three alanine residues which are unable to establish stable interactions with other residues in the protein. As in addition, the EDK residues in 15.5K and Snu13p (15,16) do not develop internal interactions, no modification of the protein structure was expected upon their mutation into AAA and this was confirmed by our CD data. So that, we are confident that the effects we observed are directly due to the amino acid substitutions.

Therefore, our data point to an essential role of the helix α3 acidic residue (E_{74} in 15.5K and E72 in Snu13p) in recognition of proteins Snu13p and 15.5K by Rsa1p and NUFIP, respectively. Interestingly, residue E_{74} is conserved in proteins 15.5K, NHP2, and SBP2 of the L7Ae family (Figure 1B), which all interact with NUFIP (2). In contrast, this residue is not conserved in the members of the L30 super-family which do

not belong to the L7Ae family (Figure 1B) (68). Accordingly, data of co-expression and yeast two-hybrid assays failed to detect an interaction between NUFIP and other members of the L30 family that contain no counterparts of residue E_{74} ((2) and data not shown). This suggests that the presence of residue E_{74} is a good predictor of the interaction with NUFIP or Rsa1p.

Amino acid sequence alignments also show the replacement of L_{69} in Snu13p by a valine residue (V_{96}) in human NHP2 (Figure 1B). In agreement with an implication of the Snu13p L_{69} residue in Van der Waals interactions with residue W_{245} of Rsap1, NHP2 does not interact with PEP in two hybrid assays. However, the $V_{96}L$ substitution in NH2P allows the interaction (Figure 1C), and the L_{71}/V mutation in 15.5K decreases its interaction with $NUFIP_{223-260}$. Altogether, these data support the idea of a role of the interaction established by residues L_{69} (Snu13p) and L_{71}(15.5K) in discrimination of Snu13p and 15.5K from other L7Ae-like proteins. Accordingly, no counterparts of residue L_{69}/L_{71} is found in SBP2, and like NHP2, SBP2 does not bind to the PEP domain of NUFIP ($NUFIP_{223-260}$) (data not shown). This suggests that amino acids other than PEP amino acids are involved in binding of NUFIP to SBP2 and NHP2 (Figure 1).

Taken together, these observations may reflect an ancestral role of NUFIP which was restricted to the biogenesis of the U4 snRNP and box C/D snoRNPs, and further acquisition of a role of NUFIP in assembly of box H/ACA snoRNPs and SBP2 mRNPs.

A loss of Snu13p-Rsa1p or 15.5K-NUFIP interaction has marked negative effects *in cellulo*

For the first time, we show *in cellulo* effects of mutations abolishing the 15.5K-NUFIP and Snu13p-Rsa1p interactions. The 15.5K mutant protein unable to bind NUFIP shows a substantially altered intracellular localization, with a stronger cytoplasmic signal that suggests a partial defect in incorporation into U4 and C/D RNPs (Figure 2). In yeast, deletion of the *RSA1* gene was known to induce a marked temperature sensitive phenotype (64). However, as Rsa1p had been proposed to have several functions (64), it was difficult to assign the observed growth defect to one given Rsa1p function. Here, based on our detailed analysis of the Rsa1p-Snu13p interaction, we could create point mutations in Rsa1p and Snu13p and dissect the functional importance of electrostatic and Van der Waals interactions formed at the Snu13p-Rsa1p interface.

Point mutation of residue E_{72} in Snu13p increased the cell doubling time by a factor of about ~10%, while point mutations of residue R_{249} and W_{245} in Rsa1p increased the cell doubling time by factors of ~25% and ~ 18%, respectively. This indicates that the electrostatic interactions developed by residues E_{72}(Snu13p) and R_{249}(Rsa1p), as well as the Van der Waals interaction established by residue W_{245}(Rsa1p) play important roles *in vivo*. We can reasonably assume that this role consist in enabling the recruitment of Rsa1p on pre-snoRNPs and therefore the association of the R2TP complex and other snoRNP core proteins on pre-snoRNPs. In agreement with this statement, we showed that *in vitro* binding of the Tah1p-Pih1p dimer to pre-snoRNP is dependent on the presence of Rsa1p and that mutation of residue E_{72} in Snu13p impairs this binding (Figure 9). Also consistent with this statement, the level of C/D box snoRNAs after $R_{249}A$ substitution in Rsa1p was very similar to that found upon deletion of the *RSA1* gene. Almost no snR13 and snR63 snoRNAs were detected in the two types of mutated cells.

The stronger effect on yeast growth of mutations in Rsa1p compared to mutations in Snu13p suggests that the Rsa1p amino acids involved in the Snu13p binding may also be involved in interaction of Rsa1p with

other target proteins. Consistent with our data, substitution of the $E_{74}D_{75}K_{76}N_{77}$ motif in protein 15.5K by a RKMT motif was previously found to abolish the binding of a recombinant protein PRP31 to the U4-15.5K complex in cell extract, as well as of the box C/D snoRNP core proteins NOP56, NOP58, Fibrillarin, and of the R2TP components proteins TIP48 and TIP49 on snoRNA-15.5K complexes in cellular extract (22).

One L7Ae residue in the counterpart of the Snu13p EDK motif develop an electrostatic interaction with residue K289 in NOP5

NUFIP and Rsa1p, bind the 15.5K and Snu13p proteins on a face opposite to that needed for RNA binding. This is consistent with the possibility to form a ternary complex containing: the snoRNA, Snu13p (or 15.5K), Rsa1p (or NUFIP) and the other assembly factors. However, one important question is to know whether the core proteins Nop56p, Nop58p and Nop1p recruited by combined the action of Rsa1p and the R2TP complex can form a stable mature snoRNP in the presence of Rsa1p or whether Rsa1p has to leave the pre-snoRNP complex in the ultimate steps of the assembly process. One way to get insight into this question was to know whether the E_{72} residue in Snu13p and E_{74} residue in15.5K which are, respectively, required for interaction with Rsa1p and NUFIP are available in mature box C/D snoRNPs or are involved in snoRNP internal protein-protein interactions.

Up to now, no 3D structure of eukaryotic snoRNPs is available to look for availability of residue E_{72}/E_{74} in the mature particle. However, the crystal structure of an archaeal box C/D sRNP has been recently established (26) and archaeal and eukaryotes C/D box RNPs are very similar. The archaeal box C/D sRNPs also contain a protein of the L7Ae family (the name of this protein family originates from the name of the archaeal protein). Its 3D structure is very similar to those established for proteins Snu13p and 15.5K (refs). A counterpart of the EDK motif is found in helix α3 of archaeal L7Ae proteins (*i.e.* $D_{74}E_{75}K_{76}$ in the *S. solfataricus* L7Ae protein) (Figure S4A). Remarkably, the L7Ae amino acids D_{74} and E_{75} of this motif and residue L_{71} corresponding to the L_{69} residue of Snu13p, are found at the interface between L7Ae and NOP5 (the archaeal homolog of NOP56/Nop56p and NOP58/), and Fibrillarine (Figure S4B). Whereas, the L7Ae D_{74} residue appears to be available for interaction in the C/D box sRNP structure, its neighbour residue E_{75} forms a stable electrostatic interaction with the highly conserved residue K_{289} of NOP5 and therefore clearly participates in the sRNP architecture (Figure S4D). If we assume similar interaction between Snu13p and Nop56p or Nop58p as found in archaea, we can conclude that part of the Snu13p-Rsa1p interaction is likely partially or totally destabilized prior completion of snoRNP assembly.

SUPPLEMENTARY DATA

Supplementary Tables S1 and S2; Supplementary Figures S1, S2, S3, and S4.

ACCESSION NUMBERS

The resonance assignments have been deposited in the BMRB with accession codes 17556. The corresponding coordinates and restraints for structure calculations have been deposited in the PDB with accession code 2LSD.

ACKNOWLEDGEMENT

C. Bobo is acknowledged for her technical assistance. A. Krol, C. Allmang-Cura are thanked for sharing data on protein SBP2. B. Maigret is thanked for fruitful discussions on molecular dynamics.

FUNDING

This work was supported by the Agence Nationale de la Recherche [ANR-06-BLAN-0208], the ARC, the french Ministère de l'Enseignement Supérieur et de la Recherche [doctoral fellowships to B.R. and R.B.]; the Centre National de la Recherche Scientifique ; the Pôle de Recherche Scientifique et Technologique "Bioingénierie" of Région lorraine. The authors are grateful to the Service Commun de Biophysicochimie des Interactions Moléculaires (SCBIM) of Nancy University for access to the CD spectrometer.

REFERENCES

1. Allmang, C., Carbon, P. and Krol, A. (2002) The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. *RNA*, **8**, 1308-1318.
2. Boulon, S., Marmier-Gourrier, N., Pradet-Balade, B., Wurth, L., Verheggen, C., Jady, B.E., Rothe, B., Pescia, C., Robert, M.C., Kiss, T. *et al.* (2008) The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *J Cell Biol*, **180**, 579-595.
3. Watkins, N.J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C. and Luhrmann, R. (2000) A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, **103**, 457-466.
4. Fatica, A. and Tollervey, D. (2002) Making ribosomes. *Curr Opin Cell Biol*, **14**, 313-318.
5. Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145-148.
6. Wahl, M.C., Will, C.L. and Luhrmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701-718.
7. Jady, B.E., Darzacq, X., Tucker, K.E., Matera, A.G., Bertrand, E. and Kiss, T. (2003) Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *EMBO J*, **22**, 1878-1888.
8. Blackburn, E.H. and Collins, K. (2011) Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol*, **3**.
9. Kiss, T., Fayet-Lebaron, E. and Jady, B.E. (2010) Box H/ACA small ribonucleoproteins. *Mol Cell*, **37**, 597-606.
10. Copeland, P.R., Fletcher, J.E., Carlson, B.A., Hatfield, D.L. and Driscoll, D.M. (2000) A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J*, **19**, 306-314.
11. Darzacq, X., Jady, B.E., Verheggen, C., Kiss, A.M., Bertrand, E. and Kiss, T. (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J*, **21**, 2746-2756.
12. Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R. and Luhrmann, R. (1999) Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J*, **18**, 6119-6133.
13. Henras, A., Henry, Y., Bousquet-Antonelli, C., Noaillac-Depeyre, J., Gelugne, J.P. and Caizergues-Ferrer, M. (1998) Nhp2p and Nop10p are

essential for the function of H/ACA snoRNPs. *EMBO J*, **17**, 7078-7090.
14. Wang, C. and Meier, U.T. (2004) Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J*, **23**, 1857-1867.
15. Oruganti, S., Zhang, Y. and Li, H. (2005) Structural comparison of yeast snoRNP and spliceosomal protein Snu13p with its homologs. *Biochem Biophys Res Commun*, **333**, 550-554.
16. Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R. and Ficner, R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell*, **6**, 1331-1342.
17. Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, **289**, 905-920.
18. Clery, A., Senty-Segault, V., Leclerc, F., Raue, H.A. and Branlant, C. (2007) Analysis of sequence and structural features that identify the B/C motif of U3 small nucleolar RNA as the recognition site for the Snu13p-Rrp9p protein pair. *Mol Cell Biol*, **27**, 1191-1206.
19. Marmier-Gourrier, N., Clery, A., Senty-Segault, V., Charpentier, B., Schlotter, F., Leclerc, F., Fournier, R. and Branlant, C. (2003) A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA*, **9**, 821-838.
20. Nottrott, S., Urlaub, H. and Luhrmann, R. (2002) Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J*, **21**, 5527-5538.
21. Granneman, S., Pruijn, G.J., Horstman, W., van Venrooij, W.J., Luhrmann, R. and Watkins, N.J. (2002) The hU3-55K protein requires 15.5K binding to the box B/C motif as well as flanking RNA elements for its association with the U3 small nucleolar RNA in Vitro. *J Biol Chem*, **277**, 48490-48500.
22. Schultz, A., Nottrott, S., Watkins, N.J. and Luhrmann, R. (2006) Protein-protein and protein-RNA contacts both contribute to the 15.5K-mediated assembly of the U4/U6 snRNP and the box C/D snoRNPs. *Mol Cell Biol*, **26**, 5146-5154.
23. Watkins, N.J., Dickmanns, A. and Luhrmann, R. (2002) Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol*, **22**, 8342-8352.
24. Schultz, A., Nottrott, S., Hartmuth, K. and Luhrmann, R. (2006) RNA structural requirements for the association of the spliceosomal hPrp31 protein with the U4 and U4atac small nuclear ribonucleoproteins. *J Biol Chem*, **281**, 28278-28286.
25. Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Luhrmann, R., Carlomagno, T. and Wahl, M.C. (2007) Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science*, **316**, 115-120.
26. Lin, J., Lai, S., Jia, R., Xu, A., Zhang, L., Lu, J. and Ye, K. (2011) Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature*, **469**, 559-563.

27. Xue, S., Wang, R., Yang, F., Terns, R.M., Terns, M.P., Zhang, X., Maxwell, E.S. and Li, H. (2010) Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Mol Cell*, **39**, 939-949.
28. Ye, K., Jia, R., Lin, J., Ju, M., Peng, J., Xu, A. and Zhang, L. (2009) Structural organization of box C/D RNA-guided RNA methyltransferase. *Proc Natl Acad Sci U S A*, **106**, 13808-13813.
29. Baker, D.L., Youssef, O.A., Chastkofsky, M.I., Dy, D.A., Terns, R.M. and Terns, M.P. (2005) RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev*, **19**, 1238-1248.
30. Charpentier, B., Muller, S. and Branlant, C. (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res*, **33**, 3133-3144.
31. Omer, A.D., Ziesche, S., Ebhardt, H. and Dennis, P.P. (2002) In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc Natl Acad Sci U S A*, **99**, 5289-5294.
32. Tran, E.J., Zhang, X. and Maxwell, E.S. (2003) Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J*, **22**, 3930-3940.
33. Chari, A. and Fischer, U. (2010) Cellular strategies for the assembly of molecular machines. *Trends Biochem Sci*, **35**, 676-683.
34. Matera, A.G., Terns, R.M. and Terns, M.P. (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol*, **8**, 209-220.
35. McKeegan, K.S., Debieux, C.M., Boulon, S., Bertrand, E. and Watkins, N.J. (2007) A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. *Mol Cell Biol*, **27**, 6782-6793.
36. Zhao, R., Kakihara, Y., Gribun, A., Huen, J., Yang, G., Khanna, M., Costanzo, M., Brost, R.L., Boone, C., Hughes, T.R. *et al.* (2008) Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. *J Cell Biol*, **180**, 563-578.
37. Neuenkirchen, N., Chari, A. and Fischer, U. (2008) Deciphering the assembly pathway of Sm-class U snRNPs. *FEBS Lett*, **582**, 1997-2003.
38. Pellizzoni, L., Yong, J. and Dreyfuss, G. (2002) Essential role for the SMN complex in the specificity of snRNP assembly. *Science*, **298**, 1775-1779.
39. Massenet, S., Pellizzoni, L., Paushkin, S., Mattaj, I.W. and Dreyfuss, G. (2002) The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. *Mol Cell Biol*, **22**, 6533-6541.
40. Darzacq, X., Kittur, N., Roy, S., Shav-Tal, Y., Singer, R.H. and Meier, U.T. (2006) Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J Cell Biol*, **173**, 207-218.
41. Kittur, N., Darzacq, X., Roy, S., Singer, R.H. and Meier, U.T. (2006) Dynamic association and localization of human H/ACA RNP proteins. *RNA*, **12**, 2057-2062.
42. Ballarino, M., Morlando, M., Pagano, F., Fatica, A. and Bozzoni, I. (2005) The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in

Saccharomyces cerevisiae. *Mol Cell Biol*, **25**, 5396-5403.

43. Leulliot, N., Godin, K.S., Hoareau-Aveilla, C., Quevillon-Cheruel, S., Varani, G., Henry, Y. and Van Tilbeurgh, H. (2007) The box H/ACA RNP assembly factor Naf1p contains a domain homologous to Gar1p mediating its interaction with Cbf5p. *J Mol Biol*, **371**, 1338-1353.
44. Li, S., Duan, J., Li, D., Ma, S. and Ye, K. (2011) Structure of the Shq1-Cbf5-Nop10-Gar1 complex and implications for H/ACA RNP biogenesis and dyskeratosis congenita. *EMBO J*, **30**, 5010-5020.
45. Walbott, H., Machado-Pinilla, R., Liger, D., Blaud, M., Rety, S., Grozdanov, P.N., Godin, K., van Tilbeurgh, H., Varani, G., Meier, U.T. *et al.* (2011) The H/ACA RNP assembly factor SHQ1 functions as an RNA mimic. *Genes Dev*, **25**, 2398-2408.
46. McKeegan, K.S., Debieux, C.M. and Watkins, N.J. (2009) Evidence that the AAA+ proteins TIP48 and TIP49 bridge interactions between 15.5K and the related NOP56 and NOP58 proteins during box C/D snoRNP biogenesis. *Mol Cell Biol*, **29**, 4971-4981.
47. Zhao, R., Davey, M., Hsu, Y.C., Kaplanek, P., Tong, A., Parsons, A.B., Krogan, N., Cagney, G., Mai, D., Greenblatt, J. *et al.* (2005) Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell*, **120**, 715-727.
48. Diebold, M.L., Fribourg, S., Koch, M., Metzger, T. and Romier, C. (2011) Deciphering correct strategies for multiprotein complex assembly by co-expression: Application to complexes as large as the histone octamer. *J Struct Biol*.
49. Fribourg, S., Romier, C., Werten, S., Gangloff, Y.G., Poterszman, A. and Moras, D. (2001) Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in E. coli. *J Mol Biol*, **306**, 363-373.
50. Romier, C., Ben Jelloul, M., Albeck, S., Buchwald, G., Busso, D., Celie, P.H., Christodoulou, E., De Marco, V., van Gerwen, S., Knipscheer, P. *et al.* (2006) Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. *Acta Crystallogr D Biol Crystallogr*, **62**, 1232-1242.
51. Bitter, G.A. and Egan, K.M. (1984) Expression of heterologous genes in Saccharomyces cerevisiae from vectors utilizing the glyceraldehyde-3-phosphate dehydrogenase gene promoter. *Gene*, **32**, 263-274.
52. Stutz, F. and Rosbash, M. (1994) A functional interaction between Rev and yeast pre-mRNA is related to splicing complex formation. *EMBO J*, **13**, 4096-4104.
53. Boulon, S., Pradet-Balade, B., Verheggen, C., Molle, D., Boireau, S., Georgieva, M., Azzag, K., Robert, M.C., Ahmad, Y., Neel, H. *et al.* (2010) HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. *Mol Cell*, **39**, 912-924.
54. Keller, R.L.J. (2004) *Computer-Aided Resonance Assignment Tutorial*. First edition ed. CANTINA Verlag, Zürich.
55. Guntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol*, **278**, 353-378.
56. Brunger, A.T. (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc*, **2**, 2728-2733.

57. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905-921.
58. de Vries, S.J., van Dijk, A.D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T. and Bonvin, A.M. (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*, **69**, 726-733.
59. Dominguez, C., Boelens, R. and Bonvin, A.M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*, **125**, 1731-1737.
60. Koradi, R., Billeter, M. and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph*, **14**, 51-55, 29-32.
61. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L. and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem*, **26**, 1781-1802.
62. MacKerell, A.D., Jr., Banavali, N. and Foloppe, N. (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, **56**, 257-265.
63. Janicki, S.M., Tsukamoto, T., Salghetti, S.E., Tansey, W.P., Sachidanandam, R., Prasanth, K.V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R.H. *et al.* (2004) From silencing to gene expression: real-time analysis in single cells. *Cell*, **116**, 683-698.
64. Kressler, D., Doere, M., Rojo, M. and Linder, P. (1999) Synthetic lethality with conditional dbp6 alleles identifies rsa1p, a nucleoplasmic protein involved in the assembly of 60S ribosomal subunits. *Mol Cell Biol*, **19**, 8633-8645.
65. Liang, W.Q. and Fournier, M.J. (1995) U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. *Genes Dev*, **9**, 2433-2443.
66. Piekna-Przybylska, D., Decatur, W.A. and Fournier, M.J. (2007) New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. *RNA*, **13**, 305-312.
67. Eckert, K., Saliou, J.M., Monlezun, L., Vigouroux, A., Atmane, N., Caillat, C., Quevillon-Cheruel, S., Madiona, K., Nicaise, M., Lazereg, S. *et al.* (2010) The Pih1-Tah1 cochaperone complex inhibits Hsp90 molecular chaperone ATPase activity. *J Biol Chem*, **285**, 31304-31312.
68. Koonin, E.V., Bork, P. and Sander, C. (1994) A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? *Nucleic Acids Res*, **22**, 2166-2167.
69. DeLano, W.L. (2010), *The PyMOL Molecular Graphics System, Version 1.3r1*. Schrödinger, LLC, New York.

TABLE AND FIGURES LEGENDS

Figure 1. Residues in 15.5K helix α3 are involved in the 15.5K-NUFIP$_{223-260}$ interaction. **(A)** Schematic representation of the human protein NUFIP and the yeast protein Rsa1p. The homologous domain between the two proteins (named hPEP in human and yPEP in yeast) is colored in gray. The various fragments of NUFIP and Rsa1p used in this study are represented. **(B)** The

RNA binding domain shared by members of the L7Ae family. Multiple amino acid sequence alignments were constructed using CLUSTALW. Secondary structure elements of human protein 15.5K are labeled on top of the sequences. Numbering corresponds to residues of 15.5K. Identical residues are highlighted in light grey and similar residues are in dark grey. **(C)** Yeast two hybrid interactions between 15.5K and NUFIP, $NUFIP_{223-260}$ or Alix as a negative control. Tests were also performed with NHP2, the ore protein of box H/ACA snoRNPs. The following residues of 15.5K were changed into alanines except when indicated. MutA: $H_{17}L_{18}T_{19}K_{20}K_{21}$; MutB: $K_{33}Q_{34}L_{35}R_{36}$; MutC: $E_{61}P_{62}E_{64}$; MutD: $E_{74}D_{75}K_{76}$; MutE: $R_{84}S_{85}K_{86}$; MutF: $K_{107}E_{108}K_{113}A_{114}$; MutG: $Q_{27}Q_{28}$; MutH: $N_{31}Y_{32}K_{33}$; MutI: E124S R125S; MutJ: $N_{40}T_{43}N_{47}$.

Figure 2. 15.5K $E_{74}A$ does not bind NUFIP *in vivo* and is not efficiently incorporated into RNPs. U2OS and U2OS LacO cells were transfected with mRFP-LacI-NUFIP and either GFP-15.5K (upper two panels), or its EDK->AAA variant (bottom two panels). WT 15.5K accumulates in nucleoli, in Cajal bodies and is also diffusely present in the nucleoplasm (top left panel). Furthermore, when mRFP1-LacI-NUFIP accumulates at the LacO sites, WT 15.5K is also enriched there (see insets in the top panel). In contrast, while the mutant 15.5K is also present in nucleoli, Cajal bodies and in the nucleoplasm, it is also present in large amount in the cytoplasm, indicating a defect in RNP assembly (bottom left panel). In addition, this mutant fails to be recruited to the LacO sites by mRFP-LacI-NUFIP (bottom panels).

Figure 3. Functional analysis of the Snu13p-Rsa1p interaction. **(A)** Comparison of the 15.5K and Snu13p proteins. 3D structure representations of the complex formed between human protein 15.5K and the 5' stem-loop of U4 snRNA (PDB accession number: 1E7K) (16), and of the *S. cerevisiae* Snu13p protein (PDB accession number: 2ALE) (15) were obtained by using the PyMOL software (69). Residues of the COOH terminus of helix α3 are indicated by black letters. On the U4-15.5K structure, residues in pink interact with PRP31 in the U4-15.5K-$PRP31_{78-333}$ complex (25). Alanine substitutions performed in Snu13p are displayed in the insert. **(B)** Yeast two-hybrid assays between Snu13p and its variants used as the bait, and the preys Rsa1p or its minimal functional fragment $Rsa1p_{230-381}$. A control was performed with an empty vector (-). Growth on medium lacking histidine but containing 3-Amino-1,2,4-triazol (3AT), a competitive inhibitor of the product of the *HIS3* reporter gene, indicates an interaction. **(C)** Co-expression and co-purification assays. The wild type or mutated His-tagged protein Snu13p was co-expressed in *E. coli* with Rsa1p. His-pull down assays from the crude cellular extracts were performed by IMAC. The soluble fraction (Input, In) and the purified proteins (Eluate, E) were fractionated by SDS-PAGE and the gel stained with Coomassie blue. The ladder (L) is indicated on the left. **(D)** Gel-shift assays. Radiolabeled U14 snoRNA was incubated with the wild type Snu13p or one of its α3 helix variants and the indicated His-tagged $Rsa1p_{230-381}$ or $Rsa1p_{230-266}$. GST is used as a negative control.

Figure 4. Study of protein-protein interaction by circular dichroism. A two-compartment cuvette was used and spectra were recorded as described in Materials and Methods. **(A)** and **(B)**, Snu13p-$Rsa1p_{230-266}$ interaction. For each interaction experiments, the CD spectra recorded before mixing the samples are in dark blue whereas the spectra in light blue (WT Snu13p) or in dark green (variants) were obtained after mixing the two compartments.

Figure 5. Functional analysis of residues of peptide Rsa1p$_{230-266}$. **(A)** The amino acid sequence of Rsa1p$_{230-266}$, which is conserved between NUFIP and Rsa1p, was mutated. Alanine substitutions in the various Rsa1p variants are underlined. **(B)** Analysis of the substitutions by EMSA in the same conditions as in Figure 3D. **(C)** Yeast two-hybrid assays using Snu13p used as the bait, and WT or variant Rsa1p$_{230-381}$ fragments of Rsa1p as the prey. Analyses were performed as in Figure 2B. **(D)** Co-expression and co-purification assays. Same experiments as in Figure 3C but with $_{His6}$Snu13p co-expressed in *E. coli* with one of the Rsa1p or Rsa1p$_{230-381}$ variants.

Figure 6. NMR structure of Rsa1p$_{238-252}$, docking and modeling of the Snu13p-Rsa1p interaction. **(A)** Backbone superimposition of the ten best structures of Rsa1p$_{238-252}$. **(B)** Side chains representation of the lowest-energy Rsa1p$_{238-252}$ structure. The helical conformation is highlighted using a cartoon representation. The hydrogen atoms are not represented. **(C)** Ribbon representation of the ten 3D models with the lowest haddock score resulting from the docking of the lowest-energy NMR structure of Rsa1p$_{238-252}$ on the X-ray structure of Snu13p. **(D)** Ionic interaction network between R_{246} and R_{249} in Rsa1p$_{238-252}$ and E_{72} and D_{73} in Snu13p. Salt bridges are shown with black dash lines. **(E)** Highlight of the Snu13p pocket hosting the W_{245} in Rsa1p$_{238-252}$. The figure has been done with PyMOL software (69).

Figure 7: Role of residue L_{69} in the Snu13p-Rsa1p interaction. Formation of U14 RNA-Snu13p-Rsa1p$_{230-266}$ complexes were analyzed by gel-shift assays as in figure 3D.

Figure 8: Effect of substitutions within the Snu13p helix α3 or the Rsa1p$_{230-266}$ domain on cell growth. **(A)** Yeast strain YPH499-*GAL::SNU13* was transformed with various recombinant pG1::*SNU13* vectors and grown in medium containing glucose. Effect of the ectopic expression of Snu13p and its variant forms on cell growth was monitored by measurement of the absorbance at 600 nm and by maintaining cell in exponential growth by successive dilutions. **(B)** Effects of ectopic expression of WT and variant Rsa1p from recombinant pG1::*RSA1* plasmid on growth of a *RSA1* KO strain. **(C)** Northern blot analysis of box C/D snoRNAs steady-state levels in cells with ectopic expression of WT and R_{249}/A variant Rsa1p from recombinant pG1::*RSA1* plasmid. RNAs were specifically detected with antisens oligonucleotide probes.

Figure 9: *In vitro* interaction of Pih1p with the pre-RNP U14 snoRNA-Snu13p-Rsa1p$_{230-381}$. **(A)** His-pull down assay with the $_{His6}$Tah1p-Pih1p protein complex. The radiolabeled U14 snoRNA was incubated with the recombinant Snu13 protein and the $_{His6}$Tah1p-Pih1p dimer. The same experiment was realized in presence of an increasing concentration of the Rsa1p$_{230-381}$ fragment of Rsa1p. The mixture was incubated with nickel-sepharose beads to allow binding of $_{His6}$Tah1p-Pih1p, and retention of the RNA was analyzed by denaturing gel electrophoresis. The input corresponds to 10% of the total quantity of radiolabeled RNA that was used in the binding assays (on the left). **(B)** His-pull down performed in the same conditions with recombinant Snu13p WT and Snu13p E_{72}/A.

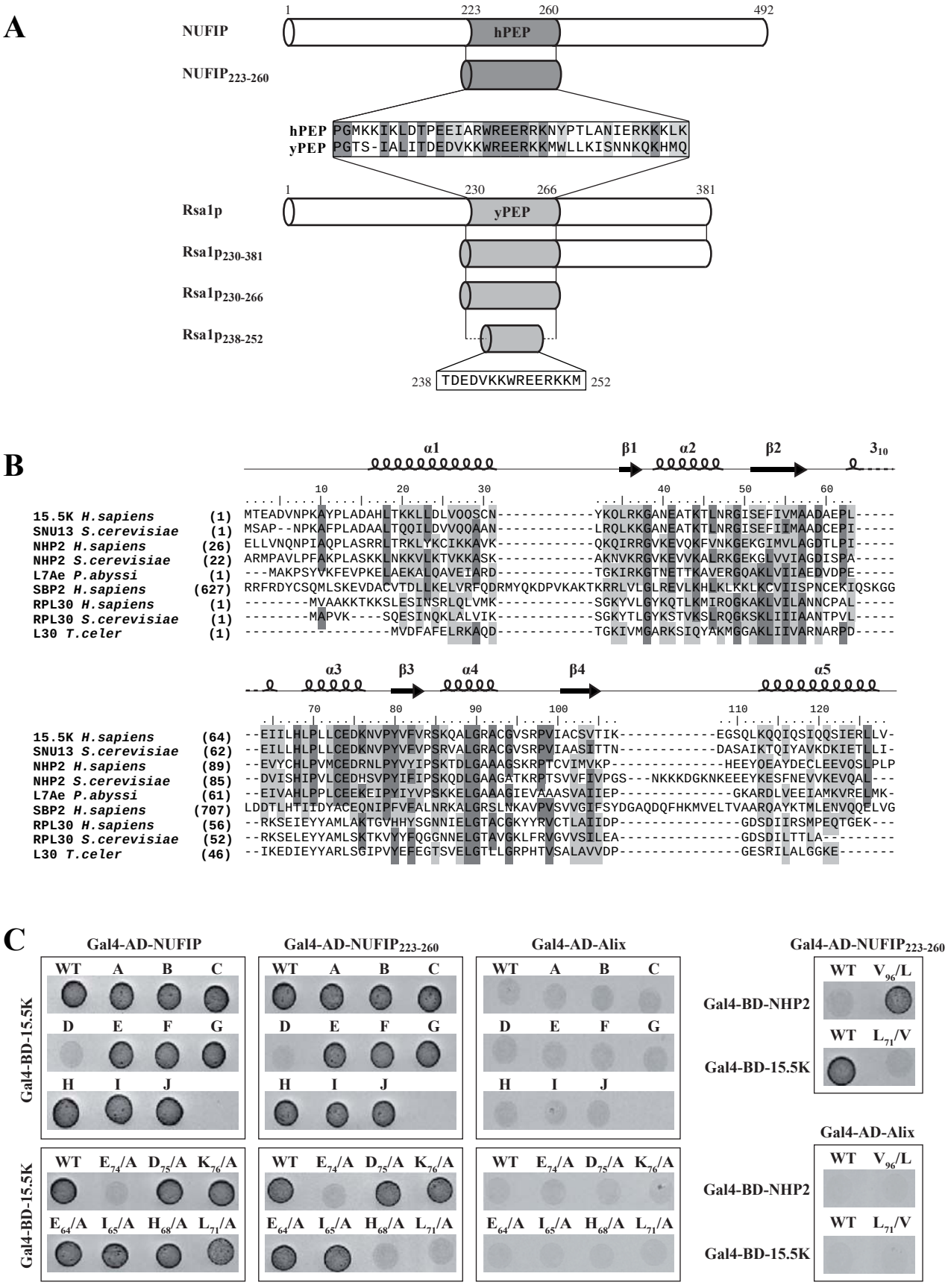


Figure 1 - Rothé *et al.*

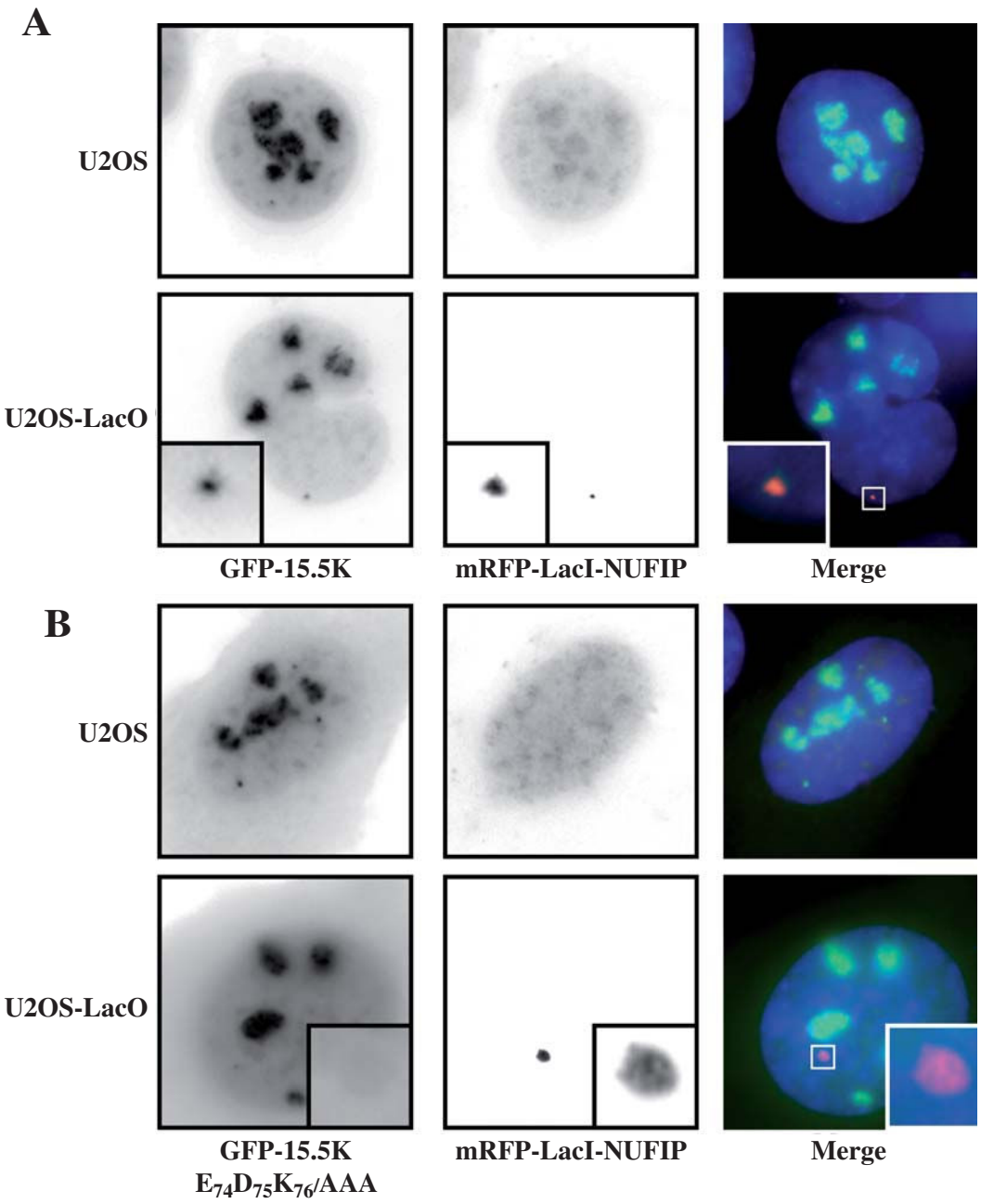
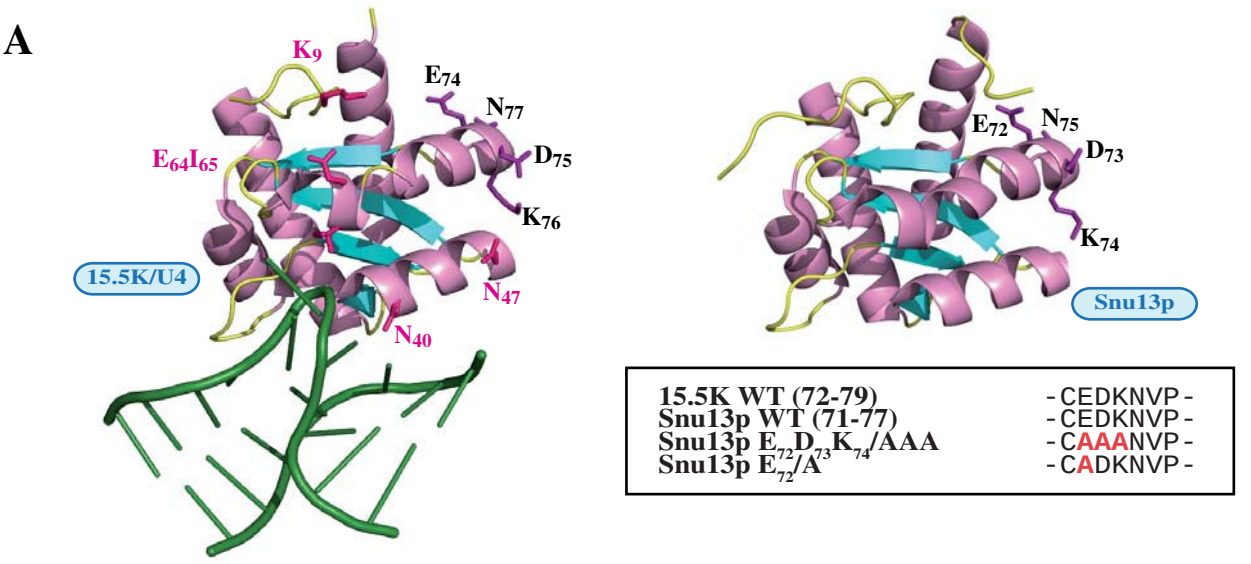
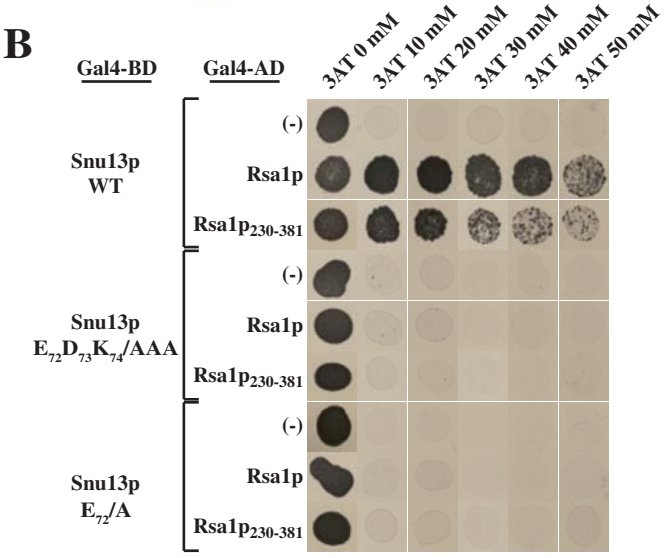
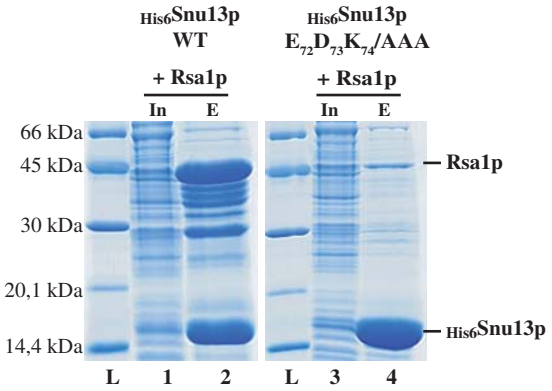


Figure 2 - Rothé *et al.*







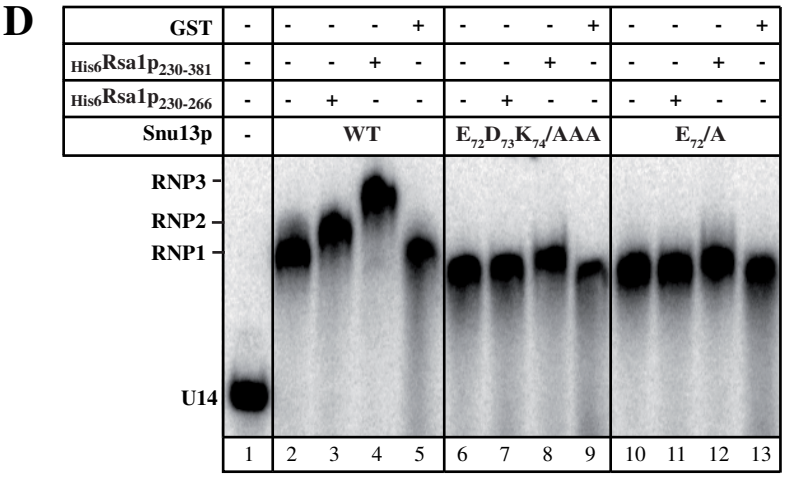


Figure 3 - Rothé *et al.*

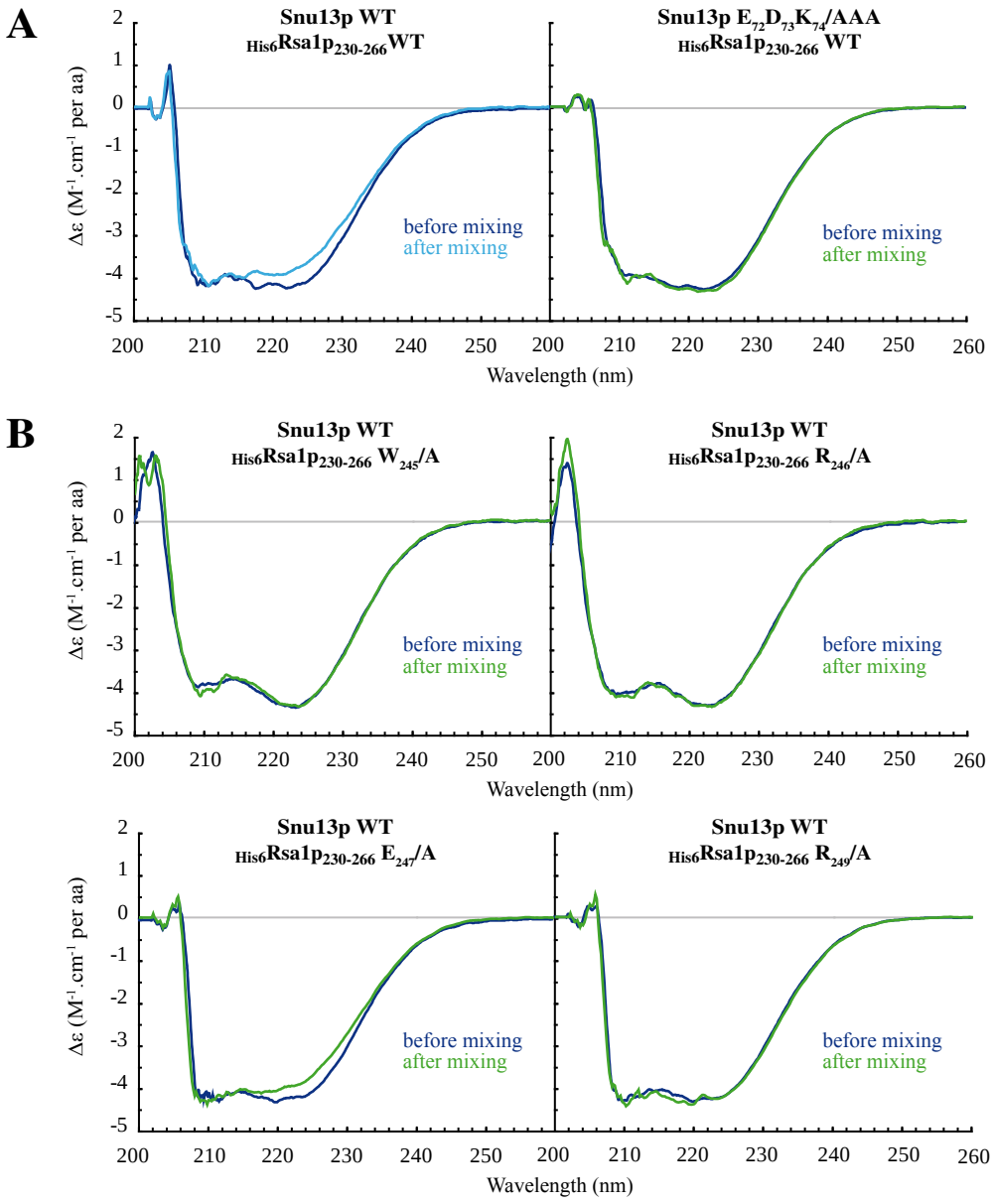


Figure 4 - Rothé *et al.*

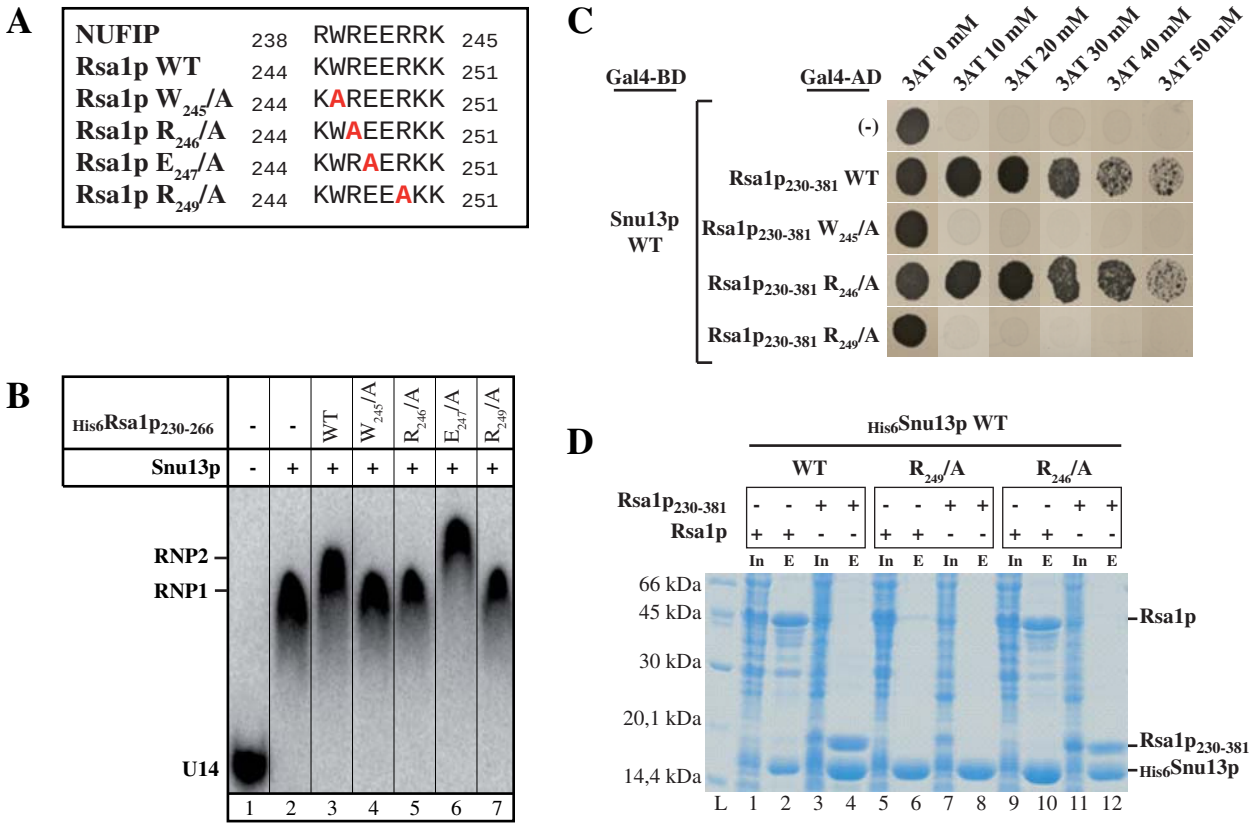


Figure 5 - Rothé *et al.*

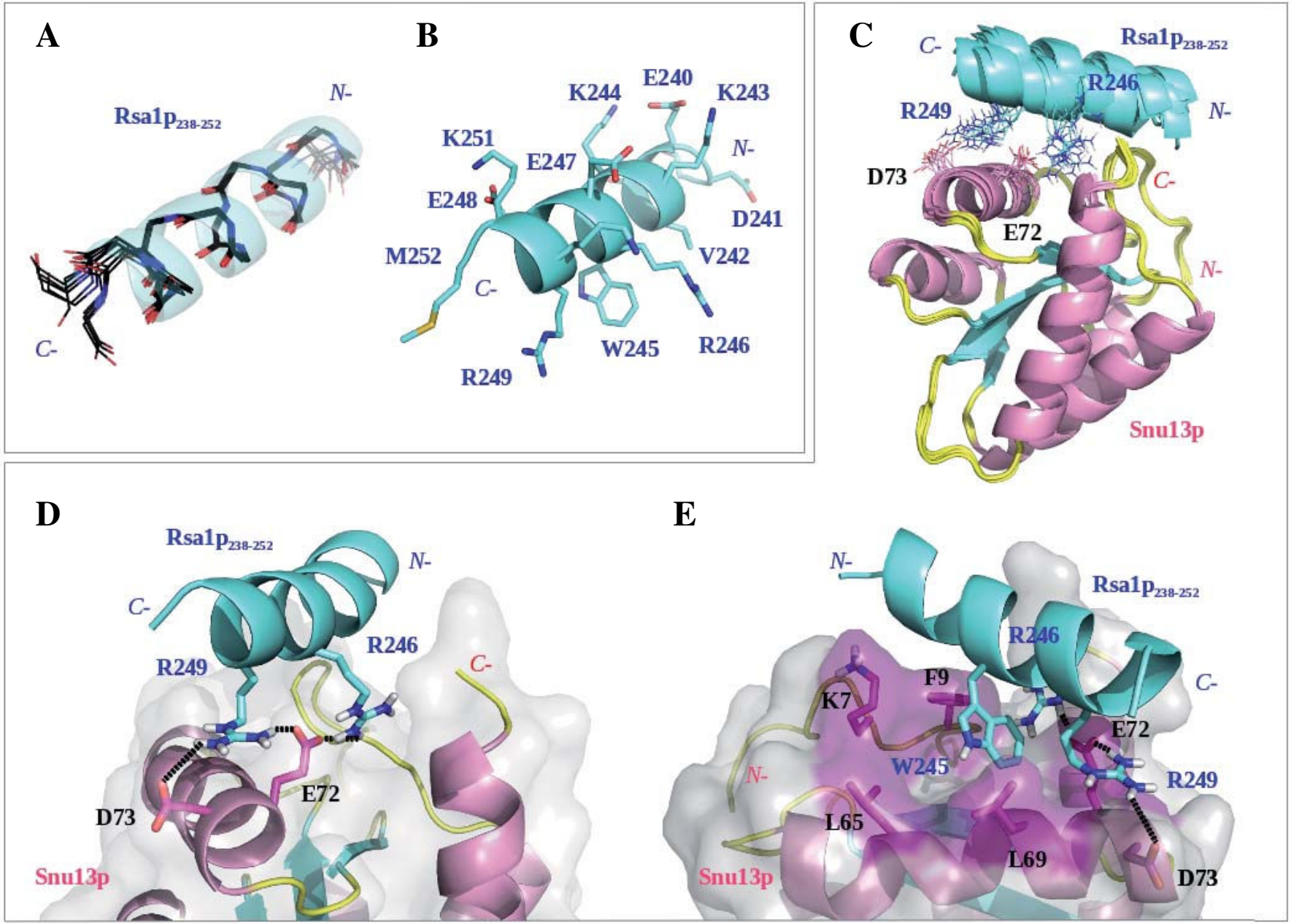


Figure 6 - Rothé *et al*.

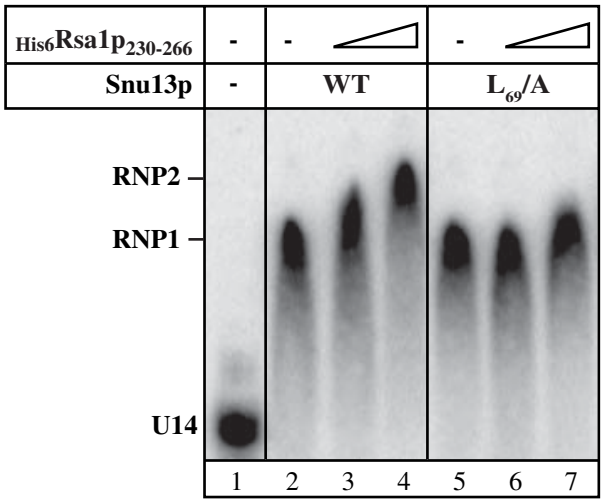


Figure 7 - Rothé *et al*.

A

Strain *Gal::SNU13*	G (h)	Growth defect (% WT)
pG1::*SNU13* -WT	3.26 ±0.04	
pG1::*SNU13* -$E_{72}D_{73}K_{74}$/AAA	3.56 ±0.05	-9.2
pG1::*SNU13* -E_{72}/A	3.58 ±0.09	-9.8
pG1::*SNU13* -L_{69}/A	3.22 ±0.09	+1.2

B

Strain Δ*RSA1*	G (h)	Growth defect (% WT)
pG1::(-)	3.74 ±0.18	-44.2
pG1::*RSA1*- WT	2.6 ±0.13	
pG1::*RSA1*- W_{245}/A	3.06 ±0.10	-17.8
pG1::*RSA1*- R_{246}/A	2.55 ±0.10	+1.7
pG1::*RSA1* -R_{249}/A	3.24 ±0.03	-24.8

C

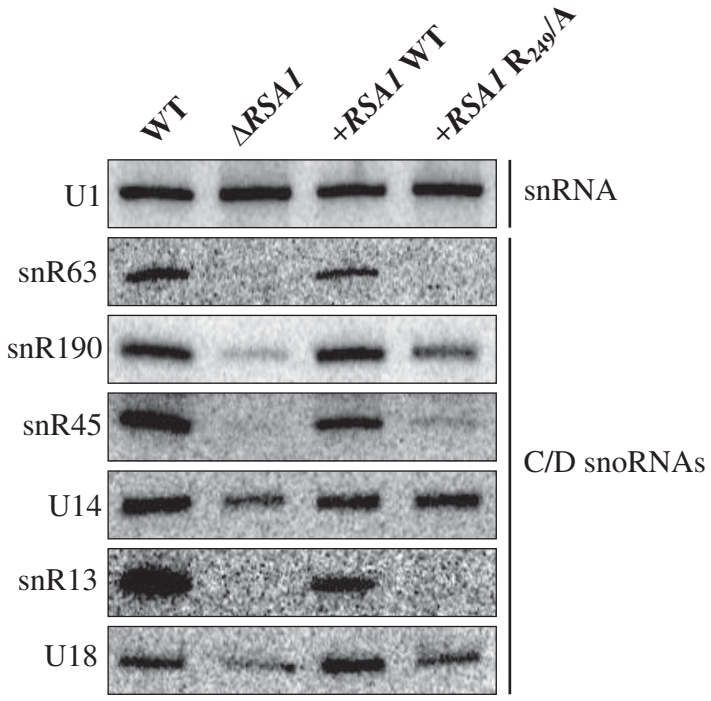


Figure 8 - Rothé *et al.*

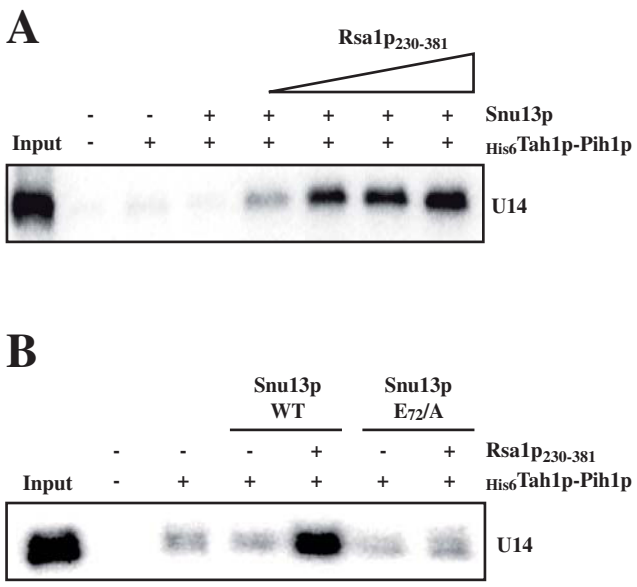


Figure 9 - Rothé *et al.*

	Snu13p	**Rsa1p$_{238-352}$**
Active residues	E72	W245 R249
Passive residues	L69 D73 N75	K243 K244 R246 E247 E248 K250 K251
Semi-flexible segments	L67 to P77	T238 to M252
Fully flexible segments	L67 to P77	W245 R249

	Docking results
RMSD from mean structure (Å)	
all backbone atoms	0.70 ± 0.19
all heavy atoms	0.99 ± 0.21
Energies (kcal/mol)	
E.inter	-341 ± 33
E.vdw	-21 ± 4
E.elec	-321 ± 31
Buried surface area (Å2)	922 ± 67
Number of AIR violation (>0.3Å)	0

Table 1- Rothé *et al*.

SUPPLEMENTARY DATA

SUPPLEMENTARY METHODS

GST pull-down assays

Recombinant fusion proteins GST-Snu13p, wild-type or variant $E_{72}D_{73}K_{74}$/AAA, (0.35 nmol), were bound on 15 µl of beads of Glutathione-Sepharose 4B (GE Healthcare©) in 100 µl of binding buffer (10 mM phosphate pH 7.0, 150 mM NaCl). A control was performed with beads bound by the GST protein. Recombinant $_{His6}Rsa1p_{230-266}$ (0.7 nmol) was added to the different beads samples, and the incubation was continued for 30 min at room temperature. Beads were washed three times with the binding buffer. Proteins were eluted by boiling in Laemmli loading buffer. Protein fractions were resolved by SDS-PAGE on 15% polyacrylamide gels and visualized by Coomassie staining.

SUPPLEMENTARY FIGURES (S1-S3)

Figure S1. GST-pull down experiments. A His-tagged version of the conserved peptide yPEP within the yeast Ras1p protein, named $_{His6}Rsa1p_{230-266}$ protein, was over-expressed in *E. coli* and purified. The obtained recombinant protein was assayed for binding to the indicated recombinant GST fusion proteins. The gel was stained with Coomassie blue.

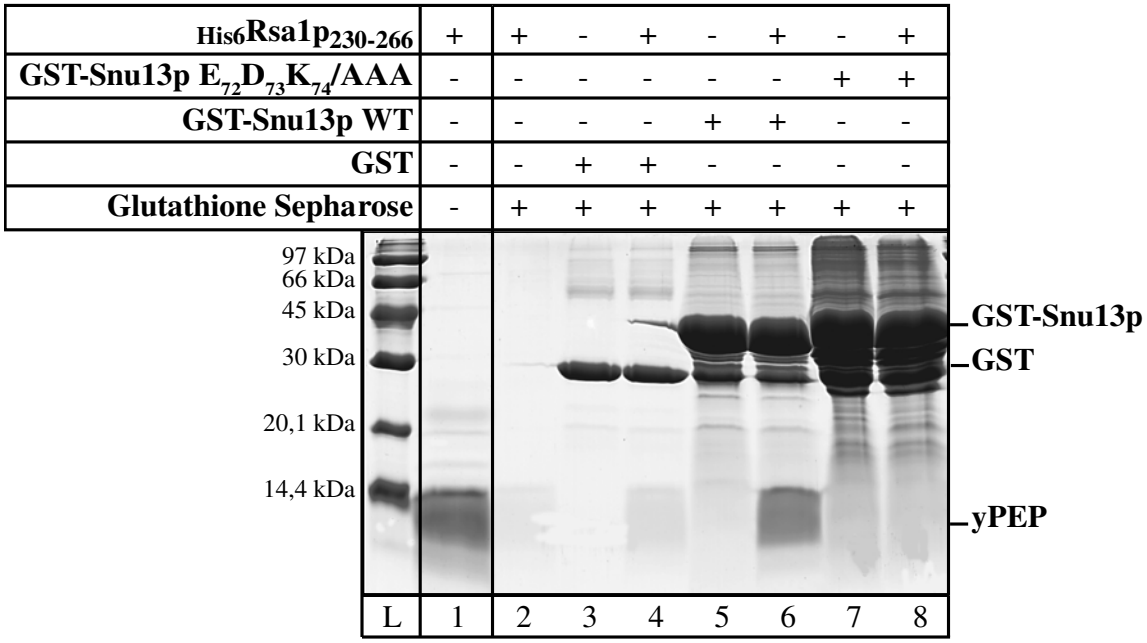


Figure S2. The PEP sequence. Multiple amino acid sequence alignments were constructed using CLUSTALW (1). Numbering corresponds to residues of the *S. cerevisiae* Rsa1p protein. Identical residues are highlighted in light grey.

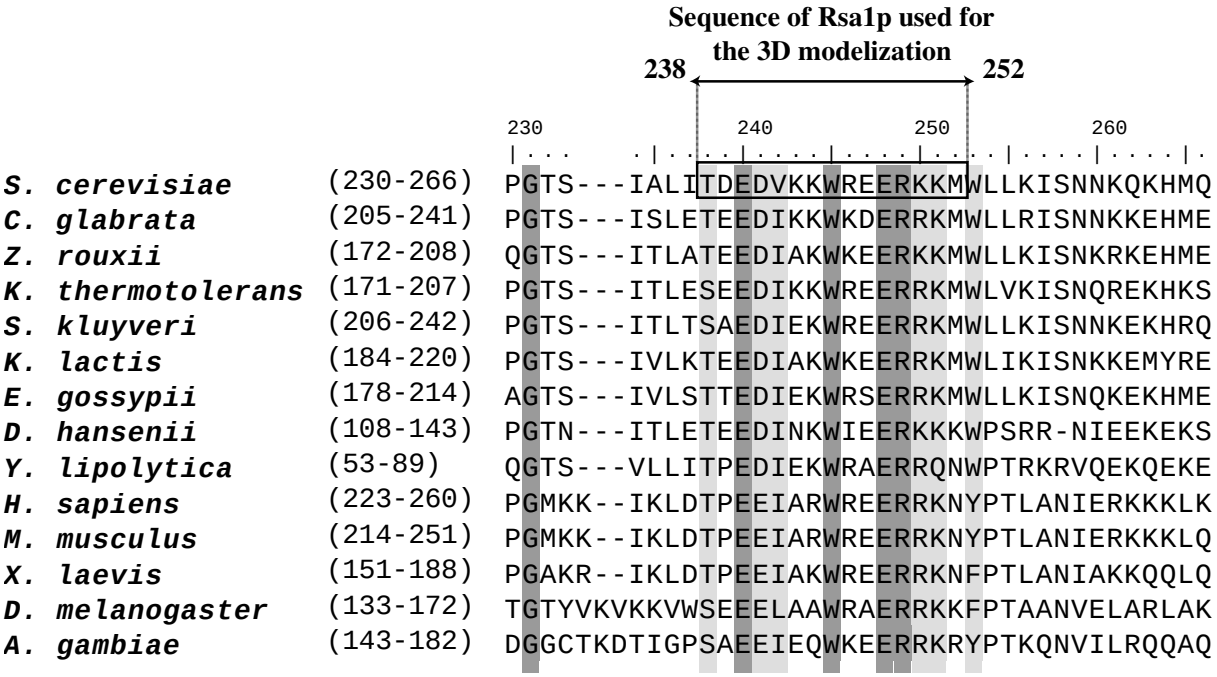


Figure S3. NOE connectivities for amide proton amide (N), alpha (α) and beta (β) protons of Rsa1p$_{238-252}$. The NOE intensities are represented by different line thicknesses, stronger lines meaning stronger intensities.

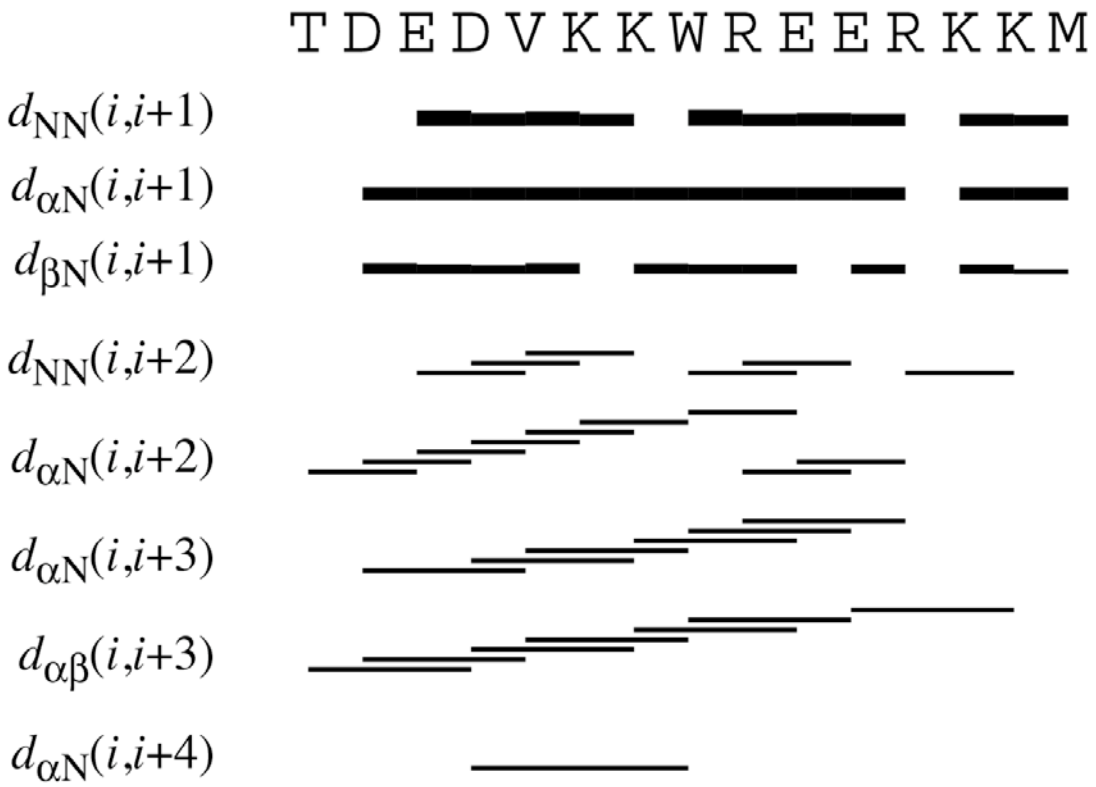


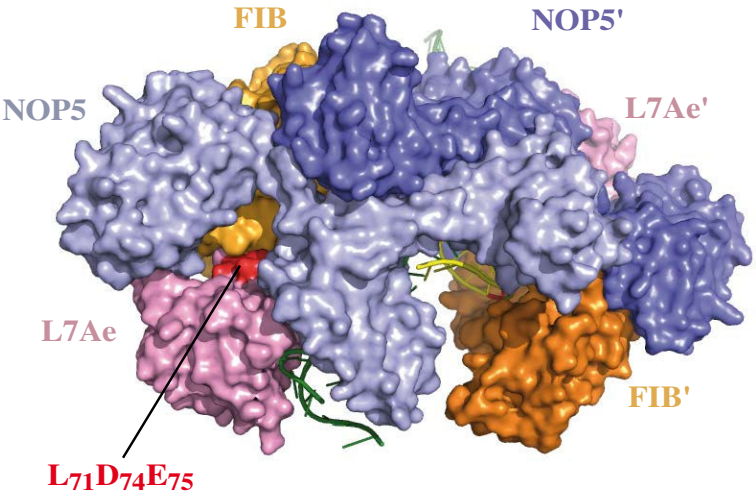
Figure S4. Comparison between archaeal box C/D sRNP and eukaryal box C/D snoRNP. (A) Multiple amino acid sequence alignments of the α3 helix in the eukaryal proteins 15.5K/Snu13p and the archaeal L7Ae proteins constructed using CLUSTALW. (B) *Sulfolobus solfataricus* box C/D sRNP 3D structure ((2); PDB : 3PLA). Amino acids L_{71}, D_{74} and E_{75} are indicated in red. (C) Interaction interface between L7Ae and NOP5 ((2); PDB: 3PLA). The black line represents an electrostatic interaction. (D) Multiple amino acid sequence alignments of a part of the eukaryal proteins NOP56, NOP58, PRP31 and the archaeal protein NOP5 constructed using CLUSTALW.

A

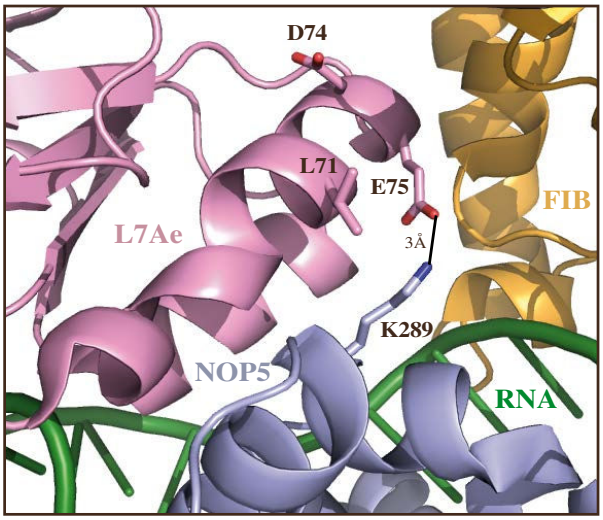
L7Ae/Snu13p/15.5K

15.5K	*H. sapiens*	59	DAEPLEIILHLPLLCEDKNVPYV
Snu13p	*S. cerevisiae*	57	DCEPIEILLHLPLLCEDKNVPYV
L7Ae	*P. abyssi*	56	DVDPEEIVAHLPPLCEEKEIPYI
L7Ae	*A. fulgidus*	52	DVDPPEIVAHLPLLCEEKNVPYI
L7ae	*M. jannaschii*	51	DVKPEEVVAHLPYLCEEKGIPYA
L7Ae	*S. solfataricus*	59	DVQPEEIVAHLPLLCDEKKIPYV

B



C



D

NOP5-NOP56/58-PRP31

NOP58	*H. sapiens*	308	LAKHAASTVQILGAEKALFRALK
Nop58p	*S. cerevisiae*	309	LAKSPASTIQILGAEKALFRALK
NOP58	*M. musculus*	308	LAKHAASTVQILGAEKALFRALK
NOP58	*X. laevis*	309	LAKHPASTVQILGAEKALFRALK
NOP58	*D. melanogaster*	309	LAKHPSSTVQILGAEKALFRALK
NOP56	*H. sapiens*	318	LAKYPASTVQILGAEKALFRALK
Nop56p	*S. cerevisiae*	325	LSKQAASTVQILGAEKALFRALK
NOP56	*M. musculus*	318	LAKYPASTVQILGAEKALFRALK
Nop56	*D. melanogaster*	319	LAKYPASTVQILGAEKALFRALK
NOP56	*X. laevis*	318	LAKYPASTVQILGAEKALFRALK
PRP31	*H. sapiens*	241	LSKMPACNIMLLGAQRKTLSGFS
Prp31p	*S. cerevisiae*	251	FSRIPSCNIASIGKNKHLSHELH
NOP5	*S. Solfataricus*	287	LAKMPASTIQVLGAEKALFRALR

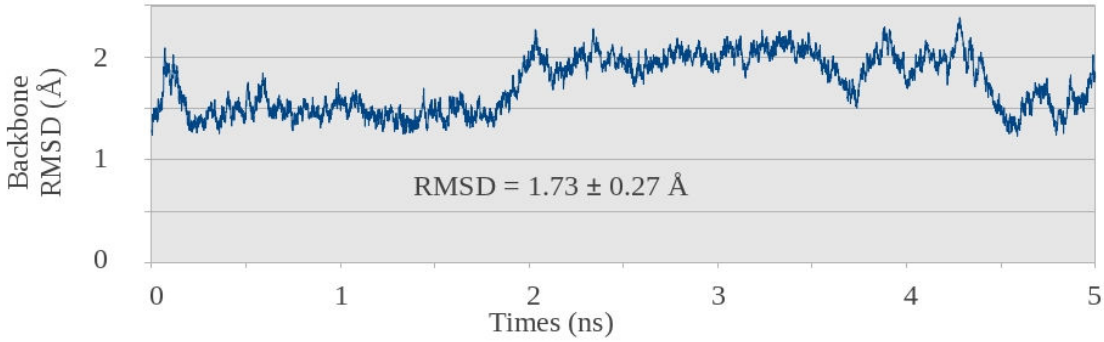
SUPPLEMENTARY TABLE (S1)

Table S1. NMR-Derived Geometrical Restraints and Structural Statistics of $Rsa1p_{238-252}$.

Restraints		Number
Distance restraints		
	total of distance restraints	169
	intra	43
	short	65
	medium	61
dihedral angles		
	ϕ	12
Violation		Occurrence
	distance violation> 0.2 Å	0
	dihedral violation> 5 degrees	0
Ramachandran statistics (PROCHECK-NMR)		
	residues in the most favorable region	96.2 %
	residues in additionally allowed regions	3.8 %
	residues in generously allowed regions	0 %
	residues in disallowed regions	0 %
Atoms (residues E_{240} to K_{250})		rmsd (Å)
	backbone	0.23 ± 0.07
	heavy	1.31 ± 0.42

Table S2. Analysis of the NAMD trajectory calculated on the Snu13p/$Rsa1p_{238-252}$ 3D model (3).
(A) Backbone RMSD calculated all along the trajectory. The reference structure was the top structure coming from HADDOCK (4,5). (B) Mean distance between $Rsa1p_{238-252}$ W_{245}-Cζ2 and Snu13p K_7-Cβ, F_9-Cζ, L_{65}-Cβ, L_{69}-Cγ and E72-Cβ all along the trajectory. (C) Salt bridges analysis between $Rsa1p_{238-352}$ R_{246} and R_{249} and Snu13p E_{72} and D_{73}. The mean distance has been measured between nitrogens of the guanidium group and oxygens of the carboxyl group. A cut-off of 3.8 Å between acceptor and donor atoms has been defined to evaluate the time occupancy of the salt bridges.

A



B

Rsa1p 238-352	Snu13p	Mean distance (Å)	Standard-Deviation
W245-Cζ2	K7 - Cβ	8.88	0.71
W245-Cζ2	F9 - Cζ	5.23	0.95
W245-Cζ2	L65 - Cβ	8.23	0.52
W245-Cζ2	L69 - Cγ	4.45	0.43
W245-Cζ2	E72 - Cβ	5.40	0.65

C

Rsa1p 238-352	Snu13p	Mean distance (Å)	Standard-Deviation	Time occupancy
R249	E72	4.99	1.20	19%
R249	D73	3.70	1.02	72%
R246	E72	3.51	0.25	89%

1. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.
2. Lin, J., Lai, S., Jia, R., Xu, A., Zhang, L., Lu, J. and Ye, K. (2011) Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature*, **469**, 559-563.
3. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L. and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem*, **26**, 1781-1802.
4. de Vries, S.J., van Dijk, A.D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T. and Bonvin, A.M. (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*, **69**, 726-733.
5. Dominguez, C., Boelens, R. and Bonvin, A.M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*, **125**, 1731-1737.

2 Données complémentaires sur le complexe Snu13p/Rsa1p

Lors de l'analyse à haut débit par co-expression des interactions entre facteurs d'assemblage des snoRNP C/D (Résultats & Discussion, Chapitre 1), nous avons mis en évidence une association forte entre les protéines Rsa1p (231-375) et Snu13p (Résultats & Discussion, Chapitre 1, paragraphe 2.3). Le rôle fonctionnel de cette interaction pour la production normale des snoRNA à boîtes C/D et pour la croissance cellulaire a clairement été mis en évidence dans le manuscrit inséré dans ce chapitre. Les données R.M.N. obtenues par Marc Quinternet décrites dans ce manuscrit ont permis d'établir la structure 3D du fragment 238-252 de la protéine Rsa1p et de montrer qu'il adopte une structure en hélice α. Comme jusque-là nous n'avions pas réussi à avoir un complexe assez stable entre notre fragment de Rsa1p et de Snu13p, l'association entre le fragment PEP de Rsa1p et Snu13p a été modélisée comme décrit dans le manuscrit à partir de cette structure du fragment 238-252 et de la structure 3D établie pour Snu13p.

Sur la base de nos données de co-expression, de modélisation et aussi des résultats de spectrométrie de masse nous avons produit et purifié de nombreux fragments de Rsa1p en complexe avec la protéine Snu13p. Aucun ne nous avaient permis d'obtenir des données de R.M.N. exploitables pour la détermination de la structure 3D. En fonction du fragment de Rsa1p utilisé pour la co-expression, nous avons rencontré des problèmes de reproductibilité de la purification, de précipitation lors de la concentration, ou de présence d'une conformation double de certains résidus de la protéine Snu13p en association avec le fragment de Rsa1p. Marc Quinternet a récemment commencé l'étude par R.M.N. du fragment 238-269 de la protéine Rsa1p synthétisé chimiquement en interaction avec la protéine Snu13p. Les premiers spectres indiquent l'absence d'équilibre entre plusieurs formes et une stabilité plus importante du complexe. Nous espérons donc pouvoir obtenir la structure 3D de ce complexe.

3 Conclusions et perspectives

3.1 Bilan

Les résultats présentés dans ce chapitre nous ont permis de décrire avec précision l'interface d'interaction entre Rsa1p et Snu13p. Elle semble être conservée dans le complexe formé par les protéines NUFIP/15.5K. Cette interaction met en jeu les résidus $W_{245}R_{246}$--R_{249}

de la protéine Rsa1p et le motif L_{69}--$E_{72}D_{73}$ de l'hélice α3 de la protéine Snu13p. Il s'établirait aussi un réseau de liaisons électrostatiques entre les deux arginines de Rsa1p et les résidus acides de Snu13p, ainsi que la formation de liaisons hydrophobes impliquant le résidu tryptophane de la protéine Rsa1p. De façon intéressante, la mutation ponctuelle des résidus E_{72} de Snu13p et R_{249} de Rsa1p, suffit à inhiber l'association des deux protéines. L'absence de cette interaction se caractérise *in vivo* par un défaut de croissance mais aussi par une diminution spécifique du taux de certains snoRNA à boîtes C/D (Manuscrit n°2).

3.2 La structure tridimensionnelle du complexe Snu13p/Rsa1p : élément de compréhension de la biogenèse des snoRNP à boîtes C/D

La détermination de la structure tridimensionnelle du complexe formé entre la protéine constitutive des snoRNP à boîtes C/D, Snu13p et le facteur d'assemblage Rsa1p est un enjeu majeur dans la compréhension des étapes de la biogenèse des snoRNP à boîtes C/D. Comme nous l'avons décrit dans le manuscrit, au niveau de la protéine Snu13p, un motif L--ED près de l'extrémité C-terminale de l'hélice α3 est impliqué dans son association avec la protéine Rsa1p (Manuscrit n°2) et cette interface est conservée chez l'homme (McKeegan *et al.*, 2009). Récemment, la structure tridimensionnelle d'une particule à boîtes C/D de *Sulfolobus sulfataricus* a été déterminée (Lin *et al.*, 2011). Il est intéressant de noter que le motif L_{72}--$E_{75}K_{76}$ de la protéine L7Ae (l'orthologue de Snu13p) se situe à l'interface avec les protéines Fibrillarine et aNop5 homologue d'archae de Nop1p et Nop56/58p. Les résidus chargés du motif L--EK établissent des interactions électrostatiques avec certains résidus de ces deux protéines notamment les résidus K289 de aNop5 et les résidus R135 et D158 de la Fibrillarine (Figure 96). Ce réseau d'interaction ressemble étonnamment à celui observé au niveau de l'interface Snu13p/Rsa1p et décrit dans la manuscrit n°2. Le motif L--ED/L--EK est donc clairement inaccessible à tout autre partenaire dans la particule mature finale.

Cette observation pose donc la question d'une mise en place d'un mécanisme séquentiel d'interaction dans un premier temps avec le facteur Rsa1p, suivit par le départ de ce facteur et son remplacement par les autres protéines « cœur ». L'obtention d'informations structurales sur les déterminants de l'interaction entre les protéines Snu13p et Rsa1p est donc très important pour comprendre les étapes successives de la biogenèse des snoRNP à boîtes C/D. C'est pour cela que l'objectif est la structure 3D d'un complexe formé entre Snu13p et

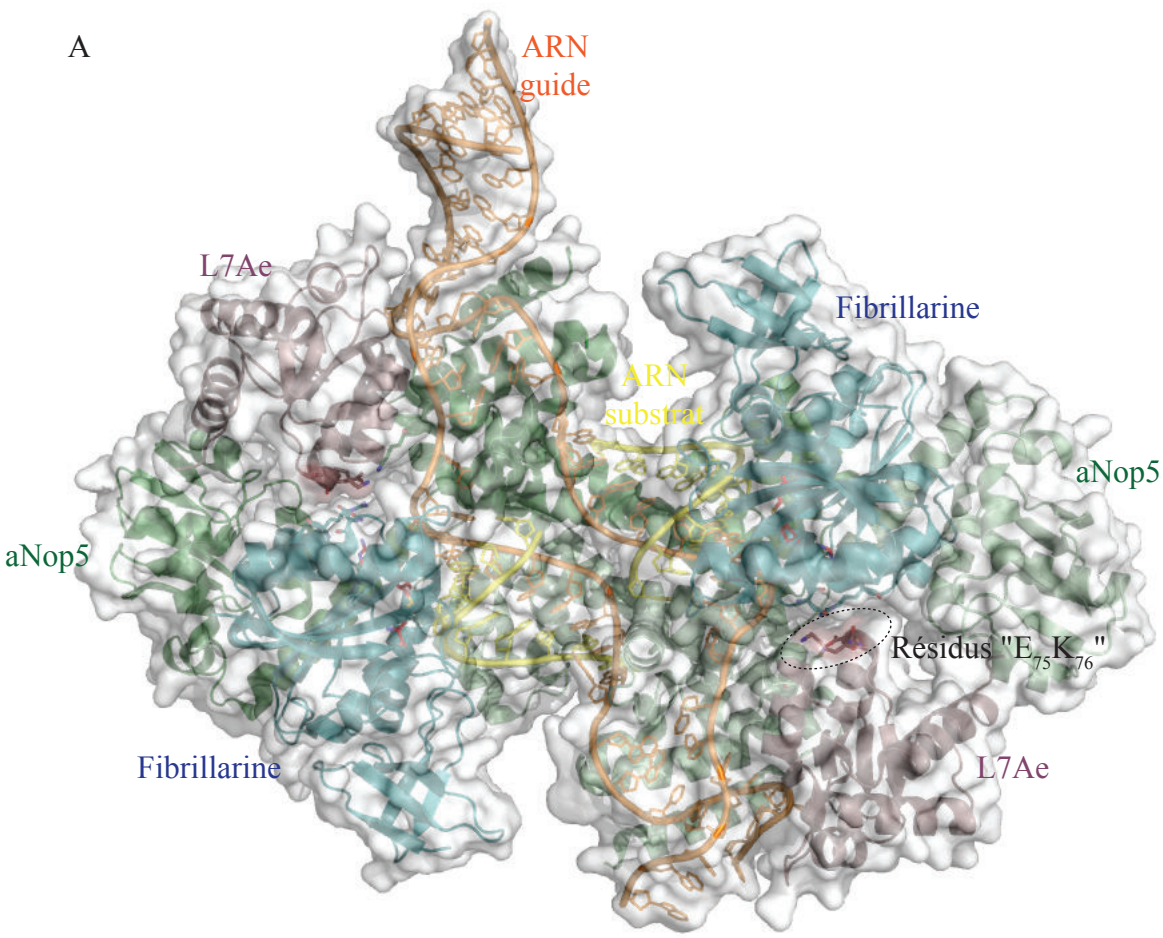


Figure 96. **Topologie des résidus chargés du motif L_{72}--$E_{75}K_{76}$ de l'extrémité C-terminale de l'hélice α3 de la protéine L7Ae dans le contexte de la structure entière de la particule C/D de *Sulfolobus sulfataricus*** (Lin *et al*., 2011). (A) Structure tridimensionnelle de la particule C/D de *Sulfolobus sulfataricus* représentée en ribbon avec une surface moléculaire en transparence. Les résidus E75 et K76 de la protéine L7Ae sont représentés en tige-bâton et projetés à la surface moléculaire. (B) Vue du réseau d'interaction electrostatique impliquant les résidus E75 et K76 de la protéine L7Ae avec les résidus K289 de la protéine aNop5 et R135/D158 de la Fibrillarine. (Lin *et al*, 2011 ; code PDB 3PLA).

un fragment de Rsa1p. L'obtention d'un complexe multimérique contenant la protéine Rsa1p devrait ensuite permettre de compléter l'analyse structurale de cette protéine plateforme qui a la capacité de se lier à de nombreux autres partenaires comme Hit1p ou encore Pih1p. Nous pourrons aussi envisager d'obtenir des informations sur les macrocomplexes formés par Rsa1p par emploi de technique de type S.A.X.S. (*« Small-angle X-ray scattering »*) ou encore par la cryo-microscopie électronique.

CHAPITRE 4

Structure tridimensionnelle de la protéine Tah1p et de son complexe avec l'extrémité C-terminale de la chaperonne Hsp90 : les déterminants de l'interaction

1 Présentation des travaux

En 2008, notre équipe en collaboration avec celle d'E. Bertrand a démontré l'implication de la chaperonne HSP90 dans la biogenèse des RNP de la famille L7Ae chez l'homme (Boulon *et al.*, 2008). Cette fonction implique vraisemblablement le complexe R2TP et notamment au sein de ce dernier, la protéine Tah1p, une protéine à domaine TPR (Boulon *et al.*, 2008; Zhao *et al.*, 2005; Zhao *et al.*, 2008). Les domaines TPR sont des motifs structuraux caractéristiques d'interaction protéine-protéine (Lamb *et al.*, 1995). Ils sont constitués de motifs TPR correspondant à une séquence de 34 acides aminés qui se structurent en une architecture de type hélice α-coude-hélice α antiparallèle. Un domaine TPR consiste en une succession de 3 à 16 motifs TPR (D'Andrea & Regan, 2003). Parmi les protéines à domaines TPR, une famille se caractérise par la présence de résidus conservés formant une pince à carboxylate ciblant spécifiquement les chaperonnes Hsp70/90 (Prasad *et al.*, 2010). Dans un crible à haut débit, la protéine Tah1p avait été identifiée comme un nouveau cofacteur de la chaperonne hsp90, agissant avec la protéine Pih1p (Eckert *et al.*, 2010; Zhao *et al.*, 2005; Zhao *et al.*, 2008). Malgré la conservation des résidus des chaperonnes Hsp70 et Hsp90 impliqués dans l'interaction avec les domaines TPR, ces domaines sont capables de reconnaitre spécifiquement l'une ou l'autre des chaperonnes Hsp70 et Hsp90 (Millson *et al.*, 2008; Scheufler *et al.*, 2000).

Afin de mieux comprendre les bases structurales de l'interaction entre Tah1p et Hsp90 et de connaitre la structure 3D de Tah1p, nous avons débuté une étude structurales par RMN de Tah1p de *S. cerevisiae* sous forme libre que j'ai terminée en collaboration avec l'équipe de F. Allain (ETH Zurich). Cette collaboration a permis de déterminer la structure tridimensionnelle de la protéine Tah1p libre. J'ai par ailleurs aussi, toujours en collaboration avec F. Allain, établi la structure 3D du complexe formé par Tah1p et un peptide correspondant à l'extrémité C-terminale d'Hsp90 dénommée Hsp82 chez *S. cerevisiae*. Les

résultats obtenus ont révélé la présence dans Tah1p d'un domaine TPR N-terminal atypique. Il est composé de 5 hélices α antiparallèles. Le domaine C-terminal de Tah1p est non structuré que ce soit à l'état libre ou à l'état lié. La qualité des structures obtenu a permis de comprendre comment la protéine Tah1p s'associait à la chaperonne Hsp90 et comment se réalise la discrimination avec son homologue Hsp70. De plus, nous avons également montré que le domaine C-terminal de la protéine Tah1p est responsable de son association avec un autre composant majeur du complexe R2TP, la protéine Pih1p. Ce domaine est nécessaire et suffisant pour l'interaction.

En parallèle à notre travail, l'équipe de W. Houry a aussi étudié la structure 3D du complexe formé par Tah1p et l'extrémité C-terminal d'Hsp90. Mais les données RMN obtenues par cette équipe n'étaient pas d'assez bonne qualité pour leur permettre de déterminer comment le peptide d'Hsp90 interagissait avec Tah1p.

2 Résultats et discussion

Aussi nous avons rassemblé l'ensemble de nos données dans un manuscrit qui a été soumis au journal *Structure*. Des résultats sur l'étude structurale de la protéine Pih1p seront également présentés en complément de l'article dans le paragraphe 2.2.

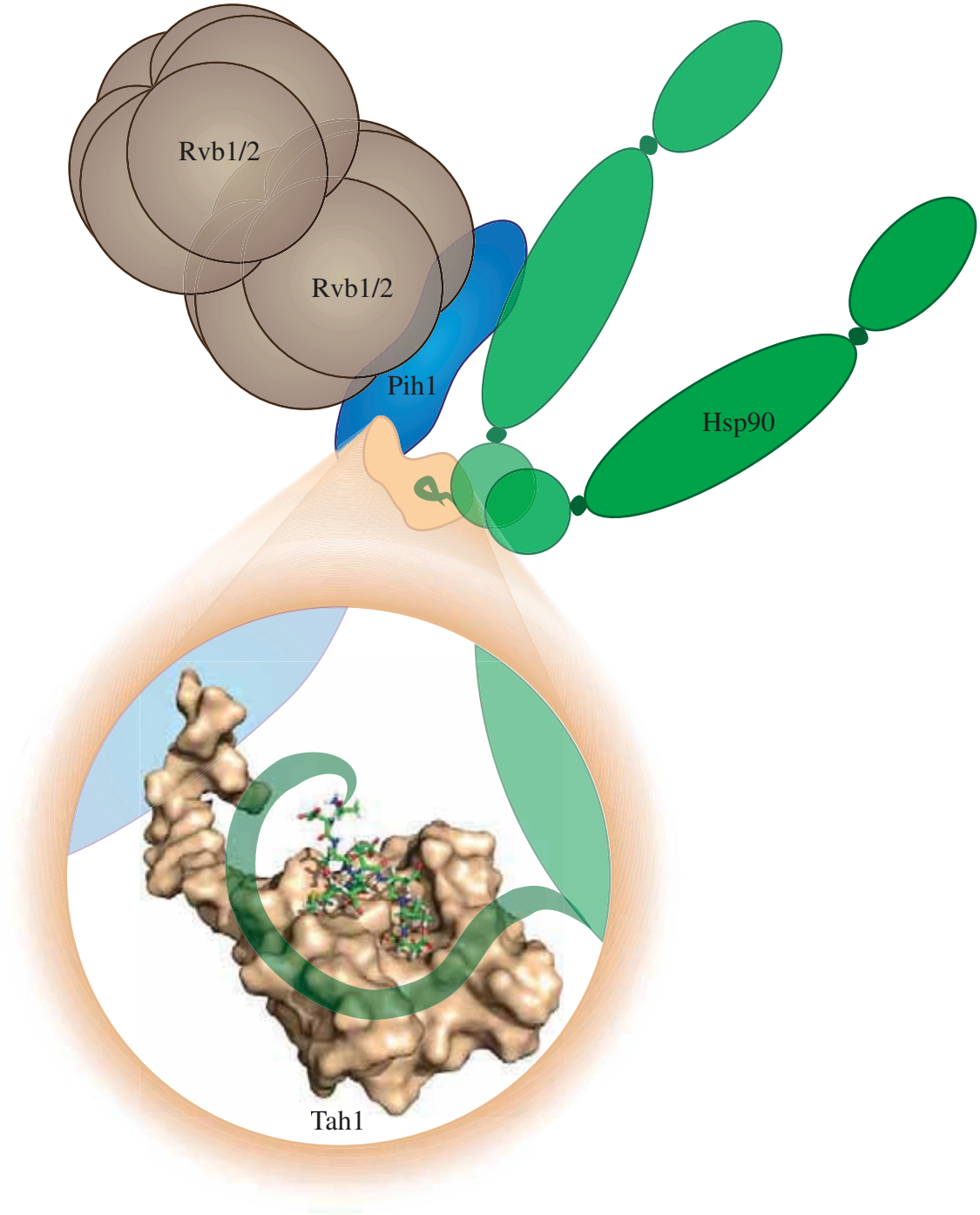
2.1 Manuscrit n°3

NMR high resolution structures of free Tah1 and Tah1 bound to the Hsp90 C-terminal tail explains how Hsp90 recognizes the R2TP complex

Régis Back, Cyril Dominguez, Benjamin Rothé, Claude Bobo, Chrystel Beaufils, Solange Moréra, Philippe Meyer, Bruno Charpentier, Christiane Branlant, Frédéric H.-T. Allain, and Xavier Manival

Soumission à Structure

Graphical Abstract
Click here to download Graphical Abstract: Fig_Tah1_Graphical.pdf



The Hsp90 hook or how to catch the R2TP complex !

Manuscript
Click here to view linked References

NMR high resolution structures of free Tah1 and Tah1 bound to the Hsp90 C-terminal tail explains how Hsp90 recognizes the R2TP complex

Régis Back[1], Cyril Dominguez[2,+], Benjamin Rothé[1], Claude Bobo[1], Chrystel Beaufils[3], Solange Moréra[4], Philippe Meyer[4*], Bruno Charpentier[1], Christiane Branlant[1], Frédéric H.-T. Allain[2§], and Xavier Manival[1§]

(1) AREMS (ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale), Nancy Université, UMR 7214 CNRS-UHP, Bâtiment Biopôle, Faculté de médecine, 9 Avenue de la Forêt de Haye, BP184, 54506 Vandoeuvre-lès-Nancy Cedex, France.

(2) IMB (Institute of Molecular Biology and Biophysics), ETH, Zurich, Suisse.

(3) LCPM, Groupe Synthèse Organique et Biostructures, UMR CNRS-INPL 7568, 1, rue Grandville, BP 20451, F- 54001 Nancy Cedex, France.

(4) LEBS (Laboratoire d'Enzymologie et Biochimie Structurales) UPR CNRS 3082, Bât. 34, 91198 Gif-sur-Yvette Cedex, France.

+ Present address: Department of Biochemistry, University of Leicester, Leicester, United Kingdom.

* Present address: Laboratoire de Biologie Moléculaire et Cellulaire des Eucaryotes - Institut de Biologie Physico-Chimique, CNRS-UPMC, 15 rue Pierre et Marie Curie 75005 Paris, France.

§ To whom correspondence should be addressed

xavier.manival@maem.uhp-nancy.fr

allain@mol.biol.ethz.ch

Short Title: NMR structures of Tah1 and its complex with Hsp90

Keywords: Tah1; Hsp90; Pih1; R2TP complex; Saccharomyces cerevisiae; TPR domain; NMR; Y2H.

2

SUMMARY (150 words)

The ubiquitous Hsp90 chaperone participates to snoRNP and RNA polymerase assembly through interaction with the R2TP complex. Yeast R2TP includes proteins Tah1, Pih1, Rvb1, and Rvb2. Tah1 links R2TP to Hsp90. Its small TPR domain includes two TPR motifs and a capping helix C. We established the high resolution solution structures of free Tah1 and Tah1 bound to the Hsp90 C-terminal peptide. In the free and bound state the TPR domain is folded, while the C-terminal domain is unfolded. In addition to a TPR solvating/stabilizing role, helix C is essential for recognition of the Hsp90 ^{704}EMEEVD^{709} motif. Indeed, it contains two residues involved in the carboxylate clamp and the Tyr82 residue which establishes a π/S-CH3 interaction with M^{705} generating a 3_{10} helix in the peptide backbone. By 2-hybrid assays, we confirmed the functional role of helix C and found that the unstructured Tah1 C-domain is sufficient to bind the Pih1 C-domain.

INTRODUCTION

The highly conserved eukaryotic Hsp90 (90-kDa heat-shock protein) protein is an ubiquitous ATP-dependent molecular chaperone which is essential for cell viability (Borkovich et al., 1989). In addition to its activity as a chaperone protein in conditions of stress, Hsp90 is required for conformational maturation of signal transduction proteins, which links the Hsp90 activity to cell cycle regulation and cancer progression (reviewed in Caplan et al., 2007; Whitesell and Lin, 2012). Hsp90 recognizes specifically a set of protein substrates designated as client proteins such as nuclear hormone receptors (androgen and estrogen receptors), transcription factors (p53, HIF-1), protein kinases (ERBB2, B-RAF, C-RAF, and CDK4), and a group of non-signaling proteins, including hTERT, a major component of telomerase (Theodoraki and Caplan, 2012). Furthermore, Hsp90 has been recently shown to play other crucial functions for proper assembly of several essential eukaryotic ubiquitous cellular machineries such as: i) RNPs containing an RNA-binding protein of the L7Ae family, namely, the C/D box snoRNPs that guide and catalyse 2'-O methylations in pre-ribosomal RNA and are required for pre-rRNA cleavages, the H/ACA snoRNPs which catalyse pseudouridine formation in the pre-rRNA, the spliceosomal U4 snRNP, the telomerase RNP responsible for telomere synthesis, and selenoprotein mRNPs (Boulon et al., 2008; Zhao et al., 2008), ii) the nuclear RNA polymerases (Boulon et al., 2010) and, iii) the phosphatidylinositol 3-kinase-related protein kinases cell-cycle control family, including proteins mTOR (mammalian Target Of Rapamycin) and SMG1 (Suppressor of Morphogenetic effect on Genitalia-1) (Horejsi et al., 2010). How Hsp90 is involved in these new highly important functions still has to be deciphered.

Progresses in the understanding of the previously identified Hsp90 functions will be useful to understand its new functions. Hsp90 reveals the activity of its client proteins by acting at a late stage when they are in a near-native state (reviewed in Pearl and Prodromou,

2006). Hsp90 is also involved in the degradation of trapped misfolded proteins via the ubiquitin-proteasome pathway (Connell et al., 2001). The Hsp90 chaperone activity depends on the ATPase activity carried by its N-terminal and middle domains (Meyer et al., 2003b), and inhibition of this activity by geldanamycin blocks cell proliferation (Whitesell and Cook, 1996). Interestingly, geldanamycin was also shown to block the newly discovered functions of Hsp90 in snoRNP and RNA polymerase assembly (Boulon et al., 2008; Zhao et al., 2008). The Hsp90 C-terminal domain is responsible for Hsp90 dimerization (Minami et al., 1994). In its dimeric form, Hsp90 acts as a molecular clamp whose transient closure allows ATP hydrolysis (Ali et al., 2006).

To achieve its specific protein maturation activity, Hsp90 cooperates with another molecular chaperone, Hsp70, and numerous co-factors called co-chaperones. Through modulation of ATP affinity or ATP hydrolysis, these co-chaperones participate in the regulation of Hsp90 ATPase cycle (reviewed in Prodromou, 2012), and in the specific recognition of protein substrates ((Southworth and Agard, 2011) and reviewed in Li et al., 2012). Many Hsp90 co-chaperones contain a TetratricoPeptide Repeat domain (TPR domains). These proteins bind to the 20 amino acid C-terminal sequence of Hsp90 that contain the conserved MEEVD motif (reviewed in Zeytuni and Zarivach, 2012).

Numerous 3D structures of Hsp90 domains or full-length Hsp90 proteins have been solved at high resolution (reviewed in Prodromou, 2012), i.e. Escherichia coli HtpG (Shiau et al., 2006), yeast Hsp90 and canine grp94 proteins (Dollins et al., 2007). The apo-enzyme is in an open V shape, while the AMP-PNP bound enzyme is in a closed twisted conformation. This closed conformation is stabilized by interaction with co-chaperones, and surprisingly, recent data revealed that in its closed conformation, the Hsp90 dimer would not enclose its client protein but provides a bipartite binding surface whose formation and disruption are coupled to the chaperone ATPase cycle (Prodromou, 2012).

The crystal structures of Hsp90 and Hsp70 C-terminal domains in complex with partner TPRs brought information on determinants ensuring TPR recognition (Scheufler et al., 2000). For instance, Hsp70 and Hsp90 recognize distinct TPR domains in the Hop (HSP-Organizing Protein) co-factor (TPR1 and TPR2A, respectively). TRP domains contain from 3 up to 16 TPR motifs (reviewed in D'Andrea and Regan, 2003). TPRs are defined by a 34 amino acids sequence with only a few conserved residues (reviewed in Zeytuni and Zarivach, 2012). Based on numerous crystal structures (Das et al., 1998; Scheufler et al., 2000; Zhang et al., 2005), TPR domains share a common helix-turn-helix structure, that packs in a parallel fashion resulting in a spiral of repeating anti-parallel α helices. The TPR domains of Hsp70/90 co-chaperones share five exposed residues which form the dicarboxylate clamp interacting with the D residue of the Hsp70 and Hsp90 EEVD motifs (Russell et al., 1999; Scheufler et al., 2000). Discrimination between Hsp90 and Hsp70 by the co-chaperone depends on the identity of the amino acid located upstream of EEVD (Scheufler et al., 2000).

To ensure the newly discovered functions in macro-complex assembly Hsp90 associates with one peculiar co-chaperone, the R2TP complex. This complex contains the two AAA+ ATPases/Helicases Rvb1 and Rvb2 (TIP49 and TIP48 in human, respectively), and proteins Pih1 (Protein Interacting with Hsp90) and Tah1 [TPR containing protein associated with Hsp90] in yeast, or Spaghetti, its counterpart in vertebrates (Boulon, 2008). Pih1 forms a stable complex with Tah1, which down-regulates Hsp90 ATPase activity (Eckert et al., 2010). As a result, the dimeric clamp of Hsp90 is open, allowing the recognition and loading of the client proteins. This R2TP property is shared by members of the early co-chaperone family (Roe et al., 2004; Southworth and Agard, 2011). Tah1 contains two TPR motifs forming a single TPR domain (Zhao et al., 2008). Amino acid sequence comparison shows that the Tah1 TPR domain and the Hop TPR2B domain share similarities suggesting a similar binding mode to Hsp90 (Millson et al., 2008). Determining how Tah1 interacts with Hsp90 is essential to explain how Tah1 tethers Hsp90 to the R2TP complex. The NMR structure of Tah1 in complex

with the MEEVD peptide corresponding to the yeast Hsp90 C-terminus was recently studied (Jimenez et al., 2012). However, in this study, due to the low number of intermolecular NOE restraints obtained for the complex, only the fold of bound Tah1 was completely determined from NMR experiments. The structure of the bound Hsp90 peptide was modelled. Consequently, the proposed structure of the complex is not of sufficient resolution to fully understand the specificity of complex formation. Furthermore, the structure of free Tah1 could not be determined and was reported to be unstable (Jimenez et al., 2012).

Here, we report the high resolution structures of both free Tah1 and Tah1 bound to the ADTEMEEVD C-terminal sequence of Hsp90 using a large panel of NMR approaches. Our data highlight how the Hsp90 terminal peptide is bound to Tah1 and explain the specificity of the interaction. We compare our structure with the structures of several previously reported TPR-Hsp90 interactions and the recently reported Hsp90-Tah1 model. We also studied the Tah1:Hsp90 and Tah1:Pih1 interactions in cellulo. Altogether, our data bring consistent information on how Hsp90 recognizes the R2TP complex.

Experimental Procedures

Biological Material used in this study

The Saccharomyces cerevisiae strains Y190 and Y187 and plasmids pACTII::PIH1, PACTII::HSP82 (WT and mutants) and pACTII::TAH1, pGBKT7::TAH1 (WT, domains and mutants) and pGBKT7::PIH1 (WT and domains) were used for Y2H assays (supplemental experimental procedures). To express Tah1-His6 a cDNA encoding full length Tah1 was cloned into plasmid pKHS (Eckert et al., 2010). Escherichia coli BL21 (DE3) (Novagen®) cells complemented with the pRare2 plasmid were used as the host cells for Tah1-His6 production. The ADTEMEEVD peptide was purchased from GenCust® (Luxembourg).

Production and purification of the Tah1 recombinant protein

Recombinant Tah1-His_6 protein was purified as described previously (Eckert et al., 2010). The Ni-Sepharose bound protein was eluted with an imidazole gradient, dialyzed in 20 mM Tris HCl (pH 7.5), 150 mM NaCl and bound again on Ni-Sepharose High Performance beads. Free Tah1 was released by cleavage with PreScission protease (5-10 U per mg of protein) overnight at 4°C. Residual impurities and the protease were removed by size exclusion chromatography using a Superdex 75 prep grade column (GE Healthcare®) equilibrated with the NMR buffer (10 mM NaH_2PO4/Na_2HPO4 (pH 7.2), 150 mM NaCl). The protein was stored in this buffer at -20°C at 1 mM.

NMR measurements

NMR spectra of free Tah1 were recorded at 288 K on a BRUKER DRX 600 US equipped with a triple resonance (^1H, ^{15}N, ^{13}C) TCI Z 5 mm cryoprobe using a ^{15}N-^{13}C labeled Tah1 sample (~1 mM). Data were processed by Topspin (Bruker®) and analyzed with the Sparky 3.113 software (Goddard and Kneller, 2002). Sequence-specific backbone, side chain and aromatic proton assignments were achieved using 2D and 3D classical experiments (supplemental experimental procedures).

NMR measurements for the assignment of the Tah1: peptide complex were performed in the same conditions as above using Bruker AVIII-500 MHz, AVIII-700 MHz, and Avance-900 MHz spectrometers. The samples were prepared at a ~ 1 mM concentration using ^{15}N-^{13}C labeled Tah1 and unlabeled yeast Hsp90 peptide at a 1:1 molar ratio. We assigned the bound Tah1 backbone, side chains and aromatic protons using 2D and 3D experiments described in supplemental experimental procedures. We assigned protons of the bound Hsp90 peptide in the Tah1:peptide complex by using 2D F1-filtered, F2-filtered TOCSY and 2D F1-filtered, F2-filtered NOESY experiments in H_2O or 100% D_2O.

Intermolecular NOEs were obtained from 3D ^{15}N F1-filtered, F2-edited NOESY-HSQC, 3D ^{13}C F1-filtered, F3-edited NOESY-HSQC experiments in H_2O or 100% D_2O, and 2D ^{13}C F1-filtered, F2-edited NOESY performed on the ^{15}N-^{13}C Tah1:peptide complex.

Structure calculation and refinement

The automated peak picking, NOE assignment and structure calculation of free and bound Tah1 were performed using the Atnos/Candid software (Herrmann et al., 2002a, b). We used: i) 3D ^{15}N and ^{13}C NOESY-HSQC, and 2D homonuclear NOE spectra recorded in D_2O for structure analysis of free Tah1, ii) 3D ^{15}N and ^{13}C NOESY-HSQC experiments for structure determination of the bound protein, and iii) 2D ^{13}C F1-filtered F2-filtered NOESY in H_2O or 100% D_2O for structure analysis of the bound peptide. Seven iterations were performed and 250 independent structures were calculated at each iteration step.

We used the CYANA software to calculate the structures of the Tah1:Hsp90 peptide complex, calculation was based on a list of automatically assigned intra-molecular NOE distance constraints that we established for the bound Tah1 and bound peptide using the Atnos/Candid software and manually assigned intermolecular distance constraints for the complex (Guntert, 2004).

The 50 structures with the lowest energy were refined using the AMBER 7.0 software with simulated annealing protocol (Case et al., 2005). We analyzed the 20 best final structures based on energy, NOE violations, and structural quality using the PROCHECK software (Laskowski et al., 1996).

Yeast two-hybrid assays (Y2H).

Recombinant plasmids obtained after cloning the various protein coding sequences into plasmids pACTII (LEU2) and pGBKT7 (TRP1), were used to transform haploid cells (strains Y190 and Y187, respectively). After mating, the diploid cells were selected on media suitable for double selection (Leu^-, Trp^-), and interaction tests were performed after plating the cells on media suitable for triple selection (Leu^-, Trp^-, His^-). Various concentrations of 3-Amino-1,2,4-

triazol (3AT) were used to evaluate the strength of the interactions. Growth was assessed after three days of incubation at 30°C.

RESULTS

Free Tah1 NMR structure reveals the shortest TPR domain and a flexible C-terminus

To study the structure of free Tah1 in solution using NMR, we produced a ^{15}N-^{13}C labeled Tah1 protein which ^{1}H-^{15}N HSQC spectrum is shown in Figure S1A. Backbone and side-chain resonance were assigned using conventional methods (Sattler et al., 1999). Structure calculation based on NOESY spectra were performed using CYANA (Guntert, 2004) followed by a refinement with AMBER (Case et al., 2005). Figure 1 shows the superposition of the 20 conformers with the lowest energy that best satisfy the experimental restraints (Figures 1A-1B, Table 1). This family of 20 conformers is well defined with an rmsd for the backbone atoms of amino acids 4-92 of 0.44 ± 0.08 Å (Table 1). The data reveals the presence of a structured N-terminal TPR domain (1-92, N-domain) flanked by a flexible short C-terminal domain (93-111, C-domain). As expected from previous predictions (Millson et al., 2008; Zhao et al., 2008), the N-terminal residues 1-67 are folded into two tandem TPR motifs (TPR1 and TPR2) each composed of a pair of antiparallel α-helices: residues 4-17 and 20-34 for TPR1, and 38-51 and 54-67 for TPR2 (Figure 1). While Tah1 TPR1 contains the highly conserved TPR residues Gly/Ala11, Ala23, Tyr27, Pro35, and the less conserved residues Leu14 and Leu30 (reviewed in Zeytuni and Zarivach, 2012), TPR2 is more degenerated (Figure 1C). TPR2 contains only one of the conserved residues of TPR motifs, Asn43, which is part of the dicarboxylate clamp. The inter-helical angles in the TPR1 and TPR2 motifs range from 155 to 168° according to MOLMOL software calculations (Koradi et al., 1996). They are highly similar to those calculated for other Hsp co-chaperones. As recently found for Tah1 bound to MEEVD (Jimenez et al., 2012), helix 2B in TPR2, is tightly associated through multiple hydrophobic contacts with a fifth helix (residues 74 to 90), designated as capping helix or helix C. These hydrophobic contacts are established by Leu84 and Val91 in helix C and Ile58 of TPR2 and by Leu80 in helix C and Leu65 of TPR2 (Figure S1B). Helix C does not display the sequence

hallmarks of the first helix of a third TPR motif. However it bears the conserved Lys79 and Arg83 residues which might participate to carboxylate clamp formation (Figure 1C). The angle defined by helix C and helix 2B is small (145°) compared to angles in TPR motifs. Therefore, in free Tah1, helix C can be considered as the capping/solubilizing helix of the shortest Tah1 TPR domain (only 2 TPR motifs). As usual, the two helices of each TPR motif interact through their hydrophobic side-chains, while their opposite hydrophilic sides are exposed to the solvent (reviewed in Zeytuni and Zarivach, 2012). The additional amphipathic helix C fully participates in the formation of a compact helix-turn-helix arrangement characteristic of TPR domains (Figure 1C). Lastly, a 3_{10} helix encompassing residues 71 to 73 was detected in some but not all the 20 structures of lowest energy (Figure 1A).

The 19 amino acid C-domain (positions 93 to 111) located immediately downstream from helix C displays the characteristic features of an unstructured region, with random coil chemical shifts and reduced ^{1}H-^{15}N heteronuclear NOE values indicative of elevated motion relative to the TPR domain (Figure S2A). Another difference between these two domains, concerns their surface charges which are predominantly negative for the C-domain and positive for the N-domain (Figure S2B). These differences suggested distinct protein binding properties for the two domains.

The two domains of Tah1 interact with different protein partners

In the R2TP-Hsp90 complex, Tah1 is bound to both Pih1 and Hsp90 (Zhao et al., 2005). Although, the Tah1 N-terminal TPR domain was expected to bind Hsp90, a recombinant protein containing the TPR1 and TPR2 motifs but not helix C failed to interact with Hsp90 (Zhao et al., 2008). Having delineated a short TPR domain (two TPR motifs + helix C) and a flexible C-terminal domain in free Tah1, we used yeast two-hybrid assays (Y2H) (Figure 2A) to test their capacities to bind to Hsp90 and Pih1. To this end, the nucleotide sequences encoding, i) full-length Tah1, ii) the C-domain, and iii) the short TPR domain were individually cloned into the Y2H plasmid pGBKT7 and the nucleotide sequences

encoding Pih1 and Hsp90 were inserted into the Y2H plasmid pACTII. Various concentrations of 3-Amino-1,2,4-triazol (3AT), were used to evaluate the strength of the interactions in yeast cells (Figure 2A). In these assays, the full length Tah1 and the short TPR domain were found to establish tight (stable at 10 mM 3AT) and almost identical interactions with Hsp90, strongly indicating that the short TPR domain represents the Hsp90 recognition domain of Tah1. This domain did not interact with Pih1 in the assays (Figure 2A), while the isolated C-terminal domain interacted tightly with this protein (Figure 2A). Therefore, the very short unstructured C-domain of Tah1 is sufficient to bind Pih1 in cellulo and the interaction is comparable to the full length Tah1.

The 253-344 C-terminal fragment of Pih1 interacts with the flexible C-domain of Tah1

Recent in vitro mass spectrometry analysis showed that the Pih1 fragment extending from position 199 to 344 binds to Tah1 in vitro (Eckert et al., 2010). To further delineate the Tah1 binding site in Pih1, we again used Y2H assays. Alignment of the yeast and human Pih1 amino acid sequences, revealed an inserted 42 amino acids sequence in yeast Pih1 that separates an N-terminal 211 amino acids sequence showing 38% similarity in the two proteins and a 92 C-terminal amino acids sequence displaying 24% of similarity in the two proteins (Figure 2B). The 1-252 Pih1 N-domain including the yeast specific sequence did not interact with Tah1 in Y2H assays, while an efficient interaction was detected with the Pih1 253-344 C-domain (Figure 2B). Therefore, our data indicate that the 92 C-terminal amino acids of Pih1 are sufficient for in vivo interaction with Tah1 (interaction mediated by the flexible C-domain of Tah1 as shown in Figure 2A).

The NMR structure of Tah1 bound to the Hsp90 C-terminal sequence reveals how the shortest TPR domain mediates the interaction

The 9 C-terminal amino acids of yeast Hsp90 (ADTEMEEVD) being sufficient to specifically interact with Tah1 (Millson et al., 2008), we titrated ^{15}N labeled Tah1 with the

peptide by increasing its concentration up to a 2:1 peptide:protein stoechiometry (Figure S1A). The ^{1}H-^{15}N HSQC spectra indicate that complex formation is achieved at a 1:1 protein:peptide stoichiometric ratio. Tah1 residues showing the largest chemical shift changes upon complex formation are residues of the carboxylate clamp (Asn12, Asn43, and Arg83) and Tyr27 and Lys50 (Figure S1C). No residues in the C-terminal domain showed a significant change in chemical shift consistent with the fact that this unstructured region does not participate in the interaction. We then determined the structure of the complex using a ^{15}N-^{13}C labeled Tah1 protein bound to an unlabeled Hsp90 peptide. Resonance assignments for the bound peptide were obtained by double filtered TOCSY and NOESY spectra. The structure was determined using a total of 2,771 NOE-derived distance restraints including 55 intermolecular distance restraints derived from filtered-edited 2D and 3D NOESY (Figure S3). In the structure of the complex, Tah1 adopts the same fold and organization as in its free form (Figure S1B). The superposition of Tah1 TPR domains from the free and bound structures reveals a backbone RMS deviation of 0.94Å (for residues 4-70) indicating no conformational change upon peptide binding. Only helix C undergoes a slight shift towards helix 2A which is due to establishment of a new hydrophobic interaction between residues Leu86 and Ile49 in addition to the Ile76-Val30 interaction (Figure S1B). The C-terminal amino acids remain unfolded in the bound state. This was confirmed by ^{1}H-^{15}N heteronuclear NOE measurements of Tah1 in the bound state (Figure S2A).

In this complex, the three N-terminal residues of Hsp90 ($A^{701}DT^{703}$) adopt a rather flexible conformation, while the six most C-terminal residues ($E^{704}MEEVD^{709}$) have a rigid conformation (Figure 3A). The $E^{707}VD^{709}$ segment adopts an extended conformation similarly to that previously observed for Hsp C-terminal peptides bound to the HOP, CHIP, or Tom71 proteins (Li et al., 2009; Scheufler et al., 2000; Zhang et al., 2005). As found in the CHIP:Hsp90 complex (Zhang et al., 2005), the main chain of the $M^{705}E^{706}E^{707}$ triplet adopts a turn stabilized by intermolecular H-bonds formed between the two main chain carbonyl groups

of M^{705} and E^{707} and the side chain of Tah1 Arg83 which brings the side chain of M^{705} close to the Tah1 helix C (Figure 4A). The backbone atoms of $E^{704}M^{705}E^{706}$ adopt a 3_{10} helix structure enabling the N-terminal extremity of the peptide to exit the protein channel (Figures 3C-D). This peculiar structure of the E^{704}MEEV^{707} peptide bound to Tah1, which is stabilized by formation of a H-bond between the main chain of E^{704} and V^{707}, is required for its accommodation into the groove of the Tah1 TPR domain (Figures 3B-C and 4).

The highly asymmetric charge distribution between the peptide and Tah1 (Figures S4A-B) results in the establishment of several salt bridges between the positive channel of Tah1 and the negative peptide. These interactions are reinforced by formation of several intermolecular H-bonds and hydrophobic interactions (Figures 4A and B) explaining the in vivo stability of the Tah1:Hsp90 and 1-92 Tah1 TPR:Hsp90 interactions (Figure 2A). Each of the residues in the EMEEVD sequence participates in the Tah1 TPR recognition. The C-terminal amino acid D^{709} plays a key role in formation of the complex. Its side chain forms H-bonds with the Asn12 and Asn43 side chains and a salt bridge with Lys8 which are all located in the TPR motifs of Tah1 (Figure 4A). The terminal main chain carboxylate group of D^{709} forms two salt bridges with Lys79 and Arg83 located in helix C. Finally, the main chain amide of D^{709} forms an H-bond with Asn43 carbonyl side chain (Figure 4A). A similar network of interactions has been observed in complexes formed by the TPR domains of Hop, PP5 and Tom 71 proteins and C-terminal peptides of chaperone proteins (Li et al., 2009; Scheufler et al., 2000; Zhang et al., 2005). The side chain of residue V^{708} is located in a hydrophobic pocket formed by the side chains of Lys8 and Lys16 and the aromatic ring of Phe15. Its main chain carbonyl group is hydrogen-bonded with the Lys79 side chain (Figure 4). The carboxylate side chains of residues E^{706} and E^{707} establish salt bridges with the side chain amino-groups of Lys50 and Lys16, respectively. The divalent sulfur group of M^{705} specifically interacts with the π electron of the aromatic carbon cycle (sp^2) of Tyr82, and its methyl group is positioned between the

methylene of Tyr82 and the aliphatic side chain of Lys79 (Figure 4). Finally, the carboxylate side chain and main chain carbonyl of E^{704} both interact with the amino side chain of Lys79. No interaction was observed between A^{701}, D^{702}, and T^{703} of Hsp90 and Tah1. Our structure shows that the two canonical TPR motifs of Tah1 and its additional helix C are involved in the interaction with the C-terminal sequence EMEEVD of yeast Hsp90. Altogether, helices 1A and 2A of the canonical TPR part contribute to the formation of one hydrophobic interaction, four salt bridges and three H-bonds, while helix C contributes to the formation of nine interactions with EMEEVD: one hydrophobic interaction, three salt bridges and five H-bonds, underlying the crucial role of this additional helix C for the interaction with Hsp90.

Residues in Tah1 helix C are essential for the interaction with Hsp90 in vivo

To confirm the functional importance of helix C in the interaction of Tah1 with Hsp90, we substituted individually the Tah1 Lys79, Tyr82, Arg83 and Hsp90 E^{706}, E^{707}, V^{708}, D^{709} by alanine and tested the effects of these mutations on the Hsp90:Tah1 interaction by Y2H assays (Figure 5). As a control, Arg77, Glu85 in the Tah1 helix C, and D^{702}, T^{703} in Hsp90 which are not involved in the protein:peptide interaction were also substituted. We controlled the expression of mutated Tah1 and Hsp90 proteins by testing their interactions with Pih1, using Y2H assays. In perfect agreement with our structure, Lys79Ala, Tyr82Ala, and Arg83Ala substitutions in Tah1 helix C abolished the interaction with Hsp90 but did not affect the Tah1:Pih1 interaction (Figure 5A). Also, in agreement with our structure, substitutions of Arg77 or Glu85 had no effect on any of the interactions. Concerning Hsp90, the D^{709}A substitution completely abolished the interaction with Tah1, confirming the crucial role of the C-terminal D^{709}. The interaction was weakened by the E^{706}A and V^{708}A substitutions, although the effect was less pronounced than for the other mutants. The E^{707}A mutation had a more limited negative effect, but E^{707} implication was evidenced by a significant decrease of the 3-AT resistance. As expected, mutations of amino acids D^{702} and T^{703} had no marked effect.

These results confirmed the functional importance of Tah1 helix C for in vivo interaction of Tah1 with Hsp90.

DISCUSSION

Here, we report the first solution structures of the Tah1 protein in a free form and in complex with the C-terminal Hsp90 peptide (ADTEMEEVD) at high resolution. Recently, a solution structure of the same protein bound to the MEEVD motif of Hsp90 was described (Jimenez et al., 2012). However, due to a limited number of intermolecular NOE information, the authors had to combine various approaches, including NMR titration assays, site-directed mutagenesis and 3D structure modeling, to try to get a 3D structure model of the MEEVD:Tah1 complex. Although we share a number of common conclusions with these authors, our data disagree with their definition of the mode of interaction of the MEEVD peptide. More precisely, both studies identified a short TPR domain formed by two TPR motifs and a additional functional capping helix C, as well as a flexible C-terminal domain in Tah1. However, our 3D structure of bound Tah1 differs subtantially from their structure (rmsd of 2.5Å) (Figure 6A). As shown by superimposition of the two structures (Figure 6A), one strong difference concerns the orientation of the Hsp90 peptide in the TPR groove. Importantly, our positioning of the peptide in the Tah1 groove was determined with 55 intermolecular NOEs (Figure S3) at a stoichiometric ratio, while Jimenez et al., only observed 5 intermolecular NOEs at a very high 1 to 10 protein to peptide molar ratio. Therefore, we are confident that the orientation of the peptide in our complex is accurate.

Furthermore, in contrast to helices of the Tah1 TPR motifs in the structure of Jimenez et al., the equivalent four helices in our structure of bound Tah1 are superimposable with the equivalent helices in other bound TPR domains, such as the CHIP TPR and Hop TPR2A domains (rmsd average of 1Å) (Figure 6B-C). The Tah1 helix C superimposed with a slight shift towards the peptide on helix 3A of the third TPR motifs of these two larger TPR domains.

Furthermore, residues of the Tah1 carboxylate clamp, namely, Lys8, Asn12, Asn43 from the TPR motifs and Lys79 and Arg83 from helix C, are superimposable with amino acids of the carboxylate clamps of the bound CHIP and Hop TPR2A domains (Figures 1C, 6B-C and 7). It should be pointed out that, in spite of main differences between their structure and our structure of the complex, Jimenez et al., experimentally observed a functional importance of the Tah1 Lys8, Asn12, Asn43, Lys79 and Arg83. Indeed, they found that individual substitution of these five Tah1 residues into alanine either reduced or abolished in vitro binding of MEEVD (Jimenez et al., 2012). Data of our Y2H assays further support that Lys79 and Arg83 of helix C play a crucial role for interaction of Tah1 with the full length Hsp90 in vivo. In our 3D structure of bound Tah1, the six C-terminal amino acids (E^{704}MEEVD^{709}) of the Hsp90 tail are involved in the interaction with Tah1. Whereas, the EEVD-Tah1 interaction is similar to that found for the human PP5 (Russell et al., 1999) and Hop proteins (Brinker et al., 2002), the interaction established by M^{705} bound to Tah1 is different from that in bound PP5 and Hop proteins but similar to that in bound CHIP protein (Figures 6B-C).

Our structure indicates that three residues of the Hsp90 EEVD motif are key players in the interaction with Tah1: namely, E706, V708 and D709 which contacts 3, 4 and 5 amino acids of Tah1, respectively. Accordingly, in our Y2H experiments, the D^{709}A mutation had a pronounced negative effect on the interaction, the E^{706}A and V^{708}A substitutions an intermediate one and the E^{707}A mutation a weak effect (Figure 5B). This lower effect of the E^{707}A mutation is consistent with formation of only one interaction between E^{707} and Tah1 (a salt bridge with Lys50). It should be noted that mutation of residue R74 in PP5, the counterpart of Lys50 in Tah1, which also forms a unique salt bridge with Hsp90, reduced but did not abrogate the interaction with Hsp90 (Russell et al., 1999). In contrast, coupling mutation of residue E^{707} with mutations of E^{706} and D^{709} in the Hsp90 tail abolished its interaction with both PP5 and Hop proteins (Ramsey et al., 2000). The lower importance of E^{707} in the

Hsp90:TPR interaction probably explains the low degree of conservation of this residue in eukaryotic Hsp90 proteins (Scheufler et al, 2000).

Our data strongly implicate the interactions established between the Tah1 helix C and the Hsp90 amino acids located upstream from EEVD in the specificity of the Tah1:Hsp90 recognition. It is already well admitted that the Hsp90 M^{705} residue, which is replaced by an isoleucine in Hsp70, is one key element in the specific recognition of TPRs by Hsp90 (Brinker et al., 2002; Scheufler et al., 2000). Furthermore, interactions between TPRs and Hsp90 amino acids located upstream of EEVD were found to differ from one TPR domain to the other (Figure 7) (Scheufler et al., 2000; Zhang et al., 2005). In the Hsp90:Tah1complex this specific interaction of residues upstream of EEVD is characterized by a peculiar conformation of the peptide backbone at the level of M^{705}. The main chain of $E^{704}MEE^{707}$ residues forms a 3_{10} helix followed by a turn, so that M^{705} is in close contact with the Tah1 helix C. In the MEEVD:Hop-TPR2A complex, the peptide chain orientation is different, so that M^{728} (equivalent to yeast Hsp90 M^{705}) is in close contact to helix 1A (Figure 7B). The peculiar orientation of the main chain in the Tah1 complex notably results from formation of a S-CH3/π interaction (Meyer et al., 2003a) between M^{705} and the Tah1 Tyr82. Similar S-CH3/π interactions are formed by M^{505} (equivalent to yeast Hsp90 M^{705}) and Phe99 in the bound CHIP TPR but another kind of interaction takes place in bound Hop TPR2A, i.e. a hydrophobic interaction between the methyl of M^{728} and Tyr236 (Figure 7A-B). The different locations of Tah1 Tyr82 and Hop TPR2A Tyr236 within their respective TPR domains induce different orientations of the Hsp90 peptide backbone within the TPR groove. Requirement of the S-CH3/π interaction in the Tah1:Hsp90 complex explains why a $M^{705}V$ mutation in yeast Hsp90 decreased the in vitro binding of Tah1 by a factor of 20 (Millson et al., 2008), and the strong importance of the Tah1 Tyr82 for Tah1:Hsp90 interaction in cellulo is confirmed by the absence of Y2H interaction between Hsp90 and Tah1 carrying a Tyr82A mutation (Figure 5A). Therefore, interactions

formed by M^{705} and specific amino acids of TPR domains are key determinants for both recognition of TPRs by Hsp90 and orientation of the Hsp90 tail within the TPR groove.

As a consequence of the E^{704}MEEVD^{709} backbone orientation, the main and side chains of the Tah1 Lys79 interact with E^{704}. The absence of detection of such additional interaction with E^{704} in previous structures might simply be explained by the very limited number of structural analyses performed with peptides longer than (M/I)EEVD (Zhang et al., 2005). However, this choice of short peptides was justified by the observation of a low influence of residues located upstream of the Hsp90 M^{705} residue on binding to TPRs (Scheufler et al., 2000). In agreement with our structural data, previous results showed that the binding affinity of Tah1 for the Hsp90 MEEVD peptide increases by a factor 30 upon addition of the $T^{703}E^{704}$ amino acids (Millson et al., 2008). The implication of E^{704} in specific interactions with some of the TPR domains may explain its rather high degree of conservation (more than 70%) in yeast species. Lys79 and Tyr82, which interact with E704 and M705, respectively, both belong to helix C. Altogether, our data point out that within the very short Tah1 TPR domain (2 TPR motifs and a unusual functional capping helix) helix C plays a crucial role in the interaction with Hsp90.

Whereas the Tah1 TPR domain adopts a highly folded structure both in free and bound Tah1, the C-terminal domain of Tah1 remains unfolded in the presence or the absence of the Hsp90 peptide. Based on in vitro binding assays, the C-terminal region of Tah1 was previously proposed to interact with Pih1 (Eckert et al., 2010; Zhao et al., 2005). However, the segment required was not precisely delineated. We found that the unfolded C-terminal segment (93-111) is sufficient to get a positive interaction with Pih1 in Y2H-assays. As these assays were performed in a homologous genetic context, we cannot exclude the possibility of a stabilization of the interaction by an endogenous protein. However, these data show that, in vivo, the unfolded domain of Tah1 is sufficient to interact with Pih1, either alone or through the help of a cellular partner. The specific biological activities carried by co-chaperones are associated to

the presence of a domain additional to the TPR domain: e.g. a phosphatase domain in PP5, a peptidylprolyl isomerase domain in FKBP52, an ubiquitin ligase domain in CHIP. Tah1 lacks such specific activity. The short Tah1 sequence which is not included in the TPR domain (position 93 to 111) probably only serves as a Pih1 binding site. The capacity to modulate negatively the Hsp90 ATPase activity is carried out by Pih1 (Eckert et al., 2010), Tah1 alone only has a very weak stimulation activity on Hsp90 (Millson et al., 2008). Therefore, the Pih1:Tah1 heterodimer can be considered as the functional entity able to recognize specifically Hsp90 and to negatively modulate its activity. Tah1 stabilizes Pih1 and mediates its association with Hsp90. However, the role of these two distinct polypeptide chains within this functional co-chaperone entity remains to be elucidated.

ACCESSION NUMBERS

The coordinates for theses free and bound Tah1 structures were deposited in April, 2012 with the PDB codes 2LSU and 2LSV respectively. The resonance assignments have been deposited in the BMRB with accession codes 18445 and 18447 for free and bound Tah1 respectively.

ACKNOWLEDGMENTS

This work was supported by the French “Agence Nationale pour la Recherche” (ANR - 06-BLAN-0208-02 and ANR 11 BSV8 015 03), the French “Centre National de la Recherche Scientifique”, Lorraine University, the “Plan Etat Région Lorraine” (Bio-engineering program) and the European EURASNET network (LSHG-CT-2005-518238). RB and BR were fellows from the French « Ministère de l’Enseignement Supérieur et de la Recherche ». Stays of RB in the FA Lab were supported by the exchange program of EURASNET. Research in the FA lab was supported by the NCCR-SNF structural biology and the ETH Zurich.

REFERENCES

Ali, M.M., Roe, S.M., Vaughan, C.K., Meyer, P., Panaretou, B., Piper, P.W., Prodromou, C., and Pearl, L.H. (2006). Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. Nature 440, 1013-1017.

Borkovich, K.A., Farrelly, F.W., Finkelstein, D.B., Taulien, J., and Lindquist, S. (1989). hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. Mol. Cell. Biol. 9, 3919-3930.

Boulon, S., Marmier-Gourrier, N., Pradet-Balade, B., Wurth, L., Verheggen, C., Jady, B.E., Rothe, B., Pescia, C., Robert, M.C., Kiss, T., et al. (2008). The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. J. Cell. Biol. 180, 579-595.

Boulon, S., Pradet-Balade, B., Verheggen, C., Molle, D., Boireau, S., Georgieva, M., Azzag, K., Robert, M.C., Ahmad, Y., Neel, H., et al. (2010). HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. Mol. Cell 39, 912-924.

Brinker, A., Scheufler, C., Von Der Mulbe, F., Fleckenstein, B., Herrmann, C., Jung, G., Moarefi, I., and Hartl, F.U. (2002). Ligand discrimination by TPR domains. Relevance and selectivity of EEVD-recognition in Hsp70 x Hop x Hsp90 complexes. J. Biol. Chem. 277, 19265-19275.

Caplan, A.J., Mandal, A.K., and Theodoraki, M.A. (2007). Molecular chaperones and protein kinase quality control. Trends Cell Biol 17, 87-92.

Case, D.A., Cheatham, T.E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005). The Amber biomolecular simulation programs. J Comput Chem 26, 1668-1688.

Connell, P., Ballinger, C.A., Jiang, J., Wu, Y., Thompson, L.J., Hohfeld, J., and Patterson, C. (2001). The co-chaperone CHIP regulates protein triage decisions mediated by heat-shock proteins. Nat Cell Biol 3, 93-96.

D'Andrea, L.D., and Regan, L. (2003). TPR proteins: the versatile helix. Trends Biochem Sci 28, 655-662.

Das, A.K., Cohen, P.W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. EMBO J 17, 1192-1199.

DeLano, W.L. (2010). The PyMOL Molecular Graphics System, Version 1.3r1. In The PyMOL Molecular Graphics System, Version 1.3r1 (New York, Schrödinger, LLC).

Dollins, D.E., Warren, J.J., Immormino, R.M., and Gewirth, D.T. (2007). Structures of GRP94-nucleotide complexes reveal mechanistic differences between the hsp90 chaperones. Mol Cell 28, 41-56.

Eckert, K., Saliou, J.M., Monlezun, L., Vigouroux, A., Atmane, N., Caillat, C., Quevillon-Cheruel, S., Madiona, K., Nicaise, M., Lazereg, S., et al. (2010). The Pih1-Tah1 cochaperone complex inhibits Hsp90 molecular chaperone ATPase activity. J. Biol. Chem. 285, 31304-31312.

Goddard, T.D., and Kneller, D.G. (2002). SPARKY 3. S.F. University of California, ed.

Guntert, P. (2004). Automated NMR structure calculation with CYANA. Methods Mol Biol 278, 353-378.

Herrmann, T., Guntert, P., and Wuthrich, K. (2002a). Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR 24, 171-189.

Herrmann, T., Guntert, P., and Wuthrich, K. (2002b). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. 319, 209-227.

Horejsi, Z., Takai, H., Adelman, C.A., Collis, S.J., Flynn, H., Maslen, S., Skehel, J.M., de Lange, T., and Boulton, S.J. (2010). CK2 phospho-dependent binding of R2TP complex to TEL2 is essential for mTOR and SMG1 stability. Mol. Cell 39, 839-850.
Jimenez, B., Ugwu, F., Zhao, R., Orti, L., Makhnevych, T., Pineda-Lucena, A., and Houry, W.A. (2012). Structure of minimal tetratricopeptide repeat domain protein Tah1 reveals mechanism of its interaction with Pih1 and Hsp90. J Biol Chem 287, 5698-5709.
Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14, 51-55, 29-32.
Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8, 477-486.
Li, J., Qian, X., Hu, J., and Sha, B. (2009). Molecular chaperone Hsp70/Hsp90 prepares the mitochondrial outer membrane translocon receptor Tom71 for preprotein loading. J. Biol. Chem. 284, 23852-23859.
Li, J., Soroka, J., and Buchner, J. (2012). The Hsp90 chaperone machinery: Conformational dynamics and regulation by co-chaperones. Biochim Biophys Acta 1823, 624-635.
Meyer, E.A., Castellano, R.K., and Diederich, F. (2003a). Interactions with aromatic rings in chemical and biological recognition. Angew Chem Int Ed Engl 42, 1210-1250.
Meyer, P., Prodromou, C., Hu, B., Vaughan, C., Roe, S.M., Panaretou, B., Piper, P.W., and Pearl, L.H. (2003b). Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. Mol Cell 11, 647-658.
Millson, S.H., Vaughan, C.K., Zhai, C., Ali, M.M., Panaretou, B., Piper, P.W., Pearl, L.H., and Prodromou, C. (2008). Chaperone ligand-discrimination by the TPR-domain protein Tah1. Biochem. J. 413, 261-268.
Minami, Y., Kimura, Y., Kawasaki, H., Suzuki, K., and Yahara, I. (1994). The carboxy-terminal region of mammalian HSP90 is required for its dimerization and function in vivo. Mol. Cell. Biol. 14, 1459-1464.
Pearl, L.H., and Prodromou, C. (2006). Structure and mechanism of the Hsp90 molecular chaperone machinery. Annu Rev Biochem 75, 271-294.
Prodromou, C. (2012). The 'active life' of Hsp90 complexes. Biochim Biophys Acta 1823, 614-623.
Ramsey, A.J., Russell, L.C., Whitt, S.R., and Chinkers, M. (2000). Overlapping sites of tetratricopeptide repeat protein binding and chaperone activity in heat shock protein 90. J. Biol. Chem. 275, 17857-17862.
Roe, S.M., Ali, M.M., Meyer, P., Vaughan, C.K., Panaretou, B., Piper, P.W., Prodromou, C., and Pearl, L.H. (2004). The Mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50(cdc37). Cell 116, 87-98.
Russell, L.C., Whitt, S.R., Chen, M.S., and Chinkers, M. (1999). Identification of conserved residues required for the binding of a tetratricopeptide repeat domain to heat shock protein 90. J. Biol. Chem. 274, 20060-20063.
Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the
structure determination of proteins in solution employing pulsed
field gradients. Progress in Nuclear Magnetic Resonance Spectroscopy 34, 93–158.
Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F.U., and Moarefi, I. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. Cell 101, 199-210.
Shiau, A.K., Harris, S.F., Southworth, D.R., and Agard, D.A. (2006). Structural Analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. Cell 127, 329-340.

Southworth, D.R., and Agard, D.A. (2011). Client-loading conformation of the Hsp90 molecular chaperone revealed in the cryo-EM structure of the human Hsp90:Hop complex. Mol Cell 42, 771-781.
Steinkellner, G., Rader, R., Thallinger, G.G., Kratky, C., and Gruber, K. (2009). VASCo: computation and visualization of annotated protein surface contacts. Bmc Bioinformatics 10.
Theodoraki, M.A., and Caplan, A.J. (2012). Quality control and fate determination of Hsp90 client proteins. Biochim Biophys Acta 1823, 683-688.
Whitesell, L., and Cook, P. (1996). Stable and specific binding of heat shock protein 90 by geldanamycin disrupts glucocorticoid receptor function in intact cells. Mol Endocrinol 10, 705-712.
Whitesell, L., and Lin, N.U. (2012). HSP90 as a platform for the assembly of more effective cancer chemotherapy. Biochim Biophys Acta 1823, 756-766.
Zeytuni, N., and Zarivach, R. (2012). Structural and functional discussion of the Tetra-Trico-Peptide Repeat, a protein interaction module. Structure 20, 397-405.
Zhang, M., Windheim, M., Roe, S.M., Peggie, M., Cohen, P., Prodromou, C., and Pearl, L.H. (2005). Chaperoned ubiquitylation--crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex. Mol. Cell 20, 525-538.
Zhao, R., Davey, M., Hsu, Y.C., Kaplanek, P., Tong, A., Parsons, A.B., Krogan, N., Cagney, G., Mai, D., Greenblatt, J., et al. (2005). Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. Cell 120, 715-727.
Zhao, R., Kakihara, Y., Gribun, A., Huen, J., Yang, G., Khanna, M., Costanzo, M., Brost, R.L., Boone, C., Hughes, T.R., et al. (2008). Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. J Cell Biol 180, 563-578.

Figure legends

Figure 1: The solution NMR structure of free Tah1. (A) Backbone view of the 20 energetically lowest conformers of free Tah1 with representation of TPR1 (turquoise), TPR2 (grey), helix C (brown) and helix 3_{10} (red). This color code is used in all figures of this manuscript. (B) PyMOL transparent surface view of the free Tah1 energetically lowest conformer with inside ribbon representation of the protein chain, same color code as in (A) (DeLano, 2010). (C) Structure based sequence alignment of various TPR domains of eukaryotic Hsp70/90 co-chaperones. The GenBank sequences of Hsp70-binding TPR domains: TPR1 (P31948) and CHIP (Q9WUD1) and Hsp90-binding TPR domains: TPR2A (P31948), TPR2B (P31948), PP5 (P53041), Tah1 (P25638), and Tom71 (P38825) from Homo sapiens (Hs), Mouse musculus (Mm) or Saccharomyces cerevisiae (Sc) were aligned using CLUSTAL W and manual optimization of the alignment. The numbering corresponds to the Tah1 sequence. Identical residues are highlighted in blue rectangles. Residues exhibiting more than

50% of conservation are in grey boxes. Residues implicated in the carboxylate clamp are marked by stars above the Tah1 sequence. The TPR consensus sequence is aligned with the TPR1 and TPR2 sequences. Secondary structure elements are represented above the Tah1 sequence with the color code used in (A).

Figure 2: In cellulo mapping of the Tah1-Pih1 and Tah1-Hsp90 interactions. (A) Yeast 2H-tests of the Pih1 and Hsp90 interaction with full length Tah1, the Tah1 TPR domain and the Tah1 unfolded C-fragment. Tah1 and its fragments (same color code as in Figure 1) were used as the baits (pGBKT7-Gal4BD) and Pih1 (left picture) or Hsp90 (right picture) was the prey (pACTII-Gal4AD). Empty pACTII plasmid (-) was used as a negative control. Strains Y190 and Y187 are auxotrophic for histidine. Activation of the HIS3 reporter gene by interaction between the bait and prey is therefore needed for growth. Similar growths at high 3-Amino-1,2,4-triazol (3AT) concentration (a competitive inhibitor of the HIS3 gene product) indicate that full length Tah1 and the Tah1 C-terminal fragment have similar in cellulo affinities for Pih1. (B) Y2H tests for in cellulo interaction of full length Pih1 and its 1-251 N-domain and 253-344 C-domain with Tah1. Pih1 and its fragments were the baits (pGBKT7-Gal4BD) and Tah1 the prey (pACTII-Gal4AD). Full length Pih1 and its 253-344 C-domain showed similar Y2H interactions with Tah1.

Figure 3: NMR structure of the Tah1:Hsp90 peptide complex. (A) Ensemble of the NMR structures of Tah1 bound to the C-terminal peptide of Hsp90. Out of 50 calculated structures, 20 were selected by lowest total energy. The same color code as in Figure 1 is used for Tah1 and the peptide is in green. (B, C, and D) Front (B), bottom (C) and side (D) views of the best lower energy conformer of the 20 Tah1:Hsp90 structures represented in (A). Bound Tah1 is represented by PyMOL transparent surface view with inside ribbon representation of the protein chain. For clarity, the unfolded C-terminal domain of Tah1 is not displayed. The

ribbon representation of the interacting peptide backbone is in green and the amino acids are labeled.

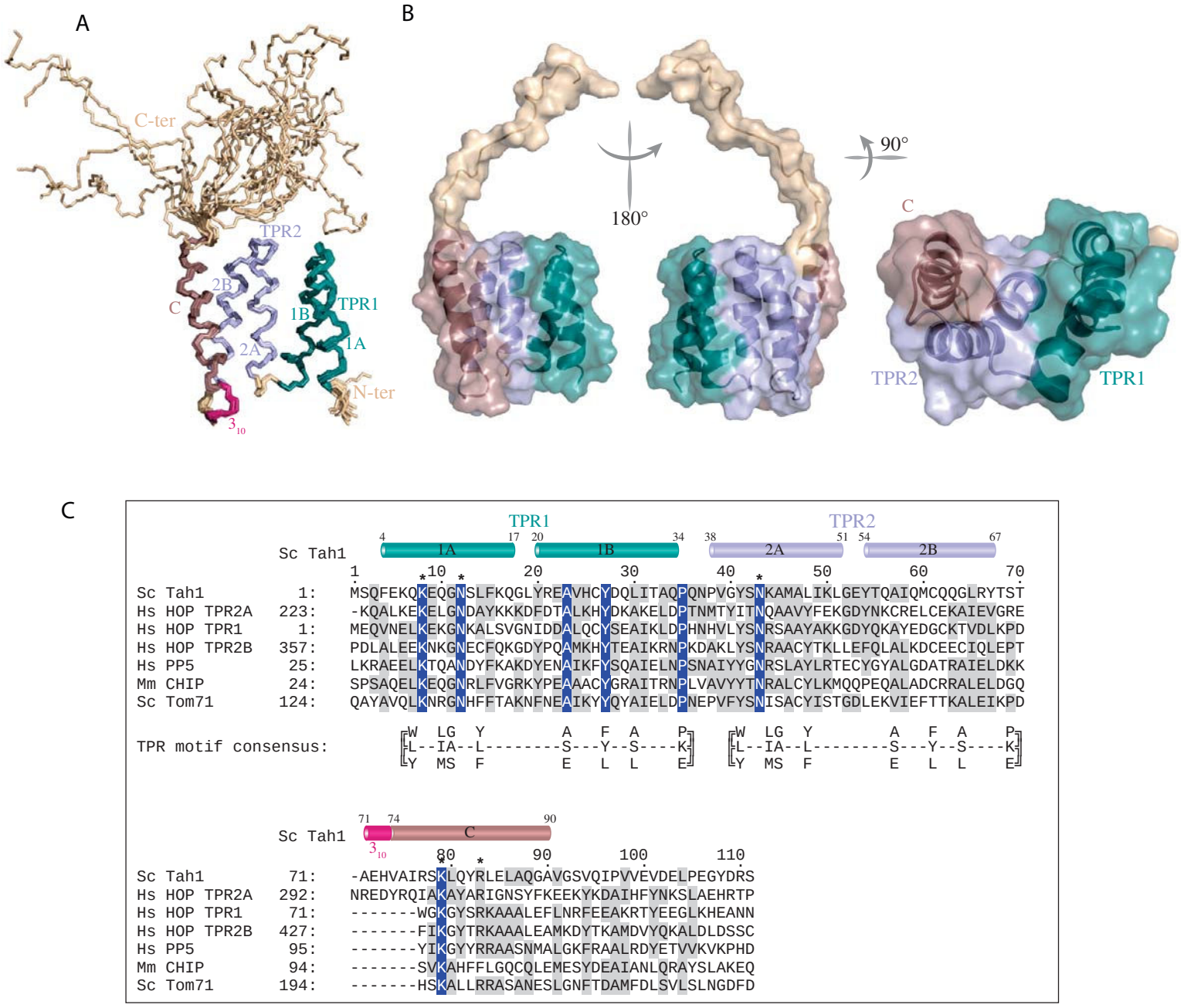
Figure 4: The six C-terminal amino acids of Hsp90 interact with the Tah1 TPR domain (A) Topology of the Tah1:Hsp90 interaction. A complete view of the interaction with ribbon representation of Tah1 and stick representation of the peptide is shown at the center. Magnified views of various areas of the interaction are shown in various insets. For clarity, some of them are rotated relative to the central panel. Hydrogen bonds and salt bridges are symbolized by dotted and full lines, respectively. Interactions with helices C and 2A and interactions with helices 2A and 1A are shown on the left and right insets, respectively. Hydrogen bonds and salt bridges are symbolized by dotted and full lines, respectively. The left and right insets show the numerous interactions established by helix C and helixes 1A-B respectively. (B) Summary of the interaction network at the Tah1:Hsp90 peptide interface. The primary sequence of the Hsp90 C-terminal peptide is developed and the amino acids labeled. Red, blue and violet lines represent salt bridges, H-bonds and Van Der Waals interactions established between Tah1 and the peptide, respectively.

Figure 5: Dissecting the Tah1 helix C and Hsp90 amino acids critical for Tah1-Hsp90 interaction in vivo (A) Tah1 Lys79, Tyr82 and Arg83 play a critical role for Tah1-Hsp90 interaction in vivo. Y2H assays were performed as in Figure 2, using 3-Amino-1,2,4-triazol (3AT) as a competitive inhibitor of the HIS3 gene product. WT Tah1 and its variants were the bait (pGBKT7-Gal4BD) and Hsp90 or Pih1 the prey (pACTII-Gal4AD). The Hsp90 interaction with Tah1 was abrogated by Arg77Ala, Lys79Ala or Arg83Ala substitution, while Arg77Ala and Glu85Ala had no negative effects. Positive 2H interactions between Tah1 variants and Pih1 (right picture) confirmed the expression of Tah1 variants. On the left, the positions of the 5 mutated amino acids are shown on the structure of the Tah1-Hsp90 complex.

(B) An essential role of Hsp90 D^{709} and an important role of Hsp90 V^{708} and E^{706} for in vivo interaction with Tah1. In the Y2H assays shown on the right, Hsp90 and its variants were used as the prey (pACT2-Gal4AD) and Tah1 or Pih1 as the bait (pGBKT7I-Gal4BD). On the left, the positions of the mutated Hsp90 amino acids are shown on the 3D structure of the complex.

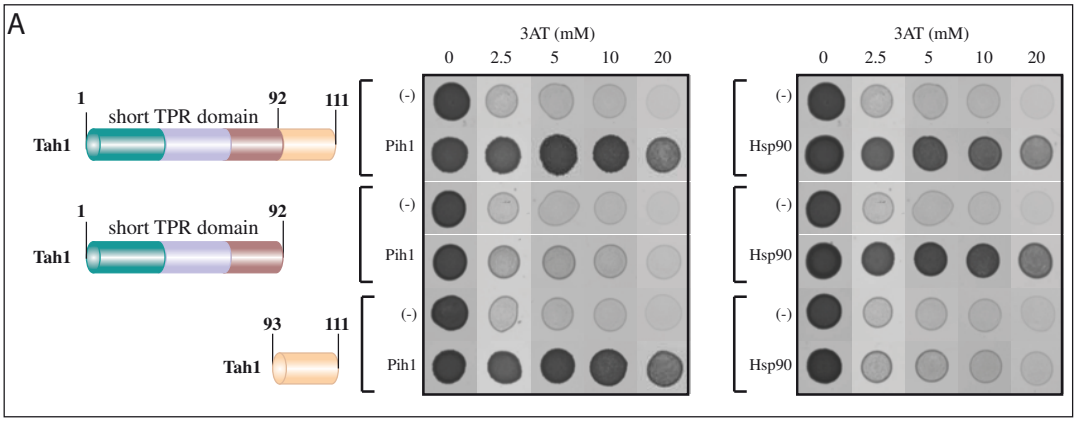
Figure 6: Comparison of bound states of Tah1 TPR domain with Hop TPR2A and CHIP TPR domains. Our yeast Tah1:ADTEMEEVD peptide structure is overlayed with structures of the yeast Tah1*-MEEVD peptide (2L6J) complex established by Jimenez et al., (Jimenez et al., 2012) (A), the human Hop TPR2A-MEEVD peptide (1ELR) (B) (Scheufler et al., 2000) and mouse CHIP-DTSRMEEVD peptide (2C2L) (C) (Zhang et al., 2005) complexes. The ribbon representation of the Tah1 TPR in wheat is overlayed on each of the other TPR domains. The Cα traces of the Hsp90 peptides are colored in pink (structure of Jimenez et al.), green (our structure), orange (TPR2A complex) and blue (CHIP complex) and the alpha carbons of peptide residues are indicated by a full black sphere. All right side views are rotated by 90°. The structure alignments were performed by PyMOL using 4 to 70 backbone residues as initial parameters. Superimposition of our Tah1 bound TPR structure is better with bound Hop TPR2A and CHIP TPR structures than with the bound Tah1* TPR (Jimenez et al., 2012).

Figure 7: Detailed structural comparison of the Tah1:ADTEMEEVD peptide complex with CHIP: DTSRMEEVD peptide (A) and TPR2A-MEEVD peptide (B) complexes. The representation and color coding are the same as in figure 6. The side-chains of residues in inter-molecular interactions are represented in stick and are labeled. The side chains of ADT and DTS residues in peptide ligands of Tah1 and CHIP, respectively, are not represented.


A
C-ter
TPR2
2B
C
1B
TPR1
2A
1A
N-ter
3 10
B
180°
90°
C
TPR2
TPR1
C
TPR1
TPR2
Sc Tah1
1A
1B
2A
2B
Sc Tah1 1: MSQFEKQKEQGNSLFKQGLYREAVHCYDQLITAQPQNPVGYSNKAMALIKLGEYTQAIQMCQQGLRYTST
Hs HOP TPR2A 223: -KQALKEKELGNDAYKKKDFDTALKHYDKAKELDPTNMTYITNQAAVYFEKGDYNKCRELCEKAIEVGRE
Hs HOP TPR1 1: MEQVNELKEKGNKALSVGNIDDALQCYSEAIKLDPHNHVLYSNRSAAYAKKGDYQKAYEDGCKTVDLKPD
Hs HOP TPR2B 357: PDLALEEKNKGNECFQKGDYPQAMKHYTEAIKRNPKDAKLYSNRAACYTKLLEFQLALKDCEECIQLEPT
Hs PP5 25: LKRAEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYALGDATRAIELDKK
Mm CHIP 24: SPSAQELKEQGNRLFVGRKYPEAAACYGRAITRNPLVAVYYTNRALCYLKMQQPEQALADCRRALELDGQ
Sc Tom71 124: QAYAVQLKNRGNHFFTAKNFNEAIKYYQYAIELDPNEPVFYSNISACYISTGDLEKVIEFTTKALEIKPD
TPR motif consensus:
W LG Y A F A P
L--IA--L--------S---Y--S----K
Y MS F E L L E
Sc Tah1
C
3 10
Sc Tah1 71: -AEHVAIRSKLQYRLELAQGAVGSVQIPVVEVDELPEGYDRS
Hs HOP TPR2A 292: NREDYRQIAKAYARIGNSYFKEEKYKDAIHFYNKSLAEHRTP
Hs HOP TPR1 71: -------WGKGYSRKAAALEFLNRFEEAKRTYEEGLKHEANN
Hs HOP TPR2B 427: -------FIKGYTRKAAALEAMKDYTKAMDVYQKALDLDSSC
Hs PP5 95: -------YIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
Mm CHIP 94: -------SVKAHFFLGQCQLEMESYDEAIANLQRAYSLAKEQ
Sc Tom71 194: -------HSKALLRRASANESLGNFTDAMFDLSVLSLNGDFD

C

FIGURE 1.


A
3AT (mM)
0 2.5 5 10 20
1
92
111
short TPR domain
Tah1
(-)
Pih1
Hsp90
93

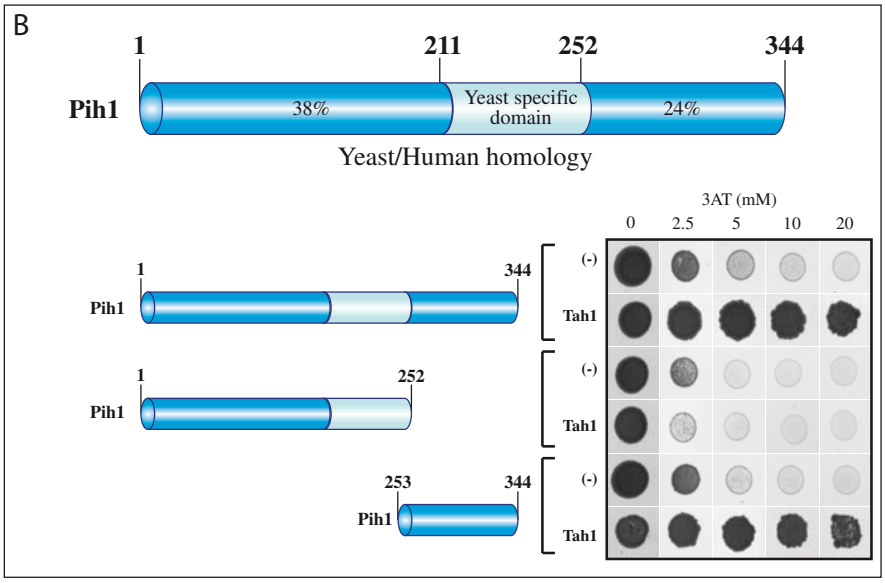

B
1
211
252
344
Pih1
38%
Yeast specific domain
24%
Yeast/Human homology
3AT (mM)
0 2.5 5 10 20
Tah1
(-)
253

FIGURE 2 .

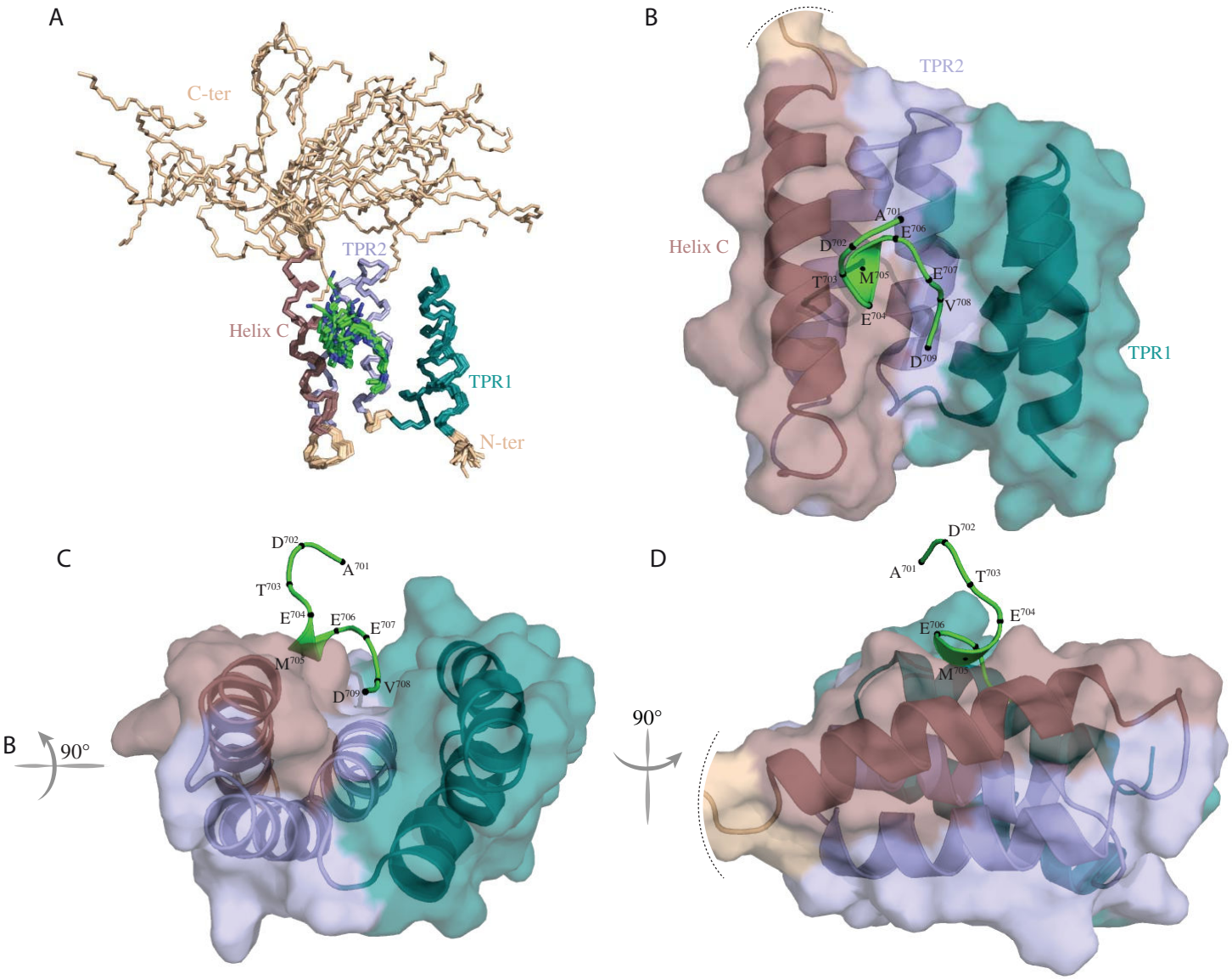
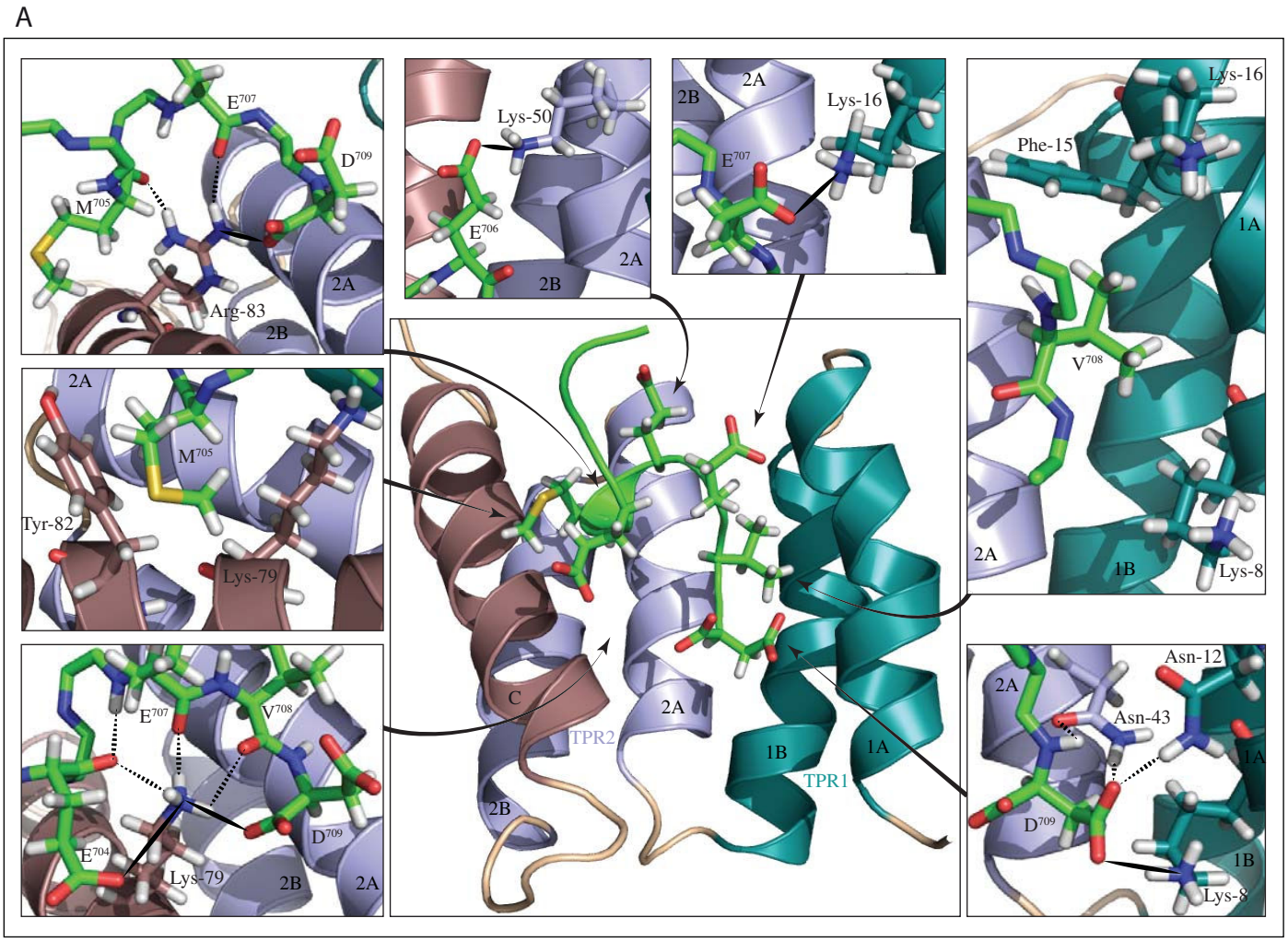

A
C-ter
TPR2
Helix C
TPR1
N-ter
B
TPR2
Helix C
A701
E706
D702
T703
M705
E707
E704
V708
D709
TPR1
C
D702
A701
T703
E704
E706
E707
M705
D709
V708
B
90°
D
D702
A701
T703
E704
E706
M705
90°

FIGURE 3.


A
E707
D709
M705
Arg-83
2A
2B
Lys-50
E706
2A
2B
2A
2B
Lys-16
E707
Lys-16
Phe-15
1A
V708
2A
1B
Lys-8
2A
M705
Tyr-82
Lys-79
C
2A
TPR2
1B
1A
TPR1
2B
E707
V708
D709
E704
Lys-79
2B
2A
2A
Asn-12
Asn-43
1A
D709
1B
Lys-8

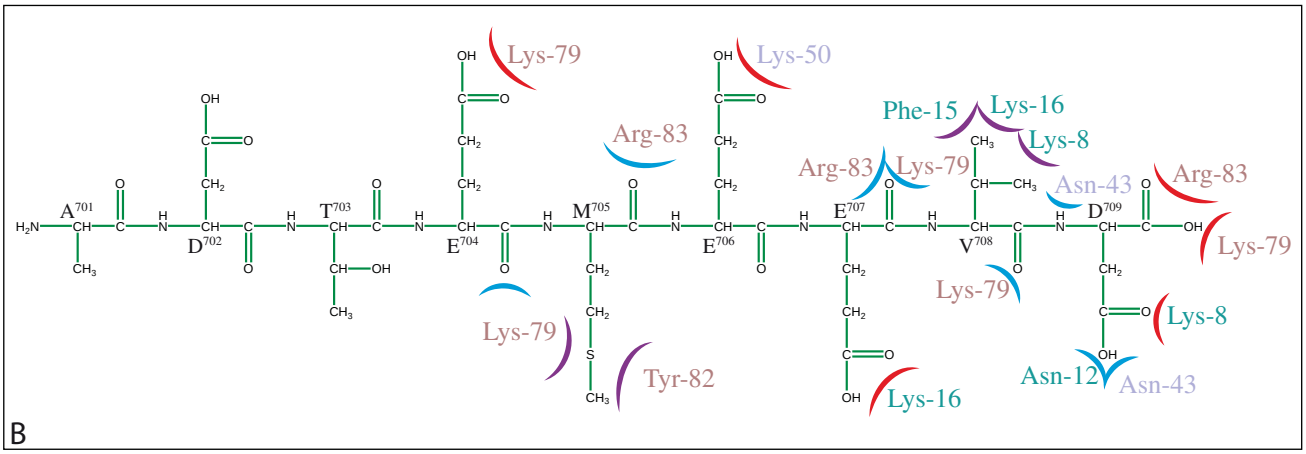
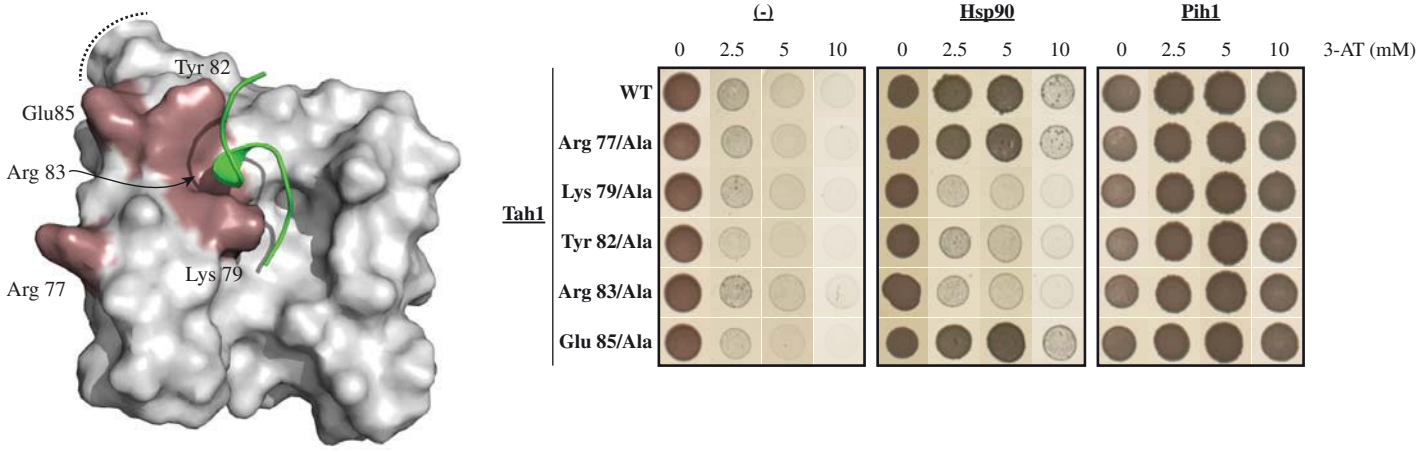

Lys-79
Lys-50
Arg-83
Phe-15
Lys-16
Lys-8
Arg-83
Lys-79
Asn-43
Arg-83
Lys-79
Lys-79
Lys-8
Lys-79
Tyr-82
Lys-16
Asn-12
Asn-43
B

FIGURE 4.

A


Tyr 82
Glu85
Arg 83
Arg 77
Lys 79
(-)
Hsp90
Pih1
0 2.5 5 10 0 2.5 5 10 0 2.5 5 10 3-AT (mM)
Tah1
WT
Arg 77/Ala
Lys 79/Ala
Tyr 82/Ala
Arg 83/Ala
Glu 85/Ala

B

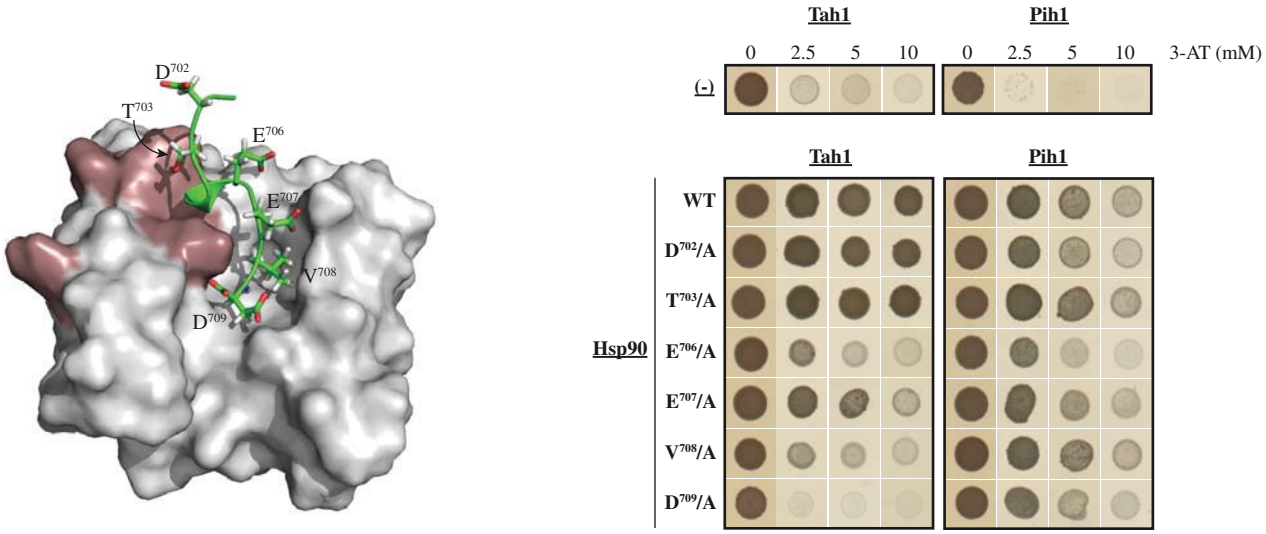

D702
T703
E706
E707
V708
D709
Tah1
Pih1
0 2.5 5 10 0 2.5 5 10 3-AT (mM)
(-)
Tah1
Pih1
Hsp90
WT
D702/A
T703/A
E706/A
E707/A
V708/A
D709/A

FIGURE 5 .

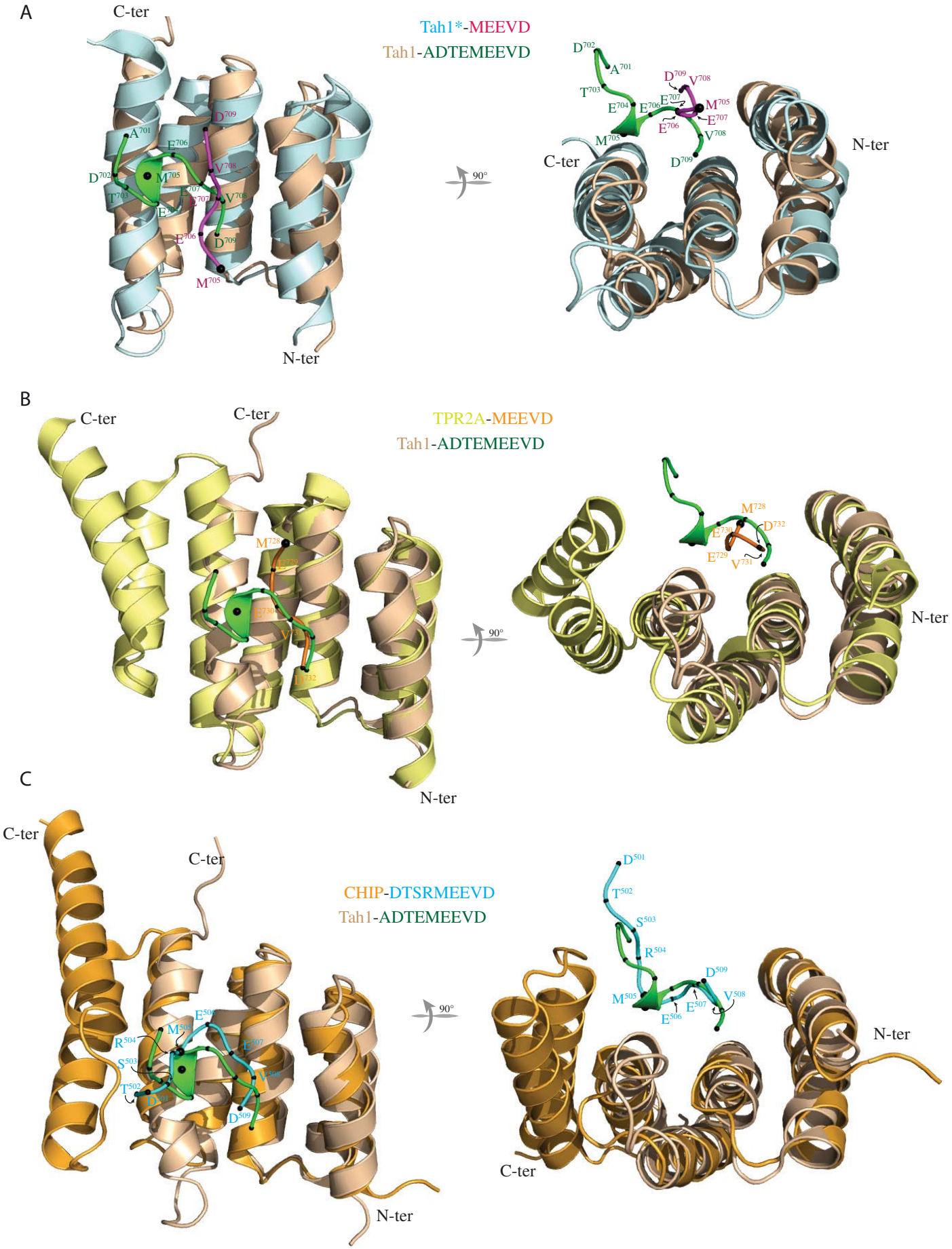

A
C-ter
Tah1*-MEEVD
Tah1-ADTEMEEVD
N-ter
90°
C-ter
N-ter
B
C-ter
C-ter
TPR2A-MEEVD
Tah1-ADTEMEEVD
90°
N-ter
N-ter
C
C-ter
C-ter
CHIP-DTSRMEEVD
Tah1-ADTEMEEVD
90°
N-ter
C-ter
N-ter

FIGURE 6.

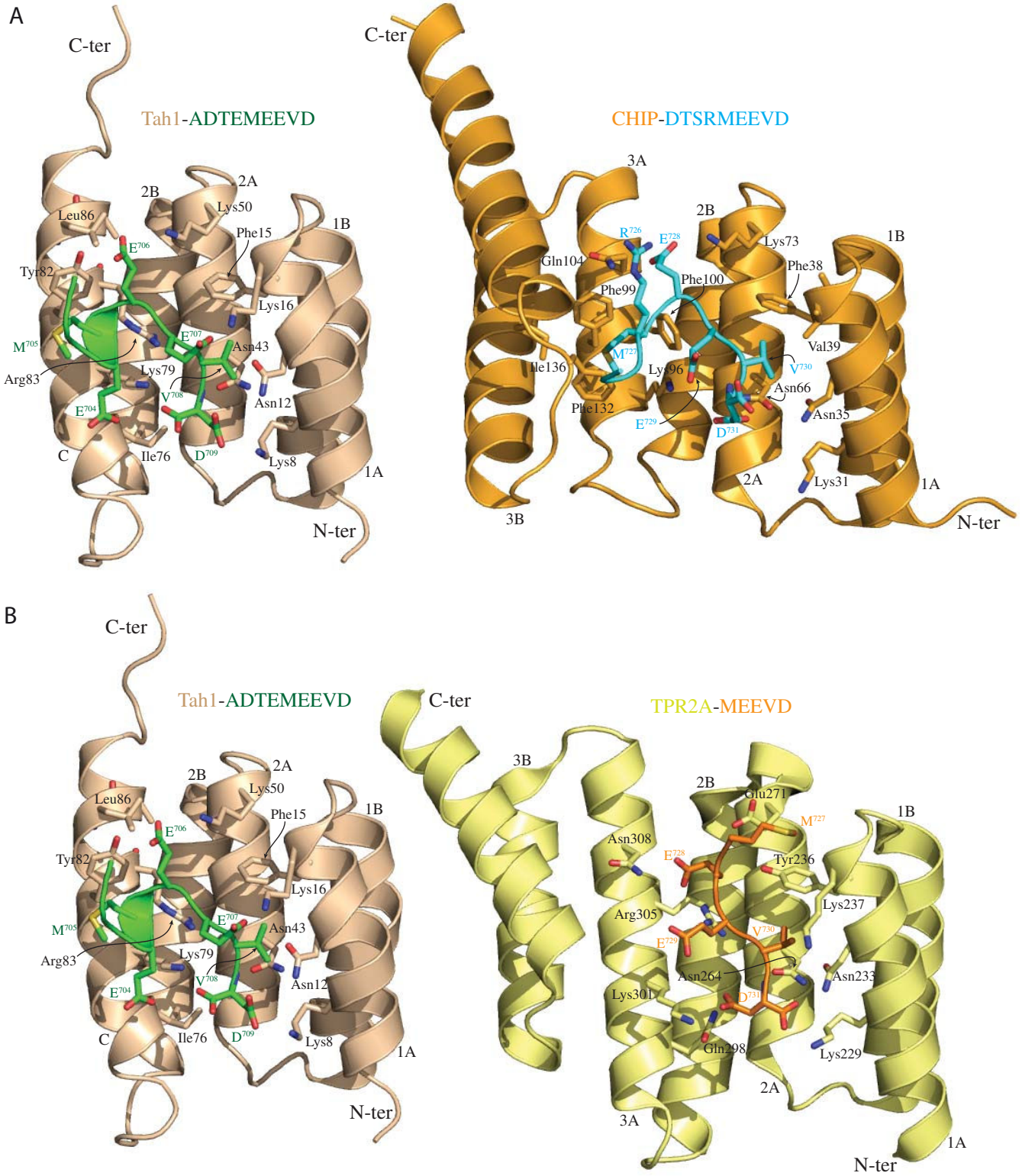

A
C-ter
Tah1-ADTEMEEVD
2B
2A
1B
1A
Leu86
Lys50
Phe15
E706
Tyr82
Lys16
M705
E707
Asn43
Arg83
Lys79
V708
Asn12
E704
C
Ile76
D709
Lys8
N-ter
C-ter
CHIP-DTSRMEEVD
3A
2B
1B
R726
E728
Lys73
Gln104
Phe38
Phe99
Phe100
M727
Val39
Ile136
Lys96
V730
Asn66
Phe132
E729
D731
Asn35
2A
Lys31
1A
3B
N-ter
B
C-ter
Tah1-ADTEMEEVD
2B
2A
1B
Leu86
Lys50
Phe15
E706
Tyr82
Lys16
M705
E707
Asn43
Arg83
Lys79
V708
Asn12
E704
C
Ile76
D709
Lys8
1A
N-ter
C-ter
TPR2A-MEEVD
3B
2B
Glu271
M727
1B
Asn308
E728
Tyr236
Lys237
Arg305
V730
E729
Asn264
Asn233
Lys301
D731
Gln298
Lys229
2A
3A
1A
N-ter

FIGURE 7.

Table

	Free form	Complex		
	Tah1 protein	Tah1 protein		Yeast Hsp90 peptide
NMR distance constraints				
• **Total Intramolecular Noes**	1786	2659		56
o **Intra-residue**	482	610		19
o **Inter-residues**				
▪ **Sequential**	448	687		25
▪ **Medium range**	411	776		12
▪ **Long range**	445	586		0
• **Total Intermolecular Noes**			55	
Structure statistics				
• **RMSD to mean structure**				
o **Backbone atoms**	0.44 ± 0.08 (a)	0.36 ± 0.10 (a)	0.42 ± 0.10 (a+b)	0,56 ± 0.21 (b)
o **Heavy atoms**	0.98 ± 0.13 (a)	0.86 ± 0.12 (a)	0.92 ± 0.12 (a+b)	1.30 ± 0.25 (b)
• **RMSD from idealized covalent geometry**				
o **Bond length (Å)**	0.0038 ± 0.0001		0.0042 ± 0.0001	
o **Bond angle (°)**	1.43 ± 0.02		1.74 ± 0.03	
• **Violations**				
o **Average number of violations (>0.4 Å)**	0.00		9.6	
o **Highest violation (Å)**	0.39		0.95	
• **Energies (kcal.mol^{-1})**				
o **Average AMBER**	-4009 ± 20		-4146 ± 17.5	
o **Average constraint violations energy**	96 ± 10		530 ± 18	
• **Ramachandran analysis**				
o **Most favoured**	70.6		74.8	
o **Additionnaly allowed region**	28.3		22.9	
o **Generously allowed region**	1.0		2.1	
o **Disallowed region**	0.1		0.2	

(a) r.m.s. deviation was calculated using residues 4-90 for the ensemble of 20 refined structures
(b) r.m.s. deviation was calculated using residues 704-709 for the ensemble of 20 refined structures

Table 1: NMR structure determination statistics of the Tah1 protein and Tah1-yeast Hsp90 complex

Supplemental Data

Supplemental Legends

Supplemental Figure S1: Mapping of the Tah1:Hsp90 peptide interaction by chemical shift perturbations and superposition of free and bound forms of Tah1. (A) ^{15}N-^{1}H HSQC spectral overlays of free (black) and bound forms (red) of Tah1 protein [1:1 protein:peptide molar ratio]. (B) Overlay of the Tah1 TPR domain in free and bound states. The ribbon representations of bound and free Tah1 are colored in light and dark brown, respectively. Only the side chains of hydrophobic amino acids involved in the interaction between the helixes 2B and C are developed. The right side view is rotated by 90. (C) Bar graphs represent normalized amide chemical shift perturbations per residues for Tah1 in complex with Hsp90 peptide. Amide chemical shift changes were normalized using the equation: $\Delta\delta^{HN}=\sqrt{((\Delta HN)^2+(\Delta N/6.51)^2)}$. The horizontal red dot line indicates the threshold for significant chemical shift changes (average +1 of standard deviation). Residues that with significant chemical shift changes are indicated (red: >0.4 ppm, orange: 0.3 to 0.4 ppm, and yellow: <0.3 ppm).

Supplemental Figure S2: Flexibility of Tah1 protein in free and bound forms and electrostatic potentials of free Tah1 protein. (A) Superposition of heteronuclear NOE intensities for each residues of Tah1 in free (black point) and bound (red square) forms. (B) Vacuum electrostatic potentials mapped on the Tah1 molecular surface in two opposite views using VASCo Plugin software (Steinkellner et al., 2009). Blue, white and red regions correspond to positive, neutral and negative electrostatic potentials respectively.

Supplemental Figure S3: Summary of the 55 inter-molecular NOEs of the Tah1:Hsp90 peptide interface. The sequence of the Hsp90 C-terminal peptide

(ADTMEEVD) was developed and the amino acids labeled using the same nomenclature and color code as in figure 1. The type of protons and the names and positions of each Tah1 residues giving NOE cross-peaks with peptide are shown by arrows.

Supplemental Figure S4: Electrostatic potentials of Tah1 protein in complex with Hsp90 peptide. (A) Vacuum electrostatic potentials mapped on the molecular surface of bound Tah1, the Hsp90 C-terminal peptide is represented in stick mode with side chains. The color code is the same as in figure S2B. (B) Vacuum electrostatic potentials mapped on the molecular surface of the Tah1:Hsp90 peptide complex. The process and the color code are like in figure S2B.

Supplemental Experimental Procedures

Biological Material used in this study

Recombinant pACTII and pGBKT7 plasmids expressing Gal4AD and Gal4BD fusion proteins, respectively, were obtained by cloning various coding sequences between the following pairs of restriction sites BamHI/XhoI and BamHI/SalI, respectively. DNA inserts were obtained by PCR amplification using appropriate primers containing the restriction sites. These amplifications were performed on genomic DNA of the Saccharomyces cerevisiae BY4741 strain (Euroscarf). The amplified yeast DNA sequences encoded: Tah1(1-111), Tah1(1-92), Tah1(93-111), Pih1(1-344), Pih1(1-252), Pih1(253-344), Hsp82(1-709) and Tah1 R77/A, Tah1 K79/A, Tah1 R83/A, Tah1 E85/A, Hsp82 D^{702}/A, Hsp82 D^{703}/A, Hsp82 E^{706}/A, Hsp82 E^{707}/A, Hsp82 V^{708}/A, Hsp82 D^{709}/A variants.

NMR measurements

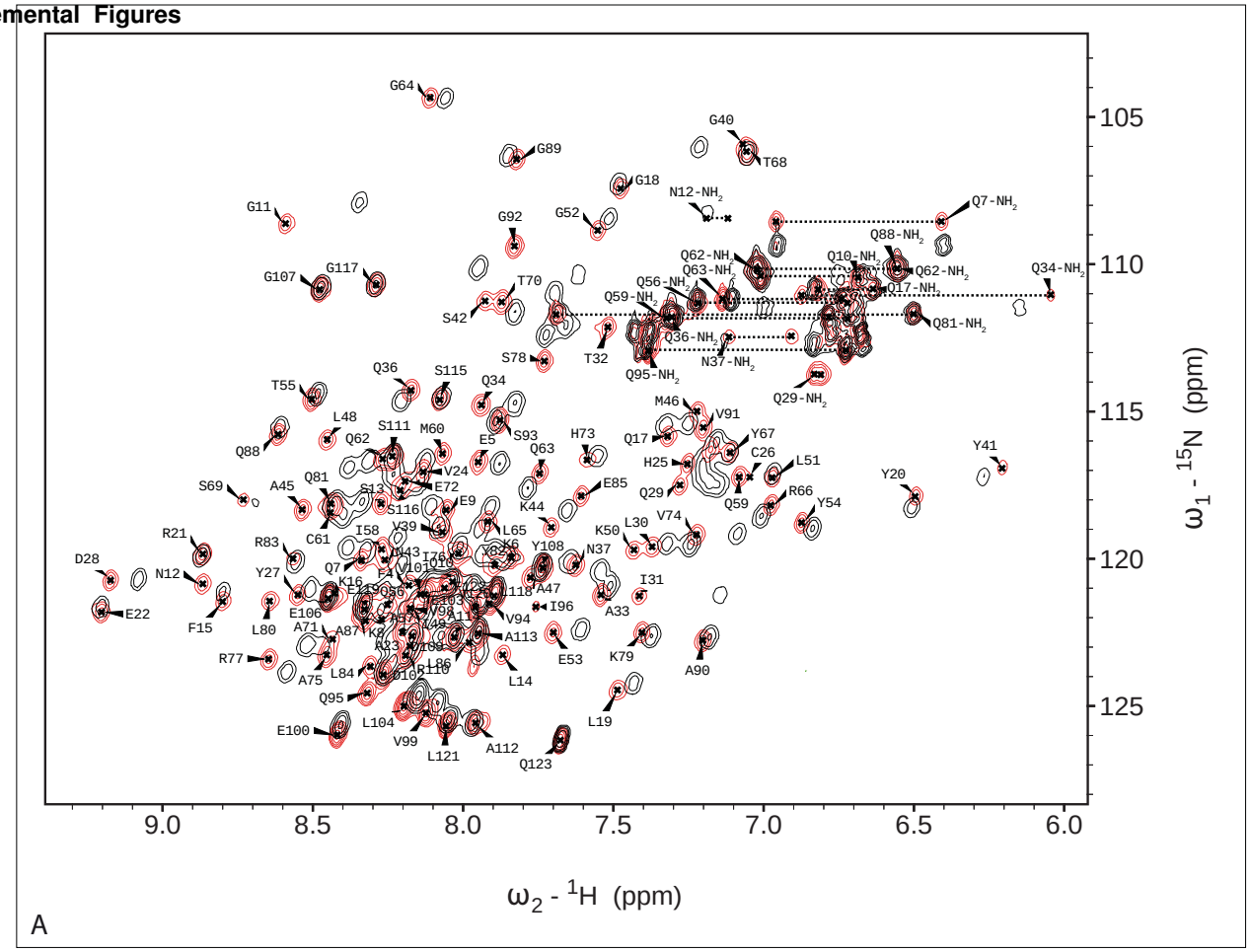
Sequence-specific backbone assignments were achieved using 2D (^{15}N-^{1}H) and (^{13}C-^{1}H) HSQCs, 3D HNCO, 3D HN(CA)CO, 3D CBCANH, and 3D CBCA(CO)NH experiments.

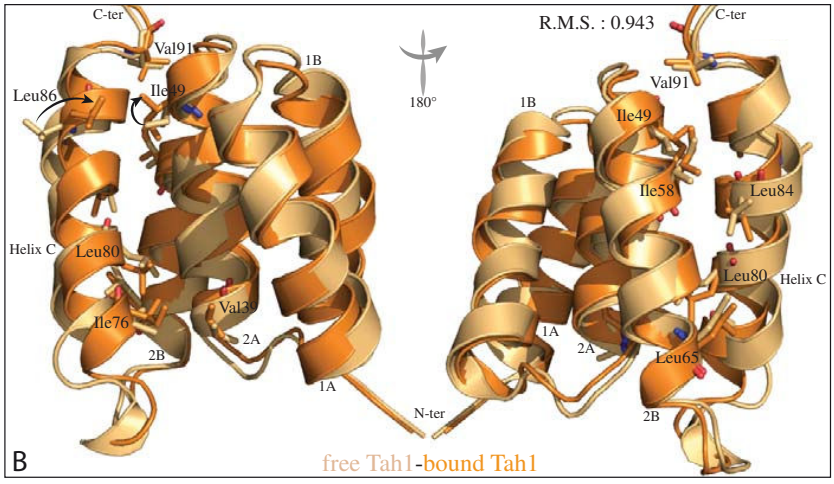
Side chain assignments were performed using 3D H(C)CH-TOCSY, 3D HNHA, 3D HBHANH, 3D HBHA(CO)NH, 3D NOESY-(^{15}N-^{1}H)-HSQC and 3D NOESY-(^{13}C-^{1}H)-HSQC experiments. Aromatic proton assignments were performed using 2D-TOCSY and 2D-NOESY in 100% D_2O.

Relaxation measurements were performed in the same sample conditions using ^{15}N labeled Tah1 protein at a ~1 mM concentration and a Bruker AVIII-600 MHz spectometer.

We assigned the bound Tah1 backbone and side chains using 2D (^{15}N-^{1}H) and (^{13}C-^{1}H) HSQC, 3D HNCA, 3D hCcccoNH, 3D HccccoNH, 3D NOESY (^{15}N-^{1}H) and (^{13}C-^{1}H) HSQC experiments, taking advantage of the backbone and side chains assignments of free Tah1. We assigned the aromatic proton of the bound Tah1 using 2D (^{13}C-^{1}H)-HSQC in H_2O and 100% D_2O as well as 3D NOESY-(^{13}C-^{1}H)-HSQC recorded in the aromatic region.

Supplemental Figures





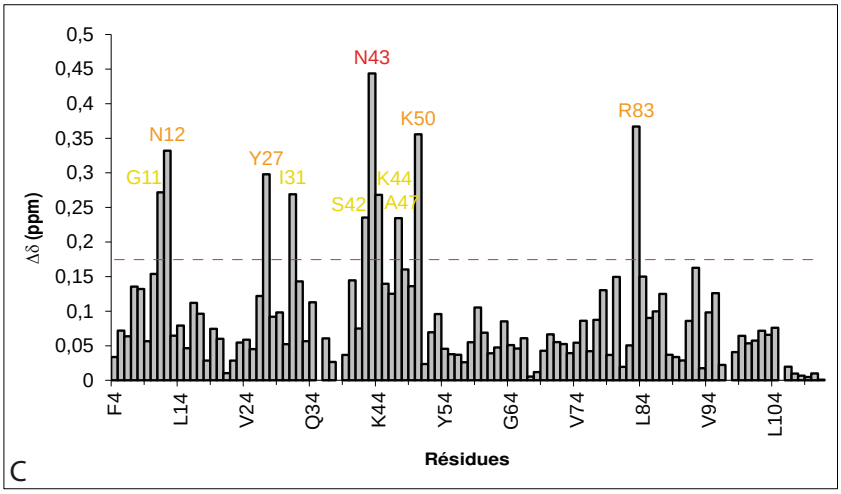
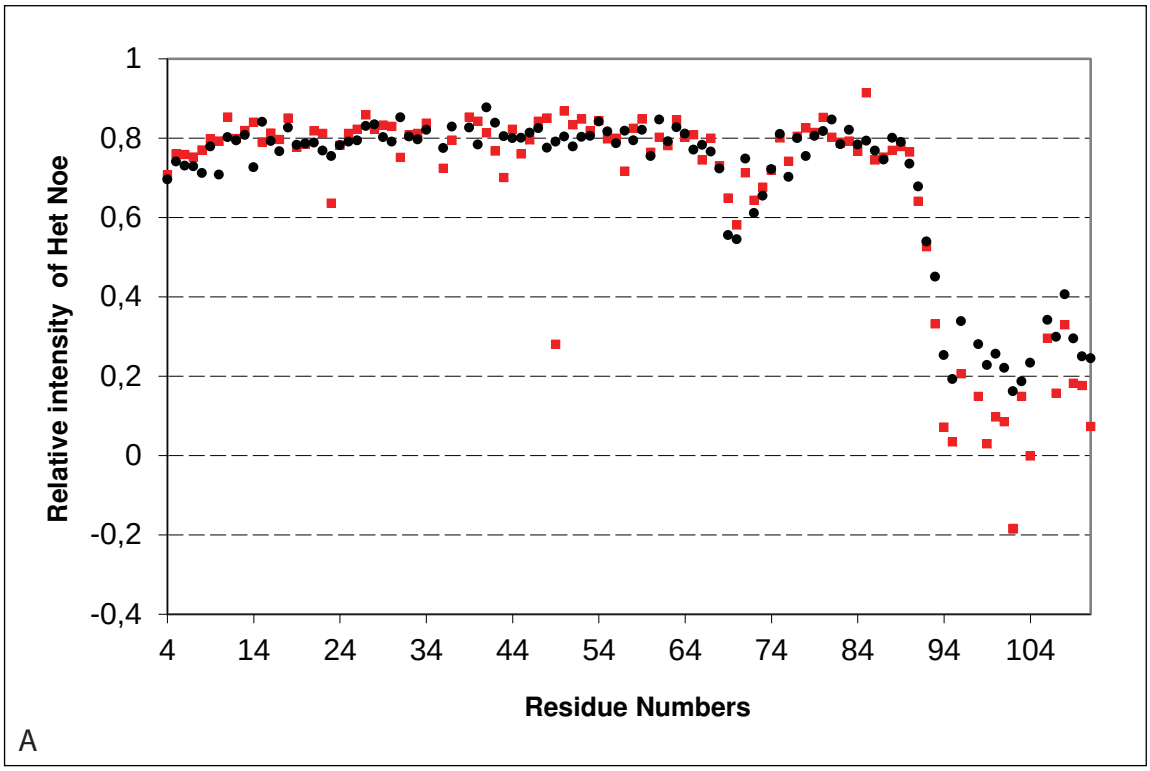


FIGURE S1.


Relative intensity of Het Noe
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
4
14
24
34
44
54
64
74
84
94
104
Residue Numbers
A

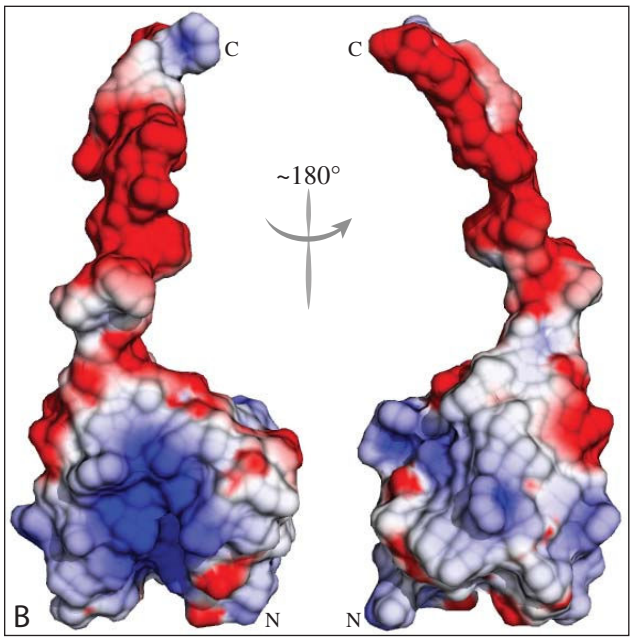

C
C
~180°
B
N
N

FIGURE S2.

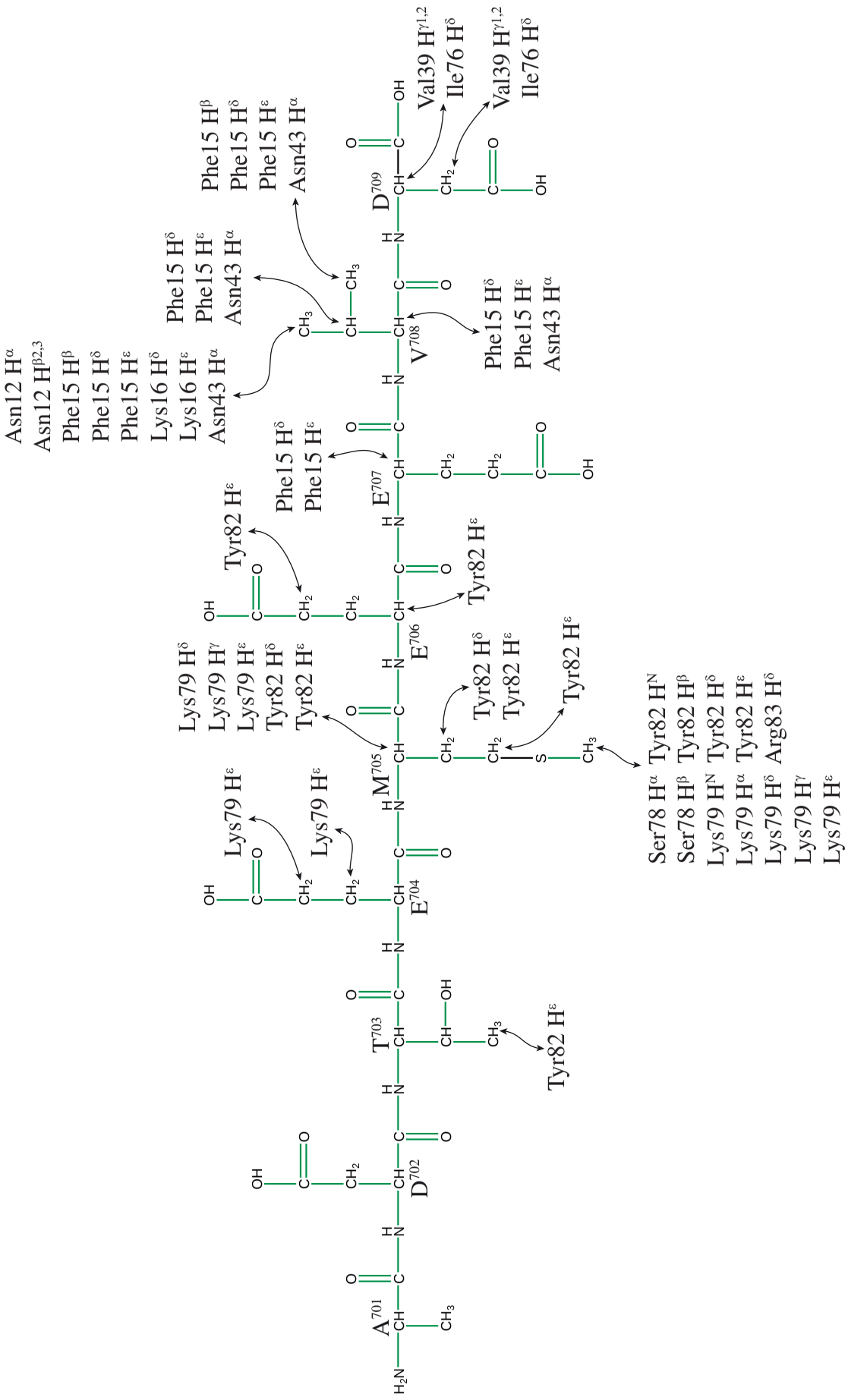

Asn12 Hα
Asn12 Hβ2,3
Phe15 Hβ
Phe15 Hδ
Phe15 Hε
Lys16 Hδ
Lys16 Hε
Asn43 Hα
Phe15 Hδ
Phe15 Hε
Asn43 Hα
Phe15 Hβ
Phe15 Hδ
Phe15 Hε
Asn43 Hα
Val39 Hγ1,2
Ile76 Hδ
Val39 Hγ1,2
Ile76 Hδ
Phe15 Hδ
Phe15 Hε
Asn43 Hα
Phe15 Hδ
Phe15 Hε
Tyr82 Hε
Tyr82 Hε
Lys79 Hδ
Lys79 Hγ
Lys79 Hε
Tyr82 Hδ
Tyr82 Hε
Tyr82 Hδ
Tyr82 Hε
Tyr82 Hε
Lys79 Hε
Lys79 Hε
Tyr82 Hε
Ser78 Hα
Ser78 Hβ
Lys79 HN
Lys79 Hα
Lys79 Hδ
Lys79 Hγ
Lys79 Hε
A701
D702
T703
E704
M705
E706
E707
V708
D709

FIGURE S3

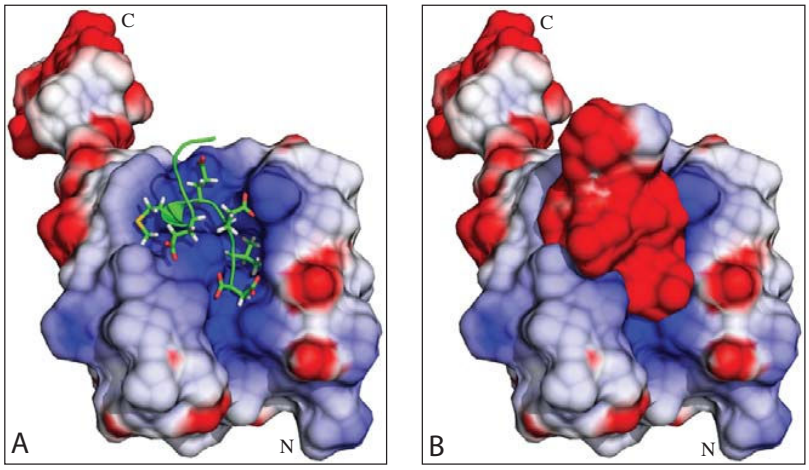


FIGURE S4.

2.2 Compléments concernant les analyses du complexe Tah1p-Hsp90 décrites dans le manuscrit n°3

L'établissement de la structure tridimensionnelle du complexe Tah1p-ADTEMEEVD que nous venons de décrire avait été réalisée sur un échantillon où seule la protéine Tah1p était doublement marquée ^{15}N ^{13}C. En effet le peptide issu d'une synthèse chimique ne pouvait pas être enrichi en ces isotopes. Afin de déterminer avec précision la structure du complexe, il était nécessaire d'obtenir un maximum d'informations à la fois sur le peptide et sur les contacts entre la protéine et le peptide. Pour cela nous avons adopté une approche généralement utilisée pour l'étude des complexes ARN-protéine en solution par R.M.N. : les expériences filtrées et/ou éditée. Elles reposent sur le marquage isotopique ^{15}N ^{13}C de la protéine pour filtrer les informations provenant de ces noyaux et ainsi obtenir des informations (les NOEs ou *« Nuclear Overhauser effect »*) sur la distance entre les atomes $^{15}N/^{13}C$ de la protéine et ^{1}H du peptide. Ceci est rendu possible par l'utilisation des expériences NOESY (« *Nuclear Overhauser Effect SpectroscopY*») filtrée-éditée. Une expérience TOCSY (*« TOtal Correlation SpectroscopY »*) filtrée-filtrée nous a permis, d'atteindre les déplacements chimiques des protons du peptide au sein du complexe. Un exemple d'informations R.M.N. obtenus par ces approches est présenté dans la figure 97.

2.3 Travaux complémentaires réalisés sur la protéine Pih1p non décrits dans le manuscrit

La protéine Pih1p fait partie du complexe R2TP (Zhao *et al.*, 2005) qui s'associe avec la chaperonne Hsp90 ((Eckert *et al.*, 2010) et Manuscrit n°3). Elle interagit aussi avec la protéine Nop58p ((Gonzales *et al.*, 2005) et nos données décrites dans le chapitre 1). Elle pourrait donc faire le lien entre la machinerie d'assemblage R2TP et les snoRNP à boîtes C/D.

2.3.1 Travaux réalisés sur l'extrémité N-terminale de Pih1p

Sur la base d'alignement de séquences, de prédictions de structures secondaires, de données de spectrométrie de masse (Eckert *et al.*, 2010) et de prédiction de solubilité nous avons pu déterminer différents domaines de la protéine Pih1p.

Nous avons ainsi défini la région 1-180 de Pih1p comme étant un domaine bien défini, soluble et ayant une séquence en acides aminés conservée chez les levures. La taille de ce fragment étant compatible avec une étude structurale par R.M.N., nous avons alors entrepris

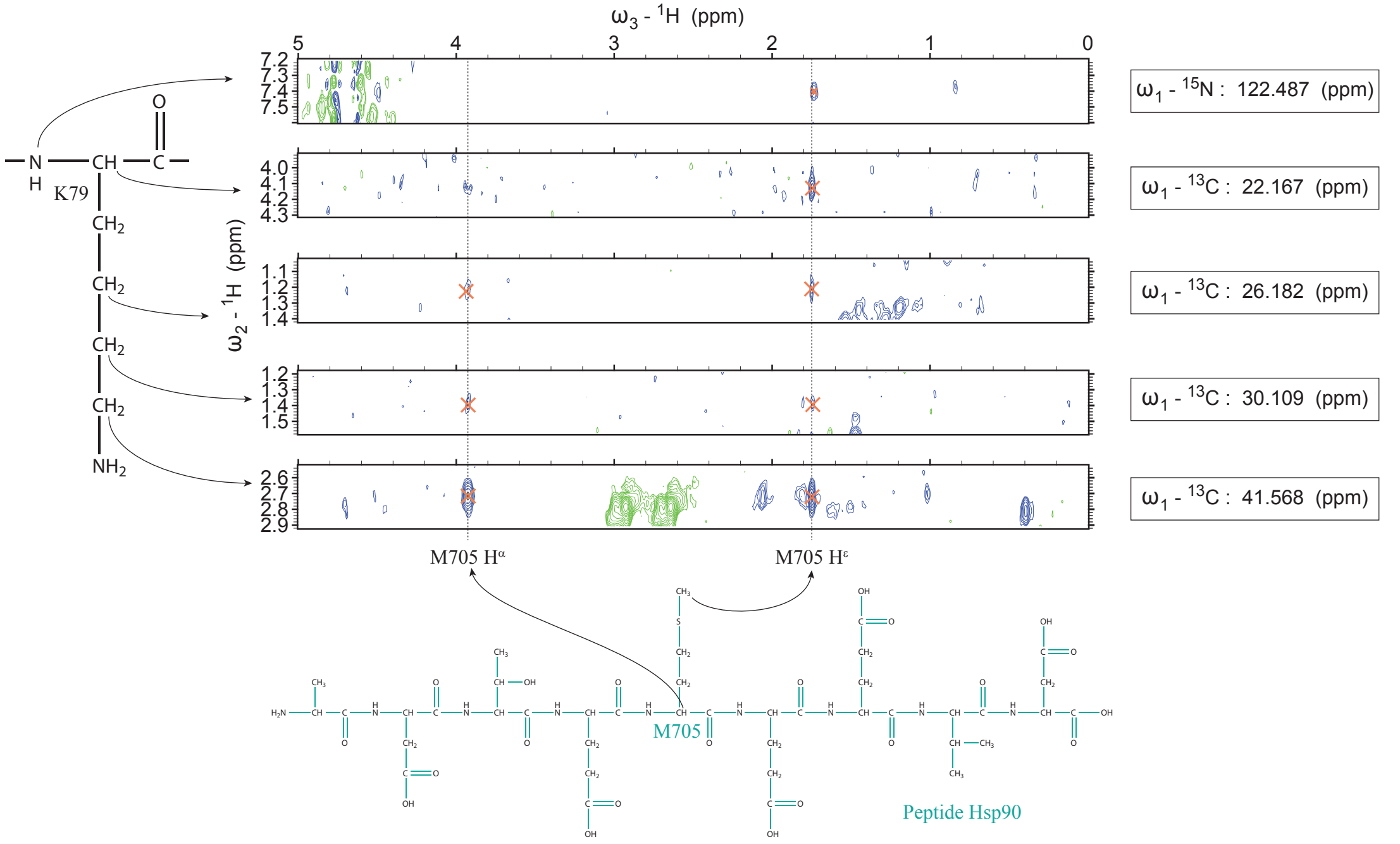


Figure 97. **Exemple de NOE intermoléculaires observés au sein du complexe Tah1p-peptide d'Hsp90.** Les expériences R.M.N. NOESY-HSQC ^{15}N filtrée-éditée et NOESY-HSQC ^{13}C filtrée-éditée nous ont permis d'observer des taches de corrélation entre les protons de la protéine Tah1p (ici ceux de la lysine 79) et du peptide (ici ceux de la méthionine 705). Sur la base de ces NOE intermoléculaires, nous avons pu déterminer la structure tridimensionnelle du complexe avec une grande précision.

sa production et sa purification en grande quantité. La protéine Pih1p (1-180) simplement et doublement marquée aux isotopes ^{15}N et ^{13}C a ainsi pu être obtenue à une concentration de 20 mg/ml (soit 1 mM). Les spectres enregistrés n'ont malheureusement pas pu être exploités sans doute liée au fait que la protéine présentait une instabilité au cours du temps dans les expériences R.M.N réalisées à 288°K (Figure 98). En effet, dans les expériences ^{15}N-^{1}H TROSY-HSQC, nous observions l'apparition de pics dans la région 130-132 ppm qui s'amplifiaient au cours du temps et aussi une modification importante de la région centrale du spectre. Une analyse par spectrométrie de masse nous a permis de constater que la protéine Pih1p (1-180) se dégrade au cours du temps en un fragment stable (14-180).

De plus, les précipités observés dans le tube R.M.N. ressemblaient à des « pastilles » solides et non pas à un précipité de type « blanc d'œuf ». La relative solubilité de la protéine et son instabilité conduisant à une structure relativement ordonné nous ont conduit à réaliser avec C. Charron (IR en cristallographie, CNRS) des tests de cristallisation sur la protéine Pih1p (1-180). Les premiers résultats ont donné lieu à l'apparition dans certaines conditions (Tableau 11) d'oursins de cristaux à la surface de la goutte. Afin d'améliorer ces résultats, nous avons redéfini différents fragments du domaine N-terminal sur la bases des données de spectrométrie de masse, de prédiction de solubilité et de prédiction de sites potentiels de clivage par les protéases. Nous avons ainsi cloné les régions codantes des fragments 15-180, 21-180, 15-208, 21-208 et 1-208 de Pih1p. Ces différents fragments ont été produits et purifiés (Figure 99). Ils ont ensuite été concentrés jusqu'à atteindre une concentration d'environ 10 mg/ml et testés en cristallisation. Comme pour le fragment 1-180, nous avons observé pour certains fragments, et notamment le fragment 1-208, des oursins de cristaux. Ces expériences ont été poursuivies au laboratoire par C. Charron sur le domaine 1-180 de Pih1p, et des cristaux ont été obtenus, notamment dans la condition D2 du kit JBS Classic 2 (Jena Bioscience) (Figure 100). Ces résultats sont encourageants nous conduisant à poursuivre l'étude par diffraction des rayons X du domaine N-terminal de la protéine Pih1p qui interagit avec Nop58p et Rsa1p.

2.3.2 Travaux réalisés sur l'extrémité C-terminale de Pih1p

Des analyses précédentes (Eckert *et al.*, 2010) et nos données de doubles hybrides (Manuscrit n°3) avaient clairement établi le fait que l'extrémité C-terminale de Pih1p s'associait avec Tah1p, en particulier la région 93-111. Nous avons alors co-exprimé le fragment 93-111 de Tah1p avec le domaine 212-344 de Pih1p. Il s'est avéré après analyse par

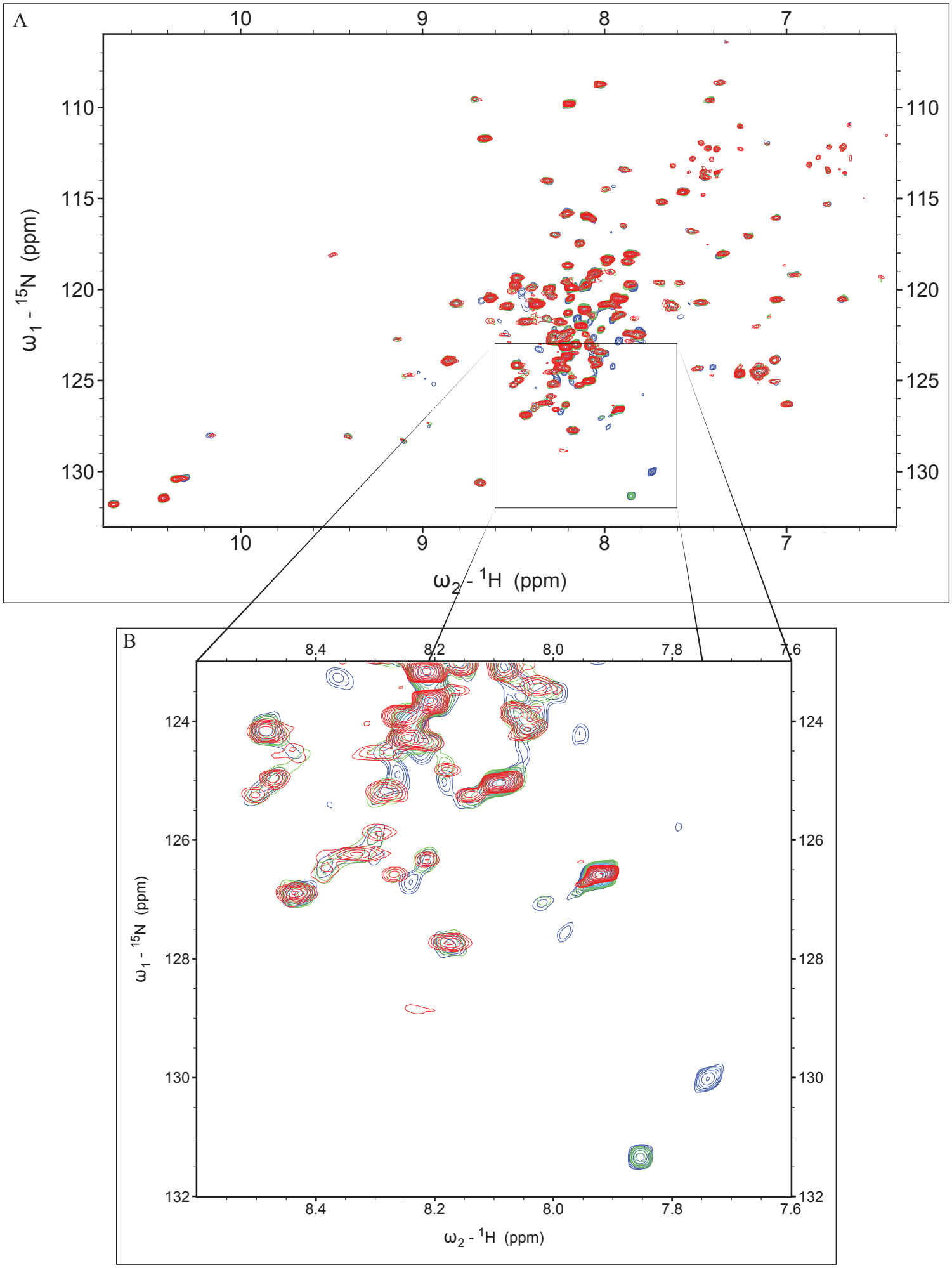


Figure 98. **Analyse par R.M.N. de la protéine Pih1p à 288°K.** L'échantillon était utilisé à une concentration de 20,2 mg/ml (1 mM) dans un tampon phosphate 10 mM, pH 6,6 ; NaCl 150 mM. (A) Superposition des spectres ^{15}N-^{1}H TROSY-HSQC de la protéine Pih1p 1-180 au temps To, To+3 jours et To+7 jours en rouge, vert et bleu respectivement. (B) Agrandissement d'une partie du spectre ^{15}N-^{1}H trosy HSQC. Le code couleur est le même que dans le panneau A.

Fournisseur	Kit	Condition n°	Sels	Tampon	pH	Agent precipitant
Molecular dimension	Structure Screen 2	19	Aucun	Na HEPES 0.1 M	7.5	10 % w/v PEG 8000, 8 % v/v ethylene glycol
		26	$(NH_4)_2SO_4$ 0,2 M	MES 0.1 M	6.5	30 % w/v PEG 5000 MME
		42	Aucun	Aucun		25 % v/v ethylene glycol
	PACT	43	NaCl 0,2 M	Tris 0.1 M	8.0	20 % w/v PEG 6000
Hampton research	Index	7	Aucun	Acide Citrique 0.1 M	3.5	3.0 M Sodium chloride
		47	Aucun	BIS-TRIS 0.1 M	6.5	28% w/v PEG MME 2,000
		71	NaCl 0,2 M	BIS-TRIS 0.1 M	6.5	25% w/v PEG 3,350
		84	$MgCl_2,(H_2O)_6$ 0,2 M	HEPES 0.1 M	7.5	25% w/v PEG 3,350
		86	$KNaC_4H_4O_6$, $(H_2O)_4$ 0,2 M	Aucun		20% w/v PEG 3,350
Jena Bioscience	JBS Classic 1	C1	NaCl 0,1 M	Bicine 100 mM	9.0	30 % w/v PEG 550 MME
	JBS Classic 2	C3	$MgCl_2$ 0,2 M	Na MES 0,1 M	6.5	25 % w/v PEG 4000
		D2	$CaCl_2$ 0,2 M	Na HEPES 0,1M	7.5	30 % w/v PEG 4000

Tableau 11. **Composition des solutions utilisées pour les tests de cristallisation de la protéine Pih1p (1-180) qui ont permis d'obtenir des cristaux en forme d'oursins.**

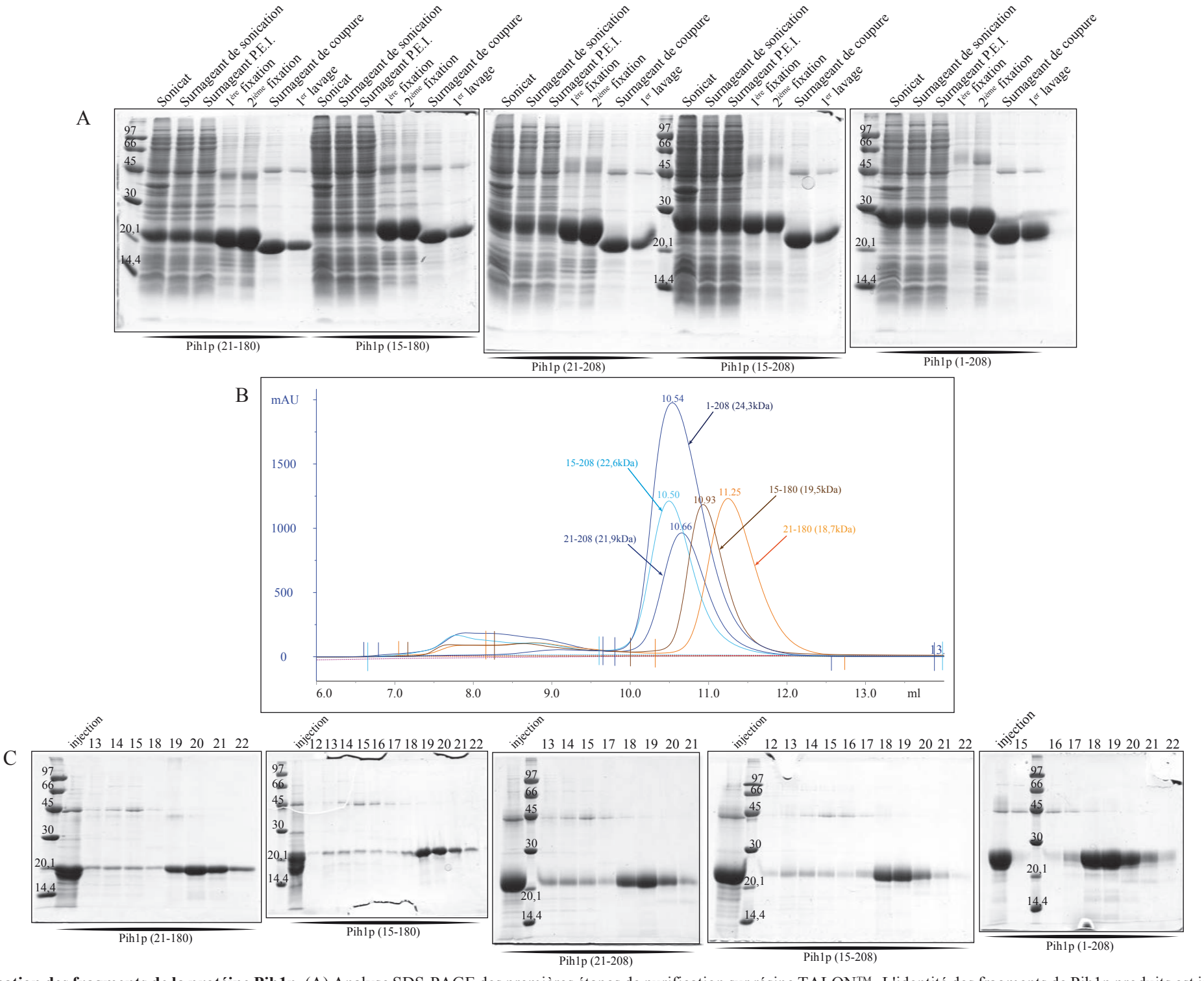


Figure 99. **Purification des fragments de la protéine Pih1p.** (A) Analyse SDS-PAGE des premières étapes de purification sur résine TALON™. L'identité des fragments de Pih1p produits est indiqué sous chaque gel. (B) Superposition des différentes chromatographies d'exclusion de taille (Superdex 75) obtenus pour les différents fragments de Pih1p. Leurs identitées sont indiquées sur les chromatogrammes. (C) Analyse par SDS-PAGE des fractions correspondantes aux pics obtenus lors des chromatographies d'exclusion de taille.

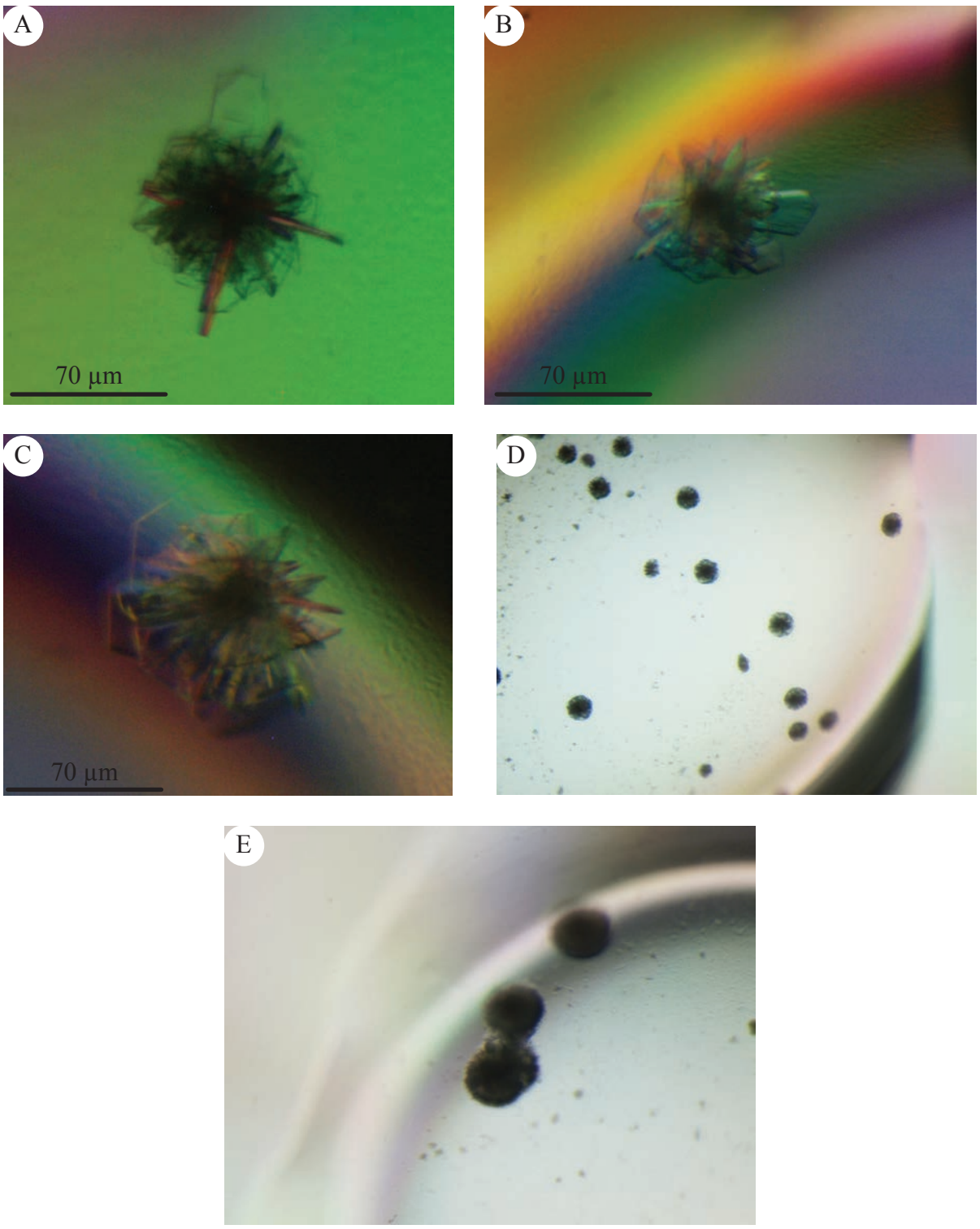


Figure 100. **Photographies des oursins et des cristaux obtenus pour le fragment Pih1p (1-180).** La protéine était utilisée à une concentration de 10 mg/ml en tampon Hepes 10 mM, pH 7,5 ; NaCl 150 mM, β-mercaptoethanol 5 mM. Les cristaux sont apparu dans la condition D2 du kit JBS Classic 2 (Jena Bioscience). Le tampon contenait de l'Hepes 0,1M, pH 7,5 ; $CaCl_2$ 0,2 M ; 30 % w/v PEG 4000 à 20°C. La cristallisation était réalisée en goutte assise. (A) Goutte de 4 µl avec un ratio échantillon protéique : solution de cristallisation de 1:1. (B) et (C) Goutte de 6 µl avec un ratio échantillon protéique : solution de cristallisation de 1:2. (D) et (E) Goutte de 6 µl avec un ratio échantillon protéique : solution de cristallisation de 2:1

spectrométrie de masse en condition native que le domaine 212-344 de Pih1p subissait une dégradation à son extrémité N-terminale pour donner lieu au fragment 256-344 et que ce fragment continuait à s'associer au fragment 93-111 de Tah1p.

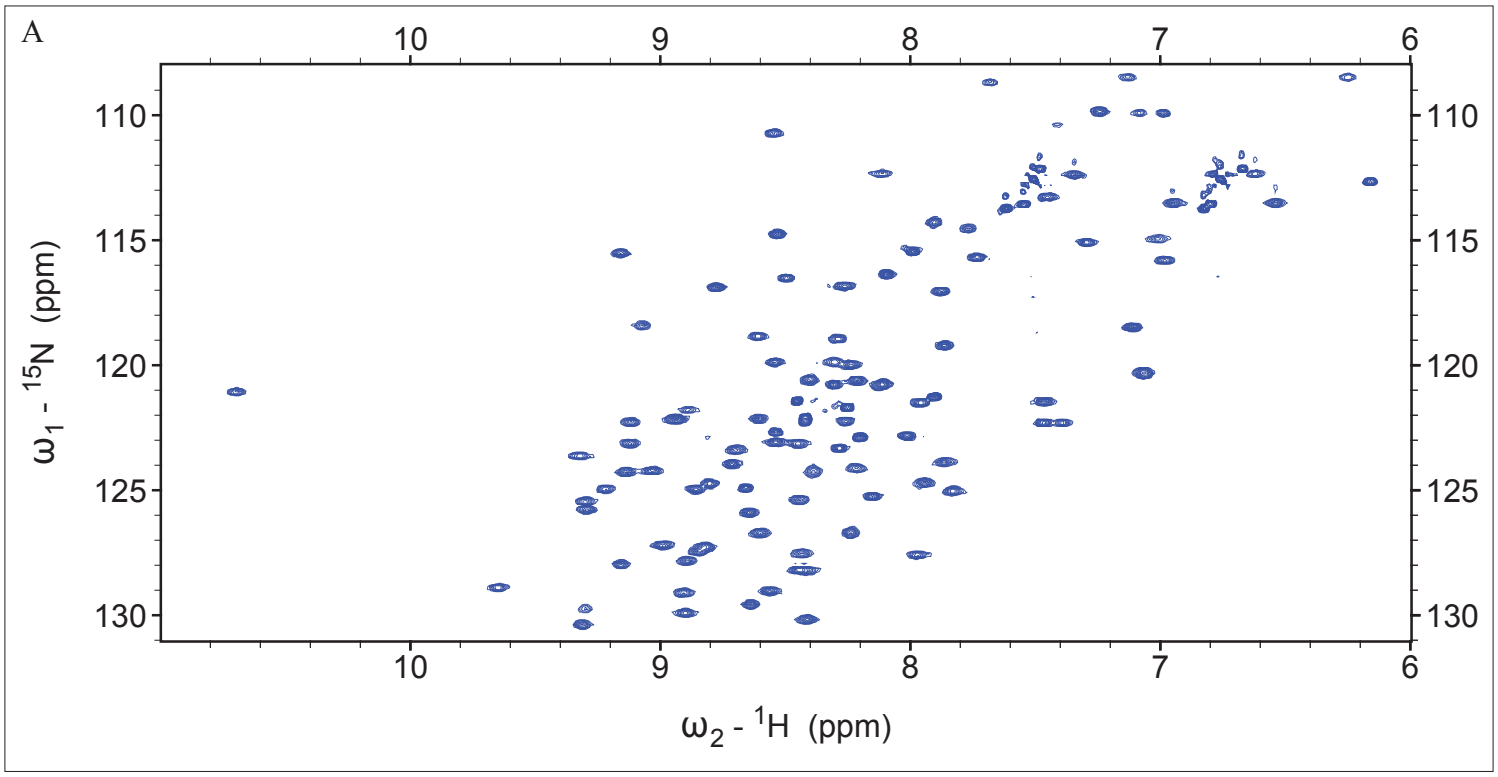
Après reclonage du fragment 256-344 de Pih1p, le complexe a été produit par co-expression avec le fragment de Tah1p (93-111) et purifié par chromatographie d'affinité grâce à l'étiquette 6xHis porté par le fragment de Tah1p. Les premières analyses réalisées par M. Quinternet sont présentées dans la figure 101. Le nombre de taches de corrélations observés dans l'expérience ^{1}H-^{15}N HSQC et leur dispersion sont tout à fait satisfaisants. Le nombre de signaux NOE, notamment au niveau des groupements NH de la chaine principale de cette expérience ^{1}H-^{1}H NOESY indique que les fragments Tah1p (93-111) et Pih1p 256-344 sont structurés. L'étude structurale de ce complexe par R.M.N. va donc être débutée.

3 Conclusion et perspectives

3.1 Bilan

Initialement, la protéine Tah1p avait été décrite comme une protéine à deux domaines, capable d'interagir avec Hsp90 par le biais de son domaine TPR et avec Pih1p par l'intermédiaire de sa région C-terminale (Zhao *et al.*, 2005; Zhao *et al.*, 2008). L'étude structurale de la protéine Tah1p a permis de définir plus précisément ces deux domaines en particulier de mettre en évidence la présence d'une cinquième hélice dans le domaine TPR qui joue un rôle majeur dans l'interaction avec le peptide C-terminal d'Hsp90. Nos données structurales ont aussi permis d'identifier les résidus mis en jeu dans l'association du domaine TPR atypique de Tah1p avec la chaperonne. Nous avons également révélé que le court segment C-terminal (93-111) de Tah1p est nécessaire et suffisant pour son interaction avec la protéine Pih1p. La protéine Tah1p assurerait donc l'association entre la protéine Pih1p et Hsp90 pouvant être en plus renforcée par quelques interactions directes mais moins stable entre ces deux protéines. Il a été montré qu'en phase stationnaire de croissance, Tah1p stabiliserait Pih1p et empêcherait sa dégradation (Zhao *et al.*, 2008).

Notre étude de la protéine Pih1p a révélé que son domaine C-terminal (256-344) est responsable de son interaction avec Tah1p, tandis que son domaine N-terminal (1-212), est capable de s'associer avec la protéine Nop58p (Manuscrit n°3, B. Rothé, travail de thèse). La capacité de Pih1p à interagir en plus avec le domaine C-terminal de la protéine Rsa1p



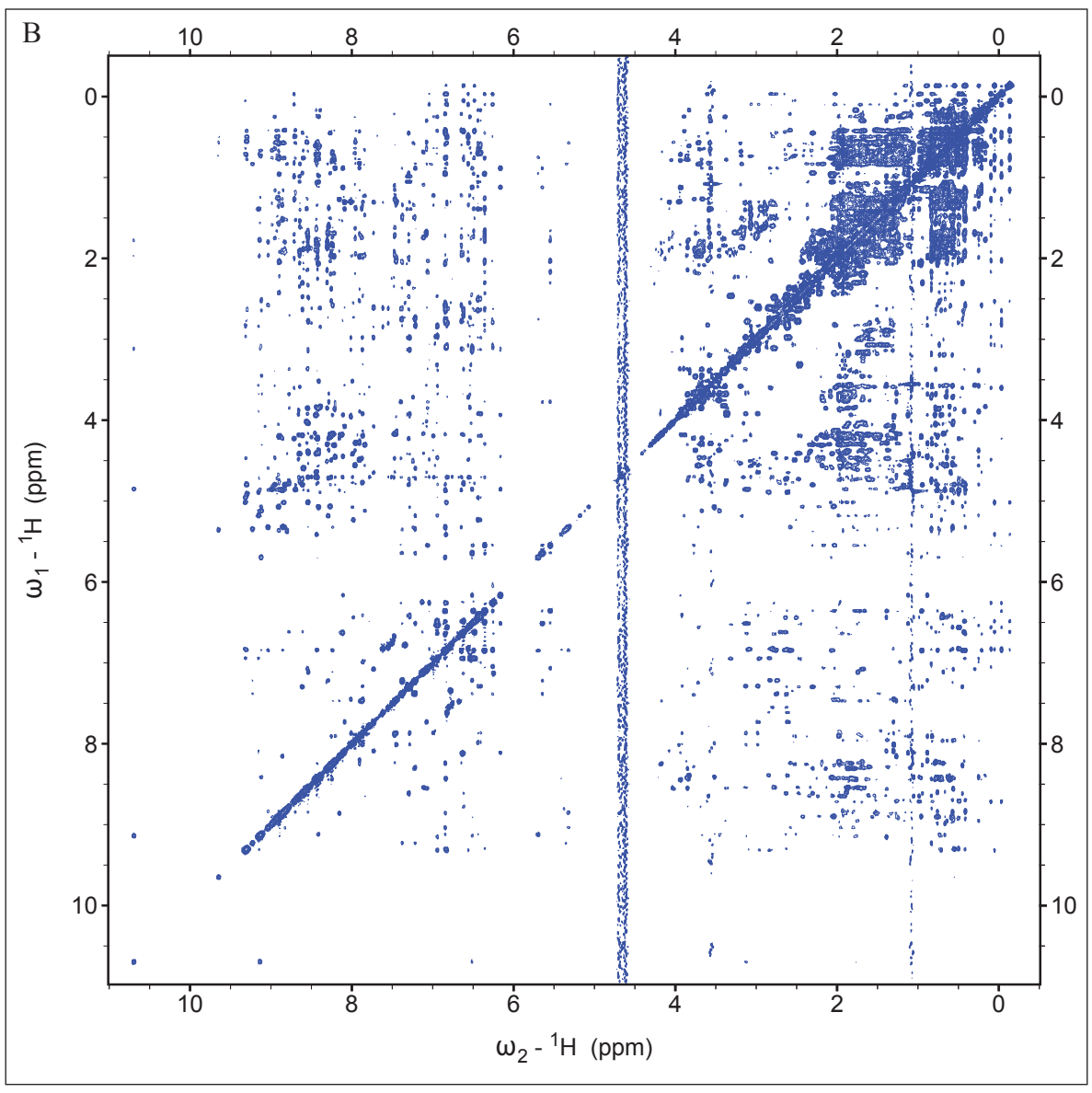


Figure 101. **Analyse par R.M.N. du complexe Tah1 (93-111)/Pih1p (256-344) à 298K.** L'échantillon a été utilisé à une concentration de 1,4 mM dans du tampon phosphate 10 mM, pH 6,4 ; NaCl 50 mM. (A) Spectre ^{1}H-^{15}N HSQC. (B) Spectre ^{1}H-^{1}H NOESY.

(Résultats et discussion, Chapitre 1, paragraphe 2.1.3) en fait une protéine centrale dans le réseau d'interaction des facteurs d'assemblages et des protéines constitutives des particules C/D. Elle semble assurer le lien entre le complexe R2TP et les autres composants impliqués dans la biogenèse des snoRNP C/D.

3.2 Etude structurale du complexe R2TP : la protéine Pih1p

Même si le rôle exact du complexe R2TP associé à Hsp90 dans la biogenèse des snoRNP à boîtes C/D n'est pas complétement élucidé ((Boulon *et al.*, 2008), B. Rothé, travail de thèse), les protéines constituant ce complexe R2TP, et notamment Pih1p établissent de nombreux contacts avec les protéines constituantes des snoRNP en particulier Nop58p (B. Rothé, travail de thèse) et des facteurs d'assemblage, en particulier Rsa1p. Une approche structurale globale du complexe R2TP semble être actuellement inabordable, c'est pourquoi l'étude des protéines Tah1p et Pih1p constituent un excellent point de départ pour aborder l'étude de la structure du complexe. Des structures des deux autres protéines RVB1 et RVB2 humaines ont été obtenus à haute résolution par diffraction des rayons X (Gorynia *et al.*, 2011). De ce fait, Pih1p est la dernière protéine du complexe dont la structure n'a pas encore été déterminée. Au cours de mon travail de thèse, j'ai pu définir différents domaines de la protéine Pih1p : le domaine N-terminal fonctionnel 1-211 qui est responsable des interactions avec Nop58p et Rsa1p et le domaine C-terminal 256-344 responsable de l'association avec Tah1p. Actuellement, grâce aux travaux que j'ai réalisé, chaque domaine fait l'objet d'une étude structurale soit par diffraction des rayons X (domaine N-terminal) soit par R.M.N. (extrémité C-terminal). L'obtention des structures 3D de ces domaines permettra de mieux comprendre les mécanismes qui régissent l'implication du complexe R2TP dans la biogenèse des snoRNP à boîtes C/D.

CONCLUSIONS
&
PERSPECTIVES

Conclusion & Perspectives

Etude structurale de la machinerie de biogenèse des snoRNP à boîtes C/D chez la levure Saccharomyces cerevisae

Les cellules eucaryotes, en plus de posséder de multiples machineries de nature protéique, font également appel à des complexes ribonucléoprotéiques composés de plusieurs protéines avec un ou plusieurs ARN non codants. Ces RNP sont des acteurs clés de nombreuses fonctions essentielles du vivant. Au cours de mon travail de thèse, nous nous sommes tout particulièrement intéressés aux relations structure/fonction des composants de la machinerie de biogenèse des snoRNP à boîtes C/D. Ces particules se caractérisent par un snoRNA à boîtes C/D possédant des séquences conservées, les boîtes C et D, mais aussi par la présence d'un ensemble de 4 protéines constitutives : les protéines Nop58p et Nop56p, l'enzyme Nop1p/Fibrillarine et une protéine appartenant à la famille des protéines L7Ae, la protéine Snu13p. Ces RNP interviennent dans un processus biologique essentiel, la maturation des ARNr ribosomiques qui conduit à la formation des ribosomes. L'action de ces RNP est par conséquent extrêmement précise, leur assemblage doit donc aussi être parfaitement contrôlé. En collaboration avec les équipes d'E. Bertrand et d'A. Krol, notre laboratoire avait découvert de nouveaux facteurs protéiques impliqués dans la biogenèse de ces particules (Boulon *et al.*, 2008). Pour comprendre le mécanisme d'action de ces différents facteurs, il était important d'apporter des informations structurales sur ces facteurs afin de pouvoir déterminer comment ils interviennent dans l'assemblage des snoRNP à boîtes C/D. C'est la mission qui m'avait été confiée pour ce travail de thèse. Il faut noter que les informations attendues devaient dépasser au niveau de leur impact le simple cadre de l'assemblage des snoRNP C/D puisque notre équipe et celles d'E. Bertrand et d'A. Krol avaient montré que plusieurs des facteurs d'assemblages sont aussi utilisés pour l'assemblage des snoRNP H/ACA, de la télomérase et des mRNP codant des sélénoprotéines.

1 La co-expression, un outil d'analyse pour la biologie structurale

Atteindre des informations structurale sur ces facteurs protéiques de levure impliquait de passer un premier goulot d'étranglement lié à la faible solubilité de la majorité des protéines étudiées. En effet, peu de ces protéines ne possédaient des caractéristiques de solubilité nécessaires à l'obtention d'échantillons compatibles avec une étude par R.M.N ou par radiocristallographie. Une étroite collaboration avec l'équipe de C. Romier nous a permis de mettre en place au laboratoire un système de co-expression dans la bactérie *E. coli*. Cette approche peut apporter une solution au problème posé par la présence des domaines non repliés ou flexibles car ces domaines peuvent se replier lors de leur association avec leurs partenaires naturels respectifs. Par ailleurs, l'association avec ces partenaires peut permettre de neutraliser les zones hydrophobes de surface qui limitent la solubilité des protéines exprimées de manière individuelle. Par co-expression une nouvelle entité est produite au sein de la bactérie. La présence d'une étiquette spécifique fusionnée à l'un des partenaires permet une purification du complexe par une chromatographie d'affinité comme s'il s'agissait d'une seule molécule. Le savoir-faire que j'avais acquis lors de mes stages successifs à Strasbourg dans le groupe de C. Romier, nous a ensuite permis d'implanter avec X. Manival une plateforme robotisée TECAN au laboratoire et de pouvoir ainsi continuer mes expériences. Ce nouvel outil et mes expériences réalisées à Strasbourg nous ont permis de réaliser des séries de criblage d'interaction à haut débit et aussi de mettre en évidence un réseau important d'interactions entre les protéines constitutives des snoRNP et leurs facteurs d'assemblages (Figure 102). Les résultats obtenus nous ont également permis de définir avec précision les domaines protéiques impliqués dans ces interactions. Couplée à une stratégie de mutagenèse dirigée (ou aléatoire), lorsqu'une interaction positive est observée, la co-expression permet d'affiner les connaissances sur les déterminants de l'interaction. Enfin, cet outil nous a aussi permis de définir les domaines des protéines pouvant être produits seuls ou en complexe sous forme soluble. Ce qui a permis de débuter un ensemble d'analyses structurales par RMN ou cristallographie des rayons X qui ont déjà produit des structures 3D importantes pour la compréhension du système d'assemblage des snoRNP et plusieurs autres sont en cours. Par ailleurs, les complexes et protéines purifiées ont aussi été utilisés par B. Rothé pour des études *in vitro* sur leur mécanisme d'action

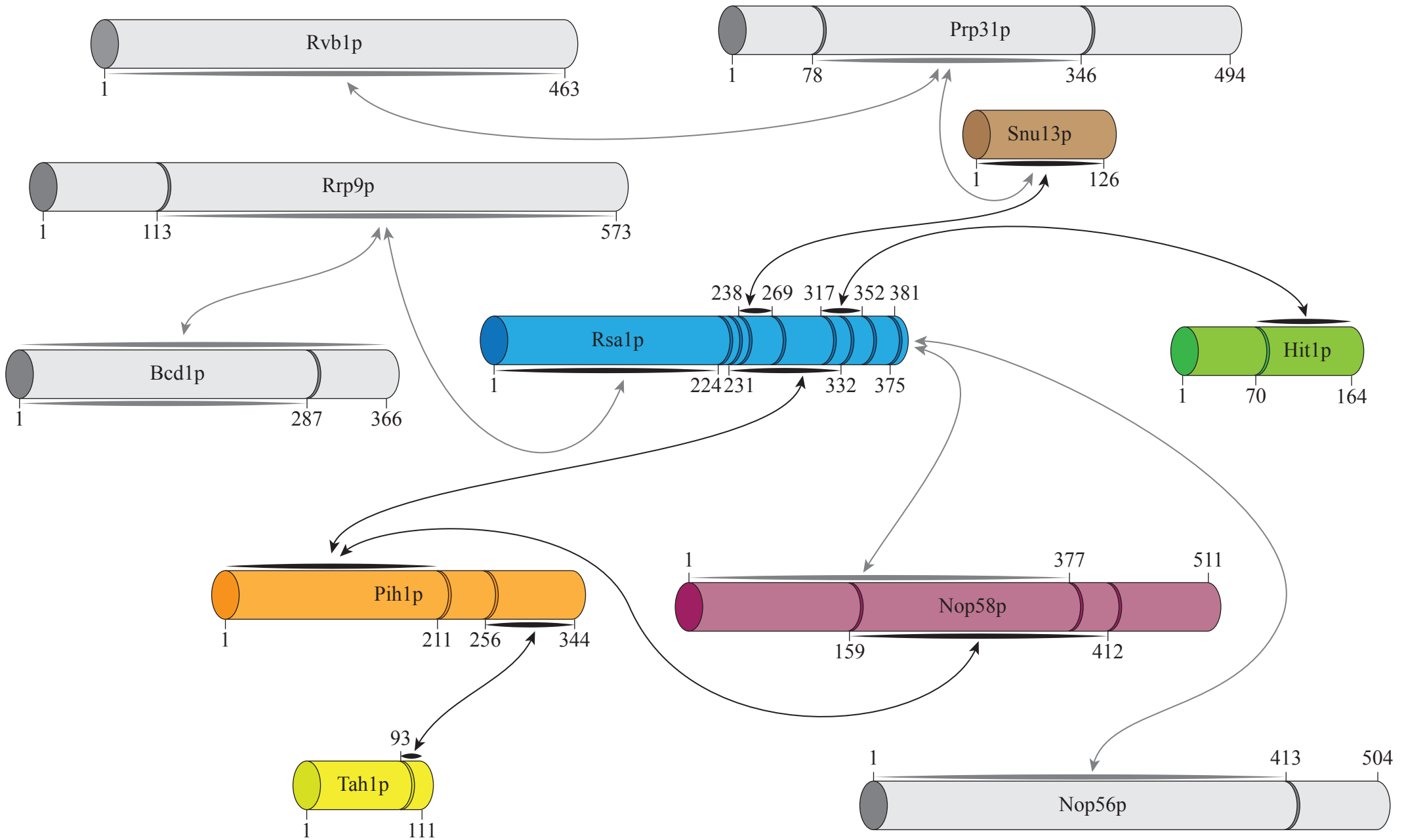


Figure 102. **Bilan des interactions mises en évidence par l'ensemble des résultats obtenu lors de la thèse.** Les protéines sont représentées sous forme de cylindre dont la longueur est proportionnelle à leur taille en acides aminés. Les bornes des différents domaines des protéines sont indiquées par les numéros des résidus. Les protéines en couleur correspondent aux protéines qui ont fait l'objet d'une étude plus appronfondie lors de la thèse.

2 La protéolyse ménagée, définition des domaines

Parfois, malgré l'obtention de complexes stables et purs à haute concentration, il n'a pas été possible d'obtenir de cristaux pour poursuivre l'étude par radiocristallographie. C'est le cas de l'hétérotrimère Snu13p/Rsa1p (231-375)/Hit1p, qui malgré des essais de cristallisation jusqu'à une concentration de 100 mg/ml n'a pas pu être cristallisé. La flexibilité intrinsèque des composants, notamment de la protéine Rsa1p et dans une moindre mesure la protéine Hit1p était sans doute un frein à la nucléation des cristaux. Pour pallier à ce problème, nous avons mis en place, en collaboration avec V. Pena du laboratoire du professeur R. Lürhmann au Max Planck Institut à Göttingen, une stratégie de protéolyse ménagée de ce complexe pour en dégager les domaines structurés. Les résultats obtenus nous ont permis de définir un sous-complexe comportant les domaines (317-375) de Rsa1p et (44-164) de Hit1p. Malheureusement, la protéolyse ménagée ne nous a pas permis de maintenir le troisième partenaire, à savoir Snu13p, au sein des complexes. Ce complexe à deux partenaires avait l'avantage d'être compatible avec une étude par R.M.N et les premiers résultats obtenus montrent une structuration des deux partenaires au sein du complexe. Néanmoins, quelques résidus étaient encore trop flexibles pour que l'attribution des signaux R.M.N. soit complète. L'emploi d'analyses par spectrométrie de masse pour l'étude des produits d'hydrolyse des fragments nous a conduit à éliminer l'extrémité N-terminale de la protéine Hit1p (44-164) et C-terminale de la protéine Rsa1p (317-375) dans les protéines recombinantes. Le nouveau complexe Rsa1p (317-352)/Hit1p (70-164) ainsi produit est actuellement en cours d'analyse et l'attribution des résidus de la chaîne principale a été entièrement réalisée par M. Quinternet au laboratoire. Nous espérons donc déterminer très rapidement la détermination de la structure tridimensionnelle du complexe.

3 La R.M.N.

Au cours de ma thèse, malgré nos efforts dans l'optimisation des conditions de purification par des analyses D.L.S. des produits aucun cristal de protéines ou de complexes protéiques utilisables en radiocristallographie n'a pu être obtenu. La R.M.N. a donc apporté une alternative de choix pour nos études structurales. Même si la masse moléculaire des échantillons qui peuvent être étudiés par RMN est limitée, cette approche nous a permis d'étudier la plupart des protéines et complexes que nous avions réussis à produire sous forme soluble. La résolution de la structure tridimensionnelle de la protéine Tah1p appartenant au

complexe R2TP seule et en complexe avec l'extrémité C-terminale de la chaperonne Hsp90 en est un bel exemple. Cette étude a été menée en collaboration avec l'équipe de F. Allain à Zurich. La protéine Tah1p adopte un repliement de type domaine TPR mais ce dernier est atypique car il ne possède que cinq hélices α. De plus, le domaine C-terminal de Tah1p n'adopte aucune conformation particulière à l'état libre mais est nécessaire et suffisant pour que Tah1p s'associe à la protéine Pih1p appartenant également au complexe R2TP. Notre analyse du complexe formé entre Tah1p et le peptide révèle un mode particulier de fixation pour ce dernier. En effet, tout comme pour le complexe formé entre la chaperonne Hsp90 et la protéine HOP (Scheufler *et al.*, 2000), le motif NH_2-EVD-COOH du peptide adopte une conformation étendue au sein du domaine TPR de la protéine et l'interaction fait intervenir les résidus de la *« carboxylate clamp »* du domaine TPR de Tah1p. Cependant, à la différence de ce qui a été observé pour le complexe Hsp90-Hop, dans le complexe Hsp90-Tah1p les acides aminés en amont du motif conservé EEVD de Hsp90 se positionnent sur l'hélice α5 de capping du domaine TPR de Tah1p.

Actuellement au laboratoire, deux complexes protéiques, Rsa1p (317-352)/Hit1p (70-164) et Snu13p/Rsa1p (231-269), sont en cours d'analyses par R.M.N. Des données structurales sur ces deux interactions pourront donc être bientôt disponibles.

4 Vers une étude structurale par SAXS

Comme cela a été précisé auparavant, à l'heure actuelle aucun cristal exploitable de protéines ou de complexes protéiques n'a pu être obtenu. Lors de l'assemblage des snoRNP C/D, la protéine Rsa1p joue le rôle de plateforme pour les protéines constitutives de la particule C/D. Elle possède donc la capacité à s'associer à différents partenaires protéiques et nous avons pu montrer que cette interaction était mise en place grâce à différentes régions de Rsa1p. Cette protéine semble contenir un enchaînement de domaines qui sont parfaitement repliés uniquement lorsqu'ils sont en interaction avec leurs partenaires respectifs. Les domaines étant reliés par des parties plus flexibles, permettant d'accommoder la fixation de différentes protéines sur Rsa1p au cours de la biogenèse des snoRNP à boîtes C/D. Si nous incluons la possibilité d'un repliement induit par l'interaction avec son partenaire, l'*« induce fiting »*, l'étude structurale de la protéine Rsa1p représente un vrai défi. Une possibilité est donc d'étudier individuellement chaque domaine en complexe avec leurs partenaires respectifs et d'utiliser ensuite l'ensemble de ces données pour reconstruire la protéine entière.

L'utilisation de l'approche S.A.X.S. devrait pouvoir nous permettre de définir l'enveloppe moléculaire de l'hétérotrimère, ce qui nous permettrait ensuite de replacer les structures atomiques des deux « sous-complexes » à l'intérieur de cette enveloppe.

5 Vers l'incorporation des ARN dans l'étude des complexes

La composante ARN est très importante dans la formation des particules C/D matures. Il est donc logique de considérer le snoRNA comme un partenaire à part entière dans l'étude structurale des complexes. Néanmoins, l'étude structurale des acides nucléiques reste, encore à ce jour, d'une difficulté supérieure à celle des protéines, notamment dans leur production et leur purification car les critères de quantité, d'homogénéité et de pureté sont élevés pour ce genre d'étude. La synthèse chimique peut permettre d'obtenir des échantillons de bonne qualité mais reste coûteuse et la taille des ARN produits reste limitée. Des systèmes de transcription *in vitro* basés sur l'utilisation de la T7 RNA Polymérase produite de façon recombinante ont été développés et testés au laboratoire. Cependant l'optimisation de telles transcriptions peut s'avérer assez coûteuse et fastidieuse et les échantillons obtenus ne sont pas toujours de qualité suffisante car l'enzyme peut marquer des arrêts de synthèse prématurés ou rajouter un ou deux nucléotides à l'extrémité 3', donnant lieu à des échantillons non homogènes incompatibles avec une étude par radiocristallographie. L'étude structurale par RMN est moins sensible à ce manque relatif d'homogénéité des échantillons, mais le problème de recouvrement des signaux des acides nucléiques implique l'utilisation de très hauts champs magnétiques, sans toutefois être une solution absolue. Tout ceci génère donc une forte limitation de la taille des cibles ARN qui peuvent être étudiés. La solution du marquage isotopique des acides nucléiques permettrait l'étude de molécules de taille plus importantes, comme cela est le cas avec les protéines, mais sa mise en place, y compris par synthèse chimique reste délicate et coûteuse. Il faudrait donc imaginer un système calqué sur celui des protéines qui permette une synthèse *in vivo* des ARN. Ainsi nous limiterions les problèmes de quantité d'ARN avec la possibilité d'utiliser des milieux de culture comportant les isotopes à incorporer dans les ARN. Récemment, l'utilisation d'ARNt *« scaffold »* pour protéger de la dégradation les ARN exprimés en système hétérologue (*E. coli*) a été proposé (Ponchon *et al.*, 2009; Ponchon & Dardel, 2011). Ce système repose sur la forte expression et stabilité des ARNt chez *E. coli*. L'ARN d'intérêt est inséré au sein de la boucle anticodon de l'ARNt, région présentant naturellement une certaine flexibilité. Il est même possible d'étiqueter l'ARN avec un aptamère capable de se lier à une matrice Séphadex par exemple,

ce qui facilite les étapes de purification. Mais la séparation de l'ARN d'intérêt de l'ARNt *« scaffold »* par l'utilisation de RNAses reste encore une étape délicate à maîtriser. Une dernière solution serait de concevoir un système de co-expression ARN-protéine chez *E. coli* où l'association stable des protéines co-exprimées sur leur cible ARN protègerait ce dernier des dégradations au sein de la bactérie. La mise au point d'un tel système représente un défi de taille mais un succès dans ce domaine ouvrirait la voie à des développements considérables pour les études structurales incluant de l'ARN ou des complexes ARN-protéines.

MATERIEL
ET
METHODES

De nombreux protocoles ont fait l'objet d'optimisation au cours de la thèse et sont donc détaillés dans la partie Résultats & Discussion du manuscrit. La grande majorité du matériel a été décrite dans la partie Matériel et Méthodes des différents articles.

1 Matériel

1.1 Les souches bactériennes

1.1.1 Pour l'amplification et l'extraction d'ADN plasmidique

***Escherichia coli* DH5α** : *sup*E44, Δ*lac*U169 (Φ80 *lacZ*ΔM15), *hsd*R17, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *rel*A1

1.1.2 Pour la surexpression de protéines recombinantes

***Escherichia coli* BL21 (DE3)** : F– *omp*T, *hsd*SB (rB–, mB–), *gal dcm* (DE3)

1.2 Les plasmides bactériens

1.2.1 Pour le clonage

Les clonages de fragments d'ADN amplifiés par la technique PCR ont été réalisés dans les plasmides pCR®2.1 fournis avec le kit « TA Cloning® Kit» et pCR®-Blunt fournis avec le kit « ZeroBlunt® TOPO® PCR Cloning Kits ».

Vecteur pCR®2.1 (Invitrogen™). Le plasmide est commercialisé sous forme linéarisé et porte un résidu thymine protubérant sur chaque extrémités 3'. Il contient des gènes de résistance à l'ampicilline et à la kanamycine. Il possède également le gène *lacZ* codant la β-galactosidase placé sous le contrôle du promoteur P*lac* et interrompu par la casette de clonage.

Vecteur pCR®-Blunt (Invitrogen™). Le plasmide est commercialisé sous forme linéarisé et présente des extrémités franches. Il contient des gènes de résistance à la kanamycine et à la Zéocine. Il possède également le gène *lacZ* codant la β-galactosidase placé sous le contrôle du promoteur P*lac* et interrompu par la casette de clonage.

1.2.2 Pour la production de protéines

Plasmide pRARE2 (Novagen®). Il permet de complémenter la souche bactérienne BL21 (DE3) d'*E. coli* par l'expression de sept ARNt eucaryotes rares pour cet organisme procaryote. Ce plasmide code pour les ARNt reconnaissant les codons AUA, AGG, AGA, CUA, CCC, CGG et GGA et porte également les gènes *argN5*, *metT*, *thrT*, *glyT* et *thrU*.

Vecteur PKHS (Eckert *et al.*, 2010). Ce vecteur contenant la séquence codante de la protéine Tah1p nous a aimablement été fourni par l'équipe de S. Morera (Gif-sur-Yvette). Il permet l'expression, par induction à l'IPTG, de protéines fusionnées à une étiquette 6xHis au niveau de leur extrémité C-terminale. Il a été utilisé pour la production de la protéine Tah1p pour les analyses par R.M.N. homo- et hétéronucléaire (publication n°3).

Vecteurs pnEA-tH, pnEA-3CH, pnCS et pnYC (Diebold *et al.*, 2011). Ces plasmides nous ont aimablement été fournis par C. Romier. Ils font l'objet d'une description détaillée dans les paragraphes 1.1.1 et 1.1.2 du chapitre 1 des Résultats & Discussion de ce manuscrit. Ils ont été utilisés pour les essais de co-expression à haut débit sur plateforme robotisée TECAN mais aussi pour la production des échantillons pour les essais de cristallogenèse et les analyses par R.M.N. Ils permettent l'expression, par induction à l'IPTG, de protéines fusionnées à une étiquette 6xHis au niveau de leur extrémité N-terminale (pour le vecteur pnEA) ou de façon native (pour les vecteurs pnCS et pnYC).

1.2.3 Pour la synthèse d'ARN *in vitro*

Vecteur pUC18 (Invitrogen™). Ce vecteur a été utilisé pour cloner certaines séquences codant des snoRNA (notamment le snoRNA U3 Δ2,3,4) et pour générer, par PCR, des matrices ADN double-brin pour la transcription *in vitro*. Il contient un gène de résistance à l'ampicilline et l'origine de réplication du vecteur pMB1. Il possède également le gène *lacZ* codant la β-galactosidase placé sous le contrôle du promoteur P*lac* et interrompu par la casette de clonage.

1.3 Les milieux de culture

Les différents milieux de culture sont autoclavés à 121°C, sous 1 bar de pression, pendant 20 minutes. Pour la préparation des milieux solides, de l'agar est additionné à raison de 17 g/l avant le passage à l'autoclave.

Les antibiotiques sont ajoutés après l'étape de stérilisation, à une concentration finale de 100 mg/l pour l'ampicilline, 50 mg/l pour la kanamycine, 30mg/l pour la spectinomycine et 25 mg/l pour le chloramphénicol.

- **Milieu riche LB** : (Luria Bertani) : tryptone (peptone pancréatique animale/peptone trypsique de caséine) 10 g/l ; extrait de levure 5 g/l ; NaCl 5 g/l. Le pH est ajusté à 7,5 avec une solution de NaOH 10 N.
- **Milieu riche 2TY** : tryptone (peptone pancréatique animale/peptone trypsique de caséine) 20 g/l ; extrait de levure 10 g/l ; NaCl 5 g/l. Le pH est ajusté à 7,5 avec une solution de NaOH 10 N.
- **Milieu riche Graffinity** : milieu 2TY ; glucose 0,25 % ; Hepes 10 mM, pH 7 ; Lactose 0,3 % ; IPTG 0,25 mM.

  La culture débute dans le milieu 2TY ; Glucose 0,5 % dans un volume correspondant à la moitié du volume final. Lorsque la DO_{600nm} atteint 1,2-1,5, un même volume de milieu 2TY ; Hépes 20 mM, pH 7 ; IPTG 0,5 mM ; Lactose 0,6 % est ajouté au premier.
- **Milieu minimum MnRMN** : (adapté du milieu N5052) : Na_2HPO_4 50 mM ; KH_2PO_4 50 mM ; Na_2SO_4 5 mM, pH 7,4 ; $MgSO_4$ 1 mM ; $^{15}NH_4Cl$ 2 g/l ; Glucose ^{13}C 4 g/l ; Biotine 1 mg/l ; Acide folique 1 mg/l ; Nicotinamide 1 mg/l ; Riboflavine (vitamine B2) 0,1 mg/l ; Thiamine 5 mg/l ; Métaux Mix I 200 µl/l ; Métaux Mix II 100 µl/l ; Antibiotiques en fonction.
- **Métaux Mix I** : $FeCl_3$ 50 mM.
- **Métaux Mix II** : $CaCl_2$ 20 mM ; $MnCl_2$ 10 mM ; $ZnSO_4$ 10 mM ; $CoCl_2$ 2 mM ; $CuCl_2$ 2 mM ; $NiCl_2$ 2 mM ; Na_2MoO_4 2 mM ; Na_2SeO_3 2 mM ; H_3BO_3 2 mM.
- **Milieu SOC** : Tryptone 2 % (p/v) ; extrait de levure 0,5 % (p/v) ; NaCl 10 mM ; KCl 2,5 mM ; MgCl2 10 mM ; glucose 20 mM.

1.4 Les tampons et solutions d'usage courant

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

Bleu de dépôt pour gel de protéines dénaturant : SDS 3 % (v/v) ; glycérol 10 % (v/v) ; bleu de bromophénol 0,1 mg/ml ; DTT 100 mM ; Tris-HCl 80 mM, pH 6,8.

Bleu de formamide : EDTA 20 mM ; bleu de bromophénol 0,05 %; bleu de xylène cyanol 0,05 % dans de la formamide désionisée.

Marqueur de taille de protéines : (GE Heathcare) : Phosphorylase B de muscle de lapin (97 kD), Albumine de sérum bovin (66 kD,), Ovalbumine d'œufs de poulet (45 kD), Anhydrase carbonique d'érythrocyte bovin (30 kD), Inhibiteur de Trypsine (20,1 kD), α-lactabumine (14,4 kD).

Marqueur de taille de protéine : (Bio-Rad, Precision Plus Protein™ All Blue Standards) : protéines recombinantes de 10, 15, 20, 25, 37, 50, 75, 100, 150 et 250 kD.

Solution de polyacrylamide pour gel non dénaturant : Acrylamide/Bis-Acrylamide (38/2) 40 % (v/v) ; TBE 1X.

Solution de polyacrylamide pour gel dénaturant : Acrylamide/Bis-Acrylamide (23,75/1,25) 25 % (v/v) ; TBE 1X ; urée 8M.

Solution I de coloration des gels SDS-PAGE : Ethanol 50 % ; Acide Acétique 10 %.

Solution II de coloration des gels SDS-PAGE : Ethanol 5 % ; Acide Acétique 7,5 %.

Solution stock de coloration des gels SDS-PAGE : Bleu de Coomassie G250 0,25 % ; Ethanol 96 %.

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

Tampon de migration SDS-PAGE : glycine 14,25 g/l ; Tris 3 g/l ; SDS 1 g/l (v/v).

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

Tampon de transcription θ : Tris-HCl 40 mM, pH 8, DTT 5 mM, Spermidine 1 mM, Triton 0,01 %, PEG 8000 80 mg/ml, $MgCL_2$ 38 mM

2 Méthodes

2.1 Techniques de génie génétique

2.1.1 Clonage

Les fragments d'ADN d'intérêt sont amplifiés par PCR en utilisant l'ADN polymérase Phusion® (Finnzymes) et un couple d'oligodéoxynucléotides amorces apportant en 5' et en 3' les sites de restriction compatibles avec le vecteur de destination selon le protocole suivant :

Matrice	50 ng
dNTP (1,25 mM)	8 µl
Oligodéoxynucléotide 5' (100 µM)	0,5 µl
Oligodéoxynucléotide 3' (100 µM)	0,5 µl
Tampon HF Phusion® 5X	10 µl
Phusion® DNA polymérase (2 U/µl)	0,2 µl

H_2O qsp 50 µl

Conditions de la PCR

Température (°C)	Temps	Nombre de cycles	Fonction
98	1 min	1	Dénaturation
98	10 sec	35	Dénaturation
55	20 sec		Hybridation
72	15 sec/kb		Elongation
72	5 min	1	Elongation
4	∞	1	Stockage

A l'issue de la PCR, deux possibilités de clonages dans les vecteurs d'expression peuvent être envisagée.

2.1.1.1 Clonage direct dans les vecteurs d'expressions bactérien

Les fragments d'ADN double-brin obtenus par PCR sont traités par l'endonucléase *DpnI* (MBI-Fermentas), purifiés sur colonnes NucleoSpin® (Macherey-Nagel) et quantifiés par mesure d'absorbance à 260 nm (NanoDrop, Thermo Fisher Scientific). 2 µg d'ADN sont ensuite digérés par double-restriction enzymatique, dans les conditions décrites par le fournisseur (MBI-Fermentas). Les produits de digestion sont à nouveau purifiés sur colonnes NucleoSpin®, dont la limite d'exclusion permet d'éliminer les produits de clivage provenant des deux extrémités du fragment d'ADN. Le vecteur de destination est préparé de la même façon et subit une étape supplémentaire de déphosphorylation de ses extrémités 5' par action de la phosphatase alcaline CIAP (MBI Fermentas). La ligation par la T4-DNA ligase (MBI-Fermentas) est ensuite réalisée à partir de 50 ng de vecteur et d'un excès molaire de trois fois de fragment. Après transformation de cellules *E. coli* DH5α super-compétentes et sélection des clones recombinants sur milieu sélectif, les plasmides sont purifiés en utilisant le protocole de mini-préparation de Macherey-Nagel que nous avons adapté sur la plate-forme robotisée TECAN Freedom EVO 200 et ensuite analysés par séquençage. Une fois la séquence vérifiée une préparation d'ADN à plus grande échelle est effectuée pour être stockée à 20°C.

2.1.1.2 Sous clonage par le pCR®-Blunt

Les ADN double-brin obtenus sont purifiés par migration sur gel agarose et extraction par NucleoSpin®. Les fragments purifiés sont ensuite insérés dans le vecteur pCR®-Blunt (ou pCR®2.1 si l'ADN polymérase utilisée lors de la PCR possédait une activité « 5' adénosine extendase ») par la T4-DNA ligase dans les conditions décrites par le fournisseur (Invitrogen®). Après transformation de cellules *E. coli* DH5α super-compétentes et sélection

des clones recombinants sur milieu sélectif, les plasmides sont purifiés en utilisant le protocole de mini-préparation de Macherey-Nagel que nous avons adapté sur la plate-forme robotisée TECAN Freedom EVO 200 et ensuite analysés par séquençage. Une fois la séquence vérifiée une préparation d'ADN à plus grande échelle est effectuée pour être stockée à 20°C. Le cas échéant, un ou plusieurs cycles de mutagenèse dirigée pourra être réalisé sur la construction afin d'éliminer des sites de restriction internes à la séquence codante pouvant être gênants pour les étapes suivantes de clonage. Le plasmide obtenu est hydrolysé par une double-restriction enzymatique dans les conditions décrites par le fournisseur (MBI-Fermentas). Les produits de digestion sont séparés sur gel d'agarose et les fragments d'intérêt extraits et purifiés par NucleoSpin® puis quantifiés par mesure d'absorbance à 260 nm (NanoDrop, Thermo Fisher Scientific). La ligation par la T4-DNA ligase (MBI-Fermentas) est ensuite réalisée à partir de 50 ng de vecteur de destination et d'un excès molaire de trois fois de fragment. Après transformation de cellules *E. coli* DH5α super-compétentes et sélection des clones recombinants sur milieu sélectif, les plasmides sont purifiés en utilisant le protocole de mini-préparation de Macherey-Nagel que nous avons adapté sur la plate-forme robotisée TECAN Freedom EVO 200 et ensuite analysés par séquençage.

2.1.2 Mutagenèse dirigée par PCR

La mutagenèse dirigée par PCR est réalisée avec l'ADN polymérase Phusion® (Finnzymes) selon le protocole suivant :

Matrice	50 ng
dNTP (1,25 mM)	3,2 µl
Oligodéoxynucléotide 5' (10 µM)	0,3 µl
Oligodéoxynucléotide 3' (10 µM)	0,3 µl
Tampon HF Phusion® 5X	4 µl
Phusion® DNA polymérase (2 U/µl)	0,5 µl
H_2O qsp	20 µl

Conditions de la mutagenèse par PCR

Température (°C)	**Temps**	**Nombre de cycles**	**Fonction**
98	5 min	1	Dénaturation
98	30 sec	30	Dénaturation
55	30 sec		Hybridation
72	15 à 30 sec/kb		Elongation
72	7 min	1	Elongation
4	∞	1	Stockage

A l'issue des cycles de PCR, 15 µl du mélange réactionnel sont traités par l'endonucléase *DpnI* (MBI Fermentas) dans les conditions recommandées par le fournisseur. Après transformation de cellules *E. coli* DH5α super-compétentes et sélection des clones recombinants sur milieu sélectif, les plasmides sont purifiés en utilisant le protocole de mini-préparation de Macherey-Nagel que nous avons adapté sur la plate-forme robotisée TECAN Freedom EVO 200 et ensuite analysés par séquençage.

2.2 Ingénierie des protéines

Les conditions décrites ci-dessous correspondent aux conditions standards, les plus couramment utilisées pour nos protéines d'intérêt. Le protocole expérimental seul sera décrit car les détails concernant les compositions des tampons, le matériel utilisé, … sont décrits dans la partie Résultats & Discussion du manuscrit.

2.2.1 Surproduction de protéines chez *E. coli*

Les cellules BL21-DE3 (complémentées ou non par le plasmide pRAR2) transformées par le(s) plasmide(s) d'intérêt (comme pnEA-3CH + pnCS + pnYC dans les cribles de co-expression) sont utilisées pour ensemencer les milieux de cultures contenant les antibiotiques adéquats. Trois protocoles de culture et d'induction peuvent être utilisés en fonction de la finalité des échantillons obtenus.

2.2.1.1 Pour les cribles de co-expression (échelle analytique)

Les cultures sont réalisées dans des Deep-Weel 24 puits (Uniplate®, Whatman®) avec l'ensemencement de 1,3 ml de milieu Graffinity (2TY ; Glucose 0,5 %) par une colonie transformée et mise sous agitation à 37°C. Lorsque la $D.O._{600nm}$ atteint la valeur de 1,2-1,5, la surproduction des protéines est induite par l'ajout de 1,3 ml de milieu (2TY ; Hépes 20 mM, pH 7 ; IPTG 0,5 mM ; Lactose 0,6 %). La culture se poursuit alors à 20°C O/N avant d'être divisée en deux plaques 24 puits lesquelles sont ensuite centrifugées 10 min à 3000 rpm.

2.2.1.2 Pour des cribles de cristallogenèse ou d'analyse par R.M.N. homonucléaire (échelle préparative)

Dans ce cas, 1 litre de milieu riche (LB ou 2TY + antibiotique(s)) est ensemencé à 0,1 unité A_{600nm}/ml à partir d'une pré-culture à saturation. Les cultures sont placées sous agitation à une température de 37 ou 30°C selon le cas. Lorsque la $D.O._{600nm}$ atteint la valeur de 0,6-

0,8, la surproduction des protéines est induite par ajout d'IPTG à la concentration finale de 1 mM. Les cultures sont alors incubées sous agitation à 20°C O/N.

2.2.1.3 Pour des analyses par R.M.N. hétéronucléaire (échelle préparative)

L'enrichissement isotopique de l'échantillon protéique pour les analyses R.M.N. hétéronucléaire nécessite l'utilisation du milieu minimum MnRMN (adapté du milieu N5052). Une première pré-culture de 500 µl de milieu riche (LB ou 2TY + antibiotique(s)) est tout d'abord ensemencée par une colonie transformée puis mise sous agitation à 37°C pendant 6 h. Elle servira à ensemencer à son tour 25 ml de milieu MnRMN (+ antibiotique(s)) qui seront incubés sous agitation à 37°C O/N. Au sein de la cuve du fermenteur de 5 L (BioStat® Plus, Sartorius), un litre de milieu MnRMN est inoculé à 0,1 unités A_{600nm}/ml à partir de la deuxième pré-culture et reste sous agitation à 37°C en maintenant l'oxygénation constante (PO_2 fixée à 50 %) et un pH constant (fixé à 7,5). Lorsque la $D.O._{600nm}$ atteint la valeur de 0,6-0,8, la surproduction des protéines est induite par ajout d'IPTG à la concentration de 1 mM finale. Les cultures sont alors incubées sous agitation à 20°C O/N en maintenant l'oxygénation constante (PO_2 fixée à 50 %) et un pH constant (fixé à 7,5).

2.2.2 Purification de protéines

2.2.2.1 Lors des cribles par co-expression

Les culots cellulaires sont repris dans 1,5 ml de tampon Tris (Tris-HCl 10 mM, pH 8,8) avec deux concentrations salines différentes (NaCl 50 ou 400 mM). Les cellules sont ensuite lysées à 4°C, par traitement aux ultra-sons (amplitude 20 %, temps 16 s, pulse 2 s ON/2 s OFF). Le lysat est centrifugé 20 min à 4000 rpm à 4°C. Le surnageant de sonication est alors incubé 1 h à 4°C sous agitation douce dans un Deep-Weel 96 puits (ABgene®) avec 50 µl de résine Sépharose (TALON® SuperFlow™, Clontech) préalablement équilibrée au tampon Tris. Le surnageant est éliminé après centrifugation 10 min à 1000 rpm. La résine est lavée deux fois par ajout de 500 µl de tampon (Tris-HCl 10 mM, pH 8,8 ; NaCl 50 ou 400 mM) suivi d'une centrifugation 5 min à 1000 rpm à 4°C. Le surnageant est alors éliminé et 30 µl de bleu de dépôt pour gel dénaturant de protéines sont ajoutés et incubés 10 min avec la résine à température ambiante. Le bleu est alors analysé par SDS-PAGE 15 %.

2.2.2.2 Lors des préparations pour des études structurales

Les culots cellulaires sont repris dans le tampon adéquat à raison de 100 unités A_{600nm}/ml. La lyse cellulaire est effectuée à 4°C par traitement aux ultra-sons à l'aide d'un sonicateur Sonifer 250 (Branson) (« Duty-cycle » 50 % ; « Outpout » 5-6). Le lysat est centrifugé 1 h à 4°C à 20000 rpm. Le surnageant de sonication est incubé avec du P.E.I. à une concentration de 0,0125 % par agitation douce à 4°C pendant 30 min. Le mélange est centrifugé à 4°C pendant 30 min à 20000 rpm. Le surnageant est alors incubé 45 min à 4°C sous agitation douce avec de la résine Sépharose (TALON® SuperFlow™, Clontech) préalablement équilibrée au tampon de casse (300 µl pour l'équivalent de 1 litre de culture initiale). Cette étape est divisée en trois incubations avec 100 µl de résine. La résine est lavée par incubation à 4°C avec 15 ml de tampon adéquat sous agitation douce. Si la démarche expérimentale nécessite de conserver l'étiquette poly-Histidines, la résine sera incubée pendant 10 min à 4°C sous agitation douce avec un tampon contenant de l'Imidazole 200 mM. En revanche, si l'étiquette doit être éliminée, la résine est tout d'abord équilibrée au tampon d'élution (Tris-HCl / Hépes-KOH 10 mM, NaCl 300 mM ; β-mercaptoéthanol 5 mM) puis incubée pendant 12 h à 4°C sous agitation douce avec 500 µl de tampon d'élution additionné de protéase GST-PreScission (produite au laboratoire) à raison de 1 unité pour 100 µg de protéine retenue sur résine (estimation sur gel). L'éluât (issu du traitement à l'Imidazole ou de la protéase) peut alors être purifié par filtration sur gel sur colonnes Superdex™ 75 ou 200 (GE Healthcare).

2.2.3 Protéolyse ménagée

Pour la préparation des enzymes, le lyophilisat commercial (de 20 mg à 25 mg, voir tableau ci-dessous) est repris dans 500 µl d'eau ultra-pure. Cette dilution correspond au stock 1:10. L'enzyme est alors aliquotée par tube de 6 µl et congelée rapidement à l'azote liquide puis stocké à -80°C.

Enzyme	Lyophilisat	Volume de reprise du lyophilisat	Concentration du stock en g/l (=1:10)
Subtilisine	25 mg	500 µl	50
Elastase	20 µg	500 µl	0,04
Thermolysine	25 mg	500 µl	50
Carboxypeptidase Y	25 µg	500 µl	0,05
Glu-C	25 µg	500 µl	0,05
Chymotrypsine	25 µg	500 µl	0,05

Pour chaque essai de protéolyse ménagée, les enzymes sont préparées dans la glace aux dilutions 1:10 (stock), 1:100 et 1:1000 afin de déterminer la concentration adéquate en protéase. Chaque mélange réactionnel est constitué de 15 µl de protéine (~0,6 µg/µl) additionnés de 4 µl de protéase (un mélange réactionnel par dilution de protéase) est incubé 30 min à température ambiante. La réaction est stoppée par l'ajout de 1 µl de PMSF (100 mM). Les produits de protéolyse sont analysés par SDS-PAGE. Le choix de la dilution d'enzyme est défini sur la base du profil de migration obtenu. La même expérience est réalisée en fixant cette fois la dilution de protéase est en faisant varier le temps d'incubation (cinétique enzymatique sur 5, 10, 20, 30, 60, 120, 300 min).

2.2.4 Analyse par D.L.S.

Les échantillons protéiques sont dialysés dans 40 ml de tampon (pH et concentration saline différentes) par agitation douce à 4°C par l'utilisation de bouton de dialyse de 25 µl (25 µl Dialysis Buttons™, Hampton Research) pendant 12 h. Les échantillons sont récupérés et centrifugés 30 min à 4°C à 13200 rpm, 15 µl sont alors analysés par un appareil D.L.S. (Zetasizer Nano-S, Malvern Instruments) dans une microcuve (QS 3mm, 12 µl, Hellma). Après une équilibration de 1 min à la température choisie (5, 15 ou 20 °C), 3 mesures sont effectuées successivement.

2.2.5 Analyse par S.L.S.

Les échantillons protéiques à une concentration de 100 µM dans leur tampon respectif sont injectés sur une colonne d'exclusion de taille (Shodex KW 8025) couplée à un appareil S.L.S. (TDA 302, Viscotek, Malvern Instruments) à température ambiante.

2.3 Synthèse *in vitro* des ARN

La transcription d'ARN *in vitro* est réalisée à l'aide d'une ARN T7 polymérase produite au laboratoire. Dans un premier temps, une matrice d'ADN double-brin linéaire, portant un promoteur T7 en amont de la séquence codant l'ARN, est générée par PCR, à l'aide de l'ADN polymérase GoTaq (Promega). La réaction de transcription est réalisée pendant 12 h à 37°C, à partir de 23,2 µl de produit PCR, dans un tampon de transcription (Tris-HCl 40 mM, pH 8, DTT 5 mM, Spermidine 1 mM, Triton 0,01 %, PEG 8000 80 mg/ml, $MgCl_2$ 38 mM) additionnés de 10 µl d'ARN polymérase T7 et 25 mM de chaque NTP. A l'issue de cette réaction, l'ADN matrice est dégradé par ajout de 7,5 unités de DNase I et incubation pendant 1 h à 37°C. Les ARN sont ensuite dénaturés 2 min à 96°C en présence de

bleu de formamide, puis purifiés sur gel de polyacrylamide dénaturant. Après visualisation des ARN par ombrage aux UV, les bandes de gel d'intérêt sont découpées et les ARN sont élués par incubation sous agitation, pendant 30 min à 37°C, dans 300 µl de tampon d'élution (Tris-HCl 10 mM, pH 7,5 ; NaCl 300 mM, EDTA 1 mM ; SDS 1 %). Les ARN élués sont ensuite isolés par précipitation alcoolique (sans ajout de sels), repris dans 20 µl d'eau ultra-pure, puis dosés par mesure de l'absorbance à 260 nm (NanoDrop, Thermo Fisher Scientific).

REFERENCES BIBLIOGRAPHIQUES

Références bibliographiques

Aittaleb M, Rashid R, Chen Q, Palmer JR, Daniels CJ, Li H (2003) Structure and function of archaeal box C/D sRNP core proteins. *Nat Struct Biol* **10:** 256-263

Aittaleb M, Visone T, Fenley MO, Li H (2004) Structural and thermodynamic evidence for a stabilizing role of Nop5p in S-adenosyl-L-methionine binding to fibrillarin. *J Biol Chem* **279:** 41822-41829

Ali MM, Roe SM, Vaughan CK, Meyer P, Panaretou B, Piper PW, Prodromou C, Pearl LH (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440:** 1013-1017

Allmang C, Carbon P, Krol A (2002) The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. *RNA* **8:** 1308-1318

Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J* **18:** 5399-5410

Amiri KA (1994) Fibrillarin-like proteins occur in the domain Archaea. *J Bacteriol* **176:** 2124-2127

Angrisani A, Turano M, Paparo L, Di Mauro C, Furia M (2011) A new human dyskerin isoform with cytoplasmic localization. *Biochim Biophys Acta* **1810:** 1361-1368

Antal M, Boros E, Solymosy F, Kiss T (2002) Analysis of the structure of human telomerase RNA in vivo. *Nucleic Acids Res* **30:** 912-920

Aravind L, Koonin EV (1999) Novel predicted RNA-binding domains associated with the translation machinery. *J Mol Evol* **48:** 291-302

Arnez JG, Steitz TA (1994) Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry* **33:** 7560-7567

Auffinger P, Westhof E (1998) Effects of Pseudouridylation on tRNA Hybridation and Dynamics: a Theoretical Approach. In *Modification and Editing of RNA*, Grosjean H, Benne R (eds), ASM Press, Washington, DC. edn, 6, pp 103-112.

Bachellerie JP, Cavaillé J, Hüttenhofer A (2002) The expanding snoRNA world. *Biochimie* **84:** 775-790

Bagni C, Lapeyre B (1998) Gar1p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element. *J Biol Chem* **273:** 10868-10873

Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, Terns MP (2005) RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev* **19:** 1238-1248

Bakin A, Lane BG, Ofengand J (1994) Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry* **33:** 13475-13483

Bakin A, Ofengand J (1993) Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* **32:** 9754-9762

Balakin AG, Smith L, Fournier MJ (1996) The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86:** 823-834

Ballarino M, Morlando M, Pagano F, Fatica A, Bozzoni I (2005) The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in Saccharomyces cerevisiae. *Mol Cell Biol* **25:** 5396-5403

Bally M, Hughes J, Cesareni G (1988) SnR30: a new, essential small nuclear RNA from Saccharomyces cerevisiae. *Nucleic Acids Res* **16:** 5291-5303

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289:** 905-920

Bardoni B, Schenck A, Mandel JL (1999) A novel RNA-binding nuclear protein that interacts with the fragile X mental retardation (FMR1) protein. *Hum Mol Genet* **8:** 2557-2566

Baserga SJ, Agentis TM, Wormsley S, Dunbar DA, Lee S (1997) Mpp10p, a new protein component of the U3 snoRNP required for processing of 18S rRNA precursors. *Nucleic acids symposium series***:** 64-67

Baudin-Baillieu A, Fabret C, Liang XH, Piekna-Przybylska D, Fournier MJ, Rousset JP (2009) Nucleotide modifications in three functionally important regions of the Saccharomyces cerevisiae ribosome affect translation accuracy. *Nucleic Acids Res* **37:** 7665-7677

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:** 5139-5147

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

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011) The structure of the eukaryotic ribosome at 3.0 A resolution. *Science* **334:** 1524-1529

Bleichert F, Gagnon KT, Brown BA, Maxwell ES, Leschziner AE, Unger VM, Baserga SJ (2009) A dimeric structure for archaeal box C/D small ribonucleoproteins. *Science* **325:** 1384-1387

Bleichert F, Granneman S, Osheim YN, Beyer AL, Baserga SJ (2006) The PINc domain protein Utp24, a putative nuclease, is required for the early cleavage steps in 18S rRNA maturation. *Proc Natl Acad Sci U S A* **103:** 9464-9469

Borkovich KA, Farrelly FW, Finkelstein DB, Taulien J, Lindquist S (1989) hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol Cell Biol* **9:** 3919-3930

Borovjagin AV, Gerbi SA (2000) The spacing between functional Cis-elements of U3 snoRNA is critical for rRNA processing. *J Mol Biol* **300:** 57-74

Bortolin ML, Bachellerie JP, Clouet-d'Orval B (2003) In vitro RNP assembly and methylation guide activity of an unusual box C/D RNA, cis-acting archaeal pre-tRNA(Trp). *Nucleic Acids Res* **31:** 6524-6535

Bortolin ML, Ganot P, Kiss T (1999) Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *EMBO J* **18:** 457-469

Boulon S, Marmier-Gourrier N, Pradet-Balade B, Wurth L, Verheggen C, Jády BE, Rothé B, Pescia C, Robert MC, Kiss T, Bardoni B, Krol A, Branlant C, Allmang C, Bertrand E, Charpentier B (2008) The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *J Cell Biol* **180:** 579-595

Boulon S, Verheggen C, Jady BE, Girard C, Pescia C, Paul C, Ospina JK, Kiss T, Matera AG, Bordonné R, Bertrand E (2004) PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. *Mol Cell* **16:** 777-787

Bousquet-Antonelli C, Henry Y, Gélugne J P, Caizergues-Ferrer M, Kiss T (1997) A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs. *EMBO J* **16:** 4770-4776

Brand RC, Klootwijk J, Planta RJ, Maden BE (1978) Biosynthesis of a hypermodified nucleotide in Saccharomyces carlsbergensis 17S and HeLa-cell 18S ribosomal ribonucleic acid. *Biochem J* **169:** 71-77

Branlant C, Krol A, Machatt MA, Pouyet J, Ebel JP, Edwards K, Kössel 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

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

Cabello-Villegas J, Nikonowicz EP (2005) Solution structure of psi32-modified anticodon stem-loop of Escherichia coli tRNAPhe. *Nucleic Acids Res* **33:** 6961-6971

Caffarelli E, Fatica A, Prislei S, De Gregorio E, Fragapane P, Bozzoni I (1996) Processing of the intron-encoded U16 and U18 snoRNAs: the conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA. *EMBO J* **15:** 1121-1131

Cahill NM, Friend K, Speckmann W, Li ZH, Terns RM, Terns MP, Steitz JA (2002) Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J* **21:** 3816-3828

Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Müller KM, Pande N, Shang Z, Yu N, Gutell RR (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3:** 2

Caplan AJ, Mandal AK, Theodoraki MA (2007) Molecular chaperones and protein kinase quality control. *Trends in cell biology* **17:** 87-92

Cavaillé J, Bachellerie JP (1996) Processing of fibrillarin-associated snoRNAs from pre-mRNA introns: an exonucleolytic process exclusively directed by the common stem-box terminal structure. *Biochimie* **78:** 443-456

Cavaillé J, Nicoloso M, Bachellerie JP (1996) Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature* **383:** 732-735

Chanfreau G, Rotondo G, Legrain P, Jacquier A (1998) Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J* **17:** 3726-3737

Chang HC, Lindquist S (1994) Conservation of Hsp90 macromolecular complexes in Saccharomyces cerevisiae. *J Biol Chem* **269:** 24983-24988

Charette M, Gray MW (2000) Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* **49:** 341-351

Charpentier B, Muller S, Branlant C (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res* **33:** 3133-3144

Charron C, Manival X, Cléry A, Senty-Ségault V, Charpentier B, Marmier-Gourrier N, Branlant C, Aubry A (2004) The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J Mol Biol* **342:** 757-773

Chen S, Smith DF (1998) Hop as an adaptor in the heat shock protein 70 (Hsp70) and hsp90 chaperone machinery. *J Biol Chem* **273:** 35194-35200

Cheung KL, Huen J, Kakihara Y, Houry WA, Ortega J (2010) Alternative oligomeric states of the yeast Rvb1/Rvb2 complex induced by histidine tags. *J Mol Biol* **404:** 478-492

Cléry A, Senty-Ségault V, Leclerc F, Raué HA, Branlant C (2007) Analysis of sequence and structural features that identify the B/C motif of U3 small nucleolar RNA as the recognition site for the Snu13p-Rrp9p protein pair. *Mol Cell Biol* **27:** 1191-1206

Collins K (2006) The biogenesis and regulation of telomerase holoenzymes. *Nat Rev Mol Cell Biol* **7:** 484-494

D'Andrea LD, Regan L (2003) TPR proteins: the versatile helix. *Trends Biochem Sci* **28:** 655-662

Darzacq X, Jády BE, Verheggen C, Kiss AM, Bertrand E, Kiss T (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J* **21:** 2746-2756

Darzacq X, Kittur N, Roy S, Shav-Tal Y, Singer RH, Meier UT (2006) Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J Cell Biol* **173:** 207-218

Das AK, Cohen PW, Barford D (1998) The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J* **17:** 1192-1199

Davis DR (1995) Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* **23:** 5020-5026

Davis DR, Poulter CD (1991) 1H-15N NMR studies of Escherichia coli tRNA(Phe) from hisT mutants: a structural role for pseudouridine. *Biochemistry* **30:** 4223-4231

Davis DR, Veltri CA, Nielsen L (1998) An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNALys, tRNAHis and tRNATyr. *J Biomol Struct Dyn* **15:** 1121-1132

De Jonge P, Klootwijk J, Planta RJ (1977) Terminal nucleotide sequences of 17-S ribosomal RNA and its immediate precursor 18-S RNA in yeast. *European journal of biochemistry / FEBS* **72:** 361-369

Decatur WA, Fournier MJ (2002) rRNA modifications and ribosome function. *Trends Biochem Sci* **27:** 344-351

Deng L, Starostina NG, Liu ZJ, Rose JP, Terns RM, Terns MP, Wang BC (2004) Structure determination of fibrillarin from the hyperthermophilic archaeon Pyrococcus furiosus. *Biochem Biophys Res Commun* **315:** 726-732

Dez C, Henras A, Faucon B, Lafontaine D, Caizergues-Ferrer M, Henry Y (2001) Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p. *Nucleic Acids Res* **29:** 598-603

Dez C, Noaillac-Depeyre J, Caizergues-Ferrer M, Henry Y (2002) Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs. *Mol Cell Biol* **22:** 7053-7065

Dezwaan DC, Freeman BC (2008) HSP90: the Rosetta stone for cellular protein dynamics? *Cell Cycle* **7:** 1006-1012

Diebold ML, Fribourg S, Koch M, Metzger T, Romier C (2011) Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. *J Struct Biol* **175:** 178-188

Dincbas-Renqvist V, Engstrom A, Mora L, Heurgue-Hamard V, Buckingham R, Ehrenberg M (2000) A post-translational modification in the GGQ motif of RF2 from Escherichia coli stimulates termination of translation. *EMBO J* **19:** 6900-6907

Donalies UE, Stahl U (2001) Phase-specific gene expression in Saccharomyces cerevisiae, using maltose as carbon source under oxygen-limiting conditions. *Current genetics* **39:** 150-155

Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005a) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21:** 3433-3434

Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005b) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol* **347:** 827-839

Dragon F, Gallagher JE, Compagnone-Post PA, Mitchell BM, Porwancher KA, Wehner KA, Wormsley S, Settlage RE, Shabanowitz J, Osheim Y, Beyer AL, Hunt DF, Baserga SJ (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417:** 967-970

Dragon F, Pogacic V, Filipowicz W (2000) In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs. *Mol Cell Biol* **20:** 3037-3048

Duan J, Li L, Lu J, Wang W, Ye K (2009) Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Mol Cell* **34:** 427-439

Dunbar DA, Wormsley S, Agentis TM, Baserga SJ (1997) Mpp10p, a U3 small nucleolar ribonucleoprotein component required for pre-18S rRNA processing in yeast. *Mol Cell Biol* **17:** 5803-5812

Dutca LM, Gallagher JE, Baserga SJ (2011) The initial U3 snoRNA:pre-rRNA base pairing interaction required for pre-18S rRNA folding revealed by in vivo chemical probing. *Nucleic Acids Res* **39:** 5164-5180

Echeverria PC, Bernthaler A, Dupuis P, Mayer B, Picard D (2011) An interaction network predicted from public data as a discovery tool: application to the Hsp90 molecular chaperone machine. *PloS one* **6:** e26044

Eckert K, Saliou JM, Monlezun L, Vigouroux A, Atmane N, Caillat C, Quevillon-Cheruel S, Madiona K, Nicaise M, Lazereg S, Van Dorsselaer A, Sanglier-Cianferani S, Meyer P,

Morera S (2010) The Pih1-Tah1 cochaperone complex inhibits Hsp90 molecular chaperone ATPase activity. *J Biol Chem* **285:** 31304-31312

Ejby M, Sørensen MA, Pedersen S (2007) Pseudouridylation of helix 69 of 23S rRNA is necessary for an effective translation termination. *Proc Natl Acad Sci U S A* **104:** 19410-19415

Epstein P, Reddy R, Busch H (1984) Multiple states of U3 RNA in Novikoff hepatoma nucleoli. *Biochemistry* **23:** 5421-5425

Faber AW, Van Dijk M, Raue HA, Vos JC (2002) Ngl2p is a Ccr4p-like RNA nuclease essential for the final step in 3'-end processing of 5.8S rRNA in Saccharomyces cerevisiae. *RNA* **8:** 1095-1101

Fatica A, Dlakić M, Tollervey D (2002) Naf1 p is a box H/ACA snoRNP assembly factor. *RNA* **8:** 1502-1514

Fatica A, Galardi S, Altieri F, Bozzoni I (2000) Fibrillarin binds directly and specifically to U16 box C/D snoRNA. *RNA* **6:** 88-95

Fatica A, Oeffinger M, Dlakic M, Tollervey D (2003) Nob1p is required for cleavage of the 3' end of 18S rRNA. *Mol Cell Biol* **23:** 1798-1807

Feder M, Pas J, Wyrwicz LS, Bujnicki JM (2003) Molecular phylogenetics of the RrmJ/fibrillarin superfamily of ribose 2'-O-methyltransferases. *Gene* **302:** 129-138

Fournier R, Brulé F, Ségault V, Mougin A, Branlant C (1998) U3 snoRNA genes with and without intron in the Kluyveromyces genus: yeasts can accommodate great variations of the U3 snoRNA 3'-terminal domain. *RNA* **4:** 285-302

Fox GW, Woese CR (1975) 5S RNA secondary structure. *Nature* **256:** 505-507

Frank J, Zhu J, Penczek P, Li Y, Srivastava S, Verschoor A, Radermacher M, Grassucci R, Lata RK, Agrawal RK (1995) A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome. *Nature* **376:** 441-444

Fribourg S, Romier C, Werten S, Gangloff YG, Poterszman A, Moras D (2001) Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in E. coli. *J Mol Biol* **306:** 363-373

Frolova LY, Tsivkovskii RY, Sivolobova GF, Oparina NY, Serpinsky OI, Blinov VM, Tatkov SI, Kisselev LL (1999) Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* **5:** 1014-1020

Frydman J, Hohfeld J (1997) Chaperones get in touch: the Hip-Hop connection. *Trends Biochem Sci* **22:** 87-92

Galardi S, Fatica A, Bachi A, Scaloni A, Presutti C, Bozzoni I (2002) Purified box C/D snoRNPs are able to reproduce site-specific 2'-O-methylation of target RNA in vitro. *Mol Cell Biol* **22:** 6663-6668

Ganot P, Bortolin ML, Kiss T (1997a) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89:** 799-809

Ganot P, Caizergues-Ferrer M, Kiss T (1997b) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev* **11:** 941-956

Gao H, Sengupta J, Valle M, Korostelev A, Eswar N, Stagg SM, Van Roey P, Agrawal RK, Harvey SC, Sali A, Chapman MS, Frank J (2003) Study of the structural dynamics of the E coli 70S ribosome using real-space refinement. *Cell* **113:** 789-801

Gaspin C, Cavaillé J, Erauso G, Bachellerie JP (2000) Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the Pyrococcus genomes. *J Mol Biol* **297:** 895-906

Gautier T, Bergès T, Tollervey D, Hurt E (1997) Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol Cell Biol* **17:** 7088-7098

Geerlings TH, Vos JC, Raue HA (2000) The final step in the formation of 25S rRNA in Saccharomyces cerevisiae is performed by 5'-->3' exonucleases. *RNA* **6:** 1698-1703

Geoghegan KF, Dixon HB, Rosner PJ, Hoth LR, Lanzetti AJ, Borzilleri KA, Marr ES, Pezzullo LH, Martin LB, LeMotte PK, McColl AS, Kamath AV, Stroh JG (1999) Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. *Analytical biochemistry* **267:** 169-184

Gerbi SA, Borovjagin A (1997) U3 snoRNA may recycle through different compartments of the nucleolus. *Chromosoma* **105:** 401-406

Ginisty H, Sicard H, Roger B, Bouvet P (1999) Structure and functions of nucleolin. *J Cell Sci* **112 (Pt 6):** 761-772

Girard JP, Lehtonen H, Caizergues-Ferrer M, Amalric F, Tollervey D, Lapeyre B (1992) GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *EMBO J* **11:** 673-682

Godin KS, Walbott H, Leulliot N, van Tilbeurgh H, Varani G (2009) The box H/ACA snoRNP assembly factor Shq1p is a chaperone protein homologous to Hsp90 cochaperones that binds to the Cbf5p enzyme. *J Mol Biol* **390:** 231-244

Gonzales FA, Zanchin NI, Luz JS, Oliveira CC (2005) Characterization of Saccharomyces cerevisiae Nop17p, a novel Nop58p-interacting protein that is involved in Pre-rRNA processing. *J Mol Biol* **346:** 437-455

Goody TA, Melcher SE, Norman DG, Lilley DM (2004) The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA* **10:** 254-264

Gorynia S, Bandeiras TM, Pinho FG, McVey CE, Vonrhein C, Round A, Svergun DI, Donner P, Matias PM, Carrondo MA (2011) Structural and functional insights into a dodecameric molecular machine - the RuvBL1/RuvBL2 complex. *J Struct Biol* **176:** 279-291

Granneman S, Kudla G, Petfalski E, Tollervey D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci U S A* **106:** 9613-9618

Granneman S, Pruijn GJ, Horstman W, van Venrooij WJ, Luhrmann R, Watkins NJ (2002) The hU3-55K protein requires 15.5K binding to the box B/C motif as well as flanking RNA elements for its association with the U3 small nucleolar RNA in Vitro. *J Biol Chem* **277:** 48490-48500

Granneman S, Vogelzangs J, Lührmann R, van Venrooij WJ, Pruijn GJ, Watkins NJ (2004) Role of pre-rRNA base pairing and 80S complex formation in subnucleolar localization of the U3 snoRNP. *Mol Cell Biol* **24:** 8600-8610

Gribun A, Cheung KL, Huen J, Ortega J, Houry WA (2008) Yeast Rvb1 and Rvb2 are ATP-dependent DNA helicases that form a heterohexameric complex. *J Mol Biol* **376:** 1320-1333

Gromadski KB, Rodnina MV (2004) Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol Cell* **13:** 191-200

Gross JD, Moerke NJ, von der Haar T, Lugovskoy AA, Sachs AB, McCarthy JE, Wagner G (2003) Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell* **115:** 739-750

Grozdanov PN, Roy S, Kittur N, Meier UT (2009) SHQ1 is required prior to NAF1 for assembly of H/ACA small nucleolar and telomerase RNPs. *RNA* **15:** 1188-1197

Gu X, Liu Y, Santi DV (1999) The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA. *Proc Natl Acad Sci U S A* **96:** 14270-14275

Gutell RR, Larsen N, Woese CR (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev* **58:** 10-26

Hamma T, Ferré-D'Amaré AR (2004) Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution. *Structure* **12:** 893-903

Hamma T, Ferré-D'Amaré AR (2006) Pseudouridine synthases. *Chem Biol* **13:** 1125-1135

Hamma T, Reichow SL, Varani G, Ferré-D'Amaré AR (2005) The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat Struct Mol Biol* **12:** 1101-1107

Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, Poustka A, Dokal I (1998) X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet* **19:** 32-38

Henras A, Dez C, Noaillac-Depeyre J, Henry Y, Caizergues-Ferrer M (2001) Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. *Nucleic Acids Res* **29:** 2733-2746

Henras A, Henry Y, Bousquet-Antonelli C, Noaillac-Depeyre J, Gélugne JP, Caizergues-Ferrer M (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J* **17:** 7078-7090

Henras AK, Bertrand E, Chanfreau G (2004) A cotranscriptional model for 3'-end processing of the Saccharomyces cerevisiae pre-ribosomal RNA precursor. *RNA* **10:** 1572-1585

Henras AK, Soudet J, Gerus M, Lebaron S, Caizergues-Ferrer M, Mougin A, Henry Y (2008) The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cellular and molecular life sciences : CMLS* **65:** 2334-2359

Henry Y, Wood H, Morrissey JP, Petfalski E, Kearsey S, 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

Hessling M, Richter K, Buchner J (2009) Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. *Nat Struct Mol Biol* **16:** 287-293

Hirabayashi N, Sato NS, Suzuki T (2006) Conserved loop sequence of helix 69 in Escherichia coli 23 S rRNA is involved in A-site tRNA binding and translational fidelity. *J Biol Chem* **281:** 17203-17211

Hirashima A, Kaji A (1973) Role of elongation factor G and a protein factor on the release of ribosomes from messenger ribonucleic acid. *J Biol Chem* **248:** 7580-7587

Hoang C, Ferré-D'Amaré AR (2001) Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **107:** 929-939

Horn DM, Mason SL, Karbstein K (2011) Rcl1 protein, a novel nuclease for 18 S ribosomal RNA production. *J Biol Chem* **286:** 34082-34087

Huang GM, Jarmolowski A, Struck JC, Fournier MJ (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

Huang L, Pookanjanatavip M, Gu X, Santi DV (1998) A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry* **37:** 344-351

Hughes JM (1996) Functional base-pairing interaction between highly conserved elements of U3 small nucleolar RNA and the small ribosomal subunit RNA. *J Mol Biol* **259:** 645-654

Hughes JM, Ares M, Jr. (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 JM, Konings DA, Cesareni G (1987) The yeast homologue of U3 snRNA. *EMBO J* **6:** 2145-2155

Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK (2003) Global analysis of protein localization in budding yeast. *Nature* **425:** 686-691

Ito K, Uno M, Nakamura Y (2000) A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature* **403:** 680-684

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* **98:** 4569-4574

Jackson RJ, Hellen CU, Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11:** 113-127

Jansen RP, Hurt EC, Kern H, Lehtonen H, Carmo-Fonseca M, Lapeyre B, Tollervey D (1991) Evolutionary conservation of the human nucleolar protein fibrillarin and its functional expression in yeast. *J Cell Biol* **113:** 715-729

Jenner L, Demeshkina N, Yusupova G, Yusupov M (2010) Structural rearrangements of the ribosome at the tRNA proofreading step. *Nat Struct Mol Biol* **17:** 1072-1078

Jha S, Dutta A (2009) RVB1/RVB2: running rings around molecular biology. *Mol Cell* **34:** 521-533

Jiang W, Middleton K, Yoon HJ, Fouquet C, Carbon J (1993) An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules. *Mol Cell Biol* **13:** 4884-4893

Johnson BD, Schumacher RJ, Ross ED, Toft DO (1998) Hop modulates Hsp70/Hsp90 interactions in protein folding. *J Biol Chem* **273:** 3679-3686

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292:** 195-202

Jones KW, Gorzynski K, Hales CM, Fischer U, Badbanchi F, Terns RM, Terns MP (2001) Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin. *J Biol Chem* **276:** 38645-38651

Kakihara Y, Houry WA (2012) The R2TP complex: discovery and functions. *Biochim Biophys Acta* **1823:** 101-107

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

Kawai G, Yamamoto Y, Kamimura T, Masegi T, Sekine M, Hata T, Iimori T, Watanabe T, Miyazawa T, Yokoyama S (1992) Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* **31:** 1040-1046

Khanna M, Wu H, Johansson C, Caizergues-Ferrer M, Feigon J (2006) Structural study of the H/ACA snoRNP components Nop10p and the 3' hairpin of U65 snoRNA. *RNA* **12:** 40-52

King TH, Decatur WA, Bertrand E, Maxwell ES, Fournier MJ (2001) A well-connected and conserved nucleoplasmic helicase is required for production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins. *Mol Cell Biol* **21:** 7731-7746

King TH, Liu B, McCully RR, Fournier MJ (2003) Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell* **11:** 425-435

Kiss T, Fayet E, Jády BE, Richard P, Weber M (2006) Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harb Symp Quant Biol* **71:** 407-417

Kiss-László Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85:** 1077-1088

Kiss-László Z, Henry Y, Kiss T (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J* **17:** 797-807

Kittur N, Darzacq X, Roy S, Singer RH, Meier UT (2006) Dynamic association and localization of human H/ACA RNP proteins. *RNA* **12:** 2057-2062

Klein DJ, Schmeing TM, Moore PB, Steitz TA (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J* **20:** 4214-4221

Knox AA, McKeegan KS, Debieux CM, Traynor A, Richardson H, Watkins NJ (2011) A weak C' box renders U3 snoRNA levels dependent on hU3-55K binding. *Mol Cell Biol* **31:** 2404-2412

Koo BK, Park CJ, Fernandez CF, Chim N, Ding Y, Chanfreau G, Feigon J (2011) Structure of H/ACA RNP protein Nhp2p reveals cis/trans isomerization of a conserved proline at the RNA and Nop10 binding interface. *J Mol Biol* **411:** 927-942

Koonin EV (1996) Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res* **24:** 2411-2415

Korostelev A, Asahara H, Lancaster L, Laurberg M, Hirschi A, Zhu J, Trakhanov S, Scott WG, Noller HF (2008) Crystal structure of a translation termination complex formed with release factor RF2. *Proc Natl Acad Sci U S A* **105:** 19684-19689

Kressler D, Doère M, Rojo M, Linder P (1999) Synthetic lethality with conditional dbp6 alleles identifies rsa1p, a nucleoplasmic protein involved in the assembly of 60S ribosomal subunits. *Mol Cell Biol* **19:** 8633-8645

Krol A, Branlant C, Lazar E, Gallinaro H, Jacob M (1981) Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs. *Nucleic Acids Res* **9:** 2699-2716

Kufel J, Allmang C, Chanfreau G, Petfalski E, Lafontaine DL, Tollervey D (2000) Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding. *Mol Cell Biol* **20:** 5415-5424

Kufel J, Allmang C, Verdone L, Beggs J, Tollervey D (2003) A complex pathway for 3' processing of the yeast U3 snoRNA. *Nucleic Acids Res* **31:** 6788-6797

Kufel J, Dichtl B, Tollervey D (1999) Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. *RNA* **5:** 909-917

Kuhn JF, Tran EJ, Maxwell ES (2002) Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res* **30:** 931-941

Lafontaine DL, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D (1998) The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev* **12:** 527-537

Lafontaine DL, Tollervey D (1999) Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA* **5:** 455-467

Lafontaine DL, Tollervey D (2000) Synthesis and assembly of the box C+D small nucleolar RNPs. *Mol Cell Biol* **20:** 2650-2659

Lamanna AC, Karbstein K (2009) Nob1 binds the single-stranded cleavage site D at the 3'-end of 18S rRNA with its PIN domain. *Proc Natl Acad Sci U S A* **106:** 14259-14264

Lamb JR, Tugendreich S, Hieter P (1995) Tetratrico peptide repeat interactions: to TPR or not to TPR? *Trends Biochem Sci* **20:** 257-259

Lane BG, Ofengand J, Gray MW (1995) Pseudouridine and O2'-methylated nucleosides. Significance of their selective occurrence in rRNA domains that function in ribosome-catalyzed synthesis of the peptide bonds in proteins. *Biochimie* **77:** 7-15

Lange TS, Ezrokhi M, Borovjagin AV, Rivera-León R, North MT, Gerbi SA (1998) Nucleolar localization elements of Xenopus laevis U3 small nucleolar RNA. *Mol Biol Cell* **9:** 2973-2985

Laurberg M, Asahara H, Korostelev A, Zhu J, Trakhanov S, Noller HF (2008) Structural basis for translation termination on the 70S ribosome. *Nature* **454:** 852-857

Lee SJ, Baserga SJ (1997) Functional separation of pre-rRNA processing steps revealed by truncation of the U3 small nucleolar ribonucleoprotein component, Mpp10. *Proc Natl Acad Sci U S A* **94:** 13536-13541

Lee SJ, Baserga SJ (1999) Imp3p and Imp4p, two specific components of the U3 small nucleolar ribonucleoprotein that are essential for pre-18S rRNA processing. *Mol Cell Biol* **19:** 5441-5452

Lestrade L, Weber MJ (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* **34:** D158-162

Leulliot N, Godin KS, Hoareau-Aveilla C, Quevillon-Cheruel S, Varani G, Henry Y, Van Tilbeurgh H (2007) The box H/ACA RNP assembly factor Naf1p contains a domain homologous to Gar1p mediating its interaction with Cbf5p. *J Mol Biol* **371:** 1338-1353

Li D, Fournier MJ (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 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

Li J, Richter K, Buchner J (2011a) Mixed Hsp90-cochaperone complexes are important for the progression of the reaction cycle. *Nat Struct Mol Biol* **18:** 61-66

Li J, Soroka J, Buchner J (2012) The Hsp90 chaperone machinery: Conformational dynamics and regulation by co-chaperones. *Biochim Biophys Acta* **1823:** 624-635

Li L, Ye K (2006) Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443:** 302-307

Li S, Duan J, Li D, Ma S, Ye K (2011b) Structure of the Shq1-Cbf5-Nop10-Gar1 complex and implications for H/ACA RNP biogenesis and dyskeratosis congenita. *EMBO J* **30:** 5010-5020

Li S, Duan J, Li D, Yang B, Dong M, Ye K (2011c) Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes Dev* **25:** 2409-2421

Liang B, Xue S, Terns RM, Terns MP, Li H (2007a) Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. *Nat Struct Mol Biol* **14:** 1189-1195

Liang B, Zhou J, Kahen E, Terns RM, Terns MP, Li H (2009a) Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. *Nat Struct Mol Biol* **16:** 740-746

Liang XH, Liu Q, Fournier MJ (2007b) rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Mol Cell* **28:** 965-977

Liang XH, Liu Q, Fournier MJ (2009b) Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA* **15:** 1716-1728

Limbach PA, Crain PF, McCloskey JA (1994) Summary: the modified nucleosides of RNA. *Nucleic Acids Res* **22:** 2183-2196

Lin J, Lai S, Jia R, Xu A, Zhang L, Lu J, Ye K (2011) Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* **469:** 559-563

Lindahl L, Archer RH, Zengel JM (1994) Alternate pathways for processing in the internal transcribed spacer 1 in pre-rRNA of Saccharomyces cerevisiae. *Nucleic Acids Res* **22:** 5399-5407

Liu S, Li P, Dybkov O, Nottrott S, Hartmuth K, Lührmann R, Carlomagno T, Wahl MC (2007) Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science* **316:** 115-120

Lorsch JR, Dever TE (2010) Molecular view of 43 S complex formation and start site selection in eukaryotic translation initiation. *J Biol Chem* **285:** 21203-21207

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. *Nucleic Acids Res* **21:** 5377-5385

Lukowiak AA, Granneman S, Mattox SA, Speckmann WA, Jones K, Pluk H, Venrooij WJ, Terns RM, Terns MP (2000) Interaction of the U3-55k protein with U3 snoRNA is mediated by the box B/C motif of U3 and the WD repeats of U3-55k. *Nucleic Acids Res* **28:** 3462-3471

Lygerou Z, Allmang C, Tollervey D, Seraphin B (1996) Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro. *Science* **272:** 268-270

Maden BE (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol* **39:** 241-303

Makhnevych T, Houry WA (2012) The role of Hsp90 in protein complex assembly. *Biochim Biophys Acta* **1823:** 674-682

Manival X, Charron C, Fourmann JB, Godard F, Charpentier B, Branlant C (2006) Crystal structure determination and site-directed mutagenesis of the Pyrococcus abyssi aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res* **34:** 826-839

Marmier-Gourrier N, Cléry A, Schlotter F, Senty-Ségault V, Branlant C (2011) A second base pair interaction between U3 small nucleolar RNA and the 5'-ETS region is required for early cleavage of the yeast pre-ribosomal RNA. *Nucleic Acids Res* **39:** 9731-9745

Marmier-Gourrier N, Cléry A, Senty-Ségault V, Charpentier B, Schlotter F, Leclerc F, Fournier R, Branlant C (2003) A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA* **9:** 821-838

Massenet S, Mougin A, Branlant C (1998) Posttranscriptional modifications in the U small nuclear RNAs. In *Modifications and Editing of RNA*, Grosjean H, Benne R (eds), ASM Press, Washington, DC edn, pp 201-228.

Matera AG, Terns RM, Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* **8:** 209-220

Matias PM, Gorynia S, Donner P, Carrondo MA (2006) Crystal structure of the human AAA+ protein RuvBL1. *J Biol Chem* **281:** 38918-38929

McKeegan KS, Debieux CM, Watkins NJ (2009) Evidence that the AAA+ proteins TIP48 and TIP49 bridge interactions between 15.5K and the related NOP56 and NOP58 proteins during box C/D snoRNP biogenesis. *Mol Cell Biol* **29:** 4971-4981

Meier UT (2006) How a single protein complex accommodates many different H/ACA RNAs. *Trends Biochem Sci* **31:** 311-315

Méreau A, Fournier R, Grégoire A, Mougin A, Fabrizio P, Lührmann R, Branlant C (1997) An in vivo and in vitro structure-function analysis of the Saccharomyces cerevisiae U3A snoRNP: protein-RNA contacts and base-pair interaction with the pre-ribosomal RNA. *J Mol Biol* **273:** 552-571

Meyer P, Prodromou C, Hu B, Vaughan C, Roe SM, Panaretou B, Piper PW, Pearl LH (2003) Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. *Mol Cell* **11:** 647-658

Meyer P, Prodromou C, Liao C, Hu B, Roe SM, Vaughan CK, Vlasic I, Panaretou B, Piper PW, Pearl LH (2004) Structural basis for recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. *EMBO J* **23:** 1402-1410

Millson SH, Vaughan CK, Zhai C, Ali MM, Panaretou B, Piper PW, Pearl LH, Prodromou C (2008) Chaperone ligand-discrimination by the TPR-domain protein Tah1. *Biochem J* **413:** 261-268

Mitchell JR, Wood E, Collins K (1999) A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402:** 551-555

Mitchell P, Petfalski E, Tollervey D (1996) The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev* **10:** 502-513

Moazed D, Noller HF (1989) Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell* **57:** 585-597

Mollapour M, Neckers L (2012) Post-translational modifications of Hsp90 and their contributions to chaperone regulation. *Biochim Biophys Acta* **1823:** 648-655

Moore T, Zhang Y, Fenley MO, Li H (2004) Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure* **12:** 807-818

Morrissey JP, Tollervey D (1993) Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. *Mol Cell Biol* **13:** 2469-2477

Mouaikel J, Verheggen C, Bertrand E, Tazi J, Bordonné R (2002) Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol Cell* **9:** 891-901

Muller S, Fourmann JB, Loegler C, Charpentier B, Branlant C (2007) Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Psi55-synthase and RNA-guided RNA:Psi-synthase activities. *Nucleic Acids Res* **35:** 5610-5624

Muller S, Leclerc F, Behm-Ansmant I, Fourmann JB, Charpentier B, Branlant C (2008) Combined in silico and experimental identification of the Pyrococcus abyssi H/ACA sRNAs and their target sites in ribosomal RNAs. *Nucleic Acids Res* **36:** 2459-2475

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

Narayanan A, Speckmann W, Terns R, Terns MP (1999) Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Mol Biol Cell* **10:** 2131-2147

Newman DR, Kuhn JF, Shanab GM, Maxwell ES (2000) Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA* **6:** 861-879

Ni J, Tien AL, Fournier MJ (1997) Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89:** 565-573

Nolivos S, Carpousis AJ, Clouet-d'Orval B (2005) The K-loop, a general feature of the Pyrococcus C/D guide RNAs, is an RNA structural motif related to the K-turn. *Nucleic Acids Res* **33:** 6507-6514

Noller HF (2006) Biochemical characterization of the ribosomal decoding site. *Biochimie* **88:** 935-941

Noller HF, Kop J, Wheaton V, Brosius J, Gutell RR, Kopylov AM, Dohme F, Herr W, Stahl DA, Gupta R, Waese CR (1981) Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res* **9:** 6167-6189

Noller HF, Woese CR (1981) Secondary structure of 16S ribosomal RNA. *Science* **212:** 403-411

Normand C, Capeyrou R, Quevillon-Cheruel S, Mougin A, Henry Y, Caizergues-Ferrer M (2006) Analysis of the binding of the N-terminal conserved domain of yeast Cbf5p to a box H/ACA snoRNA. *RNA* **12:** 1868-1882

Nottrott S, Hartmuth K, Fabrizio P, Urlaub H, Vidovic I, Ficner R, Lührmann R (1999) Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J* **18:** 6119-6133

O'Farrell HC, Pulicherla N, Desai PM, Rife JP (2006) Recognition of a complex substrate by the KsgA/Dim1 family of enzymes has been conserved throughout evolution. *RNA* **12:** 725-733

Ofengand J, Bakin A (1997) Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. *J Mol Biol* **266:** 246-268

Ofengand J, Bakin A, Wrzesinski J, Nurse K, Lane BG (1995) The pseudouridine residues of ribosomal RNA. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **73:** 915-924

Ogle JM, Brodersen DE, Clemons WM, Jr., Tarry MJ, Carter AP, Ramakrishnan V (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292:** 897-902

Ogle JM, Murphy FV, Tarry MJ, Ramakrishnan V (2002) Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **111:** 721-732

Ogle JM, Ramakrishnan V (2005) Structural insights into translational fidelity. *Annual review of biochemistry* **74:** 129-177

Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP (2000) Homologs of small nucleolar RNAs in Archaea. *Science* **288:** 517-522

Omer AD, Ziesche S, Ebhardt H, Dennis PP (2002) In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc Natl Acad Sci U S A* **99:** 5289-5294

Ooi SL, Samarsky DA, Fournier MJ, Boeke JD (1998) Intronic snoRNA biosynthesis in Saccharomyces cerevisiae depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA. *RNA* **4:** 1096-1110

Oruganti S, Zhang Y, Li H (2005) Structural comparison of yeast snoRNP and spliceosomal protein Snu13p with its homologs. *Biochem Biophys Res Commun* **333:** 550-554

Oruganti S, Zhang Y, Li H, Robinson H, Terns MP, Terns RM, Yang W (2007) Alternative conformations of the archaeal Nop56/58-fibrillarin complex imply flexibility in box C/D RNPs. *J Mol Biol* **371:** 1141-1150

Osheim YN, French SL, Keck KM, Champion EA, Spasov K, Dragon F, Baserga SJ, Beyer AL (2004) Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in Saccharomyces cerevisiae. *Mol Cell* **16:** 943-954

Panaretou B, Siligardi G, Meyer P, Maloney A, Sullivan JK, Singh S, Millson SH, Clarke PA, Naaby-Hansen S, Stein R, Cramer R, Mollapour M, Workman P, Piper PW, Pearl LH, Prodromou C (2002) Activation of the ATPase activity of hsp90 by the stress-regulated cochaperone aha1. *Mol Cell* **10:** 1307-1318

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

Pellizzoni L, Baccon J, Charroux B, Dreyfuss G (2001) The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. *Current biology : CB* **11:** 1079-1088

Petfalski E, Dandekar T, Henry Y, Tollervey D (1998) Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol Cell Biol* **18:** 1181-1189

Piekna-Przybylska D, Decatur WA, Fournier MJ (2007) New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. *RNA* **13:** 305-312

Piekna-Przybylska D, Przybylski P, Baudin-Baillieu A, Rousset JP, Fournier MJ (2008) Ribosome performance is enhanced by a rich cluster of pseudouridines in the A-site finger region of the large subunit. *J Biol Chem* **283:** 26026-26036

Pluk H, Soffner J, Lührmann R, van Venrooij WJ (1998) cDNA cloning and characterization of the human U3 small nucleolar ribonucleoprotein complex-associated 55-kilodalton protein. *Mol Cell Biol* **18:** 488-498

Pogacić V, Dragon F, Filipowicz W (2000) Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. *Mol Cell Biol* **20:** 9028-9040

Polacek N, Mankin AS (2005) The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. *Crit Rev Biochem Mol Biol* **40:** 285-311

Ponchon L, Beauvais G, Nonin-Lecomte S, Dardel F (2009) A generic protocol for the expression and purification of recombinant RNA in Escherichia coli using a tRNA scaffold. *Nature protocols* **4:** 947-959

Ponchon L, Dardel F (2011) Large scale expression and purification of recombinant RNA in Escherichia coli. *Methods* **54:** 267-273

Pradet-Balade B, Girard C, Boulon S, Paul C, Azzag K, Bordonne R, Bertrand E, Verheggen C (2011) CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *EMBO J* **30:** 2205-2218

Prasad BD, Goel S, Krishna P (2010) In silico identification of carboxylate clamp type tetratricopeptide repeat proteins in Arabidopsis and rice as putative co-chaperones of Hsp90/Hsp70. *PloS one* **5:** e12761

Prodromou C (2012) The 'active life' of Hsp90 complexes. *Biochim Biophys Acta* **1823:** 614-623

Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW, Pearl LH (1997) Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. *Cell* **90:** 65-75

Prodromou C, Siligardi G, O'Brien R, Woolfson DN, Regan L, Panaretou B, Ladbury JE, Piper PW, Pearl LH (1999) Regulation of Hsp90 ATPase activity by tetratricopeptide repeat (TPR)-domain co-chaperones. *EMBO J* **18:** 754-762

Puri T, Wendler P, Sigala B, Saibil H, Tsaneva IR (2007) Dodecameric structure and ATPase activity of the human TIP48/TIP49 complex. *J Mol Biol* **366:** 179-192

Putnam CD, Clancy SB, Tsuruta H, Gonzalez S, Wetmur JG, Tainer JA (2001) Structure and mechanism of the RuvB Holliday junction branch migration motor. *J Mol Biol* **311:** 297-310

Qu G, van Nues RW, Watkins NJ, Maxwell ES (2011) The spatial-functional coupling of box C/D and C'/D' RNPs is an evolutionarily conserved feature of the eukaryotic box C/D snoRNP nucleotide modification complex. *Mol Cell Biol* **31:** 365-374

Qu LH, Henras A, Lu YJ, Zhou H, Zhou WX, Zhu YQ, Zhao J, Henry Y, Caizergues-Ferrer M, Bachellerie JP (1999) Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. *Mol Cell Biol* **19:** 1144-1158

Ramamurthy V, Swann SL, Paulson JL, Spedaliere CJ, Mueller EG (1999) Critical aspartic acid residues in pseudouridine synthases. *J Biol Chem* **274:** 22225-22230

Rashid R, Aittaleb M, Chen Q, Spiegel K, Demeler B, Li H (2003) Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles. *J Mol Biol* **333:** 295-306

Rashid R, Liang B, Baker DL, Youssef OA, He Y, Phipps K, Terns RM, Terns MP, Li H (2006) Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol Cell* **21:** 249-260

Reddy R, Busch H (1983) Small nuclear RNAs and RNA processing. *Prog Nucleic Acid Res Mol Biol* **30:** 127-162

Reddy R, Henning D, Busch H (1980) Substitutions, insertions, and deletions in two highly conserved U3 RNA species. *J Biol Chem* **255:** 7029-7033

Reichow SL, Hamma T, Ferré-D'Amaré AR, Varani G (2007) The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res* **35:** 1452-1464

Reichow SL, Varani G (2008) Nop10 is a conserved H/ACA snoRNP molecular adaptor. *Biochemistry* **47:** 6148-6156

Reilly TH, Schmitt ME (1995) The yeast, Saccharomyces cerevisiae, RNase P/MRP ribonucleoprotein endoribonuclease family. *Molecular biology reports* **22:** 87-93

Richard P, Darzacq X, Bertrand E, Jády BE, Verheggen C, Kiss T (2003) A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J* **22:** 4283-4293

Richard P, Kiss AM, Darzacq X, Kiss T (2006) Cotranscriptional recognition of human intronic box H/ACA snoRNAs occurs in a splicing-independent manner. *Mol Cell Biol* **26:** 2540-2549

Richter K, Haslbeck M, Buchner J (2010) The heat shock response: life on the verge of death. *Mol Cell* **40:** 253-266

Richter K, Muschler P, Hainzl O, Reinstein J, Buchner J (2003) Sti1 is a non-competitive inhibitor of the Hsp90 ATPase. Binding prevents the N-terminal dimerization reaction during the atpase cycle. *J Biol Chem* **278:** 10328-10333

Richter K, Walter S, Buchner J (2004) The Co-chaperone Sba1 connects the ATPase reaction of Hsp90 to the progression of the chaperone cycle. *J Mol Biol* **342:** 1403-1413

Rife JP, Moore PB (1998) The structure of a methylated tetraloop in 16S ribosomal RNA. *Structure* **6:** 747-756

Roe SM, Ali MM, Meyer P, Vaughan CK, Panaretou B, Piper PW, Prodromou C, Pearl LH (2004) The Mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50(cdc37). *Cell* **116:** 87-98

Roovers M, Hale C, Tricot C, Terns MP, Terns RM, Grosjean H, Droogmans L (2006) Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic Acids Res* **34:** 4293-4301

Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, Hüttenhofer A (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res* **31:** 869-877

Samarsky DA, Fournier MJ (1998) Functional mapping of the U3 small nucleolar RNA from the yeast Saccharomyces cerevisiae. *Mol Cell Biol* **18:** 3431-3444

Samarsky DA, Fournier MJ, Singer RH, Bertrand E (1998) The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J* **17:** 3747-3757

Scheufler C, Brinker A, Bourenkov G, Pegoraro S, Moroder L, Bartunik H, Hartl FU, Moarefi I (2000) Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell* **101:** 199-210

Schmeing TM, Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461:** 1234-1242

Schuette JC, Murphy FVt, Kelley AC, Weir JR, Giesebrecht J, Connell SR, Loerke J, Mielke T, Zhang W, Penczek PA, Ramakrishnan V, Spahn CM (2009) GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *EMBO J* **28:** 755-765

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

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR* **44:** 213-223

Siligardi G, Panaretou B, Meyer P, Singh S, Woolfson DN, Piper PW, Pearl LH, Prodromou C (2002) Regulation of Hsp90 ATPase activity by the co-chaperone Cdc37p/p50cdc37. *J Biol Chem* **277:** 20151-20159

Singh M, Gonzales FA, Cascio D, Heckmann N, Chanfreau G, Feigon J (2009) Structure and functional studies of the CS domain of the essential H/ACA ribonucleoparticle assembly protein SHQ1. *J Biol Chem* **284:** 1906-1916

Smith JE, Cooperman BS, Mitchell P (1992) Methylation sites in Escherichia coli ribosomal RNA: localization and identification of four new sites of methylation in 23S rRNA. *Biochemistry* **31:** 10825-10834

Song H, Mugnier P, Das AK, Webb HM, Evans DR, Tuite MF, Hemmings BA, Barford D (2000) The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* **100:** 311-321

Speckmann W, Narayanan A, Terns R, Terns MP (1999) Nuclear retention elements of U3 small nucleolar RNA. *Mol Cell Biol* **19:** 8412-8421

Speckmann WA, Terns RM, Terns MP (2000) The box C/D motif directs snoRNA 5'-cap hypermethylation. *Nucleic Acids Res* **28:** 4467-4473

Spedaliere CJ, Hamilton CS, Mueller EG (2000) Functional importance of motif I of pseudouridine synthases: mutagenesis of aligned lysine and proline residues. *Biochemistry* **39:** 9459-9465

Steinberg S, Misch A, Sprinzl M (1993) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **21:** 3011-3015

Stoneley M, Willis AE (2004) Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene* **23:** 3200-3207

Suryadi J, Tran EJ, Maxwell ES, Brown BA (2005) The crystal structure of the Methanocaldococcus jannaschii multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the box C/D RNAs. *Biochemistry* **44:** 9657-9672

Szewczak LB, DeGregorio SJ, Strobel SA, Steitz JA (2002) Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem Biol* **9:** 1095-1107

Taipale M, Jarosz DF, Lindquist S (2010) HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol* **11:** 515-528

Tapia H, Morano KA (2010) Hsp90 nuclear accumulation in quiescence is linked to chaperone function and spore development in yeast. *Mol Biol Cell* **21:** 63-72

Terns M, Terns R (2006) Noncoding RNAs of the H/ACA family. *Cold Spring Harb Symp Quant Biol* **71:** 395-405

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

Torreira E, Jha S, López-Blanco JR, Arias-Palomo E, Chacón P, Cañas C, Ayora S, Dutta A, Llorca O (2008) Architecture of the pontin/reptin complex, essential in the assembly of several macromolecular complexes. *Structure* **16:** 1511-1520

Tran E, Zhang X, Lackey L, Maxwell ES (2005) Conserved spacing between the box C/D and C'/D' RNPs of the archaeal box C/D sRNP complex is required for efficient 2'-O-methylation of target RNAs. *RNA* **11:** 285-293

Tran EJ, Zhang X, Maxwell ES (2003) Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J* **22:** 3930-3940

Trepel J, Mollapour M, Giaccone G, Neckers L (2010) Targeting the dynamic HSP90 complex in cancer. *Nature reviews Cancer* **10:** 537-549

Van Hoof A, Lennertz P, Parker R (2000a) Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. *EMBO J* **19:** 1357-1365

van Hoof A, Lennertz P, Parker R (2000b) Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol Cell Biol* **20:** 441-452

Veldman GM, Klootwijk J, de Jonge P, Leer RJ, Planta RJ (1980) The transcription termination site of the ribosomal RNA operon in yeast. *Nucleic Acids Res* **8:** 5179-5192

Veldman GM, Klootwijk J, de Regt VC, Planta RJ, Branlant C, Krol A, Ebel JP (1981) The primary and secondary structure of yeast 26S rRNA. *Nucleic Acids Res* **9:** 6935-6952

Venema J, Vos HR, Faber AW, van Venrooij WJ, Raué HA (2000) Yeast Rrp9p is an evolutionarily conserved U3 snoRNP protein essential for early pre-rRNA processing cleavages and requires box C for its association. *RNA* **6:** 1660-1671

Verheggen C, Lafontaine DL, Samarsky D, Mouaikel J, Blanchard JM, Bordonné R, Bertrand E (2002) Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. *EMBO J* **21:** 2736-2745

Verschoor A, Warner JR, Srivastava S, Grassucci RA, Frank J (1998) Three-dimensional structure of the yeast ribosome. *Nucleic Acids Res* **26:** 655-661

Vidovic I, Nottrott S, Hartmuth K, Lührmann R, Ficner R (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell* **6:** 1331-1342

Villa E, Sengupta J, Trabuco LG, LeBarron J, Baxter WT, Shaikh TR, Grassucci RA, Nissen P, Ehrenberg M, Schulten K, Frank J (2009) Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc Natl Acad Sci U S A* **106:** 1063-1068

von der Haar T, Gross JD, Wagner G, McCarthy JE (2004) The mRNA cap-binding protein eIF4E in post-transcriptional gene expression. *Nat Struct Mol Biol* **11:** 503-511

Voorhees RM, Weixlbaumer A, Loakes D, Kelley AC, Ramakrishnan V (2009) Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol* **16:** 528-533

Walbott H, Machado-Pinilla R, Liger D, Blaud M, Réty S, Grozdanov PN, Godin K, van Tilbeurgh H, Varani G, Meier UT, Leulliot N (2011) The H/ACA RNP assembly factor SHQ1 functions as an RNA mimic. *Genes Dev* **25:** 2398-2408

Wandinger SK, Richter K, Buchner J (2008) The Hsp90 chaperone machinery. *J Biol Chem* **283:** 18473-18477

Wang C, Meier UT (2004) Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J* **23:** 1857-1867

Wang H, Boisvert D, Kim KK, Kim R, Kim SH (2000) Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6 A resolution. *EMBO J* **19:** 317-323

Warner JR (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* **24:** 437-440

Watkins NJ, Dickmanns A, Lührmann R (2002) Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol* **22:** 8342-8352

Watkins NJ, Gottschalk A, Neubauer G, Kastner B, Fabrizio P, Mann M, Lührmann R (1998a) Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA* **4:** 1549-1568

Watkins NJ, Lemm I, Ingelfinger D, Schneider C, Hossbach M, Urlaub H, Lührmann R (2004) Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. *Mol Cell* **16:** 789-798

Watkins NJ, Leverette RD, Xia L, Andrews MT, Maxwell ES (1996) Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA* **2:** 118-133

Watkins NJ, Newman DR, Kuhn JF, Maxwell ES (1998b) In vitro assembly of the mouse U14 snoRNP core complex and identification of a 65-kDa box C/D-binding protein. *RNA* **4:** 582-593

Watkins NJ, Ségault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Lührmann R (2000) A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103:** 457-466

Wegele H, Haslbeck M, Reinstein J, Buchner J (2003) Sti1 is a novel activator of the Ssa proteins. *J Biol Chem* **278:** 25970-25976

Wegele H, Wandinger SK, Schmid AB, Reinstein J, Buchner J (2006) Substrate transfer from the chaperone Hsp70 to Hsp90. *J Mol Biol* **356:** 802-811

Wehner KA, Gallagher JE, Baserga SJ (2002) Components of an interdependent unit within the SSU processome regulate and mediate its activity. *Mol Cell Biol* **22:** 7258-7267

Weixlbaumer A, Jin H, Neubauer C, Voorhees RM, Petry S, Kelley AC, Ramakrishnan V (2008) Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* **322:** 953-956

Whitehead SE, Jones KW, Zhang X, Cheng X, Terns RM, Terns MP (2002) Determinants of the interaction of the spinal muscular atrophy disease protein SMN with the dimethylarginine-modified box H/ACA small nucleolar ribonucleoprotein GAR1. *J Biol Chem* **277:** 48087-48093

Winkler AA, Bobok A, Zonneveld BJ, Steensma HY, Hooykaas PJ (1998) The lysine-rich C-terminal repeats of the centromere-binding factor 5 (Cbf5) of Kluyveromyces lactis are not essential for function. *Yeast* **14:** 37-48

Woese CR, Magrum LJ, Gupta R, Siegel RB, Stahl DA, Kop J, Crawford N, Brosius J, Gutell R, Hogan JJ, Noller HF (1980) Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res* **8:** 2275-2293

Wu H, Min J, Zeng H, Loppnau P, Weigelt J, Sundstrom M, Arrowsmith CH, Edwards AM, Bochkarev A, Plotnikov AN. The Crystal Structure of Human Fibrillarin in complex with SAH.

Wurm JP, Meyer B, Bahr U, Held M, Frolow O, Kötter P, Engels JW, Heckel A, Karas M, Entian KD, Wöhnert J (2010) The ribosome assembly factor Nep1 responsible for Bowen-Conradi syndrome is a pseudouridine-N1-specific methyltransferase. *Nucleic Acids Res* **38:** 2387-2398

Xia L, Watkins NJ, Maxwell ES (1997) Identification of specific nucleotide sequences and structural elements required for intronic U14 snoRNA processing. *RNA* **3:** 17-26

Xia T, SantaLucia J, Jr., Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **37:** 14719-14735

Xu Z, O'Farrell HC, Rife JP, Culver GM (2008) A conserved rRNA methyltransferase regulates ribosome biogenesis. *Nat Struct Mol Biol* **15:** 534-536

Xue S, Wang R, Yang F, Terns RM, Terns MP, Zhang X, Maxwell ES, Li H (2010) Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Mol Cell* **39:** 939-949

Xue Y, Bai X, Lee I, Kallstrom G, Ho J, Brown J, Stevens A, Johnson AW (2000) Saccharomyces cerevisiae RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. *Mol Cell Biol* **20:** 4006-4015

Yang PK, Hoareau C, Froment C, Monsarrat B, Henry Y, Chanfreau G (2005) Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. *Mol Cell Biol* **25:** 3295-3304

Yang PK, Rotondo G, Porras T, Legrain P, Chanfreau G (2002) The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis. *J Biol Chem* **277:** 45235-45242

Ye K, Jia R, Lin J, Ju M, Peng J, Xu A, Zhang L (2009) Structural organization of box C/D RNA-guided RNA methyltransferase. *Proc Natl Acad Sci U S A* **106:** 13808-13813

Young JC, Moarefi I, Hartl FU (2001) Hsp90: a specialized but essential protein-folding tool. *J Cell Biol* **154:** 267-273

Youngman EM, He SL, Nikstad LJ, Green R (2007) Stop codon recognition by release factors induces structural rearrangement of the ribosomal decoding center that is productive for peptide release. *Mol Cell* **28:** 533-543

Yu YT (2006) The most complex pseudouridylase. *Structure* **14:** 167-168

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science* **292:** 883-896

Zagorski J, Tollervey D, Fournier MJ (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

Zebarjadian Y, King T, Fournier MJ, Clarke L, Carbon J (1999) Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Mol Cell Biol* **19:** 7461-7472

Zeytuni N, Zarivach R (2012) Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* **20:** 397-405

Zhang X, Champion EA, Tran EJ, Brown BA, Baserga SJ, Maxwell ES (2006) The coiled-coil domain of the Nop56/58 core protein is dispensable for sRNP assembly but is critical for archaeal box C/D sRNP-guided nucleotide methylation. *RNA* **12:** 1092-1103

Zhao R, Davey M, Hsu YC, Kaplanek P, Tong A, Parsons AB, Krogan N, Cagney G, Mai D, Greenblatt J, Boone C, Emili A, Houry WA (2005) Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell* **120:** 715-727

Zhao R, Kakihara Y, Gribun A, Huen J, Yang G, Khanna M, Costanzo M, Brost RL, Boone C, Hughes TR, Yip CM, Houry WA (2008) Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. *J Cell Biol* **180:** 563-578

Zhou J, Liang B, Li H (2011) Structural and functional evidence of high specificity of Cbf5 for ACA trinucleotide. *RNA* **17:** 244-250

Etudes structurales des interactions protéines-protéines et ARN-protéines impliquées dans l'assemblage des snoRNP à boîtes C/D.

De nombreuses fonctions cellulaires essentielles telles que la traduction, l'épissage, la biogenèse des ribosomes et la réplication des télomères font appels aux particules RNP non codantes. La biogenèse de ces dernières chez les eucaryotes est un processus très complexes qui fait intervenir de nombreux facteurs cellulaires. La biogenèse du ribosome nécessite au moins 150 facteurs. Ceux-ci sont importants pour faciliter mais également contrôler la biogenèse de cette machinerie cellulaire essentielle qu'est le ribosome. Parmi ces facteurs, nous avons les snoRNP à boîtes C/D. Ces RNP sont impliqués dans la maturation des pré-ARNr (méthylation post-transcriptionnelle des riboses et clivages endonucléolytiques). Récemment notre laboratoire a participé à la découverte de facteurs d'assemblage de ces RNP. Il s'agit entre autre des protéines Rsa1p, du complexe R2TP (Rvb1p, Rvb2p, Tah1p et Pih1p) et de Hit1p chez la levure *Saccharomyces cerevisiae*. En utilisant une approche de co-expression à haut débit, nos travaux ont révélé un réseau complexe d'interactions entre les protéines constitutives des snoRNP et leurs facteurs d'assemblage. Couplé à une stratégie de protéolyse ménagée, la co-expression nous a permis d'obtenir des sous-complexes protéiques Snu13p/Rsa1p et Rsa1p/Hit1p qui font actuellement l'objet d'une étude structurale par RMN. En collaboration avec l'équipe de F. Allain (ETH Zurich), nous avons également déterminé la structure tridimensionnelle de la protéine Tah1p, ainsi que de son complexe avec le peptide C-terminale de la protéine chaperonne Hsp90 à haute résolution. Ces travaux ont révélé un mode d'association particulier entre le domaine TPR de la protéine et le peptide.

Mots clés : *Saccharomyces cerevisiae*, snoRNP, Snu13p, Rsa1p, Hit1p,Tah1p, Hsp90, RMN, co-expression, protéolyse ménagée.

Structural studies on protein-protein and RNA-protein interactions implicated in the C/D snoRNP biogenesis.

A lot of essential cellular functions like translation, splicing, ribosome biogenesis and telomere replication need the activity of non coding RNPs. The biogenesis of non coding RNPs in eukaryotes is a complex pathway involving numerous cellular factors. For instance, ribosome biogenesis requires more than 150 factors. They are important to facilitate and to control the biogenesis of this essential cellular machinery. These factors include the C/D box snoRNPs. These RNPs are involved in pre-rRNA maturation (post-transcriptional ribose methylation and endo-nucleolytic cleavages). Recently, our laboratory participated to the discovery of snoRNP assembly factors: the Rsa1p protein, R2TP complex (Rvb1p, Rvb2p, Tah1p and Pih1p) and Hit1p in the yeast *Saccharomyces cerevisiae*. Using a high throughput co-expression approach, we deciphered a network of interactions between RNP core proteins and the assembly factors. Coupled with a limited proteolysis strategy, the co-expression method allowed us to obtain proteins sub-complexes Snu13p/Rsa1p and Rsa1p/Hit1p which are currently studied by NMR. In collaboration with the F. Allain team (ETH Zurich), we also determined the tridimensional structure of protein Tah1p and its complex with the chaperon Hsp90 C-terminal peptide at high resolution. The data obtained reveal a particular mode of association of the Tah1p TPR domain with the Hsp90 peptide.

Keyword: *Saccharomyces cerevisiae*, snoRNP, Snu13p, Rsa1p, Hit1p, Tah1p, Hsp90, NMR, co-expression, limited proteolysis.