



HAL
open science

Décryptage des mécanismes moléculaires d'assemblage de complexes macromoléculaires ARN/protéine : la snoRNP C/D comme modèle de travail

Sana Dermouche

► **To cite this version:**

Sana Dermouche. Décryptage des mécanismes moléculaires d'assemblage de complexes macromoléculaires ARN/protéine : la snoRNP C/D comme modèle de travail. Sciences du Vivant [q-bio]. Université de Lorraine, 2022. Français. NNT : 2022LORR0110 . tel-03851830

HAL Id: tel-03851830

<https://hal.univ-lorraine.fr/tel-03851830v1>

Submitted on 9 Sep 2024

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

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



**UNIVERSITÉ
DE LORRAINE**

**BIBLIOTHÈQUES
UNIVERSITAIRES**

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 bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

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>

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 **Sana DERMOUCHE**

Décryptage des mécanismes moléculaires d'assemblage de complexes macromoléculaires ARN/protéine : la snoRNP C/D comme modèle de travail

Le 22 juin 2022

Membres du jury :

Rapporteurs :	Mme Claudine MAYER	Professeure, Université Paris Cité, Paris
	Mme Christine ALLMANG	Directrice de Recherche, CNRS, Strasbourg
Examineurs :	Mme Sarah CIANFÉRANI	Directrice de Recherche, CNRS, Strasbourg
	Mme Natacha DREUMONT	HDR, Inserm, Nancy
	M. Xavier MANIVAL	Directeur de Recherche, CNRS, Nancy Directeur de thèse
	M. Stéphane LABIALLE	Maître de Conférence, Université de Lorraine, Nancy Co-directeur de thèse
Membres invités:	M. Bruno CHARPENTIER	Professeur, CNRS, Nancy

**Laboratoire IMoPA: Ingénierie Moléculaire et Physiopathologie Articulaire
UMR 7365 CNRS-UL, Faculté de Médecine, Biopôle de l'Université de Lorraine,
9 Avenue de la Forêt de Haye, CS 50184, 54505 Vandoeuvre-lès-Nancy**

Remerciements

Je tiens à exprimer en premier lieu mes sincères remerciements à Xavier Manival et Stéphane Labialle pour avoir dirigés mon projet de thèse avec beaucoup de bienveillance. vos conseils, vos encouragements et nos échanges scientifiques, m'ont énormément aidé dans l'accomplissement de mon travail.

Je remercie chaleureusement Claudine Mayer, Christine Allmang, Natacha Dreumont et Sarah Cianférani, de faire partie de mon jury de thèse. Mes remerciements vont également à Bruno Charpentier pour m'avoir accueillie au sein de l'équipe 1 au laboratoire IMoPA lors de mes stages de Master 1 puis Master 2. Je garde de très beaux souvenirs de vos cours sur les snoRNP durant mon année de Master 1, à travers lesquels j'ai appris à aimer ce sujet de recherche.

Je remercie également tous les collaborateurs de notre groupe qui ont toujours gentiment accepté d'échanger sur mes travaux de recherche et contribuer à son avancement. Je pense particulièrement à Evolène Deslignière, Célia Plisson-Chastang, Deniz Ugurlar et Boxue Ma.

Mes vives remerciements vont à l'ensemble du personnel de laboratoire IMoPA et en particulier les personnes m'ayant formés à mes débuts au laboratoire, Hélène Marty-Capelle pour ses conseils et sa bonne humeur au quotidien, Arnaud Paul, Marc Quinternet et Marie-Eve Chagot, j'ai beaucoup appris à vos côtés.

J'ai une pensée particulière pour les anciens thésard qui sont passés par le laboratoire, Laeya, Liza et Géraud, votre écoute et nos longues discussions ont nourri en moi une philosophie de vie alliant vie professionnelle et personnelle. Je vous souhaite à tous de vous épanouir pleinement dans vos différents parcours.

L'accomplissement de ce long travail n'aurait pas été possible sans le soutien en toute épreuve de ma famille, ma maman qui a cru en moi et m'a permis de dépasser mes peurs, faisant de moi la femme que je suis aujourd'hui. Ma sœur Imène et mon frère Amine, vous avez toujours été mes exemples de réussite. Naila notre petite ange qui veille sur nous depuis là-haut. Un remerciement particulier pour mon cousin et mon grand frère Chouki, promis je viendrai vivre plus au sud à côté de la mer.

Je remercie aussi l'ensemble de ma belle-famille qui n'ont cessé de prier pour ma réussite. Merci à ma belle-maman Djamila, mon beau-père Fethi, Ikram, Abderahmane et Omar.

Un immense merci à Nadir, mon merveilleux époux et mon fidèle ami dont la présence et le soutien à mes côtés au quotidien a été sans faille. Il a fait preuve de beaucoup de compréhension et d'imagination pour que je puisse surmonter les moments difficiles au cours de ce travail de thèse.

Enfin j'aimerais dédier ce doctorat à deux personnes qui m'ont donné la force d'avancer peu importe la difficulté. Mon papa qui nous a quitté trop tôt pour pouvoir assister à cet accomplissement, tu es sans doute très fier de moi de là où tu es, je le sais car tu l'as toujours été à chacune de mes réussites. Ma fille qui va venir au monde dans quelques mois et qui m'aura accompagnée au quotidien, particulièrement durant la rédaction de ma thèse, avec tes petits coups dans mon ventre pour me rappeler que tu es là et que je ne suis pas seule. J'ai très hâte de te voir enfin !

*Liste des
Abréviations*

Liste des Abréviations

°C : degré Celsius	Kd : Constante de dissociation
2YT : 2 <i>Yeast extract and Tryptone</i>	K-turn : <i>Kink-turn</i>
Å : Angström	LB : <i>Lysogeny Broth</i> (ou Luria Bertani)
AAA+ : <i>ATPases Associated with diverse cellular Activities</i>	LSU : <i>Large SubUnit</i>
ADN : Acide Désoxyribonucléique	MHz : MégaHertz
ADP : Adénosine Di-Phosphate	min : minute
APS : PerSulfate d'Ammonium	MRP : <i>Mitochondrial RNA Processing</i>
ARN : Acide RiboNucléique	MTase : Méthyltransférase
ARNm : ARN messenger	Nop : <i>Nucleolar protein</i>
ARNr : ARN ribosomique	NOESY : <i>Nuclear Overhauser Effect Spectroscopy</i>
ARNt : ARN de transfert	NUFIP1 : <i>Nuclear FMRP Interacting Protein-1</i>
ATP : Adénosine TriPhosphate	nt : nucléotide
Bcd1p : <i>Box C/D snoRNA-1</i>	PAGE : <i>PolyAcrylamid Gel Electrophoresis</i>
CAF-1 : <i>Chromatin Assembly Factor-1</i>	PEG : <i>PolyEthylene Glycol</i>
ChIP : <i>Chromatin Immunoprecipitation</i>	pI : point Isoélectrique
CK2 : <i>Casein Kinase 2</i>	PM : Poids Moléculaire
Cryo-EM : <i>Cryo-electron microscopy</i>	ORF : <i>Open Reading Frame</i>
CS : <i>CHORD and Sgt1</i>	pb : paire de base
Da/kDa : Dalton/kiloDalton	PCR : <i>Polymerase Chain Reaction</i>
DLS : <i>Dynamic Light Scattering</i>	qPCR : PCR quantitative
dNTP : désoxyriboNucléotide TriPhosphate	PDB : <i>Protein Data Bank</i>
DO : Densité Optique	PEI : PolyEthylèneImine
DTT : <i>DiThioThreitol</i>	pH : potentiel Hydrogène
g/L : gramme par Litre	Pih1p : <i>Protein Interacting with Hsp90-1</i>
g/mol : gramme par mole	ppm : parties par million
GAR : <i>Glycine/Arginine-Rich</i>	PTC : <i>Peptidyl-Transferase Center</i>
GTP : Guanosine TriPhosphate	PTM : <i>Post-Translational Modifications</i>
h : heure	R2TP : Rvb1p/Rvb2p/Tah1p/Pih1p
HAT : <i>Histones AcetylTransferases</i>	RMN : Résonance Magnétique Nucléaire
HEPES : acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique	RMSD : <i>Root-Mean-Square Deviation</i>
Hit1p : <i>High Temperature growth factor-1</i>	RNP : <i>RiboNucleoParticle</i> (Particule RiboNucléique)
HSP90 : <i>Heat Shock Protein 90 kDa</i>	rNTP : riboNucléotide TriPhosphate
HSP70 : <i>Heat Shock Protein 70 kDa</i>	RPAP3 : <i>RNA Polymerase-associated protein-3</i>
HSQC : <i>Heteronuclear Single-Quantum Correlation</i>	rpm : rotation par minute
IPTG : IsoPropyl β -D-1-ThioGalactopyranoside	Rsa1 : <i>Ribosom assembly-1</i>
ITC : <i>Isothermal Titration Calorimetry</i>	RSC : <i>Remodel the Structure of Chromatin</i>
	Rtt106 : <i>Regulator of Ty1 Transposition 106</i>
	Rvb : <i>RuVB-like</i>
	SAM : S-Adénosyl-(L)-Méthionine
	scaRNA : <i>small Cajal Bodies RNA</i>
	SDS : <i>Sodium DodecylSulfate</i>

SILAC : *Stable Isotope Labeling with Amino acids in Cell culture*
snRNA : *small nuclear RNA*
snoRNA : *small nucleolar RNA*
snoRNP : *small nucleolar RNP*
Snu13p : *Small Nuclear ribonucleoprotein associated-13*
SSU : *Small SubUnit*
TAF9 : *Transcription Activating Factor 9*
Tah1p : *TPR containing protein Associated with HSP90-1*
TCEP : *Tris-(2-CarboxyEthyl)Phosphine*
TE : *Tris-EDTA*
TEMED : *TEtraMethylEthylèneDiamine*
TRASH : *Trafficking, Resistance and Sensing of Heavy Metals*
TIP48/49 : *TATA box-binding protein-interacting protein of 48/49 kDa*
TPR : *TetratricoPeptide Repeat*
Tris : *Trishydroxyméthylaminométhane*
U : *Uridine*
UV : *UltraViolet*
YNB : *Yeast nitrogen base*
YPD : *Yeast extract Peptone Dextrose*
YPG : *Yeast extract Peptone Galactose*
Ψ : *Pseudouridine*
ZHD : *zf-HIT domain*

Nomenclatures

- La première lettre des noms des protéines de levure est écrite en majuscule, avec le nom suivi de la lettre « p » pour protéine, sauf pour les enzymes et les protéines précédées du mot « protéine », elles ne prennent pas de « p » à la fin.
- Les complexes protéiques sont écrits avec un « / » séparant les protéines le constituant, par exemple : Rvb1/Rvb2.
- Les mots en anglais sont écrits en italique.

*Table des
Matières*

Table des Matières

Introduction

Chapitre 1 : Biogenèse des ribosomes et maturation des ARNr	1
1. Clivages et modifications chimiques des précurseurs d'ARNr	3
1.1. Organisation des gènes codant les précurseurs d'ARNr	3
1.2. Clivages nucléolytiques des pré-ARNr chez la levure <i>S. cerevisiae</i>	5
1.3. Modifications chimiques post-transcriptionnelles	6
1.3.1. La 2'- <i>O</i> -méthylation des résidus d'ARNr	8
1.3.2. La pseudouridylation	10
1.3.3. Rôle et positionnement des modifications des résidus d'ARNr	11
1.3.4. Ribosomopathies et maladies associées	12
2. Les snoRNP : acteurs essentiels dans le processus de maturation des ARNr	16
2.1. Les snoRNP à boîtes C/D	16
2.1.1. snoARN à boîtes C/D	16
a. Transcription et maturation	16
b. Organisation et structure tridimensionnelle (3D)	18
2.1.2. Les protéines cœur de la particule snoRNP C/D	21
a. La protéine Snu13p/SNU13 (15.5K)	21
b. La 2'- <i>O</i> -méthylase Nop1/Fibrillarine	24
c. Les protéines Nop56/NOP56 et Nop58/NOP58	26
2.1.3. Structure 3D d'une particule sRNP C/D d'archée	29
2.2. Les snoRNP à boîtes H/ACA	29
2.2.1. Structure d'ARN H/ACA : Structure 3D et organisation	31
2.2.2. Protéines cœur de la snoRNP à boîtes H/ACA	31
Chapitre 2 : Processus d'assemblage des particules snoRNP C/D fonctionnelles chez la levure	33
1. Décryptage de la machinerie d'assemblage des snoRNP C/D et de ses acteurs protéiques	33
1.1. Le complexe protéique Rsa1-Hit1/NUFIP1-ZNHIT3	33
1.2. La protéine Bcd1	34
1.2.1. Domaines et organisations structurales	36
1.2.2. Rôle dans les étapes d'assemblage de la snoRNP à boîtes C/D	39
1.2.3. Bcd1p et les mécanismes de rétrotransposition	41
1.2.4. Bcd1p et la protéine chaperon d'histones Rtt106	41
1.3. Le complexe protéine chaperon/co-chaperon Hsp82/HSP90-R2TP	42
1.3.1. La protéine chaperon Hsp82/HSP90	42
1.3.2. Le complexe protéique cofacteur des protéines chaperon HSP90/R2TP	45
1.3.2.1. Les ATPases Rvb1/Rvb2	48
a. Mécanismes cellulaires associés	48
b. Structures 3D et organisations quaternaires	51

1.3.2.2.	La protéine Tah1	58
1.3.2.3.	La protéine Pih1	60
1.3.3.	Le complexe R2SP chez l'homme, homologue du R2TP	61
2.	Modèle d'assemblage des snoRNP à boîtes C/D chez la levure <i>S. cerevisiae</i>	63

Objectifs

Objectifs	65
-----------	----

Matériels et méthodes

1.	Matériels	67
1.1.	Oligonucléotides	67
1.2.	Vecteurs d'expression	67
1.2.1.	Plasmides bactériens	67
1.2.2.	Plasmides de levure	67
1.3.	Souches utilisées	72
a.	Souches bactériennes : <i>Escherichia coli</i>	72
b.	Souches de levures : <i>Saccharomyces cerevisiae</i>	72
1.4.	Les milieux de cultures	73
1.5.	Les antibiotiques	74
1.6.	Les tampons utilisés	74
a.	<i>Escherichia coli</i>	74
b.	<i>S. cerevisiae</i>	74
1.7.	Les solutions	75
1.8.	Les gels d'analyses	75
1.9.	Les anticorps	76
2.	Méthodes	76
2.1.	Stratégies de génie génétique	76
a.	La PCR (<i>Polymerase Chain Reaction</i>)	76
b.	Le clonage	76
c.	Le sous-clonage	77
d.	La transformation bactérienne	77
e.	La mutagenèse dirigée	78
2.2.	Biochimie des protéines recombinantes	78
a.	Synthèse des protéines recombinantes	78
b.	Purification des protéines recombinantes	78
c.	La technique de co-élution	80
2.3.	Etude biophysique des protéines et complexes protéiques	80
a.	<i>Isothermal Titration Calorimetry</i> (ITC)	80
b.	<i>Dynamique Light Scattering</i> (DLS)	81
c.	<i>Bio-Layer Interferometry</i> (BLI)	81
2.4.	Etude structurale par Cryo-microscopie électronique	82
2.5.	Transformation des levures compétentes	82
2.6.	Système d'expression avec un promoteur inductible	82

2.7.	Préparation de l'extrait cellulaire total de levure	83
2.8.	Extraction d'ARN totaux de levure	83
2.9.	Méthodes de détection des protéines et des ARNs	84
a.	Stratégie de <i>Western Blot</i>	84
b.	RT-PCR	85
c.	PCR quantitative	85
2.10.	Méthodes d'analyse des complexes protéiques et ARN-protéines	85
a.	Méthode de co-immunoprécipitation	85
b.	Méthode d'immunoprécipitation d'ARN (RIP)	86

Résultats et discussion

Chapitre 1 : Le facteur d'assemblage Bcd1p est un partenaire du complexe ATPasique Rvb1/Rvb2 chez la levure **87**

1.	Etude du complexe formé entre Rvb1/Rvb2 et Bcd1p	87
1.1.	L'état oligomérique du complexe Rvb1/Rvb2 varie selon son état nucléotidique	87
1.2.	Reconstitution du complexe Rvb1/Rvb2/Bcd1p	93
1.2.1.	Co-élution sans nucléotide	93
1.2.2.	Co-élution en présence d'ADP ou d'ATP	95
1.2.3.	Caractérisation biophysique du complexe Rvb1/Rvb2/Bcd1p	98
a.	Diffusion dynamique de la lumière (<i>Dynamique Light Scattering</i> , DLS)	98
b.	Titration calorimétrique isotherme (ITC)	98
1.2.4.	La présence de l'ADP dans les sites actifs de Rvb1 et Rvb2 est nécessaire à la formation du complexe Rvb1/Rvb2/Bcd1p	101
2.	Etude structurale du complexe Rvb1/Rvb2/Bcd1p	103

Chapitre 2 : Analyse de la contribution du facteur Bcd1p dans l'assemblage de la particule snoRNP à boîtes C/D **108**

1.	L'interaction Bcd1p-Snu13p est essentielle pour l'assemblage des snoRNP C/D	108
2.	Mise en place au laboratoire pour l'étude des complexes d'assemblage formés par Bcd1p D72A	110
3.	L'expression de Bcd1p D72A affecte la stabilité des snoARN à boîtes C/D transcrits par l'ARN polymérase II	112
4.	Bcd1p D72A bloque l'assemblage en amont du recrutement de Nop58p sur la pré-snoRNP à boîtes C/D	112
5.	L'association de Bcd1p avec Snu13p est perturbée par la mutation D72A	113

Chapitre 3 : Etudes structurales des domaines TPR1 et TPR3 de SPAG1 et caractérisation de leurs modes d'interaction avec HSP70 et HSP90 **116**

1.	SPAG1 est une plateforme au recrutement des HSP	116
2.	La protéine SPAG1 possède-elle une activité GTPase ?	117
	Publication N°01: <i>Binding properties of the quaternary assembly protein SPAG1</i>	119

Publication N°02: *Optimizing the first TPR domain of the human SPAG1 protein provides insight into the HSP70 and HSP90 binding properties* 135

Conclusion et perspectives

1. Caractérisation de l'interaction du complexe Rvb1/Rvb2 et Bcd1p de *S. cerevisiae* 150
2. Une mutation ponctuelle D72A sur Bcd1p perturbe l'assemblage des snoRNP à boîtes C/D chez la levure *S. cerevisiae* 152
3. Caractérisation d'un complexe R2TP-like : le R2SP 157

Références bibliographiques

Références bibliographiques 158

Introduction

Chapitre 1 : Biogenèse des ribosomes et maturation des ARNr

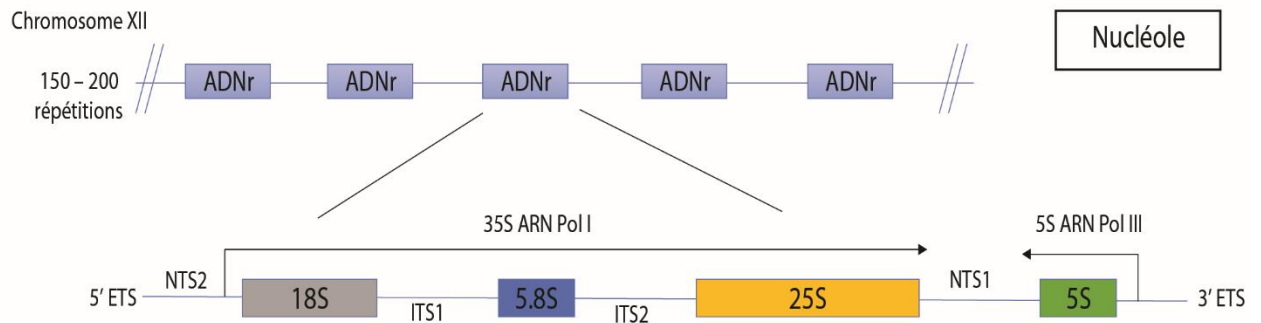
La production continue de protéines dans les cellules repose sur l'efficacité d'une machinerie complexe nommée le ribosome. Il assure dans le cytoplasme la traduction fidèle du message génétique en protéine par l'assemblage de chaque acide aminé apporté par les ARN de transfert (ARNt). L'information génétique nécessaire à la synthèse de la protéine est portée par l'ARN messager (ARNm). La production des protéines est donc assurée par cette machinerie énergivore (selon le type de cellule : 50 -1000 mmol d'ATP pour 1 g de protéines synthétisées (Pannevis and Houlihan, 1992), permettant de stimuler la croissance cellulaire. Chez toutes les espèces, les ribosomes sont constitués de deux sous-unités ribonucléoprotéiques : une grande sous-unité, responsable de la formation de la liaison peptidique, incluant trois ARN ribosomiques (ARNr), les ARNr 5S, 5,8S et 25S/28S (Levure/Homme), ainsi qu'une petite sous-unité qui assure le décodage des ARNm, constituée de l'ARNr 18S.

Des études chez la levure et chez les eucaryotes de façon plus large, ont permis d'élucider les mécanismes sous-jacents au contrôle élaboré de la biogenèse des ribosomes qui nécessite une régulation coordonnée des trois ARN polymérase (I, II et III) (Mayer and Grummt, 2006). L'ARN polymérase II (ARN pol II) est à l'origine de la transcription des gènes codants pour les protéines ribosomiques ainsi que de la quasi-totalité des facteurs impliqués en *trans* dans ce processus. L'ARN polymérase III (ARN pol III) permet la synthèse du précurseur de l'ARNr 5S et des ARNt, alors que l'ARN polymérase I assure la production du précurseur des ARNr 5,8S, 18S et 25S/28S (Figure 1A).

Durant la biogenèse des ribosomes, les pré-ARNr s'assemblent avec des protéines ribosomiques mais également avec de nombreux facteurs transitoires pour former des particules pré-ribosomiques (Figure 1B). Lors de la formation de ces complexes dans le nucléole, une série de modifications chimiques spécifiques ainsi qu'une succession de clivages endonucléolytiques sont catalysés pour la plupart par des particules snoRNP (*small nucleolar RiboNucleoprotein Particle*), permettant la formation des ARNr mature. Ils participent de façon plus large à la biogenèse des ribosomes (Henras et al., 2015; Watkins and Bohnsack, 2012) Les snoRNP sont des particules constituées d'un ARNnc (non codant) et de protéines cœur responsable de la réaction catalytique de la particule.

La transcription de l'ARN ribosomique est finement ajustée au statut énergétique cellulaire et liée à la production des protéines ribosomiques. Nous pouvons citer la régulation de la

A



B

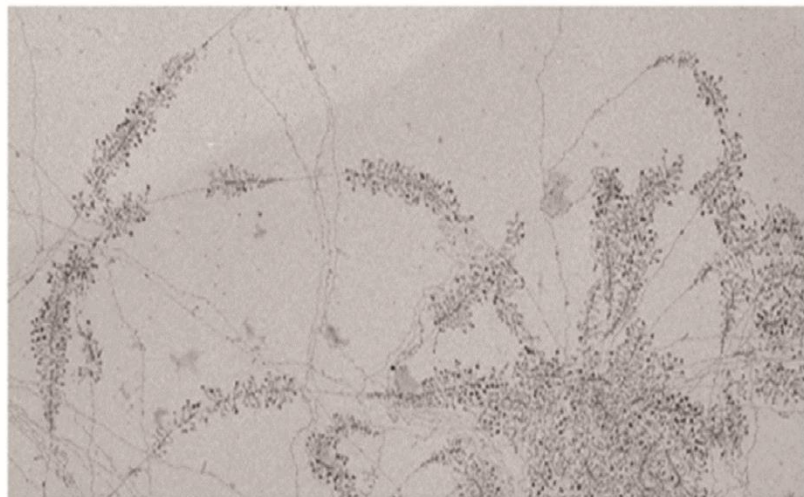


Figure 1 : (A) Organisation du locus ADNr chez *S. cerevisiae*. Les répétitions d'ADNr (150-200) sont situées sur le chromosome 12. Une seule unité répétée est transcrite par l'ARN polymérase I (ARN pol I) pour synthétiser le transcrit pré-ARNr primaire 35S qui sera traité pour produire le 18S, 5.8S mature, et des ARNr 25S, et par l'ARN polymérase III (ARN pol III) pour synthétiser l'ARNr 5S. NTS: *Nontranscribed spacer* ; ETS: *External transcribed spacer*; ITS: *Internal transcribed spacer* (Woolford and Baserga, 2013). (B) Contenus nucléolaires analysés par microscopie électronique. La transcription des unités d'ADNr répétées peut être visualisée comme des « arbres de Noël », où le tronc de l'arbre est l'ADNr, les branches sont l'ARNr et les ornements sont les boutons aux extrémités 5' qui contiennent la petite sous-unité (SSU= *Small SubUnit*) du processome (Woolford and Baserga, 2013).

transcription du pré-ARNr 35S par le complexe TOR1 (*Target Of Rapamycin 1*) (Mayer & Grummt, 2006). Il a été montré que le manque de nutriment inhibe les voies de signalisation qui requièrent la cible de la rapamycine mTOR (TOR1 et TOR2 chez les mammifères et la levure) (Gingras et al., 2001). La fixation de TOR1 au niveau du promoteur de l'ADNr permet d'activer la transcription et pourrait alors assurer une adaptation du taux de production des ribosomes en fonction des conditions environnementales. Des mécanismes contrôlant divers aspects de la biogenèse des ribosomes ont été découverts et révèlent de nouvelles connexions au contrôle du cycle cellulaire et de la taille des cellules (Lempiäinen and Shore, 2009). La régulation de la biogenèse du ribosome est maintenant décrite comme étant étroitement liée à certaines maladies, en particulier au cancer, à l'anémie et au vieillissement. Des études récentes ont réussi à approfondir et élucider les mécanismes moléculaires sous-jacents à ces connexions (Penzo et al., 2019). Parmi ces mécanismes, il a été montré que la surexpression dans les cellules cancéreuses de l'agent oncogène Myc, responsable du contrôle des étapes de la biogenèse du ribosome : *i*) augmente l'activité de la pol I en facilitant le recrutement de SL1 sur les promoteurs, *ii*) stimule la biogenèse des protéines ribosomiques par l'augmentation de la transcription de l'ARN pol II, *iii*) accélère la transcription de l'ARN pol III à travers l'activation du facteur d'initiation de la transcription de la polymérase III B (TFIIIB) (Jenq and van den Brink, 2010; White, 2005).

1. Clivages et modifications chimiques des précurseurs d'ARNr

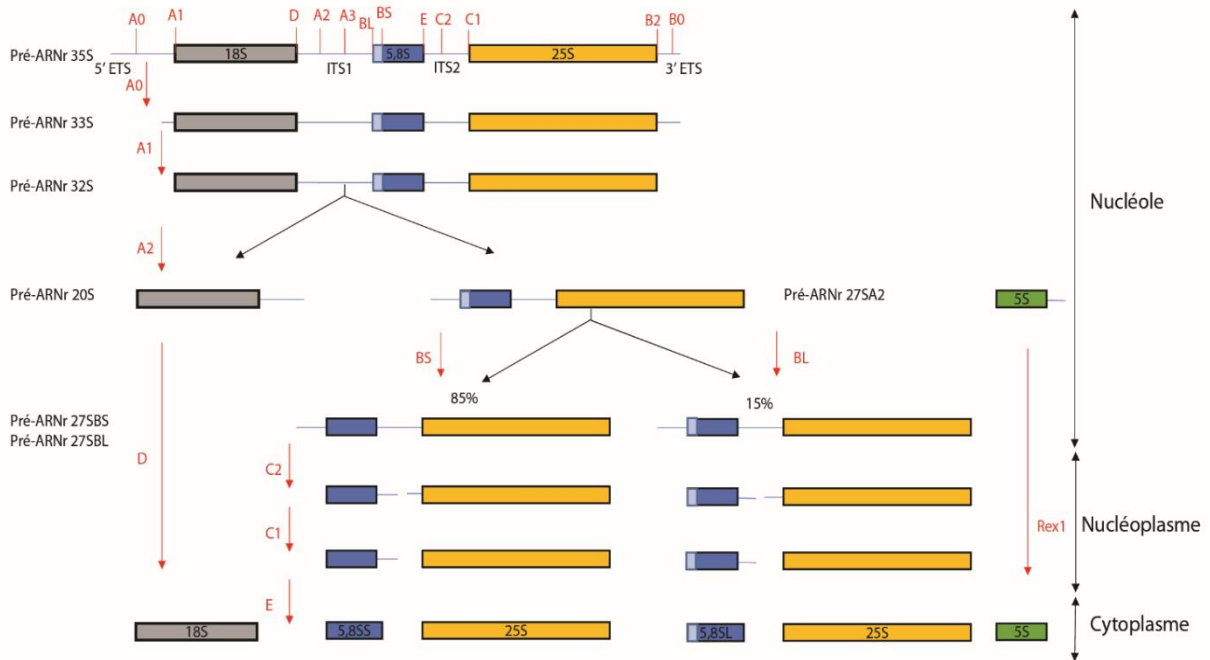
1.1. Organisation des gènes codant les précurseurs d'ARNr

Les ARNr 18S, 5,8S et 25/28S sont issus d'une même unité transcriptionnelle (Figure 1A), transcrite par l'ARN polymérase I au niveau du nucléole. Il en résulte un transcrit unique appelé pré-ARNr 35S chez la levure (Figure 2A) et pré-ARNr 45S chez l'homme.

Chez l'homme, des répétitions des unités transcriptionnelles codant les ARNr 18S, 5,8S et 28S sont réparties sur 5 paires de chromosomes distincts (les chromosomes 13, 14, 15, 21 et 22), tandis que les copies du gène codant l'ARNr 5S sont retrouvées sur le chromosome 1 (Sylvester et al., 1986). Chez les mammifères, l'ARNr 5S est transcrit dans le nucléoplasme, puis importé dans le nucléole pour former la future sous-unité ribosomique 60S.

Nous retrouvons chez la levure les segments d'ADNr localisés sur le chromosome XII sous la forme de répétitions en tandem (Petes, 1979) (Figure 1A). Les deux unités transcriptionnelles (ARN polymérases I et III) sont physiquement rapprochées, puisqu'elles sont regroupées deux

A



B

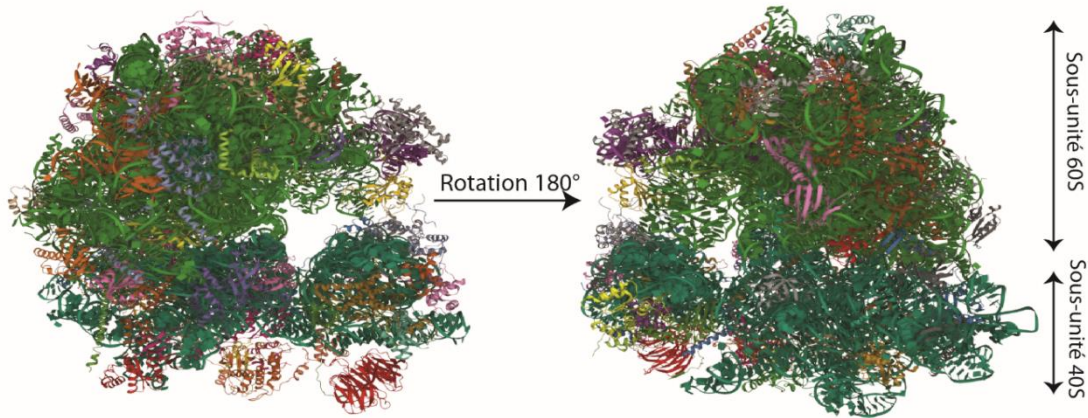


Figure 2 : (A) Voie de traitement du pré-ARNr de levure. Le pré-ARNr 35S est transcrit par l'ARN polymérase I et contient des séquences pour les ARNr 18S, 5.8S et 25S flanqués et séparés par des espaceurs transcrits externes et internes (ETS et ITS). L'ARNr pré-5S est transcrit par l'ARN polymérase III. Les séquences d'espaceur sont retirées du pré-ARNr par la série indiquée d'étapes de traitement endonucléolytique et exonucléolytique, au sein des ribosomes d'assemblage. Chaque site de traitement est indiqué. Le traitement commence dans le nucléole de la cellule, mais des étapes ultérieures se produisent dans le nucléoplasme et le cytoplasme. Notez que dans les cellules à division rapide, la majorité du pré-ARNr subit un clivage co-transcriptionnel aux sites A0, A1 et A2 avant que la transcription ne soit terminée (Adapté de Woolford and Baserga, 2013). (B) Structure 3D obtenue par cristallographie à une résolution 3.0 Å. Les ARNr des deux sous-unités sont représentés en vert, les protéines ribosomiques sont indiquées en différentes couleurs (Ben-Shem et al., 2011) ; code PDB : 4V88).

à deux. Dans cette configuration, la transcription de l'ARNr 5S par l'ARN polymérase III s'effectue directement au sein du nucléole (Haeusler & Engelke, 2006) (Figure 2A).

1.2. Clivages nucléolytiques des pré-ARNr chez la levure *S. cerevisiae*

La maturation du transcrit pré-ARNr chez *S. cerevisiae* débute par trois clivages successifs au niveau du nucléole. Le clivage de la séquence 5'-ETS aux deux sites A0 puis A1 présents dans cette région conduit à la libération de deux produits intermédiaires (Hughes and Ares, 1991), respectivement les pré-ARNr 33S et 32S (Beltrame et al., 1994). Il est suivi par un clivage au site A2 conduisant à la libération de l'ARN intermédiaire 20S (futur ARNr 18S), et du pré-ARNr 27SA2 qui donnera les futurs ARNr 25S et 5,8S (Beltrame et al., 1994; Udem and Warner, 1972) (Figure 2A).

Le pré-ARNr 27SA2 est mûri selon deux processus différents. La voie majoritaire consiste en un clivage au site A3 dans la région dite ITS1, catalysée par la RNase MRP (*Multiple RNase P*) (un ribozyme) (Henry et al., 1994 ; Lygerou et al., 1996) ayant pour produit l'intermédiaire 27SA3. Ce dernier sera ensuite dégradé au niveau de son extrémité 5' par les deux exonucléases Rrp17p (*Ribosomal RNA processing protein 17*) et Rat1p (*Ribonucleic acid-trafficking protein 1*) donnant lieu à l'intermédiaire 27SBS (Oeffinger et al., 2009) (Figure 2A). Tandis que la voie minoritaire, consiste en un clivage direct du 27SA2 par une endonucléase de la RNase MRP donnant naissance à l'intermédiaire 27SBL. La maturation se poursuit par un clivage au site C2, conduisant à la formation du pré-ARNr 7S et 26S. La séquence ITS2 présente en 5' du 26S est clivée par Rat1p, formant enfin l'ARNr 25S (Geerlings et al., 2000). L'intermédiaire 7S poursuit des étapes de clivage, donnant lieu à l'ARNr 5,8S mature. L'intermédiaire 20S est exporté du nucléole vers le cytoplasme, où il va connaître un clivage au site D à son extrémité 3' catalysé par l'endonucléase Nob1p (*Nin1 (One) Binding protein*), afin d'aboutir à la naissance de l'ARNr 18S mature (Fatica et al., 2003; Lamanna and Karbstein, 2009). L'ARNr 5S subit une maturation au niveau de son extrémité 3' au sein du nucléole par l'endonucléase Rex1 (*RNase exonuclease 1*), avant d'être transporté dans le cytoplasme pour constituer la grande sous-unité (LSU) 60S du ribosome (Lee and Nazar, 1997) (Figure 2).

La maturation des pré-ARNr a lieu au site de transcription où des études montrent la présence de larges complexes moléculaires, nommés les petites sous-unités du processome (Figure 1B), constitués de facteurs protéiques et de complexes ribonucléoprotéiques, ces derniers étant constitués d'un ARN non codant et de protéines (Kressler et al., 2010). Une analyse de ces complexes révèle la présence notamment de la snoRNP U3 (Dragon et al., 2002). Par

appariement de bases avec le pré-ARNr, la particule U3 va servir de guide permettant une meilleure accessibilité des sites A0, A1 et A2 (Kass et al., 1990). Nous pouvons noter la présence d'une autre snoRNP (Beltrame et al., 1994; Yuan et al., 2003), la snoRNP U14 qui possède également des fonctions de clivage de pré-ARNr aux sites A1 et A2, notamment chez la *Drosophile*, l'Homme et *S. cerevisiae* (Huang et al., 1992).

1.3. Modifications chimiques post-transcriptionnelles

En plus des clivages d'ARN précurseurs, les ARN ribosomiques de tous les organismes subissent d'autres modifications post-transcriptionnelles comme les introductions du groupement méthyle et des conversions de certains résidus uridine en pseudouridine (Figure 3A). L'introduction de groupements méthyle se produit à des sites spécifiques de la molécule. Les groupes méthyle dans les ARNr bactériens sont principalement attachés à diverses positions dans les bases puriques et pyrimidiques (méthylation des bases) et dans les ARNr eucaryotes principalement attachés à la position 2'-O des riboses (méthylation du ribose). Nous retrouvons 109 modifications d'ARNr chez *S. cerevisiae* (Decatur and Fournier, 2002). Les ARNr ribosomiques 18S et 25S contiennent respectivement 33 et 74 modifications chacun. Les ARNr 5,8S et 5S ne comportent qu'une seule modification chacun (Piekna-Przybylska et al., 2007). Les ARNr bactériens contiennent environ 0,7 groupement méthyle pour 100 nucléotides (Dubin and Günalp, 1967; Hayashi et al., 1966), tandis que chez les eucaryotes, ce chiffre varie de 1,2 pour la levure (Klootwijk and Planta, 1973) à 1,7 pour les cellules HeLa (Maden and Salim, 1974). La teneur en pseudouridine est de 0,25 résidu pour 100 nucléotides dans l'ARNr bactérien (Dubin and Günalp, 1967), de 0,8 dans l'ARNr de levure (Klootwijk and Planta, 1973) et 1,2 dans l'ARNr des cellules HeLa (Jeanteur et al., 1968).

Sur les 63 snoARN différents de *S. cerevisiae*, 20 constituent la famille des boîtes H/ACA et le reste fait partie de la famille des boîtes C/D. Ensemble, ils sont impliqués dans la plupart des modifications d'ARNr, dont 23 conversions en pseudouridine (Figure 3B) et 54 méthylations du ribose d'ARNr (Figure 3C). Certains snoARN agissent sur deux ou trois sites différents (Bonnerot et al., 2003). Le rôle des snoARN dans la modification de l'ARNr a d'abord été reconnu lorsque des « séquences guides internes » ont été identifiées pour être complémentaires aux régions du pré-ARNr, couvrant le nucléotide cible (Bachelierie and Cavaillé, 1997).

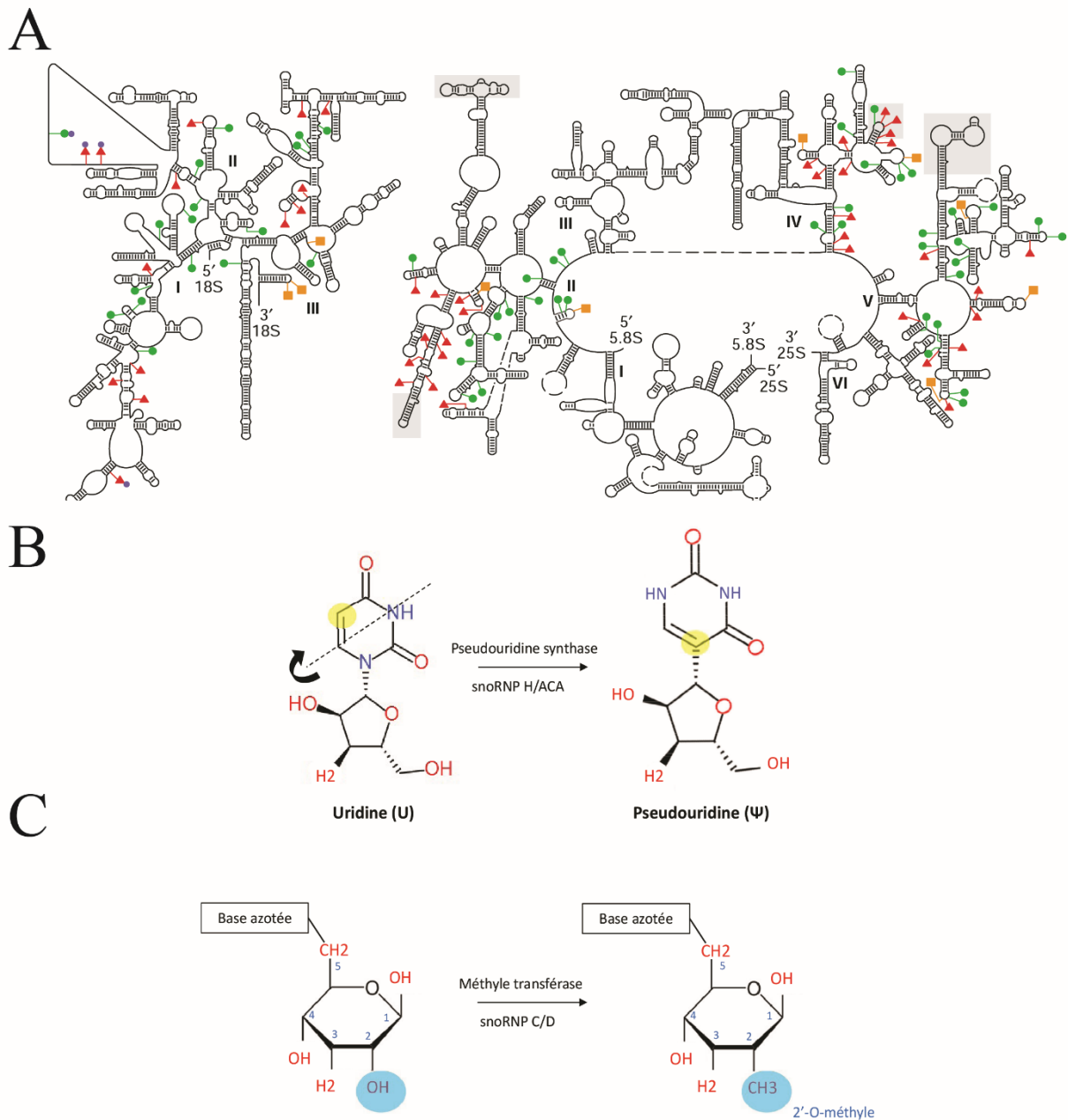


Figure 3 : Représentation des modifications chimiques les plus retrouvées dans les ARNr eucaryotes (adapté de (Decatur and Fournier, 2002)). (A) Structure secondaire des ARNr 18S (gauche) 25S et 5.8S (à droite) de levure. Les points verts indiquent les modifications de 2'-O-méthylation, les triangles rouges représentent les sites de pseudouridylation, et les méthylation de base sont indiquées en carré orange. (B) Pseudourylation des uridines en pseudouridines par une isomérisation suivant l'axe représenté en pointillé. Cette réaction est catalysée par des snoRNP H/ACA. (C) 2'-O-méthylation du ribose, catalysée par les snoRNP C/D.

Toute modification post-transcriptionnelle de l'ARNr demande une contribution énergétique considérable en termes de nombre d'enzymes nécessaires au traitement des molécules précurseurs et à l'introduction de résidus modifiés. Plus de 100 protéines transitoires sont impliquées dans la modification de l'ARNr et l'assemblage des ribosomes. Au moins six groupes sont identifiés avec les nucléases, les enzymes modificatrices, les hélicases, les GTP/ATPases, les protéines chaperon, les protéines d'assemblage ainsi que les facteurs de transport (Henras et al., 2015).

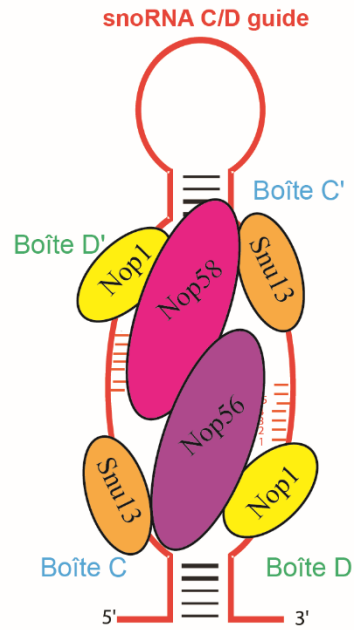
1.3.1. La 2'-O-méthylation des résidus d'ARNr

La méthylation de l'ARN à la position 2'-O-ribose (Figure 3C) est un type de modification très répandu. Elle est retrouvée également sur l'ARNt, l'ARNm et le snoARN (*small nuclear RNA*). Ces méthylations peuvent affecter le repliement 3D des ARN permettant alors une meilleure stabilité, ainsi que leur interaction avec des ligands. Cette modification a lieu majoritairement sur les pré-ARNr naissants au cours de leur transcription. La méthylation du ribose est strictement limitée aux régions constituant les futurs ARNr matures. Elles sont généralement retrouvées sur des séquences d'ARNr conservées au cours de l'évolution. Alors que *Xenopus laevis* possède un modèle de méthylation d'ARNr identique à celui des mammifères, les levures et les vertébrés partagent 31 sites communs de méthylation au niveau de l'ARNr sur un total de 55 sites de méthylation chez la levure et 46 sites pour les vertébrés (Ayadi et al., 2019). Les procaryotes quant à eux, sont pauvres en méthylation des riboses d'ARNr, avec par exemple seulement 4 sites chez *Escherichia coli* (un sur l'ARNr 16S et trois sur l'ARNr 23S) (Edward and Maden, 1998).

La majeure partie des nucléotides modifiés se situe au niveau des ARNr 18S et 25S (Piekna-Przybylska et al., 2007). Chez la levure *S. cerevisiae*, les ARNr 25S, 5,8S et 18S contiennent respectivement 74, 1 et 35 modifications. L'ARNr 5S comporte une modification chimique de conversion d'une uridine en pseudouridine, catalysée par l'enzyme Pus7p (Decatur and Schnare, 2008; Piekna-Przybylska et al., 2007).

La 2'-O-méthylation est catalysée par les petites particules ribonucléoprotéiques dites snoRNP à boîtes C/D, qui seront décrites en détail plus loin dans ce chapitre. Le snoARN à boîtes C/D qui est un des constituants de ces particules, sert de guide pour la catalyse de la position exacte sur le substrat. Il s'agit du ribose du nucléotide du pré-ARNr apparié avec le 5^{ème} nucléotide en

A



B

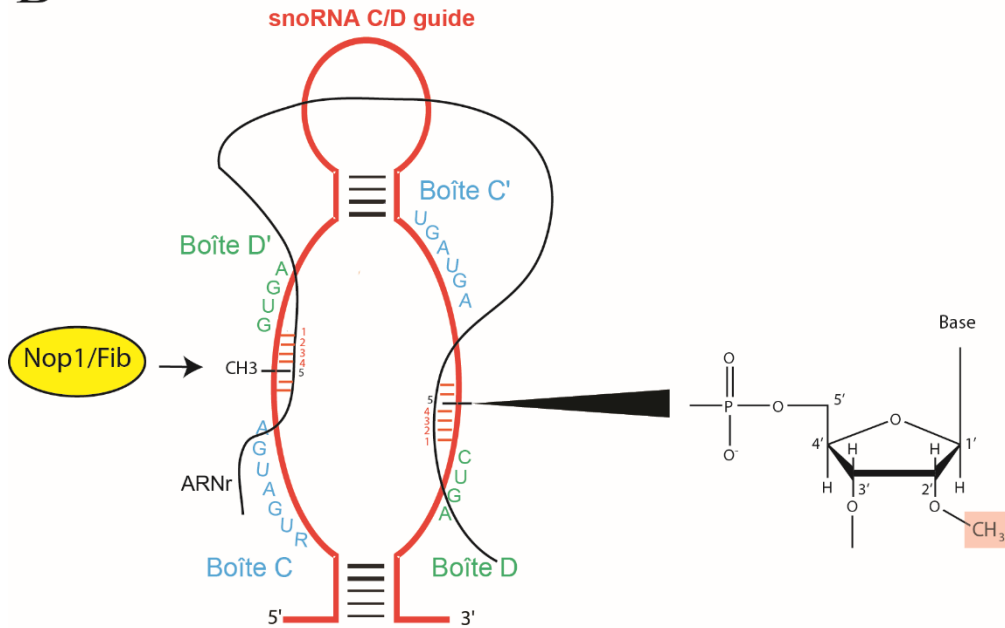


Figure 4 : La particule snoRNP à boîtes C/D

(A) Schéma d'une snoRNP à boîtes C/D composée d'un snoARN à boîtes C/D guide et des quatre protéines cœur constitutives : Snu13p, Nop1/Fibrillarine, Nop58p et Nop56p. (B) Schéma d'appariement snoARN : ARNr et méthylation de base. Méthylation d'un ARNr, représenté en noire, par un snoARN à boîtes C/D en rouge, au niveau du 5^{ème} nucléotide en amont des boîtes D et D'. Les séquences des boîtes C/D et C'/D' sont indiquées en vert et bleu.

amont des boîtes D et D' (Filipowicz et al., 1999; Kiss-László et al., 1996; Weinstein and Steitz, 1999). Un duplex est alors formé entre le snoARN guide/pré-ARNr (Figure 4B), impliquant les nucléotides situés en amont des boîtes D et D'. Cet hétéroduplex s'organise en double-hélice dans l'espace (Cavaillé et al., 1996; Kiss-László et al., 1996), facilitant l'accès de la méthylase Nop1/Fibrillarine au ribose substrat. Une large partie des snoRNP à boîtes C/D (Figure 4A) chez les eucaryotes sont responsables de la modification des ARNr. Cependant, elles sont décrites pour également modifier les snRNA (U2, U6 et U5) et la maturation de certains pré-ARNm (Maxwell and Fournier, 1995; Terns and Terns, 2002). Une récente étude indique la présence de certaines 2'-O-méthylation en faible abondance sur certains ARNm (Ayadi et al., 2019). Cette liste peut également être étendue pour les archées où les sRNP (*small RNP*) modifient des ARNt (Omer et al., 2000). Cela indique une origine ancestrale des snoRNP C/D qui serait responsable de plusieurs voies de maturation des ARN. Une étude révèle que la perte des sites méthylés chez *S. cerevisiae* affecte sa capacité d'adaptation aux conditions de stress. Cependant les mécanismes sous-jacent restent à ce jour inconnus (Esguerra et al., 2008).

1.3.2. La pseudouridylation

Dans les ARNr humains, plus de 100 nucléotides sont méthylés en position 2'-O-hydroxyle et environ 95 uridines sont converties en pseudouridine (Ofengand and Bakin, 1997; Ofengand et al., 1995). De manière intéressante, ces modifications sont regroupées au centre du ribosome dans les régions fonctionnelles des ARNr et elles montrent une conservation évolutive remarquable. Chez *E. coli*, le seul résidu pseudouridine de l'ARNr 16S (Wrzesinski et al. 1995a) et au moins une des neuf pseudouridines de l'ARNr 23S (Wrzesinski et al. 1995b) sont formés par des pseudouridines synthétases seules et spécifiques. Cependant, il semble peu probable que chacun des nombreux résidus pseudouridine des ARNr eucaryotes soit formé par sa propre synthétase spécifique. Il était devenu évident que le nucléole, en plus des snoARN C/D, contient une autre famille distincte de snoARN qui est défini par une structure secondaire universelle « *hairpin-hinge-hairpin-tail* » (Ganot et al., 1997). Ces snoARN forment des petites particules ribonucléoprotéiques connues sous le nom de snoRNP à boîtes H/ACA. Leur composition ainsi que leur organisation dans l'espace seront décrites plus loin dans ce chapitre. La réaction de pseudouridylation est catalysée par ces snoRNP H/ACA. Des travaux ont permis de mieux comprendre le mécanisme de mise en place de cette réaction d'isomérisation : les facteurs de pseudouridylation des mammifères reconnaissent de courtes séquences ribosomiques situées autour de l'uridine cible. Les snoARN H/ACA humains et de levure possèdent des complémentarités avec ces séquences ribosomiques signal (Ganot et al., 1997).

Les deux hélices qui forment le duplex snoARN-pré-ARNr entourent deux nucléotides non appariés dans la séquence d'ARNr. Le premier nucléotide non apparié (dans le sens 5'→3') est un résidu uridine qui est toujours converti en pseudouridine. Alors que les ARN guides de méthylation possèdent de longues complémentarités ininterrompues avec les séquences d'ARNr, les ARN guides de pseudouridylation contiennent deux courts motifs complémentaires, dont aucun ne pourrait à lui seul fournir des informations suffisantes pour sélectionner sans ambiguïté un site unique dans le pré-ARNr. Ces deux motifs courts de reconnaissance des snoARN à boîtes H/ACA sont réunis par une structure tige-boucle pour créer un signal de reconnaissance bipartite statistiquement significatif (Ganot et al., 1997). Les composants des snoRNP H/ACA jouent un rôle crucial dans la sélection et la catalyse du site précis de pseudouridylation ribosomique dans le pré-ARNr. En premier lieu, la pseudouridine synthétase Cbf5 chez la levure *S. cerevisiae* qui est responsable de la catalyse enzymatique, aidée par les protéines cœurs et le snoARN H/ACA qui permettent un accès au nucléotide à modifier. L'absence transitoire de certaines modifications de pseudouridine ribosomiques n'a pas de phénotype détectable dans des souches de levure (Parker et al., 1988). De plus, une souche mutante dépourvue de cinq snoARN non essentiels, parmi lesquels trois ARN guides de pseudouridylation (snR3, snR5 et snR8) responsables de la synthèse de quatre pseudouridines, ont des caractéristiques de croissance similaires à celles des cellules sauvages (Parker et al. 1988). Ces observations sont cohérentes avec l'idée que les pseudouridines ribosomiques pourraient subtilement améliorer les fonctions ribosomiques qui peuvent également se produire en leur absence, bien qu'avec une efficacité légèrement inférieure. Néanmoins, ces pseudouridines ribosomiques conféreraient dans leur pleine globalité vraisemblablement des avantages sélectifs importants au cours de l'évolution (Ganot et al., 1997).

1.3.3. Rôle et positionnement des modifications des résidus d'ARNr

Les propriétés chimiques des nucléotides modifiés n'indiquent pas à priori de rôles fonctionnels spécifiques. Un certain nombre de changements structuraux peut se produire. Ces effets comprennent des propriétés stériques altérées dans tous les cas, un potentiel de liaison hydrogène différent, une amélioration de l'empilement local des bases constituées de pseudouridine, une stabilité de la liaison phosphodiester pour la méthylation des riboses et une rigidité structurale accrue pour les régions simple et double brins (Arnez and Steitz, 2002; Auffinger and Westhof, 1998; Charette and Gray, 2000). La méthylation du 2'-OH du ribose a principalement deux conséquences. Elle favorise la configuration C3'-endo du ribose plus stable

thermodynamiquement et empêche l'hydrolyse des liaisons entre nucléotides (Williams et al., 2001). Dans chaque cas, les effets structuraux et thermodynamiques dépendent de l'environnement et peuvent s'étendre au-delà du site de modification. La distribution des modifications dans l'ARNr n'est pas aléatoire (Figure 5), particulièrement dans la LSU. Dans les deux sous-unités, la plupart des modifications se produisent dans des régions conservées et donc fonctionnellement importantes. Cette corrélation indique que les modifications influencent à la fois la structure et la fonction du ribosome (Bakin et al., 2002; Brimacombe et al., 1993; Lane et al., 1995; Smith et al., 2002). Bien que les effets sur la production de ribosomes doivent être pris en compte, les arguments en faveur d'un rôle dans la traduction même sont les plus importants comme pour les modifications du domaine V dans la LSU. Ce domaine est connu depuis longtemps pour être lié à l'activité peptidyl-transférase et à la liaison à l'ARNt (Nissen et al., 2000), et est particulièrement riche en modifications.

La plupart des modifications se produisent à l'intérieur des ARNr, à l'interface des sous-unités. Les régions riches en modifications comprennent plusieurs sites où des événements de traduction se produisent et où les sous-unités interagissent (Decatur and Fournier, 2002). Les modifications sont sensiblement absentes des régions enrichies par les protéines ribosomiques : les surfaces externes et la périphérie des régions d'interface (Ban et al., 2000; Schlutzen et al., 2000; Wimberly et al., 2000). La présence des modifications de 2'-O-méthylation est étroitement liée à la fonction du ribosome. Une étude montre que la baisse de ces modifications a un impact sur la traduction d'un groupe spécifique d'ARNm, mais aussi à la plus grande sensibilité à différents antibiotiques qui ont pour cible le blocage de la traduction (Gillot-Chafia, 2018).

1.3.4. Ribosomopathies et maladies associées

Les ribosomopathies regroupent les pathologies causées par des défauts de la biogenèse des ribosomes. Chez l'homme, ces anomalies génétiques sont la conséquence de défauts de production et de fonction des ribosomes. Il s'agit pour la plupart d'entre elles d'anomalies médullaires, où la quantité de cellules sanguines est sensiblement altérée, ainsi qu'une forte prédisposition au développement de cancer (Freed et al., 2010; Narla and Ebert, 2010). Une corrélation entre ces pathologies et des mutations recensées dans le génome, en particulier sur les gènes codant les protéines ribosomiques et des facteurs transitoires de biogenèse, a pu être identifiée. Parmi ces troubles physiologiques nous retrouvons l'anémie de Diamond-Blackfan, le syndrome de Shwachman-Diamond, le syndrome de Bowen-Conradi, l'hypoplasie du

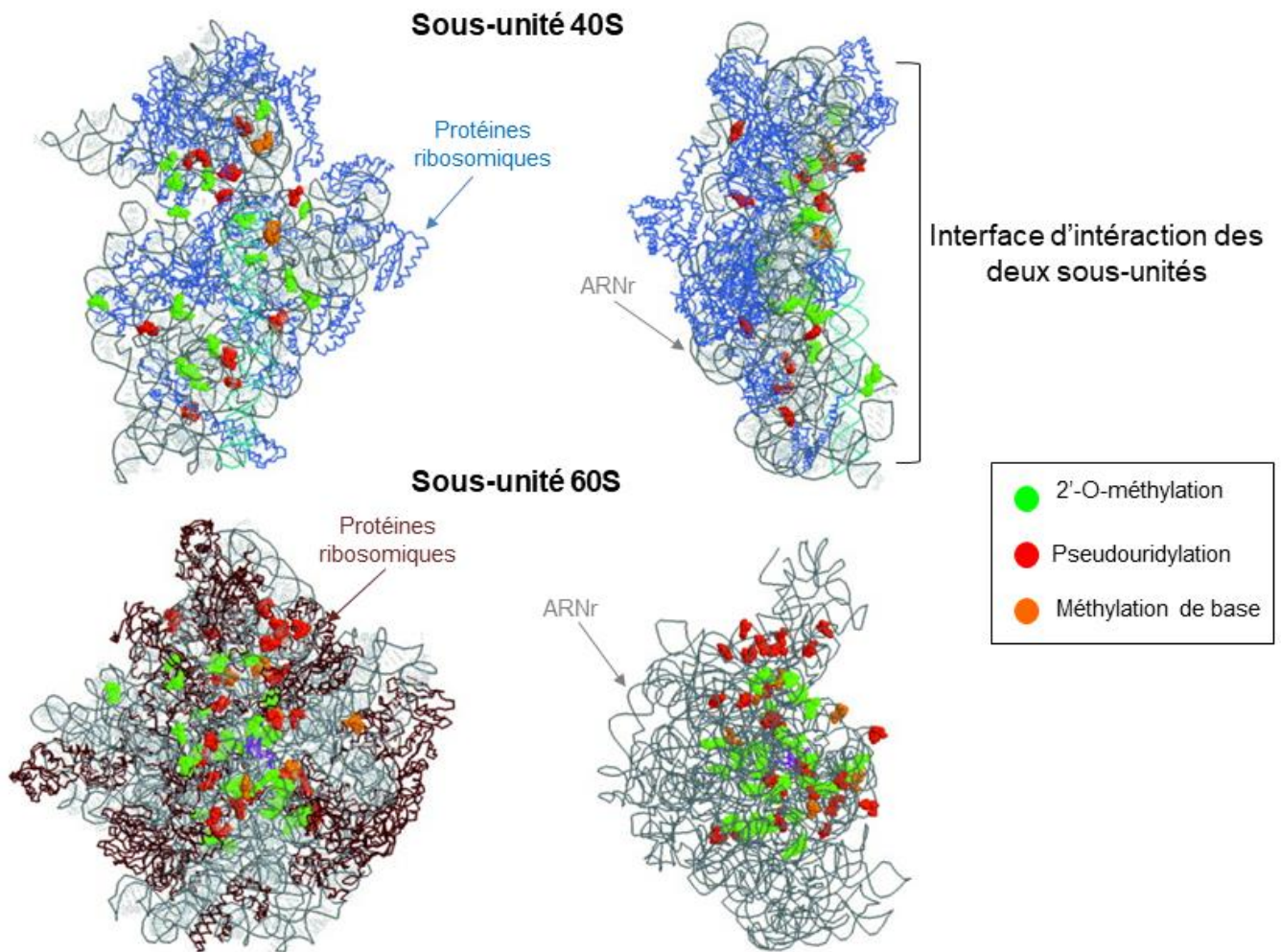


Figure 5 : Représentation des sites de modification des nucléotides d'ARNr sur le ribosome de levure.

Positions des modifications sur la structure 3D de la petite sous-unité (en haut) et de la grande sous-unité (en bas). Les modifications sont identifiées par trois couleurs (rouge, vert et orange). Deux vues des deux sous-unités sont représentées avec une rotation de 90° pour chacune (d'après (Decatur and Fournier, 2002b)).

cartilage et des cheveux, ou encore la dyskératose congénitale. A titre d'exemple, l'anémie de Diamond-Blackfan est le plus souvent (40% des patients) caractérisée par des mutations sur plusieurs gènes codant des protéines ribosomiques (RP), tels que RPL5, RPL11 ou RPL36 (Choismel et al., 2008; Ganapathi et al., 2007; Gazda et al., 2008; Robledo et al., 2008). Le syndrome de Bowen est, quant à lui, la résultante de mutations dans un gène codant une protéine transitoire à la biogenèse des ribosomes, EMG1 (*Essential protein for Mitotic Growth 1*) chez l'homme dont l'homologue chez la levure est Emg1p. Elle est responsable de la biogenèse de la sous-unité 40S (Armistead et al., 2009).

Une autre pathologie complexe connue sous le nom de syndrome de Prader-Willi (Figure 6) a été montré comme la résultante de deux phénomènes qui convergent vers la même pathologie. Pour la plupart des patients, il s'agit d'une délétion partielle des gènes du chromosome 15 chez l'homme, au niveau d'une région soumise à une régulation par empreinte génomique parentale. La région du génome délétée présente une douzaine de gènes non codants des protéines, qui codent plusieurs snoARN à boîtes C/D dont le SNORD116 qui est chez la plupart des patients atteints, toujours absent dans le cas de délétions. Il peut être présent dans les cas de disomies uniparentales maternelles ou les très rares cas d'épimutation du centre de contrôle de l'empreinte (présent mais transcriptionnellement inactif dans ces deux cas) (Ohta et al., 1999) (Figure 6) (Cavaillé et al., 2000). Il arrive que certaines personnes héritent uniquement des deux chromosomes 15 maternels (disomie uniparentale maternelle (matUPD)). Ils développent un syndrome de Prader-Willi à la naissance ou à l'âge adulte, caractérisé par une hypotonie musculaire, hyperphagie qui se développe en obésité morbide chez les enfants avec des retards de croissance facilement observables cliniquement.

Contrairement aux pathologies citées précédemment, il a été suggéré qu'une hypermodification d'ARNr pouvait conduire à une prolifération de cellules cancéreuses (Lafontaine, 2015). En effet, dans les cancers de poumon, le snoARN SNORD42 qui est un snoARN H/ACA est surexprimé. La diminution de son expression *in vitro* a permis de freiner l'évolution de la tumeur en activant des mécanismes d'apoptose (Mei et al., 2012). Dans les cellules cancéreuses du sein et du côlon, l'observation d'une corrélation inverse entre les niveaux d'expression du suppresseur de tumeur p53 et FBL (Fibrillarine) a conduit à suggérer que p53 réprime la transcription de FBL par liaison directe à son premier intron (Marcel et al., 2013). D'autres travaux rapportent une forte expression de FBL dans de nombreux cancers, soutenant la prolifération cellulaire (Su et al., 2014).

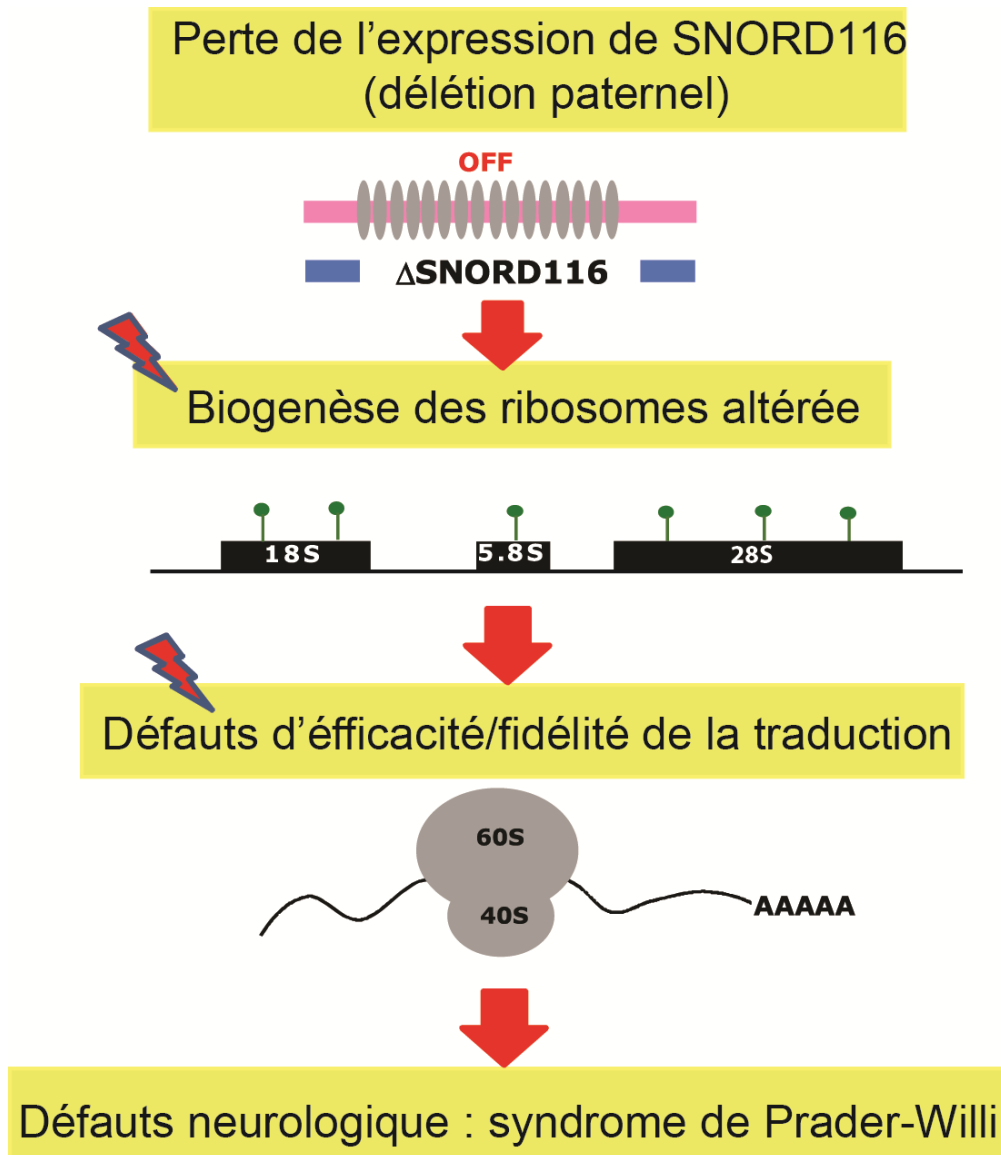


Figure 6 : Représentation schématique explicative du syndrome de Prader-Willi (adapté de Cavallé, 2017)

2. Les snoRNP : acteurs essentiels dans le processus de maturation des ARNr

Les nucléoles eucaryotes contiennent un grand nombre de petits ARN métaboliquement stables connus collectivement sous le nom de petits ARN nucléolaires (snoARN) (Bachellet et al., 1995; Maxwell and Fournier, 1995). Seulement quelques 150 espèces de snoARN ont été recensées dans les cellules humaines au début de leurs études (Lafontaine and Tollervey, 1998). Des approches et des stratégies génomiques à haut débit plus robustes ont conduit à la découverte de nombreux nouveaux membres et à une diversité surprenante de la fonctionnalité et des mécanismes d'action du snoARN (Bergeron et al., 2020; Bogard et al., 2021).

Il a été montré que ces snoARN appartiennent à deux classes structurellement et fonctionnellement distinctes (Balakin et al., 1996; Ganot et al., 1997; Lafontaine and Tollervey, 1998; Tollervey and Kiss, 1997). Nous retrouvons les ARN à boîtes C/D et les ARN à boîtes H/ACA. Au sein de chaque grande famille de snoARN, deux groupes fonctionnels distincts peuvent être discernés. Ceux requis pour le clivage du précurseur pré-ARNr comme nous l'avons décrit précédemment, avec par exemple le snR30 (snoARN H/ACA), U3 et U14 (snoARN C/D) (Hughes & Ares, 1991). Ces snoARN sont essentiels pour assurer la viabilité cellulaire. Le second groupe constitué de la grande majorité des snoARN, fonctionne comme des ARN guides pour la modification covalente notamment du pré-ARNr et sont requis pour la croissance.

Les membres de chaque classe de snoARN sont associés à des composants protéiques communs dans des petites particules ribonucléoprotéiques nucléolaires ou snoRNP.

Les protéines formant la snoRNP H/ACA ont été les premières à être décrites : chez la levure, Gar1p (Balakin et al., 1996; Ganot et al., 1997), Cbf5p (Lafontaine et al., 1998), Nhp2p, et Nop10p (Henras et al., 1998 ; Watkins et al., 1998). Une brève présentation de la composition et la structure des snoRNP à boîtes H/ACA sera faite à la fin de ce chapitre.

Les quatre protéines qui composent les snoRNP C/D sont également connues : il s'agit chez la levure de Snu13p (Nottrott et al.; 1999), Nop58p, Nop56p (Lafontaine and Tollervey, 1999) et Nop1/Fibrillarine (Gautier et al., 1997).

2.1. Les snoRNP à boîtes C/D

2.1.1. snoARN à boîtes C/D

a. Transcription et maturation

Les gènes des snoARN présentent une dichotomie d'organisation inhabituelle. Un groupe est organisé de manière classique dont la séquence codante est flanquée d'éléments du promoteur,

terminateur et amplificateur. Le deuxième groupe est tout à fait unique, avec la séquence codante du snoARN intégrée dans les introns des gènes codant pour des protéines. Ces séquences codantes introniques sont retrouvées sous forme de clusters issus de duplications locales.

La plupart des snoARN à boîtes C/D chez les vertébrés et la levure sont présents au niveau des régions introniques du génome. La transcription de ces snoARN est donc reliée et étroitement dépendante de la transcription du gène hôte, impliquant la maturation du pré-ARNm. Ce dernier code des protéines qui interviennent dans la biogenèse du ribosome, ou dans certaines fonctions nucléolaires. De récentes études ont également montré que l'expression des gènes hôtes peut être découplée de celle du snoARN, dans le but de répondre aux différents niveaux d'abondance de snoARN et aux besoins fonctionnels de certains tissus humains hautement régulés, où ils sont enrichis (Fafard-Couture et al., 2021).

Ce nouvel arrangement génétique suggère un moyen potentiel pour réguler le métabolisme de l'ARN nucléaire et coordonner la production de ribosomes avec d'autres activités cellulaires.

La transcription des gènes codants les snoARN est assurée par l'ARN pol II chez les vertébrés et la levure. Toutefois, une exception existe pour le snoARN52 de *S. cerevisiae* qui est transcrit par la machinerie de l'ARN pol III (Guffanti et al., 2006). La taille de la majorité des snoARN à boîtes C/D varie de 50 à 70 nucléotides. Il existe également d'autres snoARN C/D plus long, tel que le snoARN U3 qui fait 217 nucléotides chez l'homme et le SNORD116 qui fait 96 nucléotides.

Une maturation des snoARN en 3' est nécessaire pour l'assemblage de snoRNP à boîtes C/D fonctionnelles. Il a été montré que cette maturation caractérisée par un clivage du snoARN est concomitante avec l'assemblage pour la snoRNP (Hirose et al., 2006). En effet, certaines protéines cœur constitutives de la snoRNP mature, les protéines humaines SNU13 et Fibrillarine, sont retrouvées en association au complexe du spliceosome C1 de façon précoce (Hirose et al., 2003). Cela permet également de protéger le snoARN des dégradations exo- et endo-nucléolytiques (Weinstein and Steitz, 1999).

La maturation des snoARN est indépendante du gène hôte. Ils possèdent leurs propres signaux essentiels pour une maturation précise (Kiss and Filipowicz, 1995). Le processus commence par une libération de l'intron, du fait de l'action de l'ARN hélicase Aquarius ou IBP160 (*Intron Binding Protein 160*) présente chez l'homme. Son rôle est fondamental dans la libération des snoARN sous forme de lasso (Hirose et al., 2006) mais aussi à leur repliement correct favorisant le recrutement des protéines cœur. Chez *S. cerevisiae*, ce rôle est assuré par l'enzyme Dbr1 (Petfalski et al., 1998). Par la suite, l'action d'exonucléases telles que Rat1 ou Xrn1 à

l'extrémité 5' (Petfalski *et al.*, 1998) et de l'exosome à l'extrémité 3' (Mao *et al.*, 1995; Shibagaki *et al.*, 1995; Tsukamoto *et al.*, 1997) (Allmang *et al.*, 1999; Terns *et Terns*, 2002) est nécessaire pour produire un snoARN correctement maturé à ses extrémités. Cette maturation se fait également par des attaques exonucléolytiques des extrémités 5' et 3'. La maturation s'achève après l'action de la protéine Rrp6 (*Ribosomal RNA processing 6*) indépendamment de l'exosome (Allmang *et al.*, 1999). Chez les eucaryotes, le processus de maturation des snoARN se poursuit par l'ajout d'une coiffe m⁷G (7-méthylguanosine) à l'extrémité 5' libre, sur la position 7 de la guanine. Cette méthylation concerne uniquement les snoARN produits à partir d'unité transcriptionnelle indépendante. Cette action est assurée chez la levure par 3 enzymes différentes : Cet1 (*Capping Enzyme Triphosphatase 1*), Ceg1 (*Capping Enzyme Guanylyltransferase 1*) et Abd1 (méthyltransférase) (Mao *et al.*, 1995; Shibagaki *et al.*, 1995; Tsukamoto *et al.*, 1997).

Cette coiffe est éliminée chez la levure pour les snoARN C/D à l'exception de U3 et snR4/snR13. Une maturation nucléolytique de l'extrémité 5' est réalisée par l'enzyme Rnt1, via une reconnaissance des structures tige-boucle caractéristiques des régions 5' des pré-snoARN à boîtes C/D (Grzechnik *et al.*, 2018).

Tous les snoARN transcrits par l'ARN pol II subissent les mécanismes de terminaison de transcription (une exception est faite pour le snR52 qui lui est transcrit par l'ARN pol III). De façon très surprenante, des événements d'épissage ont lieu sur le snoARN U3 de levure, conduisant à l'élimination d'introns. Ce phénomène a lieu chez *S. cerevisiae*, *N. crassa* et *M. grisea* et, est absent chez l'homme et la majorité des organismes. Une étude réalisée au laboratoire, montre que l'épissage du U3 a lieu avant la maturation de l'extrémité 3' du snoARN U3 (Rothé *et al.*, 2017).

b. Organisation et structure tridimensionnelle (3D)

Les snoARN à boîtes C/D sont constitués de deux séquences consensus antisens nommées boîte C et boîte D. La boîte C est retrouvée au niveau de l'extrémité 5' du snoARN et comprend la séquence 5'-RUGAUGA-3'. La boîte D est située à l'autre extrémité du snoARN et est caractérisée par la séquence 5'-CUGA-3' (Bachellerie *et al.*, 2002) (Figure 4). Un second couple de boîtes dit C'/D' est retrouvé dans la région centrale du snoARN. Ces dernières sont dégénérées comparées aux séquences consensus (Kiss-László *et al.*, 1998).

L'appariement des motifs conservés « RUGAUGA » et « CUGA » situés aux extrémités impose le repliement du snoARN sur lui-même, formant une boucle interne asymétrique à 5+2 nucléotides, du fait que la boîte C soit plus longue que la boîte D (Figure 7A). Cela impose une

conformation spatiale particulière connue sous le nom de *K-turn* (*Kink-turn*) faisant référence à la courbure de 60° que présente le squelette phosphodiester (Watkins et al., 2000) (Figure 7). Le motif *K-turn* est constitué de deux hélices I et II qui entourent la boucle asymétrique, particulièrement riche en purine. Nous retrouvons dans cette boucle trois nucléotides non appariés et deux appariements non Watson-Crick (baptisées « *sheared* ») G•A et A•G. Ces éléments sont responsables de l'angle de 60° entre les deux hélices et de la protubérance du résidu uracile (Figure 7). Pour le snoRNA U14 l'angle passe à environ 30° plus caractéristique d'une conformation liée (Figure 7B). Ce motif est également stabilisé par des empilements de bases entre deux résidus de la boucle précédant le U et un résidu de chaque hélice (I et II). Le résidu U non apparié rejeté hors de la boucle forme des interactions hydrophobes avec les protéines de la famille L7Ae. La conformation particulière en C2'-*endo* de son ribose maximise l'exposition de son squelette phosphodiester (Chagot et al., 2019; Moore et al., 2004; Vidovic et al., 2000). Ce motif en *K-turn* (*K-loop*) est particulièrement présent chez les organismes hyperthermophiles, ce qui suggère un rôle de stabilisation de l'ARN face à des conditions de stress thermique. Cette structuration en *K-turn* n'est pas propre au snoARN C/D. Elle a été caractérisée initialement dans d'autres ARN tels que l'ARNr du ribosome (Klein et al., 2001), le snRNA U4 (Vidovic et al., 2000; Woźniak et al., 2005) et certains ARNm (Allmang et al., 2002; Mao et al., 1999; Winkler et al., 2001).

Comme déjà dit, le rôle du snoARN dans la fonction des snoRNP C/D est de guider précisément la position à méthyler. Un appariement snoARN C/D avec l'ARNr cible se crée 9 à 21 nucléotides en amont des boîtes D et D' créant un hétéroduplex par des interactions de type Watson-Crick. Le 5^{ème} nucléotide du snoARN C/D (Figure 4), en amont de la boîte D et D', sera apparié avec le nucléotide de l'ARN cible recevant la modification chimique, un hétéroduplex est alors formé par des interactions Watson-Crick (Kiss-László et al., 1996).

Chez les eucaryotes, un snoARN à boîtes C/D se différencie des autres par sa fonction mais aussi par sa taille et sa structure. Il s'agit du snoARN U3, c'est le plus long des snoARN à boîtes C/D (217 nucléotides chez l'homme et 333 nucléotides chez *S. cerevisiae*). Il comporte deux régions distinctes : la région 5' responsable des appariements avec les ARNr cibles qui est non structurée et la région 3' comprenant deux motifs structurés en *K-turn* pour le recrutement des protéines cœur de la snoRNP U3 (Jeppesen et al., 1988). Ces deux motifs sont formés par l'appariement des deux couples de boîtes : C'/D (dérivés du couple de boîtes C/D) et B/C (non canonique au couple de boîtes C/D) (Granneman et al., 2002; Marmier-Gourrier et al., 2003).

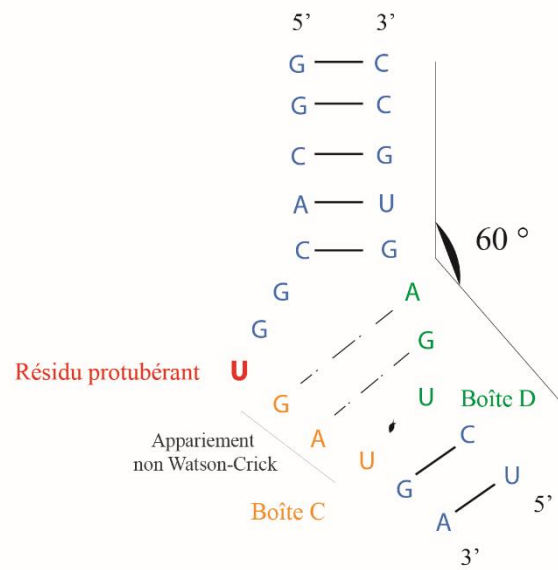
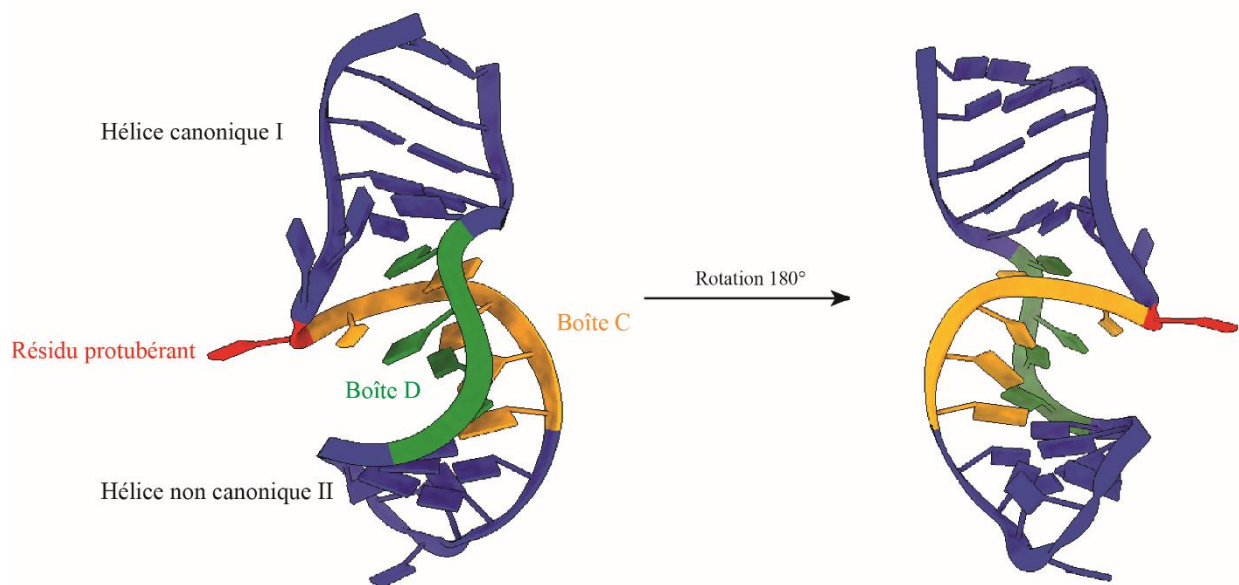
A**B**

Figure 7 : Motif en *K-turn* du snoARN à boîte C/D U14 de *S. cerevisiae*.

(A) Schématisation de la structure secondaire de la séquence représentative des boîtes consensus C et D, respectivement en orange et vert. Les liaisons hydrogènes sont représentées en noir, et les liaisons non Watson-Crick en traits discontinus. (B) Représentation de la structure tridimensionnelle du même motif en *K-turn* du snoARN U14, selon deux axes d'observation opposés. Code PDB : 6HYK (Chagot et al., 2019).

2.1.2. Les protéines cœur de la particule snoRNP C/D

En plus du snoARN à boîte C/D nécessaire pour guider précisément la position à modifier sur l'ARN cible, les snoRNP C/D sont également constituées de protéines (Levure/Homme, Figure 8) qui vont permettre sa structuration et la catalyse de la réaction de 2'-O-méthylation. Il s'agit de Snu13p/SNU13 (anciennement 15.5K chez l'homme) qui est la protéine de nucléation de l'assemblage des snoRNP C/D. Elle reconnaît le motif *K-turn* du snoARN de façon précoce lors de l'assemblage des particules. Nop56p/NOP56 (NOP=*Nucleolar Protein*) et Nop58p/NOP58 forment un hétérodimère homologue à l'homodimère de la protéine aNOP5 chez les archées. Enfin, la méthyltransférase Nop1/Fibrillarine catalyse la réaction de 2'-O-méthylation.

a. La protéine Snu13p/SNU13 (15.5K)

Initialement, il a été montré chez *Haloarcula marismortui* que cinq motifs en *K-turn* d'ARNr de la sous-unité 50S sont associés à des protéines ribosomiques. La plus connue est la protéine ribosomique L7 ainsi que ses protéines associées. Ces dernières forment une famille de *RNA-Binding-Protein* (Klein et al. 2001) qui inclut les protéines eucaryotiques et des archées suivantes : L7Ae, L30e et S12e (Koonin et al., 1997), les protéines de levure Nhp2 et Snu13 et la protéine humaine SNU13 (Nottrott et al. 1999).

La structure tridimensionnelle (3D) de la protéine humaine SNU13 a été la première de la famille des L7Ae à être résolue associée au snRNA U4 (Vidovic et al., 2000) (Figure 9A). La structure de la protéine Snu13 de levure *S. cerevisiae* a pu être résolue également quelques années plus tard (Oruganti et al., 2005). L'analyse de ces structures présentent des repliements identiques (RMSD de 0,474 Å) (Figure 9B), à savoir 5 hélices α et 4 brins β , dont l'arrangement 3D forme une topologie conservée dite en sandwich α - β - α et une architecture globale globulaire. Le motif constitué par les hélices α 2 et α 4 et le brin β 1 de SNU13 est responsable de l'interaction avec l'ARN (Figure 9A). Les points de liaison pour ce dernier concernent plus particulièrement le squelette ribose-phosphate des résidus de l'hélice II du motif en *K-turn*, ainsi que le résidu uridine (troisième résidu non apparié de la boucle interne) qui se retrouve projeté à l'extérieur de l'ARN. Il se retrouve logé dans une poche hydrophobe de la protéine Snu13 (SNU13), stabilisé par des liaisons hydrogènes avec des résidus électrostatiques essentiels (K61 et K86 chez l'homme, D54 et K79 chez les archées, E59 et R84 chez la levure) (Vidovic et al., 2000).

		<i>Levure</i>	<i>Homme</i>	<i>Archées</i>
<i>Protéines cœur constitutives de la particule snoRNP à boîtes C/D</i>		Snu13p	SNU13/15.5K	L7Ae
		Nop58p	NOP58	aNOP5
		Nop56p	NOP56	
		Nop1p	Fibrillarine	aNOP1
<i>Facteurs d'assemblages</i>	<i>Complexe R2TP</i>	Rvb1 (R1)	RuvBL1/Tip49	
		Rvb2 (R2)	RuvBL2/Tip48	
		Pih1p/Nop17p (P)	PIH1D1	
		Tah1p (T)	RPAP3	
		Bcd1p	BCD1/ZNHIT6	
		Rsa1p	NUFIP	
		Hit1p	ZNHIT3	
		Rtt106p		

Figure 8 : Tableau récapitulatif de l'ensemble des protéines cœur et des facteurs protéiques intervenant dans l'assemblage des snoRNP à boîtes C/D.

Pour chaque protéine son homologue est indiquée chez l'homme, la levure et les archées. Les facteurs d'assemblage indiqués n'ont pas d'homologues chez les archées.

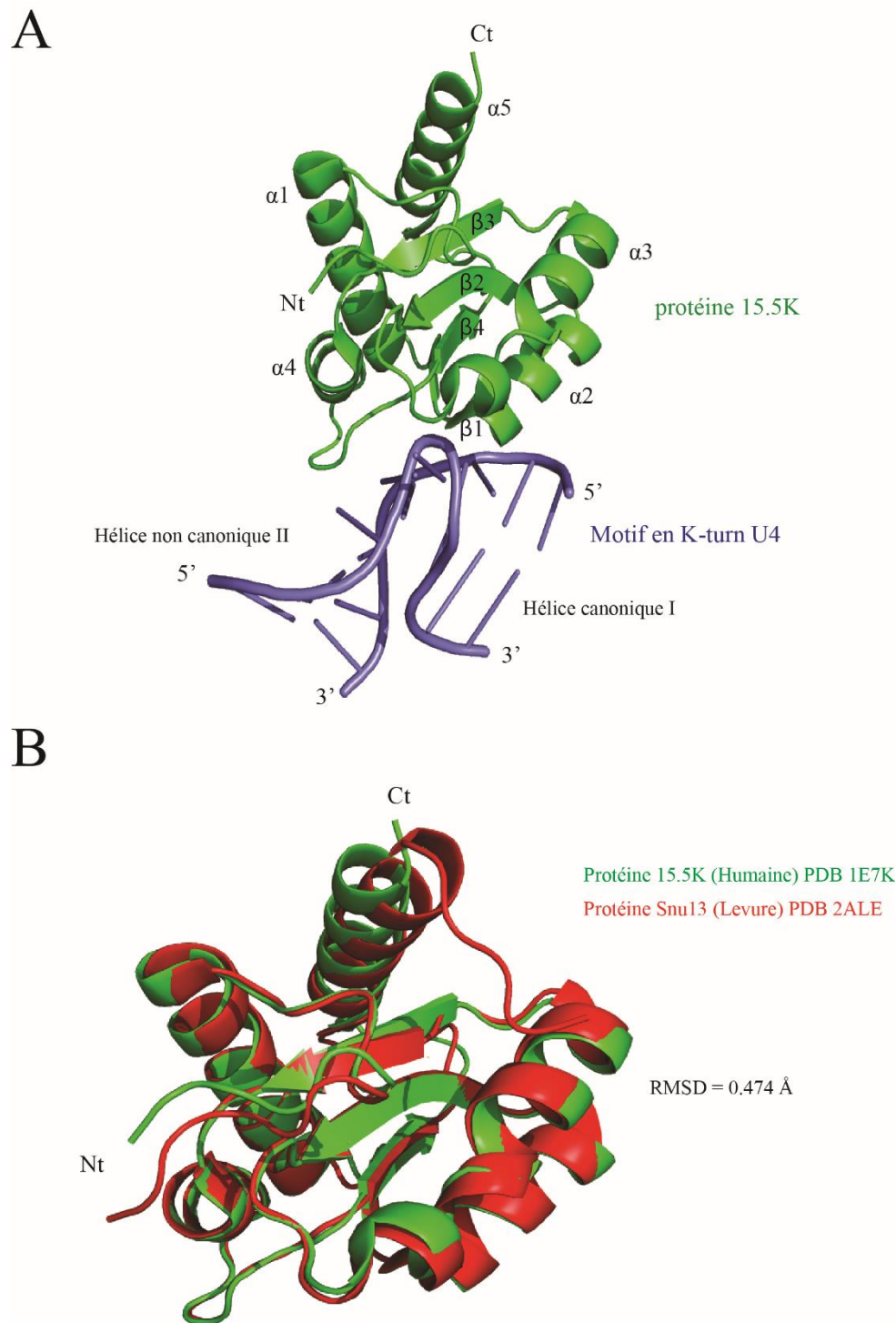


Figure 9 : La protéine cœur Snu13/L7Ae interagit avec le motif en K-turn du snoARN à boîtes C/D.

(A) Structure 3D obtenue par cristallographie aux rayons X d'un complexe entre la protéine cœur 15.5K (SNU13) humaine (vert) et le motif en *K-turn* du snARN U4 (violet) (Vidovic et al., 2000) PDB : 1E7K. (B) Alignement de structure des protéines 15.5K humaine en vert (PDB : 1E7K) et la protéine Snu13 de levure en rouge (PDB : 2ALE) avec un RMSD (*Root Mean Square Deviation*) de 0,474 Å.

L'interaction de la protéine Snu13 avec le snoARN est une étape nécessaire au recrutement des autres protéines cœur de la snoRNP C/D. Des études physico-chimiques ont permis de montrer que l'ARN libre n'est que partiellement structuré sous forme de *K-turn*, avec tout de même une conformation pré-liée caractéristique avec un angle entre les deux hélices de 30° (voir paragraphe ci-dessus). Son association avec la protéine stabilise complètement la conformation globale de l'ARN et la rend favorable à l'arrimage des autres protéines (Chagot et al., 2019; Goody et al., 2004) (Figures 7, 9A).

Il est à noter que l'absence de la protéine Snu13 chez la levure conduit à un défaut d'épissage des pré-ARNm suite à un déficit de snRNP U4 matures et fonctionnelles, un défaut de maturation des pré-ARNr, ainsi qu'à une diminution drastique de la stabilité des snoARN à boîtes C/D (Dobbyn and O'keefe, 2004). Cela conduit à dire que Snu13p/SNU13 est cruciale à la viabilité cellulaire.

b. La 2'-O-méthylase Nop1/Fibrillarine

La réaction de 2'-O-méthylation réalisée par les particules snoRNP C/D est catalysée par une enzyme qui fait partie des protéines constitutives de la particule. Cette enzyme est nommée Nop1 chez la levure et Fibrillarine chez l'homme. Son nom est dû à sa découverte pour la première fois dans des régions organisatrices nucléolaires chromosomiques putatives (NORs). Ce sont des régions fibrillaires et granulaires du nucléole de cellules humaines (Ochs et al., 1985). Chez l'homme, la Fibrillarine est constituée de deux domaines distincts : un domaine GAR (*Glycin-Arginin-Rich*) du côté N-terminal, un domaine méthyltransférase caractéristique avec un domaine de liaison aux ARN et de liaison au SAM (S-Adénosine-(L)-Méthionine), le donneur de méthyle (Figure 10A). Le domaine *Rossmann fold-Like*, composé d'un feuillet β à 7 brins centraux flanqués par 3 hélices α , constitue le domaine catalytique. Il est caractéristique des méthyltransférases SAM dépendantes de classe I (Schubert et al., 2003). Des domaines GAR similaires sont retrouvés dans d'autres protéines nucléolaires, par exemple, la nucléoline, GAR1, NSR1 et SSB1 (Jong et al., 1987; Lapeyre et al., 1987). Un polypeptide correspondant au domaine GAR de la nucléoline a été exprimé dans *Escherichia coli* et purifié. Il s'agit d'un tour β répété qui interagit de manière non spécifique avec l'ARN non empilé et déplié. Un rôle dans des interactions protéine-protéine ou de ciblage nucléolaire n'est pas à exclure.

L'enzyme 2'-O-méthylase est responsable de la substitution de l'hydrogène du groupement hydroxyle (OH) en position 2' du ribose de l'ARNr cible par un groupement méthyle issu du cofacteur SAM (Aittaleb et al., 2003; Wang et al., 2000) (Figure 10A).

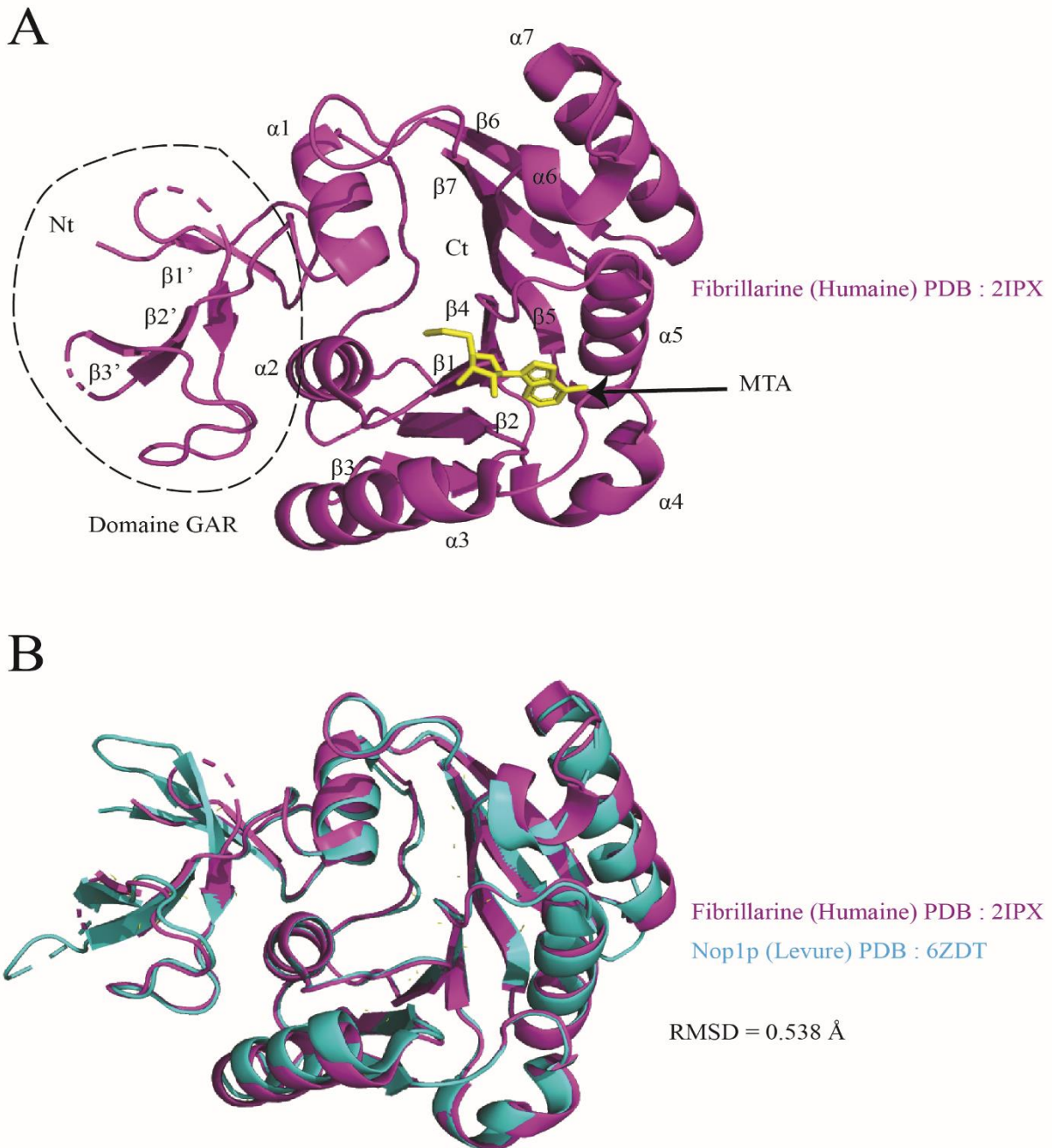


Figure 10 : Conservation de la structure tridimensionnelle de la 2'-O- méthylase, de la levure à l'homme.

(A) Structure 3D obtenue par cristallographie aux rayons X de la protéine humaine Fibrillarine PDB : 2IPX (Min et al., 2006). Le MTA (5'-deoxy-5'-MéthylThioAdénosine) représenté en jaune est un analogue du cofacteur SAM. Le domaine GAR est entouré en trait discontinu. (B) Alignement de structure des protéines Fibrillarine humaine en rose (PDB : 2IPX) et la protéine Nop1p de levure en bleu (PDB : 6ZDT) (Höfler et al., 2021) avec un RMSD (*Root Mean Square Deviation*) de 0,538 Å.

Une déprotonation du OH du ribose conduit à une attaque nucléophile de son oxygène sur le groupement méthyle du cofacteur SAM qui devient un groupement partant. Cette catalyse est rendue possible grâce au site actif de l'enzyme qui est constitué d'une triade catalytique conservée K-D-K, connue pour être essentielle à cette catalyse (Feder et al., 2003; Hager et al., 2002). L'intégration de la Fibrillarine dans les particules snoRNP C/D permet la catalyse des ARNr, en partie grâce au snoARN qui sert de guide, mais également grâce aux protéines plateformes aNOP5 des archées et Nop58/NOP58 Nop56/NOP56 chez la levure/homme. Cependant, Nop1/Fibrillarine est capable de catalyser des méthylations de protéine en dehors des particules ribonucléoprotéiques, plus exactement la méthylation de la glutamine Q105/104 de l'histone H2A chez la levure et l'homme (Tessarz et al., 2014). Une conservation fidèle de la structure 3D entre les protéines homologues Fibrillarine et Nop1 est indiquée par une superposition des deux structures 3D avec un RMSD de 0,538 Å (Figure 10B).

c. Les protéines Nop56/NOP56 et Nop58/NOP58

Chez la levure, la recherche de nouveaux composants de la machinerie de synthèse des ribosomes, a permis de mettre en évidence deux protéines nucléolaires, Nop56 et Nop58 par criblage de plusieurs mutations sur Nop1. Ces deux protéines présentent une forte homologie de séquence (45% d'identité et 63% de similarité) (Gautier et al., 1997) et sont donc sans doute issues d'un phénomène de duplication de gène chez les organismes eucaryotes. Chez les archées, nous ne retrouvons qu'une seule protéine, aNOP5. Les protéines orthologues chez l'homme sont également identifiées. Il s'agit des protéines NOP56 et NOP58. Malgré le fait qu'elles soient issues de la même protéine au cours de l'évolution, l'une ne peut pas se substituer à la fonction de l'autre. Des études ont montré que l'absence de ces protéines conduit à des phénotypes sévères pour la cellule : la déplétion de Nop58p engendre une diminution drastique du taux de snoARN à boîtes C/D et une inhibition des clivages aux sites A0, A1 et A2 dont la conséquence est un défaut de production de l'ARNr 18S (Lafontaine et Tollervey, 1999). L'absence de Nop56p présente un effet moins drastique sur la cellule, néanmoins une diminution du taux des 2'-O-méthylations des ARNr est observée.

La protéine Nop56 n'est capable de s'associer aux snoARN qu'en présence de la protéine Nop1, contrairement à Nop58p qui est capable de se lier au snoARN indépendamment du recrutement ou non de Nop1 (Lafontaine et Tollervey, 2000). Une étude utilisant des mutations des hélices I et II du *K-turn* qui bloquent le recrutement de la protéine SNU13 sur le snoARN, indique que l'association des protéines NOP58 et ensuite NOP56 nécessite le recrutement de

SNU13 au préalable, ou bien de façon simultanée (Watkins et al., 2002). En effet, NOP58 ne reconnaît que l'hétérodimère SNU13/snoARN.

Il est à noter que l'extrémité C-terminale de Nop56p et Nop58p comporte une séquence pseudo-consensus KKE/D répétée 11 et 13 fois respectivement chez la levure. Ces répétitions ne sont pas les mêmes dans la protéine NOP58 humaine et ne sont pas présentes dans la protéine des archées aNOP5. Elles sont remplacées par plusieurs zones de répétitions soit de Lysine (K) soit d'Acide Glutamique (E) chez l'homme. Ces motifs NoLS (*Nucleolar Localisation Signal*) chez la levure (Gautier et al., 1997) et l'homme (Pradet-Balade et al., 2011), sont responsables de la localisation nucléolaire de Nop56p/NOP56 et Nop58p/NOP58. Une hyperméthylation des régions C-terminales de ces protéines est essentielle à la structuration du nucléole ainsi qu'à l'adressage correct des snoRNP dans le nucléole (Colau et al., 2004; Mouaikel et al., 2002). Le clonage ainsi que la purification de la protéine recombinante Nop58 pleine taille de levure au laboratoire ne sont possibles qu'en absence des répétitions KKE/D en C-terminal.

Actuellement, aucune structure de résolution atomique n'a pu être obtenue pour les protéines eucaryotes Nop56/NOP56 et Nop58/NOP58 seules et entières. Cependant, la structure de la protéine des archées aNOP5 a pu être résolue en complexe avec la protéine aNOP1 (Figure 11) ou au sein d'une sRNP (*small RNP*, orthologue des snoRNP C/D) (Figure 12). Cette dernière est divisée en trois domaines. Un domaine N-terminal qui assure l'interaction avec l'enzyme aNOP1, et par extension Nop1/Fibrillarine chez la levure et l'homme respectivement. Il constitue le domaine catalytique dans sa globalité. La position de ce domaine dans la protéine entière est susceptible de varier en fonction des partenaires présents (Lin et al., 2011). Ensuite, un domaine NOP (Figure 11A) à l'extrémité C-terminale. Il interagit avec le sARN et plus exactement sur l'hélice II du motif en *K-turn*. Cette interaction n'exclut pas la reconnaissance de la protéine L7Ae du motif en *K-turn* (interaction avec les résidus de l'hélice I) et est même indispensable. Le domaine NOP participe avec le domaine en N-terminal à établir l'interaction avec aNOP1 (Lin et al., 2011) (Figure 11B).

La mutation d'une arginine du domaine C-terminal de la protéine aNOP5 perturbe la catalyse du cofacteur SAM par la méthyltransférase aNOP1 (Aittaleb et al., 2003). Enfin, aNOP5 possède un domaine central dit *coiled-coil* ou fagot d'hélices (Figure 11A) qui comprend deux hélices ($\alpha 4$ et $\alpha 5$) super-enroulées.

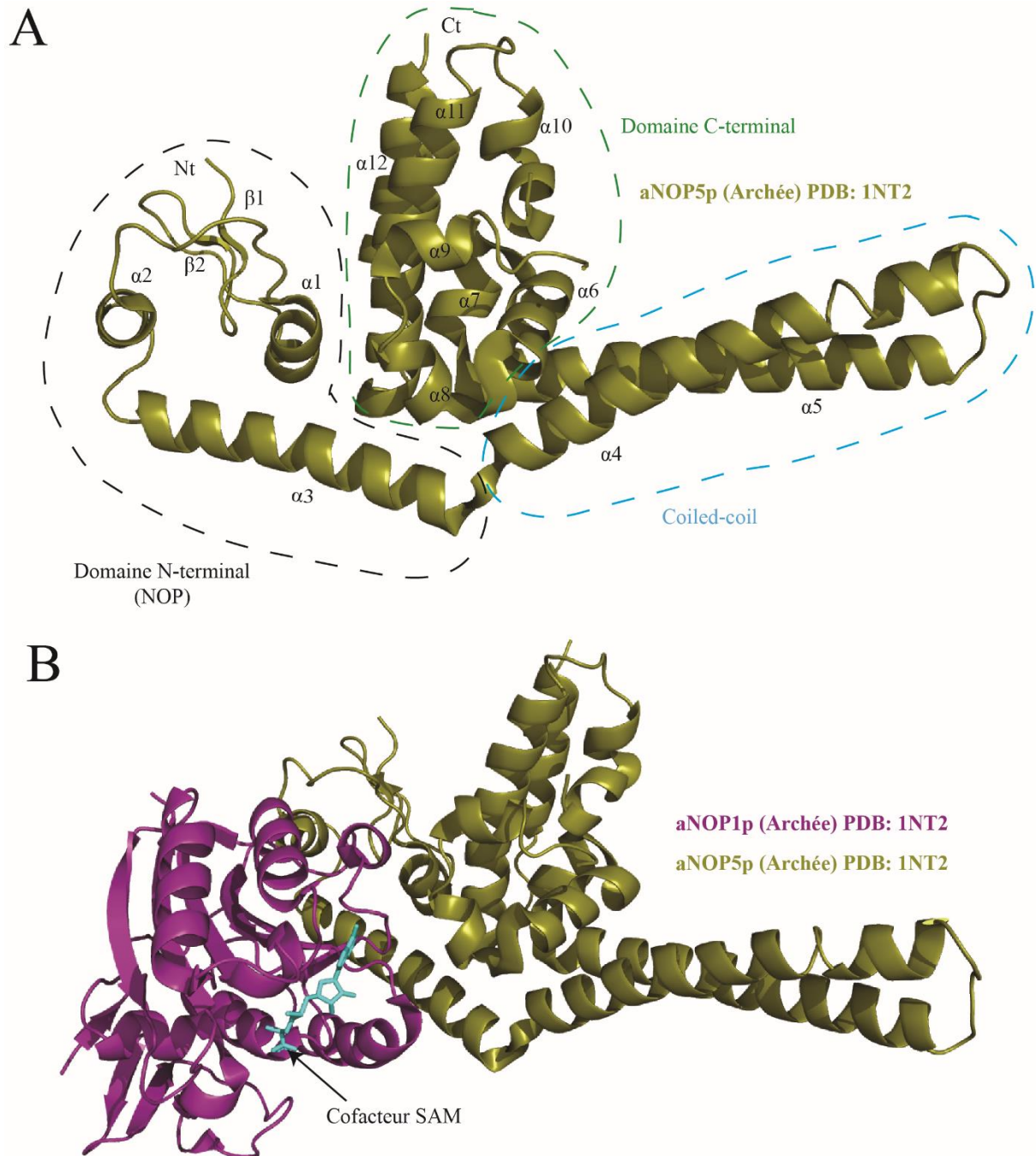


Figure 11 : Structure tridimensionnelle du complexe entre les protéines d'Archées aNOP5 et aNOP1.

(A) Structure tridimensionnelle de la protéine aNOP5 de *A. fulgidus* résolue par cristallographie aux rayons X (Aittaleb et al., 2003) PDB : 1NT2. La protéine aNOP1 a été supprimée pour des raisons de clarté. Les différents domaines qui constituent aNOP5p sont entourés par des traits discontinus.

(B) Structure tridimensionnelle du complexe aNOP5p-aNOP1 de *A. fulgidus* résolue par cristallographie aux rayons X (Aittaleb et al., 2003) PDB : 1NT2. Le cofacteur SAM est représenté en bleu. Cette interaction s'établit essentiellement par le domaine NOP de aNOP5p.

Il est responsable de l'homodimérisation de deux protéines aNOP5 chez les archées, ou de l'hétérodimérisation de Nop56p/NOP56 avec Nop58p/NOP58 (Aittaleb et al., 2003). Un petit domaine dit TIP (*TBP Interacting Protein*) est retrouvé entre les hélices $\alpha 4$ et $\alpha 5$ du domaine central. Ce domaine n'est pas retrouvé chez tous les organismes.

2.1.3. Structure 3D d'une particule sRNP C/D d'archée

L'étude structurale des snoRNP C/D eucaryotes n'a toujours pas permis de reconstituer une particule *in vitro*. Sa reconstitution par simple mélange de ses différents composants est impossible. Cela s'explique par l'importance fonctionnelle des protéines de la machinerie d'assemblage des snoRNP C/D chez les eucaryotes. Cependant, la structure 3D de la sRNP C/D des archées a pu être résolue grâce à sa reconstitution *in vitro* effectuée par plusieurs équipes (Figure 12). Etant donné la forte conservation entre les eucaryotes et les archées des éléments composants ces particules, une transposition de l'organisation structurale globale peut être faite entre la sRNP C/D et la snoRNP C/D des eucaryotes. Il existe deux modèles issus de deux hypothèses différentes qui co-existent concernant l'organisation 3D des sRNP C/D. Un assemblage en mono-sRNP (dite monomérique) (Figure 12), constitué principalement d'un sARN à deux jeux de boîtes C/D synthétique hybride, sur lequel vont s'agencer les protéines cœur par paires (une sur chaque jeu) (L7Ae, aNOP1, aNOP5) (Lin et al, 2011). L'autre possibilité, est une organisation en di-sRNP (dite dimérique), constituées d'une double mono-sRNP décrite précédemment (Lapinaite et al., 2013). Un premier modèle a été établi d'une sRNP de *sulfolobus solfataricus*. Nous y retrouvons un snoARN guide constitué d'un seul couple de boîtes C/D, la protéine aNOP1, la protéine L7Ae et les protéines aNOP5 qui interagissent ensemble par leur domaine CC. Ce mode d'interaction est très similaire à celui observé entre SNU13-PRP31-U4 (Liu et al., 2007) (Figure 12). D'autres structures de sRNP C/D ont également été résolues, notamment celles de *Pyrococcus furiosus* et *Methanococcus jannaschii* (Bleichert et al., 2009).

2.2. Les snoRNP à boîtes H/ACA

Ce sont des ribonucléoparticules qui contiennent un snoARN à boîtes H et ACA, ainsi que quatre protéines cœur chez la levure, Gar1p, Nhp2p, Nop10p et Cbf5p. Ces protéines sont fortement conservées de la levure aux mammifères. Ces RNP catalysent pour la plupart, les

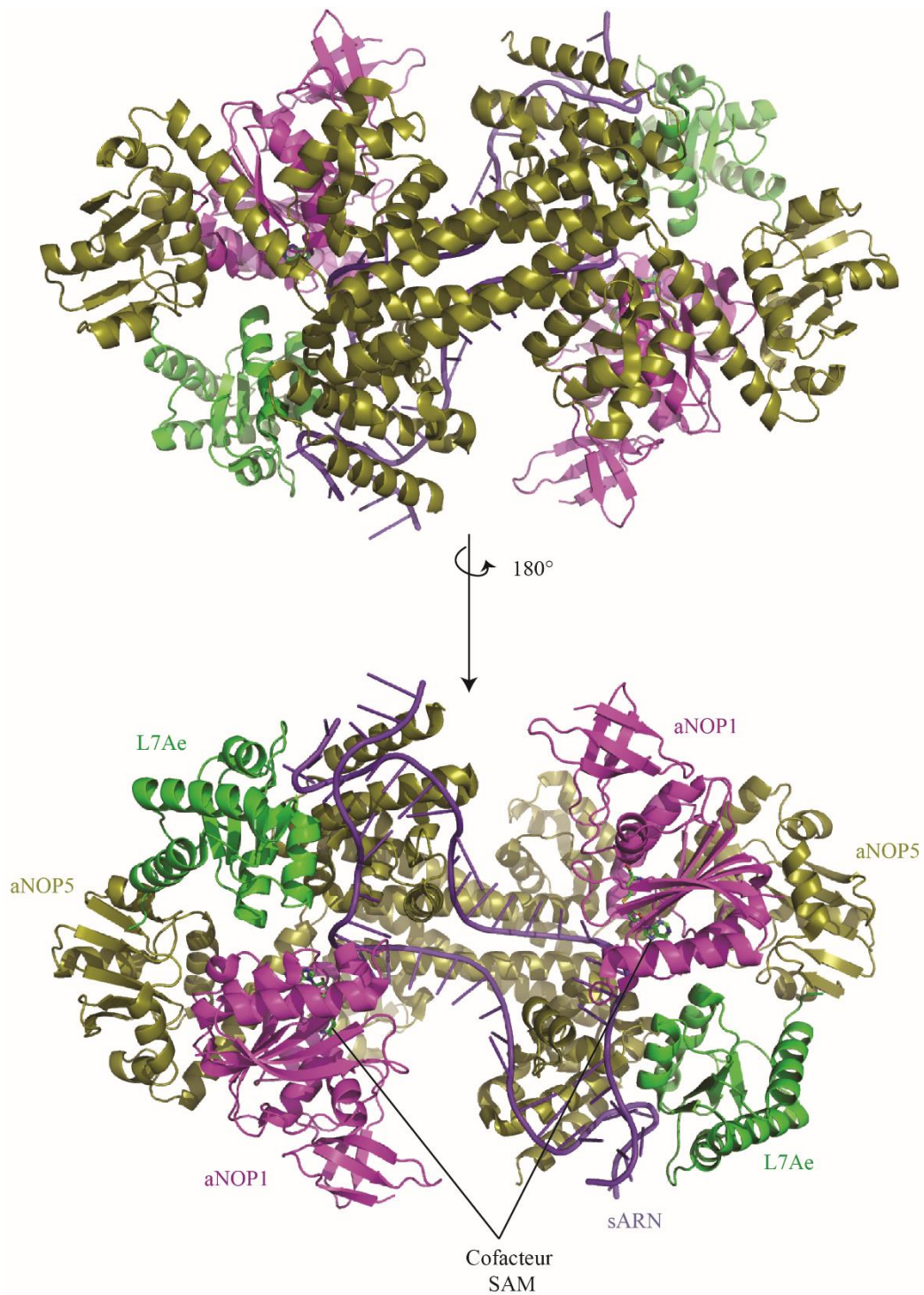


Figure 12 : Organisation structurale 3D d'une particule en mono sRNP d'archée.

Structure tridimensionnelle résolue par cristallographie aux rayons X d'une mono sRNP de *S. solfataricus* (construction synthétique) catalytiquement active. L'ARN cible a été retiré par soucis de clarté. Deux jeux de boîtes C/D sont représentées sur le sARN guide C/D (en violet). Le cofacteur SAM est illustré en bâtonnets. Les protéines identiques sont représentées de la même couleur (Lin et al., 2011) PDB : 3PLA.

réactions de pseudouridylation de nucléotides des précurseurs d'ARNr (Charette and Gray, 2000). Les domaines H/ACA sont également importants pour la fonction de quelques télomérases (Henras et al., 2004). Certains snoARN à boîtes H/ACA comme U17/snr30 participent à la maturation et donc aux clivages des ARNr (Morrissey and Tollervey, 1993).

2.2.1. Structure d'ARN H/ACA : Structure 3D et organisation

La transcription des ARN à boîtes H/ACA chez les levures et les archées se fait à partir des gènes organisés le plus souvent en unités transcriptionnelles indépendantes (mono- ou poly-cistroniques). Tandis que chez les mammifères, ces gènes sont retrouvés au niveau des parties introniques du génome codant essentiellement pour les ARNm.

Le nom H/ACA correspond à la présence de deux boîtes consensus, précédées chacune d'une structure en tige-boucle : la boîte H pour *Hinge*, présente une séquence consensus 5'-ANANNA-3' où N correspond à n'importe quelle base, et la boîte ACA qui correspond à la séquence 5'-ACA-3' située à l'extrémité 3' de l'ARN (Balakin et al., 1996; Ganot et al., 1997). Elles jouent un rôle important dans la stabilité et la maturation du snoARN ainsi que dans la catalyse de la réaction de pseudouridylation. Une bulle interne simple brin appelée poche de pseudouridylation, située au niveau de chaque tige-boucle, assure un appariement avec l'ARN cible, de façon que le résidu cible à modifier (Ψ = pseudouridine) soit localisé 14 à 16 nucléotides en amont des boîtes H ou ACA (Ganot et al., 1997).

La biogenèse des snoRNP H/ACA se fait de façon co-transcriptionnelle. L'accumulation du précurseur de snoARN à boîtes H/ACA immature à l'extrémité 5' nécessite la présence des protéines cœur de la particule snoRNP H/ACA. Durant leur biogenèse, les snoRNP H/ACA sont adressées au niveau des corps de Cajal, où certaines étapes de l'assemblage ont lieu, puis transférées au nucléole afin de catalyser les réactions de maturation des ARN cibles (Jády et al., 2004; Richard et al., 2003).

2.2.2. Protéines cœur de la snoRNP à boîtes H/ACA

La première protéine à reconnaître le snoARN à boîtes H/ACA est Nhp2p/NHP2 (*Non-Histone Protein 2*) (17 kDa). C'est la seule protéine de la famille L7Ae à ne pas exiger de motif en *K-turn* pour son interaction avec l'ARN. La structure tridimensionnelle de Nhp2p indique une conformation identique à celle des protéines de la famille L7Ae (Koo et al., 2011). Un des spécificités de reconnaissance des ARN H/ACA, vient d'une isomérisation en *trans* de la proline 83 de Nhp2p, responsable de l'interaction avec le snoARN, qui serait différente de

l'isomérisation *cis* retrouvée dans les protéines de la famille L7Ae qui reconnaissent les motifs *K-turn* d'ARN.

La protéine Gar1 (*Glycine Arginine Rich 1*) (21 kDa) est constituée d'un tonneau β central de 6 brins, ainsi que de deux autres domaines : un domaine NTE (N-terminal) dont l'absence de ses deux prolines conservées conduit à une agrégation de la protéine et la disparition des corps de Cajal (CB = *Cajal Bodies*). Tandis que le domaine CTE (C-terminal) comporte une hélice α qui assure une interaction avec la protéine cœur Cbf5. Cette interaction est mutuellement exclusive avec le facteur d'assemblage Naf1p pour Cbf5 (Darzacq et al., 2006).

La protéine Cbf5 (*Centromere Binding Factor 5*) /Dyskérine (levure/humain) (54 kDa) est essentielle à la viabilité cellulaire. Son absence induit chez la levure un défaut fonctionnel de la particule RNP en affectant la stabilité des snoARN H/ACA (Watkins et al., 1998 ; Lafontaine et al., 1998). Elle se divise en deux domaines. Un domaine PUA (*PseudoUridine synthase and Archaeosine transglycosydase*), dont la mutation est responsable du développement de la dyskératose congénitale (Heiss et al., 1998). Un domaine catalytique, qui lui permet d'assurer son rôle de pseudouridine-synthase. Sa cristallisation sous la forme libre n'a pas été possible, potentiellement en raison de la présence de motifs répétés KKE/D à son extrémité C-terminale, qui comme pour Nop58p et Nop56p, est fortement dégradée *in vitro*.

Une quatrième petite protéine cœur Nop10/NOP10 (*Nucleolar Protein 10*) (6 kDa) se retrouve en interaction avec le domaine catalytique de Cbf5/Dyskérine. Elle est composée chez la levure d'un domaine N-terminal en feuillet β et d'un domaine C-terminal en hélice α avec une région dite charnière reliant les deux. Chez les archées, le domaine N-terminal de NOP10 est constitué d'un domaine à doigt de zinc, qui comporte un atome de zinc stabilisé par quatre résidus cystéine. La structure 3D du complexe CBF5:NOP10 des archées a été résolue au laboratoire (Manival et al., 2006). Son domaine C-terminal (NOP10) se retrouve en interaction avec Cbf5 par son hélice $\alpha 6$.

Chapitre 2 : Processus d'assemblage des particules snoRNP C/D fonctionnelles chez la levure

1. Décryptage de la machinerie d'assemblage des snoRNP C/D et de ses acteurs protéiques

Contrairement aux sRNP des archées, l'assemblage des particules snoRNP C/D eucaryotes matures fait intervenir un certain nombre de facteurs protéiques cellulaires de façon transitoire agissant en *trans*. De ce fait, leur assemblage *in vitro* reste impossible. Ces facteurs d'assemblages interviennent également dans le trafic cytoplasmique et ensuite nucléaire, avant que la RNP mature n'atteigne le nucléole. A l'heure actuelle, trois modules indépendants mais agissant de façon connectés, sont recrutés de façon séquentielle : Rsa1p-Hit1p/NUFIP1-ZNHIT3, le complexe HSP70/90-R2TP et la protéine Bcd1/BCD1 (ZNHIT6) (Bizarro et al., 2014; Boulon et al., 2008; Bragantini, 2016a; Khoshnevis et al., 2019; McKeegan et al., 2007; Paul et al., 2019a; Rothé et al., 2014a).

Dans ce chapitre, les facteurs d'assemblage de la levure *S. cerevisiae* seront décrits, ainsi que leur implication dans le processus d'assemblage selon ce qui est connu dans la littérature à ce jour.

1.1. Le complexe protéique Rsa1-Hit1/NUFIP1-ZNHIT3

La protéine Rsa1 (*Ribosome assembly 1*) doit son nom à sa découverte chez la levure *S. cerevisiae*. Elle a été identifiée comme étant nécessaire à la biogenèse de la grande sous-unité ribosomique 60S, avec notamment le maintien d'un taux stable d'ARNr 25S (Kressler et al., 1999).

Les deux protéines orthologues (levure/humaine) Rsa1/NUFIP1 (*Nuclear FMRP Interacting Protein 1*) ne partage qu'un seul motif commun nommé PEP de 37 acides aminés (Rothé et al., 2014b). Des travaux de triple hybride menés au laboratoire ont également permis d'identifier la protéine Rsa1 comme partenaire du motif B/C du snoARN U3 via la protéine Snu13. L'extrémité C-terminale (230-381) de Rsa1p est suffisante pour rétablir la croissance d'une souche Δ RS1 de levure.

En parallèle, NUFIP1 a été identifiée au laboratoire d'E. Bertrand (Institut de Génétique Humaine (IGH), Montpellier) comme partenaire de la protéine cœur humaine SNU13 par un système de criblage (Boulon et al., 2008). Cette interaction Snu13p/SNU13:Rsa1p/NUFIP1 fait

intervenir plus précisément le domaine PEP de Rsa1p. La structure 3D du complexe PEP-Snu13p a été résolue au laboratoire par cristallographie aux rayons X (Figure 13A) (Bizarro et al., 2014). La résolution de ce complexe a permis de définir les résidus à la surface d'interaction. Il s'agit notamment des trois résidus E₇₂ D₇₃ K₇₄ de la protéine Snu13p et de l'arginine 249 portée par Rsa1p, qui sont conservés chez la levure et chez l'homme. La surface d'interaction de Snu13p avec Rsa1p et à l'opposé de celle responsable de son interaction avec le motif en *K-turn* du snoARN C/D. Ces données suggèrent une possible existence d'un complexe à trois partenaires : snoARN C/D / Snu13p / Rsa1p. Un domaine en doigt de zinc est retrouvé seulement sur la protéine NUFIP1 humaine, en amont du domaine PEP.

Des travaux du laboratoire d'E. Bertrand ont mis en évidence une autre protéine, Hit1/ZNHIT3 (Boulon et al., 2008). Cette dernière présente de façon similaire à NUFIP1, un rôle dans l'assemblage des snoRNP à boîtes C/D. Le laboratoire a montré que Hit1p est responsable du contrôle de la concentration cellulaire à l'équilibre de la protéine Rsa1, à travers une interaction directe entre les régions C-terminales de Hit1p et Rsa1p (Figure 13B) (Rothé et al., 2014). Ceci a pour conséquence de prévenir la dégradation de Rsa1p. La structure 3D du complexe Rsa1p:Hit1p explique bien cette fonction, où les deux hélices de Rsa1p se retrouvent imbriquées entre les cinq hélices de Hit1p (Figure 13B). Hit1p ainsi que son orthologue humain ZNHIT3 sont caractérisés par la présence d'un arrangement en double doigts à zinc chevauchants (domaine Zf-HIT, Figure 14) dans leur partie N-terminale respective (Bragantini et al., 2016). Une étude intéressante semble indiquer que NUFIP1 est un partenaire direct de la protéine LC3B (*Light Chain 3B*), qui est une protéine centrale dans le processus d'autophagie (Wyant et al., 2018). Cette interaction est essentielle pour guider le ribosome vers l'autophagosome, assurant ainsi son recyclage.

1.2. La protéine Bcd1

La protéine Bcd1 (*Box C/D snoRNA protein 1*) chez la levure ou la protéine ZNHIT6/BCD1 chez l'homme, ont été identifiées dans des expériences de criblage génétique haut débit sur des mutants associés à des défauts de biogenèse ou de stabilité d'ARN non codants. L'absence de cette protéine chez la levure se traduit par une diminution du taux de snoARN à boîtes C/D spécifiquement (Hiley et al., 2005; Peng et al., 2003). Une étude de la protéine BCD1/ZNHIT6 chez les mammifères utilisant des siARN spécifiques de cette protéine, indique une forte réduction du taux de snoARN U3 nucléolaire. Ce facteur d'assemblage interagit avec des protéines cœur (Nop58p/NOP58p, Snu13p/SNU13p) de la particule snoRNP à boîtes C/D,

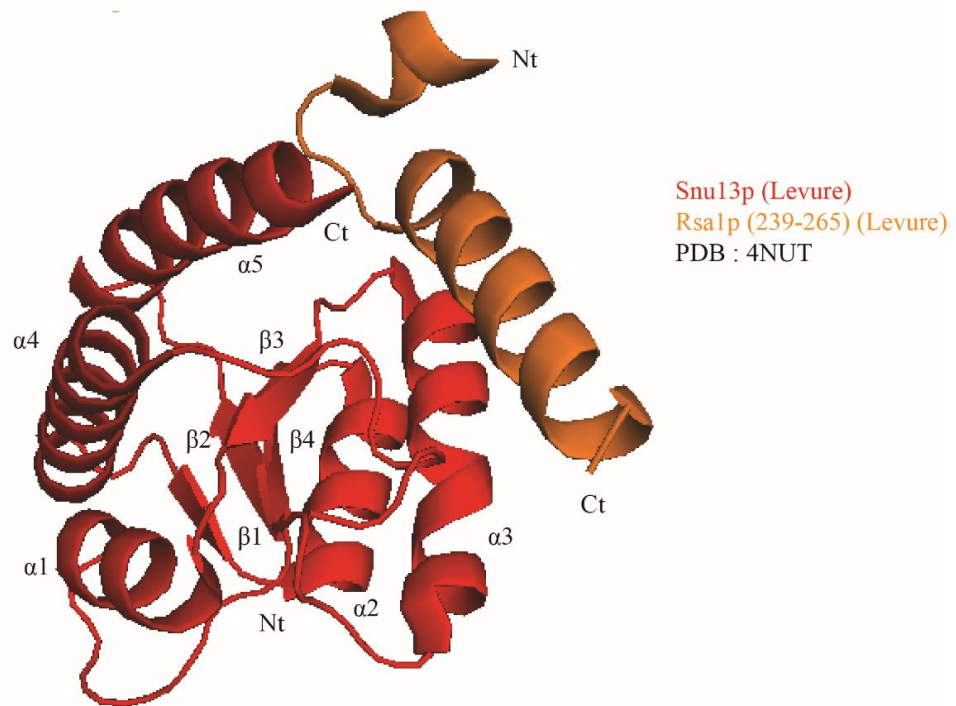
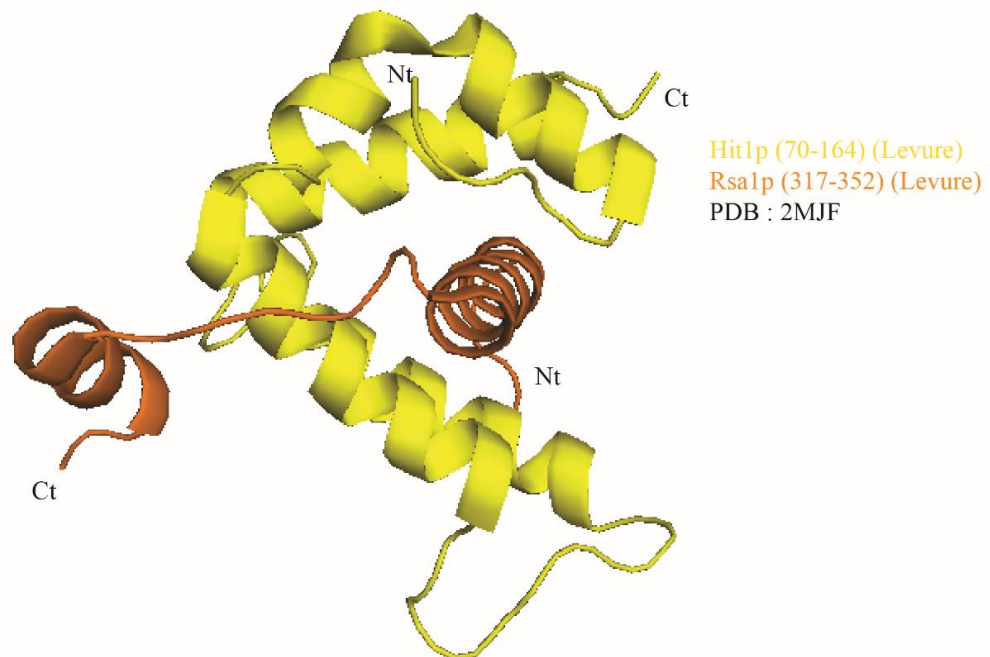
A**B**

Figure 13 : Rsa1p interagit avec Hit1p et Snu13p selon deux interfaces d'interactions différentes.

(A) Structure tridimensionnelle du complexe protéique de Snu13p avec Rsa1p₂₃₉₋₂₆₅ (domaine PEP) chez la levure *S. cerevisiae* obtenue par cristallographie aux rayons X, code PDB 4NUT (Bizarro et al., 2014).
(B) Structure tridimensionnelle du complexe Rsa1p₃₁₇₋₃₅₂ avec Hit1p₇₀₋₁₆₄ obtenue par RMN code PDB 2MJF (Rothé et al., 2014b).

suggérant ainsi un rôle de pontage entre différentes pré-snoRNP lors du processus d'assemblage (McKeegan et al., 2007; Paul et al., 2019a).

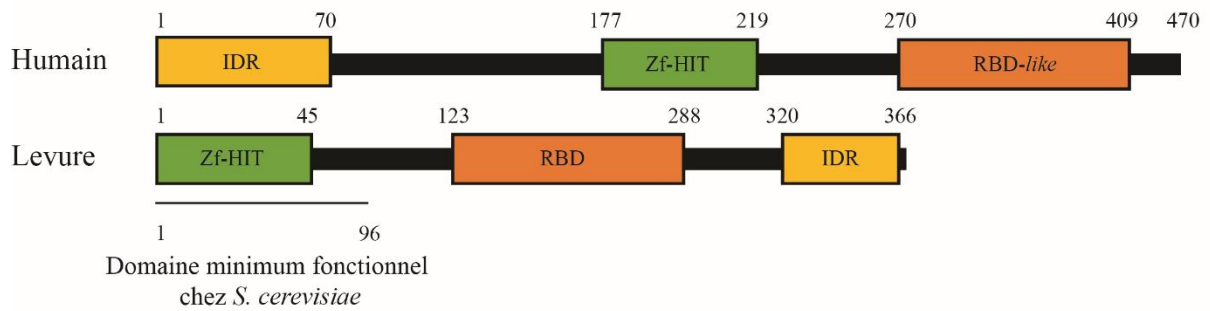
1.2.1 Domaines et organisations structurales

Selon des alignements de séquences protéiques (Clustal Omega) ainsi que des prédictions de structure (IUPred et Phyre2) (Figure 14A) (Khoshnevis et al., 2019), Bcd1p est constituée d'un domaine IDR (*intrinsically disordered region*) à son extrémité C-terminale (Bcd1p₃₁₀₋₃₆₆). Ce même domaine IDR est retrouvé chez son homologue humain BCD1/ZNHIT6 à son extrémité N-terminale (BCD1₁₋₇₀). Bcd1p et ZNHIT6 font parties des protéines de la famille Zf-HIT (*Zinc finger HIT*, (Bragantini et al., 2016), caractérisée par la présence d'un double motif en doigt à zinc, situé à l'extrémité N-terminale chez la levure (Bcd1₁₋₄₅) et en position centrale chez la protéine humaine (BCD1₁₇₇₋₂₁₉). Le domaine Zf-HIT de Bcd1p/BCD1 est essentiel pour la stabilité de la protéine chez la levure et nécessaire mais suffisant pour la croissance cellulaire (Bragantini et al., 2016), mais aussi serait en partie responsable de l'interaction avec les protéines humaines RUVBL1/RUVBL2 (Bragantini et al., 2016; Cloutier et al., 2017). Ce même domaine en doigt à zinc est retrouvé sur les protéines humaines ZNHIT2, 3 et 5 au niveau de l'extrémité N-terminale, et pour les protéines ZNHIT1 et 4 au niveau de la région C-terminale (Verheggen et al., 2015). Ces protéines font parties de la superfamille TRASH (*Trafficking, Resistance and Sensing of Heavy Metals*) caractérisée par un repliement tridimensionnel composé de deux à quatre brins β ainsi qu'une hélice α terminale. Cette superfamille est majoritairement caractéristique des eucaryotes, avec un motif fortement conservé constitué de plusieurs cystéines, allant jusqu'à 5 ou 7 pour la famille Zf-HIT. Elle est composée de 6 membres chez l'homme, dont deux protéines (ZNHIT6 et ZNHIT3) jouent un rôle dans l'assemblage de snoRNP C/D, tandis que ZNHIT1/SRCAP et ZNHIT4/INO80 font parties des complexes responsable du remodelage de la chromatine (Verheggen et al., 2015). Chez la levure, seulement trois membres constituent cette famille : Bcd1p et Hit1p, qui participent à l'assemblage des snoRNP à boîtes C/D et Vps71p qui est un facteur de remodelage de la chromatine (Bonangelino et al., 2002). Le deuxième point commun à l'ensemble de ces protéines est leur association avec les ATPases (Humain/Levure) RUVBL1/Rvb1 et RUVBL2/Rvb2.

Les structures tridimensionnelles en solution des domaines Zf-HIT (ZHD) de Bcd1p et Hit1p ont été résolues par RMN au laboratoire par B. Bragantini (Bragantini et al., 2016).

L'agencement de tous les ZHD semble être le même, avec un feuillet β constitué de deux brins

A



B

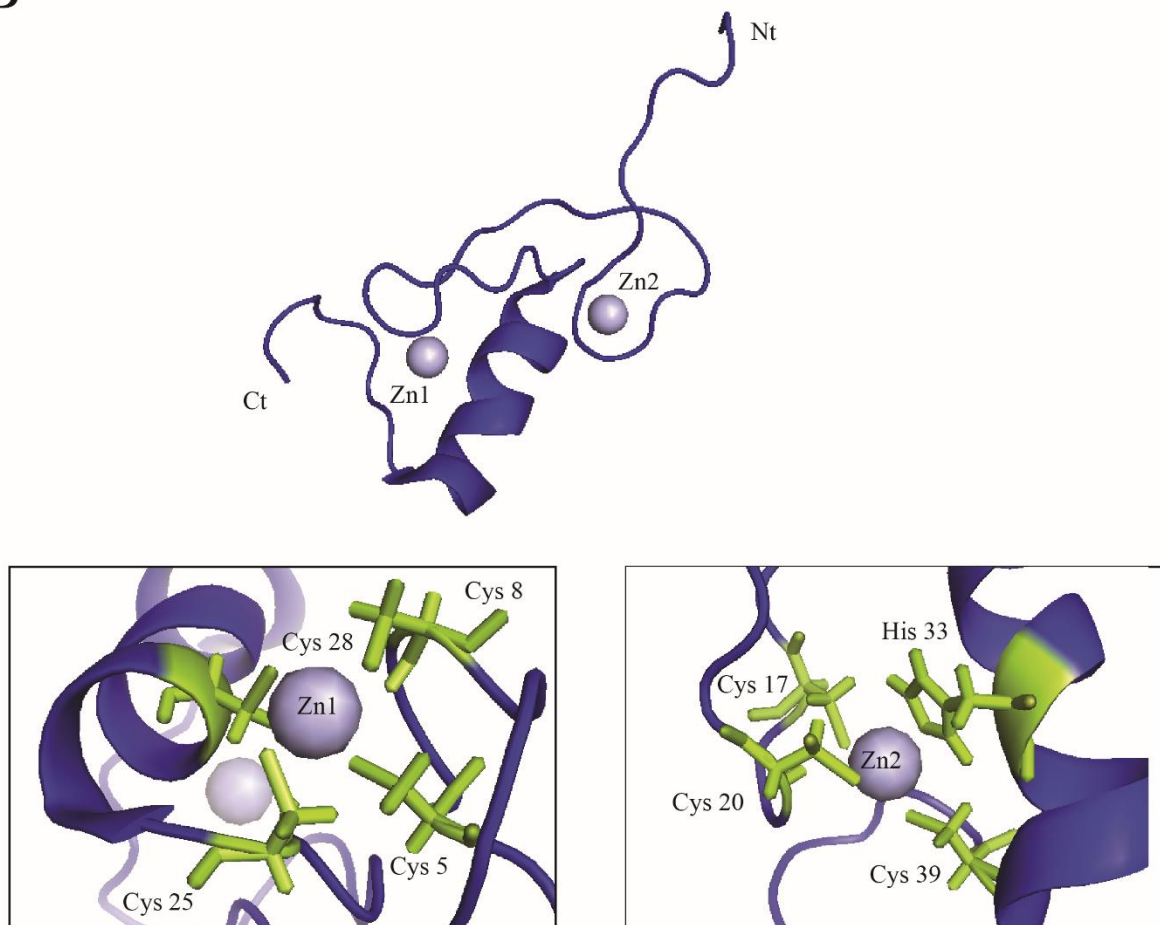


Figure 14 : La protéine Bcd1 de levure est organisée en domaines conservés avec son homologue humaine BCD1/ZNHIT6.

(A) Organisation en domaine de la protéine Bcdp/BCD1, IDR : *Intrinsically Disordered Regions*, Zf-HIT : *Zinc finger HIT*, RBD : *Rtt106 Binding Domain*.

(B) Structure tridimensionnelle du domaine ZHD de Bcd1p₁₋₄₅ résolue par RMN, code PDB : 2N94 (Bragantini, 2016b). Les acides aminés qui contribuent à la coordination des deux atomes de zinc sont représentés en vert.

antiparallèles posés sur une hélice α (Figure 14B). Le repliement correct de ce domaine nécessite la présence de deux atomes de zinc. Le premier atome de zinc est stabilisé par quatre cystéines (Cys5, Cys8, Cys25 et Cys29). Le second atome de zinc est en interaction avec trois cystéines (Cys17, Cys20, Cys39) et une histidine (His33) (Figure 14B).

La structure 3D du domaine central 120-303 de Bcd1p a également été résolue au laboratoire par RMN par B. Bragantini (Figure 15A). La séquence en acides aminés de ce dernier est similaire au domaine C-terminal de son homologue humain ZNHIT6 (Figure 14A) (Bragantini et al., 2021). Ce domaine est composé de 9 brins β et 4 hélices α , avec les brins β 4, β 5, β 6, β 7 et β 9 formant un feuillet β central torsadé. La majorité des hélices α sont regroupées sur un des côtés du feuillet β central. Ce domaine est responsable de l'interaction avec Rtt106 (une protéine chaperon d'histones) caractérisée *in vivo* par des expériences de double hybride chez la levure *S. cerevisiae* (Bragantini et al., 2021).

La structure de Bcd1p pleine taille n'a toujours pas été résolue expérimentalement. Il existe toutefois des approches informatiques capable de prédire dans certains cas les structures 3D protéiques avec une précision proche de celles issues de données expérimentales. *AlphaFold* est un logiciel de prédiction de structure protéique largement utilisé depuis son développement. Cela est essentiellement dû à sa grande fiabilité entre structure prédite et les données structurales expérimentales (Jumper et al., 2021). L'analyse de la structure de Bcd1p pleine taille par *AlphaFold* (Figure 16) indique une forte conservation des positions des résidus : 3-40, 126-170, 177-210 et 227-300. Cependant, nous retrouvons des régions déstructurées : 100-124, 212-225 et l'extrémité C-terminale 301-366. Cela apporte une flexibilité à la protéine en solution qui pourrait expliquer les échecs de cristalliser la protéine entière au laboratoire (Figure16) (Jumper et al., 2021; Varadi et al., 2022).

La structure 3D du complexe Bcd1p₁₂₀₋₃₀₃-Rtt106p₆₅₋₃₀₁ a été résolue au laboratoire par cristallographie aux rayons X (Figure 15B). Bcd1₁₂₀₋₃₀₃ interagit avec Rtt106p au niveau de son domaine PH1, faisant intervenir les résidus qui sont exposés aux solvant au niveau de l'hélice α 1 et le brin β 6. De plus, l'extrémité N-terminale des brins β 8 et β 10 de Bcd1p ainsi que la boucle en amont du brin β 9, sont partiellement enfouis avant l'interaction avec Rtt106₆₅₋₃₀₁. Un cluster hydrophobe formé à l'interface de l'hétérodimère faisant intervenir quatre résidus hydrophobes de chaque partenaire. Par ailleurs, d'autres liaisons ioniques sont retrouvées à cette interface d'interaction, impliquant des résidus chargés positivement de Bcd1p et des résidus chargés négativement de Rtt106p. Des liaisons hydrogènes sont également observées stabilisant le complexe (Bragantini et al., 2021). La comparaison de l'agencement spatial du domaine Bcd1p₁₂₀₋₃₀₃ entre sa forme libre et en interaction avec Rtt106p ne révèle pas de modification

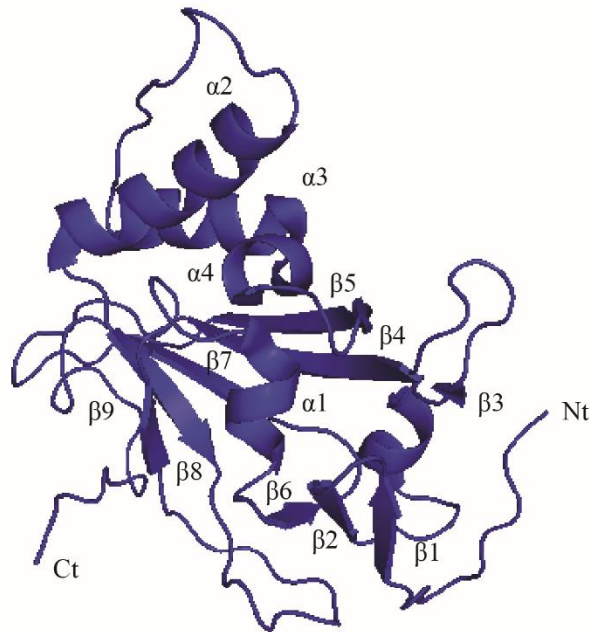
conformationnelle majeure (Bragantini et al., 2021). Rtt106p est connue pour être responsable de la déposition d'histones, en association avec la machinerie de transcription (Imbeault et al., 2008). De récentes données du laboratoire suggèrent que l'interaction Bcd1p-Rtt106p entre en compétition avec l'association de Rtt106p avec l'ARN polymérase II, en réduisant son association avec Rpb1p (la plus grande sous-unité de l'ARN polymérase II) et réduit donc son activité chaperonne au niveau des loci transcriptionnellement actifs (Bragantini et al., 2021).

1.2.2. Rôle dans les étapes d'assemblage de la snoRNP à boîtes C/D

Des expériences de double hybride réalisées au laboratoire et par l'équipe d'E. Bertrand sur des protéines de la levure *S. cerevisiae*, ont permis de caractériser des associations entre Bcd1p et des constituants du complexe R2TP : Pih1p et Rvb2. Cette interaction a également été observée avec les mêmes protéines humaines homologues (McKeegan et al., 2007). Un travail de thèse réalisé au laboratoire par D. Tiotiu a permis de mettre en évidence la région nécessaire pour la fonction de Bcd1p chez la levure *S. cerevisiae* (Tiotiu, 2016). Le fragment minimum de Bcd1p en région N-terminale (1-96) permet de restaurer la croissance, ainsi que le taux de snoARN à boîtes C/D transcrits par l'ARN pol II, dans les souches de levure dont la protéine Bcd1 endogène a été inactivée. Ce fragment comporte en partie N-terminale le domaine en doigt de zinc décrit précédemment. Toutefois, ce domaine est nécessaire, essentiellement en condition de stress thermique, mais n'est pas suffisant seul pour assurer la fonction de Bcd1p dans la cellule. Comme cela a été décrit plus haut, l'inactivation de Bcd1p se traduit par un défaut de stabilité uniquement de snoARN C/D (Peng et al., 2003), ainsi qu'à une accumulation des précurseurs d'ARNr 35S et 23S, accompagnés d'une diminution de l'expression des ARNr 18S et 25S. Ces données ont permis d'affirmer que Bcd1p est un facteur central et clé de l'assemblage des snoRNP à boîtes C/D. Un de ses rôles essentiels est le recrutement de la protéine cœur Nop58 sur la pré-snoRNP en cours d'assemblage (Paul et al., 2019a).

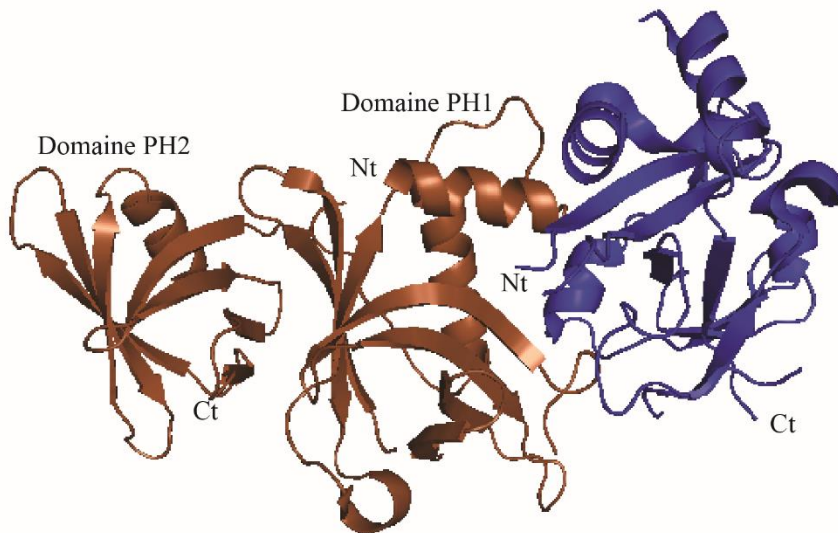
Bcd1p/BCD1 est recrutée sur les snoARN C/D de façon co-transcriptionnelle, sur les formes immatures en 3' des snoARN (U3, U16) chez la levure et chez l'homme. Elle participe également à la modulation de certaines associations lors de l'assemblage. En effet, l'absence de Bcd1p affecte drastiquement l'association entre Rsa1p et Nop58p. Cela suggère que Bcd1p participerait au rapprochement des complexes transitoires d'assemblage. Une étude récente montre que la mutation ponctuelle de l'Aspartate 72 en Alanine sur Bcd1p perturbe drastiquement la stabilité des snoARN à boîtes C/D avec une croissance des levures ralentie.

A



Bcd1p (120-303) (Levure)
PDB 6NZ2

B



Rtt106p (65-301) (Levure)
Bcd1p (120-303) (Levure)
PDB 6THL

Figure 15 : Le domaine central de Bcd1p interagit avec la protéine Rtt106.

(A) Structure tridimensionnelle du domaine Bcd1p₁₂₀₋₃₀₃ résolue par RMN code PDB 6NZ2 (Bragantini et al., 2021).

(B) Structure tridimensionnelle du complexe Rtt106p₆₅₋₃₀₁ et Bcd1p₁₂₀₋₃₀₃ résolu par cristallographie aux rayons X, code PDB 6THL (Bragantini et al., 2021).

Cette mutation est responsable d'une perte d'interaction directe entre Snu13p et Bcd1p *in vitro* (Khoshnevis et al., 2019).

1.2.3. Bcd1p et les mécanismes de rétrotransposition

Une analyse par spectrométrie de masse d'échantillon préparé au laboratoire, suggère l'implication de Bcd1p dans d'autres mécanismes cellulaires. En effet, l'analyse des protéines en association spécifique avec la partie C-terminale de Bcd1p de levure, révèle plusieurs protéines impliquées dans les mécanismes de rétrotransposition de gènes de la famille Ty1 (YD15B, YP13B, YL13B ...) ainsi que des protéines de la sous-unité de l'ARN polymérase III (RPC1, RPC5, RPC2...) (Paul, 2018). Cependant, l'absence de Bcd1p n'affecte pas la fonction de l'ARN polymérase III. La rétrotransposition est un phénomène de duplication de gènes mobiles, et cela par l'intermédiaire d'une étape impliquant l'ARN. Ils sont retrouvés chez toutes les espèces du règne eucaryote. Ce mécanisme est assuré par une particule dite VLP (*Virus-Like Particle*) qui contient l'ensemble des protéines nécessaires à la rétrotransposition (*Reverse Transcriptase* ou RT, *Intégrase* ou IN, *Protease* ou PR, ...). Chez la famille Ty1 de levure *S. cerevisiae*, ces protéines sont issues d'un seul et même transcrit primaire avec des extrémités 3' et 5' dites LTR (*Long Terminal Repeat*). L'étude de l'effet biologique de Bcd1p sur la rétrotransposition, ainsi que l'effet local de rétrotransposition au locus *SNR52* (comme déjà décrit dans le chapitre 1 est un snoARN à boîtes C/D transcrit par la pol III) se poursuit au laboratoire à l'aide de souche de levure génétiquement modifiées. Des résultats préliminaires indiquent que l'expression relative de snR52 en absence de l'extrémité C-terminale de Bcd1p baisse fortement. Cela suggère que la partie C-terminale de Bcd1p est essentielle pour la transcription ou la stabilité de snR52.

1.2.4. Bcd1p et la protéine chaperon d'histones Rtt106

La protéine Rtt106 (*Regulation of ty transposition 106*), est une protéine chaperon des histones H3 et H4 chez *S. cerevisiae*. Tout comme le complexe CAF-1 (*Chromatin Assembly Factor 1*), elle prévient l'agrégation des histones synthétisées *de novo* afin de former des nucléosomes, assurant ainsi les mécanismes de réplication, transcription et de réparation de l'ADN (Huang et al., 2005). Plus particulièrement, il s'agit d'un acteur majeur de la régulation négative de la transcription des gènes codant ces histones H3 et H4 (Silva et al., 2012; Zunder and Rine, 2012). De manière très intéressante, même si la délétion de Rtt106 n'affecte pas la stabilité des snoARN à boîtes C/D, une forte association a été identifiée au laboratoire entre Rtt106 et Bcd1p

en double hybride. Cette association s'est avérée être une interaction directe confirmée *in vitro* par co-expression chez *E. coli*. L'analyse par titration calorimétrique isotherme (ITC) montre qu'il s'agit d'une interaction endothermique avec une constante de dissociation, K_d de $1,17 \pm 0,17 \mu\text{M}$. La formation de ce complexe chez la levure suggère un lien entre la machinerie d'assemblage des snoRNP à boîtes C/D et la machinerie d'assemblage et de remodelage de la chromatine, et donc un point de coordination important entre l'activité de l'ARN polymérase II et la formation de ribosome (Bragantini et al., 2021). De manière surprenante, même si Rtt106 est associée aux snoARN à boîtes C/D, cette association semble dépendre uniquement de la présence de Bcd1p, et donc se déroule de façon co-transcriptionnelle durant l'assemblage des snoRNP C/D. Cependant, Rtt106 est faiblement recrutée aux *loci* codant les snoARN C/D (Paul, 2018). A ce jour, aucun homologue humain de la protéine Rtt106 n'a été mis en évidence chez l'homme.

1.3. Le complexe protéine chaperon/co-chaperon Hsp82/HSP90-R2TP

1.3.1. La protéine chaperon Hsp82/HSP90

Hsp82/HSP90 (*Heat Shock Protein*) est une protéine chaperon de levure et humaine hautement conservée, ubiquitaire, abondante et essentielle pour le bon fonctionnement des cellules. Leur rôle est de permettre le repliement correct de protéines, l'assemblage de complexes multimériques, d'éviter les phénomènes d'interaction non spécifique des protéines et les phénomènes d'agrégation, au sein d'un environnement intracellulaire extrêmement concentré en protéines (Makhnevych and Houry, 2012). C'est une protéine qui participe à d'autres mécanismes cellulaires, dont le remodelage de la chromatine et donc la régulation de l'expression des gènes. HSP90 est une protéine homodimérique flexible. Chaque monomère (90 kDa) est constitué de trois domaines (Figure 17A) : un domaine N-terminal responsable de la fixation de l'ATP et des co-protéines chaperon, un domaine central qui comporte une arginine catalytique essentielle à son activité ATPase et à la fixation des protéines clientes et un domaine C-terminal constituant l'interface d'homodimérisation. La structure 3D de la protéine chaperon de levure Hsc82 (PDB 6XLC) a été établie à une résolution de 3,66 Å par microscopie électronique. Elle présente l'arrangement dans l'espace de ses différents domaines constitutifs, avec une conformation fermée, en présence d'un analogue d'ATP non hydrolysable (AMP-PNP) (Figure 17A).

La structure de cette protéine chaperon montre clairement un arrangement en dimère parallèle

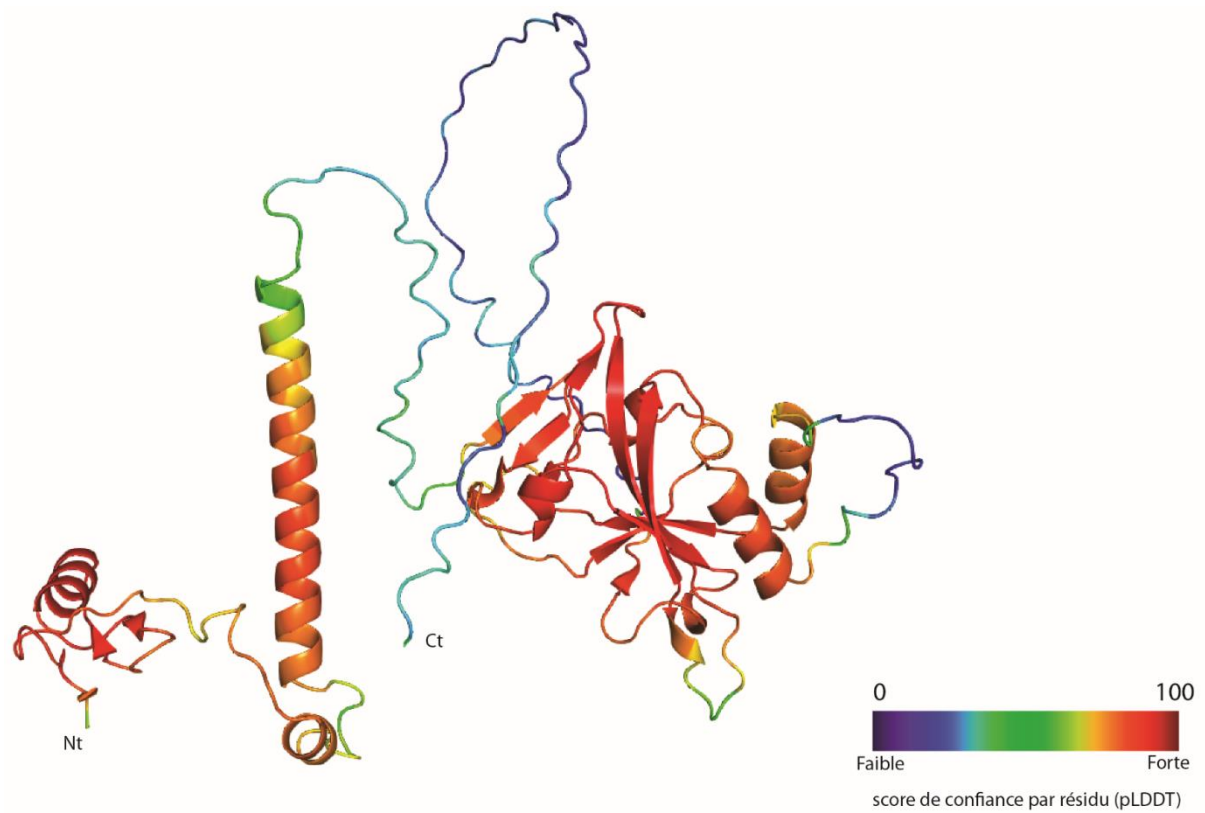


Figure 16 : Prédiction de la structure de Bcd1p de levure *S. cerevisiae* par le logiciel de prédiction *AlphaFold Monomer v2.0* (Jumper et al., 2021; Varadi et al., 2022).

AlphaFold produit un score de confiance par résidu (pLDDT = *predicted Local-Distance Difference Test*) compris entre 0 et 100. Certaines régions inférieures à 50 pLDDT peuvent être non structurées de manière isolée.

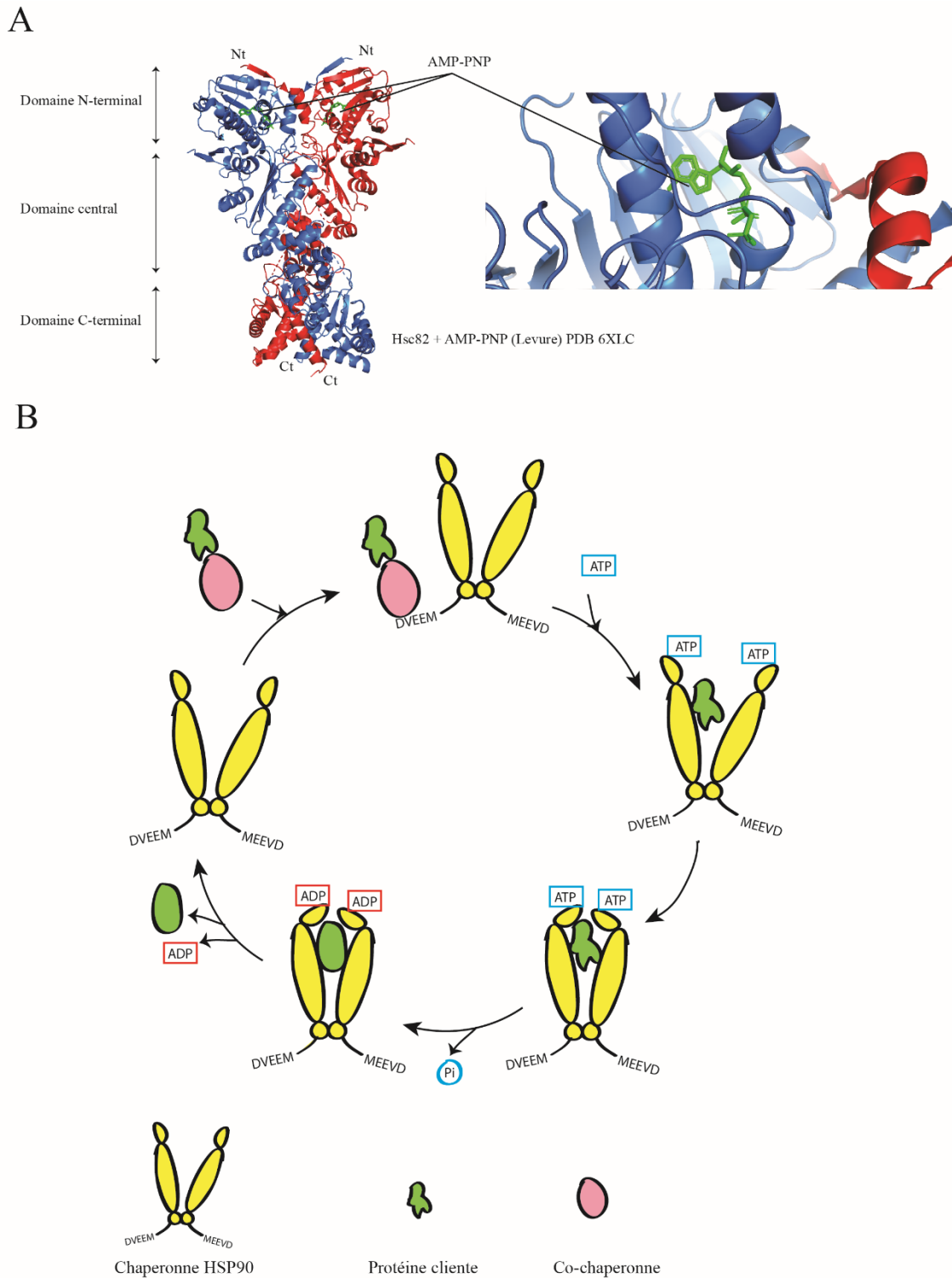


Figure 17 : Structure et cycle ATPase des protéines chaperon de la famille HSP90.

(A) Structure de la protéine chaperon Hsc82 de levure (*S. cerevisiae*) PDB 6XLC, fixant deux molécules d'AMP-PNP qui est un analogue non hydrolysable de l'ATP. Il mime l'état fermé de la protéine chaperon. Comme pour les protéines de la famille des HSP90, il s'agit d'un homodimère dont chaque monomère est composé de trois domaines.

(B) Cycle d'activité de la protéine chaperon HSP90, les protéines clientes sont apportées par une co-chaperonne qui se fixe au niveau de l'extrémité C-terminale de HSP90 (MEEVD), elles sont ensuite activées grâce à une hydrolyse d'ATP par HSP90.

dans lequel les deux domaines N et C-terminaux sont en interaction pour former un homodimère (Pearl and Prodromou, 2006). L'activité d'HSP90 impose un changement dans l'arrangement de ses deux monomères constitutifs (ouvert/fermé). Les protéines clientes sont spécifiquement apportées par des cofacteurs de la protéine chaperon, suivie d'une fixation de deux molécules d'ATP sur les domaines N-terminaux conduisant à leur rapprochement dans l'espace. Une hydrolyse d'ATP provoque le relargage d'ADP et de la protéine cliente correctement repliée. Par la suite, HSP90 retrouve sa conformation ouverte de départ. HSP90 contient également un motif hautement conservé formé des résidus EEVD (Figure 17B). Ce motif, placé à l'extrémité C-terminale stricte, est responsable du recrutement par interaction directe des protéines dites cofacteurs de la protéine chaperon. Celles-ci confèrent une spécificité d'action pour HSP90 (Pearl and Prodromou, 2006; Zeytuni and Zarivach, 2012). Elles possèdent 5 résidus basiques conservés qui forment une pince moléculaire à carboxylate (ou « *carboxylate clamp* ») capable d'interagir avec les résidus acides du motif EEVD par le biais d'un domaine TPR (*Tetratricopeptide Repeat*) (Scheufler et al., 2000a) Chaque domaine TPR est constitué en général de 6 motifs TPR. Un motif TPR est structuré en deux hélices α antiparallèles de 34 acides aminés.

Un criblage à haut débit des partenaires de Hsp82 de levure, a permis de lui associer de nouvelles fonctions. Parmi elles, l'association à un nouveau complexe a été identifié, le R2TP, constitué de quatre protéines : Rvb1, Rvb2, Tah1p, Pih1p (Zhao et al., 2005). Cette interaction avec le motif EEVD en C-terminal de Hsp82 est assurée par la présence d'un domaine TPR dans la protéine Tah1 (Back et al., 2013). Des expériences de *knock-down* du gène *TAH1* chez la levure, révèlent une dégradation de la protéine Pih1 en absence de Tah1p et donc par extension de Hsp82 sur le complexe R2TP. La protéine chaperon Hsp82 participe donc à la stabilisation de Pih1p (Zhao et al., 2008). L'inhibition par la geldanamycine d'HSP90 chez l'homme est associée à des défauts de stabilisation des protéines NOP58 et SNU13 ainsi qu'à une diminution du taux de snoARN U3 et U4 (Boulon et al., 2008).

1.3.2. Le complexe protéique cofacteur des protéines chaperon HSP90/R2TP

Un criblage à grand échelle des interactants de HSP90 a permis à l'équipe d'Houry, la mise en évidence d'un complexe multiprotéique R2TP chez *S. cerevisiae* (Zhao et al., 2005). Un complexe homologue chez l'homme a pu être également être décrit, constitué de RUVBL1 et RUVBL2 respectivement orthologues de Rvb1 et Rvb2, RPAP3 homologue de Tah1p et PIH1D1 homologue de Pih1p (Boulon et al., 2008; Jeronimo et al., 2007).

Le complexe R2TP est connu pour son rôle de cofacteur de la protéine chaperon Hsp82 chez *S. cerevisiae* et HSP90 chez l'homme. Son activité est essentielle et centrale dans divers processus cellulaires (Zhao et al., 2008). Il est notamment requis dans la biogenèse des snoRNP à boîtes C/D à travers une interaction directe entre Pih1p et la protéine cœur Nop58, ainsi qu'en association avec les protéines de la famille L7Ae (Boulon et al., 2008; Gonzales et al., 2005). Le complexe R2TP de levure a pu être reconstitué la première fois *in vitro* par co-élution (Zhao et al., 2008). Le complexe Rvb1-Rvb2 de levure a été identifié auparavant sous forme d'hétérohéxamère ou hétérododécamère selon l'ajout ou non d'ADP (Gribun et al., 2008). Des images du R2TP de levure ont été obtenues par cryo-EM (*Electronic Microscopy*) à une résolution de 8,37 Å. L'analyse des images 2D et 3D indiquent que Tah1p et Pih1p sont positionnées au niveau des domaines DII de Rvb1/Rvb2 qui forment une corbeille accueillant un seul complexe Tah1p/Pih1p pour un hexamère de Rvb1/Rvb2 (Figure 18) (Rivera-Calzada et al., 2017). L'analyse du complexe R2TP de levure indique que l'interaction de Pih1p avec Rvb1-Rvb2 est directe, tandis que Tah1p est apportée sur le R2TP par une interaction avec la protéine Pih1 (Zhao et al., 2008). De façon surprenante, le R2TP humain présente une interaction directe entre la protéine PIH1D1, par son domaine CS, et RPAP3 au niveau de sa région intrinsèquement désordonnée IDR. De plus, une interface d'interaction entre RPAP3, au niveau de son domaine C-terminal et le domaine DIII de RUVBL2, a été caractérisée. De ce fait, PIH1D1 n'établit aucune interaction directe avec le complexe hétéro-hexamérique RUVBL1/RUVBL2 (Seraphim et al., 2022). La déposition de RPAP3 seul ou RPAP3/PIH1D1 sur le complexe RUVBL1/RUVBL2 n'affecterait pas son activité ATPasique (Seraphim et al., 2022). Dans une autre étude, il a été montré que l'interaction de PIH1D1 avec RUVBL2 transforme la conformation spatiale du DII en présence d'ADP, permettant ainsi les échanges nucléotidiques (Muñoz-Hernández et al., 2019).

L'interaction du complexe R2TP avec ses protéines clientes a lieu lorsque ces dernières s'assemblent dans des complexes quaternaires, fournissant un mécanisme pour leur assemblage correct (Mao and Houry, 2017). Ce rôle de réarrangement quaternaire leur a valu le nom de PAQosome (*Particle for Arrangement of Quaternary structure*) chez l'homme lorsqu'il s'associe avec les modules préfoldines et pseudo-préfoldines (PFDL) (Houry et al., 2018).

Le mécanisme d'action du complexe R2TP n'a toutefois pas été encore élucidé. Il est sujet à d'intenses investigations. Il apparaît que RUVBL1/Rvb1-RUVBL2/Rvb2 semble être la clé catalytique des complexes d'assemblage de snoRNP à boîtes C/D. La structure cryo-EM du

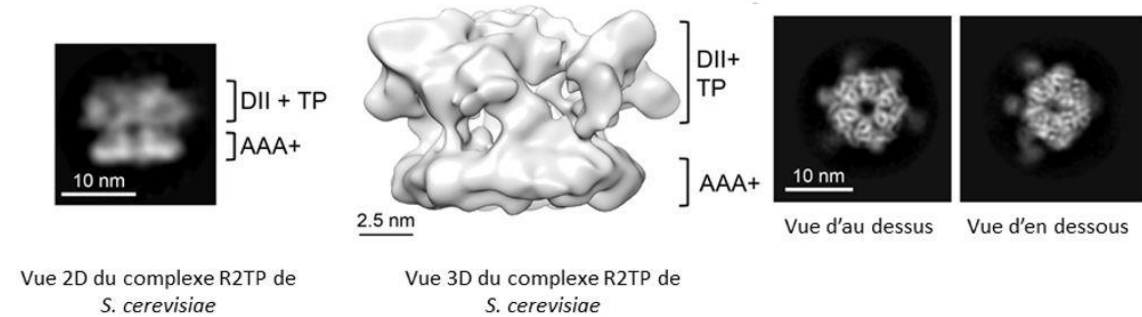
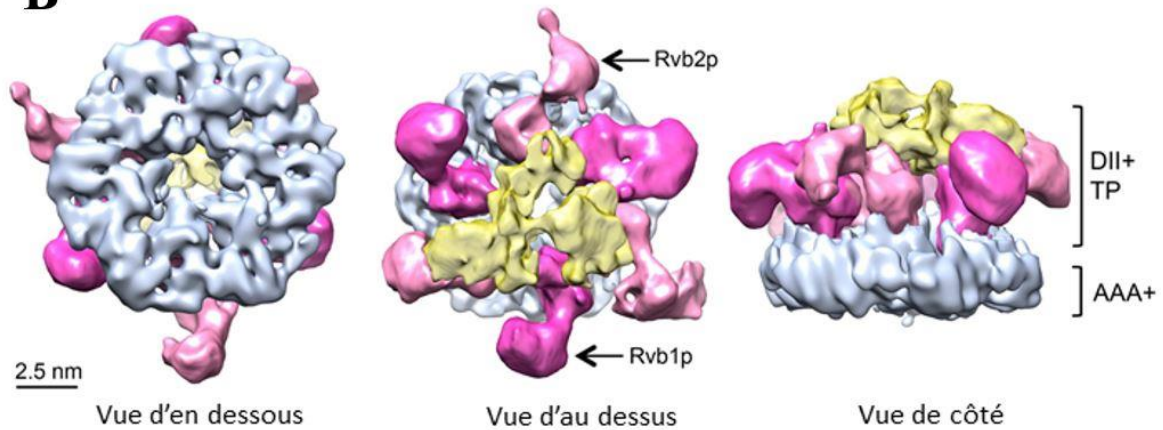
A**B**

Figure 18 : Structures 2D et 3D du complexe R2TP de levure *S. cerevisiae* (Rivera-Calzada et al., 2017)

(A) Les images 2D ont été obtenues par cryo-microscopie électronique. Elles représentent la moyenne de plusieurs images 2D sélectionnées. TP : Tah1p Pih1p, DII : domaine DII de Rvb1 et Rvb2, AAA+ domaines DI et DIII de l'anneau hexamérique des Rvb1/Rvb2.

(B) Représentation de différentes vues du complexe R2TP de levure, avec l'anneau hexamérique Rvb1/Rvb2 (DI + DIII) en bleu clair, les domaines DII en rose et les protéines Tah1 et Pih1 en jaune.

complexe R2TP-Nop58p met en évidence les interfaces d'interactions, particulièrement le positionnement de Nop58p sur le complexe R2TP (Yu et al., 2019). qui établit une interaction avec le domaine C-terminal de Pih1p (Figure 19).

1.3.2.1. Les ATPases Rvb1/Rvb2

a. Mécanismes cellulaires associés

Rvb1 et Rvb2 (dont les orthologues chez les mammifères sont RUVBL1 et RUVBL2) font parties des ATPases AAA+ (*ATPases Associated with various cellular Activities*). Elles sont associées dans plusieurs complexes protéiques nucléolaires (Figure 20). Ils possèdent un spectre d'action très large, en agissant comme protéine chaperon de complexes protéiques impliqués dans le processus de remodelage de la chromatine INO80 (*INOsitol requiring 80*) responsable de la reconnaissance des cassures double brin de l'ADN. Rvb1 et Rvb2 s'associent avec la région d'insertion du domaine ATPase de INO80. L'absence de Rvb1-Rvb2 dans ce complexe conduit à un mauvais remodelage de la chromatine (Feng et al., 2003). SWR1 (*SWi2/snf2-Related 1*) est également un complexe de remodelage de la chromatine qui contient Rvb1 et Rvb2. Le complexe SWR1 est nécessaire au remplacement du dimère d'histones H2A-H2B dans le nucléosome, par un dimère d'histones variants Htz chez la levure et H2AZ chez les mammifères (Kobor et al., 2004; Mizuguchi et al., 2004). Ce mécanisme est essentiel pour les processus de régulation de la transcription, la ségrégation des chromosomes, la progression du cycle cellulaire et en réponse aux dommages liés à l'ADN. Le complexe Rvb1-Rvb2, en s'associant au complexe TIP60 (*Tat Interactive Protein of 60 kDa*), est essentiel pour l'acétylation de la protéine p53, qui est nécessaire à sa liaison aux gènes pro-apoptotiques (Nano and Houry, 2013). Rvb1 et Rvb2 participent également à l'assemblage des snoRNP (à boîtes C/D et à boîtes H/ACA) (Eickhoff and Costa, 2017; Vickers, 2017). Elles sont retrouvées chez l'homme, associées aux quatre protéines cœur de la particule snoRNP C/D (SNU13, Fibrillarine, NOP58 et NOP56) (McKeegan et al., 2009). Les données actuelles indiquent que les protéines Rvb1 et Rvb2 jouent un rôle central dans l'assemblage et le remodelage du complexe snoRNP C/D au cours de sa biogenèse, principalement en tant que composant du complexe R2TP. Elles constituent également des composantes essentielles des granules de stress cytoplasmique des mammifères et de levures (Narayanan et al., 2019 ; Zaarur et al., 2015). Le complexe RUVBL1-RUVBL2 jouent un rôle essentiel dans la voie de signalisation PIKK (*Phosphatidylinositol 3-kinase-related protein kinase*) (Izumi et al., 2010), ainsi que

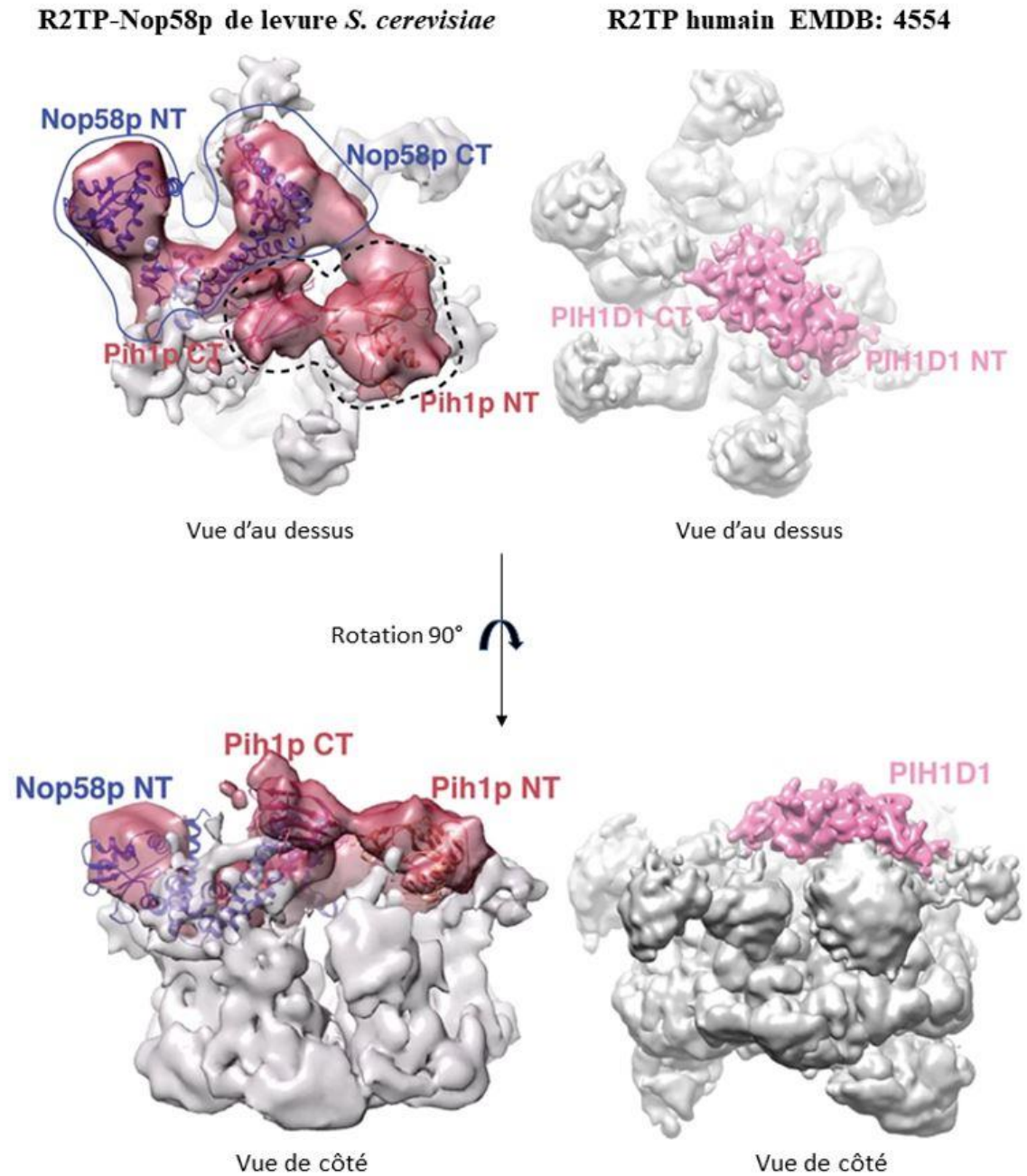


Figure 19 : Deux vues orthogonales des cartes de densité électronique du R2TP en complexe avec Nop58p₁₋₄₄₇ de levure *S. cerevisiae*, comparé au complexe R2TP humain obtenues en cryo-microscopie électronique (Yu et al., 2019).

Les domaines C-terminal et N-terminal de Nop58p sont représenté en bleu avec une superposition de la structure PDB 5WLC de Nop58p dans son potentiel emplacement sur la carte. Rvb1/Rvb2 est représenté en un anneau hexamérique en gris. Pih1p et PIH1D1 sont représentés en rose, la structure de Pih1p obtenue par cristallographie aux rayons X a été superposée à la carte de densité électronique.

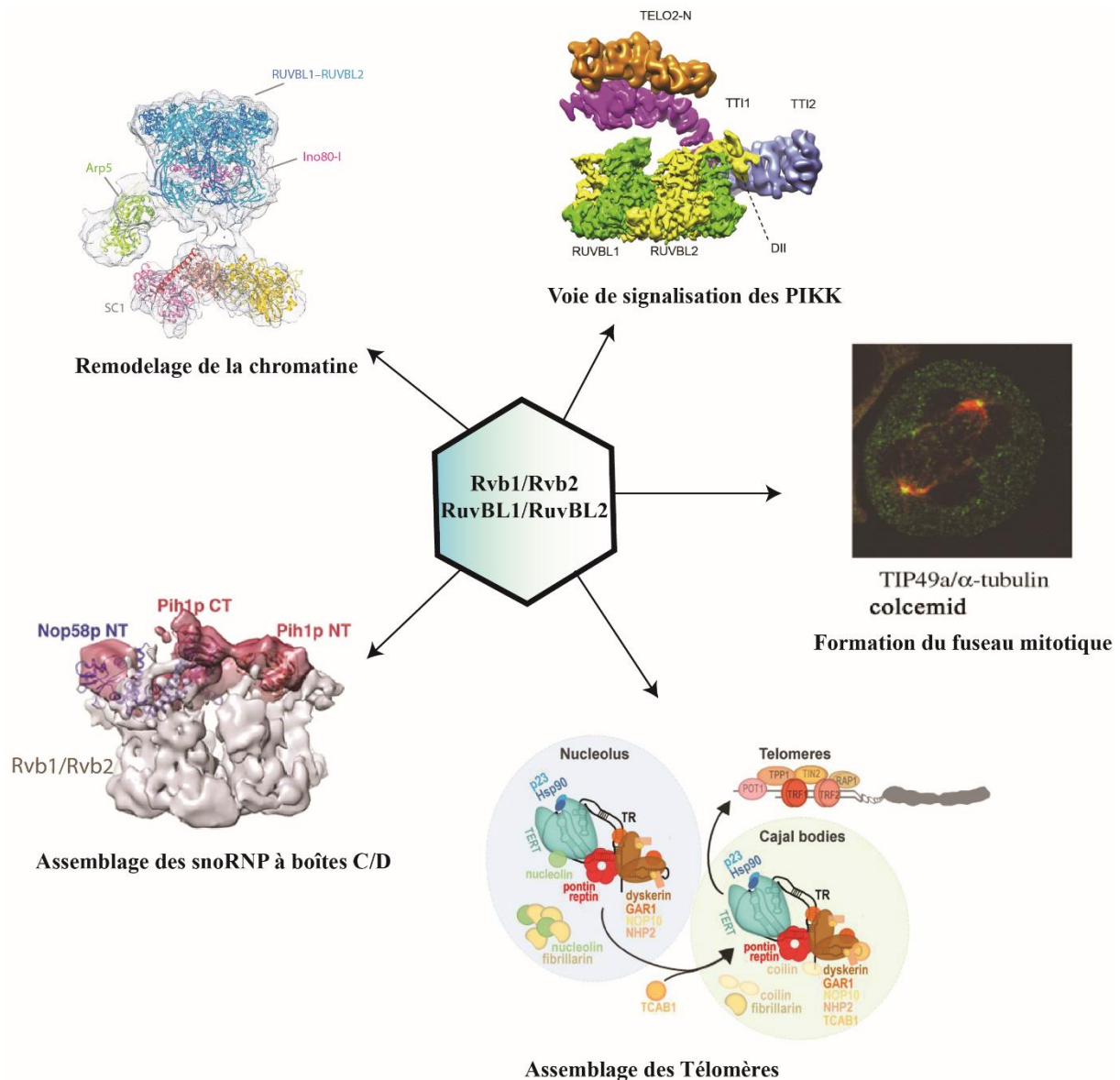


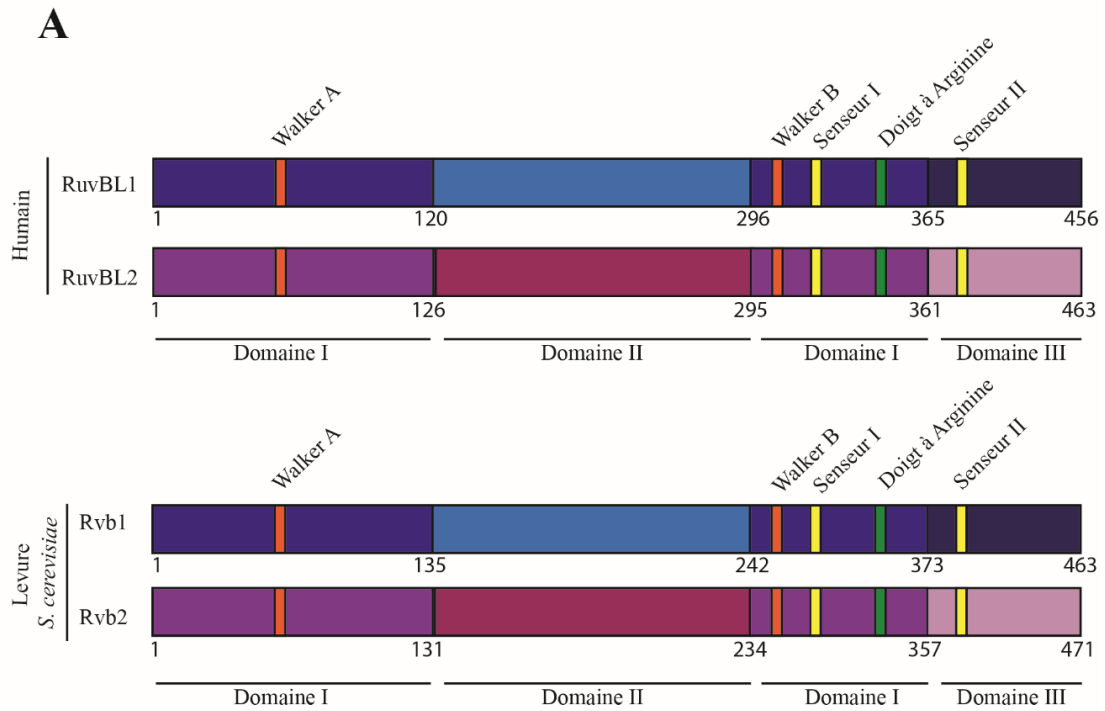
Figure 20 : Rvb1/Rvb2 et RUVBL1/RUVBL2 interviennent dans divers mécanismes moléculaires.

Ils participent à l'assemblage des snoRNP à boîtes C/D dans le cadre du complexe R2TP, et orchestrent la déposition de Nop58p aux sites d'assemblages (Yu et al., 2019). Ils interviennent dans les complexes macromoléculaires de l'assemblage du fuseau mitotique en s'associant avec la tubuline (Gartner et al., 2003). Nous les retrouvons également dans les complexes moléculaires de remodelage de la chromatine avec les complexes INO80 et SWR1 (Aramayo et al., 2018). Ils sont également présents aux sites de formation des télomères (Schořová et al., 2019).

l'assemblage de l'ARN polymérase II (Boulon et al., 2010). étude récente indique que Rvb1 et Rvb2 jouent un rôle dans le couplage de la transcription, la localisation cytoplasmique d'ARNm et la transcription de certains gènes spécifiques dans des conditions de privation de glucose chez la levure (Chen et al., 2021). Chez les mammifères, il a été montré que RUVBL1 et RUVBL2 sont liés à des mécanismes pro-tumoraux. Elles sont retrouvées surexprimées dans 80% des cancers du côlon, des poumons et dans des cellules du carcinome hépatocellulaire (Huber et al., 2008).

b. Structures 3D et organisations quaternaires

Rvb1 et Rvb2 de levure sont deux protéines orthologues avec 43% d'identité de séquence et 65% de similarité. Ce sont des protéines très conservées chez les eucaryotes, leurs homologues humains sont RUVBL1 (*Pontin/TIP49a*) et RUVBL2 (*Reptin/Tip49b*). Elles font parties de la famille des ATPases AAA+ (Neuwald et al., 1999). Rvb1 et Rvb2 contiennent des motifs de fixation et d'hydrolyse d'ATP structurellement définis (Matias et al., 2006). La séquence protéique des RUVBL/Rvb peut être divisée en trois domaines : DI, DII et DIII (Figure 21A). Les domaines DI et DIII se replient pour constituer le domaine AAA+ caractéristique de la famille des ATPases dont elles appartiennent. Le premier domaine (DI) est le domaine principal de l'interaction des nucléotides grâce au motif Walker A qui est caractérisé par la *P-loop* (*Phosphate binding loop*) $\alpha\beta\alpha$. Il consiste en cinq feuillets β entourés d'hélices α . Le motif Walker B et le doigt à arginine constituent également le domaine DI et sont responsables de l'hydrolyse de l'ATP. Le doigt à arginine établit des contacts avec le phosphate γ de l'ATP (Matias et al., 2006). Le domaine DII (qui se subdivise en DII_{int} et DII_{ext}) correspond à une insertion de 160 à 170 acides aminés dans le domaine DI entre les motifs *Walker A* et *Walker B*. Ce domaine est spécifique des protéines Rvb1/RUVBL1 et Rvb2/RUVBL2 comparées aux autres membres de la famille des AAA+. Le domaine DIII contient le sensor II. Avec le sensor I contenu dans le domaine DI, ils sont responsables de faire la distinction entre un nucléotide di- ou tri-phosphates fixé à la protéine (Figure 21B). (Snider et al., 2008). Les complexes Rvb1/Rvb2 et RUVBL1/RUVBL2 adoptent deux organisations quaternaires principales (un anneau hétéro-hexamérique composé de 3 Rvb1 ou RUVBL1 et 3 Rvb2 ou RUVBL2 (Figure 22A), un hétérododécamère formé par deux anneaux hétéro-hexamériques interagissant par les domaines DII (Figure 22B). Le domaine ATPase est situé à l'interface d'interaction du cycle oligomérique formé par Rvb1 et Rvb2. Cela suggère que l'hydrolyse efficace de l'ATP



B

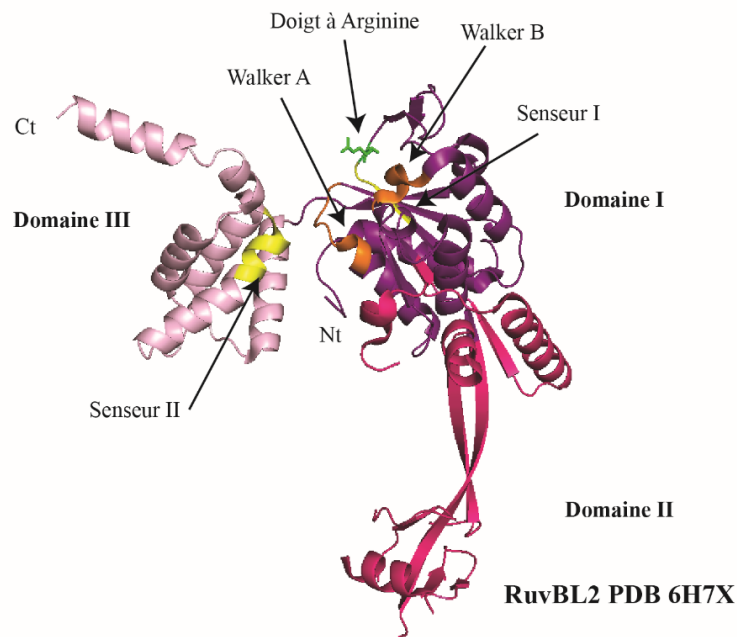


Figure 21 : Organisation des protéines Rvb1, RUVBL1 et Rvb2, RUVBL2.

(A) Les protéines humaines et de levure *S. cerevisiae* : Rvb1, RUVBL1 (Bleu) et Rvb2, RUVBL2 (Violet) sont organisées de façon similaires et se divisent en trois domaines principaux (DI, DII, DIII). Le domaine DI porte le Walker A, Walker B, doigt à arginine et le senseur I, le domaine DIII comprend le senseur II. Le domaine DII ou domaine d'insertion est retrouvé au milieu du domaine DI.

(B) Représentation 3D de la protéine RUVBL2 humaine résolue par cristallographie aux rayons X, code PDB 6H7X. Les domaines DI et DIII sont regroupés dans l'espace, tandis que le domaine DII prend une configuration allongée, essentiellement dû à ses longs feuillets β (Silva et al., 2018)

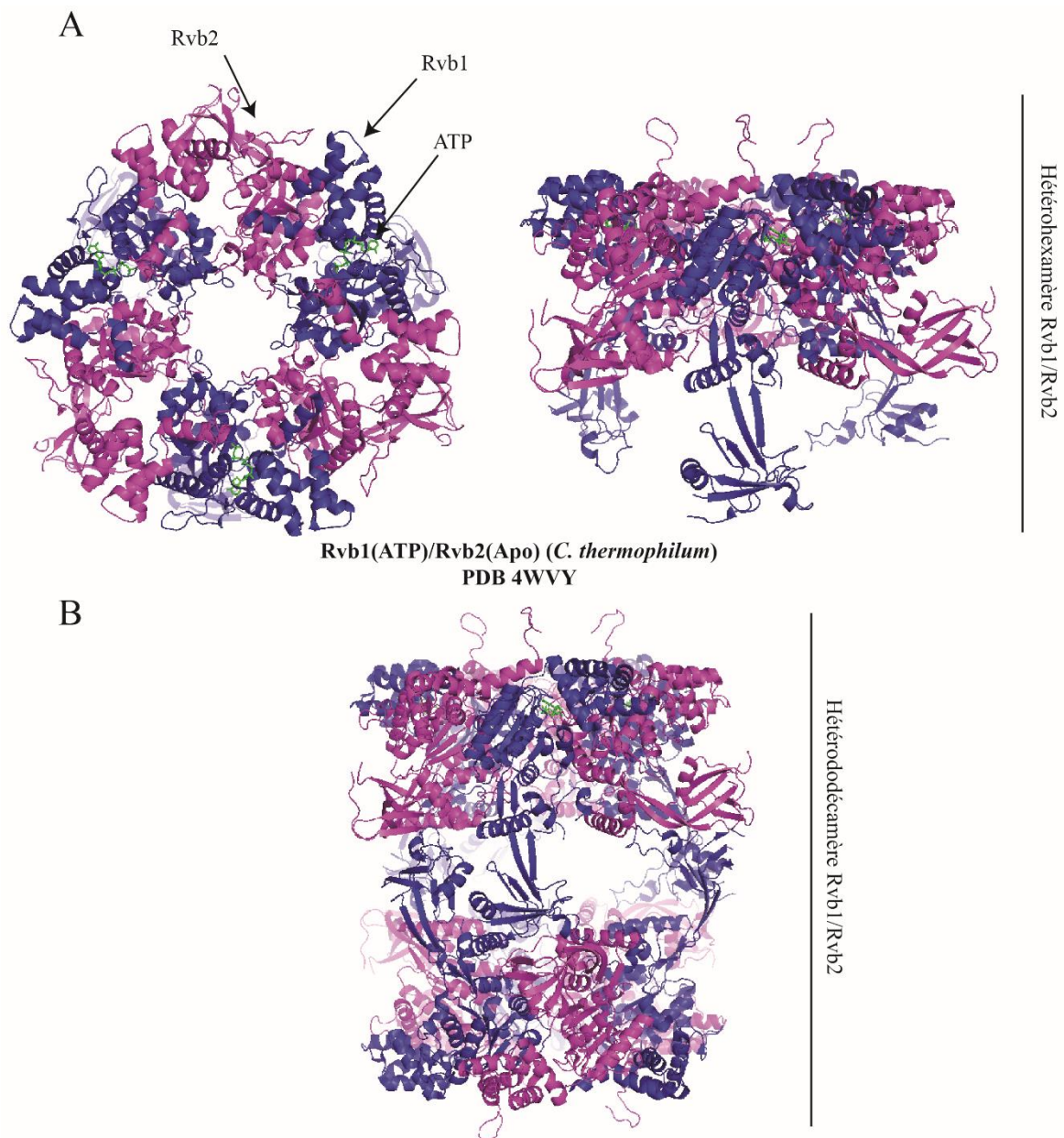


Figure 22 : Structures 3D du complexe Rvb1(Apo)/Rvb2(ATP) de *C. thermophilum* résolues par cristallographie aux rayons X, code PDB 4WVY (Lakomek et al., 2015).

Les structures quaternaires du complexe Rvb1/Rvb2 hexamérique (A) et dodécamérique (B) ont été reconstituées grâce à un logiciel de visualisation 3D (*Pymol*) par symétrie, à partir du complexe hétérodimérique Rvb1/Rvb2 (PDB 4WVY). La protéine Rvb1 contient dans sa poche catalytique une molécule d'ATP (vert) située à l'interface du dimère Rvb1/Rvb2. Trois protéines Rvb1 et trois protéines Rvb2 constituent un anneau hexamérique, avec les domaines DII projetés à l'extérieur de l'anneau. Ce sont ces derniers qui permettent l'organisation en dodécamère, ie deux hexamères Rvb1/Rvb2 qui interagissent par leurs domaines DII.

nécessite la coopération de plusieurs monomères. Les anneaux oligomériques Rvb1/Rvb2 subissent des changements de conformation au cours des cycles de liaison/hydrolyse de nucléotides. Ces changements seraient responsables du rôle de protéine chaperon des protéines Rvb1 et Rvb2 lors de l'assemblage de complexes multiprotéiques plus grands (Ewens et al., 2016; Gribun et al., 2008). L'utilisation des méthodes biophysiques et structurales dans différentes études, ont permis de montrer que Rvb1 ou RUVBL1 et Rvb2 ou RUVBL2 peuvent former de multiples oligomères, modulés notamment par leurs domaines d'insertion DII. Plus exactement les résidus arginine 180 de Rvb1 et acide glutamique 166 de Rvb2 de levure qui sont essentiels dans la dodécamérisation du complexe (Ewens et al., 2016), mais aussi par la présence d'une étiquette 6xhistidines en N-terminal chez Rvb1 et/ou Rvb2 de levure lors de leurs purifications (Cheung et al., 2010; Gribun et al., 2008; Puri et al., 2007). Ces observations semblent indiquées que les Rvbs sont capables de former différents oligomères selon le complexe ou le processus cellulaire dans lequel ils sont impliqués, mais aussi en fonction des cofacteurs qui peuvent moduler ainsi leur activité. En effet, il a été montré que le domaine d'insertion du complexe INO80 de levure qui se trouve en interaction avec le complexe Rvb1/Rvb2, est également responsable de la régulation de son activité ATPase tout en favorisant l'hétéro-dodécamérisation du complexe Rvb1/Rvb2 (Zhou et al., 2017). L'incubation *in vitro* de Rvb1 et Rvb2 de levure *S. cerevisiae* montre la formation de larges complexes qui a été montré par des expériences de sédimentation, comme étant des complexes en anneaux hétéro-hexamérique (Gribun et al., 2008), contrairement aux complexes RUVBL1/RUVBL2 qui forment des complexes hétéro-dodécamérique constitués d'un empilement de deux anneaux hétéro-hexamérique, interagissant par les domaines DII de chaque anneau (Puri et al., 2007). Cette différence d'organisation est très surprenante, en sachant que les homologues levure/humain Rvb1/RUVBL1 et Rvb2/RUVBL2 partagent 70% d'identité de séquence. Ces variations d'états oligomériques peuvent être affectés par le protocole de purification utilisé (étiquette 6xhistidines) (Tableau 1), mais aussi par l'ajout de nucléotides (ATP/ADP/ATP- γ S/AMP-PNP) dans le tampon de purification. Des expériences de spectrométrie de masse native, montre que la conformation adoptée par le complexe Rvb1/Rvb2 en présence de nucléotides (hydrolysable ou non) est un hexamère de Rvb1 et Rvb2. Toutefois, un intermédiaire de la réaction d'hydrolyse d'ATP par le complexe Rvb1/Rvb2 peut favoriser la formation de dodécamère (Zhou et al., 2017). L'ensemble des structures dodécamériques décrites dans la littérature ne sont pas toutes identiques. Certaines particules compactes ou étirés (Figure 23A) dans le sens de la hauteur ont pu être observées avec deux anneaux hexamériques

Protéines étudiés	Organisme	Étiqueté / Non-étiqueté	Etat oligomérique	Activité ATPase	Technique utilisée	Références
RUVBL2	Humain	Flag-His ₆ -RUVBL2	RUVBL1/RUVBL2 en dodécamère	Actif (stimulé par l'ADNsb)	Filtration sur gel	(Kanemaki et al., 1999)
RUVBL1	Humain	Flag-His ₆ -RUVBL1	- RUVBL1/RUVBL2 en complexe hétérohexamère - Interaction avec l'ADNsb et l'ADNdb	Inactif	Cristallographie aux rayons X	(Matias et al., 2006)
RUVBL1 RUVBL2	Humain	- His ₆ -RUVBL1 - RUVBL2-His ₆ - Co-expression : RUVBL1/His ₆ -RUVBL2	- Monomère RUVBL1 - Monomère et hexamère RUVBL2 - Dodécamère RUVBL1/RUVBL2 (interaction par les domaines DII)	Actif (Non stimulé par l'ADNsb ni l'ADNdb)	- Filtration sur gel - Marquage négatif	(Puri et al., 2007)
Rvb1 Rvb2	Levure (<i>S. cerevisiae</i>)	Non-étiqueté	- Monomère et hexamère Rvb1 - Monomère et hexamère Rvb2 - Hexamère Rvb1/Rvb2	Actif (stimulé par par ADNdb)	- Filtration sur gel - Microscopie électronique	(Gribun et al., 2008)
Rvb1 Rvb2	Levure (<i>S. cerevisiae</i>)	Rvb1/His ₆ -Rvb2	Dodécamère (double anneaux hexamériques interagissant par les domaines DII)	Actif	- Marquage négatif - Cryo-EM	(Torreira et al., 2008)
Rvb1 Rvb2	Levure (<i>S. cerevisiae</i>)	- Rvb1/His ₆ -Rvb2 - His ₆ -Rvb1/Rvb2 - Rvb1/Rvb2	- Dodécamère (His ₆ -Rvb1/Rvb2 et Rvb1/His ₆ -Rvb2) - Hexamère Rvb1/Rvb2		- Marquage négatif - Purification de complexe endogène	(Cheung et al., 2010)
Rvb1 Rvb2	Levure (<i>S. cerevisiae</i>)	- Non étiqueté	- Dodécamère + Hexamère (sans nucléotides) - Hexamère (ADP) - Hexamère (ATP) - Hexamère (ATP γ S)	Faible (complexe dodécamère Rvb1/Rvb2) Forte (Complexe hexamère Rvb1/Rvb2)	- Purification de complexe endogène - Filtration sur gel - Marquage négatif	(Jeganathan et al., 2015)
Rvb1 Rvb2	Champignon thermophile (<i>C. thermophilum</i>)	- Rvb1/His ₈ -Rvb2	- Dodécamère (flexibilité des domaines DII selon le nucléotide ajouté)	Actif	- SEC, SLS, SAXS, XL-MS, cristallographie aux rayons X	(Lakomek et al., 2015)
Rvb1 Rvb2	Levure (<i>S. cerevisiae</i>)	- Rvb1/Rvb2 - His ₆ -Rvb1/Rvb2 - Rvb1/His ₆ -Rvb2	- Hexamère (Rvb1/Rvb2) - Hexamère (Rvb1/Rvb2 +ATP) - Dodécamère (Rvb1/His ₆ -Rvb2) - Dodécamère (Rvb1/His ₆ -Rvb2 +ATP) - Hexamère (Rvb1/His ₆ -Rvb2 +ADP)	Actif (Stimulé par INO80 et la dodécamérisation)	- SEC, MS native, XL-MS - Cryo-microscopie électronique	(Zhou et al., 2017)

			- Hexamère (Rvb1/His ₆ -Rvb2 +AMP-PNP/ATP γ S)			
Rvb1 Rvb2	Champignon filamenteux <i>N. crassa</i>	- Rvb1/ His ₆ -Rvb2	- Hexamère (Rvb1/His ₆ -Rvb2) sans nucléotides - Hexamère (Rvb1/His ₆ -Rvb2) en présence d'ADP/ATP	Actif (Stimulé par l'ADNsb)	- Filtration sur gel - Dichroïsme circulaire - Mobilité électrophorétique	(Maimoni Campanella et al., 2021)

Tableau 1 : Données structurales et fonctionnelles de Rvb1, Rvb2 et RUVBL1, RUVBL2 (adapté depuis Nano, 2018).

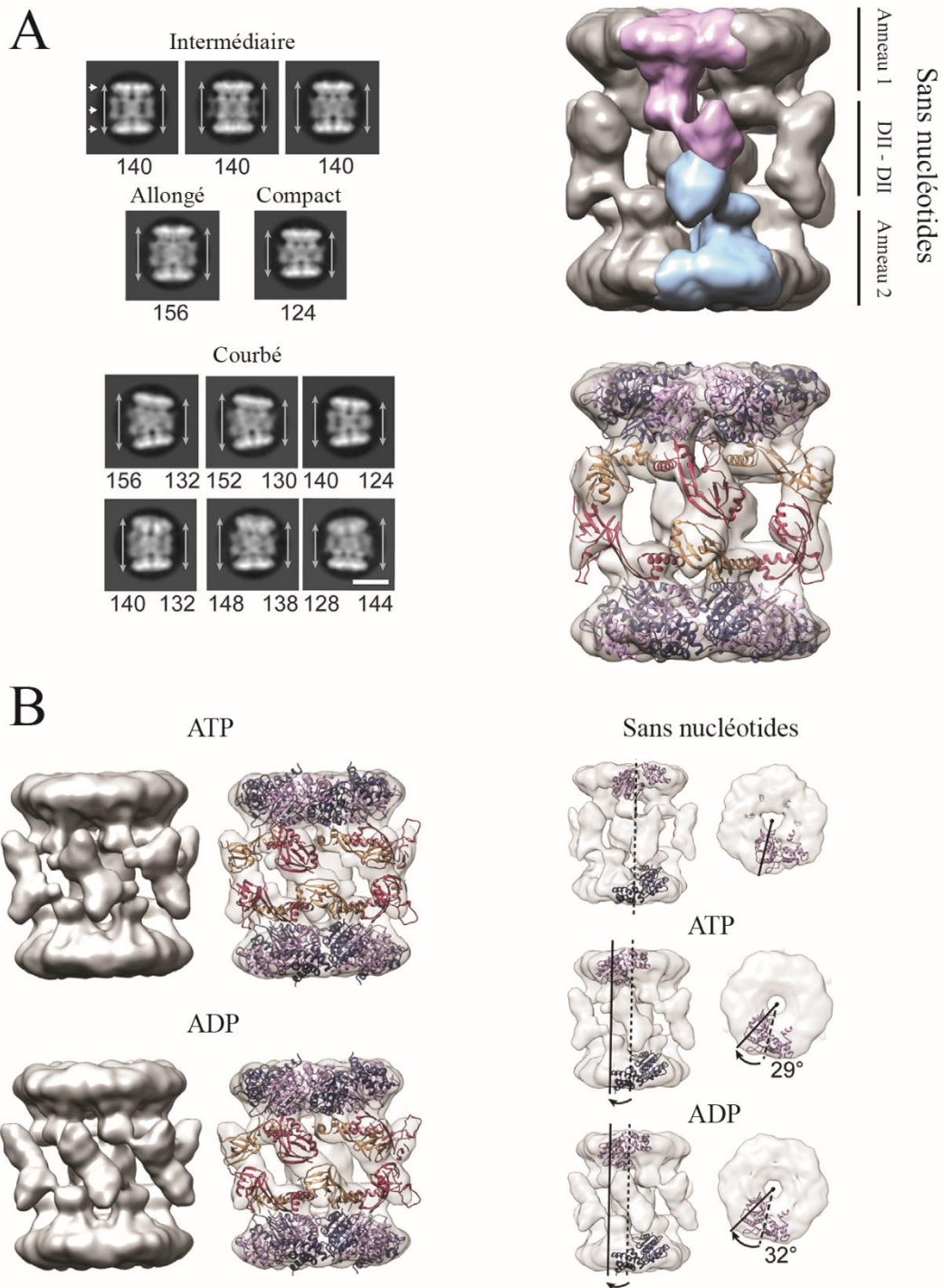


Figure 23 : Analyse des changements conformationnels du complexe dodécamérique Rvb1/Rvb2 de levure (*S. cerevisiae*) selon l'état nucléotidique : sans nucléotides (Apo), ADP ou ATP (Ewens et al., 2016).

(A) Les images obtenues par microscopie électronique après marquage négatif (gauche), correspondent aux conformations plus ou moins allongées du complexe dodécamérique Rvb1/Rvb2 sans nucléotides. Les anneaux hexamériques du dodécamère (droite) sont parfaitement alignés et parallèles.

(B) Les conformations spatiales adoptées par le complexe dodécamérique en présence de nucléotides (ADP, ATP) diffèrent de celles sans nucléotides. Nous constatons une rotation des deux anneaux hexamériques, avec une inclinaison des domaines DII les reliant. Une rotation de 32° en présence de d'ADP et 29° en ATP.

parfaitement alignés et parallèles (Jeganathan et al., 2015), mais aussi un dodécamère avec les deux anneaux hexamériques non parallèles, courbés ou décalés l'un par rapport à l'autre (Ewens et al., 2016; Silva-Martin et al., 2016). L'ajout d'ADP ou d'ATP au complexe Rvb1/Rvb2 (dodécamérique) montre également des changements conformationnels importants comparés à l'état apo du dodécamère. Une rotation des deux anneaux hexamériques de 29° pour l'ATP et 32° pour l'ADP, indique que la flexibilité du dodécamère, essentiellement permis par les domaine DII, est essentielle dans l'hydrolyse d'ATP (Ewens et al., 2016) (Figure 23B).

1.3.2.2. La protéine Tah1

Tah1 (*TPR containing associated with HSP90*) est une petite protéine de 12,5 kDa. Elle a tout d'abord été mise en évidence grâce à un crible des partenaires de HSP90 (Zhao et al., 2005). Tah1p est donc un cofacteur qui module l'activité de Hsp82, en interagissant également avec une autre protéine nommée Pih1. Le complexe formé entre HSP90/Tah1p/Pih1p serait responsable de la stabilisation de la protéine Pih1, et peut former le complexe R2TP en s'associant avec les complexes Rvb1/Rvb2. Deux motifs TPR de 34 acides aminés chacun (Figure 24A) sont retrouvés dans la structure 3D de Tah1p au niveau de son extrémité N-terminale. Plusieurs protéines sont connues pour interagir avec HSP90 et HSP70, plus particulièrement avec les résidus EEVD à l'extrémité C-terminale des protéines chaperon (Scheufler et al., 2000b). Le domaine TPR de la Tah1p est particulier comparé aux domaines TPR retrouvés chez d'autres protéines. Ce dernier est composé de seulement deux motifs TPR plus une hélice dite de « capping » soit cinq hélices α en tout, contre sept habituellement. Les résidus à l'extrémité C-terminale de Tah1p (76-111) sont déstructurés dans le cadre de la protéine libre. Ils sont responsables de l'interaction avec Pih1p, où ce domaine apparaît structuré dans le complexe Tah1p-Pih1p en formant un feuillet β intermoléculaire. RPAP3 (*RNA polymerase II-associated protein 3*) est une protéine homologue à Tah1p chez l'homme. Elle est bien plus longue (75 kDa) et possède un domaine TPR supplémentaire une région de reconnaissance à PIH1D1 (orthologue à Pih1p chez l'Homme) (Morgan et al., 2015) dont la structuration 3D dépend de son interaction. Les deux domaines TPR de RPAP3 présentent chacun 7 hélices α . La structure 3D du domaine C-terminal de RPAP3₅₃₅₋₆₆₅ a été résolue par RMN au laboratoire (Maurizy et al., 2018). Il est composé de 8 hélices α regroupées (Figure 24B), formant une structure globulaire compacte. Les résidus 541-548 forment une boucle qui protège les acides aminés hydrophobes du solvant. Cette boucle est importante pour la

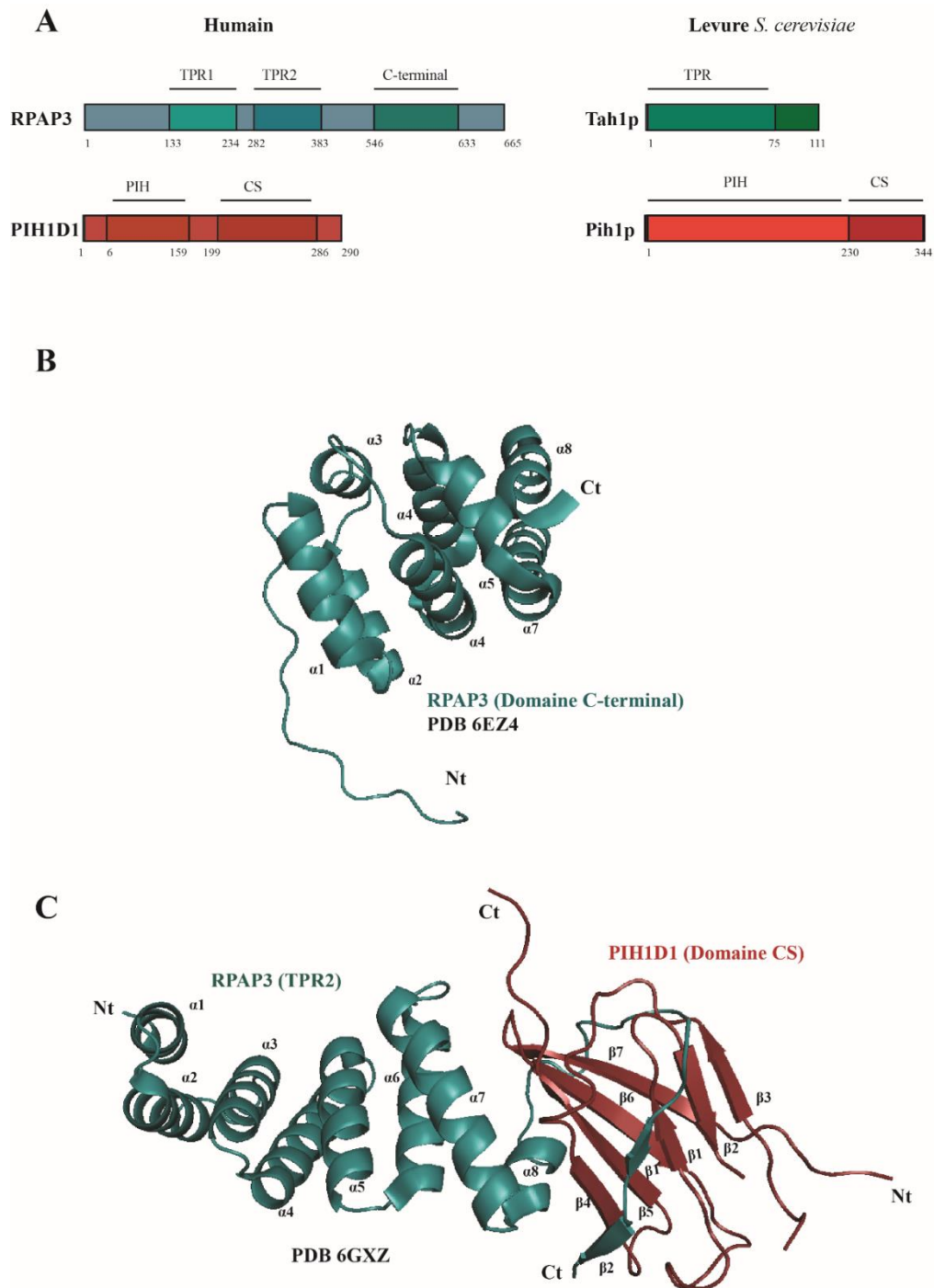


Figure 24 : Organisations des protéines humaines RPAP3 et PIH1D1.

(A) La protéine RPAP3 et son homologue de levure Tah1p, sont organisées en domaines TPR (deux domaines TPR pour RPAP3 et un pour Tah1p). RPAP3 est plus longue que Tah1p et présente un domaine C-terminal supplémentaire essentiel. La protéine PIH1D1 et son homologue de levure Pih1p sont organisées de façon similaires, en deux domaines : Domaine PIH et le domaine CS (*CHORD Sgt1*).

(B) Structure 3D du domaine C-terminal de RPAP3 résolu par RMN (PDB 6EZ4). Il présente une structure globulaire constituée de 8 hélices α ainsi qu'une longue boucle N-terminale projeté à l'extérieur (Maurizy et al., 2018).

(C) Structure du complexe PIH1D1₁₉₉₋₂₉₀/ RPAP3₂₈₁₋₄₄₅ résolue par cristallographie aux rayons X (PDB 6GXZ) (Henri et al., 2018).

solubilité du domaine C-terminal de RPAP3 (Maurizy et al., 2018). Une interaction entre la protéine RPAP3 et MONAD, une protéine impliquée dans la voie l'apoptose a été identifiée (Itsuki et al., 2008). Cette interaction met en évidence un rôle potentiel de RPAP3 dans les processus de dégradation ubiquitine dépendante par le protéasome (Itsuki et al., 2008).

Une récente étude montre que dans le cas du prion chez *S. cerevisiae* [URE3], la protéine Tah1 est retrouvée en interaction avec la protéine Ure2, essentielle pour la propagation du prion. La surexpression de Tah1p conduit à une déstabilisation du prion [URE3] indépendamment du complexe R2TP (Puri et al., 2021).

1.3.2.3. La protéine Pih1

La protéine Pih1 de levure (*Protein interacting with Hsp82*) (40 kDa) a été mise en évidence comme partenaire de la protéine chaperon Hsp82. Il a été montré que Pih1p est capable d'interagir avec Hsp82 même en absence de Tah1p (Zhao et al., 2008). Elle est constituée de deux régions intrinsèquement désordonnées essentielles pour établir des interactions avec d'autres partenaires protéiques (Rvb1/Rvb2 et Tah1p). PIH1D1 (32 kDa), anciennement nommée Nop17p est l'homologue humain de Pih1p. Elle a tout d'abord été caractérisée comme partenaire de la protéine Nop58/NOP58 dans le nucléole (Gonzales et al., 2005). Cette interaction fait intervenir le domaine NOP (constitué de la série de répétition KKE/D) à l'extrémité C-terminale de Nop58/NOP58 (Yu et al., 2019). C'est une interaction qui confirme l'implication de Pih1p dans l'assemblage des snoRNP à boîtes C/D. Elle serait également en compétition avec l'interaction de Nop58p avec le snoARN C/D (Gonzales et al., 2005). La région en position N-terminale de Pih1p/PIHD1 est constituée d'un domaine PIH (Figure 24A) qui a été retrouvé en interaction avec la protéine Tel2/TEL2 impliquée dans la stabilisation ainsi que l'activation des PIKK ou de MONAD (Kakihara and Houry, 2012). La région C-terminale de Pih1p est repliée en un domaine dit CS (*CHORD and Sgt1*) caractérisé par 7 brins β en sandwich (Figure 24C) essentiel pour son interaction avec Tah1p (Pal et al., 2014; Quinternet et al., 2015). Le même domaine CS est retrouvé chez son homologue humain PIH1D1. Une structure du complexe PIH1D1₁₉₉₋₂₉₀/RPAP3₂₈₁₋₄₄₅ a été résolue par cristallographie aux rayons X (Figure 24C). L'interface d'interaction du complexe fait intervenir un ensemble d'interactions hydrophobes et polaires, établi entre les hélices $\alpha 7$ et $\alpha 8$ de RPAP3 et les feuillettes $\beta 1$, $\beta 2$, $\beta 6$ et $\beta 7$ du domaine CS de PIH1D1 (Henri et al., 2018). Une étude par cryo-microscopie électronique et *cross-link* a permis de mettre en évidence la structure globale de Nop58 (1-447) de levure (biochimiquement stable *in vitro*) avec le complexe R2TP. Il apparaît de cette étude

que le domaine PIH de Pih1p contacte à la fois l'anneau hexamérique Rvb1/Rvb2 *via* leurs DII, mais aussi l'extrémité C-terminale de Nop58 (Yu et al., 2019).

1.3.3. Le complexe R2SP chez l'homme, homologue du R2TP

Le complexe R2SP (RUVBL1/RUVBL2/SPAG1/PIH1D2) est orthologue au complexe R2TP, avec le remplacement des protéines adaptatrices RPAP3 et PIH1D1 par respectivement SPAG1 (*Sperm associated antigen 1*) et PIH1D2 (Maurizy et al., 2018). R2TP et R2SP sont tous deux cofacteurs de protéines chaperon qui fonctionnent avec les protéines chaperon HSP70 et HSP90, facilitant ainsi l'organisation quaternaire des protéines clientes spécifiques de chaque complexe. De plus, l'expression du complexe R2SP a été montré comme étant tissu spécifique. Il serait enrichi dans les testicules qui nécessite une température de 30°C (Maurizy et al., 2018). La structure du complexe R2TP a été largement étudiée par RMN, cristallographie aux rayons X et cryo-microscopie électronique. Il n'y a pas de données tridimensionnelles à haute résolution pour le complexe R2SP à ce jour. Alors que PIH1D1 et PIH1D2 sont tous deux composés d'un domaine PIH et d'un domaine CS, SPAG1 et RPAP3 partagent un domaine C-terminal conservé ainsi que plusieurs domaines TPR (Maurizy et al., 2018) (Figure 25A). En raison de leurs similitudes organisationnelles, RPAP3/PIH1D1 et SPAG1/PIH1D2 pourraient partager des fonctions communes. Cependant, il a été proposé que SPAG1 possède une activité d'hydrolyse GTP spécifique, apportant éventuellement l'énergie nécessaire à sa fonction de cofacteur de protéine chaperon (Lin et al., 2001). Parmi les clients du R2SP, la liprine- α 2 semble un candidat majeur. Un enrichissement du complexe R2SP dans des cellules Hela conduit à une surexpression de la liprine- α 2 à 30°C, la température testiculaire. Il s'agit d'une protéine d'échafaudage s'associant à plusieurs facteurs protéiques aux rôles très variés : adhésion et migration cellulaires, organisation de la zone synaptique active (Maurizy et al., 2018). SPAG1 est une protéine de 104 kDa, homologue à la protéine RPAP3. Elle contient trois domaines TPR et un domaine RPAP3 C-terminal (Figure 25A). SPAG1 est une protéine qui fait partie de la famille des facteurs protéiques DNAAF (*Dynein Axonemal Assembly Factors*). Elles sont responsables de l'assemblage des bras de dynéine constituant les cils. Des mutations dans le gène de SPAG1 seraient responsables d'une dyskinésie ciliaire primitive (DCP), une maladie génétique rare causant des troubles respiratoires ainsi que des problèmes de fertilité chez l'Homme. La DCP se caractérise par une mauvaise formation des cils mobiles, précisément des bras de dynéines, permettant aux cellules flagellées d'être mobiles (Knowles et al., 2013).

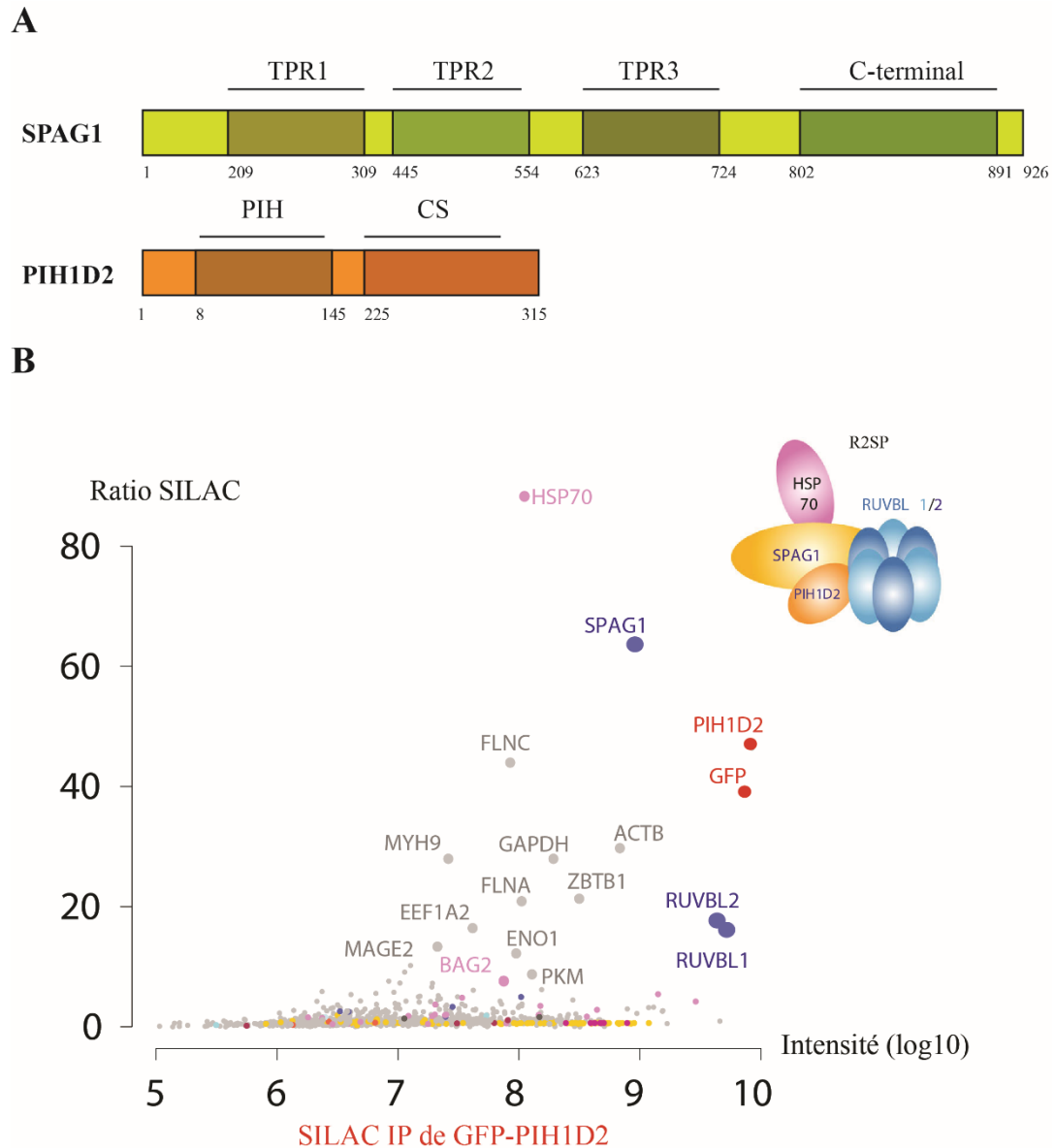


Figure 25 : Mise en évidence d'un complexe R2TP-like : R2SP (RUVBL1/RUVBL2/SPAG1/PIH1D2) (Maurizy et al., 2018).

(A) La protéine SPAG1 homologue de RPAP3, est organisée en trois domaines TPR, et un domaine C-terminal. La protéine PIH1D2 homologue de PIH1D1 comporte deux domaines : CS et PIH.

(B) *Stable isotope labeling by amino acids in cell culture* (SILAC), est une méthode de protéomique quantitative utilisant la spectrométrie de masse. Le SILAC de la protéine marquée GFP-PIH1D2 dans des cellules HeLa indique un enrichissement des protéines SPAG1, HSP70 et RUVBL1/RUVBL2, ce qui suggère l'existence d'un complexe R2TP-like nommé R2SP en association avec HSP70.

SPAG1 est également impliquée dans l'activation du signal extracellulaire régulé par la protéine kinase G (Liu et al., 2006). Pour finir, elle joue également un rôle dans les voies dépendantes des nucléotides, telle que la kinase dépendante de l'AMP et les kinases C des voies de signalisation (Huang et al., 2016). Des expériences d'IP LUMIER ont montré que la surexpression de PIH1D2 améliore l'affinité de SPAG1 pour le complexe RUVBL1/RUVBL2. Cette affinité d'interaction est similaire à celle de RPAP3 pour RUVBL1/RUVBL2 (Maurizy et al., 2018).

Le génome humain code quatre protéines avec un domaine PIH : PIH1D1, PIH1D2, PIH1D3 et DNAAF2. De façon intéressante, la duplication du gène ancestral de RPAP3 pour donner SPAG1 est suivie d'une duplication similaire de l'ancêtre de PIH1D1, générant PIH1D1 et PIH1D2 (Yamamoto et al., 2010). En plus de SPAG1 et RUVBL1/RUVBL2, d'autres partenaires sont en associations avec PIH1D2 en IP-LUMIER. Il s'agit des protéines impliquées dans divers processus cellulaires : le métabolisme de l'ADN, la transcription et la maturation de l'ARN (Maurizy et al., 2018).

2. Modèle d'assemblage des snoRNP à boîtes C/D chez la levure *S. cerevisiae*

L'ensemble des études fonctionnelles et structurales de la décennie ont permis de proposer des modèles de l'assemblage des snoRNP à boîtes C/D, ainsi que de suggérer des fonctions pour chacune des protéines impliquées dans les différentes étapes de la biogenèse (Bizarro et al., 2014; Boulon et al., 2008; Kakihara et al., 2014a; McKeegan et al., 2007, 2009; Paul et al., 2019b; Zhao et al., 2008). Toutefois, de nombreux points sont encore à élucider, à savoir, à quelles étapes les différents facteurs d'assemblage interviennent, avec quels partenaires et pour assurer quelle(s) fonction(s) précise(s). Il est à noter qu'il subsiste quelques différences dans l'assemblage de snoRNP fonctionnelles entre la levure et l'Homme. Dans cette étude, nous nous intéresserons plus particulièrement aux étapes de cette machinerie complexe chez la levure.

Plusieurs résultats suggèrent un assemblage des snoRNP de façon co-transcriptionnelle. En effet, Nop1, Nop58p et Bcd1p ont été détectées par ChIP (*Chromatin ImmunoPrecipitation*) chez la levure sur les loci codant les snoARN à boîtes C/D mais également sur quelques snoARN naissants (Paul et al., 2019b; Vincenti et al., 2007). De plus, certains facteurs d'assemblage, tel que les protéines Rsa1/NUFIP1 et Hit1/ZNHIT3 sont requis pour le recrutement des protéines cœur sur les snoARN, et plus particulièrement la protéine Snu13, au sein d'un complexe protéique SRH (Snu13-Rsa1-Hit1) (Rothé et al., 2014b). Le dépôt de

Snu13p sur le snoARN constitue la première étape pour le recrutement des autres protéines cœur au sein de la pré-particule snoRNP (Bizarro et al., 2014; Boulon et al., 2008; Gonzales et al., 2005; Kakihara et al., 2014b; King et al., 2001). Le complexe R2TP est décrit comme étant requis pour le recrutement de Nop58p dans les complexes de pré-assemblages de snoRNP à boîtes C/D (Gonzales et al., 2005; Kakihara et al., 2014b; Zhao et al., 2008).

Objectifs

Objectifs

Le mécanisme de recrutement des facteurs d'assemblage et des protéines cœur sur les snoARN lors de la formation des pré-complexes d'assemblage est encore à élucider. Il est possible que plusieurs sous-complexes protéiques soient recrutés dans un ordre précis, ou bien qu'un seul pré-complexe contenant tous les facteurs protéiques requis pour l'assemblage des particules soit formé. Par ailleurs, l'étude de la formation des pré-particules constitue une étape clé pour la suite de la compréhension de leur relargage de la particule mature et fonctionnelle. Ces mécanismes de désassemblage et de recyclage des snoRNP n'ont pas encore été étudiés.

Le réseau d'interactions mis en évidence entre les différentes protéines de la machinerie d'assemblage et les protéines cœur des snoRNP à boîtes C/D suggère que la formation de complexes multiprotéiques est nécessaire, réalisée de manière séquentielle, ordonnée et contrôlée dans la cellule par l'intervention de plusieurs facteurs protéiques.

Plusieurs données de la littérature indiquent que Bcd1p est nécessaire au dépôt de la protéine Nop58 sur la pré-snoRNP incluant au moins Snu13p, ce qui constitue une étape essentielle dans ce mécanisme d'assemblage. Nous nous sommes intéressés au laboratoire à l'étude de la région N-terminale de Bcd1p chez la levure. Comme nous l'avons décrit dans le chapitre d'introduction, cette région de Bcd1p contient le fragment minimum 1-96 qui assure une croissance normale des cellules. Elle est donc le siège de l'ensemble des interactions essentielles à la fonction de Bcd1p dans l'assemblage des snoRNP à boîtes C/D. Un alignement du fragment 45-96 de Bcd1p chez *S. cerevisiae*, avec plusieurs mêmes fragments de Bcd1p chez différentes espèces de levure, indique la conservation stricte de deux acides aminés au cours de l'évolution : l'Aspartate 72 et l'Arginine 81.

La mutation de l'Aspartate 72 en Alanine a pu facilement être obtenue. Nous avons voulu connaître l'effet de cette mutation ponctuelle sur l'assemblage dans la cellule. Son expression transitoire chez *S. cerevisiae* nous permettrait ainsi de mieux caractériser le rôle joué par Bcd1p, à travers l'analyse des associations protéine : protéine et snoARN : protéine. Cette étude sera facilitée par l'utilisation de souche de levure qui seront générées au laboratoire, exprimant les facteurs d'assemblage et des protéines cœur fusionnées à des étiquettes, permettant ainsi de les isoler par immunoprécipitation et de les détecter par *Western-Blot*.

Le complexe R2TP a fait l'objet de plusieurs études, dans le but de mieux caractériser et de dissocier le rôle joué ainsi que l'apport de chacun de ses constituants (Rvb1/Rvb2, Tah1p et Pih1p) dans la fonction finale du complexe. Une partie de ce travail de thèse a été dédiée à la

compréhension du lien fonctionnel, biochimique et structural entre Bcd1p et le complexe Rvb1/Rvb2. Des travaux menés au laboratoire indiquent que le ZHD de Bcd1p (1-45) en région N-terminale est suffisant pour interagir avec notamment avec les RUVBL1/RUVBL2 humaines, toutefois nous ignorons la contribution de ce complexe ATPasique dans les étapes de l'assemblage (changements conformationnels et activation de protéine, acheminements et libérations de cibles protéiques aux sites d'assemblage, dissociation de partenaire protéique, etc). La complexité de son étude, comme pour la plupart des facteurs d'assemblage des snoRNPs à boîtes C/D connus, vient du fait que Rvb1/RUVBL1 et Rvb2/RUVBL2 opèrent dans plusieurs voies cellulaires différentes comme mentionné précédemment, avec au moins deux stœchiométries différentes (complexe hétérohexamérique ou et hétérododécamérique). L'activité ATPasique du complexe protéine co-chaperon Rvb1/Rvb2 et ses différentes conformations qui en découlent selon la nature des nucléotides fixés sont vraisemblablement liées à la fonction de Bcd1p dans l'assemblage. Une partie de ce travail est donc dédiée à l'étude du complexe Rvb1/Rvb2/Bcd1p *in vitro* par des techniques biochimiques (chromatographie de filtration sur gel), biophysiques (ITC) et structuraux (cryo-microscopie électronique). La compréhension structurale de ce complexe, permettra d'apporter des éléments de réponses sur le mécanisme du rôle joué par Bcd1p.

Une autre partie de cette étude a été consacrée à un complexe R2TP-like (R2SP) qui comme pour le R2TP est un complexe protéique aidant l'assemblage quaternaire de clients protéiques spécifiques s'associant aux protéines chaperons HSP90 et HSP70. La compréhension de ce complexe R2TP-like apporterait des informations supplémentaires pour la compréhension de la fonction du R2TP. Dans un premier temps, nous avons étudié le rôle des domaines TPR de SPAG1 dans le recrutement des protéines chaperons HSP par le biais d'essais biochimiques (*pull down*) et biophysiques (ITC), de spectroscopie Résonance Magnétique Nucléaire (RMN) et des simulations de dynamique moléculaire (Publication N°01).

Le domaine TPR1 de SPAG1 a été plus particulièrement étudié, utilisant une approche d'optimisation de structure protéique par mutations (en combinaison avec des approches structurales et biophysiques). L'analyse à l'échelle atomistique de l'interaction entre le fragment en position C-terminale des HSP70 et HSP90, et une forme variante du TPR1 de SPAG1 a été réalisée. Cette étude a pour but de caractériser les acides aminés dans les motifs TPR responsables du positionnement des peptides HSP (Publication N°2).

*Matériels et
Méthodes*

1. Matériels

1.1. Oligonucléotides

L'ensemble des amorces désoxyribonucléotidiques utilisé durant ma thèse est répertorié dans le tableau 1, avec le numéro du tube, la séquence nucléotidique et l'usage de chacun.

1.2. Vecteurs d'expression

1.2.1. Plasmides bactériens

- Le plasmide pRARE2 (Figure M1A) exprime les 7 ARNt (Arg, Pro, Leu, Met, Thr, Gly, Ile) faiblement utilisés par la bactérie comparée aux eucaryotes et ainsi il évite une saturation du système de traduction des ARNm issus d'organismes eucaryotes. Il est résistant au chloramphénicol, avec une origine de réplication de type p15a ori.
- Le plasmide pnEA3CH (Figure M1B): est utilisé pour la surexpression des protéines recombinantes avec l'étiquette 6xHis chez *E. coli*. Il est résistant à l'ampicilline et possède une origine de réplication de type pRB322. Le gène est sous le contrôle d'un promoteur inductible à l'IPTG, celui de l'ARN polymérase du bactériophage T7. Nous retrouvons une étiquette 6 x Histidines en amont de la cassette de clonage, avec une séquence qui code pour un site de coupure à la PreScission (3C), entre l'étiquette et le gène d'intérêt.

1.2.2. Plasmides de levure

- Le plasmide pFA6a-KanMX6-pGAL1-3HA (Figure M2A): est utilisé dans la levure *S. cerevisiae* pour exprimer de façon ectopique et transitoire la protéine Bcd1, grâce à l'utilisation d'un promoteur conditionnel pGAL inductible par le galactose. La protéine Bcd1 est par ailleurs produite en fusion avec une étiquette 3HA côté N-terminal. Il est résistant à la Kanamycine (G418).
- Le plasmide p416GPD 3x FLAG (Figure M2B): est utilisé pour l'expression ectopique des protéines de levure, en fusion avec une étiquette 3x FLAG en position N-terminale. Il comporte une origine de réplication CEN/ARS, un gène de résistance à l'ampicilline ainsi qu'un marqueur d'auxotrophie URA3. Il possède un promoteur fort issu de la glycéraldéhyde 3-phosphate déshydrogénase (GAPDH).

N°	Séquence 5'→3'	Usage
E741	TCCCGGACTATGCAGGATCCTATCCATATGACGTTCCAGATT ACGCTATGGCGGTGTTGTGTGGTGTATGC	Oligo 5' de Clonage de l'ORF BCD1 de levure dans le vecteur d'expression pFA6a
E740	GCGCCTCAGCACTGAGCAGCGTAATCTGGAACGTCATATGGA TAGGATCCTCATGCAGTGAGGAAATCCA	Oligo 3' de Clonage de l'ORF BCD1 de levure dans le vecteur d'expression pFA6a
F692	AAATCAGCAGACCCTGATGCCATGGATACTACGGA	Oligo 5' interne de RVB2 de levure (utilisé pour le séquençage de la concaténation pnEA3CH Rvb2 Rvb1)
F604	AGGGCTGTTTGGTAGCGGGCCCCCTCAACTGGTGAGACC GCTCTTGCC	Oligo 5' de mutation de Fixation d'ATP sur yRvb2 K81E
F603	ACCCAGAGACTGGGAAACACCCATGGCAAGAGCGGTCTCAC CAGTTGAAGG	Oligo 3' de mutation de fixation d'ATP sur yRvb2 K81E
F587	TTGTTTCATTGATAATGTTAATATGCTAGACATTGAAATA	Oligo 5' de mutation d'hydrolyse d'ATP yRvb1 E312N
F588	ATTAACATTATCAATGAACAAAACACCTGGAAT	Oligo 3' de mutation de l'ORF de yRvb1 E312N Mutation d'hydrolyse d'ATP
F589	TTATTTATCGACAATGTCCACATGTTGGATATAGAATGT	Oligo 5' de mutation de l'ORF de yRvb2 E297N, dans les vecteurs de co-expressions (pnEA-3cH, pnCS et pnYK)
F590	GTGGACATTGTCGATAAATAATACACCAGGAAC	Oligo 3' de mutation de l'ORF de yRvb1 E297N, dans les vecteurs de co-expressions (pnEA-3cH, pnCS et pnYK)
F591	CCTTCTACAGGTGAGACAGCATTGGCTCTCGCCATTTC	Oligo 5' de mutation de l'ORF de yRvb1 K85E, dans les vecteurs de co-expressions (pnEA-3cH, pnCS et pnYK)
F592	TGCTGTCTCACCTGTAGAAGGACCACCAGCTAA	Oligo 3' de mutation de l'ORF de yRvb1 K85E, dans les vecteurs de co-expressions (pnEA-3cH, pnCS et pnYK)
F306	ATATACCTCTATACTTTAACGTCAAGGAGAAAAACCCGG	Amorce pour le séquençage pFA6a Gal 3HA Bcd1 WT/D72A
F305	CATAAGAAATTCGCTTATTTAGAAGTGCGCGC	Amorce pour le séquençage pFA6a Gal KanMx
F446	GGGGAGCTCGATCGCTTGCCTGTAACCTTACACGC	Oligo 3' de clonage du CEN/ARS dans le vecteur pFA6a avec site de restriction <i>SacI</i>
F445	GGGACTAGTATCACGTGCTATAAAAATAATTATA	Oligo 5' de clonage du CEN/ARS dans le vecteur pFA6a avec site de restriction <i>SpeI</i>
E860	GGAGATCCTCGAGTACCGCCACAACACACC	Oligo 3' d'amplification de l'ORF Bcd1 avec site de restriction <i>XhoI</i>
E858	GAGGGCTAGCAGTACGTCACCTCTTT	Oligo 5' d'amplification de l'ORF Bcd1 avec site de restriction <i>NheI</i>
E592	TACATGCACCATTGCCTTCAACTGCGTCAGATAGTTGTA	Oligo 3' de mutation de l'ORF de Bcd1 R81A, dans les vecteurs de co- expressions (pnEA-3cH, pnCS et pnYK).

E591	ACGCAGTTGAAGGCAATGGTGCATGTACAAAAGATGGAT	Oligo 5' de mutation de l'ORF de Bcd1 R81A, dans les vecteurs de co-expressions (pnEA-3cH, pnCS et pnYK).
D852	AATGCTTATGTCCAGAGGGCCGCAACTATCTGACGCAG	Oligo 5' de mutation de l'ORF de Bcd1p D72A Y73A, dans les vecteurs de co-expressions (pnEA-3cH, pnCS et pnYK)
D851	CTGCGTCAGATAGTTGGCGGCCCTCTGGACATAAGCATT	Oligo 3' de mutation de l'ORF de Bcd1 D72A Y73A, dans les vecteurs de co-expressions (pnEA-3cH, pnCS et pnYK).
9688	TGAATGACATTAGCGTGAACAA	Oligo 5' de qPCR sur ARN snR52 levure
9689	TTCAGAAGGAAGGCAACATAAG	Oligo 3' de qPCR sur ARN snR52 levure
A159	TGCCTCGCTCGGTTGTTAAT	Oligo 5' qPCR sur snR36 de levure
A160	AGCGATATCCTCGTACTCAAAGA	Oligo 3' de qPCR sur snR36 de levure
A161	TGCAGGGAATGACTAAAAGCC	Oligo 5' qPCR sur snR32 de levure
A162	TCAGAAATGATCAACACCAGGC	Oligo 3' de qPCR sur snR32 de levure
B740	GAGCCACTGAATCCAAC	Oligo 5' de qPCR exon 2 U3 levure
B741	GTACATAGGATGGGTCA	Oligo 3' qPCR exon 2 U3 levure
B868	ATGGCAGTCTGACAAGTTA	Oligo 5' U3 levure immatures en 3'
C319	CCAGTAAATAATGCCAAAAATCC	Oligo 3' de qPCR RVB2 levure
C320	AAAGCTAGTGGGAAGATTACC	Oligo 5' de qPCR RVB2 levure
9702	TCACGGTGATGAAAGACTGG	Oligo 5' de qPCR sur transcrit snR128 (U14) de levure
9703	AAGAGCGGTCACCGAGAGTA	Oligo 3' de qPCR sur transcrit snR128 (U14) de levure

Tableau M1 : Liste des amorces désoxyribonucléotidiques utilisées pour les différentes expériences de ce travail de thèse (Oligo=Oligonucléotide).

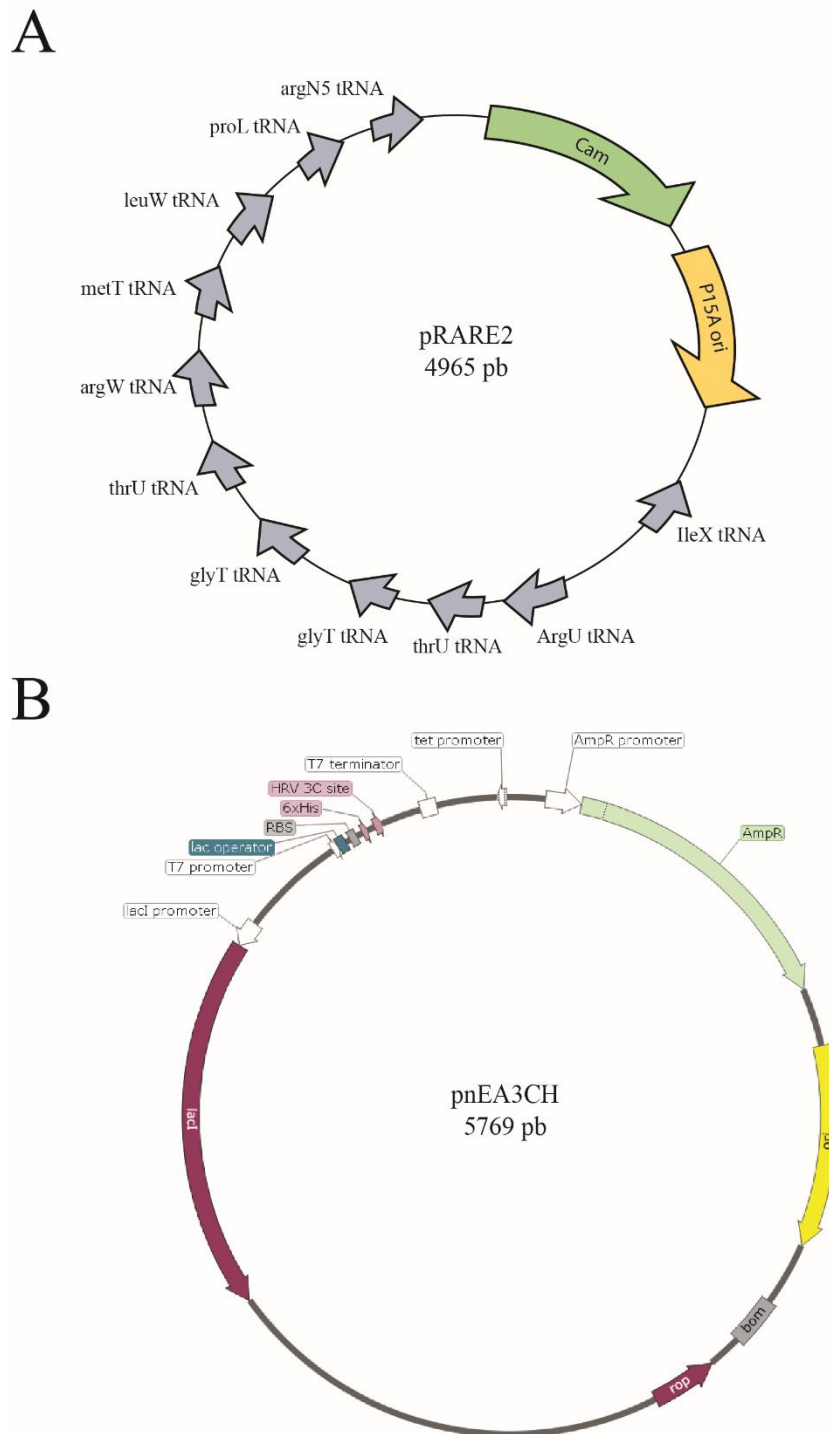
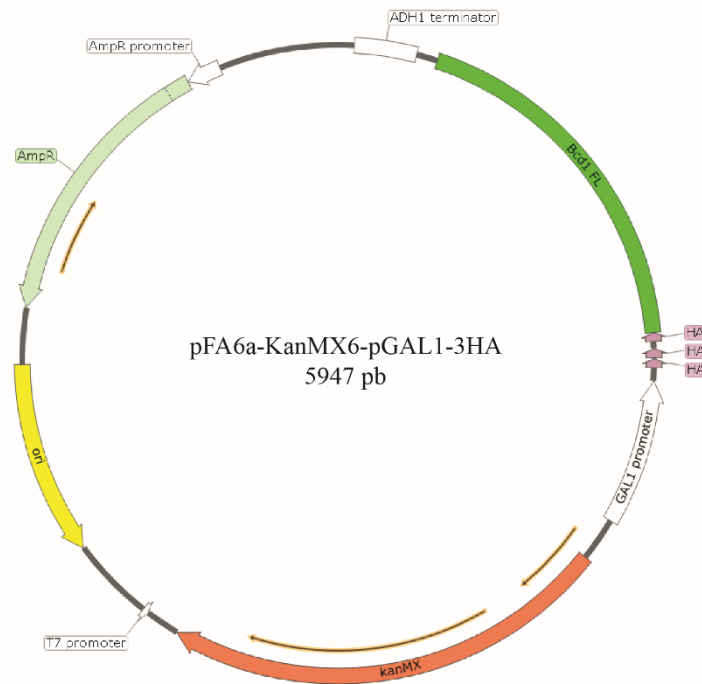


Figure M1 : Carte schématique des plasmides bactériens utilisés.

(A) vecteur pRARE2, cam : chloramphénicol, tRNA : ARN de transfert, ori : origine de réplication. (B) vecteur d'expression de protéines recombinantes chez *E. coli*, le pnEA3CH. RBS : *Ribosome Binding Site*, lacI : code pour le répresseur LacI de l'opéron lactose qui se fixe sur l'opérateur lactose en absence d'IPTG. Bom (*Basis Of Mobility*) : région de conjugaison fonctionnelle. Rop (*Repressor Of Primer*) : protéine qui maintient le plasmide à faible nombre de copies dans la cellule.

A



B

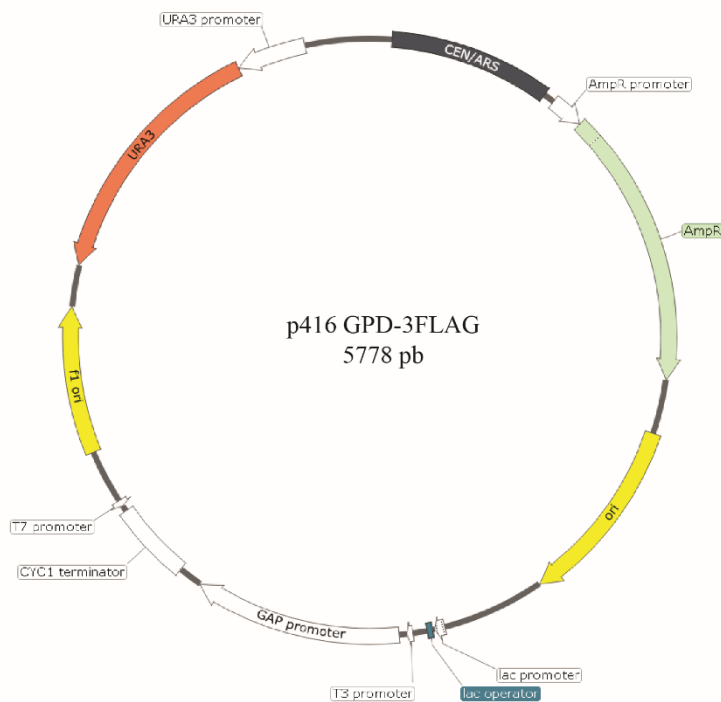


Figure M2 : Carte schématique des plasmides de levure utilisés.

(A) vecteur pFA6a GAL 3HA, ADH : *Alcool Déshydrogénase 1*, AmpR : Résistance à l'Ampicilline, ori : origine, KanMX : Kanamycine, GAL1 : Galactose 1. (B) vecteur p416 GPD-3FLAG, URA : URAcil, CEN/ARS : CENtromère (*Automatically Replicating Sequence*), lac : opéron lactose, GAPDH : Glycéraldéhyde 3-phosphate déshydrogénase.

1.3.Souches utilisées

a. Souches bactériennes : *Escherichia coli*

La souche **DH5 α** (*supE44*, Δ *lacU169* (Φ 80 *lacZ* Δ M15), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*) a été utilisée pour les étapes d'amplification et d'extraction d'ADN plasmidique relatives aux manipulations de génie génétique.

La souche **BL21 (DE3) pRARE2** (F- *ompT hsdSB* (rB-, mB-) *gal dcm* λ (DE3), [*argN5*, *argX*, *proL*, *leuW*, *metT*, *argW*, *thrT*, *glyT*, *thrU*, *argU*, *ileX*, *Camr*]) ainsi que la souche **C41 (DE3) pRARE2** (F- *ompT hsdSB* (rB- mB-) *gal dcm* (DE3), [*argN5*, *argX*, *proL*, *leuW*, *metT*, *argW*, *thrT*, *glyT*, *thrU*, *argU*, *ileX*, *Camr*]) ont servi de systèmes hôtes d'expression de protéines recombinantes nécessaires aux analyses biochimiques, biophysiques et structurales.

b. Souches de levures : *Saccharomyces cerevisiae*

La souche *S. cerevisiae* **BY4741** (Mating a ; *his3* Δ 1; *leu2* Δ 0; *met15* Δ 0; *ura3* Δ 0) ainsi que la souche *S. cerevisiae* **TAP-tag** (Mating a ; *ade2*; *arg4*; *leu2-3,112*; *trp1-289*; *ura3-52*) ont servi pour la construction de souches transformées avec le vecteur pFA6aKan-MX6 de façon à pouvoir contrôler l'expression inductible et transitoire de Bcd1p de façon ectopique dans les cellules de levure.

- **BY4741 pFA6a KanMX6 GAL1-3HA (-)**
- **BY4741 pFA6a KanMX6 GAL1-3HA-Bcd1 WT**
- **BY4741 pFA6a KanMX6 GAL1-3HA-Bcd1 D72A**
- **TAP-RVB2 pFA6a KanMX6 GAL1-3HA (-)**
- **TAP-RVB2 pFA6a KanMX6 GAL1-3HA-Bcd1 WT**
- **TAP-RVB2 pFA6a KanMX6 GAL1-3HA-Bcd1 D72A**
- **TAP-NOP58 pFA6a KanMX6 GAL1-3HA (-)**
- **TAP-NOP58 pFA6a KanMX6 GAL1-3HA-Bcd1 WT**
- **TAP-NOP58 pFA6a KanMX6 GAL1-3HA-Bcd1 D72A**
- **TAP-RSA1 pFA6a KanMX6 GAL1-3HA (-)**
- **TAP-RSA1 pFA6a KanMX6 GAL1-3HA-Bcd1 WT**
- **TAP-RSA1 pFA6a KanMX6 GAL1-3HA-Bcd1 D72A**
- **TAP-NOP56 pFA6a KanMX6 GAL1-3HA(-)**

- TAP-NOP56 pFA6a KanMX6 GAL1-3HA-Bcd1 WT
- TAP-NOP56 pFA6a KanMX6 GAL1-3HA-Bcd1 D72A
- TAP-RSA1 p416 GPD 3FLAG-SNU13 pFA6a KanMX6 GAL1-3HA (-)
- TAP-RSA1 p416 GPD 3FLAG-SNU13 pFA6a KanMX6 GAL1-3HA-Bcd1 WT
- TAP-RSA1 p416 GPD 3FLAG-SNU13 pFA6a KanMX6 GAL1-3HA-Bcd1 D72A

1.4. Les milieux de cultures

Différents milieux préparés au laboratoire ont servi pour la mise en culture des bactéries et des levures. Tous ont été stérilisés par autoclavage à 120°C pendant 20 minutes à 1 bar de pression :

- **Le milieu LB** (*Luria Bertani*) est le milieu de culture des bactéries C41, BL21(DE3) et DH5 α . Il est composé de bactotryptone à 10 g/L, d'extrait de levure à 5 g/L et de NaCl à 5 g/L avec un pH 7,5.

- **Le milieu riche YPD** (*Yeast extract Peptone Dextrose*) est le milieu de culture des levures avec le glucose (dextrose) comme source de carbone. Il est composé d'extrait de levure à 10 g/L, bactopectone à 20 g/L et glucose à 20 g/L

- **Le milieu riche YPG** (*Yeast extract Peptone Galactose*) est le milieu de culture des levures avec le galactose comme source de carbone. Il est composé d'extrait de levure à 10 g/L, bactopectone à 20 g/L et galactose à 20 g/L.

- **Le milieu sélectif YNB** (*Yeast Nitrogen Base*) est le milieu de culture des levures constitué d'une base azotée de levure sans acides aminés à 10 g/L, du glucose / galactose à 20 g/L et d'un mélange d'acides aminés à 20 g/L pour chaque acide aminé, permettant d'appliquer une sélection selon le marqueur d'auxotrophie du vecteur à sélectionner dans la levure.

1.5. Les antibiotiques

- **Ampicilline** qui a été utilisée à 100 μ g/L.
- **Chloramphénicol** qui a été utilisé à 25 μ g/L.
- **Généticine (G418)** qui a été utilisé à 200 μ g/mL.

1.6. Les tampons utilisés

a. *Escherichia coli*

- **Le tampon de lyse pour purifier une protéine recombinante** : 25 mM HEPES KOH pH 7,5 ; 300 mM NaCl ; 10 mM Imidazole ; 0,5 mM TCEP ; 5% Glycérol ; (0,5 mM ADP/ATP et 5 mM MgCl₂ sont ajoutés pour certaines purifications de protéines recombinantes).
- **Le tampon d'élution des protéines de la résine Cobalt ou Nickel** : 25 mM HEPES KOH pH 7,5 ; 300 mM NaCl ; 300 mM Imidazole ; 0,5 mM TCEP ; 5 % Glycérol (0,5 mM ADP/ATP et 5 mM MgCl₂ sont ajoutés pour certaines purifications de protéines recombinantes).
- **Le tampon de dialyse** : 25 mM HEPES KOH pH 7,5 ; 300 mM NaCl ; 1 mM DTT ; 5 % Glycérol.
- **Le tampon de chromatographie d'exclusion de taille** : 10 mM HEPES KOH pH 7,5 ; 150 mM NaCl ; 1 mM DTT ; 1 % Glycérol (0,5 mM ADP/ATP et 5 mM MgCl₂ sont ajoutés pour certaines purifications de protéines recombinantes).

b. *S. cerevisiae*

- **Tampon de lyse des levures pour la préparation d'extrait cellulaire total** : 20 mM HEPES KOH pH 7,9 ; 3 mM MgCl₂ ; 10% Glycérol ; 150 mM NaCl ; 500 nM DTT ; 0,1% Triton X100.
- **Tampon d'extraction d'acides nucléiques totaux** : 0,1 M LiCl ; 0,1 M Tris-HCl, pH 7,5 ; 0,1 mM EDTA.

1.7. Les solutions

- **TE** : 10 mM Tris-HCl, pH 7,4 ; 1 mM EDTA.
- **PBS 1X** : 10 mM Na₂HPO₄ ; 1,8 mM KH₂PO₄, pH 7,3 ; 140 mM NaCl ; 2,7 mM KCl
- **PBS Tween** : PBS 1X ; Tween20 0,1%.
- **TAE 1X** : 40 mM Tris, 20 mM Acide acétique et 1 mM EDTA.
- **Bleu de dépôt pour gel d'agarose 6X** : Glycérol (30%), Bleu de bromophénol (0,25% p/v) et Xylène cyanol (0,25 % p/v).
- **Tampon de migration SDS-PAGE** : Glycine 14,25 g/L ; Tris 3 g/l ; SDS 1 g/L.

- **Bleu de dépôt pour SDS-PAGE** : 80 mM Tris-HCl pH 6,8 ; SDS 3% ; 100 mM DTT ; 10% Glycérol ; 0,1% Bleu de bromophénol.
- **Tampon de transfert de SDS-PAGE (membrane nitrocellulose)** : 25mM Tris-HCl ; 0,96 M Glycine ; 20% Ethanol.
- **Solution de coloration des membranes de nitrocellulose au rouge ponceau** : 5% Acide acétique ; 0,1% Rouge Ponceau.
- **Solution de coloration pour SDS-PAGE** : 5% Ethanol ; 7,5% Acide acétique et Bleu de coloration (environ 200 µL).
- **Solution de déshydratation pour SDS-PAGE** : 50% Ethanol et 10% Acide acétique.

1.8. Les gels d'analyses

- **SDS-PAGE (*stacking*)** : Acrylamide 5% (500 µL) (à partir d'un stock d'Acrylamide/Bisacrylamide 29:1 à 40%), TRIS-HCl à pH 6,8 (384 µL) (50 mM), H₂O (3mL), SDS (0,1%) (40 µL), APS (10 µL) et TEMED (10 µL). 2 mL par gel.
- **SDS-PAGE 12,5 % (*Running*) (2 gels)** : Acrylamide 12,5% (2,5 mL) (à partir d'un stock d'Acrylamide/Bisacrylamide 29:1 à 40%), Tris-HCl à pH 8,8 (2 mL), H₂O (3,3 mL), SDS (0,1%) (80 µL), APS (20 µL) et TEMED (20 µL). 4 mL par gel.
- **Gel d'agarose 1,5%** : Agarose 1,5 % (1,5 g pour 100 mL de tampon), TAE 1X, Gel Red à 1/10000.

1.9. Les anticorps

- **Anticorps α -AspRS (Dps1)** : Anticorps polyclonal produit chez le lapin. Utilisé à une dilution au 1/5 000^e.
- **Anticorps α -HA-Prx** : Anticorps monoclonal produit chez le rat (Roche). Utilisé à une dilution au 1/250^e.
- **Anticorps α -PAP (TAP)-Prx** : Anticorps polyclonal produit chez le lapin (Sigma). Utilisé à une dilution au 1/2000^e.
- **Anticorps α -IgG-Prx lapin** : Anticorps polyclonal produit chez la chèvre, couplé à la peroxydase (Thermo). Utilisé à une dilution au 1/20 000^e.

2. Méthodes

2.1. Stratégies de génie génétique

a. La PCR (*Polymerase Chain Reaction*)

Des réactions de PCR ont été réalisées en utilisant des oligonucléotides spécifiques de l'ORF (*Open Reading Frame*) et qui permettent, si besoin, d'amener les sites de restriction compatibles avec le vecteur de destination en amont (5') et en aval (3') de l'ORF. La *Dream Taq Pol* (Thermo Fisher Scientific) a été utilisée (1 U/ μ L pour un volume final) dans son tampon *Taq Buffer* 10x et dans un volume final de 25 μ L. La *Phusion Hot Start II* (Thermo Fisher Scientific) utilisée à 8 mU/ μ L avec le tampon *HF 5X*. Les oligonucléotides utilisés sont à des concentrations de 100 μ M chacun. Le mélange des 4 nucléotides dNTP est ajouté à raison de 0,2 μ M pour chaque nucléotide, la matrice ADN contenant la séquence à amplifier est ajoutée à 1 ng/ μ L final. Les thermocycleurs utilisés sont le *Thermocycler 2720* (Applied Biosystems) et *SureCycler 8800* (Agilent Technologies). Les produits de la PCR sont par la suite visualisés sur gel d'agarose préparés à un pourcentage en fonction de la taille de la bande attendue. La totalité du produit PCR est mélangée au bleu de dépôt, à une concentration de 1X final et déposée sur gel. La migration est réalisée en tampon TAE 1X ou TBE 1X pendant 40 min à 100 V. Le produit PCR est par la suite purifié sur gel agarose et extrait en utilisant le kit *NucleoSpin® Gel and PCR Clean-up* (Machery-Nagel), suivi d'un dosage à 260 nm sur le *NanoDrop 2000c* (Thermo Fischer Scientific).

b. Le clonage

Il permet l'insertion du gène d'une protéine d'intérêt dans un vecteur plasmidique d'expression qui sera par la suite amplifié dans un système bactérien. Après amplification par PCR de la séquence codant la protéine d'intérêt avec les sites des enzymes de restriction choisies, le produit obtenu ainsi que le vecteur de destination sont ensuite digérés par les enzymes de restriction spécifiques. Il s'agit d'enzymes dites *FastDigest* (Thermo Fisher Scientific) utilisées à 0,5X dans un tampon *FastDigest* 10X. Un traitement du mélange réactionnel par l'enzyme *DpnI* (0,25X) à 37°C pendant 20 minutes, permet de dégrader uniquement l'ADN méthylé, contrairement aux produits PCR qui ne le sont pas. Une purification du produit PCR et du vecteur digéré est ensuite réalisée en utilisant le kit *NucleoSpin® Gel and PCR Clean-up* (Machery-Nagel). L'étape de ligature permet l'intégration de l'insert dans le vecteur de destination préalablement digéré par les mêmes enzymes de restrictions. L'insert est donc mis

en présence du vecteur linéarisé à un ratio en masse (1/3) (vecteur/insert), l'enzyme T4 DNA Ligase est ajoutée à raison de 0,5 U/ μ L final en présence de son tampon 10X. La ligature est réalisée à température ambiante pendant 2 heures, puis des bactéries DH5 α rendues compétentes sont transformées par le mélange réactionnel obtenu, puis étalées sur boîte de pétri et incubées à 37°C pendant 12 heures. La présence du vecteur plasmidique dans les colonies obtenues est vérifiée par PCR sur colonies.

c. Le sous-clonage

Il permet de cloner un gène voulu, d'un vecteur à un autre vecteur plasmidique, possédant les mêmes sites de restrictions, sans passer par une amplification du gène d'intérêt par PCR. Une simple digestion de l'insert par des enzymes de restrictions est réalisée, qui est ensuite purifiée par le kit *NucleoSpin® Gel and PCR Clean-up* (Machery-Nagel) avant d'être insérée dans le vecteur de destination qui a subi une linéarisation par les mêmes enzymes de restrictions.

d. La transformation bactérienne

Elle permet l'amplification du plasmide d'intérêt issu d'un clonage ou d'un sous-clonage. Pour cela, le mélange réactionnel est mis au contact avec des bactéries DH5 α chimio-compétentes (traitées aux CaCl₂) dans de la glace pendant 30 minutes. Cette étape va permettre aux acides nucléiques d'adhérer aux parois des cellules bactériennes. Un choc thermique à 42°C est réalisé pendant 50 secondes, ce qui va assurer la fluidification de la paroi bactérienne et donc la pénétration du vecteur dans les cellules. Puis le mélange est remis sur glace dans le but d'arrêter cette fluidification. La paroi des cellules étant perturbée, l'ajout de milieu de culture LB (140 μ L) dans le mélange et une incubation à 37°C pendant 1 heure sont nécessaires pour assurer sa régénération. Les bactéries sont ensuite étalées sur boîtes de pétri contenant du milieu de culture LB avec l'antibiotique permettant la sélection du plasmide inséré dans les cellules. Les boîtes de pétri sont incubées 12 heures à 37°C.

e. La mutagenèse dirigée

Elle consiste à apporter une mutation ponctuelle sur un gène donné, à travers une réaction PCR utilisant deux amorces complémentaires apportant la mutation désirée. La réaction est réalisée

en utilisant une enzyme ADN polymérase *Phusion*[®] (*Finnzyme*) 0,25 U/ μ L, 1,25 μ M de chaque dNTP, des deux amorces à 20 μ M chacune et 100 ng du vecteur portant la matrice ADN à muter. Après la réaction PCR, le mélange est traité par *DpnI* puis transformé dans des bactéries compétentes DH5 α qui sont misent en culture 12 heures à 37°C. Un séquençage du vecteur obtenu permet de s'assurer de la présence de la mutation insérée.

2.2. Biochimie des protéines recombinantes

a. Synthèse des protéines recombinantes

L'ensemble des protéines recombinantes utilisées ont été produites dans un système bactérien qui assure une production satisfaisante en terme de quantité de protéine en un temps de culture très court. Un ou plusieurs vecteurs portant le (les) gène(s) codant la (les) protéine(s) d'intérêt(s) est (sont) transformé(s) dans une souche de bactérie *E. coli* de type BL21 DE3 pRARE2, une autre souche a également été utilisée pour la production du complexe recombinant Rvb1-Rvb2, il s'agit de la souche C41 DE3 pRARE2. Une première culture de 20 mL est lancée pendant 12 heures à 37°C, à partir d'une seule colonie, dans un milieu de culture riche LB complétement de l'antibiotique approprié permettant la sélection du vecteur transformé dans les cellules, et de 25 mg/L de chloramphénicol qui sert à maintenir le vecteur pRARE2 dans les cellules. La pré-culture étant à saturation au bout de 12 heures, elle sert à ensemencer une plus grande culture (1L) à raison de 0,1 DO_{600nm}/mL. Cette culture est incubée à 37°C sous agitation 180 RPM, jusqu'à ce que sa densité optique atteint 0,6 – 0,8. L'IPTG est ajouté à une concentration de 0,3 mM final, ce dernier permet la surexpression de la protéine recombinante dans la cellule bactérienne. La culture est alors incubée à 20°C sous agitation 180 RPM pendant 12 heures. Les cellules bactériennes sont ensuite récupérées par centrifugation du milieu de culture à 8 000 xg pendant 20 minutes. Elles sont ensuite stockées à -20°C.

b. Purification des protéines recombinantes

Cette étape va permettre d'isoler la protéine recombinante surexprimée d'intérêt, de tous les autres composants bactériens. Le culot bactérien est repris dans un tampon de lyse des bactéries (30 mL de tampon de lyse pour un culot bactérien d'une culture de 1 L à un DO de 3 (100 DO/mL). Cette suspension bactérienne va ensuite subir une lyse par ultrasons 4 fois 1 minutes (avec un intervalle de 1 minutes pour éviter la surchauffe) à une amplitude de 50% et sur une puissance de 6, en utilisant un appareil *Sonifier 250* de chez *Branson*. Cela permet de casser les membranes cellulaires ainsi que l'ADN et ainsi libérer la protéine d'intérêt. L'élimination des

débites cellulaires se fait par une première étape de centrifugation 20 000 RPM pendant 40 minutes à 4°C (*Beckman Coulter Avanti J-20XP centrifuge™*). Le surnageant obtenu est composé d'acides nucléiques et de protéines. Il subit un traitement au polyÉthylène Imine (PEI) 0,0125% (v/v), qui est un polymère chimique chargé positivement. Il va permettre la précipitation des acides nucléiques chargés négativement qui sont ensuite éliminés du surnageant par centrifugation 20 000 RPM pendant 20 minutes à 4°C suivi d'une filtration à 0,25 µm. La purification de la protéine d'intérêt a été réalisée grâce à l'étiquette 6 x Histidines présente sur la protéine, par la technique IMAC (*Immobilized Metal Affinity Chromatography*). Elle consiste à utiliser une résine cobalt (TALON, *Clontech*) qui est composée d'une matrice sépharose sur laquelle est greffée le cobalt, qui va fixer fortement les cycles histidines de l'étiquette. Le surnageant obtenu après filtration est mis en présence de 800 µl de résine à 50% pendant 1 heure sur roue (10 RPM) à 4°C. Cette étape de fixation est répétée 3 fois avec de la résine neuve pour chaque purification de façon à épuiser la protéine d'intérêt dans le surnageant. Un tampon d'élution composé d'une forte concentration d'imidazole (300 µM) est appliqué sur la résine (2 mL pour 400 µL de résine). Ce tampon va permettre d'éluer la protéine d'intérêt par compétition avec les cycles d'histidines sur le cobalt. La coupure de l'étiquette 6 x Histidines est permise par l'action de l'enzyme PreScission 3C toute la nuit à 4°C sous agitation. Une dialyse est nécessaire afin d'éliminer l'imidazole en présence de la protéine d'intérêt. Pour assurer un clivage efficace par la 3C, l'échantillon est donc placé dans un boudin de dialyse qui assure les échanges, dans un volume de 3 à 4 L. Pour finaliser le processus de purification, l'échantillon est concentré puis centrifugé à 16 000 x g pendant 5 minutes à 4°C, puis il est injecté sur une colonne de chromatographie d'exclusion de taille Superdex 200/75 ou Superose 6 *increase*, en fonction de la taille de la/les protéines ou des complexes à purifier. Ces colonnes sont pilotées par un système AKTA Purifier. Les composants de l'échantillon seront donc séparés selon leurs tailles, plus un composé est grand plus il va sortir en premier de la colonne (il est exclu des pores de la matrice composant la colonne). Un débit d'élution (0,5 mL/min) est appliqué. Les protéines sont détectées grâce à un détecteur d'UV à 280 nm qui nous indique une intensité de chaque pic sur le chromatogramme. Les fractions sont collectées dans des tubes (2 mL pour chaque fraction) en sortie de colonne. Les fractions correspondantes aux pics les plus majoritaires dans le chromatogramme sont analysées par SDS-PAGE 10% ou 12,5% selon la taille des protéines à analyser.

c. La technique de co-élution

Elle consiste à analyser la formation d'un complexe protéique *in vitro*, issu de deux (ou plus) purifications de protéines séparées dans un tampon identique. Un mélange de deux (ou plus) protéines est réalisé *in vitro* à un ratio avec un des deux partenaires en excès. Ce mélange est laissé dans un tube pendant 2 heures minimum sur glace. Puis, il est concentré grâce à des concentrateurs de type Amicon® choisi selon la taille des protéines de l'échantillon. Une centrifugation de l'échantillon à 16 000 x g pendant 5 minutes à 4°C permet d'éliminer les potentiels agrégats protéiques formés après concentration. L'échantillon est ensuite injecté sur une colonne Superose 6 *increase*. Les fractions récupérées en sortie de colonne sont analysées sur SDS-PAGE en fonction du profil chromatographique obtenu. Dans le cas d'un complexe composé que de deux protéines, si le profil du chromatogramme indique uniquement deux pics aux mêmes volumes d'élution que chacune des deux protéines seules sur cette colonne, cela signifie qu'il n'y a pas eu formation de complexe entre les deux partenaires. Si au contraire, nous retrouvons un seul pic majoritaire dont le volume d'élution est supérieur aux deux autres volumes d'élution des protéines seules, c'est qu'il y a potentiellement formation d'un complexe. Les fractions sous ce pic sont ensuite analysées sur SDS-PAGE où nous devons retrouver deux bandes correspondantes aux deux protéines dans le cas de la formation d'un complexe.

2.3. Etude biophysique des protéines et complexes protéiques

a. Isothermal Titration Calorimetry (ITC)

La technique d'ITC permet de mesurer une constante de dissociation (K_d) d'un complexe en caractérisant ses paramètres thermodynamiques. Elle repose sur la mesure de l'échange de chaleur entre l'échantillon et le milieu, qui est due à la formation d'un complexe entre deux partenaires, accompagnée d'une absorption ou émission de chaleur. Deux cellules adiabatiques composent l'appareil. Elles sont maintenues à une chaleur identique et constante. La cellule de référence ne contient que de l'eau. Sa température est régulée de façon à la maintenir constante. Par rapport à la deuxième cellule qui va contenir l'échantillon où aura lieu la formation du complexe. Les variations de température issues de la formation du complexe sont donc précisément mesurées par l'appareil de façon indirecte. Un des deux partenaires dit titrant est mis dans une seringue qui va injecter à intervalles constants le titrant sur le deuxième partenaire qui se trouve dans la cellule de réaction. A chaque injection (40 injections de 6 μ L espacées de 200 secondes chacune), l'appareil fournit une énergie qui permet le retour à la température définie (25°C) en début d'expérience. C'est cette énergie fournie qui est mesurée par l'appareil.

L'appareil utilisé est le VP-ITC de chez *MicroCalTM* (*GE Healthcare*). L'échantillon placé dans la cellule est à une concentration de 40 μM et l'échantillon qui se trouve dans la seringue est à une concentration dix fois plus élevée, 400 μM . Le tampon des deux partenaires est identique. Il est composé de 10 mM HEPES pH 7,5 ; 150 mM NaCl ; 0,5 mM TCEP ; 0,5 mM ADP ; 5 mM MgCl_2 .

b. Dynamique Light Scattering (DLS)

Un échantillon est placé sous un faisceau de lumière qui est dévié par la nature même de l'échantillon. La diffusion de la lumière dépend de la taille de la particule. En DLS, seuls les rayons déviés à 90° sont observés. Une cuve en quartz est utilisée, elle présente des ouvertures sur trois de ses faces (QS 3 mm de chez Hellma avec un volume utile minimum de 12 μL). Cette technique nous renseigne sur l'homogénéité de l'échantillon ainsi que sur la taille des particules présentes. L'appareil utilisé est le *Zetasizer Nanoseries (Nano-S)* de *Malvern Instruments*. Une équilibration de 2 minutes de l'échantillon est appliquée, suivie de trois séries de mesures, avec 12 mesures chacune, qui seront moyennées.

c. Bio-Layer Interferometry (BLI)

C'est une technologie sans marquage permettant de mesurer les interactions biomoléculaires. Il s'agit d'une technique d'analyse optique qui analyse le motif d'interférence de la lumière blanche réfléchi par deux surfaces : une couche de protéine immobilisée sur la pointe du biocapteur et une couche de référence interne. Tout changement dans le nombre de molécules liées à la pointe du biocapteur provoque un changement dans le modèle d'interférence qui peut être mesuré en temps réel. La liaison entre un ligand immobilisé sur la surface de la pointe du biocapteur et un analyte en solution produit une augmentation de l'épaisseur optique à la pointe du biocapteur, ce qui entraîne un décalage de longueur d'onde, qui est une mesure directe du changement d'épaisseur de la couche biologique. Les interactions sont mesurées en temps réel, offrant la possibilité de surveiller la spécificité de liaison, les taux d'association et de dissociation, ou la concentration, avec précision et exactitude. Seules les molécules qui interagissent ou sont dissociées du biocapteur peuvent faire varier le profil d'interférence et générer un profil sur le système Octet[®]. Les molécules non liées n'affectent pas le motif d'interférence. Cela permet de connaître les paramètres d'affinité des partenaires et leur cinétique. Cette technique a été expérimentée au laboratoire d'Architecture et Fonction des Molécules Biologiques AFMB à Marseille. Les échantillons ont tous été préparés au laboratoire IMoPA.

2.4. Etude structurale par Cryo-microscopie électronique

L'échantillon du complexe protéique a été produit et purifié au laboratoire, dans un système bactérien *E. coli*, puis par co-élution des deux partenaires sur colonne Superose 6 *increase* sur un système AKTA *purifier*. Le tampon final de l'échantillon est 10 mM HEPES KOH pH 7,5 ; 150 mM NaCl ; 1 mM TCEP ; 1 % Glycérol 0,5 mM ADP et 5 mM MgCl₂. L'acquisition des données et leur analyse par cryo-microscopie a été faite au laboratoire de *Thermo Fisher Scientific* au Pays-Bas. Le complexe a été utilisé à une concentration de 1,14 mg/mL pour générer et cribler 12 grilles *via* le Glacios Falcon4 (Température 4°C, humidité 95%, temps de d'absorption 5 à 7 secondes, temps d'attente d'absorption 5 secondes). A l'issue de ce criblage, 10858 images ont été sélectionnées à l'aide du KriosG4 Falcon4 (360 images/heure, temps d'exposition 2,27 secondes, grossissement 215 000 fois, taille des pixel 0,58 Å).

2.5. Transformation des levures compétentes

Une culture est préparée à partir d'une pré-culture en phase stationnaire de croissance, en milieu YPG, YPD ou en milieu sélectif YNB, à une DO_{600nm} = 0,3. Nous incubons la culture à 30°C avec agitation de 180 RPM pendant 6 heures, jusqu'à ce que la DO_{600nm} soit égale à 0,8-1. Les cellules sont ensuite récupérées par centrifugation à 2 500 RPM pendant 10 minutes, puis lavées dans 10 mL d'eau. Le culot de cellules est récupéré une seconde fois par centrifugation et repris dans 1 mL d'AcLiTE (Acétate de Lithium 0.1M ; Tris/HCl 10 mM, pH 7,4 ; EDTA 1 mM). 50 µL de cellules sont incubées 30 minutes à 30°C, avec environ 1 µg de plasmide, en présence de 45 µg d'ADN de sperme de saumon préalablement dénaturé 15 minutes à 96°C et de 300 µL de PEG 4 000. Un choc thermique de 10 minutes à 42°C est appliqué pour garantir la pénétration du plasmide à travers les parois. Les cellules sont ensuite étalées sur boîte de pétri contenant le milieu sélectif correspondant au marqueur d'auxotrophie porté par le plasmide.

2.6. Système d'expression avec un promoteur inductible

Le système choisi pour l'expression conditionnelle est un système galactose et plus précisément un plasmide dont le gène *BCDI D72A* est placé sous le promoteur *GALI* à expression conditionnelle. Ce plasmide permet l'expression ou non du mutant dans la cellule selon la source de carbone fournie dans le milieu de culture. L'enzyme Galactokinase est produite suite à l'expression du gène *GALI* dans la cellule. Le promoteur de ce gène est soumis à une

régulation négative par le répresseur Gal4p qui empêche la fixation de l'ARN polymérase II. L'ajout de galactose avec le milieu de culture YPG, permet de lever la répression imposée par le Gal4p sur le promoteur *GALI* et donc libère l'expression du gène soumis au promoteur *GALI*. Dans le cadre de notre construction, il s'agit du gène qui exprime la protéine Bcd1 D72A. Avec la présence de glucose dans le milieu de culture YPD, le promoteur *GALI* est bloqué par Gal4p et le gène *BCD1* D72A n'est pas transcrit. Les souches transformées par le plasmide pFA6a GAL 3HA BCD1 D72A ont dans un premier temps été cultivées dans un milieu YPD pour une pré-culture et ensuite une culture minimum permettant la croissance des cellules. Puis le milieu de culture est remplacé par du YPG qui permet l'expression du mutant Bcd1p D72A pendant deux heures seulement jusqu'à atteindre une $DO_{600nm} = 0,8-1$. Les cellules sont ensuite centrifugées à 2 500 RPM pendant 10 minutes et stockées à $-80^{\circ}C$ après un passage rapide dans de l'azote liquide.

2.7. Préparation de l'extrait cellulaire total de levure

Des cellules de levures conservées à $-80^{\circ}C$ sont décongelées (40 à 60 unités de DO_{600nm}/mL). Les parois cellulaires sont rompues dans un premier temps par l'ajout de 300 μL du tampon de lyse complété par de l'antiprotéase 1X (*Thermoscientific*) ainsi que 300 μL de billes en verres (Sigma). 4 à 5 séries d'agitation d'une minute à 20 Hz sont effectuées sur un agitateur à bras mécanique (Retsch MM400), espacées par 1 minute sur glace. Une entaille à la base du tube est effectuée, pour permettre la récupération du lysat par centrifugation d'une minute à 3 000 x g et à $4^{\circ}C$. Un dosage du surnageant par le kit Bradford (*The Quick Start™ Bradford protein assay, Biorad*) permet de quantifier les protéines présentes. Cet extrait protéique est gardé sur la glace ou stocké à $-80^{\circ}C$ pendant une semaine maximum, après un bref passage dans de l'azote liquide.

2.8. Extraction d'ARN totaux de levure

Des cellules de levure conservées à $-80^{\circ}C$ sont décongelées (20 unités de DO_{600nm}/mL). Elles sont reprises dans 200 μL de tampon d'extraction d'acides nucléiques et 300 μL de billes en verre (Sigma). Les cellules sont broyées par agitation avec 4 à 5 séries d'agitation d'une minute à 20 Hz (Retsch MM400), espacées d'un passage d'une minute sur glace. 400 μL du mélange phénol/chloroforme/alcool isoamylique (25: 24: 1), ainsi que 25 μL de SDS10%. L'échantillon est ensuite passé au vortex pendant une minute puis centrifugé 5 minutes à 16 000 x g. Seul le surnageant est soigneusement récupéré dans un nouveau tube, auquel est ajouté 400 μL du

mélange chloroforme/alcool isoamyélique (24: 1), et le tout est centrifugé une nouvelle fois, 5 minutes à 16 000 x g. Dans un premier temps, les acides nucléiques du surnageant sont précipités par l'ajout de 100% puis 70% d'éthanol et centrifugés à chaque étape, avant d'être repris dans 30 µL d'eau ultrapure. L'ADN est ensuite éliminé par ajout de DNase I RQ1 (*Promega*) (3 UE) pendant 30 minutes à 30°C. Une précipitation au phénol/chloroforme suivie d'une précipitation à l'alcool sont à nouveau réalisées, afin de récupérer les ARN totaux dans de l'eau ultrapure qui sont dosés par la méthode spectrophotométrie à DO_{260nm} avec le nanodrop 2000c (*Thermoscientific*).

2.9. Méthodes de détection des protéines et des ARNs

a. Stratégie de *Western Blot*

Une fois l'extrait protéique obtenu, il est ensuite dosé par deux techniques similaires : la méthode du Qubit (*Qubit Protein Assay, Thermoscientific*) basée sur le principe de détection de faible quantité de protéine par fluorescence, ou par le dosage de Bradford. Il sera ensuite dénaturé dans un tampon (Tris-HCl 80 mM, pH 6,8 ; SDS 3% ; DTT 100 mM ; Glycérol 10%, Bleu de bromophénol 0,1%) pendant 5 minutes, puis il est analysé par SDS-PAGE (*Biorad*, ou *Hofer SE400*). Une fois la migration des protéines sur le gel terminée, elles sont transférées sur une membrane de Nitrocellulose (*Biorad*) grâce à un courant électrique à 100 V généré par l'appareil Mini Trans-Blot (*Biorad*) dans un tampon de transfert (Tris 7,57 g/L ; Glycine 1,35 g/L ; Ethanol 20%), pendant 1 heure. Un tampon de blocage (Lait 5% ; PBS 1X ; Tween-20 0,1%) est ensuite appliqué à la membrane sous une agitation pendant 45 minutes. Ce tampon est essentiellement composé de protéines de lait qui vont se fixer à la membrane et ainsi occuper les sites non spécifiques et les saturer. La révélation des protéines étudiées sur la membrane commence par l'hybridation d'un anticorps primaire pendant 1 heure sous agitation. Cet anticorps est dilué dans un tampon d'hybridation (PBS 1X ; Tween-20 0,1% ; BSA 3%). La membrane est lavée 3 fois pendant 15 minutes avec un tampon de rinçage (PBS 1X ; Tween-20 0,1%). En fonction de la protéine à révéler, il arrive qu'un anticorps secondaire soit nécessaire. Il est alors incubé avec la membrane de la même façon que l'anticorps primaire avant de rincer à nouveau cette dernière 3 fois pendant 15 minutes par le tampon de rinçage. Une révélation par chimioluminescence par le kit ECL+ (*GE Healthcare*) permet d'identifier les protéines d'intérêt sur la membrane par l'utilisation du système d'imagerie *Fusion Solo* (*Vilmer Lourbat*).

b. RT-PCR

1 µg d'ARN totaux est incubé avec 1 µl d'amorce poly-dT (10 µM) dans un volume de 10 µl d'eau ultrapure à 70°C pendant 5 minutes. L'échantillon est refroidi sur glace avant d'ajouter 200 unités de la rétrotranscriptase M-MLV (*Moloney Murine Leukemia Virus*) (*Promega*) avec son tampon suivi d'une incubation pendant 1 heure à 42°C. 10 µM d'amorces standards sont ajoutées à 10 µM avant un cycle d'incubation de 30 minutes à 16°C, 30 minutes à 42°C et 5 minutes à 85°C et ensuite l'ensemble est refroidi à 4°C pendant 5 minutes.

c. PCR quantitative

Une courbe standard relative est utilisée pour l'ensemble des qPCRs réalisées. Elle est composée de 6 points de gamme testés en duplicat regroupant l'ensemble des ADNc (toutes les conditions expérimentales étudiées) dilué avec un facteur de 1/10. L'amplification est réalisée grâce à la *Supermix iTaq Universal SYBER Green* (Biorad), à laquelle est ajoutée les amorces PCR à une concentration finale de 0,5 µM dans un volume final de 10 µL. L'hybridation des amorces est faite à 59°C avec 40 cycles dans un thermocycleur *Applied Biosystem StepOne*. L'amplification de chaque point expérimental d'ADNc est réalisée en triplicat.

2.10. Méthodes d'analyse des complexes protéiques et ARN-protéines

a. Méthode de co-immunoprécipitation

Les protéines contenues dans l'extrait cellulaire total de levure sont immunoprécipitées par des billes décorées d'anticorps spécifiques. Selon la cible à immunoprécipiter, différents types de billes ont été utilisées : Des billes IgG-Sepharose (GE Healthcare) pour l'étiquette TAP, des billes anti-FLAG-M2 Agarose (Sigma) pour l'étiquette FLAG, des billes magnétiques protéine G (Invitrogen) sur lesquelles sont fixés des anticorps α-HA (Roche) pour l'étiquette HA. Pour chacune des immunoprécipitations 40 µL de billes sont utilisées. Elles sont lavées 3 fois pendant 15 minutes dans le tampon de lyse cellulaire. L'extrait cellulaire est ensuite rajouté sur les billes (150 à 1000 µg de protéines de l'extrait cellulaire) sur roue à 4°C pendant 2 heures. Les billes sont ensuite lavées 3 fois avec le tampon de lyse et sont directement additionnées de 50 µL de bleu de protéine 2X. Un volume d'échantillon de l'extrait cellulaire total est préparé en prélevant 10% du volume ajoutée sur les billes, il représente l'*Input* qui permet de visualiser l'ensemble des protéines contenues dans l'extrait cellulaire total. Les échantillons sont ensuite

dénaturés à 96°C pendant 5 minutes avant d'être déposés sur gel SDS-PAGE 12,5%, puis analysés par *Western-Blot*.

b. Méthode d'immunoprécipitation d'ARN (RIP)

Afin d'analyser les ARN associés aux protéines immunoprécipitées, les billes récupérées après mise en contact avec l'extrait cellulaire total sur roue pendant 2 heures à 4°C (voir partie 2.9.a de ce chapitre), sont lavées 3 fois pendant 15 minutes, et incubées avec le tampon de lyse contenant 150 µg de protéinase K (Roche) pendant 30 minutes. Cela permet de dégrader les protéines immunoprécipitées par les billes et de libérer les acides nucléiques associés. Un échantillon Input est préparé à partir de 10% d'extrait cellulaire total. L'extraction des acides nucléiques est réalisée dans un premier temps par phénol/chloroforme/alcool isoamyélique, puis par chloroforme/alcool isoamyélique. Le surnageant est récupéré et additionné de 20 µL d'acétate de sodium 3M, 1 µL de Glycoblue (Invitrogen) et 1 mL d'éthanol 100%. Les échantillons sont incubés une nuit ou deux heures minimum à -20°C. Une centrifugation des échantillons à 16000 g pendant 45 minutes à 4°C permet de précipiter les acides nucléiques qui sont ensuite lavés par 70% d'éthanol. Le culot d'acides nucléiques récupéré est repris dans 30 µL d'eau. Afin d'éliminer l'ADN des échantillons, un traitement DNase est appliqué en rajoutant 3 unités de DNase RQ1 (Promega) dans son tampon à 37°C pendant 30 minutes. La réaction est ensuite arrêtée en ajoutant le tampon Stop RQ1 (Promega) suivi d'une incubation de 10 minutes à 65°C. Les ARN d'intérêt sont ensuite analysés par RT-qPCR (voir partie 2.8.b et 2.8.c de ce chapitre).

*Résultats et
Discussion*

Chapitre 1 : Le facteur d'assemblage Bcd1p est un partenaire du complexe ATPasique Rvb1/Rvb2 chez la levure

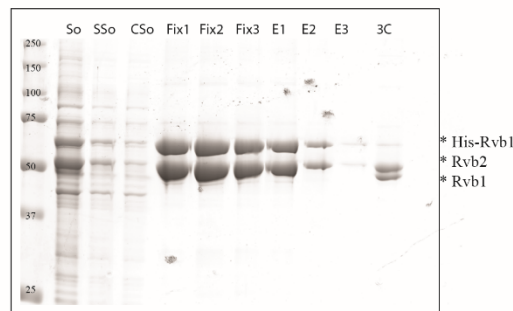
1. Etude du complexe formé entre Rvb1/Rvb2 et Bcd1p

1.1. L'état oligomérique du complexe Rvb1/Rvb2 varie selon son état nucléotidique

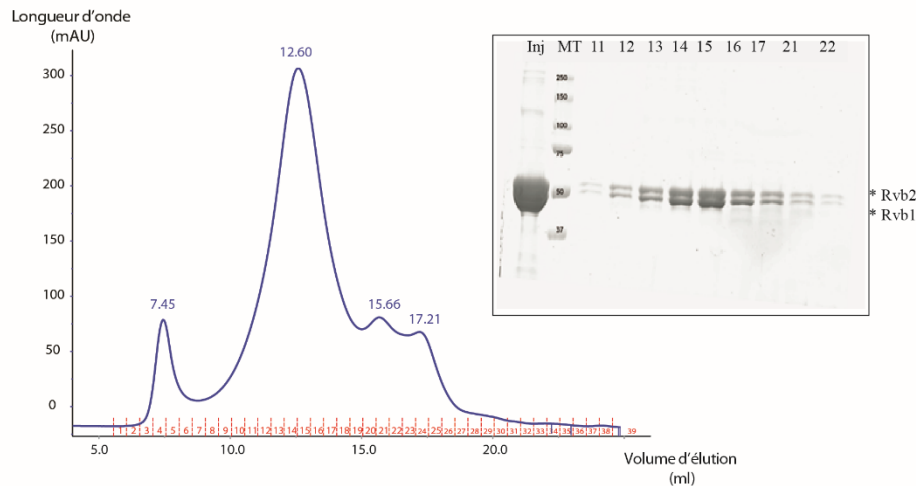
La purification du complexe ATPasique Rvb1/Rvb2 de levure a été faite de façon recombinante dans un système bactérien *E. coli*, en utilisant un vecteur d'expression pnEA3CH porteur d'une concaténation des deux séquences codantes, avec une étiquette 6xHistidines en position N-terminale sur la protéine Rvb1. Celle-ci permettra la purification et elle sera ensuite retirée grâce à un site de coupure à la protéase PreScission (3C) situé entre l'étiquette et la séquence codante. La co-expression simultanée de ces deux ATPases permet leur reconstitution directement dans *E. coli*, conduisant raisonnablement un assemblage final avec un état oligomérique le plus proche de celui observé *in vivo* dans les cellules de levure.

Nous avons dans un premier temps purifié le complexe sans nucléotide avec un tampon (10 mM HEPES KOH pH 7,5 ; 150 mM NaCl ; 1 mM DTT ; 1 % Glycérol). L'analyse par SDS-PAGE des fractions issues des différentes étapes de purification est illustrée dans la figure 26A. Rvb1 correspond à la bande du dessus avant coupure à la 3C et à la bande du dessous après coupure (dernière piste sur la droite). Nous avons ensuite analysé le complexe Rvb1/Rvb2 en chromatographie d'exclusion de taille (colonne superose 6 *increase* analytique). Le profil chromatographique (Figure 26B) indique la présence d'une forme majoritaire qui est éluée de la colonne ($V_1 = 12,60$ mL), suivie de deux autres pics de faibles intensités ($V_2 = 15,66$ mL et $V_3 = 17,21$ mL). L'analyse de ces fractions sur SDS-PAGE 10% confirme bien la présence des deux protéines Rvb1 et Rvb2 (Fractions 12→17 pour le pic majoritaire). En référence à la courbe étalon de la colonne utilisée (Figure 26C), nous avons pu déduire la masse moléculaire approximativement de chaque population protéique présente sous ces pics (la précision de la masse moléculaire prédite par la courbe étalon dépend de la forme de la protéine, plus elle est globulaire plus sa masse moléculaire prédite par la courbe étalon se rapproche de sa masse moléculaire réelle). L'espèce majoritaire à V_1 correspond à un hétérododécamère de Rvb1/Rvb2. Le second pic du chromatogramme correspond à la forme dimérique des Rvb (Rvb1/Rvb1, Rvb1/Rvb2 ou Rvb2/Rvb2) (Fractions 21 et 22). Enfin, le troisième pic correspond à l'élution des produits de dégradation des protéines et des éventuels contaminants de purification. Un pic est également présent en début du chromatogramme ($V_i = 7,45$ mL). Il

A



B



C

	Masse moléculaire (Da)	volume d'éluion (mL)	log MM
Thyroglobuline dimer	1340000	12,24	6,127104798
Thyroglobuline	670000	13,47	5,826074803
Gamma globuline	158000	15,47	5,198657087
ovalbumine	44000	16,99	4,643452676
myoglobine	17000	18,33	4,230448921
vitamine B12	1350	21,01	3,130333768
Rvb1/Rvb2 dodécamère	1137104	12,6	6,0558
Rvb1/Rvb2 dimère	121837	15,66	5,08578
Produit de dégradation / contaminants	39303	17,21	4,59443

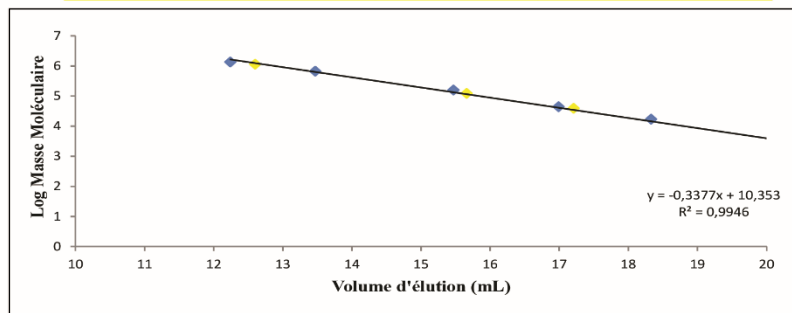


Figure 26 : Purification du complexe Rvb1/Rvb2 recombinant en l'absence de nucléotide
(A) Analyse par SDS-PAGE 10% des étapes de purification du complexe Rvb1/Rvb2. MT : Marqueur de Taille, So : Sonicat, Sso : Surnageant de sonication, Cso : Culot de sonication, Fix : Fixation sur résine TALON, E : Elution, 3C : coupure à la protéase PreScission 3C. **(B)** Chromatogramme des fractions d'éluion de Rvb1/Rvb2, sur une colonne d'exclusion de taille superose 6 *increase* pilotée par un système AKTA *purifier* : P : Pic, Vo : volume mort de la colonne, Vt : volume total de la colonne. Les fractions des pics principaux en sortie de colonne ont été analysées sur SDS-PAGE 10% (droite) : Inj : Injection. **(C)** Courbe étalon de la colonne superose 6 *increase* représentant la masse moléculaire en fonction des volumes d'éluions de chaque protéines mentionnées dans le tableau. Les produits de purification sont représentés sur la courbe et le tableau en jaune.

correspond au volume mort de la colonne, avec l'élution des agrégats et complexes de très grand poids moléculaire exclus par la colonne (Figure 26B).

Etant donné la grande diversité de conformations quaternaires du complexe Rvb1/Rvb2 selon les conditions de sa purification, nous avons testé sa purification avec des tampons différents, contenant de l'ADP ou de l'ATP (10 mM HEPES KOH pH 7,5 ; 150 mM NaCl ; 1 mM DTT ; 1 % Glycérol ; 0.5 mM ADP / ATP ; 5 mM MgCl₂). Les deux chromatogrammes (ADP/ATP) sont relativement identiques (Figures 27A et B). Leur analyse indique la présence d'un pic majoritaire ($V_1 = 13,84$ mL pour l'ADP et $V_1' = 14,28$ mL pour l'ATP) qui s'est déplacé vers la droite du chromatogramme, comparé à celui obtenu pour la condition sans nucléotide (Figure 26B). Les fractions collectées sous ces deux pics ont également été analysées sur SDS-PAGE 10% (Figure 27C). Nous constatons la présence des deux partenaires protéiques Rvb1 et Rvb2 avec une intensité égale (les fractions 17, 18 en ADP et les fractions 16, 17 en ATP). L'élution plus tardive des Rvb1/Rvb2 dans ces deux conditions comparées à l'élution sans nucléotides dans le tampon est synonyme d'un complexe de poids moléculaire plus faible. Selon la courbe étalon de la colonne utilisée (Figure 27D), nous avons pu déduire un poids moléculaire approximatifs en faveur d'un complexe hétérohexamérique (Rvb1/Rvb2) en présence d'ADP ou d'ATP. Dans ce cas, la différence de volume d'élution serait due à une conformation différente.

Nous avons constaté que selon l'ajout ou non de nucléotides dans le tampon de purification, la structure quaternaire du complexe Rvb1/Rvb2 était affectée entre hétérohexamérique et hétérododécamérique. Ces résultats sont en accords avec les données bibliographiques qui indiquent que la conformation oligomérique de ce complexe est très labile. Elle varie selon le nucléotide ajouté lors de sa purification mais également selon le partenaire protéique (Zhou et al., 2017 ; Tian et al., 2017 ; Lakomek et al., 2015). L'analyse du complexe Rvb1/Rvb2 +ADP en Diffusion Statique de la Lumière (DLS) montre un profil homogène avec une population dont la mesure moyenne du diamètre hydrodynamique (*Z-Average*) est de 16,46 nm +/- 5,7 (Figure 28, pic 1) avec un indice de polydispersité faible à 0,216.

Nous avons ensuite voulu connaître la stœchiométrie du complexe Rvb1/Rvb2 et s'il interagissait directement avec des molécules d'ADP. Pour cela, une analyse par spectrométrie de masse native a été réalisée par le Dr Evolène Deslignière au laboratoire de spectrométrie de masse bio-organique à Strasbourg (Figure 29). Le profil chromatographique de l'échantillon après passage sur une micro-colonne d'exclusion de taille révèle 3 pics : $V_{E1} = 10,42$ mL, $V_{E2} = 11,56$ mL et $V_{E3} = 12,25$ mL. L'analyse par spectrométrie de masse native de chaque pic est représentée avec les trois spectres de dessous dans la figure 29 (Intensité relative en fonction

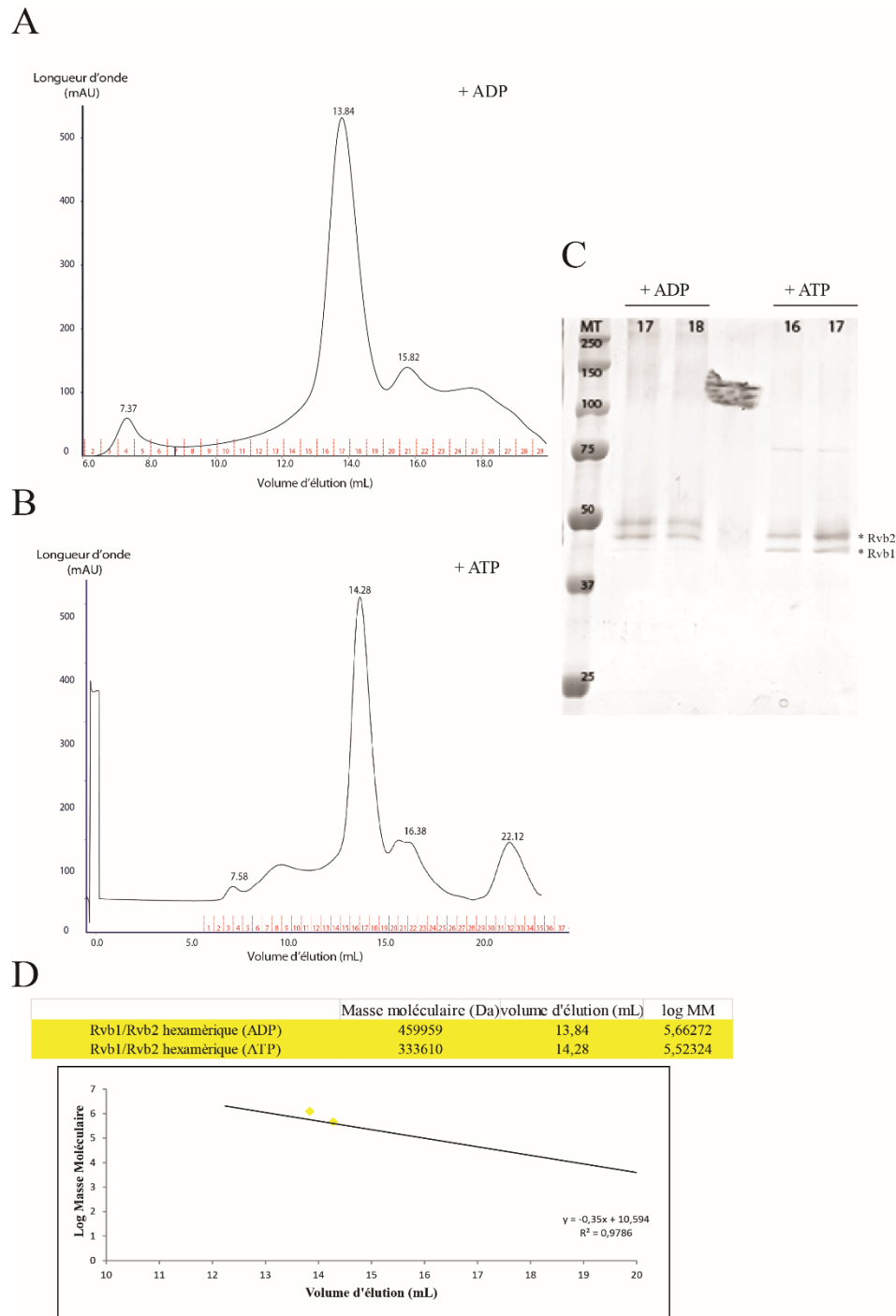


Figure 27 : Purification du complexe Rvb1/Rvb2 en présence de nucléotides (ADP/ATP) dans le tampon. (A) Chromatogramme du complexe Rvb1/Rvb2 en présence d'ADP (0,5 mM) dans le tampon de filtration sur gel en colonne superose 6 *increase* pilotée par un système AKTA *purifier*. P : Pic, Vo : Volume mort de la colonne, Vt : volume total de la colonne. (B) Chromatogramme du complexe Rvb1/Rvb2 en présence d'ATP (0,5 mM) dans le tampon de filtration sur gel en colonne superose 6 *increase* sur système AKTA *purifier*. P : Pic, Vo : Volume mort de la colonne, Vt : volume total de la colonne. (C) Analyse par SDS-PAGE 10% des fractions des deux pics en sortie de colonne, 16 et 17 pour l'ATP et 17 et 18 pour l'ADP. (D) Courbe étalon de la colonne superose 6 *increase* représentant la masse moléculaire en fonction des volumes d'élutions. Les produits de purification sont représentés sur la courbe et le tableau en jaune.

	Taille (d.n...	% Intensité	St Dev (d.n...
Z-Average (d.nm): 15,63	Pic 1: 16,46	94,5	5,706
Indice de polydispersité 0,216	Pic 2: 3778	5,5	1148
Intercept: 0,952	Pic 3: 0,000	0,0	0,000

Qualité des résultats **Good**

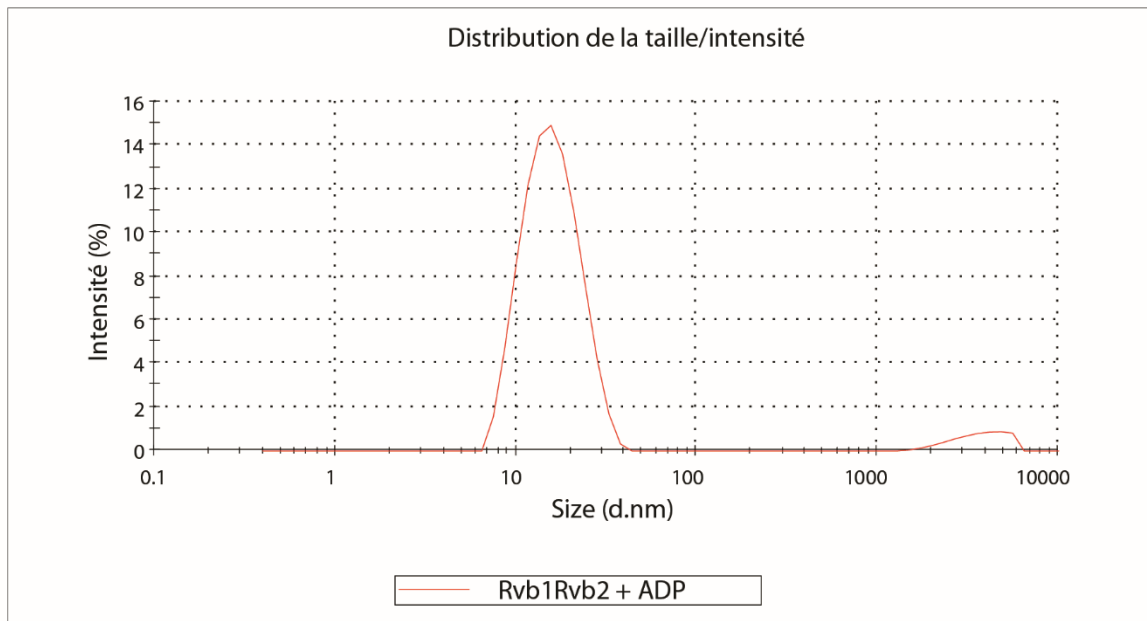


Figure 28 : Analyses biophysiques du complexe Rvb1/Rvb2 + ADP.

Résultats de la Diffusion Statique de la Lumière (*Dynamic Light Scattering* DLS) sur le complexe Rvb1/Rvb2 +ADP. La valeur de Z-Average correspond au diamètre moyen des particules en solution. Le PdI est l'indice de polydispersité et évalue la pureté de l'échantillon. Un faible PdI signifie que l'échantillon en solution est monodisperse (une seule taille de particule), et donc très pur.

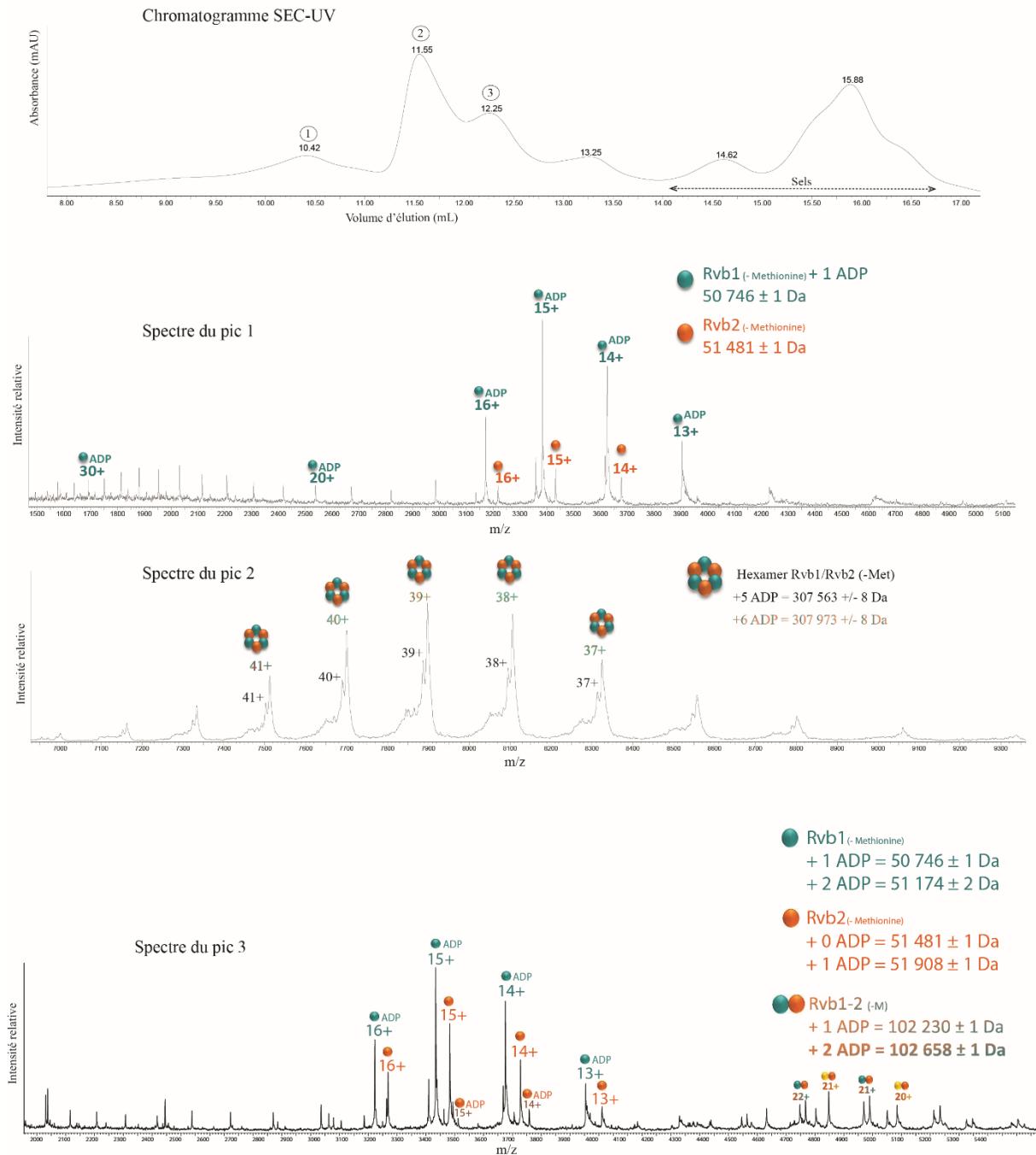


Figure 29 : Analyse du complexe Rvb1/Rvb2 purifié en présence d'ADP par spectrométrie de masse native

Chromatogramme du complexe sur colonne d'exclusion de taille BEH200 de chez Waters (haut), suivi d'une analyse des pics 1 à 3 par spectrométrie de masse native. Le pic 2 majoritaire correspond au complexe Rvb1/Rvb2. L'analyse de son enveloppe de charge révèle la présence de monomères et de deux populations d'hexamères selon le nombre de molécules de nucléotides (+ 5 ADP/ + 6 ADP). Les pics 1 et 3 correspondent respectivement à des monomères de Rvb1 et Rvb2 et des monomères avec des hétérodimères Rvb1/Rvb2. D'après les résultats obtenus au LSMBO à Strasbourg.

de la masse sur la charge, « m/z »). Le pic 1 correspond à des molécules de grandes tailles (agrégats). Son spectre indique qu'il est composé essentiellement des deux protéines Rvb1 + 1 molécules d'ADP et Rvb2. Le spectre du pic 2 correspond à un complexe hétérohexamérique 3Rvb1/3Rvb2 avec deux populations distinctes en fonction du nombre de molécule d'ADP : jusqu'à 5 ou 6 molécules d'ADP (3 Rvb1/3 Rvb2/5 ADP et 3 Rvb1/3 Rvb2/6 ADP). Le spectre du pic 3 représente des formes hétérodimériques Rvb1/Rvb2 + 1 ou 2 molécules d'ADP ainsi que des protéines libres Rvb1 et Rvb2. Ce résultat confirme bien la présence de molécules d'ADP potentiellement dans le site actif de Rvb1 et Rvb2 après leur purification de façon recombinante chez *E. coli*. Le complexe Rvb1/Rvb2 étudié comprend 5 à 6 molécules d'ADP. L'analyse par spectrométrie de masse native du complexe Rvb1/Rvb2 purifié en absence de nucléotide n'a malheureusement pas pu aboutir. Cela est essentiellement dû à sa grande instabilité *in vitro* (agrégation) à des concentrations supérieures à 35 μM testées au laboratoire.

1.2. Reconstitution du complexe Rvb1/Rvb2/Bcd1p

1.2.1. Co-élution sans nucléotide

Dans le but d'une reconstitution *in vitro* du complexe Rvb1/Rvb2/Bcd1p, nous avons dans un premier temps, purifié les partenaires protéiques séparément, le complexe Rvb1/Rvb2 sans nucléotide (20 mM Hépes pH 7,5 ; 150 mM NaCl ; 0,5 mM TCEP ; 1% Glycérol) et Bcd1p (20 mM Hépes pH 7,5 ; 150 mM NaCl ; 0,5 mM TCEP), puis nous les avons mélangés dans un ratio 1 : 6 respectivement, et concentrés par ultrafiltration par centrifugation sur membrane de concentration (Amicon 30 kDa, MerckMillipore). L'échantillon a été injecté sur une chaîne de purification (AKTA *purifier*, GE Healthcare) connectée à une colonne Superose 6 analytique (10/300 *increase*, GE Healthcare). Cette dernière va permettre de séparer les protéines et complexes protéiques contenus dans l'échantillon injecté, avec un ordre inverse de celui de leurs poids moléculaires (chromatographie d'exclusion de taille). Le profil du chromatogramme obtenu pour le mélange Rvb1, Rvb2 et Bcd1p (en bleu, Figure 30A) comprend deux pics dont les volumes d'élutions sont : $V_{P1} = 12,56 \text{ mL}$ et $V_{P2} = 16,20 \text{ mL}$. Une analyse des fractions sous ces deux pics a été faite sur un gel SDS-PAGE 12,5% (Figure 30B). Le pic P1 (Fractions 12, 13 et 14) représente l'élution du complexe Rvb1/Rvb2 seul et le pic P2 (Fractions 24, 25, 26, 27 et 28) correspond à l'élution de Bcd1p seul. De plus, la superposition des profils chromatographiques de Bcd1p seul (vert) et Rvb1/Rvb2 (rose) de la figure 26 (Figure 30A) confirme que le volume d'élution du P1 correspond au volume d'élution du complexe hétérododécamérique Rvb1/Rvb2 seul et le volume d'élution de P2 est similaire à celui de

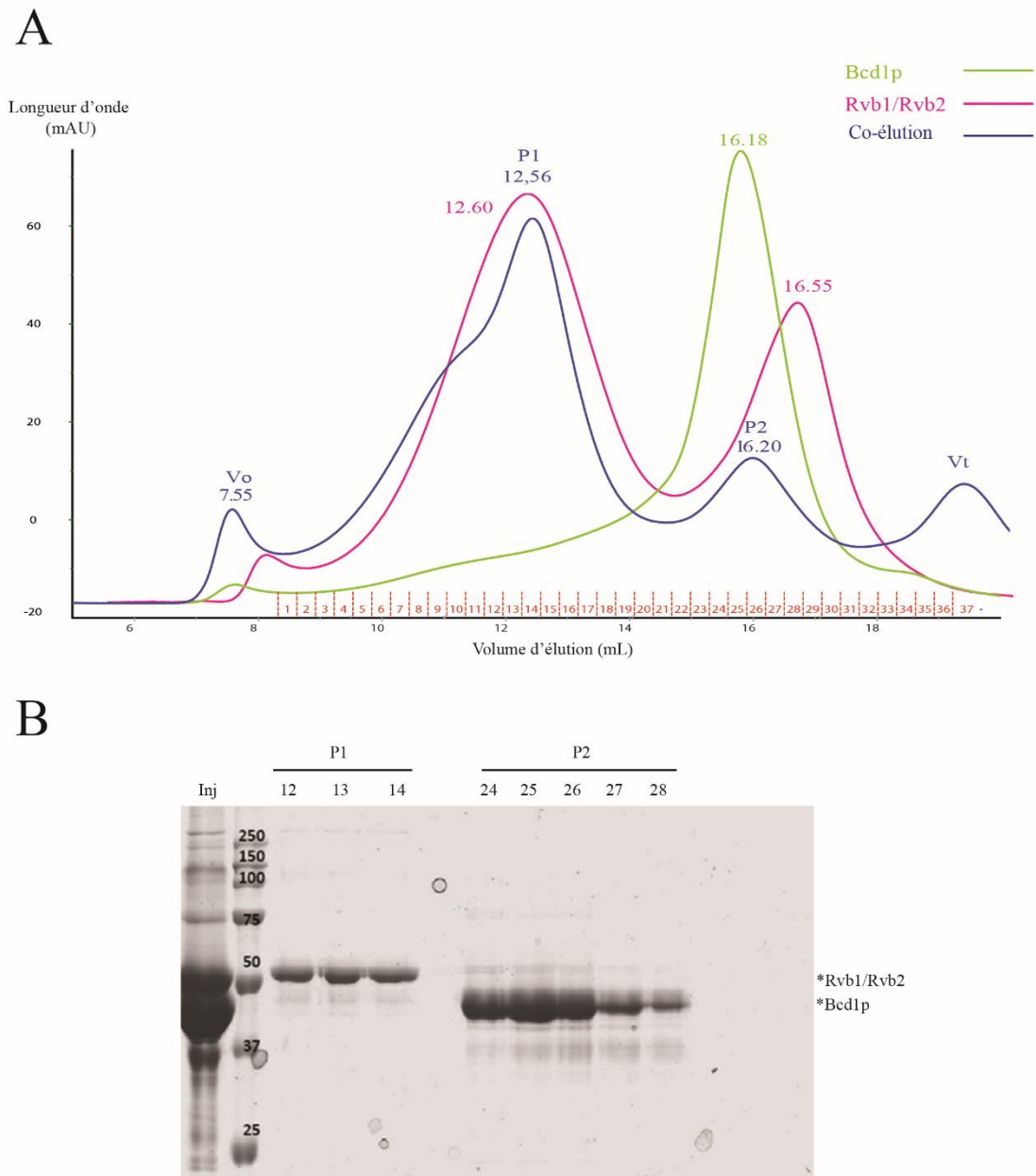


Figure 30 : Analyse de l'élution du complexe Rvb1/Rvb2 et la protéine Bcd1 sans nucléotide.

(A) Superposition de plusieurs chromatogrammes correspondant à : Bleu, l'élution du complexe du mélange Rvb1/Rvb2 et Bcd1p sur colonne superose 6 *increase*, P : Pic, Vo : volume mort de la colonne, Vt : volume total de la colonne. Vert, l'élution de Bcd1p seul sur la même colonne. Rose, l'élution du complexe Rvb1/Rvb2 seul sur la même colonne (Figure 26). (B) Le gel SDS-PAGE 12,5% correspond à l'analyse des fractions en sortie de colonne de l'élution du mélange Rvb1/Rvb2 et Bcd1p (bleu), Inj : Injection sur colonne.

Bcd1p libre. Ce résultat indique clairement que nous n'avons pas de formation de complexe entre Rvb1/Rvb2 et Bcd1p dans ces conditions expérimentales. Nous nous sommes demandés si cette interaction ne nécessitait pas une conformation particulière du complexe Rvb1/Rvb2 pour qu'elle puisse être établie ou/et la présence de nucléotides.

1.2.2. Co-élution en présence d'ADP ou d'ATP

La reconstitution *in vitro* du complexe Rvb1/Rvb2/Bcd1 s'est donc poursuivie par une optimisation des conditions expérimentales, et plus précisément par un ajout d'ADP ou d'ATP dans les tampons de purification des partenaires protéiques mais aussi lors de leur co-élution sur colonne. L'interaction des protéines homologues humaines, BCD1 et RUVBL1/RUVBL2, a en effet été identifiée exclusivement en présence d'ADP ou d'ATP par *GST pull-down* (McKeegan et al., 2007).

Les purifications de Rvb1/Rvb2 et Bcd1p ont été faites séparément dans un tampon (20 mM Hépes pH 7,5 ; 150 mM NaCl ; 0,5 mM TCEP ; 1% Glycérol 0,5 mM ADP ou ATP ; 5 mM MgCl₂). Comme pour la première condition de reconstitution du complexe, les deux partenaires Bcd1p et Rvb1/Rvb2 ont été mélangés à un ratio 1 : 6 respectivement, et concentrés *via* des systèmes d'ultrafiltration par centrifugation (Amicon 30 kDa, Merck Millipore). L'échantillon a été injecté sur une chaîne de purification (AKTA *purifier*, GE Healthcare) connectée à une colonne Superose 6 analytique (10/300 *increase*, GE Healthcare). L'analyse des deux profils chromatographiques obtenus (bleu) en présence d'ADP dans le tampon (Figure 31A) ou d'ATP (Figure 31B) montre deux pics pour chacune des deux conditions testées : P1/P'1 et P2/P'2. Les volumes d'élution sont du même ordre de grandeur en ADP ou en ATP. En fonction de la courbe de calibration de la colonne, les premiers pics P1 et P'1 correspondraient à une forme dodécamérique des Rvb1/Rvb2. Nous remarquons en effet un décalage vers la gauche des pics P1 et P'1 (bleu) avec un volume d'élution plus court (bleu) : $V_{P1} = 12,49$ mL et $V_{P'1} = 12,37$ mL par rapport aux volumes d'élution du complexe Rvb1/Rvb2 seul, en ADP ou en ATP (rose) (Figure 31). Pour rappel, les Rvb1/Rvb2 libres en présence de nucléotides sont majoritairement sous forme hexamérique (Figure 27). L'analyse des fractions sous ces pics (gel SDS-PAGE 12,5% à droite de chaque chromatogramme) montre la présence de deux bandes pour les pics P1 et P'1 qui correspondent aux masses moléculaires de Rvb1 + Rvb2 et Bcd1p. Les pics P2 et P'2 correspondent à l'élution de l'excès de Bcd1p libre et les Rvb1 et Rvb2 non liées en complexe (Figure 31). Ces résultats indiquent la co-élution des Rvb1/Rvb2 et de Bcd1p et donc la formation d'un complexe en ADP et en ATP. L'interaction de Bcd1p avec le complexe

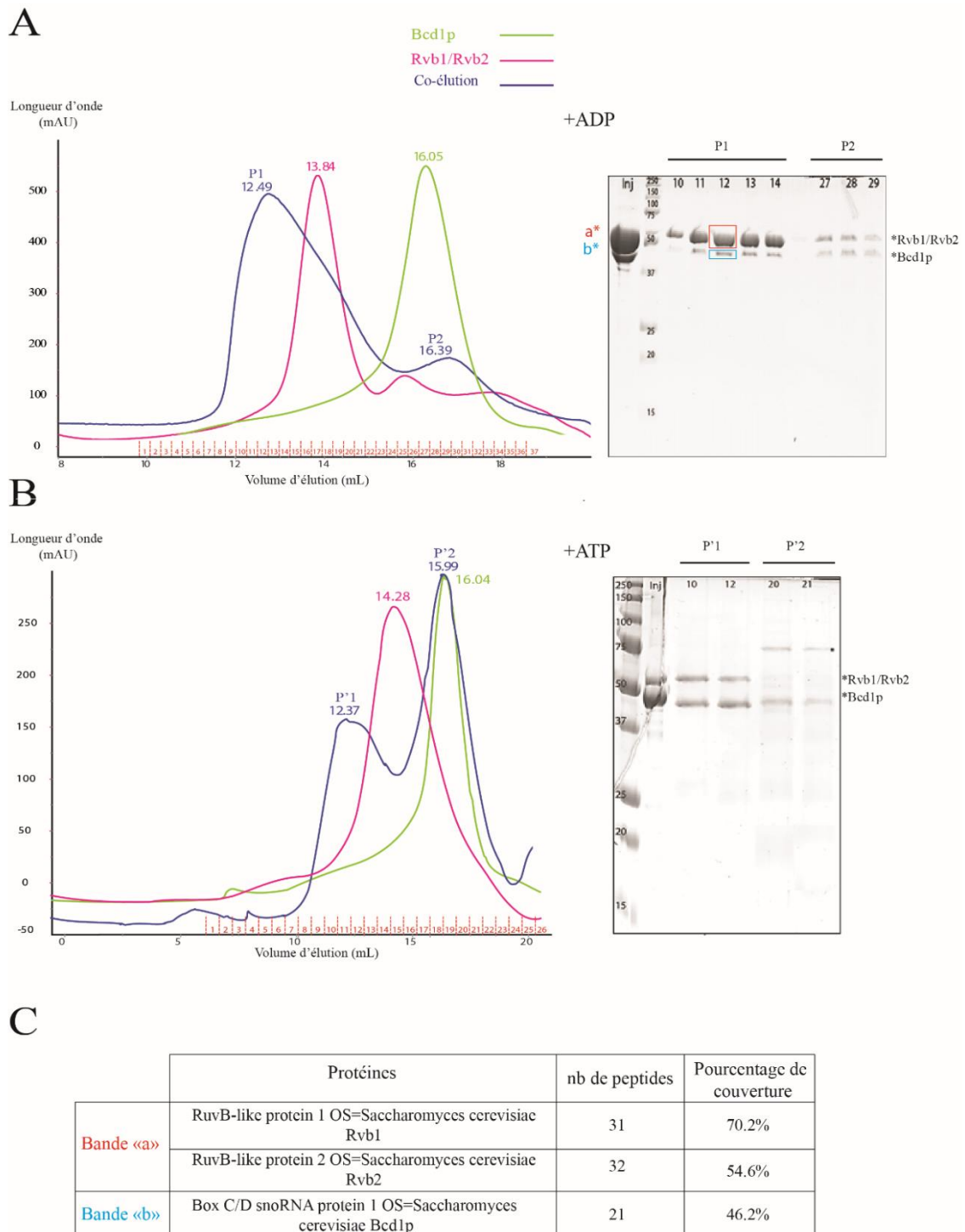


Figure 31 : Co-élution du complexe entre Rvb1/Rvb2 et Bcd1p en présence de nucléotides dans le tampon de purification (ADP/ATP).

Superposition de plusieurs chromatogrammes correspondant à : Bleu, la co-élution du complexe Rvb1/Rvb2 et Bcd1p sur colonne superose 6 *increase*, P : Pic, Vo : volume mort de la colonne, Vt : volume total de la colonne. Vert, l'élution de Bcd1p seul sur la même colonne et dans les mêmes conditions. Rose, l'élution du complexe Rvb1/Rvb2 seul sur la même colonne et dans les mêmes conditions. (A) ADP, (B) ATP. Les gels SDS-PAGE 12,5% correspondent à l'analyse des fractions des deux pics majoritaires en sortie de colonne de la co-élution de Rvb1/Rvb2 et Bcd1p (Bleu), Inj : Injection sur colonne. (C) Résultats de l'analyse protéomique des bandes « a » et « b » effectuée sur la plateforme protéomique de l'UMS 2008/US40 IBS Lor.

Rvb1/Rvb2 est donc dépendante de la présence d'ADP ou d'ATP certainement dans leurs sites actifs. Il est à noter que l'ATP s'hydrolyse spontanément en ADP *in vitro*, même en condition ATP cela pourrait être l'ADP qui serait moteur de l'interaction. Nous avons également analysé les bandes de gel SDS-PAGE qui correspondent aux masses moléculaires des deux partenaires : « a » et « b » de la fraction 12 en condition +ADP par spectrométrie de masse MS/MS (analyse protéomique) au laboratoire IMoPA. Nous retrouvons bien la présence majoritaire des trois protéines : Rvb1 (70,2% de couverture), Rvb2 (54,6% de couverture) sous la même bande et Bcd1p (46,2% de couverture) (Figure 31C).

Comme cité plus haut dans le chapitre d'introduction, une interaction a déjà été identifiée par RMN au laboratoire par B. Bragantini, entre le doigt à zinc de Bcd1p de levure et le complexe orthologue RUVBL1/RUVBL2 humain. Il a également été montré que les domaines doigt à zinc de Hit1p et Bcd1p sont fortement similaires. En effet, des expériences menées au laboratoire ont montré que la protéine chimérique ZHD_{Hit1p}-Bcd1p₄₆₋₃₆₆ est fonctionnelle chez la levure (Bragantini et al., 2016). Nous avons donc voulu tester si le domaine ZHD_{Hit1p} interagit lui aussi avec le complexe Rvb1/Rvb2 *in vitro*. Une co-élution entre Hit1p et Rvb1/Rvb2 en présence d'ADP dans le tampon a été réalisée. Pour cela, Hit1p a été co-purifiée en complexe avec le fragment de Rsa1p₂₃₁₋₃₇₅ qui assure sa stabilité *in vitro*. Les deux partenaires Rvb1/Rvb2 et Rsa1p₂₃₁₋₃₇₅/Hit1p ont été mélangés à un ratio 1 : 6 respectivement, et concentrés via des systèmes d'ultrafiltration par centrifugation (Amicon 10 kDa, MerckMillipore). L'échantillon a été injecté sur une chaîne de purification (AKTA purifier, GE Healthcare) connectée à une colonne Superose 6 analytique (10/300 increase, GE Healthcare). Le profil chromatographique obtenu (bleu) (Figure 32A) est composé de deux pics P1 avec $V_{P1} = 13,87$ mL et P2 avec $V_{P2} = 17,42$ mL. L'analyse des fractions sous les pics P1 et P2 sur gel SDS-PAGE 10% montre deux bandes qui migrent aux poids moléculaires de Rvb1 et Rvb2 pour P1 (Fractions 13 et 14) et deux bandes qui migrent aux poids moléculaires de Rsa1p₂₃₁₋₃₇₅ et Hit1p pour P2 (Fractions 25 et 26) (Figure 32B). Ce résultat indique qu'il n'y a pas de co-élution et donc pas de formation de complexe entre Rvb1/Rvb2 et Rsa1p₂₃₁₋₃₇₅/Hit1p en présence d'ADP dans le tampon. Le domaine en doigt à zinc ne semble donc pas suffisant à lui seul pour établir une interaction stable avec le complexe Rvb1/Rvb2. La contribution d'autres acides aminés contenus dans une autre région de Bcd1p serait donc nécessaire pour consolider cette interface d'interaction. Bcd1p₁₋₉₆ (porteur du domaine doigt à zinc en région N-terminale 1-45) a été décrit précédemment dans les travaux de D. Tiotu comme fragment minimum nécessaire pour la fonction de Bcd1p chez la levure. Il est probable que la région Bcd1p₄₅₋₉₆ soit importante pour la stabilisation de l'interaction entre Bcd1p et Rvb1/Rvb2. Une autre possibilité serait que

l'interaction de Rsa1p avec la partie C-terminale de Hit1p puisse gêner de manière stérique, l'interaction de la partie N-terminale de Hit1p avec les Rvb1/2.

1.2.3. Caractérisation biophysique du complexe Rvb1/Rvb2/Bcd1p

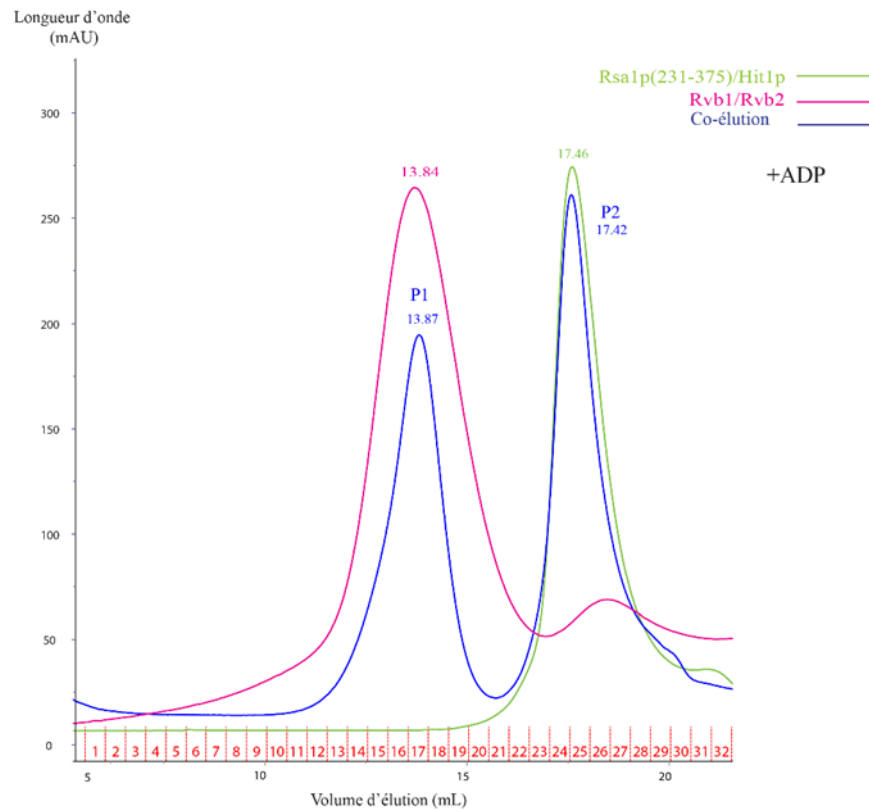
a. Diffusion dynamique de la lumière (*Dynamique Light Scattering, DLS*)

Le complexe formé entre Rvb1/Rvb2 et Bcd1p en ADP a été analysé par DLS (Figure 33A) et comparé avec le profil DLS du complexe Rvb1/Rvb2 libre toujours en ADP (Figure 28). La distribution de taille montre un profil homogène quasiment monodisperse avec une population dont la taille moyenne du diamètre hydrodynamique (pic 1) est de 19,39 nm. L'indice de polydispersité de l'échantillon est de 0,221. On note une augmentation de taille significative ($\text{Diamètre}_{\text{Rvb1/Rvb2}} = 15,46 \text{ nm}$ et $\text{Diamètre}_{\text{Rvb1/Rvb2/Bcd1p}} = 19,39 \text{ nm}$). Cela peut s'expliquer par l'interaction de plusieurs molécules de Bcd1p sur le complexe Rvb1/Rvb2 mais plus vraisemblablement par un changement conformationnel du complexe Rvb1/Rvb2 vu l'écart important. D'après les résultats de filtration sur gel (Figure 31), il s'agirait du passage de l'état hexamérique à dodécamérique des Rvb1/2.

b. Titration calorimétrique isotherme (ITC)

L'affinité ainsi que les paramètres thermodynamiques de l'interaction entre le complexe Rvb1/Rvb2 et la protéine Bcd1 purifiées en présence d'ADP dans le tampon, ont été déterminés par ITC sur l'appareil VP-ITC de l'UMS IBS Lor. L'échantillon correspondant au complexe Rvb1/Rvb2 a été déposé dans la cellule de l'appareil. La seringue a été remplie avec l'échantillon de la protéine Bcd1. Rvb1/Rvb2 ainsi que Bcd1p ont été purifiés dans le même tampon (20 mM Hépes pH 7,5 ; 150 mM NaCl ; 0,5 mM TCEP ; 1% Glycérol ; 0,5 mM ADP ; 5 mM MgCl_2) afin d'éviter toute variation thermodynamique détectée par l'appareil pouvant être dû au tampon. Le résultat montre une affinité de l'ordre du micromolaire (0,328 μM) (Figure 33B), avec un nombre de sites (N) proche de 2. Cela signifie que nous aurions probablement deux protéines Bcd1 qui interagissent avec un hétéro-complexe de Rvb1/Rvb2. ΔS et ΔH sont des paramètres thermodynamiques de variation d'entropie et d'enthalpie respectivement. La variation d'enthalpie négative de notre réaction signifie que nous sommes en présence d'une réaction exothermique. Au final ces résultats confirment bien la formation d'un

A



B

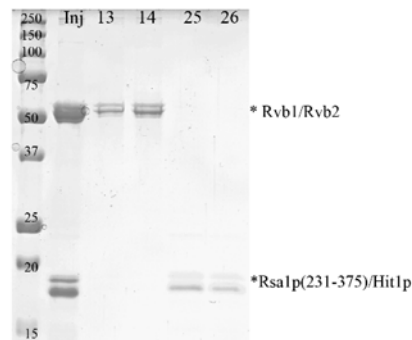


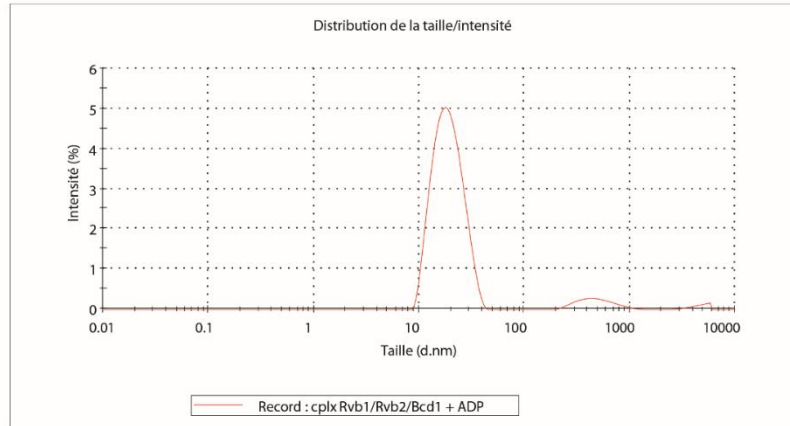
Figure 32 : Analyse de la co-élution du complexe Rvb1/Rvb2 et Rsa1p₂₃₁₋₃₇₅/Hit1p en présence de nucléotide (ADP) dans le tampon de purification.

(A) Superposition de plusieurs chromatogrammes correspondant à : Bleu, l'élution des complexes Rvb1/Rvb2 et Rsa1p₂₃₁₋₃₇₅/Hit1p sur colonne superose 6 *increase*, P : Pic. Vert, l'élution de Rsa1p₂₃₁₋₃₇₅/Hit1p seul sur la même colonne. Rose, l'élution du complexe Rvb1/Rvb2 seul sur la même colonne dans les mêmes conditions. (B) Les gels SDS-PAGE 10% correspondent à l'analyse des fractions en sortie de colonne de l'élution de Rvb1/Rvb2 et de Rsa1p₂₃₁₋₃₇₅/Hit1p (Bleu), Inj : Injection sur colonne.

A

	Taille (d.n...	% Intensité	St Dev (d.n...
Z-Average (d.nm):	19,33		
Indice de polydispersité:	0,221		
Intercept:	0,938		
	Pic 1: 19,39	92,8	6,091
	Pic 2: 492,7	5,9	189,2
	Pic 3: 4642	1,3	806,5

Qualité des résultats **Bonne**



B

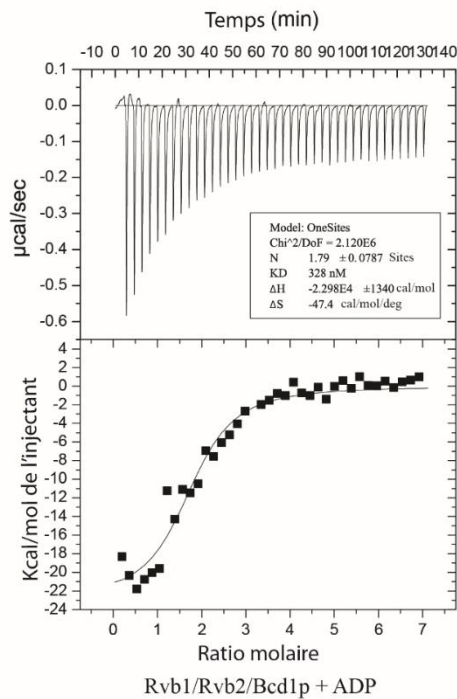


Figure 33 : Le complexe Rvb1/Rvb2/Bcd1p se forme *in vitro* en présence d'ADP

(A) Résultats de DLS sur le complexe Rvb1/Rvb2/Bcd1p +ADP. La valeur de Z-Average correspond au diamètre moyen des particules en solution. Le PDI est l'indice de polydispersité, et évalue la pureté de l'échantillon. Un faible PDI signifie que l'échantillon en solution est monodisperse (une seule taille de particule) et donc pur (B) Résultat de titration calorimétrique interne (modèle VP-ITC) à 20°C entre le complexe Rvb1/Rvb2 et Bcd1p, purifiés dans le même tampon (10 mM Hépes pH=7,5 ; 150 mM NaCl ; 1% Glycerol ; 0,5 mM TCEP ; 0,5 mM ADP).

complexe Rvb1/Rvb2/Bcd1p avec une affinité importante en présence de nucléotide (ADP) dans le tampon.

1.2.4. La présence de l'ADP dans les sites actifs de Rvb1 et Rvb2 est nécessaire à la formation du complexe Rvb1/Rvb2/Bcd1p

Dans le but d'approfondir la caractérisation de l'interaction de Bcd1p avec le complexe Rvb1/Rvb2, nous avons voulu savoir à quelle étape du cycle ATPasique l'interaction avec Bcd1p est établie. L'utilisation d'analogues non hydrolysables de l'ATP permet l'étude du cycle d'hydrolyse des ATPases. Il est à noter que ces analogues d'ATP ne résistent pas complètement à l'hydrolyse. Cela dépend du type d'analogue utilisé ainsi que des familles d'ATPases étudiées. L'ATP- γ S (adénosine 5'-(gamma-thio-triphosphate)) et l'AMP-PNP (adénylyl imidodiphosphate) sont les deux analogues les plus utilisés dans l'étude des hélicases (Lacabanne et al., 2020). Nous avons donc utilisé dans nos expériences ces deux analogues sensés figer un état pré-hydrolytique relativement stable du complexe Rvb1/Rvb2.

Nous avons dans un premier temps purifié les partenaires protéiques : Rvb1/Rvb2 et Bcd1p séparément dans un tampon contenant l'ATP- γ S ou l'AMP-PNP (20 mM Hépés pH 7,5 ; 150 mM NaCl ; 0,5 mM TCEP ; 1% Glycérol ; 0,5 mM ATP- γ S ou AMP-PNP ; 5 mM MgCl₂). Nous les avons ensuite mélangés dans un ratio 1:6 (Rvb1/Rvb2 et Bcd1p respectivement). Nous les avons concentrés via des systèmes d'ultrafiltration par centrifugation (Amicon 30 kDa, Merck Millipore). L'échantillon a été injecté sur une chaîne de purification (AKTA purifier, GE Healthcare) connectée à une colonne Superose 6 analytique (10/300 increase, GE Healthcare). Les résultats obtenus (Figures 34A, 34B) montre un profil chromatographique identique pour les deux analogues testés. Ils sont constitués de deux pics chacun : P1/P'1 et P2/P'2 (bleu) (Figures 34A et 34B). La comparaison de ces profils avec les profils d'élution de Bcd1p et Rvb1/Rvb2 non liés montre un volume d'élution identique entre $V_{P1/P'1} = V_{Rvb1/Rvb2}$ et $V_{P2/P'2} = V_{Bcd1p}$. L'analyse des fractions sous les pics P1/P'1 et P2/P'2 sur gel SDS-PAGE (Figure 34C) indique la présence de Rvb1 et Rvb2 qui migrent de façon proche dans le gel sous les pics P1 et P'1, et Bcd1p seul sous les pics P2 et P'2. La reconstitution du complexe Rvb1/Rvb2/Bcd1p n'est donc pas possible dans ces conditions. Cela signifie que les analogues ATP- γ S et l'AMP-PNP favorisent un état du complexe Rvb1/Rvb2 qui ne permettrait pas une interaction avec Bcd1p. Vu les valeurs des volumes d'élution proches de celui des Rvb1/2 en

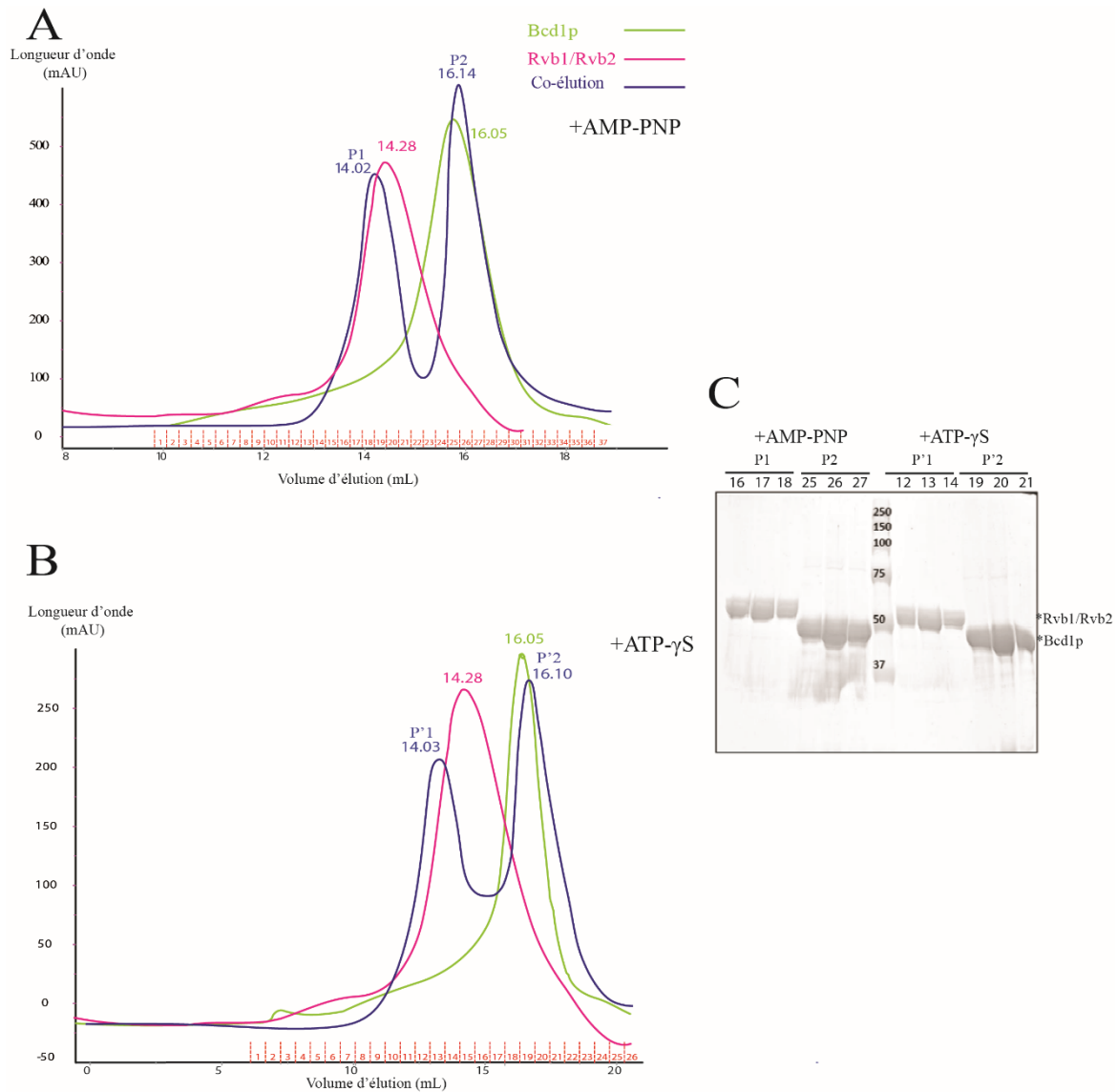


Figure 34 : Analyse de l'éluion du complexe Rvb1/Rvb2 et de Bcd1p en présence de nucléotides analogues non hydrolysables de l'ATP dans le tampon de purification (ATP-γS et AMP-PNP).

Superposition de plusieurs chromatogrammes correspondant à : Bleu, l'éluion du complexe Rvb1/Rvb2 et Bcd1p sur la colonne superose 6 *increase*, P : Pic. Vert, l'éluion de Bcd1p seul sur la même colonne. Rose, l'éluion du complexe Rvb1/Rvb2 seul sur la même colonne. (A) AMP-PNP (B) ATP-γS (C) Les gels SDS-PAGE 12.5% correspondent à l'analyse des fractions en sortie de colonne de l'éluion de Rvb1/Rvb2 et Bcd1p (Bleu), Inj : Injection sur colonne.

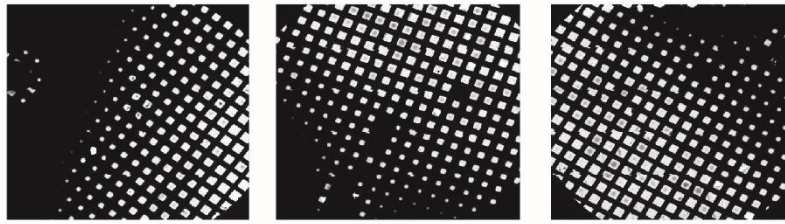
ATP (Figure 31B), il s'agit de l'état hétéro-hexamérique. Cela suggère que l'état post-hydrolyse de l'ATP (ie la présence d'ADP dans les sites actifs de Rvb1 et Rvb2) serait nécessaire à la formation du complexe Rvb1/Rvb2/Bcd1p.

2. Etude structurale du complexe Rvb1/Rvb2/Bcd1p

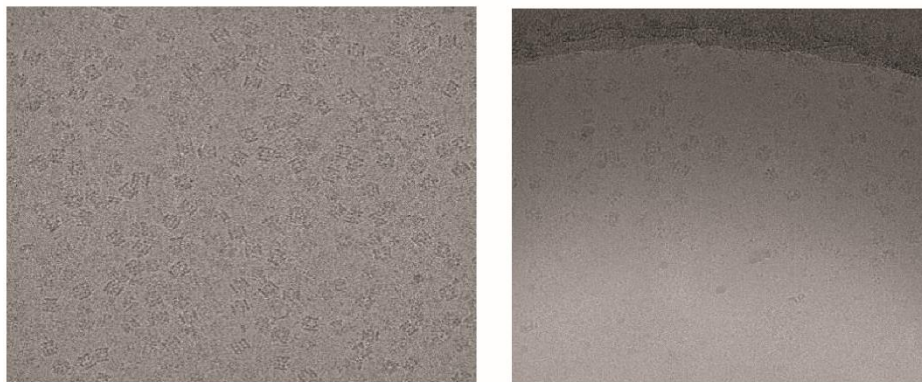
Nous avons voulu par la suite en savoir plus sur ce complexe et notamment comment le facteur protéique Bcd1 se positionne par rapport à l'hétéro-complexe Rvb1/Rvb2. Pour tester l'hypothèse où les DII seraient responsables de la formation du complexe Rvb1/Rvb2/Bcd1p, des images de microscopie électronique à hautes résolution ont été acquises sur un Titan Krios (300 kV) par une équipe de la plateforme de cryo-microscopie électronique de Thermo Fisher Scientific au Pays-Bas. L'échantillon du complexe Rvb1/Rvb2/Bcd1p que j'ai préparé a été obtenu par co-élution sur une colonne de filtration sur gel en présence d'ADP dans le tampon de purification classique. Le complexe a été analysé à une concentration de 1,14 mg/mL, sans dilution préalable. Dans un premier temps, l'observation de l'échantillon en coloration négative est une étape nécessaire car elle permet de vérifier l'homogénéité ainsi que la pureté de l'échantillon. Un atlas du criblage automatisé de 12 grilles a été effectué afin d'estimer les zones avec la bonne qualité et l'épaisseur de glace des grilles obtenues (Figure 35). Les particules sont sélectionnées à partir des micrographies qui informent sur la répartition du complexe. Cela permet également de les centrer et les aligner. La classification et le moyennage des images permettent d'améliorer le rapport signal sur bruit, et les moyennes de classe peuvent être utilisées pour calculer un modèle initial du complexe Rvb1/Rvb2/Bcd1p étudié à faible résolution. Les images moyennes triées en classes 2D (246 520 particules) (Figure 35) indique la présence d'une forme en tonneau (dodécamère) formée par les deux anneaux hexamériques qui interagissent par leurs domaines DII flexibles (hétérogénéité visible au centre des particules). Nous avons également observé des conformations courbées et droites du dodécamère Rvb1/Rvb2/Bcd1p (Figure 36). Ces caractéristiques sont similaires à celles rapportées dans des structures cryo-EM récemment publiées de dodécamères Rvb1/Rvb2 libre (Ewens et al., 2016, Silva-Martin et al., 2016).

Pour approfondir l'étude de la dynamique conformationnelle dans ce complexe, toutes les particules utilisées pour la classification 2D ont également été utilisées pour le raffinement 3D afin de générer une carte cryo-EM. Il existe 3 populations de classe 3D à haute résolution (Figure 36) : la première classe présente une résolution de 3,64 Å représentant 61% des particules analysées. Elle représente la conformation N°1. Les deux anneaux hexamériques sont parfaitement parallèles avec une augmentation de masse à l'interface externe des deux

Criblage de 12 grilles (Glacios Cryo-TEM)



Vérification de la répartition et l'homogénéité des particules



Collecte des données (Titan Krios)
Classification 2D

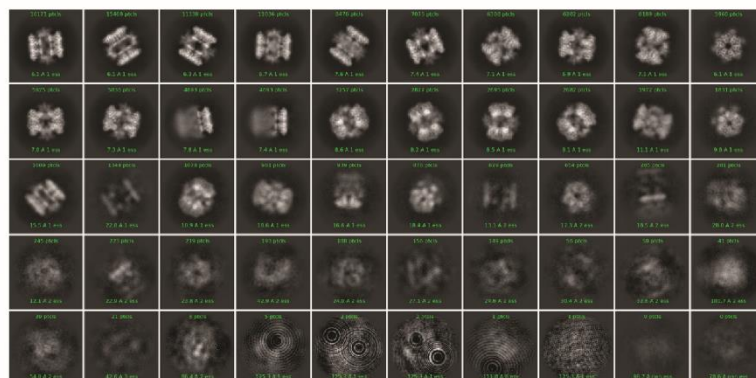


Figure 35 : Classification 2D obtenue après tri et sélection des images de cryo-microscopie électronique des particules du complexe Rvb1/Rvb2/Bcd1p.

Processus de l'analyse par cryo-microscopie électronique. 12 grilles sont analysées sur deux sessions (Glacios Cryo-TEM 200 kV, et Titan Krios 300 kV 360 images/heure). La température est de 4°C, le temps de contact de 5-7 secondes et il n'y a pas eu de dilution de l'échantillon envoyé. Le *screen Atlas* automatisé a permis de choisir la qualité de glace optimale.

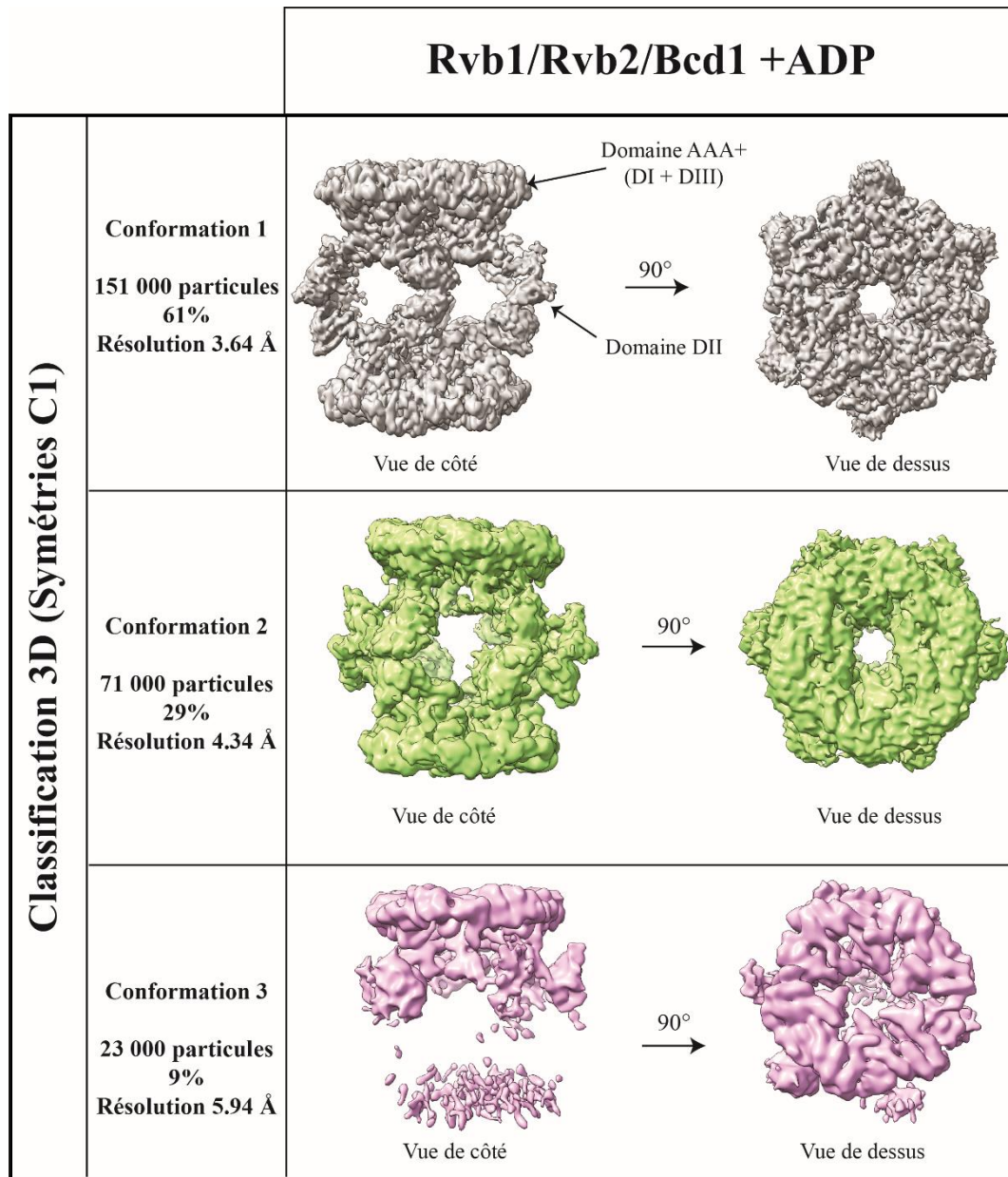


Figure 36 : L'analyse du complexe Rvb1/Rvb2/Bcd1p +ADP par cryo-microscopie électronique révèle une hétérogénéité conformationnelle.

(A) Classification 3D de symétrie C1 de la conformation 1 représentative d'une moyenne de 151 000 particules (61%). (B) Classification 3D de symétrie C1 de la conformation 2 représentative d'une moyenne de 71 000 particules (29%). (C) Classification 3D de symétrie C1 de la conformation 3 représentative d'une moyenne de 23 000 particules (9%).

hexamères, correspondant aux domaines DII qui présente une flexibilité en solution. La deuxième classe 3D correspond à la conformation N°2 qui a une résolution de 4,34 Å et qui représentant 29% des particules analysées. Il s'agit également d'un dodécamère avec deux anneaux hexamériques présentant un certain degré d'inclinaison de l'un par rapport à l'autre. Sa vue de côté au niveau de la Figure 36 montre la présence d'une densité supplémentaire au centre au niveau de l'interface de deux DII. La troisième classe 3D présente une résolution 5,94 Å représentative de 9% des particules. Elle est constituée que d'un seul hexamère avec une augmentation de masse au niveau de certains de ses domaines DII. Il est à noter la présence d'une densité de faible résolution en bas (Figure 36).

Une symétrie D3 a été imposée à la conformation 1 en utilisant comme modèle la structure cryo-EM de Rvb1/Rvb2 dodécamérique de *C. thermophilum* (code PDB 4WW4) en présence d'ADP. La résolution de cette classe 3D a pu ainsi légèrement être améliorée avec 3,18 Å (Figure 37A). Il apparaît très clairement que la conformation de l'ensemble des domaines DII de Rvb1 et Rvb2 est étendue vers le solvant.

L'analyse de la conformation 2 en coupe longitudinale (Figure 37B) montre la présence de deux zones de densité supplémentaire par rapport à la conformation N°1 au niveau de la connexion des domaines DII externes : la première à droite dans l'enveloppe dodécamérique de gauche. La deuxième au centre au fond de l'enveloppe dodécamérique de droite (voir les flèches de la Figure 37B) Ces deux nouvelles densités pourraient correspondre à la présence de la protéine Bcd1p en interaction au sein du complexe Rvb1/Rvb2. L'analyse des images est toujours en cours par notre collaboratrice Célia Plisson-Chastang du Laboratoire de Biologie Moléculaire Eucaryote (LBME) à Toulouse, afin de mieux caractériser ce complexe et de retrouver précisément la ou les position (s) de Bcd1p, ainsi que sa stœchiométrie.

Ces résultats préliminaires de cryo-EM nous permettent d'émettre une hypothèse selon laquelle l'interaction Bcd1p se fait en post-hydrolyse de l'ATP en ADP, et qu'elle entraîne l'hétérododécamérisation du complexe Rvb1/Rvb2 puisque sa conformation quaternaire non liée est hétérohexamérique. Il apparaît des résultats de cryo-EM obtenus à ce jour, que la fixation de Bcd1p sur le complexe Rvb1/Rvb2 conduit à son hétérododécamérisation.

Comme nous l'avons rapporté dans l'introduction, la protéine Nop58 au sein du complexe Rvb1/Rvb2 constituant du R2TP est amenée *via* Bcd1p aux sites d'assemblages des snoRNP à boîtes C/D. Les étapes de déposition de Nop58p sur la pré-snoRNP composée du snoARN C/D, Snu13p, Rsa1p et Hit1p, ne sont toujours pas élucidées. Par conséquent, nous ignorons par quel mécanisme Nop58p est correctement positionnée sur la snoRNP à boîtes C/D, mais aussi avec quelle chronologie les facteurs d'assemblage quittent la particule RNP.

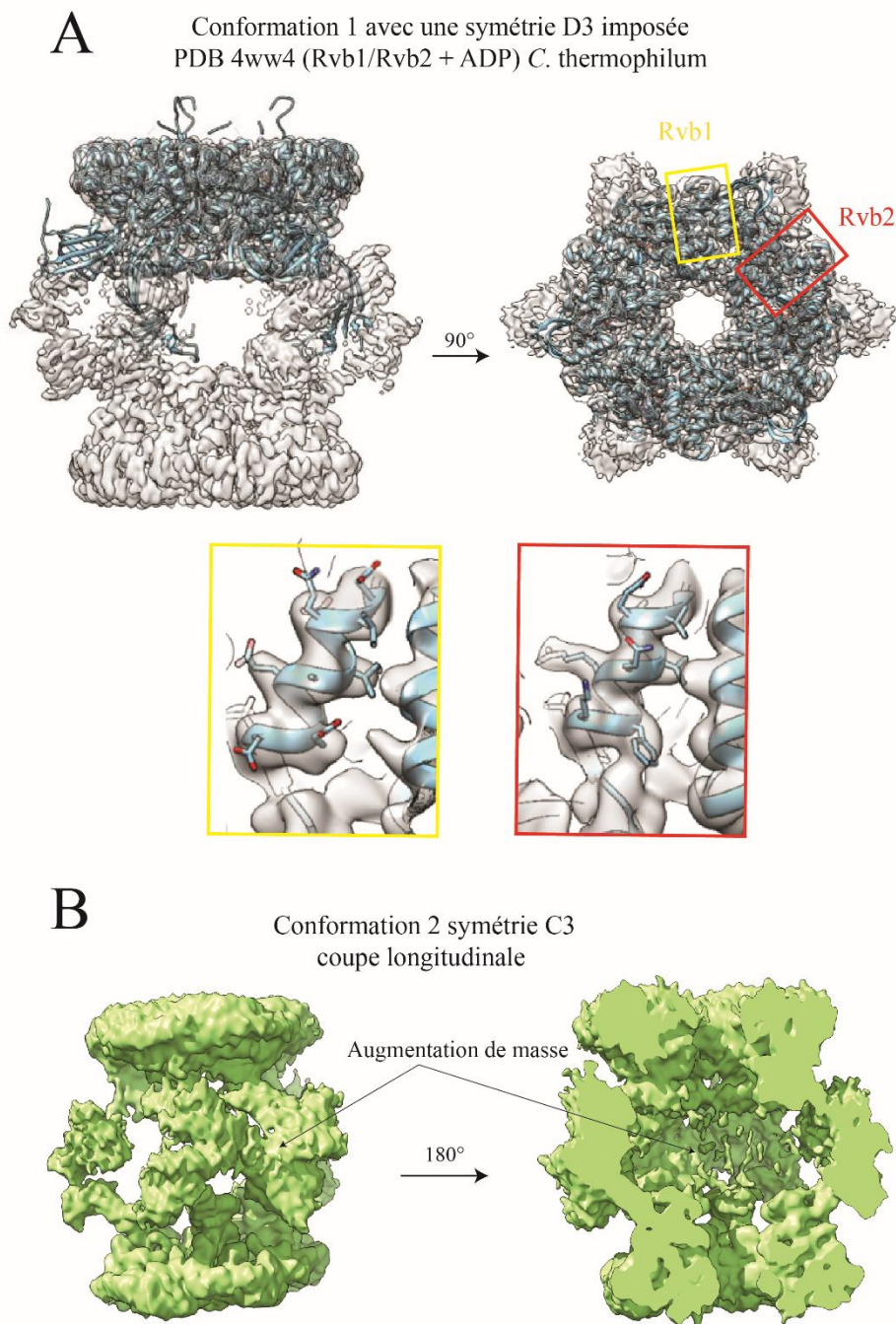


Figure 37 : Affinement et analyse des conformations 1 et 2 obtenues en cryo-microscopie électronique pour le complexe Rvb1/Rvb2/Bcd1p + ADP.

(A) Superposition de la structure PDB 4WW4 du complexe hétérododécamérique Rvb1/Rvb2 en ADP pour *C. thermophilum*, à la conformation 1 obtenue pour le complexe Rvb1/Rvb2/Bcd1p +ADP. La conservation du positionnement des structures secondaires ainsi que les chaînes latérales des acides aminés de Rvb1 et Rvb2 sont indiquées (encadrement en jaune pour Rvb1 et rouge pour Rvb2). (B) La coupe longitudinale au niveau de la conformation 2 a permis de mettre en évidence une augmentation de masse conséquente reliant deux domaines DII ainsi qu’au centre de la structure dodécamérique pouvant correspondre à Bcd1p.

Chapitre 2 : Analyse de la contribution du facteur Bcd1p dans l'assemblage de la particule snoRNP à boîtes C/D

1. L'interaction Bcd1p-Snu13p est essentielle pour l'assemblage des snoRNP C/D

A mon arrivée au laboratoire, notre équipe avait déjà mis en évidence l'importance de la protéine Bcd1 pour la croissance cellulaire des levures (B. Bragantini *et al.*, 2016), mais également sa nécessité pour le maintien d'un taux stable des snoARN à boîtes C/D dans les cellules (Paul *et al.*, 2019). Les précédentes analyses de la protéine Bcd1, à la fois fonctionnelles et structurales ont servi de base afin de poursuivre son étude de façon plus détaillée. Le fragment minimum fonctionnel Bcd1p₁₋₉₆ est suffisant pour rétablir la croissance des levures. Il est composé d'un domaine en doigt de zinc dans sa partie N-terminale (1-45) et d'une région peu structurée 46-96. Cette dernière serait responsable d'interactions essentielles mises en jeu dans la fonction de Bcd1p vis-à-vis de l'assemblage des snoRNP C/D. Dans le but de caractériser les acides aminés clés dans ce processus, nous avons procédé à l'alignement des séquences protéiques de Bcd1p de plusieurs espèces (Figure 38A). Cela nous a permis de mettre en évidence la conservation stricte de deux acides aminés chargés : l'aspartate 72 et l'arginine 81. Cette conservation phylogénétique est potentiellement due à un rôle important dans le réseau d'interaction de Bcd1p.

Afin de déterminer l'éventuelle importance fonctionnelle de ces deux acides aminés chez la levure, une approche par mutagenèse dirigée a été utilisée. Dans un premier temps, nous nous sommes concentrés sur l'étude de l'aspartate 72, dont la mutation en alanine a pu rapidement être obtenue dans un vecteur d'expression constitutif de levure, le p416 GPD. Du fait du caractère essentiel de la protéine Bcd1 pour la viabilité cellulaire, nous avons utilisé une souche de levure BY4741 où le gène qui code pour Bcd1p endogène est sous l'expression d'un promoteur conditionnel inductible au Galactose (pGAL). En présence de Galactose dans le milieu de croissance, cette souche est donc capable de produire la protéine Bcd1p endogène. Cependant, en présence de Glucose comme source de carbone dans le milieu de croissance, cette même souche n'exprime plus Bcd1p endogène. Nous l'avons transformé avec un vecteur p416 BCD1 sauvage (*Wild Type*) ou BCD1 muté D72A, en version pleine taille ou avec le gène correspondant au fragment fonctionnel minimum BCD1-96. L'expression des protéines par ce vecteur est constitutive. Les profils de croissance de ces souches ont été comparés dans deux milieux nutritifs différents : *Yeast Peptone Galactose* (YPG) et *Yeast Peptone Dextrose* (ou encore *Glucose*) (YPD) parallèlement en milieu nutritif solide (boîte de pétri) (Figure 38B) et

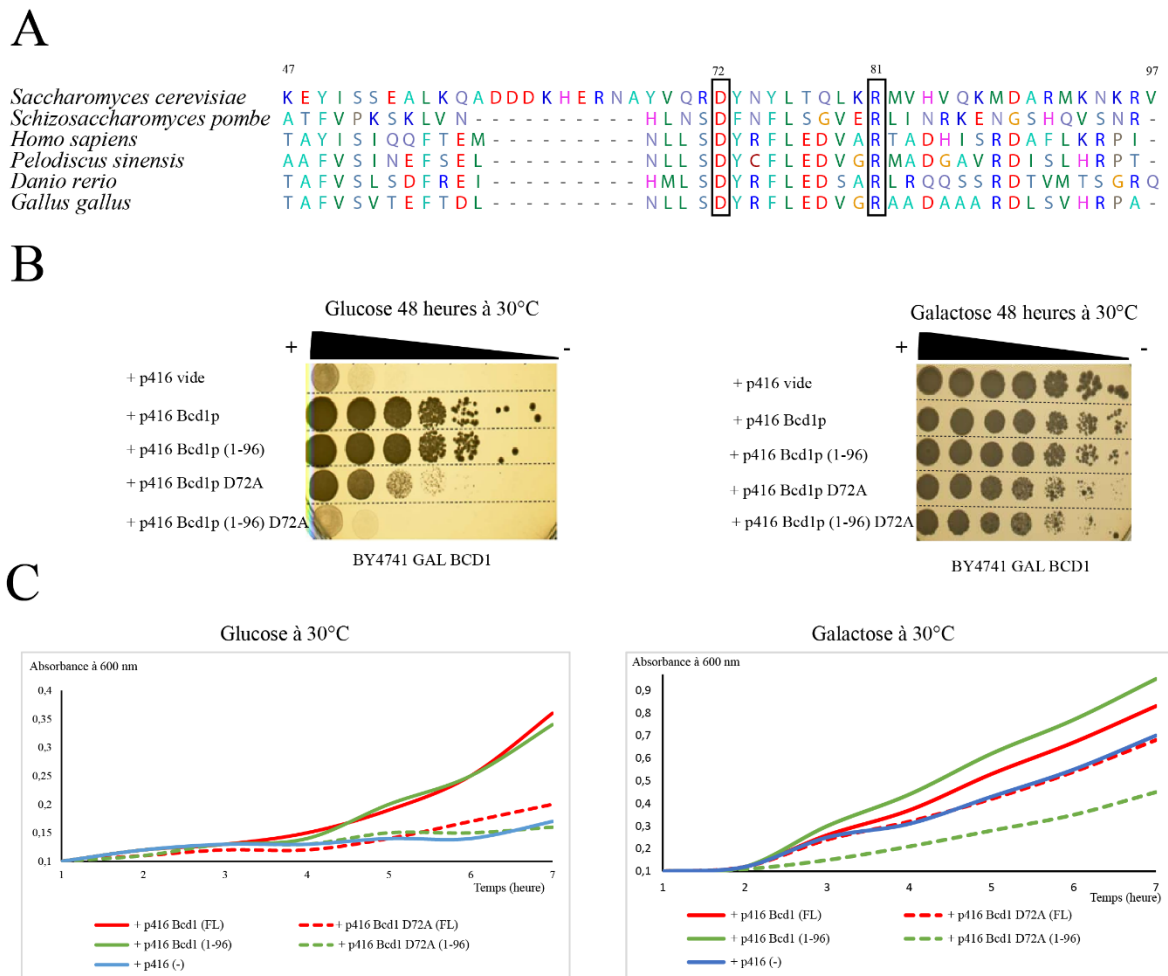


Figure 38 : La mutation D72A sur Bcd1p induit un retard de croissance des levures *S. cerevisiae*.

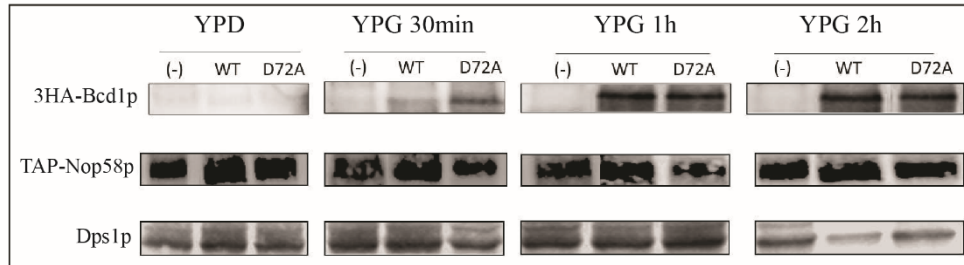
(A) Alignement de plusieurs séquences protéiques de la protéine Bcd1 chez différentes espèces, en utilisant *CLUSTAL W*. Les acides aminés strictement conservés sont encadrés. (B) Résultats de suivi de croissance sur milieu solide (en boîte de pétri), avec deux milieux nutritifs différents, YPG : *Yeast Peptone Galactose* (droite) et YPD : *Yeast Peptone Dextrose* (gauche). Une série de dilution de chaque souche est représentée sur l'axe horizontal du schéma. n=3 (C) Résultats de suivi de croissance en milieu liquide, avec deux milieux nutritifs différents, YPG : *Yeast Peptone Galactose* (droite) et YPD : *Yeast Peptone Dextrose* (gauche). n=4

en milieu liquide (mesure de densité optique à 600 nm chaque heure) (Figure 38C). De manière intéressante, pour les souches exprimant uniquement le mutant Bcd1p D72A en version pleine taille ou fragment 1-96, nous observons un ralentissement de la croissance en glucose comparée aux souches exprimant la protéine Bcd1p sauvage (Figure 38B, 38C). De manière intéressante, nous remarquons également un ralentissement de croissance en milieu nutritif galactose pour la protéine Bcd1 D72A et le fragment Bcd1₁₋₉₆ D72A comparés aux versions sauvages testées. Ce résultat a également été obtenu en parallèle de nos expériences par l'équipe de Koshnevis *et al.*, 2019, où ils constatent un retard de croissance des levures exprimant Bcd1p D72A. Ces résultats suggèrent un effet de dominance négative, qui consiste en une compétition entre Bcd1p endogène sauvage et Bcd1p D72A ectopique à former des complexes d'assemblage fonctionnels (Bcd1p endogène), qui assurent une croissance minimum, et non fonctionnels (Bcd1p D72A) qui sont abortifs et ne permettent pas un assemblage correct de snoRNP à boîtes C/D. Cette hypothèse sera approfondie et testée dans la suite de ce chapitre par la caractérisation des complexes transitoires de l'assemblage dans lesquels le mutant Bcd1p D72A est impliqué.

2. Mise en place au laboratoire pour l'étude des complexes d'assemblage formés par Bcd1p D72A

Comme nous l'avons décrit plus haut, l'expression constitutive et ectopique de Bcd1p D72A ralentie fortement la croissance conduisant à un arrêt de croissance des levures. Nous voulons nous affranchir des effets de la perte de croissance sur le plan moléculaire qui sont caractérisés essentiellement par une dégradation de la protéine cœur Nop58 et une déstabilisation du taux de snoARN à boîtes C/D. Nous avons donc utilisé pour la suite de cette étude un vecteur d'expression de Bcd1p D72A sous un promoteur conditionnel inductible pGAL : pFA6a GAL 3HA BCD1 D72A. Des souches de levures ont été générées, de façon à exprimer à la fois certaines protéines cœur ou facteurs d'assemblages avec une étiquette, et transformées par le plasmide pFA6a GAL 3HA BCD1 D72A permettant une expression inductible du mutant de Bcd1p. Cette expression de Bcd1p D72A a été testée à 30 minutes, 1 heure et 2 heures de culture en milieu galactose. Une analyse par *Western-Blot* de chacune des cultures (Figure 39A) indique que l'expression de la protéine est optimale après 1 heure d'induction YPG. Il est important de noter également que la protéine Nop58 dans ces conditions reste stable dans les cellules.

A



B

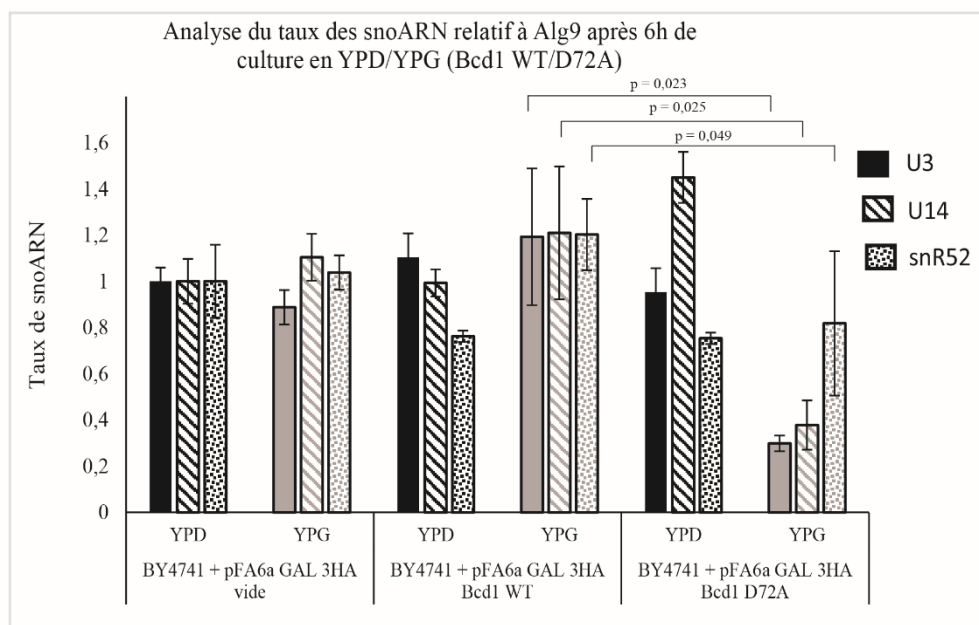


Figure 39 : L'expression de Bcd1p D72A affecte la stabilité des snoARN à boîtes C/D.

(A) Suivi de l'expression de Bcd1p dans la souche BY4741 TAP-NOP58 + pFA6a GAL 3HA BCD1 par *Western Blot* avec des temps différents d'induction de son expression. La stabilité de Nop58p a également été vérifiée en fonction des différents temps d'induction de l'expression de Bcd1p. n=3 (B) Le taux des snoARN à boîtes C/D U3, U14 et snR52 a été analysé par PCR quantitative après une induction de l'expression de Bcd1p sauvage ou mutant pendant 6 heures. La significativité statistique a été évaluée par l'utilisation d'un test t de Student non apparié bilatéral. Les valeurs p représentent les différences significatives entre deux conditions testées (significatif pour $p < 0,05$). n=3

3. L'expression de Bcd1p D72A affecte la stabilité des snoARN à boîtes C/D transcrits par l'ARN polymérase II

Dans le but de déterminer l'effet de l'expression du mutant sur la stabilité des snoARN à boîtes C/D, nous avons induit l'expression de Bcd1p D72A pendant 6 heures en YPG, puis les cellules ont été culotées, broyées et le niveau d'expression de plusieurs snoARN a été analysé par RT-qPCR : U3, U14 et snR52 (Figure 39B). Nous remarquons que l'expression du mutant de Bcd1p affecte drastiquement le taux des snoARN U3 et U14, tandis que le taux de snR52 demeure relativement stable. Il est essentiel de rappeler que U3 et U14 sont des transcrits de l'ARN polymérase II à l'inverse du snR52 qui possède la particularité d'être le seul transcrit de l'ARN polymérase III chez la levure. Il a été montré au laboratoire par A. Paul que l'association de Bcd1p avec les sous-unités de l'ARN polymérase III existe. Cette association nécessite particulièrement l'extrémité C-terminale de Bcd1p₁₁₆₋₃₆₆. La mutation D72A sur Bcd1p étant dans son extrémité N-terminale, elle ne perturbe donc pas son association avec l'ARN polymérase III, ce qui suggère que le recrutement de Bcd1p D72A puisse se faire de façon efficace sur les pré-snoRNP snR52 en cours de transcription.

4. Bcd1p D72A bloque l'assemblage en amont du recrutement de Nop58p sur la pré-snoRNP à boîtes C/D

Afin de caractériser le complexe transitoire d'assemblage bloqué par le mutant Bcd1p D72A, nous avons voulu savoir si ce dernier se trouve en association avec le snoARN en cours d'assemblage, mais aussi déterminer si le transfert de la protéine Nop58 sur le snoARN, qui est une étape clé de l'assemblage, est réalisé.

Pour cela, j'ai réalisé des expériences de RIP (RNA *ImmunoPrécipitation*) sur Bcd1p D72A et la protéine cœur Nop58. Cela consiste à immunoprécipiter les protéines par des étiquettes TAP-Nop58p et 3HA-Bcd1p D72A et ensuite d'analyser les snoARN à boîtes C/D associés par RT-qPCR. Pour cette étude, nous avons choisi d'analyser le taux des snoARN U3 matures et du snR32 et du snR36 (un snoARN H/ACA), ainsi que deux formes immatures de U3 : U3_{ext3'} et U3_{intron} identifiées au laboratoire comme étant en association avec Bcd1p (Paul *et al.*, 2019).

L'analyse des snoARN C/D après une RIP HA-Bcd1p (Figure 40A) montre que la substitution D72A entraîne une perte d'association significative de Bcd1p avec le snoARN U3 mature. Ce résultat avec le résultat de *pull-down* de Khoshnevis *et al.*, 2019, suggèrent que le recrutement de Bcd1p sur la particule RNP en cours d'assemblage, passe par l'interaction directe Snu13p-

Bcd1p. Par ailleurs, l'association du mutant Bcd1p D72A avec le snoARN snR52 ne semble pas être affectée de façon aussi importante que son association avec le transcrite U3. Le mécanisme de recrutement de Bcd1p sur ce type de snoARN (transcrit par l'ARN polymérase III) lors de l'assemblage de la snoRNP serait donc différent de ceux pour les snoARN transcrits par l'ARN polymérase II.

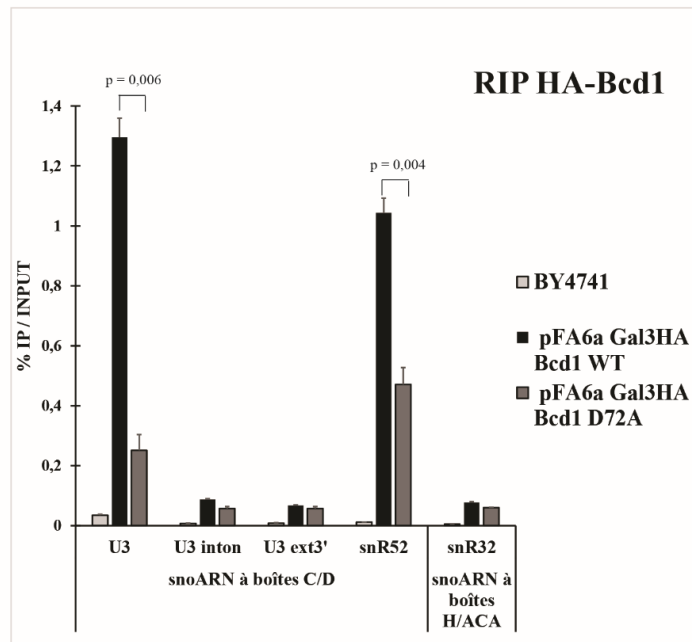
Des études menées au laboratoire nous ont permis d'attribuer à la partie C-terminale de Bcd1p, un rôle potentiel dans la rétrotransposition. Bcd1p a été identifiée en association avec des domaines de la polymérase III par analyse protéomique de ses partenaires *in vivo* (Paul, 2018). D'autre part, la protéine cœur Nop58 perd également son association avec les snoARN U3 de façon prononcée dans le cadre du mutant D72A de Bcd1p (Figure 40B), contrairement au snR52 dont l'association avec Nop58p est maintenue. Il est suggéré que l'expression de Bcd1p D72A perturbe la déposition de Nop58p sur la particule en cours d'assemblage, ce qui bloque la suite du processus et altère la formation des snoRNP C/D matures et fonctionnelles issues de transcrits Pol II, essentielles à la croissance des levures.

Des expériences menées au laboratoire ont permis de constater que l'absence de Bcd1p affecte drastiquement le recrutement de Nop58p sur le snoARN en cours d'assemblage (Paul *et al.*, 2019). Le mutant Bcd1p D72A bloque ainsi l'assemblage de la même façon et mimerait de façon plus ciblée une délétion de Bcd1p entière. Nous supposons donc que Nop58p et Bcd1p D72A se retrouvent dans des complexes de pré-assemblage uniquement protéiques.

5. L'association de Bcd1p avec Snu13p est perturbée par la mutation D72A

La protéine cœur Snu13 s'associe de façon précoce avec le snoARN lors de l'assemblage des snoRNP C/D. Elle est également retrouvée en association avec Bcd1p. Par ailleurs, nous avons constaté au laboratoire que l'absence de la protéine Bcd1p ne perturbe pas l'association de Snu13p avec le snoARN. Nous avons voulu voir si la mutation D72A sur Bcd1p affecte son association à Snu13p, pour cela nous avons exprimé le mutant dans une souche de levure exprimant la protéine Snu13 étiquetée par le biais d'un vecteur d'expression (p416 GPD) (Flag-Snu13p), que nous avons utilisée pour l'immunoprécipiter et analyser si la protéine Bcd1p D72A était associée ou non. D'après l'analyse *Western-Blot* des protéines d'intérêt (Figure 41), nous remarquons que l'association de Bcd1p avec Snu13p est perdue dans le cas de sa version mutante D72A. Comme cela a pu être montré *in vitro* par *GST-pull down* dans l'étude du mutant D72A de Bcd1p (Khoshnevis *et al.*, 2019), l'association testée entre Snu13p et Bcd1p D72A est déstabilisée.

A



B

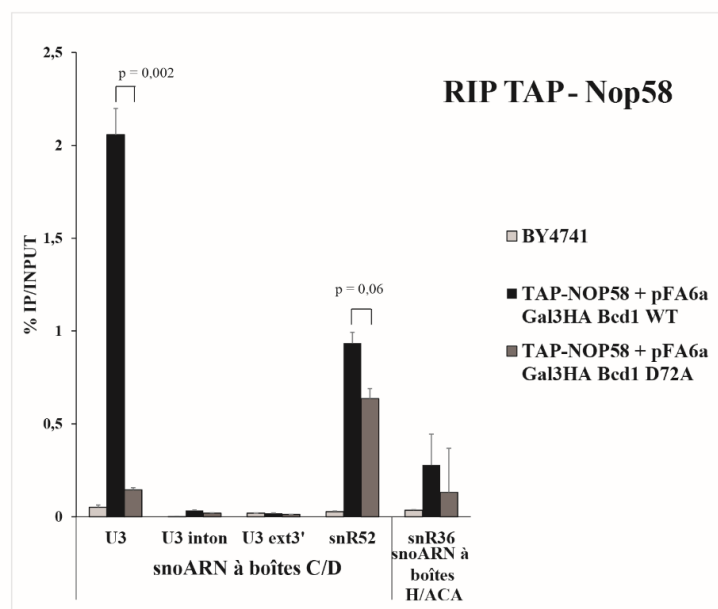


Figure 40 : Le mutant Bcd1p D72A bloque la déposition de Nop58p sur le snoARN.

(A) Analyse par RIP de l'association de Bcd1p avec les snoARN à boîtes C/D U3 mature et ses formes immatures (U3ext3' et U3 intron), le snR52 et avec le snoARN à boîtes H/ACA snR32 (contrôle négatif). L'enrichissement de Bcd1p sur ces snoARN a été déterminé par PCR quantitative à la suite d'expériences de RIP en utilisant la souche BY4741+ pFA6a GAL 3HA BCD1 WT/D72A, exprimant de façon stable la protéine Bcd1p sauvage ou mutée pendant 2 heures. (B) Analyse par RIP de l'association de Nop58p avec le snoARN à boîtes C/D U3 mature et ses formes immatures (U3ext3' et U3 intron), snR52 et avec le snoARN à boîtes H/ACA snR36 (contrôle négatif). L'enrichissement de Bcd1p sur ces snoARN a été analysé par PCR quantitative à la suite d'expériences de RIP en utilisant la souche BY4741 TAP-NOP58 + pFA6a GAL 3HA BCD1 WT/D72A, exprimant de façon stable la protéine Bcd1p sauvage ou mutée pendant 2 heures et la protéine Nop58 étiquetée TAP. La significativité statistique a été évaluée par l'utilisation d'un test t de Student non apparié bilatéral. Les valeurs p représentent les différences significatives entre deux conditions testées (significatif pour $p < 0,05$). n=3

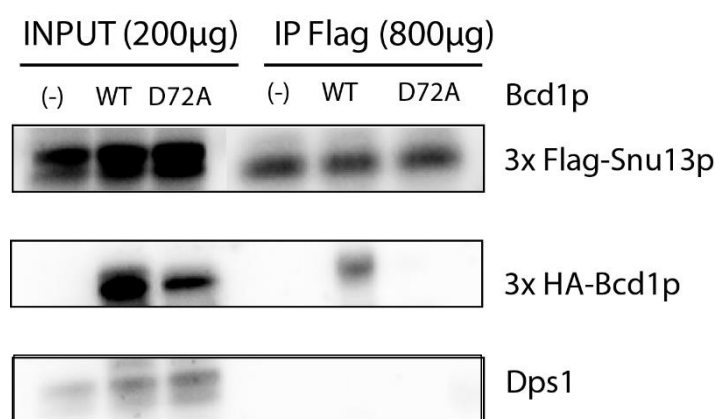


Figure 41 : L'association Bcd1p/Snu13p est déstabilisée par la mutation D72A sur Bcd1p.

Test de co-immunoprécipitation de Snu13p via son étiquette FLAG avec la protéine Bcd1p sauvage vs sa version mutante D72A identifiées par l'étiquette HA. Dps1 est une protéine qui sert de contrôle de dépôt. n=2

Cela signifie que cette interaction est potentiellement importante pour l'assemblage des snoRNP à boîtes C/D fonctionnelles. Il est probable que la déposition de Nop58p sur le snoARN est conditionnée au préalable par l'interaction Bcd1p/Snu13p.

Chapitre 3 : Etudes structurales des domaines TPR1 et TPR3 de SPAG1 et caractérisation de leurs modes d'interaction avec HSP70 et HSP90

Le complexe R2TP joue un rôle majeur dans l'assemblage des snoRNP à boîtes C/D. Les travaux de l'équipe d'E. Bertrand (IGH, Montpellier) ont permis de mettre en évidence l'existence, dans des cellules humaines, d'un complexe homologue au R2TP, ou « R2TP-like » dénommé R2SP. Ce complexe fait l'objet d'études structurales approfondies au sein du laboratoire.

Le R2SP semble présenter également un rôle de cofacteur de protéine chaperon, à travers le recrutement des protéines chaperon HSP70 et HSP90 humaines. Comme pour la protéine RPAP3, la protéine SPAG1, *via* ses trois domaines TPR en position N-terminale, est responsable du recrutement de soit HSP70 soit HSP90 au sein du complexe R2SP. En effet, des essais de production et de purification de chacun des trois domaines TPR de SPAG1 ont été réalisées avant mon arrivée au laboratoire. Alors que le domaine TPR2 est resté insoluble, les domaines TPR1 et TPR3 ont pu être obtenus en quantité suffisante pour envisager leurs études biophysiques (méthodes de dosages biochimiques et ITC) et structurales (RMN et simulations par dynamique moléculaire) *in vitro*. La caractérisation structurale du complexe R2SP permettrait de mieux comprendre le complexe R2TP essentiel à l'assemblage des snoRNP à boîtes C/D. Nous avons étudié au laboratoire le rôle des domaines TPR de SPAG1 dans le recrutement des protéines chaperon HSP par des méthodes de dosages biochimiques, ITC, spectroscopie RMN et de simulations par dynamique moléculaire.

1. SPAG1 est une plateforme au recrutement des HSP

L'alignement des séquences des protéines RPAP3 et SPAG1 humaines montre une similarité et une identité de séquence de 34% et 24% respectivement. Elles sont toutes les deux constituées de domaines TPR et d'un domaine C-terminal RPAP3-C. La protéine SPAG1 possède un domaine TPR supplémentaire (TPR2), ce qui rend SPAG1 30% plus longue que RPAP3. L'alignement suggère que le TPR1 et le TPR2 de RPAP3 sont homologues aux TPR1 et TPR3

de SPAG1. L'importance fonctionnelle de ces deux domaines TPR est justifiée par la présence de la majorité des mutations impliquées dans la PCD (*Primary Ciliary Dyskinesia*) au niveau du TPR1 et TPR3 de SPAG1. Les trois domaines TPR de SPAG1 possèdent tous un motif dit *carboxylate clamp* caractéristique des domaines TPR. Il est responsable de l'interface d'interaction avec la séquence C-terminale EEVD des HSP.

La résolution de la structure tridimensionnelle par RMN du TPR1 et TPR3 de SPAG1 indique la présence d'une partie concave qui héberge les cinq résidus de la *carboxylate clamp* (K213, N217, N247, K277 et R281 du TPR1, K627, N631, N662, K692 et R696 du TPR3).

L'affinité des domaines TPR1 et TPR3 pour HSP90 semble meilleure que celle mesurée pour HSP70. Le domaine TPR2 de SPAG1 est proposé dans cette étude comme jouant un rôle organisationnel de la protéine pleine taille SPAG1. Cela est dû à ses propriétés biochimiques d'insolubilité lorsqu'il est produit tout seul. Il ne participerait donc pas au recrutement de HSP90 et HSP70 qui sont capables de prendre en charge et délivrer des protéines clientes dans des voies d'assemblage.

L'étude de l'interface d'interaction du TPR1 de SPAG1 à l'échelle atomique n'a pas été possible en raison de la formation d'agrégats à une concentration supérieure à 150 μM nécessaire pour une analyse structurale. Pour contrecarrer cela, nous avons optimisé sa séquence par des mutations ciblées sur le domaine TPR permettant sa stabilité à des concentrations élevées. TPR1_{wt} et TPR1_{opt} sont similaires dans leurs propriétés thermodynamiques à interagir avec HSP70 et HSP90.

Ces études nous ont permis de définir les motifs exacts dans les séquences TPR responsables du recrutement des HSP, qui sont capables de prendre en charge et délivrer des protéines clientes dans des voies d'assemblage. L'existence de ces deux modes de reconnaissance (HSP90 et HSP70) pourrait prévenir le blocage total de la fonction de SPAG1 par un inhibiteur spécifique de l'un ou de l'autre type de reconnaissance des HSP.

L'interaction de la protéine pleine taille HSP70 avec SPAG1 TPR1_{opt} a également été étudiée en BLI (*BioLayer Interferometry*) au laboratoire AFMB (Architecture et Fonction des Macromolécules Biologiques) à Marseille. Nous avons bien confirmé l'interaction (données non présentées), cependant la mesure de la constante de dissociation entre HSP70 et SPAG1 TPR1_{opt} n'a pas été possible par cette technique.

2. La protéine SPAG1 possède-elle une activité GTPase ?

Afin de mieux caractériser la fonction d'hydrolyse de GTP attribuée à SPAG1, nous nous sommes intéressés à l'interaction du GTP avec SPAG1 que nous avons étudiée par titration

suivie par spectroscopie RMN. La comparaison des spectre HSQC $^1\text{H}/^{15}\text{N}$ de SPAG1 (622-926) (TPR3 + domaine C-terminal) libre et après l'ajout de 10 molaire de GTP, ne révèle aucun changement des déplacements chimiques ni d'intensité des pics $^1\text{H}/^{15}\text{N}$ de SPAG1. Cela indique qu'aucune interaction n'est détectée entre SPAG1 (622-926) et le GTP. Pour confirmer ce résultat nous avons mesuré par spectroscopie RMN spécifiquement la libération de phosphate inorganique par SPAG1(622-926) après hydrolyse du GTP marqué au ^{31}P . Après une incubation de 24 heures avec le GTP, aucun pic correspondant au ^{31}P n'apparaît entre 0 et 5 ppm sur le spectre. Cela conforte le précédent résultat, à savoir que l'hydrolyse du GTP par SPAG1 (622-926) n'est pas observée dans ces conditions.

Publications :

Research Article

Binding properties of the quaternary assembly protein SPAG1

Marie-Eve Chagot¹, Raphael Dos Santos Morais^{1,*}, Sana Dermouche¹, Dorian Lefebvre², Xavier Manival¹, Christophe Chipot^{3,4}, François Dehez^{3,4} and  Marc Quinternet²

¹Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France; ²Université de Lorraine, CNRS, INSERM, IBSLor, F-54000 Nancy, France; ³LPCT, UMR 7019, Université de Lorraine, CNRS, Vandoeuvre-lès-Nancy, France; ⁴Laboratoire International Associé CNRS and University of Illinois at Urbana-Champaign, Vandoeuvre-lès-Nancy, France

Correspondence: François Dehez (francois.dehez@univ-lorraine.fr) or Marc Quinternet (marc.quinternet@univ-lorraine.fr)

In cells, many constituents are able to assemble resulting in large macromolecular machineries possessing very specific biological and physiological functions, e.g. ribosome, spliceosome and proteasome. Assembly of such entities is commonly mediated by transient protein factors. SPAG1 is a multidomain protein, known to participate in the assembly of both the inner and outer dynein arms. These arms are required for the function of sensitive and motile cells. Together with RUVBL1, RUVBL2 and PIH1D2, SPAG1 is a key element of R2SP, a protein complex assisting the quaternary assembly of specific protein clients in a tissue-specific manner and associating with heat shock proteins (HSPs) and regulators. In this study, we have investigated the role of TPR domains of SPAG1 in the recruitment of HSP chaperones by combining biochemical assays, ITC, NMR spectroscopy and molecular dynamics (MD) simulations. First, we propose that only two, out of the three TPR domains, are able to recruit the protein chaperones HSP70 and HSP90. We then focused on one of these TPR domains and elucidated its 3D structure using NMR spectroscopy. Relying on an NMR-driven docking approach and MD simulations, we deciphered its binding interface with the C-terminal tails of both HSP70 and HSP90. Finally, we addressed the biological function of SPAG1 and specifically demonstrated that a SPAG1 sub-fragment, containing a putative P-loop motif, cannot efficiently bind and hydrolyze GTP *in vitro*. Our data challenge the interpretation of SPAG1 possessing GTPase activity. We propose instead that SPAG1 regulates nucleotide hydrolysis activity of the HSP and RUVBL1/2 partners.

Introduction

Cilia are essential elements of motile and sensitive cells. In human, cilia are crucial for the function of specific tissues, including testis, kidney or lung and their dysfunction is responsible for various pathologies, chief among which are infertility, polycystic kidney disease and primary ciliary dyskinesia (PCD) [1,2]. Cilia are operated by means of a molecular motor composed of both outer and inner dynein arms combining several dozens of proteins. Assembly of these arms requires many protein factors called DNAAFs (dynein axonemal assembly factors) [3], which are not components of the mature arm. SPAG1 belongs to this family of proteins. Several mutations in the human SPAG1 gene were shown to induce PCD [4]. To be specific, these mutations can lead to frameshift in the gene, deletion in the gene, nonsense codons, missense codons, or loss of the start codon. Interestingly enough, the R2SP complex, which combines SPAG1, PIH1D2, RUVBL1 and RUVBL2 proteins, was shown to be involved in the assembly of different protein complexes [5]. In addition, RUVBL1, absent from mature dynein arms, was proven pivotal to maintain the integrity of cilia [6].

The R2SP complex is close to the R2TP complex, with RPAP3 and PIH1D1 adaptors being replaced by SPAG1 and PIH1D2, respectively [5,7]. Both R2SP and R2TP were proposed to work in a concerted fashion with protein chaperones HSP70 and HSP90, thereby facilitating quaternary structure

*Present address: Université de Lorraine, LIBio, F-54000 Nancy, France

Received: 12 March 2019

Revised: 9 May 2019

Accepted: 22 May 2019

Accepted Manuscript online:
22 May 2019

Version of Record published:
14 June 2019

organization of client proteins [8]. It is worth noting that R2TP, together with the RPAP3 interactor WDR92 (also known as MONAD), participates in dynein assembly, but at a stage distinct from SPAG1, thus, suggesting the involvement of HSP (heat shock protein) chaperones in this biological process [9,10]. Despite similarities, R2SP and R2TP target their own specific protein clients. In addition, expression of R2SP was shown to be tissue-specific with a peculiar enrichment in testis observed for SPAG1 [5].

Although the structure of the R2TP complex and its sub-complexes has been extensively studied by NMR (nuclear magnetic resonance), X-ray and cryo-EM, high-resolution three-dimensional data for the R2SP complex are still lacking [11–16]. SPAG1/RPAP3 and PIH1D1/PIH1D2 proteins have been shown to share common features. Whereas PIH1D1 and PIH1D2 are both composed of one PIH and one CS domain, SPAG1 and RPAP3 both possess one conserved C-terminal domain, as well as several TPR domains [5]. TPR domains (for tetratricopeptide repeat) are repeats of at least two 34-residue motifs arranged in a helix-turn-helix fashion [17]. Interestingly enough, a multi-TPR protein, such as HOP, organizes the delivery of clients from HSP70 to HSP90 via simultaneous anchoring of the chaperones to dedicated TPR domains [18,19]. Although HSP70 and HSP90 are structurally very different, this anchoring is mainly driven by a common motif made of residues EEVD and located at the C-terminal end of the chaperones [20]. Similar functions have been deciphered for RPAP3 [14,21]. More precisely, we have demonstrated that RPAP3-TPR1 is likely to recruit HSP70, whereas RPAP3-TPR2 is responsible for anchoring HSP90 to a large complex involving PIH1D1. The CS domain of the latter protein strongly binds a 39-residue sequence specific to isoform 1 of RPAP3, and its PIH domain is recognized by RPAP3-TPR1. Thus, due to their organizational similarities, RPAP3:PIH1D1 and SPAG1:PIH1D2 could share common functions. However, SPAG1 has been proposed to possess a specific GTP hydrolysis activity, possibly providing the energy required for its co-chaperone function [22]. Moreover, SPAG1 seems also able to activate G-protein-mediated extracellular signal-regulated kinases (ERK1/2) by associating with protein G β 1 [23], and other experimental evidence has highlighted its role in nucleotide-dependent pathways, such as the AMP-dependent kinase, mitogen-activated protein kinase and protein kinase C signaling pathways [24]. A P-loop motif (i.e. phosphate-binding loop, also known as Walker A motif), capable of binding specifically the phosphate moiety of nucleotides, and the consensus sequence of which is (G/A)-(X)₄-GK-(T/S), was identified in the region between the last TPR domain and the C-terminal domain (from residue 781 to residue 788) [22]. It ought to be noted that the presence of this motif, not found in RPAP3, was also established in RUVBL proteins [25]. Indeed, RUVBLs can directly bind and hydrolyze nucleotides, inducing conformational changes that could putatively modify the specificity and/or the affinity of their protein clients [26].

In this work, we deciphered the binding features of SPAG1 toward the protein chaperones HSP70 and HSP90, combining biochemical, biophysical and *in silico* approaches. We specifically focused on the third TPR domain of SPAG1, exhibiting a high stability *in vitro*. Using NMR and MD (molecular dynamics) simulations, we revealed at the molecular level the structure and dynamics of SPAG1 in complex with the tail of HSPs. We also investigated the ability of SPAG1 to bind and hydrolyze GTP. With this study, we bring to the fore new data, key for a better understanding of the involvement of HSP chaperones in dynein assembly.

Experimental

Cloning

DNA sequences of human HSP70, HSP90 and HSP90-MC as well as DNA sequences of the full-length human SPAG1 and of sub-fragments 206–327, 446–570, 440–573, 622–742 and 622–926 were cloned into pnEA-3CH, pnYK and pCARGHO vectors [27,28]. The pnEA and pCARGHO vectors introduce, respectively, a 6x-His-tag and a CRD-tag at the N-terminal position of the recombinant proteins. The pnYK vector produces native recombinant proteins.

Protein co-expression assays

For each co-expression test, a Ca²⁺-competent *Escherichia coli* BL21(DE3) pRARE2 strain was co-transformed with plasmids encoding his-tagged (pnEA) and native (pnYK) proteins [27]. The positive co-transformations were selected on LB-agar medium supplemented with antibiotics. One single colony was used for IPTG-induced overexpression of recombinant proteins in a selective LB medium, overnight at 20°C. The cell pellet was then sonicated in lysis buffer [25 mM HEPES (pH 7.5), 300 mM NaCl, 0.5 mM TCEP, 10 mM imidazole]. The cell lysate was placed in contact with TALON beads (Clontech) and left stirring for 1 h at 4°C.

Beads were washed three times with the lysis buffer and proteins retained on the beads were eluted using the lysis buffer supplemented with 300 mM imidazole. Samples were analyzed by SDS–polyacrylamide gels (10% SDS-PAGE). Control experiments were also performed with native proteins purified in the absence of his-tagged partners.

Purification of recombinant proteins

Isolated recombinant proteins were overexpressed in *E. coli* BL21(DE3) pRARE2 grown in a selective LB medium or in a selective M9 medium supplemented with 0.5 g/L $^{15}\text{NH}_4\text{Cl}$ and 2.0 g/L ^{13}C -d6-glucose for NMR samples. Briefly, bacteria were transformed with the plasmid of interest, selected on an LB agar medium and placed in pre-culture in a selective LB medium at 37°C. One liter of LB or NMR growth medium was seeded with the pre-culture. When $\text{OD}_{600\text{nm}}$ reached ~ 0.6 , overexpression of the protein was induced with 0.2 mM IPTG and cell cultures were placed at 20°C for one night. Cells were harvested and sonicated in lysis buffer [25 mM HEPES (pH 7.5), 300 mM NaCl, 0.5 mM TCEP, 10 mM Imidazole]. The tagged proteins were retained on TALON (or lactose sepharose) beads that were washed three times with the lysis buffer. Elution from the beads was performed using the 3C-protease or with lysis buffer supplemented with 300 mM imidazole or D-lactose whether the 6x-His or CRD tags were kept or not. Finally, a size exclusion chromatography performed in 10 mM NaPi (pH 6.4), 150 mM NaCl, 0.5 mM TCEP on Superdex-75 or Superdex-200 columns permitted to isolate the protein of interest. The purification process was followed on SDS-PAGE (12.5%). When necessary, the final buffer was exchanged using several dilution/concentration cycles in Amicon Ultra Centrifugal Filters (Merck Millipore).

Dynamic light scattering

DLS experiments were performed at 20°C on a Zetasizer Nano ZS (Malvern Panalytical) in low-volume quartz cell. Data were processed using the Zetasizer software (Malvern) with default parameters.

Size exclusion chromatography coupled to static light scattering

Size exclusion chromatography coupled to static light scattering experiment (SEC-SLS) was performed at 30°C with an AKTA Explorer HPLC system (GE Healthcare) coupled with both RALS/LALS light scattering and refractive index detectors (Viscotek TDA302, Malvern Panalytical). The refractive index increment value (dn/dc) used to determine the molecular weight was 0.185 mL/g. The SEC column (Shodex KW-804, 7 μm , 8 mm ID \times 300 mm, void volume ~ 6 mL, total volume ~ 12.5 mL) was equilibrated with phosphate-buffered saline (pH 7.4, PBS, Gibco, Thermo Fischer Scientific). The flow rate was 0.5 mL/min. Data were processed using the OmniSEC software (v5.12, Malvern Panalytical).

Isothermal titration calorimetry

The binding of synthetic peptides DASRMEEVD (C-terminal fragment 724–732 of human HSP90) and SGPTIEEVD (C-terminal fragment 638–646 of human HSP70) to the full-length SPAG1 or sub-fragments was measured at 288 K with an ITC-200 microcalorimeter (Malvern) in 10 mM NaPi (pH 6.4), 150 mM NaCl, 0.5 mM TCEP. Peptide and protein concentrations, around 1.5 mM and 0.15 mM, respectively, were calculated using absorbance at 205 and 280 nm [29]. Peptides and proteins were dispatched into the syringe and cell, respectively. Calorimetric data were treated with ORIGIN7 to derive the dissociation constant (K_d), the stoichiometry (N), and the variations in enthalpy (ΔH), entropy (ΔS) and free energy (ΔG).

Protein–protein cross-link assays

His-tagged HSP70 and HSP90-MC proteins concentrated at 30 μM were incubated for 1 h at room temperature with SPAG1-TPRs 1 and 3 concentrated at 120 μM in 10 mM NaPi (pH 6.4), 150 mM NaCl supplemented with 20 mM EDC ((1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride, Sigma Aldrich). The cross-linking reaction was stopped by addition of LDS buffer (Expedeon). Controls without EDC were performed. Samples were analyzed on precast SDS-PAGE Gradient 4–12% (Expedeon).

HSP binding monitored by NMR

^1H - ^{13}C METHYL-SOFAST-HMQC were recorded on ^{13}C -labelled SPAG1-TPRs 1 and 3 concentrated at 30 μM in 10 mM NaPi (pH 6.4), 150 mM NaCl, 0.5 mM TCEP, 5% D_2O in the absence or in the presence of

a 3 molar excess of non-labeled his-tagged HSP70 or HSP90-MC. Spectra were collected at 293 K on a 600 MHz Avance III spectrometer equipped with a TCI cryoprobe.

NMR analysis of SPAG1-TPR3

An almost complete resonance assignment (>91%) of the free $^{13}\text{C}/^{15}\text{N}$ -labelled SPAG1-TPR3 was obtained using a 1 mM sample in 10 mM NaPi (pH 6.4), 150 mM NaCl, 0.5 mM TCEP, 5% D_2O . Classical 2D and 3D spectra were recorded at 293 K on a 600 MHz Avance III spectrometer equipped with a TCI cryoprobe. 3D structures were derived from ^1H - ^{15}N and ^1H - ^{13}C NOESYHSQC and ^1H - ^1H NOESY experiments recorded with a mixing time of 120 ms and from dihedral restraints coming from TALOS-N and PREDITOR [30,31]. CYANA3.0 was used for NOE assignment [32]. Finally, 100 structures were refined in water using RECOORD scripts [33]. The 20 structures with the lowest overall energies were selected as a representative ensemble of structures.

The NMR assignments of the HSP-bound forms of $^{13}\text{C}/^{15}\text{N}$ -labelled SPAG1-TPR3 were obtained using classical heteronuclear 3D NMR spectra recorded on samples containing 1 mM protein and 3 mM peptides. Buffer and temperature conditions were the ones described above. 2D- and 3D-filtered NOESY experiments collected with a mixing time of 120 ms were recorded to select the ^1H signals coming from the peptide before transferring them either to the protein or to the peptide signals. Filtered TOCSY experiments collected with a mixing time of 20 and 70 ms were also recorded to obtain the NMR assignment of the bound peptide.

Chemical shift perturbations (CSPs) of backbone amide groups in SPAG1-TPR3 upon HSP peptides binding were determined by comparing ^1H - ^{15}N HSQC spectra of the free and bound forms of the ^{15}N -labelled protein.

For this purpose, composite ^1H - ^{15}N chemical shifts were calculated as follows: $\Delta\delta = \sqrt{0.1(\delta^{15}\text{N})^2 + (\delta^1\text{H})^2}$.

HADDOCK modeling

HADDOCK2.2 and CNS1.3 were used to determine the NMR-driven structure of the SPAG1-TPR3 bound to SGTPIEVD and DASRMEEVD peptides [34]. The best NMR structure of SPAG1-TPR3 and the X-ray structures of the HSP70 and HSP90 peptides enclosed, respectively, in PDB entries 3Q49 and 4I2Z were used as starting structures [35,36]. Filtered NOE cross-peaks recorded on SPAG1-TPR3:peptide complexes were converted into distance restraints and injected in the docking procedure. Automated mode was chosen for the definition of the semi-flexible interfaces. 512 and 128 structures were calculated in the rigid-body and the semi-flexible simulated annealing steps, respectively. For the last 128 structures, a final flexible refinement in water was performed. Clustering of the docking solutions was performed using the FCC (Fraction of Common Contacts) algorithm with a cutoff of 0.75 [37]. The 3D structures were ranked according to their HADDOCK score.

Molecular dynamics simulation

Molecular assays were built starting from the best Haddock predicted structures subsequently solvated in a cubic box (edge 80 Å) by ~15 300 water molecules in the presence of 150 mM NaCl. Ions, protein and the three peptides were described using the CHARMM36 force field [38] including CMAP corrections [39] whereas water was represented by the TIP3P model [40]. All MD trajectories were generated with the NAMD package (version 2.13) [41]. The equation of motions was integrated by means of a multiple time step algorithm [42] using a time step of 2 and 4 fs for short- and long-range interactions, respectively. Covalent bonds involving hydrogen atoms were constrained to their equilibrium values by means of the Rattle algorithm [43]. All trajectories were carried out in the isobaric-isothermal ensemble at 300 K under 1 atm using Langevin dynamics [44] (damping coefficient 1 ps^{-1}) and the Langevin piston method [45]. Long-range electrostatic interactions were taken into account by means of the Particle Mesh Ewald (PME) algorithm [46]. Water and ions were first thermalized during 5 ns while restraining harmonically the heavy atoms of the complexes to their initial positions. Next, all positional restraints were removed. The distance between the protein residue I724 and the peptide residues Met728/Ile642 was enforced to evolve at values equal or lower to that observed by NMR using the 'colvar' module of NAMD [47]. The molecular assays were further thermalized in this condition for 50 ns. Finally, MD trajectories of 550 ns were generated without any restraints and analyzed using VMD [48].

GTP binding and hydrolysis monitored by NMR

^1H - ^{15}N BEST-TROSY-HSQC and HSQC spectra were recorded on the ^{15}N -labelled fragment 622–926 of SPAG1 concentrated at 0.1 mM in the presence or absence of a 10, 20 and 30 molar excess of GTP using a

600 MHz Avance III spectrometer equipped with a TCI cryoprobe. Data were collected at 293 K in 10 mM NaPi (pH 6.4), 150 mM NaCl, 0.5 mM TCEP, 5% D₂O supplemented or not with 5 mM MgCl₂. Affinity value was derived from CSP using an in-house python script.

GTPase activity was followed using decoupled 1D ³¹P NMR spectra recorded at 298 K in 10 mM HEPES (pH 7.5), 150 mM NaCl, 5% D₂O, 5 mM MgCl₂, 1 mM GTP supplemented or not with 0.1 mM ¹⁵N-labelled SPAG1(622–926). Samples were incubated for 24 h at 298 K before data acquisition. Spectra were recorded on a 400 MHz Avance III spectrometer equipped with a BBFO probe.

Results

SPAG1 and RPAP3 display similar structural organization

SPAG1 and RPAP3 are homologous proteins, binding, respectively, PIH1D2 and PIH1D1 to form, with RUVBL1/2 proteins, on the one hand, the R2SP complex, and, on the other hand, the R2TP complex. When aligned, the amino acid sequences of SPAG1 and RPAP3 display similarity and identity scores of 34% and 23%, respectively (Figure 1A, Supplementary Figure S1). As previously described, SPAG1 and RPAP3 both possess TPR domains and one specific C-terminal domain, which represents the signature of R2TP-like quaternary chaperones [5]. Despite these similarities, SPAG1 is 30% longer than RPAP3. This difference in size is mainly due to the presence of one supplemental TPR domain in SPAG1, compared to the two TPR domains found in RPAP3. More specifically, analysis of the amino acid alignment suggests that TPRs 1 and 2 in RPAP3 are homologous to TPRs 1 and 3 in SPAG1, leaving an orphan TPR2 in the middle of the SPAG1 sequence. It is noteworthy that five of the eight mutations described in the SPAG1 gene and involved in PCD, take place in TPRs 1 and 3, the three last mutations occurring within the N- and the C-terminal regions [4]. All of the three TPRs in SPAG1 hold five conserved residues defining a carboxylate-clamp motif, suggesting an ability to bind the EEVD sequence located at the C-terminal tail of the HSP family members [49].

SPAG1 binds two EEVD motifs with micromolar affinities

In RPAP3, TPRs 1 and 2 display significantly different affinities for HSP90 C-terminal peptides ($K_d \sim 100$ and $\sim 5 \mu\text{M}$, respectively) and similar affinities for HSP70 peptides ($K_d \sim 50 \mu\text{M}$) [14]. As previously performed for RPAP3, we determined the HSP-binding properties of SPAG1 to get insight into the specific role of its three TPR domains in the putative recruitment of HSP chaperones.

First, we performed protein co-expression assays (Figure 1B, Supplementary Figure S2A) and showed that the full-length SPAG1 co-elutes with his-tagged HSP70, HSP90 and HSP90-MC (the latter protein corresponding to HSP90 truncated of its N-terminal domain). Our results suggest that the C-terminal tails of HSP proteins are sufficient to recruit SPAG1. We further carried out ITC (isothermal titration calorimetry) analyses with the full-length SPAG1 and long peptides mimicking the HSP tails (HSP90: DASRMEEVD and HSP70: SGPTIEEVD) (Figure 1C). Using static light scattering coupled to exclusion size chromatography, we first checked that the full-length SPAG1 was monomeric (Supplementary Figure S2B). Similar K_d values of $\sim 13 \mu\text{M}$ were then measured for the two peptides. More strikingly, a stoichiometry value of ~ 2 was obtained for the two peptides, meaning that two binding sites in SPAG1, with similar affinities toward peptides, are able to provide ITC signal. These results strongly suggest that only two out of the three TPR domains are available in SPAG1 for the binding of HSP tails.

To assess this property, we tried to overexpress isolated TPR domains of SPAG1 in *E. coli*. Whereas TPRs 1 and 3 were found soluble (i.e. fragments 206–327 and 622–742) and suitable for biophysical studies, we failed to solubilize TPR2 with the desired efficiency, despite several attempts toward this end (Supplementary Figure S3A–D). This shortcoming suggests that TPR2 in SPAG1 displays hydrophobic features that cause it to dive mainly into the bacteria pellet after cell lysis and centrifugation. Insofar as recombinant TPRs 1 and 3 are concerned, their purification was successfully performed. TPR3 was found highly soluble and stable, while TPR1 exhibited a marked propensity toward aggregation and instability upon concentration (i.e. for concentration $>150 \mu\text{M}$, Supplementary Figure S3E). Next, using protein co-expression assays, we showed that isolated his-tagged TPRs 1 and 3 co-eluted with full-length HSP70, HSP90 and HSP90-MC (Figure 1D, Supplementary Figure S2A). These interaction data were strengthened, first, using NMR experiments, wherein drastic changes in the SOFAST-METHYL-HMQC spectra of TPRs 1 and 3 occurred upon addition of HSP70 and HSP90-MC (Figure 1E, Supplementary Figure S4A), and, second, using protein–protein cross-link assays (Supplementary Figure S4B). To acquire additional information about the interaction, we then conducted ITC experiments with

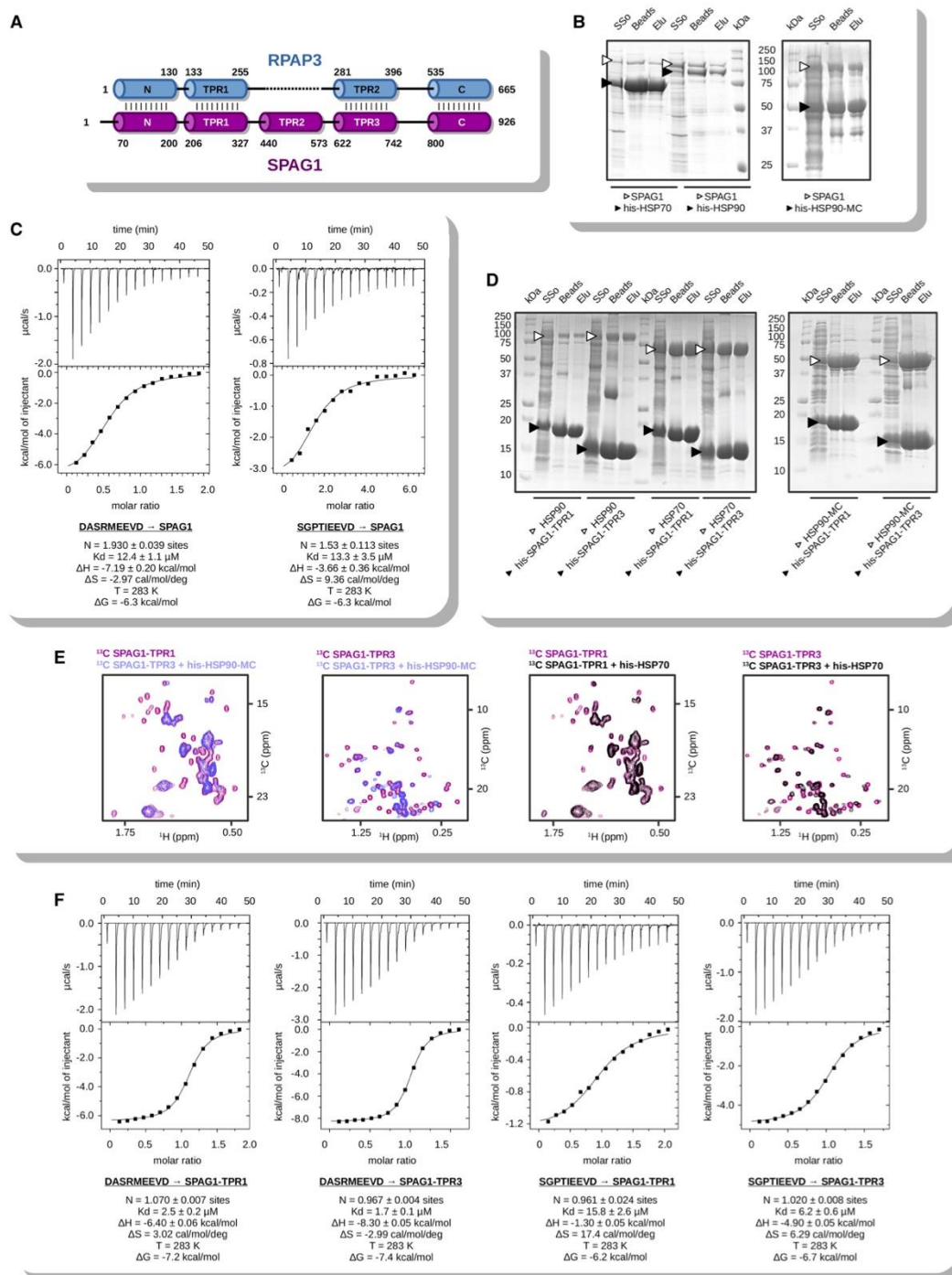


Figure 1. SPAG1 interacts with HSP70 and HSP90.

Part 1 of 2

(A) Schematic comparison between the human RPAP3 and SPAG1 proteins. Limitations of the sub-domains are indicated. (B) Protein co-expression assays between SPAG1 and his-HSP70, his-HSP90 and his-HSP90-MC. ‘SSo’, ‘Beads’ and ‘Elu’ relate respectively to the supernatant sonicate, talon beads and elution from the beads with imidazole. The co-purified proteins are indicated with arrows. (C) ITC experiments recorded between SPAG1 and peptides DASMREEVD (HSP90) and SGPTIEEVD

Figure 1. SPAG1 interacts with HSP70 and HSP90.

Part 2 of 2

(HSP70). Stoichiometry (N), dissociation constant (K_d), temperature (T), variations in enthalpy (ΔH), in entropy (ΔS) and in free energy (ΔG) are reported. (D) Protein co-expression assays between his-SPAG1-TPRs 1 and 3 and HSP90, HSP70 and HSP90-MC (E) SOFAST-METHYL-HMQC spectra of isolated TPR1 and TPR3 of SPAG1 in the absence (purple) and in the presence of his-tagged HSP90-MC (blue) and HSP70 (black). In all cases, methyl resonances are significantly shifted or highly broaden upon addition of HSP proteins. (F) ITC experiments recorded between isolated TPR1 and TPR3 of SPAG1 and peptides DASRMEEVD (HSP90) and SGPTIEEVD (HSP70). Same parameters as those described in C are reported.

the HSP peptides (Figure 1F). The stoichiometry values of ~ 1 demonstrate that each TPR domain possesses one HSP-binding site. With the HSP90 peptide, the K_d values were similar for TPR1 ($K_d = 2.5 \pm 0.2 \mu\text{M}$) and for TPR3 ($K_d = 1.7 \pm 0.1 \mu\text{M}$). With the HSP70 peptide, the K_d values were slightly higher than the ones measured with the HSP90 peptide. Specifically, a 3- and a 6-fold increase were obtained, respectively, for TPR3 ($K_d = 6.2 \pm 0.6 \mu\text{M}$) and for TPR1 ($K_d = 15.8 \pm 0.2 \mu\text{M}$).

Altogether, these data suggest that only TPRs 1 and 3 within the full-length SPAG1 could efficiently recruit the EEVD motif of the HSP family members. Indeed, considering the unexpected insolubility of the TPR2 domain, we can assume that its carboxylate-clamp motif is not accessible within the full-length protein. Last, working with isolated TPR domains slightly enhances the affinity toward HSP90 peptides but should reflect the behavior of these modules in the context of the full-length SPAG1. We, thus, tried to establish the molecular basis underlying the recognition between SPAG1 and members of the HSP family using NMR spectroscopy.

NMR analysis of the free and HSP-bound forms of SPAG1-TPR3

We set out to perform NMR experiments on TPR1. In our hands, this domain was somewhat unstable, thus, precluding a clear NMR assignment. We then focused on TPR3, fully stable and soluble when purified from a $^{13}\text{C}/^{15}\text{N}$ -enriched M9 medium.

First, most of the ^1H , ^{13}C and ^{15}N resonances in SPAG1-TPR3 were assigned and derived into a well-resolved ensemble of three-dimensional structures (Table 1). SPAG1-TPR3 is composed of seven α -helices, the first six organized into three TPR motifs (Figure 2A). Helix $\alpha 7$ plays the role of a capping helix, which may enhance the solubility by shielding several hydrophobic residues located on helix $\alpha 6$ (Figure 2B). The residues of the carboxylate-clamp, namely K627, N631, N662, K692 and R696, are located on the concave face of the domain and are ideally positioned in a cradle ready to bind the EEVD tail of HSP proteins (Figure 2C).

Next, building on our ITC experiments, we monitored by NMR the binding of HSP peptides to ^{15}N -labelled SPAG1-TPR3, using ^1H - ^{15}N CSPs. This method reveals the footprint left by the peptide on the protein domain. Strikingly, the two peptides provided very similar CSP profiles (Figure 2D). Helices $\alpha 1$, $\alpha 3$, $\alpha 5$ (which hold the five carboxylate-clamp residues) and $\alpha 7$ were similarly perturbed upon peptide binding. These data suggest a common binding mode for HSP90 and HSP70, but also, and at first sight, a minor contribution to the binding mode from residues lying upstream the EEVD sequence. Some slight differences could, nonetheless, be observed at the root of helix $\alpha 7$, when HSP70 and HSP90 footprints are compared, which could be correlated to the difference in amino acid sequence of the two peptides and to the slight difference in affinity values measured by ITC (Figure 1F). Thus, in order to understand the molecular mechanisms that drive these binding modes, we modeled SPAG1-TPR3:HSP-peptide complexes and performed MD simulations.

HADDOCK modeling and molecular dynamics simulation of the SPAG1-TPR3: HSP-peptide complexes

We started our structural study by re-assigning the NMR frequencies of ^1H , ^{15}N and ^{13}C nuclei in the HSP-bound forms of SPAG1-TPR3. We then recorded intermolecular NMR signals using filtered NOESY experiments (Figure 3). HSP90 and HSP70 peptides provided common intermolecular NOEs, in agreement with similar CSP profiles imprinted on SPAG1-TPR3. We aimed to use these unambiguous NMR data to drive a docking procedure using the HADDOCK protocol [34,50].

Toward this end, we screened the Protein Data Bank in search of three-dimensional structures of HSP peptides that could globally satisfy both intermolecular data and the CSP profiles. For the HSP90 peptide, we found entry 4I2Z, which depicts the crystal structure of the HSP90-bound form of myosin chaperone UNC-45

Table 1 NMR-derived restraints and structural statistics for the 20 best structures of SPAG1-TPR3 calculated with RECOORD scripts

	SPAG1-TPR3
NMR distances and dihedral constraints	
Total NOEs	3488
Short range ($ i-j \leq 1$)	1605
Medium range ($ i-j < 5$)	943
Long range ($ i-j \geq 5$)	940
Total dihedral angle restraints	239
ϕ	107
ψ	105
χ_1	27
Structure statistics	
Violation occurrences	0
Distance constraints ($>0.5 \text{ \AA}$)	0
Dihedral angle constraints ($>5^\circ$)	0
Deviations from idealized geometry	
Bond lengths ($\times 10^{-3} \text{ \AA}$)	2.743 ± 0.169
Bond angles ($^\circ$)	0.433 ± 0.013
Impropers ($^\circ$)	1.083 ± 0.055
RMSD to best structure (Å)	
All backbone atoms	1.17 ± 0.28
All heavy atoms	1.40 ± 0.22
Backbone atoms in secondary structures	0.45 ± 0.09
Heavy atoms in secondary structures	0.93 ± 0.07
Ramachandran statistics	
Residues in most favored regions (%)	92.1
Residues in additional allowed regions (%)	7.9
Residues in generously allowed regions (%)	0
Residues in disallowed regions (%)	0

from *Caenorhabditis elegans* [35]. We extracted the ASRMEEVD segment from this entry and fulfilled the N-terminal position with an Asp residue. For the HSP70 peptide, we found entry 3Q49, which depicts the crystal structure of the TPR domain of CHIP complexed to peptide PTIEEVD [36]. We extracted the latter segment and fulfilled the N-terminal position with a Ser and a Gly residue. Each peptide was docked onto the best NMR structure of SPAG1-TPR3, using 49 and 40 intermolecular distance restraints for HSP90 and HSP70, respectively (Supplementary Figure S5). For the two HADDOCK calculations, the FCC clustering approach (cutoff = 0.75), subsequent to the final water refinement stage, placed all the 128 calculated structures of each complex in a unique ensemble [37]. No distance restraint violations greater than 0.5 Å were detected. C α RMSD values of $0.76 \pm 0.11 \text{ \AA}$ and $0.77 \pm 0.23 \text{ \AA}$ were measured on the top-10 structures of SPAG1-TPR3 complexed with HSP90 and HSP70, respectively (Supplementary Figure S6). No significant change was observed between the bound and the unbound conformations of SPAG1-TPR3 (C α RMSDs $\sim 0.5 \text{ \AA}$).

Next, to assess the structure and the dynamics of the binding interface, a 605-ns MD simulation was performed on both SPAG1-TPR3:DASRMEEVD and SPAG1-TPR3:SGPTIEEVD immersed in a water box, using the best HADDOCK solutions as a starting point. In the first 5 ns of the simulations, distance restraints were enforced to limit the motions of backbone atoms, while permitting reorganization of the side chains. A single

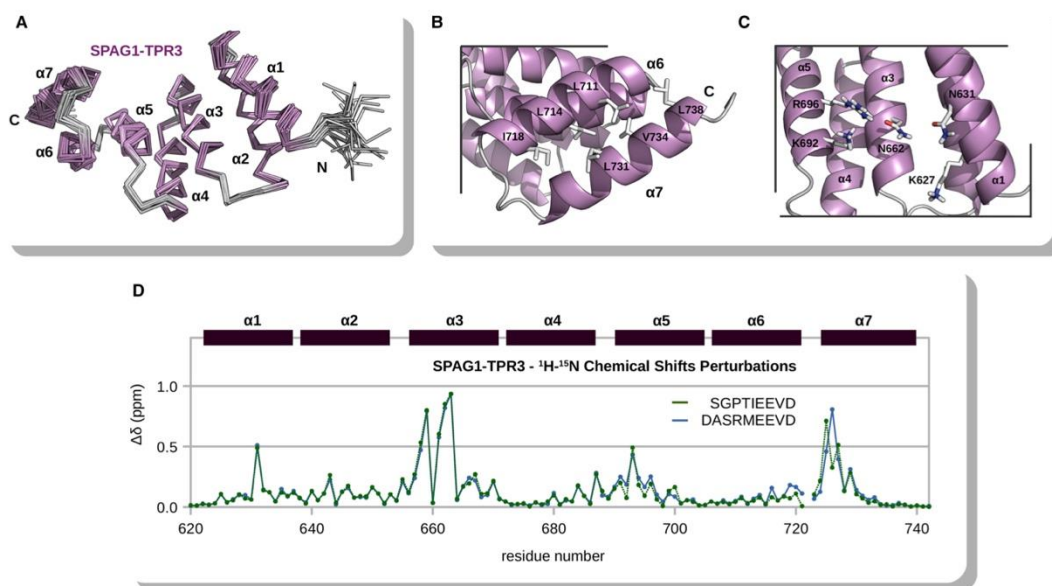


Figure 2. NMR analysis of the free and HSP-bound SPAG1-TPR3.

(A) Backbone representation of the 20 best NMR structures of SPAG1-TPR3. (B) Zoom on the capping helix $\alpha 7$. (C) Zoom on the conserved residues K627, N631, N662, K692 and R696 of the carboxylate-clamp. In (B) and (C), the best NMR structure was used and only polar hydrogens were represented. (D) Chemical shift perturbations ($\Delta\delta$ in ppm) of backbone amide groups in ^{15}N -labelled SPAG1-TPR3 in the presence of peptides DASRMEEVD (blue line) and SGPTIEEVD (green line) were plotted against the protein sequence number. The limitations of SPAG1-TPR3 α -helices are indicated on the top of the diagrams.

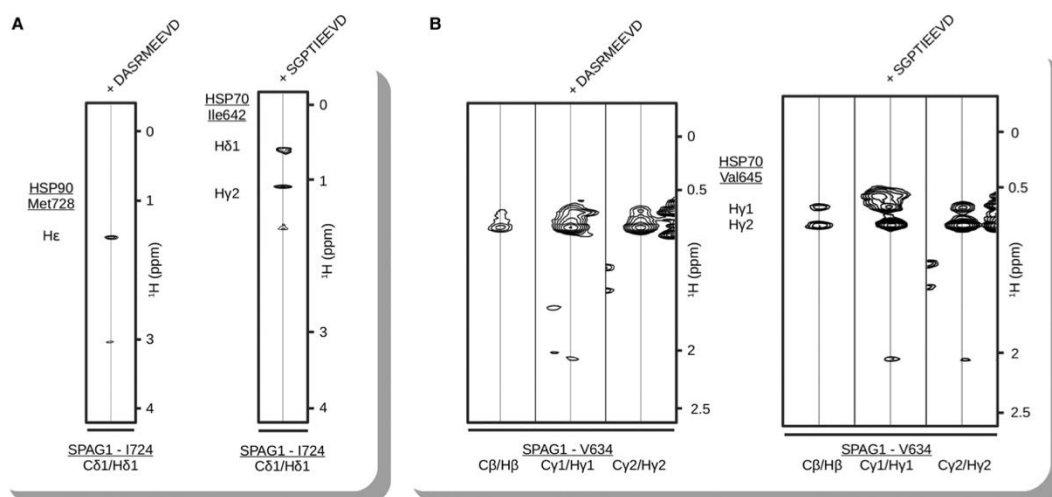


Figure 3. SPAG1-TPR3 shares a common pattern of intermolecular NOEs with the two HSP peptides.

(A) Intermolecular NOEs detected between Ile724 (A) and Val634 (B) in ^{13}C -labelled SPAG1-TPR3 and HSP peptides using 3D-filtered NOESY experiments.

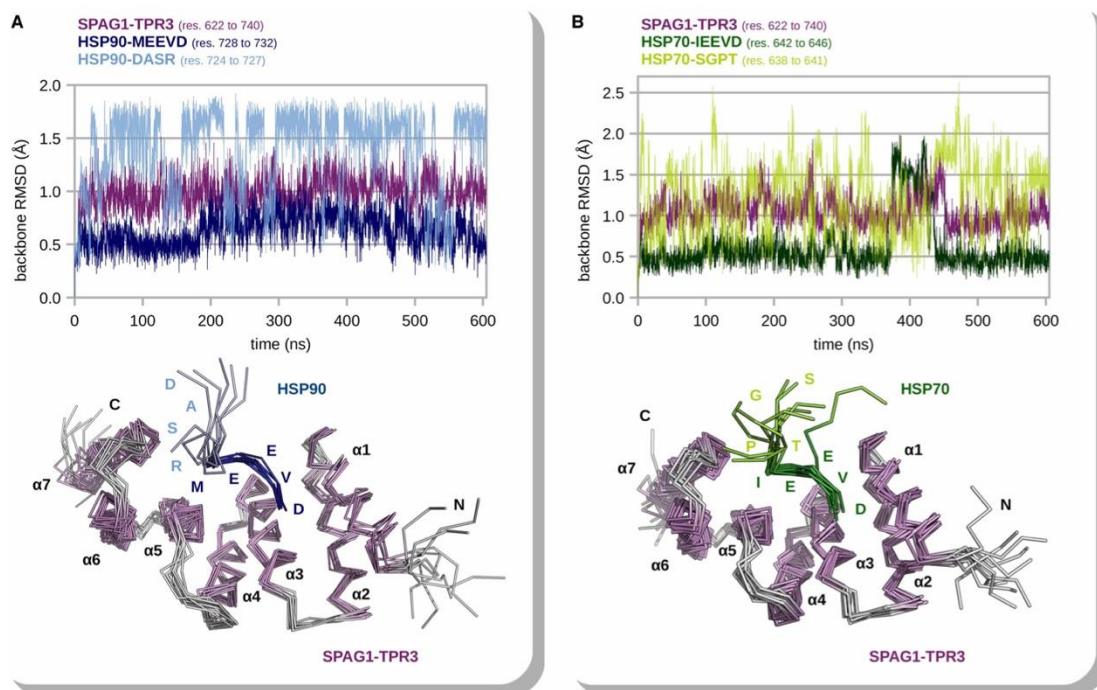


Figure 4. Evolution of backbone RMSDs in SPAG1-TPR3:HSP complexes.

(A) Analysis of the SPAG1-TPR3:DASRMEEVD complex. The backbone RMSD values of SPAG1-TPR3 (residues 622 to 740, in purple) and of the DASR and MEEVD segments in the peptide (respectively, in light and dark blue) are plotted against the simulation time. (B) Analysis of the SPAG1-TPR3:SGPTIEEVD complex. The backbone RMSD values of SPAG1-TPR3 (residues 622–740, in purple) and of the SGPT and IEEVD segments in the peptide (respectively, in light and dark green) are plotted against the simulation time. In (A) and (B), the first frame of the trajectory was the reference structure. In (A) and (B), 3D structures were extracted from the trajectories at 60-ns intervals, superimposed and represented as ribbons.

distance restraint based on intermolecular NOEs was then used to immobilize the peptide over the 50 following ns. Finally, all geometric restraints were removed for the last 550 ns of the simulations. With respective mean backbone RMSD values of $0.99 \pm 0.15 \text{ \AA}$ and $1.05 \pm 0.19 \text{ \AA}$, SPAG1-TPR3 bound to HSP90 and HSP70 peptides showed only minor deviations from the initial position along the trajectories (Figure 4, Supplementary Movies MS1 and MS2). Similarly, the MEEVD and IEEVD parts, in the HSP90 and HSP70 peptides, respectively, were shown to be stable throughout the trajectories, as depicted in the mean backbone RMSD values of $0.63 \pm 0.15 \text{ \AA}$ and $0.61 \pm 0.32 \text{ \AA}$ measured for these segments (Figure 4, Supplementary Movies MS1 and MS2). We could, nevertheless, note a short and partial release of the IEEVD segment between 370 and 430 ns.

Stability of the interface in the two complexes was ensured by a common pattern of interactions. For clarity in the following description, the one-letter code and three-letter code were, respectively, used for the denomination of residues in SPAG1-TPR3 and in HSP peptides. First, an electrostatic network was woven between the five residues K627, N631, N662, K692 and R696 of the carboxylate-clamp motif from SPAG1-TPR3 and the residues Met728 to Asp732 in HSP90 and the corresponding amino acids Ile642 to Asp646 in HSP70 (Figure 5). Specifically, the carboxylate groups of the main and side chains of Asp732/Asp646 in the HSP90/70 peptides were highly stabilized by the side chains of K627, N631, N662 and K692 in SPAG1. R696 in SPAG1 and the backbones of Met728/Ile642 and Glu730/Glu644 in the peptides contributed efficiently to the interface as well as the side chains of K669 and Glu729/Glu643. Second, hydrophobic contacts were formed between the side chains of V634 in SPAG1 and Val731/Val645 in the peptides (Figure 5). This binding mode matches well with that commonly observed in the other three-dimensional structures of HSP-bound TPR domains. Third, supplemental stabilization forces were provided by a hydrophobic pocket made of side chains from V691,

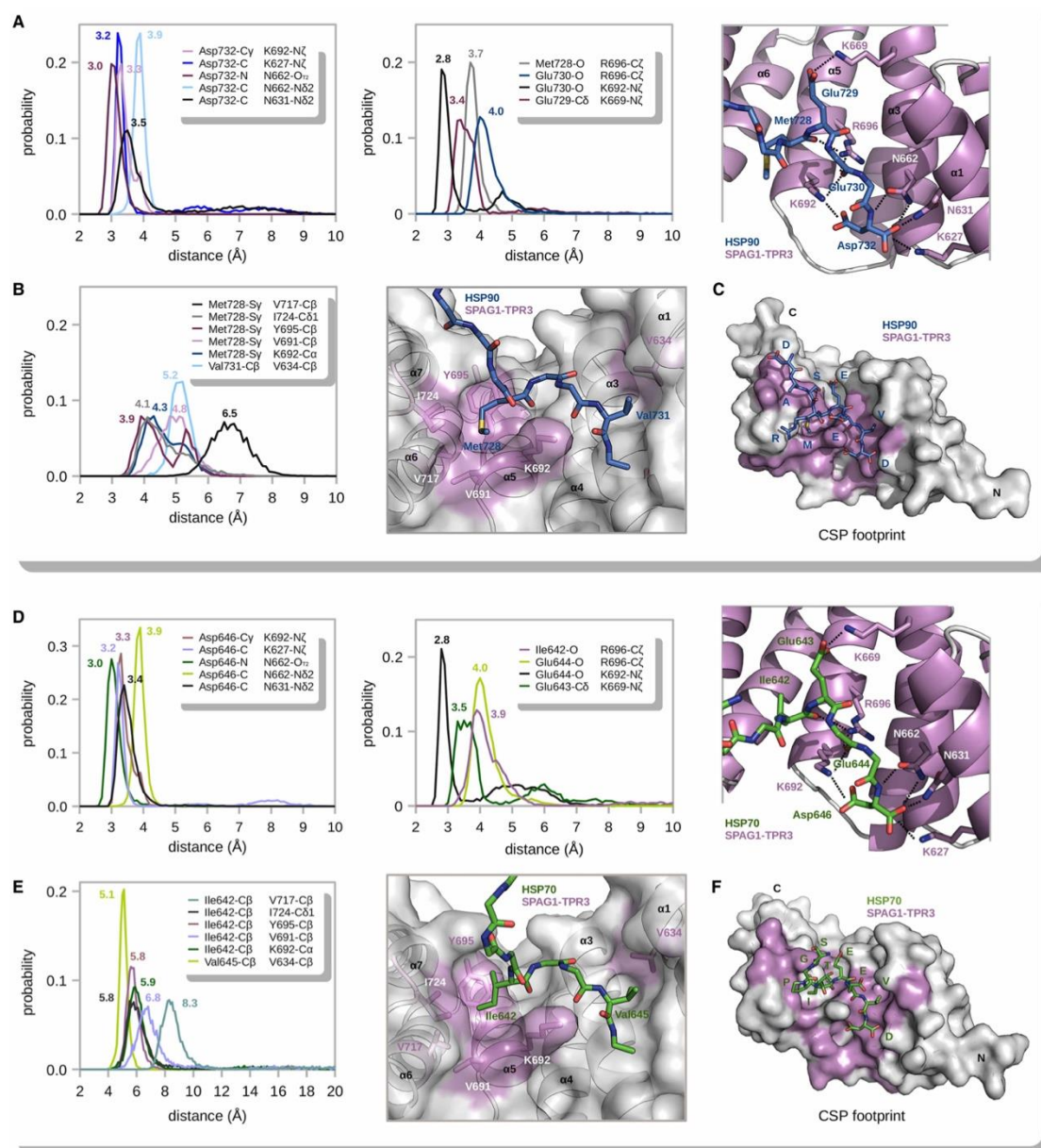


Figure 5. MD analysis of SPAG1-TPR3 in complex with the DASMEEVD and SGTPIEVD peptides.

(A,D) Distribution of intermolecular distances involving carboxylate-clamp residues in SPAG1-TPR3 and atoms in the peptides. For each distribution, the value of the most populated class of distance is indicated. A view of the polar network of interaction is shown on a representative 3D structure extracted from the trajectory. (B,E) Distribution of intermolecular distances involving residues forming hydrophobic pockets in SPAG1-TPR3 and hosted residues in the peptides. For each distribution, the value of the most populated class of distance is indicated. A view of the pockets is shown on a representative 3D structure extracted from the trajectory. (C,F) Residues in SPAG1-TPR3 displaying ^1H - ^{15}N CSP values in the 75th percentile upon peptide binding are colored in purple on the molecular surface of a representative 3D structure of the complexes (see also Figure 2). For convenience, hydrogens were not represented.

K692, Y695, V717 and I724 in SPAG1 harboring the methyl groups of Met728 and Ile642 in the HSP peptides (Figure 5). Similar pockets were also observed in other complexes, such as RPAP3, CHIP and FKBP38 bound to similar peptides [14,36,51].

However, in sharp contrast with the latter MEEVD and IEEVD segments, the DASR and the SGPT extremities in HSP90 and HSP70 exhibited more flexibility along the trajectories (Figure 4, Supplementary Movies MS1 and MS2). The backbone RMSD values of 1.39 ± 0.38 Å and 1.22 ± 0.41 Å measured in the DASR and SGPT parts of the HSP90 and HSP70 peptides reflected their higher mobility. In both complexes, the N-terminal part of the peptides (i.e. the first four residues) appeared more solvent-exposed than the C-terminal part (i.e. the five last residues). However, we were able to highlight in the trajectories significant transient salt bridges established between Arg727 in HSP90 and E721 and E726 in SPAG1-TPR3 (Supplementary Figure S7). Similar interactions could not be observed with the HSP70 peptide, the first residues of which lack strong electron attracting groups. This difference could explain the slight lower affinity measured for the HSP70 peptide in complex with SPAG1-TPR3 (Figure 1F). However, the dynamic behavior observed within the N-terminal segment of the peptides suggests that the majority of the binding energy arises from the stable interface involving both polar and hydrophobic anchoring of the (M/I)EEVD segment into the cradle of SPAG1-TPR3.

Last, the evolution of the intermolecular distances agrees nicely with the experimental NOE and CSP data (Figure 5, Supplementary Figure S6), suggesting that our MD simulations rendered a realistic representation of the three-dimensional structure of the complexes at hand.

SPAG1 does not bind and hydrolyze GTP efficiently

Having deciphered the binding mode of HSPs onto TPR domains at the atomic level, we focused next on another putative substrate of SPAG1 with the primary idea of understanding, at the same atomic scale, the nucleotidase function of the protein. Indeed, GTPase activity of SPAG1 was demonstrated by measuring the release of inorganic phosphate from GTP in the presence of *E. coli* extracts overexpressing or not a recombinant fragment of SPAG1, going from residues 577 to 821 [22]. Authors showed an increase in absorbance at 700 nm in the presence of SPAG1, suggesting a direct contribution of the overexpressed protein in GTP hydrolysis. Using GTP and ATP blot assays, SPAG1 was proposed to bind GTP, while protein sequence analysis highlighted a P-loop motif between residues 781 and 788 (ASEKGGKS). To gain additional information about this function, we aimed to monitor the strength of the binding of GTP to the stable C-terminal fragment 622–926 of SPAG1, using NMR spectroscopy.

First, the HSQC ^1H - ^{15}N spectrum that we recorded on SPAG1(622–926) overlapped well with the spectrum of SPAG1-TPR3 that corresponds to region 622–742 (Figure 6A). Two sets of supplemental peaks could also be observed in this spectrum. The first one, with NMR signals of moderate intensities spreading over the spectrum widths, displays the features of well-folded protein regions. The second one, with intense ^1H and ^{15}N resonances centered, respectively, around 8 ppm and 120 ppm, is characteristic of disordered regions in proteins. These spectral features led us to reason that SPAG1(622–926) could behave as a protein made of two independent domains devoid of strong contacts between each other. Due to the multi-modular organization of SPAG1 (Figure 1A, Supplementary Figure S1), we could reasonably assign these peaks to the C-terminal domain of the protein and to the linker upstream that makes the connection with TPR3.

Surprisingly, we did not observe any drastic change in the NMR spectrum of SPAG1(622–926) in the presence of a 10-molar excess of GTP, even when Mg^{2+} cations were added (Figure 6B). An efficient binding of the nucleotide should have significantly modified the chemical shift values and/or the intensities of those residues involved in the binding of the nucleotide. Cautious inspection of the SPAG1(622–926) spectra recorded with a huge excess of nucleotide showed, nonetheless, that few isolated NH resonances underwent observable chemical shift modifications upon addition of GTP (Figure 6B,C). These CSPs were derived into a K_d greater than 5 mM, a value for which the question of biological relevance could be raised. This result strongly suggests that the 622–926 fragment of SPAG1 was not able to bind GTP efficiently *in vitro* (at least with a sub-millimolar affinity).

As very low binding could not totally be related to low hydrolysis efficiency, we monitored the release of inorganic phosphate by SPAG1(622–926) from GTP in the presence of MgCl_2 , using ^{31}P NMR. Even after 24 h of incubation, no ^{31}P peak between 0 and 5 ppm appeared, thus, indicating that hydrolysis of GTP could not be observed (Figure 6D). These data call into question the efficiency of GTPase activity in SPAG1.

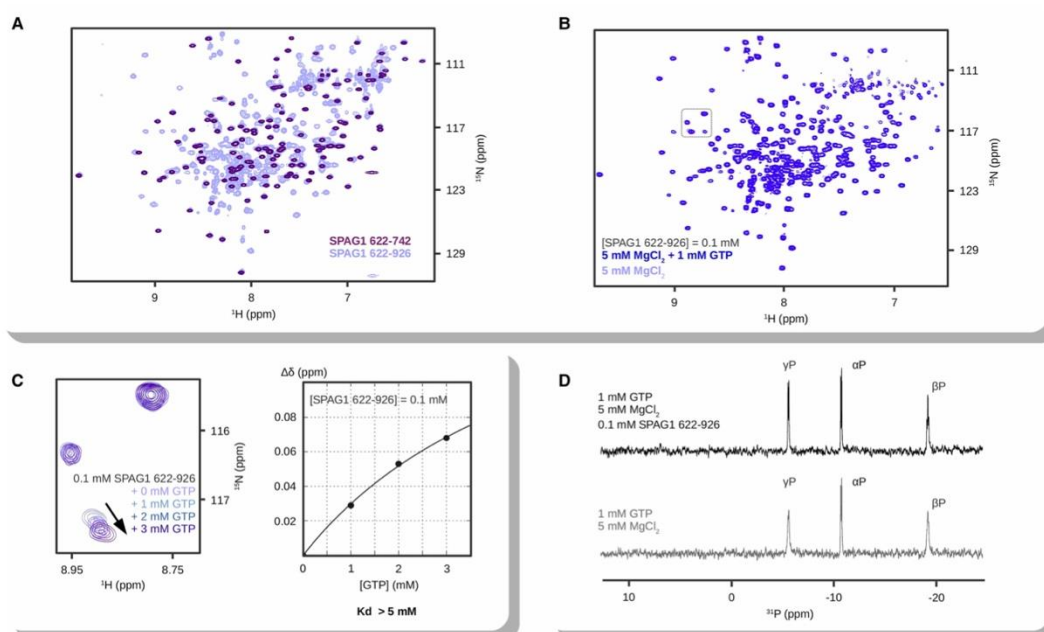


Figure 6. Insights into the GTPase activity of SPAG1.

(A) Superimposition of the ^1H - ^{15}N HSQC spectra of the ^{15}N -labelled fragments 622–742 (in purple) and 622–926 (in light blue) of SPAG1 recorded in the NMR buffer (10 mM NaPi (pH 6.4), 150 mM NaCl, 0.5 mM TCEP) at 293 K. (B) BEST-TROSY ^1H - ^{15}N HSQC spectra of SPAG1(622–926) recorded in the NMR buffer supplemented with 5 mM MgCl_2 in the presence (marine blue) or absence (light blue) of 1 mM GTP. (C) Chemical shift perturbations upon GTP addition on SPAG1(622–926) monitored in experimental conditions mentioned in (B) and curve fitting of the data. The spectral region that was studied is highlighted in the gray box drawn in (B). (D) 1D ^{31}P NMR spectra of 1 mM GTP in the presence (black) or absence (gray) of 0.1 mM ^{15}N -labelled SPAG1(622–926) recorded at 298 K in 10 mM HEPES (pH 7.5), 150 mM NaCl, 5 mM MgCl_2 and 10% D_2O . Reference (gray) and test (black) samples were prepared 24 h prior to data acquisition. Assignment of ^{31}P nuclei in GTP is indicated.

Discussion

SPAG1 is a platform for HSP proteins

SPAG1 has been proposed to be involved in the assembly of dynein arms, and thus, in the function of cilia. Quaternary assembly of this kind of macromolecular complex is supervised by proteins with chaperoning activities. We have demonstrated here the ability of SPAG1 to recruit directly HSP70 and HSP90. We have shown that SPAG1 binds the tails of HSP70 and HSP90 with comparable affinities, either for the full-length protein or for the isolated TPR domains. The affinities measured between isolated SPAG1-TPRs 1 and 3 and the DASRMEEVD peptide (HSP90) are slightly higher than the ones measured with the SGPTIEEVD peptide (HSP70), thereby indicating that none of the TPR domains seems dedicated to one specific chaperone. This idea is supported by our synergistic NMR and MD study of SPAG1-TPR3 and HSP peptides, illuminating that DASRMEEVD and SGPTIEEVD peptides can accommodate their protein receptor in a similar way. Specifically, on one side, the EEVD residues of the peptide are consensually bound to the carboxylate-clamp motif of the TPR domain and, on the other side, the Met728 and Ile642 residues (which were proposed as key residues for HSP specificity) are stably hosted in the same pocket in SPAG1-TPR3. This observation is in contrast with multi-TPR HSP co-chaperone proteins such as RPAP3 or HOP, the TPR domains of which have been proposed devoted to HSP70 or HSP90 in order to facilitate protein–client exchange inside a chaperone/co-chaperone complex [19]. For example, we have previously documented two different binding modes for HSP70 and HSP90 peptides on RPAP3-TPR2 and related these results to an enhanced affinity of RPAP3-TPR2 for HSP90. Here, our results indicate that SPAG1 could be potentially saturated by one or the other HSP proteins. It is noteworthy that previous IP-SILAC experiments using PIH1D2 as bait have revealed that HSP70

was strongly associated with SPAG1, RUVBL1, RUVBL2 and PIH1D2 proteins [5], suggesting that HSP70 could be the privileged partner of SPAG1 *in vivo*. This result also raises the question of the putative regulating effect of SPAG1 on HSP chaperones. Indeed, one might wonder whether SPAG1 is a simple carrier for HSP proteins or is able to enhance or decrease the chaperoning abilities of the latter. To address this question, it would be of interest to test the effect of the full-length SPAG1 and its sub-fragments on the ATPase activity of HSPs. We propose in the present work that SPAG1-TPR2 could not participate in the binding of the HSP C-terminal tails due to its biochemical features. Still, one might question whether TPR2 only plays an organizational role in SPAG1, facilitating the proper folding of the protein, or act as a regulatory module that modifies the ATPase activity of the HSP chaperones through contacts that do not involve its carboxylate-clamp motif. Such regulatory roles were highlighted for the TPR domain of Sec72 that is able to interact with the ATPase domain of Ssb1, a yeast HSP70 homolog lacking the EEVD motif [52] and for Sti1, whose TPR2B domain interacts with the M domain of HSP90 [19]. In the same way, the formation of HSP70 dimers, and consequently the chaperoning activity of the protein, is also differentially regulated by TPR-containing proteins, Chip and Tomm34 [53].

Is SPAG1 a GTPase?

Using the sub-fragment 577–821, SPAG1 was described years ago as a GTPase protein [22]. In our study, we failed to highlight catalytic activity for the sub-fragment 622–926 of SPAG1. Although different in size, these two sub-fragments equally contain a putative P-loop motif that lays between residues 781 and 788. Using NMR, we have demonstrated the stability and the proper folding of sub-fragment 622–926 (Figure 6A,B). In addition, sequence analysis of SPAG1 defines the region 577–621 in SPAG1 as a linker connecting TPR domains 2 and 3 devoided of any specific motif, whereas region 822–926 appears folded (Figure 6A, Supplementary Figure S1). Thus, although the use of the sub-fragment 622–926 appears totally relevant, the P-loop motif remained silent in our experiments. Moreover, the SPAG1 sequence did not reveal the presence of a G domain, which is a conserved domain in GTPases.

However, among partners, SPAG1 was also able to interact with RUVBL1 and RUVBL2 proteins, two other protein chaperones both shown to be able to hydrolyze ATP. To fulfill this function, RUVBLs proteins contain Walker A and B motifs, as well as ADP/ATP sensor motifs. The Walker A/P-loop sequence is a common motif for ATP and GTP binding. Merging these data together leads us to propose that SPAG1 might have a general enhancing effect on nucleotide hydrolysis. Indeed, we could easily assume that SPAG1 might have stimulated GTPases from *E. coli* in the past experiment [22]. In turn, this hypothesis suggests that SPAG1 could be a key activator of Walker nucleotidases, such as RUVBL1/2 proteins, which could facilitate their role in the quaternary assembly of macromolecular complexes. Interestingly enough, the C-terminal domain of SPAG1 is homologous to the C-terminal domain of RPAP3, which has been shown to be in direct contact with the RUVBL1/2 hetero-hexameric ring. Thus, this domain could be an important module in the regulation of the nucleotide hydrolysis function. As a perspective, extensive activity assays might highlight that SPAG1, through its multiple domains, is a fine tuner of the nucleotide hydrolysis activity of non-Walker and Walker nucleotidases (i.e. HSP and RUVBL proteins, respectively) within a large R2SP/HSP complex.

Abbreviations

cryo-EM, electronic cryo-microscopy; CS, Chord and Sgt1; CSP, chemical shift perturbation; DAAF, dynein axonemal assembly factors; FCC, Fraction of Common Contacts; HSP, heat shock protein; HSQC, heteronuclear single quantum coherence; IPTG, isopropyl β -D-1-thiogalactopyranoside; ITC, isothermal titration calorimetry; K_d , dissociation constant; MD, molecular dynamics; NMR, nuclear magnetic resonance; PCD, primary ciliary dyskinesia; PIH, protein interacting with HSP; TPR, tetratricopeptide repeat.

Data Depositions

Chemical shifts and 3D coordinates of the free form of SPAG1-TPR3 were deposited in the Biological Magnetic Resonance Data Bank and in the Protein Data Bank under respective accession numbers 34329 and 6I57.

Author Contribution

M.Q. conceived and coordinated the project. M.E.C, S.D. and X.M. were involved in the DNA cloning steps. M.E.C., R.D.S.M., D.L. and M.Q. performed and analyzed biochemistry and biophysics experiments. C.C., F.D. and M.Q. performed and analyzed *in silico* experiments. M.Q. wrote the original draft. All authors reviewed the manuscript.

Funding

This work was supported by the French National Center for Scientific Research (CNRS), University of Lorraine (France), the 'Région Grand Est' and the 'Agence Nationale de la Recherche' [Grant No. ANR-16-CE11-0032-02]. C.C and F.D. acknowledge the State-Region Plan 'Technological Innovations, Modeling and Personalized Medical Support' (IT2MP) and the European Regional Development Funds (ERDF) for generous support.

Acknowledgements

We thank IBSLor (UMS 2008 CNRS, INSERM, Université de Lorraine) for ITC, DLS, SEC-SLS and NMR core facilities. We thank Institut Jean Barriol (IJB, FR2843 CNRS) for ³¹P NMR core facilities.

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

- Heydeck, W., Fievet, L., Davis, E.E. and Katsanis, N. (2018) The complexity of the cilium: spatiotemporal diversity of an ancient organelle. *Curr. Opin. Cell Biol.* **55**, 139–149 <https://doi.org/10.1016/j.ceb.2018.08.001>
- Wang, L. and Dynlacht, B.D. (2018) The regulation of cilium assembly and disassembly in development and disease. *Development* **145**, dev151407 <https://doi.org/10.1242/dev.151407>
- Mitchison, H.M., Schmidts, M., Loges, N.T., Freshour, J., Dritsoula, A., Hirst, R.A. et al. (2012) Mutations in axonemal dynein assembly factor DNAAF3 cause primary ciliary dyskinesia. *Nat. Genet.* **44**, 381–389. S1–S2 <https://doi.org/10.1038/ng.1106>
- Knowles, M.R., Ostrowski, L.E., Loges, N.T., Hurd, T., Leigh, M.W., Huang, L. et al. (2013) Mutations in SPAG1 cause primary ciliary dyskinesia associated with defective outer and inner dynein arms. *Am. J. Human Genet.* **93**, 711–720 <https://doi.org/10.1016/j.ajhg.2013.07.025>
- Maurizy, C., Quintemet, M., Abel, Y., Verheggen, C., Santo, P.E., Bourguet, M. et al. (2018) The RPAP3-Cterminal domain identifies R2TP-like quaternary chaperones. *Nat. Commun.* **9**, 2093 <https://doi.org/10.1038/s41467-018-04431-1>
- Dafinger, C., Rinschen, M.M., Borgal, L., Ehrenberg, C., Basten, S.G., Franke, M. et al. (2018) Targeted deletion of the AAA-ATPase Ruvb1 in mice disrupts ciliary integrity and causes renal disease and hydrocephalus. *Exp. Mol. Med.* **50**, 75 <https://doi.org/10.1038/s12276-018-0108-z>
- Kakihara, Y. and Houry, W.A. (2012) The R2TP complex: discovery and functions. *Biochim. Biophys. Acta* **1823**, 101–107 <https://doi.org/10.1016/j.bbamcr.2011.08.016>
- Houry, W.A., Bertrand, E. and Coulombe, B. (2018) The PAQosome, an R2TP-based chaperone for quaternary structure formation. *Trends Biochem. Sci.* **43**, 4–9 <https://doi.org/10.1016/j.tibs.2017.11.001>
- Lage P. Z., Stefanopoulou, P., Styczynska-Soczka, K., Quinn, N., Mali, G., von Kriegsheim, A. et al. (2018) Ciliary dynein motor preassembly is regulated by Wdr92 in association with HSP90 co-chaperone, R2TP. *J. Cell Biol.* **217**, 2583–2598 <https://doi.org/10.1083/jcb.201709026>
- Liu, G., Wang, L. and Pan, J. (2018) Chlamydomonas WDR92 in association with R2TP-like complex and multiple DNAAFs to regulate ciliary dynein preassembly. *J. Mol. Cell Biol.* <https://doi.org/10.1093/jmcb/mjy067>
- Martino, F., Pal, M., Muñoz-Hernández, H., Rodríguez, C.F., Núñez-Ramírez, R., Gil-Carton, D. et al. (2018) RPAP3 provides a flexible scaffold for coupling HSP90 to the human R2TP co-chaperone complex. *Nat. Commun.* **9**, 1501 <https://doi.org/10.1038/s41467-018-03942-1>
- Rivera-Calzada, A., Pal, M., Muñoz-Hernández, H., Luque-Ortega, J.R., Gil-Carton, D., Degliesposti, G. et al. (2017) The structure of the R2TP complex defines a platform for recruiting diverse client proteins to the HSP90 molecular chaperone system. *Structure* **25**, 1145–1152.e4 <https://doi.org/10.1016/j.str.2017.05.016>
- Pal, M., Morgan, M., Phelps, S.E., Roe, S.M., Parry-Morris, S., Downs, J.A. et al. (2014) Structural basis for phosphorylation-dependent recruitment of Tel2 to Hsp90 by Pih1. *Structure* **22**, 805–818 <https://doi.org/10.1016/j.str.2014.04.001>
- Henri, J., Chagot, M.E., Bourguet, M., Abel, Y., Terral, G., Maurizy, C. et al. (2018) Deep structural analysis of RPAP3 and PIH1D1, two components of the HSP90 co-chaperone R2TP complex. *Structure* **26**, 1196–1209.e8 <https://doi.org/10.1016/j.str.2018.06.002>
- Quinet, M., Rothé, B., Barbier, M., Bobo, C., Saliou, J.M., Jacquemin, C. et al. (2015) Structure/function analysis of protein-protein interactions developed by the yeast Pih1 platform protein and its partners in box C/D snoRNP assembly. *J. Mol. Biol.* **427**, 2816–2839 <https://doi.org/10.1016/j.jmb.2015.07.012>
- Tian, S., Yu, G., He, H., Zhao, Y., Liu, P., Marshall, A.G. et al. (2017) Pih1p-Tah1p puts a lid on hexameric AAA⁺ ATPases Rvb1/2p. *Structure* **25**, 1519–1529.e4 <https://doi.org/10.1016/j.str.2017.08.002>
- Blatch, G.L. and Lässle, M. (1999) The tetrapeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21**, 932–939 [https://doi.org/10.1002/\(SICI\)1521-1878\(199911\)21:11<932::AID-BIES5>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1521-1878(199911)21:11<932::AID-BIES5>3.0.CO;2-N)
- Chen, S. and Smith, D.F. (1998) Hop as an adaptor in the heat shock protein 70 (Hsp70) and hsp90 chaperone machinery. *J. Biol. Chem.* **273**, 35194–35200 <https://doi.org/10.1074/jbc.273.52.35194>
- Schmid, A.B., Lagleder, S., Gräwert, M.A., Röhl, A., Hagn, F., Wandinger, S.K. et al. (2012) The architecture of functional modules in the Hsp90 co-chaperone Sti1/Hop. *EMBO J.* **31**, 1506–1517 <https://doi.org/10.1038/emboj.2011.472>
- Kryzstofinska, E.M., Evans, N.J., Thapaliya, A., Murray, J.W., Morgan, R.M.L., Martínez-Lumbreras, S. et al. (2017) Structure and interactions of the TPR domain of Sgt2 with yeast chaperones and Ybr137wp. *Front. Mol. Biosci.* **4**, 68 <https://doi.org/10.3389/fmolb.2017.00068>
- Benbahouche Nel, H., Iliopoulos, I., Török, I., Marhold, J., Henri, J., Kajava, A.V. et al. (2014) Drosophila Spag is the homolog of RNA polymerase II-associated protein 3 (RPAP3) and recruits the heat shock proteins 70 and 90 (Hsp70 and Hsp90) during the assembly of cellular machineries. *J. Biol. Chem.* **289**, 6236–6247 <https://doi.org/10.1074/jbc.M113.499608>
- Lin, W., Zhou, X., Zhang, M., Li, Y., Miao, S., Wang, L. et al. (2001) Expression and function of the HSD-3.8 gene encoding a testis-specific protein. *Mol. Hum. Reprod.* **7**, 811–818 <https://doi.org/10.1093/molehr/7.9.811>

- 23 Liu, N., Qiao, Y., Cai, C., Lin, W., Zhang, J., Miao, S. et al. (2006) A sperm component, HSD-3.8 (SPAG1), interacts with G-protein β 1 subunit and activates extracellular signal-regulated kinases (ERK). *Front. Biosci.* **11**, 1679–1689 <https://doi.org/10.2741/1913>
- 24 Huang, C., Wu, D., Khan, F.A., Jiao, X., Guan, K. and Huo, L. (2016) The GTPase SPAG-1 orchestrates meiotic program by dictating meiotic resumption and cytoskeleton architecture in mouse oocytes. *Mol. Biol. Cell* **27**, 1776–1785 <https://doi.org/10.1091/mbc.e16-02-0132>
- 25 Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) Distantly related sequences in the α - and β -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951 <https://doi.org/10.1002/j.1460-2075.1982.tb01276.x>
- 26 Ewens, C.A., Su, M., Zhao, L., Nano, N., Houry, W.A. and Southworth, D.R. (2016) Architecture and nucleotide-dependent conformational changes of the Rvb1-Rvb2 AAA* complex revealed by cryoelectron microscopy. *Structure* **24**, 657–666 <https://doi.org/10.1016/j.str.2016.03.018>
- 27 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.* **175**, 178–188 <https://doi.org/10.1016/j.jsb.2011.02.001>
- 28 Kriznik, A., Yelehe-Okouma, M., Lec, J.C., Grosheny, G., Le Cordier, H., Charron, C. et al. (2018) CRD_{SAT} generated by pCARGHO: a new efficient lectin-based affinity tag method for safe, simple, and low-cost protein purification. *Biotechnol. J.* **14**, e1800214 <https://doi.org/10.1002/biot.201800214>
- 29 Anthis, N.J. and Clore, G.M. (2013) Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci.* **22**, 851–858 <https://doi.org/10.1002/pro.2253>
- 30 Shen, Y. and Bax, A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 <https://doi.org/10.1007/s10858-013-9741-y>
- 31 Berjanskii, M.V., Neal, S. and Wishart, D.S. (2006) PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Res.* **34**, W63–W69 <https://doi.org/10.1093/nar/gkl341>
- 32 Guntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378 <https://doi.org/10.1385/1-59259-809-9:353>
- 33 Nederveen, A.J., Doreleijers, J.F., Vranken, W., Miller, Z., Spronk, C.A., Nabuurs, S.B. et al. (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**, 662–672 <https://doi.org/10.1002/prot.20408>
- 34 de Vries, S.J., van Dijk, A.D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V. et al. (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* **69**, 726–733 <https://doi.org/10.1002/prot.21723>
- 35 Gazda, L., Pokrzywa, W., Hellerschmied, D., Löwe, T., Forné, I., Mueller-Planitz, F. et al. (2013) The myosin chaperone UNC-45 is organized in tandem modules to support myofibril formation in *C. elegans*. *Cell* **152**, 183–195 <https://doi.org/10.1016/j.cell.2012.12.025>
- 36 Wang, L., Liu, Y.T., Hao, R., Chen, L., Chang, Z., Wang, H.R. et al. (2011) Molecular mechanism of the negative regulation of Smad1/5 protein by carboxyl terminus of Hsc70-interacting protein (CHIP). *J. Biol. Chem.* **286**, 15883–15894 <https://doi.org/10.1074/jbc.M110.201814>
- 37 Rodrigues, J.P., Trellet, M., Schmitz, C., Kastritis, P., Karaca, E., Melquiand, A.S. et al. (2012) Clustering biomolecular complexes by residue contacts similarity. *Proteins* **80**, 1810–1817 <https://doi.org/10.1002/prot.24078>
- 38 MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J. et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* **102**, 3586–3616 <https://doi.org/10.1021/jp973084f>
- 39 MacKerell, Jr, A.D., Feig, M. and Brooks, 3rd, C.L. (2004) Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **126**, 698–699 <https://doi.org/10.1021/ja036959e>
- 40 Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 <https://doi.org/10.1063/1.445869>
- 41 Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E. et al. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 <https://doi.org/10.1002/jcc.20289>
- 42 Tuckerman, M., Berne, B.J. and Martyna, G.J. (1992) Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990–2001 <https://doi.org/10.1063/1.463137>
- 43 Andersen, H.C. (1983) Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **52**, 24–34 [https://doi.org/10.1016/0021-9991\(83\)90014-1](https://doi.org/10.1016/0021-9991(83)90014-1)
- 44 Martyna, G.J., Tobias, D.J. and Klein, M.L. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 <https://doi.org/10.1063/1.467468>
- 45 Feller, S.E., Zhang, Y., Pastor, R.W. and Brooks, B.R. (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 <https://doi.org/10.1063/1.470648>
- 46 Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 <https://doi.org/10.1063/1.464397>
- 47 Fiorin, G., Klein, M.L. and Hémin, J. (2013) Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **111**, 3345–3362 <https://doi.org/10.1080/00268976.2013.813594>
- 48 Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38. 27–28 [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
- 49 Scheuffer, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H. et al. (2000) Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell* **101**, 199–210 [https://doi.org/10.1016/S0092-8674\(00\)80830-2](https://doi.org/10.1016/S0092-8674(00)80830-2)
- 50 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 <https://doi.org/10.1021/ja026939x>
- 51 Blundell, K.L., Pal, M., Roe, S.M., Pearl, L.H. and Prodromou, C. (2017) The structure of FKBP38 in complex with the MEEVD tetrapeptide binding-motif of Hsp90. *PLoS ONE* **12**, e0173543 <https://doi.org/10.1371/journal.pone.0173543>
- 52 Tripathi, A., Mandon, E.C., Gilmore, R. and Rapoport, T.A. (2017) Two alternative binding mechanisms connect the protein translocation Sec71-Sec72 complex with heat shock proteins. *J. Biol. Chem.* **292**, 8007–8018 <https://doi.org/10.1074/jbc.M116.761122>
- 53 Trcka, F., Durech, M., Vankova, P., Chmelik, J., Martinkova, V., Hausner, J., et al. (2019) Human stress-inducible Hsp70 has a high propensity to form ATP-dependent antiparallel dimers that are differentially regulated by cochaperone binding. *Mol. Cell. Proteomics* **18**, 320–337 <https://doi.org/10.1074/mcp.RA118.001044>

Optimizing the First TPR Domain of the Human SPAG1 Protein Provides Insight into the HSP70 and HSP90 Binding Properties

Sana Dermouche,[§] Marie-Eve Chagot,[§] Xavier Manival,[§] and Marc Quinteret*



Cite This: <https://doi.org/10.1021/acs.biochem.1c00052>



Read Online

ACCESS |



Metrics & More

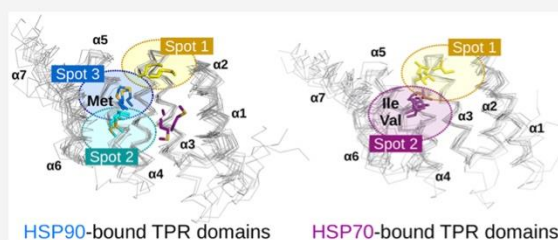


Article Recommendations



Supporting Information

ABSTRACT: Tetratricopeptide repeat domains, or TPR domains, are protein domains that mediate protein:protein interaction. As they allow contacts between proteins, they are of particular interest in transient steps of the assembly process of macromolecular complexes, such as the ribosome or the dynein arms. In this study, we focused on the first TPR domain of the human SPAG1 protein. SPAG1 is a multidomain protein that is important for ciliogenesis whose known mutations are linked to primary ciliary dyskinesia syndrome. It can interact with the chaperones RUVBL1/2, HSP70, and HSP90. Using protein sequence optimization in combination with structural and biophysical approaches, we analyzed, with atomistic precision, how the C-terminal tails of HSPs bind a variant form of SPAG1-TPR1 that mimics the wild-type domain. We discuss our results with regard to other complex three-dimensional structures with the aim of highlighting the motifs in the TPR sequences that could drive the positioning of the HSP peptides. These data could be important for the druggability of TPR regulators.



To ensure the efficient functioning of biological pathways in cells, proteins must interact through a complex network weaved within and between cellular compartments, as well as with lipid membranes and the extracellular medium. To transmit information, proteins can direct contacts with other proteins, nucleic acids, lipids, or small molecules. Thus, domains dedicated to the binding of specific types of partners have appeared in proteins, sometimes in several copies, to finely regulate their communication abilities.¹ The rise of structural biology highlighted the three-dimensional (3D) structures of many protein domains, commonly found in proteins, like domains SH2 and SH3 (Src Homology 2 and 3, respectively), PDZ (PSD-95 Dlg1 ZO-1), zinc fingers, RRM (RNA recognition motif), CRD (carbohydrate recognition domain), and many more. Fascinatingly, structural biology permits us to access, with an atomic precision, the molecular determinants that drive interactions between these domains and their biological partners. This latter point is a key point in investigating the druggability of protein targets, especially for proteins coming from pathogenic bacteria, viruses, and parasites.² To answer functional questions, structural biology studies require significant amounts of samples and their analyses in vitro through methods such as nuclear magnetic resonance (NMR), X-ray crystallography, or cryo-electron microscopy. At a first approach, proteins of interest are purified using bacterial expression systems because they are, most of the time, easy to set up. Unfortunately, not all proteins can be efficiently purified due to their intrinsic features that can make them insoluble for example. To get through this bottleneck, prokaryotic, eukaryotic, and in vitro expression systems were

developed in combination with more and more complex purification protocols.^{3–5} Another possibility is to slightly modify the wild-type sequence of a protein to facilitate expression under simple experimental conditions.⁶ This approach requires control analyses that demonstrate that the behavior of the modified protein (meaning enzymatic activity, binding capacities, conformational adaptability, etc.) does not differ from its native form.

There have been extensive studies of the TPR domain (tetratricopeptide repeat) that is present in dozens of entries in the Protein Data Bank, a worldwide library that collects 3D structures of biological macromolecules. The TPR domain consists of a repetition of at least two TPR motifs, each motif being generally made of 34 amino acids shaped into a helix–turn–helix structure.⁷ Moreover, solubilization of the domain is facilitated by a C-terminal α -helix that caps hydrophobic residues of the last TPR motif.^{7,8} To maintain the characteristic helix–turn–helix 3D shape, positions need to be conserved within the amino acid sequence, especially for aromatic residues.⁷ Thus, on the basis of sequence conservation and known 3D structures, a consensus TPR motif could have been successfully designed.⁹ The 3D features

Received: January 19, 2021

Revised: March 10, 2021

of the TPR motif give to the TPR domain the shape of a concave cradle,⁷ until forming a superhelix in case of a long juxtaposition of many copies of TPR motifs.¹⁰ From a functional point of view, the TPR domain mediates protein:protein interactions, in particular with heat shock proteins HSP70 and HSP90. More precisely, this latter interaction is allowed because of five conserved polar residues located in the cradle of the TPR domain that promote polar contacts with the four last EEVD residues of the HSPs.^{11,12} The five conserved residues in the TPR domain form a clamp that traps the two carboxylate groups of the C-terminal aspartic acids in HSPs, hence the name of “carboxylate clamp” given to this five-residue motif.¹³ In addition, the interaction is also maintained by hydrophobic contacts involving the penultimate valine in HSPs and hydrophobic residues in the TPR. Finally, residues upstream of the EEVD sequence were also shown to be important in the binding. A methionine, found in only HSP90, and an isoleucine, found in only HSP70, are both located just before the EEVD motif and can accommodate a pocket at the surface of the TPR.^{13–16} It was proposed that this Met/Ile switch represents a crucial point for HSP selectivity.¹³ Moreover, the amino acids upstream the Met/Ile residue can also be crucial for strengthening the binding of HSPs to a TPR domain.^{14,15} Nonetheless, the cradle in the TPR is not the only area that can promote direct contacts with protein partners. Indeed, the opposite face, which displays a convex curvature, can interact with a partner for regulation. It is notably the case of TPR domains in proteins Sti1 and Sec72.^{17,18}

SPAG1 (sperm-associated antigen 1) is a human protein, present in lung and testis, that possesses HSP binding abilities and is required in the preassembly of dynein arms.^{15,19,20} Mutations in the SPAG1 gene were shown to be responsible for a human syndrome, called primary ciliary dyskinesia (PCD), that affects lung and fertility functions by weakening the motility of flagellar and ciliary cells.²⁰ SPAG1 is a 104 kDa protein that exhibits at least four distinct domains, including three carboxylate-clamp TPR domains in a row.^{15,21} SPAG1 is homologous to RPAP3, the two proteins possessing a RPAP3C domain that signals their interaction capacity with two other chaperones, RUVBL1 and RUVBL2.^{22,23} We and others extensively studied RPAP3 and SPAG1 and their interactions with HSPs and RUVBLs in the context of the assembly processes of the small nucleolar ribonucleoproteins and of dynein arms.^{14,15,24–28} It has been proposed that HSPs carried out protein clients to their final destinations in these assembly pathways. For that purpose, HSPs interact with the so-called R2TP and R2SP complexes that both contained the RUVBL1 and RUVBL2 chaperones in combination, with the RPAP3/PIH1D1 or SPAG1/PIH1D2 heterodimers, the PIH-containing partner also being able to directly bind their own partners.^{22,29–34}

Previously, we demonstrated that first and third TPR domains in SPAG1 could recruit HSPs with similar affinities, whereas the second TPR was most likely unavailable for HSP binding, certainly because of its inner hydrophobicity. We deciphered with an atomic resolution the mode of binding of the third TPR domain to HSP peptides and questioned the direct GTPase activity that was initially proposed.^{15,21} In this study, we used protein engineering to solubilize the first TPR domain of SPAG1. We showed that an optimized TPR1 could mimic the wild-type version and obtained, via NMR, the precision required to finely inspect its interaction with the HSP C-terminal tails. Interestingly, TPR1 and TPR3 in SPAG1

display a distinct binding mode, especially for the HSP90 peptide. Our data are discussed with regard of the emerging interest for druggable TPR domains, with regard to known 3D structures and the functionality of the TPR domains in SPAG1.

MATERIALS AND METHODS

Homology Modeling. The wild-type amino acid sequence of the human SPAG1 protein was retrieved from the UniProt database (Q07617). The region of residues 206–327 (SPAG1-TPR1) was used for homology modeling with the I-TASSER Web server (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>).³⁵

Optimization of Protein Sequences. The best 3D model of SPAG1-TPR1 provided by I-TASSER was used as an entry in the PROSS Web server (<https://pross.weizmann.ac.il/step/pross-terms/>).⁶ Default parameters were used for protein sequence optimization.

Protein Expression and Purification. The optimized sequence of SPAG1-TPR1 was introduced into a pNEA-3CH plasmid dedicated to the overexpression of six-His-tagged recombinant proteins.³⁶ Briefly, a Ca²⁺ competent BL21(DE3) pRARE2 strain of *Escherichia coli* was transformed with the protein expression vector using a heat shock at 42 °C. Positive clones were selected on LB-agar plates supplemented with chloramphenicol and ampicillin. An isolated colony was grown in selective LB medium at 37 °C, overnight and under agitation. One liter of selective LB medium was seeded with the preculture and held at 37 °C under agitation until the optical density (OD) at 600 nm reached ~0.6. Overexpression of the recombinant protein was induced by addition of IPTG (isopropyl β -D-1-thiogalactopyranoside) at a final concentration of 0.2 mM. The culture was placed at 20 °C overnight under agitation.

Then, bacteria were harvested by centrifugation. The pellet was resuspended in lysis buffer [25 mM HEPES (pH 7.5), 300 mM NaCl, 0.5 mM TCEP, and 10 mM imidazole] and sonicated. After centrifugation, the supernatant was placed on TALON beads for 2 h at 4 °C under agitation. The beads were washed three times with lysis buffer. The elution of bound proteins was performed in lysis buffer by adding recombinant 3C PreScission protease directly on the beads. After cleavage, the non-native GPHM sequence remains at the N-terminal extremity of the recombinant protein. Finally, a gel filtration (GF) was performed in 10 mM NaPi (pH 6.4), 150 mM NaCl, and 0.5 mM TCEP with a Superdex-75 column. The GF fractions containing the protein of interest were pooled and concentrated to a protein concentration of 1 mM. The purification process was followed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE).

The same protocol was applied to obtain samples suitable for NMR spectroscopy, except that bacteria were grown in a minimal M9 medium supplemented with ¹⁵NH₄Cl and [¹²C]- or [¹³C]-d₆-glucose.

Dynamic Light Scattering (DLS). DLS experiments were performed at 20 °C in 10 mM NaPi (pH 6.4), 150 mM NaCl, and 0.5 mM TCEP on a Zetasizer Nano ZS instrument (Malvern Panalytical) in a small-volume quartz cell. Data were processed with Zetasizer software (Malvern) with default parameters.

Protein Co-expression Assays. Sequences of human HSP70 (Uniprot P0DMV8) and HSP90-MC (N domain-truncated HSP90, Uniprot P08238) proteins and of SPAG1-TPR1_{opt} were introduced into pNYK (for untagged proteins)

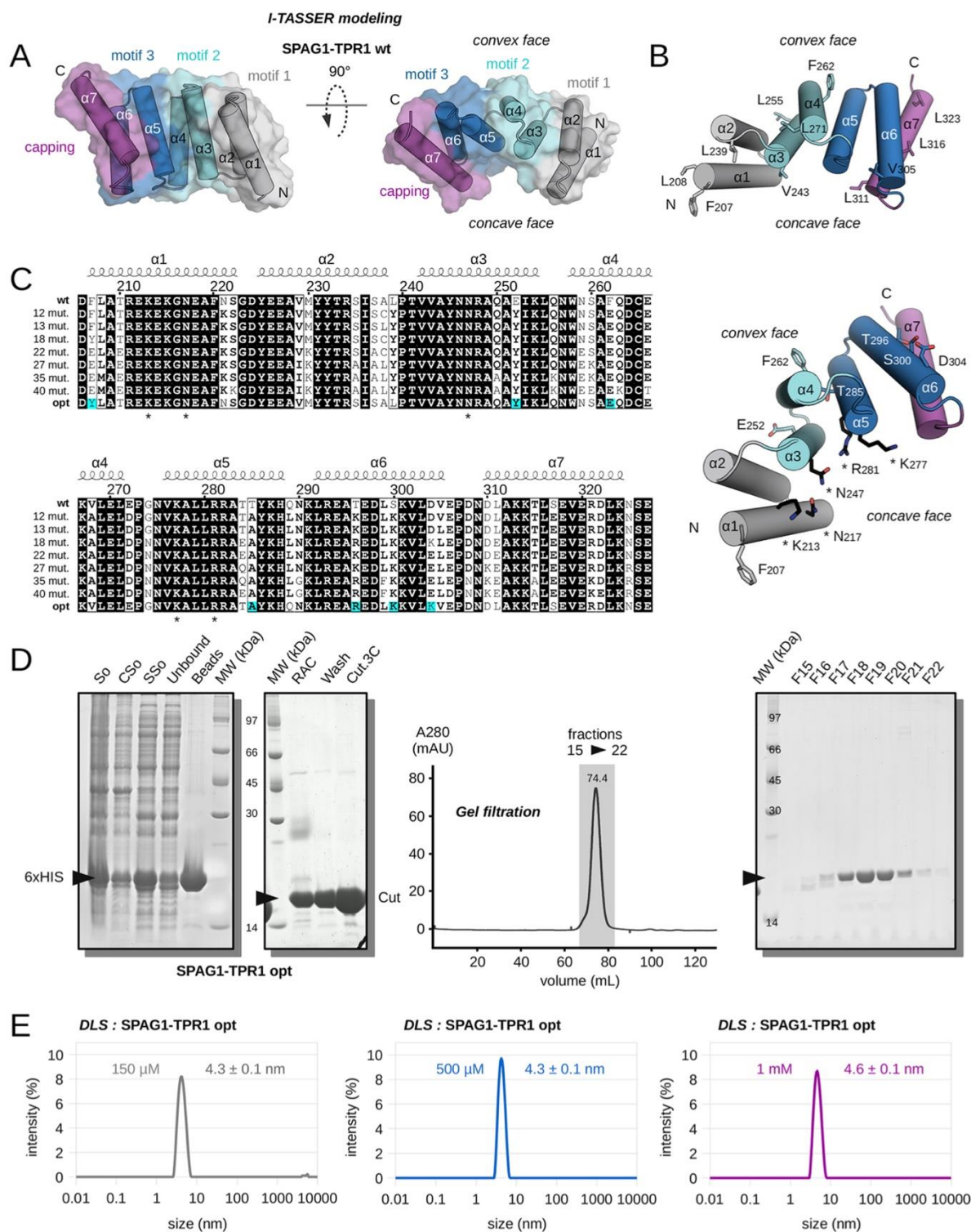


Figure 1. Optimizing the TPR1 domain in SPAG1. (A) Two views of the 3D structure of the region of residues 206–327 in SPAG1 obtained by homology with the I-TASSER software. TPR motifs and the C-terminal capping helix are highlighted using different colors. (B) Hydrophobic residues in SPAG1-TPR1_{wt} are represented using sticks on the 3D homology model. (C) Results of PROSS optimization. Sequences of the wild type, the PROSS proposals, and the in-house optimized SPAG1-TPR1 are aligned. Strictly conserved residues are in black boxes. Residues of the carboxylate-clamp motif are highlighted with an asterisk. Secondary structures coming from the 3D homology model are displayed on the top of the

C

<https://doi.org/10.1021/acs.biochem.1c00052>
Biochemistry XXXX, XXX, XXX–XXX

Figure 1. continued

alignment. On the right, carboxylate-clamp residues (black, with an asterisk) and residues selected for optimization in SPAG1-TPR1_{wt} are displayed using sticks on the 3D homology model. (D) Purification of recombinant SPAG1-TPR1_{opt}. On the SDS-PAGE profile displayed on the left, So, CSO, SSo, Unbound, Beads, MW, RAC, Wash, and Cut3C refer to the sonicate, the sonicate pellet, the sonicate supernatant, the flow-through of TALON beads, the TALON beads, the size marker, the TALON beads after 3C cleavage, the washed TALON beads after 3C cleavage, and the soluble fraction after 3C cleavage, respectively. The protein of interest is indicated using a black arrow. The gel filtration profile was obtained on a preparative Superdex-75 column. Fractions of interest are highlighted in a gray box. The maximal intensity measured at 280 nm is indicated. The SDS-PAGE gel on the right corresponds to the analysis of fractions 15–22 coming from gel filtration. (E) DLS analysis of SPAG1-TPR1_{opt} at 150, 500, and 1000 μM . The size corresponding to the highest intensity is reported.

and pNEA-3CH (for six-His-tagged proteins) vectors, respectively. The Ca²⁺ competent BL21(DE3) pRARE2 strain of *E. coli* was simultaneously transformed with the pNYK and pNEA vectors. Control experiments were performed with pNYK-transformed bacteria. Our protocol was previously described.¹⁵ Briefly, positive clones were selected on LB-agar plates, picked for preculture at 37 °C in selective LB medium, and grown in LB medium at 37 °C until the OD at 600 nm reached ~0.6. Overexpression of recombinant proteins was induced with 0.2 mM IPTG for one night at 20 °C. Cells were harvested, pelleted, and sonicated in the lysis buffer described above. Sonicates were centrifuged, and the supernatant was placed on TALON beads for 2 h at 4 °C under agitation. Proteins were eluted from the beads using lysis buffer supplemented with 300 mM imidazole. The purification process was followed on denaturant polyacrylamide gels.

Isothermal Titration Calorimetry (ITC). The binding of synthetic peptides DASRMEEVD (C-terminal amino acids of human HSP90) and SGPTIEEVD (C-terminal amino acids of human HSP70) to recombinant SPAG1-TPR1_{opt} was individually measured at 283 K on an ITC-200 microcalorimeter (Malvern) in 10 mM NaP_i (pH 6.4), 150 mM NaCl, and 0.5 mM TCEP. The peptide stock concentration was between 1.5 and 2 mM. The protein stock concentration was between 0.15 and 0.2 mM. The peptide was placed in the syringe and gradually added to the cell containing the protein solution. The dissociation constant (K_d), the stoichiometry (N), and the variations in enthalpy (ΔH), entropy (ΔS), and free energy (ΔG) were derived from the calorimetric data using the ORIGIN7 software (Malvern).

Nuclear Magnetic Resonance. NMR data were recorded at 293 K on a ¹³C- and ¹⁵N-labeled sample of SPAG1-TPR1_{opt} concentrated at 1 mM in 10 mM NaP_i (pH 6.4), 150 mM NaCl, and 0.5 mM TCEP using a 600 MHz spectrometer equipped with a cryoprobe. The usual two-dimensional (2D) and 3D NMR spectra were recorded. A majority of ¹H, ¹³C, and ¹⁵N resonances were assigned and deposited in the Biological Magnetic Resonance Bank as entry 34586.

The 3D NMR structure of SPAG1-TPR1_{opt} was calculated using distance restraints derived from ¹H–¹⁵N 3D NOESY-HSQC, ¹H–¹³C 3D NOESY-HSQC, and ¹H–¹H 2D NOESY experiments recorded with a mixing time of 120 ms, and dihedral restraints derived from TALOS-N and PREDITOR.^{37,38} The automated procedure of CYANA 3.0 was used to iteratively assign NOE cross-peaks.³⁹ After careful inspection of the final CYANA step, the distance and dihedral restraints were used to calculate 100 water-refined structures with the AMBER web portal.⁴⁰ The 20 structures with the lowest restraint energy were selected as representative structures of SPAG1-TPR1_{opt} and deposited in the Protein Data Bank as entry 7BEV.

Samples containing 1 mM ¹³C- and ¹⁵N-labeled SPAG1-TPR1_{opt} and 3 mM DASRMEEVD or SGPTIEEVD peptide were prepared in 10 mM NaP_i (pH 6.4), 150 mM NaCl, and 0.5 mM TCEP. In addition to classical 2D and 3D NMR spectra, 2D- and 3D-filtered NOESY data with a mixing time of 120 ms were collected at 600 MHz and 293 K, to highlight intermolecular cross-peaks. To evaluate the chemical shift perturbations (CSPs) induced by binding of the HSP peptide to labeled SPAG1-TPR1_{opt}, the variation of the composite ¹H–¹⁵N chemical shift was calculated, for each backbone amide group, as follows: $\Delta\delta = \sqrt{0.1(\delta^{15}\text{N})^2 + (\delta^1\text{H})^2}$ (where $\delta^{15}\text{N}$ and $\delta^1\text{H}$ are the chemical shift differences in the ¹⁵N dimension and the ¹H dimension, respectively, between the free and bound states of SPAG1-TPR1_{opt}).

Modeling of the Peptide:SPAG1-TPR1_{opt} Complexes. We used the best NMR structure of SPAG1-TPR1_{opt} and 10 structures of the DASRMEEVD and SGPTIEEVD peptides to model complex structures using HADDOCK-2.2 and CNS-1.3.⁴¹ 3D structures of the two peptides were extracted from the 3D NMR structure of RPAP3-TPR2 in complex with DTSRMEEVD and SGPTIEEVD peptides. The threonine in the HSP90 peptide was mutated into alanine using PyMOL.⁴² To drive the docking, we defined active and passive residues to build ambiguous intermolecular restraints (AIRs). In SPAG1-TPR1_{opt}, active residues are solvent-exposed residues that (i) are located on the concave face of the structure and (ii) belong to the 25% most perturbed residues in NMR titration experiments (i.e., with $\Delta\delta$ values > 75th percentile values). Passive residues are surface residue neighbors of the active residues. All residues in the peptides were defined as active. Intermolecular NOEs were turned into unambiguous restraints. Additional unambiguous distance restraints were added between residues of the carboxylate clamp in SPAG1-TPR1_{opt} and the C-terminal part of the peptides. The automated mode was chosen for the definition of the semiflexible interfaces; 2560 and 256 structures were calculated in the it0 rigid-body and the it1 semiflexible simulated annealing steps, respectively. The last 256 structures were refined in water. The docking solutions were clustered using the FCC (fraction of common contacts) algorithm with a cutoff value of 0.75.⁴³ The 3D structures of the complexes were ranked according to their HADDOCK score.

RESULTS

Sequence Optimization of SPAG1-TPR1 Greatly Enhances Its Solubility. Previously, we showed that the human SPAG1 protein contained three TPR domains.¹⁵ When produced in an isolated form, two of them were soluble, meaning the first and third domains, the second domain being not soluble. Unfortunately, due to aggregation features appearing at a protein concentration of >150 μM , we were

D

<https://doi.org/10.1021/acs.biochem.1c00052>
Biochemistry XXXX, XXX, XXX–XXX

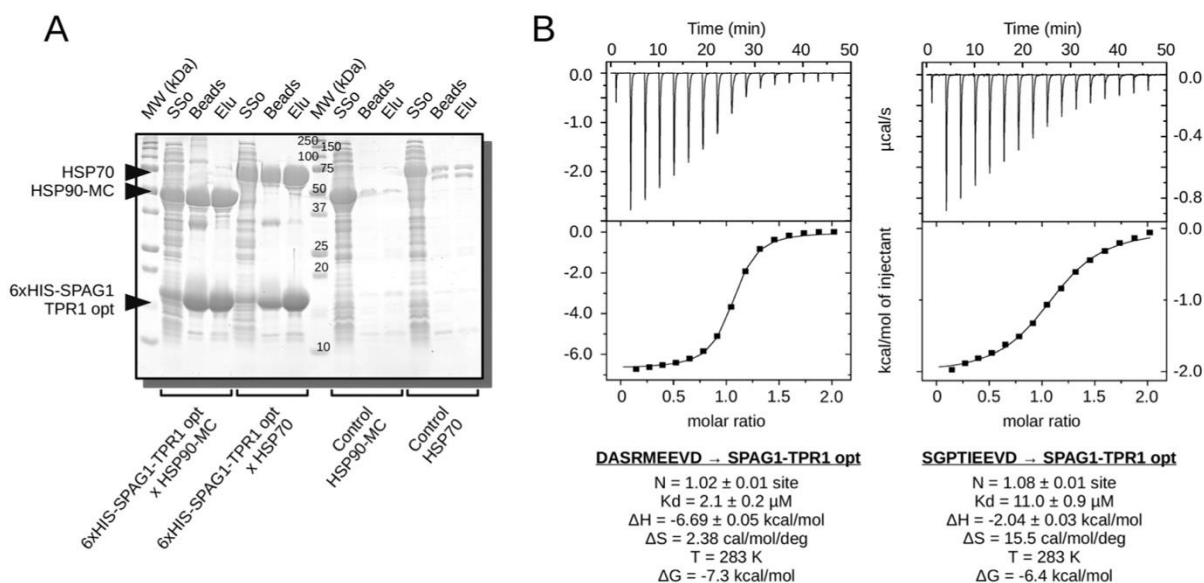


Figure 2. SPAG1-TPR1_{opt} interacts with HSPs. (A) Co-expression assays between SPAG1-TPR1_{opt} and HSP90-MC or HSP70. On the SDS–PAGE profile, SSo, Beads, Elu, and MW refer to the supernatant sonicate, the TALON beads, the soluble fraction after elution with imidazole buffer, and the size marker, respectively. Test experiments are shown on the left part of the gel, and control experiments on the right part. (B) ITC experiments recorded at 283 K with SPAG1-TPR1_{opt} and DASRMEEVD or SGPTIEEVD peptides in 10 mM NaPi (pH 6.4), 150 mM NaCl, and 0.5 mM TCEP. The protein stock concentration was between 0.15 and 0.2 mM. The peptide stock concentration was between 1.5 and 2 mM. The peptide was placed in the syringe and was gradually added to the cell containing the protein solution. *N*, *K_d*, *T*, ΔH , ΔS , and ΔG refer to the stoichiometry, the dissociation constant, the temperature, and the variations in enthalpy, entropy, and free energy, respectively.

not able to perform an in-depth analysis of SPAG1-TPR1, which has however been shown able to bind HSP70 and HSP90 through their C-terminal tail.¹⁵ Indeed, wild-type SPAG1-TPR1 (SPAG1-TPR1_{wt}) possesses residues K213, N217, N247, K277, and R281 that form a conserved carboxylate-clamp motif, which is the signature for the binding of HSP70 and HSP90 tails. Thus, we thought that study at the atomic level could provide new insights into the molecular properties of this protein domain.

As the TPR domain is a well-characterized domain in the literature, we used the I-TASSER homology modeling Web server to have a first, but potentially strong, preview of the 3D structure of SPAG1-TPR1_{wt}.³⁵ Using the amino acid sequence of the region of residues 206–327 of the human SPAG1 protein as the sole source of information, we obtained a 3D model with a *C* score value of 0.66 (Figure 1A–C). This value is sufficiently high to consider the modeling as reliable. Interestingly, the best threading 3D template was SPAG1-TPR3 [Protein Data Bank (PDB) entry 6I57]. As expected, the prediction consists of seven α -helices folded into three TPR motifs followed by a final α -helix that caps the sixth α -helix (Figure 1A). The global shape of the protein domain displays a convex face and a concave face, where carboxylate-clamp residues are available for HSP binding. Analysis of the predicted structure revealed solvent-exposed hydrophobic amino acids, especially residues F207, L208, L239, V243, L255, F262, L271, V305, L311, and L323, which could promote protein aggregation through uncontrolled intermolecular hydrophobic contacts (Figure 1B).

Next, considering this prediction, we used the PROSS Web server with the idea of designing an optimized sequence suitable for recombinant expression of SPAG1-TPR1 in *E. coli*.⁶ For this purpose, PROSS requires a 3D structure and an

amino acid sequence alignment. On the basis of the I-TASSER 3D structure of SPAG1-TPR1_{wt}, the PROSS algorithm returned seven primary sequences, starting with 12 mutations up to 40 mutations with respect to the wild-type sequence (Figure 1C). We previously demonstrated that not only α -helices 1, 3, and 5 (that hold the carboxylate-clamp residues) but also the region at the root of α -helix 7 could be involved in the recognition of HSP peptides in RPAP3-TPR2 or SPAG1-TPR3.^{14,15} In light of these data, we rationally designed an amino acid sequence of SPAG1-TPR1 based on the PROSS proposals, but only for residues distant from the carboxylate-clamp residues and located on the convex face of the domain. In addition, we tried to restrict the number of mutations introduced in the wild-type sequence to minimize their potential impact on the global features of the protein domain. Thus, we constructed a final sequence with the following seven mutations: F207Y, E252Y, F262E, T285A, T296R, S300K, and D304K (Figure 1C). This optimized sequence was introduced in a plasmid dedicated to protein overexpression in bacteria. We showed that our optimized SPAG1-TPR1 (SPAG1-TPR1_{opt}) was efficiently purified from *E. coli* (Figure 1D). Interestingly, with a mean size of 4.6 ± 0.1 nm, DLS showed that this recombinant domain remains monomeric in solution at a protein concentration of 1 mM (Figure 1E). This result strongly contrasts with the DLS experiments performed on SPAG1-TPR1_{wt} (i.e., wild-type SPAG1-TPR1), for which a mean size of ~ 80 nm was previously measured at a protein concentration of >150 μ M,¹⁵ and demonstrates the stabilization effect provided by the point mutations.

Optimized and Wild-Type SPAG1-TPR1 Behave Similarly. Because of protein sequence optimization, we obtained a highly soluble sample of SPAG1-TPR1. However,

E

https://doi.org/10.1021/acs.biochem.1c00052
 Biochemistry XXXX, XXX, XXX–XXX

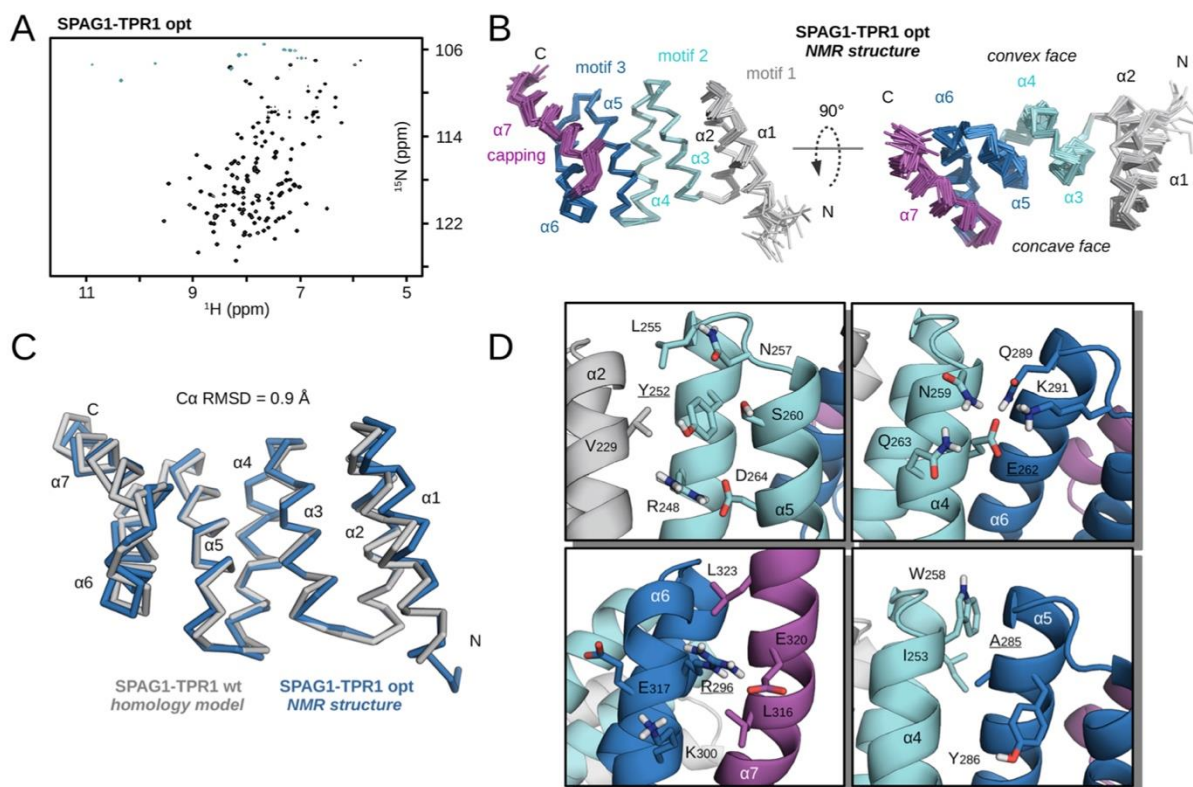


Figure 3. NMR solution structure of SPAG1-TPR1_{opt}. (A) ¹H–¹⁵N HSQC spectrum of SPAG1-TPR1_{opt} recorded at 600 MHz and 293 K in 10 mM NaPi (pH 6.4), 150 mM NaCl, and 0.5 mM TCEP. Folded resonances are colored blue. (B) Two backbone representations of the 20 best NMR structures after AMBER water refinement. The three TPR motifs and the C-terminal capping helix are highlighted. (C) Superimposition of the SPAG1-TPR1_{opt} NMR structure with the 3D homology model of SPAG1-TPR1_{wt}. (D) Structural close-up on the region of residues Y252, E262, R296, and A285 in SPAG1-TPR1_{opt}.

one could wonder whether this domain could reproduce the behavior of the wild-type domain.

To answer this question, we first performed protein co-expression assays between His-tagged SPAG1-TPR1_{opt} and human HSP70 or HSP90-MC (i.e., HSP90 with its N-terminal domain truncated). As observed for wild-type SPAG1-TPR1_{wt} we showed that SPAG1-TPR1_{opt} co-elutes specifically with HSPs (Figure 2A). We next performed isothermal titration calorimetry (ITC) experiments to access the thermodynamics of the interaction between SPAG1-TPR1_{opt} and HSP peptides (DASRMEEVD for HSP90 and SGPTIEEVD for HSP70) (Figure 2B). For the two peptides, we obtained stoichiometry values close to 1:1. The dissociation constants (K_d) were 2.1 ± 0.2 and $11.0 \pm 0.9 \mu\text{M}$ for HSP90 and HSP70, respectively. These ITC data were similar to the ITC data obtained with SPAG1-TPR1_{wt} for which we measured, under the same experimental conditions, 1:1 stoichiometries and K_d values of $2.5 \pm 0.2 \mu\text{M}$ and 15.8 ± 2.6 for HSP90 and HSP70 peptides, respectively.¹⁵ More, for each peptide, values of variations in enthalpy (ΔH), entropy (ΔS), and free energy (ΔG) are almost identical regardless of the domain tested.¹⁵ Altogether, these biochemical and biophysical data strongly support the fact that optimized and wild-type SPAG1-TPR1s behave similarly *in vitro*. Thus, they allowed us to go deeper in the structural analysis of SPAG1-TPR1.

Structural Analysis of SPAG1-TPR1_{opt}. We used NMR in solution to obtain experimental data for SPAG1-TPR1_{opt} at the atomic scale. We produced a ¹³C- and ¹⁵N-labeled SPAG1-TPR1_{opt} at a concentration of 1 mM and recorded a set of 2D and 3D spectra to obtain its resonance assignment.

In agreement with DLS data, the ¹H–¹⁵N HSQC spectrum displays the features of a monomeric and well-folded protein domain. Indeed, the cross peaks observed on this 2D spectrum are thin, with homogeneous intensities, and well-dispersed in the two dimensions (Figure 3A). Analysis of the 3D spectra permitted us to assign a large majority of ¹H, ¹³C, and ¹⁵N resonances of SPAG1-TPR1_{opt} and allowed us to access the 3D structure of the protein domain. We obtained a well-defined ensemble of 20 3D structures of SPAG1-TPR1_{opt} (for statistics, see Table 1). Our experimental structure reveals that SPAG1-TPR1_{opt} folds into a TPR domain with a seventh capping α -helix (Figure 3B), as supported by the HSP binding assays. Here again, the concave face, which is made of α -helices 1, 3, 5, and 7, exposes the five residues of the carboxylate clamp, whereas a convex face is made of α -helices 2, 4, and 6. With a root-mean-square deviation (RMSD) of 0.9 Å measured on C α atoms, this structure superimposed well with the I-TASSER model generated with the wild-type sequence of SPAG1-TPR1 (Figure 3C).

Next, we naturally focused on the optimized positions to better understand their role in the stabilization of the domain.

F

<https://doi.org/10.1021/acs.biochem.1c00052>
Biochemistry XXXX, XXX, XXX–XXX

Table 1. NMR-Derived Restraints and Structural Statistics for the 20 Best Water-Refined Structures of SPAG1-TPR1_{opt}

NMR distances and dihedral constraints	
total no. of NOEs	2919
short-range ($ i - j \leq 1$)	1357
medium-range ($ i - j < 5$)	904
long-range ($ i - j \geq 5$)	658
total no. of dihedral angle restraints	239
φ	109
ψ	110
χ_1	26
structure statistics	
no. of violation occurrences	
distance constraints ($>0.3 \text{ \AA}$)	0
dihedral angle constraints ($>5^\circ$)	0.15 ± 0.37
RMSD from the best structure (\AA)	
all backbone atoms	0.92 ± 0.20
all heavy atoms	1.28 ± 0.14
backbone atoms in secondary structures	0.51 ± 0.11
heavy atoms in secondary structures	1.12 ± 0.11
Ramachandran statistics (%)	
residues in the most favored regions	91.3
residues in additional allowed regions	8.7
residues in generously allowed regions	0
residues in disallowed regions	0

For position 207, the tyrosine residue, which is quite solvent-exposed, brings polarity to the side chain without changing the hydrophobic nature of the phenol ring coming from the initial phenylalanine. Positions 252 and 262 were of particular interest because our mutations drastically changed the chemical feature of the amino acid side chain. In SPAG1-TPR1_{opt}, Y252 is framed by V229, L255, and R248 (Figure 3D). Moreover, its hydroxyl group brings surface polarity. On the contrary, E262 in SPAG1-TPR1_{opt} promotes stable polar contacts with N259, Q263, Q289, and K291. Obviously, this polar environment appeared not favorable to host the phenylalanine present in the wild-type sequence. At position 285, the conversion of the threonine to an alanine residue, in a hydrophobic context supported by I253, W258, and Y286, cancels the polar feature of the hydroxyl group. With regard to position 296, introducing an arginine residue permitted us (i) to efficiently connect helices α_6 and α_7 through a salt bridge with E320 and (ii) to limit exposure of L316 and L323 to solvent. Finally, lysine residues introduced at positions 300 and 304 enhances the general solubility because of their side chain amine group. Thus, our NMR structure fully explains the high stability of SPAG1-TPR1_{opt}.

SPAG1-TPR1_{opt} Does Not Bind HSP C-Terminal Tails as SPAG1-TPR3 Does. Having determined the 3D structure in solution of SPAG1-TPR1_{opt} and demonstrated that it could nicely mimic the wild-type domain, we aimed to gain more precise information about the interaction with HSP90 and HSP70 C-terminal tails.

To this end, we mapped the footprint of HSP C-terminal tails on the SPAG1-TPR1_{opt} 3D structure by analyzing the perturbations of the ^1H - ^{15}N composite chemical shifts induced upon formation of the protein:peptide complex (CSP). This analysis revealed that α -helices 1, 3, and 5 as well as the α_6 - α_7 bridging region at the root of the last α -helix in SPAG1-TPR1_{opt} appear to be involved in the binding of the two peptides (Figure 4). Analysis of filtered NOESY spectra

revealed a common pattern of intermolecular NOEs in the complexes with the two peptides (Figure 4). Indeed, Met in HSP90, but also Ile in HSP70, faced I253, T284, and L311 in SPAG1. Ala in HSP90 was found close to the N309 region in SPAG1, while no intermolecular NOE could have been highlighted in this region with the HSP70 peptide (Figure 4). These NMR data reminded us the binding features observed in HSP-bound RPAP3-TPR2.¹⁴ Indeed, in addition to comparable CSP profiles obtained upon HSP peptide binding, RPAP3-TPR2 and SPAG1-TPR1_{opt} share patterns of intermolecular NOEs involving conserved positions on the peptides and on the protein domains (Figure S1). This strongly suggests that HSP peptides adopt similar conformations upon binding of RPAP3-TPR2 and SPAG1-TPR1_{opt}.

To have 3D views of these interactions, we modeled the protein:peptide complexes in a HADDOCK protocol. Thus, as entries, we used the best SPAG1-TPR1_{opt} NMR structure and 10 distinct structures of the HSP peptides extracted from the complexes with RPAP3-TPR2 (PDB entries 6FDT and 6FDP). To drive the docking, we used our ambiguous and unambiguous restraints derived from NMR data. To ensure the orientation of the peptide, we added distances locking the EEVD parts of the peptides next to the conserved carboxylate-clamp region that were shown to be essential in TPR domains for efficient HSP binding (Figure S2).^{13,14,16,27} The results of the calculations are summarized in Table 2.

We obtained, for each complex, final 3D solutions that all cluster in one group, meaning that all of the structures share 75% of common contacts. For the two complexes, the five best HADDOCK structures superimposed well, and the positioning of the peptides, which both adopt a horseshoe conformation, nicely matches with the NMR CSP footprints and intermolecular NOEs (Figure 5A,B). Due to their importance in co-chaperone selectivity, we focused our attention on the Met and Ile residues in the HSP peptides. These 3D models reveal that these two residues are both hosted by residues Q250, I253, L280, T284, and L311. We compared our 3D data with those we obtained with the third TPR in SPAG1 known to be the other module that can bind HSPs in the full-length protein.

We noticed a clear difference in the positioning of the Met side chain in the HSP90 peptide (Figure 5C). The T284L substitution in SPAG1-TPR3 may prevent, by steric hindrance, a deep hosting of the Met methyl group close to this position. The Y residue in SPAG1-TRP3, replacing L280 in SPAG1-TPR1_{opt}, seems to act as a barrier, which, in addition to the L311E and N309I substitutions that displace hydrophobicity at the bottom of SPAG1-TPR3, promotes opposite orientations of the Met side chain in the two complexes. With regard to the HSP70 peptide, its Ile residue is bulkier, more hydrophobic, and less plastic than the Met in HSP90. L280 in SPAG1-TPR1_{opt} leaves enough space to host the two methyl groups of Ile in a quite hydrophobic environment spreading from T284 to V276 (Figure 5C). For that reason, the positioning of Ile in the two complexes is not as different as in the case of the Met in HSP90. Thus, even if TPR1 and -3 recruit HSPs with similar affinities, they take distinct structural paths to bind their C-terminal ends, especially for HSP90.

DISCUSSION

Optimization of the TPR Domain. In vitro techniques used for atomistic structural studies are usually dependent on the quality of the samples. De facto, not all proteins can be

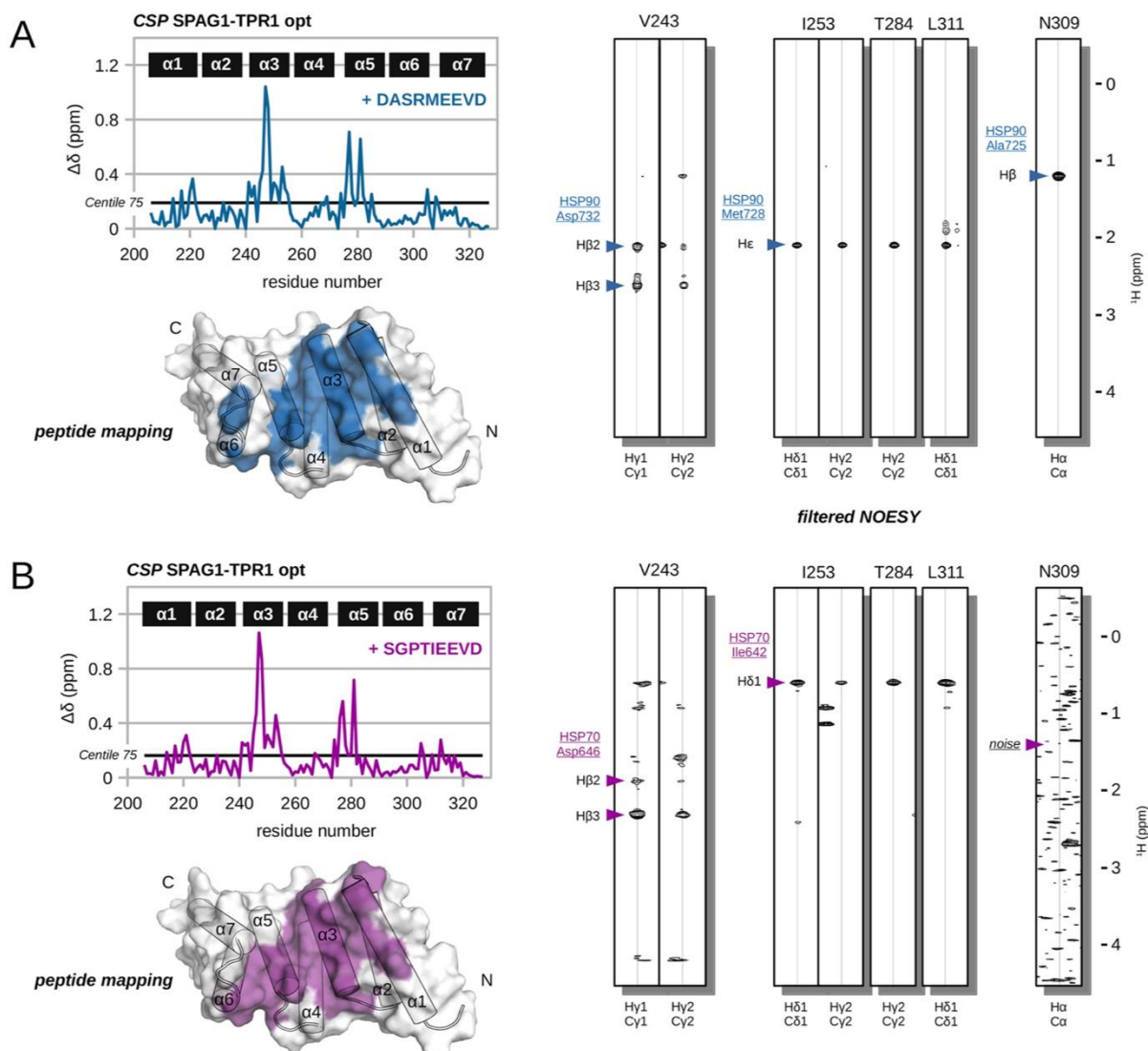


Figure 4. NMR analysis of SPAG1-TPR1_{opt} upon HSP peptide binding. (A) Chemical shift perturbation (CSP) profiles of SPAG1-TPR1_{opt} upon binding of the DASRMEEVD peptide. The ^1H - ^{15}N chemical shift variation ($\Delta\delta$) for each residue is plotted vs the sequence number. Positions of α -helices in SPAG1-TPR1_{opt} are represented on the diagram. The 75th percentile of $\Delta\delta$, which discriminates the most perturbed residues, is indicated as a line. The residues displaying $\Delta\delta$ values higher than the 75th percentile value are colored on a molecular surface representation of SPAG1-TPR1_{opt}. On the right, intermolecular NOE cross peaks are represented on strips extracted from a filtered 3D ^1H - ^{13}C NOESY-HSQC spectrum. (B) Same as panel A but for the SGPTIEEVD peptide. CSP data were recorded at 600 MHz and 293 K in 10 mM NaP₁ (pH 6.4), 150 mM NaCl, and 0.5 mM TCEP using a peptide:protein ratio of 3.

analyzed due to misbehavior, according to the experimenter's expectations, during the expression or purification process. Here, we used protein engineering to solubilize TPR1 of SPAG1.

The TPR domain is a versatile domain that has been closely studied because it mediates protein/protein interaction, especially with HSPs if the carboxylate-clamp signature is present in the sequence. The amino acids surrounding the structuring residues required for establishing the TPR motif could display a high degree of heterogeneity.^{7,9} As we demonstrated, this heterogeneity of sequence could lead to protein insolubility in SPAG1-TPR2.¹⁵ By mutating <6% of the amino acid sequences, we obtained a soluble and highly

stabilized SPAG1-TPR1 *in vitro*. Remarkably, the SPAG1-TPR1_{opt} sample was stable for more than one year at 4 °C in the NMR buffer. This demonstrates that clever modifications in TPR domains allow their structural study alone or in complex with biological partners.

One great interest of the optimization approach is to support studies on the druggability of proteins. Indeed, *in silico* methods are often used to screen a plethora of ligands toward protein targets.⁴⁴ The robust 3D structure of the target is required for an efficient screening. Enhancing the solubility of the target will facilitate the resolution of structures, which can directly be used in bioinformatic binding assays or can first serve as a starting template for homology modeling of the wild-

Table 2. Summary of Restraints Used in HADDOCK and Structural Statistics for the Five Best Structures of SPAG1-TPR1_{opt} in Complex with the DASRMEEVD and SGPTIEEVD Peptides

	SPAG1-TPR1 _{opt} :DASRMEEVD		SPAG1-TPR1 _{opt} :SGPTIEEVD	
	SPAG1-TPR1 _{opt}	DASRMEEVD	SPAG1-TPR1 _{opt}	SGPTIEEVD
active residues				
total no.	9	9		9
residues	214, 217, 220, 221, 243, 254, 277, 305, 308	all	214, 217, 220, 221, 243, 254, 277, 305, 306	all
passive residues				
total no.	20	0	20	0
residues	210, 213, 216, 225, 240–242, 246, 247, 250, 251, 275, 276, 278, 280, 281, 306, 307, 309, 310		210, 213, 216, 225, 240–242, 246, 247, 250, 251, 275, 276, 278, 280, 281, 306–309	
no. of AIRs (ambiguous intermolecular restraints)	9	9	9	9
no. of UIRs (unambiguous intermolecular restraints)		27		24
no. coming from intermolecular NOEs		17		14
statistics of the top 5 structures				
HADDOCK score		−114.959 ± 6.842		−103.447 ± 1.751
AIR violation (>0.3 Å)		0.4 ± 0.5		0
UIR violation (>0.3 Å)		0		0
RMSD from the best structure (Å)				
all backbone atoms		0.74 ± 0.15		0.67 ± 0.05
all heavy atoms		1.07 ± 0.13		0.95 ± 0.07

type protein. It also permits bench scale experiments, this step being crucial to test the real binding activity of computational ligands. Moreover, stabilized samples also appear to be essential for testing the impact of the ligand on the activity of the target, *in vitro*. Interestingly, using combined *in silico* and biochemical approaches, the TPR2A domain of the protein Hop was turned into a druggable target.⁴⁵ Indeed, authors designed an organic ligand that binds the HSP90 recognition site through contacts with the carboxylate-clamp residues. Similarly, Dictyoceratin-A and -C are two inhibitor molecules that bind the TPR1 region of RPAP3.⁴⁶ These studies open the door to a fine-tuning of the HSP90 activity via the selective blocking of its different adaptive regulators. Our novel data on SPAG1-TPR1_{opt} in complex with DASRMEEVD and SGPTIEEVD peptides improve our understanding of the binding mode of TPR domains containing a carboxylate-clamp signature and could contribute to the development of novel molecular compounds with specific modulator activity on HSP70 and/or HSP90 co-chaperones.

Because fine-tuning of the HSP activity goes hand in hand with fine knowledge of the interaction modes of the co-chaperones, we compared our 3D data with known structures available in the Protein Data Bank and in our previous studies.

Sequence Signature in the TPR Domain Might Forecast the Mode of Binding of the Last Methionine in HSP90.

With regard to the interaction between the C-terminal tail of HSP90 and TPR domains, we observed that the methionine of the MEEVD motif could adopt four main orientations, defining four families of structures (Figure 6A,B). In one of them, the Met residue points toward the solvent and does not seem to be involved in the anchoring of the peptide to the TPR domain [family 0 (Figure 6A)]. The three remaining groups define three distinct interaction spots for the methionine at the surface of the TPR domains (Figure 6B,C). On a structural alignment, we looked at primary sequences in

these patches with the aim of highlighting residues that could be the signature of one interaction mode or the other.

Family 1 of structures (Figure 6A) contains the first ever described 3D structure of an HSP90 peptide in complex with a TPR domain.¹³ In this group, the peptide is quite linear, spreading from the top to the bottom of the TPR surface. It distinguishes this family from families 2 and 3 in which the peptide adopts superimposable conformations looking like a horseshoe, extended from the right to the left of the TPR (Figure 6A). For TPR domain orientation, see Figure 6. Remarkably, the HSP90 peptides used in the structures of the family 1 are very short, as it contains only the last five MEEVD residues. On the contrary, for 3D determination of structures included in families 2 and 3, longer peptide sequences were used. The interaction spot in family 1 (spot 1) seems to be quite well conserved in families 2 and 3 (Figure 6C), making it difficult to highlight specific positions for family 1. Thus, one could wonder how the length of the peptide influences its positioning. Interaction spot 2 in family 2 looks quite heterogeneous in terms of amino acid sequences, and the corresponding residues in families 1 and 3 reproduce this heterogeneity. Nonetheless, our detailed inspection of these structural data drew our attention to position 3 in interaction spot 3, where a Thr residue is strictly conserved in structures belonging to family 3. This Thr is replaced by an Asn in family 1 and, in family 2, by bulkier residues (Leu or Lys) or a Ser, which is more hydrophilic than Thr. This spot 3, quite hydrophobic in family 3, appears to be more susceptible to hosting the methyl group of the HSP90 methionine than in the other families. From our analysis, we propose that this Thr residue, which corresponds to Thr284 in optimized SPAG1-TPR1_{opt}, may be a key residue that signals the positioning of the HSP90 methionine close to it. TPR domains that lack this Thr could be direct to other families, but clear sequence determinants still escape our analysis.

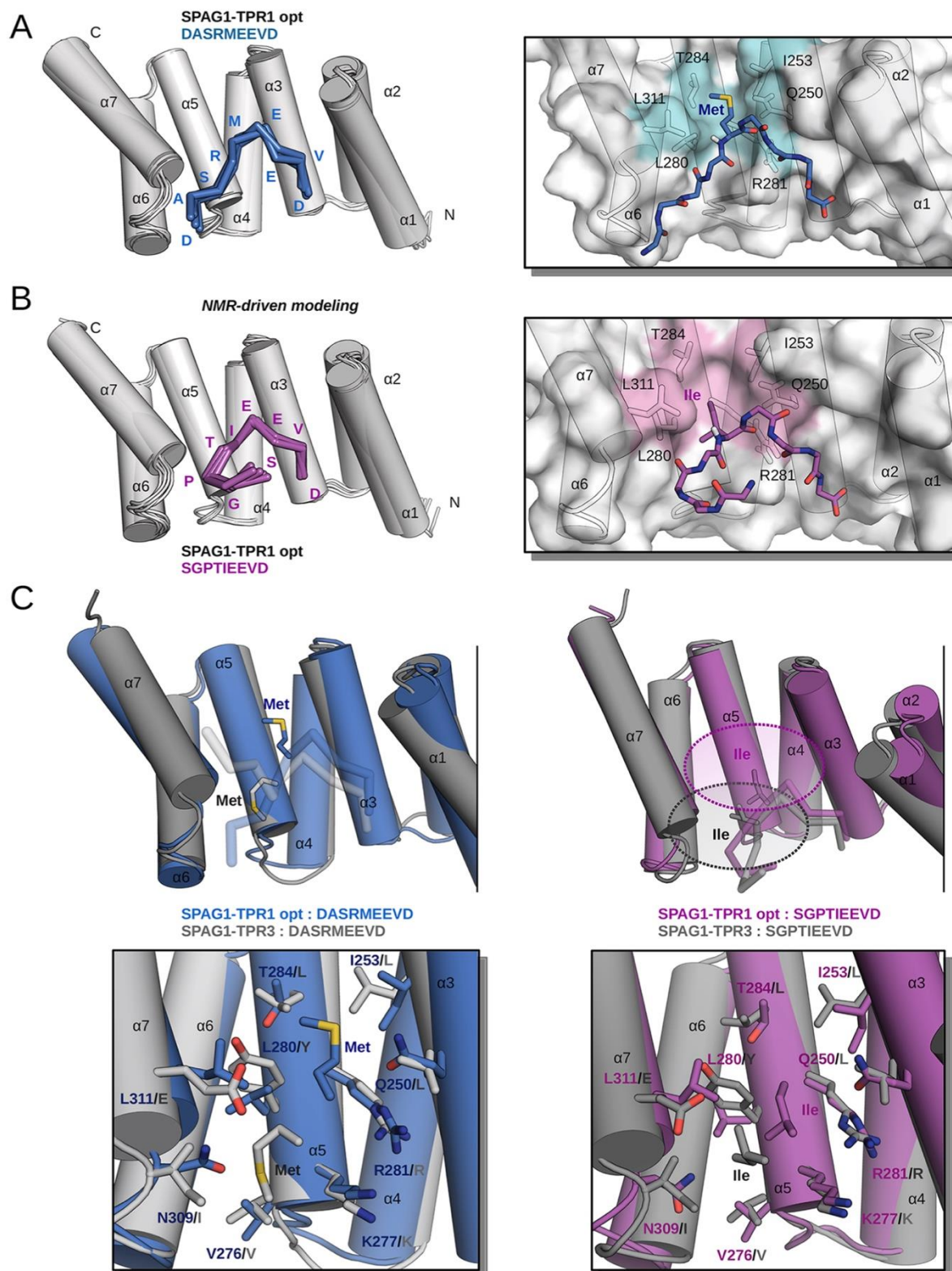


Figure 5. HADDOCK modeling of SPAG1-TPR1_{opt} in complex with HSP peptides. (A and B) Five best 3D models (i.e., presenting the best HADDOCK scores) of SPAG1-TPR1_{opt} in complex with DASRMEEVD and SGPTIEEVD, respectively, are superimposed and presented on the

J

<https://doi.org/10.1021/acs.biochem.1c00052>
Biochemistry XXXX, XXX, XXX-XXX

Figure 5. continued

left. The peptide is represented using its backbone trace. On the right are detailed views of the methionine (A) and the isoleucine (B) environments using the best HADDOCK structure of each complex. (C) Superimposition of the HSP-bound states of SPAG1-TPR1_{opt} and SPAG1-TPR3. For SPAG1-TPR1_{opt} the best HADDOCK structures were used (blue or magenta). For SPAG1-TPR3, representative structures coming from the work of Chagot and colleagues were used (light or dark gray).¹⁵ Detailed views of the environment of the methionine and the isoleucine in HSP peptides are presented. The numbering of SPAG1-TPR1_{opt} was used. Residues in the SPAG1-TPR1_{opt} and SPAG1-TPR3 complexes are labeled in blue/magenta and gray, respectively.

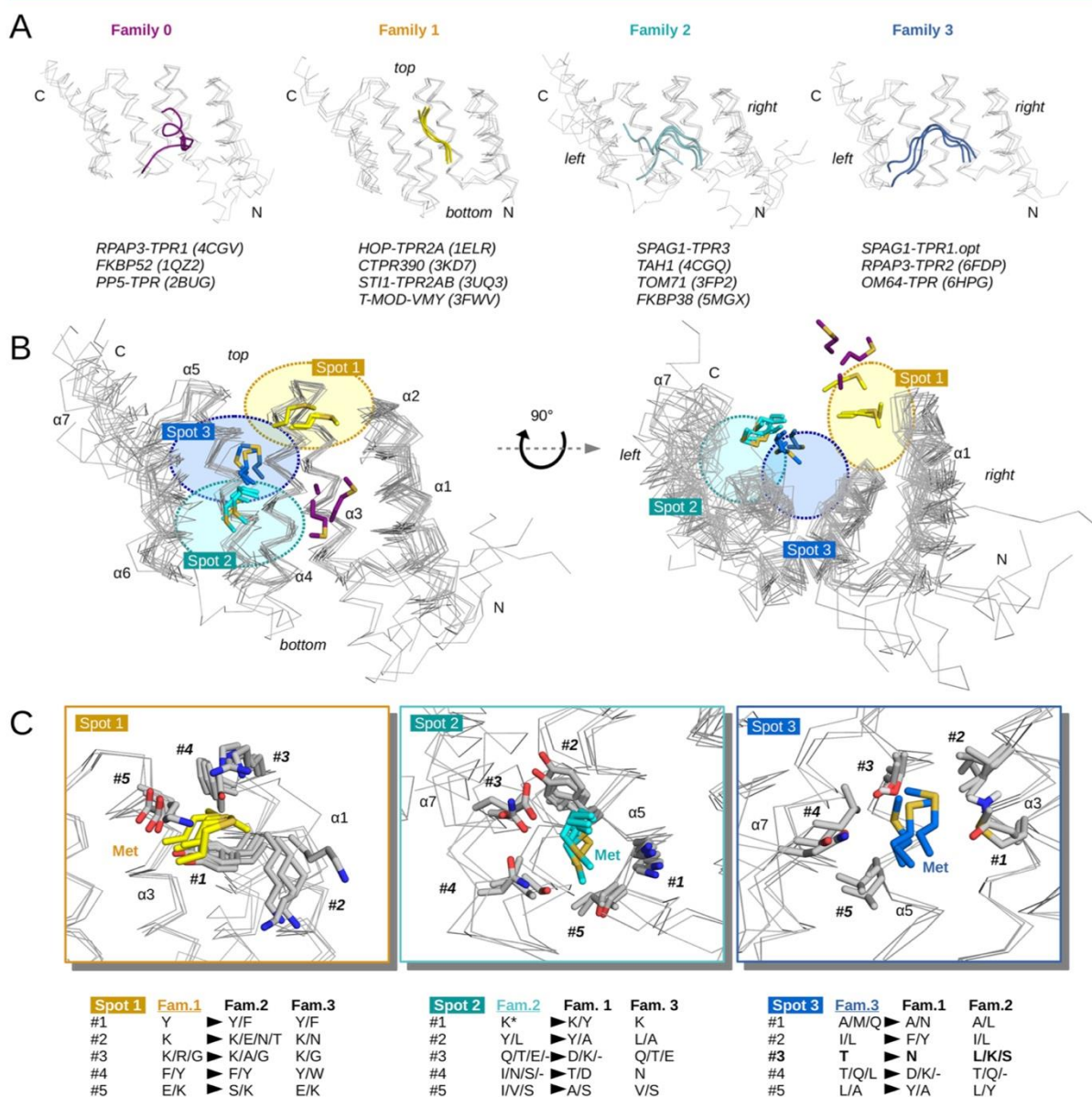


Figure 6. Structural comparison of TPR domains in complex with HSP90 peptides. (A) Families 0–3 of superimposed 3D structures are represented using the backbone traces of the TPR domain and cartoon representations of the HSP90 peptides. Arbitrary orientations of the structures are specified. When available, the name of the TPR protein and the PDB code are indicated. (B) All of the 3D structures displayed in panel A are superimposed, and for each of them, the methionine in the HSP90 peptide is represented using sticks. The three main spots of interaction involving this methionine are highlighted with colored circles. Two views are proposed. (C) Detailed views of the binding spots of each family. To look at sequence conservation on the structural alignment, five positions (#1–#5) are defined at each spot. The amino acid that occupies each of these positions in all of the structures is reported, structures being sorted by family. The residues belonging to the carboxylate-clamp motif are denoted with an asterisk.

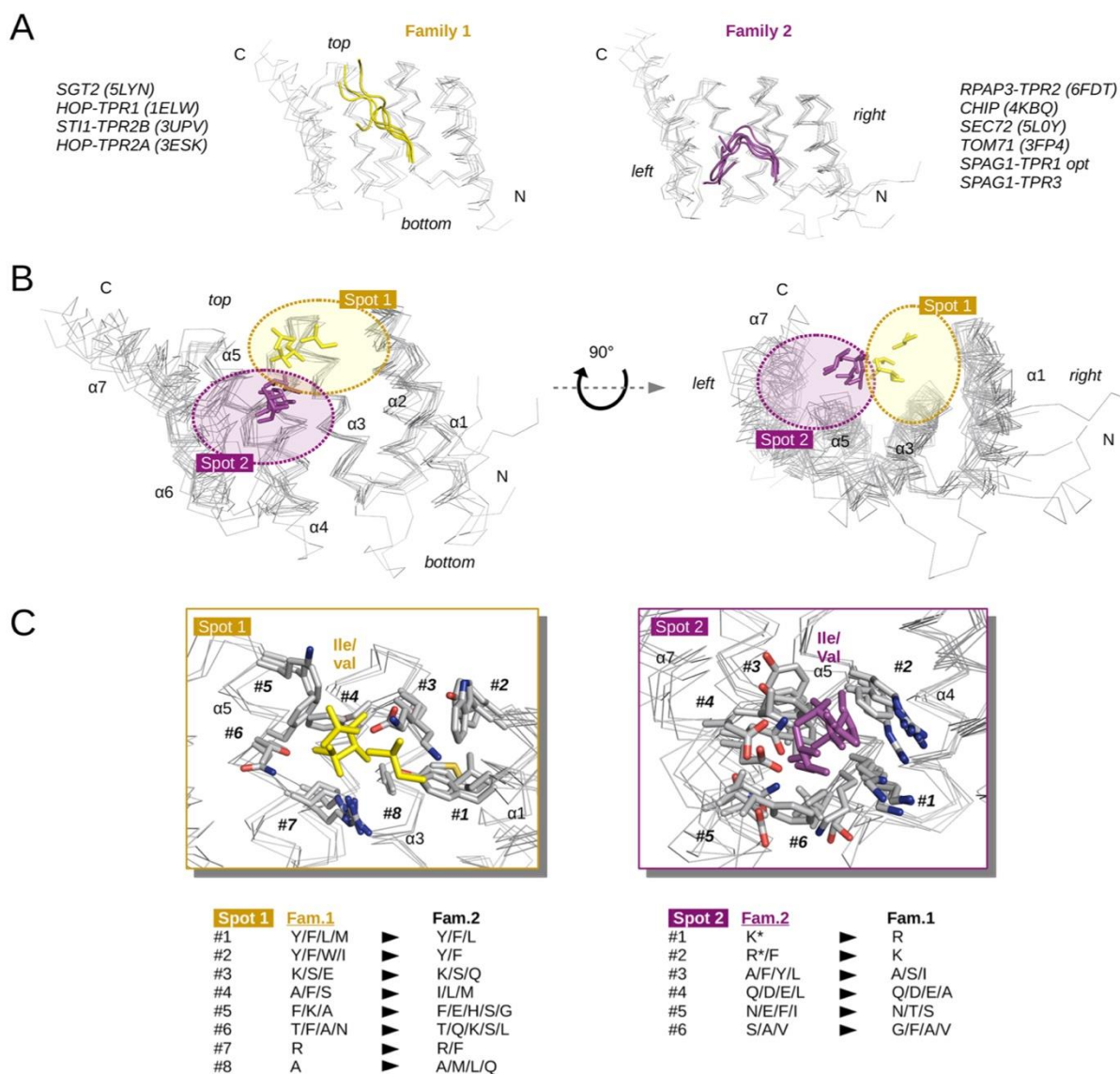


Figure 7. Structural comparison of TPR domains in complex with HSP70 peptides. (A) Families 1 and 2 of superimposed 3D structures are represented using the backbone traces of the TPR domain and cartoon representations of the HSP70 peptides. Arbitrary orientations of the structures are specified. When available, the name of the TPR protein and the PDB entry is indicated. (B) All of the 3D structures displayed in panel A are superimposed, and for each of them, the isoleucine, or the corresponding valine, in the HSP70 peptides is represented using sticks. The two main spots of interaction involving this isoleucine/valine are highlighted with colored circles. Two views are proposed. (C) Detailed views of the binding spots of each family. To look at sequence conservation on the structural alignment, up to eight positions (#1–#8) are defined at each spot. The amino acid that occupies each of these positions in all of the structures is reported, structures being sorted by family. The residues belonging to the carboxylate-clamp motif are denoted with an asterisk.

To go further, we performed the same structural analysis for TPR domains in complex with the HSP70 C-terminal tail. In the 3D structures retrieved from the Protein Data Bank, the HSP70 peptides include an Ile or a Val residue just before the EEVD end. These residues are bulkier and largely more hydrophobic than the methionine of the HSP90 end. As we showed with the comparison of 3D data obtained on TPR1_{opt} and TPR3 (Figure 5), accommodation of the Ile/Val side chains by the molecular surface of the TPR domain requires space and enough hydrophobicity to minimize exposure with

the solvent. This point is even more important because the amino acids located before the Ile/Val residue, meaning Ser, Gly, Pro, and Thr, do not possess the same interaction power as the corresponding Asp, Ala, Ser, and Arg in HSP90. Thus, compared to the DASR sequence in HSP90, the SGPT sequence in HSP70 is likely less susceptible to binding an anchoring point at the surface of the TPR domain. We classified the 3D structures into two families (Figure 7). Family 1 is close to the family 1 defined with the HSP90 peptides, because the HSP70 peptide is linear and spreads from the top

to the bottom of the protein receptor. The Val/Ile side chains are shown in the top right corner of the TPR. In family 2, Val or Ile is in the bottom left corner. For TPR domain orientation, see Figure 7. We compared TPR sequences in these two spots. Within family 1, spot 1 displays heterogeneity in terms of amino acid content. At this spot, family 2 of structures exhibits also inhomogeneous sequences. The same comments could be made for spot 2, where various amino acids occupy each position in the two families of structures. On the basis of our structural alignment, and to our mind, it was not possible to highlight sequence features that could predict the positioning of the Ile residue in the HSP70 peptide.

All of these observations can be related to the lower affinity that is generally observed for HSP70 than HSP90 for a given TPR domain. Accommodation of residues upstream from the EEVD end to the TPR domain is more problematic for the HSP70 peptide that does not display highly attractive chemical functions within its first residues. We think that the N-terminal end of the HSP70 peptide adapts the TPR surface to minimize exposure of its Ile residue to the solvent. However, having so much variability between the sequences of the TPR family is probably not the best way to achieve optimized conditions to reduce solvent exposure. On the contrary, the HSP90 peptide appears to be more susceptible to the formation of chemical bonds because of its first Asp or Arg residues. In addition, because it is smaller than Ile, its methionine offers more possibilities for accommodating the TPR surface. Thus, the HSP90 peptide sequence seems to be more adapted to sampling and finding the most favorable energetic positions on the TPR domain.

HSP Binding Properties of SPAG1. Previously, we described SPAG1 as a platform for HSP proteins through interactions with its TPR1 and TPR3. We demonstrated that the wild-type and optimized versions of SPAG1-TPR1 behaved very similarly with respect to HSP binding, especially in terms of thermodynamics, strongly suggesting that they recruit the HSP C-terminal tails with the same structural recognition mode. Despite the T285A substitution (which occurs sequentially close to but spatially quite distant from the Thr284 that was proposed as a marker of family 3 for HSP90 peptides), we can reasonably rank SPAG1-TPR1_{wt} into the same families that were described for SPAG1-TPR1_{opt} (Figures 6 and 7). Thus, the structural differences between the HSP70- and HSP90-bound forms of SPAG1-TPR1_{opt} and SPAG1-TPR3 could rationally be transposed to SPAG1-TPR1_{wt}.

This result suggests that SPAG1 needs a strong partnership with HSPs. The two chaperones, HSP70 and HSP90, are both able to hold and deliver client proteins in an assembly pathway. One could assume that the delivery of HSP clients via SPAG1 (which works in concert with PIH1D2, RUVBL1, and RUVBL1 into a so-called R2SP complex) is a crucial function for the cell expressing this protein in which two TPR domains, with similar properties, were included in the sequence. Having two structural modes of recognition for the HSPs prevents the total blocking of the SPAG1 function if an inhibitor molecule, specific from one mode or the other, enters the cell. Another possibility could be that two HSPs could be bound to SPAG1 with the aim of bringing faster, or in larger quantities, client proteins that need to be assembled in a specific pathway. It could also help the exchange of clients between HSP70 and HSP90 if the two chaperones bind SPAG1 simultaneously.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biochem.1c00052>.

Figures S1 and S2 (PDF)

Accession Codes

The NMR structure and the ¹H, ¹³C, and ¹⁵N chemical shifts of SPAG1-TPR1_{opt} were deposited in the Protein Data Bank and in the Biological Magnetic Resonance Bank as entries 7BEV and 34586, respectively. The Uniprot accession codes of human proteins SPAG1, HSP70, and HSP90 used in this study are Q07617, P0DMV8, and P08238, respectively.

■ AUTHOR INFORMATION

Corresponding Author

Marc Quinternet – Université de Lorraine, CNRS, INSERM, IBSLor, F-54000 Nancy, France; orcid.org/0000-0002-8299-7136; Email: marc.quinternet@univ-lorraine.fr

Authors

Sana Dermouche – Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France

Marie-Eve Chagot – Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France

Xavier Manival – Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.biochem.1c00052>

Author Contributions

[§]S.D. and M.-E.C. contributed equally to this work.

Funding

This work was supported by the French National Center for Scientific Research (CNRS), Université de Lorraine (France), the “Région Grand Est”, and the “Agence Nationale de la Recherche” (Grant ANR-16-CE11-0032-02).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Access to the NMR, DLS, and ITC core facilities of the Biophysics and Structural Biology department of UMS2008/US40 IBSLor was deeply appreciated (<https://umsibslor.univ-lorraine.fr/>).

■ REFERENCES

- (1) Ponting, C. P., and Russell, R. R. (2002) The Natural History of Protein Domains. *Annu. Rev. Biophys. Biomol. Struct.* 31 (1), 45–71.
- (2) Cheng, A. C., Coleman, R. G., Smyth, K. T., Cao, Q., Souillard, P., Caffrey, D. R., Salzberg, A. C., and Huang, E. S. (2007) Structure-Based Maximal Affinity Model Predicts Small-Molecule Druggability. *Nat. Biotechnol.* 25 (1), 71–75.
- (3) Structural Genomics Consortium, Architecture et Fonction des Macromolécules Biologiques, Berkeley Structural Genomics Center, China Structural Genomics Consortium, Integrated Center for Structure and Function Innovation, Israel Structural Proteomics Center, Joint Center for Structural Genomics, Midwest Center for Structural Genomics, New York Structural Genomics Research Center for Structural Genomics, Northeast Structural Genomics Consortium, Oxford Protein Production Facility, Protein Sample Production Facility, Max Delbrück Center for Molecular Medicine, and RIKEN Structural Genomics/Proteomics Initiative (2008) SPINE2-Com-

- plexes. Protein Production and Purification. *Nat. Methods* 5 (2), 135–146.
- (4) Assenberg, R., Wan, P. T., Geisse, S., and Mayr, L. M. (2013) Advances in Recombinant Protein Expression for Use in Pharmaceutical Research. *Curr. Opin. Struct. Biol.* 23 (3), 393–402.
- (5) Wingfield, P. T. (2015) Overview of the Purification of Recombinant Proteins. *Curr. Protoc. Protein Sci.* 80 (1), 6.1.1–6.1.35, DOI: 10.1002/0471140864.ps0601s80.
- (6) Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R. L., Aharoni, A., Silman, I., Sussman, J. L., Tawfik, D. S., and Fleishman, S. J. (2016) Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* 63 (2), 337–346.
- (7) D'Andrea, L. D., and Regan, L. (2003) TPR Proteins: The Versatile Helix. *Trends Biochem. Sci.* 28 (12), 655–662.
- (8) 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 (5), 1192–1199.
- (9) Main, E. R. G., Xiong, Y., Cocco, M. J., D'Andrea, L., and Regan, L. (2003) Design of Stable α -Helical Arrays from an Idealized TPR Motif. *Structure* 11 (5), 497–508.
- (10) Kajander, T., Cortajarena, A. L., Mochrie, S., and Regan, L. (2007) Structure and Stability of Designed TPR Protein Superhelices: Unusual Crystal Packing and Implications for Natural TPR Proteins. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 63 (7), 800–811.
- (11) Chen, S., and Smith, D. F. (1998) Hop as an Adaptor in the Heat Shock Protein 70 (Hsp70) and Hsp90 Chaperone Machinery. *J. Biol. Chem.* 273 (52), 35194–35200.
- (12) 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 (22), 19265–19275.
- (13) 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 (2), 199–210.
- (14) Henri, J., Chagot, M. E., Bourguet, M., Abel, Y., Terral, G., Maurizy, C., Aigueperse, C., Georgescauld, F., Vandermoere, F., Saint-Fort, R., Behm-Ansmant, I., Charpentier, B., Pradet-Balade, B., Verheggen, C., Bertrand, E., Meyer, P., Cianferani, S., Manival, X., and Quinternet, M. (2018) Deep Structural Analysis of RPAP3 and PIH1D1, Two Components of the HSP90 Co-Chaperone R2TP Complex. *Structure* 26 (9), 1196–1209.e8.
- (15) Chagot, M. E., Dos Santos Morais, R., Dermouche, S., Lefebvre, D., Manival, X., Chipot, C., Dehez, F., and Quinternet, M. (2019) Binding Properties of the Quaternary Assembly Protein SPAG1. *Biochem. J.* 476 (11), 1679–1694.
- (16) Back, R., Dominguez, C., Rothe, B., Bobo, C., Beauflis, C., Morera, S., Meyer, P., Charpentier, B., Branlant, C., Allain, F. H., and Manival, X. (2013) High-Resolution Structural Analysis Shows How Tah1 Tethers Hsp90 to the R2TP Complex. *Structure* 21 (10), 1834–1847.
- (17) Schmid, A. B., Lagleder, S., Grawert, M. A., Rohl, A., Hagn, F., Wandinger, S. K., Cox, M. B., Demmer, O., Richter, K., Groll, M., Kessler, H., and Buchner, J. (2012) The Architecture of Functional Modules in the Hsp90 Co-Chaperone Sti1/Hop. *EMBO J.* 31 (6), 1506–1517.
- (18) Tripathi, A., Mandon, E. C., Gilmore, R., and Rapoport, T. A. (2017) Two Alternative Binding Mechanisms Connect the Protein Translocation Sec71-Sec72 Complex with Heat Shock Proteins. *J. Biol. Chem.* 292 (19), 8007–8018.
- (19) Horani, A., Ustione, A., Huang, T., Firth, A. L., Pan, J., Gunsten, S. P., Haspel, J. A., Piston, D. W., and Brody, S. L. (2018) Establishment of the Early Cilia Preassembly Protein Complex during Motile Ciliogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 115 (6), E1221–E1228.
- (20) Knowles, M. R., Ostrowski, L. E., Loges, N. T., Hurd, T., Leigh, M. W., Huang, L., Wolf, W. E., Carson, J. L., Hazucha, M. J., Yin, W., Davis, S. D., Dell, S. D., Ferkol, T. W., Sagel, S. D., Olivier, K. N., Jahnke, C., Olbrich, H., Werner, C., Raidt, J., Wallmeier, J., Pennekamp, P., Dougherty, G. W., Hjeij, R., Gee, H. Y., Otto, E. A., Halbritter, J., Chaki, M., Diaz, K. A., Braun, D. A., Porath, J. D., Schueler, M., Baktai, G., Griesse, M., Turner, E. H., Lewis, A. P., Bamshad, M. J., Nickerson, D. A., Hildebrandt, F., Shendure, J., Omran, H., and Zariwala, M. A. (2013) Mutations in SPAG1 Cause Primary Ciliary Dyskinesia Associated with Defective Outer and Inner Dynein Arms. *Am. J. Hum. Genet.* 93 (4), 711–720.
- (21) Lin, W., Zhou, X., Zhang, M., Li, Y., Miao, J., Wang, L., Zong, S., and Koide, S. S. (2001) Expression and Function of the HSD-3.8 Gene Encoding a Testis-Specific Protein. *Mol. Hum. Reprod.* 7 (9), 811–818.
- (22) Maurizy, C., Quinternet, M., Abel, Y., Verheggen, C., Santo, P. E., Bourguet, M. C. F., Paiva, A., Bragantini, B., Chagot, M. E., Robert, M. C., Abeza, C., Fabre, P., Fort, P., Vandermoere, F. M. F., Sousa, P., Rain, J. C., Charpentier, B., Cianferani, S., Bandejas, T. M., Pradet-Balade, B., Manival, X., and Bertrand, E. (2018) The RPAP3-Cterminal Domain Identifies R2TP-like Quaternary Chaperones. *Nat. Commun.* 9 (1), 2093.
- (23) Rodríguez, C. F., and Llorca, O. (2020) RPAP3 C-Terminal Domain: A Conserved Domain for the Assembly of R2TP Co-Chaperone Complexes. *Cells* 9 (5), 1139 DOI: 10.3390/cells9051139.
- (24) Chagot, M. E., Jacquemin, C., Branlant, C., Charpentier, B., Manival, X., and Quinternet, M. (2015) (1)H, (15)N and (13)C Resonance Assignments of the Two TPR Domains from the Human RPAP3 Protein. *Biomol. NMR Assignments* 9 (1), 99–102.
- (25) Manival, X., Jacquemin, C., Charpentier, B., and Quinternet, M. (2015) (1)H, (15)N and (13)C Resonance Assignments of the Yeast Pih1 and Tah1 C-Terminal Domains Complex. *Biomol. NMR Assignments* 9 (1), 71–73.
- (26) 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 (8), S698–S709.
- (27) Pal, M., Morgan, M., Phelps, S. E., Roe, S. M., Parry-Morris, S., Downs, J. A., Polier, S., Pearl, L. H., and Prodromou, C. (2014) Structural Basis for Phosphorylation-Dependent Recruitment of Tel2 to Hsp90 by Pih1. *Structure* 22, 805.
- (28) Fabczak, H., and Osinka, A. (2019) Role of the Novel Hsp90 Co-Chaperones in Dynein Arms' Preassembly. *Int. J. Mol. Sci.* 20 (24), 6174.
- (29) Boulon, S., Pradet-Balade, B., Verheggen, C., Molle, D., Boireau, S., Georgieva, M., Azzag, K., Robert, M. C., Ahmad, Y., Neel, H., Lamond, A. I., and Bertrand, E. (2010) HSP90 and Its R2TP/Prefoldin-like Co-chaperone Are Involved in the Cytoplasmic Assembly of RNA Polymerase II. *Mol. Cell* 39 (6), 912–924.
- (30) Benbahouche, N. E. H., Iliopoulos, I., Torok, I., Marhold, J., Henri, J., Kajava, A. V., Farkas, R., Kempf, T., Schnolzer, M., Meyer, P., Kiss, I., Bertrand, E., Mechler, B. M., and Pradet-Balade, B. (2014) Drosophila Spag Is the Homolog of RNA Polymerase II-Associated Protein 3 (RPAP3) and Recruits the Heat Shock Proteins 70 and 90 (Hsp70 and Hsp90) during the Assembly of Cellular Machineries. *J. Biol. Chem.* 289 (9), 6236–6247.
- (31) Martino, F., Pal, M., Muñoz-Hernandez, H., Rodríguez, C. F., Nunez-Ramirez, R., Gil-Carton, D., Degliesposti, G., Skehel, J. M., Roe, S. M., Prodromou, C., Pearl, L. H., and Llorca, O. (2018) RPAP3 Provides a Flexible Scaffold for Coupling HSP90 to the Human R2TP Co-Chaperone Complex. *Nat. Commun.* 9 (1), 1501.
- (32) Muñoz-Hernández, H., Pal, M., Rodríguez, C. F., Fernández-Leiro, R., Prodromou, C., Pearl, L. H., and Llorca, O. (2019) Structural Mechanism for Regulation of the AAA-ATPases RUVBL1-RUVBL2 in the R2TP Co-Chaperone Revealed by Cryo-EM. *Sci. Adv.* 5 (5), No. eaaw1616.

- (33) Yamaguchi, H., Oda, T., Kikkawa, M., and Takeda, H. (2018) Systematic Studies of All PIH Proteins in Zebrafish Reveal Their Distinct Roles in Axonemal Dynein Assembly. *eLife* 7, e36979.
- (34) Horejsi, Z., Stach, L., Flower, T. G., Joshi, D., Flynn, H., Skehel, J. M., O'Reilly, N. J., Ogradowicz, R. W., Smerdon, S. J., and Boulton, S. J. (2014) Phosphorylation-Dependent PIH1D1 Interactions Define Substrate Specificity of the R2TP Cochaperone Complex. *Cell Rep.* 7 (1), 19–26.
- (35) Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: Protein Structure and Function Prediction. *Nat. Methods* 12 (1), 7–8.
- (36) 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.* 175, 178.
- (37) Berjanskii, M. V., Neal, S., and Wishart, D. S. (2006) PREDITOR: A Web Server for Predicting Protein Torsion Angle Restraints. *Nucleic Acids Res.* 34 (Web Server issue), W63–W9.
- (38) Shen, Y., and Bax, A. (2013) Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *J. Biomol. NMR* 56 (3), 227–241.
- (39) Guntert, P. (2004) Automated NMR Structure Calculation with CYANA. *Methods Mol. Biol.* 278, 353–378.
- (40) Bertini, I., Case, D. A., Ferella, L., Giachetti, A., and Rosato, A. (2011) A Grid-Enabled Web Portal for NMR Structure Refinement with AMBER. *Bioinformatics* 27 (17), 2384–2390.
- (41) 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: Struct., Funct., Genet.* 69 (4), 726–733.
- (42) Schrödinger, LLC (2015) *The PyMOL Molecular Graphics System*, ver. 1.8.
- (43) Rodrigues, J. P., Trellet, M., Schmitz, C., Kastiris, P., Karaca, E., Melquiond, A. S., and Bonvin, A. M. (2012) Clustering Biomolecular Complexes by Residue Contacts Similarity. *Proteins* 80 (7), 1810–1817.
- (44) Kandai, G., Acencio, M. L., and Lemke, N. (2015) Prediction of Druggable Proteins Using Machine Learning and Systems Biology: A Mini-Review. *Front. Physiol.*, 6.
- (45) Darby, J. F. (2020) Solution Structure of the Hop TPR2A Domain and Investigation of Target Druggability by NMR, Biochemical and in Silico Approaches. *Sci. Rep.* 10, 16000.
- (46) Kawachi, T., Tanaka, S., Fukuda, A., Sumii, Y., Setiawan, A., Kotoku, N., Kobayashi, M., and Arai, M. (2019) Target Identification of the Marine Natural Products Dictyocerin-A and -C as Selective Growth Inhibitors in Cancer Cells Adapted to Hypoxic Environments. *Mar. Drugs* 17 (3), 163.

*Conclusion et
Perspectives*

1. Caractérisation du complexe Rvb1/Rvb2 et Bcd1p de *S. cerevisiae*

Plusieurs études s'intéressent au rôle joué par les ATPases Rvb1/Rvb2 et leurs homologues humains RUVBL1/RUVBL2. Cela se justifie par le fait qu'elles sont recrutées dans un large panel de voies moléculaires différentes qui suggèrent un rôle central et essentiel dans la cellule. Malgré les progrès significatifs au cours des dernières années sur le plan structural et biochimique de leurs études, une compréhension claire et précise du fonctionnement reste en suspens. L'étude plus précise de leur recrutement sur les différents mécanismes biologiques dans lesquels elles sont impliquées, ainsi que la compréhension du rôle de la liaison et/ou de l'hydrolyse de l'ATP sur leur conformation et fonction, permettrait de décrire les étapes clés de leur fonctionnement en prenant comme modèle l'assemblage de la snoRNP C/D.

Une partie de ce travail de thèse a donc été de chercher à mieux caractériser l'implication de ce complexe protéique dans le processus d'assemblage des snoRNP à boîtes C/D par l'étude biochimique et biophysique de leur interaction avec le facteur d'assemblage essentiel Bcd1p.

Nos expériences de co-élution par chromatographie d'exclusion de taille ont permis de mettre en évidence l'interaction stable entre Bcd1p et le complexe Rvb1/Rvb2. Cependant, cette interaction n'a pu être obtenue qu'en présence de nucléotides ADP et potentiellement ATP dans le milieu réactionnel. Nous avons pu montrer que ces nucléotides influencent fortement l'état oligomérique du complexe Rvb1/Rvb2 libre, qui va connaître des réarrangements importants de sa structure quaternaire. L'ADP et l'ATP sélectionnent une conformation en hétérohexamère (3 Rvb1 + 3 Rvb2) stable dans les conditions expérimentales utilisées dans nos expériences. L'ajout d'ATP en présence de magnésium dans le tampon conduit à son hydrolyse rapide par le complexe Rvb1/Rvb2 donnant un état post-hydrolyse où nous aurions de l'ADP lié à Rvb1 et Rvb2 car il n'y a pas de cofacteur à libérer des sites actifs. Cet état post-hydrolyse, en présence de Bcd1p, favoriserait donc une conformation spatiale hétérododécamérique du complexe Rvb1/Rvb2 en interaction avec Bcd1p. Nos données de cryo-EM montrent clairement cette organisation dodécamérique des Rvb1/Rvb2 en présence de Bcd1p.

Ce résultat a également été appuyé par l'utilisation des deux analogues non hydrolysables de l'ATP : l'AMP-PNP et l'ATP- γ S. Leur fixation dans les sites actifs de Rvb1 et Rvb2 mime un état de pré-hydrolyse de l'ATP. D'après les résultats obtenus par filtration sur gel, la seule fixation de l'ATP ne permet pas l'interaction de Bcd1p avec le complexe Rvb1/Rvb2.

Il a été montré qu'il existe des acides aminés conservés chez les protéines humaines RUVBL1 et RUVBL2, et Rvb1 et Rvb2 de la levure *S. cerevisiae*, au niveau des motifs Walker A et Walker B, dont la mutation entraîne le blocage de la fixation de l'ATP mais aussi l'inhibition

de son hydrolyse. Les mutations spécifiques K85E sur Rvb1 et K81E sur Rvb2 empêchent la fixation de l'ATP, et les mutations E312N sur Rvb1 et E297N sur Rvb2 inhibent l'hydrolyse de l'ATP. L'expression transitoire de ces formes mutantes dans la levure permettrait une étude plus approfondie de l'implication du complexe Rvb1/Rvb2 dans l'assemblage des snoRNP à boîtes C/D, et ainsi une caractérisation des complexes intermédiaires formés *in vivo*. Des constructions de vecteurs permettant l'expression de ces mutants ont déjà été entamées au laboratoire.

Il a été montré au laboratoire par RMN que cette interaction est également possible entre RUVBL1/RUVBL2 et le domaine doigt à zinc de Bcd1p. Il a également pu être montré par comparaison de structures 3D réalisées au laboratoire que ce domaine doigt à zinc est fortement similaire au doigt à zinc de Hit1p. Les deux appartiennent à la famille *Zf-HIT-Domain* (ZHD). Cependant, les données expérimentales présentés dans ce travail montrent bien une absence de co-élution entre Hit1p/Rsa1p et Rvb1/Rvb2 en présence d'ADP ou d'ATP. Cela peut s'expliquer par la nécessité d'autres facteurs en plus de Bcd1p qui assurent une stabilisation de l'interface d'interaction entre Rvb1/Rvb2 et Bcd1p. Toutefois la présence du fragment Rsa1p₂₃₁₋₃₇₅ pourrait perturber cette interaction. Le recours à la coexpression du complexe Hit1p/Rsa1p se justifie par le fait que la protéine Hit1 se dégrade partiellement lorsqu'elle est exprimée toute seule chez *E. coli*. D'autres conditions de purification de Hit1p seul sans dégradation ou en la limitant sont donc à tester par la suite pour s'assurer de ces hypothèses. Nous pourrions notamment coexprimer directement Hit1p avec les Rvb1/Rvb2 ou produire Hit1p avec une étiquette du type GST (Glutathion-S-Transférase).

La formation du complexe Rvb1/Rvb2/Bcd1p a pu être caractérisé par la détermination des paramètres thermodynamiques en ITC. Il s'agit d'une interaction relativement forte étant donné la gamme sub-micromolaire de sa constante de dissociation. Le nombre de site estimé est de 2 ce qui suggèrerait la présence de deux Bcd1p dans le complexe.

La détermination de la stœchiométrie du complexe Rvb1/Rvb2/Bcd1p en présence d'ADP a été une de nos préoccupations dans sa caractérisation. Pour cela, plusieurs échantillons ont été envoyés à nos collaborateurs à Strasbourg (LSMBO) pour une analyse par spectrométrie de masse native et par masse photométrique. Cependant, l'absence de signal pour le complexe n'a pas permis de déterminer la stœchiométrie du complexe.

L'étude du complexe Rvb1/Rvb2/Bcd1p s'est poursuivie par sa caractérisation structurale par cryo-microscopie électronique au laboratoire de *ThermoFisher Scientific* au Pays-Bas. Les données obtenues de classification 3D ont révélé une organisation quaternaire de Rvb1/Rvb2 en hétérododécamère pour la majorité des images analysées et certainement en complexe avec

Bcd1p pour 30% des cas (conformation N°2). L'exploitation de ces données est toujours en cours par notre collaboratrice Célia Plisson-Chastang au laboratoire de biologie moléculaire, cellulaire et du développement à Toulouse. Avant d'en arriver à un modèle de la structure 3D du complexe, elle consiste, dans un premier temps, en l'identification de la ou les position(s) de Bcd1p sur le complexe hétérododécamérique Rvb1/Rvb2, ainsi que la détermination de la stœchiométrie par la détermination du nombre de protéines Bcd1 dans le complexe.

Ce changement conformationnel important du complexe Rvb1/Rvb2 en interaction (hétérododécamère) ou pas (hétérohéxamère) avec Bcd1p, affecte la mesure de la constante de dissociation (Kd) obtenue dans cette étude par ITC. Il est à noter que le dégagement de chaleur mesuré au cours des titrations par Bcd1p sur Rvb1/Rvb2 dans la cellule représente à la fois les liaisons mises en jeu pour l'interaction de deux hexamères formant un dodécamère via leurs domaines DII, ainsi que les liaisons permettant l'interaction de Bcd1p au sein du complexe Rvb1/Rvb2.

Un modèle d'assemblage est proposé dans cette étude (Figure 42), il illustre essentiellement les changements stœchiométriques du complexe Rvb1/Rvb2 en interaction ou pas avec Bcd1p, selon l'état nucléotidique *in vitro*. Le complexe R2TP aurait pour rôle la stabilisation de la protéine cœur Nop58. L'association de Bcd1p avec R2TP-Nop58p permettrait ainsi la déposition de Nop58p aux sites d'assemblage de la particule RNP. La dissociation des facteurs d'assemblage et la libération de la particule snoRNP à boîtes C/D (étape E) (Figure 42) reste encore non caractérisée.

2. Une mutation ponctuelle D72A sur Bcd1p perturbe l'assemblage des snoRNP à boîtes C/D chez la levure *S. cerevisiae*

L'évolution des organismes eucaryotes a été suivie d'une évolution et d'une adaptation des mécanismes cellulaires, par une complexification des machineries moléculaires, permettant une régulation plus fine des processus essentiels dans la cellule. L'assemblage des particules snoRNP à boîtes C/D a été longtemps étudié, la difficulté de la compréhension de ces étapes d'assemblage vient du fait de la nécessité de plusieurs facteurs protéiques qui agissent en *trans* sur le processus d'assemblage, mais aussi sont retrouvés, pour la plupart, dans d'autres machineries cellulaires différentes. Le complexe R2TP est un exemple parfait qui illustre cela. La diversité fonctionnelle de ce complexe est essentiellement permise par d'autres cofacteurs qui vont favoriser son recrutement à des sites spécifiques dans la cellule. Bcd1p est un facteur

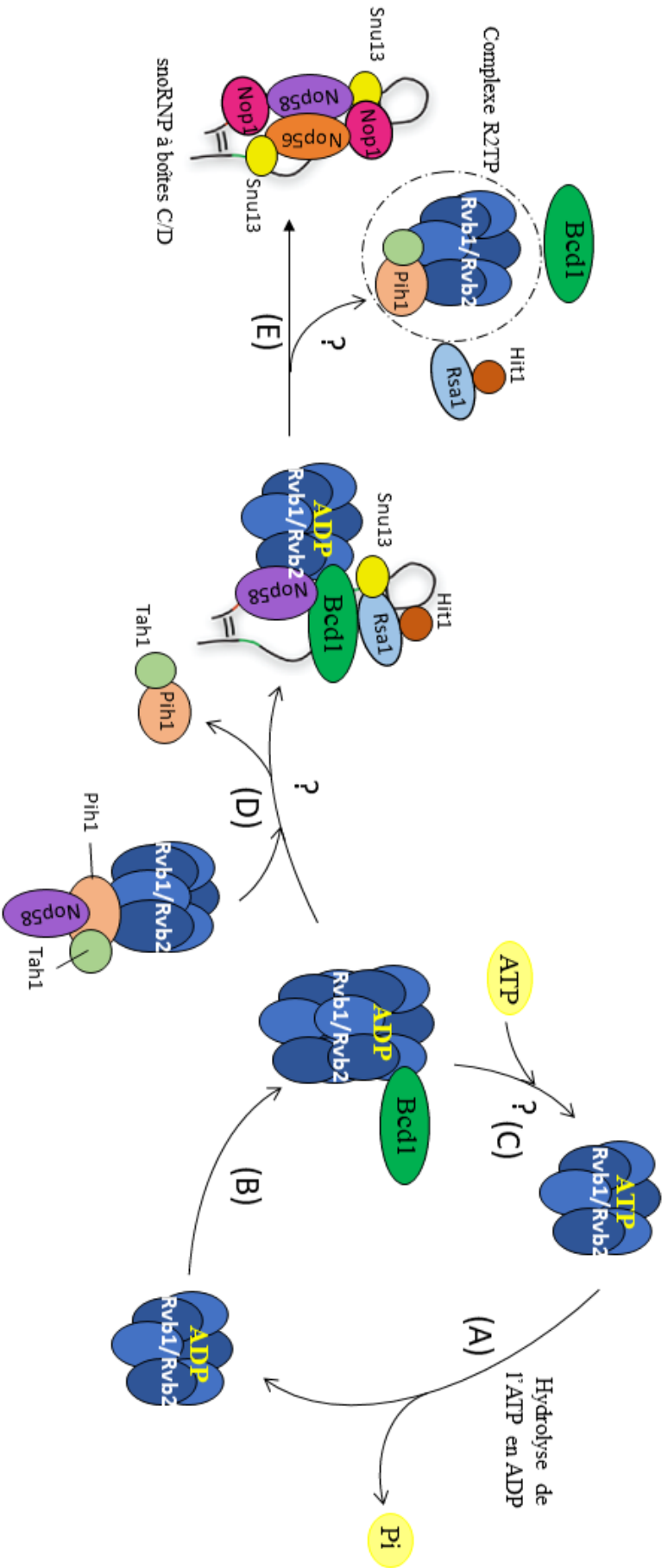


Figure 42 : Modèle d'assemblage des snoRNP à boîtes C/D et le cycle nucléotidique du complexe Rvb1/Rvb2/Bcd1p.

(A) L'ATP contenu dans les sites actifs du complexe Rvb1/Rvb2 est hydrolysée en ADP suivi de la libération de Phosphate (Pi). (B) Le recrutement de Bcd1p sur le complexe Rvb1/Rvb2 en présence d'ADP dans les sites actifs, induit potentiellement la dodécamérisation de ce dernier. (C) Le mécanisme de recyclage du complexe Rvb1/Rvb2/Bcd1p est inconnu, il implique probablement la dissociation de Bcd1p du complexe Rvb1/Rvb2 et la division du dodécamère en hexamère. (D) Le complexe R2TP amène la protéine cœur Nop58 aux sites d'assemblage. Bcd1p recrutée par le complexe Rvb1/Rvb2, s'associe au complexe protéique de pré-assemblage R2TP-Nop58, induisant la dissociation de Tah1p-Pih1p, cependant le mécanisme moléculaire précis de ce remaniement quaternaire du complexe d'assemblage (Rvb1/Rvb2/Nop58/Bcd1p) reste pour l'heure peu décrit. Ce complexe s'associe sur la pré-snoRNP (Snu13p/snoARN) assurant ainsi la déposition de Nop58p. (E) L'assemblage se poursuit par la dissociation des facteurs d'assemblage (complexe R2TP, Rsa1p-Hit1p et Bcd1p), et le recrutement des protéines cœur Nop1p et Nop56p, constituant ainsi une snoRNP à boîtes C/D mature et fonctionnelle.

essentiel dans cette voie d'assemblage. Son absence affecte drastiquement la croissance des levures, mais aussi le recrutement correct d'autres protéines et complexes protéiques clients du R2TP aux sites d'assemblage des snoRNP à boîtes C/D.

Nous avons vu dans ce travail de thèse qu'il existe une conservation de certains acides aminés sur la protéine Bcd1p au cours de l'évolution, plus exactement au niveau de la région 1-96 identifiée comme étant fonctionnellement suffisante chez la levure *S. cerevisiae*. La mutation de seulement un acide aminé (Aspartate 72 en Alanine) a montré un effet drastique sur l'assemblage des snoRNP à boîtes C/D. En effet, l'expression de ce mutant chez la levure conduit à un fort ralentissement de la croissance des cellules et cela malgré l'expression de la version endogène.

Une analyse plus approfondie de son effet sur le plan moléculaire indique une perte de la stabilité des snoARN à boîtes C/D, qui serait potentiellement due à un assemblage incomplet et abortif des particules ribonucléoprotéiques dans la cellule. L'étude de ce mutant a également été le centre d'intérêt de l'équipe de H. Ghalei, où ils montrent par le biais d'expérience *in vitro* de *GST-pull down* que Bcd1p D72A perd son interaction avec Snu13p (Khoshnevis et al., 2019). Par ailleurs, l'investigation détaillée de l'effet de ce mutant dans la cellule n'a pas été étudiée mis à part une baisse de la croissance des cellules.

Cet effet important sur la croissance des levures nous a poussé à réfléchir à une façon d'étudier les associations protéiques perturbées dans ce processus d'assemblage par le mutant Bcd1p D72A, en s'affranchissant des effets connus dus au ralentissement de la croissance des levures (Kakihara et al., 2014). Pour cela, son expression de façon induite et transitoire, permettrait ainsi son intégration dans les étapes d'assemblage de snoRNP à boîtes C/D néoformées dans la cellule et donc leur caractérisation par Western-Blot.

Un travail de thèse mené au laboratoire (Paul, 2018) montre qu'il existe un lien entre la protéine Bcd1 et la protéine cœur Nop58 (Paul, 2018). L'absence de Bcd1p dans la cellule semble bloquer la déposition de Nop58p sur le snoARN à boîtes C/D au niveau du site d'assemblage de la future snoRNP. Dans un premier temps, nous avons voulu savoir si cette association de la protéine cœur Nop58 avec le snoARN est également affectée par l'expression du mutant Bcd1p D72A. Les résultats d'immunoprécipitation d'ARN (RIP) indiquent que Nop58p n'est plus recrutée sur les snoARN à boîtes C/D testés. De plus, Bcd1p D72A perd également son association avec ces ARN. Il existerait donc un lien fonctionnel étroit entre Nop58p et Bcd1p pour la déposition de Nop58p aux sites d'assemblages des snoRNP à boîtes C/D. Ces résultats ont été obtenus en analysant les snoARN à boîtes C/D transcrits par l'ARN polymérase II.

Cependant, en ce qui concerne le cas particulier du snoARN snR52 transcrit par l'ARN polymérase III, sa stabilité ne semble pas être affectée par l'expression du mutant Bcd1p D72A. Par ailleurs, l'association du mutant sur le snR52 diminue de moitié, ce qui n'empêche pas la déposition de la protéine cœur Nop58p sur le snoARN snR52. Ce résultat suggère l'existence d'un mécanisme alternatif et/ou parallèle de l'assemblage des snoRNP à boîtes C/D mettant en jeu le seul transcrit ARN polymérase III snR52. Des travaux antérieurs réalisés au laboratoire montre qu'une association existe entre Bcd1p et les sous-unités de l'ARN polymérase III, plus exactement avec la région C-terminale de Bcd1p. La mutation D72A sur Bcd1p ne semble donc pas perturber cette association, qui va alors permettre un assemblage correct de la snoRNP snR52. Toutefois cette snoRNP n'est pas suffisante pour assurer les modifications correctes des ARNr et par extension l'assemblage de ribosome fonctionnel essentiel à la croissance des levures. Cette hypothèse est appuyée par le fait que des études indiquent que l'absence de la snoRNP snR52 n'affecte pas de façon notable la modification des ARNr (Baxter-Roshek et al., 2007). Afin de confirmer cette hypothèse, il est important de s'assurer que l'association de Bcd1p D72A et les sous-unités de la polymérase III sont maintenues dans ces conditions. Ces expériences seront réalisées au laboratoire par double hybride.

Nos résultats suggèrent également une perte d'association entre Snu13p et le mutant Bcd1p D72A. Cela pourrait en partie expliquer l'absence de Nop58p sur le snoARN. Des modèles d'assemblage de la particule snoRNP à boîtes C/D établis au laboratoire semblent indiquer qu'une pré-snoRNP constituée essentiellement du snoARN, Snu13p (qui reconnaît de façon précoce le *K-turn* des snoARN), Rsa1p et Hit1p se forme dans un premier temps. Bcd1p a donc été proposée comme facteur assurant le lien entre cette pré-particule et le complexe protéique ou riboprotéique qui véhicule la protéine cœur Nop58p. L'interaction directe Bcd1p-Snu13p favoriserait donc ce rapprochement aux sites d'assemblages des snoRNP à boîtes C/D. La confirmation de cette hypothèse nécessite des résultats supplémentaires. En effet, nous ignorons si l'interaction de Snu13p et Rsa1p sur le snoARN sont maintenues quand le mutant Bcd1p D72A est exprimé. Des résultats obtenus au laboratoire suggèrent qu'en absence d'expression de Bcd1p dans la cellule, Nop56p interagit de façon précoce avec la pré-snoRNP (snoARN à boîtes C/D et Snu13p), cela bloque l'assemblage de snoRNP C/D fonctionnelles. Une des perspectives de ce projet serait de tester si Nop56p est associée au snoARN en exprimant le mutant Bcd1p D72A. Des souches ont été générées à cet effet au laboratoire pour permettre de répondre à ces questions par des expériences d'immunoprécipitation. Nous avons également voulu savoir si l'association Bcd1p D72A et Nop58p est déstabilisée, cependant des optimisations des immunoprécipitations sont encore nécessaires (Figure 43).

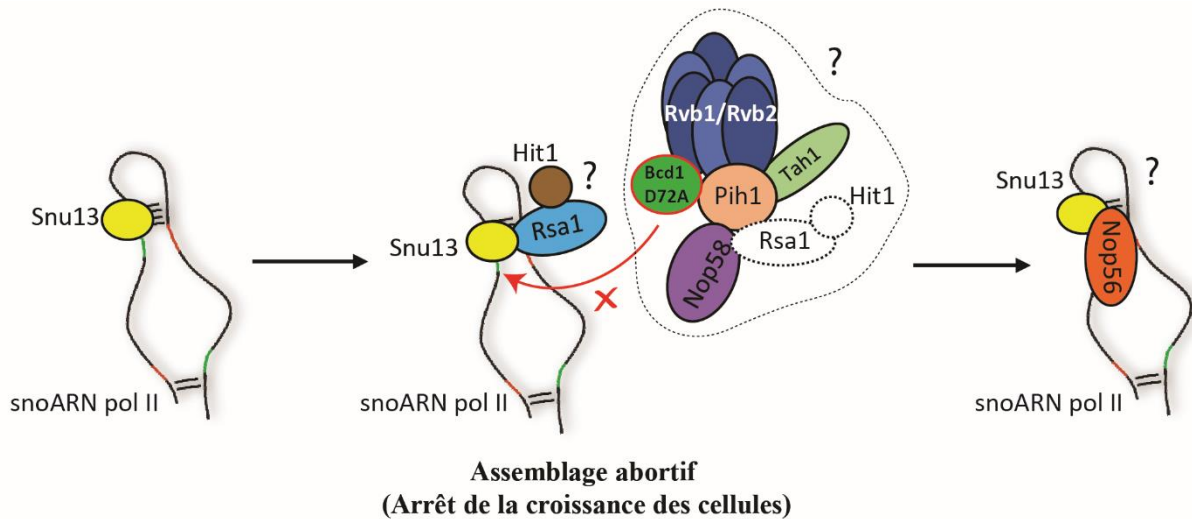


Figure 43 : Modèle de blocage de l'assemblage de snoRNP à boîtes C/D lors de l'expression du mutant Bcd1p D72A chez *S. cerevisiae*.

L'assemblage se caractérise par l'arrivée précoce de Snu13p sur le snoARN à boîtes C/D favorisant la structuration du K-turn. La pré-snoRNP (snoARN à boîtes C/D, Snu13p, Rsa1p et Hit1p) a été caractérisé précédemment au laboratoire comme étant stable biochimiquement *in vitro*, cependant nous ignorons si ce complexe ARN/protéine se forme au cours de l'assemblage dans la cellule. La mutation D72A sur Bcd1p bloque son association au snoARN à boîtes C/D et à Snu13p. La caractérisation des partenaires protéiques de ce mutant nécessite d'avantage de données expérimentales (co-immunoprécipitation). Nop56p s'associe de façon précoce au snoARN à boîtes C/D en absence de Bcd1p, il est intéressant de tester par RIP si cette association a lieu lors de l'expression de Bcd1p D72A.

3. Caractérisation d'un complexe R2TP-like : le R2SP

Ce travail a débuté par la volonté au laboratoire de caractériser fonctionnellement et structuralement l'extrémité C-terminale de la protéine humaine RPAP3 (Maurizy et al., 2018). En plus de RPAP3, le génome humain code pour deux protéines comprenant un domaine RPAP3-Cter : SPAG1 et CCDC103. SPAG1 est une protéine multi-domaines, connue pour participer à l'assemblage des bras de dynéine interne et externe des cils cellulaires. Ces bras sont nécessaires au fonctionnement des cellules sensibles et mobiles. Avec RUVBL1, RUVBL2 et PIH1D2, SPAG1 est un élément clé du R2SP, un complexe protéique assistant le repliement quaternaire de clients protéiques de façon tissu spécifique. Dans cette étude, nous avons analysé le rôle des domaines TPR de SPAG1 dans le recrutement des protéines chaperon HSP70/HSP90. Premièrement, nous proposons que seuls deux, des trois domaines TPR, sont capables de recruter HSP70 et HSP90, les domaines TPR1 et TPR3 pour lesquels la structure 3D a été déterminée par RMN au laboratoire.

La combinaison de plusieurs approches d'optimisation de séquences protéiques via le serveur web PROSS (Goldenzweig et al., 2016) avec des techniques structurales et biophysiques ont permis une analyse avec une précision atomistique des interfaces d'interactions de ces domaines TPR avec l'extrémité C-terminale des protéines HSP70 et HSP90. Des acides aminés dans les séquences des TPR ont été identifiés comme responsables et essentiels au bon positionnement des peptides des HSPs. Malgré la grande similarité entre HSP70 et HSP90, leur recrutement sur les domaines TPR étudiés de SPAG1 est différent. L'existence de ces deux modes de reconnaissance des HSPs empêcherait potentiellement le blocage total de la fonction de SPAG1 par une molécule inhibitrice. Il est également suggéré que ces deux HSPs puissent se lier à SPAG1 dans le but d'apporter plus rapidement et/ou en plus grande quantité les clients protéiques appropriés. Ces deux domaines TPR pourraient offrir une autre possibilité qui serait un transfert direct entre un client de HSP70 et HSP90 avant d'être pris en charge au final par le R2SP.

De plus, les résultats de notre étude suggèrent que SPAG1 ne possède pas d'activité GTPase proprement dite, mais elle participe à l'activation de nucléotidases telles que RUVBL1/RUVBL2 qui pourraient faciliter leur rôle dans l'assemblage quaternaire de complexes macromoléculaires. Des mesures d'activité ATPasiques pourraient mettre en évidence le rôle de SPAG1 dans la régulation précise de l'hydrolyse de nucléotides dans les larges complexes R2SP et R2SP/HSP.

Références
Bibliographiques

Références Bibliographiques

Aittaleb, M., Rashid, R., Chen, Q., Palmer, J.R., Daniels, C.J., and Li, H. (2003). Structure and function of archaeal box C/D sRNP core proteins. *Nat Struct Mol Biol* 10, 256–263. <https://doi.org/10.1038/nsb905>.

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. <https://doi.org/10.1017/S1355838202020034>.

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., & Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *The EMBO journal*, 18(19), 5399-5410.

Aramayo, R.J., Willhoft, O., Ayala, R., Bythell-Douglas, R., Wigley, D.B., and Zhang, X. (2018). Cryo-EM structures of the human INO80 chromatin-remodeling complex. *Nat Struct Mol Biol* 25, 37–44. <https://doi.org/10.1038/s41594-017-0003-7>.

Armistead, J., Khatkar, S., Meyer, B., Mark, B.L., Patel, N., Coghlan, G., Lamont, R.E., Liu, S., Wiechert, J., Cattini, P.A., et al. (2009). Mutation of a Gene Essential for Ribosome Biogenesis, EMG1, Causes Bowen-Conradi Syndrome. *The American Journal of Human Genetics* 84, 728–739. <https://doi.org/10.1016/j.ajhg.2009.04.017>.

Arnez, J.G., and Steitz, T.A. (2002). Crystal Structure of Unmodified tRNA^{Gln} Complexed with Glutamyl-tRNA Synthetase and ATP Suggests a Possible Role for Pseudo-Uridines in Stabilization of RNA Structure (American Chemical Society).

Auffinger, P., and Westhof, E. (1998). Effects of Pseudouridylation on tRNA Hydration and Dynamics: a Theoretical Approach. In *Modification and Editing of RNA*, (John Wiley & Sons, Ltd), pp. 103–112.

Ayadi, L., Galvanin, A., Pichot, F., Marchand, V., and Motorin, Y. (2019). RNA ribose methylation (2'-O-methylation): Occurrence, biosynthesis and biological functions. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1862, 253–269. <https://doi.org/10.1016/j.bbagr.2018.11.009>.

Bachelierie, J.-P., and Cavaillé, J. (1997). Guiding ribose methylation of rRNA. *Trends in Biochemical Sciences* 22, 257–261. [https://doi.org/10.1016/S0968-0004\(97\)01057-8](https://doi.org/10.1016/S0968-0004(97)01057-8).

Bachelierie, J.-P., Nicoloso, M., Qu, L.-H., Michot, B., Caizergues-Ferrer, M., Cavaille, J., and Renalier, M.-H. (1995). Novel intron-encoded small nucleolar RNAs with long sequence complementarities to mature rRNAs involved in ribosome biogenesis. *Biochem. Cell Biol.* 73, 835–843. <https://doi.org/10.1139/o95-091>.

Bachelierie, J.-P., Cavaillé, J., and Hüttenhofer, A. (2002). The expanding snoRNA world. *Biochimie* 84, 775–790. [https://doi.org/10.1016/S0300-9084\(02\)01402-5](https://doi.org/10.1016/S0300-9084(02)01402-5).

Back, R., Dominguez, C., Rothé, B., Bobo, C., Beaufile, C., Moréra, S., Meyer, P., Charpentier, B., Branlant, C., Allain, F.H.-T., et al. (2013). High-Resolution Structural Analysis Shows How Tah1 Tethers Hsp90 to the R2TP Complex. *Structure* 21, 1834–1847. <https://doi.org/10.1016/j.str.2013.07.024>.

- Bakin, A., Lane, B.G., and Ofengand, J. (2002). Clustering of Pseudouridine Residues Around the Peptidyltransferase Center of Yeast Cytoplasmic and Mitochondrial Ribosomes (American Chemical Society).
- Balakin, A.G., Smith, L., and Fournier, M.J. (1996). The RNA World of the Nucleolus: Two Major Families of Small RNAs Defined by Different Box Elements with Related Functions. *Cell* 86, 823–834. [https://doi.org/10.1016/S0092-8674\(00\)80156-7](https://doi.org/10.1016/S0092-8674(00)80156-7).
- 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 Å Resolution. *Science* 289, 905–920. <https://doi.org/10.1126/science.289.5481.905>.
- Beltrame, M., Henry, Y., and 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 Research* 22, 4057–4065. <https://doi.org/10.1093/nar/22.20.4057>.
- Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011). The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. *Science* 334, 1524–1529. <https://doi.org/10.1126/science.1212642>.
- Bergeron, D., Fafard-Couture, É., and Scott, M.S. (2020). Small nucleolar RNAs: continuing identification of novel members and increasing diversity of their molecular mechanisms of action. *Biochemical Society Transactions* 48, 645–656. <https://doi.org/10.1042/BST20191046>.
- Bizarro, J., Charron, C., Boulon, S., Westman, B., Pradet-Balade, B., Vandermoere, F., Chagot, M.-E., Hallais, M., Ahmad, Y., Leonhardt, H., et al. (2014). Proteomic and 3D structure analyses highlight the C/D box snoRNP assembly mechanism and its control. *J Cell Biol* 207, 463–480. <https://doi.org/10.1083/jcb.201404160>.
- Bleichert, F., Gagnon, K.T., Brown, B.A., Maxwell, E.S., Leschziner, A.E., Unger, V.M., and Baserga, S.J. (2009). A Dimeric Structure for Archaeal Box C/D Small Ribonucleoproteins. *Science* 325, 1384–1387. <https://doi.org/10.1126/science.1176099>.
- Bogard, B., Francastel, C., and Hubé, F. (2021). Systematic Identification and Functional Validation of New snoRNAs in Human Muscle Progenitors. *Non-Coding RNA* 7, 56. <https://doi.org/10.3390/ncrna7030056>.
- Bonangelino, C.J., Chavez, E.M., and Bonifacino, J.S. (2002). Genomic screen for vacuolar protein sorting genes in *Saccharomyces cerevisiae*. *Mol Biol Cell* 13, 2486–2501. <https://doi.org/10.1091/mbc.02-01-0005>.
- Bonnerot, C., Pintard, L., and Lutfalla, G. (2003). Functional Redundancy of Spb1p and a snR52-Dependent Mechanism for the 2'-O-Ribose Methylation of a Conserved rRNA Position in Yeast. *Molecular Cell* 12, 1309–1315. [https://doi.org/10.1016/S1097-2765\(03\)00435-0](https://doi.org/10.1016/S1097-2765(03)00435-0).
- Boulon, S., Marmier-Gourrier, N., Pradet-Balade, B., Wurth, L., Verheggen, C., Jády, B.E., Rothé, B., Pescia, C., Robert, M.-C., Kiss, T., et al. (2008). The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *Journal of Cell Biology* 180, 579–595. <https://doi.org/10.1083/jcb.200708110>.
- 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. *Molecular Cell* 39, 912–924. <https://doi.org/10.1016/j.molcel.2010.08.023>.

Bragantini, B. (2016a). Caractérisation structurale et fonctionnelle de la protéine Bcd1, impliquée dans la biogenèse des snoRNP à boîtes C/D chez la levure *Saccharomyces cerevisiae*. thesis. Université de Lorraine.

Bragantini, B. (2016b). Caractérisation structurale et fonctionnelle de la protéine Bcd1, impliquée dans la biogenèse des snoRNP à boîtes C/D chez la levure *Saccharomyces cerevisiae*. thesis. Université de Lorraine.

Bragantini, B., Tiotiu, D., Rothé, B., Saliou, J.-M., Marty, H., Cianférani, S., Charpentier, B., Quinternet, M., and Manival, X. (2016). Functional and Structural Insights of the Zinc-Finger HIT protein family members Involved in Box C/D snoRNP Biogenesis. *Journal of Molecular Biology* 428, 2488–2506. <https://doi.org/10.1016/j.jmb.2016.04.028>.

Bragantini, B., Charron, C., Bourguet, M., Paul, A., Tiotiu, D., Rothé, B., Marty, H., Terral, G., Hessmann, S., Decourty, L., et al. (2021). The box C/D snoRNP assembly factor Bcd1 interacts with the histone chaperone Rtt106 and controls its transcription dependent activity. *Nat Commun* 12, 1859. <https://doi.org/10.1038/s41467-021-22077-4>.

Brimacombe, R., Mitchell, P., Osswald, M., Stade, K., and Bochkariov, D. (1993). Clustering of modified nucleotides at the functional center of bacterial ribosomal RNA. *The FASEB Journal* 7, 161–167. <https://doi.org/10.1096/fasebj.7.1.8422963>.

Cavaillé, J. (2017). Box C/D small nucleolar RNA genes and the Prader-Willi syndrome: a complex interplay. *WIREs RNA* 8, e1417. <https://doi.org/10.1002/wrna.1417>.

Cavaillé, J., Hadjiolov, A.A., and Bachellerie, J.-P. (1996). Processing of Mammalian rRNA Precursors at the 3' End of 18S rRNA. *European Journal of Biochemistry* 242, 206–213. <https://doi.org/10.1111/j.1432-1033.1996.0206r.x>.

Cavaillé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.-P., Brosius, J., and Hüttenhofer, A. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proceedings of the National Academy of Sciences* 97, 14311–14316. <https://doi.org/10.1073/pnas.250426397>.

Chagot, M.-E., Quinternet, M., Rothé, B., Charpentier, B., Coutant, J., Manival, X., and Lebars, I. (2019). The yeast C/D box snoRNA U14 adopts a “weak” K-turn like conformation recognized by the Snu13 core protein in solution. *Biochimie* 164, 70–82. <https://doi.org/10.1016/j.biochi.2019.03.014>.

Charette, M., and Gray, M.W. (2000). Pseudouridine in RNA: What, Where, How, and Why. *IUBMB Life* 49, 341–351. <https://doi.org/10.1080/152165400410182>.

Chen, Y.S., Tracy, S., Harjono, V., Xu, F., Moresco, J.J., Yates, J.R., and Zid, B.M. (2021). Rvb1/Rvb2 proteins couple transcription and translation during glucose starvation.

Cheung, K.L.Y., Huen, J., Kakihara, Y., Houry, W.A., and Ortega, J. (2010). Alternative oligomeric states of the yeast Rvb1/Rvb2 complex induced by histidine tags. *J. Mol. Biol.* 404, 478–492. <https://doi.org/10.1016/j.jmb.2010.10.003>.

Choesmel, V., Fribourg, S., Aguisa-Touré, A.-H., Pinaud, N., Legrand, P., Gazda, H.T., and Gleizes, P.-E. (2008). Mutation of ribosomal protein RPS24 in Diamond-Blackfan anemia results in a ribosome biogenesis disorder. *Human Molecular Genetics* 17, 1253–1263. <https://doi.org/10.1093/hmg/ddn015>.

Cloutier, P., Poitras, C., Durand, M., Hekmat, O., Fiola-Masson, É., Bouchard, A., Faubert, D., Chabot, B., and Coulombe, B. (2017). R2TP/Prefoldin-like component RUVBL1/RUVBL2 directly interacts

with ZNHIT2 to regulate assembly of U5 small nuclear ribonucleoprotein. *Nat Commun* 8, 15615. <https://doi.org/10.1038/ncomms15615>.

Colau, G., Thiry, M., Leduc, V., Bordonné, R., and Lafontaine, D.L.J. (2004). The Small Nucleolar RNA Cap Trimethyltransferase Is Required for Ribosome Synthesis and Intact Nucleolar Morphology. *Molecular and Cellular Biology* 24, 7976–7986. <https://doi.org/10.1128/MCB.24.18.7976-7986.2004>.

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. *Journal of Cell Biology* 173, 207–218. <https://doi.org/10.1083/jcb.200601105>.

Decatur, W.A., and Fournier, M.J. (2002a). rRNA modifications and ribosome function. *Trends in Biochemical Sciences* 27, 344–351. [https://doi.org/10.1016/S0968-0004\(02\)02109-6](https://doi.org/10.1016/S0968-0004(02)02109-6).

Decatur, W.A., and Fournier, M.J. (2002b). rRNA modifications and ribosome function. *Trends in Biochemical Sciences* 27, 344–351. [https://doi.org/10.1016/S0968-0004\(02\)02109-6](https://doi.org/10.1016/S0968-0004(02)02109-6).

Decatur, W.A., and Schnare, M.N. (2008). Different Mechanisms for Pseudouridine Formation in Yeast 5S and 5.8S rRNAs. *Molecular and Cellular Biology* 28, 3089–3100. <https://doi.org/10.1128/MCB.01574-07>.

Dobbyn, H.C., and O’keefe, R.T. (2004). Analysis of Snu13p mutations reveals differential interactions with the U4 snRNA and U3 snoRNA. *RNA* 10, 308–320. <https://doi.org/10.1261/rna.5970404>.

Dragon, F., Gallagher, J.E.G., Compagnone-Post, P.A., Mitchell, B.M., Porwancher, K.A., Wehner, K.A., Wormsley, S., Settlage, R.E., Shabanowitz, J., Osheim, Y., et al. (2002). A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* 417, 967–970. <https://doi.org/10.1038/nature00769>.

Dubin, D.T., and Günalp, A. (1967). Minor nucleotide composition of ribosomal precursor, and ribosomal, ribonucleic acid in *Escherichia coli*. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* 134, 106–123. [https://doi.org/10.1016/0005-2787\(67\)90094-9](https://doi.org/10.1016/0005-2787(67)90094-9).

Edward, B., and Maden, H. (1998). Eukaryotic rRNA methylation: the calm before the Sno storm. *Trends in Biochemical Sciences* 23, 447–450. [https://doi.org/10.1016/S0968-0004\(98\)01304-8](https://doi.org/10.1016/S0968-0004(98)01304-8).

Eickhoff, P., and Costa, A. (2017). Escorting Client Proteins to the Hsp90 Molecular Chaperone. *Structure* 25, 964–965. <https://doi.org/10.1016/j.str.2017.06.007>.

Esguerra, J., Warringer, J., and Blomberg, A. (2008). Functional importance of individual rRNA 2’-O-ribose methylations revealed by high-resolution phenotyping. *RNA* 14, 649–656. <https://doi.org/10.1261/rna.845808>.

Ewens, C.A., Su, M., Zhao, L., Nano, N., Houry, W.A., and Southworth, D.R. (2016). Architecture and Nucleotide-Dependent Conformational Changes of the Rvb1-Rvb2 AAA+ Complex Revealed by Cryoelectron Microscopy. *Structure* 24, 657–666. <https://doi.org/10.1016/j.str.2016.03.018>.

Fafard-Couture, É., Bergeron, D., Couture, S., Abou-Elela, S., and Scott, M.S. (2021). Annotation of snoRNA abundance across human tissues reveals complex snoRNA-host gene relationships. *Genome Biol* 22, 172. <https://doi.org/10.1186/s13059-021-02391-2>.

Fatica, A., Oeffinger, M., Tollervey, D., and Bozzoni, I. (2003). Cic1p/Nsa3p is required for synthesis and nuclear export of 60S ribosomal subunits. *RNA* 9, 1431–1436. <https://doi.org/10.1261/rna.5130503>.

Feder, M., Pas, J., Wyrwicz, L.S., and Bujnicki, J.M. (2003). Molecular phylogenetics of the RrmJ/fibrillarin superfamily of ribose 2'-O-methyltransferases. *Gene* 302, 129–138. [https://doi.org/10.1016/S0378-1119\(02\)01097-1](https://doi.org/10.1016/S0378-1119(02)01097-1).

Feng, Y., Lee, N., and Fearon, E.R. (2003). TIP49 Regulates β -Catenin-Mediated Neoplastic Transformation and T-Cell Factor Target Gene Induction via Effects on Chromatin Remodeling. *Cancer Res* 63, 8726–8734. .

Filipowicz, W., Pelczar, P., Pogacic, V., and Dragon, F. (1999). Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification. *Acta Biochimica Polonica* 46, 377–389. https://doi.org/10.18388/abp.1999_4171.

Freed, E.F., Bleichert, F., Dutca, L.M., and Baserga, S.J. (2010). When ribosomes go bad: diseases of ribosome biogenesis. *Mol. Biosyst.* 6, 481–493. <https://doi.org/10.1039/B919670F>.

Ganapathi, K.A., Austin, K.M., Lee, C.-S., Dias, A., Malsch, M.M., Reed, R., and Shimamura, A. (2007). The human Shwachman-Diamond syndrome protein, SBDS, associates with ribosomal RNA. *Blood* 110, 1458–1465. <https://doi.org/10.1182/blood-2007-02-075184>.

Ganot, P., Bortolin, M.-L., and Kiss, T. (1997). Site-Specific Pseudouridine Formation in Preribosomal RNA Is Guided by Small Nucleolar RNAs. *Cell* 89, 799–809. [https://doi.org/10.1016/S0092-8674\(00\)80263-9](https://doi.org/10.1016/S0092-8674(00)80263-9).

Gartner, W., Rossbacher, J., Zierhut, B., Daneva, T., Base, W., Weissel, M., Waldhäusl, W., Pasternack, M.S., and Wagner, L. (2003). The ATP-dependent helicase RUVBL1/TIP49a associates with tubulin during mitosis. *Cell Motility* 56, 79–93. <https://doi.org/10.1002/cm.10136>.

Gautier, T., Bergès, T., Tollervey, D., and Hurt, E. (1997). Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Molecular and Cellular Biology* 17, 7088–7098. <https://doi.org/10.1128/MCB.17.12.7088>.

Gazda, H.T., Sheen, M.R., Vlachos, A., Choesmel, V., O'Donohue, M.-F., Schneider, H., Darras, N., Hasman, C., Sieff, C.A., Newburger, P.E., et al. (2008). Ribosomal Protein L5 and L11 Mutations Are Associated with Cleft Palate and Abnormal Thumbs in Diamond-Blackfan Anemia Patients. *The American Journal of Human Genetics* 83, 769–780. <https://doi.org/10.1016/j.ajhg.2008.11.004>.

Geerlings, T.H., Vos, J.C., and Raué, H.A. (2000). The final step in the formation of 25S rRNA in *Saccharomyces cerevisiae* is performed by 5' \rightarrow 3' exonucleases. *RNA* 6, 1698–1703. <https://doi.org/10.1017/S1355838200001540>.

Gillot-Chafia, S. (2018). Impact de la composition du ribosome sur la fidélité de la traduction. phdthesis. Université Paris Saclay (COMUE).

Gingras, A.-C., Raught, B., and Sonenberg, N. (2001). Regulation of translation initiation by FRAP/mTOR. *Genes Dev.* 15, 807–826. <https://doi.org/10.1101/gad.887201>.

Gonzales, F.A., Zanchin, N.I.T., Luz, J.S., and Oliveira, C.C. (2005). Characterization of *Saccharomyces cerevisiae* Nop17p, a Novel Nop58p-Interacting Protein that is Involved in Pre-rRNA Processing. *Journal of Molecular Biology* 346, 437–455. <https://doi.org/10.1016/j.jmb.2004.11.071>.

Goody, T.A., Melcher, S.E., Norman, D.G., and Lilley, D.M.J. (2004). The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA* 10, 254–264. <https://doi.org/10.1261/rna.5176604>.

Granneman, S., Pruijn, G.J.M., Horstman, W., Venrooij, W.J. van, Lührmann, 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 *. *Journal of Biological Chemistry* 277, 48490–48500. <https://doi.org/10.1074/jbc.M206631200>.

Gribun, A., Cheung, K.L.Y., Huen, J., Ortega, J., and Houry, W.A. (2008). Yeast Rvb1 and Rvb2 are ATP-Dependent DNA Helicases that Form a Heterohexameric Complex. *Journal of Molecular Biology* 376, 1320–1333. <https://doi.org/10.1016/j.jmb.2007.12.049>.

Grzechnik, P., Szczepaniak, S.A., Dhir, S., Pastucha, A., Parslow, H., Matuszek, Z., Mischo, H.E., Kufel, J., and Proudfoot, N.J. (2018). Nuclear fate of yeast snoRNA is determined by co-transcriptional Rnt1 cleavage. *Nat Commun* 9, 1783. <https://doi.org/10.1038/s41467-018-04094-y>.

Guffanti, E., Percudani, R., Harismendy, O., Soutourina, J., Werner, M., Iacovella, M.G., Negri, R., and Dieci, G. (2006). Nucleosome Depletion Activates Poised RNA Polymerase III at Unconventional Transcription Sites in *Saccharomyces cerevisiae* *. *Journal of Biological Chemistry* 281, 29155–29164. <https://doi.org/10.1074/jbc.M600387200>.

Hager, J., Staker, B.L., Bügl, H., and Jakob, U. (2002). Active Site in RrmJ, a Heat Shock-induced Methyltransferase *. *Journal of Biological Chemistry* 277, 41978–41986. <https://doi.org/10.1074/jbc.M205423200>.

Hayashi, Y., Osawa, S., and Miura, K. (1966). The methyl groups in ribosomal RNA from *Escherichia coli*. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* 129, 519–531. [https://doi.org/10.1016/0005-2787\(66\)90067-0](https://doi.org/10.1016/0005-2787(66)90067-0).

Heiss, N.S., Knight, S.W., Vulliamy, T.J., Klauck, S.M., Wiemann, S., Mason, P.J., Poustka, A., and 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. <https://doi.org/10.1038/ng0598-32>.

Henras, A., Henry, Y., Bousquet-Antonelli, C., Noaillac-Depeyre, J., Gélugne, J. P., & Caizergues-Ferrer, M. (1998). Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *The EMBO journal*, 17(23), 7078-7090.

Henras, A.K., Dez, C., and Henry, Y. (2004). RNA structure and function in C/D and H/ACA s(no)RNPs. *Curr Opin Struct Biol* 14, 335–343. <https://doi.org/10.1016/j.sbi.2004.05.006>.

Henras, A.K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A., and Gleizes, P.-E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *WIREs RNA* 6, 225–242. <https://doi.org/10.1002/wrna.1269>.

Henri, J., Chagot, M.-E., Bourguet, M., Abel, Y., Terral, G., Maurizy, C., Aigueperse, C., Georgescauld, F., Vandermoere, F., Saint-Fort, R., et al. (2018). Deep Structural Analysis of RPAP3 and PIH1D1, Two Components of the HSP90 Co-chaperone R2TP Complex. *Structure* 26, 1196-1209.e8. <https://doi.org/10.1016/j.str.2018.06.002>.

Henry, Y., Wood, H., Morrissey, J. P., Petfalski, E., Kearsey, S., & Tollervey, D. (1994). The 5' end of yeast 5.8 S rRNA is generated by exonucleases from an upstream cleavage site. *The EMBO journal*, 13(10), 2452-2463.

Hiley, S.L., Babak, T., and Hughes, T.R. (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 Research* 33, 3048–3056. <https://doi.org/10.1093/nar/gki608>.

Hirose, T., Shu, M.-D., and Steitz, J.A. (2003). Splicing-Dependent and -Independent Modes of Assembly for Intron-Encoded Box C/D snoRNPs in Mammalian Cells. *Molecular Cell* 12, 113–123. [https://doi.org/10.1016/S1097-2765\(03\)00267-3](https://doi.org/10.1016/S1097-2765(03)00267-3).

Hirose, T., Ideue, T., Nagai, M., Hagiwara, M., Shu, M.-D., and Steitz, J.A. (2006). A Spliceosomal Intron Binding Protein, IBP160, Links Position-Dependent Assembly of Intron-Encoded Box C/D snoRNP to Pre-mRNA Splicing. *Molecular Cell* 23, 673–684. <https://doi.org/10.1016/j.molcel.2006.07.011>.

Höfler, S., Lukat, P., Blankenfeldt, W., and Carlomagno, T. (2021). High-resolution structure of eukaryotic Fibrillarin interacting with Nop56 amino-terminal domain. *RNA* 27, 496–512. <https://doi.org/10.1261/rna.077396.120>.

Houry, W.A., Bertrand, E., and Coulombe, B. (2018). The PAQosome, an R2TP-Based Chaperone for Quaternary Structure Formation. *Trends in Biochemical Sciences* 43, 4–9. <https://doi.org/10.1016/j.tibs.2017.11.001>.

Huang, C., Wu, D., Khan, F.A., Jiao, X., Guan, K., and Huo, L. (2016). The GTPase SPAG-1 orchestrates meiotic program by dictating meiotic resumption and cytoskeleton architecture in mouse oocytes. *MBoC* 27, 1776–1785. <https://doi.org/10.1091/mbc.e16-02-0132>.

Huber, O., Ménard, L., Haurie, V., Nicou, A., Taras, D., and Rosenbaum, J. (2008). Pontin and Reptin, Two Related ATPases with Multiple Roles in Cancer. *Cancer Res* 68, 6873–6876. <https://doi.org/10.1158/0008-5472.CAN-08-0547>.

Hughes, J. M. X., & Ares Jr, M. (1991). Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *The EMBO journal*, 10(13), 4231–4239.

Imbeault, D., Gamar, L., Rufiange, A., Paquet, E., and Nourani, A. (2008). The Rtt106 Histone Chaperone Is Functionally Linked to Transcription Elongation and Is Involved in the Regulation of Spurious Transcription from Cryptic Promoters in Yeast *♦. *Journal of Biological Chemistry* 283, 27350–27354. <https://doi.org/10.1074/jbc.C800147200>.

Itsuki, Y., Saeki, M., Nakahara, H., Egusa, H., Irie, Y., Terao, Y., Kawabata, S., Yatani, H., and Kamisaki, Y. (2008). Molecular cloning of novel Monad binding protein containing tetratricopeptide repeat domains. *FEBS Letters* 582, 2365–2370. <https://doi.org/10.1016/j.febslet.2008.05.041>.

Izumi, N., Yamashita, A., Iwamatsu, A., Kurata, R., Nakamura, H., Saari, B., Hirano, H., Anderson, P., and Ohno, S. (2010). AAA+ Proteins RUVBL1 and RUVBL2 Coordinate PIKK Activity and Function in Nonsense-Mediated mRNA Decay. *Science Signaling* 3, ra27–ra27. <https://doi.org/10.1126/scisignal.2000468>.

Jády, B.E., Bertrand, E., and Kiss, T. (2004). Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *J Cell Biol* 164, 647–652. <https://doi.org/10.1083/jcb.200310138>.

Jeanteur, P., Amaldi, F., and Attardi, G. (1968). Partial sequence analysis of ribosomal RNA from HeLa cells: II. Evidence for sequences of non-ribosomal type in 45 and 32 s ribosomal RNA precursors. *Journal of Molecular Biology* 33, 757–775. [https://doi.org/10.1016/0022-2836\(68\)90318-5](https://doi.org/10.1016/0022-2836(68)90318-5).

Jeganathan, A., Leong, V., Zhao, L., Huen, J., Nano, N., Houry, W.A., and Ortega, J. (2015). Yeast Rvb1 and Rvb2 Proteins Oligomerize As a Conformationally Variable Dodecamer with Low Frequency. *Journal of Molecular Biology* 427, 1875–1886. <https://doi.org/10.1016/j.jmb.2015.01.010>.

Jeppesen, C., Stebbins-Boaz, B., and Gerbi, S.A. (1988). Nucleotide sequence determination and secondary structure of *Xenopus* U3 snRNA. *Nucleic Acids Research* 16, 2127–2148. <https://doi.org/10.1093/nar/16.5.2127>.

Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Thérien, C., Bergeron, D., Bourassa, S., Greenblatt, J., et al. (2007). Systematic Analysis of the Protein Interaction Network for the Human Transcription Machinery Reveals the Identity of the 7SK Capping Enzyme. *Molecular Cell* 27, 262–274. <https://doi.org/10.1016/j.molcel.2007.06.027>.

Jong, A.Y., Clark, M.W., Gilbert, M., Oehm, A., and Campbell, J.L. (1987). *Saccharomyces cerevisiae* SSB1 protein and its relationship to nucleolar RNA-binding proteins. *Molecular and Cellular Biology* 7, 2947–2955. <https://doi.org/10.1128/mcb.7.8.2947-2955.1987>.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.

Kakahara, Y., and Houry, W.A. (2012). The R2TP complex: Discovery and functions. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1823, 101–107. <https://doi.org/10.1016/j.bbamcr.2011.08.016>.

Kakahara, Y., Makhnevych, T., Zhao, L., Tang, W., and Houry, W.A. (2014a). Nutritional status modulates box C/D snoRNP biogenesis by regulated subcellular relocalization of the R2TP complex. *Genome Biol* 15, 404. <https://doi.org/10.1186/s13059-014-0404-4>.

Kakahara, Y., Makhnevych, T., Zhao, L., Tang, W., and Houry, W.A. (2014b). Nutritional status modulates box C/D snoRNP biogenesis by regulated subcellular relocalization of the R2TP complex. *Genome Biol* 15, 404. <https://doi.org/10.1186/s13059-014-0404-4>.

Kanemaki, M., Kurokawa, Y., Matsu-ura, T., Makino, Y., Masani, A., Okazaki, K., Morishita, T., and Tamura, T. (1999). TIP49b, a New RuvB-like DNA Helicase, Is Included in a Complex Together with Another RuvB-like DNA Helicase, TIP49a *. *Journal of Biological Chemistry* 274, 22437–22444. <https://doi.org/10.1074/jbc.274.32.22437>.

Kass, S., Tyc, K., Steitz, J.A., and Sollner-Webb, B. (1990). The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. *Cell* 60, 897–908. [https://doi.org/10.1016/0092-8674\(90\)90338-F](https://doi.org/10.1016/0092-8674(90)90338-F).

Khoshnevis, S., Dreggors, R.E., Hoffmann, T.F.R., and Ghalei, H. (2019). A conserved Bcd1 interaction essential for box C/D snoRNP biogenesis. *Journal of Biological Chemistry* 294, 18360–18371. <https://doi.org/10.1074/jbc.RA119.010222>.

King, T.H., Decatur, W.A., Bertrand, E., Maxwell, E.S., and Fournier, M.J. (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. *Molecular and Cellular Biology* 21, 7731–7746. <https://doi.org/10.1128/MCB.21.22.7731-7746.2001>.

- Kiss, T., and Filipowicz, W. (1995). Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes Dev.* 9, 1411–1424. <https://doi.org/10.1101/gad.9.11.1411>.
- Kiss-László, Z., Henry, Y., Bachellerie, J.-P., Caizergues-Ferrer, M., and Kiss, T. (1996). Site-Specific Ribose Methylation of Preribosomal RNA: A Novel Function for Small Nucleolar RNAs. *Cell* 85, 1077–1088. [https://doi.org/10.1016/S0092-8674\(00\)81308-2](https://doi.org/10.1016/S0092-8674(00)81308-2).
- Klein, D. J., Schmeing, T. M., Moore, P. B., & Steitz, T. A. (2001). The kink-turn: a new RNA secondary structure motif. *The EMBO journal*, 20(15), 4214-4221.
- Klootwijk, J., and Planta, R.J. (1973). Analysis of the Methylation Sites in Yeast Ribosomal RNA. *European Journal of Biochemistry* 39, 325–333. <https://doi.org/10.1111/j.1432-1033.1973.tb03130.x>.
- Knowles, M.R., Daniels, L.A., Davis, S.D., Zariwala, M.A., and Leigh, M.W. (2013). Primary Ciliary Dyskinesia. *Recent Advances in Diagnostics, Genetics, and Characterization of Clinical Disease*. *Am J Respir Crit Care Med* 188, 913–922. <https://doi.org/10.1164/rccm.201301-0059CI>.
- Kobor, M.S., Venkatasubrahmanyam, S., Meneghini, M.D., Gin, J.W., Jennings, J.L., Link, A.J., Madhani, H.D., and Rine, J. (2004). A Protein Complex Containing the Conserved Swi2/Snf2-Related ATPase Swr1p Deposits Histone Variant H2A.Z into Euchromatin. *PLOS Biology* 2, e131. <https://doi.org/10.1371/journal.pbio.0020131>.
- Koo, B.-K., Park, C.-J., Fernandez, C.F., Chim, N., Ding, Y., Chanfreau, G., and 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. *Journal of Molecular Biology* 411, 927–942. <https://doi.org/10.1016/j.jmb.2011.06.022>.
- Koonin, E.V., Mushegian, A.R., Galperin, M.Y., and Walker, D.R. (1997). Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Molecular Microbiology* 25, 619–637. <https://doi.org/10.1046/j.1365-2958.1997.4821861.x>.
- Kressler, D., Linder, P., and de la Cruz, J. (1999). Protein trans-Acting Factors Involved in Ribosome Biogenesis in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* 19, 7897–7912. <https://doi.org/10.1128/MCB.19.12.7897>.
- Kressler, D., Hurt, E., and Baßler, J. (2010). Driving ribosome assembly. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1803, 673–683. <https://doi.org/10.1016/j.bbamcr.2009.10.009>.
- Lafontaine, D.L.J. (2015). Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nat Struct Mol Biol* 22, 11–19. <https://doi.org/10.1038/nsmb.2939>.
- Lafontaine, D.L.J., and Tollervey, D. (1998). Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends in Biochemical Sciences* 23, 383–388. [https://doi.org/10.1016/S0968-0004\(98\)01260-2](https://doi.org/10.1016/S0968-0004(98)01260-2).
- Lafontaine, D.L.J., and Tollervey, D. (1999). Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA* 5, 455–467. <https://doi.org/10.1017/S135583829998192X>.
- Lafontaine, D.L.J., Preiss, T., and Tollervey, D. (1998). Yeast 18S rRNA Dimethylase Dim1p: a Quality Control Mechanism in Ribosome Synthesis? *Molecular and Cellular Biology* 18, 2360–2370. <https://doi.org/10.1128/MCB.18.4.2360>.

Lakomek, K., Stoehr, G., Tosi, A., Schmailzl, M., and Hopfner, K.-P. (2015). Structural Basis for Dodecameric Assembly States and Conformational Plasticity of the Full-Length AAA+ ATPases Rvb1-Rvb2. *Structure* 23, 483–495. <https://doi.org/10.1016/j.str.2014.12.015>.

Lamanna, A.C., and Karbstein, K. (2009). Nob1 binds the single-stranded cleavage site D at the 3'-end of 18S rRNA with its PIN domain. *Proceedings of the National Academy of Sciences* 106, 14259–14264. <https://doi.org/10.1073/pnas.0905403106>.

Lane, B.G., Ofengand, J., and Gray, M.W. (1995). Pseudouridine and O²-methylated nucleosides. Significance of their selective occurrence in rRNA domains that function in ribosome-catalyzed synthesis of the peptide bonds in proteins. *Biochimie* 77, 7–15. [https://doi.org/10.1016/0300-9084\(96\)88098-9](https://doi.org/10.1016/0300-9084(96)88098-9).

Lapeyre, B., Bourbon, H., and Amalric, F. (1987). Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. *PNAS* 84, 1472–1476. <https://doi.org/10.1073/pnas.84.6.1472>.

Lapinaite, A., Simon, B., Skjaerven, L., Rakwalska-Bange, M., Gabel, F., and Carlomagno, T. (2013). The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* 502, 519–523. <https://doi.org/10.1038/nature12581>.

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. <https://doi.org/10.1038/nature09688>.

Lin, W., Zhou, X., Zhang, M., Li, Y., Miao, S., Wang, L., Zong, S., and Koide, S.S. (2001). Expression and function of the HSD-3.8 gene encoding a testis-specific protein. *Molecular Human Reproduction* 7, 811–818. <https://doi.org/10.1093/molehr/7.9.811>.

Liu, N., Qiao, Y., Cai, C., Lin, W., Zhang, J., Miao, S., Zong, S., Koide, S.S., and Wang, L. (2006). A sperm component, HSD-3.8 (SPAG1), interacts with G-protein beta 1 subunit and activates extracellular signal-regulated kinases (ERK). *Front Biosci* 11, 1679–1689. <https://doi.org/10.2741/1913>.

Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Lührmann, 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. <https://doi.org/10.1126/science.1137924>.

Maden, B.E.H., and Salim, M. (1974). The methylated nucleotide sequences in HeLa cell ribosomal RNA and its precursors. *Journal of Molecular Biology* 88, 133–152. [https://doi.org/10.1016/0022-2836\(74\)90299-X](https://doi.org/10.1016/0022-2836(74)90299-X).

Maimoni Campanella, J.E., Ramos Junior, S.L., Rodrigues Kiraly, V.T., Severo Gomes, A.A., de Barros, A.C., Mateos, P.A., Freitas, F.Z., de Mattos Fontes, M.R., Borges, J.C., and Bertolini, M.C. (2021). Biochemical and biophysical characterization of the RVB-1/RVB-2 protein complex, the RuvBL/RVB homologues in *Neurospora crassa*. *Biochimie* 191, 11–26. <https://doi.org/10.1016/j.biochi.2021.08.002>.

Makhnevych, T., and Houry, W.A. (2012). The role of Hsp90 in protein complex assembly. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1823, 674–682. <https://doi.org/10.1016/j.bbamcr.2011.09.001>.

Manival, X., Charron, C., Fourmann, J.-B., Godard, F., Charpentier, B., and Branlant, C. (2006). Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10

complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Research* *34*, 826–839. <https://doi.org/10.1093/nar/gkj482>.

Mao, H., White, S.A., and Williamson, J.R. (1999). A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nat Struct Mol Biol* *6*, 1139–1147. <https://doi.org/10.1038/70081>.

Mao, X., Schwer, B., and Shuman, S. (1995). Yeast mRNA cap methyltransferase is a 50-kilodalton protein encoded by an essential gene. *Molecular and Cellular Biology* *15*, 4167–4174. <https://doi.org/10.1128/MCB.15.8.4167>.

Marcel, V., Ghayad, S.E., Belin, S., Therizols, G., Morel, A.-P., Solano-González, E., Vendrell, J.A., Hacot, S., Mertani, H.C., Albaret, M.A., et al. (2013). p53 Acts as a Safeguard of Translational Control by Regulating Fibrillarin and rRNA Methylation in Cancer. *Cancer Cell* *24*, 318–330. <https://doi.org/10.1016/j.ccr.2013.08.013>.

Marmier-Gourrier, N., Cléry, A., Senty-Ségault, 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. <https://doi.org/10.1261/rna.2130503>.

Matias, P.M., Gorynia, S., Donner, P., and Carrondo, M.A. (2006). Crystal Structure of the Human AAA+ Protein RuvBL1 *. *Journal of Biological Chemistry* *281*, 38918–38929. <https://doi.org/10.1074/jbc.M605625200>.

Maurizy, C., Quinternet, M., Abel, Y., Verheggen, C., Santo, P.E., Bourguet, M., C. F. Paiva, A., Bragantini, B., Chagot, M.-E., Robert, M.-C., et al. (2018). The RPAP3-Cterminal domain identifies R2TP-like quaternary chaperones. *Nat Commun* *9*, 2093. <https://doi.org/10.1038/s41467-018-04431-1>.

Maxwell, E., and Fournier, M. (1995). THE SMALL NUCLEOLAR RNAs. *Annual Review of Biochemistry* *64*, 897–934. <https://doi.org/10.1146/annurev.bi.64.070195.004341>.

Mayer, C., and Grummt, I. (2006). Ribosome biogenesis and cell growth: mTOR coordinates transcription by all three classes of nuclear RNA polymerases. *Oncogene* *25*, 6384–6391. <https://doi.org/10.1038/sj.onc.1209883>.

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. *Molecular and Cellular Biology* *27*, 6782–6793. <https://doi.org/10.1128/MCB.01097-07>.

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. *Molecular and Cellular Biology* <https://doi.org/10.1128/MCB.00752-09>.

Mei, Y.-P., Liao, J.-P., Shen, J., Yu, L., Liu, B.-L., Liu, L., Li, R.-Y., Ji, L., Dorsey, S.G., Jiang, Z.-R., et al. (2012). Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. *Oncogene* *31*, 2794–2804. <https://doi.org/10.1038/onc.2011.449>.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S., and Wu, C. (2004). ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. *Science* <https://doi.org/10.1126/science.1090701>.

Moore, T., Zhang, Y., Fenley, M.O., and Li, H. (2004). Molecular Basis of Box C/D RNA-Protein Interactions: Cocystal Structure of Archaeal L7Ae and a Box C/D RNA. *Structure* *12*, 807–818. <https://doi.org/10.1016/j.str.2004.02.033>.

Morgan, R.M.L., Pal, M., Roe, S.M., Pearl, L.H., and Prodromou, C. (2015). Tah1 helix-swap dimerization prevents mixed Hsp90 co-chaperone complexes. *Acta Cryst D* 71, 1197–1206. <https://doi.org/10.1107/S1399004715004551>.

Morrissey, J.P., and Tollervey, D. (1993). Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. *Molecular and Cellular Biology* 13, 2469–2477. <https://doi.org/10.1128/mcb.13.4.2469-2477.1993>.

Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J., and Bordonné, R. (2002). Hypermethylation of the Cap Structure of Both Yeast snRNAs and snoRNAs Requires a Conserved Methyltransferase that Is Localized to the Nucleolus. *Molecular Cell* 9, 891–901. [https://doi.org/10.1016/S1097-2765\(02\)00484-7](https://doi.org/10.1016/S1097-2765(02)00484-7).

Muñoz-Hernández, H., Pal, M., Rodríguez, C.F., Fernandez-Leiro, R., Prodromou, C., Pearl, L.H., and Llorca, O. (2019). Structural mechanism for regulation of the AAA-ATPases RUVBL1-RUVBL2 in the R2TP co-chaperone revealed by cryo-EM. *Science Advances* <https://doi.org/10.1126/sciadv.aaw1616>.

Nano, N., and Houry, W.A. (2013). Chaperone-like activity of the AAA+ proteins Rvb1 and Rvb2 in the assembly of various complexes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20110399. <https://doi.org/10.1098/rstb.2011.0399>.

Narayanan, A., Meriin, A., Andrews, J.O., Spille, J.-H., Sherman, M.Y., and Cisse, I.I. (2019). A first order phase transition mechanism underlies protein aggregation in mammalian cells. *ELife* 8, e39695. <https://doi.org/10.7554/eLife.39695>.

Narla, A., and Ebert, B.L. (2010). Ribosomopathies: human disorders of ribosome dysfunction. *Blood* 115, 3196–3205. <https://doi.org/10.1182/blood-2009-10-178129>.

Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E.V. (1999). AAA+: A Class of Chaperone-Like ATPases Associated with the Assembly, Operation, and Disassembly of Protein Complexes. *Genome Res.* 9, 27–43. <https://doi.org/10.1101/gr.9.1.27>.

Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* 289, 920–930. <https://doi.org/10.1126/science.289.5481.920>.

Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R., & Lührmann, R. (1999). Functional interaction of a novel 15.5 kD [U4/U6·U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *The EMBO journal*, 18(21), 6119–6133.

Ochs, R.L., Lischwe, M.A., Spohn, W.H., and Busch, H. (1985). Fibrillarin: a new protein of the nucleolus identified by autoimmune sera. *Biology of the Cell* 54, 123–133. <https://doi.org/10.1111/j.1768-322X.1985.tb00387.x>.

Oeffinger, M., Zenklusen, D., Ferguson, A., Wei, K.E., El Hage, A., Tollervey, D., Chait, B.T., Singer, R.H., and Rout, M.P. (2009). Rpl17p Is a Eukaryotic Exonuclease Required for 5' End Processing of Pre-60S Ribosomal RNA. *Molecular Cell* 36, 768–781. <https://doi.org/10.1016/j.molcel.2009.11.011>.

Ofengand, J., and Bakin, A. (1997). Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaeobacteria, mitochondria and chloroplasts. Edited by D. E. Draper. *Journal of Molecular Biology* 266, 246–268. <https://doi.org/10.1006/jmbi.1996.0737>.

Ofengand, J., Bakin, A., Wrzesinski, J., Nurse, K., and Lane, B.G. (1995). The pseudouridine residues of ribosomal RNA. *Biochem. Cell Biol.* *73*, 915–924. <https://doi.org/10.1139/o95-099>.

Ohta, T., Gray, T.A., Rogan, P.K., Buiting, K., Gabriel, J.M., Saitoh, S., Muralidhar, B., Bilienska, B., Krajewska-Walasek, M., Driscoll, D.J., et al. (1999). Imprinting-Mutation Mechanisms in Prader-Willi Syndrome. *The American Journal of Human Genetics* *64*, 397–413. <https://doi.org/10.1086/302233>.

Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R., and Dennis, P.P. (2000). Homologs of Small Nucleolar RNAs in Archaea. *Science* *288*, 517–522. <https://doi.org/10.1126/science.288.5465.517>.

Oruganti, S. vidya, Zhang, Y., and Li, H. (2005). Structural comparison of yeast snoRNP and spliceosomal protein Snu13p with its homologs. *Biochemical and Biophysical Research Communications* *333*, 550–554. <https://doi.org/10.1016/j.bbrc.2005.05.141>.

Pal, M., Morgan, M., Phelps, S.E.L., Roe, S.M., Parry-Morris, S., Downs, J.A., Polier, S., Pearl, L.H., and Prodromou, C. (2014). Structural Basis for Phosphorylation-Dependent Recruitment of Tel2 to Hsp90 by Pih1. *Structure* *22*, 805–818. <https://doi.org/10.1016/j.str.2014.04.001>.

Pannevis, M.C., and Houlihan, D.F. (1992). The energetic cost of protein synthesis in isolated hepatocytes of rainbow trout (*Oncorhynchus mykiss*). *J Comp Physiol B* *162*, 393–400. <https://doi.org/10.1007/BF00258960>.

Parker, R., Simmons, T., Shuster, E.O., Siliciano, P.G., and Guthrie, C. (1988). Genetic analysis of small nuclear RNAs in *Saccharomyces cerevisiae*: viable sextuple mutant. *Molecular and Cellular Biology* *8*, 3150–3159. <https://doi.org/10.1128/mcb.8.8.3150-3159.1988>.

Paul, A. (2018). Rôle de la protéine Bcd1p/BCD1 dans les étapes précoces de la biogenèse des snoRNP à boîtes C/D eucaryotes. phdthesis. Université de Lorraine.

Paul, A., Tiotiu, D., Bragantini, B., Marty, H., Charpentier, B., Massenet, S., and Labialle, S. (2019a). Bcd1p controls RNA loading of the core protein Nop58 during C/D box snoRNP biogenesis. *RNA* *25*, 496–506. <https://doi.org/10.1261/rna.067967.118>.

Paul, A., Tiotiu, D., Bragantini, B., Marty, H., Charpentier, B., Massenet, S., and Labialle, S. (2019b). Bcd1p controls RNA loading of the core protein Nop58 during C/D box snoRNP biogenesis. *RNA* *25*, 496–506. <https://doi.org/10.1261/rna.067967.118>.

Pearl, L.H., and Prodromou, C. (2006). Structure and Mechanism of the Hsp90 Molecular Chaperone Machinery. *Annual Review of Biochemistry* *75*, 271–294. <https://doi.org/10.1146/annurev.biochem.75.103004.142738>.

Peng, W.-T., Robinson, M.D., Mnaimneh, S., Krogan, N.J., Cagney, G., Morris, Q., Davierwala, A.P., Grigull, J., Yang, X., Zhang, W., et al. (2003). A Panoramic View of Yeast Noncoding RNA Processing. *Cell* *113*, 919–933. [https://doi.org/10.1016/S0092-8674\(03\)00466-5](https://doi.org/10.1016/S0092-8674(03)00466-5).

Penzo, M., Montanaro, L., Tréré, D., and Derenzini, M. (2019). The Ribosome Biogenesis—Cancer Connection. *Cells* *8*, 55. <https://doi.org/10.3390/cells8010055>.

Petfalski, E., Dandekar, T., Henry, Y., and Tollervy, D. (1998). Processing of the Precursors to Small Nucleolar RNAs and rRNAs Requires Common Components. *Molecular and Cellular Biology* *18*, 1181–1189. <https://doi.org/10.1128/MCB.18.3.1181>.

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. <https://doi.org/10.1261/rna.373107>.

Pradet-Balade, B., Girard, C., Boulon, S., Paul, C., Azzag, K., Bordonné, R., ... & Verheggen, C. (2011). CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *The EMBO journal*, *30*(11), 2205-2218.

Puri, A., Singh, P., Kumar, N., Kumar, R., and Sharma, D. (2021). Tah1, A Key Component of R2TP Complex that Regulates Assembly of snoRNP, is Involved in De Novo Generation and Maintenance of Yeast Prion [URE3]. *Journal of Molecular Biology* *433*, 166976. <https://doi.org/10.1016/j.jmb.2021.166976>.

Puri, T., Wendler, P., Sigala, B., Saibil, H., and Tsaneva, I.R. (2007). Dodecameric structure and ATPase activity of the human TIP48/TIP49 complex. *J. Mol. Biol.* *366*, 179–192. <https://doi.org/10.1016/j.jmb.2006.11.030>.

Quinternet, M., Rothé, B., Barbier, M., Bobo, C., Saliou, J.-M., Jacquemin, C., Back, R., Chagot, M.-E., Cianférani, S., Meyer, P., et al. (2015). Structure/Function Analysis of Protein–Protein Interactions Developed by the Yeast Pih1 Platform Protein and Its Partners in Box C/D snoRNP Assembly. *Journal of Molecular Biology* *427*, 2816–2839. <https://doi.org/10.1016/j.jmb.2015.07.012>.

Richard, P., Darzacq, X., Bertrand, E., Jády, B.E., Verheggen, C., and Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J* *22*, 4283–4293. <https://doi.org/10.1093/emboj/cdg394>.

Rivera-Calzada, A., Pal, M., Muñoz-Hernández, H., Luque-Ortega, J.R., Gil-Carton, D., Degliesposti, G., Skehel, J.M., Prodromou, C., Pearl, L.H., and Llorca, O. (2017). The Structure of the R2TP Complex Defines a Platform for Recruiting Diverse Client Proteins to the HSP90 Molecular Chaperone System. *Structure* *25*, 1145-1152.e4. <https://doi.org/10.1016/j.str.2017.05.016>.

Robledo, S., Idol, R.A., Crimmins, D.L., Ladenson, J.H., Mason, P.J., and Bessler, M. (2008). The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *RNA* *14*, 1918–1929. <https://doi.org/10.1261/rna.1132008>.

Rothé, B., Saliou, J.-M., Quinternet, M., Back, R., Tiotiu, D., Jacquemin, C., Loegler, C., Schlotter, F., Peña, V., Eckert, K., et al. (2014a). Protein Hit1, a novel box C/D snoRNP assembly factor, controls cellular concentration of the scaffolding protein Rsa1 by direct interaction. *Nucleic Acids Research* *42*, 10731–10747. <https://doi.org/10.1093/nar/gku612>.

Rothé, B., Saliou, J.-M., Quinternet, M., Back, R., Tiotiu, D., Jacquemin, C., Loegler, C., Schlotter, F., Peña, V., Eckert, K., et al. (2014b). Protein Hit1, a novel box C/D snoRNP assembly factor, controls cellular concentration of the scaffolding protein Rsa1 by direct interaction. *Nucleic Acids Research* *42*, 10731–10747. <https://doi.org/10.1093/nar/gku612>.

Rothé, B., Manival, X., Rolland, N., Charron, C., Senty-Ségault, V., Branlant, C., and Charpentier, B. (2017). Implication of the box C/D snoRNP assembly factor Rsa1p in U3 snoRNP assembly. *Nucleic Acids Research* *45*, 7455–7473. <https://doi.org/10.1093/nar/gkx424>.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F.U., and Moarefi, I. (2000a). Structure of TPR Domain–Peptide Complexes: Critical Elements in the Assembly of the Hsp70–Hsp90 Multichaperone Machine. *Cell* *101*, 199–210. [https://doi.org/10.1016/S0092-8674\(00\)80830-2](https://doi.org/10.1016/S0092-8674(00)80830-2).

- Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F.U., and Moarefi, I. (2000b). Structure of TPR Domain–Peptide Complexes: Critical Elements in the Assembly of the Hsp70–Hsp90 Multichaperone Machine. *Cell* 101, 199–210. [https://doi.org/10.1016/S0092-8674\(00\)80830-2](https://doi.org/10.1016/S0092-8674(00)80830-2).
- Schlutzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., et al. (2000). Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution. *Cell* 102, 615–623. [https://doi.org/10.1016/S0092-8674\(00\)00084-2](https://doi.org/10.1016/S0092-8674(00)00084-2).
- Schořová, Š., Fajkus, J., Závěská Drábková, L., Honys, D., and Prochazkova Schrupfova, P. (2019). The Plant Pontin and Reptin Homologues, RuvBL1 and RuvBL2a, Colocalize with TERT and TRB Proteins In Vivo, and Participate in Telomerase Biogenesis. *The Plant Journal* 98. <https://doi.org/10.1111/tpj.14306>.
- Schubert, H.L., Blumenthal, R.M., and Cheng, X. (2003). Many paths to methyltransfer: a chronicle of convergence. *Trends in Biochemical Sciences* 28, 329–335. [https://doi.org/10.1016/S0968-0004\(03\)00090-2](https://doi.org/10.1016/S0968-0004(03)00090-2).
- Seraphim, T.V., Nano, N., Cheung, Y.W.S., Aluksanasuwan, S., Colleti, C., Mao, Y.-Q., Bhandari, V., Young, G., Höll, L., Phanse, S., et al. (2022). Assembly principles of the human R2TP chaperone complex reveal the presence of R2T and R2P complexes. *Structure* 30, 156-171.e12. <https://doi.org/10.1016/j.str.2021.08.002>.
- Shibagaki, Y., Gotoh, H., Kato, M., and Mizumoto, K. (1995). Localization and In Vitro Mutagenesis of the Active Site in the *Saccharomyces cerevisiae* mRNA Capping Enzyme1. *The Journal of Biochemistry* 118, 1303–1309. <https://doi.org/10.1093/oxfordjournals.jbchem.a125023>.
- Silva, A.C., Xu, X., Kim, H.-S., Fillingham, J., Kislinger, T., Mennella, T.A., and Keogh, M.-C. (2012). The Replication-independent Histone H3-H4 Chaperones HIR, ASF1, and RTT106 Co-operate to Maintain Promoter Fidelity. *Journal of Biological Chemistry* 287, 1709–1718. <https://doi.org/10.1074/jbc.M111.316489>.
- Silva, S.T.N., Brito, J.A., Arranz, R., Sorzano, C.Ó.S., Ebel, C., Douth, J., Tully, M.D., Carazo, J.-M., Carrascosa, J.L., Matias, P.M., et al. (2018). X-ray structure of full-length human RuvB-Like 2 - mechanistic insights into coupling between ATP binding and mechanical action. *Sci Rep* 8, 13726. <https://doi.org/10.1038/s41598-018-31997-z>.
- Silva-Martin, N., Daudén, M.I., Glatt, S., Hoffmann, N.A., Kastiris, P., Bork, P., Beck, M., and Müller, C.W. (2016). The Combination of X-Ray Crystallography and Cryo-Electron Microscopy Provides Insight into the Overall Architecture of the Dodecameric Rvb1/Rvb2 Complex. *PLOS ONE* 11, e0146457. <https://doi.org/10.1371/journal.pone.0146457>.
- Smith, J.E., Cooperman, B.S., and Mitchell, P. (2002). Methylation sites in *Escherichia coli* ribosomal RNA: Localization and identification of four new sites of methylation in 23 S rRNA (American Chemical Society).
- Snider, J., Thibault, G., and Houry, W.A. (2008). The AAA+ superfamily of functionally diverse proteins. *Genome Biol* 9, 216. <https://doi.org/10.1186/gb-2008-9-4-216>.
- Su, H., Xu, T., Ganapathy, S., Shadfai, M., Long, M., Huang, T.H.-M., Thompson, I., and Yuan, Z.-M. (2014). Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene* 33, 1348–1358. <https://doi.org/10.1038/onc.2013.89>.

Sylvester, J.E., Whiteman, D.A., Podolsky, R., Pozsgay, J.M., Respass, J., and Schmickel, R.D. (1986). The human ribosomal RNA genes: structure and organization of the complete repeating unit. *Hum Genet* 73, 193–198. <https://doi.org/10.1007/BF00401226>.

Terns, M.P., and Terns, R.M. (2002). Small Nucleolar RNAs: Versatile trans-Acting Molecules of Ancient Evolutionary Origin. *Gene Expression* 10, 17–39. <https://doi.org/10.3727/096020197390059>.

Tessarz, P., Santos-Rosa, H., Robson, S.C., Sylvestersen, K.B., Nelson, C.J., Nielsen, M.L., and Kouzarides, T. (2014). Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* 505, 564–568. <https://doi.org/10.1038/nature12819>.

Tian, S., Yu, G., He, H., Zhao, Y., Liu, P., Marshall, A.G., Demeler, B., Stagg, S.M., and Li, H. (2017). Pih1p-Tah1p Puts a Lid on Hexameric AAA+ ATPases Rvb1/2p. *Structure* 25, 1519–1529.e4. <https://doi.org/10.1016/j.str.2017.08.002>.

Tiotiu, D. (2016). Analyse fonctionnelle des protéines Hit1 et Bcd1 impliquées dans la biogenèse des snoRNP à boîtes C/D eucaryotes. phdthesis. Université de Lorraine.

Tollervey, D., and Kiss, T. (1997). Function and synthesis of small nucleolar RNAs. *Current Opinion in Cell Biology* 9, 337–342. [https://doi.org/10.1016/S0955-0674\(97\)80005-1](https://doi.org/10.1016/S0955-0674(97)80005-1).

Torreira, E., Jha, S., López-Blanco, J.R., Arias-Palomo, E., Chacón, P., Cañas, C., Ayora, S., Dutta, A., and Llorca, O. (2008). Architecture of the Pontin/Reptin Complex, Essential in the Assembly of Several Macromolecular Complexes. *Structure* 16, 1511–1520. <https://doi.org/10.1016/j.str.2008.08.009>.

Tsukamoto, T., Shibagaki, Y., Imajoh-Ohmi, S., Murakoshi, T., Suzuki, M., Nakamura, A., Gotoh, H., and Mizumoto, K. (1997). Isolation and Characterization of the Yeast mRNA Capping Enzyme β Subunit Gene Encoding RNA 5'-Triphosphatase, Which Is Essential for Cell Viability. *Biochemical and Biophysical Research Communications* 239, 116–122. <https://doi.org/10.1006/bbrc.1997.7439>.

Udem, S.A., and Warner, J.R. (1972). Ribosomal RNA synthesis in *Saccharomyces cerevisiae*. *Journal of Molecular Biology* 65, 227–242. [https://doi.org/10.1016/0022-2836\(72\)90279-3](https://doi.org/10.1016/0022-2836(72)90279-3).

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* 50, D439–D444. <https://doi.org/10.1093/nar/gkab1061>.

Verheggen, C., Pradet-Balade, B., and Bertrand, E. (2015). SnoRNPs, ZNHIT proteins and the R2TP pathway. *Oncotarget* 6, 41399–41400. .

Vickers, N.J. (2017). Animal Communication: When I'm Calling You, Will You Answer Too? *Current Biology* 27, R713–R715. <https://doi.org/10.1016/j.cub.2017.05.064>.

Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R., and Ficner, R. (2000). Crystal Structure of the Spliceosomal 15.5kD Protein Bound to a U4 snRNA Fragment. *Molecular Cell* 6, 1331–1342. [https://doi.org/10.1016/S1097-2765\(00\)00131-3](https://doi.org/10.1016/S1097-2765(00)00131-3).

Vincenti, S., Chiara, V.D., Bozzoni, I., and Presutti, C. (2007). The position of yeast snoRNA-coding regions within host introns is essential for their biosynthesis and for efficient splicing of the host pre-mRNA. *RNA* 13, 138–150. <https://doi.org/10.1261/rna.251907>.

Wang, C., Kaufmann, J.A., Sanchez-Ross, M.G., and Johnson, K.M. (2000). Mechanisms of N-Methyl-d-aspartate-Induced Apoptosis in Phencyclidine-Treated Cultured Forebrain Neurons. *J Pharmacol Exp Ther* 294, 287–295. .

Watkins, N.J., and Bohnsack, M.T. (2012). The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *WIREs RNA* 3, 397–414. <https://doi.org/10.1002/wrna.117>.

WATKINS, N. J., GOTTSCHALK, A., NEUBAUER, G., KASTNER, B., FABRIZIO, P., MANN, M., & LUeHRMANN, R. E. I. N. H. A. R. D. (1998). Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *Rna*, 4(12), 1549-1568.

Watkins, N.J., Ségault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C., and Lührmann, R. (2000). A Common Core RNP Structure Shared between the Small Nucleolar Box C/D RNPs and the Spliceosomal U4 snRNP. *Cell* 103, 457–466. [https://doi.org/10.1016/S0092-8674\(00\)00137-9](https://doi.org/10.1016/S0092-8674(00)00137-9).

Watkins, N.J., Dickmanns, A., and 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. *Molecular and Cellular Biology* 22, 8342–8352. <https://doi.org/10.1128/MCB.22.23.8342-8352.2002>.

Weinstein, L.B., and Steitz, J.A. (1999). Guided tours: from precursor snoRNA to functional snoRNP. *Current Opinion in Cell Biology* 11, 378–384. [https://doi.org/10.1016/S0955-0674\(99\)80053-2](https://doi.org/10.1016/S0955-0674(99)80053-2).

Williams, D.J., Boots, J.L., and Hall, K.B. (2001). Thermodynamics of 2'-ribose substitutions in UUCG tetraloops. *RNA* 7, 44–53. <https://doi.org/10.1017/S1355838201001558>.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P., Vornrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* 407, 327–339. <https://doi.org/10.1038/35030006>.

Winkler, W.C., Grundy, F.J., Murphy, B.A., and Henkin, T.M. (2001). The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. *RNA* 7, 1165–1172. .

Woolford, J.L., Jr., and Baserga, S.J. (2013). Ribosome Biogenesis in the Yeast *Saccharomyces cerevisiae*. *Genetics* 195, 643–681. <https://doi.org/10.1534/genetics.113.153197>.

Woźniak, A.K., Nottrott, S., Kühn-Hölsken, E., Schröder, G.F., Grubmüller, H., Lührmann, R., Seidel, C. a. M., and Oesterhelt, F. (2005). Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements. *RNA* 11, 1545–1554. <https://doi.org/10.1261/rna.2950605>.

Wrzesinski, J., Nurse, K., Bakin, A., Lane, B. G., & Ofengand, J. (1995). A dual-specificity pseudouridine synthase: an *Escherichia coli* synthase purified and cloned on the basis of its specificity for psi 746 in 23S RNA is also specific for psi 32 in tRNA (phe). *Rna*, 1(4), 437-448.

Wrzesinski, J., Bakin, A., Nurse, K., Lane, B. G., & Ofengand, J. (1995). Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from *Escherichia coli*. *Biochemistry*, 34(27), 8904-8913.

Wyant, G.A., Abu-Remaileh, M., Frenkel, E.M., Laqtom, N.N., Dharamdasani, V., Lewis, C.A., Chan, S.H., Heinze, I., Ori, A., and Sabatini, D.M. (2018). NUFIP1 is a ribosome receptor for starvation-induced ribophagy. *Science* 360, 751–758. <https://doi.org/10.1126/science.aar2663>.

Yamamoto, R., Hirono, M., and Kamiya, R. (2010). Discrete PIH proteins function in the cytoplasmic preassembly of different subsets of axonemal dyneins. *Journal of Cell Biology* 190, 65–71. <https://doi.org/10.1083/jcb.201002081>.

Yu, G., Zhao, Y., Tian, S., Rai, J., He, H., Spear, J., Sousa, D., Fan, J., Yu, H.-G., Stagg, S.M., et al. (2019). Yeast R2TP Interacts with Extended Termini of Client Protein Nop58p. *Sci Rep* 9, 20228. <https://doi.org/10.1038/s41598-019-56712-4>.

Yuan, G., Klämbt, C., Bachellerie, J., Brosius, J., and Hüttenhofer, A. (2003). RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Research* 31, 2495–2507. <https://doi.org/10.1093/nar/gkg361>.

Zeytuni, N., and Zarivach, R. (2012). Structural and Functional Discussion of the Tetra-Trico-Peptide Repeat, a Protein Interaction Module. *Structure* 20, 397–405. <https://doi.org/10.1016/j.str.2012.01.006>.

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. <https://doi.org/10.1016/j.cell.2004.12.024>.

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. *Journal of Cell Biology* 180, 563–578. <https://doi.org/10.1083/jcb.200709061>.

Zhou, C.Y., Stoddard, C.I., Johnston, J.B., Trnka, M.J., Echeverria, I., Palovcak, E., Sali, A., Burlingame, A.L., Cheng, Y., and Narlikar, G.J. (2017). Regulation of Rvb1/Rvb2 by a Domain within the INO80 Chromatin Remodeling Complex Implicates the Yeast Rvbs as Protein Assembly Chaperones. *Cell Reports* 19, 2033–2044. <https://doi.org/10.1016/j.celrep.2017.05.029>.

Zunder, R.M., and Rine, J. (2012). Direct Interplay among Histones, Histone Chaperones, and a Chromatin Boundary Protein in the Control of Histone Gene Expression. *Molecular and Cellular Biology* 32, 4337–4349. <https://doi.org/10.1128/MCB.00871-12>.

Décryptage des mécanismes moléculaires d'assemblage de complexes macromoléculaires ARN/protéine : la snoRNP C/D comme modèle de travail

Chez les eucaryotes la production de ribosomes fonctionnels implique une modification fréquente, la méthylation en 2'OH des riboses du pré-ARNr. Un complexe ARN: protéine catalyse cette réaction, nommé snoRNP à boîtes C/D. L'assemblage de ses protéines cœur chez la levure *S. cerevisiae* (Snu13p, Nop58p, Nop56p, Nop1p) avec son ARN guide (snoARN à boîtes C/D), fait intervenir des protéines d'assemblage de façon transitoire, assurant une chronologie finement régulée. La protéine Bcd1p est un facteur essentiel de cet assemblage. Elle a été identifiée au laboratoire comme étant nécessaire au dépôt correct de Nop58p sur le snoARN à boîtes C/D. Durant ce travail de thèse nous avons cherché à caractériser les complexes transitoires précoces de l'assemblage, comprenant la protéine Bcd1. La machinerie ATPasique Rvb1/Rvb2 est étroitement liée au rôle de Bcd1p. La caractérisation biochimique et structurale du complexe Rvb1/Rvb2/Bcd1p suggère une flexibilité conformationnelle importante.

Mots clés : Bcd1p, Assemblage des snoRNP à boîtes C/D, snoARN, Rvb1/Rvb2, *S. cerevisiae*

Deciphering the molecular mechanisms of the assembly of macromolecular complexes : the snoRNP as a working model

In eukaryotes, the production of functional ribosomes involves a frequent modification, the 2'OH methylation of the pre-rRNA riboses. An RNA/protein complex catalyzes this reaction, termed the C/D box snoRNP. The assembly of its core proteins in the yeast *S. cerevisiae* (Snu13p, Nop58p, Nop56p, Nop1p) with its guide RNA (snoRNA with C/D boxes), involves assembly proteins in a transient manner, ensuring a fine regulated chronology. The Bcd1p protein is an essential factor in this assembly. It has been identified in the laboratory as being necessary for the correct deposition of Nop58p on the C/D box snoRNA. During this thesis work we sought to characterize the early transitional complexes of assembly, including the Bcd1 protein. The Rvb1/Rvb2 ATPase machinery is closely linked to the role of Bcd1p. Structural biochemical characterization of the Rvb1/Rvb2/Bcd1p complex suggests significant conformational flexibility

Key words : Bcd1p, Box C/D snoRNP Assembly, snoRNA, Rvb1/Rvb2, *S. cerevisiae*