



HAL
open science

Caractérisation structurale et fonctionnelle de la protéine Bcd1, impliquée dans la biogenèse des snoRNP à boîtes C/D chez la levure *Saccharomyces cerevisiae*

Benoît Bragantini

► **To cite this version:**

Benoît Bragantini. Caractérisation structurale et fonctionnelle de la protéine Bcd1, impliquée dans la biogenèse des snoRNP à boîtes C/D chez la levure *Saccharomyces cerevisiae*. Biochimie [q-bio.BM]. Université de Lorraine, 2016. Français. NNT : 2016LORR0295 . tel-01752418

HAL Id: tel-01752418

<https://hal.univ-lorraine.fr/tel-01752418>

Submitted on 2 Jun 2018

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

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



AVERTISSEMENT

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

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

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

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

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

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

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

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

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

Benoît BRAGANTINI

**Caractérisation structurale et fonctionnelle de la protéine
Bcd1, impliquée dans la biogenèse des snoRNP à boîtes
C/D chez la levure *Saccharomyces cerevisiae*.**

Le 12 décembre 2016

Membres du jury :

Rapporteurs :	Mme Françoise GUERLESQUIN	Directrice de recherche, CNRS, Marseille
	Mme Bérengère PRADET-BALADE	Chargée de recherche HDR, CNRS, Montpellier
Examineurs :	M. Marc GRAILLE	Directeur de recherche, CNRS, Palaiseau
	Mme Pascale TSAN	Maître de conférences, Université de Lorraine, Nancy
	M. Bruno CHARPENTIER	Professeur, Université de Lorraine, Nancy Directeur de thèse
	M. Xavier MANIVAL	Chargé de recherche, CNRS, Nancy Co-directeur de thèse

Remerciements

Même si les fins de grandes aventures sont généralement synonymes de tristesse, elles représentent toutefois l'opportunité de regarder en arrière le chemin parcouru et de repenser à toutes les rencontres qui ont jalonné notre parcours. Le moment est ainsi venu pour moi aujourd'hui de me prêter à cet exercice de souvenir me permettant au passage de remercier toutes ces personnes qui ont beaucoup compté durant ces cinq années.

Je tiens donc pour commencer à remercier Mme Christiane Branlant, M. Bruno Charpentier et M. Xavier Manival qui ont accepté de m'accueillir dans leur équipe du laboratoire IMoPA, fraîchement débarqué dans son nouveau bâtiment aux couloirs verts, et plus particulièrement Xavier et Bruno qui deviendront par la suite mes directeurs de thèse. C'est ainsi, en cet hiver 2012, que j'ai pu mettre pour la première fois les pieds dans le monde de la recherche scientifique avec ma « formatrice paillasse » Clémence Jacquemin et monsieur RMN, plus connu sous le nom de Marc Quinternet, tout juste papa (pour la 1^{ère} fois). Ce stage m'aura également permis de rencontrer mes futurs prédécesseurs au rang de doctorant, Régis Back et Benjamin Rothé, m'ayant ouvert la voie des études sur la fameuse Bcd1 (et tant d'autres).

Je me dois également de remercier chaleureusement Marie-Eve Chagot, arrivée dans l'équipe l'année suivante, et dont l'aide, les conseils et la bonne humeur m'ont été d'une incommensurable bénédiction (difficile de finir une thèse en 3 ans sans ça !). Je vais ensuite m'empresse de remercier également Benjamin Chagot (on ne sais jamais, je me méfie des crises de jalousie) pour ses nombreuses pauses café dans mon bureau, toujours ponctuées de discussions scientifiques inégalables.

Je remercie Christophe Charron qui a réalisé les travaux de cristallographie sur le sujet Bcd1-Rtt106 (un projet en apparence simple et rapide) et Aurélien Thureau (ou d'après certains Thureauélien) du Synchrotron Soleil qui aura lui effectué les analyses par SAXS. Je remercie également et vivement l'équipe de Sarah Cianferani du Laboratoire de Spectrométrie de Masse Bio-Organique (LSMBO) de Strasbourg et en particulier Guillaume Terral, avec qui nous avons eu une grande collaboration scientifique autour de mes projets. Pour ce qui est des autres collaborations grâce auxquelles nos projets ont pu avancer à grands pas, je dois également remercier les équipes de Pedro Matias de l'IBET (Oeiras, Portugal) pour leur apport sur le sujet des protéines RuvBL1/2 et de Georges Mer de la Mayo Clinic (Rochester, MN, USA) pour leurs travaux sur la protéine Rtt106.

Merci également à tous ceux qui ont pu m'aider d'une façon ou d'une autre durant tout ce temps passé au laboratoire, je pense notamment à Val, Flo, Florence, Lylia, Christelle, Hélène, Valentin et MuMu pour le bon côté de la force, et Sab, Sèv et Alex pour le côté obscur affilié à l'enzymologie. Afin d'éviter tout oubli regrettable, je conclurais ce paragraphe par un joli : et tout les autres membres des équipes 1 et 3 du laboratoire IMoPA (qu'on ne vienne pas me dire que j'ai oublié quelqu'un).

Et parce que sans eux nous ne serions rien, je remercie Jocelyne Charbonnel (sans faute cette fois-ci), Josiane Fournier et Christophe Muller (LE GhostBuster).

Je tiens également à adresser mes remerciements à Marc Graille et Pascale Tsan pour avoir accepté de suivre l'avancée de mon travail tout au long de mon doctorat en faisant partie de mon comité de suivi de thèse, puis par la suite en étant tout les deux membres examinateurs de mon jury de thèse. De la même façon, je remercie Françoise Guerlesquin et Bérengère Pradet-Balade pour avoir accepté d'être rapporteurs de ma thèse. Je remercie donc les différents membres de mon jury avec qui j'ai pu échanger sur mon sujet de thèse ainsi que sur l'univers de la recherche en général, ce qui m'a permis de mieux appréhender la suite de ma carrière.

Bien évidemment, et comme dans l'adversité il faut se serrer les coudes, il me faut remercier ici tous mes camarades de galère. Je vais donc commencer par mon compatriote (enfin presque n'exagérons rien, picards et champenois ne sont après tout que voisins) des premiers instants qui aura réussi l'immense prouesse de soutenir sa thèse avant moi... Merci donc à mon ami Johnny (toujours avec deux « n ») qui ne m'aura pas lâché tout au long de mon cursus universitaire (ou peut-être est-ce l'inverse ?), depuis notre jeunesse insouciante du temps de la licence rémoise aux côtés de deux drôles de zigotos, Valentin et un étrange petit homme vietnamien (TQN), jusqu'à notre maturité scientifique inconsciente. Merci également aux thésards déjà en faction pour leur accueil, Antoine (grand président), Topi (un kiné à ce qu'il paraît), Jean-Guitoune (homme des bois et maître du propeller), Titi (un vosgien pas comme les autres... enfin si tout compte fait probablement comme les autres), Baloo (parce qu'il en faut peu), Décébal (mon top co-auteur pour le moment), Samia, Mathieu (un grand connaisseur dans le domaine du genou) et tous les autres bien entendu, comme Omar (dit OM-OM) le post-doc restaurateur. Mais il ne faut pas en rester là car si ceux-ci m'ont précédé, d'autres auront bientôt eu l'immense honneur de marcher dans mes pas... Je veux ici parler d'ARN-O (autre infortuné doctorant affecté à Bcd1), de notre divin J-C (toujours rayonnant), puis Fanny (seule rescapée de la gente féminine au sein de ce petit monde enzymatique) ainsi que Flo (le tombeur bodybuildé), Roger (alias G-RO, ou peut-être était-ce l'inverse je ne sais plus) et Adeline. Ce fut un plaisir de partager tout ce temps (voir même plus) avec vous tous.

Tant que nous sommes dans le registre des amitiés, je remercie également l'association A-Team of BioSE, fondée par l'illustre Jérémy Lagrange, dont je fus un temps trésorier puis simple vice-président ainsi que tous ses membres et particulièrement mes collègues de bureaux Léo (une présidente), Mathilde, Amel (prêtresse du fouet), Cécile, J-C (encore lui) et ARN-O (idem) pour le premier et Laurie (une autre présidente), à nouveau Amel et Cécile, le grand Paul, la petite Daf' et Fanny (si si encore elle) pour le second. Je profite de ce paragraphe pour ajouter un merci à Mélu, ma voisine doctorante qui fait (entre autres choses) de très bons gâteaux. Et je pense que je dois quand même remercier Tonio pour m'avoir maintenu en vie toutes ces années. Vivre ailleurs n'aurait pas été aussi simple.

Et pour finir je vais ajouter un gros remerciement général à toute ma famille qui a dû désespérer de ne plus me voir pendant ces années de thèse, car voyez-vous le travail d'un doctorant, même s'il n'en a pas l'air, est finalement assez prenant. Bon disons le clairement, ça ne devrait pas vraiment s'arranger par la suite... Je conçois que ça n'a malgré tout pas dû être évident pour vous, sachant en plus que vous ne compreniez rien à ce que je faisais, et probablement pas pourquoi je voulais le faire ! Merci.

Nomenclatures

- Les noms des protéines seront toujours cités comme suit :
archée/levure/humain.
Par exemple : L7Ae/Snu13p/SNU13.
- Le nom des protéines de levure sera toujours suivi d'un « p » signifiant « protéine », sauf dans les cas où le terme « protéine » précède le nom de la protéine.
Par exemple : Snu13p. La protéine Snu13.
- Les noms des constituants d'un complexe seront toujours séparés par « : »
Par exemple : Le complexe Bcd1p:Rtt106.
Cas particulier : Afin de simplifier la lecture, le complexe entre les protéines Rvb1 et Rvb2 sera nommé Rvb1/2 (RuvBL1/2 dans le cas des protéines humaines).
- Les termes non français seront écrits en italique.
- Les acides aminés seront abrégés avec le code à une lettre.

Liste
des
Abréviations

Abréviations

°C	: degré Celsius
2YT	: 2 <i>Yeast extract and Tryptone</i>
Å	: Angström
AAA+	: <i>ATPases Associated with diverse cellular Activities</i>
ADN	: Acide Désoxyribonucléique
ADP	: Adénosine Di-Phosphate
APS	: PerSulfate d'Ammonium
ARN	: Acide RiboNucléique
ARNm	: ARN messenger
ARNr	: ARN ribosomique
ARNt	: ARN de transfert
ATP	: Adénosine TriPhosphate
Bcd1p	: <i>Box C/D snoRNA-1</i>
CAF-1	: <i>Chromatin Assembly Factor-1</i>
CCS	: <i>Collision Cross Surfaces</i>
CCT	: <i>Chaperonin-Containing T</i>
CD	: <i>Circular Dichroism</i>
CK2	: <i>Casein Kinase 2</i>
cps	: coups par seconde
Cryo-EM	: <i>Cryo-electron microscopy</i>
CS	: <i>CHORD and Sgt1</i>
CTP	: Cytidine TriPhosphate
Da/kDa	: Dalton/kiloDalton
DLS	: <i>Dynamic Light Scattering</i>
Dmax	: Distance maximale
DMSO	: DiMéthylSulfOxyde
dNTP	: désoxyriboNucléotide TriPhosphate
DO	: Densité Optique
DSC	: <i>Differential Scanning Calorimetry</i>
DTT	: <i>DiThioThreitol</i>
ECP51/54	: <i>51/54 kDa erythrocyte cytosolic protein</i>
EDTA	: Ethylène Diamine TétrAcétate
g/mg/µg/ng	: gramme/milligramme/microgramme/nanogramme
g/L	: gramme par Litre
g/mol	: gramme par mole
GAR	: <i>Glycine/Arginine-Rich</i>
GTP	: Guanosine TriPhosphate
h	: heure
HAT	: <i>Histones AcetylTransferases</i>
HEPES	: acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
HDX	: <i>Hydrogen-Deuterium Exchange</i>
Hit1p	: <i>High Temperature growth factor-1</i>
HIR	: <i>Histone Regulatory</i>
HMQC	: <i>Heteronuclear Multiple-Quantum Correlation</i>
HPLC	: <i>High Performance Liquide Chromatography</i>
HSP90	: <i>Heat Shock Protein 90 kDa</i>

HSQC	: <i>Heteronuclear Single-Quantum Correlation</i>
IMAC	: <i>Immobilized Metal Affinity Chromatography</i>
IPTG	: <i>IsoPropyl β-D-1-ThioGalactopyranoside</i>
ITC	: <i>Isothermal Titration Calorimetry</i>
K	: Degré Kelvin
kb	: kilobase
Kd	: Constante de dissociation
K-turn	: <i>Kink-turn</i>
L/mL/ μ L	: Litre/milliLitre/microLitre
LB	: <i>Lysogeny Broth</i> (ou Luria Bertani)
LSU	: <i>Large SubUnit</i>
M/mM/ μ M	: Molaire (mole par litre)/milliMolaire/microMolaire
mA	: miliAmpère
MALLS	: <i>Multi-Angle Laser Light Scattering</i>
mCi/mL	: miliCurie par milliLitre
MHz	: MégaHertz
min	: minute
MRP	: <i>Mitochondrial RNA Processing</i>
MTase	: Méthyltransférase
mU/ μ L	: miliUnité par microLitre
nm/ μ m	: nanomètre/micromètre
Nop	: <i>Nucleolar protein</i>
NOESY	: <i>Nuclear Overhauser Effect Spectroscopy</i>
NoLS	: <i>Nucleolar Localisation Signal</i>
NUFIP1	: <i>Nuclear FMRP Interacting Protein-1</i>
nt	: nucléotide
PAGE	: <i>PolyAcrylamid Gel Electrophoresis</i>
PEG	: <i>PolyEthylene Glycol</i>
pI	: point Isoélectrique
PM	: Poids Moléculaire
ORF	: <i>Open Reading Frame</i>
OB-fold	: <i>Oligonucleotide/Oligosaccharide Binding-fold</i>
pb	: paire de base
PCR	: <i>Polymerase Chain Reaction</i>
PDB	: <i>Protein Data Bank</i>
PEI	: PolyÉthylèneImine
pH	: potentiel Hydrogène
PH	: <i>Pleckstrin Homology</i>
Pi	: Phosphate inorganique
Pih1p	: <i>Protein Interacting with Hsp90-1</i>
ppm	: parties par million
PTC	: <i>Peptidyl-Transferase Center</i>
PTM	: <i>Post-Translational Modifications</i>
R2TP	: Rvb1p/Rvb2p/Tah1p/Pih1p
RCNA	: <i>Replication-Coupled Nucleosome Assembly</i>
Rg	: Rayon de giration
RINA	: <i>Replication-Independant Nucleosome Assembly</i>
RMN	: Résonance Magnétique Nucléaire
RMSD	: <i>Root-Mean-Square Deviation</i>
RNP	: <i>RiboNucleoParticle</i> (Particule RiboNucléique)

rNTP	: riboNucléotide TriPhosphate
RPAP3	: <i>RNA Polymerase-associated protein-3</i>
rpm	: rotation par minute
Rrp9	: <i>Ribosomal RNA Processing-9</i>
Rsa1	: <i>Ribosom assembly-1</i>
RSC	: <i>Remodel the Structure of Chromatin</i>
Rtt106	: <i>Regulator of Ty1 Transposition 106</i>
Rvb1/2	: <i>RuVB-like</i>
S	: Svedberg
SAH	: S-Adénosyl-(L)-Homocystéine
SAM	: S-Adénosyl-(L)-Méthionine
SANS	: <i>Small-Angle Neutron Scattering</i>
SAXS	: <i>Small-Angle X-ray Scattering</i>
scaRNA	: <i>small Cajal Bodies RNA</i>
SDS	: <i>Sodium DodecylSulfate</i>
SILAC	: <i>Stable Isotope Labeling with Amino acids in Cell culture</i>
SMN	: <i>Survival of MotoNeuron</i>
snRNA	: <i>small nuclear RNA</i>
snoRNA	: <i>small nucleolar RNA</i>
snoRNP	: <i>small nucleolar RNP</i>
Snu13p	: <i>Small Nuclear ribonucleoprotein associated-13</i>
SOB	: <i>Super Optimal Broth</i>
SOC	: <i>SOB with Catabolic repression</i>
SPR	: <i>Surface plasmon resonance</i>
SSU	: <i>Small SubUnit</i>
SWI/SNF	: <i>SWItch/Sucrose Non-Fermentable</i>
TAF9	: <i>Transcription Activating Factor 9</i>
Tah1p	: <i>TPR containing protein Associated with HSP90-1</i>
TCEP	: <i>Tris-(2-CarboxyEthyl)Phosphine</i>
TEMED	: <i>TEtraMethylEthylèneDiamine</i>
™	: TradeMark
T _m	: Température de fusion
TRASH	: <i>Trafficking, Resistance and Sensing of Heavy Metals</i>
TROSY	: <i>Transverse Relaxation Optimized SpectroscopY</i>
TIP48/49	: <i>TATA box-binding protein-interacting protein of 48/49 kDa</i>
TOCSY	: <i>TOTAL Correlation SpectroscopY</i>
TPR	: <i>TetratricoPeptide Repeat</i>
TRIP3	: <i>Thyroid Receptor-Interacting Protein-3</i>
Tris	: <i>Trishydroxyméthylaminométhane</i>
TTP	: <i>Thymidine TriPhosphate</i>
U	: Uridine
UV	: UltraViolet
V	: Volt
v/v	: Volume à volume
ZHD	: <i>zf-HIT domain</i>

Table
des
Matières

Table des Matières



AVANT-PROPOS	1
--------------------	---

Introduction

1. Les snoRNP à boîtes C/D	6
1.1. Les snoRNA à boîtes C/D	6
1.1.1. Organisation des gènes codant les snoRNA à boîtes C/D	6
1.1.2. Transcription et maturation des snoRNA à boîtes C/D	7
1.1.3. Les boîtes C et D	8
1.1.4. Les snoRNA à boîtes C/D possèdent un motif structural en K-turn	9
1.2. Composition des snoRNP à boîtes C/D matures	13
1.2.1. La famille L7Ae (la protéine Snu13/SNU13)	15
1.2.2. La 2'-O-méthylase	18
1.2.3. Les protéines Nop56/NOP56 et Nop58/NOP58	22
1.2.4. Structure 3D des particules sRNP d'archées	26
1.2.4.1. Hypothèse de la mono-sRNP	27
1.2.4.2. Hypothèse de la di-sRNP	33
1.2.4.3. Cas particulier de la snoRNP U3	37
2. Biogenèse des snoRNP à boîtes C/D eucaryotes : la machinerie d'assemblage	39
2.1. Le complexe Hsp82-R2TP	40
2.1.1. La chaperonne Hsp82/HSP90	40
2.1.2. Le complexe R2TP	42
2.1.2.1. La protéine Tah1/RPAP3	42
2.1.2.2. La protéine Pih1/PIH1D1	45
2.1.2.3. Les protéines Rvb1/RuvBL1 et Rvb2/RuvBL2	46
2.1.2.3.1. Organisation structurale	47
2.1.2.3.2. Les protéines Rvb1/2 sont des composants de nombreux complexes nucléaires	51
2.2. Le complexe Rsa1p:Hit1p/NUFIP1:ZNHIT3	54
2.3. La protéine Bcd1/BCD1	58
2.3.1. Connaissances au début de ma thèse	58
2.3.2. Les domaines en doigts à zinc de la famille zf-HIT (ZHD)	59
2.3.3. Lien entre les ZHD et les protéines RuvBL1/2	59
3. Les chaperonnes d'histone	61
3.1. Composition et structure de la chromatine	61
3.2. Assemblage des nucléosomes	61
3.2.1. Assemblage des nucléosomes couplé à la réplication (RCNA)	65
3.2.2. Assemblage des nucléosomes indépendant de la réplication (RINA)	65
3.3. Les protéines liées aux nucléosomes	66
3.3.1. Modification des histones	66
3.3.2. Chaperonnes d'histones	66

3.3.2.1. Le complexe CAF-1	66
3.3.2.2. Le complexe FACT	67
3.3.2.3. La chaperonne d'histone Rtt106.....	67
Objectifs	72

Matériels et Méthodes

1. Matériels.....	75
1.1. Les oligonucléotides	75
1.2. Les souches bactériennes d'Escherichia coli.....	75
1.3. Les milieux de culture bactérienne.....	75
1.4. Les antibiotiques.....	79
1.5. Les plasmides bactériens	79
1.6. Les tampons généraux	81
1.7. Les tampons utilisés lors de l'étude des protéines recombinantes	83
1.8. Les différents types de gels	84
1.9. Les peptides	84
2. Méthodes	85
2.1. Le génie génétique.....	85
2.1.1. Clonage	85
2.1.1.1. La PCR (Polymerase Chain Reaction)	85
2.1.1.2. La digestion par des enzymes de restriction.....	86
2.1.1.3. Le sous-clonage.....	87
2.1.1.4. Préparation de vecteur linéarisé.....	87
2.1.1.5. La ligation.....	87
2.1.1.6. Transformation de bactéries	88
2.1.1.7. Amplification de vecteurs.....	88
2.1.2. Mutagenèse dirigée par PCR	88
2.1.2.1. La mutagenèse dirigée par Rolling-circle.....	89
2.1.2.2. La mutagenèse dirigée par Overlap-PCR	89
2.1.3. Transcription in vitro	89
2.1.4. Marquage radioactif.....	90
2.1.4.1. Marquage radioactif des ARN.....	90
2.1.4.2. Marquage radioactif de l'ADN.....	90
2.2. Biochimie des protéines.....	90
2.2.1. Production de protéines recombinantes chez E. coli	91
2.2.2. Purification de protéines recombinantes	91
2.2.3. La technique de co-expression (crible)	93
2.2.4. La protéolyse ménagée	94
2.2.5. Le His-pull-down	94
2.2.6. La technique de retard sur gel.....	95
2.3. Analyses biophysiques des protéines	95
2.3.1. ITC (Isothermal Titration Calorimetry).....	95
2.3.2. Dichroïsme circulaire.....	96
2.3.3. DLS (Dynamic Light Scattering).....	97

2.3.4.	DSC (Differential Scanning Calorimetry)	97
2.4.	La RMN (Résonance Magnétique Nucléaire)	97
2.4.1.	Préparation des échantillons	97
2.4.2.	Acquisition des spectres.....	98
2.4.3.	Analyse des spectres et calcul de structures	98
2.4.4.	Titrages suivis par RMN.....	98
2.4.5.	Tests d'interaction par suivi RMN des méthyles.....	99

Résultats et Discussion

1.	Etude des doigts à zinc de la famille zf-HIT	101
1.1.	Travaux publiés	101
1.2.	Publication	102
1.3.	Résultats complémentaires à la publication.....	127
1.3.1.	Coordination des atomes de zinc	127
1.3.2.	Etude des ZHD par dichroïsme circulaire.....	127
1.3.3.	Domaine minimal fonctionnel de Bcd1p	127
1.3.4.	Analyse de la stabilité de la protéine Bcd1	133
1.4.	Les ZHD sont des modules d'interaction avec les protéines RuvBL1:2	136
1.4.1.	Identification d'une interaction entre ZHD et RuvBL par RMN.....	136
1.4.2.	Analyse par l'approche Blitz de l'interaction entre ZHD et RuvBL	140
2.	Recherche de partenaires associés à Bcd1p	140
2.1.	Partenaires protéiques	140
2.1.1.	Isolement d'un complexe Nop58p:Rvb1p:Rvb2p:Bcd1p.....	140
2.1.1.1.	Co-expression des protéines Nop58opt(1-447), Rvb1, Rvb2 et Bcd1	143
2.1.1.2.	Interaction via le domaine DII _{ext} de Rvb2 ?.....	143
2.1.1.3.	Interaction via le domaine DIII de Rvb2p ?	145
2.1.1.4.	Interaction entre les protéines Rvb1 et Rvb2 par leurs domaines DII.....	148
2.1.2.	La protéine Rtt106	148
2.2.	Partenaires nucléiques	151
3.	Etude du complexe formé entre les protéines Bcd1 et Rtt106	151
3.1.	Caractérisation biophysique du complexe Bcd1p-Rtt106p(65-320)	151
3.2.	Détermination de l'interface d'interaction entre les protéines Bcd1 et Rtt106	155
3.2.1.	Détermination des domaines minimaux permettant l'interaction.....	157
3.2.1.1.	Approche par protéolyse ménagée et analyse par spectrométrie de masse	157
3.2.1.2.	Etude par ITC et RMN de l'interaction des fragments de Bcd1p et Rtt106p.....	161
3.2.1.2.1.	Etude des fragments de Bcd1p	161
3.2.1.2.2.	Etude des fragments de Rtt106	167
3.2.2.	Caractérisation des régions à l'interface de l'interaction Bcd1p-Rtt106p	169
3.2.2.1.	Etude par spectrométrie de masse.....	169
3.2.2.1.1.	Echange hydrogène-deutérium (HDX)	169
3.2.2.1.1.1.	Analyse par HDX de la protéine Rtt106	171
3.2.2.1.1.2.	Analyse par HDX de la protéine Bcd1	174
3.2.2.1.2.	Mobilité ionique et spectrométrie de masse.....	176
3.2.2.1.3.	Cross-link	176
3.2.2.2.	Titrages suivis par RMN.....	178

3.2.2.2.1. Titrage de la protéine Rtt106(65-320)	179
3.2.2.2.2. Titrage de la protéine Bcd1(120-303)	181
3.3. Etudes structurales tridimensionnelles	181
3.3.1. Analyse structurale de la protéine Bcd1(120-303)	183
3.3.1.1. Détermination de la structure 3D en solution de la protéine Bcd1(120-303) par RMN.....	183
3.3.1.2. Surface d'interaction avec la protéine Rtt106(65-320)	189
3.3.2. Analyse Structurale du complexe entre les protéines Bcd1 et Rtt106	193
3.3.2.1. Analyse SAXS (Small Angle X-ray Scattering)	193
3.3.2.2. Structure par cristallographie aux rayons X	196
3.3.2.2.1. Apport de la RMN : expérience de NOE hétéronucléaires.....	196
3.3.2.2.2. Le complexe Rtt106p(65-301):Bcd1p(120-303)D220-225	199
3.4. Compétition entre Bcd1 et les histones	199

Conclusions et Perspectives

1. Le domaine minimum fonctionnel de la protéine Bcd1	204
2. Interaction ZHD-RuvB1/2	205
3. Le complexe Nop58p-Rvb1p-Rvb2p-Bcd1p	206
4. Structure 3D en solution de la protéine Bcd1(120-303)	207
5. Interaction entre les protéines Bcd1 et Rtt106	208
6. Vers une structure par diffraction des rayons X du complexe Rtt106p/Bcd1p.....	209
7. Test de compétition entre Bcd1p et les histones pour l'interaction avec Rtt106p	210
8. La protéine BCD1 humaine	211

Références Bibliographiques

AVANT

PROPOS

AVANT PROPOS

De l'ADN aux protéines : le ribosome

Les protéines représentent aujourd'hui le principal moteur des différentes fonctions biologiques réalisées par les cellules. Ces protéines doivent être synthétisées par les cellules à partir de l'information génétique contenue dans l'ADN sous la forme de séquences codantes, les ORF ou phases ouvertes de lecture. La mécanique conduisant à la production des protéines à partir de ces ORF est complexe et nécessite le passage par une séquence ARN, appelée ARN messenger (ARNm). Cet ARN joue le rôle d'élément de transition entre l'ADN et la machinerie de production des protéines. Cette machinerie implique le ribosome dont la fonction est de décoder l'information contenue dans l'ARNm sous forme d'une succession de nucléotides et de catalyser la synthèse des liaisons peptidiques entre les acides aminés. Pour cela, le ribosome requiert l'intervention de « clés de décryptage » permettant d'attribuer un acide aminé à une séquence de trois nucléotides de l'ARNm (codon), appelées ARN de transfert (ARNt). Il existe une large variété d'ARNt permettant de recouvrir l'ensemble des possibilités de codons.

Composition des ribosomes eucaryotes

Les ribosomes ont été observés pour la première fois en 1955 par microscopie électronique (Palade, 1955). Ce sont des superstructures ribonucléoprotéiques constituées de deux sous-unités, elles-mêmes constituées d'ARN ribosomiques (ARNr) et de protéines associées. Chez les eucaryotes, ces deux sous-unités sont appelées 60S (grande sous-unité ou LSU « *Large SubUnit* ») et 40S (petite sous-unité ou SSU « *Small SubUnit* ») selon leur coefficient de sédimentation et forment le ribosome 80S. Chez l'Homme, la LSU est constituée des 3 ARNr 28S (25S chez la levure), 5S et 5,8S et de 47 protéines, tandis que la SSU n'est constituée que de l'ARNr 18S et de 33 protéines (Nakao *et al.*, 2004) (Figure 1A).

L'ARNm en cours de traduction est fixé par la SSU, lieu où va pouvoir se faire la reconnaissance codon-anticodon entre l'ARNm et l'ARNt (Schmeing et Ramakrishnan, 2009). Le ribosome possède trois sites de fixation aux ARNt, et chaque ARNt va transiter par chacun de ces sites (Figure 1B). Dans un premier temps, l'ARNt aminoacylé (lié covalamment à un acide aminé) se fixe au site A (Aminoacyl-ARNt), qui est le site d'arrivée. Après création de la liaison peptidique et translocation du ribosome, l'ARNt occupe le site P (Peptidyl-ARNt), où il est lié à la chaîne polypeptidique naissante. Enfin, le site E (*Exit*) est le

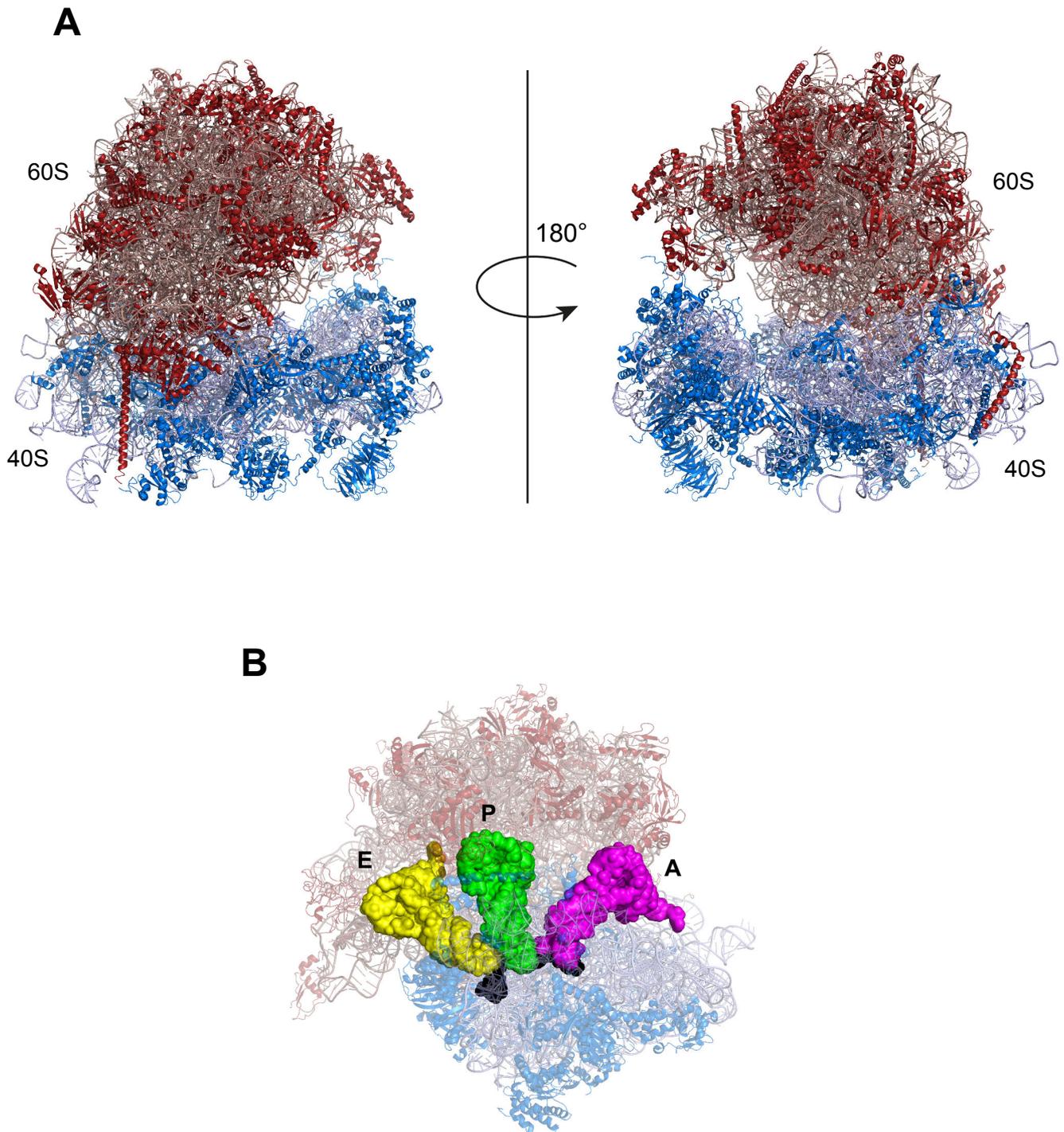


Figure 1. Structure du ribosome. (A) Structure cristallographique à 3,0 Å du ribosome 80S de *Saccharomyces cerevisiae* (Ben-Shem *et al.*, 2011 ; code PDB : 4V88). La grande sous-unité 60S est représentée en nuances de rouge (ARN et protéines). La petite sous-unité 40S est représentée en nuances de bleu (ARN et protéines). (B) Structure obtenue par cryo-EM à 6,4 Å du ribosome 70S de *Thermus thermophilus* (Schuette *et al.*, 2009 ; code PDB : 4V68). Le code couleur entre les deux sous-unités est le même que pour (A), avec les ARNt et ARNm représentés en surface. Les ARNt sont colorés en magenta pour l'ARNt au site A (Aminoacyl-ARNt), en vert pour celui au site P (Peptidyl-ARNt) et en jaune au site E (*Exit*). L'ARNm est coloré en noir (3 codons simplement).

site où l'ARNt (sans son acide aminé) est placé après formation de la liaison peptidique et avant son éjection du ribosome. La formation de la liaison peptidique se fait au niveau du site PTC (*Peptidyl-Transferase Center*) présent au sein de la LSU du ribosome. Ce site se situe de telle sorte que les extrémités 3' des ARNt aminoacylés positionnés aux sites A et P s'y retrouvent. Le dernier acide aminé de la chaîne polypeptidique naissante encore porté par l'ARNt au site P se retrouve ainsi à proximité de l'acide aminé qui le succèdera et qui est lui présent à l'extrémité 3' de l'ARNt au site A (Schmeing et Ramakrishnan, 2009).

Maturation des pré-ARNr et biogenèse des ribosomes

L'obtention de ribosomes fonctionnels passe au préalable par plusieurs étapes de maturation des pré-ARNr. Les ARNr 28S (25S chez la levure), 5,8S et 18S sont codés au sein d'un même cluster dans le génome (ADNr) (Woolford et Baserga, 2013; Lafontaine, 2015). Après transcription de cet ADNr par l'ARN polymérase I dans le nucléole, ces trois futurs ARNr constituent dans un premier temps un seul et même pré-ARNr, le pré-ARNr 47S (35S chez la levure). Ce précurseur ARN va subir plusieurs étapes successives de clivages pour finalement aboutir aux trois ARNr 28S, 5,8S et 18S. L'ARNr 5S est lui transcrit par l'ARN polymérase III dans le nucléoplasme. Les étapes précoces de clivages du précurseur ARNr impliquent notamment des particules ribonucléoprotéiques (RNP), présentes dans le nucléole des cellules eucaryotes, appelées snoRNP (*small nucleolar RNP*). Ces particules sont catégorisées en deux types, les snoRNP à boîtes C/D (U3 et U8), et celles à boîtes H/ACA (U17). Le mécanisme de maturation fait également intervenir pour la voie de maturation des ARNr 5,8S et 28S une RNase organisée sous la forme d'une RNP, la MRP et plusieurs autres activités exonucléolytiques comme celle portée par le complexe multiprotéique de l'exosome.

En plus des processus nucléolaires de maturation des ARNr par clivage du pré-ARNr, les ARNr sont modifiés chimiquement et l'assemblage des protéines ribosomiques débute. Globalement, la biogenèse du ribosome nécessite l'intervention d'environ 200 facteurs protéiques d'assemblage ainsi qu'un grand nombre de snoRNA (de 75 chez la levure à plus de 200 chez l'Homme) (Woolford et Baserga, 2013; Lafontaine, 2015). La machinerie d'assemblage du ribosome est donc extrêmement complexe et fait intervenir un grand nombre de facteurs dont beaucoup sont essentiels à la viabilité cellulaire.

Modifications chimiques des ARNr

Les ARNr composant le ribosome eucaryote contiennent un grand nombre de nucléotides modifiés, de 100 pour la levure à près de 200 chez l'Homme (Piekna-Przybylska

et al., 2007). Ces modifications sont de deux types, les méthylations de bases ou des riboses et les pseudouridylations (Figure 2). Ces modifications sont pour beaucoup apportées par les snoRNP à boîtes C/D qui catalysent la réaction de 2'-O-méthylation des riboses et les snoRNP à boîtes H/ACA qui convertissent les uridines en pseudouridines. En majorité, ces modifications sont positionnées au niveau des sites importants pour la fonctionnalité du ribosome, tels que les sites A, P et E de liaison aux ARNt et à l'ARNm, le PTC, le tunnel de sortie du polypeptide néoformé ou l'interface entre les deux sous-unités (Figure 2) (Decatur et Fournier, 2002). Ces modifications ont donc des rôles importants dans la structure du ribosome, ainsi que dans sa fonction.

Ribosomes spécialisés

Si les protéines constituant le ribosome ainsi que les modifications des ARNr sont connues, il semble qu'il puisse exister une hétérogénéité dans la stœchiométrie des protéines ribosomiques (Slavov *et al.*, 2015) et dans la présence des modifications des ARNr (Marcel *et al.*, 2013; Lafontaine, 2015). Ceci conduirait à l'existence de ribosomes spécialisés (Xue et Barna, 2012; Marcel *et al.*, 2013; Marcel *et al.*, 2015; Dinman, 2016). Ainsi, la composition en protéines et en nucléotides modifiés des ARNr permettrait la production de ribosomes aux fonctions distinctes, comme une spécificité d'action en tant que monosome ou polysome (Slavov *et al.*, 2015). Des travaux ont également montré une certaine interdépendance des modifications des ARNr, pouvant être le reflet de l'existence de différentes voies de spécialisation des ribosomes.

Ce concept d'hétérogénéité et de spécialisation des ribosomes pourrait également être à l'origine des différentes pathologies liées au ribosome. En effet, des dysfonctionnements de ces voies de spécialisation pourraient conduire à des phénotypes pathologiques appelés ribosomopathies. De plus, il a été montré que dans des cas de cancer, les ARNr sont retrouvés hypermodifiés (Lafontaine, 2015). Ceci suggère également que la méthylation partielle, due au fait que tous les sites de méthylation connus des ARNr ne soient pas systématiquement modifiés, est ce qui prévaut dans les conditions physiologiques normales et non-pathologiques.

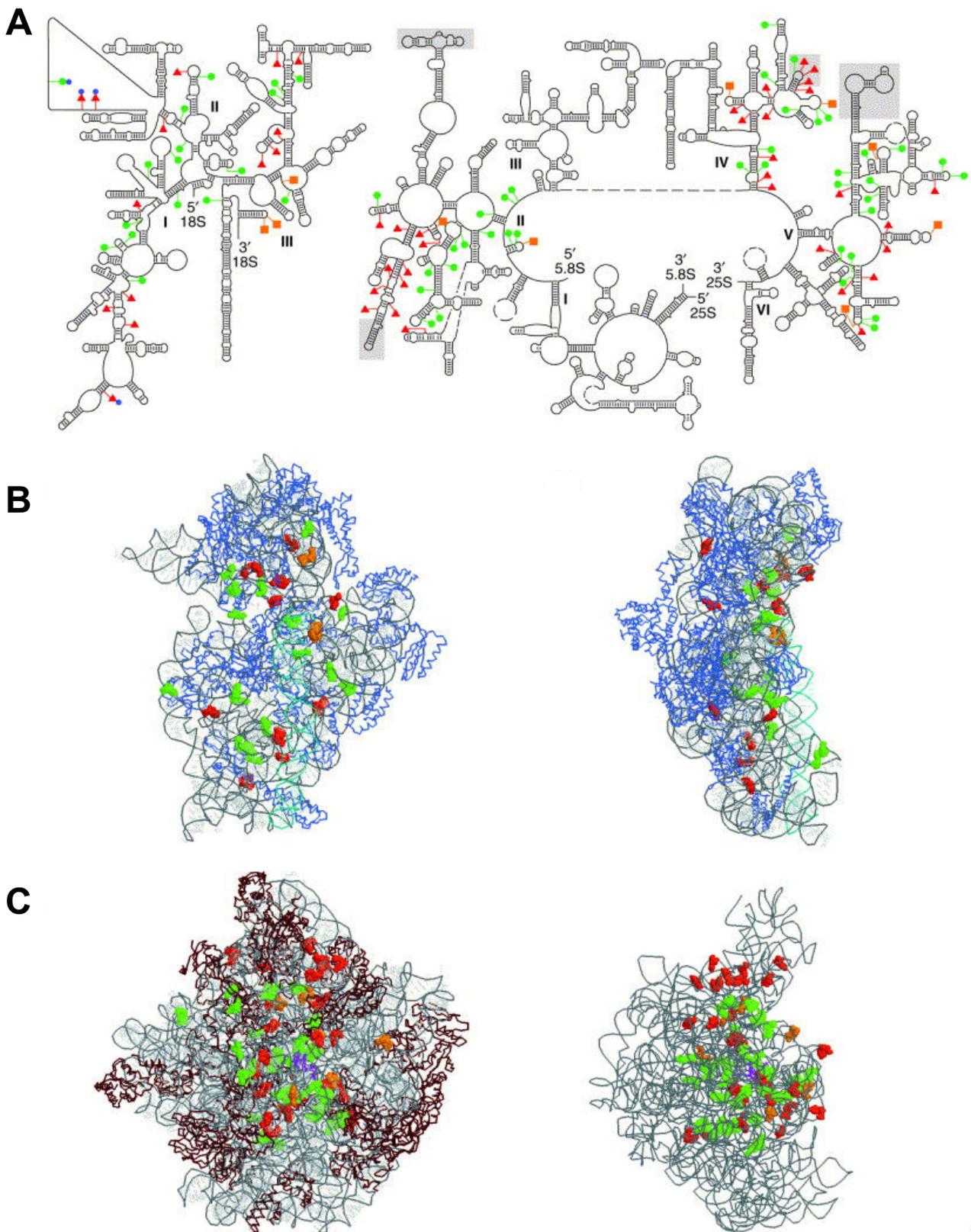


Figure 2. Sites de modification sur le ribosome de levure. (A) Schéma des structures secondaires des ARNr 18S (à gauche) et 25S-5,8S (à droite). Les pseudouridines sont symbolisées par des triangles rouges, les méthylations de bases par des carrés oranges et les 2'-*O*-méthylations des riboses par des ronds verts. (B) Positions des modifications sur la structure 3D de la petite sous-unité et (C) de la grande sous-unité (d'après Decatur et Fournier, 2002).

Introduction

1. Les snoRNP à boîtes C/D

Les snoRNP à boîtes C/D sont des particules constituées d'un ARN (snoRNA) et de quatre protéines. Les protéines au sein de ces particules s'agencent autour du snoRNA qui sert également de matrice de recrutement (guide) de la cible ARN à méthyler (Cavaillé *et al.*, 1996). La réaction de méthylation (2'-O-méthylation) est catalysée par une enzyme constitutive de la particule mature et cible notamment les ARNr (Matera *et al.*, 2007), mais également certains UsnRNA par des particules à boîtes C/D particulières, les scaRNP (Darzacq *et al.*, 2002) ou même des ARNt chez les archées (Dennis *et al.*, 2001).

Les scaRNA, composant ARN des scaRNP, sont retrouvés au sein des corps de Cajal dans le nucléoplasme des cellules eucaryotes, et participent aux modifications post-transcriptionnelles des UsnRNA du spliceosome transcrits par l'ARN polymérase II (Henras *et al.*, 2004). Certains scaRNA ont la particularité de pouvoir posséder les boîtes C/D et H/ACA sur un même ARN (Darzacq *et al.*, 2002).

Les snoRNA à boîtes C/D jouent généralement le rôle de guide pour assurer la méthylation d'un autre ARN cible (ARNr ou UsnRNA) et représentent aujourd'hui près de 83% des snoRNA canoniques connus chez l'Homme (227 sur les 275 connus). Les autres snoRNA canoniques (48 sur 275) n'ont actuellement pas de cible ARN connue et sont appelés snoRNA orphelins (Jorjani *et al.*, 2016). Ces snoRNA orphelins possèdent des fonctions autres que la fonction canonique des snoRNA, comme un rôle dans la régulation de l'épissage et de la maturation des pré-ARNm, comme par exemple l'ARN SNORD27 qui est impliqué dans l'épissage alternatif du pré-ARNm codant le facteur de transcription E2F7 ou encore en tant que facteurs associés à la chromatine (Dupuis-Sandoval *et al.*, 2015; Falaleeva *et al.*, 2016).

1.1. Les snoRNA à boîtes C/D

1.1.1. Organisation des gènes codant les snoRNA à boîtes C/D

La majorité des snoRNA à boîtes C/D chez les vertébrés (et quelques-uns chez la levure) sont codés dans le génome par des séquences introniques (Brown *et al.*, 2008). La transcription des gènes codant les snoRNA est donc dépendante de la transcription du gène hôte contenant cette séquence, ce qui permet l'expression coordonnée des ARN hébergés et protéines. Dans beaucoup de cas, les protéines codées par les séquences hôtes des gènes de

snoRNA introniques sont impliquées dans la biogenèse des ribosomes ou ont des fonctions nucléolaires.

Chez les vertébrés et la levure, la plupart des snoRNA introniques sont retrouvés à raison d'un seul snoRNA par intron (Chen *et al.*, 2008). En revanche, chez d'autres organismes eucaryotes, les gènes codant des snoRNA sont souvent retrouvés au sein de clusters introniques. Beaucoup de ces clusters sont constitués de gènes homologues, probablement issus de duplications locales. Chez les plantes, les clusters de gènes de snoRNA sont plus souvent hétérologues (Brown *et al.*, 2003).

Certains snoRNA ont également la particularité d'être codé dans le génome au sein de séquences introniques dont le gène hôte ne code pas une protéine mais un ARN non codant (Tycowski et Steitz, 2001; Bachellerie *et al.*, 2002). Cette particularité identifiée à l'origine chez la drosophile (gène *Uhg*) semble pouvoir être une organisation spécialement dédiée à l'expression de snoRNA avec un snoRNA par intron (Huang *et al.*, 2005a).

Les gènes codant des snoRNA non introniques peuvent être divisés en deux types, les gènes indépendants possédant leur propre promoteur et les gènes regroupés en cluster de gènes placés sous le contrôle d'un même promoteur (Brown *et al.*, 2003).

1.1.2. Transcription et maturation des snoRNA à boîtes C/D

Les gènes codant les snoRNA à boîtes C/D sont généralement transcrits par l'ARN polymérase II chez les vertébrés et chez la levure. Il existe toutefois des exceptions. Par exemple, chez la levures le gène *SNR52* codant le snoRNA à boîtes C/D snR52 est le seul gène codant un snoRNA à être transcrit par l'ARN polymérase III (Guffanti *et al.*, 2006).

La maturation des snoRNA à boîtes C/D introniques chez les vertébrés est couplée à l'épissage. Pour cela, les gènes des snoRNA à boîtes C/D sont toujours situés 70 nucléotides en amont du site 3' d'épissage (Hirose et Steitz, 2001; Hirose *et al.*, 2003). Les protéines cœur SNU13 et Fibrillarine s'associent d'ailleurs avec le complexe spliceosomal C1 (Hirose *et al.*, 2003). De plus, la protéine IBP160 (*Intron Binding Protein of 160 kDa*), qui est un facteur d'épissage de type hélicase présent dans le complexe C1, interagit avec la séquence située 33 à 40 nucléotides en amont du point de branchement quelle que soit cette séquence (Hirose *et al.*, 2006). Ainsi le fait que cette protéine interagisse avec le snoRNA situé proche du site de branchement interfère avec l'assemblage de la snoRNP. IBP160 est en effet requise pour l'assemblage des snoRNP à boîtes C/D. Cette protéine pourrait donc être le lien entre l'épissage et l'assemblage des snoRNP à boîtes C/D (Hirose *et al.*, 2006). La maturation de

ces snoRNA à boîtes C/D passe ensuite majoritairement par des étapes de clivages exonucléolytiques à partir des introns déjà excisés, et semble indépendante de l'épissage du pré-ARNm (Eliceiri, 1999). Les étapes de maturation des snoRNA introniques sont indépendantes du gène hôte, ce qui signifie que les snoRNA contiennent dans leur séquence les signaux nécessaires à leur maturation précise (Kiss et Filipowicz, 1995).

La voie majoritaire de maturation des snoRNA introniques implique dans un premier temps l'enzyme Dbr1 qui va réaliser le débranchement de l'intron comprenant le snoRNA, alors sous forme d'un lasso après l'étape d'épissage (Ooi *et al.*, 1998). La production de snoRNA introniques est donc dépendante de l'enzyme Dbr1 dans de nombreux cas. Une autre voie de maturation moins répandue génère, elle, un clivage endonucléolytique (Petfalski *et al.*, 1998; Ooi *et al.*, 1998; Weinstein et Steitz, 1999). Par la suite et dans chacune de ces deux voies, des attaques exonucléolytiques des extrémités 5' et 3' permettront de réduire celles-ci jusqu'à l'obtention d'un snoRNA précisément maturé aux extrémités correctes (Petfalski *et al.*, 1998; Weinstein et Steitz, 1999). Chez la levure, protéines Rat1 et Xrn1 sont les deux enzymes responsables des événements de clivages exonucléolytiques 5'→3' (Petfalski *et al.*, 1998). Les épisodes de clivages exonucléolytiques 3'→5' sont, eux, pris en charge par l'exosome (Allmang *et al.*, 1999; Terns et Terns, 2002), puis pour finir par la protéine Rrp6 (Allmang *et al.*, 1999). La génération de ces extrémités précises requiert la présence très précoce des protéines spécifiques aux snoRNP liées au snoRNA afin de le protéger des coupures exonucléolytiques (Watkins *et al.*, 1996; Terns et Terns, 2002).

Les snoRNA provenant d'unités transcriptionnelles polycistroniques indépendantes sont maturés en premier lieu par l'endonucléase Rnt1. Cette enzyme est l'orthologue de la RNase III bactérienne, et participe notamment à la maturation des snoRNA U14 et snR190 (Chanfreau *et al.*, 1998). Ces deux snoRNA sont en effet co-transcrits en un précurseur di-cistronique devant subir l'action endonucléolytique de la protéine Rnt1 afin de libérer les deux cistrons. Cette maturation se fait de concert avec l'enzyme Rat1 (Chanfreau *et al.*, 1998; Weinstein et Steitz, 1999).

1.1.3. Les boîtes C et D

Le nom de snoRNA à boîtes C/D vient du fait que deux séquences consensus anti-sens particulières appelées « boîtes » sont retrouvées chez ces ARN. La boîte C, située dans la région 5' du snoRNA, correspond à la séquence 5'-RUGAUGA-3' et la boîte D, située dans la région 3', à la séquence 5'-CUGA-3' (Figure 3A) (Bachellerie *et al.*, 2002). Ces deux

séquences anti-sens vont s'associer entre elles et ainsi induire le repliement du snoARN sur lui-même, puisque les deux boîtes sont situées aux extrémités 5' et 3'. L'appariement des séquences des deux boîtes va résulter en la formation d'une boucle interne asymétrique (Watkins *et al.*, 2000). Les snoRNA contiennent généralement un second couple de boîtes, appelé C'/D' et pouvant être dégénéré par rapport aux séquences consensus, dans la région centrale du snoRNA (Figures 3 et 4) (Kiss-László *et al.*, 1998).

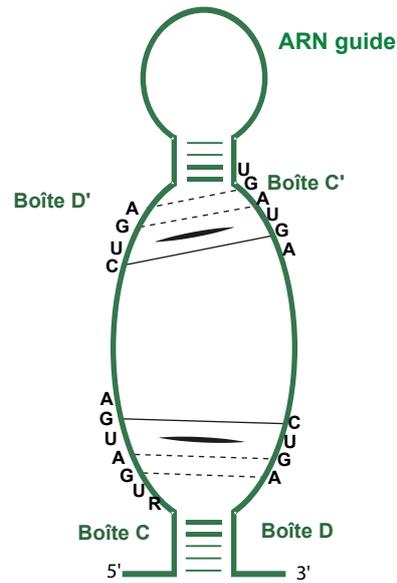
La spécificité d'un guide snoRNA pour son ARN cible repose sur une parfaite complémentarité de 10 à 17 nucléotides entre les deux séquences et au sein de laquelle se situe le nucléotide devant recevoir la modification (Figure 3B) (Kiss-László *et al.*, 1996). La séquence du snoRNA complémentaire à la cible se situe en amont des boîtes D et D' et son appariement par des interactions Watson-Crick avec sa cible va générer la formation d'un duplex ARN en double hélice. Pour les snoRNA à boîtes C/D, le 5^{ème} nucléotide situé en amont des boîtes D et D' fait toujours face au nucléotide de l'ARNr devant subir la méthylation dans le cadre du duplex formé entre le snoRNA et son ARN cible (Figure 3B) (Kiss-László *et al.*, 1996).

1.1.4. Les snoRNA à boîtes C/D possèdent un motif structural en *K-turn*

L'appariement des deux boîtes C et D (ou C' et D'), et le fait que la boîte C soit plus longue que la boîte D vont induire la formation d'une structure particulière de l'ARN appelé *K-turn* (Figures 4A et B). La formation de ce motif structural de l'ARN est favorisée par la présence de la protéine L7Ae/Snu13/SNU13 (alias 15.5K) (archée/levure/humain) (Watkins *et al.*, 2000; Marmier-Gourrier, 2003).

Le motif en *K-turn* se caractérise par une organisation de type hélice-boucle-hélice comprenant environ 15 nucléotides (Vidovic *et al.*, 2000). Dans cette structuration, deux hélices I et II encadrent une boucle interne dissymétrique (5+2) riche en purines (Figure 4). Cette organisation génère la formation d'une courbure de l'ARN d'environ 60°, notamment par la présence de trois nucléotides de la boîte C non appariés (en rouge et bleu sur la Figure 5) ainsi que deux appariements non Watson-Crick G•A et A•G successifs (en magenta sur la Figure 5). L'hélice I, précédant le coude, est une hélice canonique tandis que l'hélice II est non-canonique du fait de la présence d'un appariement U•U (en vert sur la Figure 5) faisant suite aux deux appariements G•A et A•G (Figures 5A à C). Ce motif structural est stabilisé par plusieurs empilements de bases (forces de *stacking*). En effet, de tels empilements de bases sont visibles entre le premier résidu non apparié de la boîte C et

A



B

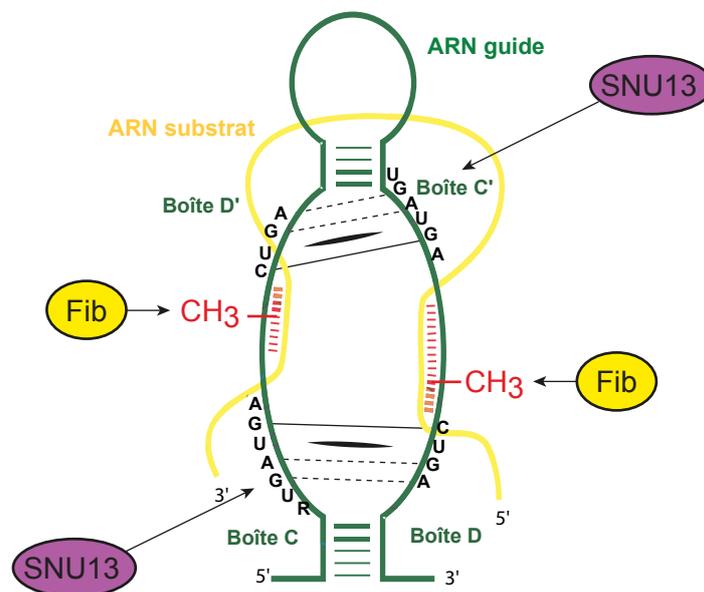


Figure 3. Le snoRNA eucaryote. (A) Schéma d'un snoRNA eucaryote avec deux jeux de boîtes C/D et C'/D'. (B) Schéma du snoRNA eucaryote avec l'ARN substrat en jaune. L'emplacement du site de 2'-O-méthylation est indiqué (CH₃).

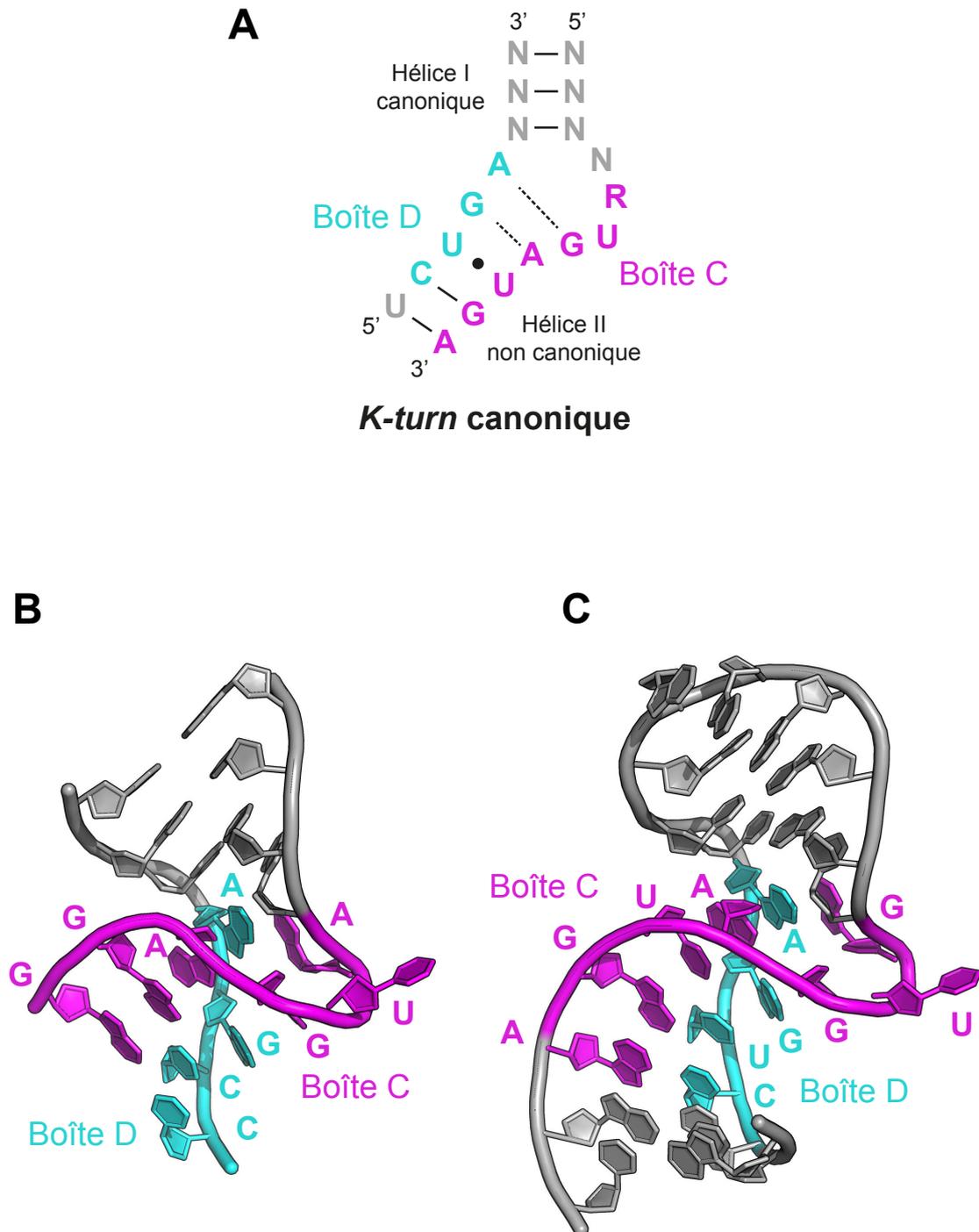


Figure 4. Formation des motifs en *K-turn* et *K-loop* par les boîtes C/D. (A) Représentation schématique de la structure secondaire de la séquence consensus du motif en *K-turn* canonique des snoRNA. Les nucléotides colorés en magenta correspondent à la boîte C et ceux en cyan à la boîte D. Les couleurs correspondent à celles des structures en (B) et (C). (B) Structure tridimensionnelle du motif en *K-turn* du snRNA U4 humain (code PDB : 1E7K) au sein du complexe avec la protéine SNU13 (Vidovic *et al.*, 2000), dont la représentation a été masquée pour des raisons de clarté. (C) Structure tridimensionnelle d'un motif *K-loop* de la séquence minimale d'un ARN à boîtes C/D synthétique comprenant une tige terminée par une boucle GAAA (code PDB : 1RLG) en complexe avec la protéine L7Ae (Moore *et al.*, 2004), dont la représentation a également été masquée pour des raisons de clarté.

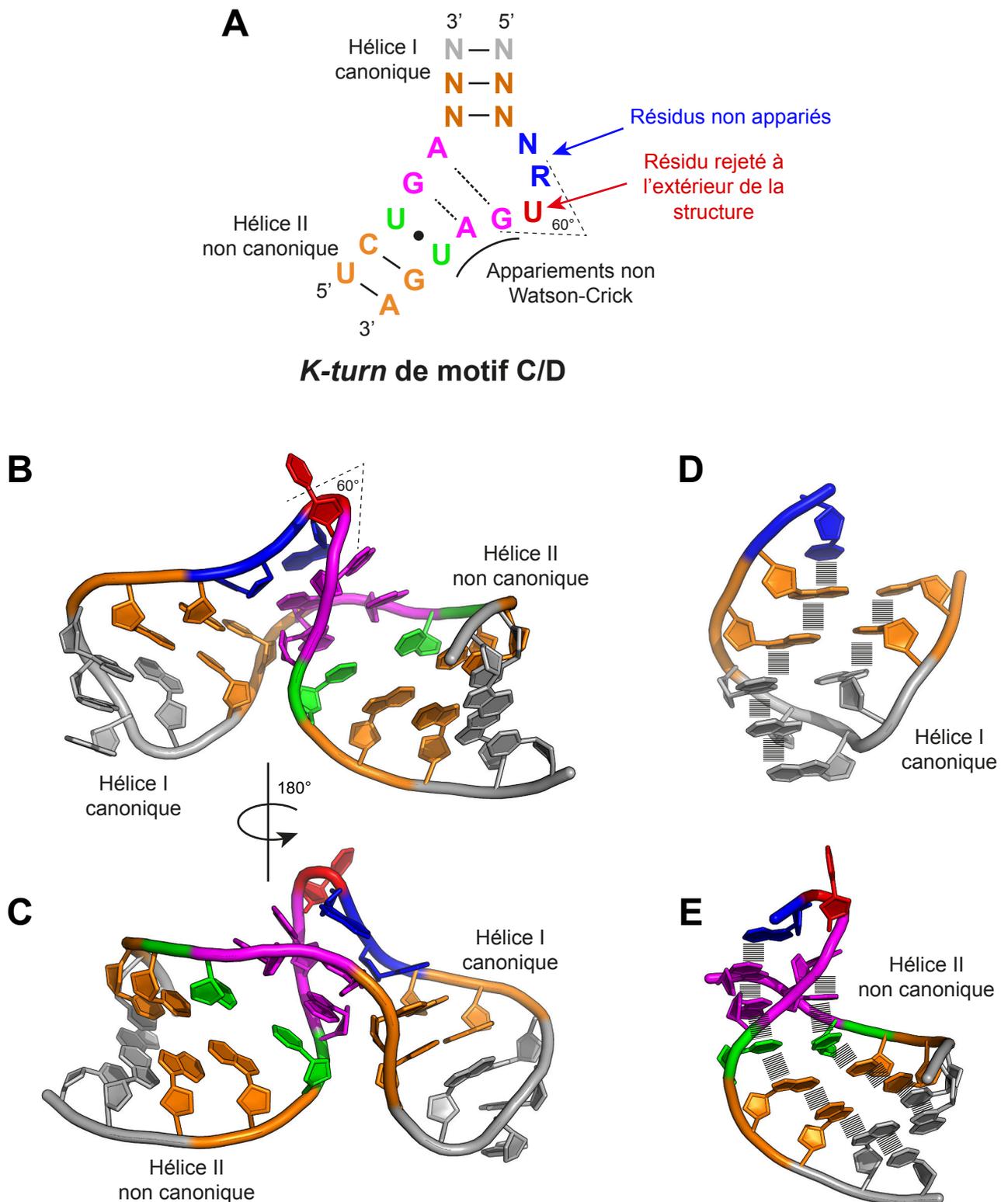


Figure 5. Le motif en *K-turn*. (A) Représentation schématique de la structure secondaire de la séquence consensus du motif en *K-turn* d'un motif C/D d'ARN. Les nucléotides colorés en orange (et gris) correspondent aux hélices I et II. Les résidus non appariés sont colorés en bleu et le résidu uridine rejeté hors de la structure de l'ARN en rouge. Les résidus réalisant des appariements non Watson-Crick, aussi appelées *sheared pairs*, sont colorés en magenta (paires A•G) ou en vert (paire U•U). Les résidus colorés en orange correspondent à des appariements Watson-Crick classiques. Les couleurs correspondent à celles des structures en dessous. Les panneaux (B) et (C) représentent la structure tridimensionnelle d'un motif *K-loop* de la séquence minimale d'un ARN à boîtes C/D synthétique (Moore *et al.*, 2004) selon deux axes d'observation différents de 180° verticalement. (D) et (E) Zooms sur les hélices I et II. Les forces de *stacking* sont représentées en pointillés.

l'hélice I (Figure 5D), ainsi qu'entre le deuxième et l'hélice II (Figure 5E). De plus, les paires G•A et A•G font également partie de l'empilement de bases avec l'hélice II (Figure 5E). Le seul résidu à ne pas prendre part à la stabilisation de cette structure est le troisième résidu non apparié (U), qui est lui rejeté hors de la structure (en rouge sur la Figure 5), pour l'interaction *via* une poche hydrophobe avec le représentant de la famille de la protéine L7Ae. Ce résidu uridine fait en plus partie d'un groupe de cinq nucléotides (sur les 7 composant la boucle interne asymétrique) dont la conformation du ribose est de type 2'-endo (Figure 6), contrairement à la conformation 3'-endo usuelle plus stable (Moore *et al.*, 2004). Le passage de la conformation 3'-endo à la conformation 2'-endo maximise l'exposition des squelettes phosphodiester de ces résidus de façon à favoriser leur interaction avec d'autres bases ou avec le représentant de la famille de la protéine L7Ae. Chez les sRNP/snoRNP à boîtes C/D, ce sont en effet les protéines de la famille L7Ae qui se retrouvent en interaction directe avec le motif en *K-turn* du snoRNA ou *K-loop* du sRNA, et ceci par le biais de ces spécificités d'ordre structural qu'il génère. Il a également été démontré que la formation totale en *K-turn* est dépendante de l'association avec son partenaire protéique de la famille L7Ae. Lorsque le snoRNA se trouve libre en solution, un échange dynamique s'établit entre les formes non-structurée et structurée en *K-turn* (Marmier-Gourrier, 2003; Falb *et al.*, 2010).

Une variante du motif en *K-turn* est particulièrement retrouvée chez les archées, où l'hélice I, entre les boîtes C' et D', est remplacée par seulement 3 nucléotides. Ce raccourcissement va générer la formation d'une boucle terminale (Figure 4C). Ce motif, appelé *K-loop*, est similaire au *K-turn* tant par sa séquence que par sa structure, mais n'est constitué que d'une simple tige fermée par une boucle terminale (Nolivos *et al.*, 2005). Ce motif conserve toutefois sa capacité d'interaction avec la protéine L7Ae. L'intérêt de ce motif serait d'améliorer la stabilité de l'ARN, notamment aux fortes températures auxquelles peuvent être soumis les organismes hyperthermophiles chez qui il est principalement retrouvé.

1.2. Composition des snoRNP à boîtes C/D matures

Les particules snoRNP à boîtes C/D matures et fonctionnelles se composent donc d'un snoRNA à boîtes C/D (guide) pouvant recruter un ou deux ARN cibles, ainsi que de quatre protéines distinctes. Ces protéines constitutives des particules matures sont appelées protéines cœur. La première, la protéine Snu13 chez la levure, sert de centre nucléateur à l'assemblage de la particule et aide à la stabilisation de la conformation du snoRNA en *K-turn*. La réaction de 2'-O-méthylation réalisée par la snoRNP est catalysée par une enzyme, appelée Nop1 chez

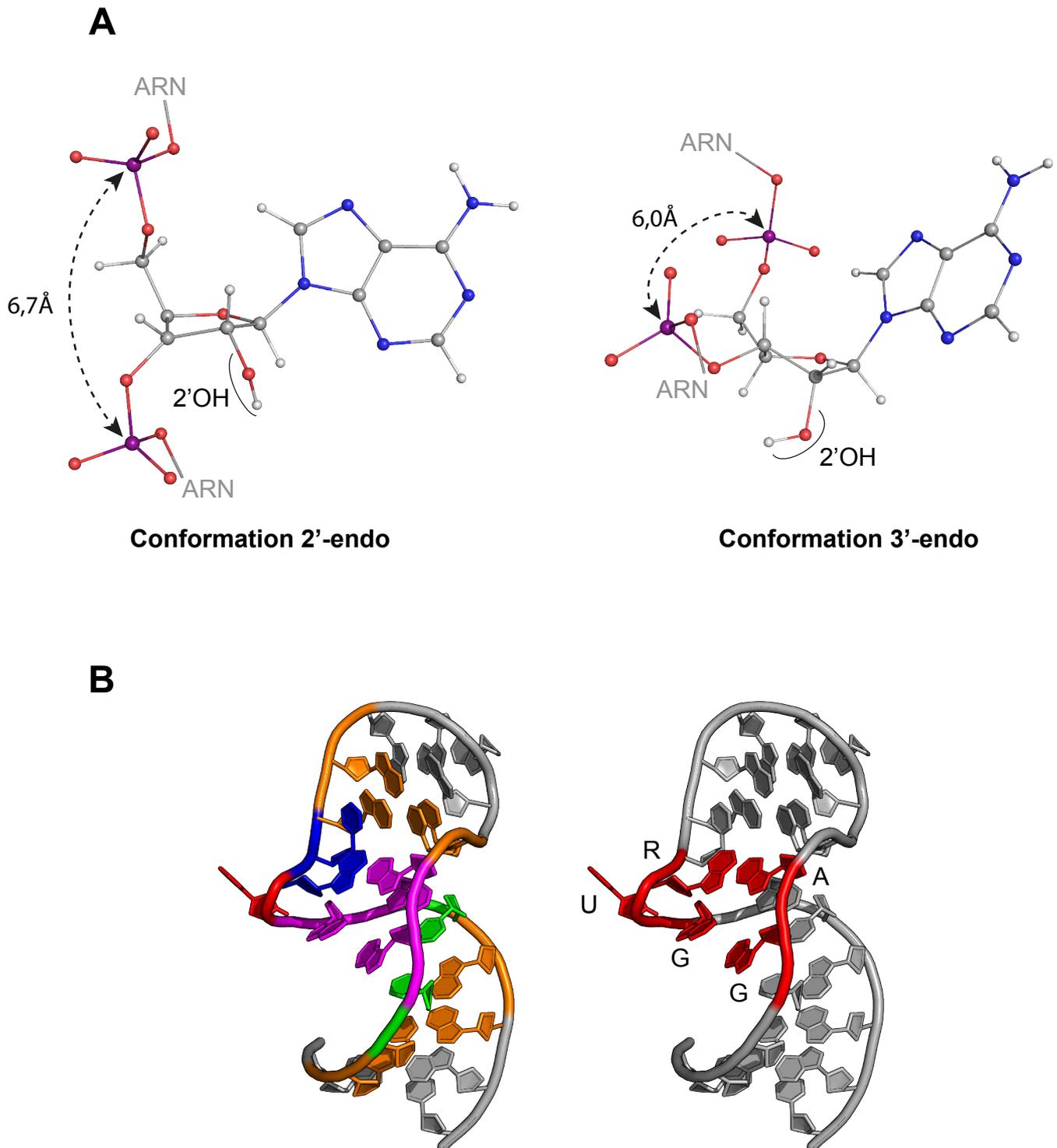


Figure 6. 5 nucléotides de l'ARN en *K-turn* ou *K-loop* ont une conformation de type 2'-endo. **(A)** Représentation des conformations 2'-endo (à gauche) et 3'-endo (à droite) du ribose d'un résidu adénosine. Les deux distances indiquées séparent les deux groupements phosphate encadrant le nucléoside (les atomes de phosphate sont colorés en violet). **(B)** Report sur la structure de l'ARN (le code couleur de la représentation de gauche correspond à celui de la figure 3) des 5 résidus annotés possédant une conformation 2'-endo (en rouge sur la représentation de droite).

la levure. Enfin, les deux protéines Nop56 et Nop58 de levure forment un hétérodimère impliqué dans la structuration de la particule. Les différents noms des protéines cœur de la snoRNP à boîtes C/D en fonction des organismes archée, levure et humain sont résumés dans la Table 1.

1.2.1. La famille L7Ae (la protéine Snu13/SNU13)

Initialement appelée 15.5K en raison de sa taille, la protéine SNU13 humaine a été la première de sa famille (L7Ae) à voir sa structure élucidée en complexe avec un ARN (le snRNA U4) (Vidovic *et al.*, 2000). Par la suite d'autres protéines de cette famille issues de divers organismes, comme les archées (Moore *et al.*, 2004) ou la levure *S. cerevisiae* (Oruganti *et al.*, 2005), ont fait l'objet d'études structurales. Toutes les structures connues des protéines de cette famille présentent des repliements identiques (Figure 7), constitués d'une structure en sandwich α - β - α , comprenant 5 hélices α et 4 brins β . La reconnaissance et l'interaction avec l'ARN sont rendues possibles par la particularité structurale du motif en *K-turn* impliquant le rejet du résidu uridine (troisième résidu non apparié) vers l'extérieur de la structure ARN. Ce résidu peut alors être stabilisé au sein d'une poche hydrophobe de la protéine SNU13, contenant également deux résidus polaires importants dans cette stabilisation, les résidus E61 et K86 chez l'Homme (Figure 7A). Ces résidus sont également retrouvés chez la protéine L7Ae des archées (D54 et K79) (Figure 7B) et chez la protéine Snu13 de levure (E59 et R84). Ajouté à cela, le squelette ribose-phosphate des résidus de l'hélice II et les résidus de la boucle interne de l'ARN interagissent préférentiellement avec les résidus contenus dans les éléments de structures secondaires de la protéine SNU13, à savoir les hélices α_2 et α_4 , ainsi que le brin β_1 et les boucles β_1 - α_2 , β_2 - α_3 et α_4 - β_4 (Figure 7). Parmi ces éléments, l'hélice α_2 présente un motif conservé NExxK, dont les résidus asparagine, glutamine et lysine participent à la stabilisation de l'ARN en établissant des contacts avec les résidus guanine des deux paires non Watson-Crick G•A et A•G (en magenta sur la Figure 7). Le motif VSRP compris dans la boucle α_4 - β_4 de la protéine SNU13 (motif IEVP chez la protéine L7Ae des archées) établit des contacts avec les résidus non appariés de la boucle interne de l'ARN (en bleu et rouge sur la Figure 7). Il a été montré que la présence de l'hélice II de l'ARN n'est pas requise pour la liaison avec la protéine SNU13, mais indispensable au recrutement des autres protéines cœur des snoRNP à boîtes C/D, dont notamment NOP58 (Watkins *et al.*, 2002). De plus, la protéine SNU13 est également nécessaire au recrutement de ces autres protéines cœur. La formation du complexe

Type de protéine	Archée	Levure <i>Saccharomyces cerevisiae</i>	Humain
Protéines cœur	L7Ae	Snu13	SNU13/15.5K
	aNOP1	Nop1	Fibrillarine
	aNOP5	Nop56	NOP56
	aNOP5	Nop58	NOP58
Composants du complexe R2TP	x	Rvb1	RuvBL1/TIP49
	x	Rvb2	RuvBL2/TIP48
	x	Tah1	RPAP3
	x	Pih1/Nop17	PIH1D1
Facteurs d'assemblage	x	Rsa1	NUFIP1
	x	Hit1	TRIP3
	x	Bcd1	BCD1

Table 1. Les protéines liées aux sRNP/snoRNP à boîtes C/D et les protéines de la machinerie d'assemblage de ces RNP.

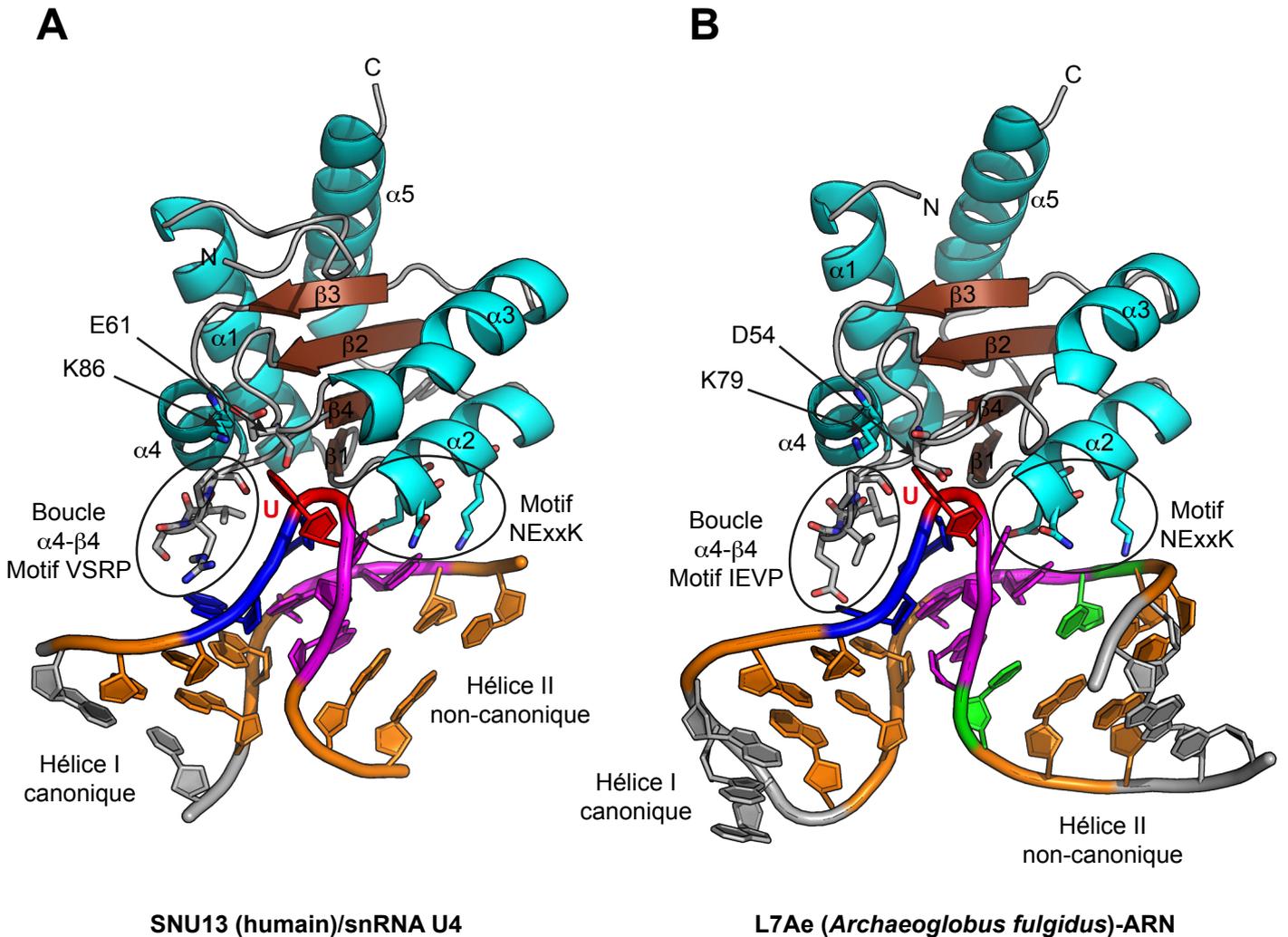


Figure 7. Le complexe L7Ae-ARN. Les panneaux (A) et (B) représentent les structures 3D obtenues par cristallographie aux rayons X sur les complexes SNU13-*K-turn* (code PDB : 1E7K ; Vidovic *et al.*, 2000) et L7Ae-*K-loop* (code PDB : 1RLG ; Moore *et al.*, 2004) respectivement. Le code couleur des ARN correspond à celui des figures précédentes. Les hélices α des protéines sont colorées en cyan et les brins β en marron. Les acides aminés importants dans l'interaction avec l'ARN et dans la stabilisation du *K-turn* sont représentés en bâtonnet et annotés.

ARN:SNU13 constitue donc la première étape de la formation de la particule, et semble donc créer une plateforme d'arrimage pour les autres protéines constitutives.

Enfin, il est à noter que protéine L7Ae a une plus faible spécificité vis-à-vis de sa cible ARN puisque, contrairement à ses homologues eucaryotes qui n'interagissent qu'avec les motifs en *K-turn*, elle est capable d'interagir indifféremment avec les motifs en *K-turn* et en *K-loop* (Charron *et al.*, 2004). Cette différence dans la capacité à reconnaître ou non les motifs en *K-loop* doit pouvoir s'expliquer structuralement par la divergence de résidus composant la boucle $\alpha 4\text{-}\beta 4$, où le motif VSRP chez les eucaryotes (entièrement conservé entre levure et humain) est remplacé par un motif IEVP chez l'archée *Archaeoglobus fulgidus* (dont les résidus E et V sont strictement conservés chez d'autres organismes archées) (Hamma et Ferré-D'Amaré, 2004).

1.2.2. La 2'-O-méthylase

Pour les snoRNP à boîtes C/D, la réaction de 2'-O-méthylation est catalysée par une enzyme, appelée aNOP1/Nop1/Fibrillarine (archée/levure/humain). Le nom de cette protéine chez l'Homme vient de sa découverte initiale dans les régions fibrillaires du nucléole de cellules humaines (Ochs *et al.*, 1985). La Fibrillarine humaine se divise en deux domaines distincts, un domaine GAR riche en glycines et arginines situé du côté N-terminal et un domaine MTase (Méthyltransférase) dans sa région C-terminale (Wang *et al.*, 2000; Rodriguez-Corona *et al.*, 2015). Le domaine GAR de la Fibrillarine humaine est nécessaire à sa localisation nucléaire et est méthylyé sur certains résidus arginine. L'organisation du domaine catalytique de cette enzyme (domaine MTase ne comprenant pas les brins $\beta 1'$ à $\beta 3'$ sur les représentations de la Figure 8) est similaire aux structures de types *Rossmann fold* (ou Pli de Rossmann) de type méthyltransférase (Wang *et al.*, 2000). Cependant, l'analyse de sa séquence comparée à celles d'autres méthyltransférases spécifiques de la 2'-O-méthylation possédant un *Rossmann fold* classique montre une divergence (Feder *et al.*, 2003) probablement signe que les 2'-O-méthyltransférases autonomes (de la famille SPOUT) et les méthyltransférases guidées par les snoRNA (de la famille Fibrillarine) auraient une origine évolutive indépendante. Ce domaine *Rossmann fold-Like* se compose d'un feuillet central de 7 brins β flanqué de part et d'autre par 3 hélices α (Figures 8A et B), formant une structure en sandwich $\alpha\text{-}\beta\text{-}\alpha$. Ce type de topologie est d'ailleurs caractéristique des méthyltransférases SAM dépendantes de classe I, tandis que la famille SPOUT appartient à la classe IV (Schubert *et al.*, 2003). Au sein de la RNP mature, cette enzyme est capable de substituer

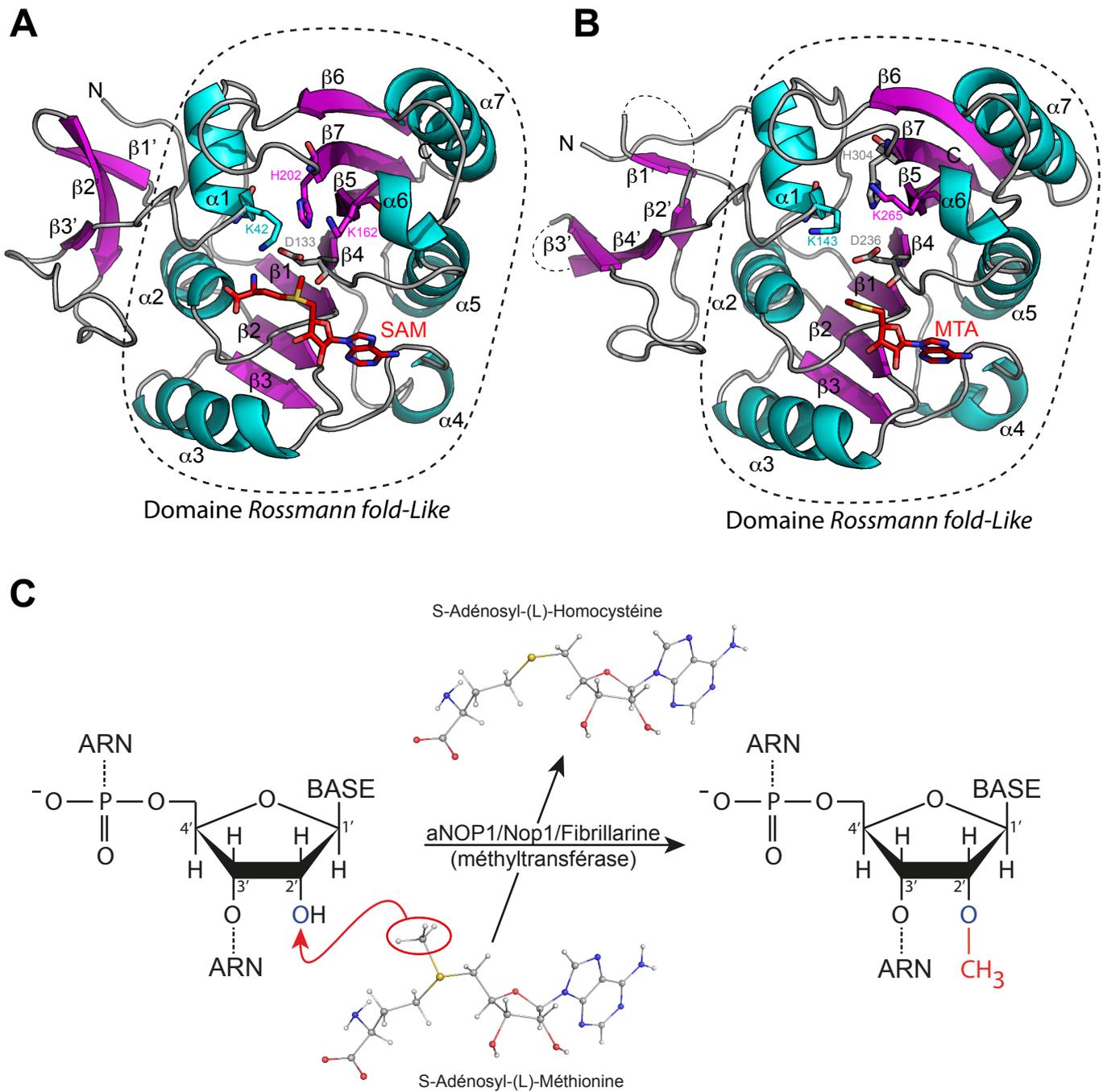


Figure 8. L'enzyme 2'-*O*-méthyltransférase Fibrillarine. (A) Structure tridimensionnelle de la protéine aNOP1 de l'archée *Archaeoglobus fulgidus* en complexe avec la protéine aNOP5 (non représentée) déterminée par Aittaleb *et al.*, 2003 (code PDB : 1NT2). Le cofacteur SAM (S-Adénosyl-(L)-méthionine) est représenté en bâtonnets rouges. (B) Structure tridimensionnelle du domaine MTase de la protéine Fibrillarine humaine (code PDB : 2IPX) par Wu *et al.*, en 2006 (non publié). Le MTA (5'-deoxy-5'-méthylthioadénosine), un analogue de la SAM est représenté en bâtonnets rouges. (C) Schéma de la réaction de 2'-*O*-méthylation du ribose d'un ARN par l'enzyme aNOP1/Nop1/Fibrillarine (archée/levure/humain) couplée à la SAM comme cofacteur donneur de groupement méthyle. Cette réaction substitue l'atome d'hydrogène porté par l'oxygène du carbone C2' du ribose par un groupement méthyle apporté par le cofacteur SAM (dont la structure est représentée en lignes et sphères). Le groupement méthyle partant porté par l'atome de soufre (en jaune) est entouré en rouge.

l'hydrogène du groupement hydroxyle (-OH) porté par le carbone 2' du ribose de l'ARN cible par un groupement méthyle issu de son cofacteur S-Adénosyl-(L)-Méthionine (SAM) (Figure 8C) (Wang *et al.*, 2000; Aittaleb *et al.*, 2003). L'activité de cette protéine, comme les autres membres de la famille des 2'-O-méthyltransférases de ribose guidées par les snoRNA, repose sur la présence d'une tétrade conservée (K-D-K-H) proche de la poche de fixation du substrat (Figures 8A et B et 9B) et dont la triade K-D-K est essentielle à la catalyse (Hager *et al.*, 2002; Feder *et al.*, 2003). La position du site actif de cette enzyme à proximité du site à méthyler (Figure 9) est précisément définie par la structuration globale de la particule et particulièrement grâce à l'ARN guide et au domaine catalytique (N-terminal) des protéines aNOP5. Ce site se retrouve en effet toujours 5 nucléotides en amont de la boîte D (symbolisé par une étoile rouge sur la Figure 9A). Il a toutefois récemment été montré que les boîtes dégénérées C'/D' pouvaient être responsables de la méthylation d'une même région de l'ARNr cible sur plusieurs sites (van Nues et Watkins, 2016). Le mécanisme proposé pour cette réaction de 2'-O-méthylation est qu'après une déprotonation du groupement hydroxyle du ribose, une attaque nucléophile soit réalisée par cet atome d'oxygène déprotoné sur le groupement méthyle de la SAM dans une réaction de type S_N2 . Ce groupement méthyle est un bon groupement partant, car il est porté par un atome de soufre électropositif. Il est alors transféré à la place de l'hydrogène du groupement hydroxyle du ribose (Hager *et al.*, 2002). La déprotonation du groupement hydroxyle est principalement due à la proximité des chaînes latérales de la lysine 60 et de l'aspartate 153 (Figure 9B) chez *Sulfolobus solfataricus*, équivalents de la lysine 143 et de l'aspartate 236 de la Fibrillarine humaine (Figure 8B) (Hager *et al.*, 2002; Aittaleb *et al.*, 2003).

La Fibrillarine n'est active qu'au sein d'une structure de type sRNP/snoRNP. Sous forme libre, elle n'est en effet pas en mesure de catalyser la réaction de méthylation du substrat ARN. Au sein de la sRNP elle établit en effet des interactions avec les ARN guide et substrat (Figure 9), grâce aux interactions avec les autres protéines cœur SNU13 et NOP5 (Wang *et al.*, 2000; Galardi *et al.*, 2002). La protéine Nop1/Fibrillarine est néanmoins capable de réaliser la méthylation de protéines. Il a notamment été montré que ces protéines de levure et humaines étaient des histones méthyltransférases (Tessarz *et al.*, 2014). Plus particulièrement, il a été montré qu'elles étaient responsables de la méthylation des glutamines Q105 (levure) et Q104 (humaine) de l'histone H2A. Chez la levure, les histones H2A présentant une méthylation au niveau de la glutamine 105 sont fortement retrouvées au niveau de la région d'ADNr 35S (Tessarz *et al.*, 2014).

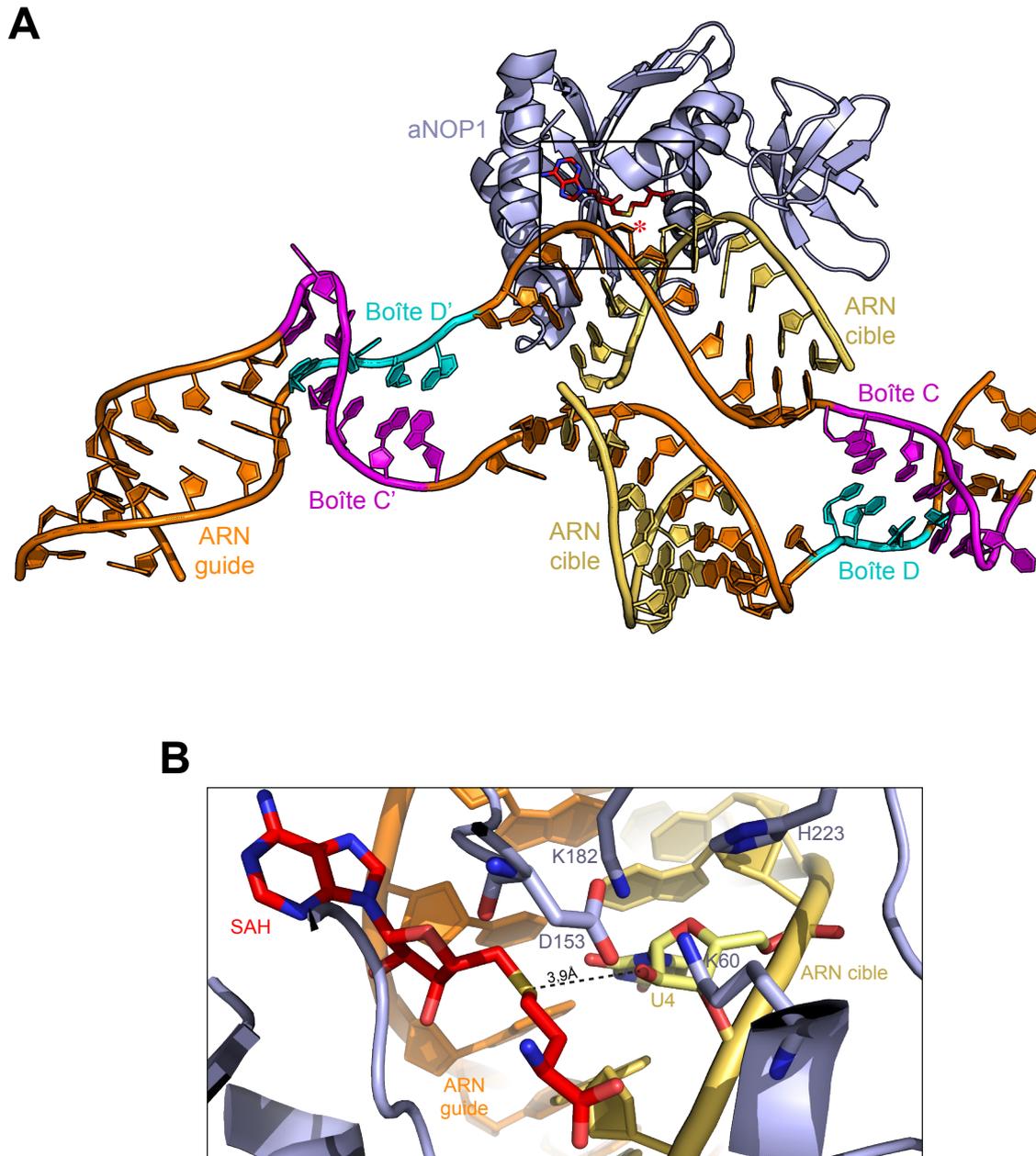


Figure 9. 2'-O-méthylation de l'ARN cible. (A) Structure tridimensionnelle de la protéine aNOP1 (2'-O-méthyltransférase) en complexe avec les ARN guide (orange) et cible (jaune). Cette représentation est issue de la structure de la particule sRNP complète de *Sulfolobus solfataricus* obtenue par Lin *et al.*, 2011 (code PDB : 3PLA) où les autres protéines coeur ont été cachées par soucis de clarté. L'étoile rouge symbolise le 5^{ème} nucléotide en amont de la boîte D, apparié avec le nucléotide à modifier. Dans cette structure, la particule a été cristallisée en présence de SAH (en rouge) et non de SAM, ce qui explique que la méthylation du substrat n'ait pas eu lieu. Le SAH est le produit de la réaction après transfert du groupement méthyle du SAM. La zone située dans le cadre noir est agrandie dans le panneau (B) et les résidus de la tétrade conservée (K-D-K-H) de aNOP1 sont représentés en bâtonnets. Le groupement hydroxyle en position 2' du ribose du nucléotide U4 est à proximité de l'atome de soufre du SAH, atome sur lequel est normalement greffé le groupement méthyle à transférer sur le cofacteur SAM.

1.2.3. Les protéines Nop56/NOP56 et Nop58/NOP58

Les deux dernières protéines constitutives des snoRNP à boîtes C/D, les protéines de levure Nop56 et Nop58, partagent une forte homologie (Gautier *et al.*, 1997). En effet, ces deux protéines possèdent 45% d'identité entre leurs séquences. Toutes deux sont localisées dans le nucléole des cellules eucaryotes. Ces deux protéines paralogues semblent être issues d'un phénomène de duplication de gène ayant eu lieu chez les eucaryotes uniquement (Omer *et al.*, 2002). En effet, chez les archées ces deux protéines ne correspondent qu'à une seule et même protéine, la protéine aNOP5. Toutefois si une seule protéine chez les archées peut remplacer ces deux protéines, chez les eucaryotes l'une n'est pas en mesure de compléter un défaut de l'autre (Gautier *et al.*, 1997).

Les régions C-terminales des protéines Nop56 et Nop58 contiennent des séquences répétées KKE/D (répétée 11 et 13 fois respectivement) qui ne sont pas retrouvées chez les protéines humaines (Gautier *et al.*, 1997). En revanche, chez les protéines humaines ces séquences sont remplacées par des portions successives fortement acides ou basiques, assurant en tant que NoLS la localisation nucléolaire (Pradet-Balade *et al.*, 2011), ce qui n'est pas le cas des protéines de levure. La protéine Cbf5 de levure, qui est l'enzyme assurant la réaction de pseudouridylation au sein des particules snoRNP à boîtes H/ACA, possède également ce type de séquence répétée KKE/D. Il a été montré que les protéines Nop56, Nop58 et Cbf5 étaient capables d'interagir avec la protéine Tgs1 (hyperméthylase) *via* leurs domaines C-terminaux contenant les répétitions KKE/D (Mouaikel *et al.*, 2002). La délétion de la protéine Tgs1 ou des régions C-terminales des protéines Nop56 ou Nop58 affecte la structure du nucléole et la localisation des snoRNP (Colau *et al.*, 2004). Il est donc possible qu'au cours de l'évolution l'interaction entre la protéine Tgs1 et les répétitions KKE/D situées aux extrémités C-terminales des protéines Nop56 et Nop58 ait conservé un double rôle dans l'hyperméthylation de la coiffe et dans le transport des snoRNP (Colau *et al.*, 2004; Pradet-Balade *et al.*, 2011). L'exportine CRM1 semble interagir avec la protéine Tgs1 et provoquer sa dissociation de Nop58, ce qui pourrait avoir une importance dans la localisation de cette protéine et donc des snoRNP (Pradet-Balade *et al.*, 2011). L'équivalent chez les archées des protéines Nop56 et Nop58, la protéine aNOP5, ne possédant pas ces répétitions KKE/D, il est possible que l'émergence des triméthylases de petits ARN chez les eucaryotes ait conduit à une co-évolution avec leur substrat et donc mené à l'acquisition de ces motifs répétés (Colau *et al.*, 2004).

Actuellement, aucune structure des protéines Nop56 ou Nop58 eucaryotes à résolution atomique n'est disponible. Toutefois, plusieurs structures de la protéine aNOP5 ont été déterminées chez différentes archées, et peuvent raisonnablement être transposées aux eucaryotes compte tenu de la forte similarité de séquence entre ces protéines. Il est tout de même à noter qu'aucune de ces structures n'a été déterminée sur la protéine aNOP5 seule (et entière), mais toujours en présence de la protéine aNOP1 *a minima* ou des autres composants de la sRNP. La structure de la protéine aNOP5 d'archée présente trois domaines distincts (Figure 10A), un domaine N-terminal, un domaine C-terminal (aussi appelé domaine NOP) et un domaine central reliant les deux précédents appelé domaine *coiled-coil* (super hélice) (Aittaleb *et al.*, 2003). Ce domaine en *coiled-coil* est constitué de deux hélices ($\alpha 4$ et $\alpha 5$) super-enroulées. Celui-ci permet la dimérisation entre deux protéines aNOP5 chez les archées ou entre les protéines Nop56 et Nop58 chez la levure en formant un fagot de 4 hélices super-enroulées. Ce domaine a un rôle structurel, puisqu'il sépare les deux couples de boîtes C/D. Il est à noter qu'une insertion (domaine Tip) est présente entre les deux hélices ($\alpha 4$ et $\alpha 5$) du domaine *coiled-coil* chez certains organismes par rapport à d'autres (Figure 10).

Le domaine N-terminal est lui responsable de l'interaction avec l'enzyme aNOP1 (Fibrillarine), avec quelques résidus du domaine C-terminal (Aittaleb *et al.*, 2003) dans le cadre d'un dimère isolé entre les deux protéines aNOP1 et aNOP5. Cependant, l'agencement des domaines de la protéine aNOP5 les uns par rapport aux autres diffère entre les structures établies dans la littérature (Figure 10B). Plus particulièrement, le domaine N-terminal peut se retrouver dans une position différente en fonction des cas. Il semble que la position de ce domaine soit dépendante des partenaires présents au contact de la protéine aNOP5. En effet, dans la structure établie par Lin *et al.*, 2011, le domaine N-terminal bascule à l'opposé du domaine C-terminal (NOP) par rapport au domaine *coiled-coil*, ce domaine NOP se retrouve ainsi éloigné de la protéine aNOP1 (Figures 10B et 11A). Il est possible que le positionnement de ce domaine soit dirigé par la présence ou non du substrat dans la particule (Lin *et al.*, 2011).

Le domaine C-terminal (NOP) de la protéine aNOP5 est capable d'interagir avec l'hélice II et le motif en *K-turn* du sRNA de manière concomitante à la protéine L7Ae (Figure 11). De plus, l'hélice $\alpha 9$ ' de aNOP5 se retrouve positionnée entre les deux brins de l'ARN au niveau de l'hélice II (Figure 11B), et semble d'ailleurs stabilisée par des contacts avec l'ARN car dans la structure de aNOP5 sans ARN cette région est déstructurée (invisible en densité) (Ye *et al.*, 2009).

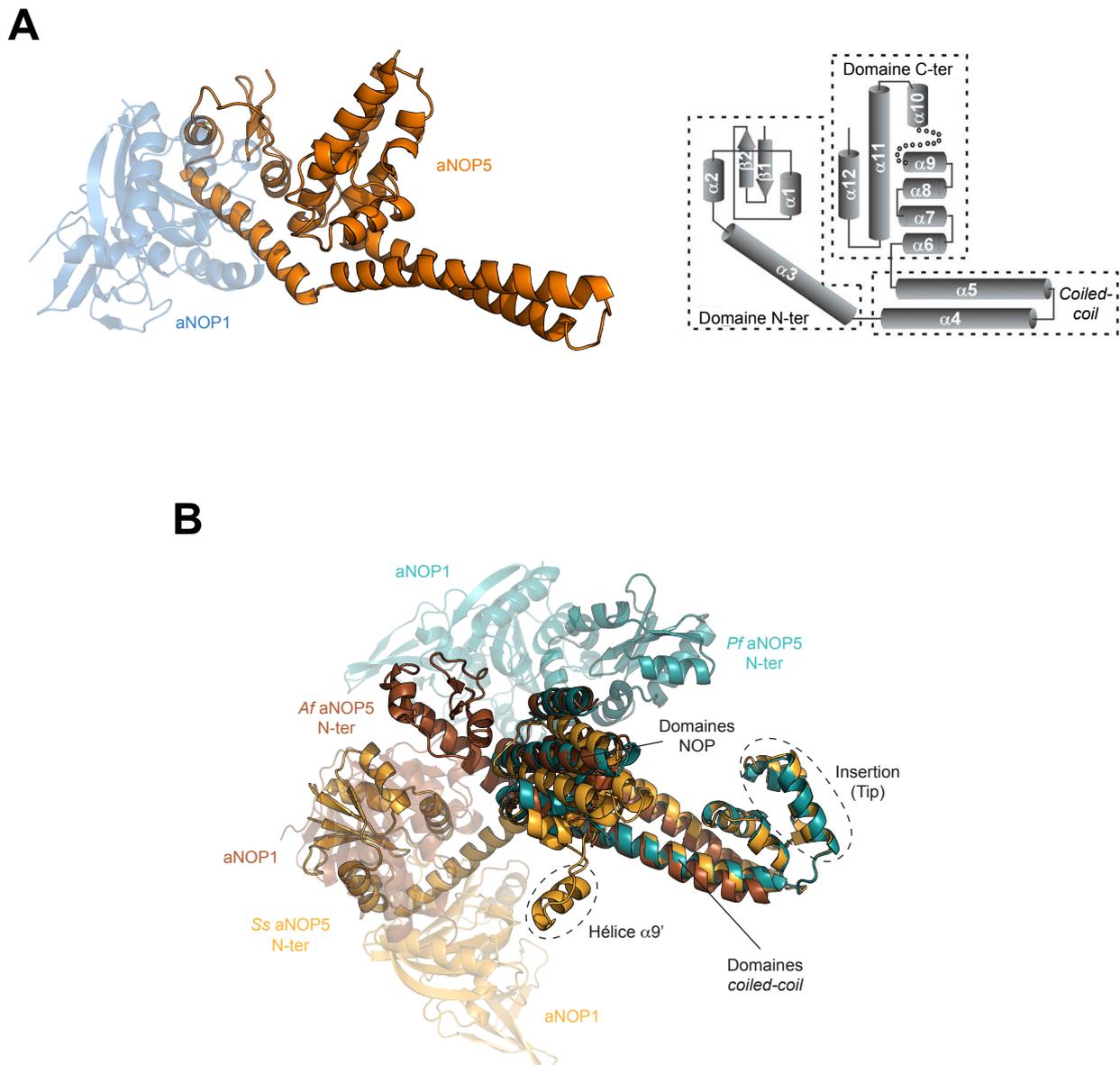


Figure 10. Structure et organisation des protéines aNOP5. (A) Structure tridimensionnelle du complexe entre les protéines aNOP5 (en orange) et aNOP1 (en bleu transparent) de *Archaeoglobus fulgidus* (Aittaleb *et al.*, 2003 ; code PDB : 1NT2). Le diagramme topologique correspondant à la protéine aNOP5 est représenté à droite. La région entre les hélices α_9 et α_{10} n'est pas représentée car invisible en densité électronique lors de la détermination de la structure par cristallographie aux rayons X. (B) Superposition par leurs domaines *coiled-coil* de trois structures entre les protéines aNOP5 et aNOP1 issues des organismes *Archaeoglobus fulgidus* (*Af*, en marron ; code PDB : 1NT2), *Pyrococcus furiosus* (*Pf* en bleu ; code PDB : 2NNW) et *Sulfolobus solfataricus* (*Ss*, en jaune ; code PDB : 3PLA).

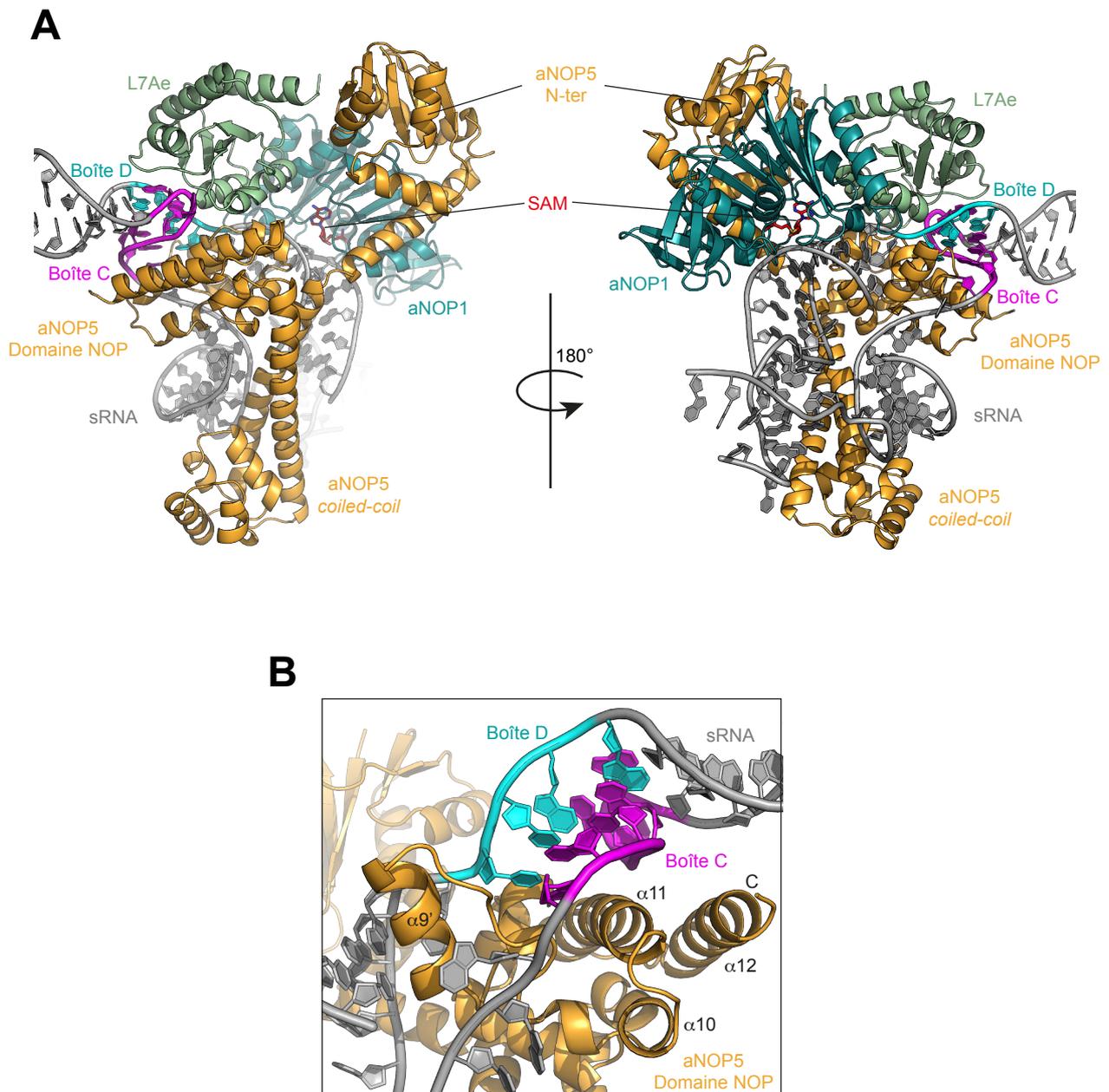


Figure 11. Interactions entre les protéines aNOP5 et l'ARN dans les sRNP à boîtes C/D d'archées. (A) Structure tridimensionnelle d'une demi-particule sRNP de *Sulfolobus solfataricus*, issue de la structure complète dont l'autre moitié n'est pas représentée dans un souci de clarté (Lin *et al.*, 2011 ; code PDB : 3PLA). (B) Zoom de la structure (A) sur le domaine NOP de la protéine aNOP5 en contact avec l'ARN.

Chez la levure, la protéine Nop56 ne s'associe avec les snoRNA qu'en présence de la protéine Nop1 (Lafontaine et Tollervey, 2000). En revanche, les protéines Nop58 et Nop1 sont capables de se lier au snoRNA indépendamment les unes des autres. Chez les archées, seule la protéine L7Ae peut se lier individuellement au sRNA. Les protéines aNOP5 et aNOP1, que ce soit seules ou ensemble, ne sont en effet pas capable de s'associer au sRNA en absence de la protéine L7Ae (Omer *et al.*, 2002).

1.2.4. Structure 3D des particules sRNP d'archées

Comme décrit précédemment, les particules sRNP/snoRNP à boîtes C/D se composent des protéines L7Ae/Snu13/SNU13, aNOP1/Nop1/Fibrillarine et aNOP5/Nop56-Nop58/NOP56-NOP58 (archée/levure/humain). De nombreux travaux ont porté sur la reconstitution d'une particule mature et fonctionnelle *in vitro* et sur la détermination de la structure 3D d'une telle particule. Cependant, jusqu'à aujourd'hui aucune structure de particule eucaryote n'a pu être déterminée, principalement car chez les eucaryotes l'assemblage d'une telle particule ne peut pas se faire simplement par mise en présence des différents constituants. En effet, le concours d'autres facteurs, dits facteurs d'assemblage, est nécessaire à l'obtention de particules actives. Ce n'est toutefois pas le cas des sRNP d'archées, sur lesquelles plusieurs équipes ont pu obtenir des informations structurales ces dernières années (Ye *et al.*, 2009; Bleichert *et al.*, 2009; Xue *et al.*, 2010; Lin *et al.*, 2011; Lapinaite *et al.*, 2013; Kornprobst *et al.*, 2016). La similarité de séquence entre les composants des sRNP d'archées et des snoRNP eucaryotes nous permet donc encore une fois d'utiliser le modèle des archées afin d'obtenir des informations probablement transposables aux eucaryotes. De plus, les caractéristiques structurales et fonctionnelles des snoRNP à boîtes C/D semblent avoir été conservées durant l'évolution entre les archées et les eucaryotes (Qu *et al.*, 2011) et étaient probablement présentes chez un ancêtre commun.

Même si plusieurs structures de sRNP sont désormais connues, deux hypothèses majeures persistent toujours, car toutes ces structures sont loin d'être similaires. Les informations dont nous disposons aujourd'hui mènent soit vers une particule monomérique (mono-sRNP) constituée d'un sRNA à deux jeux de boîtes C/D sur lequel vont s'agencer les protéines cœur par paires (L7Ae, aNOP1 et aNOP5), soit vers une particule dimérique (di-sRNP) où toutes les quantités citées précédemment sont doublées.

1.2.4.1. Hypothèse de la mono-sRNP

La première structure d'une sRNP (*Sulfolobus solfataricus*) déterminée par cristallographie aux rayons X (Ye *et al.*, 2009) a été obtenue en utilisant un demi sRNA (guide) de 35 nucléotides ne comprenant qu'un couple de boîtes C/D et une région guide de 10 nucléotides en amont de la boîte D (Figure 12). Dans cette structure, les domaines *coiled-coil* des protéines aNOP5 induisent la formation d'un dimère entre deux aNOP5 et sont au centre de la structure, qu'ils divisent en deux. D'une part, et liés aux domaines NOP des protéines aNOP5, les protéines L7Ae interagissent avec les demi-sRNA sur le motif en *K-turn*. D'autre part, les domaines N-terminaux des protéines aNOP5 sont chacun en contact avec une protéine aNOP1 contenant elle-même une molécule de SAM. Cette structure semble montrer qu'une sRNP contient bien deux exemplaires des protéines L7Ae, aNOP5 et aNOP1, et utilise deux jeux de boîtes C/D agencés symétriquement. Cependant, cette structure ne représente pas un état actif de la particule, puisque les ARN guides sont situés à l'opposé des enzymes aNOP1. La distance entre le cofacteur SAM de la protéine aNOP1 et le 5^{ème} résidu en amont de la boîte D sur l'ARN guide (coloré en noir) est en effet de l'ordre de 60-70 Å. La possibilité qu'il puisse exister des conformations distinctes de particules actives et non-actives est en accord avec le fait que les différentes structures connues de la protéine aNOP5 présentent de fortes différences dans la disposition des domaines les uns par rapport aux autres (Figure 10B). La flexibilité apparente entre les domaines N-terminal et *coiled-coil* pourrait donc être à l'origine d'un grand mouvement entre les domaines et ainsi rapprocher l'enzyme aNOP1 du substrat recruté par l'ARN guide et permettre le passage d'une conformation à l'autre (Ye *et al.*, 2009).

Par la suite, une autre structure de sRNP de *Pyrococcus furiosus* a été déterminée avec cette fois la présence d'un ARN substrat (Figure 13), mais toujours en présence de demi sRNA guide (Xue *et al.*, 2010). Ces travaux ont montré l'importance de l'hélice $\alpha 9'$ de la protéine aNOP5, et plus particulièrement du motif conservé GAEK présent sur cette hélice, dans la fusion de l'hélice de l'ARN guide et le positionnement de l'ARN substrat. Cette structure est similaire à la précédente dans l'organisation générale, avec toutefois une inversion de position entre les domaines N-terminaux des protéines aNOP5 et les protéines aNOP1. Ce changement de position de l'enzyme ne la rapproche toutefois pas suffisamment du substrat pour que cette particule soit active (environ 45 Å entre le cofacteur SAM et le résidu cible représenté en rouge).

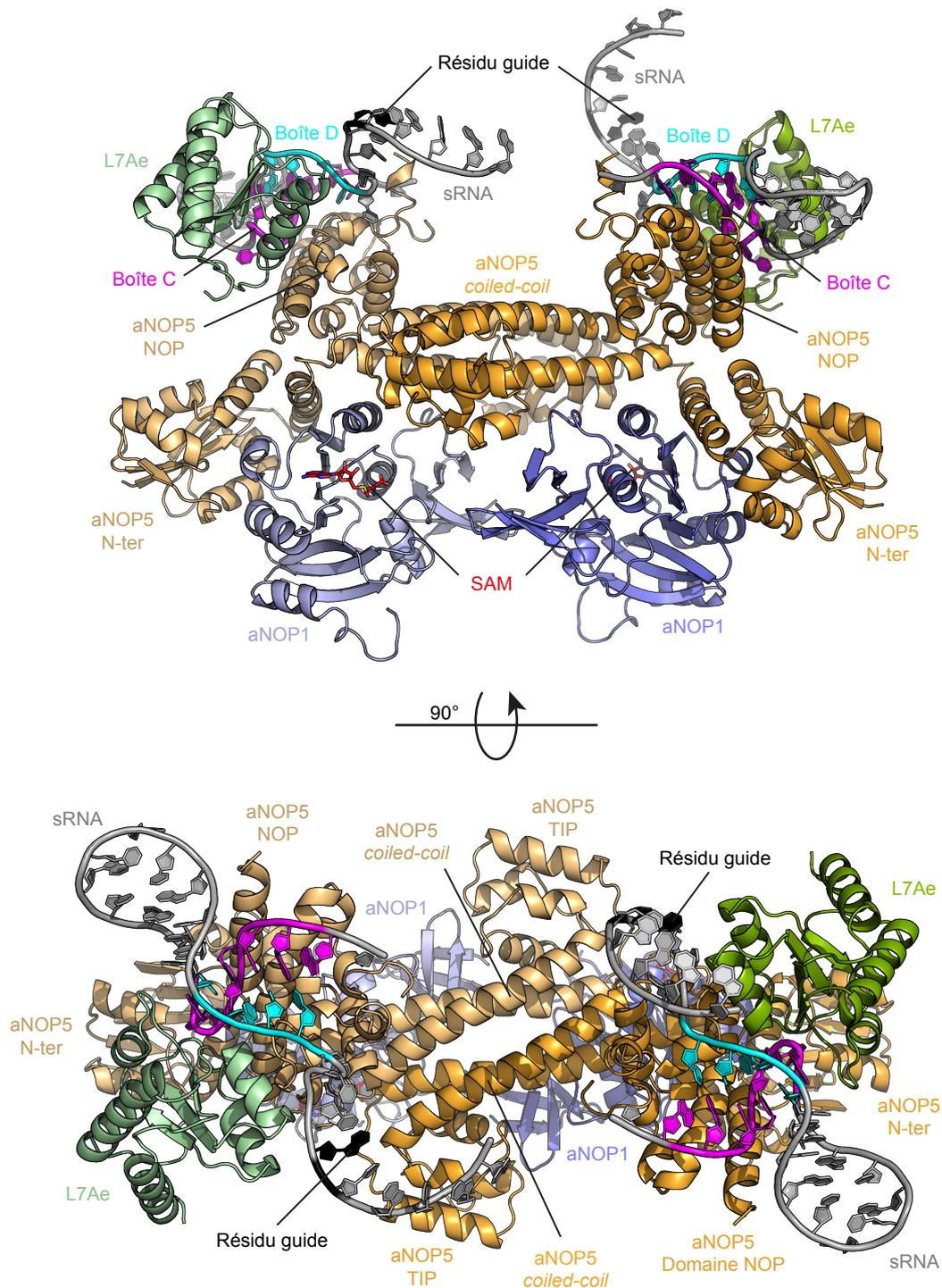


Figure 12. Structure tridimensionnelle d'une sRNP de l'archée *Sulfolobus solfataricus* avec deux demi-sRNA C/D et sans substrat en vue de côté (haut) ou du dessus (bas). Cette structure a été obtenue par cristallographie aux rayons X (Ye *et al.*, 2009 ; code PDB : 3ID5). Les résidus colorés en noir sur la structure du sRNA (ARN guide) correspondent aux 5^{èmes} résidus en amont des boîtes D, devant servir de guide à la 2'-*O*-méthylation de l'ARN substrat. Cette structure présente un agencement symétrique des composants de la particule, tous retrouvés par paires.

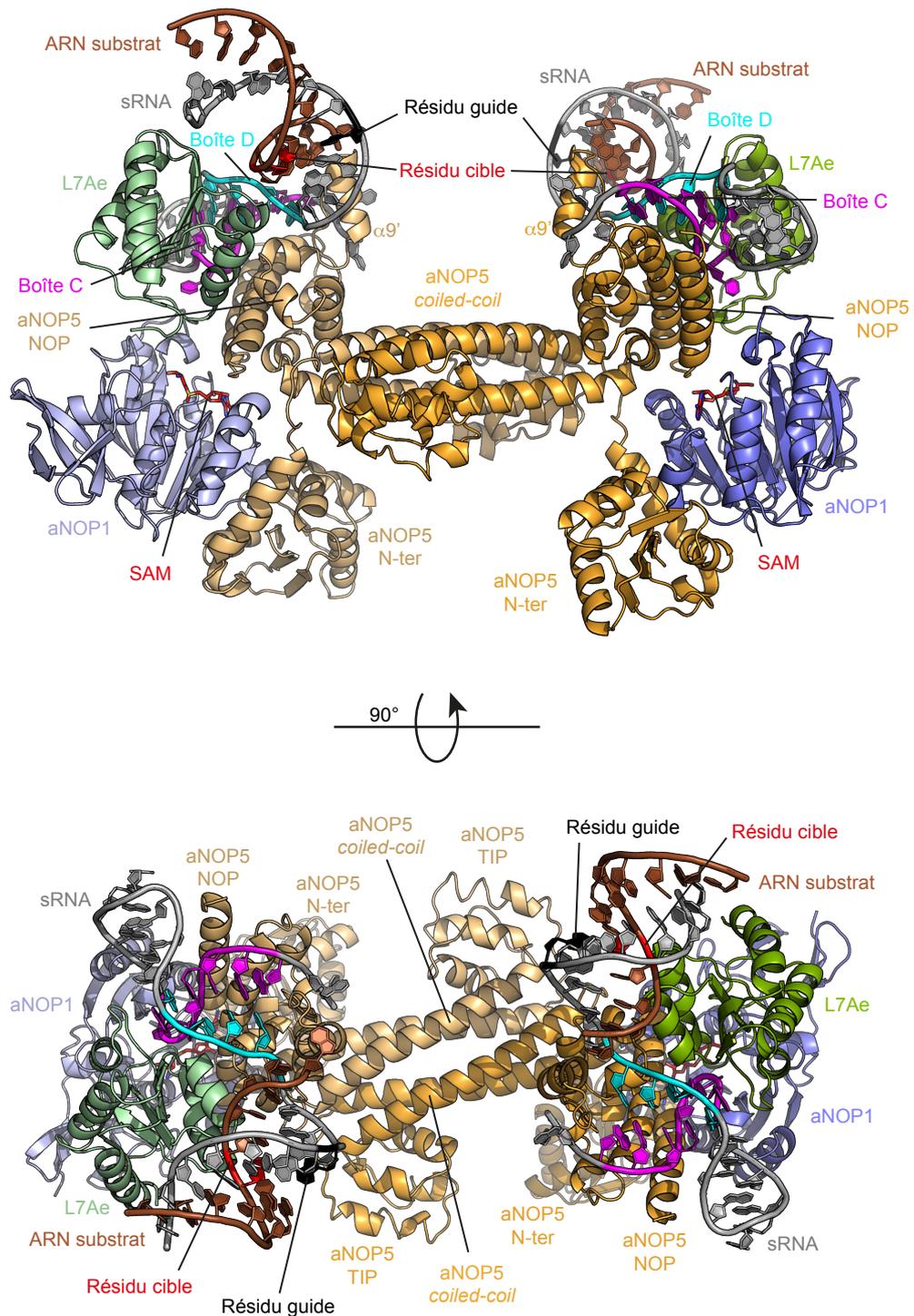


Figure 13. Structure tridimensionnelle d'une sRNP de l'archée *Pyrococcus furiosus* avec deux demi-sRNA C/D et avec le substrat en vue de côté (haut) ou du dessus (bas). Cette structure a été obtenue par cristallographie aux rayons X (Xue *et al.*, 2010 ; code PDB : 3NMU). Les résidus colorés en noir sur la structure du sRNA (ARN guide) correspondent aux 5^{èmes} résidus en amont des boîtes D, devant servir de guide à la 2'-O-méthylation de l'ARN substrat. Les résidus de l'ARN cible colorés en rouge correspondent aux résidus cibles devant accepter le transfert d'un groupement méthyle.

Plus récemment, une nouvelle structure de sRNP de *Sulfolobus solfataricus* a été déterminée avec, cette fois-ci, la présence de deux ARN substrats et un seul sRNA guide de 80 nucléotides (Figure 14) comprenant les deux jeux de boîtes C/D et C'/D' (Lin *et al.*, 2011). Des tests d'activité ont par ailleurs montré que cette particule est bien active et cible spécifiquement les sites attendus sur l'ARN cible. Dans cette structure de sRNP active, l'homodimère des deux protéines aNOP5 sert de plateforme au reste des composants de la particule, en adoptant une forme de berceau. La position du domaine NOP de aNOP5 et de la protéine L7Ae en interaction avec le *K-turn* de l'ARN guide est identique à ce qui est observé dans les précédentes structures (Figures 12-14). L'interaction entre le domaine N-terminal de la protéine aNOP5 et la protéine aNOP1 est également la même que pour les précédentes structures, mais l'orientation du domaine N-terminal de aNOP5 par rapport au domaine *coiled-coil* est différente. Ce changement de position du module catalytique aNOP5-Nter/aNOP1 situe le site catalytique de aNOP1 au contact du sillon mineur des deux hétéroduplex ARN formés par l'ARN guide et l'ARN substrat (Figure 14). Dans le cas de cette sRNP symétrique, les deux enzymes aNOP1 sont placées au contact de l'ARN dans une conformation active, ce qui suppose que les deux séquences guides du sRNA à boîtes C/D peuvent fonctionner simultanément. De plus, le fait qu'aucun contact ne soit apparent entre les deux duplex suggère que les deux séquences guides puissent fonctionner indépendamment. Dans cette particule, l'ARN est donc positionné dans un canal central formé par les protéines cœur, avec au-dessous les domaines *coiled-coil* en dimère des protéines aNOP5, aux extrémités les domaines NOP des protéines aNOP5 (bloquant l'ARN avec leurs hélices $\alpha 9'$) et les protéines L7Ae et sur les côtés les protéines aNOP1 (Figure 15A). Les hélices $\alpha 9'$ des protéines aNOP5 se retrouvent bloquées au centre de la structure ARN entre les deux brins du guide, au niveau de l'hélice II, où elles verrouillent d'un côté le duplex ARN-guide/ARN-substrat et de l'autre le couple de boîtes C/D. La spécificité du site de méthylation dépend à la fois de la capacité de aNOP1 à reconnaître l'ARN double brin, à la très faible mobilité du module catalytique une fois placé en position active et au positionnement correct du substrat (Lin *et al.*, 2011).

La comparaison de ces différentes structures de sRNP semble indiquer que ces particules possèdent plusieurs conformations menant à des sRNP actives ou non-actives (Figure 15B). La conformation semble être déterminée par la présence ou non du substrat lié à l'ARN guide et dirigée par des mouvements entre les différents domaines des protéines

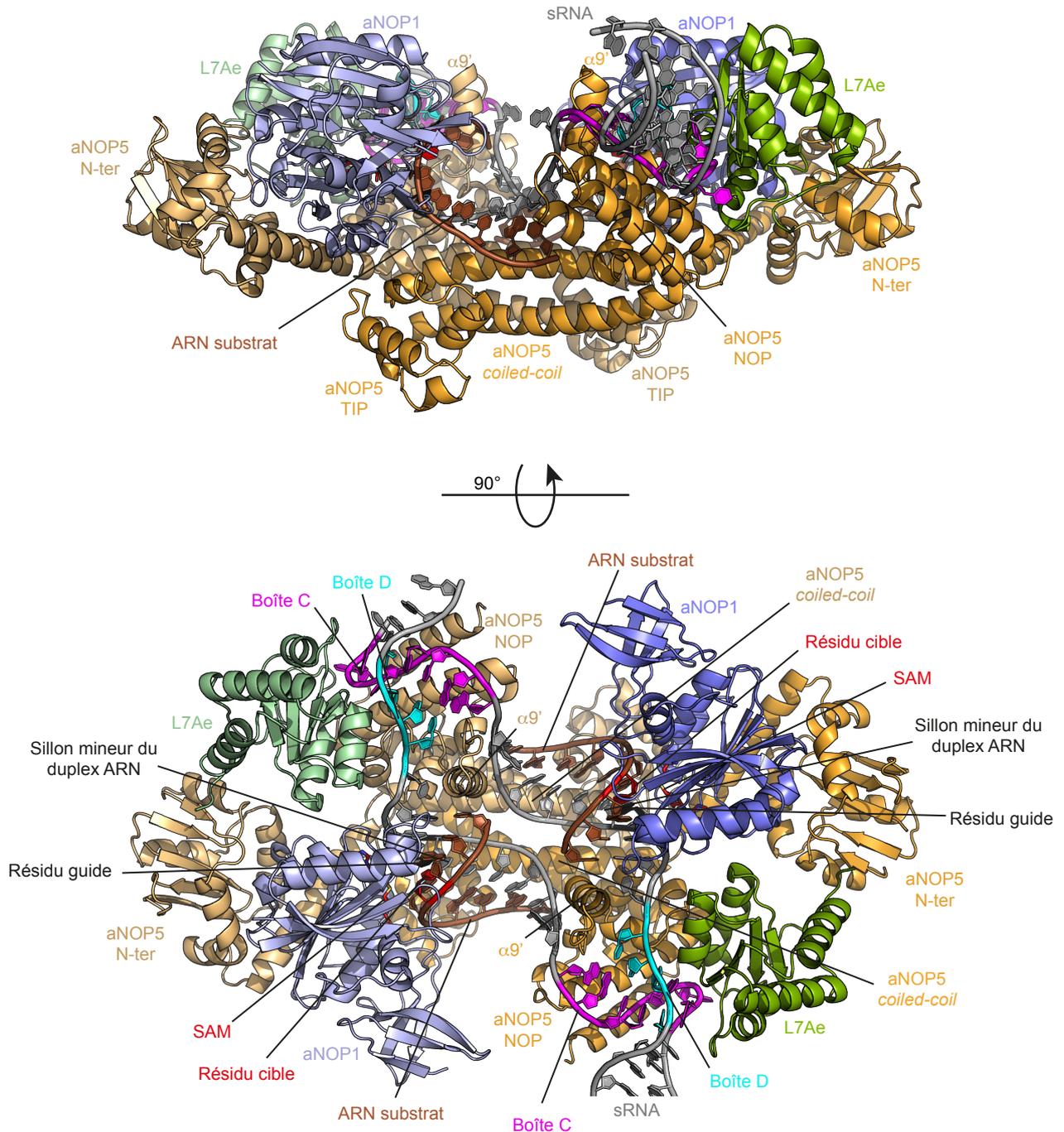


Figure 14. Structure tridimensionnelle d'une sRNP de l'archée *Sulfolobus solfataricus* avec un seul sRNA C/D comprenant deux jeux de boîtes et avec le substrat en vue de côté (haut) ou du dessus (bas). Cette structure a été obtenue par cristallographie aux rayons X (Lin *et al.*, 2011 ; code PDB : 3PLA). Les résidus colorés en noir sur la structure du sRNA (ARN guide) correspondent aux 5^{èmes} résidus en amont des boîtes D, devant servir de guide à la 2'-O-méthylation de l'ARN substrat. Les résidus de l'ARN cible colorés en rouge correspondent aux résidus cibles devant accepter le transfert d'un groupement méthyle.

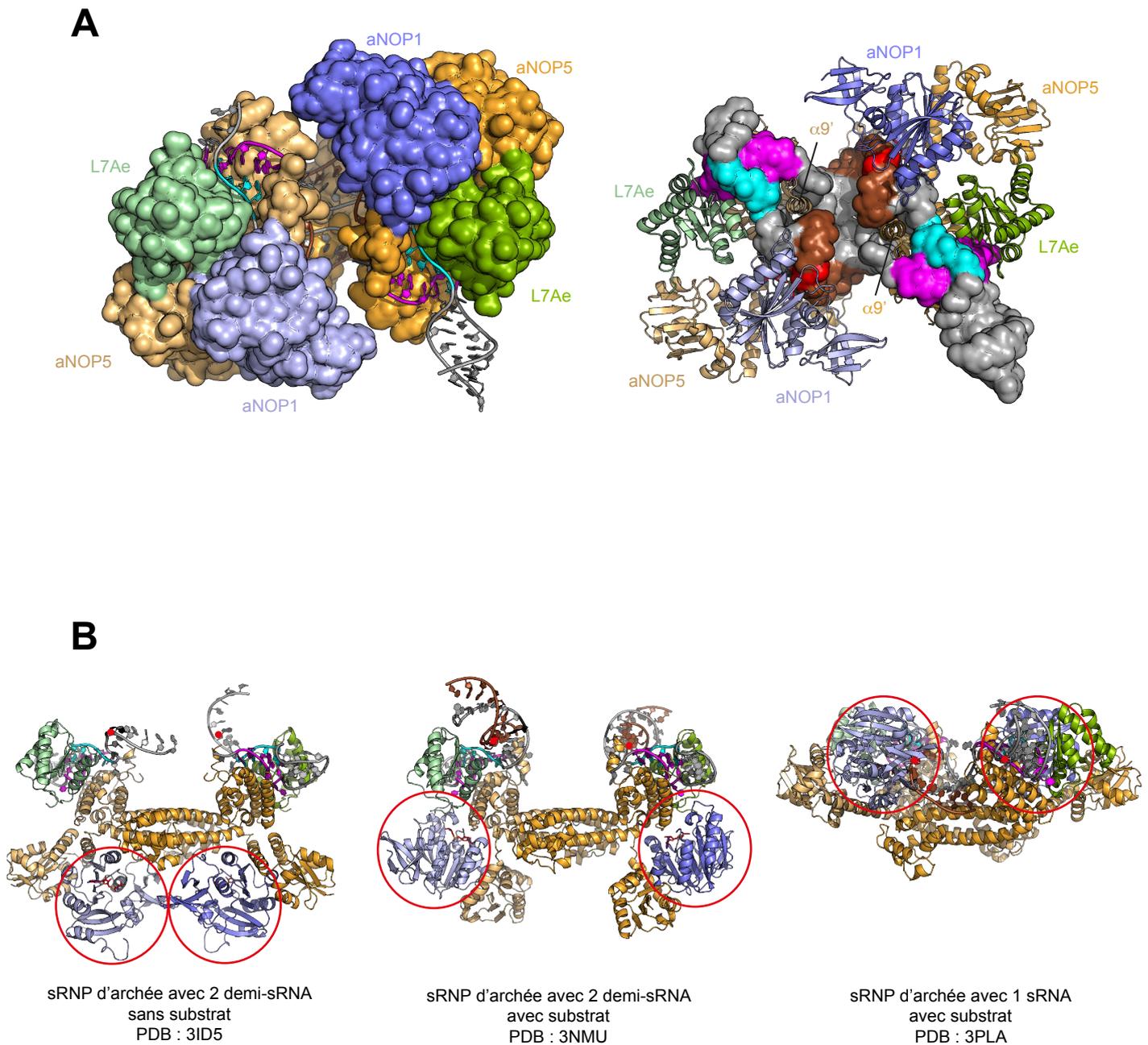


Figure 15. Hypothèse de la mono-sRNP à boîtes C/D. (A) Structure tridimensionnelle de la particule sRNP de *Sulfolobus solfataricus* obtenue par Lin *et al.*, 2011 (code PDB : 3PLA). Les protéines sont représentées en surface et les ARN en cartoon (à gauche) et inversement (à droite). (B) Vue de côté des trois structures tridimensionnelles de mono-sRNP d'archées. Les résidus à modifier (ou résidus guides dans le 1^{er} cas) sont symbolisés par des points rouges et les enzymes aNOP1 sont encerclées en rouge.

aNOP5, ce qui rapproche les protéines aNOP1 du site à méthyler (points et cercles rouges sur la Figure 15B).

1.2.4.2. Hypothèse de la di-sRNP

Une analyse par microscopie électronique (coloration négative) d'une sRNP de l'archée *Methanocaldococcus jannaschii* a pour la première fois mis en évidence une structure dimérique (di-sRNP) avec une organisation différente du modèle de la mono-sRNP (Bleichert *et al.*, 2009). Les résultats obtenus par l'équipe de S. Baserga montrent en effet des particules de taille supérieure à celle attendue en accord avec le modèle de la mono-sRNP, et catalytiquement actives. Si l'enveloppe obtenue permet de positionner les protéines cœur, il n'est en revanche pas possible d'y placer l'ARN sans ambiguïté. Dans cette architecture ce sont les protéines aNOP5 qui dirigent le positionnement des sRNA et ainsi maintiennent la conformation de la di-sRNP (Bleichert *et al.*, 2009), en pontant les deux parties de la di-sRNP par leurs domaines *coiled-coil* (Figure 16A). Dans cette conformation, les deux sRNA sont positionnés parallèlement l'un à l'autre et orientés vers des directions opposées. Des deux modèles possibles, symétrique ou asymétrique, seul le modèle asymétrique est cohérent avec une particule active puisque dans le modèle symétrique les enzymes aNOP1 sont trop éloignée du site de méthylation (Xue *et al.*, 2010). Selon le modèle asymétrique la catalyse est possible, mais seulement par deux enzymes sur les quatre que compte la particule, avec une sur chaque sRNA (Figure 16A). Une telle di-sRNP devrait donc pouvoir réaliser la 2'-O-méthylation de deux substrats simultanément, et à partir de deux sRNA guide différents.

Par la suite d'autres travaux ont montré un agencement en di-sRNP chez les trois espèces d'archées *Sulfolobus solfataricus*, *Pyrococcus abyssi* et *Archaeoglobus fulgidus* suggérant que ce type d'organisation est conservé chez l'ensemble des archées (Bower-Phipps *et al.*, 2012). Dans cette étude, les auteurs ont travaillé avec un seul sRNA (sR8) portant deux couples de boîtes C/D et C'/D' (et non pas deux demi-sRNA). A la différence des travaux de Lin *et al.*, 2011, le sRNA utilisé ici contient une boucle apicale (motif en *K-loop*) qui est un motif omniprésent chez les archées. Il ne s'agit donc plus de deux oligonucléotides mimant un sRNA. Une récente étude sur une sRNP de l'archée *Methanococcus jannaschii* a montré que la présence des boucles apicales sur les motifs *K-loop* des sRNA semblait être un déterminant de la conformation en di-sRNP (Yip *et al.*, 2016). Dans ce modèle, les deux sRNA sont positionnés de façon parallèle et interagissent entre eux par leurs boucles terminales. Cette interaction semble stabiliser l'architecture en di-sRNP.

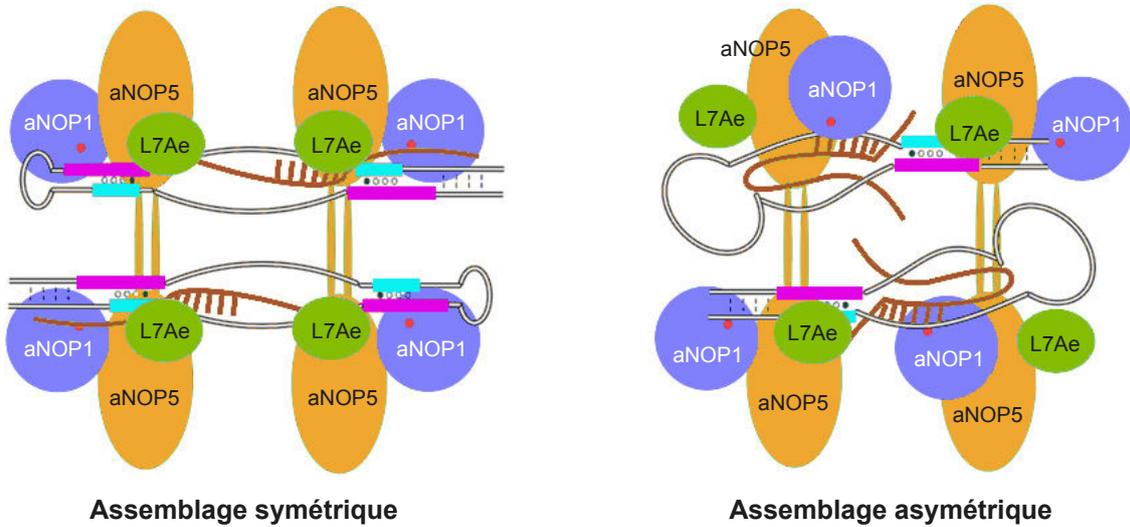
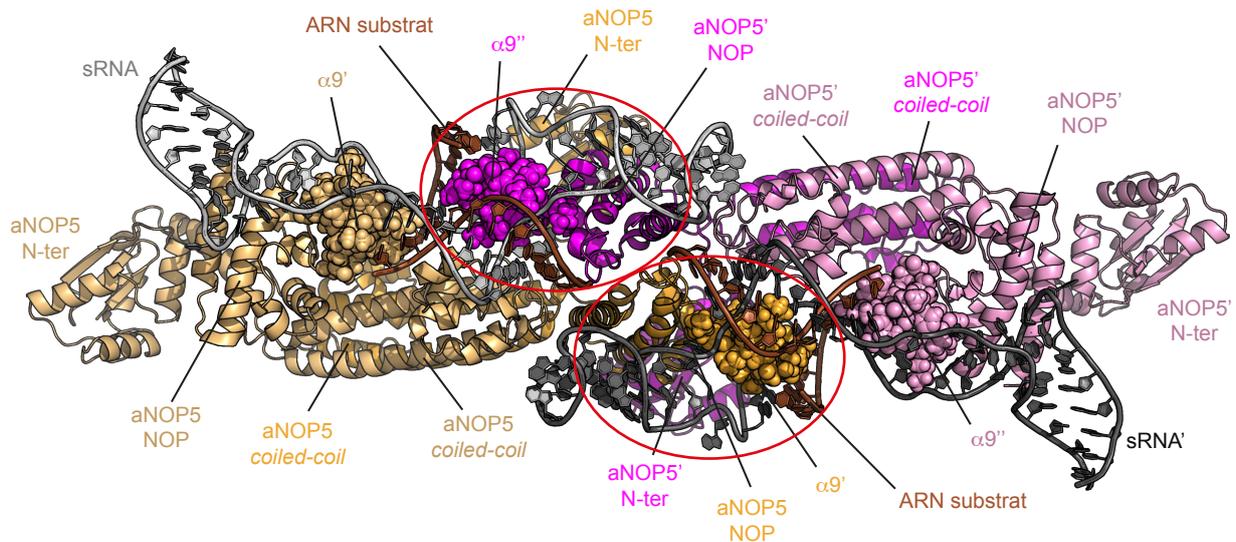
A**B**

Figure 16. Hypothèse de la di-sRNP à boîtes C/D. (A) Schémas théoriques des structures de di-sRNP symétrique (à gauche) et asymétrique (à droite) basés sur les observations en microscopie électronique (Bleichert *et al.*, 2009) et les structures à résolution atomique obtenues par cristallographie aux rayons X (d'après Xue *et al.*, 2010). (B) Structure tridimensionnelle de la di-sRNP de *Pyrococcus furiosus* (Lapinaite *et al.*, 2013; code PDB : 4BY9). Seules les protéines aNOP5 et les ARN guides et substrats ont été représentés. Les composants dont le nom est suivi d'un prime appartiennent à la seconde RNP. Les domaines NOP (contenant l'hélice $\alpha 9'$) sont représentés en sphère. Les domaines NOP des protéines aNOP5 d'une RNP interagissant avec l'ARN de l'autre RNP sont encadrés en rouge.

Plus récemment, une autre équipe a également montré l'existence de particules actives et inactives sous forme de di-sRNP comme entités fonctionnelles chez l'archée *Pyrococcus furiosus* en utilisant une approche combinée entre RMN et SANS (*Small-Angle Neutron Scattering*) (Lapinaite *et al.*, 2013). Dans cette étude, la particule a été reconstituée avec un sRNA comprenant deux jeux de boîtes C/D (deux motifs en *K-turn/K-loop*) et une boucle terminale. Les auteurs ont identifié deux conformations distinctes, en absence (apo-sRNP) et en présence (holo-sRNP) du substrat ARN, ce qui est cohérent avec l'hypothèse selon laquelle la présence du substrat modifie la conformation de la particule vers une position active. Contrairement à ce qui avait été proposé suite aux observations de microscopie électronique, dans la conformation active (holo-sRNP) les sRNA ne sont pas très éloignés l'un de l'autre et positionnés symétriquement l'un au-dessus de l'autre. Ici, les deux sRNA sont proches l'un de l'autre par leur boucle terminale, ce qui se rapproche des données obtenues par Yip *et al.*, 2016. De plus, les domaines *coiled-coil* des protéines aNOP5 des deux monomères ne sont pas parallèles mais plutôt alignés (Figures 16B et 17). Dans cette structure, les domaines NOP (encerclés en rouge sur la Figure 16B) d'une des deux protéines aNOP5 de chaque RNP (situés vers l'intérieur de la structure) interagissent avec une protéine aNOP1 et avec l'ARN de l'autre RNP, ainsi qu'avec le domaine NOP de l'autre. D'ailleurs, l'hélice $\alpha 9'$ comprise dans ces domaines NOP verrouille l'ARN de l'autre RNP sur un des deux motifs en *K-turn* (Figure 16B). Cette structure suggère que seulement deux protéines aNOP1 soient en position pour pouvoir catalyser la réaction sur le substrat, une de chaque RNP (Figure 17), malgré le fait que quatre substrats soient fixés sur les quatre guides. Il s'agit dans chaque RNP de la protéine aNOP1 proche du second motif C/D, *i.e* le motif C'/D' (côté intérieur du sRNA). Pour que l'autre site de méthylation, *i.e* le motif C/D, soit actif sur les deux RNP, il faut donc qu'un changement de conformation ait lieu. Des tests d'activité sur cette di-sRNP ont mis en évidence une méthylation à des sites spécifiques et très contrôlée (Lapinaite *et al.*, 2013). En effet, le substrat correspondant au motif C'/D' est toujours modifié en premier et le substrat correspondant au motif C/D ensuite (*in vitro*, avec cette particule et dans les conditions expérimentales utilisées par les auteurs). Dans le cas où seul le substrat du motif C/D est présent, il n'est pas modifié. Il semble donc qu'il faille une certaine quantité de produit issu de la méthylation au site C'/D' avant que la particule puisse modifier le substrat associé à la séquence guide du site C/D. Ceci est en accord avec des observations décrites dans la littérature où la perte d'une modification entraîne d'autres. La cellule semble donc avoir

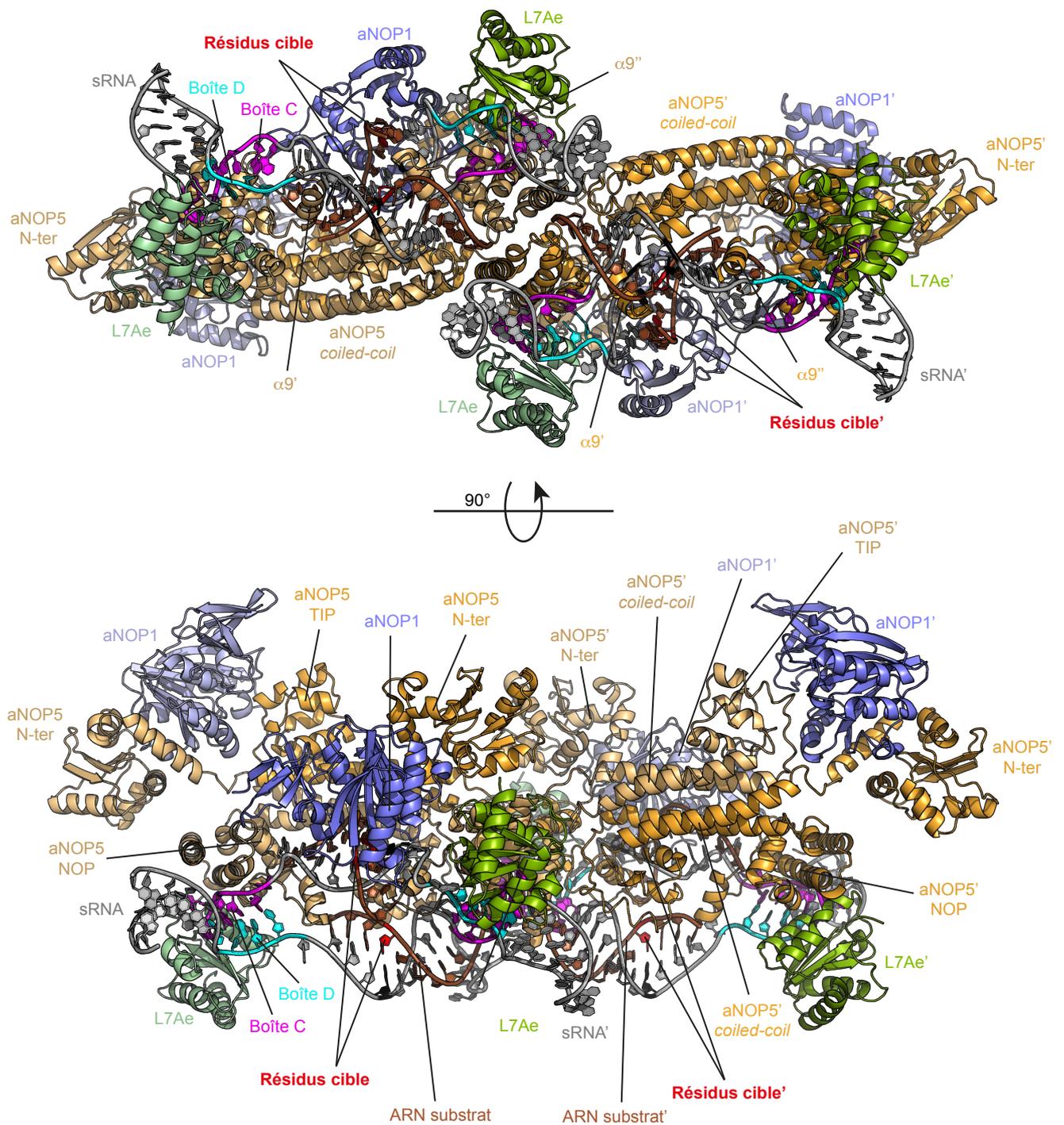


Figure 17. Structure tridimensionnelle d'une particule sRNP de l'archée *Pyrococcus furiosus* avec substrat adoptant une conformation de type di-sRNP (Lapinaite *et al.*, 2013; code PDB : 4BY9). Les composants dont le nom est suivi d'un prime appartiennent à la seconde RNP.

combiné différentes séquences guides sur un même sRNA pour pouvoir réguler la modification de certains sites par d'autres.

1.2.4.3. Cas particulier de la snoRNP U3

La snoRNP U3 est une snoRNP particulière, puisqu'elle ne catalyse pas la 2'-O-méthylation d'un ARN, mais participe à la machinerie intervenant dans le processome 90S ou pré-ribosome 90S, aux étapes précoces de clivage du pré-ARNr. Le snoRNA de cette particule est d'ailleurs particulier, puisqu'il possède des motifs C'/D et B/C, ainsi que des structures secondaires tige-boucle supplémentaires, notamment une hélice entre les deux jeux de boîtes (hélice 3) et une extrémité interne (boucle) en forme de T (hélice 4). Ajouté à cela, cette particule contient également une protéine supplémentaire qui lui est spécifique, la protéine Rrp9/U3-55K (Watkins *et al.*, 2000). La structure de cette particule a dernièrement été établie au sein du pré-ribosome 90S par cryo-EM chez le champignon *Chaetomium thermophilum* (Kornprobst *et al.*, 2016). Cette structure présente une organisation légèrement différente des autres déjà observées avec les protéines Nop1 éloignées de l'ARN (Figure 18), ce qui, dans ce cas, n'est pas étonnant puisque cette particule ne réalise pas de méthylation. Il s'agit ici d'une conformation de type mono-sRNP. Il est à noter que dans cette particule, le snoRNA est positionné très différemment de ce qui a pu être observé pour les autres mono-sRNP (Figure 14). En effet, dans cette structure l'ARN n'entretient pas de contacts avec les domaines *coiled-coil* des protéines Nop56/58 comme c'était le cas avec les protéines aNOP5 dans la mono-sRNP classique de *Sulfolobus solfataricus* (Lin *et al.*, 2011), et la surface d'interaction entre l'ARN et les protéines Nop56/58 est deux fois inférieure à ce qui est observable avec la structure de la particule de l'archée.

L'orientation de l'ARN par rapport aux domaines *coiled-coil* est d'ailleurs différente entre les deux structures. Cette différence est probablement due à la présence de l'hélice 3 du snoRNA U3, qui implique la formation d'un coude entre les deux jeux de boîtes B/C et C'/D. La structure de type mono-sRNP observée pour la snoRNP U3 est donc probablement induite par la structure particulière du snoRNA U3 et par la présence de la protéine Rrp9, absente chez les snoRNP à boîtes C/D conventionnelles.

La question sur l'existence de particules mono ou dimérique n'est donc toujours pas réellement résolue, même s'il paraît envisageable que la cible d'une particule sRNP/snoRNP

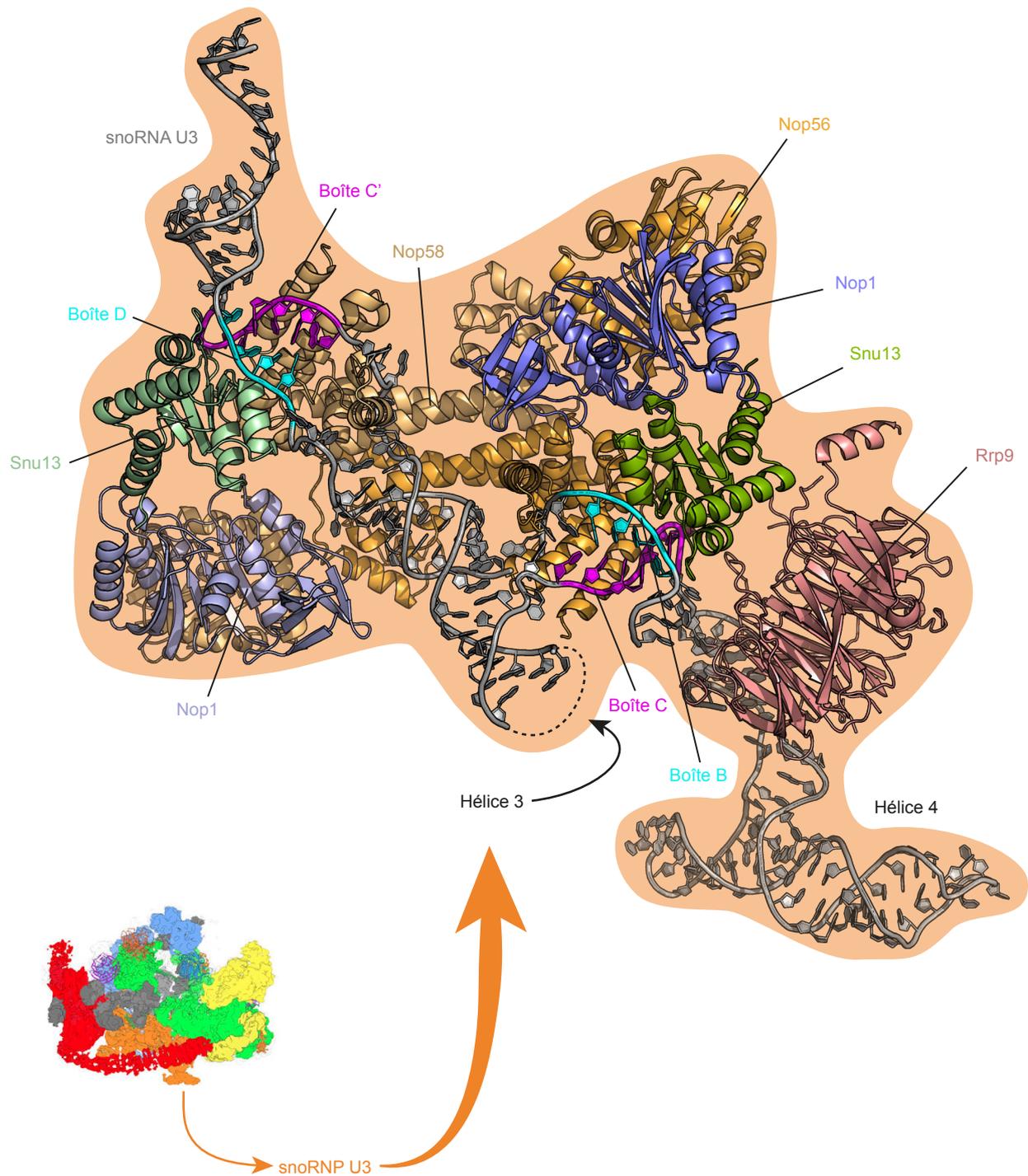


Figure 18. Structure tridimensionnelle de la particule U3 du champignon *Chaetomium thermophilum*. Cette particule U3 a été isolée de la structure du pré-ribosome 90S (en bas à gauche) obtenue par cryo-EM (Kornprobst *et al.*, 2016). Le positionnement des protéines et ARN sur la base des structures cristallographiques connues laisse place à quelques collisions mais permet toutefois d’appréhender l’organisation globale de la particule.

soit un déterminant majeur menant vers une organisation structurale de type mono- ou di-sRNP. De plus, il semble possible que la structure secondaire du sRNA en *K-turn* ou en *K-loop* soit le déterminant structural menant à la formation d'une mono- ou di-sRNP. Etant donné la prévalence des motifs *K-turn* et *K-loop* dans les règnes Eucaryote et Archée, les particules mono-sRNP seraient majoritairement présentes chez les organismes eucaryotes et les di-sRNP chez les archées.

2. Biogenèse des snoRNP à boîtes C/D eucaryotes : la machinerie d'assemblage

S'il est possible de reconstituer une particule sRNP à boîtes C/D d'archée active *in vitro* par auto-assemblage en mélangeant simplement en solution les différents composants d'une particule mature (Omer *et al.*, 2002) ou par co-expression de tous les composants (Peng *et al.*, 2014), il est impossible d'obtenir par ces méthodes une particule eucaryote active (Peng *et al.*, 2014). La biogenèse de ces particules eucaryotes requiert en effet l'intervention d'autres facteurs protéiques, appelés facteurs d'assemblage (Watkins *et al.*, 2004; McKeegan *et al.*, 2007; Boulon *et al.*, 2008; Rothe *et al.*, 2014a; Bizarro *et al.*, 2015). Ainsi, différents facteurs nécessaires à la biogenèse des particules snoRNP à boîtes C/D ont été identifiés. Ces facteurs ne sont pas directement impliqués dans le processus de maturation mais sont des candidats potentiels pour contrôler et organiser les événements d'assemblage des constituants du cœur des particules matures. Dans un premier temps, l'implication des hélicases Rvb1 et Rvb2 (RuvBL1 et RuvBL2 chez l'humain) a été montrée être cruciale dans la production des snoRNA à boîtes C/D (Newman *et al.*, 2000) et dans la localisation des particules snoRNP à boîtes C/D chez la levure (King *et al.*, 2001). Il semble donc que ces protéines soient impliquées dans les phases précoces de la biogenèse des snoRNP à boîtes C/D. Ces deux protéines font également partie d'un complexe à quatre composants chez la levure avec les protéines Tah1 et Pih1 (RPAP3 et PIH1D1 chez l'Homme), appelé complexe R2TP (pour Rvb1, Rvb2, Tah1p et Pih1p), considéré comme cofacteur de la chaperonne Hsp82/HSP90 (Zhao *et al.*, 2005). Par la suite, l'importance d'autres facteurs a été démontrée. Il s'agit ici des protéines Rsa1/NUFIP1 (McKeegan *et al.*, 2007; Boulon *et al.*, 2008), Hit1/TRIP3 (Rothe *et al.*, 2014; Bizarro *et al.*, 2014b) et Bcd1/BCD1 (Peng *et al.*, 2003), qui seront détaillées par la suite.

2.1. Le complexe Hsp82-R2TP

2.1.1. La chaperonne Hsp82/HSP90

Les protéines chaperonnes de type HSP90 (*Heat Shock Proteins* de 90 kDa) sont des protéines très abondantes dans les cellules. Elles représentent 1 à 2% des protéines cellulaires humaines et peuvent même monter jusqu'à 4 à 6% en période de stress (Prodromou, 2016). Ces protéines sont également très conservées dans les différents règnes du vivant (hormis les archées) (Taipale *et al.*, 2010). Dans les cellules humaines, deux formes variantes de la protéine HSP90 existent, la protéine HSP90 β qui est essentielle et exprimée de manière constitutive, alors que la protéine HSP90 α est exprimée en réponse à un stress thermique (Prodromou, 2016). Ce système est transposable chez la levure où deux isoformes se côtoient, la protéine Hsc82 qui est exprimée constitutivement tandis que l'expression de la protéine Hsp82 est induite par un état de stress thermique (Zhao *et al.*, 2005). Les rôles de ces chaperonnes sont de permettre le repliement correct d'autres protéines, l'assemblage de complexes multimériques, ainsi que d'éviter les phénomènes d'interaction aspécifique des protéines, voire d'agrégation, dans un environnement intracellulaire extrêmement concentré en protéine, en particulier dans le cytoplasme des cellules de mammifère où la concentration est de l'ordre de 300 mg/mL (Taipale *et al.*, 2010; Makhnevych et Houry, 2012). De plus, Hsp82 semble être impliquée dans des phénomènes de remodelage de la chromatine, et donc dans la régulation de l'expression des gènes, et pourrait ainsi affecter la physiologie cellulaire (Zhao *et al.*, 2005). Cette forte multiplicité fonctionnelle de Hsp82 est expliquée par le fait qu'en comptant ses protéines clientes (substrats) et ses cofacteurs, elle interagit physiquement ou génétiquement avec 10% du protéome (Zhao *et al.*, 2005; Makhnevych et Houry, 2012). Cette chaperonne serait donc un élément central dans un très large panel de voies cellulaires. En effet, il a été montré que la machinerie impliquant la chaperonne HSP90 fait intervenir une autre protéine chaperonne, HSP70, ainsi que de nombreux cofacteurs et se retrouve impliquée dans l'activation et la stabilisation de plus de 200 protéines dans les cellules de mammifères (Makhnevych et Houry, 2012).

La protéine Hsp82/HSP90 est composée de trois domaines, un domaine N-terminal comprenant le site de fixation de l'ATP, un domaine central et un domaine C-terminal responsable de la dimérisation de la protéine et de l'interaction avec les cofacteurs (Taipale *et al.*, 2010). Le dimère d'HSP90 adopte une conformation ouverte (en « V ») favorable à l'interaction avec les cofacteurs (Figure 19) (Li *et al.*, 2012). Lors de la fixation de l'ATP et de la protéine cliente, la conformation du dimère d'HSP90 change et un deuxième point de

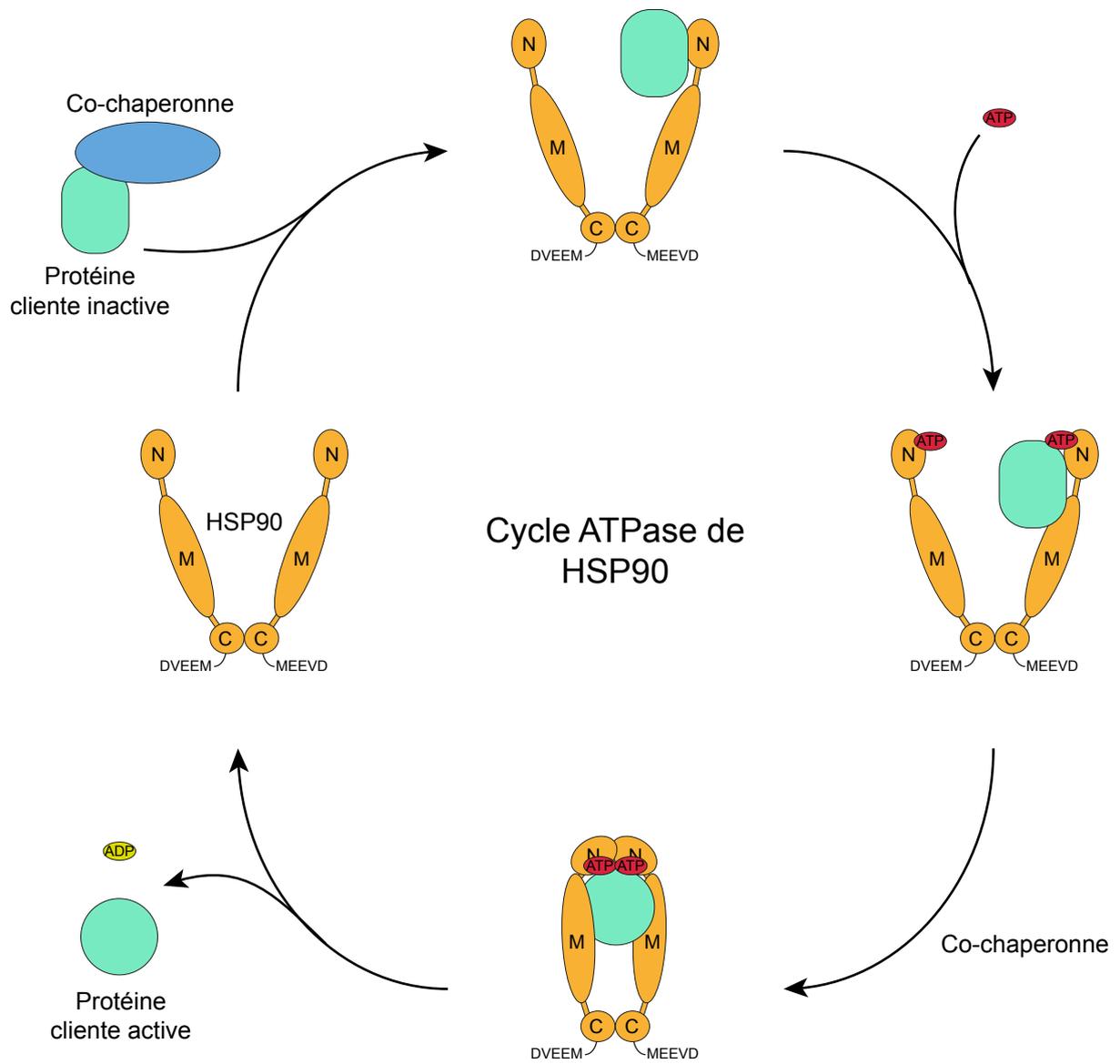


Figure 19. Le cycle de la chaperone HSP90.

dimérisation apparaît dans le domaine N-terminal, ce qui favorise l'hydrolyse de l'ATP, au sein d'une structure désormais en position « fermée ». Une fois l'hydrolyse de l'ATP effectuée, le dimère reprend sa conformation ouverte initiale et se dissocie de l'ADP et du Phosphate inorganique (Pi). Le domaine N-terminal est également connu pour avoir un rôle important dans la reconnaissance des protéines clientes (Taipale *et al.*, 2010), au même titre que le domaine central (Meyer *et al.*, 2003). Ainsi, la protéine HSP90 suit un cycle dépendant de la fixation de l'ATP et des différents cofacteurs (tels que les protéines Tah1-Pih1 du complexe R2TP, HOP/Sti1 ou Cdc37) ou des protéines clientes (Figure 19) (Richter *et al.*, 2003; Li *et al.*, 2012). Les cofacteurs de HSP90 interagissent généralement avec un peptide conservé situé dans le domaine C-terminal (MEEVD chez HSP90 ou IEEVD chez HSP70) par le biais d'un domaine TPR (*Tetratricopeptide Repeat*) (Scheufler *et al.*, 2000).

2.1.2. Le complexe R2TP

Comme mentionné précédemment, le complexe R2TP se compose des protéines Rvb1, Rvb2, Tah1 et Pih1. Ce complexe est considéré comme un cofacteur de la chaperonne Hsp82/HSP90 (Zhao *et al.*, 2005), mais est également requis pour la biogenèse des snoRNP à boîtes C/D ou dans d'autres processus biologiques tels que l'apoptose, des voies de signalisation (PIKK) ou l'assemblage de l'ARN polymérase II et potentiellement des autres ARN polymérases (Boulon *et al.*, 2010; Kakihara et Houry, 2012).

2.1.2.1. La protéine Tah1/RPAP3

La protéine Tah1 a initialement été identifiée en tant que partenaire des protéines Pih1 (Ito *et al.*, 2001), puis Hsp82 (Zhao *et al.*, 2005). Son interaction avec la chaperonne Hsp90 se fait par l'intermédiaire de son domaine TPR, d'où lui vient son nom (*TPR-containing protein Associated with HSP90*). L'interaction entre ces deux protéines par le domaine TPR de Tah1p a été d'abord supposée d'après la forte similitude entre ce domaine TPR et celui de la protéine Sti1, également cofacteur de la chaperonne Hsp82 (Scheufler *et al.*, 2000). Cependant, et contrairement à Sti1p, Tah1p interagit spécifiquement avec les chaperonnes Hsp82/Hsc82 (ou avec leurs orthologues humains HSP90 α/β), mais pas avec la chaperonne Ssa1 (HSP70 chez l'Homme). Cette spécificité vient de la forte implication de la méthionine du motif MEEVD de Hsp82 (ou également de Hsc82) dans l'interaction avec Tah1p (Millson *et al.*, 2008; Back *et al.*, 2013), alors que cette méthionine est remplacée par une valine chez Ssa1 (VEEVD).

La protéine Tah1 est une petite protéine, dont le domaine TPR atypique, qui représente 92 des 111 acides aminés de la protéine, est constitué de 2 motifs TPR et d'une hélice C-terminale qui permet la liaison à Hsp82 (Jiménez *et al.*, 2012; Back *et al.*, 2013). Ce domaine TPR de la protéine Tah1 de 5 hélices (Figure 20A) se distingue des autres domaines TPR classiques composés classiquement de 7 hélices (Back *et al.*, 2013; Morgan *et al.*, 2015). La protéine RPAP3 (humaine) est, elle, plus longue (665 résidus) et possède un domaine TPR supplémentaire, ainsi qu'un domaine C-terminal (Morgan *et al.*, 2015). Les résidus C-terminaux de la protéine Tah1, qui apparaissent comme étant déstructurés dans le cadre de la protéine libre, constituent une région suffisante à l'interaction avec la protéine Pih1, et apparaissent complètement structurés au sein du complexe Tah1p:Pih1p (Figure 20A) (Back *et al.*, 2013; Pal *et al.*, 2014). De plus, il semble que l'interaction de Tah1p avec Pih1p, ainsi que la présence de Hsp82, stabilisent Pih1p, qui seule est très fortement déstabilisée par sa région C-terminale (Zhao *et al.*, 2008; Jiménez *et al.*, 2012; Quinternet *et al.*, 2015). L'interaction entre Tah1p et Hsp82 n'est pas perturbée, que la protéine Pih1 soit présente ou non, ce qui signifie que la région d'interaction de Tah1p avec Pih1p est différente de celle de Tah1p avec Hsp82. La protéine Pih1 établit également des contacts avec Hsp82 dans le complexe Hsp82:Tah1p:Pih1p (Quinternet *et al.*, 2015). Cependant, la présence du complexe Tah1p:Pih1p inhibe l'activité ATPase de la chaperonne (Eckert *et al.*, 2010), alors que la présence de Tah1p seule sur Hsp82 provoque une légère hausse de cette activité ATPasique (Millson *et al.*, 2008; Eckert *et al.*, 2010). La présence de la protéine Pih1 avec Tah1p semble donc responsable de l'inhibition de l'activité ATPase de Hsp82. La présence du complexe co-chaperon Tah1p:Pih1p sur Hsp82 empêche l'arrivée d'un autre cofacteur à domaine TPR. Il est donc impossible d'obtenir des complexes co-chaperons mixtes sur un dimère de Hsp82 (Morgan *et al.*, 2015). Pour expliquer cela, l'une des hypothèses est que deux protéines Tah1 puissent former un homodimère par l'intermédiaire de leurs hélices $\alpha 5$ (Figure 20A) (Pal *et al.*, 2014; Morgan *et al.*, 2015). Ceci aurait pour résultat de former localement des domaines TPR à 6 hélices, imbriqués l'un dans l'autre. La formation de ce dimère de Tah1p pourrait ainsi former un complexe de stœchiométrie 2:2 avec Hsp82. Cette hypothèse est notamment basée sur le fait que la protéine humaine RPAP3 possède deux domaines TPR, dont chacun semble pouvoir contacter le peptide C-terminal de HSP90 (Pal *et al.*, 2014).

Il a également été suggéré que la protéine RPAP3 (humaine) était un partenaire de la protéine MONAD (protéine de type *WD40 repeat*), et pourrait dans le cadre de cette

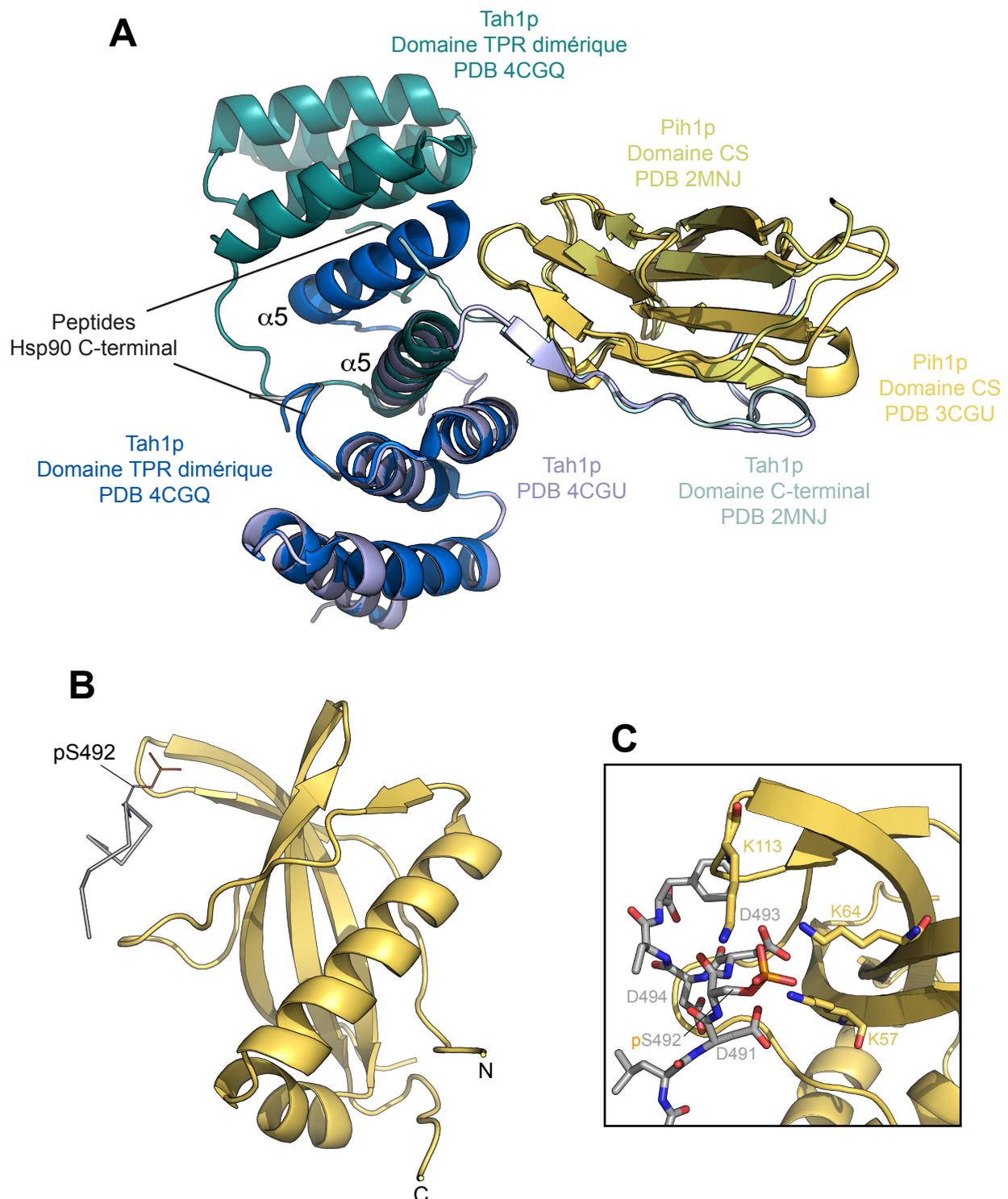


Figure 20. Structures tridimensionnelles des protéines Tah1 et Pih1. (A) Superposition des structures 3D du domaine CS de la protéine Pih1 avec la région C-terminale de Tah1p (Quinternet *et al.*, 2015; code PDB : 2MNJ) ou avec Tah1p entière et le peptide SRMEEVD de Hsp82 (Pal *et al.*, 2014; code PDB : 4CGU), ainsi qu'un dimère du domaine TPR de Tah1p (Morgan *et al.*, 2015; code PDB : 4CGQ). (B) Structure 3D du domaine PIH de la protéine Pih1 en complexe avec le peptide SELDpSDDEF de Tel2 phosphorylé sur la sérine (Pal *et al.*, 2014; code PDB : 4CSE). Le peptide est représenté en ruban gris et la chaîne latérale de la sérine phosphorylée en ligne. (C) Zoom sur le peptide de la structure présentée en (B). Le peptide, dont les résidus DpSDD sont annotés, est représenté en bâtonnets, ainsi que les trois lysines de Pih1p impliquées dans la stabilisation du groupement phosphate de la sérine.

interaction être impliquée dans la voie de l'apoptose en tant que modulateur (Itsuki *et al.*, 2008).

2.1.2.2. La protéine Pih1/PIH1D1

La protéine Pih1 (*Protein Interaction with Hsp90*, ou anciennement appelée Nop17) a initialement été identifiée comme un partenaire nucléolaire de la protéine Nop58 (Gonzales *et al.*, 2005), et plus précisément *via* la région C-terminale de Nop58p (contenant le domaine NOP et les répétitions KKE/D). Son implication dans la biogenèse des snoRNP à boîtes C/D a par la suite été démontrée chez l'Homme (McKeegan *et al.*, 2007). Il ne s'agit pas d'une protéine essentielle, mais sa déplétion provoque un phénotype thermosensible chez la levure, ainsi qu'un défaut de localisation des 4 protéines constitutives des snoRNP à boîtes C/D (Gonzales *et al.*, 2005). L'interaction entre Pih1p et Nop58p semble avoir un rôle dans la régulation de l'activité des snoRNP à boîtes C/D, puisque celle-ci entre en compétition avec l'interaction entre Nop58p et le snoRNA (Gonzales *et al.*, 2005). Cette régulation semble également dépendre des conditions de croissance et de la disponibilité en nutriment, qui régissent la localisation du complexe R2TP dans le cytoplasme lorsque la croissance est faible ou dans le noyau en cas de forte croissance (Kakihara *et al.*, 2014). Pih1p a également été identifiée comme étant un partenaire de la protéine Rrp43, qui est une sous-unité de l'exosome connue pour être impliquée dans des étapes de maturation des ARNr ou dans la dégradation des ARNm (Gonzales *et al.*, 2005).

Comme mentionné dans la partie précédente, Pih1p seule est une protéine instable et compte au nombre des cofacteurs de Hsp82 lorsqu'elle se trouve en complexe avec la protéine Tah1 (Zhao *et al.*, 2005), dont la présence semble jouer un rôle dans la stabilisation de Pih1p (Zhao *et al.*, 2008). Il a également été montré que la protéine Pih1 serait capable de former un complexe avec la chaperonne Hsp82 même en absence de Tah1p, ce qui montre que Hsp82 peut également stabiliser Pih1p (Zhao *et al.*, 2008). Ce rôle dans la stabilisation de la protéine Pih1 pourrait être le lien entre Hsp82 et la biogenèse des snoRNP. Ces interactions ont également été retrouvées entre les protéines humaines HSP90, RPAP3 et PIH1D1 (Boulon *et al.*, 2008).

La protéine Pih1 de levure contient deux régions intrinsèquement désordonnées, importantes dans le cadre d'interactions protéine-protéine (Paci *et al.*, 2012), l'une pour l'interaction avec les protéines Rvb1/Rvb2 et l'autre avec la protéine Tah1. La région C-terminale de la protéine Pih1 s'apparente à un domaine CS (*CHORD and Sgt1*) composé de

7 brins β en sandwich (Pal *et al.*, 2014), responsable de l'interaction avec la région C-terminale de la protéine Tah1 (en jaune sur la Figure 20A) (Quinternet *et al.*, 2015). Ce type de domaine est retrouvé chez d'autres co-chaperonnes de Hsp82, telle que Sgt1 (Zhang *et al.*, 2008). La région N-terminale de Pih1p/PIH1D1, appelée domaine PIH (Figure 20B), est requise pour l'interaction avec la protéine Tel2/TEL2 (Hořejší *et al.*, 2010; Pal *et al.*, 2014; Hořejší *et al.*, 2014), sur des sites spécifiquement phosphorylés par la kinase CK2 (*Casein Kinase 2*). Le motif DpSDD reconnu par le domaine PIH contient des résidus acides entourant la sérine phosphorylée (Figure 20C) (Pal *et al.*, 2014; Hořejší *et al.*, 2014). La protéine Tel2/TEL2 est une protéine impliquée dans l'activation et la stabilisation des PIKK ou de la protéine MONAD (Kakihara et Houry, 2012). Pih1p/PIH1D1 pourrait donc jouer le rôle d'intermédiaire entre la voie des PIKK (ou d'autres voies) et le complexe Hsp82:R2TP.

Plus récemment, une nouvelle fonction de la protéine Pih1 aurait été identifiée, en lien avec le protéasome. En effet, en l'absence de Tah1p, Pih1p forme un complexe avec la protéine Rpn8 (chez la levure), qui est une sous-unité du protéasome (Paci *et al.*, 2016). Seuls les 30 derniers acides aminés de la protéine Rpn8 sont nécessaires à l'interaction avec le domaine C-terminal de Pih1p (domaine CS). Par le biais de cette interaction avec la sous-unité Rpn8 du protéasome, la région C-terminale de Pih1p semble pouvoir agir comme un degron, c'est-à-dire une séquence protéique agissant sur la régulation des taux de dégradation des protéines, dans le cas présent en dirigeant Pih1p dans une voie de dégradation ubiquitine-indépendante (Paci *et al.*, 2016) par le protéasome, lorsqu'elle n'est pas en complexe avec la protéine Tah1.

2.1.2.3. Les protéines Rvb1/RuvBL1 et Rvb2/RuvBL2

Les protéines Rvb1 et Rvb2 sont deux protéines paralogues disposant d'une forte similarité de séquence entre elles. Toutes deux sont membres de la famille des protéines AAA+ (*ATPases Associated with diverse cellular Activities*), connues pour former généralement des structures en anneau hexamérique (Jha et Dutta, 2009; Nano et Houry, 2013). Les protéines appartenant à cette famille présentent des motifs conservés responsables de la liaison de nucléotides et de leur hydrolyse. Les complexes formés par ces deux protéines sont d'ailleurs connus pour avoir des activités ATPase et hélicase (Gribun *et al.*, 2008). Ces deux protéines interviennent dans de nombreuses voies cellulaires et sont d'ailleurs impliquées dans un grand nombre de complexes protéiques. Elles ont ainsi été identifiées par plusieurs équipes travaillant sur différents processus cellulaires et de différents organismes.

Pour cela, plusieurs noms différents leur ont été attribués au fil de leurs découvertes dans diverses conditions. Les protéines Rvb1/Rvb2 de levure sont donc également connues chez l'Homme sous les noms de TIP49/TIP48, RuvBL1/RuvBL2, Pontin/Reptin ou encore ECP54/ECP51. Dans la suite du manuscrit, ces protéines seront appelées Rvb1p/Rvb2p (ou Rvb1/2) pour les protéines de levure et RuvBL1/RuvBL2 (ou RuvBL1/2) pour les protéines humaines.

2.1.2.3.1. Organisation structurale

Les protéines Rvb1/2 sont organisées en trois domaines DI, DII et DIII (Matias *et al.*, 2006; Cheung *et al.*, 2010). Les domaines DI et DIII sont fortement liés l'un à l'autre et forment le domaine AAA+ (Figure 21A), qui rattache ces protéines à la famille du même nom. Ces deux domaines contiennent en effet des déterminants typiques des membres de la famille des AAA+. Le domaine DI contient ainsi les motifs Walker A et B, ainsi qu'un doigt à arginine et le sensor I. Le domaine DIII contient lui le sensor II. Ces motifs sont classiquement retrouvés chez les protéines capables de lier et hydrolyser l'ATP. Qui plus est, le domaine DI présente un repliement en sandwich $\alpha/\beta/\alpha$ similaire au *Rossmann fold*, connu pour être un repliement spécifique de la liaison avec des nucléotides. Les deux motifs Walker A et B au sein du *Rossmann fold* sont responsables de la fixation et de l'hydrolyse du nucléotide (Figures 21B et C) (Walker *et al.*, 1982). Les motifs sensor I et II permettent de détecter si le nucléotide fixé est un di- ou tri-phosphate (Snider *et al.*, 2008). Enfin, le doigt à arginine permet une hydrolyse efficace de l'ATP en établissant des contacts avec le groupement phosphate γ de l'ATP (Matias *et al.*, 2006; Snider *et al.*, 2008). Le domaine DII des protéines Rvb1/2 correspond à une insertion d'environ 170 acides aminés dans le domaine DI (Figure 21A), entre les motifs Walker A et B. Ce domaine d'insertion constitue une particularité des protéines Rvb1/2 par rapport aux autres membres de la famille AAA+.

Les domaines DI et DIII des protéines Rvb1/2 s'arrangent en un anneau hexamérique (Figure 22A), constitué d'une alternance des deux protéines (Lakomek *et al.*, 2015). L'absence des domaines DII n'influe d'ailleurs pas sur la formation de cet anneau hexamérique (Gorynia *et al.*, 2011). Il a également été montré que la protéine RuvBL1 était capable de former un homo-hexamère *in vitro* (Matias *et al.*, 2006), lequel ne présente toutefois pas d'activité ATPase. Dans la structure en anneau hexamérique, le doigt à arginine d'un monomère interagit avec l'ATP du monomère voisin précédent et joue un rôle dans son hydrolyse (Matias *et al.*, 2006; Snider *et al.*, 2008). Ce motif participe donc à la

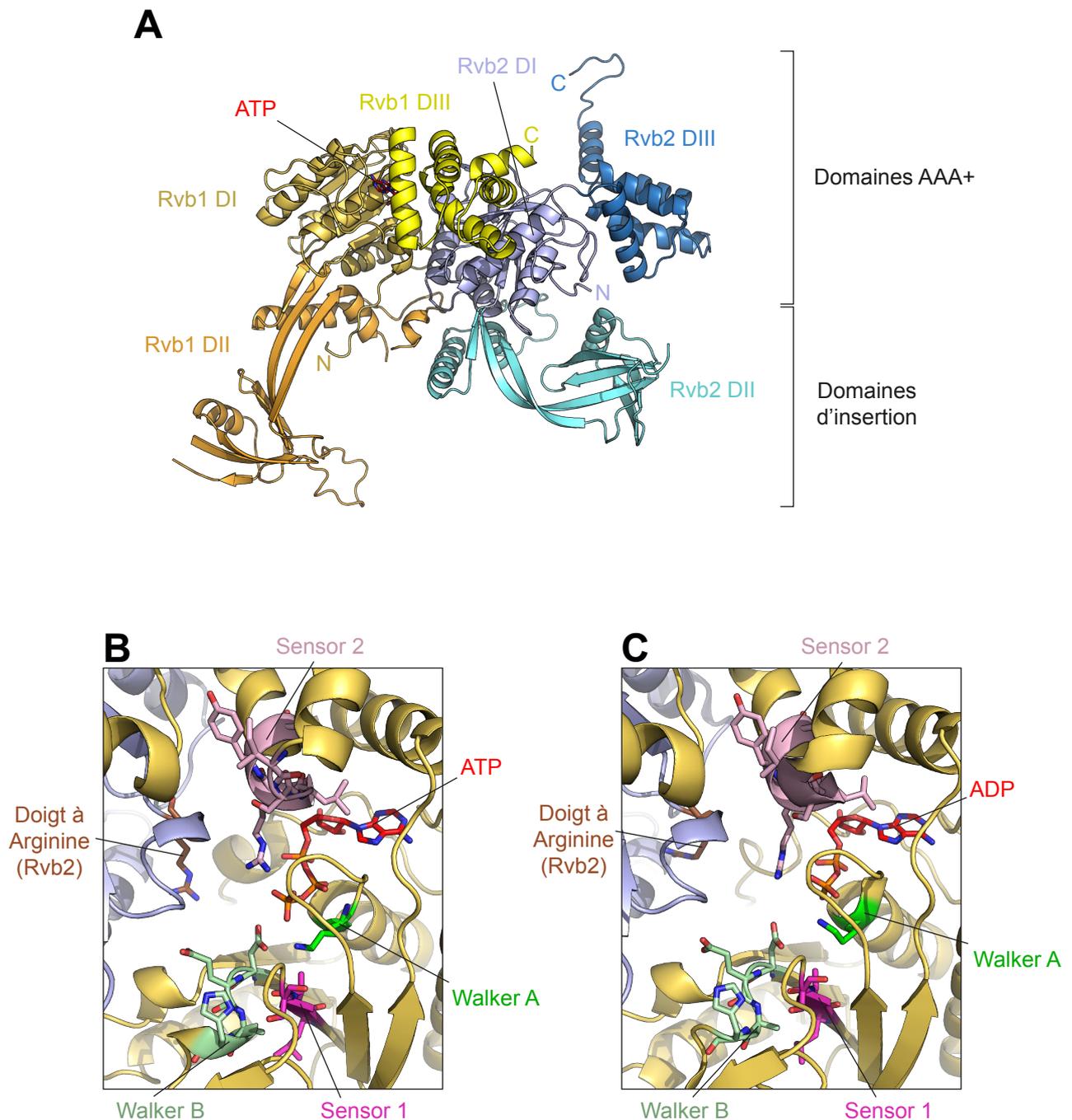


Figure 21. Structure tridimensionnelle de l'hétérodimère formé par les protéines Rvb1/2. (A) La structure du dimère chez *Chaetomium thermophilum* a été obtenue par diffraction des rayons X (Lakomek *et al.*, 2015; code PDB : 4WVY). La protéine Rvb1 est représentée en nuances de jaune et la protéine Rvb2 en nuances de bleu. Une molécule d'ATP, colorée en rouge, est fixée à la protéine Rvb1. (B) Zoom sur le site de fixation de l'ATP (code PDB : 4WVY) ou (C) de l'ADP (code PDB : 4WW4) sur la protéine Rvb1. Les motifs particuliers impliqués dans la fixation ou l'hydrolyse de l'ATP sont représentés en bâtonnets et colorés spécifiquement. Le motif en doigt à arginine provient de la protéine Rvb2 et se retrouve positionné différemment en fonction de la présence d'ATP ou d'ADP sur Rvb1p.

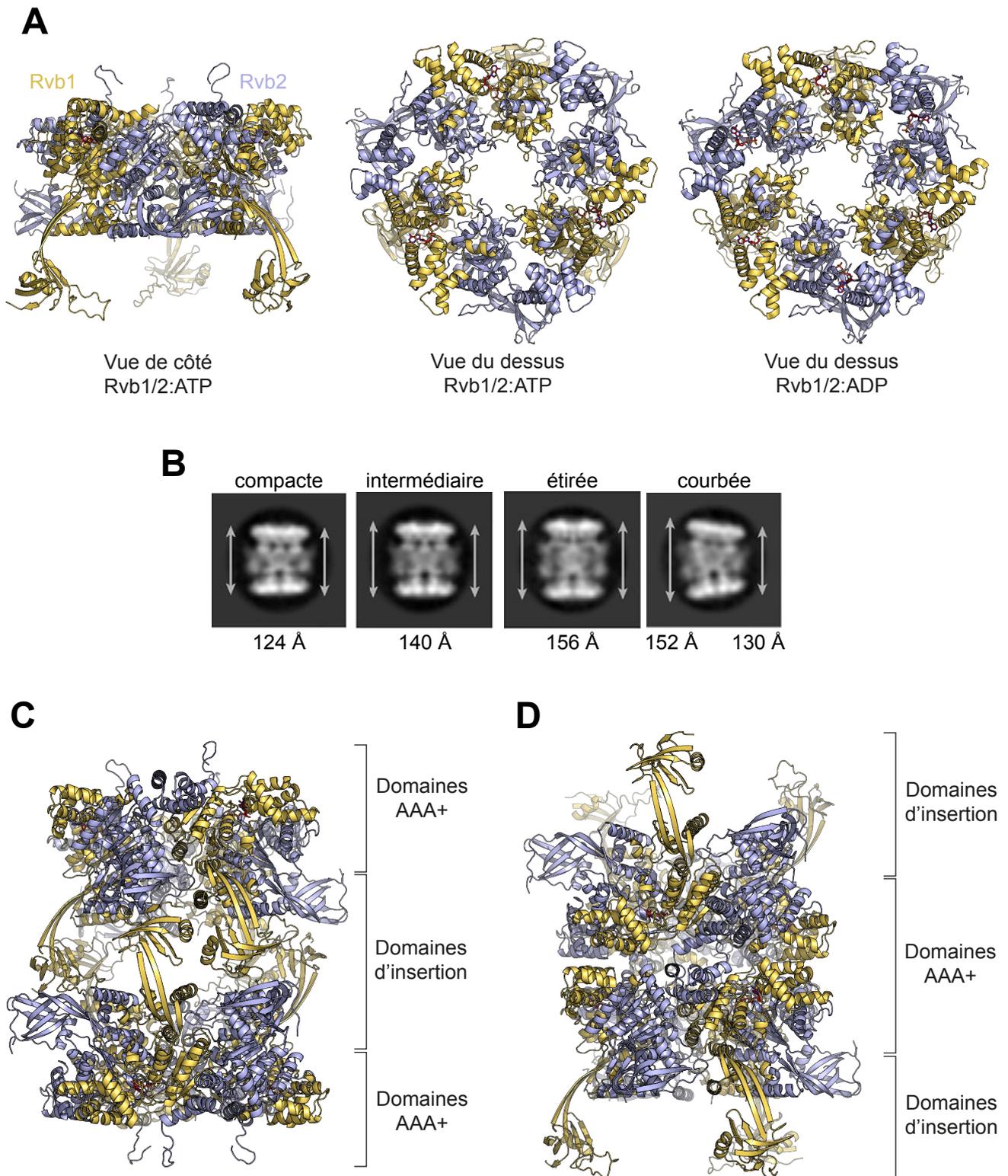


Figure 22. Conformation en hexamère et dodécamère des protéines Rvb1/2. (A) Structure en hexamère des protéines Rvb1/2 de *Chaetomium thermophilum* (Lakomek *et al.*, 2015; code PDB : 4WVY). (B) Différentes conformations adoptées par les structures en dodécamères de levure obtenues par cryo-EM (Ewens *et al.*, 2016). (C) Structure en dodécamère des protéines Rvb1/2 (Lakomek *et al.*, 2015; code PDB : 4WVY) avec interaction entre les hexamères par les domaines d'insertion ou (D) par les domaines AAA+ (anneau).

communication entre les sous-unités. Cette forme de coopération entre les monomères semble indiquer que la formation d'une structure hétéro-hexamérique est nécessaire à l'activité ATPasique de ces protéines (Matias *et al.*, 2006).

L'utilité du domaine DII des protéines RuvBL1/2 n'a toujours pas été totalement élucidée. Il a tout de même été montré que l'absence de ces domaines conduit à une augmentation substantielle de la consommation (hydrolyse) en ATP, ainsi que de l'activité hélicase de ces protéines (Gorynia *et al.*, 2011; Nano et Houry, 2013). Cela suggère que les domaines d'insertion DII pourraient agir comme des régulateurs de l'activité du complexe RuvBL1/2. De plus, il a été montré que ce domaine adoptait une conformation de type OB-fold (*Oligonucleotide/Oligosaccharide Binding*) (Petukhov *et al.*, 2012) et pouvait interagir avec des acides nucléiques (Matias *et al.*, 2006).

Des expériences de chromatographie d'exclusion de taille et de spectrométrie de masse ont par la suite mis en évidence la possibilité que l'état d'oligomérisation des protéines RuvBL1/2 pouvait être plus important que les formes en hexamères observées précédemment et que la stœchiométrie des protéines pouvait varier (Niewiarowski *et al.*, 2010). Des études de cryo-microscopie électronique (cryo-EM) sur le complexe Rvb1/2 ont également montré que son état d'oligomérisation correspondait à un dodécamère (Figure 22B) constitué d'un empilement de deux hexamères formés par les deux protéines Rvb1/2 en proportion stœchiométrique (Torreira *et al.*, 2008; López-Perrote *et al.*, 2012; Jeganathan *et al.*, 2015; Silva-Martin *et al.*, 2016; Ewens *et al.*, 2016). Cependant, si de nombreuses observations de complexes dodécamériques ont été réalisées, toutes ne sont pas identiques. En effet, il semble que de nombreux facteurs puissent influencer sur la structuration globale des dodécamères. Il est donc possible d'observer des formes compactes ou étirées dans le sens de la hauteur avec les deux anneaux hexamériques parfaitement superposés et parallèles (López-Perrote *et al.*, 2012; Jeganathan *et al.*, 2015) ou décalés l'un par rapport à l'autre (Silva-Martin *et al.*, 2016), ou encore courbés et donc non-parallèles (Ewens *et al.*, 2016) (Figure 22). Le point commun entre toutes ces différentes observations faites par cryo-EM est que les deux hexamères sont toujours en interaction par leurs domaines DII, et non par leurs anneaux.

La première reconstitution à l'échelle atomique d'un dodécamère entre les protéines RuvBL1/2 a été réalisée sur des protéines aux domaines DII tronqués (Gorynia *et al.*, 2011). Dans ce cas, les deux anneaux se superposent symétriquement. Cependant, d'autres structures à l'échelle atomique par diffraction des rayons X ont par la suite été obtenues sur les protéines entières issues d'un champignon thermophile (*Chaetomium thermophilum*) et montrent un

empilement non symétrique des deux anneaux (Lakomek *et al.*, 2015; Silva-Martin *et al.*, 2016). Dans ces deux cas, il existe un décalage d'environ 60° entre les deux anneaux, conduisant les protéines Rvb1 à se retrouver face aux protéines Rvb2 entre les deux anneaux (Figures 22C et D). Cependant, dans les cas mentionnés précédemment, les unités asymétriques observables dans les cristaux correspondent à des hétérodimères Rvb1/2. La reconstitution des hexamères a donc été réalisée par application d'une symétrie axiale d'un facteur 3. Les hexamères sont reliés entre eux par dyade cristallographique (symétrie), ce qui a permis aux auteurs de reformer des dodécamères. Cependant, à cause de l'empilement cristallin, il est impossible de discriminer avec certitude laquelle des deux conformations est possible entre des empilements d'hexamères par les anneaux (Figure 22D) ou par les domaines DII (Figure 22C). L'analyse par SAXS des protéines Rvb1/2 a par ailleurs montré qu'il existait probablement un mélange en solution entre les deux formes dodécamériques (Lakomek *et al.*, 2015; Silva-Martin *et al.*, 2016). Étonnamment, la forme dodécamérique par contact entre les anneaux n'a jamais été observée en cryo-EM.

Il semble donc que la forme privilégiée lors de l'association des protéines Rvb1/2 soit un dodécamère où l'interaction entre les deux anneaux hétéro-hexamériques est médiée par des contacts entre les domaines DII des protéines Rvb1 et ceux des protéines Rvb2. Les différentes formes observées par cryo-EM (compactes, étirées, décalées ou courbées) doivent donc être dues à des déplacements des domaines DII (Ewens *et al.*, 2016). Il apparaît que les formes prises par ces complexes sont conditionnées par la présence ou non de nucléotides (ATP ou ADP). La conformation prise par le complexe Rvb1/2 semble donc dépendre du cycle de l'hydrolyse de l'ATP (Ewens *et al.*, 2016).

2.1.2.3.2. Les protéines Rvb1/2 sont des composants de nombreux complexes nucléaires

Les protéines Rvb1/2 sont retrouvées au sein de nombreux complexes nucléaires et sont ainsi impliquées dans de nombreuses voies cellulaires.

Le complexe histone acétyl-transférase TIP60 humain (NuA4 chez la levure) est un complexe de remodelage de la chromatine impliqué dans la transcription, la réparation de l'ADN et l'apoptose (Ikura *et al.*, 2000; Nano et Houry, 2013). En cas de dommages de l'ADN, le variant d'histone H2AX se retrouve phosphorylé, et sert ainsi de marqueur d'ADN endommagé. La protéine Tip60 est un facteur de remodelage des nucléosomes contenant l'histone phospho-H2AX. Les protéines Tip60 et RuvBL1 sont importantes pour la régulation

à la baisse des phospho-H2AX dans les cellules après un phénomène de dommage de l'ADN (Jha *et al.*, 2008; Nano et Houry, 2013). En effet, l'absence d'une de ces deux protéines conduit à une augmentation de H2AX phosphorylée. La protéine RuvBL1 est également nécessaire à l'activité histone H4 acétyltransférase du complexe TIP60. Tout cela semble indiquer que la protéine RuvBL1 est requise pour la formation du complexe TIP60. Cette implication du complexe TIP60 dans la réponse aux dommages de l'ADN est reliée à la voie apoptotique, car en cas de défaut de ce complexe, les cellules perdent leur capacité d'apoptose (Ikura *et al.*, 2000). De plus, ce complexe est requis pour l'acétylation de la protéine p53. L'acétylation de cette protéine est nécessaire à sa liaison aux gènes pro-apoptotiques (Nano et Houry, 2013).

Chez la levure, les complexes de remodelage chromatinien INO80 et SWR1 contiennent également les protéines Rvb1/2. Le complexe INO80 est impliqué dans la régulation de la transcription, la réplication et la réparation des cassures de l'ADN double brin (Nano et Houry, 2013). Il permet notamment le retrait des nucléosomes entourant l'ADN double brin au niveau des cassures de cet ADN, afin qu'il puisse être réparé. Son recrutement dépend de la phosphorylation du variant d'histone H2AX. Les protéines Rvb1/2 sont requises pour l'intégrité structurale et fonctionnelle du complexe INO80 (Jónsson *et al.*, 2004). Le complexe SWR1 de levure (aussi connu sous le nom de SRCAP chez l'Homme) est également impliqué dans le remodelage de la chromatine, mais à l'inverse du complexe INO80 il permet le remplacement des dimères d'histones H2A/H2B par des dimères contenant les variants Htz1 (levure) et H2AZ (Homme). Le rôle des protéines Rvb1/2 dans ce complexe est encore inconnu, mais il est probable que ce soit également un rôle dans l'assemblage (Nano et Houry, 2013). Ces deux complexes ont été analysés par cryo-EM (Figure 23). Dans les deux cas, la forme des enveloppes permet de replacer facilement les protéines Rvb1/2, grâce à leur conformation en anneaux. Il se trouve que les structures obtenues montrent que les protéines Rvb1/2 s'agencent en anneau hexamérique au sein du complexe SWR1 (Figure 23B) (Nguyen *et al.*, 2013), alors qu'elles apparaissent sous forme de dodécamère dans le complexe INO80 (Figure 23A) (Tosi *et al.*, 2013). Ces résultats montrent qu'il ne doit pas exister une conformation unique adoptée par les protéines Rvb1/2 *in vivo*, mais que leurs caractéristiques permettent d'adapter leur degré d'oligomérisation à la fonction qui leur est demandée dans chaque complexe spécifiquement.

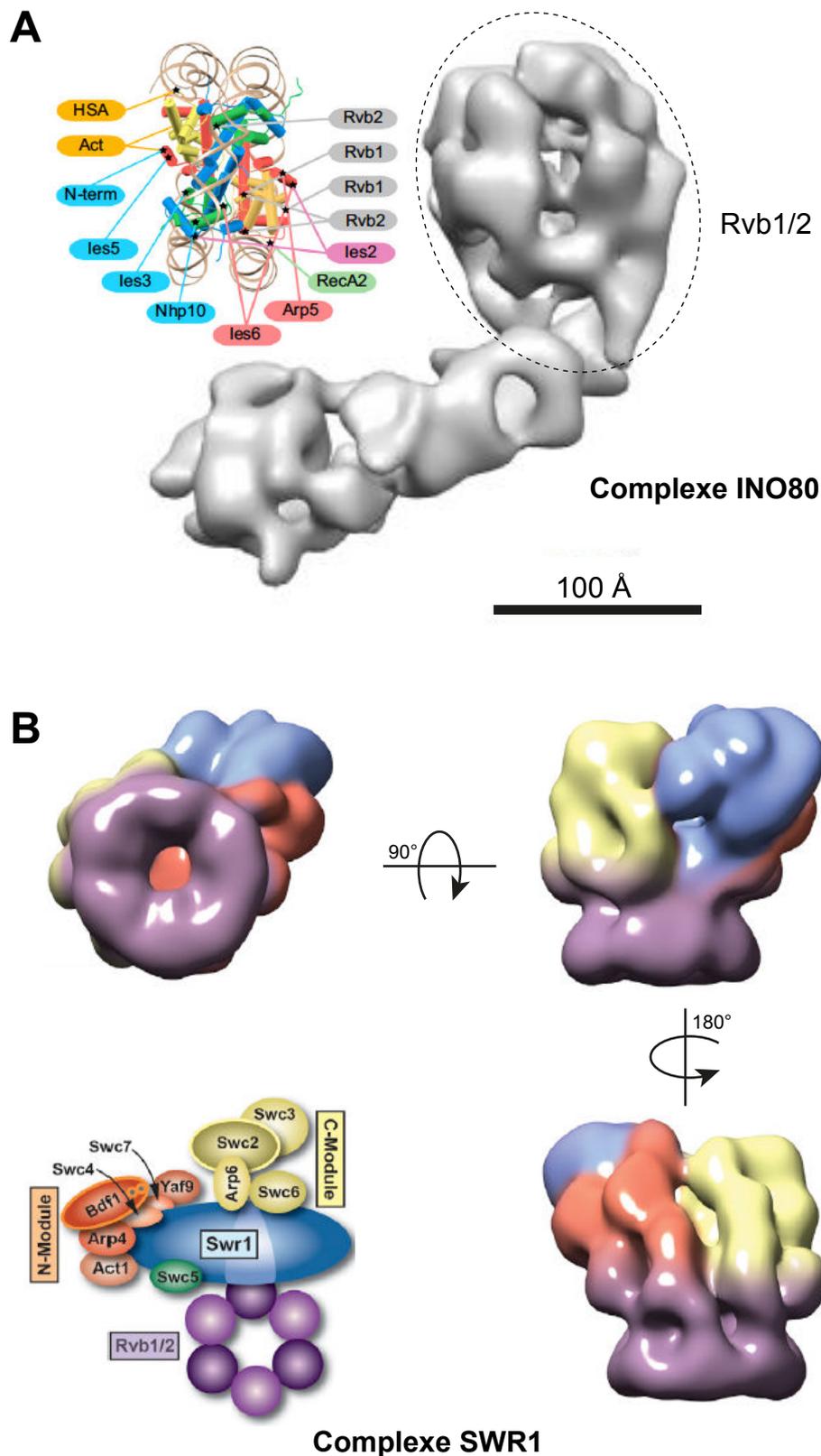


Figure 23. Structures obtenues par cryo-EM des complexes de remodelage chromatiniens INO80 (A) et SWR1 (B) chez la levure. (A) La région entourée en pointillés sur le complexe INO80 correspond aux protéines Rvb1/2. Le schéma à gauche représente les sites de fixation des différentes protéines du complexe sur le nucléosome (d'après Tosi *et al.*, 2013). **(B)** Le complexe SWR1 est représenté sous plusieurs angles. Les protéines sur le schéma en bas à gauche ont le même code couleur que sur la structure cryo-EM (d'après Nguyen *et al.*, 2013).

Chez l'Homme, les protéines RuvBL1/2 ont été identifiées comme étant des membres du complexe de la télomérase et ayant un rôle essentiel dans l'assemblage de ce complexe et dans le maintien de la stabilité de l'ARN associé (Venteicher *et al.*, 2008). Une étude de protéomique a récemment montré que les protéines Rvb1/2 étaient également associées avec le complexe CCT (*Chaperonin-Containing T*) qui aide au repliement des protéines, la particule 19S de régulation du protéasome et le complexe de l'ARN polymérase II (Lakshminarasimhan *et al.*, 2016). L'implication des protéines Rvb1/2 a été montrée dans plusieurs voies de signalisation telles que les voies de mTOR et des PIKK (Nano et Houry, 2013). Elles ont un rôle dans l'assemblage et la stabilisation des PIKK, en tant que membre du complexe R2TP associé au complexe *Prefoldin-like* (Kakihara et Houry, 2012), lequel intervient dans la biogenèse des particules ribonucléoprotéiques aux niveaux protéomiques et génomiques (Lakshminarasimhan *et al.*, 2016). En effet, l'absence des protéines Pih1 et Tah1 provoque une dérégulation des gènes impliqués dans la biogenèse des RNP, ainsi qu'un défaut d'association des protéines constitutives des RNP avec les protéines Rvb1/2. Le complexe R2TP est donc un complexe conservé impliqué dans un très grand nombre de voies cellulaires (Figure 24).

2.2. Le complexe Rsa1p:Hit1p/NUFIP1:ZNHIT3

La protéine Rsa1 (*Ribosome assembly 1*) a été découverte par l'équipe de P. Linder lors un crible génétique visant à rechercher des mutations conférant un phénotype synthétique létale lorsqu'elles étaient combinées à des mutations de l'hélicase Dbp6 (Kressler et al., 1999). L'appellation Rsa1p vient du fait que sa délétion provoque une diminution de la quantité de sous-unités ribosomiques 60S, et également une diminution du taux d'ARNr 25S. L'ensemble de ces effets est détectable à 30°C mais apparaît plus marqué à 37°C. Cette protéine a par la suite été identifiée par l'équipe de B. Charpentier lors d'un crible par le système triple hybride visant à identifier les protéines se fixant au motif B/C du snoRNA U3. De plus, des expériences au laboratoire ont montré que la région C-terminale (230-381) de cette protéine est capable de compléter le défaut de croissance d'une souche de levure Δ RSAl. Parallèlement, l'équipe d'E. Bertrand (IGM, Montpellier) entreprenait la recherche par criblage par le système double hybride de nouveaux partenaires protéiques pour la protéine humaine SNU13. Ces travaux ont mis en évidence un nouveau facteur d'intérêt, la protéine NUFIP1. Un travail collaboratif a débuté entre les deux équipes permettant de démontrer que les protéines Rsa1p (levure) et NUFIP1 (Homme) étaient impliquées dans les

(Potentielles) Fonctions du complexe R2TP

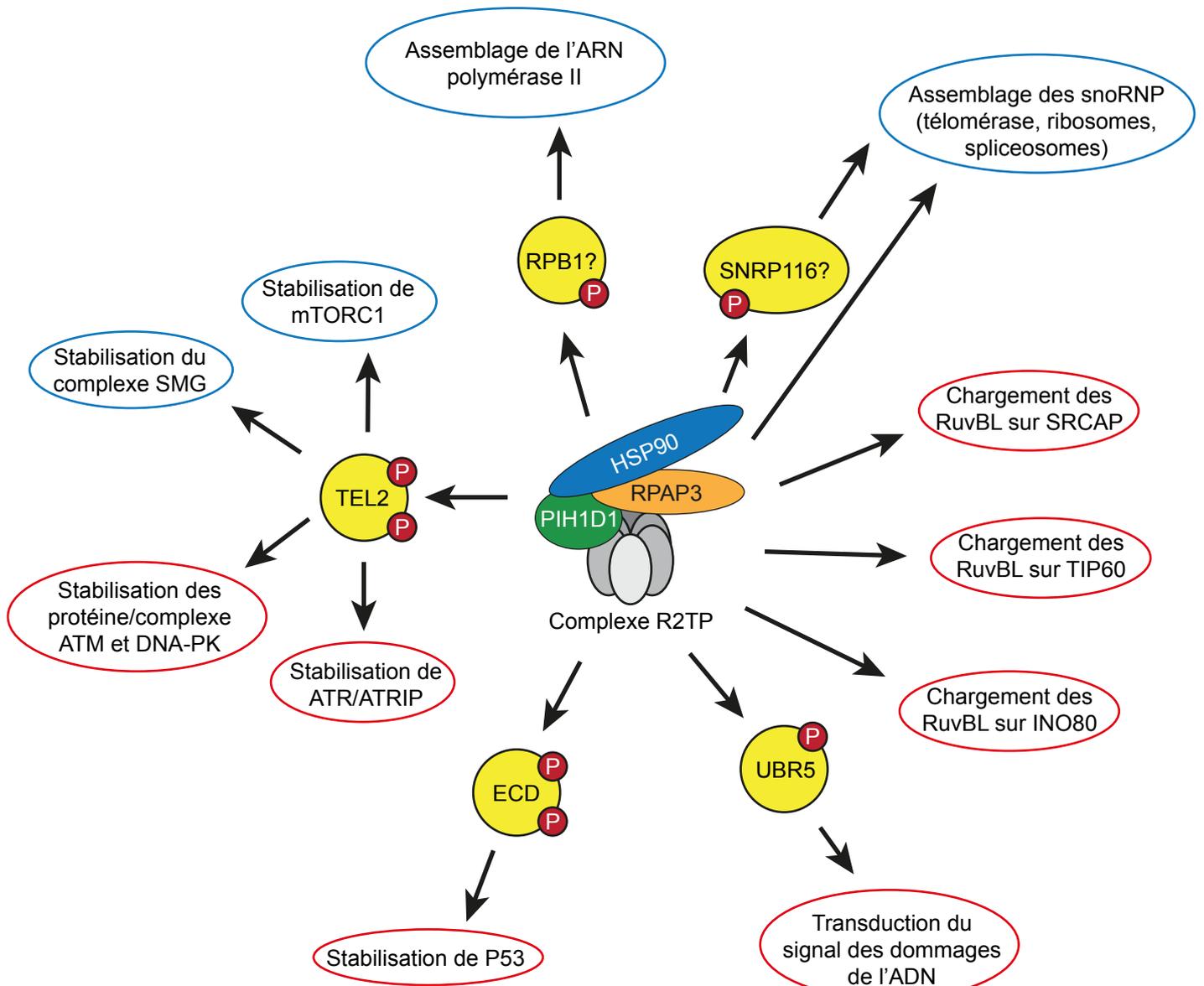


Figure 24. Schéma récapitulant les fonctions connues (cercles bleus) et potentielles (cercles rouges) du complexe R2TP (d'après von Morgen *et al.*, 2015).

voies de biogenèse des snoRNP à boîtes C/D. Celles-ci réalisent des connexions entre les différentes protéines cœur telles que les protéines Snu13/SNU13 ou Nop58/NOP58 des snoRNP à boîtes C/D lors de leur biogenèse (McKeegan *et al.*, 2007; Boulon *et al.*, 2008; Rothe *et al.*, 2014a; Rothe *et al.*, 2014b; Bizarro *et al.*, 2014). NUFIP1 permet également de réguler l'interaction entre les protéines RuvBL1/2 et les protéines cœur (McKeegan *et al.*, 2009). Outre son implication dans la biogenèse des snoRNP C/D, la protéine NUFIP1 apparaît être impliquée, avec les complexes HSP90:R2TP et SMN auxquels elle est retrouvée associée, dans la biogenèse des snoRNP à boîtes H/ACA, de la télomérase et de la particule UsnRNP U4 (Boulon *et al.*, 2008; Bizarro *et al.*, 2015).

La protéine NUFIP1 ne partage avec Rsa1p qu'une courte région de 37 acides aminés conservés, nommée PEP, responsable de l'interaction avec la protéine Snu13/SNU13 (Boulon *et al.*, 2008; Rothe *et al.*, 2014a). D'après la structure 3D, obtenue au laboratoire par radiocristallographie du complexe entre le domaine PEP de Rsa1p et la protéine Snu13, celui-ci interagit avec les hélices α_3 et α_5 de Snu13p (Figure 25B) (Bizarro *et al.*, 2014). Ce site d'interaction sur Snu13p est situé à l'opposé de la région de Snu13p qui établit des contacts avec l'ARN (Figure 25B), et doit donc permettre l'existence d'un hétérotrimère ARN:Snu13p:Rsa1p. La protéine humaine NUFIP1 possède un domaine en doigt à zinc en amont de son domaine PEP, qui n'est pas retrouvé chez la protéine Rsa1 de levure.

Un autre facteur impliqué dans la voie de biogenèse des snoRNP à boîtes C/D a ensuite été identifié au laboratoire, il s'agit de la protéine Hit1 (Rothe *et al.*, 2014b). Cette protéine était connue pour jouer un rôle important pour la croissance des levures à haute température, mais pas en condition normale de croissance (Kawakami *et al.*, 1992). Elle est un partenaire directe de la protéine Rsa1, qu'elle contribue à stabiliser (Rothe *et al.*, 2014b). La fonction décrite de cette protéine est donc d'assurer le maintien du taux d'expression de Rsa1p et de l'état fonctionnel de Rsa1p en évitant sa dégradation. Elle a donc un rôle indirect sur la biogenèse des RNP. Le site d'interaction de Hit1p sur Rsa1p se situe dans la région C-terminale de cette dernière, (317-325) (Figure 25C), tandis que le domaine PEP d'interaction avec Snu13p se situe plus en amont (230-266) (Figure 25B) (Rothe *et al.*, 2014b). L'interaction entre Rsa1p et Hit1p présente une très forte imbrication des 5 hélices de Hit1p avec 2 hélices de Rsa1p (Figure 25C). Il a de plus été montré que l'hétérotrimère formé par les trois protéines Snu13:Rsa1:Hit1 était en mesure d'interagir avec les snoRNA à boîtes C/D et avec la protéine Nop58. La protéine Hit1 possède dans sa région N-terminale un domaine en doigt à zinc de la famille zf-HIT (ZHD). Ce domaine (1-46) n'est pas contenu dans la

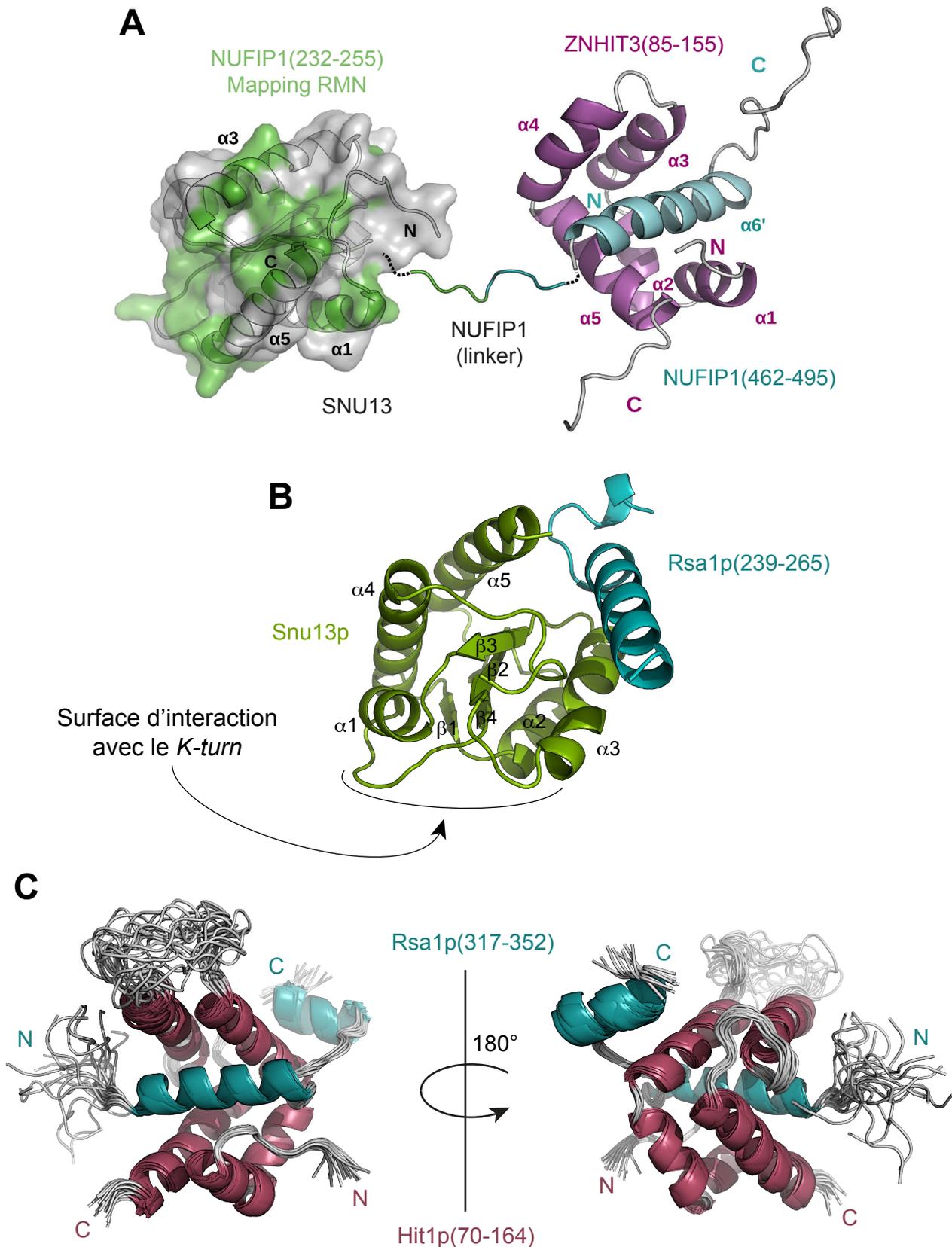


Figure 25. Structures tridimensionnelles des complexes entre les protéines Rsa1/NUFIP1 et Snu13/SNU13 ou Rsa1/NUFIP1 et Hit1/ZNHIT3. (A) Structure 3D du complexe NUFIP1(462-495):ZNHIT3(85-155) (à droite) et cartographie par RMN des résidus touchés par l'interaction de NUFIP1 sur SNU13 (d'après Quinternet *et al.*, 2016; code PDB : 5L85). (B) Structure 3D du complexe Rsa1p(239-265):Snu13p (Bizarro *et al.*, 2014; code PDB : 4NUT). (C) Structure 3D du complexe Rsa1p(317-352):Hit1p(70-164) (d'après Rothé, Saliou *et al.*, 2014; code PDB : 2MJF).

région nécessaire à l'interaction avec Rsa1p (70-164), qui est la région fonctionnelle. Ce ZHD ne semble donc pas utile à la fonction de Hit1p, mais représente toutefois la seule région de la séquence proche de celle de son orthologue humaine, la protéine ZNHIT3 (ou TRIP3), ce qui a d'ailleurs permis de l'identifier (Quinternet *et al.*, 2016). Comme je le décrirai dans la partie Résultats, malgré cette grande différence au niveau de la séquence, les protéines Hit1 et ZNHIT3 possèdent une très grande homologie de structure, et partagent le même motif d'interaction avec leurs partenaires Rsa1 et NUFIP1 (Figure 25A) (Quinternet *et al.*, 2016), et donc probablement la même fonction.

2.3. La protéine Bcd1/BCD1

2.3.1. *Connaissances au début de ma thèse*

La protéine Bcd1 ou *Box C/D snoRNA 1* (BCD1 ou ZNHIT6 chez l'Homme) a été identifiée par crible génétique haut débit de mutants de levures associés à des défauts de biogenèse d'ARN non codants. Sa présence s'est révélée essentielle à la viabilité cellulaire chez la levure et à l'accumulation spécifique des snoRNA à boîtes C/D dans leur ensemble (Peng *et al.*, 2003; Hiley, 2005). La protéine Bcd1 est donc impliquée dans la biogenèse des snoRNA et/ou des snoRNP à boîtes C/D. Sa fonction précise est toutefois encore inconnue aujourd'hui. Par la suite, l'équipe de J. Watkins, qui a travaillé sur la protéine BCD1 humaine, a pu identifier des interactions entre BCD1 et la protéine RuvBL2 en présence d'ADP (McKeegan *et al.*, 2007) et avec le complexe RuvBL1/2 en présence d'ATP (McKeegan *et al.*, 2009). D'autres interactions avec la protéine NUFIP1 et avec le facteur TAF9 (facteur d'initiation de la transcription TFIID, sous-unité 9) lui ont également été attribuées, malgré le faible signal des expériences mises en œuvre. Ces données n'ont jamais pu être validées ultérieurement (McKeegan *et al.*, 2007). Plus tard, une approche de protéomique fonctionnelle SILAC (*Stable Isotope Labeling with Amino acids in Cell culture*) a montré que la protéine BCD1 était en association avec les protéines RuvBL1/2, NUFIP1 et ZNHIT3 (Bizarro *et al.*, 2014). Une expérience de TAP-Tag menée sur la protéine Bcd1 au laboratoire et analysée par spectrométrie de masse a permis de mettre en évidence l'existence de ce potentiel complexe transitoire (Rothé, 2012). En effet, les résultats présentent plusieurs peptides associés à Bcd1p provenant des protéines Nop58, Rvb1 et Rvb2. Il serait donc envisageable qu'un complexe transitoire comprenant les protéines Nop58, Rvb1, Rvb2 et Bcd1 puisse exister au cours du processus de biogenèse des snoRNP à boîtes C/D.

Comme Hit1p, Bcd1p contient un ZHD dans sa région N-terminale. Ce domaine est retrouvé chez la protéine humaine BCD1, au centre de la séquence. La protéine humaine est plus longue (470 acides aminés pour 366 chez la levure). Sa région N-terminale (1-215) n'a aucune homologie avec la séquence de levure.

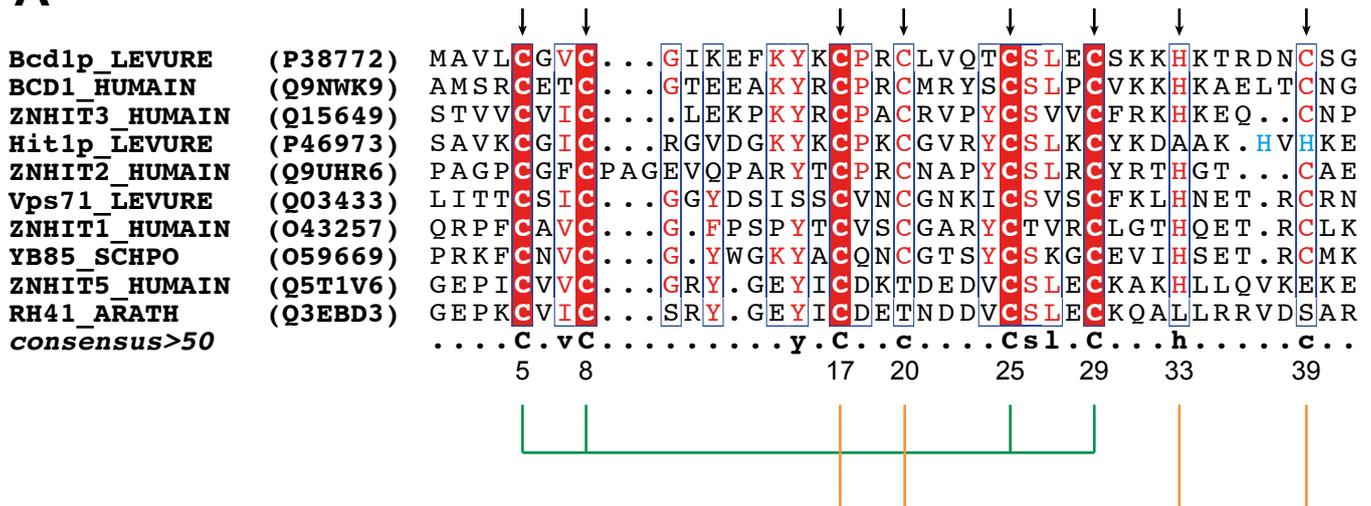
2.3.2. Les domaines en doigts à zinc de la famille zf-HIT (ZHD)

La famille zf-HIT a été initialement décrite suite à la découverte des doigts à zinc de la protéine Hit1, d'où lui vient son nom (Kawakami *et al.*, 1992). Cette famille est elle-même membre de la superfamille TRASH (*T*rafficking, *R*esistance and *S*ensing of *H*eavy *M*etals), laquelle est caractérisée généralement par un repliement en « *treble clef* » (Ettema *et al.*, 2003). Ce repliement particulier se compose de 2 ou 4 brins β , suivis par une hélice α C-terminale. Il existe également chez les représentants de cette superfamille un motif fortement conservé de 3 ou 4 cystéines, allant jusqu'à 5 ou 7 pour les membres de la famille zf-HIT (Figure 26A). Elle est très majoritairement représentée dans le règne eucaryote (99%) et très peu chez les archées. Aucun membre n'est retrouvé chez les procaryotes. Elle compte 6 membres chez l'Homme nommés ZNHIT1 à 6 (Figure 26B) (Verheggen *et al.*, 2015). La levure compte, elle, trois représentants de cette famille, les protéines Bcd1, Hit1 et Vps71 (Figure 26A). Deux protéines sur les 6 humaines sont impliquées dans la voie de biogenèse des snoRNP à boîtes C/D (ZNHIT3 et 6). Les protéines ZNHIT1 et 4 sont des constituants des complexes de remodelage de la chromatine SRCAP et INO80 respectivement. Les deux dernières restent encore mal caractérisées.

2.3.3. Lien entre les ZHD et les protéines RuvBL1/2

Il est intéressant de noter que les protéines contenant des ZHD sont pour beaucoup retrouvées au sein de complexes nucléaires ou de voies de biogenèse auxquels appartiennent également les protéines RuvBL1/2 (Figure 26B) (Verheggen *et al.*, 2015). En effet, les protéines ZNHIT3 et 6 appartiennent à la voie de biogenèse des snoRNP à boîtes C/D, au même titre que le complexe R2TP contenant les protéines RuvBL1/2. De plus, une interaction directe a été montrée entre la protéine ZNHIT6 (BCD1) et la protéine RuvBL2 (McKeegan *et al.*, 2007). Comme mentionné précédemment, les protéines ZNHIT1 et 4 sont respectivement membres des complexes SRCAP et INO80. Or ces complexes contiennent également les protéines RuvBL1/2. L'orthologue chez la levure de ZNHIT4, la protéine Ies2, établit des contacts directs avec les protéines RuvBL1/2 (Tosi *et al.*, 2013). Toutes ces données

A



B

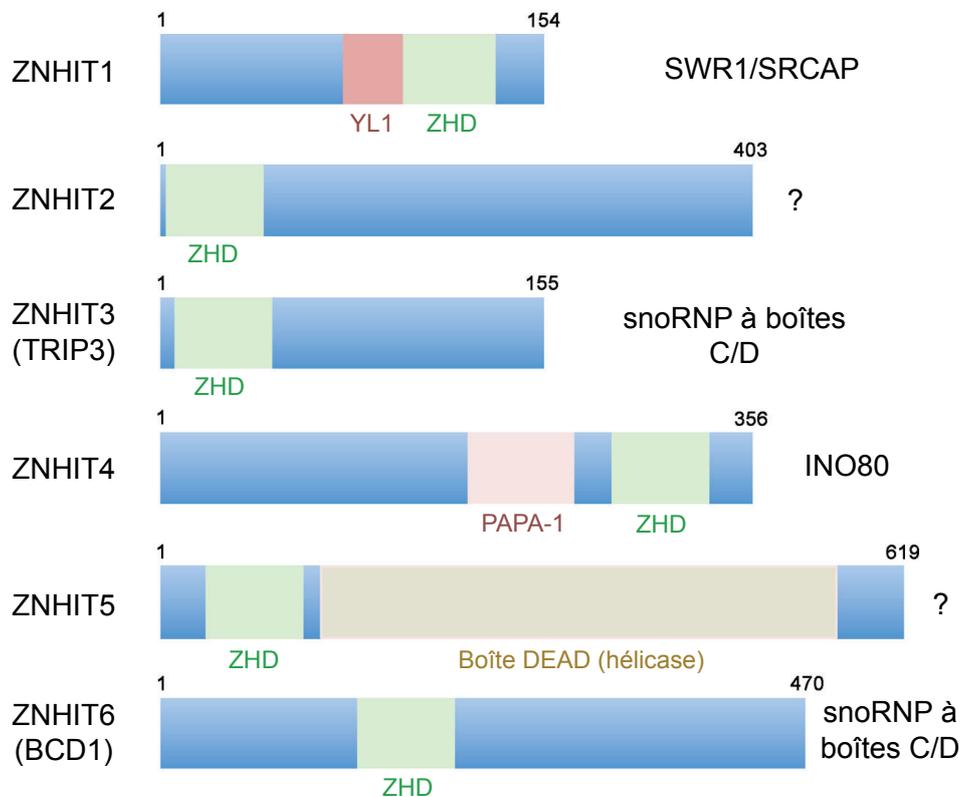


Figure 26. Les doigts à zinc de la famille zf-HIT (ZHD). (A) Alignement de séquences des différents membres de la famille zf-HIT réalisé avec le programme Esript 3.0 (Robert et Gouet, 2014). Les résidus cystéine et histidine de Bcd1p coordonnant les atomes de zinc sont précisés par des flèches noires pointant vers le bas au-dessus des séquences et numérotés en dessous. Les traits sous l'alignement rejoignent les résidus impliqués dans la coordination du premier atome de zinc (vert) et du second (orange). Les histidines impliquées dans la coordination du second atome de zinc du domaine zf-HIT de la protéine Hit1p sont colorés en bleu. Les séquences proviennent de différents organismes (levure : *Saccharomyces cerevisiae*, humain : *Homo sapiens*, SCHPO : *Schizosaccharomyces pombe*, ARATH : *Arabidopsis thaliana*). Les numéros d'accèsion à la base de données UniProtKB sont donnés entre parenthèses pour chaque protéine. (B) Représentation schématique de la séquence des 6 protéines ZNHIT, membres de la famille zf-HIT retrouvées chez l'Homme (d'après Verheggen *et al.*, 2015).

suggèrent que les ZHD sont des co-facteurs clés des protéines RuvBL1/2 (Verheggen *et al.*, 2015), qui pourraient avoir évolué de manière concomitante aux RuvBL1/2.

3. Les chaperonnes d'histone

3.1. Composition et structure de la chromatine

Au sein du noyau des cellules eucaryotes, l'ADN est assemblé sous forme de chromatine. Il s'agit d'une structure ADN-protéine, constituée d'un enchaînement de nucléosomes, permettant à l'ADN d'adopter une forme plus compacte. Dans un nucléosome, l'ADN s'enroule autour d'un complexe constitué de protéines particulières appelées histones. Chaque nucléosome contient une séquence de 147 pb (paire de bases) d'ADN, ainsi que 8 histones de quatre types différents. Dans cet octamère, les histones sont agencées en un tétramère d'histones H3-H4 (Figure 27A) et deux dimères d'histones H2A-H2B (Figure 27B) (Dahlin *et al.*, 2015). L'ADN effectue approximativement deux tours autour d'un octamère d'histones (Ramakrishnan, 1997). La chromatine structurée en nucléosomes se retrouve sous forme d'une fibre desserrée (Kowalski et Pałyga, 2016). Une autre histone, l'histone H1 se lie à l'ADN des nucléosomes et provoque une réorganisation de leur part en fibre de chromatine dite de 30 nm (Ramakrishnan, 1997; Chen *et al.*, 2014; Kowalski et Pałyga, 2016). L'absence ou la présence d'histone H1 sur la chromatine a donc une importance cruciale sur sa structure (Figure 28), qui se trouve ainsi respectivement sous forme détendue et donc active, ou condensée et alors inactive (Kowalski et Pałyga, 2016). L'histone H1 participe ainsi à la modulation de l'expression des gènes.

La chromatine contient l'information épigénétique et correspond à la forme de l'ADN nécessaire aux mécanismes de transcription, de réparation de l'ADN ou de réplication. La structure de la chromatine doit donc être finement régulée pour satisfaire tous ces processus. Cette régulation est en partie orchestrée par la présence ou l'absence de modifications post-traductionnelles sur les histones (Dahlin *et al.*, 2015).

3.2. Assemblage des nucléosomes

L'assemblage des nucléosomes est régi par deux grands types de mécanismes (Figure 29), liés aux raisons pour lesquelles l'assemblage des nucléosomes est requis (Dahlin *et al.*, 2015). Le premier mécanisme est associé à la réplication de l'ADN durant la phase S (Replication-Coupled Nucleosome Assembly ou RCNA), où l'ADN parental double brin doit

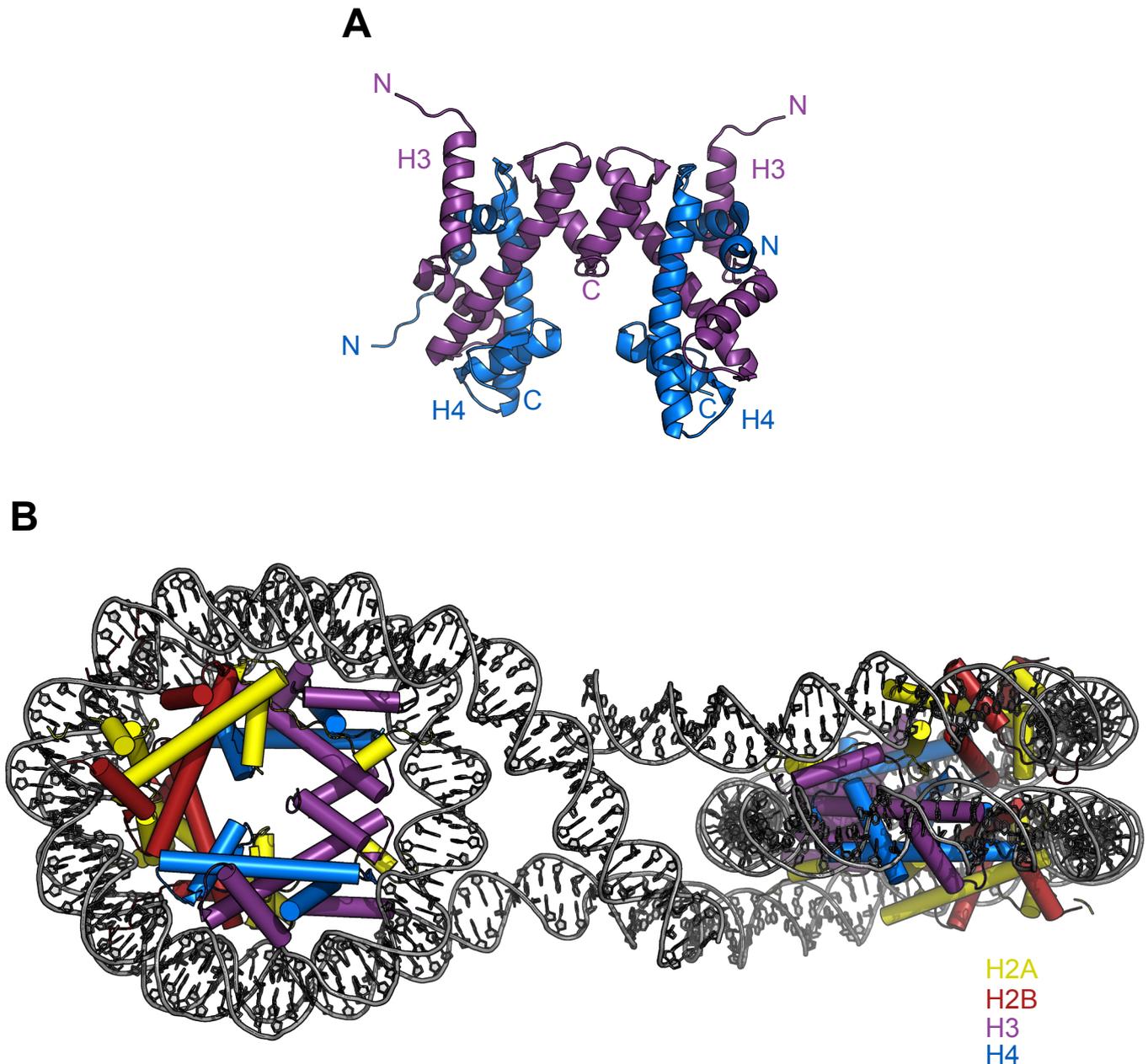


Figure 27. Structures 3D des nucléosomes. (A) Structure tridimensionnelle d'un tétramère d'histones H3-H4 compris dans (B) la structure tridimensionnelle de deux nucléosomes contenant chacun un tétramère d'histones H3-H4 (violet et bleu) et deux dimères d'histones H2A-H2B (jaune et rouge) (Schlach et al., 2005; code PDB : 1ZBB). Les hélices des protéines sont représentées en cylindres.

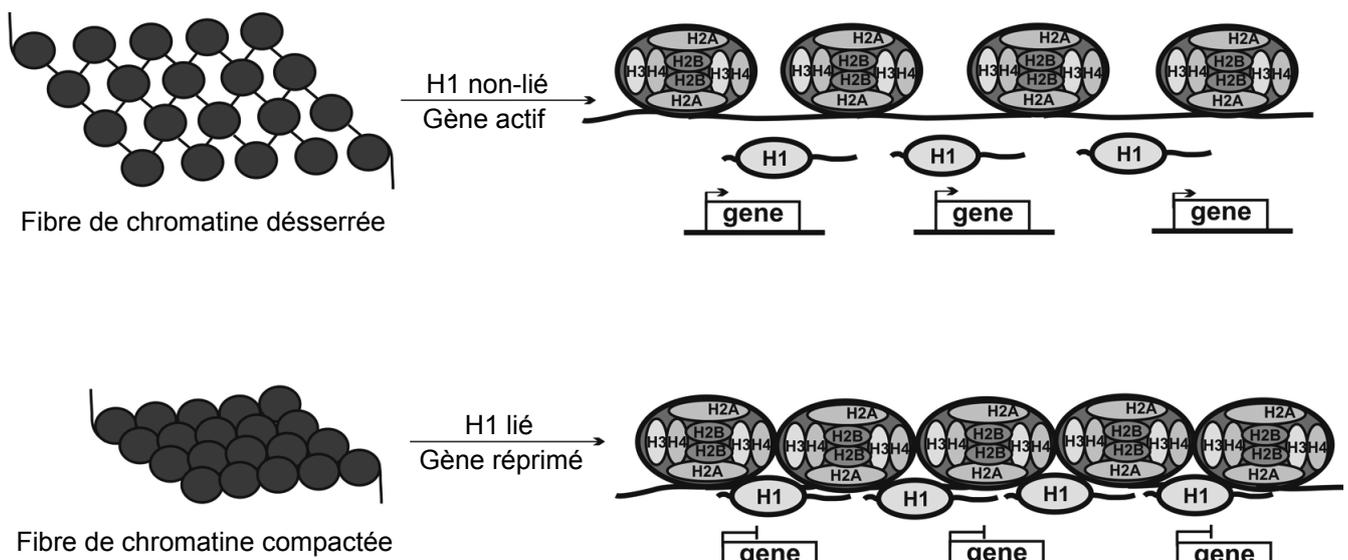
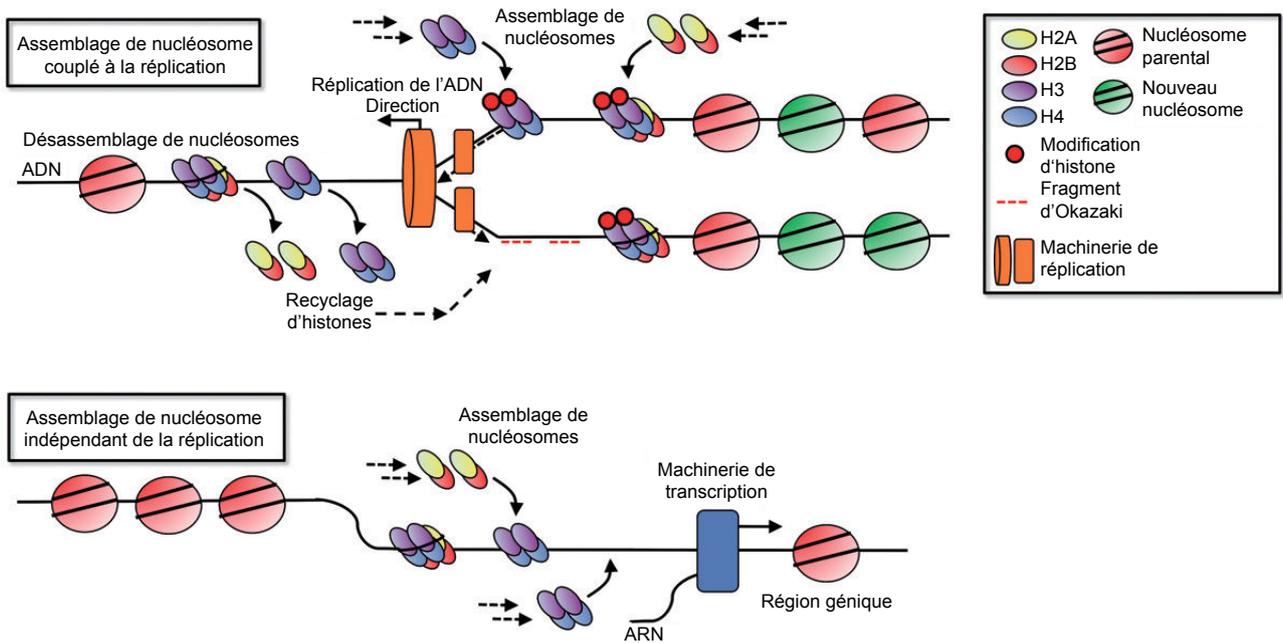


Figure 28. Importance de l'histone H1. (A) Schéma représentant l'effet de la fixation de l'histone H1 sur la chromatine (d'après Kowalski et Pałyga, 2016).

A



B

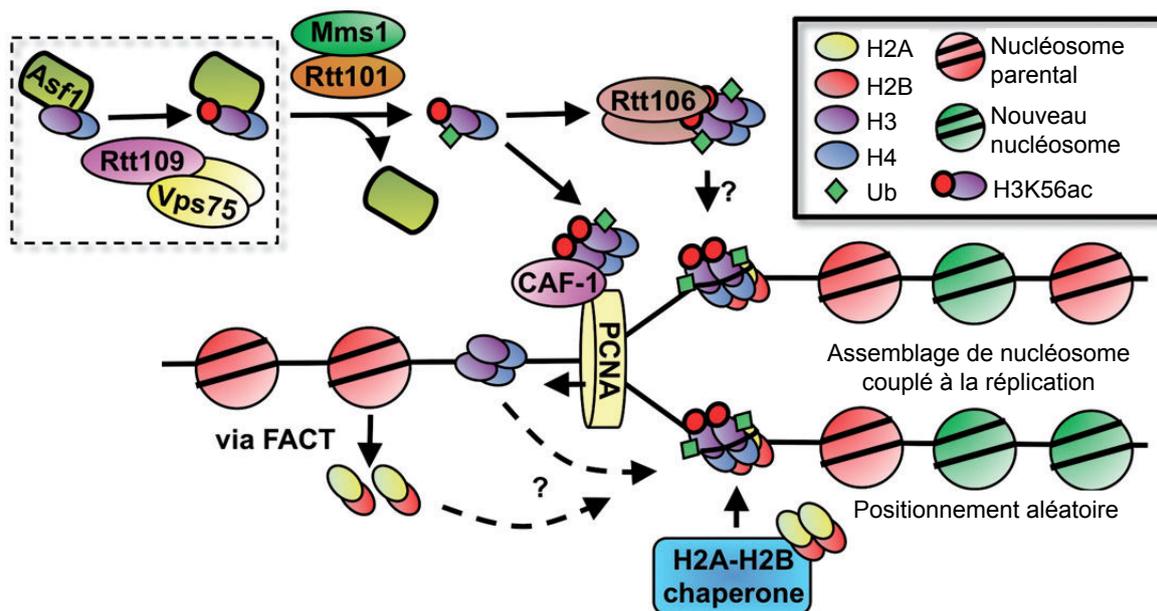


Figure 29. Assemblage de la chromatine. (A) Schéma des déplacements d'histones lors de l'assemblage de nucléosomes couplé à la réplication (RCNA) en haut et de l'assemblage de nucléosomes indépendamment de la réplication en bas. Les couleurs des histones correspondent à celles de la figure précédente (d'après Dahlin *et al.*, 2015). (B) Schéma des voies menant à l'assemblage de nucléosomes couplé à la réplication (RCNA). Les ronds rouges représentent des acétylations et les losanges verts des ubiquitylations.

donner naissance à deux nouveaux ADN double brin contenant chacun un brin parental et un brin néoformé. Les nucléosomes de l'ADN parental doivent donc être retirés pour permettre la réplication, tandis que de nouveaux nucléosomes sont mis en places sur les nouveaux ADN. Le second mécanisme n'est pas lié à la réplication, mais à d'autres processus nécessitant le retrait temporaire des nucléosomes de l'ADN, tels que la transcription de gènes ou la réparation de l'ADN (Dahlin *et al.*, 2015). Il s'agit de l'assemblage de nucléosomes indépendant de la réplication (*Replication-Independent Nucleosome Assembly* ou RINA).

3.2.1. Assemblage des nucléosomes couplé à la réplication (RCNA)

Lors de la phase S, les nucléosomes présents sur l'ADN génomique doivent être désassemblés pour laisser la place à la machinerie de réplication (Dahlin *et al.*, 2015). Une fois la réplication de l'ADN effectuée, les histones provenant des nucléosomes désassemblés du brin parental peuvent être recyclées lors de la formation de nouveaux nucléosomes sur les brins néo-synthétisés (Figure 29A). Les autres nucléosomes nécessaires sont produits à partir d'histones nouvellement synthétisées. Le démantèlement des nucléosomes commence toujours par le départ des dimères d'histones H2A-H2B, suivi de celui du tétramère d'histones H3-H4. L'assemblage se fait ensuite en sens inverse, avec l'arrivée du tétramère H3-H4 puis des dimères H2A-H2B pour former un nucléosome complet. Le mécanisme de RCNA est similaire entre la levure et l'Homme, quoique plus complexe chez l'Homme notamment par la présence de plusieurs variant de l'histone H3 inexistants chez la levure (Dahlin *et al.*, 2015).

3.2.2. Assemblage des nucléosomes indépendant de la réplication (RINA)

Ce mode d'assemblage met en œuvre un phénomène dit d'échange d'histones, permettant aux cellules de remplacer des histones endommagées par d'autres et ainsi de réguler la chromatine de manière dynamique sans avoir recours à la réplication (Dahlin *et al.*, 2015). Ce mécanisme est particulièrement utile aux cellules ne se divisant pas et ne pouvant donc pas utiliser le RCNA (Figure 29A). Les systèmes d'assemblage de nouveaux dimères H2A-H2B ou de leur échange apparaissent fonctionner de façons similaires dans les deux différents types d'assemblages RCNA et RINA. Ces histones sont en effet plus aisément interchangeables que les histones H3-H4 formant des tétramères très stables. Le système RINA est également utilisé dans les cas de réparations de l'ADN, où l'échange d'histones occupe une part importante du mécanisme. En effet, la présence de dommages de l'ADN est

signalée par le remplacement de certaines histones par des formes variantes (Dahlin *et al.*, 2015).

3.3. Les protéines liées aux nucléosomes

De nombreuses classes de protéines sont retrouvées liées aux histones. En effet, ceux-ci ont bien souvent besoin d'être accompagnées par des protéines chaperonnes ou des complexes chaperons. De plus, les modifications post-traductionnelles des histones ou PTM (*Post-Translational Modifications*) sont essentielles aux processus liés à la chromatine. Ces PTM jouent généralement des rôles de signalisation.

3.3.1. *Modification des histones*

Plusieurs types de PTM peuvent être introduits au niveau des histones, comme la méthylation, l'acétylation, la phosphorylation, la sumoylation ou l'ubiquitinylation. L'ajout ou le retrait de ces PTM sur les histones sont réalisés par des enzymes de modification des histones, tels que les histones acétyltransférases (HAT) ou déacétylases (Dahlin *et al.*, 2015). L'enzyme Rtt109 par exemple est une HAT responsable de la formation de H3K56ac (histones H3 acétylées sur la lysine 56), typique des histones H3 nouvellement synthétisées.

3.3.2. *Chaperonnes d'histones*

Les histones sont des protéines particulières ayant une forte tendance à l'agrégation si elles ne sont pas prises en charge par d'autres facteurs. Il s'agit du principal intérêt des chaperonnes d'histones. Leur variété est liée à la nécessité qu'ont les histones à être guidées différemment en fonction des étapes d'assemblage/désassemblage des nucléosomes, et leur spécificité liée à la présence de PTM sur les histones. Par exemple, les protéines Vps75 et Asf1 sont nécessaires au fonctionnement de l'HAT Rtt109, dans le cadre du RCNA (Figure 29B). Vps75p guide l'acétylation de l'histone H3 sur la lysine 56 par Rtt109, sur l'hétérotrimère Asf1p-H3-H4 en tant que substrat.

3.3.2.1. *Le complexe CAF-1*

Ce complexe (*Chromatin Assembly Factor-1*) est un chaperon des histones H3-H4 nouvellement synthétisées, et permet leur adressage et leur assemblage sur l'ADN. Ce complexe intervient après la chaperonne Asf1p dans l'ordre chronologique de circulation des

histones H3-H4 vers l'ADN (Dahlin *et al.*, 2015). Durant le RCNA, le dépôt de tétramères H3-H4 nouvellement synthétisés par le complexe CAF-1 empêche la formation de nucléosomes mixtes, contenant un dimère d'histones parental et un autre nouvellement synthétisé (Figure 29B).

3.3.2.2. Le complexe FACT

Ce complexe intervient suite au complexe CAF-1 et à l'arrivée du tétramère d'histones H3-H4 sur l'ADN, en y adressant les dimères d'histones H2A-H2B (Dahlin *et al.*, 2015). Ce complexe intervient dans les deux mécanismes RCNA et RINA. Il a d'ailleurs également été décrit comme étant impliqué dans l'assemblage des histones H3-H4 sur les nucléosomes après la transcription de gènes (Figure 29B).

3.3.2.3. La chaperonne d'histone Rtt106

La protéine Rtt106 (*Regulator of Tyl Transposition 106*) est également une chaperonne des histones H3-H4, ayant *a priori* la même fonction que le complexe CAF-1, avec lequel elle est d'ailleurs capable d'interagir par l'intermédiaire de la protéine Cac1 (Huang *et al.*, 2005a). Elle est également un partenaire du complexe HIR (*H*istone *R*egulatory), où elle se retrouve au centre des interactions et permet notamment de recruter les sous-complexes inhibiteur RSC (*Remodel the Structure of Chromatin*) ou activateur SWI/SNF (*SW*itch/*S*ucose *N*on-*F*ermentable) de la transcription (Figure 31A) (Amin *et al.*, 2012). Rtt106p paraît impliquée dans la formation de l'hétérochromatine et dans la régulation de l'épigénétisme (Huang *et al.*, 2005b). Elle permet l'assemblage des nucléosomes en facilitant l'assemblage des histones H3-H4 en tétramères (Fazly *et al.*, 2012) et est également impliquée dans l'élongation de la transcription et la régulation du taux d'adressage de l'histone H3 sur les régions en cours de transcription (Imbeault *et al.*, 2008). La protéine Rtt106, au même titre que le complexe CAF-1, reconnaît préférentiellement l'histone H3 acétylée sur la lysine K56 (Li *et al.*, 2008), ce qui renforce sa capacité à assembler l'ADN après réplication sous forme de nucléosomes. Cette PTM pourrait donc être à l'origine de la régulation de la voie RCNA par l'intermédiaire de ces chaperonnes.

La protéine Rtt106 contient 3 domaines distincts (Figure 30A). La région N-terminale contient un domaine de dimérisation (Figure 30B) (Fazly *et al.*, 2012; Su *et al.*, 2012). Un domaine structuré en double domaine PH (*P*leckstrin *H*omology) est présent dans la région centrale (Figure 30C) (Liu *et al.*, 2010; Su *et al.*, 2012; Zunder *et al.*, 2012); il s'agit d'un

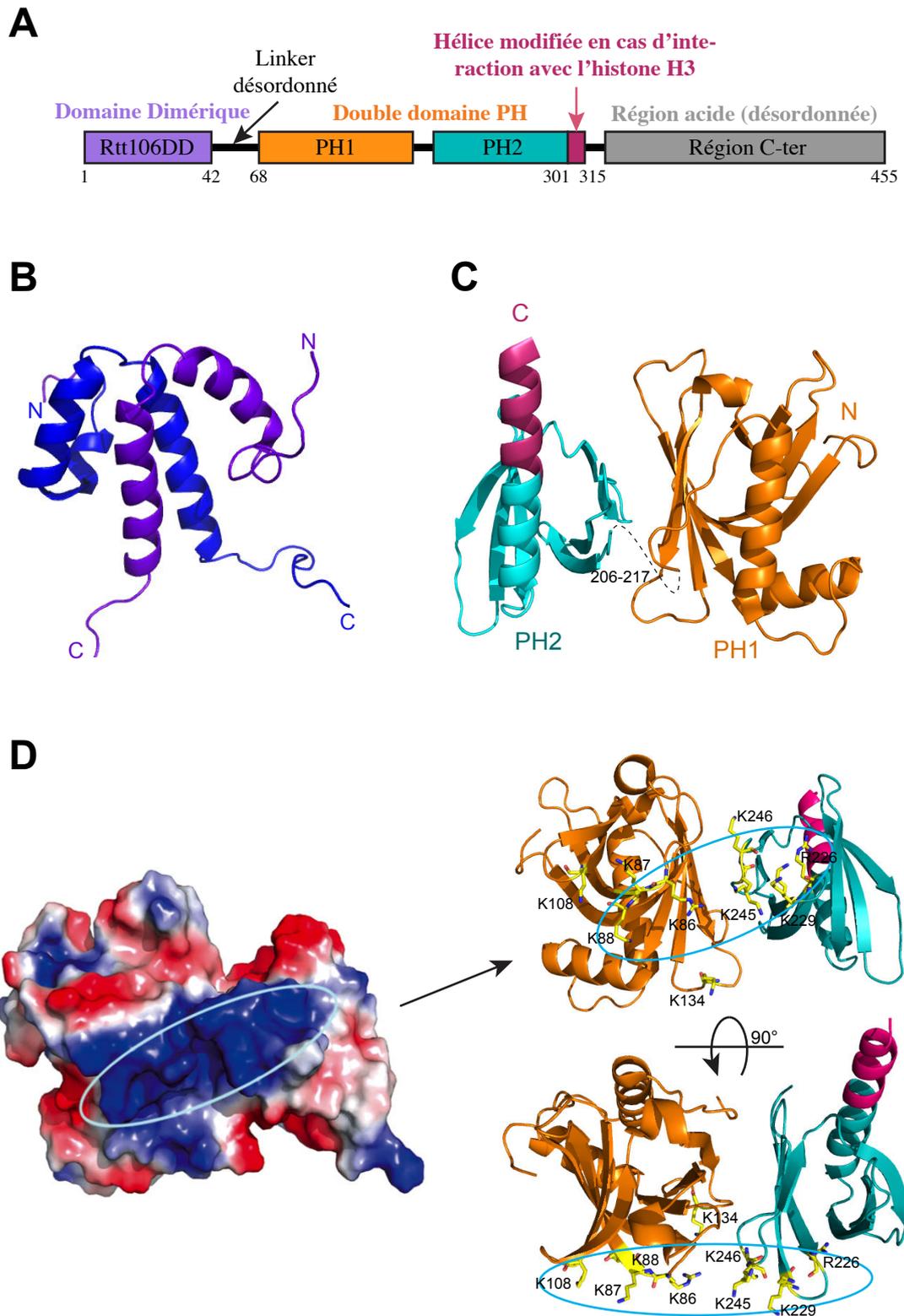
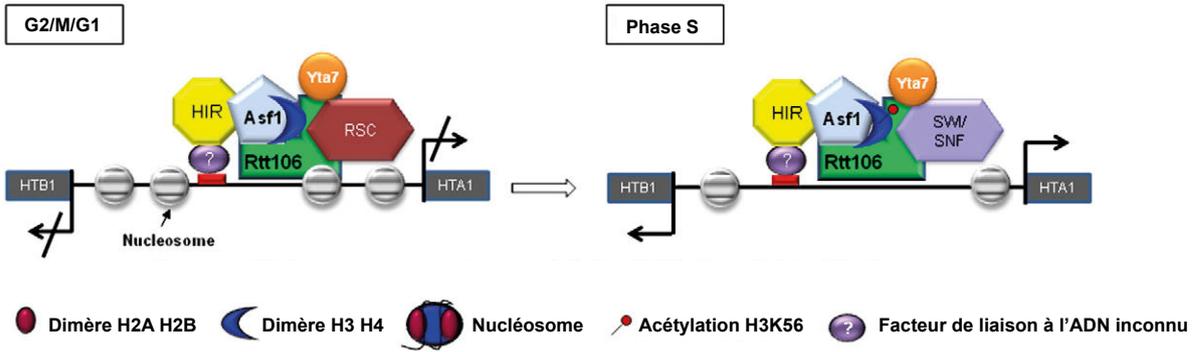


Figure 30. La protéine Rtt106. (A) Représentation schématique des différents domaines de la protéine Rtt106. (B) Structure tridimensionnelle du dimère formé par le domaine dimérique de Rtt106 (Rtt106DD) obtenue par RMN (Su *et al.*, 2012; code PDB : 2LH0). (C) Structure tridimensionnelle du domaine tandem PH (*Pleckstrin Homology*) de Rtt106p(68-301) obtenue par diffraction des rayons X (Su *et al.*, 2012; code PDB : 3TVV). (D) Potentiel électronique de surface du domaine double PH de la protéine Rtt106 formant un sillon électro-positif (à gauche) et représentation des résidus positivement chargés impliqués dans la formation de ce sillon (à droite) (Liu *et al.*, 2010; Su *et al.*, 2012).

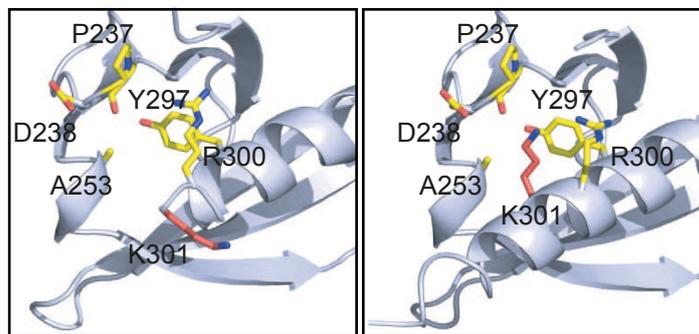
type de domaine connu pour être responsable d'interactions protéine-protéine. Enfin, sa région C-terminale est acide et apparemment déstructurée (Su et al., 2012). Il semble que le domaine PH soit capable de se lier à l'ADN par l'intermédiaire d'une région formant un sillon positivement chargé pouvant accueillir une double hélice d'ADN (Figure 30D) (Liu et al., 2010). L'interaction de Rtt106p avec les histones H3-H4 est elle aussi réalisée par ce double domaine PH (Figure 31C), bien que le domaine de dimérisation semble également capable de lier l'histone H3 non modifiée avec une faible affinité (Su et al., 2012). Ce domaine PH est capable d'interagir avec l'histone H3 non modifiée, mais l'affinité entre les deux partenaires est très fortement augmentée en présence d'acétylation sur la lysine 56 de l'histone H3. Lors de l'interaction entre les deux protéines, ce groupement se retrouve projeté dans une poche au sein du domaine PH2 de la protéine Rtt106 faisant intervenir les résidus D238, A253, Y297, R300 et K301 (Figure 31B). Ceci favorise une forme plus courte de l'hélice $\alpha 5$ de la protéine Rtt106, permettant une fixation thermodynamiquement plus favorable entre les deux partenaires (Su *et al.*, 2012). Il a toutefois été également montré que des résidus du domaine PH1 étaient importants dans l'interaction avec l'histone H3, et particulièrement les résidus S80 et R86 (Zunder *et al.*, 2012).

Le domaine en double PH de la protéine Rtt106 est très similaire à celui de la protéine Pob3, qui est une des sous-unités du complexe FACT. Toutefois, Pob3p n'est pas capable d'interagir directement avec les histones H3-H4, contrairement à Rtt106p. Cela est notamment dû au fait que l'une des différences entre les deux protéines est une prolongation de l'hélice α située en position C-terminale de Rtt106p que ne possède pas Pob3p (Figure 32). Or cette extension de l'hélice est un site important dans l'interaction avec l'histone H3. En comparaison à Rtt106p, la protéine Pob3 possède également deux hélices α supplémentaires dans le motif PH1. A l'heure actuelle, aucune protéine orthologue de Rtt106p chez l'Homme n'a été identifiée (Liu *et al.*, 2010; Su *et al.*, 2012; Zunder *et al.*, 2012).

A



B



C

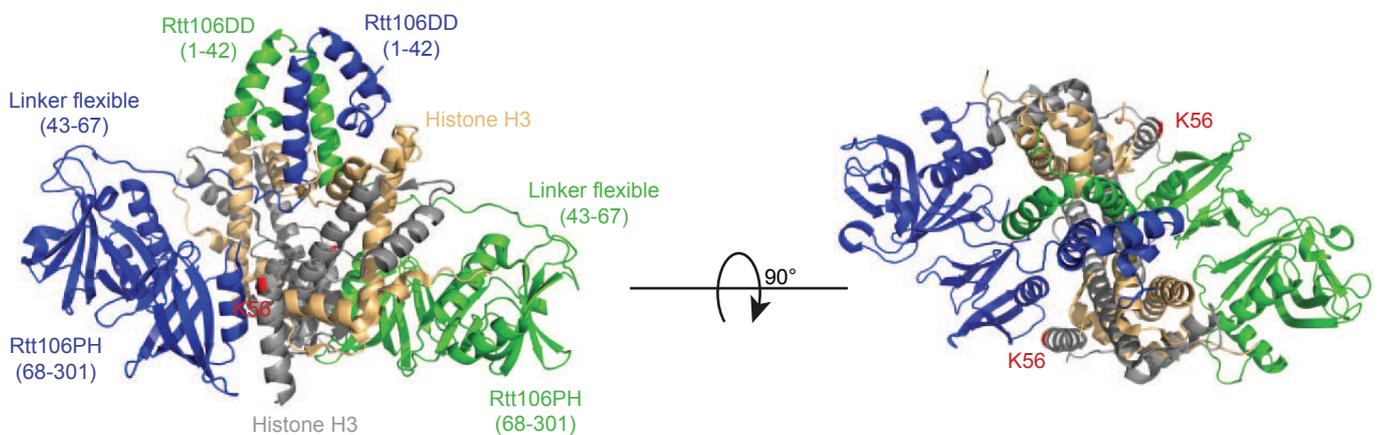


Figure 31. Lien entre Rtt106 et les histones H3-H4. (A) La protéine Rtt106 est retrouvée chez la levure *Saccharomyces cerevisiae* associée au complexe HIR, dont le rôle est de réguler la transcription des gènes d'histones en fonction du cycle cellulaire. La protéine Rtt106 peut notamment recruter le complexe RSC, un répresseur de la transcription des gènes d'histones ou le complexe SWI/SNF, un activateur de la transcription des gènes d'histones (Amin *et al.*, 2012). (B) Représentations de la poche d'interaction théorique pour la lysine 56 acétylée de l'histone H3. La conformation de gauche, avec l'hélice $\alpha 5$ plus courte, est favorable à l'interaction. Dans la conformation de droite (censée être en absence de l'histone H3) la lysine 301 bloque l'accès à la poche. L'hélice $\alpha 5$ est alors plus longue (Su *et al.*, 2012). (C) Modèle structural de la protéine Rtt106 dimérique en complexe avec les histones H3 et H4 (Su *et al.*, 2012).

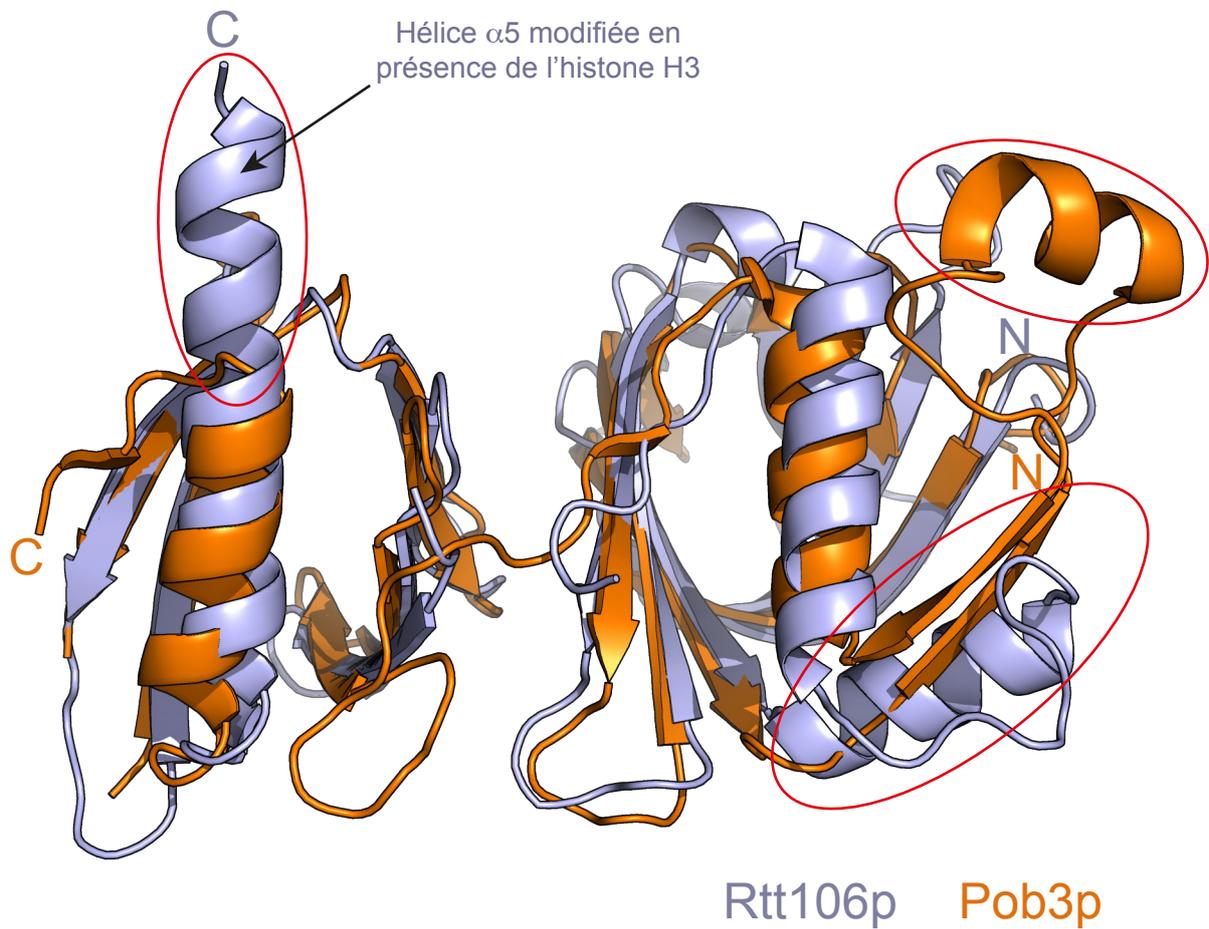


Figure 32. Superposition des structures des domaines en double PH des protéines Rtt106 (en gris; code PDB : 3TVV) et Pob3 (en orange; code PDB : 2GCJ). Les trois différences majeures entre les deux structures sont encerclées en rouge.

Objectifs

OBJECTIFS

La protéine Bcd1 a été identifiée comme étant nécessaire à la viabilité cellulaire chez la levure en agissant sur la biogenèse des snoRNP à boîtes C/D. Pour autant, sa fonction au sein du noyau reste encore inconnue. La seule donnée que nous ayons est que la réduction de son taux d'expression engendre une forte baisse du taux de snoRNA, mais uniquement ceux à boîtes C/D. Aucun lien ne lui a été trouvé avec les snoRNP à boîtes H/ACA. Cependant, cette protéine n'est pas retrouvée comme constituant des particules snoRNP à boîtes C/D matures. Cela signifie donc que Bcd1p doit être un facteur d'assemblage spécifique de ces particules. Nous ne savons toutefois pas à quelle étape de l'assemblage elle intervient. Plusieurs associations, avec la protéine Pih1 par exemple ou des interactions avec les protéines Rvb1/2, qui sont des facteurs impliqués dans l'assemblage des snoRNP à boîtes C/D ont bien été identifiées, mais les données sur le réseau d'interaction entretenu par Bcd1p dans cette voie restent encore très vagues. Qui plus est, nous ignorons également si cette protéine ne possède qu'une fonction restreinte au seul assemblage des snoRNP à boîtes C/D, ou si elle participe à d'autres fonctions cellulaires encore non identifiées.

A mon arrivée au laboratoire, la seule information connue sur la séquence de la protéine Bcd1 était la présence d'un domaine en doigt à zinc de la famille zf-HIT (ZHD) dans sa région N-terminale. Une interaction avec les protéines RuvBL1/2 chez l'Homme avait également été montrée par l'équipe de N. Watkins par *GST-pull down*. Enfin l'étude des bases de données issues d'approches à haut débit laissait soupçonner une association avec la chaperonne d'histone Rtt106p, association qui a ensuite été validée par double hybride au laboratoire. Le but de mon travail a donc été de caractériser la protéine Bcd1 structuralement et fonctionnellement par des approches de biochimie et de biologie structurale.

1) Le point de départ de mon travail a été de déterminer la structure tridimensionnelle en solution du ZHD de Bcd1p par RMN. L'outil RMN nous a effectivement semblé être le plus approprié à l'étude d'un petit domaine de 45 acides aminés. La forte similarité de séquence entre les différents ZHD et leur appartenance à des protéines présentes dans différents complexes nucléaires nous a conduit à penser que leur fonction pouvait être commune et d'une grande importance. Un membre de cette famille est également un autre facteur d'assemblage des snoRNP à boîtes C/D, la protéine Hit1. Ce dernier présentait

toutefois un motif séquentiel différent des autres ZHD au niveau des deux derniers résidus coordonnant le second atome de zinc, de type HxH au lieu de HxxxxxC (où les x représentent n'importe quel résidu). Pour cela, nous avons choisi de déterminer en parallèle la structure 3D de ce domaine. En plus de l'étude structurale de ces domaines, nous avons voulu identifier quelle(s) est (sont) leur(s) fonction(s). Pour cela, nous avons travaillé de concert avec Decebal Tiotiu, également en thèse au laboratoire, qui a étudié l'importance du ZHD de Bcd1p sur la croissance des levures ainsi que sur le taux des snoRNA à boîtes C/D. Les doigts à zinc étant connus pour être des modules d'interaction protéine-acide nucléique ou protéine-protéine, des tests d'interaction ont été menés entre le ZHD de Bcd1p et des cibles ARN, ADN ou protéine. En ce qui concerne les cibles protéiques, nous avons principalement choisi de tester les protéines Rvb1/2, qui semblent avoir un lien évolutif avec les ZHD et font également partie des facteurs d'assemblage des snoRNP à boîtes C/D en tant que membres du complexe R2TP.

Si le ZHD de Bcd1p s'est avéré important pour sa fonction, il n'est pas suffisant pour l'assurer seul. Une autre partie de mon travail a été d'identifier et d'étudier structurellement un fragment minimum fonctionnel correspondant aux résidus 1 à 96. Le but étant de pouvoir par la suite étudier *in vivo* et *in vitro* ce fragment.

2) La protéine Bcd1 semble être au centre du vaste réseau d'interactions visant à assurer la biogenèse correcte des snoRNP à boîtes C/D. Ainsi, la compréhension de ce réseau et la connaissance de toutes ces interactions doivent pouvoir nous permettre de progresser dans la compréhension du mécanisme menant à des snoRNP fonctionnelles. Avant mon arrivée au laboratoire, un complexe entre les protéines Nop58, Rvb1, Rvb2 et Bcd1 venait d'être mis en évidence par une approche de purification en tandem menée sur Bcd1p et analysé par spectrométrie de masse. Une analyse par co-expression des partenaires de ce complexe a donc été entreprise dans le but de vérifier la formation d'un tel complexe *in vitro*, et le cas échéant, de caractériser les interactions mises en jeu.

3) Comme mentionné plus haut, une association entre les protéines Bcd1 et Rtt106 venait d'être identifiée par double hybride au laboratoire par D. Tiotiu. Après avoir validé cette interaction par co-expression chez *E. coli*, mon but a été d'en caractériser les paramètres thermodynamiques. Par la suite, nous avons voulu déterminer les régions précises de ces deux protéines permettant de réaliser l'interaction. Pour cela, nous nous sommes appuyés sur la technique de protéolyse ménagée analysée par spectrométrie de masse. Ceci nous a permis de

mettre en évidence un domaine structuré de la protéine Bcd1, le domaine 120-303, lequel a par la suite fait l'objet d'une étude structurale par RMN. Plusieurs techniques ont par la suite été combinées afin de caractériser plus précisément cette interaction, comme de l'échange hydrogène/deutérium suivi par spectrométrie de masse, du *cross-link*, de la mobilité ionique, des titrages suivis par RMN ou encore du SAXS. Notre but final étant d'obtenir des informations à l'échelle atomique, une étude structurale du complexe par diffraction des rayons X a également été initiée.

Afin de répondre à ces différentes problématiques, les travaux menés dans le cadre de ce travail de thèse ont été effectués en collaboration avec Decebal Tiotiu, qui a réalisé les études de croissance de levure, de double hybride et de Northern blot. De plus, d'autres collaborations avec des équipes françaises ou étrangères ont été réalisées. J'ai notamment pu profiter du réseau de laboratoires français associés au nôtre dans le cadre d'un projet ANR. Ainsi, les analyses par spectrométrie de masse ont été effectuées par Guillaume Terral de l'équipe de Sarah Cianferani, au laboratoire LSMBO de Strasbourg. Les expériences de SAXS ont, elles, été menées au synchrotron SOLEIL par Aurélien Thureau. Nos études sur les protéines RuvBL1/2 ont été possibles grâce à un échange avec l'équipe de Pedro Matias de l'IBET au Portugal, qui nous a fourni ces protéines recombinantes. Enfin, les travaux débutés sur l'implication potentielle de la protéine Bcd1 sur les nucléosomes par l'intermédiaire de la protéine Rtt106 sont réalisés en collaboration avec l'équipe de Georges Mer à la Mayo Clinic aux Etats-Unis (Minnesota).

Matériels

&

Méthodes

1. Matériels

1.1. Les oligonucléotides

La liste des oligonucléotides utilisés pour le travail de thèse est consultable dans les Tables 2 et 3.

1.2. Les souches bactériennes d'*Escherichia coli*

L'amplification et l'extraction d'ADN plasmidique nécessaire aux étapes de génie génétique ont été réalisées avec la souche **DH5 α** (*supE44*, Δ *lacU169* (Φ 80 *lacZ* Δ M15), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*).

La surexpression de protéines recombinantes pour les analyses biochimiques, biophysiques ou structurales a été réalisée avec la souche **BL21 (DE3) [pRARE2]** (*F⁻ ompT hsdS_B (r_B⁻ m_B⁻) gal dcm* λ (DE3), [*argN5*, *argX*, *proL*, *leuW*, *metT*, *argW*, *thrT*, *glyT*, *thrU*, *argU*, *ileX*, *Cam^r*]).

La surexpression de protéines recombinantes avec incorporation de sélénométhionines pour la cristallographie aux rayons X a été réalisée avec la souche **B834 (DE3) [pRARE2]** (*F⁻ ompT hsdS_B (r_B⁻ m_B⁻) gal dcm met* λ (DE3), [*argN5*, *argX*, *proL*, *leuW*, *metT*, *argW*, *thrT*, *glyT*, *thrU*, *argU*, *ileX*, *Cam^r*]).

1.3. Les milieux de culture bactérienne

Tous les milieux de culture sont stérilisés par autoclavage à 120°C pendant 20 min à une pression de 1 bar. Dans le cas des milieux solides (en boîtes de pétri), de l'agar (17 g/L) est ajouté au milieu liquide avant la stérilisation. Les antibiotiques sont ajoutés le cas échéant après la stérilisation, une fois la température descendue en dessous de 50°C.

Le milieu riche LB (*Lysogeny Broth* ou Luria-Bertani) est utilisé classiquement pour la surexpression de protéines recombinantes (souche BL21) ou l'amplification d'ADN plasmidique (souche DH5 α). Il se compose de tryptone (10 g/L), d'extrait de levure (5 g/L) et de NaCl (5 g/L). Le pH est ensuite ajusté à 7,5.

Le milieu riche 2TY est un milieu plus riche que le LB. Les quantités de tryptone (*Tryptone*) et d'extrait de levure (*Yeast extract*) sont doublées (2TY). Il se compose donc de tryptone (16 g/L), d'extrait de levure (10 g/L) et de NaCl (5 g/L). Le pH est ensuite ajusté à 7.

Le milieu riche SOC (*SOB with Catabolic repression*) est utilisé lors de l'étape de régénération après les transformations de bactéries par des plasmides. Il se compose de milieu

n°	Séquence 5' - 3'	Utilisation
6360	GGATATCCATATGGAGTTTGCTGCTGAAAAAT	Oligo 5' de clonage de hBCD1 avec site NdeI
6361	GGGAGATCTTCAATTTTCATTGCCAACGTTCTT	Oligo 3' de clonage de hBCD1 avec site BglII
8677	GGGGGATCCTTAAACATAGAAATCCACTCCACGCT	Oligo 3' de clonage de Bcd1p(x-168) avec site BamHI
8691	CCCGCATATGGATGAAATTTCCGAAACCA	Oligo 5' de clonage de Rtt106(65-320) avec site NdeI
9475	GGATATCCATATGAAGGGGATGCAGCGATCATCGCAA	Oligo 5' de clonage de Bcd1p(139-x) avec site NdeI
9476	GGGGGATCCTTAAATCCATATATGCGTGGTGTGTA	Oligo 3' de clonage de Bcd1p(x-255) avec site BamHI
9477	GGATATCCATATGGCGTTGAAACAGGCGGACGACAAG	Oligo 5' de clonage de Bcd1p(54-x) avec site NdeI
9478	GGGGGATCCTTATATATCGTATCTTCTCTTCTTGAA	Oligo 3' de clonage de Bcd1p(x-115) avec site BamHI
9479	GGATATCCATATGCGTGTAGCAGCGAGTGTCACTGG	Oligo 5' de clonage de Bcd1p(120-x) avec site NdeI
9480	GGATATCCATATGAATGCTTATGTCCAGAGGGACTAC	Oligo 5' de clonage de Bcd1p(66-x) avec site NdeI
9599	GGATATCCATATGATGGCGGTGTGAGTGGTGTAAAGCGAATAAAAAG AGTTC	Oligo 5' de mutation de Bcd1p C5S+C8S avec site NdeI
9621	GGGGGATCCTTATTTCCCTTTTAAAGTAGTCGTCATTTTAGC	Oligo 3' de clonage de Rtt106(x-301) avec site BamHI
9622	GGGGGATCCTTATTTAGCTTTTCAATTCCCTCAGA	Oligo 3' de clonage de Rtt106(x-315) avec site BamHI
9623	GGATATCCATATGTCGGAAACCAATACTATTTTC	Oligo 5' de clonage de Rtt106(68-x) avec site BamHI
9634	GGATATCCATATGAATTGCTCAGGTCAAACACATGATCC	Oligo 5' de clonage de Bcd1p(37-x) avec site NdeI
9635	GGGGGATCCTTGAGAGTCTCTTCCAGCGTCTTC	Oligo 3' de clonage de Bcd1p(x-331) avec site BamHI
9646	GGATATCCATATGTACGATATAGATGAGGATGATCGTGTAGC	Oligo 5' de clonage de Bcd1p(113-x) avec site NdeI
9647	GGGGGATCCTTACTTGTTCAGTGTACTCGTATGCTC	Oligo 3' de clonage de Bcd1p(x-315) avec site BamHI
9648	GGGGGATCCTTACTTCAATTTGTTTCCCTTTTAAACGTA	Oligo 3' de clonage de yRtt160p (x-304) avec site BamHI
9696	GGATATCCATATGCGTGTAGCAGCGAGTGTGAGAGGATA	Oligo 5' de clonage de Bcd1p(120-x) avec site NdeI
9697	GGGGGATCCTTAAATCCTTCATTTGTTTCCCTTTTAAACGTAAGTCGTC	Oligo 3' de clonage de yRtt106(x-305) avec site BamHI
9801	GGATATCCATATGAGGAGCAAGTGGGACAAAACGATG	Oligo 5' de clonage de Bcd1p(149-315) avec site NdeI
9802	GGATATCCATATGGCCATGTCAAGGTGTGAGACTTGT	Oligo 5' de clonage de hBCD1 zf (216-260) avec site NdeI
9803	GGGGGATCCTTATTTATCTCGAAGTCCATTACATGT	Oligo 3' de clonage de hBCD1 zf (216-260) avec site BamHI
9806	GGGGGATCCTTAAATCATGTGTTTGACCTGAGCAATT	Oligo 3' de clonage de Bcd1p(1-45) avec site BamHI
9807	GGGGGATCCTTATTTGTTTCGCTTTTCCCTTATGAACATG	Oligo 3' de clonage de yHit1p(1-46) avec site BamHI
A252	GGGGGATCCTTATTTGAGAGTCTCTTCCAGCGTCTTC	Oligo 3' de clonage de Bcd1p(x-331) avec site BamHI
A563	GGGGGATCCTTAGTCAGCCTCTGTTCATGGCAACGAA	Oligo 3' de clonage de Bcd1p(x-292) avec site BamHI
A564	GGGGGATCCTTATTTCCCTGGTGTAGCACCTCGTAGCC	Oligo 3' de clonage de Bcd1p(x-303) avec site BamHI
A580	GGATATCCATATGAGGAGGGGGGTGACCTGTTTGATG	Oligo 5' de clonage de Bcd1p(130-x) avec site BamHI
A611	GGATATCCATATGAGGAGGGGGGTGAACTGTTTGATG	Oligo 5' de clonage de Bcd1p(130-x) avec site NdeI
A684	GGATATCCATATGGAGTGTGAGAGGATAAATTAGGAGG	Oligo 5' de clonage de Bcd1p(124-x) avec site NdeI
A685	GGGGGATCCTTAGCCCTCTATGCAAGAGCTTGTTC	Oligo 3' de clonage de Bcd1p(x-220) avec site BamHI
A686	GGATATCCATATGGAGGAAAGAACACAGATCCTACAG	Oligo 5' de clonage de Bcd1p(228-x) avec site NdeI
A687	GGGGGATCCTTAAATGTCAATAAACTTGTTCAGTGT	Oligo 3' de clonage de Bcd1p(x-319) avec site BamHI
A688	GGATATCCATATGAAGGAGTATATATCGAGTGAGCGC	Oligo 5' de clonage de Bcd1p(47-x) avec site NdeI
A703	GGATATCCATATGACCAAGGAAGTTTATGAAGGTGAA	Oligo 5' de clonage de RuvBL1(127-x) avec site NdeI
A704	GGGGGATCCTTATTTCTTTCTTTTGTGCACATCCCC	Oligo 3' de clonage de RuvBL1(x-233) avec site BamHI
A705	GGATATCCATATGACTAAGGAAGTTTATGAGGGTGAA	Oligo 5' de clonage de Rvb1(136-x) avec site NdeI
A706	GGGGGATCCTTATTTCTTTTCTTATGCACTTCGCC	Oligo 3' de clonage de Rvb1(x-242) avec site BamHI
A707	GGATATCCATATGGAGACGGAGATCATCGAAGGGGAG	Oligo 5' de clonage de RuvBL2(134-x) avec site NdeI
A708	GGGGGATCCTTACTCCTTGCCTTCTGGAGCTCCCC	Oligo 3' de clonage de RuvBL2(x-237) avec site BamHI
A709	GGATATCCATATGGAGACAGAATTGATTGAAGGTGAA	Oligo 5' de clonage de Rvb2(132-x) avec site NdeI
A710	GGGGGATCCTTATGTTTTCTTTTTCAGTTCGCC	Oligo 3' de clonage de Rvb2(x-234) avec site BamHI
A777	GGATATCCATATGAAGATTGAGGAGGTGAAGAGCACT	Oligo 5' de clonage de RuvBL1 avec site NdeI
A778	GGGGGATCCTTACTTCACTGACTTATCCTGCTGGTC	Oligo 3' de clonage de RuvBL1 avec site BamHI
A779	GGATATCCATATGGCAACCGTTACAGCCACAACCAAA	Oligo 5' de clonage de RuvBL2 avec site NdeI
A780	GGGGGATCCTTAGGAGGTGTCCATGGTATAGCCTTT	Oligo 3' de clonage de RuvBL2 avec site BamHI
A964	GGGGGATCCTTACCGCTTGTCTTCATTCTAGCATC	Oligo 3' de clonage de Bcd1p(x-96) avec site BamHI

Table 2. Liste des oligonucléotides utilisés au cours du doctorat (partie 1).

n°	Séquence 5' - 3'	Utilisation
B096	TCCAACCCATTGCTGGGTCTTCTGGGATAAACTCATTTCAC	Oligo 3' de mutation de Rtt106(65-320) D206-216 + GSSG
B097	GTGAAATGAGTTTATCCAGAAGACCCAAACGAATGGGTGGA	Oligo 5' de mutation de Rtt106(65-320) D206-216 + GSSG
B373	AATCTATGTCCTATGCAGGAGCTTTTCAAACATGTC	Oligo 3' de mutation de Bcd1p(120-303) D172-177
B374	GACATGTTTGAAAAGCTCCTGCATAGGACATAGAAT	Oligo 5' de mutation de Bcd1p(120-303) D172-177
B375	ACAAGCTCTTGCATAGAGACCAAAGAGGAAAGAACA	Oligo 3' de mutation de Bcd1p(120-303) D220-225
B376	TGTTCTTTCTCTTTGGTCTCTATGCAAGAGCTTGT	Oligo 5' de mutation de Bcd1p(120-303) D220-225
B388	GTAACAGACTTTGAAAAGCCGCTCGAATATATAAGAAAAG	Oligo 5' de mutation de Rtt106(65-320) C185A
B389	CTTCTTTATATATTCGACGGCCTTTTCAAAGTCTGTTAC	Oligo 3' de mutation de Rtt106(65-320) C185A
B390	CAAGCCATTTTGACGGGTGCTAAAATCTCCAACCCATTG	Oligo 5' de mutation de Rtt106(65-320) F198A
B391	GAATGGGTTGGAGATTTTAGCACCCGTCAAAATGGCTTG	Oligo 3' de mutation de Rtt106(65-320) F198A
B392	CAGCGATCATCGAAAACGCGAGCAAGTGGGACAAAACG	Oligo 5' de mutation de Bcd1p(120-303) R149A
B393	CGTTTTGTCCTACTGCTCGCTTTTGGCATGATCGCTG	Oligo 3' de mutation de Bcd1p(120-303) R149A
B394	GTATGGAGCGTGGAGTGGGCTCTATGCTCTATGCAGGAA	Oligo 5' de mutation de Bcd1p(120-303) I166A
B395	TTCTGTCATAGGACATAGAGCCCACTCCACGCTCCATAC	Oligo 3' de mutation de Bcd1p(120-303) I166A
B396	AAGGAGCTTTTCAAACATGCCAGCCACCGTATCAAAGAG	Oligo 5' de mutation de Bcd1p(120-303) V183A
B397	CTCTTTGATACGGTGGCTGGCATGTTTGAAGAGTCCCTT	Oligo 3' de mutation de Bcd1p(120-303) V183A
B398	ACAGTGATCGAGTTTCCCGCATATTCGTTGCCATGACA	Oligo 5' de mutation de Bcd1p(120-303) T283A
B399	TGTCATGGCAACGAATATCGCGGAAACTCGATCACTGT	Oligo 3' de mutation de Bcd1p(120-303) T283A
B448	TCGGAAACCAATACTATTGCCAATTAGAAGGTGTTTCT	Oligo 5' de mutation de Rtt106(65-320) F74A
B449	AGAAACACCTTCTAATTTGGCAATAGTATTTGGTTTCCGA	Oligo 3' de mutation de Rtt106(65-320) F74A
B450	AGAAAAAGCTTGATTTGGCGTCTATCTTTCCAATGTT	Oligo 5' de mutation de Rtt106(65-320) V92A
B451	AACATTTGAAAAGATAGAAGCCCAAATCAAGCTTTTCTCT	Oligo 3' de mutation de Rtt106(65-320) V92A
B616	GGGGATCCTTATCGCTTCAACTGCGTCAGATAGTT	Oligo 3' de clonage de Bcd1p(x-81) avec site BamHI
B617	GGATATCCATATGGAATTTGCAGCCGAAAATGAAGGT	Oligo 5' de clonage de hBCD1opt(1-x) avec site NdeI
B618	GGGGATCCTTAGTTTTCTGTTGCCACATTTTTGGT	Oligo 3' de clonage de hBCD1opt(x-470) avec site BamHI
B619	GGATATCCATATGGCAATGAGCCGTTGTGAAACCTGT	Oligo 5' de clonage de hBCD1opt(216-x) avec site NdeI
B620	GGGGATCCTTAAATATATTTCTGCTGGCTCTGCGG	Oligo 3' de clonage de hBCD1opt(x-360) avec site BamHI
B942	GGATATCCATATGCAGCGCCCTTCTGTGCTGTCTGT	Oligo 5' de clonage de l'ORF de ZNHIT1(113-x) avec site NdeI
B943	GGGGATCCTTACACAGTCCACTTCAGACACCTGGT	Oligo 3' de clonage de l'ORF de ZNHIT1(x-154) avec site BamHI
B944	GGGGATCCTTAGTAGAAGTTTTCTGCGCAGGAGCC	Oligo 3' de clonage de l'ORF de ZNHIT2(x-46) avec site BamHI
B945	GGGGATCCTTAAGTTTCAAGGTTGCACTGTTCTTT	Oligo 3' de clonage de l'ORF de ZNHIT3(x-46) avec site BamHI
B958	GGATATCCATATGTCAGGCCCGCTGCTGTGCCCGGCTGT	Oligo 5' de clonage de l'ORF de ZNHIT4(307-x) avec site NdeI
B946	GGGGATCCTTACGTAGCCAAAAGGGGGATCCAGG	Oligo 3' de clonage de l'ORF de ZNHIT4(x-356) avec site BamHI
B947	GGATATCCATATGCAGCGCTGGGCAGAACAGGGGAA	Oligo 5' de clonage de l'ORF de ZNHIT5(98-x) avec site NdeI
B948	GGGGATCCTTATGATTTCTCTTCTTTTCTTTAAC	Oligo 3' de clonage de l'ORF de ZNHIT5(x-146) avec site BamHI
B978	GGATATCCATATGAGTTATAATGAGCAGGAGATTAAG	Oligo 5' de clonage de l'ORF DIExt de Rvb2(358-x) avec site NdeI
B979	GGATATCCATATGGACAGCGCTAGATCTGTTAAGTAT	Oligo 5' de clonage de l'ORF DIExt de Rvb2(432-x) avec site NdeI
B980	GGATATCCATATGAGCATTCAGACCAGCGATCCGAAT	Oligo 5' de clonage de l'ORF de Rvb2opt (N-ter) avec site NdeI
B981	GGGGATCCTTATTCGGTGGTATCCATTGCATCCGG	Oligo 3' de clonage de l'ORF de Rvb2opt (C-ter) avec site BamHI
B982	GGATATCCATATGAGCTACAACAGCAAGAGATCAA	Oligo 5' de clonage de l'ORF de Rvb2opt DIII (358-x) avec site NdeI
B983	GGATATCCATATGGATAGCGCAGCTAGCGTTAAATAT	Oligo 5' de clonage de l'ORF de Rvb2opt a-hook (432-x) avec site NdeI
B984	GGATATCCATATGGCAACCGTTACCAGCAACCACCAA	Oligo 5' de clonage de l'ORF de RuvBL2opt (N-ter) avec site NdeI
B985	GGGGATCCTTAGCTGGTATCCATGGTTTCACTTT	Oligo 3' de clonage de l'ORF de RuvBL2opt (C-ter) avec site BamHI
B986	GGATATCCATATGTATAGCGAAAAGATACCAAACA	Oligo 5' de clonage de l'ORF de RuvBL2opt DIII (362-x) avec site NdeI
B987	GGATATCCATATGGATGAAAGCGGTAGCACCCAGTAT	Oligo 5' de clonage de l'ORF de RuvBL2opt a-hook (435-x) avec site NdeI
B988	GGATATCCATATGAAAATCGAAGAGGTGAAAAGCACC	Oligo 5' de clonage de l'ORF de RuvBL1opt (N-ter) avec site NdeI
B989	GGGGATCCTTATTTTCATGATTTTATCTCTGCTGATC	Oligo 3' de clonage de l'ORF de RuvBL1opt (C-ter) avec site BamHI
	CTAGTTATGCTCAGCGG	pET-RP
	TAATACGACTCACTATAGGG	T7
	dIdCdIdCdIdCdIdCdIdCdIdCdIdCdIdCdIdC	Poly-dIdC

Table 3. Liste des oligonucléotides utilisés au cours du doctorat (partie 2).

SOB (*Super Optimal Broth*), soit de tryptone (20 g/L), d'extrait de levure (5 g/L), de NaCl (0,5 g/L), de KCl (2,5 mM), de MgCl₂ (10 mM) et de MgSO₄ (10 mM) auquel est ajouté du glucose (20 mM).

Le milieu riche graffinity est utilisé lors des cribles de co-expression à haut débit. Ce milieu permet l'automatisation de l'induction ainsi qu'une synchronisation. Il s'utilise en deux parties. Dans un premier temps, la culture est réalisée dans un milieu 2LB (composition du milieu LB avec les trois composants au double de leur quantité en LB) supplémenté de glucose (0,5%). Le volume de ce premier milieu correspond à la moitié du volume final de la culture. Durant cette première partie de la culture, les cellules croissent normalement en présence de glucose. Dans un second temps (lorsque la DO atteint environ 1,2) est ajouté le même volume de milieu 2LB supplémenté de lactose (0,6%), d'IPTG (1 mM final) et d'HEPES à pH 7 (20 mM final). Le fait de diluer le volume total par 2 ramène la DO autour de 0,6 (phase exponentielle). Le second milieu contient du lactose (et non du glucose), qui sera incorporé par les cellules une fois le glucose épuisé. A partir de ce moment, le lactose permettra l'induction de l'expression de la protéine d'intérêt. La présence d'IPTG dans ce second milieu permet en plus de s'assurer de la bonne surexpression de la protéine.

Le milieu minimum M9 (utilisé pour la RMN) ne contient que des sels de Na₂HPO₄ (H₂O)₁₂ (15 g/L), de KH₂PO₄ (3 g/L) et de NaCl (0,5 g/L). Ce milieu permet de contrôler les sources de carbone et d'azote et au besoin d'utiliser des isotopes marqués uniformément au carbone 13 (¹³C) et/ou à l'azote 15 (¹⁵N). La culture se fait alors dans 950 mL de milieu M9 à pH 7,2 (ensemencés avec 50 mL de pré-culture) auquel sont ajoutés du MgSO₄ (2 mM final), des vitamines (mélange à 1 000X), du mélange Métaux Mix I (2‰), du mélange Métaux Mix II (1‰), du NH₄Cl (marqué ou non au ¹⁵N à 1 g/L final) et du glucose (marqué ou non au ¹³C à 2 g/L final).

Le mélange de **vitamines** contient de la Biotine (1 mg/mL), de la Nicotinamide (1 mg/mL), de la Riboflavine (0,1 mg/mL) et de la Thiamine (5 mg/mL).

Le mélange **Métaux Mix I** contient du FeCl₃ (50 mM).

Le mélange **Métaux Mix II** contient du CaCl₂ (20 mM), du MnCl₂ (10 mM), du ZnSO₄ (10 mM), du CoCl₂ (2 mM), du CuCl₂ (2 mM), du NiCl₂ (2 mM), du Na₂MoO₄ (2 mM), du Na₂SeO₃ (2 mM) et du H₃BO₃ (2 mM).

Le milieu minimum M9 (pour le marquage avec Se-Met) est utilisé pour la surexpression de protéines recombinantes sélénées en vue d'une étude structurale par cristallographie aux rayons X (lorsqu'il n'y a pas d'homologie structurale connue). Ce milieu ne contient que les sels de KH₂PO₄ (6 g/L) et de Na₂HPO₄ (H₂O)₁₂ (35,8 g/L). Il permet l'ajout

un à un les 20 acides aminés nécessaires à la croissance des bactéries en remplaçant la méthionine par de la séléno-méthionine. La culture se fait ensuite dans 900 mL de milieu M9 à pH 7,4 (ensemencés avec 100 mL de pré-culture) auquel sont ajoutés du NH_4Cl (2 g/L), du glucose (4 g/L), du MgSO_4 (2 mM), des vitamines (mélange à 1 000X), du mélange Métaux Mix I (2‰), du mélange Métaux Mix II (1‰), du mélange Acide Aminé Mix I (1%, soit 40 mg/L de chaque acide aminé), du mélange Acide Aminé Mix II (1,1% soit 44 mg/L de chaque acide aminé) et de la séléno-méthionine (40 mg/L).

Le mélange **Acide Aminé Mix I** contient les acides aminés Ala, Arg-HCl, Asp, Cys, Glu, Gln, Gly, His, Leu, Ile, Lys-HCl, Pro, Ser, Thr et Val (4 mg/mL).

Le mélange **Acide Aminé Mix II** contient les acides aminés aromatiques Tyr, Trp et Phe (4 mg/mL). Le pH est ajusté à 8 avec du NaOH.

1.4. Les antibiotiques

L'ampiciline est utilisée à 100 mg/L final.

Le chloramphénicol est utilisé à 25 mg/L final.

La kanamycine est utilisée à 50 mg/L final.

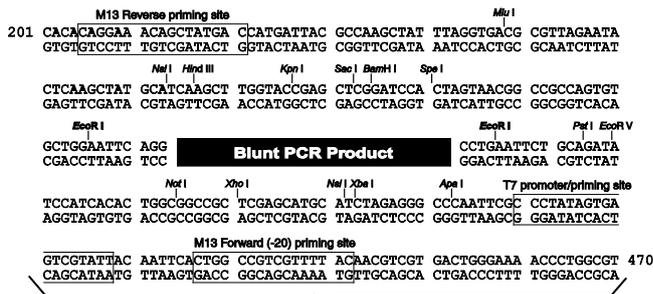
La spectinomycine est utilisée à 30 mg/L final.

1.5. Les plasmides bactériens

Le vecteur **pCR[®]-Blunt** (Invitrogen[™]) est souvent utilisé dans la première étape de clonage, juste après la réaction de PCR. Il a la particularité d'être commercialisé sous forme linéaire et possède des extrémités franches (Figure 33). Ainsi, il est possible d'y insérer un produit amplifié par PCR sans avoir à le digérer au préalable par des enzymes de restriction. Ce vecteur possède des gènes de résistance à la kanamycine et à la zéocine, ainsi que la séquence *lacZ α* codant le peptide α de la β -galactosidase placée sous contrôle d'un promoteur *Plac*. Le promoteur *Plac* est séparé de la séquence *lacZ α* par la cassette de clonage.

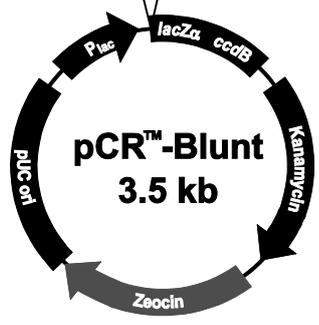
Le plasmide **pRARE2** (Figure 33) permet d'améliorer la traduction des ARNm issus d'organismes eucaryotes en protéines par les souches bactériennes. Il contient des séquences codant 7 ARNt peu utilisés par la bactérie contrairement aux eucaryotes. Ces ARNt reconnaissent les codons AUA (Ile), AGG (Arg), AGA (Arg), CUA (Leu), CCC (Pro), CGG (Arg) et GGA (Gly). Ce plasmide possède un gène de résistance au chloramphénicol.

A



Ce plasmide pCR-Blunt contient :

- Le gène *lacZα*.
- Le promoteur *P_{lac}*.
- Des gènes de résistance à la kanamycine et à la zéocine.
- Une origine de réplication pUC ori.



B

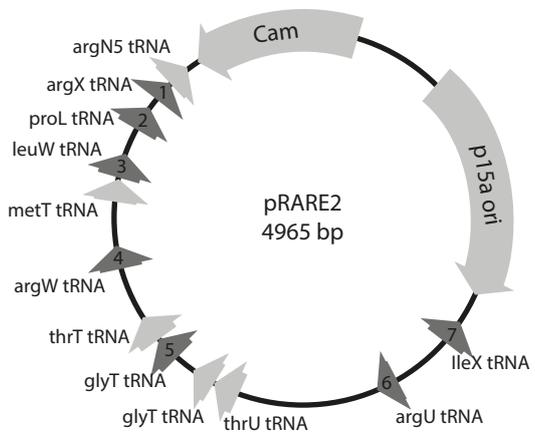


Figure 33. Carte des plasmides (1). (A) Schéma et carte de restriction du plasmide commercial pCR-Blunt (Invitrogen). (B) Le plasmide pRARE2 contient les séquences codantes des ARNt sous représentés chez les bactéries (Arg, Pro, Leu, Gly et Ile), pour la surexpression de protéines recombinantes eucaryotes.

Les vecteurs **pnEA-3CH**, **pnCS** et **pnYK** (Diebold *et al.*, 2011) sont utilisés pour la surexpression de protéines recombinantes chez *E. coli* dans le cadre de la stratégie de co-expression (Figure 34). Ces trois vecteurs possèdent des gènes de résistance différents, l'ampicilline pour le pnEA-3CH, la spectinomycine pour le pnCS et la kanamycine pour le pnYK ainsi que des origines de répliation compatibles. De plus, le vecteur pnEA-3CH présente une séquence codant une étiquette 6xHistidines en amont de la cassette de clonage, ce qui se traduit par la présence de cette étiquette à l'extrémité N-terminale de la protéine surexprimée. Une séquence codant le site de coupure à la PreScission (3C) est également présente entre l'étiquette et la cassette de clonage. Les pnCS et pnYK sont sans étiquette avec une traduction qui commence à une méthionine. La cassette de clonage de ces trois vecteurs est placée sous le contrôle du promoteur du bactériophage T7 inductible à l'IPTG.

1.6. Les tampons généraux

Les différents tampons et solutions utilisés pour les travaux sont listés ci-après.

Bleu de dépôt sur SDS-PAGE : Tris-HCl à pH 6,8 (62,5 mM), SDS (2%), β -mercaptoéthanol (100 μ M), Glycérol (25%), Bleu de bromophénol (0,01% p/v).

Bleu de formamide : EDTA (20 mM), Bleu de bromophénol (0,05% p/v) et Bleu de xylène cyanol (0,05% p/v) dans du formamide désionisé.

Bleu natif de dépôt sur gel d'agarose (6X) : Glycérol (30%), Bleu de bromophénol (0,25% p/v) et xylène cyanol (0,25% p/v).

Bleu natif CSB de dépôt pour retard sur gel : HEPES-KOH à pH 7,9 (20 mM), Glycérol (40%), Bleu de bromophénol (0,05% p/v) et Bleu de xylène cyanol (0,03% p/v).

Marqueur de poids moléculaire (GE Healthcare) : Phosphorylase B de muscle de lapin (96 kDa), Albumine de sérum bovin (66 kDa), Ovalbumine d'œuf de poulet (45 kDa), Anhydrase carbonique d'érythrocyte bovin (30 kDa), Inhibiteur de trypsine (20 kDa) et α -lactalbumine (14,4 kDa).

Solution stock de bleu de coloration pour SDS-PAGE : Ethanol (20%), Acide acétique (10%), G250 (0,5 g/L) et R250 (0,5 g/L).

Solution de coloration pour SDS-PAGE : Ethanol (5%), Acide acétique (7,5%) et Bleu de coloration (environ 200 μ L).

Solution de déshydratation pour SDS-PAGE : Ethanol (50%) et Acide acétique (10%).

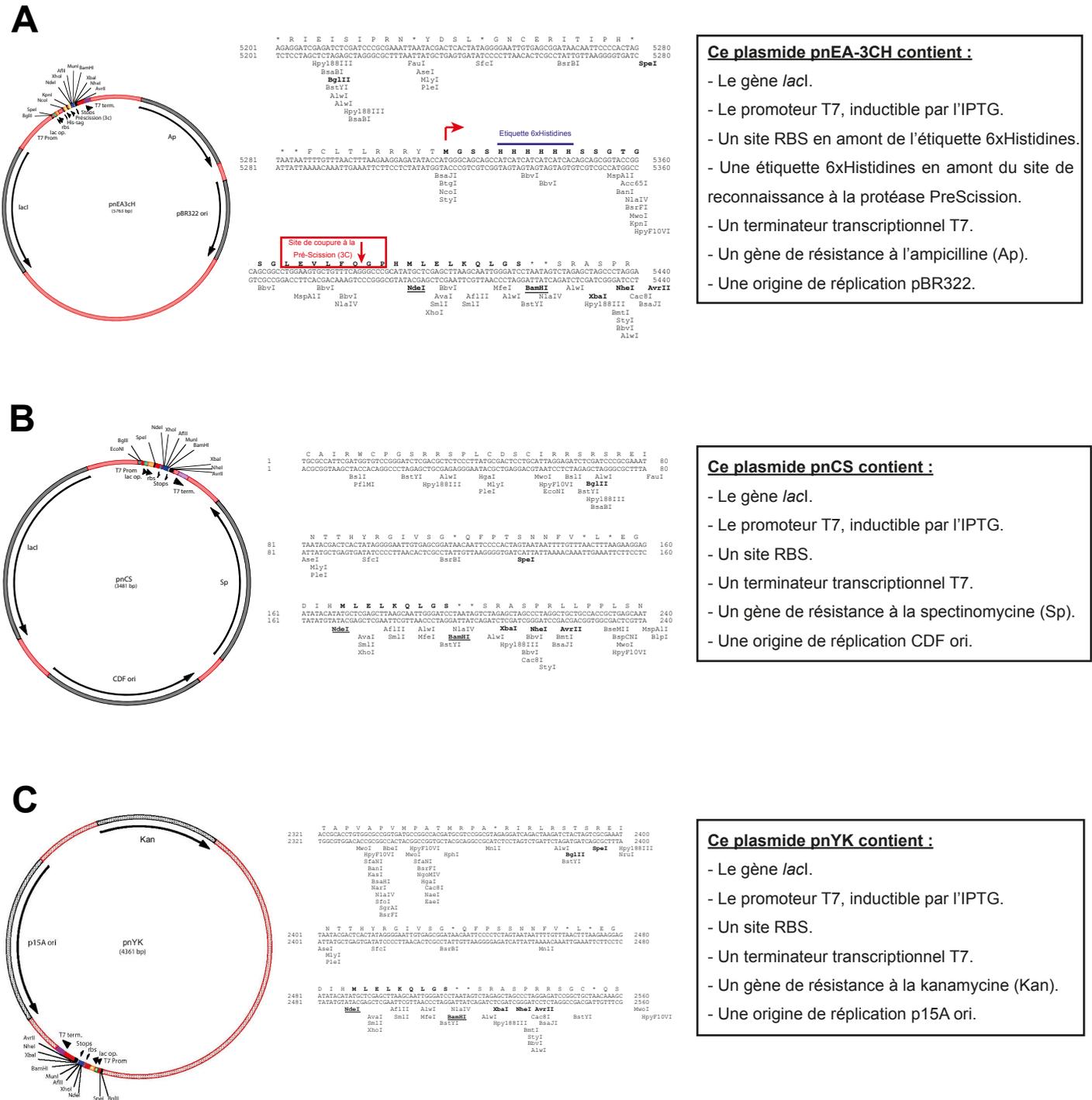


Figure 34. Carte des plasmides (2). Représentation schématique des plasmides (A) pNEA-3CH, (B) pncS et (C) pnYK, ainsi que leurs cassettes de clonage. Le plasmide pNEA-3CH contient un site de coupure spécifique de la protéase PreScission (3C) (site de clivage indiqué par la flèche rouge sur la cassette) et une étiquette 6xHistidines (en bleu). La méthionine initiatrice et la séquence en acide aminé traduits sont représentées en gras. Les sites de restriction utilisés pour le clonage de l'ORF (*NdeI* et *BamHI*) sont soulignés en gras ceux utilisés pour la concaténation (*BgIII*, *SpeI*, *XbaI*, *NheI* et *AvrII*) sont en gras. Les étoiles représentent des codons stop (d'après Diebold *et al.*, 2011).

TAE (1X) : Tris (40 mM), Acide acétique (20 mM) et EDTA (1 mM).

Tampon D : HEPES-KOH à pH 7,9 (20 mM), KCl (150 mM), MgCl₂ (1,5 mM).

Tampon d'éluion ARN : Tris-HCl à pH 7,5 (10 mM), NaCl (300 mM), EDTA (1 mM) et SDS (1%).

Tampon de migration pour gel dénaturant (ARN) : TBE (0,5X).

Tampon de migration pour gel retard : TBE ou TB 20X (2,5%) et Glycérol (5%).

Tampon de migration pour SDS-PAGE : Tris (3,09 g/L), Glycine (14,4 g/L) et SDS 20% (5%).

Tampon de transcription (3X) : Tris-HCl à pH 8 (120 mM), DTT (15 mM), Spermidine (3 mM), Triton (0,03%), PEG 8 000 (240 mg/mL) et MgCl₂ (60 mM).

TBE (1X): Tris (100 mM), Acide borique (89 mM) et EDTA (1 mM).

TB (1X): Tris (100 mM) et Acide borique (89 mM).

1.7. Les tampons utilisés lors de l'étude des protéines recombinantes

Le tampon de lyse bactérienne : HEPES ou Tris à pH 7,5-8 (25 mM), NaCl (150 à 300 mM), Imidazole (10 à 40 mM), TCEP (0,5 mM).

Le tampon d'éluion : HEPES ou Tris à pH 7,5-8 (25 mM), NaCl (150 à 300 mM), Imidazole (300 mM), TCEP (0,5 mM).

Le tampon de dialyse : HEPES ou Tris à pH 7,5-8 (20 mM), NaCl (150 à 300 mM), DTT (1 à 2 mM).

Le tampon de chromatographie d'exclusion de taille : HEPES, Tris ou Phosphate (NaPi) à pH 6,4-8 (20 mM), NaCl (0 à 300 mM), DTT (1 à 5 mM) et/ou TCEP (0,5 mM).

Le tampon bas sel de lyse bactérienne (crible de co-expression) : HEPES à pH 7,5 (25 mM), NaCl (50 mM), Imidazole (10 mM).

Le tampon haut sel de lyse bactérienne (crible de co-expression) : HEPES à pH 7,5 (25 mM), NaCl (400 mM), Imidazole (10 mM).

Le tampon de RuvBL1 : Tris-HCl pH 8 (20 mM), NaCl (150 mM), Glycérol (10%), MgCl₂ (2 mM), ADP (200 µM) et β-mercaptoéthanol (4 mM).

Le tampon de RuvBL2 : NaKPi pH 7,7 (50 mM), NaCl (300 mM), Glycérol (10%), MgCl₂ (3 mM), ADP (300 µM) et TCEP (0,5 mM).

Le tampon de RuvBL1:2 : Tris-HCl pH 8 (20 mM), NaCl (250 mM), Glycérol (10%), MgCl₂ (5 mM), ADP (300 µM) et TCEP (0,5 mM).

1.8. Les différents types de gels

Gel d'acrylamide pour retard sur gel : Acrylamide (6 ou 8%) (à partir d'un stock d'Acrylamide/Bisacrylamide 19:1 à 40%), TBE (ou TB) 0,5X, Glycérol (20%), APS et TEMED. 40 mL par gel.

Gel dénaturant (6%) pour la purification des ARN : Acrylamide (6%) (à partir d'un stock d'Acrylamide/Urée à 25%), TBE 1X, APS et TEMED et complément en urée (stock 8 M). 20 mL par gel.

SDS-PAGE 12,5% : Acrylamide 12,5% (à partir d'un stock d'Acrylamide/Bisacrylamide 29:1 à 40%), Tris-HCl à pH 8,8 (375 mM), SDS (0,1%), APS et TEMED. 4 mL par gel.

SDS-PAGE *stacking* : Acrylamide 5% (à partir d'un stock d'Acrylamide/Bisacrylamide 29:1 à 40%), Tris-HCl à pH 6,8 (50 mM), SDS (0,1%), APS et TEMED. 2 mL par gel.

1.9. Les peptides

Deux peptides de synthèse correspondant à de petites régions des protéines d'intérêts ont été commandés chez Proteogenix (Strasbourg). Ces peptides ont été utilisés pour tester des interactions protéine-protéine, dans le cadre d'expériences d'ITC ou de titration par RMN.

Le peptide **Rtt106M α 2** correspond à l'hélice α 2 (161-173) de la protéine Rtt106(65-320), soit à la séquence NKENTLNQFKNLG (pI de 8,59 et PM de 1519,6 g/mol⁻¹). La pureté de ce peptide est de 96,25%.

Le peptide **H3K56ac** correspond à la séquence de l'histone H3 de levure pouvant interagir avec la protéine Rtt106, soit la séquence IRRFQKacSTELL (pI de 10,84 et PM de 1390,6 g/mol⁻¹). Cette séquence contient une modification sur la lysine 56 (acétylation) permettant d'augmenter son affinité pour la protéine Rtt106. La pureté de ce peptide est de 96,01%.

2. Méthodes

2.1. Le génie génétique

Les techniques de génie génétique comprennent ici toutes les étapes nécessaires à la création de constructions de vecteurs dédiés à la production de protéines recombinantes et à la préparation d'échantillons d'acides nucléiques radiomarqués.

2.1.1. *Clonage*

Le but des étapes de clonage est d'insérer la séquence codant la protéine d'intérêt dans un vecteur d'expression et d'amplifier ce vecteur.

2.1.1.1. *La PCR (Polymerase Chain Reaction)*

Dans un premier temps, la séquence codante doit être isolée et amplifiée avec l'éventuel ajout de séquences correspondant à des sites de restriction compatibles avec le vecteur de destination en amont (5') et en aval (3'). Pour cela, la technique utilisée est la PCR, qui consiste en l'amplification d'une séquence à partir d'amorces oligodésoxyribonucléotidiques. Ces amorces contiennent les séquences des sites de restriction ainsi qu'une partie de la séquence à amplifier, sur laquelle elle viendra s'hybrider. L'une correspond au début de la séquence (extrémité 5') et l'autre à la fin de la séquence (extrémité 3'). Une amorce 5' doit pouvoir s'hybrider sur le brin sens de l'ADN et une amorce 3' sur le brin anti-sens. Le principe de la PCR repose sur trois grandes étapes, la dénaturation, l'hybridation et l'élongation. L'étape de dénaturation sert à séparer les deux brins de la matrice ADN par l'action d'une température élevée (98°C). Puis, l'étape d'hybridation permet aux amorces de se lier à la séquence complémentaire au sein de la matrice. Enfin, l'étape d'élongation se déroule à une température optimale au fonctionnement de la polymérase utilisée (72°C), et permet à celle-ci de construire le brin complémentaire à la séquence de la matrice en partant de l'amorce. Les polymérases utilisées pour les PCR proviennent d'organismes thermophiles et sont résistantes aux fortes températures. Elles sont donc capables de résister aux températures requises par la PCR. Les appareils utilisés au laboratoire pour les PCR sont des thermocycleurs : 2720 Thermocycler (Applied Biosystems) et SureCycler 8800 (Agilent Technologies).

Lors de l'étape de PCR, la matrice (séquence ADN contenant la séquence à amplifier) est utilisée à 1 ng/μL final et les amorces à 1 μM final. Les nucléotides utilisés pour la

polymérisation sont apportés sous forme d'un mélange des quatre dNTP (200 nM final). Le milieu réactionnel contient également 10% de DMSO, ce qui permet d'améliorer l'amplification notamment en cas de pourcentage élevé en bases G-C. Deux enzymes différentes ont été utilisées, la Phusion Hot Start II (ThermoFisher Scientific) utilisée à 8 mU/μL avec le tampon HF 5X et la Drem Taq (ThermoFisher Scientific) utilisée à 25 mU/μL avec le tampon correspondant (*green buffer*) à 10X pour les PCR sur colonies (polymérase de plus faible fidélité par rapport à la Phusion Hot Start II).

La PCR est donc constituée de trois étapes à des températures différentes. La température de l'étape d'hybridation dépend des T_m (température de fusion) des amorces et est généralement comprise entre 50°C et 80°C. Ces trois étapes sont répétées 35 fois pour une PCR classique et 25 fois pour une PCR sur colonies. Ce cycle est précédé d'une première dénaturation de 5 à 10 min à 98°C et est suivi d'une étape de 5 à 10 min à 72°C. Les étapes de dénaturation et d'hybridation durent 30 s (le temps de l'étape d'hybridation peut être augmenté si la matrice est de l'ADN génomique). La durée de l'étape d'élongation dépend elle de la taille de la séquence à amplifier (1 min/kb).

Les amplifications par PCR sont ensuite visualisées sur gel agarose. La migration est faite en tampon TBE ou TAE 1X 45 min à 120 V.

Dans le cas d'une construction dans le pCR[®]-Blunt, le produit PCR est purifié avec le kit NucleoSpin[®] Gel and PCR Clean-up (Machery-Nagel). Il est ensuite dosé à 260 nm avec un NanoDrop 2000c (ThermoFischer Scientific). Le produit PCR est ensuite inséré dans le vecteur pCR[®]-Blunt par une étape de ligation.

2.1.1.2. La digestion par des enzymes de restriction

Il est possible de passer outre cette étape dans le pCR[®]-Blunt. Dans ce cas le produit PCR est digéré par des enzymes de restriction (*NdeI* et *BamHI* pour les vecteurs pnEA-3CH, pnCS et pnYK). Les deux enzymes de restriction sont des FastDigest (ThermoFischer Scientific) et sont utilisées à 0,5X en présence de tampon FastDigest 10X. L'enzyme *DpnI* FastDigest est ajoutée au milieu réactionnel (0,25X). Cette enzyme clive l'ADN sur des sites [GA^{m6}↓TC]. Elle est donc spécifique d'une forme de méthylation retrouvée sur l'ADN produit par les bactéries. Comme la PCR est réalisée *in vitro*, les produits obtenus ne sont pas méthylés et ne seront donc pas touchés par la *DpnI*. La réaction se fait pendant 20 min à 37°C. Une fois la réaction terminée, le produit est purifié avec le kit NucleoSpin[®] Gel and PCR Clean-up (Machery-Nagel). Il est ensuite dosé à 260 nm avec un NanoDrop 2000c

(ThermoFischer Scientific). Le produit PCR digéré est ensuite inséré dans le vecteur de destination, préalablement digéré lui aussi avec les mêmes enzymes de restriction, par une étape de ligature.

2.1.1.3. Le sous-clonage

Dans le cas d'une construction déjà existante à cloner dans un nouveau vecteur possédant les mêmes sites de restriction, une étape de sous-clonage est suffisante. Pour cela, la construction existante est digérée par les enzymes de restriction pendant 20 min à 37°C. Puis l'insert est purifié et dosé avant d'être inséré dans le nouveau vecteur.

2.1.1.4. Préparation de vecteur linéarisé

Le vecteur vide est d'abord digéré par les enzymes de restriction (10 µg de vecteur dans un volume de 100 µL 10 min à 37°C). Il se retrouve alors linéarisé avec des extrémités cohésives, à l'inverse du pCR[®]-Blunt. L'enzyme *FastAP* (phosphatase) est ensuite ajoutée au volume réactionnel (10 µL) qui est à nouveau incubé 10 min à 37°C, de manière à déphosphoryler les extrémités du vecteur linéarisé, afin d'éviter une recircularisation.

Ce volume réactionnel est supplémenté par un volume d'eau et deux volumes de phénol/chloroforme (phase inférieure). Le mélange est vortexé pendant 1 min puis centrifugé 5 min à 16 000 g à température ambiante. Au surnageant est ajouté 1/10 (v/v) d'acétate de sodium 3 M. Trois volumes d'éthanol à 100% sont ensuite ajoutés, puis le tout est placé à -80°C pendant 15 min et centrifugé à 16 000 g 30 min à froid. Le surnageant est éliminé et 100 µL d'éthanol à 70% sont ajoutés pour laver le culot. Après une centrifugation de 5 min à 16 000 g à froid, le surnageant est à nouveau éliminé et le culot est laissé à sécher à température ambiante. Le culot est ensuite repris dans de l'eau et dosé au NanoDrop.

2.1.1.5. La ligature

Une fois l'insert purifié, il est inséré dans le vecteur de destination. Pour cela, il est placé en présence de vecteur linéarisé (5 ng/µL final) à un ratio de un pour trois (insert/vecteur) et d'une enzyme, la T4 DNA Ligase (0,5 U/µL final) avec son tampon (10X). La réaction se fait à température ambiante pendant 2 h. Le volume réactionnel est ensuite utilisé pour transformer des bactéries DH5α. La présence du plasmide dans les bactéries est vérifiée par PCR sur colonies avec des amorces spécifiques de la construction (dans le cas

d'un clonage dans le pCR[®]-Blunt) ou avec les amorces universelles T7 et pET-RP (dans le cas des autres vecteurs). Les séquences T7 et pET-RP sont présentes de part et d'autre de la cassette de clonage des vecteurs d'expression et permettent donc l'amplification de toute la cassette avec la séquence insérée. Ceci permet de limiter les faux négatifs, car l'amplification doit toujours avoir lieu et la discrimination entre les conditions avec ou sans insert se fait en fonction de la taille des bandes sur gel agarose.

2.1.1.6. Transformation de bactéries

La transformation de bactéries compétentes permet l'incorporation d'un (ou plusieurs) vecteur(s) dans une bactérie. Pour cela, le vecteur est placé au contact de bactéries (50 µL) à froid pendant 20 min, de façon à ce que celui-ci adhère à la paroi des bactéries. Un choc thermique est alors réalisé à 42°C pendant 40 s, puis le milieu est repassé à froid pendant 5 min. Ce choc thermique génère des mouvements de diffusion dans la paroi lipoprotéique des bactéries et permet l'entrée du vecteur. Les bactéries sont ensuite régénérées pendant une heure avec du milieu riche SOC (100 µL) à 37°C et étalées sur boîtes de pétri stérilement. Les bactéries sont ensuite incubées une nuit dans une étuve à 37°C.

2.1.1.7. Amplification de vecteurs

Les bactéries DH5α permettent la multiplication des plasmides. Pour amplifier ces vecteurs, une colonie de bactéries transformées est mise en culture dans 4 mL de milieu 2TY pendant une journée à 37°C sous agitation. Ensuite, les bactéries sont centrifugées à 4 000 g pendant 10 min et le surnageant est éliminé. Puis, l'ADN plasmidique est extrait et purifié en suivant le protocole du kit de Mini-préparation *NucleoSpin[®] 8 plasmid* (Machery-Nagel) avec un robot pipeteur automatisé Tecan[®] *freedom evo 200*. L'ADN plasmidique des mini-préparations est ensuite séquencé par la société GATC Biotech. Les vecteurs peuvent ensuite être amplifiés et purifiés avec un protocole de midi-préparation, ce qui permet d'obtenir de plus grandes quantités avec une plus grande pureté.

2.1.2. Mutagenèse dirigée par PCR

La réalisation de mutants des protéines étudiées se fait par mutagenèse dirigée. Cette technique permet de modifier spécifiquement un (ou plusieurs) nucléotide(s) dans la séquence

codant la protéine. Deux méthodes différentes basées sur la PCR peuvent être utilisées, le *Rolling-circle* ou l'*Overlap-PCR*.

2.1.2.1. La mutagenèse dirigée par *Rolling-circle*

La technique du *Rolling-circle* permet d'échanger, ajouter ou retirer un nucléotide ou une petite séquence directement sur un vecteur. Pour cela, la mutation à réaliser est apportée par les amorces de la PCR. Les deux amorces doivent être complémentaires l'une de l'autre et des séquences du vecteur encadrant le site à muter. Le protocole de mutagenèse est le même que celui d'une PCR classique avec une quantité d'amorce utilisée plus faible (100 nM) pour limiter l'hybridation des amorces entre-elles. Le DMSO est utilisé à 5% et le nombre de cycle de PCR est placé à 30. De plus, comme le vecteur entier est amplifié la concentration de dNTP est augmentée à 250 nM final et la durée d'élongation est augmentée en fonction de la longueur du vecteur. Cette technique permet d'obtenir un mutant directement dans le vecteur d'expression sans passer par une étape de ligature.

2.1.2.2. La mutagenèse dirigée par *Overlap-PCR*

La mutagenèse par *Overlap-PCR* est principalement utilisée pour les ajouts ou délétions de séquence ou la création de protéines chimériques issues de deux protéines différentes. Cette méthode nécessite trois PCR distinctes et quatre amorces. Deux amorces en partie complémentaires entre elles apportent la modification (amorces 2 et 3) et les deux autres amorces s'hybrident sur les extrémités 5' et 3' de la séquence et apportent les sites de restriction (amorces 1 et 4). Dans un premier temps, deux PCR sont réalisées avec les amorces 1 et 2 d'une part et 3 et 4 d'autre part. Ces deux PCR vont amplifier chacune une partie de la séquence totale jusqu'au site modifié. Enfin, une troisième PCR avec les amorces 1 et 4 permet d'obtenir la séquence totale. Cette dernière PCR utilise comme matrice un mélange des produits des deux précédentes PCR à 0,5 nM final.

2.1.3. Transcription *in vitro*

La transcription *in vitro* consiste en la production d'un ARN à partir d'une matrice ADN (produite et amplifiée par PCR). Pour que la réaction de transcription soit possible, la matrice ADN linéaire doit posséder en amont de la séquence codant l'ARN un site promoteur T7 et la séquence ARN doit commencer par au moins de G. Cette matrice est ensuite mise en présence de l'ARN polymérase issue du phage T7 (produite au laboratoire) correspondant au

promoteur, de tampon de transcription 3X et de rNTP (4,16 mM final) et incubée une nuit à 37°C. Une fois la transcription terminée, l'échantillon est traité à la DNase *RQI* pendant 45 min à 37°C de façon à éliminer la matrice ADN. L'ARN est ensuite purifié sur gel de polyacrylamide (6%) dénaturant. Pour cela l'échantillon est supplémenté de bleu de formamide, puis les ARN sont dénaturés à 96°C pendant 2 min avant d'être déposés sur gel. La migration se fait pendant 1 h à 30 mA (constant). La bande correspondant à l'ARN est ensuite visualisée par ombrage aux UV (254 nm) et découpée. L'ARN est alors élué du gel par action du tampon d'éluion ARN (300 µL) pendant une nuit à froid puis 30 min à 37°C (pour re-solubiliser le SDS). L'ARN est alors purifié avec du phénol/chloroforme et par précipitation alcoolique (sans sels), puis repris dans 20 µL d'eau.

2.1.4. Marquage radioactif

2.1.4.1. Marquage radioactif des ARN

Le marquage radioactif des ARN se fait par ajout d'un groupement phosphate radioactif [γ - $^{32}\text{PO}_4$] à l'extrémité 5' de l'ARN. Pour cela, l'ARN doit être au préalable déphosphorylé. 1 µg d'ARN est donc déphosphorylé par action de la *SAP* (dans son tampon) 1 h à température ambiante. L'ARN déphosphorylé est par la suite purifié avec du phénol/chloroforme et précipité dans de l'alcool avec de l'acétate de sodium (300 mM final).

Le marquage radioactif est réalisé par action de la kinase *PNK* (dans son tampon) à partir d'ATP radiomarqué sur le phosphate γ [γ - ^{32}P -ATP]. La réaction se fait pendant 1 h à 37°C, puis l'ARN est purifié sur colonne ARN (*mini quick spin columns* – Roche).

2.1.4.2. Marquage radioactif de l'ADN

L'ADN est radiomarqué grâce à l'incorporation du [α - ^{32}P -dTTP] lors de la PCR. Pour cela, le protocole PCR est classique sauf en ce qui concerne les dNTP. Les dATP, dCTP et dGTP sont utilisés à 200 µM. Le dTTP est ajouté sous les formes marquée et non-marquée. Pour cela, le dTTP est utilisé à 20 µM et le [α - ^{32}P -dTTP] à 10 mCi/mL.

2.2. Biochimie des protéines

Cette partie de biochimie traite des aspects de préparation d'échantillons de protéines recombinantes *in vitro* et de leur utilisation pour des expériences d'interaction. Les protocoles décrits ici sont des standards utilisés au cours du travail de thèse.

2.2.1. Production de protéines recombinantes chez *E. coli*

La production de protéines recombinantes se fait en utilisant un système bactérien (temps de génération très court). Le (ou les) vecteur(s) contenant la séquence codant la protéine à surexprimer est (sont) d'abord utilisé(s) pour transformer une souche de bactérie *E. coli* de type BL21 DE3 (culture classique) ou B834 pour les cultures séléniées (utilisées afin de produire des séléno-protéines pour la cristallographie aux rayons X). Une colonie de bactérie transformée est ensuite utilisée pour ensemercer une pré-culture (50 à 100 mL pour 1 L de culture). Les cultures en milieu minimum M9 (RMN ou Se-met) sont d'abord lancées dans 10 mL de milieu riche LB (à 37°C sous agitation) puis une pré-culture de milieu minimum M9 est inoculée avec 1-2 mL. La pré-culture est mise à incuber une nuit (pour atteindre la saturation) à 37°C sous agitation. Cette pré-culture est ensuite utilisée pour ensemercer une culture à raison de 0,1 DO_{600nm}/mL pour les cultures en milieu riche. Dans le cas des cultures en milieu minimum, toute la pré-culture est utilisée pour l'ensemencement (50 mL par litre de culture). Les cultures sont mises à incuber à 37°C sous agitation jusqu'à atteindre une DO_{600nm} comprise entre 0,6 et 1. La surexpression de la protéine d'intérêt est alors induite par l'ajout d'IPTG à 0,3 mM. A ce moment précis, la culture est passée à 20°C toujours sous agitation pendant une nuit. Les milieux sont alors centrifugés à 4 000 g pendant 40 min (ou 8 000 g pendant 10 min) et les surnageants éliminés. Le culot bactérien est ensuite conservé à -20°C.

2.2.2. Purification de protéines recombinantes

Toutes les étapes décrites ci-après sont réalisées à froid (4°C ou dans la glace).

L'étape de purification de protéines recombinantes permet d'extraire les protéines produites des bactéries et de les purifier en éliminant tout autre composé issu de la bactérie. Pour cela, la première étape est la lyse des bactéries. Celles-ci sont d'abord resuspendues dans du tampon de lyse (1 mL pour 100 DO_{600nm}), puis cette suspension est soumise aux ultrasons (3 à 5 fois 2 min à une amplitude de 50% et sur une puissance de 5-6) avec un appareil Sonifier 250 de chez Branson. Ce traitement casse les membranes bactériennes et l'ADN. Afin d'éliminer les gros objets (débris cellulaires), une première étape de centrifugation est réalisée à 20 000 rpm (50 000 g) pendant 45 min à 4°C (Beckman Coulter Avanti J-20XP centrifuge™). Le surnageant est récupéré et subit un traitement au PEI (PolyÉthylène Imine) entre 0,0125 et 0,05% (v/v). Ce composé très long, fortement chargé positivement, fixe les

molécules chargées négativement comme les acides nucléiques et les fait précipiter. Pour éviter qu'il n'entraîne également les protéines chargées négativement, le pourcentage utilisé doit être calibré. Ces éléments précipités sont ensuite éliminés avec une seconde étape de centrifugation à 20 000 rpm (50 000 g) de 30 min à 4°C. Le surnageant est ensuite filtré à 0,45 µm pour éviter de boucher les colonnes.

La protéine d'intérêt est ensuite fixée sur résine par l'intermédiaire de son étiquette 6xHistidines présente à son extrémité N-terminale pour effectuer une IMAC (chromatographie d'affinité pour ions métalliques immobilisés). Pour cela, nous avons utilisé deux types de résines, la résine nickel et la résine TALON (cobalt). Ces résines contiennent du nickel ou du cobalt greffé à la matrice (sépharose). Ces métaux interagissent avec le cycle imidazole des histidines et fixent fortement les histidines de l'étiquette. La résine TALON est plus spécifique mais avec une capacité de fixation moins importante que la résine nickel et permet donc une meilleure discrimination entre les protéines portant l'étiquette 6xHistidines par rapport aux autres protéines (notamment *d'E. coli*, contenant elles-aussi des histidines mais pas en cluster comme l'étiquette). Le surnageant contenant la protéine est donc placé en présence de résine (de 300 µL à 1,2 mL à 50%) pendant 45 min. Cette étape est répétée 2 à 5 fois selon la quantité de protéine présente, jusqu'à épuisement de la protéine d'intérêt dans le surnageant. Une fois les fixations terminées, deux choix sont possibles. L'étiquette de la protéine est coupée par action de la PreScission (3C) (produite au laboratoire) pendant une nuit à 4°C dans du tampon de lyse soit directement sur la résine, soit en solution. Le fait de couper à la 3C permet de séparer la protéine de la résine. L'étiquette y reste elle fixée. Malheureusement, dans certains cas la protéine précipite lors de cette étape ou la coupure reste inefficace. Pour palier à cela, il est possible d'éluer la protéine avec son étiquette par action de tampon d'élution (1 mL pour 300 µL de résine) pendant 20 min à trois reprises. Ce tampon contient 300 mM d'imidazole qui va agir comme compétiteur des histidines pour la fixation du nickel (ou cobalt). Ceci permet d'éluer la protéine de la résine sans couper l'étiquette. Par la suite, la protéine doit être dialysée de façon à éliminer l'imidazole, car ce dernier inhibe la 3C. Cette étape se fait sur la nuit dans un grand volume de tampon de dialyse (2 à 3 L) sous agitation. La protéine est placée dans un boudin de dialyse permettant les échanges de tampon.

Suivant le degré de pureté de la protéine à ce moment, une étape de Q-sépharose (colonne échangeuse d'anions) peut être nécessaire. Pour cela, la protéine est injectée sur colonne Q-sépharose, couplée à un système AKTA FPLC, où elle va se fixer par des

interactions électrostatiques. L'élution de la protéine se fait par le passage d'un gradient en sel (NaCl) allant de 50 mM à 1 M. La protéine est éluée à une concentration en sel qui lui est spécifique, ce qui permet de la séparer de nombreux contaminants ayant des propriétés physico-chimiques différentes. La protéine est détectée par un système UV à 260 nm. Les différentes fractions présentant une forte intensité aux UV sont analysées par SDS-PAGE 12,5%.

Pour finir, la protéine est concentrée à moins de 5 mL, puis centrifugée 5-10 min à 16 000 g avant d'être injectée sur colonne de chromatographie d'exclusion de taille Superdex 75 ou 200 en fonction de la taille de la protéine (installée sur un système AKTA PRIME plus). Durant cette étape, les composés présents en solution vont être séparés en fonction de leur taille (plus un composé est gros, plus il traverse la colonne rapidement car il est exclu des pores). Le débit du tampon traversant la colonne (de base à 1 mL /min) peut être diminué pour améliorer la séparation de protéines de tailles proches (jusqu'à 0,2 mL/min). La protéine est détectée par un système UV à 260 nm. Les différentes fractions présentant une forte intensité aux UV sont analysées par SDS-PAGE 12,5%.

2.2.3. La technique de co-expression (crible)

La technique de crible par co-expression permet de tester des interactions protéine-protéine. Pour cela, chaque protéine doit être exprimée par un vecteur différent et possédant un gène de résistance à un antibiotique différent et une origine de répllication compatible. Les bactéries sont donc co-transformées par ces différentes constructions, puis des colonies sont mises en culture de 4 mL de milieu riche graffinity en *deep well*. Après culture, les milieux sont séparés en deux et sont centrifugés à 4 000 g pendant 10 min. Les deux culots de chaque condition sont traités avec deux tampons de lyse différents (haut sel et bas sel). Une fois les culots re-solubilisés, les cellules sont traités aux ultrasons avec un appareil Vibra-cell 75043 de chez Bioblock Scientific équipé d'une sonde à quatre têtes (amplitude de 30%, cycles on/off de 2 s) deux fois 1 min, puis centrifugées à 4 000 g pendant 20 min. Le surnageant est alors placé en plaque 96 puits avec 40 µL de billes magnétiques (contenant du cobalt), à l'aide d'un robot pipeteur automatisé Tecan® *freedom evo 200*. La plaque est ensuite posée sur une plaque métallique (les billes magnétiques sont alors attirées vers le fond des puits. Le surnageant est ensuite éliminé par pipetage, la résine est lavée avec du tampon de lyse et 30 µL de bleu de dépôt SDS-PAGE sont ajoutés à la résine et incubés 10 min à température

ambiante. Le bleu est ensuite récupéré par pipetage selon la même méthode. Les résultats sont analysés manuellement sur SDS-PAGE 12,5%.

2.2.4. La protéolyse ménagée

La protéolyse ménagée est une technique permettant la protéolyse dirigée d'une protéine. Cela permet d'isoler les différents domaines structuraux d'une protéine. Pour cela, la protéine est placée en présence d'une protéase dans des conditions évitant une protéolyse trop agressive, c'est-à-dire ménagée. Il faut donc optimiser ces conditions pour chaque protéine.

La première étape est un test de différentes protéases (subtilisine, thermolysine, carboxypeptidase Y, elastase, glu-C, chymotrypsine, trypsine et thrombine) à trois concentrations différentes (d'un facteur 10 d'intervalle). Pour ce premier test, le temps d'incubation est défini à 30 min. 15 μ L de protéine (stock à 0,6 mg/mL) sont mis en présence de 4 μ L de protéase à chaque concentration. A l'issue de ce test, les réactions sont arrêtées par ajout de bleu de dépôt (dénaturation de l'enzyme) et les résultats analysés sur SDS-PAGE 12,5%. Les protéases générant des fragments de protéine de petite taille sont retenues pour la seconde étape.

La seconde étape est un test cinétique. Les différentes protéases sont cette fois-ci utilisées à différents temps d'incubation (5 min, 10 min, 20 min, 30 min, 1 h, 2 h et 5 h). Une nouvelle fois, les résultats sont analysés sur SDS-PAGE 12,5%, après avoir stoppé la réaction avec du bleu de dépôt.

Les meilleures conditions (permettant l'obtention de bandes intenses et de taille intermédiaire sur gel) sont utilisées sur une plus grosse quantité de protéine pour que le résultat soit analysé par spectrométrie de masse afin de déterminer précisément les séquences des fragments obtenus.

2.2.5. Le His-pull-down

La technique de *pull-down* permet d'observer une interaction directe entre deux protéines *in vitro*. Pour cela, l'une des deux protéines est purifiée en conservant l'étiquette 6xHistidines et l'autre est purifiée sans l'étiquette. La première est ensuite incubée en présence de résine TALON[®] (50 μ L) pendant une heure sur roue, puis l'excès de protéine non accrochée est éliminé par lavage de la résine au tampon. L'autre partenaire est alors incubé en présence de la résine sur laquelle est fixée la première protéine (1 h sur roue). La résine est ensuite lavée et le résultat du *pull-down* est analysé sur SDS-PAGE 12,5%.

2.2.6. La technique de retard sur gel

Le gel retard est une technique utilisée pour tester des interactions protéine-acide nucléique. Pour cela, l'acide nucléique testé doit être radiomarqué. L'acide nucléique est utilisé à 20-40 cps final par dépôt, avec du tampon D 1X final. Ce mélange est chauffé 5 min à 70°C, puis laissé refroidir 5min à température ambiante. 250 ng d'ARNt sont ajoutés ainsi que la protéine testée (5 µL dans du tampon D). Cette solution est incubée 20-30 min à 4°C avant l'ajout de 10 µL de bleu CSB. L'échantillon (de 19 µL) est ensuite déposé sur gel acrylamide 6 ou 8%. La migration se fait à 100 V constant pendant 2,5-3 h, puis le gel est séché pendant 45 min et la radioactivité est révélée avec un imageur laser Typhoon 9400 (GE Healthcare).

2.3. Analyses biophysiques des protéines

2.3.1. ITC (Isothermal Titration Calorimetry)

La technique d'ITC est utilisée pour mesurer une constante d'affinité (K_d) entre deux molécules *in vitro* et déterminer les paramètres thermodynamiques de cette interaction ainsi que la stœchiométrie. Le principe de cette technique repose sur les échanges de chaleurs réalisés par l'échantillon avec le milieu. En effet, dans la plupart des cas, la formation d'un complexe entre deux molécules se traduit par une absorption ou un dégagement de chaleur. L'appareil ITC se compose de deux cellules (cuves) adiabatiques, l'une où est réalisée l'expérience et l'autre qui est la cellule référence. Le principe de base est qu'à tout moment les deux cellules doivent être à la même température, choisie au début de l'expérience. Comme la cuve référence ne contient que de l'eau et n'est pas touchée, sa température ne doit pas varier. Si l'interaction entre deux partenaires dans la cellule échantillon modifie la température de celle-ci, l'appareil s'emploie aussitôt à ramener la température à sa valeur normale. Le résultat de l'expérience d'ITC est donc le reflet de l'énergie fournie par l'appareil afin de retourner à la température de base après chaque injection de titrant. En effet, plusieurs injection successives de titrant sont réalisées. Pour cela, l'un des deux partenaires testés (le titrant) est placé dans une seringue à injection automatique. L'autre partenaire (titré) est placé dans la cellule échantillon. L'appareil ajoute automatiquement et à intervalles constants le même volume de titrant dans la cellule. Les intervalles entre chaque injection doivent permettre un retour à l'équilibre. Les concentrations des deux partenaires doivent permettre

de se positionner de part et d'autre de l'état de stœchiométrie dans la cellule entre la première injection de titrant et la dernière (de part et d'autre du point d'inflexion de la sinusoïde caractérisant l'interaction).

Deux appareils ITC différents ont été utilisés, le VP-ITC et l'ITC₂₀₀ de chez MicroCal™ (GE Healthcare). Le partenaire titré (dans la cellule échantillon) est placé à une concentration de 50 µM (VP-ITC) ou 20-40 µM (ITC₂₀₀) et le titrant est placé dans la seringue à une concentration dix fois supérieure (500 µM pour le VP-ITC et 200-400 µM pour l'ITC₂₀₀). Le tampon utilisé est composé de phosphate (NaPi) à différentes concentrations en sel (entre 0 et 150 mM) en présence ou non de réducteur (TCEP). Les expériences ont été réalisées à 20°C (10°C dans le cas des expériences avec un peptide). Le nombre d'injections varie entre les deux appareils. Avec le VP-ITC, 50 injections de 5 µL sur 5 s sont espacées de 300 s. Dans le cas des expériences avec l'ITC₂₀₀, 17 injections de 2,5 µL sur 5 s sont réalisées et espacées de 180 s.

2.3.2. Dichroïsme circulaire

Les analyses par dichroïsme circulaire permettent d'avoir des informations sur la composition en structures secondaires d'un échantillon. Le principe de cette technique repose sur la propriété qu'ont les molécules chirales à absorber différemment la lumière polarisée circulairement droite et la lumière polarisée circulairement gauche. Cela se traduit par la déviation de la résultante des deux polarisations observée pour chaque longueur d'onde. Ce phénomène est notamment important pour les structures secondaires présentes dans les protéines ou acides nucléiques (observé dans le domaine des ultraviolets).

Les analyses par dichroïsme circulaire en cuve simple (1 mm) sur un échantillon unique en solution se font à une concentration en protéine de 5-10 µM dans du tampon phosphate avec TCEP. Le NaCl est remplacé par du NaF pour cette expérience, car le chlore parasite les résultats. Lors des analyses en double cuve (2 fois 4,37 mm), deux échantillons en solution sont placés dans une cuve à deux compartiments séparés. Après une analyse des deux échantillons séparés (le faisceau traversant les deux compartiments), les deux échantillons sont mélangés et la seconde analyse se fait dans la même cuve mais avec les échantillons en présence l'un de l'autre. Pour ce type d'analyse, les deux protéines sont utilisées à des concentrations de 125 nM (500 µL). Ces expériences ont été réalisées avec un spectromètre CD Chirascan-Plus de chez Applied Biophysics, qui est accessible *via* la plateforme SCBIM de la FR BMCT 3209 au Biopôle de l'Université de Lorraine.

2.3.3. DLS (Dynamic Light Scattering)

La technique de DLS (Diffusion dynamique de la lumière) repose sur la diffusion de la lumière par un échantillon en solution, dont seuls les rayons déviés à 90° sont observés. Pour cela, la cuve utilisée (QS 3 mm, 12 µL, Hellma) possède des fenêtres sur trois de ces faces. La diffusion de la lumière par les particules en solution dépend de la taille de ces particules. Le DLS permet donc de déterminer la taille des particules en solution, ainsi que l'homogénéité de l'échantillon.

L'analyse DLS se fait avec un appareil Zetasizer Nanoseries (Nano-S) de Malvern Instruments à température variable avec 15 µL d'échantillon. L'échantillon est centrifugé 5-10 min à 16 000 g au préalable. Après une équilibration de 1 à 2 min, 3 séries d'une dizaine de mesures sont réalisées. Ces mesures sont ensuite moyennées.

2.3.4. DSC (Differential Scanning Calorimetry)

La calorimétrie différentielle à balayage (ou DSC) est une technique d'analyse thermique d'un échantillon en solution. Cette technique est utilisée pour déterminer la température de fusion (T_m) et/ou de précipitation d'une protéine, ce qui reflète sa stabilité thermique. Le système de détection est similaire à celui de l'ITC.

L'échantillon (600 µL) est injecté dans la cuve d'un appareil VP-DSC de chez MicroCal™ et va subir une augmentation de la température de 20°C à 110°C à raison de 1°C/min. La concentration de l'échantillon est comprise entre 10 et 100 µM en fonction de la taille de l'objet étudié.

2.4. La RMN (Résonance Magnétique Nucléaire)

2.4.1. Préparation des échantillons

Les échantillons protéiques pour la RMN sont purifiés et concentrés comme expliqué avant. La concentration d'un échantillon protéique en vue d'études structurales doit être supérieure 500 µM. Les échantillons sont ensuite centrifugés à 16 000 g pendant 10 min. 5 à 10% de D₂O sont ajoutés à l'échantillon. Celui-ci est ensuite placé dans un tube RMN de 5 mm (environ 550 µL) ou de 3 mm (environ 200 µL), puis légèrement centrifugé.

2.4.2. Acquisition des spectres

Les spectres RMN ont été enregistrés sur un spectromètre RMN Bruker Avance III de 600 MHz équipé d'une cryosonde TCI à triple résonance ^1H , ^{13}C et ^{15}N à l'aide du logiciel TopSpin® 3.0. Les attributions séquentielles des protéines et de leurs chaînes principales sont réalisées avec les spectres HSQC ^1H - ^{15}N , HNCA, HNCACB, CBCA(CO)NH, HNHA, HNCO et HN(CA)CO. Les attributions des chaînes latérales sont elles réalisées avec les spectres HSQC ^1H - ^{13}C , H(C)CH-TOCSY, (H)CCH-TOCSY, NOESY-HSQC- ^{15}N et NOESY-HSQC- ^{13}C pour les chaînes aliphatiques et HSQC ^1H - ^{13}C aromatique et NOESY-HSQC- ^{13}C aromatique pour les chaînes latérales des résidus aromatiques. Les spectres NOESY servent également à l'obtention des contraintes de distances NOE pour le calcul de structure. L'état de protonation des atomes d'azote contenus dans les chaînes latérales des résidus histidines des doigts à zinc a été observé avec des spectres HSQC ^1H - ^{15}N *long range*.

2.4.3. Analyse des spectres et calcul de structures

Les spectres RMN ont été analysés avec les logiciels Sparky et CARRA pour l'attribution des fréquences de résonance et l'acquisition des contraintes NOE (pics sélectionnés manuellement à partir des spectres NOESY). La détermination des angles dièdres des liaisons peptidiques a été réalisée avec le programme TALOS+. Ces angles, alliés aux déplacements chimiques connus de la protéine permettent d'établir des prédictions de structure et sont des contraintes utilisées pour le calcul de structure. Le calcul de structure est dans un premier temps réalisé avec le logiciel CYANA 3.97, puis le calcul de structure dans une boîte d'eau est réalisé avec le programme CNS. Suite à ce calcul, les 20 structures de plus basses énergies sur les 250 calculées sont conservées et considérées comme les plus représentatives. Les structures sont ensuite visualisées avec le logiciel PyMol (Schrödinger, LLC).

2.4.4. Titrages suivis par RMN

Le titrage suivi par RMN permet de visualiser une interaction entre deux protéines et de déterminer (si les conditions sont favorables) l'affinité entre les deux partenaires (Kd) ainsi que les régions impliquées dans l'interaction, à l'échelle de l'acide aminé. Dans cette expérience, seul l'un des deux partenaires est sélectivement et uniformément marqué à l'azote 15. De cette façon, seule la protéine marquée est observée sur le spectre RMN

HSQC ^1H - ^{15}N . Le but est d'observer des modifications de déplacement ou d'intensité au niveau des pics correspondant aux résidus dont l'environnement chimique est modifié suite à l'ajout du partenaire, entre les spectres de la protéine marquée libre et liée.

Les expériences de titrages suivis par RMN ont été réalisées dans des tubes RMN de 3 mm (200 μL). La première étape consiste en l'acquisition d'un premier spectre HSQC ^1H - ^{15}N de référence, où seule le partenaire marqué à l'azote 15 est présent (à une concentration autour de 400 μM). Une fois ce premier spectre enregistré, une série d'autres spectres aux mêmes paramètres sont enregistrés après ajouts successifs de petits volumes du partenaire non marqué (pour ne pas faire trop chuter la concentration du premier partenaire). Pour cela, la concentration stock du partenaire non marqué doit être de l'ordre de 1 mM.

Les spectres obtenus avec différents ratios de concentrations entre les deux partenaires peuvent ensuite être comparés entre eux et à la référence. Les pics correspondants aux résidus impliqués dans l'interaction doivent subir des modifications de déplacement chimique ou d'intensité de plus en plus important avec l'ajout en concentrations croissantes du partenaire.

Les spectres enregistrés pour les expériences de titrage suivi par RMN avec les protéines Bcd1 et Rtt106 sont des TROSY (*Transverse Relaxation Optimized Spectroscopy*), ce qui permet d'augmenter la résolution et la sensibilité pour de larges biomolécules. L'expérience de TROSY permet de sélectionner la composante de couplage pour laquelle les mécanismes de relaxation ont pratiquement disparu, conduisant à un seul pic étroit.

2.4.5. Tests d'interaction par suivi RMN des méthyles

Dans le cas de protéines de faible concentration, la technique de suivi des méthyles par RMN permet d'identifier une interaction entre deux protéines, dont l'une est sélectivement marquée au carbone 13 (^{13}C). Avec cette technique, la taille du partenaire non marqué n'a pas d'importance, contrairement à celle de la protéine marquée qui doit être dans la gamme étudiable par RMN. Cette expérience est réalisée en tube RMN de 3 mm (200 μL) et est basée sur l'expérience Sofast HMQC ^1H - ^{13}C , où le nombre de points de la dimension carbone est restreint à 1 (spectre 1D). De plus, l'acquisition dans la dimension proton est calibrée sur la région des méthyles. Cette expérience permet donc d'observer sélectivement les protons liés à des atomes de carbone marqués au ^{13}C , et ainsi de sélectionner spécifiquement les méthyles de la protéine marquée en s'affranchissant du signal de l'autre protéine, plus grosse et non marquée. Tout cela permet de réduire fortement le temps d'expérimentation, et ainsi de pouvoir augmenter le nombre de scans de manière significative afin d'améliorer le ratio du

signal sur bruit. Il est donc possible de travailler sur des concentrations très faibles pour la RMN (de l'ordre de la dizaine de μM).

Au même titre que le titrage suivi par RMN, la première étape est l'acquisition d'un spectre référence avec uniquement la protéine marquée au ^{13}C . Puis, le partenaire potentiel est ajouté à équimolarité pour l'acquisition du second spectre. La présence de différences significatives entre les deux spectres est le signe d'une très probable interaction entre les deux protéines testées.

Dans le cas des tests d'interaction avec les protéines RuvBLs, les tampons utilisés sont listés dans la partie 1.7.

Résultats

&

Discussion

1. Etude des doigts à zinc de la famille zf-HIT

1.1. Travaux publiés

Les domaines en doigt à zinc de la famille zf-HIT (ou ZHD pour *Zf-HIT Domain*) sont notamment retrouvés chez les protéines Bcd1 et Hit1, faisant toutes deux partie des facteurs de la machinerie d'assemblage des snoRNP à boîtes C/D. Afin de déterminer quelle pouvait être l'importance de ces domaines pour la fonction de ces protéines, nous avons entrepris des études fonctionnelles chez la levure ainsi que des études structurales par RMN.

Dans un premier temps, des constructions plasmidiques exprimant des variants de ces protéines (délétions de séquences) ont été utilisées lors de tests fonctionnels par complémentation. La protéine Bcd1 étant essentielle à la viabilité cellulaire, la suppression de l'expression du gène codant cette protéine empêche toute croissance. Une souche où le gène codant la protéine Bcd1 est placé sous le contrôle d'un promoteur répressible par la tétracycline (*tetO₇::BCD1*) a été créée au laboratoire. Cette souche permet ainsi d'observer en condition de répression du gène *BCD1*, l'impact sur la croissance des levures de l'expression ectopique de différents variants. Ces expériences (travaux de thèse de Decebal Tiotiu) ont montré que la présence du ZHD de Bcd1p est importante pour la fonction essentielle de cette protéine, car les mutants de délétion de ce domaine ne peuvent pas restaurer la croissance des levures aux températures élevées (Figure 2 de la publication). Ce domaine n'est toutefois pas suffisant pour assumer cette fonction, car exprimé seul il ne restaure pas la croissance quelle que soit la température. En ce qui concerne la protéine Hit1, son rôle fonctionnel est la stabilisation de la protéine Rsa1 (Rothe *et al.*, 2014). Or, son domaine ZHD ne fait pas partie de la séquence nécessaire à l'interaction avec Rsa1p et sa fonction biologique n'a pas été identifiée.

Afin de progresser dans notre compréhension du rôle fonctionnel de ce domaine, nous avons résolu par RMN les structures tridimensionnelles des deux ZHD des protéines Bcd1 et Hit1 membres de la famille zf-HIT (Figures 3 et 8 de la publication). Ces deux structures sont en accord avec celles déjà connues d'autres membres de cette famille (ZNHIT2, code PDB : 1X4S ; ZNHIT3, code PDB : 2YQQ et ZNHIT5, code PDB : 2YQP). En effet l'agencement spatial global de tous les ZHD semble être le même : un feuillet β constitué de deux brins antiparallèles qui se positionne sur une hélice α . Cet agencement verrouille le (ou les) atome(s) de zinc entre ces deux structures secondaires.

Ces premiers résultats nous ont conduit à réaliser plusieurs variants portant des substitutions d'acides aminés et dont les effets ont été analysés par des tests de

complémentation de la croissance des levures *tetO₇::BCD1* (Figure 4 de la publication). Tout d'abord, les cystéines de la protéine Bcd1 coordonnant les atomes de zinc ont été mutées en sérines. Ces variants nous ont permis d'observer l'importance de la structuration tridimensionnelle de ce domaine sur sa fonction, car l'absence d'atomes de zinc empêche le repliement de ce domaine. Cet effet a été vérifié par dichroïsme circulaire (Figure S4 de la publication). De plus, des variants chimériques ont été construits afin de tester si le remplacement du ZHD de Bcd1p par un autre ZHD (Figure 7 de la publication), ou même par un doigt à zinc d'une autre famille, affectait ou non la fonctionnalité de la protéine.

Les domaines en doigt à zinc étant généralement connus pour être médiateur d'interactions protéine–acide nucléique (ou protéine–protéine), des expériences de retard de migration sur gel ont été menées afin de tester *in vitro* l'interaction potentielle de ce domaine avec l'ARN ou l'ADN. Ces expériences ont montré que la protéine Bcd1 est effectivement capable de lier l'ARN (probablement de façon aspécifique), mais que le ZHD n'est pas impliqué dans cette liaison (Figures 5 et 6 de la publication).

1.2. Publication



Functional and Structural Insights of the Zinc-Finger HIT protein family members Involved in Box C/D snoRNP Biogenesis

Benoit Bragantini^{1,†}, Decebal Tiotiu^{1,†}, Benjamin Rothé¹,
Jean-Michel Saliou^{3,4}, H  l  ne Marty¹, Sarah Cianf  rani^{3,4},
Bruno Charpentier¹, Marc Quinternet² and Xavier Manival¹

1 - Ing  nierie Mol  culaire et Physiopathologie Articulaire (IMoPA), UMR 7365, CNRS Universit   de Lorraine, Biop  le, Campus Biologie-Sant  , 9 Avenue de la For  t de Haye, CS 50184, 54505 Vand  uvre-l  s-Nancy, France

2 - FR 3209, CNRS-Universit   de Lorraine, Bioing  nierie Mol  culaire, Cellulaire et Th  rapeutique, Biop  le, Campus Biologie-Sant  , CS 50184, 54505 Vand  uvre-l  s-Nancy Cedex, France

3 - BioOrganic Mass Spectrometry Laboratory (LSMBO), IPHC, Universit   de Strasbourg, 25 rue Becquerel, 67087 Strasbourg, France

4 - IPHC, UMR 7178, CNRS, 67087 Strasbourg, France

Correspondence to Marc Quinternet and Xavier Manival: marc.quinternet@univ-lorraine.fr; xavier.manival@univ-lorraine.fr
<http://dx.doi.org/10.1016/j.jmb.2016.04.028>

Edited by Mingjie Zhang

Abstract

Zf–HIT family members share the zf–HIT domain (ZHD), which is characterized by a fold in “treble-clef” through interleaved CCCC and CCHC ZnF motifs that both bind a zinc atom. Six proteins containing ZHD are present in human and three in yeast proteome, all belonging to multimodular RNA/protein complexes involved in gene regulation, chromatin remodeling, and snoRNP assembly. An interesting characteristic of the cellular complexes that ensure these functions is the presence of the RuvBL1/2/Rvb1/2 ATPases closely linked with zf–HIT proteins. Human ZNHIT6/BCD1 and its counterpart in yeast Bcd1p were previously characterized as assembly factors of the box C/D snoRNPs. Our data reveal that the ZHD of Bcd1p is necessary but not sufficient for yeast growth and that the motif has no direct RNA-binding capacity but helps Bcd1p maintain the box C/D snoRNAs level in steady state. However, we demonstrated that Bcd1p interacts nonspecifically with RNAs depending on their length. Interestingly, the ZHD of Bcd1p is functionally interchangeable with that of Hit1p, another box C/D snoRNP assembly factor belonging to the zf–HIT family. This prompted us to use NMR to solve the 3D structures of ZHD from yeast Bcd1p and Hit1p to highlight the structural similarity in the zf–HIT family. We identified structural features associated with the requirement of Hit1p and Bcd1p ZHD for cell growth and box C/D snoRNA stability under heat stress. Altogether, our data suggest an important role of ZHD could be to maintain functional folding to the rest of the protein, especially under heat stress conditions.

   2016 Elsevier Ltd. All rights reserved.

Introduction

Zinc-binding proteins are one of the largest protein superfamilies in eukaryotes and represent up to 10% of the human proteome [1]. These proteins contribute to functionally important protein–protein or protein–nucleic acid interactions for regulation of gene expression including transcription regulation, chromatin remodeling, and mRNA stability and processing [2]. The most common types of coordination environment for zinc-finger (ZnF) motifs are the CCCC and CCHH type (C for cysteine and H for histidine), and the classification of the zinc-binding

proteins in different subfamilies relies on a combination of these ZnF motifs. In the TRASH (trafficking, resistance, and sensing of heavy metals) superfamily, the members of the zf–HIT family share the zf–HIT domain (ZHD), which is characterized by a fold in “treble-clef” [3,4]. The name of this family derives from the yeast protein Hit1, which is the first protein for which ZHD was defined [5]. The globular 3D fold of the ZHD was described for the first time in the ZNHIT2 protein and consists of two or four β -strands followed by a C-terminal short α -helix packed against the second β -sheet [4]. Amino acid sequence analysis revealed that the zf–HIT family is

divided into three distinct subgroups (Fig. 1). The first subgroup includes proteins with the interleaved CCCC and CCHC ZnF motifs [first zinc finger (ZnF1) and second zinc finger (ZnF2)] that bind two zinc atoms. The CCHC ZnF motif is not conserved in the second and third subgroups, leading to the proposal that these subgroups bind a single zinc ion [4]. This type of classification is also consistent with results of phylogenetic analysis of zf-HIT-containing proteins. Indeed, the first subgroup includes proteins with a single ZHD with two ZnF motifs, whereas the second and third subgroups include a ZHD consisting of a single ZnF motif. The second and third groups are also characterized by the presence of extra domains such as the PAPA-1 homology sequence and the DEAD-box helicase, respectively [4].

ZHD is mainly found in nuclear proteins involved in gene regulation, chromatin remodeling, and pre-snoRNP assembly, but its precise function remains largely unknown. Six Zinc finger HIT domain-containing proteins (ZNHIT1 to 6) are present in the human proteome and three in the yeast *Saccharomyces cerevisiae* proteome (Fig. 1) [6]. In the rest of the paper, mammal proteins are written in uppercase and yeast proteins in lowercase. The ZNHIT1 protein specifically associates with class I histone deacetylase, which was suggested to reduce the acetylation level of histone H4 in the CDK6 promoter region [7]. Moreover, ZNHIT1 protein is associated with helicases TIP49/RuvBL1 and TIP48/RuvBL2 (here called RuvBL1/2) within the complex SRCAP (SNF2-related CBP activator protein; aka hINO80). Its counterpart SWR1 complex (SWR-C) in yeast [8] contains the putative homolog of ZNHIT1 (i.e., the vacuolar protein sorting-associated protein 71, Vps71p, also Swc6p) and proteins Rvb1/2, which are the yeast counterparts of RuvBL1/2. Affinity purification and mass spectrometry (AP-MS) analysis showed that

the second protein ZNHIT2 of the human zf-HIT family associates with RuvBL1/2 and RPAP3 (Tah1p in yeast) [9,10]. ZNHIT2 has no known homolog in yeast. The third protein ZNHIT3 (Hit1p in yeast) or human thyroid hormone receptor interacting protein 3 (TRIP3) is a cofactor of the hepatocyte nuclear factor 4 α , which is a transcription factor and a member of the steroid hormone receptor superfamily [11]. Mutations in the hepatocyte nuclear factor 4 α gene are known to cause maturity-onset diabetes of the young [12]. Recently, ZNHIT3(TRIP3)/Hit1p in human and yeast, respectively, was reported as new cofactor of the platform protein NUFIP1 (Rsa1p in yeast) involved in box C/D small nuclear ribonucleoprotein (snoRNP) assembly and tightly associated with RuvBL1/2 [13,14]. Like ZNHIT1, the ZNHIT4 (INO80B) protein was reported to be a core component of the INO80 complex, which also includes RuvBL1/2 [15,16]. Recent structural data indicate that the yeast counterpart of ZNHIT4, Ies2p (Iln eighty subunit 2), interacts with proteins Rvb1/2 to recruit these proteins within the INO80 complex [17]. ZNHIT5 (also DDX59 for DEAD box polypeptide 59) protein is a relatively uncharacterized member of the DEAD-box-containing RNA helicase family of proteins that is known to play a critical role in many aspects of RNA metabolism. Finally, the last member of the family, ZNHIT6 (also BCD1 for box C/D snoRNA protein 1), is the ortholog of yeast Bcd1p that was identified by high-throughput genetic screening of yeast mutants associated with defects in non-coding RNA biogenesis [18]. Sequence analysis of the 366 aa of Bcd1p revealed the presence of a ZHD, which is conserved in human and is similar to the one found in the N-terminal part of the TRIP3/Hit1p protein (Fig. 1 and Supplementary Fig. S1). Taken together, these data emphasize the close connection between the zf-HIT proteins and the RuvBL1/2/Rvb1/2 ATPases.

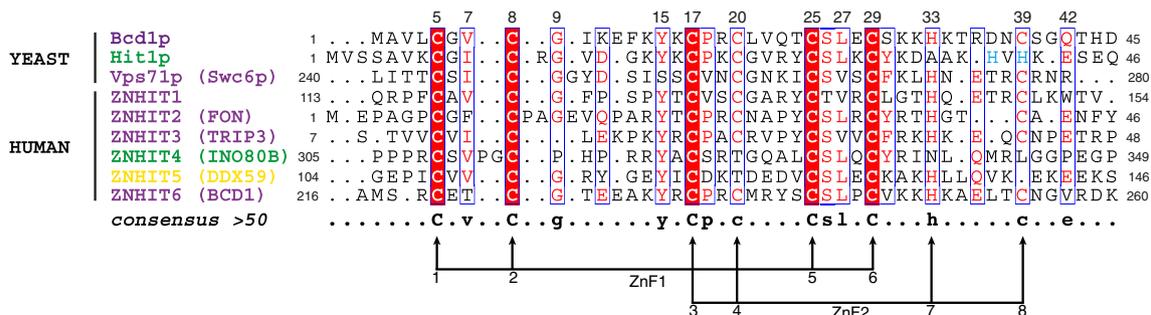


Fig. 1. Sequence comparison between members of the zf-HIT family. Multiple-sequence alignment of different known ZHD sequences obtained from Pfam (Bateman et al., 2004) and made with the ESPript 3.0 program [54] and manually optimized. The cysteine and histidine residues of Bcd1p coordinating zinc ions are indicated by black arrows pointing upwards toward the bottom of the sequences. Numbering at the top corresponds to yeast Bcd1p. Histidine residues involved in the coordination of the second zinc ion in the ZHD of Hit1p are in blue. The three subgroups are in different colors corresponding with the name of the sequences: purple, green, and yellow, for respectively, the first, second, and third subgroups [4].

TIP49(RuvBL1)/Rvb1 and TIP48(RuvBL2)/Rvb2 are two homologous, highly conserved ATP-binding proteins belonging to the AAA+ (ATPases associated with diverse cellular activities) superfamily (for review, see Ref. [19]). They are crucial for a large number of cellular processes and are involved in the assembly of an eclectic set of protein or ribonucleo-protein complexes in the cytoplasm and the nucleus, including various ATP-dependent chromatin remodeling machines such as INO80, SWR1, and p400 [20–24], the histone acetyltransferase TIP60 complex [20], the box H/ACA and box C/D ribonucleo-protein particles [25–27], transcription regulation complexes [28,29], the telomerase complex [30], the RNA polymerase II [31], PIKKs (phosphatidylinositol-3-kinase-like kinases) such as mTOR and SMG1 [32,33], and axonemal dynein [34].

In the recent years, our team has made a particular effort to decipher the mechanisms through which snoRNPs are assembled. Each box C/D snoRNP includes a unique box C/D guide RNA, but they all share a common set of four core proteins, SNU13/Snu13p, NOP58/Nop58p, NOP56/Nop56p, and Fibrillarlin/Nop1, in human/yeast, respectively. The box C/D snoRNPs are catalysts for ribose 2'-O-methylation during the biosynthesis and processing of the pre-rRNA, and these modifications take place in functional regions such as decoding and peptidyl-transferase centers [35]. Concerning the snoRNP biogenesis, the individual depletion of the proteins RuvBL1/Rvb1 and RuvBL2/Rvb2 affects the cellular level of box C/D snoRNAs and leads to incorrect localization of the snoRNP core proteins, both in human and yeast [36,37]. In yeast, the ability of Rvb2 to bind and/or hydrolyze ATP is essential for the accumulation of box C/D snoRNAs [38].

Interestingly, ZNHIT3/Hit1p, which was recently described as a new snoRNP assembly factor specific to the box C/D pathway, is also required for the proper accumulation of box C/D snoRNAs [13,14]. Protein Bcd1 also contributes to the specific stability of the box C/D snoRNAs in *Saccharomyces cerevisiae* and is essential for cellular viability [18,39]. In addition, RNA immunoprecipitation experiments revealed that BCD1 is specifically associated with nucleoplasm precursors of snoRNAs U3 and U8 [40,41]. RNAi-mediated depletion of BCD1 resulted in a significant reduction in the amounts of U3, U8, and U14 but also in a change in the location of U3 snoRNP [40]. These associations probably take place in large assembly complexes, since BCD1 protein is co-purified with proteins NOP58, NOP56, NUFIP1, PIH1D1, and RuvBL1/2 [40]. Pull-down assays using recombinant BCD1 allowed partial mapping of these interactions [40,41]. The truncated 1–360 form of protein BCD1, which contains the ZHD, interacts *in vitro* with Fibrillarlin, NUFIP1, and RuvBL1/2, and these interactions are modulated by ATP. The presence of the C-terminal

domain of BCD1 is required to establish an association with SNU13. Finally, no direct interaction has been observed with PIH1D1. Recently, BCD1 was shown in close association in cells with RuvBL1/2 in an early protein-only complex of the snoRNP biogenesis pathways [13]. Altogether, these results show that BCD1/Bcd1p is a major factor in the biogenesis of box C/D snoRNP.

Here, we report the results of new functional and structural analysis of the Bcd1 and Hit1 yeast proteins that advance our understanding of the box C/D snoRNP biogenesis pathway. Through domain-mapping experiments, we show that ZHD of Bcd1p (ZHD-Bcd1p) is necessary but not sufficient for its function, even if it is not involved in direct RNA binding. Interestingly, we demonstrate that Bcd1 is however able to nonspecifically bind different RNAs in a size-dependent manner. Our data reveal that zinc coordination is nonetheless crucial for the ability of Bcd1p to maintain the box C/D snoRNAs in steady state. Thus, the mutation of the ZnF2 of Bcd1p exhibits a temperature-sensitive phenotype. Using NMR, we solved the high-resolution structures of the ZHD of both Bcd1p (ZHD-Bcd1p) and Hit1p (ZHD-Hit1p) to highlight the determinants of yeast ZHD proteins involved in the box C/D snoRNP assembly. Our data also demonstrate that these two domains adopt similar folds and that their swap did not alter the function of Bcd1p. All these data are discussed with respect to past classification and genetic evolution of ZHD proteins.

Results

Human and yeast sequence analysis of zf-HIT family

Respectively, three and six members of the zf-HIT family are expressed in the yeast *S. cerevisiae* and in human. Multiple sequence alignment confirmed the conservation of eight residues, including seven cysteines and one histidine, which may coordinate two zinc ions with the following spacing: C₁xxC₂-C₅xxxC₆ interleaved with C₃xxC₄-H₇xxxxxC₈ (Fig. 1). The first, second, fifth, and sixth positions correspond to zinc ion ligands in the ZnF1, and the third, fourth, seventh, and eighth positions correspond to the ZnF2. Nevertheless, residues corresponding to the fourth, seventh, and eighth ligand positions are not strictly conserved in Hit1, ZNHIT4, and ZNHIT5 proteins. This observation raised the question of the fold of the ZnF2 motif in these proteins.

In addition to the eight zinc-coordinating residues, other amino acids are significantly conserved (consensus line, Fig. 1), for example, a Val or hydrophobic residue in position 7, a Gly residue in position 9, a Tyr residue in position 15, a Ser residue in position

26, a Leu or Val residue in position 27, and a Glu residue in position 42.

Remarkably, two factors involved in the biogenesis of box C/D snoRNP, proteins Hit1 and Bcd1 and their

respective counterparts TRIP3 and BCD1 in human, belong to the zf-HIT family. Given our team's interest in the assembly of these particles in yeast, we focused our work on Hit1p and Bcd1p.

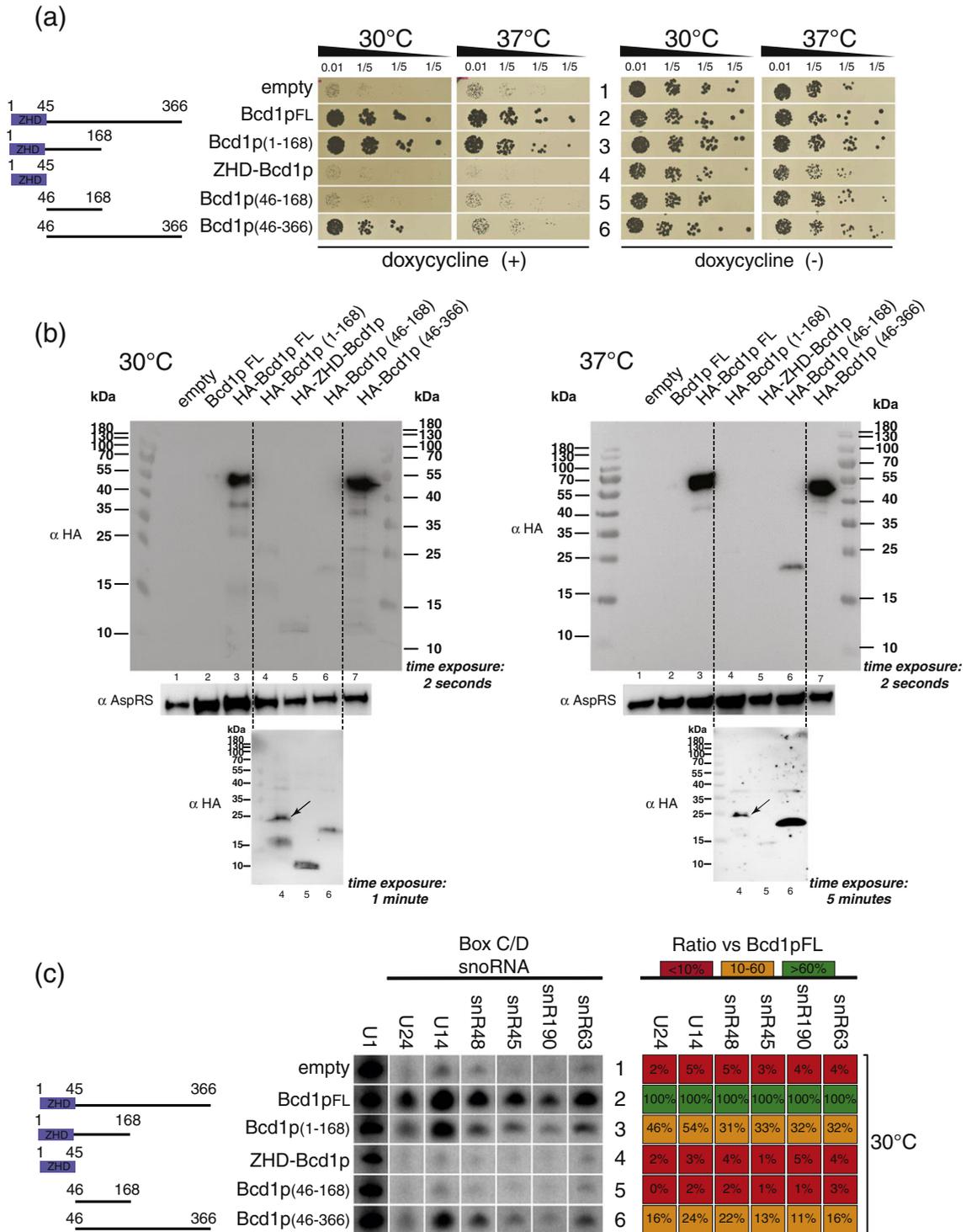


Fig. 2 (legend on next page)

The ZHD of Bcd1p is required for the functionality of the protein

Previously, we showed that the Rsa1p-binding domain of Hit1p, but not its ZHD, is required for the function of Hit1p in snoRNP biogenesis [14]. To test the functional importance of the ZHD of Bcd1p in this pathway, we designed several fragments based on sequence conservation and secondary structure prediction (Fig. 2a). First, each fragment was tested in a complementation assay by ectopic expression in the yeast strain *tetO_z::BCD1* in which expression of the endogenous Bcd1p is repressed upon treatment with doxycycline [18]. The presence of doxycycline (Fig. 2a, left panel *versus* right panel) completely inhibited growth when the cells were transformed with the empty vector p413-TEF (row 1), confirming that Bcd1p expression is essential for cell viability, at both 30 °C and 37 °C. By contrast, the ectopic expression of the full-length 1–366 Bcd1p (row 2) restored growth similar to that in the untreated cells (Fig. 2a, left panel *versus* right panel). Only fragment 1-168 that includes the ZHD allowed complementation at both temperatures (row 3). Even so, the ZHD was not sufficient for complementation (row 4) but was apparently required to maintain the cell viability, since no growth was observed with fragment 46-168 (row 5). This fragment-based approach to complementation assays, which was not previously performed on Bcd1p, revealed that lack of ZHD did not have the same drastic effect since fragment 46-366 allowed partial growth recovery, although only at 30 °C (row 6). The different fragments are not expressed at same levels as shown by western blot analysis of extracts prepared from cells expressing an HA-tagged version of these fragments (Fig. 2b). The amounts of fragments 1-168 (lanes 4) and 46-168 (lanes 6) are similar but much lower

(~3-fold change at 30 °C) than the amounts of full-length Bcd1p (Bcd1p-FL) (lanes 3); fragment 46–366 (lanes 7) is expressed even at a higher level (~1.4-fold change at 30 °C). Nevertheless, only expression of fragment 1–168 allowed an apparent growth similar to the one obtained upon expression of Bcd1p-FL. This showed that the first 168 aa of the N-terminal end represent a functional domain *in vivo*. Interestingly, we showed that the ZHD plays an essential role in the function of the protein and consequently in cell viability.

The ZHD of Bcd1p contributes to the steady-state level of the box C/D snoRNAs

Previously, T.R. Hughes' team demonstrated that Bcd1p depletion causes a specific and drastic decrease in the amount of box C/D snoRNAs in the yeast strain *tetO_z::BCD1* upon doxycycline treatment [18,39]. We next checked for a correlation between the ability of each fragment to complement the growth defect of the *tetO_z::BCD1* strain and to maintain the amount of box C/D snoRNAs. The steady-state levels of a selected series of box C/D snoRNAs were analyzed by Northern blot hybridization on total RNA extracted from cells expressing the different constructs at 30 °C (Fig. 2c). U1 snRNA was used as negative control and as the internal reference for quantification. The resulting data are in agreement with those obtained in the complementation assays (Fig. 2a). Indeed, the absence of Bcd1p (Fig. 2c, row 1) led to a 25-fold average decrease in the levels of the box C/D snoRNAs, while complementation with Bcd1p-FL restored a high level of these RNAs (row 2). Fragment 1–168 (row 3) and to a lesser extent fragment 46-366 (row 6) restored the level of snoRNAs partially but sufficiently to ensure cell growth. These results

Fig. 2. Functional analysis of Bcd1p domains. (a) Complementation assays. Different fragments of Bcd1p were expressed in *tetO_z::BCD1* (R1158) strain transformed with the recombinant p413-TEF(*HIS3*) vector. In this strain, expression of the endogenous gene *BCD1* is repressed in the presence of doxycycline [18]. The panel on the left represents the different fragments, and ZHD is identified by a blue rectangle. Transformed cells were harvested from His⁻ plates, and a stock dilution with $A_{600} = 1$ per mL was prepared in sterile water. The first sample was diluted to $A_{600} 0.01$, four successive cascade dilutions with a factor of 1/5 were prepared, and 10 μ L of each samples were spotted on solid selective (His⁻) media prepared in the presence or absence of 50 μ g/mL doxycycline. Cell growth was assessed after 72 h at 30 °C and 37 °C. The strain transformed with the empty vector p413TEF was used as a negative control (empty). (b) Western blot analysis of equal amounts of total cell lysates from *tetO_z::BCD1* (R1158) strain transformed with the recombinant p413-TEF(*HIS3*)-HA, grown at 30 °C and 37 °C and expressing the different HA-tagged fragments. Tagged proteins were detected with anti-HA (α -HA) antibodies. AspRS protein used as loading control was detected using specific anti-AspRS (α -AspRS) antibodies. (c) Northern blot analysis and heat map representation of steady-state levels of six representative box C/D snoRNAs at 30 °C during the exponential growth phase. Total RNAs were extracted from recombinant *tetO_z::BCD1* (R1158) cells used in the complementation assays, and 10 μ g were fractionated on 6% polyacrylamide denaturing gels. After transfer to Zeta-Probe membrane (Biorad), RNAs were detected by hybridization with specific 5' ³²P-radiolabeled oligonucleotides probes (see [Materials and Methods](#)) and visualized with a Phosphorimager (Typhoon 9410, GE Healthcare). U1 snoRNA was used as an internal control. The figure was built by cutting the bands obtained at the same exposure from the Northern blot membrane. The radioactivity in each snoRNA band was quantified using ImageQuant software version 5.2 (Molecular Dynamics). The percentage of snoRNA in the different total RNAs extracts was calculated from the radioactivity in each band relative to the radioactivity in the band of U1 snRNA. The percentage of each snoRNA relative to the WT cells is presented on the right.

suggest that the contribution of Bcd1p to the expression of box C/D snoRNAs relies at least in part on its ZHD.

NMR solution structure of the ZHD of Bcd1p

To learn more about the minimal fragment 1-168 of Bcd1p, we sought to produce a recombinant form using a standard immobilized metal ion affinity chromatography (IMAC) approach in *Escherichia coli*. Unfortunately, after several attempts, we were still not able to perform full purification. To go further in our functional analysis of ZHD-Bcd1p, we decided to solve the NMR solution structure of fragments 1-45 of Bcd1p.

We first checked for the presence of zinc ions in ZHD-Bcd1p *in vitro*. Native MS analysis revealed the presence of two zinc ions for a total monoisotopic mass of 5465.39 ± 0.01 Da, which is in agreement with the expected mass (ZHD-Bcd1p - 3H + 2Zn, $5337.5 - 3.03 + 130.76 = 5465.23$ Da) (Supplementary Fig. S2A). The need for zinc ions for correct

folding of ZHD-Bcd1p was then highlighted using a large excess of EDTA that completely deteriorated the ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) NMR spectrum of the protein (Supplementary Fig. S3A).

The use of two temperatures (298 K and 313 K) was required to reach an almost complete NMR assignment. Next, a long range ^1H - ^{15}N HSQC spectrum showed a neutral tautomeric form of the imidazole ring of His33, with the protonation of nitrogen $\text{N}^{\delta 1}$ [42]. Thus, inter-proton distance restraints derived from 2D and 3D nuclear Overhauser enhancement spectroscopy (NOESY) spectra, together with zinc-coordination restraints (see [Materials and Methods](#)), enabled us to access the solution structure of ZHD-Bcd1p.

The 20 most representative conformers obtained after water refinement with RECOORD scripts are shown in Fig. 3a, and the statistics for the calculation are listed in Table 1. With RMSD values of 0.44 ± 0.06 Å and 1.28 ± 0.13 Å, respectively, for the backbone and heavy atoms of the ordered

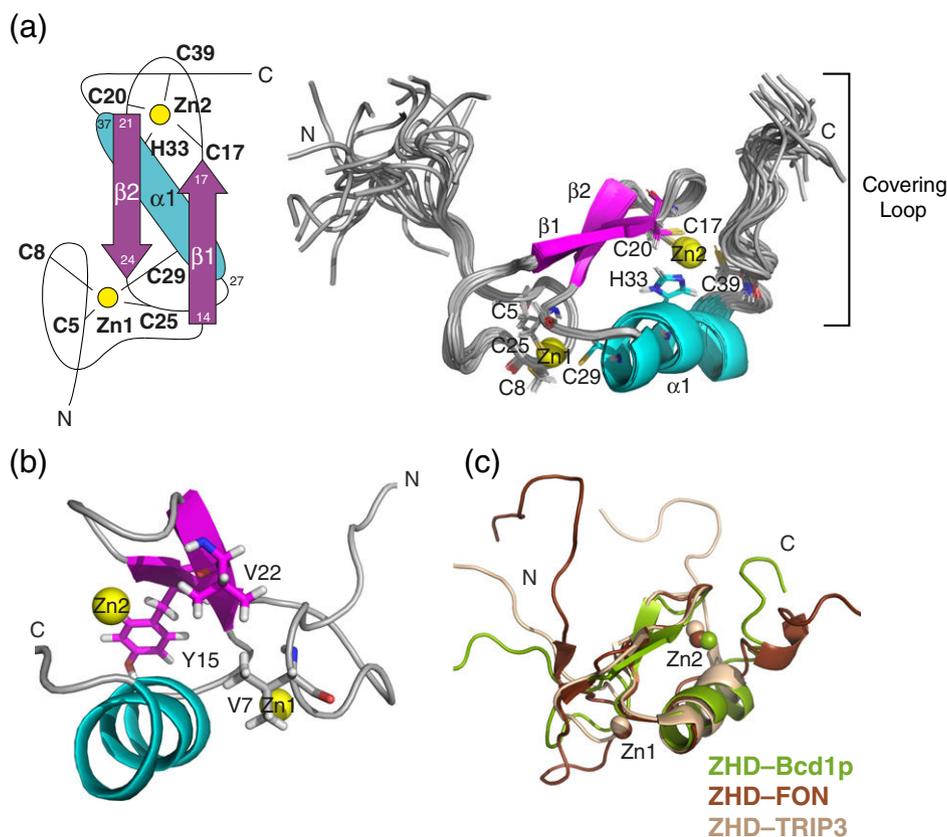


Fig. 3. Solution structures of the ZHD of Bcd1p. (a) Topology diagrams of ZHD-Bcd1p and cartoon representation of the 20 NMR structures with the lowest energies. Zinc ions are in yellow and secondary structures for α -helices in cyan and of β -strands in magenta. Cysteine and histidine residues involved in the zinc coordination are labeled and their side chains are represented by lines. (b) Ribbon representation of the first NMR structure of ZHD-Bcd1p. The side chains of residues Val7, Tyr15, and Val22 are labeled and represented by sticks. (c) Superimposition of the ZHD structures of Bcd1p (1-45) (green), FON (1-46) (brown, PDB code 1X4S), and TRIP3 (7-46) (wheat, PDB code 2YQQ).

Table 1. Statistics of the 20 final solution structures of the Bcd1p and Hit1p ZHDs. The RMSD were calculated using residues 5-39 for ZHD-Bcd1p and 6-41 for ZHD-Hit1p

	ZHD-Bcd1p	ZHD-Hit1p
<i>NMR distances constraints</i>		
Total NOE	902	932
Intra-residue	204	208
Inter-residue		
Sequential ($ i - j = 1$)	248	266
Medium-range ($ i - j < 5$)	158	156
Long-range ($ i - j > 5$)	292	302
Zinc-related restraints	27	26
Dihedral angle restraints	9	10
Average constraints/residues	18.8	19.0
<i>Structure statistics</i>		
Average number of noe violations ($>0.5 \text{ \AA}$)	0.1 ± 0.3	0.0
<i>Deviations from idealized geometry</i>		
Bond length ($.10\text{-}3 \text{ \AA}$)	3.629 ± 0.261	3.604 ± 0.157
Bond angle ($^\circ$)	0.607 ± 0.031	0.550 ± 0.034
Impropers ($^\circ$)	1.394 ± 0.173	1.147 ± 0.073
Mean CNS energy (kcal mol^{-1})	-1348.40 ± 47.2	-1502 ± 16.3
<i>Ramachandran plot statistics (%)</i>		
Residues in most favored regions	67.8	68.6
Residues in additional allowed regions	28.9	29.9
Residues in generously allowed regions	2.0	1.2
Residues in disallowed regions	1.3	1.3
<i>Average CNS pairwise RMSD (\AA)</i>		
Backbone atoms	0.44 ± 0.06	0.28 ± 0.06
Heavy atoms	1.28 ± 0.13	1.30 ± 0.14

region 5-39, the ensemble of NMR structures is well defined.

In detail, this ZHD comprises successively a structured loop (5–13), a two-stranded antiparallel β -sheet (14–17 and 21–24) and an α -helix (27–37) (Fig. 3a). The first zinc ion is located between the N-terminal structured loop and the very beginning of the α -helix, and its binding site is formed by Cys5, Cys8, Cys25, and Cys29 (Fig. 1). The second zinc ion is located between the loop connecting the two β -strands and the α -helix, and its binding site is formed by Cys17, Cys20, His33, and Cys39 (Fig. 3a). The average distance between the two zinc ions is $12.82 \pm 0.08 \text{ \AA}$.

We observed that the second zinc ion is protected by a covering loop in the C-terminal part of ZHD-Bcd1p (37-45), creating a more crowded neighborhood for the second ion than for the first (Fig. 3a). The loop is stabilized in part by the side chain of Arg19, which belongs to the connecting loop between the two β -strands. Moreover, the second zinc finger motif clamps the β -sheet over the α -helix and thus could be needed for the stabilization of secondary structures of ZHD-Bcd1p. It contrasts slightly with the first zinc finger motif, which appears more as an appendix positioned at the beginning of the fold (Fig. 3a). However, as expected, the two zinc finger motifs are arranged in a treble-clef-type domain, which is locked by hydrophobic contacts involving residues Val7 and Val22 (Fig. 3b).

Coordination of zinc ions is crucial for the function of the ZHD of Bcd1p in snoRNP biogenesis

With the help of our structural data highlighting the binding mode of zinc ions in ZHD-Bcd1p, we decided to check if the capacity of Bcd1p to coordinate zinc ions contributed to protein functionality. Thus, we introduced several combinations of point mutations in the full-length protein that consisted of the substitution of the first two cysteines of ZnF1 and/or ZnF2 by Serine (Fig. 4a). Each variant was expressed in *E. coli*, purified, and analyzed by circular dichroism to estimate protein secondary structure. All the variants were soluble and stable. Supplementary Fig. S4 showed that both single, mutated proteins (ZHD_{C5S+C8S}-Bcd1p and ZHD_{C17S+C20S}-Bcd1p) lost typical helical secondary structure, compared to wild-type (WT) Bcd1p, but remained identical to each other. The double-mutated variant (ZHD_{C5S+C8S+C17S+C20S}-Bcd1p) showed much less secondary structure than each of the single-mutated variants. In addition, we substituted the solvent-exposed amino acid Tyr15 by alanine (Figs. 1, 3b, and 4a), this residue being conserved in yeast only in ZHD involved in the box C/D snoRNP assembly pathway (i.e., Hit1p and Bcd1p, but not Vps71p).

We then tested the contribution of the substituted residues to cell growth (Fig. 4b) and to the steady-state level of box C/D snoRNAs (Fig. 4c). Completion of full-length proteins with mutations in

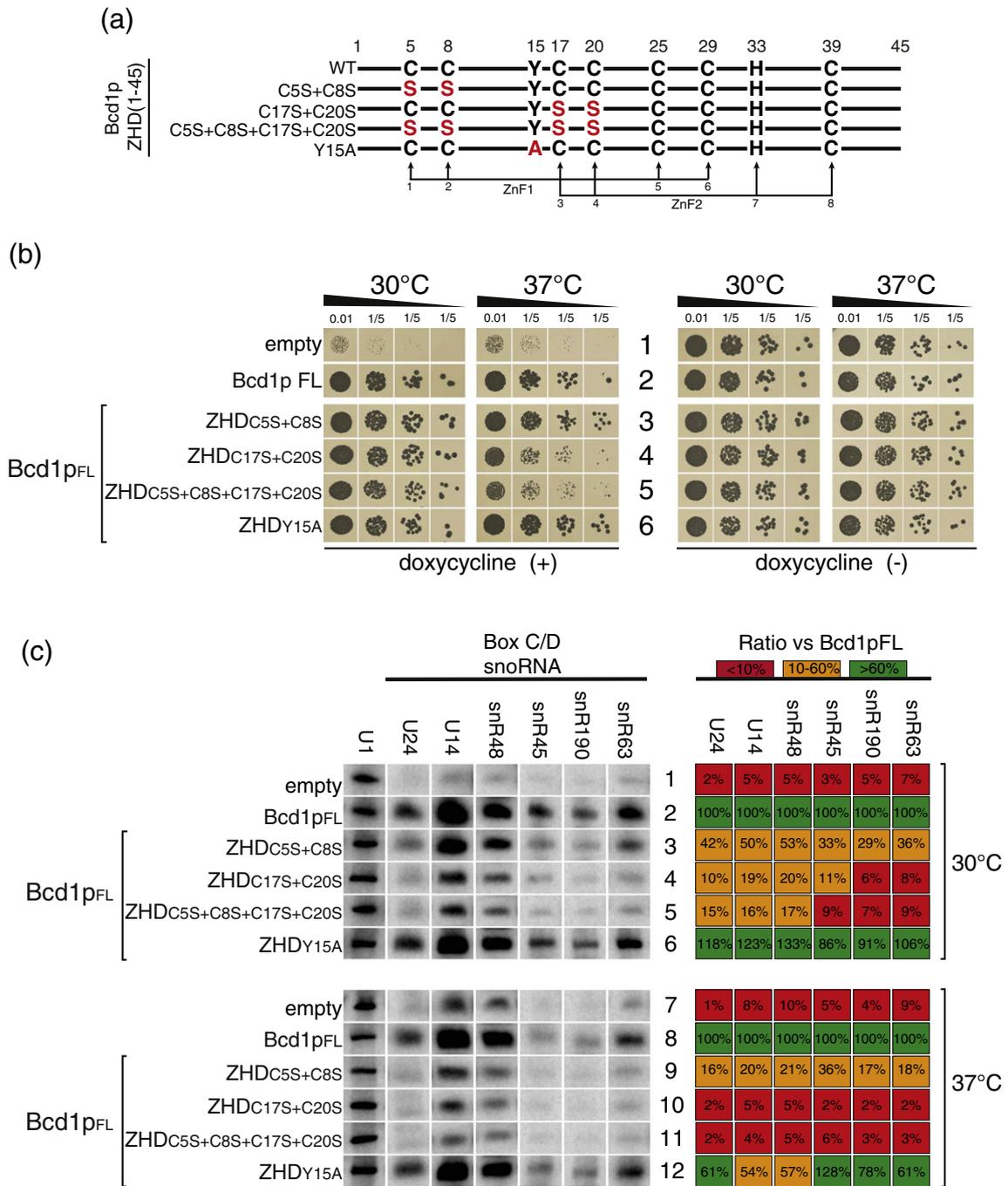


Fig. 4. Functional analysis of the ZHD of Bcd1p in the full-length context. (a) Representation of mutations in ZHD. Cysteine (C) residues substituted by serine (S), and tyrosine (Y) residue substituted by alanine (A) in four variants of Bcd1p-FL. (b) Complementation assays performed as in Fig. 2a for variants of the ZHD. (c) Northern blot analysis of steady-state levels of different box C/D snoRNAs performed as in Fig. 2b.

one of the two zinc fingers allowed growth comparable to that of the WT at 30 °C (rows 3–4) but significantly decreased the amounts of box C/D snoRNAs (rows 3–4). Interestingly, ZnF2 mutations led to a greater reduction in the amount of snoRNA than the ZnF1 mutations, but no cumulative effects

were observed either on growth or snoRNA levels, when the two ZnF motifs were mutated (Fig. 4b–c, compare row 3 with rows 4 and 5). Remarkably, in terms of snoRNA amounts, point mutations in ZnF1 and ZnF2 reproduced the behavior of fragments 1–168 and 46–366, respectively (Fig. 2b, rows 3

and 6 versus Fig. 4c, rows 3 and 4). More impressively, a growth defect was observed at 37 °C when the ZnF2 was modified, again without a cumulative effect in the double mutants (Fig. 4b, rows 4 and 5). First, we observed a global drop in snoRNA amounts with increasing temperature. Then, the remarkable growth defect in the ZnF2 mutant at 37 °C was even more clearly associated with a drastic reduction in the amounts of snoRNAs (Fig. 4c, row 10). These data clearly reinforce the role played by the ZHD of Bcd1p in snoRNP biogenesis. More precisely, the heat sensitivity induced by mutations in ZnF2 strongly suggests that this motif contributes more significantly to the ZHD function than ZnF1. In agreement with this hypothesis, observation of the 3D structure of ZHD-Bcd1p revealed that the ZnF2 motif was involved in the correct positioning of the secondary structures (Fig. 3a).

Second, we wondered if the part of functionality carried by the ZHD in Bcd1p could be based on interactions with protein partners. Thanks to its conservation and solvent exposure, the Tyr15 residue, which is intrinsically able to establish polar and/or hydrophobic contacts, was a target of choice to address this question (Fig. 3b). However, we observed that the Y15A mutation did not significantly impair growth or snoRNAs amounts whatever the conditions tested (Fig. 4b, row 6, and Fig. 4c rows 6 and 12). A single mutation was probably not sufficient to conclude on the protein binding properties of ZHD-Bcd1p, and we thus leave this question open. Since no clear effect was observed upon mutation of residue with protein-interface properties, we next hypothesized that the ZHD interacts with other entities such as nucleic acids or simply maintains a functional entity of folding for the Bcd1p-FL.

The ZHD of Bcd1p is not a DNA/RNA interaction module

Examination of the electrostatic potential of ZHD-Bcd1p derived from our 3D structure revealed a mainly positive charge distribution (Supplementary Fig. S5A) featuring nucleic acid binding properties shared by numerous zinc finger proteins. To check whether the ZHD of Bcd1p contributed to nucleic acid binding, we used electrophoretic mobility shift assays (EMSA) to assess its ability to bind *in vitro* a 20-mer polydIdC oligonucleotide or a box C/D snoRNA (Fig. 5). Recombinant His-tagged full-length and truncated Bcd1p, that is, ZHD-Bcd1p and fragment 46-366 lacking the ZHD (Bcd1p-ΔZHD), were produced in *E. coli* and purified by IMAC. In addition, the ZHD of Hit1p was tested for the sake of comparison. As shown in Fig. 5a, at the highest concentration (20 μM), the Bcd1p-FL significantly retarded migration of the

radiolabeled oligonucleotide dIdC (RNP1). Bcd1p-ΔZHD caused the same band shift as Bcd1p-FL, suggesting that its ZHD is not required for nucleic acid binding. This is in agreement with the fact that neither ZHD-Bcd1p nor ZHD-Hit1p displayed binding properties for the polydIdC oligonucleotide. Interestingly, a 10-fold lower concentration (2 μM) of Bcd1p and Bcd1p-ΔZHD was sufficient to allow complete complex formation with the U14 box C/D snoRNA (RNP2, Fig. 5b).

These data show that Bcd1p exhibits a medium RNA-binding affinity. Furthermore, it appears that the ZHD of Bcd1p, although necessary for cell growth and for maintaining high levels of snoRNAs, is not involved in direct RNA interactions.

Bcd1p interacts nonspecifically with RNAs but binding relies on RNA length

Given the interaction observed *in vitro* between Bcd1p and the U14 snoRNA, we investigated on the specificity of Bcd1p toward this family of RNAs by EMSA. By comparison with U14, we first observed that Bcd1p exhibits a similar affinity for the U3Δ2,3,4 snoRNA, a functional truncated version of U3 previously characterized by our group (Fig. 6a). The U3 snoRNA contains two K-turns that are formed, respectively, by the C/D motif and by the U3-specific B/C motif. To verify if these K-turn motifs are responsible for the binding of Bcd1p, we used two variants of the U3Δ2,3,4 RNA mutated for the B box (variant U3Δ2,3,4-mutB) or for the D box (variant U3Δ2,3,4-mutD). These mutations are known to abolish the formation of the corresponding K-turns and the recruitment of Snu13p [43]. As shown in the Fig. 6a, the affinity of Bcd1p is similar for the WT and the two U3 RNA variants, suggesting that Bcd1p has no specificity for K-turn-containing RNAs.

Then, we tested its association with various RNAs other than box C/D snoRNAs. For this purpose, we used an approach of competition by EMSA (Fig. 6b). Bcd1p was incubated simultaneously with the radiolabeled U14 snoRNA and with an increasing concentration of unlabeled RNA competitor. The unlabeled U14 snoRNA was used as control competitor. As observed in the Fig. 6b, a concentration of 500 nM of unlabeled U14 is necessary to fully inhibit the association of Bcd1p with the radiolabeled RNA. As other competitors, we used molecules from various RNA families, that is, the SeIN mRNA motif that is dedicated to the incorporation of selenocysteine during the translation, yeast tRNAs, the snR36 box H/ACA snoRNA, the U4/U6 snRNA duplex that is a component of the spliceosome, and a fragment of the genomic RNA of the Rous sarcoma virus (RSV). Interestingly, we noticed that the capacity of these molecules to compete with the radiolabeled U14 is proportional to their size. Indeed, the smallest molecules SeIN (59 nt) and

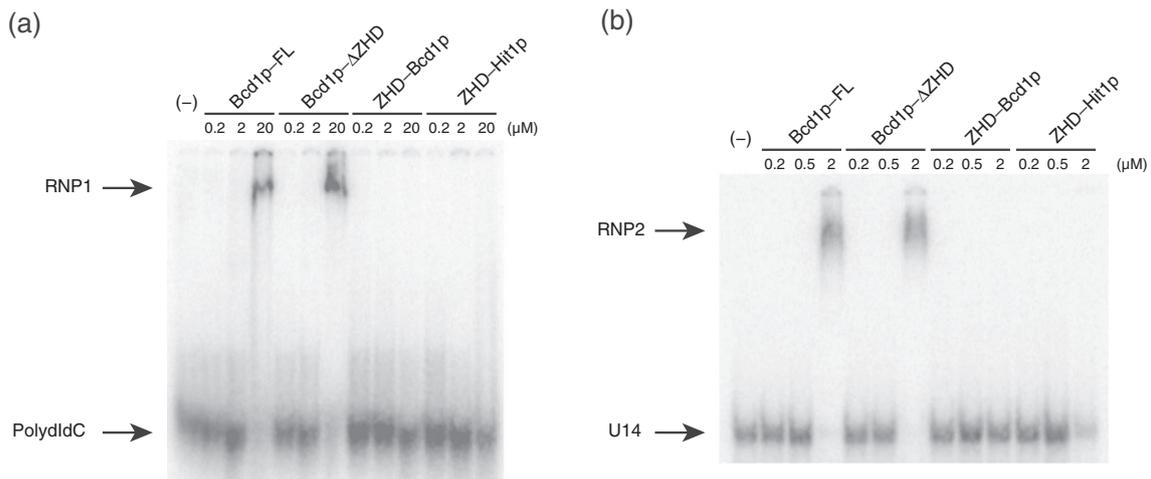


Fig. 5. Bcd1p associates with nucleic acids but has higher affinity for RNA. (a) Bcd1p has very low affinity for a DNA polymer. ³²P-labeled 20-mer dIdC oligonucleotide was incubated with increasing amounts (0.2; 2; 20 mM) of Bcd1p-FL, Bcd1p-ΔZHD lacking the ZHD, ZHD-Bcd1p, and ZHD-Hit1p in conditions described in [Materials and Methods](#). The complexes formed were fractionated by EMSA using 8% polyacrylamide native gel. The positions of the Bcd1p:polydIdC (RNP1) complex and the free polydIdC are indicated on the left side of the autoradiogram. (b) Bcd1p affinity for U14 snoRNA. ³²P-labeled U14 snoRNA was incubated with increasing amounts (0.2; 0.5; 2 mM) of the proteins used in panel A. The complexes formed were fractionated like in panel A, except that 6% acryl native gel was used. The positions of the Bcd1p:U14 RNA (RNP2) complex and the free U14 RNA (U14) are indicated on the left side of the autoradiogram.

tRNAs (70 nt) do not compete, the snR36 snoRNA (182 nt), which has an intermediate size and is the closest to U14 (128 nt), competes similarly as U14 (Fig. 6b). Finally, the two largest molecules, U4/U6 snRNAs (248 nt) and RSV (266 nt) show the maximal competitor effect, since a concentration of 125 nM is sufficient to prevent the complex formation with radiolabeled RNA.

Collectively, these data show that Bcd1p has RNA-binding capacity, likely by interacting with the phosphate backbone of RNA. It exhibits no apparent sequence specificity, because interaction is driven by the length of the RNA molecule.

The ZHD of Hit1p is functional in Bcd1p

As hypothesized above, we next wondered if the role of ZHD-Bcd1p lies in its ability to maintain functional folding to the rest of the protein. To test this hypothesis, we investigated if the ZHD of Bcd1p could be swapped with another ZHD. Genetic constructs expressing chimeric proteins of Bcd1p encompassing yeast ZHD of Hit1p and Vps71p were generated to perform complementation assays (Fig. 7a). As control, we fused Bcd1p to the C-terminal extremity of the double zinc-finger motif of Rph1p, a protein that does not belong to the zf-HIT family. The ZHDs of Hit1p and Vps71p exhibited the ability to restore the cell growth at 30 °C, while the ZnF-Rph1p did not (Fig. 7b, rows 3 and 4). The amounts of snoRNA obtained upon the expression of the Hit1p chimera were close to those obtained with Bcd1p-FL, except

snoRNA snR45 (Fig. 7c, rows 2 and 3). For the Vps71p chimera, the amounts were similar to those obtained with fragment 46–366 of Bcd1p (compare Fig. 2c, row 6 and Fig. 7c, row 4). Surprisingly, the Rph1p chimera, despite expressed at equivalent levels compared to the WT protein (Fig. 7d), displayed an important growth defect at 30 °C associated with very low amounts of snoRNAs. Finally, under heat stress (37 °C), only the Hit1p chimera enabled similar growth to the Bcd1p-FL (Fig. 7b, rows 2 and 3).

The full functionality of ZHD-Hit1p in Bcd1p observed under normal and stress conditions leads us to several conclusions. The presence of yeast ZHDs at the N-terminal part of Bcd1p appears to be required for adequate production of snoRNP for efficient cell growth. However, the presence of zinc ions does not contribute to the functionality of the protein, since the Rph1p chimera is not functionally effective. Moreover, it demonstrates that a swap of the N-terminal zinc finger domain could be more destabilizing for the function of the protein than a deletion. Thus, we can assume that the ZHD maintains functional folding to the entire Bcd1p.

The result obtained under heat stress conditions, which implied efficient production of snoRNP by active assembly factors, reinforces the hypothesis of a crucial role for Bcd1p in this adaptive pathway. This could explain why Vps71p, which has not been shown to be involved in snoRNP biogenesis, does not fully retrieve complementation capacity at 37 °C through its ZHD when it is fused to Bcd1p. These data suggest that only Bcd1p and Hit1p share some

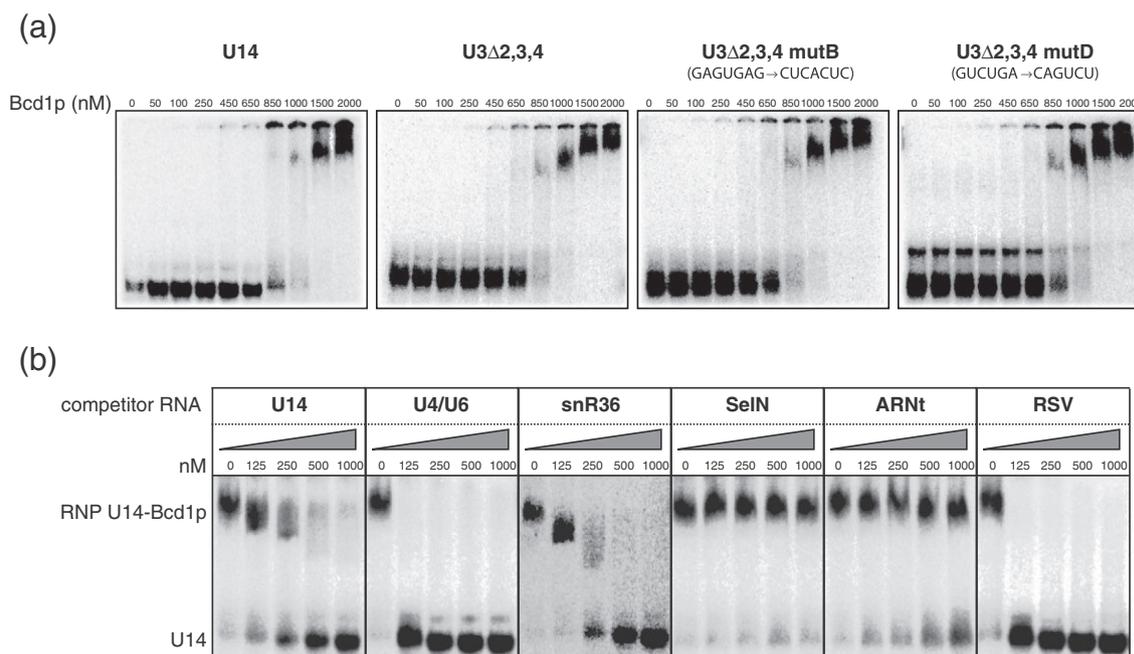


Fig. 6. *In vitro* Bcd1p interacts nonspecifically with RNAs, and affinity relies on the length of the RNA molecule. (a) EMSA between recombinant Bcd1p and radiolabeled U14, U3Δ2,3,4, U3Δ2,3,4 mutB et U3Δ2,3,4 mutD snoRNAs. RNAs were incubated with an increasing range of Bcd1p (from 50 nM to 2 μM). Mutations of B and D boxes totally disrupt the respective B/C and C/D motifs as previously shown by Ref. [43]. (b) Measurement by EMSA of the affinity of Bcd1p for a set of RNAs of variable length. The affinity of Bcd1p to each RNA was assessed by competition assay between the radiolabeled U14 snoRNA and unlabeled RNA competitors. Recombinant Bcd1p was incubated at a concentration of 1 μM, with 5 fmol of radiolabeled U14 snoRNA and an increasing concentration of unlabeled RNA (from 125 nM to 1 μM). The unlabeled U14 snoRNA was used as control competitor. The competitors are of different length: SeIN mRNA (59 nt), tRNA (70 nt), U14 box C/D snoRNA (128 nt), snR36 box H/ACA snoRNA (182 nt), U4/U6 snRNAs duplex (248 nt), and RSV viral RNA (266 nt corresponding to the fragment 691–956 [55]).

common features of their respective ZHD that appear to be functionally interchangeable.

NMR solution structure of the ZHD of Hit1p

Given the functional redundancy of Bcd1p and Hit1p ZHDs, we decided to solve the solution structure of ZHD–Hit1p (fragment 1–46) to identify shared structural determinants with ZHD–Bcd1p. Until now, Hit1p has been classified as a protein with a single ZnF motif. However, detailed observation of the sequence alignment enabled us to identify not only a potential CCCC zinc finger (Cys8, Cys11, Cys28, and Cys32) but also a hypothetical second zinc coordination motif with a CCHH pattern (Cys20, Cys23, His39, and His41) (Fig. 1). Confirming the presence of these two zinc ions in ZHD–Hit1p will be important because it would be the only member of the zf-HIT superfamily to exhibit a singular narrow HxH motif at the C terminus. First, native MS analysis of ZHD–Hit1p revealed the ability of the protein to bind predominantly two zinc ions (Supplementary Fig. S2B). Second, a ¹⁵N-labeled sample of ZHD–Hit1p containing high concentrations of reducing

agents (DTT + TCEP) produced a homogeneous and well-dispersed ¹H–¹⁵N HSQC NMR spectrum, with the expected number of resonances (considering that ZHD–Hit1p is monomeric) (Supplementary Fig. S3B). As before, a large excess of EDTA was responsible for the deterioration of the ¹H–¹⁵N HSQC spectrum. Taken together, these data showed that ZHD–Hit1p trapped two zinc ions with one CCCC and, certainly, one CCHH motif.

Next, we obtained the almost complete NMR assignment of ZHD–Hit1p, and a long range ¹H–¹⁵N HSQC spectrum revealed a neutral tautomeric form with protonated nitrogen N⁵¹ for His39 and His41. The neutral form determined for the two imidazole rings was in strong agreement with the ability of this domain to bind two zinc ions. Consequently, as for the ZHD–Bcd1p, the NMR solution structure of ZHD–Hit1p was determined using NOE-derived data supplemented with zinc cluster restraints.

The solution structure of the 20 most representative conformers of ZHD–Hit1p obtained after water refinement with RECOORDS scripts are presented in Fig. 8a. The statistics used for the calculation are listed in Table 1. After exclusion of the disordered

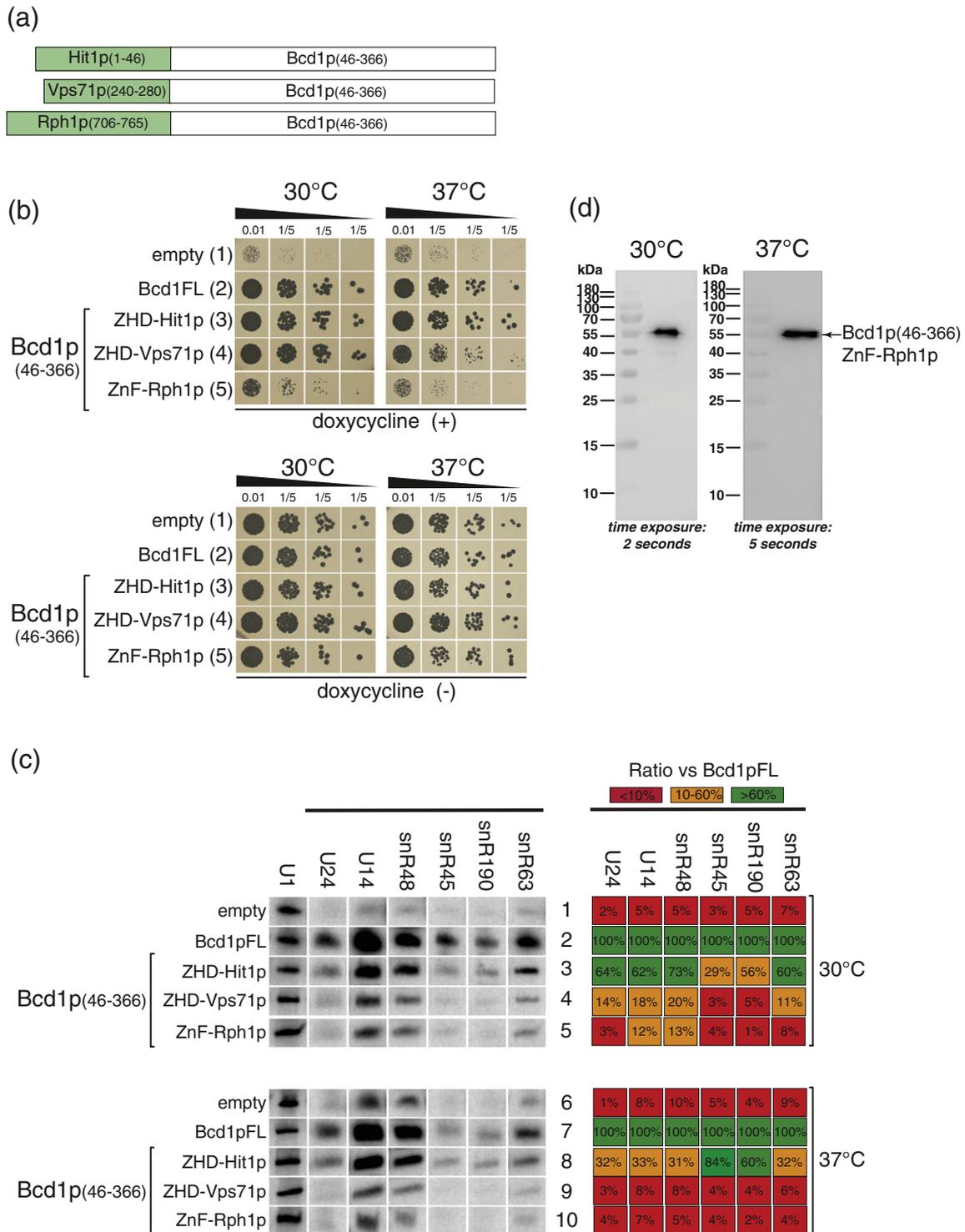


Fig. 7. Functional analysis of chimeric Bcd1p containing different ZnF motifs. (a) Representation of the chimeric proteins for which the endogenous ZHD of Bcd1p was replaced by different yeast ZnF domains (green rectangles): ZHD-Vps71p (residues 240–280) and ZnF-Rph1p (residues 706–765). (b and c) Complementation assays and Northern blot analysis performed as in Fig. 2. (d) Western blot analysis of the amounts of chimeric protein containing the Rph1p ZHD. Detection of a HA-tagged form was performed in conditions described in Fig. 2b.

case of ZHD–Bcd1p, we showed that ZHD–Hit1p was not able to bind nucleic acids (Fig. 5).

Hence, the similar conformation found for ZHD–Hit1p and ZHD–Bcd1p is in good agreement with the results of the domain swapping experiments and confirms the functionality of the chimeric protein ZHD–Hit1p–Bcd1p.

Discussion

Refinement of the organization of the zf-HIT family

As shown in Fig. 3c, the NMR structure of ZHD–Bcd1p is very similar to the two other known structures of the zf-HIT family of the first subgroup, that is, the human proteins ZNHIT2/FON (PDB entry 1X4S) and ZNHIT3/TRIP3 (PDB entry 2YQQ). The protein FON exhibits a 3_{10} -helix, containing the last cysteine of the second zinc-coordinating cluster at the C-terminal position, and a shorter α -helix compared to the one present in Bcd1p. Bcd1p and Hit1p also lack the additional β -sheet present at the N-terminal part of the FON protein. The structure of Bcd1p is more similar to protein FON (C α -RMSD \sim 0.8 Å) than to protein TRIP3 (C α -RMSD \sim 1.1 Å). Similarly, the ZHD–Hit1p 3D structure is closer to that of FON (C α -RMSD \sim 0.7 Å) than to that of TRIP3 (C α -RMSD \sim 1.1 Å). Overall, our structural data confirm similar folding in the members of the zf-HIT family.

According to the sequence analysis depicted in Fig. 1, the ZHDs of Bcd1p and Hit1p were predicted to coordinate two and one zinc ions, respectively. He and collaborators thus proposed that Hit1p belongs to the second subgroup whose members bind a single zinc atom (Fig. 1) [4]. Our MS and NMR analyses revealed that ZHD–Hit1p was able to coordinate two zinc ions with the pattern C $_1$ xxC $_2$ -C $_5$ xxxC $_6$ -C $_3$ xxC $_4$ -H $_7$ xH $_8$. This points to a restrained HxH motif not found in any other members of the zf-HIT family. Finally, we demonstrated experimentally that Hit1p belongs to the first subgroup with Bcd1p and Vps71p.

An exhaustive search of the databases identified the solution structure of a member of the zf-HIT family, ZNHIT5 [or DEAD (Asp-Glu-Ala-Asp) box polypeptide 59]. As suggested by the sequence alignment in Fig. 1, this structure confirms the presence of a single ZnF motif (PDB code: 2YQP). Even in the presence of only one zinc ion, the 3D structure of ZNHIT5 is typical of a ZHD, that is, a treble-clef fold very close to ZHD–Bcd1p (C α -RMSD \sim 0.9 Å). However, the missing zinc ion corresponds to the one that, in ZHD–Bcd1, locks the β -sheet against the α -helix (Fig. 3), strongly suggesting that the comparable organization of secondary structures in ZNHIT5 does not depend only on metal

coordination. Nonetheless, the close inspection of the 3D structure of ZHD–ZNHIT5 did not reveal any strong stabilizing contacts except one salt bridge between Glu123 and His135. Interestingly, we showed that cysteine to serine substitutions in the ZnF2 motif of ZHD–Bcd1p promoted heat sensitivity in yeast growth correlated with a marked decrease in snoRNA amounts (Figs. 3 and 4). We reasonably link this feature with the destabilization of the tertiary structures. Although we do not know the exact function of ZNHIT5, we can assume that its ZHD contributes to the stability of the protein in the face of heat stress. More importantly, we can hypothesize that coordination of a second zinc ion conveys a functional advantage, in the form of thermal stability, to other members of the zf-HIT family. It is easy to see that whatever the external conditions, a snoRNP assembly factor such as Bcd1p needs to be relatively stable to fulfill its crucial function.

Due to its sequence homology with ZNHIT5, it is very likely that ZNHIT4 has only one ZnF motif too (Fig. 1). In this case, in human, ZNHIT4 and 5 would be the only members of this family with a single zinc finger.

Evolution of Bcd1p and Hit1p zinc fingers

Our past and present analyses in yeast revealed that the two ZHDs do not contribute to the function of Bcd1p and Hit1p in the same way. Indeed, under standard growth conditions, this domain is necessary for the functioning of Bcd1p (Fig. 2) but is not required for the functioning of Hit1p [14]. The sequence comparison between these two proteins and their human counterparts illustrates this difference. The ZHD of Bcd1p showed 90% sequence similarity with that of the human protein BCD1, while the ZHD of Hit1p showed only 30% sequence similarity with that of TRIP3 (Fig. 1). The presence of the same motif in these two proteins involved together in snoRNP biogenesis pathway raises the question of functional redundancy. One possibility is that such redundancy involves lower selection pressure on one of the two ZHD molecules and, hence, divergence for the ZHD sequences of TRIP3/Hit1p.

Moreover, the ZHD sequences of Bcd1p and Hit1p showed 36% similarity, whereas their similarity to the ZHD of Vps71p—the third member of this family in yeast—was only 24% and 10%, respectively. Nonetheless, the data resulting from the functional assays performed with the chimeric proteins showed that among the zf-HIT family, only ZHD–Hit1p can efficiently substitute for the ZHD–Bcd1p in terms of growth and snoRNA amounts (Fig. 7). All these observations, especially the functionality of Hit1p lacking its ZHD at 30 °C [14] and 37 °C (data not shown), corroborate the hypothesis under which TRIP3/Hit1p, although still able to ensure the

functionality of Bcd1p, has started to diverge from other ZHDs.

Hypothesis of a functional link between the members of zf-HIT family and the ATPases TIP49(RuvBL1)/Rvb1-TIP48(RuvBL2)/Rvb2 in human and yeast

In the literature, among the three yeast and the six human proteins of the zf-HIT family, we found a pair of common functional partners, TIP49(RuvBL1)/Rvb1 and TIP48(RuvBL2)/Rvb2 ATPases [6]. These proteins are expressed in different structural forms (monomers, hexamers, and dodecamers) specific to the different multicomponent complexes they compose [44]. This structural diversity confers a specific biological activity to each complex and may reflect the acquisition of novel functions during evolution with the assistance of associated factors such as nucleotides, binding partners, or post-translational modifications. The examples we can cite in the context of this work are Vps71p and ZNHIT4 for the yeast chromatin remodeling complexes SWR-C and INO80-C, respectively, and Bcd1p/Hit1p for the yeast box C/D pre-snoRNPs. One possibility is that the proteins BCD1/Bcd1 and/or TRIP3/Hit1 are able, on one hand, to bind/regulate TIP49/Rvb1 and TIP48/Rvb2 ATPases through their ZHD, and on the other hand, to recruit the platform protein NUFIP1/Rsa1p via the C terminus of TRIP3/Hit1p [14]. Their specific recruitment on the pre-snoRNP to assemble might be enabled by the interaction between NUFIP1/Rsa1p and the core protein SNU13/Snu13p [13,45], which specifically recognizes the snoRNAs through their box C/D motif [46,47]. In this context, the RNA-binding properties of Bcd1p may help to accommodate the complex on the nascent pre-snoRNA once this one has reached a length compatible for binding (Fig. 6).

A functional link between these ATPases and BCD1/Bcd1p can be hypothesized. In yeast, a two-hybrid assay identified an association between Bcd1p and Rvb2, while a direct ATP-dependent interaction has been shown between BCD1 and the TIP49/TIP48 complex [40,41]. One possibility will be to test whether these zinc finger proteins help regulate the biological activity of the TIP49/TIP48 complex and its assembly in different structural forms. The possibility of a direct interaction between ZHDs and the domain II of ATPases could be tested. Indeed, an analysis of dodecameric TIP49/Rvb1 and TIP48/Rvb2 structures indicated that the most likely potential interaction domain is domain II of TIP48/Rvb2. A structural study of domain II of TIP48/Rvb2 interacting with Bcd1p, with or without ATP, would be of interest as knowledge of the structure could explain how the Rvb1/2 proteins are able to re-structure pre-snoRNP and modulate protein-protein interactions within it.

The Bcd1p-FL is required for complete functionality

Throughout our experimental procedure, we mainly focused on the N-terminal part of Bcd1p, that is, its ZHD and fragment 1–168, to demonstrate the functional interest of this region. However, we also demonstrated the importance of having a full-length protein to preserve the same accumulation of snoRNAs as that observed for Bcd1p WT. Since the Bcd1p-FL could be replaced by a Hit1p chimera, we assumed that the ZHD stabilized the functional folding of the entire protein. Interestingly, the secondary structure predictions for Bcd1p-FL emphasize the presence of several β -strands and α -helices for the fragment Δ ZHD of Bcd1p, but the overall 3D organization does not match any known or conserved domain, suggesting a new RNA-binding domain (Supplementary Fig. S1). Thus, the ZHD resembles a satellite acting as a fold stabilizer. Nonetheless, we experimentally proved that this unknown fold is able to direct RNA binding. The medium affinity of this binding is in agreement with the role of Bcd1p in the large family of box C/D snoRNAs. In addition, protein partners have already been described in databanks as being associated with Bcd1p using Affinity Capture-RNA, tandem affinity purification, or the two-hybrid assay [48]. Some of these proteins have strong functional links with nucleic acids. We can cite GIS2 and SHE2 gene products, which both bind mRNA; proteins Rtt106 and Ahc2, which are both linked with histone acetylation; MSA2 or MOT3 gene products, which are putative transcriptional factors; and the SFH1 subunit of the chromatin structure-remodeling complex. It would be of great interest to access the 3D structure of Bcd1p full length. This would allow: (i) a better understanding of the role played by its ZHD and (ii) the identification of new surfaces possibly involved in RNA or protein binding.

To go further, Bcd1p is also annotated as a potential partner for proteins of the chaperoning system (Ssa1 to Ssa3, Ssb1, Ssc3, T-complex protein 1 subunit zeta) [48]. We showed that under heat stress, the ZHD of the protein significantly contributes to cell growth. We wonder if this phenotype relies on direct contacts between this domain and proteins of the chaperone family.

Finally, the organization of the mammalian BCD1 diverges significantly from the yeast Bcd1p, whereas similar functions have been described in the two [13,40,41]. It is certain that they share a common ZHD (Fig. 1) but, as often, many sequence insertions are observed in the yeast protein (Supplementary Fig. S1). However, the region downstream from the ZHD is quite nicely conserved, suggesting the ability of the human protein to directly bind snoRNAs through a common 3D fold in the C-terminal domain.

More remarkably, the mammalian Bcd1p counterparts exhibit a supplemental N-terminal region of around 200 residues located upstream from the ZHD and predicted to contain several α -helices. We consequently wonder if this region corresponds to a mammalian gain of function, a hypothesis that remains to be investigated.

Materials and Methods

Plasmid constructs and protein purification

For protein production in *E. coli*, PCR-amplified fragments were cloned between the *NdeI* and *BamHI* restriction sites of the expression vector pNEA-3CH, leading to the production of fusion proteins with an N-terminal His₆-tag and a PreScission cleavage site (details on the pNEA-3CH on request).

For NMR studies on ZHDs, *E. coli* BL21 (DE3) pRARE2 cells were transformed with pNEA-3CH recombinant plasmids, and uniformly labeled recombinant proteins were produced by growing the cells under agitation in M9 minimal medium containing 1 g/L ¹⁵N ammonium chloride (U-¹⁵N 98%, Aldrich) and 2 g/L ¹³C-glucose (U-¹³C 99%, Isotec), 100 μ g/mL ampicillin, 25 μ g/mL chloramphenicol, and 100 μ M zinc sulfate at 37 °C to an absorbance of 0.6–0.8 at 600 nm (A_{600}), and expression was then induced with 1 mM IPTG for 16 h at 20 °C. The fused proteins were purified using a 5-mL HiTrap TALON column (GE Healthcare) and digested overnight in solution with PreScission Protease at 4 °C to remove the N-terminal His₆-tag. This step was followed by preparative gel filtration (HiLoad 16/60 Superdex 75, GE Healthcare) on an AKTA prime system in 10 mM phosphate buffer (pH 6.4), 150 mM NaCl, and TCEP 0.5 mM. The ZnFs of Bcd1p and Hit1p were concentrated to ~0.8 mM and ~2.2 mM, respectively. For the NMR measurement, 3 mM deuterated DTT and 10% D₂O (vol/vol) were added to the protein samples.

For the production of the proteins used for gel shift assays, *E. coli* BL21 (DE3) pRARE2 cells were transformed and grown under agitation in LB medium containing 100 μ g/mL ampicillin and 10 μ g/mL chloramphenicol at 37 °C to A_{600} 0.6–0.8, and expression was then induced with 0.3 mM IPTG for 16 h at 20 °C. The fusion proteins were purified using TALON resin (GE Healthcare) and digested overnight in solution with PreScission Protease at 4 °C to remove the N-terminal His₆-tag. This step was followed by preparative size-exclusion chromatography (HiLoad 16/60 Superdex 75, GE Healthcare) on an AKTA prime plus system in 20 mM Hepes-KOH buffer (pH 7.9), 150 mM KCl, and 1.5 mM MgCl₂.

For expression in *S. cerevisiae*, Bcd1p and Bcd1p domains (1–45, 1–168, 46–168, and 46–366) were amplified by PCR from *S. cerevisiae* genomic DNA and introduced between the *BamHI* and *SalI* restriction sites of the p413TEF (CEN/ARS, *HIS3*) expression vector. Constructs expressing chimeric proteins Bcd1p_{46–366}ZHD-Hit1p, Bcd1p_{46–366}ZHD-Vps71p, and Bcd1p_{46–366}ZHD-Rph1p were obtained by cloning the ZHD of protein Hit1p, Vps71p, and Rph1p in the recombinant

p413TEF::Bcd1p_{46–366}. For expression of fragment with an HA tag at their N-terminal ends, *BamHI*–*SalI* fragments from the recombinant p413TEF vectors were subcloned into a p413TEF vector containing an HA tag introduced upstream of the *BamHI* site. The expression level of each HA-tagged fragments was estimated by western blot analysis in previously described conditions using rabbit commercial anti-HA antibodies [14]. Anti-AspRS antibodies were used to detect Asp aminoacyl tRNA synthetase as internal control. *S. cerevisiae* strain *tetO₇::BCD1* (R1158) [18] was used for the complementation assays.

NMR spectroscopy and resonance assignments

The NMR experiments on ¹³C/¹⁵N-labeled ZHD–Bcd1p and ZHD–Hit1p samples were performed at 298 K and 313 K on a Bruker Avance III 600MHz spectrometer equipped with a TCI cryoprobe. Spectra were processed with the program TopSpin 3.0 (Bruker) and analyzed with the program Sparky [49]. An almost complete resonance assignment was achieved using combinations of 2D ¹H–¹H NOESY, ¹H–¹⁵N and ¹H–¹³C HSQC spectra, and of 3D HNHA, HNCACB, CBCA(CO)NH, HNCO, H(C)CH–TOCSY, ¹H–¹⁵N and ¹H–¹³C NOESY-HSQC spectra. All NOESY spectra were recorded with a mixing time of 200 ms. The protonation state of the imidazole rings of histidine residues (and consequently, of which the two nitrogen atoms coordinates zinc ion) was determined using long range ¹H–¹⁵N HSQC experiments [42].

Calculation of zinc finger structure

The automated procedure of CYANA 3.0 [50] was used to determine the NMR solution structure of each ZHD. Distances between protons were derived from the 2D and 3D NOESY experiments mentioned above. To ensure approximate tetrahedral zinc coordination, we added distance constraints for Zn–C _{β} (3.4–3.6 Å), Zn–S _{γ} (2.2–2.4 Å), Zn–N _{ϵ_2} (1.9–2.1 Å), S _{γ} –S _{γ} (3.7–3.95 Å), and S _{γ} –N _{ϵ_2} (3.7–3.95 Å) [51]. The final set of restraints resulting from the last CYANA iteration was carefully checked and used for water refinement with RECOORD scripts [52]. Water refinement was also performed in the presence of distance and dihedral constraints around the zinc-coordinating residues. Among the 250 structures calculated, the 20 structures with the lowest overall energy were selected as the most representative. Secondary structures were determined with dss (defines secondary structure), and all figures involving 3D structures were drawn with the program PyMOL [53].

EMSA and competition assays

RNA transcripts were synthesized by *in vitro* transcription with T7 RNA polymerase using a PCR fragment carrying the RNA coding sequence in previously described conditions [45]. The 20-mer dIdC oligodeoxyribonucleotide was purchased from Sigma. The oligodeoxyribonucleotide and the transcribed U14 RNA were 5' end labeled with T4 polynucleotide kinase in the presence of [γ -³²P]ATP and purified with a mini Quick Spin RNA

column (Roche) or by gel electrophoresis. For EMSA, WT or variant Bcd1p and ZHD-Hit1p at different concentrations indicated in figure legends were incubated for 20 min at 4 °C with 5 fmoles of ³²P-radiolabeled U14 RNA or polydIdC in buffer D (20 mM Hepes-KOH (pH 7.9), 150 mM KCl, and 1.5 mM MgCl₂). The nucleic acid-protein complexes formed were resolved by native gel electrophoresis on 8% and 6% acrylamide gels without EDTA for 20-mer dIdC and U14 RNA, respectively. Bands containing labeled nucleic acids were visualized with a Phosphorimager (Typhoon 9410, GE healthcare). For the competition assays, the protein and the radiolabeled RNA were mixed at the same time with the different concentrations of unlabeled RNA competitors as indicated in figure legends.

Yeast complementation tests

To test whether fragments of Bcd1p can complement depletion of the endogenous Bcd1p protein, we transformed *S. cerevisiae* strain *tetO₇::BCD1* (R1158) with recombinant p413 plasmids encoding various WT and mutated fragments of Bcd1p. After transformation, recombinant clones were collected from His⁻ plates and serial dilutions were prepared in sterile water before spotting on selective His⁻ plates in the absence or presence of 50 µg/mL doxycycline. Cell growth was assessed after 72 h at 30 °C and 37 °C. Clones of the cells transformed with the empty vector p413TEF were used as negative control.

Measurement of snoRNAs levels in yeast by Northern blot

Total RNAs were extracted from exponentially growing yeast cultures at 30 °C. Northern blot analyses were performed in standard conditions using 10 µg of total RNAs and specific ³²P-radiolabeled antisense probes: OG-U1 (5'-GACCAAGGAGTTTGCATCAATGA-3'), OG-snrR63 (5'-AAGGTTATGTTGGCCACTCA-3'), OG-snrR190 (5'-CGAGGAAAGAAGACACACATTATC-3'), OG-snrR45 (5'-CCTCAGATCGCTCCGAGA-3'), U14 (5'-TCACTCAGACATCCTAG-3'), OG-snrR48 (5'-GAATGGAGAGTACTTAAACTTCACATCC-3'), OG-snrR52 (5'-TTCAGAAGGAAGGCAACATAAG-3'), and OG-U24 (5'-GATCTTGGTGATAATTGGTA-3'). Bands containing snoRNAs and U1 snRNA were visualized with a Phosphorimager (Typhoon 9410, GE healthcare). The radioactivity present in each snoRNA and U1 snRNA band was quantified using the ImageQuant software version 5.2 (Molecular Dynamics).

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jmb.2016.04.028>.

Accession numbers

Chemical shifts and coordinates of ZHD-Bcd1p and ZHD-Hit1p have been deposited in the Biological Magnetic Resonance Data Bank (<http://www.bmrb.wisc.edu/>) and in the RCSB Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>) under accession numbers 26691/2N94 and 25881/2N95, respectively.

Acknowledgments

This work was supported by the Centre National de la Recherche Scientifique (CNRS), University of Lorraine (UL), the Agence Nationale de la Recherche [ANR-11-BSV8-01503], the PRST IMTS of Plan Etat Région Lorraine, and the Région Alsace for the funding of mass spectrometer. We thank the SCBIM platform of FR3209 CNRS BMCT and UL for NMR facilities. We thank Benjamin Bardiaux for technical support with the ZnF-HIT water-refinement procedure, Alexandre Kriznik for Circular Dichroism technical assistance, and Stéphane Erb for the acquisition of mass spectra.

Received 28 December 2015;

Received in revised form 20 April 2016;

Accepted 23 April 2016

Available online 30 April 2016

Keywords:

zinc finger protein;
Bcd1p;
Hit1p;
box C/D snoRNA;
NMR structure

†B.B. and D.T. contributed equally to this work.

Present address: B. Rothé, Ecole polytechnique fédérale de Lausanne (EPFL) SV ISREC, Station 19, CH-1015 Lausanne, Switzerland.

Present address: J.-M. Saliou, Lille Center for Infection and Immunity, Institut Pasteur de Lille, UMR 8204 CNRS-Université Lille Nord de France, INSERM U1019, 1 rue du Professeur Calmette, 59800 Lille, France.

Abbreviations used:

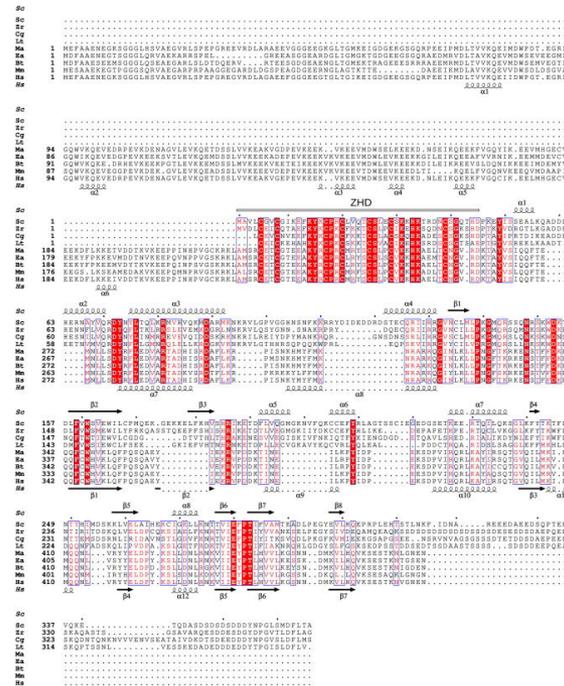
ZnF, zinc-finger; ZHD, zf-HIT domain; AP-MS, affinity purification and mass spectrometry; TRIP3, thyroid hormone receptor interacting protein 3; snoRNP, small nucleolar ribonucleoprotein; ZnF1, first zinc finger; ZnF2, second zinc finger; HSQC, heteronuclear single quantum coherence; NOESY, nuclear Overhauser enhancement spectroscopy; Bcd1p-FL, full-length Bcd1p; WT, wild-type; RSV, Rous sarcoma virus; SeIN, Selenoprotein N.

References

- [1] C. Andreini, L. Banci, I. Bertini, A. Rosato, Counting the zinc-proteins encoded in the human genome, *J. Proteome Res.* 5 (2006) 196–201.
- [2] C. Andreini, L. Banci, I. Bertini, A. Rosato, Zinc through the three domains of life, *J. Proteome Res.* 5 (2006) 3173–3178.
- [3] T.J. Ettema, M.A. Huynen, W.M. de Vos, J. van der Oost, TRASH: a novel metal-binding domain predicted to be involved in heavy-metal sensing, trafficking and resistance, *Trends Biochem. Sci.* 28 (2003) 170–173.

- [4] F. He, T. Umehara, K. Tsuda, M. Inoue, T. Kigawa, T. Matsuda, et al., Solution structure of the zinc finger HIT domain in protein FON, *Protein Sci.* 16 (2007) 1577–1587.
- [5] K. Kawakami, B.K. Shafer, D.J. Garfinkel, J.N. Strathern, Y. Nakamura, Ty element-induced temperature-sensitive mutations of *Saccharomyces cerevisiae*, *Genetics*. 131 (1992) 821–832.
- [6] C. Verheggen, B. Pradet-Balade, E. Bertrand, SnoRNPs, ZNHIT proteins and the R2TP pathway, *Oncotarget*. (2015).
- [7] Z. Yang, Y. Cao, X. Zhu, Y. Huang, Y. Ding, X. Liu, Znhit1 causes cell cycle arrest and down-regulates CDK6 expression, *Biochem. Biophys. Res. Commun.* 386 (2009) 146–152.
- [8] A.J. Morrison, X. Shen, Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes, *Nat. Rev. Mol. Cell Biol.* 10 (2009) 373–384.
- [9] C. Jeronimo, D. Forget, A. Bouchard, Q. Li, G. Chua, C. Poitras, et al., Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme, *Mol. Cell.* 27 (2007) 262–274.
- [10] M. Taipale, G. Tucker, J. Peng, I. Krykbaeva, Z.Y. Lin, B. Larsen, et al., A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways, *Cell.* 158 (2014) 434–448.
- [11] H. Iwahashi, K. Yamagata, I. Yoshiuchi, J. Terasaki, Q. Yang, K. Fukui, et al., Thyroid hormone receptor interacting protein 3 (trip3) is a novel coactivator of hepatocyte nuclear factor-4alpha, *Diabetes*. 51 (2002) 910–914.
- [12] K. Yamagata, H. Furuta, N. Oda, P.J. Kaisaki, S. Menzel, N.J. Cox, et al., Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1), *Nature*. 384 (1996) 458–460.
- [13] J. Bizarro, C. Charron, S. Boulon, B. Westman, B. Pradet-Balade, F. Vandermoere, et al., Proteomic and 3D structure analyses highlight the C/D box snoRNP assembly mechanism and its control, *J. Cell Biol.* 207 (2014) 463–480.
- [14] B. Rothe, J.M. Saliou, M. Quinternet, R. Back, D. Tiotiu, C. Jacquemin, et al., Protein Hit1, a novel box C/D snoRNP assembly factor, controls cellular concentration of the scaffolding protein Rsa1 by direct interaction, *Nucleic Acids Res.* (2014).
- [15] L. Chen, Y. Cai, J. Jin, L. Florens, S.K. Swanson, M.P. Washburn, et al., Subunit organization of the human INO80 chromatin remodeling complex: an evolutionarily conserved core complex catalyzes ATP-dependent nucleosome remodeling, *J. Biol. Chem.* 286 (2011) 11,283–11,289.
- [16] J. Jin, Y. Cai, T. Yao, A.J. Gottschalk, L. Florens, S.K. Swanson, et al., A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex, *J. Biol. Chem.* 280 (2005) 41,207–41,212.
- [17] A. Tosi, C. Haas, F. Herzog, A. Gilmozzi, O. Berninghausen, C. Ungewickell, et al., Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex, *Cell.* 154 (2013) 1207–1219.
- [18] W.T. Peng, M.D. Robinson, S. Mnaimneh, N.J. Krogan, G. Cagney, Q. Morris, et al., A panoramic view of yeast noncoding RNA processing, *Cell.* 113 (2003) 919–933.
- [19] S. Jha, A. Dutta, RVB1/RVB2: running rings around molecular biology, *Mol. Cell.* 34 (2009) 521–533.
- [20] T. Ikura, V.V. Ogrzyzko, M. Grigoriev, R. Groisman, J. Wang, M. Horikoshi, et al., Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis, *Cell.* 102 (2000) 463–473.
- [21] Z.O. Jonsson, S. Jha, J.A. Wohlschlegel, A. Dutta, Rvb1p/Rvb2p recruit Arp5p and assemble a functional Ino80 chromatin remodeling complex, *Mol. Cell.* 16 (2004) 465–477.
- [22] G. Mizuguchi, X. Shen, J. Landry, W.H. Wu, S. Sen, C. Wu, ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex, *Science*. 303 (2004) 343–348.
- [23] A.V. Samuelson, M. Narita, H.M. Chan, J. Jin, E. de Stanchina, M.E. McCurrach, et al., p400 is required for E1A to promote apoptosis, *J. Biol. Chem.* 280 (2005) 21,915–21,923.
- [24] X. Shen, G. Mizuguchi, A. Hamiche, C. Wu, A chromatin remodelling complex involved in transcription and DNA processing, *Nature*. 406 (2000) 541–544.
- [25] Y. Kakiyama, W.A. Houry, The R2TP complex: discovery and functions, *Biochim. Biophys. Acta.* 2012 (1823) 101–107.
- [26] R. Machado-Pinilla, D. Liger, N. Leulliot, U.T. Meier, Mechanism of the AAA+ ATPases pontin and reptin in the biogenesis of H/ACA RNPs, *RNA*. 18 (2012) 1833–1845.
- [27] D.R. Newman, J.F. Kuhn, G.M. Shanab, E.S. Maxwell, Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription, *RNA*. 6 (2000) 861–879.
- [28] Z.O. Jonsson, S.K. Dhar, G.J. Narlikar, R. Auty, N. Wagle, D. Pellman, et al., Rvb1p and Rvb2p are essential components of a chromatin remodeling complex that regulates transcription of over 5% of yeast genes, *J. Biol. Chem.* 276 (2001) 16,279–16,288.
- [29] H. Ohdate, C.R. Lim, T. Kokubo, K. Matsubara, Y. Kimata, K. Kohno, Impairment of the DNA binding activity of the TATA-binding protein renders the transcriptional function of Rvb2p/Tih2p, the yeast RuvB-like protein, essential for cell growth, *J. Biol. Chem.* 278 (2003) 14,647–14,656.
- [30] A.S. Venteicher, Z. Meng, P.J. Mason, T.D. Veenstra, S.E. Artandi, Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly, *Cell.* 132 (2008) 945–957.
- [31] S. Boulon, B. Pradet-Balade, C. Verheggen, D. Molle, S. Boireau, M. Georgieva, et al., HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II, *Mol. Cell.* 39 (2010) 912–924.
- [32] Z. Horejsi, H. Takai, C.A. Adelman, S.J. Collis, H. Flynn, S. Maslen, et al., CK2 phospho-dependent binding of R2TP complex to TEL2 is essential for mTOR and SMG1 stability, *Mol. Cell.* 39 (2010) 839–850.
- [33] H. Takai, Y. Xie, T. de Lange, N.P. Pavletich, Tel2 structure and function in the Hsp90-dependent maturation of mTOR and ATR complexes, *Genes Dev.* 24 (2010) 2019–2030.
- [34] H. Omran, D. Kobayashi, H. Olbrich, T. Tsukahara, N.T. Loges, H. Hagiwara, et al., Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins, *Nature*. 456 (2008) 611–616.
- [35] W.A. Decatur, M.J. Fournier, rRNA modifications and ribosome function, *Trends Biochem. Sci.* 27 (2002) 344–351.
- [36] N.J. Watkins, I. Lemm, D. Ingelfinger, C. Schneider, M. Hossbach, H. Urlaub, et al., Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multi-protein complex, *Mol. Cell.* 16 (2004) 789–798.
- [37] R. Zhao, Y. Kakiyama, A. Gribun, J. Huen, G. Yang, M. Khanna, et al., Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation, *J. Cell Biol.* 180 (2008) 563–578.

- [38] T.H. King, W.A. Decatur, E. Bertrand, E.S. Maxwell, M.J. Fournier, A well-connected and conserved nucleoplasmic helicase is required for production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins, *Mol. Cell. Biol.* 21 (2001) 7731–7746.
- [39] S.L. Hiley, T. Babak, T.R. Hughes, Global analysis of yeast RNA processing identifies new targets of RNase III and uncovers a link between tRNA 5' end processing and tRNA splicing, *Nucleic Acids Res.* 33 (2005) 3048–3056.
- [40] K.S. McKeegan, C.M. Debieux, S. Boulon, E. Bertrand, N.J. Watkins, A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly, *Mol. Cell. Biol.* 27 (2007) 6782–6793.
- [41] K.S. McKeegan, C.M. Debieux, N.J. Watkins, Evidence that the AAA+ proteins TIP48 and TIP49 bridge interactions between 15.5K and the related NOP56 and NOP58 proteins during box C/D snoRNP biogenesis, *Mol. Cell. Biol.* 29 (2009) 4971–4981.
- [42] J.G. Pelton, D.A. Torchia, N.D. Meadow, S. Roseman, Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIgIc, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques, *Protein Sci.* 2 (1993) 543–558.
- [43] N. Marmier-Gourrier, A. Clery, V. Senty-Segault, B. Charpentier, F. Schlotter, F. Leclerc, et al., A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA, *RNA.* 9 (2003) 821–838.
- [44] K.L. Cheung, J. Huen, W.A. Houry, J. Ortega, Comparison of the multiple oligomeric structures observed for the Rvb1 and Rvb2 proteins, *Biochem. Cell Biol.* 88 (2010) 77–88.
- [45] B. Rothe, R. Back, M. Quinternet, J. Bizarro, M.C. Robert, M. Blaud, et al., Characterization of the interaction between protein Snu13p/15.5K and the Rsa1p/NUFIP factor and demonstration of its functional importance for snoRNP assembly, *Nucleic Acids Res.* 42 (2014) 2015–2036.
- [46] I. Vidovic, S. Nottrott, K. Hartmuth, R. Luhrmann, R. Ficner, Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment, *Mol. Cell.* 6 (2000) 1331–1342.
- [47] N.J. Watkins, V. Segault, B. Charpentier, S. Nottrott, P. Fabrizio, A. Bachi, et al., A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP, *Cell.* 103 (2000) 457–466.
- [48] C. UniProt, UniProt: a hub for protein information, *Nucleic Acids Res.* 43 (2015) D204–D212.
- [49] T.D. Goddard, D.G. Kneller, in: University of California SF (Ed.), *Sparky 3*, 2002.
- [50] P. Guntert, Automated NMR structure calculation with CYANA, *Methods Mol. Biol.* 278 (2004) 353–378.
- [51] B. Wang, S.L. Alam, H.H. Meyer, M. Payne, T.L. Stemmler, D.R. Davis, et al., Structure and ubiquitin interactions of the conserved zinc finger domain of Npl4, *J. Biol. Chem.* 278 (2003) 20,225–20,234.
- [52] A.J. Nederveen, J.F. Doreleijers, W. Vranken, Z. Miller, C.A. Spronk, S.B. Nabuurs, et al., RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank, *Proteins.* 59 (2005) 662–672.
- [53] W.L. DeLano, The PyMOL Molecular Graphics System, Version 1.3r1. The PyMOL Molecular Graphics System, Version 13r1, Schrödinger, LLC, New York, 2010.
- [54] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Res.* 42 (2014) W320–W324.
- [55] A. Bar, V. Marchand, G. Khoury, N. Dreumont, A. Mouglin, N. Robas, et al., Structural and functional analysis of the Rous sarcoma virus negative regulator of splicing and demonstration of its activation by the 9G8 SR protein, *Nucleic Acids Res.* 39 (2011) 3388–3403.



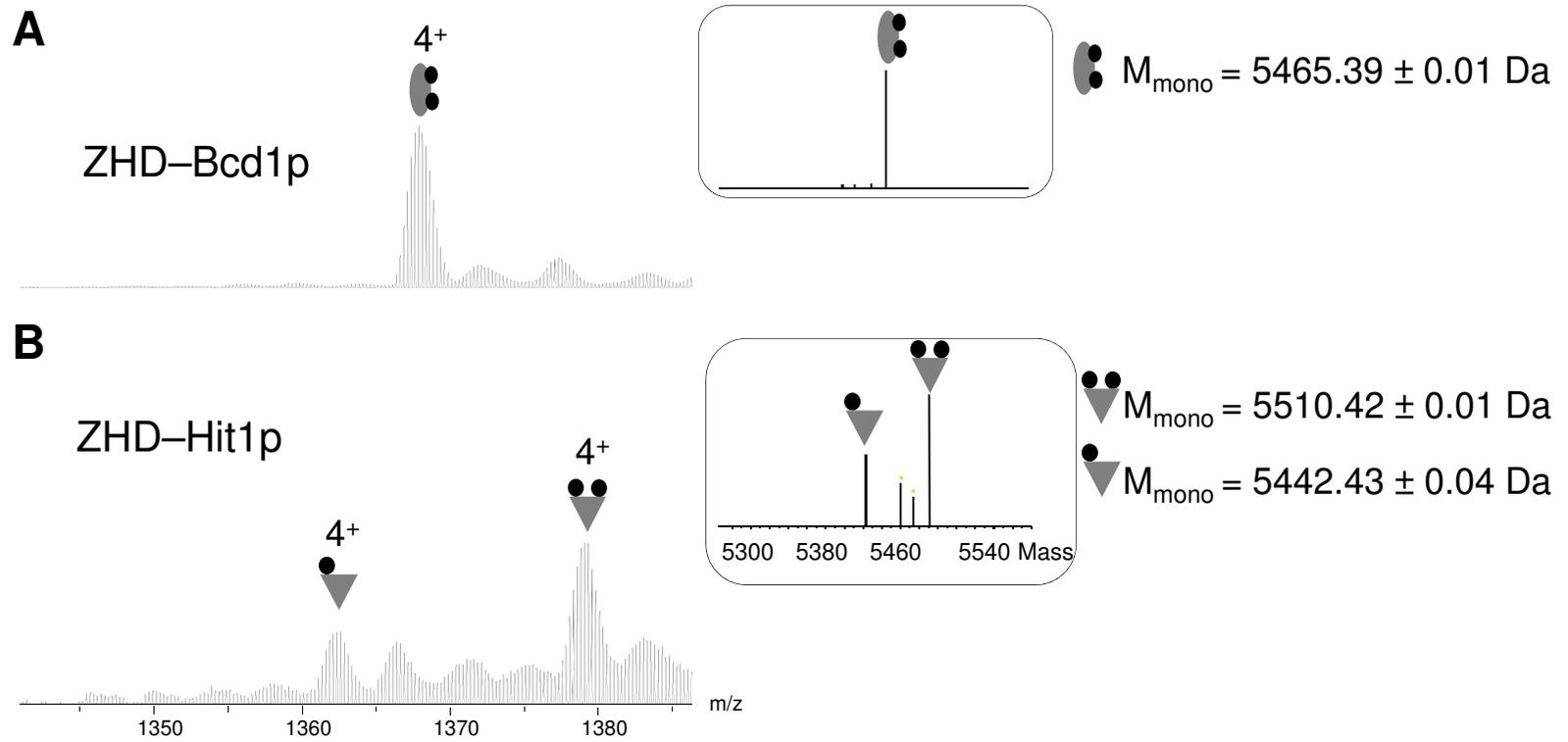
Supplementary Fig. S1. Amino-acid alignments built with Clustal Omega using sequences of the human (BCD1) and the yeast BCD1 (Bcd1p) proteins [1]. The numbering of the human protein and the ZHDs are indicated. Secondary structure predictions obtained with PsiP...

Benoit Bragantini, Decebal Tiotiu, Benjamin Rothé, Jean-Michel Saliou, H el ene Marty, Sarah Cianf erani, Bruno Charpentier, Marc Quinternet, Xavier Manival

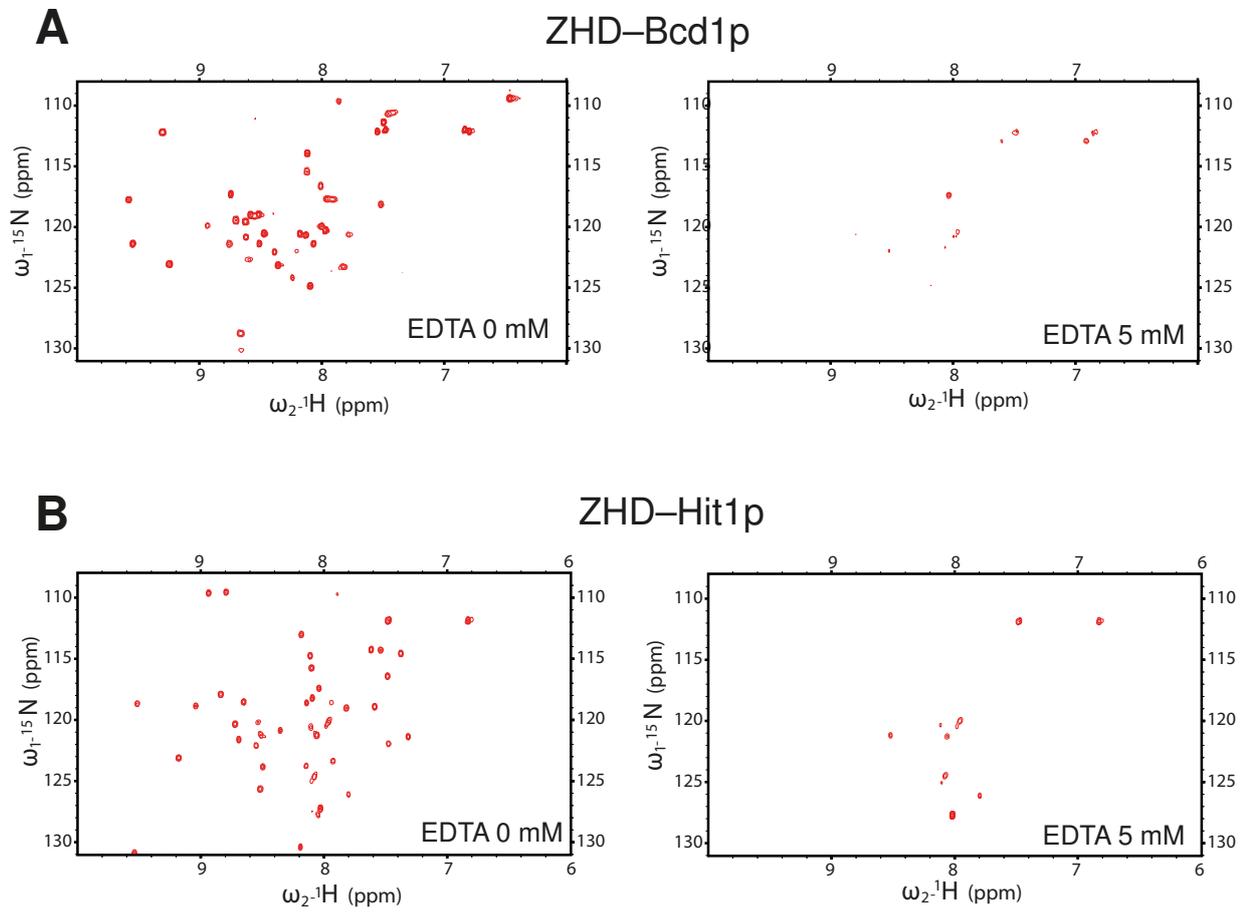
Functional and Structural Insights of the Zinc-Finger HIT protein family members Involved in Box C/D snoRNP Biogenesis

Journal of Molecular Biology, Volume 428, Issue 11, 2016, 2488–2506

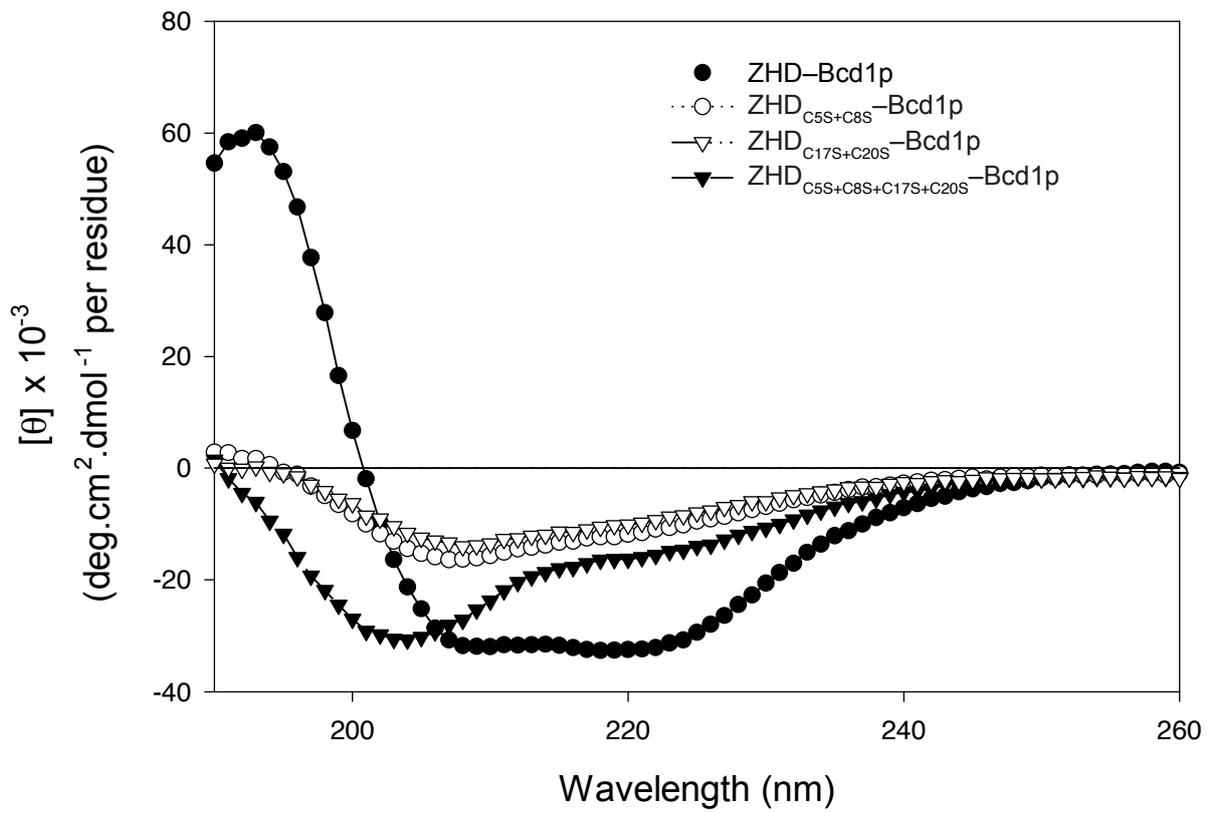
<http://dx.doi.org/10.1016/j.jmb.2016.04.028>



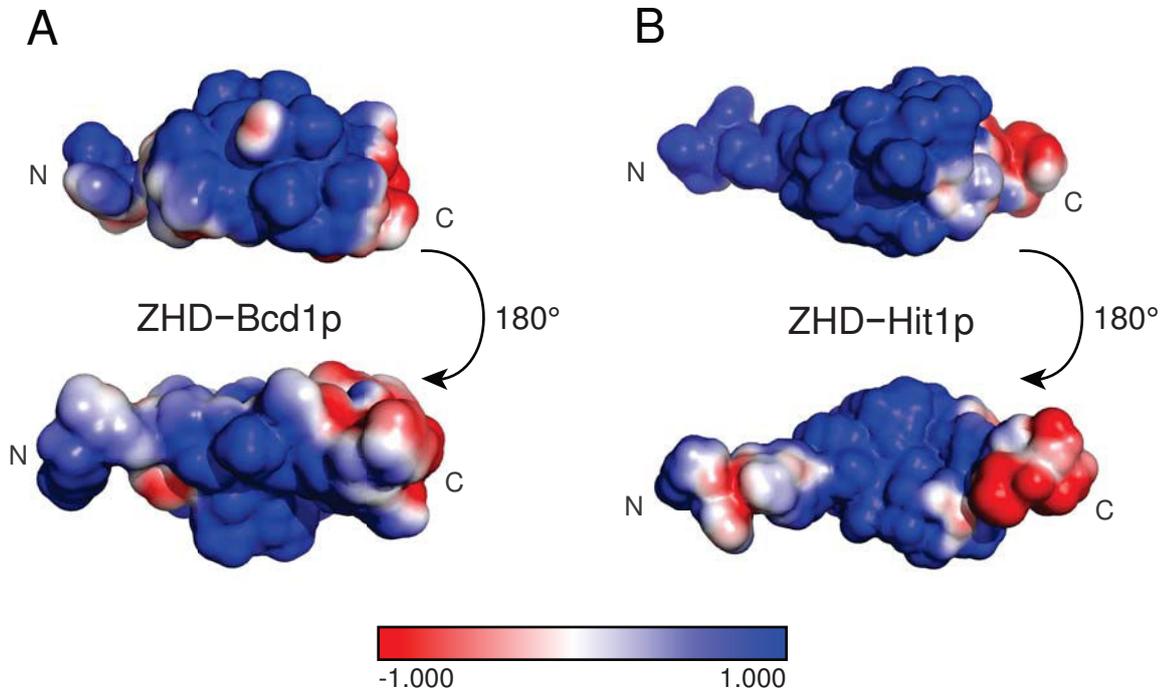
Supplementary Figure S2



Supplementary Figure 3



Supplementary Figure S4



Supplementary Figure 5

1.3. Résultats complémentaires à la publication

1.3.1. Coordination des atomes de zinc

Si les cystéines ne peuvent coordiner les atomes de zinc que par leur S_{γ} , les histidines, elles, ont deux possibilités : le $N_{\delta 1}$ ou le $N_{\epsilon 2}$ du cycle imidazole (Figure 35). L'identification de l'atome coordonnant le zinc a été réalisée par RMN en suivant la méthode décrite par (Pelton *et al.*, 1993) par l'utilisation de spectres HSQC 1H - ^{15}N *long range*. L'empreinte spectrale des histidines sur ce type de spectre reflète l'état de protonation des atomes d'azote du cycle imidazole. Cette expérience a montré que les histidines coordonnant les atomes de zinc des protéines Bcd1 et Hit1 le font par l'intermédiaire de leur azote $N_{\epsilon 2}$ (Figure 35). Selon la littérature, la majorité des domaines en doigt à zinc présentant une coordination par une histidine le font *via* le $N_{\epsilon 2}$ (Barraud *et al.*, 2012).

1.3.2. Etude des ZHD par dichroïsme circulaire

L'ajout d'EDTA (5 mM) au ZHD de Bcd1p se traduit par une déstructuration de la protéine visible sur son spectre de dichroïsme circulaire (Figure 36A). Le variant C5S+C8S+C17S+C20S du ZHD de Bcd1p présente un spectre de dichroïsme circulaire très similaire à celui du ZHD de Bcd1p en présence d'EDTA (Figure 36B). Ceci valide le fait que ce variant est totalement déplié. Il est donc un bon candidat pour tester l'importance fonctionnelle du repliement en doigt à zinc du ZHD de Bcd1p. Grâce aux expériences menées sur ce variant, nous avons pu déterminer que le rôle du ZHD pour la protéine Bcd1 ne repose pas uniquement sur un motif particulier de la séquence primaire. En effet, le variant C5S+C8S+C17S+C20S réduit la croissance des levures à haute température et conduit à une forte diminution du taux de snoRNA à boîtes C/D, malgré une faible modification de la structure primaire (Figure 4 de la publication).

1.3.3. Domaine minimal fonctionnel de Bcd1p

La région minimale de Bcd1p permettant de restaurer la croissance dans une souche *tetO₇::BCD1* a été déterminée au laboratoire par des tests de complémentation d'un défaut de croissance cellulaire à partir d'une série de fragments de la protéine. Ceux-ci ont été définis sur la base d'analyses de prédictions de structures secondaires (Figure 2 de la publication). Une première région correspondant aux résidus 1 à 168 a été identifiée comme pouvant porter

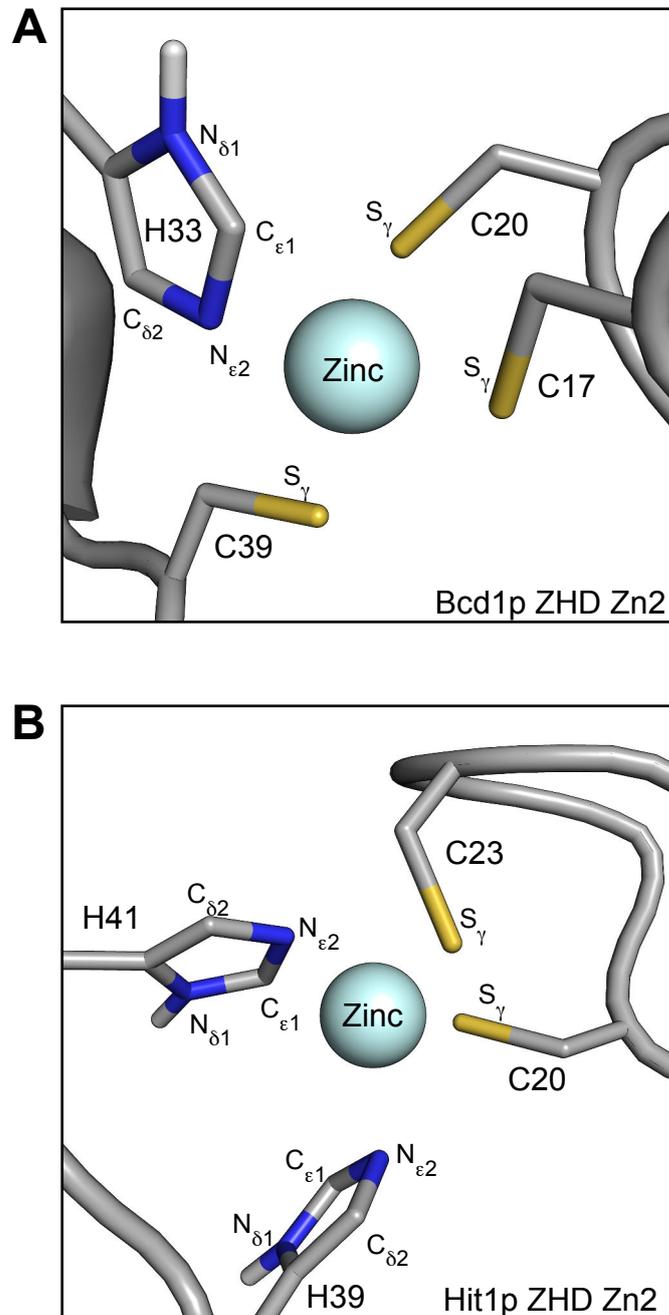


Figure 35. Coordination du zinc par les ZHD des protéines Bcd1 (A) et Hit1 (B). Zoom sur les atomes de zinc du deuxième motif en doigt à zinc de chacune des deux protéines (respectivement codes PDB : 2N94 et 2N95). Les résidus Cystéine et Histidine impliqués dans la coordination de l'atome de zinc sont représentés en bâtonnets (chaînes latérales uniquement). Les atomes de soufre sont colorés en jaune et les atomes d'azote en bleu. La coordination des atomes de zinc se fait par l'intermédiaire de l'azote N ϵ 2 de chaque Histidine (non-protoné).

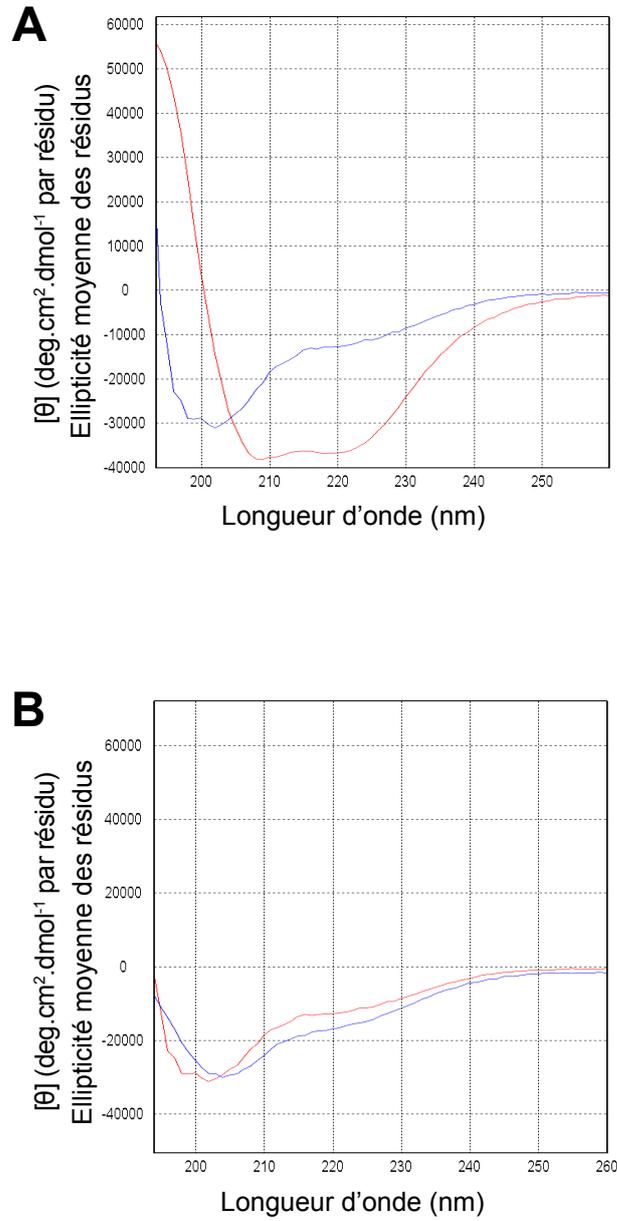


Figure 36. Analyse de l'impact de la double mutation sur le ZHD de la protéine Bcd1. (A) Spectres de dichroïsme circulaire du ZHD de Bcd1p sans (courbe rouge) ou avec 5 mM d'EDTA (courbe bleue). (B) Comparaison des spectres de dichroïsme circulaire du ZHD de Bcd1p sauvage avec 5 mM d'EDTA (courbe rouge) et du mutant C5S+C8S+C17S+C20S sans EDTA (courbe bleue).

la fonction essentielle de Bcd1p en permettant de maintenir la viabilité cellulaire et la stabilité des snoARN (Figure 2 de la publication). Malgré sa capacité à restaurer la croissance, il a été impossible de produire ce fragment *in vitro* de façon recombinante chez *E. coli*. Afin de pouvoir travailler à partir d'une construction exprimant une protéine soluble et stable, nous avons déterminé une séquence plus adaptée. Pour cela, nous nous sommes dans un premier temps basé sur un alignement entre les séquences protéiques de Bcd1p de la levure *Saccharomyces cerevisiae* et de son homologue hBCD1 humaine (Figure 37A). Il apparaît sur cet alignement que des séquences supplémentaires sont insérées chez la levure par rapport à la séquence humaine (60-68, 95-99, 115-129, 174-181, 249-266 et 320-366 principalement). Nous avons alors réduit la taille du fragment 1-168 en prenant en compte cet alignement. Ainsi, nous avons choisi d'exprimer un nouveau fragment allant des résidus 1 à 115 s'affranchissant de l'insertion 115-129. En se basant cette fois-ci également sur la prédiction des structures secondaires de la protéine Bcd1 (Figure 37B), une seconde construction a été créée. Le fragment produit, comprenant les résidus 1 à 96, se termine après l'hélice α 61-92, au niveau de la seconde insertion chez la levure (65-99).

Ces deux nouveaux fragments ont fait l'objet de tentatives de surexpression chez *E. coli* à partir du plasmide pnEA-3CH, ajoutant une étiquette 6xHistidines séparée de la séquence de la protéine par un site de coupure à la PreScission (Figure 34). Contrairement au fragment 1-168, ces fragments plus courts sont surexprimés sous forme soluble (Figure 38). Cependant, quatre bandes apparaissent après électrophorèse sur gel pour le fragment 1-115 (Figure 38A) et deux pour le fragment 1-96 (Figure 38B). Cela indique que même si ces deux fragments sont solubles, ils sont peu stables et subissent des clivages protéolytiques. Cette protéolyse apparaît vraisemblablement dans la région C-terminale, puisque la région N-terminale contient le ZHD dont nous savons par expérience qu'il est très stable. Un autre argument soutenant cette hypothèse est que toutes ces bandes de protéolyse présentent une migration électrophorétique plus rapide après la coupure par la PreScission de l'étiquette 6xHistidines située à l'extrémité N-terminale de Bcd1p (Figures 38A et B, indiquées par des flèches sur les pistes Coupure par la PreScission).

Les fragments Bcd1p(1-115) et Bcd1p(1-96) ont par la suite été utilisés pour réaliser des tests de complémentation de la levure *tetO₇::BCD1* (travaux réalisés par Decebal Tiotiu, Hélène Marty et Arnaud Paul au laboratoire). Les résultats obtenus montrent que ces deux fragments permettent tout comme le fragment Bcd1p(1-168) de restaurer la croissance des

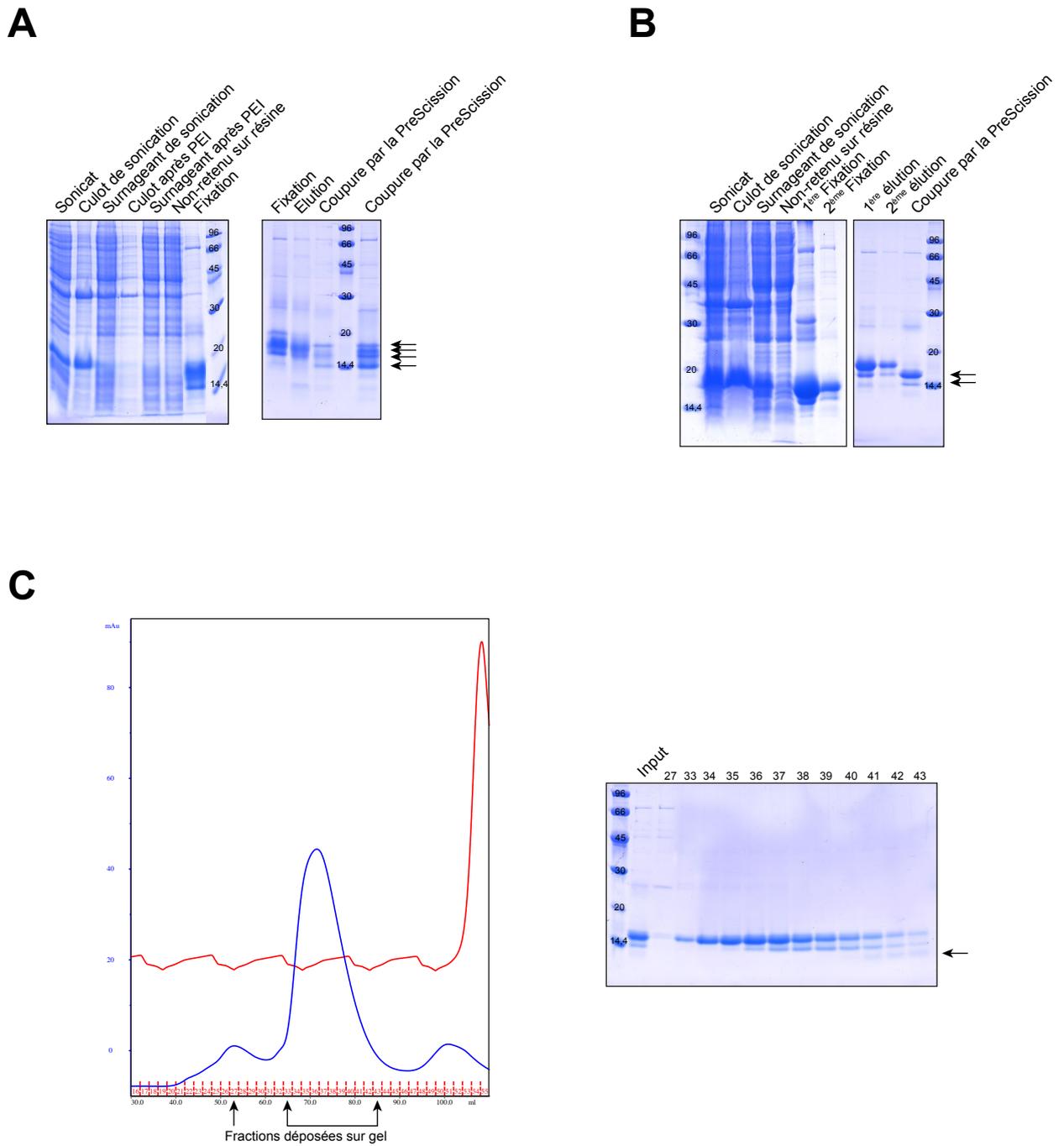


Figure 38. Purification des protéines Bcd1(1-115) et Bcd1(1-96). Etapes de purification des protéines Bcd1(1-115) (A) et Bcd1(1-96) (B) possédant toutes deux une étiquette 6xHistidines fusionnée en N-terminal analysées par SDS-PAGE 12,5%. Les flèches indiquent les bandes dont la taille est plus basse après coupure par la PreScission. (C) Chromatogramme de la chromatographie d'exclusion de taille effectuée avec une colonne Superdex S75 (16:60) sur la protéine Bcd1(1-96) et dépôt des fractions correspondantes sur SDS-PAGE 12,5%. La courbe bleue sur le chromatogramme correspond à l'UV à 280 nm et la courbe rouge à la conductimétrie. La flèche à droite du gel montre la bande la plus basse obtenue dans les fractions les plus tardives (40 à 43).

levures. Nous avons donc conclu que la région minimale importante pour la fonction de la protéine Bcd1 est comprise entre les résidus 1 et 96.

Etant donné que lors de sa purification le fragment 1-96 présente deux ou trois bandes distinctes sur gel (Figure 38C), une analyse par spectrométrie de masse a été réalisée afin de déterminer l'identité des séquences du fragment de Bcd1p de poids moléculaires plus faibles. Cette analyse a été réalisée par nos collaborateurs du laboratoire LSMBO (Laboratoire de Spectrométrie de Masse BioOrganique) de Strasbourg. Les résultats obtenus montrent qu'en plus de la séquence 1-96, un fragment correspondant à la séquence 1-81 est présent dans l'échantillon purifié (Figure 39A), ainsi qu'une troisième espèce correspondant à la séquence 1-67. Cette troisième espèce doit correspondre à la bande de plus faible poids moléculaire observée à partir de la fraction 40 (Figure 38C). La possibilité que le fragment 1-81 puisse lui aussi compléter le défaut de croissance de la levure *tetO₇::BCD1* n'a pas encore été vérifiée. Une analyse par RMN de la protéine Bcd1(1-96) a été réalisée. Le spectre HSQC ¹H-¹⁵N de cet échantillon montre une hétérogénéité dans l'intensité des pics (Figure 39B). En effet, une bonne moitié des pics est plus intense que le reste. Après analyse, le spectre HSQC ¹H-¹⁵N du ZHD de Bcd1p se superpose parfaitement avec une partie de celui du fragment 1-96 (les tâches correspondant aux résidus du ZHD sont symbolisées par des flèches sur la Figure 39B). Ceci signifie que l'environnement chimique du ZHD ne change pas entre la construction 1-45 et la construction 1-96. De plus, les tâches de corrélation les plus intenses retrouvés sur le spectre HSQC ¹H-¹⁵N de Bcd1p(1-96) ne sont pas présentes sur le spectre du ZHD. Elles appartiennent donc à la région 46-96 et sont d'ailleurs toutes situées au centre du spectre (entre 7,5 et 8,5 ppm). Ces données complétées par une analyse de la flexibilité des résidus par RMN prouve que la région 46-96 n'est pas repliée dans le contexte du fragment Bcd1p(1-96), ce qui est probablement la raison pour laquelle cette séquence se dégrade en plusieurs fragments.

1.3.4. Analyse de la stabilité de la protéine Bcd1

Afin de savoir si le rôle du ZHD de Bcd1p ne serait pas dans la stabilisation de la protéine à haute température, les différents fragments de Bcd1p ont été analysés par DSC. Les résultats de DSC montrent que la protéine Bcd1 entière n'est pas très stable, avec un T_m d'environ 46°C et une apparition de précipité (identifiable par une brutale chute du signal Cp en dessous de 0) vers 60°C (Figure 40). La perte du ZHD n'a pas d'effet sur la stabilité de la protéine. Le doigt à zinc, lui, reste stable à très haute température. Ceci n'est pas étonnant en

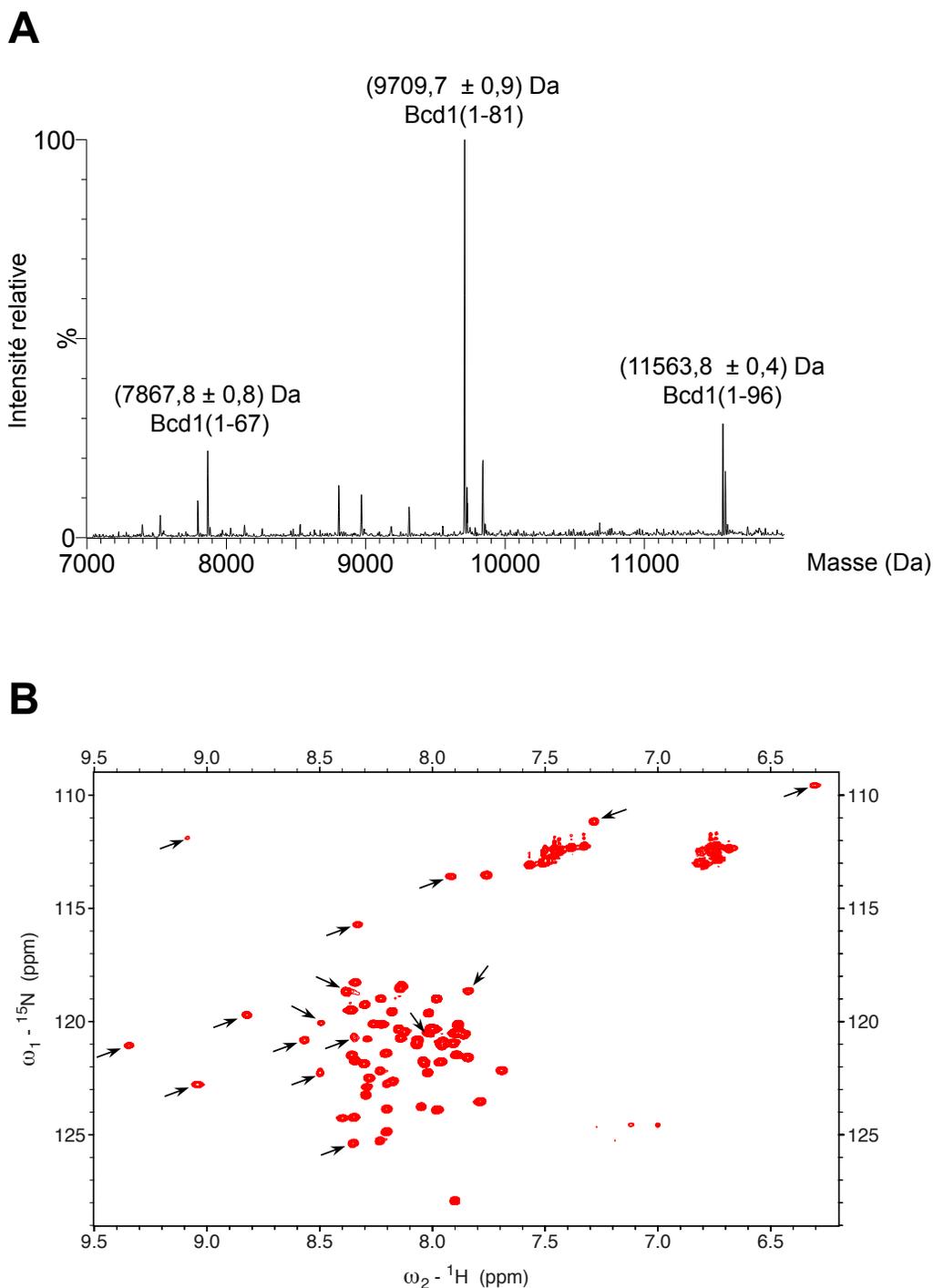


Figure 39. Analyse physico-chimique de la protéine Bcd1(1-96). (A) Spectrométrie de masse sur l'échantillon Bcd1(1-96). Les trois pics majoritaires correspondent à trois espèces de masses distinctes en solution. Les masses observées de ces trois espèces sont indiquées au-dessus de chaque pic, avec la séquence compatible de Bcd1p déterminée à partir de ces masses. (B) Spectre HSQC ${}^1\text{H}$ - ${}^{15}\text{N}$ enregistré à 293K de la protéine Bcd1(1-96) préalablement produite avec un marquage uniforme à l'azote 15. Les flèches indiquent des tâches correspondant à des résidus de la région 1-45 (ZHD). Le spectre a été enregistré en tampon Phosphate (NaPi) à 10 mM à pH 6,4 avec 150 mM de NaCl et 1 mM de DTT, sur la protéine Bcd1(1-96) concentrée à 150 μM . Ce spectre a été analysé avec le logiciel Sparky (Goddard et Kneller, 2002).

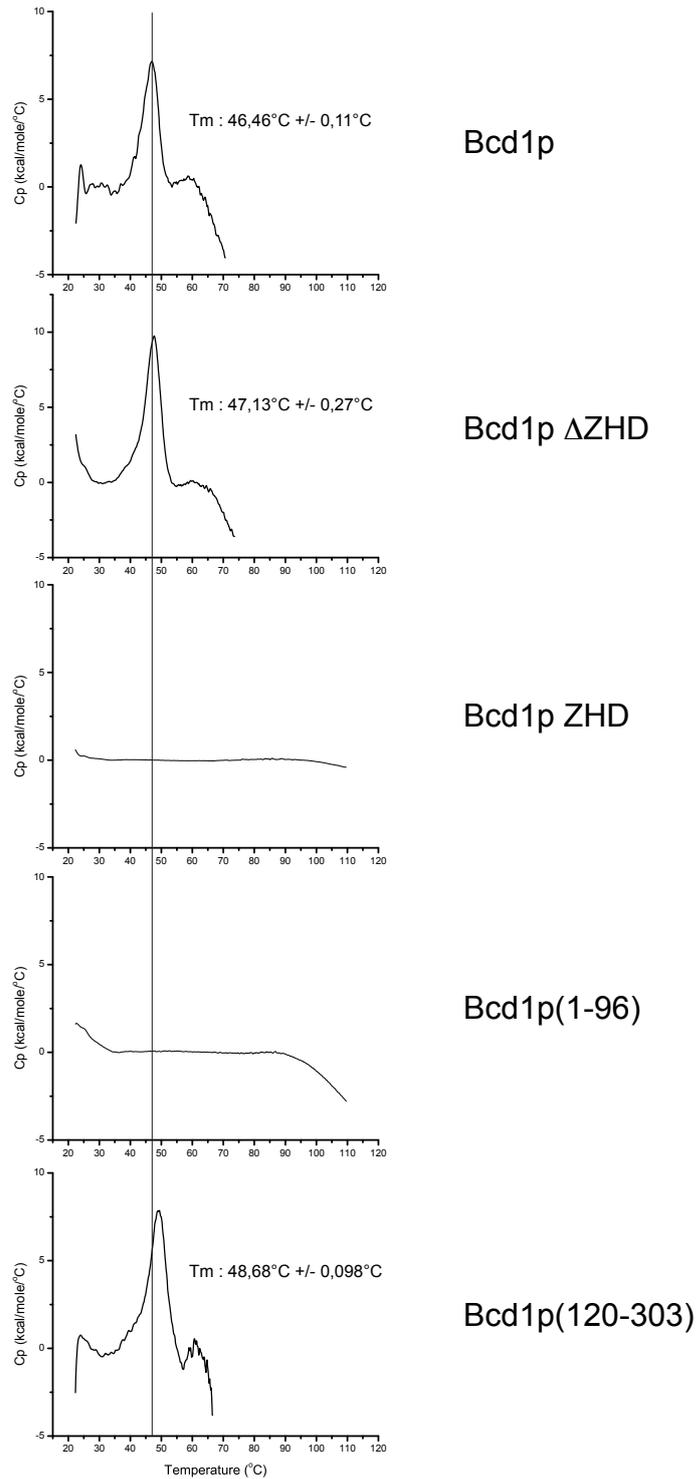


Figure 40. Analyse de la stabilité de la protéine Bcd1 et de ses fragments par expériences de DSC. Les fragments testés sont indiqués à la droite de chaque graphe. Les pics correspondent à la température de fusion (Tm) mesurée de la construction. Le trait vertical indique la valeur de Tm pour la protéine Bcd1 entière. Une forte baisse du Cp en dessous de 0 indique que l'échantillon précipite.

considérant que la structure de ce domaine est très compacte et que les expériences RMN ont été réalisées à 40°C pendant plusieurs jours d'affilée sans perte importante de signal. Le fragment 1-96, qui contient le ZHD plus quarante résidus, est lui aussi très stable et ne commence à précipiter qu'à 90°C. Au final, le ZHD est donc très stable mais ne semble pas conférer de stabilité supplémentaire au reste de la protéine. Il est donc possible que le ZHD se comporte comme un module structuralement indépendant. Enfin, le fragment correspondant à la région 120-303 présente un profil similaire à celui de la protéine entière avec un T_m comparable (autour de 47°C).

1.4. Les ZHD sont des modules d'interaction avec les protéines RuvBL1:2

En prenant en compte l'interaction directe connue entre les protéines humaines BCD1 et RuvBL2 (McKeegan *et al.*, 2007; McKeegan *et al.*, 2009), ainsi que la propension qu'ont les protéines dotées d'un ZHD à se retrouver au sein de complexes comprenant les protéines RuvBL1 et RuvBL2 (RuvBL1/2), nous avons voulu vérifier si les ZHD sont des modules d'interaction avec les protéines RuvBL1 et RuvBL2. Pour cela nous avons travaillé en collaboration avec l'équipe de P. Matias (Laboratoire IBET, Oeiras, Portugal), qui nous a fourni les protéines humaines RuvBL1 et RuvBL2 qui sont les orthologues des protéines Rvb1 et Rvb2 de levure. Deux approches ont été utilisées, l'une par RMN au laboratoire, et l'autre réalisée par l'équipe de P. Matias par la technique de Blitz qui est basée sur la résonance plasmonique de surface (SPR).

1.4.1. Identification d'une interaction entre ZHD et RuvBL par RMN

Pour tester l'interaction entre les ZHD et les protéines RuvBL1/2, nous avons utilisé des ZHD marqués uniformément au ^{13}C et utilisé les propriétés des groupements méthyle ($-\text{CH}_3$). Ceux-ci présentent des pics plus facilement isolables par rapport aux groupements $-\text{CH}$ ou $-\text{CH}_2$ par exemple, de par leur déplacement chimique élevé et leur faible abondance dans les protéines. De plus, le fait que trois atomes d'hydrogène chimiquement équivalents soient présents sur ces carbones augmente fortement l'intensité des pics et permet de travailler à plus faibles concentrations.

Des tests d'interaction entre les protéines RuvBL1/2 humaines et différents ZHD issus de protéines humaines ou de levure ont donc été réalisés. Dans un premier temps, le ZHD de la protéine Bcd1 ou celui de son homologue humain BCD1 (ZNHIT6) ont été testés contre la protéine RuvBL2 ou le complexe RuvBL1/2. Les résultats de ces expériences montrent une

forte modification des spectres RMN observés avec la protéine (ZHD) seule (marquée au ^{13}C) et ceux observés avec la même protéine à laquelle ont été ajoutés la protéine RuvBL2 ou le complexe RuvBL1/2 non marqués (Figure 41). Le fait que les spectres soient modifiés suite à l'ajout des protéines RuvBL indique que les protéines Bcd1 ou BCD1 interagissent, par le domaine ZHD, avec la protéine RuvBL2 seule ou le complexe RuvBL1/2. Toutefois, le signal RMN ne disparaît pas totalement comme nous aurions pu nous y attendre compte tenu de la taille théorique très importante d'un complexe constitué du ZHD et des RuvBL. Cela pourrait éventuellement être dû à des propriétés dynamiques du domaine d'interaction présent sur les protéines RuvBL. Lorsque la même expérience est réalisée avec le domaine 120-303 de la protéine Bcd1 ne contenant pas le ZHD, aucune modification n'apparaît sur le spectre RMN enregistré pour Bcd1p seule ou pour Bcd1p en présence du complexe RuvBL1/2 (Figure 42A). Ce résultat négatif indique une absence d'interaction entre les deux entités et démontre l'interaction directe entre le ZHD des protéines Bcd1 ou BCD1 avec les protéines RuvBL2 et RuvBL1/2. Le fait qu'une interaction est observée même entre des protéines d'organismes différents doit probablement refléter une très forte conservation évolutive des protéines Rvb. De plus, ce résultat amène le questionnement sur une potentielle conservation de l'interaction entre les protéines Rvb et des domaines ZHD d'autres protéines. Pour cela, l'interaction du ZHD de la protéine Hit1 avec le complexe RuvBL1/2 a également été testée par la même technique et donne également un résultat positif (Figure 42B). Ceci apporte donc du poids à l'hypothèse d'un lien co-évolutif entre les protéines Rvb/RuvBL et l'un de ses cofacteurs, le doigt à zinc de la famille ZHD.

Enfin, pour vérifier que le domaine ZHD n'interagit pas de façon non spécifique avec n'importe quelle protéine, un test a été réalisé avec le complexe Snu13p:Rsa1p(230-375):Hit1p dont le poids moléculaire (environ 50 kDa) est proche de celui d'un monomère de Rvb1p ou Rvb2p. Cette expérience ne montre pas de différence significative entre les deux spectres RMN (Figure 42C), ce qui démontre la spécificité d'interaction entre le ZHD et les protéines RuvBL. Pour finir, afin de nous assurer qu'il n'existait pas d'interaction intra-protéine entre le ZHD de Bcd1p et sa région C-terminale, un test a été réalisé entre ce ZHD et le domaine 120-303 de Bcd1p (Figure 42D) qui ne montre aucune différence significative entre les deux spectres RMN. Le ZHD de Bcd1p n'interagit donc pas avec la région C-terminale de Bcd1p.

Les interactions précédemment identifiées entre les ZHD des protéines Bcd1 et Hit1 et les protéines RuvBL2 ou le complexe RuvBL1/2 nous ont conduit à penser que les ZHD pourraient être des modules généraux d'interaction avec les protéines RuvBL. Pour vérifier

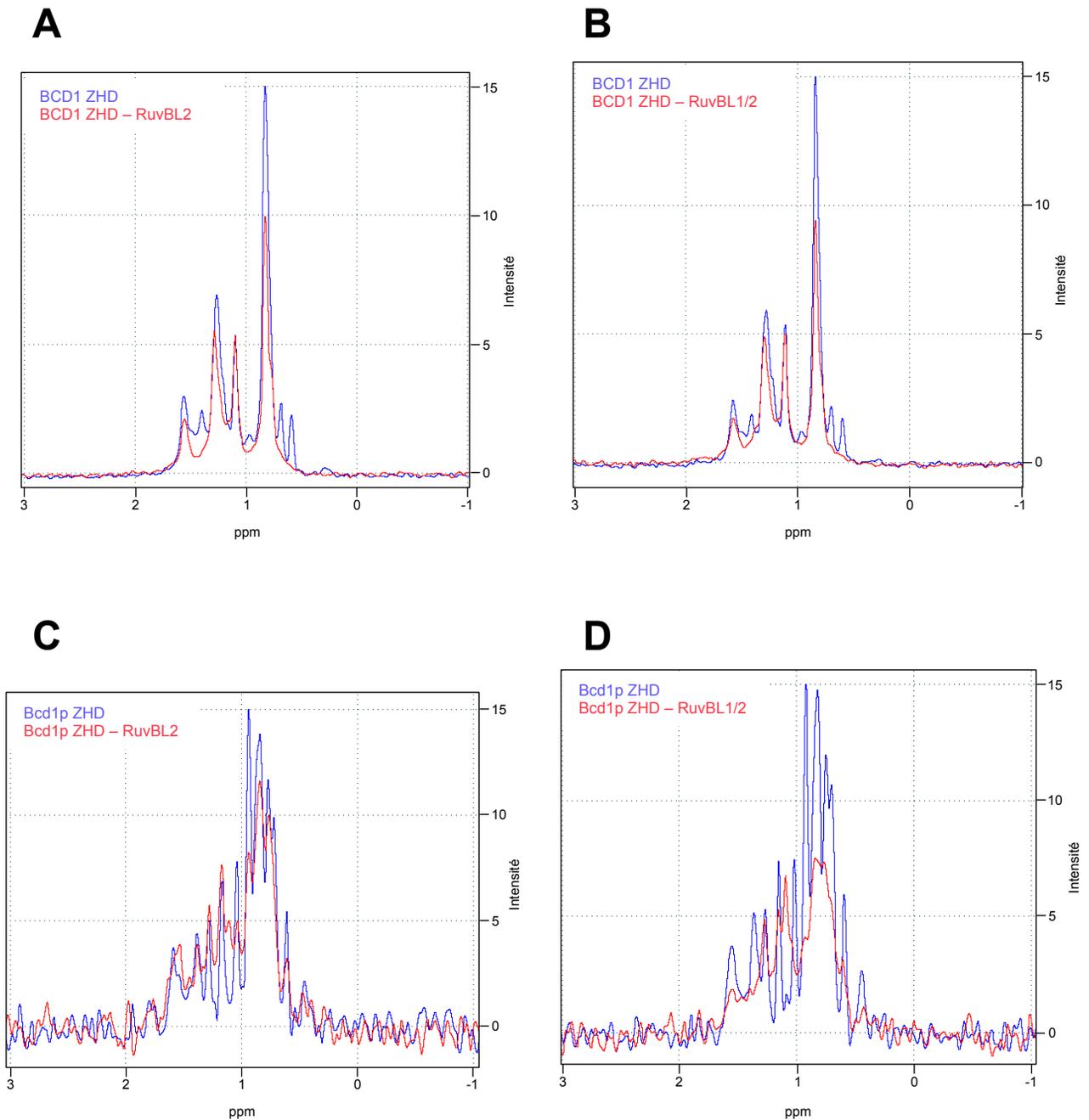


Figure 41. Tests d'interaction suivis par RMN entre les ZHD des protéines BCD1 et Bcd1 et les protéines RuvBL. Les spectres RMN 1D présentés ont été enregistrés sur les protons portés par des atomes ^{13}C dont le déplacement chimique se situe entre 2 ppm et -1 ppm (groupements méthyle) avec le logiciel TopSpin 3.0. Les courbes bleues ont été acquises avec les ZHD de BCD1 (A) et (B) ou Bcd1p (C) et (D) seuls et les courbes rouges avec les ZHD en présence des protéines RuvBL2 (A) et (C) ou RuvBL1/2 non marquées (B) et (D).

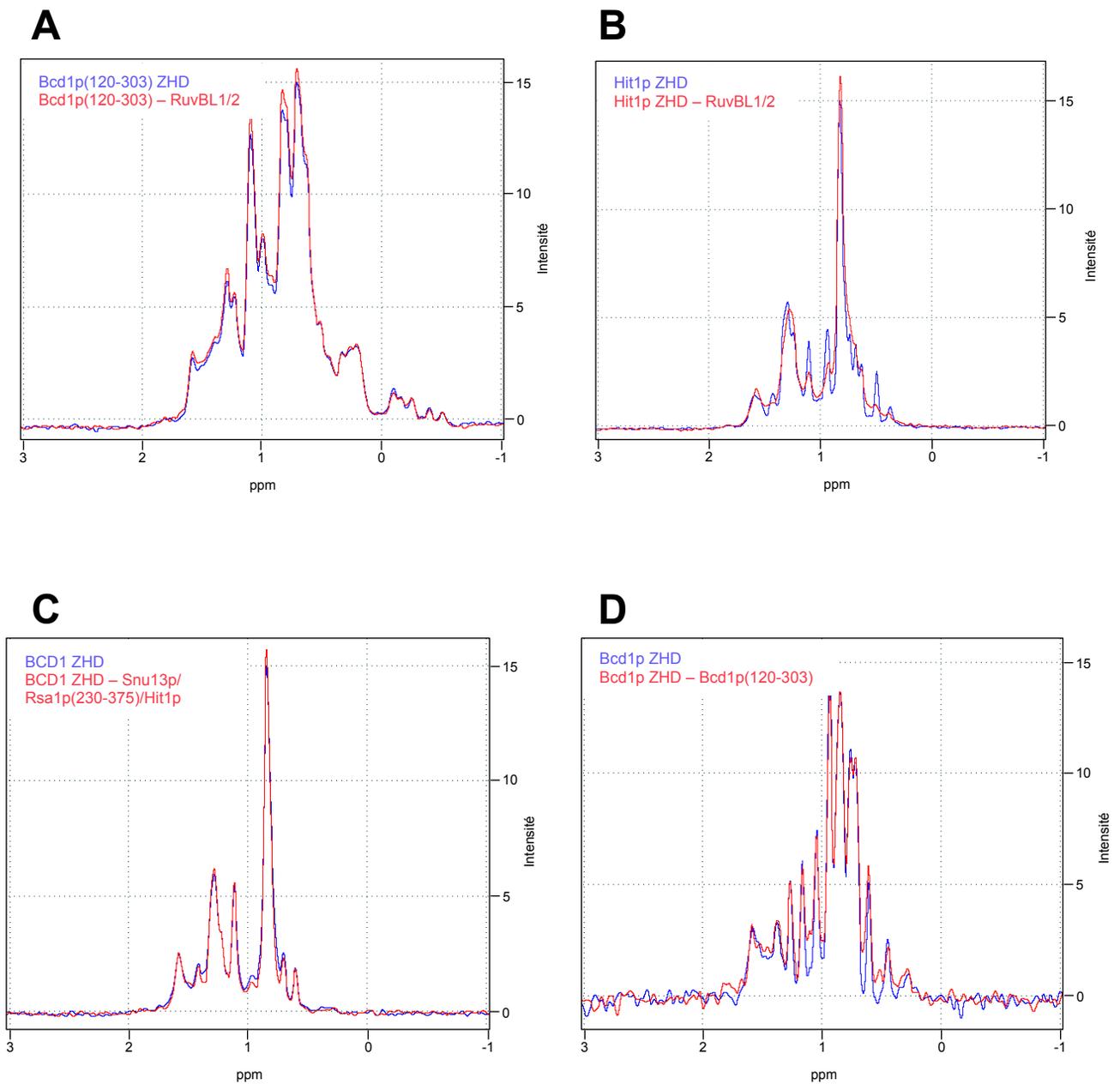


Figure 42. Tests d'interaction suivis par RMN (suite). Les spectres RMN 1D présentés ont été enregistrés sur les protons portés par des atomes ^{13}C dont le déplacement chimique se situe entre 2 ppm et -1 ppm (groupements méthyles) avec le logiciel TopSpin 3.0. Comme pour la figure 41, les courbes bleues représentent la protéine marquée au ^{13}C et les courbes rouges cette même protéine supplémentée d'une autre non marquée entre lesquelles est testée l'interaction. **(A)** Bcd1p(120-303) sans (bleu) et avec (rouge) RuvBL1/2. **(B)** Hit1p ZHD sans (bleu) et avec (rouge) RuvBL1/2. **(C)** BCD1 ZHD sans (bleu) et avec (rouge) le complexe Snu13p/Rsa1p(230-375)/Hit1p. **(D)** Bcd1p ZHD sans (bleu) et avec (rouge) Bcd1p(120-303).

cette hypothèse, de nouveaux tests ont été entrepris avec d'autres ZHD, les ZHD des protéines ZNHIT2 (double doigt à zinc) et ZNHIT5 (simple doigt à zinc) (Figure 43). Les résultats montrent que le ZHD de la protéine ZNHIT2 interagit avec le complexe RuvBL1:2 (Figure 43A), mais pas avec la protéine RuvBL1. Le spectre avec la protéine RuvBL2 présente quelques changements, mais beaucoup moins importants qu'avec le complexe RuvBL1:2. Il est donc difficile de conclure quant à une interaction entre les protéines ZNHIT2 et RuvBL2 par cette technique. Les mêmes expériences ont été menées avec le ZHD de la protéine ZNHIT5 (Figure 43B), mais aucune modification n'apparaît entre les spectres sans ou avec les protéines RuvBL. Il ne semble donc pas y avoir d'interaction entre le ZHD de la protéine ZNHIT5 et les protéines RuvBL.

1.4.2. Analyse par l'approche Blitz de l'interaction entre ZHD et RuvBL

Pour vérifier si les résultats obtenus par la méthode RMN étaient probants, nous avons envoyé des échantillons des protéines Bcd1 (levure) et BCD1 (humain) à l'équipe de P. Matias, qui a pu valider par la technique de Blitz nos observations. Une interaction est en effet bien visible entre RuvBL2 et Bcd1p entière (Figure 44A) du fait de l'augmentation du signal de Blitz entre 180 s et 270 s, et non avec le fragment 120-303 de Bcd1p qui ne contient pas le ZHD. Les analyses sur les ZHD et les protéines humaines sont en cours de réalisation.

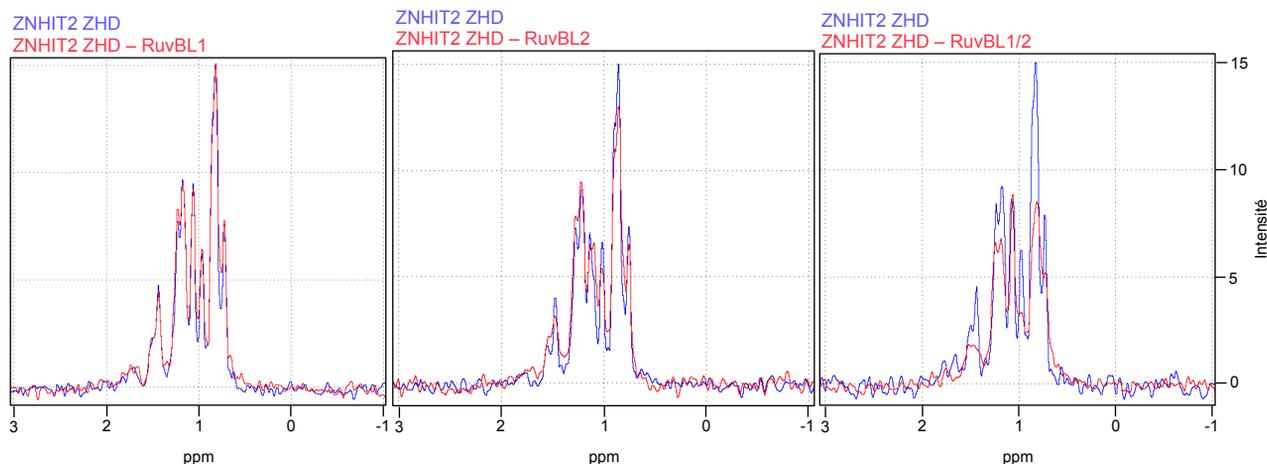
2. Recherche de partenaires associés à Bcd1p

2.1. Partenaires protéiques

2.1.1. Isolement d'un complexe Nop58p:Rvb1p:Rvb2p:Bcd1p

Des résultats préliminaires obtenus au laboratoire d'expériences chez la levure *S. cerevisiae* de TAP-Tag sur la protéine Bcd1 associées à une analyse par spectrométrie de masse (effectuée par le LSMBO de Strasbourg) ont montré une potentielle association de Bcd1p avec les protéines Nop58, Rvb1 et Rvb2 (Figure 44B) (Rothé, 2012). Compte tenu de l'interaction directe décrite entre les protéines humaines BCD1 et RuvBL2 (McKeegan *et al.*, 2007) et de nos résultats démontrant une interaction entre le ZHD de Bcd1p et la protéine RuvBL2, nous avons émis l'hypothèse de l'existence d'un complexe composé des quatre protéines Nop58:Rvb1:Rvb2:Bcd1. Nous avons voulu vérifier s'il était possible de former un tel complexe dans le système hétérologue de co-expression chez *E. coli*. Malheureusement, il nous a été impossible d'obtenir la construction génétique exprimant la protéine Nop58 entière

A



B

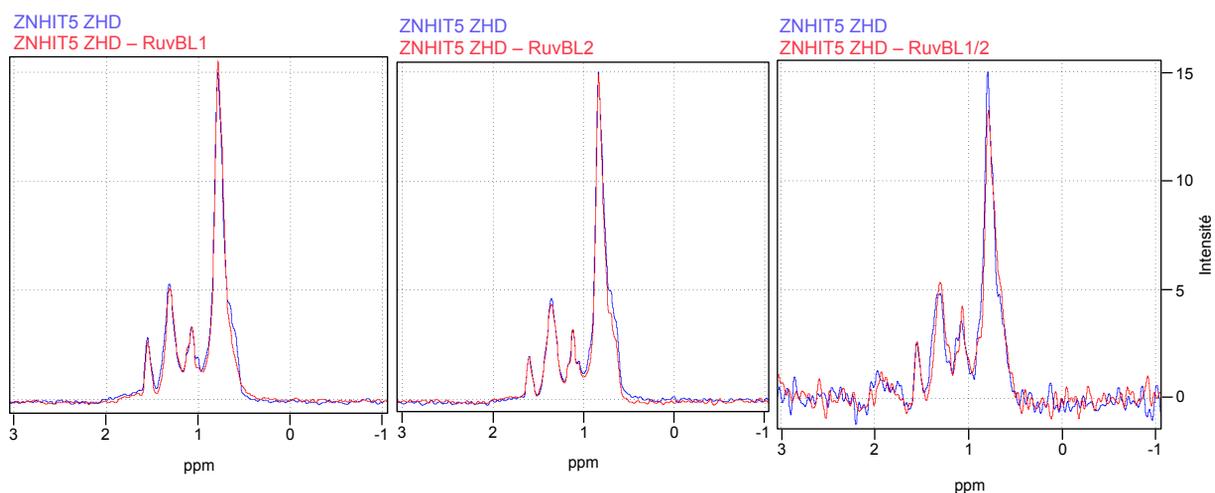
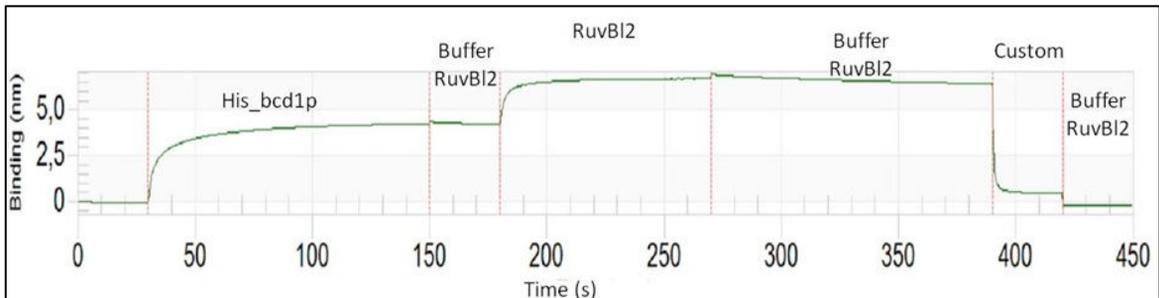


Figure 43. Tests d'interaction suivis par RMN entre les ZHD des protéines ZNHIT et les protéines RuvBL. Les spectres RMN 1D présentés ont été enregistrés sur les protons portés par des atomes ^{13}C dont le déplacement chimique se situe entre 2 ppm et -1 ppm (groupements méthyle) avec le logiciel TopSpin 3.0. Les courbes bleues ont été acquises avec les ZHD de ZNHIT2 (A) ou ZNHIT5 (B) seuls et les courbes rouges avec les ZHD en présence des protéines RuvBL1, RuvBL2 ou RuvBL1/2 non marquée.

A



B

	Protéine	Numero d'accession	Masse moléculaire (Da)	Contrôle Peptides uniques	TAP-BCD1 Peptides uniques
	Bcd1	P38772	42 316	-	2
	Rsa1	Q08932	43 981	-	-
	Hit1	P46973	18 358	-	-
Complexe R2TP	Pih1	P38768	39 503	-	-
	Tah1	P25638	12 493	-	-
	Rvb1	Q03940	50 436	-	1
	Rvb2	Q12464	51 595	-	1
snoRNP à boîtes C/D	Snu13	P39990	13 551	-	-
	Nop1	P15646	34 195	-	-
	Nop56	Q12460	56 849	-	-
	Nop58	A6ZPE5	57 176	-	2

Figure 44. Recherche de partenaires de Bcd1p. (A) Expérience de Blitz (technique basé sur la SPR) entre les protéines Bcd1 et RuvBL2, réalisée par l'équipe de P. Matias de l'IBET (Portugal). La protéine His-Bcd1 est dans un premier temps fixée sur puce par l'intermédiaire de son étiquette 6xHistidines (signal observé entre 30 et 150 s). Puis la protéine RuvBL2 est ajoutée entre 180 et 270 s. L'augmentation du signal montre une interaction entre les deux protéines. Enfin, entre 270 et 390 s du tampon est passé sur l'échantillon. Durant ce temps, le signal diminue très faiblement, ce qui nous indique que l'interaction entre les deux protéines est stable car la vitesse de dissociation est faible. (B) Analyse par spectrométrie de masse de l'association entre Bcd1p, les facteurs d'assemblage et les composants des snoRNP C/D *in vivo*. Bcd1p a été purifiée à partir de l'étiquette protA du TAP-tag et les protéines co-immunosélectionnées ont été analysées de façon exhaustive par spectrométrie de masse. La souche de levure BY4741, n'exprimant pas de protéine étiquetée, a été utilisée comme contrôle négatif. Le numéro d'accession à la base de donnée UniProtKB/Swiss-Prot est précisé pour chaque protéine (Rothé, 2012).

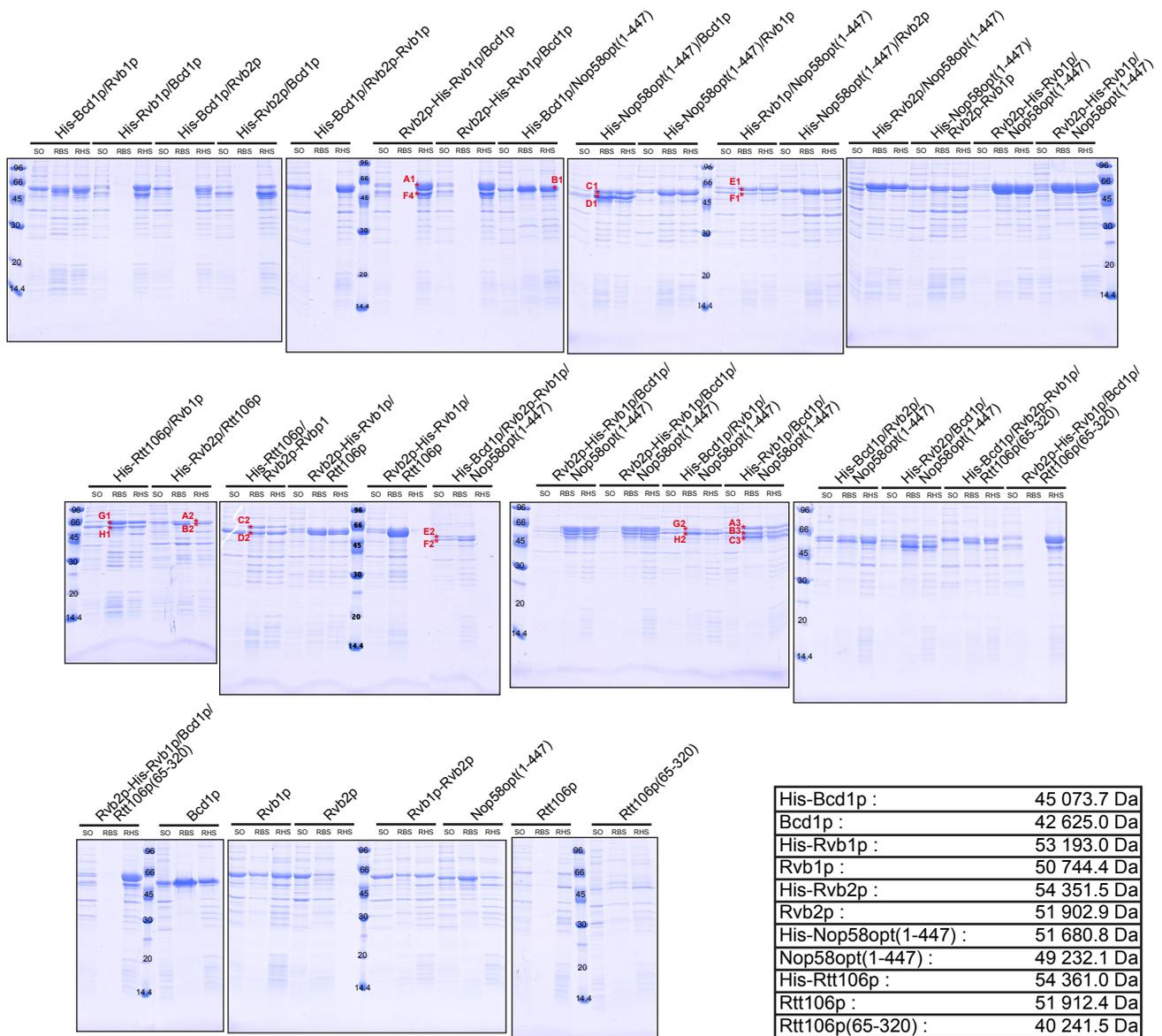
(1-511). Nous avons donc travaillé sur un variant plus court du côté C-terminal, le fragment 1-447, produit à partir d'une séquence ADN optimisée (opt) pour la surexpression chez *E. coli*.

2.1.1.1. Co-expression des protéines Nop58opt(1-447), Rvb1, Rvb2 et Bcd1

Des tests de co-expression ont donc été réalisés sur ces quatre protéines par des combinaisons d'expression deux à deux, à trois ou toutes ensemble. Les résultats de ces tests ainsi que les analyses par spectrométrie de masse sont présentés en Figure 45. Les contrôles négatifs de fixation aspécifique sur résine montrent que les protéines Rvb1 et Bcd1 restent toujours fixées à la résine, ainsi que la protéine Nop58opt(1-447) dans la condition à bas sel. Par contre, la protéine Rvb2 ne semble interagir que très peu avec la résine à bas sel. Le résultat du test dans la condition à haut sel n'a pu être établi du fait d'un problème d'expérimentation. Les résultats de cette campagne de co-expression sont donc difficilement interprétables. Il est possible toutefois qu'une interaction puisse s'établir entre les protéines Bcd1 et Nop58 (Figure 45, bandes B1, C1 et D1). Cependant, les analyses de ces bandes par spectrométrie de masse présentent des résultats incohérents, puisque la protéine Rvb1 est détectée dans la bande D1 par exemple, alors que cette protéine ne faisait pas partie de la condition testée. Les résultats de spectrométrie de masse sont probablement faussés, possiblement par une contamination de l'appareil par les échantillons précédemment analysés. Les conditions avec les quatre protéines Nop58opt(1-447), Rvb1, Rvb2 et Bcd1 présentent quatre bandes dans certaines conditions, mais là encore aucune conclusion ne peut être effectuée du fait de la forte fixation aspécifique sur la résine de certaines de ces protéines. Ces résultats restent donc à être vérifiés par d'autres méthodes, et si possible avec la protéine Nop58 entière. Par exemple, la production de protéines Bcd1 et Rvb fusionnées avec une étiquette Flag est à l'étude, afin de pouvoir travailler avec une autre résine limitant les interactions aspécifiques avec ces protéines.

2.1.1.2. Interaction via le domaine DII_{ext} de Rvb2 ?

Dans la structure hexamérique des protéines Rvb1 et Rvb2, les domaines DII_{ext} (136-242 et 132-234, respectivement) sont projetés vers l'extérieur de l'anneau hexamérique (Figure 22). Cela fait donc de ces domaines des candidats à d'éventuelles interactions avec des partenaires. Nous avons donc voulu vérifier si le domaine DII_{ext} de Rvb2p ne constituait



Bande	Protéines identifiées	Couverture de séquence %
A1	Rvb2	85
	His-Rvb1	84
F4	Bcd1	70
	His-Rvb1	51
B1	Nop58opt	80
	His-Bcd1	73
C1	Bcd1	78
	His-Nop58opt	56
D1	Rvb1	78
	His-Nop58opt	61
E1	Nop58opt	79
	His-Rvb1	51

Bande	Protéines identifiées	Couverture de séquence %
F1	Rtt106	84
G1	Rvb1	75
	His-Rtt106	71
H1	Rvb2	84
	His-Rtt106	69
A2	His-Rvb2	87
B2	Rtt106	79
	Rvb1	74
C2	Rvb2	84
	His-Rtt106	55
D2	Rvb1	71
	Rvb2	80

Bande	Protéines identifiées	Couverture de séquence %
E2	Nop58opt	81
	His-Bcd1	63
F2	Rvb1	48
	Rvb1	84
G2	Nop58opt	82
	His-Bcd1	50
H2	His-Bcd1	72
	Nop58opt	64
A3	His-Rvb1	23
B3	Rvb2	19
C3	His-Rvb1	76
	Rvb2	80
	Nop58opt	56

Figure 45. Expérience de co-expression entre les protéines Nop58opt(1-447), Rvb1 et 2, Bcd1 et Rtt106 de levure. L'analyse des résultats de co-expression est faite par dépôts sur SDS-PAGE 12,5%. Pour chaque condition testée, les dépôts sur gel correspondent au sonicat (SO), aux résines à bas sel (RBS) et à haut sel (RHS). Les concentrations en sel sont de 50 mM à bas sel et de 400 mM à haut sel. Le tableau à droite récapitule les poids moléculaires de chaque protéine utilisée dans ce test. Les abréviations His et opt signifient étiquette 6xHistidines et protéine exprimée à partir d'un gène optimisé pour la production dans *E. coli* respectivement. Le tableau en bas synthétise les résultats de spectrométrie de masse protéomique (réalisés par le LSMBO de Strasbourg).

pas le domaine d'interaction avec les protéines Bcd1 ou Nop58. Pour cela, une analyse par *His-Pull-down* a été entreprise, mais les résultats ne montrent aucune interaction directe entre le domaine DII_{ext} de Rvb2p et les protéines Bcd1 ou Nop58 (Figure 46A). En effet, lorsque la protéine Rvb2 DII_{ext} est incubée en présence d'une résine sur laquelle est déjà fixée la protéine Bcd1 ou Nop58opt(1-447), aucune trace de Rvb2p DII_{ext} ne subsiste après lavage de la résine. En ce qui concerne Nop58p, étant donné que nous n'avons utilisé que la forme tronquée 1-447, il est possible que l'interaction mette en jeu des résidus de la région C-terminale qui a été supprimée. En revanche, il paraît clair que la protéine Bcd1 n'interagit pas avec Rvb2p au niveau de son domaine DII_{ext}. Afin de vérifier si le fragment correspondant au domaine DII_{ext} de Rvb2p est correctement replié, celui-ci a été produit avec un marquage aux ¹³C et ¹⁵N et analysé par RMN. Le spectre HSQC ¹H-¹⁵N de cette protéine est caractéristique d'une protéine repliée (Figure 46B). Nous observons une bonne dispersion des pics sur le spectre avec une grande homogénéité d'intensité, même si certains pics au centre du spectre apparaissent plus intenses que les autres et appartiennent probablement à une petite région désordonnée. L'absence d'interaction observée par *Pull-down* entre ce domaine et les protéines Bcd1 et Nop58opt(1-447) n'est donc pas due à un problème de repliement de ce dernier. De plus un test d'interaction suivi par RMN a été réalisé entre le domaine DII_{ext} de Rvb2p et le ZHD de la protéine Bcd1 (dont nous savons qu'il interagit avec la protéine Rvb2 entière) et il ne montre pas non plus d'interaction (Figure 47A). Les deux spectres de la Figure 47A se superposent en effet fortement, ce qui indique qu'il n'y a pas eu de modification de l'environnement chimique des groupements méthyle du ZHD de Bcd1p après l'ajout de Rvb2p DII_{ext}, comme c'était le cas lors de l'ajout de la protéine RuvBL2 (Figure 41).

2.1.1.3. Interaction via le domaine DIII de Rvb2p ?

Le domaine DIII (allant des résidus 358 à 471) de Rvb2p présente dans son extrémité C-terminale une extension dirigée vers le solvant, appelée α -hook qui est constituée d'une hélice α et d'une queue dépliée (*hook*) (Figure 22). Cette extension n'est pas retrouvée chez Rvb1p et pourrait elle-aussi être un site de fixation potentiel pour des partenaires tel que Bcd1p. Un test d'interaction entre ce domaine DIII et le ZHD de Bcd1p a été réalisé par RMN, mais ne montre aucune interaction (Figure 47B). Cependant, le domaine DIII de Rvb2p seul, sans le domaine DI n'est pas très stable et se dégrade très rapidement. Un problème de repliement pourrait être la cause du résultat d'interaction négatif observé en RMN. Le spectre

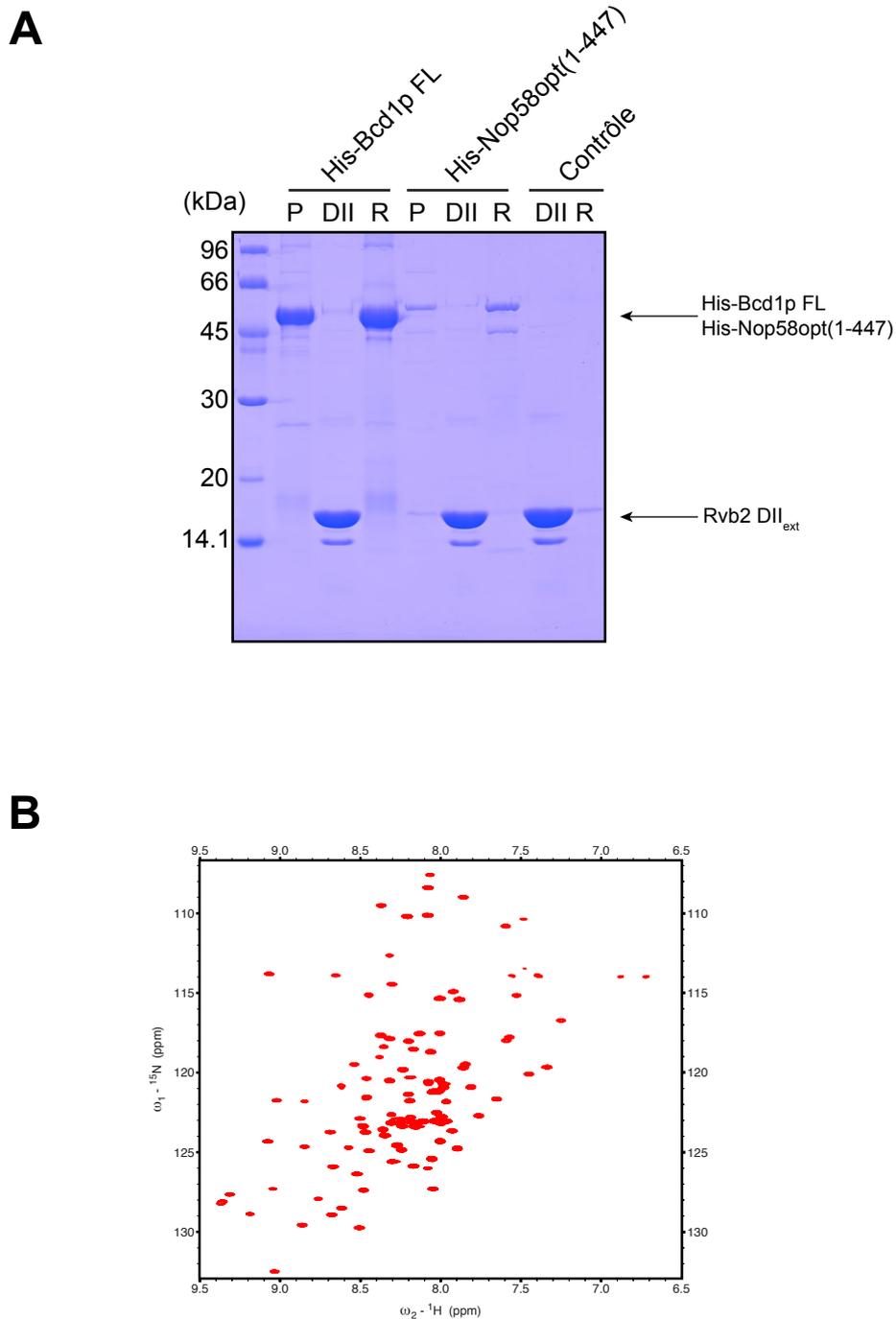


Figure 46. Analyse du domaine DII_{ext} de la protéine Rvb2 de levure. (A) His-Pull-down entre les protéines His-Bcd1 (entière) ou His-Nop58opt(1-447) et Rvb2 DII_{ext}, dans du tampon de lyse classique (voir partie matériels et méthodes). Les différents échantillons sont fractionnés sur SDS-PAGE 12,5%. La piste P correspond à la protéine avec étiquette His avant le Pull-down, la piste DII à la protéine Rvb2 DII_{ext} non étiquetée avant le Pull-down et la piste R à la résine après le Pull-down. Les protéines correspondant aux bandes sur le gel sont indiquées à droite. (B) HSQC ${}^1\text{H}$ - ${}^{15}\text{N}$ de la protéine Rvb2 DII_{ext} (concentrée à 1 mM) dans un tampon contenant 10 mM de phosphate (NaPi) à pH 6,4 et 150 mM de NaCl et 0,5 mM de TCEP.

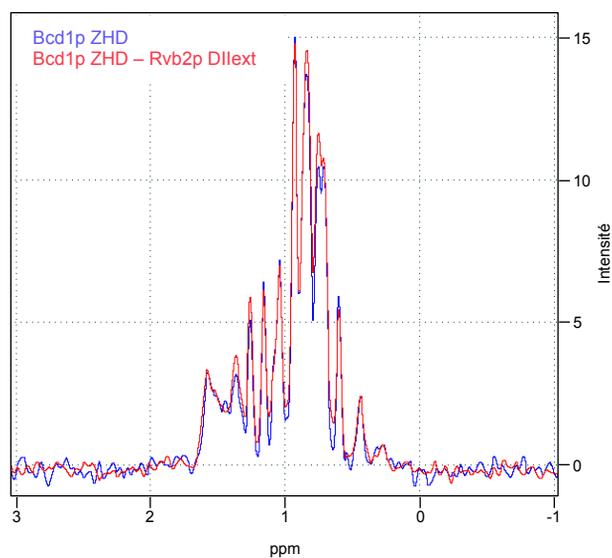
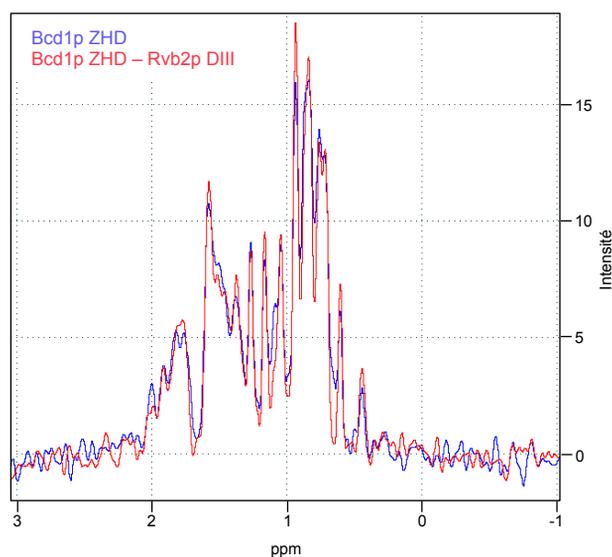
A**B**

Figure 47. Tests d'interaction suivis par RMN entre le ZHD de Bcd1p et les domaines DII et DIII de Rvb2p. Les spectres RMN 1D présentés ont été enregistrés sur les protons portés par des atomes ^{13}C dont le déplacement chimique se situe entre 2 ppm et -1 ppm (groupements méthyle) avec le logiciel TopSpin 3.0. Comme pour la figure 41, la courbe bleue représente la protéine marquée au ^{13}C et la courbe rouge cette même protéine supplémentée d'une autre non-marquée entre lesquelles est testée l'interaction. **(A)** ZHD de Bcd1p sans (bleu) et avec (rouge) Rvb2p DII_{ext}. **(B)** ZHD de Bcd1p sans (bleu) et avec (rouge) Rvb2p DIII.

HSQC ^1H - ^{15}N de ce domaine est en effet caractéristique d'une protéine dépliée avec beaucoup de pics autour de 8 ppm en proton et une forte hétérogénéité d'intensité des pics (Figure 48A). Les investigations sur le domaine d'interaction des ZHD sur les protéines Rvb doivent encore être poursuivies.

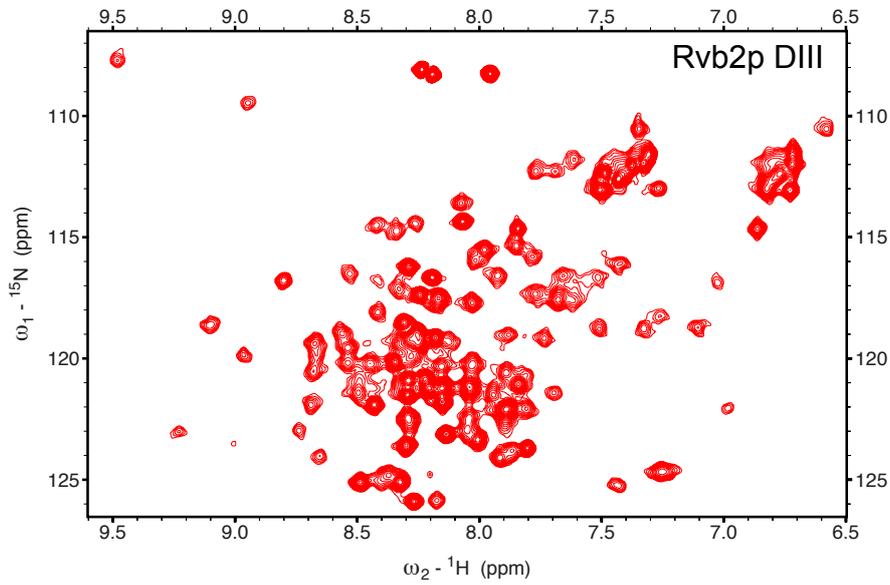
2.1.1.4. Interaction entre les protéines Rvb1 et Rvb2 par leurs domaines DII

Selon la littérature, la formation d'un macro-complexe dodécamérique entre les protéines Rvb1 et Rvb2 pourrait se faire par l'intermédiaire des domaines DII_{ext} de chacune de ces deux protéines (Niewiarowski *et al.*, 2010; Lakomek *et al.*, 2015). Nous avons donc voulu voir si une interaction directe entre ces deux domaines seuls était possible. Malheureusement, le domaine DII_{ext} de Rvb1p présente là encore un spectre HSQC ^1H - ^{15}N typique d'une protéine entièrement dépliée (Figure 48B), bien que pourtant très bien exprimé chez *E. coli* et soluble lors de la purification. Il ne nous a donc pas été possible de tester cette interaction par RMN.

2.1.2. La protéine Rtt106

Une association entre les protéines Bcd1 et Rtt106 a été mise en évidence au laboratoire par B. Rothé et D. Tiotiu suite à un test de criblage en double hybride chez la levure sur des partenaires potentiels de Bcd1p répertoriés dans les bases de données (Figure 49A). Par la suite, l'interaction directe entre ces deux protéines a été démontrée par co-expression (Figure 49B). Deux bandes sont bien visibles sur gel lors de la co-expression des deux protéines aux deux concentrations en sel et quelle que soit la protéine portant l'étiquette 6xHistidines. Une analyse protéomique des bandes sur gel nous a permis de nous assurer que ces bandes correspondent bien aux protéines d'intérêt. Plus particulièrement, cette interaction met en jeu le domaine central de la protéine Rtt106, contenant le double domaine PH (région 65-320). Nous pouvons noter que les bandes sur gel apparaissent plus intenses et avec une meilleure stœchiométrie dans le cas de la protéine Rtt106(65-320), par rapport à la protéine Rtt106 entière aux deux concentrations en sel (Figure 49B). De plus, une seule bande est visible pour la condition His-Bcd1p/Rtt106p(65-320) à 50 mM de sel, contrairement à la même condition à 400 mM de sel. Cela nous a conduit à penser qu'il était préférable de travailler à haute concentration en sel avec les protéines Bcd1 et Rtt106(65-320).

A



B

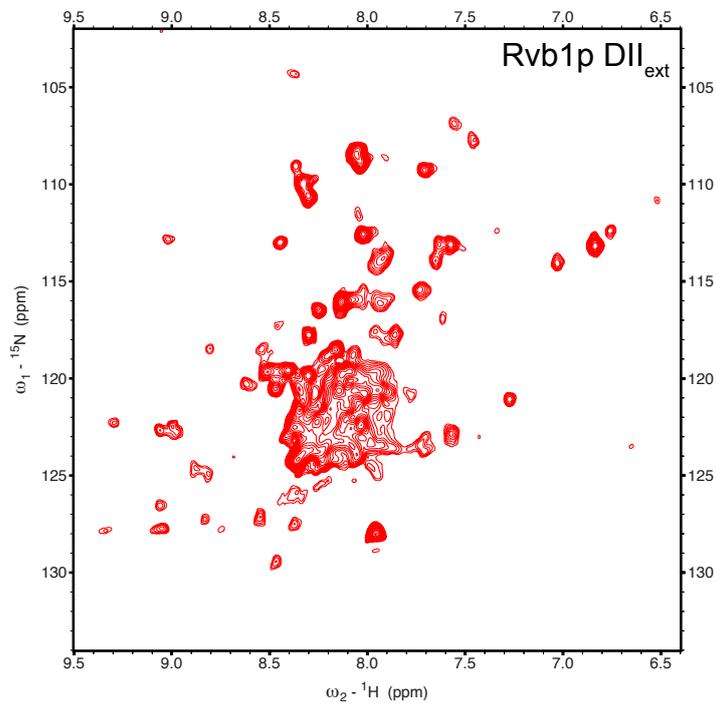
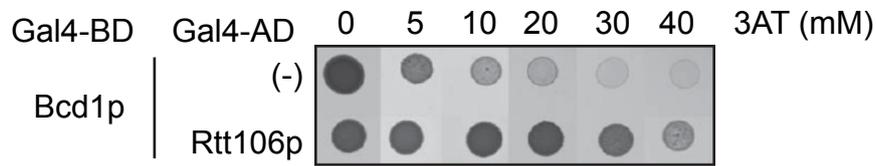


Figure 48. HSQC ${}^1\text{H}$ - ${}^{15}\text{N}$ des domaines DIII de Rvb2p (A) et DII_{ext} de Rvb1p (B). Ces expériences ont été réalisées à 298K en tampon phosphate (10 mM) à pH 6,4 et avec 150 mM de NaCl et 0,5 mM de TCEP.

A



B

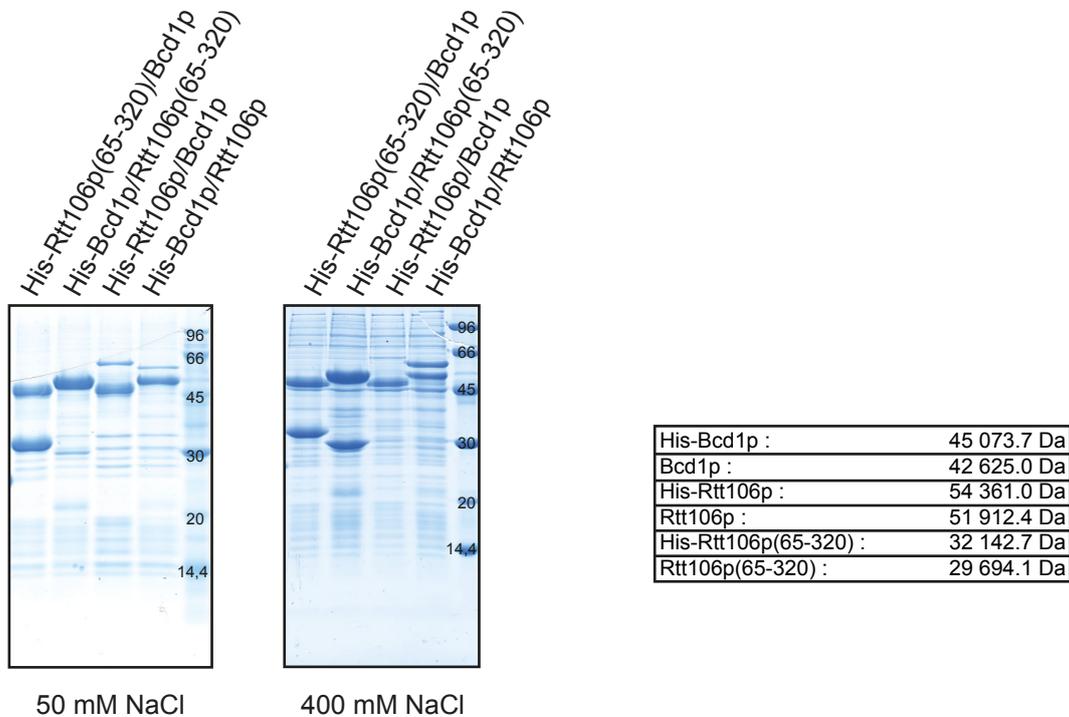


Figure 49. Identification de l'interaction entre les protéines Bcd1 et Rtt106. (A) Tests d'interaction par double-hybride entre Bcd1p et Rtt106p, respectivement exprimées à partir des vecteurs pGBKT7 (Gal4-BD) et pACT2 (Gal4-AD). Les tests d'interaction ont été réalisés par croisement des souches Y190 (pACT2) et Y187 (pGBKT7), et étalement sur milieu sélectif sans histidine, en présence d'une gamme de concentration croissante en 3-aminotriazol (Rothé, 2012). (B) Analyse des résultats de co-expression entre les protéines Bcd1 et Rtt106 par SDS-PAGE 12,5%. La protéine Rtt106 est soit entière tronquée au double domaine PH central (65-320). Les deux gels correspondent aux conditions à bas sel (gel de gauche) et à haut sel (gel de droite). Les poids moléculaires des différentes protéines utilisées dans cette expérience sont listés dans le tableau à droite.

2.2. Partenaires nucléiques

L'observation du potentiel électrostatique de surface des ZHD des protéines Bcd1 et Hit1 révèle que ces deux protéines possèdent une zone chargée positivement (voir Figure S5 de la publication). Une telle charge positive en surface est souvent le reflet d'un domaine de liaison avec des acides nucléiques. Nous avons déterminé que les ZHD de ces deux protéines n'étaient pas des modules de fixation aux acides nucléiques. En revanche, la protéine Bcd1 est quant à elle bien capable de se lier avec l'oligonucléotide poly-dIdC et avec les ARN (voir Figure 5 de la publication). La liaison avec l'ADN a également été testée entre la protéine Bcd1 ou ses différents domaines et la séquence ADN codant le snoRNA U14. Les résultats de gel retard montrent que la protéine Bcd1 n'est pas capable de se lier à cette séquence ADN (Figure 50). En effet, aucun retard n'est observé à une concentration de 2 μ M de protéine, contrairement à ce que nous avons observé avec la séquence ARN correspondante. A la concentration de 20 mM, les protéines semblent s'agréger dans les poches du gel et empêchent ainsi la migration, mais sans qu'il n'y ait pourtant d'interaction spécifique ou significative avec l'ADN.

3. Etude du complexe formé entre les protéines Bcd1 et Rtt106

Pour commencer, les deux protéines recombinantes Bcd1 et Rtt106(65-320) ont été produites et purifiées chez *E. coli*. Si la protéine Rtt106(65-320) est bien exprimée et stable (Figure 51), la protéine Bcd1 est elle aussi bien exprimée, mais a tendance à se dégrader rapidement (Figure 52). Cette dégradation est visible sur SDS-PAGE par la présence de bandes qui augmentent en nombre et en intensité au cours du temps (données non présentées) et qui correspondent à des produits de plus petite taille. Après avoir testé les deux conditions de purification conservant ou non l'étiquette 6xHistidines, il semble que la présence de cette étiquette fusionnée à la région N-terminale de la protéine améliore sensiblement la stabilité de celle-ci. Toutefois, si le fait que Bcd1p se dégrade est gênant pour réaliser des études structurales, cela est compatible avec des analyses physico-chimiques telles que celles par ITC ou dichroïsme circulaire.

3.1. Caractérisation biophysique du complexe Bcd1p-Rtt106p(65-320)

Les paramètres thermodynamiques et l'affinité de l'interaction entre la protéine Bcd1 entière (1-366) et le double domaine PH (65-320) de Rtt106p ont été déterminés par ITC sur

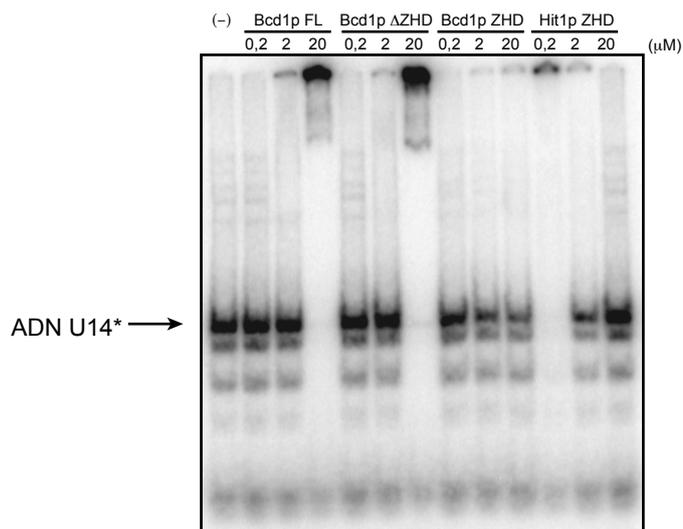
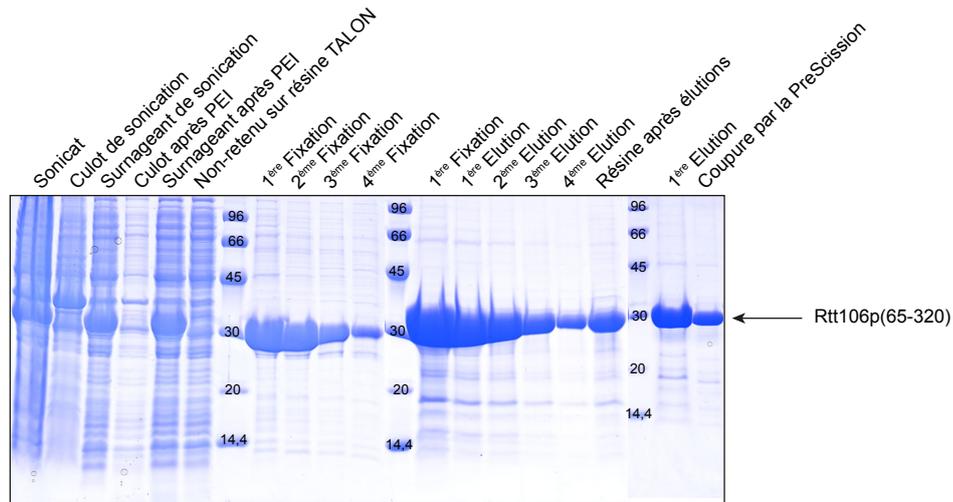


Figure 50. Bcd1p n'est pas capable de se lier à l'ADN. Expérience de retard sur gel (EMSA) avec la séquence ADN codant le snoRNA U14 radiomarquée au ³²P. L'astérisque signifie que l'ADN est radiomarqué.

A



B

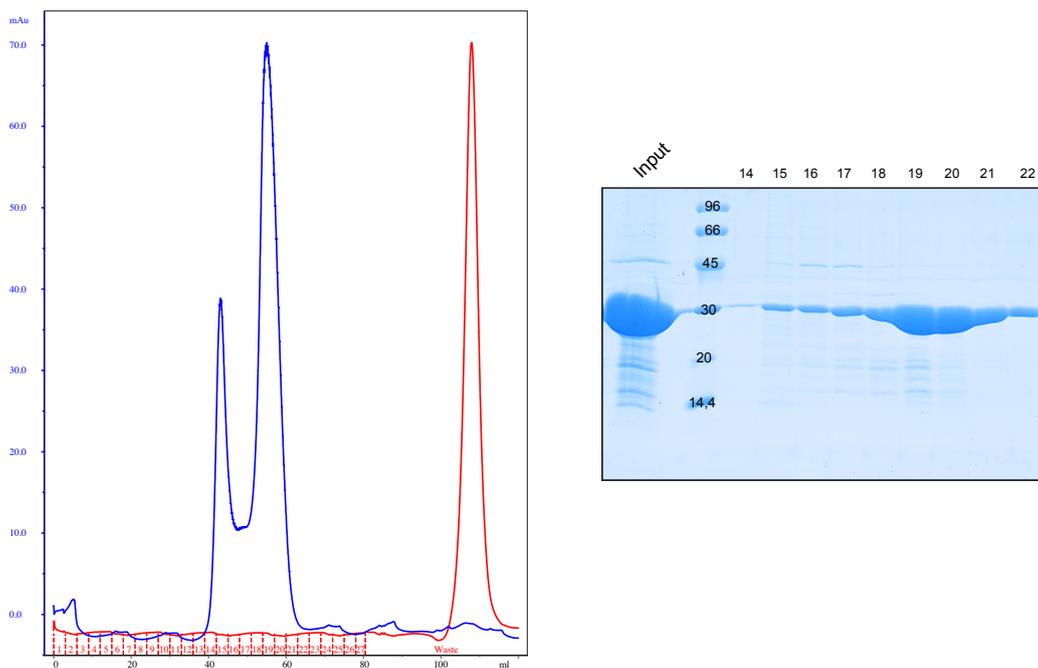
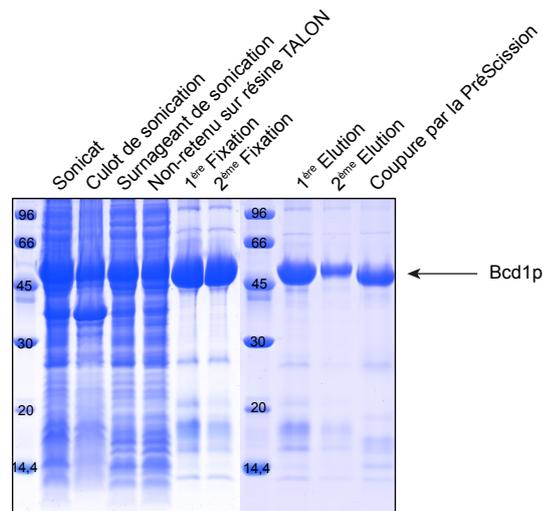


Figure 51. Purification de la protéine Rtt106(65-320). (A) Les différentes étapes de purification sont analysées sur SDS-PAGE 12,5%. (B) Chromatogramme de la chromatographie d'exclusion de taille effectuée avec une colonne Superdex S75 (16:60) sur la protéine Rtt106(65-320) et dépôt des fractions correspondantes sur SDS-PAGE 12,5%. La courbe bleue sur le chromatogramme correspond à l'UV à 280 nm et la courbe rouge à la conductimétrie.

A



B

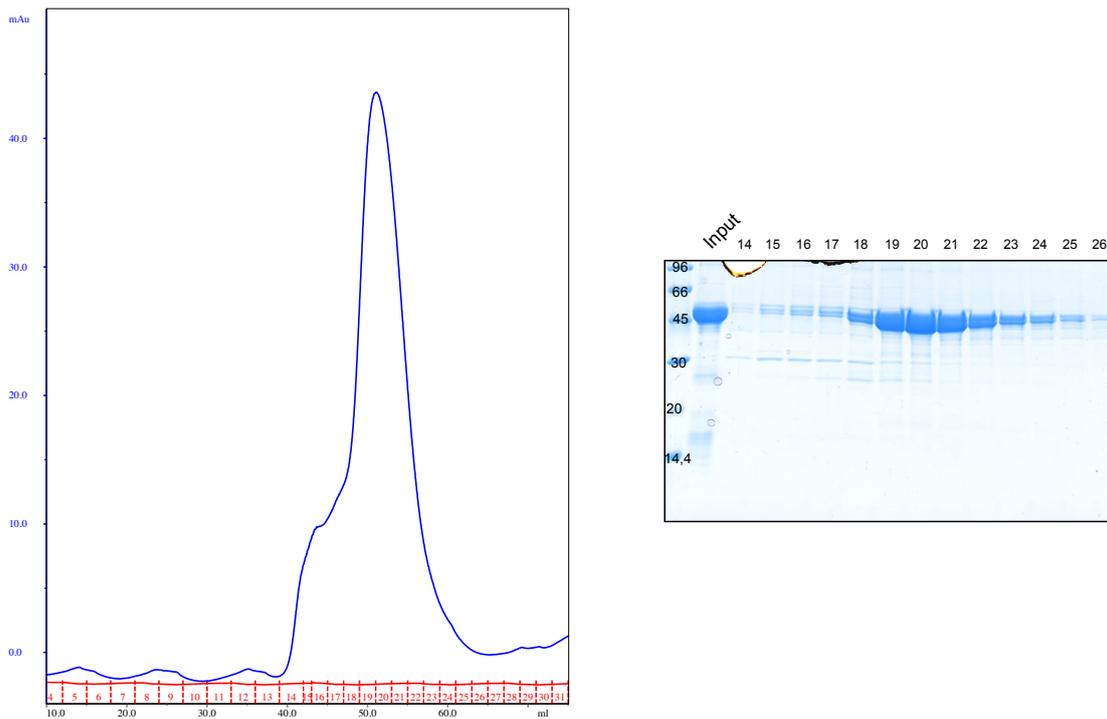


Figure 52. Purification de la protéine Bcd1 entière. (A) Les différentes étapes de purification sont analysées sur SDS-PAGE 12,5%. (B) Chromatogramme correspondant à la chromatographie d'exclusion de taille effectuée avec une colonne Superdex S75 (16:60) sur la protéine Bcd1 et dépôt des fractions correspondantes sur SDS-PAGE 12,5%. La courbe bleue sur le chromatogramme correspond à l'UV à 280 nm et la courbe rouge à la conductimétrie.

deux types d'appareil de mesure (VP-ITC et ITC₂₀₀). Les résultats montrent une affinité de l'ordre du micromolaire (0,94 μ M avec le VP-ITC et 1,17 μ M avec l'ITC₂₀₀) avec un ratio (N) proche de 1:1 (0,932 avec le VP-ITC et 0,985 avec l'ITC₂₀₀) (Figures 53A et B). Les deux appareils donnent des résultats similaires en terme d'affinité et de paramètres thermodynamiques. Les légères différences observées peuvent être dues à des variations de mesure inhérentes à chacun des deux appareils utilisés, à de légères différences entre les échantillons utilisés pour les deux expériences (tampon, pureté, ...) ou au fait que les étiquettes 6xHistidines avaient été conservées pour le test avec le VP-ITC (Figure 53A) ou retirées avec l'autre test (Figure 53B). Ceci nous montre tout de même que la présence de cette étiquette 6xHistidines en N-terminale de chacune des deux protéines ne gêne *a priori* pas l'interaction entre celles-ci.

L'interaction entre ces deux protéines semble faire intervenir une modification de structure secondaire. En effet, les résultats de dichroïsme circulaire (Figure 53C) montrent une importante différence entre les spectres des deux protéines séparées (courbe bleue) et mélangées (courbe rouge). Cette différence est le signe que l'interaction entre les deux partenaires implique une modification au niveau des structures secondaires de l'un ou l'autre des partenaires (ou des deux). Ici, la différence est marquée par une baisse du signal autour de 208 nm, ce qui témoigne généralement de l'existence d'une perte d'hélice α dans l'échantillon. La très forte hausse après mélange aux longueurs d'onde inférieures à 200 nm est difficilement interprétable, car cette zone est très sensible aux petites variations de tampon.

3.2. Détermination de l'interface d'interaction entre les protéines Bcd1 et Rtt106

Les travaux sur l'importance que peut avoir l'interaction entre ces deux protéines d'un point de vue biologique passent entre autre par la possibilité de produire des variants de ces protéines dont les mutations affectent la formation du complexe tout en épargnant leurs structures. L'obtention de tels variants pourrait permettre d'étudier l'effet de la perte de l'interaction sur les fonctions biologiques communes auxquelles participent ces deux protéines. Pour cela, nous avons voulu déterminer quels sont les résidus situés à l'interface entre les deux protéines, et plus particulièrement quels résidus participent activement à la formation ou au maintien du complexe.

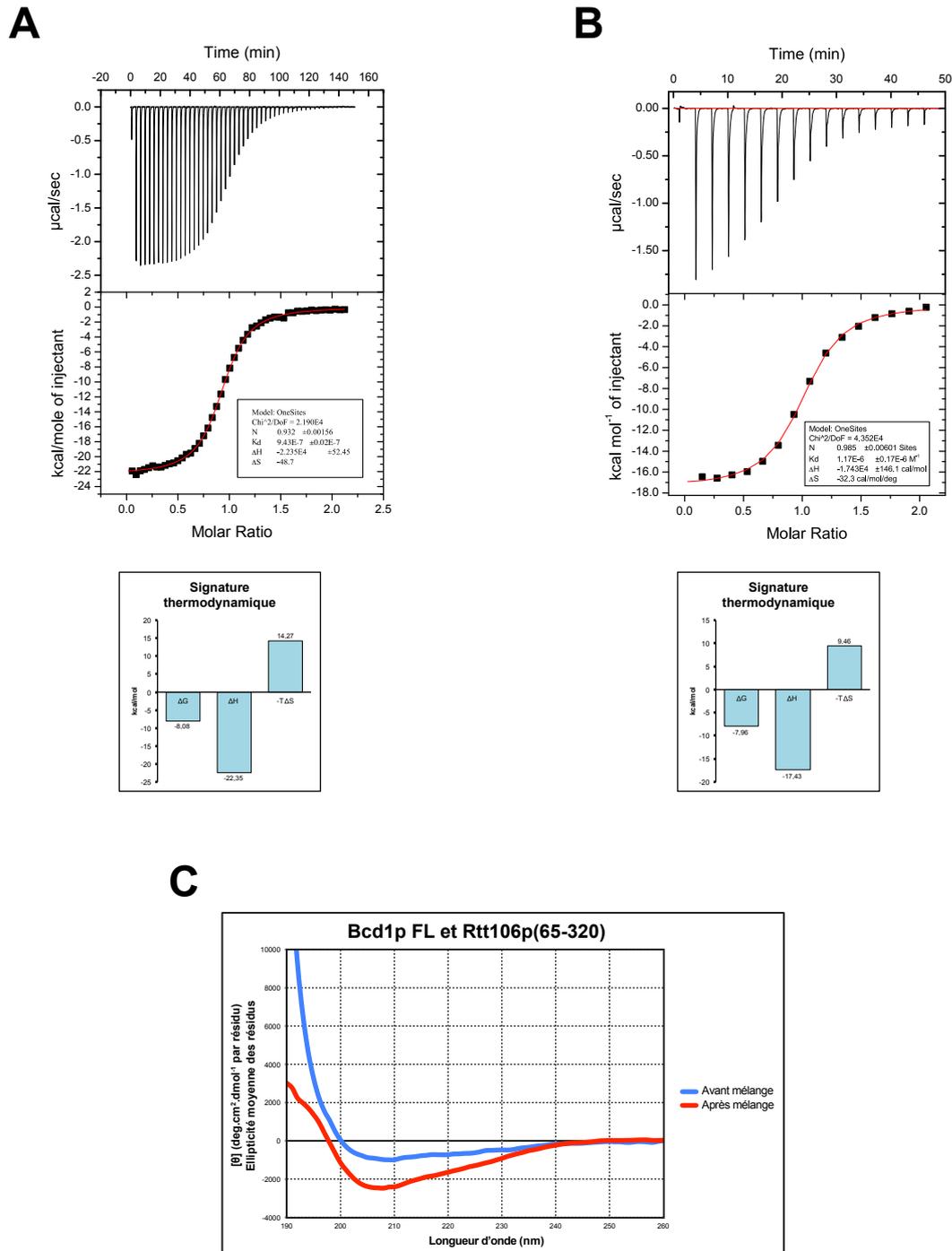


Figure 53. Caractérisation biophysique du complexe formé entre les protéines Bcd1 et Rtt106. Résultats d'expériences d'ITC avec les appareils VP-ITC (A) ou ITC₂₀₀ (B). Ces deux expériences ont été réalisées à 20°C avec les protéines Rtt106(65-320) et Bcd1 entière (FL) avec les étiquettes 6xHis-tidines conservées (A) ou non (B). La protéine Rtt106(65-320) est placée dans la seringue à une concentration de 500 μM (VP-ITC) ou 400 μM (ITC₂₀₀) et la protéine Bcd1 dans la cellule à une concentration de 50 μM (VP-ITC) ou 40 μM (ITC₂₀₀). Les affinités obtenues sont de 0,94 μM (VP-ITC) et 1,17 μM (ITC₂₀₀). Les paramètres thermodynamiques sont représentés sous les résultats ITC par des diagrammes en barres. (C) Résultats de dichroïsme circulaire en double cuve sur les protéines Bcd1 entière (FL) et Rtt106(65-320). Le spectre bleu représente le résultat obtenu avec les deux protéines séparées (compartiments de cuve distincts) et le rouge le résultat après mélange des deux compartiments.

3.2.1. Détermination des domaines minimaux permettant l'interaction

Les protéines ont souvent à leur charge différentes interactions possibles avec plusieurs partenaires, ou même plusieurs fonctions biologiques plus ou moins distinctes. De ce fait, il est fréquent de rencontrer des protéines constituées de plusieurs modules plus ou moins distincts les uns des autres que ce soit fonctionnellement ou structurellement. Cela signifie potentiellement que dans les protéines étudiées, un domaine puisse être isolé du reste de la protéine sans que sa structure et/ou sa fonction ne soit altérée(s). De façon à travailler sur un complexe le plus petit et stable possible, et ainsi favoriser la manipulation des protéines et éliminer les régions non-impliquées dans l'interaction, nous avons entrepris d'identifier la séquence la plus courte possible de chacune des deux protéines permettant à chacune de conserver l'interaction. Pour ce faire, et dans un premier temps, nous avons exploité une approche par protéolyse ménagée du complexe.

3.2.1.1. Approche par protéolyse ménagée et analyse par spectrométrie de masse

Le but de la protéolyse ménagée est de découper la protéine étudiée par des protéases de façon mesurée (ménagée) pour conserver les domaines structuraux. En effet, en faible quantité les protéases vont dans un premier temps cliver les liaisons peptidiques de la protéine cible au niveau des séquences non structurées qui sont plus facilement accessibles. Ainsi, nous pouvons espérer que les régions repliées ne seront pas attaquées et seront séparées les unes des autres.

Dans le cas d'un complexe, l'interface entre les deux protéines est protégée et les deux éventuels domaines structuraux interagissant ensemble n'en forment plus qu'un. De ce fait, lors de la protéolyse ménagée les protéases vont cliver les deux protéines autour des deux domaines complexés. Le but est de trouver les conditions adéquates permettant de supprimer le plus possible de séquences inutiles à l'interaction tout en évitant d'aller trop loin dans l'hydrolyse et de perdre l'interaction.

Les tests de co-expression nous avaient déjà donné l'information que le double domaine PH de la protéine Rtt106 (la région 65-320) permettait de conserver l'interaction avec la protéine Bcd1. Nous avons donc décidé de débiter les premiers tests avec ce domaine plutôt qu'avec la protéine entière. En revanche, n'ayant aucune indication concernant la protéine Bcd1, nous avons employé en première approche la protéine entière.

Le complexe entre les protéines Bcd1 et Rtt106(65-320) a donc été produit et purifié par co-expression des deux partenaires. Sur ce complexe, l'activité de différentes protéases a été testée (Subtilisine, Thermolysine, Carboxypeptidase Y, Elastase, Glu-C, Chymotrypsine, Trypsine et Thrombine). Différents tests de concentration en enzyme et de cinétique de réaction sur le complexe où l'étiquette 6xHistidines est positionnée sur l'une ou l'autre des deux protéines ont été réalisés (données non présentées). Une condition permettant d'obtenir une dégradation conséquente des protéines de départ a été déterminée. Il s'agit d'une coupure à la trypsine à une concentration de 0,1 $\mu\text{g}/\mu\text{L}$ durant 5 h sur le complexe His-Rtt106p(65-320):Bcd1p (Figure 54A). Une fois cette condition identifiée, une protéolyse ménagée du complexe a été réalisée sur un échantillon suffisamment concentré pour permettre une analyse par spectrométrie de masse en conditions natives. Les résultats de spectrométrie de masse ont montré que la protéine Rtt106(65-320) n'était presque pas clivée. Les masses mesurées correspondent à un fragment légèrement plus court que la taille de départ (65-304) (Figure 54B), ce qui est logique puisque ce fragment est déjà limité au domaine structural double PH ayant déjà été cristallisé (Liu *et al.*, 2010). En ce qui concerne la protéine Bcd1, plusieurs fragments se dégagent (Figure 54A). Mais globalement, la région conservant l'interaction avec Rtt106p apparaît être la région C-terminale, et plus précisément la région 113-315. Cette même expérience a également été réalisée par G. Terral du LSMBO qui a analysé le complexe par spectrométrie de masse native avant et après digestion tryptique (Figure 55A, spectres du haut et du bas respectivement). Ces analyses montrent bien une baisse de la masse du complexe entre les deux spectres, représentés par des pics bleus (avant digestion) et gris (après digestion). Le complexe obtenu après digestion a également été analysé par spectrométrie de masse dénaturante, afin de déterminer les bornes précises des fragments obtenus. Ces fragments, représentés schématiquement sur la Figure 55B, montrent également que la protéine Rtt106 (en vert) n'est pas beaucoup touchée par la digestion puisque seuls quelques résidus manquent du côté C-terminal. Ces résultats montrent aussi que la région de Bcd1 suffisante à l'interaction est située du côté C-terminal, avec des fragments débutant aux bornes 111, 112 et 113. Toutefois, une autre borne apparaît par rapport aux précédents résultats. La borne 150 semble effectivement pouvoir conserver l'interaction avec la protéine Rtt106 (Figure 55B).

Suite à ces résultats, les fragments Bcd1p(113-315) et Bcd1p(149-315) ont été clonés afin de poursuivre la recherche de la région la plus courte possible conservant l'interaction. Ces deux fragments comprennent en effet la borne C-terminale la plus courte identifiée

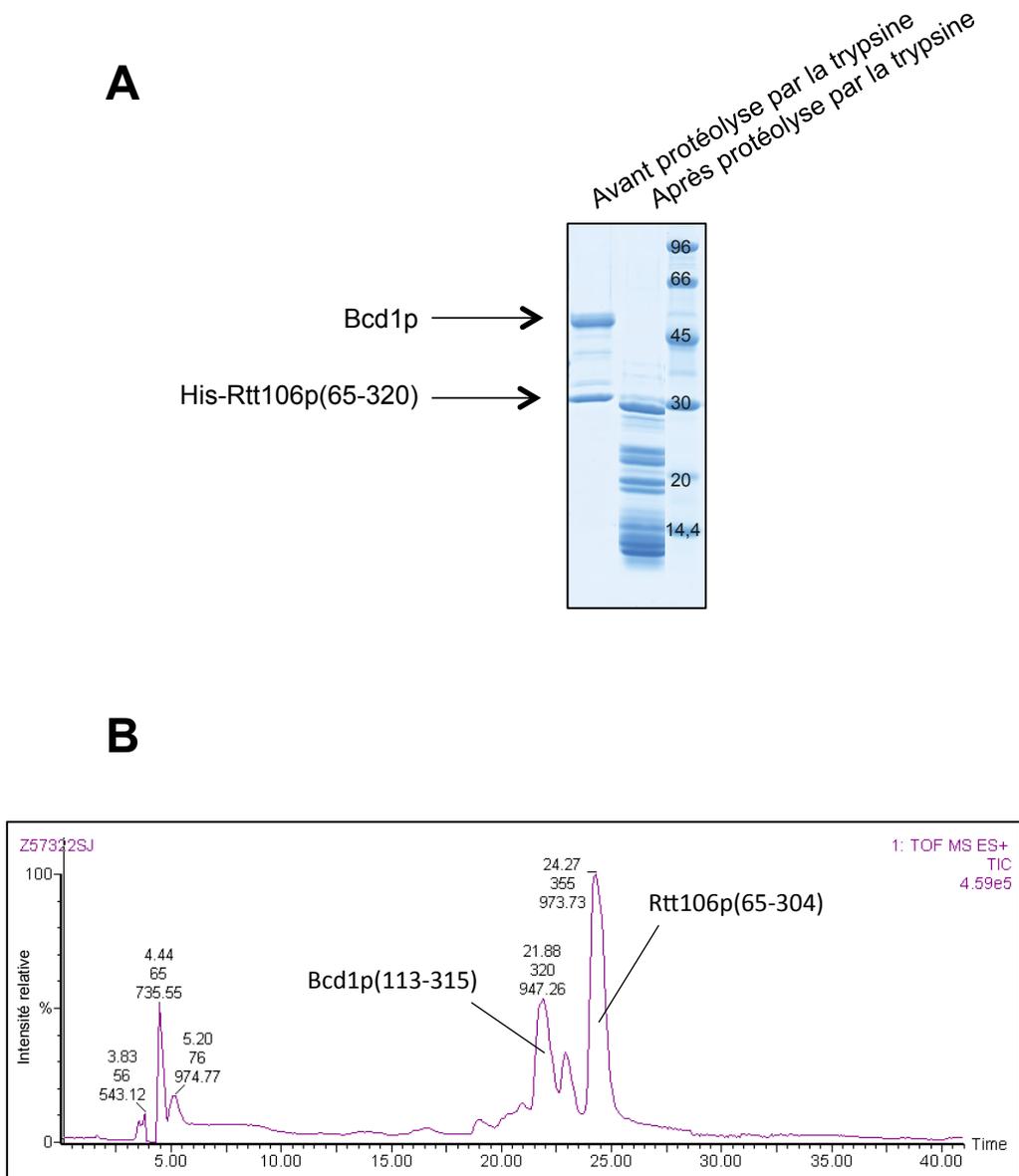
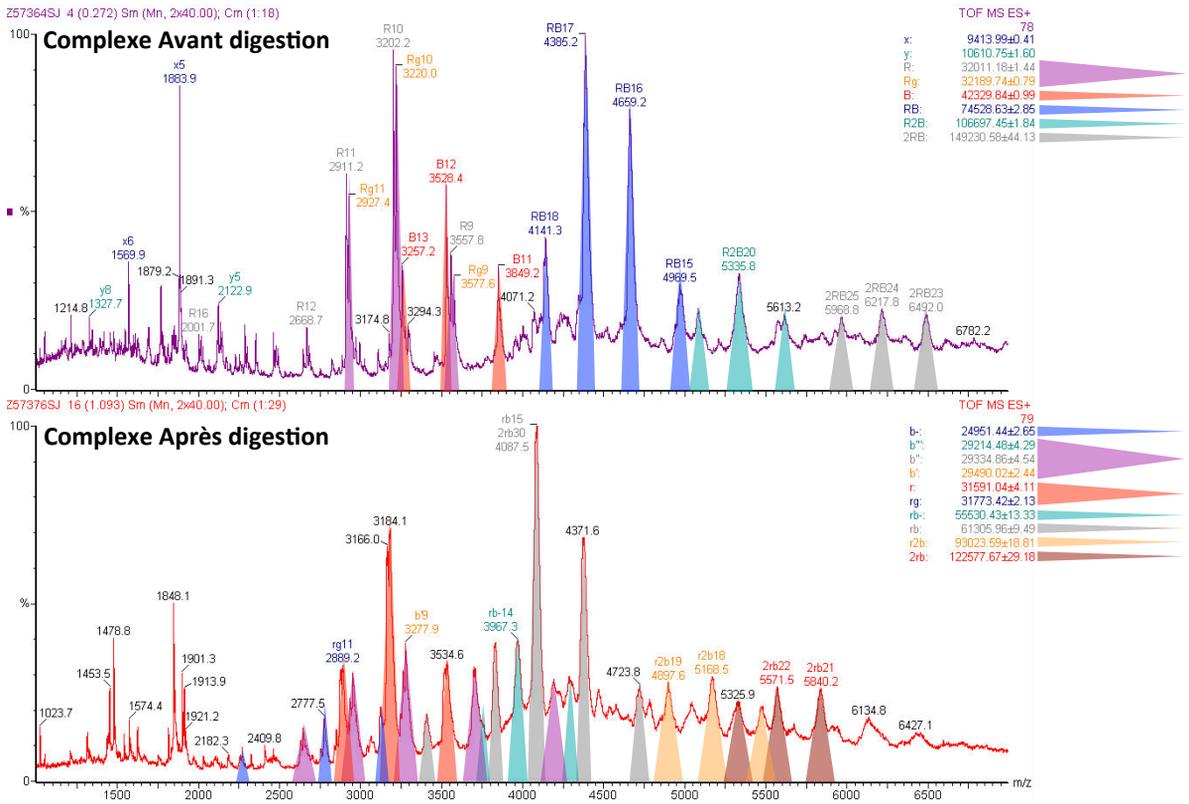


Figure 54. Détermination des régions constituant l'interface entre les protéines Bcd1 et Rtt106. (A) Protéolyse ménagée par la trypsine du complexe His-Rtt106p(65-320):Bcd1p. La réaction a été réalisée pendant 5 h à température ambiante en présence de trypsine (0,1 $\mu\text{g}/\mu\text{L}$) et arrêtée par ajout de PMSF. (B) Spectre de masse sur l'échantillon de complexe soumis à la protéolyse à la trypsine. Les deux pics majoritaires correspondent à des fragments de Rtt106p et Bcd1p.

A



B

Rtt106p(65-320) :

MGSSHHHHHSSSGTGSGLVFLFQGHMDEISETNTIFKLEGVSVLSPLRKKLDLVFVLSNVDGSPVITLLKGNDRLSIYQLNKNIKMASFLPVPEKPNLIYLFMYTTSCEDNKFSPEPVVMTLNKEN

TLNQFKNLGLLDSNVTDFEKCVEYIRKQAILTGFKISNPFVNSLVDLDAEKINSFHLQCHRGTKEGTYLFLPDHIIFGFKPKILLFDASDIESITYSSITRLTFNASLVTKDGEKYEFMSIDQTEY

AKIDDDYVVRKQMKDKMSSEELKAKSKSKG

R : 32011,5 (full-Met) et Rg: 32189,5 (full-Met+gluco)

x : 31455,9

Bcd1p :

MAVLGCVGCIKEFKYKPCRLVQTCSLKSKKHKTRDNCSSGQTHDPKEYISSEALKQADDDKHERNAVYQRDYNLTLQKRMVHVQKMDARMKKNRVLGPGVGHNSNFKRRYDIDEDDRDSTECQR

B : 42333.7 (full = 1-366)

b : 29643.1 / 113-366

bⁿ : 29486.9 / 112-366

b^m : 29330.7 / 111-366

bⁿⁿ : 29210.7 / 111-362

IIRRGVNCMLPLFGMKQRSSQNRSKWDKTMDFLVFVSVVWILCPMQEKGEKELFKHVSHRIKETDFLVQGMGNVVFQKCFEYRLAGTSSCIEGEDGSETKEERTQILQKSLGKLFYTKTFPYNTHIM

b : 24948,9 / 150-366

DSKKLVELAIHEKCI GELLKNTTVIEFPTIFVAMTEADLPEGYEVLHQEPRPLEHTSTLNKFI DNAREEEDAEDSQPTEEPVQKETQDASDS DSDSDDDYNPGLSMDFLTA

Figure 55. Spectrométrie de masse sur le complexe His-Rtt106p(65-320):Bcd1p après protéolyse ménagée à la trypsine. (A) Spectres natifs obtenus avant et après la digestion à la trypsine. Les masses mesurées pour chaque pic sont indiquées à droite des spectres, et les abréviations sont reprises dans les séquences. R2B : 2His-Rtt106p(65-320) + Bcd1p. 2RB : 2(His-Rtt106p(65-320) + Bcd1p). R : His-Rtt106p(65-320). Rg His-Rtt106p(65-320) + gluconoylation. B : Bcd1p. RB : His-Rtt106p(65-320) + Bcd1p. (B) Analyse des fragments de protéolyse ménagée par spectrométrie de masse dénaturante. Les traits verts et bleus représentent les fragments définis par les masses.

(Figure 54B) ainsi que les deux bornes N-terminales 113 et 150 (Figure 55B). De plus, d'autres bornes N- ou C-terminales, déterminées par l'analyse de la séquence et des prédictions de structures secondaires (Figure 37), ont également fait l'objet d'analyses. Ces bornes sont les résidus 120 pour le côté N-terminal et 331, 303 et 292 pour le côté C-terminal.

3.2.1.2. Etude par ITC et RMN de l'interaction des fragments de Bcd1p et Rtt106p

3.2.1.2.1. Etude des fragments de Bcd1p

Ces différentes bornes ayant été déterminées, plusieurs constructions génétiques ont été réalisées permettant l'expression de différentes versions de fragments de Bcd1p. Pour commencer, nous avons raccourci la protéine du côté N-terminal, avec les constructions 120-366 et 149-366. Ces deux fragments ont été produits et purifiés et ne posent aucun problème de solubilité ou de stabilité. Pour savoir si ces fragments raccourcis en N-terminal permettent toujours de conserver l'interaction avec la protéine Rtt106, une analyse par ITC a été réalisée. Les mesures par ITC de l'interaction de la protéine Rtt106(65-320) avec ces deux nouveaux fragments montrent clairement que le fragment 120-366 est toujours suffisant pour permettre l'interaction (Figure 56B), tandis que le fragment 149-366, lui, ne le permet plus (Figure 56C). Comme le fragment 120-366 conserve l'interaction avec la même affinité que la protéine entière (Kd de 1,7 μ M à 20°C avec 150 mM de NaCl), il est très probable que toute l'interface d'interaction du côté de la protéine Bcd1 soit comprise dans ce fragment 120-366. Par la suite, des fragments de la protéine Bcd1 plus courts du côté C-terminal ont été produits et analysés de la même façon. Tout comme le fragment 120-366, le fragment 120-331 conserve lui aussi l'interaction (Figure 56D) avec même une affinité accrue (Kd de 260 nM à 20°C sans NaCl). Cette augmentation de l'affinité est très probablement due au fait que cette expérience ait été réalisée sans NaCl contrairement aux autres. Le fragment 120-303 donnent lui aussi un résultat similaire par ITC (Figure 56E), avec une affinité comparable à la protéine entière (Kd de 1,16 μ M à 20°C avec 150 mM de NaCl) ou au résultat obtenu avec le fragment 120-331 sans NaCl (Kd de 127 nM à 20°C sans NaCl) (Figure 56F). Le dernier fragment que nous souhaitons tester, Bcd1p(120-292) était quant à lui non-exprimé, ce qui suggère que la structure du domaine d'interaction est affectée par l'élimination des 11 résidus du côté C-terminal. Le fragment le plus court de Bcd1p permettant de conserver l'interaction avec la protéine Rtt106 est donc le fragment 120-303.

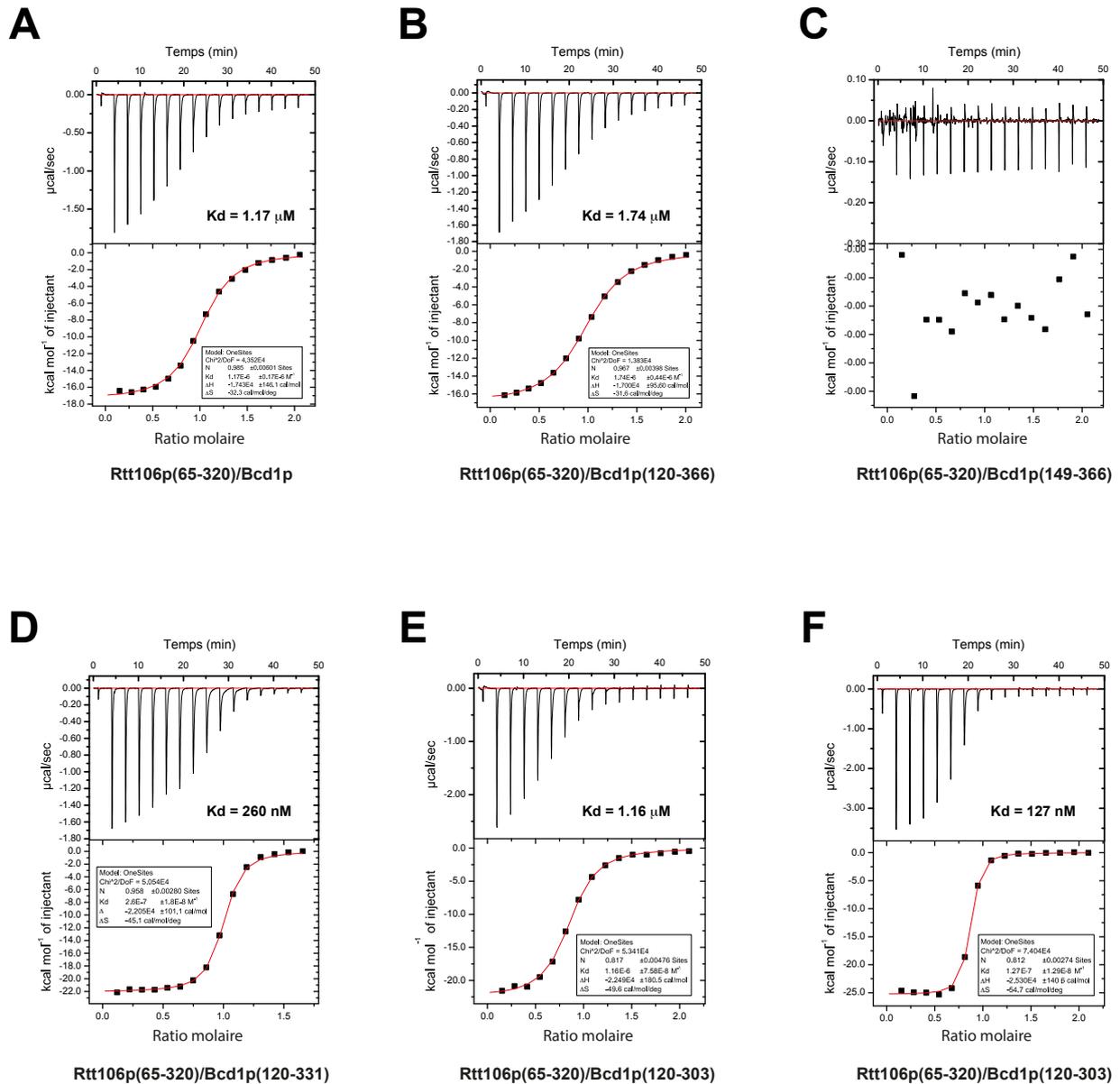


Figure 56. Recherche des constructions de la protéine Bcd1 conservant l'interaction avec la protéine Rtt106. Résultats d'expériences d'ITC effectuées avec l'ITC₂₀₀. Ces expériences ont été réalisées à 20°C avec dans chaque cas la protéine Rtt106(65-320) dans la seringue à une concentration de 400 μM et la construction de la protéine Bcd1 dans la cellule à une concentration de 40 μM . Les fragments utilisés sont indiqués pour chaque expérience sous le résultat. Les affinités (K_d) pour chaque expérience sont indiquées sur le thermogramme (le cas échéant).

Afin de valider ces résultats d'ITC et de vérifier l'impact de ces délétions successives sur le repliement du domaine d'interaction de Bcd1p, une étude par RMN a été entreprise. Pour cette étude, les différents fragments de Bcd1p débutant aux résidus 120 ou 149 ont été produits avec un marquage uniforme à l'azote 15. Des spectres HSQC ^1H - ^{15}N de ces fragments ont été réalisés et montrent que tous les fragments qui débutent au résidu 120 (Figures 57A, C et E) présentent des spectres superposables (Figures 57G et H). Les résidus des régions successivement supprimés à l'extrémité C-terminale prennent l'apparence de pics très intenses situés aux centres des spectres. Ceci montre que les résidus allant de 303 à 366 ne sont pas structurés, et qu'il est préférable de poursuivre les études, notamment structurales, sur le fragment la plus courte (120-303). L'analyse des spectres des fragments 149-331 (Figure 58A) et 149-315 (Figure 58C) montre que ceux-ci sont bien structurés avec une bonne dispersion et homogénéité des pics sur les spectres. Là encore, la perte des 15 résidus C-terminaux entre les fragments se traduit par la perte de pics intenses situés au centre du spectre (Figure 58E), ce qui est le signe que des résidus non structurés ont été supprimés sans toucher au repliement du reste du domaine. De manière intéressante, la superposition des spectres obtenus avec les fragments 120-331 et 149-331 (Figure 59A) ou 120-315 et 149-315 où seul le côté N-terminal est réduit (Figure 59B) montre une très grande différence entre les spectres. Cela signifie que retirer les résidus 120 à 149 modifie fortement l'environnement chimique d'une grande partie de la protéine sans pour autant que le repliement ne soit détruit. Toutefois, il est impossible de dire sur la seule base de ces données si la perte de ces 29 résidus entraîne un changement drastique de la structure de la protéine ou si ces 29 résidus sont simplement en interaction avec le restant de la structure. Cependant, même si ces résidus n'avaient pas d'incidence sur le repliement du domaine, les résultats d'ITC nous indiquent qu'ils sont importants pour la formation d'un complexe avec la protéine Rtt106.

Pour nous assurer de la validité des résultats obtenus par ITC, nous avons décidé de vérifier également par RMN que les fragments débutant au résidu 149 ne sont effectivement pas capables d'interagir avec la protéine Rtt106. Pour cela, la protéine Rtt106(65-320) non marquée à l'azote 15 a été ajoutée à chaque fragment de Bcd1p marquée à l'azote 15. La taille d'un complexe entre les protéines Bcd1 et Rtt106 devenant trop grosse pour être visualisée par RMN, la perte du signal de la protéine Bcd1 en présence de la protéine Rtt106 signifie que ces deux protéines interagissent l'une avec l'autre. Ainsi, il est possible d'observer de cette façon une perte de signal avec les fragments 120-331 (Figure 57B), 120-315 (Figure 57D) et 120-303 (Figure 57F) de Bcd1p, mais les spectres demeurent inchangés avec les fragments 149-331 (Figure 58B) et 149-315 (Figure 58D). Ces résultats obtenus par RMN confirment

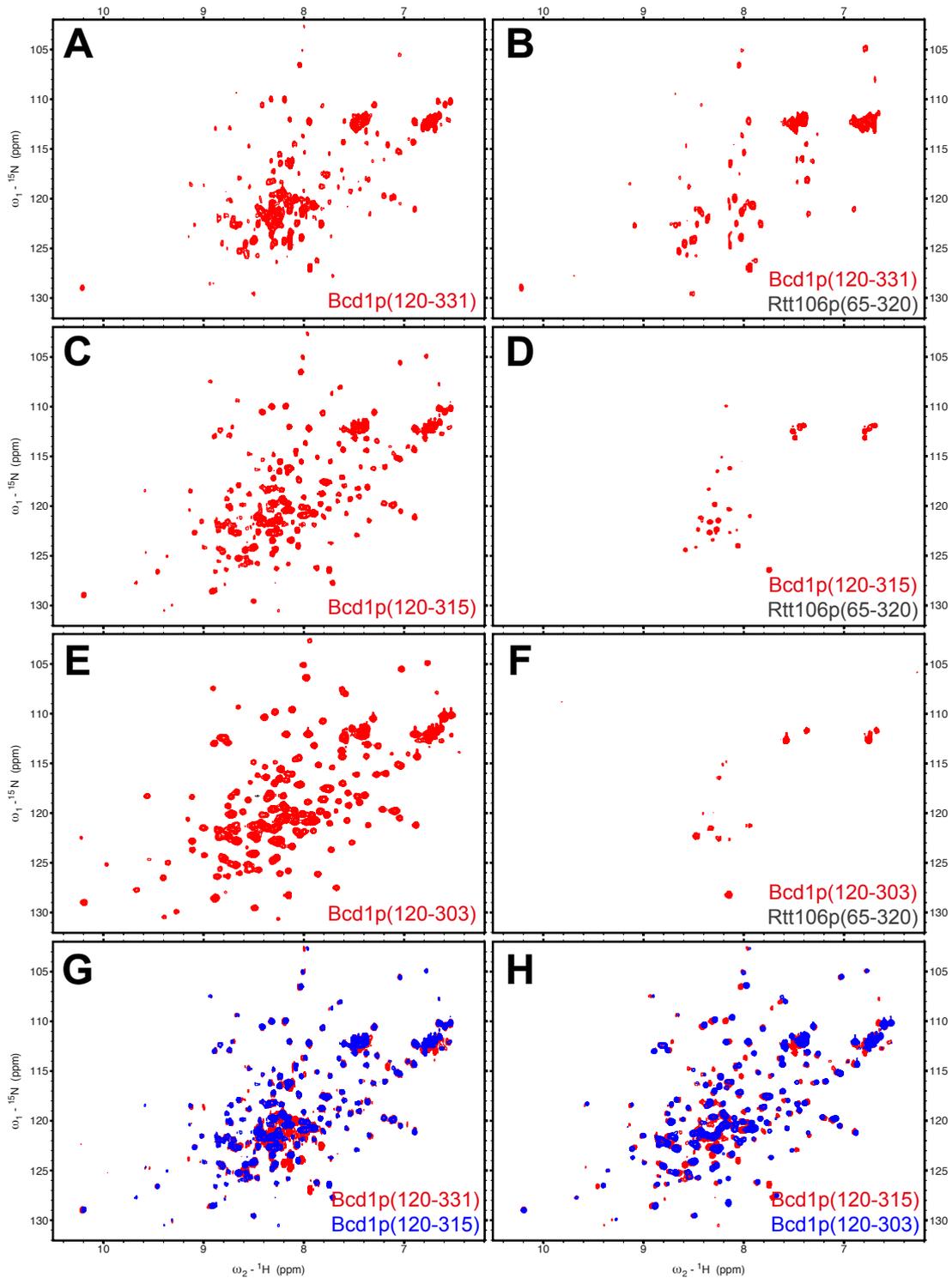


Figure 57. Spectres HSQC ^1H - ^{15}N des fragments de Bcd1p débutant au résidu 120. Spectres enregistrés à 293K sur un spectromètre Bruker 600 MHz équipé d'une crysonde TCI à triple résonance. Les protéines sont utilisées à 200 μM en tampon phosphate 10 mM à pH 6,4 avec 150 mM de NaCl et 1 mM de DTT. Les spectres ont été analysés avec le logiciel Sparky (Goddard et Kneller, 2002).

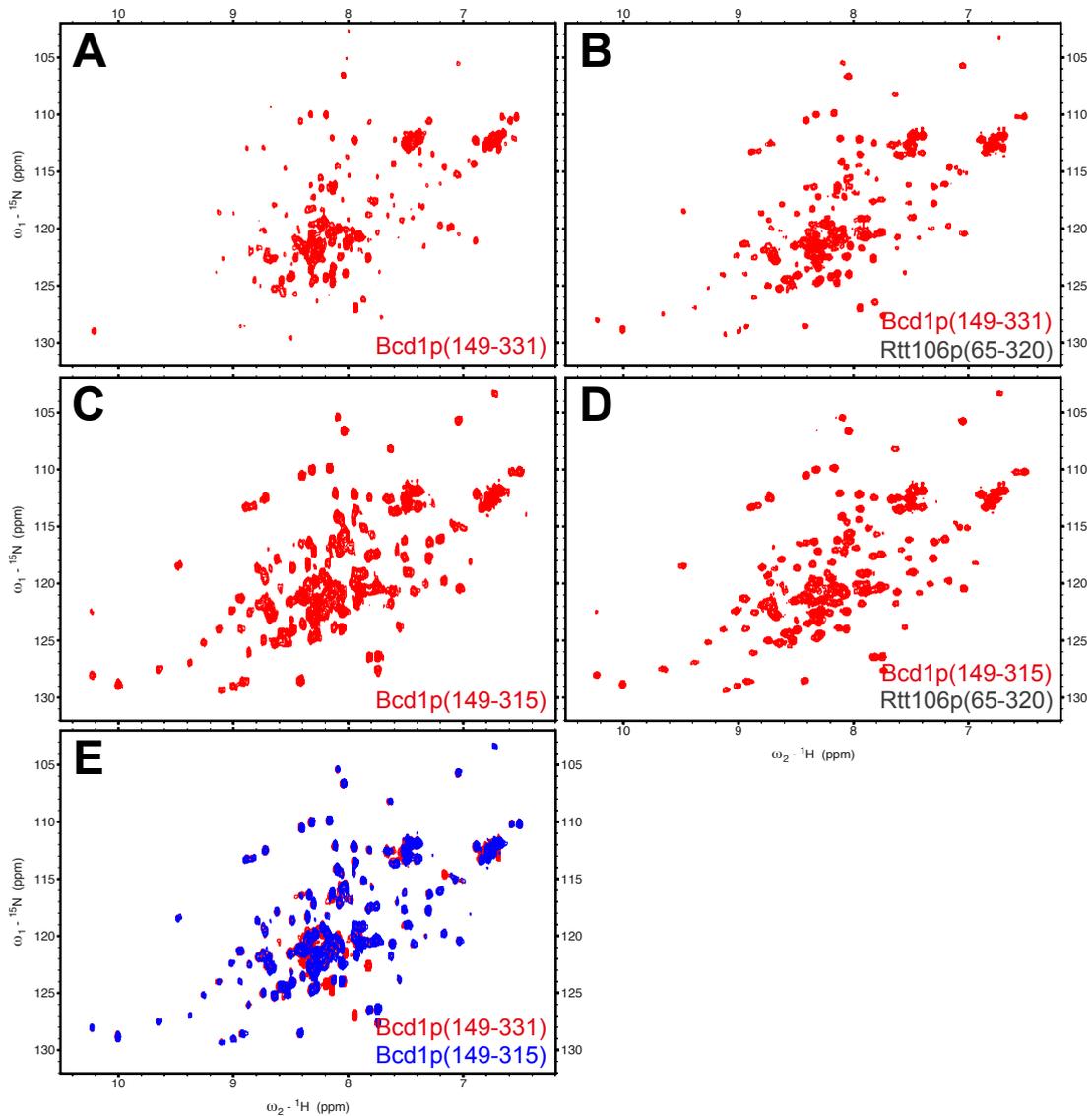


Figure 58. Spectres HSQC ^1H - ^{15}N des fragments de Bcd1p débutant au résidu 149. Spectres réalisés à 293K sur un spectromètre Bruker 600 MHz équipé d'une cryosonde TCI à triple résonance. Les protéines sont utilisées à 200 μM en tampon phosphate 10 mM à pH 6,4 avec 150 mM de NaCl et 1 mM de DTT. Les spectres ont été analysés avec le logiciel Sparky (Goddard et Kneller, 2002).

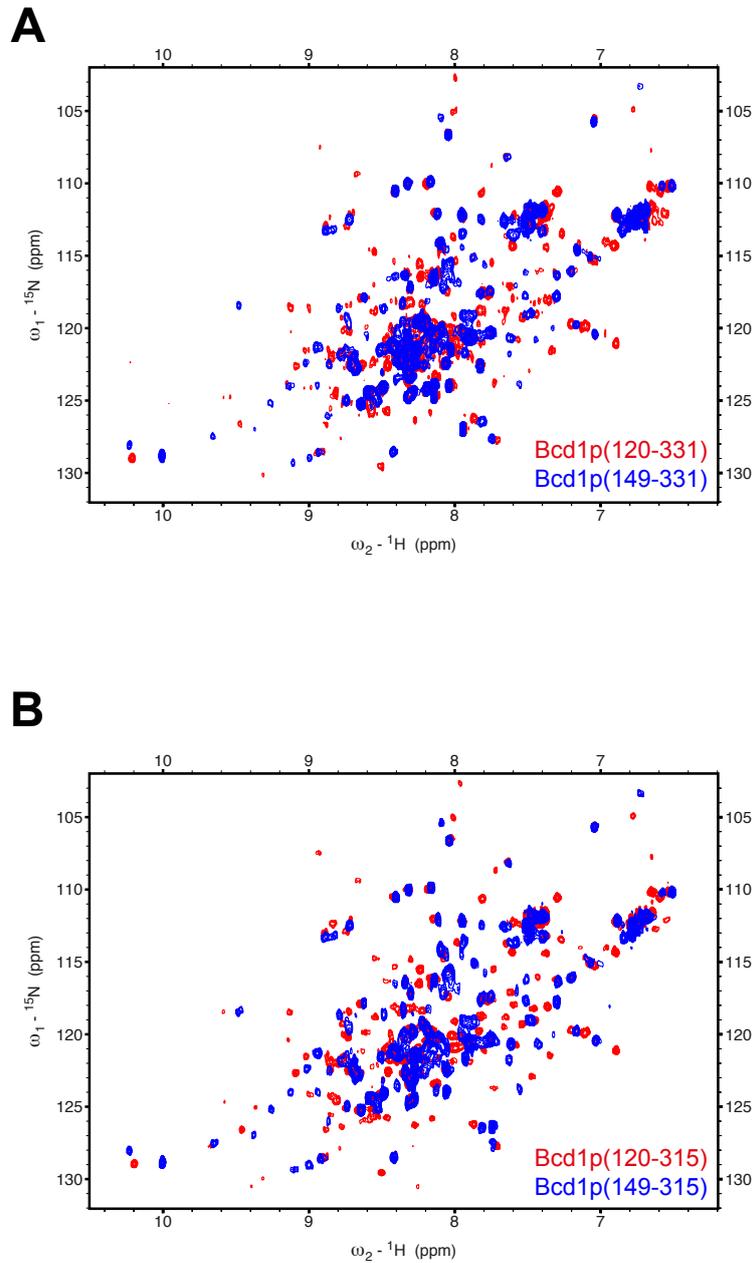


Figure 59. Comparaison des spectres HSQC ^1H - ^{15}N des fragments de Bcd1p débutant au résidu 120 ou au résidu 149. Spectres réalisés à 293K sur un spectromètre Bruker 600 MHz équipé d'une crysonde TCI à triple résonance. Les protéines sont utilisées à 200 μM en tampon phosphate 10 mM à pH 6,4 avec 150 mM de NaCl et 1 mM de DTT. Les spectres ont été analysés avec le logiciel Sparky (Goddard et Kneller, 2002).

donc ceux obtenus au préalable par ITC. Les résultats obtenus par RMN et ITC sur les fragments débutant au résidu 149 sont toutefois contraires à ceux obtenus par l'analyse par spectrométrie de masse des produits de la protéolyse ménagée du complexe Rtt106p(65-320):Bcd1p (Figure 55) où le fragment Bcd1p(150-366) semble toujours permettre l'interaction. La différence entre les deux résultats vient du fait que dans le cas de la protéolyse ménagée le complexe est déjà formé avant d'être digéré (jusqu'au résidu 150 de Bcd1p) et dans les cas de l'ITC et de la RMN les fragments de Bcd1p n'ont déjà plus la région 120-149 avant d'être mis en présence de Rtt106p. Cela pourrait signifier que la région 120-149 soit nécessaire à l'interaction mais qu'une fois l'interaction établie, elle devient dispensable.

Il est à noter que le fragment 130-303 a également été testé. Celui-ci donne un spectre RMN similaire à celui obtenu avec le fragment 120-303 (Figures 60A et C) et interagit avec la protéine Rtt106 (Figure 60B). Cependant, la qualité du spectre HSQC ^1H - ^{15}N de ce fragment est moins bonne que celle du fragment 120-303 (Figure 60C). En effet, la différence en nombre de pics entre les deux HSQC est supérieure à 10 (différence attendue), ce qui signifie que certaines tâches de corrélations ne sont plus visibles. Pour cette raison, nous avons choisi de travailler par la suite sur le fragment 120-303.

3.2.1.2.2. Etude des fragments de Rtt106

De la même façon que pour la protéine Bcd1, plusieurs fragments de la protéine Rtt106 ont été produits et testés. Tout d'abord, nous avons réalisé une construction génétique exprimant une version légèrement plus courte des côtés N et C terminaux, le fragment Rtt106p(68-304). Le choix de ce fragment fut guidé par nos résultats de spectrométrie de masse pour définir la borne 304 du côté C-terminal, et des données de la littérature pour la borne 68 du côté N-terminal (Su *et al.*, 2012). Ce fragment conserve toujours l'interaction avec la protéine Bcd1 entière (Figure 61B) ou avec le fragment 120-366 (Figure 61C), mais l'affinité se trouve très fortement réduite avec respectivement des Kd de 33,7 μM et 48,5 μM au lieu de 1,17 μM et 1,74 μM à 20°C en présence de 150 mM de NaCl. Compte tenu de cette perte d'affinité, ce fragment ne sera pas retenu. Un autre fragment a par la suite été choisi d'après la littérature (Su *et al.*, 2012). Ce fragment (65-301) est plus court du côté C-terminal uniquement. Contrairement au fragment précédent, celui-ci en plus de conserver l'interaction avec la protéine Bcd1 (Figure 61D) conserve également la même affinité avec un Kd de

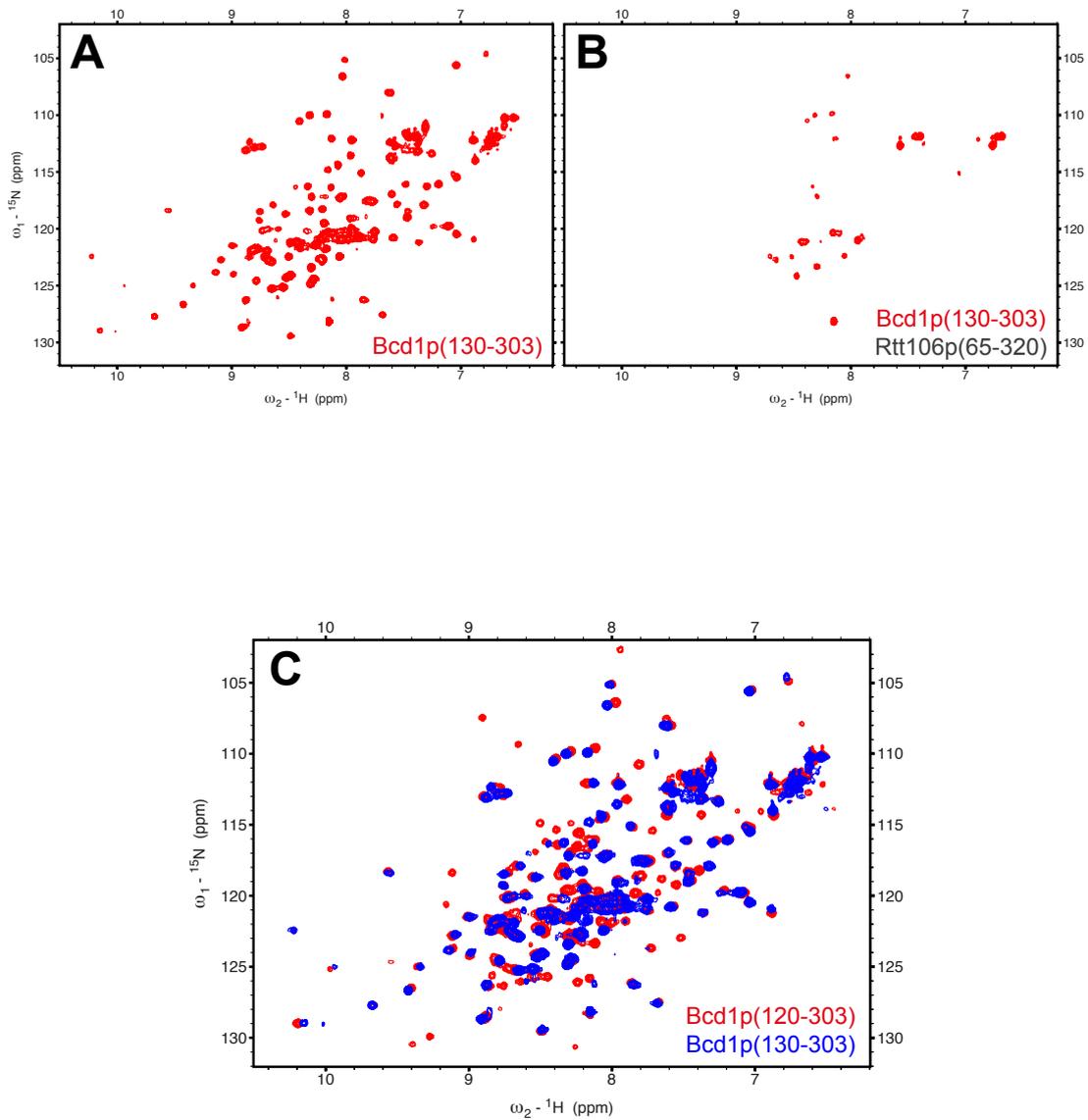


Figure 60. Analyse par spectre HSQC ^1H - ^{15}N du fragment 130-303 de Bcd1p seul (A) ou en présence de Rtt106p(65-320) (B) et comparaison entre les fragments 130-303 et 120-303 (C). Spectres réalisés à 293K sur un spectromètre Bruker 600 MHz équipé d'une cryosonde TCI à triple résonance. Les protéines sont utilisées à 200 μM en tampon phosphate 10 mM à pH 6,4 avec 150 mM de NaCl et 1 mM de DTT. Les spectres ont été analysés avec le logiciel Sparky (Goddard et Kneller, 2002).

0,94 μM à 20°C en présence de 150 mM de NaCl. Ce fragment (65-301) plus court fera plus tard l'objet d'études structurales.

Au niveau de la structure 3D de la protéine Rtt106, les deux domaines PH bien structurés sont reliés entre eux par un linker non visible sur la structure cristallographique car trop flexible. Nous nous sommes demandés si ce linker, apparemment plus long que nécessaire pour simplement relier les deux domaines PH, ne pourrait pas être responsable de l'interaction avec la protéine Bcd1. Pour cela, les résidus 206 à 216 de Rtt106p ont été remplacés par un motif GSSG. Ce nouveau fragment a également été testé avec la protéine Bcd1, mais la suppression de ces quelques résidus entre les deux domaines PH n'a pas d'impact sur l'interaction (Figure 61E). Même si ce fragment plus court n'a pas d'influence sur l'interaction avec la protéine Bcd1, nous avons choisi de l'écarter pour éviter les biais pouvant être dus à l'altération de la séquence d'origine. De plus, le thermogramme obtenu lors de l'expérience d'ITC est plus bruité que pour les fragments testés et une légère bande de dégradation apparaît sur gel lors de la purification de ce fragment contrairement au fragment 65-301 par exemple.

En conclusion, le fragment stable le plus court que nous avons pu caractériser de la protéine Rtt106 qui permet de conserver l'interaction avec la protéine Bcd1 sans en affecter l'affinité ou la qualité est donc le fragment 65-301.

3.2.2. Caractérisation des régions à l'interface de l'interaction Bcd1p-Rtt106p

De façon à identifier le plus précisément possible les régions ou résidus impliqués dans l'interaction entre les deux protéines Bcd1 et Rtt106, plusieurs techniques ont été mises en œuvre. Ces techniques sont basées sur la spectrométrie de masse et ont alors été réalisées par le LSMBO de Strasbourg et sur la RMN au laboratoire.

3.2.2.1. Etude par spectrométrie de masse

3.2.2.1.1. Echange hydrogène-deutérium (HDX)

Les expériences d'échange hydrogène-deutérium permettent au sein d'une protéine d'intérêt d'identifier par spectrométrie de masse les peptides dont la cinétique d'échange entre l'hydrogène et le deutérium a été modifiée. Ces modifications peuvent être de deux types, une baisse ou une augmentation de la vitesse d'incorporation du deutérium entre les espèces libres et complexées. Une augmentation de cette vitesse se traduit par une différence de masse

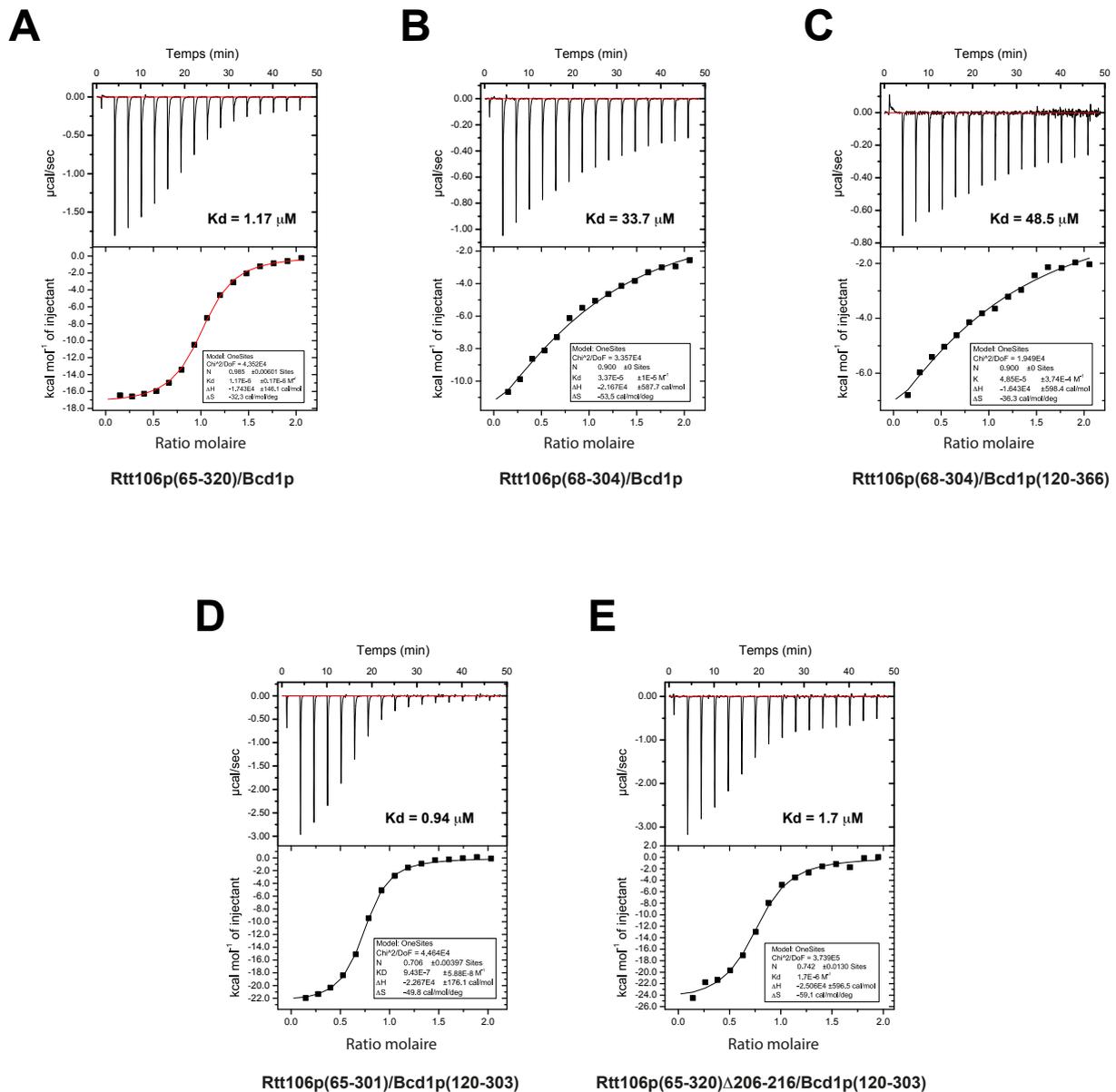


Figure 61. Recherche des fragments de la protéine Rtt106 conservant l'interaction avec la protéine Bcd1. Résultats d'expériences d'ITC réalisées avec l'ITC₂₀₀ à 20°C avec dans chaque cas un fragment de la protéine Rtt106 dans la seringue à une concentration de 400 μM et la protéine Bcd1 dans la cellule à une concentration de 40 μM . Les fragments utilisés sont indiqués pour chaque expérience sous le résultat. Les affinités (Kd) pour chaque expérience sont indiquées sur le thermogramme. Le fragment Rtt106p(65-320)Δ206-216 correspond à la séquence de la protéine Rtt106(65-320) dont la boucle (206-216) a été remplacée par les résidus GSSG.

mesurée (protéine libre – protéine liée) négative. Dans ce cas de figure, l'interaction entre les deux partenaires génère un changement conformationnel où une région de la protéine se retrouve plus exposée au solvant que dans sa forme libre. *A contrario*, une diminution de la vitesse d'incorporation du deutérium, provoquée par le masquage, dans le complexe, d'une région de la protéine auparavant exposée au solvant, se traduit par une différence de masse mesurée positive.

3.2.2.1.1.1. Analyse par HDX de la protéine Rtt106

L'analyse par spectrométrie HDX a été réalisée par Guillaume Terral (LSMBO, Strasbourg) sur le fragment Rtt106p(65-320) libre ou en complexe avec la protéine Bcd1 entière. Les résultats de cette expérience (Figure 62A) montrent que certains peptides présentent une différence de deutération plus grande (supérieure à 1,5 Da) entre les analyses effectuées sur la forme libre ou en complexe, à savoir des peptides couvrant les régions 159-175, 175-182 et 188-195. Ces régions sont reportées en orange sur la structure 3D de la protéine Rtt106 (Figure 62B). L'ensemble de ces peptides correspond majoritairement aux deux hélices successives α_2 et α_3 et à la boucle entre celles-ci. Il est à noter qu'après la digestion trypsique, les peptides identifiés par spectrométrie de masse recouvrent 88% de la séquence totale de Rtt106p(65-320). Certaines régions minoritaires de la protéine ne sont donc pas analysables. C'est le cas de la séquence 183-187 correspondant à la première moitié de l'hélice α_3 qu'aucun peptide ne recouvre. Il est donc possible que cette partie soit également touchée par l'interaction avec la protéine Bcd1.

De façon intéressante, l'une des deux hélices discutées dans le paragraphe précédent – l'hélice α_2 – constitue l'une des principales différences structurales entre les protéines Rtt106 et Pob3 (Figure 63B), puisque deux brins β sont ajoutés de part et d'autre de cette hélice formant un feuillet β antiparallèle dans la structure de Pob3p. Cette dernière est en effet un homologue structural de la protéine Rtt106 chez la levure et qui porte également un double domaine PH. Qui plus est, la protéine Pob3, malgré sa très forte similarité de structure avec la protéine Rtt106 n'est pas capable d'interagir avec la protéine Bcd1 (Figure 63A). Puisque Pob3p n'est pas en mesure d'interagir avec Bcd1p malgré sa forte similarité de structure avec Rtt106p, il est envisageable que la différence structurale au niveau de l'hélice α_2 soit un déterminant essentiel de l'interaction Bcd1p/Rtt106p. Cette hypothèse serait donc en adéquation avec les résultats de spectrométrie de masse ayant montrés que cette même hélice était probablement située proche ou à l'interface entre les deux partenaires. Toutefois, un titrage suivi par RMN a été réalisé entre la protéine Bcd1(120-303) et un peptide de

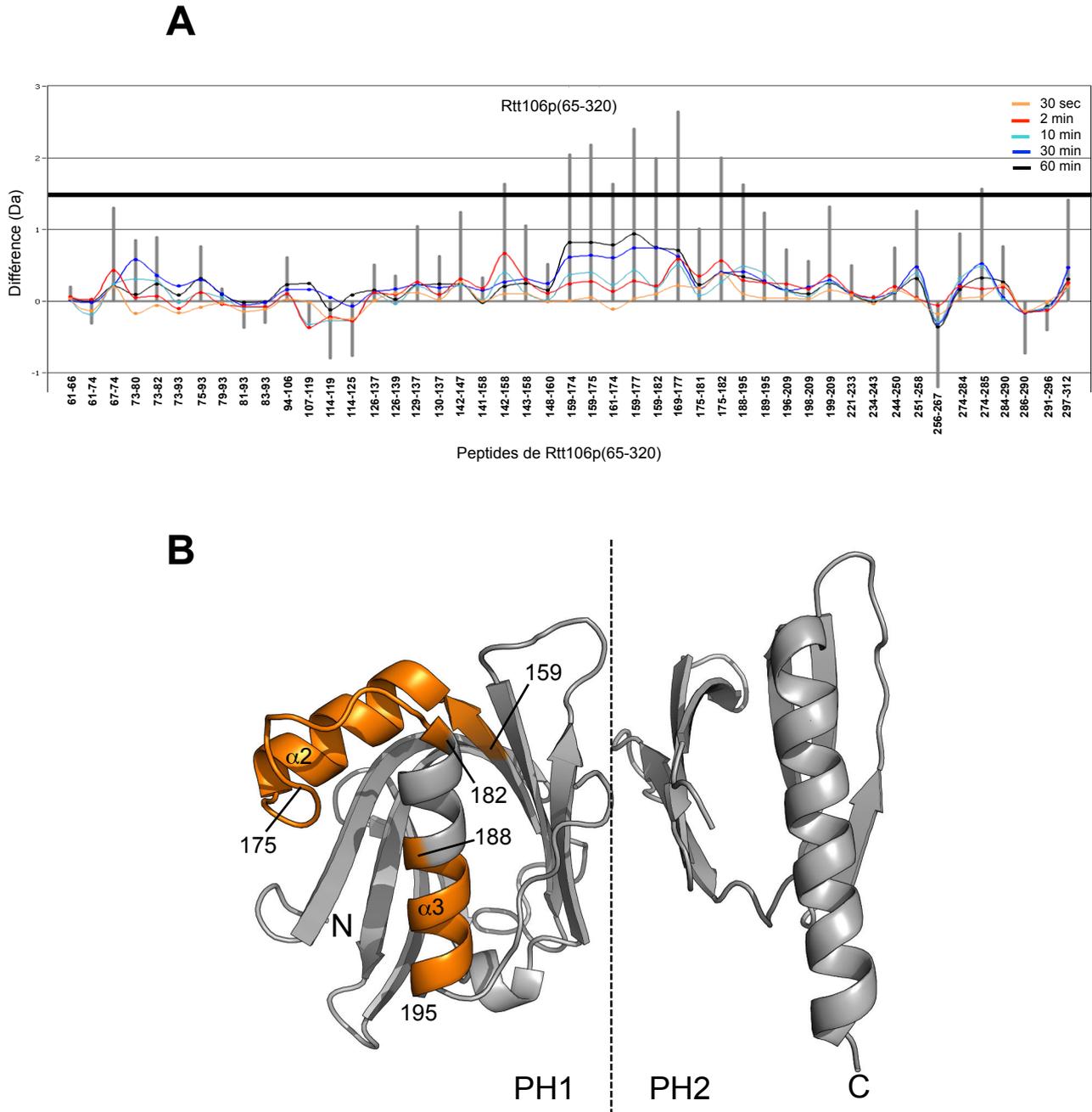


Figure 62. Echange Hydrogène-Deutérium analysé par spectrométrie de masse sur la protéine Rtt106. (A) Résultat de l'expérience d'HDX sur la protéine Rtt106 (réalisée par G. Terral du laboratoire LSMBO). La protéine est d'abord mise en présence de deutérium, puis digérée à la trypsine. Les différents peptides obtenus après digestion sont analysés par spectrométrie de masse. Les différentes courbes reflètent l'incorporation du deutérium après différents temps d'incubation (30 s, 2 min, 10 min, 30 min et 60 min), par la différence de masse des peptides de la protéine Rtt106, entre la protéine Rtt106 seule et la protéine Rtt106 en complexe avec Bcd1p. Le trait noir horizontal représente le seuil de significativité. (B) Report des peptides dont la différence de masse est la plus grande lors de l'expérience d'HDX (représentés en orange) sur la structure 3D de la protéine Rtt106 (code PDB : 3TVV ; Su *et al.*, 2012). Les bornes des peptides sont indiquées en abscisses.

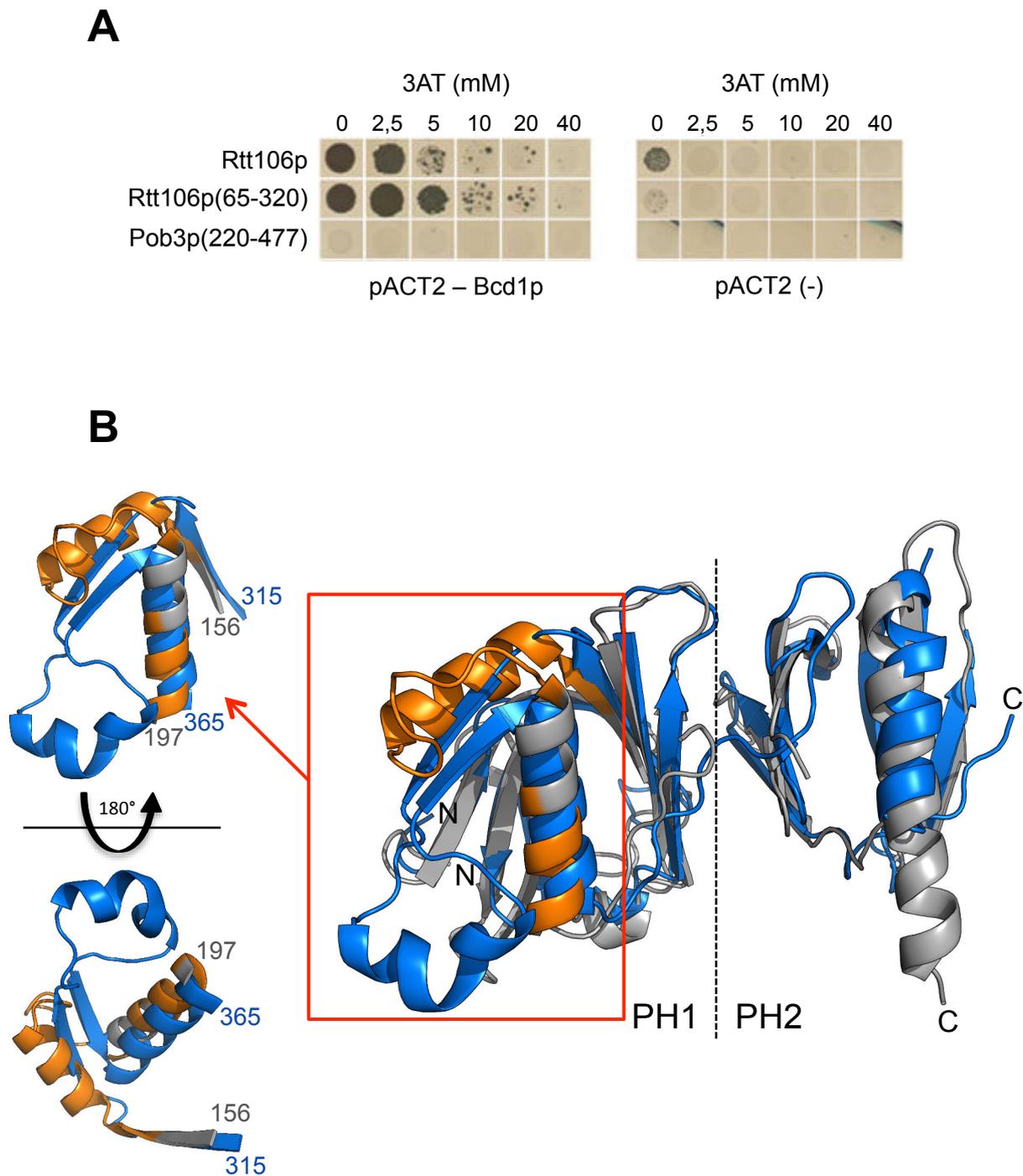


Figure 63. Différences biologiques et structurales entre les protéines Rtt106 et Pob3. (A) Résultat d'une expérience de double hybride réalisée par D. Tiotiu au laboratoire entre la protéine Bcd1 et les protéines Rtt106 ou Pob3. (B) Superposition des structures 3D des protéines Rtt106 (code PDB : 3TVV; Su *et al.*, 2012) en gris et orange (même code couleur que pour la figure précédente) et Pob3 (code PDB : 2GCJ; VanDemark *et al.*, 2006) en bleu. La région encadrée en rouge est représentée isolée à gauche sous deux angles différents.

13 résidus correspondant à l'hélice $\alpha 2$ de Rtt106p (Rtt106M $\alpha 2$), sans qu'aucune interaction ne soit observée (données non présentées). Cette hélice seule ne semble donc pas suffisante à l'interaction avec Bcd1p, ou du moins lorsqu'elle est étudiée hors du contexte de la protéine.

3.2.2.1.1.2. Analyse par HDX de la protéine Bcd1

L'expérience a également été réalisée sur la protéine Bcd1 entière par Guillaume Terral. Avec cette protéine, le recouvrement de séquence par les peptides obtenus après digestion trypsique est de 97%. Il est à noter que la couverture est meilleure que pour Rtt106p. En prenant en compte les peptides subissant le plus de différences entre les analyses effectuées sur les formes libre et complexée, plusieurs régions se distinguent. Tout d'abord, six peptides présentent une différence de vitesse de deutération positive et élevée, ceux-ci correspondent aux régions 139-158, 169-192, 220-226, 242-259, 272-281 et 329-361 (Figure 64). De plus, la région 209-217 présente, elle, une différence de vitesse de deutération fortement négative. Cela peut signifier que cette région est, après interaction avec la protéine Rtt106, plus accessible au solvant. Cette information est cohérente avec les résultats de dichroïsme circulaire obtenus précédemment qui montrent une modification de structure secondaire lors de l'interaction entre les deux partenaires.

Parmi les six régions citées précédemment et semblant plus protégées après la formation du complexe, cinq sont comprises dans le domaine 120-303 déterminé précédemment comme domaine d'interaction minimum. Le dernier peptide (329-361) est lui en dehors de ce domaine, mais appartiendrait à une région C-terminale non repliée. Il s'agit donc probablement d'un artefact de l'expérience dû à la flexibilité de cette région.

En ce qui concerne les cinq autres régions, celles-ci ne se suivent pas dans la séquence comme cela était observé lors de l'étude de la protéine Rtt106. Cela peut signifier que plusieurs régions de la protéine Bcd1 sont en contact avec la protéine Rtt106, ou que ces cinq séquences se retrouvent proches dans le contexte de la protéine repliée et forment une seule région de contact. Enfin, comme cette expérience analyse l'enrichissement en deutérium de peptides, elle n'a pas une résolution à l'échelle de l'acide aminé. Cela signifie que lorsqu'un peptide donne un résultat positif, il est possible que seul un résidu de ce peptide soit touché. Ce ne sont donc certainement pas tous les résidus de chaque peptide qui sont importants mais seulement un ou quelques uns. L'hypothèse d'un seul domaine en contact avec la protéine Rtt106 constitué de plusieurs résidus venant de différentes zones de la séquence semble donc la plus probable. De plus et pour aller dans ce sens, la zone d'interaction définie sur la

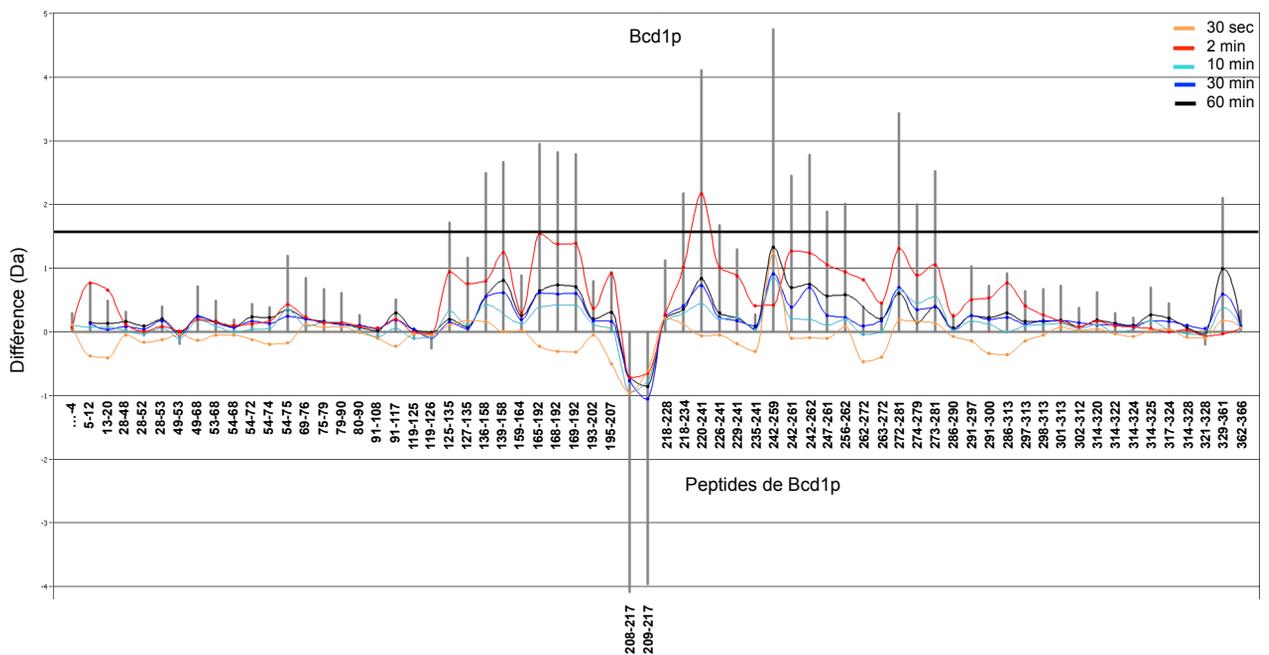


Figure 64. Echange Hydrogène-Deutérium analysé par spectrométrie de masse sur la protéine Bcd1. (A) Résultat de l'expérience d'HDX sur la protéine Bcd1 (réalisée par G. Terral du laboratoire LSMBO). La protéine est d'abord mise en présence de deutérium, puis digérée à la trypsine. Les différents peptides obtenus après digestion sont analysés par spectrométrie de masse. Les différentes courbes reflètent l'incorporation du deutérium après différents temps d'incubation (30 s, 2 min, 10 min, 30 min et 60 min), par la différence de masse des peptides de la protéine Bcd1, entre la protéine Bcd1 seule et la protéine Bcd1 complexée à Rtt106p. Le trait noir horizontal représente le seuil de significativité.

protéine Rtt106 est assez restreinte. Pour finir, nous ne pouvons pas exclure un changement conformationnel de la protéine Bcd1 lors de son interaction avec Rtt106p comme le suggèrent les résultats de dichroïsme circulaire.

3.2.2.1.2. Mobilité ionique et spectrométrie de masse

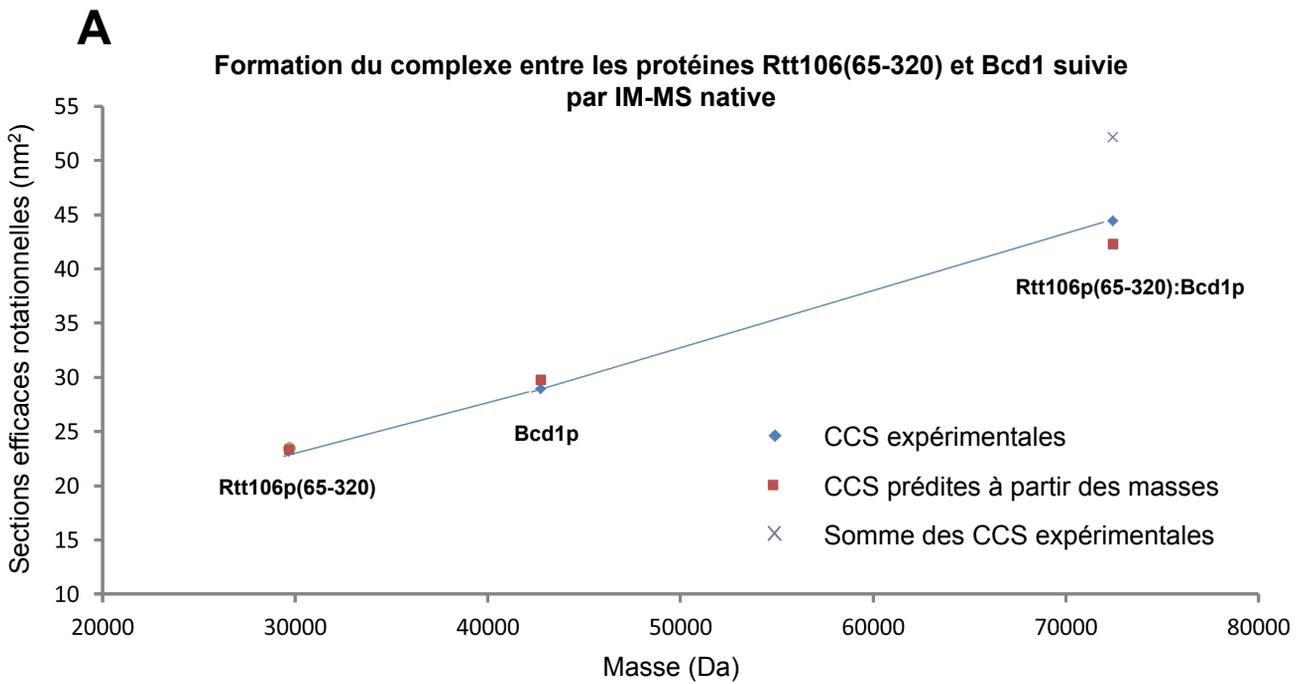
La mobilité ionique permet de séparer les objets en solution en fonction de leur taille (masse, conformation, compaction) et de leur charge. Ici cette technique a été couplée à la spectrométrie de masse pour la détection des différentes formes (Travaux réalisés par Guillaume Terral au LSMBO de Strasbourg). La mobilité ionique se traduit par l'observation des sections efficaces rotationnelles ou CCS pour *Collision Cross Sections*. Ces sections efficaces rotationnelles se traduisent par l'empreinte moyenne en deux dimensions que peut donner un objet en solution.

Il est possible de déterminer de façon théorique quelle sera la CCS d'une protéine à partir de sa masse : 29,8 nm² pour Bcd1p et 23,4 nm² pour Rtt106p(65-320) (2^{ème} colonne du tableau, Figure 65), ou à partir d'une structure 3D connue : 23,5 nm² (colonne de droite). Dans le cas des protéines Bcd1 entière ou Rtt106(65-320) libres, les CCS théoriques sont très proches des CCS mesurées : 29 nm² pour Bcd1p et 23,2 nm² pour Rtt106p(65-320) (Figure 65A).

La CCS attendue pour le complexe Bcd1p:Rtt106p(65-320), calculée en faisant la somme des CCS des deux partenaires libres est de 53,3 nm² (Figure 65A). Cependant, la CCS mesurée est de 44,4 nm² (Figure 65A). Cette différence de CCS peut être due à une forme d'imbrication des deux partenaires l'un dans l'autre et/ou d'une compaction lors de l'interaction ou encore un changement conformationnel. Ce résultat peut être en accord avec celui de l'analyse HDX effectuée sur la protéine Bcd1 et celui de dichroïsme circulaire sur le complexe, où une modification structurale importante semble intervenir lors de l'interaction entre les deux protéines. Il n'est toutefois pas possible d'exclure un changement conformationnel du côté de la protéine Rtt106.

3.2.2.1.3. Cross-link

Cette dernière technique utilisée en spectrométrie de masse (Travaux réalisés par Guillaume Terral du LSMBO de Strasbourg) permet de lier de manière covalente entre elles deux protéines dont nous voulons caractériser l'interaction. Pour cela un agent pontant, le BS3 (bis-sulfosuccinimidyl suberate), est mis en présence du complexe. Cet agent possède à



Protéine	CCS expérimentales (nm ²)	CCS prédites à partir des masses (nm ²)	CCS déterminées à partir d'une structure connue
Bcd1	29,0 ± 1,5	29,8	-
Rtt106(65-320)	23,2 ± 1,9	23,4	23,5
Σ CCS _{exp}	53,3 ± 3,4	53,2	-
Rtt106(65-320):Bcd1	44,4 ± 1,4	42,3	-

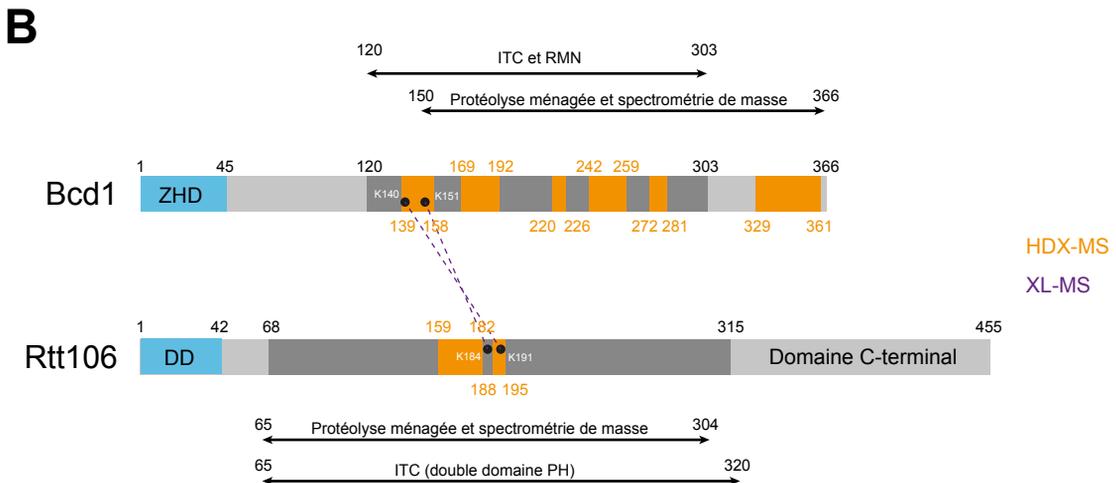


Figure 65. Mobilité ionique et Cross-Link sur le complexe Bcd1p:Rtt106p. (A) Mesure de la mobilité ionique (effectuées par G. Terral au LSMBO) des protéines Rtt106(65-320), Bcd1 et du complexe Rtt106p(65-320):Bcd1p. La mobilité ionique est liée aux sections efficaces rotationnelles ou CCS (*Collisions Cross Sections*). Le graphe représente les CCS en fonction de la masse mesurée de la protéine. Les données sont reportées dans le tableau. (B) Synthèse des résultats obtenus en spectrométrie de masse sur le complexe Rtt106p(65-320):Bcd1p. Les traits en pointillés violets relient deux lysines pontées par l'agent BS3 lors d'expériences de *Cross-Link* (XL-MS).

ses deux extrémités un ester N-hydroxysulfosuccinimide (NHS) (espacés de 11,4 Å) qui est réactif aux amines comme celles de la chaîne latérale d'un résidu lysine. Si deux de ces résidus sont proches dans l'espace et accessibles, le BS3 va pouvoir réagir avec les deux lysines en réalisant un pont covalent (ou *cross-link*) entre les résidus. Après digestion trypsique, les deux peptides contenant les deux lysines pontées resteront liés et identifiables par spectrométrie de masse.

Dans le cas du complexe Rtt106p(65-320):Bcd1p, deux de ces ponts covalents ont été identifiés. Ils relient les lysines 140 et 151 de Bcd1p avec les lysines 191 et 184 de Rtt106p respectivement (Figure 65B). Les deux lysines de la protéine Rtt106 impliquées dans ce résultat de *cross-link* appartiennent toutes deux à l'hélice $\alpha 3$. Cette hélice avait déjà été identifiée par les résultats de HDX précédents, comme impliquée dans l'interaction avec Bcd1p de manière directe ou indirecte (Figure 62), ce qui ajoute du crédit au fait que cette hélice soit importante dans l'interaction de Rtt106p avec son partenaire. Les deux lysines pontées pour la protéine Bcd1 sont dans la région N-terminale du domaine 120-303, et sont également comprises dans un peptide identifié par HDX (Figure 64). De plus, les différents tests menés sur la protéine Bcd1 lors de l'optimisation des fragments minimums conservant l'interaction avec Rtt106p montrent que la région 130-149 est nécessaire (Figures 58 et 60). Tous ces résultats qui sont synthétisés sur la Figure 65B sont donc en accord les uns avec les autres et indiquent qu'à *minima* la région N-terminale du domaine 120-303 de la protéine Bcd1 est essentielle à l'interaction avec la protéine Rtt106 et qu'elle réalise probablement des contacts directs avec celle-ci.

3.2.2.2. Titrages suivis par RMN

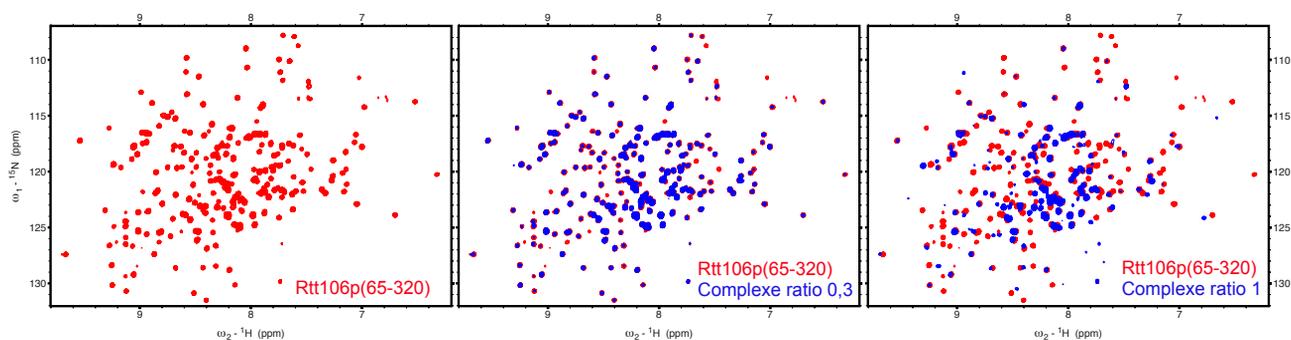
Les analyses par spectrométrie de masse ont permis d'identifier des peptides impliqués dans l'interaction entre les deux protéines Bcd1 et Rtt106. Afin d'obtenir une résolution plus précise, à l'échelle de l'acide aminé, une approche de titrage suivi par RMN a été mise en œuvre. Dans notre cas, le complexe entre les deux protéines étant d'un poids moléculaire élevé pour la RMN (51,5 kDa), les deux protéines ont été produites en milieu minimum contenant de l'eau deutérée ($^2\text{H}_2\text{O}$). De plus, les spectres ont été enregistrés en TROSY (une variante de l'expérience HSQC), ce qui permet d'augmenter la résolution.

3.2.2.2.1. Titrage de la protéine Rtt106(65-320)

Lors du titrage de la protéine Rtt106(65-320) marquée ^{15}N par la protéine Bcd1(120-303) aucune tâche de corrélation ne se déplace, mais beaucoup perdent en intensité avec l'ajout de Bcd1p en proportion croissante (Figure 66A). Au ratio 1:1, une partie non négligeable des pics a disparu, ou au moins fortement perdu en intensité par rapport à la référence (Figure 66A). Nous notons également l'apparition de quelques nouveaux pics. La disparition de pics suivie de l'apparition de nouveaux pics est caractéristique d'un échange lent en RMN entre les deux états. Comme l'échange est lent, les pics observés correspondent soit aux résidus des espèces libres, soit aux résidus des espèces liées. Ainsi, en augmentant le ratio entre les deux partenaires nous modifions les proportions des deux pics (libre et lié), ce qui conduit à la disparition de l'un au profit de l'autre. Comme dans ce cas nous ne pouvons pas suivre le déplacement des pics, nous avons regardé la baisse d'intensité des pics lors de l'ajout de la protéine Bcd1(120-303). Le graphique présenté en Figure 66B représente le ratio établi entre les pics des spectres de Rtt106p(65-320) en présence de Bcd1p(120-303) à raison d'un ratio de 0,3 (0,3 Bcd1p pour 1 Rtt106p) et libre (référence). Sur ce graphique, chaque point représente une tâche de corrélation, et donc un résidu (en abscisses) et le ratio de ses intensités entre les deux spectres (en ordonnées). Seuls les pics sans chevauchement ont été analysés dans cette expérience pour conserver la plus grande fiabilité possible des résultats. Sur ce graphique (Figure 66B), deux régions majeures apparaissent distinctement. Ces deux régions correspondent aux domaines PH1 et PH2. Les médianes des valeurs pour chacun de ces domaines sont représentées par des droites horizontales orange pour le domaine PH1 et verte pour le domaine PH2. Ces médianes sont de 0,41 pour le PH1 (résidus 68 à 204) et 0,58 pour le PH2 (résidus 217 à 299). Il y a donc un écart significatif entre les valeurs de ces deux domaines. Les deux autres régions apparaissant sur le graphique et ne faisant partie d'aucun domaine PH correspondent à la boucle entre les deux domaines PH (205-216) et à l'extension terminale (300-320) de la protéine. Ces deux régions étant beaucoup plus flexibles et supposées non touchées par l'interaction avec la protéine Bcd1, il est cohérent que leurs valeurs de ratio soient élevées (proches de 1). Au sein des deux domaines PH, certains résidus présentent des écarts importants par rapport aux autres. Ces résidus correspondent à des pics très faibles en intensité sur le spectre qui ont déjà totalement disparus (ou presque) au ratio 0,3. C'est la raison pour laquelle ces résidus présentent de faibles valeurs de ratio.

Le résultat de ce titrage montre que l'interaction avec la protéine Bcd1 fait intervenir le domaine PH1 qui comprend les peptides déjà identifiés par spectrométrie de masse et pas le

A



B

Titration de Rtt106p(65-320) par Bcd1p(120-303) au ratio de 0,3

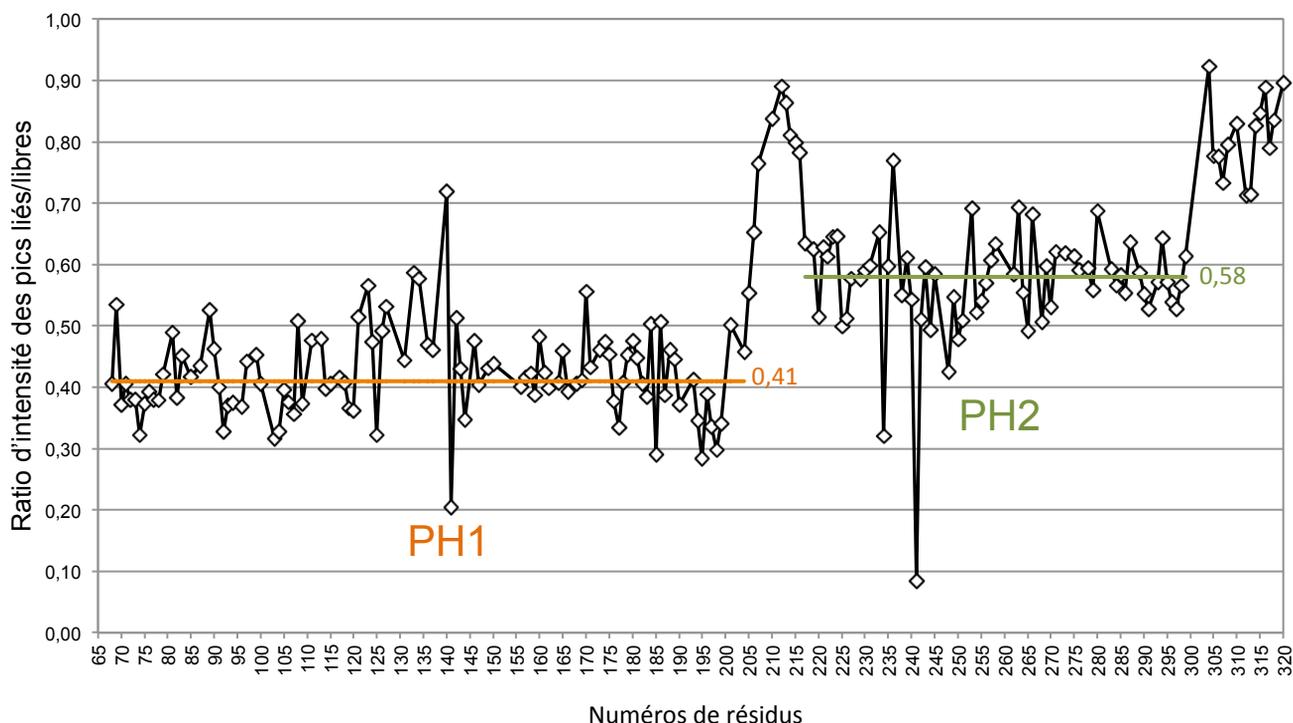


Figure 66. Titration suivie par RMN de la protéine Rtt106(65-320) par la protéine Bcd1(120-303). (A) Comparaison des spectres TROSY-HSQC ^1H - ^{15}N obtenus pour la protéine Rtt106(65-320) marquée ^2H - ^{13}C - ^{15}N seule (spectre de gauche) ou en présence de la protéine Bcd1(120-303) ^2H aux ratios de 0,3 (spectre central) et 1 (spectre de droite). (B) Graphique représentant le ratio d'intensité des pics obtenus pour chaque résidu entre l'expérience TROSY-HSQC ^1H - ^{15}N au ratio 0,3 et l'expérience référence. Les droites orange et verte représentent les médianes des valeurs pour les domaines PH1 (résidus 68 à 204) et PH2 (résidus 217 à 299) respectivement. Les deux zones non comprises dans les domaines PH correspondent à la boucle entre les deux domaines (205 à 216) et à l'extension terminale (300 à 320) de la protéine Rtt106.

reste de la protéine Rtt106. A ce stade, il ne semble pas possible de définir avec plus de précision un ou des résidus important(s) au sein du domaine PH1 pour l'interaction avec Bcd1p.

3.2.2.2.2. *Titration de la protéine Bcd1(120-303)*

La même expérience de titration par RMN a été réalisée dans l'autre sens avec la protéine Bcd1 marquée. Le résultat de ce titration présente des résultats similaires à ceux obtenus avec la protéine Rtt106, avec des pics disparaissant peu à peu dû au régime d'échange lent (Figure 67A). Le graphique montrant les valeurs de ratio entre les expériences à un ratio de 0,3 (liées) et de référence (libres) en fonction des résidus (Figure 67B) ne fait, cette fois-ci, pas ressortir une région distincte d'interaction sur la protéine Bcd1. Cependant, même s'il n'est pas non plus possible d'isoler un ou quelques résidu(s) important(s) pour l'interaction, plusieurs petites régions se distinguent. Celles-ci sont retrouvées en diverses positions de la séquence et qui sont encerclées en rouge sur la Figure 67B (régions 124-128, 137-151, 181-186, 244-258 et 278-284).

Ces données, comme pour les résultats obtenus par spectrométrie de masse HDX (Figure 64), peut signifier que plusieurs régions distinctes de la protéine Bcd1 sont situées à l'interface avec Rtt106 ou que la structure 3D de la protéine Bcd1 fait que toutes ces régions espacées dans la séquence se retrouvent proches les unes des autres dans l'espace et forment une seule et même zone d'interaction.

3.3. Etudes structurales tridimensionnelles

Tous les travaux décrits précédemment sur l'analyse de l'interface entre les protéines Bcd1 et Rtt106 nous ont permis d'identifier des domaines d'interaction et plus ou moins précisément des régions impliquées dans l'interaction. Cependant il n'a pas été possible par de telles méthodes de déterminer des résidus d'importance. La détermination de la structure tridimensionnelle du complexe formé par ces deux protéines permettrait d'identifier de tels résidus. Compte tenu de la taille de ce complexe, la cristallographie aux rayons X semble la technique la plus appropriée.

Cependant, et dans un premier temps, nous avons cherché à déterminer la structure 3D du domaine 120-303 de la protéine Bcd1 en solution par RMN. La structure 3D de ce domaine permettrait de pouvoir analyser les résultats de spectrométrie de masse et de titration par RMN. De plus, les résultats décrits précédemment tendent à montrer un possible

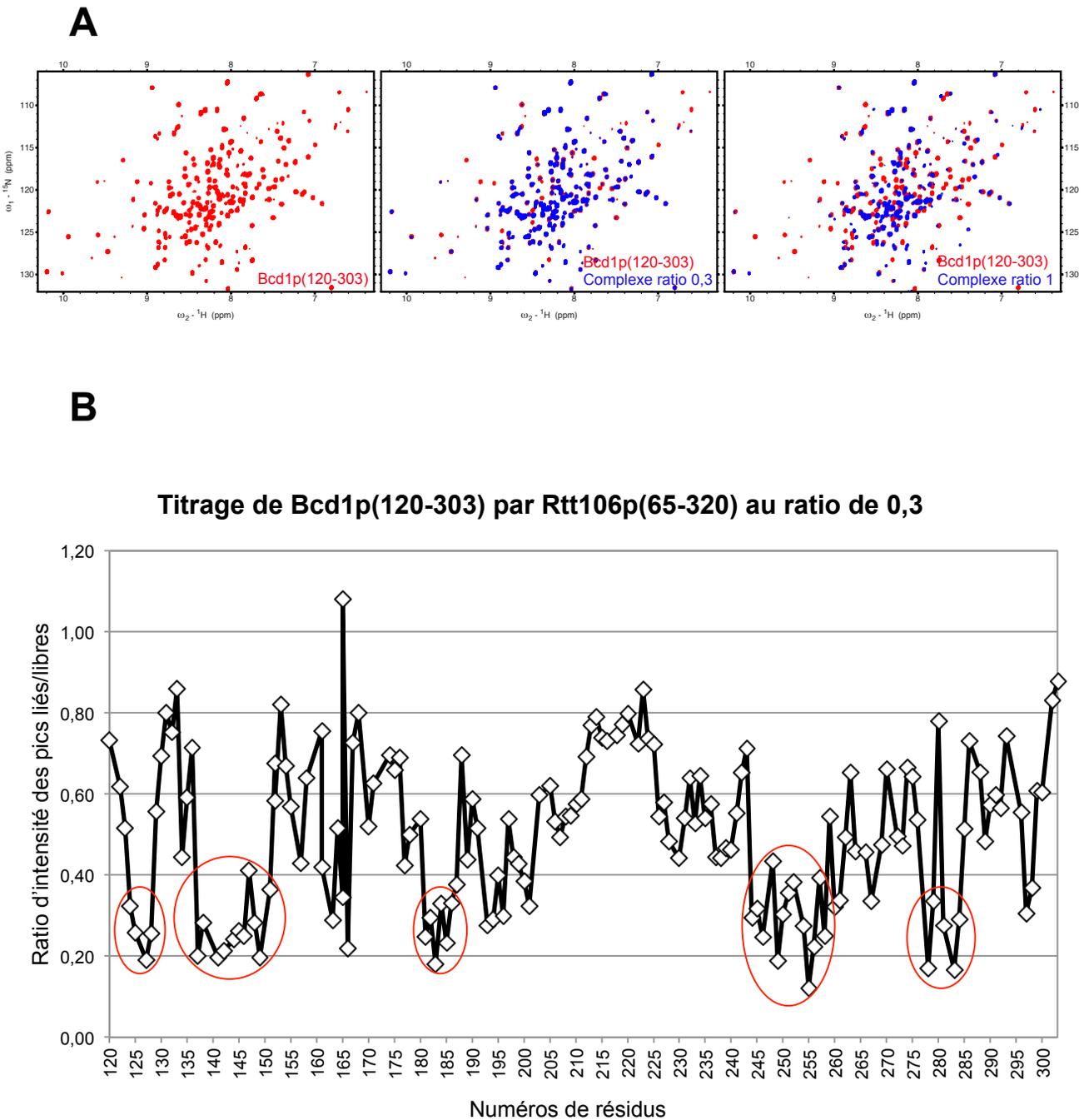


Figure 67. Titration suivie par RMN de la protéine Bcd1(120-303) par la protéine Rtt106(65-320). (A) Comparaison des spectres TROSY-HSQC ^1H - ^{15}N obtenus pour la protéine Bcd1(120-303) ^2H - ^{13}C - ^{15}N seule (spectre de gauche), ou en présence de la protéine Rtt106(65-320) ^2H aux ratios de 0,3 (spectre central) et 1 (spectre de droite). (B) Graphique représentant le ratio d'intensité des pics obtenus pour chaque résidu entre l'expérience TROSY-HSQC ^1H - ^{15}N au ratio 0,3 et l'expérience référence. Les cercles rouges mettent en évidence les régions de Bcd1 les plus touchées par l'interaction.

changement de conformation de la protéine Bcd1 suite à l'interaction avec la protéine Rtt106. Connaître la structure de la forme libre pourrait donc également être intéressant en comparaison de la structure du même domaine au sein du complexe formé avec Rtt106p.

3.3.1. Analyse structurale de la protéine Bcd1(120-303)

3.3.1.1. Détermination de la structure 3D en solution de la protéine Bcd1(120-303) par RMN

Le domaine 120-303 de la protéine Bcd1 est un domaine stable et bien exprimé chez *E. coli* de façon recombinante (Figure 68A). Le rendement de la purification de ce domaine est d'environ 26 mg par litre de culture pour une préparation en milieu minimum et marquée aux isotopes ^{13}C et ^{15}N . Ce domaine apparaît monomérique d'après le volume d'éluion en chromatographie d'exclusion de taille et homogène d'après le profil DLS (Figures 68B et C). Après analyses, il apparaît que les spectres sont de meilleure qualité dans un tampon phosphate à pH 6,4 contenant 300 mM de NaCl. La protéine étudiée compte 188 acides aminés incluant les résidus restant de l'étiquette 6xHistidines après coupure par la PreScission. Pour réaliser l'attribution des chaînes principales nous avons choisi de travailler sur un échantillon totalement deutéré qui a été produit dans un milieu contenant 100% de $^2\text{H}_2\text{O}$. L'attribution des chaînes latérales a été réalisée à partir d'échantillons 65% deutéré ou non deutéré. Les spectres nécessaires à l'attribution de la protéine ont été enregistrés sur un spectromètre 600 MHz à cryosonde triple résonance disponible sur la plateforme de Biologie Structurale de la FR BMCT 3209.

L'HSQC ^1H - ^{15}N présentée en Figure 69 montre l'empreinte de la protéine Bcd1(120-303). Sur ce spectre, tous les acides aminés sont représentés par une tâche de corrélation correspondant à la corrélation entre l'azote et le proton (N-H) de la liaison peptidique, hormis les résidus proline qui n'ont pas de proton amide et le résidu Q143 qui n'est pas visible. Les autres tâches présentes en haut à droite sur le spectre correspondent aux corrélations N-H présentes dans les chaînes latérales de certains résidus et ne sont pas annotées.

Suite à ce travail, 83,2% des noyaux de la protéine (^1H - ^{13}C - ^{15}N) ont été attribués. Cette attribution a permis le calcul de structures avec l'utilisation de spectres NOESY obtenus sur les noyaux azotes, carbones aliphatiques et aromatiques, sur des échantillons deutérés ou non. De plus, afin d'améliorer la résolution des spectres NOESY comportant beaucoup de pics, des expériences NOESY ont également été enregistrées sur un spectromètre à haut champ

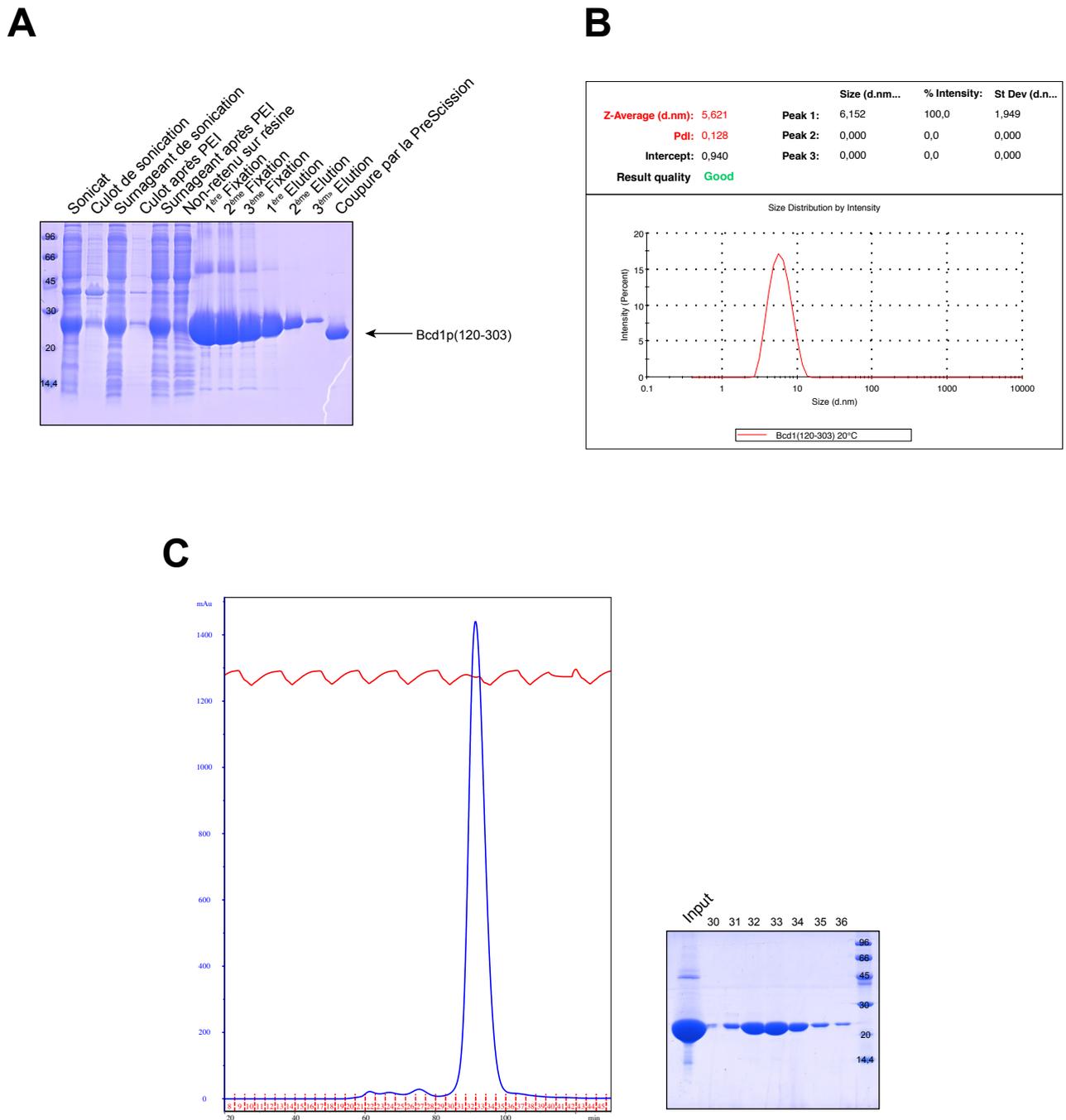


Figure 68. Purification de la protéine Bcd1(120-303). (A) Etapes de purification de la protéine Bcd1(120-303) analysées sur SDS-PAGE 12,5%. (B) Résultats de DLS sur la protéine Bcd1(120-303). La valeur de Z-Average correspond au diamètre moyen des particules en solution. Le PdI est l'indice de polydispersité, 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. (C) Chromatogramme de la chromatographie d'exclusion de taille effectuée avec une colonne Superdex S75 (16:60) sur la protéine Bcd1p(120-303) en fin de purification et dépôt des fractions correspondantes sur SDS-PAGE 12,5%. La courbe bleue sur le chromatogramme correspond à l'UV à 280 nm et la courbe rouge à la conductimétrie.

(950 MHz) de l'ICSN (Institut de Chimie des Substances Naturelles) de Gif-sur-Yvette (France) dans le cadre d'un projet TGIR (Très Grandes Infrastructures de Recherche).

6300 contraintes NOE ont été établies à partir des différents spectres NOESY utilisés (Table 4), dont 677 n'ont pas pu être attribuées par le logiciel CYANA 3.97. Suite au calcul, les 20 structures de plus basses énergies ont été conservées et présentent des RMSD de $1,80 \pm 0,49$ Å sur les chaînes principales et de $2,28 \pm 0,38$ Å sur les chaînes latérales. Le calcul de structure doit toutefois encore être affiné, car trop de contraintes NOE restent sans attribution (677 sur 6300). Le calcul doit également encore être réalisé dans une boîte d'eau avant d'être définitif. La superposition de ces 20 structures est présentée en Figure 70A. Le résultat de ce calcul de structure montre que ce domaine de la protéine Bcd1 est bien replié et compact, avec une forte imbrication des éléments de structure secondaire (Figure 70A). Ce domaine présente un feuillet central vrillé composé de cinq brins β (les brins $\beta 5$ à $\beta 9$) pivotants. La région N-terminale est composée de deux feuillets β antiparallèles eux-mêmes constitués des brins $\beta 1$ et $\beta 2$ pour le premier et $\beta 3$ et $\beta 4$ pour le second (Figure 70B). Ces deux feuillets encadrent un tour d'hélice α correspondant à l'hélice $\alpha 1$. Entre les brins $\beta 6$ et $\beta 7$ du feuillet central se trouvent trois hélices, la petite hélice $\alpha 2$ de cinq résidus et les deux plus grandes hélices $\alpha 3$ et $\alpha 4$. Ces deux hélices de 11 et 13 résidus respectivement, sont plaquées contre une face du feuillet central (Figure 70B). La superposition des structures est moins bonne sur l'hélice $\alpha 4$ et sur la boucle $\alpha 3$ - $\alpha 4$ reliant ces deux hélices. Malheureusement, très peu de contraintes NOE sont présentes dans cette région, il y a donc moins de contraintes de distances disponibles pour restreindre les possibilités de positionnement de cette partie de la structure. L'hélice $\alpha 3$ est plaquée contre les brins $\beta 5$ et $\beta 6$. La boucle entre les brins $\beta 8$ et $\beta 9$ contient également la petite hélice $\alpha 5$ de 5 résidus. Enfin, la dernière hélice $\alpha 6$ de trois résidus se situe vers l'extrémité C-terminale.

La région 120-149 (région N-terminale) comprend les brins $\beta 1$ et $\beta 2$, ainsi que l'hélice $\alpha 1$. Cette région, bien qu'effectuant de nombreux contacts avec le reste de la protéine, apparaît comme un module indépendant situé sous le feuillet β central. Même si elle est bien repliée et s'intègre parfaitement dans la structure de ce domaine, il semble tout à fait possible de la supprimer sans pour autant affecter la structure du reste du domaine. Ceci explique les résultats obtenus lors des tests effectués sur les différents fragments de la protéine Bcd1 (partie 3.2.1.2.1). En effet, l'HSQC ^1H - ^{15}N du fragment 149-303 était caractéristique d'une protéine toujours repliée. De plus, le spectre de ce fragment n'était pas superposable avec celui du fragment 120-303. Etant donné la position de la région 120-149 dans la structure du

Contraintes NOE	
Nombre total de pics NOE	6300
Pics NOE non-attribués	677
NOE Séquentiels ($ i - j \leq 1$)	3425
NOE à moyennes distances ($1 < i - j < 5$)	606
NOE à longues distances ($ i - j \geq 5$)	984
CYANA target function	3,28
Diagramme de Ramachandran (%)	
Régions favorables	81,5
Régions permises	18,4
Régions généreusement permises	0,1
Régions non-permises	0,0
RMSD moyens (Å)	
Chaînes principales	1,80 ± 0,49
Chaînes latérales	2,28 ± 0,38

Table 4 : Statistiques RMN des 20 structures en solution de la protéine Bcd1(120-303) calculées avec le logiciel CYANA 3.97 (Schrödinger, LLC). Les RMSD ont été calculés sur les résidus 126 à 299.

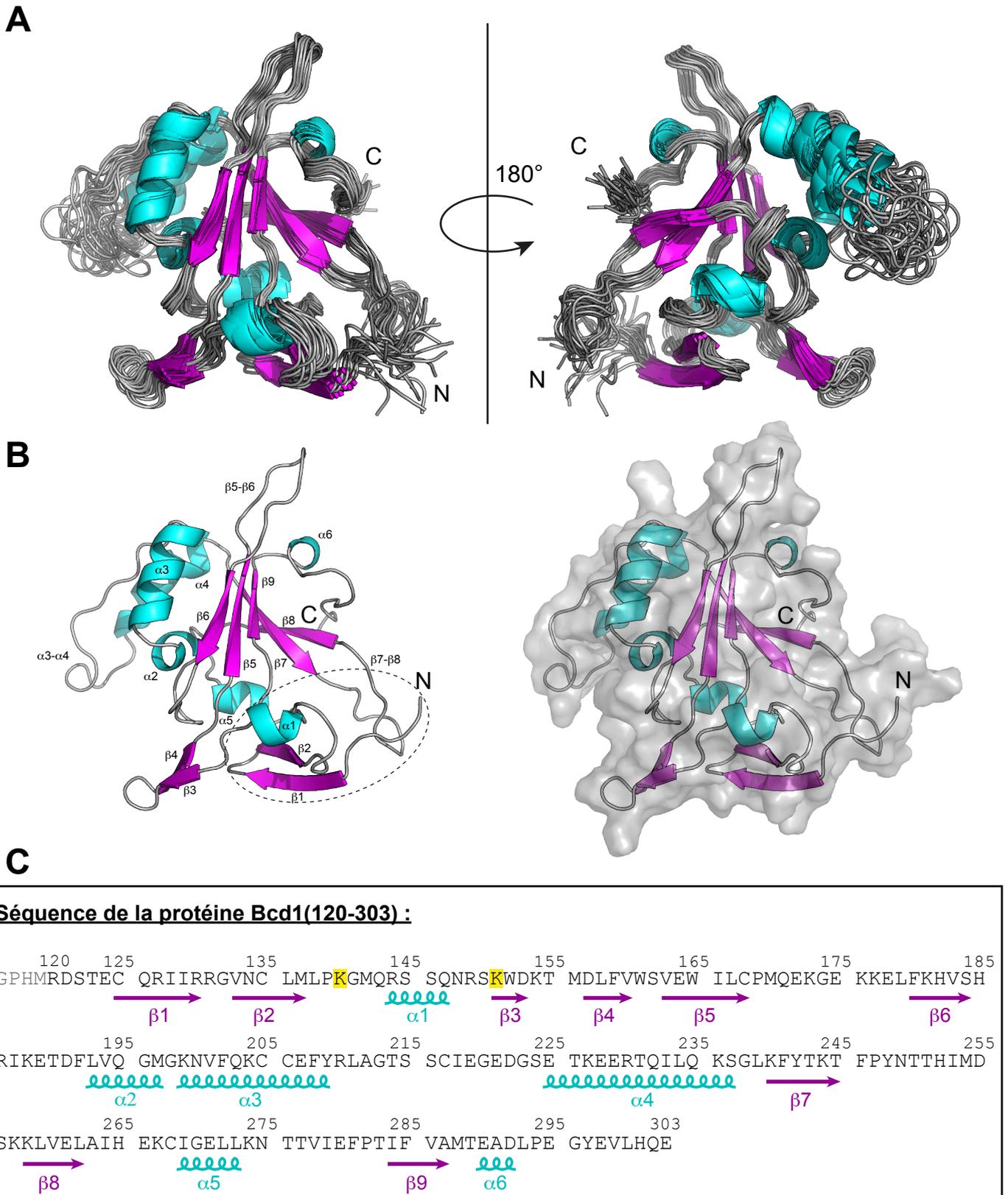


Figure 70. Structure 3D en solution de la protéine Bcd1(120-303). (A) Superposition des 20 structures calculées avec CYANA présentant la plus basse énergie, représentées en cartoon, avec coloration des structures secondaires. Les hélices α sont colorées en cyan et les brins β en magenta. (B) Représentation uniquement de la structure de plus basse énergie, avec annotation des différents éléments de structure secondaire (à gauche) et dans l'enveloppe de surface (à droite). Les régions annotées $\beta 5-\beta 6$, $\alpha 3-\alpha 4$ et $\beta 7-\beta 8$ correspondent à des boucles. La région encadrée en pointillés correspond à la séquence 120-149. (C) Séquence de la protéine Bcd1(120-303) avec report des éléments de structure secondaire en dessous. Les quatre premiers résidus (GPHM, en gris et non numérotés) correspondent aux résidus de l'étiquette 6xHistidines restants après coupure par la 3C. La séquence de la protéine débute au résidu R120. Les résidus surlignés en jaune correspondent aux lysines réalisant des ponts avec Rtt106p lors des expériences de cross-link.

domaine, il est compréhensible que les déplacements chimiques de nombreux autres résidus soient affectés lors de la suppression de ce fragment. En effet, celui-ci se situe au contact d'une large zone du reste de la protéine, comprenant en particulier les boucles $\alpha 5$ - $\beta 9$ et $\beta 7$ - $\beta 8$, ainsi que le feuillet constitué des brins $\beta 3$ et $\beta 4$.

Si la protéine Bcd1(120-303) possède un repliement compact et globulaire, elle présente toutefois des disparités en fonction de l'axe d'observation choisi. En effet, si la protéine est orientée face au feuillet β central (Figure 71A) sa largeur maximale est de 39 Å, et sa longueur de 54,4 Å. En revanche, si l'axe d'observation est celui du feuillet central, les largeurs et longueurs maximales sont de 57,2 Å et 55,3 Å respectivement (Figure 71B), ce qui donne au final une forme globalement carrée. Cela est dû au fait que les éléments de structure secondaire ne s'agencent pas tout autour du feuillet central, mais uniquement sur une des faces de ce feuillet, ainsi que sur une extrémité vers laquelle pointent les brins $\beta 6$ et $\beta 7$.

Enfin, l'analyse du potentiel électrostatique de ce domaine révèle une répartition non homogène des résidus chargés à la surface de la protéine. Cette représentation montre une face de la protéine fortement chargée positivement (Figure 72B à gauche), tandis que la face opposée est plus fournie en résidus chargés négativement (Figure 72B à droite).

3.3.1.2. Surface d'interaction avec la protéine Rtt106(65-320)

Il est désormais possible d'intégrer les données obtenues précédemment sur l'interface entre les protéines Bcd1 et Rtt106 au regard de la structure 3D de la protéine Bcd1(120-303). Les peptides déterminés lors de l'expérience d'HDX ont été représentés sur la structure RMN de la protéine Bcd1(120-303) (Figure 73A). La région en bleu correspond au peptide dont l'incorporation du deutérium était accélérée après formation du complexe (Figure 64). Ce peptide correspond à la boucle $\alpha 3$ - $\alpha 4$ (Figure 70B), qui apparaît très flexible sur la structure RMN (Figure 70). Il est donc possible que ce résultat obtenu en HDX soit un artéfact dû à la forte flexibilité de cette zone. Les peptides représentés en orange sur la Figure 73A correspondent à une incorporation du deutérium ralentie et touchent majoritairement une face de la protéine. Les résultats de la titration RMN ont aussi été reportés sur la structure RMN (Figure 73B). Seuls les résidus dont les pics de l'HSQC ^1H - ^{15}N ont subi une forte baisse d'intensité entre les HSQC avec les formes libre ou en complexe ont été représentés. Les résidus colorés en orange sont ceux dont le ratio d'intensité (forme liée/forme libre) est inférieur ou égal à 0,3 et ceux en rouge à 0,2 (Figure 67). Ces résidus se concentrent tous dans la même zone de la protéine. La comparaison des deux résultats montre une plus grande zone

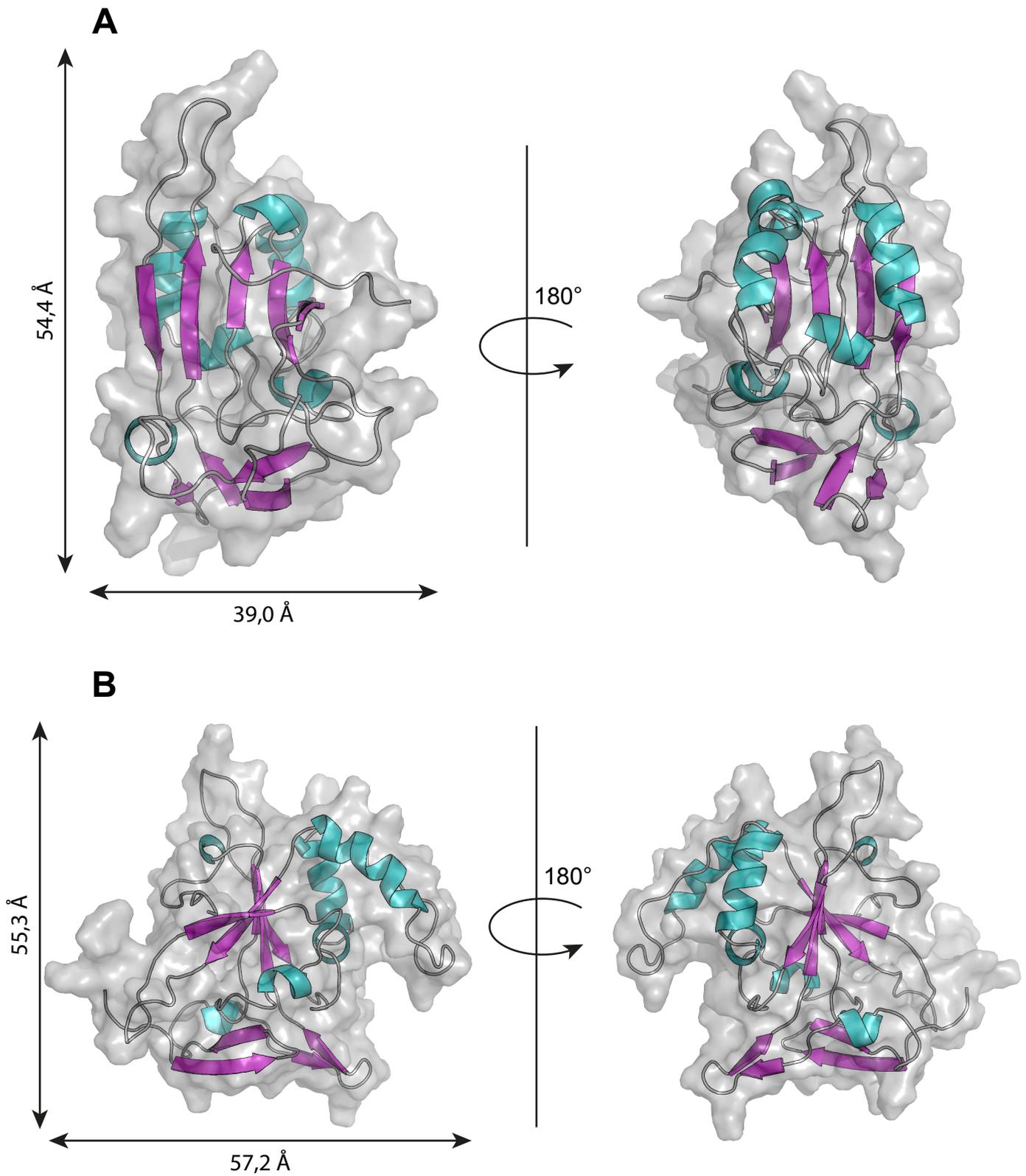


Figure 71. Dimensions de la structure 3D de la protéine Bcd1(120-303). Représentations en cartoon de la structure 3D de la protéine Bcd1(120-303) dans une enveloppe de surface (semi-transparente) selon deux angles d'observation. Les longueurs des deux côtés sont indiquées sur les doubles flèches dans chaque cas.

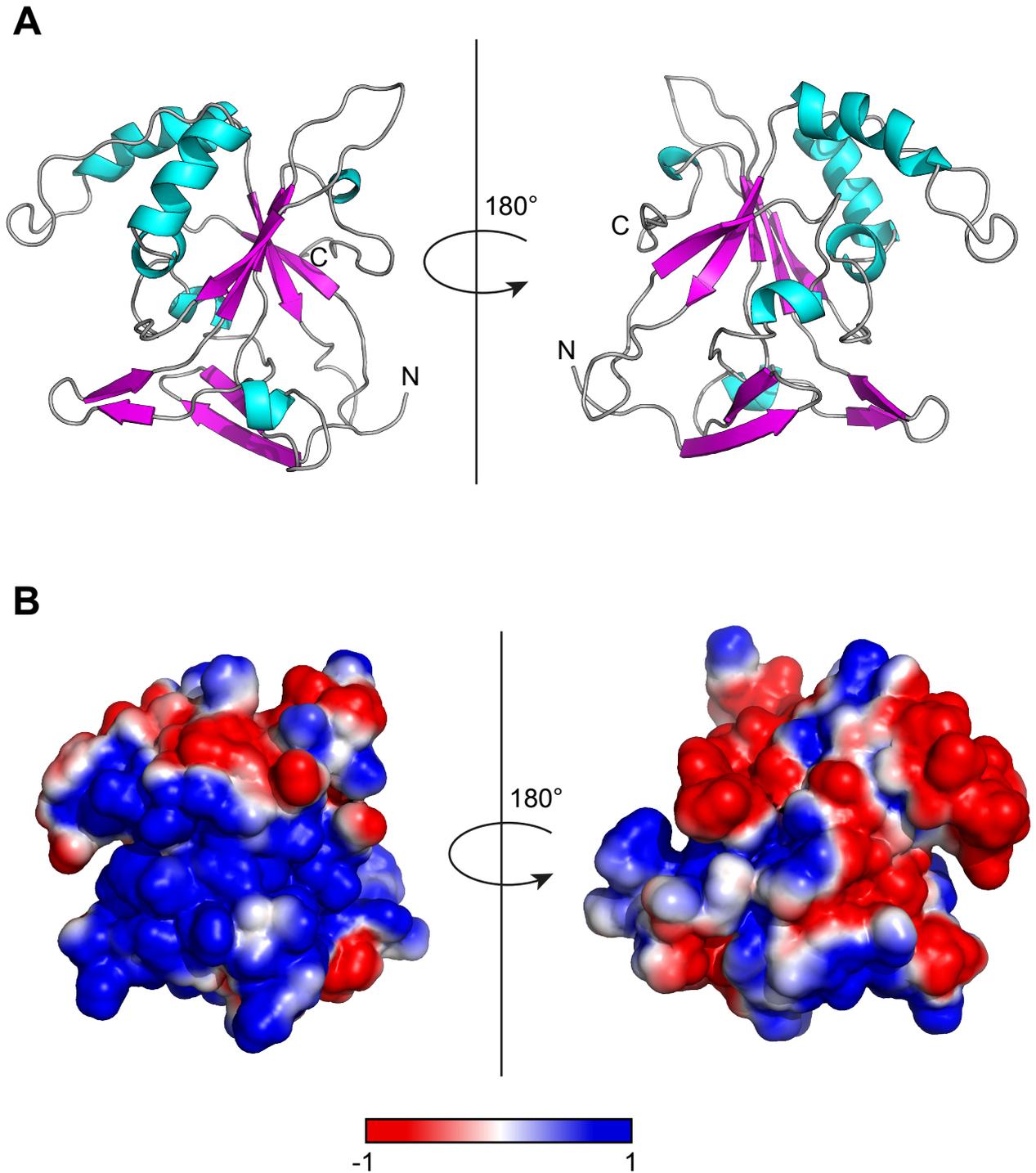


Figure 72. Potentiel électrostatique de la protéine Bcd1(120-303). (A) Représentation en cartoon, de la protéine Bcd1p(120-303) et (B) représentation du potentiel électrostatique des mêmes faces, réalisée avec le module APBS du logiciel PyMOL (Schrödinger, LLC). L'échelle du potentiel électrostatique (allant de -1 à 1 kT/e) est représentée en-dessous.

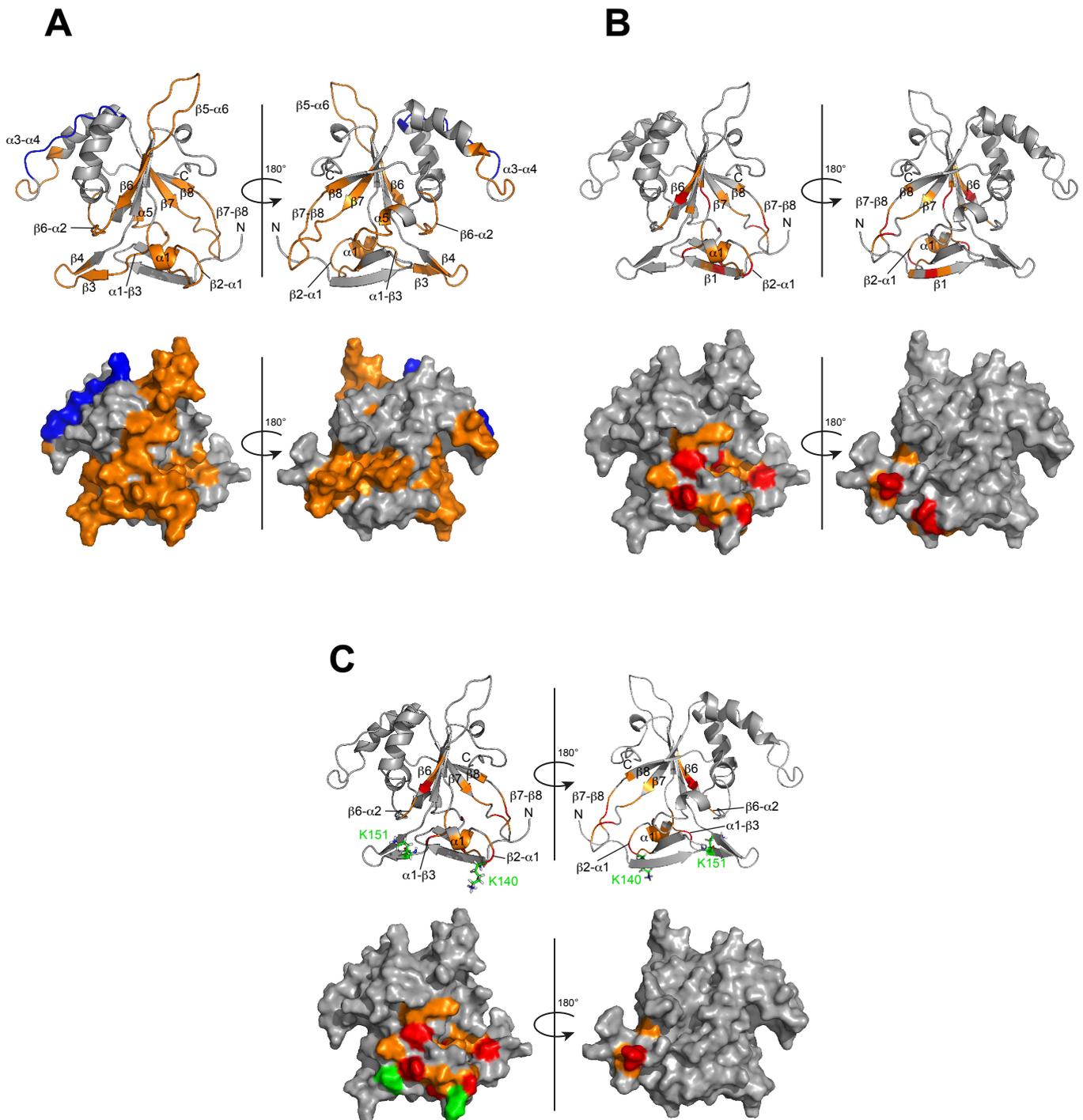


Figure 73. Régions de la protéine Bcd1(120-303) impliquées dans l'interaction avec la protéine Rtt106. (A) Représentation sur la structure RMN de la protéine Bcd1(120-303) des peptides dont l'écart de vitesse de deutération est plus important lors de l'expérience d'HDX. Les peptides représentés en orange sont ceux dont la différence de vitesse avant et après formation du complexe est positive (vitesse ralentie après formation du complexe). Le peptide en bleu est celui dont cette différence est négative (vitesse accélérée après formation du complexe). (B) Représentation sur la structure RMN de la protéine Bcd1(120-303) des résidus dont la perte d'intensité des pics correspondants de l'HSQC est la plus grande lors de la titration par RMN. Les résidus en orange sont ceux dont le ratio est inférieur ou égal à 0,3 (différence prononcée) et ceux en rouge à 0,2 (différence très prononcée). (C) Synthèse des représentations (A) et (B). Seuls les résidus communs aux deux expériences ont été conservés. Les lysines K140 et K151 réalisant des ponts avec Rtt106 lors de l'expérience de *cross-link* sont représentées en bâtonnets et colorées en vert.

d'interaction potentielle avec la technique HDX, ce qui est probablement dû à la faible résolution de cette technique avec des peptides de l'ordre d'une quinzaine d'acides aminés. Sur la représentation de la Figure 73C, seuls les résidus communs aux deux techniques ont été conservés et comprennent l'hélice α_1 , les boucles β_2 - α_1 , α_1 - β_3 , β_6 - α_2 et β_7 - β_8 et les brins β_6 , β_7 et β_8 . Etant donné la convergence des résultats, il semble donc que l'interaction avec la protéine Rtt106 se fasse dans cette zone, où sont également comprises les deux lysines K140 et K151, situées dans la boucle β_2 - α_1 et le brin β_3 respectivement, effectuant des pontages avec Rtt106p lors de l'expérience de *cross-link* (Figures 65B et 73C).

3.3.2. Analyse Structurale du complexe entre les protéines Bcd1 et Rtt106

3.3.2.1. Analyse SAXS (Small Ange X-ray Scattering)

Puisque les structures des protéines libres Bcd1(120-303) et Rtt106(65-301) sont connues désormais, nous avons entrepris une analyse de ces deux protéines libres ou en complexe par les techniques de SAXS et MALLS (Multi-Ange Laser Light Scattering). L'intérêt de la technique de SAXS est qu'elle permet d'obtenir une enveloppe de particules en solution sans restriction de taille ou nécessité de marquage, tandis que le MALLS renseigne sur la masse moléculaire des particules en solution. Les protéines Bcd1(120-303) et Rtt106(65-301) libres ou en complexe ont donc été analysées au synchrotron SOLEIL de Gif-sur-Yvette, sur la ligne SWING par Aurélien Thureau dans le cadre d'une collaboration.

Les résultats de l'expérience de MALLS se présentent sous forme d'un pic UV et de masses détectées tout au long du profil d'élution. La masse moléculaire de l'échantillon analysé est donnée par la valeur de masse observée sous le pic correspondant à la protéine sur le profil d'élution. La masse observée doit être stable sous le pic (plateau). L'analyse MALLS de la protéine Bcd1(120-303) nous indique que la masse moléculaire de l'échantillon est de 25,6 kDa (Figure 74A), avec un second plateau à 24,7 kDa (écart négligeable). La masse théorique de ce fragment étant de 21,8 kDa, les résultats de MALLS montrent que la protéine est présente sous forme de monomère en solution à 10 mg/mL (450 μ M). Le profil d'élution R_g et I_0 , obtenu après passage sur colonne HPLC suivie d'une analyse SAXS, montre que le R_g sous le pic correspondant à Bcd1p(120-303) est stable avec des valeurs comprises entre 21 et 22 Å (Figure 74B). A partir de ces données, le logiciel Saxes_MoW2 nous donne une estimation de la masse sous le pic de 22,5 kDa, ce qui est cohérent avec les résultats de MALLS précédents. Ces résultats ont ensuite été analysés avec le logiciel Primus, qui a calculé une valeur de R_g de $20,2 \text{ \AA} \pm 1,6 \text{ \AA}$, ainsi qu'un I_0 de $0,015 \text{ cm}^{-1}$. En prenant ces

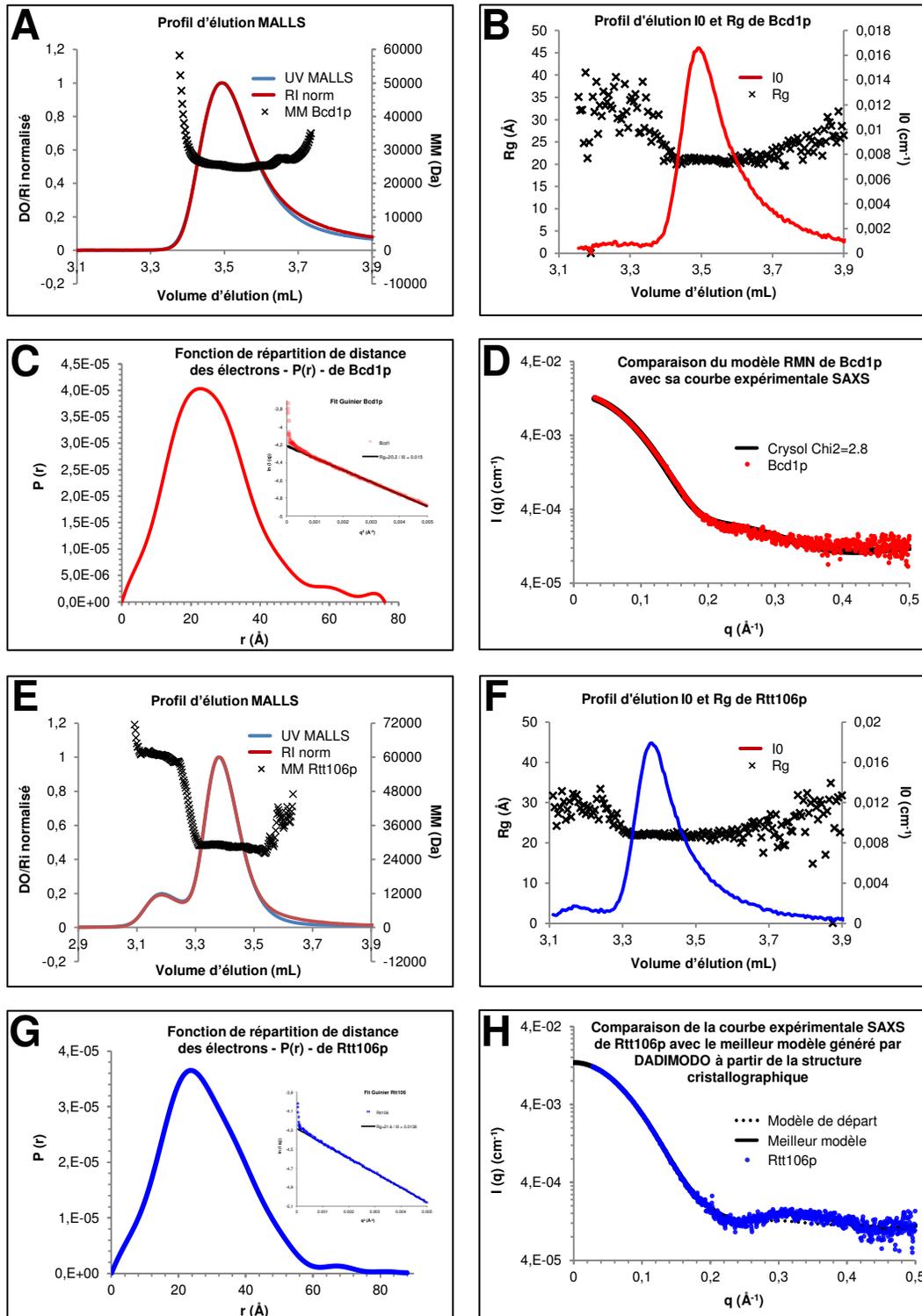


Figure 74. Analyses SAXS des protéines Bcd1(120-303) (A à D) et Rtt106(65-301) (E à H) libres. Les expériences SAXS ont été enregistrées et traitées par Aurélien Thureau au synchrotron SOLEIL. Le même tampon a été utilisé pour tous les échantillons et était constitué de Tris 10 mM à pH 8, de NaCl à 150 mM et de 10% de glycérol.

valeurs pour bases, le fit Guinier obtenu montre une bonne superposition entre les points expérimentaux et la droite théorique, avec toutefois une remontée aux tout petits angles indiquant une probable agrégation de la protéine lors de l'acquisition des données (Figure 74C). La courbe de fonction de répartition des distances des électrons (Figure 74C) calculée pour un R_g de 21 Å et un I_0 de $0,015 \text{ cm}^{-1}$ permet d'estimer le D_{max} – plus longue distance possible entre deux électrons de la protéine – à 76 Å. Ce résultat est en accord avec la structure RMN, dont la plus longue distance observable est de 71 Å, sachant que le D_{max} tient également en compte la couche de solvation dans la structure. Enfin, la courbe SAXS expérimentale a été comparée avec la courbe théorique calculée à partir de la structure RMN avec le logiciel Crysol (Figure 74D). Les deux courbes se superposent bien. Le χ^2 calculé entre les deux courbes est de 2,81, ce qui est une bonne valeur pour la ligne SWING. Ces résultats montrent que la structure RMN est compatible avec les données SAXS.

L'analyse MALLS de la protéine Rtt106(65-301) révèle que la masse de l'échantillon est comprise entre 28,2 et 29,3 kDa (Figure 74E), ce qui est proche de la masse théorique de 27,6 kDa. Ce fragment de la protéine Rtt106 est donc monomérique en solution à 10 mg/mL (360 μM). Les valeurs de R_g sont stable sous le pic du profil d'élution R_g et I_0 et comprises entre 21 et 22 Å (Figure 74F) et la masse estimée avec le logiciel Saxes_MoW2 est de 27,9 kDa, ce qui est très proche de la valeur théorique et de celles obtenues en MALLS. Le R_g calculé est de $21,6 \text{ Å} \pm 1,6 \text{ Å}$, ainsi qu'un I_0 de $0,0138 \text{ cm}^{-1}$. Le fit Guinier obtenu se superpose bien avec la droite théorique avec là encore une remontée aux plus petits angles, dont les points ne seront pas pris en compte pour la suite. La fonction de répartition des distances des électrons indique que le D_{max} est de 88 Å (Figure 74G). Il a été impossible de comparer la courbe SAXS expérimentale avec la courbe théorique Crysol basée sur la structure cristallographique, puisque celle-ci ne comprend pas les résidus 206 à 216 pour cause d'une absence de densité électronique. Pour pouvoir comparer la structure cristallographique avec les données SAXS, cette boucle a été reconstituée avec le logiciel Modeller (modèle de départ), puis des calculs avec le logiciel Dadimodo ont permis d'améliorer la position de cette boucle de sorte que la courbe théorique qui découle de la structure correspondent à la courbe SAXS (Figure 74H). Une solution se superposant bien avec la courbe expérimentale (χ^2 de 1,29) a pu être déterminée (meilleure solution). Les résultats de SAXS sur la protéine Rtt106(65-301) sont donc compatibles avec la structure cristallographique connue.

L'analyse du complexe Rtt106p(65-301):Bcd1p(120-303) par MALLS donne une masse observée comprise entre 37,8 et 38,6 kDa, ce qui est bien plus faible que la masse théorique de 49,4 Å (Figure 75A). Cet écart entre les valeurs théorique et expérimentale est probablement dû au fait que le complexe n'est pas formé à 100% en permanence, mais qu'un échange dynamique doit avoir lieu en solution. Le calcul de masse par SAXS donne également des valeurs incorrectes. Le Rg calculé par SAXS pour ce complexe est de $27,6 \text{ \AA} \pm 2,1 \text{ \AA}$ (Figure 75B). Pour un complexe à 100% formé, le Rg devrait être égal à la racine carrée de la somme des carrés des Rg des protéines libres, soit ici de 29,6 Å, or ce n'est ici pas le cas. Enfin le Dmax calculé est de 90 Å (Figure 75C). Ces résultats de SAXS sur le complexe sont donc ininterprétables car ce complexe n'est pas formé à 100% en solution dans les conditions de tampon utilisées.

3.3.2.2. Structure par cristallographie aux rayons X

Plusieurs tests de cristallisation ont été réalisés sur les complexes Rtt106p(65-320):Bcd1p(120-331), Rtt106p(65-320):Bcd1p(120-303) et Rtt106p(65-301):Bcd1p(120-303) mais sans succès, que ce soit à partir de protéines co-exprimées ou non. Ce domaine de la protéine Rtt106 ayant déjà été cristallisé et structuralement caractérisé par d'autres équipes, le problème de cristallogenèse devait probablement venir de la protéine Bcd1. Afin de vérifier si la protéine Bcd1(120-303) ne possédait pas encore des régions non repliées ou très flexibles, une étude RMN de NOE hétéronucléaires a été entreprise.

3.3.2.2.1. Apport de la RMN : expérience de NOE hétéronucléaires

Cette expérience de NOE hétéronucléaires permet de déterminer la flexibilité de chaque acide aminé de la protéine en solution. Comme attendu, les résidus situés aux extrémités N et C-terminales sont très mobiles (Figure 76A). Cependant, les résultats de cette expérience montrent également que deux régions internes de la protéine sont bien plus mobiles que le reste de la protéine (sous un ratio de 0,4). Ces deux régions comprennent les résidus 172 à 178 et 213 à 225 (Figure 76A). Ces deux régions correspondent à deux boucles sur la structure RMN, l'une entre les brins $\beta 5$ et $\beta 6$ et l'autre entre les hélices $\alpha 3$ et $\alpha 4$ respectivement (Figure 76B). La boucle entre les brins $\beta 7$ et $\beta 8$ (249-256) présente également une mobilité mais dans une moindre mesure que les deux autres avec des ratios compris entre 0,4 et 0,6.

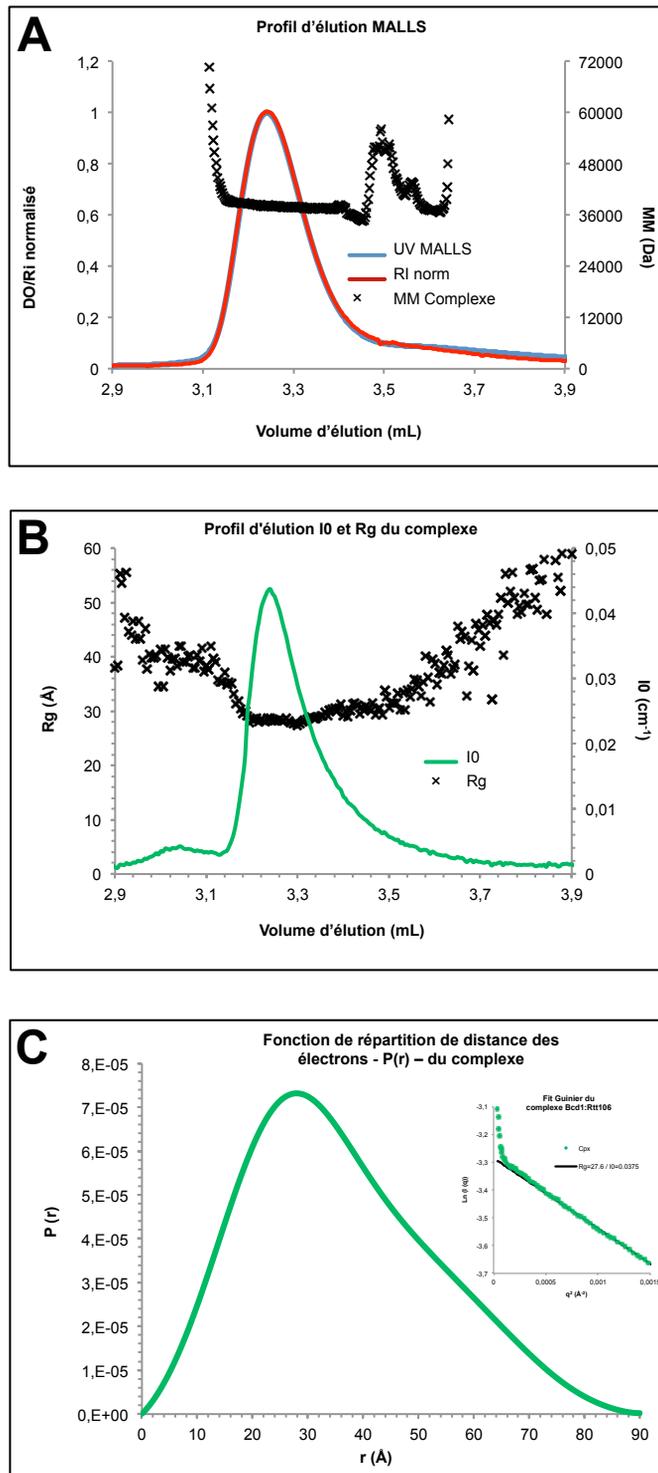
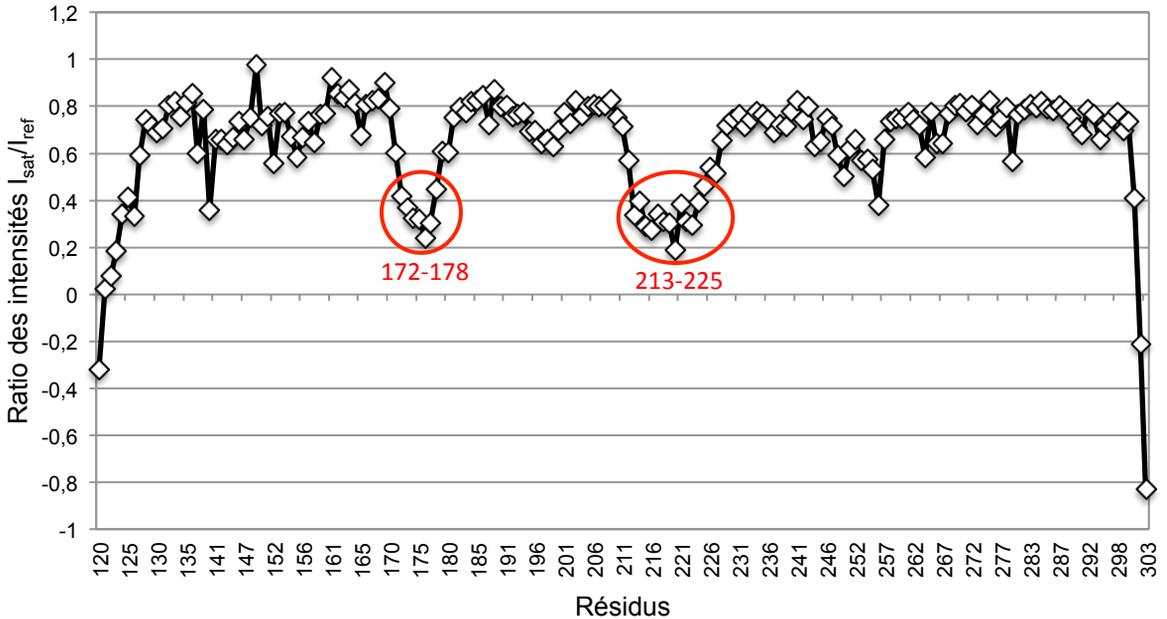


Figure 75. Analyses SAXS du complexe Rtt106p:Bcd1p. Les expériences SAXS ont été enregistrées et traitées par Aurélien Thureau au synchrotron SOLEIL. Le même tampon a été utilisé pour tous les échantillons et était constitué de Tris 10 mM à pH 8, de NaCl à 150 mM et de 10% de glycérol.

A

NOE hétéronucléaires sur la protéine Bcd1(120-303)



B

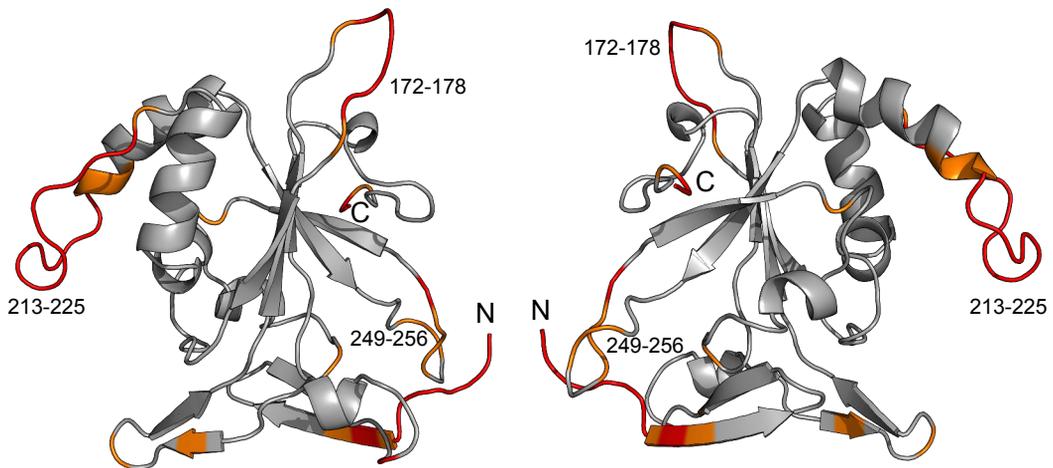


Figure 76. Etude de la flexibilité des résidus de la protéine Bcd1(120-303). (A) Expérience de NOE hétéronucléaires sur la protéine Bcd1(120-303). Le graphe représente le ratio d'intensité entre les pics de l'expérience HSQC ^1H - ^{15}N avec saturation de la magnétisation et ceux de la référence en fonction de chaque résidu. (B) Représentation sur la structure RMN de Bcd1p(120-303) des résultats de l'expérience de NOE hétéronucléaires. Les résidus dont le ratio est inférieur à 0,6 sont colorés en orange et ceux dont le ratio est inférieur à 0,4 sont colorés en rouge. Les trois boucles mobiles sont annotées.

Les deux boucles (172-178) et (213-225) apparaissant comme fortement mobiles, nous avons décidé de supprimer chacune de ces boucles dans la protéine Bcd1. Les fragments Bcd1p(120-303) Δ 172-177 et Bcd1p(120-303) Δ 220-225 portant des délétions des séquences codant ces deux régions ont donc été réalisées. Ces deux mutations doivent limiter la flexibilité des protéines variantes sans pour autant nuire à la structure 3D globale de la protéine. Les tests d'expression de ces deux nouvelles protéines mutées et recombinantes ont montré que seule la seconde est exprimée chez *E. coli*. Il est donc probable que malgré sa forte flexibilité, la boucle 172-177 joue un rôle dans la structuration et/ou la stabilité de la protéine. La protéine Bcd1(120-303) Δ 220-225 est en revanche parfaitement exprimée et soluble chez *E. coli* (rendement du même ordre que la protéine sauvage), et représentait donc un bon candidat pour de nouveaux essais de cristallisation.

3.3.2.2.2. *Le complexe Rtt106p(65-301):Bcd1p(120-303) Δ 220-225*

De nouveaux tests de cristallisation ont donc été réalisés avec cette nouvelle protéine variante et ont pu aboutir à l'identification de conditions permettant la cristallisation de ce complexe à 25 mg/mL. Il semble donc que la flexibilité de la boucle 220-225 était problématique pour la cristallisation. Les cristaux obtenus sont sous forme de plaquettes très fines dans l'une des trois dimensions. Il reste donc encore à optimiser les conditions pour la cristallogénèse de façon à obtenir des cristaux de meilleure qualité avant de pouvoir réaliser les acquisitions de spectres de diffraction au synchrotron. Le dépôt sur SDS-PAGE de plusieurs de ces cristaux montre que les deux protéines sont bien présentes dans ces cristaux, et visiblement de façon stœchiométrique (Figure 77C). Il semble donc que les conditions de cristallisation utilisées aient permis de piéger les deux protéines sous leur forme complexée, ce qui nous a été impossible d'obtenir en solution lors des études par SAXS.

Comme nous ignorons actuellement s'il sera possible de phaser les données cristallographiques en utilisant la technique du remplacement moléculaire puisque seule la structure cristallographique de la protéine Rtt106(65-301) est connue, une production du complexe Rtt106p(65-301):Bcd1p(120-303) Δ 220-225 sélénié a été réalisée (Figures 77A et B).

3.4. Compétition entre Bcd1 et les histones

Comme décrit dans la partie Introduction, la protéine Rtt106 est capable d'interagir avec les histones H3-H4. Nos données complètent avec Bcd1p le registre des interactants de

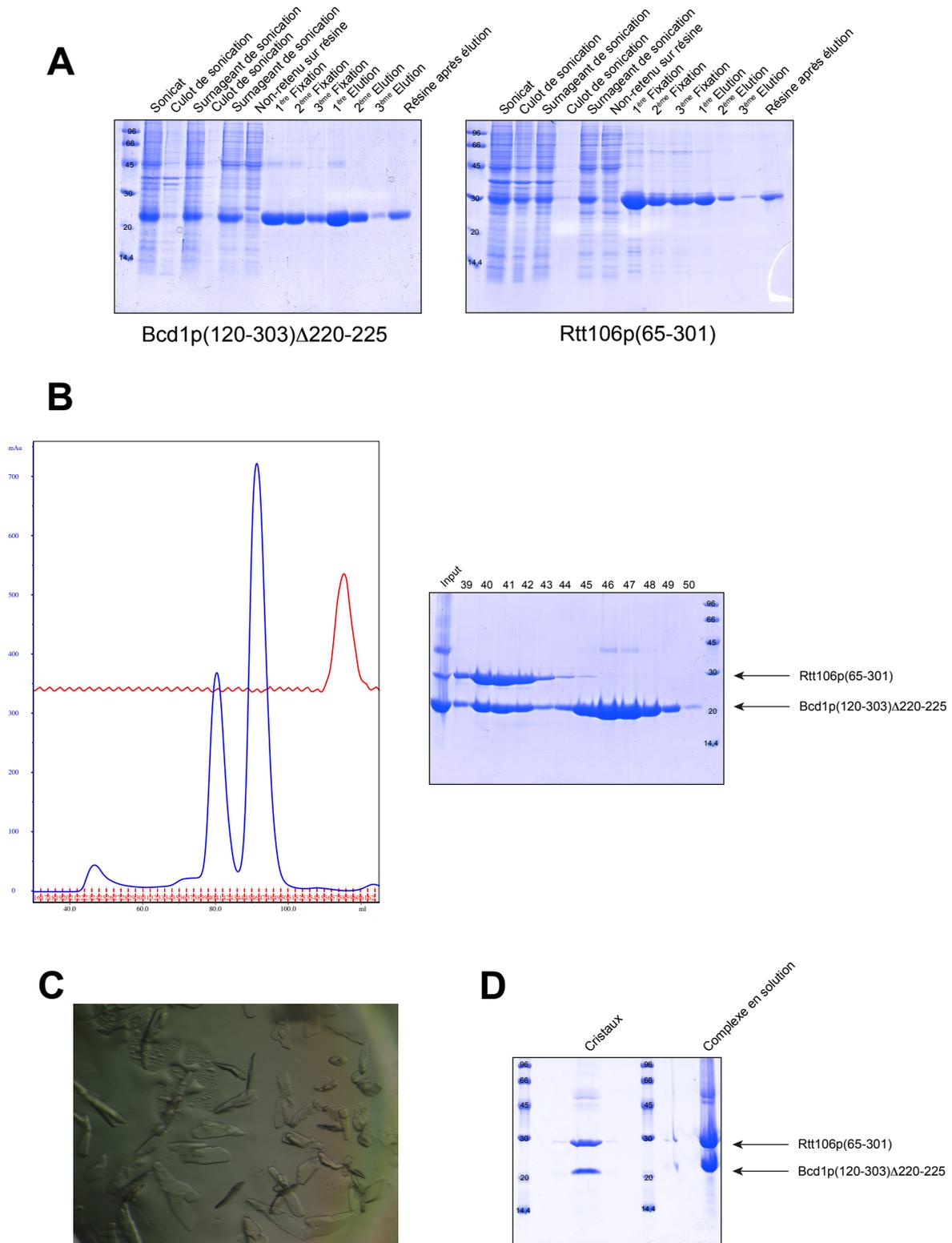


Figure 77. Analyse du complexe Rtt106p(65-301):Bcd1p(120-303)Δ220-225 sélénié. (A) Etapes de purification des protéines Rtt106(65-301) et Bcd1(120-303)Δ220-225 ayant incorporées des résidus sélénio-méthionine. **(B)** Chromatogramme de la chromatographie d'exclusion de taille (Superdex 200, colonne 16:60) effectuée sur le mélange des deux protéines et dépôt des fractions correspondantes sur SDS-PAGE 12,5%. La courbe bleue sur le chromatogramme correspond à l'UV à 280 nm et la rouge à la conductimétrie. Le débit du tampon à travers la colonne est de 0,2 mL/min. Le premier pic correspond au complexe Rtt106p(65-301):Bcd1p(120-303)Δ220-225 et le second à la protéine Bcd1p(120-303)Δ220-225 libre en excès. **(C)** Photographie d'une goutte contenant des cristaux du complexe Rtt106p(65-301):Bcd1p(120-303)Δ220-225. **(D)** Dépôt sur SDS-PAGE 12,5% de cristaux du complexe Rtt106p(65-301):Bcd1p(120-303)Δ220-225 (isolés de la goutte de cristallisation et lavés). Le dépôt complexe en solution correspond au complexe Rtt106p(65-301):Bcd1p(120-303).

cette chaperonne d'histone. La question se pose donc quant au fait que les deux interactions Rtt106p:histones et Rtt106p:Bcd1p soient compatibles ou exclusives. En effet, les travaux présentés précédemment montrent que la protéine Bcd1 doit interagir sur le domaine PH1 de la protéine Rtt106, or le site d'interaction principal de l'histone H3 par la lysine K56 acétylée semble se situer sur le domaine PH2 (Figures 30 et 31) (Su *et al.*, 2012), même si le domaine PH1 semble tout de même impliqué dans l'interaction avec H3 (Zunder *et al.*, 2012). Pour cela, un test préliminaire d'interaction a été réalisé par ITC avec les protéines Rtt106(65-301), Bcd1(120-303) et un peptide de l'histone H3 acétylé sur la lysine 56 (H3K56ac). Ce peptide (IRRFQKacSTELL) a déjà été utilisé par l'équipe de G. Mer (Mayo Clinic, Rochester, Minnesota) qui a déterminé que ce peptide était suffisant pour une interaction avec la protéine Rtt106 (Su *et al.*, 2012).

Deux expériences ITC ont été réalisées entre le peptide H3K56ac et la protéine Rtt106(65-301) (Figure 78B) ou le complexe déjà formé entre les protéines Rtt106(65-301) et Bcd1(120-303) (Figure 78C). Ces deux expériences sont comparées avec une expérience témoin (Figure 78A) réalisée entre les protéines Rtt106(65-301) et Bcd1(120-303). Les résultats de ces expériences montrent que l'affinité entre le peptide H3K56ac et Rtt106p(65-301) est de 714 μM , ce qui est du même ordre que l'affinité entre le peptide et le complexe Rtt106p(65-301):Bcd1p(120-303), qui est de 649 μM . Nous pouvons noter qu'il s'agit d'une valeur d'affinité plutôt faible, en accord avec les résultats obtenus par l'équipe de G. Mer (Su *et al.*, 2012). Le fait que ces deux expériences donnent des résultats de Kd similaires tend à démontrer qu'il y a bien interaction entre le peptide et Rtt106, et que cette interaction est toujours possible dans le cadre du complexe Rtt106p:Bcd1p déjà formé. De plus, si la protéine Bcd1 se dissociait de Rtt106p lors de la fixation du peptide, une absorption de chaleur du même ordre que le dégagement observé dans l'expérience témoin devrait être visible, ce qui n'est pas le cas. En effet, si l'on compare les paramètres thermodynamiques calculés lors de ces expériences (Figure 78D), nous constatons que les valeurs des paramètres thermodynamiques sont similaires entre les deux expériences réalisées avec le peptide (histogrammes bleu et orange). Les valeurs des paramètres thermodynamiques de l'expérience témoin (histogramme violet) sont quant à elles, environ deux fois supérieures aux deux autres. Si la protéine Bcd1 s'était dissociée de Rtt106p, les valeurs des paramètres thermodynamiques de l'expérience avec le peptide et le complexe devraient être égales à la différence de celles entre l'expérience réalisée avec le peptide et Rtt106p et l'expérience témoin, ce qui n'est pas le cas. Ce test ITC montre donc qu'il y a bien formation d'un

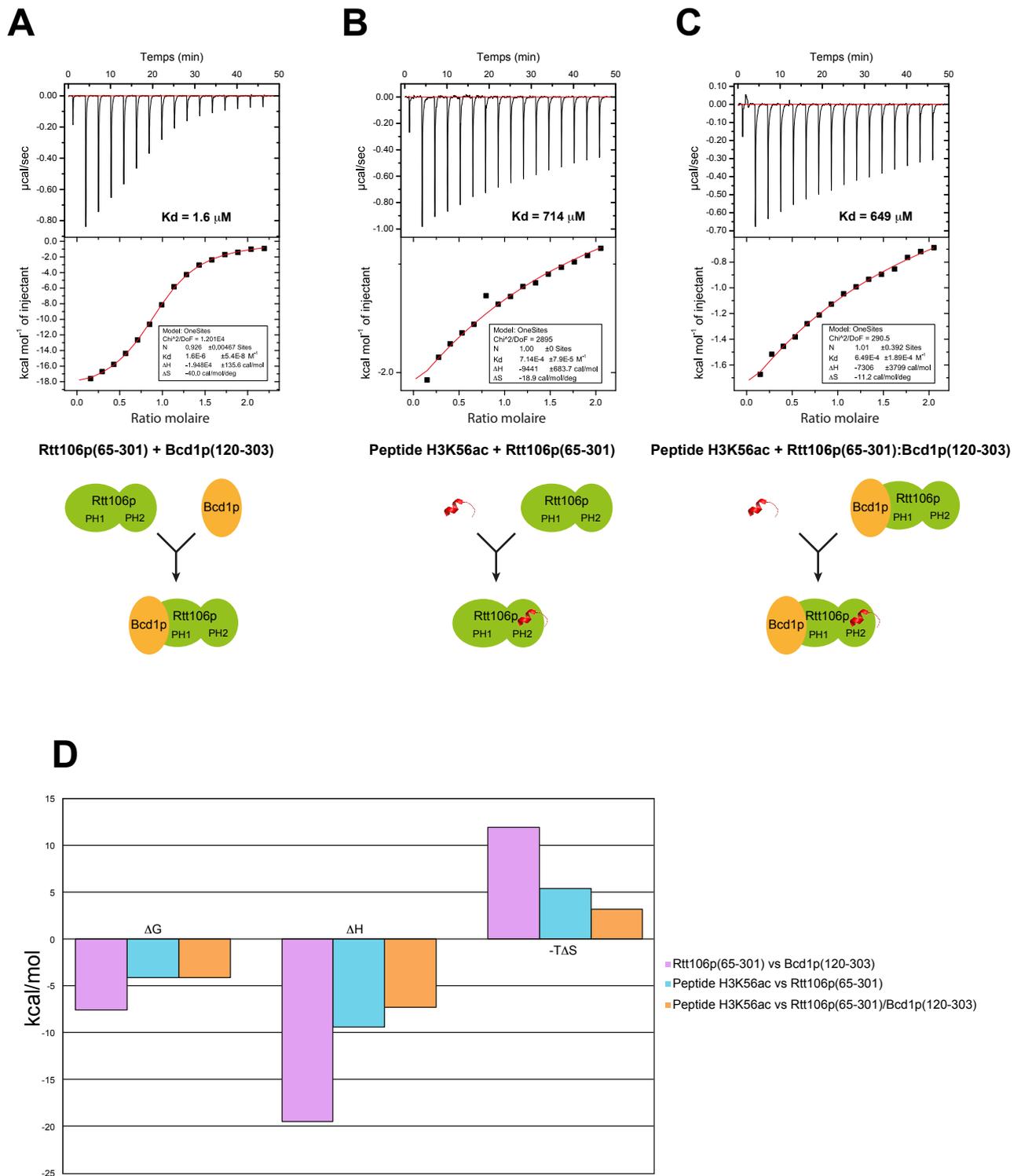


Figure 78. Possibilité d'un complexe à trois partenaires entre Bcd1p, Rtt106p et un peptide de l'histone H3. Résultats d'ITC entre (A) les protéines Rtt106(65-301) et Bcd1(120-303), (B) le peptide H3K56ac et la protéine Rtt106(65-301) et (C) le peptide H3K56ac et le complexe Rtt106p(65-301):Bcd1p(120-303). Dans les expériences (B) et (C) le peptide est placé dans la seringue à une concentration de 2 mM et le partenaire testé est placé dans la cuve à une concentration de 200 µM. Les schémas sous les résultats d'ITC représentent les complexes formés lors des différentes expériences. (D) Paramètres thermodynamiques enregistrés lors des trois expériences ITC (A), (B) et (C).

complexe entre les protéine Rtt106(65-301), Bcd1(120-303) et le peptide H3K56ac *in vitro*. Cela dit, nous avons travaillé ici avec un peptide de 11 résidus qui est loin de représenter l'encombrement stérique de l'histone H3 qui contient 136 acides aminés et se retrouve généralement au sein d'une structure tétramérique (H3-H4) ou plus grosse encore dans le cadre des nucléosomes (ADN enroulé autour d'un octamère d'histones). Des travaux complémentaires doivent donc encore être menés avec les histones entières ou des nucléosomes, afin de déterminer si l'interaction de Bcd1p avec ces entités est mutuellement exclusive ou compatible.

Conclusions

&

Perspectives

Conclusions et Perspectives

1. Le domaine minimum fonctionnel de la protéine Bcd1

La protéine Bcd1 est connue pour être essentielle à la viabilité cellulaire chez la levure. Nous avons montré que seule sa région N-terminale (résidus 1 à 96), comprenant le domaine en doigt à zinc (ZHD), était suffisante pour assurer sa fonction essentielle, puisque ce fragment permet de restaurer la croissance d'une souche de levure *tetO₇::BCD1* cultivée en présence de doxycycline, ainsi que le taux de snoRNA à boîtes C/D. Nous avons également montré que la protéine Bcd1 était capable d'interagir avec une faible affinité avec l'ARN *in vitro* et que cette association n'était pas dépendante de la séquence de l'ARN mais de sa taille qui doit être supérieure à ~100 nt. Toutefois, ce domaine N-terminal 1-96 ne semble pas être le moteur de cette interaction, ou n'est pas suffisant et cette interaction ARN-protéine ne semble pas faire intervenir le ZHD. En effet, nous avons testé l'interaction *in vitro* par EMSA entre l'ARN U14 et les fragments de Bcd1p 1-96 et 113-366, couvrant la quasi totalité de la séquence, sans obtenir de résultat positif. Il est en effet possible que l'absence de repliement observée pour la région 46-96 du fragment 1-96 par RMN puisse être la cause de cette absence d'interaction. Il est également possible que cette interaction nécessite la présence de la protéine entière, avec des déterminants d'interaction éparpillés sur toute la séquence. Il est toutefois à noter que l'interaction entre Bcd1p et les ARN, qui paraît aspécifique, n'est peut être pas physiologique. Puisque le fragment 1-96 suffit à compléter l'absence de Bcd1p entière, l'interaction de celle-ci avec les ARN ne semble pas essentielle, même si nous ne pouvons pas exclure qu'un autre facteur puisse aider cette interaction *in vivo*, par stabilisation de la région 46-96 de Bcd1p par exemple. Lors de l'analyse par spectrométrie de masse du fragment Bcd1p(1-96), un autre fragment plus court c'est dégagé, le fragment 1-81. Il serait intéressant de voir si ce fragment conserve le même effet sur la restauration de la fonction essentielle de Bcd1p, et si le fait d'être plus court le rend plus stable *in vitro*. Le fragment Bcd1p(1-96), suffisant pour assurer la fonction essentielle de la protéine, est déstructurée entre les résidus 46 et 96. La région 1-46 comprend le ZHD qui est structuré mais non-essentiel dans le cadre de la protéine entière. Il est donc possible que la fonction essentielle de Bcd1p nécessite la présence d'un partenaire pour être effective, et que la région non structurée représente le site d'interaction de ce partenaire.

2. Interaction ZHD-RuvB1/2

Nous avons montré par des expériences de titrage suivies par RMN que les ZHD des protéines Bcd1, Hit1, ZNHIT6 et ZNHIT2 étaient capables d'interagir *in vitro* avec le complexe formé par les protéines RuvBL1:2 humaines, et plus particulièrement avec RuvBL2. Les ZHD et les protéines Rvb1/2 présentent une forte conservation de séquence entre les organismes eucaryotes, au point que les ZHD de levure interagissent également avec les protéines RuvBL1/2 humaines. Ces résultats portent à croire que l'interaction entre ZHD et les protéines de la famille Rvb remonte à un ancêtre commun aux eucaryotes. Si les ZHD présentent une grande similarité de séquence entre eux, ce n'est pas le cas du reste de la séquence des protéines auxquelles ils appartiennent. Il semble donc que seul le ZHD ait été conservé dans l'évolution tandis que le reste des séquences était modifié. De plus, d'avantage de protéines contiennent des ZHD chez les eucaryotes dits les plus évolués, avec par exemple 3 représentants chez la levure *Saccharomyces cerevisiae* pour 6 chez l'Homme. Au cours de l'évolution et avec l'apparition du noyau, les protéines Rvb semblent avoir pris une importance croissante dans le métabolisme cellulaire, puisqu'elles sont retrouvées dans de nombreux complexes nucléaires. Si les ZHD sont des partenaires directs des protéines Rvb, il est donc possible que les gènes de ceux-ci aient pu être dupliqués, de manière à constituer un panel d'interactants des Rvb leur permettant d'assurer leurs différentes activités cellulaires.

Nous avons également montré que le ZHD de Bcd1p n'était pas nécessaire dans des conditions de croissance optimales, mais que la déplétion de Bcd1p avait un impact négatif sur la croissance des levures et le taux de snoRNA à boîtes C/D en conditions de stress thermique. Les protéines Rvb sont connues pour être des composants du complexe R2TP, qui est lui-même un co-chaperon de la chaperonne Hsp82. Il est donc probable que le ZHD de Bcd1p permette de recruter la chaperonne Hsp82, par le biais du complexe R2TP, pour améliorer le repliement et la stabilité des facteurs impliqués dans la biogenèse des snoRNP à boîtes C/D qui sont affectés par le stress thermique.

Si nous avons démontré l'existence d'interactions entre les ZHD et les protéines Rvb1/2 par RMN, ces résultats doivent encore être validés par une autre méthode, ce qui sera réalisé prochainement par nos collaborateurs de l'IBET (équipe de P. Matias) par la technique de Blitz. L'objectif serait par la suite de déterminer la structure cristallographique d'un complexe entre le ZHD de ZNHIT6 et RuvBL2, ou éventuellement avec un fragment plus long de ZNHIT6 correspondant au fragment 1-81 de levure, ce qui nous donnerait des informations structurales précises sur l'interface. De plus, la structure des protéines

RuvBL1/2 n'a actuellement jamais pu être déterminée en complexe avec un partenaire à l'échelle atomique. Cela pourrait nous apporter des renseignements sur le fonctionnement de ces protéines, encore sujet à de nombreux questionnement. Qui plus est, la présence d'un partenaire en interaction pourrait modifier les symétries cristallines obtenues jusqu'à présent. Ceci pourrait permettre de discriminer les symétries réelles des symétries cristallines. De tels résultats aideraient à déterminer l'agencement réel des protéines RuvBL1/2 entre elles, et principalement l'interaction entre hexamères dans le cadre d'un complexe dodécamérique.

Sachant que la protéine Bcd1 de levure interagit avec la protéine Rvb2, nous avons voulu déterminer le domaine d'interaction présent au niveau de Rvb2p. Pour cela, nous avons tenté d'identifier des domaines périphériques de Rvb2p en nous basant sur les structures connues des protéines RuvBL1/2. Nous avons ainsi produit de façon recombinante deux fragments correspondant aux domaines DII_{ext} et DIII. Les tests d'interaction avec Bcd1p que nous avons réalisé sur ces deux domaines, que ce soit par *His-Pull down* ou par RMN, se sont tous révélés négatifs. Même si le domaine DII_{ext} nous paraissait être un bon candidat pour des interactions protéine-protéine, ce n'est manifestement pas le cas avec les Rvb. En ce qui concerne le domaine DIII, il est possible que son expression sans le domaine DI, avec lequel il effectue de nombreux contacts, ait eu un impact négatif sur son repliement et sur sa stabilité. Nous avons en effet constaté que la protéine s'agrège rapidement lors des analyses par RMN. Nous n'avons donc jusque-là pas été en mesure de déterminer quel domaine de Rvb2p représente le site d'interaction avec Bcd1p. D'autres analyses plus poussées sont donc nécessaires pour caractériser plus précisément cette interaction. Il est en effet envisageable que la structure en hexamère de la protéine soit un facteur essentiel dans la reconnaissance de la protéine Bcd1, ou que les domaines utilisés pour les analyses d'interaction n'aient pas été optimaux.

3. Le complexe Nop58p-Rvb1p-Rvb2p-Bcd1p

Les tests de co-expression qui visaient à reproduire le complexe formé par les protéines Nop58, Rvb1, Rvb2 et Bcd1 ne nous ont malheureusement pas permis de mettre en évidence des interactions évidentes entre ces différentes protéines. Cette technique ne semble finalement pas être adaptée à l'étude de ces protéines, dont plusieurs réalisent des contacts aspécifiques avec la résine de sépharose utilisée pour retenir l'étiquette 6xHistidines. Nous avons également des doutes sur le repliement des protéines Rvb1/2 dans ces conditions de

production. Ces protéines sont connues pour être difficiles à produire et à purifier. Or les expériences de co-expression peuvent se résumer à un début de purification classique, dont le protocole n'est pas optimisé pour l'étude des protéines Rvb1/2. En effet, la production de ces protéines dans leur bonne conformation est souvent conditionnée par les conditions d'expression et surtout de tampon lors de la purification. La présence d'ADP ou d'ATP pourrait notamment s'avérer être un paramètre très important. La production des protéines Rvb1/2 n'est pas le seul obstacle à l'étude de ce complexe, puisque nous n'avons pas été en mesure de travailler avec la protéine Nop58 de pleine taille, mais uniquement avec le fragment 1-447. Les résidus C-terminaux manquant (448-511) de Nop58p pourraient s'avérer essentiels pour les contacts avec ses partenaires, dont les protéines Rvb. L'étude du complexe Nop58p-Rvb1p-Rvb2p-Bcd1p *in vitro* devra donc passer par des étapes préalables d'optimisation des conditions de production/purification des protéines Rvb1/2 et Nop58 entière, en passant éventuellement par une expression en système eucaryote (cellules d'insecte) par l'intermédiaire du baculovirus. Cette expérience de co-expression nous a toutefois fourni une piste quant à une éventuelle interaction entre Bcd1p et Nop58p(1-447). Des analyses par d'autres techniques devront être réalisées afin de vérifier si une interaction entre les deux protéines est possible ou non *in vitro*.

4. Structure 3D en solution de la protéine Bcd1(120-303)

Des approches de protéolyse ménagée suivies d'analyses par spectrométrie de masse, ainsi que l'analyse de spectres HSQC nous ont permis de délimiter un domaine de la région C-terminale de Bcd1p, le fragment 120-303. Ce dernier est stable, replié et permet de conserver l'interaction avec le partenaire Rtt106p. La structure tridimensionnelle en solution de cette protéine ne semble pas avoir d'homologue connu. Elle présente un feuillet β central vrillé composé de 5 brins β contre une face duquel viennent se plaquer deux hélices α . La région 120-148 est probablement indépendante du reste du domaine, puisque d'autres fragments étudiés débutent à la borne 149 et apparaissent repliés en RMN. Cette petite région est constituée d'un feuillet de deux brins β antiparallèles et d'une petite hélice α , et se situe sous le feuillet central. Cette région semble tout de même très importante, puisque sans elle l'interaction avec la protéine Rtt106 est impossible.

La structure de ce domaine par RMN est actuellement encore en cours de raffinement. Quelques erreurs d'attribution ont depuis été corrigées et de nouveaux noyaux ont été attribués. Les statistiques de la structure sont donc en cours d'amélioration, avec notamment

une baisse du nombre de contraintes NOE sans attribution (initialement de 677) et une augmentation du nombre de NOE (désormais autour de 11 000). Toutes ces améliorations ont déjà eu un effet sur la qualité de la structure, ainsi que sur les valeurs de RMSD. L'obtention de la structure définitive nécessite encore des calculs dans un programme (CNS, AMBER...) prenant en compte la présence des molécules d'H₂O.

5. Interaction entre les protéines Bcd1 et Rtt106

Une grande partie de mon travail de thèse a porté sur l'étude de l'interaction entre les protéines Bcd1 et Rtt106. Après avoir validé l'aspect direct de cette interaction par co-expression analysée par spectrométrie de masse et en avoir déterminé les paramètres thermodynamiques et d'affinité par ITC, les fragments minimums permettant de la conserver ont été définis. Avec des approches croisées de protéolyse ménagée suivies d'analyses par spectrométrie de masse, d'ITC et de RMN, nous avons identifié les fragments stables, repliés et plus courts possibles n'altérant pas l'interaction. Ces fragments Bcd1p(120-303) et Rtt106p(65-301) ont également fait l'objet d'analyses par SAXS. Cette approche mise en œuvre par A. Thureau au synchrotron SOLEIL a montré que ce complexe n'était pas présent à 100% en solution. Les données de SAXS n'ont donc pas pu être utilisées pour définir une enveloppe de surface du complexe. Afin de vérifier si les vitesses d'association/dissociation (k_{on}/k_{off}) de ce complexe empêchent effectivement d'obtenir en solution un complexe formé à près de 100%, une analyse par SPR (Résonance Plasmonique de Surface) devra être effectuée. Cette analyse pourrait également nous permettre de tester et d'identifier des conditions de tampon permettant de limiter la vitesse de dissociation du complexe et donc d'obtenir une plus grande proportion de complexe en solution. De plus, une nouvelle expérience de SAXS pourrait être menée dans des conditions où l'un des deux partenaires est à une concentration largement supérieure (x10) au K_d, de façon à déplacer l'équilibre en faveur de la formation du complexe lors de l'analyse et ainsi obtenir des données exploitables. Une autre possibilité serait d'étudier ce complexe chez d'autres espèces de levures, comme la levure thermophile, où il pourrait posséder des caractéristiques différentes comme une plus grande stabilité. A terme, l'obtention d'une enveloppe SAXS du complexe, associée à la connaissance des structures 3D des deux partenaires libres et aux données de spectrométrie de masse sur le complexe (échange hydrogène/deutérium, mobilité ionique, *cross-link*) et de RMN (titrages), pourrait nous permettre de générer un modèle de l'interaction entre ces deux protéines.

La région 120-148 revêt un intérêt particulier : elle est essentielle à la formation du complexe *in vitro*, puisque sans elle aucune interaction n'est visible. Cependant, une fois le complexe formé et soumis à une digestion trypsique, les analyses par spectrométrie de masse montrent que cette région peut être supprimée sans que le complexe ne soit affecté. La formation du complexe doit donc requérir la présence de déterminants spécifiques présents dans cette région 120-148, lesquels ne sont apparemment plus nécessaires au maintien de la forme complexée entre les deux partenaires.

6. Vers une structure par diffraction des rayons X du complexe

Rtt106p/Bcd1p

Les données issues des expériences de dichroïsme circulaire et de spectrométrie de masse pourraient toutefois aller dans le sens d'une réorganisation structurale lors de la formation du complexe. Dans ce cas, la seule alternative serait l'obtention d'une structure par diffraction des rayons X.

Les études menées par RMN sur la protéine Bcd1(120-303), et principalement les NOE hétéronucléaires, nous ont montré que deux régions (boucles) de celle-ci sont très mobiles. Nous avons alors choisi de supprimer certains résidus de ces boucles afin de limiter la flexibilité de la protéine. Si le mutant de l'une des deux boucles (boucle 172-177) n'est pas exprimé, l'autre (boucle 220-225) est quant à lui exprimé, soluble et stable dans le temps. Les derniers essais de cristallisation menés par C. Charron au laboratoire sur le complexe Rtt106p(65-301):Bcd1p(120-303) Δ 220-225 où le fragment de Bcd1p est muté au niveau de la deuxième boucle flexible ont permis d'obtenir des cristaux protéiques contenant les deux partenaires. Il semble donc que la rigidification au niveau de cette boucle 220-225 ait permis la cristallisation. Les cristaux de ce complexe doivent encore être améliorés avant toute analyse de diffraction aux rayons X.

Nous n'avons, lors de ces travaux, travaillé qu'avec le double domaine PH de Rtt106p. Mais il est connu que la protéine Rtt106 possède dans sa région N-terminale un domaine de dimérisation (Su *et al.*, 2012). Il pourrait être intéressant de tester si l'interaction Bcd1p:Rtt106 ne serait pas impactée par la présence de ce domaine N-terminal de Rtt106p induisant sa dimérisation.

7. Test de compétition entre Bcd1p et les histones pour l'interaction avec Rtt106p

La protéine Rtt106 est une chaperonne d'histone connue pour interagir avec l'histone H3, et avec une affinité accrue pour l'histone H3 acétylée sur la lysine 56 (H3K56ac), par le biais de son domaine en double PH. L'interaction de Rtt106p avec l'histone H3K56ac est réalisée par le domaine PH2 principalement (Su *et al.*, 2012), mais il a également été montré que le domaine PH1 était impliqué dans l'interaction (Zunder *et al.*, 2012). Nous nous sommes donc demandés si les interactions Bcd1p:Rtt106 et Rtt106:H3 étaient mutuellement exclusives ou non. Afin d'apporter un début de réponse à cette question, nous avons réalisé des expériences d'ITC avec le complexe Rtt106p(65-301) et un peptide de l'histone H3 de 11 acides aminés contenant la lysine K56 acétylée, dont le site de fixation se situe sur le domaine PH2 de Rtt106p (Su *et al.*, 2012). Ce premier test nous indique que la protéine Bcd1(120-303) et le peptide de H3K56ac peuvent interagir simultanément avec la protéine Rtt106(65-301), ce qui paraît cohérent avec les résultats de RMN suggérant que Bcd1p interagit avec le domaine PH1 de Rtt106p. Nous n'avons ici testé qu'un peptide de l'histone H3, ce qui est loin de représenter la réalité de la protéine entière, laquelle est généralement en complexe avec d'autres histones pour former des hétérotétramères, voire même au sein d'un nucléosome. Pour pouvoir continuer les investigations sur ce sujet, nous avons entamé une collaboration avec l'équipe de G. Mer (Mayo Clinic, Rochester, USA), ayant l'expertise des histones et de la protéine Rtt106. Leurs premières études sur ce sujet consisteront à observer l'effet de l'ajout de la protéine Bcd1 sur le complexe nucléosome-Rtt106p.

La protéine humaine TAF9 (facteur d'initiation de la transcription TFIID, sous-unité 9) humaine a été identifiée comme étant un facteur d'assemblage potentiel des snoRNP à boîtes C/D, interagissant potentiellement avec la protéine BCD1 (McKeegan *et al.*, 2007). De façon intéressante cette protéine, ainsi que son orthologue chez la levure, la protéine Taf9 (ou encore Taf_{II}17), possèdent un domaine *histone fold* dont le repliement est semblable à celui des histones (Gangloff *et al.*, 2001). La protéine Taf9 pourrait donc être un bon candidat pour tester des interactions avec les protéines Bcd1, puisque leurs orthologues humaines semblent interagir entre elles, et Rtt106. Un tel lien entre ces protéines pourrait représenter un couplage entre la biogenèse des snoRNP à boîtes C/D et la transcription, notamment des gènes codant les snoRNA à boîtes C/D.

Même si elle ouvre de grandes perspectives, notamment par rapport à son importance vis-à-vis des nucléosomes, l'interaction entre les protéines Rtt106 et Bcd1 n'est pas à l'origine du caractère essentiel de Bcd1p. En effet, le fragment Bcd1p(1-96) est suffisant à restaurer la croissance suite à la répression de l'expression de Bcd1p endogène. Or aucun acide aminé de ce fragment n'est représenté dans le fragment Bcd1p(120-303) permettant l'interaction avec Rtt106.

8. La protéine BCD1 humaine

Nos tentatives d'études *in vitro* de la protéine humaine BCD1 (ou ZNHIT6) se sont pour le moment avérées infructueuses. En nous basant sur des données de la littérature (McKeegan *et al.*, 2007), nous avons tenté de produire de façon recombinante en système bactérien le fragment hBCD1(1-360) mais nous n'avons pu obtenir que de très faibles quantités (moins d'1 mg par litre de culture). Les futurs travaux sur cette protéine nécessiteront donc encore de l'optimisation des conditions d'expression, tel que la séquence des fragments exprimés ou une expression dans un système eucaryote (cellules d'insecte). Si malgré cela la protéine humaine reste difficile à étudier, il restera la possibilité d'étudier la protéine BCD1 d'autres mammifères en transposant les résultats obtenus à l'Homme.

Références

Bibliographiques

Références bibliographiques

- A**ittaleb M, Rashid R, Chen Q, Palmer JR, Daniels CJ, Li H. Structure and function of archaeal box C/D sRNP core proteins. *Nature Structural Biology*, 2003, **10** : 256-263.
- Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *The EMBO Journal*, 1999, **18** : 5399-5410.
- Amin AD, Vishnoi N, Prochasson P. A global requirement for the HIR complex in the assembly of chromatin. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 2012, **1819** : 264-276.
- B**achellerie J-P, Cavallé J, Hüttenhofer A. The expanding snoRNA world. *Biochimie*, 2002, **84** : 775-790.
- Back, R., Dominguez, C., Rothé, B., Bobo, C., Beaufils, C., Moréra, S., Meyer, P., Charpentier, B., Branlant, C., Allain, F.H.-T., Manival, X. High-Resolution Structural Analysis Shows How Tah1 Tethers Hsp90 to the R2TP Complex. *Structure*, 2013, **21** : 1834-1847.
- Barraud, P., Schubert, M., Allain, F.H.-T. A strong ¹³C chemical shift signature provides the coordination mode of histidines in zinc-binding proteins. *Journal of Biomolecular NMR*, 2012, **53** : 93-101.
- Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., Yusupov, M. The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science (New York, N.Y.)*, 2011, **334** : 1524-1529.
- Bizarro, J., Charron, C., Boulon, S., Westman, B., Pradet-Balade, B., Vandermoere, F., Chagot, M.-E., Hallais, M., Ahmad, Y., Leonhardt, H., Lamond, A., Manival, X., Branlant, C., Charpentier, B., Verheggen, C., Bertrand, E. Proteomic and 3D structure analyses highlight the C/D box snoRNP assembly mechanism and its control. *The Journal of Cell Biology*, 2014, **207** : 463-480.
- Bizarro, J., Dodré, M., Huttin, A., Charpentier, B., Schlotter, F., Branlant, C., Verheggen, C., Massenet, S., Bertrand, E. NUFIP and the HSP90/R2TP chaperone bind the SMN complex and facilitate assembly of U4-specific proteins. *Nucleic Acids Research*, 2015, **43** : 8973-8989.
- Bleichert, F., Gagnon, K.T., Brown, B.A., Maxwell, E.S., Leschziner, A.E., Unger, V.M., Baserga, S.J. A Dimeric Structure for Archaeal Box C/D Small Ribonucleoproteins. *Science*, 2009, **325** : 1384-1387.

- Boulon, S., Marmier-Gourrier, N., Pradet-Balade, B., Wurth, L., Verheggen, C., Jády, B.E., Rothé, B., Pescia, C., Robert, M.-C., Kiss, T., Bardoni, B., Krol, A., Branlant, C., Allmang, C., Bertrand, E., Charpentier, B. The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *The Journal of Cell Biology*, 2008, **180** : 579-595.
- Boulon, S., Pradet-Balade, B., Verheggen, C., Molle, D., Boireau, S., Georgieva, M., Azzag, K., Robert, M.-C., Ahmad, Y., Neel, H., Lamond, A.I., Bertrand, E. HSP90 and Its R2TP/Prefoldin-like Cochaperone Are Involved in the Cytoplasmic Assembly of RNA Polymerase II. *Molecular Cell*, 2010, **39** : 912-924.
- Bower-Phipps, K.R., Taylor, D.W., Wang, H.-W., Baserga, S.J. The box C/D sRNP dimeric architecture is conserved across domain Archaea. *RNA*, 2012, **18** : 1527-1540.
- Brown, J.W.S., Echeverria, M., Qu, L.-H. Plant snoRNAs: functional evolution and new modes of gene expression. *Trends in Plant Science*, 2003, **8** : 42-49.
- Brown, J.W.S., Marshall, D.F., Echeverria, M. Intronic noncoding RNAs and splicing. *Trends in Plant Science*, 2008, **13** : 335-342.
- C**availlé, J., Nicoloso, M., Bachellerie, J.-P. Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature*, 1996, **383** : 732-735.
- Chanfreau, G., Rotondo, G., Legrain, P., Jacquier, A. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *The EMBO Journal*, 1998, **17** : 3726-3737.
- Charron, C., Manival, X., Cléry, A., Senty-Ségault, V., Charpentier, B., Marmier-Gourrier, N., Branlant, C., Aubry, A. The Archaeal sRNA Binding Protein L7Ae has a 3D Structure Very Similar to that of its Eukaryal Counterpart While Having a Broader RNA-binding Specificity. *Journal of Molecular Biology*, 2004, **342** : 757-773.
- Chen, C.-L., Chen, C.-J., Vallon, O., Huang, Z.-P., Zhou, H., Qu, L.-H. Genomewide Analysis of Box C/D and Box H/ACA snoRNAs in *Chlamydomonas reinhardtii* Reveals an Extensive Organization Into Intronic Gene Clusters. *Genetics*, 2008, **179** : 21-30.
- Chen, P., Zhu, P., Li, G. New insights into the helical structure of 30-nm chromatin fibers. *Protein & Cell*, 2014, **5** : 489-491.
- Cheung, K.L.Y., Huen, J., Houry, W.A., Ortega, J. Comparison of the multiple oligomeric structures observed for the Rvb1 and Rvb2 proteins This paper is one of a selection of papers published in this special issue entitled 8th International Conference on AAA Proteins and has undergone the Journal's usual peer review process. *Biochemistry and Cell Biology*, 2010, **88** : 77-88.
- Colau, G., Thiry, M., Leduc, V., Bordonné, R., Lafontaine, D.L.J. The Small Nucle(ol)ar RNA Cap Trimethyltransferase Is Required for Ribosome Synthesis and Intact Nucleolar Morphology. *Molecular and Cellular Biology*, 2004, **24** : 7976-7986.
- Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, 1988, **16** : 10881-10890.

- D**ahlin, J.L., Chen, X., Walters, M.A., Zhang, Z. Histone-modifying enzymes, histone modifications and histone chaperones in nucleosome assembly: Lessons learned from Rtt109 histone acetyltransferases. *Critical Reviews In Biochemistry And Molecular Biology*, 2015, **50** : 31-53.
- Darzacq, X., Jády, B.E., Verheggen, C., Kiss, A.M., Bertrand, E., Kiss, T. Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *The EMBO Journal*, 2002, **21** : 2746-2756.
- Decatur, W.A., Fournier, M.J. rRNA modifications and ribosome function. *Trends in Biochemical Sciences*, 2002, **27** : 344-351.
- Dennis, P.P., Omer, A., Lowe, T. A guided tour: small RNA function in Archaea. *Molecular Microbiology*, 2001, **40** : 509-519.
- Diebold, M.-L., Fribourg, S., Koch, M., Metzger, T., Romier, C. Deciphering correct strategies for multiprotein complex assembly by co-expression: Application to complexes as large as the histone octamer. *Journal of Structural Biology*, 2011, **175** : 178-188.
- Dinman, J.D. Pathways to Specialized Ribosomes: The Brussels Lecture. *Journal of Molecular Biology*, 2016, **428** : 2186-2194.
- Dupuis-Sandoval, F., Poirier, M., Scott, M.S. The emerging landscape of small nucleolar RNAs in cell biology. *Wiley Interdisciplinary Reviews: RNA*, 2015, **6** : 381-397.
- E**ckert, K., Saliou, J.-M., Monlezun, L., Vigouroux, A., Atmane, N., Caillat, C., Quevillon-Chérueil, S., Madiona, K., Nicaise, M., Lazereg, S., Dorselaer, A.V., Sanglier-Cianférani, S., Meyer, P., Moréra, S. The Pih1-Tah1 Cochaperone Complex Inhibits Hsp90 Molecular Chaperone ATPase Activity. *Journal of Biological Chemistry*, 2010, **285** : 31304-31312.
- Eliceiri, G.L. Small nucleolar RNAs. *Cellular and Molecular Life Sciences CMLS*, 1999, **56** : 22-31.
- Ettema, T.J., Huynen, M.A., de Vos, W.M., van der Oost, J. TRASH: a novel metal-binding domain predicted to be involved in heavy-metal sensing, trafficking and resistance. *Trends in biochemical sciences*, 2003, **28** : 170-173.
- Ewens, C.A., Su, M., Zhao, L., Nano, N., Houry, W.A., Southworth, D.R. Architecture and Nucleotide-Dependent Conformational Changes of the Rvb1-Rvb2 AAA+ Complex Revealed by Cryoelectron Microscopy. *Structure*, 2016, **24** : 657-666.
- F**alaleeva, M., Pages, A., Matuszek, Z., Hidmi, S., Agranat-Tamir, L., Korotkov, K., Nevo, Y., Eyras, E., Sperling, R., Stamm, S. Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. *Proceedings of the National Academy of Sciences*, 2016, **113** : E1625-E1634.

Falb, M., Amata, I., Gabel, F., Simon, B., Carlomagno, T. Structure of the K-turn U4 RNA: a combined NMR and SANS study. *Nucleic Acids Research*, 2010, **38** : 6274-6285.

Fazly, A., Li, Q., Hu, Q., Mer, G., Horazdovsky, B., Zhang, Z. Histone Chaperone Rtt106 Promotes Nucleosome Formation Using (H3-H4)₂ Tetramers. *Journal of Biological Chemistry*, 2012, **287** : 10753-10760.

Feder, M., Pas, J., Wyrwicz, L.S., Bujnicki, J.M. Molecular phylogenetics of the RrmJ/fibrillarlin superfamily of ribose 2'-O-methyltransferases. *Gene*, 2003, **302** : 129-138.

Galardi, S., Fatica, A., Bachi, A., Scaloni, A., Presutti, C., Bozzoni, I. Purified Box C/D snoRNPs Are Able To Reproduce Site-Specific 2'-O-Methylation of Target RNA In Vitro. *Molecular and Cellular Biology*, 2002, **22** : 6663-6668.

Gautier, T., Bergès, T., Tollervey, D., Hurt, E. Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Molecular and Cellular Biology*, 1997, **17** : 7088-7098.

Goddard, T.D., Kneller, D.G. *SPARKY 3.114*, University of California, San Francisco.

Gonzales, F.A., Zanchin, N.I.T., Luz, J.S., Oliveira, C.C. Characterization of *Saccharomyces cerevisiae* Nop17p, a Novel Nop58p-Interacting Protein that is Involved in Pre-rRNA Processing. *Journal of Molecular Biology*, 2005, **346** : 437-455.

Gorynia, S., Bandejas, T.M., Pinho, F.G., McVey, C.E., Vornrhein, C., Round, A., Svergun, D.I., Donner, P., Matias, P.M., Carrondo, M.A. Structural and functional insights into a dodecameric molecular machine – The RuvBL1/RuvBL2 complex. *Journal of Structural Biology*, 2011, **176** : 279-291.

Gribun, A., Cheung, K.L.Y., Huen, J., Ortega, J., Houry, W.A. Yeast Rvb1 and Rvb2 are ATP-Dependent DNA Helicases that Form a Heterohexameric Complex. *Journal of Molecular Biology*, 2008, **376** : 1320-1333.

Guffanti, E., Ferrari, R., Preti, M., Forloni, M., Harismendy, O., Lefebvre, O., Dieci, G. A Minimal Promoter for TFIIC-dependent in Vitro Transcription of snoRNA and tRNA Genes by RNA Polymerase III. *Journal of Biological Chemistry*, 2006, **281** : 23945-23957.

Hager, J., Staker, B.L., Bügl, H., Jakob, U. Active Site in RrmJ, a Heat Shock-induced Methyltransferase. *Journal of Biological Chemistry*, 2002, **277** : 41978-41986.

Hamma, T., Ferré-D'Amaré, A.R. Structure of Protein L7Ae Bound to a K-Turn Derived from an Archaeal Box H/ACA sRNA at 1.8 Å Resolution. *Structure*, 2004, **12** : 893-903.

Henras, A.K., Dez, C., Henry, Y. RNA structure and function in C/D and H/ACA s(no)RNPs. *Current Opinion in Structural Biology*, 2004, **14** : 335-343.

- Hiley, S.L. Global analysis of yeast RNA processing identifies new targets of RNase III and uncovers a link between tRNA 5' end processing and tRNA splicing. *Nucleic Acids Research*, 2005, **33** : 3048-3056.
- Hirose, T., Ideue, T., Nagai, M., Hagiwara, M., Shu, M.-D., Steitz, J.A. A Spliceosomal Intron Binding Protein, IBP160, Links Position-Dependent Assembly of Intron-Encoded Box C/D snoRNP to Pre-mRNA Splicing. *Molecular Cell*, 2006, **23** : 673-684.
- Hirose, T., Shu, M.-D., Steitz, J.A. Splicing-Dependent and -Independent Modes of Assembly for Intron-Encoded Box C/D snoRNPs in Mammalian Cells. *Molecular Cell*, 2003, **12** : 113-123.
- Hirose, T., Steitz, J.A. Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proceedings of the National Academy of Sciences*, 2001, **98** : 12914-12919.
- Hořejší, Z., Stach, L., Flower, T.G., Joshi, D., Flynn, H., Skehel, J.M., O'Reilly, N.J., Ogrodowicz, R.W., Smerdon, S.J., Boulton, S.J. Phosphorylation-Dependent PIH1D1 Interactions Define Substrate Specificity of the R2TP Cochaperone Complex. *Cell Reports*, 2014, **7** : 19-26.
- Hořejší, Z., Takai, H., Adelman, C.A., Collis, S.J., Flynn, H., Maslen, S., Skehel, J.M., de Lange, T., Boulton, S.J. CK2 Phospho-Dependent Binding of R2TP Complex to TEL2 Is Essential for mTOR and SMG1 Stability. *Molecular Cell*, 2010, **39** : 839-850.
- Huang, S., Zhou, H., Katzmann, D., Hochstrasser, M., Atanasova, E., Zhang, Z. Rtt106p is a histone chaperone involved in heterochromatin-mediated silencing. *Proceedings of the National Academy of Sciences of the United States of America*, 2005b, **102** : 13410-13415.
- Huang, Z.-P., Zhou, H., He, H.-L., Chen, C.-L., Liang, D., Qu, L.-H. Genome-wide analyses of two families of snoRNA genes from *Drosophila melanogaster*, demonstrating the extensive utilization of introns for coding of snoRNAs. *RNA*, 2005a, **11** : 1303-1316.
- I**kura, T., Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., Nakatani, Y. Involvement of the TIP60 Histone Acetylase Complex in DNA Repair and Apoptosis. *Cell*, 2000, **102** : 463-473.
- Imbeault, D., Gamar, L., Rufiange, A., Paquet, E., Nourani, A. 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*, 2008, **283** : 27350-27354.
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences*, 2001, **98** : 4569-4574.
- Itsuki, Y., Saeki, M., Nakahara, H., Egusa, H., Irie, Y., Terao, Y., Kawabata, S., Yatani, H., Kamisaki, Y. Molecular cloning of novel Monad binding protein containing tetratricopeptide repeat domains. *FEBS Letters*, 2008, **582** : 2365-2370.

- Jeganathan, A., Leong, V., Zhao, L., Huen, J., Nano, N., Houry, W.A., Ortega, J. Yeast Rvb1 and Rvb2 Proteins Oligomerize As a Conformationally Variable Dodecamer with Low Frequency. *Journal of Molecular Biology*, 2015, **427** : 1875-1886.
- Jha, S., Dutta, A. RVB1/RVB2: Running Rings around Molecular Biology. *Molecular Cell*, 2009, **34** : 521-533.
- Jha, S., Shibata, E., Dutta, A. Human Rvb1/Tip49 Is Required for the Histone Acetyltransferase Activity of Tip60/NuA4 and for the Downregulation of Phosphorylation on H2AX after DNA Damage. *Molecular and Cellular Biology*, 2008, **28** : 2690-2700.
- Jiménez, B., Ugwu, F., Zhao, R., Ortí, L., Makhnevych, T., Pineda-Lucena, A., Houry, W.A. Structure of Minimal Tetratricopeptide Repeat Domain Protein Tah1 Reveals Mechanism of Its Interaction with Pih1 and Hsp90. *Journal of Biological Chemistry*, 2012, **287** : 5698-5709.
- Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, 1999, **292** : 195-202.
- Jónsson, Z.O., Jha, S., Wohlschlegel, J.A., Dutta, A. Rvb1p/Rvb2p Recruit Arp5p and Assemble a Functional Ino80 Chromatin Remodeling Complex. *Molecular Cell*, 2004, **16** : 465-477.
- Jorjani, H., Kehr, S., Jedlinski, D.J., Gumienny, R., Hertel, J., Stadler, P.F., Zavolan, M., Gruber, A.R. An updated human snoRNAome. *Nucleic Acids Research*, 2016, **44** : 5068-5082.
- K**akahara, Y., Houry, W.A. The R2TP complex: Discovery and functions. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2012, **1823** : 101-107.
- Kakahara, Y., Makhnevych, T., Zhao, L., Tang, W., Houry, W.A. Nutritional status modulates box C/D snoRNP biogenesis by regulated subcellular relocalization of the R2TP complex. *Genome Biology*, 2014, **15** : 404.
- Kawakami, K., Shafer, B.K., Garfinkel, D.J., Strathern, J.N., Nakamura, Y. Ty element-induced temperature-sensitive mutations of *Saccharomyces cerevisiae*. *Genetics*, 1992, **131** : 821-832.
- King, T.H., Decatur, W.A., Bertrand, E., Maxwell, E.S., Fournier, M.J. A Well-Connected and Conserved Nucleoplasmic Helicase Is Required for Production of Box C/D and H/ACA snoRNAs and Localization of snoRNP Proteins. *Molecular and Cellular Biology*, 2001, **21** : 7731-7746.
- Kiss, T., Filipowicz, W. Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes & Development*, 1995, **9** : 1411-1424.
- Kiss-László, Z., Henry, Y., Bachellerie, J.-P., Caizergues-Ferrer, M., Kiss, T. Site-Specific Ribose Methylation of Preribosomal RNA: A Novel Function for Small Nucleolar RNAs. *Cell*, 1996, **85** : 1077-1088.
- Kiss-László, Z., Henry, Y., Kiss, T. Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *The EMBO Journal*, 1998, **17** : 797-807.

- Kornprobst, M., Turk, M., Kellner, N., Cheng, J., Flemming, D., Koš-Braun, I., Koš, M., Thoms, M., Berninghausen, O., Beckmann, R., Hurt, E. Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome. *Cell*, 2016, **166** : 380-393.
- Kowalski, A., Pałyga, J. Modulation of chromatin function through linker histone H1 variants. *Biology of the Cell*, 2016 : n/a-n/a.
- Kressler, D., Doère, M., Rojo, M., Linder, P. Synthetic Lethality with Conditional dbp6 Alleles Identifies Rsa1p, a Nucleoplasmic Protein Involved in the Assembly of 60S Ribosomal Subunits. *Molecular and Cellular Biology*, 1999, **19** : 8633-8645.
- Lafontaine, D.L.J., Tollervey, D. Synthesis and Assembly of the Box C+D Small Nucleolar RNPs. *Molecular and Cellular Biology*, 2000, **20** : 2650-2659.
- Lafontaine, D.L.J. Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nature Structural & Molecular Biology*, 2015, **22** : 11-19.
- Lakomek, K., Stoehr, G., Tosi, A., Schmailzl, M., Hopfner, K.-P. Structural Basis for Dodecameric Assembly States and Conformational Plasticity of the Full-Length AAA+ ATPases Rvb1-Rvb2. *Structure*, 2015, **23** : 483-495.
- Lakshminarasimhan, M., Boanca, G., Banks, C.A.S., Hattem, G.L., Gabriel, A.E., Groppe, B.D., Smoyer, C., Malanowski, K.E., Peak, A., Florens, L., Washburn, M.P. Proteomic and Genomic Analyses of the Rvb1 and Rvb2 Interaction Network upon Deletion of R2TP Complex Components. *Molecular & Cellular Proteomics*, 2016, **15** : 960-974.
- Lapinaite, A., Simon, B., Skjaerven, L., Rakwalska-Bange, M., Gabel, F., Carlomagno, T. The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature*, 2013, **502** : 519-523.
- Li, J., Soroka, J., Buchner, J. The Hsp90 chaperone machinery: Conformational dynamics and regulation by co-chaperones. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2012, **1823** : 624-635.
- Li, Q., Zhou, H., Wurtele, H., Davies, B., Horazdovsky, B., Verreault, A., Zhang, Z. Acetylation of Histone H3 Lysine 56 Regulates Replication-Coupled Nucleosome Assembly. *Cell*, 2008, **134** : 244-255.
- Lin, J., Lai, S., Jia, R., Xu, A., Zhang, L., Lu, J., Ye, K. Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature*, 2011, **469** : 559-563.
- Liu, Y., Huang, H., Zhou, B.O., Wang, S.-S., Hu, Y., Li, X., Liu, J., Zang, J., Niu, L., Wu, J., Zhou, J.-Q., Teng, M., Shi, Y. Structural Analysis of Rtt106p Reveals a DNA Binding Role Required for Heterochromatin Silencing. *Journal of Biological Chemistry*, 2010, **285** : 4251-4262.
- López-Perrote, A., Muñoz-Hernández, H., Gil, D., Llorca, O. Conformational transitions regulate the exposure of a DNA-binding domain in the RuvBL1–RuvBL2 complex. *Nucleic Acids Research*, 2012, **40** : 11086-11099.

- M**akhnevych, T., Houry, W.A. The role of Hsp90 in protein complex assembly. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2012, **1823** : 674-682.
- Marcel, V., Catez, F., Diaz, J.-J. Ribosome heterogeneity in tumorigenesis: the rRNA point of view. *Molecular & Cellular Oncology*, 2015, **2** : e983755.
- 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., Bourdon, J.-C., Jordan, L., Thompson, A., Tafer, Y., Cong, R., Bouvet, P., Saurin, J.-C., Catez, F., Prats, A.-C., Puisieux, A., Diaz, J.-J. p53 Acts as a Safeguard of Translational Control by Regulating Fibrillarin and rRNA Methylation in Cancer. *Cancer Cell*, 2013, **24** : 318-330.
- Marmier-Gourrier, N. A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA*, 2003, **9** : 821-838.
- Matera, A.G., Terns, R.M., Terns, M.P. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature Reviews Molecular Cell Biology*, 2007, **8** : 209-220.
- Matias, P.M., Gorynia, S., Donner, P., Carrondo, M.A. Crystal Structure of the Human AAA+ Protein RuvBL1. *Journal of Biological Chemistry*, 2006, **281** : 38918-38929.
- McKeegan, K.S., Debieux, C.M., Boulon, S., Bertrand, E., Watkins, N.J. A Dynamic Scaffold of Pre-snoRNP Factors Facilitates Human Box C/D snoRNP Assembly. *Molecular and Cellular Biology*, 2007, **27** : 6782-6793.
- McKeegan, K.S., Debieux, C.M., Watkins, N.J. Evidence that the AAA+ Proteins TIP48 and TIP49 Bridge Interactions between 15.5K and the Related NOP56 and NOP58 Proteins during Box C/D snoRNP Biogenesis. *Molecular and Cellular Biology*, 2009, **29** : 4971-4981.
- Meyer, P., Prodromou, C., Hu, B., Vaughan, C., Roe, S.M., Panaretou, B., Piper, P.W., Pearl, L.H. Structural and Functional Analysis of the Middle Segment of Hsp90: Implications for ATP Hydrolysis and Client Protein and Cochaperone Interactions. *Molecular Cell*, 2003, **11** : 647-658.
- Millson, S.H., Vaughan, C.K., Zhai, C., Ali, M.M.U., Panaretou, B., Piper, P.W., Pearl, L.H., Prodromou, C. Chaperone ligand-discrimination by the TPR-domain protein Tah1. *Biochemical Journal*, 2008, **413** : 261-268.
- Moore, T., Zhang, Y., Fenley, M.O., Li, H. Molecular Basis of Box C/D RNA-Protein Interactions. *Structure*, 2004, **12** : 807-818.
- Morgan, R.M.L., Pal, M., Roe, S.M., Pearl, L.H., Prodromou, C. Tah1 helix-swap dimerization prevents mixed Hsp90 co-chaperone complexes. *Acta Crystallographica Section D: Biological Crystallography*, 2015, **71** : 1197-1206.
- Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J., Bordonné, R. Hypermethylation of the Cap Structure of Both Yeast snRNAs and snoRNAs Requires a Conserved Methyltransferase that Is Localized to the Nucleolus. *Molecular Cell*, 2002, **9** : 891-901.

- N**akao, A., Yoshihama, M., Kenmochi, N. RPG: the Ribosomal Protein Gene database. *Nucleic Acids Research*, 2004, **32** : D168-170.
- Nano, N., Houry, W.A. Chaperone-like activity of the AAA+ proteins Rvb1 and Rvb2 in the assembly of various complexes. *Phil. Trans. R. Soc. B*, 2013, **368** : 20110399.
- Newman, D.R., Kuhn, J.F., Shanab, G.M., Maxwell, E.S. Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA*, 2000, **6** : 861-879.
- Nguyen, V.Q., Ranjan, A., Stengel, F., Wei, D., Aebersold, R., Wu, C., Leschziner, A.E. Molecular Architecture of the ATP-Dependent Chromatin-Remodeling Complex SWR1. *Cell*, 2013, **154** : 1220-1231.
- Niewiarowski, A., Bradley, A.S., Gor, J., McKay, A.R., Perkins, S.J., Tsaneva, I.R. Oligomeric assembly and interactions within the human RuvB-like RuvBL1 and RuvBL2 complexes. *Biochemical Journal*, 2010, **429** : 113-125.
- Nolivos, S., Carpousis, A.J., Clouet-d'Orval, B. The K-loop, a general feature of the Pyrococcus C/D guide RNAs, is an RNA structural motif related to the K-turn. *Nucleic Acids Research*, 2005, **33** : 6507-6514.
- O**chs, R.L., Lischwe, M.A., Spohn, W.H., Busch, H. Fibrillarin: a new protein of the nucleolus identified by autoimmune sera. *Biology of the Cell*, 1985, **54** : 123-133.
- Omer, A.D., Ziesche, S., Ehardt, H., Dennis, P.P. In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proceedings of the National Academy of Sciences*, 2002, **99** : 5289-5294.
- Ooi, S.L., Samarsky, D.A., Fournier, M.J., Boeke, J.D. Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA. *RNA*, 1998, **4** : 1096-1110.
- Oruganti, S. vidya, Zhang, Y., Li, H. Structural comparison of yeast snoRNP and spliceosomal protein Snu13p with its homologs. *Biochemical and Biophysical Research Communications*, 2005, **333** : 550-554.
- P**aci, A., Liu, P.X.H., Zhang, L., Zhao, R. The Proteasome Subunit Rpn8 Interacts with the Small Nucleolar RNA Protein (snoRNP) Assembly Protein Pih1 and Mediates Its Ubiquitin-independent Degradation in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 2016, **291** : 11761-11775.
- Paci, A., Liu, X.H., Huang, H., Lim, A., Houry, W.A., Zhao, R. The Stability of the Small Nucleolar Ribonucleoprotein (snoRNP) Assembly Protein Pih1 in *Saccharomyces cerevisiae* Is Modulated by Its C Terminus. *Journal of Biological Chemistry*, 2012, **287** : 43205-43214.

- Pal, M., Morgan, M., Phelps, S.E.L., Roe, S.M., Parry-Morris, S., Downs, J.A., Polier, S., Pearl, L.H., Prodromou, C. Structural Basis for Phosphorylation-Dependent Recruitment of Tel2 to Hsp90 by Pih1. *Structure*, 2014, **22** : 805-818.
- Palade, G.E. A small particulate component of the cytoplasm. *The Journal of Biophysical and Biochemical Cytology*, 1955, **1** : 59-68.
- Pelton, J.G., Torchia, D.A., Meadow, N.D., Roseman, S. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated III₂Glc, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques. *Protein Science : A Publication of the Protein Society*, 1993, **2** : 543-558.
- Peng, W.-T., Robinson, M.D., Mnaimneh, S., Krogan, N.J., Cagney, G., Morris, Q., Davierwala, A.P., Grigull, J., Yang, X., Zhang, W., Mitsakakis, N., Ryan, O.W., Datta, N., Jovic, V., Pal, C., Canadien, V., Richards, D., Beattie, B., Wu, L.F., Altschuler, S.J., Roweis, S., Frey, B.J., Emili, A., Greenblatt J.F., Hughes, T.R. A panoramic view of yeast noncoding RNA processing. *Cell*, 2003, **113** : 919–933.
- Peng, Y., Yu, G., Tian, S., Li, H. Co-Expression and Co-Purification of Archaeal and Eukaryal Box C/D RNPs. *PLoS ONE*, 2014, **9**.
- Petfalski, E., Dandekar, T., Henry, Y., Tollervey, D. Processing of the Precursors to Small Nucleolar RNAs and rRNAs Requires Common Components. *Molecular and Cellular Biology*, 1998, **18** : 1181-1189.
- Petukhov, M., Dagkessamanskaja, A., Bommer, M., Barrett, T., Tsaneva, I., Yakimov, A., Quéval, R., Shvetsov, A., Khodorkovskiy, M., Käs, E., Grigoriev, M. Large-Scale Conformational Flexibility Determines the Properties of AAA+ TIP49 ATPases. *Structure*, 2012, **20** : 1321-1331.
- Piekna-Przybylska, D., Decatur, W.A., Fournier, M.J. New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. *RNA (New York, N.Y.)*, 2007, **13** : 305-312.
- Pradet-Balade, B., Girard, C., Boulon, S., Paul, C., Azzag, K., Bordonné, R., Bertrand, E., Verheggen, C. CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *The EMBO Journal*, 2011, **30** : 2205-2218.
- Prodromou, C. Mechanisms of Hsp90 regulation. *Biochemical Journal*, 2016, **473** : 2439-2452.
- Qu, G., Nues, R.W. van, Watkins, N.J., Maxwell, E.S. The Spatial-Functional Coupling of Box C/D and C'/D' RNPs Is an Evolutionarily Conserved Feature of the Eukaryotic Box C/D snoRNP Nucleotide Modification Complex. *Molecular and Cellular Biology*, 2011, **31** : 365-374.
- Quinternet, M., Chagot, M.-E., Rothé, B., Tiotiu, D., Charpentier, B., Manival, X. Structural Features of the Box C/D snoRNP Pre-assembly Process Are Conserved through Species. *Structure*, 2016, **0**.

Quinternet, M., Rothé, B., Barbier, M., Bobo, C., Saliou, J.-M., Jacquemin, C., Back, R., Chagot, M.-E., Cianféroni, S., Meyer, P., Branlant, C., Charpentier, B., Manival, X. Structure/Function Analysis of Protein–Protein Interactions Developed by the Yeast Pih1 Platform Protein and Its Partners in Box C/D snoRNP Assembly. *Journal of Molecular Biology*, 2015, **427** : 2816-2839.

Ramakrishnan, V. Histone structure and the organization of the nucleosome. *Annual review of biophysics and biomolecular structure*, 1997, **26** : 83–112.

Richter, K., Muschler, P., Hainzl, O., Reinstein, J., Buchner, J. Sti1 Is a Non-competitive Inhibitor of the Hsp90 ATPase BINDING PREVENTS THE N-TERMINAL DIMERIZATION REACTION DURING THE ATPASE CYCLE. *Journal of Biological Chemistry*, 2003, **278** : 10328-10333.

Robert, X., Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 2014, **42** : W320-W324.

Rodriguez-Corona, U., Sobol, M., Rodriguez-Zapata, L.C., Hozak, P., Castano, E. Fibrillarin from Archaea to human. *Biology of the Cell*, 2015, **107** : 159-174.

Rothe, B., Back, R., Quinternet, M., Bizarro, J., Robert, M.-C., Blaud, M., Romier, C., Manival, X., Charpentier, B., Bertrand, E., Branlant, C. Characterization of the interaction between protein Snu13p/15.5K and the Rsa1p/NUFIP factor and demonstration of its functional importance for snoRNP assembly. *Nucleic Acids Research*, 2014a, **42** : 2015-2036.

Rothe, B., Saliou, J.-M., Quinternet, M., Back, R., Tiotiu, D., Jacquemin, C., Loegler, C., Schlotter, F., Pena, V., Eckert, K., Morera, S., Dorsselaer, A.V., Branlant, C., Massenet, S., Sanglier-Cianferani, S., Manival, X., Charpentier, B. Protein Hit1, a novel box C/D snoRNP assembly factor, controls cellular concentration of the scaffolding protein Rsa1 by direct interaction. *Nucleic Acids Research*, 2014, **42** : 10731-10747.

Rothé, B. *Étude des processus de biogenèse des petites particules ribonucléoprotéiques nucléolaires à boîtes C/D (snoRNP C/D) chez la levure Saccharomyces cerevisiae. Caractérisation fonctionnelle et structurale d'une machinerie dédiée à l'assemblage de ces RNP.*

Schalch, T., Duda, S., Sargent, D.F., Richmond, T.J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 2005, **436** : 138-141.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F.U., Moarefi, I. Structure of TPR Domain–Peptide Complexes. *Cell*, 2000, **101** : 199-210.

Schmeing, T.M., Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature*, 2009, **461** : 1234-1242.

Schubert, H.L., Blumenthal, R.M., Cheng, X. Many paths to methyltransfer: a chronicle of convergence. *Trends in Biochemical Sciences*, 2003, **28** : 329-335.

Schuette, J.-C., Murphy, F.V., Kelley, A.C., Weir, J.R., Giesebrecht, J., Connell, S.R., Loerke, J., Mielke, T., Zhang, W., Penczek, P.A., Ramakrishnan, V., Spahn, C.M.T. GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *The EMBO journal*, 2009, **28** : 755-765.

Silva-Martin, N., Daudén, M.I., Glatt, S., Hoffmann, N.A., Kastritis, P., Bork, P., Beck, M., Müller, C.W. The Combination of X-Ray Crystallography and Cryo-Electron Microscopy Provides Insight into the Overall Architecture of the Dodecameric Rvb1/Rvb2 Complex. *PLOS ONE*, 2016, **11** : e0146457.

Slavov, N., Semrau, S., Airoidi, E., Budnik, B., van Oudenaarden, A. Differential Stoichiometry among Core Ribosomal Proteins. *Cell Reports*, 2015, **13** : 865-873.

Snider, J., Thibault, G., Houry, W.A. The AAA+ superfamily of functionally diverse proteins. *Genome biology*, 2008, **9** : 1.

Su, D., Hu, Q., Li, Q., Thompson, J.R., Cui, G., Fazly, A., Davies, B.A., Botuyan, M.V., Zhang, Z., Mer, G. Structural basis for recognition of H3K56-acetylated histone H3–H4 by the chaperone Rtt106. *Nature*, 2012, **483** : 104-107.

Su, D., Hu, Q., Li, Q., Thompson, J.R., Cui, G., Fazly, A., Davies, B.A., Botuyan, M.V., Zhang, Z., Mer, G. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nature Reviews Molecular Cell Biology*, 2010, **11** : 515-528.

Terns, M.P., Terns, R.M. Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expression*, 2002, **10** : 17-39.

Tessarz, P., Santos-Rosa, H., Robson, S.C., Sylvestersen, K.B., Nelson, C.J., Nielsen, M.L., Kouzarides, T. Glutamine methylation in Histone H2A is an RNA Polymerase I dedicated modification. *Nature*, 2014, **505** : 564-568.

Torreira, E., Jha, S., López-Blanco, J.R., Arias-Palomo, E., Chacón, P., Cañas, C., Ayora, S., Dutta, A., Llorca, O. Architecture of the Pontin/Reptin Complex, Essential in the Assembly of Several Macromolecular Complexes. *Structure*, 2008, **16** : 1511-1520.

Tosi, A., Haas, C., Herzog, F., Gilmozzi, A., Berninghausen, O., Ungewickell, C., Gerhold, C.B., Lakomek, K., Aebersold, R., Beckmann, R., Hopfner, K.-P. Structure and Subunit Topology of the INO80 Chromatin Remodeler and Its Nucleosome Complex. *Cell*, 2013, **154** : 1207-1219.

Tycowski, K.T., Steitz, J.A. Non-coding snoRNA host genes in *Drosophila*: expression strategies for modification guide snoRNAs. *European Journal of Cell Biology*, 2001, **80** : 119-125.

VanDemark, A.P., Blanksma, M., Ferris, E., Heroux, A., Hill, C.P., Formosa, T. The Structure of the yFACT Pob3-M Domain, Its Interaction with the DNA Replication Factor RPA, and a Potential Role in Nucleosome Deposition. *Molecular Cell*, 2006, **22** : 363-374.

- van Nues, R.W., Watkins, N.J. Unusual C'/D' motifs enable box C/D snoRNPs to modify multiple sites in the same rRNA target region. *Nucleic Acids Research*, 2016 : gkw842.
- Venteicher, A.S., Meng, Z., Mason, P.J., Veenstra, T.D., Artandi, S.E. Identification of ATPases Pontin and Reptin as Telomerase Components Essential for Holoenzyme Assembly. *Cell*, 2008, **132** : 945-957.
- Verheggen, C., Pradet-Balade, B., Bertrand, E., Verheggen, C., Pradet-Balade, B., Bertrand, E. SnoRNPs, ZNHIT proteins and the R2TP pathway. *Oncotarget*, 2015, **6** : 41399-41400.
- Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R., Ficner, R. Crystal Structure of the Spliceosomal 15.5kD Protein Bound to a U4 snRNA Fragment. *Molecular Cell*, 2000, **6** : 1331-1342.
- von Morgen, P., Hořejší, Z., Macurek, L. Substrate recognition and function of the R2TP complex in response to cellular stress. *Frontiers in Genetics*, 2015, **6**.
- W**alker, J.E., Saraste, M., Runswick, M.J., Gay, N.J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO Journal*, 1982, **1** : 945.
- Wang, H., Boisvert, D., Kim, K.K., Kim, R., Kim, S.-H. Crystal structure of a fibrillarin homologue from *Methanococcus jannaschii*, a hyperthermophile, at 1.6 Å resolution. *The EMBO Journal*, 2000, **19** : 317-323.
- Watkins, N.J., Lemm, I., Ingelfinger, D., Schneider, C., Holsbach, M., Urlaub, H., Lührmann, R. Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. *Molecular cell*, 2004, **16** : 789–798.
- Watkins, N.J., Ségault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C., Lührmann, R. A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, 2000, **103** : 457–466.
- Watkins, N.J., Dickmanns, A., Lührmann, R. Conserved Stem II of the Box C/D Motif Is Essential for Nucleolar Localization and Is Required, Along with the 15.5K Protein, for the Hierarchical Assembly of the Box C/D snoRNP. *Molecular and Cellular Biology*, 2002, **22** : 8342-8352.
- Watkins, N.J., Leverette, R.D., Xia, L., Andrews, M.T., Maxwell, E.S. Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA*, 1996, **2** : 118-133.
- Weinstein, L.B., Steitz, J.A. Guided tours: from precursor snoRNA to functional snoRNP. *Current Opinion in Cell Biology*, 1999, **11** : 378-384.
- Woolford, J.L., Baserga, S.J. Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics*, 2013, **195** : 643-681.

Xue, S., Barna, M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nature reviews. Molecular cell biology*, 2012, **13** : 355-369.

Xue, S., Wang, R., Yang, F., Terns, R.M., Terns, M.P., Zhang, X., Maxwell, E.S., Li, H. Structural Basis for Substrate Placement by an Archaeal Box C/D Ribonucleoprotein Particle. *Molecular Cell*, 2010, **39** : 939-949.

Ye, K., Jia, R., Lin, J., Ju, M., Peng, J., Xu, A., Zhang, L. Structural organization of box C/D RNA-guided RNA methyltransferase. *Proceedings of the National Academy of Sciences*, 2009, **106** : 13808-13813.

Yip, W.S.V., Shigematsu, H., Taylor, D.W., Baserga, S.J. Box C/D sRNA stem ends act as stabilizing anchors for box C/D di-sRNPs. *Nucleic Acids Research*, 2016, **44** : 8976-8989.

Zhang, M., Botër, M., Li, K., Kadota, Y., Panaretou, B., Prodromou, C., Shirasu, K., Pearl, L.H. Structural and functional coupling of Hsp90- and Sgt1-centred multi-protein complexes. *The EMBO Journal*, 2008, **27** : 2789-2798.

Zhao, R., Davey, M., Hsu, Y.-C., Kaplanek, P., Tong, A., Parsons, A.B., Krogan, N., Cagney, G., Mai, D., Greenblatt, J., Boone, C., Emili, A., Houry, W.A. Navigating the Chaperone Network: An Integrative Map of Physical and Genetic Interactions Mediated by the Hsp90 Chaperone. *Cell*, 2005, **120** : 715-727.

Zhao, R., Kakihara, Y., Gribun, A., Huen, J., Yang, G., Khanna, M., Costanzo, M., Brost, R.L., Boone, C., Hughes, T.R., Yip, C.M., Houry, W.A. Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. *The Journal of Cell Biology*, 2008, **180** : 563-578.

Zunder, R.M., Antczak, A.J., Berger, J.M., Rine, J. Two surfaces on the histone chaperone Rtt106 mediate histone binding, replication, and silencing. *Proceedings of the National Academy of Sciences*, 2012, **109** : E144-E153.

Caractérisation structurale et fonctionnelle de la protéine Bcd1, impliquée dans la biogenèse des snoRNP à boîtes C/D chez la levure *Saccharomyces cerevisiae*.

La protéine Bcd1 est un facteur nucléaire essentiel à la viabilité cellulaire de la levure *Saccharomyces cerevisiae*. Il est décrit comme requis pour assurer la stabilité des snoRNA à boîtes C/D. Ces petits ARN non codants s'assemblent à un jeu de 4 protéines invariables pour former les snoRNP à boîtes C/D qui sont des acteurs cruciaux de la biogenèse des ribosomes. En effet, quelques-unes de ces particules participent aux mécanismes assurant la maturation du précurseur des ARN ribosomiques et la grande majorité des autres particules sont des catalyseurs de la modification par 2'-*O*-méthylation des riboses. Bcd1p n'est pas présente au sein des particules matures, mais fait partie de ses facteurs d'assemblage, au même titre que les sous-complexes Rsa1p:Hit1p et R2TP (Rvb1p:Rvb2p:Tah1p:Pih1p). Notre analyse de différents fragments de Bcd1p a dans un premier temps montré que sa région N-terminale (résidus 1 à 96) suffit à lui conférer son caractère essentiel. Cette région comprend un domaine à double doigt à zinc de la famille zf-HIT, également présent chez un autre facteur d'assemblage des snoRNP à boîtes C/D, la protéine Hit1. Nous avons résolu la structure 3D en solution de ces doigts à zinc et montré que ce sont des modules d'interaction avec les protéines Rvb1/2. Dans un second temps nous avons identifié la région C-terminale (résidus 120 à 303) de la protéine Bcd1 comme étant suffisante pour interagir avec la chaperonne d'histone Rtt106p. La structure 3D en solution de ce domaine a été déterminée par RMN. Différentes approches de cinétique d'échange hydrogène/deutérium et d'expériences de *cross-link* suivies par des analyses par spectrométrie de masse, des expériences de titrage par RMN et de SAXS nous ont permis d'obtenir des informations sur les surfaces d'interaction de chacune de ces deux protéines. Un fragment, défini à partir des données de RMN de Bcd1p libre, nous a permis d'obtenir des cristaux du complexe Bcd1p:Rtt106p ouvrant la perspective de résoudre sa structure 3D par diffraction aux rayons X. De plus, des études fonctionnelles ont débuté visant à déterminer l'importance de la formation de ce complexe sur la biogenèse des snoRNP à boîtes C/D et l'impact de Bcd1p sur l'interaction entre Rtt106p et les nucléosomes.

Mots clés : Facteurs d'assemblage, Bcd1p, Hit1p, doigt à zinc, snoRNP à boîtes C/D, Rtt106p, chaperonne d'histone, RMN, spectrométrie de masse, SAXS, structure 3D.

Structural and functional characterization of protein Bcd1, implicated in box C/D snoRNP biogenesis in the yeast *Saccharomyces cerevisiae*.

The protein Bcd1 is a nuclear factor essential for the cellular viability of the yeast *Saccharomyces cerevisiae*. It is described as required to ensure box C/D snoRNA stability. These small non-coding RNAs associate with an invariable set of 4 proteins to form the box C/D snoRNPs that are crucial players in ribosome biogenesis. Indeed, some of these particles participate in mechanisms for the maturation of the ribosomal RNA precursor (pre-rRNA) and the vast majority of the other particles are catalysts of 2'-*O*-methylation of riboses. Bcd1p is not present in mature particles, but is one of the assembly factors in addition to the Rsa1p:Hit1p and R2TP (Rvb1p:Rvb2p:Tah1p:Pih1p) sub-complexes. Our analysis of the different Bcd1p fragments has firstly shown that the essential function of Bcd1p relies on its N-terminal region (residues 1 to 96). It comprises a double zinc finger domain from the zf-HIT family, also present in another box C/D snoRNP assembly factor, the protein Hit1. We solved the 3D solution structure of these two zinc fingers and showed that these are modules for the interaction of Bcd1p with the Rvb1/2 proteins. Secondly, we identified the C-terminal region (residues 120 to 303) of Bcd1p as being sufficient to interact with the histone chaperone Rtt106p. The 3D solution structure of this domain of Bcd1p was determined by NMR. Different approaches of hydrogen/deuterium kinetic exchange and cross-link experiments followed by mass spectrometry analysis, NMR titration, and SAXS allowed us to obtain information about the interaction surfaces on each of the two proteins. A fragment defined from NMR data on the free Bcd1p allowed us to obtain crystals of the Bcd1p:Rtt106p complex, opening the perspective to solve its 3D structure by X-ray diffraction. Furthermore, functional studies started in order to determine the importance of this complex formation in box C/D snoRNP biogenesis and the impact of Bcd1p on the interaction of Rtt106p with nucleosomes.

Keywords: Assembly factor, Bcd1p, Hit1p, zinc finger, box C/D snoRNP, Rtt106p, histone chaperone, NMR, mass spectrometry, SAXS, 3D structure.