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

Nancy-Université
Université Henri Poincaré

UFR de Sciences et Techniques Biologiques

Ecole Doctorale BioSE (Biologie-Santé-Environnement)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE l'UNIVERSITE HENRI POINCARE

Mention ARN : Biogenèse et Structure / Fonction

par **Jean-Baptiste FOURMANN**

Analyse structurale et fonctionnelle des particules sRNP à boîtes H/ACA, catalysant la formation ciblée des pseudouridines dans les ARN d'archaea

Soutenance publique le 13 Novembre 2009

Membres du jury :

Rapporteurs :	**M. Yves HENRY**	**Directeur de recherche, UMR-CNRS, Toulouse**
	M. Nicolas LEULLIOT	**Professeur, Université Paris Descartes, Paris**
Examinateur :	**M. Alain KROL (Président)**	**Directeur de Recherche, CNRS, Strasbourg**
	M^{me} Christiane BRANLANT	**Directeur de Recherche, UMR-CNRS, Nancy**
	M. Bruno CHARPENTIER	**Professeur, Université Henri Poincaré, Nancy** **Directeur de thèse**

Laboratoire AREMS : ARN, RNP, maturation-structure-fonction, Enzymologie Moléculaire et Structurale UMR 7214 CNRS-UHP
Faculté des Sciences & Techniques - 54506 Vandoeuvre-lès-Nancy

Remerciements

Je tiens à exprimer ma gratitude à Christiane Branlant, directeur du laboratoire, pour m'avoir accueilli et conseillé scientifiquement.

Je tiens particulièrement à remercier mon directeur de thèse Bruno Charpentier, pour son encadrement, sa bonne humeur, sa rigueur et dans sa transmission d'un savoir faire aussi bien expérimental que dans l'enseignement. Dire que j'ai beaucoup appris avec lui est bien peu en comparaison de tout ce qu'il m'a apporté.

Je remercie Yves Henry et Nicolas Leulliot d'avoir acceptés d'être rapporteur de ce mémoire. Je remercie également Alain Krol qui a accepté de juger ce travail.

Sébastien qui m'a permis de faire ce travail et partagé de nombreuses découvertes et la stimulation mutuelle dans les différents thèmes de recherches que nous avons abordés en commun.

Magali qui a réussi à me transmettre sa rigueur au travail et sa joie de vivre communicative.

Xavier M. qui m'a fait découvrir le laboratoire lors d'un stage et partagé sa passion pour la structure et les protéines.

Le laboratoire.

Christine pour son aide technique précieuse.

Valérie, Véro, Antoine, Ben et Nico pour leurs conseils avisés et leur aide dans mon travail. Les personnes qui ont « souffert » en même temps que moi Sylvain, Xavier R. et Maryline. Les nouveaux venus Anne-Sophie qui a repris le flambeau, Véro et Thomas. Toutes les autres personnes du laboratoire que je remercie pour leur sympathie et leur gentillesse.

Mes amis.

Clarisse, que je remercie très affectueusement, m'a supporté, aimé et soutenu ; sa famille.

Ma mère, qui bien que je ne l'ai pas toujours éclairée sur mes tenants et mes aboutissants et pour qui cela n'a pas toujours été facile de comprendre, m'a toujours aidé et encouragé.

Mon père qui aurait été très heureux, moi je ne l'oublie pas.

A tous, merci.

Liste des abréviations

A : adénine
aa : acide aminé
AA : Anémie aplastique
AD : autosomal dominant
ADN : acide désoxyribonucléique
ADNc : acide désoxyribonucléique complémentaire
AR : autosomal récessif
ARN : acide ribonucléique
ARNm : acide ribonucléique messager
ARNnc : ARN non codant
ARN pol : ARN polymérase
ARNr : acide ribonucléique ribosomique
ARNt : acide ribonucléique de transfert
ARNtm : acide ribonucléique transfert messager
ATP : adénosine 5' triphosphate
C : cytosine
°C : degré celsius
Ci : curie
CB : « Cajal Bodies », Corps de Cajal
CMCT : 1-cyclohexyl-3-(2-morpholinoéthyl) carbodiimide métha-p-toluène
cpm : coups par minute
cps : coups par seconde
Ct : extrémité C-terminale
Da, KDa : dalton, kilodalton
db : double brin
DC : Dyskératose Congénitale
ddNTP : didésoxyribonucléotide triphosphate
DNase : désoxyribonucléase
dNTP : désoxyribonucléotide triphosphate
DO : densité optique
DTT : dithiothréitol
EDTA : éthylène diamine tétraacétate
ETS : external transcribed sequence
5-FU : 5-fluoro-uracile
g/ mg/ µg/ ng : gramme/ milligramme/ microgramme/ nanogramme
G : guanine
gRNA : ARN guide
GST : glutathione S-transferase
h : heure
HH : syndrôme de Hoyeraal-Hreidrasson
I : inosine
IPTG : isopropyl β-D-thiogalactoside

Kb : kilobase
l/ ml/ µl : litre/ millilitre/ microlitre
LB : Luria Bertani
LSU : large subunit ribosomal
M/ mM/ µM/ nM : molaire/ millimolaire/ micromolaire/ nanomolaire
min : minute
NMP : ribonucléotide monophosphate
NTP : ribonucléotide triphosphate
nt : nucléotide
Nt : extrémité N-terminale
ORF : phase ouverte de lecture (Open Reading Frame)
pb : paire de base
PCR : amplification en chaine par une polymérase (polymerase chain reaction)
PEG : polyéthylène glycol
pGal : promoteur galactose dépendant
pré-ARNm : précurseur de l'ARN messager
pré-ARNr : précurseur de l'ARN ribosomique
pré-ARNt : précurseur de l'ARN de transfert
qsp : quantité suffisante pour
R : purine (adénine ou guanine)
RMN : résonance magnétique nucléaire
RNase : ribonucléase
RNasin : inhibiteur des ribonucléases
rNTP : ribonucléotide triphosphate
RT : reverse transcriptase
RX : rayons X
s : seconde
S : Svedberg
sb : simple brin
scaRNA : petit ARN contenu dans les Cajal Bodies (Small Cajal Bodies RNA)
scaRNP : particule ribonucléoprotéique contenant un ou des petits ARN contenus dans les Cajal Bodies (Small Cajal Bodies ribonucleoprotein particule)
SDS : sodium dodécylsulfate
SELEX : Systematic Evolution of Ligands by EXponential enrichment
sRNA : petit ARN (Small RNA)
sRNP : particule ribonucléoprotéique contenant un ou des petits ARN (Small ribonucleoprotein particule)
snRNA : petit ARN nucléaire (Small nuclear RNA)
snRNP : particule ribonucléoprotéique contenant un ou des petits ARN nucléaires (Small nuclear Ribonucleoprotein Particule)
snoRNA : petit ARN nucléolaire (Small nucleolar RNA)
snoRNP : particule ribonucléoprotéique contenant un ou des petits ARN nucléolaires (Small nucleolar ribonucleoprotein particule)

spliceosome : complexe nucléaire ribonucléoprotéique dans lequel ont lieu les réactions d'épissage
SSU : small subunit ribosomal
structure 2D : structure secondaire
structure 3D : structure tertiaire
T : thymine
TLC : chromatographie sur couche mince
U (UE) : unité enzymatique
UsnRNA : « uridine rich small nuclear ribonucleic acid »
UsnRNA : « uridine rich small nuclear ribonucleoprotein particle »
UTR : untranslated region
Ψ : pseudouridine
Y : pyrimidine (cytosine, uracile ou thymine)

SOMMAIRE

Table des matières

Remerciements ..... *i*

Liste des abréviations ..... *ii*

Première partie : Introduction

Avant propos *1*

I. Les Archaea : un intérêt scientifique grandissant pour ce domaine du vivant *5*

A. Intérêt d'étudier les Archaea **5**

1. Des caractéristiques partagées avec celles des bactéries et des eucaryotes **5**

2. Une grande diversité d'espèces **6**

B. *Pyrococcus abyssi* **7**

II. Les nucléotides modifiés dans les ARN : les résidus pseudouridines ... *8*

A. Propriétés des résidus pseudouridines (Ψ) **9**

B. Localisation et rôles des pseudouridines dans les ARN **10**

1. Les pseudouridines dans les ARNr **10**

1.1. Les modifications dans la petite sous-unité ribosomique 11

1.2. Les modifications de la grande sous-unité ribosomique 12

2. Les pseudouridines dans les snRNA **13**

2.1. L'épissage des pré-ARNm 13

2.2. Les snRNA 14

3. Les pseudouridines dans les ARNt **16**

3.1. Les positions des résidus Ψ dans l'ARNt 17

3.2. La position 55 des ARNt 18

III. Les enzymes de modification : les ARN:pseudouridine-synthases *19*

A. Découverte et classification des ARN:pseudouridine-synthases **20**

B. Mécanisme catalytique de la réaction de pseudouridylation **21**

1. Deux mécanismes catalytiques proposés **21**

2. Le 5-fluorouracile (5-FU), un composé efficace pour inhiber les ARN:Ψ-synthases ? 23

3. 5-FU et incertitude sur la nature du mécanisme catalytique 24

C. Structure 3D des ARN:Ψ-synthases 26

D. Les motifs structuraux et le site actif des ARN:Ψ-synthases 27

1. Le motif I assure un rôle structural 28

2. Le motif II intervient directement dans la catalyse 28

2.1. L'histidine 43 (aCBF5 H77) 28

2.2. L'aspartate 48 catalytique (aCBF5 D82) 29

2.3. La tyrosine 76 (aCBF5 Y110) 29

3. Le motif III participe à la catalyse 30

3.1. La tyrosine 179 (aCBF5 Y179) 30

3.2. L'arginine 181 (aCBF5 R181) 31

3.3. L'arginine 202 (aCBF5 R202) 31

E. Spécificité de reconnaissance des ARN:Ψ-synthases 32

1. Reconnaissance du substrat par le domaine catalytique 32

1.1. TruB et RluA : le mécanisme d'extrusion de base 33

1.1.1. Une enzyme site-spécifique : TruB 33

1.1.2. Une enzyme à double spécificité : RluA 35

1.1.3. Le mécanisme d'extrusion de base 36

1.2. Une enzyme région spécifique : TruA 37

1.3. Une enzyme site-spécifique : RsuA 38

1.4. Une enzyme région-spécifique : PUS10 39

1.5. Une enzyme site-spécifique : TruD 40

2. Reconnaissance du substrat par les domaines additionnels 41

2.1. Le domaine PUA 41

2.2. Le domaine S4 42

2.3. Le domaine THUMP 43

2.4. Le domaine TruD 44

3. Cas particuliers de reconnaissance 44

IV. Les systèmes de modification reposant sur les ARN guides 44

A. Le nucléole et la maturation du pré-ARN ribosomique 45
1. Chez les Eucaryotes 45
2. Chez les Archaea 46
B. Les sno/sRNP à boîtes C/D 47
1. Propriétés et rôles des 2'-*O*-méthylations 47
2. Chez les Eucaryotes 48
2.1. Les ARN guides à boîtes C/D 48
2.2. Le motif en Kink-turn (K-turn) 49
2.3. Les protéines associées aux snoRNA à boîtes C/D 51
2.4. Biogenèse des snoRNP à boîtes C/D 52
3. Chez les Archaea 54
4. Reconstitution de particules à boîtes C/D d'archaea : information structurales et comparaison avec le système eucaryote 56
C. Les sno/sRNP à boîtes H/ACA 58
1. L'ARN guide à boîtes H/ACA 59
1.1. Chez les Eucaryotes 59
1.2. Les scaRNA et l'ARN de la télomérase 61
1.3. Chez les Archaea 62
2. Les protéines associées aux RNP à boîtes H/ACA 63
2.1. Chez les Eucaryotes 63
2.2. Chez les Archaea 65
2.3. Localisation des gènes codant les protéines H/ACA sur les génomes d'archaea 66
3. Evolution des s/snoRNP 66
4. La biogenèse des particules snoRNP H/ACA 67
4.1. Chez les Eucaryotes 67
4.1.1. Schéma général de la biogenèse 67
4.1.2. Les facteurs associés à la biogenèse 68
4.2. Chez les Archaea 70
5. Reconstitution de particules à boîtes H/ACA 70
5.1. Chez les Eucaryotes 70
5.2. Chez les Archaea 71

6. Vue d'ensemble de la structure 3D d'une sRNP à boîtes H/ACA d'archaea .. 73

6.1. Les protéines associées aux sRNA à boîtes H/ACA d'archaea dans les structures 3D résolues en absence d'ARN 73

6.2. Structure 3D de la sRNP à boîtes H/ACA 75

6.3. Interactions entre aCBF5 et le sRNA H/ACA 75

6.4. Interaction entre aCBF5 et aNOP10 76

6.5. Interactions entre aCBF5 et aGAR1 77

6.6. Interactions entre aNOP10, L7Ae et le sRNA 77

6.7. Positionnement du substrat dans la sRNP 78

6.8. Changement de conformation lié à l'assemblage d'une sRNP H/ACA 80

6.8.1. Remodelage de la poche de pseudouridylation par L7Ae 80

6.8.2. Positionnement du substrat au sein du site actif de aCBF5 80

7. La télomérase : une RNP indispensable à la biologie de la cellule 81

7.1. Les télomères et la réplication des chromosomes linéaires 81

7.2. Structure de l'extrémité des télomères 83

7.3. Mécanisme de synthèse des télomères par la télomérase 84

7.4. Le complexe shelterine 84

7.5. La télomérase et ses protéines 85

7.6. La biogenèse de la télomérase 86

8. La Dyskératose Congénitale humaine, une maladie génétique rare 88

8.1. La forme liée au chromosome X 90

8.1.1. Etudes réalisées sur des cellules humaines portant de mutations dans le gène *DKC1* 90

8.1.2. Etudes réalisées sur des souris portant des mutations dans le gène *DKC1* humain 91

8.1.3. Bilan des études menées chez l'homme et chez la souris 92

8.2. La forme autosomale dominante (AD) 92

8.2.1. Mutations touchant la séquence de l'ARN de la télomérase (hTERC) 92

8.2.2. Mutations touchant la séquence de la transcriptase inverse de la télomérase (hTERT) 93

8.2.3. Mutations touchant la séquence de la protéine TIN2 93

8.3. La forme autosomale récessive (AR) 94

8.3.1. Mutation touchant la séquence de la protéine NOP10 94
8.3.2. Mutations touchant la séquence de la protéine NHP2 94
8.4. Les autres syndromes liés à la DC 95
V. Objectifs de la thèse 96

Deuxième partie : Matériel & Méthodes

I. Matériel 100
A. Liste des oligonucléotides 100
B. Les souches utilisées 100
1. Souches de bactéries 100
2. Souches de levure 100
3. Souches d'archaea 101
C. Les plasmides utilisés 101
D. Milieux de culture, tampons et solutions d'usage courant 102
II. Méthodes 103
A. Méthodes sur les protéines, les ARN et mesure de l'activité ARN:Ψ-synthase 103

Chapitre d'ouvrage :

Reconstitution of Archaeal H/ACA sRNPs and test of their activity.

Bruno CHARPENTIER, Jean-Baptiste FOURMANN and Christiane BRANLANT

Methods in Enzymology, 2007 volume 425 : 389-405.

B. Séparation des nucléotides modifiés ou non par chromatographie sur couche mince en deux dimensions 104
C. Analyse de la stabilité des ARN par Northern Blot 104
D. Analyse de la structure secondaire des macromolécules biologiques par dichroïsme circulaire 106
E. Numéros d'accès à la PDB 107

Troisième partie : Résultats & Discussions

I. Recherche de déterminants ARN qui contribuent à l'activité ou à l'assemblage des particules sRNP à boîtes H/ACA 108

A. Déterminants présents dans les sRNA et dans leurs ARN substrats 108

1. Présentation du travail 108

Publication n°1 :

Combined *in silico* and experimental identification of the *Pyrococcus abyssi* H/ACA sRNAs and their target sites in ribosomal RNAs.

Sébastien MULLER, Fabrice LECLERC, Isabelle BEHM-ANSMANT, **Jean-Baptiste FOURMANN**, Bruno CHARPENTIER and Christiane BRANLANT

Nucleic Acids Research, 2008 volume 36(8) : 5459-5475.

2. Résultats complémentaires 110

2.1. Etude *in vitro* de l'activité du sRNA Hsp3 de *H. halobium* 110

2.2. Influence sur l'activité de la sRNP Pab91 de l'identité des résidus non appariés et contigus au résidu U ciblé 112

B. Recherche de déterminants ARN contribuant à l'affinité de aCBF5 pour les sRNA à boîtes H/ACA 114

1. Contexte 114

2. Résultats 114

2.1. L'importance de la boîte ACA dans les sRNA 114

2.2. La séquence GAG contenue dans la boucle apicale n'est pas un déterminant général pour la fixation de aCBF5 aux sRNA 116

2.3. Test de l'influence d'un résidu sortant, situé à la jonction entre l'hélice 2 et la poche de pseudouridylation, sur la fixation de aCBF5 117

2.4. Contribution de l'hélice H2 dans l'assemblage et dans l'activité des sRNP118

2.5. L'hélice H1 est un déterminant important de la fixation de aCBF5 au sRNA 119

2.5.1. Influence de la permutation des hélices H1 des sRNA Pab91 et Pab21 sur la fixation de aCBF5 par l'approche de retard de migration sur gel 119

2.5.2. Influence de la permutation des hélices H1 des sRNA Pab91 et Pab21 sur la conformation globale des sRNA étudiée par dichroïsme circulaire 121

2.5.3. Modélisation de l'hélice H1 et de la poche de pseudouridylation des sRNA Pab91 et Pab21 124

2.5.4. Le taux de complexe formé entre aCBF5 et un sRNA ne détermine pas l'efficacité de la réaction de modification 125

2.6. La contribution des sRNA à l'interaction séquentielle entre aGAR1 et aCBF5 126

3. Discussion 128

3.1. L'affinité apparente de aCBF5 pour l'ARN est variable en fonction de l'identité du sRNA 128

3.2. Une séquence CCA ne peut se substituer à la séquence ACA d'un sRNA . 129

3.3. Le rôle de l'hélice H1 des sRNA 130

3.4. L'activité du sRNA Pab21 serait limitée par la séquence de sa poche de pseudouridylation 131

3.5. L'effet de aGAR1 sur la stabilité de la sRNP 132

II. Etude des protéines des sRNP à boîtes H/ACA 134

A. Etude du rôle des domaines des protéines aNOP10 et aCBF5 dans l'assemblage et dans l'activité des particules 134

1. Présentation du travail 134

Publication n°2 :

Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity.

Xavier MANIVAL, Christophe CHARRON, **Jean-Baptiste FOURMANN**, Bruno CHARPENTIER and Christiane BRANLANT

Nucleic Acids Research, 2006 volume 34(3) : 826-839.

2. Résultats complémentaires 135

2.1. L'effet positif du domaine PUA de aCBF5 sur l'activité de sRNP guidées par les sRNA Pab160 et Pab105 135

2.2. Importance de la dimérisation du complexe aCBF5-aNOP10 pour sa fixation sur le sRNA Pab105 à double motif tige-boucle 136

B. Analyse comparative des deux activités ARN:Ψ-synthase de aCBF5 : l'activité guidée et l'activité non guidée par un ARN 138

1. Présentation du travail 138

Publication n°3 :

Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Ψ55-synthase and RNA-guided RNA:Ψ-synthase activities.

Sébastien MULLER*, **Jean-Baptiste FOURMANN***, Christine LOEGLER, Bruno CHARPENTIER and Christiane BRANLANT

* Contribution équivalente au travail

Nucleic Acids Research, 2007 volume 35(16) : 5610-5624.

2. Discussion complémentaire 140

2.1. La présence de la séquence CCA 140

2.2. Le rôle des protéines aNOP10 et aGAR1 140

2.3. Le rôle de la position 54 et de aPUS10 141

C. Etude approfondie de la protéine aCBF5 : rôle de la boucle β7/β10 et de résidus du site actif dans le turn-over de la particule et dans la spécificité de substrat.. 143

1. Contexte et stratégie d'étude 143

2. Choix des mutations introduites dans la séquence codant aCBF5 de *P. abyssi* 144

2.1. Les substitutions de résidus de la boucle β7/β10 de aCBF5 144

2.2. Les substitutions de résidus du site actif de aCBF5 146

3. Résultats 147

3.1. Le résidu R143 est un déterminant important de l'interaction de aCBF5 avec aGAR1 147

3.2. Analyse de l'activité ARN:Ψ-synthase de sRNP assemblées avec des variants de la boucle β7/β10 de aCBF5 148

3.3. Analyse de l'activité ARN:Ψ-synthase de sRNP assemblées avec des variants du site actif de aCBF5 150

3.4. Analyse de l'incorporation de l'ARN substrat au sein des sRNP sauvage et variantes 150

3.5. Mise au point d'une méthode permettant d'analyser le déroulement de l'étape de libération du produit de la réaction de pseudouridylation 152

3.6. Le variant R143A de aCBF5 permet de stimuler l'activité globale de la particule sans nécessiter la présence de aGAR1 153

3.7. Les substitutions Y179A et P140AP141A n'empêchent pas une fixation stable de l'ARN substrat à la particule mais affectent l'activité catalytique 154

3.8. Recherche des résidus de aCBF5 contribuant à la sélection par la sRNP du substrat à modifier 155

3.8.1. Utilisation de substrats RNA-U/C dont le résidu U cible de la réaction est substitué par un C 155

3.8.2. Production et utilisation d'un substrat pseudouridylé, RNA-Ψ 157

3.9. Etude des variants de aCBF5 dans l'activité de modification non guidée .. 158

4. Compilation des résultats et discussion 160

4.1. Analyse des variants du site actif de aCBF5 160

4.2. Analyse des variants de la boucle β7/β10 de aCBF5 162

D. Importance fonctionnelle des interactions établies, au sein de la sRNP H/ACA, entre aNOP10 et L7Ae et entre aNOP10 et l'ARN 165

1. Présentation du travail 165

Publication n°4 :

Interactions between proteins aNOP10 and L7Ae organize archaeal box H/ACA sRNPs for stable substrate RNA binding and optimal RNA:pseudouridine-synthase activity.

Jean-Baptiste FOURMANN*, Magali BLAUD*, Christiane BRANLANT and Bruno CHARPENTIER

* Contribution équivalente au travail

En préparation pour soumission à publication.

2. Résultats complémentaires 168

2.1. Rupture de la liaison entre L7Ae et aNOP10 par l'allongement de l'hélice H2 des sRNA Pab21 et Pab91 168

2.2. Développement d'une technique d'empreintes sur gel pour analyser le complexe sRNP-ARN substrat 168

E. Utilisation des sRNA H/ACA d'archaea et des snoRNA H/ACA de levure comme modèle d'étude de la Dyskératose Congénitale 174

1. Intérêt et contexte 174

2. Effet des substitutions R34A et R34W de Nop10p sur la croissance des levures et sur la stabilité des snoRNA à boîtes H/ACA 175

Quatrième partie : Conclusion & Perspectives

Conclusion & Perspectives 178

Cinquième partie : Bibliographie

Références Bibliographiques 185

AVANT-PROPOS

Avant propos

Si les ARN les plus abondants dans les cellules ont été identifiés il y a plus de 50 ans, de nouvelles espèces sont encore caractérisées aujourd'hui. Parallèlement à l'enrichissement du catalogue des ARN cellulaires, des études visent à établir leur registre de fonction et de comprendre précisément leur mécanisme d'action. Il est à présent clairement établi que l'ARN participe à des mécanismes fondamentaux à la vie cellulaire tels que la traduction, l'épissage, ou le maintien de l'intégrité de l'extrémité des chromosomes. Ainsi, l'ARN de transfert (ARNt) est la molécule nécessaire au décodage des ARN messagers (ARNm) dans la traduction. L'ARNt est utilisé comme adaptateur au sein du ribosome pour apporter l'acide aminé correspondant à celui « inscrit » dans la séquence de l'ARNm puis de servir de point d'ancrage au peptide naissant. Le mécanisme d'épissage des pré-ARNm repose quant à lui sur l'appariement entre des petits ARN nucléaires (snRNA) avec certaines régions conservées contenues dans les introns du pré-ARNm et qui ne sont pas retrouvées dans l'ARNm mature. Les snRNA, assistés par des protéines, forment sur le pré-ARNm un macrocomplexe appelé le spliceosome, au sein duquel sont définies les positions exactes des introns à exciser. Enfin, certains petits ARN nucléolaires (snoRNA) s'apparient avec le précurseur des ARN ribosomiques (ARNr) pour remanier localement la structure du pré-ARNr et aboutir à des clivages qui sont essentiels à la maturation des ARNr. La télomérase, qui permet d'ajouter des séquences d'ADN aux séquences répétées des télomères à l'extrémité des chromosomes, est composée d'un ARN et de protéines. La réaction de synthèse des télomères dépend d'une étroite collaboration entre les protéines et l'ARN pour faire de la télomérase un complexe ARN-protéines catalytiquement efficace.

Il est actuellement admis que le monde vivant, tel que nous le connaissons aujourd'hui, ait été précédé par un monde primordial où l'ARN jouait à la fois le rôle de matériel génétique et celui de catalyseur (Gilbert, 1986). Cette théorie repose essentiellement sur l'argumentaire suivant :

- l'ARN possède de multiples potentialités structurales et fonctionnelles.

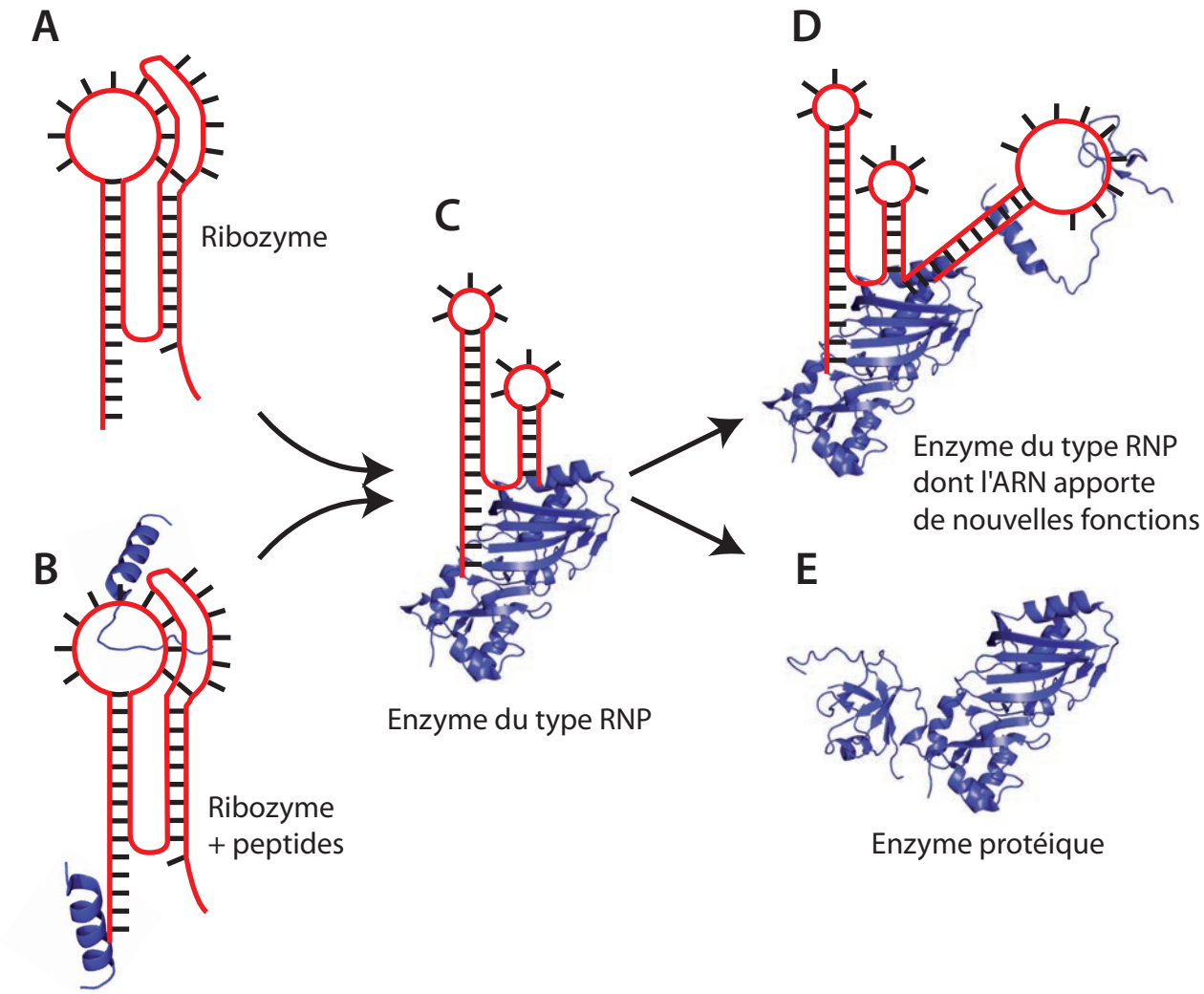


Figure i1 : Modèle de l'évolution de la biocatalyse. (Adapté de Cech, 2009).
A. La catalyse employée par les ribozymes primordiaux est uniquement basée sur l'emploi d'ARN (le squelette ribose phosphate est représenté en rouge et les bases en noir).
B. La présence de divers peptides (indiqués en bleu) dans l'environnement confère une valeur sélective aux ribozymes qui a pu conduire la formation de RNP.
C. L'enzyme de type RNP est un état avancé de la catalyse biologique dans lequel la synthèse des polypeptides est réalisée de façon reproductible, ce qui augmente la stabilité et la fonction des RNP, en y incluant le système le plus abouti qui correspond à la machinerie ribosomique de traduction des ARN messagers.
D. Les enzymes du type RNP ont continué à évoluer, par exemple des sous-unités protéiques ou des fragments d'ARN ont été ajoutés puis conservés dans le but d'apporter plus de spécificité à une RNP donnée. C'est à cette catégorie qu'appartiennent les snoRNP à boîtes C/D et H/ACA.
E. Dans la plupart des cas, ce sont les enzymes purement protéiques qui possèdent le rôle catalytique telle que l'ARN:pseudouridine-synthase TruB d'*E. coli*.

- l'ARN peut porter toute une variété d'activités catalytiques comme l'atteste la découverte des ribozymes (Zaug et Cech, 1986), la sélection *in vitro* parmi des séquences aléatoire d'ARN à activité ligase ou ADN polymérase (pour revues, (Joyce, 2002, 2007)), (Robertson et Scott, 2007) et la découverte que la biosynthèse des protéines dans le ribosome est catalysée par l'ARN (et non par des protéines) (Ban *et al.*, 2000).

D'un point de vue chronologique, le monde ARN primordial aurait d'abord été remplacé par un monde bipolaire ARN-protéine lors de l'apparition de la machinerie de transcription préfigurant notre monde tripolaire ADN-ARN-protéines (pour revue, (Freeland *et al.*, 1999)). Cette proposition est basée sur le constat que dans le monde biologique contemporain, une grande diversité d'espèces ARN est impliquée dans le métabolisme général de l'ARN ainsi que dans la transcription, ce qui contraste avec la très faible implication de l'ARN dans la synthèse de l'ADN. De plus, la synthèse dans tous les organismes actuels des désoxyribonucléotides par réduction des ribonucléotides, et de la thymine par méthylation de l'uracyle, sont deux arguments supplémentaires en faveur d'une apparition de l'ADN postérieure à celle de l'ARN (pour revue, (Freeland *et al.*, 1999)).

Aujourd'hui, des fonctions essentielles sont assurées par des complexes ARN-protéines, baptisés particules ribonucléoprotéiques (RNP). Un schéma évolutif du ribozyme primordial vers les complexes ARN-protéine contemporains est proposé (pour revue, (Cech, 2009)) :

- Dans le monde ARN primordial, les ribozymes primitifs ont pu s'associer à des peptides environnants qui leur ont permis d'étendre leur registre structural et leur répertoire fonctionnel (Fig. i1 A et B).

- Ces ribozymes associés aux peptides auraient ainsi évolué en RNP dont l'activité enzymatique est portée soit par l'ARN, soit par une protéine associée (Fig. i1 C).

- Dans certains cas, les RNP pourraient évoluer indépendamment par ajout d'éléments extérieurs du type ARN qui contribuent directement à la fonction ou par apport de nouvelles protéines (ou domaines protéiques) dans le complexe pour aboutir aux RNP contemporaines (Fig. i1 D).

- A d'autres occasions, l'avantage conféré par l'ARN dans la catalyse est insuffisant, et ce sont les protéines enzymatiques qui ont conservé la fonction pour conduire aux enzymes actuelles (Fig. i1 E) (pour revue, (Cech, 2009)).

Lors de ce travail je me suis intéressé à une classe de RNP présentes chez les archaea et les eucaryotes, les RNP à boîtes H/ACA. Celles-ci permettent la conversion post-transcriptionnelle de résidus uridines en pseudouridines dans les ARN. L'activité catalytique ARN:pseudouridine-synthase est portée par une protéine, invariablement présente avec trois autres protéines au sein de chaque particule, mais dont le substrat est défini par le constituant ARN propre à chaque RNP. Ainsi, dans chaque particule, un ARN sert de guide pour l'enzyme protéique en s'appariant avec une séquence ARN cible. Cependant, chez les archaea, l'enzyme protéique de ces RNP possède une dualité fonctionnelle puisqu'elle est capable de modifier certains ARN cibles sans l'aide de l'ARN guide. L'étude chez ces organismes des particules RNP à boîtes H/ACA et de la pseudouridine-synthase qui leur est associée pouvait apporter des informations sur l'évolution des RNP et du rôle fonctionnel des ARN.

Au début de mon travail de thèse, Bruno Charpentier au Laboratoire et l'équipe de M. et R. Terns venaient de mettre au point un système de reconstitution *in vitro* de particules actives sRNP à boîtes H/ACA d'archaea. Les structures 3D de la plupart des ARN:pseudouridine-synthases bactériennes étaient connues. L'ensemble de ces données structurales et la disposition d'un outil pour l'étude fonctionnelle *in vitro* constituaient les bases sur lesquelles nous pouvions débuter un travail d'analyse

de la relation entre la structure et la fonction de ces particules. Par la suite et au cours de ma thèse, plusieurs équipes dont la nôtre, ont établi les structures 3D des protéines libres de ces particules puis de plusieurs complexes protéines-ARN.

Le chapitre d'introduction comporte quatre parties en lien avec mon travail de thèse. La première partie expose tout d'abord l'intérêt de l'étude des systèmes provenant des archaea et présente brièvement l'organisme sur lequel ont porté mes travaux de thèse, à savoir *Pyrococcus abyssi*. Puis le second point est dédié au nucléotide modifié pseudouridine (Ψ) et à ses multiples rôles dans les ARN. Le troisième thème aborde ensuite les enzymes de modification, en particulier celles qui permettent l'isomérisation des résidus uridine en pseudouridine. Enfin, le dernier chapitre de l'introduction fera l'état des connaissances sur les deux systèmes de modification qui reposent sur un ARN guide, à savoir les s/snoRNP à boîtes C/D et ceux à boîtes H/ACA. Cette partie abordera également les connaissances sur la télomérase et sur la maladie humaine appelée la Dyskératose Congénitale.

INTRODUCTION

I. Les Archaea : un intérêt scientifique grandissant pour ce domaine du vivant

A. Intérêt d'étudier les Archaea

Le monde du vivant se décompose en 3 domaines (Fig. i2). Les Archaea sont des organismes procaryotes et constituent l'un de ces domaines du vivant. L'organisation génétique et certains de leurs mécanismes moléculaires présentent des caractéristiques similaires à celles employées chez les Bactéries. Cependant, la plupart des macrocomplexes accomplissant les grandes fonctions cellulaires sont comparables à ceux adoptées chez les Eucaryotes, rendant les recherches menées sur ces organismes particulièrement enrichissantes. Il existe quatre grandes familles représentées au sein de ces organismes : les Euryotes, les Crénotes, les Nanoarchaeotes et les Korachaeotes (Fig. i2) (pour revue, (Allers et Mevarech, 2005)).

1. Des caractéristiques partagées avec celles des bactéries et des eucaryotes

Les Archaea sont des organismes unicellulaires dépourvus de noyau dont la morphologie et la taille sont comparables à celles des Bactéries. Ils possèdent certaines caractéristiques originales et propres à ce règne du vivant. Leur membrane est généralement constituée d'une couche unique de lipides qui sont composés de chaînes isoprénoïdes attachées par des liaisons de type éther à des molécules de glycérol (pour revue, (van de Vossenberg *et al.*, 1998)). Ces constituants renforcent la résistance des membranes plasmiques vis-à-vis de la dégradation, de la température élevée, de la forte concentration saline et du pH. Ces propriétés rendent les membranes lipidiques d'archaea plus favorables à la vie et à la survie dans des environnements extrêmes contrairement à la bicouche lipidique de type ester des bactéries et des eucaryotes (pour revue, (van de Vossenberg *et al.*, 1998)).

L'organisation chromosomique rappelle celle présente chez la majorité des bactéries avec une seule origine de réplication par chromosome mais avec des protéines de la machinerie de réplication qui partagent plutôt des similitudes avec leurs homologues eucaryotes (pour revue, (Grabowski et Kelman, 2003)). L'ADN

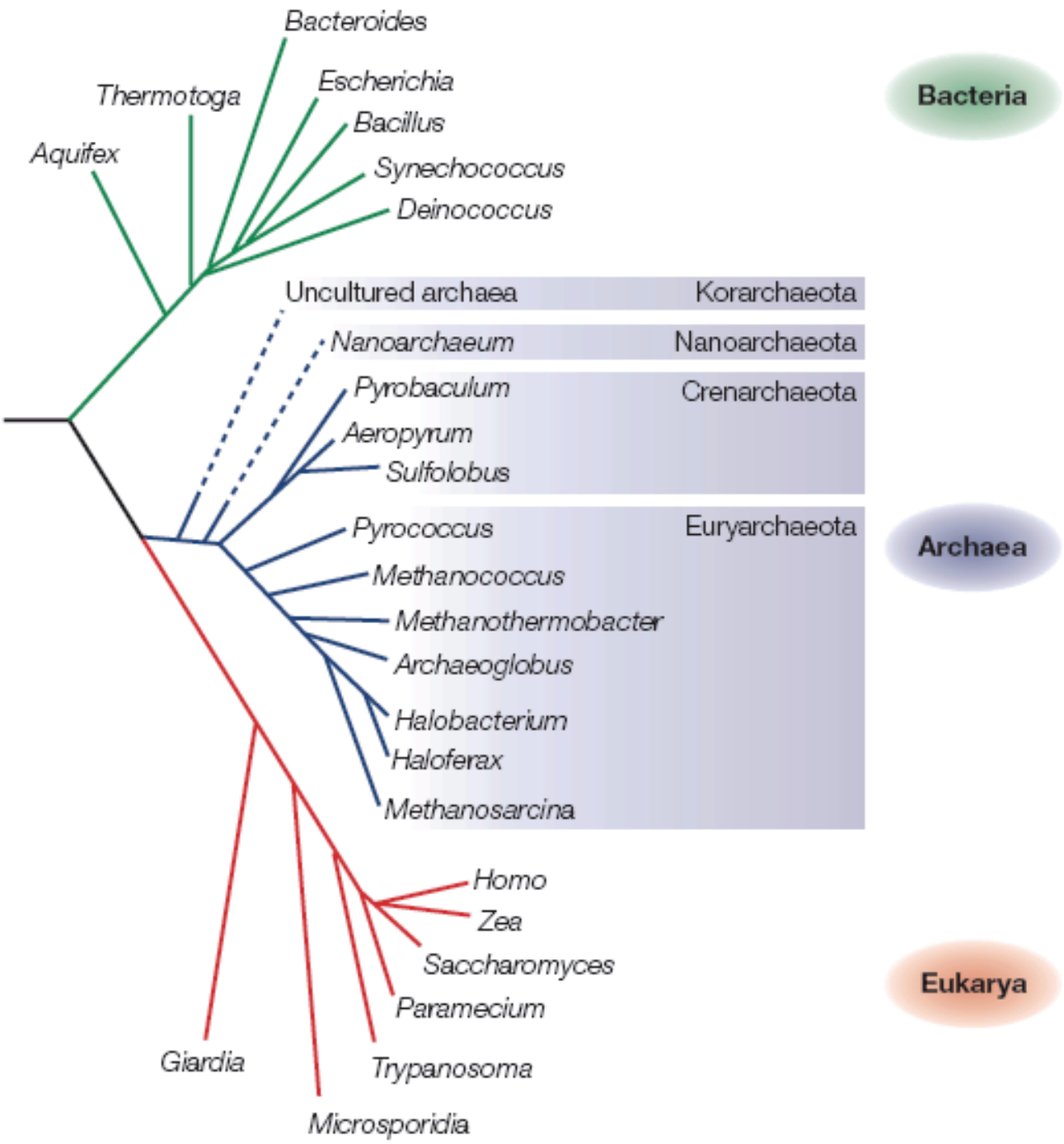


Figure i2 : Arbre phylogénétique universel. Adapté des travaux de Carl Woese et collaborateurs dans la revue (Allers et Mevarech, 2005).
L'arbre est basé sur la comparaison des séquences des ARNr de la petite sous-unité ribosomique. Il décrit une division du vivant en trois domaines primaires : Bactéries, Archaea et Eucaryotes. Le domaine Archaea est divisé en deux domaines principaux, les euryotes (euryarchaeota) et les crénotes (crenarchaeota). L'existence de phyla supplémentaires aux deux précédents n'est pas encore clairement établie.

chromosomique circulaire est compacté différemment selon la famille d'archaea examinée. Chez les euryotes, l'ADN génomique contient des histones qui ont une structure similaire mais une séquence raccourcie par rapport à leurs homologues présents chez les eucaryotes ; tandis que chez les crénotes les protéines sont de type histone-like mais elles sont complètement différentes des histones eucaryotes ou euryotes (pour revue, (White et Bell, 2002)).

La transcription repose sur l'emploi de l'ARN polymérase II proche de celle qui est présente chez les eucaryotes et sur des facteurs qui reconnaissent la boîte TATA. Néanmoins, la présence chez les archaea de protéines homologues aux facteurs de régulation de la transcription bactériens indique que celle-ci reposerait sur les principes de la régulation bactérienne (pour revue, (Bell et Jackson, 2001)).

De nombreux facteurs impliqués dans l'initiation de la traduction sont similaires à ceux présent chez les eucaryotes, ce qui indique que le mécanisme de traduction serait de nature eucaryotique. De plus, l'initiation de la traduction chez les archaea et les eucaryotes utilise une méthionine tandis que chez les bactéries c'est une formyl méthionine qui est employée. Néanmoins, seuls les archaea et les bactéries produisent des ARNm polycistroniques et la reconnaissance de l'ARNm par le ribosome se fait par appariement de l'ARN 23S avec une séquence du type Shine-Dalgarno présente dans la région 5' non traduite, quelques résidus en amont du codon d'initiation. De plus comme chez les bactéries, la traduction chez les archaea est couplée à la transcription (pour revue, (Dennis, 1997)).

Les gènes codant les enzymes impliquées dans les fonctions du métabolisme central, la conversion d'énergie et la biosynthèse, ont une origine bactérienne, tandis que leur mode de régulation est plutôt du type eucaryote (Rivera *et al.*, 1998).

2. Une grande diversité d'espèces

Regroupant une grande variété d'espèces, les archaea occupent une large diversité de niches écologiques, notamment celles caractérisées par des conditions de

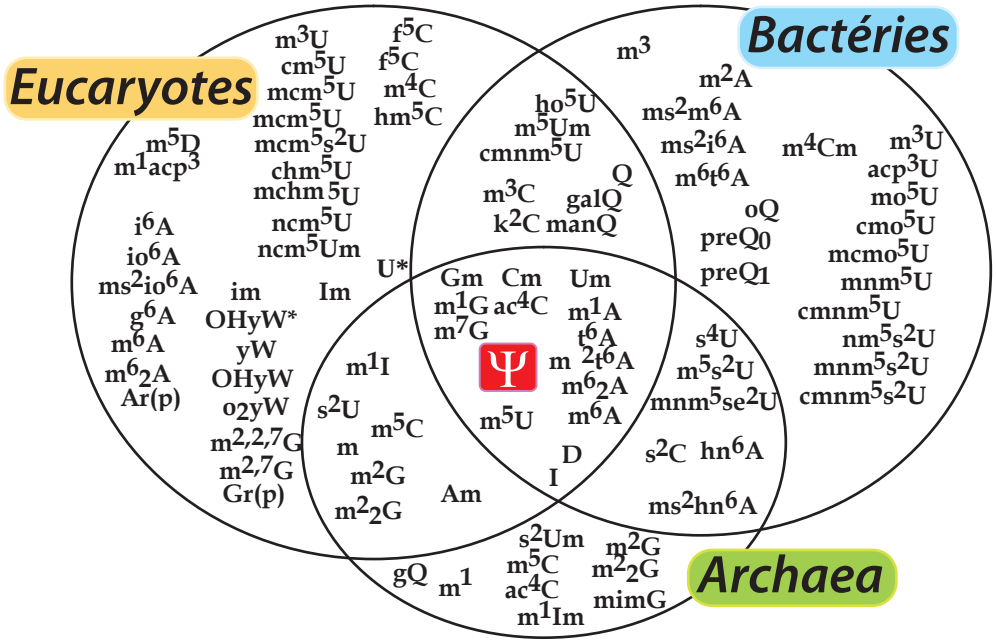
vie extrêmes telles que la vie dans les marais salins, les profondeurs océaniques, les mines de fer, les zones arides et acides, les réservoirs de pétrole, les lacs gelés de l'antarctique ou encore les zones volcaniques. Les organismes hyper-thermophiles sont naturellement résistants face aux hautes températures (pour revue, (Stetter, 2006)). Certaines protéines produites par ces organismes ont un intérêt biotechnologique comme par exemple la technologie de l'amplification d'ADN par PCR qui repose essentiellement sur les propriétés de thermostabilité de l'ADN polymérase employée comme par exemple la *Pfu* provenant de l'archae *Pyrococcus furiosus*. De plus, il est possible de produire et de purifier les protéines de ces organismes avec une facilité accrue par rapport à certaines protéines eucaryotes, ce qui permet d'envisager de réaliser des études *in vitro* dans un but expérimental ou structural. A l'heure actuelle, la contrepartie de la facilité relative des études *in vitro* des systèmes d'archaea hyperthermophiles est la difficulté rencontrée dans la mise en œuvre du travail *in vivo*, en particulier de génétique moléculaire. En effet, les conditions de culture et de sélection sont plus compliquées ce qui pousse à envisager ce type de recherche dans des organismes mésophiles de ce domaine du vivant.

Néanmoins, la cristallisation des protéines ou des complexes ARN-protéines d'archaea permet d'obtenir des informations précieuses nous permettant de mieux comprendre et d'appréhender les structures des complexes analogues trouvés chez les eucaryotes. La première structure cristallographique du ribosome est celle de l'archaea d'*Haloarcula marismortui* qui a permis de faire des avancées majeures dans la compréhension des mécanismes liés à la traduction des ARNm en protéines (Ban *et al.*, 2000). C'est pour ces nombreux avantages que beaucoup d'équipes employant la radiocristallographie ou la RMN s'intéressent particulièrement aux composants produits chez les archaea hyperthermophiles.

B. *Pyrococcus abyssi*

Cette partie présente l'organisme sur lequel ont porté mes études réalisées au laboratoire, il s'agit de l'archaea *Pyrococcus abyssi*. Avec une température optimale de

A



B

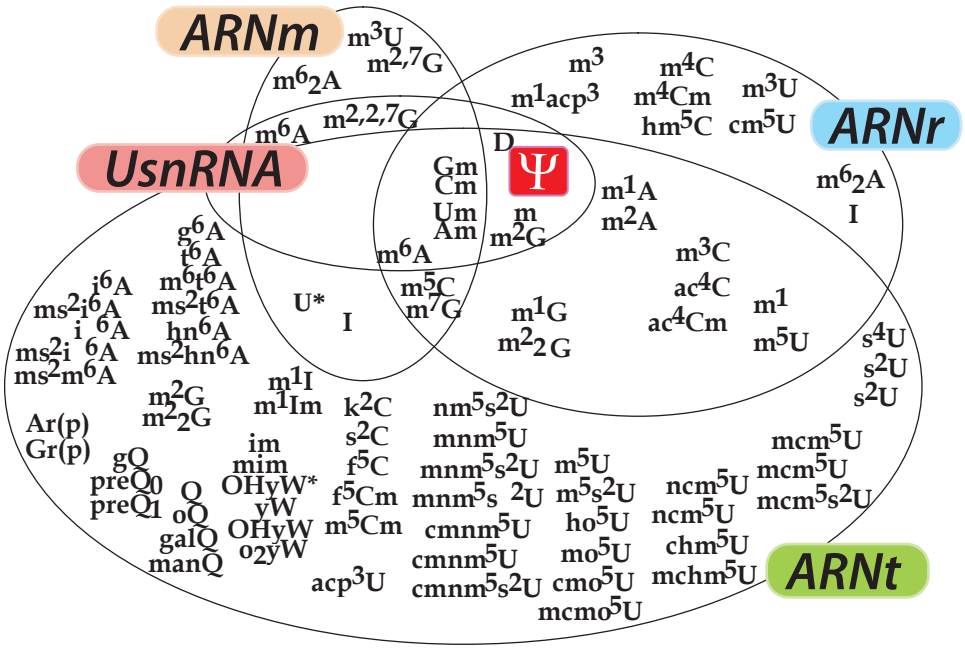


Figure i3 : Distribution des nucléotides modifiés dans les trois domaines du vivant (A) et par type d'ARN (B). (D'après la revue, Motorin et Grosjean, 1998).
U* correspond à l'uridine formée par déamination d'une cytosine par un mécanisme d'édition post-transcriptionnel. Signification des abréviations : ac (acéthylation), c (carboxylation), f (formylation), g (glycinylation), gal (galactosylation), h (hydroxylation), i (isopentanylation), k (lysidinylation), m (méthylation), man (mannosylation), n (norvalinylation), o (oxyacéthylation), p (propylation), r(p) (2'-*O*-ribosyl phosphorylation), s (thiolation), t (thréonylation) ; A (adénine), C (cytosine), D (dihydouridine), G (guanine), I (inosine), im (wyinosine), oQ (epoxyqueuosine), Q (queuosine), T (thymine), U (uracyle), yW (wybutosine) et Ψ (pseudouridine).

croissance égale à 96°C, *P. abyssi* appartient aux archaea hyperthermophiles. En effet, il a été découvert en 1989 par une équipe d'océanologues franco-japonais dans des cheminées volcaniques sous-marines abyssales (à 2000 mètres de profondeur), d'où son nom d'espèce *abyssi*, dans le bassin nord Fidjien de l'océan pacifique (Auzende et Urabe, 1989). C'est un organisme hétérotrophe (nécessité de se procurer d'autres sources de nourriture que par soi-même) et anaérobie (vie sans oxygène) qui fait partie de l'ordre des Thermococcales. Il est considéré comme un organisme du type gram négatif qui peut se déplacer et vivre à de très fortes pressions atmosphériques (20 MPa). Son génome est composé d'un chromosome unique et circulaire de 1,77 Mpb et contient 1788 gènes annotés (Cohen *et al.*, 2003). Son temps de doublement est de 33 minutes dans les conditions optimales de croissance. Les conditions de température particulières pour la croissance de *P. abyssi*, comparées à celles d'une bactérie par exemple, expliquent pourquoi nos tests *in vitro* ont été réalisés à une température élevée.

II. Les nucléotides modifiés dans les ARN : les résidus pseudouridines

En plus des quatre nucléotides A, C, G et U, les ARN peuvent contenir divers nucléotides modifiés, présentant des modifications chimiques sur la base ou le ribose. Ces nucléotides modifiés sont répartis dans l'ensemble des trois domaines du vivant (Fig i3 A). Une centaine de résidus modifiés différents ont été identifiés dans les ARN non codants : ARNr, ARNt, ARNm, UsnRNA, ARNtm (Fig. i3 B) (pour revues, (Limbach *et al.*, 1994; Motorin et Grosjean, 1998)). Ces nucléotides modifiés sont synthétisés de façon post-transcriptionnelle. Parmi toutes les classes d'ARN, c'est dans l'ARNt que l'on retrouve le plus grand nombre de modifications. En effet, jusqu'à 26% des résidus d'un ARNt peuvent être modifiés. A l'inverse, dans l'ARNm moins de 1% des résidus ont été recensés comme étant modifiés (Rozenski *et al.*, 1999). Certaines modifications peuvent être réalisées par une seule enzyme (pour la méthylation ou la pseudouridylation) alors que d'autres plus complexes nécessitent

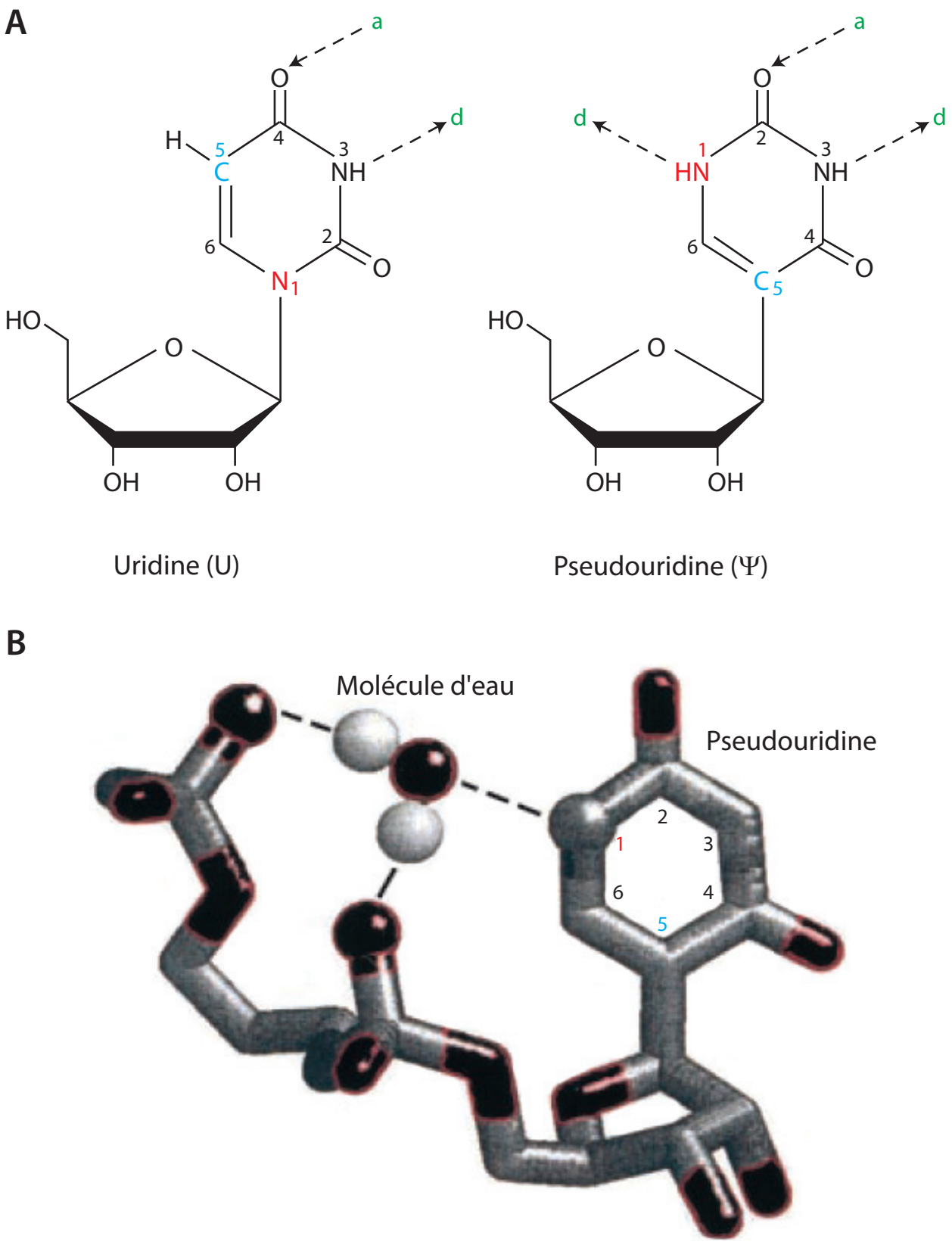


Figure i4 : Le résidu pseudouridine (Ψ). (Adapté de Charette et Gray, 2000).
A. Comparaison des structures des résidus uridine et pseudouridine. Les liaisons hydrogène possibles dans le sens accepteur (a) ou donneur (d) sont indiquées par des flèches.
B. Vue schématique d'une liaison hydrogène médiée par une molécule d'eau entre le groupement NH1 de la Ψ et les atomes d'oxygènes appartenant à son propre groupement phosphate et à celui du nucléotide situé en 5'. Ces interactions permettent de rigidifier le squelette sucre-phosphate et d'augmenter la capacité d'interaction par empilement de base. La base en 5' n'est pas représentée. Les atomes d'oxygènes sont colorés en rouge.

une cascade de réactions enzymatiques (pour la synthèse de la queuosine ou de l'archaeosine) (Slany *et al.*, 1993; Van Lanen *et al.*, 2005).

Dans un génome donné, une proportion importante de gènes est dédiée à la modification des ARN dont les gènes codant les enzymes permettant la pseudouridylation. La conservation de certaines modifications, tant au niveau de l'identité qu'au niveau de la position des nucléotides modifiés au cours de l'évolution, suggère un rôle essentiel de ces modifications dans le fonctionnement cellulaire (pour revue, (Björk, 1995)). Il arrive que certaines mutations conduisent à des erreurs de modification qui entraînent l'apparition de maladies appartenant à la famille des myopathies mitochondriales qui parfois peuvent être fatales chez l'homme. Par exemple, une mutation dans les $ARNt^{leu}$ ou $ARNt^{lys}$ mitochondriaux conduit à la non-addition d'un groupement taurine(τ) sur l'uridine en position 34 de l'anti-codon (Yasukawa *et al.*, 2000a; Yasukawa *et al.*, 2000b; Suzuki *et al.*, 2002).

Le ribonucléotide modifié le plus fréquemment retrouvé dans les ARN est le résidu pseudouridine (5-β-D ribofuranosyluracile, Ψ) qui est un isomère de l'uracile (1-β-D ribofuranosyluracile, U). Mon travail de thèse ayant porté sur la machinerie assurant la conversion d'une uridine en pseudouridine, seules les propriétés de ce résidu sont développées dans les chapitres qui suivent.

A. Propriétés des résidus pseudouridines (Ψ)

La pseudouridine est, avec les nucléotides méthylés en position 2' OH, un des nucléotides modifiés les plus fréquemment rencontrés dans les ARN (Fig. i4 A). Elle a été détectée en tant que nouveau nucléotide dès 1957 et fut longtemps considérée comme le « cinquième ribonucléotide » (Davis et Allen, 1957). En tant qu'isomère de l'uridine, la pseudouridine possède donc la même masse ce qui la rend indiscernable de l'uridine par spectrométrie de masse. Par contre, elle se distingue de l'uridine en possédant des propriétés chimiques et chromatographiques différentes (Cohn, 1960). L'emploi de diverses méthodologies analytiques révélant ces propriétés a ainsi permis de caractériser et de déterminer la structure de ce résidu (Fig. i4 A) (Cohn,

Organisme	Nombre de mole de Ψ /mole d'ARNr			
	Petite sous-unité	Grande sous-unité	Analyse	Références
ARNr cytoplasmique				
E. coli	1	9	Seq. (100)	Bakin *et al.*, 1994
B. subtilis	1	5	Seq. (40)	Niu et Ofengand, 1999
H. halobium	0	4	Seq. (70)	Bakin et Ofengand, 1995
S. solfataricus	5	4	Total	Noon *et al.*, 1998
S. acidocaldarius	nd	6	Seq. (?)	Massenet *et al.*, 1999
P. abyssi	**3**	**12**	**Seq. (25)**	**Muller *et al.*, 2008**
A. fulgidus	1	5	Seq. (?)	Tang *et al.*, 2002
H. marismortui	nd	3	Seq. et Total	Kirpekar *et al.*, 2005
S. cerevisiae	14	30	Seq. (99)	Bakin *et al.*, 1994
D. melanogaster	nd	57	Seq. (59)	Ofengand et Bakin, 1997
X. laevis	44	~52	Total	Maden, 1990
M. musculus	~36	~57	Seq. (~50)	Ofengand et Bakin, 1997
H. sapiens	~36	~57	Seq. (~50)	Ofengand et Bakin, 1997
ARNr mitochondrial				
S. cerevisiae	nd	1	Seq. (99)	Bakin *et al.*, 1994
M. musculus	nd	1	Seq. (17)	Ofengand et Bakin, 1997
H. sapiens	nd	1	Seq. (41)	Ofengand et Bakin, 1997
T. brucei	nd	6	Seq. (98)	Ofengand et Bakin, 1997
ARNr chloroplastique				
Z. mays	nd	4	Seq. (29)	Ofengand et Bakin, 1997
ARNr 5.8S				
S. cerevisiae	1		Seq. (100)	Rubin, 1973
M. musculus	2		Seq. (100)	Nazar *et al.*, 1976
H. sapiens	2		Seq. (100)	Khan *et al.*, 1977

Tableau i1 : Nombre de résidus Ψ présents dans les ARNr des petites et grandes sous-unités des ribosomes cytoplasmiques, mitochondriaux et chloroplastiques ainsi que dans les ARNr 5S de différents organismes provenant des trois domaines du vivant.

Les organismes appartenant aux archaea sont surlignés en vert. L'analyse des ARNr a été réalisée soit par séquençage de l'ARN (Seq.) et détermination précise de la position (méthode au CMCT), soit par l'étude de la composition totale en nucléotides modifiés reposant sur la spectrométrie de masse (Total). Le nombre indiqué entre parenthèse à côté de "Seq." traduit le pourcentage de la séquence des ARNr analysée. "nd" signifie non déterminé.

1960). Ses propriétés de migration chromatographique varient par rapport à celles observées pour l'uridine, ce qui permet ainsi de séparer aisément ces deux résidus (Keith, 1995) (voir la partie Matériels et Méthodes, Fig. m1).

Une pseudouridine est formée par isomérisation post-transcriptionnelle d'un résidu uracile (Johnson et Soll, 1970) et cela sans apport d'énergie ou de co-facteur (Arena *et al.*, 1978). La structure chimique du résidu Ψ autorise la formation de 3 liaisons hydrogènes au lieu de 2 pour le résidu U (Fig. i4A). Le site donneur d'hydrogène supplémentaire provient du groupement imino (-NH-) qui est libre dans le cas d'une pseudouridine et qui est relié au ribose dans le cas de l'uracile.

Les résidus Ψ peuvent jouer un rôle dans la stabilisation des interactions ARN-ARN ou ARN/protéines. En effet, des études menées par RMN ont montré que le groupement imino supplémentaire de la Ψ stabilisait l'interaction entre hélices d'ARN via une molécule d'eau. Ceci favoriserait la formation d'hélices ARN à la structure unique, qui ne pourrait être obtenue en présence d'une uridine (Newby et Greenbaum, 2002a). Par rapport à un résidu U, la pseudouridine confère également la propriété d'augmenter la stabilité thermique de duplexes ARN-ARN le contenant (Hall et McLaughlin, 1991; Newby et Greenbaum, 2001). Il a aussi été montré par simulations de dynamique moléculaire, que la pseudouridine diminuait la mobilité de la base voisine en rigidifiant le squelette ribose-phosphate (Auffinger et Westhof, 1996). Cela peut s'effectuer grâce à l'intervention du groupement imino libre de la Ψ qui coordine une molécule d'eau qui elle-même interagit avec le phosphate du nucléotide (Fig. i4B) (Charette et Gray, 2000). L'ensemble de ces propriétés singulières du résidu Ψ pourraient expliquer sa prévalence dans les structures ARN.

B. Localisation et rôles des pseudouridines dans les ARN

1. Les pseudouridines dans les ARNr

Les ARN ribosomiques contiennent un grand nombre de résidus pseudouridines s'élevant de 1 à 3% des résidus totaux de l'ARNr (pour revue, (Motorin et Grosjean, 1998)). Le nombre de résidus Ψ dans les ARNr est également

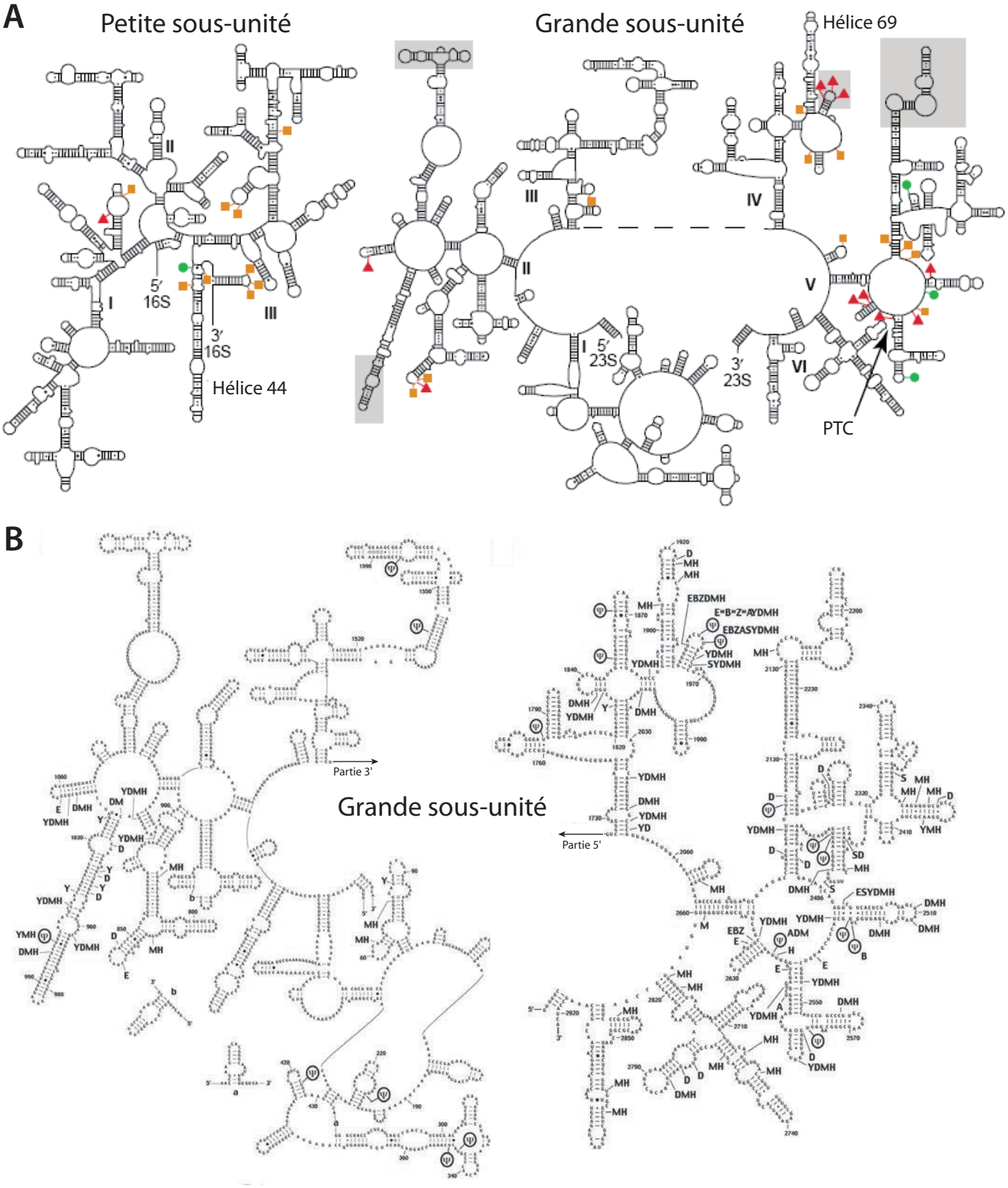


Figure i5 : Distribution des nucléotides modifiés dans les ARNr d'*E. coli* et de *H. marismortui*. (D'après Decatur *et al.*, 2002 ; Ofengand *et al.*, 2002).

A. Structures secondaires des ARNr 16S et 23S. Les résidus Ψ sont représentés par des triangles rouges, les 2'-*O*-méthylations par des cercles verts et les méthylations de bases (mN) par des carrés oranges. Les motifs ARN grisés correspondent aux parties non visibles dans le cristal.

B. La structure secondaire et la numérotation de l'ARNr 23S de *H. marismortui* sont utilisées (voir l'adresse suivante : http://www.rna.ccbb.utexas.edu/). Les positions des résidus pseudouridines sont indiqués chez *H. marismortui* (Ψ), *E. coli* (E), *B. subtilis* (B), *Z. mays* (Z), *H. halobium* (A), *S. acidocaldarius* (S), *S. cerevisiae* (Y), *D. melanogaster* (D), *M. musculus* (M) et *H. sapiens* (H).

plus important dans les organismes les plus évolués sur le plan phylogénétique (Tab. i1). On peut souligner que le nombre de résidus pseudouridines dans les ARNr mitochondriaux et chloroplastiques est comparable à celui observé chez les bactéries et quelques uns ont également été détectés dans l'ARNr 5S (Tab. i1). La localisation des résidus modifiés dans la structure 3D des ribosomes indique qu'ils sont pour la plupart situés au niveau des zones fonctionnelles (Decatur et Fournier, 2002), et cela est valables pour les ARNr présents dans des organismes des 3 domaines du vivant : *E. coli* (Fig. i5 A), *H. marismortui* (Fig. i5 B) ou *S. cerevisiae* (Fig. i6) (Decatur et Fournier, 2002; Ofengand, 2002; Del Campo *et al.*, 2005). Les modifications sont présentes à l'intérieur de la structure et aux interfaces d'interaction entre les sous-unités du ribosome. Elles sont néanmoins absentes des zones de fixations des protéines ribosomiques telles que la surface externe et la périphérie du ribosome. L'absence de nucléotides modifiés dans ces régions indique que l'établissement de la plupart des interactions ARN-protéine ne dépendrait pas directement de ces modifications (Decatur et Fournier, 2002). Il semble que la présence de ces résidus ne soit pas indispensable à l'activité du ribosome mais pourraient constituer un avantage sélectif pour certaines espèces (Ofengand *et al.*, 2001).

1.1. Les modifications dans la petite sous-unité ribosomique

Les résidus Ψ de l'ARNr de la petite sous-unité sont distribués de façon assez homogène dans la structure 2D, mais dans la structure 3D, ils sont regroupés pour la plupart dans les sites A, P, E, les sites d'interaction entre sous-unités du ribosome et dans le tunnel de fixation de l'ARNm (Fig. i5 et i6) (Decatur et Fournier, 2002). Il a été proposé que la région du tunnel de l'ARNm contiendrait des modifications qui pourraient influencer la fixation de l'ARNm dans le ribosome et la vitesse de lecture par le ribosome (Schluenzen *et al.*, 2000). On peut souligner qu'un seul résidu Ψ a été localisé dans l'ARNr 16S de *E. coli*, les autres modifications sont des méthylations (Fig. i5).

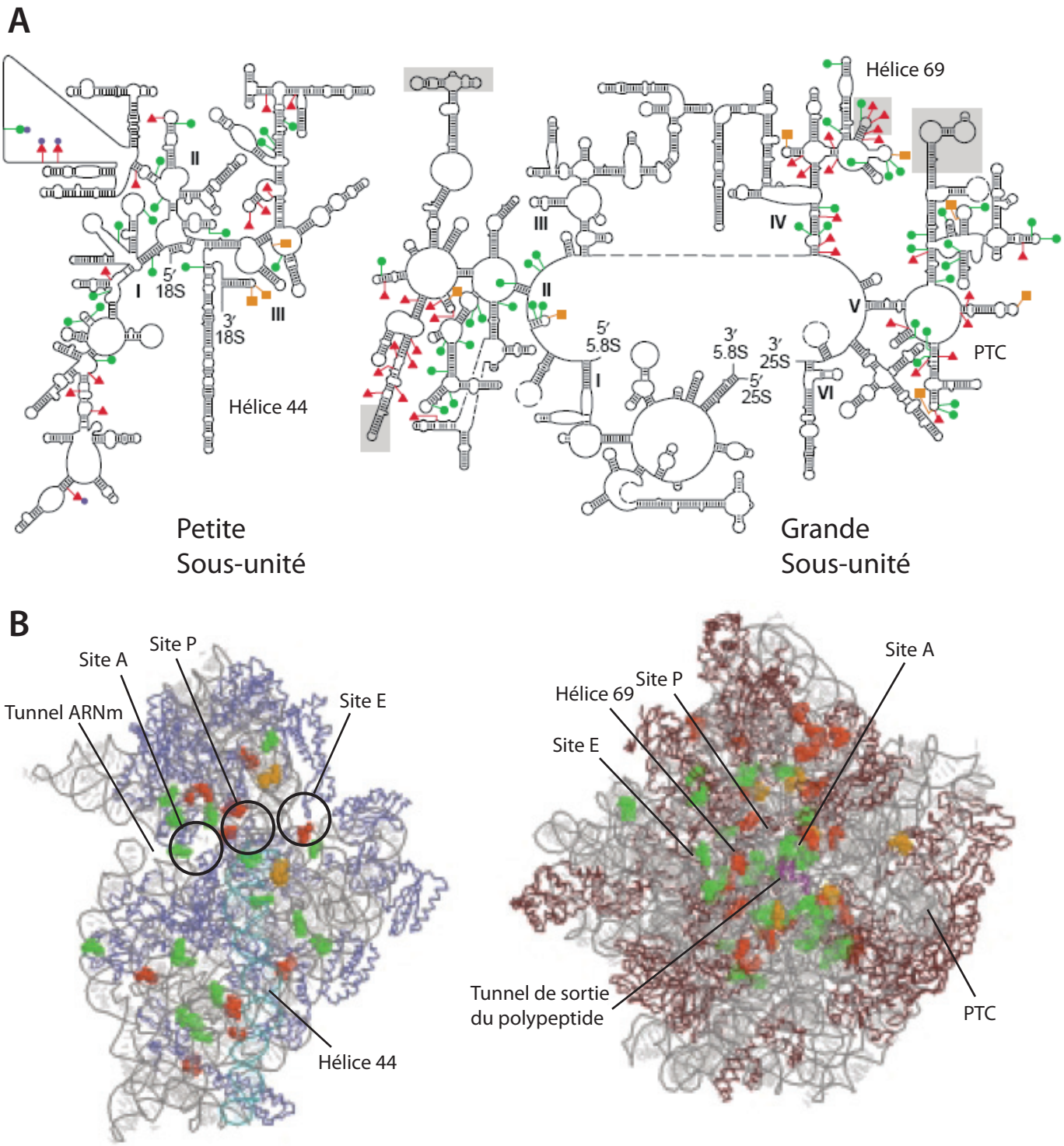


Figure i6 : Distribution des nucléotides modifiés dans les ARNr de *S. cerevisiae*. (D'après Decatur *et al.*, 2002).

A. Structures secondaires des ARNr 18S et 25S. Les résidus Ψ sont représentés par des triangles rouges, les 2'-*O*-méthylations par des cercles verts et les méthylations de bases (mN) par des carrés oranges. Les motifs ARN grisés correspondent aux parties non visibles dans le cristal.

B. Structures tridimentionnelles des petites et grandes sous-unités du ribosome de *S. cerevisiae* modélisées par analogie avec, respectivement, les structures 3D disponibles de *T. thermophilus* (PDB = 1FJF) et de *H. marismortui* (PDB = 1FFK et 1FFZ). Les nucléotides modifiés sont représentés selon le même code couleur qu'en **A**. Les chaînes protéiques sont respectivement indiquées en bleu et en magenta pour la petite et la grande sous-unité du ribosome.

1.2. Les modifications de la grande sous-unité ribosomique

Si l'on considère à présent l'ARNr de la grande sous-unité, on s'aperçoit que les modifications ne sont pas localisées au hasard mais sont concentrées principalement dans les régions fonctionnellement importantes, ceci aussi bien chez *E. coli* que chez *S. cerevisiae* (Fig. i5 et i6) (Branlant *et al.*, 1979; Branlant *et al.*, 1981b; Veldman *et al.*, 1981; Decatur et Fournier, 2002). Ces régions incluent les sites A, P, E, le centre peptidyl transférase (PTC), le tunnel de sortie du peptide en cours de synthèse et l'interface entre les deux sous-unités ribosomiques (Fig. i6) (Maguire et Zimmermann, 2001; Decatur et Fournier, 2002). En outre, de nombreuses positions de résidus Ψ sont conservées entre espèces ce qui suggère une importance structurale et/ou fonctionnelle de ces résidus au sein du ribosome (Fig. i5 B).

L'ARNr de la grande sous-unité se compose de plusieurs domaines fonctionnels (Fig. i7 A) (Ban *et al.*, 2000). Le domaine V de l'ARNr 23S constitue le centre peptidyl transférase (PTC). Celui-ci joue un rôle capital dans le transfert de l'acide aminé apporté par un ARNt sur la chaîne polypeptidique naissante. C'est dans cette région fonctionnellement importante du PTC que se regroupent de nombreux nucléotides modifiés (Fig. i7 B) (Veldman *et al.*, 1981; King *et al.*, 2003). Il a été montré chez la levure que les résidus Ψ du PTC étaient importants pour la structure et l'activité du ribosome (King *et al.*, 2003). En effet, la suppression de 6 résidus Ψ du PTC par déplétion des composants de la machinerie de synthèse, influencerait le repliement ou la dynamique de la conformation du domaine V (Fig. i7 C). Les changements de structure sembleraient se limiter à ce domaine. La structure du domaine VI, qui est en contact avec le domaine V, n'est pas modifiée ; néanmoins, la croissance de ces levures est tout de même ralentie (King *et al.*, 2003).

Le domaine IV de l'ARNr 23S fait partie du centre de décodage du ribosome (Fig. i7 A). Celui-ci participe à la translocation de l'ARNt et à la fidélité de la

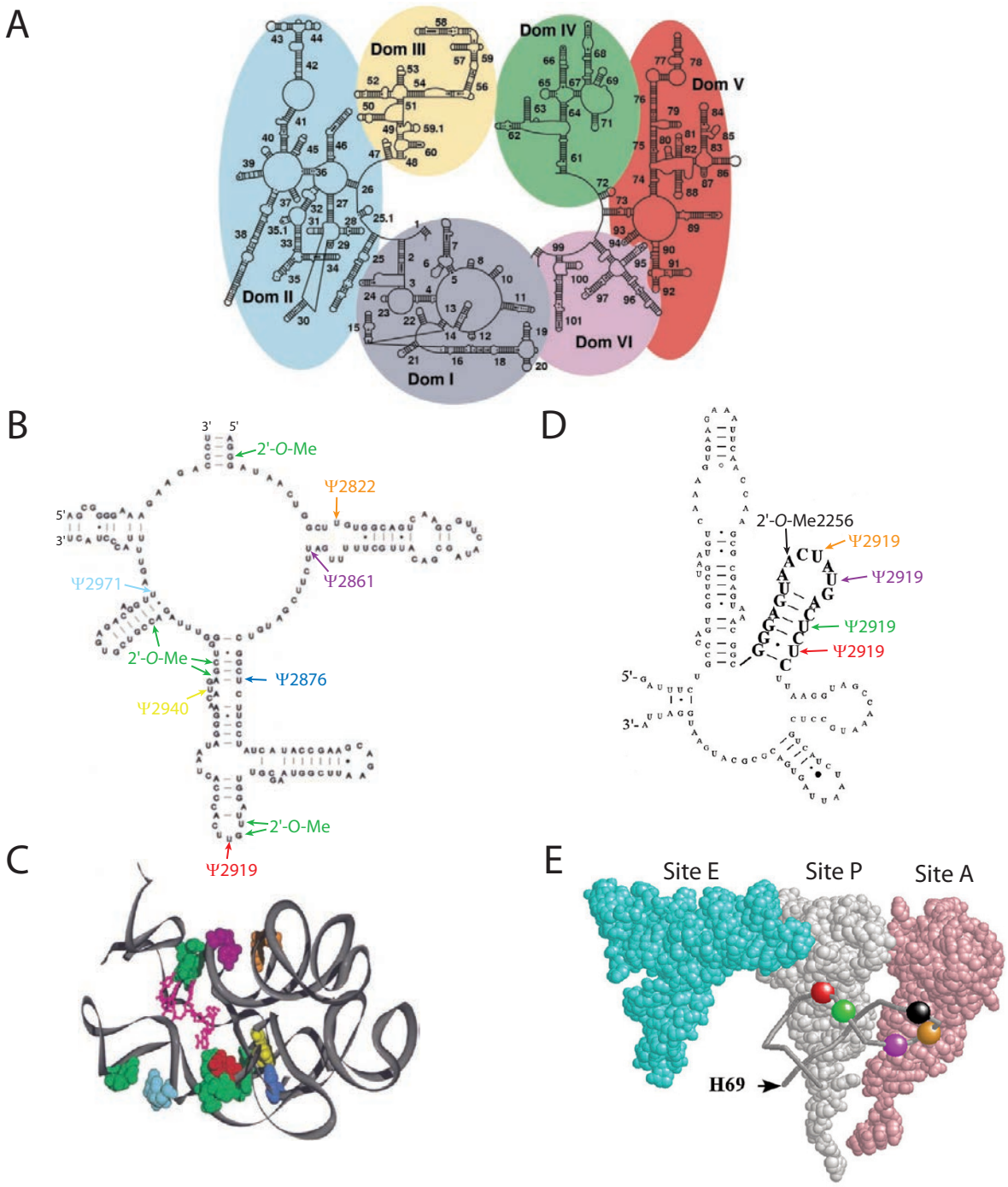


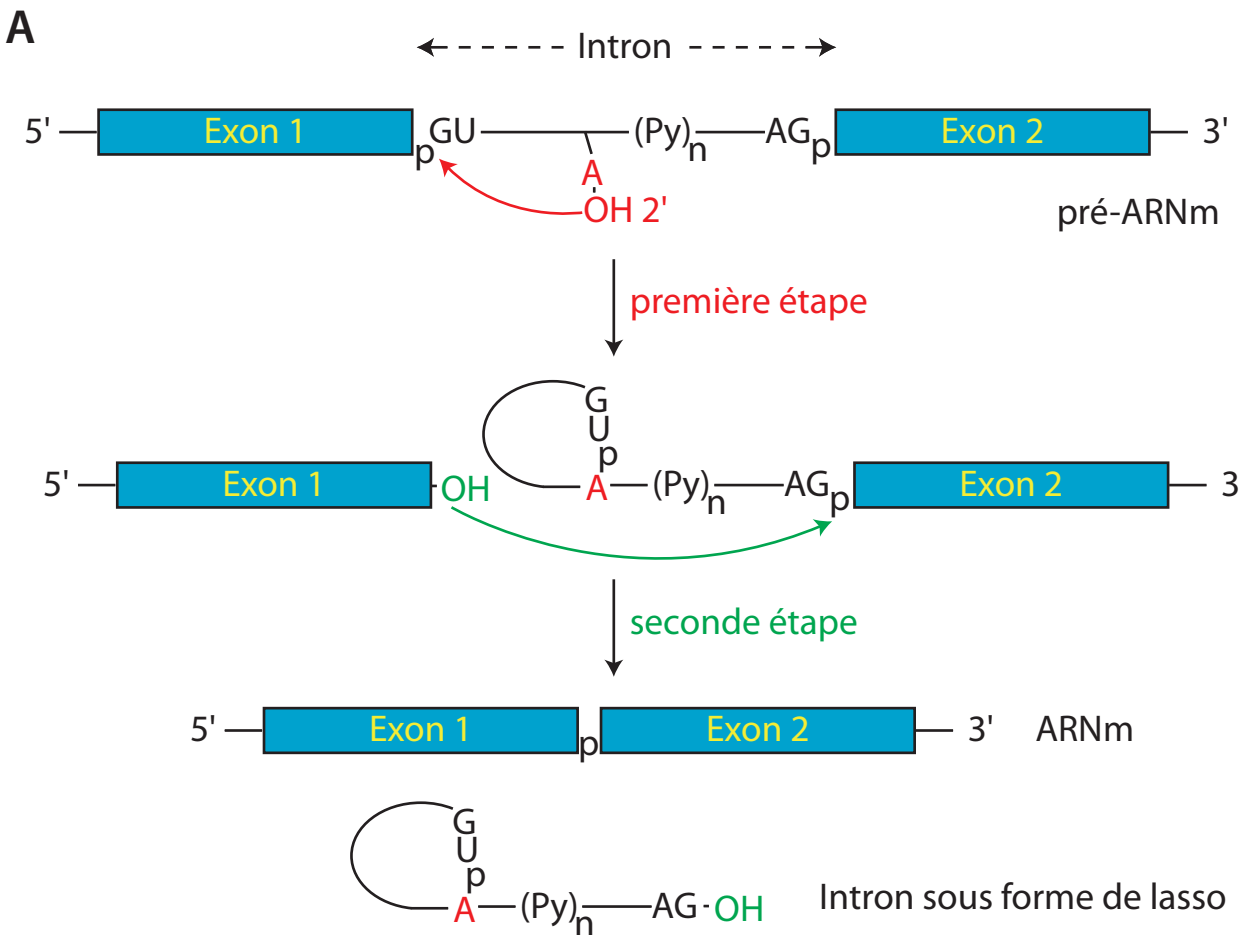
Figure i7 : Organisation de l'ARNr et localisation de certains nucléotides modifiés.
A. Structure secondaire de l'ARN ribosomique 23S d'*E. coli* déduite de la structure 3D de la grande sous-unité du ribosome obtenu par radiocristallographie. Les 6 domaines sont indiqués. (Ban *et al.,* 2000).
B. Structure secondaire de la région du centre peptidyl transférase (PTC) du domaine V de l'ARNr 25S de *S. cerevisiae*. Les résidus 2'-*O*-méthylés sont indiqués en vert et les 6 résidus Ψ sont représentés dans les autres couleurs : Ψ2822 en orange, Ψ2861 en violet, Ψ2876 en bleu, Ψ2919 en rouge, Ψ2940 en jaune et Ψ2971 en cyan. (King *et al.,* 2003).
C. Arrangement 3D prédit pour les nucléotides modifiés du PTC. Les modifications (Ψ et 2'-*O*-Me) de levure ont été colorées sur la structure du PTC de *Haloarcula marismortui*, en complexe avec la puromycine (magenta). Le code couleur des nucléotides modifiés est le même qu'en B.
D. Structure secondaire de l'hélice H69 du domaine IV de l'ARNr 25S de *S. cerevisiae*. Le résidu 2'-*O*-méthylé 2256 est indiqué en noir et les 4 résidus Ψ sont représentés dans les autres couleurs : Ψ2258 en jaune, Ψ2260 en bleu, Ψ2264 en vert et Ψ2266 en rouge. (Liang, X. H. *et al.,* 2007).
E. Interaction de l'hélice H69 avec le site A et P du ribosome. Les nucléotides modifiés sont indiqués dans la même couleur qu'en A.

traduction. La structure du ribosome lié à la fois à un ARNt et à un ARNm indique que l'ARNt est situé à l'interface des sous-unités 30S et 50S et qu'il est majoritairement fixé par l'ARNr (Yusupov *et al.*, 2001). L'hélice H69 de l'ARNr 23S forme avec l'hélice H44 de l'ARNr 16S un pont entre les sous-unités 30S et 50S du ribosome (Fig. i6). Il est intéressant de souligner que certains nucléotides modifiés sont également conservés dans cette région charnière située à l'interface des deux sous-unités ribosomiques (Ofengand *et al.*, 2001; Yusupov *et al.*, 2001). En effet, l'hélice H69 contient 5 nucléotides modifiés chez la levure (4 résidus Ψ et 1 résidu 2'-*O*-méthylé) (Fig. i7 D). Ces résidus sont importants pour la biogenèse du ribosome, son activité dans la traduction et pour la croissance cellulaire (Liang, X. H. *et al.*, 2007). Il a été montré que la déplétion d'au moins 3 résidus Ψ affectait la croissance des levures (Liang, X. H. *et al.*, 2007). Chez *E. coli*, l'hélice H69 contient 3 des 4 résidus Ψ et leur absence affecte également la croissance des bactéries (Ejby *et al.*, 2007). Dans le cas où les 5 nucléotides modifiés ont été supprimés, le taux d'ARNr mature chute de 50%, le taux de traduction est réduit par une plus faible assimilation des acides aminés dans les protéines naissantes (chute de 60%) et la lecture correcte des codons stop est très fortement affectée (seulement 30% des codons stops sont reconnus en tant que tels) (Ejby *et al.*, 2007; Liang, X. H. *et al.*, 2007). En absence de résidus Ψ et 2'-*O*-Me, la structure de l'ARNr est modifiée, ce qui indique que les modifications de ces nucléotides jouent un rôle important pour stabiliser la structure de l'ARNr (Fig. i7 E) (Liang, X. H. *et al.*, 2007).

La déplétion des résidus Ψ dans les ARNr n'est pas létale pour la levure mais conduit à un sévère phénotype de retard de croissance (Zebarjadian *et al.*, 1999), ce qui indique que ce ne serait pas la présence de 1 ou 2 résidus Ψ pris isolément dans l'ARNr qui contribuerait à la structuration et à la fonctionnalité de l'ARNr mais bien une partie ou l'ensemble des résidus Ψ d'une même région de l'ARNr.

2. Les pseudouridines dans les snRNA

2.1. L'épissage des pré-ARNm



B

	snRNA U1			snRNA U2			snRNA U4			snRNA U5			snRNA U6		
	Ψ	2'*O*	mN	Ψ	2'*O*	mN	Ψ	2'*O*	mN	Ψ	2'*O*	mN	Ψ	2'*O*	mN
H. sapiens	2	2	-	14	9	-	3	4	-	3	5	-	nd	nd	nd
R. norvegeticus	2	2	-	16	9	1	3	4	-	3	5	-	3	8	2
M. musculus	2	2	-	nd	nd	nd	3	4	-	3	5	-	3	8	2
D. melanogaster	2	2	-	nd	nd	nd	2	2	-	4	5	-	nd	nd	nd
C. elegans	1	nd	nd	9	nd	nd	2	nd	nd	3	nd	nd	1	nd	nd
S. pombe	1	-	-	7	4	1	-	-	-	4	3	-	2	4	1
S. cerevisiae	2	nd	nd	3	nd	nd	-	nd	nd	1	nd	nd	-	nd	nd

Figure i8 : La réaction d'épissage et tableau récapitulatif du nombre de résidus modifiés présents dans les UsnRNA de diverses espèces.

A. L'épissage se déroule en deux étapes de trans-estérification. La première implique l'attaque nucléophile par le groupement OH en 2' du résidu adénine de la boîte de branchement sur le phosphate 3' de l'exon 1 du pré-ARNm. La seconde met en jeu l'attaque nucléophile par l'extrémité 3' OH de l'exon 1 sur le phosphate situé entre l'extrémité 3' de l'intron et la partie 5' de l'exon 2. Il en résulte la formation d'un intron en lasso qui sera dégradé ou maturé, ainsi que la formation de l'ARNm mature. Les exons sont encadrés. L'intron est représenté en noir. Le résidu adénine de la boîte de branchement nécessaire à la première étape est représenté en rouge. L'extrémité 3'OH de l'exon 1 nécessaire à la seconde étape est représentée en vert. (Py)n correspond à une séquence polypyrimidine.

B. Ψ, 2'*O* et mN indiquent respectivement des résidus pseudouridines, 2'-*O*-méthylés et des méthylation de bases, "nd" signifie non déterminé. (D'après la revue Massenet *et al.*, 1998).

L'épissage de l'ARNm a lieu dans le noyau et se déroule au cours d'une série de deux réactions de trans-estérification. La première est initiée par l'attaque nucléophile du groupement 2' OH du ribose de l'adénosine de la boîte de branchement de l'intron sur le phosphate situé à la jonction entre l'exon 1 en 5' et l'intron. Ceci conduit à la libération de l'exon 1 avec une extrémité 3'OH et un intermédiaire réactionnel en forme de lasso constitué de l'intron lié à l'exon 2 par une liaison 5'-2' phosphodiester (Fig. i8 A). La seconde réaction est réalisée par l'attaque nucléophile du groupement 3'OH libre de l'exon 1 sur le phosphate de l'extrémité 5' du deuxième exon, ce qui entraîne le clivage de la jonction intron-exon 2 et la ligature des exons 1 et 2. L'intron est libéré sous forme de lasso avant d'être dégradé par des RNases ou dirigé dans une voie de maturation (pour revue, (Wahl *et al.*, 2009)).

Une machinerie de composition complexe et dynamique appelée le spliceosome doit s'assembler sur le pré-ARNm avant le déroulement des deux réactions catalytiques. Le spliceosome est composé de protéines et d'ARN (pour revue, (Wahl *et al.*, 2009)), en particulier de UsnRNA (small nuclear RNA) qui sont des ARN non codant de séquences riches en U. Dans le cas de la majorité des pré-ARNm, 5 UsnRNA participent à la formation du splicesome et sont dénommés U1, U2, U4, U5 et U6. Chacun s'associe avec des protéines spécifiques conduisant à la formation des 5 snRNP. Les UsnRNA subissent des modifications post-transcriptionnelles telles que l'ajout d'une coiffe en 5' et son hyperméthylation, et la modification de certains résidus de leur séquence (Epstein *et al.*, 1980; Branlant *et al.*, 1981a; Krol *et al.*, 1981a; Krol *et al.*, 1981b; Branlant *et al.*, 1982).

2.2. Les snRNA

Les snRNA possèdent un nombre limité de nucléotides modifiés recensés parmi les espèces analysées. Les principales modifications retrouvées sont la 2'-*O*-méthylation, la pseudouridylation et la méthylation de bases (m6A principalement) (pour revue, (Massenet *et al.*, 1998)). La figure i8 B répertorie les résidus modifiés dans les snRNA. On constate qu'il y a un nombre plus élevé de modifications dans

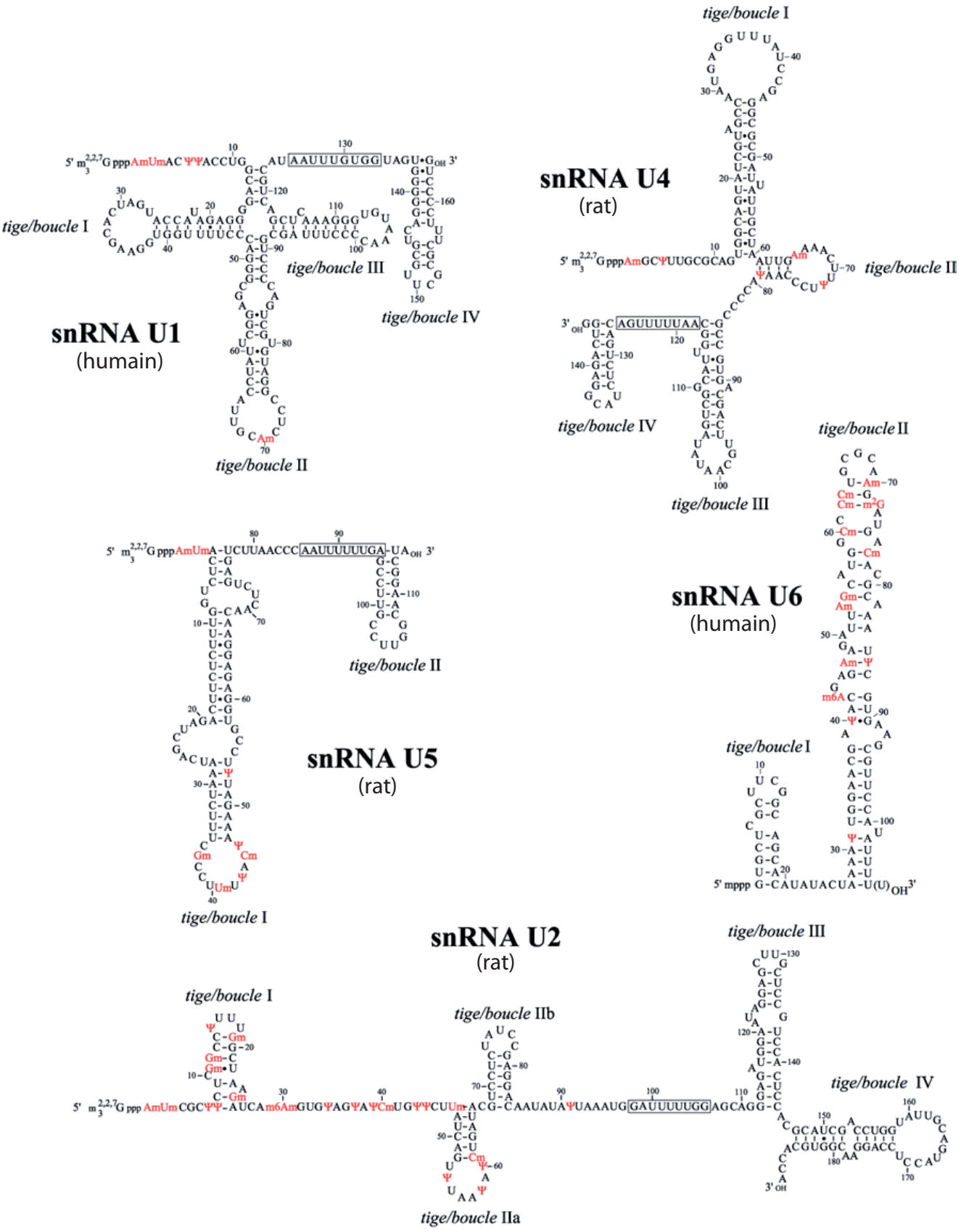


Figure i9 : Modèles de structure secondaire des snRNA spliceosomaux U1, U2, U4, U5 et U6 portant des nucléotides modifiés.

Les structures secondaires représentées sont celles établies pour les snRNA : U1 (Branlant *et al.,* 1981) ; U2 (Branlant *et al.,* 1982) ; U4 (Krol *et al.,* 1981b) ; U5 (Krol *et al.,* 1981a) et U6 (Epstein *et al.,* 1980). La séquence permettant la fixation des protéines Sm est encadrée dans les différents snRNA (Branlant *et al.,* 1982). Les nucléotides modifiés sont indiqués en rouge et le nom de l'espèce est indiqué sous le snRNA.

les séquences de UsnRNA produits dans les organismes les plus évolués au niveau phylogénétique. Les snRNA U1 et U2 sont modifiés au niveau de leur région 5' terminale, les snRNA U4 et U6 sont majoritairement modifiés dans leur région centrale et le snRNA U5 comporte essentiellement des modifications au niveau de l'hélice et de la boucle terminale de sa structure tige-boucle I (Fig. i9). De façon intéressante, les modifications des UsnRNA spliceosomaux sont concentrées dans les régions contractant des appariements intermoléculaires qui sont formés lors de la réaction d'épissage entre les snRNA et l'ARN pré-messager ou entre les snRNA. Ces résidus Ψ ont vraisemblablement des fonctions stabilisatrices d'éléments de structure secondaire, ou des rôles favorisant des conformations particulières des ARN.

De tous les snRNA, c'est le snRNA U2 qui est le plus intensément modifié (Fig. i8 B et i9) (Massenet *et al.*, 1999). Les 6 résidus uracile de la région d'interaction avec le site de branchement et dans la région 3' adjacente sont convertis en résidus pseudouridine (Ψ34, Ψ37, Ψ39, Ψ41, Ψ43, Ψ44). Trois de ces résidus pseudouridines (Ψ34, Ψ41 et Ψ43) sont conservées dans le snRNA U2 de *S. cerevisiae* (Ψ35, Ψ42 et Ψ44) (Massenet *et al.*, 1999). Chez *S. cerevisiae*, la position Ψ35 est modifiée par l'enzyme Pus7p (Behm-Ansmant *et al.*, 2003; Ma *et al.*, 2003). La Ψ44 est quant à elle catalysée par l'enzyme Pus1p (Massenet *et al.*, 1999) . Enfin, la Ψ42 du snRNA U2 est convertie par la snoRNP H/ACA snR81 appartenant au système guidé de modification (Ma *et al.*, 2005). Le snRNA U2 joue deux rôles majeurs dans l'épissage. Il s'apparie avec l'ARNm en reconnaissant la boîte de branchement dans la première étape de l'épissage (Fig. i10 A), et il se lie au snRNA U6 dans la seconde étape de l'épissage (Fig. i10 B).

Chez la levure, le résidu conservé Ψ35 (Ψ34 chez les mammifères) du snRNA U2 est adjacent, au sein de l'appariement U2-pre-mRNA, à l'adénine très conservée et représentant le point de branchement de la structure en lasso de l'intron (A36 ou A35 chez les mammifères) (Fig. i10 C et D). La pseudouridine 35 induit un changement local de conformation de l'ARNm ce qui a pour conséquence d'expulser

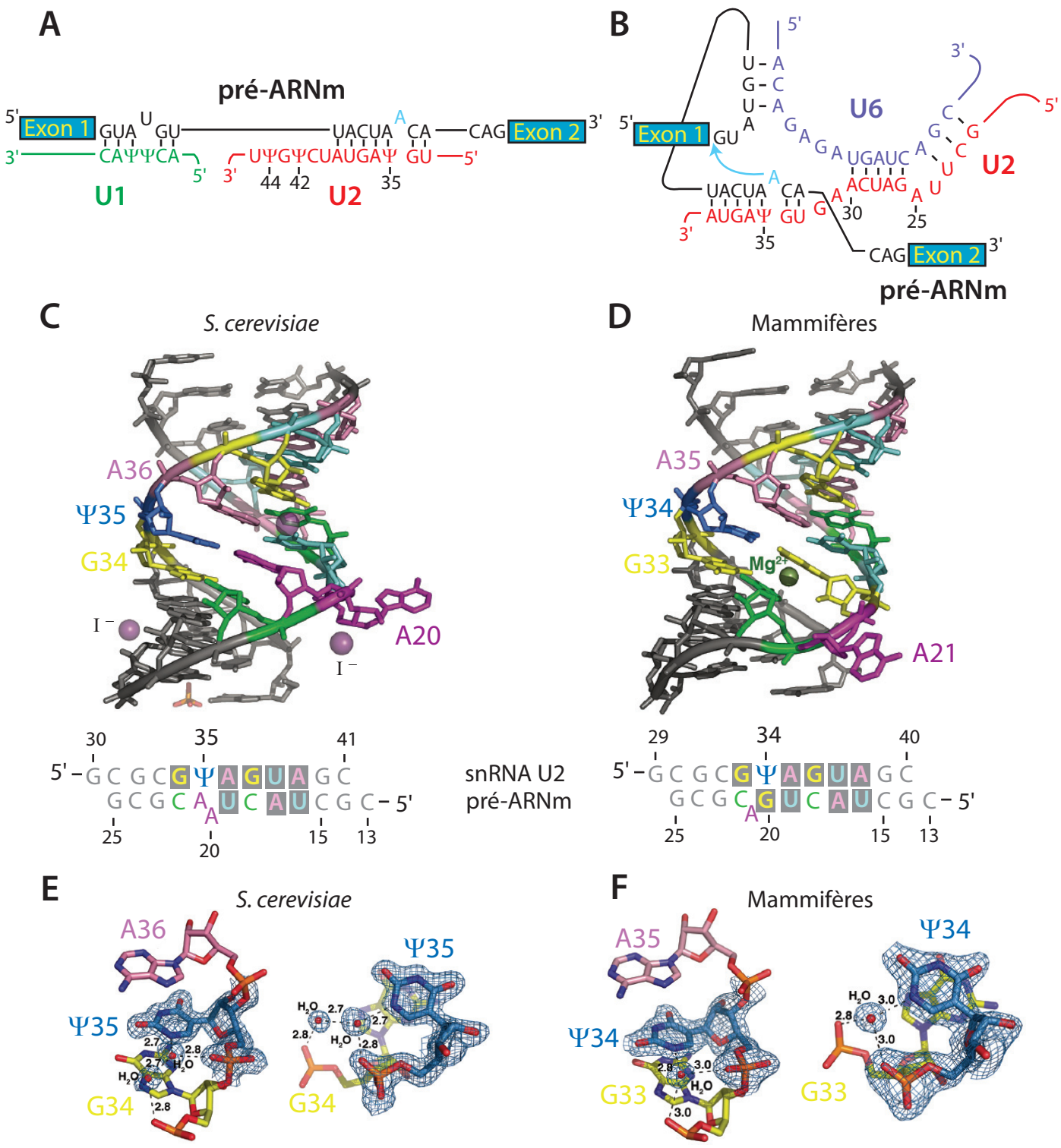


Figure i10 : Rôle de la pseudouridine 34/35 du snRNA U2 dans l'expulsion du résidu A de la boîte de branchement du pré-ARNm. (Newby et Greenbaum, 2001 ; Lin et Kielkopf, 2008).
A. Appariement entre un pré-ARNm et les snRNA U1 et U2.
B. Appariement observé entre un pré-ARNm et les snRNA U2 et U6 dans la seconde étape de l'épissage (complexe B). L'appariement du snRNA U2 avec la séquence de la boîte de branchement conduit à la mise en position extra-hélicale du résidu A responsable de l'attaque nucléophile.
C. et **D.** Structures 3D du snRNA U2 apparié à la séquence de la boîte de branchement d'un pré-ARNm chez la levure et chez les mammifères respectivement. Les nucléotides conservés dans chaque duplexe sont colorés : G en jaune, C en vert, A en magenta, U en cyan et Ψ en bleu.
E. et **F.** Interactions du type empilement de base et hydratation liées à la présence du résidu Ψ chez la levure et chez les mammifères respectivement. La carte de densité électronique autour du résidu Ψ et des molécules d'eau est représentée en bleu. La vue de droite correspond à une rotation de 90° autour de l'axe des x par rapport à la vue de gauche.

le résidu A de l'hélice, le plaçant ainsi dans une position favorable pour réaliser l'attaque nucléophile de la première réaction de l'épissage (Newby et Greenbaum, 2001). Le simple remplacement du résidu U35 en Ψ35 est suffisant pour déplacer le résidu A36 de l'ARNm d'un état empilé en faveur d'une conformation extra hélicale (Newby et Greenbaum, 2001). En effet, il a été montré expérimentalement que les résidus pseudouridines du snRNA U2 étaient indispensables à l'assemblage de la snRNP ainsi qu'à la réaction d'épissage (Zhao et Yu, 2004b). A partir des données expérimentales un premier modèle a été établi dans lequel le résidu Ψ35 participe à l'expulsion du résidu A de l'ARNm (Newby et Greenbaum, 2002b). Ce modèle a été confirmé par la structure 3D obtenue en présence d'un résidu Ψ en position 34/35 (Lin et Kielkopf, 2008) (Fig. i10 C et D). De manière intéressante, il apparaît que le résidu Ψ34/35 présent dans le snRNA U2 favoriserait son interaction avec le résidu localisé en amont par des interactions du type empilement de base (Fig. i10 E) ou du type double pontage par molécules d'eau (Fig. i10 F), ce qui augmente la stabilité du duplexe (Lin et Kielkopf, 2008).

3. Les pseudouridines dans les ARNt

Dans les trois domaines du vivant, les ARNt présentent tous au moins un résidu Ψ (pour revues, (Limbach *et al.*, 1994; Motorin et Grosjean, 1998)). Les 3 boucles de la structure en feuille de trèfle des ARNt sont les plus modifiées (Fig. i11). Ainsi, la dénomination de deux de ces boucles repose sur la nature des résidus modifiés présents au sein de leurs séquences. En effet, la boucle D contient très fréquemment des résidus dihydrouridines (D) aux positions 16, 17 et 20 ; tandis que la boucle TΨC comporte une ribothymine (T) et une pseudouridine (Ψ) aux positions 54 et 55 (Rozenski *et al.*, 1999). La boucle anti-codon contient également certaines positions conservées fréquemment pseudouridylées comme les positions 31, 32, 38 et 39. Le nombre important de nucléotides modifiés différents et la conservation de certaines des positions modifiées suggèrent que ces modifications jouent un rôle très important dans les ARNt, en particulier dans le mécanisme de traduction.

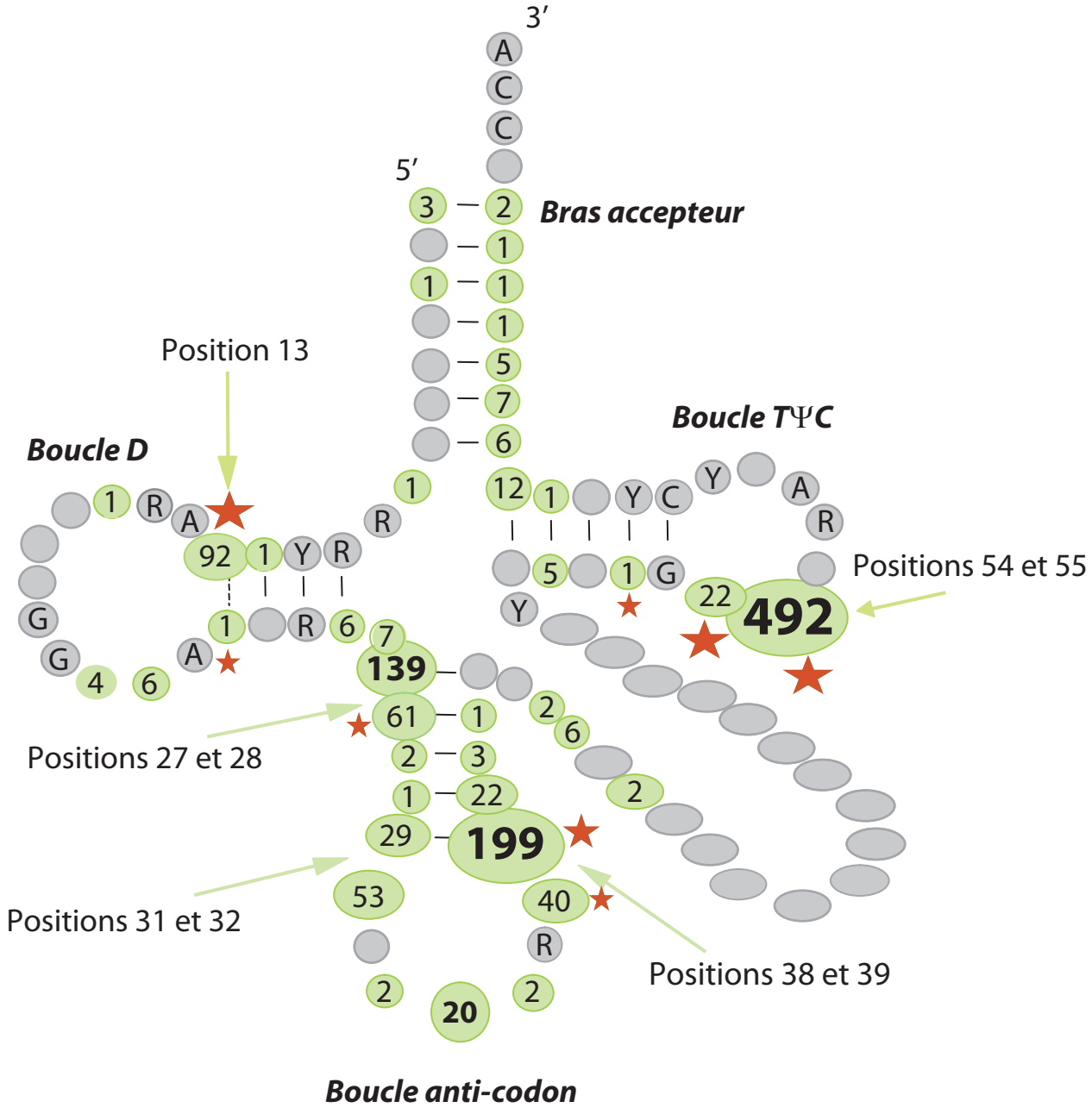


Figure i11 : Localisation et fréquences de détection des résidus Ψ dans les ARNt provenant des trois domaines du vivant. (D'après la banque de données de modifications dans les ARNt : http://medlib.med.utah.edu/RNAmods/ ; Rozenski *et al.,* 1999).
La banque de donnée regroupe les informations concernant 571 séquences d'ARNt dont 59 seulement proviennent d'archaea. Le nombre d'ARNt contenant un résidu Ψ est indiqué pour chaque position. Les positions pseudouridylées chez les archaea sont indiquées par une étoile rouge dont la taille est proportionnelle au nombre de résidus Ψ détectés. Les nucléotides concervés dans plus de 90% des cas sont indiqués (R = purine, Y = pyrimidine).

3.1. Les positions des résidus Ψ dans l'ARNt

La pseudouridine en position 13 est bien conservée mais son rôle n'est pas établi avec certitude. Une étude a montré que son absence (cumulée à la perte d'une autre modification post-transcriptionnelle m^2A37), conduit à des perturbations structurales très limitées sur la boucle anti-codon et que ces résidus ne contribuaient pas à la reconnaissance par l'aminoacyl ARNt-synthétase (Madore *et al.*, 1999).

Les résidus Ψ27 et Ψ28 structureraient l'ARNt de façon à ce que celui-ci soit reconnu efficacement par la machinerie de traduction. En effet, des études ont montrées qu'en absence du résidu Ψ27 l'appariement avec le résidu G43 est perdu (Fig. i11), ce qui conduit à une mauvaise lecture des codons par un ARNt dont la conformation est altérée (Schultz et Yarus, 1994a, b).

Les résidus Ψ31 et Ψ32 auraient pour rôle de stabiliser les paires de bases 31-39 et 32-38 situées à l'extrémité de la tige de la boucle anticodon (Fig. i11). En effet, le résidu Ψ31 diminue la mobilité conformationnelle de la paire de base 31-39. Cette dernière devient ainsi thermodynamiquement équivalente à une paire G-C, c'est à dire que sa stabilité est augmentée par comparaison à une paire A-U (Auffinger et Westhof, 1998). Quant au résidu Ψ32, il permet d'établir deux liaisons hydrogènes sur son oxygène receveur et renforce la stabilité de la tige-boucle anticodon (Auffinger *et al.*, 1999).

Les nucléotides 34, 35 et 36 constituent l'anti-codon des ARNt. Dans certains cas, la dégénérescence du code génétique et la redondance des codons impose un choix multiple de codon de l'ARNm pour un même ARNt. Par exemple, l'$ARNt^{Ile}$ porte l'anti-codon UAU et peut reconnaître à la fois les codons AUA de l'isoleucine et AUG de la méthionine (Fig. i12). Pour éviter une mauvaise lecture par ces ARNt, il a été montré que des résidus pseudouridines aux positions 34 et 36 des $ARNt^{Ile}$ permettaient de réaliser un appariement spécifique avec le codon de l'isoleucine des ARNm et ainsi de discriminer le codon méthionine (Fig. i12) (Szweykowska-Kulinska *et al.*, 1994; Senger *et al.*, 1997). Dans ce cas, ces résidus Ψ jouent le rôle d'anti-déterminants en empêchant la reconnaissance du codon méthionine.

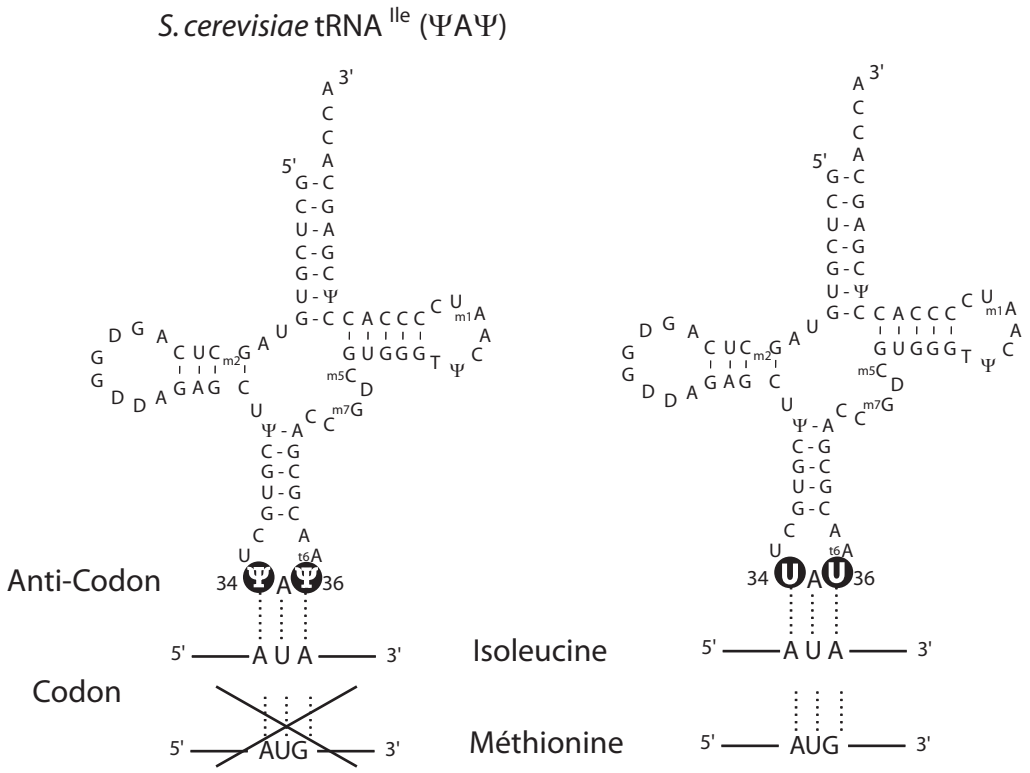


Figure 12 : Rôle des résidus pseudouridine 34 et 36 dans les ARNt. (Adapté à partir des données obtenues par Senger *et al.,* 1997).
Les résidus Ψ34 et Ψ36 jouent le rôle d'anti-déterminant en empêchant la reconnaissance du codon méthionine (AUG) et permettent à l'ARNt isoleucine de *S. cerevisiae* portant l'anticodon (ΨAΨ) de reconnaître uniquement le codon isoleucine.

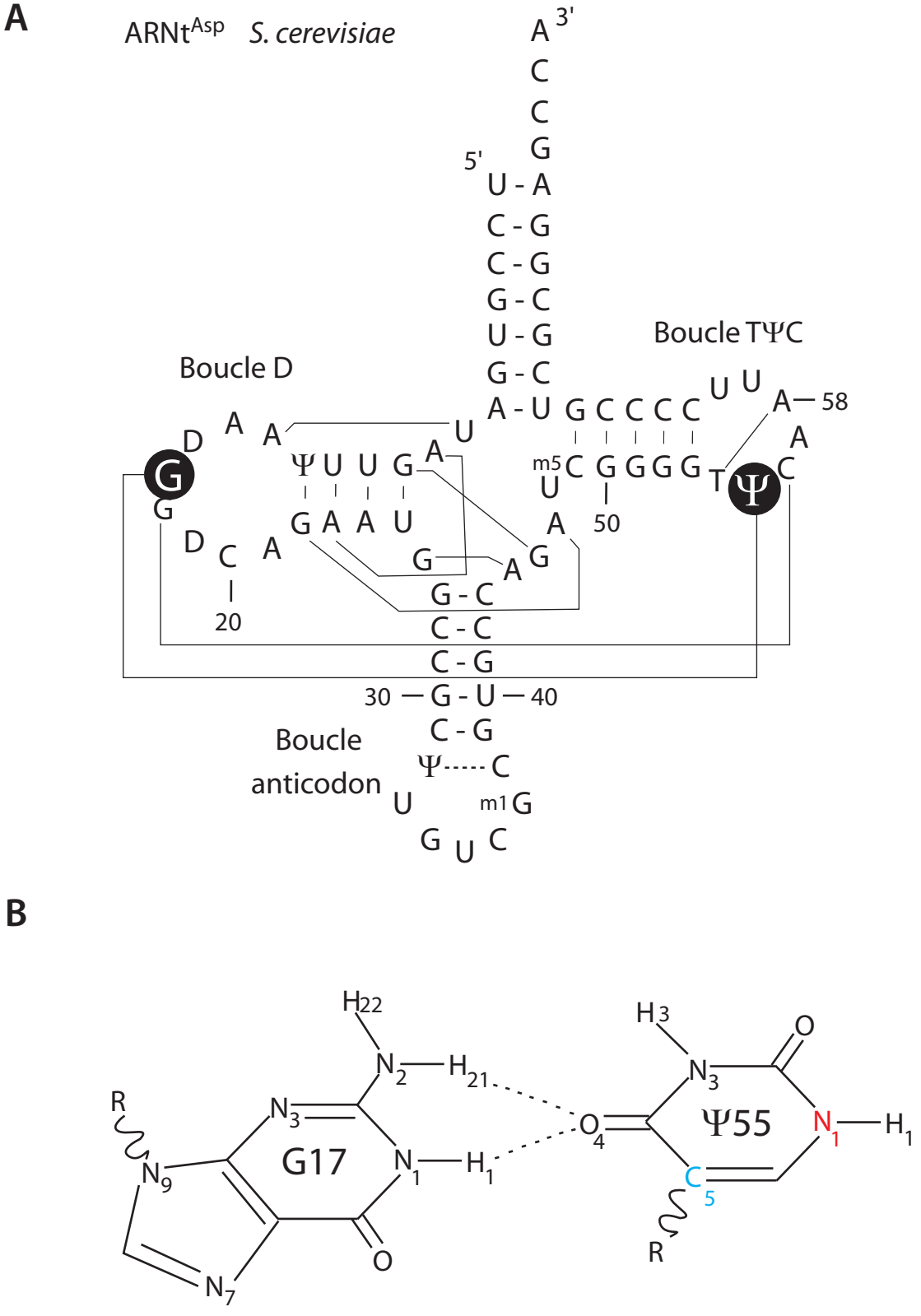


Figure i13 : Rôle du résidu pseudouridine 55 dans les ARNt. (Adapté de Auffinger et Westhof, 1999).

A. Structure 2D en feuille de trèfle de l'ARNt^{Asp} de *S. cerevisiae* et des liaisons hydrogènes observées dans le repliement 3D en "L" inversé.

B. Représentation schématique de l'interaction existant entre les résidus G17 et Ψ55 de l'ARNt^{Asp} de *S. cerevisiae*.

La présence de résidus modifiés aux positions Ψ38, Ψ39 et Ψ40 est importante pour le décodage de l'ARNm par l'ARNt en empêchant le décalage du cadre de lecture (Urbonavicius *et al.*, 2001). Il a été montré *in vivo* que les résidus Ψ présents à ces trois positions permettaient d'augmenter la fidélité de lecture de l'anti-codon par les ARNt, d'accroître l'efficacité de traduction lors de l'élongation et d'influencer la vitesse de sélection de l'ARNt correspondant au codon de l'ARNm (Li *et al.*, 1997). Ces résidus modifiés contribuent à la stabilisation de la tige-boucle anti-codon en favorisant l'empilement des bases des résidus de cette région. Ainsi, la tige-boucle anticodon d'un ARNt contenant le résidu Ψ39 est thermodynamiquement plus stable qu'une même tige-boucle dépourvue de la modification (Durant et Davis, 1997, 1999). Il a également été montré par RMN et spectroscopie que les résidus Ψ39 et s^2U34 (thiolation) de l'$ARNt^{Lys}$ cytoplasmique, qui sert d'amorce à la rétro transcription du virus HIV-1 humain, stabilisaient l'interaction entre la région anticodon de l'ARNt et la séquence de l'ARN viral (Bilbille *et al.*, 2009).

3.2. La position 55 des ARNt

La position 55 des ARNt possède un intérêt direct avec mon travail de thèse et certaines données bibliographiques complémentaires à celles présentées dans ce chapitre sont données dans la partie Résultats. C'est aussi le résidu le plus fréquemment retrouvé dans les ARNt. En effet, 90% des ARNt des 3 domaines du vivant possèdent une pseudouridine en position 55 (Fig. i11). Cette position est modifiée par TruB chez *E. coli* (Nurse *et al.*, 1995) ou Pus4p chez *S. cerevisiae* (Becker *et al.*, 1997). Chez les archaea, il est proposé que l'enzyme aCBF5 convertie seule ou en présence de aNOP10 et de aGAR1 la position 55 des ARNt (Roovers *et al.*, 2006; Gurha *et al.*, 2007; Muller *et al.*, 2007b). Toujours chez les archaea, l'enzyme aPUS10 est également capable de modifier les positions 54 et 55 des ARNt (Roovers *et al.*, 2006; Gurha et Gupta, 2008).

L'invalidation du gène de Pus4p n'entraîne aucun phénotype particulier de croissance suggérant que la modification de la position 55 n'est pas essentielle à la

viabilité cellulaire (Becker *et al.*, 1997). Des travaux sur l'ARNt d'*E. coli* ont montré que la Ψ55 n'avait aucune influence sur le décalage du cadre de lecture du ribosome ou sur la traduction (Urbonavicius *et al.*, 2003). Ces résultats sont tout de même surprenants étant donné qu'il s'agit de la position modifiée la plus conservée des ARNt.

Le repliement des ARNt sous la forme d'un « L » inversé conduit à une interaction entre les résidus G17 et Ψ55 (Fig. i13 A). Un des rôles de la Ψ55 est de stabiliser cette interaction tertiaire qui est nécessaire à la structure globale adoptée par un ARNt. Cette interaction se produit par l'intermédiaire de deux liaisons hydrogènes particulières qui sont appelées « liaisons hydrogènes bifurquées ». Elles proviennent des hydrogènes liés aux atomes d'azotes 1 et 2 du résidu G17 et sont dirigées vers le même atome receveur, l'oxygène 4, de la pseudouridine (Fig. i13 B) (Auffinger et Westhof, 1999). Par les liaisons hydrogènes qu'il établit, le résidu Ψ55 universellement conservé apparait être « bloqué » dans une conformation particulière qui pourrait ainsi contribuer à la stabilité de la boucle TΨC (Auffinger et Westhof, 1999).

III. Les enzymes de modification : les ARN:pseudouridine-synthases

Les nucléotides modifiés sont synthétisés par des enzymes de modifications de l'ARN. Dans le cas du résidu pseudouridine, ces enzymes sont appelées ARN:pseudouridine-synthases.

Des études comparatives de séquences de domaines protéiques, menées sur des génomes d'eucaryotes, d'archaea et de bactéries, ont montré que de 3% à 11% de la capacité codante d'un génome pouvait conduire à la production d'enzymes de modification de l'ARN (Anantharaman *et al.*, 2002). La cellule possède ainsi la ressource génétique pour les activités enzymatiques assurant les réactions de modifications et il est à présent clairement établi que beaucoup d'énergie est dépensée pour introduire des dizaines de nucléotides modifiés dans les ARN. On

peut dénombrer plus d'une centaine d'enzymes différentes catalysant la modification des nucléotides de l'ARN (pour revue, (Ferre-D'Amare, 2003)). Il semble donc que l'ensemble des enzymes de modification appartiennent au jeu minimal de biomolécules nécessaires à la vie cellulaire (Mushegian et Koonin, 1996).

A partir de ce chapitre, j'emploierai indifféremment les termes d'ARN cible ou d'ARN substrat pour qualifier les ARN qui sont modifiés par les enzymes ou par les complexes enzymatiques de type RNP.

A. Découverte et classification des ARN:pseudouridine-synthases

Les premières ARN:Ψ-synthases ont été identifiées par voie biochimique en testant, après fractionnement chromatographique, les activités ARN:Ψ-synthases grâce à une technique de mesure du tritium (^{3}H) libéré. Cette stratégie consiste à utiliser des ARN substrats ayant incorporé des uraciles qui contiennent un atome de tritium en position 5 (Cortese *et al.*, 1974). En effet, la réaction d'isomérisation des résidus U conduit à une libération d'un atome d'hydrogène présent initialement à cette position qui peut alors être dosé (voir la partie III.C, Fig. i14 étape 6). C'est ainsi qu'a été découvert la première enzyme qui a initialement été appelée Pseudouridylate Synthase I puis a été renommée TruA (Arena *et al.*, 1978).

La seconde stratégie employée est basée sur la recherche par homologie de séquence. Le gène candidat dont la séquence codante présente des similarités avec des ARN:Ψ-synthases connues est cloné dans un vecteur d'expression qui permet la surproduction de la protéine chez *E. coli* puis sa purification. La capacité de la protéine purifiée, à modifier des substrats ARN produits par transcription *in vitro*, est alors testée.

Ces deux premières approches sont complémentaires et aboutissent à la caractérisation d'une ARN:Ψ-synthase et de l'identification de ses substrats (pour revue, (Ansmant et Motorin, 2001)). Deux approches expérimentales sont utilisées pour démontrer la fonction d'une ARN:Ψ-synthase. D'une part, lorsque le gène est

Famille	Organisme	Mode d'identification	Spécificité de substrat	Références
TruA				
TruA (HisT, PSU-I)	*E. coli*	Biochimique	Ψ_{38}, Ψ_{39}, Ψ_{40} ARNt	Kammen *et al.*, 1988
PSU-I	*S. typhimurium*	Biochimique	Ψ_{38}, Ψ_{39}, Ψ_{40} ARNt	Arena *et al.*, 1978
PSU-I	*B. taurus*	Biochimique	Ψ_{38}, Ψ_{39}, Ψ_{40} ARNt	Green *et al.*, 1982
Pus1	*S. cerevisiae*	Homologie de séquence TruA	$\Psi_{26\text{-}28}$, $\Psi_{34\text{-}36}$, Ψ_{65}, Ψ_{67} ARNt	Simos *et al.*,1996 /Motorin *et al.*,1998
			Ψ_{44} snRNA U2	Massenet *et al.*, 1999
mPus1	*M. musculus*	Homologie de séquence Pus1	Ψ_{27}, $\Psi_{34, 36}$, (Ψ_{32}?) ARNt	Chen et Patton, 1999
			Ψ_{1}ARNt, Ψ_{44} snRNA U2	Behm-Ansmant *et al.*, 2006
Pus1	*C. elegans*	Homologie de séquence Pus1	Ψ_{27} ARNt	Patton et Padgett, 2003
Pus2	*S. cerevisiae*	Homologie de séquence TruA	Ψ_{27}, Ψ_{28} ARNt mitochondriaux	Behm-Ansmant *et al.*, 2007
Pus3 (Deg1)	*S. cerevisiae*	Homologie de séquence TruA	Ψ_{38}, Ψ_{39} ARNt cytoplasmiques et mitochondriaux	Lecointe *et al.*, 1998
TruB				
TruB	*E. coli*	Biochimique	Ψ_{55} ARNt	Nurse *et al.*, 1995
Pus4 (Ynl292)	*S. cerevisiae*	Homologie de séquence TruB	Ψ_{55} ARNt cytoplasmiques et mitochondriaux	Becker *et al.*, 1997a
Cbf5	*S. cerevisiae*	Homologie de séquence TruB	ARNr cytoplasmiques	Lafontaine *et al.*, 1998a
RluA				
RluA	*E. coli*	Biochimique	Ψ_{746} ARNr 23S + Ψ_{32} ARNt	Wrzesinski *et al.*, 1995b
RluC (YceC)	*E. coli*	Homologie de séquence TruA	Ψ_{955}, Ψ_{2504}, Ψ_{2580} ARNr 23S	Conrad *et al.*, 1998
RluD (YfiI, SfhB)	*E. coli*	Biochimique	Ψ_{1911}, Ψ_{1915}, Ψ_{1917} ARNr 23S	Raychaudhuri *et al.*, 1998
TruC (YqcB)	*E. coli*	Homologie de séquence RsuA	Ψ_{65} ARNt	Del Campo *et al.*, 2001
Pus5 (Ylr165)	*S. cerevisiae*	Homologie de séquence RluA	Ψ_{2819} (Ψ_{2580} *E. coli*) ARNr 21S mitochondrial	Ansmant *et al.*, 2000
Pus6 (Ygr169)	*S. cerevisiae*	Homologie de séquence RluA	Ψ_{31} ARNt cytoplasmiques et mitochondriaux	Ansmant *et al.*, 2001
Pus8 (Rib2)	*S. cerevisiae*	Homologie de séquence RluA	Ψ_{32} ARNt cytoplasmiques	Behm-Ansmant *et al.*, 2004
Pus9 (Ydl036)	*S. cerevisiae*	Homologie de séquence RluA	Ψ_{32} ARNt mitochondriaux	Behm-Ansmant *et al.*, 2004
RsuA				
RsuA	*E. coli*	Biochimique	Ψ_{516} ARNr 16S	Wrzesinski *et al.*, 1995a
RluB (YciL)	*E. coli*	Homologie de séquence RluB B. subtilis	Ψ_{2605} ARNr 23S	Del Campo *et al.*, 2001
RluB	*B. subtilis*	Homologie de séquence RsuA	Ψ_{2633} (Ψ_{2605} *E. coli*) ARNr 23S	Niu et Ofengand, 1999
RluE (YmfC)	*E. coli*	Homologie de séquence RsuA	Ψ_{2457} ARNr 23S	Conrad *et al.*, 1999 Del Campo *et al.*, 2001
RluF (YjbC)	*E. coli*	Homologie de séquence RsuA	Ψ_{2604} ARNr23S	Del Campo *et al.*, 2001
TruD				
TruD (YgbO)	*E. coli*	Biochimique	Ψ_{13} ARNt	Kaya et Ofengand, 2003
Pus7 (Yor243)	*S. cerevisiae*	Banque ORF fusionnées GST	Ψ_{35} du snRNA U2	Ma *et al.*, 2003
			Ψ_{13}, Ψ_{35} ARNt	Behm-Ansmant *et al.*, 2003
Pus10				Watanabe et Gray 2000 Roovers *et al.*, 2006 Gurha et Gupta, 2008
Pus10 / PsuX	*P. furisus*	Homologie de séquence	Ψ_{54} , Ψ_{55} ARNt	

Tableau i2 : ARN:Ψ-synthases caractérisées à ce jour et approches utilisées pour leur caractérisation.

Le code couleur utilisé pour le nom des enzymes correspond au mode d'identification : biochimie en rouge, par homologie de séquence en vert et par utilisation d'une banque d'ORF fusionnées à la GST en bleu.

non essentiel à la viabilité cellulaire, il est possible de l'invalider et de repérer l'absence de résidus Ψ aux positions attendues. La modification à ces positions peut ensuite être rétablie en complémentant la souche invalidée par expression ectopique du gène. D'autre part, dans le cas de gènes essentiels, la région codante du gène n'est pas invalidée mais placée sous contrôle d'un promoteur dont l'activité peut être contrôlée. Dans ce cas, les modifications des substrats potentiels sont analysées après avoir réprimé l'expression du gène endogène.

La troisième approche est basée sur le criblage à haut débit de produits de l'expression de gènes. Il s'agit de rechercher une enzyme pour une position identifiée comme étant modifiée. L'une des stratégies possible a consisté à employer une banque d'expression de *S. cerevisiae* dont les ORF ont été individuellement fusionnées à la GST. L'activité des protéines de fusion purifiées provenant de cette banque a été testée sur le snRNA U2 produit *in vitro*. Cet ARN substrat contenait un résidu (U35) dont le phosphate lié au carbone 5' était radiomarqué (Zhao et Yu, 2004a). Par cette méthode, les auteurs ont ainsi caractérisé le gène codant l'ARN:Ψ-synthase Pus7p responsable de la modification de la position 35 du snRNA U2 (Ma *et al.*, 2003).

Les premières ARN:Ψ-synthases identifiées étaient des enzymes d'*E. coli* et ont initialement été classées en 4 familles distinctes sur la base d'un motif commun de fixation de l'uridine : TruA (tRNA PseudoUridine-synthase A), TruB, RluA (rRNA Large subunit PseudoUridine-synthase A) et RsuA (rRNA Small subunit PseudoUridine-synthase A) (Koonin, 1996). Puis vinrent s'ajouter 2 nouvelles classes d'enzymes : TruD et Pus10 (ARN:PseudoUridine-Synthase 10) (Watanabe et Gray, 2000; Kaya et Ofengand, 2003). Actuellement, les ARN:Ψ-synthases sont regroupées en 6 familles (Tab. i2).

B. Mécanisme catalytique de la réaction de pseudouridylation

1. Deux mécanismes catalytiques proposés

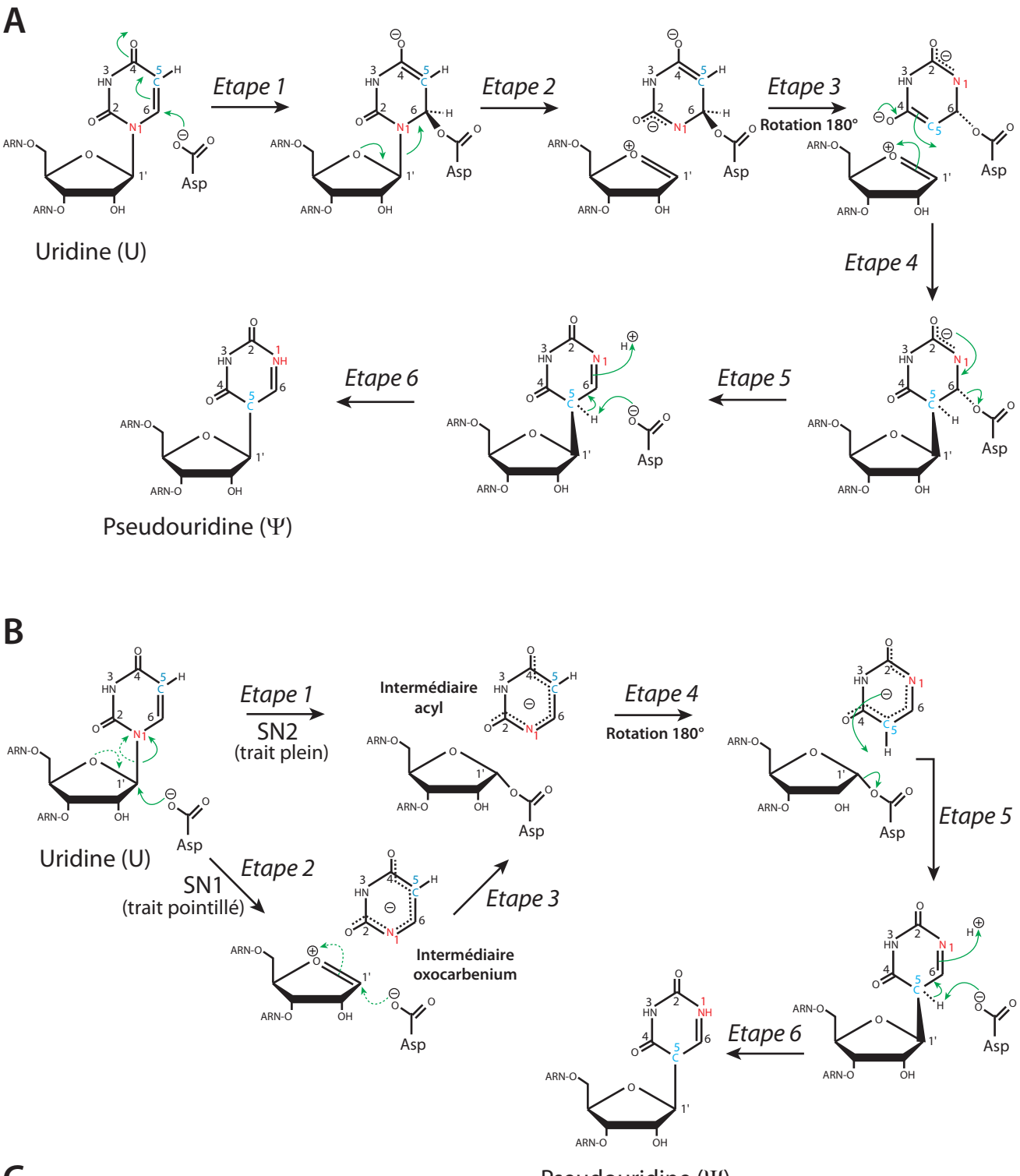


Figure i14 : Mécanismes catalytiques proposés pour la réaction de pseudouridylation. (Adapté de Hamilton *et al.,* 2006).
Deux mécanismes différents reposent, l'un sur l'attaque nucléophile en position C6 de l'uridine (**A**), l'autre sur l'attaque nucléophile de la position C1' du ribose de l'uridine (**B**). En **B**. les traits pointillés correspondent à la réaction SN1, tandis que les traits pleins indiquent la réaction de SN2. **C.** Structure de la 5-fluorouridine (5-FU) et de la 5-fluoro-6-hydroxypseudouridine (5-F-6-OHΨ).

Le mécanisme catalytique de la réaction de pseudouridylation proposé est basé sur la réactivité de l'acide aspartique agissant comme résidu catalytique. Celui-ci est parfaitement conservé parmi les différentes familles d'ARN:Ψ-synthases (Huang *et al.*, 1998). Actuellement, il subsiste toujours 2 mécanismes possibles pour l'isomérisation de l'uridine en pseudouridine (Hamilton *et al.*, 2006).

Le premier mécanisme est du type « addition de Michaël » et il correspond à l'addition d'un nucléophile sur une double liaison conjuguée avec un groupement électroattracteur. L'aspartate catalytique de l'enzyme réalise une attaque nucléophile sur le C6 de l'uracile, forme une liaison covalente et conduit à délocaliser les électrons sur l'oxygène du C4 (Fig. i14 A, étape 1). Il y a ensuite rupture de la liaison N-glycosidique entre la base et le ribose (étape 2) puis rotation de 180° de l'uracile (étape 3). Une nouvelle liaison se crée entre le C5 de la base et le C1' du ribose dont la force motrice est la relocalisation des électrons qui proviennent de l'oxygène du C4 (étape 4). Les deux dernières étapes conduisent au détachement de l'aspartate de la base (étape 5) et enfin une neutralisation des charges par élimination d'un proton pour former la pseudouridine (étape 6).

Le second mécanisme catalytique repose sur une réaction de substitution nucléophile d'ordre 1 ou 2 (SN1 ou SN2) avec la formation d'un intermédiaire acyl. L'aspartate catalytique accompli une attaque nucléophile sur le C1' du ribose puis 2 possibilités sont envisageables (Fig. i14 B). Soit une SN2 a lieu et la liaison N-gylcosidique entre la base et le ribose est rompue en une étape (étape 1). Soit une SN1 se produit en formant un intermédiaire oxocarbenium et l'uracile se détache du ribose en 2 étapes (étapes 2 et 3). Lors de la rotation de 180°, la base doit rester proche du ribose ; elle est probablement stabilisée par le site actif de l'enzyme car selon ce mécanisme catalytique, l'uracile n'est plus liée de façon covalente à l'aspartate. La suite du mécanisme se déroule d'une façon similaire au premier mécanisme pour aboutir à la formation de la pseudouridine (étapes 4 à 6).

2. Le 5-fluorouracile (5-FU), un composé efficace pour inhiber les ARN:Ψ-synthases ?

De nombreuses études montrent que la présence de 5-FU bloque l'action des ARN:Ψ-synthases (Fig. i14 C). Historiquement, le 5-FU a d'abord été incorporé uniformément à des ARNt lors de leur transcription *in vitro*. Ces ARNt modifiés, formaient un complexe stable avec les enzymes Pus4p et Pus7p de *S. cerevisiae* (Samuelsson, 1991). Le 5-FU a aussi la propriété de former *in vitro* un complexe covalent entre l'enzyme TruA et un ARNt modifié avec du 5-FU, comme l'a montré une étude dans laquelle ce complexe était préservé lors de sa migration sur un gel SDS-PAGE (Gu *et al.*, 1999). Dans des études réalisées *in vivo* à partir de cellules HeLa en cultures, il a été montré que l'incorporation de 5-FU dans le snRNA U2 bloquait la pseudouridylation de sites importants pour l'épissage (Ψ34, Ψ37, Ψ39, Ψ43) (Lenz *et al.*, 1994; Zhao et Yu, 2007).

Lors d'un crible effectué sur une collection de 4676 souches haploïdes de *S. cerevisiae*, 138 mutants ont été identifiées comme étant hypersensibles au 5-FU (Gustavsson et Ronne, 2008). La plupart des mutants les plus sensibles au 5-FU ne possèdent plus certaines enzymes de modification de l'ARNt, dont l'enzyme Pus1p. La maturation post-transcriptionnelle de l'ARNt joue un rôle important dans la signalisation *in vivo* du 5-FU. Ainsi, il a été proposé que la combinaison de l'effet du 5-FU et la perte de certaines enzymes de modification de l'ARNt conduise à la déstabilisation de l'ARNt en développant un phénotype de sensibilité à la température (Gustavsson et Ronne, 2008). Les effets observés par l'addition de 5-FU suggèrent également que la déstabilisation de l'ARNt contribuerait à la cytotoxicité du 5-FU dans les cellules saines. Il a ainsi été proposé que l'effet du 5-FU dans les thérapies anti-cancéreuses pourrait expliquer l'augmentation de l'hyperthermie (Kouloulias *et al.*, 2005).

De manière intéressante, une étude indique qu'en présence d'un substrat dont le résidu cible est un groupement 5-FU, TruB d'*E. coli* ne forme pas de complexe covalent même de façon transitoire et n'est que peu inhibée par ce dernier, contrairement à RluA et à TruB de la bactérie hyperthermophile *T. maritima*

(Phannachet et Huang, 2004; Spedaliere et Mueller, 2004). Ces observations semblent contredire les résultats montrant que le 5-FU inhibait les ARN:Ψ-synthases. Il apparaît ainsi qu'il existerait de subtiles différences dans la géométrie du site actif des ARN:Ψ-synthases, y compris entre les membres d'une même famille, ce qui pourrait expliquer la différence de comportement du 5-FU vis-à-vis de ces enzymes (Spedaliere et Mueller, 2004).

3. 5-FU et incertitude sur la nature du mécanisme catalytique

Les ARN:Ψ-synthases partagent de nombreuses similitudes structurales et de nombreux résidus conservés, comme nous le verrons dans la partie suivante, ce qui suggère que ces familles d'enzymes opèrent selon un mécanisme chimique universel (pour revue, (Hamma et Ferre-D'Amare, 2006)). Et pourtant, cette « simplicité apparente » du mode opératoire est plus complexe que prévu.

Les mesures de tritium libéré du substrat ARNt 5-FU de TruA indiquent que c'est plutôt la voie de l'attaque sur le C6 qui serait employée dans le mécanisme catalytique (Fig. i14 A) (Gu *et al.*, 1999). En outre, les auteurs précisent que la réaction sans fluor devrait se dérouler de la même façon qu'en présence de 5-FU (Gu *et al.*, 1999). Il était envisageable que d'autres résidus que l'aspartate puissent jouer un rôle de nucléophile dans la réaction catalytique, comme par exemple la cystéine. Il a donc été montré par mutagenèse dirigée que les résidus cystéines des enzymes TruB et RluA ne sont pas requis dans la catalyse, ce qui exclut leur participation dans une addition de Michaël sur le C6 de l'uridine (Fig. i14 A) (Ramamurthy *et al.*, 1999b).

Dans la structure cristalline du complexe issu de l'interaction entre TruB et le bras TΨC de l'ARNt contenant un 5-FU en position 55, le composé présent porte un atome de fluor sur le C5 (remplaçant l'hydrogène éliminé en fin de réaction) ainsi qu'une molécule d'eau liée au C6, et correspond donc au 5-fluoro-6-hydroxypseudouridine ou 5-F-6-OHΨ (Fig. i14 C) (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003; Phannachet et Huang, 2004; Phannachet *et al.*, 2005). Des études de

Famille	Enzyme	Organisme	Substrat	Auteurs	N° PDB
TRUA	TruA	*E. coli*		Foster *et al.,* 2000	1DJ0
			ARNt entier	Hur et Stroud, 2007	2NQP 2NR0/E
		T. thermophilus		Dong *et al.,* 2006	1VS3
TRUB	TruB	*E. coli*	ARNt TSL	Hoang et Ferre d'Amarre, 2001	1K8W
				Pan *et al.,* 2003	1R3F
		T. maritima	ARNt TSL	Pan *et al.,* 2003	1R3E
				Phannachet et Huang, 2004	1ZE2
				Phannachet et Huang, 2004	1ZE1
	TruB Y76F	*T. maritima*	ARNt TSL	Phannachet *et al.,* 2005	2AB4
	TruB D48N	*E. coli*	ARNt TSL	Hoang *et al.,* 2005	1ZL3
	aCBF5	*M. jannaschii*		Hamma *et al.,* 2005	2APO
		P. abyssi		**Mannival *et al.,* 2006**	**2AUS**
		P. furiosus		Rashid *et al.,* 2006	2EY4
			Particule	Li et Ye, 2006	2HVY
				Liang, B. *et al.,* 2007	2RFK
				Duan *et al.,* 2009	3HAX/Y
				Liang *et al.,* 2009	3HJW/Y
RLUA	RluA	*E. coli*	ARNt ACSL	Hoang *et al.,* 2006	2I82
	RluC	*E. coli*		Mizutani *et al.,* 2004	1V9K
	RluD	*E. coli*		Del Campo *et al.,* 2004	1QYU
				Mizutani *et al.,* 2004	1V9F
				Sivaraman *et al.,* 2004	1PRZ
RSUA	RsuA	*E. coli*	U et UMP	Sivaraman *et al.,* 2002	1KSK/L/V
		H. influenzae		Matte *et al.,* 2005	1VIO
	RluE	*E. coli*		Pan *et al.,* 2007	2OML/W
	RluF	*E. coli*		Sunita *et al.,* 2006	2GML
			Fragment ARNr	Alian *et al.,* 2009	3DH3
TRUD	TruD	*E. coli*		Ericsson *et al.,* 2004	1SZW
				Hoang et Ferre d'Amarre, 2004	1SB7
				Kaya *et al.,* 2004	1SI7
PUS10	PUS10	*H. sapiens*		McCleverty *et al.,* 2007	2V9K

Tableau i3 : Bilan des ARN:Ψ-synthases dont la structure tridimentionnelle a été déterminée.
La numérotation des variants enzymatiques de TruB correspond à celle employée chez *E. coli*.

biochimie et de spectrométrie de masse ont confirmé la formation de 5-F-6-OHΨ en présence de TruB (Spedaliere et Mueller, 2004). Ces deux résultats sont en faveurs du mécanisme de l'attaque du C6 par l'aspartate (Fig. i14 A). Ces auteurs proposent que certains sites actifs d'ARN:Ψ-synthases disposeraient d'un espace suffisant pour qu'une molécule d'eau se positionne entre le résidu catalytique et le C6, conduisant à l'hydratation directe de la base. Tandis que d'autres sites actifs avec une distance plus courte entre l'aspartate et le C6 favoriseraient la formation d'un ester hydrolysable (Spedaliere et Mueller, 2004).

Toutes ces données sont plutôt en faveur du mécanisme reposant sur l'attaque du C6 de l'uracile (Fig. i14 A). Néanmoins, une étude complémentaire basée sur la détection d'^{18}O par spectrométrie de masse a révélé que l'adduit C6-aspartate catalytique ne contenait pas d'^{18}O (Spedaliere *et al.*, 2004). Cela indique que l'hydratation du produit résultant de la réaction en présence de 5-FU ne peut pas découler d'un mécanisme basé sur une addition C6 de Michaël puis hydrolyse de l'ester et proviendrait du second mécanisme (Fig. i14 B) (Spedaliere *et al.*, 2004). La première explication possible est que le produit réarrangé et hydraté en présence de 5-FU (5-F-6-OHΨ) ne provienne pas de la voie catalytique « normale » qui a lieu lorsqu'il s'agit de convertir un résidu uracile (possédant un atome d'hydrogène qui peut être éliminé en fin de réaction contrairement à l'atome de fluor) en résidu pseudouridine (Spedaliere *et al.*, 2004). La seconde possibilité serait l'attaque d'une molécule d'eau activée (OH^-) sur le C6 de la 5-fluoropseudouridine pour former le produit 5-F-6-OHΨ (Spedaliere *et al.*, 2004).

L'impossibilité de figer dans un cristal un complexe covalent enzyme-substrat et l'observation d'une formation rapide du composé 5-F-6-OHΨ avant même la cristallisation ont concouru à proposer la possibilité d'une attaque de l'aspartate catalytique sur le C1' du ribose comme mécanisme catalytique (Fig. i14 B) (Phannachet et Huang, 2004).

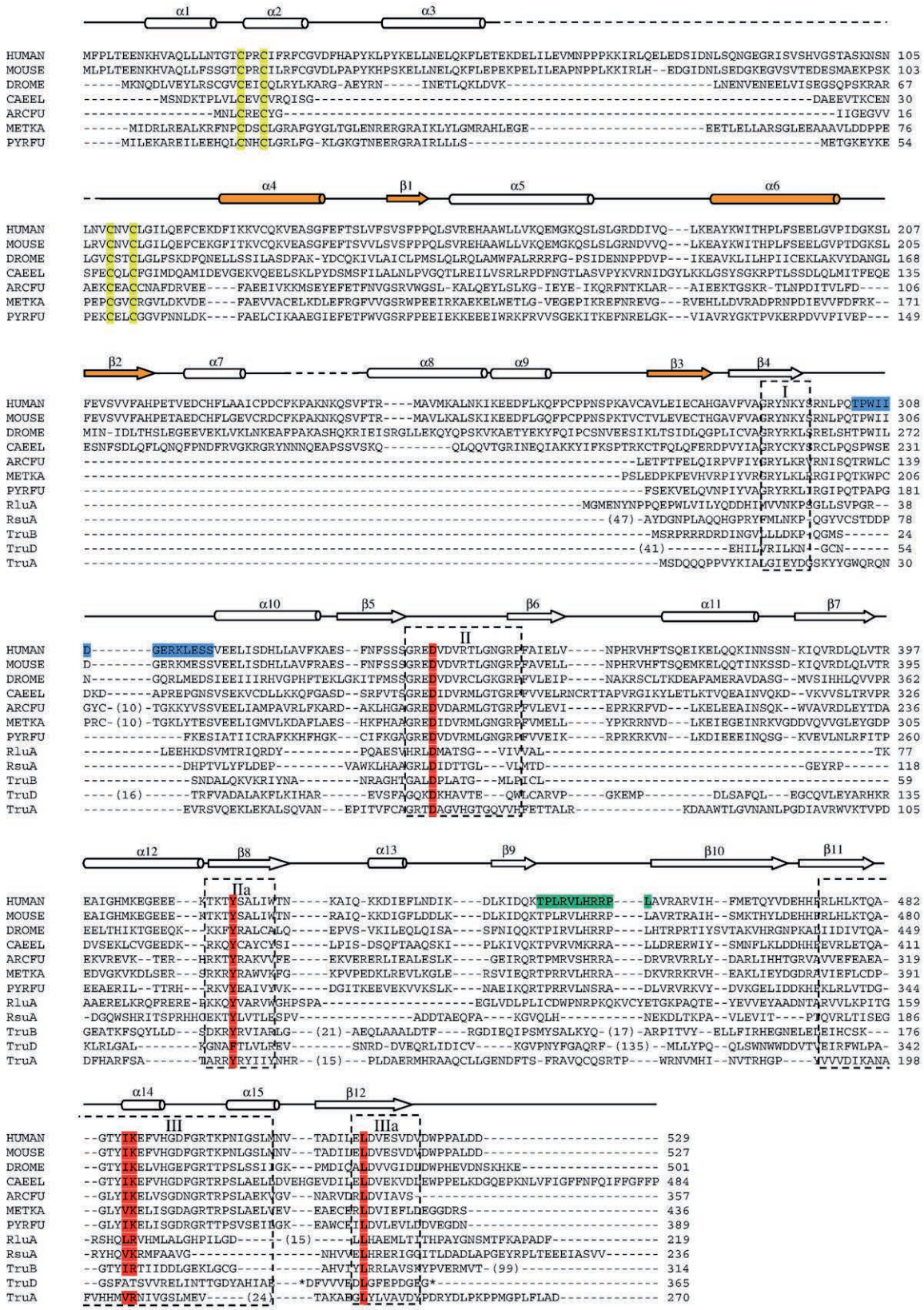


Figure i15 : Alignement de séquence des ARN:Ψ-synthases. (McClerverty *et al.,* 2007).
Les éléments de structures secondaires appartiennent à PUS10 humaine avec les hélices α représentées par des cylindres, les brins β par des flèches ; les éléments colorés en orange appartiennent au domaine THUMP et les lignes brisées correspondent à la structure désordonnée de la protéine. Les cystéines qui coordonnent le zinc sont en jaune, les boucles du pouce et de l'index en vert et bleu respectivement. Les homologues sont montrés pour la souris (MOUSE), la drosophile (DROME), *C. elegans* (CAEEL), et pour les Archaea *M. kandleri* (METKA) et *P. furiosus* (PYRFU). L'alignement de PUS10 est montré avec les autres membres des ARN:Ψ-synthases issus de *E. coli* : RluA, RsuA, TruB, TruD et TruA. Les motifs I, II, IIa, III et IIIa sont représentés par des encadrés en pointillé avec les résidus catalytiques en rouge. Les espaces sont représentés par un tiret et les larges insertions qui ont été omises sont indiquées par un nombre de résidu entre parenthèse.

A ce jour, la question du type de mécanisme catalytique de la réaction de pseudouridylation n'est toujours pas résolue, même si l'attaque en C6 rassemble néanmoins le plus grand nombre d'arguments en sa faveur (Fig. i14 A).

C. Structure 3D des ARN:Ψ-synthases

Sur la base de l'alignement des séquences d'ARN:Ψ-synthases appartenant à chacune des six familles, 5 motifs dont la séquence est conservée peuvent être dégagés (Fig. i15). Il s'agit des motifs I, II, IIa, III et IIIa, permettant ainsi de regrouper toutes ces enzymes au sein de cette classe d'enzymes possédant des caractéristiques structurales communes (Del Campo *et al.*, 2004).

Chacune des 6 familles d'ARN:Ψ-synthases comportent au moins un représentant dont la structure 3D a été résolue (Tab. i3). La structure 3D de TruA a été la première structure établie parmi les ARN:Ψ-synthase (Foster *et al.*, 2000). Puis, par ordre chronologique, ont été décrites les structures de TruB, RsuA, RluC, RluD, TruD, aCBF5, RluA, RluF, RluE et PUS10 (Tab. i3). Seules les structures 3D de RluA, RluF, TruA, TruB et aCBF5 ont été obtenues sous forme de complexes avec un ARN. Il s'agissait de segments d'ARN mimant respectivement la tige-boucle anticodon (ACSL), un segment de l'ARNr, la tige-boucle du bras TΨC (TSL), un ARNt entier et enfin un sRNA à boîtes H/ACA (Tab. i3).

Malgré la faible conservation de séquence entre les différentes familles d'ARN:Ψ-synthases, la superposition des structures établies révèle que celles-ci possèdent 2 domaines : un domaine catalytique et un domaine additionnel qui est propre à chacune des familles d'enzymes, à l'exception de TruA qui contient uniquement un domaine catalytique mais qui est active sous forme dimérique (pour revues, (Hamma et Ferre-D'Amare, 2006; Hur *et al.*, 2006)). Le domaine catalytique est constituée de 2 feuillets β anti-parallèles qui sont reliés entre-eux par des liaisons hydrogènes, pour former au site actif un plancher de 8 brins β flanqués d'hélices et de boucles (Fig. i16 A). L'aspartate catalytique est systématiquement au centre d'une

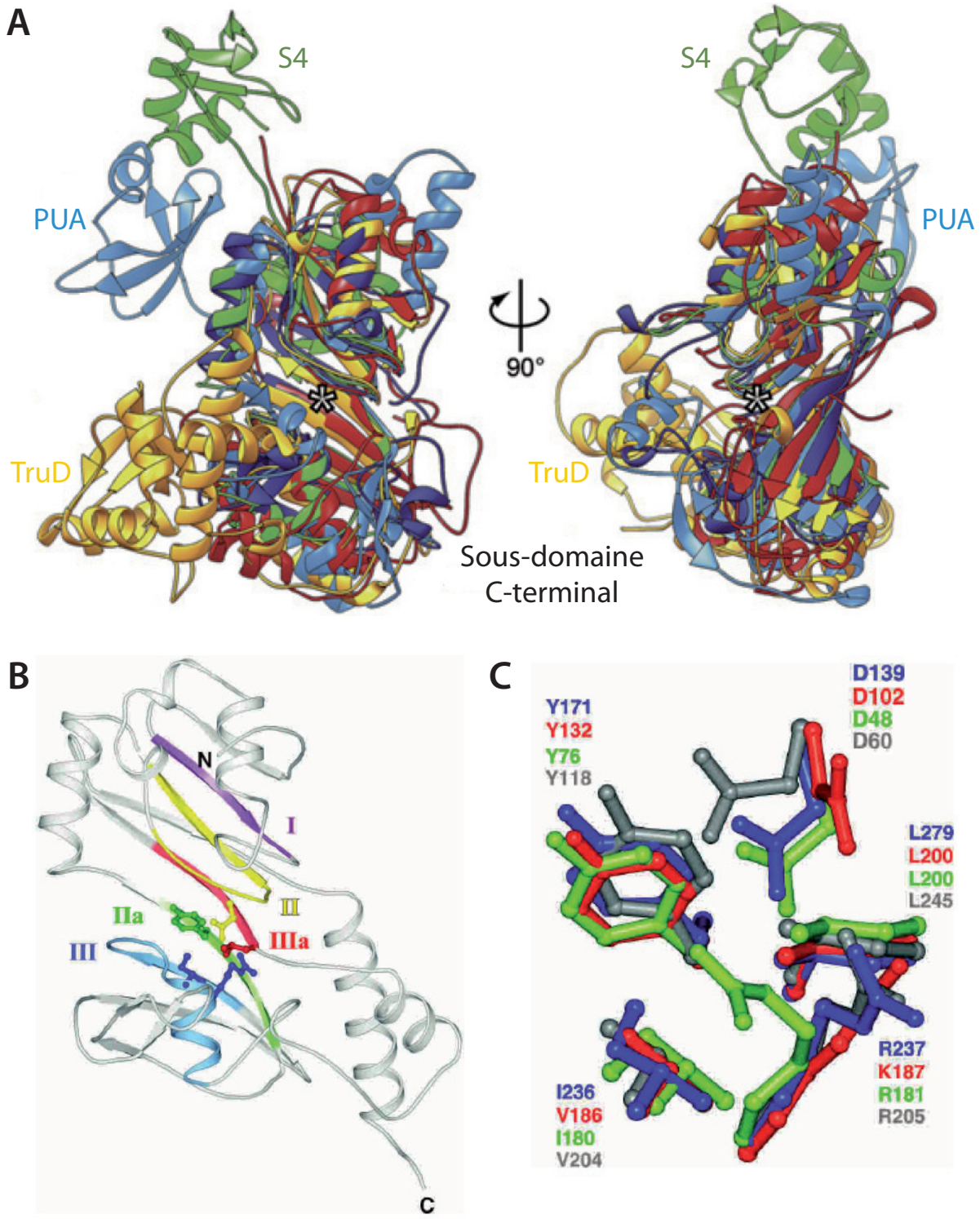


Figure i16 : Comparaison des structures 3D de différents membres des familles d'ARN:Ψ-synthases et présentation du domaine catalytique et de résidus conservés du site actif. (Ferre-d'Amarre, 2006 ; Del Campo *et al.*, 2004).

A. Superposition des modèles en ruban des structures 3D de TruA (rouge), TruB (bleu clair), TruD (jaune), RluA (bleu foncé) et RsuA (vert). Le domaine central est très similaire et permet la mise en exergue i) des domaines périphériques TruD, PUA et S4 appartenant respectivement aux familles TruD, TruB et RsuA et ii) du sous-domaine C-terminal. Le site catalytique est indiqué par un astérisque.

B. Structure 3D de l'enzyme RluD (famille RluA) mettant en évidence la structure du domaine catalytique. Les différents motifs sont représentés : motif I (violet), motif II (jaune), motif IIa (vert), motif III (bleu) et motif IIIa (rouge).

C. Superposition des chaînes latérales des résidus du site actif de RsuA (rouge), TruB (vert), TruA (gris) et RluD (bleu).

crevasse dont la surface est couverte de résidus basiques. Il est orienté de telle façon que sa chaîne latérale soit dirigée vers le solvant. Des domaines additionnels, présentant une variété de repliements, viennent se greffer aux extrémités N et C terminales tels que les domaines PUA, S4, TruD ou THUMP (Fig. i16 A) (pour revues, (Hamma et Ferre-D'Amare, 2006; Hur *et al.*, 2006)).

Dans la partie III.D, le domaine catalytique contenant le site actif sera détaillé, puis la partie III.E présentera la spécificité de reconnaissance de ces enzymes avec l'intervention des domaines additionnels.

D. Les motifs structuraux et le site actif des ARN:Ψ-synthases

Les motifs I, II, IIa, III et IIIa apportent des résidus conservés dans le site actif et ils contribuent au repliement du cœur hydrophobe des ARN:Ψ-synthases par l'intermédiaire de résidus apolaires (Fig. i16 B) (Spedaliere *et al.*, 2000; Mueller, 2002; Del Campo *et al.*, 2004). Cinq résidus du site actif sont conservés et leur chaîne principale se superpose dans les enzymes RsuA, TruB, TruA et RluD (Fig. i16 C) (Del Campo *et al.*, 2004). Dans un premier temps, voici une description du rôle général joué par ces résidus. Comme nous l'avons déjà vu précédemment, le résidu catalytique est un acide aspartique (motif II), un résidu basique (arginine ou lysine du motif III) forme un pont salin avec l'aspartate catalytique et avec une tyrosine (motif IIa). Cette tyrosine permet de réaliser une interaction avec l'uracile cible par empilement de bases. Il y a aussi deux résidus hydrophobes conservés dont le rôle est moins connu : une leucine (motif IIIa) et le second résidu qui peut être une isoleucine ou une valine (motif III) (Del Campo *et al.*, 2004; Hamma et Ferre-D'Amare, 2006). La liste des résidus est répertoriée dans le tableau i4.

Pour recentrer sur le travail effectué au cours de ma thèse, la suite de la description porte sur le site actif de la famille d'ARN:Ψ-synthase TruB. La superposition des structures 3D de membres de la famille TruB révèle la position conservée des résidus du site actif qui ont été décrits comme jouant un rôle important

	ARN:pseudouridine-synthase							Rôle proposé à partir des études structurales
Motif	**RluA**	**RsuA**	**TruA**	**TruD**	**PUS10**	**TruB**	**aCBF5**	
I	Lys28	Lys67	Tyr18	Lys50	Lys295	Lys19	Lys53	?
II	-	-	-	Phe75	-	His43	His77	Stabilisation de de base
	Arg62	Arg100	Arg58	-	-	-	-	Stabilisation de de base
	Asp64	Asp102	Asp60	Asp80	Asp344	Asp48	Asp82	Acide aspartique catalytique
	Gly69	Gly107	Gly65	Thr85	Thr349	Gly53	Gly87	?
IIa	Lys94	Lys130	Arg116	Asn129	Lys412	Lys74	Lys108	?
	Tyr96	Tyr132	Tyr118	Phe131	Tyr414	Tyr76	Tyr110	Empilement de base avec l'uridine cible
III	His162	His184	His202	-	-	-	-	?
	-	-	-	Phe345	Tyr485	Tyr179	Tyr179	Empilement de base avec l'uridine cible
	Leu164	Val186	Val204	Ala330	Ile486	Ile180	Ile180	?
	Arg165	Lys187	Arg205	Thr331*	Lys487	Arg181	Arg181	Pont salin avec l'acide aspartique catalytique
IIIa	Leu195	Leu200	Leu245	Leu33	Leu494	Leu200	Leu200	?
	-	-	-	Lys21*	-	-	-	Pont salin avec l'acide aspartique catalytique
	-	Arg202	-	-	-	Arg202	Arg202	Contacte le phosphate 5' de l'uridine cible

Tableau i4 : Localisation et rôle suggéré par la structure 3D des résidus appartenant aux 5 motifs conservés entre les 6 familles d'ARN:Ψ-synthases.
La numérotation des résidus provient de l'alignement de séquence des ARN:Ψ-synthases réalisé par McCleverty *et al.,* 2007 de la figure i15.
Les résidus sur fond vert correspondent aux 5 résidus les mieux conservés. La protéine aCBF5 sur fond bleu est un représentant de la famille TruB, dont la numérotation des résidus correspond à celle de l'archae *P. abyssi*. Les protéines RluA, RsuA, TruA, TruD et TruB portent les numérotations d'*E. coli* ; PUS10 est numérotée selon l'espèce *H. sapiens*. * : le résidu Lys21 de TruD compenserait l'absence du résidu basique (Arg181 chez TruB) qui est remplacé par le résidu Thr331.

dans l'activité de l'enzyme (Fig. i17 A et B). Dans la suite du manuscrit, la numérotation des résidus correspond à celle employée pour TruB d'*E. coli*, celle de aCBF5 de *P. abyssi* est indiquée entre parenthèse.

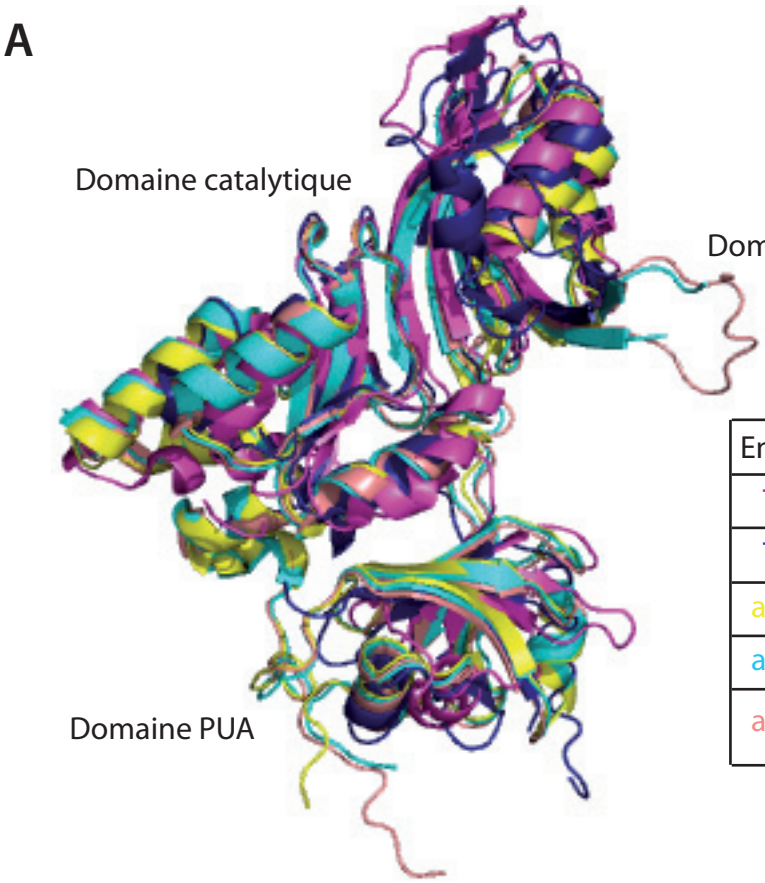
1. Le motif I assure un rôle structural

Une investigation a été menée pour déterminer le rôle joué par certains résidus conservés du motif I à la fois sur TruB et sur RluA (Spedaliere *et al.*, 2000). L'hypothèse initialement proposée par les auteurs reposait sur l'implication de résidus du motif I dans la catalyse. Les résidus Lys19 (aCBF5 K53) (Fig. i17 B) et Pro20 (aCBF5 P54) présents dans le motif I de TruB ont été mutés. La lysine pouvait jouer un rôle de donneur acide/base dans la catalyse et la proline via sa conformation unique parmi les acides aminés était supposée orienter la chaîne latérale de la lysine adjacente dans une position favorable à son rôle catalytique (Spedaliere *et al.*, 2000). La substitution de la lysine ou de la proline diminue de façon dramatique la stabilité de TruB ou de RluA. Ces résultats indiquent que la conformation globale de l'enzyme repose sur la structuration du motif I. Celui-ci jouerait donc un rôle plutôt structural que directement en lien avec la réaction catalytique (Spedaliere *et al.*, 2000).

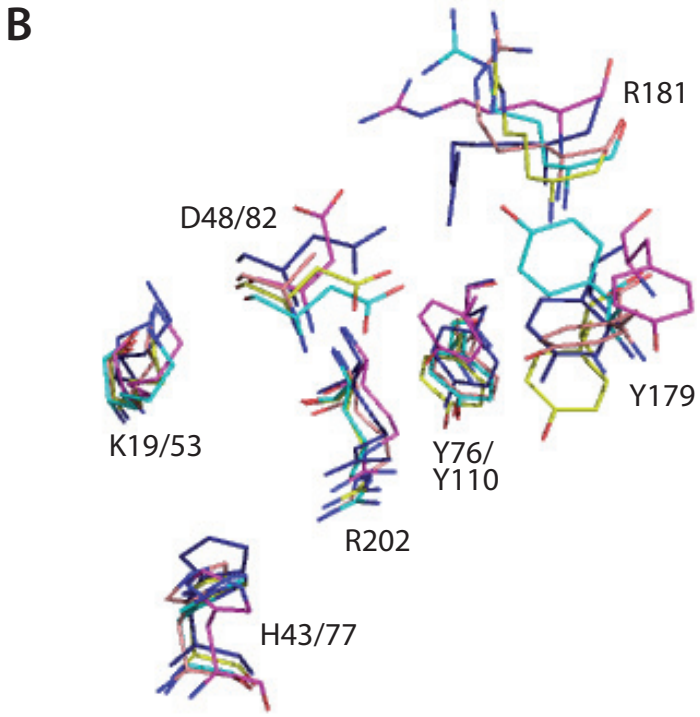
2. Le motif II intervient directement dans la catalyse

2.1. L'histidine 43 (aCBF5 H77)

Les propriétés de la chaîne latérale de l'histidine 43 en font un candidat pour agir en tant qu'acide/base générale ou donneur/accepteur de liaison hydrogène lors de la réaction (Fig. i17 B et i18) (Hamilton *et al.*, 2005). L'établissement de la structure 3D de TruB permet de constater que l'His43 est située proche du résidu U cible de la réaction. L'His43 s'empile sur la paire de base réverse Hoogsteen U54 : A58 ce qui permet d'exposer le résidu U55 cible de TruB (voir le chapitre III.E.1.) (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003). Des études biochimiques confirment que



Enzyme	Substrat	Organisme
TruB	libre	*T. maritima*
TruB	ARNt TSL	*T. maritima*
aCBF5	libre	*P. abyssi*
aCBF5	sRNA H/ACA	*P. furiosus*
aCBF5	sRNA H/ACA ARN substrat	*P. furiosus*



Résidu	*E. coli*	*P. abyssi*
Lysine	19	53
Histidine	43	77
Aspartate	48	82
Tyrosine	76	110
Tyrosine	179	179
Arginine	181	181
Arginine	202	202

Figure i17 : Les ARN:Ψ-synthases de la famille TruB.
A. Superposition des structures 3D des enzymes de la famille TruB. TruB libre (rose foncé, 1R3F), TruB liée au bras TΨC d'un ARNt (bleu foncé, 1R3E), aCBF5 libre (jaune, 2AUS), aCBF5 liée au sRNA (bleu clair, 2HVY) et aCBF5 liée à une sRNP tronquée (rose clair, 2RFK).
B. Superposition des résidus du site actif des enzymes de la famille TruB selon la même orientation qu'en **A** : K19/K53, H43/H77, D48/D82, Y76/Y110, Y179, R181, R202. Il faut noter que le résidu D82 présent dans la structure de Liang, B. *et al.,* 2007 est une alanine. Même code couleur qu'en **A**. Réalisé à l'aide du logiciel PyMol à partir des structures de Pan *et al.,* 2003 ; Manival *et al.,* 2006 ; Li et Ye, 2006 ; Liang, B. *et al.,* 2007.

l'His43 pourrait remplir ce type de fonction. En effet, la substitution H43F ne conduit qu'à très peu de perte d'activité de TruB indiquant que la fonction reposerait essentiellement sur la présence d'un noyau imidazole ou indole. A l'inverse, les substitutions H43A ou H43G abolissent l'activité mais qui peut être partiellement restaurée en présence de noyaux imidazoles (Hamilton *et al.*, 2005). Ainsi, le rôle de l'histidine 43 serait de favoriser l'exposition au site actif de l'enzyme du résidu U cible de la réaction.

2.2. L'aspartate 48 catalytique (aCBF5 D82)

La substitution de cet acide aspartique parfaitement conservé par une alanine ou une cystéine abolit l'activité des ARN:Ψ-synthases TruA, TruB et RluA tout en préservant leur capacité à se lier à l'ARNt substrat, ce qui a conduit à définir ce résidu comme le résidu catalytique (Fig. i17 B et i18) (Huang *et al.*, 1998; Ramamurthy *et al.*, 1999a; Ramamurthy *et al.*, 1999b). Néanmoins, lorsque l'Asp48 est remplacé par un analogue d'acide sulfinique (substitution D48C et sous une forme oxydée), l'activité de TruB est partiellement restaurée (30 fois moins active que TruB sauvage) (Hamilton *et al.*, 2005). Dans ce variant D48C oxydé, le groupement acide sulfinique (R-SOOH) est une base peu efficace et un moins bon nucléophile que l'acide aspartique (R-COOH). La vitesse de réaction (kcat) de TruB D48C oxydé a été montré être dépendante du pH, ce qui souligne un rôle de base générale de l'Asp48 pour déprotoner de dernier intermédiaire du mécanisme et générer la pseudouridine en plus de son rôle de nucléophile joué dans l'étape 6 du mécanisme (Fig. i14 A ou B étape 6) (Hamilton *et al.*, 2005).

2.3. La tyrosine 76 (aCBF5 Y110)

Lors de l'incubation du variant de TruB Y76F avec un substrat mimant la tige-boucle TΨC, aucune formation de pseudouridine n'a été détectée. Par contre, en présence de 5-FU, celui-ci est réarrangé et hydraté en 5-F-6-OHΨ, comme avec

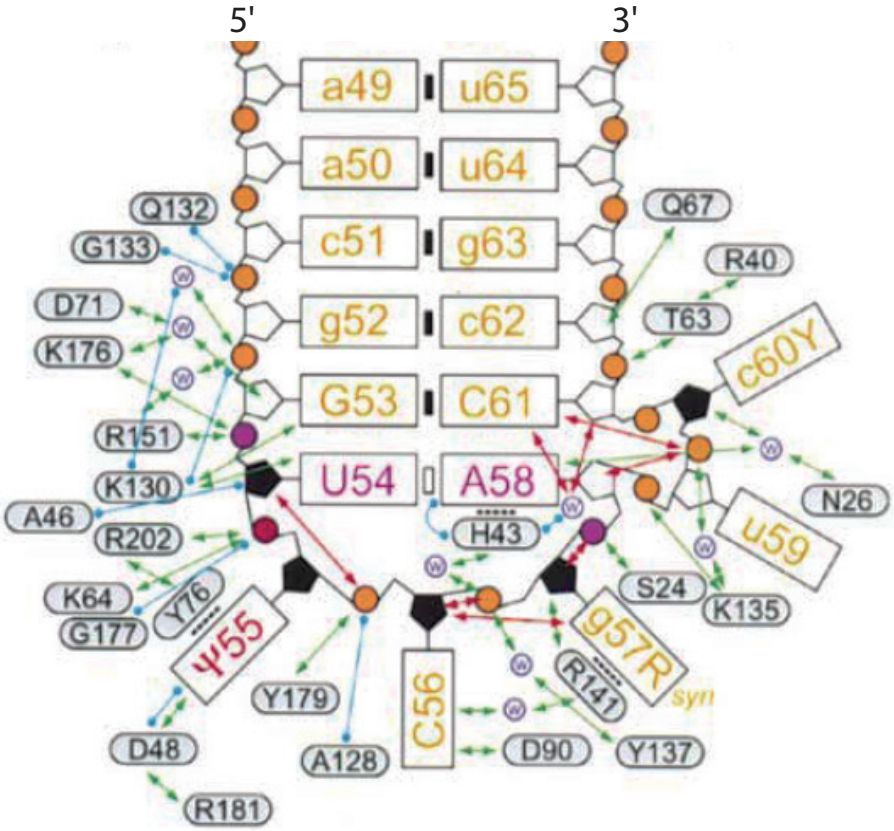


Figure i18 : Représentation schématique de la tige-boule TΨC de l'ARNt en interaction avec TruB et de l'ensemble des interactions protéine-ARN et ARN-ARN identifiées. (Hoang et Ferre-d'Amarre, 2001).
Structure du complexe TruB-ARNt d'*E. coli*. Les résidus conservés au sein de l'ARN sont écrits en majuscule et R et Y désignent respectivement la conservation d'un résidu purine ou pyrimidine.

l'enzyme TruB sauvage (Phannachet *et al.*, 2005). Dans la structure 3D de TruB sauvage, la tyrosine 76 se retrouve empilée sur l'uridine cible et son groupement phénolique permet la formation d'une liaison hydrogène avec l'atome de fluor du 5-FU qui a été transformé en 5-F-6-OHΨ (Fig. i17 B et i18) (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003; Phannachet et Huang, 2004). Dans la structure cristalline du variant de TruB Y76F, le composé 5-F-6-OHΨ est également formé, très peu de 5-F-6-OHΨ est observé lors de l'emploi du variant Y76A et aucun résidu Ψ n'est détecté à partir du substrat sauvage et des variants de TruB Y76F ou Y76A (Phannachet *et al.*, 2005). Ainsi, il a été conclu que la tyrosine 76 joue un double rôle dans le site actif. Elle maintiendrait l'intégrité de la structure du site actif par son cycle aromatique et agirait sous forme de base générale via son groupement OH^- pour la déprotonation finale requise dans l'étape 6 du mécanisme catalytique (Fig. i14 A ou B étape 6) (Phannachet *et al.*, 2005). Néanmoins, comme nous l'avons vu dans le paragraphe précédent, l'équipe d'Eugène Mueller propose que ce rôle de base générale soit joué par l'aspartate catalytique (Hamilton *et al.*, 2005). Ainsi, l'identité de résidu assurant ce rôle est toujours controversée.

3. Le motif III participe à la catalyse

A ce jour dans la littérature, aucune étude biochimique n'a été menée sur les résidus du motif III ou IIIa. Ceci est lié à des difficultés pour produire des protéines TruB variantes stables portant des substitutions de résidus appartenant à ce motif. Leur rôle proposé repose donc sur l'étude des structures 3D des ARN:Ψ-synthases et en particulier de TruB qui est l'enzyme bactérienne la plus proche de aCBF5 d'archae.

3.1. La tyrosine 179 (aCBF5 Y179)

Les structures 3D obtenues pour la protéine TruB révèlent que le résidu tyrosine 179 interagit avec le phosphate 3' du U cible de la réaction (Fig. i17 B et i18)

(Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003). Dans la structure 3D obtenue par Hoang et Ferre d'Amarre, le noyau phényl est empilé avec le ribose du résidu U cible (Hoang et Ferre-D'Amare, 2001) ; tandis que dans la structure résolue par Pan et collaborateurs, il s'empile avec la base de l'uracile cible (Pan *et al.*, 2003). Ainsi le rôle proposé pour le résidu Tyr179 serait de stabiliser et d'orienter le U cible vers le résidu catalytique (Pan *et al.*, 2003). Il a également été proposé que la Tyr179 pourrait avoir, par le biais du groupement OH^-, une fonction en tant qu'acide/base générale dans l'étape 6 de la réaction d'isomérisation (Fig. i14 A ou B étape 6) (Hamilton *et al.*, 2005).

3.2. L'arginine 181 (aCBF5 R181)

Les structures 3D de TruB indiquent que le résidu arginine 181 réalise un pont salin avec l'aspartate catalytique pouvant activer le potentiel nucléophile de l'aspartate en déprotonant le carboxylate (Fig. i17 B et i18) (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003). Cependant cette interaction serait perdue lorsque l'aspartate réalise l'attaque nucléophile en créant un intermédiaire covalent soit avec la base, soit avec le ribose de l'uridine cible de la réaction (Hoang *et al.*, 2005). Il est proposé que la perte du pont salin nécessiterait une réorganisation structurale qui compenserait la perte de la charge et serait favorable à la formation une liaison hydrogène pour l'Arg181, ce réarrangement pourrait ainsi se propager au domaine du pouce de TruB (Hoang *et al.*, 2005). Le résidu R181 pourrait constituer un « loquet » du site actif qui indiquerait au domaine du pouce que la catalyse est terminée (Hoang *et al.*, 2005).

3.3. L'arginine 202 (aCBF5 R202)

Les structures 3D de TruB indiquent que le résidu arginine 202 interagit directement avec résidu U cible de la réaction (Fig. i17 B et i18) (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003) ainsi qu'avec le résidu tyrosine 76 (Hoang et Ferre-D'Amare, 2001). Aucun rôle n'a explicitement été proposé pour l'arginine 202.

E. Spécificité de reconnaissance des ARN:Ψ-synthases

Il existe 2 systèmes de reconnaissance par lesquels les ARN:Ψ-synthases et les 2'-*O*-méthylases reconnaissent spécifiquement leur substrat. Dans le premier système, qui concerne des enzymes de composition totalement protéique, celles-ci reconnaissent le résidu à modifier au sein d'une séquence ou d'une structure particulière de l'ARN. Le second système alternatif est basé sur l'utilisation d'un ARN guide qui s'hybride avec le substrat et c'est cet appariement qui définit la position exacte du résidu à modifier (abordé dans le chapitre IV). Dans ce cas, une variété d'ARN guide s'associe avec un jeu commun de protéines dont l'une est une enzyme qui va réaliser la modification et les autres sont supposées être présentes pour stabiliser cette architecture et faciliter les changements conformationnels de l'ARN qui sont nécessaires au processus de catalyse (pour revues, (Bachellerie *et al.*, 2002; Ferre-D'Amare, 2003; Hamma et Ferre-D'Amare, 2006)). Dans le cas du premier système que je vais décrire dans ce chapitre, les enzymes peuvent appartenir à l'une des 3 catégories suivantes :

- Les enzymes dites site spécifique qui ne modifient qu'un seul résidu U dans une classe d'ARN voir dans un seul ARN.

- Les enzymes dites région spécifique qui modifient plusieurs positions proches au sein d'un même ARN.

- Les enzymes à spécificité multiple qui modifient plusieurs résidus U dans des contextes différents au sein d'un ou plusieurs types d'ARN. Il est également possible de séparer, parmi cette classe d'enzyme, celles qui reconnaissent un seul ou plusieurs type(s) d'ARN. On distingue dès lors une spécificité pour un substrat donné et non plus pour un résidu U localisé dans un contexte précis.

1. Reconnaissance du substrat par le domaine catalytique

La reconnaissance d'un ARN en tant que substrat repose sur des déterminants moléculaires présents sur l'ARN et qui sont propres à chaque enzyme. Ils permettent

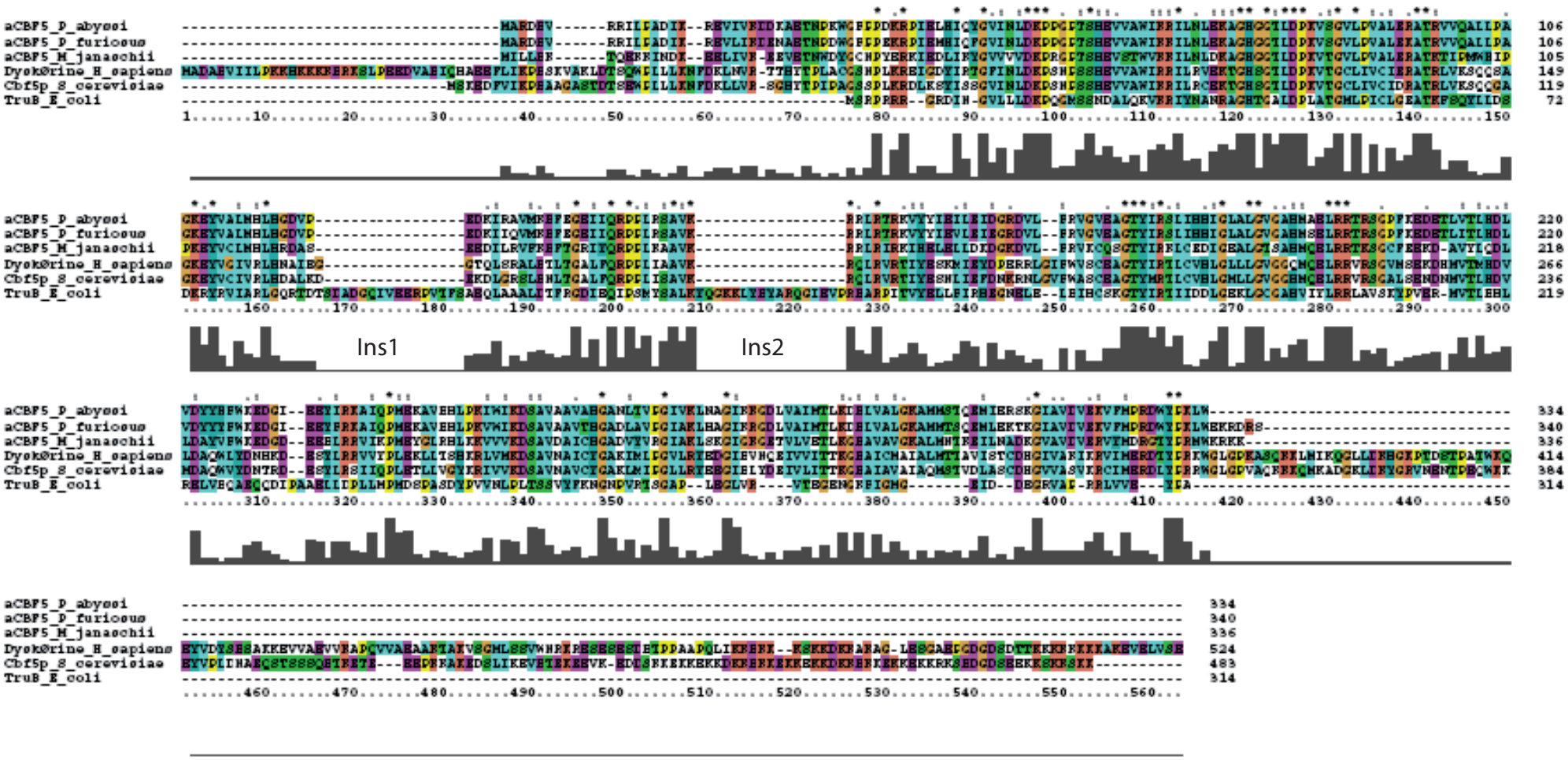


Figure i19 : Alignement de séquences des protéines de la famille d'ARN:Ψ-synthases TruB.
aCBF5 chez *P. abyssi*, *P. furiosus*, *M. janaschii* ; Dyskérine chez l'homme ; Cbf5p chez la levure et TruB chez *E. coli*. Lorsqu'ils sont conservés, les résidus basiques sont indiqués en orange foncé, les résidus acides en violet, les résidus polaires neutres en vert, les résidus hydrophobes en bleu, la glycine en orange clair et la proline en jaune. Les résidus strictement conservés portent une étoile au-dessus de l'alignement ; lorsque les résidus sont conservés avec un changement de résidu au sein de la même classe 2 points sont indiqués et un point lorsque des résidus sont conservés mais présentent un changement de classe. La fréquence de conservation des résidus au sein des différentes séquences est représentée sous l'alignement par un histogramme. Alignement réalisé à l'aide du logiciel Clustal X2 version 2.0.10 et à partir des séquences disponibles.

la formation d'un complexe enzyme-substrat ARN stable. Bien souvent le placement du résidu à modifier dans le site actif participe à la stabilisation de ce complexe et constitue un principe général sur lequel repose la spécificité de reconnaissance. Dans le cas des ARN:Ψ-synthases, l'extrusion du résidu U ciblé dans la structure ARN constitue une étape clé.

1.1. TruB et RluA : le mécanisme d'extrusion de base

1.1.1. Une enzyme site-spécifique : TruB

L'alignement de séquence de membres d'ARN:Ψ-synthases de la famille TruB révèle une grande conservation des résidus (Fig. i19). Quelques variations existent néanmoins, telles que 2 séquences d'insertion notées Ins1 et Ins2 présentent uniquement chez TruB. De plus, les séquences qui sont présentes en Nt et Ct sont propres à la famille Cbf5p eucaryote et sont donc absentes des protéines d'archaea ou de bactéries. Ces séquences sont nécessaires à la localisation nucléolaire ou à l'interaction des protéines avec les microtubules (voir le chapitre IV.C.2.1.).

TruB d'*E. coli* et son homologue Pus4p de *S. cerevisiae* appartiennent à la classe des enzymes dont la reconnaissance du substrat est site spécifique. Chez les bactéries et les eucaryotes, ces enzymes modifient uniquement la position 55 de tous les ARNt cellulaires possédant un résidu U à cette position et à l'exception des ARNt initiateurs de la traduction (Nurse *et al.*, 1995; Becker *et al.*, 1997). Les études visant à comprendre les déterminants de cette spécificité pour les enzymes TruB et Pus4p ont révélé que pour être modifié, le substrat doit correspondre à la boucle TΨC flanquée de 2 paires G-C pour la refermer et que l'identité des nucléotides U54, U55, C56 et A58 doit être strictement conservée (Fig. i20 A) (Becker *et al.*, 1997; Gu *et al.*, 1998). Des études *in vivo* ont confirmé que la perte de la structure tertiaire de l'ARNt qui s'établissait entre la boucle D et TΨC n'affectait pas la modification de la position 55 mais que des changements dans la structure de la boucle TΨC bloquaient complètement l'activité de TruB (Grosjean *et al.*, 1996).

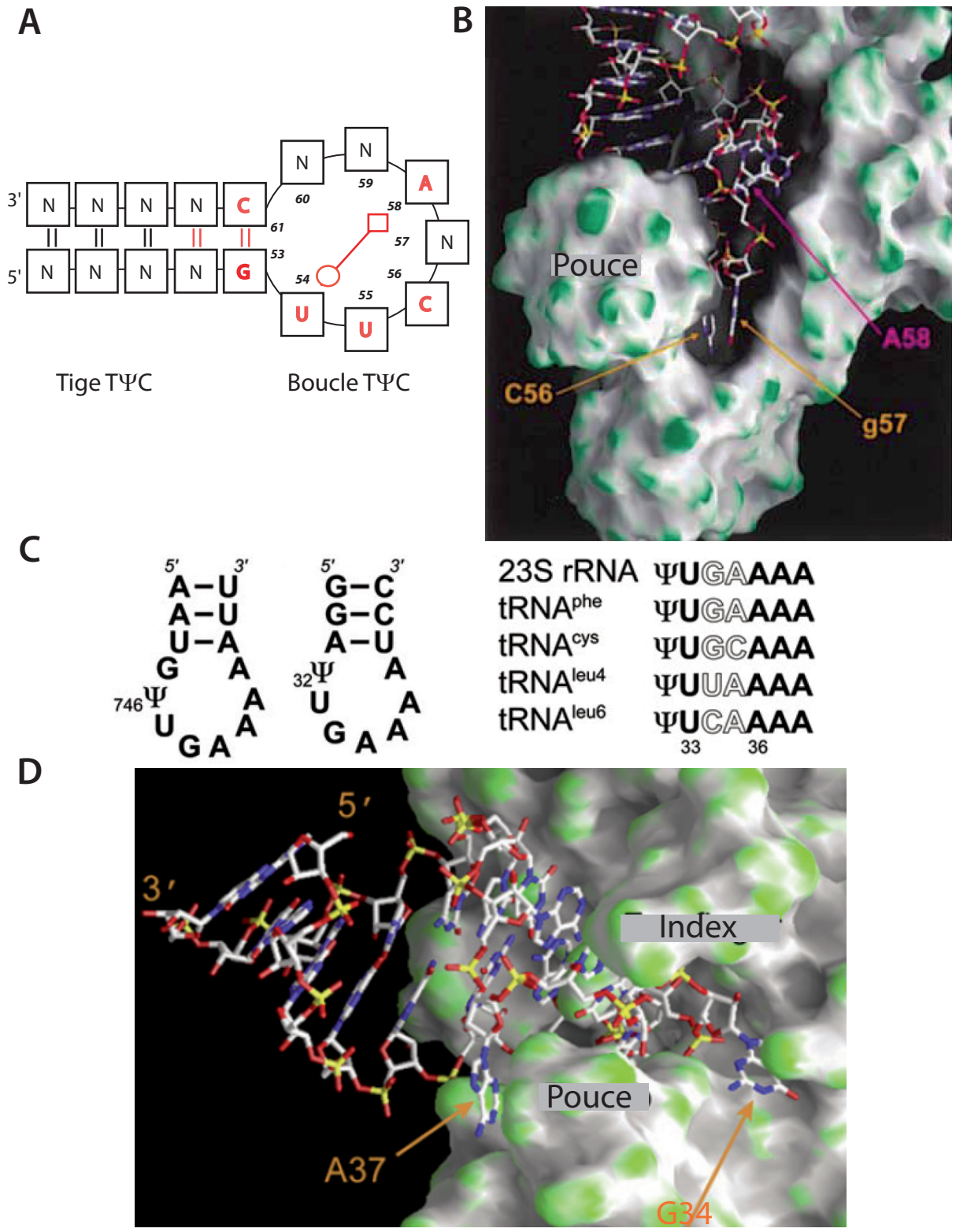


Figure i20 : Reconnaissance spécifique des substrats par TruB et RluA. (D'après Hoang et Ferre d'Amarre, 2001 (TruB) et Hoang *et al.,* 2006 (RluA)).

A. Séquence consensus du bras TΨC de l'ARNt reconnue par TruB obtenue grâce à l'étude de variants des différentes positions (d'après Gu *et al.,* 1998). Les bases et les liaisons dont l'altération affecte significativement la pseudouridylation de la position 55 sont représentées en rouge.

B. Bases moléculaires de la fixation de la tige-boucle TΨC de l'ARNt par l'enzyme TruB. Représentation 3D de l'interaction de TruB via son domaine pouce avec le grand sillon du bras TΨC. TruB accommode la tige-boucle TΨC dans une crevasse profonde formée notamment par le domaine pouce. La surface moléculaire de TruB est ici représentée et colorée selon l'incurvation (les zones vertes correspondent à des surfaces convexes et les zones grises à des surfaces concaves).

C. Séquences cibles de RluA dans l'ARNr ou dans l'ARNt^{Phe} d'*E. coli* et alignement de séquence selon les sites de pseudouridylation de RluA.

D. Représentation 3D de la surface accessible au solvant de RluA colorée selon le même mode qu'en **B**.

Les nombreuses structures 3D de TruB complexé à un fragment d'ARNt mimant la boucle TΨC ont permis de révéler les interactions établies entre TruB et sa cible (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003; Phannachet et Huang, 2004). La surface moléculaire de TruB est hautement complémentaire à la structure préformée de la tige-boucle TΨC de l'ARNt (Fig. i20 B). Celle-ci est imbriquée dans un profond renfoncement à la surface de la protéine TruB au niveau d'une crevasse principalement constituée de l'association du domaine pouce et de résidus 40 à 67 provenant à la fois du motif II et de la région en aval du motif II reconnaissant principalement le sillon majeur de l'ARN (Hoang et Ferre-D'Amare, 2001). Le domaine pouce est majoritairement constitué des résidus provenant des éléments d'insertion Ins1 (résidus 90 à 105) et Ins2 (résidus 135 à 150) qui sont les 2 principaux déterminants qui différentient la famille TruB/Pus4p, qui modifient les ARNt, de la famille Cbf5p/Dyskérine qui sont associées aux snoRNA H/ACA (Fig. i19).

La fixation du substrat induit des changements de conformation de TruB. En particulier, le pouce qui est totalement désordonné dans la forme libre devient ordonné lors de la fixation de l'ARN substrat tout comme le domaine PUA qui lui se déplace d'environ 6 Å (Pan *et al.*, 2003). La fixation de TruB au substrat se dérouleraît en 2 étapes :

- Une fixation rigide par complémentarité de forme suivie d'une structuration du pouce au contact de l'ARN.

- Une fois l'ARN et la protéine liés, une reconnaissance plus spécifique et des ajustements des résidus du site actif ont lieu.

Parmi plus de 5000 protéines purifiées de *S. cerevisiae*, Pus4p a été identifiée pour fixer préférentiellement une tige-boucle ARN riche en A qui est requise pour la réplication du virus à ARN simple brin de la mosaïque du brome de plante (Zhu *et al.*, 2007). Il a ensuite été montré que Pus4p inhibait de façon dramatique l'infection systémique de ce virus chez la plante ainsi que son encapsidation (Zhu *et al.*, 2007). Dans cette situation, il a été proposé que Pus4p entrerait en compétition avec la tige-boucle ARN naturelle du virus car celle-ci ne s'éloigne pas tellement de la séquence

minimale reconnue par Pus4p pour modifier la position 55 des ARNt, à l'exception que la boucle n'est composée que de 3 nucléotides (AUA) puis refermée par 2 paires de bases GC (Zhu *et al.*, 2007). De plus, lorsque le résidu catalytique de Pus4p est substitué, celle-ci conserve sa capacité d'inhibition de la réplication virale à un taux comparable à Pus4p sauvage, ce qui indique que Pus4p pourrait jouer un rôle de protéine chaperonne en se fixant à l'ARN (Zhu *et al.*, 2007).

1.1.2. Une enzyme à double spécificité : RluA

RluA appartient à la classe des enzymes à spécificité multiple. En effet, RluA modifie la position 32 des ARNt ainsi que la position 746 de l'ARNr 23S d'*E. coli* (Wrzesinski *et al.*, 1995b). Ces sites de pseudouridylation sont situés dans des boucles ou dans des tige-boucles qui se révèlent être des substrats reconnus et modifiés par RluA (Fig. i20 C) (Hamilton *et al.*, 2006). Des études *in vivo* indiquent que la suppression de la structure tertiaire de l'ARNt affecte la modification de la position 32 indiquant ainsi que RluA est sensible à la structure tertiaire de l'ARNt (Grosjean *et al.*, 1996). On peut noter que des structures d'autres membres de la familles RluA ont été résolues mais en absence de substrat tel que RluC (Mizutani *et al.*, 2004) ainsi que RluD (Del Campo *et al.*, 2004; Mizutani *et al.*, 2004; Sivaraman *et al.*, 2004).

La structure 3D de RluA complexée à un fragment d'ARN mimant la boucle anti-codon des ARNt révèle un mode de fixation par « pincement » de l'ARN qui s'insère entre le domaine pouce et index de RluA (Fig. i20 D) (Hoang *et al.*, 2006). Contrairement à TruB, les résidus C-terminaux du motif I ainsi que l'hélice adjacente, qui constituent le domaine index de RluA, forment une saillie qui interagit avec le sillon mineur de la boucle anti-codon. Les brins β7, β8 et l'hélice insérée entre ces 2 brins constituent le domaine pouce de RluA qui se positionne contre la face du sillon majeur de l'ARN (Hoang *et al.*, 2006). La fixation à l'ARN de RluA conduit à une réorganisation totale de la boucle anti-codon de l'ARNt par rapport à sa structure libre en solution (Shi et Moore, 2000) et ceci est détaillé dans le paragraphe suivant.

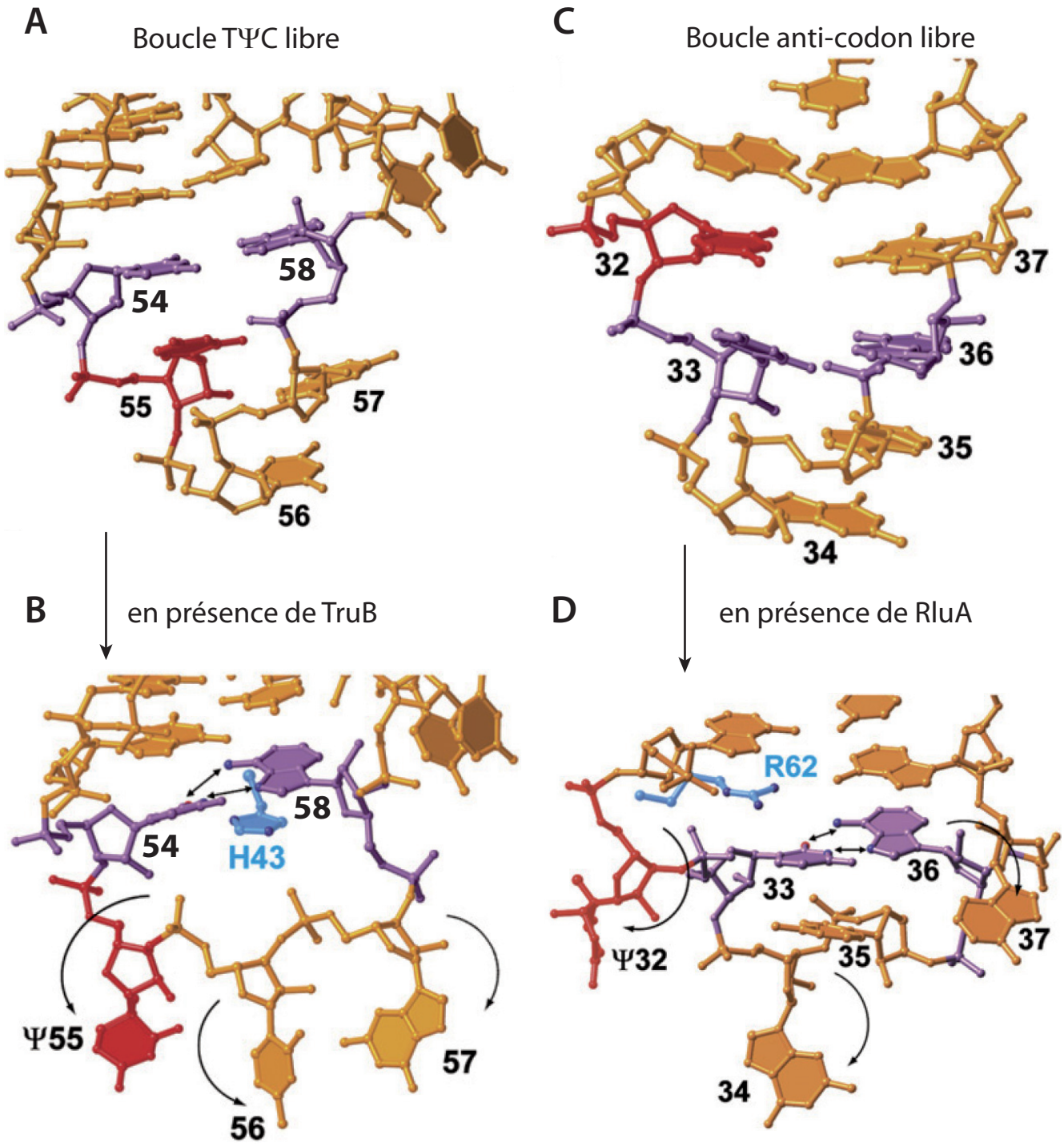


Figure i21 : Mécanisme d'extrusion de bases mis en évidence par l'étude de la reconnaissance de l'ARN par TruB et RluA. (D'après Hamma et Ferre d'Amarre, 2006).
A. et **C.** Structures 3D des boucles TΨC et anti-codon sous leur forme libre provenant de la structure 3D établie pour l'ARNt^{Phe} de *S. cerevisiae* (Shi et Moore, 2000). Les résidus U cibles de TruB et de RluA sont représentés en rouge.
B. Structure 3D de la boucle TΨC lorsqu'elle est liée à TruB (Hoang et Ferre d'Amarre, 2001). Les 3 nucléotides 55, 56 et 57 sont projetés hors de l'hélice lors de la fixation de TruB à l'ARNt. Le résidu H43 (bleu clair) de TruB reconnaît la paire de base réverse Hoogsteen (violet) et occupe l'espace laissé par le résidu G18 de la boucle D dans la structure repliée de l'ARNt libre.
D. Structure 3D de la boucle anticodon liée à RluA (Hoang *et al.,* 2006). Les 3 nucléotides 32, 34 et 37 sont expulsés hors de l'hélice et les 3 autres 33, 35 et 36 adoptent une nouvelle position empilée lors de la fixation de RluA à l'ARNt. Le résidu R62 (bleu clair) de RluA s'intercale entre la paire de base réverse Hoogsteen (violet) et la paire de base du dessus et occupe l'espace laissé par le nucléotide 32 expulsé de l'hélice.

1.1.3. Le mécanisme d'extrusion de base

Parmi les 6 classes d'ARN:Ψ-synthases deux d'entre-elles partagent le même mécanisme de reconnaissance de l'ARN : il s'agit du mécanisme d'extrusion de base employé à la fois par TruB et RluA (Fig. i21) (pour revue, (Hamma et Ferre-D'Amare, 2006)).

La reconnaissance par TruB de la boucle TΨC induit un changement de conformation des nucléotides U55 (cible de la réaction), C56 et G57 qui sont projetés hors de l'hélice formée par la tige T et sont ainsi accessibles pour des interactions avec des résidus de TruB (Fig. i21 A et B) (Hoang et Ferre-D'Amare, 2001). Cette conformation particulière de l'ARN est favorisée par la présence du résidu conservé histidine 43 qui s'intercale dans l'hélice ARN et plus particulièrement en s'empilant sur le résidu A58. La paire réverse Hoogsteen U54:A58 (i.e. interaction en *trans* de la face Watson-Crick du U avec la face Hoogsteen du A) étant invariante dans les ARNt, il a été proposé que cette paire de base caractéristique constituerait un déterminant moléculaire essentiel pour les contacts spécifiques entre le bras TΨC et TruB (Hoang et Ferre-D'Amare, 2001). En effet, l'importance de cette paire réverse Hoogsteen dans la reconnaissance du substrat par TruB corrobore l'impossibilité pour TruB de modifier les $ARNt^{Met}$ initiateurs qui ne portent pas la paire U54:A58.

Lors de la fixation de RluA, les résidus U32 (cible de la réaction), G34 et A37 sont projetés hors de l'hélice (Fig. i21 C et D). Le résidu A36 occupe l'espace laissé vacant par le résidu A37. Le résidu U33 forme un appariement réverse Hoogsteen avec le résidu A36 repositionné dans l'hélice. Enfin, le résidu A35 pivote de 180° et s'empile avec le résidu U33 (Hoang *et al.*, 2006). Par analogie avec le résidu H43 dans le cas de TruB, c'est ici la chaîne latérale de l'Arg62 de RluA qui s'insère dans l'hélice ARN et qui provoque ce changement de conformation de la boucle anti-codon.

Ce mécanisme d'extrusion de base facilite l'accès de l'enzyme au résidu à modifier et serait indispensable à la catalyse pour ces 2 classes d'ARN:Ψ-synthases. Par ailleurs, ce type de mécanisme est également observé pour des enzymes qui requièrent un accès à une base de l'ADN lors d'événements de modification ou de

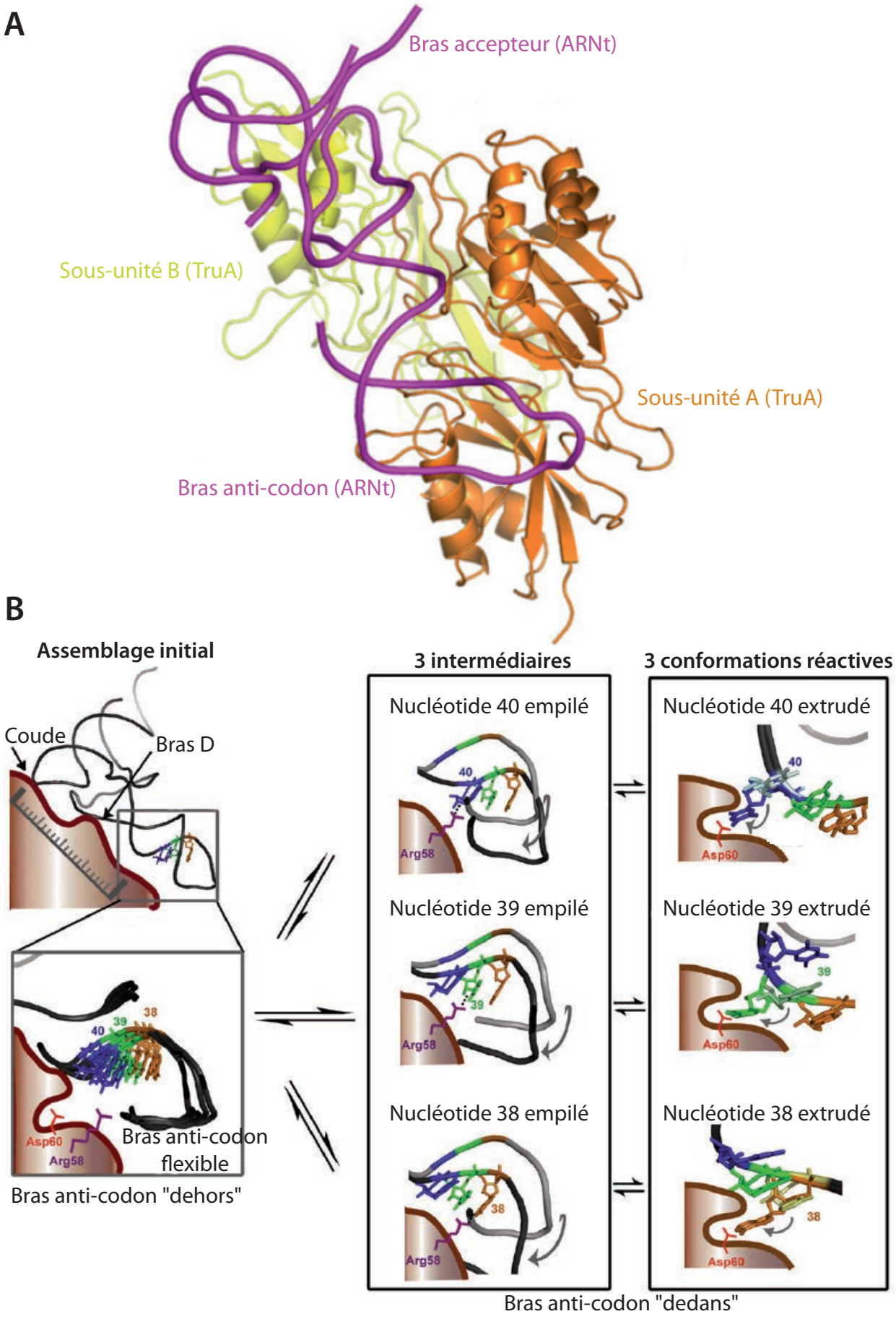


Figure i22 : Modèle proposé pour la reconnaissance du nucléotide cible par TruA. (D'après Hur et Stroud, 2007).

A. Structure 3D de TruA sous sa forme dimérique (monomères jaune et orange) en présence de son ARNt substrat (violet).

B. Modèle de reconnaissance de l'ARNt par TruA. Compte tenu de la flexibilité du bras anti-codon lors de l'assemblage initial, trois intermédiaires peuvent être formés selon l'identité du nucléotide adjacent au résidu arginine 58, ce qui conduit à trois conformations réactives.

réparation, telle que les ADN méthyl-transférases (pour revue, (Hamma et Ferre-D'Amare, 2006)).

1.2. Une enzyme région spécifique : TruA

TruA appartient à la classe des enzymes dite région-spécifique. En effet, TruA modifie les positions 38, 39 et 40 dans l'hélice du bras anti-codon des ARNt (Kammen *et al.*, 1988).

Plusieurs structures 3D de TruA ont été établies en absence de substrat (Foster *et al.*, 2000; Dong *et al.*, 2006), mais trois l'ont été sous forme de dimère et en présence d'un ARNt entier différent (Fig. i22 A) (Hur et Stroud, 2007). Dans ces structures, la boucle anti-codon adopte une conformation différente ce qui a permis aux auteurs de modéliser trois états représentatifs du mode de reconnaissance de l'ARNt par TruA (Hur et Stroud, 2007). TruA interagit sous forme de dimère avec le coude (jonction entre les boucles D et TΨC) et le bras D de l'ARNt par le biais d'interactions électrostatiques et par reconnaissance de la forme générale de la chaîne ribose phosphate de l'ARNt. Ces interactions orientent l'ARNt et positionnent le bras anti-codon à proximité du site actif (Fig. i22 B). Puis, en passant d'une conformation dite ouverte vers une conformation fermée, le résidu Arg58 vient interagir avec l'un des trois nucléotides : 38, 39 ou 40 et provoque son extrusion vers le résidu catalytique Asp60. Le rôle critique de l'Arg58 a été confirmé par des mesures d'activité du variant R58A qui est inactif tout en conservant la même affinité pour le substrat ARNt (Hur et Stroud, 2007). Le choix du nucléotide ne semble dépendre que de la flexibilité et donc des oscillations possibles de l'hélice du bras anti-codon. En effet, ces auteurs ont montré que la stabilisation et la déstabilisation de cette hélice affectaient le taux de modification de chacun des trois nucléotides. La présence d'un seul résidu Ψ modifie la stabilité de l'hélice du bras anti-codon et réduit le taux de lecture positive du codon par l'anti-codon. Un résidu Ψ altère donc le positionnement des nucléotides à proximité du résidu Arg58, et ainsi favorise ou au contraire empêche une seconde puis une troisième modification dans cette région de

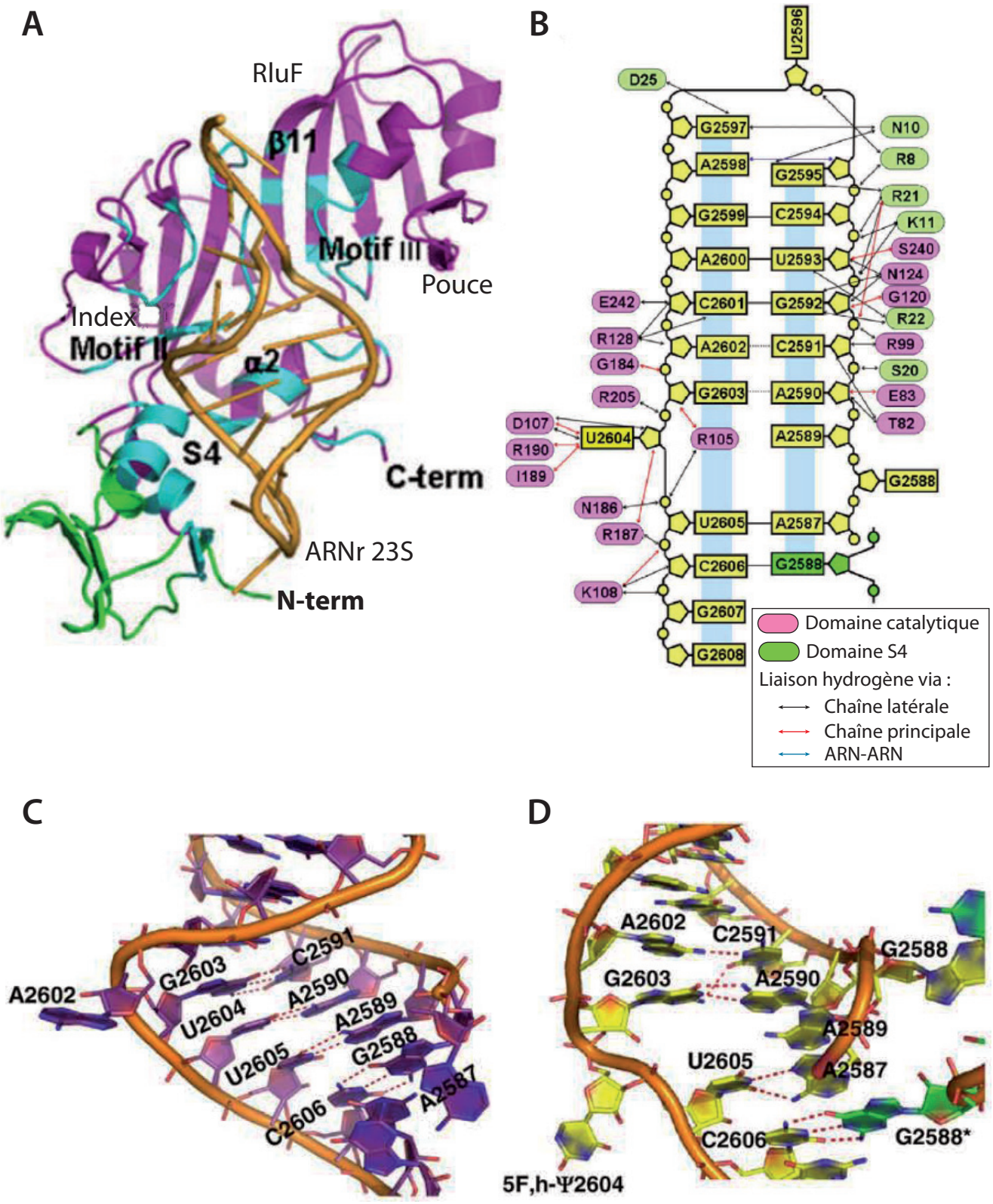


Figure i23 : Structure 3D de RluF (famille RsuA) en complexe avec un fragment d'ARNr 23S. (D'après Alian *et al.*, 2009).

A. Structure 3D de RluF avec son domaine catalytique (violet), son domaine S4 (vert) en complexe avec un ARNr substrat de 22 nts (orange). Les résidus de RluF en contact avec l'ARN sont représentés en cyan.

B. Représentation schématique de l'ARNr de 22 mer présent dans la structure 3D en complexe avec RluF. Les liaisons hydrogènes et les ponts salins sont représentés. L'arginine 105 prend la place du nucléotide cible U2604 qui est expulsé hors de l'hélice vers le résidu catalytique Asp107.

C. et **D.** Représentation tri dimensionnelle du fragment d'ARNr 23S en absence de RluF (structure dans l'ARN ribosomique seul, (**C.**)), ou en présence de RluF (**D.**) et contenant le résidu cible U2604 modifié en 5-fluoro-6-hydroxypseudouridine qui est projeté en dehors de l'hélice ARN.

l'hélice (Hur et Stroud, 2007). En résumé, les enzymes de la famille TruA ont évolué vers une reconnaissance de la flexibilité de l'hélice anti-codon. Ce mode de reconnaissance permet de distinguer les hélices stables, leur évitant une stabilisation excessive, des hélices peu stables pour lesquelles il est impossible de renforcer leur stabilité par la modification de résidus U en Ψ. Ce mode d'action permet donc d'ajuster finement la stabilité du bras anti-codon de chaque ARNt (Hur et Stroud, 2007).

1.3. Une enzyme site-spécifique : RsuA

RsuA appartient à la classe des enzymes dite site spécifique car elle modifie uniquement la position 516 de l'ARNr 16S d'*E. coli* (Wrzesinski *et al.*, 1995a). Cependant, son mode de fonctionnement est distinct de celui de TruB. La structure 3D de RsuA a été résolue sous forme libre mais contenait dans le site actif soit une molécule d'UMP ou une uracile (Sivaraman *et al.*, 2002). L'enzyme a également été obtenue sous forme de dimère dans un cristal (Matte *et al.*, 2005) mais aucun cocristal d'un complexe entre RsuA et un fragment d'ARNr n'a été résolu. Ni l'uracile, ni l'UMP ne peuvent être modifiés par les ARN:Ψ-synthases. Par modélisation moléculaire, il a été montré qu'un trinucléotide, dont le nucléotide central était superposé à l'UMP, produisait des chocs par encombrement stérique avec les résidus environnants du site actif (Sivaraman *et al.*, 2002). Cette observation suggère que le manque d'activité vis-à-vis de l'UMP pourrait résulter d'un mode de fixation différent par rapport à un substrat ARN plus long. D'autres membres de la famille RsuA ont été cristallisés mais également en absence de substrat : RluF (Sunita *et al.*, 2006) et RluE (Pan *et al.*, 2007).

Très récemment, le cocristal entre d'un complexe entre RluF qui modifie la position U2604 de l'ARNr 23S d'*E. coli* et un fragment d'ARNr a été résolue (Fig. i23) (Alian *et al.*, 2009). En plus des domaines pouce et index présents pour « pincer » l'ARN, le fragment d'ARNr 23S de 22 nts utilisé est fixé dans un sillon formé par les motifs II et III de RluF (Fig. i23 A). Une boucle du motif III se lie au sillon majeur de

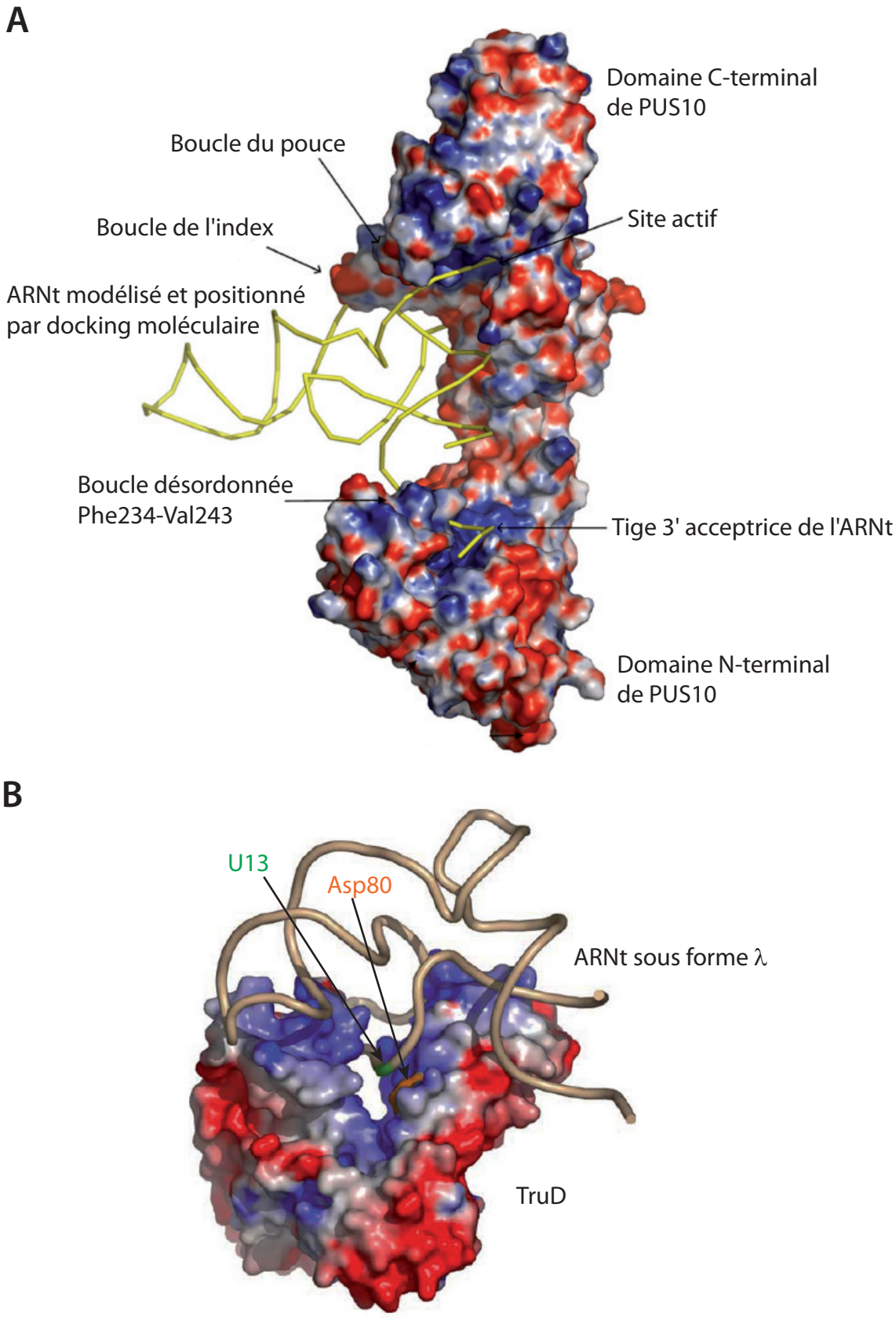


Figure i24 : Modèles d'interactions des enzymes PUS10 et TruD avec la structure d'un ARN substrat, obtenus par modélisation moléculaire. (McCleverty *et al.,* 2007 ; Ericsson *et al.,* 2004).
A. La surface électrostatique de PUS10 est représentée. L'ARNt a été modélisé et positionné par docking moléculaire dans le site actif. Dans ce modèle, les boucles du pouce et de l'index ont été réarrangées pour accommoder l'ARNt. La couleur bleue indique les charges positives et la rouge les charges négatives.
B. Représentation de la structure 3D de TruD en interaction avec un substrat ARNt modélisé sous forme λ. L'aspartate 80 catalytique est représenté en orange et le résidu cible U13 de l'ARNt est coloré en vert. Le bleu indique les charges positives et le rouge les charges négatives.

l'hélice d'ARN, tandis qu'une boucle du motif II contenant l'Asp107 catalytique ainsi que l'Arg105 conservée s'associe au sillon mineur du substrat. Le groupement guanidinium de l'Arg105 s'intercale entre les bases de l'ARN G2603 et U2605, prenant ainsi la place du résidu U2604 qui est projeté en dehors de l'hélice dans le site actif de RluF (Fig. i23 B et D) (Alian *et al.*, 2009). En effet, cette portion de l'ARNr 23S se structure sous forme d'hélice avec le résidu A2602 projeté vers le solvant, contrairement au résidu U2604 cible de RluF qui s'empile dans l'hélice d'ARN (Fig. i23 C). La fixation du domaine S4 à l'ARN induit un changement de conformation du domaine catalytique, en particulier l'hélice α2 viendrait ainsi se positionner dans le sillon majeur de l'ARN en débobinant sa partie Ct. C'est le déplacement de cette hélice α2 qui repositionnerait le résidu A2602 dans l'hélice ARN et ainsi permettrait à RluF d'expulser le résidu cible U2604 libérant son accès au site actif pour le modifier (Alian *et al.*, 2009).

1.4. Une enzyme région-spécifique : PUS10

L'enzyme PUS10 appartient aux enzymes de classe dite région-spécifique car elle modifie les résidus U54 et U55 des ARNt (Watanabe et Gray, 2000; Roovers *et al.*, 2006; Gurha *et al.*, 2007; Gurha et Gupta, 2008). Il est intéressant de souligner que PUS10 est la dernière classe d'ARN:Ψ-synthase à avoir été identifiée et c'est actuellement la seule à n'avoir aucun représentant chez les bactéries. Les études biochimiques ont été réalisées avec la protéine produite chez les archaea dont certaines propriétés pourraient être spécifiques à ce domaine du vivant, en particulier son lien fonctionnel avec aCBF5 pour la modification de la position 55 qui sera détaillé dans la partie II.B.2 des résultats.

La structure 3D de PUS10 humaine (notée FLJ32312) résolue sous forme libre indique que cette enzyme se replie en 2 domaines contenant un domaine pouce et index (Fig. i24 A) (McCleverty *et al.*, 2007). Les auteurs ont établi par modélisation moléculaire que PUS10 était suffisamment large pour accommoder la tige acceptrice de l'ARNt lorsque la boucle TΨC est insérée dans le site actif de l'enzyme et serré par

le pouce et l'index de PUS10. Des mouvements accompagnent parfois la fixation du substrat à d'autres ARN:Ψ-synthases, ce qui ne semble pas nécessaire dans le cas de PUS10 (Fig. i24 A). Le modèle place également la tige acceptrice de l'ARNt proche d'une boucle basique (Phe234 à Val243) qui est désordonnée dans la structure. Cette boucle pourrait se structurer lors de la fixation de l'ARNt pour former de nouvelles interactions afin de stabiliser le complexe enzyme-substrat (McCleverty *et al.*, 2007).

Par ailleurs, une étude a montré que PUS10 humaine jouait un rôle dans l'induction de l'apotose provoqué par la molécule Trail qui appartient à la famille du TNF et qui induit une cytotoxicité sélective des cellules tumorales. Ce facteur transduit le signal apoptotique d'un membre de la famille Bcl-2 (la protéine Bid) à la mitochondrie conduisant au recyclage du cytochrome C (Aza-Blanc *et al.*, 2003). Cela suggère que PUS10 possède une double fonction chez les eucaryotes et qui serait absente chez les archaea.

1.5. Une enzyme site-spécifique : TruD

TruD appartient à la classe des enzymes dite site-spécifique car elle modifie uniquement la position 13 des ARNt (Kaya et Ofengand, 2003). Les 3 structures 3D de TruD obtenues l'ont été en absence de substrat (Ericsson *et al.*, 2004; Hoang et Ferre-D'Amare, 2004; Kaya *et al.*, 2004). Néanmoins, un modèle d'interaction a été réalisé à partir de la structure de TruD et un ARNt modélisé sous sa forme λ (Fig. i25 B) (Ericsson *et al.*, 2004). En effet, il semble que la modification des nucléotides du bras D nécessite un dépliement partiel de l'ARNt le faisant passer d'une forme en L inversé à une forme λ comme observé dans la structure de l'ArcTGT (tRNA-guanine trans glycosylase) modifiant le nucléotide G15 (Ishitani *et al.*, 2003). Dans le modèle de TruD avec la forme λ d'ARNt, la boucle D de ce dernier est insérée dans un creux formé par les 2 domaines de TruD avec le nucléotide cible U13 proche de l'aspartate 80 catalytique du site actif (Ericsson *et al.*, 2004).

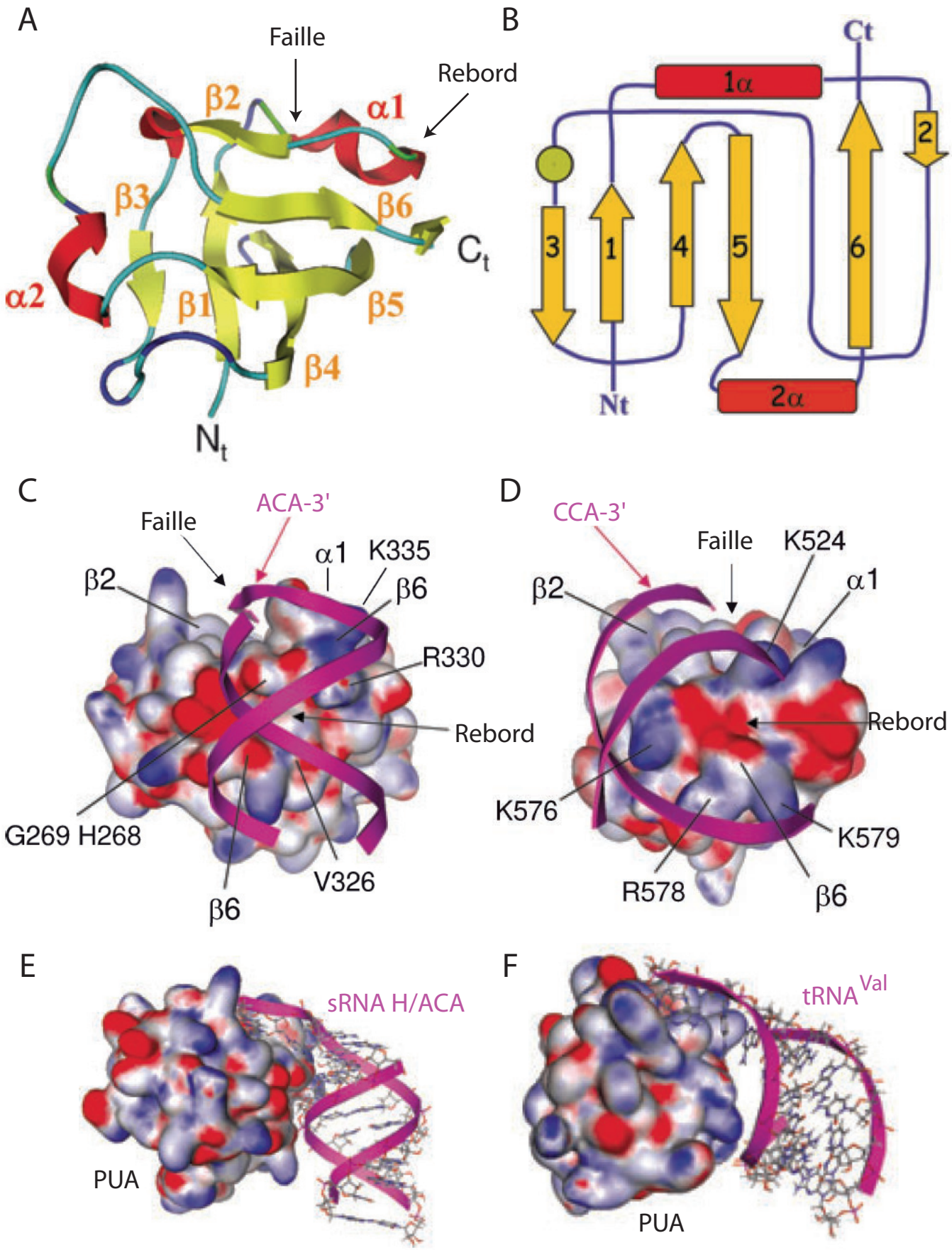


Figure i25 : Le domaine PUA : structure et mode d'interaction avec l'ARN. (D'après Pérez-Arellano *et al.,* 2007).

A et B. Structure du domaine PUA de l'enzyme ArcTGT de *P. horikoshii* (Ishitani *et al.,* 2003). Les hélices α sont indiquées en jaune, l'hélice 3.10 en vert et les brins β en rouge.

C et E. Domaine PUA de aCBF5 de *P. furiosus* en interaction avec un sRNA H/ACA (Li et Ye, 2006).

D et F. Domaine PUA de l'ArcTGT en interaction avec un $ARNt^{Val}$ (Ishitani *et al.,* 2003).

C à F. Les couleurs représentent la surface électrostatique. Le rouge correspond aux charges négatives et le bleu aux charges positives.

2. Reconnaissance du substrat par les domaines additionnels

En plus de l'interaction du domaine catalytique avec l'ARN cible, les ARN:Ψ-synthases possèdent un domaine additionnel qui est généralement utilisé comme point d'ancrage de l'enzyme à l'ARN. Parmi les 6 classes d'enzymes, 4 domaines supplémentaires au domaine catalytique sont actuellement décrits. Il s'agit du domaine PUA pour la famille TruB, le domaine S4 pour les familles RluA et RsuA, le domaine TruD propre à l'enzyme TruD et enfin le domaine THUMP spécifique à PUS10. Seul TruA ne possède pas de domaine dédié à la fixation à l'ARN. En effet, dans le cas de TruA, celle-ci interagit directement avec l'ARN via son domaine catalytique et sous forme de dimère. Le rôle de ces domaines semble être confiné à la reconnaissance de l'ARN et ainsi décharger le domaine catalytique à son rôle de catalyse et à la reconnaissance fine du résidu cible de la réaction.

2.1. Le domaine PUA

Le domaine PUA n'est pas seulement présent au sein des protéines de la famille TruB, il existe également dans d'autres protéines dont la fonction est en lien avec le métabolisme des ARN. Parmi celles-ci, l'ArcTGT (Archaeosine tRNA Guanine Transglycosylase) est une enzyme de modification de l'ARNt (Aravind et Koonin, 1999). Cette enzyme a été prise en compte pour définir l'acronyme utilisé pour ce domaine (PseudoUridine and Archaeosine tRNA guanine transglycosylase). Il a été initialement proposé que les domaines PUA de la famille TruB et ArcTGT partageraient un même rôle fonctionnel qui serait basé sur un mode de reconnaissance commun de l'ARN (Ishitani *et al.*, 2003). Dans la structure 3D de l'ArcTGT en présence de sa cible, un ARNt entier, plusieurs résidus du domaine PUA entrent en interaction à différents endroits du squelette phosphate de l'ARNt (Fig. i25). En outre, l'extrémité 3' CCA de l'ARNt est empilée avec 2 phénylalanines du domaine PUA (Ishitani *et al.*, 2003). Ces 2 résidus sont hautement conservés dans les ArcTGT. De façon assez surprenante, il a été montré que l'extrémité CCA n'était pas nécessaire à la reconnaissance de l'ARNt par ArcTGT (Watanabe et al., 2000).

Une explication possible proviendrait du fait que l'extrémité 3' CCA des ARNt soit ajoutée de façon post-transcriptionnelle par une enzyme nommée CCase (Sprinzl et Cramer, 1979; Tomita *et al.*, 2006). Une autre possibilité pourrait être apportée par les données indiquant que le domaine PUA de l'ArcTGT de *P. furiosus* n'est pas requis pour la formation d'archaeosine en position 15 de l'ARNt, mais que sa présence contribue à augmenter la spécificité de reconnaissance de l'ARNt de 75 fois et à accroître de 4 fois la vitesse de réaction (Sabina et Soll, 2006).

Le domaine PUA interagit avec l'ARN toujours par l'une de ses boucles qui est riche en glycine et qui est située entre l'hélice α1 et le brin β2, et par certains résidus toujours localisés près de la boucle du brin β6. Les brins β2–β1 forment une faille d'un côté du rebord qui est capable d'accommoder un ARN sb (Fig. i25 A et B). Il est intéressant de souligner que le domaine PUA est capable de reconnaître de l'ADN db en faisant varier la composition des résidus de α1, β2 et β6 (pour revue, (Perez-Arellano *et al.*, 2007)). Deux modes différents de reconnaissance peuvent être distingués et cela dépend de la séquence impliquée dans l'interaction avec l'ARN. Les domaines PUA de aCBF5 ou de TruB se fixent latéralement à l'ARN db en reconnaissant d'abord le sillon mineur ; tandis que le domaine PUA de l'ArcTGT reconnaît l'extrémité de la tige acceptrice db à travers le sillon majeur (Fig. i25 C à F) (pour revue, (Perez-Arellano *et al.*, 2007)).

2.2. Le domaine S4

Le domaine S4 est présent dans plusieurs classes d'enzymes en lien avec le métabolisme de l'ARN telles que la protéine bactérienne ribosomique S4 d'où lui vient son nom, les ARN:Ψ-synthases RusA et certaines enzymes de la famille RluA, la protéine de choc thermique associée au ribosome Hsp15 ou encore la thréonyl ARNt synthéthase (Aravind et Koonin, 1999). Ce domaine a ainsi été prédit comme étant un domaine de fixation à l'ARN (Aravind et Koonin, 1999).

Les premières études par radiocristallographie des membres de la famille RsuA ou RluA n'ont pas permis de valider le rôle joué par le domaine S4. En effet, ces études ont été réalisées en absence d'ARN substrat. Néanmoins, elles ont tout de même permis de montrer que le domaine S4 de RsuA est relié au domaine catalytique par un fragment qui est mobile et qui semble ainsi pouvoir s'ajuster lors de la fixation à l'ARN (Sivaraman *et al.*, 2002; Matte *et al.*, 2005). Ultérieurement, l'enzyme RluA d'*E. coli* a été cocristallisée en présence d'un ARN substrat mais cette protéine était uniquement composée du domaine catalytique, tronqué de son domaine S4 (Hoang *et al.*, 2006). Par contre, les structures 3D de RluF libre (famille RsuA) et associée à un substrat consistant en un fragment de l'ARN 23S ont été établies récemment et montrent que le domaine S4 interagit directement avec l'ARN (Alian *et al.*, 2009). En effet, environ 1/3 de la surface d'interaction RluF-ARN provient de la fixation du domaine S4 à l'ARN (représenté en cyan dans la figure i23 A). Ce domaine possède un motif hélice-tour-hélice qui interagit avec le sillon majeur de la tige ARN, tandis que d'autres résidus interagissent avec la boucle de la structure formée par l'ARNr (Alian *et al.*, 2009).

2.3. Le domaine THUMP

Le domaine THUMP (THioUridine synthase, Methylases and PSUSs) est retrouvé dans PUS10 mais également dans d'autres enzymes de modification de l'ARN telle que l'enzyme ThiI responsable de la modification 4-Thiouridine en position 8 de certains ARNt bactériens. Dans PUS10, ce domaine contient une fissure composée de nombreux résidus basiques qui serait favorable pour accommoder la tige acceptrice 3' de l'ARNt (Fig. i24 A). De plus, plusieurs résidus conservés constituant ce domaine, sont prédits pour être impliqués dans la fixation de l'ARNt (McCleverty *et al.*, 2007). Le domaine THUMP possède un motif en doigt de zinc inaccessible au solvant qui apparaît jouer un rôle structural plutôt que d'intervenir directement dans la réaction de pseudouridylation (McCleverty *et al.*, 2007).

2.4. Le domaine TruD

Le domaine TruD est spécifiquement retrouvé dans la famille TruD / Pus7p. Son rôle n'a pas encore été totalement élucidé. Néanmoins, ce domaine possède une zone fortement chargée positivement qui fait face au domaine catalytique (Ericsson *et al.*, 2004; Hoang et Ferre-D'Amare, 2004; Kaya *et al.*, 2004). Il a donc été proposé que le domaine TruD soit un domaine de fixation à l'ARN et en particulier qu'il pourrait contribuer à stabiliser l'ARNt sous une forme λ qui semble être nécessaire à la modification des nucléotides du bras D (Fig. i24 B) (Ericsson *et al.*, 2004).

3. Cas particuliers de reconnaissance

Certaines positions conservées peuvent être parfois modifiées soit par une enzyme, soit guidées par le système snoRNA H/ACA. Par exemple la position 966 de l'ARNr 25S de *S. cerevisiae* est modifiée soit par Pus7p, soit par une snoRNP portant l'ARN guide snR43 (Decatur et Schnare, 2008). La comparaison des séquences des enzymes aPUS7 des archaea de l'ordre des Sulfolobales et des Desulfurococcales a révélé des substitutions de résidus dans le site actif qui abolissent l'activité catalytique (Muller *et al.*, 2009). Chez ces archaea, la position 35 de l'ARNt est modifiée par un complexe sRNP Sso1 à boîtes H/ACA qui remplace l'action de aPUS7 déficiente (Muller *et al.*, 2009).

IV. Les systèmes de modification reposant sur les ARN guides

Comme nous l'avons vu dans le chapitre précédent les systèmes de modification purement enzymatiques reposent sur l'utilisation d'une enzyme qui va modifier de un à plusieurs nucléotides cibles dans le meilleur des cas. Les systèmes s'appuyant sur des ARN guides, qui appartiennent à la classe des ARN non codant, vont pouvoir être adoptés pour modifier des dizaines de nucléotides cibles tout en employant la même enzyme de modification. Il suffit à l'organisme de coder une seule enzyme pour des dizaines de positions à modifier, contrairement aux systèmes

purement enzymatiques qui nécessitent de coder pratiquement une enzyme par position à modifier.

Les systèmes de modification basés sur l'utilisation des ARN guides existent à la fois chez les Eucaryotes et les Archaea, mais aucun n'a été identifié chez les Bactéries. Dans ces systèmes, les modifications sont catalysées par des complexes formés entre un ARN guide et des protéines, appelés snoRNP (small nucleolar RiboNucleoParticule) ou scaRNP (small cajal bodies RiboNucleoParticule) chez les Eucaryotes et sRNP (small RiboNucleoParticule) chez les Archaea. L'ARN guide de ces particules permet de cibler la position exacte de la modification et celle-ci est déterminée par son appariement avec l'ARN cible ; c'est donc l'identité de l'ARN guide qui va varier en fonction des positions à modifier. L'une des protéines associée à l'ARN guide porte l'activité catalytique nécessaire à la formation de la modification, les autres protéines peuvent être qualifiées de « protéines accessoires » même si leur rôle est souvent capital pour que la réaction de modification puisse avoir lieu. On distingue ainsi deux grandes familles de sno-sca/sRNP qui sont réparties en fonction du type de modification réalisée: les RNP à boîtes C/D qui catalysent les 2'-*O*-méthylations et les RNP à boîtes H/ACA qui catalysent les pseudouridylations (pour revues, (Bachellerie *et al.*, 2002; Kiss, 2002; Omer *et al.*, 2003)). Au sein de ces deux familles, les sno/sRNP à boîtes C/D et les sno/sRNP à boîtes H/ACA sont les plus abondantes dans les cellules et participent dans le nucléole à la biogenèse des ribosomes. Après avoir rappelé brièvement les principes de la maturation de l'ARNr, je détaillerai dans cette partie la structure, la biogenèse et la fonction des snoRNP eucaryotes et sRNP d'archaea à boîtes C/D et H/ACA. Le cas des scaRNP sera traité au travers de l'exemple de la télomérase humaine.

A. Le nucléole et la maturation du pré-ARN ribosomique

1. Chez les Eucaryotes

Le nucléole est une structure nucléaire proéminente et dynamique qui est visible en microscopie conventionnelle par contraste de phase, ce qui lui a valu son

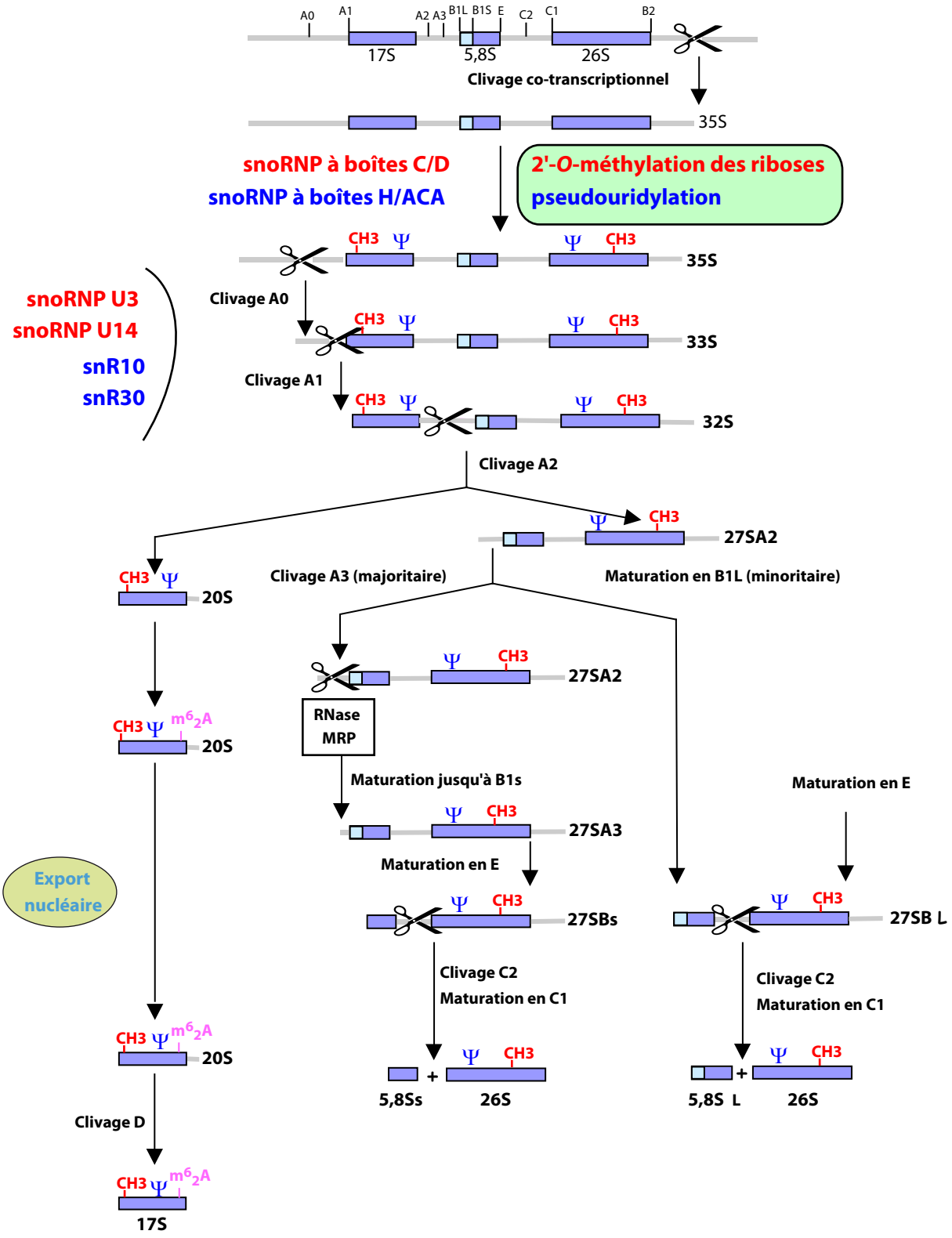


Figure i26 : Maturation du précurseur 35S des ARNr de *S. cerevisiae*. (Adapté de Venema et Tollervey, 1999).
Le schéma donne un aperçu des nombreuses étapes permettant la maturation du précurseur pré-ARNr en trois ARNr, 17S, 5.8S et 26S. Les snoRNA à boîtes C/D et les 2'-*O*-méthylations sont représentés en rouge tandis que les snoRNA à boîtes H/ACA et les pseudouridylations sont colorés en bleu.

nom d'organite bien qu'il ne soit entouré d'aucune membrane. Il est connu depuis les années 1960 comme étant le centre de synthèse et d'assemblage des ARNr qui sont transcrits par l'ARN polymérase I (pour revue, (Hernandez-Verdun et Louvet, 2004)). Chez les eucaryotes, le nucléole contient les gènes codant les ARNr. Ce sont des gènes répétés, regroupés dans des régions chromosomiques appelées régions organisatrices du nucléole (NOR) autour desquelles s'ordonne le nucléole.

Les gènes des ARNr sont transcrits sous forme d'un long précurseur, le pré-ARNr, qui est ensuite clivé pour générer les ARNr matures 18S, 5.8S et 25S/26S qui s'associent à de nombreuses protéines et à l'ARNr 5S (transcrit par l'ARN polymérase III) pour composer le ribosome. La majorité des constituants du ribosome s'assemblent au niveau du nucléole pour former les particules pré-ribosomiques. Celles-ci traversent le nucléoplasme et sortent du noyau pour rejoindre le cytoplasme où se déroule la maturation finale des ribosomes (Fig. i31) (pour revues, (Venema et Tollervey, 1999; Fromont-Racine *et al.*, 2003)). Durant la maturation du pré-ARNr, celui-ci est clivé de façon séquentielle par une cascade de digestions endo- et exonucléolytiques (Fig. i26). Les clivages précoces requièrent la présence de snoRNA à boîtes C/D tels que les snoRNA à boîtes C/D U3 et U14 ainsi que les snoRNA à boîtes H/ACA snR10 et snR30. Tous ceux-ci auraient un rôle de chaperon de l'ARN. Parallèlement à la maturation du précurseur, les séquences des ARNr sont modifiées par l'activité catalytique des snoRNP C/D et H/ACA.

2. Chez les Archaea

Le schéma général du processus de maturation des ARNr chez les archaea est différent de celui décrit précédemment chez les eucaryotes. En effet, les ARNr d'archaea sont situés sur un même transcrit pré-ARNr (Fig. i27 A) (Tang *et al.*, 2002b). Celui-ci est maturé lors de deux étapes de clivage par des endonucléases. Elles reconnaissent une structure d'un motif du type boucle-hélice-boucle (BHB) qui est composé deux trinucléotides en simple brin, séparé par une hélice de 4 paires de bases (Fig. i27 A) (Thompson et Daniels, 1990). Les endonucléases sont constituées

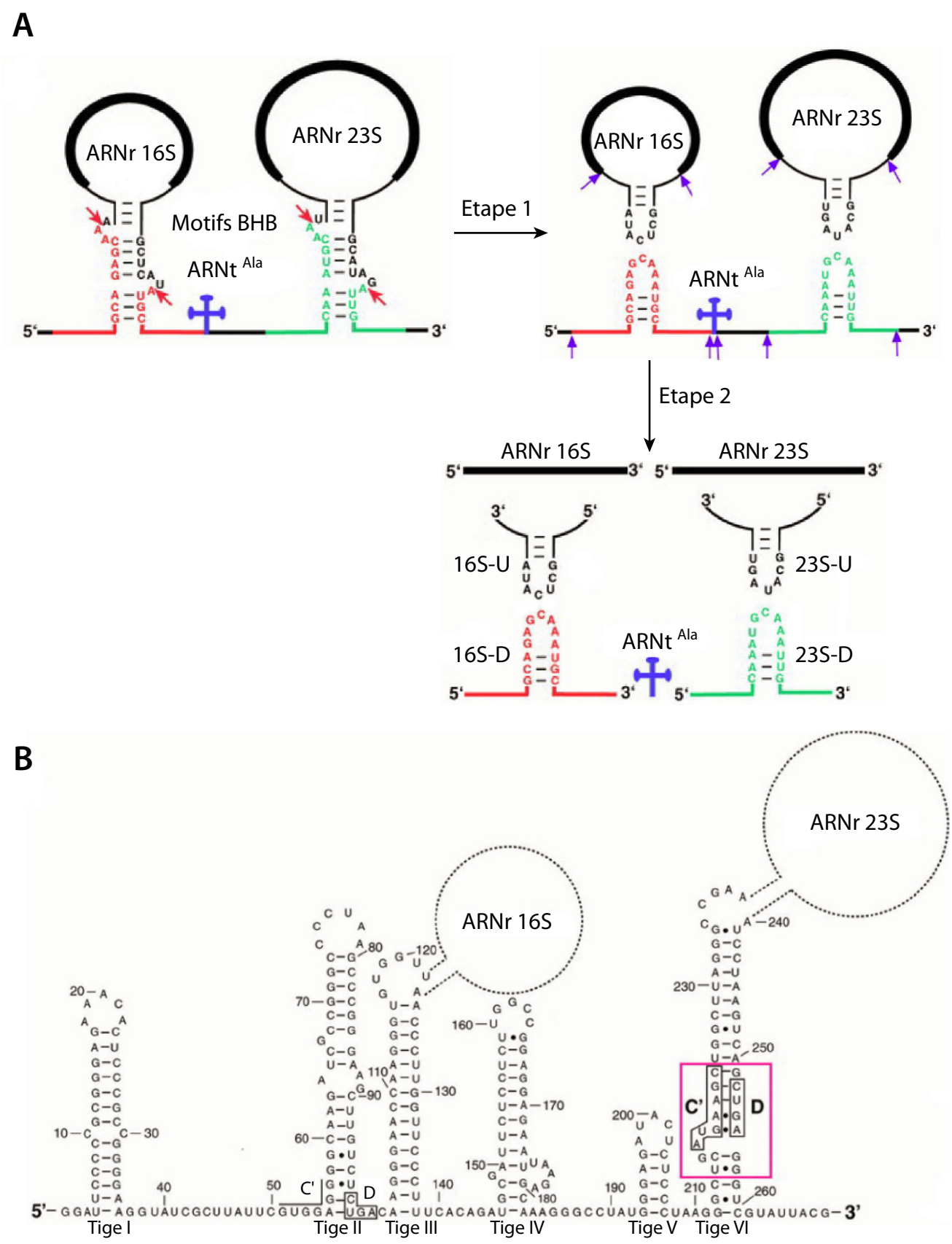


Figure i27 : Maturation du précurseur pré-ARNr d'archaea. (D'après Tang *et al.,* 2002b).
A. Modèle de maturation du pré-ARNr proposé chez *Archaeoglobus fulgidus*. Les ARNr matures sont représentés par des lignes épaisses et noires et les précurseurs par des traits fins. Les sites de clivages dans les motifs BHB (boucle-hélice-boucle) sont indiqués par des flèches rouges et les autres étapes de clivages par des flèches violettes. Les ARN espaceurs provenant de la maturation du pré-ARNr sont respectivement indiqués par 16S/23S-D (Down) ou 16S/23S-U (Up).
B. Modèle de structure secondaire du pré-ARNr 16/23S-D de *Sulfolobus solfataricus*. Le motif en K-turn fonctionnel fixant la protéine L7Ae *in vitro* est encadré en rouge. La localisation des ARNr 16S et 23S précédant leur clivage lors de l'étape 1 est représentée en traits pointillés.

d'un domaine catalytique et d'un domaine structural (Li et Abelson, 2000). Elles sont actives sous la forme d'homodimère et clivent dans chacune des deux boucles du motif BHB (Fig. i27 A) (Li et Abelson, 2000). La première étape de la maturation consiste à séparer le pré-ARNr 16S et le pré-ARNr 23S et elle se termine par une circularisation des pré-ARNr 16S et 23S par ligature intramoléculaire (Fig. i27 A) (Tang *et al.*, 2002b). La seconde phase se déroule grâce à d'autres endonucléases qui vont libérer les ARNr 16S et 23S matures. Les ARNr 16S-D ou 23S-D peuvent contenir un motif en K-turn de type C/D, que je décrirai dans la partie IV.B.2.2, sur lequel se fixe la protéine L7Ae qui pourrait, en formant ce complexe, avoir un rôle actif dans la maturation des ARNr d'archaea (Fig. i27 B) (Tang *et al.*, 2002b). Chez certaines espèces d'archaea, il est possible de retrouver un gène d'ARNt sur le transcrit pré-ARNr qui est également maturé par des clivages endonucléolytiques lors de la seconde phase de clivage (Tang *et al.*, 2002b).

B. Les sno/sRNP à boîtes C/D

1. Propriétés et rôles des 2'-*O*-méthylations

La méthylation en position 2' du ribose renforce la résistance de l'ARN vis-à-vis de l'hydrolyse par les RNases. En effet, la présence du groupement 2'-*O*-méthyl bloque l'activité des RNases qui clivent au niveau du phosphate 3', telle que la RNase T2. La capacité d'établir une liaison avec un hydrogène du groupement 2'-*O*-hydroxyle qui est hydrophile est diminuée par le groupement méthyle qui lui est hydrophobe (Fig. i28 A) (pour revue, (Bachellerie *et al.*, 2002)). Il a été proposé que chez les Archaea, le nombre de 2'-*O*-méthylations présentes dans l'ARN augmente avec la température optimale de croissance de l'espèce (Omer *et al.*, 2000).

Il est probable que les 2'-*O*-méthylations des riboses influencent la conformation locale de l'ARNr. Leur présence favorise en effet la conformation 3' endo du ribose, ce qui facilite l'empilement des nucléotides et donc les appariements de bases et aussi les interactions avec les protéines associées. Comme dans le cas des pseudouridines, les résidus 2'-*O*-méthylés des ARNr sont concentrés dans les régions

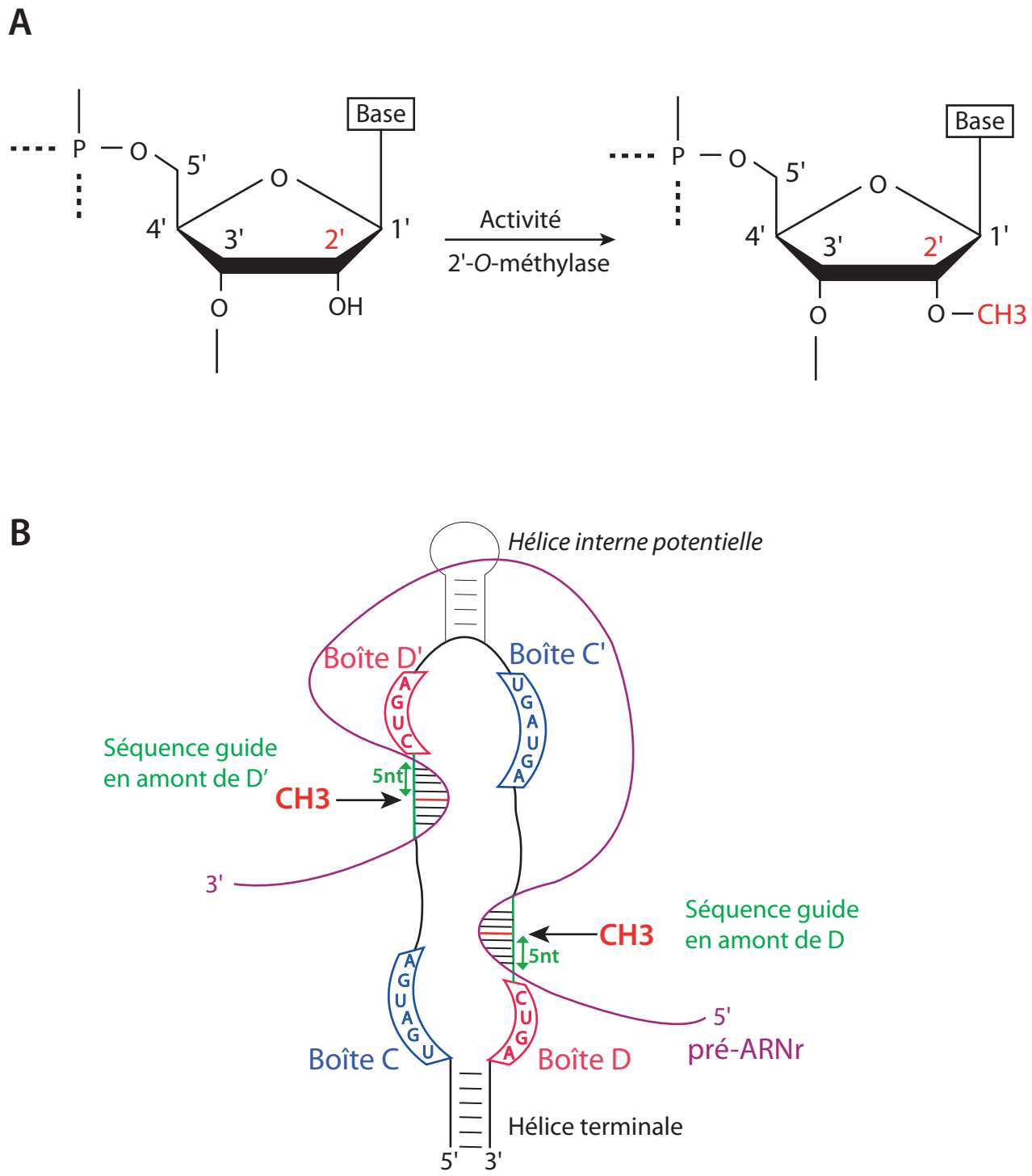


Figure i28 : Les snoRNA à boîtes C/D, guides de méthylation en 2'-*O* des riboses.
A. Structure d'un ribonucléotide modifié par une activité 2'-*O*-méthylase, additionnant un groupement méthyle en 2' du ribose.
B. Représentation schématique d'un snoRNA à boîtes C/D et du principe de la 2'-*O*-méthylation guidée par cet ARN. Un duplexe formé entre le snoRNA et son ARN cible pré-ARNr est présenté. Deux couples de boîtes C/D et C'/D' sont présentés respectivement aux extrémités 5' et 3' et au centre de la séquence du snoRNA. Sont également représentés les séquences guides en amont de la boîte D et en amont de la boîte D', complémentaires à deux séquences du pré-ARNr. Le nucléotide dont le ribose sera méthylé est toujours celui qui s'apparie avec le 5^{e} nucléotide en amont des boîtes D ou D'. La méthylation du ribose apparaît en rouge.

fonctionnellement importantes et hautement conservées (Fig. i5 et i6) (Decatur et Fournier, 2002). Alors que l'absence d'un site ou de quelques sites de méthylation n'a pas de conséquence phénotypique, l'absence d'une grande partie des positions modifiées provoque un défaut d'assemblage des sous-unités ribosomiques (pour revues, (Bachellerie *et al.*, 2002; Omer *et al.*, 2003)). Les 2'-*O*-méthylations contribuent donc à la modulation de la biogenèse et de l'activité des ribosomes. Il est également envisageable que la fonction des snoRNA soit associée à l'importance de l'appariement de l'ARN guide plutôt qu'au résidu modifié lui-même et donc de jouer le rôle de molécule chaperonne. Ainsi, c'est l'appariement entre l'ARNr et le snoRNA qui permettrait de favoriser ou d'empêcher un repliement donné de l'ARNr.

2. Chez les Eucaryotes

2.1. Les ARN guides à boîtes C/D

Les snoRNA à boîtes C/D, responsables de la 2'-*O*-méthylation, représentent la première classe de snoRNA à avoir été découverte (Tollervey, 1987; Tyc et Steitz, 1989). Leur taille est comprise entre 75 et 100 nucléotides (Kiss-Laszlo *et al.*, 1998; Lowe et Eddy, 1999). Ils sont caractérisés par la présence de deux séquences conservées : les boîtes C (5'-UGAUGA-3') et D (5'-CUGA-3'), qui sont situées respectivement aux extrémités 5' et 3' de ces ARN (Fig. i28 B) (Tyc et Steitz, 1989). Ils comportent souvent un second jeu de boîtes C' et D', dont les séquences sont moins conservées et sont alors situées dans une portion interne des snoRNA (Kiss-Laszlo *et al.*, 1998). La présence de l'ensemble de ces boîtes est nécessaire à la stabilité des snoRNA dans la cellule (Huang *et al.*, 1992). Le couple de boîtes C et D participe au signal de maturation de ces snoRNA (Watkins *et al.*, 1996). En plus, ils comportent une à deux séquences, dites anti-sens, d'une taille de 10 à 21 nucléotides. Ces séquences sont complémentaires à l'ARN cible et c'est au niveau de l'hélice formée entre le snoRNA et l'ARN substrat qu'a lieu la méthylation, très précisément à la cinquième paire de bases, en amont de la boîte D ou de la boîte D' (Fig. i28 B) (Cavaille *et al.*, 1996; Kiss-Laszlo *et al.*, 1996). Ainsi, l'existence des deux boîtes D et D'

et des deux séquences anti-sens permet à un même snoRNA de guider la réaction de méthylation à une ou deux positions sur un même ARN substrat ou sur deux ARN distincts. Une mutation qui empêche la formation de la paire de bases entre le snoRNA et sa cible au niveau du nucléotide à modifier entraîne un défaut de méthylation (Omer *et al.*, 2002). La nature de la séquence qui précède les boîtes D et D' est donc le seul élément déterminant le ou les sites de méthylation.

Certains snoRNA à boîtes C/D ciblent d'autres ARN que le pré-ARNr tel que l'ARNm. Ainsi, le snoRNA à boîtes C/D HBII-52, exprimé dans le cerveau humain et conservé chez la souris (MBII-52), présente une complémentarité de séquence avec l'ARNm du récepteur 5-HT2C de la sérotonine au niveau d'une région subissant un phénomène d'édition (adénine en inosine) et un épissage alternatif (Cavaille *et al.*, 2000). Ceci suggère une régulation possible de ces étapes post-transcriptionnelles par les snoRNA dans le syndrome de Prader-Willi qui est une maladie génétique résultant d'une déficience de l'expression de gènes paternels situés dans la région 15q11-q13 du chromosome 15 humain (Cavaille *et al.*, 2000). De nombreux snoRNA à boîtes C/D ne possèdent aucune cible potentielle détectée dans les ARNr, ceci suggère que : i) plusieurs snoRNA à boîtes C/D pourraient être impliqués dans la modification d'ARNm, et ii) la fonction de ces snoRNA soit autre que de guider des évènements de méthylation et pourraient servir en tant que molécules chaperonnes (pour revue, (Bachellerie *et al.*, 2002)).

2.2. Le motif en Kink-turn (K-turn)

Dans la majorité des snoRNA à boîtes C/D, une hélice est formée par appariement des extrémités 5' et 3' du snoRNA permettant la juxtaposition des boîtes C et D. Ce couple a la capacité de former un motif structural particulier : le motif K-turn (Watkins *et al.*, 2000; Marmier-Gourrier *et al.*, 2003; Moore *et al.*, 2004). La première structure 3D de ce motif a été établie par analyse radiocristallographique du complexe formé entre la protéine 15.5 kD et le snRNA U4 de la machinerie d'épissage (Fig. i29 A et B) (Vidovic *et al.*, 2000). Le motif en K-turn conventionnel est

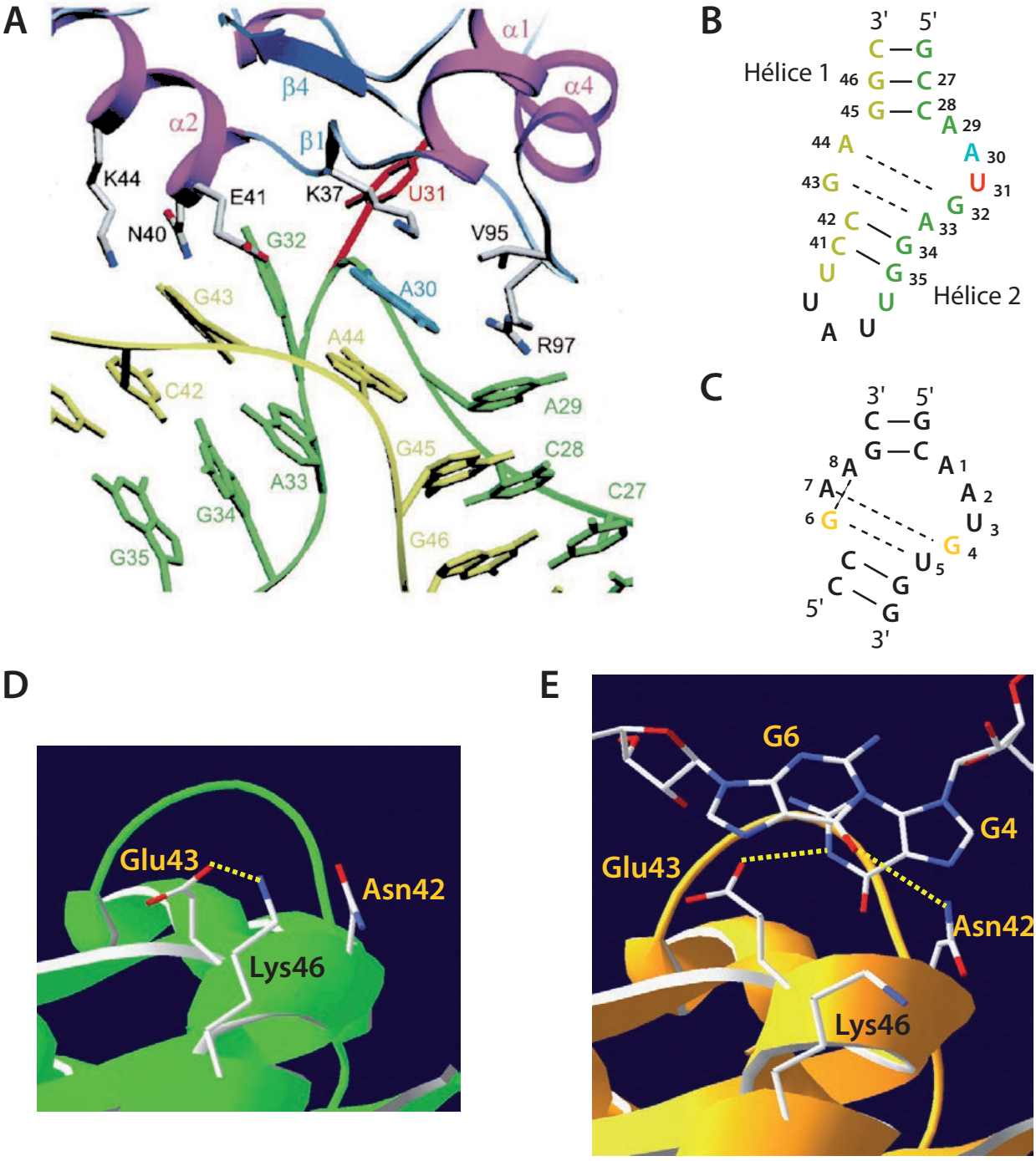


Figure i29 : Représentation du motif en K-turn et du complexe résultant de son association avec la protéine ribosomique L7Ae. (Vidovic *et al.,* 2000 ; Ban *et al.,* 2000 ; Charron *et al.,* 2004).
A. Structure 3D de la boucle interne de la structure tige-boucle 5'-terminale du snRNA U4 en interaction avec la protéine 15,5 kD. Les brins des deux hélices ARN (1 et 2) sont respectivement représentés en jaune et vert. Le résidu U31 en rouge est le nucléotide projeté hors de la structure dans la poche formée par la protéine.
B. Représentation 2D du motif en K-turn de la boucle terminale du snRNA U4 humain selon le même code couleur qu'en A. Les paires de bases G:A et A:G (sheared pairs) sont schématisées par des pointillés.
C. Représentation 2D du motif en K-turn KT-15 de l'ARNr 23S de *H. marismortui* (Klein *et al.,* 2001). Les nucléotides G4 et G6 qui sont visibles dans le panneau E sont représentés en jaune foncé.
D. Structure 3D établie au laboratoire, de la zone d'interaction au K-turn de L7Ae sous forme libre (en vert). Dans cette situation, le résidu de L7Ae Glu43 réalise un pont salin avec le résidu Lys46.
E. Structure 3D de la même zone d'interaction au K-turn KT-15 de L7Ae sous forme liée (en jaune). Dans cette situation, le pont salin formé entre les résidus Glu43 et Lys46 est rompu au profit des interactions entre les résidus Asn42, Glu43 et la base G4 de l'ARN.

composé d'une quinzaine de nucléotides structurés sous la forme hélice 1-boucle interne-hélice 2 (Klein *et al.*, 2001). La boucle interne asymétrique (5+2 nucléotides) est hautement structurée et une courbure de 120° est induite dans le squelette phosphate-ribose entre les hélices 1 et 2 (Klein *et al.*, 2001). L'hélice 1 se termine au niveau de la boucle interne asymétrique avec 2 paires de bases G-C de type Watson-Crick, tandis que l'hélice 2 débute avec 2 paires G•A non Watson-Crick, appelées « paires partagées » (sheared pairs) (Klein *et al.*, 2001). Un résidu de la boucle interne (souvent une uridine) est projeté en dehors de l'hélice ARN pour venir se loger dans une poche de la protéine (Vidovic *et al.*, 2000). Un motif en K-turn non conventionnel est appelé K-loop ou pseudo-K-turn car celui-ci se forme lorsqu'il est situé au niveau d'une boucle apicale et dans ce cas il ne possède qu'une seule hélice flanquée des deux paires G•A (Charron *et al.*, 2004; Hamma et Ferre-D'Amare, 2004). Lors de la fixation d'une protéine (Snu13p/15.5 kD/L7Ae) qui reconnaît un motif ARN en K-turn ou K-loop, ce dernier subit des changements de conformations contrairement à la protéine qui ne subit que très peu de modifications (Fig. i29 C à E). C'est donc l'ARN qui adapterait sa conformation à celle de la protéine qui dirigerait un phénomène d'« induced-fit » (Charron *et al.*, 2004).

Le motif en K-turn est un motif de structuration de nombreux ARN. En effet, ce type de motif est retrouvé, en plus des sno/sRNA à boîtes C/D (Watkins *et al.*, 2000; Kuhn *et al.*, 2002), dans les ARNr (Ban *et al.*, 2000), le snRNA U4 (Vidovic *et al.*, 2000), les sRNA à boîtes H/ACA (Rozhdestvensky *et al.*, 2003) et l'élément SECIS des ARNm de sélénocystéine (Walczak *et al.*, 1996). La séquence de chacun de ces motifs détermine l'identité de la protéine s'y fixant spécifiquement. Le complexe formé sert alors de site de nucléation autour duquel d'autres protéines viennent spécifiquement s'assembler pour former une RNP spécifique aux ARNr (Klein *et al.*, 2001), au snRNA U4 (Nottrott *et al.*, 1999), aux snoRNA C/D (Watkins *et al.*, 2000), ou à l'élément SECIS des ARNm des sélénoprotéines (Fletcher *et al.*, 2001).

Protéines coeur des snoRNP à boîtes C/D

Eucaryotes		Archaea
Levure	**Homme**	
Nop1p	Fibrillarine	aNOP1 (aFIB)
Snu13p	15.5 kD	L7Ae
Nop56p	NOP56	aNOP56/58 (aNOP5)
Nop58p	NOP58	

Tableau i5 : Comparatif des différentes protéines composant les snoRNP à boîtes C/D chez les Eucaryotes (levure et homme) et chez les Archaea.

2.3. Les protéines associées aux snoRNA à boîtes C/D

Pour faciliter la compréhension, le tableau i5 répertorie la correspondance entre les noms donnés aux protéines associées aux snoRNA à boîtes C/D chez l'homme, la levure et les archaea.

La protéine Nop1 / Fibrillarine

La première protéine associée aux snoRNA à boîtes C/D à avoir été identifiée est la fibrillarine chez l'homme ou Nop1p chez la levure (Tab. i5) (Ochs *et al.*, 1985; Schimmang *et al.*, 1989; Henriquez *et al.*, 1990). La fibrillarine est une protéine composée de 321 résidus (36 kDa) qui possède 67% d'identité avec Nop1p (Aris et Blobel, 1991). C'est une protéine basique de type non-histone qui s'accumule majoritairement dans les Corps de Cajal et dans les régions fibrillaires du nucléole, d'où lui vient son nom (Ochs *et al.*, 1985). Elle est indispensable à la viabilité cellulaire (Schimmang *et al.*, 1989) et est requise pour la stabilité de certains des snoRNA à boîtes C/D (Lafontaine et Tollervey, 2000). Elle contient un domaine GAR, riche en enchaînements glycine/arginine. Nop1p possède un motif caractéristique des méthyltransférases dépendantes du cofacteur S-adénosyl méthionine (SAM) (Niewmierzycka et Clarke, 1999) et des mutations dans ce domaine inhibent la formation de 2'-*O*-Me dans le pré-ARNr aux positions ciblées par les snoRNP à boîtes C/D. Ceci a permis de montrer que c'est cette protéine qui porte l'activité catalytique de ces snoRNP (Tollervey *et al.*, 1993). La délétion du gène codant Nop1p est létale, ce qui conduit en particulier, à l'abolition de la production de l'ARNr 18S et à l'accumulation d'un intermédiaire aberrant de maturation (23S), obtenu en absence des clivages précoces aux sites A0, A1 et A2 (Fig. i26) (Tollervey *et al.*, 1991).

Les protéines Nop56 et Nop58

Les protéines Nop56 (56 kDa) et Nop58 (58 kDa) ont initialement été identifiées chez la levure par un crible de létalité synthétique à partir d'une souche exprimant un variant de la protéine Nop1 (Tab. i5) (Gautier *et al.*, 1997). Elles sont essentielles à la viabilité cellulaire et ont une localisation majoritairement nucléolaire

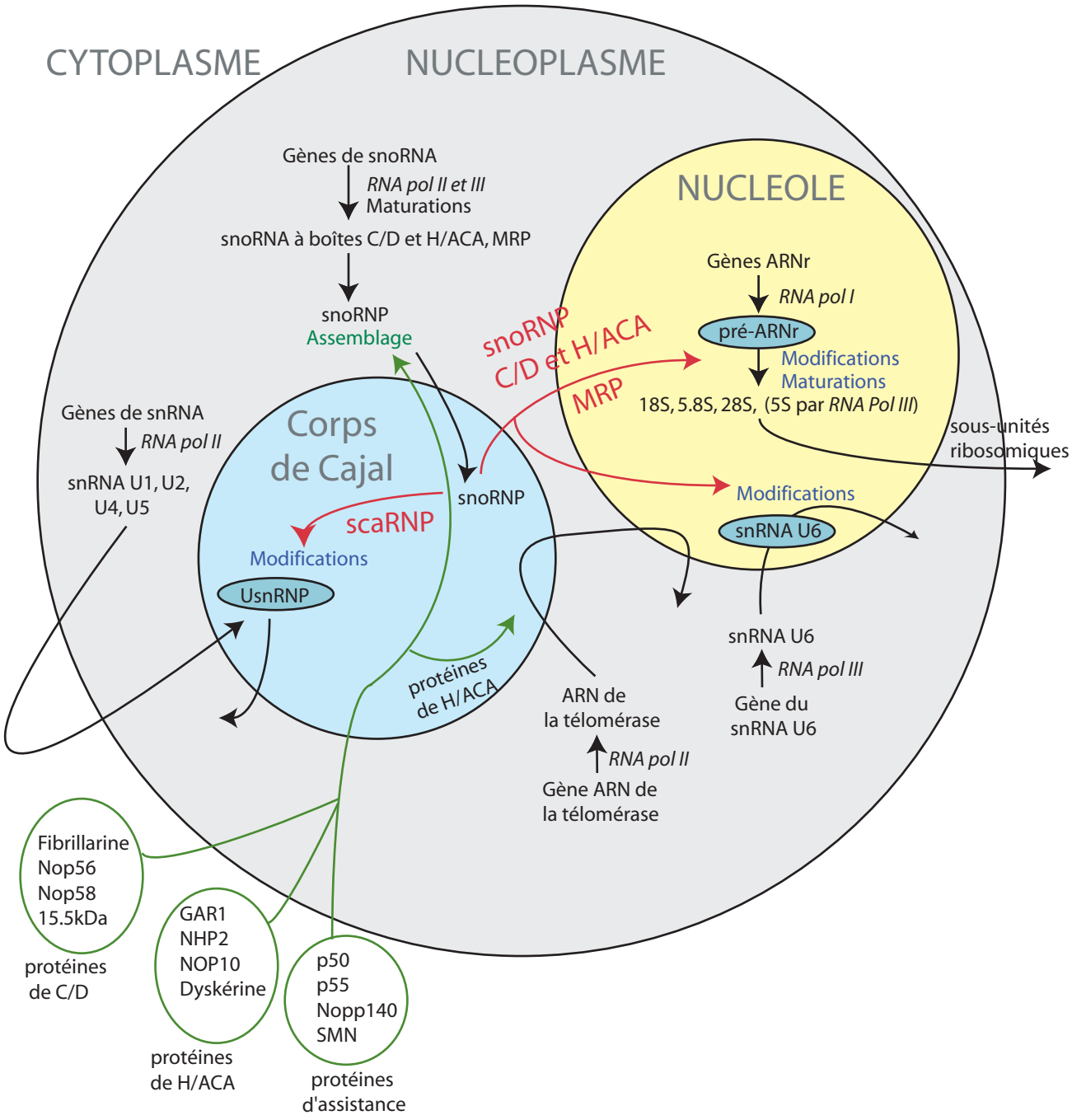


Figure i30 : Biogenèse, trafic et fonctions des snoRNA, snRNA, ARNr et de la télomérase chez les vertébrés. (Adapté de Filipowicz et Pogacic, 2002).
Les snoRNP fonctionnelles sont indiquées en rouge, leurs cibles et leur actions sont précisées en bleu. Des protéines jouant un rôle dans la biogenèse et le trafic des snoRNP, dont la liste n'est pas exhaustive, sont indiquées en vert.

(Gautier *et al.*, 1997). Leur région Ct est riche en séquences répétées KKE/D qui ne sont pas requises pour leur localisation nucléolaire (Gautier *et al.*, 1997). Les protéines Nop56 et Nop58 présentent entre elles une forte homologie de séquence (45% d'identité) indiquant qu'elles résultent vraisemblablement d'une duplication d'un gène ancestral commun. Malgré cette forte homologie existant entre Nop56p et Nop58p, une cellule de levure n'exprimant plus Nop56p ne peut pas être complémentée par Nop58p et vice-versa (Gautier *et al.*, 1997). Il a ainsi été suggéré que la formation d'un complexe entre Nop1p et Nop56p était nécessaire pour que les étapes d'assemblage du ribosome puissent avoir lieu (Gautier *et al.*, 1997). L'absence de production des protéines Nop56 ou Nop58 conduit, comme pour Nop1p, à un défaut de maturation des ARNr aux sites de clivage A0, A1 et A2 (Fig. i26) (Lafontaine et Tollervey, 1999, 2000). Cependant, seule Nop58p est requise pour la stabilité des snoRNA à boîtes C/D (Lafontaine et Tollervey, 1999).

La protéine Snu13 / 15.5 kD

Les protéines Snu13 (13 kDa) de levure et son homologue 15.5 kD (15,5 kDa) humaine ont été initialement été découvertes comme étant associées à la snRNP spliceosomale U4/U6•U5 (Tab. i5) (Gottschalk *et al.*, 1999; Nottrott *et al.*, 1999; Stevens et Abelson, 1999) puis sa présence dans les snoRNP C/D a été démontrée (Watkins *et al.*, 2000). Les couples de boîtes C/D des snoRNA forment un motif en K-turn qui fixe la protéine Snu13/15.5 kD (Fig. i29) (Watkins *et al.*, 2000). En fait, c'est la fixation de la protéine Snu13/15.5 kD qui favorise la structuration de l'ARN en K-turn (Marmier-Gourrier *et al.*, 2003). La déplétion de Snu13p conduit spécifiquement à une réduction du taux de snoRNA à boîtes C/D et celle-ci est suivie rapidement par une diminution drastique du taux d'ARN cellulaires (Watkins *et al.*, 2000).

2.4. Biogenèse des snoRNP à boîtes C/D

D'une manière générale, les snoRNA transcrits dans le noyau s'associent avec des protéines qui doivent être importées du cytoplasme vers le noyau (Fig. i30). Les

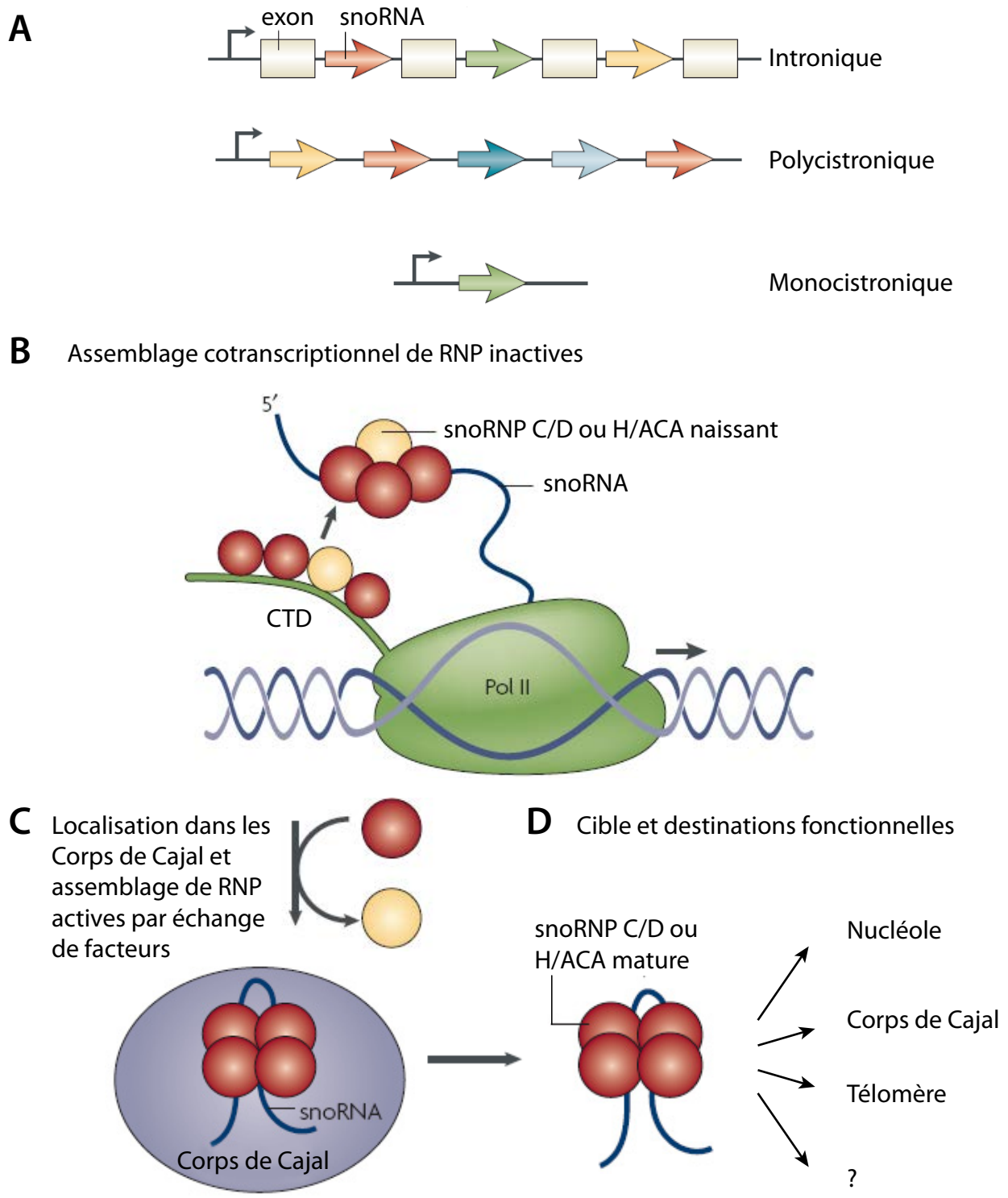


Figure i31 : Couplage entre la synthèse, l'assemblage et le trafic des snoRNP à boîtes C/D et H/ACA. (D'après Matera *et al.*, 2007).

A. Organisation des structures génomiques codant les snoRNA à boîtes C/D et H/ACA. Les séquences codant ces ARN peuvent être introniques, polycistroniques ou monocistroniques.

B. Assemblage cotranscriptionnel d'une pré-RNP inactive (biogenèse précoce). Une partie des protéines cœur (rouge) et de facteurs d'échange (jaune) est associée avec le domaine C-terminal (CTD) de l'ARN polymérase II (Pol II) au site de transcription de l'ARN guide et s'assemblent sur l'ARN naissant, formant une pré-RNP métaboliquement stable mais inactive.

C. Maturation en RNP fonctionnelle dans les Corps de Cajal (biogenèse tardive). La pré-RNP se déplace dans les Corps de Cajal. Cette migration est conditionnée par sa composition en protéine. Les étapes terminales dans la formation des complexes fonctionnels ont alors lieu, incluant les modifications finales de l'ARN et le remplacement du facteur d'échange par la ou les protéines cœurs manquantes.

D. Sites cibles dans lesquels les snoRNP vont s'accumuler pour assurer leur fonction.

protéines vont se fixer avec les snoRNA naissants aux sites de transcription pour former une snoRNP (pour revue, (Filipowicz et Pogacic, 2002)). Ce schéma général de la biogenèse est valable pour les snoRNP à boîtes C/D et celles à boîtes H/ACA qui font l'objet du chapitre IV.C.1.1 de l'introduction.

Chez les eucaryotes, les pré-snoRNA à boîtes C/D sont transcrits par l'ARN polymérase II et leurs gènes sont localisés soit dans des introns d'ARNm, ce qui est pratiquement la généralité chez les vertébrés, soit possèdent comme de nombreux cas chez la levure, leur propre promoteur (Fig. i31 A) (pour revue, (Matera *et al.*, 2007)). L'assemblage des pré-snoRNP à boîtes C/D se déroule de façon co-transcriptionnelle tout en étant couplé à l'épissage du pré-ARNm (Fig. i31 B) (Verheggen *et al.*, 2002; Hirose *et al.*, 2003; Morlando *et al.*, 2004). La biogenèse des snoRNP à boîtes C/D implique un grand nombre de facteurs d'assemblages, ces protéines interviennent de façon transitoire durant ce processus et ne sont pas retrouvées dans les particules matures.

Il existe une machinerie moléculaire d'assemblage pour les RNP portant les membres de la famille de la protéine L7Ae. Cette machinerie existe chez l'homme et la levure et est composé de la protéine adaptatrice appelée NUFIP (Nuclear FMRP Interacting Protein) chez l'homme et Rsa1 (Ribosome Assembly 1) chez la levure. Elle est proposée jouer un rôle central en tant que plate-forme d'assemblage des pré-snoRNP à boîtes C/D en facilitant les contacts entre les différentes protéines cœurs des particules et en recrutant le complexe formé entre la chaperonne HSP90 et le complexe R2TP (Zhao *et al.*, 2005; Boulon *et al.*, 2008; Zhao *et al.*, 2008). Le complexe R2TP contient 4 protéines qui interagissent entre elles, il s'agit chez l'homme des protéines hSPAGH, hPIH1, et les AAA^+ ATPases Pontine (TIP48) et Reptine (TIP49), et respectivement chez la levure : Tah1p, Pih1p, Rvb1p et Rvb2p (Zhao *et al.*, 2005; McKeegan *et al.*, 2007; Boulon *et al.*, 2008; Zhao *et al.*, 2008). La protéine HSP90 (Heat Schock Proteins 90) forme un complexe stable avec hSPAGH/Tah1p (Zhao *et al.*, 2008) et les protéines NUFIP/Rsa1p établissent des interactions avec hPIH/Pih1 (Boulon *et al.*, 2008). Il est proposé que le complexe R2TP en association avec protéines NUFIP/Rsa1p et HSP90 intervienne lors de la biogenèse des snoRNP en

favorisant le déroulement des étapes de remodelage des pré-snoRNP pour aboutir aux particules matures (Boulon *et al.*, 2008; Zhao *et al.*, 2008).

D'autres facteurs seraient impliqués dans les mécanismes d'assemblage, l'ARN hélicase IBP160 (Intron Binding Protein 160) est décrite pour se fixer en amont des introns contenant les snoRNA à boîtes C/D. Cette activité serait requise pour débuter l'assemblage de la pré-snoRNP à boîtes C/D (Hirose *et al.*, 2006). Un autre facteur, la protéine BCD1 (Box C/D RNA 1) est nécessaire à l'accumulation spécifique des snoRNA à boîtes C/D (Peng *et al.*, 2003; McKeegan *et al.*, 2007). Elle interagit notamment avec la machinerie transcriptionnelle mais n'a jamais été retrouvée associée de façon stable avec des particules à boîtes C/D matures ou en cours d'assemblage (Fig. i31 C) (Peng *et al.*, 2003; McKeegan *et al.*, 2007). C'est dans les Corps de Cajal qu'ont lieu les dernières étapes de la maturation des snoRNP telles que l'hyperméthylation de la coiffe, la maturation de l'extrémité 3' des snoRNA et l'échange de certains facteurs protéiques (Verheggen *et al.*, 2001; Verheggen *et al.*, 2002). Puis les snoRNP matures sont dirigées vers leurs différents lieux d'action à l'exception des scaRNP, retenues dans les Corps de Cajal, et dont les caractéristiques seront détaillées dans la partie IV.C.1.2. (Fig. i31 D).

3. Chez les Archaea

Les sRNA (small RNA) à boîtes C/D d'archaea ont été découverts plus tardivement que leurs homologues eucaryotes (Gaspin *et al.*, 2000; Omer *et al.*, 2000; Tang *et al.*, 2002a). La taille de ces sRNA est en moyenne de 50 à 60 nucléotides de long, ce qui est significativement plus court que les snoRNA eucaryotes, mais sRNA et snoRNA C/D adoptent avec une structure similaire (Fig. i28 B). Il a été montré qu'un sRNA à boîtes C/D d'archae microinjecté dans des oocytes de xénope se localisait dans les Corps de Cajal et était fonctionnel (Speckmann *et al.*, 2002). Tous les sRNA de la famille à boîte C/D comportent les deux couples de boîtes C/D et C'/D' et leur mode d'action pour cibler la position pour la 2'-*O*-méthylation repose sur la même règle que pour les snoRNA C/D : le 5^{e} nucléotide en amont des boîtes D

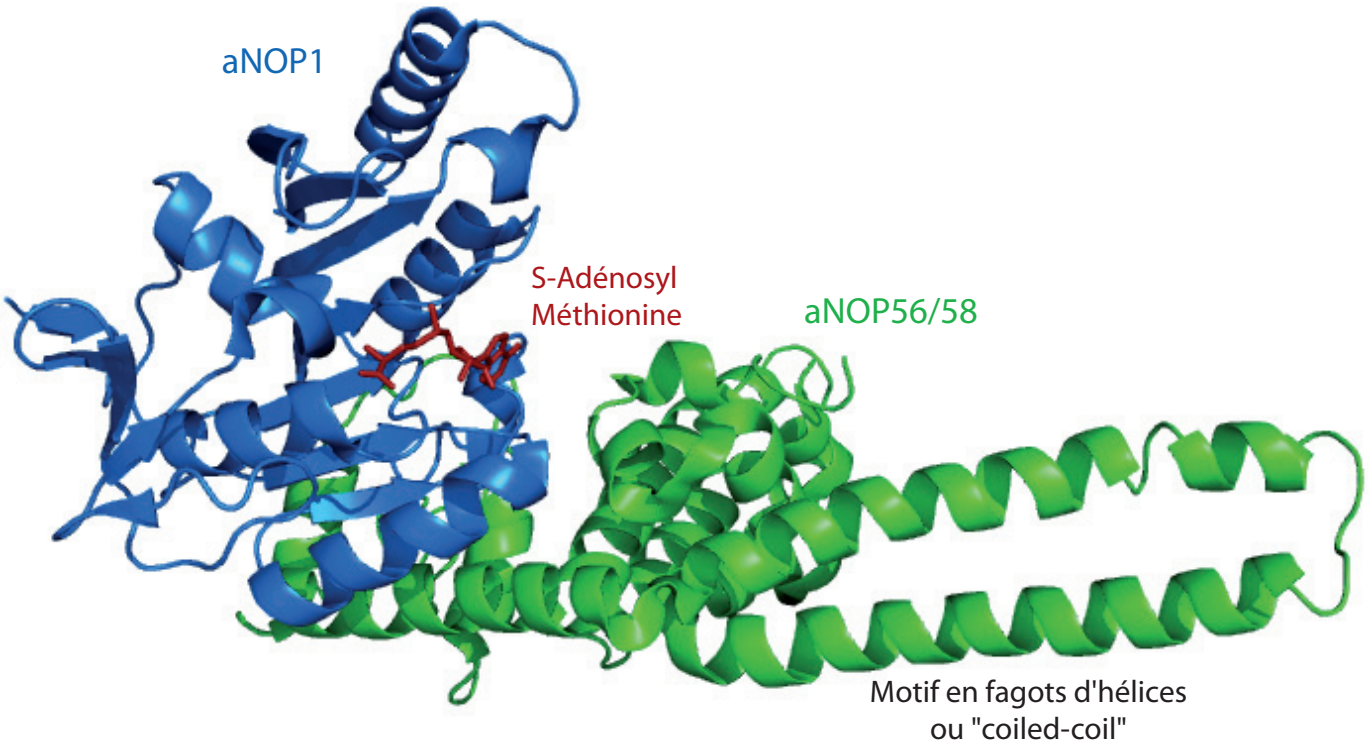


Figure i32 : Structure 3D du complexe protéique aNOP1-aNOP56/58 d'*Archaeoglobus fulgidus*.
(D'après Aittaleb *et al.,* 2003).
aNOP1 (aFIB) est représentée en bleu, aNOP56/58 (aNOP5) en vert et le cofacteur S-adénosyl méthionine en rouge. Réalisé à l'aide du logiciel Pymol à partir de la structure obtenues par Aittaleb *et al.,* 2003 (PBB=1NT2).

ou D' présent dans l'appariement entre le snoRNA et l'ARN cible sera modifié (Omer *et al.*, 2000). Chaque sRNA C/D a potentiellement la capacité de cibler 2 positions à modifier que ce soit sur le même substrat ou sur 2 substrats différents. La distance de 12 à 13 nts entre les motifs C/D et C'/D' est conservée chez les sRNA d'archaea et l'altération d'un seul de ces nucléotides « espaceurs » affecte grandement l'activité de méthylation (Tran *et al.*, 2005). Une étude a montré que certains sRNA à boîtes C/D de *P. furiosus* seraient circulaires, ce qui permettrait ainsi de protéger les deux extrémités, mais sans pour autant qu'une séquence *cis* reconnue spécifiquement par une exonucléase et une ligase ne soit décrite (Starostina *et al.*, 2004).

Les protéines associées aux snoRNA à boîtes C/D eucaryotes possèdent des homologues chez les archaea, il s'agit des protéines aNOP1, aNOP56/58 (aNOP5) et L7Ae (Tab. i5). Le cocristal du complexe aNOP1-aNOP56/58 a été résolu en présence de son cofacteur S-adénosyl méthionine (Fig. i32) (Aittaleb *et al.*, 2003).

L'enzyme aNOP1 comporte 210 résidus et ne possède pas le domaine GAR présent chez son homologue eucaryote. Elle affiche un domaine de fixation à la S-adénosyl méthionine dont le repliement est caractéristique des méthyles transférases et qui consiste en 7 brins β pris en sandwich entre 6 hélices α (Wang *et al.*, 2000; Aittaleb *et al.*, 2003).

La protéine aNOP56/58 contient 261 acides aminés et ne possède pas de répétitions du type KKE/D, ce qui est en accord avec le fait que ses homologues eucaryotes sont fonctionnels sans ce motif (Gautier *et al.*, 1997). A ce sujet, il est tout de même intéressant de souligner que chez certaines espèces telle que *P. abyssi*, aNOP56/58 possède des répétitions KKE/D en Ct. La protéine aNOP56/58 composée de 2 brins β et de 12 hélices α s'organise autour de 2 domaines : le domaine Nt (α1 à α3 et β1-β2) qui interagit avec aNOP1 et le domaine Ct (α6 à α12) qui fixe l'ARN (Aittaleb *et al.*, 2003). Les 2 domaines sont reliés entre eux par les hélices α4 et α5 qui constituent un motif en fagot d'hélices (coiled-coil) (Fig. i33). En effet, aNOP56/58 possède un domaine de fixation à l'ARN chargé positivement, le

domaine conservé NOP, qui se replie de façon autonome et qui conserve sa capacité à se fixer à l'ARN ainsi qu' à L7Ae liée à l'ARN (Hardin *et al.*, 2009).

Le complexe aNOP1-aNOP56/58 forme un hétérodimère en solution et dimérise via le motif en fagot d'hélices des deux protéines aNOP56/58. En effet, la délétion ou certaines mutations ponctuelles dans ce motif abolissent la dimérisation de aNOP56/58 (Aittaleb *et al.*, 2003; Rashid *et al.*, 2003). Il est suggéré que la dimérisation des protéines associées aux sRNA à boîtes C/D ainsi que l'organisation symétrique des protéines soient des pré-requis à l'assemblage et à la fonctionnalité de ces sRNP (Fatica et Tollervey, 2003; Rashid *et al.*, 2003). Une autre étude montre également que si le motif en fagot d'hélice n'est pas essentiel pour l'assemblage des particules sRNP à boîtes C/D, il est capital pour l'activité de méthylation de ces dernières (Zhang *et al.*, 2006). Un alignement de séquence révèle que les résidus du motif en fagot d'hélices sont conservés chez les homologues eucaryotes des protéines d'archaea suggérant que la dimérisation des protéines Nop56 et Nop58 pourrait avoir lieu d'une façon similaire à aNOP56/58 (Aittaleb *et al.*, 2003).

La structure de la protéine L7Ae sera détaillée dans la partie IV.C.2.2.

4. Reconstitution de particules à boîtes C/D d'archaea : information structurales et comparaison avec le système eucaryote

Contrairement à l'assemblage des sRNP à boîtes H/ACA qui sera décrit dans la partie suivante, l'assemblage conduisant aux sRNP à boîtes C/D s'effectue dans un ordre séquentiel. C'est-à-dire que la présence de certaines protéines est un pré-requis pour que les autres puissent se fixer à la RNP (pour revues, (Fatica et Tollervey, 2003; Omer *et al.*, 2003; Reichow *et al.*, 2007), C'est en particulier le cas de l'enzyme aNOP1 qui, contrairement à l'enzyme aCBF5 des sRNP H/ACA, n'est pas une protéine de liaison à l'ARN (cf IV.C.6.1).

- L7Ae est la première protéine à s'associer au motif en K-turn du sRNA (Fig. i33 B piste 2, RNP I) (Omer *et al.*, 2002; Rashid *et al.*, 2003; Tran *et al.*, 2003). Comme

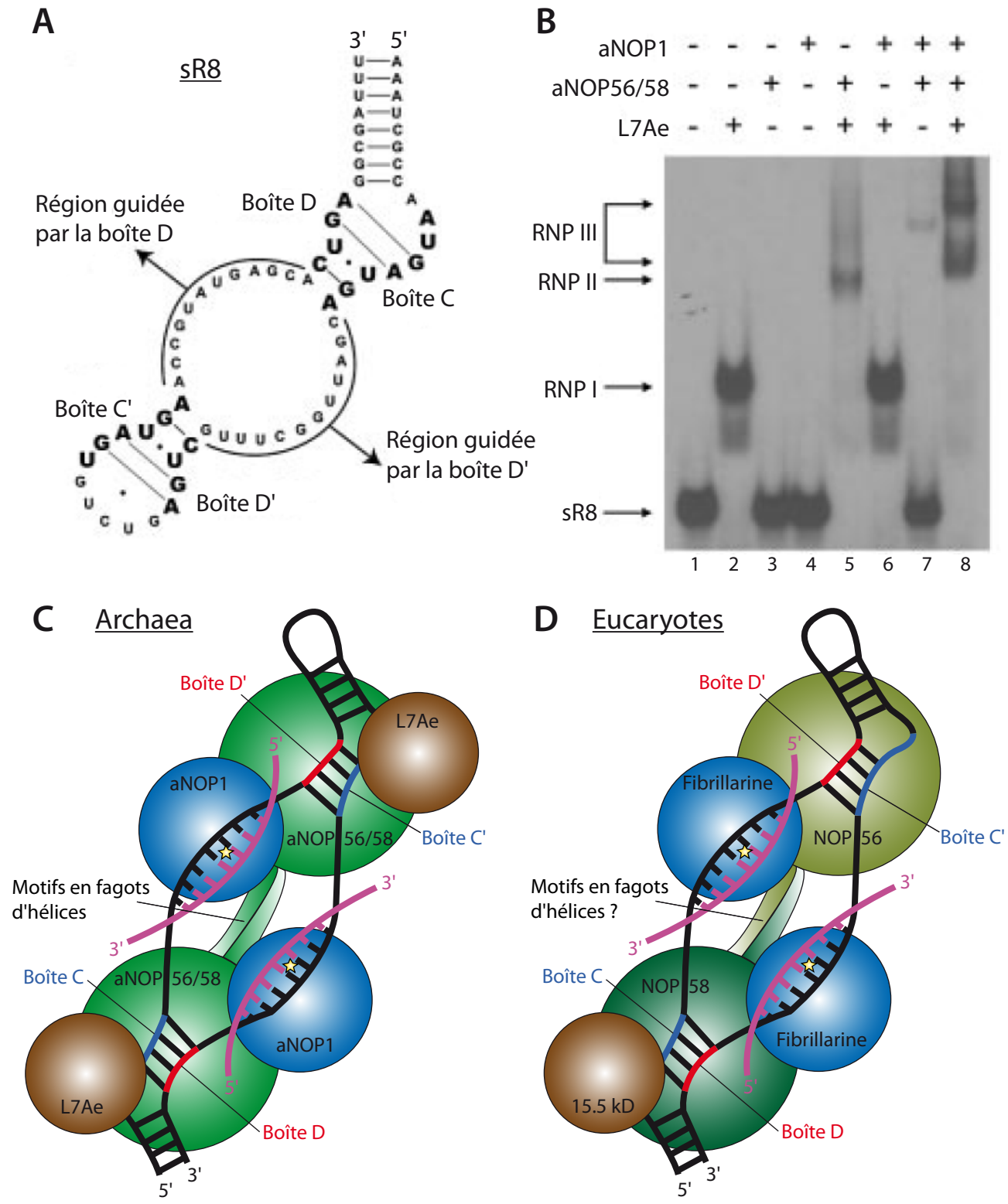


Figure i33 : Assemblage de s/snoRNP à boîtes C/D actives dans la 2'-*O*-méthylation. (Tran *et al.,* 2003 ; Hardin *et al.,* 2006 ; Reichow *et al.,* 2007).

A. Structure schématique du sRNA sR8 à boîtes C/D de *M. jannaschii*.

B. Reconstitution de la sRNP sR8 à boîtes C/D et analyse des complexes formés par retard de migration sur gel. La position de migration des sous-complexes (RNP I et RNP II) et du complexe entier (RNP III) est indiquée ainsi que l'ordre des protéines ajoutées (+ ou -).

C. Organisation du complexe sRNP à boîtes C/D chez les archaea. Les protéines composant ces RNP sont L7Ae (brun), aNOP56/58 (vert) et aNOP1 (bleu), le sRNA est en noir et le site de méthylation est représenté par une étoile jaune, l'ARN substrat est en violet, les boîtes couples de boîtes C et C' en bleu et les boîtes D et D' en rouge.

D. Organisation du complexe snoRNP à boîtes C/D chez les eucaryotes. Les protéines composant ces RNP sont 15.5 kD/Snu13p (brun), NOP56 (vert clair) et NOP58 (vert foncé) et la Fibrillarine/Nop1p (bleu).

nous l'avons déjà mentionné pour la protéine 15.5 kD, la fixation de L7Ae permet de structurer le motif en K-turn et c'est la protéine qui sert de site de nucléation autour duquel vont s'associer aNOP56/58 et aNOP1. Chez les eucaryotes, c'est la protéine Snu13/15.5 kD qui joue un rôle analogue (Watkins *et al.*, 2000; Schultz *et al.*, 2006).

- Aucune interaction directe n'a été détectée entre aNOP1/fibrillarine et l'ARN (Fig. i33 B piste 4) (Fatica *et al.*, 2000). Une interaction a été observée mais elle était de faible affinité et ce résultat n'a jamais été retrouvé pour d'autres snoRNA à boîtes C/D ce qui indique qu'il pourrait s'agir d'une exception (Cahill *et al.*, 2002; Watkins *et al.*, 2002). Le complexe sRNA-L7Ae-aNOP56/58 doit être préformé pour que aNOP1 puisse se fixer à la sRNP (Omer *et al.*, 2002; Rashid *et al.*, 2003; Tran *et al.*, 2003). Cela corrobore l'observation que ni aNOP1 et ni Nop1p ne présentent une activité *in vitro* guidée par l'ARN en absence des autres protéines mais ne fonctionnent qu'une fois intégrées dans une s/snoRNP à boîtes C/D (Wang *et al.*, 2000; Galardi *et al.*, 2002).

- Une fois L7Ae liée au sRNA, aNOP56/58 peut se fixer à ce complexe (Fig. i33 B piste 5, RNP II) (Omer *et al.*, 2002; Rashid *et al.*, 2003; Tran *et al.*, 2003).

- Enfin, aNOP1 peut s'incorporer dans le complexe formé par le sRNA lié aux protéines L7Ae et aNOP56/58 (Fig. i33 B piste 8, RNP III) (Omer *et al.*, 2002; Rashid *et al.*, 2003; Tran *et al.*, 2003). Deux rôles sont proposés pour la protéine aNOP56/58. Sa zone de fixation à aNOP1 est proche du site actif de aNOP1, donc aNOP56/58 pourrait orienter le duplexe sRNA-ARN substrat pour la catalyse (Aittaleb *et al.*, 2003). Elle pourrait également servir de molécule adaptatrice pour réaliser un pont entre le complexe sRNA-L7Ae et l'enzyme aNOP1, cela par le biais de son domaine NOP conservé (Hardin *et al.*, 2009). Néanmoins, une étude récente de localisation *in vivo* dans les cellules eucaryotes indique que les protéines NOP56 et fibrillarine interagissent ensemble avant d'être co-assemblées sur la snoRNP à boîtes C/D (Lechertier *et al.*, 2009). Il est donc possible que l'assemblage *in vivo* des snoRNP C/D soit différent du modèle d'assemblage *in vitro* d'une sRNP C/D d'archaea.

Les données biochimiques obtenues par différents groupes de recherche indiquent que l'architecture des sRNP à boîtes C/D d'archaea serait symétrique avec une copie de chaque protéine assemblée sur les motifs C/D et C'/D' (Fig. i33 C) (Omer *et al.*, 2002; Rashid *et al.*, 2003; Tran *et al.*, 2003). L'architecture de type symétrique joue un rôle crucial dans l'activité de 2'-*O*-méthylation, même s'il est possible d'obtenir des sRNP asymétriques fonctionnelles (Hardin et Batey, 2006; Singh *et al.*, 2008). De plus, pour les espèces hyperthermophiles, une température élevée (70°C) permet de remodeler la structure de la sRNP et de faciliter son assemblage (Gagnon *et al.*, 2006).

Les snoRNP à boîtes C/D eucaryotes sont prédites pour adopter une architecture pseudo-symétrique, similaire à celle des sRNP d'archaea (Fig. i33 D). En effet, une étude employant l'approche de pontages aux U.V. réalisée à partir de snoRNP reconstituées dans des oocytes de xénope indique que la protéine Nop56 interagit uniquement avec la boîte C' tandis que la protéine Nop58 se fixe exclusivement sur la boîte C, suggérant que ces 2 protéines possèdent des sites de fixation spécifiques sur les snoRNA (Cahill *et al.*, 2002). Cette même étude montre également qu'il y aurait une copie de la fibrillarine fixée à chaque couple de boîtes C/D et C'/D' (Cahill *et al.*, 2002). La protéine Snu13/15.5 kD se fixe seulement au motif C/D pour initier l'assemblage de la snoRNP (Watkins *et al.*, 2000; Szewczak *et al.*, 2002). Néanmoins, il est parfaitement envisageable que la protéine Snu13/15.5 kD soit recrutée au motif C'/D' par des interactions protéines-protéines. En effet, des mutations réalisées dans la protéine 15.5 kD indiquent que cette protéine contribue à l'architecture de la snoRNP en réalisant des interactions avec NOP56 et NOP58 (Schultz *et al.*, 2006).

C. Les sno/sRNP à boîtes H/ACA

Après avoir abordé le système de modification guidé par les ARN à boîtes C/D, je vais maintenant aborder le thème central sur lequel a porté la majeure partie

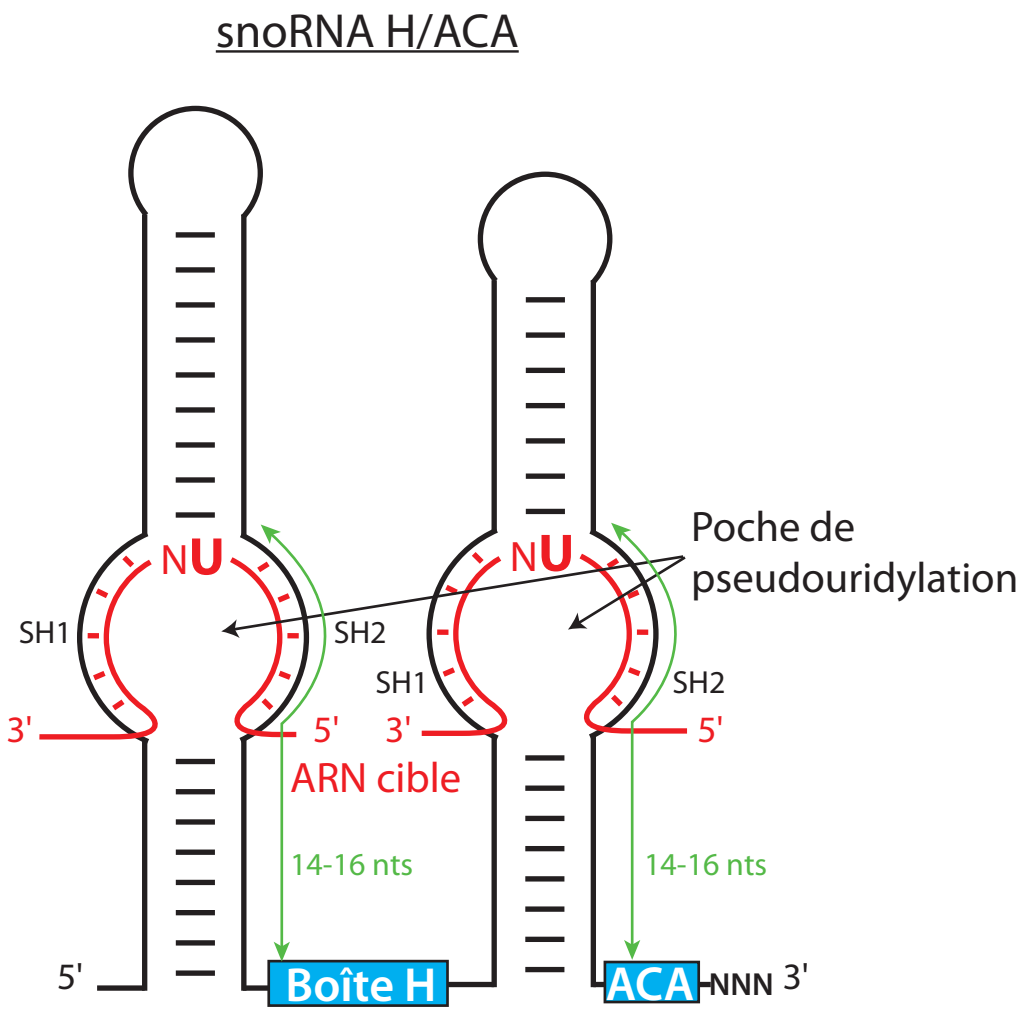


Figure i34 : Représentation schématique de la structure secondaire des snoRNA à boîtes H/ACA eucaryotes et appariement de chacune des poches de pseudouridylation avec un ARN cible.

Les boîtes H et ACA sont encadrées. Les éléments antisens notés s1 et s2, présents dans les boucles internes de chacune des deux tige-boucles présentent des complémentarités de séquences avec des régions de l'ARN cible, conduisent à la formation des hélices SH1 et SH2. La distance entre le résidu U cible de la réaction et les boîtes H ou ACA est comprise entre 14 et 16 nts.

de mon travail de recherche durant ma thèse : le système de modification guidé par les sno/sRNP à boîtes H/ACA.

1. L'ARN guide à boîtes H/ACA

1.1. Chez les Eucaryotes

Les snoRNA à boîtes H/ACA constituent la deuxième grande classe d'ARN guides à avoir été découverte chez les eucaryotes à partir d'alignements de séquences (Balakin *et al.*, 1996). La grande majorité sert de guides pour la réaction de pseudouridylation. Leur taille varient entre 130 et 600 nts (Ganot *et al.*, 1997b). Ils possèdent une séquence 5'-ACA-3' conservée, située trois nucléotides en amont de l'extrémité 3' qui est dénommée boîte ACA (Balakin *et al.*, 1996). Ils sont constitués de deux structures tige-boucle qui sont toutes les deux essentielles à l'activité de pseudouridylation (Fig. i34) (Bortolin *et al.*, 1999). Elles sont séparées par une région charnière en simple brin qui contient la boîte H (Hinge) de séquence conservée 5'-AnAnnA-3' (Ganot *et al.*, 1997b). Les boîtes H et ACA sont nécessaires à la stabilité et à la maturation des snoRNA ainsi qu'à l'activité de pseudouridylation (Balakin *et al.*, 1996; Ganot *et al.*, 1997b; Bortolin *et al.*, 1999). Chaque tige-boucle contient une boucle interne appelée la poche de pseudouridylation. Ses deux brins sont complémentaires à l'ARN cible qui possède deux résidus non appariés à cette poche dont le résidu U à convertir en pseudouridine (Ganot *et al.*, 1997a; Ni *et al.*, 1997). Cet appariement formé par la partie 5' de la poche de pseudouridylation du snoRNA et l'élément antisens 3' de l'ARN cible est noté SH1 pour substrate-helix 1. De la même façon, le second appariement formé entre la région 3' de la poche de pseudouridylation et l'élément antisens 5' de l'ARN cible est appelé SH2 pour substrate-helix 2 (Fig i34). Il existe une contrainte pour la distance située entre le site de modification et l'une des 2 boîtes. En effet, le résidu U cible de la réaction est toujours localisé 14 à 16 nucléotides en amont de la boîte H ou ACA (Fig. i34) (Ganot *et al.*, 1997a; Ni *et al.*, 1997). Depuis la découverte des snoRNA, les méthodes biochimiques et/ou bioinformatiques sont employées pour rechercher de nouveaux candidats de

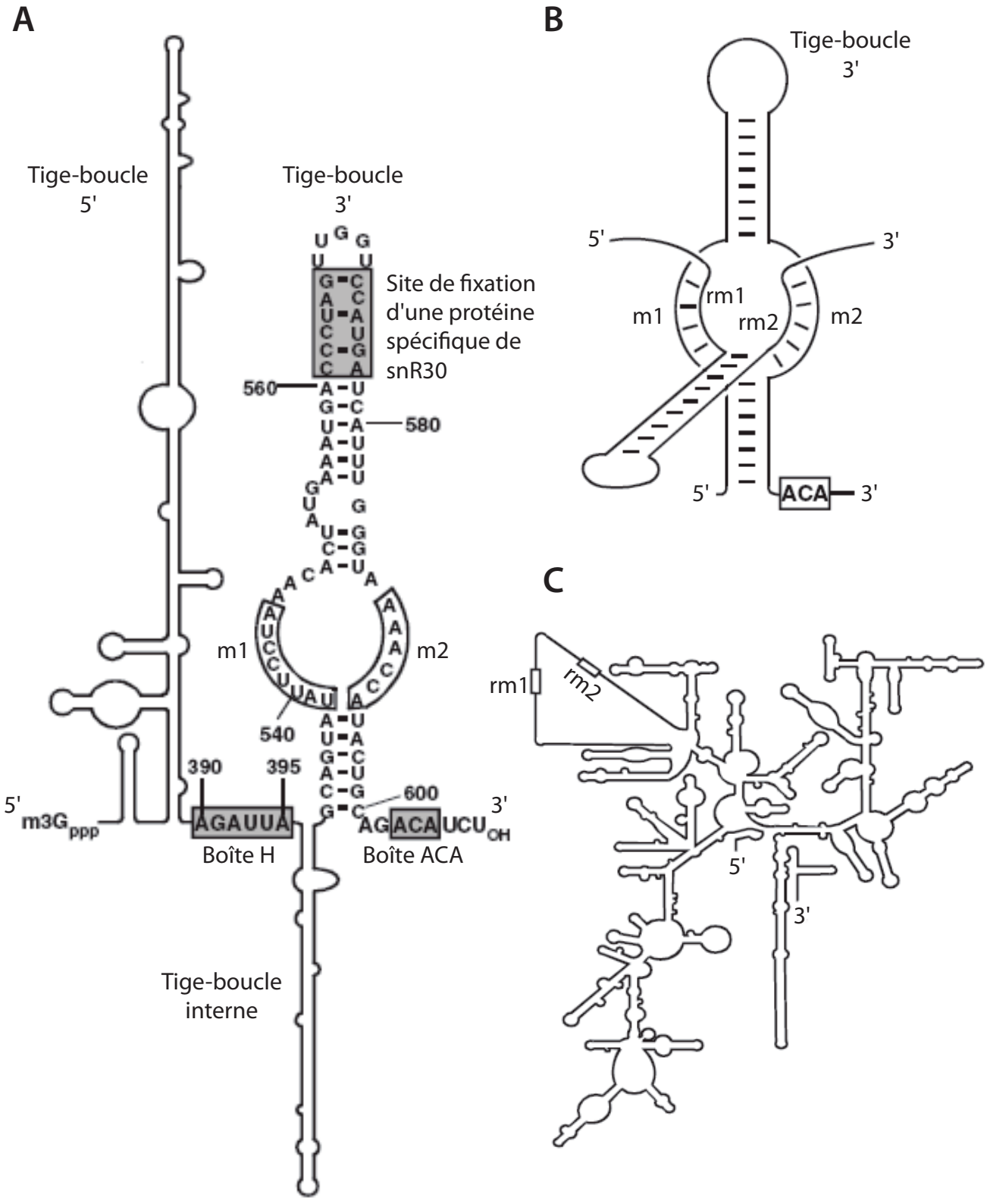


Figure i35 : Le snoRNA à boîte H/ACA snR30 de levure possède des éléments essentiels pour cibler le pré-ARNr 18S. (Fayet-Lebaron *et al.,* 2009).

A. Modèle de structure 2D du snoRNA snR30 de levure. Les tige-boucles 5', 3' et internes sont indiquées ainsi que les boîtes H et ACA. Les séquences m1 et m2 sont encadrées ainsi que le site récemment identifié et proposé représenter un site de fixation d'une protéine spécifique de snR30.

B. Représentation schématique de l'interaction entre le pré-ARNr18S et la tige-boucle 3' de snR30. Les séquences rm1 et rm2 correspondent aux séquences présentes sur le pré-ARNr 18S qui s'apparient aux séquences m1 et m2 de snR30.

C. Structure 2D schématique de l'ARNr 18S de levure. La position des séquences rm1 et rm2 est indiquée.

snoRNA H/ACA dans les génomes de nombreux organismes. Il a ainsi été identifié 82 snoRNA H/ACA chez l'homme qui ciblent 97 sites de pseudouridylation dans l'ARNr (Kiss *et al.*, 2004; Schattner *et al.*, 2006). Mais quelques résidus identifiés en tant que pseudouridines dans l'ARNr ne possèdent ni d'enzyme, ni de snoRNA associés et sont désignés en tant que positions orphelines (Kiss *et al.*, 2004).

Certains snoRNA H/ACA d'organismes eucaryotes qui ont divergés très précocement (*Euglena gracilis* ou *Trypanosoma*) dérogent à la règle du double motif tige-boucle. En effet, plusieurs snoRNA H/ACA sont composés d'une seule tige-boucle et d'une boîte AGA ou d'un triple motif tige-boucle avec une seule boîte H et ACA (Russell *et al.*, 2004; Myslyuk *et al.*, 2008). La structure de ces snoRNA H/ACA particuliers semble se rapprocher des sRNA H/ACA d'archaea (Fig. i38).

Quelques snoRNA H/ACA possèdent une autre fonction que celle jouée en tant qu'ARN guide de la pseudouridylation. Le snoRNA à boîtes H/ACA U17 chez l'homme et snR30 chez la levure ciblent une séquence présente dans l'ARNr et sont essentiels à la production de l'ARNr 18S mature (Fig. i35 A) (Morrissey et Tollervey, 1993; Atzorn *et al.*, 2004). Cet ARN guide servirait de molécule chaperonne qui faciliterait le repliement correct du pré-ARNr lors de la biogenèse des ribosomes (Atzorn *et al.*, 2004). En effet, il a été démontré que la partie inférieure de la poche de pseudouridylation (notée m1 et m2) de la tige-boucle 3' de snR30 s'appariait à deux courtes séquences conservées de l'ARNr 18S (notées rm1 et rm2) lors de la maturation de celui-ci (Fig. i35 B et C) (Fayet-Lebaron *et al.*, 2009). En outre, le snoRNA snR30 porte une séquence essentielle à la maturation de l'ARNr 18S, localisée dans la partie distale de sa tige-boucle 3', qui fixerait une protéine spécifique à ce snoARN et additionnelle aux 4 autres protéines cœurs (Cbf5p, Nhp2p, Nop10p et Gar1p) (Fig. i35 A) (Fayet-Lebaron *et al.*, 2009). La fixation de protéines spécifiques à snR30 lui permettrait d'intervenir dans la maturation de l'ARNr en le distinguant des snoRNA H/ACA classiques qui guident la pseudouridylation des ARN.

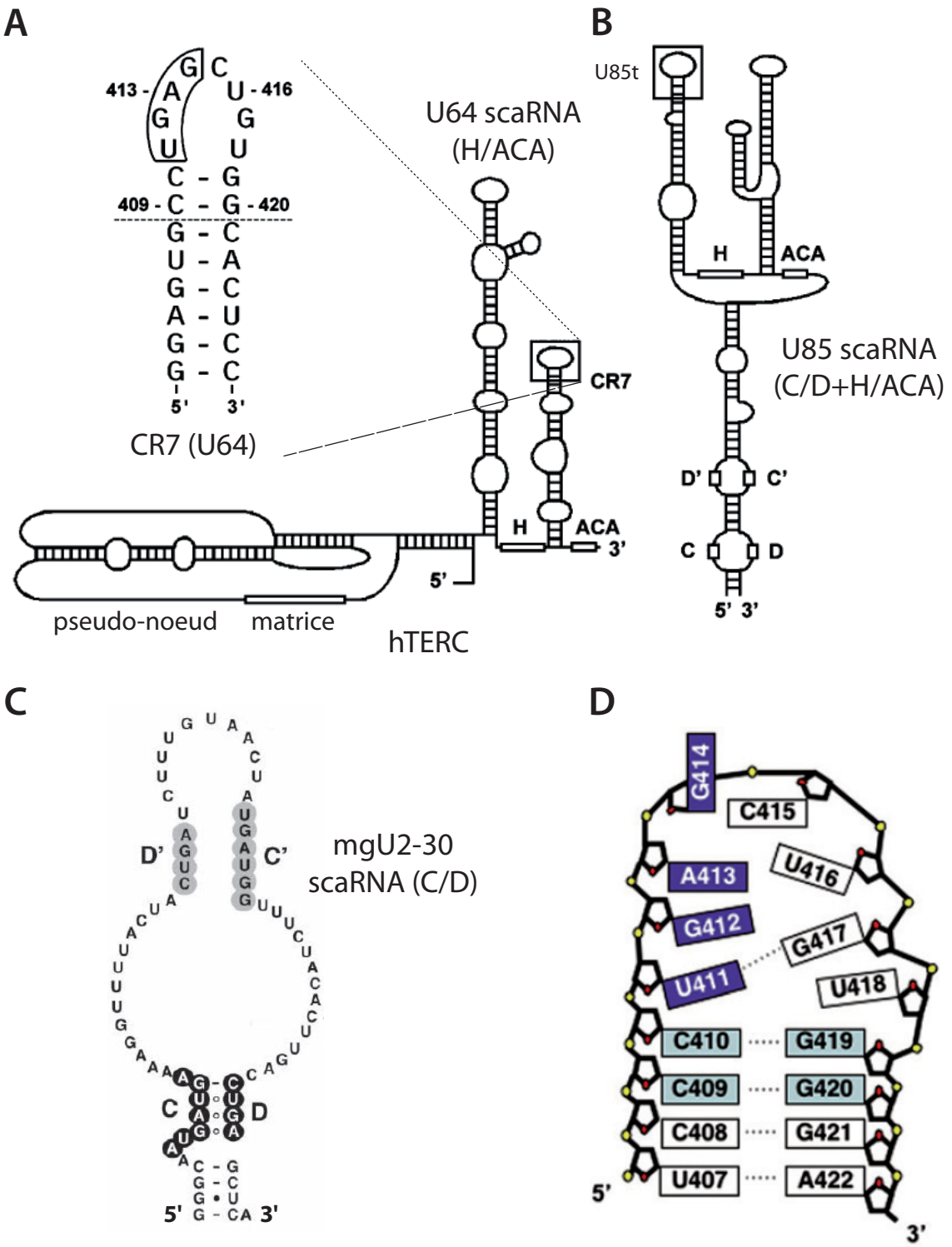


Figure i36 : Les scaRNA et leur signal de localisation dans les Corps de Cajal. (Theimer *et al.,* 2007).

A. à C. Structure secondaire de hTERC qui contient le scaRNA U64 H/ACA humain (**A.**), du scaRNA U85 C/D et H/ACA humain (**B.**) et du scaRNA mgU2-30 C/D humain (**C.**). Les séquences encadrées correspondent aux 4 résidus UGAG de la boîte CAB contenue dans la région conservée CR7 de hTERC. Les nucléotides représentés sous la ligne pointillée correspondent à des séquences non naturelles ajoutées à l'ARN dont la structure a été établie par RMN (en **D.**).

D. Structure du signal de localisation dans les Corps de Cajal de hTERC établie par RMN. Les nucléotides du motif CR7 représentés en bleu foncé correspondent aux 4 résidus de la boîte CAB de hTERC. Les nucléotides en bleu clair correspondent aux résidus qui atténuent la localisation. Les nucléotides représentés en blanc n'ont aucun rôle sur la localisation de hTERC. Ces données ont été obtenues à partir de résultats de mutagenèse et de comparaison de séquences chez les Vertébrés.

1.2. Les scaRNA et l'ARN de la télomérase

Les scaRNA (Small Cajal Bodies specific RNA) constituent une classe de snoRNA et peuvent comporter soit un motif snoRNA à boîte C/D (ex. mgU2-30), soit un motif H/ACA (ex. U64 de hTERC), soit pour certains les deux à la fois (ex. U85) (Fig i36 A à C). Ils sont localisés dans les Corps de Cajal, dans lesquels ils guident la modification des snRNA lorsqu'ils y transitent, à l'exception de hTERC qui sera détaillé dans le paragraphe suivant (Fig. i30) (Darzacq *et al.*, 2002). Il a été montré que les scaRNA possédaient une boîte CAB, située dans l'une des boucles apicales de l'ARN, de séquence (5'-ugAG-3') dont les nucléotides A et G sont mieux conservés et que sa mutation conduisait à relocalisation du scaRNA dans le nucléole (Richard *et al.*, 2003). Elle n'est requise ni pour l'expression, ni pour la stabilité des scaRNA (Richard *et al.*, 2003).

L'ARN de la télomérase est dénommé TERC (Telomerase RNA Component). C'est un transcrit de l'ARN polymérase II et qui porte une coiffe triméthylguanosine en 5' (Mitchell *et al.*, 1999a; Jady *et al.*, 2004). Celui-ci possède une partie commune dans son domaine 5', appelée pseudo-nœud, qui est conservée depuis les organismes ciliés en passant par les levures jusqu'aux vertébrés (Fig. i37) (Chen et Greider, 2004). Chez l'homme, le domaine 3' de TERC forme le motif scaRNA U64 à boîtes H/ACA (Fig. i36 A) (Mitchell *et al.*, 1999a). La scaRNP U64 s'assemble *in vitro* avec les protéines cœurs associées aux H/ACA (Mitchell *et al.*, 1999b; Dragon *et al.*, 2000; Dez *et al.*, 2001). Le scaRNA U64 ne cible pas de séquence connue dans l'ARNr mais est essentiel *in vivo* pour l'accumulation et la maturation en 3' de hTERC ainsi que pour l'activité de la télomérase (Mitchell *et al.*, 1999a). La protéine WDR79/TCAB1 (Telomerase CAjal Bodies protein 1) a récemment été identifiée comme étant responsable de la localisation dans les Corps de Cajal de hTERC (Tycowski *et al.*, 2009; Venteicher *et al.*, 2009). Cette protéine fixe spécifiquement la boîte CAB (Fig. i36D) des scaRNA C/D et H/ACA ce qui concourt à leur localisation dans les Corps de Cajal (Theimer *et al.*, 2007; Tycowski *et al.*, 2009). Le domaine 5' de TERC est indispensable à l'activité télomérase. Il contient le pseudo-nœud ainsi que la matrice nécessaire à l'appariement avec l'ADN sb des répétitions présentes à l'extrémité des

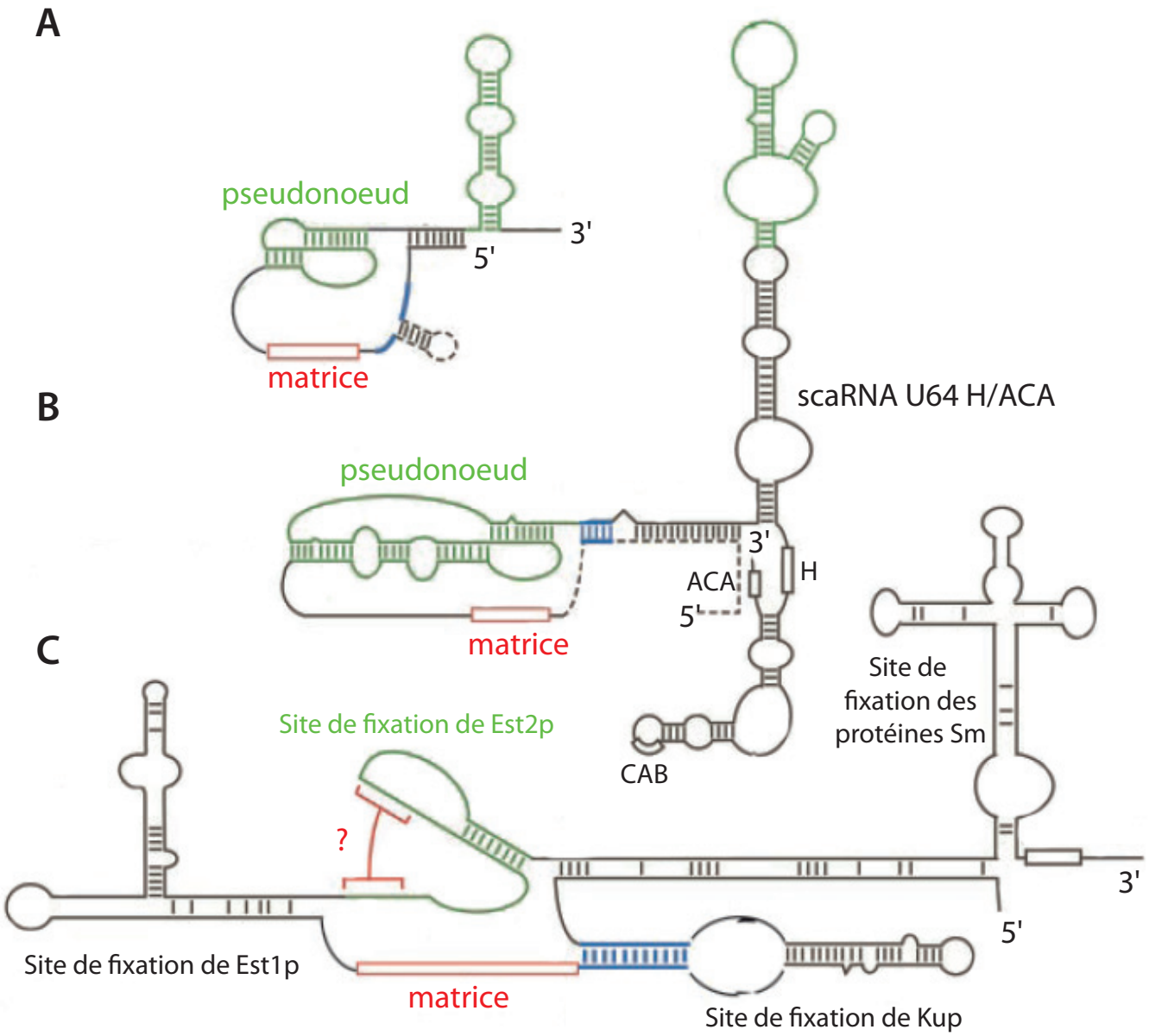


Figure i37 : Modèle de structure secondaire de l'ARN de la télomérase TERC chez différents organismes. (Adapté de Chen et Greider, 2004).
Les éléments en vert représentent les structures conservées qui interagissent avec TERT. Les structures qui ne sont pas conservées chez toutes les espèces d'un même groupe sont représentées en pointillés. La région définissant les limites de la matrice est représentée en rouge.
A. Structure de TERC chez les ciliés composée de 150 nucléotides.
B. Structure de TERC (ou hTERC chez l'homme) de vertébrés composée de 500 nucléotides.
C. Structure de TERC de levures (TLC1) composée de 1300 nucléotides.

chromosomes (Fig. i37 B) (Chen et Greider, 2005). Le pseudo-nœud contient une structure conservée en triple hélice qui contribue à l'activité de la télomérase mais pas à la fixation des protéines associées à la télomérase (Qiao et Cech, 2008). En effet, certains groupements 2'OH provenant de cette triple hélice participeraient à la catalyse de la télomérase et pourraient orienter la matrice vers le site actif de l'enzyme (Qiao et Cech, 2008). Il est intéressant de souligner que l'ARN de la télomérase de levure (TCL1) est dépourvu de motif snoRNP H/ACA en 3' mais que celui-ci est remplacé par un site de fixation aux protéines Sm, les mêmes protéines Sm qui interviennent dans l'épissage en se fixant aux snRNA (Fig. i37 C) (pour revue, (Cech, 2004)).

1.3. Chez les Archaea

Les sRNA à boîtes H/ACA d'Archaea ont été identifiés postérieurement à leurs homologues eucaryotes. Les premiers sRNA H/ACA ont été découverts chez *Archaeoglobus fulgidus* et il a été montré expérimentalement que ces sRNA guidaient bien la formation de pseudouridines (Tang *et al.*, 2002a). A la différence des snoRNA eucaryotes, les sRNA H/ACA d'archaea sont hétérogènes en taille. En effet, ils peuvent être constitués d'une, de deux ou de trois structures en tige-boucle (Fig. i38) (Rozhdestvensky *et al.*, 2003). Les sRNA ne possèdent pas de boîte H mais uniquement une boîte ACA (5'-AnA-3') directement en aval de chaque motif tige-boucle. Ils ont la particularité de posséder un motif en K-turn (ou pseudo-K-turn) dans l'hélice H2 de chaque tige-boucle. Le motif en K-turn est toujours situé à 5 ou 6 paires de bases du résidu U cible de la réaction, tandis que la boîte ACA est localisée comme chez les snoARN H/ACA à 14 ou 15 pb de ce même résidu (Fig. i38) (Rozhdestvensky *et al.*, 2003).

Au laboratoire, S. Muller a recherché des gènes pouvant coder des sRNA H/ACA sur des génomes entiers d'archaea, à l'aide des logiciels de bioinformatique ERPIN et RNAMOT. Ces recherches sont basées sur la présence d'un motif en K-turn, des contraintes de distance et de taille évoquées ci-dessus (mais autorisant un

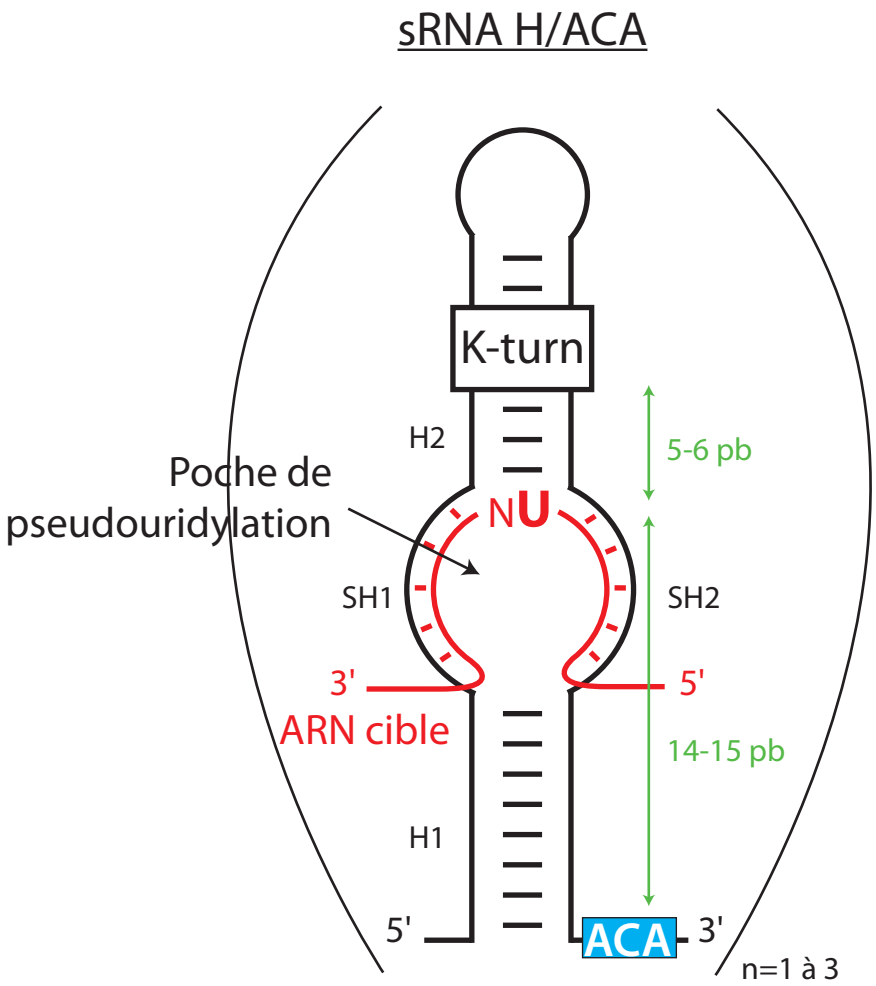


Figure i38 : Représentation schématique de la structure secondaire des sRNA à boîtes H/ACA d'archaea.
Le motif de base, qui peut être répété jusqu'à 3 fois, est composé d'une structure tige-boucle comprenant toutes les caractéristiques décrites ci-dessous. La boîte ACA et le motif en K-turn sont encadrés. Les éléments antisens notés s1 et s2, présents dans les boucles internes de chacune des deux tige-boucles présentent des complémentarités de séquences avec des régions de l'ARN cible et conduisent à la formation des hélices SH1 et SH2. La distance entre le résidu U cible de la réaction et la boîte ACA est comprise entre 14 et 15 paires de bases et la distance séparant ce résidu U cible du motif en K-turn est de 5 à 6 paires de bases. Les hélices ARN H1 et H2 sont également indiquées.

degré de variabilité) ainsi que sur l'existence d'appariements des sRNA candidats avec un ARN cible (Muller *et al.*, 2007a). J'ai participé à la caractérisation du rôle fonctionnel des sRNA de *P. abyssi* identifiés par ce travail et ils seront présentés dans la partie I.A des résultats.

2. Les protéines associées aux RNP à boîtes H/ACA

2.1. Chez les Eucaryotes

Parmi les protéines associées aux snoRNA à boîtes H/ACA, on distingue i) les protéines associées transitoirement et qui interviennent dans la biogenèse des snoRNP et ii) les 4 protéines cœur associées de façon stable et qui sont directement en lien avec la réaction de pseudouridylation. Ces dernières sont appelées Cbf5p, Gar1p, Nop10p et Nhp2p chez la levure ; et respectivement, Dyskérine, GAR1, NOP10 et NHP2 chez l'homme : (Tab. i6) (pour revue, (Meier, 2005)).

La protéine Cbf5 /Dyskérine

Initialement, la protéine Cbf5 (483 résidus chez la levure/ 524 résidus chez l'homme) a été identifiée chez la levure pour interagir *in vivo* mais avec une faible affinité avec les centromères et *in vivo* et *in vitro* avec les microtubules, d'où le nom donné à Cbf5p (Centromere Binding Factor 5) (Jiang *et al.*, 1993). L'analyse de sa séquence a révélé la présence de 10 motifs répétés KKE/D dans sa partie Ct (Fig. i19). La protéine Cbf5 entière est essentielle à la viabilité cellulaire, mais une protéine avec une extrémité réduite à 3 motifs conduit à une croissance cellulaire ralentie suggérant que Cbf5p interviendrait dans le cycle cellulaire (Jiang *et al.*, 1993). Son interaction *in vitro* avec les microtubules se fait par l'intermédiaire de la région Nt de Cbf5p (Jiang *et al.*, 1993).

Plus tard, l'activité ARN:Ψ-synthase a été attribuée à cette protéine en raison de sa forte homologie de séquence avec TruB et de la présence dans sa séquence des

Protéines coeur des snoRNP à boîtes H/ACA

Eucaryotes		Archaea
Levure	**Homme**	
Cbf5p	Dyskérine	aCBF5
Nhp2p	NHP2	L7Ae
Nop10p	NOP10	aNOP10
Gar1p	GAR1	aGAR1

Tableau i6 : Comparatif des différentes protéines composant les snoRNP à boîtes H/ACA chez les Eucaryotes (levure et homme) et chez les Archaea.

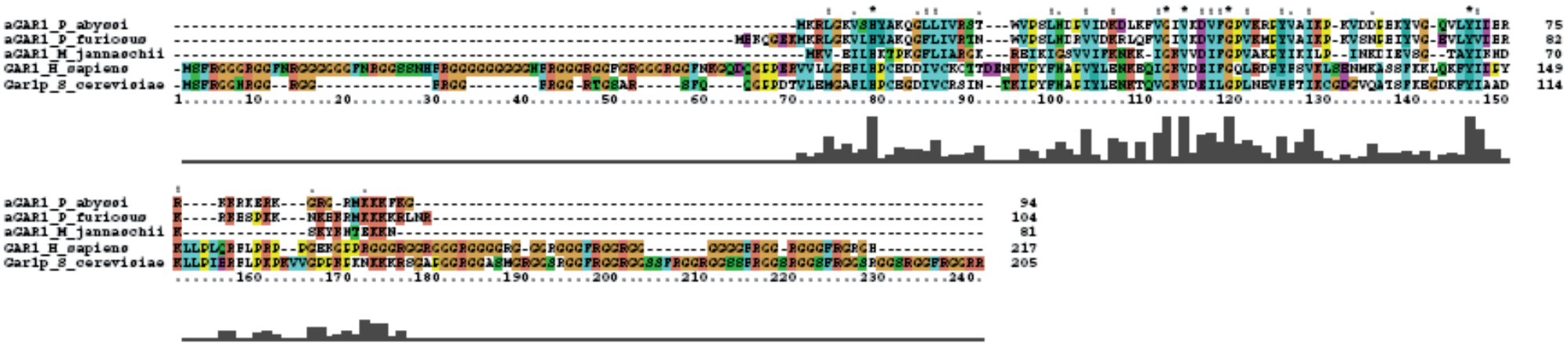
motifs I et II caractéristiques de cette classe d'enzymes (cf III.D) (Koonin, 1996). La substitution de l'acide aspartique catalytique par une alanine abolit l'activité *in vivo* de pseudouridylation dans les ARNr (Zebarjadian *et al.*, 1999). La localisation nucléolaire prédominante de Cbf5p suggérait que son rôle principal dans la cellule n'était pas de participer aux mécanismes de ségrégation des chromosomes. Son implication dans la biogenèse de l'ARNr a ainsi été démontrée (Cadwell *et al.*, 1997).

Enfin, il a été montré que Cbf5p était associée spécifiquement aux snoRNA à boîte H/ACA (Watkins *et al.*, 1998). Par expression conditionnelle chez la levure, car Cbf5p est essentielle à la viabilité cellulaire, il a été montré que la déplétion de Cbf5p inhibait dramatiquement le taux de pseudouridylation dans l'ARNr, faisait chuter le taux d'ARNr mature, et déstabilisait spécifiquement les snoRNA à boîtes H/ACA (Lafontaine *et al.*, 1998).

La protéine Gar1/GAR1

La protéine Gar1 (205 résidus chez la levure et 217 résidus chez l'homme) est une protéine nucléolaire essentielle à la viabilité cellulaire (Tab. i6). Elle contient à chacune de ses extrémités, deux domaines renfermant des répétitions du motif GAR (Glycine-Arginine Rich) (Fig. i39) (Girard *et al.*, 1992). Ces répétitions ne sont essentielles, ni pour la croissance chez la levure, ni pour l'adressage nucléolaire (Girard *et al.*, 1994). En effet, Gar1p possède un domaine central hautement conservé au cours de l'évolution (Fig. i39) (Watanabe et Gray, 2000). L'absence de cette protéine chez la levure induit une inhibition de la production de l'ARNr mature, empêche la pseudouridylation du pré-ARNr mais pas des ARNt (Bousquet-Antonelli *et al.*, 1997). Cela indiquait que Gar1p était associée aux snoRNA H/ACA (Balakin *et al.*, 1996; Ganot *et al.*, 1997b) et qu'elle jouait un rôle dans la pseudouridylation site-spécifique des ARNr (Bousquet-Antonelli *et al.*, 1997). Néanmoins, Gar1p n'est pas requise pour l'accumulation des snoRNA à boîtes H/ACA (Dez *et al.*, 2001). Une étude a montré que Gar1p, produite par transcription/traduction *in vitro*, pouvait se

Famille GAR1



Famille NOP10



Figure i39 : Alignement de séquences des protéines des familles GAR1 et NOP10.
aGAR1 et aNOP10 chez *P. abyssi*, *P. furiosus*, *M. janaschii* ; GAR1 et NOP10 chez l'homme ; Gar1p et Nop10p chez la levure. Lorsqu'ils sont conservés, les résidus basiques sont indiqués en orange foncé, les résidus acides en violet, les résidus polaires neutres en vert, les résidus hydrophobes en bleu, la glycine en orange clair et la proline en jaune. Les résidus strictement conservés portent une étoile au-dessus de l'alignement ; lorsque les résidus sont conservés avec un changement de résidu au sein de la même classe 2 points sont indiqués et un point lorsque des résidus sont conservés mais présentent un changement de classe. La fréquence de conservation des résidus au sein des différentes séquences est représentée sous l'alignement par un histogramme. Alignement réalisé à l'aide du logiciel Clustal X2 version 2.0.10 et à partir des séquences disponibles.

fixer directement aux snoRNA H/ACA par l'intermédiaire de son domaine central (Bagni et Lapeyre, 1998).

La protéine Nop10/NOP10

La protéine Nop10 (NucleOlar Protein 10) est essentielle à l'activité de pseudouridylation des snoRNA H/ACA et elle est requise pour la stabilité des snoRNA H/ACA (Henras *et al.*, 1998). C'est une petite protéine de 58 résidus chez la levure et de 64 résidus chez l'homme (Tab. i6 ; Fig. i39). Elle est constituée d'un domaine Nt qui se structure sous la forme de 2 brins β et d'un domaine Ct qui est constitué d'une hélice α et qui sont reliés par une région charnière (linker) (Hamma *et al.*, 2005; Khanna *et al.*, 2006; Reichow et Varani, 2008).

La protéine Nhp2/NHP2

La protéine Nhp2 (Non-Histone chromosome Protein 2-like) est composée de 156 résidus chez la levure et de 153 résidus chez l'homme (Tab. i6 ; Fig. i40). Elle possède une fonction essentielle à la physiologie de la cellule (Kolodrubetz et Burgum, 1991). Nhp2p est primordiale à l'activité de pseudouridylation et est requise pour la stabilité des snoRNA à boîtes H/ACA (Henras *et al.*, 1998; Watkins *et al.*, 1998). Cette protéine ne fixerait pas de motif ARN défini, mais s'associerait de façon non spécifique aux ARN (Henras *et al.*, 2001). L'accumulation nucléolaire de Nhp2 dépend de son association avec les snoRNA H/ACA (Henras *et al.*, 2001).

2.2. Chez les Archaea

Les protéines associées aux snoRNA à boîtes H/ACA eucaryotes possèdent des homologues chez les archaea. Il s'agit des protéines aCBF5, aNOP10, aGAR1 et L7Ae qui seront détaillées dans la partie IV.C.6 de l'introduction.

Famille L7Ae / NHP2

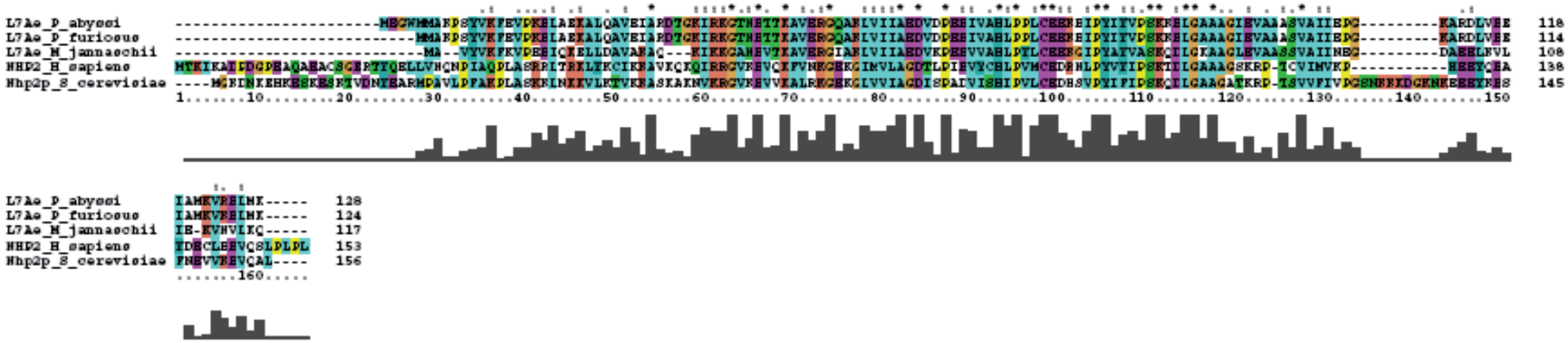


Figure i40 : Alignement de séquences des protéines de la famille L7Ae / NHP2.
L7Ae chez *P. abyssi*, *P. furiosus*, *M. janaschii* ; NHP2 chez l'homme ; Nhp2p chez la levure. Lorsqu'ils sont conservés, les résidus basiques sont indiqués en orange foncé, les résidus acides en violet, les résidus polaires neutres en vert, les résidus hydrophobes en bleu, la glycine en orange clair et la proline en jaune. Les résidus strictement conservés portent une étoile au-dessus de l'alignement ; lorsque les résidus sont conservés avec un changement de résidu au sein de la même classe 2 points sont indiqués et un point lorsque des résidus sont conservés mais présentent un changement de classe. La fréquence de conservation des résidus au sein des différentes séquences est représentée sous l'alignement par un histogramme. Alignement réalisé à l'aide du logiciel Clustal X2 version 2.0.10 et à partir des séquences disponibles.

2.3. Localisation des gènes codant les protéines H/ACA sur les génomes d'archaea

La question posée ici est de savoir si les gènes codant les protéines associées aux sRNP à boîtes H/ACA sont localisés sur un même opéron ou sur plusieurs unités transcriptionnelles dans les génomes d'archaea. Des éléments de réponses sont disponibles et ils indiquent que : i) le gène codant aGAR1 est sur le même opéron que TFIIB, ce qui suggère une régulation couplée de l'expression des composants de la machinerie de transcription et de traduction (Watanabe et Gray, 2000), et ii) des gènes impliqués dans la réplication sont contigus sur le chromosome (en clusters) et selon la même orientation par rapport à des regroupements de gènes intervenant dans la traduction dont le gène codant la protéine aNOP10 (Berthon *et al.*, 2008). Il existerait ainsi un couplage entre la réplication de l'ADN et la traduction (Berthon *et al.*, 2008). Néanmoins, il serait intéressant de connaître la localisation des ARN:Ψ-synthases et la possibilité d'un couplage avec d'autres machineries moléculaires. La question reste ouverte : existe-t-il une coordination de l'expression de ces gènes chez les archaea ?

3. Evolution des s/snoRNP

La protéine L7Ae possède une fonction double chez les Archaea en tant que constituant des sRNP C/D et H/ACA et en tant que composant du ribosome. Ceci suggère que la RNP ancestrale commune aux complexes de modification de nucléotides guidés par un ARN aurait ses origines évolutives parmi les composants de la machinerie primitive de la traduction (Fig. i41) (pour revue, (Tran *et al.*, 2004)). Une hypothèse propose que le motif en K-turn présent dans la machinerie de traduction primitive aurait pu fonctionner comme l'un des éléments moteur du repliement de l'ARNr. Comme je l'ai déjà évoqué (cf IV.B.2.2), ce motif ARN est fortement stabilisé par la fixation d'une protéine, suggérant que la participation des protéines et la formation de RNP s'est effectuée tôt dans l'évolution (Goody *et al.*, 2004). Ensuite, deux possibilités d'évolution ont pu exister :

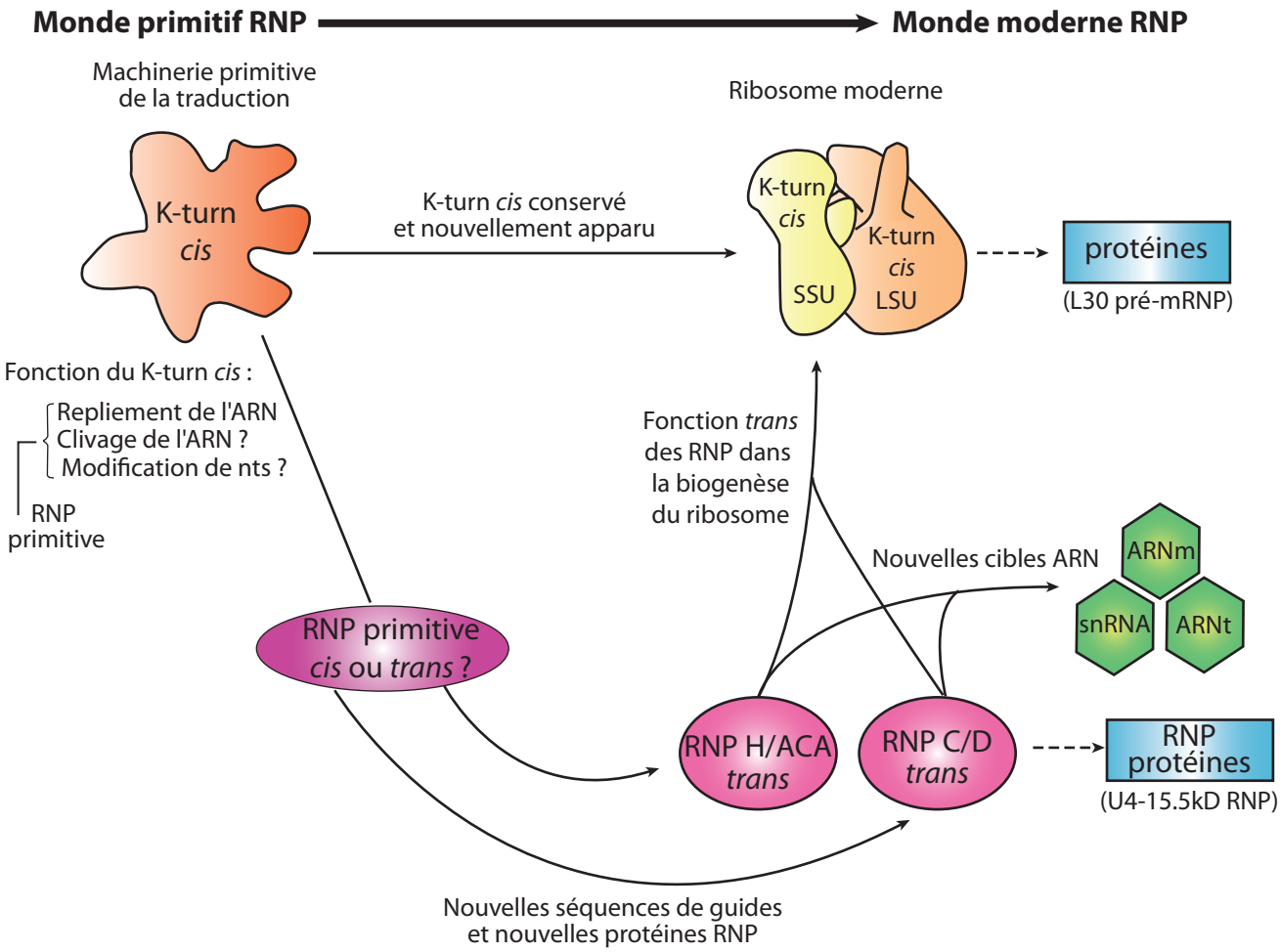


Figure i41 : Evolution des complexes de modification de nucléotides guidés par l'ARN. (Adapté de Tran *et al.,* 2004).
Le K-turn *cis* présent dans la machinerie de traduction primitive aurait évolué suivant deux voies. Soit, il aurait été conservé dans la machinerie de traduction primitive et il se retrouve dans les sous-unités du ribosome actuel. Soit, il aurait été intégré dans des RNP primitives qui elles-même ont évolué en RNP à boîtes C/D et H/ACA. Voir les explications dans le texte.

- Soit un ARN ou une RNP a dû évoluer en tant que complexe primitif qui serait le précurseur des RNP à boîtes C/D et H/ACA actuelles (Fig. i41). Dans un premier temps, cette RNP primitive a pu fonctionner en tant qu'élément *cis* dans la machinerie de traduction puis aurait évolué en diverses RNP qui sont actuellement les RNP à boîtes C/D et H/ACA (pour revue, (Tran *et al.*, 2004)).

- Alternativement la RNP primitive serait d'abord devenue un complexe *trans* indépendant avant d'évoluer en deux familles structurales et fonctionnelles distinctes de RNP jouant le rôle de guide (Fig. i41) (pour revue, (Tran *et al.*, 2004)). Ces deux familles ont ensuite continué d'évoluer indépendamment. L'organisation des motifs ARN (simple motif tige-boucle H/ACA chez les Archaea contre un double motif chez les Eucaryotes) ainsi que la distribution des protéines associées (duplication du gène ancestral de aNOP56/58 d'archaea pour 2 protéines distinctes Nop56 et Nop58 chez les Eucaryotes) sont des exemples d'adaptations réalisées au cours de l'évolution de ces RNPs.

4. La biogenèse des particules snoRNP H/ACA

4.1. Chez les Eucaryotes

Comme déjà mentionné dans le paragraphe des snoRNA à boîtes C/D, de nombreux snoRNA à boîtes H/ACA eucaryotes sont contenus dans les introns des ARN pré-messagers qui sont transcrits par l'ARN polymérase II (Fig. i30 et i31) (pour revue, (Bachellerie *et al.*, 2002)). Chez la levure, la majorité des séquences de snoRNA correspondent à des unités transcriptionnelles individuelles (Maxwell et Fournier, 1995) ou sont regroupées sur un transcrit polycistronique (Qu *et al.*, 1999).

4.1.1. Schéma général de la biogenèse

Le modèle actuel de biogenèse et d'assemblage des snoRNP H/ACA chez les eucaryotes repose sur la formation d'un trimère cytoplasmique contenant les protéines Dyskérine-NOP10-NHP2 (Fig. i42) (Wang et Meier, 2004). Il a été suggéré

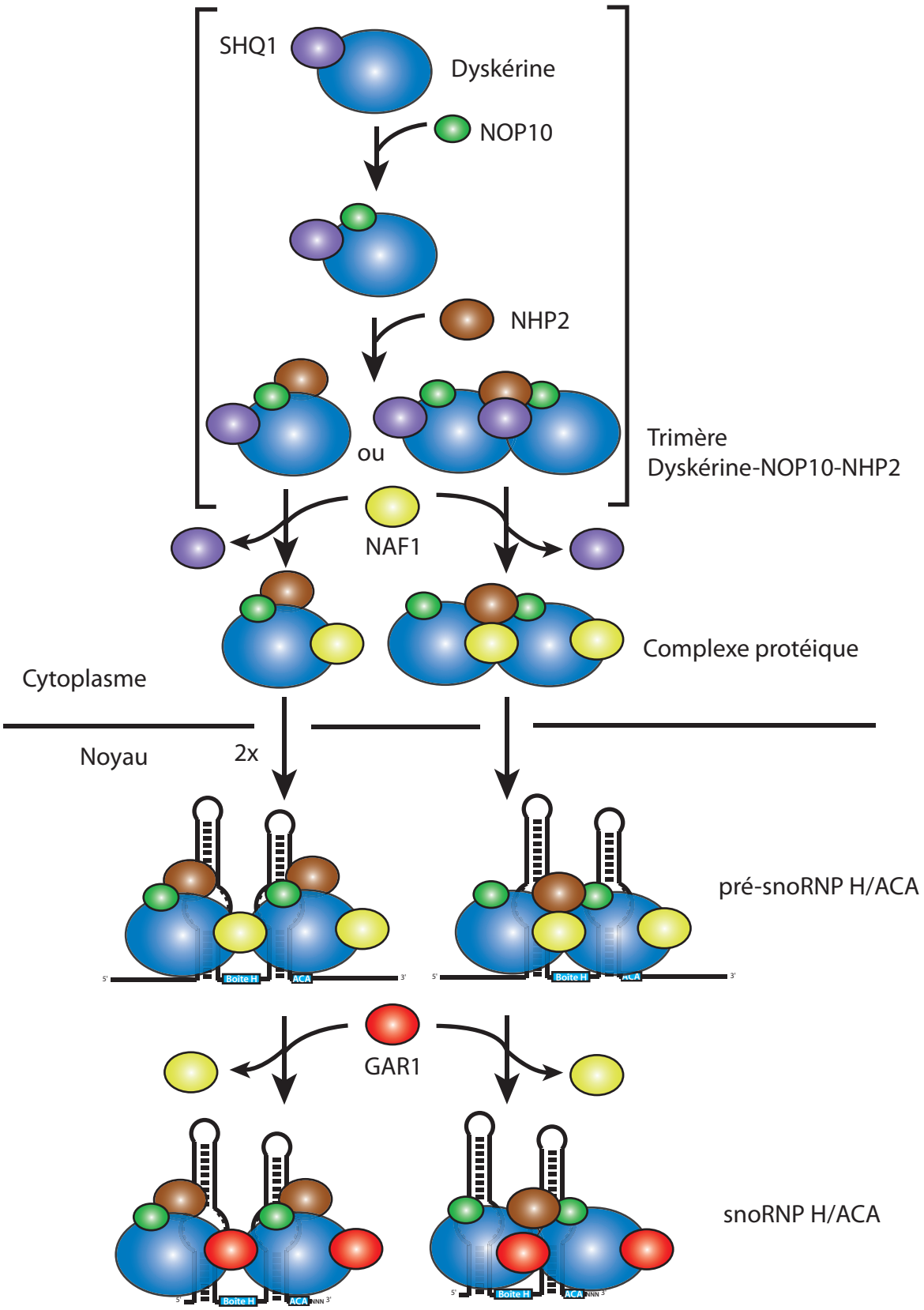


Figure i42 : Voie d'assemblage schématique des snoRNP à boîtes H/ACA chez les eucaryotes. (Adapté de Wang et Meier, 2004 et complété par les données récentes de Darzacq *et al.,* 2006 ; Richard *et al.,* 2006 ; Leulliot *et al.,* 2007 ; Grozdanov *et al.,* 2009).
Dans le cytoplasme les protéines Dyskérine, NOP10 et NHP2 forment un trimère associé à SHQ1 qui est liée à la Dyskérine. La protéine SHQ1 est échangée avec NAF1 qui se lie à la Dyskérine et ce complexe protéique migre dans le noyau au site de transcription du pré-snoRNA à boîtes H/ACA. La protéine GAR1 remplace NAF1 pour former la snoRNP mature.

que SHQ1 (Small nucleolar RNA H/ACA Quantitative accumulation) accompagne la Dyskérine depuis le moment de sa synthèse dans le cytoplasme jusqu'à son association avec NOP10 et NHP2 pour former le trimère (Grozdanov *et al.*, 2009). Puis SHQ1 laisserait sa place à NAF1 (Nuclear Association Factor 1) qui fait la navette du noyau vers le cytoplasme où elle interagit avec la Dyskérine pour la stabiliser (Fig. i42) (Darzacq *et al.*, 2006). Ce complexe serait ensuite importé dans le noyau via la séquence NLS de la Dyskérine ou de NAF1 et va s'assembler sur le snoRNA naissant (Fig. i42) (Darzacq *et al.*, 2006). En outre, il a été montré que chez la levure, Naf1p interagissait avec le domaine CTD phosphorylé de l'ARN pol II et que Naf1p était localisé sur le transcrit naissant (Fig. i42) (Fatica *et al.*, 2002). Plusieurs études indiquent que le trimère Dyskérine-NOP10-NHP2 se fixerait très tôt et de manière co-transcriptionnelle sur le transcrit naissant, tandis que GAR1 arriverait plus tardivement (Henras *et al.*, 1998; Ballarino *et al.*, 2005; Yang *et al.*, 2005; Richard *et al.*, 2006). Il est suggéré que c'est l'assemblage correct de la snoRNP lors de la transcription qui déterminerait la formation de l'extrémité 3' du snoRNA (Ballarino *et al.*, 2005). En effet, le pré-snoRNA est maturé par des mécanismes exonucléolytiques de 5' vers 3' et de 3' vers 5' ayant lieu dans le nucléoplasme. Les précurseurs des snoRNA ne porteraient pas de coiffe en 5', car l'ajout d'une coiffe $m^{2,2,7}G$ inhibe la maturation de pré-snoRNA en extrait de cellules HeLa (Kiss et Filipowicz, 1995). Enfin, l'interaction de la Dyskérine avec NAF1 serait perdue au profit d'une interaction avec la protéine GAR1 pendant la maturation de la snoRNP ou pendant son transit vers les Corps de Cajal (Fig. i42) (Verheggen *et al.*, 2001; Darzacq *et al.*, 2006; Richard et Kiss, 2006).

4.1.2. Les facteurs associés à la biogenèse

Comme l'indique le schéma général de la biogenèse, certains facteurs protéiques tels que Shq1p et Naf1p chez la levure s'associent transitoirement à la pré-snoRNP, ils sont essentiels à la croissance cellulaire et leur déplétion fait chuter spécifiquement le taux des snoRNA H/ACA ainsi que le taux de l'ARNr mature 18S

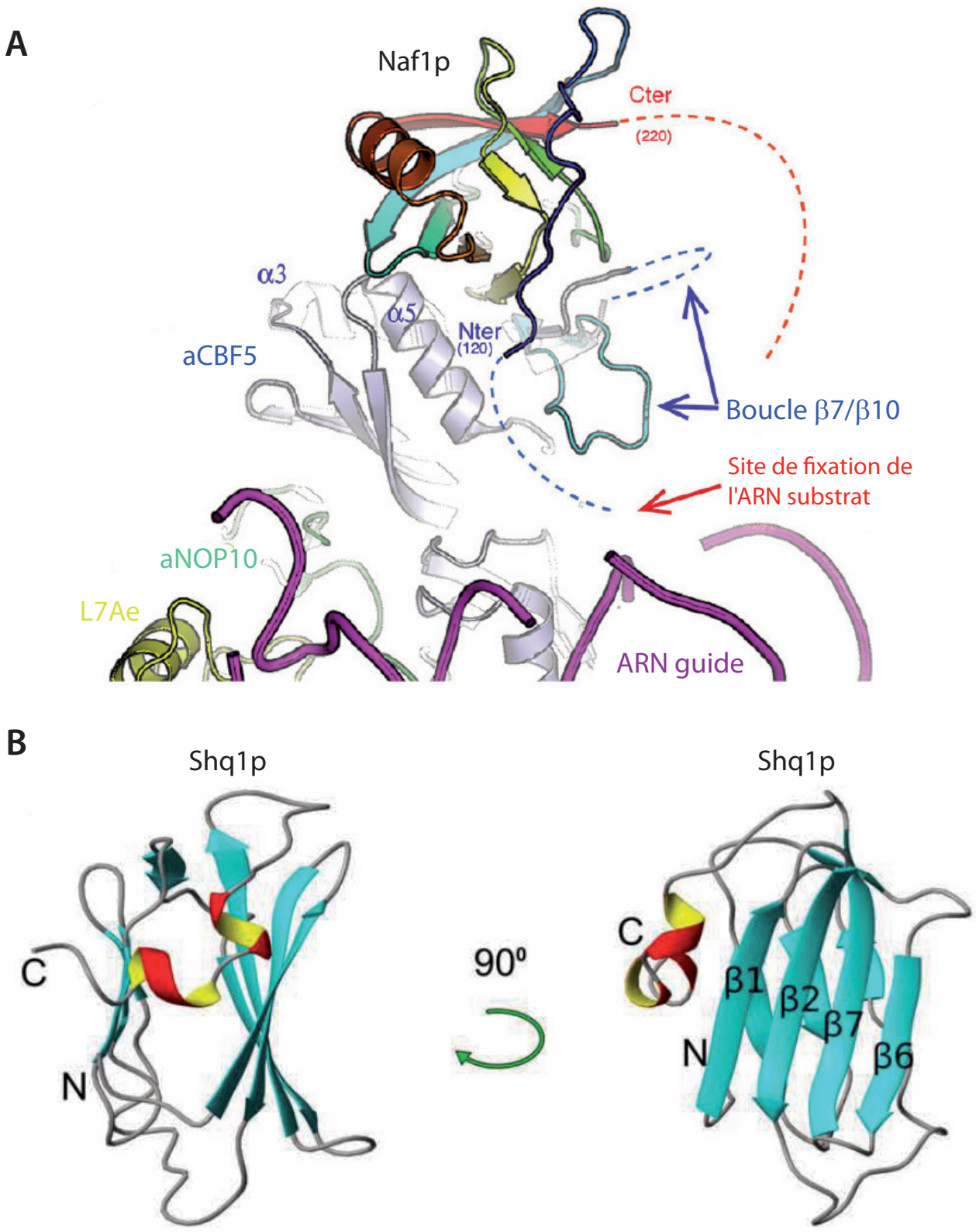


Figure i43 : Structure 3D des protéines Naf1p et Shq1p de levure. (Leulliot *et al.,* 2007 ; Godin *et al.,* 2009).

A. Modèle basé sur la structure 3D de la protéine Naf1p tronquée de ses extrémités Nt et Ct qui a été placée dans le complexe aCBF5-aNOP10-L7Ae-ARN guide d'archaea. La boucle β7/β10 de aCBF5 est représentée dans une conformation fermée et modélisée par des lignes brisées pour la conformation ouverte. Les extensions Nt et Ct de Naf1p, non présentes dans la forme cristallisée, sont respectivement représentées par des traits pointillés bleus et rouges.

B. Structure en solution du domaine CS de Shq1p selon deux orientations. Les brins β sont colorés en cyan tandis que l'hélice α est représentée en rouge et jaune.

(Dez *et al.*, 2002; Fatica *et al.*, 2002; Yang, P. K. *et al.*, 2002). NAF1 humaine est également requise pour l'accumulation des sno/scaRNA H/ACA (Hoareau-Aveilla *et al.*, 2006). Le remplacement de NAF1 par GAR1 sur la Dyskérine est rendu possible par la présence du domaine central de Naf1p qui présente de très fortes homologies structurales avec la protéine aGAR1 d'archaea (Leulliot *et al.*, 2007). Cette dernière interagit avec la boucle de chargement du substrat β7/β10 de aCBF5 (Fig. i43 A). Ces observations suggèrent que Naf1p pourrait interagir avec la boucle β7/β10 de Cbf5p et moduler sa capacité à charger et à libérer l'ARN substrat. De plus les extensions Nt et Ct de Naf1p pourraient contribuer à inhiber l'enzyme pendant son transport en l'empêchant de charger des ARN qui ne représentent pas des substrats pour la particule (Leulliot *et al.*, 2007).

D'autres facteurs joueraient le rôle de protéines chaperonnes qui organiseraient les snoRNP H/ACA durant la modification de l'ARNr telle que la protéine NOPP140/Srp40p qui s'associe de façon réversible avec Cbf5p (Meier et Blobel, 1994; Wang *et al.*, 2002). La protéine Shq1 possède à son extrémité N-terminale un domaine CS (pour CHORD-containing proteins and SGT1), essentiel à sa fonction. Bien que son repliement soit similaire à celui retrouvé dans des protéines co-chaperonnes de la protéine chaperonne Hsp90p, le domaine CS n'interagit pas avec Hsp90p *in vitro* (Fig. i43 B) (Godin *et al.*, 2009; Singh *et al.*, 2009). L'activité chaperonne portée par le domaine Ct serait donc indépendante de Hsp90p et serait requise pour le repliement correct et l'assemblage des composants des snoRNP H/ACA (Godin *et al.*, 2009; Singh *et al.*, 2009). Le complexe SMN (Suvival MotoNeuron), qui intervient dans l'assemblage des snRNP, interagit avec les protéines GAR1 et Fibrillarine, ce qui suggère qu'il pourrait avoir une fonction dans la biogenèse de plusieurs classes de RNP (Pellizzoni *et al.*, 2001). La contribution du complexe HSP90-R2TP-NUFIP/Rsa1p à la biogénèse de RNP dont les snoRNP C/D et H/ACA a été proposée récemment (Boulon *et al.*, 2008). Comme je l'ai mentionné précédemment pour les snoRNP C/D, la protéine NUFIP/Rsa1 servirait de plateforme d'interaction entre plusieurs protéines. Dans les cas des snoRNP H/ACA, celle-ci permettrait le recrutement de NHP2. En effet, les protéines NUFIP et NHP2

interagissent dans des tests de double hybride et l'interaction de NUFIP avec des snoRNA H/ACA a été observée lors d'expériences d'immunoprécipitations dans des extraits de cellules HeLa. Ces données suggèrent donc un rôle possible de NUFIP et de son complexe associé HSP90-R2TP lors de la biogenèse des snoRNP H/ACA (Boulon *et al.*, 2008). De plus, l'inhibition de l'activité de la protéine HSP90 conduit i) à une baisse de la stabilité des protéines 15.5 kD et NOP58 mais aussi de la protéine NHP2, ii) à une baisse de la stabilité des ARN U4 snRNA, U3 snoRNA et de l'ARN télomérique hTERC. Il est ainsi proposé que HSP90, par son interaction avec le complexe R2TP-NUFIP participerait à la biogenèse des snoRNP C/D et H/ACA en contrôlant la structuration des protéines des RNP au moment de leur assemblage sur l'ARN (Boulon *et al.*, 2008). L'interaction détectée entre Cbf5p et un facteur de transcription de l'ARN polymérase I, suggère que la modification du pré-ARNr pourrait avoir lieu de façon concomitante à sa transcription (Cadwell *et al.*, 1997).

4.2. Chez les Archaea

Aucuns homologues aux protéines Naf1p et Shq1p n'ont été trouvés chez les Archaea, ce qui suggère que leur fonction dans la biogenèse des snoRNP H/ACA est spécifique aux Eucaryotes (Darzacq *et al.*, 2006). Aucune donnée concernant les événements de maturation conduisant aux sRNP H/ACA matures d'archaea n'existe à ce jour. Néanmoins, le développement ces dernières années d'outils de génétique moléculaire chez les archaea pourraient permettre dans le futur d'étudier la biogenèse des sRNP H/ACA *in vivo*.

5. Reconstitution de particules à boîtes H/ACA

5.1. Chez les Eucaryotes

Historiquement, la première observation d'une snoRNP à boîtes H/ACA par microscopie électronique a suggéré une forme composée par deux motifs distincts, ce qui a conduit les auteurs à proposer un modèle de bi-particule (Watkins *et al.*, 1998).

Puis, plusieurs équipes ont reconstituées une particule snoRNP à boîtes H/ACA à partir d'un snoRNA transcrit *in vitro* puis incubé soit dans des extraits nucléaires de cellules HeLa (Dragon *et al.*, 2000; Wang *et al.*, 2002), ou soit avec les protéines de ces particules produites par transcription et traduction dans des extraits de réticulocytes de lapins (Wang et Meier, 2004). Les résultats obtenus ont permis de montrer une association spécifique des 4 protéines Dyskérine, NOP10, NHP2 et GAR1 aux snoRNA à boîtes H/ACA dont le snoRNA U64 de hTERC (Dragon *et al.*, 2000; Wang *et al.*, 2002; Wang et Meier, 2004). Néanmoins, les particules reconstituées par ces approches ne présentaient qu'une très faible activité ARN:Ψ-synthase. Un autre système a été développé permettant de reconstituer sur un transcrit H/ACA d'intérêt une particule avec une forte activité ARN:Ψ-synthase. Ce système est basé dans un premier temps sur la purification des snoRNP de levure par affinité par l'intermédiaire de l'étiquette TAP (Ma *et al.*, 2005), puis la préparation est traitée à la nucléase micrococcale permettant de dégrader les ARN. La fraction protéique obtenue permet de reconstituer une snoRNP fonctionnelle après incubation avec un transcrit *in vitro* de snoRNA H/ACA (Ma *et al.*, 2005; Karijolich *et al.*, 2007; Xiao *et al.*, 2009). Néanmoins pour l'instant, aucun système de reconstitution basé sur l'emploi des composants protéiques purifiés n'a été décrit. La raison principale est que la production des protéines de ces particules sous forme recombinante est très difficile à obtenir chez *E. coli*.

5.2. Chez les Archaea

Les premières reconstitutions *in vitro* de RNP à boîtes H/ACA réalisées à partir des composants protéiques purifiés à homogénéité ont été obtenues en employant des protéines d'archaea produites sous forme recombinantes chez *E. coli* et d'ARN transcrits *in vitro* (Baker *et al.*, 2005; Charpentier *et al.*, 2005). Bruno Charpentier a développé au laboratoire les conditions pour la reconstitution des particules sRNP H/ACA de *P. abyssi* en employant le sRNA Pab91 comme modèle

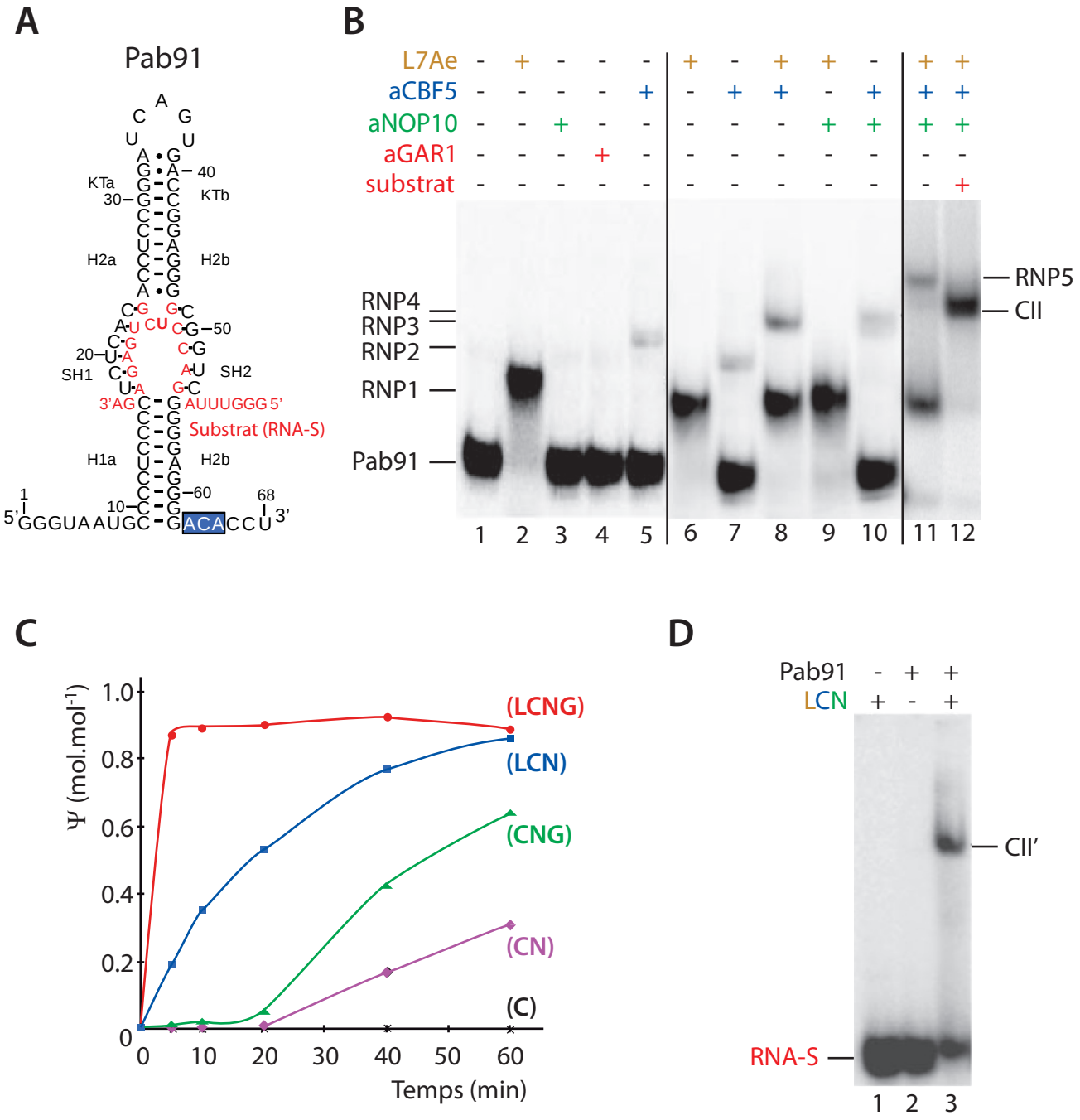


Figure i44 : Reconstitution de particules sRNP à boîtes H/ACA d'archaea, actives pour la réaction de pseudouridylation (Charpentier *et al.,* 2005).

A. Structure schématique du sRNA Pab91 à boîtes H/ACA de *P. abyssi* associé à un ARN substrat (RNA-S).

B. Association des protéines L7Ae (L), aCBF5 (C), aNOP10 (N) ou aGAR1 (G) au sRNA Pab91 à boîtes H/ACA radiomarqué en 5' par la technique de retard de migration sur gel. La position des complexes (RNP1 à RNP5) et du complexe en présence de substrat (CII) est indiquée ainsi que le mélange de protéines ou du substrat ajoutés (+ ou -).

C. Etude cinétique de la formation de résidus pseudouridines au cours du temps. Les conditions d'incubation de différents mélanges de protéines en présence du sRNA Pab91 et de l'ARN substrat sont indiquées : L7Ae (L), aCBF5 (C), aNOP10 (N) et aGAR1 (G).

D. Etude de la formation du complexe CII' obtenu par incubation des 3 protéines LCN, du sRNA Pab91 non radiomarqué et de l'ARN substrat RNA-S radiomarqué en 5'.

(Fig. i44 A). Les premières études basées sur l'emploi de ce système ont conduit aux résultats et conclusions qui suivent :

- Les techniques de retard de migration sur gel ont permis d'identifier que les protéines L7Ae (RNP1) et aCBF5 (RNP2) se fixent directement au sRNA contrairement aux protéines aNOP10 et aGAR1 (Fig. i44 B pistes 1 à 7) (Baker *et al.*, 2005; Charpentier *et al.*, 2005). Comme prévu, L7Ae se fixe sur le motif en K-turn et la mutation de ce dernier empêche la liaison de L7Ae au sRNA (Baker *et al.*, 2005; Charpentier *et al.*, 2005). De façon intéressante, la mutation de la boîte ACA empêche la fixation de aCBF5 à l'ARN guide indiquant que aCBF5 se lie directement à la boîte ACA (Baker *et al.*, 2005; Charpentier *et al.*, 2005). Des études analogues indiquent que la boîte ACA et la poche de pseudouridylation sont 2 points d'ancrage de aCBF5 au sRNA (Baker *et al.*, 2005).

- Les protéines L7Ae et aCBF5 peuvent former un complexe de plus haut poids moléculaire sur le sRNA (Fig. i44 B piste 8, RNP3).

- B. Charpentier a montré également que le recrutement de la protéine aNOP10 nécessitait la fixation préalable de aCBF5 sur le sRNA (Fig. i44 B pistes 9 et 10, RNP4). Le complexe formé par aCBF5-aNOP10 sur le sRNA constitue l'ensemble minimal capable de recruter l'ARN substrat (CI) pour que la particule ait une activité ARN:Ψ-synthase détectable (Fig. i44 C, CN) (Charpentier *et al.*, 2005).

- Les protéines L7Ae, aCBF5, aNOP10 fixés au sRNA (Fig. i44 B piste 11, RNP5) permettent la formation d'un complexe unique en présence de l'ARN substrat (Fig. i44 B piste 12, CII) dont l'activité ARN:Ψ-synthase est grandement améliorée. La cinétique ainsi obtenue se caractérise par une augmentation régulière du taux de pseudouridines formées, et cela sans un temps de latence qui est propre à l'absence de L7Ae (Fig. i44 C, LCN) (Charpentier *et al.*, 2005).

- Enfin un complexe stable est détecté en présence des 4 protéines, du sRNA et du substrat (CIII) ce qui permet d'obtenir une activité maximale pour la sRNP H/ACA reconstituée (Fig. i44 C, LCNG) (Baker *et al.*, 2005; Charpentier *et al.*, 2005).

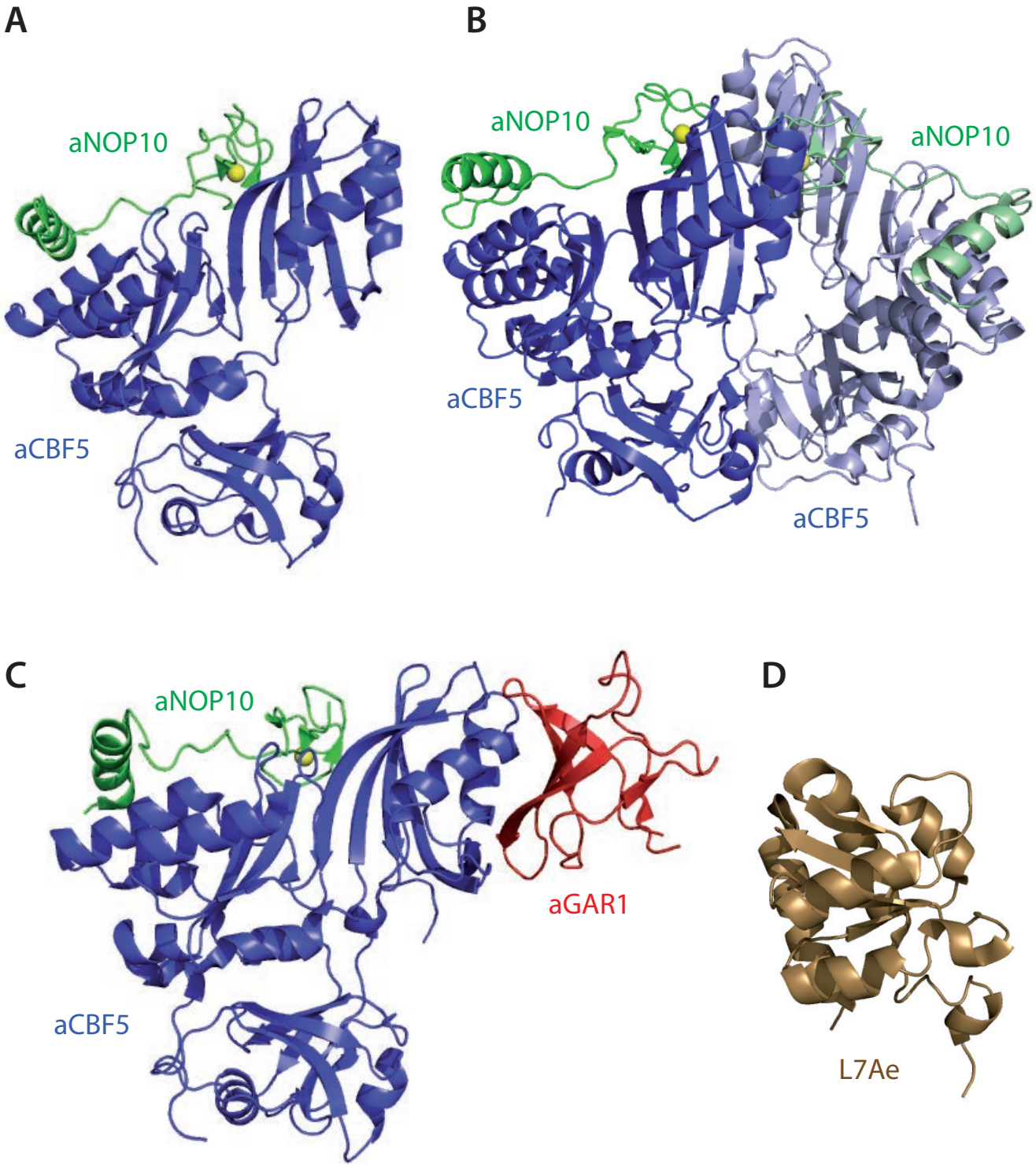


Figure i45 : Structure 3D des protéines aCBF5, aNOP10, aGAR1 et L7Ae d'archaea.
aCBF5 est représentée en bleu, aNOP10 en vert, l'atome de zinc en jaune et aGAR1 en rouge.
A. Complexe aCBF5-aNOP10, PDB=2APO (Hamma *et al.,* 2005).
B. Hétérotétramère aCBF5-aNOP10 / aNOP10-aCBF5, PDB=2AUS (Manival *et al.,* 2006).
C. Complexe aCBF5-aNOP10-aGAR1, PDB=2EY4 (Rashid *et al.,* 2006).
D. Structure de la protéine ribosomique L7Ae, PBB=1PXW (Charron *et al.,* 2004).
Représentation des structures réalisée à l'aide du logiciel Pymol à partir des coordonnées PDB indiquées.

- L'emploi de l'ARN substrat radiomarqué permet de confirmer que celui-ci est bien incorporé au sein de la particule et il est dénommé le complexe CII', qui ne se forme uniquement qu'en présence d'ARN guide et des protéines L7Ae, aCBF5 et aNOP10 (Fig. i44 D) (Charpentier *et al.*, 2005).

Deux autres équipes ont ensuite également utilisé ces systèmes de reconstitution (Hamma *et al.*, 2005; Duan *et al.*, 2009). Il est intéressant de souligner que des mesures d'activité ARN:Ψ-synthase ont été réalisées dans des conditions saturantes en substrat de « multiple turn-over » ou en conditions de « single turn-over » dans lesquelles le substrat était limitant (Duan *et al.*, 2009).

6. Vue d'ensemble de la structure 3D d'une sRNP à boîtes H/ACA d'archaea

6.1. Les protéines associées aux sRNA à boîtes H/ACA d'archaea dans les structures 3D résolues en absence d'ARN

Les protéines associées aux sRNA H/ACA d'archaea ont initialement été découvertes par des alignements de séquences (Fig. i19, i39 et i40) (Henras *et al.*, 1998; Watkins *et al.*, 1998; Watanabe et Gray, 2000). Plusieurs complexes ont été obtenus à partir des protéines seules (L7Ae, aCBF5-aNOP10 et aCBF5-aNOP10-aGAR1) ou en présence d'ARN (sRNA+LCNG, sRNA tronqué dans son hélice H2+CNG, sRNA+ARN substrat+CN, sRNA+ARN substrat+LCN et sRNA+ARN substrat+LCNG).

<u>aCBF5</u>

Les différentes structures 3D de aCBF5 en absence d'ARN révèlent que cette protéine de 334 résidus est composée de deux domaines et sa structure globale se superpose fidèlement avec celle de TruB (Fig. i45 A à C) (Hamma *et al.*, 2005; Manival *et al.*, 2006; Rashid *et al.*, 2006). Le domaine catalytique de aCBF5, situé en Nt, est composé de 250 résidus répartis sur une face comportant 8 brins β qui sont décorés par plusieurs boucles et 7 hélices α qui flanquent le site actif. Le domaine PUA,

localisé en Ct, comporte 84 résidus répartis sous la forme d'un pseudo-tonneau de 5 brins β dont chaque côté est coiffé par une hélice α. Le domaine PUA est enveloppé par une extension de 33 résidus provenant du domaine Nt de aCBF5 qui forme un élément de structure secondaire (feuillet β) permettant de stabiliser aCBF5 (Manival *et al.*, 2006).

aNOP10

La protéine aNOP10 de 60 résidus adopte une forme très allongée et possède 2 domaines reliés par une longue région charnière (linker) (Fig. i45 A à C). La région Nt contient un motif en doigt de zinc à 4 cystéines qui est globalement structuré sous la forme d'un feuillet composé de 4 brins β. La partie Ct de aNOP10 contient une hélice α (Hamma *et al.*, 2005; Manival *et al.*, 2006; Rashid *et al.*, 2006).

aGAR1

C'est une protéine de 94 résidus constituée d'un seul domaine qui est structuré sous la forme d'un tonneau exclusivement composé de 6 brins β (Fig. i45 C) (Rashid *et al.*, 2006).

L7Ae

Cette protéine constituée de 128 résidus se replie sous la forme d'un domaine globulaire qui s'organise autour d'un feuillet central contenant 4 brins β qui est entouré de 5 hélices α (Fig. i45 D) (Ban *et al.*, 2000; Charron *et al.*, 2004; Hamma et Ferre-D'Amare, 2004; Moore *et al.*, 2004). L7Ae possède une face chargée positivement et l'autre chargée négativement ce qui lui permet de se fixer au motif en K-turn avec une affinité apparente de 20 à 200 nM en fonction du type de motif en K-turn (K-turn ou K-loop) et dont la surface d'interaction avec l'ARN est de 1500 Å^2

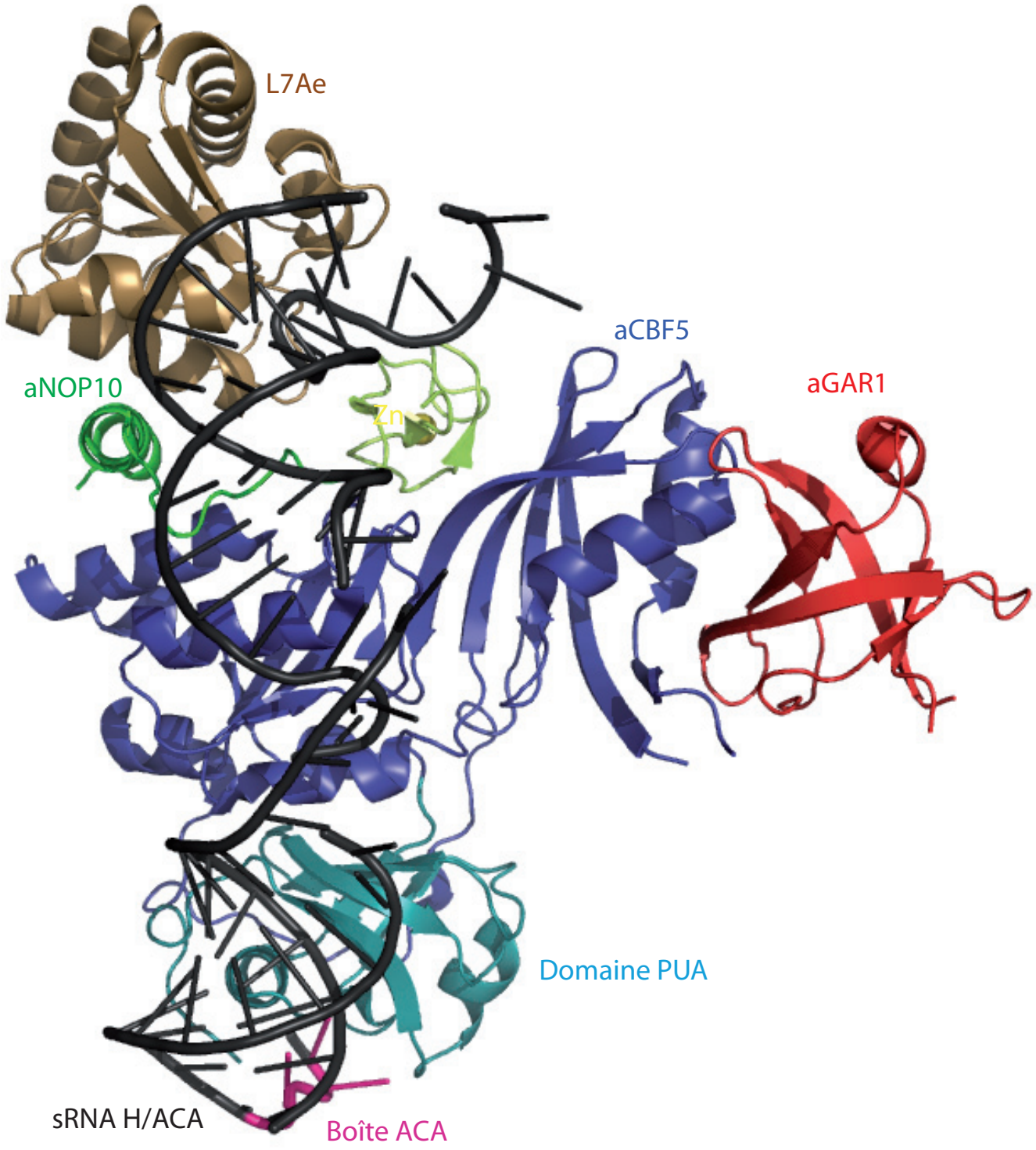


Figure i46 : Structure 3D de la sRNP H/ACA d'archaea. (Li et Ye, 2006).
Le domaine catalytique de aCBF5 est représenté en bleu, son domaine PUA en cyan, aGAR1 en rouge, le fragment Nt de aNOP10 contenant le motif en doigt de zinc en jaune, son fragment Ct en vert, L7Ae en brun, le sRNA H/ACA en noir et la boîte ACA en magenta. Réalisé à l'aide du logiciel Pymol à partir des coordonnées PDB=2HVY.

(Rozhdestvensky *et al.*, 2003; Charron *et al.*, 2004; Hamma et Ferre-D'Amare, 2004). La reconnaissance d'un motif en K-loop reposerait sur la présence du tri-peptide L87-E88-V89 chez L7Ae qui est substitué par V-S-R chez son homologue eucaryote (Snu13p/15.5kD) qui ne peut reconnaître efficacement les motifs en K-loop. C'est l'arginine qui serait le résidu de L7Ae qui discriminerait la reconnaissance de la seconde hélice présente dans le motif en K-turn (Hamma et Ferre-D'Amare, 2004).

6.2. Structure 3D de la sRNP à boîtes H/ACA

La structure complète des 4 protéines L7Ae, aCBF5, aNOP10 et aGAR1 de *P. furiosus* assemblées sur un sRNA dérivé de Afu46 mais en absence d'ARN substrat, a été résolue par radiocristallographie (Li et Ye, 2006). Les protéines de cette sRNP se superposent avec les complexes protéiques obtenus en absence d'ARN. Cette sRNP de 94 kDa s'inscrit dans un triangle équilatéral avec des côtés de 80 à 90 Å de longueur et environ 20 Å d'épaisseur (Fig. i46). Le domaine catalytique de aCBF5 se situe au centre du complexe. Son domaine PUA, L7Ae et aGAR1 constituent chacun un des sommets du triangle et aNOP10 est prise en sandwich entre L7Ae et aCBF5 (Li et Ye, 2006; Ye, 2007). Le sRNA s'étend sur 80 Å de long et il est entièrement recouvert par les protéines qui sont fixées sur une face (Li et Ye, 2006; Liang, B. *et al.*, 2007). Une organisation similaire de la sRNP est envisagée chez les snoRNP H/ACA eucaryotes (Li et Ye, 2006).

6.3. Interactions entre aCBF5 et le sRNA H/ACA

De nombreux contacts sont établis entre le domaine PUA et l'ARN (Fig. i46). Les 1er et 3e résidus des boîtes ACA et H sont universellement conservés et sont critiques pour la stabilité de l'ARN. La 1ère adénine (A58) est dans une conformation *syn* et s'insère dans une poche de la protéine formée entre le ribose du sillon mineur de l'ARN et les résidus His268, Gly269, Ala270 et Trp 337. La formation de cette poche est en partie induite par la fixation de l'ARN (Li et Ye, 2006). La 3e adénine

(A60), également en conformation *syn*, s'empile au-dessus de l'hélice α8 de aCBF5 et recouvre partiellement le ribose du résidu C59. Le nucléotide A60 forme deux liaisons hydrogènes avec le squelette de aCBF5, imposant ainsi une spécificité à la séquence. Au contraire, le résidu C59 adopte une conformation ouverte suggérant une spécificité moindre et cette conformation peut être modulée par d'autres interactions (Li et Ye, 2006). La fixation de aCBF5 du haut de l'hélice 1 et du bas de l'hélice 2 du sRNA place la poche de pseudouridylation à proximité du sillon catalytique de aCBF5 (Fig. i46) (Li et Ye, 2006).

Chez la levure, des expériences de retard de migration sur gel et d'empreintes sur gel ont montré que Cbf5p se fixait sur l'une ou l'autre des tige-boucles d'un snoRNA et que les boîtes H et ACA n'étaient pas strictement nécessaires à l'assemblage *in vitro* de Cbf5p mais que chacune contribuait à augmenter l'affinité de Cbf5p au snoRNA (Normand *et al.*, 2006). Chez les archaea, notre équipe a contribué à montrer l'interaction stable formée entre aCBF5 et l'ARN (Baker *et al.*, 2005; Charpentier *et al.*, 2005; Youssef *et al.*, 2007; Baker *et al.*, 2008) et que celle-ci reposait fortement sur le domaine PUA de aCBF5 (Manival *et al.*, 2006). Les données de nos travaux pour cette question seront détaillées dans les parties I.B et II.A des résultats.

6.4. Interaction entre aCBF5 et aNOP10

Sous une forme libre, aNOP10 adopte une structure hautement dynamique (Reichow et Varani, 2008) alors qu'en présence de aCBF5, la conformation de aNOP10 est stabilisée par de nombreuses interactions avec des résidus conservés du domaine catalytique de aCBF5 avec une surface d'interaction de 2800 Å² (Fig. i45 et i46) (Hamma *et al.*, 2005; Manival *et al.*, 2006; Rashid *et al.*, 2006). En particulier, la région charnière de aNOP10 initialement peu structurée, va s'organiser lors de l'interaction avec aCBF5 et positionnerait certains résidus de aCBF5 présents à proximité du résidu catalytique. Plusieurs réseaux d'interactions entre aNOP10 et aCBF5 ont ainsi été proposés (Hamma *et al.*, 2005; Rashid *et al.*, 2006). Nous avons contribué à la mise en évidence expérimentale des acides aminés de aNOP10

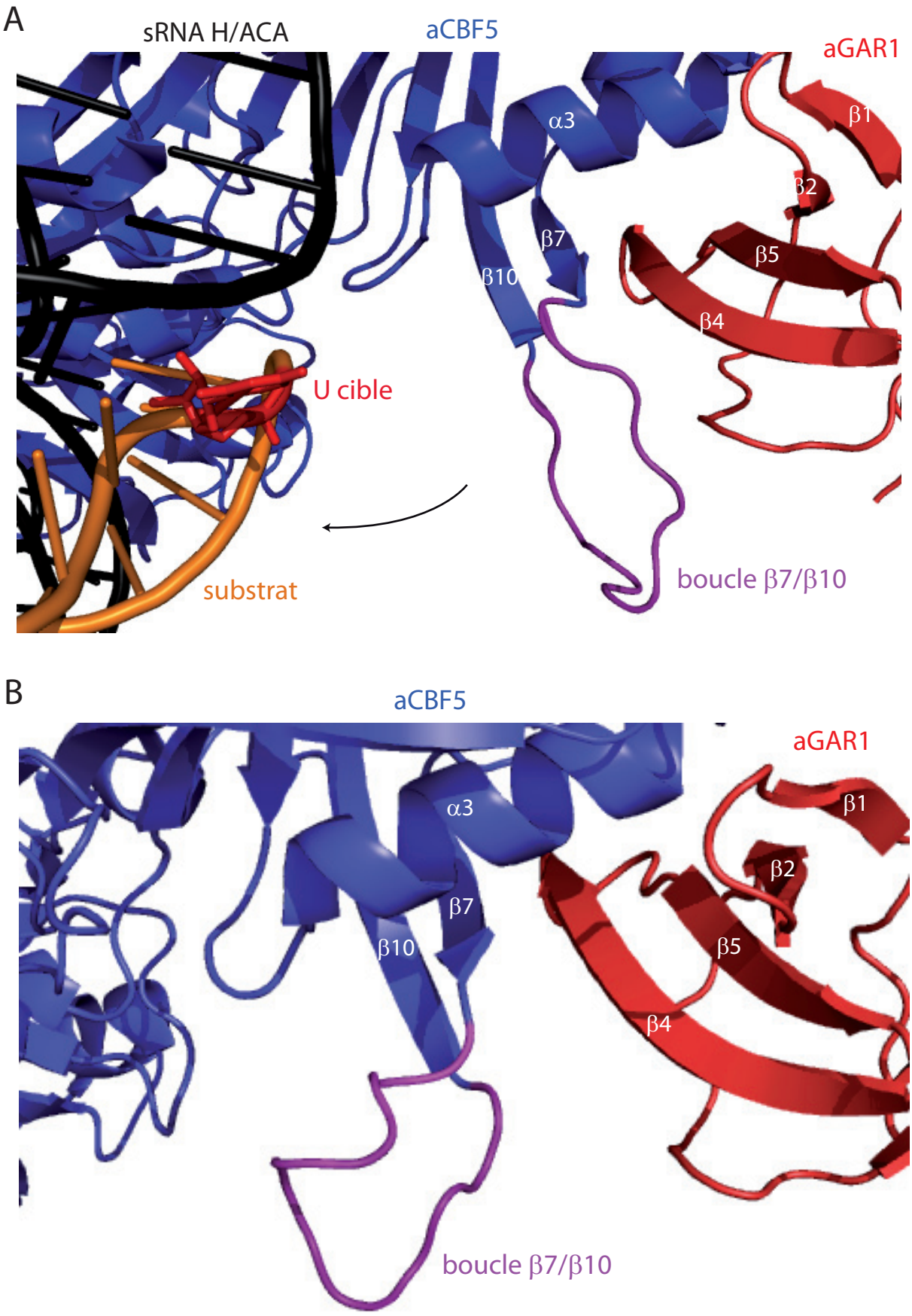


Figure i47 : La boucle β7/β10 de aCBF5 s'oriente suivant deux conformations : l'une ouverte ou l'autre fermée.
aCBF5 est représentée en bleu, sa boucle β7/β10 en violet, aGAR1 en rouge, le sRNA H/ACA en noir et le substrat en orange.
A. Conformation ouverte de la boucle β7/β10 de aCBF5, PDB=2RFK (Liang *et al.,* 2007).
B. Conformation fermée de la boucle β7/β10 de aCBF5, PDB=2EY4 (Rashid *et al.,* 2006).
Réalisé à l'aide du logiciel Pymol à partir des coordonnées PDB indiquées.

pouvant influencer l'activité de aCBF5 ainsi que des rôles fonctionnels joués par les domaines Nt et Ct de aNOP10. Ces informations seront détaillées dans la partie II.B des résultats.

6.5. Interactions entre aCBF5 et aGAR1

Deux boucles formées par les brins β1/β2 et β4/β5 de aGAR1 lui permettent d'interagir fortement avec la face de aCBF5 constituée par les hélices α3, α5 et le brin β7 avec une surface d'interaction de l'ordre de 1500 Å² (Li et Ye, 2006; Rashid *et al.*, 2006; Liang, B. *et al.*, 2007). Bien que aCBF5 ne possède pas les brins β8/β9, qui chez TruB sont responsables de l'accrochage de la boucle TΨC de l'ARNt, la boucle formée par les brins β7/β10 de aCBF5 est bien conservée et suffisamment longue pour interagir avec l'ARN cible (Rashid *et al.*, 2006). L'analyse des structures 3D révèle que la boucle β7/β10 de aCBF5 s'organise sous 2 états conformationnels, une conformation dite « ouverte » qui permettrait l'entrée de l'ARN substrat dans le site actif de aCBF5 (Li et Ye, 2006; Liang, B. *et al.*, 2007; Ye, 2007) et une conformation dite « fermée » qui stabiliserait l'appariement entre le sRNA et sa cible (Fig. i47) (Rashid *et al.*, 2006; Ye, 2007). Il est intéressant de souligner que la boucle β7/β10 de aCBF5 est désorganisée dans la plupart des structures 3D (Hamma *et al.*, 2005; Li et Ye, 2006; Manival *et al.*, 2006). Nos travaux concernant cette partie seront détaillés dans la partie II.C des résultats.

6.6. Interactions entre aNOP10, L7Ae et le sRNA

En absence des autres composants des sRNP, les protéines L7Ae et aNOP10 n'interagissent pas entre-elles (Baker *et al.*, 2005; Charpentier *et al.*, 2005). Par contre au sein de la sRNP, des interactions sont observées entre ces 2 protéines (Li et Ye, 2006). Il a également été montré par retard de migration sur gel que la protéine aNOP10 n'interagissait pas directement avec un sRNA (Baker *et al.*, 2005; Charpentier *et al.*, 2005; Hamma *et al.*, 2005), mais que dans la structure 3D, des

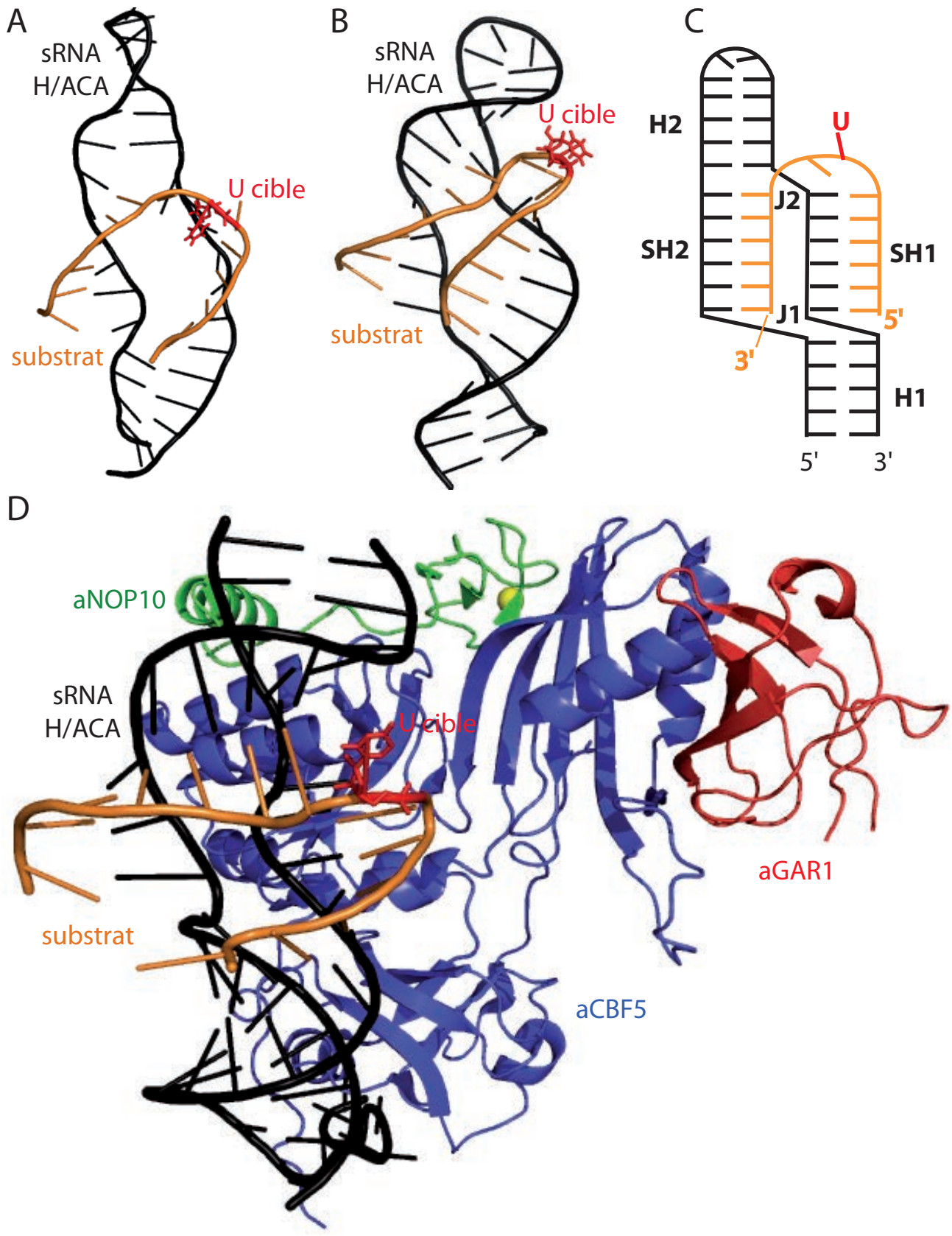


Figure i48 : Structures 3D de complexes H/ACA contenant un ARN substrat.
aCBF5 est représentée en bleu, aNOP10 en vert, aGAR1 en rouge, le sRNA H/ACA en noir et le substrat en orange.
A. et **B.** Structure 3D du complexe entre un sRNA et son substrat, PDB=2PCW (Jin *et al.*, 2007) (**A**), PBB=2P89 (Wu et Feigon, 2007) (**B**).
C. Représentation schématique d'un sRNA apparié à son ARN substrat.
D. Structure 3D d'une sRNP consistant en un sRNA tronqué de sa partie apicale, de l'ARN substrat et des protéines aCBF5, aNOP10 et aGAR1, PBD=2RFK (Liang *et al.*, 2007).
Représentations réalisées à l'aide du logiciel Pymol à partir des coordonnées PDB disponibles.

interactions sont observées entre aNOP10 et l'hélice 2 du sRNA (Fig. i46) (Li et Ye, 2006). Nous avons contribué à mettre en évidence que aNOP10 interagissait avec L7Ae et le sRNA et que ces interactions sont importantes pour la conformation de la particule. Ces informations seront détaillées dans la partie II.D des résultats.

6.7. Positionnement du substrat dans la sRNP

L'ARN guide ne remplit pas seulement la fonction de recrutement, par le biais d'une complémentarité de séquence, d'un ARN substrat, mais il contribue aussi à positionner ce substrat dans le site actif de aCBF5. En absence de substrat, la poche de pseudouridylation du sRNA est partiellement appariée pour former une hélice irrégulière (Khanna *et al.*, 2006) ou non structurée (Li et Ye, 2006). Plusieurs structures 3D d'un complexe sRNA-ARN substrat ont été déterminées par RMN ou par radiocristallographie et elles indiquent que le substrat se fixe sur une seule face du sRNA, celle qui n'est pas occupée par les protéines (Fig. i48) (Jin *et al.*, 2007; Liang, B. *et al.*, 2007; Wu et Feigon, 2007). Très récemment, les structures de sRNP comprenant le substrat en présence des jeux de protéines CN, LCN et LCNG ont été publiées et seront détaillées et discutées dans la partie des résultats (Fig. i49) (Duan *et al.*, 2009; Liang *et al.*, 2009).

L'hélice résultant de l'appariement entre la partie 5' de l'ARN substrat et la partie 3' de la poche de pseudouridylation du sRNA est appelée SH1 (Fig. i48 C). Celle-ci s'empile de façon coaxiale avec l'hélice H1 basale du sRNA. L'hélice résultant de l'appariement de l'élément 3' du substrat et de la partie en 5' de la poche de pseudouridylation du sRNA est appelée SH2 (Fig. i48 C). Celle-ci s'empile de façon co-axiale avec l'hélice H2 apicale du sRNA (Jin *et al.*, 2007; Liang, B. *et al.*, 2007; Wu et Feigon, 2007). On peut donc distinguer deux hélices discontinues qui imposent une courbure de 120° à 180° à l'ARN substrat. L'ARN substrat adopte donc respectivement une forme en « omega, Ω », « V » ou « U » (Jin *et al.*, 2007; Liang, B. *et al.*, 2007; Wu et Feigon, 2007). L'existence de différentes formes empruntées par l'ARN substrat reflète une certaine variabilité dans la structure des ARN qui doit être

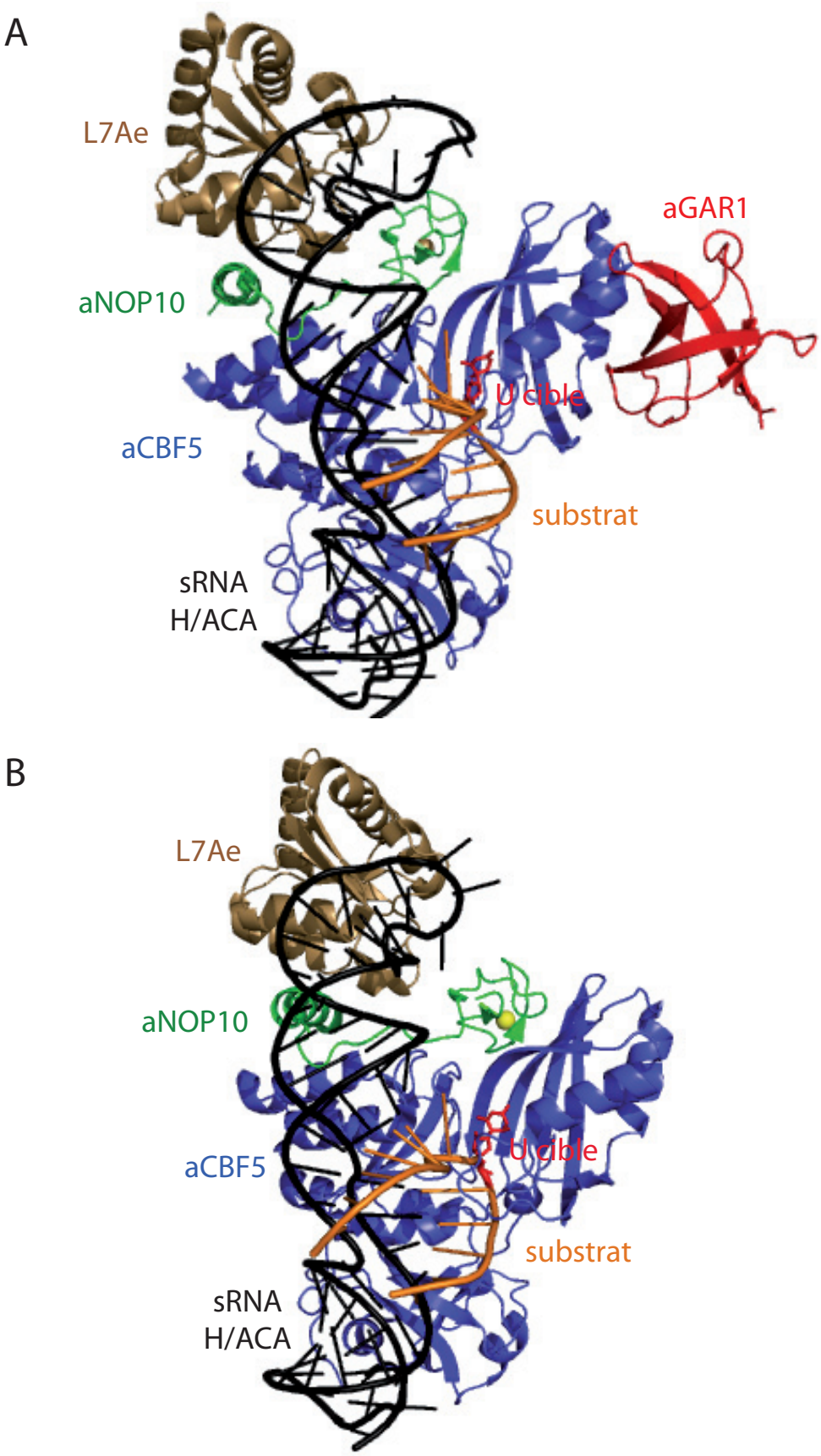


Figure i49 : Structures 3D de complexes H/ACA contenant un ARN substrat.
L7Ae est représentée en brun, aCBF5 bleu, aNOP10 en vert, aGAR1 en rouge, le sRNA H/ACA en noir, le substrat en orange et son résidu U cible en rouge.
A. Structure 3D du complexe formé entre un sRNA, un ARN substrat et les protéines L7Ae, aCBF5, aNOP10 et aGAR1, PBD=3HAY (Duan *et al.,* 2009).
B. Structure 3D du complexe formé entre un sRNA, un ARN substrat et les protéines L7Ae, aCBF5, aNOP10, PBD=3HJW (Jiang *et al.,* 2009).
Représentations réalisées à l'aide du logiciel Pymol à partir des coordonnées PDB disponibles.

accommodée par la particule (pour revue, (Li, 2008)). L'appariement sRNA-ARN substrat conduit à la formation de deux jonctions triple-hélices caractéristiques (J1 et J2) qui positionnent l'uridine vers l'extérieur comme cela a été observé lors du mécanisme d'extrusion de bases par les ARN:Ψ-synthases (Fig. i48 C) (Wu et Feigon, 2007). Il est à noter également que la situation au niveau de la jonction J2 mime la conformation existant au niveau de la tige-boucle TΨC d'un ARNt en interaction avec TruB. Au final les 4 protéines ne seraient requises ni pour l'appariement de l'ARN substrat au sein du sRNA, ni pour l'extrusion de l'uridine. En outre, il a été montré qu'un ARN substrat pouvait s'apparier en absence de protéines avec un sRNA avec un assez faible Kd≈ 200-300 µM (Wu et Feigon, 2007) ou plus élevé Kd≈1 µM (Jin *et al.*, 2007). Cela indique néanmoins que les protéines participent activement au recrutement du substrat dans la sRNP.

Dans la structure 3D d'une RNP obtenue en présence des protéines aCBF5, aNOP10 et aGAR1, d'un ARN substrat et d'un sRNA tronqué de son hélice apicale incluant le motif en K-turn, le résidu U cible est positionné à 11 Å du site actif de aCBF5, ce qui ne peut conduire à sa modification (Fig. i48 D) (Liang, B. *et al.*, 2007). Ces données structurales corroborent les observations du rôle crucial joué par L7Ae et son site de fixation pour l'activité des particules (Charpentier *et al.*, 2005).

Lors de l'association de la particule avec le substrat, aCBF5 établit de nouveaux contacts avec la poche de pseudouridylation du sRNA qui devient ordonnée. Les interactions entre aCBF5 et l'hélice SH2 sont particulièrement nombreuses, contrairement à l'hélice SH1 qui n'interagit pas avec les protéines (Liang, B. *et al.*, 2007). Les longs substrats, parfois jusqu'à plusieurs milliers de nucléotides comme les séquences des ARNr, peuvent se fixer par la face dépourvue de protéine et former les hélices SH1 et SH2 avec le sRNA, sans la nécessité d'ARN hélicases (pour revue, (Li, 2008)). Après modification, la snoRNP se décroche de l'ARN modifié par un mécanisme encore mal compris mais ne mettant à contribution aucune protéine (voir la partie II.C des résultats). Chez les eucaryotes, seuls les snoRNA essentiels dans la maturation du précurseur des ARNr nécessitent une

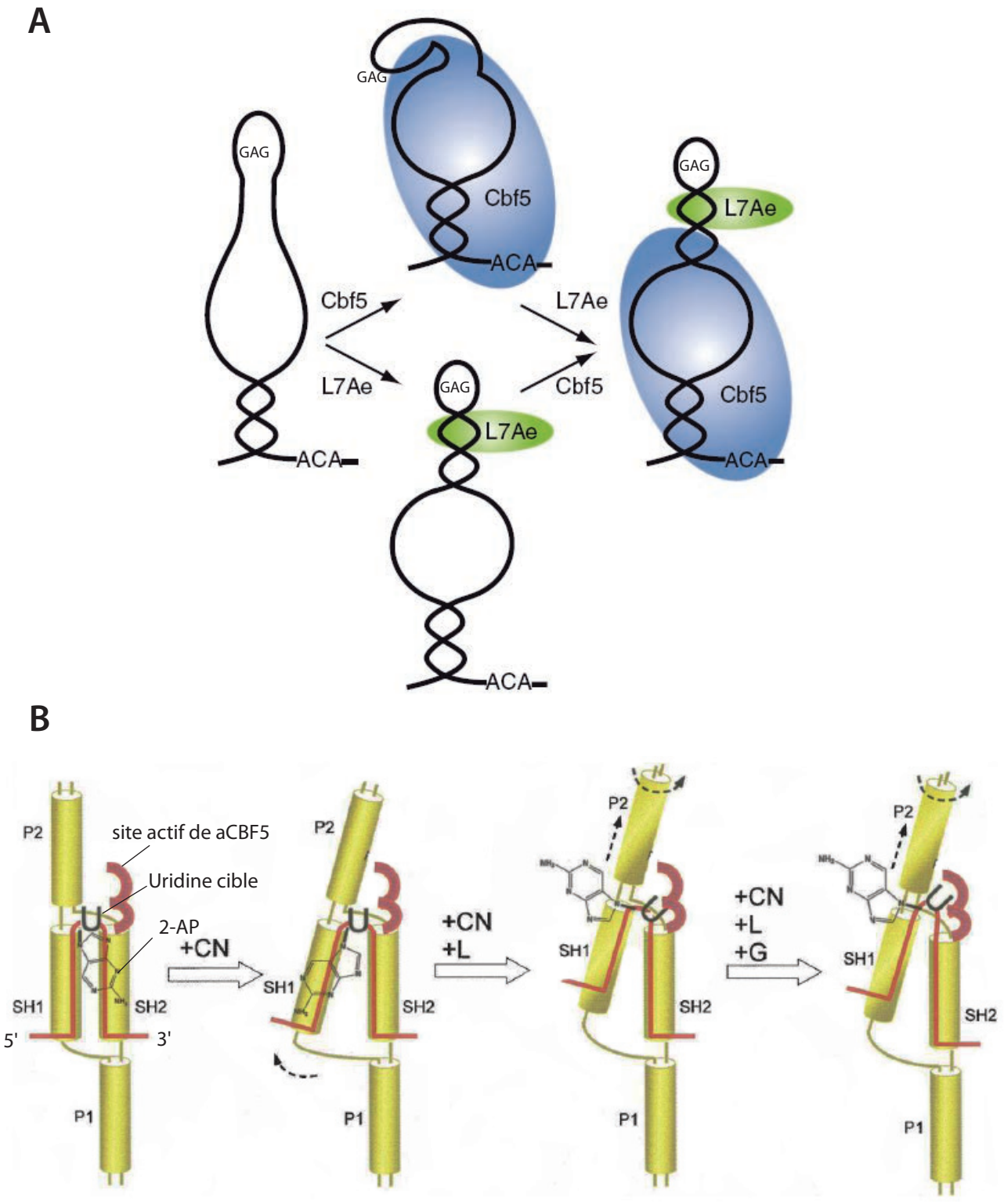


Figure i50 : Modèles d'assemblage d'une sRNP H/ACA.
A. L'hélice H2 du sRNA n'est pas stabilisée en absence de L7Ae. aCBF5 interagit avec la boîte ACA, l'hélice H1, la poche de pseudouridylation et un site unique localisé dans la partie 5' de la boucle apicale (séquence GAG). La fixation de L7Ae au K-turn induit la formation de l'hélice H2 et remodèle la poche de pseudouridylation. L7Ae se fixe sur le K-turn et chasse aCBF5 de cette zone 5' apicale. (Youssef *et al.,* 2007).
B. Différentes conformations de l'ARN substrat (en rouge) en fonction de la composition en protéines de la sRNP. Le résidu U cible est indiqué ainsi que le résidu 2-AP. Les cylindres jaunes représentent les hélices du sRNA (P1 et P2) et du duplexe sRNA-ARN substrat (SH1 et SH2). Détail dans le texte. (Liang *et al.,* 2008).

activité ARN hélicase pour leur libération de l'ARN, comme par exemple l'ARN hélicase Rok1 (Rescuer Of *kem1*) dans la cas de la snoRNP snR30 (Bohnsack *et al.*, 2008).

6.8. Changement de conformation lié à l'assemblage d'une sRNP H/ACA

6.8.1. Remodelage de la poche de pseudouridylation par L7Ae

L'équipe de M. Terns a proposé que aCBF5 reconnaisse une séquence 5'-GAG-3' située dans la partie 5' de la boucle apicale des sRNA et qui pourrait avoir un rôle similaire à celui de la boîte CAB des scaRNA eucaryotes (Fig. i50 A) (Baker *et al.*, 2005). Cette interaction représenterait un point d'ancrage à l'ARN supplémentaire à l'interaction entre le domaine PUA et la boîte ACA (Li et Ye, 2006; Baker *et al.*, 2008), permettant ainsi de stabiliser le complexe initial formé entre aCBF5 et le sRNA (Baker *et al.*, 2008). En effet, le taux de complexe aCBF5-sRNA est plus faible lorsque la séquence GAG est mutée (Baker *et al.*, 2005). Un modèle a été proposé selon lequel L7Ae occuperait la boucle apicale et chasserait aCBF5 de ce site par compétition (Fig. i50 A) (Youssef *et al.*, 2007). Par contre, L7Ae jouerait un rôle significatif dans la structuration de l'hélice H2 du sRNA (Youssef *et al.*, 2007) et dans la fixation du substrat en modelant la poche de pseudouridylation (Liang, B. *et al.*, 2007; Youssef *et al.*, 2007). Nos données concernant cette reconnaissance potentielle de la partie apicale du sRNA par aCBF5 et de l'effet à distance de la fixation de L7Ae seront présentées dans la partie I.B et II.D des résultats.

6.8.2. Positionnement du substrat au sein du site actif de aCBF5

L'emploi d'un ARN substrat contenant le fluorophore 2-aminopurine (2-AP) à la place du résidu non apparié directement en 3' du U cible de la réaction, remplacé lui-même par un groupement 5-fluorouridine (5-FU) afin de bloquer le déroulement de la réaction de pseudouridylation, permet de suivre les changements de conformation du substrat au sein de différentes sRNP H/ACA partiellement

assemblées, représentant des sous-complexes sRNP (Fig. i50 B) (Liang, B. *et al.*, 2007; Liang *et al.*, 2008). L'intensité de fluorescence du 2-AP est en lien direct avec son état conformationnel dans l'ARN. En effet, le résidu 2-AP émet une fluorescence maximale lorsqu'il est empilé à d'autres résidus de l'ARN, l'émission de fluorescence devient très faible alors que ce résidu est par exemple expulsé en dehors du squelette ribose-phosphate. Les changements de fluorescence observés par comparaison de sous-complexes sRNP en interaction avec l'ARN substrat traduisent l'existence d'une conformation différente de l'ARN substrat au sein de ces sous-complexes. Ainsi, tout changement de positionnement du résidu 5-FU contigu peut être détecté et suivi par les propriétés de fluorescence du résidu 2-AP. A partir des résultats obtenus dans les études employant cette stratégie et les données structurales des sous-complexes sRNP H/ACA, l'équipe de H. Li a proposé que L7Ae permettait, en orientant l'hélice H2 vers le site actif de aCBF5, un positionnement du substrat pouvant conduire à sa modification (Fig. i50 B) (Liang, B. *et al.*, 2007; Liang *et al.*, 2008). La protéine aNOP10 stabiliserait l'ARN H/ACA en se fixant sur l'hélice H2 du sRNA et donc par un effet indirect positionnerait le substrat dans la poche de pseudouridylation. La protéine aGAR1 interviendrait dans le placement de l'ARN substrat par son interaction avec la boucle β7/β10 de aCBF5 (Fig. i50 B) (Liang *et al.*, 2008).(Liang *et al.*, 2008) Il est important de souligner que dans ces études, le sRNA et le substrat sont d'abord pré-incubés entre-eux, ce qui peut alors donner lieu à la formation d'un complexe sRNA-ARN substrat, avant que les protéines ne soient ajoutées.

7. La télomérase : une RNP indispensable à la biologie de la cellule

7.1. Les télomères et la réplication des chromosomes linéaires

Chez tous les eucaryotes, les télomères sont composés de centaines voire de milliers de nucléotides répétés et placés en tandem dans la région terminale de chaque chromosome. Chez l'homme, l'unité fondamentale de ces éléments répétés est la séquence TTAGGG qui présente entre 5000 et 12000 copies (pour revue, (Rufer et Nabholz, 2003)).

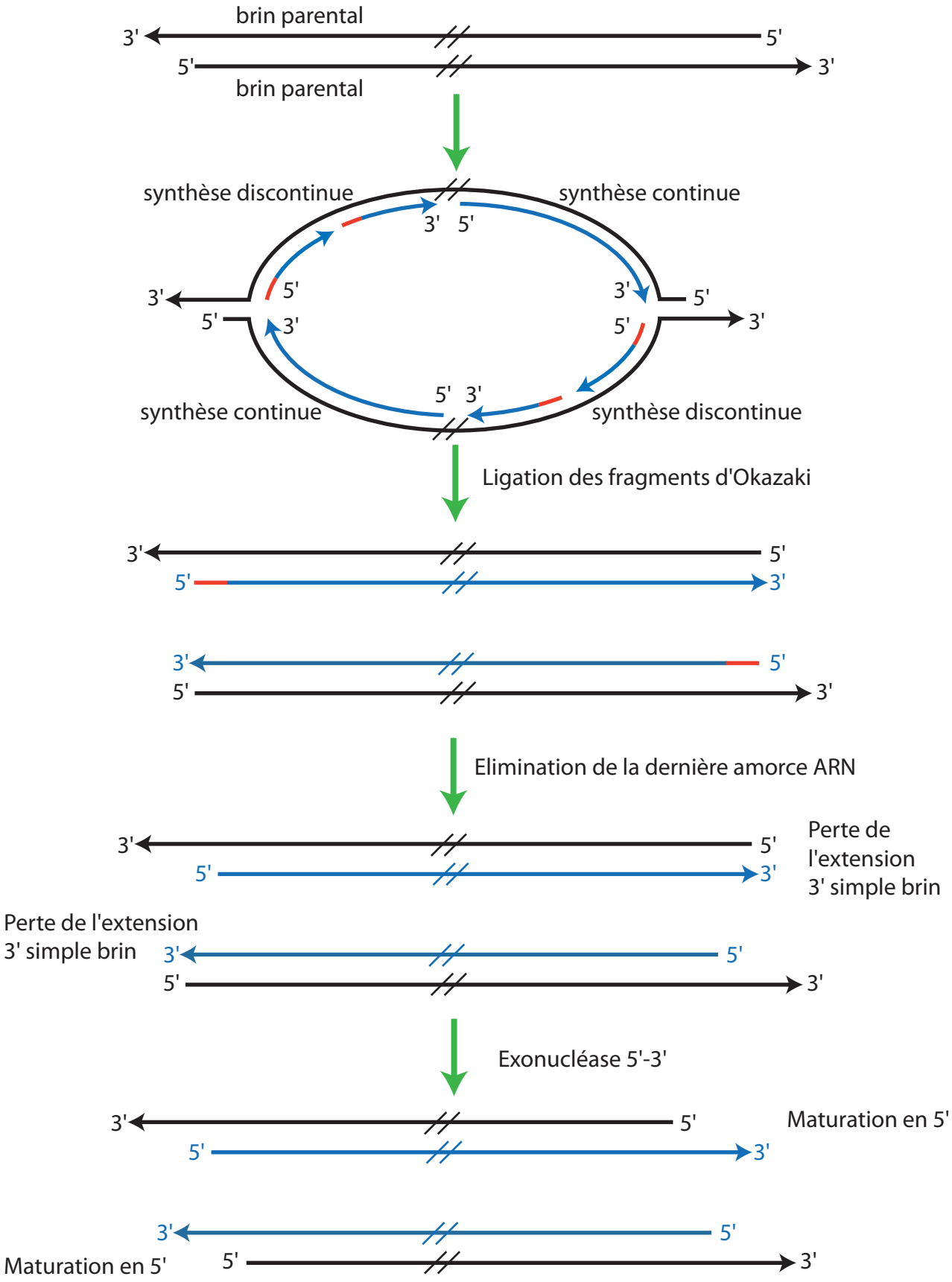


Figure i51 : La réplication des chromosomes linéaires conduit au raccourcissement des brins d'ADN. (Adapté de LeBel et Wellinger, 2004).
L'ADN parental (en noir) est répliqué (en bleu) selon un mécanisme semi-conservatif qui ne permet pas de régénérer une extension 3' simple brin à l'extrémité du brin synthétisé de façon continue. En revanche, à l'extrémité du brin synthétisé de façon discontinue, l'élimination de la dernière amorce ARN (en rouge) produit un brin raccourci en 5'. Un mécanisme de maturation en 5' par une exonucléase 5'-3' raccourcit les brins parentaux laissant ainsi des extrémités 3' protubérantes mais aussi une taille raccourcie des brins d'ADN qui s'amplifie à chaque cycle de réplication.

Les télomères servent à limiter l'effet de la perte d'un fragment de séquence chromosomique qui a lieu à chaque étape de la division cellulaire, ceci du fait de la synthèse par fragments d'un des deux brins de l'ADN lors de sa réplication (Fig. i51). La synthèse du brin retardé implique en effet de petits fragments d'ARN servant d'amorce à l'ADN polymérase pour former les fragments d'Okazaki. La séquence ADN 3'-terminale du brin d'ADN en cours de réplication, qui est complémentaire au fragment d'Okazaki, est perdue à chaque réplication du fait d'une activité nucléase, d'où une perte de longueur des télomères à chaque division cellulaire. Ce phénomène, associé aux fragments d'Okazaki, entraîne un raccourcissement qui, pour les télomères humains, est de 50 à 200 pb par division cellulaire (Fig. i51) (pour revue, (Kahn, 2005)).

Pour éviter de perdre des séquences de gènes importants pour la cellule ou l'organisme, un enchaînement de séquences répétées est donc ajouté à chaque extrémité des chromosomes dans les cellules germinales. La perte de ces séquences répétées limite l'effet négatif que pourrait avoir la perte de séquences codantes. L'enchaînement des séquences répétées des télomères n'est pas catalysé par les ADN polymérases classiques, mais repose sur l'activité rétrotranscriptase de la télomérase (voir IV.C.7.3). Dans un grand nombre de cellules somatiques l'activité télomérase est peu ou pas présente. De ce fait la longueur des télomères est réduite au cours des divisions cellulaires. Seul un certain nombre de cellules somatiques à multiplication rapide comme les cellules épithéliales présentent une activité télomérase (pour revue, (Gilson et Geli, 2007)).

Même si la présence des séquences télomériques permet de compenser la perte de longueur des chromosomes pendant un grand nombre de divisions cellulaires, la longueur de l'ADN télomérique ne doit pas être trop réduite afin d'éviter les phénomènes de recombinaisons entre chromosomes. De ce fait, un système de recensement de la longueur des chromosomes existe dans les cellules eucaryotes et à partir d'une certaine réduction en taille des télomères, les cellules perdent leur capacité à se diviser. C'est ce qui se passe dans le vieillissement tissulaire ; d'où l'importance pour tout tissu épithélial, en particulier, les épithélia pulmonaires, des muqueuses des

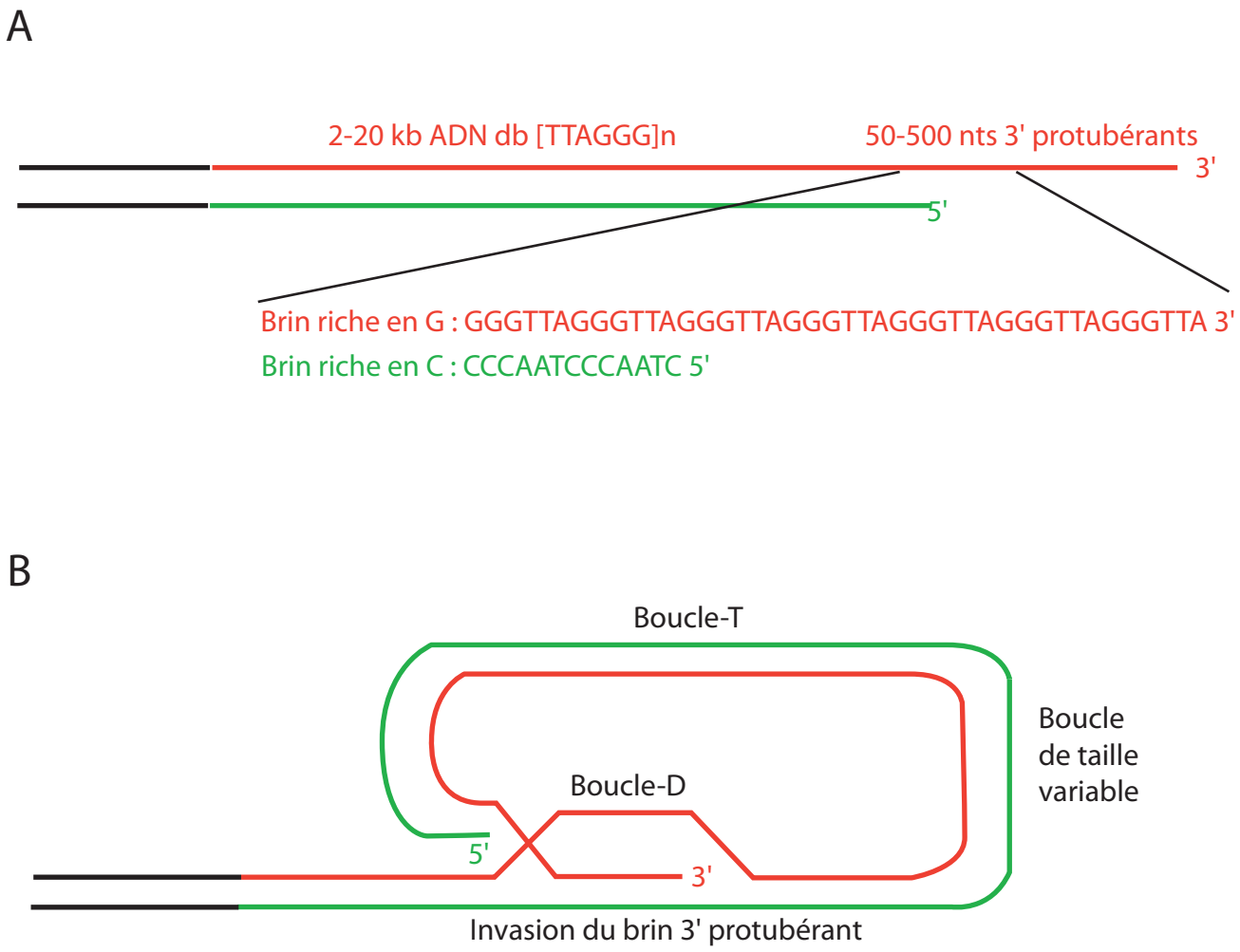


Figure i52 : Modèles de structure des télomères humains. (Adapté de Palm et De Lange, 2008).
A. Modèle classique d'une extrémité chromosomique avec une extension en 3' de 50 à 500 nucléotides sur le brin riche en G (en rouge). Le brin complémentaire est riche en C (en vert).
B. Modèle schématique de la structure en boucle-T, obtenue par le repliement du double brin sur lui-même. L'extension 3' protubérante simple brin s'insère dans la portion double brin et engendre la boucle-D.

appareils digestif et respiratoire et les épithélia cardiaques et des vaisseaux sanguins (cellules endothéliales) de préserver une activité télomérase (pour revue, (Gilson et Geli, 2007)).

En sens inverse, une cellule qui a perdu le contrôle de sa division cellulaire par acquisition d'un oncogène, ou perte d'un proto-oncogène, continue à se diviser même lorsque la longueur des télomères devient trop courte, conduisant à l'apparition de remaniements chromosomiques et généralement aboutissant à l'apoptose d'une majorité des cellules ayant acquis ces propriétés anormales de division cellulaire. Seules des cellules qui ont été capables d'activer efficacement leur production de télomérase deviennent quasiment immortelles. Cette étape de réactivation de la télomérase est donc à la base du développement tumoral.

7.2. Structure de l'extrémité des télomères

Des protéines sont associées aux télomères, forment des structures qui protègent les extrémités des chromosomes. Il s'agit des protéines constituant « le capuchon télomérique » appelé aussi le complexe shelterine (pour revue, (de Lange, 2005)). Dans un modèle «linéaire» de l'extrémité des chromosomes, le capuchon télomérique est formé de l'extension d'ADN simple brin et de protéines spécifiques qui lui sont associées (Fig. i52 A). Cependant, en 1999, ce modèle a été remis en cause. Pour la première fois, des observations en microscopie électronique ont montré que les extrémités des chromosomes de mammifères sont en fait repliées sur elles-mêmes, formant une énorme boucle terminale, appelée boucle T (Telomere loop) (Griffith *et al.*, 1999). Ces boucles sont créées par une insertion de l'extrémité 3' simple brin dans la portion double brin du télomère (Fig. i52 B). Il y a ainsi un déplacement d'une partie de l'ADN télomérique double brin, qui devient simple brin à l'intérieur de la boucle même, formant une boucle D (Displacement loop) (Griffith *et al.*, 1999). La formation de ces structures est possible par la présence, en 3' du brin riche en G (synthèse d'ADN discontinue lors de la réplication), d'extensions terminales simple brin relativement longues chez les mammifères et par l'association

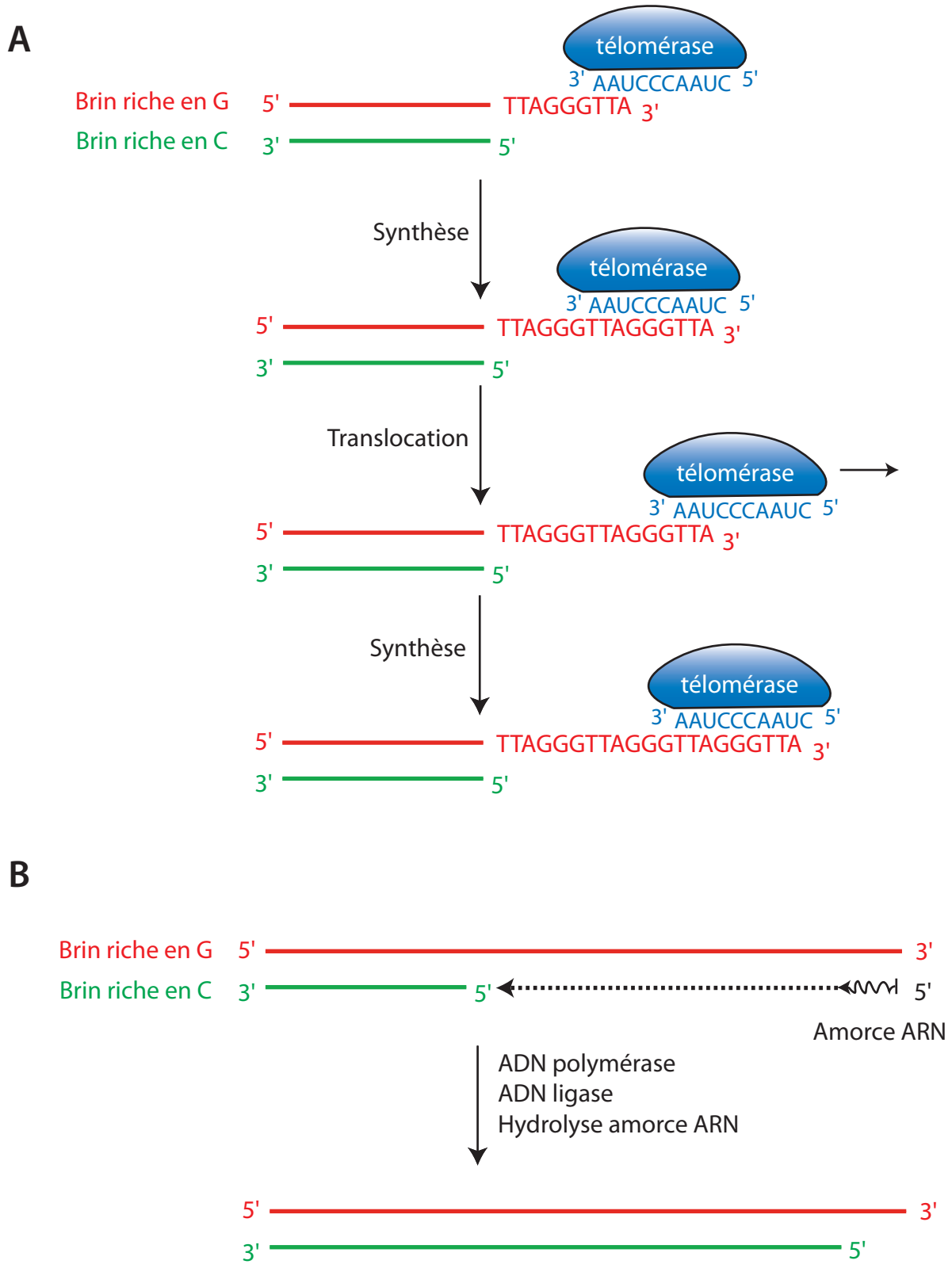


Figure i53 : Modèle proposé pour la synthèse des télomères. (D'après Vermeesch et Price, 1994).
A. Synthèse des télomères à séquence riche en G à l'extrémité 3' de l'ADN, par la translocation de la télomérase (synthèse discontinue de ce brin).
B. Modèle simplifié pour la synthèse des télomères à séquence riche en C à l'extrémité 5' (synthèse continue de ce brin).

avec certaines protéines spécifiques des télomères. Ces structures terminales sont donc différentes des extrémités chromosomiques issues des cassures double-brin survenant à d'autres endroits dans l'ADN. Ainsi la structure particulière des télomères n'est pas reconnue par la machinerie de réparation (pour revue, (LeBel et Wellinger, 2004)).

7.3. Mécanisme de synthèse des télomères par la télomérase

Le complexe de la télomérase synthétise l'extrémité des chromosomes selon le mécanisme présenté schématiquement en figure i53. L'ARN de la télomérase possède une séquence matrice qui est partiellement complémentaire à celle des répétitions de l'extrémité du brin riche en G des chromosomes (Fig. i53A). Ces deux séquences s'hybrident sur quelques nucléotides et par transcription réverse, la télomérase ajoute à l'extrémité de l'ADN, la séquence complémentaire à celle présente sur son ARN matrice. Le duplexe ADN-ARN subit ensuite une translocation par un mécanisme de glissement conduisant à une nouvelle hybridation et à un nouveau cycle de synthèse (Vermeesch et Price, 1994). La synthèse du brin riche en C est effectuée de façon conventionnelle par une ADN polymérase sans l'intervention de la télomérase (Fig. i53B).

7.4. Le complexe shelterine

La shelterine est un complexe multiprotéique qui s'associe spécifiquement avec les séquences télomériques, protégeant ainsi l'extrémité des chromosomes des dommages de l'ADN, et régulant la maintenance des télomères (pour revue, (de Lange, 2005)). Chez l'homme, le complexe shelterine comporte 6 protéines TRF1, TRF2, POT1, TIN2, TPP1 et RAP1 (Fig. i54 A). De plus, les télomères comportent un grand nombre de protéines qui, contrairement au complexe shelterine, ont aussi des fonctions non-télomériques. La spécificité du complexe shelterine pour l'ADN télomérique provient de la reconnaissance des répétitions TTAGGG par 3 de ses

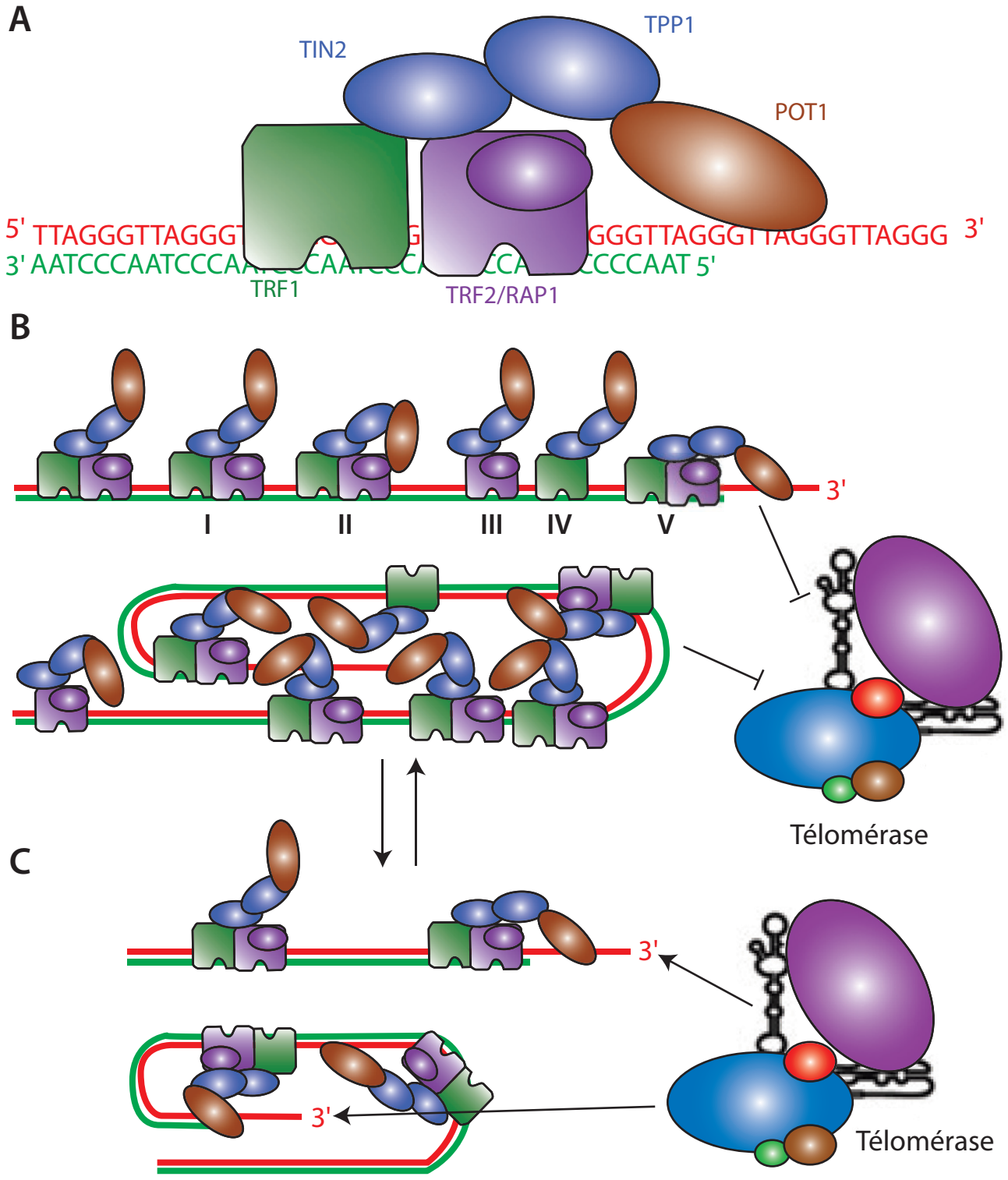


Figure i54 : Le complexe shelterine protège l'extrémité des chromosomes et régule l'action de la télomérase. (Adapté de De Lange, 2005).

A. Schéma de la shelterine sur l'ADN télomérique. Le brin riche en G est en rouge, le brin riche en C est en vert. La protéine POT1 peut fixer l'extrémité 3' simple brin. Les protéines TRF1 et 2 fixent l'ADN db.

B. Complexes et sous-complexes shelterine potentiels formés sur l'ADN télomérique. (**I**) Les 6 protéines avec POT1 non fixée à l'ADN sb. (**II**) Comme en (I) avec POT1 qui interagit avec TRF2. (**III**) Complexe TRF2/RAP1/TIN2/TPP1/POT1. (**IV**) Complexe TRF1/TIN2/TPP1/POT1. (**V**) Les 6 sous-unités du complexe shelterine avec POT1 fixée à l'ADN sb. De longs télomères recrutent plus de comlexes shelterine et la télomérase est bloquée.

C. Des télomères courts recrutent peu de complexes shelterine et la télomérase n'est plus bloquée.

composants : les protéines TRF1 et 2 (Telomeric Repeat binding Factor 1 et 2) fixent la partie des télomères (ADN db) sous forme de duplexe, tandis que la protéine POT1 (Protection Of Telomeres 1) peut se lier aux répétitions simple brin TTAGGG présentes dans la boucle-D de l'extrémité 3' protubérante. TRF1 et TRF2 recrutent les 4 autres membres du complexe shelterine : TIN2 (TRF2-and TRF1-Interacting Nuclear protein 2), RAP1 (Repressor/Activator Protein 1), TPP1 (autrefois connu sous TINT1, PTOP ou PIP1) et POT1. Le complexe shelterine permet ainsi à la cellule de distinguer l'extrémité naturelle des ses chromosomes, de cassures ou de réparations dans l'ADN et régule la maintenance des télomères en inhibant la télomérase (pour revues, (Bianchi et Shore, 2008; Palm et de Lange, 2008)).

En dépit des variations de la taille des télomères dans une cellule, la taille moyenne des télomères des cellules exprimant la télomérase est inscrite dans une gamme bien définie pour une espèce donnée. Cela indique qu'il existe un équilibre entre l'élongation des télomères et leur raccourcissement (Fig. i54 B et C). Cet équilibre provient de la régulation de la télomérase par des protéines qui fixent en *cis* l'ADN télomérique. La clé de cette régulation repose sur un contrôle négatif dans lequel le produit de la télomérase, l'ADN télomérique, fixe un inhibiteur de la télomérase en quantité proportionnelle à la taille des télomères, et qui est le complexe shelterine (pour revues, (Bianchi et Shore, 2008; Palm et de Lange, 2008)).

7.5. La télomérase et ses protéines

La télomérase est un complexe ribonucléoprotéique de PM de 1 à 2 MDa composé de l'ARN TERC (Telomerase RNA Component) qui est détaillé dans la partie III.C.1.2 de l'introduction, et d'un nombre important de protéines aux fonctions très hétérogènes et pas toujours identifiées telles que la transcriptase réverse TERT (Telomerase Reverse Transcriptase), les protéines cœurs Dyskérine, NHP2, NOP10 et GAR1 qui sont associées aux RNP H/ACA, des hnRNP, des protéines Sm, des ATPases et des protéines chaperonnes (Fu et Collins, 2007).

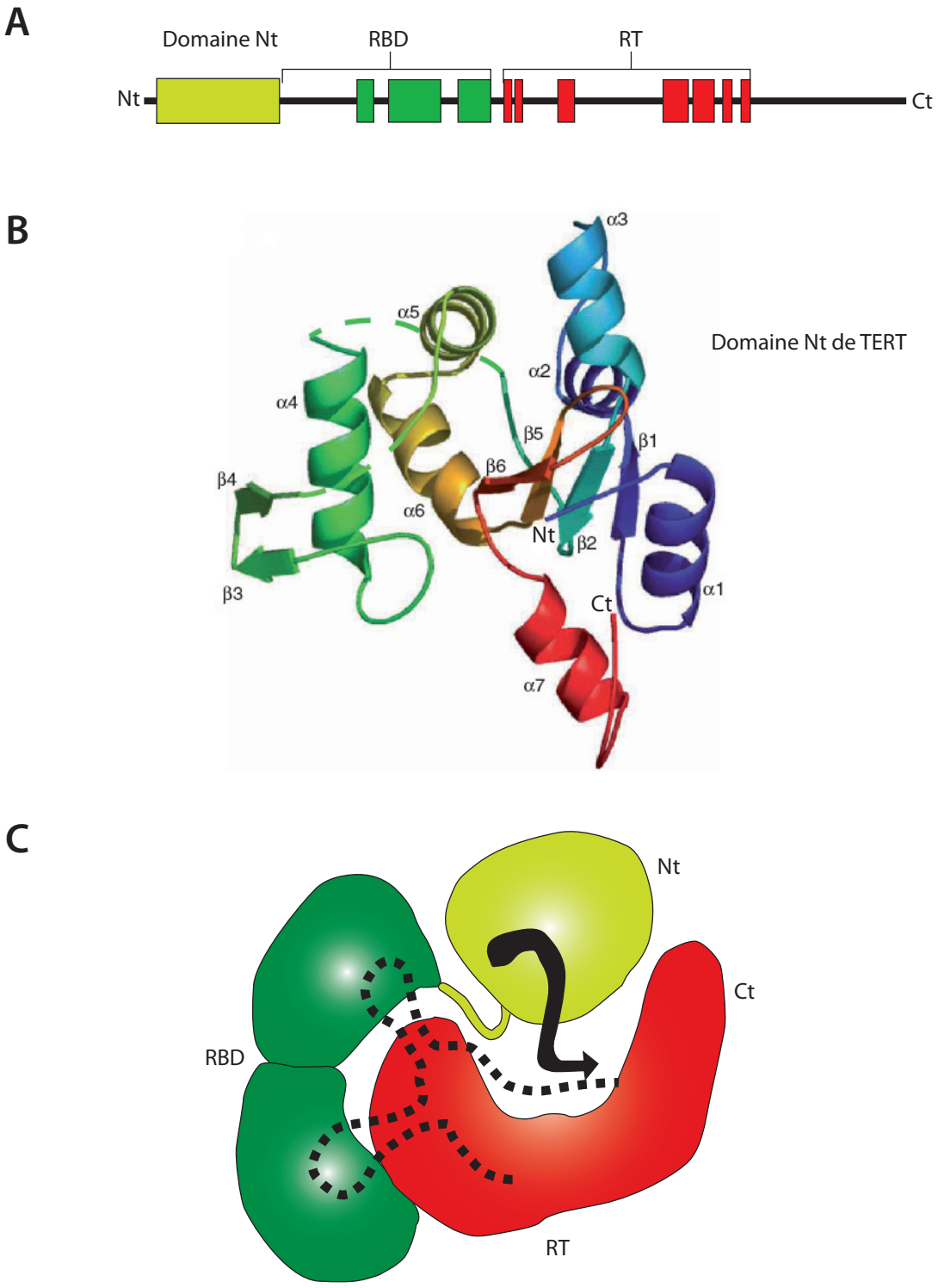


Figure i55 : Les domaines de la protéine TERT de la télomérase. (Adapté de Jacobs *et al.*, 2006).
A. Schématisation de la séquence de TERT indiquant les différents domaines Nt, région centrale (RBD, fixation à l'ARN) et Ct (RT, réverse transcriptase).
B. Structure 3D du domaine Nt de TERT de la télomérase de *Tetrahymena thermophilus*. Les lignes en pointillées représentent les boucles non retrouvées dans la carte électronique de densité.
C. Modèle de la contribution du domaine Nt au mécanisme de la télomérase. Le domaine Nt (en jaune) fixe une portion de l'ADN télomérique (flèche épaisse) qui est distant de l'extrémité 3' terminale (pointe de la flèche), et qui s'étend dans le domaine réverse transcriptase de TERT (en rouge). Le domaine Nt interagit aussi avec TERC (pointillé). Ces interactions sont cruciales pour la catalyse mais ne contribuent que faiblement à la stabilisation du complexe, ce qui est accompli principalement par le domaine central de fixation à l'ARN de TERT (en vert).

La transcriptase réverse TERT, codée par le gène EST2 (Ever Shorter Telomere gene 2) (Meyerson *et al.*, 1997), est composée de 3 domaines : le domaine Nt, la région centrale et le domaine Ct (Fig. i55 A). La structure du domaine Nt a été résolue par radiocristallographie (Fig. i55 B) (Jacobs *et al.*, 2006). Ces auteurs ont montré que ce domaine, qui n'est pas retrouvé dans les autres classes de transcriptases réverses, il participerait à la fixation du complexe à l'ADN télomérique sb et à l'ARN TERC, et il joue un rôle fondamental dans l'activité de la télomérase (Jacobs *et al.*, 2006). La région centrale constitue un domaine de fixation à l'ARN, tandis que le domaine Ct porte le site actif de transcriptase inverse (Fig. i55C).

Chez la levure, les protéines Est1, Est2 et Est3 sont associées à la télomérase et sont nécessaires à sa fonction *in vivo* (Lingner *et al.*, 1997). C'est Est2p qui porte l'activité catalytique de transcriptase réverse (Counter *et al.*, 1997). Les protéines Sm et la protéine Ku font également partie du complexe télomérase chez la levure et possèdent leur propre site de fixation sur l'ARN télomérique (Fig. i37 C) (pour revue, (Cech, 2004)).

La télomérase n'existe pas chez les Bactéries ou les Archaea. En effet, dans ces deux domaines du vivant, les génomes sont sous une forme circulaire, à l'exception de quelques espèces bactériennes qui pour répliquer l'extrémité de leur chromosome linéaire, possèdent leurs propres mécanismes basés sur des phénomènes de recombinaison. Donc le problème de séquence manquante à chaque réplication ne se pose pas pour les organismes procaryotes.

7.6. La biogenèse de la télomérase

Deux protéine de la famille des ATPases AAA+ (ATPases for Associated with various cellular Activities) et essentielles à l'assemblage de la télomérase ont été identifiées pour interagir avec la Dyskérine et TERT (Fig. i56). Il s'agit de la Pontine (TIP49) et de la Reptine (TIP48) dont les homologues chez la levure sont respectivement Rvb1p et Rvb2p (Venteicher *et al.*, 2008). Ces protéines sont

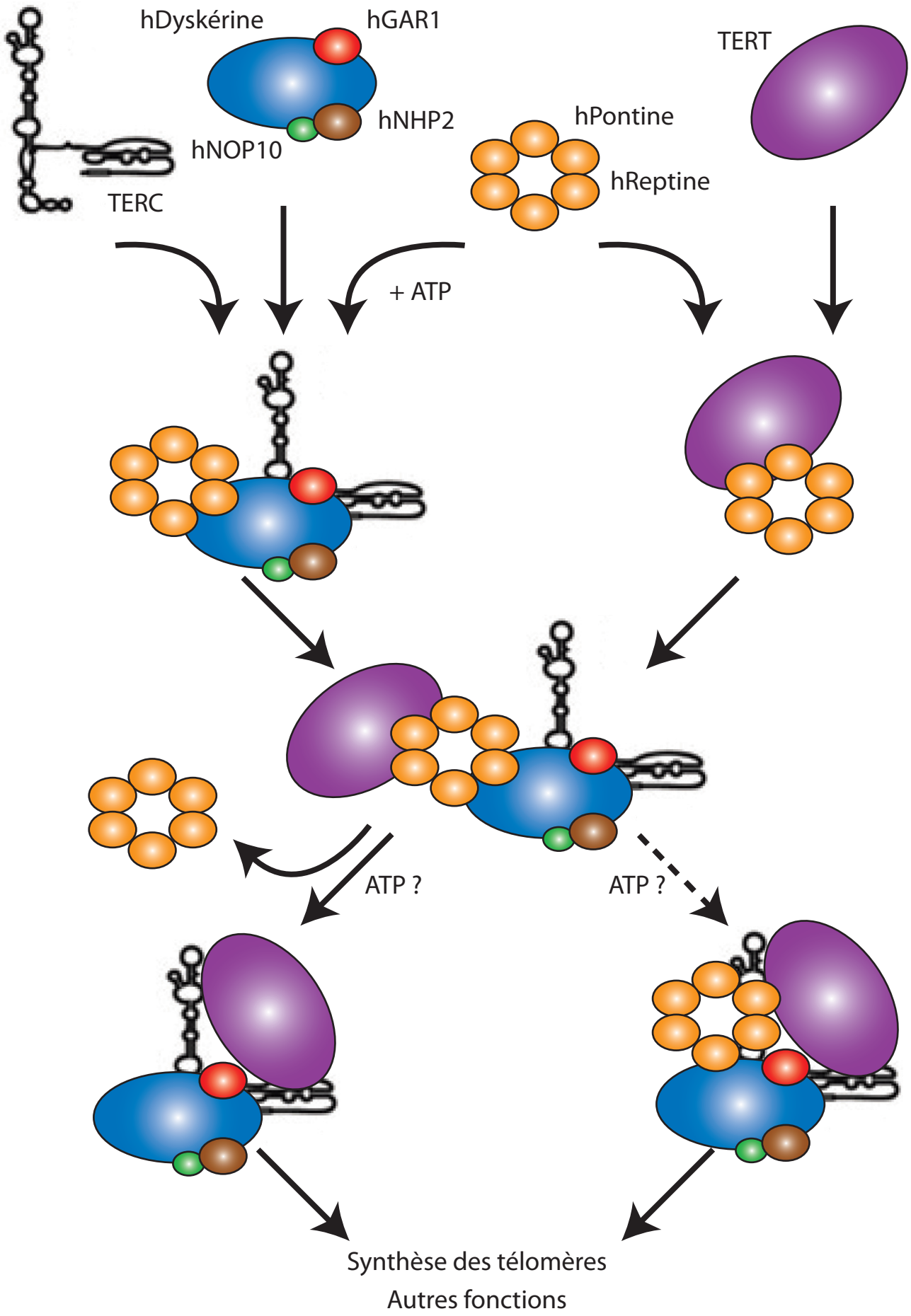


Figure i56 : Schéma d'assemblage de la RNP H/ACA sur l'ARN de la télomérase hTERC. Proposition d'un rôle fonctionnel des protéines d'assemblage Pontine et Reptine. (Adapté de Venteicher *et al.,* 2008). Voir le texte pour les détails.

fonctionnelles sous la forme d'anneaux hexamériques (Venteicher *et al.*, 2008). Ces deux protéines interviennent conjointement pour assembler et stabiliser la RNP TERT ce qui nécessite l'activité ATPase de la Pontine. Il est intéressant de souligner que ces deux protéines sont également retrouvées dans le complexe R2TP de la machinerie d'assemblages des snoRNA, confirmant que la télomérase pourrait être assemblée par cette même machinerie (cf IV.B.2.4) (McKeegan *et al.*, 2007; Boulon *et al.*, 2008; Venteicher *et al.*, 2008). La protéine NAF1 contribue également à la biogénèse de la télomérase. En effet, elle est requise pour la maturation de l'ARN de la télomérase (Hoareau-Aveilla *et al.*, 2006) et est inclue dans le complexe de maturation de la télomérase humaine (Venteicher *et al.*, 2009).

Les composants de la télomérase hTERC et hTERT sont majoritairement localisés dans le noyau et une proportion plus faible dans le nucléole car hTERT possède un signal de localisation nucléolaire situé dans les 15 premiers résidus de son extrémité Nt (Mitchell *et al.*, 1999a; Yang, Y. *et al.*, 2002). De plus, hTERC possède un signal de maturation à son extrémité 3' qui comprend i) une tige stable de 2 pb située en dessous de la boîte CAB, ii) un appariement U•G et iii) un résidu U non apparié (Fig. i57 A) (Theimer *et al.*, 2007). L'association de la protéine hTERT avec hTERC semblent être un pré-requis à la localisation du complexe de la télomérase dans les Corps de Cajal (Tomlinson *et al.*, 2008). C'est dans les Corps de Cajal qu'aurait lieu des évènements de maturation de la télomérase comme l'hyperméthylation de la coiffe et l'assemblage de protéines avec hTERC (Jady *et al.*, 2004).

La protéine TCAB1 (Telomerase Cajal Bodies protein 1), majoritairement localisée dans les Corps de Cajal, s'assemble avec le complexe de la télomérase comprenant au minimum les protéines hTERC, hTERT, Dyskérine, NAF1, Pontine et Reptine (Venteicher *et al.*, 2009). Elle interagit avec la boîte CAB de TERT, mais également avec les protéines TERC et la Dyskérine mais pas avec NAF1, la Pontine ou la Reptine, suggérant que la protéine TCAB1 fait partie du complexe actif de la télomérase plutôt que d'un complexe pré-télomérase (Fig. i57 B 1 et 2) (voir aussi IV.C.1.2 de l'introduction) (Venteicher *et al.*, 2009). Il est proposé que l'ARN hTERC

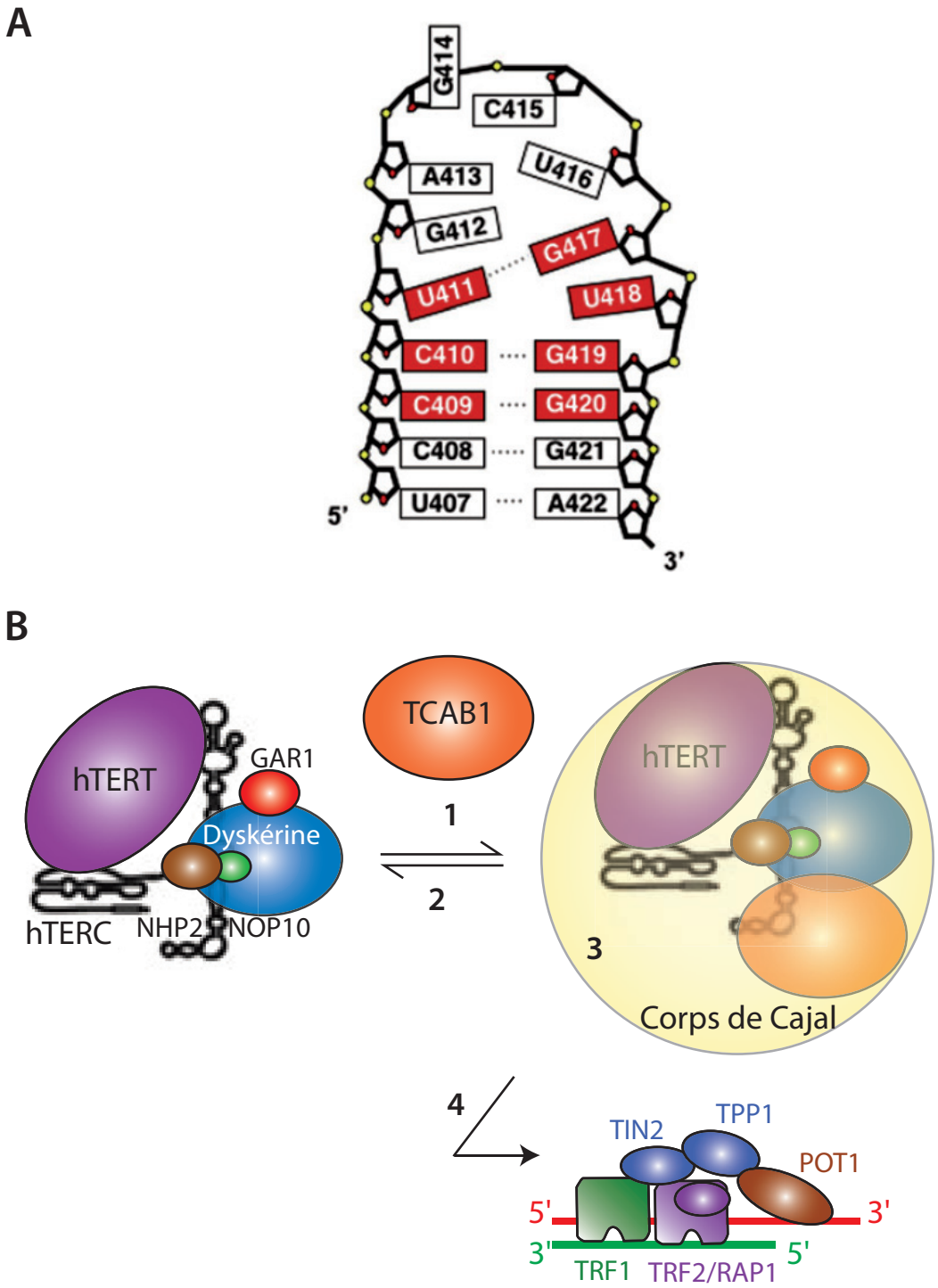


Figure i57 : Signal de maturation et modèle d'assemblage de la télomérase humaine. (Adaptés de Theimer *et al.,* 2007 et de Venteicher *et al*., 2009).
A. Signal de maturation spécifique à l'ARN hTERC. Les résidus du motif CR7 représentés en rouge correspondent à ceux qui ont un rôle significatif pour la maturation correcte de hTERC. Les nucléotides représentés en blanc n'ont aucun rôle sur la maturation de hTERC. Ces données ont été obtenues à partir de résultats de mutagenèse et de comparaison de séquences chez les Vertébrés.
B. La protéine TCAB1 est essentielle à la localisation de la télomérase dans les Corps de Cajal et à la synthèse des télomères. Modèle pour la fonction de TCAB1 dans la voie de synthèse des télomères. 1/ et 2/ transit de la télomérase entre le nucléoplasme et les Corps de Cajal, 3/ localisation de la télomérase dans les Corps de Cajal et 4/ localisation de la télomérase sur les télomères.

et tous les scaRNA soient retenus dans les Corps de Cajal par l'intermédiaire de la protéine TCAB1 qui favoriserait donc les étapes d'assemblage et de maturation de la télomérase (Venteicher *et al.*, 2009). Il est également proposé que TCBA1 facilite l'adressage du complexe de la télomérase vers les télomères (Fig. i57 B 3 et 4) (Venteicher *et al.*, 2009).

Par ailleurs, il a été montré que le complexe SMN interagissait avec hTERT et GAR1 (Bachand *et al.*, 2002; Whitehead *et al.*, 2002), suggérant que le complexe SMN pourrait avoir un rôle dans la biogenèse de la télomérase et des snoRNP H/ACA (voir aussi IV.C.4.1.2 de l'introduction) (Tomlinson *et al.*, 2006).

8. La Dyskératose Congénitale humaine, une maladie génétique rare

La dyskératose congénitale (DC) ou syndrome de Zinsser-Cole-Engman est une maladie génétique rare qui a été décrite pour la première fois sur le plan dermatologique par Zinsser en 1906. Il s'agit d'une maladie humaine qui est polymorphe sur le plan clinique et variable dans son mode de transmission. On estime sa prévalence à une personne sur un million. Certains symptômes peuvent apparaître tardivement, ce qui rend difficile son diagnostique clinique. Néanmoins, depuis 1995, l'équipe d'Inderjeet Dokal a ouvert un registre international au Royal London Hospital qui recense tous les cas de patients atteints par cette maladie (données cliniques et prélèvements sanguins). Il recense plus de 280 familles à travers le monde (pour revue, (Walne et Dokal, 2008)).

Les manifestations cliniques de la DC apparaissent en général pendant l'enfance, d'abord par des changements au niveau de la pigmentation de la peau (aspect réticulé et hyperpigmenté de la peau), de la dystrophie des ongles et d'une leucoplasie des muqueuses buccales (apparition de plaques blanchâtres à la surface de l'épithélium ou des muqueuses) vers 10 ans (Sirinavin et Trowbridge, 1975). La défaillance de la moelle osseuse se développe avant l'âge de 20 ans et 85% des patients ont développé des anormalités de la moelle osseuse avant 30 ans ce qui représente la cause majeure de décès (pour revues, (Dokal, 2000; Hoareau-Aveilla *et*

Symptômes	Pourcentage des patients affectés
Atteintes cutanées	89 %
Dystrophie des ongles	88 %
Aplasie médullaire (défaillance de la moelle osseuse)	86 %
Leucokératose	78 %
Larmoiement anormal	30 %
Atteintes neurologiques (retard mental)	25 %
Maladies pulmonaires	20 %
Retard de croissance	19 %
Anomalies dentaires	17 %
Sténoses de l'oesophage	17 %
Perte de cheveux prématurés	16 %
Hyperhidrose	15 %
Cancers	10 %
Cirrhose du foie	7 %
Coordination musculaire défaillante	7 %
Hypogonadisme	6 %
Microcéphalie	6 %
Restriction urinaire	5 %
Ostéoporose précoce	5 %
Surdité	1 %

Tableau i7 : Fréquence de différents symptômes chez les patients atteints de Dyskératose Congénitale (d'après Walne *et al.*, 2008).

al., 2008; Vulliamy et Dokal, 2008)). Cette maladie se caractérise ainsi par des altérations des tissus à fort taux de renouvellement : la peau et la moelle osseuse. Certains cas peuvent parfois aboutir à des cas de cancers. Les différents symptômes provenant de la DC sont répertoriés dans le tableau i7. Deux types de traitements existent pour lutter contre les défaillances de la moelle osseuse. Le premier est basé sur l'emploi de stéroïdes anabolisants (érythropoïétine) qui stimulent les fonctions hématopoïétiques de certains patients pour une période donnée. Le second traitement, qui est le principal, est la greffe allogénique (provenant d'un autre individu) de cellules souches hématopoïétiques (pour revue, (Dokal et Vulliamy, 2003)).

La Dyskératose Congénitale fait intervenir deux composants majeurs dans la cellule. Il s'agit du complexe de la télomérase, des snoRNP et scaRNP à boîtes H/ACA. Les symptômes de cette maladie peuvent donc provenir d'un défaut dans la maintenance des télomères et/ou dans la biogenèse des ribosomes pouvant former ainsi une balance entre ces deux éléments, ce qui ne facilite pas l'interprétation et l'étude de la DC chez l'homme (pour revue, (Meier, 2003)). Les observations actuelles suggèrent néanmoins que la DC pourrait être une défaillance de la fonction de la télomérase plutôt qu'un défaut de pseudouridylation ou d'épissage si la biogenèse des snRNP est perturbée par des défauts de scaRNP. La Dyskératose Congénitale peut exister sous trois formes distinctes : liée au chromosome X, autosomale dominante (AD) ou autosomale récessive (AR) (Tab. i8). Néanmoins, pour plus de 60% des patients présentant les symptômes de la DC, les facteurs génétiques impliqués sont toujours inconnus à ce jour (pour revue, (Walne et Dokal, 2008)). En France, « l'Association Française de la Dyskératose Congénitale (AFDC) » a été créée en avril 2007 dont le site internet peut être consulté à l'adresse suivante http://www.afdc.asso.fr. La première « étude épidémiologique sur la Dyskératose Congénitale et ses formes atypiques » a débuté fin 2008 et s'étendra sur une période de 2 ans.

Maladie	Dyskératose Congénitale		
Forme	Liée à l'X	Autosomale dominante	Autosomale récessive
Composant touché	Dyskérine	Télomérase (TERC+TERT) hTINF2	NOP10 NHP2 TERT
RNP affectée	Télomérase H/ACA	Télomérase Shelterine	Télomérase H/ACA

Tableau i8 : Tableau récapitulatif des différentes formes de la Dyskératose Congénitale, les facteurs impliqués et le type de RNP affecté.

8.1. La forme liée au chromosome X

C'est en 1998 que pour la première fois un gène responsable de la DC a été identifié. L'équipe d'I. Dokal a analysé l'ADNc de 28 patients et elle a déterminé que le gène à l'origine de la maladie se situait sur le locus Xq28 du chromosome X (Heiss *et al.*, 1998). Il s'agit du gène *DKC1* qui code la Dyskérine. La DC-X est la forme la plus courante (30% à 40% des cas). Dans le gène *DKC1*, il a été identifié 39 mutations faux-sens, 3 qui conduisent à des défauts d'épissages, 1 située dans le promoteur, 1 délétion d'un résidu et 1 mutation silencieuse (Marrone *et al.*, 2005). La structure 3D de la sRNP d'archaea indique que la majorité des mutations identifiés dans la Dyskérine humaine sont regroupées sur une face du domaine PUA de la protéine aCBF5 (Fig. i58) (pour revue, (Vulliamy et Dokal, 2008)). Une fois le gène *DKC1* identifié, il s'est rapidement posé la question de savoir comment des mutations individuelles dans la Dyskérine peuvent conduire à un défaut d'expression ou d'activité variables en fonction des tissus (Luzzatto et Karadimitris, 1998).

8.1.1. Etudes réalisées sur des cellules humaines portant de mutations dans le gène *DKC1*

Les études menées sur des lignées de cellules humaines porteuses de mutations dans *DKC1* n'indiquent aucun changement dans le taux de pseudouridylation des ARNr et dans le taux et la cinétique de maturation des ARNr (Mitchell *et al.*, 1999b; Wong et Collins, 2006). A l'inverse, le taux de hTERC chute et la taille des télomères est raccourcie (Mitchell *et al.*, 1999b; Wong et Collins, 2006). Néanmoins, l'activation de la télomérase dans des cellules humaines dans lesquelles il existe des mutations dans *DKC1* ne permet pas l'élongation des télomères ou leur maintenance de façon stable (Wong et Collins, 2006).

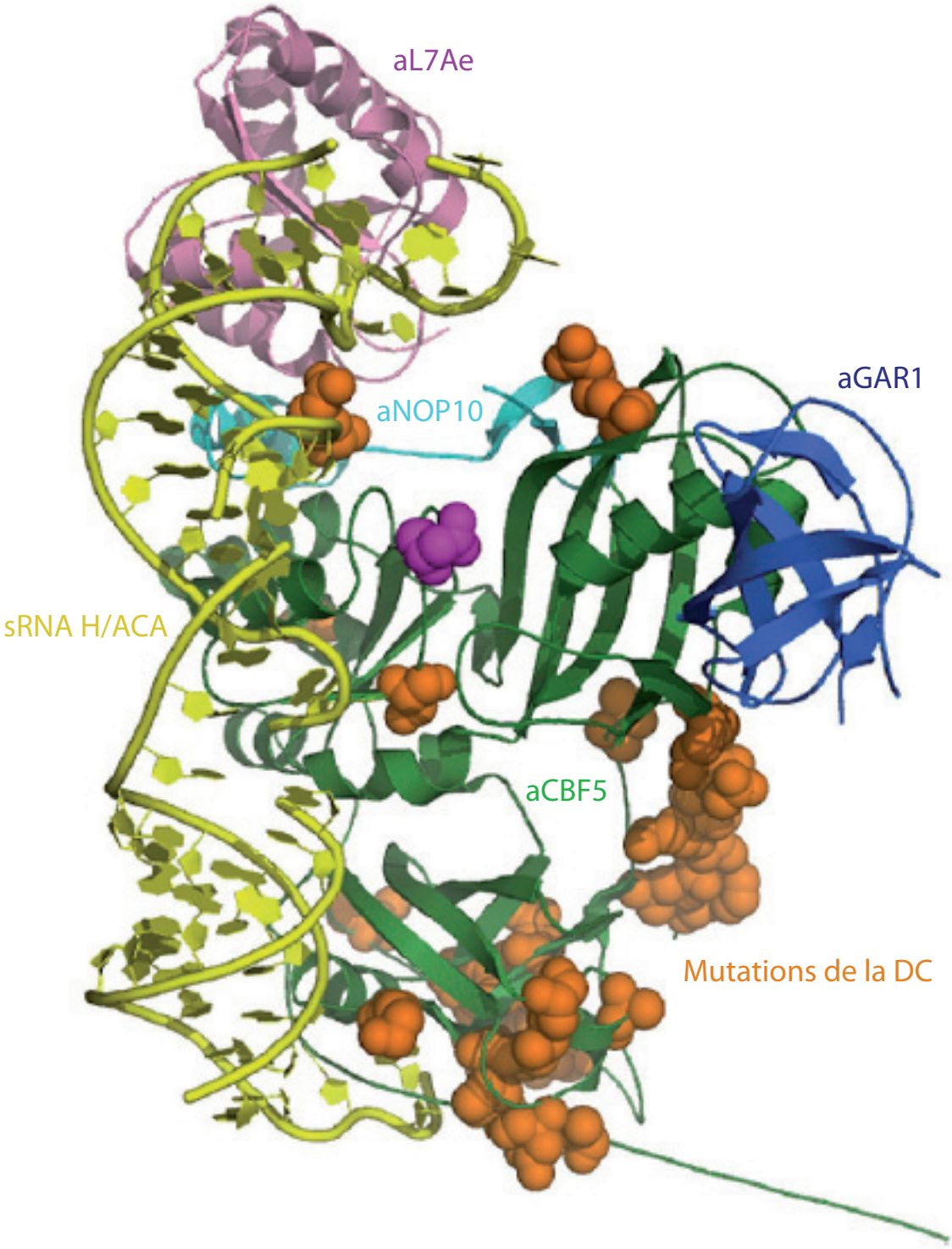


Figure i58 : Les mutations de la Dyskérine humaine transposées dans la particule sRNP à boîtes H/ACA d'archaea. Tiré de Vulliamy et Dokal, 2008 à partir de la structure de Li et Ye, 2006.
aCBF5 est représentée en vert, aGAR1 en bleu, aNOP10 en cyan et L7Ae en rose, l'ARN H/ACA jaune, le résidu catalytique en magenta et les chaînes latérales des résidus mutés équivalents à ceux identifiés dans le cas de la DC humaine sont matérialisés par des sphères oranges.

8.1.2. Etudes réalisées sur des souris portant des mutations dans le gène *DKC1* humain

Chez des souris portant des mutations dans le gène *DKC1* humain, les générations précoces portent les mêmes symptômes que chez l'homme (défaillance de la moelle osseuse et signes cutanés), le taux de pseudouridylation dans les ARNr est réduit, la maturation des ARNr est altérée et les ribosomes deviennent hypersensibles à certaines drogues qui inhibent la traduction. Alors que les générations tardives de ces souris présentent des défauts dans la fonction de la télomérase (réduction de la taille des télomères et chute de l'activité de la télomérase) (Ruggero *et al.*, 2003). Les auteurs ont ainsi suggéré que l'initiation de la DC serait liée aux défauts des ribosomes et que l'aggravation de la maladie proviendrait plus tardivement par la défaillance du complexe télomérase.

La substitution A353V dans la Dyskérine correspond à la mutation la plus fréquemment retrouvée dans la DC. Une seconde étude menée sur des cellules embryonnaire de souris a montré que la substitution A353V conduisait à l'accumulation de l'ARN de la télomérase, des snoRNA H/ACA, des défauts dans l'activité de la télomérase et dans la maturation des ARNr (Mochizuki *et al.*, 2004).

La réduction du taux de pseudouridylation des ARNr induite par des mutations dans la Dyskérine ne perturbe pas l'activité de la machinerie de la synthèse protéique (Yoon *et al.*, 2006). Par contre, des souris portant des mutations dans la Dyskérine, présentent un défaut très spécifique de la traduction initiée à partir des sites d'entrées internes (IRES) du ribosome sur les ARNm (Yoon *et al.*, 2006). Cela pourrait relier certains cas de cancers humains de la DC car de nombreux facteurs anti-apoptotiques possèdent des éléments IRES dans leur 5'UTR qui leur permettent l'initiation de la traduction lors de conditions de stress (cancers…), procurant aux cellules un avantage pour leur survie. L'ensemble de ces travaux menés chez la souris indiquent que la réduction du taux de pseudouridylation des ARNr, liée aux mutations dans la Dyskérine, affecterait la traduction par mécanisme IRES d'ARNm cellulaires à fonction cellulaire importante (Yoon *et al.*, 2006).

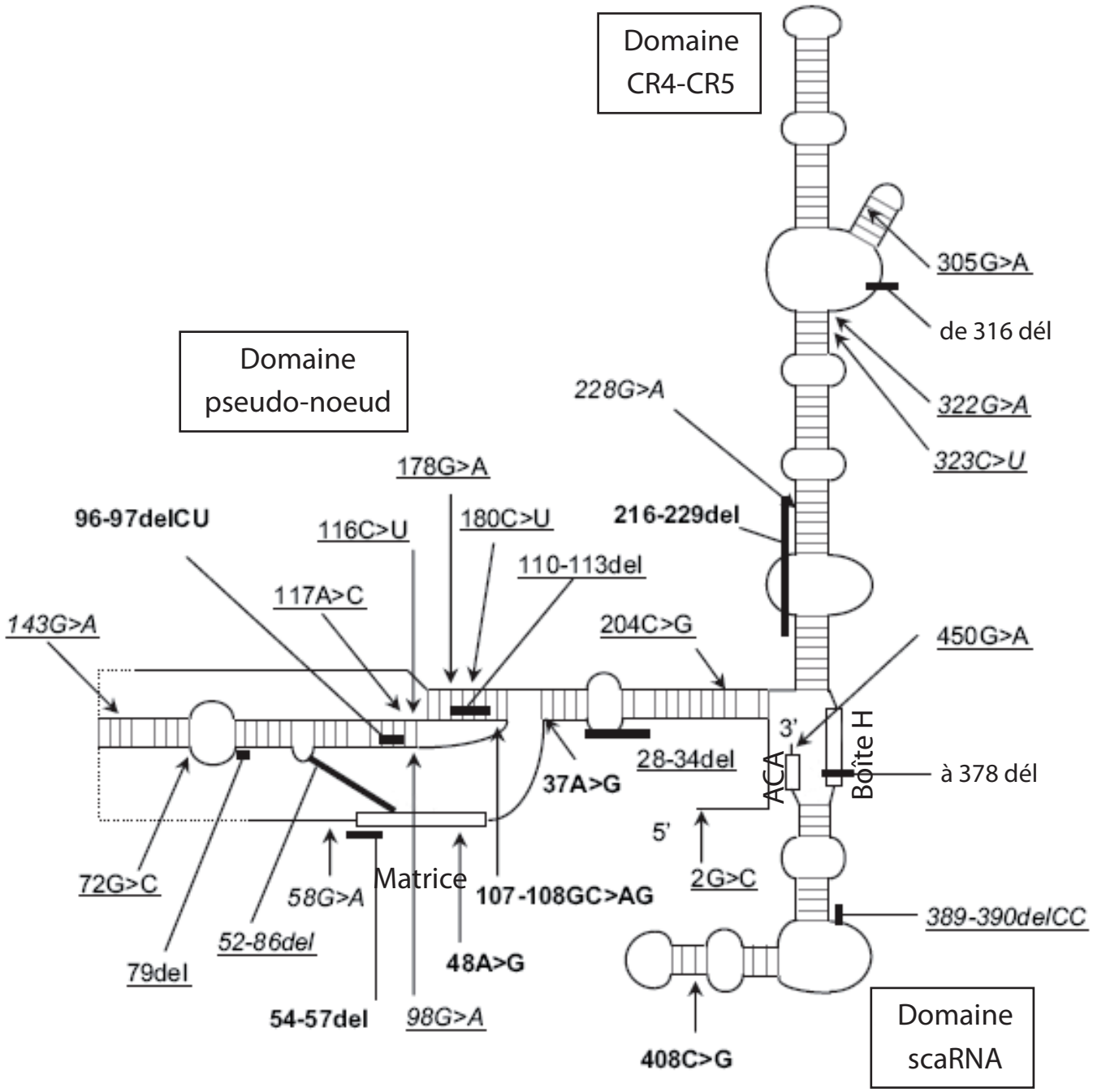


Figure i59 : Localisation des mutations sur l'ARN de la télomérase qui conduisent à la forme AD de la DC et d'autres pathologies. (Vulliamy et Dokal, 2008).
Les mutations identifiées chez des patients atteint de la DC sont indiquées en gras, de l'Anémie Aplasique soulignées, de la myélodysplasie, de la thrombocytose, de la fibrose pulmonaire soulignées et en italique et les individus normaux en italique. Les mutations ponctuelles sont représentées par des flèches et les délétions par des barres.

8.1.3. Bilan des études menées chez l'homme et chez la souris

Le modèle souris et le modèle humain semblent toutefois différer en ce qui concerne i) la nécessité de la Dyskérine pour la fonction de la télomérase et ii) la taille des télomère beaucoup plus importante chez la souris que chez l'homme (Wong et Collins, 2006). Dans la maladie humaine, il est difficile de dissocier les effets sur l'activité des snoRNP H/ACA (effet sur le ribosome), des effets sur la télomérase par l'intermédiaire du motif U64 H/ACA de hTERT (raccourcissement des télomères).

8.2. La forme autosomale dominante (AD)

La DC à mode de transmission autosomal dominant semble représenter 30% à 40% des cas (pour revue, (Vulliamy et Dokal, 2008)).

8.2.1. Mutations touchant la séquence de l'ARN de la télomérase (hTERC)

L'existence d'un lien entre la DC et la télomérase a été établie pour la première fois par l'équipe de Dokal en 2001. Une étude réalisée sur 3 familles atteintes de DC, indique que les mutations dans les trois domaines de l'ARN hTERC (Fig. i59) d'un seul allèle de l'ARN de la télomérase sont suffisantes pour induire la maladie (Vulliamy *et al.*, 2001).

Il a ensuite été découvert que la taille des télomères était significativement plus courte dans la 2^{e} génération de familles possédant des mutations dans TERC et que le phénomène de raccourcissement semblait s'amplifier au cours des générations (Vulliamy *et al.*, 2004). C'est un taux réduit de TERC qui limiterait la maintenance des télomères (Wong et Collins, 2006). Des mutations dans la partie scaRNP H/ACA (Fig. i59) de TERC réduisent considérablement l'assemblage de la scaRNP H/ACA (Trahan et Dragon, 2009).

8.2.2. Mutations touchant la séquence de la transcriptase inverse de la télomérase (hTERT)

La mutation d'un résidu parfaitement conservé dans la transcriptase réverse humaine (K902N) induit le phénotype de la DC. Cette mutation conduit à la perte de l'activité *in vitro* de transcriptase réverse et au raccourcissements *in vivo* des télomères (Armanios *et al.*, 2005). La présence d'un seul allèle mutant (K902N) n'inhibe pas la fonction de la télomérase sauvage (2^{e} allèle sauvage) *in vitro* mais réduit son activité au cours du temps. Cette mutation provoque le phénomène d'haploinsuffisance de la télomérase. La présence de l'enzyme sauvage seule (1 allèle) n'est pas suffisante pour avoir l'activité nécessaire au non déclanchement de la maladie (Armanios *et al.*, 2005). Néanmoins, une autre étude montre que dans deux familles consanguines, d'autres mutations homozygotes dans TERT conduisent à la DC avec un profil de transmission autosomal récessif (activité télomérase réduite et raccourcissement des télomères) (Marrone *et al.*, 2007).

8.2.3. Mutations touchant la séquence de la protéine TIN2

Des mutations dans le gène codant la protéine TIN2 ont été trouvées chez des patients (Savage *et al.*, 2008). Elle est l'un des constituants du complexe shelterine, ayant ainsi un lien direct avec le complexe télomérase (Fig. i54). Ces mutations conduisent à un raccourcissement de la taille des télomères et provoquent un phénotype sévère de DC chez ces patients. Dans ce cas, il est intéressant de noter que le taux d'ARN TERC n'est pas réduit (Walne *et al.*, 2008). Comme la taille des télomères diminue progressivement avec l'âge, un parent atteint de la forme AD fournit à ses enfants des télomères plus petits. Les enfants qui hériteront à la fois de télomères plus courts et de mutations dans l'ARN télomérique seront malheureusement porteurs d'une maladie plus sévère. Aucun cas d'enfants héritant de télomères raccourcies sans mutation dans l'ARN télomérique et sans développer les symptômes de la DC n'a été répertorié (Shay et Wright, 2004). Sur le modèle de

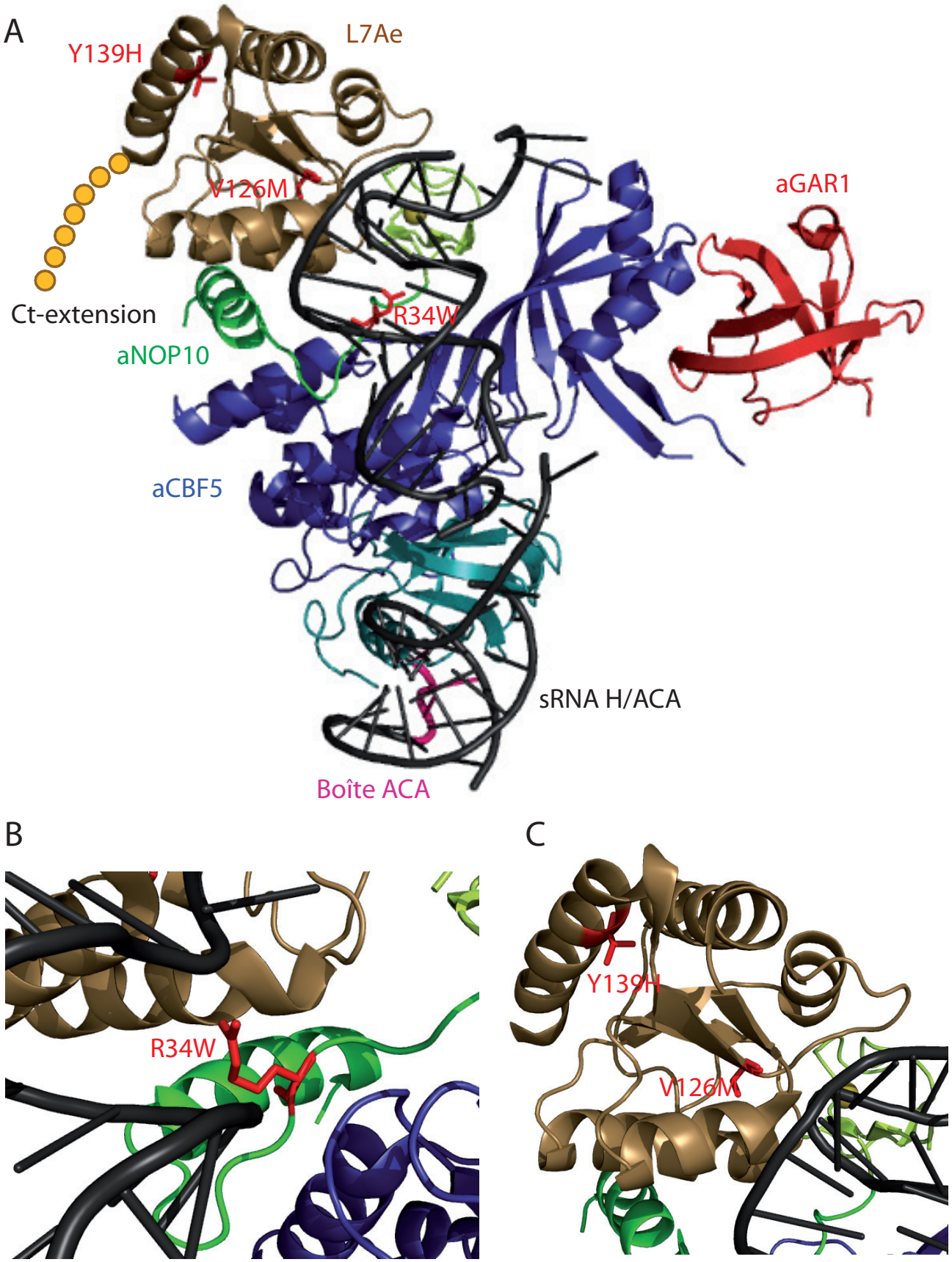


Figure i60 : Les mutations conduisant la forme AR de la DC ont été transposées dans la particule H/ACA d'archaea.

A. aCBF5 est représentée en bleu, aGAR1 en rouge, aNOP10 en vert et L7Ae en brun, l'ARN H/ACA en noir, la boîte ACA en magenta et les équivalents des mutations identifiées dans des cas de la DC-AR humaine sont représentées en rouge dans L7Ae (NHP2 humaine Y139H et V126M) et aNOP10 (NOP10 humaine R34W).

B et **C.** Vues rapprochées.

Réalisé à l'aide du logiciel Pymol à partir de la structure PBD=2HVY (Li et Ye, 2006).

transmission AD, il est possible de prédire et d'anticiper la maladie en se basant sur l'arbre généalogique des familles à risques (Marrone *et al.*, 2005).

8.3. La forme autosomale récessive (AR)

Certains facteurs génétiques conduisant à la forme AR ont récemment été identifiés, suggérant que la forme DC-AR représente 15 à 20% des cas.

8.3.1. Mutation touchant la séquence de la protéine NOP10

La substitution R34W a été détectée dans la protéine NOP10 de 3 individus d'une famille issue de mariages consanguins (Fig. i60 A et B) (Walne *et al.*, 2007). Cette mutation conduit au profil classique de la DC : la taille des télomères et le taux de TERC sont affectés mais aucune activité de pseudouridylation n'a été évaluée. Cette mutation est prédite pour avoir une conséquence fonctionnelle importante car elle se situe dans une région conservée de la protéine et le changement d'un acide aminé basique en un résidu aromatique pourrait changer la structure de la protéine (Walne *et al.*, 2007). Dans les structures 3D d'archaea, le résidu R34 de aNOP10 interagit directement soit avec aCBF5 (Rashid *et al.*, 2006), soit avec le sillon majeur de l'hélice H2 du sRNA (Li et Ye, 2006). Il est proposé que l'addition d'un noyau aromatique dans la structure de NOP10 devrait interrompre l'interaction entre la Dyskérine et NOP10 et l'ARN guide et conduire à réduire l'activité de la snoRNP (Walne *et al.*, 2007). Nous avons donc cherché à savoir quel pouvait être le rôle joué par cette mutation dans l'assemblage et l'activité des particules d'archaea. Ce volet sera détaillé dans les parties II.D et E des résultats.

8.3.2. Mutations touchant la séquence de la protéine NHP2

Lors d'une analyse de gènes de patients atteints de la DC, dont la provenance génétique n'était pas connue, des mutations dans la protéine NHP2 ont été détectées

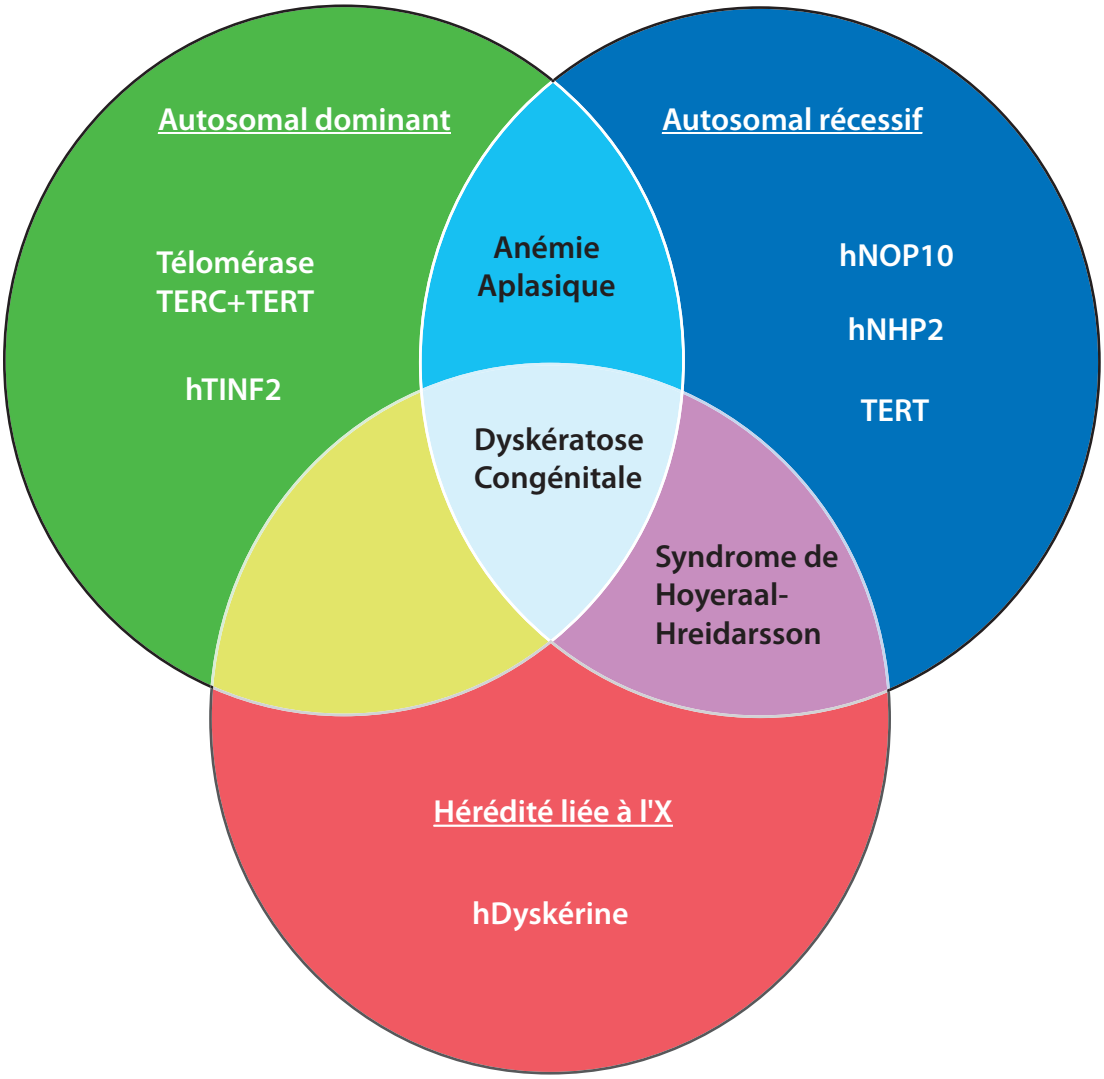


Figure i61 : Schéma représentant le chevauchement clinique et génétique entre la Dyskératose Congénitale, l'Anémie Aplasique et le syndrome de Hoyeraal-Hreidarsson. (Adapté de Marrone *et al.*, 2005).

(Vulliamy *et al.*, 2008). Il s'agit des mutations V126M, Y139H et X154R. Dans le cas de la mutation X154R, le codon stop (154^{e} position) est substitué par un codon arginine, conduisant donc à un allongement de la protéine de 51 résidus en Ct (204 résidus alors que NHP2 sauvage est constituée de 153 résidus) (Fig. i60 A et C). Ces mutations induisent un raccourcissement de la taille des télomères et conduisent à une chute du taux de TERC (Vulliamy *et al.*, 2008). Dans cette même étude, aucune mutation dans GAR1 n'a été trouvée conduisant à l'une des formes de la DC.

8.4. Les autres syndromes liés à la DC

L'Anémie Aplasique (AA) est un grave dérèglement de la moelle osseuse qui affecte entre 2 à 5 personnes par million et par an (Fig. i61). Chez 5 à 10% des patients, cette maladie est constitutive c'est-à-dire qu'elle est familiale ou associée à d'autres anormalités somatiques (Vulliamy *et al.*, 2002). Dans ce cas, des facteurs génétiques sont mis en jeu. Chez certains patients, les défaillances de la moelle osseuses peuvent apparaître avant les manifestations muco-cutanées conduisant ainsi à un diagnostique de l'AA (pour revue, (Dokal et Vulliamy, 2003)). Des mutations dans TERC (Vulliamy *et al.*, 2002) (Fig. i59) ainsi que dans TERT (Yamaguchi *et al.*, 2005) ont été identifiées chez des patients anémiques. Comme pour la DC, les mutations dans TERC ou TERT conduisent à la réduction de la taille des télomères.

Le syndrome de Hoyeraal-Hreidarsson (HH) est l'une des formes les plus sévères de la DC dans laquelle les individus affectés meurent avant de développer les signes muco-cutanés caractéristiques de la DC (Fig. i61). En effet, les patients décèdent avant 5 ans alors que les signes cliniques de la DC se déclarent aux alentours de 8-10 ans (pour revue, (Dokal, 2000)). Des mutations ont été retrouvées dans le gène *DKC1* codant la Dyskérine, A353V, S121G et R158W et la sévérité de la maladie pourrait s'expliquer par la proximité de ces mutations avec le site actif de la Dyskérine (Alter, 2000; Yaghmai *et al.*, 2000; Vulliamy et Dokal, 2008). Des mutations dans TERC ont également été détectées (Fig. i60) (Marrone *et al.*, 2007). Les maladies

AA ou le syndrome de HH conduisent à des symptômes et impliquent des facteurs génétiques qui semblent chevaucher ceux de la DC (Fig. i61).

V. Objectifs de la thèse

Comme tout mécanisme biologique complexe, la réaction de pseudouridylation englobe de nombreux évènements moléculaires qui peuvent être identifiés grâce à la caractérisation d'espèces intermédiaires formées lors de son déroulement. Ces intermédiaires peuvent s'accumuler suite au blocage ou au ralentissement d'une étape du mécanisme réactionnel, provoqués soit par à un facteur exogène (une drogue par exemple) ou soit par un facteur endogène (une mutation dans l'un des composants de la machinerie intervenant dans le mécanisme). Les intermédiaires accumulés peuvent alors être caractérisés sur le plan structural et fonctionnel. La comparaison des données obtenues sur ces intermédiaires permet ensuite de reconstituer la séquence des grandes étapes du déroulement du mécanisme. A ce stade, la compréhension des bases moléculaires du passage d'un intermédiaire à l'autre requiert des méthodes d'investigations plus pointues, capables par exemple de suivre en temps réel le déroulement de l'une des étapes du mécanisme.

Dans le cas de la réaction de pseudouridylation, la structure de plusieurs sous complexes d'assemblage de sRNP à boîtes H/ACA et de complexes sRNP-ARN substrat a été établie (Li et Ye, 2006; Liang, B. *et al.*, 2007; Duan *et al.*, 2009; Liang *et al.*, 2009). La comparaison de l'ensemble de ces structures 3D et les données expérimentales de biochimie (Youssef *et al.*, 2007; Baker *et al.*, 2008) ou de fluorescence (Liang *et al.*, 2008) obtenues lors de l'étude des différents complexes permet de proposer un modèle en 5 grandes étapes du mécanisme de la réaction de pseudouridylation (Fig. i62) :

- Etape n°1 : biogenèse de la particule consistant en l'assemblage des protéines (P) sur l'ARN guide (sRNA) pour former la particule sRNP (Fig. i62 étape 1).

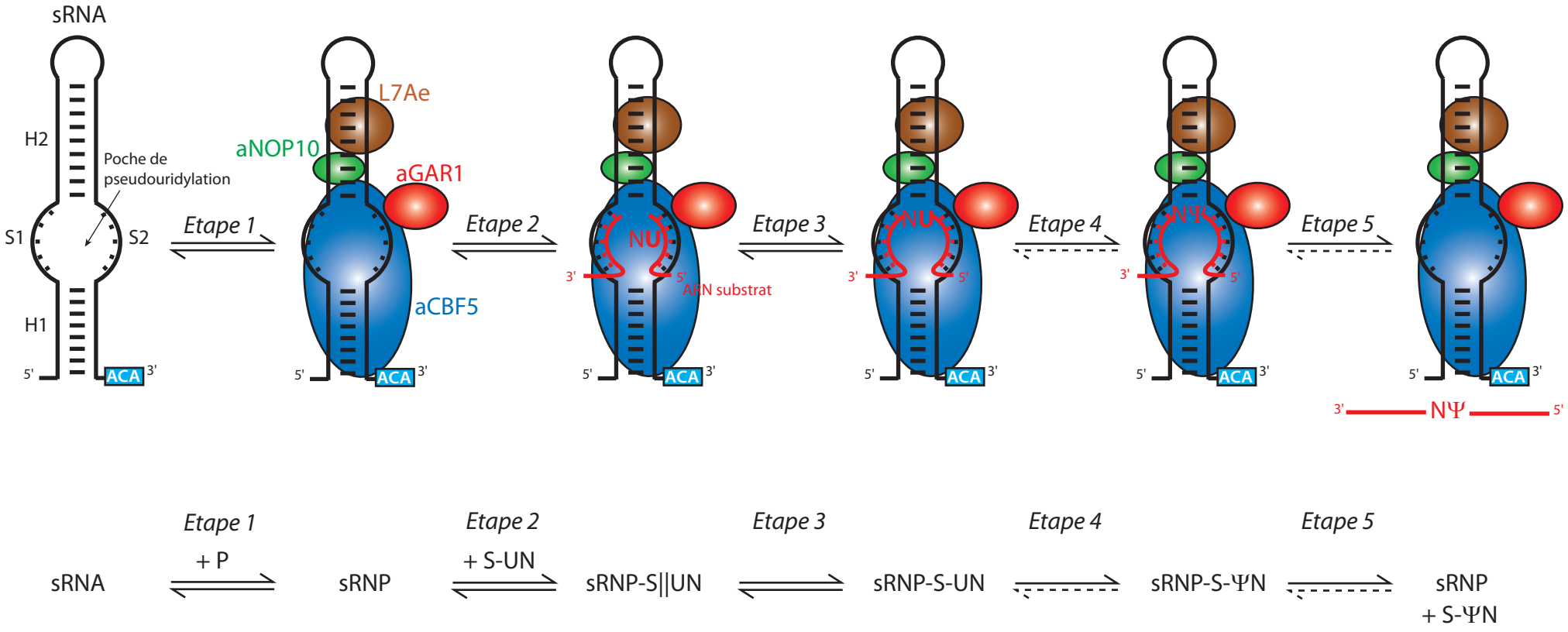


Figure r62 : Représentation schématique des différentes étapes du mécanisme d'une réaction de pseudouridylation dans le système guidé.
(Adapté à partir des résultats obtenus par Liang *et al.*, 2008).
Le mécanisme de la réaction de pseudouridylation se décompose en 5 grandes étapes.
Etape n°1 : association des protéines (P) L7Ae (en brun), aCBF5 (en bleu), aNOP10 (en vert) et aGAR1 (en rouge) au sRNA à boîtes H/ACA (en noir) remodelant la structure de l'ARN (sRNP).
Etape n°2 : hybridation de l'ARN substrat (S-UN) (en rouge) avec la poche de pseudouridylation dont les deux nucléotides en simple brin ne sont pas encore positionnés dans le site catalytique de aCBF5 (sRNP-S||UN).
Etape n°3 : après repositionnement du substrat, les deux nucléotides simple brin sont placés dans le site actif de aCBF5 (sRNP-S-UN).
Etape n°4 : réaction de catalyse du substrat qui aboutit à l'isomérisation du résidu U cible en résidu Ψ (sRNP-S-ΨN).
Etape n°5 : libération du produit de la réaction, le substrat modifié (sRNP+S-ΨN).

- Etape n°2 : entrée de l'ARN substrat (S-NU) dans la particule et appariement de sa séquence à la poche de pseudouridylation mais dont les deux nucléotides en simple brin ne sont pas positionnés de façon optimale dans le site actif de aCBF5 (sRNP-S||UN) (Fig. i62 étape 2). Ce type d'intermédiaire peut être obtenu lors d'un assemblage effectué en absence de la protéine L7Ae et avec un sRNA tronqué du motif en K-turn (Liang, B. *et al.*, 2007).

- Etape n°3 : repositionnement de l'uridine cible qui vient se loger à proximité de l'aspartate catalytique du site actif de aCBF5 (sRNP-S-UN) (Fig. i62 étape 3). Cet évènement peut être suivi par spectroscopie de fluorescence lorsque l'ARN substrat porte un résidu 2-AP à la position immédiatement en 3' de l'uridine cible (Liang *et al.*, 2008).

- Etape n°4 : déroulement du mécanisme catalytique qui permet la conversion de l'uridine en pseudouridine (sRNP-S-ΨN) (Fig. i62 étape 4).

- Etape n°5 : libération du substrat modifié (sRNP+S-ΨN) (Fig. i62 étape 5).

Au cours de mon travail de thèse, je me suis intéressé aux déterminants moléculaires sur lesquels reposait de déroulement de plusieurs de ces étapes: celle de l'assemblage des particules (étape 1), de la fixation du substrat (étapes 2+3), de la réaction de catalyse (étape 4), et de la libération de l'ARN substrat (étape 5) qui conditionne le turn-over de l'enzyme sRNP.

Lorsque je suis arrivé au laboratoire, Bruno Charpentier et l'équipe de M. et R. Terns venaient de mettre au point un système de reconstitution *in vitro* de particules d'archaea sRNP à boîtes H/ACA actives. Par cette approche d'assemblage *in vitro* et de mesure de l'activité des particules, certains déterminants présents sur l'ARN et le rôle fonctionnel général des protéines venaient juste d'être identifiés. Les structures 3D de la plupart des ARN:Ψ-synthases bactériennes étaient connues et celles-ci pouvaient servir de base pour débuter des études de relation structure-fonction de l'ARN:Ψ-synthase aCBF5 d'archaea. Les informations structurales sur les particules sRNP H/ACA des archaea ont nettement progressé ces dernières années : au cours

de mon travail de thèse, plusieurs équipes dont la nôtre, ont résolu la structure de complexes entre protéines de ces particules ou entre ARN et protéines. Ces données très importantes ont facilité l'avancée de nos études visant à déterminer le rôle exact de chacune des protéines dans l'assemblage des particules et dans la réaction de pseudouridylation.

Compte tenu de l'absence, au démarrage de ma thèse, de données structurales sur les ARN guides à boîtes H/ACA, le premier volet de mon travail a consisté à préciser les déterminants moléculaires présents sur ces ARN et qui étaient importants pour l'assemblage et/ou l'activité des particules. En particulier, nous avions constaté de grandes différences d'affinité de la protéine aCBF5 pour sa fixation à divers sRNA à boîtes H/ACA. J'ai donc recherché quelles pouvaient en être les raisons.

Le second axe de recherche visait à essayer de comprendre le rôle respectif des protéines associées aux sRNA à boîtes H/ACA et de leurs différents domaines dans l'assemblage et dans l'activité des particules. Mes études se sont ensuite orientées vers 4 points:

1) Plusieurs équipes avaient constaté que l'enzyme aCBF5 avait la capacité de modifier la position 55 des ARNt. De plus, sur la base de l'analyse des structures 3D des complexes protéiques aCBF5-aNOP10 et aCBF5-aNOP10-aGAR1 des réseaux d'interaction entre résidus des protéines aCBF5 et aNOP10 ont été proposés (Hamma *et al.*, 2005; Rashid *et al.*, 2006). Nous avons alors entrepris une analyse comparative des activités ARN-guidée et non guidée de aCBF5 et étudié, pour ces deux activités, l'importance de résidus impliqués dans les réseaux d'interaction entre aCBF5 et aNOP10.

2) Comme nous l'avons vu dans l'introduction, les études sur l'activité des ARN:Ψ-synthases ont été effectuées sur les protéines d'*E. coli* et la nature des mécanismes moléculaires à l'origine de la réaction de pseudouridylation n'est toujours pas tranchée. La stabilité en solution des protéines d'archaea nous permettait d'envisager la production de variants de ces protéines qui n'avaient encore jamais été produites chez les enzymes bactériennes telle que TruB. Nous avons ainsi cherché à préciser le rôle d'acides aminés conservés dans le site actif de l'enzyme aCBF5 ainsi que le rôle de la boucle β7/β10 dans la catalyse et dans la réaction de pseudouridylation ainsi que dans la sélection par la particule de la séquence spécifique à modifier.

3) Par ailleurs, la structure 3D de la particule entière a montré que les protéines L7Ae et aNOP10 interagissaient entre elles dans la sRNP. Cette observation était d'autant plus surprenante car l'incubation de ces deux protéines ensemble et en absence de sRNA ne conduisait à la formation d'aucun complexe stable. Nous avons étudié l'importance fonctionnelle de cette interaction formée entre aNOP10 et L7Ae au sein de la sRNP.

4) Enfin, l'identification de la substitution R34W dans la protéine NOP10 humaine conduisant à la forme autosomale récessive de la maladie génétique de la Dyskératose Congénitale, nous a poussé à étudier l'effet de cette substitution sur l'assemblage, la structure et l'activité *in vitro* des sRNP H/ACA d'archaea. Dans ce cadre, nous avons également utilisé les snoRNA H/ACA de levure comme modèle d'étude *in vivo* du rôle joué par cette mutation dans l'activité des snoRNP H/ACA.

MATERIEL & METHODES

I. Matériel

A. Liste des oligonucléotides

Les oligonucléotides utilisés pour les clonages et les mutagenèses dirigées des séquences des protéines sont répertoriés dans le tableau m1, tandis que ceux utilisés pour l'obtention des matrices ADN db pour les réactions de transcription *in vitro* des sRNA et de leur cible sont répertoriés dans le tableau m2.

B. Les souches utilisées

1. Souches de bactéries

Les souches d'*E. coli* **DH5α** (*sup*E44, Δ*lac*U169 (Φ80 *lac*ZΔM15), *hsd*R17, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *rel*A1), **XL1-blue** (*rec*A1, *end*A1, *gyr*A48, *thi*-1, *hsd*R17, *sup*E44, *rel*A1, *lac*F' [*pro*AB, *lacIq*, *lac*ZΔM15, Tn10(Tetr)]) et **TOP 10 F'** (Invitrogen) (F', *end*A1, *rec*A1, *hsd*R17 (rk-,mk+), *sup*E44, *thi*-1, *gyr*A96, *rel*A1, Φ80*lac*ZΔM15, Δ(*lac*ZY A- *arg*F), U169 λ-) ont été utilisées pour la production et la purification de plasmides.

Les souches **BL21 CodonsPlus** (DE3) RIL d'*E. coli* (F-, *omp*T, *hsd*S(rB-mB-), *dcm*+, Tetr, *gal* λ(DE3), *end*A, Hte[*arg*U *ile*Y *leu*W Camr]) ont été utilisées pour la production de toutes les protéines recombinantes.

2. Souches de levure

Les souches JG540 (*Matα, ade2, ade3, leu2, lys2.1, trp1.1, tyr7.1, ura3, can1, rnt1-*Δ) GAL::Nop10p ou GAL::Nhp2p et la souche YDL401 (*Matα, his3, ade3, gal2, trp1.1, ura3*) GAL::Cbf5p nous ont été aimablement fournies par Y. Henry (Henras *et al.*, 1998; Lafontaine *et al.*, 1998).

N°	Séquence	Description
3201	GGATCCATGAAAAGGCTTGGTAAGGT	Amplification aGAR1 *P. abyssi.* Amont, site *Bam* HI
3202	CTCGAGTCAGCCTTTGAACTTTTTCTT	Amplification aGAR1 *P. abyssi.* Aval, site *Xho* I
3203	AGATCTATGAGGTTTAGAATTAGGAAAT	Amplification aNOP10 *P. abyssi* . Amont, site *Bgl* II
3204	CTCGAGTCATTTTTCCTTCCTCCCAA	Amplification aNOP10 *P. abyssi* . Aval, site *Xho* I
3224	GGATCCATGGCGAGAGATGAGGTC	Amplification aCBF5 *P. abyssi.* Amont, site *Bam* HI
3225	CTCGAGTCACCAGAGCTTCGGATA	Amplification aCBF5 *P. abyssi.* Aval, site *Xho* I
3394/95	ACCGCTAACCTTAGGGGCGAGGGTTCCACCATG	Mutagenèse aCBF5 D82A. Amont/aval
3943	GGATCCTTGCCTAAGATATGGATAAAG	Amplification domaine PUA aCBF5. Amont, site *Bam* HI
3944	CTCGAGTCAATGTTCAACAGCTTTCTCC	Amplification domaine Nt aCBF5. Aval, site *Xho* I
4011/12	GGTCGAGGGTTCCACCAGCACCGGCTTTCTCCAAGTT	Mutagenèse aCBF5 H77A. Amont/aval
4018	CTCGAGTTACGCCGGGTATTCAACC	Hybride: aCBF5-PUATruB, fixation fin ORF TruB
4019	CAATGGAGAAAGCTGTTGAAGACTACCCGGTGGTGAAT	Hybride: aCBF5-PUATruB, fixation début PUA TruB
4020	CTCGAGTTACCACTTCCGAGGGTAAGT	Hybride: aCBF5-PUADyskérine, fixation fin PUA
4021	AATGGAGAAAGCTGTTGAATCTCATAAACGGCTGGTTAT	Hybride: aCBF5-PUADyskérine, fixation début PUA Dysk.
4022	CGAAGCTGGACTGTCCAT	Hybride: TruB-PUAaCBF5, fixation aval Domaine Nt TruB
4023	ATGGACAGTCCAGCTTCGCATTTGCCTAAGATATGGATA	Hybride: TruB-PUAaCBF5, fixation début PUA aCBF5
4024	GGATCCATGAGTCGTCCTCGTCGT	Amplification TruB *E. coli.* Amont, site *Bam* HI
4025	TTCAACAGCTTTCTCCATTGG	Amplification TruB *E. coli.* Aval, site *Xho* I
4406/07	TTCTCACCCGAGGATCCAGCTGGTGAATACAGAAGGAG	Mutagenèse aNOP10 Y41A. Amont/aval
4408/09	GGTGAATACAGAAGGAGGGCTAAGAGGGAACTCCTGGG	Mutagenèse aNOP10 L48A. Amont/aval
4410	AGATCTATGAGGTTTGCTATTAGGAAATGCCCA	Mutagenèse aNOP10 R4A. Amont, site *Bgl* II
4411	AGATCTATGAGGTTTAGAGCTAGGAAATGCCCAAAGTGC	Mutagenèse aNOP10 I5A. Amont, site *Bgl* II
4412/13	GCCCACCCACCAAGGTTCGCTCCCGAGGATCCATACGG	Mutagenèse aNOP10 S36A. Amont/aval
4456/57	TTTTTCCTTCCTCCCAGCTCCCAGGAGTTCCCT	Mutagenèse aNOP10 I55A. Amont/aval
4458/59	GAGCGTGTATCTACCTGACTTTGGGCATTTCCT	Mutagenèse aNOP10 C11S. Amont/aval
4460	CTCGAGTCAGTGGGCTACTTTAGTCTTTTC	Fragment aNOP10 Nt1. Aval, site *Xho* I
4461	AGATCTCCACCAAGGTTCTCACCCGAGGAT	Fragment aNOP10 Ct1. Amont, site *Bgl* II
4709/10	GAGGATCCAGCTGGTGAAGCCAGAAGGAGGTTGAAGAGG	Mutagenèse aNOP10 Y41A et Y44A. Amont/aval
4711/12	AAGACTAAAGTAGCCCACGCACCAAGGTTCTCACCCGAG	Mutagenèse aNOP10 P32A. Amont/aval
4796/97	GAGGATCCATACGGTGAAGCCAGAAGGAGGTTGAAGAG	Mutagenèse aNOP10 Y44A. Amont/aval
4798/99	CTCCCGGCTGGAAAGGAGGCCGTTGCGTTGATGCACCTT	Mutagenèse aCBF5 Y110A. Amont/aval
4815	CTCGAGTCATGGTGGGTGGGCTACTTTAGTC	Fragment aNOP10 Nt2. Aval, site *Xho* I
4816	AGATCTAGGTTCTCACCCGAGGATCCAT	Fragment aNOP10 Ct2. Amont, site *Bgl* II
4852/53	TGCCCAAAGTGCGGTAGAGCCACGCTCAAGGAGACCTG	Mutagenèse aNOP10 Y14A. Amont/aval
4854/55	GAAAAGACTAAAGTAGCCGCCCCACCAAGGTTCTCACC	Mutagenèse aNOP10 H31A. Amont/aval
4856/57	GGGGTCATAAACCTTGACGCGCCTCCAGGCCCTACTAG	Mutagenèse aCBF5 K53A. Amont/aval
4858/59	AAGGAGAGATAATTCAGAGG/AGGACGAGGAAGGTCTA	Mutagenèse aCBF5 délétion 140-150. Amont/aval
4860/61	GAAGGAGAGATAATTCAGAGGGCCGCCCTAAGAAGTGC	Mutagenèse aCBF5 P140A,P141A. Amont/aval
4862/63	GAAGGAGAGATAATTCAGGCGCCCCCCCTAAGAAGTGC	Mutagenèse aCBF5 R139A. Amont/aval
4864/65	CACATGGCCGAGCTTAGGGCCACTAGGAGTGGTCCCTT	Mutagenèse aCBF5 R202A. Amont/aval
4866/67	GAGGCTGGGACTTACATTCGGTCATTGATTCATCACATT	Mutagenèse aCBF5 R181A. Amont/aval
4868/69	GGGGTTGAGGCTGGGACTGCCATTAGGTCATTGATTCAT	Mutagenèse aCBF5 Y179A. Amont/aval
5127	AGATCTGCCCACCCACCAAGGTTCTCA	Fragment aNOP10 Ct3. Amont, site *Bgl* II
5129/30	CAGAGGCCCCCCCTAGCAAGTGCCGTGAAAAG	Mutagenèse aCBF5 R143A. Amont/aval
5216	GGATCCATGATAATTGAAAAAGCTGGGAA	Amplification aPUS10 *P. abyssi* . Amont, site *Bam* HI
5217	GTCGACTCAAACCTCTCGAGAATTTCC	Amplification aPUS10 *P. abyssi* . Aval, site *Sal* I.
5218/19	GTTGAGGCTGGGACTTTCATTAGGTCATTGA	Mutagenèse aCBF5 Y179F. Amont/aval
5220/21	CCGGCTGGAAAGGAGTTCGTTGCGTTGATG	Mutagenèse aCBF5 Y110F. Amont/aval
5621/22	GAGGAGATAGTTGCCGGCCTTCCACCGGGCCTGTGAGGA	Mutagenèse L7Ae H70A, L74A. Amont/aval
5627/28	CAGAGGCCCCCCCCTAATAAGTGCCGTGAAA	Mutagenèse aCBF5 R143I. Amont/aval
5629/30	CAGAGGCCCCCCCCTAGAAAGTGCCGTGAAA	Mutagenèse aCBF5 R143E. Amont/aval
5753/74	AGCCCACCCACCATGGTTCTCACCCGA	Mutagenèse aNOP10 R34W. Amont/aval
5755/56	AGCCCACCCACCAGCGTTCTCACCCGAG	Mutagenèse aNOP10 R34A. Amont/aval
5959/60	GGCCTGTGAGGCGAAGGAGATACC	Mutagenèse L7Ae E77A. Amont/aval
5961	GGATCCATGCATTTGATGTACACTTTGGG	Amplification yNop10p *S. cerevisiae* . Amont, site *Bam* HI
5962	GTCGACCTATTGGCCTGGTACCAAACC	Amplification yNop10p *S. cerevisiae* . Aval, site *Sal* I
6067/68	CTCACCCAGCTGCATTTTCTCCAGAC	Mutagenèse yNop10p R34A. Amont/aval
6069/70	CTCACCCAGCTTGGTTTTCTCCAGAC	Mutagenèse yNop10p R34W. Amont/aval
6417	GGATCCATGTTTCTCCAGTATTACCTCAAC	Amplification hNOP10 Humain. Amont, site *Bam* HI
6418	GTCGACTCAGAGGACAGGGCGCGG	Amplification hNOP10 Humain. Aval, site *Sal* I
6419	AGATCTATGGCGGATGCGGAAGTAATT	Amplification hDyskérine Humain. Amont, site *Bgl* II
6420	GTCGACCTACTCAGAAACCAATTCTACC	Amplification hDyskérine Humain. Aval, site *Sal* I
6488	GGATCCATGGGTAAAGACAACAAGGAAC	Amplification yNhp2p *S. cerevisiae* . Amont, site *Bam* HI
6489	CTCGAGTCATAAAGCTTGAACTTCTTTGA	Amplification yNhp2p *S. cerevisiae* . Aval, site *Xho* I
6490	GGATCCATGACCAAAATAAAGGCAGATC	Amplification hNHP2 Humain. Amont, site *Bam* HI
6491	GTCGACTCATAGGGGTAGGGGCAGG	Amplification hNHP2 Humain. Aval, site *Sal* I
6694/95	CCCATCCTGCTGCGTTCTCCCCAG	Mutagenèse hNOP10 R34A. Amont/aval
6696/97	CCCATCCTGCTTGGTTCTCCCCAG	Mutagenèse hNOP10 R34W. Amont/aval

Tableau m1 : Liste des oligonucléotides employés dans le but de cloner ou de muter des séquences codant des protéines.

3. Souches d'archaea

L'ADN génomique de la souche GE5 de *Pyrococcus abyssi* nous a été aimablement fourni par F. Matsunaga et A. Hecker du laboratoire de P. Forterre à Orsay.

C. Les plasmides utilisés

Plasmide pCR2.1 (Invitrogen) **:** Le clonage de fragments d'ADN obtenus par amplification par la technique de PCR a été réalisé dans le plasmide pCR2.1 fourni avec le kit " PCR T/A cloning kit". Le plasmide est commercialisé sous forme linéarisée et porte un résidu thymine protubérant sur chaque extrémité 3'. Il contient des gènes de résistance à l'ampicilline et à la kanamycine. Il possède également le gène *lacZ* codant la β-galactosidase sous le contrôle du promoteur P*lac* et interrompu par la cassette de clonage. Ce plasmide a été utilisé pour vérifier la séquence des produits de PCR qui ont été utilisés pour exprimer tous les sRNA à boîtes H/ACA selon le protocole décrit dans la revue (Charpentier *et al.*, 2007).

Plasmide pGEX-6P-1 (Amersham Pharmacia Biotech) : ce plasmide a été utilisé pour la production de protéines recombinantes L7Ae, aCBF5, aNOP10, aGAR1, aPUS10, TruB et de tous leurs variants protéiques, fusionnés à l'étiquette GST en vue d'une purification sur résine glutation-sépharose selon le protocole décrit dans la revue (Charpentier *et al.*, 2007).

Plasmide pG1 (LGC Standards) : ce plasmide a été utilisé pour l'expression *in vivo* de protéines chez la levure. Les ORF sont insérées entre les sites de restriction *Bam*HI et *Sal*I. Le plasmide possède une origine de réplication « *Ori*C » et « CEN/ARS » pour se répliquer respectivement chez la bactérie et chez la levure. Il

N°	Séquence	Description
3239	TAATACGACTCACTATAGGGTGGCCCCCGGA	Amplification Pab160 *P. abyssi* . Amont, **PT7**
3241	TAATACGACTCACTATAGGGTAATGCCCCTC	Amplification Pab91 *P. abyssi* . Amont, **PT7**
3357/58	TAATACGACTCACTATAGGGTTTAGACCGTCGTGAG	Matrice substrat Pab91, RNA-L. Amont/aval, **PT7**
3391	CCCCTCCCCTCTCACACCTCCGGGATCAGCGACCGGAGG	Variant Pab-91 mt K-turn. Amont
3392	GGTGTCCCCTCCCCGACCGCCCTCCGGTCGCTGATCCCG	Variant Pab-91 mt K-turn. Aval
3397/98	TAATACGACTCACTATAGGGTTTAGACCGCCG	Matrice substrat mt Pab91, RNA-L U/C. Amont/aval, **PT7**
3406	AGGTGTCCCCTCCCCGAC	Amplification Pab91 *P. abyssi* . Aval, ACA+3
3408	AGGACACCCCTCCCCGAC	Amplification Pab91 *P. abyssi* . Aval, mtACA+3
3409	CCCCCTCTCACACCTCCGGGATCAGTGACCGGA	Variant Pab-91 H1-4. Amont
3410	CCCCCGACCGCCCTCCGGTCACTGATCCCGGAG	Variant Pab-91 H1-4. Aval
3438	TCTCTCACGACGGTCTAAACCCTATAGTGAGTCGTAT	Matrice substrat Pab91, RNA-S. Aval, **PT7**
3439	TCTCTCACGGCGGTCTAAACCCTATAGTGAGTCGTAT	Matrice substrat mt Pab91, RNA-U/C. Aval, **PT7**
3440/41	TAATACGACTCACTATAGGGTAGACCGTCGTGAGACAGGT	Matrice substrat Pab91, RNAr. Amont/aval, **PT7**
3557	AGGTGTCCCCCGACCGCCC	Amplification variant Pab91 H1-4, Aval, ACA+3
3669	TAATACGACTCACTATAGGGTAATGCCCCCTCTCAC	Amplification variant Pab91 H1-4, Amont **PT7**
3829/30	TAATACGACTCACTATAGGGTGGAGCGTGCG	Matrice substrat Pab160, RNA-16S922. Amont/aval, **PT7**
4222	CCCCCCTCCCCCCTCTCACACCTCCGGGATCAGTGACCG	Variant Pab-91 H1+4. Amont
4223	CCCCCCTCCCCCCGACCGCCCTCCGGTCACTGATCCCGG	Variant Pab-91 H1+4. Aval
4224	TAATACGACTCACTATAGGGTAATGCCCCCCT	Amplification variant Pab91 H1+4, Amont **PT7**
4225	AGGTGTCCCCCCTCCCCCCGAC	Amplification variant Pab91 H1+4, Aval, ACA+3
4472	TAATACGACTCACTATAGGGACCTGCCGCCCC	Amplification Pab105 *P. abyssi* . Amont, **PT7**
4473	TATAGGCCCGGCTTATCCA	Amplification Pab105 *P. abyssi* . Aval, AUA+3
4474	TAATACGACTCACTATAGGGGCTCGGTCTACC	Amplification Pab35 *P. abyssi* . Amont, **PT7**
4475	TGTAGGCCGGCAACAGA	Amplification Pab35 *P. abyssi* . Aval, ACA+3
4476	TAATACGACTCACTATAGGGCCGACAGCGAG	Amplification Pab35 long *P. abyssi* . Amont, **PT7**
4477/78	CCGCGGGGGAGGAGCGAGCGGCGAGCTATGACTCCCCTT	Mutagenèse Pab160 en Pf9 *P. furiosus* (GAG). Amont/ava
4479	AATTGTGCCCCCGGGGAGCGAA	Amplification Pab160 *P. abyssi* . Aval, ACA+3
4480	AATACAGCCCCCGGGGAGCGAA	Amplification Pab160 *P. abyssi* . Aval, mtACA+3
4596/97	TAATACGACTCACTATAGGGTTGTCCTCCCG	Matrice substrat Pab105 Ψ2795 ARNr 23S.Amont/aval,**PT**
4598/99	TAATACGACTCACTATAGGGATCCTGCCGGA	Matrice substrat Pab105 Ψ27 ARNr 16S. Amont/aval, **PT7**
4642	TGGTGTAGGCCCGGCTTATCCA	Amplification Pab105 *P. abyssi* . Aval, ACA+3
4766	TGTCCGCCCGGCGACGTCGC	Variant Pab105 TB 5'. Aval, ACA+0
4767	TAATACGACTCACTATAGGGCCCGGTCTCCGC	Variant Pab105 TB 3'. Amont, **PT7**
4817	TAATACGACTCACTATAGGGTATGCCCCCGCAAGCG	Amplification Pab40 *P. abyssi* . Amont, **PT7**
4818	ATCTGTAGCCCGCCCGTAACACC	Amplification Pab40 *P. abyssi* . Aval, ACA+3
4819	TAATACGACTCACTATAGGGGATGGGCCCGGCTCCC	Amplification Pab21 *P. abyssi* . Amont, **PT7**
4820	ACTGTAGGCCCGGCCGGAACCCC	Amplification Pab21 *P. abyssi* . Aval, ACA+2
4821/22	TAATACGACTCACTATAGGGAAGGAATTGGC	Matrice substrat Pab21 Ψ891 ARNr 16S. Amont/aval, **PT7**
4880	AATTGGGCCCCCGGGGAG	Amplification Pab160 *P. abyssi* . Aval, CCA+3
4891	TTGGCGCCCGGGCCG	Réverse transcription sur ARNt Asp *P. abyssi* . Aval
4981/82	ATGCCCCTCCCCTCTCACCICCTCCGGGATCAGTGACC	Mutagenèse variant Pab91CU bulge. Amont/aval

Tableau m2 : Liste des oligonucléotides employés dans le but de cloner ou de muter des séquences codant des sRNA ou des ARN substrats.

contient le gène *TRP1* qui permet de sélectionner les levures transformées par auxotrophie et un gène de résistance à l'ampicilline Amp^R.

D. Milieux de culture, tampons et solutions d'usage courant

Milieu LB (Luria Bertani) : peptone pancréatique animale 5 g/l ; peptone trypsique de caséine 5 g/l ; extrait de levure 5 g/l ; NaCl 10 g/l. Le pH est ajusté à 7,5 avec NaOH 10 N. Le milieu de culture est ensuite autoclavé 20 min à 121°C sous une pression de 1 bar.

Milieu solide LB : l'agar est ajouté à raison de 17 g par litre de milieu de culture LB avant stérilisation à l'autoclave.

Milieu LBA/LBK : l'ampicilline ou la kanamycine (100 mg/ml) est ajoutée après autoclave dans le milieu de culture LB à une concentration finale de 100 µg/ml.

Milieu sélectif YNB (yeast nitrogen base) : Base azotée de levure sans acide aminé (10 g/l) ; galactose ou glucose (20 g/l). Un mélange d'acides aminés ne contenant pas les acides aminés permettant la sélection par auxotrophie des souches de levure transformées est ensuite ajouté (20 g/l de chaque acide aminé).

TE : Tris-HCl 10 mM, pH 7,4 ; EDTA 1 mM.

Solution de polyacrylamide dénaturante : Acrylamide/Bisacrylamide (23,75/1,25) 25 % (p/v) ; TBE 1X ; urée 8 M.

Solution de polyacrylamide native : Acrylamide/Bisacrylamide (38/2) 40 % (p/v) ; TBE 1X.

Solution de bleu de dépôt agarose (6X) : EDTA 125 mM ; SDS 2 % (v/v) ; glycérol 25 % (v/v) ; bleu de bromophénol 0,12 % (p/v).

Solution de bleu de coloration de gel de protéine : Ethanol 20% (v/v) ; Acide Acétique 10% (v/v) ; 2 g/l Bleu de Coomassie R250.

N°	Séquence	Description
5070	GGGCCCGGCTCTCACACCTCCGGGATCAGTGACCGGAGG	Variant Pab91 H1 Pab21. Amont
5071	AGGCCCGGCGACCGCCCTCCGGTCACTGATCCCGGAGGT	Variant Pab91 H1 Pab21. Aval
5072	TAATACGACTCACTATAGGGTAATGGGGCCCGGCTC	Amplification Pab91 H1 Pab21. Amont, **PT7**
5073	AGGTGTAGGCCCGGCGACCGC	Amplification Pab91 H1 Pab21. Aval, ACA+3
5074	CCCCTCCCCTCCCGCCCTCTCCGGGGAATCGTGAACCGG	Variant Pab21 H1 Pab91. Amont
5075	CCCCTCCCCCGGAACCCCCGGTTCACGATTCCCCGGAGA	Variant Pab21 H1 Pab91. Aval
5076	TAATACGACTCACTATAGGGGATCCCCTCCCCTCCC	Amplification Pab21 H1 Pab91. Amont, **PT7**
5077	ACTGTCCCCTCCCCCGGAACC	Amplification Pab21 H1 Pab91. Aval, ACA+3
5205	TGTCGCCCGGGCCGGGATTTGAA	Amplification ARNt Asp *P. abyssi* . Aval, ACA
5206	CGCCCGGGCCGGGATTTGAAC	Amplification ARNt Asp *P. abyssi* . Aval, ΔCCA
5226	**TAATACGACTCACTATAG**CCCGGGTG	Amplification ARNt Asp *P. abyssi* . Amont, **PT7**
5227	TGGCGCCCGGGCCGGGATTTGAACC	Amplification ARNt Asp *P. abyssi* . Aval, CCA
5623	GGGCCCGGCTCCCGCCCTCTCCCCCGGGGGGGAATCGTGAA	Variant Pab21 H2+4. Amont
5624	AGGCCCGGCCGGAACCCCCCCCGGGGTTCACGATTCCCCC	Variant Pab21 H2+4. Aval
5625	GGGCCCGGCTCCCGCCCTCTCCCCCCCGGGGGGGGAATCG	Variant Pab21 H2+8. Amont
5626	AGGCCCGGCCGGAACCCCCCCCCCGGGGGGTTCACGATTC	Variant Pab21 H2+8. Aval
5891	TAATACGACTCACTATAGGGAAAGCCCAGG	Amplification Pab19 *P. abyssi* . Amont, **PT7**
5892	TTCTATACCCAGGGGAGGAG	Amplification Pab19 *P. abyssi* . Aval, AUA+3
5893/94	TAATACGACTCACTATAGGGCCGAGAGGAGGIGCAT	Matrice substrat Pab19 Ψ1017 ARNr 16S. Amont/aval,
5895/96	TAATACGACTCACTATAGGGCCGAGAGGAGGIGIATGGCC	Matrice substrat Pab19 C1019U
5897/98	TAATACGACTCACTATAGGGCCGAGAGGAGGIGATGGCC	Matrice substrat Pab19 ΔC1019
6071	CGCCCGCCACGAGGGTTTCCCGGCCGACAGGGCGTGCCG	Reconstitution TB Hsp3. Amont
6072	CGCCCGCAACACTAACGACCGGCCTCATCCGTGGGCGGC	Reconstitution TB Hsp3. Aval
6073	TAATACGACTCACTATAGGGCCCGCCCGCCACGAGG	Amplification Hsp3. Amont, **PT7**
6074	CGGTGTCGCCCGCAACACTAACGAC	Amplification Hsp3. Aval, ACA+3
6075/76	TAATACGACTCACTATAGGGTAACIAIGACC	Matrice subtrat Hsp3 Ψ2016/2018. Amont/aval, **PT7**
6077/78	TAATACGACTCACTATAGGGTAACIAGACCC	Matrice subtrat Hsp3 Ψ2016Δ2018. Amont/aval, **PT7**
6079/80	TAATACGACTCACTATAGGGTAGACCGICIGTGAGA	Matrice subtrat Pab91 RNAr 5'-UNU-3'. Amont/aval, **PT7**
6425/26	TAATACGACTCACTATAGGGTAGACCGIAGTGAGAC	Matrice substrat Pab91 RNAr 5'-UN-3' UA. Amont/aval,
6427/28	TAATACGACTCACTATAGGGTAGACCGIGGTGAGAC	Matrice substrat Pab91 RNAr 5'-UN-3' UG.Amont/aval,
6429/30	TAATACGACTCACTATAGGGTAGACCGIIGTGAGAC	Matrice substrat Pab91 RNAr 5'-UN-3' UU. Amont/aval,
6431/32	TAATACGACTCACTATAGGGTAGACCGICAGTGAGA	Matrice substrat Pab91 RNAr 5'-UNA-3'. Amont/aval, **PT7**
6433/34	TAATACGACTCACTATAGGGTAGACCGICGGTGAGA	Matrice substrat Pab91 RNAr 5'-UNG-3'. Amont/aval, **PT7**
6435/36	TAATACGACTCACTATAGGGTAGACCAICGTGAGAC	Matrice substrat Pab91 RNAr 5'-AUN-3'. Amont/aval, **PT7**
6437/38	TAATACGACTCACTATAGGGTAGACCCICGTGAGAC	Matrice substrat Pab91 RNAr 5'-GUN-3'. Amont/aval, **PT7**
6439/40	TAATACGACTCACTATAGGGTAGACCIICGTGAGAC	Matrice substrat Pab91 RNAr 5'-UUN-3'. Amont/aval, **PT7**
6447	GGTACCTTGCCCTGTGCCTCGCTCG	Amplification snR36 *S. cerevisiae* . Amont, site *Kpn* I
6698	ACTGGAGGCCCGGCCGGAAC	Amplification Pab21. Aval, CCA+2
6699	ACACAAGGCCCGGCCGGAAC	Amplification Pab21. Aval, mtACA+2
6700	AGGTGGCCCCTCCCCGACCG	Amplification Pab91. Aval, CCA+3
6967	CTCGAGTGATATGAGACGTTTCGATGACAAT	Amplification snR36 mtSH2 *S. cerevisiae* . Aval, site *Xho* I
6968/69	CAATTAGACTTCTAAGAGCCACGAGGATATC	Mutagenèse SH1 TB3' snR36 Ψ2875
6970/71	CAATTAGACTTCTAAGAGCACGAGGATATCG	Mut. SH1 TB3' snR36 Ψ2875+3nts ARNr 25S. Amont/aval
6973	CGCTTACCGAATTCTGCTTCG	Réverse transcription ARNr 25S 2894 à 2914 levure.Aval

Tableau m2. (suite)

Solution de décoloration de gel de protéine : Ethanol 20% (v/v); Acide Acétique 10% (v/v).

Tampon de gel de migration de gel de protéine : Glycine 14,25 g/l ; Tris 3 g/l ; SDS 1 g/l.

Tampon RNase P1 2x : Acétate d'ammonium 100 mM, pH 5,2.

Tampon de migration RII pour TLC : Isopropanol 10 ml ; Sulfate d'ammonium 300 g ; Na2HPO4/NaH2PO4 pH 6,8 0,1 M final (soit 127,5 ml Na2HPO4 mélangé à 123,5 ml NaH2PO4) ; H2O qsp 500 ml.

II. Méthodes

A. Méthodes sur les protéines, les ARN et mesure de l'activité ARN:Ψ-synthase

Le chapitre d'ouvrage suivant présente différentes techniques basées sur la production des ARN et des protéines ainsi que les méthodes d'assemblage des particules et de la mesure de leur activité.

Chapitre d'ouvrage :

Reconstitution of Archaeal H/ACA sRNPs and test of their activity.

Bruno CHARPENTIER, **Jean-Baptiste FOURMANN** and Christiane BRANLANT

Methods in Enzymology, 2007 volume 425 : 389-405.

CHAPTER SIXTEEN

Reconstitution of Archaeal H/ACA sRNPs and Test of their Activity

Bruno Charpentier, Jean-Baptiste Fourmann, *and* Christiane Branlant

Contents

1. Introduction 390
2. Overview of the Issues that can be Addressed by the Method 393
3. The Basic Principles of the Experiments 393
4. Materials and Reagents 394
   4.1. Chemicals 394
   4.2. General reagents 394
   4.3. Equipment 394
   4.4. Enzymes 395
   4.5. Buffers 395
5. Methods 396
   5.1. Production/purification of the archaeal proteins aCBF5, aNOP10, L7Ae, and aGAR1 396
   5.2. Production of the box H/ACA guide sRNAs 396
   5.3. Production of the RNA substrate 398
   5.4. Assembly of box H/ACA sRNPs and their characterization by EMSA 399
   5.5. Analysis of the formation of an RNA duplex between the guide sRNA and the RNA substrate 400
   5.6. Measurement of the RNA:Ψ-synthase activity of the reconstituted sRNPs 400
   5.7. Quantification of the data 401
6. General Comments 403
   6.1. Choice of the size of the RNA substrate 403
   6.2. No activity is detected 403
References 404

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, Nancy Université, Faculté des Sciences et Techniques, Vandoeuvre-les-Nancy, France

Methods in Enzymology, Volume 425
ISSN 0076-6879, DOI: 10.1016/S0076-6879(07)25016-5
© 2007 Elsevier Inc.
All rights reserved.

Abstract

Conditions for the reconstitution of archaeal sRNPs active in RNA-guided post-transcriptional modification of RNAs (2′-*O*-methylation or pseudouridylation) were recently developed. This has opened a vast field of research on structure–function relationships of the sRNP components. We present here an efficient method for *in vitro* reconstitution of H/ACA sRNPs with an active RNA-guided pseudouridylation activity. They are assembled from an *in vitro* transcribed H/ACA sRNA with recombinant L7Ae, aCBF5, aNOP10, and aGAR1 proteins. The protocol to test the activity of the assembled particles by the method of the nearest neighbor is also described. The combination of *in vitro* assembly of H/ACA sRNPs together with time course analysis of pseudouridine formation in the target RNA is useful for structure–function analysis of H/ACA sRNPs and also to identify the target sites of H/ACA sRNAs discovered by computer analysis of archaeal genomes. Furthermore, this efficient *in vitro* pseudouridylation system can be used for generation of pseudouridine residues at defined positions in RNAs.

1. Introduction

Isomerization of uridine into pseudouridine (5-ribosyluracil, Ψ) is one of the most frequent posttranscriptional modifications encountered in cellular RNAs (Rozenski *et al.*, 1999). The functional importance of Ψ in RNA folding (Davis, 1995; Newby and Greenbaum, 2002), ribosome synthesis (Badis *et al.*, 2003; King *et al.*, 2003; Mengel-Jorgensen *et al.*, 2006; Meroueh *et al.*, 2000; Sumita *et al.*, 2005), pre-mRNA splicing (Yang *et al.*, 2005; Yu *et al.*, 1998), and the specific recognition of anticodon by tRNAs (Grosjean *et al.*, 1998) is now illustrated by several experimental data.

In both Eukarya and Archaea, U to Ψ conversions in RNAs, which are termed pseudouridylations, can be catalyzed by a single protein with a specific RNA:Ψ-synthase activity or by an H/ACA ribonucleoprotein complex (RNP). These RNPs contain at least four proteins associated with a small RNA. The RNA component is named H/ACA snoRNA or H/ACA scaRNA in Eukarya, depending on its nuclear localization (nucleolus or Cajal bodies, respectively), and H/ACA sRNA in Archaea (Bachellerie *et al.*, 2002; Henras *et al.*, 2004; Kiss, 2002; Omer *et al.*, 2003). All these RNA components contain one or several elements folded into a characteristic stem-loop structure, and they carry conserved ANA trinucleotide sequences. The RNA component of the RNPs functions as a guide RNA that base pairs with the target RNA. Through this interaction, the guide RNA defines the position to be modified by the RNA:Ψ-synthase activity, that is carried by one of its protein partners (Fig. 16.1). A high degree of conservation exists between the eukaryal and the archaeal

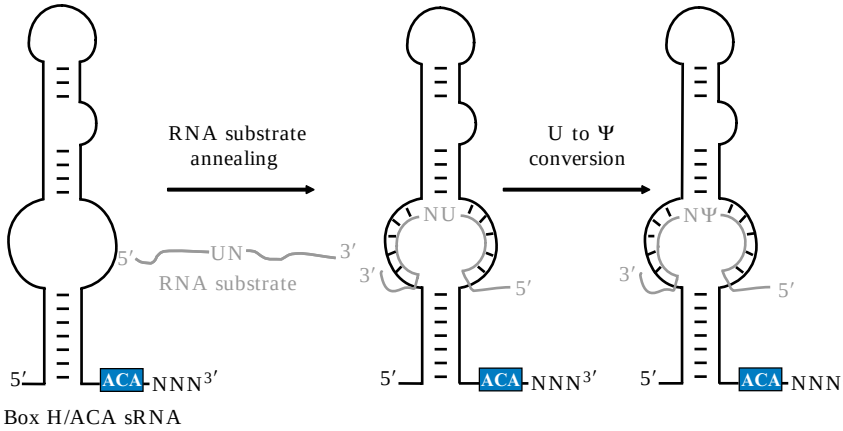


Figure 16.1 Generic scheme explaining the RNA-guided function of H/ACA sRNAs. Two segments (the antisense elements) of the internal loop present in the irregular stem-loop structure of the sRNA pseudouridylation pocket base pair with the sRNA substrate. A dinucleotide 5′-UN-3′ is single stranded, and the U residue is converted in a Ψ residue by the RNA:Ψ-synthase activity carried by the aCBF5 protein component of the sRNP.

H/ACA RNPs. They all contain a protein that belongs to the TruB family of RNA:Ψ-synthases, NAP57/CBF5/Dyskerin in eukaryal H/ACA snoRNPs/scaRNPs, and its counterpart aCBF5 in Archaea. They contain three other proteins, NOP10, GAR1, and NHP2, in Eukarya and their homologs, aNOP10, aGAR1, and the L7Ae, in Archaea. Protein L7Ae is, in addition, a ribosomal protein. In Eukarya, as well as in Archaea, the characteristic stem-loop structures of the H/ACA guide RNAs are essential for their function. Their internal loop is complementary to the target RNA (Fig. 16.1) and is called the pseudouridylation pocket (Ganot *et al.*, 1997; for review, Bachellerie *et al.*, 2002). A pair of residues remains unpaired in the RNA duplex. The targeted residue is the 5′ residue in this pair. H/ACA snoRNAs most frequently contain two stem-loop structures that guide modification at two distinct positions in rRNAs that are linked by a hinge sequence carrying the ANANNA sequence (the H box). The ACA sequence is located downstream from the second stem-loop structure. It is always found three nucleotides upstream of the snoRNA 3′ end. In Archaea, the sRNA structure is more variable: it can include one, two, or even three stem-loop structures containing a pseudouridylation pocket. Here also, each of these stem-loop structures is flanked by an ANA trinucleotide (more generally ACA).

In addition to the importance of Ψ formation, the H/ACA RNP architecture is important by itself (Meier, 2005). Indeed, some of the box H/ACA snoRNPs are essential for the maturation of the pre-rRNA

(Bachellerie *et al.*, 2002). Furthermore, the vertebrate telomerase RNA, which is required for the synthesis of chromosome extremities, contains an H/ACA snoRNA domain (Mitchell *et al.*, 1999). This domain binds the H/ACA snoRNP proteins (Dez *et al.*, 2001; Pogacic *et al.*, 2000). Mutations in the gene encoding the telomerase RNA or the NAP57/CBF5 enzyme lead to bone marrow failure dyskeratosis congenita (Armanios *et al.*, 2005; Heiss *et al.*, 1998; Vulliamy *et al.*, 2001). This stresses the necessity to investigate the structure of the H/ACA RNP domains. Unfortunately, reconstitution of eukaryal snoRNPs is difficult to achieve mainly because of the low solubility of the eukaryal snoRNP proteins produced as recombinant proteins (Wang and Meier, 2004).

However, active archaeal H/ACA sRNPs can be reconstituted by use of purified recombinant proteins and an *in vitro* transcribed sRNAs (Baker *et al.*, 2005; Charpentier *et al.*, 2005). This opens the possibility of studying structure–function relationships for the various components of these RNPs. Information was already gained by use of this reconstitution system. It was shown that the ACA trinucleotide of archaea H/ACA sRNAs is required for H/ACA sRNPs assembly (Baker *et al.*, 2005; Charpentier *et al.*, 2005). In particular, it is needed for association of protein aCBF5 with the guide sRNA. Protein aCBF5 is required for protein aNOP10 recruitment in the sRNP (Baker *et al.*, 2005; Charpentier *et al.*, 2005). This minimal particle containing only two of the sRNP proteins has some RNA-guided RNA:Ψ-synthase activity (Charpentier *et al.*, 2005). However, the kinetics of the pseudouridylation is markedly increased by addition of the L7Ae and aGAR1 recombinant proteins (Charpentier *et al.*, 2005). When the full set of recombinant proteins is used, the targeted U residue can be totally converted into a Ψ residue in the substrate RNA within 10 min.

The 3D structures of components and, more recently, of a fully assembled sRNP (a H/ACA sRNA and the four proteins) were very recently established (Hamma *et al.*, 2005; Khanna *et al.*, 2006; Li and Ye, 2006; Manival *et al.*, 2006; Rashid *et al.*, 2006); several hypotheses formulated on the basis of these 3D structures can now be tested by use of the *in vitro* reconstitution system. Testing the effect of mutations in the H/ACA sRNA and/or the proteins will provide a deep understanding of the catalytic mechanism of H/ACA sRNPs and also of the relative roles of the various components of the H/ACA sRNPs in the assembly of catalytically active particles. Information will also be gained on the molecular basis for target specificity of H/ACA sRNPs. Noticeably also, the *in vitro* assembly system offers the possibility of defining the specificity of H/ACA sRNAs identified by computational analyses of archaeal genomes (see Chapter 15 by Muller *et al.*). Indeed, computer analyses now allow the identification of putative H/ACA sRNA genes in archaeal genomes with a high degree of reliability (see Chapter 15 by Muller *et al.*). However, it is often more difficult to

predict their target site with accuracy. The *in vitro* assembly system described in this chapter can be used to test the activity of the H/ACA sRNA on its putative targets after production of the putative H/ACA sRNA and its putative target RNA sequences by *in vitro* transcription and of the H/ACA sRNP proteins as soluble recombinant proteins in *E. coli*.

2. Overview of the Issues that can be Addressed by the Method

- Structure–function analysis of the protein and RNA components of H/ACA sRNPs
- Investigation on the binding of a target RNA (designated as the RNA substrate) to the H/ACA sRNPs
- Investigation of the catalytic mechanism of the H/ACA sRNPs
- Identification of the RNA substrate for H/ACA sRNAs isolated by biochemical or by computational approaches

In addition, because RNA-directed Ψ formation is very efficient in this *in vitro* system, an important use of this method can be the production of RNAs with Ψ residues at defined positions. This application will require the design of sRNA guides adapted to the RNA substrate. This will be facilitated by the data accumulated in the course of the structure–function analysis of H/ACA sRNPs.

3. The Basic Principles of the Experiments

RNPs formed on incubation of an sRNA with the proteins are characterized by electrophoresis mobility shift assays (EMSA). For instance, we characterized various sRNPs assembled on incubation of an H/ACA sRNA (Pab91) and different sets of proteins (aCBF5, L7Ae, aNOP10). Their coding sequences were PCR amplified from the archaeon *P. abyssi* genomic DNA (Charpentier *et al.*, 2005). Examples of EMSA experiments performed with this experimental system are shown in Fig. 16.2B and C, as a reference for scientists who want to compare their data with the ones of these reconstitution experiments.

The RNA:Ψ-synthase activity of the assembled sRNPs is measured by the nearest neighbor approach by use of thin-layer chromatography (TLC) for fractionation of the 3′-labeled mononucleotides. Residue Ψp has a distinct mobility compared with Up in the TLC system used (Keith, 1995) (examples are shown in Fig.16.5).

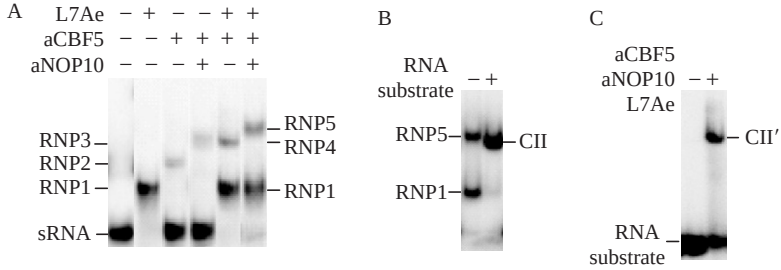


Figure 16.2 Examples of sRNP assembly experiments and RNA substrate binding by use of components of the Pab91 H/ACA sRNP from *P. abyssi*. (A) sRNPs were assembled by incubation of components of the *P. abyssi* archaeon: the radiolabeled guide sRNA Pab91 was incubated with either L7Ae, aCBF5, L7Ae-aCBF5, aCBF5-aNOP10, or L7Ae-aCBF5-aNOP10 proteins to obtain the RNP1, RNP2, RNP3, RNP4, and RNP5 complexes, respectively. The complexes are fractionated by electrophoresis mobility shift analysis (EMSA). (B) In the presence of an unlabeled RNA fragment which sequence is complementary to the pseudouridylation pocket of the sRNA Pab91, a unique CII complex is formed. (C) A unique CII′ complex is also obtained on incubation of a radiolabeled RNA substrate with the unlabeled sRNA and the three proteins L7Ae, aCBF5, and aNOP10.

4. Materials and Reagents

4.1. Chemicals

The four ribonucleotides triphosphate ATP, GTP, CTP, UTP (100 m*M* each, Fermentas).

[γ-^{32}P]ATP (10 mCi/ml, 3000 Ci/mmol; Amersham Biosciences) and [α-^{32}P]NTP (10 or 20 mCi/ml, 800 Ci/mmol; Amersham Biosciences)

4.2. General reagents

RNase-free water obtained from an ionization system (e.g., Milli-Q), sterilized and filtered through a 0.22-μm membrane filter (Millex, Millipore)

Glycerol 87% (Merck)

A phenol/chloroform solution

Na-Acetate (0.3 *M*)

RNasin (Amersham Biosciences)

Total yeast tRNAs (0.5 μg/μl, Roche)

A 40% Acrylamide/Bis-acrylamide solution (38:2)

4.3. Equipment

16 × 15 cm plates for vertical electrophoresis and 1- and 1.5-mm thick spacers

Plastic sheets precoated by a thin layer (0.1 mm) of cellulose (Macherey-Nagel)

The acrylamide gels used for electrophoresis mobility shift assays (EMSA) and the cellulose sheets used for 1D and 2D TLC were visualized and quantified on a PhosphorImager (Typhoon 9410, Amersham Biosciences) with the ImageQuant software.

4.4. Enzymes

Commercial T7 RNA polymerase (20 U/μl, Ambion) or purified from *E. coli* BL21 strain (Zawadzki and Gross, 1991)

Calf intestinal alkaline phosphatase (CIP, MBI-Fermentas, 0.1 U/μl) and its 10× reaction buffer (100 m*M* $MgCl_2$, 100 m*M* Tris-HCl, pH 7.5)

T4 Polynucleotide kinase (PNK, MBI-Fermentas 10 U/μl) and its 10× reaction buffer (100 m*M* $MgCl_2$, 50 m*M* DTT, 1 m*M* spermidine, 1 m*M* EDTA, 500 m*M* Tris-HCl, pH 7.6)

RNAse-free DNase I (10 U/μl, Amersham Biosciences)

RNase T2 (Invitrogen, 0.1 U/μl)

4.5. Buffers

20× TBE: 1780 m*M* borate, 40 m*M* EDTA, 1780 m*M* Tris-borate, pH 8.3

3× transcription buffer (TrB): 60 m*M* $MgCl_2$, 400 m*M* HEPES-KOH, pH 7.6

10× Ambion transcription buffer: 100 m*M* DTT, 20 m*M* spermidine, 60 m*M* $MgCl_2$, 400 m*M* Tris-HCl, pH 7.9

10× buffer D: 150 m*M* KCl, 1.5 m*M* $MgCl_2$, 0.2 m*M* EDTA, 20 m*M* HEPES, pH 7.9

Denaturing gel loading solution (i.e., formamide dye): 20 m*M* EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol in deionized formamide (Merck)

Elution buffer (EB): 1 m*M* EDTA, 300 m*M* NaCl, 1% SDS, 10 m*M* Tris-HCl, pH 7.5

Complex separation loading dye (Cs): 40% glycerol, 0.03% bromophenol blue, 0.03% xylene cyanol, 20 m*M* HEPES-KOH, pH 7.9

EMSA running buffer: 115 ml glycerol 87%, 50 ml 20× TBE, total volume 2 liters

2× RNase T2 buffer: 100 m*M* NH_4-Ac, pH 4.6

Solvent N1 for 1D TLC: 50 ml isobutyrate, 1.1 ml ammoniac 25%, H_2O to 80 ml

Solvent N2 for the second dimension in 2D TLC: 70 ml isopropanol, 15 ml HCl 37%, H_2O to 100 ml

5. Methods

5.1. Production/purification of the archaeal proteins aCBF5, aNOP10, L7Ae, and aGAR1

The ORFs encoding the archaeal proteins aCBF5, aNOP10, L7Ae, and aGAR1 are PCR amplified from the genomic DNA. As an example, we will describe the application of the methods to the *P. abyssi* archaeon. However, they are of general use and can be applied to other archaeons. The ORFs with GenBank accession nos. CAB49444 (aCBF5), CAB49761 (aNOP10), C75109 (L7Ae), and CAB49230 (aGAR1) were PCR amplified from the *P. abyssi* GE5 strain genomic DNA (Charpentier *et al.*, 2005).

The PCR-amplified DNA fragment can be cloned into an *E. coli* expression vector. We used plasmid pGEX-6P-1 (Pharmacia) to produce GST fusion proteins. M. Terns and colleagues produced the proteins from *P. furiosus* as His-tagged proteins (Baker *et al.*, 2005). The recombinant GST-aCBF5, GST-aNOP10, GST-L7Ae, and GST-aGAR1 proteins are routinely produced in *E. coli* BL21 CodonPlus cells (Novagen) and purified under native conditions in the PBS buffer (140 m*M* NaCl, 2.7 m*M* KCl, 10 m*M* Na_2HPO_4, 1.8 m*M* KH_2PO_4, pH 7.3) with glutathione-Sepharose 4B (Pharmacia) as previously described (Charron *et al.*, 2004). The GST part of the fusion proteins are directly cleaved on beads with 80 U of PreScission protease (Pharmacia) per ml of glutathione-Sepharose bead suspension. Cleavage is performed overnight at 4° followed by a 15 min incubation at 65°. The contaminant *E. coli* proteins are precipitated in these conditions, and they are eliminated by a 20-min centrifugation at 16,000*g*. The quality of the protein preparation is assessed by SDS-PAGE and Coomassie staining. Absorbance at 280 nm ($A_{280\ nm}$) is measured for each sample, and the protein concentration is calculated by use of their theoretical ε_{280nm}. The proteins are stored at −80° at a 4.5 μM concentration in 150 m*M* NaCl, 1 m*M* EDTA, 1 m*M* DTT, glycerol 10%, 50 m*M* Tris-HCl, pH 7.

5.2. Production of the box H/ACA guide sRNAs

The guide sRNAs are synthesized by T7 RNA polymerase with, as the template, PCR-amplified DNA fragments containing the H/ACA sRNA coding sequence under the control of a T7 RNA polymerase promoter (Fig. 16.3). These fragments are obtained by PCR amplification by use of an archaeal genomic DNA as the template and as the forward primer, an oligonucleotide carrying at its 5′ extremity a T7 RNA polymerase promoter sequence (5′ TAATACGACTCACTATAGGG 3′). Alternately, the DNA sequence of an sRNA can be obtained by use of two complementary oligonucleotides as the template for PCR amplification. The sequence of

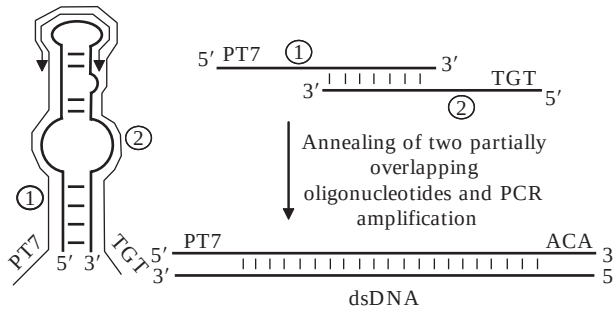


Figure 16.3 **Strategy used to obtain the DNA fragment for T7 *in vitro* transcription of a H/ACA sRNA with a single hairpin structure (see text for details).**

the T7 RNA polymerase promoter is included in the left template oligonucleotide (named 1 in Fig. 16.3), or it is introduced by a forward primer in a second PCR step. This second approach for production of the template RNA sequence opens the possibility to introduce mutations within the RNA sequence. Classical protocols are used for the PCR amplifications. Purification of the amplified fragments is not necessary for subsequent transcription with the T7 RNA polymerase.

The transcription reaction is performed in a 30-μl solution obtained by mixing 9 μl of H_2O, 3 μl of 10× transcription buffer, 10 μl of a mix of the four NTPs (12.5 m*M* each), 6 μl of the PCR reaction, 1 μl of Rnasin, and 1 μl of T7 RNA polymerase (20 U). When T7 RNA polymerase purified from *E. coli* is used, the reaction is performed in 1× TrB buffer. Incubation is performed at 37° for at least 4 h; routinely incubation is done overnight. Then, the DNA template is hydrolyzed by 1 μl of DNase I (10 U) at 37° for 1 h. After addition of 15 μl of the denaturing gel loading solution and heating at 95° for 2 min, the products of the transcription reaction are fractionated on a 1-mm thick, 8% denaturing polyacrylamide-8*M* urea gel prepared in 1× TBE buffer. RNAs are visualized by UV shadowing, the gel slices are cut with a razor blade, and the RNA molecules are eluted from the gel piece by an overnight incubation in 300 μl of elution buffer EB at 4°. The elution solution is collected and transferred into a microcentrifuge tube. The RNA transcripts are precipitated at −80° after addition of 1 ml of ethanol. The precipitated RNAs are collected by microcentrifugation at top speed for 10 min at 4°, washed with ethanol 70%, and dried by leaving the tube open on the bench. The dry pellet is dissolved in 20 μl of H_2O. The RNA concentration is quantified by measurement of the $A_{260\ nm}$. Milli-Q H_2O is added to get a 200 ng/μl concentration and RNA is stored at −80°.

Binding of the proteins on the sRNAs is analyzed by EMSA (see later) by use of a radiolabeled sRNA. To this end, the unlabeled sRNA transcripts

prepared as previously described are labeled at their 5′ end after dephosphorylation of their 5′ triphosphate extremities by calf intestinal phosphatase (CIP). Dephosphorylated sRNAs are prepared by the following procedure: 5 μl (1 μg) of sRNA is treated by 1 μl of CIP enzyme (0.1 U) in 10 μl of the 1× CIP buffer for 1 h at 37°. The reaction is stopped by addition of 150 μl of Na-Ac (0.3 *M*), followed by a phenol-chloroform extraction, and the dephosphorylated RNAs are ethanol precipitated. The dried pellets are dissolved in 10 μl of Milli-Q H_2O and stored at −80°.

For 5′-end labeling, 2.5 μl (250 ng) of the dephosphorylated RNA preparation are mixed with 1 μl of 10× PNK buffer, 1 μl of [γ-^{32}P]ATP (10 mCi/ml, 3000 Ci/mmol), 1 μl of T4 PNK (10 U/μl), and 4.5 μl H_2O (total volume 10 μl), and incubated for 45 min at 37°. The reaction is stopped by addition of 10 μl of denaturing gel loading solution. After heating at 96° for 2 min, the sample is loaded on a 1-mm thick 8% polyacrylamide-8 *M* urea denaturing gel. The radiolabeled band is cut, the 5′-end labeled sRNA is eluted overnight at 4° with the procedure described previously, followed by ethanol precipitation. The pellet is dissolved in Milli-Q H_2O to get 10,000 cpm/μl, and the labeled RNA is stored at −20°.

5.3. Production of the RNA substrate

The unlabeled RNA substrate, complementary to the sRNA pseudouridylation pocket (Fig. 16.1), is synthesized by T7 RNA polymerase with the same protocol as for sRNA production. For RNA:Ψ-synthase activity measurements, the RNA substrate needs to be labeled. This labeling is obtained by incorporation of a radiolabeled NTP during the transcription reaction. The choice of the labeled nucleotide depends on the sequence of the RNA substrate. To demonstrate the occurrence of the U to Ψ conversion in the substrate by the nearest-neighbor method on the basis of RNase T2 digestion, a labeled phosphorus atom should be present 3′ to the targeted U residue (Fig. 16.4). Thus, the identity of the residue 3′ to the targeted U residue defines the identity of the NTP to be used in the labeling reaction. For instance, in the case of the *P. abyssi* Pab91 sRNA, the needed labeling was obtained by incorporation of [α-^{32}P]CTP in the course of transcription (Fig. 16.4). The following conditions are used for transcription: the reaction is performed by mixing 5 μl of 3× TrB buffer, 0.5 μl of RNasin, 3 μl of a mix of the ATP, UTP, GTP nucleotides (20 m*M* each), 1 μl of CTP (2 m*M*), 1 μl of [α-^{32}P]CTP (800 Ci/mmol), 1.5 μl of BSA (2 μg/μl), 1 μl of the PCR reaction, and 1.5 μl of the T7 RNA polymerase. Incubation is performed at 37° for 2 h, and the DNA is digested with 10 U of DNase I for 1 h. The labeled transcripts are fractionated by electrophoresis on the same type of polyacrylamide denaturing gel as described previously and visualized by autoradiography. The radiolabeled bands are cut, and the labeled RNAs

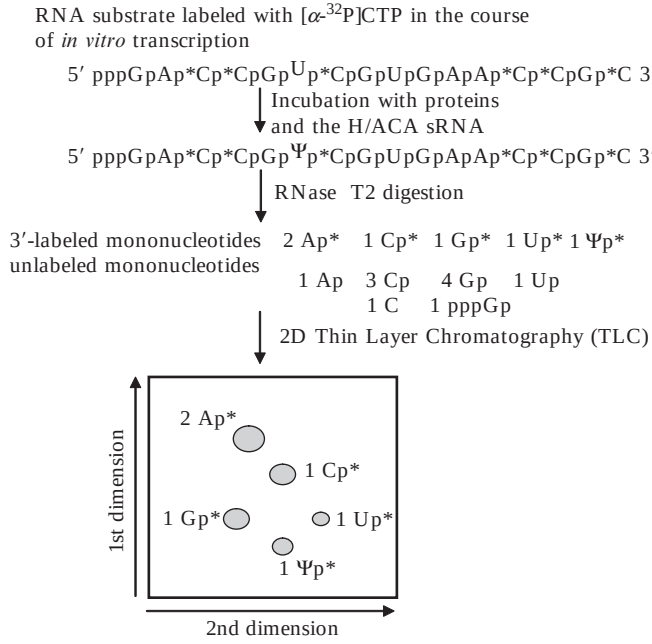


Figure 16.4 **Detection of Ψ residues on a 2D TLC by the technique of the nearest neighbors. The radiolabeled phosphates of the RNA substrate sequence are marked by an asterisk.**

are eluted overnight as described previously. After ethanol precipitation, RNAs are diluted in Milli-Q H_2O to get 10,000 cpm/μl.

5.4. Assembly of box H/ACA sRNPs and their characterization by EMSA

To avoid nonspecific binding, a mixture of yeast tRNAs is used as competitor. Thus, the ^{32}P-labeled sRNA (10,000 cpm, approximately 50 fmoles) together with 500 ng of yeast tRNA mixture are added in 1× D buffer in a final volume of 3.5 μl. The RNAs are denatured by a 5-min incubation at 65°. After a brief centrifugation, they are renatured by slow cooling at room temperature. The mixture of the recombinant proteins aCBF5, aNOP10, L7Ae, and aGAR1 at a 900 n*M* concentration each is prepared in 1× D buffer. RNPs are formed by adding 1 μl of the protein mixture to the 3.5-μl RNA solution (each recombinant protein is at a 200 n*M* final concentration). After incubation for 10 min at 65°, 4 μl of complex separation loading dye (Cs) is added. The formation of RNA–protein complexes is tested by electrophoresis at room temperature in 1.5-mm thick 6% nondenaturing polyacrylamide gels. For 40 ml of gel solution, 1 ml of 20× TBE is mixed

with 2.3 ml of 87% glycerol, 6 ml of 40% acrylamide/bisacrylamide (38/2), and Milli-Q H_2O is added to 40 ml. Electrophoresis is carried out at 6.5 V/cm for 2 h.

5.5. Analysis of the formation of an RNA duplex between the guide sRNA and the RNA substrate

The association of the sRNP with the target RNA can be tested by two alternative procedures. In a first procedure, the guide sRNA is labeled, and the RNA substrate is unlabeled. This unlabeled substrate is added to the RNP assembly reaction described in the preceding section. The quantity of RNA substrate required to observe a complete shift of the electrophoretic mobility of the assembled RNP is variable, depending on the RNA guide and the substrate used. For the Pab91 sRNP, 15 ng (~2 pmol) of target RNA is used. Incubation and analysis by EMSA are carried out as described in the preceding section (Fig. 16.2B).

In the second procedure, the labeled RNA is the RNA substrate. The RNP assembly is performed with an unlabeled sRNA, and thus binding of its labeled target RNA is monitored (Fig. 16.2C). In that case, the RNP is assembled by combining in 1× D buffer, 200 ng of unlabeled sRNA (~10 pmol), 500 ng of tRNAs, and 200 n*M* of each protein in the presence of the labeled RNA substrate (10,000 cpm, ~150 fmoles). The RNP/RNA substrate complex is resolved as described previously by gel electrophoresis in nondenaturing conditions.

5.6. Measurement of the RNA:Ψ-synthase activity of the reconstituted sRNPs

The RNA:Ψ-synthase activities of the reconstituted sRNPs are tested by the nearest-neighbor approach (Fig. 16.4). The substrate RNA, uniformly labeled by the strategy explained in the section "Production of the RNA Substrate," is incubated with the modification machinery (the guide sRNA and the proteins). The incubation is performed at 65°, which corresponds to the optimal temperature for the Ψ-synthase activity of reconstituted archaeal H/ACA sRNPs (data not shown). The RNA substrate is next totally digested by an RNase to get labeled 3′-mononucleotides that can be fractionated by 2D or 1D TLC.

The unlabeled sRNA (10 pmoles, ~2.2 m*M*) and ^{32}P-labeled target RNA (150 fmoles, ~30 n*M*) are mixed at room temperature in buffer D and treated as described for EMSA assays. The proteins (200 n*M* each) are added at room temperature. The final 4.5-μl solution is incubated at 65°. For time course experiments, the reaction volume is scaled up to 45 μl, by use of the same RNA and protein concentrations, and the reaction is started

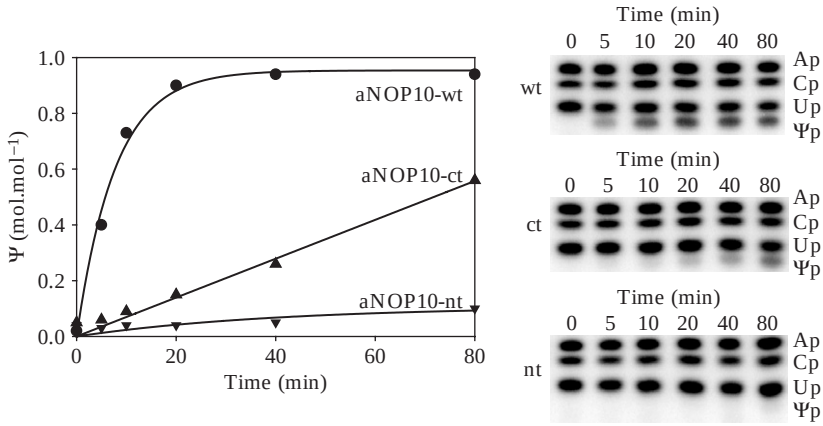


Figure 16.5 **Time course analysis of Ψ formation in the RNA substrate of the Pab91 sRNA. The unlabeled Pab91 sRNA was mixed with the labeled RNA substrate and the four proteins L7Ae, aCBF5, aNOP10, and aGAR1. The amounts of Ψ residues formed were estimated by 1D TLC. They are expressed in moles of Ψ residue formed per moles of RNA substrate. Variants of the aNOP10 protein were tested: either a fragment carrying the N-terminal residues (aNOP10-nt) or the C-terminal residues (aNOP10-ct) was tested.**

by addition of the proteins after incubation of the RNA solution at 65° for 10 min. At time intervals, 4.5-μl aliquots are collected. For each aliquot, the reaction is stopped by the addition of 150 μl of Na-Ac (0.3 *M*), 2 μg tRNAs, which is immediately followed by extraction with 150 μl of phenol/chloroform (1:1). The extracted RNAs are ethanol precipitated.

The dried pellet of the modified RNA targets is dissolved in 8 μl of 1× T2 buffer containing 0.4 U of T2 RNase (0.1 U/μl) for 12 h at 37°. The resulting 3′-mononucleotides (3 μl) are chromatographed on thin-layer cellulose plates as described elsewhere (Keith, 1995). For reactions performed on RNA substrates with no 3′-labeled guanosine, 1D chromatography in solvent N1 is sufficient to fractionate the other labeled 3′-mononucleotides (Fig.16.5). If 3′-labeled G residues are obtained after RNase T2 digestion, a 2D TLC fractionation is required, and migration in the second dimension is performed in solvent N2.

5.7. Quantification of the data

The dried gels of the EMSA experiments and the plates used for TLC are analyzed with a PhosphorImager (Typhoon 9410, Amersham Biosciences). The amounts of radioactivity in the bands of gels and in the spots are estimated with the ImageQuant software.

5.7.1. Quantification of the amount of assembled sRNP

The percentage of RNA in each RNP is calculated, taking into account the percentage of radioactivity in each band on the gel relative to the total amount of radioactivity in the lane. As can be seen in Fig. 16.2A, >90% of the Pab91 sRNA is shifted into a unique RNP1 complex in the presence of protein L7Ae. Incubation of this sRNA with protein aCBF5 leads to formation at a low yield of the RNP2 complex (~7%). The direct binding of aCBF5 is more efficient on other sRNAs with yields of RNP2 formation >50% (Fourmann *et al.*, unpublished results). The Pab91 sRNA is shifted partially into a complex of low electrophoretic mobility (RNP5) on its incubation with the three proteins aCBF5, aNOP10, and L7Ae. The residual RNP1 complex remains the more abundant (>50%). Nevertheless, complex RNP5 can be formed at higher yields (>80%) with other guide sRNAs (Fourmann *et al.*, unpublished results).

In the presence of the unlabeled RNA substrate, a unique complex we named CXII is formed (Fig. 16.2B). We named CII′ the complex detected on incubation of the proteins, the unlabeled guide sRNA, and the radiolabeled RNA substrate (Fig. 16.2C). In various experiments, we found that the amount of this complex was variable with a maximal value of ~40%.

5.7.2. Quantification of the Ψ formed

A calculation table, prepared with the classical Excel software, can be used to determine the number of moles of Ψ formed per mole of RNA substrate. A table with 5 columns (A, B, C, D, and E) and 5 lanes (1, 2, 3, 4, and 5) is built (Fig. 16.6). This table is filled for each assay in the time course analysis of Ψ formation. The number of 3′ end-labeled nucleotides (A, C, G, U, and Ψ), which are expected to be obtained after complete digestion of the RNA substrate by the T2 RNase, are used to fill the cells of the B column. For the example given in Fig. 16.4, the values are of 2 for cell B1, and 1 for the B2, B3, and B4 cells. A value of 1 is used for the expected number of Ψ nucleotide. The values experimentally obtained by quantification of the TLC spots by the ImageQuant software are used to fill the cells in column C: for example, the value of the quantification of the Ap spot is written in cell C2. The D column will give the ratios between the expected nucleotide numbers and the measured radioactivity (Cx/Dx). When incorporation of radioactive residues in the substrate has occurred uniformly, the C1/B1, C2/B2, C3/B3, and C4/B4 values should be very similar. Discrepancies between these values in column D may reflect a bias in the labeling reaction and/or some degradation of the RNA in the course of the experiment. Therefore, it is very important to take these values into consideration.

The number of moles of Ψ per mole of RNA substrate is calculated as the mean value determined by taking into account the B and C values for each nucleotide. To this end, for each determination, the C5 value is

	A	B	C	D	E
	Nucleotides	Number of 3′ labeled nucleotides (1)	Values of the Phosphor Imaging quantification (2)	Relative quantity	Yield of Ψ formation (3)
1	A	B1	C1	C1/B1	E1 = C5 × (B1/C1)
2	C	B2	C2	C2/B2	E2 = C5 × (B2/C2)
3	G	B3	C3	C3/B3	E3 = C5 × (B3/C3)
4	U	B4	C4	C4/B4	E4 = C5 × (B4/C4)
5	Ψ	1	C5		E5 = (E1 + E2 + E3 + E4)/4

(1) Number of 3′ end-labeled nucleotides calculated from the nucleotide sequence of the RNA substrate taking into account the identity of the labeled nucleotide used for the *in vitro* transcription. The number of Ψ written in this cell is 1, based on the assumption that only one U residue is specifically modified within the RNA substrate sequence.

(2) Values of the PhosphorImager quantification, corrected for background.
Values are given in Volume by the Molecular analyst software.

(3) The value in moles of Ψ formed per moles of RNA substrate is obtained in cell E5. If only one Ψ is present in the RNA substrate sequence, the maximal values in this column cannot be higher than ~1 for each labeled nucleotide.

Figure 16.6 **Calculation table in Excel format used for the determination of the amount of Ψ formed.**

multiplied by the theoretical B value for one given nucleotide and divided by the experimental C value measured for this residue. One E value is thus obtained (column E), and then the mean value of the E numbers is calculated in cell E5.

In the conditions used in the experiment displayed in Fig. 16.5, ~90% of the RNA substrate molecules are modified after 20 min of incubation.

6. General Comments

6.1. Choice of the size of the RNA substrate

We tested the effect of the RNA substrate length on the activity of the reconstituted Pab91 H/ACA sRNP. We found that addition of 19nt at the 3′ extremities of the sequences complementary to the antisense element of the guide RNA does not change the amount of Ψ residue formed after a 1-h incubation. However the kinetics of the reaction is slower in the presence of the extensions (Charpentier *et al.*, 2005).

6.2. No activity is detected

The binding of each protein should be tested by EMSA, in particular the binding of proteins L7Ae and aCBF5 that bind the sRNA directly (Fig. 16.2A).

ACKNOWLEDGMENTS

J.-B. F. is a fellow from the French Ministère Délégué à l'Enseignement Supérieur et à la Recherche. The work was supported by the Centre National de la Recherche Scientifique (CNRS), the French Ministère de la Recherche et des Nouvelles Technologies (MRNT), the Actions Concertées Incitatives (ACI) Microbiologie and BCMS of the French MRNT, and the Bioingénierie PRST of the Conseil Régional Lorrain.

REFERENCES

Armanios, M., Chen, J. L., Chang, Y. P., Brodsky, R. A., Hawkins, A., Griffin, C. A., Eshleman, J. R., Cohen, A. R., Chakravarti, A., Hamosh, A., and Greider, C. W. (2005). Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. *Proc. Natl. Acad. Sci. USA* **102,** 15960–15964.

Bachellerie, J. P., Cavaille, J., and Huttenhofer, A. (2002). The expanding snoRNA world. *Biochimie* **84,** 775–790.

Badis, G., Fromont-Racine, M., and Jacquier, A. (2003). A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *RNA* **9,** 771–779.

Baker, D. L., Youssef, O. A., Chastkofsky, M. I., Dy, D. A., Terns, R. M., and Terns, M. P. (2005). RNA-guided RNA modification: Functional organization of the archaeal H/ACA RNP. *Genes Dev.* **19,** 1238–1248.

Charpentier, B., Muller, S., and Branlant, C. (2005). Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res.* **33,** 3133–3144.

Charron, C., Manival, X., Charpentier, B., Branlant, C., and Aubry, A. (2004). Purification, crystallization preliminary X-ray diffraction data of L7Ae sRNP core protein from *Pyrococcus abyssi*. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 122–124.

Davis, D. R. (1995). Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* **23,** 5020–5026.

Dez, C., Henras, A., Faucon, B., Lafontaine, D., Caizergues-Ferrer, M., and Henry, Y. (2001). Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p. *Nucleic Acids Res.* **29,** 598–603.

Ganot, P., Bortolin, M. L., and Kiss, T. (1997). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89,** 799–809.

Hamma, T., Reichow, S. L., Varani, G., and Ferre-D'Amare, A. R. (2005). The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat. Struct. Mol. Biol.* **12,** 1101–1107.

Heiss, N. S., Knight, S. W., Vulliamy, T. J., Klauck, S. M., Wiemann, S., Mason, P. J., Poustka, A., and Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat. Genet.* **19,** 32–38.

Henras, A. K., Dez, C., and Henry, Y. (2004). RNA structure and function in C/D and H/ACA s(no)RNPs. *Curr. Opin. Struct. Biol.* **14,** 335–343.

Keith, G. (1995). Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie* **77,** 142–144.

Khanna, M., Wu, H., Johansson, C., Caizergues-Ferrer, M., and Feigon, J. (2006). Structural study of the H/ACA snoRNP components Nop10p and the 3′ hairpin of U65 snoRNA. *RNA* **12,** 40–52.

King, T. H., Liu, B., McCully, R. R., and Fournier, M. J. (2003). Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol. Cell* **11,** 425–435.

Kiss, T. (2002). Small nucleolar RNAs: An abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109,** 145–148.

Li, L., and Ye, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443,** 302–307.

Manival, X., Charron, C., Fourmann, J. B., Godard, F., Charpentier, B., and Branlant, C. (2006). Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res.* **34,** 826–839.

Meier, U. T. (2005). The many facets of H/ACA ribonucleoproteins. *Chromosoma* **114,** 1–14.

Mengel-Jorgensen, J., Jensen, S. S., Rasmussen, A., Poehlsgaard, J., Iversen, J. J., and Kirpekar, F. (2006). Modifications in *Thermus thermophilus* 23 S ribosomal RNA are centered in regions of RNA-RNA contact. *J. Biol. Chem.* **281,** 22108–22117.

Meroueh, M., Grohar, P. J., Qiu, J., SantaLucia, J., Jr., Scaringe, S. A., and Chow, C. S. (2000). Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of *Escherichia coli* 23S rRNA. *Nucleic Acids Res.* **28,** 2075–2083.

Mitchell, J. R., Cheng, J., and Collins, K. (1999). A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3′ end. *Mol. Cell. Biol.* **19,** 567–576.

Newby, M. I., and Greenbaum, N. L. (2002). Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. *Proc. Natl. Acad. Sci. USA* **99,** 12697–12702.

Omer, A. D., Ziesche, S., Decatur, W. A., Fournier, M. J., and Dennis, P. P. (2003). RNA-modifying machines in archaea. *Mol. Microbiol.* **48,** 617–629.

Pogacic, V., Dragon, F., and Filipowicz, W. (2000). Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. *Mol. Cell Biol.* **20,** 9028–9040.

Rashid, R., Liang, B., Baker, D. L., Youssef, O. A., He, Y., Phipps, K., Terns, R. M., Terns, M. P., and Li, H. (2006). Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol. Cell* **21,** 249–260.

Rozenski, J., Crain, P. F., and McCloskey, J. A. (1999). The RNA Modification Database: 1999 update. *Nucleic Acids Res.* **27,** 196–197.

Sumita, M., Desaulniers, J. P., Chang, Y. C., Chui, H. M., Clos, L., 2nd, and Chow, C. S. (2005). Effects of nucleotide substitution and modification on the stability and structure of helix 69 from 28S rRNA. *RNA* **11,** 1420–1429.

Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P. J., and Dokal, I. (2001). The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* **413,** 432–435.

Wang, C., and Meier, U. T. (2004). Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.* **23,** 1857–1867.

Yang, C., McPheeters, D. S., and Yu, Y. T. (2005). Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280,** 6655–6662.

Yu, Y. T., Shu, M. D., and Steitz, J. A. (1998). Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J.* **17,** 5783–5795.

Zawadzki, V., and Gross, H. J. (1991). Rapid and simple purification of T7 RNA polymerase. *Nucleic Acids Res.* **19,** 1948.

B. Séparation des nucléotides modifiés ou non par chromatographie sur couche mince en deux dimensions

La méthode de digestion des ARN et du fractionnement des nucléotides est présentée dans le chapitre d'ouvrage ci-dessus et la carte de migration sur plaque TLC des différentes espèces est présentée dans la figure m1 (Keith, 1995).

C. Analyse de la stabilité des ARN par Northern Blot

Extraction des ARN totaux

Les cellules transformées sont cultivées sur milieu YNB contenant du galactose, puis l'expression du gène d'intérêt est réprimée en milieu YNB supplémenté en glucose pendant 24h. Les cellules sont centrifugées à 4°C 10 min à 2500 rpm. Pour l'extraction d'ARN, les cellules sont reprises dans 200 µl de tampon d'extraction (Tris/HCl 50 mM, pH 7,5 ; NaCl 100 mM ; EDTA 10 mM), et 300 µl de billes de verre sont ajoutés. Chaque échantillon est agité fortement au vortex dans un microtube 2 fois 1 min, en étant placé dans la glace 2 min entre chaque agitation. Puis, 200 µl de tampon d'extraction, 50 µl de SDS 10% et 400 µl de phénol/ chloroforme sont ajoutés au broyat. Le mélange est agité fortement au vortex pendant 1 min et incubé 10 min à 65°C. Une seconde extraction phénol/ chloroforme.est réalisée. Les ARN sont précipités à l'acétate de sodium 0,3M en présence d'éthanol 100%, le précipité est lavé, séché et repris dans de l'eau stérile. La quantité d'ARN est déterminée par la mesure de l'absorbance à 260 nm.

Séparation des ARN

5 µg d'ARN totaux sont dénaturés 2 min à 96°C puis refroidis dans la glace et ensuite déposés sur un gel de polyacrylamide dénaturant 6% en présence d'un

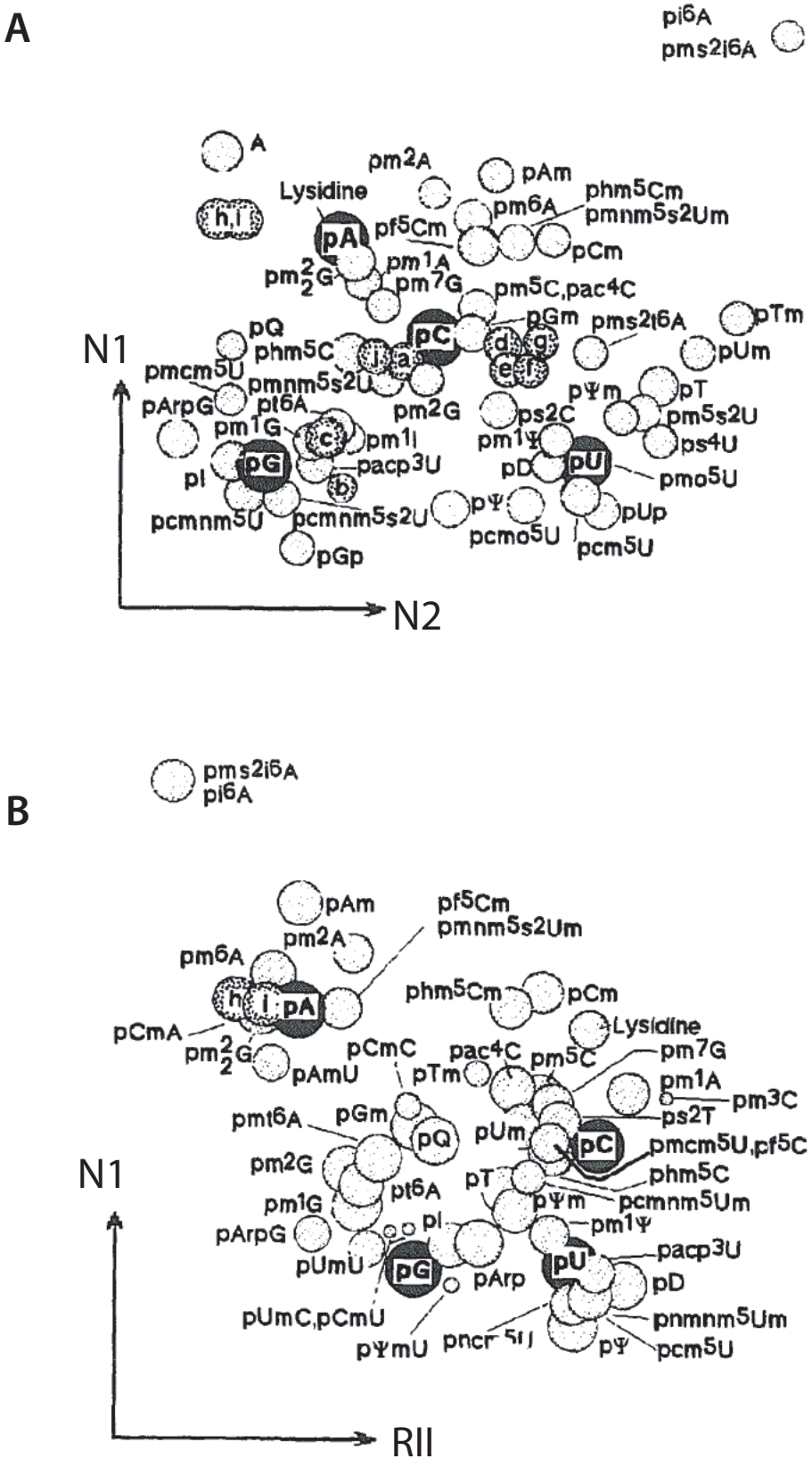


Figure m1 : Positionnement relatif en chromatographie bi-dimentionnelle sur couche mince des spots correspondant aux nucléotides majeurs et nucléotides modifiés de l'ARN. (Keith, 1995).

Deux distributions différentes sont observées selon l'utilisation de tampon N2 (**A.**) ou RII (**B.**) pour la seconde dimension de séparation par chromatographie.

volume de bleu de formamide. La migration a lieu à 30 mA à température ambiante durant 1h dans le tampon TBE 0,5X stérile.

Transfert sur membrane de nylon Hybond N+ (Electroblotting)

Après la migration, les ARN sont transférés par un champ électrique (Biorad Trans Blot SD SemiDry Transfer Cell) sur une membrane de nylon (Hybond N+, 0,45 µm, Pharmacia) en tampon TBE 0,5x pendant 30 min à une intensité de 3 mA/cm2. Les ARN sont fixés de façon covalente à la membrane par pontage photochimique aux UV pendant 5 min.

Marquage de la sonde

Les oligonucléotides sont radiomarqués selon le protocole suivant : [γ-^{32}P]ATP (GE Healthcare, 10 mCi/ml, 3000 Ci/mmol) 1 µl ; oligonucléotide (15 pmol/µl soit environ 100 ng/µl) 1 µl. Tampon de phosphorylation 10X (Tris/HCl 0,5 M, pH 7,6 ; MgCl2 100 mM ; DTT 50 mM ; Spermidine 1 mM ; EDTA 1 mM) 1 µl ; T4 polynucléotide kinase (MBI Fermentas, 10 U/µl) 1 µl ; H2O stérile qsp 10 µl. Incuber 1h à 37°C. La T4 PNK est inactivée en ajoutant 10 µl d'eau stérile et en incubant 10 min à 65°C.

Pré-hybridation

La membrane est pré-hybridée 1 heure à 37°C sous agitation dans un tube à hybridation en présence de 30 ml de tampon de pré-hybridation 6x composé de 30 ml de SSPE 20x (NaCl 175,3 g ; Na2HPO4, 2H2O 27,6 g ; EDTA 7,4 g ; H2O qsp 1 l, ajusté à pH 7,4) ; 5 ml de réactif de Denhardt 100X (Ficoll 5 g ; PolyVinylPyrrolidone 5g ; BSA en poudre 5 g ; H2O qsp 250 ml) ; 5 ml de SDS 10% ; H2O qsp 100 ml ; préchauffé à 37°C.

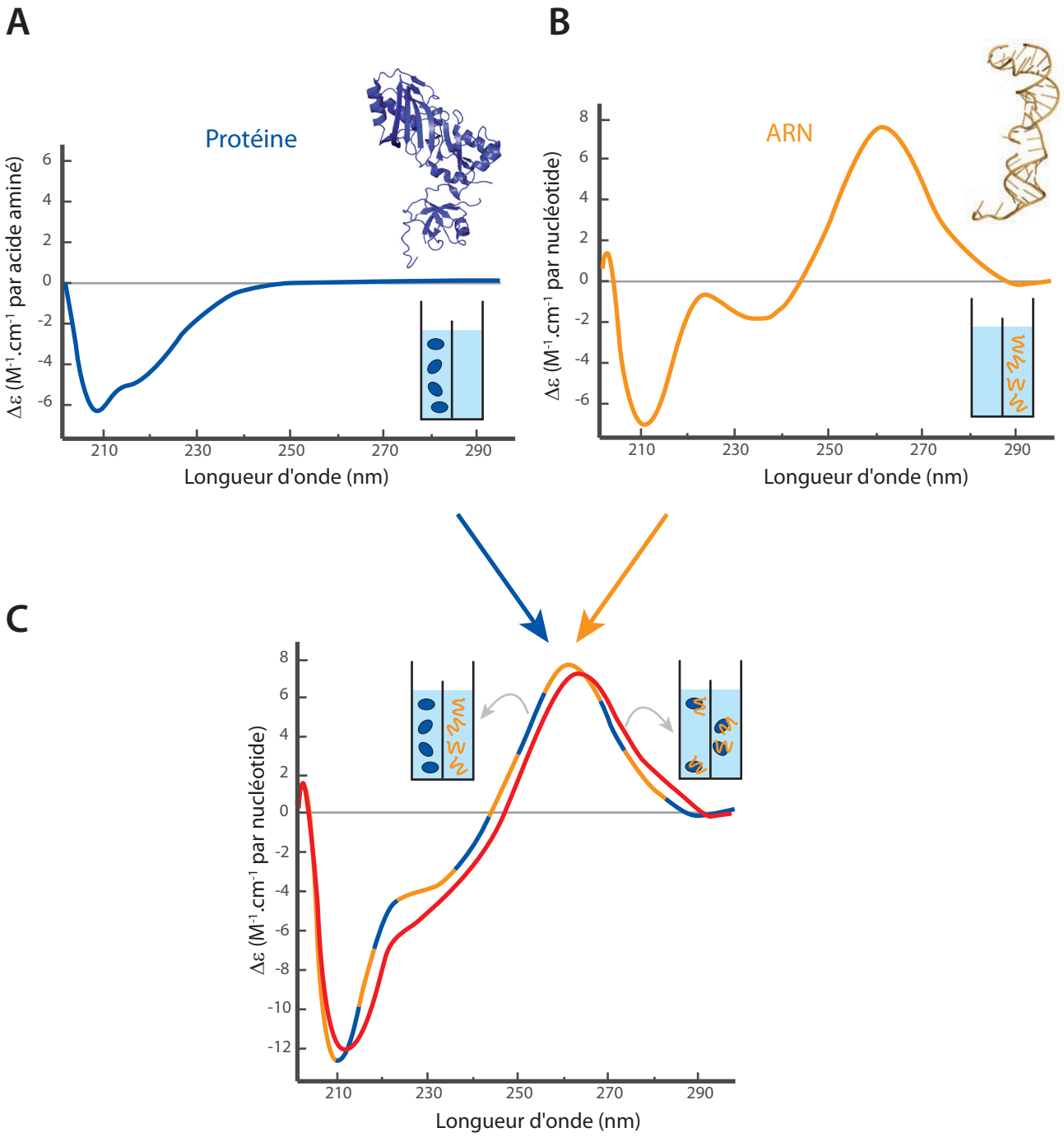


Figure m2 : Principe de l'étude des interactions ARN-protéines par la technique de dichroïsme circulaire.

A. Spectre caractéristique de la protéine aCBF5 de *P. abyssi* (en bleu) qui est placée dans un compartiment de la cuve et du tampon dans l'autre compartiment.

B. Spectre caractéristique de la molécule d'acide nucléique Pab21 (en orange) qui est placée dans un compartiment de la cuve et du tampon dans l'autre compartiment.

C. Etude par dichroïsme circulaire de l'interaction entre l'ARN et la protéine. Une mesure est effectuée lorsque les deux composés sont placés dans la cuve mais dans des compartiments différents (spectre en orange et bleu). Après mélange un nouveau spectre est mesuré (en rouge). Les différences observées et mesurées sont alors caractéristiques de la formation du complexe ARN-protéine. Dans la région comprise entre 240 et 300 nm, les modifications du spectre sont spécifiquement attribuées à un changement de la structure globale de l'ARN Pab21, provoqué par la fixation de la protéine aCBF5.

Hybridation

L'oligonucléotide marqué en 5' est dénaturé par incubation 2 min à 96°C puis refroidissement dans la glace. Puis, 1 ml de tampon de pré-hybridation est ajouté à l'oligonucléotide marqué et le mélange est filtré à l'aide d'une seringue et d'un filtre de 2 µm (de façon à réduire le bruit de fond sur la membrane) puis est ajouté au tampon dans le tube à hybridation. L'hybridation est réalisée sur la nuit à 37°C sous agitation dans un four à hybridation. La membrane est alors lavée 3 à 4 fois pendant 5 min avec 15 ml de tampon de lavage SSPE 6x chauffé à 42°C et placée sous agitation. La membrane humide est emballée dans un plastique hermétique avant d'être exposée en cassette d'un système de type PhosphorImager (Typhoon, Amersham Biosciences).

Déshybridation et réhybridation

Il est possible de réutiliser la membrane en éliminant la sonde fixée par deux lavages de la membrane avec une solution de SDS 0,1% portée à ébullition, sous agitation pendant 10 min. La membrane peut à nouveau être utilisée pour une hybridation avec une autre sonde en suivant le protocole décrit précédemment (pré-hybridation, hybridation, lavages et exposition).

D. Analyse de la structure secondaire des macromolécules biologiques par dichroïsme circulaire

Le dichroïsme circulaire est une technique qui permet d'appréhender la structure des macromolécules en solution. Les spectres de dichroïsme circulaire ont été enregistrés grâce à un appareil appelé dichrographe (Jobin-Yvon Mark VI) à une vitesse de mesure de 2 nm/s, dans une cuve en quartz à deux compartiments de 0,875 cm de trajet optique, à température ambiante maintenue à 21°C et avec une concentration finale en protéine et ARN de 2 µM soit à stœchiométrie 1 : 1, dans le tampon Tris-HCl 10 mM (pH 7,5), NaCl 150 mM.

Les spectres caractéristiques d'une protéine (Fig. m2 A) ou d'un ARN (Fig. m2 B) sont enregistrés en plaçant la molécule d'intérêt dans un compartiment et le tampon dans l'autre compartiment. Pour réaliser le spectre d'un complexe ARN-protéine, la protéine est placée dans le premier compartiment et l'ARN dans le second compartiment (Fig. m2 A et B). Les spectres dichroïques sont enregistrés avant et après mélange des deux solutions disposées dans la cuve (Fig. m2 C). La mesure du spectre du tampon est réalisée avant chaque expérience et les valeurs sont ensuite soustraites des données recueillies pour les échantillons. Deux spectres sont effectués pour chaque mesure et sont ensuite moyennés pour augmenter le rapport signal/bruit. Les résultats sont présentés en tant que valeurs normalisées de $\Delta\varepsilon$ sur la base de la masse moléculaire moyenne d'un résidu acide aminé (110 Da) pour les protéines ou d'un nucléotide moyen (330 Da) pour les interactions ARN-protéines. La sensibilité de l'appareil utilisé est de l'ordre de $\delta(DA) = 10^{-6}$.

E. Numéros d'accès à la PDB

Les codes des structures 3D des ARN:Ψ-synthases déposés dans la PDB (Protein Data Bank) sont indiqués dans le tableau i3 de la partie introduction.

RESULTATS & DISCUSSION

I. Recherche de déterminants ARN qui contribuent à l'activité ou à l'assemblage des particules sRNP à boîtes H/ACA

A. Déterminants présents dans les sRNA et dans leurs ARN substrats

1. Présentation du travail

Au laboratoire, Sébastien Muller et Fabrice Leclerc avaient développé une approche bioinformatique performante pour la recherche de sRNA à boîtes H/ACA dans les génomes d'archaea. S. Muller a ainsi identifié 7 sRNAs H/ACA potentiels dans le génome de *P. abyssi* et a confirmé leur expression par une analyse par Northern Blot. Il a prédit les séquences cibles de ces sRNA au sein des séquences des ARNr 16S et 23S et il a vérifié expérimentalement la modification aux positions prédites par la méthode de RT/CMCT basée sur l'emploi de l'agent modifiant CMCT suivi d'un traitement alcalin et de la détection des résidus Ψ par transcription réverse. La démonstration de l'identité du sRNA responsable de chacune de ces modifications a été effectuée après reconstitution *in vitro* de chacune des sRNP à boîtes H/ACA suivi par un test de leur activité de modification de fragments d'ARN produits *in vitro* et portant les sites prédits dans les ARNr. J'ai participé au clonage de certains sRNA, à la reconstitution *in vitro* des sRNP et j'ai effectué des tests de leur activité. L'ensemble des données obtenues nous ont apporté des informations sur plusieurs caractéristiques générales des sRNP à boîtes H/ACA telles que les contraintes fonctionnelles qui dirigent leur activité au sein des particules :

- La distance entre la séquence ACA et le résidu U cible de la réaction doit être de 14 ou 15 nucléotides.

- L'appariement formé avec le brin 3' de la poche pseudouridylation est plus important que celui formé avec le brin 5'.

- La distance entre le motif en K-turn qui fixe L7Ae et la partie supérieure de la poche pseudouridylation doit être de 5 ou 6 résidus.

- Le dinucléotide simple brin 5'-UN-3' qui se forme lors de l'interaction entre les deux brins de la poche pseudouridylation et l'ARN substrat n'est pas strictement requis

pour l'activité. La modification *in vitro* d'un substrat peut aussi avoir lieu en présence d'un trinucléotide non apparié 5'-UNN-3'.

De plus cette étude a permis d'identifier deux positions modifiées (Ψ2585 et Ψ2603) dans l'ARNr 23S de *P. abyssi* qui ne sont guidées par aucun des 7 sRNA H/ACA. La position U2603 peut être modifiée *in vitro* par le mélange de protéines aCBF5, aNOP10 et aGAR1 en absence de tout ARN guide. Ainsi, ces résultats indiquent que aCBF5 a la capacité de modifier une séquence de l'ARNr par une activité non guidée. Un deuxième exemple de cette activité sera observé dans le cas de modification de la position U55 des ARNt (article n°3, (Muller *et al.*, 2007b)) La modification à la position U2585 reste énigmatique : soit un sRNA H/ACA qu'il reste à identifier permet de guider la pseudouridylation à cette position, soit une ARN:Ψ-synthase agit à cette position sans l'aide d'ARN guide et nos résultats indiquent qu'il ne s'agirait pas de aCBF5.

Ces résultats sont rassemblés dans un article publié dans la revue *Nucleic Acids Research* et je suis co-auteur :

Publication n°1 :

Combined *in silico* and experimental identification of the *Pyrococcus abyssi* H/ACA sRNAs and their target sites in ribosomal RNAs.

Sébastien MULLER, Fabrice LECLERC, Isabelle BEHM-ANSMANT, **Jean-Baptiste FOURMANN**, Bruno CHARPENTIER and Christiane BRANLANT

Nucleic Acids Research, 2008 volume 36(8) : 5459-5475.

Published online 27 February 2008 *Nucleic Acids Research, 2008, Vol. 36, No. 8* ***2459–2475***
doi:10.1093/nar/gkn077

Combined *in silico* and experimental identification of the *Pyrococcus abyssi* H/ACA sRNAs and their target sites in ribosomal RNAs

Sébastien Muller, Fabrice Leclerc, Isabelle Behm-Ansmant, Jean-Baptiste Fourmann, Bruno Charpentier and Christiane Branlant*

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP, Nancy Université, Faculté des Sciences et Techniques, 54506 Vandoeuvre-lès-Nancy, France

Received January 15, 2008; Revised February 7, 2008; Accepted February 8, 2008

ABSTRACT

How far do H/ACA sRNPs contribute to rRNA pseudouridylation in Archaea was still an open question. Hence here, by computational search in three *Pyrococcus* genomes, we identified seven H/ACA sRNAs and predicted their target sites in rRNAs. In parallel, we experimentally identified 17 Ψ residues in *P. abyssi* rRNAs. By *in vitro* reconstitution of H/ACA sRNPs, we assigned 15 out of the 17 Ψ residues to the 7 identified H/ACA sRNAs: one H/ACA motif can guide up to three distinct pseudouridylations. Interestingly, by using a 23S rRNA fragment as the substrate, one of the two remaining Ψ residues could be formed *in vitro* by the aCBF5/aNOP10/aGAR1 complex without guide sRNA. Our results shed light on structural constraints in archaeal H/ACA sRNPs: the length of helix H2 is of 5 or 6 bps, the distance between the ANA motif and the targeted U residue is of 14 or 15 nts, and the stability of the interaction formed by the substrate rRNA and the 3′-guide sequence is more important than that formed with the 5′-guide sequence. Surprisingly, we showed that a sRNA–rRNA interaction with the targeted uridine in a single-stranded 5′-UNN-3′ trinucleotide instead of the canonical 5′-UN-3′ dinucleotide is functional.

INTRODUCTION

Conversion of uridine into pseudouridine (Ψ) residues is one of the most abundant post-transcriptional modifications of tRNAs, rRNAs and UsnRNAs (1). Compared to U residues, Ψ residues can form an additional hydrogen bond at the N1–H position. Furthermore, the carbon–carbon link between the ribose and the base limits the flexibility of the ribose backbone of Ψ residues, which favours and stabilizes base-pair interactions (2). A role of Ψ residues in stabilization of the tRNA 3D structure has also been well documented (3), and the functional importance of Ψ residues in U2 snRNA for the activity of splicing complexes was demonstrated (4–7). Eukaryal rRNAs contain a large number of Ψ residues compared to bacterial rRNAs and they are concentrated in rRNA regions expected to play important functional roles, in particular in domains IV and V, which are directly involved in the peptidyl transferase activity (8–10). Taken individually, pseudouridylations in rRNAs are not essential for cell growth. However, the complete abolition of Ψ formation in rRNAs is deleterious for ribosome assembly and activity (10–12). Recent data suggest their possible involvement in: (i) subunit interaction (12–14), (ii) the translocation step (14,15), (iii) translation termination (12,16) and (iv) folding of 23S rRNA in an active form at the peptidyl transferase centre (PTC) (11,12,17). The large number of pseudouridylation sites in eukaryal rRNAs compared to bacterial rRNAs is explained by the use of different catalysts: stand-alone enzymes carrying both the RNA recognition capability and the catalytic activity are used in bacteria (10,18), whereas U to Ψ conversions are catalyzed by H/ACA snoRNPs in eukarya (19,20). The H/ACA snoRNPs contain four proteins and an H/ACA snoRNA that defines the targeted U residue by base-pair interaction with the rRNA. Recent data revealed a similar RNA-guided system in archaea (21–24). Most of the eukaryal H/ACA snoRNAs contain two characteristic stem-loop structures, with an internal loop (pseudouridylation pocket), that is complementary to two nucleotide stretches bordering the targeted U residue. Each of the stem-loops is flanked by a conserved motif (H and

*To whom correspondence should be addressed. Tel: +33 3 83 68 43 03; Fax: +33 3 83 68 43 07; Email: christiane.branlant@maem.uhp-nancy.fr

© 2008 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ACA, respectively). In the following part of this manuscript, one stem-loop structure containing a pseudouridylation pocket flanked by one of the conserved 3′ sequence will be denoted an H/ACA motif. In the RNA duplex formed between the H/ACA motif and the substrate rRNA sequence, the targeted U residue and its 3′ nucleotide are both single-stranded. The archaeal sRNA counterparts of the snoRNAs have more diverse architectures. They are composed of one, two or three H/ACA motifs (21–26). In addition, each archaeal H/ACA stem-loop structure contains a K-turn or a K-loop which binds protein L7Ae (22). The conserved 3′-flanking sequence is an ANA trinucleotide (most frequently an ACA trinucleotide).

As successful reconstitutions of active H/ACA sRNPs were achieved using an *in vitro* transcribed H/ACA sRNA and the recombinant archaeal H/ACA sRNP proteins (23,24), strong progresses were recently made in the understanding of the H/ACA sRNP structure and function. Like the eukaryal H/ACA snoRNPs, the archaeal H/ACA sRNPs contain four proteins. Protein aCBF5 is the RNA: Ψ-synthase, aNOP10 is required for aCBF5 activity, L7Ae binds the K-turn or K-loop motif, and aGAR1 may facilitate the H/ACA sRNP turnover (22–24). Both L7Ae and aGAR1 strongly reinforce the sRNP activity (23). The crystal structures of H/ACA sRNP protein complexes and of an entire H/ACA sRNP were recently solved at high resolution (27–31). The ANA sequence is needed for binding of aCBF5 to the guide RNA (28,29,31). In the crystal structure, aCBF5 also interacts with the pseudouridylation pocket, helix H1 and helix H2 of the H/ACA motif (31). Recently a 3D structure obtained for an H/ACA sRNP devoid of the L7Ae protein and bound to its RNA substrate, revealed contacts between aCBF5 and the target rRNA–H/ACA sRNA duplex (30). However, little is known on the structural constraints required for formation of an active rRNA target–H/ACA sRNA interaction. The two NMR structures established for a complex formed between an H/ACA stem-loop structure and a small target RNA revealed the capability of the two RNAs to interact together, in the absence of protein, at the high concentration used for NMR analysis (32,33). In these structures, the heterologous helix formed by interaction of the RNA target with the 3′-guide element of the sRNA is stacked on helix H1, while the heterologous helix formed by interaction of the RNA target with the 5′-guide element of the sRNA is stacked on helix H2.

At present, in contrast to this extensive knowledge on H/ACA sRNPs, little is known on the number and location of Ψ residues in archaeal rRNAs. Their presence has only been investigated in *Halobacterium halobium* (34), *Haloarcula marismortui* (35,36), *Sulfolobus acidocaldarius* (37) and *Archaeoglobus fulgidus* (21). Unfortunately, the search for putative H/ACA sRNA genes by computational analysis was made for other archaeal species: *Pyrococcus furiosus* (22,24,38), *Methanococcus jannaschii* (26) and *Thermococcus kodakarensis* (25). The utilization of RNomics approaches for the search of H/ACA sRNAs in archaea is even more scarce, it was only applied to the *A. fulgidus* (21) and *Sulfolobus solfataricus* (39) species: three H/ACA sRNAs and one single sRNA were identified, respectively. For a better definition of the rules that govern the H/ACA sRNP specificities and efficiencies, it was of high importance to identify all the putative H/ACA sRNAs of a given species as well as Ψ residues in its rRNAs and then, to try to assign the detected Ψ residues to the detected H/ACA sRNAs.

To this end, here, we used different experimental approaches in order to define the target sites of the putative H/ACA sRNAs that we identified by a computational analysis of the genomic sequences of three *Pyrococcus* species. These three species, *Pyrococcus abyssi*, *Pyrococcus furiosus* and *Pyrococcus horikoshii*, are hyperthermophiles (optimal growth temperature between 95 and 100°C). As a first step, we developed and used various computational approaches to identify all the putative H/ACA sRNAs of these species and their putative target sites in rRNAs. Then, to test for the presence of Ψ residues at the predicted sites in 16S and 23S rRNAs, we applied the RT-CMCT approach to large segments of the *P. abyssi* 16S and 23S rRNAs that were including the predicted pseudouridylation sites. Finally, to confirm the role of the identified H/ACA motifs in formation of the identified Ψ residues, we tested the activities of reconstituted H/ACA sRNPs on small rRNA fragments containing the expected target U residues. Finally, on the basis of the data obtained, we drew conclusions on structural constraints to which H/ACA sRNAs and the H/ACA sRNA–rRNA interactions are subjected.

MATERIALS AND METHODS

Extraction of inter-coding-regions (ICR) from *Pyrococcus* genomes

The sequences and annotations of the archaeal genomes *Pyrococcus abyssi GE5*, *Pyrococcus horikoshii OT3* and *Pyrococcus furiosus DSM3638* were downloaded in Fasta and Genbank formats from NCBI ftp site, ftp://ftp.ncbi.nih.gov/genomes/Bacteria. In each genome, the positions of DNA sequences corresponding to ORFs or template sequences of known stable RNAs (rRNA, tRNA, RNaseP and 7S RNA) were listed in a table denoted 'position table'. Based on this table, the remaining segments of the genomes were extracted automatically from the genomic sequences, by using a script written in awk and perl languages (ExtractICR). Only sequences exceeding 15 nts were collected. ExtractICR flanks each of the linking sequences by their two 15-nt long bordering sequences. These exported sequences, denoted ICRs, were formatted and assembled in data bases, using Readseq (ftp://iubio.bio.indiana.edu/molbio/readseq) and Formatdb (NCBI toolkit).

Selection of conserved ICRs and search for H/ACA sRNA genes

A Blast version adapted for multi-aligment was used to compare the ICR sequences. First, repeated elements (more than 50-nt long segments repeated several times in the genome) were removed by comparison of all the

ICRs extracted from one given genome. Then, the ICRs from one given pyrococcal species were compared to those of the two other species. Based on a statistical analysis, the criteria retained to consider that an ICR shows a significant degree of conservation in the three species was the presence of at least a stretch of 18 nts with an identical sequence in these species or a 21-nt long sequence with only one base-pair substitution in one or two of the three species. The selected ICRs were then aligned using Clustal-W (40). GeneMarks (41) was used for elimination of the conserved ICRs corresponding to mis-annotated ORFs. An RNAMOT (42) descriptor, that was designed for C/D box sRNA gene detection, was used to identify these genes in the conserved ICRs. Finally, another RNAMOT descriptor was used to screen for the presence of H/ACA sRNA genes. It was based on some of the known structural features of the archaeal H/ACA motifs: the presence of two helices H1 (at least 7 bps) and H2 (at least 5 bps) flanking an internal loop (each strand including 5 to 11 nts). The length of the apical loop was allowed to vary between 8 and 35 nts. The descriptor included an ACA trinucleotide flanking helix H1. Only H/ACA motifs having at least one putative target site in rRNAs or tRNAs that fitted to the rules defined for H/ACA sRNA–rRNA interactions were retained. To this end, RNAMOT descriptors were built for each putative H/ACA motif as described in (25).

Search for H/ACA sRNA genes in entire genomic sequences

Then, based on the results obtained with RNAMOT, we used the ERPIN (43) software for searches with higher constraints in the entire *Pyrococcus* genomes. ERPIN builds helix and single-strand lod score profiles from sequence alignments and screens genome sequences for occurrences of these profiles. The H/ACA sRNAs that we identified by screening the conserved ICRs, together with the H/ACA sRNAs identified experimentally in *A. fulgidus* (21), were used to build the ERPIN profile as described in (25). As above, for each new candidate detected by this approach, RNAMOT was used to predict target sites in rRNAs (25).

The nucleotide sequences of all the H/ACA sRNAs detected in this study and the positions of their template sequences in the archaeal genomes are accessible at http://tagc.univ-mrs.fr/erpin/.

P. abyssi cultures

P. abyssi strain GE5 cells were grown as described (44) at 95°C in Vent Sulfothermophiles Medium (20 g/l NaCl, 0.25 g/l KCl, 0.05 g/l NaBr, 0.5 g/l $SrCl_2$-$6H_2O$, 0.08 g/l boric acid, 3 g/l PIPES, 1 g/l yeast extracts, 4 g/l peptone, 1 g/l Resazurine, 200 g/l $MgSO_4$, 50 g/l $CaCl_2$, 50 g/l KH_2PO_4 pH 6.8). *P. abyssi* cell growth was stopped at the end of the exponential phase.

Total RNA isolation

The *P. abyssi* cells were collected by centrifugation. After washing in 1 M sorbitol, 25 mM Hepes, pH 7.0, they were frozen and stored at −80°C. The method described by Chomczynski and Sacchi (1986) (45) was used for extraction of total RNA. About 10^{10} cells were dissolved in 4 ml of solution D (4 M guanidinium thiocyanate, 25 mM sodium citrate pH 7.0, 0.5% sarcosyl, 0.1 M β-mercaptoethanol). The extracted RNAs were recovered by phenol/chloroform extraction, followed by ethanol precipitation using 0.3 M sodium acetate. They were dissolved in bi-distilled water and quantified.

Northern blot analysis

About 10 µg of *P. abyssi* total RNA and 5′-end labelled DNA size markers (100 bp DNA ladder, MBI Fermentas) were fractionated in parallel on 6% denaturing polyacrylamide gel (8 M urea, 0.5× TBE buffer). After electrotransfer on a Hybond-N+ membrane (Amersham) and by UV-crosslinking, a pre-hybridization was carried out for 1 h at 58°C in SSPE buffer (0.9 M NaCl, 47 mM Na_2HPO_4-$2H_2O$, 6 mM EDTA pH 7.4, containing 1 g/l Ficoll, 1 g/l polyvinylpyrolidone, 1 g/l BSA, 0.5% SDS). Oligonucleotide probes complementary to the predicted H/ACA sRNAs (Table S1 in Supplementary data) were 5′-end labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase for 1 h at 37°C. Hybridization was carried out at 58°C for 16 h in the presence of 100 ng of the labelled oligonucleotide. The membranes were washed four times in SSPE buffer at 42°C for 5 min and the hybridization bands were visualized on a Typhoon 9410 (Amersham Biosciences).

Mapping of pseudouridine (Ψ) residues in the *P. abyssi* rRNAs

The *N*-cyclohexyl-*N*′-(2-morpholinoethyl)-carbodiimide metho-*p*-toluolsulfonate (CMCT) RNA treatment was adapted from Bakin and Ofengand (1993) (46), as previously described (47). CMCT modifications were performed on 5 µg of *P. abyssi* total RNA. Positions of CMCT modifications were identified by primer extension analysis, using the AMV RT (QBiogene, USA). The sequences of the 5′-end labelled primers that we used are given in Table S1 (Supplementary data). RNA sequencing was done on 20 µg of *P. abyssi* total RNA.

Recombinant protein production

The recombinant GST-L7Ae, GST-aNOP10, GST-aGAR1, GST-aCBF5 proteins were produced in *E. coli* BL21 CodonPlus cells (Novagen). The GST-tag was cleaved by the precision protease in the course of purification performed on Glutathione-Sepharose 4B (Pharmacia), as previously described (23).

In vitro transcription of H/ACA sRNAs and their rRNA substrates

The Pab19, Pab21, Pab35, Pab40, Pab91, Pab105 and Pab160 sRNA sequences were PCR amplified from the *P. abyssi* GE5 genomic DNA using a forward primer containing the T7 RNA polymerase promoter and a second primer complementary to the expected 3′ end of the sRNA (Table S1 in Supplementary data). The amplified DNA fragments were cloned in the pTAdv vector (Clontech) and sequenced. *In vitro* transcriptions with the T7 RNA polymerase were performed as

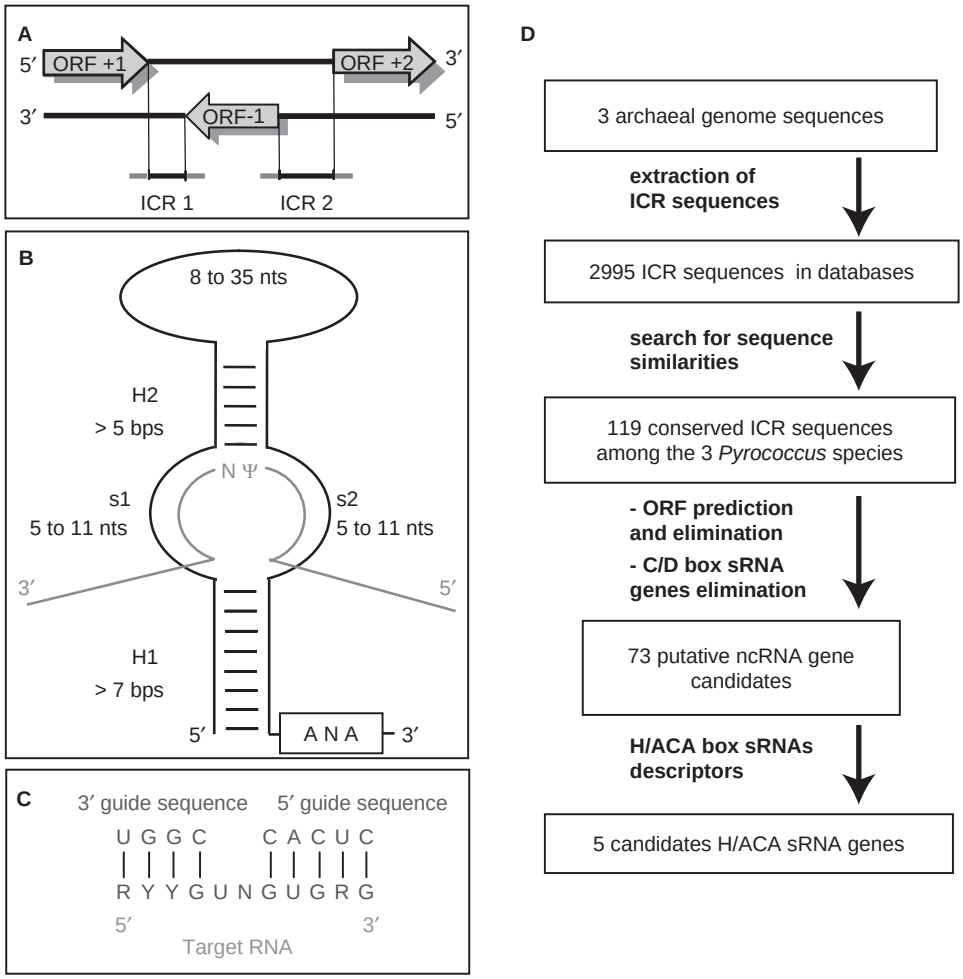


Figure 1. The computational process used for the search of H/ACA genes in ICRs. Panel A: ICR definition. ICRs include 15 bps upstream and downstream of the segments linking the ORFs or template sequences for known RNAs. Panel B: the helices H1 and H2, and loop size criteria used to design the RNAMOT descriptor devoted to the search for H/ACA sRNA genes are shown. Panel C: Example of base-pair criteria used to search for possible rRNA targets of the candidate H/ACA motifs. Panel D: the 5 computer based-steps used for the search of H/ACA genes: (1) sequences and annotations were downloaded, (2) sequences of ICRs were extracted from the genomic sequences and assembled in databases, (3) after elimination of ICRs containing repeated sequences, the ICR sequences from one species were compared to those of other species by using Blast, 119 conserved ICRs were selected, (4) protein-coding sequences and known C/D box sRNA genes were filtered out of the conserved sequences and (5) RNAMOT descriptors were used for the search of H/ACA genes in ICRs, five putative genes were detected.

previously described (23). The H/ACA sRNA transcripts were purified by gel electrophoresis.

Template DNA encoding the RNA targets derived from *P. abyssi* rRNAs were obtained by annealing two complementary oligonucleotides. *In vitro* transcription was performed as previously described (23), in the presence of 20 μCi [α-^{32}P] NTP (800 Ci/mmol), 0.13 mM of the same NTP and 4 mM of each of the three other NTPs. When using RNase P1 for the digestion, [α-^{32}P]UTP was used for labelling. When RNase T2 was used, the identity of the [α-^{32}P]NTP used was defined by the residue located 3′ to the targeted U residue. The sequences of the *in vitro* transcribed substrates are given in Table S2 in Supplementary data.

H/ACA sRNP reconstitution and test of their activity

As previously described (48), the unlabelled sRNA (4 pmol) and [α-^{32}P] NTP-labelled targets (150 fmol) were mixed with the four L7Ae, aCBF5, aNOP10 and aGAR1 recombinant proteins at a 200 nM concentration, at room temperature in buffer D (150 mM KCl; 1.5 mM $MgCl_2$; 0.2 mM EDTA; 20 mM HEPES, pH 7.9). The mixture was incubated at 65°C for 80 min. Then, formation of Ψ residue was detected after either T2 or P1 RNase digestion. The 3′-mono phosphate nucleotides produced by RNase T2 digestion or the 5′-monophosphate residues obtained after P1 RNase digestion were fractionated by thin-layer chromatography (TLC) (mono or 2D) as previously described, using the N1 buffer for 1D TLC and N1-N2 or N1-R2 buffers for 2D TLC (48). The radioactivity in the spots or the bands was quantified with a phosphorimager, by using the ImageQuant software. When digestion was achieved with RNase T2, the yield of Ψ formation (expressed in mol of Ψ residue per mol of target RNA) was estimated taking into account the total number of residues located 5′ to the incorporated

labelled nucleotide. After P1 RNase digestion, we only took into consideration the total number of U residues in the target RNA.

RESULTS

Analysis of conserved ICRs in *Pyrococcus* identifies five conserved putative H/ACA sRNAs

In order to make a link between H/ACA sRNAs and pseudouridylation in rRNAs, in the *P. abyssi* species, we first made a complete identification of H/ACA–sRNA motifs in *P. abyssi*. This was done in two steps because, when we started this study, only a limited number of archaeal H/ACA sRNAs had been identified. As a first step, we applied a phylogenetic approach to the *Pyrococcus* genus. The idea was that DNA segments, which link ORFs and/or template sequences for stable known RNAs (rRNAs, tRNAs, RNase P, 7S RNA) (Inter-Coding-Regions, ICRs) and that bear long stretches of conserved sequences in three *Pyrococcus* species *P. abyssi*, *P. furiosus* and *P. horikoshii*, may correspond to genes for functional non-coding RNAs. Based on the few H/ACA sRNAs known at that time (three from *A. fulgidus*) (21) some characteristic features of H/ACA sRNA motifs could be delineated. They were used to build an RNAMOT descriptor for the search of H/ACA sRNA genes. Then, RNAMOT profiles were built for the search of the possible rRNA target sites of each candidate H/ACA motif. By using this approach (see the details in Materials and Methods), we detected five putative H/ACA sRNA genes, that were common to the three species. Four of them have been recently characterized in *P. furiosus* by the use of a computational approach based on G/C content analysis, namely, the Pf1, Pf3, Pf6 and Pf7 sRNAs (22,49). The counterparts that we identified in *P. abyssi* and *P. horikoshii* are designated as Pab21, Pab105, Pab35 and Pab40 sRNAs (*P. abyssi*) and Pho21, Pho 105, Pho35 and Pho40 sRNAs (*P. horikoshii*), respectively (Figure 2). The fifth common H/ACA sRNA, which had not been characterized by other teams, is denoted Pab91, Pfu91 and Pho91 in *P. abyssi*, *P. furiosus* and *P. horikoshii*, respectively. We previously used it to settle conditions for *in vitro* reconstitution of active H/ACA sRNPs (23). Taking into account the numerous compensatory base-pair mutations in the three species studied and in *T. kodakarensis*, we could propose relevant secondary structures for each of the five sRNAs (Figure 2). Only for sRNA Pab21 and for motif 2 in sRNA Pab40, it was difficult to make a choice between two possible 2D structures that were both containing a K-loop motif (Figure 2). Production of the H/ACA sRNAs Pab21, Pab35, Pab40, Pab91 and Pab105 in *P. abyssi* was verified by Northern blot analysis (Figure 3), and two forms of Pab21 sRNA with or without the C/D box motif were found to be present *in vivo* (Figure 3).

Screening of complete genomes with ERPIN identifies two additional common putative H/ACA sRNAs in Pyrococci

Some H/ACA motifs may have escaped detection when using the above comparative approach. Therefore, we took advantage of the increased knowledge on H/ACA sRNAs, which was brought by the 15 pyrococcal H/ACA sRNAs that we identified, to develop another computational approach based on the ERPIN software (25,43). To this end, the sequences of the 3 known H/ACA sRNAs from *A. fulgidus* and the 15 pyrococcal H/ACA sRNAs were aligned by ERPIN, on the basis of their proposed 2D structures and conserved sequence elements. As recently described (25), by comparison of such a profile with complete archaeal genomic sequences, ERPIN can predict H/ACA sRNA genes with a high degree of efficiency. By using different sets of constraints for the lengths of the 3′- and 5′-guide strands of the pseudouridylation pocket, we detected two other putative H/ACA sRNAs common to the three *Pyrococcus* species. One of them was also recently detected in *P. furiosus* (sRNA Pf9) by another approach (24). Its counterparts in *P. abyssi* and *P. horikoshii* were denoted Pab160 and Pho160, respectively (Figure 2F). The second H/ACA sRNA identified (Pab19, Pfu19 and Pho19 in *P. abyssi*, *P. furiosus* and *P. horikoshii*, respectively) had not been detected previously, probably because of its extended 5′-guide sequence (Figure 2). Nucleotide sequence conservation in the three species and secondary structure conservation by compensatory base changes strongly suggest the presence of a small 5′ stem-loop structure upstream from the H/ACA motif in sRNA Pab19. Note that the Pab19 pseudouridylation pocket is larger than usual due to the length of its 5′-guide sequence (15 nts) (Figure 2F). By applying the ERPIN search approach with other kinds of relaxed constraints to each of the 3 *Pyrococcus* genomes, we did not find any other putative H/ACA sRNA gene. Therefore, we concluded that *Pyrococcus* species probably contain only seven H/ACA sRNAs, which are highly conserved in this genus. In addition, they are also conserved in *Thermococcus kodakarensis* that belongs to the same order as the *Pyrococcus* genus (25).

The seven pyrococcal H/ACA sRNAs are predicted to target 21 sites in rRNAs

We used the RNAMOT software (25) to predict possible target sites in rRNAs for each of the seven identified putative H/ACA sRNAs, and this in the three species studied. One up to four distinct target sites were predicted for each putative pseudouridylation pocket (Figure 4). Taken together, 14 and 7 target sites were predicted in 23S and 16S rRNAs, respectively. Note that motif 1 in sRNA Pab35 and sRNA Pab160 are both predicted to guide modifications at position 922 in 16S rRNA and at position 2672 in 23S rRNA (Table 1). Only 8 of the 17 sites that we predicted for the five previously identified *P. furiosus* H/ACA sRNAs had been proposed (22). Strong conservation of the possibility to form the rRNA–sRNA interactions are observed in the three species (Figure 4). Note that in the non-canonical interaction proposed for sRNA Pab19, two single-stranded residues are found 3′ to the targeted U residue (5′-UGC-3′) (Figure 4G). We verified that none of the seven identified H/ACA sRNAs can act on tRNAs. In addition, no other putative H/ACA

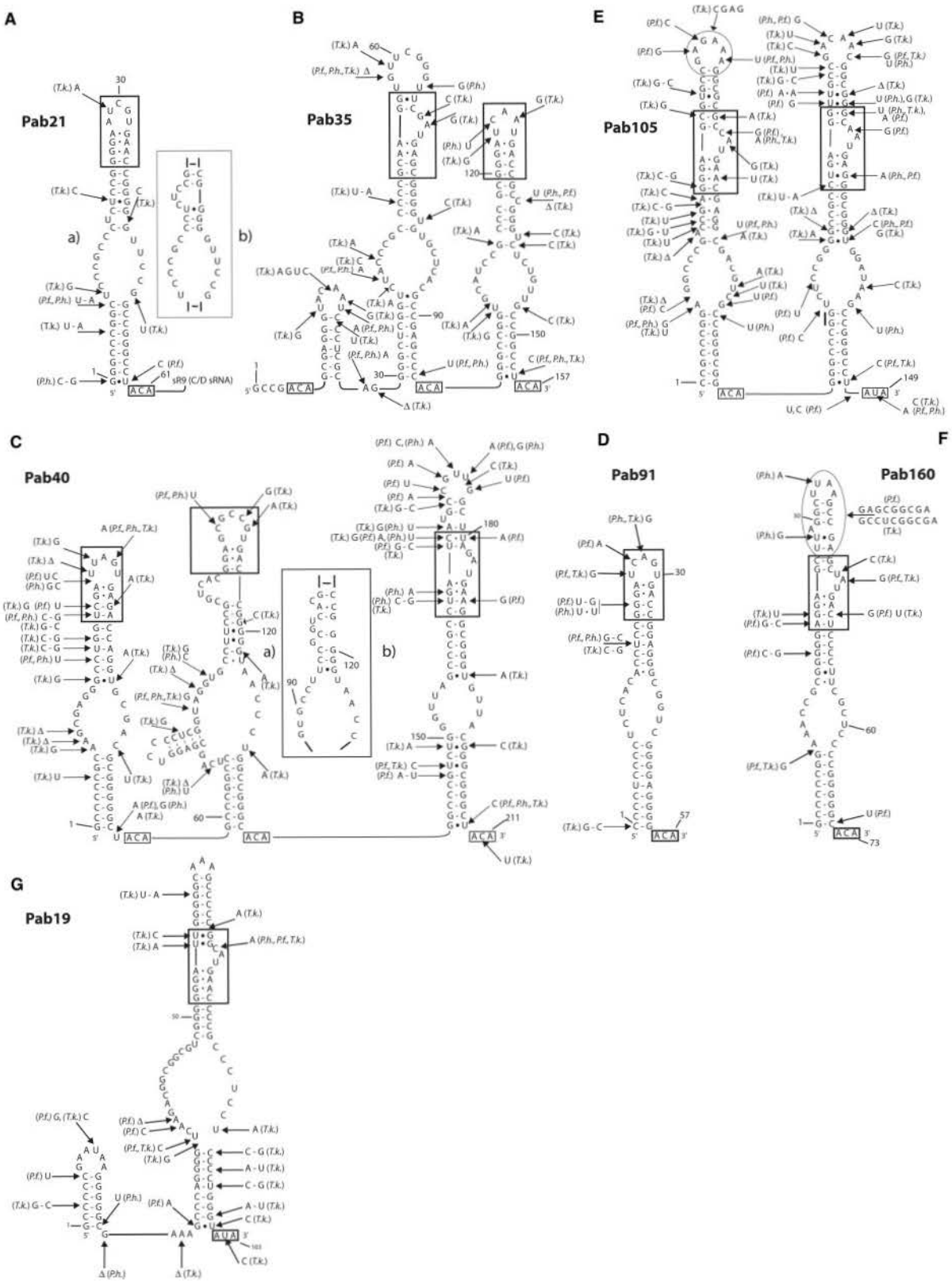


Figure 2. Sequences and proposed secondary structures of the seven pyrococcal H/ACA sRNA candidates. The sequences and proposed secondary structures for the seven candidate *P. abyssi* sRNAs, Pab21 (**A**), Pab35 (**B**), Pab40 (**C**), Pab91 (**D**), Pab105 (**E**), Pab160 (**F**) and Pab19 (**G**) sRNAs, are shown. The ANA sequence at the 3′ end of the RNA and the K-turn or K-loop motif in the apical part of the H/ACA motif are boxed. Base substitutions in the *P. furiosus (P.f.)*, *P. horikoshii* (*P.h.*) and *T. kodakarensis* (*T.k.*) (25) sRNAs are shown. The two putative foldings, which can be proposed for sRNA Pab21 and for motif 2 in sRNA Pab40, are shown (insets in panels A and C). Only the conformation shown in the entire molecule turned to be functional.

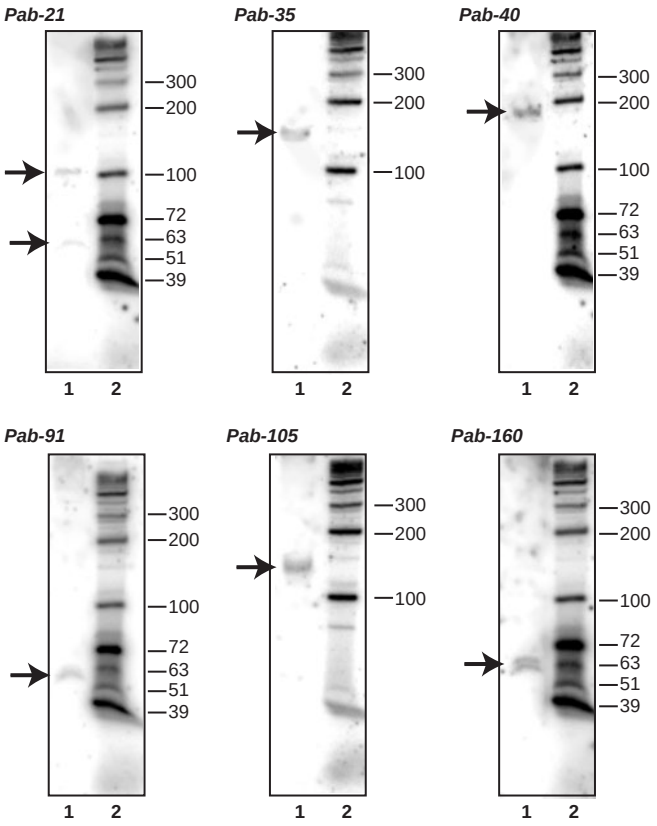


Figure 3. Northern blot analysis of the candidate *P. abyssi* H/ACA sRNAs. The expression of each gene encoding a putative H/ACA sRNA was tested by Northern blot analysis of *P. abyssi* total RNA extracts, as described in Materials and Methods (Lane 1 in each panel). The nucleotide sequences of the specific 5′-end radio-labelled probes used for each sRNA are given in Table S1 in Supplementary data. A 5′-end labelled DNA ladder was loaded in Lane 2. The detected transcripts are shown with an arrow on the left of each autoradiogram.

sRNA that may guide modification of a tRNA was detected in any of the three species.

Experimental search for Ψ residues at the predicted target sites in rRNAs

We used the CMCT-RT approach (47) to test for the presence of Ψ residues at the rRNA positions predicted to be targeted by the seven putative H/ACA sRNAs. To this end, total RNA from *P. abyssi* was treated with CMCT with or without further alkaline incubation. Positions of the alkaline-resistant Ψ residues were detected by primer-extension analyses using a large series of oligonucleotide primers (Table S1 in Supplementary data). Altogether, the 16S and 23S rRNA segments that were probed represented 20 and 27% of the entire molecules, respectively. In particular, large parts of the 23S rRNA domains located at the peptidyl-transferase centre (domains IV and V) were analysed. We identified 3 and 11 Ψ residues in *P. abyssi* 16S and 23S rRNAs, respectively, 8 of them are located in domain V of 23S rRNA.

The experimental analysis confirmed Ψ formation at 12 of the 21 predicted sites (positions 27, 891, 922 in 16S rRNA, and positions 1932, 2549, 2554, 2588, 2672, 2685, 2697, 2794 and 2930 in 23S rRNA) (Table 1 and Figure 5). In addition, the absence of Ψ formation was clearly demonstrated at six of the predicted positions (positions 892, 995 and 1122 in 16S rRNA and 278, 2250, and 2552) (Table 1 and Figure 5). Note that three of these unmodified positions were previously proposed to be modified in *P. furiosus* (22). Based on the present data, they are unlikely to be modified in this species. For three of the positions predicted to be modified in *P. abyssi*, the experimental analysis was obscured by the presence of an RT pause at the level of the stop expected after CMCT modification (positions 1017 in 16S rRNA, 1377 and 2016 in 23S rRNA). Therefore, we could not determine whether U to Ψ conversion occurred at these positions. However, for one of them (2016 in 23S rRNA), formation of a Ψ residue was detected at the corresponding position in two other archaeal species: *H. halobium* (34) and *H. marismortui* (35,36), and in nearly all bacterial and eukaryal organisms which were studied up to now (10).

Interestingly, in the course of this analysis, Ψ residues were found at two positions that were not predicted to be targeted by any of the identified H/ACA sRNAs (positions 2585 and 2603 in 23S rRNA). Even by relaxing the constraints in the ERPIN search, we could not detect putative H/ACA sRNAs for these positions. Therefore, their formation may be sRNA independent.

The seven H/ACA sRNPs are active *in vitro*

To get experimental supports for the relationships that we established between Ψ residues in *P. abyssi* rRNAs and the identified H/ACA sRNAs, we used the H/ACA sRNP *in vitro* reconstitution method which was developed in the laboratory (23). To this end, each of the seven identified H/ACA sRNAs was transcribed, and the four recombinant aCBF5, aNOP10, aGAR1 and L7Ae proteins were produced. Assembly of individual proteins and different combinations of them on the H/ACA sRNAs was tested by electrophoresis-mobility shift assay (EMSA). Each of the identified H/ACA sRNA could be assembled into a sRNP (data not shown), which was a clear demonstration that they were true H/ACA sRNAs. Then, we measured the activities of the reconstituted particles on all their predicted target sites. To this end, we used small *P. abyssi* rRNA fragments containing from 18 up to 31 nts, most of them had a single target U residue located in the middle of the molecule, except 3 substrates that were containing 2 or 3 target U residues, because of the close vicinity of these U residues in 23S rRNA (see Table S2 in Supplementary data). Therefore, the activity of one up to three distinct sRNPs was tested on a given RNA substrate. This activity was measured by the nearest-neighbour method as previously described (23,48). T2 RNase digestion was performed on RNA substrates labelled with [α-^{32}P]ATP, CTP, GTP or UTP. P1 RNase hydrolysis was done on [α-^{32}P]UTP labelled RNA substrates. Each RNA substrate was incubated with the four core proteins (LCNG) at 65°C as

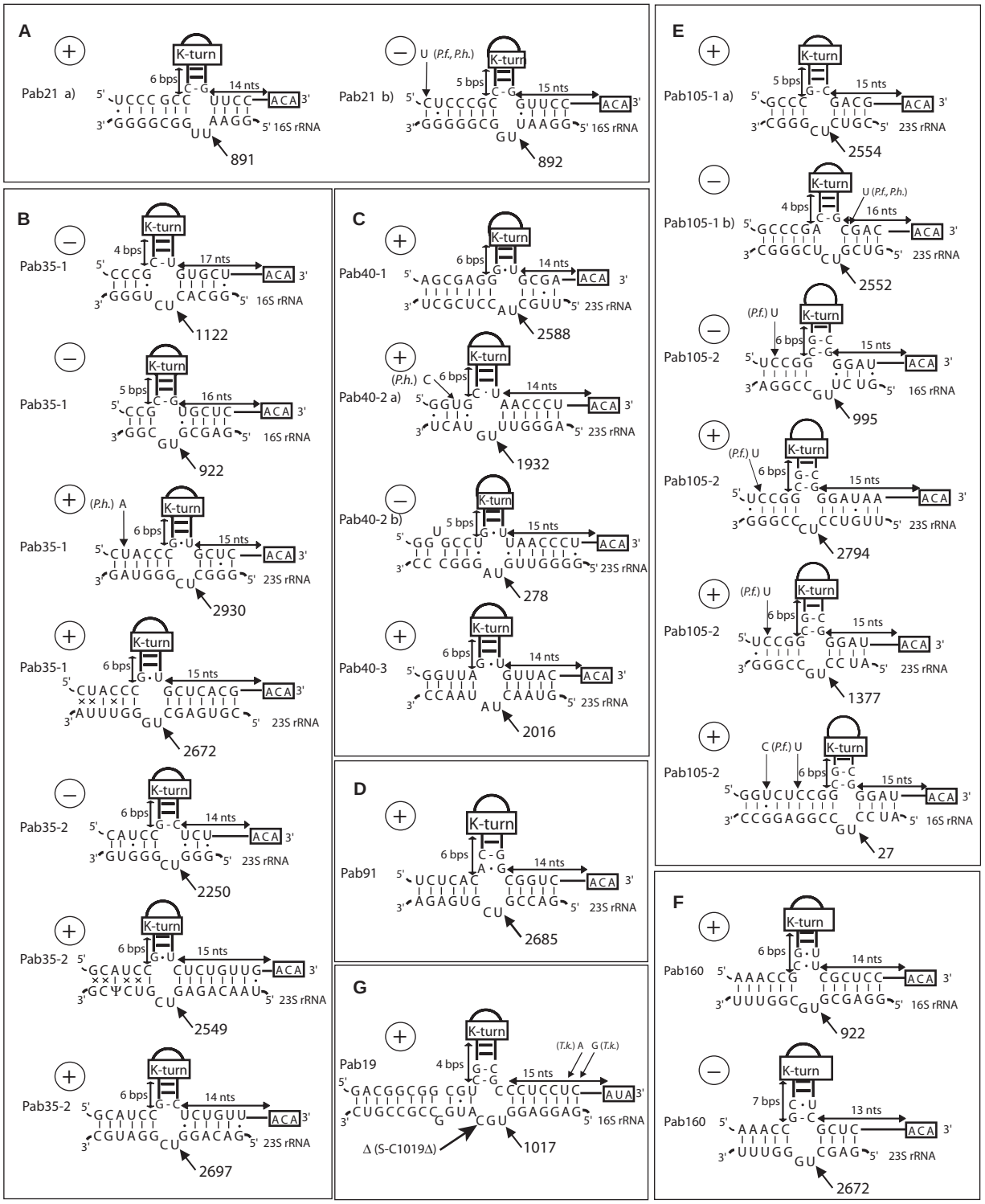


Figure 4. Predicted rRNAs target sites of the 11 H/ACA motifs. Interactions between each H/ACA motif and its putative rRNA target sequences are shown for the seven candidates sRNAs Pab21 (**A**), Pab35 (**B**), Pab40 (**C**), Pab91 (**D**), Pab105 (**E**), Pab160 (**F**) and Pab19 (**G**). The expected targeted U residues are indicated by arrows and numbers giving their positions in the 16S or 23S rRNAs. The distance between the K-turn motif and the pseudouridylation pocket is given in bps, the length between the ANA sequence and the pseudouridylation pocket is given in nts. The two possible intermolecular interactions proposed for sRNA Pab21 and motif 2 in sRNA Pab40 result from the two possible alternative conformations of these H/ACA motifs (Figure 2 Panels A and C). The interactions found to be functional in *in vitro* assays are indicated by (+), the inactive ones are shown by (−).

Table 1. Predicted and experimentally identified target sites of the identified *P. abyssi* H/ACA sRNAs and comparison with *P. furiosus*

P. abyssi H/ACA	Predicted targets	RT-CMCT analysis	*In vitro* activity	*P. furiosus* counterparts	Predicted targets
Pab21	16S rRNA 891	+	++	Pf1	16S rRNA 879
	16S rRNA 892*	–	–		
Pab35	motif 1: 16S rRNA 1122	–	–	Pf6	
	motif 1: 16S rRNA 922	+	–		
	motif 1: 23S rRNA 2930	+	++		motif 1: 23S rRNA 2953
	motif 1: 23S rRNA 2672	+	++		motif 1: 23S rRNA 2695
	motif 2: 23S rRNA 2250	–	–		
	motif 2: 23S rRNA 2549	+	++		motif 2: 23S rRNA 2572
	motif 2: 23S rRNA 2697*	+	+++		motif 2: 23S rRNA 2720*
Pab40	motif 1: 23S rRNA 2588*	+	+++	Pf7	motif 1: 23S rRNA 2611*
	motif 2: 23S rRNA 278*	–	–		
	motif 2: 23S rRNA 1932	+	+		
					motif 2: 23S rRNA 2701
	motif 3: 23S rRNA 2016*	?	+++		motif 3: 23S rRNA 2039*
Pab91	23S rRNA 2685	+	+++	Pfu91	23S rRNA2708
Pab105	motif 1: 23S rRNA 2552*	–	–	Pf3	
	motif 1: 23S rRNA 2554	+	++		motif 1: 23S rRNA 2577
	motif 2: 16S rRNA 995	–	–		
	motif 2: 16S rRNA 27*	+	++		motif 2: 16S rRNA 15*
	motif 2: 23S rRNA 1377	?	+++		motif 2: 23S rRNA 1400
	motif 2: 23S rRNA 2794	+	+++		motif 2: 23S rRNA 2817
Pab160	23S rRNA 2672	+	–	Pf9	
	23S rRNA 922*	+	+++		16S rRNA 910*
Pab19	16S rRNA1017	?	+++	Pfu19	16S rRNA 1005

The predicted target positions in 16S and 23S rRNAs are indicated for each of the candidates H/ACA motifs. Detection by CMCT-RT analysis of a Ψ residue at the predicted position in *P. abyssi* rRNAs is indicated by '+' in the second lane. '?' indicates the presence of an RT pause that obscured the analysis. Detection of an *in vitro* activity of the reconstituted H/ACA sRNP at the predicted site is indicated by '+','++' or '+++' in the third lane. The number of '+' is proportional to the rate of modification detected after a 80 min incubation (25–50%, 50–80% and 80–100%, respectively). Column 5 gives the name of the *P. furiosus* counterpart sRNAs (22,24). The target positions that were previously predicted for these *P. furiosus* Pf1, Pf3, Pf6, Pf7 and Pf9 sRNAs (22,24) are indicated by an asterisk in column 2. Based on the rules that we established from our *P. abyssi* experimental data, we predicted modification positions for both the previously identified *P. furiosus* sRNAs and the two additional sRNAs detected in this study (column 6). The validated previously proposed modified positions are marked by an asterisk.

previously described, in the presence or the absence of the H/ACA sRNA (23). Then, the RNAs were extracted, digested, and fractionated by thin layer chromatography (as described in Materials and Methods) (Figure 6).

Only 15 of the 21 predicted sites were found to be modified by the reconstituted sRNPs and the rates of U to Ψ conversions were ranging from 30 to 100% (Figure 6 and Table 1). The rRNA segments predicted to be modified by two distinct H/ACA motifs were in fact modified by only one of them (motif 1 in sRNA Pab35 modifies position 2672 in 23S rRNA and sRNA Pab160 acts at position 922 in 16S rRNA). The data obtained confirmed that each of the 11 H/ACA motifs is capable to guide pseudouridylation. Even the Pab19 H/ACA motif, which has a large internal loop and forms a non-canonical interaction with its substrate, was active (Figure 6D). Deletion of the additional single-stranded C residue, which is located 3′ to the 5′-UN-3′dinucleotide (mutant S-C1019Δ) (Figures 4G and 6D), showed that the Pab19 sRNP modifies the WT and mutated RNA substrates at very similar rates. Among the two possible conformations of motif 2 in sRNA Pab40, only conformation denoted a in Figure 2C is active *in vitro*. Interestingly, this structure includes an additional stem-loop in the 5′-guide strand and a poorly stable helix 2 (Figure 2C). This may explain the low yield of U to Ψ conversion found for this H/ACA motif as compared to the other *P. abyssi* H/ACA motifs (Figure 6C).

Some of the H/ACA motifs can guide modification at more than one position in rRNAs

Some of the H/ACA motifs can guide modifications at two or even three distinct positions. For instance, altogether, motifs 1 and 2 in sRNA Pab35 can guide U to Ψ conversion at positions 2672, 2930, 2549 and 2697 in 23S rRNA. Motif 1 in sRNA Pab105 only guides modification at one position (2554 in 23S rRNA), while motif 2 in this sRNA can act at positions 1377 and 2794 in 23S rRNA and 27 in 16S rRNA. Therefore, both sRNAs Pab35 and Pab105 can guide modifications at four positions in rRNAs. Interestingly, we found that all the H/ACA sRNAs that contain a single H/ACA motif (Pab21, Pab91, Pab160 and Pab19) guide modification at a unique position in rRNAs. Altogether, the 11 H/ACA motifs of the 7 *P. abyssi* sRNAs can guide U to Ψ conversion at 15 sites in the *P. abyssi* rRNAs (Table 1). Presence of Ψ residue was detected at 12 of these sites (Figure 5 and Table 1). As mentioned above, for technical reasons, modification could not be tested experimentally at the three other positions. However, the strong activity measured *in vitro* at these three rRNA positions is a strong argument for the occurrence of these modifications *in vivo*.

Constraints on the H/ACA sRNA structure and H/ACA sRNA–target RNA interaction

Based on the above data, we tried to define rules for H/ACA sRNA structure and interactions. In several of

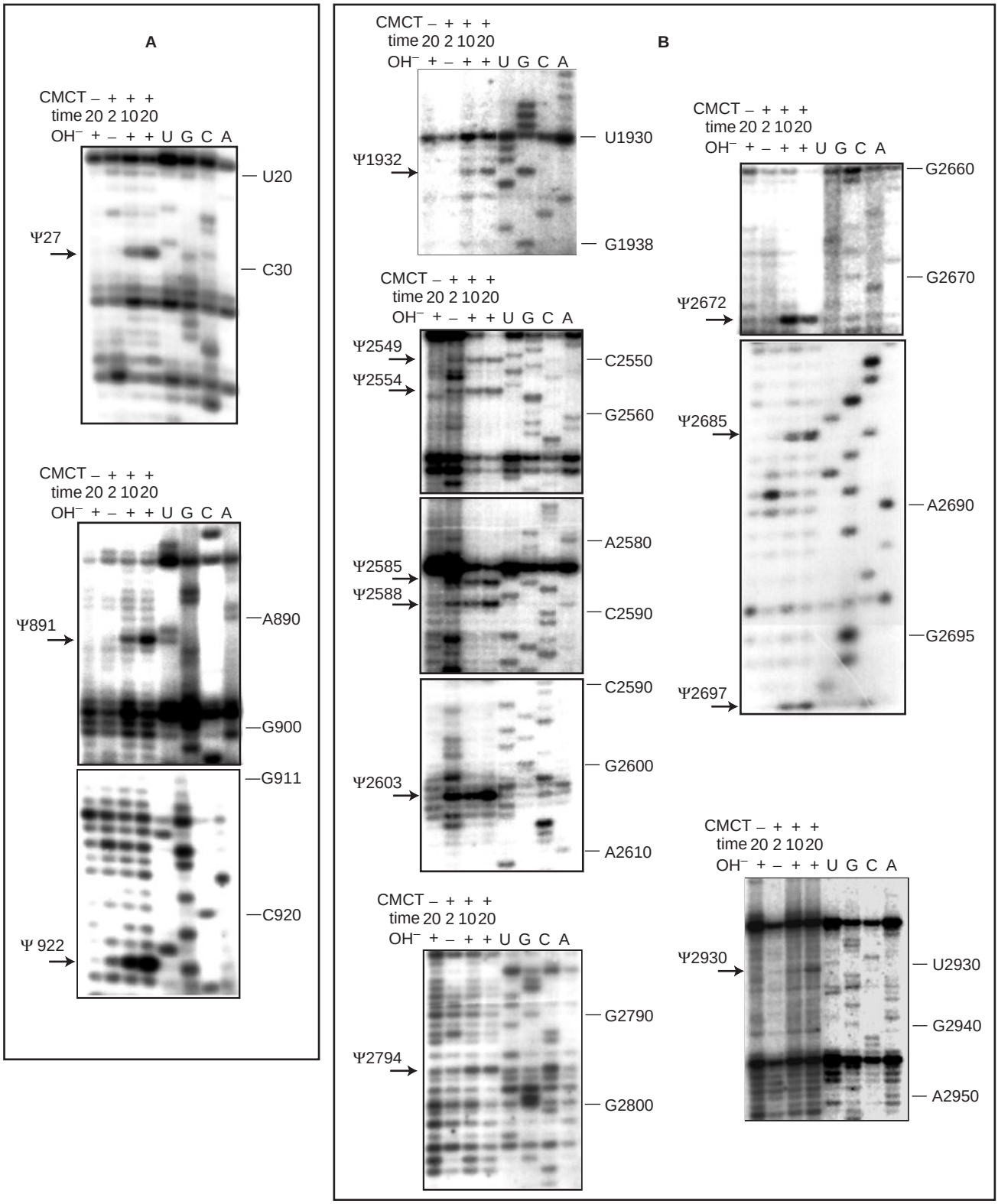


Figure 5. Identification of Ψ residues in the *P. abyssi* 16S (Panel A) and 23S (Panel B) rRNAs by the RT-CMCT method. The *P. abyssi* total RNA was treated in the absence (–) or the presence of CMCT (+), for 2, 10, or 20 min as indicated on the top of the lanes. The CMCT treatment was (+) or was not (–) followed by an alkaline treatment at pH 10.4 as indicated above the lanes. The appearance of an RT stop of increasing intensity after this treatment reveals the presence of a Ψ residue. They are indicated by horizontal arrows. Lanes U, G, C and A correspond to the RNA sequencing ladder. Positions of residues in 16S or 23S rRNAs are given on the right side of the autoradiograms.

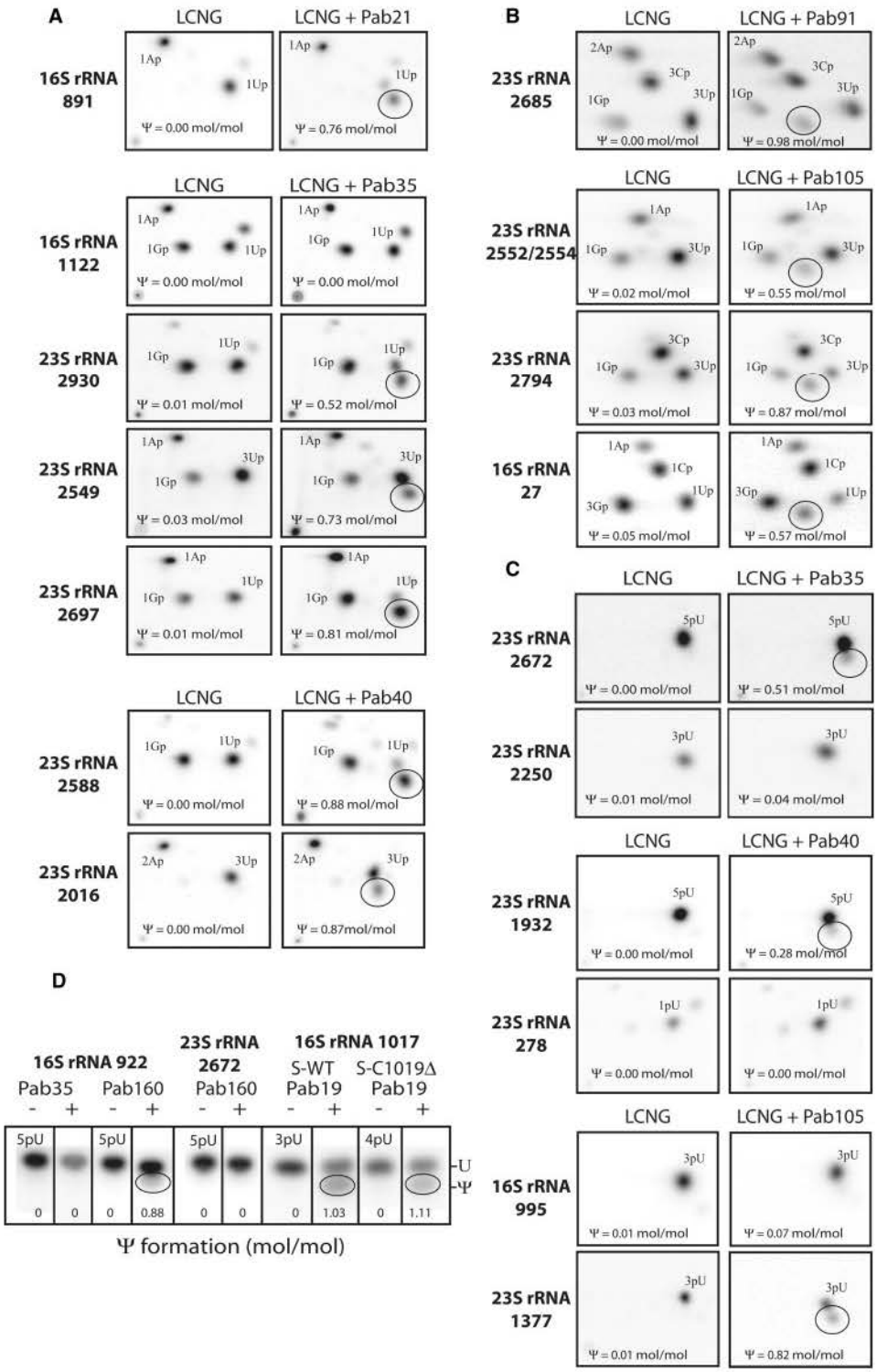


Figure 6. *In vitro* activity tests of the seven reconstituted H/ACA sRNPs. H/ACA sRNP particles were assembled by using 4 pmol of *in vitro* transcribed H/ACA sRNA, the four recombinant proteins L7Ae (L), aCBF5 (C), aNOP10 (N) and aGAR1 (G) (LCNG) (200 nM each) and 150 fmol of radio-labelled *in vitro* transcribed RNA substrate as previously described (23). The putative target position in the RNA substrate is given on the left side of the chromatograms. After a 80-min incubation at 65°C, the RNAs were extracted and digested with T2 (Panels A and B) or P1 nuclease (Panels C and D). The nucleotide 3′ or 5′ monophosphate, that are respectively released, were fractionated by 2D (Panel A, B and C) or 1D (Panel D) thin layer chromatography. When RNase T2 was used for the digestion, the RNA template was labelled by incorporation of [α-^{32}P] ATP, CTP, GTP or UTP depending on the identity of the residue located 3′ to the targeted U residue. When P1 nuclease was used for digestion, labelling was achieved by [α-^{32}P] UTP incorporation. 2D TLC on cellulose plates were performed using either the N1-N2 (Panels A and C), or the N1-R2 (Panel B) buffers and the N1 buffer was used for 1D TLC. Positions of Ap, Cp, Gp, Up and Ψp or pU and pΨ in the chromatograms are indicated as well as their numbers in the substrate RNAs. The Ψ spots are circled. The indicated numbers of Ψ moles formed per mole of RNA substrates were calculated as explained in Materials and Methods.

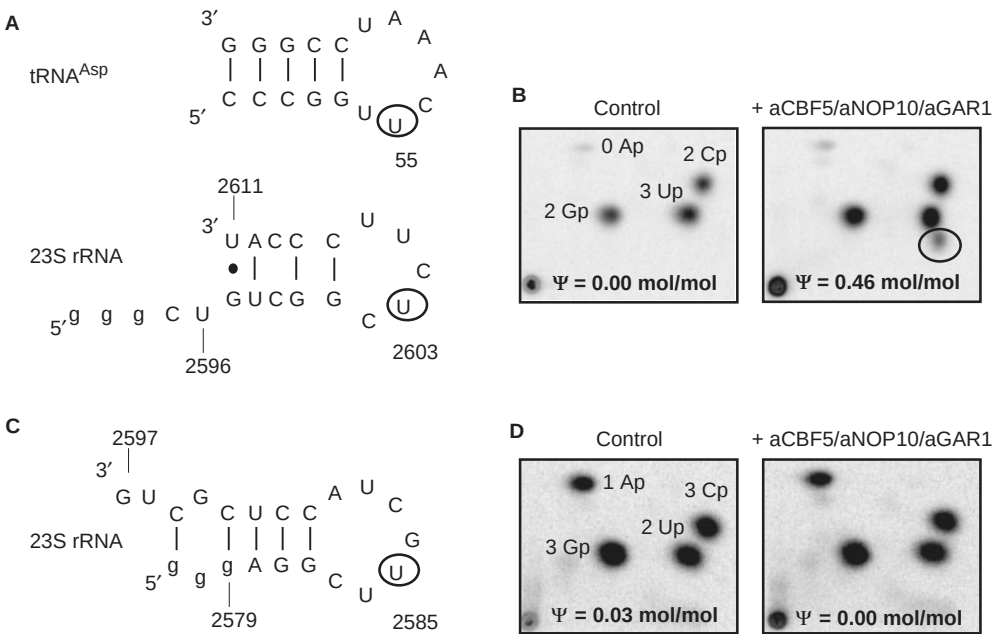


Figure 7. The aCBF5/aNOP10/aGAR1 complex catalyses U to Ψ *in vitro* conversion at position 2603 in a 23S rRNA fragment. Panel A: Comparison of the 2D structure of the TΨC arm of the *P. abyssi* tRNA^{Asp} with the 2D structure that can be formed by the *P. abyssi* 23S rRNA fragment used to test the pseudouridylation activity of the aCBF5/aNOP10/aGAR1 at position 2603 in this rRNA. The three G residues located at the 5′ extremity of the rRNA substrate arose from the T7 promoter. The U residues targeted by the aCBF5/aNOP10/aGAR1 complex are circled (Panel B). Panel B: The 20-nt long RNA substrate (150 fmol) was incubated with the aCBF5/aNOP10/aGAR1 protein complex (200 nM each). The experiment was performed in conditions described in Figure 6, except the absence of guide RNA. A control experiment was performed in the absence of the protein complex. Positions of the Ap, Cp, Gp, Up and Ψp spots are indicated on the chromatograms, as well as their specific ratios in the target RNA. The Ψ spot is circled. The yield of Ψ formation was assessed and is indicated. Panel C: Proposed secondary structure for the 23S rRNA fragment used to test the *in vitro* activity of the aCBF5/aNOP10/aGAR1 protein complex at position 2585 in the *P. abyssi* 23S rRNA. Panel D: Absence of *in vitro* activity of the aCBF5/aNOP10/aGAR1 complex at position 2585 in the *P. abyssi* 23S rRNA.

the predicted sRNA–target RNA interactions that turned to be non-functional, helix H2 was too short (4 bps) or too long (7 bps). The most frequent length in functional interactions (14 out of 15) is of 6 bps. Some of the non-functional interactions also showed too long or too short distances between the ACA trinucleotide and the targeted U residue (17 or 13 nts). The distance in the active interactions is of 15 or 14 nts. Interestingly also, the base-pair interaction established with the sRNA 3′-guide sequence is more important than that formed with the 5′-guide sequence. This is exemplified by the efficient modifications at positions 2549 and 2672 in 23S rRNA, which are guided by the H/ACA motifs 1 and 2 of sRNA Pab35, respectively (Figure 4). In both cases, the 5′-guide sequence forms a weak base-pair interaction with the rRNA substrate. On the contrary, in spite of a canonical length of helix H2 (6 bps) and a canonical distance between the ACA trinucleotide and the target U residue (15 nts), motif 2 in sRNA Pab105 did not guide modification at position 995 in 16S rRNA, probably because of the low stability of the interaction formed with the sRNA 3′-guide sequence (4 bps including two G.U pairs). Therefore, the length of helix H2, the distance between the ACA triplet and the targeted residue as well as the stability of the base-pair interaction established by the 3′-guide sequence and the rRNA substrate are essential criteria for activity. In contrast, unexpectedly, the presence of the 5′-UN-3′ single-stranded dinucleotide between the two intermolecular interactions is not a strict rule.

One of the two orphan Ψ residues in 23S rRNA can be formed *in vitro* without guide sRNA

As we detected no guide H/ACA sRNA for residues Ψ2585 and Ψ2603 in *P. abyssi* 23S rRNA, and as the aCBF5/aNOP10/aGAR1 complex can modify position 55 in tRNAs in the absence of guide sRNA (50–52), we tested the *in vitro* activity of this complex at these two 23S rRNA positions. The assays were performed using two 20-nt long rRNA fragments (Table S1) containing, respectively, residue U2585 or U2603 (Figure 7). The aCBF5/aNOP10/aGAR1 complex was active on the rRNA fragment containing residue U2603 (46% yield) (Figure 7), but not on that containing residue U2585. Interestingly, we observed that the sequence containing the residue 2603 can be folded into a stem-loop structure showing some similarity with the TΨC stem-loop of tRNA (Figure 7). We concluded that formation of residue Ψ2603 may be catalyzed by the free aCBF5-aNOP10-aGAR1 complex, while an unidentified catalyst may act at position U2585.

DISCUSSION

Application of the various computational approaches that we developed for the search of H/ACA sRNAs in archaeal genomes turned to be highly efficient, since two of the seven sRNAs detected in *P. furiosus* were not found by other approaches previously applied to one of this species (22,24). Our blind detection of C/D box

sRNAs in conserved ICRs allowed the detection of 45 out of the 49 known C/D box sRNAs which are conserved in the three pyrococcal species. This finding illustrates the efficiency of our ICR-based approach for the search of non-coding RNAs of unknown structural characteristics. Up to now, computational predictions of the H/ACA sRNA–rRNA targets were not reliable. This is illustrated by the numerous corrections that we made in previous predictions. Our experimental search for Ψ residues in *P. abyssi* rRNAs, together with the use of the H/ACA sRNP reconstitution and activity assays, turned to be essential to identify the sites which are really targeted by the identified sRNAs. However, by using the structural rules that we established for functional sRNA–rRNA interactions, computational predictions will be more reliable.

A high number of Ψ residues in *P. abyssi* rRNAs as compared to other archaea

Based on experimental analysis of Ψ residues in rRNAs and reconstitution of H/ACA sRNPs, the *P. abyssi* rRNA regions that we studied (20 and 27% of the 16S and 23S rRNAs, respectively) probably contain 17 Ψ residues. These rRNA regions were selected because they contain the highest number of post-transcriptionally modified residues in all living organisms (10). However, we cannot exclude the possibility that some additional Ψ residues are formed without the use of guide RNA, in *P. abyssi* rRNA segments located outside of these regions. Up to now, 17 Ψ residues is the highest number of Ψ residues found in archaeal rRNAs. Only three and four Ψ residues were found upon complete analysis of the *H. marismortui*, *H. halobium* 16S and 23S rRNAs, respectively (34–36). Six Ψ residues were detected in domains II, IV and V of the *S. acidocaldarius* 23S rRNA and five Ψ residues were found in *A. fulgidus* rRNAs when looking for H/ACA sRNA target sites (21,37). Interestingly, the two halophile species, *H. marismortui* and *H. halobium*, which have the smallest number of Ψ residues, grow at 50 and 42°C, respectively. *S. acidocaldarius* and *A. fulgidus* grow at 80 and 83°C, respectively, whereas *P. abyssi* optimally grows at 98°C. Therefore, as already proposed for archaeal 2′-*O*-methylations (53), there may be a correlation between the number of Ψ residues in archaeal rRNAs and the growth temperature of the organism. Confirmation of this statement requires further analysis on a larger number of archaeal species growing at different temperatures.

Ψ residues in *P. abyssi* 23S rRNA are concentrated at the PTC

Interestingly, 2 and 8 of the 13 Ψ residues expected to be present in *P. abyssi* 23S rRNA are located in the functional domains IV and V, respectively (Figure 8). Noticeably, residue 2016 in domain IV corresponds to one of two highly conserved Ψ residues in stem-loop structure 69 (SLS69). Its conservation is probably explained by the high functional importance of SLS69: (i) it contacts both the 16S rRNA and the tRNA bound at the acceptor site (A site) (12,14), (ii) it is located at the subunit interface and, (iii) it was proposed to play a role in tRNA translocation (9,54,55) and in translation termination (12,16). Formation of two Ψ residues in the terminal loop of SLS69, including the *P. abyssi* Ψ2016 counterpart, was found to confer a growth advantage in *S. cerevisiae* (12,17).

Eight Ψ residues are concentrated in domain V of the *P. abyssi* 23S rRNA. This domain is expected to be directly involved in the catalytic activity of the PTC. Interestingly, these 8 Ψ residues are all located in helical segments (Figure 8). They may stabilize the conformation of this 23S rRNA region at the high growth temperature of *P. abyssi*. A cooperative effect of Ψ residues in stabilizing RNA conformation, was previously proposed (10). A need for this cooperative effect may explain the presence of 4, 2 and 2 Ψ residues in close vicinity in helices 89, 90 and 93, respectively. Only three of these eight Ψ residues were detected in several eukarya, including *S. cerevisiae* (Figure 8) (34,56).

Altogether, four pseudouridylation sites in *P. abyssi* 23S rRNAs are highly conserved in the large eukaryal rRNAs (positions 2016 in domain IV, 2549, 2588 and 2603 in domain V) (Figure 8). In addition, residues Ψ2930 detected in domain VI is also present in mouse and human rRNAs. Position 2698 in *P. abyssi* 23S rRNA corresponds to a frequently pseudouridylated position in eukarya. Instead of U2698, U2697 is converted into a Ψ residue in *P. abyssi*. Altogether, this comparison of *P. abyssi* and eukaryal pseudouridylation in rRNAs suggests that conserved pseudouridylation sites in the large rRNA may have appeared early in evolution. Interestingly, one of the conserved Ψ residues in domain V (Ψ2603) is not guided by an H/ACA sRNA in *P. abyssi*.

A base substitution in sRNA Pab40 may generate a different specificity as compared to sRNAs Afu4, Pf7 and Pho40

Like the *P. abyssi* Pab40 sRNA, the Afu4, Pf7, and Pho40 sRNAs contain three H/ACA motifs. The target sites proposed for motifs 1 and 3 in Afu4 and Pf7 sRNAs (21,22) are identical to the ones determined experimentally for motifs 1 and 3 in Pab40 sRNA, respectively. In contrast, whereas motif 2 in the Afu4 and Pf7 sRNAs were proposed to guide modification at positions 2601 and 302 in the *A. fulgidus* and *P. furiosus* 23S rRNAs, respectively (corresponding to positions 2647 and 278 in *P. abyssi* 23S rRNA, respectively), we found that motif 2 in sRNA Pab40 is active at position 1932 in *P. abyssi* 23S rRNA. A difference of specificity between motif 2 in Pab40 and motifs 2 in Afu4, PFf7, and Pho40 sRNAs can be explained by a point mutation in the K-turn motif (a G to U substitution at position 85 as referred to the Pab40 numbering, Figure S1). Indeed, this base substitution modifies the secondary structure of motif 2 and its guiding specificity (Figure S1). Nevertheless, the target sites previously proposed for motifs 2 in Afu4 and Pf7 sRNAs are not in agreement with the structural rules that we established for active H/ACA sRNA–rRNA interactions. According to these rules, positions 2632 in the *A. fulgidus* 23S rRNA and 2701 in the *P. furiosus* 23S rRNA, that both correspond to position 2678 in *P. abyssi* 23S rRNA, are expected to be the true target

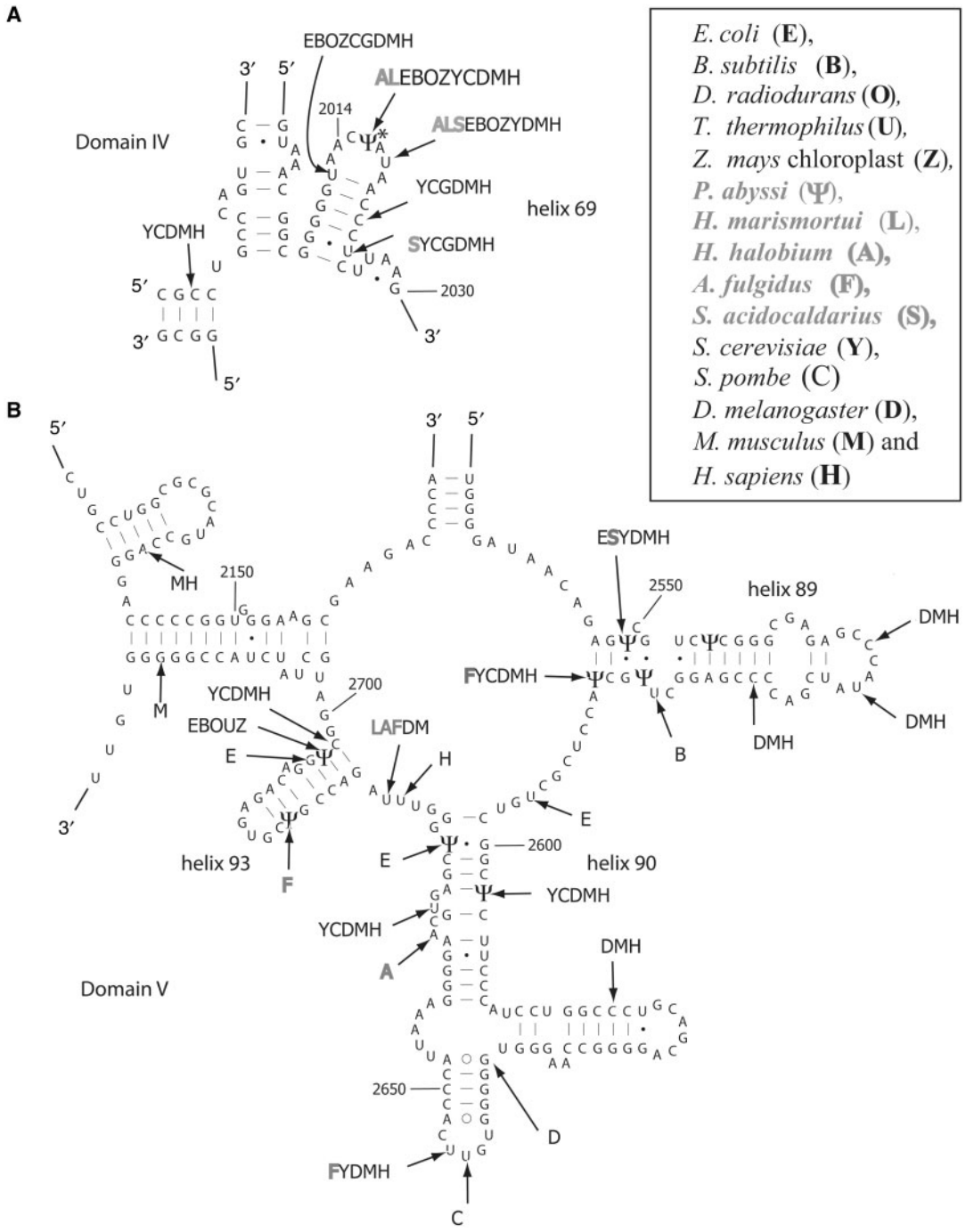


Figure 8. Location of Ψ residues in domains IV (A) and V (B) of the *P. abyssi* 23S rRNA. The secondary structure of the *P. abyssi* 23S rRNA is adapted from one of the *Thermococcus celer* 23S rRNA (M67497) (Gutell website, www.rna.icmb.utewas.edu). Numbering of residues is that of *P. abyssi* 23S rRNA. The Ψ residues detected in this work are indicated by Ψ symbol. The one in domain IV, which could not be detected in 23S rRNA because of an RT pause but was formed *in vitro*, is marked by an asterisk. Positions of pseudouridylation in *E. coli*, *B. subtilis*, *D. radiodurans*, *T. thermophilus*, *Z. mays* chloroplasts, *H. marismortui*, *H. halobium*, *A. fulgidus*, *S. acidocaldarius*, *S. cerevisiae*, *D. melanogaster*, *M. musculus* and *H. sapiens* are indicated by arrows marked by E, B, O, U, Z, L, A, F, S, Y, D M and H, respectively. Archaeal species are shown in grey.

sites (Figure S1). U 2678 is not converted into a Ψ residue in *P. abyssi*, however, this is the case in *H. marismortui* (35,36), *H. halobium* (34), *Drosophila melanogaster* and *Mouse musculus* (10) (Figure 8). The absence of modification at position 2678 in the *P. abyssi* 23S rRNA shows how a single point mutation can dramatically modify the specificity of an H/ACA motif.

Note that two other proposed target sites in *P. furiosus* rRNAs (corresponding to positions 892 in 16S rRNA and 2552 in *P. abyssi* 23S rRNA and which were expected to be guided by sRNAs Pf1 and Pf3 respectively, Table 1), are invalidated by our experimental rules. In addition, some *bona fide* target sites were not previously predicted in *P. furiosus* rRNAs (Table 1). Altogether, these data

strengthen the importance to verify experimentally the proposed H/ACA target sites.

Poor conservation of pseudouridylation sites between archaeal orders

Except for motif 2 in sRNA Pab40, H/ACA sRNAs and their target sites are conserved in the three *Pyrococcus* species studied as well as in the phylogenetically related *T. kodakarensis* species (25). In contrast, conservation of Ψ positions is poor between species of different orders. Only one of the three Ψ residues found in the *H. halobium* and *H. marismortui* 23S rRNAs and only one of the six Ψ residues found in the *S. acidocaldarius* 23S rRNA are conserved in *P. abyssi* (Figure 8). *A. fulgidus* is more closely related to *Pyrococcus* species than *Sulfolobus*, *Halobacterium*, and *Haloarcula* species. Accordingly, our experimental data show that three of the six Ψ residues detected in the *A. fulgidus* 23S rRNA are conserved in *P. abyssi* as well as one modified position in 16S rRNA.

The four Ψ residues detected in 16S rRNA are located in functional areas of the 30S subunit

The location of Ψ residues in the bacterial and eukaryal SSU rRNAs is rather variable from one species to the other. Interestingly, two of the four Ψ residues that we detected in the *P. abyssi* 16S rRNA (positions 27 and 891), are located within or very close to the essential central pseudoknot of the 16S rRNA (57). They may play a role in its formation or stability. This central pseudoknot is in close vicinity of the P site (58). Residue Ψ922 also belongs to a 16S rRNA segment located at the P site, and residue Ψ1017 belongs to a segment involved in A site formation. Hence, the four Ψ residues present in the *P. abyssi* 16S rRNA are located at or very near the A and P sites. Noticeably, no Ψ residue was detected in the 16S rRNA from *H. volcanii* while *S. solfataricus* 16S rRNA was estimated to contain five Ψ residues based on mass spectrometry analysis (59).

The aCBF5/aNOP10/aGAR1 complex may act on rRNA without guide sRNA

Among the small *P. abyssi* rRNA substrates that we used to test the activity of reconstituted sRNPs (Table S2), only that containing residue U2603 was modified in the absence of H/ACA sRNA (Figure 7). Therefore, the aCBF5/aNOP10/aGAR1 complex may be a specific catalyst for this position in 23S rRNA. The stem-loop structure formed by the small RNA substrate used in the assay is not the one expected to be formed in the 50S subunit (Figures 7 and 8). However, we cannot exclude the possibility that it is formed at some stage during 23S rRNA synthesis or 50S subunit assembly. Although a stable stem-loop structure could also be formed by the small substrate containing residue U2585 (Figure 7), the aCBF5/aNOP10/aGAR1 complex did not modify it. No activity was detected with the recombinant Pus10 enzyme (data not shown). Two other RNA: Ψ synthases are expected to be present in archaea, the *E. coli* TruA and TruD (Pus7) homologues (60). The implication of Pus7 is unlikely, since residue 2585 is not located in a sequence that fits the consensus sequence recognized by this enzyme (RSUNΨAR (R = purine, S = G/C, N = any nucleotide) (61 and our unpublished data). Therefore, it would be interesting to test the activity of the TruA homologue at position 2585 in *P. abyssi* 23S rRNA.

In agreement with the absence of *E. coli* RluE, RluB and RluC homologues in *P. abyssi*, the pseudouridylations at the three positions modified by these enzymes in *E. coli* are catalyzed by one H/ACA sRNA, Pab35.

Structural determinants for H/ACA sRNA specificity

Our data shed light on structural determinants of the sRNA-rRNA interaction that will be useful for the identification of new H/ACA sRNAs and their target sites. They can be explained taking structural data into account: (i) NMR analysis of the H/ACA snoRNA–rRNA interaction showed that the P1S and P2S intermolecular helices formed with the 3′- and 5′-guide sequences, respectively, can be coaxially stacked on helices H1 and H2 of the sRNA (32,33), and in this structure, the substrate is folded into a U-shape, with the targeted U residue protruding in the middle (32), (ii) based on the sRNP crystal structure, the H2-P2S pseudo-helix likely interacts with protein L7Ae bound to the K-turn or K-loop (22–24) and to a lesser extent with aNOP10 and aCBF5 through helix H2 (31), (ii) the various X-ray structures which have been established show that several interactions are formed between the aCBF5 and aNOP10 proteins (27–31) and the L7Ae and aNOP10 proteins (31) and (iv) finally, the H1-P1S pseudo-helix is expected to interact with protein aCBF5 bound to the ACA triplet (23,24,30,31). Therefore, it is highly conceivable that variations of the distances that separate the ACA box and the targeted uridine residue on the one hand (14 or 15-nt long in archaea, *versus* 16-nt long in eukarya) (62), and the K-turn and targeted U residue on the other hand (5 or 6 bps), disturb the positioning of aCBF5 relative to the targeted U residue and/or prevent some protein–protein interactions in the core protein structure (23,51). Importantly, we noticed that helix H2 never contains more than one bulged nucleotide.

Noticeably, for the first time, we show a greater importance of the rRNA interaction formed with the 3′-guide sequence (P1S), compared to that formed with the 5′-guide sequence (PS2). This difference can be explained by the close contact of aCBF5 with the 3′-guide sequence, compare to its limited contact with the 5′-guide sequence in the sRNP 3D structure (30). Remarkably, also the minimal length of the overall rRNA–sRNA base-pair interaction seems to be shorter in archaea compared to eukaryal system. This decreased stringency in archaea may explain why, in archaea but not in eukarya, a given H/ACA motif can guide Ψ formation at up to three different positions in rRNAs.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

S. Muller and J-B Fourmann were doctoral fellows supported by the French Ministère de la Recherche et des Nouvelles Technologies (MRNT). The work was supported by the Centre National de la Recherche Scientifique (CNRS), the MRNT and the Pôle de Recherche Scientifique et Technologique (PRST) « Bioingénierie » of Région Lorraine. Fujihiko Matsunaga and Patrick Forterre (UMR CNRS 8621, Orsay, France) are warmly thanked for their generous gift of *P. abyssi* cells. J. Ugolini is acknowledged for her efficient technical assistance. Funding to pay the Open Access publication charge was provided by CNRS.

Conflict of interest statement. None declared.

REFERENCES

1. Rozenski,J., Crain,P.F. and McCloskey,J.A. (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res.*, **27**, 196–197.
2. Arnez,J.G. and Steitz,T.A. (1994) Crystal structure of unmodified $tRNA^{Gln}$ complexed with glutaminyl- tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry*, **33**, 7560–7567.
3. Auffinger,P. and Westhof,E. (2001) An extended structural signature for the tRNA anticodon loop. *RNA*, **7**, 334–341.
4. Yu,Y.T., Shu,M.D. and Steitz,J.A. (1998) Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J*, **17**, 5783–5795.
5. Newby,M.I. and Greenbaum,N.L. (2002) Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat. Struct. Biol.*, **9**, 958–965.
6. Segault,V., Will,C.L., Sproat,B.S. and Luhrmann,R. (1995) In vitro reconstitution of mammalian U2 and U5 snRNPs active in splicing: Sm proteins are functionally interchangeable and are essential for the formation of functional U2 and U5 snRNPs. *EMBO J.*, **14**, 4010–4021.
7. Yang,C., McPheeters,D.S. and Yu,Y.T. (2005) Ψ35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **280**, 6655–6662.
8. Veldman,G.M., Klootwijk,J., de Regt,V.C., Planta,R.J., Branlant,C., Krol,A. and Ebel,J.P. (1981) The primary and secondary structure of yeast 26S rRNA. *Nucleic Acids Res.*, **9**, 6935–6952.
9. Decatur,W.A. and Fournier,M.J. (2002) rRNA modifications and ribosome function. *Trends Biochem. Sci.*, **27**, 344–351.
10. Ofengand,J. (2002) Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett.*, **514**, 17–25.
11. King,T.H., Liu,B., McCully,R.R. and Fournier,M.J. (2003) Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol. Cell*, **11**, 425–435.
12. Liang,X.H., Liu,Q. and Fournier,M.J. (2007) rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Mol. Cell*, **28**, 965–977.
13. Meroueh,M., Grohar,P.J., Qiu,J., SantaLucia,J.,Jr., Scaringe,S.A. and Chow,C.S. (2000) Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of *Escherichia coli* 23S rRNA. *Nucleic Acids Res.*, **28**, 2075–2083.
14. Mengel-Jorgensen,J., Jensen,S.S., Rasmussen,A., Poehlsgaard,J., Iversen,J.J. and Kirpekar,F. (2006) Modifications in *Thermus thermophilus* 23 S ribosomal RNA are centered in regions of RNA-RNA contact. *J. Biol. Chem.*, **281**, 22108–22117.
15. Sumita,M., Desaulniers,J.P., Chang,Y.C., Chui,H.M., Clos,L.,2nd and Chow,C.S. (2005) Effects of nucleotide substitution and modification on the stability and structure of helix 69 from 28S rRNA. *RNA*, **11**, 1420–1429.
16. Ejby,M., Sorensen,M.A. and Pedersen,S. (2007) Pseudouridylation of helix 69 of 23S rRNA is necessary for an effective translation termination. *Proc. Natl. Acad. Sci. USA*, **104**, 19410–19415.
17. Badis,G., Fromont-Racine,M. and Jacquier,A. (2003) A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *RNA*, **9**, 771–779.
18. Hamma,T. and Ferre-D'Amare,A.R. (2006) Pseudouridine synthases. *Chem. Biol.*, **13**, 1125–1135.
19. Kiss,T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
20. Decatur,W.A. and Fournier,M.J. (2003) RNA-guided nucleotide modification of ribosomal and other RNAs. *J. Biol. Chem.*, **278**, 695–698.
21. Tang,T.H., Bachellerie,J.P., Rozhdestvensky,T., Bortolin,M.L., Huber,H., Drungowski,M., Elge,T., Brosius,J. and Huttenhofer,A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. USA*, **99**, 7536–7541.
22. Rozhdestvensky,T.S., Tang,T.H., Tchirkova,I.V., Brosius,J., Bachellerie,J.P. and Huttenhofer,A. (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.*, **31**, 869–877.
23. Charpentier,B., Muller,S. and Branlant,C. (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res.*, **33**, 3133–3144.
24. Baker,D.L., Youssef,O.A., Chastkofsky,M.I., Dy,D.A., Terns,R.M. and Terns,M.P. (2005) RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev.*, **19**, 1238–1248.
25. Muller,S., Charpentier,B., Branlant,C. and Leclerc,F. (2007) A Dedicated Computational Approach for the Identification of Archaeal H/ACA sRNAs. *Methods Enzymol.*, **425**, 355–387.
26. Thebault,P., de Givry,S., Schiex,T. and Gaspin,C. (2006) Searching RNA motifs and their intermolecular contacts with constraint networks. *Bioinformatics*, **22**, 2074–2080.
27. Hamma,T., Reichow,S.L., Varani,G. and Ferre-D'Amare,A.R. (2005) The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat. Struct. Mol. Biol.*, **12**, 1101–1107.
28. Manival,X., Charron,C., Fourmann,J.B., Godard,F., Charpentier,B. and Branlant,C. (2006) Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res.*, **34**, 826–839.
29. Rashid,R., Liang,B., Baker,D.L., Youssef,O.A., He,Y., Phipps,K., Terns,R.M., Terns,M.P. and Li,H. (2006) Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol. Cell*, **21**, 249–260.
30. Liang,B., Xue,S., Terns,R.M., Terns,M.P. and Li,H. (2007) Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. *Nat. Struct. Mol. Biol.*
31. Li,L. and Ye,K. (2006) Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature*, **443**, 302–307.
32. Wu,H. and Feigon,J. (2007) H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *Proc. Natl Acad. Sci. USA*, **104**, 6655–6660.
33. Jin,H., Loria,J.P. and Moore,P.B. (2007) Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Mol. Cell*, **26**, 205–215.
34. Ofengand,J. and Bakin,A. (1997) Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. *J. Mol. Biol.*, **266**, 246–268.
35. Kirpekar,F., Hansen,L.H., Rasmussen,A., Poehlsgaard,J. and Vester,B. (2005) The archaeon *Haloarcula marismortui* has few modifications in the central parts of its 23S ribosomal RNA. *J. Mol. Biol.*, **348**, 563–573.
36. Del Campo,M., Recinos,C., Yanez,G., Pomerantz,S.C., Guymon,R., Crain,P.F., McCloskey,J.A. and Ofengand,J. (2005)

Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of *Haloarcula marismortui* and *Deinococcus radiodurans*. *RNA*, **11**, 210–219.
37. Massenet,S., Ansmant,I., Motorin,Y. and Branlant,C. (1999) The first determination of pseudouridine residues in 23S ribosomal RNA from hyperthermophilic Archaea *Sulfolobus acidocaldarius*. *FEBS Lett*., **462**, 94–100.
38. Klein,R.J., Misulovin,Z. and Eddy,S.R. (2002) Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc. Natl Acad. Sci. USA*, **99**, 7542–7547.
39. Zago,M.A., Dennis,P.P. and Omer,A.D. (2005) The expanding world of small RNAs in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol. Microbiol*., **55**, 1812–1828.
40. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*., **22**, 4673–4680.
41. Besemer,J., Lomsadze,A. and Borodovsky,M. (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res*., **29**, 2607–2618.
42. Gautheret,D., Major,F. and Cedergren,R. (1993) Modeling the three-dimensional structure of RNA using discrete nucleotide conformational sets. *J. Mol. Biol*., **229**, 1049–1064.
43. Gautheret,D. and Lambert,A. (2001) Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J. Mol. Biol*., **313**, 1003–1011.
44. Charbonnier,F., Erauso,G., Barbeyron,T., Prieur,D. and Forterre,P. (1992) Evidence that a plasmid from a hyperthermophilic archaebacterium is relaxed at physiological temperatures. *J Bacteriol*, **174**, 6103–6108.
45. Chomczynski,P. and Sacchi,N. (1986) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem*., **162**, 156–159.
46. Bakin,A. and Ofengand,J. (1993) Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center – analysis by the application of a new sequencing technique. *Biochemistry*, **32**, 9754–9762.
47. Motorin,Y., Muller,S., Behm-Ansmant,I. and Branlant,C. (2007) Identification of modified residues in RNAs by reverse transcription-based methods. *Methods Enzymol*., **425**, 21–53.
48. Charpentier,B., Fourmann,J.B. and Branlant,C. (2007) Reconstitution of archaeal H/ACA sRNPs and test of their activity. *Methods Enzymol*., **425**, 389–405.
49. Schattner,P. (2002) Searching for RNA genes using base-composition statistics. *Nucleic Acids Res*., **30**, 2076–2082.
50. Roovers,M., Hale,C., Tricot,C., Terns,M.P., Terns,R.M., Grosjean,H. and Droogmans,L. (2006) Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic Acids Res*., **34**, 4293–4301.
51. Muller,S., Fourmann,J.B., Loegler,C., Charpentier,B. and Branlant,C. (2007) Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA: Ψ55-synthase and RNA-guided RNA:Ψ-synthase activities. *Nucleic Acids Res*.
52. Gurha,P., Joardar,A., Chaurasia,P. and Gupta,R. (2007) Differential roles of archaeal box H/ACA proteins in guide RNA-dependent and independent pseudouridine formation. *RNA Biol*, **4**, 101–109.
53. Dennis,P.P., Omer,A. and Lowe,T. (2001) A guided tour: small RNA function in Archaea. *Mol. Microbiol*., **40**, 509–519.
54. Harms,J., Schluenzen,F., Zarivach,R., Bashan,A., Gat,S., Agmon,I., Bartels,H., Franceschi,F. and Yonath,A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, **107**, 679–688.
55. Yusupov,M.M., Yusupova,G.Z., Baucom,A., Lieberman,K., Earnest,T.N., Cate,J.H. and Noller,H.F. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science*, **292**, 883–896.
56. Branlant,C., Krol,A., Machatt,M.A., Pouyet,J., Ebel,J.P., Edwards,K. and Kossel,H. (1981) Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. *Nucleic Acids Res*., **9**, 4303–4324.
57. Juzumiene,D.I. and Wollenzien,P. (2001) Arrangement of the central pseudoknot region of 16S rRNA in the 30S ribosomal subunit determined by site-directed 4-thiouridine crosslinking. *RNA*, **7**, 71–84.
58. Bullard,J.M., van Waes,M.A., Bucklin,D.J., Rice,M.J. and Hill,W.E. (1998) Regions of 16S ribosomal RNA proximal to transfer RNA bound at the P-site of Escherichia coli ribosomes. *Biochemistry*, **37**, 1350–1356.
59. Noon,K.R., Bruenger,E. and McCloskey,J.A. (1998) Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus*. *J. Bacteriol*., **180**, 2883–2888.
60. Watanabe,Y. and Gray,M.W. (2000) Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic Acids Res*., **28**, 2342–2352.
61. Behm-Ansmant,I., Urban,A., Ma,X., Yu,Y.-T., Motorin,Y. and Branlant,C. (2003) The *Saccharomyces cerevisiae* U2 snRNA: pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA: Ψ-synthase also acting on tRNAs. *RNA*, **9**, 1371–1382.
62. Ganot,P., Caizergues-Ferrer,M. and Kiss,T. (1997) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev*., **11**, 941–956.

SUPPLEMENTARY DATA

Table S1. The oligonucleotides used in this study

Panel A. Oligonucleotides used to test the expression of the predicted H/ACA sRNAs in *P. abyssi* total RNA extract by Northern blot analysis. The names of the tested sRNA, the 5'-3' sequence of the oligonucleotides and their hybridization positions in the sRNAs are given. In panel B, the sequences of the sense and antisense oligonucleotides used to amplify the sRNA template sequences are shown. Sequences of T7 RNA polymerase promoters are underlined. Panel C shows the oligonucleotide primers that were used for identification of the CMCT modifications in *P. abyssi* 16S and 23S rRNAs by reverse transcription.

Table S2. The rRNA fragments used as substrates of the reconstituted H/ACA sRNPs or the aCBF5-aNOP10-aGAR1 complex

Positions of the putative targetted U residues in 16S and 23S rRNAs are given in the first column. The sequences of the RNA substrates used in the assays are given in column 2. The 3 G residues at the 5' end are generated by the T7 promoter sequence. The putative targetted U residues are in bold. The two orphan Ψ residues with no corresponding sRNA are marked by asterisks.

Figure S1. How a point mutation in an H/ACA sRNA can modify its target specificity

Panel A. the 2D structure that we proposed for motif 2 in sRNA Pho40 from *P. horikoshii* is shown. Base substitutions in sRNAs Pf7 from *P. furiosus,* Afu4 from *A. fulgidus* and Tko1 from *T. kodakarensis,* preserve the possibility to form this 2D

structure. Panel B, in contrast, due to a G to U substitution, the K-turn motif present in Pho40, Pf7, Afu4 and Tko1 sRNAs cannot be formed in sRNA Pab40. One stable 2D structure can be proposed for motif 2 in this sRNA, however, it contains no K-turn or K-loop (structure a). The only structure containing a K-turn which can be proposed (structure b) has a lower thermodynamic stability. However, its formation may be favoured by binding of protein L7Ae. According to the proposed folding for motifs 2 in sRNAs Pho40, Pf7, Tko1 and Afu4, these motifs can guide pseudouridylations at positions 2673, 2701, 2682 and 2632 in their respective 23S rRNA targets (Panel C). When folded according to structure a, motif 2 in sRNA Pab40 can target residue 1932 in the *P. abyssi* 23S rRNA (Panel D).

A

Northern Blotting		
sRNA	**5' - 3' Sequence**	**Hybridization position in the H/ACA sRNA**
Pab21	CGGAACCCCCGGTTCACGAT	47-28
Pab35	TTGATCCCGCGGGCGGATGCA	127-107
Pab40	AGGGTTACCCCCGGTCACGGC	130-110
Pab91	GACCGCCCTCCGGTCACTGATC	45-24
Pab105	CGACGTCGCCTCGTTCATG	64-46
Pab160	GGAGCGAAGGGGAGTCATAGCT	62-41

B

	PCR for Cloning	
sRNA	**sense oligo (5' - 3' sequence)**	**antisense oligo (5' - 3' sequence)**
Pab19	TAATACGACTCACTATAGGGAAAGCCCAGGGGTCAA	TTCTATACCCAGGGGAGGAG
Pab21	TAATACGACTCACTATAGGGGATGGGCCCGGCTCCCGCCCTCTCC	ACTGTAGGCCCGGCCGGAACCCC
Pab35	TAATACGACTCACTATAGGGGCTCGGTCTACCC	TGTAGGCCGGCAACAGA
Pab40	TAATACGACTCACTATAGGGTATGCCCCCGCAAGCGAGGGCCTGG	ATCTGTAGCCCGCCCGTAACACC
Pab91	TAATACGACTCACTATAGGGTAATGCCCCTCCCCTCTC	AGGTGTCCCCTCCCCGAC
Pab105	TAATACGACTCACTATAGGGACCTGCCGCCCGGA	TATAGGCCCGGCTTATCCA
Pab160	TAATACGACTCACTATAGGGTGGCCCCCGGAAACCG	AAAAATTGTGCCCCCGGG

C

RT-CMCT analysis		
Hybridization position	**5' - 3' Sequence**	**Area analysed**
16S rRNA 57-76	CCCTTGACTCGCATGGCTTA	16S rRNA 20-55
16S rRNA 958-939	CGGTGAGGTTCCCGGCGTTG	16S rRNA 859-926
16S rRNA 1065-1044	ACCACACTTAAGTGGACGCCTC	16S rRNA 961-1038
16S rRNA 1176-1154	CTGACCTACCGTCGCCCGCCCCT	16S rRNA 1089-1147
23S rRNA 327-309	GTTCCAGCGGACTTCGGCT	23S rRNA 254-298
23S rRNA 1409-1389	ATTCTCACCCGACTACGCTG	23S rRNA 1313-1386
23S rRNA 1989-1971	CACCCGGTTGTACCCGGGC	23S rRNA 1918-1965
23S rRNA 2135-2113	CATGCGCGCCAGGCAGAGGTTTC	23S rRNA 2018-2100
23S rRNA 2285-2265	AGGAGTGGGCGGTGTCTCATG	23S rRNA 2224-2261
23S rRNA 2544-2523	GTTATCCCCAGGGTAGCTTTTC	23S rRNA 2420-2515
23S rRNA 2593-2568	GAGGTAGCAAGCCTCGGGGTCGATA	23S rRNA 2515-2566
23S rRNA 2667-2647	CGTTCCCCTTTAATGGGTGA	23S rRNA 2565-2640
23S rRNA 2722-2704	CAACACCCCCGGTAGATAG	23S rRNA 2640-2700
23S rRNA 2860-2841	GCCGCTTCGGCCTTAGATGC	23S rRNA 2760-2835
23S rRNA 3029-3008	GGGGGACAGGACCTCGGGCGA	23S rRNA 2923-3000

Table S1

RNA Substrates for *in vitro* assays	
rRNA targets	**5' - 3' sequence**
16S rRNA 27	GGGAUCC**U**GCCGGAGGCCAC
16S rRNA 891 and 892	GGGAAGGAA**UU**GGCGGGGGAG
16S rRNA 922	GGGUGGAGCG**U**GCGGUUUA
16S rRNA 995	GGGAGGUCU**U**GCCGGACG
16S rRNA 1017	GGGCCGAGAGGAGG**U**GCAUGGCCGCCGUCA
16S rRNA 1122	GGGAGGCAC**U**CUGGGGG
23S rRNA 278	GGGUGGGGUUG**U**AGGGCCCCCG
23S rRNA 1377	GGGCGGAUCC**U**GCCGGGAGU
23S rRNA 1932	GGGAAAGGGUU**U**GUACUGGG
23S rRNA 2016	GGGGGUAAC**U**AUAACCCUC
23S rRNA 2250	GGGCUCCGGG**U**CGGGUGG
23S rRNA 2549/2552/2554	GGGAUAACAGAG**U**CG**U**C**U**CGGGCGAGA
23S rRNA 2585***** and 2588	GGGAGGCU**U***GC**U**ACCUCGCUG
23S rRNA 2603*****	GGGCUGUCGGC**U***CUUCCCAU
23S rrRNA 2672	GGGUGAGC**U**GGGUUUA
23S rRNA 2685	GGGUAGACCG**U**CGUGAGACAGGUCGG
23S rRNA 2697	GGGAGACAGG**U**CGGAUGCUA
23S rRNA 2794	GGGUUGUCC**U**CCCGGGCA
23S rrRNA 2930	GGGAGGGC**U**CGGGUAGAA

Table S2

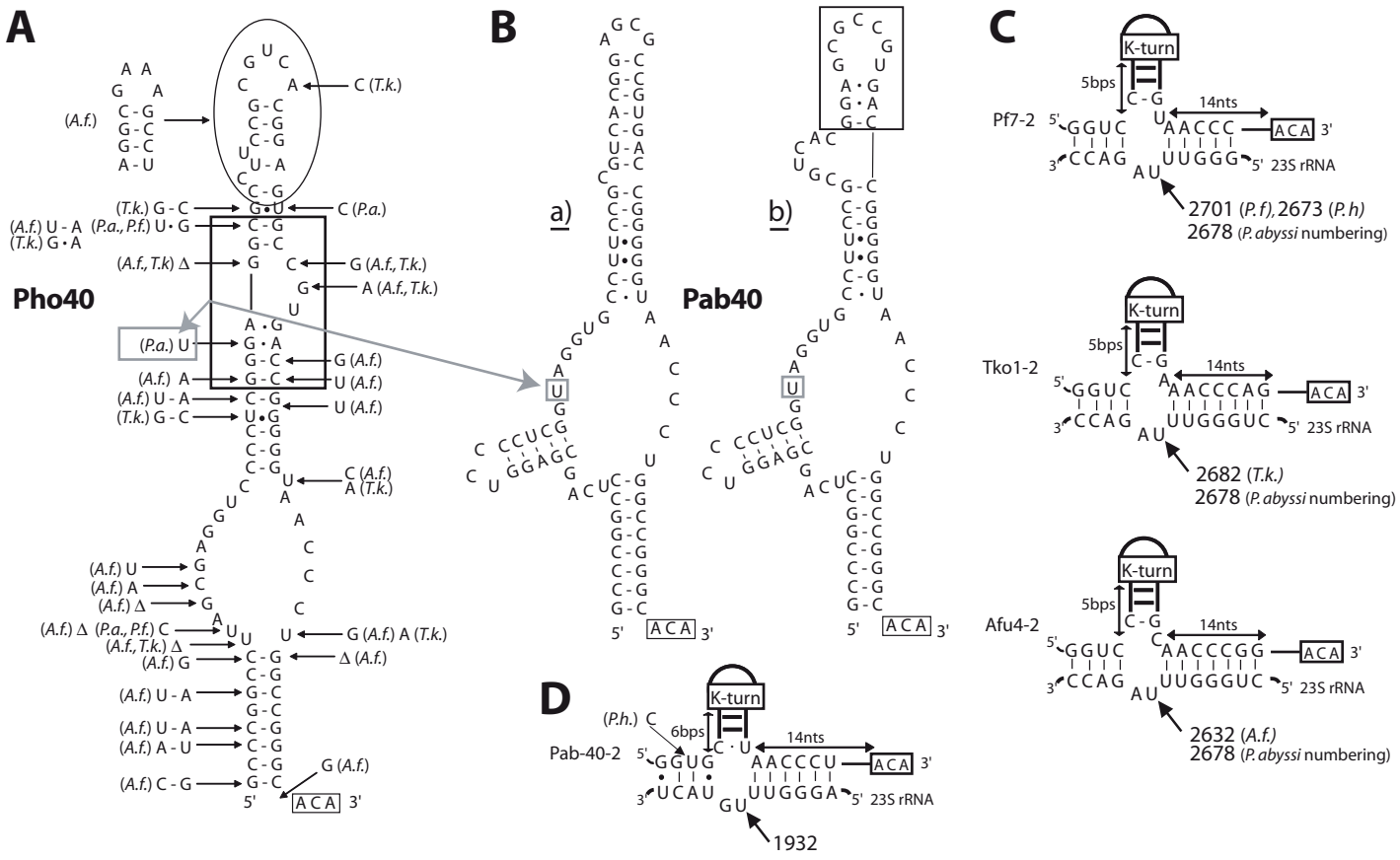

A
B
C
D
Pho40
Pab40
a)
b)
(A.f.)
(T.k.) G - C
(A.f.) U - A
(P.a., P.f.) U • G
(T.k.) G • A
(A.f., T.k) Δ
(P.a.) U
(A.f.) A
(A.f.) U - A
(T.k.) G - C
(A.f.) U
(A.f.) A
(A.f.) Δ
(A.f.) Δ
(P.a., P.f.) C
(A.f., T.k.) Δ
(A.f.) G
(A.f.) U - A
(A.f.) U - A
(A.f.) A - U
(A.f.) C - G
C (T.k.)
C (P.a.)
G (A.f., T.k.)
A (A.f., T.k.)
G (A.f.)
U (A.f.)
U (A.f.)
C (A.f.)
A (T.k.)
G (A.f.) A (T.k.)
Δ (A.f.)
G (A.f.)
ACA 3'
5'
K-turn
5bps
14nts
Pf7-2
Tko1-2
Afu4-2
23S rRNA
2701 (P. f), 2673 (P. h)
2678 (P. abyssi numbering)
2682 (T.k.)
2678 (P. abyssi numbering)
2632 (A.f.)
2678 (P. abyssi numbering)
(P.h.) C
6bps
Pab-40-2
1932

Figure S1

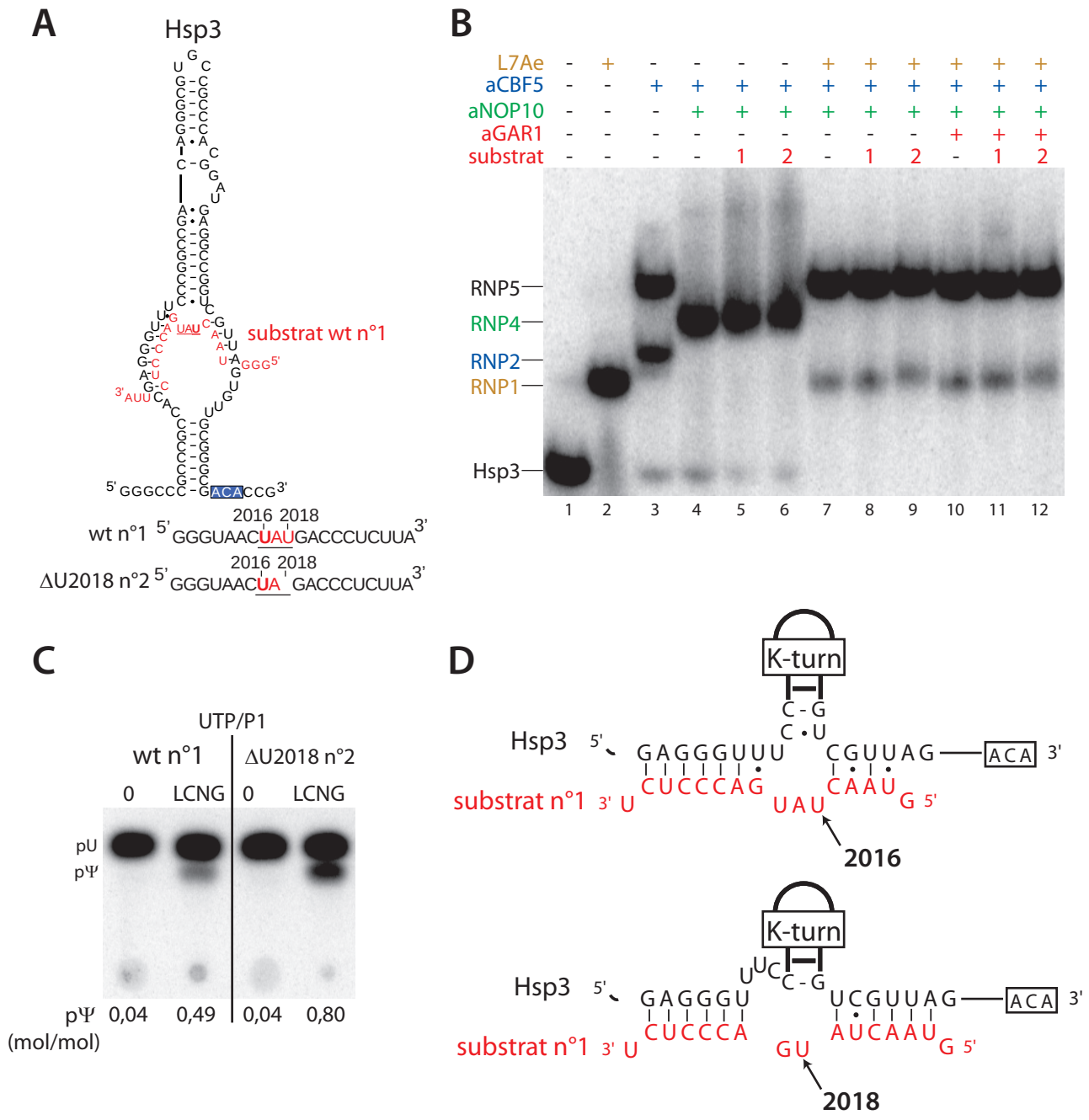


Figure r1 : Etude de l'assemblage et de l'activité de la sRNP Hsp3.

A. Structure schématique du sRNA Hsp3 à boîtes H/ACA de *H. Halobium sp* apparié au substrat sauvage n°1 qui comporte 3 nucléotides non appariés à la poche de pseudouridylation dont le résidu cible correspond à la position 2016 de l'ARNr 25S. Le substrat appelé ΔU2018 n°2 correspond à un substrat non naturel dans lequel la position U2018 a été supprimée de façon à ce que l'hybridation de ce substrat avec le sRNA Hsp3 induise la formation conventionnelle de 2 nucléotides non appariés à la poche de pseudouridylation. Les nucléotides non appariés sont en rouge et soulignés. Le résidu cible de la réaction est en gras.

B. Résultats d'expériences de retard de migration sur gel réalisées en présence du sRNA Hsp3 radiomarqué en 5' et différentes combinaisons de protéines ou de substrats indiquées.

C. Tests d'activités réalisés en présence du sRNA Hsp3, des protéines indiquées et des substrats wt n°1 ou ΔU2018 n°2 uniformément radiomarqués à l'UTP, pendant 80 min à 65°C. Après digestion à la RNase P1, seuls les résidus U et Ψ sont visualisés par chromatographie après migration en 1 dimension.

D. Schématisation de l'appariement entre le sRNA Hsp3 et son substrat selon une forme conventionnelle (U2016 cible) et selon un glissement qui pourrait s'opérer (U2018 cible).

2. Résultats complémentaires

2.1. Etude *in vitro* de l'activité du sRNA Hsp3 de *H. halobium*

Les uridines localisées aux positions 2016 et 2018 dans l'ARNr 25S de *H. halobium* sont converties toutes deux en pseudouridines. Les séquences des nucléotides au voisinage de ces positions ne correspondent à aucune des séquences reconnues par les enzymes de modification connues. S. Muller a donc recherché par bioinformatique des candidats sRNA qui pouvaient guider ces modifications. Le seul sRNA détecté chez *H. halobium* pouvant guider ces modifications est le sRNA Hsp3 (Fig. r1 A). La prédiction de la modification U2016 guidée par le sRNA Hsp3 indiquait qu'elle reposait sur un appariement avec 3 résidus (5'-U_{2016}AU-3') non appariés à la poche de pseudouridylation. Ainsi au moins deux sRNA identifiés chez deux archaea, Pab19 chez *P. abyssi* (article n°1, (Muller *et al.*, 2008)) et Hsp3 chez *H. halobium*, pouvaient guider la modification d'un résidu U par le biais d'une interaction conduisant à un trinucléotide 5'-UNN-3' en simple brin. Nous avons voulu tester *in vitro* l'activité du sRNA Hsp3.

Nous avons cloné puis transcrit ce sRNA ainsi que deux ARN substrats. Le substrat n°1 correspond à un fragment de l'ARNr 25S portant la séquence naturelle des positions 2011 à 2028 et son appariement au sRNA doit conduire à la formation des 3 nucléotides non appariés à la poche de pseudouridylation (5'-U_{2016}AU_{2018}-3') (Fig. r1 A). Le substrat n°2 correspond à un substrat variant avec une délétion du résidu U2018 permettant de retrouver un appariement conventionnel à 2 résidus non appariés à la poche de pseudouridylation (5'-U_{2016}A-3') (Fig. r1 A).

Ne disposant pas de protéines recombinantes de *H. halobium*, nous avons donc reconstitué la sRNP Hsp3 à partir des protéines de *P. abyssi* en incubant différentes combinaisons de ces protéines avec le transcrit sRNA Hsp3 radiomarqué, et analysé la formation des RNP par retard de migration sur gel (Fig. r1 B). Nous constatons que i) les différents complexes retardés RNP1, RNP2, RNP4 et RNP5 sont obtenus (Fig. r1 B), ii) que les protéines de *P. abyssi* s'assemblent sur le sRNA Hsp3 mais deux points sont à souligner. Le premier concerne la fixation de aCBF5 à Hsp3 qui conduit à la

formation de deux complexes retardés alors qu'un seul site de fixation est présent (Fig. r1 B piste 3). Dans toutes nos expériences réalisées au laboratoire, seuls les ARNt et les sRNA à double motif tige-boucle ont conduit à la formation de deux complexes suite à la fixation de aCBF5. Il est possible que dans le cas du sRNA Hsp3 formé d'une seule tige-boucle, cela provienne de l'utilisation d'une protéine aCBF5 d'une autre espèce que celle de *H. halobium*, ou que aCBF5 reconnaisse deux conformations de Hsp3 qui seraient en équilibre en solution. Néanmoins, dès l'ajout de aNOP10, un seul complexe retardé est détecté, indiquant que le complexe aCBF5-aNOP10 conduit à la formation d'une population homogène de RNP (Fig. r1 B piste 4). Un seul complexe majoritaire RNP5 est formé en présence des protéines aCBF5-aNOP10 et L7Ae (LCN) dont la quantité et la migration électrophorétique est inchangée en présence des deux ARN substrats (Fig. r1 B, pistes 7-9). La présence additionnelle de aGAR1 (combinaison LCNG) n'a pas d'effet et le complexe RNP5 est toujours présent (Fig. r1 B, pistes 10-12).

Nous avons radiomarqué uniformément à l'[α-^{32}P]UTP chacun des 2 substrats, et incubé ceux-ci avec le mélange de protéines LCNG pendant 80 min à 65°C puis digéré les ARN par traitement à la RNase P1 qui libère des nucléotides 5' monophosphate, ce qui permet de visualiser après chromatographie en une dimension tous les résidus U et Ψ indépendamment de leur positionnement au sein de la séquence de l'ARN substrat (Fig. r1 C). Nous constatons que le substrat n°1 est modifié à un taux de 49% confirmant que l'appariement à 3 résidus peut conduire à la réaction de pseudouridylation. Le substrat n°2 est modifié à un taux 1,6 fois plus élevé (Fig. r1 C). Nous concluons que le sRNA Hsp3 en association avec les protéines de *P. abyssi* est capable de guider la modification d'un substrat à 3 résidus non appariés à la poche de pseudouridylation et que le même substrat qui ne possède plus que 2 nucléotides non appariés est modifié à un taux plus élevé. Ces résultats rejoignent ceux obtenus avec le sRNA Pab19 de *P. abyssi* (article n°1, (Muller *et al.*, 2008)).

Dans le cas de Hsp3, la modification de la position U2018 reste énigmatique. En particulier, la question de savoir si *in vitro,* le sRNA Hsp3 peut guider à lui seul la modification des uridines des positions 2016 et 2018 reste sans réponse. Il est possible que le substrat puisse s'hybrider au sRNA selon un mode d'appariement alternatif conduisant à la formation du dinucléotide (5'-U_{2018}G-3') en simple brin (Fig. r1 D). Ainsi, les positions U2016 et U2018 pourraient être modifiées à partir du même sRNA Hsp3 mais par deux types d'appariements.

2.2. Influence sur l'activité de la sRNP Pab91 de l'identité des résidus non appariés et contigus au résidu U ciblé

Nous nous sommes intéressés au rôle potentiel du résidu conventionnellement non apparié qui est adjacent en 3' du résidu U cible de la réaction. Nous avons testé l'effet de la présence des 4 nucléotides possibles à cette position sur l'activité des particules sRNP Pab91 (Fig. r2 A). Nous constatons que quelque soit la nature du nucléotide, l'activité de la sRNP est élevé (>80%). Il apparaît toutefois que la présence d'un résidu pyrimidique soit légèrement plus favorable à la réaction de modification du résidu contigu. Il est possible que le plus faible encombrement stérique des résidus pyrimidiques soit un déterminant favorable à la réaction de modification. (Fig. r2 A). Nous concluons que la nature du résidu non apparié en 3' du résidu U ciblé n'influe pas fortement sur l'activité *in vitro* de la sRNP.

Nous avions montré que *P. abyssi* et *H. halobium* produisaient des sRNA capables de modifier des positions dans les ARNr dont l'appariement à la poche de pseudouridylation conduit à la formation de 3 résidus non appariés. Les sRNA en question, Pab19 et Hsp3, gardent la capacité de modifier des ARN substrats qui s'apparient à leur poche de pseudouridylaion par un appariement canonique laissant 2 résidus non appariés (article n°1, (Muller *et al.*, 2008) et Fig. r1 A et C). Nous nous sommes demandé si un sRNA modifiant sa séquence cible par le biais d'un appariement de type canonique (5'-UN-3') avec deux résidus non appariés, comme

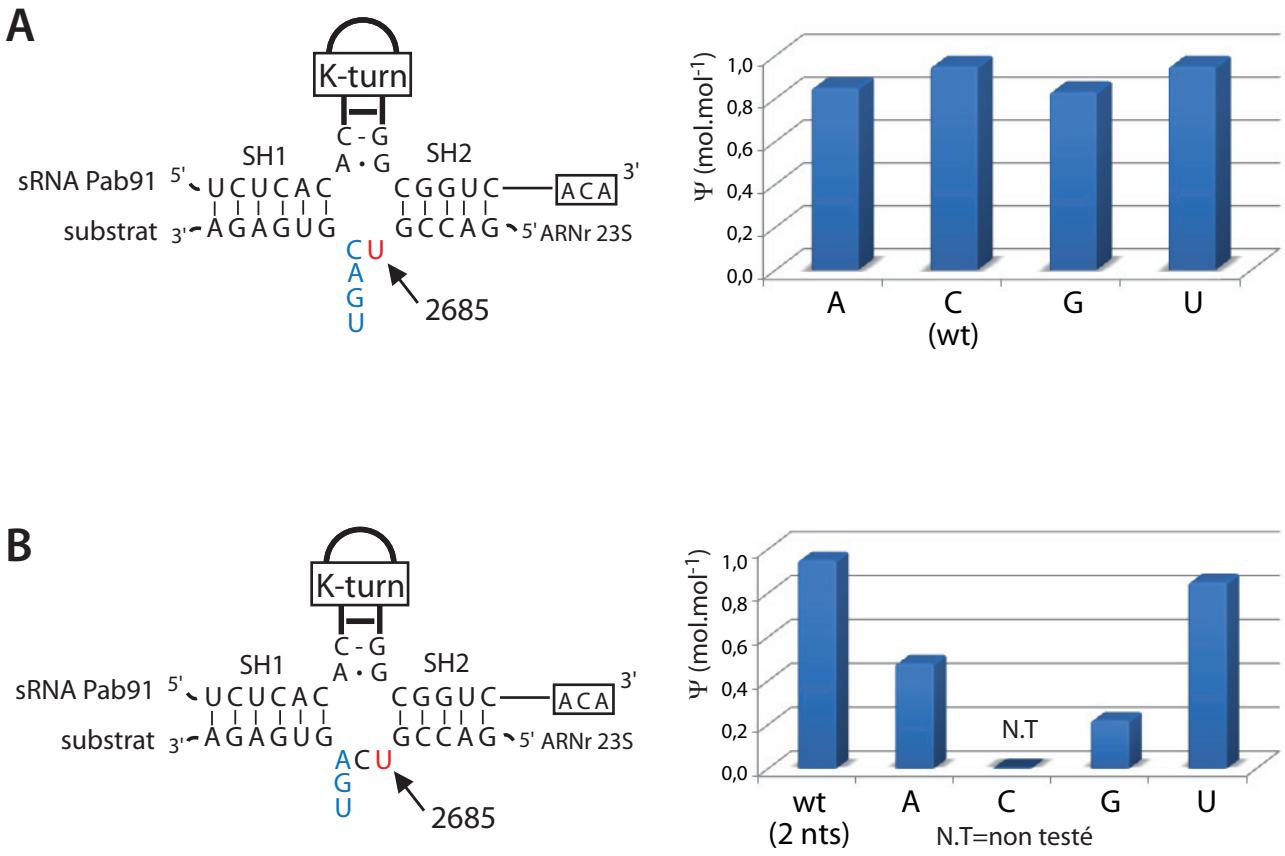


Figure r3 : Etude du rôle de l'identité de résidus ne s'appariant pas en 3' du résidu modifié du substrat du sRNA Pab91.

A. Importance du second nucléotide non apparié dans la séquence 5'-UN-3'.

A gauche, structure schématique des interactions réalisées entre le sRNA Pab91 de *P. abyssi* et divers substrats qui ciblent la position 2685 (en rouge) de l'ARNr 23S. Les nucléotides en bleu correspondent à ceux qui sont substitués. Le substrat naturel possède un résidu C en tant que nucléotide non apparié à la poche de pseudouridylation ; les autres substitutions correspondent à des substrats non naturels. A droite, résultats des tests d'activité réalisés en présence de Pab91, ses substrats dont le nucléotide modifié est indiqué, les protéines LCNG, pendant 80 min à 65°C.

B. Importance du dernier résidu non appariée à la poche de pseudouridylation dans la séquence 5'-UCN-3'.

A gauche, structure schématique des interactions réalisées entre le sRNA Pab91 de *P. abyssi* et divers substrats non naturels qui ciblent la position 2685 (en rouge) de l'ARNr 23S et qui contiennent 3 nucléotides non appariés à la poche de pseudouridylation. Les nucléotides en bleu correspondent à ceux qui sont substitués. A droite, résultats des tests d'activité réalisés en présence de Pab91, ses substrats dont le nucléotide modifié est indiqué, les protéines LCNG, pendant 80 min à 65°C. Il est à noter que l'insertion du résidu C n'a pas été testée.

par exemple le sRNA Pab91 employé comme modèle au laboratoire, avait la possibilité de modifier une séquence ARN avec trois résidus non appariés.

La capacité de Pab91 à guider la modification de la position U2685 a été testée avec plusieurs substrats dont l'identité du troisième résidu non apparié dans la séquence 5'-U2685CN-3' était variable (Fig. r2 B). L'insertion d'un résidu U ne modifie que très peu l'activité de la sRNP, 85% contre 95% pour le substrat sauvage à 2 nucléotides non appariés (Fig. r2 B). La présence d'un résidu A ou G conduit à une baisse de l'activité avec un taux de modification de 48% et 22% respectivement (Fig. r2 B). La diminution de l'activité est probablement due à l'encombrement stérique plus important de ces purines par rapport aux pyrimidines. De plus, nous avons également observé que le sRNA Pab160, modifiant naturellement la position U922 par un appariement canonique, est capable de modifier un substrat contenant les 3 nucléotides 5'-U922GU-3' non appariés et ce, avec une activité comparable à celle obtenue pour son substrat sauvage à 2 nucléotides non appariés 5'-U922G-3' (données non montrées).

Nous concluons qu'un sRNA qui modifie naturellement un substrat à partir d'un appariement canonique à 2 nucléotides non appariés est capable de modifier un substrat à 3 résidus non appariés mais que l'activité dépend alors de l'identité du résidu situé à la 3e position du trinucléotide, la présence d'une base pyrimidique étant plus favorable. Par conséquent, nous proposons que les sRNA à boîtes H/ACA d'archaea seraient capables de guider la modification de substrats à 3 résidus non appariés à la poche de pseudouridylation du type 5'-UNY-3' (Y pour pyrimidine). Il serait intéressant de vérifier si les snoRNA eucaryotes ont la possibilité de modifier des substrats à trois résidus 5'-UNN-3' non appariés à la poche de pseudouridylation ou si cette faculté est restreinte aux sRNA d'archaea. Les techniques de génétique de la levure pourraient permettre de tester cette hypothèse.

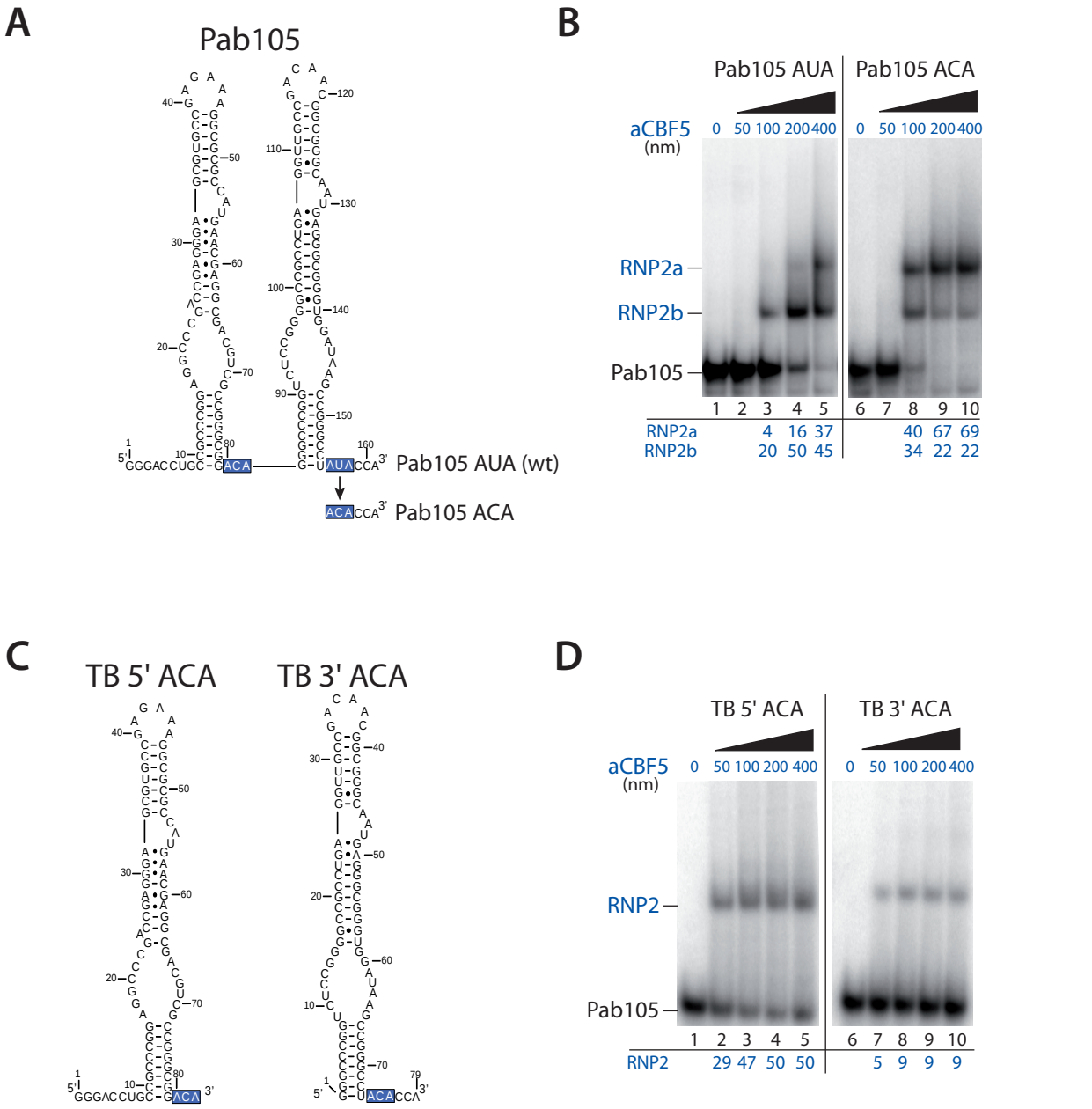


Figure r3 : Etude par expériences de retard de migration sur gel de l'influence de chacune des boîtes ACA du sRNA Pab105 sur la fixation directe de la protéine aCBF5.

A. Structure schématique du sRNA Pab105 à boîtes H/ACA de *P. abyssi* dont la séquence sauvage de la tige-boucle 3' porte la séquence AUA à la place de la boîte ACA. Cette séquence est substituée par la séquence ACA dans le variant nommé Pab105 ACA.

B. Résultats d'expériences de retard de migration sur gel réalisées en présence de variants du sRNA Pab105 radiomarqués en 5' et de aCBF5. Le taux de complexe formé entre une tige-boucle du sRNA et aCBF5 (RNP2b) ou entre les deux tige-boucles du sRNA et aCBF5 (RNP2a) est indiqué sous chaque piste.

C. Structure schématique des deux tige-boucles du sRNA Pab105 à boîtes H/ACA de *P. abyssi* produites séparément par transcription *in vitro*.

D. Résultats d'expériences de retard de migration sur gel réalisées en présence de l'une ou de l'autre des tige-boucles du sRNA Pab105 radiomarquées en 5' et de aCBF5. Le taux de complexe formé entre les tige-boucles du sRNA et aCBF5 (RNP2) est indiqué sous chaque piste.

B. Recherche de déterminants ARN contribuant à l'affinité de aCBF5 pour les sRNA à boîtes H/ACA

1. Contexte

La boîte ACA et la poche pseudouridylation sont deux éléments indispensables au sRNA pour son association avec aCBF5 (Baker *et al.*, 2005; Charpentier *et al.*, 2005). Certains sRNA ne portent néanmoins pas la séquence canonique ACA et la séquence AUA est parfois présente dans des sRNA à deux motifs en tige boucle (article n°1, (Muller *et al.*, 2008)). J'ai donc voulu comparer le rôle fonctionnel des séquences ACA et AUA. De plus, nous avons montré que aCBF5 se liait à des sRNA portant une séquence ACA mais avec une affinité variable selon le sRNA utilisé (article n°4, Fourmann *et al.,* en préparation). Ainsi, le contexte environnant pouvait moduler le rôle fonctionnel de cet élément. J'ai donc recherché des déterminants moléculaires autres qui pourraient être requis pour une fixation efficace de aCBF5 à certains sRNA à boîtes H/ACA. En particulier, j'ai comparé 2 sRNA de *P. abyssi*, Pab21 pour lequel un fort taux de complexe est formé avec aCBF5 et Pab91 pour lequel aCBF5 possède une faible affinité apparente.

Enfin, la séquence 5'-GAG-3' située dans la boucle apicale du sRNA Pf9 a été proposée servir de point d'ancrage supplémentaire à aCBF5 mais seulement au sein du sous-complexe sRNA-aCBF5. En effet, l'interaction serait perdue lors de l'association des autres protéines à la sRNP (Baker *et al.*, 2005; Youssef *et al.*, 2007; Baker *et al.*, 2008). J'ai donc testé l'effet de cette séquence lorsqu'elle était ajoutée dans des sRNA utilisés comme modèle au laboratoire et qui ne la portaient pas naturellement.

2. Résultats

2.1. L'importance de la boîte ACA dans les sRNA

Nous avons débuté cette étude en employant le sRNA Pab105 qui s'organise en un double motif comprenant une tige-boucle 5' suivie en 3' d'une séquence ACA et d'une tige-boucle 3' terminée par une séquence AUA (Fig. r3 A). Nous avons

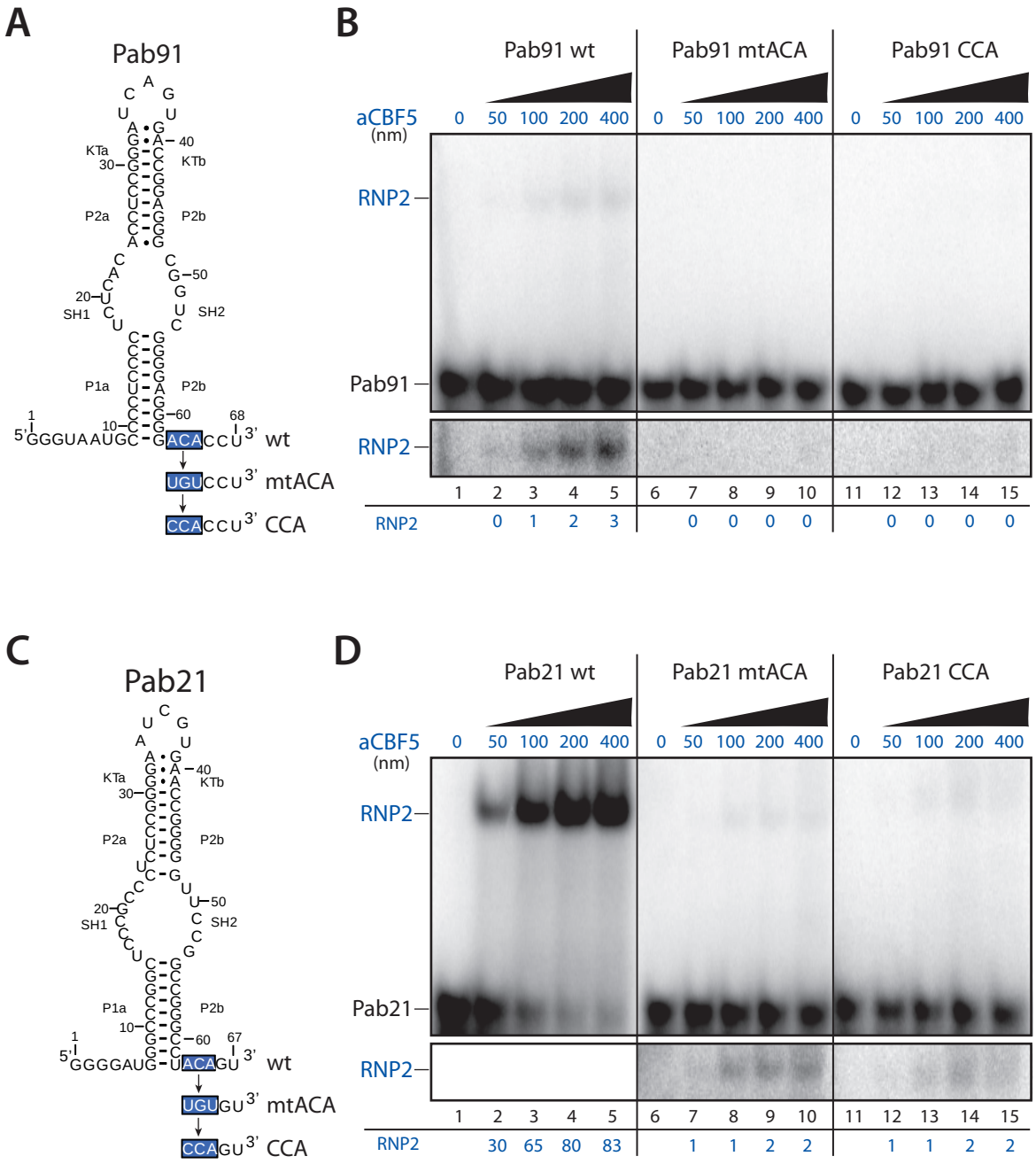


Figure r4 : Etude par expériences de retard de migration sur gel de l'influence de la boîte ACA des sRNA Pab91 et Pab21 sur la fixation directe de la protéine aCBF5.

A. Structure schématique du sRNA Pab91 à boîtes H/ACA de *P. abyssi* et des ses variants dont la boîte ACA a été mutée.

B. Résultats d'expériences de retard de migration sur gel réalisées en présence de variants du sRNA Pab91 radiomarqués en 5' et de aCBF5. Le taux de complexe formé entre le sRNA et aCBF5 (RNP2) est indiqué sous chaque piste. L'encadré présent sous le gel correspond à la zone de migration du complexe RNP2 qui a été surexposée au phosphoimager.

C. Structure schématique du sRNA Pab21 à boîtes H/ACA de *P. abyssi* et des ses variants dont la boîte ACA a été mutée.

D. Résultats d'expériences de retard de migration sur gel réalisées en présence de variants du sRNA Pab21 radiomarqués en 5' et de aCBF5. Le taux de complexe formé entre le sRNA et aCBF5 (RNP2) est indiqué sous chaque piste. L'encadré présent sous le gel correspond à la zone de migration du complexe RNP2 qui a été exposé plus longement au phosphoimager.

remplacé la séquence naturelle AUA par la séquence canonique ACA puis nous avons incubé le sRNA sauvage ou variant avec une gamme de concentration de la protéine aCBF5 (Fig. r3 B). Deux complexes sont formés, RNP2a et RNP2b, correspondant à la fixation, respectivement, d'une et deux molécules de aCBF5 sur ce sRNA. Nous constatons qu'à une même concentration de aCBF5, les taux des complexes RNP2a et RNP2b sont beaucoup plus élevés pour le sRNA Pab105 qui possède 2 boîtes ACA par rapport à celui qui contient une boîte ACA et une boîte AUA (Fig. r3 B). Nous concluons que pour ce sRNA, la substitution de la séquence naturelle AUA en ACA a un effet positif sur l'affinité de aCBF5 pour l'ARN.

Nous avons alors cherché à savoir quelle était le motif de l'une ou de l'autre des deux tige-boucles qui contribuait le plus à la fixation de aCBF5. Nous avons fabriqué deux nouvelles constructions qui permettent de transcrire séparément chacune des deux tige-boucles de Pab105. Elles sont nommées TB 5' et TB 3' et possèdent chacune une boîte ACA (Fig. r3 C). Nous avons incubé chacun de ces transcrits avec une gamme de concentration de aCBF5 et analysé le taux de complexes formés par retard de migration sur gel (Fig. r3 D). Nous constatons que dans les deux cas un seul complexe retardé RNP2 est formé. La protéine aCBF5 possède une meilleure affinité pour la tige-boucle 5' par rapport au motif 3' (Fig. r3 D), bien que l'ajout de la boîte ACA dans Pab105 entier permette d'augmenter l'affinité de aCBF5 au sRNA (Fig. r3 B). Nous concluons que la TB 5' possède des déterminants favorables à la fixation de aCBF5. Nous émettons également l'hypothèse qu'il pourrait exister une coopérativité entre les deux tige-boucles pour augmenter la liaison de aCBF5 au sRNA Pab105 entier.

Au laboratoire, B. Charpentier a montré que aCBF5 possédait une faible affinité pour le sRNA Pab91 et que la mutation de sa boîte ACA conduisait à la perte de la fixation de aCBF5 (Fig. r4 B pistes 1 à 10) (Charpentier *et al.*, 2005). Nous avons également montré que le remplacement de la séquence CCA de l'ARNt^{Asp} par la séquence ACA conduisait à augmenter fortement l'affinité de aCBF5 pour l'ARNt modifié (article n° 3, (Muller *et al.*, 2007b)). Nous avons alors analysé l'effet du remplacement de la boîte ACA de deux sRNA par la séquence CCA : le sRNA Pab91

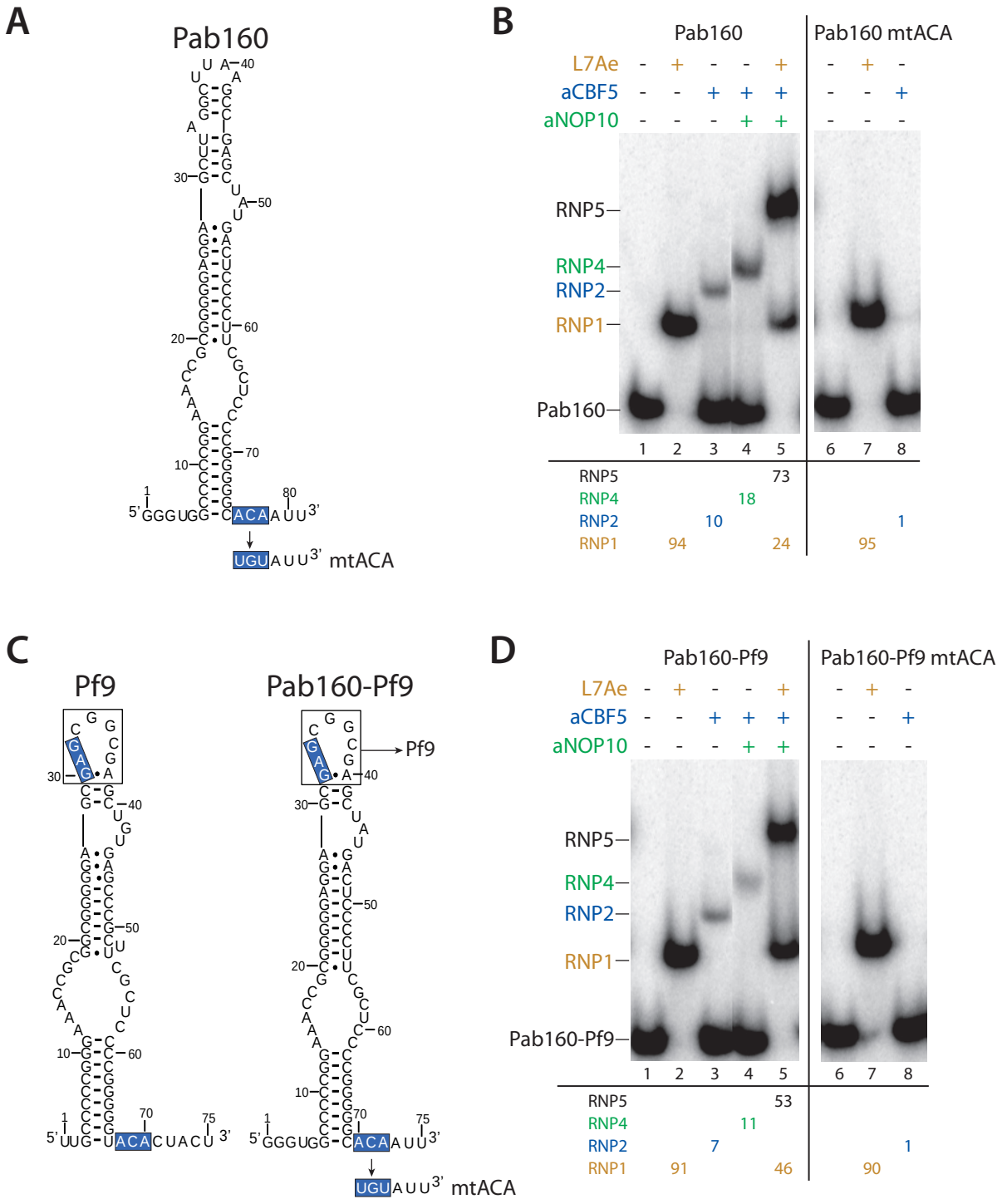


Figure r5 : Etude par expériences de retard de migration sur gel de l'importance de la boucle apicale du sRNA Pf9 inséré dans son homologue fonctionnel Pab160 sur la fixation directe des protéines.

A. Structure schématique du sRNA Pab160 à boîtes H/ACA de *P. abyssi* et du variant muté dans la boîte ACA.

B. Résultats d'expériences de retard de migration sur gel réalisées en présence du sRNA Pab160 radiomarqué en 5' et de différentes combinaisons de protéines L7Ae (brun), aCBF5 (bleu), aNOP10 (vert). Le taux de complexe formé entre le sRNA et les différentes protéines est indiqué sous chaque piste.

C. Structure schématique du sRNA Pf9 de *P. furiosus* et du sRNA composite Pab160-Pf9 à boîtes H/ACA de *P. abyssi* et de *P. furiosus* et du variant muté dans la boîte ACA. La séquence encadrée correspond à celle provenant du Pf9 de *P. furiosus* et contient la boîte GAG.

D. Résultats d'expériences de retard de migration sur gel réalisées en présence du sRNA Pab160-Pf9 radiomarqué en 5' et de différentes combinaisons de protéines L7Ae (brun), aCBF5 (bleu), aNOP10 (vert). Le taux de complexe formé entre le sRNA et les différentes protéines est indiqué sous chaque piste.

pour lequel aCBF5 possède une faible affinité et le sRNA Pab21 qui fixe fortement aCBF5 (Fig. r4 A et C). Aucun complexe stable n'est formé entre aCBF5 et les sRNA Pab91 et Pab21 possédant la séquence CCA (Fig. r4 B et D pistes 11 à 15). Des résultats similaires ont été obtenus avec le sRNA Pab160 comportant une séquence CCA (données non montrées). Nous concluons que dans le cas des sRNA, une séquence CCA ne peut se substituer à une séquence ACA.

2.2. La séquence GAG contenue dans la boucle apicale n'est pas un déterminant général pour la fixation de aCBF5 aux sRNA

Il a été montré que la présence de la séquence 5'-GAG-3' dans la boucle apicale du sRNA Pf9 de *P. furiosus* favorisait la fixation de aCBF5 aux sRNA d'archaea mais seulement dans le cadre d'un sous-complexe sRNA-aCBF5 (Baker *et al.*, 2005; Baker *et al.*, 2008). Nous disposions de l'homologue fonctionnel de Pf9 chez *P. abyssi*, le sRNA Pab160 (Fig. r5 A). La fixation de aCBF5 à Pab160 ne conduit pas à la formation d'un fort taux de complexe, contrairement à Pf9 (Fig. r5 B piste 3) (Baker *et al.*, 2005). Nous avons donc réalisé la construction d'un ARN chimérique, nommé Pab160-Pf9, qui est composé pour sa partie basale de la structure de Pab160 et de la boucle apicale de Pf9 située au-dessus du motif en K-turn (Fig. r5 C).

Nous avons testé la fixation des protéines à l'ARN chimérique Pab160-Pf9 par retard de migration sur gel (Fig. r5 D). Nous observons que le taux de complexe RNP2 formé entre aCBF5 et Pab160-Pf9 est comparable à celui obtenu entre aCBF5 et Pab160 (Fig. r5 B et D piste 3). Le taux comparable des complexes RNP1, RNP4, et RNP5, formés entre les sRNA Pab160 et Pab160-Pf9, indique que la conformation du sRNA n'a pas été perturbée par la substitution des séquences (Fig. r5 D pistes 2, 4 et 5). La mutation de la boîte ACA ne permet plus d'observer de complexes stables entre aCBF5 et les deux sRNA (Fig. r5 B et D piste 8). Nous concluons que la fixation de aCBF5 sur le sRNA Pab160 est limitée et que la présence de la séquence GAG dans la boucle apicale ne permet pas d'améliorer cette faible affinité apparente. Ainsi, la séquence GAG ne serait pas un déterminant général de la fixation de aCBF5 à ce

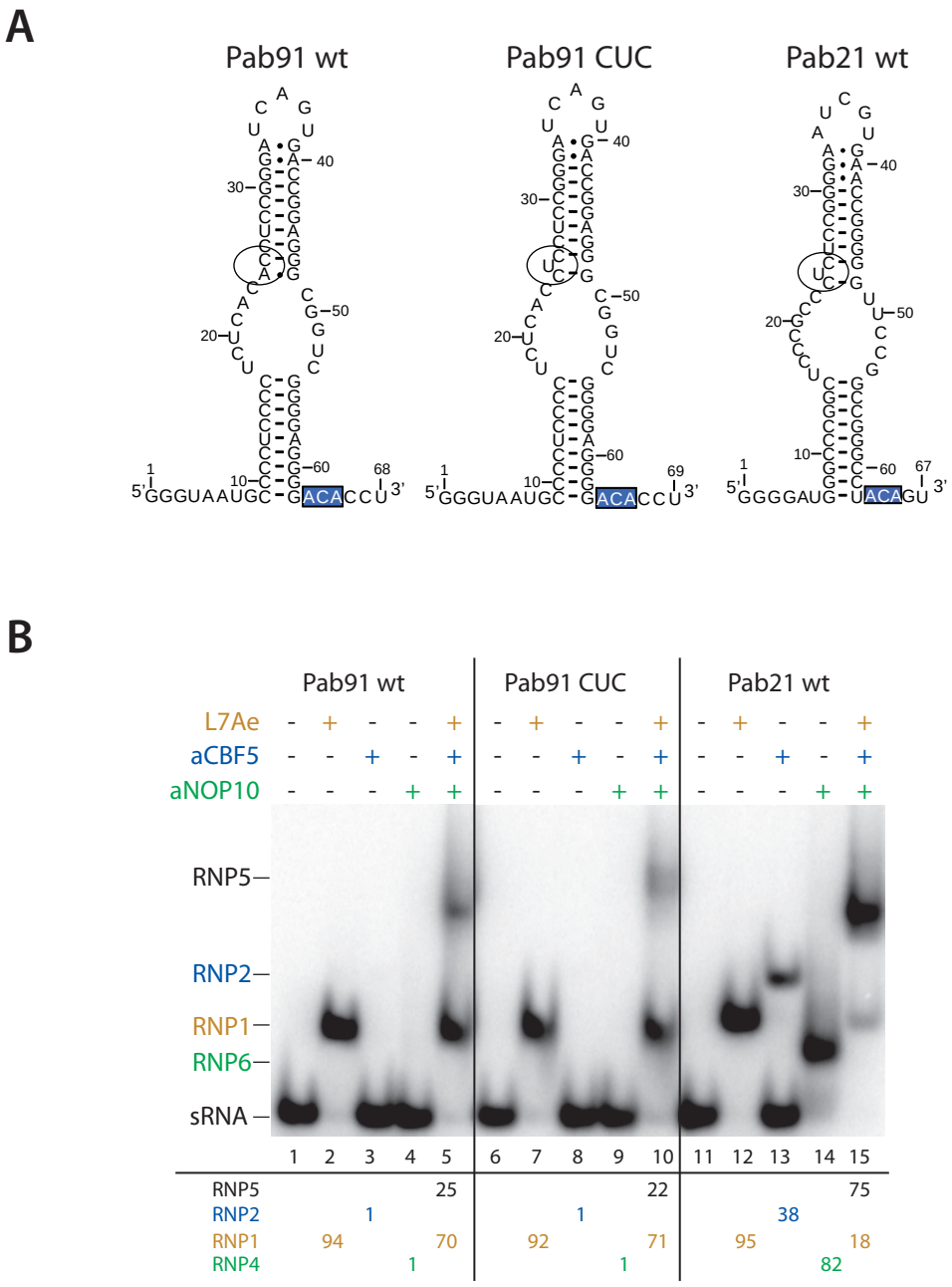


Figure r6 : Etude par expériences de retard de migration sur gel de l'importance du nucléotide extrudé à la base de l'hélice 2 de sRNA à boîtes H/ACA sur la fixation directe des protéines.
A. Structures schématiques des sRNA Pab91 et Pab21 sauvages, et du sRNA variant Pab91 CUC qui porte la séquence 5'-CUC-3' qui est entourée et qui est identique à celle de Pab21.
B. Résultats d'expériences de retard de migration sur gel réalisées a partir des différents sRNA indiqués, radiomarqués en 5' et de différentes combinaisons de protéines L7Ae (brun), aCBF5 (bleu), aNOP10 (vert). Le taux de complexe formé entre le sRNA et les différentes protéines est indiqué sous chaque piste.

sRNA. Des données complémentaires d'empreintes sur l'ARN renforcent cette conclusion (article n°4, Fourmann *et al.*, en préparation).

2.3. Test de l'influence d'un résidu sortant, situé à la jonction entre l'hélice 2 et la poche de pseudouridylation, sur la fixation de aCBF5

Nous avions constaté que les sRNA d'archaea Pab105 tige-boucle 5' (Fig. 3 C), Pab21 (Fig. r4 C) et Pf9 (Fig. r5 C) (Baker *et al.*, 2005), qui fixent fortement la protéine aCBF5 possèdent un résidu extrudé situé une paire de base au-dessus de la jonction entre l'hélice H2 et la poche de pseudouridylation. Par contre, d'autres sRNA d'archaea tels que la tige-boucle 3' de Pab105 (Fig. r3 C), Pab91 (Fig. r4 A) et Pab160 (Fig. r5 A), possèdent une faible affinité apparente envers aCBF5 et ne présentent aucun résidu sortant de leur hélice H2. Nous avons émis l'hypothèse que ce résidu « en bulge » pourrait être un déterminant de la fixation de aCBF5 aux sRNA. Nous avons introduit la séquence du sRNA Pab21 5'-CUC-3' dont le résidu U est extrudé, à la position équivalente dans le sRNA Pab91 (Fig. r6 A), puis testé la fixation des protéines par retard de migration sur gel (Fig. r6 B). Nous constatons que le sRNA Pab91 possédant la séquence CUC ne permet pas la formation d'un taux plus élevé complexe avec la protéine aCBF5 que le sRNA Pab91 sauvage (Fig. r6 B pistes 3 et 8), ce qui nous permet de conclure qu'un résidu extrudé à cette position n'est pas un déterminant pour la fixation de aCBF5. De plus, nous avions observé qu'un complexe RNP6 était formé suite à l'incubation de aNOP10 avec le sRNA Pab21 (Fig. r6 B piste 14). Nous constatons que le résidu extrudé n'est pas non plus un déterminant de la fixation de aNOP10 au sRNA puisque qu'aucun complexe RNP6 n'est formé avec le sRNA variant Pab91 CUC (Fig. r6 B piste 9).

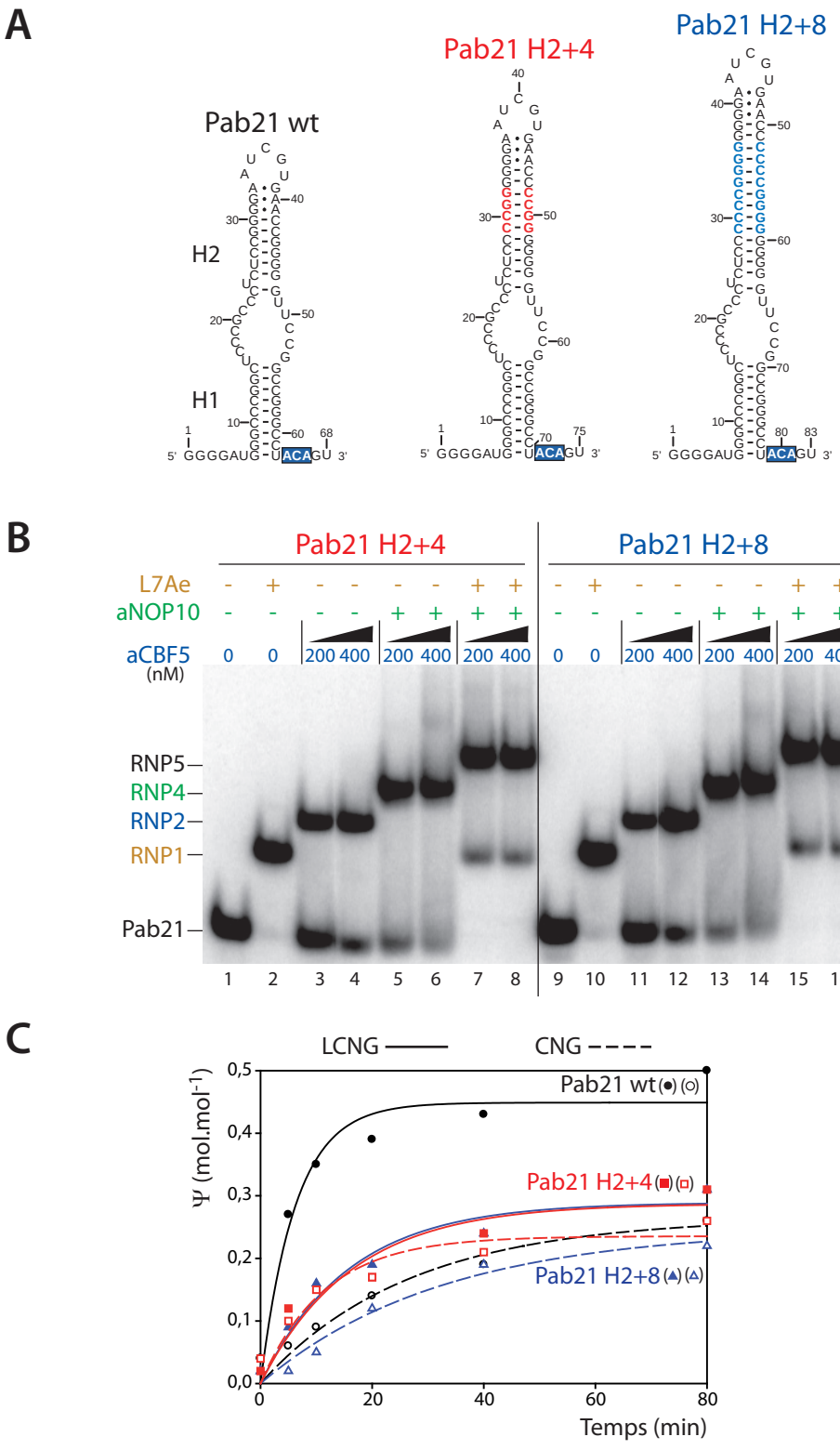


Figure r7 : Etude de l'effet de l'allongement de l'hélice H2 du sRNA Pab21 sur l'assemblage et l'activité de la particule.

A. Représentation schématique des différentes constructions sauvage (Pab21 wt), allongée de 4 paires de bases dans l'hélice H2 (Pab21 H2+4) et allongée de 8 paires de bases dans l'hélice H2 (Pab21 H2+8).

B. Résultats d'expériences de retard de migration sur gel réalisées à partir des sRNA Pab21 H2+4 et Pab21 H2+8 radiomarqués en 5', et de différentes conditions de protéines.

C. Etudes cinétiques du taux de résidus Ψ formés en employant les sRNA Pab21 wt (noir), Pab21 H2+4 (rouge) ou Pab21 H2+8 (bleu) et les jeux de protéines LCNG (trait plein) ou CNG (trait pointillé) et une température de 65°C.

2.4. Contribution de l'hélice H2 dans l'assemblage et dans l'activité des sRNP

Les modèles de structures secondaire des différents sRNA identifiés chez les archaea indiquaient que la taille de l'hélice H2 séparant le motif en K-turn de la poche de pseudouridylation était conservée et qu'une taille de 6 pb était optimale pour l'activité (article n°1, (Muller *et al.*, 2008)). Nous avons testé l'effet de la modification de la taille de l'hélice H2 dans le but d'éloigner le site de fixation de la protéine L7Ae de la poche de pseudouridylation. En outre, les sRNA variants permettaient d'étudier la contribution de L7Ae à l'activité des particules. Pour cela, nous avons introduit soit 4 pb (variant Pab21 H2+4), soit 8 pb (variant Pab21 H2+8) dans la région centrale de l'hélice H2 (Fig. r7 A). La variation de taille de 4 pb correspond environ à ½ tout d'hélice et pouvait donc contribuer à déplacer un élément *cis* comme le K-turn sur une autre face de l'hélice H2, tandis que l'introduction des 8 pb permettait d'éloigner le site de fixation de L7Ae d'environ un tour complet d'hélice de la poche de pseudouridylation. Nous avons choisi d'introduire une alternance de paires de base C-G et G-C car les 2 hélices de Pab21 portent naturellement ce type de séquences (Fig. r7 A).

Nous avons testé les deux variants par retard de migration sur gel et nous observons la formation des complexes RNP1, RNP2, RNP4 et RNP5, ce qui indique que les paires de bases ajoutées n'affectent pas l'assemblage de la sRNP (Fig. r7 B). Cependant, nous constatons que l'effet positif de L7Ae sur l'activité est perdu (Fig. r7 C). Des mesures de dichroïsme circulaire réalisées au laboratoire par M. Blaud et A.-S. Tillault, ont confirmé que L7Ae se fixait sur ces variants ARN mais que la conformation du sRNA au sein notamment de la sRNP LCN était différente (données non montrées). Ainsi, la présence de L7Ae, fixée sur le sRNA, n'est pas suffisante pour permettre une activité optimale de la particule mais son orientation par rapport au reste des protéines de la RNP apparaît être un déterminant majeur pour l'activité des particules. Ces données font l'objet de la discussion des résultats complémentaires à l'article n°4 en préparation (cf partie II.D des résultats).

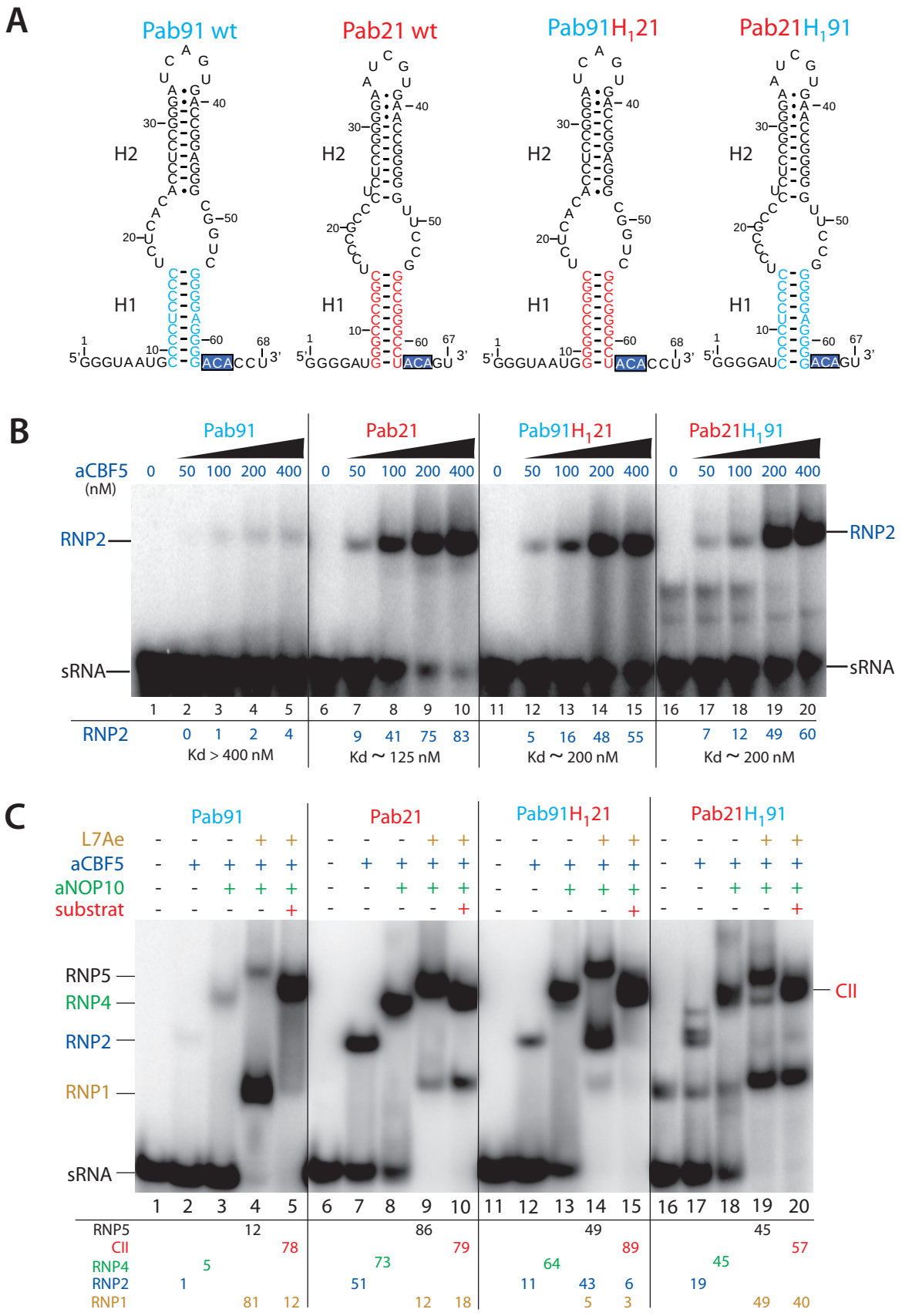


Figure r8 : Etude par expériences de retard de migration sur gel du rôle de l'hélice 1 de sRNA à boîtes H/ACA sur le taux de fixation de la protéine aCBF5 aux sRNA.

A. Structures schématiques des sRNA Pab91 et Pab21 sauvages, et des sRNA chimériques Pab91H$_1$21 et Pab21H$_1$91qui portent respectivement la séquence de l'hélice 1 du Pab21 et du Pab91.

B. et **C.** Résultats d'expériences de retard de migration sur gel réalisées à partir des différents sRNA indiqués, radiomarqués en 5' (0,1 nM) et de différentes combinaisons de protéines L7Ae (brun), aCBF5 (bleu), aNOP10 (vert) et l'ARN substrat (rouge). Le taux de complexe formé entre le sRNA et les différentes protéines est indiqué sous chaque piste.

2.5. L'hélice H1 est un déterminant important de la fixation de aCBF5 au sRNA

Comme le montre nos résultats avec Pab91 et Pab21, des éléments discrets présents sur certains sRNA favorisent la formation de forts taux de complexes aCBF5-ARN. Très peu d'éléments de séquences sont conservés dans les sRNA à boîtes H/ACA. Excepté la boîte ACA qui est très conservée, les autres parties du sRNA ont une séquence qui est variable, ce qui est peu propice à une reconnaissance séquence spécifique par les protéines de la sRNP. Néanmoins, les structures 3D récentes de la sRNP ont montré que aCBF5 reconnaissait l'hélice H1 par des interactions dites non spécifiques : des résidus chargés positivement de aCBF5 établissent des contacts électrostatiques et des liaisons hydrogènes avec le squelette ribose-phosphate chargé négativement de l'ARN (Li et Ye, 2006; Liang, B. *et al.*, 2007). Nous avons comparé la séquence des 9 paires de bases qui constituent l'hélice H1 des sRNA Pab91 et Pab21 (Fig. r8 A). Nous avons constaté que pour Pab91, tous les résidus C sont situés sur le brin 5' de l'hélice alors que les résidus G sont localisés sur le brin complémentaire 3', pouvant ainsi créer une différence dans la courbure globale intrinsèque à chaque hélice d'ARN (Fig. r8 A). Par contre, pour Pab21, nous observons une alternance entre les paires de bases C-G de part et d'autre de l'hélice H1 (Fig. r8 A). Nous avons émis l'hypothèse que la conformation du sRNA pourrait être modifiée par cette disparité du positionnement des résidus C et G dans l'hélice et de ce fait et influencerait le taux de fixation de aCBF5 au sRNA.

2.5.1. Influence de la permutation des hélices H1 des sRNA Pab91 et Pab21 sur la fixation de aCBF5 par l'approche de retard de migration sur gel

Nous avons donc créé deux ARN chimériques, l'un nommé Pab91$H_1$21 pour lequel l'hélice H1 de Pab91 a été substituée par celle de Pab21 et le second appelé Pab21$H_1$91 dans lequel c'est l'hélice H1 de Pab21 qui a été substituée par celle de Pab91 (Fig. r8 A). Nous avons testé la fixation de la protéine aCBF5 sur ces sRNA par retard de migration sur gel et estimé une constante de dissociation apparente (Kd) entre chaque sRNA et aCBF5 (Fig. r8 B). Nous constatons qu'à une concentration de

aCBF5 de 400 nM une très faible quantité de sRNA est déplacée dans le complexe RNP2 (4%), ce qui indique que aCBF5 a peu d'affinité pour le sRNA Pab91 (Fig. r8 B, pistes 2-5). Par contre, le Kd apparent de aCBF5 pour le Pab21 est de l'ordre de 125 nM, ce qui reflète une affinité apparente bien supérieure de cette protéine pour ce sRNA (Fig. r8 B, pistes 7-10). L'étude de l'affinité de aCBF5 avec l'un ou l'autre des sRNA chimériques conduit à un Kd apparent qui est de l'ordre de 200 nM (Fig. r8 B). Ainsi, la présence de l'hélice H1 de Pab21 améliore nettement la fixation de aCBF5 sur le sRNA Pab91. Nous avons vérifié que les séquences introduites ne modifiaient pas la formation des différentes RNP (Fig. r8 C, pistes 11 à 14 et 16 à 19). Dans toutes les constructions chimériques, les séquences des poches de pseudouridylation de Pab21 ou Pab91 n'ont pas été modifiées, mais la substitution de l'hélice H1 pouvait modifier la position des nucléotides et ainsi altérer leur capacité à interagir d'une part avec la protéine aCBF5 et d'autre part avec l'ARN substrat. Nous avons finalement réalisés des tests en présence de l'ARN substrat RNA-S (substrat de Pab91) à la sRNP Pab91$H_1$21 et nous constatons que celui-ci est incorporé dans la particule RNP5 conduisant à la formation d'un fort taux de complexe CII (Fig. r8 C pistes 15, 89%). L'ajout du substrat Ψ891 à la particule RNP5 assemblée sur Pab21 ou Pab21$H_1$91 conduit à la formation d'un complexe CII (57%) de mobilité électrophorétique plus élevée que celle de RNP5 ainsi que d'un complexe retardé similaire à RNP1 (40%) (Fig. r8 C pistes 20).

Ainsi, la permutation des hélices H1 entre les ARN guide Pab21 et Pab91 autorise la formation de la sRNP et conserve la spécificité de fixation des protéines et de l'ARN substrat de manière comparable aux ARN sauvages. De plus, la substitution de l'hélice H1 de Pab91 par celle de Pab21 permet d'augmenter l'affinité de aCBF5 et à l'inverse, le remplacement de l'hélice H1 de Pab21 par celle de Pab91 conduit à une baisse très forte de l'affinité apparente de aCBF5 pour le sRNA chimérique. Nous concluons que l'hélice H1 est un déterminant très important pour la fixation de aCBF5 aux sRNA à boîtes H/ACA.

2.5.2. Influence de la permutation des hélices H1 des sRNA Pab91 et Pab21 sur la conformation globale des sRNA étudiée par dichroïsme circulaire

Par la technique de retard de migration sur gel, il est possible de visualiser une interaction stable entre une ou plusieurs protéines avec un ARN et il est possible de distinguer différents complexes formés entre les protéines et l'ARN. Cependant, une des limites de cette technique est le manque d'informations structurales. En effet, pour un complexe de composition connue, des changements de structure ou de la conformation des différents partenaires peuvent parfois ne pas modifier la vitesse de migration des complexes. La technique de dichroïsme circulaire (cf Matériel et Méthodes Fig. m2) est alors mise en œuvre pour appréhender les modifications de la structure secondaire des ARN qui s'opèrent lors de la formation des complexes RNP. Cette approche a été développée au laboratoire par Magali Blaud pour l'étude de l'assemblage séquentiel des sRNP à boîtes H/ACA (article n°4, Fourmann *et al.,* en préparation).

J'ai alors utilisé cette approche pour étudier l'impact de la permutation des hélices H1 des sRNA Pab91 et Pab21 sur la conformation de ces sRNA et étudié l'effet de la fixation de la protéine aCBF5 à ces sRNA (Fig. r9). L'appareil mesure une variation d'absorption (ΔA) dont la valeur est normalisée sous la forme $\Delta\varepsilon$ en tenant compte de la concentration de l'échantillon, du trajet optique et enfin du nombre de résidus dont est constitué la molécule d'intérêt. Ainsi, les spectres de dichroïsme circulaire exprimés sous la forme de mesures de $\Delta\varepsilon$ en fonction de la longueur d'onde λ employée sont comparables entre eux, quelles que soient les conditions d'enregistrement des spectres et la taille de la protéine ou de l'ARN étudié.

<u>Etude de la protéine aCBF5 en solution par dichroïsme circulaire</u>

Nous avons mesuré un premier spectre dichroïque de la protéine aCBF5 seule en solution (Fig. r9 courbes en orange). Ce spectre est négatif dans une plage de longueur d'onde comprise entre 200 et 240 nm et est caractéristique de chaque protéine. Ainsi, ce spectre spécifique rend compte de la composition en structures

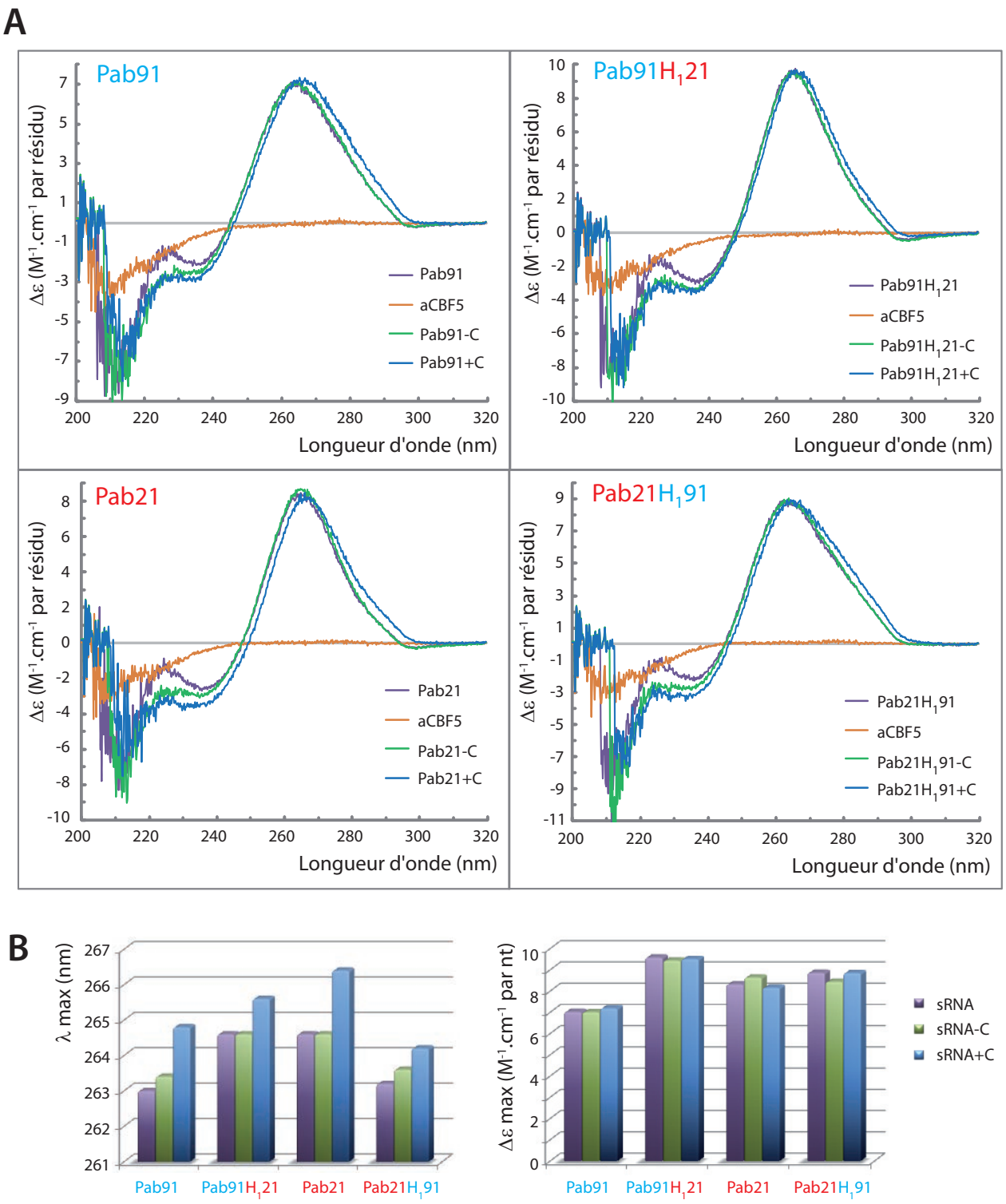


Figure r9 : Etude par dichroïsme circulaire de l'interaction entre aCBF5 et les sRNA Pab91, Pab91H$_1$21, Pab21 et Pab21H$_1$91.

A. Etude des changements de la conformation des ARN lors de l'association de la protéine aCBF5 aux sRNA à boîtes H/ACA Pab91, Pab91H$_1$21, Pab21 et Pab21H$_1$91. En violet : spectre du sRNA seul en solution. En orange : spectre de la protéine aCBF5 seule en solution. En vert : spectre du sRNA et de aCBF5 dans des compartiments séparés de la cuve. En bleu : spectre mesuré après mélange du sRNA et de aCBF5 dans la cuve. Les spectres de dichroïsme circulaire ont été enregistrés à partir du dichrographe circulaire Jobin-Yvon Mark VI à une vitesse de mesure de 2 nm/s, dans une cuve en quartz à deux compartiments de 0,875 cm de trajet optique, à température ambiante et avec une concentration finale en protéine et ARN de 2 µM dans le tampon Tris-HCl 10 mM (pH 7,5), NaCl 150 mM. Les résultats sont présentés en tant que valeurs normalisées de $\Delta\varepsilon$.

B. Histogrammes représentant les valeurs des pics positifs obtenus dans la région comprise entre 240 et 300 nm (λmax) et de l'amplitude du pic ($\Delta\varepsilon$ max) à cette valeur de λmax mesurée pour les 4 sRNA, à partir des spectres dichroïques de A dont le code couleur est inchangé.

secondaires (hélices α et feuillets β) de la protéine aCBF5. Il présente deux pics négatifs aux longueurs d'onde 208 et 222 nm, dus à la présence des hélices α mais modifiés par la contribution des feuillets β. Ce spectre montre ainsi que aCBF5 est soluble dans nos conditions expérimentales et qu'elle est dans une conformation en accord avec la structure décrite par les données cristallographiques. Il est important de souligner qu'aucune valeur de Δε n'est mesurée de 240 à 300 nm pour aCBF5. En particulier, l'absence de signal à la longueur d'onde de 260 nm indique que l'échantillon de aCBF5 utilisé dans ces expériences ne contient pas d'acides nucléiques.

Etude de la structure secondaire des sRNA par dichroïsme circulaire

Nous avons mesuré un second spectre par dichroïsme circulaire de chaque sRNA seul en solution (Fig. r9 courbes en violet). Les 4 sRNA présentent chacun un spectre caractéristique qui contient une zone de Δε positive aux longueurs d'ondes comprises entre 240 à 300 nm qui est caractéristique des acides nucléiques. Pour caractériser ces spectres, nous avons mesuré pour chacun la longueur d'onde maximale appelée λ max et l'intensité du signal enregistré à cette longueur d'onde nommée ici Δε max (Fig. r9 B).

La valeur de λ max est représentative de la courbure d'une molécule d'acide nucléique. Ainsi, dans le cas des sRNA dont la structure est formée par une succession tige-boucle-tige-boucle, cela peut rendre compte notamment de la présence d'une courbure de l'axe général du sRNA. Nous constatons que les valeurs de λ max sont très proches ou identiques pour des sRNA possédant la même hélice H1 : sRNA Pab91 (λ max=263 nm) et Pab21$H_1$91 (λ max =263,2 nm) d'une part ; et sRNA Pab21 (λ max=264,6 nm) et Pab91$H_1$21 (λ max=264,6 nm) d'autre part (Fig. r9 B). Ces résultats indiquent que les sRNA sauvages Pab91 et Pab21 ne présentent pas la même courbure de l'ARN et que les couples de sRNA (Pab91 et Pab21$H_1$91) et (Pab21 et Pab91$H_1$21) possèdent une courbure similaire. Nous concluons que l'hélice

H1 est un déterminant majeur qui influence la courbure globale d'un sRNA à boîtes H/ACA.

L'amplitude du pic, traduite par la valeur de Δε à la longueur d'onde correspondant au λ max (Δεmax), nous renseigne sur l'état d'empilement des bases, une diminution de Δεmax indique un empilement moins important des bases dans l'ARN (stacking). Nous constatons que la valeur de Δε max de Pab91 (Δε max=7.02 $M^{-1}.cm^{-1}$ par résidu) est plus faible que celle de Pab21 (Δε max=8.32 $M^{-1}.cm^{-1}$ par résidu). Cela indique que les hélices ARN de Pab91 s'empilent moins que celles de Pab21 (cf I.B.2.5.3).

Etude des interactions ARN-protéine par dichroïsme circulaire

La démarche consiste à placer l'un des 4 sRNA dans l'un des deux compartiments de la cuve de mesure et la protéine, ici aCBF5, dans l'autre compartiment. Un spectre de dichroïsme circulaire est alors mesuré. Celui-ci correspond, en théorie, à la somme des deux spectres de chacun des composants mesurés individuellement (cf Matériel et Méthodes Fig. m2 ; Fig. r9 courbes en vert).

Les contenus des deux compartiments de la cuve sont ensuite mélangés et une nouvelle mesure est réalisée (Fig. r9 courbes en bleu). Ce spectre rend compte, par comparaison au précédent, de la modification de la structure globale engendrée par la formation du complexe sRNA-aCBF5. L'association de aCBF5 au sRNA peut être suivie par la modification du spectre dans la région spécifique des acides nucléiques entre 240 et 300 nm. Pour les 4 sRNA, l'ajout de aCBF5 conduit à un décalage important de λ max vers les plus grandes longueurs d'ondes et à une faible variation de l'intensité du pic positif (δ(Δε)) (Fig. r9 B). Cela indique que la fixation de aCBF5 aux sRNA provoque une variation importante de la courbure des sRNA et une faible modification de l'empilement des bases. Cet effet est discuté en détail dans l'article n°4 (Fourmann *et al.,* en préparation). L'application de cette approche nous a permis de conclure ici que aCBF5 interagit avec chacun des 4 sRNA et qu'elle provoque une

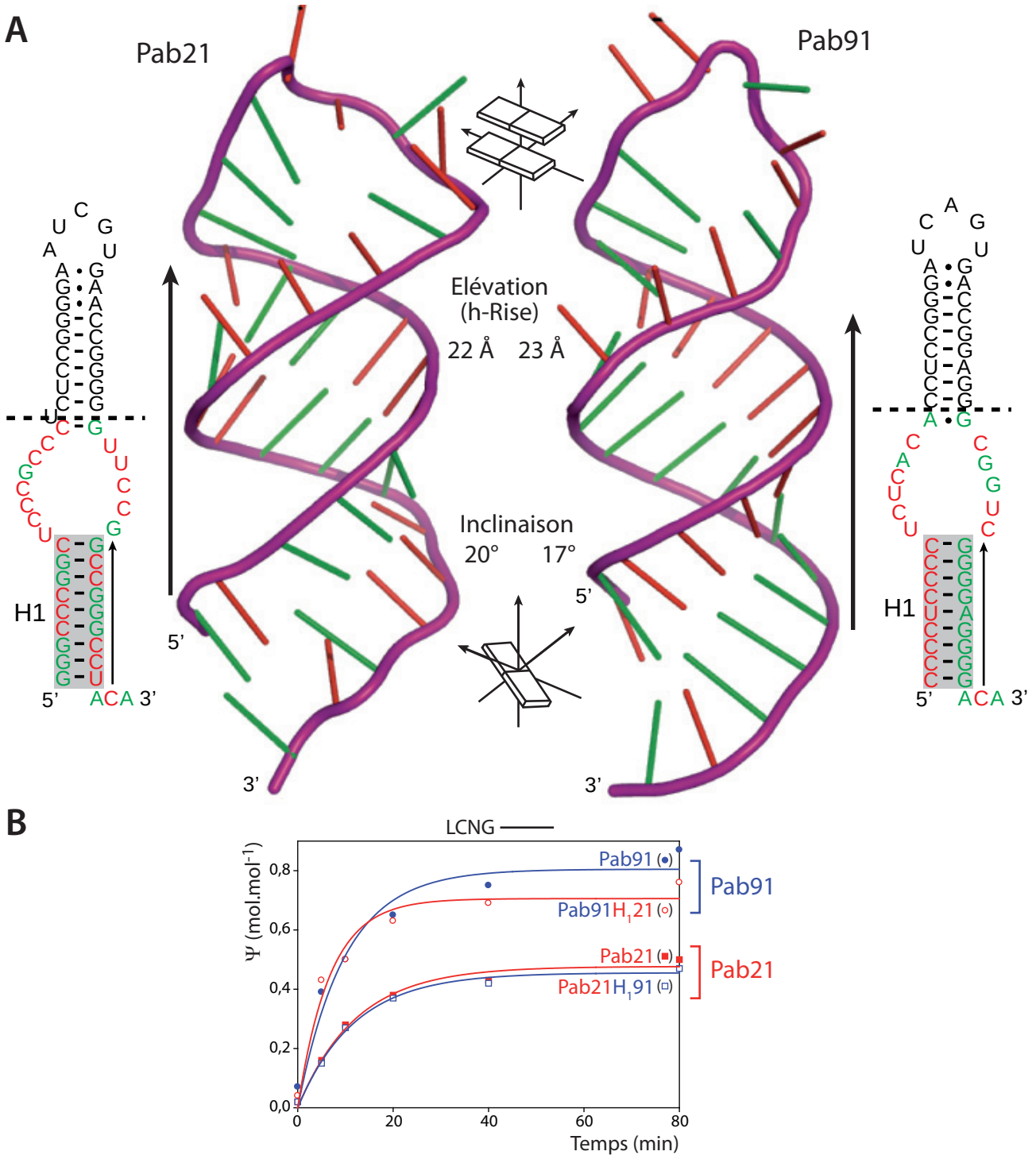


Figure r10 : Modèles des formes tronqués des sRNA Pab21 et Pab91 obtenus par dynamique moléculaire et étude cinétique du taux de résidus Ψ formés par emploie des ARN chimériques.
A. Modélisation réalisée au laboratoire par Fabrice Leclerc. Les sRNA comportent les parties basales incluant l'hélice H1, la boucle interne et la boîte ACA. Les modèles 3D ont été construits avec des paramètres d'hélice standard (Mc-Sym); la boucle terminale (assimilée à la boucle interne de la forme non tronquée) a été extraite d'une structure 3D correspondant à une dodecaloop de l'ARNr 23S (PDB=1J5A) choisie pour sa similarité de séquence et proportion purines (en vert)/pyrimidines (en rouge). Une simulation par dynamique moléculaire (CHARMM) de 5ns (en plus de l'équilibration du système 500ps) a été effectuée pour chaque ARN immergé dans une boîte d'eau contenant des contre-ions (ions Na+). Les mesures d'élévation et d'inclinaison de l'hélice ont été réalisées sur la structure "moyenne" (coordonnées moyennées sur l'ensemble de la simulation et otpmisées par minimisation d'énergie). Les valeurs d'élévation et d'inclinaison hélicoïdales ont été calculées sur les 9 paires de base de l'hélice (X3DNA).
B. Les études cinétiques ont été réalisées en employant le jeu de protéines L7Ae, aCBF5, aNOP10, aGAR1 (LCNG, en trait plein) à 65°C. Les conditions expérimentales correspondent à l'emploi des sRNA Pab91 (bleu, cercle rempli) et Pab91H$_1$21 (rouge, cercle vide) avec le substrat RNAr ou des sRNA Pab21 (rouge, carré rempli) et Pab21H$_1$91 (bleu, carré vide) avec le substrat Ψ891.

modification de la courbure de ces ARN qui serait caractéristique de la sous-particule RNP2. Nous concluons également que la courbure est plus ou moins marquée dans le complexe RNP2 en fonction de l'identité de H1.

2.5.3. Modélisation de l'hélice H1 et de la poche de pseudouridylation des sRNA Pab91 et Pab21

En parallèle, au laboratoire Fabrice Leclerc a réalisé une modélisation moléculaire par homologie d'une région des sRNA Pab91 et Pab21 comprenant l'hélice H1, la poche de pseudouridylation, deux résidus en 5' des hélices H1 et la boîte ACA en 3' (Fig. r10 A). Pour faciliter ce travail, les hélices H1 ont été fermées par des boucles identiques dans les deux sRNA. Les détails de la méthodologie sont indiqués dans la légende de la figure r10. L'élévation (h-Rise), qui correspond à une hauteur d'un tour complet d'hélice et mesurée entre 2 résidus homologues dans chacune des hélices H1, est de 22 Å pour Pab21 et de 23 Å pour Pab91. L'hélice H1 de Pab91 est donc un peu plus étirée que celle de Pab21. Ces données d'élévation obtenues concordent avec les valeurs de $\Delta\varepsilon$ provenant des mesures réalisées par dichroïsme circulaire pour les sRNA Pab91 et Pab21 (cf I.B.2.5.2).

Par rapport à un axe hélicoïdal, l'inclinaison des hélices ARN modélisées est de 20° pour le sRNA Pab21 et de 17° pour Pab91. La courbure du sRNA Pab21 est donc plus marquée que celle de Pab91. Les résultats des λ max obtenus par dichroïsme circulaire indiquaient également une modification prononcée de la courbure entre Pab21 et Pab91 (cf I.B.2.5.2). La différence entre les séquences des hélices H1 pourraient induire une variation de la stabilité de ces deux sRNA. Ainsi, une forte stabilité de la partie basale de l'hélice H1 associée à l'étirement de cette hélice dans le sRNA Pab91 pourrait donc être défavorable à la fixation de aCBF5 sur ce sRNA. Des études complémentaires de modélisation des sRNA Pab91 et Pab21 entiers ainsi que les ARN chimériques sont envisagées pour compléter ces données.

2.5.4. Le taux de complexe formé entre aCBF5 et un sRNA ne détermine pas l'efficacité de la réaction de modification

Nous avons montré que la permutation des hélices H1 entre Pab91 et Pab21 ne perturbait ni l'association des protéines à ces ARN, ni la fixation spécifique des ARN substrats respectifs à chacune des poches de pseudouridylation (Fig. r8). Nous avons cherché à savoir si la substitution des hélices H1 influençait le taux de pseudouridylation des substrats. Les sRNA Pab91 et Pab91$H_1$21 ont été testés pour leur capacité à guider la modification du substrat dénommé RNAr (Matériels et Méthodes) spécifique à Pab91, tandis que les sRNA Pab21 et Pab21$H_1$91 ont été utilisés avec le substrat Ψ891 (Matériels et Méthodes) spécifique à Pab21. Les activités ARN:Ψ-synthase ont été mesurées en présence des protéines L7Ae, aCBF5, aNOP10 et aGAR1 (LCNG) (Fig. r10 B).

Nous constatons que les cinétiques de pseudouridylation obtenues par les sRNA Pab91 et Pab91$H_1$21 d'une part, et Pab21 et Pab21$H_1$91 d'autre part, sont situées dans une gamme comparable (Fig. r10 B). Les ARN chimériques Pab91$H_1$21 et Pab21$H_1$91 ont la même capacité pour guider la modification de leur ARN substrat par rapport à leur équivalent sauvages respectifs Pab91 et Pab21. Nous concluons que l'hélice H1 n'influe pas sur la cinétique et le taux de modifications. Ainsi le taux de complexe formé entre aCBF5 et le sRNA, tel que nous l'avons appréhendé par les expériences de retard de migration sur gel, n'est pas déterminant dans la réaction de pseudouridylation. Ainsi, malgré un faible de taux de complexe RNP2, la sRNP Pab91 est capable de modifier à 100% un ARN substrat, ce qui n'est pas le cas du sRNA Pab21, qui bien que formant un fort taux de complexe RNP2, conduit à une activité de la sRNP qui atteint un plateau et ne peut modifier plus de 50 % des ARN substrats. Les déterminants de cette limitation ne sont pas en lien avec l'hélice H1. La réaction de modification par la sRNP Pab21 est sans doute limitée au niveau d'une ou plusieurs des étapes ultérieures à l'étape 1 du mécanisme (Fig. i62).

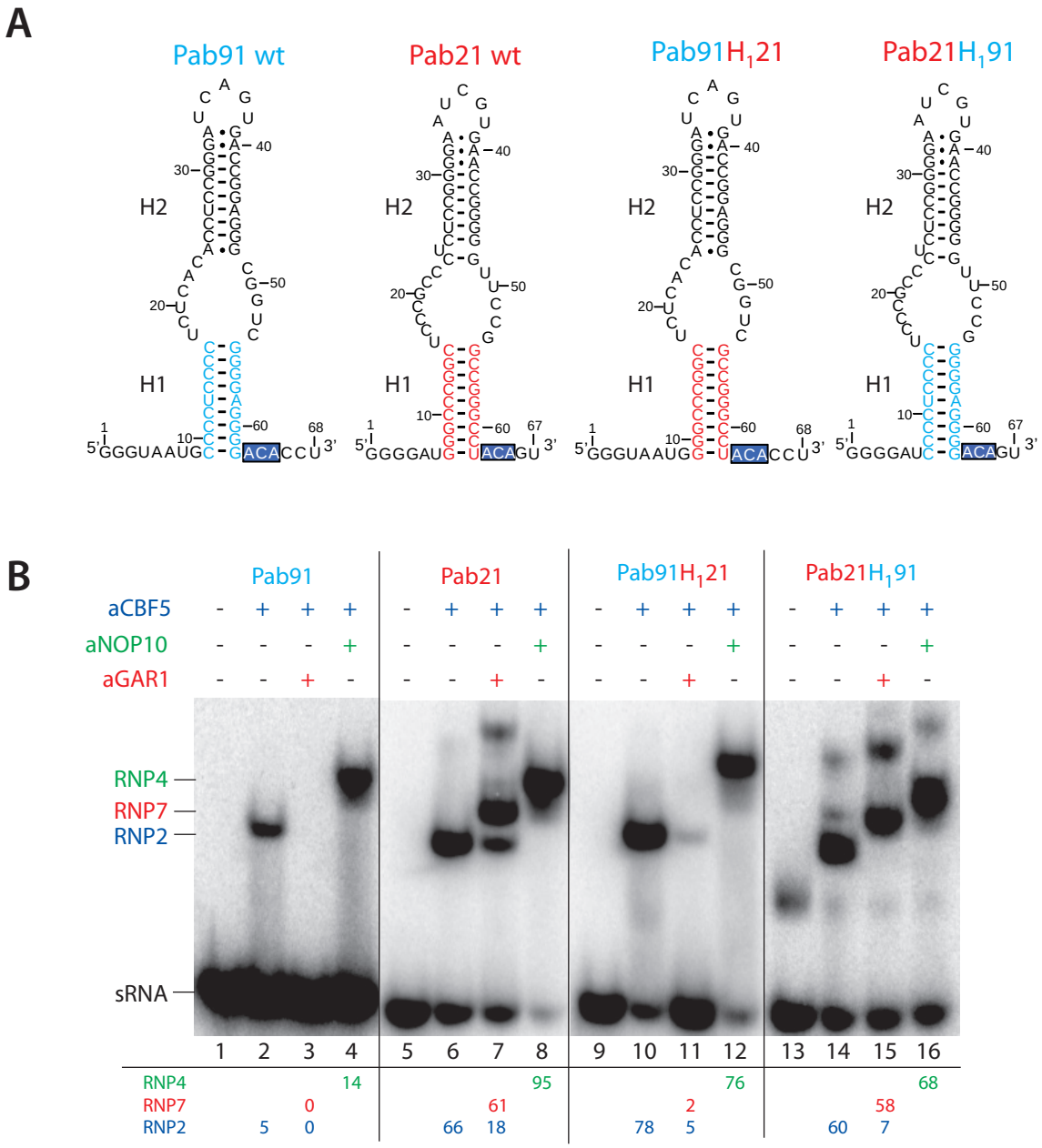


Figure r11 : Etude par expériences de retard de migration sur gel du rôle de l'hélice 1 de sRNA à boîtes H/ACA sur le taux de complexe formé entre l'ARN et le couple de protéines aCBF5-aGAR1.

A. Structures schématiques des sRNA Pab91 et Pab21 sauvages, et des sRNA chimériques Pab91H$_1$21 et Pab21H$_1$91qui portent respectivement la séquence de l'hélice 1 du Pab21 et du Pab91. Identique à la figure r9 A, mais placés ici pour rappel.

B. Résultats d'expériences de retard de migration sur gel réalisées par emploi des différents sRNA indiqués, radiomarqués en 5' et de différentes combinaisons de protéines aCBF5 (bleu), aNOP10 (vert) et aGAR1 (rouge). Le taux de complexe formé entre le sRNA et les différentes protéines est indiqué sous chaque piste.

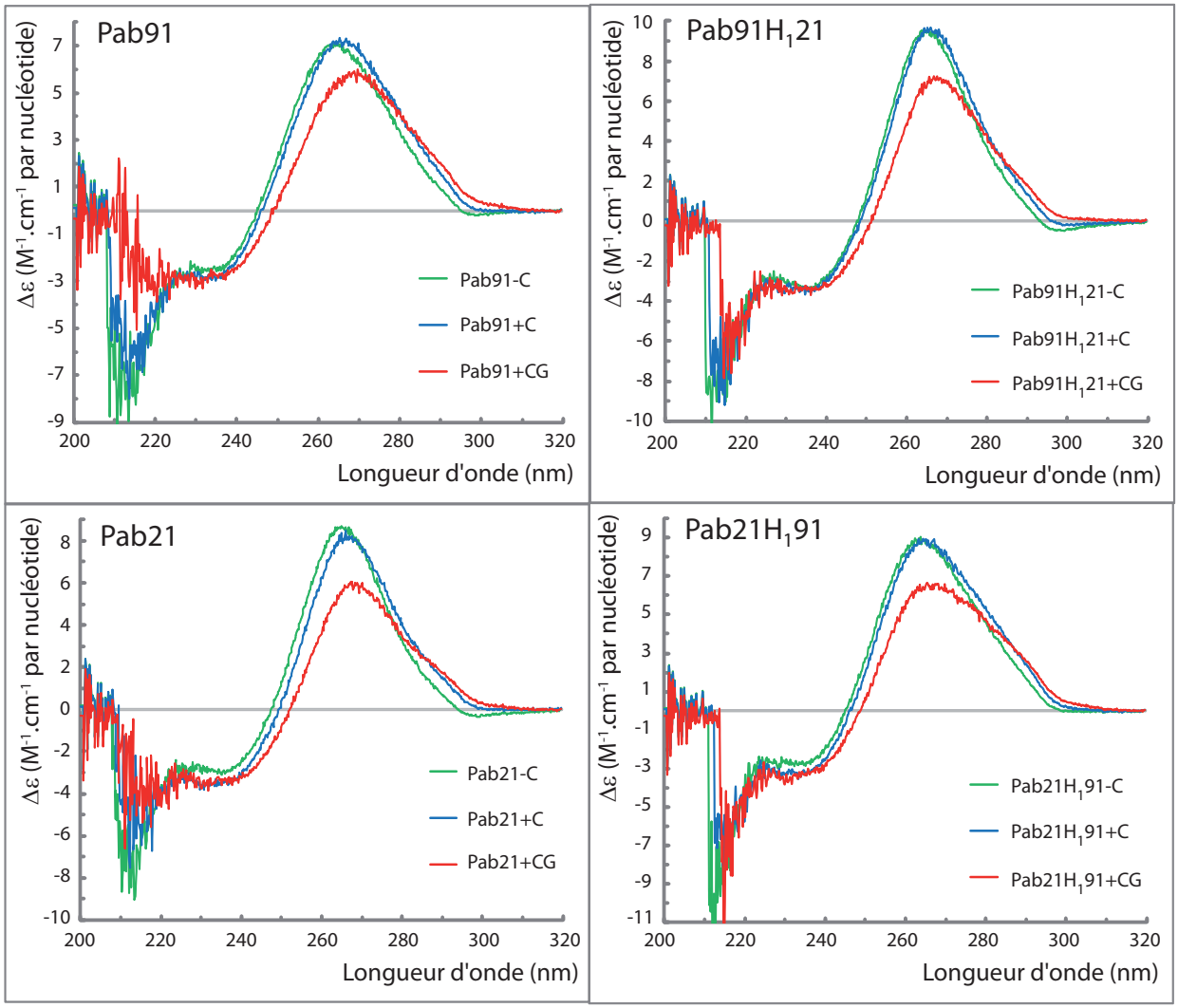
2.6. La contribution des sRNA à l'interaction séquentielle entre aGAR1 et aCBF5

Au cours de la recherche de déterminants ARN favorables à la fixation de aCBF5, nous avons employé les ARN chimériques décrits précédemment dans le but de tester la stabilité du complexe aCBF5-aGAR1 fixé au sRNA. En effet, nous avions constaté qu'en présence de aGAR1, le complexe formé entre le sRNA Pab91 et aCBF5 était déstabilisé et n'était plus observé par retard de migration sur gel, ce que nous avions qualifié par « l'effet négatif de aGAR1 sur l'assemblage de la sRNP » (Fig. r11 piste 3) (Charpentier *et al.*, 2005). Par la suite, nous avons montré qu'un complexe stable pouvait se former entre le sRNA Pab21 et les protéines aCBF5 et aGAR1 (Fig. r11 piste 7). Comme nous l'avons vu précédemment, Pab21 permet de former un plus fort taux de complexe en présence de aCBF5 que Pab21. Mais est-ce qu'un fort taux de complexe formé entre un sRNA et aCBF5 est un élément suffisant pour détecter une interaction stable entre aCBF5 et aGAR1 ?

<u>Etude de la fixation de la protéine aGAR1 au complexe RNP2 par retard de migration sur gel</u>

Pour tenter de répondre à cette question, nous avons employé les ARN chimériques présentés dans les paragraphes précédents. Malgré un fort taux de complexe formé entre aCBF5 et le sRNA Pab91$H_1$21 (78%), aucun complexe sRNA-aCBF5-aGAR1 (RNP7) (2%) n'est détecté et la présence de aGAR1 conduit à une déstabilisation des complexes aCBF5-sRNA (Fig. r11 piste 10 et 11). A l'inverse, un complexe stable du type Pab21$H_1$91-aCBF5-aGAR1 est observé (58%) (Fig. r11 piste 15). Nous concluons que le fort taux de fixation de aCBF5 au sRNA n'est pas l'élément majeur dans le recrutement stable de aGAR1 dans la sRNP. Nous concluons également que le sRNA Pab21 possède des déterminants qui ne sont pas présents dans son hélice H1 et qui lui permettent de recruter efficacement aGAR1 dans le sous-complexe sRNA-aCBF5.

A



B

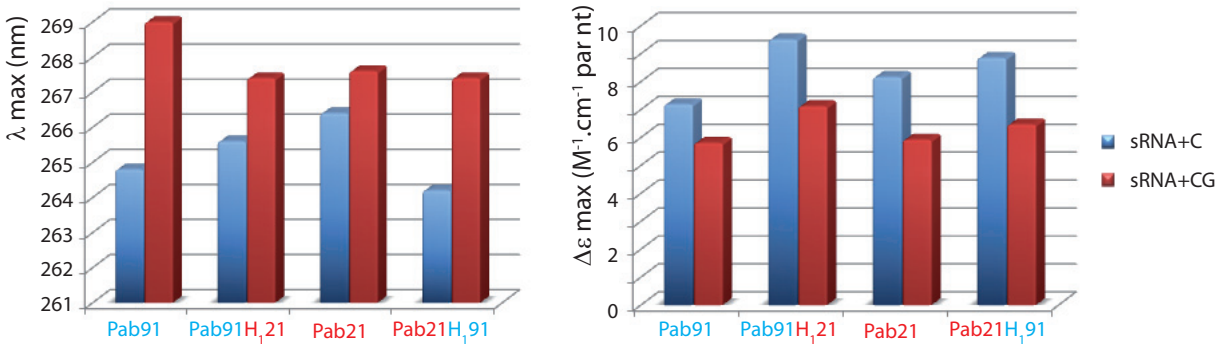


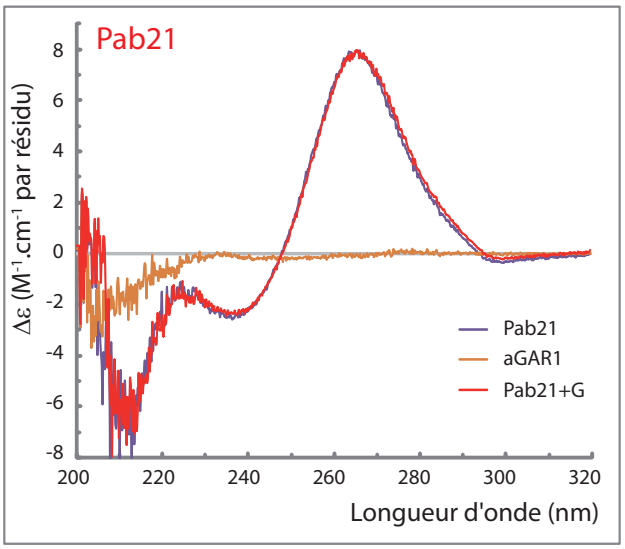
Figure r12 : Etude par dichroïsme circulaire de l'interaction entre le complexe aCBF5-aGAR1 et les sRNA Pab91, Pab91H$_1$21, Pab21 et Pab21H$_1$91.
Les conditions expérimentales sont identiques à celles employées dans la figure r11.
A. Etude du changement de conformation des sRNA indiqués lors de l'association avec le complexe aCBF5-aGAR1. En violet : spectre du sRNA seul en solution. En vert : spectre du sRNA et de aCBF5 dans des compartiments séparés de la cuve. En bleu : spectre mesuré après mélange du sRNA et de aCBF5 dans la cuve. En rouge : spectre mesuré après l'addition séquentielle de la protéine aGAR1 au complexe sRNA-aCBF5.
B. Histogrammes représentant les valeurs des pics positifs obtenus dans la région comprise entre 240 et 300 nm (λmax) et de l'amplitude du pic (Δε max) à cette valeur de λmax mesurée pour les 4 sRNA, à partir des spectres dichroïques de A dont le code couleur est inchangé.

Etude de la fixation de la protéine aGAR1 au complexe RNP2 par dichroïsme circulaire

Nous avons également vu dans le chapitre précédent qu'une interaction faible observée entre le sRNA Pab91 et aCBF5 par retard de migration sur gel, pouvait néanmoins être étudiée par dichroïsme circulaire, comme en témoigne l'effet de la présence de aCBF5 sur la valeur λ max du spectre dichroïque obtenu avec Pab91 (Fig. r9 B). En effet, les conditions employées entre ces deux techniques diffèrent au niveau de la concentration protéique qui est 10 fois plus importante dans le cas du dichroïsme circulaire et à stoechiométrie 1 : 1 avec l'ARN. Pour celui-ci, la concentration pourrait être supérieure au Kd de l'interaction entre aCBF5 et aGAR1 ce qui pourrait expliquer que cette interaction puisse être détectable par dichroïsme circulaire (2 μM) et non par retard de migration sur gel (200 nM).

Nous avons donc recherché par dichroïsme circulaire si la protéine aGAR1 pouvait interagir avec un complexe Pab91-aCBF5. Les spectres de dichroïsme circulaire des composés sRNA, aCBF5 (courbe en vert) présents dans la même cuve mais placés dans deux compartiments séparés, et des composés mélangés dans la même cuve sRNA-aCBF5 (coloré en bleu) ont été reportés sur le graphe à titre indicatif pour visualiser l'effet de la fixation de aCBF5 sur la conformation des sRNA (Fig. r12 A). L'ajout séquentiel de la protéine aGAR1 au complexe sRNA-aCBF5 (RNP2) conduit à un important décalage de λ max vers les plus hautes longueurs d'ondes ainsi qu'à une très forte diminution de Δε max traduisant empilement bien moindre des paires de bases de l'ARN lors de l'association de aGAR1 au complexe sRNA-aCBF5 déjà formé (Fig. r12 A, en rouge). De la même façon, des résultats comparables sont obtenus avec les 3 autres sRNA Pab91$H_1$21, Pab21 et Pab21$H_1$91 (Fig. r12 B à D). Nous constatons donc que la structure globale des 4 sRNA serait remodelée par une forte diminution de l'empilement des paires de bases et par une importante modification de la courbure de l'ARN, lors de l'ajout séquentiel de la protéine aGAR1 au complexe RNP2.

A



B

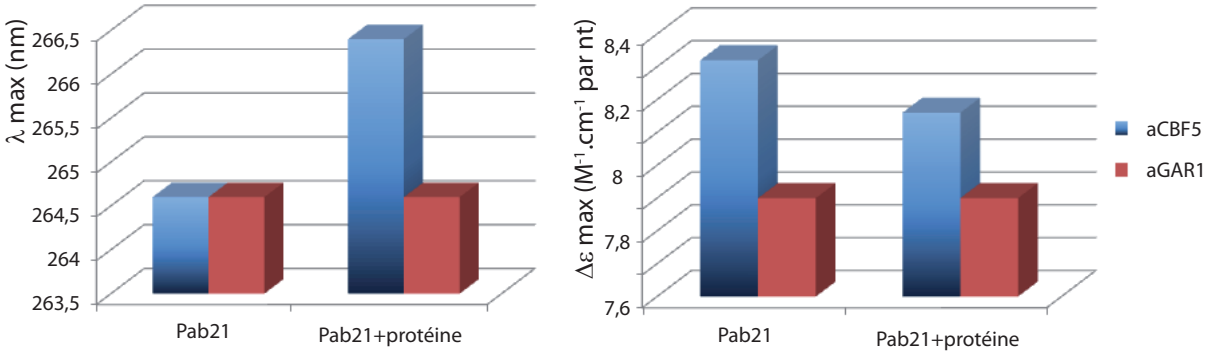


Figure r13 : Etude par dichroïsme circulaire de l'interaction entre le aGAR1 et le sRNA Pab21.
Les conditions expérimentales sont identiques à celles employées dans la figure r11.
A. Etude de l'interaction entre le sRNA Pab21 et la protéine aGAR1 par dichroïsme circulaire. En violet : spectre du Pab21 seul en solution. En orange : spectre mesuré de la protéine aGAR1 seule en solution. En rouge : spectre mesuré après mélange du Pab21 et de aGAR1 dans la cuve.
B. Histogrammes représentant les valeurs des pics positifs obtenus dans la région comprise entre 240 et 300 nm (λmax) et de l'amplitude du pic (Δε max) à cette valeur de λmax mesurée pour l'interaction entre le sRNA Pab21et la protéine aGAR1, à partir des spectres dichroïques de **A** (en rouge) ou tiré d'une autre expérience d'interaction entre le sRNA Pab21 et aCBF5 (en bleu).

Etude de la fixation de la protéine aGAR1 au sRNA Pab21 par dichroïsme circulaire

Aucun complexe stable n'est détecté entre un sRNA et la protéine aGAR1 par retard de migration sur gel (Baker *et al.*, 2005; Charpentier *et al.*, 2005). Nous avons tout de même réalisé un contrôle par mesure de spectres de dichroïsme circulaire pour montrer que aGAR1 n'interagit pas avec un sRNA en absence de la protéine aCBF5 et donc que les effets observés dans la figure r12 proviennent bien de l'interaction spécifique entre aGAR1 et aCBF5. La figure r13 présente le spectre mesuré d'une solution contenant à la fois le sRNA Pab21 et la protéine aGAR1 (Fig. r13 A et B) qui est comparée à celui contenant Pab21 et la protéine aCBF5 (Fig. r13 B). Nous constatons une variation de λ max uniquement lors du mélange de aCBF5 et de Pab21, par contre le mélange entre Pab21 avec aGAR1 ne modifie en rien la structure du sRNA (pas de variation de λ max et de Δε max) (Fig. r13 B). Les résultats de retard de migration sur gel et de dichroïsme circulaire conduisent à la même conclusion, aGAR1 n'interagit pas de façon stable avec le sRNA Pab21.

3. Discussion

3.1. L'affinité apparente de aCBF5 pour l'ARN est variable en fonction de l'identité du sRNA

La fixation de aCBF5 sur le sRNA Pab105 à 2 motifs tige-boucles induit la formation de 2 complexes distincts (Fig. r5 B, RNP2a, RNP2b). Deux complexes sont également retrouvés pour le sRNA Pf3 à double motif de *P. furiosus* (Baker *et al.*, 2005). Chez la levure, l'emploi de Cbf5p et du snoRNA H/ACA snR3 à double motif conduit directement à la formation d'un seul complexe contenant deux molécules de Cbf5p (Normand *et al.*, 2006). De plus, Cbf5p n'est pas capable de se lier à un seul motif du type sRNA d'archaea, dans ce cas la tige-boucle ne comporte aucune extension autre que la boîte ACA (Normand *et al.*, 2006). Cela suggère que dans les snoRNA à boîtes H/ACA eucaryotes, les déterminants pour la fixation de Cbf5p seraient contenus dans l'ensemble des deux tige-boucles. Tandis que chez les archaea, ces déterminants seraient présents dans chaque motif tige-boucle et que les

deux motifs faciliteraient la reconnaissance de aCBF5 pour le sRNA par un phénomène coopératif.

3.2. Une séquence CCA ne peut se substituer à la séquence ACA d'un sRNA

La séquence ACA est un déterminant crucial pour la fixation de aCBF5 au motif en tige-boucle du sRNA. La substitution de la boîte ACA par la séquence CCA est défavorable à la fixation de aCBF5 aux sRNA. Au moment de la réalisation de ces travaux, la structure d'une sRNP H/ACA n'était pas connue et nous avions montré que le domaine PUA reconnaissait la boîte ACA des sRNA (article n°2, (Manival *et al.*, 2006)) et que la substitution de la séquence CCA des ARNt par la boîte ACA augmentait grandement l'affinité de aCBF5 pour l'ARNt (article n°3, (Muller *et al.*, 2007b)). La structure 3D de la sRNP H/ACA indique que le domaine PUA de aCBF5 présente une poche de reconnaissance pour chacune des deux adénines, ce qui conduirait à une forte spécificité du domaine PUA pour la séquence ACA (Li et Ye, 2006). La séquence $C_{74}C_{75}A_{76}$ des ARNt partage les nucléotides C_{75} et A_{76} avec les deux derniers résidus de la séquence ACA. Le domaine PUA de l'enzyme ArcTGT, qui modifie les ARNt, n'emploie pas de poche de reconnaissance pour le nucléotide C_{74} (Ishitani *et al.*, 2003). Par contre les nucléotides C_{75} et A_{76} interagissent par empilement de bases avec des résidus aromatiques du domaine PUA de l'ArcTGT (Ishitani *et al.*, 2003). Ces résidus aromatiques sont absents du domaine PUA de aCBF5 (Li et Ye, 2006), suggérant que celui-ci devrait être positionné d'une façon équivalente sur un ARNt et sur un ARN guide. Un ajustement devrait néanmoins être opéré pour les résidus de la poche de reconnaissance du domaine PUA pour accommoder le résidu C_{74} de la séquence CCA d'un ARNt.

Par contre, les raisons pour lesquelles une séquence CCA n'est pas fonctionnelle lorsqu'elle est présente sur un sRNA ne sont pas élucidées. Il est possible qu'une interaction reposant sur un seul résidu A ne soit pas suffisante pour assurer la stabilité de l'interaction de la protéine sur le sRNA. Ainsi, la séquence

AUA avec deux résidus A est fonctionnelle, bien que moins efficace que la séquence ACA (Fig. r5). Néanmoins, nous avons testé la formation d'un complexe RNP2 sur plusieurs sRNA formés d'une seule tige-boucle et portant la séquence AUA à leur extrémité 3'. Cela n'altère en rien l'affinité de aCBF5 pour ces sRNA (données non montrées). Il est en effet possible que sur les sRNA à double motif, comme nos données sur Pab105 le suggèrent, des phénomènes de coopérativité facilitent la fixation de la deuxième molécule de aCBF5. Il serait intéressant de tester la formation d'un complexe RNP2a sur Pab105 portant une séquence CCA à son extrémité 3'.

3.3. Le rôle de l'hélice H1 des sRNA

L'affinité apparente de aCBF5 pour le sRNA est augmentée par le remplacement de l'hélice H1 de Pab91 par celle de Pab21 tandis qu'elle est diminuée par le remplacement de l'hélice H1 de Pab21 par celle de Pab91. La combinaison des résultats obtenus par les techniques de retard de migration sur gel (Fig. r8), de dichroïsme circulaire (Fig. r9), des mesures d'activité (Fig. r10) et de modélisation (Fig. r10) nous ont permis de conclure que l'hélice H1 est un déterminant majeur pour la reconnaissance par aCBF5 des sRNA à boîtes H/ACA.

Par ailleurs, il a été montré qu'un sRNA uniquement composé de l'hélice H1 et de sa boîte ACA n'était pas capable de fixer aCBF5 mais que pour y parvenir il fallait que la poche de pseudouridylation soit présente sur le sRNA (Baker *et al.*, 2005). Des protections observées dans la région basale de sRNA ou de snoRNA dans des expériences d'empreintes des complexes RNP2 par des sondes chimiques et enzymatiques suggèrent que l'hélice H1 de sRNA et de snoRNA est directement en contact avec aCBF5 (Normand *et al.*, 2006; Youssef *et al.*, 2007) (article n°4, Fourmann *et al.,* en préparation). Ces conclusions ont été validées par la résolution des diverses structures 3D de sRNP (Li et Ye, 2006; Liang, B. *et al.*, 2007; Duan *et al.*, 2009; Liang *et al.*, 2009). Les interactions entre aCBF5 et l'hélice H1 des sRNA sont pour la plupart réalisées entre des résidus chargés de aCBF5 et des groupements phosphates de l'ARN, ce qui ne contribue pas à une spécificité de la séquence des bases, ce qui est

en accord avec le fait que la séquence varie d'un sRNA à l'autre (Li et Ye, 2006; Baker *et al.*, 2008). Néanmoins, la conformation de l'hélice H1 pourrait influencer la formation de ces contacts et par conséquence influencer le taux de formation des complexes entre aCBF5 et un sRNA, permettant ainsi à aCBF5 une reconnaissance du sRNA de type structure-spécifique. Il est aussi possible que l'orientation de l'hélice H1 soit cruciale pour placer la boîte ACA et la poche de pseudouridylation dans un registre permettant le contact simultané du domaine PUA avec la boîte ACA et du domaine catalytique de aCBF5 avec la poche de pseudouridylation.

Par ailleurs, la taille de l'hélice H1 qui est conservée dans les sRNA d'archaea conduit à une distance entre la boîte ACA et le résidu U cible comprise entre 15 à 16 pb (article n°1, (Muller *et al.*, 2008)). Nous avons construit deux variants de l'hélice H1 du sRNA Pab91, l'un présentant un allongement de 4 pb qui permet toujours la fixation de aCBF5 et conserve une partie de son activité (30%), et l'autre portant une hélice H1 raccourcie de 4 pb qui ne permet ni la fixation de aCBF5, ni la détection d'une activité de modification (données non montrées). Ces deux variants possèdent une distance entre la boîte ACA et le résidu U cible de respectivement 11 pb et 19 pb. Il a été montré que, tant que les appariements des deux hélices SH1 et SH2 étaient maintenus, une particule pouvait modifier un ARN substrat avec des sRNA comportant une distance de 12 à 17 pb comprise entre la boîte ACA et le résidu U cible (Liang *et al.*, 2009). Nos résultats indiquent que de plus grandes variations dans la taille de l'hélice H1 conduisent à une perte d'activité des particules.

3.4. L'activité du sRNA Pab21 serait limitée par la séquence de sa poche de pseudouridylation

Nous avons constaté que la cinétique enzymatique de la sRNP Pab21 atteignait un plateau autour de 50% de résidus Ψ formés, et cela quelque soit l'identité de l'hélice H1 (Fig. r10 B). Pourtant les sRNA Pab21 et Pab21$H_1$91 permettent de fixer les protéines avec une affinité élevée. Nous proposons que la réaction catalysée pas les sRNP Pab21 comporte une étape limitante qui, *in vitro*,

conduit à une modification partielle des ARN substrats à la position U891. Nos résultats indiquent que l'assemblage de la particule (étape 1 du mécanisme Fig. i62) est très efficace et ne peut représenter cette étape limitante. Il s'agirait donc d'une étape ultérieure en lien avec la fixation du substrat et le déroulement de la catalyse (étapes 2 à 4 du mécanisme, Fig. i62). De ce fait, nous proposons que la poche de pseudouridylation de Pab21 soit responsable de cette limitation. Il est possible que le substrat conduise à la formation de deux type d'appariements (cf Fig. 4 A ; article n°1 (Muller *et al.*, 2008)). L'appariement fonctionnel permet la modification du résidu U891 dans le dinucléotide en simple brin 5'-$U_{891}U_{892}$-3'. Dans l'appariement non fonctionnel obtenu par un décalage d'un nucléotide, la modification du résidu U892 dans le dinucléotide 5'-U_{892}G-3' n'a pas lieu (Muller *et al.*, 2008). L'observation d'un taux de modification limité à 50% est compatible avec l'hypothèse que ces deux appariements soient en compétition et que seul l'un des deux peut conduire à une pseudouridylation.

3.5. L'effet de aGAR1 sur la stabilité de la sRNP

Le dernier point abordé dans cette partie concerne l'effet de aGAR1 sur la dissociation du complexe formé entre aCBF5 et le sRNA. Nous avons montré qu'un fort taux de fixation de aCBF5 à un sRNA n'était pas un élément indispensable pour recruter aGAR1 dans la sRNP de façon stable. Nous détectons la formation de complexes entre le sRNA Pab91 et les protéines aCBF5-aGAR1 par dichroïsme circulaire mais pas par retard de migration sur gel. Comme déjà mentionné dans le paragraphe I.B.2.6, la concentration en protéine employée dans ces deux approches expérimentales est différente. Dans le cas des gels retards, la concentration en protéine est de 200 nM et en sRNA (0,1 nM), alors qu'une concentration en protéines 10 fois plus importante est utilisée dans les mesures par dichroïsme circulaire (2 µM) et 2000 fois plus importante en ARN (2 µM), et ainsi dépasser la valeur correspondante au Kd de l'interaction entre aCBF5 et aGAR1. Cette augmentation de concentration pourrait expliquer que nous pouvons détecter la fixation de aGAR1 au

complexe préformé Pab91-aCBF5. Ces concentrations plus élevées en protéines sont comparables avec celles employées dans les expériences de retard de migration sur gel visant à étudier l'interaction de aCBF5 avec le sRNA Pf9 (Baker *et al.*, 2005; Youssef *et al.*, 2007; Baker *et al.*, 2008). Ces auteurs observent un déplacement total du sRNA dans un complexe RNP2 pour une concentration de 2 µM en aCBF5 alors qu'un taux plus faible de complexe est détecté à la concentration de 200 nM (Baker *et al.*, 2005; Youssef *et al.*, 2007; Baker *et al.*, 2008).

De plus, la forte diminution de l'empilement des paires de bases et la modification de la courbure de l'ARN observé lors de l'ajout de aGAR1 (Fig. r12), pourrait provenir de l'interaction réalisée entre aGAR1 et la boucle β7/β10 de aCBF5. L'interaction entre aGAR1 et le complexe sRNA-aCBF5 conduirait à des changements de conformation, qui dans le cas de Pab91, auraient un impact sur la stabilité de l'interaction de aCBF5 avec l'ARN, et ce, même en présence des autres protéines aNOP10 et L7Ae. Etant donné le lien fonctionnel entre le pouce de aCBF5 et la poche de pseudouridylation, nous proposons que l'effet déstabilisateur de aGAR1 observé en fonction de l'identité du sRNA serait dépendant de la séquence de la poche de pseudouridylation.

II. Etude des protéines des sRNP à boîtes H/ACA

A. Etude du rôle des domaines des protéines aNOP10 et aCBF5 dans l'assemblage et dans l'activité des particules

1. Présentation du travail

Xavier Manival et Christophe Charron ont résolu au laboratoire la structure 3D de l'hétérodimère aCBF5-aNOP10 par radiocristallographie à une résolution de 2.1 Å. Ces données étaient importantes car ces protéines représentaient le couple minimal nécessaire à l'activité des sRNP à boîtes H/ACA (Charpentier *et al.*, 2005). La structure a révélé une forme allongée de aNOP10, avec un motif en doigt de zinc dans le domaine N-terminal et une hélice α à l'extrémité C-terminale. Les deux régions sont intimement en contact avec le domaine catalytique de aCBF5. Nous nous sommes posés la question de savoir quelle était l'importance relative de chacun de ces deux domaines dans l'assemblage et dans l'activité de la particule.

J'ai évalué la capacité de chaque domaine de aNOP10 et de aCBF5 pour stimuler l'activité des sRNP à boîtes H/ACA. Les données ont montré que le domaine C-terminal de aNOP10 restaurait une partie de l'activité de la sRNP, bien que celui-ci ne se lie pas de façon stable à aCBF5. J'ai aussi montré que le domaine C-terminal de aCBF5 (le domaine PUA), qui est enveloppé par une extension N-terminale de aCBF5 dans la structure 3D, jouait un rôle crucial pour la fixation de aCBF5 au sRNA. La perte de l'activité des particules assemblées avec aCBF5 tronquée de son domaine PUA pouvait être partiellement rétablie par la complémentation de ce domaine produit individuellement et que sa fixation dépendait de la boîte ACA du sRNA.

Ces résultats ont été publiés dans *Nucleic Acids Research* et je suis co-auteur.

Publication n°2 :

Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity.

Xavier MANIVAL, Christophe CHARRON, **Jean-Baptiste FOURMANN**, Bruno CHARPENTIER and Christiane BRANLANT

Nucleic Acids Research, 2006 volume 34(3) : 826-839.

826–839 *Nucleic Acids Research, 2006, Vol. 34, No. 3*
doi:10.1093/nar/gkj482

Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5–aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity

Xavier Manival, Christophe Charron, Jean-Baptiste Fourmann, François Godard, Bruno Charpentier* and Christiane Branlant

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 UHP-CNRS, Université des Sciences et Techniques Henri Poincaré Nancy I, 54506 Vandoeuvre-Lès-Nancy cedex, France

Received December 8, 2005; Revised and Accepted January 14, 2006

ABSTRACT

In archaeal rRNAs, the isomerization of uridine into pseudouridine (Ψ) is achieved by the H/ACA sRNPs and the minimal set of proteins required for RNA:Ψ-synthase activity is the aCBF5–aNOP10 protein pair. The crystal structure of the aCBF5–aNOP10 heterodimer from *Pyrococcus abyssi* was solved at 2.1 Å resolution. In this structure, protein aNOP10 has an extended shape, with a zinc-binding motif at the N-terminus and an α-helix at the C-terminus. Both motifs contact the aCBF5 catalytic domain. Although less efficiently as does the full-length aNOP10, the aNOP10 C-terminal domain binds aCBF5 and stimulates the RNA-guided activity. We show that the C-terminal domain of aCBF5 (the PUA domain), which is wrapped by an N-terminal extension of aCBF5, plays a crucial role for aCBF5 binding to the guide sRNA. Addition of this domain in *trans* partially complement particles assembled with an aCBF5ΔPUA truncated protein. In the crystal structure, the aCBF5–aNOP10 complex forms two kinds of heterotetramers with parallel and perpendicular orientations of the aNOP10 terminal α-helices, respectively. By gel filtration assay, we showed that aNOP10 can dimerize in solution. As both residues Y41 and L48 were needed for dimerization, the dimerization likely takes place by interaction of parallel α-helices.

INTRODUCTION

Ribose 2′-*O*-methylations and conversion of uridines into pseudouridines (Ψ) are the two most frequent post-transcriptional modifications in RNAs (1,2). Several of these modifications were found to be of functional importance for RNA folding or activity (3–6). In bacteria, U to Ψ conversions are catalyzed by RNA:Ψ-synthases consisting of a unique polypeptide, that ensures the RNA substrate recognition and provides the catalytic activity (7–9). In archaea and eukarya, the pseudouridylation and 2′-*O*-methylation reactions can be catalyzed by ribonucleoprotein particles (denoted snoRNPs in eukarya and sRNPs in archaea) [for review, (10,11)]. They consist of a guide RNA (C/D box RNAs for 2′-*O*-methylation and H/ACA box RNAs for pseudouridylation) and a set of 4 to 5 proteins. Recent reconstitution experiments demonstrated the involvement of proteins L7Ae, aCBF5, aNOP10 and aGAR1 in formation of fully active archaeal H/ACA sRNPs (12,13). This revealed a high degree of conservation between archaeal H/ACA sRNPs and eukaryal H/ACA snoRNPs. Indeed, eukaryal snoRNPs contain a set of proteins (Nhp2p, Cbf5p/Dyskerin, Nop10p and Gar1p) very similar to that of archaeal H/ACA sRNPs (14–16).

*To whom correspondence should be addressed. Tel: +33 3 83 68 43 16; Fax: +33 3 83 68 43 07; Email: bruno.charpentier@maem.uhp-nancy.fr
Present address:
François Godard, Institut de Biochimie et de Génétique Cellulaire (IBGC), UMR5095 Université Bordeaux2- CNRS, 1 rue Camille Saint-Saens, 33077 Bordeaux cedex, France

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

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

In both C/D and H/ACA RNPs, the base pair interactions formed by the guide RNAs and the target RNAs define the positions to be modified [reviewed in (10,11)]. The H/ACA pseudouridylation guides have highly conserved characteristic features. As well in archaea, as in eukarya, they contain characteristic stem–loop structures with an internal loop that is complementary to the target RNA (17,18). This stem–loop structure is always flanked by an ANA trinucleotide (more generally ACA). The ACA trinucleotide of archaea H/ACA sRNAs is required for H/ACA sRNPs assembly (12,13). It is needed to recruit protein aCBF5 carrying the conserved motifs I and II found in bacterial RNA:Ψ-synthases. The amounts of aCBF5–sRNA complexes formed by Baker *et al*. (12) were much higher as the ones obtained in our previous study (13). We do not know whether this difference is due to the use by Baker *et al*. (12) of a *Pyrococcus furiosus* aCBF5 protein carrying an additional His-tag sequence, and/or to peculiar properties of the modified *P.furiosus* H/ACA sRNA that was used by these authors. Binding of protein aCBF5 to the guide sRNA is required for recruitment of protein aNOP10 in the RNP (12,13). We showed that this pair of proteins is the minimal set of proteins required to get an RNA guided RNA:Ψ-synthase activity (13). Indeed, using an *in vitro* transcribed H/ACA sRNA, an RNA substrate, and the recombinant proteins aNOP10 and aCBF5, we were able to reconstitute particles that catalyzed U to Ψ conversion at the expected position in the RNA substrate (13). This was at low yields, and addition of the L7Ae and aGAR1 recombinant proteins led to the reconstitution of a fully active H/ACA sRNP (more active than the one obtained by Baker *et al*. (12), this may be due to the presence of His-tags in all the proteins used by these authors). In spite of the presence of the conserved RNA:Ψ-synthase motifs I and II in protein aCBF5, this protein has no detectable RNA:Ψ-synthase activity alone (13). Thus, association of aCBF5 with aNOP10 and the H/ACA guide sRNA is required to stimulate the aCBF5 activity. It was thus of high importance to understand how binding of aNOP10 to aCBF5 may contribute to the aCBF5 activation. As we showed that the aCBF5 and aNOP10 proteins were able to interact together and to form a stable complex in the absence of the guide RNA, we crystallized this complex and solved its 3D structure by X-ray crystallography. The archaeal aCBF5 enzymes share sequence similarity with the bacterial TruB RNA:Ψ-synthase, which is involved in Ψ55 formation in all bacterial tRNAs (19). The 3D structure of the unbound *Escherichia coli* TruB (*ec*TruB) protein has been solved (20), as well as, those of the *E.coli* and *Thermotoga maritima* TruB (*tm*TruB) proteins bound to a tRNA fragment (20,21). Thus, we compared these various structures to that established for the aNOP10–aCBF5 complex. Finally, based on the established aCBF5–aNOP10 structure, we tested the functional importance of some of the aNOP10 and aCBF5 domains by using site-directed mutagenesis. The mutated proteins were used for *in vitro* reconstitution of sRNP particles and the RNA:Ψ-synthase activities of reconstituted particles were measured. In this paper, we describe the 3D structure of the *Pyrococcus abyssi* aCBF5–aNOP10 complex at 2.1 Å resolution, as well as, investigations on the functional roles of the N- and C-terminal domains of aNOP10 and the C-terminal domain of aCBF5.

MATERIALS AND METHODS

Oligonucleotides

Fragments of the aNOP10 and aCBF5 oral reading frames (ORFs) were obtained by PCR amplification with pairs of primers containing some of the primers already described (13). For fragment aNOP10-Nt production, we used the aNOP10-5′ and aNOP10-Nt-3′ primers (5′-CTCGAGTCAGTGGGCTACTTTAGTCTTTTC-3′, the underlined sequence introduces a restriction site). For production of fragment aNOP10-Ct, the primers used were: aNOP10-Ct-5′ (5′-AGATCTCCACCAAGGTTCTCACCCGAGGAT-3′) and aNOP10-3′. For fragment aCBF5ΔPUA production, the amplification was done with primers aCBF5-5′ and aCBF5ΔPUA-3′ (5′-CTCGAGTCAATGTTCAACAGCTTTCTCC-3′), and for fragment PUA production, we used primers PUA-5′ (5′-GGATCCTTGCCTAAGATATGGATAAAG-3′) and aCBF5-3′. PCR amplifications were also used to generate mutations within the aNOP10 C-terminal domain: variants aNOP10L48A and aNOP10Y41A were obtained with the oligonucleotide pairs 5′-GGTGAATACAGAAGGAGGGCTAAGAGGGAACTCCTGGGA-3′ and 5′-TCCCAGGAGTTCCCTCTTAGCCCTCCTTCTGTATTCACC-3′, and 5′-TTCTCACCCGAGGATCCAGCTGGTGAATACAGAAGGAGG-3′ and 5′-CCTCCTTCTGTATTCACCAGCTGGATCCTCGGGTGAGAA-3′, respectively. DNA templates for T7 *in vitro* transcription of the WT Pab91 and PAb91-mtACA sRNAs and for the RNA-S substrate were obtained by PCR amplification as described previously (13).

Recombinant protein production

Recombinant pGEX-6P-1 (Pharmacia) vectors producing the full-length *P.abyssi* proteins L7Ae, aCBF5, aNOP10 and aGAR1 fused to glutathione *S*-transferase (GST) were previously constructed (13). The DNA fragments encoding truncated aNOP10 (aNOP10-Nt and aNOP10-Ct) and aCBF5 (aCBF5ΔPUA and PUA) proteins were PCR-amplified as described above and cloned between the BamHI and XhoI restriction sites of plasmid pGEX-6P-1. The resulting plasmids are denoted pGEX-6P1-aNOP10-Nt, pGEX-6P1-aNOP10-Ct and pGEX-6P1-aCBF5ΔPUA and pGEX-6P1-PUA, respectively. The recombinant GST-fusion proteins were produced in *E.coli* BL21 CodonPlus cells (Novagen). They were purified under native conditions using Glutathione–Sepharose 4Fast Flow (Pharmacia), and the archaeal proteins were eluted from the beads after cleavage with the PreScission Protease as described previously (22).

Conditions used for protein crystallization

Protein aCBF5 was mixed with protein aNOP10 at a molar ratio of 1:1.2. Crystals of the *P.abyssi* aCBF5–aNOP10 complex were grown by vapor phase diffusion in hanging drops. Drops were made at 293 K by mixing 1 µl of the concentrated proteins mixture (70 mg ml^{-1}) and 3 µl of a reservoir solution containing 1 M di-potassium phosphate and 100 mM sodium acetate at pH 5.5. Crystals were flash-frozen in liquid ethane in the mother liquor with addition of 30% (v/v) glycerol as a cryoprotectant. The aCBF5–aNOP10 crystals were in space group $P2_12_12$ with unit-cell parameters $a = 136.4$ Å, $b = 136.8$ Å, $c = 59.3$ Å. There was two aCBF5–aNOP10

complexes per asymmetric unit, consistent with a Matthews's coefficient of 3.27 Å^3 Da^{-1} and the solvent content is of 62.1%. Crystals of SeMet-incorporated proteins were obtained under conditions similar to those for the wild-type complex.

Data collection and structure determination

SAD data from a (SeMet)-labeled crystal flash-frozen at 100 K were collected at the absorption peak up to 2.7 Å resolution at the BM14 beamline of the European Synchrotron Radiation Facility (ESRF, Grenoble). The dataset was indexed and scaled using HKL2000 (23). Eighteen of the 20 possible selenium sites were found and refined at 2.7 Å resolution using SOLVE (24), which produced a mean figure of merit of 0.41 and an overall score of 70. After density modification with RESOLVE, the mean figure of merit was of 0.73. The experimental electron density map was clearly interpretable and allowed tracing of both the aNOP10 and aCBF5 molecules, which is consistent with the positions of all 18 selenium sites used for phasing. The model was built using O (25) and refined by CNS (26) with the maximum likelihood target using amplitude and experimental phase distributions.

A native dataset at 2.1 Å resolution was collected at 100 K on beamline ID14-4 at the ESRF. Data were indexed and scaled up using XDS (27).

The initial aCBF5–aNOP10 model was improved by alternating cycles of simulated annealing refinement against the native dataset using CNS (26) and manual remodeling using O (25). The final model was refined to an R_{factor} of 22.4% and an R_{free} of 25.2% (Table 1) and includes residues 12–138 and 153–334 of aCBF5 (chain A), residues 12–138 and 151–334 of aCBF5 (chain C), residues 4–56 of aNOP10 (chain B and D), 5 phosphate ions, 2 zinc ions and 250 water molecules.

Table 1. Data collection and refinement statistics

Data collection		
Datasets	Native	SeMet
Wavelength (Å)	0.9686	0.9777
Resolution (Å)[a]	30–2.1 (2.2–2.1)	50–2.7 (2.8–2.7)
R_{sym} (%)[a,b]	6.8 (50.8)	10.8 (41.1)
Multiplicity	7.2	12.9
$I/\sigma I$[a]	15.6 (4.9)	9.7 (2.1)
Completeness (%)[a]	99.8 (100)	95.7 (96.9)
Refinement and model quality		
Resolution (Å)	30–2.1	
Reflections work set	62 200	
Reflections test set	3334	
R_{factor}(%)[c]	22.4	
R_{free}(%)[d]	25.2	
No. of protein atoms	5777	
No. of water molecules	250	
R.m.s.d. bonds (Å)	0.006	
R.m.s.d. angles (°)	1.49	
Ramachandran plot (%)		
Most favored	91.7	
Favored	8.0	
Allowed	0	
Disallowed[e]	0.3	

[a]The number in parentheses is for the highest resolution shell.
[b]$R_{sym} = \sum_{hkl}^{i} |I^i_{hkl} - \langle I_{hkl}\rangle| / \sum_{hkl} |I_{hkl}|$ where i is the number of reflections hkl.
[c]$R_{factor} = \sum_{hkl} ||F_o\,(hkl)| - |F_o\,hkl|| / \sum_{hkl} |F_o\,(hkl)|$.
[d]R_{free} was calculated using 5% of the reflections left out of refinement.
[e]Residue Glu94 ($\phi = 43°$; $\psi = -115°$) in aCBF5 is located in the i + 1 position of a βII′ turn.

Gel filtration experiments

Gel filtration experiments were performed using an AKTA explorer high-performance liquid chromatography (HPLC) (Amersham Biosciences) system, Superdex 75 or 200 HR columns (Amersham Biosciences) at a constant flow rate (0.5 ml/min) using 20 mM HEPES buffer (pH 7.5), 1 mM DTT, 300 mM NaCl. Proteins were detected at 280 nm using a dual-wavelength ultraviolet (UV) detector (260–280 nm). SDS–PAGE gel analysis of each fraction was performed to confirm the assignments of the retention peaks. For molecular mass estimation, the LMW Gel Filtration Calibration Kit and HMW Gel Filtration Calibration Kit (Amersham Biosciences) were applied to the Superdex 75 and 200 column, respectively.

Electrophoresis mobility shift assays (EMSA)

The DNA sequence of the sRNAs Pab91-WT and Pab91-mtACA were PCR-amplified from *P.abyssi* GE5 genomic DNA, using a forward primer containing a T7 promoter (13). These DNA fragments were used as templates for *in vitro* transcription of ^{32}P-labeled sRNAs (13). sRNP complexes were formed as described previously (13). Briefly, 50 fmol of a labeled sRNA was mixed in buffer D [150 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA and 20 mM HEPES (pH 7.9)] with recombinant full-length or truncated proteins (200 nM each) and the mixture was incubated for 10 min at 65°C. The sRNPs formed were resolved in 6% non-denaturing polyacrylamide gels. To test for the Pab91 sRNP target RNA association (complex CII), 2.5 pmol of the previously described (13) RNA-S target RNA, which is complementary to the pseudouridylation pocket of sRNA Pab91, were added in the incubation mixture. The dried gels were analyzed with a phosphorimager (Typhoon 9410, Amersham Biosciences) and the amounts of radioactivity in the bands were estimated with the ImageQuant software. The percentage of radioactivity in each of the RNP band as referred to the total radioactivity in the lane was systematically established. All the gel shift experiments were repeated three times and representative examples are shown in the figures.

In vitro RNA:pseudouridylation assays

The RNA:Ψ-synthase activities of the reconstituted sRNPs were tested by the nearest-neighbor approach (7). Conditions for time course analysis of the *in vitro* pseudouridylation reaction are detailed in (13). Typically, 4 pmol of unlabeled sRNA and 150 fmol of target RNA, that was labeled by [α-^{32}P]CTP incorporated in the course of transcription, were mixed at 65°C with the wild-type or variant L7Ae, aCBF5, aNOP10 and aGAR1 proteins. Aliquots were collected at different time points and the reaction was stopped by phenol–chloroform extraction followed by ethanol precipitation. The recovered target RNAs were digested with RNase T2. The resulting 3′-mononucleotides were subjected to monodimensional chromatography on thin-layer cellulose plates, as described previously (13). The radioactivity in the bands were quantified with a phosphorimager (Amersham Biosciences) using the ImageQuant software. The amounts of Ψ formation as a function of time were determined taking into account the total number of U residues in the target RNA.

Coordinates

The coordinates of the *P.abyssi* aCBF5–aNOP10 complex have been deposited in the Protein Data Bank (accession code 2AUS).

RESULTS

Crystals and structure determination

The *P.abyssi* proteins aCBF5 and aNOP10 were mixed at a molar ratio of 1:1.2 and good quality crystals were obtained (see Materials and Methods for details on crystallization conditions and crystal analysis). The structure was solved by single anomalous diffraction (SAD). The initial phases were calculated to 2.7 Å, by using the SAD data collected from a selenomethionine (SeMet)-labeled crystal of the aCBF5–aNOP10 complex. Selenium was found at 18 of the 20 possible selenium sites. Phase improvement by means of density modification resulted in a well defined electron density map, that allowed us to build a 3D model for both the aCBF5 and aNOP10 molecules. The final structure of the aCBF5–aNOP10 complex was refined against the dataset collected from a native crystal to an R_{factor} of 22.4% and an R_{free} of 25.2%, including all reflections between 30 and 2.1 Å resolution (Table 1). The asymmetric unit was found to contain two aCBF5–aNOP10 complexes, 5 phosphate ions, 2 zinc ions and 250 water molecules. Superimposition of the two independent aCBF5–aNOP10 heterodimers in the asymmetric unit yields a r.m.s deviation of 1.08 Å for 1424 backbone atoms. In the established structure, over 91% of the residues are within the most favored regions in a Ramachandran plot, as defined by *PROCHECK* (28). Only well-ordered Glu94 of both aCBF5 molecules have unfavorable main-chain torsion angles ($\phi = 43°$; $\psi = -115°$), but these torsion angles are consistent with the location of Glu94 in the i + 1 position of a βII′ turn (29,30).

Structure overview of the aCBF5–aNOP10 complex

In the crystal, the aNOP10 structure consists of a zinc-binding motif at the N-terminal end, a linker region and a α-helix at the C-terminal end (Figure 1A). The overall structure presents no structural homology with previously determined protein structures, as shown by a Dali search (http://www.ebi.ac.uk/dali/). However, the zinc-binding motif (Figure 1B), that contains a cluster of four cysteines (C8-X_2-C11-X_8-C20-X_2-C23), can be superimposed with a similar cluster present in the *P.abyssi* factor aIF2 (residues 56–81). The deviation for 26 Cα atoms is of 0.41 Å r.m.s. The aIF2 factor is the archaeal counterpart of the eukaryotic translation initiation factor IF2-α (19). The electron density map of aNOP10 clearly shows the presence of a tetra-coordinated metal and a zinc atom was detected by X-ray fluorescence measurement. Moreover, alignment of the aNOP10 sequences from various archaeal species strongly indicates the presence of a zinc knuckle in all members of this protein family. Furthermore, the four cysteines are strictly conserved (Figure 1C). In contrast, this cysteine cluster is not found in the eukaryal Nop10 proteins (data not shown). In the crystal, each of the aNOP10 proteins are bound with one aCBF5 molecule, and two independent aCBF5–aNOP10 heterodimers are each related to an identical heterodimer via crystallographic 2-fold axes. They form heterotetramers by interaction of two aNOP10 proteins (Figure 1D and E).

The structure of aCBF5 consists of two distinct domains (Figure 2A): a large catalytic domain and a small C-terminal domain corresponding to a PUA domain (pseudouridine synthase and archaeosine transglycosylase domain) (31). The overall fold of aCBF5 is similar to that of the *tm*TruB enzyme (20), with a 2.95 Å r.m.s. deviation for 257 Cα atoms (Figure 2C). Comparison of the aCBF5 and *tm*TruB catalytic domains revealed two major differences: the deletion of two sequences of 17 residues, and the insertion of an additional α-helix (11 residues) in protein aCBF5 compared to *tm*TruB. Interestingly, this additional α-helix is also found in *ec*TruB (helix α6) (21). The two Ins1 and Ins2 sequences, which are deleted in aCBF5, are present in both *tm*TruB and *ec*TruB and are located in the thumb-loop. These Ins1 and Ins2 sequences are considered to be characteristic features of the TruB RNA:Ψ-synthase family (21). Comparison of the aCBF5 and *tm*TruB C-terminal domains reveals several differences. First, in aCBF5 the sequence between the β-strands β16 and β17 is lacking 4-residues compared to that in *tm*TruB and the resulting truncated sequence forms an additional α-helix (α9). Secondly, seven additional residues extend the C-terminus of aCBF5 compared to that of *tm*TruB. Finally, 33 additional residues are present at the N-terminus of aCBF5. They are wrapping the C-terminal PUA domain and form one additional β-strand (β18), that is located at one extremity of the β-sheet (Figure 2A). To superimpose the C-terminal domain of aCBF5 with that of the RNA-free *ec*TruB, a rigid body movement of the tip of the C-terminal domain toward the catalytic domain of aCBF5 with a maximum displacement of 5 Å is needed (Figure 2B). Superimposition of aCBF5 with the RNA-bound *ec*TruB requires a second rigid body movement of the C-terminal domain toward the RNA of the *ec*TruB–RNA complex, with a maximum displacement of 5 Å at the tip of the C-terminal domain (Figure 2C). The latter movement is also needed to superimpose the *ec*TruB RNA-free form with the RNA-bound *tm*TruB (20).

aCBF5, dyskerin and TruB share a common active site

The aCBF5 catalytic domain folds into an extensive β-sheet. At one side of the faces of the sheet, a cleft is delineated by helices α2 and α5 and loops β11-α5, β2-β3 and α2-β4. According to previous analysis on the TruB enzyme, the active site lies at the bottom of this cleft (21). Indeed, an aspartate residue (Asp82 in aCBF5, Aps48 in *ec*TruB and Asp125 in the human dyskerin) is conserved in all known RNA:Ψ-synthases (32,33), and it was already shown to be essential for RNA:Ψ-synthase activity (13), as well in the TruB enzymes (34), as in protein aCBF5 (35).

In the TruB enzymes, two tyrosine residues (Tyr76 and Tyr179 in *ec*TruB) were shown to play an important role in orientation of the target uracil base (21). These residues are conserved in aCBF5. In the RNA-bound *ec*TruB, Tyr179 stacks against the sugar of the uracil substrate, and the aromatic ring of Tyr76 makes a stacking interaction with the uracil base (21). Both interactions are proposed to orientate the C6 atom towards the carboxylate of the catalytic aspartate

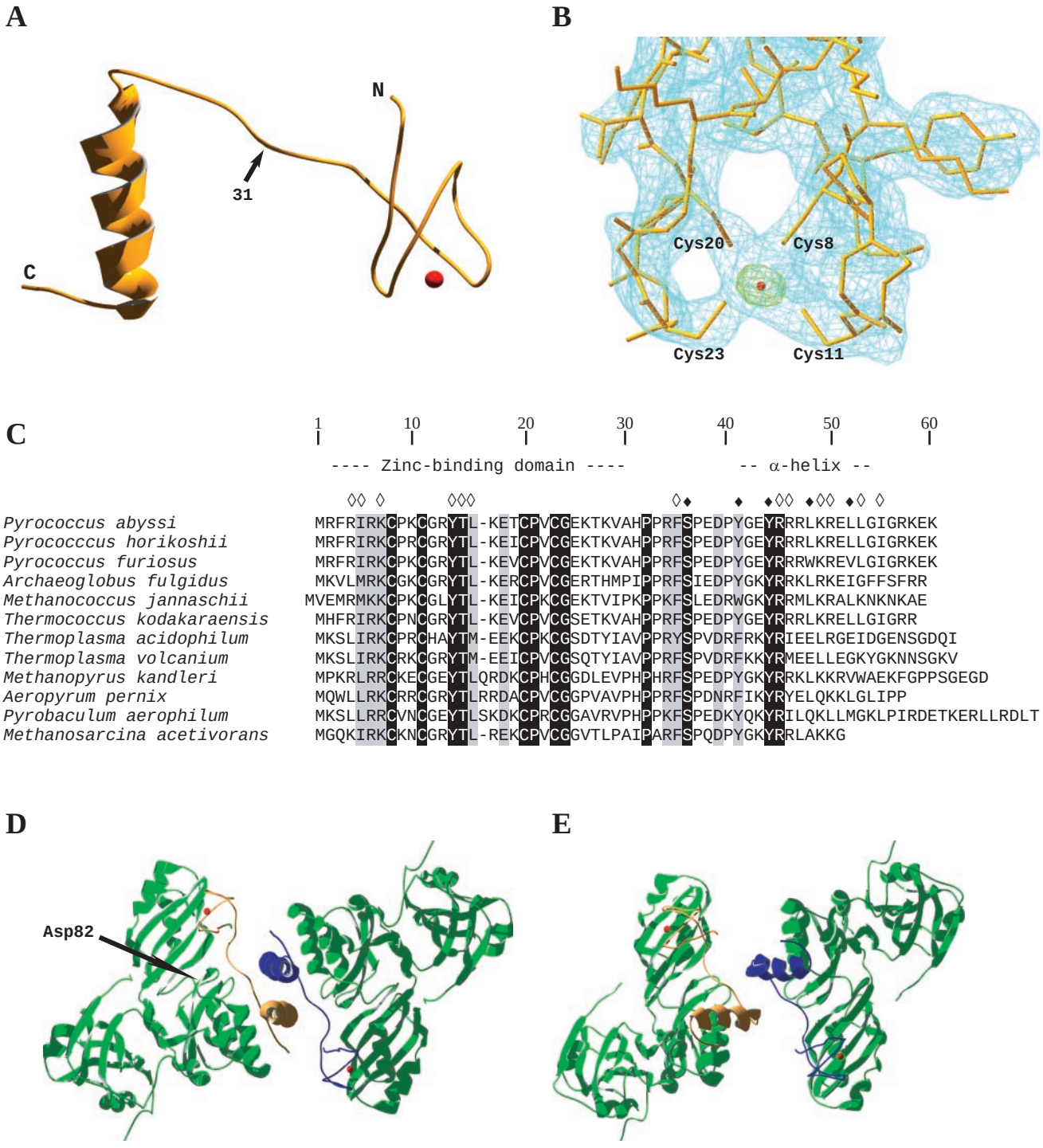


Figure 1. The *P.abyssi* aNOP10 3D structure in aCBF5–aNOP10 crystals (**A**) Ribbon representation. The Zn^{2+} ion detected in the crystal structure is represented by a red ball. The arrow indicates the position of the C-terminal end truncation in the aNOP10-Nt variant protein produced in this study (see Figure 4A). (**B**) Close-up view of the 2Fo-Fc electron density map (blue), contoured at 1.2 σ, in the region of the zinc-binding domain. The zinc atom is drawn in red. The simulated annealing omitted Fo-Fc map (green) of the zinc ion is contoured at 5 σ. (**C**) Sequence alignments of archaeal aNOP10 proteins. Residues involved in the side-chain interactions between proteins aNOP10 and aCBF5 and in the aNOP10 homodimerization are indicated by white and black diamonds, respectively. Sequences are retrieved from the SWISS-PROT database. (**D** and **E**) Ribbon representation of the two distinct aCBF5–aNOP10 heterotetramers formed by aNOP10 dimerization and found in the crystal.

to favor the reaction (21). It is also proposed that Tyr179 acts as a general base for the proton abstraction, which is required in the last step of Ψ formation (36). An Arg residue (Arg181 in *ec*TruB) also participates to the active site by formation of a salt bridge and two hydrogen bonds with the catalytic Asp48 residue (21). The superimposition of the 3D structure of the aCBF5 catalytic domain on to those of the RNA-free and RNA-bound *ec*TruB structures (20) shows the presence

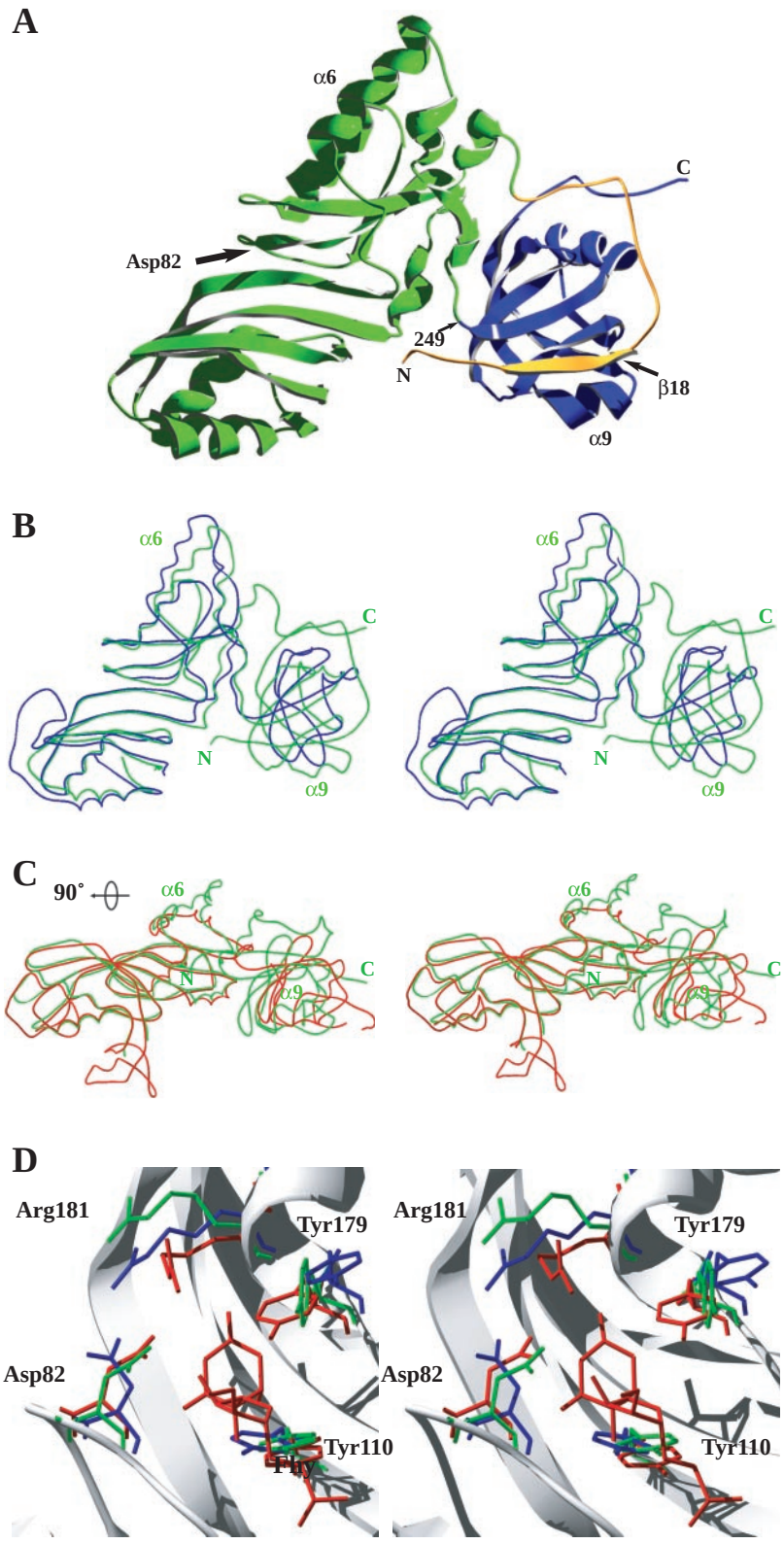


Figure 2. The *P.abyssi* aCBF5 structure in aCBF5–aNOP10 crystals and its comparison with the *E.coli* and *T.maritima* TruB 3D structure (*ec*TruB and *tm*TruB, respectively) (**A**) Ribbon representation of the established aCBF5 structure. The position of the C-terminal extremity (249) of the variant aCBF5 protein aCBF5ΔPUA produced in this study is represented (see Figure 4). The N-terminal extension is drawn in yellow. (**B**) Superimposition of the aCBF5 3D structure (green) with that of the free *ec*TruB (blue) (20). (**C**) Superimposition of the aCBF5 3D structure (green) with that of the RNA-bounded *tm*TruB (red) (20). Note that aCBF5 is oriented differently (rotation by 90°) in (B) and (C). (**D**) Superposition of the catalytic residues of protein aCBF5 (green) with the corresponding residues from the RNA-free *ec*TruB protein (blue) (20) and the RNA-associated *tm*TruB protein (red) (20). The aCBF5 amino acids numbering is used. The position identified for 5-fluoro-6-hydroxypseudouridine (5FhΨ) after catalysis by the *ec*TruB protein is indicated (21).

in the active site of aCBF5 of four residues (Asp82, Tyr110, Tyr179 and Arg181, respectively), which are homologous to the *ec*TruB Asp48, Tyr76, Tyr179 and Arg181 residues, respectively. The two sets of four residues have similar locations in the active sites of both enzymes (Figure 2D). As, the conformations of these residues are modified upon RNA binding on the *ec*TruB enzyme, it was also interesting to compare the conformations of their lateral chains in the available 3D structures. Upon RNA binding, the Asp48 conformation is modified, as it forms an ionic interaction with residue Arg181 (20). On the other hand, Tyr179 moves toward and clamps the uracil ring into a productive orientation for catalysis (20). Interestingly, the conformation of the catalytic Asp82 residue in the aCBF5 structure is closer to that of the catalytic Asp48 residue in the RNA-bound forms of *ec*TruB and *tm*TruB, than to the one of this residue in the unbound form of *ec*TruB (Figure 2D). Remarkably also, the position of Tyr179 in aCBF5 is closer to that in the bound forms of *ec*TruB and *tm*TruB, than to that in free *ec*TruB. Thus, in the aCBF5–aNOP10 complex, residues of the aCBF5 catalytic site likely have a conformation, which is favorable for catalysis of the U to Ψ conversion.

aNOP10 interacts with the catalytic domain of aCBF5

In the crystal structure, the aCBF5–aNOP10 interface (1292 Å^2) involves 26% of the total solvent-accessible surface area of aNOP10 and 9% of the total solvent-accessible surface area of aCBF5. Indeed, the surface of aNOP10 is strongly involved in the heterodimer interface and half of this interface is nonpolar. The side-chain interactions formed in the crystal are both of hydrophobic and of ionic types (Figure 3). Two distinct hydrophobic surfaces are observed at the heterodimer interface. The first one includes the hydrophobic residues at the surface of the zinc-binding motif. They are involved in formation of a hydrophobic cluster with hydrophobic residues from the contiguous β-strands β4 and β11 of protein aCBF5. The second hydrophobic surface involves residues on one face of the C-terminal α-helix in protein aNOP10 and hydrophobic residues of the helices α1, α6 and α7 in protein aCBF5. Besides these hydrophobic contacts, basic residues of the C-terminal α-helix and at the surface of the zinc-binding domain of protein aNOP10 form ionic interactions with acidic residues in protein aCBF5. These aCBF5 residues belong to helix α6 and to the loops located upstream of β11 and α6, respectively. Moreover, one phosphate ion is found at the interface between proteins aNOP10 and aCBF5 and it interacts with the side-chains of residues Arg46 and Arg50 from aNOP10 and His225 from aCBF5.

Distinct roles of the aNOP10 N- and C-terminal domains

In our recent study (13), we showed that a complex designated RNP3 (Figure 4B, lane 2) is formed by direct interaction between proteins aCBF5 and L7Ae and the Pab91 sRNA. In the presence of protein aNOP10, complex RNP3 is converted into an RNP5 complex (Figure 4B, lane 4). As protein aNOP10 did not show direct sRNA binding activity, protein aCBF5 is proposed to drive the incorporation of protein aNOP10 into the sRNP (12,13). In addition, protein aNOP10 is required for incorporation of the RNA substrate

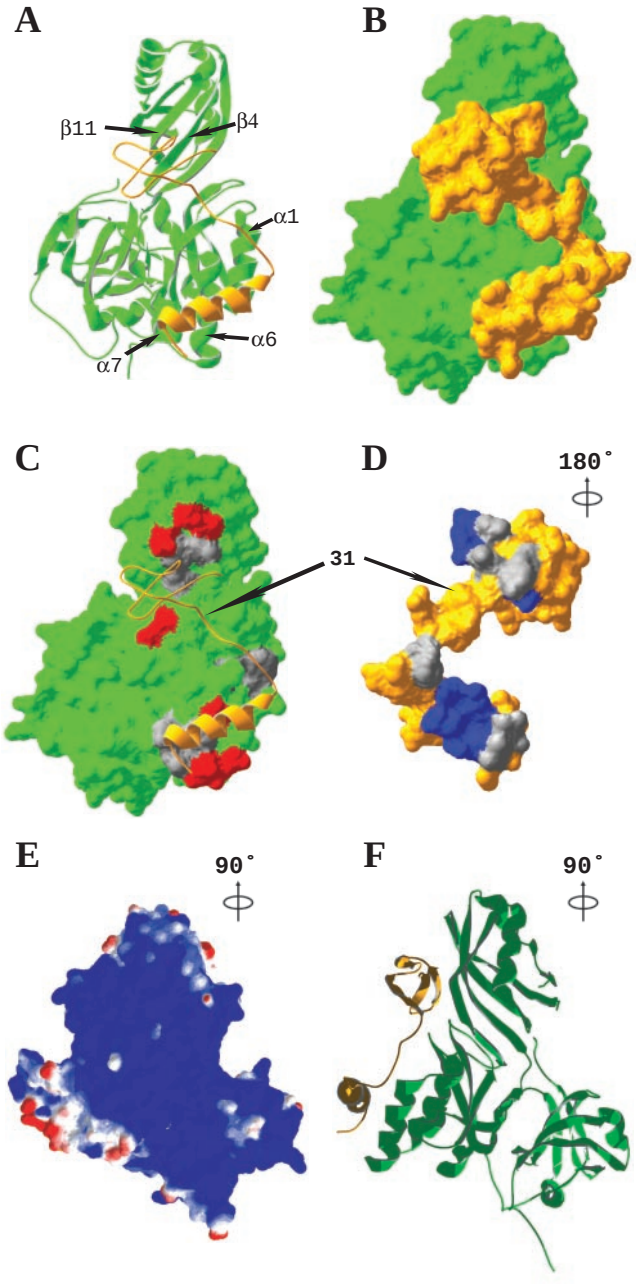


Figure 3. Interface between proteins aNOP10 and aCBF5 in the crystal (**A**) Ribbon representation of the aCBF5–aNOP10 heterodimer and (**B**) molecular surface of the aNOP10 (orange) and aCBF5 (green) surface in the crystal. (C, D) Hydrophobic (gray), acidic (red) and basic (blue) residues at the aCBF5–aNOP10 interfaces are represented in aCBF5 (**C**) and aNOP10 (**D**). (**E**) Calculated electrostatic potential for the aCBF5–aNOP10 complex mapped on its molecular surface. Positive and negative potential are drawn in blue and red, respectively. (**F**) Ribbon representation of the aCBF5–aNOP10 complex in the same orientation as in (E). Views in (A–C) have the same orientation; view in (D) is rotated by 180° and the aCBF5–aNOP10 structure in (E and F) is rotated by 90°.

(Figure 4B, lanes 3 and 5) (13) leading to the formation of the CII complex (marked by asterisks in Figure 4B). This complex has a higher electrophoretic mobility compared to complex RNP5 (Figure 4B, lanes 4 and 5). As both the N- and C-terminal fragments of aNOP10 interact with aCBF5

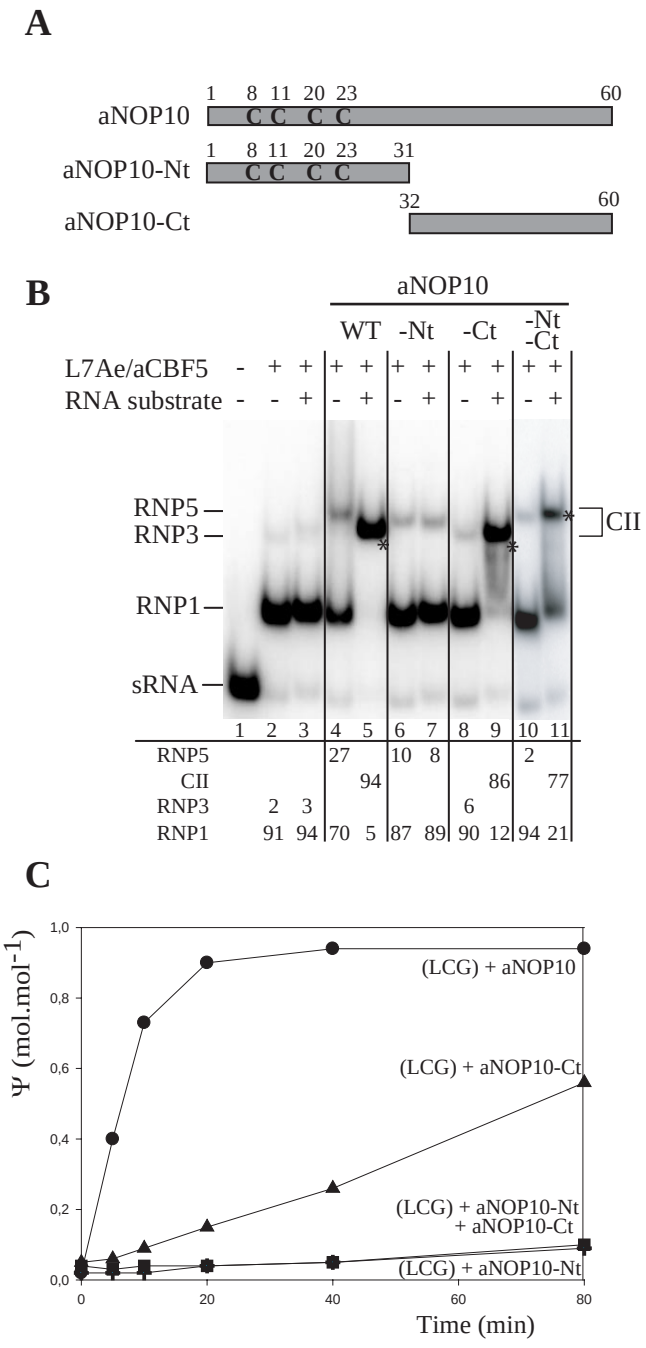


Figure 4. Analysis of the functional roles of the N- and C-terminal domains of protein aNOP10. (**A**) Schematic representation of the aNOP10 fragments used in this study. The four cysteine residues of the zinc finger domain of protein aNOP10 are indicated. (**B**) Analysis by gel mobility shift assays of complex CII formation. The radio-labeled Pab91 sRNA (50 fmol) was incubated with the L7Ae and aCBF5 proteins (200 nM each) and, as indicated on top of the lanes, either the WT aNOP10 protein (lanes 4 and 5), the Nt or Ct aNOP10 fragments (lanes 6–9) or both fragments (lanes 10 and 11). Formation of complex CII (marked by an asterisk) was tested in the presence of the RNA substrate, as indicated on top of the lanes. The percentages of Pab91 sRNA present in each complexes as referred to the total RNA amount in the lane is indicated below each lane. (**C**) Time course analyses of Ψ formation in the RNA substrate. The unlabeled Pab91 sRNA was mixed with the labeled RNA substrate and the three L7Ae, aCBF5 and aGAR1 proteins (LCG) plus either the full-length aNOP10, fragment aNOP10-Nt, fragment aNOP10-Ct or the two aNOP10 fragments. The amounts of Ψ residue formed were estimated by 1D TLC. They are expressed in moles of Ψ residue formed per moles of RNA substrate.

according to our 3D structure analysis, we tested whether one of them or both are required to recruit the RNA substrate during complex CII formation. To this end, two aNOP10 fragments (aNOP10-Nt and aNOP10-Ct) were produced and purified. The aNOP10-Nt fragment contained the zinc-binding domain and the aNOP10-Ct fragment carried the 29 C-terminal residues with the α-helix (Figure 4A). Interestingly, the formation of an RNP5-like complex (containing proteins L7Ae, aCBF5 and aNOP10-Nt) demonstrated the association of fragment aNOP10-Nt with the aCBF5 protein bound to the guide. However, no CII complex was detected, upon addition of the RNA substrate (Figure 4B, lanes 6 and 7). On the contrary, although no RNP5-like complex was formed upon incubation of protein aCBF5 with the guide RNA and the aNOP10-Ct fragment, a complex expected to correspond to the CII complex was assembled in the presence of the RNA substrate (Figure 4B, lanes 8 and 9). When formation of complex CII was tested in the presence of the two aNOP10-Nt and aNOP10-Ct fragments, a complex with an electrophoretic mobility lower than that of complex CII was obtained (Figure 4B, lane 11). The RNA:Ψ-synthase activity of the reconstituted sRNPs was measured by use of the nearest-neighbor approach described in Materials and Methods. According to the data obtained, the aNOP10-Ct fragment was sufficient to get an active particle (Figure 4C). Indeed, after a 80 min incubation of the guide sRNA and the RNA substrate with the L7Ae, aCBF5, aGAR1 and aNOP10-Ct protein set, more than 50% of the RNA substrates were modified. In contrast, a very low activity was obtained when using the aNOP10-Nt fragment. Hence, the C-terminal domain of aNOP10 alone is able to stimulate the aCBF5 activity. Nevertheless, the kinetics of the reaction was slower as compared to that obtained for the full-length aNOP10 protein. Surprisingly, in the presence of the two aNOP10-Nt and aNOP10-Ct fragments, the rate of the reaction was very low. It was similar to that found for the aNOP10-Nt fragment alone. Thus, addition of the aNOP10-Nt fragment counteracted the positive effect of the aNOP10-Ct fragment. Taken together, we concluded that the full stimulation of the aCBF5 activity requires the aNOP10 C- and N-terminal domains and the integrity of the linker fragment.

We then tested whether the presence of the Zn^{2+} ion was important for sRNP activity. To this end, a aNOP10C11S variant protein lacking one of the four cysteine residues required for Zn^{2+} binding was produced. The mutation had no effect on RNP5 and CII complex formation, and on the catalytic property of the reconstituted particles (data not shown). We, thus, concluded that the presence of a Zn^{2+} ion in aNOP10 is not required for RNP assembly and activity.

The PUA domain is required for aCBF5 binding to the sRNA

To test whether the PUA domain has a functional role in H/ACA sRNP assembly and function or whether the aCBF5 catalytic domain is sufficient for production of active sRNPs, we produced an aCBF5ΔPUA protein missing the C-terminal PUA domain (Figure 5A) and studied sRNP assembly with this protein. As shown in Figure 5, upon incubation of the

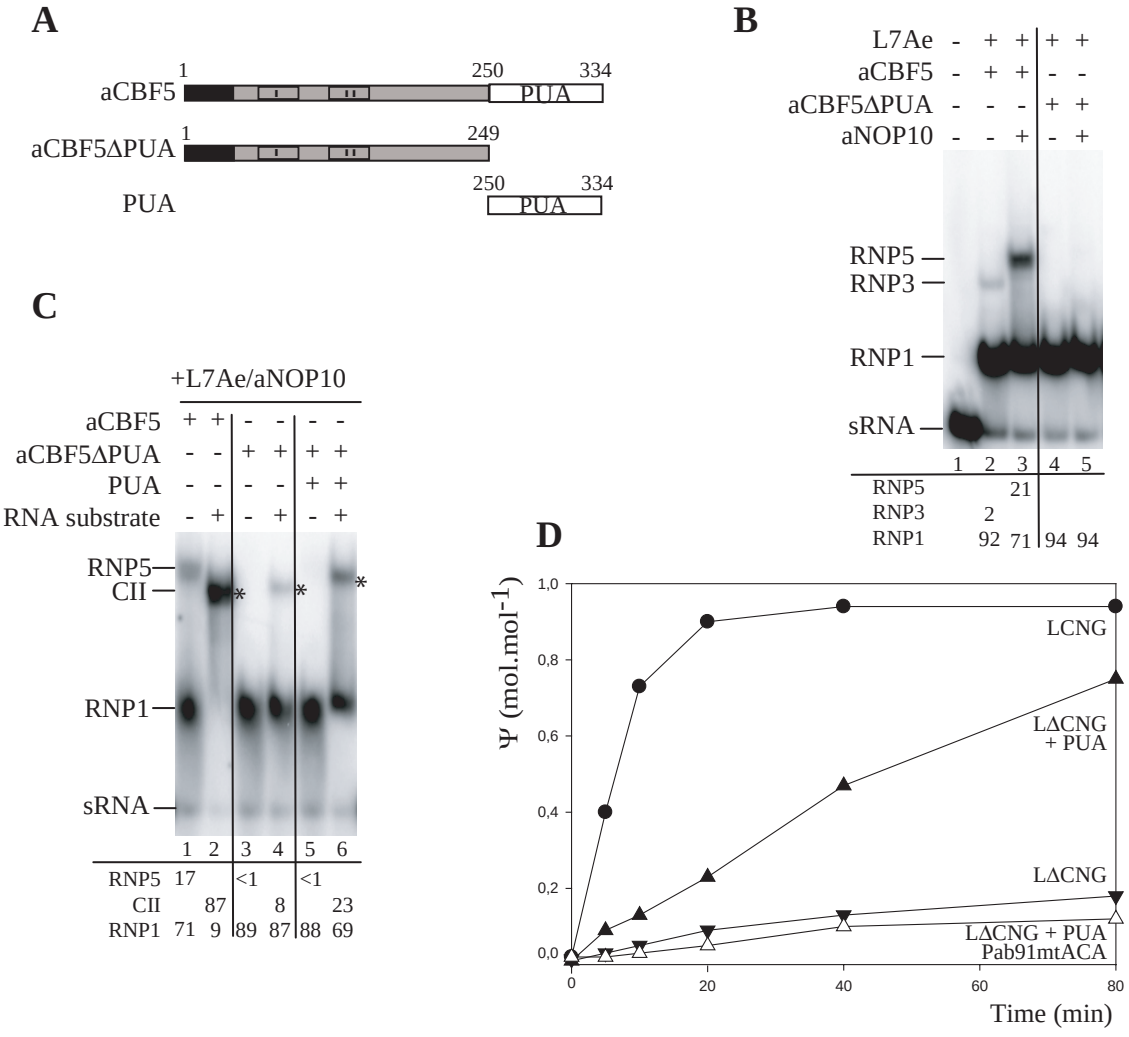


Figure 5. PUA *trans*-complementation of the aCBF5 activity. (**A**) Schematic representation of the aCBF5 protein with the N-terminal extension found in the RNA-guided Ψ-synthases (black box), the highly conserved motifs I and II in the catalytic domain, and the pseudouridine synthase archeosine transglycosylase domain, PUA (white) are represented. The two aCBF5 fragments used in this study (aCBF5ΔPUA and PUA) are shown below. (**B**) Analysis by EMSA of the complexes formed with the radio-labeled Pab91 sRNA (50 fmol), proteins L7Ae and aNOP10 and either the full-length aCBF5 or the aCBF5ΔPUA fragment (200 nM each). The RNP complexes identified by phosphorimager analysis are numbered according to (13). The percentages of radioactivity in each complexes as compared to the total amounts of radioactivity in the lanes are given below the lanes. (**C**) Analysis by EMSA of complex CII formed with the wild-type or variant aCBF5 proteins. Complex CII formation was tested by incubation of the Pab91 sRNA with proteins L7Ae and aNOP10 (200 nM) and either the full-length aCBF5 protein (lanes 1 and 2) or the truncated aCBF5ΔPUA protein (200 nM) (lanes 3–6) in the presence (lanes 5 and 6) or absence (lanes 3 and 4) of the PUA fragment (200 nM). A 50 molar excess of unlabeled RNA substrate was added to the mixture (as indicated above each lane). Complexes were fractionated by gel electrophoresis. The nomenclature is according to (13) and the CII complexes are marked by an asterisk. The percentages of Pab91 sRNA present in each RNP as referred to the total RNA amount in the lane are given below the lanes. (**D**) Time course analyses of *in vitro* pseudouridylation reactions. The unlabeled WT Pab91 (filled symbols) or Pab91mtACA (open symbols) sRNA (4 pmol) was incubated at 65°C with the labeled RNA substrate (150 fmol) and various combinations of the core proteins : L7Ae (L), aCBF5 (C) or aCBF5ΔPUA (ΔC) with or without the PUA fragment.

radio-labeled sRNA Pab91 with the aCBF5ΔPUA and L7Ae proteins, only the L7Ae-containing complex (RNP1) was formed (Figure 5B, lane 4). The same result was obtained in the presence of aNOP10 (Figure 5B, lane 5).

In addition, when the RNA substrate was added to the incubation mixture containing the guide sRNA and the L7Ae, aCBF5ΔPUA and aNOP10 protein set, the CII complex was formed at very low yield (Figure 5C, lane 4). Accordingly, time course analysis of the pseudouridylation reaction revealed a very low efficiency of particles reconstituted with the truncated aCBF5 protein. Only 10% of the RNA substrate molecules were modified after a 80 min incubation (Figure 5D). We, thus, concluded that the PUA domain has a crucial role for aCBF5 binding to the Pab91 sRNA and as a result for proper H/ACA sRNP assembly. Interestingly, when the PUA domain was added in *trans* in the reaction mixture, a higher amount of CII complex was formed (Figure 5C, compare lane 6 with lane 4) and the rate of the reaction was enhanced. After a 80 min incubation, about 70% of the RNA substrate molecules were modified (Figure 5D). When the same assay was repeated with the variant Pab91mtACA, containing an ACA to UGU substitution (13), the addition of the PUA domain had no positive effect on the sRNP activity (Figure 5D). We concluded that

the PUA domain of protein aCBF5 likely interacts with the conserved ACA sequence in the H/ACA sRNA.

The aNOP10 Tyr41 and Leu48 residues are necessary for aNOP10 dimerization

As described above, the aCBF5–aNOP10 heterodimers were found in two kinds of heterotetramer structures in the crystal (Figure 1D and E). Formation of these heterotetramers relies on aNOP10 dimerization. In one heterotetramer, the aNOP10 dimerization interface is formed of the two C-terminal α-helices, which are parallel to each other (Figure 6A), and the interface buries a quite large solvent-accessible surface area of 1507 $Å^2$. This interface involves the hydrophobic residues Tyr41, Leu48 and Leu52, which are located on one face of the α-helix (Figure 6C). In addition, two phosphate ions are bound at the N-terminal end of the α-helices and form hydrogen bonds with the Ser36 residues of both aNOP10 molecules. In the second heterotetramer, the aNOP10 dimerization interface buries a much lower solvent-accessible surface area (798 $Å^2$). The two α-helices are perpendicular to each other and only the N-terminal part of these helices are involved in the interface (Figure 6B). The Tyr41 and Tyr44 residues from both α-helices form a small aromatic cluster (Figure 6D) and no phosphate ion is bound unlike in the other dimer.

Analytical gel filtration analysis (Figure 6E and F) showed that the pure aNOP10 protein migrates as two peaks corresponding to a monomer–dimer equilibrium with a high proportion of the monomer form. These results suggested that in the absence of protein aCBF5, protein aNOP10 forms homodimers in solution. When the aNOP10 and aCBF5 proteins were mixed at a 1.2:1 ratio, two main peaks were obtained, in addition to the peak corresponding to the residual free aNOP10 protein (Figure 6E). They corresponded to aCBF5–aNOP10 heterodimers and heterotetramer, e.g. an heterodimer of two aCBF5–aNOP10 heterodimers. Therefore, in accordance with the crystal structure, the aNOP10-aCBF5 heterodimers can form heterotetramers in solution, likely via the aNOP10–aNOP10 interaction.

In order to get an idea of the kind of aNOP10 homodimers formed in solution, we produced two aNOP10 protein variants. In variant aNOP10Y41A, residue Tyr41 involved in both types of aNOP10 dimers was mutated, and in variant aNOP10L48A we mutated residue Leu48, which is expected to be only involved in the formation of homodimers with two parallel helices. As shown by gel filtration assays, both amino acids substitutions abolished homodimer formation, which is in favor of the formation of homodimers with parallel helices in solution.

DISCUSSION

As stated in the introduction, the aCBF5 RNA:Ψ-synthase of archaeal H/ACA sRNPs is inactive by itself. It needs association with the aNOP10 protein and the H/ACA sRNA to become active. The present study brings important information for understanding this peculiar property of the archaeal aCBF5 enzyme. This is as much important, as the observed peculiarities of aCBF5 compared to other RNA:Ψ synthases, are likely shared with its eukaryotic counterparts (Cbf5p/Dyskerin).

Similarities and differences between the aCBF5 and TruB 3D structures

As the TruB enzyme, protein aCBF5 consists of a large catalytic domain and a small C-terminal PUA domain. However, one important difference between aCBF5 and TruB is the presence of a 33 residues additional N-terminal sequence that wraps the PUA domain. This additional sequence is found in all the characterized archaeal aCBF5 enzymes and is also present in their eukaryal counterparts. To test for its function we tried to produce an aCBF5 protein truncated of this N-terminal additional sequence. This was not possible because the truncated protein turned to be unstable when produced in *E.coli*. Hence, we concluded that the 33 N-terminal residues of the *P.abyssi* aCBF5 protein are important for the correct folding and/or the stability of the enzyme. Wrapping of the PUA domain by the N-terminus may also be important for its precise positioning relative to the catalytic domain. This may be as much important as our data demonstrate a role of the PUA domain in aCBF5 binding to the sRNA. Previous (12,13) and present results strongly support the idea that aCBF5 association with the guide sRNA depends upon the ACA conserved motif of H/ACA sRNAs. Interestingly, the PUA domain was originally found in the archeosine tRNA-guanine transglycosylase (ArcTGT) (31), that modifies G15 of the D-arm of tRNA (37,38). The 3D structure analysis of a tRNA bound ArcTGT revealed the interactions of the PUA domain with the acceptor stem and the CCA terminus of the tRNA (39). In addition, in TruB, the PUA domain is also expected to interact with the tRNA acceptor arm (40). Although the observed interactions essentially concern the backbone of the RNA, PUA domains seem to develop preferential interactions with NCA trinucleotide sequences (21). The ACA trinucleotide in H/ACA sRNAs is located at a constant distance of the targeted uridine in the sRNA–target RNA complex (14 to 16 bp flanked by the ACA motif) [(41) and S. Muller, F. Leclerc and C. Branlant, unpublished data]. This constant distance may represent a key determinant for a good positioning of the targeted uridine in the active site.

Concerning RNA binding, it is interesting to note that because the aCBF5–aNOP10 interaction masks the negative charges present on the surface of protein aCBF5, the surface formed by association of proteins aCBF5 and aNOP10 is mostly positively charged (Figure 3E). This surface electrostatic potential is highly favorable for a large interaction with the guide RNA and the RNA substrate. Hence, masking negative charges on the aCBF5 surface may be one of the multiple roles played by protein aNOP10. It might also be that the aNOP10 interaction in the aCBF5–aNOP10 complex is responsible for the difference of conformation of residues Tyr179 and Asp82 in the active site, as compared to those of their counterparts in the RNA-free *ec*TruB enzyme (20). Based on the observed conformation of these two residues in the aCBF5–aNOP10 complex, a more limited induced fit should be needed for substrate binding, as compared to the *ec*TRuB enzyme. This may explain our previous observation of a high catalytic efficiency of the Pab91 reconstituted sRNP when the small RNA-S substrate was used (100% of RNA modification after a 5 min incubation of the RNA–protein mixture) (13).

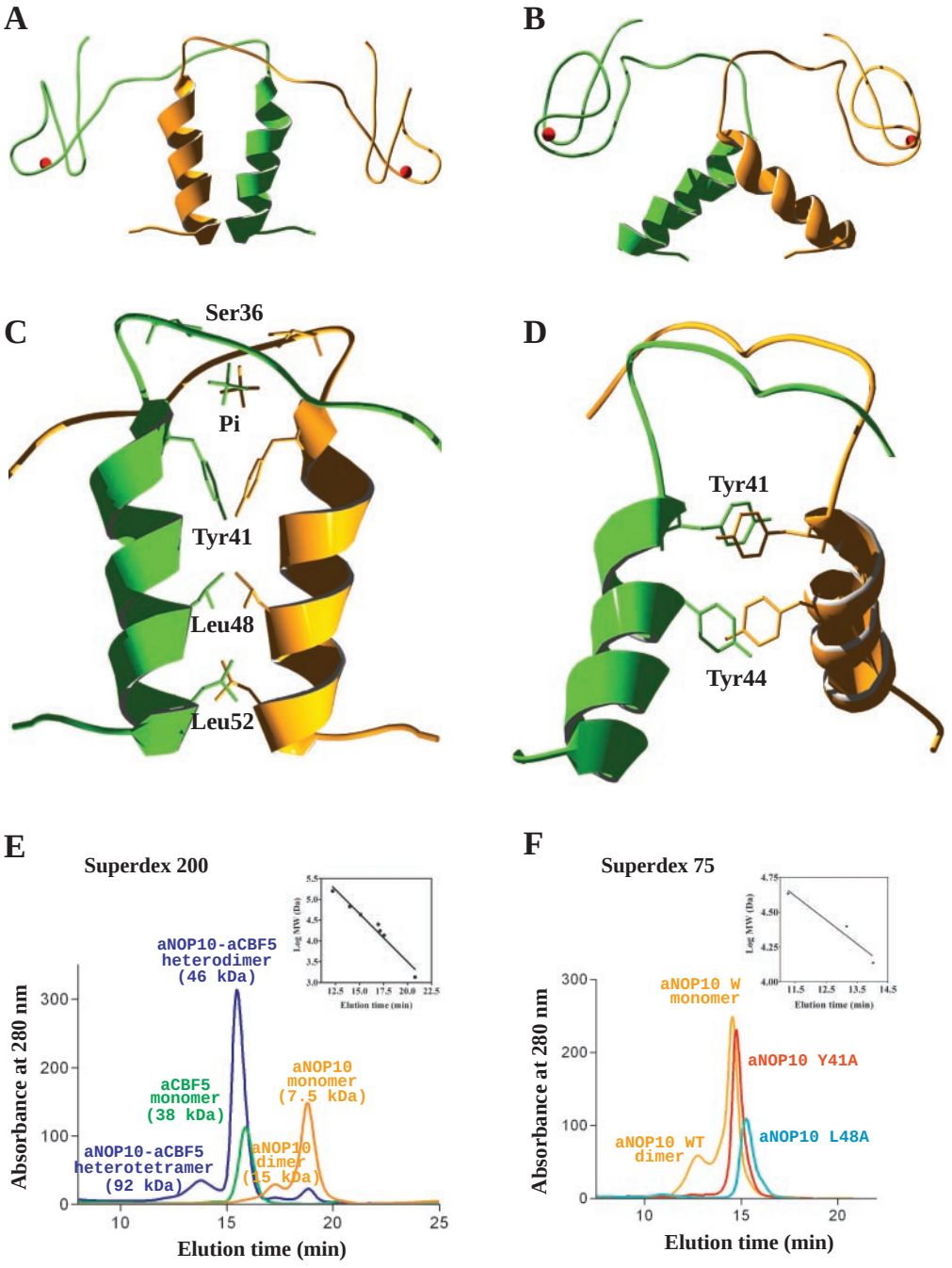


Figure 6. Analysis of protein aNOP10 dimerization by gel filtration (**A**, **B**, **C** and **D**) Ribbon representation of the two distinct aNOP10 dimers (A) and (B) found in the crystal structure and their corresponding close-up views (C) and (D). (**E**) Dimer formation analysis by gel filtration. The purified aNOP10 protein and aCBF5 protein alone, or a mixture of these two proteins in a 1.2:1 molar ratio were incubated as described in Materials and Methods and were analyzed by gel filtration on a Superdex 200. The elution profiles are represented in orange, green and blue, respectively. The inset represents the elution times of standard molecular markers as a function of the log of their molecular weights (Da): bovin gamma globulin (158 000 Da), BSA (67 000 Da), chicken ovalbumin (43 000 Da), chymotrypsinogen A (25 000), equine myoglobin (17 500 Da), Ribonuclease A (13 700) and Vitamin B12 (1350 Da). (**F**) Effects of the Y41A and L48A amino acids substitutions in aNOP10 on the dimerization properties. Fractionation was performed on a Superdex 75 column. The elution profiles of the WT, L48A and Y41A variant proteins are drawn in red, green and blue, respectively. The inset represents elution times of standard molecular markers as a function of the log of their molecular weights (Da): chicken ovalbumin (43 000 Da), chymotrypsinogen A (25 000) and Ribonuclease A (13 700).

The numerous interactions between aCBF5 and aNOP10 play differential roles in aCBF5 activation

Our crystal structure reveals a very elongate shape of protein aNOP10. This very elongated shape allows the interaction of protein aNOP10 with the two lobs of the aCBF5 catalytic domain. Such an elongated shape was already found for several ribosomal proteins (42). In spite that both the N- and C-terminal domains of protein aNOP10 can form

ionic and hydrophobic interactions with aCBF5 (Figure 3), the C-terminal part of protein aNOP10 does not form a stable complex with protein aCBF5 bound to the sRNA (Figure 4B). For integration in the RNP complex the aNOP10 C-terminal domain needs the presence of the RNA substrate and particles reconstituted with this fragment are active (60% of the WT activity with a lag period in the kinetics) (Figure 4C). In contrast, the aNOP10 N-terminal domain that contains the Zn^{2+} binding domain has the capability to bind protein aCBF5, however, active particles are not generated with this domain (Figure 4B and C). In addition, mutation of one of the four cysteines (C11) of the Zn^{2+} binding site had no effect on the RNA:Ψ-synthase activity. This is in contrast to finding of a requirement of Zn^{2+} ions for activity of the *Saccharomyces cerevisiae* Pus1p RNA:Ψ-synthase (43). The present data show that the aNOP10 Zn^{2+} ion is not essential for sRNP activity. However, the aNOP10 N-terminal domain may reinforce the sRNP activity by: (i) a role in anchoring the aNOP10 protein on to the aCBF5 protein, (ii) modifying the conformation of the aCBF5 active site and/or (iii) remodeling the structure of the guide RNA, as very recent proposed for the eukaryal Nop10 protein (44). Indeed, after submission of this article, an NMR spectroscopy study of the interaction between the yeast Nop10p protein and one stem–loop structure of a human H/ACA snoRNA was published by the J. Feigon's laboratory (44). The authors propose that the folded N-terminal region of Nop10p interacts weakly with the snoRNA, in the region at the junction between the lower stem and the pseudouridylation pocket. As the pseudouridylation pocket is partially structured in absence of proteins (44), such an interaction may participate in the opening of the pseudouridylation pocket to favor duplex formation between the guide and the substrate RNAs. In light of this hypothesis, it will be interesting to test whether the lower sRNP activity, that we observed when using the aNOP10 C-terminal domain as compared to the full-length aNOP10 protein, is in part due to the absence of remodeling of the guide RNA structure by the aNOP10 N-terminal domain. Indeed, in the absence of the aNOP10 N-terminal domain, only 60% of the RNA substrate molecules were modified and a lag period was observed in the kinetics of the reaction (Figure 4C).

As the RNA-duplex formed by the guide RNA and the RNA substrate is able to generate a stable interaction between protein aCBF5 and the aNOP10 C-terminal domain, leading to a functional particle, we conclude that this RNA duplex is able to supply, at least partially, for the absence of the anchoring aNOP10 N-terminal domain. Surprisingly, no sRNP activity was observed when the C- and N-terminal domains of protein aNOP10 were present together as separate entities. Several explanations can be proposed for this observation. First, an aberrant structure may be formed by interaction of the two halves of aNOP10. Secondly, the aNOP10 N-terminal domain may interact with both lobs of the aCBF5 catalytic domain, thus, avoiding the binding of the aNOP10 C-terminal domain. Finally, and more simply there may be a negative effect of the charged carboxylate at the extremity of the aNOP10 N-terminal domain. It may interfere with the aCBF5 active site, for instance by formation of an ion pair with the catalytic Asp82 residue.

Is the aNOP10 dimerization of functional importance?

Surprisingly, our crystals contained two kinds of aNOP10 dimers (Figure 1D and E; Figure 6). This suggests that the two kinds of dimers may exist in equilibrium in solution. However, according to our site-directed mutagenesis analysis only dimers with parallel α-helices are stable in solution. Interestingly, in the crystal, these dimers with parallel α-helices contained two phosphate ions hydrogen-bonded with residue S36. This opens the question of a possible involvement of dimer formation in RNA binding. As no phosphate ion was found in the dimers with perpendicular helices, the involvement of a conformational transition from one dimer conformation to the other one may participate to the dissociation of the modified substrate RNA.

During the time that this paper was evaluated, the 3D structures of the aCBF5–aNOP10 complex from *Methanococcus jannaschii*, and of the Nop10 proteins from *S.cerevisiae* and *M.jannaschii* have been solved (44,45). Our data on the *P.abyssi* aCBF5–aNOP10 complex are in perfect agreement with those obtained for the *M.jannaschii* complex. Moreover, our data reinforce the idea that the C-terminal α-helix of aNOP10 is stabilized upon interaction with aCBF5 (45). By inspection of the data obtained by Hamma *et al*. (45) for the *M.jannaschii* aCBF5–aNOP10 complex (PDB ID 2apo), we observed that the asymmetric unit contains only one heterodimer that is related to an identical heterodimer via a crystallographic 2-fold axis. Interestingly, the quaternary structure for the heterotetramer found in their crystal corresponds to the *P.abyssi* heterotetramer in which the two aNOP10 α-helices are perpendicular to each other. In the *M.jannaschii* quaternary structure, the aromatic cluster located at the interface of the aNOP10 helices is formed by the Trp42 and Tyr45 residues, corresponding to the Tyr41 and Tyr44 residues found in the *P.abyssi* aNOP10 cluster (Figure 1C). Thus, both the *P.abyssi* and the *P.jannaschii* aCBF5–aNOP10 complexes are found as heterotetramers in crystal structures.

In connection with a possible role of these aCBF5–aNOP10 tetramers, it should be noticed, that in eukarya most of the guide H/ACA snoRNAs contain two stem–loop structures with a pseudouridylation pocket. As they both have to be associated with a CBF5–NOP10 complex, one may imagine a synergistic binding of the two complexes by heterotetramer formation. Although the presence of two stem–loop structures is not a general rule in archaeal H/ACA sRNAs, some of these sRNAs also have two stem–loop structures [(41) and S. Muller, F. Leclerc and C. Branlant, unpublished data].

Conservation of aCBF5 and aNOP10 in eukarya

The present data are important for the understanding of the structure of eukaryal H/ACA snoRNPs and their mechanism of action. Indeed, the archaeal and eukaryal CBF5 and NOP10 proteins are highly conserved in the two kingdoms of life and, up to now, no reconstitution of active the H/ACA snoRNP was achieved using purified recombinant proteins and *in vitro* transcribed RNAs (46). This is essentially due to the difficulty to produce the eukaryal proteins CBF5 and GAR1 of H/ACA snoRNP complexes in a soluble form. Such difficulty also impairs their crystallization and 3D structure determination. However, based on the present data, one may model the 3D structure of the eukaryal CBF5–NOP10

complex. Interestingly, a linked has been established between a genetic disease, designated as X-linked dyskeratosis congenita (DC), and mutations found in the Dyskerin gene (47). The most frequent mutations (48) are located in the PUA domain and the N-terminal extension that wraps the PUA domain. In agreement with the present data, this strengthens the importance of this domain in H/ACA snoRNP assembly. In vertebrates, in addition to snoRNAs and scaRNAs (49), the telomeric RNA also contains an H/ACA domain that binds the complete set of NHP2, CBF5/Dyskerin, NOP0 and GAR1 proteins from H/ACA snoRNPs (50,51). Patients suffering from DC were found to have a decreased level of telomerase (52). We are presently generating the most prevalent mutations found in the Dyskerin encoding gene into the aCBF5 gene in order to test for their effects on aCBF5 binding to the sRNA, sRNP assembly and sRNP activity.

ACKNOWLEDGEMENTS

Christine Loegler is warmly thanked for technical assistance. The authors are grateful to ESRF for access to the BM14 beamline and the authors thank Hassan Belrhali for assistance with data collection. J.-B.F. is a fellow from the French Ministère délégué à l'Enseignement supérieur et à la Recherche. The work was supported by the Centre National de la Recherche Scentifique (CNRS), the French Ministère de la Recherche et des Nouvelles Technologies (MRNT), the Actions Concertées Incitatives (ACI) Microbiologie and BCMS of the french MRNT and the Bioingénierie PRST of the Conseil Régional Lorrain. Funding to pay the Open Access publication charges for this article was provided by the Service Commun de Documentation de I'UHP-Nancy I.

Conflict of interest statement. None declared.

REFERENCES

1. Cavaille,J. and Bachellerie,J.P. (1998) SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic Acids Res*., **26**, 1576–1587.
2. Massenet,S., Mougin,A. and Branlant,C. (1998) In Grosjean,H. and Benne,R. (eds), *The Modification and Editing of RNA*. ASM Press, Washington, DC, pp. 201–228.
3. Yang,C., McPheeters,D.S. and Yu,Y.T. (2005) Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in *Saccharomyces cerevisiae*. *J. Biol. Chem*., **280**, 6655–6662.
4. King,T.H., Liu,B., McCully,R.R. and Fournier,M.J. (2003) Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol. Cell*, **11**, 425–435.
5. Lapeyre,B. and Purushothaman,S.K. (2004) Spb1p-directed formation of Gm2922 in the ribosome catalytic center occurs at a late processing stage. *Mol. Cell*, **16**, 663–669.
6. Donmez,G., Hartmuth,K. and Luhrmann,R. (2004) Modified nucleotides at the 5′ end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA*, **10**, 1925–1933.
7. Motorin,Y., Keith,G., Simon,C., Foiret,D., Simos,G., Hurt,E. and Grosjean,H. (1998) The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA*, **4**, 856–869.
8. Nurse,K., Wrzesinski,J., Bakin,A., Lane,B.G. and Ofengand,J. (1995) Purification, cloning, and properties of the tRNA psi 55 synthase from *Escherichia coli*. *RNA*, **1**, 102–112.
9. Wrzesinski,J., Nurse,K., Bakin,A., Lane,B.G. and Ofengand,J. (1995) A dual-specificity pseudouridine synthase: an *Escherichia coli* synthase purified and cloned on the basis of its specificity for psi 746 in 23S RNA is also specific for psi 32 in tRNA(phe). *RNA*, **1**, 437–448.
10. Omer,A.D., Ziesche,S., Decatur,W.A., Fournier,M.J. and Dennis,P.P. (2003) RNA-modifying machines in archaea. *Mol. Microbiol*., **48**, 617–629.
11. Bachellerie,J.P., Cavaille,J. and Huttenhofer,A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
12. Baker,D.L., Youssef,O.A., Chastkofsky,M.I., Dy,D.A., Terns,R.M. and Terns,M.P. (2005) RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev*., **19**, 1238–1248.
13. Charpentier,B., Muller,S. and Branlant,C. (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res*., **33**, 3133–3144.
14. Henras,A., Henry,Y., Bousquet-Antonelli,C., Noaillac-Depeyre,J., Gelugne,J.P. and Caizergues-Ferrer,M. (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J*., **17**, 7078–7090.
15. Watkins,N.J., Gottschalk,A., Neubauer,G., Kastner,B., Fabrizio,P., Mann,M. and Luhrmann,R. (1998) Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA*, **4**, 1549–1568.
16. Lafontaine,D.L., Bousquet-Antonelli,C., Henry,Y., Caizergues-Ferrer,M. and Tollervey,D. (1998) The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev*., **12**, 527–537.
17. Ganot,P., Caizergues-Ferrer,M. and Kiss,T. (1997) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev*., **11**, 941–956.
18. Ni,J., Tien,A.L. and Fournier,M.J. (1997) Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell*, **89**, 565–573.
19. Watanabe,Y. and Gray,M.W. (2000) Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic Acids Res*., **28**, 2342–2352.
20. Pan,H., Agarwalla,S., Moustakas,D.T., Finer-Moore,J. and Stroud,R.M. (2003) Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proc. Natl Acad. Sci. USA*, **100**, 12648–12653.
21. Hoang,C. and Ferre-D'Amare,A.R. (2001) Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell*, **107**, 929–939.
22. Charron,C., Manival,X., Charpentier,B., Branlant,C. and Aubry,A. (2004) Purification, crystallization and preliminary X-ray diffraction data of L7Ae sRNP core protein from *Pyrococcus abyssi*. *Acta. Crystallogr. D*, **60**, 122–124.
23. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*., **276**, 307–326.
24. Terwilliger,T.C. and Berendzen,J. (1999) Automated MAD and MIR structure solution. *Acta. Crystallogr. D Biol. Crystallogr*, **55**, 849–861.
25. Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta. Crystallogr. A*, **47**, 110–119.
26. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al*. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta. Crystallogr. D*, **54**, 905–921.
27. Kabsch,W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr*., **26**, 795–800.
28. Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr*., **26**, 283–291.
29. Hutchinson,E.G. and Thornton,J.M. (1994) A revised set of potentials for beta-turn formation in proteins. *Protein Sci*., **3**, 2207–2216.
30. Gunasekaran,K., Ramakrishnan,C. and Balaram,P. (1996) Disallowed Ramachandran conformations of amino acid residues in protein structures. *J. Mol. Biol*., **264**, 191–198.

31. Aravind,L. and Koonin,E.V. (1999) Novel predicted RNA-binding domains associated with the translation machinery. *J. Mol. Evol.*, **48**, 291–302.
32. Huang,L., Pookanjanatavip,M., Gu,X. and Santi,D.V. (1998) A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry*, **37**, 344–351.
33. Gu,X., Liu,Y. and Santi,D.V. (1999) The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA. *Proc. Natl Acad. Sci. USA*, **96**, 14270–14275.
34. Ramamurthy,V., Swann,S.L., Paulson,J.L., Spedaliere,C.J. and Mueller,E.G. (1999) Critical aspartic acid residues in pseudouridine synthases. *J. Biol. Chem.*, **274**, 22225–22230.
35. Zebarjadian,Y., King,T., Fournier,M.J., Clarke,L. and Carbon,J. (1999) Point mutations in yeast CBF5 can abolish *in vivo* pseudouridylation of rRNA. *Mol. Cell. Biol.*, **19**, 7461–7472.
36. Phannachet,K., Elias,Y. and Huang,R.H. (2005) Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase. *Biochemistry*, **44**, 15488–15494.
37. Watanabe,M., Nameki,N., Matsuo-Takasaki,M., Nishimura,S. and Okada,N. (2001) tRNA recognition of tRNA-guanine transglycosylase from a hyperthermophilic archaeon, *Pyrococcus horikoshii*. *J. Biol. Chem.*, **276**, 2387–2394.
38. Ishitani,R., Nureki,O., Fukai,S., Kijimoto,T., Nameki,N., Watanabe,M., Kondo,H., Sekine,M., Okada,N., Nishimura,S. *et al.* (2002) Crystal structure of archaeosine tRNA-guanine transglycosylase. *J. Mol. Biol.*, **318**, 665–677.
39. Ishitani,R., Nureki,O., Nameki,N., Okada,N., Nishimura,S. and Yokoyama,S. (2003) Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell*, **113**, 383–394.
40. Ferré-D' Amaré,A.R. (2003) RNA-modifying enzymes. *Curr. Opin. Struct. Biol.*, **13**, 49–55.
41. Tang,T.H., Bachellerie,J.P., Rozhdestvensky,T., Bortolin,M.L., Huber,H., Drungowski,M., Elge,T., Brosius,J. and Huttenhofer,A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA*, **99**, 7536–7541.
42. Ramakrishnan,V. and Moore,P.B. (2001) Atomic structures at last: the ribosome in 2000. *Curr. Opin. Struct. Biol.*, **11**, 144–154.
43. Arluison,V., Hountondji,C., Robert,B. and Grosjean,H. (1998) Transfer RNA-pseudouridine synthetase Pus1 of *Saccharomyces cerevisiae* contains one atom of zinc essential for its native conformation and tRNA recognition. *Biochemistry*, **37**, 7268–7276.
44. Khanna,M., Wu,H., Johansson,C., Caizergues-Ferrer,M. and Feigon,J. (2006) Structural study of the H/ACA snoRNP components Nop10p and the 3′ hairpin of U65 snoRNA. *RNA*, **12**, 40–52.
45. Hamma,T., Reichow,S.L., Varani,G. and Ferré-D'Amaré,A.R. (2005) The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nature Struct. Biol.*, **12**, 1101–1107.
46. Wang,C. and Meier,U.T. (2004) Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.*, **23**, 1857–1867.
47. Heiss,N.S., Knight,S.W., Vulliamy,T.J., Klauck,S.M., Wiemann,S., Mason,P.J., Poustka,A. and Dokal,I. (1998) X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nature Genet.*, **19**, 32–38.
48. Meier,U.T. (2005) The many facets of H/ACA ribonucleoproteins. *Chromosoma*, **114**, 1–14.
49. Darzacq,X., Jady,B.E., Verheggen,C., Kiss,A.M., Bertrand,E. and Kiss,T. (2002) Cajal body-specific small nuclear RNAs: a novel class of 2′-O- methylation and pseudouridylation guide RNAs. *EMBO J.*, **21**, 2746–2756.
50. Dez,C., Henras,A., Faucon,B., Lafontaine,D., Caizergues-Ferrer,M. and Henry,Y. (2001) Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p. *Nucleic Acids Res.*, **29**, 598–603.
51. Pogacic,V., Dragon,F. and Filipowicz,W. (2000) Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. *Mol. Cell. Biol.*, **20**, 9028–9040.
52. Mochizuki,Y., He,J., Kulkarni,S., Bessler,M. and Mason,P.J. (2004) Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. *Proc. Natl Acad. Sci. USA*, **101**, 10756–10761.

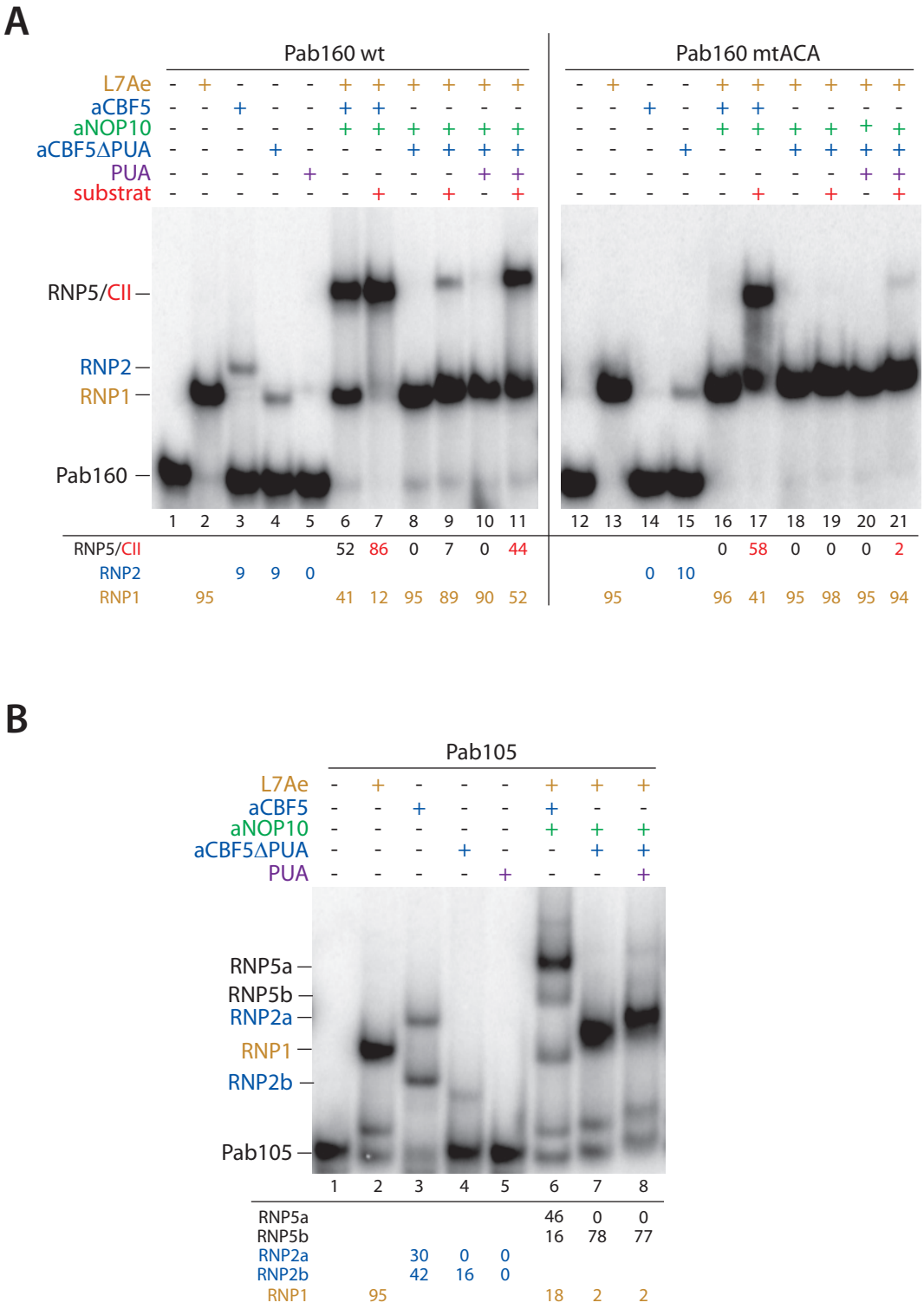


Figure r14 : Etude par la technique de retard de migration sur gel de l'assemblage du domaine PUA sur les sRNA Pab160 et Pab105.
Résultats d'expériences de retard de migration sur gel réalisées en présence des sRNA Pab160 (**A.**) ou Pab105 (**B.**) radiomarqués en 5' sauvage (wt) ou muté dans leur boîte ACA (mtACA) et de différentes conditions de protéines ou de substrats (indiqués par + ou -).

2. Résultats complémentaires

2.1. L'effet positif du domaine PUA de aCBF5 sur l'activité de sRNP guidées par les sRNA Pab160 et Pab105

Nous avons réalisé des tests avec la protéine aCBF5ΔPUA (tronquée du domaine PUA) et le domaine PUA produit séparément sur l'activité guidée par les sRNA Pab160 (simple motif) et Pab105 (double motif) (Fig. r14). Certaines données obtenues avec ces sRNA sont équivalentes à celles déjà publiées sur le sRNA Pab91 (article n°2, (Manival *et al.*, 2006)). Cependant des données singulières ont été obtenues avec ces sRNA. Nous constatons que la protéine aCBF5ΔPUA se fixe directement au sRNA Pab160, avec un taux de complexe formé identique à celui obtenu avec aCBF5 entière (Fig. r14 A pistes 3 et 4). De plus, ce complexe subsiste lorsque le sRNA est muté dans sa boîte ACA, alors que le complexe formé entre le sRNA et aCBF5 entière n'est plus observé (Fig. r14 A pistes 3 et 4). La protéine aCBF5ΔPUA se fixe également sur le sRNA Pab105 mais en ne formant qu'un seul complexe retardé, au lieu de 2 pour aCBF5 entière, et son taux est plus faible que celui produit par aCBF5 entière (Fig. r14 B pistes 3 et 4). Nous concluons que la protéine aCBF5 tronquée de son domaine PUA possède toujours la capacité de se lier à certains sRNA à boîtes H/ACA. En corolaire, nous concluons que les sRNA ont des comportements différents vis-à-vis de la fixation de la protéine aCBF5 ou d'un fragment de cette protéine. Là encore, la séquence de la poche de pseudouridylation semble être le déterminant majeur, sur lequel repose la stabilité des interactions entre l'ARN et le domaine catalytique de aCBF5.

Nous avons aussi testé par retard de migration sur gel la fixation directe du domaine PUA sur les sRNA Pab160 et Pab105 (Fig. r14 A et B piste 5), sur le sRNA Pab21 (données non montrées), et Pab91 (données non montrées). Aucun des sRNA testés ne conduit à la formation d'un complexe stable en présence du seul domaine PUA. Ces données sont confirmées par les résultats obtenus par l'équipe de R. et M. Terns qui n'observe pas non plus de complexe stable en employant le domaine PUA produit isolément (Baker *et al.*, 2008). Bien que le domaine PUA soit un domaine de fixation à l'ARN, nous concluons que ce dernier sous une forme isolée ne peut

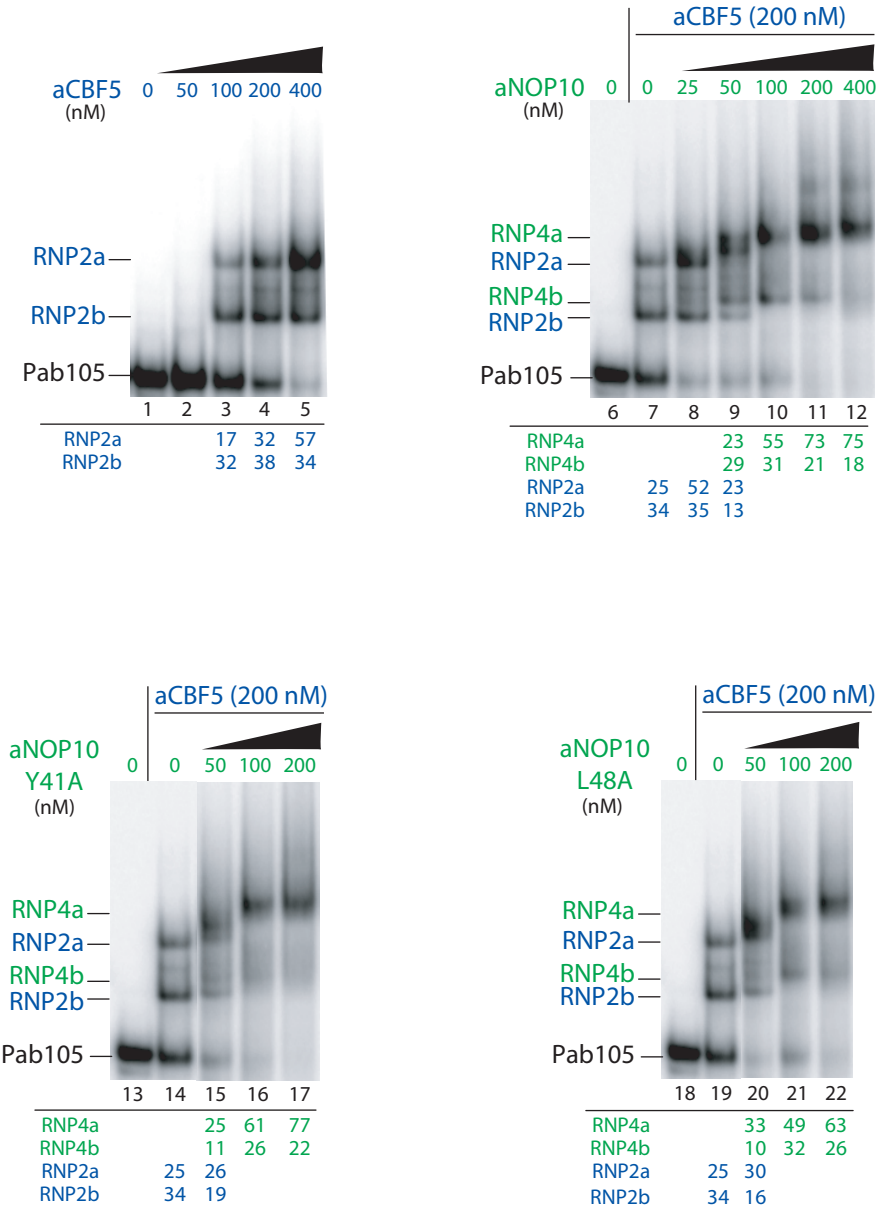


Figure r15 : Etudes par des expériences de retard de migration sur gel de la fixation de la protéine aNOP10 sauvage, ses variants Y41A ou L48A sur le sRNA Pab105 en présence d'une concentration fixe en aCBF5.

Résultats d'expériences de retard de migration sur gel réalisées en présence du sRNA Pab105 radiomarqué en 5' sauvage, d'une gamme de la concentration de la protéine aNOP10 sauvage, ses variants Y41A ou L48A et de la protéine aCBF5 fixée à la concentration de 200 mM. Les différents complexes formés sont quantifiés et la valeur est indiquée sous chacune des pistes.

former un complexe stable avec un sRNA H/ACA. L'assemblage de la sRNP Pab160 avec les protéines L7Ae, aCBF5 ou aCBF5ΔPUA, aNOP10, le domaine PUA et l'ARN substrat se déroule d'une manière similaire à celui observé pour la sRNP Pab91 (Fig. r14 A) (article n°2, (Manival *et al.*, 2006)). Brièvement, lorsque le domaine PUA est ajouté en *trans* à la sRNP formée par les protéines L7Ae, aCBF5ΔPUA et aNOP10, la complémentation par le domaine PUA permet de former un complexe CII stable (Fig. r14 A pistes 6 à 11). Cette complémentation n'est plus effective lorsque le sRNA est muté dans sa boîte ACA (Fig. r14 A pistes 16 à 21). L'assemblage sur Pab105 peut aussi être complémenté, et ce, même en absence de substrat. Cependant un seul complexe est détecté au lieu des deux obtenus en présence de aCBF5 entière (Fig. r14 B pistes 6 à 8).

Dans le but de montrer que l'élément de structure secondaire qui se forme entre l'extrémité Nt de aCBF5 et son domaine PUA était important pour la complémentation du domaine PUA, nous avons essayé de produire une protéine aCBF5 tronquée de son domaine PUA et de l'extension N-terminale qui enveloppe le domaine PUA, mais cette protéine n'était pas soluble et donc aucun test n'a pu être effectué.

2.2. Importance de la dimérisation du complexe aCBF5-aNOP10 pour sa fixation sur le sRNA Pab105 à double motif tige-boucle

Nous avons montré par la technique de gel filtration que le complexe aCBF5-aNOP10 formait un dimère en solution et que cette dimérisation dépendait de la présence des résidus Y41 et L48 de aNOP10 (article n°2, (Manival *et al.*, 2006)). Nous avons étudié l'importance de cette capacité de dimérisation de l'hétérodimère aCBF5-aNOP10 sur la formation de RNP avec le sRNA Pab105 qui comporte deux motifs tige-boucle (Fig. r15). Par la technique de retard de migration sur gel, nous avons tout d'abord réalisé une gamme de la concentration de aCBF5 (Fig. r15 pistes 1 à 5). Comme je l'ai décrit précédemment, deux complexes retardés RNP2a et RNP2b sont formés (Fig. r3). Nous constatons qu'à 200 mM, il se forme autant de complexe

RNP2a (32%) que de complexe RNP2b (38%). C'est donc à cette concentration en protéine aCBF5 que nous avons effectué les expériences suivantes.

Nous avons ensuite réalisé une gamme de concentrations croissantes de la protéine aNOP10 (de 25 à 400 mM) tout en maintenant la concentration de la protéine aCBF5 à 200 mM. Après incubation de ce mélange avec le sRNA Pab105, nous observons l'apparition progressive des deux complexes RNP4a et 4b, indiquant que chacune des tige-boucle de Pab105 fixerait un jeu de protéines aCBF5-aNOP10 (Fig. r15 pistes 6 à 12). L'emploi des variants Y41A et L48A de aNOP10 conduit à la formation des deux complexes RNP4a et RNP4b à un taux similaire à celui observée pour aNOP10 wt (Fig. r15 pistes 13 à 22). Nous concluons que la perte de capacité à former un dimère $[aCBF5\text{-}aNOP10]_2$ n'empêche pas la fixation des hétérodimères aCBF5-aNOP10 sur chaque motif tige-boucle du sRNA Pab105. Ainsi, la formation des sRNP H/ACA à double motif résulterait de la fixation d'un complexe LCNG sur chacune des tige-boucles mais que ces complexes n'entretiendraient pas d'interactions entre eux.

B. Analyse comparative des deux activités ARN:Ψ-synthase de aCBF5 : l'activité guidée et l'activité non guidée par un ARN

1. Présentation du travail

La question était de savoir quelle enzyme parmi les 4 ARN:Ψ-synthases alors décrites chez les archaea, aPUS7, aPUS10, aTRUA et aCBF5 catalysait la modification en position 55 dans le bras TΨC des ARNt. Nous avons décidé de tester si aCBF5 pouvait être active sur les ARNt d'archaea, mais aussi ceux de la levure *S. cerevisiae*. Ainsi S. Muller au laboratoire, conjointement à une autre équipe (Roovers *et al.*, 2006), a observé que l'enzyme aCBF5 était capable de modifier la position 55 dans les ARNt sans la nécessité d'intervention d'un ARN guide. Nous avons constaté que la présence des protéines aNOP10 et aGAR1 renforçait cette activité non guidée, particulièrement lorsque la séquence terminale CCA dans l'ARNt est absente. Par ailleurs, S. Muller a montré que la présence du bras anticodon favorisait la fixation de aCBF5 sur l'ARNt, et que la modification d'un substrat correspondant au bras accepteur et au bras TΨC de l'ARNt était fortement renforcée par la présence des protéines aNOP10 et de aGAR1.

Les structures 3D établies pour les divers complexes contenant aCBF5 et aNOP10 ont suggéré l'existence possible de deux réseaux d'interactions entre chaînes latérales d'acides aminés présents dans le complexe aCBF5-aNOP10 (Hamma *et al.*, 2005; Rashid *et al.*, 2006). J'ai évalué l'importance fonctionnelle de chacun de ces réseaux pour les activités guidées et non guidées de aCBF5. J'ai produit les variants protéiques de aNOP10 et de aCBF5 contenant les substitutions des acides aminés qui ont été proposés participer à la formation de ces deux réseaux d'interaction, et nous avons évalué les effets des mutations sur les deux types d'activités. Selon les mutations réalisées, des effets différents sont observés sur les activités guidées et non guidées :

- Il avait été proposé que le résidu H43 de TruB (l'homologue de H77 de aCBF5) jouait un rôle essentiel pour l'extrusion du résidu cible U55 dans les ARNt (Hoang et Ferre-D'Amare, 2001). Nous avons montré que la substitution H77A de

aCBF5 n'affectait pas son activité guidée, tandis qu'elle supprimait son activité non guidée d'ARNt:Ψ55-synthase.

- De façon intéressante, nous avons également montré que le résidu H31 de aNOP10 était requis pour l'activité guidée, tandis que le résidu Y14 du motif en doigt de zinc de aNOP10 permettait de renforcer l'activité non guidée de aCBF5.

Ainsi, nos résultats indiquent que des acides aminés ont été conservés au cours de l'évolution pour préserver la dualité des activités guidée et non guidée du complexe aCBF5-aNOP10.

Ces résultats ont été publiés dans *Nucleic Acids Research* et je suis co-premier auteur.

Publication n°3 :

Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Ψ55-synthase and RNA-guided RNA:Ψ-synthase activities.

Sébastien MULLER*, **Jean-Baptiste FOURMANN***, Christine LOEGLER, Bruno CHARPENTIER and Christiane BRANLANT

* Contribution équivalente au travail

Nucleic Acids Research, 2007 volume 35(16) : 5610-5624.

5610–5624 *Nucleic Acids Research, 2007, Vol. 35, No. 16* *Published online 17 August 2007*
doi:10.1093/nar/gkm606

Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Ψ55-synthase and RNA-guided RNA:Ψ-synthase activities

Sébastien Muller, Jean-Baptiste Fourmann, Christine Loegler, Bruno Charpentier* and Christiane Branlant

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP, Nancy Université, Université Henri Poincaré-BP 239, 54506 Vandoeuvre-Lès-Nancy cedex, France

Received July 3, 2007; Revised and Accepted July 25, 2007

ABSTRACT

Protein aNOP10 has an essential scaffolding function in H/ACA sRNPs and its interaction with the pseudouridine(Ψ)-synthase aCBF5 is required for the RNA-guided RNA:Ψ-synthase activity. Recently, aCBF5 was shown to catalyze the isomerization of U55 in tRNAs without the help of a guide sRNA. Here we show that the stable anchoring of aCBF5 to tRNAs relies on its PUA domain and the tRNA CCA sequence. Nonetheless, interaction of aNOP10 with aCBF5 can counterbalance the absence of the PUA domain or the CCA sequence and more generally helps the aCBF5 tRNA:Ψ55-synthase activity. Whereas substitution of the aNOP10 residue Y14 by an alanine disturbs this activity, it only impairs mildly the RNA-guided activity. The opposite effect was observed for the aNOP10 variant H31A. Substitution K53A or R202A in aCBF5 impairs both the tRNA:Ψ55-synthase and the RNA-guided RNA:Ψ-synthase activities. Remarkably, the presence of aNOP10 compensates for the negative effect of these substitutions on the tRNA: Ψ55-synthase activity. Substitution of the aCBF5 conserved residue H77 that is expected to extrude the targeted U residue in tRNA strongly affects the efficiency of U55 modification but has no major effect on the RNA-guided activity. This negative effect can also be compensated by the presence of aNOP10.

INTRODUCTION

Isomerization of uridine into pseudouridine (5-ribosyluracil, Ψ) is a prevalent post-transcriptional modification of cellular RNAs (1) and this modified residue is frequently found in functionally important RNA regions. Consistently, an important role of Ψ residues was demonstrated for specific codon–anticodon recognition and ribosome function (2–8). Furthermore, one of the highly conserved Ψ residue in U2 snRNA was found to play a key role in pre-mRNA splicing (9,10).

Two types of catalysts can generate Ψ residues at specific sites in a pre-existing polyribonucleotide chains. One type consists in a single protein with an RNA:Ψ-synthase activity. It recognizes specifically one or more RNA segments or motifs and converts the targeted U residues within these segments into Ψ residues. A large panel of such RNA:Ψ-synthases are found in most cell types and they are up to now the unique system detected in bacteria (11). These bacterial RNA:Ψ-synthases were the first to be characterized (12–15). The search for their homologs by inspection of various genomes led to the classification of these enzymes into five families, TruA, TruB, TruD, RsuA, RluA (16–18). Based on sequence alignment and 3D structure analysis of the catalytic domains of several RNA:Ψ-synthases, five conserved structural motifs (I, II, IIa, III and IIIa) were identified and they represent the signature for RNA:Ψ-synthases activity (16,19,20). Although only a limited number of amino acids in these motifs are conserved, these motifs have highly conserved conformations in the 3D structures established for RNA:Ψ-synthases (11). The few conserved amino acids play important roles for catalysis: a universally conserved aspartate residue in motif II is the essential catalytic amino acid (17,21–24). A conserved basic residue (Arg or Lys) in motif III makes a salt bridge with the catalytic Asp residue. A conserved Tyr residue in motif IIa has a stacking interaction with the uracyl residue to be modified in the target RNA (25). Two other

*To whom correspondence should be addressed. Tel: 33 3 83 68 43 16; Fax: 33 3 83 68 43 07; Email: bruno.charpentier@maem.uhp-nancy.fr

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

© 2007 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

hydrophobic residues, a leucine in motif IIIa and an isoleucine or a valine residue in motif III, are also highly conserved, but their functional implication in the catalytic mechanism is less understood (19,25).

In the various RNA:Ψ-synthases identified up to now, the conserved catalytic domain is associated with a large variety of additional domains, which are likely involved in the specific recognition of the target RNA motif. One of these domains, the PUA domain (PseudoUridine and Archaeosine tRNA guanine transglycosylase), has largely been studied. It is found at the C-terminus of members of the TruB family of RNA:Ψ-synthases (25–30). As deduced from the crystal structure of the complex formed between the *Escherichia coli* TruB enzyme and the T stem-loop structure of tRNA (31), the PUA domain is expected to play a key role in RNA recognition by interacting with the CCA 3′-terminal sequence of tRNAs (25). This proposed interaction of TruB with a sequence common to all tRNAs fits with the general activity of TruB at position 55 of all elongator tRNAs.

The second type of catalyst for U to Ψ conversion consists in RNP complexes containing a small RNA with H/ACA motifs (the H/ACA snoRNAs and scaRNAs in Eukarya and H/ACA sRNAs in archaea) and a well-defined set of proteins (32–34). One of these proteins denoted aCBF5 in archaea, Cbf5p in yeast and CBF5/NAP57/Dyskerin in human displays the characteristic motifs I, II, IIa, III and IIIa found in RNA:Ψ-synthases and belongs to the TruB family of RNA:Ψ-synthase (16,19). In each RNP, the RNA component base pairs with one or more targeted RNA sequences and through this interaction acts as a guide to specify the uridyl residues that will be modified (32–34). Knowledge on the mechanism of the H/ACA RNPs activity is largely based on data obtained with the archaeal H/ACA RNPs that could be reconstituted using an *in vitro* transcribed RNA and recombinant proteins (35,36). In addition to the protein bearing the signature of RNA:Ψ-synthases, each H/ACA RNP contains three distinct proteins (NOP10, GAR1 and NHP2 in human; Nop10p, Gar1p and Nhp2p in yeast; and aNOP10, aGAR1 and L7Ae in archaea). The three archaeal aCBF5, aNOP10 and aGAR1 proteins form an heterotrimer in solution (35,36). The 3D structures of the aCBF5-aNOP10 heterodimer and aCBF5-aNOP10-aGAR1 heterotrimer complexes were solved by X-ray crystallography (27–29).

Very recently, the crystal structure of an entire archaeal box H/ACA sRNP was determined (30). Through reconstitution experiments, we showed that in archaea the aCBF5–aNOP10 complex is the minimal set of proteins required to get an RNA-guided RNA:Ψ-synthase activity (36). However, the two other proteins aGAR1 and L7Ae both increase the efficiency of the catalytic reaction. Formation of an H/ACA sRNA–aCBF5–aNOP10 complex depends upon aCBF5 association with the sRNA and the conserved ACA motif in the sRNA plays an essential role for this association (35,36). More precisely, the PUA domain of aCBF5 interacts with the ACA motif of H/ACA sRNAs (27,30). According to our previous reconstitution assays, formation of an sRNA–aCBF5–aNOP10 complex is required for binding of the RNA substrate and recent structural studies on NOP10 revealed its capability to bind RNA (37). By *in vitro* reconstitution of H/ACA sRNPs with truncated aNOP10 proteins, we showed that its C-terminal half is sufficient for sRNP activity, whereas its N-terminal half reinforces the aCBF5-aNOP10 interaction (27). The L7Ae protein binds a K-turn motif present in each stem-loop structures of the archaeal H/ACA sRNAs (35,36,38,39). The direct interaction of L7Ae with aNOP10 and the sRNA stabilizes the association of the other proteins with the sRNA (30). In light of the higher rate of pseudouridylation observed in the presence of aGAR1, we proposed that protein aGAR1 might enhance the turnover of the reaction catalyzed by the sRNP (36).

A recent *in vitro* study showed that in addition to its central role in the RNA guided activity of H/ACA sRNPs, the archaeal aCBF5 enzyme has the ability to convert residue U55 of an *in vitro* transcribed tRNA into a Ψ residue (40). This activity seems to be increased in the presence of the L7Ae, aNOP10 and aGAR1 set of proteins (40). Altogether, these data revealed the ability of aCBF5 to act without guide RNA in case of a tRNA substrate. However, an *in vivo* dual role of aCBF5, as rRNA:Ψ-synthase when included in H/ACA sRNPs and as tRNA:Ψ55-synthase when acting as a free protein, is still not demonstrated since another archaeal RNA:Ψ-synthase PusX/Pus10 is also able to catalyze U to Ψ conversion at position 55 in tRNAs. Nevertheless, the capability of aCBF5 to act as a free protein and in sRNPs represented an opportunity: (i) to delineate the tRNA structures needed to get an activity of aCBF5 in the absence of guide RNA, (ii) to compare the aNOP10 and aGAR1 requirements for the guided and non guided activity of aCBF5, (iii) to compare the aCBF5 amino acid requirement for its guided and non-guided activity. Here, by site-directed mutagenesis of the *Pyrococcus abyssi* and *Saccharomyces cerevisiae* tRNAs, we show that a truncated version of tRNA including the acceptor and TΨC stem-loop structures linked by a flexible sequence is sufficient to get aCBF5 activity in the absence of a guide RNA. In the absence of the 3′ terminal CCA, the aCBF5 activity becomes strictly dependent on either aNOP10 or aGAR1. Furthermore, by site-directed mutagenesis of both aNOP10 and aCBF5 we demonstrate distinct amino acid requirements for the RNA-guided and non RNA-guided activities of aCBF5. The results obtained are discussed based on the 3D structures of the archaeal H/ACA sRNP proteins and their complexes.

MATERIALS AND METHODS

Production of protein variants

The aCBF5 D82A and CΔ(PUA) protein variants carrying, respectively, an Asp to Ala substitution at position 82 and a truncation of the PUA domain, were previously described (27,36). The PCR approach was used for site-directed mutagenesis of protein aCBF5 (K53A, H77A and R202A variants) and aNOP10 (Y14A, H31A and P32A variants). The sequence of the oligonucleotides used for the PCR reactions are available upon request.

Purification of recombinant proteins

Wild-type and variant aCBF5 and aNOP10 proteins as well as the aGAR1 and TruB proteins were produced as GST fusion proteins as previously described (36). The GST moiety was removed by cleavage with the PreScission protease. The various proteins were stored at −80°C in the following buffer 150 mM NaCl, 1 mM EDTA, 1 mM DTT, glycerol 10%, 50 mM Tris–HCl pH 7.

In vitro transcription of tRNAs, H/ACA sRNA and the sRNP target RNA

Wild-type and mutated tRNA transcripts, the Pab91 H/ACA sRNA and the RNA target RNA-S (36) were produced by *in vitro* transcription of PCR products. The forward primers used for PCR amplification generated the T7 RNA polymerase promoter. DNA templates for amplifications of sequences coding for the WT yeast $tRNA^{Asp}$ and variants of this tRNA (ΔT-$tRNA^{Asp}$, ΔD-$tRNA^{Asp}$,) were a generous gift of C. Florentz (IBMC, Strasbourg). Additional yeast $tRNA^{Asp}$ variants (Δa.c.-$tRNA^{Asp}$, Δacceptor-$tRNA^{Asp}$, Δacceptor-$tRNA^{Asp}$-U55C) were produced by site-directed mutagenesis in the course of the PCR amplification. The sequence coding for the WT *P. abyssi* $tRNA^{Asp}$ was PCR amplified using genomic DNA from the *P. abyssi* GE5 strain. Sequences encoding the variants $tRNA^{Asp}$-ACA and $tRNA^{Asp}$-ΔCCA were obtained by using a reverse primer carrying the mutated sequence. Sequences encoding the WT semi-$tRNA^{Asp}$ or the variant semi-$tRNA^{Asp}$-ΔCCA or semi-$tRNA^{Asp}$-U55C were obtained by hybridization of two complementary oligonucleotides. The other archaeal $tRNA^{Asp}$ variants ($tRNA^{Asp}$-U55C) were produced by site-directed mutagenesis in the course of a PCR amplification. DNA templates for production of the *P. abyssi* Pab91 sRNA and its target RNA (RNA-S) were previously described (36). The RNA transcripts were purified by denaturing gel electrophoresis. Labeling was done at their 5′ end for EMSA experiments or during *in vitro* transcription by incorporation of [α-^{32}P]CTP for the pseudouridylation assays. Conditions for transcription and labeling were previously described (41).

Electrophoresis mobility shift assays (EMSA)

EMSA were performed in the conditions previously described (36,41). Briefly, radiolabeled RNAs (50 fmol) were incubated in buffer D [150 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA and 20 mM HEPES (pH 7.9)] with proteins at a 200 nM concentration, at 65°C for 10 min. To test for the association of the Pab91 sRNP with its target RNA (complex CII), 2.5 pmol of RNA-S were added. Complexes were visualized by autoradiography after fractionation by non-denaturing polyacrylamide gel electrophoresis.

In vitro pseudouridylation assay

The tRNA:Ψ55-synthase activities of purified aCBF5 or H/ACA–RNP complexes were measured by the nearest-neighbor approach in the conditions previously described (36,41). The RNA-guided activity of reconstituted H/ACA sRNPs was measured as previously described by mixing ∼4 pmol of Pab91 sRNA with ∼150 fmol of the [α-^{32}P]CTP-labeled target RNA-S (36). The non-RNA-guided reaction of aCBF5 on tRNA was tested on ∼50 fmol of [α-^{32}P]CTP-labeled tRNAs. For both RNA-guided and non-RNA-guided reactions, samples were pre-incubated at 65°C and the reaction was started by addition of the proteins (200 nM each). Aliquots were collected at different time points, the RNAs were extracted and then digested by RNase T2. The resulting 3′-mononucleotides were fractionated by thin-layer cellulose chromatography. The radioactivity in the spots was quantified using the ImageQuant software after exposure of a phosphorimager screen. The amount of Ψ residue formed was determined taking into account the total number of U residues in the tRNA or RNA-S molecules. Some of the values given in the text are expressed with standard errors of the mean values. All the time-course analyses were repeated at least three times and gave reproducible data. Some control experiments were performed with the *E. coli* TruB recombinant protein. The same procedure was used to measure the activity, except that the reaction was performed at 37°C.

Mapping the site of pseudouridylation in tRNAs

We used the *N*-cyclohexyl-*N*′-(2-morpholinoethyl)-carbodiimide metho-*p*-toluolsulfonate (CMCT) modification protocol of (42), adapted by (43). CMCT modifications were performed with 0.5 μg of *in vitro* transcribed archaeal or yeast $tRNA^{Asp}$. Positions of CMCT modifications were identified by primer extension analysis, using the AMV RT (QBiogene, USA) in the conditions previously described (44) and the primer oligonucleotides OG4891 for archaeal $tRNA^{Asp}$ and OG5302 for *S. cerevisiae* $tRNA^{Asp}$ (Figure 1A1 and B1). Primers were 5′-end labeled using [γ-^{32}P]ATP (3000 Ci/mmol) and T4 polynucleotide kinase. RNA sequencing was done with 4 μg of *in vitro* transcribed RNA [for details on the technique, see (45)].

RESULTS

The kinetics of U to Ψ conversion at position 55 in tRNAs is enhanced in the presence of proteins aNOP10 and aGAR1

Previous studies revealed a U to Ψ conversion activity of the *P. furiosus* aCBF5 protein at position 55 in the *P. furiosus* $tRNA^{Asp}$(GUC), i.e. a tRNA:Ψ55-synthase activity. Such an activity is enhanced in the presence of the aNOP10, aGAR1 and L7Ae proteins set (40). For a deeper understanding of these activation properties, here we quantified the relative stimulation effects of proteins aNOP10, aGAR1 and L7Ae alone and in various combinations, on the tRNA:Ψ55-synthase activity of aCBF5. The assays were performed with the *P. abyssi* recombinant proteins aCBF5, aNOP10, aGAR1 and L7Ae, and with both *P. abyssi* $tRNA^{Asp}$(GUC) and *S. cerevisiae* $tRNA^{Asp}$(GUC) substrates (Figure 1A1 and B1). We used the nearest neighbor and the RT–CMCT approaches in order to verify that modifications occurred only at position 55 in tRNAs. For the

nearest-neighbor approach, both substrates were labeled by incorporation of [α-^{32}P]CTP during transcription. After the pseudouridylation reactions, the tRNA transcripts were digested with RNase T2 and the products were fractionated by 2D chromatography (Figure 1A2 and B2). CMCT modifications followed by alkaline treatments were performed on cold RNA transcripts, and positions of modifications were identified by reverse transcription (Figure 1A3 and B3) using oligonucleotide primers that were complementary to the 3′ terminal sequence of tRNAs (Figure 1A1 and B1). The pseudouridylation reaction was performed at 65°C, for 80 min using a 200 nM concentration of each protein. Time point experiments were performed on the *P. abyssi* tRNA^{Asp} with: (i) aCBF5 alone, (ii) aCBF5 in the presence of aNOP10 or aGAR1, or of both proteins and (iii) aCBF5 in the presence of aNOP10 and L7Ae, or the three L7Ae, aNOP10, aGAR1 proteins. Aliquots of the reaction mixtures were analyzed after 5, 10, 20, 40 and 80 min of incubation (Figures 1A4 and S1). For comparison, a time point experiment was performed at 37°C with a recombinant *E. coli* TruB enzyme (Figure 1A2).

Altogether, the results obtained demonstrated that protein aCBF5 is able to modify the *P. abyssi* and *S. cerevisiae* tRNA^{Asp} at the same high level after 80 min of incubation (~0.85 mol ± 0.05 and ~0.9 mol ± 0.05 of Ψ per mol of tRNA, respectively). In both cases, only a slight increase of the level of modification was observed in the presence of aNOP10 and aGAR1 (Figure 1A2 and B2). However, as evidenced by the time point experiments, although the kinetics of the reaction with aCBF5 at 65°C is very fast compared to that for TruB at 37°C, it was markedly increased in the presence of aNOP10 or aGAR1. On the contrary, protein L7Ae showed no stimulatory effect (see Supplementary Data, Figure S1), which is in contrast to data obtained for the sRNA-guided activity of aCBF5 (35,36). The CMCT analysis did confirm that aCBF5 acted only at position 55 in both tRNAs. Finally, we confirmed the general capability of aCBF5 to act at position 55 in tRNAs by using other archaeal and bacterial transcripts as the substrates (data not shown).

Altogether, we concluded that aCBF5 has the ability to modify various tRNAs at position 55 and this independently of the organisms from which they originate. The reaction is very rapid and is nearly complete after a 15-min incubation. However, the kinetics of this reaction is markedly increased in the presence of aNOP10 or aGAR1. In this case, the reaction is completed after a 7-min incubation. Furthermore, the catalytic property of aCBF5, alone or acting in the presence of aNOP10 and aGAR1 is highly specific of position 55.

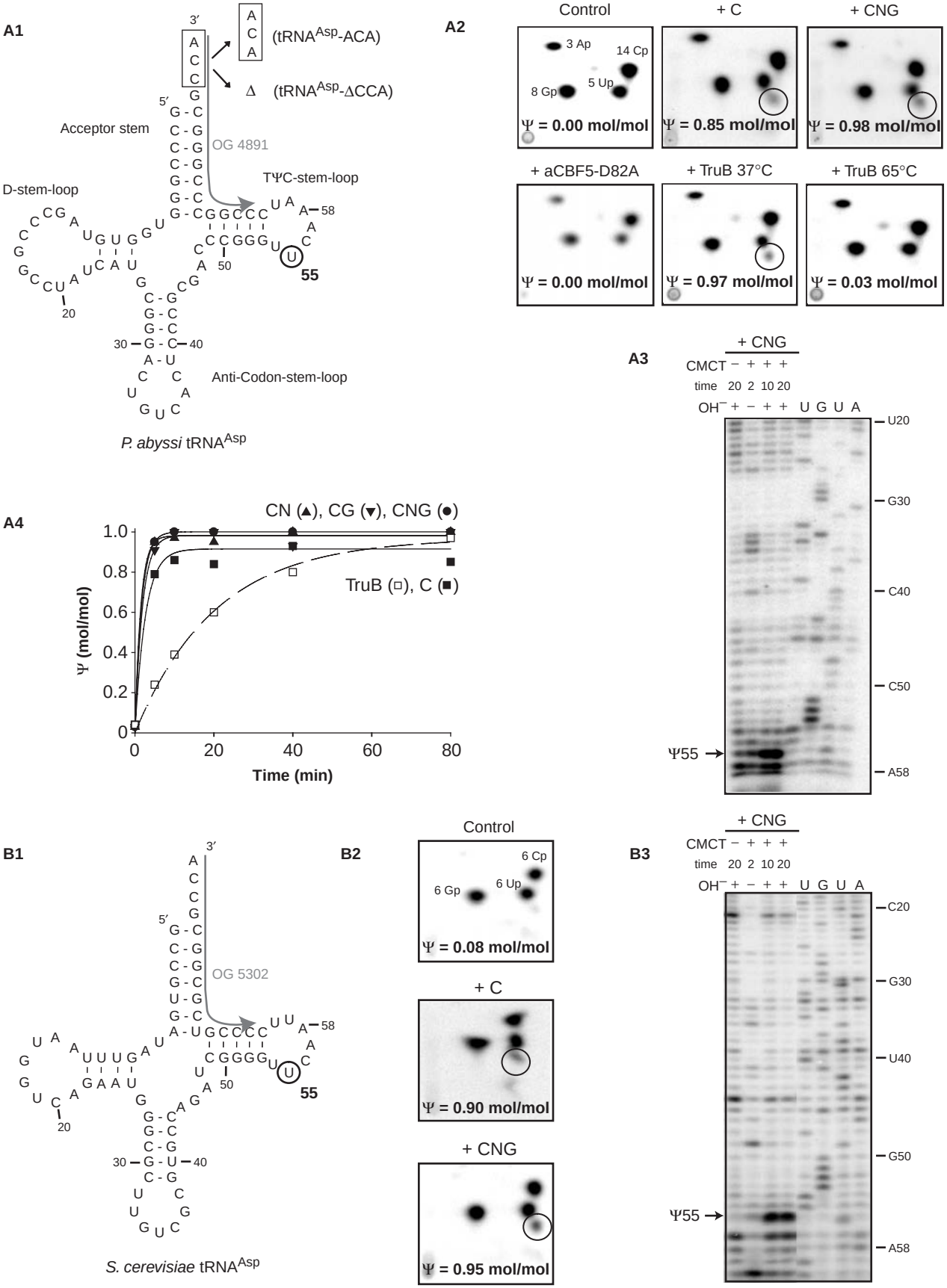
Proteins aNOP10 and aGAR1 have a strong stimulatory effect on Ψ55 formation in the absence of the aCBF5 PUA domain or of the tRNA 3′ terminal CCA motif

Our previous data on the sRNA guided activity of aCBF5 revealed a strong importance of the PUA domain of aCBF5 for this activity and a negative effect of mutations in the ACA sequence of the H/ACA sRNA (27). The recent crystal structure of an H/ACA sRNP demonstrated a direct contact of the PUA domain with the ACA motif (30). Therefore, we tested whether the PUA domain was also required for the activity of aCBF5 at position 55 in tRNAs. As shown in Figure 2, no U to Ψ conversion occurred in the *P. abyssi* tRNA^{Asp} when it was incubated with an aCBF5 protein lacking the PUA domain (CΔ(PUA)). Interestingly, this tRNA:Ψ55-synthase activity was partially restored in the presence of the aNOP10 and aGAR1 proteins pair (CΔ(PUA)NG). However, the kinetics was slower and the efficiency of modification after 80 min was lower. The individual aNOP10 and aGAR1 proteins had similar but very low capabilities to restore the tRNA:Ψ55-synthase activity of the truncated aCBF5 protein (Figure 2).

As previous data had revealed the absence of aCBF5 activity on a tRNA lacking the 3′ terminal CCA motif (40), we tested whether aNOP10 or aGAR1 or both proteins can compensate for the absence of this motif (Figure 3). We found in this case that individual aNOP10 or aGAR1 can restore a strong aCBF5 activity (CN and CG, respectively) at position 55 of tRNAs and this activity was reinforced to some extent when the two proteins were present (CNG). Remarkably, modification with the CNG complex was almost complete after a 10-min incubation, even in the absence of the CCA motif (Figure 3). The truncated tRNA was modified at a slower rate and lower yield, when TruB was used as the catalyst. Taken together, the data revealed a very important role of the PUA domain for the aCBF5 tRNA:Ψ55-synthase activity and showed the capability of the aNOP10–aGAR1 protein pair to compensate partially for its absence. Moreover, the presence of the auxiliary proteins fully restores the aCBF5 activity in the absence of the tRNA CCA 3′ terminal motif.

The aNOP10–aGAR1 protein pair can compensate for severe truncations in tRNAs

As we found that aNOP10 and aGAR1 can compensate for CCA deletion, we wondered whether they can compensate for more severe truncations in tRNAs. As aCBF5 was acting on the *S. cerevisiae* tRNA^{Asp}, and as several variants of this tRNA had already been produced (Figure 4A1), we tested the capability of aCBF5 (C) and the aCBF5–aNOP10–aGAR1 complex (CNG) to convert U55 into a Ψ residue in these truncated tRNAs (Figure 4A2). In RNAs ΔT-tRNA^{Asp}, Δa.c.-tRNA^{Asp} and ΔD-tRNA^{Asp}, the T stem-loop, the anticodon stem-loop, and the D stem-loop were deleted, respectively. The acceptor stem was truncated in RNA Δacceptor-tRNA^{Asp}, but the CCA 3′-terminal motif was conserved (Figure 4A1). Obviously, no Ψ formation was detected in the ΔT-tRNA^{Asp} substrate which was lacking residue U55. Only a very low level of modification was detected in the Δa.c.-tRNA^{Asp} when aCBF5 was used alone, while a high level of modification was detected when aNOP10 and aGAR1 were present in the incubation mixture (Figure 4A2). Similarly, no modification of the Δacceptor-tRNA^{Asp} substrate was observed when the incubation was performed with aCBF5 alone, whereas almost a complete modification was detected for these two


A1
(tRNA^Asp-ACA)
Δ (tRNA^Asp-ΔCCA)
Acceptor stem
OG 4891
TΨC-stem-loop
D-stem-loop
Anti-Codon-stem-loop
P. abyssi tRNA^Asp
A2
Control
+ C
+ CNG
3 Ap
14 Cp
8 Gp
5 Up
Ψ = 0.00 mol/mol
Ψ = 0.85 mol/mol
Ψ = 0.98 mol/mol
+ aCBF5-D82A
+ TruB 37°C
+ TruB 65°C
Ψ = 0.00 mol/mol
Ψ = 0.97 mol/mol
Ψ = 0.03 mol/mol
A3
+ CNG
CMCT − + + +
time 20 2 10 20
OH⁻ + − + + U G U A
U20
G30
C40
C50
Ψ55
A58
A4
CN (▲), CG (▼), CNG (●)
TruB (□), C (■)
Ψ (mol/mol)
Time (min)
B1
OG 5302
S. cerevisiae tRNA^Asp
B2
Control
6 Cp
6 Gp
6 Up
Ψ = 0.08 mol/mol
+ C
Ψ = 0.90 mol/mol
+ CNG
Ψ = 0.95 mol/mol
B3
+ CNG
CMCT − + + +
time 20 2 10 20
OH⁻ + − + + U G U A
C20
G30
U40
G50
Ψ55
A58

truncated tRNAs in the presence of aNOP10 and aGAR1 (Figure 4A2). Modification only occurred at position 55 in the truncated substrates since it disappeared after U to C mutation at this position (Δacceptor-tRNA^{Asp}-U55C).

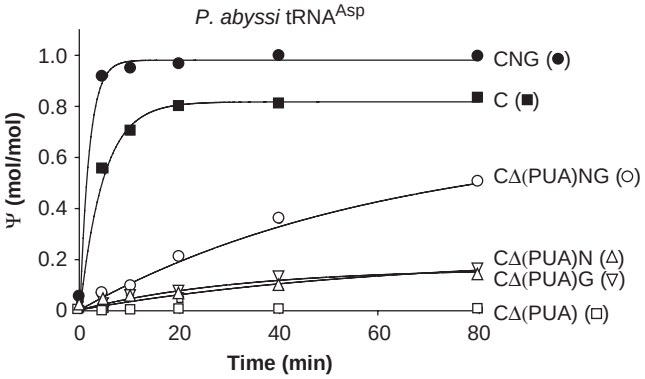


Figure 2. The PUA domain of aCBF5 is required for Ψ55 formation. Time-course experiments of Ψ55 formation were performed as in Figure 1 panel A4, using either the full-length aCBF5 (C) or aCBF5 truncated of its PUA domain [CΔ(PUA)], in the presence or the absence of the aNOP10 and/or aGAR1 proteins [CNG, CΔ(PUA)N, CΔ(PUA)G and CΔ(PUA)NG].

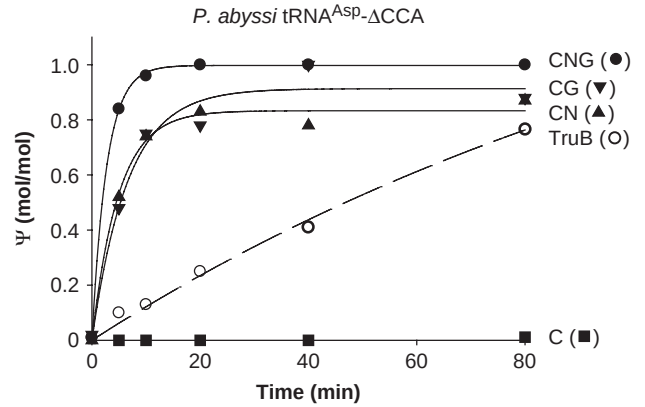


Figure 3. The aCBF5–aNOP10–aGAR1 complex can efficiently modify a tRNA lacking a 3′ CCA sequence. Time-course experiments of Ψ55 formation in the variant tRNA^{Asp}-ΔCCA by the *E. coli* TruB enzyme or various combinations of the aCBF5 (C), aNOP10 (N) and aGAR1 (G) proteins (same conditions as in Figure 2).

Interestingly, the ΔD-tRNA^{Asp} was modified by aCBF5 at a high level in absence of the auxiliary proteins (Figure 4A2). As expected from previous data (46), TruB was able to modify a tRNA lacking the acceptor stem without the need for auxiliary proteins. Based on the activity of aCBF5 alone on the ΔD-tRNA^{Asp}, we concluded that aCBF5 alone, like TruB, does not need the tRNA 3D structure for activity. However, in contrast to TruB, to act alone aCBF5 needs the presence of its target T stem-loop but also the anticodon stem-loop and the acceptor stem.

Roovers *et al.* (40) had found that aCBF5 does not modify a truncated *P. furiosus* tRNA^{Asp} that contains the acceptor stem fused to the T stem-loop. We confirmed that for an equivalent *P. abyssi* substrate, no activity was restored upon addition of the aNOP10–aGAR1 protein pair (data not shown). However, as aCBF5 was active on the yeast ΔD-tRNA^{Asp} substrate, we postulated that flexibility between the acceptor stem and the T stem-loop might be an important feature for aCBF5 activity. Therefore, we produced a truncated RNA substrate (semi-tRNA^{Asp}) by truncation of the *P. abyssi* tRNA^{Asp}, which contained a UGAC single-stranded sequence acting as a flexible link between the acceptor stem and the T stem-loop (Figure 4B1). The aCBF5 protein alone had a modest but significant activity on this substrate (mean value $<0.2\,mol.mol^{-1}$). This activity was abolished in the absence of the 3′ terminal CCA sequence (Figure 4B2). However, in the presence of the aNOP10–aGAR1 protein pair both the semi-tRNA^{Asp} and semi-tRNA^{Asp}-ΔCCA lacking the 3′ terminal CCA sequence were almost completely modified. As expected, TruB was active on both substrates in the absence of auxiliary proteins. Hence, we concluded that the aCBF5–aNOP10–aGAR1 heterotrimer, but not aCBF5 alone, has catalytic properties similar to that of TruB.

aNOP10 reinforces the formation of aCBF5–tRNA complexes

As the aNOP10 stimulatory effect on the aCBF5 activity seemed to consist in an increase of the kinetics of the reaction (36), and as aNOP10 was known to bind RNA (30,37), we hypothesized that aNOP10 might facilitate the association of aCBF5 with the tRNA substrate. To test

Figure 1. The *P. abyssi* aCBF5 protein modifies position 55 in *P. abyssi* and *S. cerevisiae* tRNA^{Asp}. (**A1**) Secondary structure of the *P. abyssi* tRNA^{Asp} used in this study. Residue U55 is circled. The CCA sequence at the 3′ end is boxed. The CCA to ACA substitution and the CCA deletion in variants tRNA^{Asp}-ACA and tRNA^{Asp}-ΔCCA, are shown. (**A2**) 2D Thin-layer-chromatography (TLC) analysis of Ψ formation in *P. abyssi* tRNA^{Asp}. The *P. abyssi* tRNA^{Asp} transcript was labeled by the incorporation of [α-^{32}PCTP] during transcription. The ^{32}P-labeled tRNA substrate (50 fmol) was mixed with either the *P. abyssi* aCBF5 protein (C), or its aCBF5-D82A variant, or the aCBF5-aNOP10–aGAR1 complex (CNG), or the *E. coli* TruB enzyme (200 nM each). Incubation was performed for 80 min at 65°C for the archaeal enzymes and at 37 or 65°C for TruB. After T2 RNase digestion, the amount of Ψ residue formed was measured by 2D TLC and expressed in moles per moles of tRNA. The data shown are representative examples of at least three independent experiments. (**A3**) CMC/RT analysis for the detection of Ψ formation at position 55 in the *P. abyssi* tRNA^{Asp}. The unlabeled tRNA (2 μg, 80 pmol) was incubated with proteins aCBF5, aNOP10 and aGAR1 (CNG, 200 mM each) for 80 min at 65°C. Treatment with CMCT was carried out for 2, 10 or 20 min as described in Materials and Methods section with (+) or without (−) alkaline treatment. Positions of Ψ residues were identified by primer extension analysis using oligonucleotide OG 4891 (represented by the arrow in panel A1). Lanes U, G, C and A correspond to the RNA sequencing ladder. (**A4**) Time course of Ψ55 formation by aCBF5 protein in the *P. abyssi* tRNA^{Asp} and effect of the addition of proteins aNOP10 and aGAR1. The *P. abyssi* tRNA^{Asp} substrate was labeled as in panel A2 and incubated with various protein combinations: aCBF5 alone (C), aCBF5 and aNOP10 (CN), aCBF5 and aGAR1 (CG), a set of the three proteins (CNG) or TruB. For each time point, an aliquot of the reaction mixture was analyzed by 2D TLC. The amount of Ψ residue was estimated as in panel A2. (**B1**) Secondary structure of the *S. cerevisiae* tRNA^{Asp}. The sequence of oligonucleotide OG 5302 used as the primer for reverse transcription is indicated by an arrow. (**B2**) 2D-TLC analysis of Ψ formation by aCBF5 in the *S. cerevisiae* tRNA^{Asp} (same conditions as in panel A2). (**B3**) Detection of Ψ55 formation in the yeast tRNA^{Asp} by the CMC/RT approach (same conditions as in panel A3).

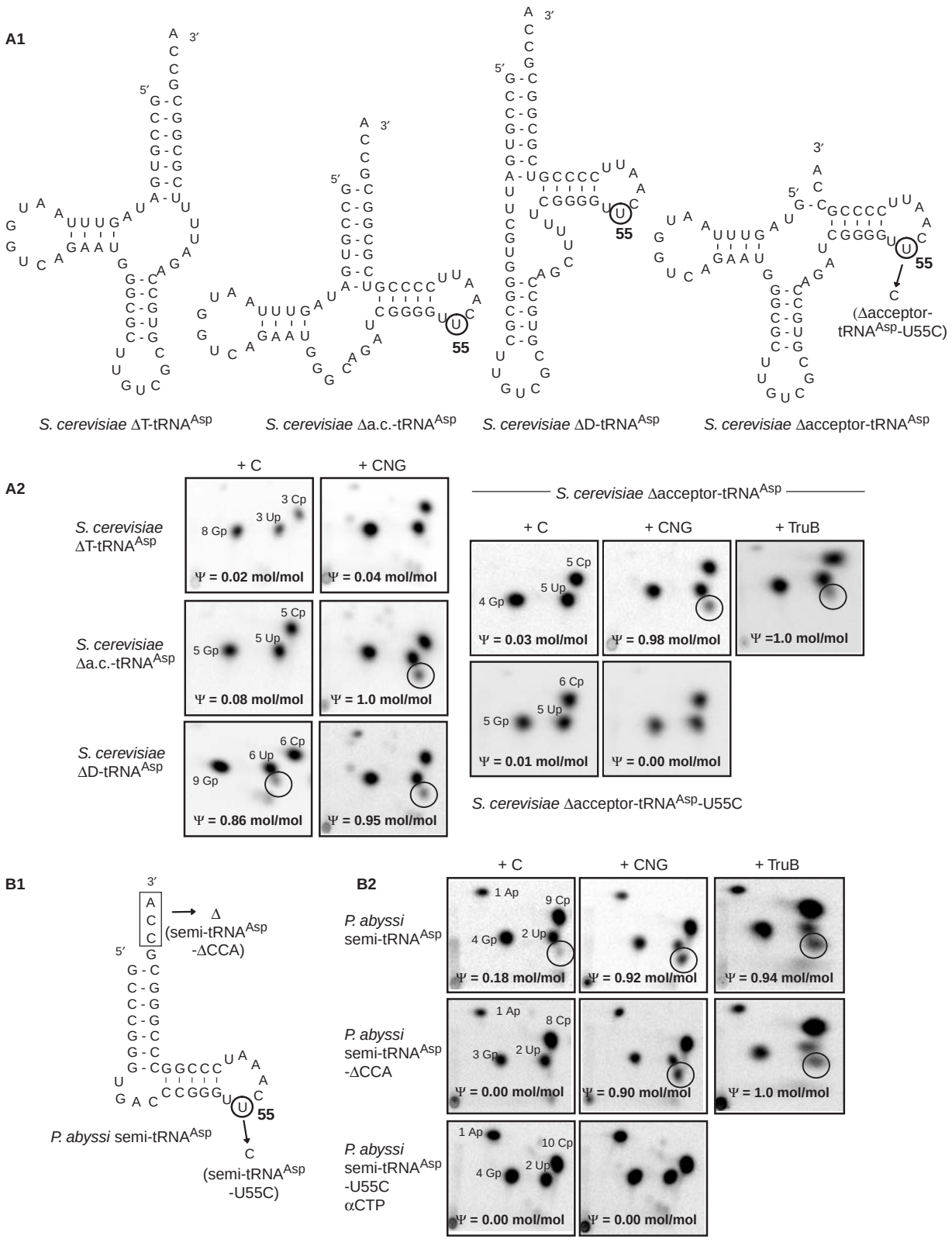


Figure 4. tRNA structural requirements for Ψ55 formation by aCBF5 or the aCBF5–aNOP10–aGAR1 complex. (**A1**) Secondary structure of the yeast truncated tRNAs^{Asp} used in this study. The DNA templates for transcription of the ΔT-tRNA^{Asp}, Δa.c.-tRNA^{Asp}, ΔD-tRNA^{Asp} and Δacceptor-tRNA^{Asp} were obtained by deletion of the T stem-loop, anticodon stem-loop, D stem-loop and acceptor stem sequences of the *S. cerevisiae* tRNA^{Asp} template sequence (Figure 1B1). (**B1**) The template sequence for the semi-tRNA^{Asp} was obtained by successive site-directed mutagenesis of the *P. abyssi* tRNA^{Asp} template sequence. (Figure 1A1). The variant semi-tRNA^{Asp}-ΔCCA with a deletion of the CCA sequence and the variant semi-tRNA^{Asp}-U55C with a U to C substitution are shown. (**A2** and **B2**) 2D-TLC analysis of Ψ formation by aCBF5, the aCBF5–aNOP10–aGAR1 complex or TruB, in the various truncated tRNAs. The same experiments conditions as in Figure 1A2 and B2 were used.

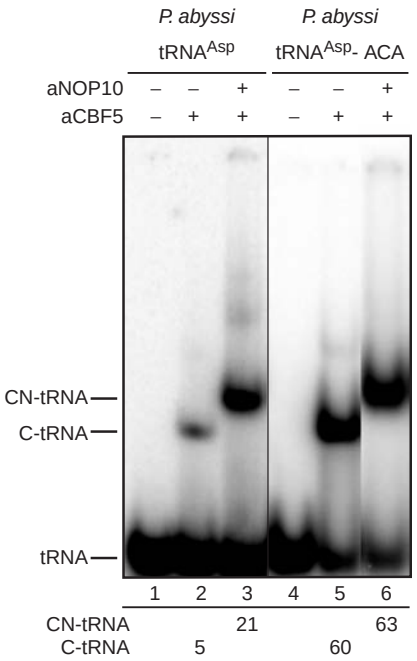


Figure 5. Analysis by electrophoresis mobility shift assays (EMSA) of the complexes formed between protein aCBF5 and archaeal tRNAs. The radiolabeled $tRNA^{Asp}$ (lanes 1–3) or $tRNA^{Asp}$-ACA (lanes 4–6) (50 fmol) were incubated with aCBF5 protein (C) (200 nM), in the presence or the absence of aNOP10 (N) (200 nM) as indicated on top of each lane. Positions of the various complexes C–tRNA and CN–tRNA are indicated on the left side of the autoradiogram. The amount of the complexes was estimated as described in Materials and Methods section.

this possibility, we estimated by EMSA experiments the level of complexes formed by the association of aCBF5 with a ^{32}P-labeled *P. abyssi* $tRNA^{Asp}$. The tests were performed in the presence or absence of aNOP10 (Figure 5). The level of complex formation was increased by a factor of ~4 in the presence of aNOP10 (Figure 5, compare lane 2 with lane 3), whereas no significant increase was observed in the presence of aGAR1 (data not shown). As aCBF5 is known to interact with the ACA conserved sequence of H/ACA sRNAs, we compared the affinity of aCBF5 for the WT $tRNA^{Asp}$ to that for a tRNA carrying an ACA triplet instead of the CCA triplet. The affinity of aCBF5 was indeed largely increased in the presence of an ACA triplet (Figure 5, compare lane 2 with lane 5) and aNOP10 did not increase the level of complex formation in this case (Figure 5, lane 6). Therefore, the stimulatory effect of aNOP10 on the kinetics of aCBF5 action on tRNAs may in part be due to its capability to facilitate the aCBF5–tRNA interaction. As previously observed for the aCBF5–RNA-guided activity, the stimulatory effect of aGAR1 on the aCBF5 activity is likely of another nature.

The conserved H31 and P32 residues in aNOP10 have different implications for the RNA-guided and non-RNA-guided activities of aCBF5

Based on the 3D structures established for the aCBF5–aNOP10 dimer and the aCBF5–aNOP10–aGAR1 trimer, the direct interactions between some conserved residues in the central flexible linker of aNOP10 (Figure 6A) with conserved residues of aCBF5 were proposed to influence the sRNA-guided activity of aCBF5 (28,29). In particular, the aNOP10 residues Y14 and H31 were found to form hydrogen bonds with residues E199 and R201 which are present in the β12 strand of aCBF5 (Figure 6B) (29). These interactions were proposed to influence the orientation of the side-chain of residue R202 that is located on the opposite face of the β12 strand. Interestingly, in the crystal structure of the complex formed between the *E. coli* TruB RNA:Ψ-synthase and a tRNA T stem-loop, the counterpart of residue R202 was proposed to interact with the phosphate on the 5′ side of the targeted uridine (31). On the other hand, in the aCBF5–aNOP10 complex, residue P32 in the aNOP10 flexible linker was found to interact by Van der Waals interaction with the conserved residue P54, which is located in motif I of aCBF5 (Figure 6C). This interaction was proposed to facilitate the formation of an hydrogen bond between residue K53 and the carbonyl oxygen of the catalytic residue D82 (28).

In light of these structural data, we used alanine substitutions to test for the relative importance of residues Y14, H31 and P32 for the RNA-guided and non-RNA-guided activities of aCBF5. Aiming at this, the stimulatory effects of each aNOP10 variant (Y14A, H31A and P32A) on the aCBF5 activity were tested on both the *P. abyssi* $tRNA^{Asp}$ lacking the CCA sequence (Figure 6D) and the RNA-S substrate (RNA target) of the *P. abyssi* Pab91 sRNPs (Figure 6F). To this end, sRNPs were reconstituted with each of the variant aNOP10 proteins. Whereas the H31A substitution had only a limited effect on the kinetics and the yield of Ψ55 formation in the archaeal $tRNA^{Asp}$ (Figure 6D), the Y14A substitution strongly slowed down the kinetics of this modification and the P32A substitution had a negative effect on the modification yield (~0.7 mol of Ψ per mole of tRNA, instead of ~0.9 mol of Ψ per mole of tRNA in the presence of the WT aNOP10 protein). Interestingly, the Y14A and P32A variants only had a limited effect on the sRNA-guided activity of aCBF5 (Figure 6F). The H31A mutation was in contrast strongly deleterious for this activity: only ~0.2 mol of Ψ per mole of target RNA was obtained after a 80 min incubation. Hence, mutations in aNOP10 have marked different effects on the RNA-guided and non-RNA-guided activities of aCBF5.

To try to define the reasons for the observed differences, we tested by EMSA the ability of the aNOP10 variant proteins to reinforce the aCBF5 association with the *P. abyssi* $tRNA^{Asp}$ and with the Pab91 sRNA and its target RNA sequence (Figure 6E and G). In agreement with the time-course results, both Y14A and P32A aNOP10 variants were less efficient (~0.5-fold) than the H31A aNOP10 variant and the WT aNOP10 protein for increasing the aCBF5 association with the tRNA (Figure 6E). On the contrary, only the H31A variant had a dramatic effect on the yield of complex CII (sRNA-L7Ae-aCBF5-aNOP10-RNA target) formation, suggesting that the H31A mutation has an important negative effect on substrate recruitment in the H/ACA sRNP. Therefore, we concluded that whereas residues

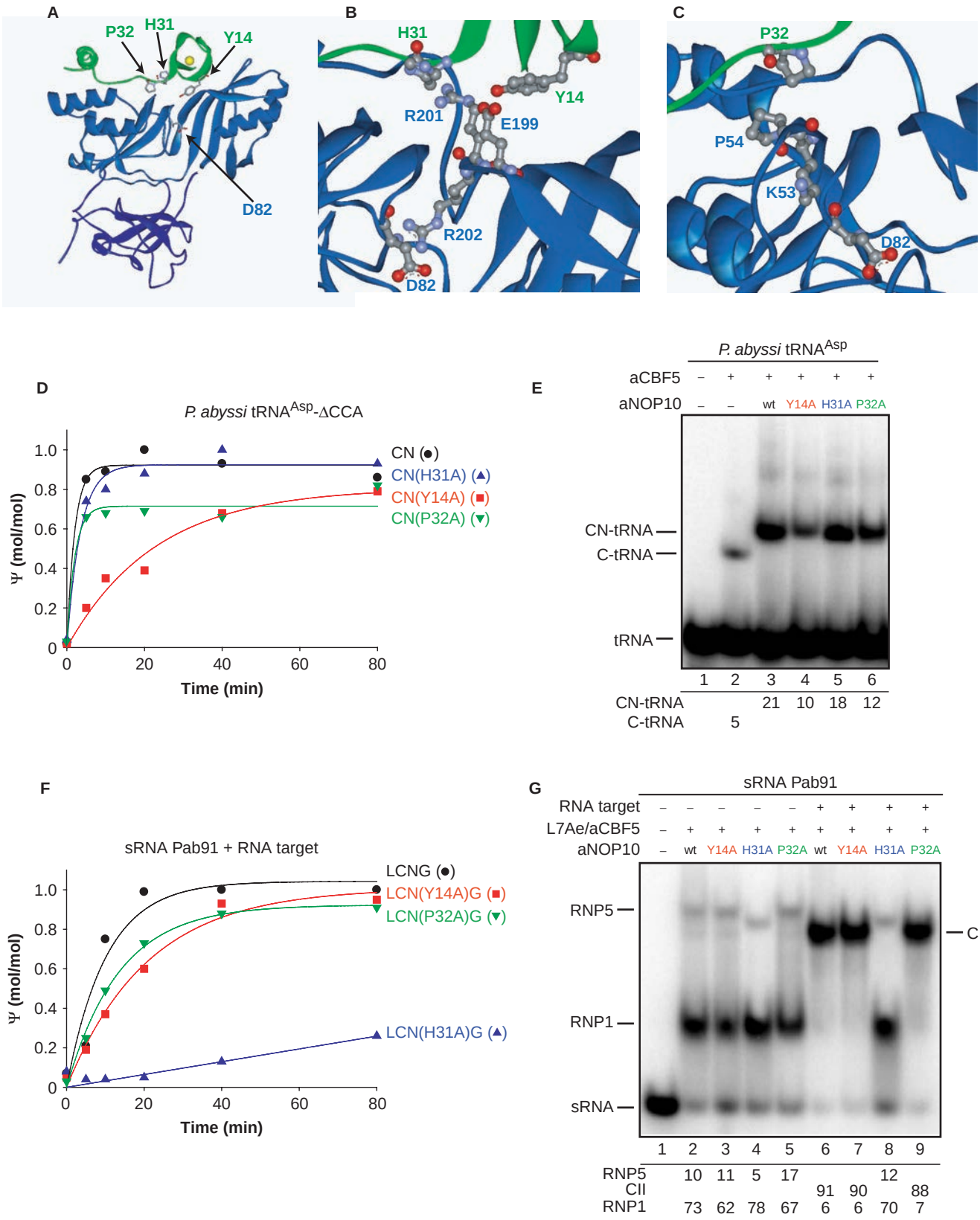


Figure 6. Role of the conserved residues in the aNOP10 protein for the non-RNA-guided and RNA-guided activities of aCBF5. (**A**) Ribbon representation of the heterodimer formed by interaction of aCBF5 (blue) with aNOP10 protein (green). The side chains of the aNOP10 residues P32, H31 and Y14 are represented, as well as the aCBF5 catalytic D82. (**B**) The residues expected to be involved in a network of interactions between aNOP10 and aCBF5 (29) are indicated on the *P. abyssi* structure. (**C**) Residues proposed to be involved in another network of interaction between aNOP10 and aCBF5 (28) are shown on the *P. abyssi* structure. (**D**) Time course of Ψ55 formation by aCBF5 (C) in the presence of the wild-type (N)

Y14 and P32 are highly important for the aCBF5 tRNA:Ψ55-synthase activity, residue H31 plays an essential role in substrate recruitment in the sRNA-guided activity of aCBF5.

aNOP10 and aGAR1 can compensate for R202A and K53A mutations in aCBF5 for the non-guided but not for the RNA-guided activity

Next, we tested the role of the aCBF5 residues R202 and K53 whose functions in the catalytic activity were proposed to be modulated by the aNOP10 residues H31 and Y14, and by residue P32, respectively (28,29). The aCBF5 variants R202A and K53A were produced, and the effect of each substitution was tested on both RNA-guided and non-RNA-guided activities. As evidenced in Figure 7A1, both R202A and K53A substitutions in aCBF5 abolished its activity at position 55 in the *P. abyssi* $tRNA^{Asp}$. However, this activity was fully restored upon addition of aNOP10 and aGAR1 in the incubation mixture (Figure 7A3). Here again, as evidenced by EMSA, the defect of activity observed for CBF5 alone was at least in part due to its low affinity for the tRNA (Figure 7A2). The addition of aNOP10 enhanced the association of aCBF5 with the tRNA, which likely explains the restoration of the tRNA:Ψ55-synthase activity. Hence, residues R202 and K53 likely play a role in tRNA association and aNOP10 can compensate for their mutations. In contrast, aNOP10 and aGAR1 were not able to compensate the negative effect of the R202A and K53A mutations in aCBF5 for the Pab91 sRNP activity (Figure 7B1), and here again, the low activity of the reconstituted sRNPs could be explained by a defect in substrate recruitment as the CII complex was formed at lower yields and displays a peculiar electrophoretic mobility (Figure 7B2, lanes 7 and 9). Hence, we concluded that residues K53 and R202 are important for tRNA binding in the non-RNA-guided system and for recruitment of the RNA substrate in the RNA-guided system. However, in this latter case, aNOP10 and aGAR1 cannot compensate for the absence of these residues. Therefore, residues R202 and K53 appear to play a more crucial role in the RNA-guided than in the non-RNA-guided system.

For both the RNA-guided and non-RNA-guided activities, the aCBF5 conserved residue H77 is dispensable in the presence of aNOP10 and aGAR1

In the *E. coli* TruB enzyme, residue H43 was proposed to be involved in the flipping of residue U55 out of the helical stack (31). The counterpart of residue H43 in aCBF5 is residue H77. As expected, the H77A mutation in aCBF5 almost completely impaired the aCBF5 activity on tRNA (Figure 7A1) and this is also likely to be explained by the lower affinity of aCBF5 for the tRNA (Figure 7A2). However, this activity could be restored in the presence of aNOP10 and aGAR1 (Figure 7A3) and we showed that protein aNOP10 restored an efficient association of aCBF5 with the tRNA (Figure 7A2). Interestingly, the H77A mutation had a very limited effect on the Pab91 sRNP activity (Figure 7B1) consistent with an efficient recruitment of the RNA substrate observed for this variant (Figure 7B2). Hence, we concluded that residue H77 does not play an essential role for the aCBF5 RNA:Ψ-synthase activity, provided that the aNOP10 and aGAR1 proteins are present.

DISCUSSION

The present data shed light on common and specific properties of aCBF5 and TruB. They also bring a comparative analysis of the sRNA-guided and non-guided activities of aCBF5. Finally, they delineate the role of aNOP10 in the aCBF5 non-guided activity.

The major role of the PUA domain in tRNA recognition by aCBF5

Interestingly, we found that aCBF5 does not modify the Δa.c.-$tRNA^{Asp}$ and Δacceptor-$tRNA^{Asp}$ but can act on the ΔD-$tRNA^{Asp}$. Therefore, like for TruB, the 3D structure of tRNA is not required for recognition of the targeted U residue in tRNAs by aCBF5. However, in addition to the T stem-loop, both the anticodon stem-loop and acceptor stem contain determinants which are necessary for aCBF5 activity on tRNAs. By these extended structural requirements, aCBF5 differs from TruB, that is able to use the T stem-loop alone as a substrate (46). This different behavior may be due to variations in both the catalytic and PUA domains of aCBF5 compared to the ones of TruB. The importance of the PUA domain for the aCBF5 tRNA:Ψ55-synthase activity is demonstrated by abolition of this activity upon PUA truncation. Previous data also revealed the strong importance of the PUA domain for H/ACA sRNP activity (27). Consistently, in the H/ACA sRNP structure, numerous protein contacts take place with the minor groove of the P1 stem and the conserved ACA motif (30). Although no 3D structure was established for an H/ACA sRNP bound to its substrate, recent NMR data (47,48) suggest that the heterologous helix P1S, which is formed by interaction of the 3′ strand of the pseudouridylation pocket and the target RNA, is coaxially stacked on helix P1 of the sRNA and altogether these two helices contain a rather conserved number of base pairs (10–12).

or variants H31A (filled triangle), Y14A (filled square), P32A (filled inverted triangle) of aNOP10 protein in the tRNA-ΔCCA. Experiments were performed as in Figure 1A4. (**E**) Analysis by EMSA of the complexes formed with the radiolabeled $tRNA^{Asp}$, the aCBF5 protein with or without the WT or variant Y14A, H31A and P32A aNOP10 proteins. (**F**) Time-course analysis of the RNA-guided Ψ formation in an RNA target. The previously described Pab91 sRNA target (36) was radiolabeled during transcription by incorporation of [α-^{32}P]CTP. It was incubated at 65°C with the unlabeled *P. abyssi* Pab91 sRNA, and the three *P. abyssi* proteins L7Ae (L), aCBF5 (C) and aGAR1 (G) with the WT (N) (filled circle) or variant aNOP10 proteins (indicated as in panel D). After T2 RNase digestion, the amount of Ψ formation was estimated by 1D-TLC analysis. (**G**) Analysis by EMSA of the amount of the RNP5 and CII complexes (36), formed with the radio-labeled Pab91 sRNA (50 fmol), proteins L7Ae, aCBF5 and either the WT or variants aNOP10. The formation of complex CII was obtained by incubation of the guide sRNA with the protein mixture and a 50 molar excess of an unlabeled RNA target, which carries a *P. abyssi* 23S rRNA fragment complementary to the Pab91 pseudouridylation pocket.

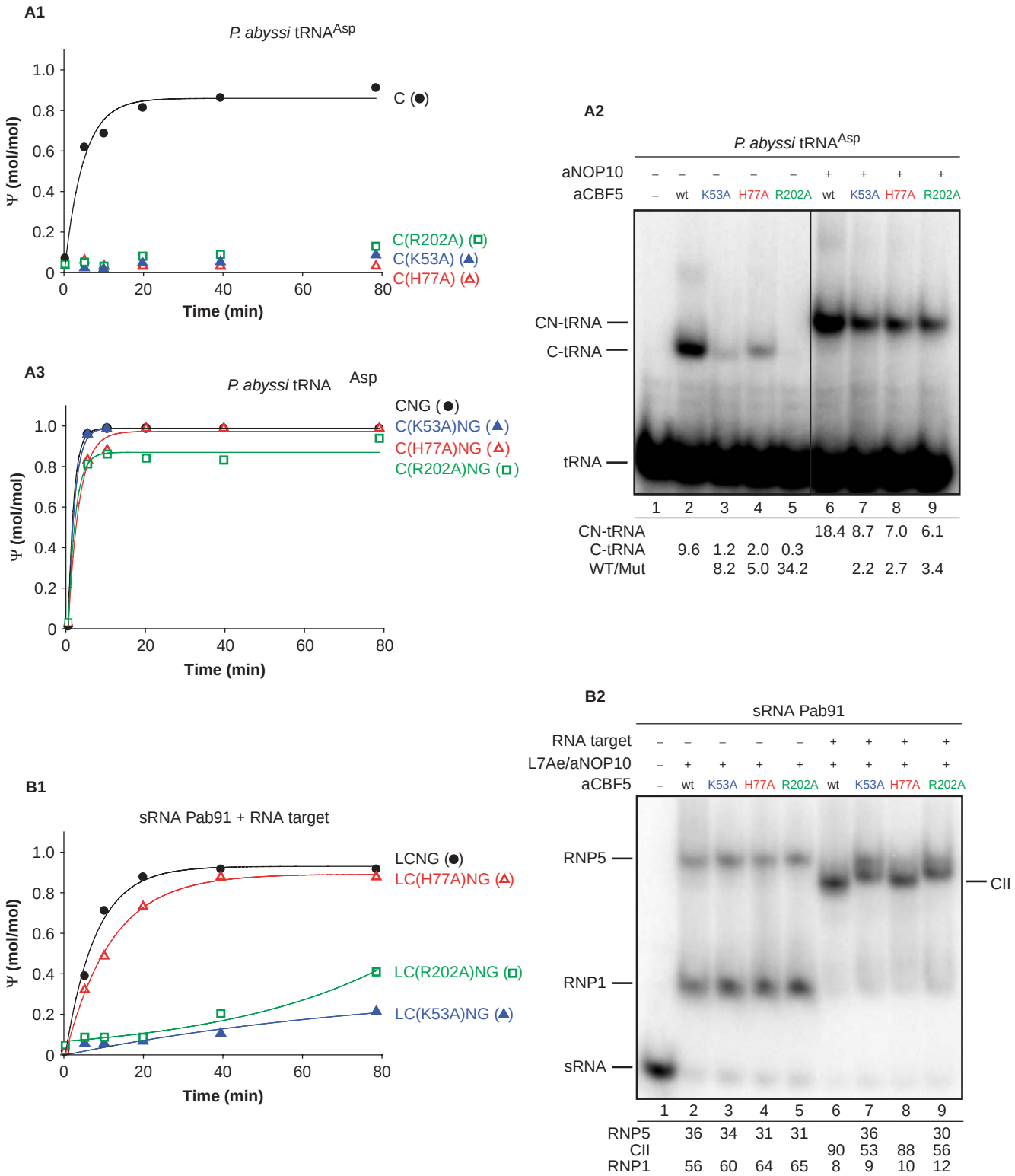


Figure 7. Effect of the substitution of residues in the aCBF5 catalytic site on its non-RNA-guided and RNA-guided activities. (**A1**) Time-course analysis of the non-RNA-guided Ψ55 formation in $tRNA^{Asp}$ by the wild type (C) (filled circle), or K53A, H77A and R202A variant aCBF5 proteins (filled triangle, open triangle, open square, respectively). Analysis was performed as in figure 1. (**A2**) Analysis by EMSA of the complexes formed with the radiolabeled $tRNA^{Asp}$, the WT (C) or variant aCBF5 proteins in the presence or absence of aNOP10 (N). The percentages of bound tRNA in the presence of WT or mutated aCBF5 and in the absence (C–tRNA) or presence of aNOP10 (CN–tRNA) are given as well as the ratios between tRNA bound to the mutated aCBF5 proteins versus that bound to WT aCBF5. (**A3**) The same experiments as in panel A1 were performed but in the presence of the auxiliary proteins aNOP10 (N) and aGAR1 (G). (**B1**) Time-course analysis of an RNA-guided Ψ formation carried out as in Figure 6F with the wild type or variant K53A, H77A and R202A aCBF5 proteins. (**B2**) Analysis by EMSA of the RNP5 and CII complexes. The experiments were performed as in Figure 6G, except that the variant aCBF5 proteins are used.

The heterologous helix P2S, that is formed with the 5′ strand of the pseudouridylation pocket, is basically stacked with the upperstem of the sRNA (P2) (47,48). The similarity between the T stem-loop structure recognized by TruB and the P2S heterologous helix formed by the guide RNA and the RNA substrate was pointed out (30,31). The tRNA acceptor stem flanked by the CCA motif would correspond to helix P1 of the sRNA and its conserved ACA motif, respectively. Based on this assumption, the finding that helix P1S, but not P2S, is stacked on the P1 stem of the sRNA can explain the requirement of a flexible link between the T and acceptor stem for aCBF5 activity. On the basis of this model of interaction between aCBF5 and tRNAs, the CCA motif of the tRNA, referred to the PUA domain, likely has the same position as the conserved ACA motif of H/ACA sRNAs (30). Therefore, like in TruB, the PUA domain of aCBF5 may interact with both the T and acceptor stems and the CCA motif, which explains the requirement of the acceptor stem and the 3′ terminal CCA for activity of the free aCBF5 protein (Figures 3 and 4). It should be pointed out that although TruB is active on the T stem-loop alone, the solved 3D structures of the RNA free and RNA bound TruB revealed, that upon tRNA binding, the PUA domain in TruB moves towards the active site and also makes interactions with residues of the acceptor stem (25). In the H/ACA sRNP structure, the two conserved adenosines of the ACA motif developed both stacking interactions and hydrogen bounds with several residues of the aCBF5 PUA domain. In particular, the 5′ proximal A residue inserts into an RNA-protein hybrid pocket formed between ribose moieties of the RNA minor groove and protein residues (30). Pocket formation is induced by RNA binding. Our gel-shift assays demonstrate a lower affinity of aCBF5 for a tRNA with an authentic CCA 3′ terminal sequence compared to a mutated tRNA with an ACA 3′ terminal sequence. Hence, an important issue would be to know how the RNA–protein interaction takes place when the proximal A residue in ACA is replaced by a C residue. The lower affinity for a CCA sequence compared to an ACA sequence may explain the needed interaction with the anticodon stem-loop for full activity of the free aCBF5 enzyme. This interaction may mimic that of aCBF5 with helix P2 in the sRNP. Interestingly, ACA to CCA substitution in a sRNA decreased substantially the amounts of aCBF5–sRNA complexes (data not shown). As aCBF5 is expected to have evolved from a pre-existing free standing enzyme, it would be interesting to investigate by 3D structure analysis of an aCBF5–tRNA complex, how the PUA domain of the pre-existing enzyme evolved in order to recognize both tRNAs and sRNAs.

It should also be pointed out that another possible explanation for the absence of activity on tRNAs of an aCBF5 without PUA domain may be the low thermostability of the truncated protein. Indeed, in contrast to the full-length aCBF5, the truncated aCBF5 cannot be enriched in the *E. coli* cellular extract by heating at 65°C. It precipitated with the *E. coli* proteins. Therefore, the PUA domain of aCBF5, which is wrapped by the N-terminal extremity of the molecule, likely plays an important role for aCBF5 thermostability.

Proteins aNOP10 and aGAR1 but not L7Ae reinforce the activity of aCBF5 on tRNAs

Whereas the individual addition of aNOP10 or aGAR1 stimulates the kinetics of aCBF5 activity on tRNA (Figure 1), no stimulatory effect was detected in the presence of L7Ae (Figure S1). A reasonable explanation is the absence of stable interaction between aCBF5 and L7Ae, while both aNOP10 and aGAR1 directly interact with this protein, as well in the presence, as in the absence of sRNA (35,36). Indeed, a direct interaction is formed between aNOP10 and L7Ae in the H/ACA sRNP structure (30), but this interaction is not stable by itself, it depends upon the binding of L7Ae on the K-turn or K-loop structure of the H/ACA sRNA. Based on these observations, it is reasonable to propose that aCBF5 can either act alone on tRNAs or in association with aNOP10 and aGAR1 *in vivo*. *In vitro*, the stimulatory effects of both aNOP10 and aGAR1 are particularly strong for tRNAs lacking the CCA sequence (Figure 3) and each of these proteins can compensate for the absence of this sequence. Therefore, it might be that in archaeal cells, the aCBF5-aNOP10, aCBF5-aGAR1 or aCBF5–aNOP10–aGAR1 complexes would modify archaeal tRNA precursors, before activity of the CCA-terminal transferase (49).

Our gel-shift assays strongly suggest that the activation property of aNOP10 on aCBF5 results in part from its capability to increase tRNA binding to aCBF5. In the 3D structure of an H/ACA sRNP, aNOP10 only forms a few contacts with the guide sRNA (30). However, aNOP10 is required for an efficient association of the substrate RNA on the complex formed by the guide sRNA and aCBF5 (36). Similarly, the tight interaction that aNOP10 established with aCBF5 (27–30) probably favors the tRNA recruitment (Figure 5). This aNOP10 activity together with the peculiarity of the aCBF5 PUA domain may compensate for the absence of the Ins1 and Ins2 segments in the aCBF5 catalytic domain as compared to the TruB catalytic domain. The Ins1 and Ins2 segments play a crucial role for TruB activity, since Ins2 forms the floor of the TΨC loop binding cavity and the Ins1 a binding thumb clamping this loop in the cavity (31). The disappearance of these two segments in aCBF5 is likely required to accommodate in the active site the sRNP four-helix-bundle structure formed upon annealing of a target RNA to the sRNA pseudouridylation pocket. The lack of Ins1 and Ins2 segments in aCBF5 may explain the strong requirement for the CCA 3′'-terminal motif, the acceptor stem and anticodon stem-loop for full activity of the free aCBF5 protein on tRNAs. It also likely explains the enhancer effect on activity of aNOP10, which increases the tRNA affinity. Up to now, the role of aGAR1 in the sRNA-guided system is less understood. GAR1 interacts with the RNA substrate in eukaryal snoRNP (50). However, this interaction seems to depend upon an additional domain present in the eukaryal protein as

compared to the archaeal protein. The protein aGAR1 also develops strong contact with the aCBF5 catalytic domain (29,30). Nonetheless, this may not help to reinforce the capability of aCBF5 to bind tRNA. In addition, aGAR1 was proposed to increase the turnover of the H/ACA sRNP (36). Therefore, this activity may also be needed to reinforce the aCBF5 tRNA:Ψ55-synthase activity.

Different networks of interaction are involved in the guided and non-guided activities of aCBF5

As mentioned above, the limited number of contacts between aNOP10 and the RNA in the H/ACA sRNP (30) suggests that the observed reinforcement of tRNA association in the presence of aNOP10 largely results from subtle conformational changes in the aCBF5 catalytic domain upon aNOP10–aCBF5 interaction. Superimposition of the 3D structures of the free and RNA-bound TruB enzyme (51) with the 3D structure of aCBF5 within the aCBF5–aNOP10 heterodimer shows a comparable positioning of the catalytic residues (27–30). In the absence of 3D structure of the free aCBF5 protein, we assume that in the free aCBF5 enzyme, an induced fit process has to take place upon tRNA binding for the optimization of the active site conformation. Interestingly, our data show that distinct residues of aNOP10 are involved in activation of the RNA-guided and non-guided activities of aCBF5 (Figure 6). Two conserved residues Y14 and H31 of aNOP10 were proposed to act in a common network of interactions linking aNOP10 to residue R202 in the aCBF5 catalytic site (29). Our data strongly suggest that the correct positioning of residue R202 in aCBF5 within the H/ACA sRNP involves residue H31 but not residue Y14, whereas the correct positioning of residue R202 for efficient tRNA:Ψ55-synthase activity depends upon residue Y14. Hence, conservation of these two residues in the course of evolution may reflect the conservation of the two enzymatic activities of aCBF5. Our data demonstrate a functional role in both the tRNA:Ψ55-synthase activity and the sRNA-guided activities of the conserved K residue in the aCBF5 motif I (K53 in aCBF5 and its counterpart K19 in *E. coli* TruB) (Figure 7). This residue is involved in the binding of the tRNA substrate and in the recruitment of the target RNA of the sRNP. Our data also show that the conserved P32 residue in aNOP10, which was proposed to link aNOP10 to the aCBF5 motif I (28), is essential, neither for the tRNA:Ψ55-synthase activity nor for the sRNA-guided activity. Another illustration of the difference between the aCBF5 RNA-guided and non-guided activities is the capability of aNOP10 to compensate for the absence of the lateral chains of residues K53 and R202 in the aCBF5 tRNA:Ψ55-synthase activity, but not in the sRNA-guided aCBF5 activity (Figure 7A3 and B1). Similarly, the substitution of residue K19 only has a mild effect on the TruB activity and leads to a thermosensitivity of the enzyme (52). Interestingly, the mutation of the conserved proline in motif I in the yeast Cbf5p also leads to a thermosensitivity of this enzyme (52,53). In light of these data, the positive effect of aNOP10 on the K53A aCBF5 variant may consist in a structural stabilization of motif I in the aCBF5–aNOP10 complex, allowing the reaction to proceed at the 65°C incubation temperature.

Residue R202 is specifically conserved in members of the TruB family. Its lateral chain is expected to interact with the 5′ phosphate of the targeted U residue (31). In agreement with these findings, we found that the R202A substitution affects tRNA association with aCBF5 and the binding of the target RNA on H/ACA sRNPs (Figure 7). Here again, protein aNOP10 can compensate for the R202A substitution. Another residue in the TruB catalytic center (Y76 from motif II) is expected to interact with the targeted U residue and is proposed to act as a general base during the catalytic mechanism (54). Other residues in aCBF5, like Y76 and/or Y179 which are expected to be located in the vicinity of the targeted U residue, can participate to the stabilization of the targeted residue. Interaction of aNOP10 with aCBF5 may favor their activity by subtle conformational changes in the aCBF5 active site.

Conservation of the crucial flipping mechanism in aCBF5

Upon tRNA binding on TruB, the uracil base at position 55 in the TΨC loop is extruded from the tRNA structure by a base-flipping mechanism that opens the tRNA 3D structure. Residue H43 probably plays a crucial role in this activity (31). Our observation that the mutation of residue H77, the counterpart of H43 in aCBF5, abolishes its tRNA:Ψ55-synthase activity strongly suggests that a similar mechanism is used for tRNA modification by aCBF5. Interestingly, aNOP10 can compensate for the H77A substitution. It would be highly interesting to determine by structural analysis how residue U55 is flipped out in the aCBF5 H77A variant associated with aNOP10. We also found that the RNA-guided activity is not impaired by the H77A substitution and this may be due to the fact that extrusion of the targeted U residue in H/ACA sRNPs is mainly mediated by the interaction between the guide and the target RNAs (47,48).

In spite of the absence of requirement of residue H77 for H/ACA sRNP activities, it should be noted that this residue is present in the eukaryal orthologs of aCBF5. This raises the question of a possible activity of the eukaryal Cbf5p/Dyskerin enzyme at position 55 in some eukaryal tRNAs. In connection with this observation, it should be pointed out that Cbf5 was found to co-purify with the selenocysteine tRNA in the protist *Euglena gracilis* (55).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

I. Motorin is thanked for helpful discussion. S.M. and J.–B.F. are fellows from the French Ministère de la Recherche et des Nouvelles Technologies (MNRT).

The work was supported by the Centre National de la Recherche Scientifique (CNRS), the MNRT and the PRST « Bioingénierie » of Région Lorraine. Funding to pay the Open Access publication charges for this article was provided by CNRS.

Conflict of interest statement. None declared.

REFERENCES

1. Rozenski,J., Crain,P.F. and McCloskey,J.A. (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res.*, **27**, 196–197.
2. Decatur,W.A. and Fournier,M.J. (2002) rRNA modifications and ribosome function. *Trends Biochem. Sci.*, **27**, 344–351.
3. King,T.H., Liu,B., McCully,R.R. and Fournier,M.J. (2003) Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol. Cell*, **11**, 425–435.
4. Sumita,M., Desaulniers,J.P., Chang,Y.C., Chui,H.M., Clos,L. II and Chow,C.S. (2005) Effects of nucleotide substitution and modification on the stability and structure of helix 69 from 28S rRNA. *RNA*, **11**, 1420–1429.
5. Badis,G., Fromont-Racine,M. and Jacquier,A. (2003) A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *RNA*, **9**, 771–779.
6. Meroueh,M., Grohar,P.J., Qiu,J., SantaLucia,J.Jr, Scaringe,S.A. and Chow,C.S. (2000) Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of *Escherichia coli* 23S rRNA. *Nucleic Acids Res.*, **28**, 2075–2083.
7. Kirpekar,F., Hansen,L.H., Rasmussen,A., Poehlsgaard,J. and Vester,B. (2005) The archaeon *Haloarcula marismortui* has few modifications in the central parts of its 23S ribosomal RNA. *J. Mol. Biol.*, **348**, 563–573.
8. Del Campo,M., Recinos,C., Yanez,G., Pomerantz,S.C., Guymon,R., Crain,P.F., McCloskey,J.A. and Ofengand,J. (2005) Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of Haloarcula marismortui and *Deinococcus radiodurans*. *RNA*, **11**, 210–219.
9. Yu,Y.T., Shu,M.D. and Steitz,J.A. (1998) Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J.*, **17**, 5783–5795.
10. Yang,C., McPheeters,D.S. and Yu,Y.T. (2005) Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **280**, 6655–6662.
11. Hamma,T. and Ferre-D'Amare,A.R. (2006) Pseudouridine synthases. *Chem. Biol.*, **13**, 1125–1135.
12. Kammen,H.O., Marvel,C.C., Hardy,L. and Penhoet,E.E. (1988) Purification, structure, and properties of *Escherichia coli* tRNA pseudouridine synthase I. *J. Biol. Chem.*, **263**, 2255–2263.
13. Nurse,K., Wrzesinski,J., Bakin,A., Lane,B.G. and Ofengand,J. (1995) Purification, cloning, and properties of the tRNA Psi 55 synthase from *Escherichia coli*. *RNA*, **1**, 102–112.
14. Wrzesinski,J., Bakin,A., Nurse,K., Lane,B.G. and Ofengand,J. (1995) Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from *Escherichia coli*. *Biochemistry*, **34**, 8904–8913.
15. Wrzesinski,J., Nurse,K., Bakin,A., Lane,B.G. and Ofengand,J. (1995) A dual-specificity pseudouridine synthase: An *Escherichia coli* synthase purified and cloned on the basis of its specificity for Ψ746 in 23S RNA is also specific for Ψ32 in $tRNA^{phe}$. *RNA*, **1**, 437–448.
16. Koonin,E.V. (1996) Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res.*, **24**, 2411–2415.
17. Kaya,Y. and Ofengand,J. (2003) A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA*, **9**, 711–721.
18. Behm-Ansmant,I., Urban,A., Ma,X., Yu,Y.-T., Motorin,Y. and Branlant,C. (2003) The *Saccharomyces cerevisiae* U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. *RNA*, **9**, 1371–1382.
19. del Campo,M., Ofengand,J. and Malhotra,A. (2004) Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of *Escherichia coli*. *RNA*, **10**, 231–239.
20. Kaya,Y., Del Campo,M., Ofengand,J. and Malhotra,A. (2004) Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *J. Biol. Chem.*, **279**, 18107–18110.
21. Huang,L., Pookanjanatavip,M., Gu,X. and Santi,D.V. (1998) A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalist. *Biochemistry*, **37**, 344–351.
22. Ramamurthy,V., Swann,S.L., Paulson,J.L., Spedaliere,C.J. and Mueller,E.G. (1999) Critical aspartic acid residues in pseudouridine synthases. *J. Biol. Chem.*, **274**, 22225–22230.
23. Conrad,J., Niu,L., Rudd,K., Lane,B.G. and Ofengand,J. (1999) 16S ribosomal RNA pseudouridine synthase RsuA of *Escherichia coli*: deletion, mutation of the conserved Asp102 residue, and sequence comparison among all other pseudouridine synthases. *RNA*, **5**, 761–763.
24. Raychaudhuri,S., Niu,L., Conrad,J., Lane,B.G. and Ofengand,J. (1999) Functional effect of deletion and mutation of the *Escherichia coli* ribosomal RNA and tRNA pseudouridine synthase RluA. *J. Biol. Chem.*, **274**, 18880–18886.
25. Ferre-D'Amare,A.R. (2003) RNA-modifying enzymes. *Curr. Opin. Struct. Biol.*, **13**, 49–55.
26. Aravind,L. and Koonin,E.V. (1999) Novel predicted RNA-binding domains associated with the translation machinery. *J. Mol. Evol.*, **48**, 291–302.
27. Manival,X., Charron,C., Fourmann,J.B., Godard,F., Charpentier,B. and Branlant,C. (2006) Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res.*, **34**, 826–839.
28. Hamma,T., Reichow,S.L., Varani,G. and Ferre-D'Amare,A.R. (2005) The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat. Struct. Mol. Biol.*, **12**, 1101–1107.
29. Rashid,R., Liang,B., Baker,D.L., Youssef,O.A., He,Y., Phipps,K., Terns,R.M., Terns,M.P. and Li,H. (2006) Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol. Cell*, **21**, 249–260.
30. Li,L. and Ye,K. (2006) Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature*, **443**, 302–307.
31. Hoang,C. and Ferre-D'Amare,A.R. (2001) Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell*, **107**, 929–939.
32. Bachellerie,J.P., Cavaille,J. and Huttenhofer,A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
33. Kiss,T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
34. Omer,A.D., Ziesche,S., Decatur,W.A., Fournier,M.J. and Dennis,P.P. (2003) RNA-modifying machines in archaea. *Mol. Microbiol.*, **48**, 617–629.
35. Baker,D.L., Youssef,O.A., Chastkofsky,M.I., Dy,D.A., Terns,R.M. and Terns,M.P. (2005) RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev.*, **19**, 1238–1248.
36. Charpentier,B., Muller,S. and Branlant,C. (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res.*, **33**, 3133–3144.
37. Khanna,M., Wu,H., Johansson,C., Caizergues-Ferrer,M. and Feigon,J. (2006) Structural study of the H/ACA snoRNP components Nop10p and the 3′ hairpin of U65 snoRNA. *RNA*, **12**, 40–52.
38. Rozhdestvensky,T.S., Tang,T.H., Tchirkova,I.V., Brosius,J., Bachellerie,J.P. and Huttenhofer,A. (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif

for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.*, **31**, 869–877.
39. Hamma,T. and Ferre-D'Amare,A.R. (2004) Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution. *Structure*, **12**, 893–903.
40. Roovers,M., Hale,C., Tricot,C., Terns,M.P., Terns,R.M., Grosjean,H. and Droogmans,L. (2006) Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic Acids Res.*, **34**, 4293–4301.
41. Charpentier,B., Fourmann,J.B. and Branlant,C. (2007) Reconstitution of archaeal H/ACA sRNPs and test of their activity. *Methods Enzymol.*, **425**, 389–405.
42. Bakin,A. and Ofengand,J. (1993) Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center - analysis by the application of a new sequencing technique. *Biochemistry*, **32**, 9754–9762.
43. Massenet,S., Ansmant,I., Motorin,Y. and Branlant,C. (1999) The first determination of pseudouridine residues in 23S ribosomal RNA from hyperthermophilic Archaea *Sulfolobus acidocaldarius*. *FEBS Lett.*, **462**, 94–100.
44. Mougin,A., Gregoire,A., Banroques,J., Segault,V., Fournier,R., Brule,F., Chevrier-Miller,M. and Branlant,C. (1996) Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *RNA*, **2**, 1079–1093.
45. Motorin,Y., Muller,S., Behm-Ansmant,I. and Branlant,C. (2007) Identification of modified residues in RNAs by reverse transcription-based methods. *Methods Enzymol.*, **425**, 21–53.
46. Gu,X., Yu,M., Ivanetich,K.M. and Santi,D.V. (1998) Molecular recognition of tRNA by tRNA pseudouridine 55 synthase. *Biochemistry*, **37**, 339–343.
47. Jin,H., Loria,J.P. and Moore,P.B. (2007) Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Mol. Cell*, **26**, 205–215.
48. Wu,H. and Feigon,J. (2007) H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *Proc. Natl Acad. Sci. USA*, **104**, 6655–6660.
49. Tomita,K., Ishitani,R., Fukai,S. and Nureki,O. (2006) Complete crystallographic analysis of the dynamics of CCA sequence addition. *Nature*, **443**, 956–960.
50. Wang,C. and Meier,U.T. (2004) Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.*, **23**, 1857–1867.
51. Pan,H., Agarwalla,S., Moustakas,D.T., Finer-Moore,J. and Stroud,R.M. (2003) Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proc. Natl Acad. Sci. USA*, **100**, 12648–12653.
52. Spedaliere,C.J., Hamilton,C.S. and Mueller,E.G. (2000) Functional importance of motif I of pseudouridine synthases: mutagenesis of aligned lysine and proline residues. *Biochemistry*, **39**, 9459–9465.
53. Zebarjadian,Y., King,T., Fournier,M.J., Clarke,L. and Carbon,J. (1999) Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Mol. Cell. Biol.*, **19**, 7461–7472.
54. Phannachet,K., Elias,Y. and Huang,R.H. (2005) Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase. *Biochemistry*, **44**, 15488–15494.
55. Russell,A.G., Schnare,M.N. and Gray,M.W. (2004) Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis*. *RNA*, **10**, 1034–1046.

Supplementary data

Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Ψ55-synthase and RNA-guided RNA:Ψ-synthase activities

S. Muller, J.-B. Fourmann, C. Loegler, B. Charpentier, and C. Branlant

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 UHP-CNRS, Université des Sciences et Techniques Henri Poincaré Nancy I, 54506 Vandoeuvre-Lès-Nancy cedex, France.

Figure S1: Analysis of the influence of protein L7Ae on Ψ55 formation in *P. abyssi* $tRNA^{Asp}$. The tRNAs were labeled by incorporation of [α-^{32}PCTP] during transcription. The ^{32}P-labeled tRNA substrate (50 fmol) was mixed with aCBF5 (C), aCBF5-aNOP10 (CN), or the aCBF5-aNOP10-aGAR1 complex (CNG) in the absence or the presence of protein L7Ae (L) (LCN and LCNG). Incubation was performed for 80 min at 65°C. After T2 RNase digestion, the amount of Ψ residue formed was measured by 2D TLC and expressed in moles per moles of tRNA.

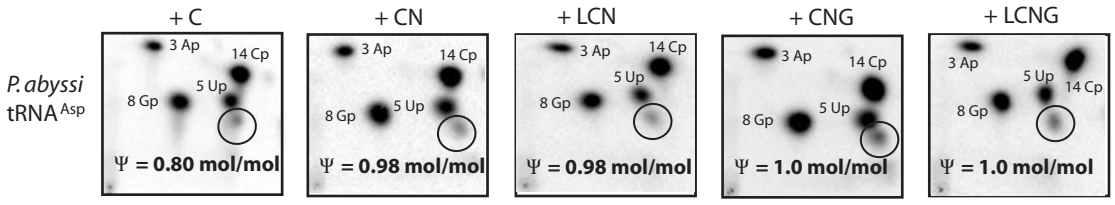


Figure S1

Muller et al.

2. Discussion complémentaire

2.1. La présence de la séquence CCA

Nos données montrent que l'association de aCBF5 avec aNOP10 et aGAR1 permet de pallier l'absence du triplet CCA dans l'ARNt et permet une action efficace de aCBF5 sur des ARNt tronqués comme un demi-ARNt restreint au bras accepteur et au bras TΨC. Ces observations sont intéressantes et renforcent la possibilité de modification d'ARN autres que des ARNt par l'activité de aCBF5, non guidée par un ARN, mais aidée par les protéines aNOP10 et aGAR1. Notre observation de la modification à la position U2603 de l'ARNr 23S de *P. abyssi* par aCBF5 en absence de guide en est un premier exemple (article n°1, (Muller *et al.*, 2008)). Etant donné que nous n'avons pas identifié de séquence CCA ou ACA dans la séquence de l'ARNr 23S au voisinage du résidu U2603, il est probable que les protéines aNOP10 et/ou aGAR1 jouent un rôle majeur pour l'activité de modification. Cette possibilité pourra être testée facilement *in vitro*.

2.2. Le rôle des protéines aNOP10 et aGAR1

Nos données font apparaître un effet positif de la présence de aNOP10 et aGAR1 sur la fixation de aCBF5 sur les ARNt. Dans le cas de aNOP10, nous avons observé que :

- aNOP10 renforce le taux de fixation de aCBF5 aux ARNt, en particulier si ces derniers sont dépourvus de leur séquence CCA 3' terminale.

- sur certains ARNt, nous avons observé la fixation de 2 molécules de aCBF5, la seconde molécule de aCBF5 pouvant être recrutée par le biais d'une interaction non spécifique. La présence de aNOP10 conduit à une stœchiométrie de 1 dimère aCBF5-aNOP10 par ARNt.

Il est proposé que l'interaction de aNOP10 avec aCBF5 concourt à la structuration du site actif de aCBF5 (Hamma *et al.*, 2005). Ainsi, cet évènement favoriserait les interactions spécifiques du domaine catalytique avec l'ARNt à modifier, au détriment d'interactions aspécifiques. D'autre part, il n'est pas exclu que

aNOP10 interagisse aussi avec l'ARN dans les complexes ARN-aCBF5-aNOP10. Cette interaction peut aussi conduire à favoriser la formation des complexes enzyme-substrat spécifiques.

Nous avons également observé un effet stimulateur de aGAR1, similaire à celui de aNOP10, sur l'activité de modification par aCBF5 d'un ARNt tronqué de la séquence CCA 3' terminale. Il apparaît que les deux effets sont cumulatifs. Les différentes structures 3D des sRNP indiquent que la protéine aGAR1 n'interagit pas directement avec l'ARN. Par contre, l'interaction de aGAR1 avec aCBF5 pourrait aussi induire un changement de conformation favorisant la fixation spécifique de aCBF5 à son substrat. En particulier, l'interaction de la boucle β7/β10 de aCBF5 avec aGAR1 pourrait maintenir le site actif de aCBF5 dans une conformation ouverte favorisant à la fois l'entrée et la sortie du substrat.

Néanmoins, le rôle respectif joué par chacune des protéines aNOP10 et aGAR1 pourrait être variable selon les espèces d'archaea. En effet, chez *M. jannashii*, l'effet positif de aNOP10 sur l'activité de aCBF5 est nettement moins marqué que chez *P. abyssi* alors qu'à l'inverse, aGAR1 semble être bien plus prépondérante à l'activité de aCBF5 chez *M. jannashii* que chez *P. abyssi* (Gurha et al., 2007).

2.3. Le rôle de la position 54 et de aPUS10

Comme nous le décrivons dans l'article n°3, la position 55 des ARNt peut être modifiée *in vitro* par une ARN:Ψ-synthase qui est associée aux ARN guides dans les sRNP à boîtes H/ACA (Roovers *et al.*, 2006; Gurha *et al.*, 2007; Muller *et al.*, 2007b). Chez les archaea, aCBF5 peut jouer un rôle équivalent à TruB/Pus4p et fonctionner en tant qu'enzyme protéique classique. Cependant, il semble que la paire de base reverse Hoogsteen U54 : A58, qui est reconnue par TruB, ne soit pas nécessaire pour la modification de la position U55 par aCBF5 (Gurha et al., 2007).

Il est intéressant de noter qu'une seconde enzyme est également capable de convertir le résidu U55 en Ψ55, il s'agit de aPUS10 qui modifie également la position adjacente U54 (Roovers *et al.*, 2006; Gurha et Gupta, 2008). Alors pourquoi la cellule

possèderait-t-elle 2 enzymes qui ciblent la même position dans les ARNt ? Une hypothèse est que la modification du résidu U55 par aCBF5 soit conservée car la présence de cette enzyme doit être maintenue du fait de son rôle essentiel dans le système guidé (Roovers *et al.*, 2006; Gurha *et al.*, 2007; Muller *et al.*, 2007b). La possibilité d'effectuer une modification Ψ55 par aCBF5, même si elle est partiellement redondante avec celle de aPUS10, doit représenter un avantage pour les cellules car les résidus de aCBF5 que nous avons pu identifier comme importants pour l'activité non guidée ont été conservés.

Il existe cependant une différence entre les deux enzymes. Comme je l'ai déjà décrit, pour modifier efficacement l'ARNt, aCBF5 a besoin de la séquence CCA 3' terminale de l'ARNt et son substrat minimal correspond au bras accepteur et la tige-boucle TΨC (Roovers *et al.*, 2006; Gurha *et al.*, 2007; Muller *et al.*, 2007b). Par contre, la tige-boucle TΨC est suffisante pour que aPUS10 modifie la position U55 (Roovers *et al.*, 2006), propriété que PUS10 partage avec TruB.

C. Etude approfondie de la protéine aCBF5 : rôle de la boucle β7/β10 et de résidus du site actif dans le turn-over de la particule et dans la spécificité de substrat

1. Contexte et stratégie d'étude

Comme déjà mentionné précédemment, les analyses *in vitro* ont montré que le complexe aCBF5-aNOP10 était le couple minimal de protéines conduisant à une particule active et que la présence des protéines L7Ae et/ou aGAR1 augmentait fortement l'activité des sRNP (Charpentier *et al.*, 2005). Au début de cette étude et jusqu'à très récemment, le rôle exact joué par aGAR1 était inconnu. Une région structurale particulière prend en charge l'ARNt et qui est dénommé le pouce (ou thumb) dans l'ARNt:Ψ-synthase TruB, possède un équivalent raccourci chez aCBF5. Celui-ci correspond à la séquence s'organisant en une boucle comprise entre le brin β7 et le brin β10 (Fig. r16 A). Les structures 3D de TruB libre ou en complexe avec l'ARN qui étaient disponibles au début de cette étude indiquaient que i) la boucle du pouce possède différentes conformations, l'une dite « ouverte » et l'autre dite « fermée » et ii) le pouce dans sa conformation fermée établit des contacts avec le substrat (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003). De plus, la protéine aGAR1 interagit avec cette boucle de aCBF5 (Rashid *et al.*, 2006). Sur la base de ces observations, il a été proposé que le pouce de TruB serait un élément important pour le turn-over de l'enzyme et que dans le cas des sRNP, aGAR1 en ayant la capacité de modifier l'orientation du pouce de aCBF5 participerait donc à l'activité de l'ARN:Ψ-synthase (Hoang *et al.*, 2005; Rashid *et al.*, 2006; Liang *et al.*, 2008). Au moment où nous avons débuté ce travail, aucune information sur la structure 3D d'une sRNP libre ou en interaction avec l'ARN substrat positionné favorablement pour le déroulement de la réaction enzymatique n'était disponible. Sur la base de l'analyse de la structure tridimensionnelle du complexe aCBF5-aNOP10-aGAR1 (Rashid *et al.*, 2006), nous avons émis l'hypothèse que les résidus R139, P140, P141, R143 de la boucle β7/β10 de aCBF5 pouvaient représenter des déterminants importants de la relation entre aGAR1 et aCBF5.

L'enzyme bactérienne TruB avait servi de modèle pour l'étude du site actif des ARN:Ψ-synthases (Huang *et al.*, 1998; Ramamurthy *et al.*, 1999a; Ramamurthy *et al.*, 1999b; Hamilton *et al.*, 2005; Phannachet *et al.*, 2005). Les premières structures 3D de l'enzyme TruB liée au bras TΨC de l'ARNt ont révélé la façon dont elle se fixait à l'ARNt et l'identité des acides aminés qui interagissaient avec le substrat et qui pourraient jouer un rôle dans la fixation du substrat et/ou dans la catalyse (Fig. r16 B) (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003). Cependant, la substitution de certains résidus chez TruB conduit à une protéine très instable qui ne peut alors être produite en abondance et être étudiée *in vitro* (Spedaliere *et al.*, 2000). De ce fait, l'étude de plusieurs résidus conservés du site actif n'avait pu être réalisée. Compte tenu de la forte stabilité des protéines d'archaea, nous avons pensé que de nouveaux variants d'ARN:Ψ-synthase pouvaient être produits puis étudiés à partir de aCBF5. Ainsi, aCBF5 pouvait représenter une enzyme modèle pour l'étude du mécanisme catalytique des ARN:Ψ-synthases (article n°3, (Muller *et al.*, 2007b)).

Ainsi, nous avons créé des variants de aCBF5 portant des substitutions de résidus de la boucle β7/β10 et du site actif et nous avons évalué l'effet des mutations sur l'assemblage des sRNP, la fixation du substrat et l'activité des particule. De plus, nous avons mis au point un test, incluant des étapes de purification sur gel des produits de la réaction, pour évaluer le turn-over des sRNP vis-à-vis d'un substrat correspondant à un fragment d'ARNr. L'activité des particules reconstituées avec les variants de aCBF5 a été analysée par ce test.

2. Choix des mutations introduites dans la séquence codant aCBF5 de *P. abyssi*

2.1. Les substitutions de résidus de la boucle β7/β10 de aCBF5

La boucle β7/β10 de aCBF5 correspond à la séquence comprise entre les positions 140 à 150 et qui représente la version raccourcie du pouce de TruB (Fig. r16 A). Nous avons éliminé la totalité de cette séquence et appelé ce variant de aCBF5

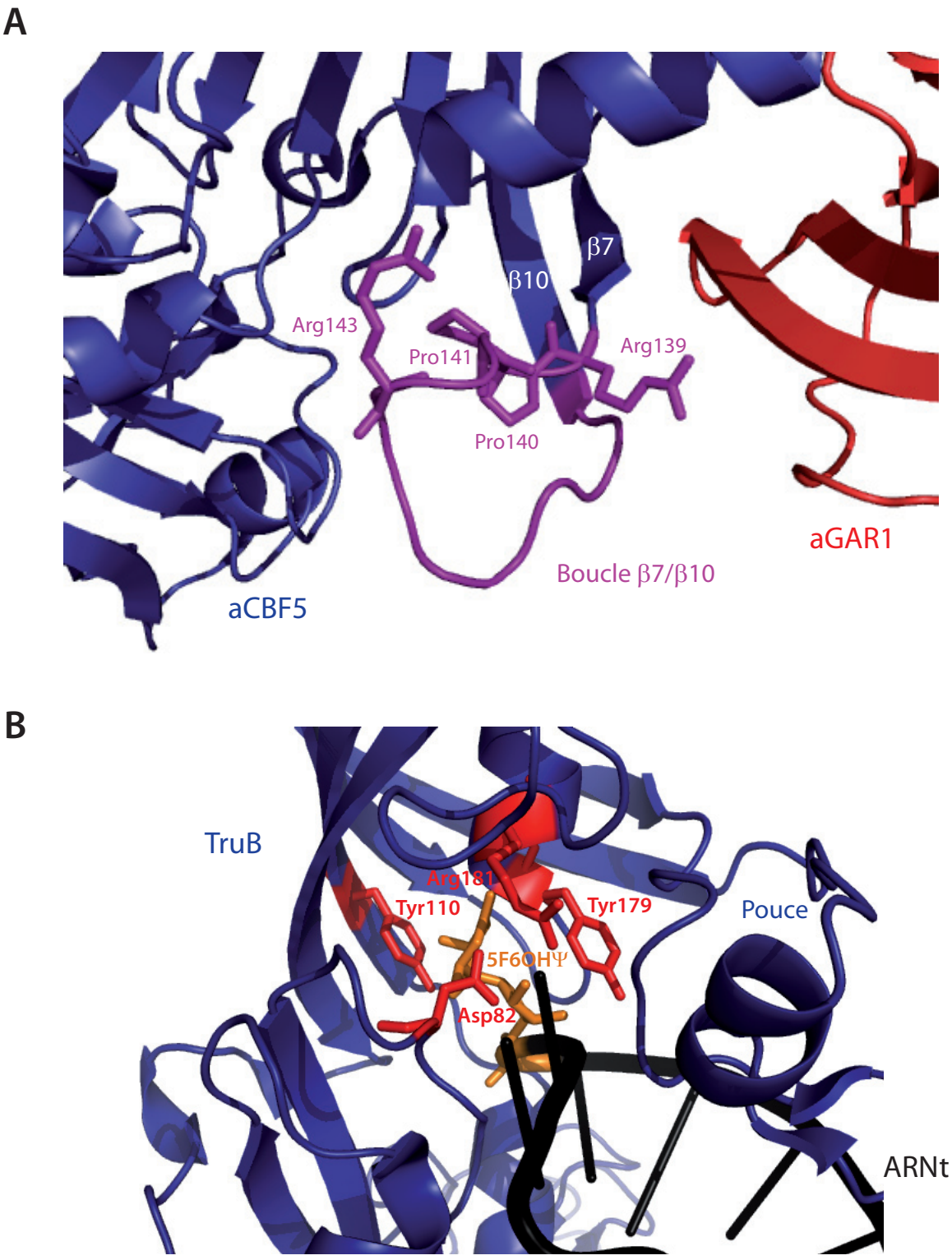


Figure r16 : Stuctures 3D des complexes protéiques aCBF5-aNOP10-aGAR1 libre et de TruB liée au bras TΨC de l'ARNt.

A. Vue d'une partie de la structure 3D du complexe aNOP10 (vert)-aCBF5 (bleu)-aGAR1 (rouge). La boucle β7/β10 de aCBF5 et les résidus R139, P140, P141 et R143 sont colorés en violet. (PDB=2EY4) (Rashid *et al.,* 2006).

B. Vue d'une partie de la structure 3D du complexe TruB (bleu)-boucle TΨC de l'ARNt (noir). Les résidus de TruB D82, Y110, Y179 et R181 sont colorés en rouge et correspondent à la numérotation de aCBF5 de *P. abyssi*. Le résidu U55 cible de la réaction est converti en 5-fluoro-6-hydroxy pseudouridine (5-F-6-OHΨ) lors de la formation du cristal et est représenté en orange. (PDB=1R3E) (Pan *et al.,* 2003).

Les vues 3D ont été réalisées grâce au logiciel Pymol à partir des coordonnée PDB mentionnées.

Δ140-150. Celui-ci a été employé dans le but d'observer la contribution de l'intégralité de cette boucle dans l'activité de particule.

L'arginine 139 est localisée juste en amont de la boucle β7/β10 de aCBF5. Ce résidu est bien conservé dans la famille TruB à l'exception de TruB d'*E. coli* qui possède une isoleucine à cette position (Fig. i19). Le résidu R139 est situé à l'interface entre aCBF5 et aGAR1 et participerait au positionnement du pouce de aCBF5 (Fig. r16 A) (Rashid *et al.*, 2006). Nous l'avons substitué en alanine pour tester le rôle que la chaîne latérale de ce résidu pouvait jouer par le biais de son interaction avec aGAR1.

Les deux prolines aux positions 140 et 141 sont positionnées à la base de la boucle β7/β10 de aCBF5 et sont très conservées dans la famille TruB à l'exception de la proline 141 qui est remplacée par une sérine dans la séquence de TruB d'*E. coli* (Fig. i19). Dans la structure contenant la boucle en position ouverte, ces 2 résidus réalisent des interactions hydrophobes avec des acides aminés de aGAR1 (Li et Ye, 2006). Par contre, dans la structure comportant la boucle en conformation fermée, ces interactions ne sont plus retrouvées (Rashid *et al.*, 2006). Ces 2 prolines contribueraient donc à l'orientation de la boucle β7/β10 (Fig. r16 A). Nous avons substitué ces 2 prolines par des alanines. La conformation de la boucle β7/β10 pouvait être modifiée par ces substitutions et conduire à une désorganisation des interactions formées avec aGAR1 lorsque cette boucle est en position ouverte.

La chaîne latérale de l'arginine 143 est orientée vers le site actif de aCBF5 lorsque la boucle est en position fermée (Fig. r16 A) (Rashid *et al.*, 2006). Cela suggère que la longue chaîne latérale de ce résidu pourrait établir des interactions avec le substrat lorsqu'il est présent dans le site actif de aCBF5. Ce résidu pourrait ainsi participer de façon directe au mécanisme réactionnel. Dans la famille TruB, le résidu R143 est conservé chez les archaea mais est remplacé par une isoleucine chez les eucaryotes ou par une tyrosine chez TruB d'*E. coli* (Fig. i19). Nous avons réalisé 3 substitutions pour ce résidu (Fig. r16 A). La substitution R143A avait pour but de modifier la taille et la charge de la chaine latérale de l'arginine. Le remplacement

R143I permettait de conserver une longue chaîne latérale, mais non chargée. Le variant R143E apporte une charge négative à la place de la charge positive de l'arginine, ainsi en cas d'interaction du résidu à la position 143 avec l'ARN, le glutamate devrait repousser l'ARN également chargé négativement.

2.2. Les substitutions de résidus du site actif de aCBF5

L'acide aspartique 82 correspond au résidu catalytique de l'enzyme aCBF5 (Fig. r16 B). Il est parfaitement conservé parmi tous les membres de la famille TruB et chez toutes les espèces (Fig. i19). Le variant D82A portant une substitution de ce résidu par une alanine a déjà été étudié au laboratoire (Charpentier *et al.*, 2005). Néanmoins il a servi de témoin dans cette étude.

Les tyrosines 110 et 179 sont des résidus parfaitement conservés dans le site actif des enzymes de la famille TruB (Fig. i19). Dans les structures 3D de TruB liée à l'ARNt, les chaînes latérales de ces deux résidus réalisent des interactions par empilement de base avec l'uridine cible de l'ARNt ce qui suggère un rôle important dans la reconnaissance du substrat ou dans le mécanisme réactionnel (Fig. r16 B) (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003). Ils ont également été proposés pour jouer un rôle de base générale dans le mécanisme catalytique (Hamilton *et al.*, 2005; Phannachet *et al.*, 2005). Pour chacune de ces deux tyrosines, nous avons réalisé 2 types de substitutions :

-soit par une alanine (variants Y110 A et Y179A) qui conduit à l'élimination de la chaîne latérale qui est composée d'un noyau phényle et de l'hydroxyle chargé négativement.

- soit par une phénylalanine (variants Y110F et Y179F) qui permet de conserver la partie phényle de la tyrosine tout en supprimant l'hydroxyle chargé négativement.

Dans la structure de TruB, l'arginine 181 réalise un pont salin avec l'aspartate catalytique et pourrait activer l'aspartate en tant que nucléophile en déprotonant le

groupement carboxyle (Fig. i14, ; r16 B) (Hoang et Ferre-D'Amare, 2001). Ce résidu est parfaitement conservé dans la famille TruB et chez toutes les espèces (Fig. i19). Dans la forme libre de TruB, ce résidu forme une seule liaison hydrogène avec l'aspartate catalytique D48, tandis que dans la forme de TruB liée à l'ARN, un pont salin est formé entre les résidus R181 et D48 (Pan *et al.*, 2003). Ce changement de conformation induit par l'interaction avec l'ARN substrat est important pour permettre le placement des groupements catalytiques dans un alignement optimal pour réaliser la réaction catalytique et pour refermer la cavité du site actif (Pan *et al.*, 2003). La perte du pont salin conduit nécessairement à un réarrangement structural pour compenser la charge partielle non satisfaite et la capacité de l'arginine 181 à former des liaisons hydrogènes. Il a été proposé que ce réarrangement pourrait facilement se propager au pouce (Hoang *et al.*, 2005). Nous avons substitué ce résidu par une alanine (variant R181A) pour déterminer son rôle dans la catalyse et/ou son lien fonctionnel avec le pouce de aCBF5.

J'ai cloné, produit puis caractérisé l'ensemble des variants de la boucle β7/β10 et du site actif de aCBF5 par l'emploi de plusieurs techniques biochimiques. Les résultats obtenus nous ont permis d'identifier des résidus qui sont cruciaux à une étape donnée de la réaction de pseudouridylation (Fig. i62) telle que celle du chargement d'un substrat spécifique au sein de la sRNP ou celle de la sortie efficace du substrat modifié de la sRNP. J'ai en particulier isolé un variant de la boucle de aCBF5 qui n'a pas besoin de la présence de aGAR1 pour conduire à une activité optimale de la sRNP reconstituée. J'ai également identifié un résidu du site actif de aCBF5 qui joue un rôle capital dans la sélection du substrat.

3. Résultats

3.1. Le résidu R143 est un déterminant important de l'interaction de aCBF5 avec aGAR1

La substitution d'un résidu dans une séquence protéique peut conduire à une déstabilisation de la protéine ou avoir un effet délétère sur l'interaction avec son

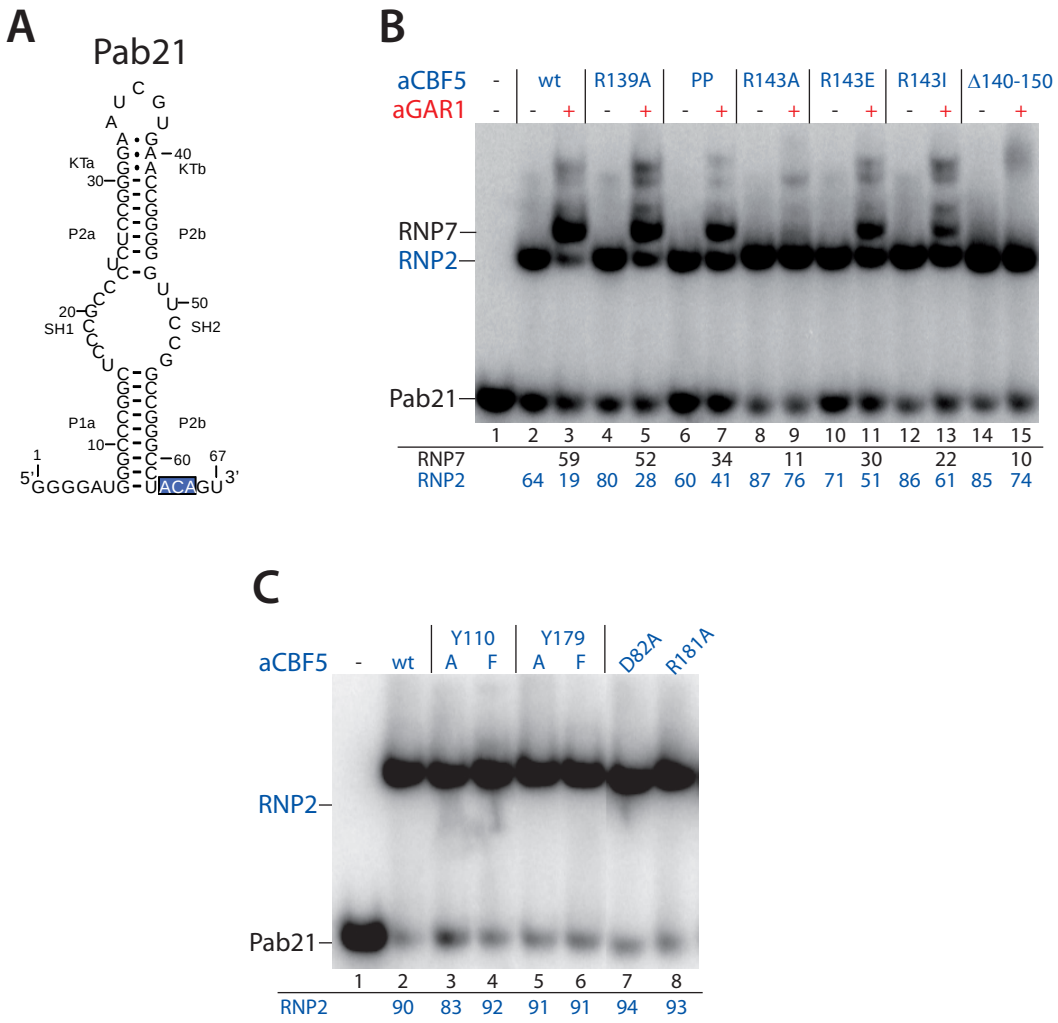


Figure r17 : Etude par expériences de retard de migration sur gel de la fixation directe sur le sRNA Pab21 de variants de aCBF5 et de l'hétérodimère aCBF5-aGAR1.

A. Pour rappel, structure schématique du sRNA Pab21 à boîtes H/ACA de *P. abyssi*.

B. Résultats d'expériences de retard de migration sur gel réalisées en présence du sRNA Pab21 marqué en 5', de variants de aCBF5 (bleu) et de aGAR1 (rouge). Le variant PP correspond à aCBF5 P140A, P141A. Le taux de complexe formé entre le sRNA et aCBF5 (RNP2) ou entre le sRNA et les protéines aCBF5 et aGAR1 (RNP7) est indiqué sous chaque piste.

C. Résultats d'expériences de retard de migration sur gel réalisées en présence du sRNA Pab21 marqué en 5' et des variants de aCBF5 (bleu). Le taux de complexe formé entre le sRNA et aCBF5 (RNP2) est indiqué sous chaque piste.

partenaire ARN ou protéine. J'ai donc effectué des expériences de retard de migration sur gel pour tester si les différentes substitutions affectaient ou non l'interaction entre aCBF5 et le sRNA Pab21 (Fig. r17). Nous constatons que chaque variant de aCBF5 conserve sa capacité à se fixer directement à Pab21 avec un taux de complexe RNP2 formé supérieur à 90%, qui est comparable à celui observé pour aCBF5 sauvage (Fig. r17 B pistes paires et C pistes 1 à 7). Nous concluons que les différentes mutations introduites dans aCBF5 n'affectent pas sa capacité de fixation au sRNA.

Ensuite, les variants du pouce de aCBF5 ont été analysés dans des expériences de retard de migration sur gel mais en présence de la protéine aGAR1 et du sRNA Pab21. Le but de cette expérience était de vérifier si les variants de aCBF5 avaient conservé leur capacité d'interaction avec aGAR1. Nous avons donc suivi la formation du complexe RNP7 qui correspond au couple de protéines aCBF5-aGAR1 lié au sRNA (Fig. r17 B). Des complexes RNP7 stables sont détectés pour tous les variants de aCBF5 à l'exception du variant R143A et de la forme tronquée Δ140-150 qui forment des complexes plus diffus et donc moins stables (Fig. r17 B, pistes 9 et 15). Nous concluons que le pouce est un élément important pour l'interaction de aCBF5 avec aGAR1 et que le résidu R143 est un déterminant majeur pour la stabilité de cette interaction.

3.2. Analyse de l'activité ARN:Ψ-synthase de sRNP assemblées avec des variants de la boucle β7/β10 de aCBF5

Dans ces expériences, nous avons mesuré le taux de formation des résidus Ψ en employant les différents variants de aCBF5 et les cinétiques obtenues sont représentées sur les graphiques de la figure r18. En présence du mélange de protéines L7Ae, aCBF5, aNOP10 et aGAR1 (LCNG), l'activité de pseudouridylation de la sRNP Pab91 est maximale avec aCBF5 sauvage puisque la majorité de l'ARN substrat est modifiée après 20 min d'incubation à 65°C (Fig. r18 A). En absence de

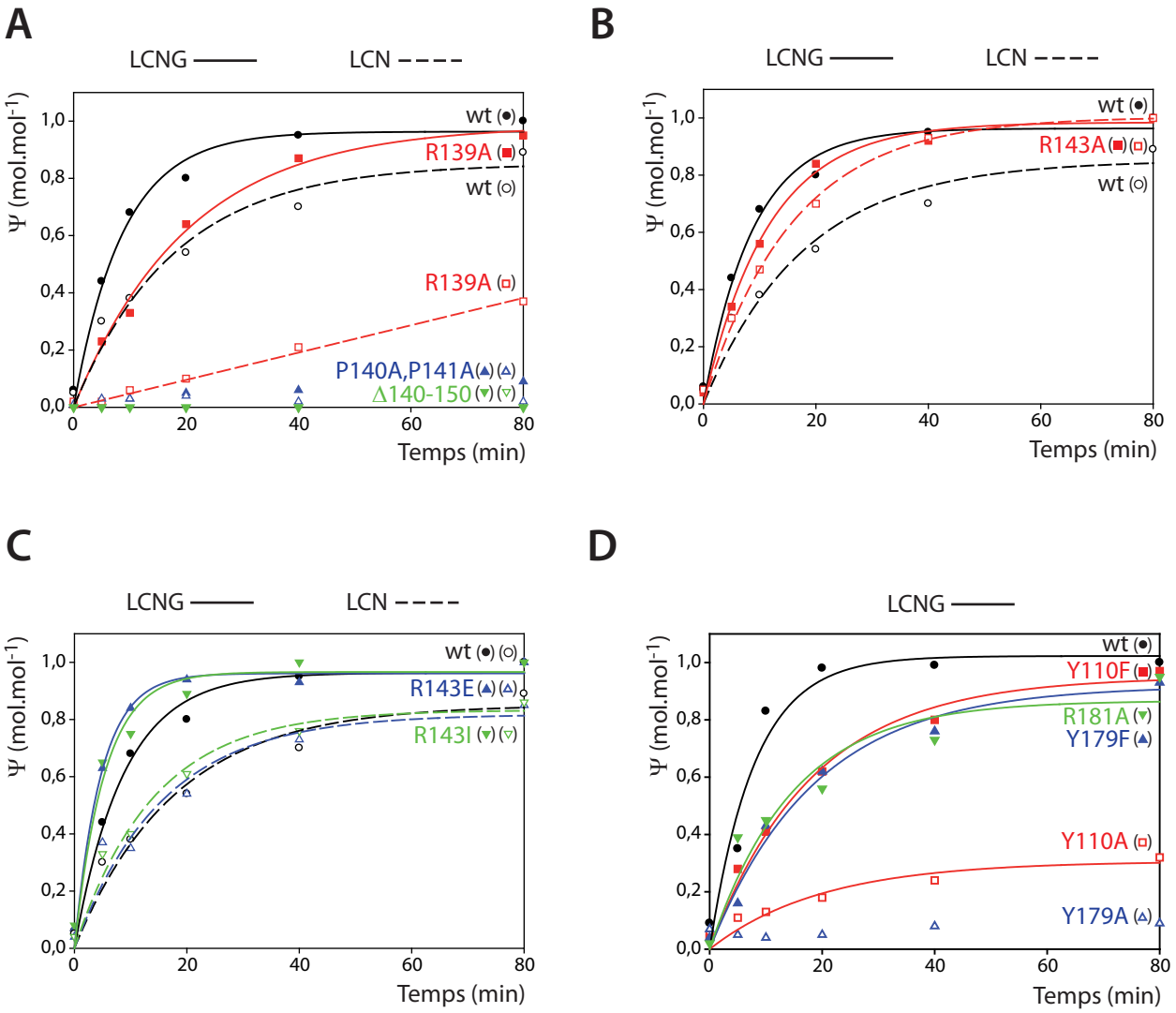


Figure r18 : Etude cinétique du taux de résidus Ψ formés en employant des variants de aCBF5. Les tests ont été réalisés en employant la combinaison de protéines soit L7Ae, aCBF5, aNOP10 et aGAR1 (LCNG, en trait plein), soit L7Ae, aCBF5 et aNOP10 (LCN, en trait pointillé). Les conditions expérimentales correspondent à une température de 65°C, à l'emploi du sRNA Pab91 et de son substrat RNAr d'une taille de 26 nucléotides qui correspond aux positions 2678 à 2700 de l'ARNr 23S . Les résultats obtenus avec aCBF5 sauvage sont représentés en noir.
A. à **C.** Etudes cinétiques des variants du pouce de aCBF5 R139A (rouge), P140A P141A (bleu), Δ140-150 (vert) (**A)** ; R143A (rouge) (**B**) ; R143E (bleu) et R143I (vert) (**C**).
D. Etudes cinétiques des variants du site actif de aCBF5 Y110A et Y110F (rouge), Y179A et Y179F (bleu), R181A (vert).

aGAR1, la réaction est plus lente mais la majorité des molécules de substrat sont modifiées après 80 min d'incubation.

L'étude de l'effet des substitutions par la mesure de l'activité en présence ou non de aGAR1 a été testée. Pour cela, nous avons réalisé les tests d'activités des sRNP portant l'ensemble des 4 protéines (LCNG) ou en absence de aGAR1 (LCN) (Fig. r18 A-C). Dans l'une ou l'autre de ces conditions, les variants P140AP141A et Δ140-150 de aCBF5 conduisent à un taux très faible de résidus Ψ formés (Fig. r18 A). Nous concluons que la boucle 140-150 et que les deux prolines 140 et 141 de aCBF5 jouent un rôle capital dans l'activité de la particule. Par contre, en absence de aGAR1, la substitution R139A de aCBF5 entraine une activité réduite mais qui est en partie restaurée lors de l'ajout de la protéine aGAR1 (Fig. r18 A). Nous concluons que le résidu R139 de aCBF5 ne joue pas un rôle majeur dans l'activité d'une sRNP H/ACA assemblée à partir des 4 protéines.

De façon remarquable, l'activité mesurée avec le variant R143A est similaire en présence ou absence de aGAR1. De plus, la cinétique obtenue avec la particule variante est comparable à celle observée avec aCBF5 sauvage dans les conditions optimales d'activité, c'est-à-dire en présence de aGAR1 (Fig. r18 B). Nous en déduisons que la substitution de cette arginine par une alanine conduit à une sRNP qui ne nécessiterait plus la présence de aGAR1 pour fonctionner de façon optimale et par conséquent que l'arginine 143 contribue fortement à la vitesse de réaction. Nous avons alors cherché à savoir quel élément de la chaîne latérale de l'arginine 143 était le plus important pour l'activité des particules : sa taille ou sa charge positive ? Nous avons donc substitué l'arginine par un acide glutamique (variant R143E) dont la charge est négative, ou par une isoleucine (variant R143I) qui comporte une longue chaîne latérale aliphatique non chargée. Les résultats obtenus indiquent que les variants R143E et R143I se comportent comme aCBF5 sauvage (Fig. r18 C). Nous concluons que l'absence d'une longue chaîne latérale à la position 143 conduirait à une sRNP H/ACA qui n'aurait plus besoin de aGAR1 pour exercer une activité optimale.

3.3. Analyse de l'activité ARN:Ψ-synthase de sRNP assemblées avec des variants du site actif de aCBF5

Le remplacement de l'enzyme sauvage par l'un des variants Y110A ou Y179A conduit à une très faible activité des particules (Fig. r18 D). Cela indique que ces deux résidus sont importants pour l'activité des sRNP à boîtes H/ACA. De façon intéressante, les variants Y110F et Y179F conduisent à un taux de résidus Ψ formés légèrement inférieur à celui obtenu pour aCBF5 sauvage (Fig. r18 D). Le simple remplacement de l'alanine en phénylalanine pour les résidus Y110 et Y179, permet de restaurer une grande partie de l'activité des particules, dont la cinétique est cependant ralentie (la majorité de l'ARN substrat est modifié après 80 min d'incubation). Cela indique que dans le mécanisme d'isomérisation, ces résidus ont un rôle majoritairement structural par leur groupement phényle et que leur groupement hydroxyle n'est sans doute pas impliqué dans la catalyse. La cinétique obtenue avec le variant R181A est également légèrement ralentie par rapport à celle de la protéine sauvage (Fig. r18 D), ce qui indique que ce résidu n'est pas essentiel dans la catalyse.

3.4. Analyse de l'incorporation de l'ARN substrat au sein des sRNP sauvage et variantes

Tout au long de ce travail, les tests d'activité sont effectués dans des conditions dans lesquelles l'ARN substrat est en quantité très minoritaire par rapport à la quantité de sRNP formées, avec un rapport molaire de ~70 sRNA pour 1 ARN substrat. Ainsi, chaque particule sRNP assemblée modifie une seule molécule de substrat et n'effectue pas de multiple turn-over. Les cinétiques de modification obtenues à partir des sRNP Pab91-L7Ae-aCBF5-aNOP10 sauvages indiquent qu'au bout de 10 min d'incubation, moins de 40% des ARN substrats sont modifiés alors qu'au bout de 60 min, la réaction est terminée pour la majorité d'entre eux (>80% de résidus Ψ formés) (Fig. r18 A-C).

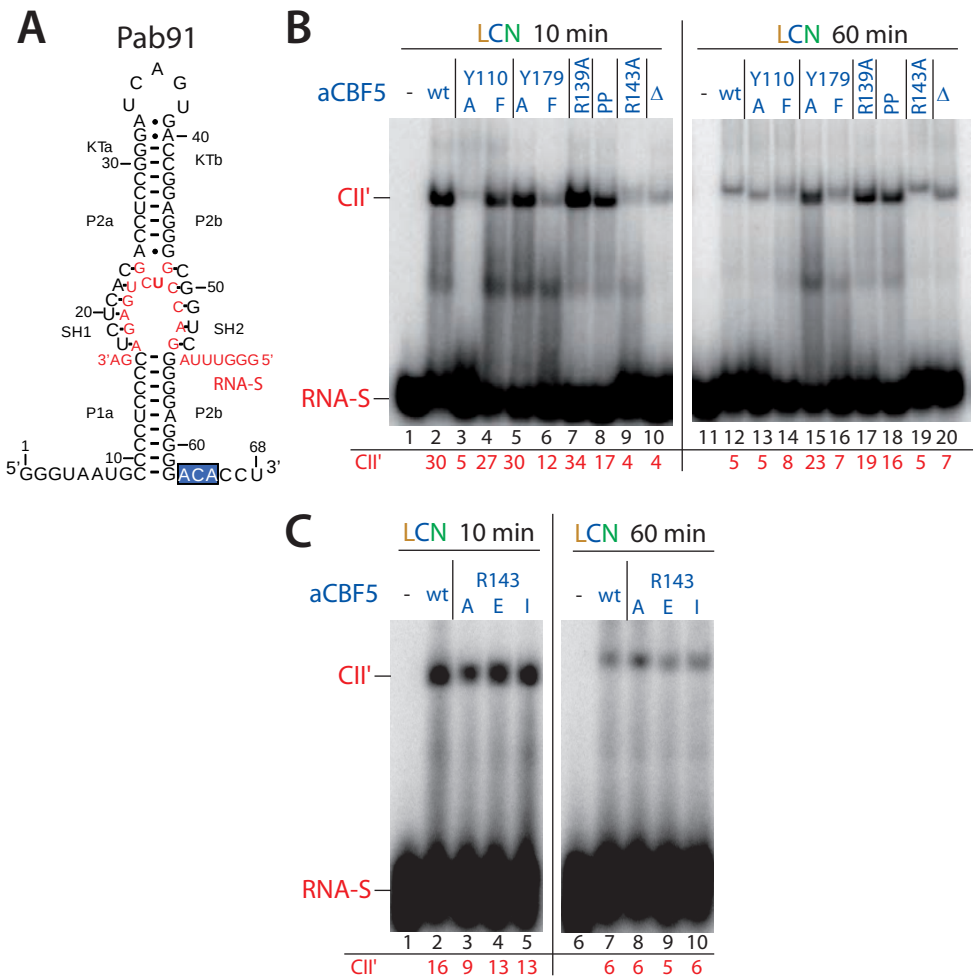


Figure r19: Etude par expériences de retard de migration sur gel de la formation des complexes CII'.

A. Pour rappel, structure schématique du sRNA Pab91 à boîtes H/ACA de *P. abyssi* en interaction avec son substrat RNA-S.

B. et **C.** Résultats d'expériences de retard de migration sur gel réalisées en présence du substrat RNA-S marqué en 5', du sRNA Pab91 non radiomarqué et du lot de protéines L7Ae, aCBF5, aNOP10 (LCN). L'enzyme aCBF5 sauvage et les variants de aCBF5 employés sont indiqués au-dessus du gel. Le taux de complexe formé entre le sRNA, le substrat RNA-S et les 3 protéines L7Ae, aCBF5 et aNOP10 (CII') est indiqué sous chaque pistes. Chaque expérience a consisté en une incubation soit de 10 minutes soit de 60 minutes avant l'électrophorèse.

Nous avons souhaité compléter les données de cinétiques de l'activité ARN:Ψ-synathase par l'analyse des complexes sRNP-ARN substrat, encore appelés CII' (Charpentier *et al.*, 2005; Charpentier *et al.*, 2007). Ceux-ci peuvent être observés par retard de migration sur gel. Nous avons analysé le taux de ces complexes présents après 10 min ou 60 min d'incubation des composants de la sRNP Pab91 (le sRNA Pab91 et le mélange de protéines L7Ae, aCBF5 et aNOP10) avec son substrat RNA-S radiomarqué en 5' (Fig. r19). Les échantillons ont été déposés sur gel de polyacrylamide non dénaturant et soumis à électrophorèse. Les complexes CII' présents après 10 minutes d'incubation correspondent aux sRNP qui seraient encore en cours de réalisation de réaction de pseudouridylation, à un stade postérieur à l'étape 2, mais précédant le déroulement de l'étape 5 du mécanisme (Fig. i62). Après 60 minutes d'incubation dans ces conditions, la totalité de l'ARN substrat est modifiée et l'absence de complexe CII' reflèterait le fait que la grande majorité des particules a franchi l'étape 5 du mécanisme et que l'ARN substrat modifié a été libéré des particules.

En employant aCBF5 sauvage, un complexe CII' est présent à 10 min et disparait au bout de 60 min, reflétant l'état d'avancement de la réaction de pseudouridylation (Fig. r19 B pistes 2 et 12, C pistes 2 et 7). Les variants Y110F, R143E, R143I et (R181A, non montré) de aCBF5 se comportent comme l'enzyme sauvage (Fig. r19 B pistes 4 et 14, C pistes 4, 5, 9 et 10), ce qui confirme les cinétiques de modification obtenues dans la figure r18. Nous notons que l'emploi des variants Y110A, Y179F, R143A et Δ140-150 conduit à pas ou peu de complexe CII' à 10 min et à 60 min (Fig. r19 B pistes 3/13, 6/16, 9/19, 10/20, C pistes 3 et 8). Les cinétiques obtenues avec les variants Y110A et Δ140-150 sont en accord avec ces résultats, la très faible activité mesurée pour ces variants pourrait provenir d'un très faible taux de complexe CII' formé avec l'ARN substrat ou que ce complexe est très instable. Par contre, dans le cas des variants R143A et Y179F qui permettent la formation d'un fort taux de résidus Ψ, l'observation d'un faible taux de CII' est surprenante. Pour le variant Y179F, nous privilégions l'hypothèse d'une perte de stabilité de l'interaction du substrat avec le site actif mais le temps d'interaction entre les partenaires est

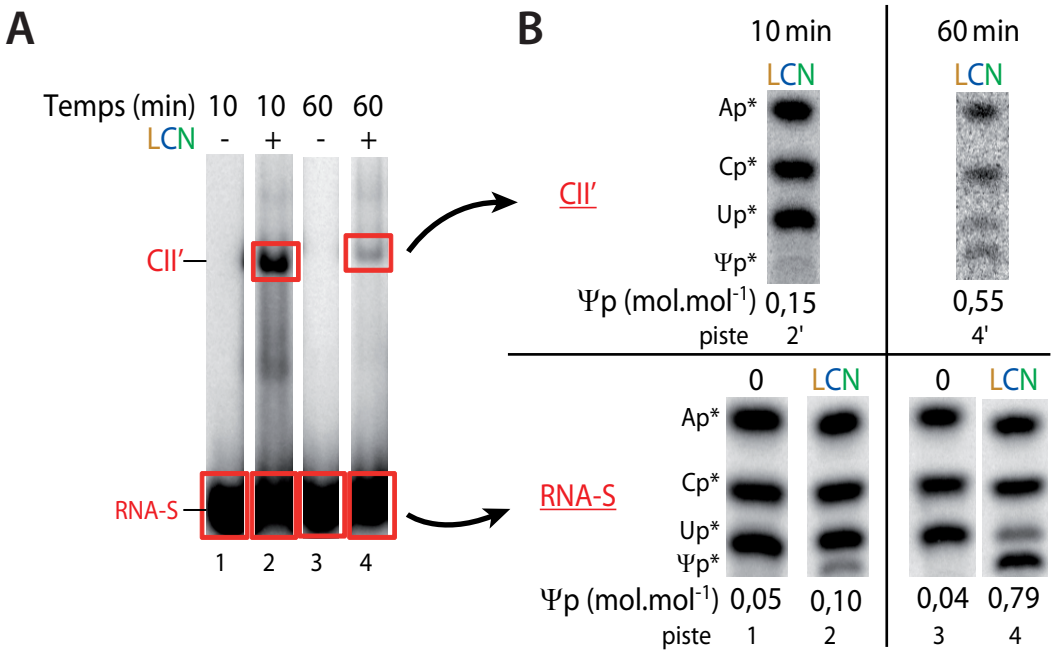


Figure r20 : Principe de la méthode d'étude du taux de substrat RNA-S modifié et libéré par la sRNP Pab91 à boîtes H/ACA.

A. Résultats d'expériences de retard de migration sur gel réalisées en présence du substrat RNA-S uniformément radiomarqué à l'[α- ^{32}P]CTP, de l'ARN guide Pab91 et en présence ou non des trois protéines L7Ae, aCBF5 et aNOP10 (LCN). L'électrophorèse a été effectuée après des temps d'incubation de 10 min ou de 60 min.

B. Taux de pseudouridine formé dans le substrat RNA-S présent soit dans la fraction libre, soit dans le complexe CII'. Pour cela, les bandes correspondant aux complexes CII' et à l'ARN libre de **B.** ont été découpées individuellement (cadres en rouge dans **A**), éluées du gel puis soumis à la digestion enzymatique de la RNase T2. Les produites de digestion ont été déposés sur plaque de verre/cellulose pour subir une chromatographie en 1 dimension. L'intensité des spots correspondant aux radionucléotides 5' phosphate (p*) a été quantifiée au phosphoimager et le taux de pseudouridine formé dans chacune des pistes est indiquée.

néanmoins suffisant pour que la réaction catalytique ait le temps de se dérouler. Notre interprétation des données concernant le variant R143A est donnée dans la partie suivante des résultats. Enfin, les variants Y179A, R139A et P140AP141A forment un complexe CII' qui est toujours présent au bout de 60 min (Fig. r19 B pistes 5/15, 7/17, 8/18). Une explication possible serait que le substrat puisse rester bloqué dans ces sRNP variantes, qu'il soit modifié (comme c'est le cas avec le variant R139A) ou non (cas du variant Y179A et P140AP141A). Pour vérifier cette hypothèse, nous avons développé une procédure expérimentale permettant de tester si le substrat modifié est libéré de la particule.

3.5. Mise au point d'une méthode permettant d'analyser le déroulement de l'étape de libération du produit de la réaction de pseudouridylation

Dans la méthodologie du plus proche voisin que nous employons pour mesurer l'activité des particules, les ARN substrat marqués sont extraits par traitement au phénol/chloroforme. Il est donc possible, bien que modifiés, que ceux-ci n'aient pas été libérés des particules. La technique que nous avons mise au point repose sur : i) le fractionnement des ARN libres et ceux complexés à la particule (CII') par électrophorèse non dénaturante (Fig. r20 A), i) le découpage des bandes correspondant au complexe CII' ou aux ARN substrat libres (Fig. r20 A encadrés en rouge), ii) l'élution des ARN contenus dans ces bandes (Fig. r20 de A vers B), iii) la digestion par la RNase T2 des ARN, iv) le fractionnement des ARN par TLC en une dimension (Fig. r20 B) et v) l'analyse du taux de pseudouridine formé dans ces ARN (Fig. r20 B).

La figure r21 présente les résultats obtenus en employant le sRNA Pab91, son substrat RNA-S uniformément radiomarqué à l'[α-^{32}P]CTP et les 3 protéines sauvages L7Ae, aCBF5 et aNOP10 (LCN). Des essais sans protéines ont été réalisés et ils indiquent qu'aucun complexe CII', dans ces conditions d'incubation, ne se forme (Fig. r20 B pistes 1 et 3). Les complexes CII' ont été analysés après 10 minutes (Fig. r20 B piste 2) et 60 minutes d'incubation (Fig. r20 B piste 4).

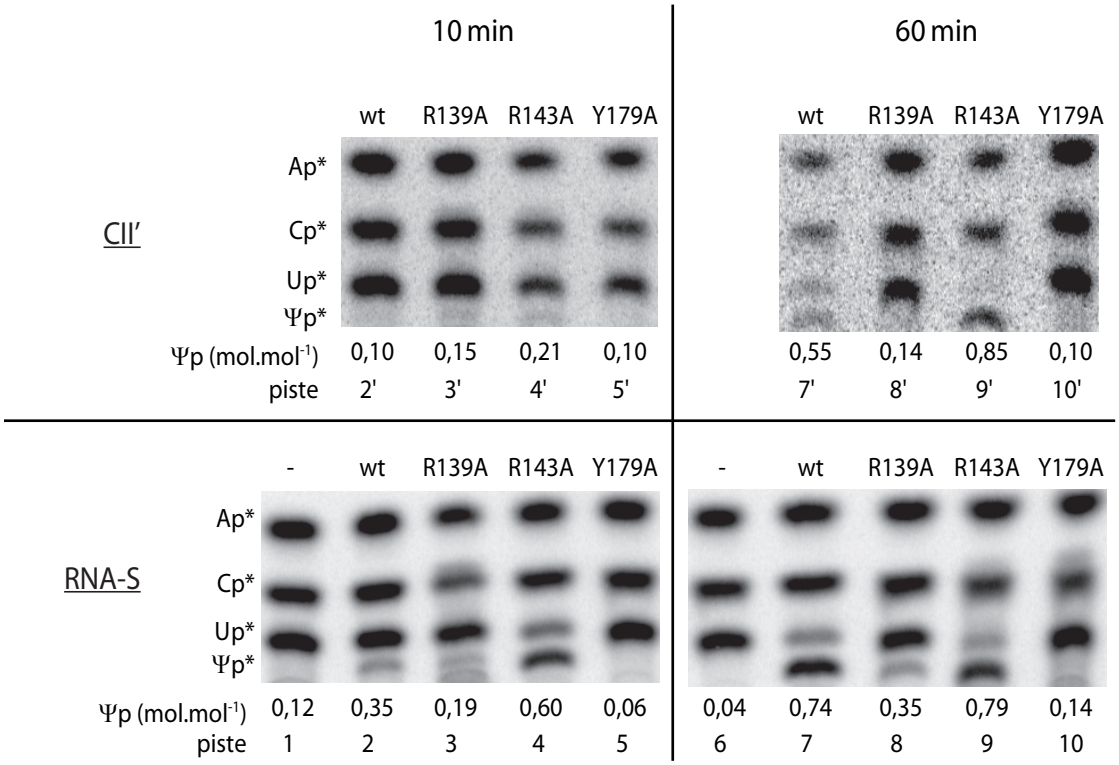


Figure r21 : Effet des substitutions R139A, R143A et Y179A sur le taux de substrat RNA-S modifié et libéré par la sRNP Pab91 à boîtes H/ACA.

Les expériences ont été réalisées suivant le principe décrit en figure r20. Les complexes CII' ont été formés aux temps d'incubation de 10 minutes et 60 minutes par expériences de retard de migration sur gel, en présence du substrat RNA-S uniformément radiomarqué à l'[α-^{32}P]CTP, du sRNA Pab91 non radiomarqué, des trois protéines L7Ae, aCBF5 et aNOP10 (LCN) dont aCBF5 est soit l'enzyme sauvage soit l'un des variant indiqués au dessus de chaque piste. Les ARN libres ou présents au sein des complexes CII' sont individuellement élués des bandes du gel les contenant. Les ARN sont digérés à la RNase T2 puis fractionnés par chromatographie en 1 dimension. Le taux de pseudouridine formé est quantifié au phosphoimager et est indiqué.

Le taux de résidus Ψ formés au sein du complexe CII' au bout de 10 minutes d'incubation indique que 15% des molécules de substrat présentes au sein de ces complexes CII' sont modifiées (Fig. r20 B piste 2'). Une partie des molécules de substrat a déjà été modifiée et libérée car 10% de résidus Ψ sont détectés parmi les molécules de la fraction des ARN libre (Fig. r20 B piste 2).

Au bout de 60 minutes, une très faible proportion des molécules d'ARN substrat sont encore en interaction avec les sRNP (Fig. r20 A piste 4), et 55% des molécules d'ARN substrat encore présentes au sein des complexes résiduels CII' sont modifiées (Fig. r20 B piste 4'). Un taux de modification de 79% est mesuré pour les ARN présents dans la fraction libre ce qui indique que la réaction de modification a eu lieu et que la majorité des substrats modifiés ont été libérés des sRNP (Fig. r20 B piste 4).

Par cette méthode, nous montrons donc que dans nos conditions expérimentales, les particules réalisent un cycle complet de la réaction de pseudouridylation. Les ARN substrat sont pris en charge par les particules, ils y sont modifiés puis libérés (Fig. i62, étapes 2 à 5).

3.6. Le variant R143A de aCBF5 permet de stimuler l'activité globale de la particule sans nécessiter la présence de aGAR1

Nous avons analysé le taux de résidus Ψ formé dans les molécules présentes dans le complexe CII' et dans la fraction d'ARN libre dans le cas du variant de aCBF5 R143A (Fig. r21). Malgré un taux de complexe CII' plus faible pour ce variant (Fig. r19 C piste 3), nous constatons que les résidus U sont rapidement convertis en résidus Ψ (Fig. r21). En effet, dès 10 minutes d'incubation, 60% (35% pour le wt) des molécules d'ARN libre sont modifiées (Fig. r21 piste 4). Au bout de 60 minutes, 79% des ARN libres sont modifiés ce qui est proche de la valeur obtenue avec les particules sauvages (74%) (Fig. r21 pistes 7 et 9).

Les taux de résidus Ψ formés qui sont plus élevés pour le variant R143A que pour l'enzyme sauvage confirment les données des mesures de cinétique. Ainsi, i) la réaction de pseudouridylation est plus rapide avec aCBF5 R143A qu'avec aCBF5 sauvage, le faible taux de complexe CII' observé dans la figure r19 est en accord avec une vitesse de réaction élevée, plus de la moitié des molécules de substrat libres étant déjà modifiée au bout de 10 min, ii) le substrat pris en charge par la sRNP contenant le variant R143A subit un cycle de réaction complet et n'est pas accumulé dans un complexe abortif avec la particule, iii) cette vitesse élevée est indépendante de la présence de aGAR1 et est le résultat de la substitution du résidu R143.

Les variants de aCBF5 R143E et R143I ont également été analysés par cette approche et les résultats obtenus confirment les tests d'activités (données non montrées).

3.7. Les substitutions Y179A et P140AP141A n'empêchent pas une fixation stable de l'ARN substrat à la particule mais affectent l'activité catalytique

La technique par purification sur gel appliquée aux particules contenant le variant Y179A confirme les résultats des tests d'activité car aucun résidu Ψ n'est détecté, soit dans le complexe CII', soit dans la fraction d'ARN libre (Fig. r21 pistes 5, 5', 10 et 10'). Bien qu'un taux de complexe CII' équivalent soit obtenu avec les particules sauvages et celles portant la substitution Y179A, il semblerait que les ARN substrat soient bloqués dans le complexe CII' formé en présence de aCBF5 Y179A, et restent non modifiés. Après 60 min d'incubation, le complexe CII' est toujours présent (Fig. r19 B) et les ARN substrat ne sont pas modifiés (Fig. r21, piste 10').

L'analyse du variant P140AP141A a conduit aux mêmes observations et conclusions : aucun résidu Ψ n'est formé dans le complexe CII' au temps 10 min et 60 min, ni dans les ARN libres, ce qui implique que pour ce variant également, les substrats resteraient bloqués dans les sRNP sans y être modifiés (données non montrées).

Nous avons réalisé des incubations de ces particules variantes Y179A et P140A P141A pendant plusieurs jours à température ambiante. Nous avons observé qu'après ces temps d'incubation, le complexe CII' était toujours détecté par retard de migration sur gel et qu'aucun résidu Ψ ne s'était formé (données non montrées). Ces deux variants de aCBF5 sont de très bons candidats pour réaliser des études par radiocristallographie des complexes sRNP-ARN substrat.

Le variant R139A était le troisième variant pour lequel le taux de complexe CII' reste élevé après 60 min d'incubation (Fig. r19), ce qui est accord avec le fait qu'en absence de aGAR1, l'activité reste faible. Nos résultats obtenus en figure r21 indiquent que la particule variante permet une isomérisation lente et progressive des résidus U en Ψ en absence de aGAR1 (Fig. r21 pistes 3, 3', 8 et 8'). Cependant, cette activité de pseudouridylation est toujours stimulée par la présence de aGAR1.

3.8. Recherche des résidus de aCBF5 contribuant à la sélection par la sRNP du substrat à modifier

3.8.1. Utilisation de substrats RNA-U/C dont le résidu U cible de la réaction est substitué par un C

B. Charpentier avait montré que la substitution de l'uridine cible par une cytosine (U/C) dans l'ARN substrat ne permettait plus de former de complexes CII' stables (Charpentier *et al.*, 2005). Cela indiquait que les appariements SH1 et SH2 formés entre le sRNA et l'ARN substrat n'étaient pas suffisants pour permettre la formation d'un complexe stable et que la présence du résidu U était un déterminant important pour la sélection du substrat à modifier par la sRNP. Il était fort probable qu'un ou des résidus du site actif de aCBF5 participent à cette discrimination en faveur d'une séquence portant un résidu U à modifier.

Pour approfondir cette question, nous avons analysé la formation des complexes CII' formés par l'interaction des sRNP Pab91-LCN sauvage et variantes avec deux types de substrats qui différent par leur taille (Fig. r22 A) :

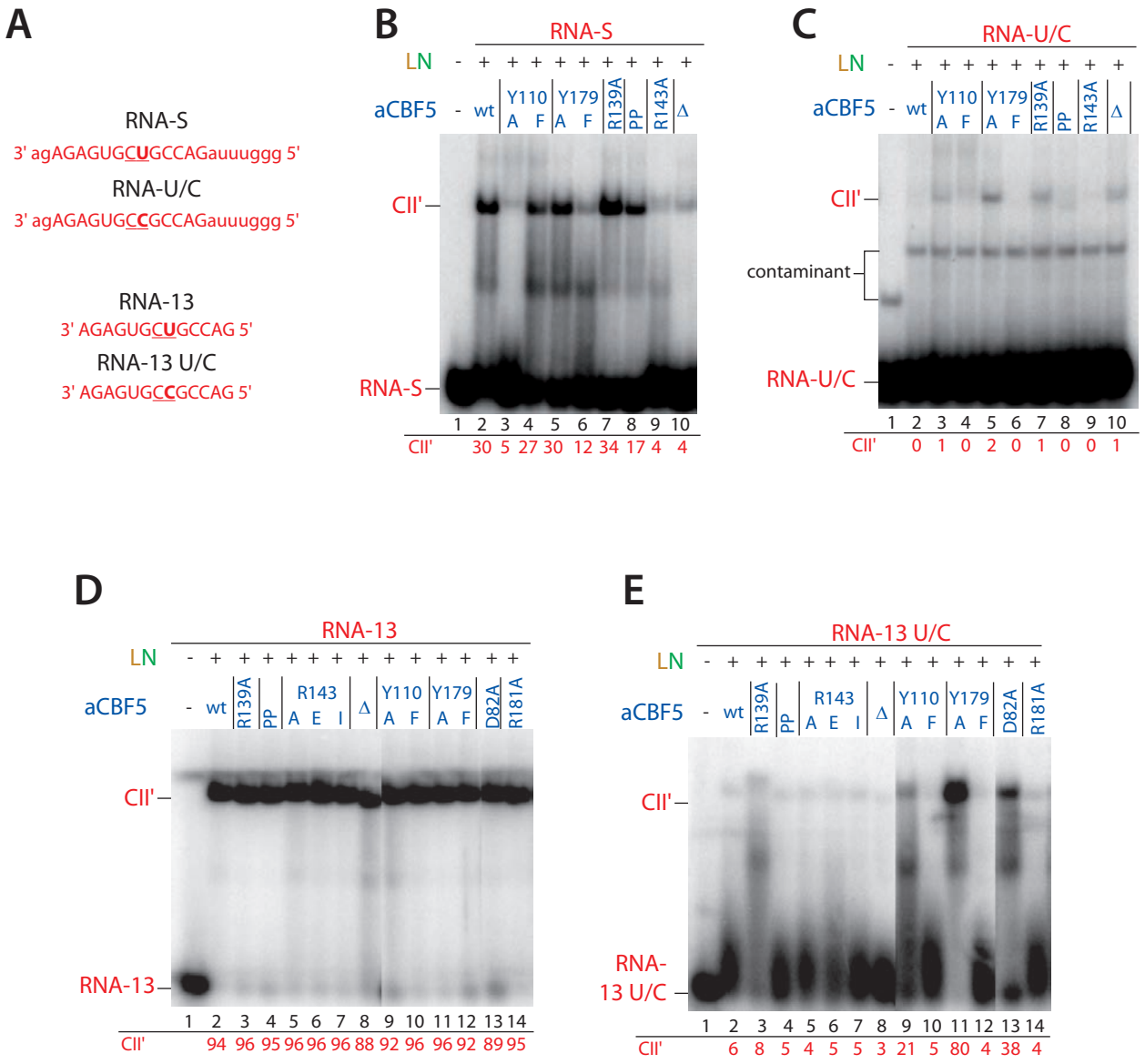


Figure r22 : Etude par des expériences de retard de migration sur gel de la formation des complexes CII' entre des particules sauvages et variantes et d'un substrat dont le résidu U cible de la réaction est substitué en C.

A. Structure des différents substrats ARN employés. Le substrat dont la séquence est basée sur celle de l'ARN RNA-S et dont le résidu U cible de la réaction est substitué en C est dénommé RNA-U/C ; l'ARN d'une taille de 13 mer et dont le résidu U cible de la réaction est muté en C est dénommé RNA-13 U/C. Les 2 nucléotides non appariés contenant le résidu U cible sont soulignés et les nucléotides des extrémités 5' et 3' qui ne sont pas appariés à la poche de pseudouridylation sont en minuscule.

B. à **E.** Résultats d'expériences de retard de migration sur gel réalisées en présence des substrats indiqués radiomarqués en 5' RNA-S (**B**), RNA-U/C (**C**), RNA-13 (**D**) et RNA-13 U/C (**E**), du sRNA Pab91 non radiomarqué et du mélange de protéines L7Ae, aCBF5, aNOP10 (LCN) dont l'enzyme aCBF5 sauvage et ses variants employés sont indiqués au-dessus des gels. Le taux de complexe formé entre le sRNA, le substrat et les 3 protéines LCN (CII') est indiqué sous chaque pistes. Dans chaque expérience, les échantillons ont été incubés 10 minutes à 65°C. La présence d'une bande contaminante (**B.**) corresponderait à une fraction des ARN sous une conformation aberrante.

- L'ARN RNA-S d'une taille de 22 nts, qui est un fragment de la région de l'ARNr 23S modifiée à la position U2685 par Pab91, et le variant RNA-U/C, qui porte la substitution U_{2685}/C, ont été produits par transcription *in vitro* (Fig. r22 A).

- L'ARN RNA-13 correspondant à une séquence raccourcie de 13 mers restreinte exclusivement à la région d'appariement de Pab91, et le variant RNA-13-U/C portant la substitution U_{2685}/C ont été synthétisés chimiquement (Fig. r22 A).

Comme attendu (Charpentier *et al.*, 2005), nous constatons que les ARN RNA-U/C sont bien discriminés comme des non substrats par les particules car aucun complexe CII' stable n'est détecté (Fig. r22 C, piste 2). Cependant alors que les sRNP portant les protéines aCBF5 variantes conservent cette propriété de discrimination, un très faible taux de complexe CII' est observé avec les particules assemblées avec les variants de aCBF5 Y110A, Y179A, R139A, et Δ140-150 (Fig. r22 C pistes 3, 5, 7 et 10), suggérant que ces résidus participeraient à la sélection du substrat pouvant être modifiées. Afin de vérifier ces données, nous avons testé l'effet des substitutions des résidus de aCBF5 sur le taux de complexe CII' formé avec les substrats RNA-13 et RNA-13-U/C. Un très fort taux de complexe est formé avec le substrat RNA-13 sauvage puisque la totalité de cet ARN est déplacé dans les complexes CII' (Fig. r22 D piste 2). De façon surprenante, aucune des substitutions n'affecte la formation des complexes CII' formés avec le substrat RNA-13 (Fig. r22 D, pistes 3-14). Comme dans le cas du substrat plus long RNA-S, la substitution du résidu U cible par un résidu C affecte dramatiquement le taux de formation des complexes CII' (Fig. 22 E, piste 2). Ainsi, les sRNP conservent leur capacité de discriminer parmi les substrats de très petite taille ceux portant un résidu U à modifier. La substitution Y179A affecte fortement cette propriété (Fig. r22 E, piste 11) et à un degré moindre les substitutions Y110A, D82A (Fig. r22 E, pistes 9 et 13). Dans le cas de la substitution R139A le profil de migration est très diffus (Fig. r22 E, piste 3).

Nous en concluons que le résidu Y179 de aCBF5, et à un moindre degré les résidus D82 et Y110, jouent un rôle dans la reconnaissance du résidu U cible de la

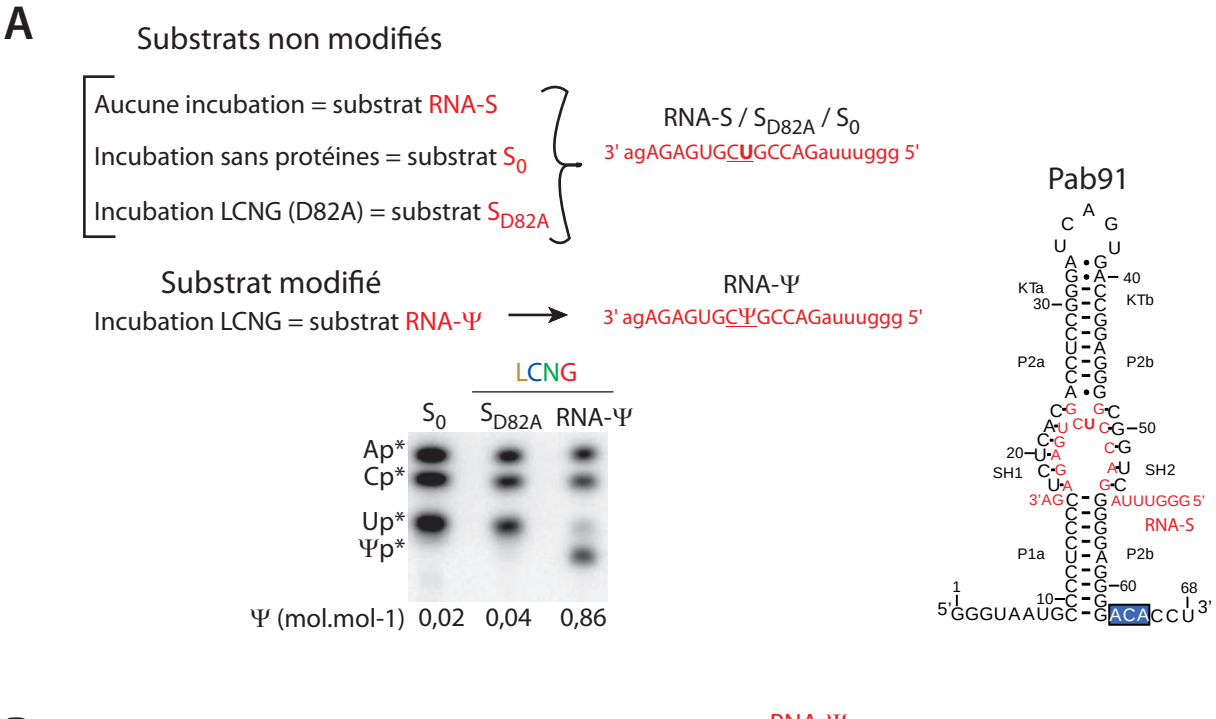


Figure r23: Pseudouridylation *in vitro* d'un ARN substrat et étude par expériences de retard de migration sur gel de son incorporation dans une sRNP.

A. Mesure du taux de résidus Ψ formés dans différents ARN substrats. Trois substrats ont été préparés. Le premier, nommé RNA-Ψ, a été modifié à un taux de 86% suite à l'incubation de l'ARN RNA-S avec les protéines LCNG. Le second substrat, nommé S_{D82A}, est non modifié mais a été incubé avec la sRNP Pab91-LCNG portant le variant inactif de aCBF5 D82A. Le troisième substrat, nommé S_0, n'est pas modifié mais a été préparé par incubation de l'ARN RNA-S sans protéines.

B. Résultats d'expériences de retard de migration sur gel réalisées en présence des différents substrats uniformément radiomarqués, du sRNA Pab91 non radiomarqué et du lot de protéines L7Ae, aCBF5, aNOP10 (LCN). Les différents substrats employés et les variants de aCBF5 sont indiqués au-dessus des gels. Le taux de complexe formé entre le sRNA, le substrat et les 3 protéines L7Ae, aCBF5 et aNOP10 (CII') est indiqué sous chaque pistes. Chaque expérience provient d'une incubation de 10 minutes à 65°C.

réaction. Cela est en accord avec le positionnement de ces résidus autour du résidu U dans les structures 3D de TruB (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003).

3.8.2. Production et utilisation d'un substrat pseudouridylé, RNA-Ψ

L'effet délétère sur le taux de complexe CII' de la substitution du U cible de la modification par un C suggérait que la présence d'un résidu Ψ pouvait conduire au même effet et, de ce fait, empêcher la réassociation du substrat modifié à la particule. Nous avons donc testé si un substrat déjà pseudouridylé pouvait donner lieu à un complexe stable CII' avec une sRNP à boîtes H/ACA (Fig. r23). J'ai donc produit par transcription *in vitro* le substrat RNA-S uniformément radiomarqué à l'[α-^{32}P]CTP, puis incubé celui-ci en présence du sRNA Pab91 et des 4 protéines L7Ae, aCBF5, aNOP10 et aGAR1 pendant 80 minutes à 65°C, et enfin analysé le taux de résidus Ψ formé dans ce substrat (Fig. r23 A). Dans ces conditions expérimentales, le substrat RNA-S a été modifié à 86% et a été nommé RNA-Ψ (Fig. r23 A). En parallèle, les molécules de substrat RNA-S ont été traitées dans les mêmes conditions que celles décrites précédemment mais en absence de protéines. Après extraction, les ARN obtenus appelés S_0 ont été utilisés afin de vérifier que le traitement *in vitro* n'affectait pas leur capacité à se fixer aux sRNP. Le substrat RNA-S a également été incubé avec les particules portant le variant inactif aCBF5 D82A. Les ARN extraits désignés S_{D82A} et non modifiés (Fig. r23A) ont été utilisés comme second témoin positif.

Les deux substrats S_0 et S_{D82A} conservent leur capacité à s'incorporer dans les particules pour former un complexe CII' (Fig. r23 B pistes 3 et 5). Par contre, le substrat pseudouridylé RNA-Ψ ne permet plus la détection de complexe CII'. Que ce soit pour aCBF5 wt ou les variants de aCBF5, l'ARN RNA-Ψ utilisé comme substrat ne forme qu'une très faible quantité de complexes CII' (Fig. r23 B pistes 6-17). Nous concluons qu'une sRNP H/ACA a la capacité de sélectionner un substrat non modifié d'un substrat pseudouridylé et qu'aucun des variants de aCBF5 que nous avons produit ne conduit à la perte de cette propriété de sélection.

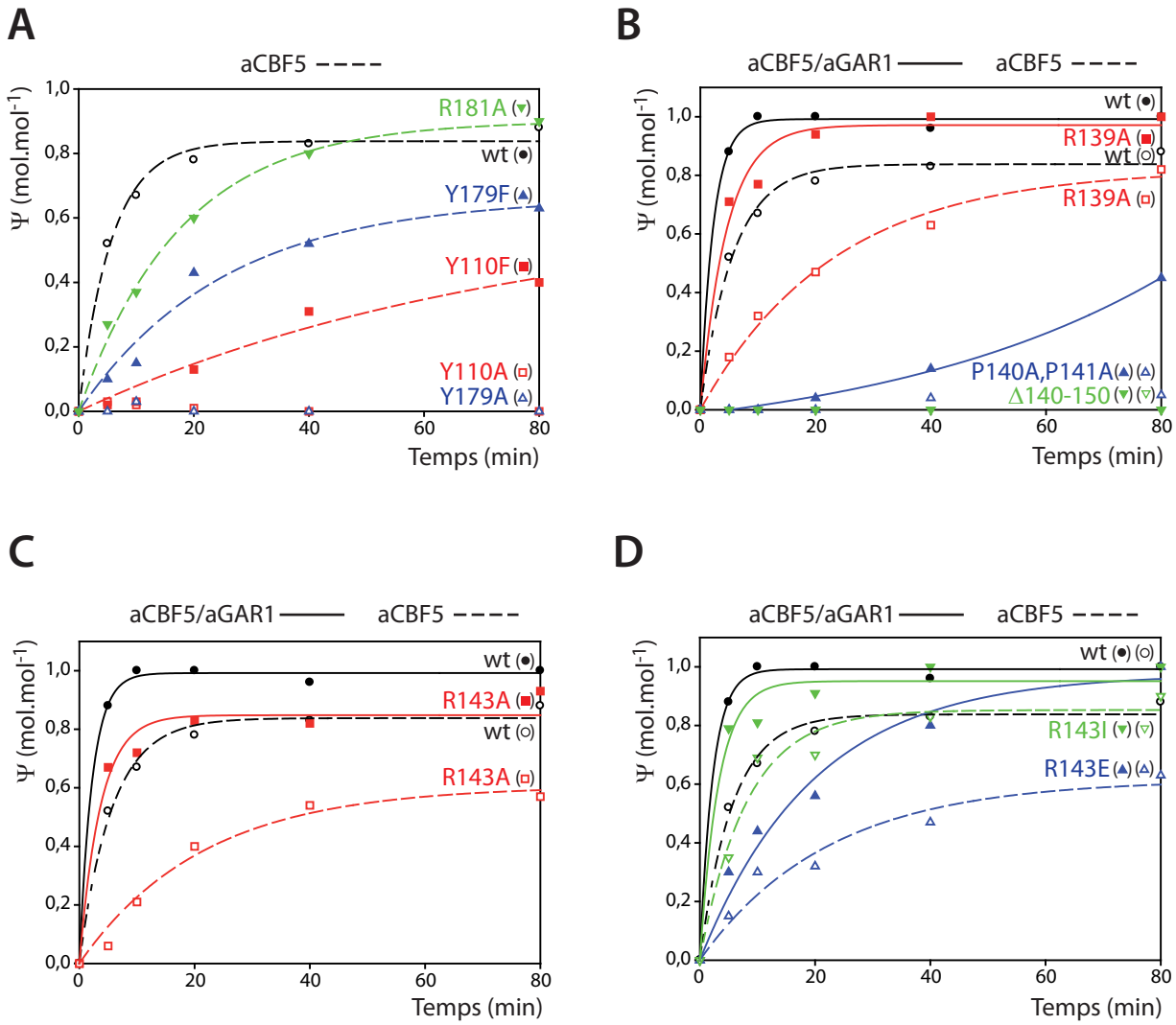


Figure r24 : Etude cinétique du taux de résidus Ψ formés en employant des variants de aCBF5. Les tests ont été réalisés en employant le jeu de protéines aCBF5 et aGAR1 (trait plein) ou avec l'enzyme aCBF5 seule (trait pointillé). Les conditions expérimentales correspondent à l'emploi du l'ARNt^{Asp} de *P. abyssi* et une température de 65°C. Les résultats obtenus avec aCBF5 sauvage sont représentés en noir.

A. Etudes cinétiques des variants de aCBF5 Y110A et Y110F (rouge), Y179A et Y179F (bleu), R181A (vert).

B. Etudes cinétiques des variants de aCBF5 R139A (rouge), P140A,P141A (bleu), Δ140-150 (vert).

C. Etudes cinétiques du variant de aCBF5 R143A (rouge).

D. Etudes cinétiques des variants de aCBF5 R143E (bleu) et R143I (vert).

3.9. Etude des variants de aCBF5 dans l'activité de modification non guidée

L'étude de résidus de aCBF5 et de aNOP10 nous a déjà révélé qu'il existait des différences de fonctionnement entre le système guidé et le système non guidé (article n°3, (Muller *et al.*, 2007b)). Nous avons souhaité compléter cette étude et avons donc analysé la modification d'un ARNt à la position U55 par l'activité non guidé des variants du site actif et de la boucle β7/β10 de aCBF5 (Fig. r24).

Mesure des activités obtenues avec les variants de aCBF5 dans le système non guidé

Les substitutions des résidus Y110 et Y179 du site actif conduisent à des cinétiques de modification comparables à celles obtenus avec le système guidé (Fig. r24 A). En effet, la substitution en alanine abolit l'activité de aCBF5, par contre sa substitution en phénylalanine permet de restaurer une partie de l'activité ARN:Ψ-synthase (Fig. r24 A). Néanmoins, l'ajout des protéines aNOP10 et aGAR1 aux variants de aCBF5 Y110A, Y110F et Y179F permettent de restaurer une activité similaire à celle obtenue pour le complexe CNG sauvage mais aucune activité n'est décelée pour le complexe CNG contenant le variant de aCBF5 Y179A (données non montrées). Nous concluons que les résidus Y110 et Y179 jouent un rôle similaire dans les deux systèmes de modification guidé et non guidé, ce qui est compatible avec un mécanisme catalytique commun à ces deux systèmes. La cinétique obtenue avec le variant R181A est comparable à celle mesurée avec aCBF5 sauvage (Fig. r24 A). Nous concluons que le résidu R181 de aCBF5 n'est pas non plus essentiel dans l'activité non guidée.

De la même manière, les variants du pouce de aCBF5 R139A, P140AP141A et Δ140-150 se comportent d'une façon équivalente à celle observée dans le système guidé (Fig. r24 B-D). En présence ou en absence de aGAR1, la substitution P140AP141A et la délétion Δ140-150 abolissent presque totalement l'activité non guidée (Fig. r24 B). En présence de aGAR1, l'effet de la substitution R139A n'a que peu d'effet sur l'activité. En absence de aGAR1, ce variant modifie l'ARNt plus

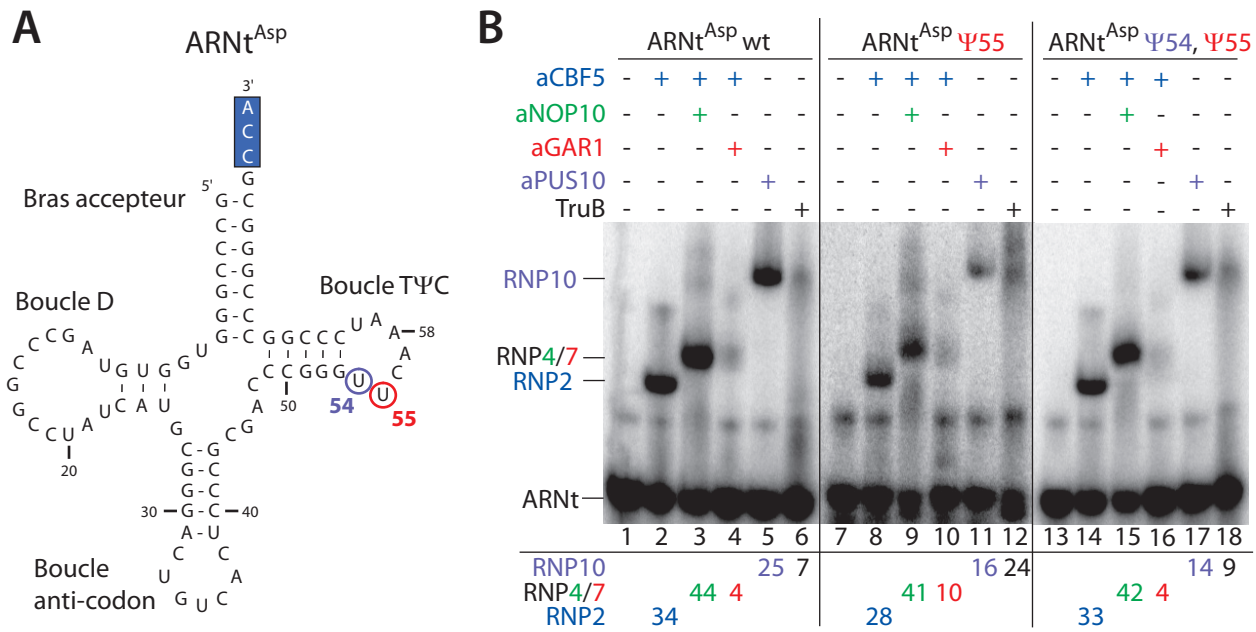


Figure r25 : Etude par expériences de retard de migration sur gel de la fixation des protéines à des ARNt déjà pseudouridylés.

A. Structure 2D de l'ARNt^{Asp} de *P. abyssi*. Les résidus modifiés par les enzymes aCBF5 (U55 en rouge) et aPUS10 (U54 en violet et U55) sont entourés.

B. Résultats d'expériences de retard de migration sur gel réalisées en présence des ARNt indiqués en haut des pistes. Ceux-ci ont été uniformément radiomarqués à l'[α-^{32}P]CTP et incubés avec les protéines aCBF5, aNOP10, aGAR1, aPUS10 et TruB comme indiqués au-dessus du gel. L'ARNt wt ne contient pas de résidu Ψ ; l'ARNt Ψ55 a été modifié par le jeu de protéines aCBF5, aNOP10 et aGAR1 pendant 80 min à 65°C et contient 100% de résidu Ψ en position 55 ; l'ARNt Ψ54,Ψ55 a été modifié par l'enzyme aPUS10 pendant 80 min à 65°C et contient 100% de résidu Ψ en position 55 mais aucune quantification n'a pu être réalisée pour la position 54 car le marquage à l'[α-^{32}P]CTP permet de révéler uniquement la position adjacente en 5' qui correspond à la position 55. Le taux de complexe formé entre l'ARNt et les différentes protéines est indiqué sous chaque piste. Dans chaque expérience, les échantillons ont été incubés 10 minutes à 65°C à l'exception de TruB qui a été incubée 10 min à 37°C.

lentement que aCBF5 wt. Les substitutions R143A ou R143E ralentissent l'activité non guidée des particules (Fig. r24 C et D), contrairement au variant R143I qui modifie l'ARNt avec une cinétique semblable à celle de aCBF5 sauvage (Fig. r24 D).

Nous concluons que la boucle β7/β10 de aCBF5 joue un rôle capital à la fois pour l'activité guidée et l'activité non guidée de aCBF5. Par contre, la conformation du site actif et celle du pouce de aCBF5 pourraient être différentes au sein des complexes ARNt-aCBF5 et des complexes sRNP-ARN substrat. En effet, la substitution R143A permet de stimuler l'activité guidée (Fig. r18 B) mais par contre affecte l'activité non guidée. A l'inverse le variant R143E n'affecte pas l'activité guidée (Fig. r18 C) mais ralentit l'activité non guidée. Dans les deux systèmes, il semble que la taille de la chaîne latérale et non pas la charge soit prépondérante dans la reconnaissance du substrat ARN car aucune des deux activités n'est changée avec lorsque le résidu R143 est substitué en isoleucine (variant R143I).

<u>Etude de l'effet de résidus pseudouridines introduits dans un ARNt, sur le taux de fixation des protéines par retard de migration sur gel</u>

J'ai ensuite testé la spécificité de aCBF5 et de ses variants vis-à-vis d'un ARNt pseudouridylé en position Ψ55. L'ARNt^{Asp} a été transcrit en présence d'[α-^{32}P]CTP, puis incubé 80 minutes avec le mélange de protéines aCBF5, aNOP10 et aGAR1, ce qui a conduit à un taux proche de 100% de modification des ARNt en position 55. L'ARN ainsi modifié est nommé ARNt^{Asp} Ψ55 (Fig. r25 A). En parallèle, l'ARNt a été incubé en présence de l'enzyme aPUS10 qui modifie les positions 54 et 55 mais seul le taux de résidus Ψ formés en position 55 a pu être quantifié car le marquage à l'[α-^{32}P]CTP ne permet pas de révéler la modification à la position 54 (Fig. r25 A). J'ai vérifié que l'ARN obtenu, appelé ARNt^{Asp} Ψ54,Ψ55, était modifié à la position U55 à un taux proche de 100%.

Les différents ARNt modifiés à la position U55 ont été incubés avec des jeux de protéines distincts et les complexes formés ont été fractionnés par électrophorèse

en gel non dénaturant (Fig. r25 B). Nous constatons que des complexes stables sont formés avec les protéines aCBF5 ou aCBF5-aNOP10 (Fig. r25 B pistes 2, 3, 8, 9, 14, 15). Par contre, le taux de complexe aCBF5-aGAR1 stable détecté est très faible, soulignant ainsi l'effet négatif de aGAR1 sur la stabilité de l'interaction entre aCBF5 et les ARNt, cet effet est comparable à celui observé dans le cas de plusieurs sRNA (Fig. r25 B pistes 4, 10 et 16). De façon intéressante, le taux de complexe formé par aPUS10 et les deux ARNt modifiés est diminué d'un facteur >1.5 par rapport à celui obtenu pour aPUS10-ARNt wt, suggérant que la présence d'un résidu Ψ pourrait déstabiliser l'interaction de aPUS10 avec les ARNt (Fig. r25 B pistes 5, 11 et 17). Les résultats observés avec l'enzyme TruB sont difficilement interprétables car aucun complexe suffisamment stable ne peut être observé sur gel (Fig. r25 B pistes 6, 12 et 18). Nous concluons que la présence de résidus Ψ aux positions 54 et 55 n'empêche pas la fixation des enzymes aCBF5 et aPUS10 mais que la présence d'un résidu Ψ à la position 55 pourrait néanmoins influencer la fixation de aPUS10 à l'ARNt.

4. Compilation des résultats et discussion

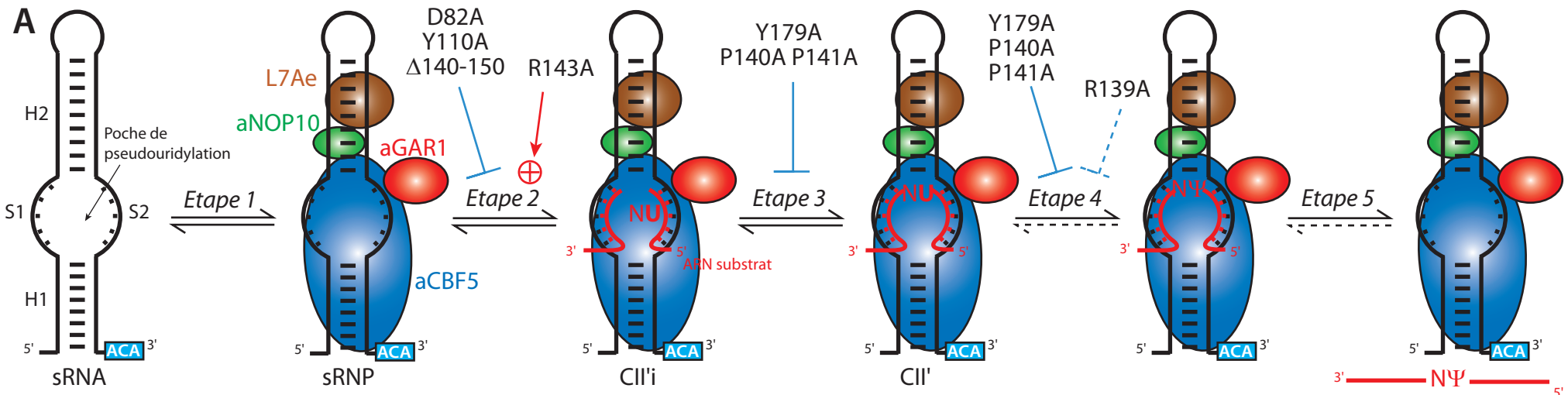
Le bilan de l'ensemble des résultats que nous avons obtenus est présenté dans la figure r26. Nous avions connaissance de ces résultats au moment où les structures 3D de sRNP comportant les protéines L7Ae, aCBF5, aNOP10, le sRNA et l'ARN substrat ont été publiées (Fig. r27) (Duan *et al.*, 2009; Liang *et al.*, 2009). Ces structures permettent de visualiser le positionnement du substrat et de ses contacts avec les résidus du le site actif de aCBF5. L'effet des différentes substitutions est discuté en s'appuyant sur ces données structurales.

4.1. Analyse des variants du site actif de aCBF5

Dans TruB, le résidu équivalent à la tyrosine 110 (Y76) a été proposé pour interagir par empilement de base avec l'uridine cible de la réaction (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003). Cette hypothèse est confirmé par les structures 3D

des particules contenant un ARN substrat dans lesquelles le résidu équivalent Y110 chez aCBF5 de *P. furiosus* établit également des interactions avec l'uracile cible (Fig. r27 A et B) (Duan *et al.*, 2009; Liang *et al.*, 2009). Nos données expérimentales indiquent que le résidu Y110 de aCBF5 possède un rôle très important dans la catalyse à la fois dans l'activité guidée et non guidée (Fig. r26). La substitution de ce résidu en alanine réduit très fortement l'activité des particules mais son remplacement par une phénylalanine permet de restaurer celle-ci. Le rôle principal de la tyrosine 110 de aCBF5 serait donc d'interagir par empilement de base avec le résidu U cible plutôt que d'intervenir dans la catalyse en tant que base générale (Phannachet *et al.*, 2005),

La tyrosine 179 correspond à la deuxième tyrosine conservée dans le site actif des ARN:Ψ-synthases de la famille TruB. Comme pour Y110, l'hypothèse d'un empilement du noyau de Y179 avec la base de l'uridine cible de la réaction a été formulée de même que le rôle de ce résidu dans la catalyse en tant que base générale (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003). Les structures 3D récentes de la sRNP avec le substrat indiquent que ce résidu (Y182 chez *P. furiosus*) réalise des interactions électrostatiques avec le phosphate 3' du résidu U cible de la réaction (Fig. r27 A et B) (Duan *et al.*, 2009; Liang *et al.*, 2009). Ainsi, le positionnement de ce résidu vis-à-vis de l'ARN substrat n'est pas le même dans le complexe TruB-ARNt et dans une sRNP associée à son ARN cible. Nos résultats montrent que le résidu Y179 possède aussi un rôle très important dans la catalyse guidée et non guidée de aCBF5 (Fig. r26). La substitution de ce résidu en alanine abolit l'activité des particules mais son remplacement par une phénylalanine permet de restaurer celle-ci. De plus, ce résidu interviendrait dans le positionnement du substrat car nous avons montré qu'en son absence, le substrat reste bloqué dans la sRNP sans y être modifié (Fig. r26 étape 3 ou 4). Le résidu Y179 jouerait un rôle dans la discrimination du substrat par les sRNP H/ACA car le variant Y179A, et dans une moindre mesure les variants D82A et Y110A, permettent l'association stable de la particule avec un substrat dont le résidu U est remplacé par un C (Fig. r26). Ce complexe peut correspondre au complexe CII' dans lequel le substrat est positionné de façon optimale pour le



B

	wt	Variants du site actif de aCBF5						Variants de la boucle β7/β10 de aCBF5					
		D82A	Y110A	Y110F	Y179A	Y179F	R181A	R139A	P140A P141A	R143A	R143E	R143I	Δ140 -150
CII' 10 min	+	-	-	+	+	-	+	+	+	-	+	+	-
CII' 60 min	-	-	-	-	+	-	-	+	+	-	-	-	-
Turn-over sur gel	+	NT	NT	NT	-	NT	NT	+	-	+	+	+	NT
Activité guidée	+	-	-	+	-	+	+	+	-	+	+	+	-
Activité non-guidée	+	-	-	+	-	+	+	+	-	+	-	+	-
Discrimination de substrat	+	U/C	U/C	+	U/C	+	+	+	+	+	+	+	+
Etape bloquée ou influencée	**0**	**≤2**	**≤2**	**0**	**<4**	**0**	**0**	**4**	**<4**	**≥2**	**0**	**0**	**2**

Figure r26 : Récapitulatif des données expérimentales obtenues avec les variants de aCBF5 et de leur effets sur les étapes du schéma réactionnel de la réaction de pseudouridylation.

A. Schéma réactionnel de la réaction de pseudourdylation et matérialisation de l'étape affectée par les substitutions effectuées dans aCBF5 (inhibition : _|_ en bleu) ou (stimulation : ->+ en rouge).

B. Tableau récapitulatif des différents effets observés à partir des techniques employées pour l'étude des variants de aCBF5. NT signifie non testé, U/C correspond à la perte de la discrimination d'un ARN substrat dont le résidu U cible de la réaction est substitué par un résidu C, le numéro de l'étape bloquée ou influencée correspond à l'étape à partir de laquelle le variant de aCBF5 inhibe ou stimule la réaction de pseudouridylation (0 indique qu'aucune étape n'est influencée).

déroulement de la réaction catalytique. Il peut correspondre également au complexe intermédiaire CIIi' dans lequel le substrat est en interaction avec la poche de pseudouridylation mais le résidu cible n'est pas positionné dans le site actif de aCBF5. Il est également possible que les complexes formés avec le variant Y179A représentent d'autres intermédiaires de l'étape 2 ou 3 du mécanisme réactionnel.

Les résultats obtenus avec les substitutions des résidus Y110 et Y179 suggèrent que ce pourrait être l'aspartate catalytique joue le rôle de base générale, comme cela a été proposé pour TruB (Hamilton *et al.*, 2005).

Il a été proposé qu'en fin de la catalyse par TruB, l'interaction entre le résidu R181 et l'aspartate catalytique soit modifiée, conduisant à un réarrangement structural de l'enzyme qui pourrait se propager au domaine du pouce (Hoang *et al.*, 2005). Dans une structure 3D de la sRNP avec le substrat, le résidu R181 réalise des contacts électrostatiques avec l'aspartate catalytique (Fig. r27 A) (Duan *et al.*, 2009), comme dans TruB liée à l'ARNt (Hoang et Ferre-D'Amare, 2001; Pan *et al.*, 2003). Cela indique que le rôle joué par ce résidu dans aCBF5 pourrait être le même que dans TruB. Nos résultats montrent que ce résidu n'est pas essentiel pour les activités guidées et non guidées (Fig. r26).

4.2. Analyse des variants de la boucle β7/β10 de aCBF5

Parallèlement à nos études, plusieurs équipes ont observé que la délétion d'une partie ou de la totalité de la boucle du pouce de aCBF5 conduisait à une perte totale de l'activité (Duan *et al.*, 2009; Liang *et al.*, 2009) mais ne modifiait pas les propriétés de fixation de aCBF5 au sRNA (Hamma *et al.*, 2005). Il est proposé que la conformation de la boucle serait remodelée suite à la fixation de aGAR1 à aCBF5 et impliquerait un changement de conformation de l'ARN substrat (Liang *et al.*, 2008). Les structures de complexes sRNP-ARN substrat révèlent que de nombreux résidus basiques de la boucle β7/β10 de aCBF5 réalisent des contacts électrostatiques avec l'ARN substrat (Duan *et al.*, 2009; Liang *et al.*, 2009). Nos résultats confirment que la

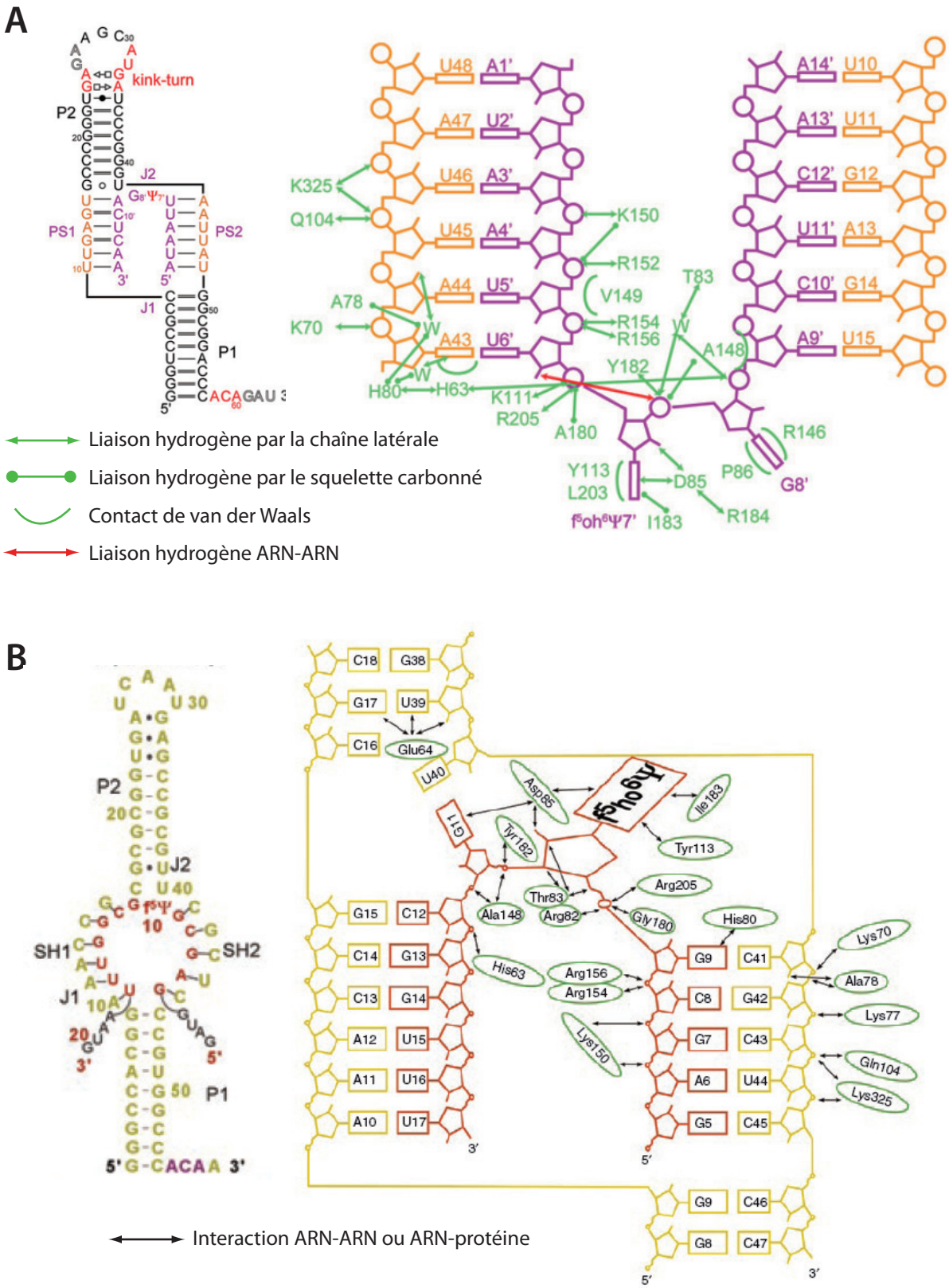


Figure r27 : Interactions obervées entre des résidus de aCBF5 et l'ARN dans deux structures 3D de sRNA H/ACA associés à un ARN substrat.

La numérotation des résidus de aCBF5 correspond à celle des résidus présents chez *P. furiosus*. Pour avoir la numérotation des résidus chez *P. abyssi*, il faut retrancher 3 à celle qui est indiquée.

A. Structure de la sRNP aCBF5/L7Ae/aNOP10-sRNA-ARN substrat de *P. furiosus* d'après Duan *et al.*, 2009.

B. Structure de la sRNP aCBF5/L7Ae/aNOP10-sRNA-ARN substrat de *P. furiosus* d'après Liang *et al.*, 2009.

Couleurs employées pour les sRNA (orange ou jaune) et pour les ARN substrats (violet ou rouge).

boucle de aCBF5, comprise entre les des résidus 140 à 150, est capitale dans l'activité guidée ou non guidée. En effet, nous avons montré que la délétion de la totalité de la boucle était dramatique pour l'activité des particules (Fig. r26).

Dans les structures 3D de aCBF5 en interaction ou non avec aGAR1, la chaîne latérale du résidu R139 de aCBF5 pointe vers aGAR1 (Duan *et al.*, 2009). Il avait été proposé que l'interaction de ce résidu avec aGAR1 pourrait orienter la boucle β7/β10 de aCBF5 (Rashid *et al.*, 2006). Cependant, il a été montré que la substitution de R139 en glutamine conduisait à un taux de résidus Ψ formé comparable à celui observé pour aCBF5 sauvage (Duan *et al.*, 2009). Nos résultats indiquent que le variant R139A n'affecte que peu ou pas l'activité et que les effets observés peuvent être compensés par l'addition de la protéine aGAR1 (Fig. r26). Ce résidu ne joue donc pas un rôle capital dans l'activité des particules et d'autres résidus de aCBF5 pourraient être impliqués dans l'interaction avec aGAR1 conduisant à orienter la boucle β7/β10.

Les structures 3D indiquent également que lorsque la boucle de aCBF5 est en position ouverte, les prolines 140 et 141 réalisent des interactions hydrophobes avec des acides aminés de aGAR1 (Li et Ye, 2006) et que lorsque la boucle est dans une conformation fermée, ces interactions ne sont plus retrouvées (Rashid *et al.*, 2006; Duan *et al.*, 2009; Liang *et al.*, 2009). Il avait donc été proposé que ces 2 prolines puissent contribuer à l'orientation de la boucle β7/β10 de aCBF5. Nous avons montré que la double mutation P140A P141A conduisait à une activité très faible qui n'est pas ou peu compensée par aGAR1 et que le substrat était bloqué dans la sRNP sans y être modifié (Fig. r26). Aucune interaction entre l'ARN substrat et ces deux résidus n'est observé dans les différentes structures 3D (Fig. r27). Nous proposons que compte tenu des propriétés structurales de ces résidus, ils jouent un rôle capital dans la conformation de la boucle β7/β10 de aCBF5. Ainsi, le changement de conformation du pouce provoqué par leur absence permet la formation d'un complexe stable entre la sRNP et l'ARN substrat mais la réaction catalytique n'a pas lieu (Fig. r26).

Une des deux structures 3D de la sRNP en présence de substrat indique que le résidu R143 (R146 chez *P. furiosus*) interagit avec la base du nucléotide non apparié situé directement en 3' du U cible de la réaction (Fig. r27 A) (Duan *et al.*, 2009). Nous avons montré que dans le système guidé, la substitution du résidu R143 en alanine conduisait à une formation plus rapide de résidus Ψ que pour aCBF5 sauvage (Fig. r26). Les substitutions R143I (Fig. r26) et R143Q (Duan *et al.*, 2009) ne modifient pas la cinétique de la réaction de modification par rapport à celle obtenue avec aCBF5 sauvage. Ce serait donc l'absence d'une chaîne latérale longue à cette position et donc la perte de la possibilité de contact avec le résidu en 3' du U ciblé qui conduirait à cet effet bénéfique sur la réaction. Par contre, dans le cas du système non guidé, les substitutions R143A et R143I n'affectent pas la cinétique de modification contrairement à la substitution R143E qui conduit à une baisse de l'activité non guidée (Fig. r26). Ceci souligne à nouveau que le mécanisme des réactions de pseudouridylation guidées et non guidées par l'ARN repose sur des déterminants différents.

D. Importance fonctionnelle des interactions établies, au sein de la sRNP H/ACA, entre aNOP10 et L7Ae et entre aNOP10 et l'ARN

1. Présentation du travail

Dès les premières expériences de reconstitution des sRNP H/ACA d'archaea, il est apparu que les protéines aNOP10 et L7Ae étaient des facteurs importants pour l'activité des particules (Baker *et al.*, 2005; Charpentier *et al.*, 2005). En effet :

- nous avons montré que l'hétérodimère aNOP10-aCBF5 représentait le couple minimal de protéine pour détecter, dans les conditions expérimentales que nous avions mises au point, la pseudouridylation d'un ARN substrat par l'activité de aCBF5 guidée par un sRNA (Baker *et al.*, 2005; Charpentier *et al.*, 2005), et nous avons démontré que des résidus de aNOP10 influençaient directement cette activité (article n°2, (Manival *et al.*, 2006)).

- en présence de L7Ae, l'activité guidée de aCBF5 est fortement stimulée (Baker *et al.*, 2005; Charpentier *et al.*, 2005) et celle-ci résulterait d'un effet à longue distance : la fixation de L7Ae au motif en K-turn s'accompagnerait d'un remodelage de la poche de pseudouridylation facilitant ainsi l'activité de la particule (Liang, B. *et al.*, 2007; Youssef *et al.*, 2007).

Alors que les deux protéines aNOP10 et L7Ae n'interagissent pas entre elles lorsqu'elles sont incubées ensemble en solution (Baker *et al.*, 2005; Charpentier *et al.*, 2005), une interaction aNOP10-L7Ae est observée dans les structures cristallines des sRNP libres ou en interaction avec l'ARN substrat (Li et Ye, 2006; Liang, B. *et al.*, 2007; Duan *et al.*, 2009; Liang *et al.*, 2009). Les interactions entre aNOP10 et aCBF5 d'une part et entre aNOP10 et L7Ae d'autre part contribuent à assurer une cohésion entre ces trois protéines qui sont empilées du côté concave de la courbure adoptée par la tige-boucle irrégulière d'un sRNA à simple motif. L'adoption de cette conformation repose sur deux points d'ancrage des protéines sur l'ARN : l'un correspond à l'interaction du domaine PUA de aCBF5 avec la boîte ACA et le second à l'autre extrémité du sRNA correspond à l'interaction de L7Ae avec le motif en K-turn (Li et Ye, 2006).

Nous nous sommes fixés comme objectif de tester l'importance structurale et fonctionnelle de l'interaction entre L7Ae et le sRNA et entre L7Ae et aNOP10. Des résidus de l'interface d'interaction entre ces deux protéines dans la sRNP ont été substitués en alanines dans le but d'abolir la possibilité d'association stable entre les deux protéines, sans affecter ni l'interaction de aNOP10 à aCBF5, ni celle de L7Ae avec le motif en K-turn. J'ai testé l'assemblage des particules avec les protéines variantes que j'ai produites et j'ai mesuré l'activité des sRNP variantes. Les particules assemblées ont été également caractérisées sur le plan structural. J'ai exploité les propriétés de réactivité chimique du plomb vis-à-vis de l'ARN pour i) détecter les changements structuraux du sRNA lors de l'assemblage des protéines en différents sous-complexes d'une sRNP, ii) sonder la conformation du sRNA dans les sRNP portant les protéines variantes. En parallèle Magali Blaud a étudié la conformation du sRNA dans l'assemblage séquentiel de la sRNP et dans les sRNP variantes par dichroïsme circulaire. Ces analyses apportent des informations quant à la contribution des différentes protéines sur la conformation de l'ARN guide et elles nous ont permis de vérifier que dans les sRNP, les protéines L7Ae et aNOP10 variantes n'interagissaient pas entre elles. Ainsi les effets observés sur l'activité étaient en lien avec la perte de l'interaction aNOP10-L7Ae.

D'autre part, il a récemment été montré qu'une mutation dans le gène *NOP10* humain conduisait à la maladie génétique humaine nommée la Dyskératose Congénitale (Walne *et al.*, 2007) (cf IV.C.8.3.1 de l'introduction). Il s'agit de la substitution en tryptophane du résidu arginine 34 qui est un résidu parfaitement conservé chez les protéines NOP10 d'eucaryotes et d'archaea. Dans la structure 3D de la sRNP d'archaea, ce résidu établit des contacts avec l'hélice H2 du sRNA (Li et Ye, 2006). Compte tenu du rôle crucial de aNOP10 pour l'activité des sRNP H/ACA, nous avons testé l'importance du résidu R34. J'ai produit les protéines variantes aNOP10 portant les substitutions R34W et R34A et nous avons analysé l'effet de cette substitution sur l'assemblage, l'activité et la structuration des particules.

Les données obtenues montrent que :

- l'activité de la particule dépend fortement de l'interaction entre les protéines aNOP10 et L7Ae. La mise en place de cette interaction apparaît être cruciale pour l'interaction stable de l'ARN substrat avec la particule. Il s'agit donc d'un déterminant important contribuant aux effets bénéfiques de L7Ae sur l'activité des particules.

- la protéine aNOP10 est un constituant important des particules. En plus de son rôle stimulateur de l'activité de aCBF5, son interaction avec L7Ae et avec l'ARN contribue à la conformation des particules.

- la mutation R34W n'affecte que très peu l'activité des RNP H/ACA alors que les données de dichroïsme circulaire indiquent que la structure secondaire du sRNA est modifiée au sein des particules variantes. Nos données suggèrent donc que l'activité des particules snoRNP H/ACA ne serait pas modifiée chez un patient porteur de la forme AR de la Dyskératose congénitale.

Ces données ont été rédigées sous forme d'un article qui va être soumis pour publication et dont je suis premier auteur.

Publication n°4 :

Interactions between proteins aNOP10 and L7Ae organize archaeal box H/ACA sRNPs for stable substrate RNA binding and optimal RNA:pseudouridine-synthase activity.

Jean-Baptiste FOURMANN*, Magali BLAUD*, Christiane BRANLANT and Bruno CHARPENTIER

* Contribution équivalente au travail

En préparation pour soumission à publication.

Interactions between proteins aNOP10 and L7Ae organize archaeal box H/ACA sRNPs for stable substrate RNA binding and optimal RNA:pseudouridine-synthase activity

J.–B. Fourmann*, M. Blaud*, C. Branlant, and B. Charpentier[§]

AREMS, Nancy Université, UMR 7567 CNRS-UHP Nancy I, Faculté des Sciences et Techniques, Bld des Aiguillettes, BP 239, 54506 Vandoeuvre-lès-Nancy Cedex. France.

* Equal contribution

§ to whom correspondence should be addressed

Tel : 33 3 83 68 43 16

Fax : 33 3 83 68 43 07

Email : bruno.charpentier@maem.uhp-nancy.fr

Abstract

Multiple RNA-guided pseudouridine synthases, organized as H/ACA ribonucleoprotein particles (RNPs) containing a guide sRNA and four proteins, catalyze site-specific post-transcriptional isomerization of uridines into pseudouridines. In a fully assembled archaeal particle, the sRNA is anchored by the pseudouridine synthase aCBF5 and ribosomal protein L7Ae. Here, we analyzed the role of interactions formed by the third protein aNOP10 with L7Ae and with the sRNA for RNP assembly and activity. aNOP10 was found to influence conformation changes occurring in the sRNA during a step by step RNP reconstitution which was monitored by chemical probing and circular dichroism spectroscopy analyses. Our data support a model in which aNOP10-L7Ae and aNOP10-sRNA interactions architecture the sRNP for stable binding of the substrate RNA. Nonetheless, the intrinsic catalytic activity of sRNPs is only dramatically affected by lack of the aNOP10-L7Ae interaction.

INTRODUCTION

Pseudouridine (Ψ) is the most common form of modified ribonucleotide found in non-coding RNAs. A large number of Ψ residues is located in functional regions of transfer RNAs (tRNAs), ribosomal RNAs (rRNAs) and spliceosomal small nuclear RNAs (UsnRNAs). While the precise role played by most Ψ residues is not clearly defined, some Ψs that are conserved in rRNAs were demonstrated to be critical for ribosome function [1-5] and one highly conserved Ψ residue in U2 snRNA was found to be crucial for spliceosome biogenesis [6-8].

Site-specific post-transcriptional isomerization of uridine (U) residues into Ψ residues is catalyzed by a variety of specific RNA:Ψ-synthases [9-12]. In bacteria, tRNAs and rRNAs are exclusively modified by single-polypeptide enzymes carrying both RNA recognition capability and RNA modification activity (for review, [13]). As an alternative to protein stand-alone enzymes, numerous catalysts achieve U to Ψ conversions in eukaryal rRNAs and UsnRNAs, and in archaeal rRNAs. They consist in multiple distinct H/ACA ribonucleoprotein particles (RNPs) which are denoted H/ACA small nucleolar (sno)RNPs in eukaryotes and H/ACA sRNPs in archaea (for review, [14,15]). Each RNP comprises a unique non-

coding RNA (snoRNA in eukaryotes or sRNA in archaea) and a common pseudouridine synthase (Cbf5 in yeast, NAP57 or dyskerin in vertebrates, aCBF5 in archaea). H/ACA RNPs contains also three other essential and evolutionary conserved core proteins, denoted Nop10/NOP10, Gar1/GAR1, and Nhp2/NHP2 in H/ACA snoRNPs, and respectively, aNOP10, aGAR1, and L7Ae in H/ACA sRNPs (for review, [15-17]). In this RNP-based modification system, the RNA component functions as a guide and forms base pair interactions with a target RNA sequence in order to specify the residue to be modified by the RNA:Ψ-synthase.

Most of our information on the structure and function of H/ACA RNPs stems from works performed on the archaeal particles. The H/ACA sRNAs identified in archaeal genomes share common features with eukaryal snoRNA but differ by their heterogeneous architectures [18-22]. Despite a great variety of length and sequences, the vast majority of eukaryal H/ACA snoRNAs fold into a conserved structure comprising two basic units linked by a single-stranded ANANNA sequence (the H box) and featuring a single-stranded ACA sequence (the ACA box) at their 3' tail [23,24]. The archaeal H/ACA sRNAs contain often only one but up to three characteristic basic units, and a single-stranded ANA sequence (ACA being the most frequently observed) separating each units and present at their 3' tail (the ACA box). As well in snoRNAs as in sRNAs, each basic unit consists in an irregular hairpin interrupted by a large internal loop featuring a lower helix P1, an internal loop called a pseudouridylation pocket, an upper helix P2, and an apical loop (as displayed in **Fig. 1 and S1**). An apical K-turn or K-loop motif is present on top of helix P2 in all the archaeal basic units (**Fig. 1 & S1**). Two sequences (s1 and s2) on both sides of the internal loop establish two distinct base-pair interactions (SH1 and SH2) with sequences flanking the target U. In such complex, the U to be modified and its 3' contiguous residue are unpaired resulting in a three-helix junction around the exposed 5'-UN-3' di-nucleotide [25-27]. Recently, solution structures of pseudouridylation pockets associated or not to a substrate RNA were solved by NMR spectroscopy [28,29]. In these studies, RNA duplexes were formed in absence of proteins by annealing RNA oligoribonucleotides derived from substrate rRNA sequences with a pseudouridylation pocket carried by shortened hairpins derived from the 5' or 3' basic units of the human U65 H/ACA snoRNA. The obtained complexes display a unique structure in solution with the substrate in a U- or Ω-shape interacting on one face of the snoRNA.

Box H/ACA sRNPs can be assembled *in vitro* by using recombinant proteins purified from *E. coli* and *in vitro* transcribed guide sRNAs. Reconstituted sRNPs pseudouridylate substrate RNAs consisting in simple target oligoribonucleotides [18,19,30]. Information on the role of the core proteins and the conserved RNA elements for sRNP assembly and activity [18,19,31,32] was gained by using this reconstitution methodology. Only proteins aCBF5 and L7Ae were found to interact directly and stably with the sRNA [18,19]: L7Ae binds the apical K-turn or K-loop motif, while aCBF5 requires the ACA tri-nucleotide for association to the sRNA. In absence of sRNA, aCBF5 forms an heterotrimer by interacting with proteins aNOP10 and aGAR1 [18,19,33]. Hence, binding of aCBF5 to the guide sRNA is required for recruitment of proteins aNOP10 and aGAR1 within the particles. Although the RNA-guided activity of aCBF5 relies on its interaction with aNOP10,

sub-sRNPs formed with the pair of proteins aCBF5-aNOP10 display only low activity [19,31,34]. Addition of L7Ae to this minimal proteins set allows formation of fully-active sRNPs [18,19]. The fourth protein aGAR1 is not required for total modification of substrate RNA, but its presence reinforces the kinetics of Ψ formation, probably by facilitating reaction turnover [19].

Crystal structures have been obtained for the RNA-free and –bound forms of L7Ae [35-37], for aNOP10 [38], for isolated heterodimeric and heterotrimeric complexes formed between the other core proteins, i.e. aNOP10-aCBF5 and aNOP10-aCBF5-aGAR1 [31,33,39], and, for a complete H/ACA sRNP including the four proteins and a single-hairpin H/ACA sRNA [40]. In the later, proteins L7Ae, aNOP10, and aCBF5 are stacked on each other, with aNOP10 sandwiched by the two other proteins. This protein platform binds one side of the H/ACA sRNA. In accordance with biochemical data, protein aCBF5 is anchored to the ACA sequence by its PseudoUridine synthase and Archaeosine transglycosylase (PUA) domain whereas its catalytic domain contacts the pseudouridylation pocket, protein aNOP10 is bound to the catalytic domain of aCBF5, and protein L7Ae is tightly associated with the apical K-turn. Two features were not observed in biochemical studies: L7Ae contacts aNOP10 and few residues of aNOP10 contact sRNA upper helix P2. In the RNP, the overall sRNA helical axis is slightly bent with the proteins facing the concave side. This structure is preserved in the 3D structure of the sRNP associated with the substrate RNA [41,42].

Despite these advances, molecular bases of the aforementioned crucial roles played by proteins aNOP10 and L7Ae for H/ACA sRNP activity are not clarified. Structural data on aCBF5-aNOP10 and aCBF5-aNOP10-aGAR1 complexes suggested that aNOP10 might organize the catalytic center of aCBF5 [33,39] and aNOP10 residues were indeed found to influence aCBF5 activity [34]. RNA binding activity of L7Ae was suggested to remodel the sRNA pseudouridylation pocket [32], and protein L7Ae was also proposed to have a crucial role in proper placement of the substrate RNA in the catalytic center [43,44].

Here, we studied the interactions engaged by aNOP10 with L7Ae and with the guide sRNA. We evaluated their respective role for sRNP assembly, substrate RNA binding and activity. In particular, we tested the role of aNOP10 residue R34, which contacts upper helix P2 in sRNP and R34W substitution is associated with a form of dyskeratosis congenita (DC). Conformation of the sRNA in various sub-RNPs and in sRNPs with a mutated aNOP10-L7Ae interface was investigated by chemical probing and circular dichroism (CD) spectroscopy.

RESULTS

Substitution of residues present in the aNOP10-L7Ae interface impairs activity of H/ACA sRNPs

In the sRNP crystal structure, aNOP10 helix α1 contacts L7Ae helix α3 [40] (**Fig. 1a**). L7Ae residues T45, A69, H70, P73, and L74 constitute an hydrophobic surface for stacking of aNOP10 residues Y41 and Y44. In addition, L7Ae residue E77 is proposed to form an ionic interaction with aNOP10 residue R47. aNOP10 and L7Ae residues that are engaged in the protein interface were substituted by alanines in order to test their importance for sRNP function. This end, sRNPs were reconstituted by incubation of a set of the wild type proteins aCBF5,

aNOP10, and L7Ae with the *in vitro* transcribed *P. abyssi* sRNA Pab21 (**Fig. 1b**) or Pab91 (**Fig. S1**). The Ψ-synthase activities of these wild-type sRNPs were compared with the ones of sRNPs carrying one protein variant of either L7Ae or aNOP10 (**Fig. 1c** and **S1c**). Substitution in aNOP10 of residue R47 (aNOP10 variant R47A) or of the two tyrosines Y41 and Y44 (aNOP10 variant Y41AY44A), or substitution in L7Ae of the three residues H70, L74, and E77 (L7Ae variant H70AL74AE77A) affected strongly the time course of substrate modification by the sRNAs. Effect of the mutations on the sRNP activity was not due to important modification in the folding of variant proteins. Indeed, CD spectra of proteins with alanine substitutions were not changed as compared with the ones of wild type proteins (data not shown). Previous works showed that aCBF5 allows *in vitro* modification of position U55 in tRNAs, i.e. the position modified in *E. coli* tRNAs by TruB [12,34]. Importantly, presence of aNOP10 becomes crucial for Ψ55 formation in a tRNA lacking the CCA 3' extension [34]. Here, we observed that aNOP10 variants R47A and Y41AY44A were able to stimulate the non RNA-guided activity of aCBF5 at the same rate as the wild-type aNOP10 (**Fig. S2a-b**). Since L7Ae does not influence such non RNA-guided activity [34], data obtained with the aNOP10 variants were consistent with the fact that the aNOP10-L7Ae interface is not involved in the non RNA-guided activity. Hence, the important role played by aNOP10 residues Y41, Y44, and R47 was only observed in presence of L7Ae, i.e. in the context of H/ACA sRNPs.

Chemical probing of sub-sRNPs

We next intended to analyze the contribution of each protein aNOP10 and L7Ae, and of their interactions on sRNP conformation. As a first approach, we used $Pb2^{+}$ which mostly induces cleavages in the RNA phosphate backbone of single-stranded regions and allows the sensitive detection of conformational changes in RNA molecules [45]. Formation of high amounts of an homogeneous population of protein-RNA complexes is a prerequisite for probing experiments. *P. abyssi* guide sRNAs Pab21 (**Fig. 1**) and Pab91 (**Fig. S1**) share a single-hairpin conformation with an apical K-loop motif which is efficiently recognized by protein L7Ae, as shown by EMSA experiments ([19], and see lanes 2 in **Fig. 1d** & **S1b**). Interestingly, while only low amounts of RNA-protein complexes are formed upon incubation of Pab91 with aCBF5 alone or in combination with aNOP10 and L7Ae (**Fig. S1b** and [19]), high amounts (>85%) of RNA-protein complexes were formed with sRNA Pab21 (**Fig. 1d**, lane 4). Therefore, the various sub-complexes of the Pab21 sRNP were formed by incubation of various proteins sets with sRNA Pab21 radio-labeled at its 5'- (**Fig. 2a**) or 3'-end (**Fig. 2b**), and we analyzed accessibility of the sRNA to hydrolysis induced by Pb (II) acetate. In absence of proteins, the strongest cleavages occurred in phosphodiester bounds between residues predicted to be single stranded: i.e. between residues A33, A34, U35, and C36 in the apical loop of the K-loop motif, between residues of the pseudouridylation pocket C17, C18, and C19 of the s1 strand, and U50, C51, C52 of the s2 strand (lanes 4). Three strong cleavages between residues from positions 24 to 27 were also observed in the lower part of helix P2 within the P2a strand. Finally, another feature of

the naked Pab21 sRNA was the accessibility of its ACA box to lead with a strong cleavage between residues C64 and A65.

By comparison with the cleavage profiles obtained with the naked Pab21 sRNA, presence of L7Ae (lanes 5, scheme L in panel c) conducted to strong protections of the cleavages occurring in the apical K-loop motif (KTa and KTb sequences). These data are in accord with the binding of L7Ae to the K-loop motif. Beside these protections, L7Ae RNA binding conducted also to an enhancement in the efficiency of lead cleavage in the apical loop between residues A34, U35, C36. The lower part of stem P2 is probably slightly destabilized in the L7Ae-sRNA complexes since two new but low abundant cleavages occurred between residues G46, G47, and G48.

In the aCBF5-Pab21 complex (lanes 7, scheme C in panel c), the ACA box was found less accessible to lead cleavages. This was awaited since the PUA domain of aCBF5 binds the ACA motif and the lower stem P1 [31,40,43,46]. We also observed a protection of the cleavages between residues U50, C51, and C52 of strand s2. These data are totally in accord with a footprint of the catalytic domain of aCBF5, which is positioned on the pseudouridylation pocket [46]. In addition, cleavages in strand s1 between residues C15, U16, and C17 were more abundant, suggesting that the folding of the pseudouridylation pocket is modified upon aCBF5 binding. The nature of the interactions taking place in this sub-sRNP between residues of the aCBF5 catalytic domain and the pseudouridylation pocket is an open question since aCBF5-sRNA contacts were only identified in the crystal structure of the fully assembled sRNP [40] free or containing the substrate RNA [41-43]. Moreover, additional contacts are proposed to take place in the aCBF5-sRNA sub-complex as compared to the fully assembled sRNP [46].

In presence of the two RNA binding proteins L7Ae and aCBF5 (lanes 8, scheme LC in panel c), the pattern of cleavages in the upper part of sRNA Pab21 was found unchanged as compared with the ones observed for the L7Ae-Pab21 complex. Probing features of the lower part of sRNA observed in the aCBF5-Pab21 complex were also conserved in the L7Ae-aCBF5-Pab21 complex, confirming binding of both proteins L7Ae and aCBF5 to the sRNA.

In presence of the aCBF5-aNOP10 dimer (lanes 10, scheme CN in panel C), the cleavage pattern was basically the same as the one observed in presence of aCBF5 alone. We noticed only a slight protection of the cleavage occurring between residues C19 and G20 in strand s1. This subtle change in the cleavage pattern within the pseudouridylation pocket could correspond to an indirect effect related to the folding of the aCBF5 catalytic center upon its interaction with aNOP10 [33,34,39].

In the sub-sRNP formed in presence of the proteins set L7Ae, aCBF5, and aNOP10 (lanes 13, scheme LCN in panel c), a limited number of cleavages were still detected. Only the cleavages occurring in the apical loop and cleavages in strand s1 at the junction with helix P1 were preserved. Hence, presence of aNOP10 together with the two RNA binding proteins aCBF5 and L7Ae conducted to significant changes in the accessibility of the probe to the sRNA. This could be related to formation of the contacts between proteins aNOP10 and L7Ae that were observed within the crystal structure of the fully assembled sRNP. Based on these data, we postulated that formation of an interaction between proteins aNOP10 and L7Ae could be monitored by analyzing disappearance of the Pb^{2+}-induced

cleavages in strand P2a at the basis of helix P2, and in the two strands s1 and s2 of the pseudouridylation pocket (marked by orange dots along the gels in **Fig. 2a**-**b**).

Conformation of the sRNA is influenced by the presence of aNOP10 and L7Ae within sub-RNPs

We next used far U.V. circular dichroism (CD) spectroscopy in order to analyze conformation changes taking place in the sRNA during a sequential reconstitution of sub-sRNPs with proteins aCBF5, aNOP10, and L7Ae (**Fig. 3**). Indeed, CD is a powerful method to monitor changes in RNA structure in solution [47] and because the significant spectral features of the RNA and protein occur at different wavelengths, this technique can be used to examine structural changes in either an RNA or a protein upon complex formation (Fasman, 1996). The CD spectrum of free sRNA Pab21 displayed a positive signal in the 250-300 nm range with a maximal $\Delta\varepsilon$ value at a wavelength of 265 nm which is referred as the λmax wavelength. These features indicate a significant folding of the sRNA in a A-form helix. In presence of the proteins, CD spectra profiles were slightly modified which is indicative of changes in the tertiary structure of the sRNA in the various sub-RNPs but with a preservation of the helical elements. Whereas the λ max of the signal was unchanged upon L7Ae binding, the amplitude of the peak was reduced ($\delta(\Delta\varepsilon)$ = -0.81 $M^{-1}.cm^{-1}$ per nt) (**Fig. 3a**, **3c**). The opposite occurred upon aCBF5 binding since the λ max was displaced from 265 to 267.2 nm but the amplitude of the peak was unchanged as compared to the spectrum obtained with the naked sRNA (**Fig. 3a**). Modification of sRNA conformation was not surprising for these two RNA binding proteins. Yet, their respective binding influenced differently sRNA conformation. We explain the CD changes in presence of L7Ae as a result of the folding of the K-turn that occurs upon L7Ae binding [48]. In the archaeal sRNP H/ACA, aCBF5's PUA domain is anchored to the lower part of the sRNA and stability of the aCBF5-sRNA complex probably relies also on contacts of the catalytic domain with helix P2, which basal part is wrapped around the catalytic domain in H/ACA sRNPs 3D structures [40,43]. Given that most of the contacts established between aCBF5 and the sRNA are conserved in the aCBF5-sRNA complex [46] and since conformation of helix P1 axis was never found particularly changed upon aCBF5 binding, we hypothesized that the major conformational effects linked to aCBF5 sRNA binding came from a modification of the tertiary structure of helix P2.

Interestingly, binding of aNOP10 to the Pab21-aCBF5 complex conducted both to a displacement of the λ max from 267.2 to 268.4 $M^{-1}.cm^{-1}$ (**Fig. 3a**) and to a reduction in the $\Delta\varepsilon$ value ($\delta(\Delta\varepsilon)$ = -1.6 $M^{-1}.cm^{-1}$ per nt) (**Fig. 3c**). Hence, presence of aNOP10 likely induced an additional modification of helix P2 conformation as compared to the one present in the aCBF5-Pab21 sub-complex.

Addition of L7Ae to the aCBF5-aNOP10-Pab21 complex resulted in an additional increase in λmax from 268.4 to 269 nm, but more remarkably in the additional decrease of the RNA signal ($\delta(\Delta\varepsilon)$ = -1.21 $M^{-1}.cm^{-1}$ per nt). Since we showed that binding of L7Ae to the K-loop motif on the free Pab21 and to the Pab21-aCBF5 complex induced a lower decrease ($\delta(\Delta\varepsilon)$ = - 0.81 and = - 0.78 $M^{-1}.cm^{-1}$ per nt, respectively), we concluded that the effect of the presence of L7Ae on the sRNA conformation relied on the

presence of aNOP10. We concluded that interaction between L7Ae and aNOP10 stabilize sRNA conformation. In addition or alternatively this interaction could also favor the positioning of aNOP10 within the sRNP.

Residues engaged in the aNOP10-L7Ae interface are important determinants of the sRNP conformation

Having evidences on the crucial role played by aNOP10 and L7Ae to architecture sRNP, we next tested effects of the alanine substitutions in residues of the aNOP10-L7Ae interface on: i) the assembly of the various Pab21 sub-sRNPs (**Fig. 1d**), ii) the accessibility of the sRNA toward Pb^{2+}-induced cleavages (**Fig. 2**), iii) the sRNA conformation established by CD analysis (**Fig. 3b-c**). The various complexes assembled with the wt or mutated proteins were formed at similar amounts. We noticed a slight modification in the electrophoretic mobility of the sub-RNPs assembled with aCBF5 and aNOP10 variant R47A, which migrated faster than the wild-type sRNP (**Fig. 1d**, lane 9). Chemical probing analyses showed that for the L7Ae-, aCBF5-, L7Ae-aCBF5-, and aCBF5-aNOP10-sRNA complexes, substitution of the interface residues did not modify the cleavage pattern of the Pab21 sRNA (**Fig. 2a-b**, lanes 6, 9, 11-12). But remarkably, the probing signature characteristic of the L7Ae-aCBF5-aNOP10-Pab21 complex was not observed in presence of either the L7Ae or the aNOP10 variants. Indeed lead still induced cleavages in the s2 and P2a strands of the variant sRNPs, but not anymore in strand s1 (**Fig. 2a-b**, lanes 13-16). Accordingly, CD spectra established for these variant sRNPs display $\delta(\Delta\varepsilon)$ values comparable as the one measured for sRNPs containing L7Ae and aCBF5 but lacking aNOP10 (**Fig. 3d**). In view of all these data, we concluded that the substitutions of aNOP10 residues R47 ,or Y41 and Y44, or of L7Ae residues H70, L74, and E77 impair the aNOP10-L7Ae interaction resulting in a change of sRNA conformation within the L7Ae-aCBF5-aNOP10-sRNA sRNP.

Conformation of an sRNP assembled with aNOP10 carrying a substitution found in a form of dyskeratosis congenita

As shown by the CD analyses, binding of aNOP10 to the Pab21-aCBF5 complex modifies the sRNA conformation (**Fig. 3a**, **3c**). Such conformational effect could be explained by formation of an interaction between aNOP10 and helix P2 in the fully assembled sRNP [40], or by an indirect effect based on a stabilization of the catalytic domain of aCBF5 influencing its interaction with the sRNA.. In order to collect more evidence in favor of one of these two possibilities, we substituted residue R34 in aNOP10 which contacts helix P2 in the crystal structure (**Fig. 1a**). Interestingly, some patients with the classical form of the rare inherited disorder dyskeratosis congenita (DC) express a mutated R34W NOP10 protein [49]. Hence, we next tested the impact of the R34W substitution in the archaeal aNOP10 protein on CD spectra of Pab21 sub-sRNPs (**Fig. 3c-d**). Unlike the wild type aNOP10, addition of the aNOP10 variant R34W to the Pab21-aCBF5 complex did not change the RNA positive signal of the CD spectrum (**Fig. 3c**). Addition of L7Ae to the Pab21-aCBF5-aNOP10R34W complex conducted to a slight reduction of the $\delta(\Delta\varepsilon)$ value (~ 0.75 $M^{-1}.cm^{-1}$ per nt). Comparable results were obtained with variant R34A. From these data we concluded that the aNOP10-RNA contact mediated by residue R34 helps the sRNA. to

acquire its conformation in the fully assembled particle. Interestingly, interactions between aNOP10 variants R34W or R34A and L7Ae are preserved since the probing signature of the LCN complex is observed (**Fig. 2d**). Nevertheless, particles with the modified RNA conformation induced by R34 substitutions are able to pseudouridylate the substrate RNA albeit with a slower activity (**Fig. 4a** and **S1f**).

aNOP10-L7Ae and aNOP10-sRNA interactions help stable binding of sRNP's substrate RNA

We previously showed that aNOP10 was crucial for binding of the substrate RNA to sRNPs containing protein aCBF5 [19]. We next investigated the role of residues of the aNOP10-L7Ae interface and residue R34 for the binding of the substrate RNA to sRNPs assembled with proteins L7Ae, aCBF5, and aNOP10. Detection of the interaction between sRNPs and an oligoribonucleotide representing their substrate RNA [19] was done by EMSA (**Fig. 4** and **S1**): a retarded CII' complex was formed upon incubation of the radio-labeled substrate RNA S (**Fig. 1b**) with the unlabeled L7Ae-aCBF5-aNOP10-Pab21 sRNP (**Fig. 4b**, lane 2). The amount of CII' complexes were reduced ~2-fold for sRNPs carrying mutations in residues engaged in the aNOP10-L7Ae interface (**Fig. 4a**, lanes 3-5). These data are in good agreement with the relative effects of the mutations on sRNPs activities (**Fig. 1c** & **S1c**). Surprisingly, contrary to variants of the aNOP10-L7Ae interface, the amounts of CII' complexes formed by sRNPs assembled with aNOP10 variants R34A or R34W did not parallel the Ψ-synthase activities. Indeed, substitution of residue R34 had a mild effect on sRNPs activity (**Fig. 4a** and **S1f**) despite detection of very low amounts of CII' complexes (**Fig. 4b**, lanes 6-7). Since these mutated sRNPs displayed high activity and since amounts of complex CII' remain low when the particle is inactive, i.e. when catalytic residue D82 is mutated (Fig. **4b**, lanes 13-14), we concluded that the two R34 mutations affected the stability of CII' complexes, not their formation.

DISCUSSION

Biochemical and crystallographic data on H/ACA RNPs allowed progress in our understanding of the molecular mechanism of the RNA-guided Ψ-synthase activity. Whereas aCBF5, aNOP10, and L7Ae were early found to represent the minimal set of proteins for efficient *in vitro* activity of archaeal H/ACA sRNPs [18,19], their respective and cooperative contribution are still unclear. In the recently established crystal structures of a fully assembled particle, the helical axis of the guide sRNA is bent toward the protein complex [40-42]. In such context, the two antisense elements of the pseudouridylation pocket are exposed at the convex side of the sRNA. In this work we performed lead probing and CD analyses and observed different conformation of the sRNA in various sub-complexes. The principal finding is that addition of aNOP10 to a aCBF5-sRNA complex substantially influences sRNA conformation (Fig. 3). The crystal structure of a sRNP lacking both a K-turn motif and protein L7Ae, showed that helix P2 is differently oriented as compared with the fully assembled sRNP [43]. This sub-RNP with an inaccurately orientated helix P2 allows substrate RNA binding but not with a correct positioning for its modification [43]. Hence, change in helix P2

orientation and misplacement of the substrate RNA were proposed to result from the absence of protein L7Ae [43]. Importantly, addition of L7Ae to a aNOP10-aCBF5-sRNA complex conduct to another sRNA conformation changes (**Fig. 2** & **3**). Since such conformation change was not observed when interaction between aNOP10 and L7Ae was impaired, we propose that the aNOP10-L7Ae interaction taking place within the sRNP likely locks the sRNA in the conformation found in the sRNP's crystal structure. Importantly, our data demonstrate that the sRNA conformation relying on the aNOP10-L7Ae interaction is optimal for sRNP activity (**Fig. 1c**) and support a model in which binding of the substrate RNA relies on proper conformation of the sRNA. Binding of L7Ae to the apical K-turn motif was proposed to indirectly influence substrate RNA binding by remodeling sRNA pseudouridylation pocket[32]. This was deduced from probing data obtained upon binding of L7Ae on the *P. furiosus* sRNA Pf9. We did not observe changes in the pattern of lead cleavages occurring in strands s1 and s2 of the L7Ae-Pab21 sRNA sub-complex (**Fig. 2**, lanes 5), suggesting that binding of L7Ae to sRNA Pab21 has no detectable effect on the folding of the pseudouridylation pocket. We checked whether this was also the case for the sub-complexes formed between L7Ae and Pab160, i.e. the functional homolog of Pf9 in *P. abyssi*, as well as with two other *P. abyssi* sRNAs Pab91 and Pab19 (**Fig. S3**). Interestingly, only binding of L7Ae to Pab160 conducted to more abundant cleavages in strand s2 (**Fig. S3**, lane 15). Hence, remodeling of the pseudouridylation pocket might not be a general feature of the binding of L7Ae to a guide sRNA. Although it is still possible that L7Ae, by itself, has a beneficial effect for substrate binding, we propose that for the majority of the sRNAs tested, the positive effect of L7Ae on sRNP activity is dependent on the presence of aNOP10.

It is likely that correct positioning of the substrate RNA observed in the crystal structure [41,42] relies on the aNOP10-L7Ae. Indeed, recent data showed that presence of L7Ae and aNOP10 favor proper docking of the substrate RNA [44] but the role of aNOP10-L7Ae interaction *per se* was not tested for such effects. Only low amounts of stable sRNP-substrate RNA complexes could be observed in absence of aNOP10-L7Ae interaction (**Fig. 4**). This is contradictory to other studies for which binding of the proteins to a preformed sRNA-substrate RNA duplex conducts to formation of a stable sRNP-substrate complex [43]. Here, we employed a different methodology in which binding of the substrate RNA necessitates an assembled sRNP [19,30]. It is possible that stable binding of the substrate RNA to a pre-formed sRNP requires its correct positioning in the active site. Hence, a preformed sRNP with inadequate conformation cannot stably bind the substrate RNA.

Is the interaction of aNOP10 with L7Ae a general feature of the RNA guided system? The three residues H70, L74 and E77 we showed to be important for aNOP10 binding are conserved in NHP2 which replaces L7Ae in eukaryal H/ACA snoRNPs. The current model of H/ACA snoRNP assembly proposes that a Dyskerin(CBF5)/NOP10/NHP2 complex is formed in the cytoplasm and associates with the assembly factor NAF1 before its loading on the nascent snoRNAs [50]. This raises the question of whether interactions between NOP10 and NHP2 are formed in the context of the snoRNP and/or in the RNA unbound state of the protein complex.

We observed that substitution of aNOP10 residue R34 preserves formation of the aNOP10-L7Ae interaction (**Fig. 3**) and conducts also to a change in sRNP conformation. Surprisingly, sRNPs remain active suggesting that intrinsic activity of these variant sRNPs is not changed as compared with the wt sRNPs. Surprisingly, only low amounts of stable sRNP-substrate RNA complexes could be characterized. Either a low fraction of the variant sRNPs can form active complexes and after several reaction turnovers the substrate population can be modified, or despite an accurate positioning of the substrate in the variant sRNP, the complexes are not stable enough to be isolated by EMSA. Importantly, it would be of interest to verify the low importance of the aNOP10-RNA interaction for sRNPs activity within the cells and whether this feature is conserved for eukaryal H/ACA snoRNPs. This is important in regards of the association existing between R34W substitution and dyskeratosis congenita [49]. It remains also be investigate whether the shortened telomeres and the low levels of telomerase RNA component found in homozygous R34W patients, could result from an improper conformation of the H/ACA RNP motif present in the telomerase.

METHODS

H/ACA sRNP reconstitution, EMSA and *in vitro* pseudouridylation assay

As previously described [30], sRNPs were reconstituted by incubation of *in vitro* transcribed and ^{32}P end-labeled sRNA (~50 fmol) with purified recombinant proteins (200 nM each) in buffer D [150 mM KCL; 1.5 mM $MgCl_2$, 0.2 mM EDTA and 20 mM HEPES (pH 7.9)] at 65°C for 10 min. To test for the association of the sRNA with its substrate RNA S, formation of complex CII' was analyzed [19,30] formation). To this end, sRNPs were assembled by incubation of unlabeled sRNA (2 pmol) with labeled oligoribonucleotide S (150 fmol) in presence of the proteins (20 nM each). Complexes were visualized by autoradiography after fractionation by non-denaturing polyacrylamide gel electrophoresis.

Conditions for time course analysis of the RNA:Ψ-synthase activities of the reconstituted sRNPs were previously detailed [30]. Briefly, 4 pmol of unlabeled sRNA and 150 fmol of substrate RNA that was labeled by incorporation of [α-^{32}P]CTP or [α-^{32}P]UTP in the course of transcription, were mixed at 65°C with the various sets of protein. Aliquots were collected at different time points and the reaction was stopped by phenol-chloroform extraction followed by ethanol precipitation. The recovered substrate RNAs were digested with RNase T2. The resulting 3'-mononucleotides were fractionated by a monodimensional thin layer chromatography. The radioactivity in the bands was quantified with a phosphorimager, by using the ImageQuant software. Amounts of the formed Ψ residues were determined taking into account the total number of U residues in the sRNA's substrate RNA.

Chemical probing

sRNP were assembled on ^{32}P-end labeled sRNA in the same conditions as described in the preceding paragraph. sRNAs were labeled either at their 5' end with T4 polynucleotide kinase and [γ-^{32}P]ATP or at their 3' end with T4 RNA ligase and ^{32}pCp. For lead (II) footprinting,

the reactions were initiated by addition of 40 mM Pb(II) acetate (Merck) freshly prepared in sterile water and incubated for 5 min at room temperature. Reactions were stopped by addition of 0.1 M EDTA and precipitated with 1 ml of ethanol at -80°C. A phenol/chloroform extraction was next performed and the treated RNA samples were ethanol precipitated in the presence of 0.3 M sodium acetate and 2 µg total yeast tRNAs. After centrifugation, the pellet was washed with 80% ethanol, and the dried RNA pellets were dissolved in the RNA loading dye [95% (v/v) formamide, 20 mM EDTA, 0.05% (w/v) each bromophenol blue and xylene cyanol]. RNA alkaline hydrolysis ladders (cleavage after each nucleotide) were generated by incubating the labeled RNA in 1 M sodium carbonate at pH 9 for 3 min at 96°C. In order to generate RNase T1 ladders (cleavage after each guanosine), the labeled RNA was incubated in 1 M sodium hydroxyde citrate in presence of 2 µg of tRNAs for 5 min at 65°C and then treated by 1 U of RNase T1 for 10 min at 65°C. The cleavage products were separated on 10 % polyacrylamide (acrylamide:bis ratio 38:1) 8 M urea-containing gels and visualized by phosphorimaging.

Circular dichroism spectral analysis

The CD spectra were recorded using a Jobin-Yvon Mark VI circular dichrograph at a scan speed of 0.2 nm/s. Quartz spare split-compartment cuvettes with 0.435 cm path length per compartment were used and spectra were scanned in the range of 200-320 nm. The final protein and/or RNA concentration was 2 µM at 25 °C in 10 mM Tris-HCl (pH 7.5) with 150 mM NaCl. The protein and RNA spectra alone were recorded by placing the relevant solution in one compartment of the cuvette and the buffer solution in the other one. For studies of the protein-RNA interaction and Sub-sRNPs formation, a sequential analysis was realized. Firstly, the protein solution (aCBF5) was placed in one compartment and the sRNA in the other one. The CD spectra were recorded before and after mixing the cuvette contents. Secondly, other proteins of the sRNP (aNOP10 and L7Ae) were added in a sequential way and the spectrum of each sub-sRNP was measured after five minutes of mixing. Blanks were run for each spectrum or sequential interaction and subtracted from the raw data. Two spectra were averaged to increase the signal-to-noise ratio. Results are presented as normalized $\Delta\varepsilon$ values on the basis of the mean residue mass of 330 Da. Tacking into account a sensitivity of $\delta(\Delta A) = 10^{-6}$ for the apparatus, the nucleotide concentration and the optical path-length of the cuvette, measurements were at a precision of $\delta(\Delta\varepsilon) = +/- 0.03\ M^{-1}.cm^{-1}$ per nt.

ACKNOWLEDGMENTS

J.-B. F was a doctoral fellow supported by the French Ministère de la Recherche et des Nouvelles Technologies (MRNT). M. B. was a post-doctoral fellow supported by the Centre National de la Recherche Scientifique (CNRS) and by the Agence National de la Recherche (ANR). C. Loegler is acknowledged for her technical assistance. The work was supported by the Agence Nationale de la Recherche (ANR-08-BLAN-0036-01), the CNRS, the MRNT and the Pôle de Recherche Scientifique et Technologique (PRST) “Bioingénierie” of Région lorraine. The authors are grateful to the Service Commun de Biophysicochimie des Interactions Moléculaires (SCBIM) of Nancy University for access for the CD analyses.

REFERENCES

1. Decatur, W.A. & Fournier, M.J. rRNA modifications and ribosome function. *Trends Biochem Sci* **27**, 344-51. (2002).
2. Ofengand, J. Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett* **514**, 17-25 (2002).
3. Badis, G., Fromont-Racine, M. & Jacquier, A. A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *Rna* **9**, 771-9 (2003).
4. King, T.H., Liu, B., McCully, R.R. & Fournier, M.J. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell* **11**, 425-35 (2003).
5. Ejby, M., Sorensen, M.A. & Pedersen, S. Pseudouridylation of helix 69 of 23S rRNA is necessary for an effective translation termination. *Proc Natl Acad Sci U S A* **104**, 19410-5 (2007).
6. Yu, Y.T., Shu, M.D. & Steitz, J.A. Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *Embo J* **17**, 5783-95 (1998).
7. Newby, M.I. & Greenbaum, N.L. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat. Struct. Biol.* **9**, 958-65 (2002).
8. Yang, C., McPheeters, D.S. & Yu, Y.T. Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280**, 6655-62 (2005).
9. Koonin, E.V., Tatusov, R.L. & Rudd, K.E. Protein sequence comparison at genome scale. *Methods in Enzymology* **266**, 295-322 (1996).
10. Kaya, Y. & Ofengand, J. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *Rna* **9**, 711-21 (2003).
11. Watanabe, Y. & Gray, M.W. Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in Euglena gracilis, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic Acids Res* **28**, 2342-52 (2000).
12. Roovers, M. et al. Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic Acids Res* **34**, 4293-301 (2006).
13. Ferre-D'Amare, A.R. RNA-modifying enzymes. *Curr Opin Struct Biol* **13**, 49-55 (2003).
14. Decatur, W.A. & Fournier, M.J. RNA-guided nucleotide modification of ribosomal and other RNAs. *J Biol Chem* **278**, 695-8 (2003).
15. Omer, A.D., Ziesche, S., Decatur, W.A., Fournier, M.J. & Dennis, P.P. RNA-modifying machines in archaea. *Mol. Microbiol.* **48**, 617-629. (2003).
16. Bachellerie, J.P., Cavaille, J. & Huttenhofer, A. The expanding snoRNA world. *Biochimie* **84**, 775-90 (2002).
17. Kiss, T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109**, 145-8. (2002).
18. Baker, D.L. et al. RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev* **19**, 1238-48 (2005).
19. Charpentier, B., Muller, S. & Branlant, C. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res* **33**, 3133-44 (2005).
20. Muller, S. et al. Combined in silico and experimental identification of the Pyrococcus abyssi H/ACA sRNAs and their target sites in ribosomal RNAs. *Nucleic Acids Res* **36**, 2459-75 (2008).
21. Rozhdestvensky, T.S. et al. Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res* **31**, 869-77 (2003).
22. Tang, T.H. et al. Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proc Natl Acad Sci U S A* **99**, 7536-41. (2002).
23. Ganot, P., Caizergues-Ferrer, M. & Kiss, T. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev* **11**, 941-56 (1997).
24. Balakin, A.G., Smith, L. & Fournier, M.J. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86**, 823-34 (1996).
25. Ni, J., Tien, A.L. & Fournier, M.J. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**, 565-573 (1997).
26. Ganot, P., Bortolin, M.L. & Kiss, T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**, 799-809 (1997).
27. Bortolin, M.L., Ganot, P. & Kiss, T. Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *Embo J* **18**, 457-69 (1999).
28. Jin, H., Loria, J.P. & Moore, P.B. Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Mol Cell* **26**, 205-15 (2007).
29. Wu, H. & Feigon, J. H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *Proc Natl Acad Sci U S A* **104**, 6655-60 (2007).
30. Charpentier, B., Fourmann, J.B. & Branlant, C. Reconstitution of archaeal H/ACA sRNPs and test of their activity. *Methods Enzymol* **425**, 389-405 (2007).
31. Manival, X. et al. Crystal structure determination and site-directed mutagenesis of the Pyrococcus abyssi aCBF5-aNOP10 complex reveal crucial

roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res* **34**, 826-39 (2006).
32. Youssef, O.A., Terns, R.M. & Terns, M.P. Dynamic interactions within sub-complexes of the H/ACA pseudouridylation guide RNP. *Nucleic Acids Res* **35**, 6196-206 (2007).
33. Rashid, R. et al. Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol Cell* **21**, 249-60 (2006).
34. Muller, S., Fourmann, J.B., Loegler, C., Charpentier, B. & Branlant, C. Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Psi55-synthase and RNA-guided RNA:Psi-synthase activities. *Nucleic Acids Res* **35**, 5610-24 (2007).
35. Hamma, T. & Ferre-D'Amare, A.R. Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution. *Structure (Camb)* **12**, 893-903 (2004).
36. Charron, C. et al. The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J Mol Biol* **342**, 757-73 (2004).
37. Suryadi, J., Tran, E.J., Maxwell, E.S. & Brown, B.A., 2nd. The Crystal Structure of the Methanocaldococcus jannaschii Multifunctional L7Ae RNA-Binding Protein Reveals an Induced-Fit Interaction with the Box C/D RNAs(,). *Biochemistry* **44**, 9657-9672 (2005).
38. Reichow, S.L. & Varani, G. Nop10 is a conserved H/ACA snoRNP molecular adaptor. *Biochemistry* **47**, 6148-56 (2008).
39. Hamma, T., Reichow, S.L., Varani, G. & Ferre-D'Amare, A.R. The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat Struct Mol Biol* **12**, 1101-7 (2005).
40. Li, L. & Ye, K. Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443**, 302-7 (2006).
41. Duan, J., Li, L., Lu, J., Wang, W. & Ye, K. Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Mol Cell* **34**, 427-39 (2009).
42. Liang, B. et al. Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. *Nat Struct Mol Biol* **16**, 740-6 (2009).
43. Liang, B., Xue, S., Terns, R.M., Terns, M.P. & Li, H. Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. *Nat Struct Mol Biol* (2007).
44. Liang, B. et al. Long-distance placement of substrate RNA by H/ACA proteins. *RNA* **14**, 2086-94 (2008).
45. Lindell, M., Romby, P. & Wagner, E.G. Lead(II) as a probe for investigating RNA structure in vivo. *Rna* **8**, 534-41 (2002).
46. Baker, D.L. et al. Determination of protein-RNA interaction sites in the Cbf5-H/ACA guide RNA complex by mass spectrometric protein footprinting. *Biochemistry* **47**, 1500-10 (2008).
47. Sosnick, T.R., Fang, X. & Shelton, V.M. Application of circular dichroism to study RNA folding transitions. *Methods Enzymol* **317**, 393-409 (2000).
48. Marmier-Gourrier, N. et al. A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *Rna* **9**, 821-38 (2003).
49. Walne, A.J. et al. Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. *Hum Mol Genet* **16**, 1619-29 (2007).
50. Darzacq, X. et al. Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J Cell Biol* **173**, 207-18 (2006).

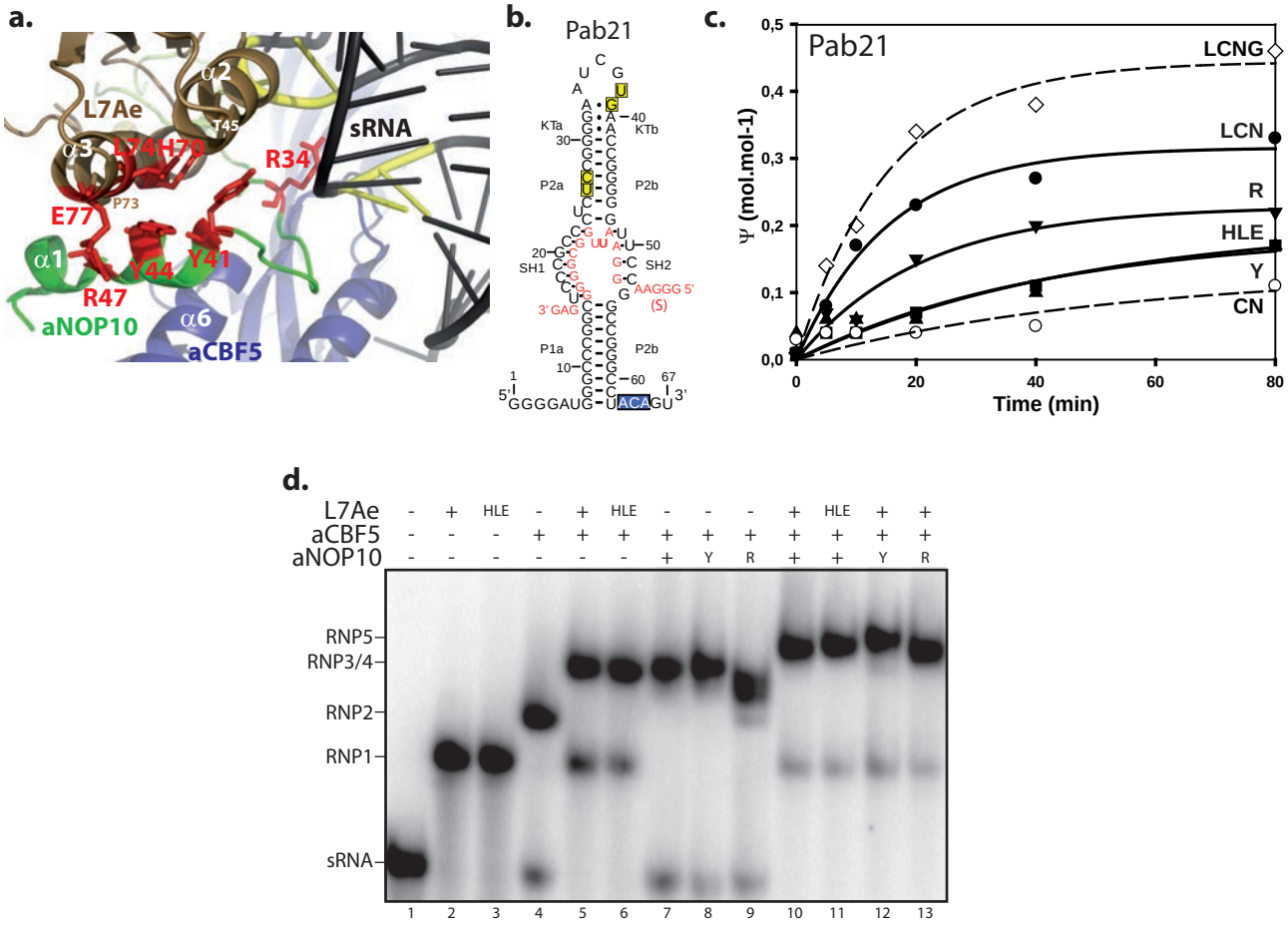


Figure 1 : Effect of mutations in residues implicated in the aNOP10-L7Ae interface on H/ACA sRNP assembly and activity. (**a**) Close-up view of the contact region between protein L7Ae (displayed in gold) and protein aNOP10 (in green). Residues from proteins of the *P. abyssi* archaeon and that are conserved in *P. furiosus* are indicated on the *P. furiosus* sRNP structure (PDB 2HVY). Residues in red were substituted by alanines. Regions of the sRNA contacted by residue R34 within the sRNP are in yellow. (**b**) Secondary structure model of *P. abyssi* Pab21 sRNA and of its interaction with the target RNA. Pab21 possess an apical K-loop and a pseudouridine pocket that targets rRNA modification at position 891 in 16S rRNA. The ACA sequence at the 3' end is boxed. The strands of the various motifs are indicated: strands KTa and KTb of the apical K loop motif, strands P2a and P2b of helix P2, strands P1a and P1b of helix P1. The RNA substrate (S) used *in vitro* is in red with the U residue converted into a Ψ residue in bold character. Helices SH1 and SH2 are formed upon interaction of the RNA substrate with the two strands s1 and s2 of the pseudouridylation pocket. In the sRNA sequence, residues contacted by the lateral chain of aNOP10's R34 aminoacid are boxed in yellow. (**c**) Time course analysis of the RNA-guided Ψ formation in RNA substrate. The RNA substrate was radio-labeled during *in vitro* transcription by incorporation of [α-^{32}P]UTP. It was incubated at 65°C with the unlabeled guide sRNA, and the protein set aCBF5-aNOP10 (CN), or L7Ae-aCBF5-aNOP10 (LCN), or for Pab21 L7Ae-aCBF5-aNOP10-aGAR1 (LCNG). Variants of proteins L7Ae (triple variant H70AL74AE77A, HLE), or variants of aNOP10 protein R47A (R), or Y41AY44A (Y) were used for the reaction. After T2 RNase digestion, the amount of Ψ formation was estimated by 1D-TLC analysis. (**d**) Electrophoresis mobility shift analysis (EMSA) of the various sub-complexes formed upon incubation of the radiolabeled sRNA Pab21 (50 fmol) with various set of the proteins (200 nM each). Variants of the proteins (HLE, Y or R) were used as indicated on top of each lane. The amounts of radioactivity in the bands were estimated wit the ImageQuant software. The percentage of RNA in each RNP was calculated from the radioactivity in each band relative to the total radioactivity in the lane.

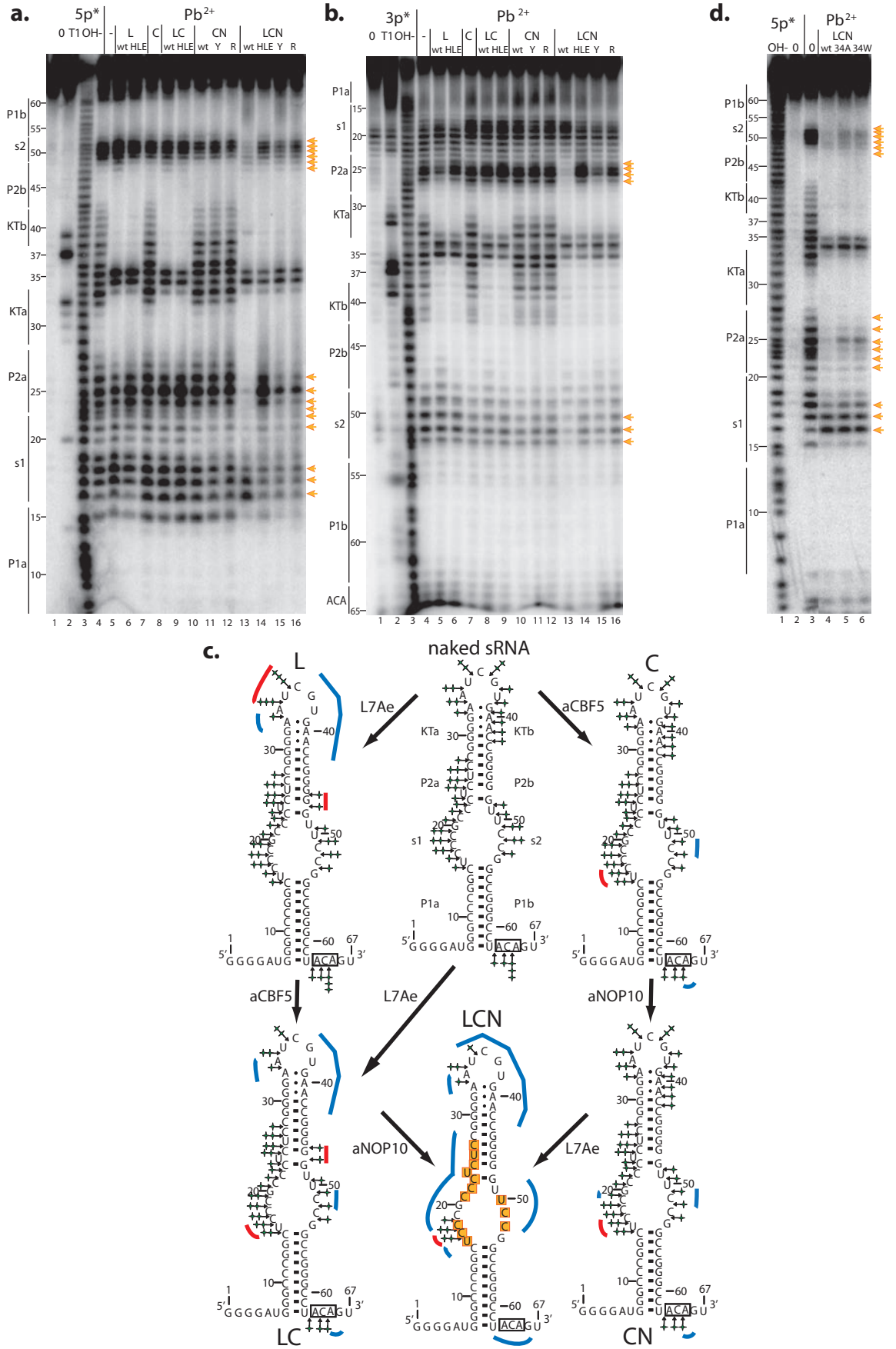


Figure 2 : Chemical probing of sub-complexes of the Pab21 sRNP. (**a**) (**b**) Footprinting of proteins L7Ae, aCBF5 and aNOP10 on sRNA Pab21. Reactions with lead were carried out on 5'- (panel a) or 3'- (panel b) 32P end-labeled sRNA. Samples were fractionated on 10 % polyacrylamide 8 M denaturing gels. The protein set use for assembly of the probed protein-RNA complexes is indicated on top of each lane: L7Ae (L), aCBF5 (C), L7Ae-aCBF5 (LC), aCBF5-aNOP10 (CN), and L7Ae-aCBF5-aNOP10 (LCN). Identity of the aNOP10 or L7Ae variant present in the complexes is indicated with the same code as in Fig. 1. Lane OH- and T1 correspond, respectively, to an alkaline hydrolysis ladder, and RNase T1 digestion ladder obtained in denaturing conditions. (**c**) Summary of probing data. Positions of the phosphodiester bounds cleaved upon incubation with lead of sRNA Pab21 (naked sRNA) and of the various sub-complexes L, C, LC, CN, and LCN are indicated by green stars on the secondary structure model of sRNA Pab21. The number of stars is proportional to the efficiency of the cleavages. Regions that display protection are highlighted by blue lines and region that display an increased lead-induced digestion are highlighted by red lines. Residues cleaved by lead in naked RNA and complexes L, C, LC, and CN, but specifically protected from lead-induced cleavages in the LCN complex are labeled in orange. (**d**) Probing of the LCN complex assembled with aNOP10 variants R34A or R34W.

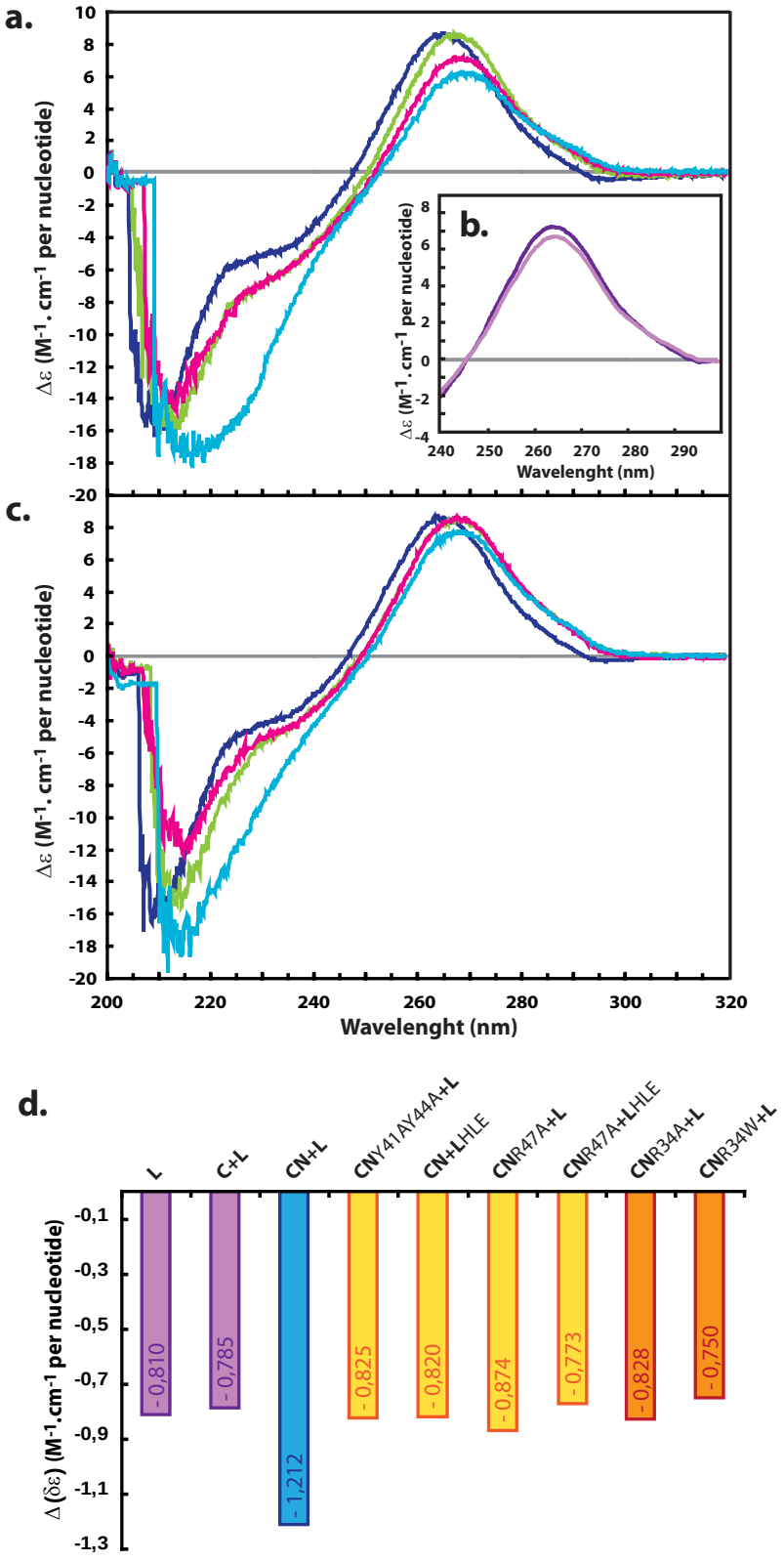


Figure 3 : Different conformational changes of sRNA Pab21 are associated with the sequential binding of various proteins. (**a**) CD spectra of sRNA Pab21 obtained in the presence of protein aCBF5 either in separate cell compartments (dark blue) or after mixing the protein and the sRNA (green). Spectra with sequential addition of proteins to the Pab21-aCBF5 complex: aNOP10 (magenta) added first and then L7Ae (light blue). (**b**) CD spectra of sRNA Pab21 obtained with the sRNA and protein L7Ae in separate compartments (dark purple) and after mixing (purple). (**c**) Same experiment as in panel (a) but with aNOP10 variant R34W. (**d**) Quantification of the decrease in the peak amplitude of the specific RNA positive signal at λmax wavelength, exclusively generated by addition of L7Ae to different sRNA-protein(s) complexes. The resulting $\delta(\Delta\varepsilon)$ values are displayed as a histogram for each complex. sRNP assembly were achieved by incubation of Pab21-aCBF5 with variants of L7Ae (H70AL74AE77A designated as HLE) or aNOP10 (R47A, Y41AY44A, R34A and R34W designated respectively R, Y, A, and W).

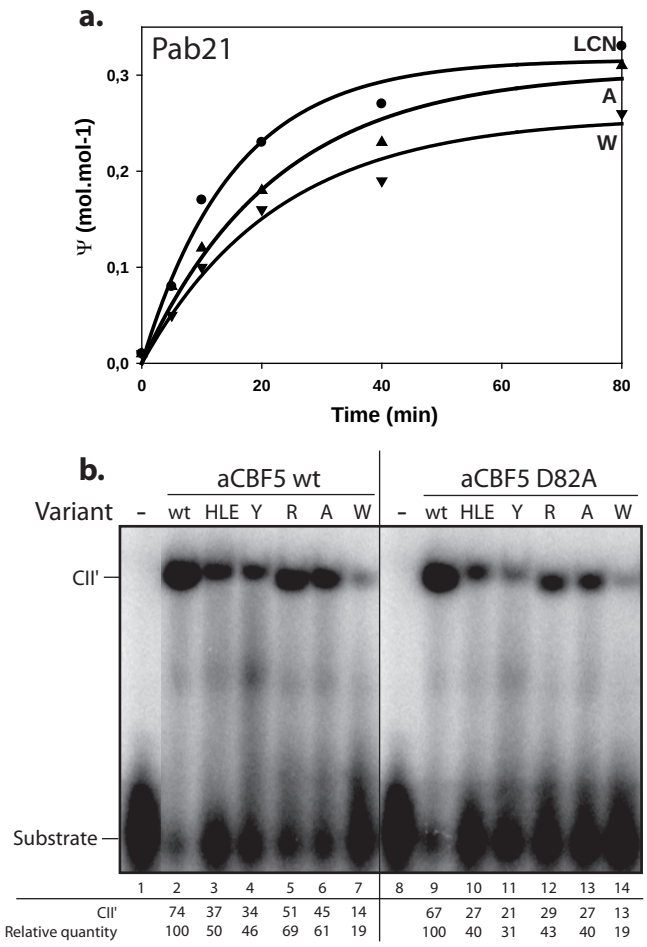


Figure 4 : Effect of mutations in proteins L7Ae and aNOP10 on RNA substrate binding to the sRNP. (**a**) Effect of substitution of aNOP10's R34 residue on sRNP activity. Time course analysis of the Pab21 sRNA-guided Ψ formation in RNA substrate was performed as in Fig. 1c. Pab21 sRNPs were assembled with proteins L7Ae, aCBF5, and either the wild-type aNOP10 protein (LCN) or a aNOP10 variant with a substituted R34 residue (A for R34A and W for R34W). (**b**) Analysis by EMSA of the complexes formed between the Pab21 sRNP and its radio-labeled RNA substrate S. Identity of the aNOP10 or L7Ae variant present in the complexes is indicated with the same code as in panel (a) and in Fig. 2. The yield of complex CII' is indicated below each lane. It represents the percentage of sRNA present in each CII' complex as estimated by radioactivity measurement. Quantities relative to the one obtained with the wild-type sRNP (with a value of 100) were calculated.

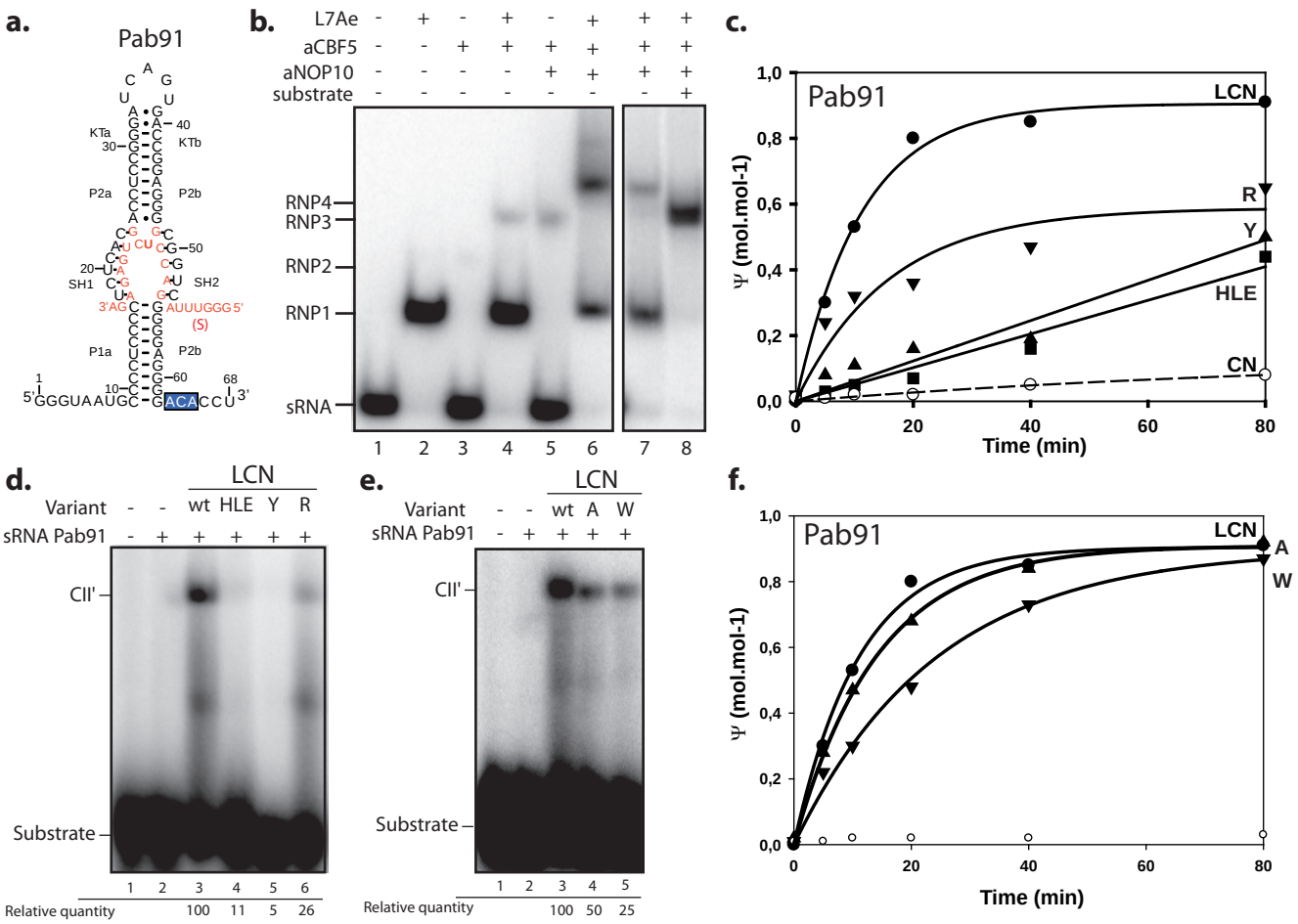


Figure S1 : Effect of mutations in proteins L7Ae and aNOP10 on Pab21 H/ACA sRNP assembly, RNA substrate binding and activity. (**a**) Secondary structure model of *P. abyssi* Pab91 sRNA and of its interaction with the target RNA. Pab91 possess an apical K-loop and a pseudouridine pocket that targets rRNA modification at position 2685 in 23S rRNA. The ACA sequence at the 3' end is boxed. The strands of the various motifs are indicated: strands KTa and KTb of the apical K loop motif, strands P2a and P2b of helix P2, strands P1a and P1b of helix P1. The RNA substrate (S) used *in vitro* is in red with the U residue converted into a Ψ residue in bold character. Helices SH1 and SH2 are formed upon interaction of the RNA substrate with the two strands s1 and s2 of the pseudouridylation pocket. (**b**) Electrophoresis mobility shift analysis (EMSA) of the various sub-complexes formed upon incubation of the radiolabeled sRNA Pab91 (50 fmol) with various set of the proteins (200 nM each). (**c**) (**f**) Time course analysis of the RNA-guided Ψ formation in RNA substrate. The RNA substrate was radio-labeled during *in vitro* transcription by incorporation of [α-^{32}P]CTP. It was incubated at 65°C with the unlabeled guide sRNA, and the protein set aCBF5-aNOP10 (CN), or L7Ae-aCBF5-aNOP10 (LCN). Variants of proteins L7Ae (triple variant H70AL74AE77A, HLE), or variants of aNOP10 protein R47A (R), Y41AY44A (Y), R34A (A), or R34W (W) were used for the reaction. After T2 RNase digestion, the amount of Ψ formation was estimated by 1D-TLC analysis. (**d**) (**e**) Analysis by EMSA of the complexes formed between the Pab91 sRNP and its radio-labeled RNA substrate S. Identity of the aNOP10 or L7Ae variant present in the complexes is indicated with the same code as in panel (c) and (f). The yield of complex CII' is indicated below each lane. It represents the percentage of sRNA present in each CII' complex as estimated by radioactivity measurement. Quantities relative to the one obtained with the wild-type sRNP (with a value of 100) were calculated.

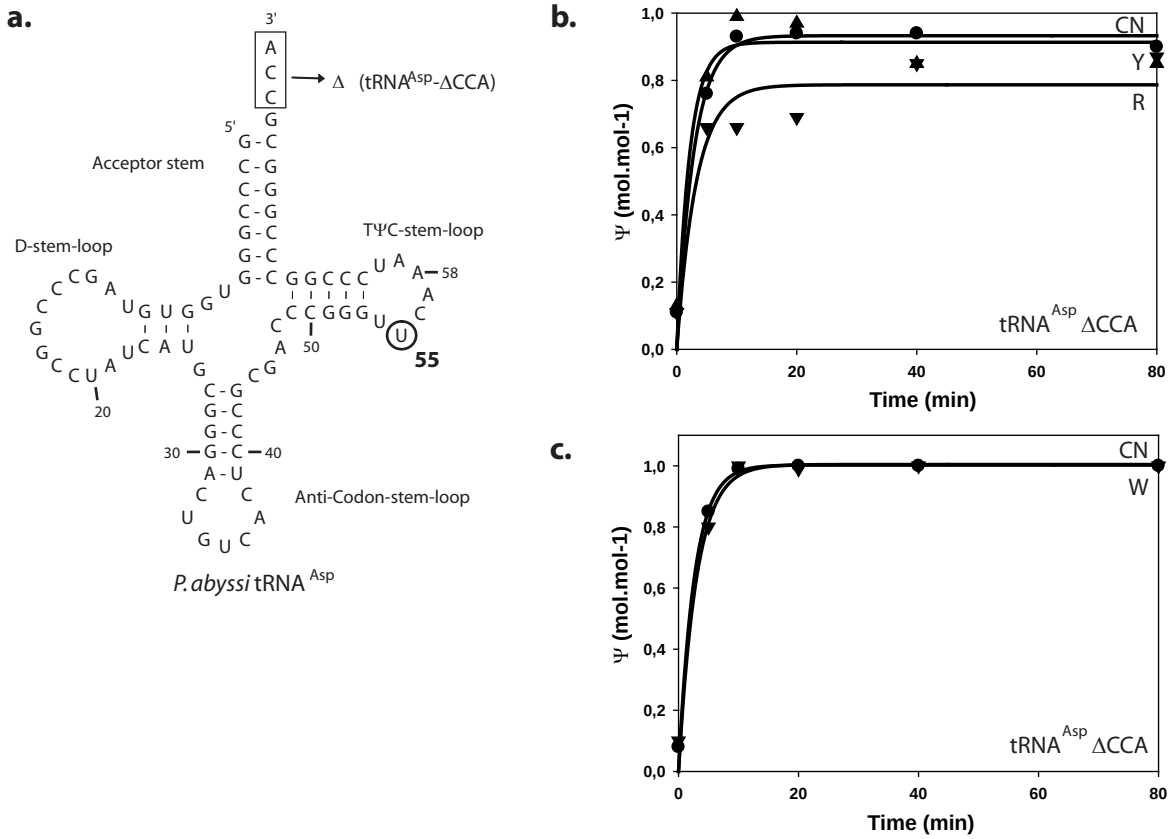


Figure S2 : Effect of mutations in protein aNOP10 on the rate of Ψ55 modification in tRNA by aCBF5. (**a**) Secondary structure models of *P. abyssi* tRNA^{Asp}. Residue U55 is circled. The CCA sequence at the 3' end is boxed. The CCA deletion in variants tRNA^{Asp}-ΔCCA is shown. (**b**) (**c**) Time course analysis of Ψ55 formation in tRNA. The tRNA^{Asp}-ΔCCA substrate was radiolabeled during *in vitro* transcription by incorporation of [α-^{32}P]CTP. It was incubated at 65°C with the protein set aCBF5-aNOP10 (CN). Variants of protein aNOP10 R47A (R), Y41AY44A (Y), R34A (A), or R34W (W) were used for the reaction. After T2 RNase digestion, the amount of Ψ formation was estimated by 1D-TLC analysis.

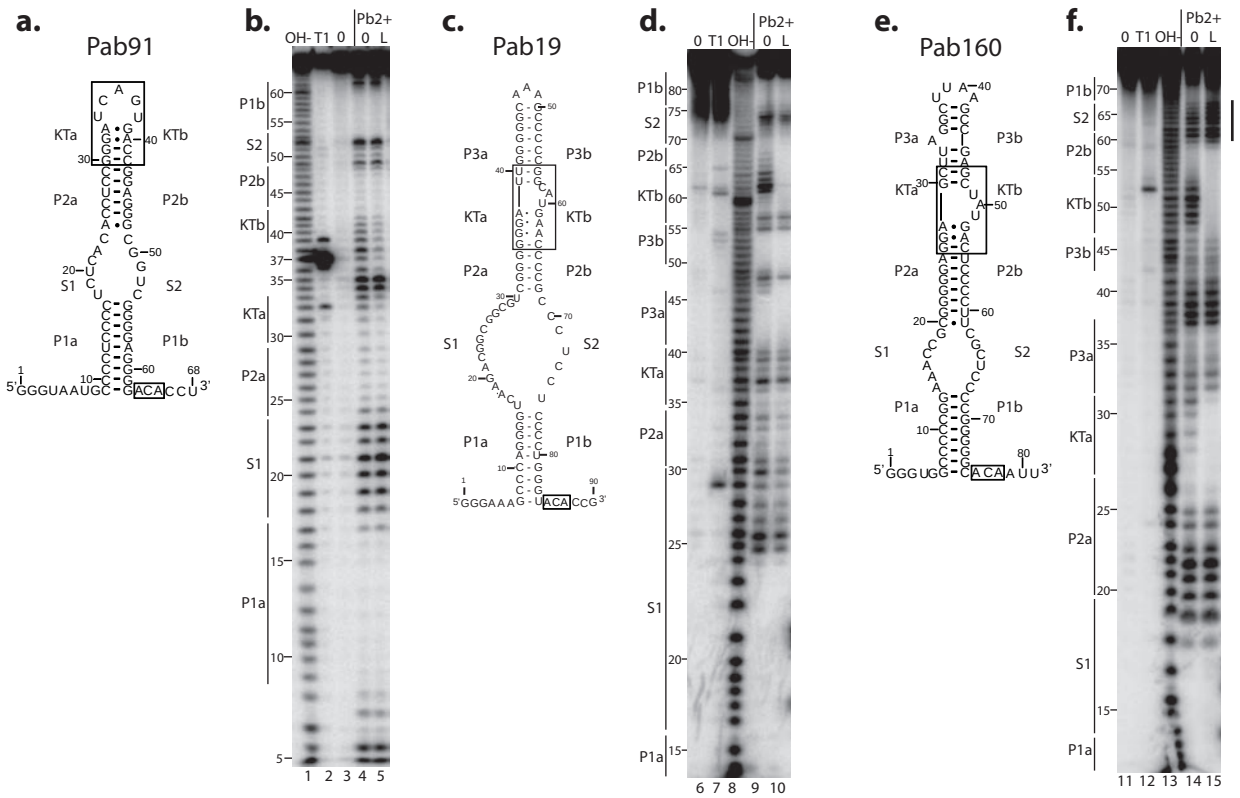


Figure S3 : Chemical probing of sub-complexes formed by the association of L7Ae with various sRNA. (**a**) (**c**) (**e**) Secondary structure models of *P. abyssi* Pab91, Pab19 and Pab160 sRNAs. (**b**) (**d**) (**f**) Footprinting of protein L7Ae (L) on sRNAs Pab91, Pab19, and Pab160. Reactions with lead were carried out on 5'-^{32}P end-labeled sRNA as in Fig. 2. Samples were fractionated on 10 % polyacrylamide 8 M denaturing gels. Lane OH- and T1 correspond, respectively, to an alkaline hydrolysis ladder, and RNase T1 digestion ladder obtained in denaturing conditions.

2. Résultats complémentaires

2.1. Rupture de la liaison entre L7Ae et aNOP10 par l'allongement de l'hélice H2 des sRNA Pab21 et Pab91

Les variants protéiques de L7Ae et de aNOP10 employés dans l'article n°4 n'établissent plus de contact entre eux au sein des particules ce qui affecte fortement l'activité des particules. Les variants Pab21 H2+4 et H2+8, présentés dans la figure r7 du chapitre I.B.2.4 de la partie Résultats, portent une hélice H2 rallongée conduisant à un éloignement du motif en K-turn de la poche de pseudouridylation. Ainsi, l'interaction entre L7Ae et aNOP10 pouvait également être perdue dans ces particules variantes. Les données obtenues avec les variants du sRNA Pab21 et celles obtenues avec les variants des protéines L7Ae et aNOP10 vont dans le même sens : en absence d'association stable aNOP10-L7Ae, l'assemblage des particules et l'incorporation du substrat ne sont pas altérés contrairement à l'activité (Fig. r7). Ces données renforcent notre proposition que l'effet stimulateur de L7Ae sur l'activité des particules ne repose pas exclusivement sur sa fixation au motif en K-turn, ce qui a été proposé par une autre équipe (Liang, B. *et al.*, 2007; Youssef *et al.*, 2007), mais aussi sur son interaction avec aNOP10 (article n°4, Fourmann *et al.,* en préparation).

2.2. Développement d'une technique d'empreintes sur gel pour analyser le complexe sRNP-ARN substrat

Un des objectifs de notre équipe est l'étude de la structure et de la cinétique de formation des complexes RNP durant le déroulement du mécanisme de pseudouridylation, en particulier le complexe sRNP-RNA substrat. La détermination de la structure 3D d'un tel complexe a représenté une avancée importante dans le domaine des recherches sur les RNP à activité catalytique (Liang, B. *et al.*, 2007; Duan *et al.*, 2009; Liang *et al.*, 2009). Avant d'avoir connaissance de ces données, nous avions cherché à caractériser la structure du complexe sRNP-ARN substrat et des essais de cristallisation d'une sRNP Pab91 reconstituée *in vitro* en présence de l'ARN substrat ont été effectués par X. Manival au laboratoire. Ces essais n'ont pas abouti à l'obtention de cristaux. En parallèle, j'ai cherché à développer une approche pour

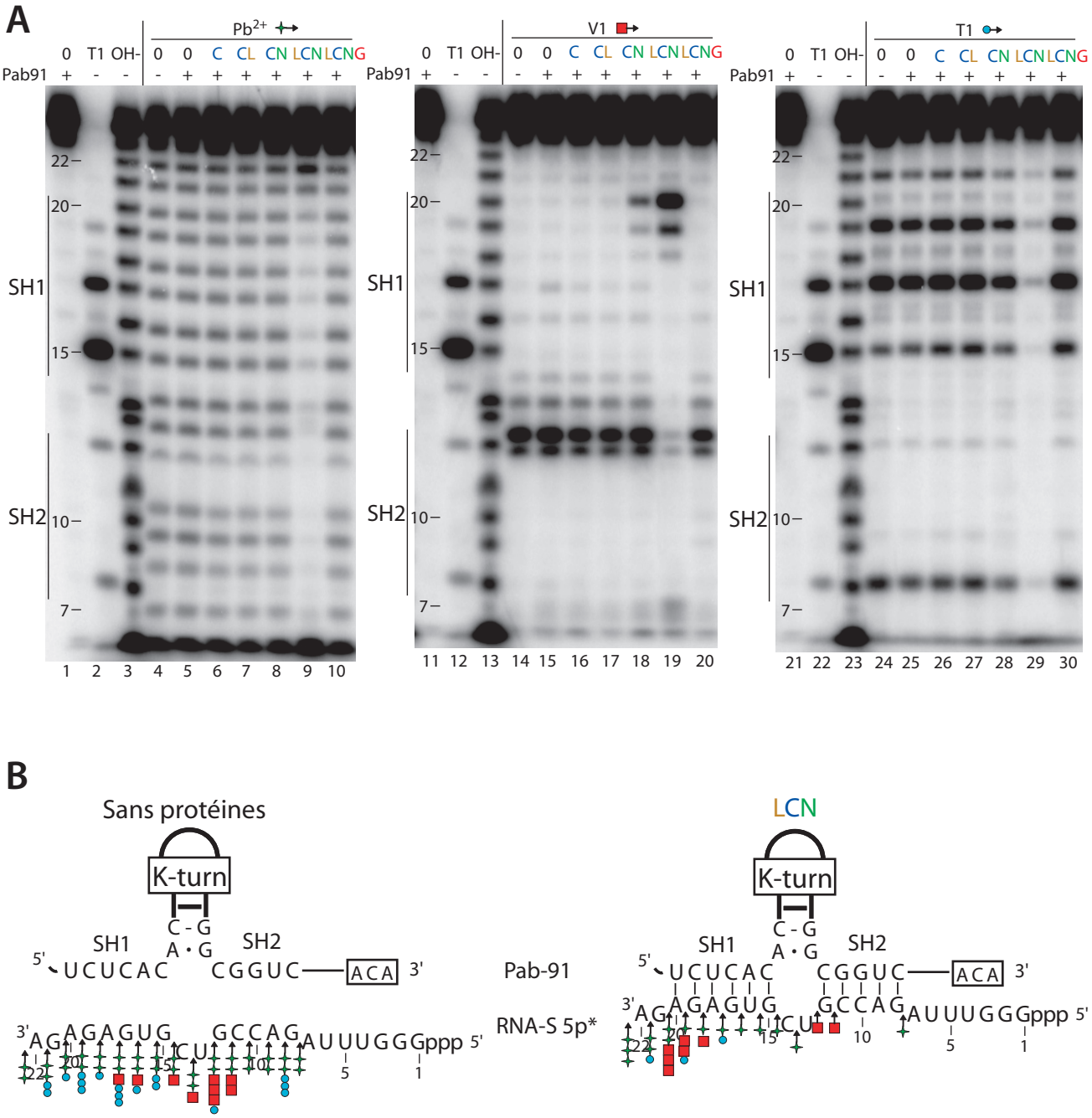


Figure r28 : Etude de la fixation de l'ARN substrat à la sRNP par la technique d'empreinte sur gel réalisée avec des sondes chimiques et enzymatiques.

A. Résultats des expériences d'empreintes sur gel réalisées en présence de l'ARN substrat radiomarqué en 5', du sRNA Pab91 non radiomarqué, de différentes combinaisons des protéines L7Ae (L), aCBF5 (C), aNOP10 (N) et aGAR1 (G). Les combinaisons de protéines et d'ARN sont incubées pendant 10 min à 65°C, puis les complexes formés sont incubés avec le plomb ou l'une des RNases indiquées (Pb^{2+} : plomb, V1 : RNase V1 ou T1 : RNase T1) pendant 2 min à température ambiante. Après extraction au phénol/chlorophorme et précipitation alcoolique, les échantillons sont fractionnés sur un gel de polyacrylamide-urée 12% qui est exposé au phosphoimager. La piste indiquée OH- correspond à une hydrolyse alcaline réalisée de façon à cliver chacun des nucléotides. La piste notée T1 correspond à la digestion en condition dénaturante de l'ARN par la RNase T1 de façon à visualiser tous les résidus G.

B. Schémas représentant la position des différents clivages et leur fréquence sur l'ARN substrat lorsqu'il est incubé avec les différents complexes indiqués formés avec le sRNA Pab91 (1 unité correspond à un clivage de faible intensité, 2 unités correspondent à un clivage de moyenne intensité et 3 unité représentent un clivage de forte intensité).

caractériser en solution les complexes du mécanisme réactionnel et de pouvoir ensuite comparer les données avec celles obtenues avec des complexes assemblés avec les protéines variantes. Dans cette optique, j'ai testé si la technique d'empreinte sur gel de l'ARN substrat lorsqu'il est en interaction avec la sRNP pouvait représenter une approche intéressante pour la caractérisation des étapes finales du mécanisme de pseudouridylation (Fig. i62 étapes 2 à 5).

Recherche des conditions d'obtention du complexe CII' par empreintes sur gel

Dans ces expériences d'empreintes, j'ai employé l'activité nucléase du plomb et des RNases V1 et T1 sur le complexe CII' qui est formé suite à l'interaction de l'ARN substrat RNA-S avec la sRNP Pab91 assemblée avec les protéines L7Ae, aCBF5 et aNOP10 (LCN) (Fig. r28). Le choix de l'ARN guide Pab91 plutôt que Pab21 s'explique par le fait que Pab91 permet de former un taux plus important de complexe CII' que Pab21, et donc les empreintes de la particule sRNP sur l'ARN substrat étaient potentiellement plus faciles à observer. J'ai également testé l'effet de la présence de aGAR1 qui comme nous l'avons vu a un effet délétère sur la quantité de complexes CII' obtenus avec Pab91 (Fig. r11 et partie I.B.3.6 des résultats).

Nous avions déjà utilisé le plomb dans notre étude précédente (article n°4, Fourmann *et al.,* en préparation). Cet agent chimique représente une sonde bien adaptée pour étudier les changements de conformation dans les ARN : sa spécificité d'activité est assez large, avec une préférence pour les régions d'ARN en simple brin ou peu ordonnées. La RNase V1 est une enzyme qui hydrolyse l'ARN dans ses régions en double brin ou présentant des empilements de bases. La RNase T1 digère l'ARN avec une très forte spécificité pour les résidus G en simple brin.

Après formation des complexes par incubation de l'ARN substrat radiomarqué à son extrémité 5', du sRNA Pab91 et des protéines en différentes combinaison pendant 10 min à 65°C, les échantillons ont été incubés 2 min à température ambiantes avec la sonde chimique ou chacune des deux sondes

enzymatiques. Après extraction et précipitation des ARN, ceux-ci ont été fractionnés par électrophorèse sur gel dénaturant (cf Matériel et Méthodes de l'article n°4). L'incubation de l'ARN substrat en absence de protéines et de l'une ou l'autre des sondes chimique ou enzymatique montre que l'ARN est très peu dégradé dans les conditions expérimentales employées (Fig. r28 A pistes 1, 11 et 21). Deux échelles de taille ont été préparées. La digestion par la RNase T1 dans des conditions dénaturantes permet de révéler tous les résidus G de l'ARN (Fig. r28 A pistes 2, 12 et 22). L'hydrolyse alcaline de l'ARN crée une échelle de taille avec une résolution au nucléotide près (Fig. r28 A pistes 3, 13 et 23).

Comme mentionné ci-dessus, notre but était de caractériser le complexe CII' (obtenu avec la combinaison de protéines LCN) par l'emploi des différentes sondes. Nous constatons qu'en absence de sRNA ou de protéines, ainsi que dans les conditions de protéines C, CL, CN et LCNG, le même profil de clivage est obtenu (Fig. r28 A pistes 3-8, 10, 14-18, 20 et 24-28, 30). Cela indique que dans ces conditions expérimentales, l'ARN substrat est largement non structuré, seuls les résidus aux positions 11 et 12 pourraient être engagés dans un appariement ou dans un empilement de bases comme en témoigne la présence de clivages à ces positions par la RNase V1. Ainsi, l'ARN substrat n'est pas apparié au sRNA et les combinaisons de protéines employées ne permettent pas au substrat d'être pris en charge par la sRNP. Les clivages observés sont reportés sur la séquence de l'ARN substrat qui n'est pas appariée à Pab91 (Fig. r28 B, schéma de droite).

L'emploi des 3 protéines LCN conduit à i) une protection presque totale de tous les résidus vis-à-vis des activités nucléases du plomb et de la RNase T1, et ii) une apparition de clivages à la RNase V1 entre les résidus 19 à 21 correspondant à la formation de l'hélice SH1 (Fig. r28 A pistes 9, 19 et 29). Ces deux effets sont cohérents avec la disparition des régions en simple brin de l'ARN substrat au profit de la formation des hélices SH1 et SH2 lorsqu'il est apparié à la poche de pseudouridylation. La protection de la zone autour des résidus 11 à 14 vis-à-vis de l'activité de la RNase V1, peut correspondre au changement de conformation opéré par l'ARN substrat lorsqu'il s'associe à la poche de pseudouridylation mais pourrait

également provenir d'une perte d'accessibilité de l'hélice SH2 pour la RNase V1 dans le complexe sRNP-ARN substrat. Nous concluons que ces observations rendent compte de la prise en charge du substrat par la sRNP assemblée avec les protéines LCN et de la formation des appariements entre le substrat et la poche de pseudouridylation (Fig. r28 B).

En présence du couple de protéines aCBF5-aNOP10, nous observons une apparition de clivages à la RNase V1 entre les nucléotides 19 à 21 et une légère protection des coupures engendrées par la RNase T1 au niveau des G des positions appariées dans l'hélice SH1 17, 19 et 21, démontrant à nouveau que les RNP assemblées avec ce couple de protéines permettent la formation d'un faible taux de complexe CII' (Fig. r28 A pistes 8, 18, 28).

Ces résultats permettent de conclure que l'utilisation du jeu de protéine LCN représente les meilleures conditions expérimentales pour visualiser la présence de l'ARN substrat dans la sRNP et que la sonde la mieux adaptée pour détecter ces complexes est la RNase V1. De plus, la conservation du profil des clivages et protections, observée lors de l'incubation de la RNase V1 avec les complexes CII' à différents temps (10 min et 60 min), indique que la conformation de l'ARN substrat au sein du complexe CII' reste très stable (Fig. r29 A et B, flèches ou encadrée oranges et vertes). Nous pouvions alors examiner l'effet de la présence des protéines variantes sur ce profil électrophorétique caractéristique du complexe CII'.

<u>Schématisation du profil de clivage du substrat dans la structure 3D de la sRNP et emploi des variants de L7Ae et de aNOP10</u>

La figure r30 présente deux vues de la structure 3D du complexe sRNP-ARN substrat (LCNG, sRNA et ARN substrat) sur lesquelles nous avons matérialisé les coupures par la RNase V1 qui sont protégées ou qui apparaissent dans la séquence de l'ARN substrat lorsqu'il est présent au sein du complexe CII'. Le substrat présent dans cette figure correspond à celui employé par Duan et collaborateurs. Nous

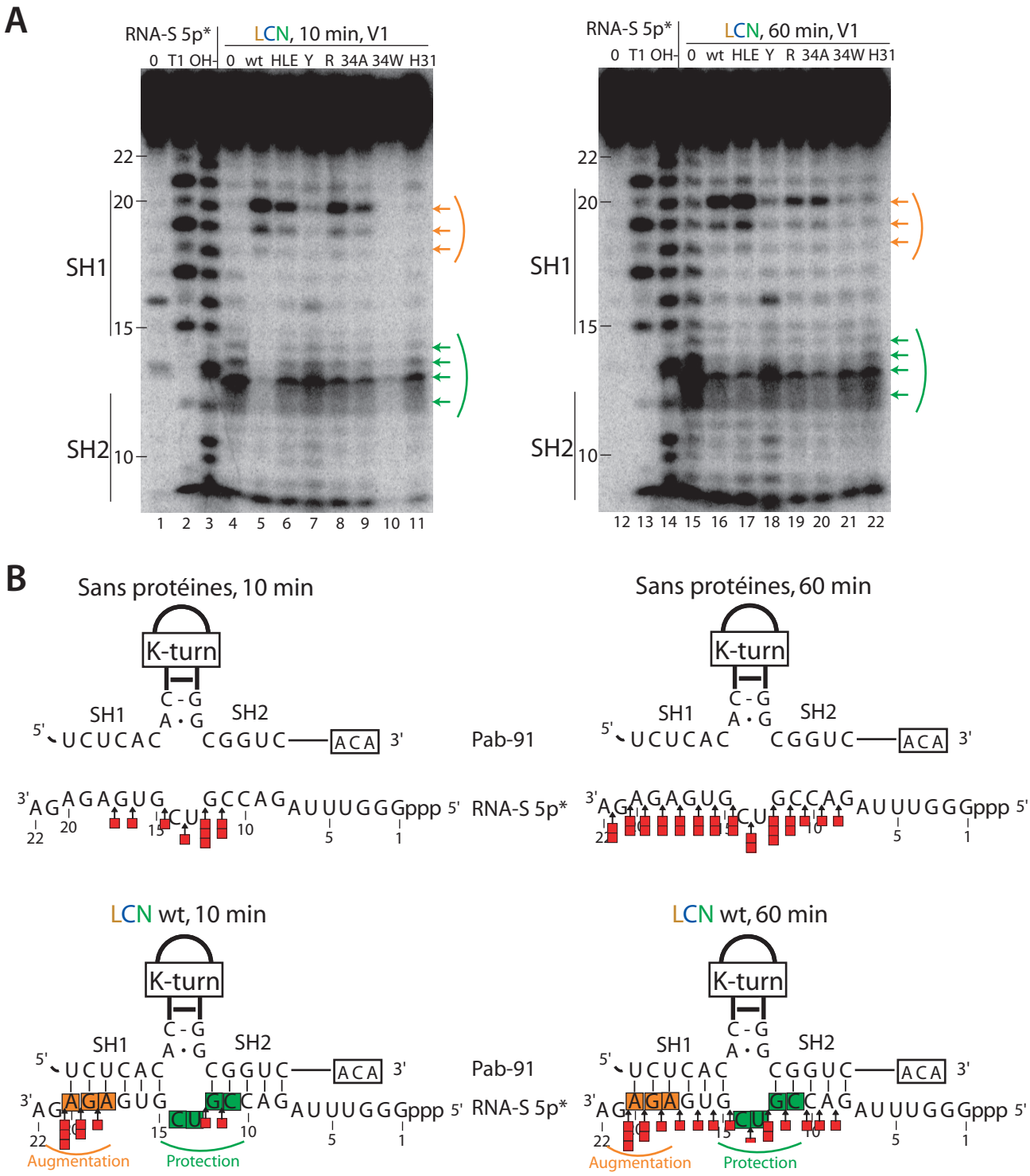


Figure r29 : Etude de l'effet de substitutions dans L7Ae et dans aNOP10 sur la fixation de l'ARN substrat à la sRNP au cours du temps par la technique d'empreinte sur gel réalisée avec des sondes enzymatiques.

A. Résultats des expériences d'empreintes sur gel réalisées en présence de l'ARN substrat radiomarqué en 5', du sRNA Pab91 non radiomarqué, les 3 protéines L7Ae, aCBF5 et aNOP10. Les protéines et les ARN sont incubés pendant 10 min ou 60 min à 65°C, puis les complexes formés et analysé selon la même méthodologie détaillée dans la figure r28.

B. Schémas représentant la position des différents clivages et leur fréquence sur l'ARN substrat lorsqu'il est incubé avec les différents complexes indiqués formés avec le sRNA Pab91. Les nucléotides encadrés sur fond orange correspondent à ceux dont l'intensité de clivage à la RNase V1 est augmentée, alors que les nucléotides encadrés sur fond vert correspondent à ceux dont l'intensité de clivage à la RNase V1 est diminuée (protection).

constatons que, si la sRNP Pab91 adopte une structure similaire à celle observée dans la structure 3D présentée, les protections des nucléotides 11 à 14 se situent dans une zone majoritairement constituée par le site actif de aCBF5 et que les augmentations de la réactivité des nucléotides 19 à 21 sont localisées à l'extrémité 3' de SH1 qui est bien accessible au solvant et donc à la RNase V1. Nous concluons que le profil de clivage du substrat est en accord avec la structure 3D. Ainsi, la méthodologie employée par empreintes sur gel permettrait de caractériser la stabilité des complexes CII' au cours du temps et en fonction des variants protéiques utilisés.

Nous avons donc analysé les profils de digestion par la RNase V1 de l'ARN substrat dans les complexes CII' assemblés après 10 min et 60 min avec les variants des protéines L7Ae (variant HLE) ou de aNOP10 (variants Y41A Y44A, R47A, R34A, R34W et H31A) (Fig. r29 A). Le variant H31A de aNOP10 est utilisé comme témoin négatif car nous avions montré qu'en sa présence, l'activité guidée de aCBF5 était très fortement et spécifiquement affectée (article n°3, (Muller *et al.*, 2007b)). Les variants peuvent être classés en deux groupes selon leur profil de clivage obtenu :

- Le premier groupe comprend les variants HLE, R47A et R34A qui permettent une incorporation modérée de l'ARN substrat à la sRNP ou la formation d'une interaction de l'ARN substrat à l'ARN guide moins stable (article n°4, Fourmann et *al.*, en préparation). Le profil électrophorétique est identique à celui obtenu avec les sRNP sauvages, la fréquence plus faible de coupures aux positions 19 à 21 et la protection moins importante des nucléotides 11 à 14 proviennent du plus faible taux de complexe CII' formé avec ces sRNP variantes par rapport aux sRNP sauvages (Fig. r29 A pistes 6, 8, 9, 17, 19 et 20).

- Le second groupe rassemble les variants Y41A Y44A, R34W et H31A pour lesquels aucun complexe CII' ne peut être détecté par retard de migration sur gel (article n°4, Fourmann *et al.*, en préparation). Ceci est cohérent avec le profil de clivage obtenu avec ces variants qui est proche de celui obtenu en absence de protéines, lorsque l'ARN substrat est libre (Fig. r29 A pistes 7, 10, 11, 18, 21 et 22). Une formation transitoire d'un complexe ARN-protéine pourrait être détectée par la

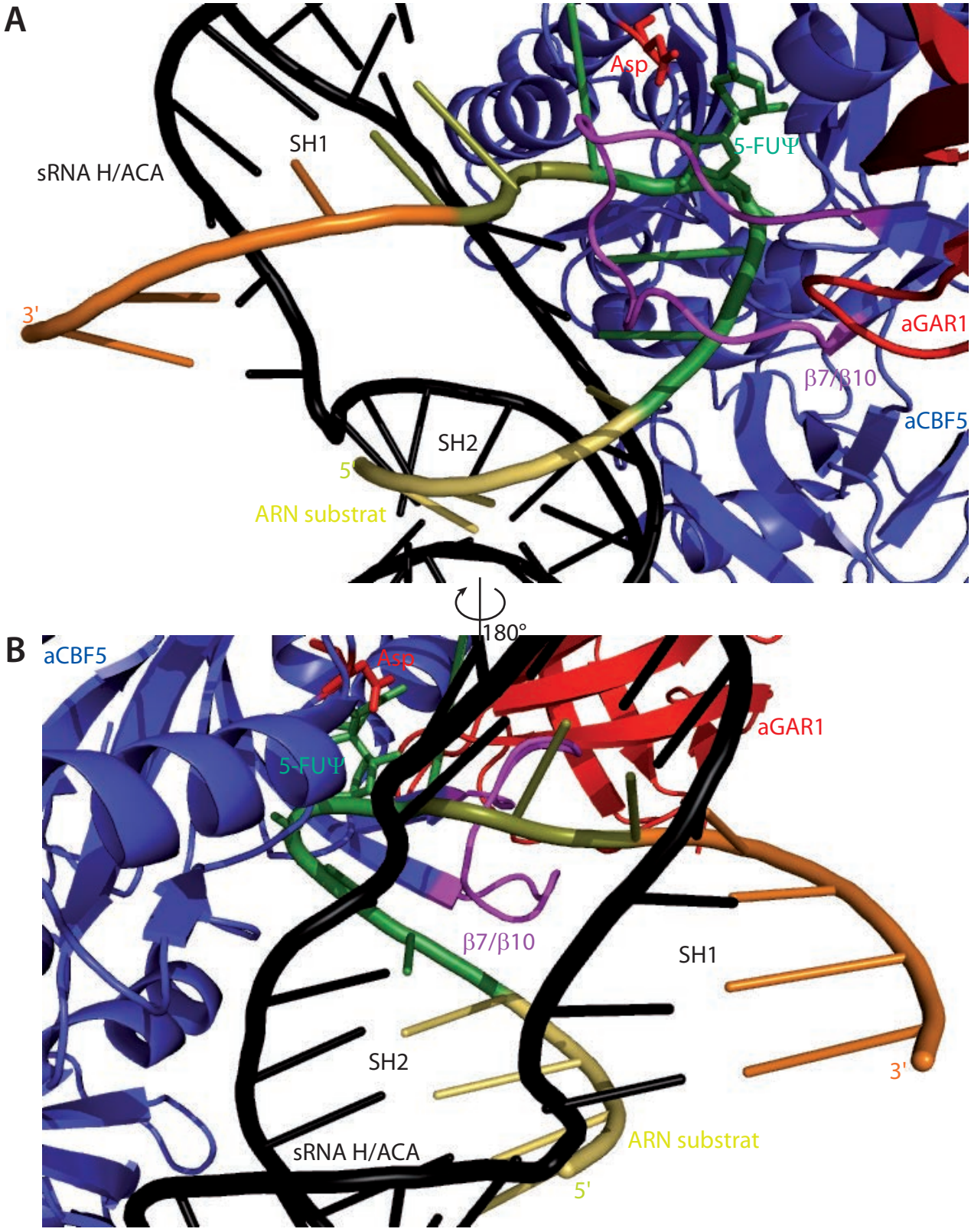


Figure r30 : Stuctures 3D du complexe L7Ae-aCBF5-aNOP10-aGAR1 lié au sRNA et à l'ARN substrat. (D'après Duan et al., 2009).

A. Vue détaillée d'une partie de la structure 3D de la particule comportant les protéines L7Ae, aCBF5 (bleu), aNOP10 et aGAR1 (rouge), le sRNA H/ACA (noir) et l'ARN substrat (jaune). La boucle β7/β10 de aCBF5 est colorée en violet et l'aspartate catalytique en rouge. Les régions de l'ARN substrat colorées en orange correspondent à celles dont l'intensité de clivage à la RNase V1 augmente, tandis que les régions de l'ARN substrat qui sont colorées en vert correspondent à celles dont l'intensité de clivage à la RNase V1 diminue (protection) d'après les observations de la figure r29. Adapté grâce au logiciel Pymol à partir des coordonnées PDB=3HAY

B. Vue détaillée d'une partie de la même particule ayant subit une rotation de 180° selon l'axe verical.

réactivité d'une sonde chimique ou enzymatique. Nos tests avec les variants de ce groupe indiquent que dans le cas de la RNase V1, il n'y a pas de formation de complexe transitoire. Nous concluons que ces mutations pourraient affecter le déroulement de l'étape 2 du mécanisme (Fig. r26) en empêchant l'incorporation du substrat dans la particule, (article n°4, Fourmann *et al.,* en préparation).

Les substitutions, permettant de conserver la formation d'un complexe CII', n'ont pas d'effet apparent le profil d'activité de la RNase V1. Ainsi, par rapport aux protéines sauvages, ces substitutions ne conduisent à aucun changement de conformation de l'ARN substrat lorsqu'il est incorporé dans la particule. Il serait intéressant d'analyser par cette approche les variants du site actif de aCBF5 et ceux de la boucle β7/β10.

E. Utilisation des sRNA H/ACA d'archaea et des snoRNA H/ACA de levure comme modèle d'étude de la Dyskératose Congénitale

1. Intérêt et contexte

Nous avons vu au cours de l'introduction que la Dyskératose congénitale possédait 3 formes :

- des mutations dans le gène *DKC1*, qui code la Dyskérine humaine, sont responsables de la forme liée au chromosome X (Heiss *et al.*, 1998),

- l'une des formes autosomale récessive de la maladie a été identifiée comme étant due à une mutation dans le gène codant la protéine NOP10, conduisant à la production de la forme variante R34W de la protéine (Walne *et al.*, 2007),

- la forme autosomale dominante de la maladie est causée par des mutations dans des gènes de composants de la télomérase humaine ou du complexe de protection des télomères (Vulliamy *et al.*, 2001; Armanios *et al.*, 2005; Savage *et al.*, 2008).

Comme nous l'avons vu dans la partie introduction, chez l'homme, l'ARN de la télomérase possède un domaine H/ACA qui recrute les protéines associées aux snoRNP à boîtes H/ACA. Les mutations dans la Dyskérine ou dans NOP10 aboutissent à une perte de la stabilité de l'ARN de la télomérase et à un raccourcissement de la taille des télomères. Néanmoins, l'activité des autres snoRNP H/ACA pourrait aussi être affectée par ces mutations et, par voie de conséquence, la biogenèse des ribosomes pourrait être perturbée, ce qui contribuerait à augmenter la sévérité de la maladie.

Nous avons voulu exploiter le fait que contrairement à l'ARN de la télomérase humaine, celui de levure ne comporte pas de domaine RNP H/ACA (cf partie IV.C.1.2 de l'introduction ; Fig. i37). Ainsi, les effets *in vivo* des mutations dans les composants de RNP H/ACA qui conduisent à la DC pouvaient être analysés exclusivement sur l'activité des snoRNP H/ACA. En particulier chez la levure, la

perte d'activité de la snoRNP snR30, qui est essentielle pour la biogenèse des ribosomes, pouvait conduire à un fort effet phénotypique. Le résidu R34 de NOP10 est conservé dans les protéines aNOP10 d'archaea et Nop10p de levure. Nos données présentées dans le chapitre II.D (article n°4, Fourmann *et al.*, en préparation) indiquent que la substitution R34W conduisant à la forme AR de la DC affecte peu l'activité d'une sRNP à boîtes H/ACA d'archaea bien qu'elle induise des changements de conformation du sRNA au sein des RNP. Nous avons donc voulu tester si l'expression du variant R34W de la protéine Nop10p affectait ou non l'activité des snoRNP H/ACA chez *S. cerevisiae*. Nous avons donc analysé l'effet de l'expression des variants R34W et R34A sur la croissance des cellules et sur le taux de snoRNA.

2. Effet des substitutions R34A et R34W de Nop10p sur la croissance des levures et sur la stabilité des snoRNA à boîtes H/ACA

La protéine Nop10 est essentielle à la croissance chez la levure. Nous avons utilisé une souche exprimant de façon conditionnelle le gène *NOP10* afin de tester la complémentation d'un arrêt de la production de la protéine endogène par une production ectopique soit de la protéine Nop10p de levure ou NOP10 humaine soit de leurs variants R34W et R34A. Nous avons employé la souche de *S. cerevisiae* GAL::Nop10p, exprimant le gène *NOP10* à partir du promoteur GAL. En présence de galactose dans le milieu de culture, l'activateur transcriptionnel GAL10 permet la production de la protéine endogène. Lors de cultures en présence de glucose, cet activateur transcriptionnel est absent et la production de la protéine endogène n'est pas effective. Les cellules de la souche GAL::Nop10p ont été transformées par des vecteurs d'expression recombinants, construits à partir du vecteur pG1 (TRP1/CEN) dans lequel les séquences codant les protéines Nop10 de levure et NOP10 humaine, ainsi que leurs variants R34A et R34W, ont été clonées. Les clones obtenus sur milieu Trp-/Galactose ont été dilués suivant une gamme de concentration et une fraction a été déposée sur milieu solide Trp-/Glucose permettant de réprimer la production de Nop10p endogène. Les résultats de croissance après 3 jours d'incubation à 3

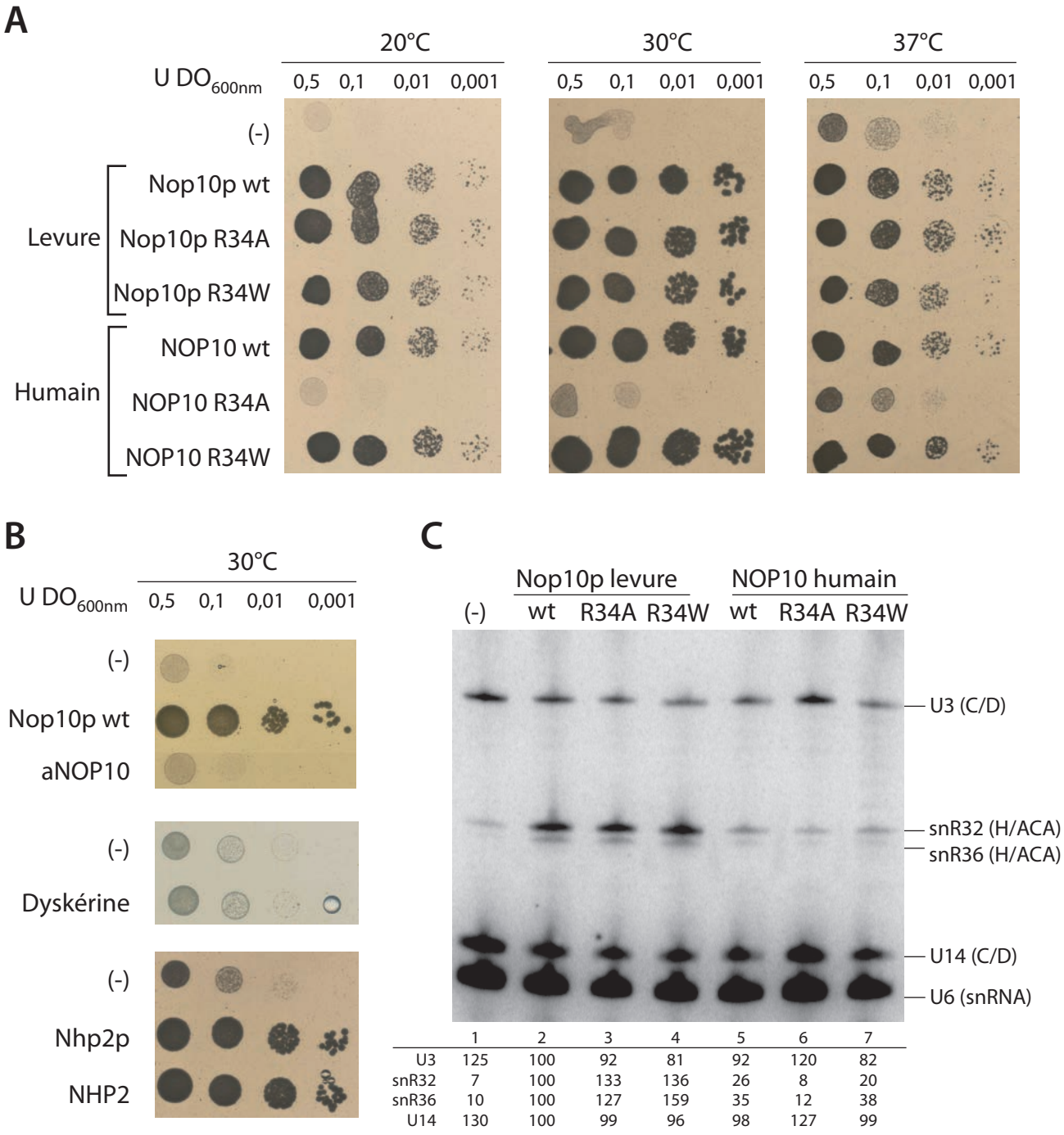


Figure r31 : Effet des mutations R34A et R34W introduites dans la protéine Nop10p de *S. cerevisiae* sur la croissance des cellules et sur la stabilité des snoRNA à boîtes H/ACA.

A. et **B.** Etude de la croissance des souches GAL::Nop10p, GAL::Cbf5p et GAL::Nhp2p transformées par le plasmide pG1 dans lequel chacun des gènes wt ou mutés codant l'une des protéines de levure, humaine ou d'archae ont été clonés. Les cellules ont été étalées sur boîtes en présence de glucose pour réprimer la production de Nop10p, Cbf5p ou Nhp2p à partir des gènes endogènes pendant 3 à 5 jours. La croissance des souches a été testée à 20, 30 et 37°C et selon une gamme de dilution (0,5 à 0,001 unité de DO_{600nm}). Un contrôle négatif a été réalisé en utilisant la même souche transformée par le plasmide pG1 dans lequel aucun gène n'a été cloné (-).

C. Etude de la stabilité de snoRNA à boîtes C/D ou H/ACA et du snRNA U6 par Northern Blot, en employant les variants R34A et R34W de Nop10p de levure ou de NOP10 humaine. Un contrôle négatif a été réalisé en utilisant la même souche transformée par le plasmide pG1 dans lequel aucun gène n'a été cloné (-). Chaque piste correspond à un dépôt de 5 µg d'ARN totaux. Le snRNA U6 est utilisé comme contrôle par rapport auquel le pourcentage de stabilité des snoRNA sont rapportés sous chaque piste.

températures différentes sont présentés (Fig. r31 A). Nous constatons que la production ectopique des protéines de levures Nop10 wt, Nop10 R34A et R34W, et des protéines humaines NOP10 wt et R34W permet la complémentation de l'absence de production de la protéine Nop10 endogène (Fig. r31 A). La densité cellulaire est similaire dans tous ces essais, indiquant que la croissance des cellules n'est pas affectée par les substitutions d'acides aminés. Seul le variant de la protéine humaine NOP10 R34A ne conduit pas à la viabilité cellulaire (Fig. r31 A). Nous concluons que la mutation R34W introduite dans les gènes codant Nop10p de levure ou NOP10 humaine ne conduit à aucun phénotype marquant de croissance des levures. Par ailleurs, nous avons testé la complémentation par la protéine aNOP10 d'archae (Fig. r31 B). Nous constatons que la protéine aNOP10 d'archae ne peut se substituer à la protéine Nop10p de levure.

Des données obtenues au laboratoire montrent que la croissance des cellules de levure est possible même lorsque le snoRNA U3, pourtant essentiel à la survie cellulaire, est produit à un très faible taux (N. Rolland, communication personnelle). De ce fait, nous avons souhaité vérifier si les substitutions bien que permettant la complémentation, n'affectaient pas tout de même la stabilité des snoRNA H/ACA. Nous avons analysé le taux des snoRNA H/ACA par Northern Blot. Les cellules transformées par les différents plasmides ont été cultivées en milieu liquide Trp-/Glucose pendant 24h, puis les ARN totaux ont été extraits et 5 µg ont été fractionnés par électrophorèse en gel d'acrylamide dénaturant. Les hybridations ont été effectuées avec des oligonucléotides radiomarqués et complémentaires des snoRNA à boîtes H/ACA snR32 et snR36, ainsi que les snoRNA à boîtes C/D U3 et U14, et le snRNA U6 qui ont été utilisés comme témoins. Comme attendu, nous constatons qu'en présence du plasmide dans lequel aucun gène n'a été cloné, le taux des snoRNA à boîtes H/ACA chute très fortement (Fig. r31 C piste 1). La complémentation par la protéine de levure Nop10 wt ou l'un de ses variant R34A et R34W, permet de rétablir le taux des snoRNA à boîtes H/ACA (Fig. r31 C pistes 2 à 4). A l'inverse, l'emploi des protéines humaines NOP10 wt, R34A et R34W, conduit à la chute spécifique du taux de snoRNA à boîtes H/ACA (Fig. r31 C pistes 5 à 7). Ces

résultats indiquent que la protéine NOP10 humaine, malgré qu'elle permette la viabilité cellulaire, a un effet négatif sur la stabilité des snoRNA à boîtes H/ACA, qui est indépendant d'un l'effet des substitutions R34A et R34W.

Nous concluons que la substitution R34W dans la protéine Nop10 de levure ne conduit à aucun effet sur la croissance des cellules et que la stabilité des snoRNP à boîtes H/ACA n'est pas modifiée. Ces données corroborent les résultats obtenu *in vitro* par reconstitution des sRNP H/ACA, suggérant que la mutation R34W n'affecte ni l'assemblage ni l'activité des snoRNP H/ACA et des sRNP H/ACA. Il reste à vérifier si la substitution R34W conduit aussi à des effets structuraux sur les snoRNP H/ACA. Nous proposons donc que chez l'homme, la substitution R34W aurait essentiellement un impact sur l'assemblage et/ou l'activité de la télomérase et pas sur l'activité des snoRNP H/ACA. Des tests d'assemblages de la télomérase à partir des protéines produites par traduction *in vitro* dans des systèmes acellulaires permettraient vérifier cette proposition.

Nous avons réalisé des tests sur d'autres souches telles que les souches GAL::Cbf5p et GAL::Nhp2p dans le but de tester la complémentation du défaut de production des protéines Cbf5 et Nhp2 par les protéines humaines Dyskérine et NHP2, respectivement. Nous constatons que la protéine Dyskérine ne complémente pas la souche GAL::Cbf5p alors que NHP2 complémente la souche GAL::Nhp2p (Fig. r31 B). Il est donc envisageable de tester dans la levure l'effet des mutations dans NHP2 conduisant à la Dyskératose Congénitale afin d'analyser si l'activité des snoRNP H/ACA est modifiée ou non.

CONCLUSION & PERSPECTIVES

Conclusion & Perspectives

Nos travaux ont permis d'avancer dans la compréhension globale de l'organisation des particules sRNP H/ACA, de définir le rôle de chacune des protéines les constituant et d'apporter des éléments de réponse à certaines grandes questions liées au déroulement de la réaction de pseudouridylation. De plus, nous avons démontré la possibilité pour les sRNP H/ACA d'archaea de modifier un résidu lorsque celui-ci est présent dans un trinucléotide en simple brin (5'-UNN-3') formé suite à l'appariement de l'ARN substrat à la poche de pseudouridylation de l'ARN guide.

1) Quel bilan faut-il tirer des données que nous avons obtenues lors des travaux effectués à partir des connaissances que nous avions des différentes structures 3D ?

Comme déjà mentionné dans l'introduction, lorsque j'ai débuté mon travail de thèse, seules les structures 3D de la plupart des ARN:Ψ-synthases bactériennes étaient connues et le complexe entre TruB et le bras TΨC d'un ARNt était la référence du mode d'interaction avec l'ARN des membres de cette famille de protéines. Puis ont été résolues les structures 3D de protéines libres ou en complexes des particules des archaea, des particules sRNP libres ou, depuis très récemment, en association avec l'ARN substrat. Ces données ont conforté les conclusions et interprétations de nos résultats de biochimie et des analyses fonctionnelles qui étaient en cours, comme par exemple le positionnement de aCBF5 sur l'ARN. Elles ont aussi été très riches en information pour dégager de nouvelles pistes d'investigation. Ainsi, nous avons :

- identifié des résidus impliqués dans un réseau d'interaction entre aNOP10 et aCBF5, ces deux protéines étant le couple minimal pour détecter une activité guidée *in vitro*.

- mis en évidence que le rôle positif de L7Ae sur l'activité des particules repose pour une part non négligeable sur son interaction avec la protéine aNOP10. A la publication de la structure 3D d'une sRNP H/ACA entière (Li et Ye, 2006), la présence d'une interaction entre ces deux protéines était une surprise puisque qu'aucun complexe aNOP10-L7Ae n'avait été détecté en absence d'ARN (Baker *et al.*, 2005; Charpentier *et al.*, 2005). Nous avons souhaité étudier le rôle de cette interaction puisqu'alors l'effet positif de L7Ae sur l'activité des sRNP n'était pas décrit comme étant en lien avec aNOP10 et semblait provenir du remodelage de l'ARN provoqué par la fixation de L7Ae sur le motif en K-turn (Liang, B. *et al.*, 2007; Youssef *et al.*, 2007). Nos données suggèrent qu'une fixation stable de l'ARN substrat à la particule repose sur la mise en place de l'interaction L7Ae-aNOP10 (article n°4, Fourmann *et al.*, en préparation). A ce stade nous souhaiterions vérifier si la perte de stabilité que nous observons avec le variant de L7Ae H70A L74A E77A et les variants de aNOP10 Y41A Y44A et R47A, repose sur un problème de positionnement de l'ARN substrat au sein des particules, pouvant correspondre à un défaut des étapes 2 et/ou 3 de la réaction de pseudouridylation (Fig. i62). A cette fin, nous pourrons employer la méthode de spectrométrie de fluorescence développée par l'équipe de H. Li (Liang, B. *et al.*, 2007) et basée sur l'emploi d'un ARN portant un résidu 2-AP à la position 3' contiguë au résidu U cible.

A présent, les données structurales récentes des complexes formés entre une sRNP H/ACA et un ARN substrat peuvent aider à éclaircir notre compréhension du mécanisme de la réaction catalytique. Le positionnement du substrat et les contacts établis entre celui-ci et les résidus du site actif et de la boucle β7/β10 de aCBF5 sont connus. Néanmoins, dans chacune des structures 3D, le résidu U cible est converti en 5-fluoro-6-hydroxypseudouridine (5-F-6-OH Ψ) (Duan *et al.*, 2009; Liang *et al.*, 2009). Ainsi, les contacts sont ceux existant dans une structure figée à une étape finale de la réaction catalytique. L'incertitude qui résidait au sujet du mécanisme catalytique, à savoir soit une attaque du résidu catalytique en C1' du ribose ou soit en position C6

de l'uracile existe donc toujours (Fig. i14). Toutefois, l'organisation du site actif et le positionnement du substrat portant le résidu 5-F-6-OH Ψ sont similaires dans les complexes sRNP-ARN substrat et TruB-ARNt, de plus, l'effet de certaines substitutions ont des effets analogues lorsqu'elles sont introduites dans aCBF5 ou dans TruB. Ces observations indiquent que le mécanisme catalytique de aCBF5 est similaire à celui de TruB et probablement aux autres ARN:Ψ-synthases. Tous les variants produits dans ce travail ont montré une très grande solubilité ce qui n'est pas le cas pour tous les variants de TruB. Cette caractéristique propre à la protéine d'archaea est intéressante et permettrait de poursuivre des études en analysant des variants encore jamais isolés du site actif d'une ARN:Ψ-synthase. Pour les variants sélectionnés, il sera nécessaire d'effectuer des mesures des constantes de vitesses (k_{on} et k_{off}). Nous envisageons de développer une méthodologie basée sur le FRET (Fluorescent Resonance Energy Transfer) pour suivre l'entrée du substrat dans la particule à partir de deux fluorophores, l'un placé sur le substrat et l'autre sur l'ARN guide. Leur rapprochement dans l'espace suite à la fixation du substrat à la particule conduira au phénomène de FRET. L'apparition et l'extinction de fluorescence permettront d'exprimer les valeurs de k_{on} et k_{off}. La vitesse k_{on} représente la vitesse apparente de réalisation des étapes 2 à 4 de la réaction de pseudouridylation (Fig. i62), et k_{off} la vitesse apparente de l'étape 5. Dans des conditions de single turn-over (lorsque la quantité de substrat est limitante), la vitesse de réaction de aCBF5 dans le système guidé est estimée à 0,32 min^{-1} (Duan *et al.*, 2009), ce qui est comparable à TruB qui modifie les ARNt avec un kcat de 0,24 min^{-1} (Gu *et al.*, 1998). Les valeurs de k_{on} et k_{off} pourront donner une idée plus précise de l'identité de l'étape limitante du mécanisme. De plus, les vitesses pourront aussi être mesurées pour les variants de aCBF5 et ainsi contribueront à mieux comprendre le rôle précis des résidus dans le mécanisme de la réaction. Certains variants nous semblent très intéressants à étudier en particulier ceux portant des substitutions aux positions Y110 et Y179 qui conduisent à des effets délétères sur l'activité. Le comportement du variant R143A est intrigant et il sera intéressant de confirmer l'identité de l'étape stimulée lors de la réaction catalysée par les sRNP portant ce variant du pouce de aCBF5.

Les structures 3D ont mis en évidence que le résidu thréonine 83 (ou 80 chez *P. abyssi*) aurait un rôle majeur dans la reconnaissance, la stabilisation et/ou la libération du substrat dans le site actif de aCBF5 (Fig. r27) (Duan *et al.*, 2009; Liang *et al.*, 2009). En effet, ce résidu interagit soit directement par liaison hydrogène, soit via une molécule d'eau, avec les nucléotides en simple brin de l'ARN substrat. Il sera donc intéressant de caractériser l'effet de substitutions de ce résidu par les méthodes déjà employées au cours de ce travail de thèse et celle permettant la mesure des constantes cinétiques.

Enfin, l'organisation structurale d'une RNP à double motif tige-boucle reste une question à résoudre dans le futur. Les sRNP à double motifs tige-boucle d'archaea s'organiseraient à partir d'un jeu de chacune des protéines par tige-boucle et une architecture similaire est suggérée être adoptée par les snoRNP H/ACA chez les eucaryotes (Li et Ye, 2006). Néanmoins, le rôle joué par les domaines additionnels des protéines eucaryotes reste à ce jour inconnu.

2) Par quel mécanisme le substrat modifié est libéré des particules?

Une des limitations de la radiocristallographie est que cette méthode renseigne sur un état figé d'une structure dans un cristal. Néanmoins, les diverses structures 3D ont permis de caractériser des complexes dans lesquels la conformation de la boucle β7/β10 de aCBF5 n'était pas dans le même état. Il a été observé qu'il existait une interaction entre des résidus de aGAR1 et cette boucle. Enfin, le rôle positif de aGAR1 sur la vitesse de la réaction a été démontré. Ces données ont conduit à proposer un modèle dans lequel le aGAR1 modulerait le placement de la boucle β7/β10 entre une position ouverte et fermée, conditionnant l'entrée et la sortie du substrat. La question qui subsiste est de savoir quel message moléculaire rend compte de l'avancée de la réaction afin que le passage à une forme ouverte se produise une fois que le résidu Ψ est formé, ce qui autorisera la libération du produit de la réaction. L'emploi de nouvelles méthodologies est nécessaire pour effectuer un

suivi du déroulement la réaction, en particulier des méthodes de cinétique structurales. Au cours de ce travail, deux méthodes ont été développées dans ce but :

- la méthode d'analyse de la formation des complexes sRNP-ARN substrat (complexes CII') par empreinte sur gel (cf II.D.2.2 des résultats) permet de suivre la présence de ces complexes au cours du temps.

- la méthode d'analyse du turn-over sur gel (cf II.C.3.5 des résultats) nous renseigne sur la libération effective ou non du substrat de la particule.

L'utilisation de ces techniques en combinaison permet d'identifier l'étape du mécanisme limitée par exemple par la présence d'une substitution. Néanmoins, ces deux méthodes ont des limites cinétiques qui les rendent peu adaptées pour des analyses en temps réel. Pour progresser dans la compréhension de la cinétique structurale de ces particules, de nouvelles techniques sont en cours de mise au point au laboratoire, basées sur le suivi de fluorescence. Ce travail sera poursuivi par A.-S. Tillault, étudiante en thèse.

Si la fonction de aGAR1 est de moduler la position de la boucle β7/β10 du pouce, le mécanisme par lequel le substrat modifié est libéré de la poche de pseudouridylation reste mystérieuse. Cela nécessite la rupture des appariements SH1 et SH2 et ce, sans l'intervention d'une activité hélicase exogène. Nos données obtenues sur l'ARN:Ψ-synthase aCBF5 suggèrent que certains résidus de son site actif participent activement à la discrimination de l'ARN substrat. En effet, nous montrons que le résidu Y179 de aCBF5, et dans une moindre mesure les résidus D82 et Y110, seraient en partie responsables de la reconnaissance spécifique d'un substrat sauvage vis-à-vis d'un substrat dont le résidu U cible est substitué en C. Aucun des variants du site actif de aCBF5 que nous avons produits dans ce travail n'a cependant pas permis de mettre en évidence des acides aminés requis pour la discrimination entre un substrat portant un résidu U et un substrat portant le résidu Ψ. Si de tels acides aminés sont identifiés, ils pourraient représenter des déterminants contribuant

à l'étape de libération du substrat modifié. Pour parachever ces études de spécificité, nous envisageons d'employer la technique basée sur le suivi de fluorescence ainsi que l'emploi de différents ARN substrats dont le résidu U cible serait substitué par d'autres résidus tels que la ribothymine ou la désoxyuridine.

3) Les activités guidées et non guidées de aCBF5

Nous avons identifié que aCBF5 peut catalyser efficacement, sans avoir recours à un sRNA H/ACA, la formation du résidu Ψ55 dans les ARNt élongateurs et dans ce système non guidé, chacune des protéines aNOP10 et aGAR1 stimule cette activité. Nous avons démontré que aNOP10 contribuait de façon différente à l'activité de aCBF5 dans le système guidé (sRNP H/ACA) et le système non guidé (article n°3, (Muller *et al.*, 2007b)). La contribution de aGAR1 n'a pas encore été analysée en détail : est-ce qu'il s'agit uniquement d'un effet en lien avec la boucle β7/β10 du pouce de aCBF5 ou bien est-ce qu'un réseau d'interaction influencerait le site actif de aCBF5 ?

Nous avons identifié une seconde position pouvant être modifiée par aCBF5 en absence d'ARN guide : il s'agit du résidu Ψ2603 dans l'ARNr 23S et nous pourrions également tester le rôle respectif des protéines aNOP10 et aGAR1 pour l'efficacité de la réaction de modification à cette position. Pour l'instant les tests ont uniquement été réalisés en présence du complexe aCBF5-aNOP10-aGAR1. Le rôle des protéines aNOP10 et aGAR1 pourrait se révéler être différent pour le substrat ARNr en comparaison au substrat ARNt pour la raison que la séquence CCA est absente dans la séquence de l'ARNr. Nous avons montré qu'un ARNt tronqué de la séquence CCA pouvait être modifié à la position 55 mais uniquement lorsque aNOP10 ou aGAR1 sont présentes.

Un autre point qu'il nous semble intéressant d'aborder concerne l'accession du site actif de aCBF5 au résidu U à modifier lors d'un mécanisme guidé et non guidé. Pour les ARN:Ψ-synthases de la famille TruB, une histine (H43 chez TruB et H77

chez aCBF5 *P. abyssi*) est importante pour l'extrusion du résidu U et nous avons contribué à ces données (article n°3, (Muller *et al.*, 2007b)). La structure d'une particule sRNP en association avec un substrat ARN fait apparaître qu'une autre histidine conservée H60 de aCBF5 est située à proximité du résidu U (Duan *et al.*, 2009) et celle-ci permettrait de reconnaître des ARNt dépourvus de la paire de base réverse Hoogsteen U54 : A58 des ARNt (Gurha *et al.*, 2007). Il serait intéressant d'étudier le rôle de chacune de ces histines sur l'activité non guidée à la position 2603 de l'ARNr 23S et de tester l'importance de H60 pour l'activité guidée. Il est possible que contrairement à H77, H60 soit cruciale pour l'activité de aCBF5 au sein des particules.

Enfin, nous n'avons pas encore abordé le lien fonctionnel entre les ARN:Ψ-synthases aPUS10 et aCBF5. La première modifie les positions 54 et 55 et la seconde la position 55 dans les ARNt. aPUS10 n'apparaît pas modifier la position Ψ2603 dans l'ARNr. Néanmoins, une interrogation subsiste : qu'en est-il de toutes ces activités *in vivo* ? Des approches de génétique chez l'archaea *Thermococcus kodakaraensis* sont actuellement employées en collaboration avec R. Fujikane du laboratoire de P. Forterrre (IGM, Orsay). Des cellules invalidées dans les gènes codant aCBF5 et aPUS10 sont en cours d'analyse afin de tester leur capacité à modifier les positions Ψ54, Ψ55 dans les ARNt et Ψ2603 dans l'ARNr 23S. De plus, la génétique chez les archaea reste un défi majeur qui permettrait d'aborder le rôle joué *in vivo* par les ARN:Ψ-synthases, les protéines associées, les résidus Ψ, et tester l'effet de mutations dans ces protéines.

REFERENCES BIBLIOGRAPHIQUES

Références Bibliographiques

Aittaleb, M., R. Rashid, Q. Chen, J. R. Palmer, C. J. Daniels et H. Li (2003). "Structure and function of archaeal box C/D sRNP core proteins." Nat Struct Biol 10(4): 256-63.

Alian, A., A. Degiovanni, S. L. Griner, J. S. Finer-Moore et R. M. Stroud (2009). "Crystal Structure of an RluF-RNA Complex: A Base-Pair Rearrangement Is the Key to Selectivity of RluF for U2604 of the Ribosome." J Mol Biol.

Allers, T. et M. Mevarech (2005). "Archaeal genetics - the third way." Nat Rev Genet 6(1): 58-73.

Alter, B. P. (2000). "Molecular medicine and bone marrow failure syndromes." J Pediatr 136(3): 275-6.

Anantharaman, V., E. V. Koonin et L. Aravind (2002). "Comparative genomics and evolution of proteins involved in RNA metabolism." Nucleic Acids Res 30(7): 1427-64.

Ansmant, I., S. Massenet, H. Grosjean, Y. Motorin et C. Branlant (2000). "Identification of the Saccharomyces cerevisiae RNA:pseudouridine synthase responsible for formation of psi(2819) in 21S mitochondrial ribosomal RNA." Nucleic Acids Res 28(9): 1941-6.

Ansmant, I. et I. Motorin (2001). "[Identification of RNA modification enzymes using sequence homology]." Mol Biol (Mosk) 35(2): 248-67.

Aravind, L. et E. V. Koonin (1999). "Novel predicted RNA-binding domains associated with the translation machinery." J Mol Evol 48(3): 291-302.

Arena, F., G. Ciliberto, S. Ciampi et R. Cortese (1978). "Purification of pseudouridylate synthetase I from Salmonella typhimurium." Nucleic Acids Res 5(12): 4523-36.

Aris, J. P. et G. Blobel (1991). "cDNA cloning and sequencing of human fibrillarin, a conserved nucleolar protein recognized by autoimmune antisera." Proc Natl Acad Sci U S A 88(3): 931-5.

Armanios, M., J. L. Chen, Y. P. Chang, R. A. Brodsky, A. Hawkins, C. A. Griffin, J. R. Eshleman, A. R. Cohen, A. Chakravarti, A. Hamosh et C. W. Greider (2005). "Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita." Proc Natl Acad Sci U S A 102(44): 15960-4.

Atzorn, V., P. Fragapane et T. Kiss (2004). "U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production." Mol Cell Biol 24(4): 1769-78.

Auffinger, P., S. Louise-May et E. Westhof (1999). "Molecular dynamics simulations of solvated yeast tRNA(Asp)." Biophys J 76(1 Pt 1): 50-64.

Auffinger, P. et E. Westhof (1996). "H-bond stability in the tRNA(Asp) anticodon hairpin: 3 ns of multiple molecular dynamics simulations." Biophys J 71(2): 940-54.

Auffinger, P. et E. Westhof (1998). "Hydration of RNA base pairs." J Biomol Struct Dyn 16(3): 693-707.

Auffinger, P. et E. Westhof (1999). "Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes." J Mol Biol 292(3): 467-83.

Auzende, J.-M. et T. Urabe (1989). "Le cadre géologique d'un site hydrothermal actif: la campagne "Starmer 1" du submersible Nautile dans le bassin Nord Fidjien." C. R. Acad. Sci. Paris 302: 1787-1795.

Aza-Blanc, P., C. L. Cooper, K. Wagner, S. Batalov, Q. L. Deveraux et M. P. Cooke (2003). "Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening." Mol Cell 12(3): 627-37.

Bachand, F., F. M. Boisvert, J. Cote, S. Richard et C. Autexier (2002). "The product of the survival of motor neuron (SMN) gene is a human telomerase-associated protein." Mol Biol Cell 13(9): 3192-202.

Bachellerie, J. P., J. Cavaille et A. Huttenhofer (2002). "The expanding snoRNA world." Biochimie 84(8): 775-90.

Bagni, C. et B. Lapeyre (1998). "Gar1p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element." J Biol Chem 273(18): 10868-73.

Baker, D. L., N. T. Seyfried, H. Li, R. Orlando, R. M. Terns et M. P. Terns (2008). "Determination of protein-RNA interaction sites in the Cbf5-H/ACA guide RNA complex by mass spectrometric protein footprinting." Biochemistry 47(6): 1500-10.

Baker, D. L., O. A. Youssef, M. I. Chastkofsky, D. A. Dy, R. M. Terns et M. P. Terns (2005). "RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP." Genes Dev 19(10): 1238-48.

Bakin, A., J. A. Kowalak, J. A. McCloskey et J. Ofengand (1994). "The single pseudouridine residue in Escherichia coli 16S RNA is located at position 516." Nucleic Acids Res 22(18): 3681-4.

Bakin, A. et J. Ofengand (1995). "Mapping of the 13 pseudouridine residues in Saccharomyces cerevisiae small subunit ribosomal RNA to nucleotide resolution." Nucleic Acids Res 23(16): 3290-4.

Balakin, A. G., L. Smith et M. J. Fournier (1996). "The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions." Cell 86(5): 823-34.

Ballarino, M., M. Morlando, F. Pagano, A. Fatica et I. Bozzoni (2005). "The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in Saccharomyces cerevisiae." Mol Cell Biol 25(13): 5396-403.

Ban, N., P. Nissen, J. Hansen, P. B. Moore et T. A. Steitz (2000). "The complete atomic structure of the large ribosomal subunit at 2.4 A resolution." Science 289(5481): 905-20.

Becker, H. F., Y. Motorin, R. J. Planta et H. Grosjean (1997). "The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs." Nucleic Acids Res 25(22): 4493-9.

Behm-Ansmant, I., C. Branlant et Y. Motorin (2007). "The Saccharomyces cerevisiae Pus2 protein encoded by YGL063w ORF is a mitochondrial tRNA:Psi27/28-synthase." RNA 13(10): 1641-7.

Behm-Ansmant, I., H. Grosjean, S. Massenet, Y. Motorin et C. Branlant (2004). "Pseudouridylation at position 32 of mitochondrial and cytoplasmic tRNAs requires two distinct enzymes in Saccharomyces cerevisiae." J Biol Chem 279(51): 52998-3006.

Behm-Ansmant, I., S. Massenet, F. Immel, J. R. Patton, Y. Motorin et C. Branlant (2006). "A previously unidentified activity of yeast and mouse RNA:pseudouridine synthases 1 (Pus1p) on tRNAs." RNA 12(8): 1583-93.

Behm-Ansmant, I., A. Urban, X. Ma, Y. T. Yu, Y. Motorin et C. Branlant (2003). "The Saccharomyces cerevisiae U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs." RNA 9(11): 1371-82.

Bell, S. D. et S. P. Jackson (2001). "Mechanism and regulation of transcription in archaea." Curr Opin Microbiol 4(2): 208-13.

Berthon, J., D. Cortez et P. Forterre (2008). "Genomic context analysis in Archaea suggests previously unrecognized links between DNA replication and translation." Genome Biol 9(4): R71.

Bianchi, A. et D. Shore (2008). "How telomerase reaches its end: mechanism of telomerase regulation by the telomeric complex." Mol Cell 31(2): 153-65.

Bilbille, Y., F. A. Vendeix, R. Guenther, A. Malkiewicz, X. Ariza, J. Vilarrasa et P. F. Agris (2009). "The structure of the human tRNALys3 anticodon bound to the HIV genome is stabilized by modified nucleosides and adjacent mismatch base pairs." Nucleic Acids Res.

Björk, G. R., Ed. (1995). Biosynthesis and function of modified nucleosides. tRNA: structure, biosynthesis and function. Washigton DC, ACM press.

Bohnsack, M. T., M. Kos et D. Tollervey (2008). "Quantitative analysis of snoRNA association with pre-ribosomes and release of snR30 by Rok1 helicase." EMBO Rep 9(12): 1230-6.

Bortolin, M. L., P. Ganot et T. Kiss (1999). "Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs." Embo J 18(2): 457-69.

Boulon, S., N. Marmier-Gourrier, B. Pradet-Balade, L. Wurth, C. Verheggen, B. E. Jady, B. Rothe, C. Pescia, M. C. Robert, T. Kiss, B. Bardoni, A. Krol, C. Branlant, C. Allmang, E. Bertrand et B. Charpentier (2008). "The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery." J Cell Biol 180(3): 579-95.

Bousquet-Antonelli, C., Y. Henry, P. G'Elugne J, M. Caizergues-Ferrer et T. Kiss (1997). "A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs." Embo J 16(15): 4770-6.

Branlant, C., A. Krol, J. P. Ebel, H. Gallinaro, E. Lazar et M. Jacob (1981a). "The conformation of chicken, rat and human U1A RNAs in solution." Nucleic Acids Res 9(4): 841-58.

Branlant, C., A. Krol, J. P. Ebel, E. Lazar, B. Haendler et M. Jacob (1982). "U2 RNA shares a structural domain with U1, U4, and U5 RNAs." EMBO J 1(10): 1259-65.

Branlant, C., A. Krol, M. A. Machatt et J. P. Ebel (1979). "Structural study of ribosomal 23 S RNA from Escherichia coli." FEBS Lett 107(1): 177-81.

Branlant, C., A. Krol, M. A. Machatt, J. Pouyet, J. P. Ebel, K. Edwards et H. Kossel (1981b). "Primary and secondary structures of Escherichia coli MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs." Nucleic Acids Res 9(17): 4303-24.

Cadwell, C., H. J. Yoon, Y. Zebarjadian et J. Carbon (1997). "The yeast nucleolar protein Cbf5p is involved in rRNA biosynthesis and interacts genetically with the RNA polymerase I transcription factor RRN3." Mol Cell Biol 17(10): 6175-83.

Cahill, N. M., K. Friend, W. Speckmann, Z. H. Li, R. M. Terns, M. P. Terns et J. A. Steitz (2002). "Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP." Embo J 21(14): 3816-28.

Cavaille, J., K. Buiting, M. Kiefmann, M. Lalande, C. I. Brannan, B. Horsthemke, J. P. Bachellerie, J. Brosius et A. Huttenhofer (2000). "Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization." Proc Natl Acad Sci U S A 97(26): 14311-6.

Cavaille, J., M. Nicoloso et J. P. Bachellerie (1996). "Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides." Nature 383(6602): 732-5.

Cech, T. R. (2004). "Beginning to understand the end of the chromosome." Cell 116(2): 273-9.

Cech, T. R. (2009). "Crawling out of the RNA world." Cell 136(4): 599-602.

Charette, M. et M. W. Gray (2000). "Pseudouridine in RNA: what, where, how, and why." IUBMB Life 49(5): 341-51.

Charpentier, B., J. B. Fourmann et C. Branlant (2007). "Reconstitution of archaeal H/ACA sRNPs and test of their activity." Methods Enzymol 425: 389-405.

Charpentier, B., S. Muller et C. Branlant (2005). "Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation." Nucleic Acids Res 33(10): 3133-44.

Charron, C., X. Manival, A. Clery, V. Senty-Segault, B. Charpentier, N. Marmier-Gourrier, C. Branlant et A. Aubry (2004). "The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity." J Mol Biol 342(3): 757-73.

Chen, J. et J. R. Patton (1999). "Cloning and characterization of a mammalian pseudouridine synthase." RNA 5(3): 409-19.

Chen, J. L. et C. W. Greider (2004). "An emerging consensus for telomerase RNA structure." Proc Natl Acad Sci U S A 101(41): 14683-4.

Chen, J. L. et C. W. Greider (2005). "Functional analysis of the pseudoknot structure in human telomerase RNA." Proc Natl Acad Sci U S A 102(23): 8080-5; discussion 8077-9.

Cohen, G. N., V. Barbe, D. Flament, M. Galperin, R. Heilig, O. Lecompte, O. Poch, D. Prieur, J. Querellou, R. Ripp, J. C. Thierry, J. Van der Oost, J. Weissenbach, Y. Zivanovic et P. Forterre (2003). "An integrated analysis of the genome of the hyperthermophilic archaeon Pyrococcus abyssi." Mol Microbiol 47(6): 1495-512.

Cohn, W. E. (1960). "Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics." J Biol Chem 235: 1488-98.

Conrad, J., L. Niu, K. Rudd, B. G. Lane et J. Ofengand (1999). "16S ribosomal RNA pseudouridine synthase RsuA of Escherichia coli: deletion, mutation of the conserved Asp102 residue, and sequence comparison among all other pseudouridine synthases." RNA 5(6): 751-63.

Conrad, J., D. Sun, N. Englund et J. Ofengand (1998). "The rluC gene of Escherichia coli codes for a pseudouridine synthase that is solely responsible for synthesis of pseudouridine at positions 955, 2504, and 2580 in 23 S ribosomal RNA." J Biol Chem 273(29): 18562-6.

Cortese, R., H. O. Kammen, S. J. Spengler et B. N. Ames (1974). "Biosynthesis of pseudouridine in transfer ribonucleic acid." J Biol Chem 249(4): 1103-8.

Counter, C. M., M. Meyerson, E. N. Eaton et R. A. Weinberg (1997). "The catalytic subunit of yeast telomerase." Proc Natl Acad Sci U S A 94(17): 9202-7.

Darzacq, X., B. E. Jady, C. Verheggen, A. M. Kiss, E. Bertrand et T. Kiss (2002). "Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs." EMBO J 21(11): 2746-56.

Darzacq, X., N. Kittur, S. Roy, Y. Shav-Tal, R. H. Singer et U. T. Meier (2006). "Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells." J Cell Biol 173(2): 207-18.

Davis, F. F. et F. W. Allen (1957). "Ribonucleic acids from yeast which contain a fifth nucleotide." J Biol Chem 227(2): 907-15.

de Lange, T. (2005). "Shelterin: the protein complex that shapes and safeguards human telomeres." Genes Dev 19(18): 2100-10.

Decatur, W. A. et M. J. Fournier (2002). "rRNA modifications and ribosome function." Trends Biochem Sci 27(7): 344-51.

Decatur, W. A. et M. N. Schnare (2008). "Different mechanisms for pseudouridine formation in yeast 5S and 5.8S rRNAs." Mol Cell Biol 28(10): 3089-100.

Del Campo, M., Y. Kaya et J. Ofengand (2001). "Identification and site of action of the remaining four putative pseudouridine synthases in Escherichia coli." RNA 7(11): 1603-15.

Del Campo, M., J. Ofengand et A. Malhotra (2004). "Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of Escherichia coli." RNA 10(2): 231-9.

Del Campo, M., C. Recinos, G. Yanez, S. C. Pomerantz, R. Guymon, P. F. Crain, J. A. McCloskey et J. Ofengand (2005). "Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of Haloarcula marismortui and Deinococcus radiodurans." RNA 11(2): 210-9.

Dennis, P. P. (1997). "Ancient ciphers: translation in Archaea." Cell 89(7): 1007-10.

Dez, C., A. Henras, B. Faucon, D. Lafontaine, M. Caizergues-Ferrer et Y. Henry (2001). "Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p." Nucleic Acids Res 29(3): 598-603.

Dez, C., J. Noaillac-Depeyre, M. Caizergues-Ferrer et Y. Henry (2002). "Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs." Mol Cell Biol 22(20): 7053-65.

Dokal, I. (2000). "Dyskeratosis congenita in all its forms." Br J Haematol 110(4): 768-79.

Dokal, I. et T. Vulliamy (2003). "Dyskeratosis congenita: its link to telomerase and aplastic anaemia." Blood Rev 17(4): 217-25.

Dong, X., Y. Bessho, R. Shibata, M. Nishimoto, M. Shirouzu, S. Kuramitsu et S. Yokoyama (2006). "Crystal structure of tRNA pseudouridine synthase TruA from Thermus thermophilus HB8." RNA Biol 3(3): 115-22.

Dragon, F., V. Pogacic et W. Filipowicz (2000). "In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs." Mol Cell Biol 20(9): 3037-48.

Duan, J., L. Li, J. Lu, W. Wang et K. Ye (2009). "Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase." Mol Cell 34(4): 427-39.

Durant, P. C. et D. R. Davis (1997). "The effect of pseudouridine and pH on the structure and dynamics of the anticodon stem-loop of tRNA(Lys,3)." Nucleic Acids Symp Ser(36): 56-7.

Durant, P. C. et D. R. Davis (1999). "Stabilization of the anticodon stem-loop of tRNALys,3 by an A+-C base-pair and by pseudouridine." J Mol Biol 285(1): 115-31.

Ejby, M., M. A. Sorensen et S. Pedersen (2007). "Pseudouridylation of helix 69 of 23S rRNA is necessary for an effective translation termination." Proc Natl Acad Sci U S A 104(49): 19410-5.

Epstein, P., R. Reddy, D. Henning et H. Busch (1980). "The nucleotide sequence of nuclear U6 (4.7 S) RNA." J Biol Chem 255(18): 8901-6.

Ericsson, U. B., P. Nordlund et B. M. Hallberg (2004). "X-ray structure of tRNA pseudouridine synthase TruD reveals an inserted domain with a novel fold." FEBS Lett 565(1-3): 59-64.

Fatica, A., M. Dlakic et D. Tollervey (2002). "Naf1 p is a box H/ACA snoRNP assembly factor." RNA 8(12): 1502-14.

Fatica, A., S. Galardi, F. Altieri et I. Bozzoni (2000). "Fibrillarin binds directly and specifically to U16 box C/D snoRNA." RNA 6(1): 88-95.

Fatica, A. et D. Tollervey (2003). "Insights into the structure and function of a guide RNP." Nat Struct Biol 10(4): 237-9.

Fayet-Lebaron, E., V. Atzorn, Y. Henry et T. Kiss (2009). "18S rRNA processing requires base pairings of snR30 H/ACA snoRNA to eukaryote-specific 18S sequences." EMBO J.

Ferre-D'Amare, A. R. (2003). "RNA-modifying enzymes." Curr Opin Struct Biol 13(1): 49-55.

Filipowicz, W. et V. Pogacic (2002). "Biogenesis of small nucleolar ribonucleoproteins." Curr Opin Cell Biol 14(3): 319-27.

Fletcher, J. E., P. R. Copeland, D. M. Driscoll et A. Krol (2001). "The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2." RNA 7(10): 1442-53.

Foster, P. G., L. Huang, D. V. Santi et R. M. Stroud (2000). "The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I." Nat Struct Biol 7(1): 23-7.

Freeland, S. J., R. D. Knight et L. F. Landweber (1999). "Do proteins predate DNA?" Science 286(5440): 690-2.

Fromont-Racine, M., B. Senger, C. Saveanu et F. Fasiolo (2003). "Ribosome assembly in eukaryotes." Gene 313: 17-42.

Fu, D. et K. Collins (2007). "Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation." Mol Cell 28(5): 773-85.

Gagnon, K. T., X. Zhang, P. F. Agris et E. S. Maxwell (2006). "Assembly of the archaeal box C/D sRNP can occur via alternative pathways and requires temperature-facilitated sRNA remodeling." J Mol Biol 362(5): 1025-42.

Galardi, S., A. Fatica, A. Bachi, A. Scaloni, C. Presutti et I. Bozzoni (2002). "Purified box C/D snoRNPs are able to reproduce site-specific 2'-O-methylation of target RNA in vitro." Mol Cell Biol 22(19): 6663-8.

Ganot, P., M. L. Bortolin et T. Kiss (1997a). "Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs." Cell 89(5): 799-809.

Ganot, P., M. Caizergues-Ferrer et T. Kiss (1997b). "The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation." Genes Dev 11(7): 941-56.

Gaspin, C., J. Cavaille, G. Erauso et J. P. Bachellerie (2000). "Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the Pyrococcus genomes." J Mol Biol 297(4): 895-906.

Gautier, T., T. Berges, D. Tollervey et E. Hurt (1997). "Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis." Mol Cell Biol 17(12): 7088-98.

Gilbert, W. (1986). "Origin of life : The RNA world." Nature 319(6055): 618.

Gilson, E. et V. Geli (2007). "How telomeres are replicated." Nat Rev Mol Cell Biol 8(10): 825-38.

Girard, J. P., C. Bagni, M. Caizergues-Ferrer, F. Amalric et B. Lapeyre (1994). "Identification of a segment of the small nucleolar ribonucleoprotein-associated protein GAR1 that is sufficient for nucleolar accumulation." J Biol Chem 269(28): 18499-506.

Girard, J. P., H. Lehtonen, M. Caizergues-Ferrer, F. Amalric, D. Tollervey et B. Lapeyre (1992). "GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast." Embo J 11(2): 673-82.

Godin, K. S., H. Walbott, N. Leulliot, H. van Tilbeurgh et G. Varani (2009). "The Box H/ACA snoRNP Assembly Factor Shq1p is a Chaperone Protein Homologous to Hsp90 Cochaperones that Binds to the Cbf5p Enzyme." J Mol Biol.

Goody, T. A., S. E. Melcher, D. G. Norman et D. M. Lilley (2004). "The kink-turn motif in RNA is dimorphic, and metal ion-dependent." RNA 10(2): 254-64.

Gottschalk, A., G. Neubauer, J. Banroques, M. Mann, R. Luhrmann et P. Fabrizio (1999). "Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP." EMBO J 18(16): 4535-48.

Grabowski, B. et Z. Kelman (2003). "Archeal DNA replication: eukaryal proteins in a bacterial context." Annu Rev Microbiol 57: 487-516.

Green, C. J., H. O. Kammen et E. E. Penhoet (1982). "Purification and properties of a mammalian tRNA pseudouridine synthase." J Biol Chem 257(6): 3045-52.

Griffith, J. D., L. Comeau, S. Rosenfield, R. M. Stansel, A. Bianchi, H. Moss et T. de Lange (1999). "Mammalian telomeres end in a large duplex loop." Cell 97(4): 503-14.

Grosjean, H., J. Edqvist, K. B. Straby et R. Giege (1996). "Enzymatic formation of modified nucleosides in tRNA: dependence on tRNA architecture." J Mol Biol 255(1): 67-85.

Grozdanov, P. N., S. Roy, N. Kittur et U. T. Meier (2009). "SHQ1 is required prior to NAF1 for assembly of H/ACA small nucleolar and telomerase RNPs." RNA.

Gu, X., Y. Liu et D. V. Santi (1999). "The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA." Proc Natl Acad Sci U S A 96(25): 14270-5.

Gu, X., M. Yu, K. M. Ivanetich et D. V. Santi (1998). "Molecular recognition of tRNA by tRNA pseudouridine 55 synthase." Biochemistry 37(1): 339-43.

Gurha, P. et R. Gupta (2008). "Archaeal Pus10 proteins can produce both pseudouridine 54 and 55 in tRNA." RNA 14(12): 2521-7.

Gurha, P., A. Joardar, P. Chaurasia et R. Gupta (2007). "Differential roles of archaeal box H/ACA proteins in guide RNA-dependent and independent pseudouridine formation." RNA Biol 4(2): 101-9.

Gustavsson, M. et H. Ronne (2008). "Evidence that tRNA modifying enzymes are important in vivo targets for 5-fluorouracil in yeast." RNA 14(4): 666-74.

Hall, K. B. et L. W. McLaughlin (1991). "Properties of a U1/mRNA 5' splice site duplex containing pseudouridine as measured by thermodynamic and NMR methods." Biochemistry 30(7): 1795-801.

Hamilton, C. S., T. M. Greco, C. A. Vizthum, J. M. Ginter, M. V. Johnston et E. G. Mueller (2006). "Mechanistic investigations of the pseudouridine synthase RluA using RNA containing 5-fluorouridine." Biochemistry 45(39): 12029-38.

Hamilton, C. S., C. J. Spedaliere, J. M. Ginter, M. V. Johnston et E. G. Mueller (2005). "The roles of the essential Asp-48 and highly conserved His-43 elucidated by the pH dependence of the pseudouridine synthase TruB." Arch Biochem Biophys 433(1): 322-34.

Hamma, T. et A. R. Ferre-D'Amare (2004). "Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution." Structure (Camb) 12(5): 893-903.

Hamma, T. et A. R. Ferre-D'Amare (2006). "Pseudouridine synthases." Chem Biol 13(11): 1125-35.

Hamma, T., S. L. Reichow, G. Varani et A. R. Ferre-D'Amare (2005). "The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs." Nat Struct Mol Biol 12(12): 1101-7.

Hardin, J. W. et R. T. Batey (2006). "The bipartite architecture of the sRNA in an archaeal box C/D complex is a primary determinant of specificity." Nucleic Acids Res 34(18): 5039-51.

Hardin, J. W., F. E. Reyes et R. T. Batey (2009). "Analysis of a Critical Interaction within the Archaeal Box C/D Small Ribonucleoprotein Complex." J Biol Chem 284(22): 15317-24.

Heiss, N. S., S. W. Knight, T. J. Vulliamy, S. M. Klauck, S. Wiemann, P. J. Mason, A. Poustka et I. Dokal (1998). "X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions." Nat Genet 19(1): 32-8.

Henras, A., C. Dez, J. Noaillac-Depeyre, Y. Henry et M. Caizergues-Ferrer (2001). "Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p." Nucleic Acids Res 29(13): 2733-46.

Henras, A., Y. Henry, C. Bousquet-Antonelli, J. Noaillac-Depeyre, J. P. Gelugne et M. Caizergues-Ferrer (1998). "Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs." Embo J 17(23): 7078-90.

Henriquez, R., G. Blobel et J. P. Aris (1990). "Isolation and sequencing of NOP1. A yeast gene encoding a nucleolar protein homologous to a human autoimmune antigen." J Biol Chem 265(4): 2209-15.

Hernandez-Verdun, D. et E. Louvet (2004). "[The nucleolus: structure, functions, amd associated diseases]." Med Sci (Paris) 20(1): 37-44.

Hirose, T., T. Ideue, M. Nagai, M. Hagiwara, M. D. Shu et J. A. Steitz (2006). "A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing." Mol Cell 23(5): 673-84.

Hirose, T., M. D. Shu et J. A. Steitz (2003). "Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells." Mol Cell 12(1): 113-23.

Hoang, C., J. Chen, C. A. Vizthum, J. M. Kandel, C. S. Hamilton, E. G. Mueller et A. R. Ferre-D'Amare (2006). "Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure." Mol Cell 24(4): 535-45.

Hoang, C. et A. R. Ferre-D'Amare (2001). "Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme." Cell 107(7): 929-39.

Hoang, C. et A. R. Ferre-D'Amare (2004). "Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold." RNA 10(7): 1026-33.

Hoang, C., C. S. Hamilton, E. G. Mueller et A. R. Ferre-D'Amare (2005). "Precursor complex structure of pseudouridine synthase TruB suggests coupling of active site perturbations to an RNA-sequestering peripheral protein domain." Protein Sci 14(8): 2201-6.

Hoareau-Aveilla, C., M. Bonoli, M. Caizergues-Ferrer et Y. Henry (2006). "hNaf1 is required for accumulation of human box H/ACA snoRNPs, scaRNPs, and telomerase." RNA 12(5): 832-40.

Hoareau-Aveilla, C., Y. Henry et T. Leblanc (2008). "[Dyskeratosis congenita, a disease caused by defective telomere maintenance]." Med Sci (Paris) 24(4): 390-8.

Huang, G. M., A. Jarmolowski, J. C. Struck et M. J. Fournier (1992). "Accumulation of U14 small nuclear RNA in Saccharomyces cerevisiae requires box C, box D, and a 5', 3' terminal stem." Mol Cell Biol 12(10): 4456-63.

Huang, L., M. Pookanjanatavip, X. Gu et D. V. Santi (1998). "A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst." Biochemistry 37(1): 344-51.

Hur, S. et R. M. Stroud (2007). "How U38, 39, and 40 of many tRNAs become the targets for pseudouridylation by TruA." Mol Cell 26(2): 189-203.

Hur, S., R. M. Stroud et J. Finer-Moore (2006). "Substrate recognition by RNA 5-methyluridine methyltransferases and pseudouridine synthases: a structural perspective." J Biol Chem 281(51): 38969-73.

Ishitani, R., O. Nureki, N. Nameki, N. Okada, S. Nishimura et S. Yokoyama (2003). "Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme." Cell 113(3): 383-94.

Jacobs, S. A., E. R. Podell et T. R. Cech (2006). "Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase." Nat Struct Mol Biol 13(3): 218-25.

Jady, B. E., E. Bertrand et T. Kiss (2004). "Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal." J Cell Biol 164(5): 647-52.

Jiang, W., K. Middleton, H. J. Yoon, C. Fouquet et J. Carbon (1993). "An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules." Mol Cell Biol 13(8): 4884-93.

Jin, H., J. P. Loria et P. B. Moore (2007). "Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA." Mol Cell 26(2): 205-15.

Johnson, L. et D. Soll (1970). "In vitro biosynthesis of pseudouridine at the polynucleotide level by an enzyme extract from Escherichia coli." Proc Natl Acad Sci U S A 67(2): 943-50.

Joyce, G. F. (2002). "The antiquity of RNA-based evolution." Nature 418(6894): 214-21.

Joyce, G. F. (2007). "Structural biology. A glimpse of biology's first enzyme." Science 315(5818): 1507-8.

Kahn, A. (2005). "[Telomeres, diseases and aging]." Med Sci (Paris) 21(5): 451-2.

Kammen, H. O., C. C. Marvel, L. Hardy et E. E. Penhoet (1988). "Purification, structure, and properties of Escherichia coli tRNA pseudouridine synthase I." J Biol Chem 263(5): 2255-63.

Karijolich, J., D. Stephenson et Y. T. Yu (2007). "Biochemical purification of box H/ACA RNPs involved in pseudouridylation." Methods Enzymol 425: 241-62.

Kaya, Y., M. Del Campo, J. Ofengand et A. Malhotra (2004). "Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold." J Biol Chem 279(18): 18107-10.

Kaya, Y. et J. Ofengand (2003). "A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya." RNA 9(6): 711-21.

Keith, G. (1995). "Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography." Biochimie 77(1-2): 142-4.

Khan, M. S. et B. E. Maden (1977). "Nucleotide sequence relationships between vertebrate 5.8 S ribosomal RNAs." Nucleic Acids Res 4(7): 2495-505.

Khanna, M., H. Wu, C. Johansson, M. Caizergues-Ferrer et J. Feigon (2006). "Structural study of the H/ACA snoRNP components Nop10p and the 3' hairpin of U65 snoRNA." RNA 12(1): 40-52.

King, T. H., B. Liu, R. R. McCully et M. J. Fournier (2003). "Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center." Mol Cell 11(2): 425-35.

Kirpekar, F., L. H. Hansen, A. Rasmussen, J. Poehlsgaard et B. Vester (2005). "The archaeon Haloarcula marismortui has few modifications in the central parts of its 23S ribosomal RNA." J Mol Biol 348(3): 563-73.

Kiss-Laszlo, Z., Y. Henry, J. P. Bachellerie, M. Caizergues-Ferrer et T. Kiss (1996). "Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs." Cell 85(7): 1077-88.

Kiss-Laszlo, Z., Y. Henry et T. Kiss (1998). "Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA." Embo J 17(3): 797-807.

Kiss, A. M., B. E. Jady, E. Bertrand et T. Kiss (2004). "Human box H/ACA pseudouridylation guide RNA machinery." Mol Cell Biol 24(13): 5797-807.

Kiss, T. (2002). "Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions." Cell 109(2): 145-8.

Kiss, T. et W. Filipowicz (1995). "Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns." Genes Dev 9(11): 1411-24.

Klein, D. J., T. M. Schmeing, P. B. Moore et T. A. Steitz (2001). "The kink-turn: a new RNA secondary structure motif." Embo J 20(15): 4214-21.

Kolodrubetz, D. et A. Burgum (1991). "Sequence and genetic analysis of NHP2: a moderately abundant high mobility group-like nuclear protein with an essential function in Saccharomyces cerevisiae." Yeast 7(2): 79-90.

Koonin, E. V. (1996). "Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases." Nucleic Acids Res 24(12): 2411-5.

Kouloulias, V., G. Plataniotis, J. Kouvaris, C. Dardoufas, C. Gennatas, N. Uzunoglu, C. Papavasiliou et L. Vlahos (2005). "Chemoradiotherapy combined with intracavitary hyperthermia for anal cancer: feasibility and long-term results from a phase II randomized trial." Am J Clin Oncol 28(1): 91-9.

Krol, A., C. Branlant, E. Lazar, H. Gallinaro et M. Jacob (1981a). "Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs." Nucleic Acids Res 9(12): 2699-716.

Krol, A., H. Gallinaro, E. Lazar, M. Jacob et C. Branlant (1981b). "The nuclear 5S RNAs from chicken, rat and man. U5 RNAs are encoded by multiple genes." Nucleic Acids Res 9(4): 769-87.

Kuhn, J. F., E. J. Tran et E. S. Maxwell (2002). "Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein." Nucleic Acids Res 30(4): 931-41.

Lafontaine, D. L., C. Bousquet-Antonelli, Y. Henry, M. Caizergues-Ferrer et D. Tollervey (1998). "The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase." Genes Dev 12(4): 527-37.

Lafontaine, D. L. et D. Tollervey (1999). "Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability." Rna 5(3): 455-67.

Lafontaine, D. L. et D. Tollervey (2000). "Synthesis and assembly of the box C+D small nucleolar RNPs." Mol Cell Biol 20(8): 2650-9.

LeBel, C. et R. J. Wellinger (2004). "[The structures at the termini of chromosomes: what is there hidden under the cap ?]." Med Sci (Paris) 20(2): 207-12.

Lechertier, T., A. Grob, D. Hernandez-Verdun et P. Roussel (2009). "Fibrillarin and Nop56 interact before being co-assembled in box C/D snoRNPs." Exp Cell Res 315(6): 928-42.

Lecointe, F., G. Simos, A. Sauer, E. C. Hurt, Y. Motorin et H. Grosjean (1998). "Characterization of yeast protein Deg1 as pseudouridine synthase (Pus3) catalyzing the formation of psi 38 and psi 39 in tRNA anticodon loop." J Biol Chem 273(3): 1316-23.

Lenz, H. J., D. J. Manno, K. D. Danenberg et P. V. Danenberg (1994). "Incorporation of 5-fluorouracil into U2 and U6 snRNA inhibits mRNA precursor splicing." J Biol Chem 269(50): 31962-8.

Leulliot, N., K. S. Godin, C. Hoareau-Aveilla, S. Quevillon-Cheruel, G. Varani, Y. Henry et H. Van Tilbeurgh (2007). "The box H/ACA RNP assembly factor Naf1p contains a domain homologous to Gar1p mediating its interaction with Cbf5p." J Mol Biol 371(5): 1338-53.

Li, H. (2008). "Unveiling substrate RNA binding to H/ACA RNPs: one side fits all." Curr Opin Struct Biol 18(1): 78-85.

Li, H. et J. Abelson (2000). "Crystal structure of a dimeric archaeal splicing endonuclease." J Mol Biol 302(3): 639-48.

Li, J., B. Esberg, J. F. Curran et G. R. Bjork (1997). "Three modified nucleosides present in the anticodon stem and loop influence the in vivo aa-tRNA selection in a tRNA-dependent manner." J Mol Biol 271(2): 209-21.

Li, L. et K. Ye (2006). "Crystal structure of an H/ACA box ribonucleoprotein particle." Nature 443(7109): 302-7.

Liang, B., E. J. Kahen, K. Calvin, J. Zhou, M. Blanco et H. Li (2008). "Long-distance placement of substrate RNA by H/ACA proteins." RNA 14(10): 2086-94.

Liang, B., S. Xue, R. M. Terns, M. P. Terns et H. Li (2007). "Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex." Nat Struct Mol Biol.

Liang, B., J. Zhou, E. Kahen, R. M. Terns, M. P. Terns et H. Li (2009). "Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA." Nat Struct Mol Biol.

Liang, X. H., Q. Liu et M. J. Fournier (2007). "rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity." Mol Cell 28(6): 965-77.

Limbach, P. A., P. F. Crain et J. A. McCloskey (1994). "Summary: the modified nucleosides of RNA." Nucleic Acids Res 22(12): 2183-96.

Lin, Y. et C. L. Kielkopf (2008). "X-ray structures of U2 snRNA-branchpoint duplexes containing conserved pseudouridines." Biochemistry 47(20): 5503-14.

Lingner, J., T. R. Cech, T. R. Hughes et V. Lundblad (1997). "Three Ever Shorter Telomere (EST) genes are dispensable for in vitro yeast telomerase activity." Proc Natl Acad Sci U S A 94(21): 11190-5.

Lowe, T. M. et S. R. Eddy (1999). "A computational screen for methylation guide snoRNAs in yeast." Science 283(5405): 1168-71.

Luzzatto, L. et A. Karadimitris (1998). "Dyskeratosis and ribosomal rebellion." Nat Genet 19(1): 6-7.

Ma, X., C. Yang, A. Alexandrov, E. J. Grayhack, I. Behm-Ansmant et Y. T. Yu (2005). "Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism." EMBO J 24(13): 2403-13.

Ma, X., X. Zhao et Y. T. Yu (2003). "Pseudouridylation (Psi) of U2 snRNA in S. cerevisiae is catalyzed by an RNA-independent mechanism." EMBO J 22(8): 1889-97.

Maden, B. E. (1990). "The numerous modified nucleotides in eukaryotic ribosomal RNA." Prog Nucleic Acid Res Mol Biol 39: 241-303.

Madore, E., C. Florentz, R. Giege, S. Sekine, S. Yokoyama et J. Lapointe (1999). "Effect of modified nucleotides on Escherichia coli tRNAGlu structure and on its aminoacylation by glutamyl-tRNA synthetase. Predominant and distinct roles of the mnm5 and s2 modifications of U34." Eur J Biochem 266(3): 1128-35.

Maguire, B. A. et R. A. Zimmermann (2001). "The ribosome in focus." Cell 104(6): 813-6.

Manival, X., C. Charron, J. B. Fourmann, F. Godard, B. Charpentier et C. Branlant (2006). "Crystal structure determination and site-directed mutagenesis of the Pyrococcus abyssi aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity." Nucleic Acids Res 34(3): 826-39.

Marmier-Gourrier, N., A. Clery, V. Senty-Segault, B. Charpentier, F. Schlotter, F. Leclerc, R. Fournier et C. Branlant (2003). "A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA." Rna 9(7): 821-38.

Marrone, A., A. Walne et I. Dokal (2005). "Dyskeratosis congenita: telomerase, telomeres and anticipation." Curr Opin Genet Dev 15(3): 249-57.

Marrone, A., A. Walne, H. Tamary, Y. Masunari, M. Kirwan, R. Beswick, T. Vulliamy et I. Dokal (2007). "Telomerase reverse-transcriptase homozygous mutations in autosomal recessive dyskeratosis congenita and Hoyeraal-Hreidarsson syndrome." Blood 110(13): 4198-205.

Massenet, S., Y. Motorin, D. L. Lafontaine, E. C. Hurt, H. Grosjean et C. Branlant (1999). "Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA." Mol Cell Biol 19(3): 2142-54.

Massenet, S., A. Mougin et C. Branlant, Eds. (1998). Posttranscriptionnal modifications in the U Small Nuclear RNAs. Washington, DC, ASM press.

Matera, A. G., R. M. Terns et M. P. Terns (2007). "Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs." Nat Rev Mol Cell Biol 8(3): 209-20.

Matte, A., G. V. Louie, J. Sivaraman, M. Cygler et S. K. Burley (2005). "Structure of the pseudouridine synthase RsuA from Haemophilus influenzae." Acta Crystallogr Sect F Struct Biol Cryst Commun 61(Pt 4): 350-4.

Maxwell, E. S. et M. J. Fournier (1995). "The small nucleolar RNAs." Annu Rev Biochem 64: 897-934.

McCleverty, C. J., M. Hornsby, G. Spraggon et A. Kreusch (2007). "Crystal structure of human Pus10, a novel pseudouridine synthase." J Mol Biol 373(5): 1243-54.

McKeegan, K. S., C. M. Debieux, S. Boulon, E. Bertrand et N. J. Watkins (2007). "A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly." Mol Cell Biol 27(19): 6782-93.

Meier, U. T. (2003). "Dissecting dyskeratosis." Nat Genet 33(2): 116-7.

Meier, U. T. (2005). "The many facets of H/ACA ribonucleoproteins." Chromosoma 114(1): 1-14.

Meier, U. T. (2006). "How a single protein complex accommodates many different H/ACA RNAs." Trends Biochem Sci 31(6): 311-5.

Meier, U. T. et G. Blobel (1994). "NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria." J Cell Biol 127(6 Pt 1): 1505-14.

Meyerson, M., C. M. Counter, E. N. Eaton, L. W. Ellisen, P. Steiner, S. D. Caddle, L. Ziaugra, R. L. Beijersbergen, M. J. Davidoff, Q. Liu, S. Bacchetti, D. A. Haber et R. A. Weinberg (1997). "hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization." Cell 90(4): 785-95.

Mitchell, J. R., J. Cheng et K. Collins (1999a). "A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end." Mol Cell Biol 19(1): 567-76.

Mitchell, J. R., E. Wood et K. Collins (1999b). "A telomerase component is defective in the human disease dyskeratosis congenita." Nature 402(6761): 551-5.

Mizutani, K., Y. Machida, S. Unzai, S. Y. Park et J. R. Tame (2004). "Crystal structures of the catalytic domains of pseudouridine synthases RluC and RluD from Escherichia coli." Biochemistry 43(15): 4454-63.

Mochizuki, Y., J. He, S. Kulkarni, M. Bessler et P. J. Mason (2004). "Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing." Proc Natl Acad Sci U S A 101(29): 10756-61.

Moore, T., Y. Zhang, M. O. Fenley et H. Li (2004). "Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA." Structure 12(5): 807-18.

Morlando, M., M. Ballarino, P. Greco, E. Caffarelli, B. Dichtl et I. Bozzoni (2004). "Coupling between snoRNP assembly and 3' processing controls box C/D snoRNA biosynthesis in yeast." EMBO J 23(12): 2392-401.

Morrissey, J. P. et D. Tollervey (1993). "Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis." Mol Cell Biol 13(4): 2469-77.

Motorin, Y. et H. Grosjean, Eds. (1998). Chemical structures and classification of posttranscriptionnally modified nucleosides in RNA. The modification and editing of RNA. Washington, DC, ASM Press.

Motorin, Y., G. Keith, C. Simon, D. Foiret, G. Simos, E. Hurt et H. Grosjean (1998). "The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity." RNA 4(7): 856-69.

Mueller, E. G. (2002). "Chips off the old block." Nat Struct Biol 9(5): 320-2.

Muller, S., B. Charpentier, C. Branlant et F. Leclerc (2007a). "A dedicated computational approach for the identification of archaeal H/ACA sRNAs." Methods Enzymol 425: 355-87.

Muller, S., J. B. Fourmann, C. Loegler, B. Charpentier et C. Branlant (2007b). "Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Psi55-synthase and RNA-guided RNA:Psi-synthase activities." Nucleic Acids Res 35(16): 5610-24.

Muller, S., F. Leclerc, I. Behm-Ansmant, J. B. Fourmann, B. Charpentier et C. Branlant (2008). "Combined in silico and experimental identification of the Pyrococcus abyssi H/ACA sRNAs and their target sites in ribosomal RNAs." Nucleic Acids Res 36(8): 2459-75.

Muller, S., A. Urban, A. Hecker, F. Leclerc, C. Branlant et Y. Motorin (2009). "Deficiency of the tRNATyr:Psi 35-synthase aPus7 in Archaea of the Sulfolobales order might be rescued by the H/ACA sRNA-guided machinery." Nucleic Acids Res 37(4): 1308-22.

Mushegian, A. R. et E. V. Koonin (1996). "A minimal gene set for cellular life derived by comparison of complete bacterial genomes." Proc Natl Acad Sci U S A 93(19): 10268-73.

Myslyuk, I., T. Doniger, Y. Horesh, A. Hury, R. Hoffer, Y. Ziporen, S. Michaeli et R. Unger (2008). "Psiscan: a computational approach to identify H/ACA-like and AGA-like non-coding RNA in trypanosomatid genomes." BMC Bioinformatics 9: 471.

Nazar, R. N., T. O. Sitz et H. Busch (1976). "Sequence homologies in mammalian 5.8S ribosomal RNA." Biochemistry 15(3): 505-8.

Newby, M. I. et N. L. Greenbaum (2001). "A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture." RNA 7(6): 833-45.

Newby, M. I. et N. L. Greenbaum (2002a). "Investigation of Overhauser effects between pseudouridine and water protons in RNA helices." Proc Natl Acad Sci U S A 99(20): 12697-702.

Newby, M. I. et N. L. Greenbaum (2002b). "Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine." Nat Struct Biol 9(12): 958-65.

Ni, J., A. L. Tien et M. J. Fournier (1997). "Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA." Cell 89(4): 565-73.

Niewmierzycka, A. et S. Clarke (1999). "S-Adenosylmethionine-dependent methylation in Saccharomyces cerevisiae. Identification of a novel protein arginine methyltransferase." J Biol Chem 274(2): 814-24.

Niu, L., B. G. Lane et J. Ofengand (1999). "Cloning and characterization of the 23S RNA pseudouridine 2633 synthase from Bacillus subtilis." Biochemistry 38(2): 629-35.

Noon, K. R., E. Bruenger et J. A. McCloskey (1998). "Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile Sulfolobus solfataricus." J Bacteriol 180(11): 2883-8.

Normand, C., R. Capeyrou, S. Quevillon-Cheruel, A. Mougin, Y. Henry et M. Caizergues-Ferrer (2006). "Analysis of the binding of the N-terminal conserved domain of yeast Cbf5p to a box H/ACA snoRNA." RNA 12(10): 1868-82.

Nottrott, S., K. Hartmuth, P. Fabrizio, H. Urlaub, I. Vidovic, R. Ficner et R. Luhrmann (1999). "Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA." Embo J 18(21): 6119-33.

Nurse, K., J. Wrzesinski, A. Bakin, B. G. Lane et J. Ofengand (1995). "Purification, cloning, and properties of the tRNA psi 55 synthase from Escherichia coli." RNA 1(1): 102-12.

Ochs, R. L., M. A. Lischwe, W. H. Spohn et H. Busch (1985). "Fibrillarin: a new protein of the nucleolus identified by autoimmune sera." Biol Cell 54(2): 123-33.

Ofengand, J. (2002). "Ribosomal RNA pseudouridines and pseudouridine synthases." FEBS Lett 514(1): 17-25.

Ofengand, J. et A. Bakin (1997). "Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts." J Mol Biol 266(2): 246-68.

Ofengand, J., A. Malhotra, J. Remme, N. S. Gutgsell, M. Del Campo, S. Jean-Charles, L. Peil et Y. Kaya (2001). "Pseudouridines and pseudouridine synthases of the ribosome." Cold Spring Harb Symp Quant Biol 66: 147-59.

Omer, A. D., T. M. Lowe, A. G. Russell, H. Ebhardt, S. R. Eddy et P. P. Dennis (2000). "Homologs of small nucleolar RNAs in Archaea." Science 288(5465): 517-22.

Omer, A. D., S. Ziesche, W. A. Decatur, M. J. Fournier et P. P. Dennis (2003). "RNA-modifying machines in archaea." Mol Microbiol 48(3): 617-29.

Omer, A. D., S. Ziesche, H. Ebhardt et P. P. Dennis (2002). "In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex." Proc Natl Acad Sci U S A 99(8): 5289-94.

Palm, W. et T. de Lange (2008). "How shelterin protects mammalian telomeres." Annu Rev Genet 42: 301-34.

Pan, H., S. Agarwalla, D. T. Moustakas, J. Finer-Moore et R. M. Stroud (2003). "Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit." Proc Natl Acad Sci U S A 100(22): 12648-53.

Pan, H., J. D. Ho, R. M. Stroud et J. Finer-Moore (2007). "The crystal structure of E. coli rRNA pseudouridine synthase RluE." J Mol Biol 367(5): 1459-70.

Pellizzoni, L., J. Baccon, B. Charroux et G. Dreyfuss (2001). "The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1." Curr Biol 11(14): 1079-88.

Peng, W. T., M. D. Robinson, S. Mnaimneh, N. J. Krogan, G. Cagney, Q. Morris, A. P. Davierwala, J. Grigull, X. Yang, W. Zhang, N. Mitsakakis, O. W. Ryan, N. Datta, V. Jojic, C. Pal, V. Canadien, D. Richards, B. Beattie, L. F. Wu, S. J. Altschuler, S. Roweis, B. J. Frey, A. Emili, J. F. Greenblatt et T. R. Hughes (2003). "A panoramic view of yeast noncoding RNA processing." Cell 113(7): 919-33.

Perez-Arellano, I., J. Gallego et J. Cervera (2007). "The PUA domain - a structural and functional overview." FEBS J 274(19): 4972-84.

Phannachet, K., Y. Elias et R. H. Huang (2005). "Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase." Biochemistry 44(47): 15488-94.

Phannachet, K. et R. H. Huang (2004). "Conformational change of pseudouridine 55 synthase upon its association with RNA substrate." Nucleic Acids Res 32(4): 1422-9.

Qiao, F. et T. R. Cech (2008). "Triple-helix structure in telomerase RNA contributes to catalysis." Nat Struct Mol Biol 15(6): 634-40.

Qu, L. H., A. Henras, Y. J. Lu, H. Zhou, W. X. Zhou, Y. Q. Zhu, J. Zhao, Y. Henry, M. Caizergues-Ferrer et J. P. Bachellerie (1999). "Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast." Mol Cell Biol 19(2): 1144-58.

Ramamurthy, V., S. L. Swann, J. L. Paulson, C. J. Spedaliere et E. G. Mueller (1999a). "Critical aspartic acid residues in pseudouridine synthases." J Biol Chem 274(32): 22225-30.

Ramamurthy, V., S. L. Swann, C. J. Spedaliere et E. G. Mueller (1999b). "Role of cysteine residues in pseudouridine synthases of different families." Biochemistry 38(40): 13106-11.

Rashid, R., M. Aittaleb, Q. Chen, K. Spiegel, B. Demeler et H. Li (2003). "Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles." J Mol Biol 333(2): 295-306.

Rashid, R., B. Liang, D. L. Baker, O. A. Youssef, Y. He, K. Phipps, R. M. Terns, M. P. Terns et H. Li (2006). "Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita." Mol Cell 21(2): 249-60.

Raychaudhuri, S., J. Conrad, B. G. Hall et J. Ofengand (1998). "A pseudouridine synthase required for the formation of two universally conserved pseudouridines in ribosomal RNA is essential for normal growth of Escherichia coli." RNA 4(11): 1407-17.

Reichow, S. L., T. Hamma, A. R. Ferre-D'Amare et G. Varani (2007). "The structure and function of small nucleolar ribonucleoproteins." Nucleic Acids Res 35(5): 1452-64.

Reichow, S. L. et G. Varani (2008). "Nop10 is a conserved H/ACA snoRNP molecular adaptor." Biochemistry 47(23): 6148-56.

Richard, P., X. Darzacq, E. Bertrand, B. E. Jady, C. Verheggen et T. Kiss (2003). "A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs." EMBO J 22(16): 4283-93.

Richard, P., A. M. Kiss, X. Darzacq et T. Kiss (2006). "Cotranscriptional recognition of human intronic box H/ACA snoRNAs occurs in a splicing-independent manner." Mol Cell Biol 26(7): 2540-9.

Richard, P. et T. Kiss (2006). "Integrating snoRNP assembly with mRNA biogenesis." EMBO Rep 7(6): 590-2.

Rivera, M. C., R. Jain, J. E. Moore et J. A. Lake (1998). "Genomic evidence for two functionally distinct gene classes." Proc Natl Acad Sci U S A 95(11): 6239-44.

Robertson, M. P. et W. G. Scott (2007). "The structural basis of ribozyme-catalyzed RNA assembly." Science 315(5818): 1549-53.

Roovers, M., C. Hale, C. Tricot, M. P. Terns, R. M. Terns, H. Grosjean et L. Droogmans (2006). "Formation of the conserved pseudouridine at position 55 in archaeal tRNA." Nucleic Acids Res 34(15): 4293-301.

Rozenski, J., P. F. Crain et J. A. McCloskey (1999). "The RNA Modification Database: 1999 update." Nucleic Acids Res 27(1): 196-7.

Rozhdestvensky, T. S., T. H. Tang, I. V. Tchirkova, J. Brosius, J. P. Bachellerie et A. Huttenhofer (2003). "Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea." Nucleic Acids Res 31(3): 869-77.

Rubin, G. M. (1973). "The nucleotide sequence of Saccharomyces cerevisiae 5.8 S ribosomal ribonucleic acid." J Biol Chem 248(11): 3860-75.

Rufer, N. et M. Nabholz (2003). "[Telomerase, elixir of life for human cells?]." Med Sci (Paris) 19(3): 345-50.

Ruggero, D., S. Grisendi, F. Piazza, E. Rego, F. Mari, P. H. Rao, C. Cordon-Cardo et P. P. Pandolfi (2003). "Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification." Science 299(5604): 259-62.

Russell, A. G., M. N. Schnare et M. W. Gray (2004). "Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in Euglena gracilis." RNA 10(7): 1034-46.

Sabina, J. et D. Soll (2006). "The RNA-binding PUA domain of archaeal tRNA-guanine transglycosylase is not required for archaeosine formation." J Biol Chem 281(11): 6993-7001.

Samuelsson, T. (1991). "Interactions of transfer RNA pseudouridine synthases with RNAs substituted with fluorouracil." Nucleic Acids Res 19(22): 6139-44.

Savage, S. A., N. Giri, G. M. Baerlocher, N. Orr, P. M. Lansdorp et B. P. Alter (2008). "TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita." Am J Hum Genet 82(2): 501-9.

Schattner, P., S. Barberan-Soler et T. M. Lowe (2006). "A computational screen for mammalian pseudouridylation guide H/ACA RNAs." RNA 12(1): 15-25.

Schimmang, T., D. Tollervey, H. Kern, R. Frank et E. C. Hurt (1989). "A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability." EMBO J 8(13): 4015-24.

Schluenzen, F., A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi et A. Yonath (2000). "Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution." Cell 102(5): 615-23.

Schultz, A., S. Nottrott, N. J. Watkins et R. Luhrmann (2006). "Protein-protein and protein-RNA contacts both contribute to the 15.5K-mediated assembly of the U4/U6 snRNP and the box C/D snoRNPs." Mol Cell Biol 26(13): 5146-54.

Schultz, D. W. et M. Yarus (1994a). "tRNA structure and ribosomal function. I. tRNA nucleotide 27-43 mutations enhance first position wobble." J Mol Biol 235(5): 1381-94.

Schultz, D. W. et M. Yarus (1994b). "tRNA structure and ribosomal function. II. Interaction between anticodon helix and other tRNA mutations." J Mol Biol 235(5): 1395-405.

Senger, B., S. Auxilien, U. Englisch, F. Cramer et F. Fasiolo (1997). "The modified wobble base inosine in yeast tRNAIle is a positive determinant for aminoacylation by isoleucyl-tRNA synthetase." Biochemistry 36(27): 8269-75.

Shay, J. W. et W. E. Wright (2004). "Telomeres in dyskeratosis congenita." Nat Genet 36(5): 437-8.

Shi, H. et P. B. Moore (2000). "The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited." RNA 6(8): 1091-105.

Simos, G., H. Tekotte, H. Grosjean, A. Segref, K. Sharma, D. Tollervey et E. C. Hurt (1996). "Nuclear pore proteins are involved in the biogenesis of functional tRNA." EMBO J 15(9): 2270-84.

Singh, M., F. A. Gonzales, D. Cascio, N. Heckmann, G. Chanfreau et J. Feigon (2009). "Structure and Functional Studies of the CS Domain of the Essential H/ACA Ribonucleoparticle Assembly Protein SHQ1." J Biol Chem 284(3): 1906-16.

Singh, S. K., P. Gurha et R. Gupta (2008). "Dynamic guide-target interactions contribute to sequential 2'-O-methylation by a unique archaeal dual guide box C/D sRNP." RNA 14(7): 1411-23.

Sirinavin, C. et A. A. Trowbridge (1975). "Dyskeratosis congenita: clinical features and genetic aspects. Report of a family and review of the literature." J Med Genet 12(4): 339-54.

Sivaraman, J., P. Iannuzzi, M. Cygler et A. Matte (2004). "Crystal structure of the RluD pseudouridine synthase catalytic module, an enzyme that modifies 23S rRNA and is essential for normal cell growth of Escherichia coli." J Mol Biol 335(1): 87-101.

Sivaraman, J., V. Sauve, R. Larocque, E. A. Stura, J. D. Schrag, M. Cygler et A. Matte (2002). "Structure of the 16S rRNA pseudouridine synthase RsuA bound to uracil and UMP." Nat Struct Biol 9(5): 353-8.

Slany, R. K., M. Bosl, P. F. Crain et H. Kersten (1993). "A new function of S-adenosylmethionine: the ribosyl moiety of AdoMet is the precursor of the cyclopentenediol moiety of the tRNA wobble base queuine." Biochemistry 32(30): 7811-7.

Speckmann, W. A., Z. H. Li, T. M. Lowe, S. R. Eddy, R. M. Terns et M. P. Terns (2002). "Archaeal guide RNAs function in rRNA modification in the eukaryotic nucleus." Curr Biol 12(3): 199-203.

Spedaliere, C. J., J. M. Ginter, M. V. Johnston et E. G. Mueller (2004). "The pseudouridine synthases: revisiting a mechanism that seemed settled." J Am Chem Soc 126(40): 12758-9.

Spedaliere, C. J., C. S. Hamilton et E. G. Mueller (2000). "Functional importance of motif I of pseudouridine synthases: mutagenesis of aligned lysine and proline residues." Biochemistry 39(31): 9459-65.

Spedaliere, C. J. et E. G. Mueller (2004). "Not all pseudouridine synthases are potently inhibited by RNA containing 5-fluorouridine." RNA 10(2): 192-9.

Sprinzl, M. et F. Cramer (1979). "The -C-C-A end of tRNA and its role in protein biosynthesis." Prog Nucleic Acid Res Mol Biol 22: 1-69.

Starostina, N. G., S. Marshburn, L. S. Johnson, S. R. Eddy, R. M. Terns et M. P. Terns (2004). "Circular box C/D RNAs in Pyrococcus furiosus." Proc Natl Acad Sci U S A 101(39): 14097-101.

Stetter, K. O. (2006). "History of discovery of the first hyperthermophiles." Extremophiles 10(5): 357-62.

Stevens, S. W. et J. Abelson (1999). "Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins." Proc Natl Acad Sci U S A 96(13): 7226-31.

Sunita, S., H. Zhenxing, J. Swaathi, M. Cygler, A. Matte et J. Sivaraman (2006). "Domain organization and crystal structure of the catalytic domain of E.coli RluF, a pseudouridine synthase that acts on 23S rRNA." J Mol Biol 359(4): 998-1009.

Suzuki, T., T. Wada, K. Saigo et K. Watanabe (2002). "Taurine as a constituent of mitochondrial tRNAs: new insights into the functions of taurine and human mitochondrial diseases." EMBO J 21(23): 6581-9.

Szewczak, L. B., S. J. DeGregorio, S. A. Strobel et J. A. Steitz (2002). "Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP." Chem Biol 9(10): 1095-107.

Szweykowska-Kulinska, Z., B. Senger, G. Keith, F. Fasiolo et H. Grosjean (1994). "Intron-dependent formation of pseudouridines in the anticodon of Saccharomyces cerevisiae minor tRNA(Ile)." EMBO J 13(19): 4636-44.

Tang, T. H., J. P. Bachellerie, T. Rozhdestvensky, M. L. Bortolin, H. Huber, M. Drungowski, T. Elge, J. Brosius et A. Huttenhofer (2002a). "Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus." Proc Natl Acad Sci U S A 99(11): 7536-41.

Tang, T. H., T. S. Rozhdestvensky, B. C. d'Orval, M. L. Bortolin, H. Huber, B. Charpentier, C. Branlant, J. P. Bachellerie, J. Brosius et A. Huttenhofer (2002b). "RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing." Nucleic Acids Res 30(4): 921-30.

Theimer, C. A., B. E. Jady, N. Chim, P. Richard, K. E. Breece, T. Kiss et J. Feigon (2007). "Structural and functional characterization of human telomerase RNA processing and cajal body localization signals." Mol Cell 27(6): 869-81.

Thompson, L. D. et C. J. Daniels (1990). "Recognition of exon-intron boundaries by the Halobacterium volcanii tRNA intron endonuclease." J Biol Chem 265(30): 18104-11.

Tollervey, D. (1987). "A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA." EMBO J 6(13): 4169-75.

Tollervey, D., H. Lehtonen, M. Carmo-Fonseca et E. C. Hurt (1991). "The small nucleolar RNP protein NOP1 (fibrillarin) is required for pre-rRNA processing in yeast." EMBO J 10(3): 573-83.

Tollervey, D., H. Lehtonen, R. Jansen, H. Kern et E. C. Hurt (1993). "Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly." Cell 72(3): 443-57.

Tomita, K., R. Ishitani, S. Fukai et O. Nureki (2006). "Complete crystallographic analysis of the dynamics of CCA sequence addition." Nature 443(7114): 956-60.

Tomlinson, R. L., E. B. Abreu, T. Ziegler, H. Ly, C. M. Counter, R. M. Terns et M. P. Terns (2008). "Telomerase reverse transcriptase is required for the localization of telomerase RNA to cajal bodies and telomeres in human cancer cells." Mol Biol Cell 19(9): 3793-800.

Tomlinson, R. L., T. D. Ziegler, T. Supakorndej, R. M. Terns et M. P. Terns (2006). "Cell cycle-regulated trafficking of human telomerase to telomeres." Mol Biol Cell 17(2): 955-65.

Trahan, C. et F. Dragon (2009). "Dyskeratosis congenita mutations in the H/ACA domain of human telomerase RNA affect its assembly into a pre-RNP." RNA 15(2): 235-43.

Tran, E., J. Brown et E. S. Maxwell (2004). "Evolutionary origins of the RNA-guided nucleotide-modification complexes: from the primitive translation apparatus?" Trends Biochem Sci 29(7): 343-50.

Tran, E., X. Zhang, L. Lackey et E. S. Maxwell (2005). "Conserved spacing between the box C/D and C'/D' RNPs of the archaeal box C/D sRNP complex is required for efficient 2'-O-methylation of target RNAs." RNA 11(3): 285-93.

Tran, E. J., X. Zhang et E. S. Maxwell (2003). "Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs." Embo J 22(15): 3930-40.

Tyc, K. et J. A. Steitz (1989). "U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus." EMBO J 8(10): 3113-9.

Tycowski, K. T., M. D. Shu, A. Kukoyi et J. A. Steitz (2009). "A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles." Mol Cell 34(1): 47-57.

Urbonavicius, J., Q. Qian, J. M. Durand, T. G. Hagervall et G. R. Bjork (2001). "Improvement of reading frame maintenance is a common function for several tRNA modifications." EMBO J 20(17): 4863-73.

Urbonavicius, J., G. Stahl, J. M. Durand, S. N. Ben Salem, Q. Qian, P. J. Farabaugh et G. R. Bjork (2003). "Transfer RNA modifications that alter +1 frameshifting in general fail to affect -1 frameshifting." RNA 9(6): 760-8.

van de Vossenberg, J. L., A. J. Driessen et W. N. Konings (1998). "The essence of being extremophilic: the role of the unique archaeal membrane lipids." Extremophiles 2(3): 163-70.

Van Lanen, S. G., J. S. Reader, M. A. Swairjo, V. de Crecy-Lagard, B. Lee et D. Iwata-Reuyl (2005). "From cyclohydrolase to oxidoreductase: discovery of nitrile reductase activity in a common fold." Proc Natl Acad Sci U S A 102(12): 4264-9.

Veldman, G. M., J. Klootwijk, V. C. de Regt, R. J. Planta, C. Branlant, A. Krol et J. P. Ebel (1981). "The primary and secondary structure of yeast 26S rRNA." Nucleic Acids Res 9(24): 6935-52.

Venema, J. et D. Tollervey (1999). "Ribosome synthesis in Saccharomyces cerevisiae." Annu Rev Genet 33: 261-311.

Venteicher, A. S., E. B. Abreu, Z. Meng, K. E. McCann, R. M. Terns, T. D. Veenstra, M. P. Terns et S. E. Artandi (2009). "A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis." Science 323(5914): 644-8.

Venteicher, A. S., Z. Meng, P. J. Mason, T. D. Veenstra et S. E. Artandi (2008). "Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly." Cell 132(6): 945-57.

Verheggen, C., D. L. Lafontaine, D. Samarsky, J. Mouaikel, J. M. Blanchard, R. Bordonne et E. Bertrand (2002). "Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments." EMBO J 21(11): 2736-45.

Verheggen, C., J. Mouaikel, M. Thiry, J. M. Blanchard, D. Tollervey, R. Bordonne, D. L. Lafontaine et E. Bertrand (2001). "Box C/D small nucleolar RNA trafficking involves small nucleolar RNP proteins, nucleolar factors and a novel nuclear domain." EMBO J 20(19): 5480-90.

Vermeesch, J. R. et C. M. Price (1994). "Telomeric DNA sequence and structure following de novo telomere synthesis in Euplotes crassus." Mol Cell Biol 14(1): 554-66.

Vidovic, I., S. Nottrott, K. Hartmuth, R. Luhrmann et R. Ficner (2000). "Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment." Mol Cell 6(6): 1331-42.

Vulliamy, T., R. Beswick, M. Kirwan, A. Marrone, M. Digweed, A. Walne et I. Dokal (2008). "Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita." Proc Natl Acad Sci U S A 105(23): 8073-8.

Vulliamy, T., A. Marrone, I. Dokal et P. J. Mason (2002). "Association between aplastic anaemia and mutations in telomerase RNA." Lancet 359(9324): 2168-70.

Vulliamy, T., A. Marrone, F. Goldman, A. Dearlove, M. Bessler, P. J. Mason et I. Dokal (2001). "The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita." Nature 413(6854): 432-5.

Vulliamy, T., A. Marrone, R. Szydlo, A. Walne, P. J. Mason et I. Dokal (2004). "Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC." Nat Genet 36(5): 447-9.

Vulliamy, T. J. et I. Dokal (2008). "Dyskeratosis congenita: the diverse clinical presentation of mutations in the telomerase complex." Biochimie 90(1): 122-30.

Wahl, M. C., C. L. Will et R. Luhrmann (2009). "The spliceosome: design principles of a dynamic RNP machine." Cell 136(4): 701-18.

Walczak, R., E. Westhof, P. Carbon et A. Krol (1996). "A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs." RNA 2(4): 367-79.

Walne, A. J. et I. Dokal (2008). "Dyskeratosis Congenita: a historical perspective." Mech Ageing Dev 129(1-2): 48-59.

Walne, A. J., T. Vulliamy, R. Beswick, M. Kirwan et I. Dokal (2008). "TINF2 mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes." Blood 112(9): 3594-600.

Walne, A. J., T. Vulliamy, A. Marrone, R. Beswick, M. Kirwan, Y. Masunari, F. H. Al-Qurashi, M. Aljurf et I. Dokal (2007). "Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10." Hum Mol Genet 16(13): 1619-29.

Wang, C. et U. T. Meier (2004). "Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins." EMBO J 23(8): 1857-67.

Wang, C., C. C. Query et U. T. Meier (2002). "Immunopurified small nucleolar ribonucleoprotein particles pseudouridylate rRNA independently of their association with phosphorylated Nopp140." Mol Cell Biol 22(24): 8457-66.

Wang, H., D. Boisvert, K. K. Kim, R. Kim et S. H. Kim (2000). "Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6 A resolution." EMBO J 19(3): 317-23.

Watanabe, Y. et M. W. Gray (2000). "Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in Euglena gracilis, an early diverging

eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria." Nucleic Acids Res 28(12): 2342-52.

Watkins, N. J., A. Dickmanns et R. Luhrmann (2002). "Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP." Mol Cell Biol 22(23): 8342-52.

Watkins, N. J., A. Gottschalk, G. Neubauer, B. Kastner, P. Fabrizio, M. Mann et R. Luhrmann (1998). "Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure." Rna 4(12): 1549-68.

Watkins, N. J., R. D. Leverette, L. Xia, M. T. Andrews et E. S. Maxwell (1996). "Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D." RNA 2(2): 118-33.

Watkins, N. J., V. Segault, B. Charpentier, S. Nottrott, P. Fabrizio, A. Bachi, M. Wilm, M. Rosbash, C. Branlant et R. Luhrmann (2000). "A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP." Cell 103(3): 457-66.

White, M. F. et S. D. Bell (2002). "Holding it together: chromatin in the Archaea." Trends Genet 18(12): 621-6.

Whitehead, S. E., K. W. Jones, X. Zhang, X. Cheng, R. M. Terns et M. P. Terns (2002). "Determinants of the interaction of the spinal muscular atrophy disease protein SMN with the dimethylarginine-modified box H/ACA small nucleolar ribonucleoprotein GAR1." J Biol Chem 277(50): 48087-93.

Wong, J. M. et K. Collins (2006). "Telomerase RNA level limits telomere maintenance in X-linked dyskeratosis congenita." Genes Dev 20(20): 2848-58.

Wrzesinski, J., A. Bakin, K. Nurse, B. G. Lane et J. Ofengand (1995a). "Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from Escherichia coli." Biochemistry 34(27): 8904-13.

Wrzesinski, J., K. Nurse, A. Bakin, B. G. Lane et J. Ofengand (1995b). "A dual-specificity pseudouridine synthase: an Escherichia coli synthase purified and cloned on the basis of its specificity for psi 746 in 23S RNA is also specific for psi 32 in tRNA(phe)." RNA 1(4): 437-48.

Wu, H. et J. Feigon (2007). "H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification." Proc Natl Acad Sci U S A 104(16): 6655-60.

Xiao, M., C. Yang, P. Schattner et Y. T. Yu (2009). "Functionality and substrate specificity of human box H/ACA guide RNAs." RNA 15(1): 176-86.

Yaghmai, R., A. Kimyai-Asadi, K. Rostamiani, N. S. Heiss, A. Poustka, W. Eyaid, J. Bodurtha, H. C. Nousari, A. Hamosh et A. Metzenberg (2000). "Overlap of dyskeratosis congenita with the Hoyeraal-Hreidarsson syndrome." J Pediatr 136(3): 390-3.

Yamaguchi, H., R. T. Calado, H. Ly, S. Kajigaya, G. M. Baerlocher, S. J. Chanock, P. M. Lansdorp et N. S. Young (2005). "Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia." N Engl J Med 352(14): 1413-24.

Yang, P. K., C. Hoareau, C. Froment, B. Monsarrat, Y. Henry et G. Chanfreau (2005). "Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly." Mol Cell Biol 25(8): 3295-304.

Yang, P. K., G. Rotondo, T. Porras, P. Legrain et G. Chanfreau (2002). "The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis." J Biol Chem 277(47): 45235-42.

Yang, Y., Y. Chen, C. Zhang, H. Huang et S. M. Weissman (2002). "Nucleolar localization of hTERT protein is associated with telomerase function." Exp Cell Res 277(2): 201-9.

Yasukawa, T., T. Suzuki, N. Ishii, T. Ueda, S. Ohta et K. Watanabe (2000a). "Defect in modification at the anticodon wobble nucleotide of mitochondrial tRNA(Lys) with the MERRF encephalomyopathy pathogenic mutation." FEBS Lett 467(2-3): 175-8.

Yasukawa, T., T. Suzuki, T. Ueda, S. Ohta et K. Watanabe (2000b). "Modification defect at anticodon wobble nucleotide of mitochondrial tRNAs(Leu)(UUR) with pathogenic mutations of mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes." J Biol Chem 275(6): 4251-7.

Ye, K. (2007). "H/ACA guide RNAs, proteins and complexes." Curr Opin Struct Biol 17(3): 287-92.

Yoon, A., G. Peng, Y. Brandenburger, O. Zollo, W. Xu, E. Rego et D. Ruggero (2006). "Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita." Science 312(5775): 902-6.

Youssef, O. A., R. M. Terns et M. P. Terns (2007). "Dynamic interactions within sub-complexes of the H/ACA pseudouridylation guide RNP." Nucleic Acids Res 35(18): 6196-206.

Yusupov, M. M., G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate et H. F. Noller (2001). "Crystal structure of the ribosome at 5.5 A resolution." Science 292(5518): 883-96.

Zaug, A. J. et T. R. Cech (1986). "The intervening sequence RNA of Tetrahymena is an enzyme." Science 231(4737): 470-5.

Zebarjadian, Y., T. King, M. J. Fournier, L. Clarke et J. Carbon (1999). "Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA." Mol Cell Biol 19(11): 7461-72.

Zhang, X., E. A. Champion, E. J. Tran, B. A. Brown, 2nd, S. J. Baserga et E. S. Maxwell (2006). "The coiled-coil domain of the Nop56/58 core protein is dispensable for sRNP assembly but is critical for archaeal box C/D sRNP-guided nucleotide methylation." RNA 12(6): 1092-103.

Zhao, R., M. Davey, Y. C. Hsu, P. Kaplanek, A. Tong, A. B. Parsons, N. Krogan, G. Cagney, D. Mai, J. Greenblatt, C. Boone, A. Emili et W. A. Houry (2005). "Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone." Cell 120(5): 715-27.

Zhao, R., Y. Kakihara, A. Gribun, J. Huen, G. Yang, M. Khanna, M. Costanzo, R. L. Brost, C. Boone, T. R. Hughes, C. M. Yip et W. A. Houry (2008). "Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation." J Cell Biol 180(3): 563-78.

Zhao, X. et Y. T. Yu (2004a). "Detection and quantitation of RNA base modifications." RNA 10(6): 996-1002.

Zhao, X. et Y. T. Yu (2004b). "Pseudouridines in and near the branch site recognition region of U2 snRNA are required for snRNP biogenesis and pre-mRNA splicing in Xenopus oocytes." RNA 10(4): 681-90.

Zhao, X. et Y. T. Yu (2007). "Incorporation of 5-fluorouracil into U2 snRNA blocks pseudouridylation and pre-mRNA splicing in vivo." Nucleic Acids Res 35(2): 550-8.

Zhu, J., K. Gopinath, A. Murali, G. Yi, S. D. Hayward, H. Zhu et C. Kao (2007). "RNA-binding proteins that inhibit RNA virus infection." Proc Natl Acad Sci U S A 104(9): 3129-34.

Résumé

L'isomère de l'uridine (U), la pseudouridine (Ψ), est le nucléoside le plus fréquemment rencontré dans les ARN. La réaction de pseudouridylation dans les ARN est catalysée par des ARN:Ψ-synthases qui, soit fonctionnent comme des enzymes protéiques, soit sont portées chez les eucaryotes et les archaea par des particules ribonucléoprotéiques (RNP). Chaque RNP est composée d'un RNA guide dit à boîtes H/ACA (sno/sca/sRNA chez les eucaryotes et sRNA chez les archaea) et d'un ensemble invariable de protéines : l'enzyme ARN:Ψ-synthase Dyskérine/aCBF5 et les 3 protéines NOP10/aNOP10, GAR1/aGAR1 et NHP2/L7Ae. L'ARN guide assure la reconnaissance de l'ARN cible à modifier par appariement de bases. Les sRNA H/ACA identifiés chez l'archae *Pyrococcus abyssi* ont servi de modèles pour des études de structure-fonction basées sur i) la mesure de l'activité de sRNP H/ACA reconstituées avec les composants protéiques et nucléiques purifiés, ii) l'analyse de la structure 2D des ARN au sein des RNP par des sondes chimiques et enzymatiques et par dichroïsme circulaire, iii) les diverses structures 3D disponibles. Les déterminants moléculaires présents sur les ARN guides ont été précisés, ainsi que le rôle fonctionnel des différentes protéines, de leurs domaines structuraux et de résidus conservés. Nous montrons que l'association entre aNOP10 et L7Ae est cruciale pour l'activité des RNP. Nous avons identifié que aCBF5 pouvait catalyser la formation des résidus Ψ55 dans les ARNt et Ψ2603 dans l'ARN 23S sans l'intervention de sRNA. Nous avons étudié le rôle de résidus conservés de aNOP10 ainsi que du site actif et de la boucle β7/β10 de aCBF5 dans les activités guidée et non guidée de aCBF5.

Mots-Clés : archaea, *Pyrococcus abyssi*, sRNP H/ACA, ARN guide, ARN:Ψ-synthase, réaction de pseudouridylation, protéines aCBF5/L7Ae/aNOP10/aGAR1.

Title: Structural and functional analysis of box H/ACA sRNP particles, which catalyze the targeted formation of pseudouridines in archaeal RNAs.

Summary

Pseudouridine (Ψ), uridine's isomer, is the most abundant modified nucleoside found in structured RNAs. The reaction of pseudouridylation is either catalyzed by a limited number of standalone RNA:Ψ-synthases, or in eukaryotes and archaea by multiple ribonucleoprotein particles (the so-called, box H/ACA sno/sca/sRNPs). Each RNP is composed of a specific box H/ACA RNA used as a guide to define the U residue for modification, and a common set of four proteins– the RNA:Ψ-synthase Dyskerin/aCBF5, and the auxiliary proteins NOP10/aNOP10, GAR1/aGAR1, NHP2/L7Ae (respectively in human and archaea). Initially, 7 guide RNAs were identified in the archaeon *P. abyssi*, which were used as models for structure-function analyses. Activities of RNPs reconstituted with purified protein and RNA components were measured; RNA 2D structure within RNPs was investigated by chemical and enzymatic probing assays as well as by circular dichroism. By taking into consideration the various 3D structures recently resolved, we were able to pinpoint the RNA molecular determinants and to clarify the role played by each protein, their domains, and some of their conserved residues. Hence, interaction between aNOP10 and L7Ae was found to be crucial for RNP activity. Moreover, we show that the enzyme aCBF5 catalyzes pseudouridylation at the position 55 in tRNAs and the position 2603 in 23S rRNA without the use of any guide sRNA. The role of residues conserved in aNOP10, and in the active site and the β7/β10 loop of aCBF5 for the RNA and non-RNA guided activities were analyzed.

Key Words: archaea, *Pyrococcus abyssi*, H/ACA sRNP, guide RNA, RNA:Ψ-synthase, reaction of pseudouridylation, proteins aCBF5/L7Ae/aNOP10/aGAR1.