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

FACULTE DES SCIENCES & TECHNIQUES

U.F.R. Sciences et Techniques Biologiques
Ecole Doctorale Biologie, Santé, Environnement

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I

en Biologie Moléculaire

par **Alan URBAN**

Caractérisation biochimique, structurale et fonctionnelle des ARN:pseudouridine synthases Pus7 de la levure *Saccharomyces cerevisiae* et d'archaea thermophiles

Soutenance publique le 15 Septembre 2008

Membres du Jury :

Rapporteurs :

Mme B. GOLINELLI-PIMPANEAU	Chargé de Recherche CNRS, Gif-sur-Yvette
Mme C. FLORENTZ	Professeur, Université Louis Pasteur, Strasbourg

Examinateur :

M. M. HELM	Docteur, Université Heidelberg, Heidelberg

Directeurs de thèse :

Mme C. BRANLANT	Directeur de Recherche CNRS, Nancy
M. I. MOTORINE	Professeur, Université Henri Poincaré, Nancy

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

Nancy-Université
Université Henri Poincaré

FACULTE DES SCIENCES & TECHNIQUES

U.F.R. Sciences et Techniques Biologiques
Ecole Doctorale Biologie, Santé, Environnement

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I

en Biologie Moléculaire

par **Alan URBAN**

Caractérisation biochimique, structurale et fonctionnelle des ARN:pseudouridine synthases Pus7 de la levure *Saccharomyces cerevisiae* et d'archaea thermophiles

Soutenance publique le 15 Septembre 2008

Membres du Jury :

Rapporteurs :

Mme B. GOLINELLI-PIMPANEAU	Chargé de Recherche CNRS, Gif-sur-Yvette
Mme C. FLORENTZ	Professeur, Université Louis Pasteur, Strasbourg

Examinateur :

M. M. HELM	Docteur, Université Heidelberg, Heidelberg

Directeurs de thèse :

Mme C. BRANLANT	Directeur de Recherche CNRS, Nancy
M. I. MOTORINE	Professeur, Université Henri Poincaré, Nancy

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

Remerciements

Je tiens tout d'abord à remercier chaleureusement Madame Christiane Branlant, Directrice de l'UMR CNRS-UHP 7567, pour m'avoir accueilli dans son laboratoire et avoir accepté de co-dirigé mon travail de thèse. Je lui exprime toute ma gratitude pour m'avoir ouvert une porte dans le vaste monde de l'ARN que j'ai pu explorer sous différents angles et dans les meilleures conditions, en utilisant l'importante panoplie d'outils et de compétences disponibles au laboratoire. Je la remercie pour m'avoir toujours aidé à trouver les meilleures pistes de recherche au cours de ce travail et m'avoir guidé sans relâche lors de la rédaction de ce manuscrit et des articles qu'il contient.

Je voudrais également remercier Monsieur Iouri Motorine qui a co-dirigé ce travail de thèse. Un immense merci pour m'avoir, avec I Behm-Ansmant, initié au travail de recherche dès mes débuts au laboratoire et pour m'avoir ensuite soutenu tout au long de ce projet. Je pense aujourd'hui que sa rigueur, ses conseils avisés et ses connaissances de la thématique m'ont permis de progresser jour après jour. J'ai beaucoup appris à ses côtés tant d'un point de vue scientifique, que personnel.

Je suis très sensible au grand honneur que me font Mesdames Béatrice Golinelli-Pimpaneau et Catherine Florentz, d'accepter d'être rapporteurs de ce mémoire. Je remercie aussi sincèrement Monsieur Mark Helm qui a accepté d'examiner cette thèse.

Je tiens tout particulièrement à remercier Isabelle Behm-Ansmant qui m'a initié à un grand nombre de techniques de biologie moléculaires, avec qui j'ai largement collaboré pour la découverte des activités de Pus7p sur les ARNt et qui m'a beaucoup aidé lors de la rédaction de ce manuscrit. Je remercie aussi très vivement Sébastien Muller, qui avait découvert le sRNA H/ACA d'archaea spécifique des ARNt par approche informatique et avec qui j'ai participé à la démonstration de son activité, et enfin Christophe Charron qui m'a initié à la cristallogenèse et avec qui j'ai débuté l'étude structurale de l'enzyme Pus7 de *P. abyssi*.

D'une manière générale, je voudrais exprimer toute mon amitié à « l'équipe du Box 3 » qui a fortement marqué mes années thèse, en particulier à Jean-Michel, Xavier, Isabelle et Sébastien. J'ai beaucoup apprécié le travail dans cette ambiance stimulante que nous avons réussi à créer au fil du temps : alliant amitié et professionnalisme qui s'est avérée très efficace ! Je voudrais aussi remercier sincèrement les stagiaires qui ont travaillé à mes côtés en particulier, Elise, Benjamin et Anne : je ne sais pas si travailler avec moi vous a donné l'envie de continuer dans la recherche mais je vous souhaite bon courage pour vos thèses respectives. Merci également aux autres personnes que j'ai côtoyé à la paillasse pour leur aide, leurs conseils et leurs idées : Lionnel, François, Claude, Mathieu, Régis et Magali.

Remerciements

Je suis particulièrement reconnaissant envers Christiane Branlant, Iouri Motorine, Bruno Charpentier, I Behm-Ansmant et Xavier Manival qui ont su me communiquer leur enthousiasme pour la recherche, leur savoir-faire dans des domaines aussi divers que la biochimie des protéines, la biologie moléculaire ou la biologie structurale. Vos conseils ont été essentiels à l'avancée de mon travail.

Je remercie tout spécialement le trio de choc sans qui le MUST'In ne serait pas possible : Xavier, Jean-Michel et Mounir, merci pour cette expérience de vie extraprofessionnelle et de m'avoir offert votre confiance. Notre périple a fait évoluer mon regard sur le monde…

Merci également à tous ceux qui ont compris qu'il était important d'expliquer au grand public comment se déroule notre métier et qui ont participé aux projets de vulgarisation scientifique et autres journées portes ouvertes: Nicolas Beck, Henri-Claude Grégoire ex Directeur du CIES de Lorraine, Fabrice N., Nathalie P., Sylvain M., tous les moniteurs du groupe « L'ADN démystifié » et tous le personnel du laboratoire qui s'est investi de près ou de loin dans ces opérations.

Merci aux autres personnes que je n'ai pas encore cité et qui ont également contribué à la bonne ambiance et qui m'ont permis de comprendre que la thèse n'était pas qu'un simple travail, c'est avant tout une tranche de vie qui forge un jeune homme et le prépare à de nombreux aspects de son futur métier. Je pense notamment à François, Alexandre K., Athanase, Lionel, François G, Séverine B., Valérie, Karine, Rémi, Hadidja, Corinne, Delphine, Christophe G., Benoît, Mathias et les autres qui se reconnaîtrons sans doute.

Je souhaiterai également rendre hommage à Jackie qui nous a quitté. Son aide technique a contribué à la réussite de bon nombre de mes expériences mais je pense surtout aux discussions extraprofessionnelles qui ont été une source intarissable de rires et de bonheurs. Encore Merci.

J'ai une pensée particulière pour mes parents qui m'ont offert la chance de faire des études supérieures et qui, avec ma sœur Jessica, m'ont toujours soutenus dans mes projets personnels et professionnels. Maman, c'est grâce à toi que j'ai pris confiance en moi et que j'ai pu progresser chaque jour en étant attentif aux conseils d'autrui. Papa, j'ai suivi ton exemple en apprenant à me battre et à aller jusqu'au bout de mes projets malgré les embûches. Je vous dédie ce travail. Je remercie également les autres membres de ma famille et ma belle-famille (et en particulier Jean-Michel et Agnès) pour m'avoir soutenu et avoir toujours cru en moi. Merci aussi à Aurélie qui a su m'accueillir, me comprendre et m'épauler dans les derniers instants de la rédaction de ce manuscrit.

Aurore, pour tout l'amour que tu m'as donné durant ces années et pour tout ce que tu m'as apporté chaque jour passé ensemble, un simple merci ne serait pas suffisant...

Abréviations

Abréviations & Anglicismes

A : adénine
aa : acide aminé
ADN : acide désoxyribonucléique
ADNc : acide désoxyribonucléique complémentaire
ARN : acide ribonucléique
ARNm : acide ribonucléique messager
ARNnc : ARN non codant
ARN pol : ARN polymérase
ARNr : acide ribonucléique ribosomique
ARNt : acide ribonucléique de transfert
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
C-ter : extrémité C-terminale
Da, KDa : dalton, kilodalton
db : double brin
ddNTP : didésoxyribonucléotide triphosphate
DNase : désoxyribonucléase
dNTP : désoxyribonucléotide triphosphate
DO : densité optique
DTT : dithiothréitol
EDTA : éthylène diamine tétraacétate
5FU : 5-fluoro-uridine
g/ mg/ µg/ ng : gramme/ milligramme/ microgramme/ nanogramme
G : guanine
h : heure
I : inosine
IPTG : isopropyl β-D-thiogalactoside
kan^r : gène de résistance à la kanamycine
Kb : kilobase
l/ ml/ µl : litre/ millilitre/ microlitre
LB : Luria Bertani
M/ mM/ µM/ nM : molaire/ millimolaire/ micromolaire/ nanomolaire
min : minute
NMP : ribonucléotide monophosphate
NTP : ribonucléotide triphosphate
nt : nucléotide
N-ter : extrémité N-terminale
pb : paire de base
PEG : polyéthylène glycol
pGalS : 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
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
SDS	: sodium dodécylsulfate
structure 2D	: structure secondaire
structure 3D	: structure tertiaire
T	: thymine
TLC	: chromatographie sur couche mince
U (UE)	: unité enzymatique
Ψ	: pseudouridine
Y	: pyrimidine

Anglicisme: certains termes ne sont pas ou sont rarement traduits :

ETS	: external transcribed sequence
GST	: glutathione S-transferase
LUCA	: Last Universal Common Ancester ou dernier ancêtre commun universel
LSU	: large subunit
ORF	: phase ouverte de lecture (Open Reading Frame)
PCR	: amplification en chaine par une polymérase (polymerase chain reaction)
SSU	: small subunit
SELEX	: Systematic Evolution of Ligands by EXponential enrichment
gRNA	: ARN guide
sRNA	: petit ARN (Small RNA)
sRNP	: particule ribonucléoprotéique contenant un ou des petits ARN (Small ribonucleoprotein particule)
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)
snoRNA	: petit ARN nucléolaire (Small nucleolar RNA)
snRNA	: petit ARN nucléaire (Small nuclear RNA)
snoRNP	: particule ribonucléoprotéique contenant un ou des petits ARN nucléolaires (Small nucleolar ribonucleoprotein particule)
snRNP	: particule ribonucléoprotéique contenant un ou des petits ARN nucléaires (Small nuclear Ribonucleoprotein Particule)
UsnRNA	: « uridine rich small nuclear ribonucleic acid »
UsnRNA	: « uridine rich small nuclear ribonucleoprotein particle »
spliceosome	: complexe nucléaire ribonucléoprotéique dans lequel ont lieu les réactions d'épissage
UTR	: untranslated region

Résumé

Résumé

Titre : Caractérisation biochimique, structurale et fonctionnelle des ARN:pseudouridine synthases Pus7 de la levure *Saccharomyces cerevisiae* et d'archaea thermophiles

La conversion des résidus U en pseudouridine (Ψ) est la modification la plus fréquente dans les ARN. Cette réaction est catalysée par des ARN:Ψ-synthases qui fonctionnent soit seules, soit au sein de particules snoRNP H/ACA composées de 4 protéines et d'un snoRNA.

Nous avons démontré que l'ARN:Ψ-synthase Pus7p de *S. cerevisiae* est capable d'agir sur les ARNt cytoplasmiques (position 13 dans 10 ARNt et position 35 dans le pré-ARNt^{Tyr} contenant un intron) (Behm-Ansmant et al., 2003). En parallèle, j'ai recherché les requis en séquence et en structure pour qu'un ARN soit substrat de Pus7p. J'ai également réalisé une étude préliminaire de l'importance des domaines additionnels de Pus7p par rapport à l'enzyme TruD d'*E. coli*.

Comme les systèmes enzymatiques responsables de la formation de trois résidus Ψ du snRNA U2 de *S. cerevisiae* avaient été identifiés, nous avons pu débuter une étude de l'importance fonctionnelle de ces résidus pour la réaction d'épissage en utilisant l'approche globale par puces à ADN que l'équipe de J. Beggs (Edinburgh University) avait mise au point pour *S. cerevisiae*. Nous avons ainsi observé une baisse de l'efficacité d'épissage de certains ARN pré-messagers lorsque plusieurs gènes codant des U2 snRNA:Ψ-synthases sont inactivés simultanément.

L'ARN:Ψ-synthase Pus7p est présente dans les 3 domaines du vivant et en particulier chez les archaea où l'activité de cette enzyme n'avait jamais été étudiée. Nous avons découvert l'existence de 2 systèmes enzymatiques différents pour la modification d'une même position d'un ARNt chez 2 espèces d'archaea thermophiles : une enzyme seule chez *Pyrococcus abyssi* contre un système de guide sRNP H/ACA chez *Sulfolobus solfataricus* dans lequel la protéine Pus7 est mutée au niveau d'acides aminés requis pour l'activité. Finalement, nous avons produit chacune de ces enzymes en grandes quantités et nous avons réalisé des tests de cristallisation. Des cristaux de Pus7 de *P. abyssi* ont été obtenus et une mesure complète de données de diffraction de rayons X a été réalisée au laboratoire à 2,5 Å de résolution.

Mots clés : pseudouridine, ARN:Ψ-synthase, Pus7, tRNA, archaea, *Pyrococcus abyssi*.

Table des Matières

Table des matières

Remerciements…….. I
Abréviations…….. III
Résumé….....V
Table des matières……VII

Introduction

1. Les modifications post-transcriptionnelles : cas du résidu pseudouridine 1
2. Localisation et distribution des résidus Ψ dans les ARN 6
A. Cas des ARN de transfert 6
B. Cas des ARN ribosomiques 6
C. Cas des UsnRNA 14
1. Rôles des UsnRNA 14
2. La réaction d'épissage 17
3. Localisation des résidus Ψ dans les UsnRNA 19
3. Rôles des résidus pseudouridines dans les ARN 24
A.Rôles dans les ARN de transfert 24
1. Résidu Ψ en position 13 26
2. Résidu Ψ en positions 27 et 28 27
3. Résidus Ψ en positions 31 et 32 29
4. Résidus Ψ en positions 34 et 36 30
5. Résidu Ψ en position 35 31
6. Résidus Ψ en positions 38 et 39 32
7. Résidu Ψ en position 55 34
B. Rôles dans les ARN ribosomiques 35
C. Rôle dans les UsnRNA 39
1. Interactions UsnRNA/ARN pré-messager 39
2. Interactions UsnRNA/UsnRNA 47
4. Caractérisation, classification, mécanismes enzymatiques et structure 3D des ARN:Ψ-synthases 48
A. Caractérisation des ARN:Ψ-synthases 48
B. Classification des ARN:Ψ-synthases 50
C. Mécanisme enzymatique 52
D. Structure tridimensionnelle des ARN:Ψ-synthases 54
E. La spécificité de reconnaissance des ARN:Ψ-synthases 58
1. ARN:Ψ-synthases ne nécessitant pas de guide ARN 58
1.1. Classification des ARN:Ψ-synthases 58
1.1.1. Classification générale des ARN:Ψ-synthases 58
1.1.2. ARN:Ψ-synthases agissant sur les ARNt et UsnRNA 60
1.1.3. ARN:Ψ-synthases agissant sur les ARNr 63
1.2. Mode de reconnaissance du substrat ARN 65
1.2.1. Site-spécificité de TruB 65

1.2.2. Double spécificité de substrat de RluA 68
1.2.3. Le mécanisme d'extrusion de base 70
1.2.4. Région-spécificité de TruA 72
1.2.5. Rôle des domaines additionnels N et C terminaux 74
2. ARN:Ψ-synthases associées à des guides ARN 76
2.1. Présentation générale 79
2.2. Structure des snoRNA H/ACA 80
2.2.1. Les snoRNA H/ACA eucaryotes 80
2.2.2. Les sRNA H/ACA d'archaea 82
2.3. Caractéristiques des 4 protéines des sno/sRNP à boîtes H/ACA 84
2.3.1. CBF5 (Cbf5p / aCBF5 / NAP57 / Dyskérine) 84
2.3.2. NOP10 (Nop10p / aNOP10) 85
2.3.3. Nhp2p / L7Ae 86
2.3.4. GAR1 (Gar1p / aGAR1) 86
2.4. Structure des sno/sRNP H/ACA 87
2.4.1. Les snoRNP H/ACA eucaryotes 87
2.4.2. Les snRNP H/ACA d'archae 90
2.5. Attribution des sno/sRNA à boîtes H/ACA aux résidus Ψ présent dans les différentes classes d'ARN 95
2.5.1. Situation dans les UsnRNA 95
2.5.2. Situation dans les ARNt 96
2.5.3. Situation dans les ARNr 96
5. Objectifs de la thèse 97

Matériel et Méthodes

1. Bioinformatique 100
A. Matériel informatique 100
B. Logiciels utilisés 100
2. Biologie Moléculaire 100
A. Matériel 100
1. Les souches utilisées 100
1.1. Souches de bactéries 100
1.2. Souches de levure 101
1.3. Souches d'archaea 101
2. Les plasmides utilisés 101
3. ADN génomique de *P. abyssi* et *S. solfataricus* 102
4. Milieux de culture, tampons et solutions d'usage courant 102
B. Méthodes 103
1. Méthodes concernant l'ADN 103
1.1. Extraction d'ADN génomique de la levure *S. cerevisiae* 103
1.2. Réaction de polymérisation en chaîne (PCR) 103
1.3. Clonage de gènes 105

1.3.1. Conditions d'hydrolyse de l'ADN 105
1.3.2. Purification des fragments ADN 105
1.3.3. Conditions de phosphorylation/déphosphorylation 105
1.3.4. Conditions de ligation 105
1.4. Transformation d'*E. coli.* 106
1.5. Analyse des clones recombinants 106
1.5.1. PCR sur colonie 106
1.5.2. Analyse rapide du contenu plasmidique 106
1.6. Préparation d'ADN plasmidique 107
1.7. Séquençage d'ADN 107
1.8. Mutagenèse par PCR 107
1.9. Transformation de *S. cerevisiae.* 109
1.10. Sporulation de levure, dissection des spores et vérification par répliques au velours 110
2. Méthodes concernant les ARN 111
2.1. Extraction d'ARN de *S. cerevisiae* 111
2.2. Extraction d'ARN de bactéries et d'archaea 111
2.3. Analyse des résidus Ψ dans les ARN par la méthode au CMCT 112
2.3.1. Modification des ARN par le CMCT 112
2.3.2. Analyse par extension d'amorce avec la transcriptase inverse 112
2.3.2.1. Marquage des oligodéoxynucléotides 112
2.3.2.2. Elongation par la transcriptase inverse 113
2.4. Production de transcrits *in vitro* 115
2.5. Evaluation de l'activité de pseudouridylation 116
3. Méthodes concernant les protéines 119
3.1. Préparation d'extraits cellulaires 119
3.1.1. Préparation d'extraits cellulaires de levure 119
3.1.2. Préparation d'extraits cellulaires de P. abyssi et S. solfataricus 119
3.2. Production des protéines recombinantes 120
3.3. Purification des protéines recombinantes 120
3.3.1. Sonication 120
3.3.2. Purification des protéines portant l'étiquette « His-6x » : chromatographie d'affinité sur ions métalliques immobilisés (IMAC) 120
3.4. Caractérisation des protéines 121
3.4.1. Analyse par électrophorèse en gel de polyacrylamide en conditions dénaturantes 121
3.4.2. Analyse par Western-Blot 122
3.4.3. Analyse de la formation de complexes ARN/protéine 123
3. Biologie structurale 124
A. Matériel 124
1. Souches bactériennes 124
2. Milieux de culture 124
B. Méthodes 125

1. Production des protéines 125
1.1. Système de production 125
1.2. Système d'expression 126
1.3. Conditions de culture 126
1.3.1. Culture en batch 126

Publication n°1……..127

Optimisation of expression and purification of the recombinant Yol066 (Rib2) protein from *Saccharomyces cerevisiae*.

A. Urban, I. Ansmant and Y. Motorin

J Chromatogr B Analyt Technol Biomed Life Sci., 786(1-2),187-195, 2003

1.3.2. Culture en fermenteur 139
1.3.3. Production de protéines sélénométhionylées (Se-Met) 139
2. Purification des protéines 140
3. Concentration et déssalage des protéines 142
4. Caractérisation des protéines 142
4.1. Dosage des protéines par la méthode de Bradford 142
4.2. Spectrométrie de masse 142
4.2.1. Principe 142
4.2.2. Contrôle qualité par spectrométrie de masse ESI dénaturante 143
5. Cristallographie R-X 143
5.1. Introduction à la cristallographie des rayons X 143
5.2. Cristallogenèse 144
5.2.1. Principe 144
5.2.2. Recherche des conditions de nucléation 145
5.2.3. Optimisation des conditions initiales pour cristalliser des monocristaux pour la collecte de données de diffraction des R-X 145
5.3. Collecte de données de diffraction 145
4. Transcriptomique (Splicing Array) 146
A. Matériel 146
1. Définition et type de puces utilisées 146
2. Milieux de culture, tampons et solutions spécifiques 146
B. Méthodes 147
1. Optimisation des conditions expérimentales 147
2. Condition de culture des différentes souches de levure 147
3. Préparation des cibles et hybridation des puces 148
3.1. Extraction des ARN totaux 148
3.2. Mesure de la concentration et contrôle de la qualité des ARN 149
3.3. Transcription inverse et incorporation des marqueurs fluorescents 150
4. Hybridation et lavage 157

5. Acquisition et analyse des données 155
5.1. Acquisition des données brutes 155
5.1.1. Lecture sur le scanner 155
5.1.2. Extraction des données numériques 156
5.2 Transformation des données 156
5.2.1. Filtrage primaire des données brutes 156
5.2.2. Le calcul des ratios d'expression 157
5.2.3. La normalisation par rapport à la moyenne globale des intensités 157
5.3. Analyse et interprétation des résultats 158

Résultats et Discussion

PARTIE 1
Caractérisation des activités ARNt:Ψ-synthase de l'enzyme Pus7p de *S. cerevisiae*

1. Contexte 160
2. Démarche expérimentale 161
3. Résultats (travail réalisé en collaboration avec I. Behm-Ansmant) 165

Publication n°2……166

The *Saccharomyces cerevisiae* U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Ψ-synthase also acting on tRNAs

Behm-Ansmant I., Urban A., Ma X., Yu YT., Motorin, Y. and Branlant C.
RNA, 9(11), 1371-82, 2003

PARTIE 2
Etude des requis en sequence et structure pour qu'un arn soit substrat de l'enzyme Pus7p de *S. cerevisiae*

1. Contexte 180
2. Démarche et Résultats 180

Publication n°3……182

RNA sequence and 2D structure features required for efficient activity of the *Saccharomyces cerevisiae* RNA:Ψ-synthase Pus7p

Alan Urban, Christiane Branlant and Yuri Motorin

3. Résultats complémentaires 223

PARTIE 3
Etude de l'importance des domaines protéiques qui sont additionnels dans l'ARN:Ψ-synthase Pus7p par rapport à l'enzyme TruD

1. Identification des domaines additionnels de Pus7p par rapport à TruD 228
1.1. Stratégie utilisée 228
1.2. Résultats obtenus par alignements multiples de séquence 229
2. Délétions des domaines additionnels en vue de tester leur importance fonctionnelle 232
2.1. Raffinement des alignements de séquence par la prise en compte de la structure 2D des protéines 232
2.2. Production de gènes codant des protéines ou chaque insertion est individuellement éliminée 234
3. Production et purification des mutants Δ1, Δ3, Δ4 et Δ5 de l'enzyme Pus7p de *S. cerevisiae* 238
4. Test de l'activité des mutants Δ1, Δ3, Δ4 et Δ5 de l'enzyme Pus7p de *S. cerevisiae* 240
5. Optimisation des conditions de production et purification de la protéine Pus7 de S. cerevisiae et du variant Pus7Δ5 en vue d'études structurales 242
5.1. Optimisation des conditions de production 242
5.2. Optimisation des conditions de purification 245
5.3. Production, purification et cristallisation du mutant Pus7Δ5 de *S. cerevisiae* 247

PARTIE 4
Etude du rôle fonctionnel possible des résidus Ψ dans le snRNA U2 de *S. cerevisiae*

1. Contexte 250
2. Démarche expérimentale 251
3. Résultats 254
A. Production et caractérisation des souches de S. cerevisiae portant des délétions individuelles ou multiples des gènes PUS1, PUS7 et snR81 254
1. Souche ΔPus1ΔPus7 254
1.1. Approche utilisée pour interrompre les ORF Pus1 et Pus7 dans une même souche 254
1.2. Analyse phénotypique et génotypique de la souche ΔPus1ΔPus7 255
1.3. Caractérisation de la souche ΔPus1ΔPus7 par analyse de l'activité ARN:Ψ-synthase sur les ARNt^{Asp}, ARNt^{Val} et pré-ARNt^{Tyr} 258
2. Souches ΔsnR81, Δ1ΔsnR81, Δ7ΔsnR81 et Δ1Δ7ΔsnR81 260
2.1. Approche utilisée pour interrompre le gène codant le snoRNA H/ACA dans les différentes souches 260

2.2. Caractérisation des souche sélectionnées après interruption du gène codant le snoRNA H/ACA snR81 263
B. Analyse du phénotype de croissance des souches mutées 265
C. Analyse globale de l'effet de l'absence des résidus Ψ dans le snRNA U2 sur l'épissage des ARN pré-messagers de S. cerevisiae par l'approche puces à ADN 266
1. Analyse de l'épissage des ARN-prémessagers par Splicing Array 266
2. Analyse et interprétation des données 272

PARTIE 5
Caractérisation des homologues de l'enzyme Pus7p de levure chez les archaea thermophiles

1. Le domaine Archaea 278
2. Contexte des études réalisées sur les enzymes Pus7p d'archaea 280
3. Démarche et Résultats 280

Publication n°4……..……281

Deficiency of the tRNATyr:Ψ35-synthase aPus7 in Archaea of the Sulfolobales order is rescued by the H/ACA sRNA-guided machinery

Sébastien Muller[1], Alan Urban[1], Arnaud Hecker[2], Fabrice Leclerc[1], Christiane Branlant[*,1] and Yuri Motorin[1]
Soumis à Nucleic Acid Research

PARTIE 6
Etude structurale de l'ARN:Ψ-synthase aPus7 de *P. abyssi* et *S. solfataricus*

1. Bilan des connaissances structurales sur les ARN:Ψ-synthases TruD/Pus7p 328
2. Résultats 329

Publication n°5…….…….……331

Expression, purification, crystallization and preliminary diffraction studies of the *Pyroccocus abyssi* tRNA:pseudouridine synthase aPus7

Alan Urban[1], Sebastien Muller[1], Jean-Michel Saliou[2], Sarah Sanglier[2], Yuri Motorin[1], Christiane Branlant[1] and Christophe Charron[1*]
Préparé pour soumission à Acta Crystallographica Section D

3. Résultats complémentaires 347
3.1. Production, purification et cristallisation de la protéine Pab aPus7 sélénométhionylée 347
3.1. Production, purification et cristallisation de la protéine Sso aPus7 348

Perspectives

1. Recherche de nouvelles ARN:Ψ-synthases 351
2. Etude de la spécificité d'action des ARN:Ψ-synthases 357
3. Poursuite des études structurales sur les ARN:Ψ-synthases Pus7p de *S. cerevisiae* et d'archaea 359

Références Bibliographiques

……..……..361

Introduction

1. Les modifications post-transcriptionnelles : cas du résidu pseudouridine

Tous les ARN métaboliquement stables, ARNm, ARNr, ARNt et UsnRNA, dans toutes les espèces vivantes, eucaryotes, archaea et bactéries, subissent des modifications post-transcriptionnelles dont certaines sont indispensables à leur activité *in cellulo* (Fig. 1). La connaissance de la seule séquence d'un ARN, telle qu'elle peut être déduite de celle de l'ADN, est insuffisante pour décrire sa structure et sa fonction, qui dépendent intimement de la présence de ces nucléotides modifiés. Actuellement, 107 modifications différentes sont recensées (Limbach et al., 1994; Motorin & Grosjean, 1998; Rozenski et al., 1999). La structure et la nomenclature de l'ensemble des nucléotides modifiés sont disponibles sur la page web : http://library.med.utah.edu/RNAmods/. Les modifications sont souvent concentrées au niveau des régions fonctionnellement importantes des ARN et ce dans les trois domaines du vivant (Fig. 1A). Ainsi, à ce jour, les Eucaryotes comportent 65 modifications différentes répertoriées contre 52 pour les Bactéries et 46 pour les Archaea (Fig. 1A). Dans le cas des ARNt, 91 nucléotides modifiés distincts, dont les plus fréquents sont représentés en Figure 2, ont été identifiés, ils représentent jusqu'à 26% de la séquence totale de ces ARN. Les ARNr comportent également un grand nombre de modifications différentes, soit 31 contre 11 pour les snRNA et 13 pour les ARNm (Rozenski et al., 1999). Notons que moins de 1% des résidus des ARNm seraient modifiés. Cependant, les analyses des ARNm et celles d'ARNnc autres que ceux évoqués ci-avant sont encore incomplètes. Les modifications résultent soit de réactions enzymatiques simples ne faisant intervenir qu'une seule enzyme, pour la méthylation (de bases ou de riboses), la déamination, l'isomérisation, la thiolation et la réduction de base lors de la formation de résidus 5-méthylcytidine (m5C), 2'-*O*-méthylés (Nm), d'inosine (I), de pseudouridine (Ψ), de 4-thiouridine (s4U) et de dihydrouridine (D), soit de réactions complexes impliquant une cascade d'activités enzymatiques, c'est notamment le cas de la formation de la wybutosine (yW) (Fig. 2) (Garcia & Goodenough-Lashua, 1998). Par contre, la formation des résidus queunine (Q) et archaeosine (gG) est particulière parce qu'elle implique, au préalable, la biosynthèse d'hétérocycles modifiés incorporés par la suite dans l'ARN par un mécanisme d'échange de bases.

Parmi l'ensemble des modifications post-transcriptionelles, la conversion de résidus U en pseudouridine (5β-D-ribofuranosyluracile ou Ψ) ou pseudouridylation (Fig. 2) et la 2'-*O*-méthylation des riboses (ou Nm) sont les des deux modifications les plus fréquentes (Davis & Allen, 1957; Cohn, 1960). L'isomérisation d'un résidu uridine en pseudouridine est catalysée par une classe d'enzymes particulières, les ARN:pseudouridine-synthases (ou ARN:Ψ-synthases) qui ne nécessitent ni énergie, ni co-facteur (Arena et al., 1978 ; Green et al., 1982 ; Kammen et al., 1988 ; Samuelsson & Olsson, 1990 ; Nurse et al., 1995 ; Wrzesinski et al., 1995a). Deux systèmes de pseudouridylation différents existent selon que la reconnaissance du résidu U est effectuée par l'ARN:Ψ-synthase elle-même ou qu'elle repose sur un appariement

ARN-ARN entre l'ARN cible et un ARN guide appelé snoRNA à boîtes H/ACA. Ces deux systèmes de pseudouridylation seront décrits plus en détails par la suite.

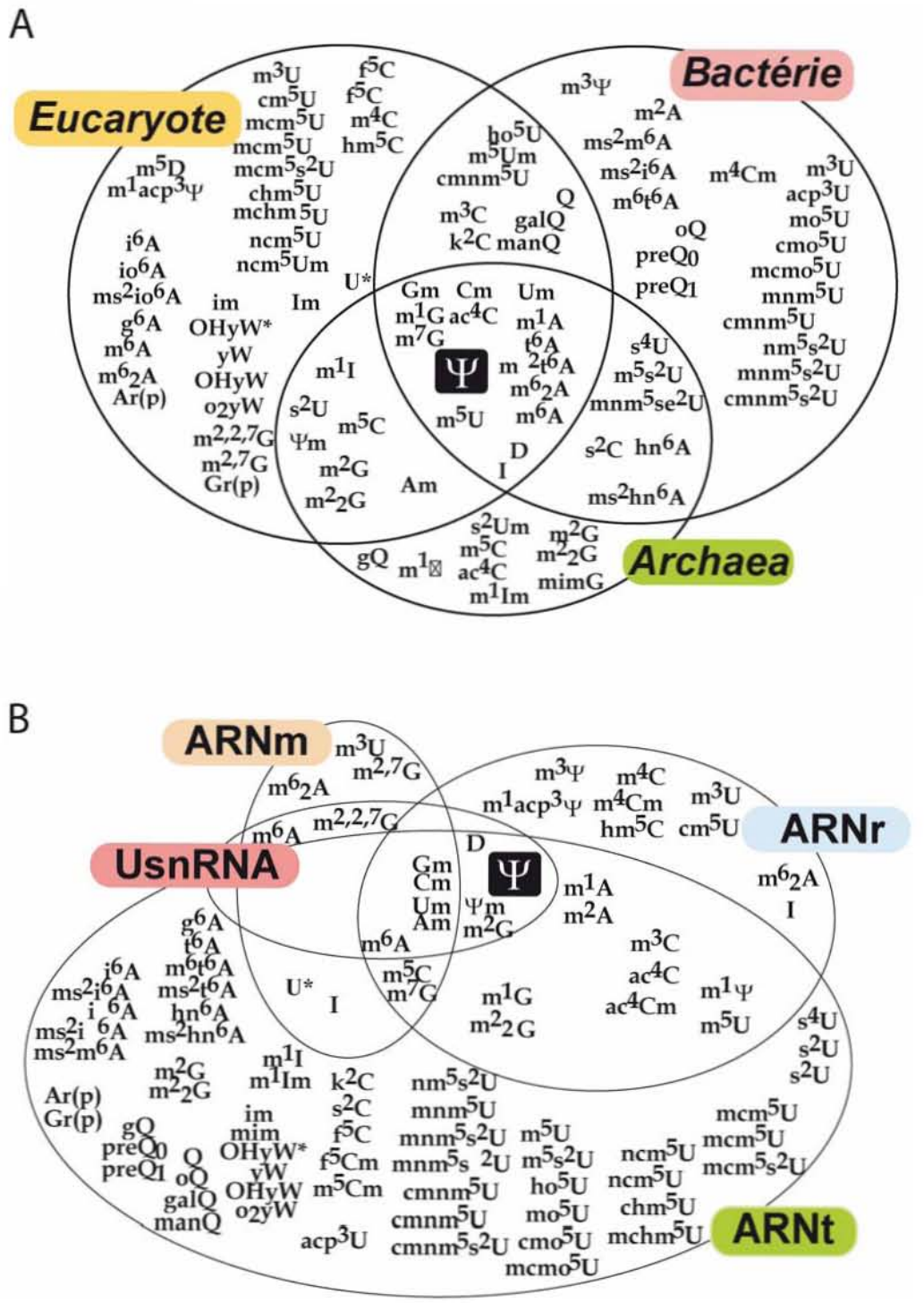


Figure 1 : Distribution des nucléotides modifiés dans les trois domaines du vivant (A) et par type d'ARN (B) (d'après, Motorin & Grosjean, 1998).
U* correspond à l'uridine formée par déamination d'une cytosine par un mécanisme d'édition post-transcriptionnel.

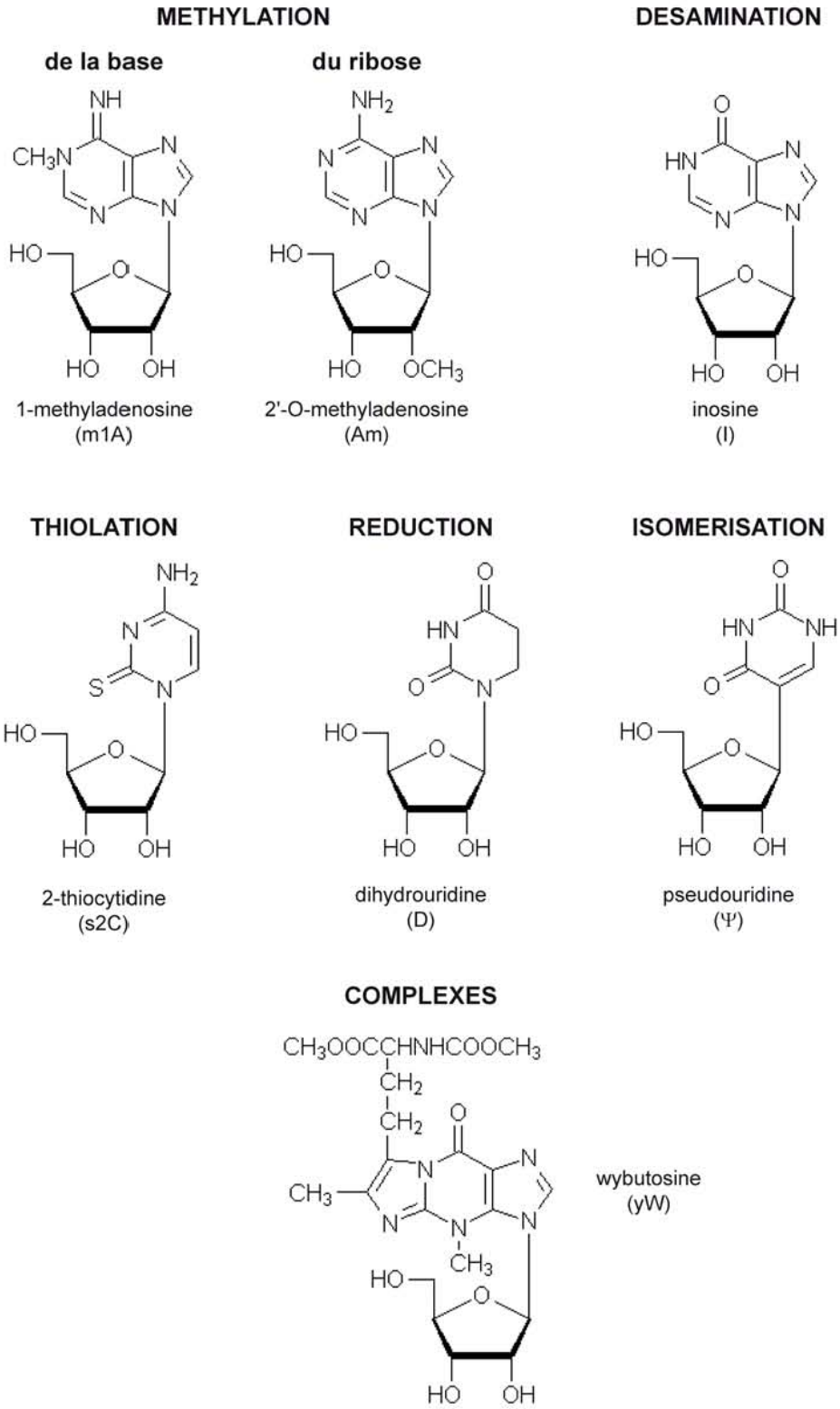


Figure 2 : Structure chimique de quelques nucléotides modifiés (d'après Limbach et al., 1994).

Le mécanisme réactionnel implique la rupture de la liaison N-glycosidique entre la base et le ribose, suivie d'une rotation de la base de 180° sur elle-même, aboutissant à la formation d'une liaison carbone-carbone entre le C_1' du ribose et le C_5 du cycle pyrimidique (Fig. 3) (Johnson & Soll, 1970; Ciampi et al., 1977; Kammen et al., 1988). Par conséquent, le résidu pseudouridine possède des propriétés différentes de celles d'un résidu uridine. Bien que la liaison C-C confère une plus grande flexibilité par rapport à la liaison N-glycosidique, les études réalisées en RMN, cristallographie en RX et simulation en dynamique moléculaire, montrent une augmentation de la rigidité du squelette ribose-phosphate. En effet, la liaison hydrogène supplémentaire restreint la mobilité et les conformations de bases présentes en amont du résidu Ψ (Charette & Gray, 2000). De plus, la liaison hydrogène supplémentaire favorise une conformation 3'endo du ribose qui restreint une conformation *anti* de la base et donc favorise l'empilement des bases dans l'environnement proche du brin portant le résidu Ψ (Davis et al., 1998). Les résidus Ψ contribueraient ainsi à stabiliser une structure ARN.

La proportion importante de gènes consacrés aux modifications des ARN dans le génome d'un organisme et en particulier de gènes dédiés à la formation de résidus pseudouridine (Björk, 1995), ainsi que la conservation de certaines des modifications tant au niveau de l'identité chimique qu'au niveau de la position des nucléotides modifiés au cours de l'évolution suggèrent un rôle essentiel de ces modifications dans le fonctionnement cellulaire. Cependant, en dépit de nombreuses études, l'importance fonctionnelle du résidu pseudouridine comme des autres modifications n'est pas complètement établie.

Lors de ma thèse, mon travail a été centré sur la caractérisation de nouveaux résidus Ψ et des systèmes enzymatiques les générant dans les ARNt et les UsnRNA de la levure *S. cerevisiae*. L'introduction aura pour objectif d'apporter les informations essentielles à l'analyse de ce travail sous forme de 4 grands chapitres. Ainsi, nous débuterons par la localisation et la distribution des résidus Ψ dans les ARN. Nous détaillerons ensuite les différents rôles prouvés ou suggérés de ces résidus pour chaque classe d'ARN. Dans la partie suivante, nous décrirons les ARN:Ψ-synthases tant d'un point de vue structural que fonctionnel en faisant le lien entre mécanisme d'action et spécificité de cette classe particulière d'enzyme. Finalement, nous présenterons les mécanismes de reconnaissance du résidu U ciblé en distinguant les deux systèmes de pseudouridylation qui reposent ou non sur des snoRNA à boîtes H/ACA.

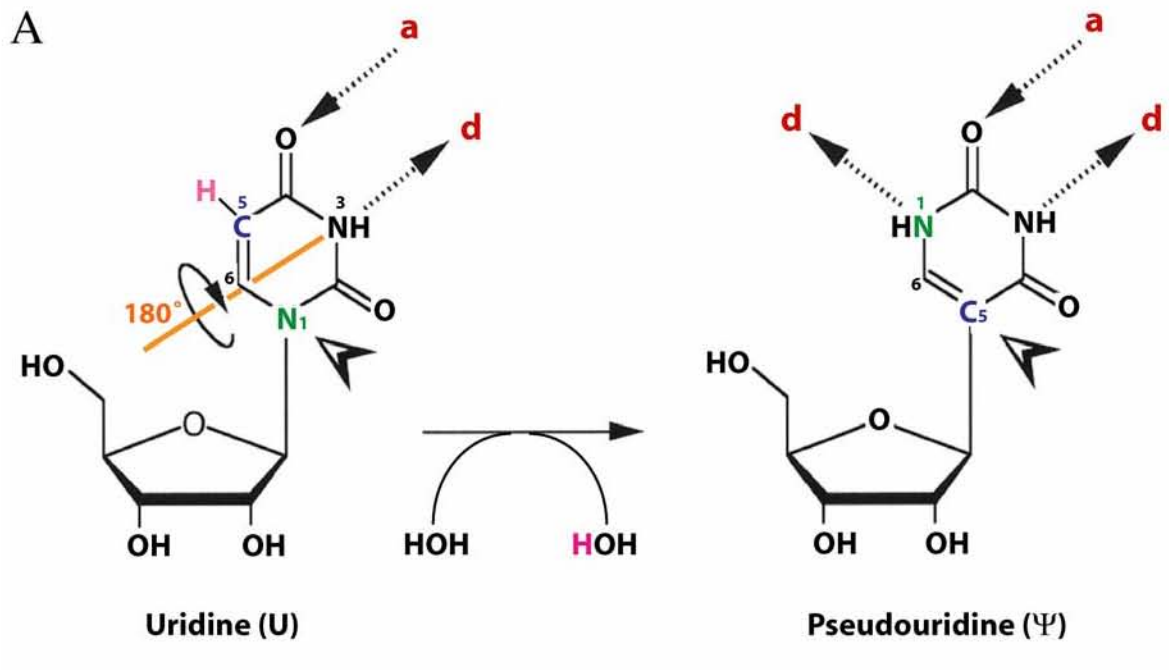


Figure 3 : Conversion par isomérisation d'un résidu uridine en pseudouridine (d'après Charette & Gray, 2000).
Les liaisons hydrogène possibles dans le sens accepteur (a) ou donneur (d) sont indiquées par des flèches. L'axe (N3-C6), autour duquel est effectuée la rotation de 180°, est présenté en orange. Les différences entre résidus U (1-β-D-ribofuranosyluracil) et Ψ (5-β-D-ribofuranosyluracil) sont indiquées en couleur dans la configuration *anti* du groupe glycosyl.

2. Localisation et distribution des résidus Ψ dans les ARN

Les résidus pseudouridines sont principalement présents dans les ARNt, les ARNr et les UsnRNA. Cependant, ils ont également été retrouvés dans d'autres ARN tels que :

- l'ARNtm 10Sa chez *E. coli* (Ray & Apirion, 1979).
- les snoRNA à boîtes C/D U3 et U8 impliqués dans le clivage du pré-ARNr (Wise et al., 1983; Reddy et al., 1985).
- le snoRNA à boîtes C/D snR4 guide de méthylation chez la levure *S. cerevisiae* (Reddy, 1988).
- le snoRNA à boîtes H/ACA snR8, guide de pseudouridylation chez la levure *S. cerevisiae* (Wise et al., 1983).

A. Cas des ARN de transfert

Les techniques d'analyse des nucléotides modifiés ont permis de mettre en évidence que les régions les plus modifiées sont les trois boucles terminales des hélices qui forment la structure en feuille de trèfle des ARNt. En effet, deux de ces boucles tirent directement leur nom des nucléotides modifiés conservés qu'elles contiennent : la boucle D dans laquelle on trouve très fréquemment des résidus dihydrouridines (D) (positions 16, 17, 20 et 20A) et la boucle TΨC qui comporte, dans plus de 90% des cas, une ribothymidine (T) (position 54) et une pseudouridine (Ψ) (position 55) qui est presque universelle (Fig. 4).

Les résidus pseudouridines sont très fréquents aux positions 13, 27, 28 et 39, et moins fréquents aux positions 31, 32, 38 et 54 (Fig. 4) (Sprinzl et al., 1998). Les Archaea, pour lesquels 59 ARNt provenant principalement d'*H. volcanii* (41 séquences) ont été analysés, présentent des résidus Ψ aux positions 13, 39, 54 et 55 (la position 54 étant très souvent une 1-methylpseudouridine) mais également, dans des cas exceptionnels, aux positions 22, 28, 38 et 52 (Grosjean et al., 1995).

Le résidu Ψ35 est présent dans tous les $ARNt^{Tyr}$ cytoplasmiques eucaryotes. Sa synthèse est intron-dépendante (Johnson & Abelson, 1983; Choffat et al., 1988) et elle dépend plus de la longueur de l'intron que de sa séquence (Choffat et al., 1988; Pienkowska et al., 2002).

B. Cas des ARN ribosomiques

L'ARNr qui représente à lui seul environ 80% de l'ARN total d'une cellule contient une proportion de nucléotides modifiés comprise entre 1 et 3%. Ces modifications sont essentiellement de deux sortes : les méthylations en 2'-*O*- des riboses et les résidus Ψ. Si l'on observe la répartition du nombre de résidus Ψ présents dans les ARNr de la petite sous-unité 16-18S et de la grande sous-unité 23-28S du ribosome, on constate qu'il y a une grande variation entre

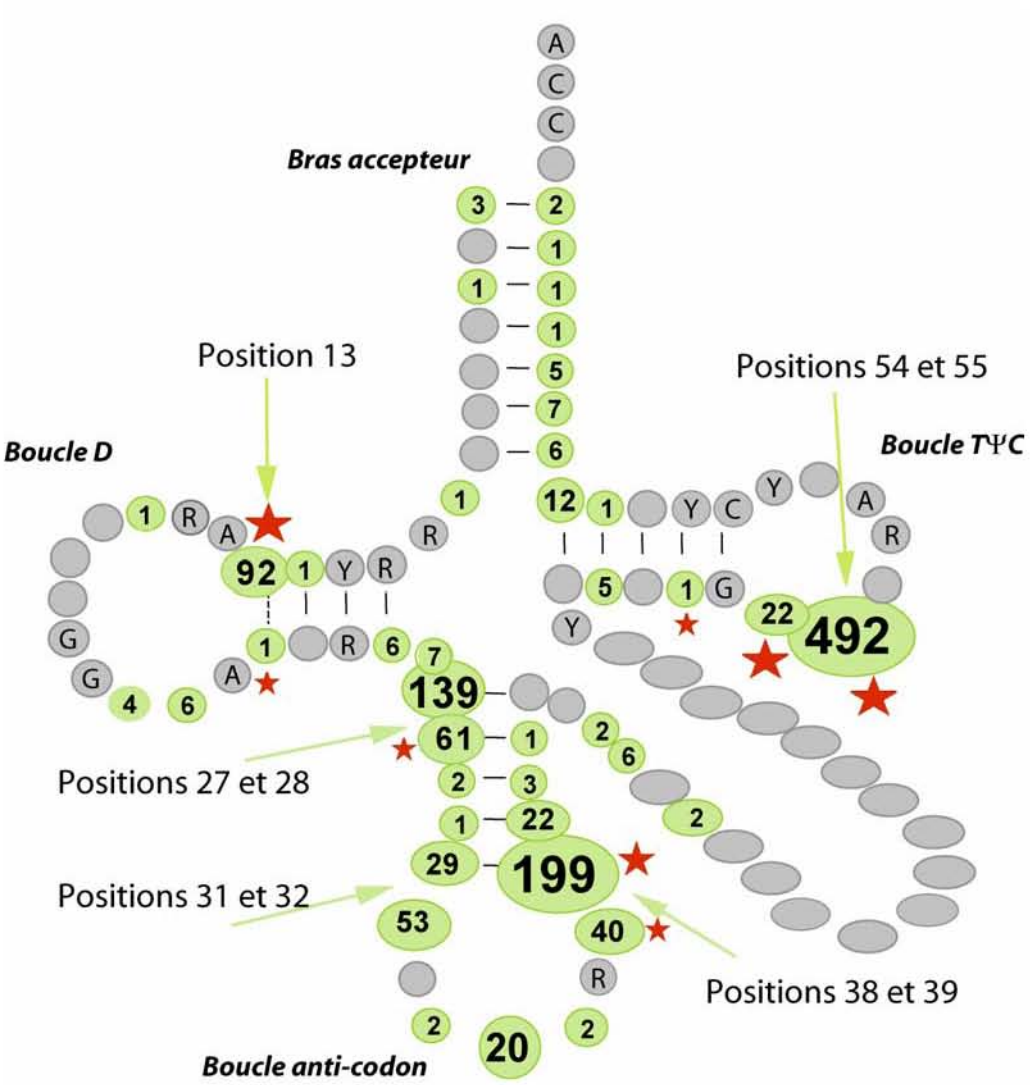


Figure 4 : Localisation et fréquences des résidus Ψ aux différentes positions des ARNt dans les trois domaines du vivant. D'après la banque de données de modifications dans les ARNt (Rozenski et al., 1999).

Ces données regroupent les informations concernant 571 séquences d'ARNt. 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 conservés dans plus de 90% des cas sont indiqués (R = purine, Y = pyrimidine).

les 3 domaines du vivant (Tableau 1). En effet, les bactéries possédent le plus faible nombre de résidus Ψ (10 pour *E. coli*), tandis que les eucaryotes sont les plus richement dotés avec un peu moins de 100 résidus Ψ détectés dans les ARNr humains. Les Archaea se situeraient dans une situation intermédiaire, car bien que certains organismes ne semblent posséder qu'un

nombre limité de résidus Ψ, l'analyse partielle des ARNr de *P. abyssi* a révélé la présence d'au moins 15 résidus Ψ (Muller et al., 2008). Il est à noter, que le nombre de résidus pseudouridines dans les ARNr mitochondriaux et chloroplastiques est comparable à celui observé chez les bactéries (Tableau 1). Quelques résidus Ψ ont également été mis en évidence dans l'ARNr 5S de certains organismes (Miyazaki, 1974; Szymanski et al., 2000; Szymanski et al., 2002).

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
T. thermophilus	1	nd	Seq. (100)	Guymon et al., 2005
C. acetobutylicum	1	nd	Total (16S)	Emmerechts et al., 2007
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., 1998
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 5.8S				
S. cerevisiae	**1**		**Seq. (100)**	**Rubin, 1973**
M. musculus	2		Seq. (100)	Nazar et al., 1976 ; Khan et al., 1977
H. sapiens	2		Seq. (100)	
ARNr 5S				
S. cerevisiae	**1**		**Seq. (100)**	**Miyazaki, 1974**
E. gracilis	1		Seq. (100)	Szymanski et al., 2000
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

Tableau 1 : 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 bactéries (en vert), archaea (en bleu) et eucaryotes (en jaune). D'après (Ofengand, 2002)**.**

L'analyse des ARNr a été faite soit par séquençage de l'ARN (Seq.) et détermination précise de la position (voir technique basée sur l'utilisation de CMCT et de la transcriptase inverse développée dans

la partie Matériels & Méthodes), 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.

Les résidus Ψ ne sont pas répartis de façon équitable entre les ARNr des 2 sous-unités du ribosome. En effet, la petite sous-unité contient peu de résidus Ψ et ces derniers ne sont pas conservés entre espèces. Ainsi, le seul résidu Ψ trouvé dans l'ARNr 16S d'*E. coli* semble conservé chez les bactéries (Fig. 5A, triangle rouge) mais n'a pas d'homologue chez les eucaryotes. Par ailleurs, si l'on s'intéresse aux bactéries, on constate que les quelques modifications post-transcriptionnelles présentes dans l'ARN 16S sont localisées au niveau des régions de l'ARNr contractant des interactions soit avec l'ARNr 23S, soit avec les molécules d'ARNt fixées au niveau des sites A et P (Fig. 5)

Concernant l'ARNr de la grande sous-unité, la localisation des résidus Ψ sur la structure tridimensionnelle du ribosome révèle une concentration élevée de ces résidus dans les régions fonctionnellement importantes (Fig. 5 et 6) (Charette & Gray, 2000; Decatur & Fournier, 2002 ; Ofengand, 2002). Ces régions sont impliquées dans les interactions avec les molécules d'ARNt au niveau des sites A, P et dans une moindre mesure au site E, à proximité du tunnel de sortie du peptide en cours de synthèse, au niveau du centre peptidyl-transférase directement impliqué dans la catalyse ou au niveau des sites de fixation des facteurs d'élongation. Le grand ARNr contient des résidus Ψ très conservés dans l'évolution (Branlant et al., 1981b) (Fig. 6 et 7). Ainsi, au sein de l'ARN de la grande sous-unité des trois domaines du vivant, les modifications sont principalement regroupées dans les domaines II, IV et V (Fig. 7) (Branlant et al., 1981a; Decatur & Fournier, 2002). Les domaines IV et V comportent la majorité des résidus Ψ. Le domaine V est situé au niveau du centre peptidyl-transférase et participe à la catalyse. Le domaine IV est le centre de décodage en interaction avec à la fois l'ARNm et la tige-boucle anti-codon de l'ARNt. Il est à noter que l'hélice 69 du domaine IV présente les résidus Ψ les mieux conservés au cours de l'évolution (Fig. 7). Or cette hélice, de même qu'une région du domaine II, est précisément à l'interface entre les deux sous-unités (Merryman et al., 1999a; Merryman et al., 1999b). Les domaines II et IV, bien que distants du domaine V sur le plan de la structure secondaire, sont en fait très proches du site de formation de la liaison peptidique. Dans le cas de la levure *S. cerevisiae*, il a été démontré que les ARNr 5S et 5,8S possèdent chacun un unique résidu Ψ, aux positions 50 et 73 respectivement (Rubin, 1973; Miyazaki, 1974), qui sont très conservées chez les hémi-ascomycètes (Decatur & Schnare, 2008). Nous pouvons conclure que même si la position des résidus Ψ n'est pas toujours identique entre espèces d'un point de vue séquence, leur localisation spatiale, principalement dans des régions de haute importance fonctionnelle, est très conservée ce qui confirme l'importance des résidus Ψ.

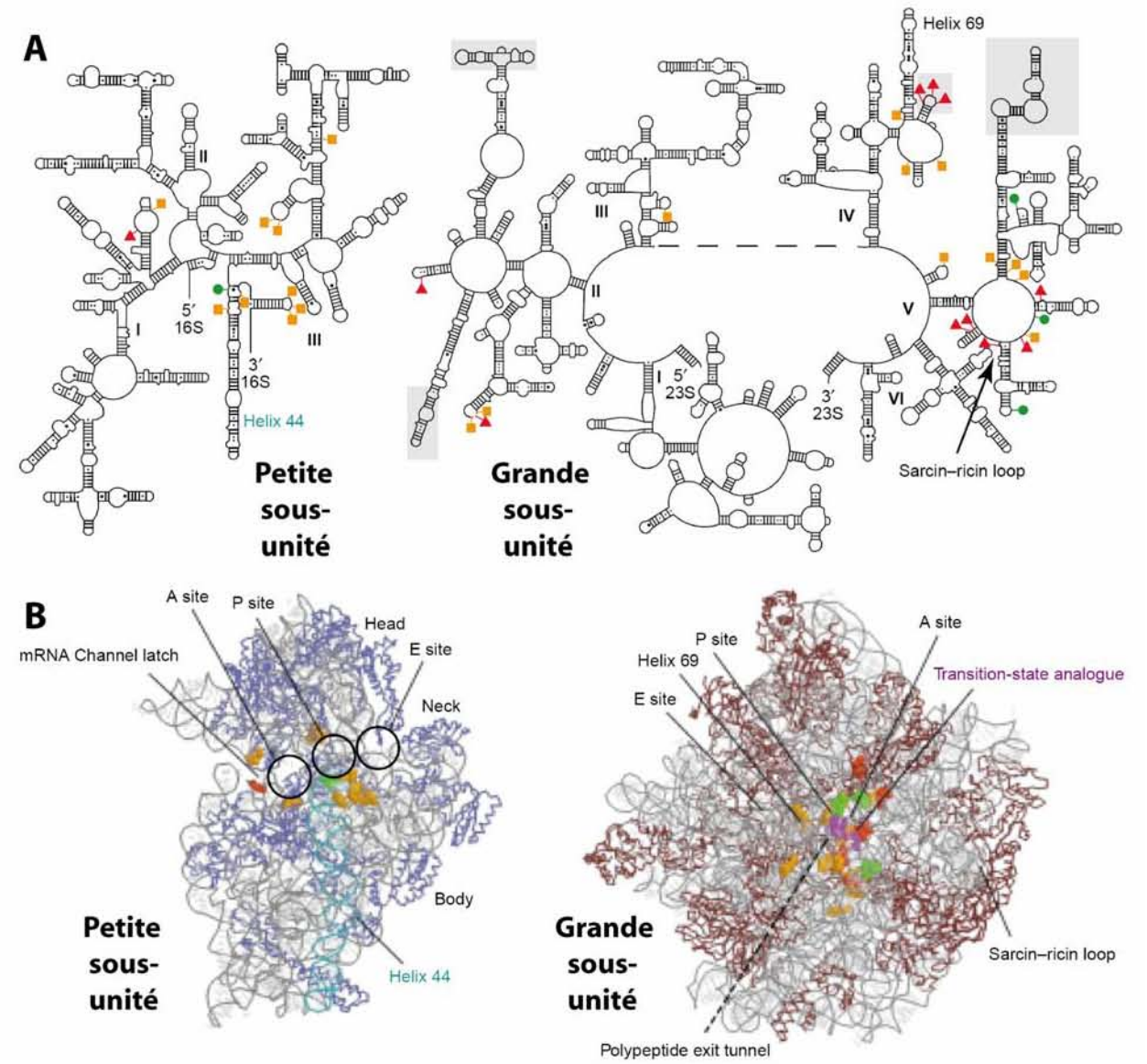


Figure 5 : Distribution des nucléotides modifiés dans les ARNr d'*E. coli* (d'après (Decatur & Fournier, 2002).

A. Structures secondaires des ARNr 16S et 23S. Les résidus Ψ sont représentés par des triangles rouges, les 2'-*O*-méthylations (Nm) par des cercles verts et les méthylations de bases (mN) par des carrés orange.

B. Structures tridimensionnelles des petites et grandes sous-unités du ribosome d'*E. coli* 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. L'analogue de l'état de transition est montré en violet.

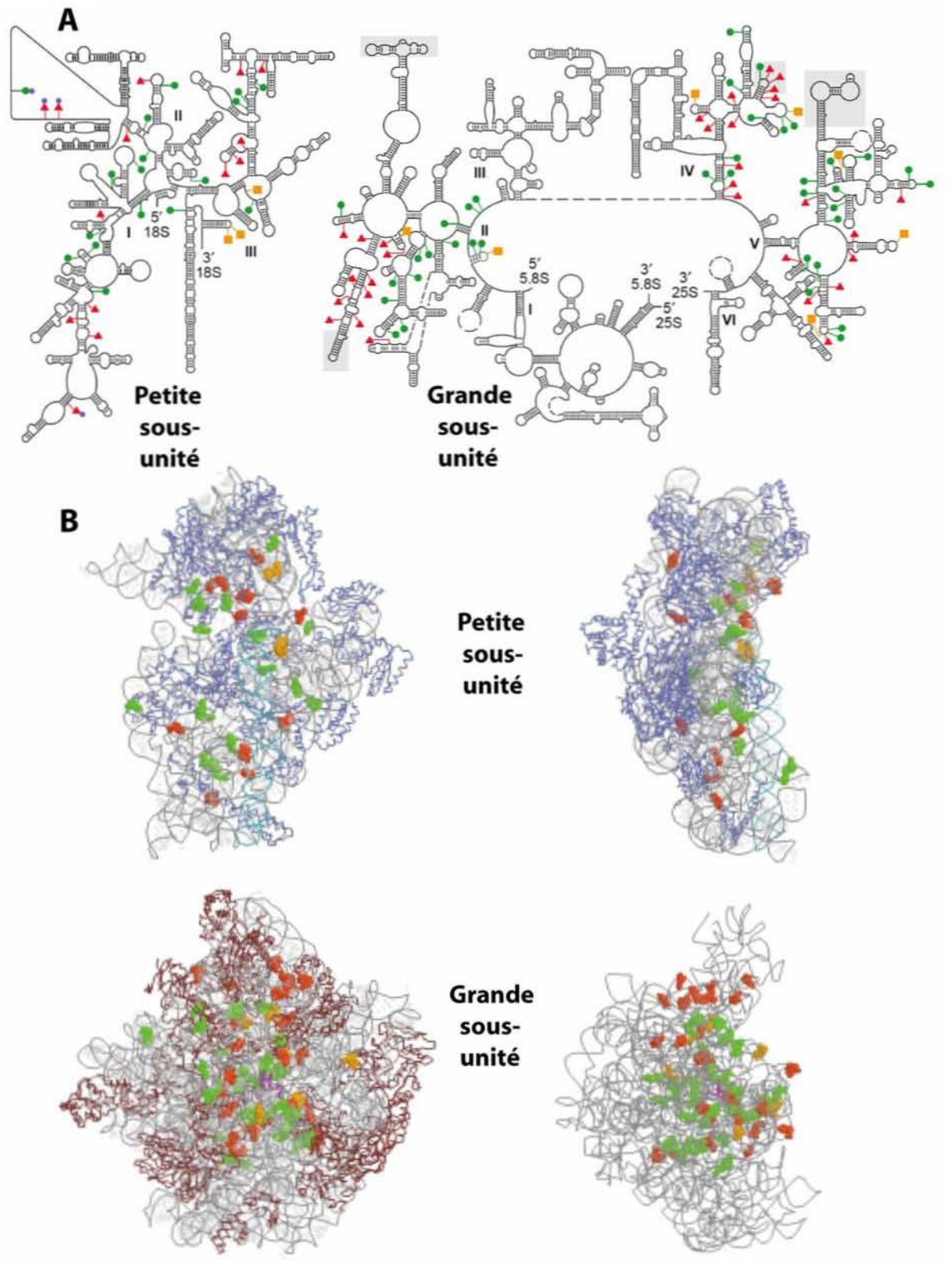


Figure 6 : Distribution des nucléotides modifiés dans les ARNr de *S. cerevisiae.* D'après (Decatur & Fournier, 2002). Voir légende de la Figure 4. Les panneaux B. correspondent à deux vues différentes de la petite sous-unité ainsi qu'à deux vues de la grande sous-unité.

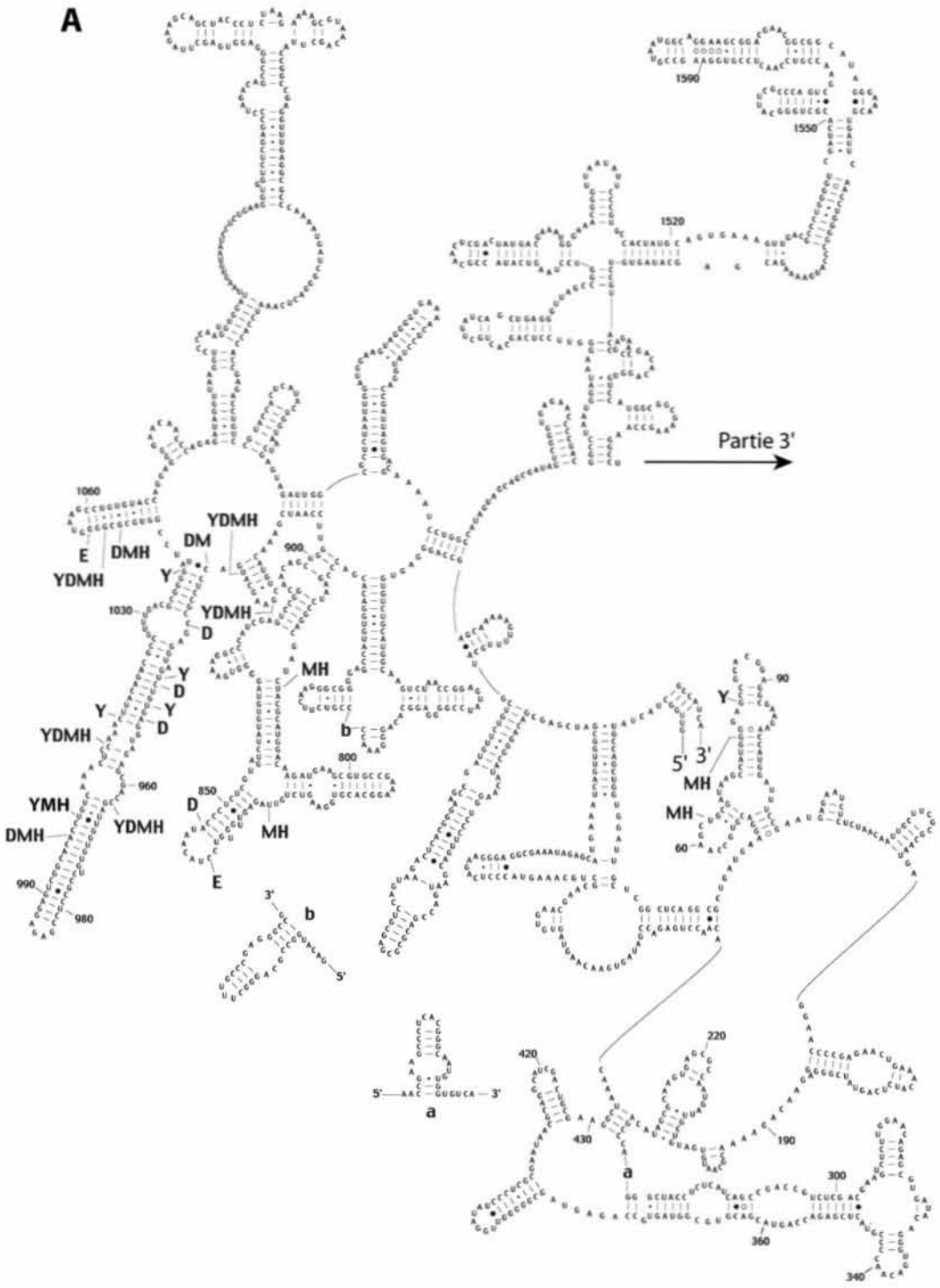


Figure 7 : Localisation des résidus Ψ dans le grand ARN de la grande sous-unité ribosomique chez les diverses espèces où ils ont été étudiés (d'après (Ofengand, 2002) et corrigé par (Kirpekar et al., 2005). La structure secondaire et la numérotation de *H. marismortui* sont utilisées (voir à l'adresse suivante : http://www.rna.ccbb.utexas.edu/). Les positions des résidus Ψ 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) ont été obtenues à partir des références indiquées en Tableau 1. La séquence est divisée en région 5' (A) et région 3' (B).

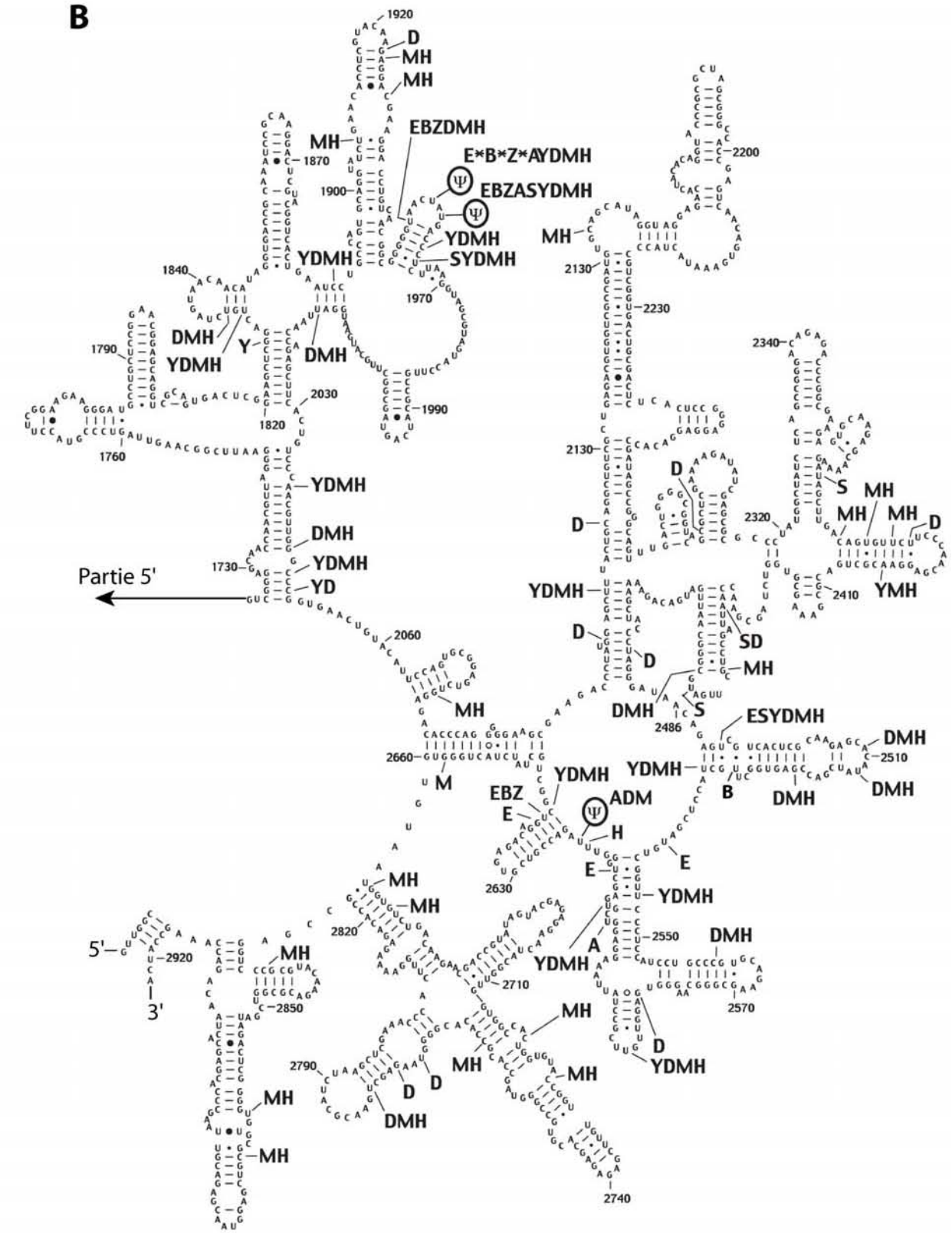


Figure 7 suite : voir légende page précédente

C. Cas des UsnRNA

1. Rôles des UsnRNA

Chez les organismes supérieurs (eucaryotes), les gènes qui codent des protéines sont constitués d'une suite d'exons et d'introns alternés. Les pré-ARNm transcrits à partir de ces gènes vont subir un processus de maturation comprenant la fixation d'une coiffe en 5', la polyadénylation en 3' et l'épissage pour finalement donner lieu à l'ARNm dit mature. L'épissage est un processus qui se produit dans le noyau de la cellule et qui consiste en l'excision des introns et en la ligation des exons. L 'ARNm mature, constitué des seuls exons, est alors exporté vers le cytoplasme pour être traduit en protéine. L'épissage est assuré par un macrocomplexe ribonucléoprotéique appelé spliceosome (Moore et al., 1993) qui fait intervenir 5 petits ARN nucléaires (snRNA) (Fig. 8) : U1, U2, U4, U5 et U6, sous forme de particules ribonucléoprotéiques appelées UsnRNP (snRNA associé à des protéines spécifiques), ainsi que de nombreuses autres protéines (Will & Luhrmann, 1997); (Will & Luhrmann, 2001).

Les ARN pré-messagers maturés au sein du spliceosome présentent un certain nombre d'éléments structuraux essentiels au mécanisme d'épissage (Fig. 9) (Shapiro & Senapathy, 1987) : les jonctions exon-intron en 5' et en 3', ainsi qu'un élément interne de l'intron, la boîte de branchement, située à moins d'une centaine de nucléotides de l'extrémité 3' de l'intron chez la levure. La séquence de ces 3 éléments structuraux est très conservée dans toutes les espèces, et notamment chez *S. cerevisiae*, où la conservation des séquences est la plus stricte. En plus de ce spliceosome majeur qui est universel chez les eucaryotes, il existe un spliceosome mineur, qui sert à l'épissage d'introns mineurs, mis en évidence chez les métazoaires et les plantes, pour lesquels les dinucléotides 5'GU et 3'AG caractéristiques des introns majeurs sont remplacés par les dinucléotides 5'AU et 3'AC (Jackson, 1991; Hall & Padgett, 1994). Ces spliceosomes mineurs mettent en jeu une autre série de UsnRNP : les snRNP U11, U12, U4atac et U6atac, qui remplacent respectivement les snRNP U1, U2, U4 et U6 (pour revues, Mount, 1996; Tarn & Steitz, 1997). La snRNP U5, est commune aux spliceosomes mineurs et majeurs.

La réaction d'épissage proprement dite s'effectue en deux temps (Fig. 9), avec tout d'abord une attaque nucléophile du 2'-OH du ribose de l'adénosine de la boite de branchement sur le phosphate de la jonction exon-intron situé en 5'. Après cette coupure, le groupement 3'-OH libéré au niveau de l'exon 1 attaque le phosphate de la jonction intron-exon 2. Les produits de cette réaction sont d'une part les deux exons ligués correctement et d'autre part, l'intron cyclisé au niveau de l'adénosine de la boite de branchement. En raison de sa forme particulière, cette forme de l'intron est appelée lasso (ou lariat); celui-ci sera dégradé par des RNases (Moore et al., 1993).

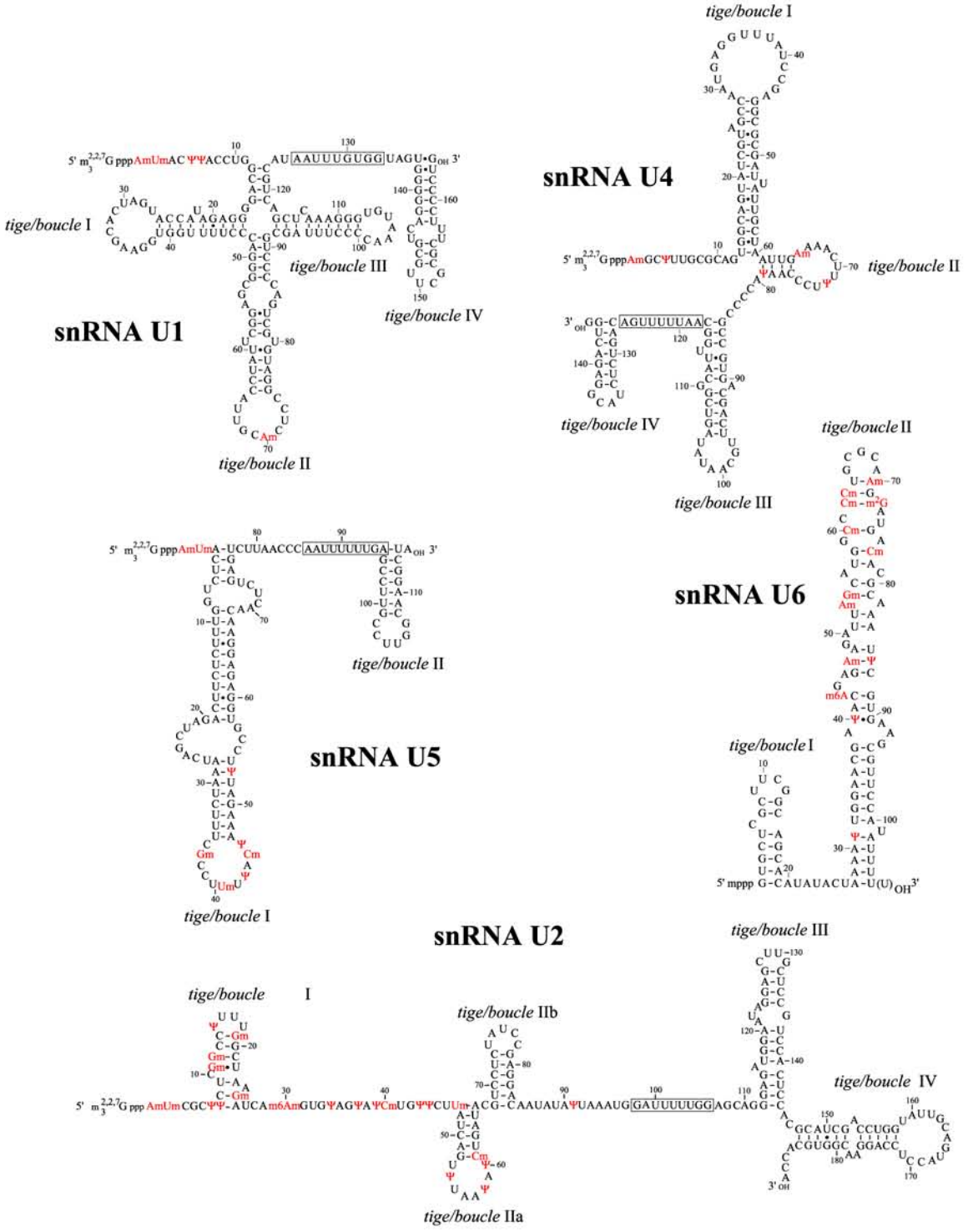


Figure 8 : Structures secondaires des snRNA spliceosomaux U1, U2, U4, U5 et U6.

Les structures secondaires représentées sont celles des snRNA de rat établies par : (Branlant et al., 1980), (Mount & Steitz, 1981) pour le snRNA U1, (Branlant et al., 1982), pour le snRNA U2, (Krol et al., 1981a) pour le snRNA U4, (Branlant et al., 1983) pour le snRNA U5 et Annie Mougin pour le snRNA U6 (résultats non publiés). La séquence permettant la fixation des protéines Sm est encadrée dans les différents snRNA (Branlant et al., 1982). Les nucléotides modifiés identifiés sont indiqués en rouge.

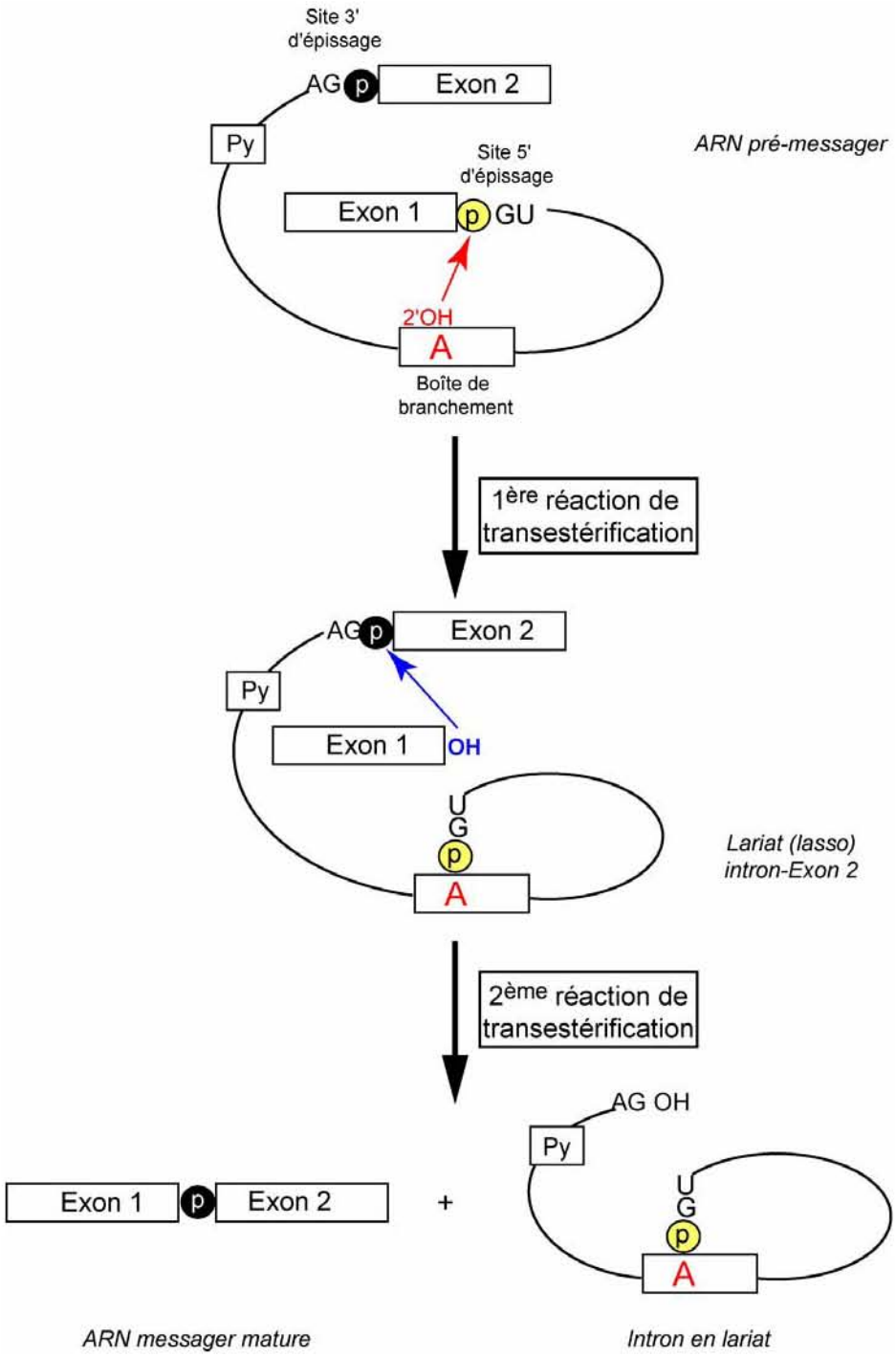


Figure 9 : Les différentes étapes du processus d'épissage des ARN pré-messagers nucléaires (Moore et al., 1993).
« Py » représente la séquence polypyrimidine, cette séquence, fréquemment présente entre le positions -15 et -5 en amont du site 3' d'épissage, n'est pas indispensable chez la levure (Rymond & Rosbash, 1985).

2. La réaction d'épissage

Le processus d'épissage est un mécanisme dynamique qui suit une succession précise d'évènements au cours de laquelle les différents snRNP s'assemblent et se désassemblent du spliceosome de façon séquentielle. La réaction débute par la fixation de la snRNP U1 au niveau du site 5' d'épissage, par appariement de bases entre l'extrémité 5' du snRNA U1 et l'ARN pré-messager (Fig. 10), et par la mise en place d'interactions entre les protéines de la snRNP U1 et l'ARN pré-messager et entre les protéines de la snRNP U1 et les facteurs d'épissage (Newman, 1994); (Madhani & Guthrie, 1994); (Nilsen, 1998). La snRNP U2 se fixe ensuite de façon ATP dépendante à l'ARN pré-messager : cette étape nécessite la fixation préalable de facteurs protéiques, tels que U2AF et SF1 (Zamore & Green, 1989; Zamore & Green, 1991; Kramer, 1992), ainsi que l'appariement du snRNA U2 avec la boîte de branchement de l'intron (Madhani & Guthrie, 1994; Newman, 1994). On aboutit alors au complexe B, l'adénosine responsable de la première attaque nucléophile est située en boucle latérale dans l'appariement snRNA U2-point de branchement (Query et al., 1994) (Fig. 10).

Au sein de la snRNP U4/U6, les snRNA U4 et U6 forment un duplex d'ARN par appariements intermoléculaires. Par association avec la snRNP U5, la snRNP U4/U6 forme une tri-snRNP [U4/U6-U5] qui est associée au complexe B, ce qui conduit après plusieurs transitions structurales à la formation du complexe A1 catalysant la 1ère étape de la réaction d'épissage (Will & Luhrmann, 1997); (Will & Luhrmann, 2001). Dans ce complexe, la boucle terminale du snRNA U5, dont la séquence est très conservée, interagit avec l'extrémité 3' de l'exon 1, puis après de nouvelles transitions structurales, ayant lieu après la 1ère étape de la réaction d'épissage, cette boucle interagit avec les 2 extrémités des exons (Cortes et al., 1993); (Newman & Norman, 1991); (Newman & Norman, 1992); (Newman et al., 1995), ce qui permet de rapprocher les deux exons, afin de les liguer.

Les changements conformationnels ayant lieu après association de la tri-snRNP au complexe B sont importants, puisque le duplex d'ARN U4/U6 est dissocié au profit d'une association U2/U6 (Fig. 10) et le snRNA U6 contracte un appariement avec l'extrémité 5' de l'intron après déstabilisation de l'interaction entre le snRNA U1 et l'ARN pré-messager (Hausner et al., 1990); (Wu & Manley, 1991); (Wolff & Bindereif, 1993); (Lesser & Guthrie, 1993).

Les interactions ARN/ARN mises en jeu lors de l'assemblage des spliceosomes mineurs sont de même nature que celles décrites pour les spliceosomes majeurs (Tarn & Steitz, 1996); (Yu & Steitz, 1997); (Kolossova & Padgett, 1997); (Incorvaia & Padgett, 1998). La grande conservation dans toutes les espèces, de la structure formée entre le snRNA U2 et le snRNA U6, et leurs homologues U12 et U6atac, ainsi que leur ressemblance avec le site catalytique des introns de groupe II suggèrent que cette structure intervient directement dans la catalyse (Jacquier, 1998).

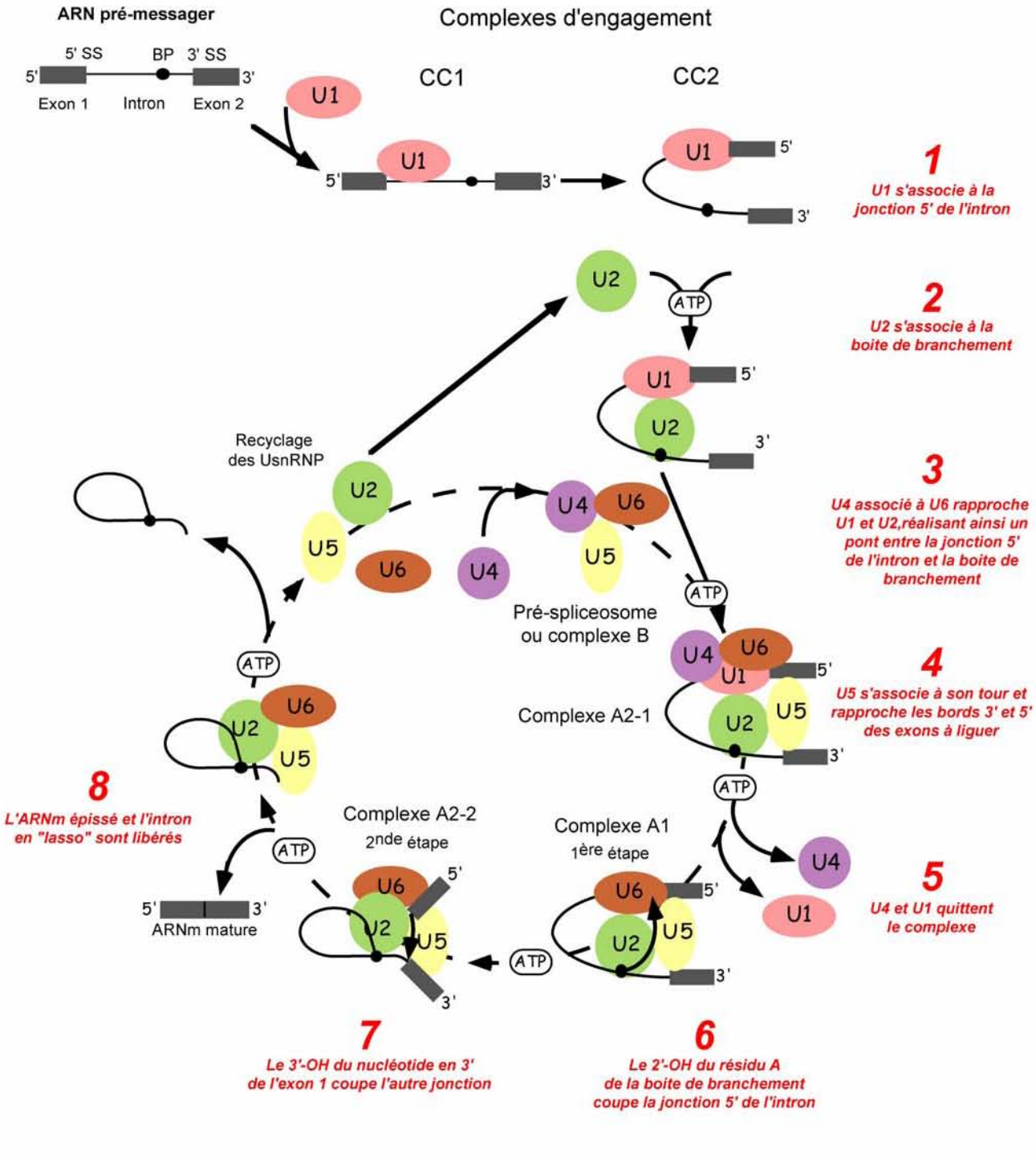


Figure 10 : Représentation schématique du cycle d'assemblage, d'action et de dissociation du spliceosome chez *S. cerevisiae*.

5'SS et 3'SS représentent respectivement, les sites 5' et 3' d'épissage et BP, la boîte de branchement. Les étapes ATP dépendantes sont indiquées par ATP encerclé. Les noms des complexes suivent la nomenclature de *S. cerevisiae*. Les principales étapes de la réaction sont numérotées en rouge.

3. Localisation des résidus Ψ dans les UsnRNA

Les UsnRNA possèdent un nombre limité de nucléotides modifiés, parmi lesquels des résidus 2'-O-méthylés, des résidus Ψ et quelques méthylations de bases (m^6A principalement). Ces modifications ont été étudiées dans un certain nombre d'espèces (Massenet et al., 1998) : (1) chez les vertébrés: poulet, souris, homme et rat (Harada et al., 1980); (Branlant et al., 1981b); (Kato & Harada, 1981); (Krol et al., 1981a); (Branlant et al., 1982) ; (Branlant et al., 1983); (Reddy, 1988), (2) chez *Drosophila melanogaster* (Myslinski et al., 1984), (3) chez des plantes (Kiss et al., 1988), (Solymosy & Pollak, 1993), et (4) chez les levures *S. pombe* (Gu et al., 1996) et *S. cerevisiae* (Massenet et al., 1999) et une fois encore, la proportion de modification des snRNA est plus élevée chez les organismes plus complexes (vertébrés, plantes et insectes) que chez les levures (Tableau 2).

	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

Tableau 2 : Récapitulatif du nombre de résidus modifiés présents dans les UsnRNA de diverses espèces.
Ψ, 2'O et mN, indiquent respectivement des résidus pseudouridines, 2'-O-méthylés et des méthylations de bases. D'après (Massenet et al., 1998). « nd » signifie non déterminé.

Concernant la distribution des résidus Ψ dans les UsnRNA, on constate que la plupart des résidus sont regroupés au niveau de régions fonctionnellement importantes et donc bien conservées dans ces ARN. Ainsi, le snRNA U1 comporte au maximum 2 résidus qui sont localisés dans la région 5' qui s'apparie avec le site 5' d'épissage de l'ARN pré-messager. Dans le cas du snRNA U2, les résidus Ψ sont également concentrés dans la région 5' terminale (Fig. 8). Comme nous l'avons vu, les snRNA U2 et U6 sont appariés dans la forme active des spliceosomes. Les segments des snRNA U2 et U6 impliqués dans les appariements hétérologues sont fortement modifiés (Fig. 11). C'est notamment le cas pour les hélices Ia, II et III, nécessaires à la réaction d'épissage (Datta & Weiner, 1991) ; (Wu & Manley, 1991) ; (Sun & Manley, 1995).

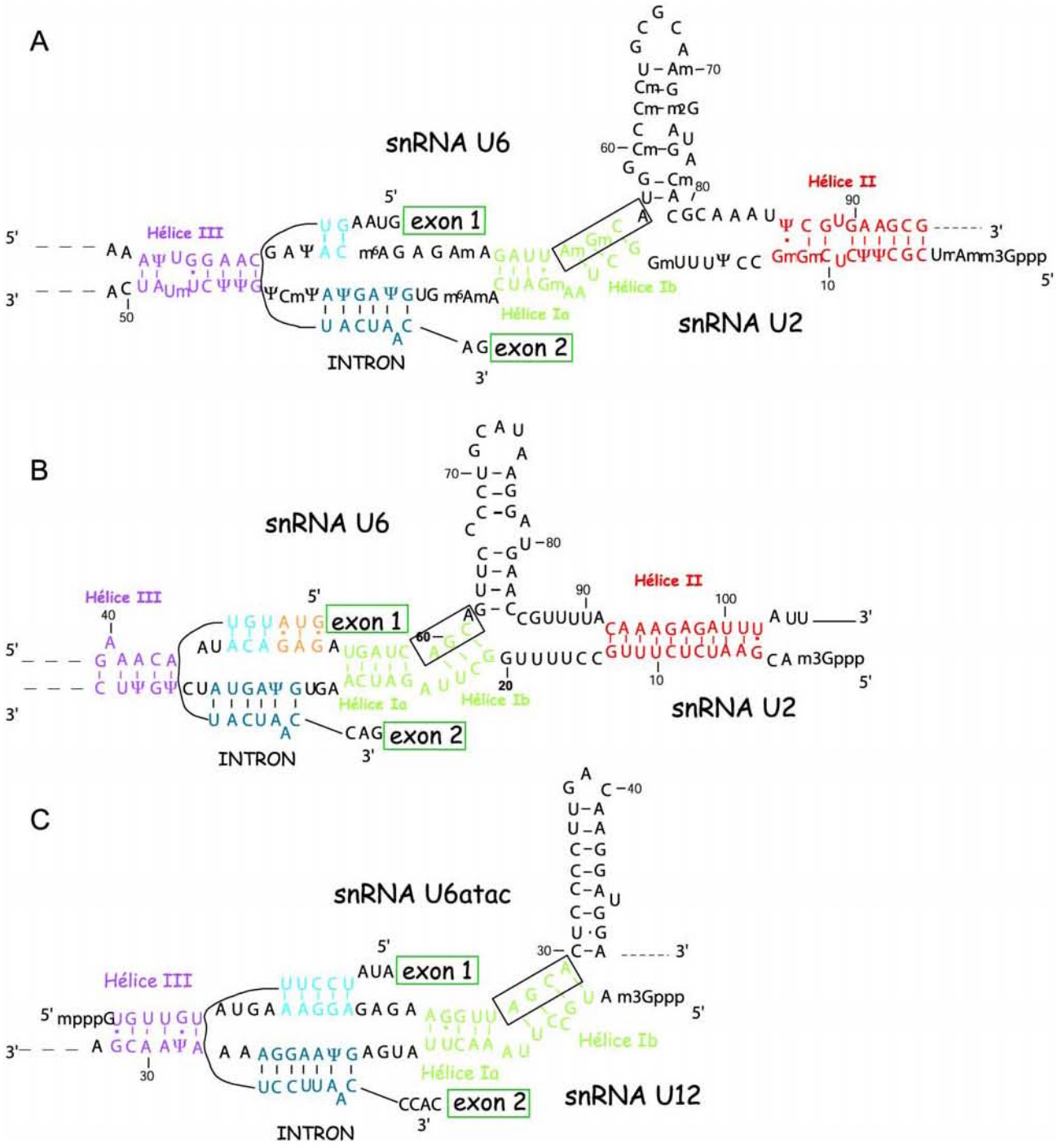


Figure 11 : Les interactions ARN/ARN impliquant les snRNA U6 et U2 de vertébrés (A) ou de *S. cerevisiae* (B) et les snRNA U6atac et U12 (C), mises en place au sein des complexes spliceosomaux.

La séquence AGCA conservée dans les snRNA U6 et U6atac est encadrée. Les hélices I, II et III sont respectivement représentées en vert, rouge et violet. L'appariement de bases entre le snRNA U2 ou U12 et la boîte de branchement est indiqué en bleu foncé et l'appariement de bases entre le snRNA U6 ou U6atac et l'ARN pré-messager en bleu clair.

Ainsi, pour le snRNA U6 de vertébrés, deux résidus Ψ sur les trois présents sont impliqués dans la formation des hélices II et III lors de l'interaction U2/U6 (Fig. 11A), le troisième résidu Ψ étant situé immédiatement en amont de la séquence phylogénétiquement conservée ACAGAG. Il est à noter que le résidu Ψ impliqué dans la formation de l'hélice III entre les snRNA U2 et U6 est également impliqué dans la formation de l'hélice III entre les snRNA U4 et U6 (Fig. 12A).

Le snRNA U4 de vertébrés contient trois résidus Ψ, le snRNA U4atac n'en contient qu'un seul et le snRNA U4 de la levure *S. cerevisiae* n'en contient aucun. De même, le snRNA U6 de vertébrés, qui comporte de nombreuses modifications post-transcriptionnelles, contient 3 résidus Ψ, alors que les snRNA U6atac et U6 de la levure *S. cerevisiae* n'en contiennent aucun (Massenet et al., 1998). Nous ne parlerons donc ici que des snRNA U4 et U6 des spliceosomes majeur et mineur de vertébrés. Les résidus Ψ de ces ARN sont localisés au niveau des régions impliquées dans des interactions ARN-ARN, soit de type U2/U6, soit de type U4/U6 (U4atac/U6atac) (Fig. 12). Sur les trois résidus Ψ présents dans le snRNA U4, deux sont impliqués dans la formation d'hélices hétérologues avec le snRNA U6, un au niveau de l'hélice II et l'autre au niveau de l'hélice III (Fig. 68). Le troisième est dans une région non appariée. Le seul résidu Ψ présent dans le snRNA U4atac est également impliqué dans la formation de l'hélice II avec le snRNa U6atac (Fig. 12B). Dans le cas du snRNA U6 de vertébrés, sur les trois résidus Ψ présents, un seul est impliqué dans une interaction avec le snRNA U4. Ce résidu Ψ est impliqué dans la formation de l'hélice III (Fig. 12A).

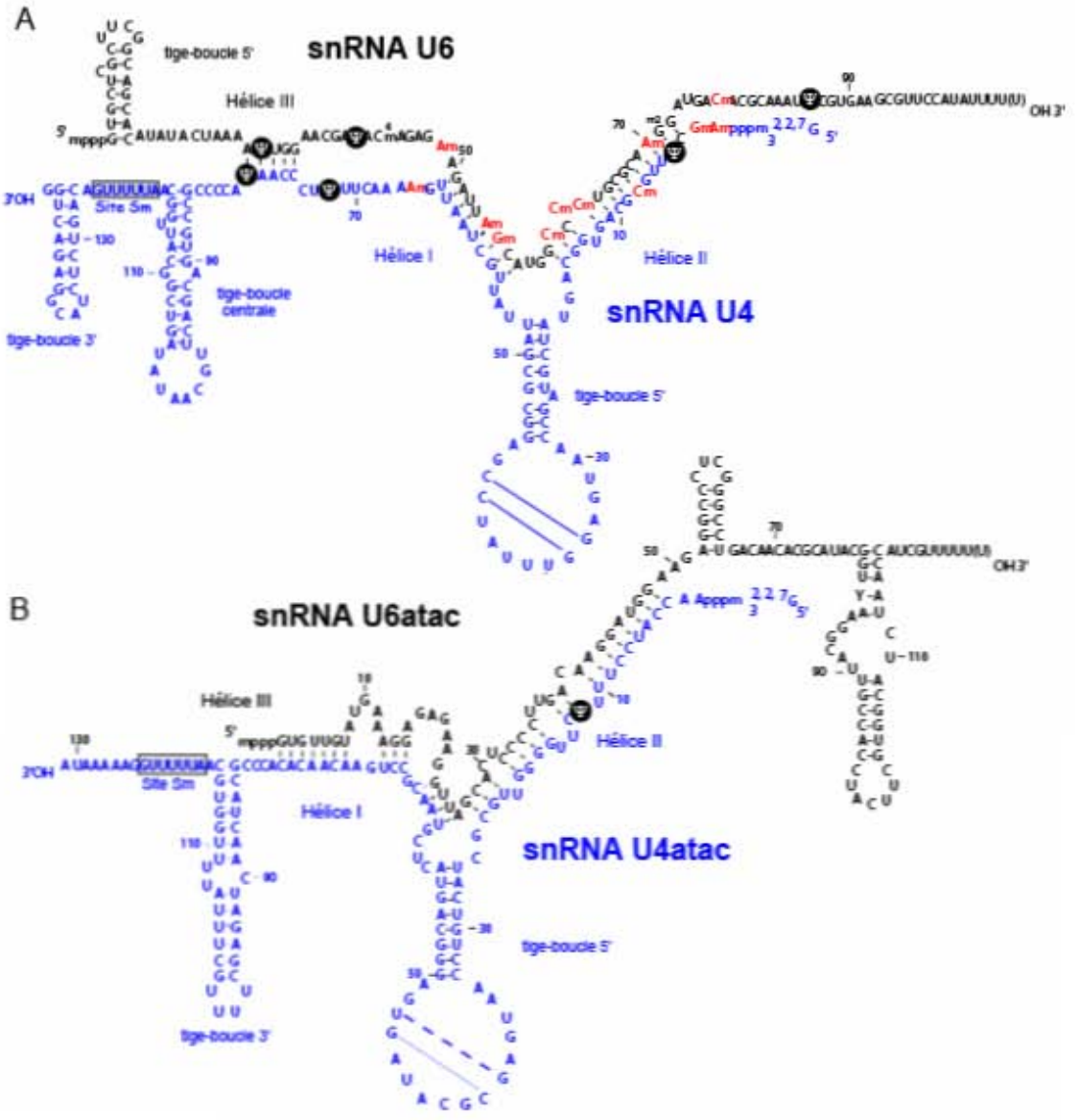


Figure 12 : Interactions ARN-ARN entre les snRNA U4 et U6 des spliceosomes majeurs et mineurs de vertébrés.
Les résidus Ψ sont indiquéspar des boules noires, les nucléotides autres modifiés sont reprsentés en rouge (Madhani & Guthrie, 1994); (Massenet et al., 1998) ; (Wu & Krainer, 1999). L'interaction entre les snRNA U4 et U6 de la levure *S. cerevisiae* n'est pas représentée du fait de l'absence de résidus Ψ dans ces deux ARN.

Concernant le snRNA U5, on notera que les modifications post-transcriptionnelles sont concentrées au niveau de la boucle terminale de la structure tige-boucle 1. Les modifications de la boucle terminale sont très conservées au cours de l'évolution, comme illustré dans la

Figure 13, ce qui suggère une importance fonctionnelle de ces résidus (Szkukalek et al., 1995; pour, revue, Massenet et al., 1998).

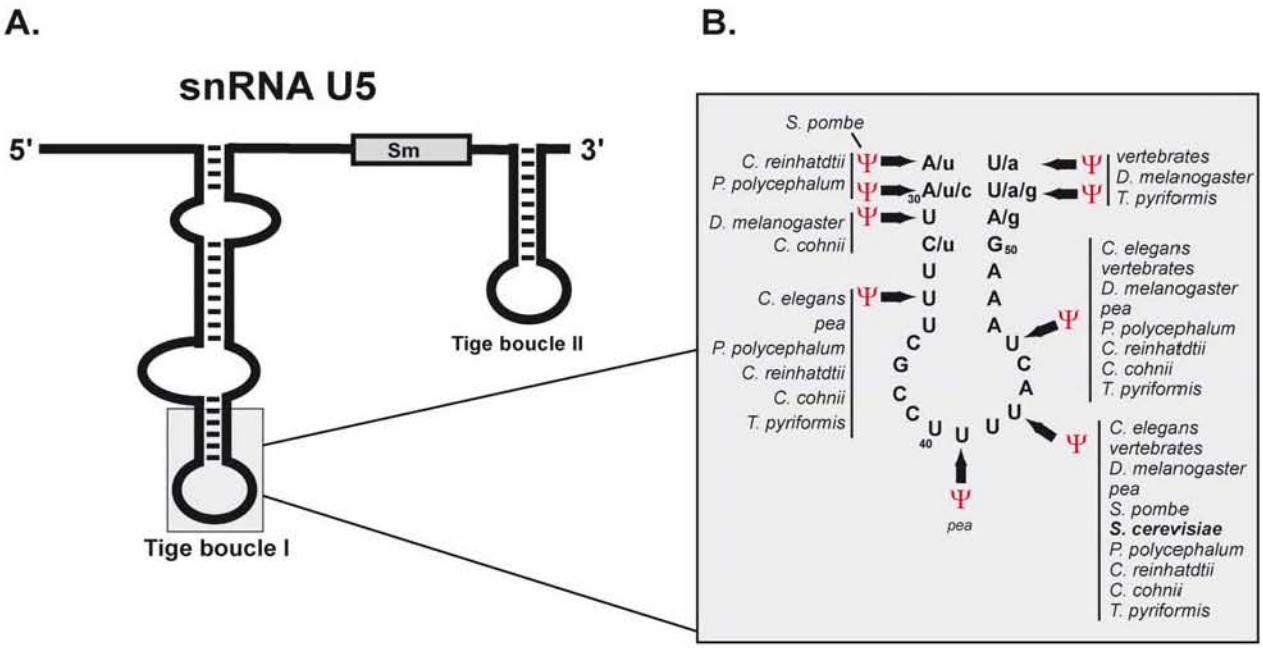


Figure 13 : Comparaison des nucléotides modifiés présents dans le snRNA U5 de différentes espèces. D'après Patton et Padgett, 2005.
A. Structure schématique du snRNA U5. La région où se localisent les résidus Ψ est encadrée en gris.
B. Détails de l'extrémité de la tige boucle I. Les nucléotides sont numérotés d'après la séquence du snRNA U5 humain. Les nucléotides pouvant varier entre les différentes espèces sont indiqués en minuscules.

On observe donc que les modifications des UsnRNA spliceosomaux sont concentrées dans les régions contractant des appariements intermoléculaires lors de la formation des complexes spliceosomaux ou de la catalyse, qu'il s'agisse d'interactions snRNA/ARN pré-messager ou snRNA/snRNA.

3. Rôles des résidus pseudouridines dans les ARN

Comme nous l'avons évoqué précédemment, les résidus Ψ sont présents dans de nombreux types d'ARN. De part leur localisation et leur conservation, il semblerait que certains résidus Ψ jouent un rôle important, bien que parfois mal compris, dans la cellule. Historiquement, c'est dans les années 60 que débutèrent les expérimentations sur la comparaison *in vitro* des propriétés de molécules d'ARNt extraits soit de souches d' *E. coli* sauvages, soit de souches déficientes et/ou mutées pour des ARN:pseudouridine-synthases (souche *his*T (Roth et al., 1966). Cependant, aucune information concernant la fonction de ces résidus n'a pu être déterminée en raison de la faible sensibilité des techniques employées.

Ce n'est que quelques années plus tard, avec l'arrivée de nouvelles méthodes de purification et de séparation des ARNt telles la chromatographie en phase reverse (Kelmers et al., 1965) et l'électrophorèse en gel de polyacrylamide (Ikemura, 1973), que les premiers résultats marquants ont été publiés. Ainsi, il a été démontré que le degré de modification des ARNt pouvait influencer la fidélité et l'efficacité de la traduction (aminoacylation, reconnaissance codon-anticodon…), intervenir indirectement, dans la régulation de la transcription de certains opérons (Dirheimer, 1983; Nishimura, 1983; Kersten, 1984; Bjork et al., 1987; Persson & Bjork, 1993; Björk, 1995, Björk, 1998), mais aussi, intervenir dans l'initiation de la transcription inverse de l'ARN du virus HIV-1 (virus du SIDA) (Isel et al., 1993; Marquet, 1998; Bajji et al., 2002).

Concernant le rôle des résidus Ψ dans les ARNr, il faudra attendre les années 90, avec les premières reconstitutions *in vitro* des sous-unités ribosomales pour comprendre l'importance fonctionnelle des résidus Ψ dans l'assemblage du ribosome et la synthèse protéique. Il apparaît alors que bien que les résidus Ψ ne soient pas indispensables individuellement (Ofengand et al., 2001a), ils confèrent un avantage sélectif (Green & Noller, 1996 ; Khaitovich et al., 1999). Concernant le rôle des résidus Ψ dans les UsnRNA, ce sont les expériences de reconstitution de spliceosome actif (Segault et al., 1995; Yu et al., 1998; Zhao & Yu, 2004) mais également la biologie structurale avec RMN dans les années 2000 (Newby & Greenbaum, 2002) et plus récemment la cristallographie par R-X (Lin & Kielkopf, 2008) qui ont permis de démontrer l'importance des résidus Ψ dans la réaction d'épissage des ARN pré-messagers.

A. Rôles dans les ARN de transfert

Il apparaît que la délétion des ARN:Ψ-synthases responsables de la formation des résidus Ψ dans les ARNt n'affecte pas la réaction d'amino-acylation (Wittenberg & Uhlenbeck, 1985) et n'affecte pas ou peu la viabilité de la cellule (Kammen et al., 1988 ; Simos et al., 1996 ; Becker et al., 1997a ; Lecointe et al., 1998 ; Motorin & Grosjean, 1998). En effet, seule l'absence des résidus pseudouridines aux positions 38, 39 et 40, après délétion des gènes

codant les enzymes TruA et Pus3 chez respectivement, *E. coli* et *S. cerevisiae* entraîne un léger phénotype de croissance (Cortese et al., 1974, Carbone, 1991 ; Tsui et al., 1991 ; Lecointe et al., 1998). Cependant, la fonction cellulaire de ces résidus Ψ reste encore peu comprise (Charette & Gray, 2000).

L'influence des résidus Ψ sur le repliement des ARNt a été mesurée dans de nombreuses études en comparant les caractéristiques structurales d'ARNt produits par transcription *in vitro* et donc dépourvus de modifications post-transcriptionnelles, à celles d'ARNt naturels. Pour cela, différentes méthodes biophysiques ont été utilisées : la fusion thermique, la mesure du rayon de Stokes par gel filtration, le dichroïsme circulaire, l'analyse par sondes chimiques ou enzymatiques de la structure en solution, la diffraction des rayons X, la RMN ou bien encore la spectrofluorimétrie (Perret et al., 1990a ; Chen et al., 1993 ; Derrick & Horowitz, 1993 ; Arnez & Steitz, 1994 ; Yue et al., 1994 ; Serebrov et al., 1998). D'une manière générale, ces études ont révélé que la présence des nucléotides modifiés stabilise les structures secondaires et tertiaires des ARNt. En effet, on constate que plusieurs résidus Ψ présent dans les ARNt sont capables de s'apparier au nucléotide situé en 5' par l'intermédiaire d'une molécule d'eau phénomène appelé « water bridge » (Fig. 14) : le NH supplémentaire du résidu Ψ interagit avec l'oxygène de la molécule d'eau qui, elle-même, interagit avec les oxygènes O_5' du nucléotide en 5' du résidu Ψ et O_3' du résidu Ψ. Cela se traduit par une rigidification locale de la chaîne ribose/phosphate et donc de la structure globale de l'ARNt (Auffinger & Westhof, 1996; 1997 ; Charette & Gray, 2000; Koike et al., 2004).

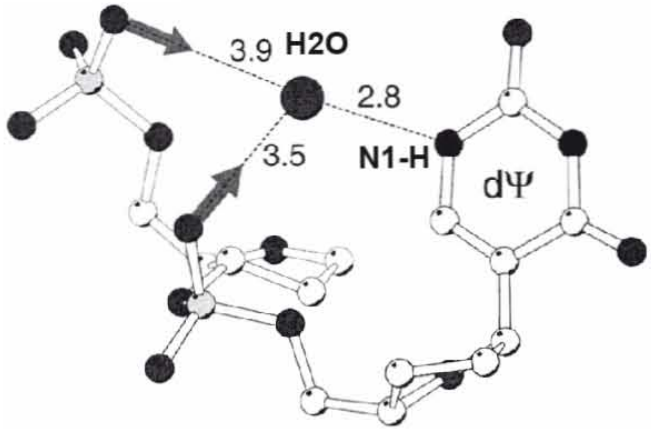


Figure 14 : Modélisation de l'interaction de type « water bridge » dans les ARNt.
La présence d'un résidu Ψ dans une molécule d'ARNt permet la formation, via une molécule d'eau, d'une liaison hydrogène entre la position N_1H de la base et la chaîne phosphate du brin qui comporte ce résidu (Charette & Gray, 2000).

Un autre effet résulterait de la stabilisation de l'ancrage des ions Mg^{2+} au sein des ARNt et d'une réorganisation de la répartition spatiale des molécules d'eau à l'intérieur et à proximité des molécules d'ARNt (Agris, 1996). La présence de résidus Ψ au niveau des jonctions entre

région simple-brin et région double-brin jouerait un rôle dans la stabilisation des hélices d'ARN (Auffinger & Westhof, 1998).

La figure 15 présente la localisation des résidus Ψ dont le rôle sera détaillé dans les paragraphes suivants.

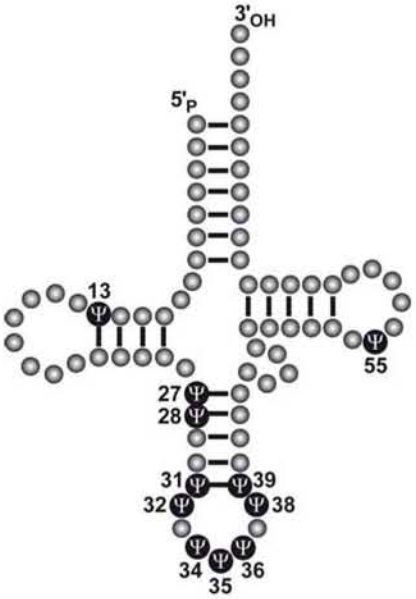


Figure 15 : Localisation des résidus Ψ présent dans les ARNt dont les rôles ont été étudiés. Représentation schématique de la structure en feuille de trèfle d'un ARNt. Les résidus Ψ dont le rôle est discuté sont indiqués en noir.

1. Résidu Ψ en position 13

Bien que ce résidu soit très conservé, son rôle n'est pas établi avec certitude mais il a été proposé que ce résidu renforce la stabilité de l'appariement non Watson-Crick Ψ-G ou Ψ-U entre les positions 13 et 22 du bras D. Notons que le nucléotide en position 22 est souvent impliqué dans un appariement triple avec la base en position 46 de la boucle variable (Auffinger & Westhof, 1998) (Fig. 16).

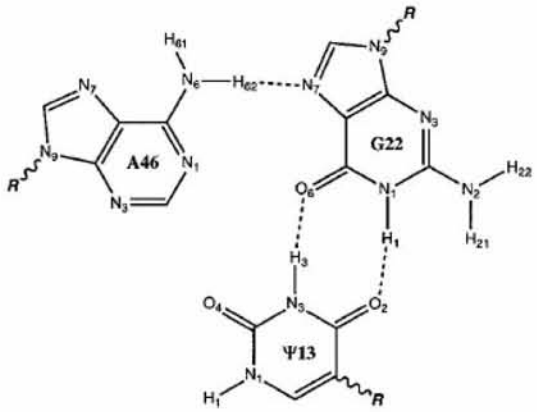


Figure 16 : Représentation de l'appariement triple dans lequel le résidu Ψ13 serait impliqué. (Auffinger & Westhof, 1999).

En utilisant un système permettant la surproduction de variants de l'ARNt^{Glu} *in vivo* chez *E. coli*, il a été démontré que l'absence de résidu Ψ en position 13, cumulée à celle de 2 autres modifications post-transcriptionelles (s^2U et m^2A), n'avait que peu d'effet sur la structure de la boucle anticodon. Il apparaît également que ces modifications ne jouent pas un rôle majeur dans la reconnaissance de l'ARNt par la glutamyl-ARNt synthétase (GluRS) et dans le mécanisme d'aminoacylation (Madore et al., 1999).

2. Résidu Ψ en positions 27 et 28

Dans les ARNt, le résidu Ψ27 s'apparie au résidu 43 pour former une paire de base qui stabilise la région apicale de la tige-boucle de l'anticodon. La fréquence d'apparition d'une paire de bases Ψ-A à ces positions est de l'ordre de 20%, tandis que la paire de bases U-A n'est présente que dans 11% des cas (les paires G-C et C-G représentent 44% des appariements contre 25% de paires G.U, G.Ψ, A-U ou A-Ψ). La pseudouridylation de la position 27 pourrait stabiliser la paire de base U-A énergétiquement moins favorable (Auffinger & Westhof, 1998). Schultz et Yarus (1994) ont étudié l'ARNt su7G36 d'*E. coli*, cet ARN dérive de l'ARNt^{Trp}, sauf qu'il possède un anticodon de séquence CUG, anticodon complémentaire du codon glutamine CAG. Cet ARN a la particularité de pouvoir se comporter en tant qu'ARNt suppresseur. En effet, la formation d'une paire de bases G·U au niveau de la première position du codon permet la lecture d'un codon UAG (ambre) de terminaison. Cette activité suppresseur peut être mesurée *in vivo* précisément, et Schultz et Yarus ont noté que la substitution de la paire de base C27-G43 par des paires de bases A-U, U-A ou G-C avait pour conséquence une diminution de l'activité suppresseur de l'ARNt su7G36. Cette activité peut être restaurée en utilisant une paire de base Ψ-A (Schultz & Yarus, 1994a, 1994b). Ces résultats semblent indiquer que les ARNt qui contiennent la paire de base Ψ27-A43 sont capables de mieux se structurer et par conséquent d'être reconnus plus efficacement par la machinerie traductionnelle.

En étudiant les conséquences de l'absence de la tige-boucle D sur le repliement tridimensionnel de certains ARNt mitochondriaux, l'équipe d'Agris a pu mettre en évidence que la présence du résidus Ψ27 dans l'ARNt^{Met}(CAU) mitochondrial bovin, très proche de l'ARNt mitochondrial humain, augmentait la stabilité thermodynamique de cet ARN en présence d'ions Mg^{2+} (Jones et al., 2006).

Récemment, il a été démontré chez l'homme que certaines maladies rares étaient liées à la présence de mutations dans des ARNt mitochondriaux. Ainsi, la mutation dans l'ARNt^{Ile} mitochondrial du résidu U27 en C (Fig. 17A) a pour conséquence une maladie appelée ophtalmoplégie se caractérisant par une faiblesse des muscles extraoculaires. Ce phénotype s'explique par une forte diminution de l'aminoacylation de cet ARNt et donc de la synthèse protéique. Bien que cela n'ait pas été discuté par les auteurs de cet article, il est important de noter que le résidu U présent en position 27 est naturellement isomérisé en Ψ dans cet ARNt. Il

a été démontré que cette diminution pouvait être compensée par la restauration d'un appariement correct entre les positions 27 et 43 (double mutation U27C/A43G). Ce résultat confirme l'importance, non pas de l'identité du résidu en position 27, mais de l'appariement U(Ψ)27-A43 qui pourrait avoir un rôle essentiel dans la catalyse de la réaction d'aminoacylation (Kelley et al., 2000).

L'absence de résidus Ψ en positions 27 et 28 des ARNt cytoplasmiques et mitochondriaux est constatée chez les patients atteints du syndrome MLASA associant myopathie mitochondriale, acidose lactique et anémie sidéroblastique et se traduisant principalement par de graves dérèglements de la chaîne respiratoire mitochondriale. Il a été démontré qu'un facteur de cette maladie autosomale récessive rare est liée à la présence de la mutation C256T dans l'exon 3 du gène PUS1 humain (Bykhovskaya et al., 2004). En analysant chimiquement les ARNt totaux et l'activité ARN:Ψ-synthase dans des extraits cellulaires réalisés à partir de lignées cellulaires de patients atteints de cette maladie, il a été observé que cette mutation conduit à la production d'une protéine Pus1p inactive. La protéine mutée dans une région très conservée chez toutes les enzymes de la famille (mutation R116W) présente une perte de l'activité ARN:Ψ-synthase au niveau de la position 28 dans l'ARNt^{Ser}(UGA) et des positions 27 et 28 dans l'ARNt^{Lys}(UUU) mitochondrial (Fig. 17B). Cependant, sa localisation cellulaire, principalement dans le noyau mais également dans le cytoplasme et les mitochondries, n'est pas affectée (Patton et al., 2005).

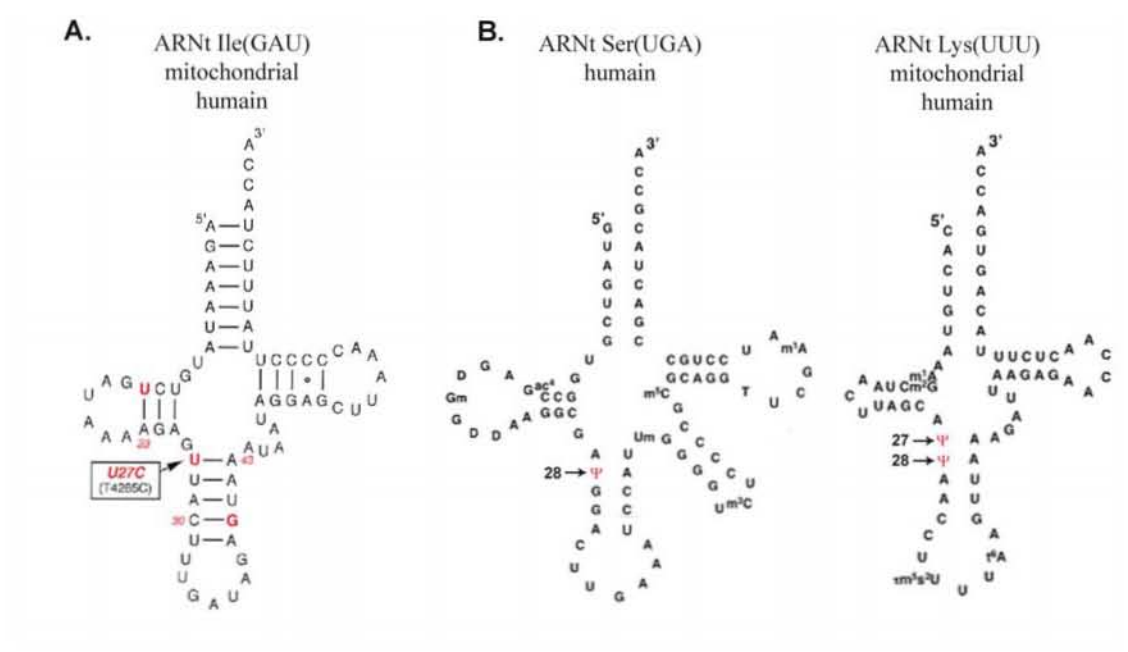


Figure 17 : Structure secondaire des différents ARNt dont les résidus Ψ27/28 sont impliqués dans des pathologies humaines.

A. Structure secondaire de l'ARNt^{Ile} mitochondrial humain. La mutation responsable de l'ophtalmoplégie est encadrée et la position du résidu est précisée. D'après Kelley et al., 2000.

B. **Structure secondaire des ARNt^{Ser} et ARNt^{Lys} mitochondriaux humain. Les résidus Ψ impliqués dans le syndrome MLASA sont indiqués en rouge et leurs positions respectives sont précisées.**

3. Résidus Ψ en positions 31 et 32

Les résidus Ψ31 et Ψ32 sont localisés dans une région charnière entre la tige et la boucle de l'anticodon. Ainsi, le résidu Ψ31 pourrait augmenter la stabilité de la paire de base 31-39 située à l'extrémité de la tige anticodon par l'intermédiaire d'une molécule d'eau, diminuant ainsi la mobilité conformationelle de cette paire de base qui devient thermodynamiquement équivalente à celle des paires de bases G-C ou C-G (Auffinger & Westhof, 1998).

Comme nous l'avons mentionné en début de chapitre, le résidu Ψ32 est également impliqué dans la structuration tertiaire de l'ARNt en stabilisant la tige boucle de l'anticodon grâce à un renforcement du lien entre le squelette nucléotidique et la pseudouridine via une molécule d'eau. L'hydratation induite par la présence de résidus Ψ contribue également à la stabilisation de cette interaction 32-38 (Auffinger & Westhof, 1998).

L'analyse par RMN de la structure de l'ARNt^{Phe} d' *E. coli* montre que l'appariement de type Watson-Crick entre les résidus Ψ32 et A38 est responsable d'une augmentation de la stabilité de la tige boucle de l'anticodon dans une conformation de type tri-loop (Fig. 18B) (Cabello-Villegas & Nikonowicz, 2005). En comparaison, les structures 3D d'ARNt obtenues par radiocristallographie révèlent que la première et la dernière base de la boucle de l'anticodon (positions 32 et 38, respectivement) forment une pseudo-paire de bases avec une seule liaison hydrogène dite bifurquée du type (Ψ32)O_4….N_4(C38) pour les ARNt^{Asp} (Westhof et al., 1985) et l'ARNt^{Phe} de levure (Kim et al., 1974; Robertus et al., 1974) (Fig. 18 A et C). Il a été proposé que cet appariement pourrait créer une transition conformationelle entre la tige et la boucle et faciliter la conformation particulière très contrainte de la boucle anticodon appelée U-Turn (Auffinger & Westhof, 1999).

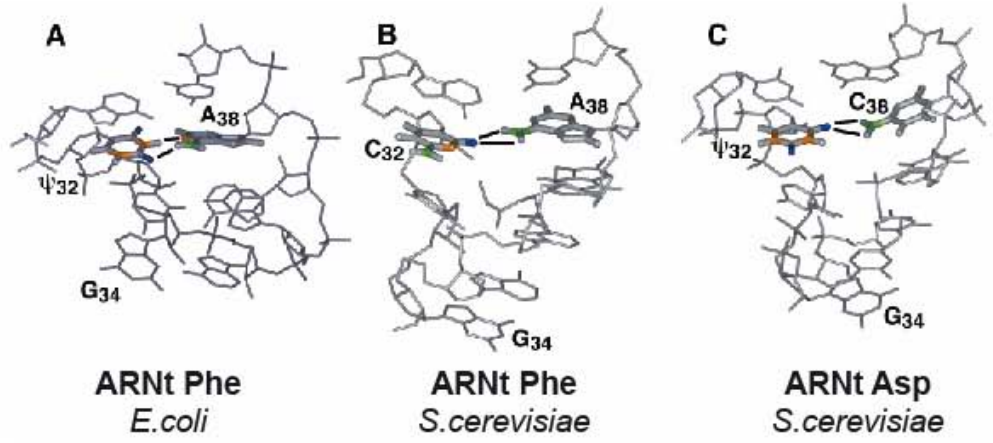


Figure 18 : Structure tridimensionnelle de la boucle anticodon de l'ARNt^{Phe} d'*E. coli* (A), de l'ARNt^{Phe} de *S. cerevisiae* (B) et de l'ARNt^{Asp} de *S. cerevisiae* (C). Les résidus Ψ32 et A38 sont

impliqués dans un appariement de type Watson-Crick (A) tandis que les nucléotides C32-A38 (B) et Ψ32-C38 (C) sont appariés par une liaison hydrogène bifurquée. La boucle anticodon de $ARNt^{Phe}$ d'*E. coli* adopte une conformation de type triloop (A) tandis que les $ARNt^{Phe}$ (B) et $ARNt^{Asp}$ de *S. cerevisiae* adoptent une conformation en U-Turn. (Cabello-Villegas & Nikonowicz, 2005)

Dans toutes les structures d'ARNt connues, cette pseudo-paire de bases allonge l'hélice de la tige anticodon et prend part à la structure de cette région. Elle est, de ce fait, importante pour la fonction de l'ARNt. En effet, une mutation du résidu Ψ32 en C32 induit une perte de la capacité de discrimination des codons par l'$ARNt^{Gly}$ d'*E. coli* (Lustig et al., 1993; Claesson et al., 1995).

4. Résidus Ψ en positions 34 et 36

La reconnaissance spécifique des ARNt par les aminoacyl ARNt synthétases (aaRS) fait intervenir des déterminants de structure primaire, secondaire et tertiaire, à savoir: la boucle anticodon et la base discriminatrice, le bras D et la conformation en L respectivement. De plus, la position 34 de l'anticodon permet d'accommoder la dégénérescence du code génétique à la spécificité de la reconnaissance codon/anticodon, grâce aux appariements dit « Wobble ». Ainsi, la reconnaissance de plusieurs codons par un même ARNt est rendue possible par l'établissement d'interactions non Watson et Crick entre la position 34 et le nucléotide en 3' du codon (Crick, 1966). Or, les nucléotides modifiés peuvent jouer le rôle de déterminants ou d'antidéterminants pour la reconnaissance des ARNt par les aaRS (Björk, 1995). Il a ainsi été suggéré que les groupements C_2=O et N_3-H du résidu Ψ34 couplés à la conformation particulière de la boucle anticodon liée à la présence du résidu Ψ36 sont des éléments d'identité pour l'IleRS (Senger et al., 1997). En effet, le résidu Ψ36 peut servir à stabiliser la structure de la boucle de l'anticodon en interagissant par une liaison du type (Ψ_{36})N1-H….O_2(U_{33}) avec l'uridine universellement conservée en position 33 (Auffinger & Westhof, 1996, 1998).

De plus, il a été démontré que les résidus Ψ modulent la lecture des codons, comme c'est le cas pour deux $ARNt^{Ile}$ de *S. cerevisiae*, l'$ARNt^{Ile}$ majeur possédant un anticodon de séquence IAU et l'ARNt mineur un anticodon ΨAΨ (Szweykowska-Kulinska et al., 1994), pour lesquels une substitution de la base en position 34 de l'anticodon de l'ARNt entraîne une baisse de l'efficacité d'aminoacylation de ces ARNt (Senger et al., 1997). Cet anticodon ΨAΨ a la capacité de reconnaître le codon AUA, mais pas le codon AUG, contrairement à l'anticodon UAU qui, lui, peut interagir avec les 2 codons. Il a été proposé que le résidu Ψ en position 34 inhibe, par sa rigidité, l'appariement du résidu Ψ36 avec la guanosine (Grosjean et al., 1997; Senger et al., 1997; Curran, 1998) (Fig. 19). Le rôle des résidus Ψ en position 34 et 36 serait donc d'assurer la fidélité de la traduction par un contrôle qualité aux 2 étapes clés que sont le chargement de l'ARNt par les aaRS et l'interaction codon/anticodon.

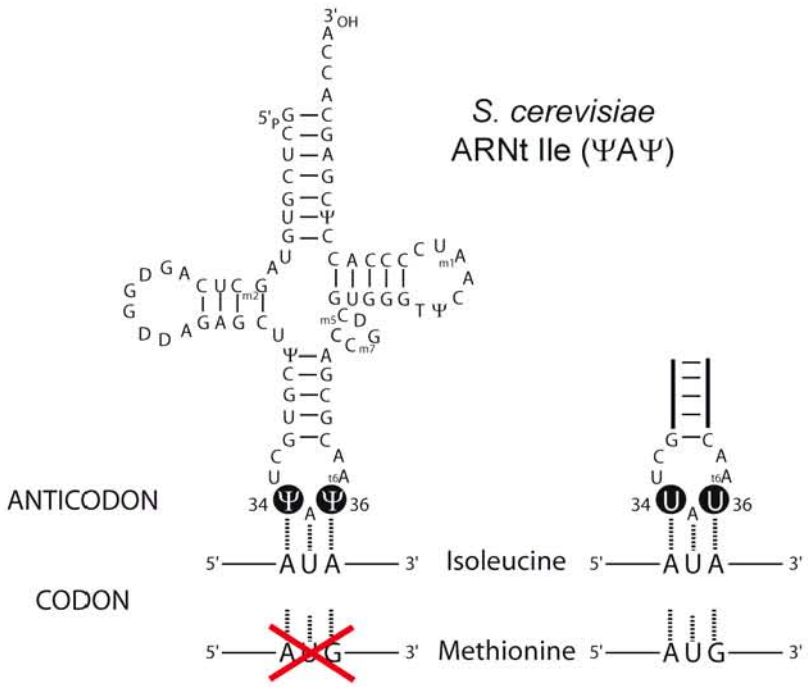


Figure 19 : Rôle des résidus Ψ34 et Ψ36 dans l'ARNt^{Ile}(ΨAΨ) de *S. cerevisiae*.
Les résidus Ψ34 et Ψ36 jouent le rôle d'antidéterminants en empêchant la reconnaissance du codon méthionine. (Senger et al., 1997).

5. Résidu Ψ en position 35

Comme dans le cas des résidus Ψ34 et Ψ36, les rôles du résidu Ψ35 sont en liaison étroite avec sa localisation en position centrale de l'anticodon de l'ARNt^{Tyr}. Ainsi, il a été démontré qu'une substitution du résidu Ψ35 en U35 dans l'ARNt^{Tyr}(GΨA) de levure a pour conséquence le doublement du Km de la réaction d'aminoacylation, indiquant la formation d'interactions spécifiques entre la synthétase et les deux groupements imino de la Ψ35 (Bare & Uhlenbeck, 1986). Cependant, cette faible augmentation du Km de la réaction d'aminoacylation suggère que la fonction première du résidu Ψ35 serait d'augmenter la spécificité des interactions codon-anticodon et non d'augmenter l'efficacité de la reconnaissance ARNt/ARNt aminoacyl synthétase (Auffinger & Westhof, 1998). En effet, il a été montré expérimentalement que la présence d'un résidu Ψ35 augmente l'activité suppresseur de l'ARNt^{Tyr}(UΨA) de *S. cerevisiae* en augmentant la stabilité de l'interaction anticodon(Ψ)-codon(A) (Johnson & Abelson, 1983). De même, dans l'ARNt^{Tyr}(GΨA) de plantes, le résidu Ψ35 est impliqué dans l'efficacité de la reconnaissance des codons de terminaison UAG et UAA, l'anticodon non modifié GUA ne pouvant pas reconnaître ces codons STOP (Zerfass & Beier, 1992). Plus récemment, un résidu Ψ35 a été observé dans l'ARNt^{Asn}(GΨU) mitochondrial des échinodermes. La présence de ce résidu Ψ35 permet la

reconnaissance des codons asparagine AAC et AAU, mais également du codon lysine AAA en tant que codon asparagine, tandis que le variant ayant un U35 ne peut décoder que les codons AAC et AAU (Tomita et al., 1999). Dans ce cas bien précis, la présence de ce résidu Ψ35 est donc un facteur d'altération du code génétique des mitochondries des échinodermes (Tomita et al., 1999).

Par ailleurs, le remplacement d'une uridine par une pseudouridine en position 35 des ARNt conduit à la substitution d'une interaction faible (U35)C5-H…O2'(U33) par une liaison hydrogène forte (Ψ35)N1-H…O2'(U33), ce qui entraîne une stabilisation supplémentaire de la structure de la boucle de l'anticodon qui adopte la conformation en « U-turn ». Cet structuration de l'ARN permet d'augmenter l'efficacité de la traduction (pour revue (Auffinger & Westhof, 1998, 2001) (Fig. 56).

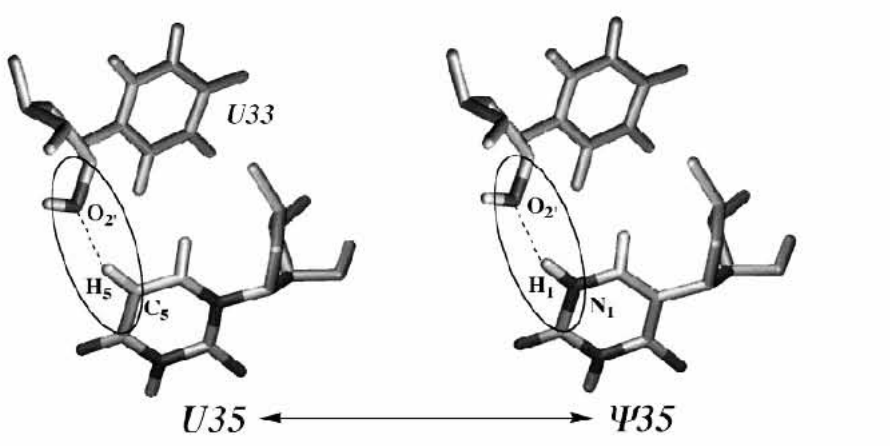


Figure 20 : Stabilisation de la boucle anticodon par le résidu Ψ35.
A gauche: interaction C5-H…O2' déduite des structures cristallines (Comarmond et al., 1986; Westhof et al., 1988; Benas et al., 2000) et des simulations de dynamique moléculaire (Auffinger et al., 1996; Auffinger & Westhof, 1996). A droite : la substitution U35 par Ψ35 renforce la structure de la boucle anticodon en remplaçant l'interaction C-H…O par une liaison N-H…O.

6. Résidus Ψ en positions 38 et 39

La comparaison des structures cristallines de l'ARNt^{Gln} naturel d'*E. coli* et du même ARN produit par transcription *in vitro* a permis de mettre en évidence que la présence des résidus Ψ38 et Ψ39 entraîne une réorganisation des molécules d'eau équivalente à celle observée pour le résidu Ψ32 (Arnez & Steitz, 1994). Comme nous l'avons évoqué précédemment, la formation de ce « water-bridging » se traduit par une rigidification locale de la chaîne ribose/phosphate et donc par une augmentation de la stabilité de la tige-boucle de l'anticodon et par conséquent de la structure globale de l'ARNt (Auffinger & Westhof, 1996).

Le résidu Ψ39 contribue également à la stabilisation de l'appariement 32-38, en augmentant l'empilement des bases de la tige de l'anticodon. Ainsi, la tige anticodon de l'ARNt^{Lys3} humain, contenant un résidu Ψ en position 39, est thermodynamiquement plus stable que la tige correspondante non modifiée (Durant & Davis, 1997, 1999). L'équipe

d'Agris s'est attachée à comprendre le rôle respectif des groupements imino du résidu Ψ39 en étudiant la stabilité thermique, ainsi que la capacité de se fixer au site P du ribosome, d'aptamères ARN synthétiques correspondants à la structure tige-boucle de l'anticodon (ASL) de l'ARNt^{Phe} de *S. cerevisiae* (Yarian et al., 1999). Ils ont ainsi pu mettre en évidence le rôle du groupement imino N3-H du résidu Ψ39 dans la stabilisation de la tige anticodon.

Chez la levure *S. cerevisiae*, le gène *PUS3* code la pseudouridine-synthase Pus3p, responsable de la synthèse des résidus pseudouridines en positions 38 et 39 des ARNt (Lecointe et al., 1998). Il a été démontré que dans une souche de levure délétée du gène *PUS3* (DEG1), l'efficacité de lecture des codons STOP par les ARNt suppresseurs naturels est diminuée (Lecointe et al., 2002). De plus, il apparaît que l'absence de résidu Ψ38 dans l'ARNt^{Gly}(GCC) et de résidu Ψ39 dans l'ARNt^{Leu}(UAG) réduit les événements de décalage du cadre de lecture de type +1 (frameshift +1) aux sites A et P du ribosome (Fig. 21). Il apparaît donc que, chez la levure, la présence des résidus Ψ en positions 38 et 39 du bras anticodon des ARNt favorise la translecture de certains codons par les ARNt non-sens naturels, ainsi que le décalage du cadre de lecture au niveau de séquences dites « glissantes » (Lecointe et al., 2002).

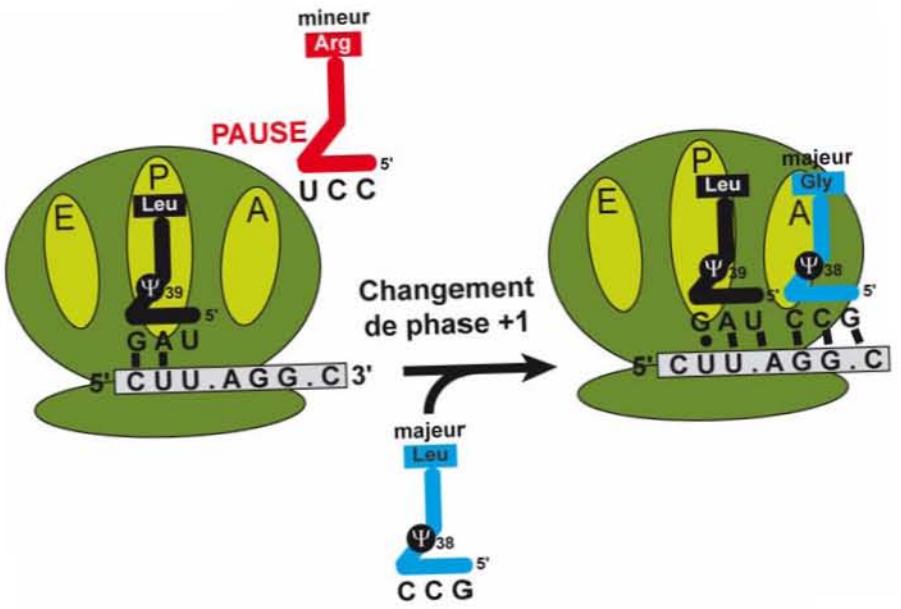


Figure 21 : Mécanisme de décalage du cadre de lecture de type +1 du rétrotransposons Ty1 chez la levure *S. cerevisiae*. (Lecointe et al., 2002).

La séquence responsable du décalage du cadre de lecture est CUU.AGG.C dans laquelle le codon CUU dit « glissant » est reconnu par l'ARNt^{Leu}(UAG) qui porte un résidu Ψ39. Le codon AGG est mal reconnu par l'ARNt^{Arg}(CCU) mineur situé dans le site A du ribosome, ce qui entraîne une pause dans l'élongation de la chaîne polypeptidique. Le codon GGC est alors reconnu par l'ARNt^{Gly}(CCG) majeur qui porte le résidu Ψ38, aboutissant au dipeptide Leu-Gly à la place de Leu-Arg. Ce résultat démontre que le résidu Ψ39 est capable d'influencer l'interaction codon-anticodon.

Plus récemment, il a été possible de sélectionner par la technique du phage display, une série de peptides se liant à un ARNt raccourci correspondant à la tige boucle anticodonde l'ARNt^{Lys3} humain. Il a été démontré, grâce à des spectres d'atténuation de fluorescence, que la présence de plusieurs nucléotides modifiés dont le résidu Ψ39 augmente l'affinité pour ces peptides dont la séquence est très proche de celle retrouvée chez les ARN:Ψ-synthases ou certaines protéines du virus HIV1 (REV, gp160). Cela suggère que le résidu Ψ39 pourrait moduler les interactions avec l'ARNr, l'aminoacyl ARNt synthétase ou d'autres facteurs protéiques lors des étapes de traduction (Eshete et al., 2007). Enfin, les résidus Ψ39 et Ψ32 pourraient, en synergie obligatoire avec d'autres modifications du bras anticodon, contribuer à la stabilisation du motif en U-turn de la boucle anticodon, modulant ainsi la lecture du codon (Tworowska & Nikonowicz, 2006)

7. Résidu Ψ en position 55

Ce résidu est présent dans plus de 90% des ARNt (soit la fréquence de résidu Ψ la plus élevée) et est à l'origine de la dénomination « bras TΨC ». Un défaut de croissance à 42°C des souches d'*E. coli* présentant une délétion du gène TruB est observée (Gutgsell et al., 2000). Néanmoins, l'absence du gène homologue (PUS4) chez la levure *S. cerevisiae* ne donne lieu à aucun phénotype de croissance (Becker et al., 1997a). De plus, l'absence du résidu Ψ en position 55 des ARNt n'affecte ni l'aminoacylation, ni la traduction *in vitro*. Ce résidu semble donc n'intervenir que dans la structuration de l'ARNt (pour revue, (Auffinger & Westhof, 1998). Le groupement imino N3-H du résidu Ψ55 est connu pour interagir avec un atome d'oxygène anionique du résidu 58. De plus, le proton du groupement imino N1-H du résidu Ψ55 interagit avec une molécule d'eau, elle-même en interaction avec des atomes d'oxygène anioniques adjacents (Westhof et al., 1988). Il apparaît donc que le résidu Ψ55 universellement conservé est « bloqué » dans une conformation particulière par les liaisons hydrogène qu'il établit et pourrait ainsi contribuer à la stabilité de la boucle TΨC. Par ailleurs, le résidu Ψ55 est également impliqué dans une interaction avec le nucléotide G17 de la boucle D de l'ARNt^{Asp} de *S. cerevisiae* (Fig. 22).

Tout comme il avait été reporté pour l'interaction entre G18 et Ψ55 dans l'ARNt^{Phe} de *S. cerevisiae* (Quigley, 1976), il a été démontré que l'atome (Ψ55)O4 a une forte tendance à former des liaisons hydrogène "bifurquées" avec les atomes H1 et H21 du G17, plutôt qu'une simple liaison hydrogène avec l'atome H1.

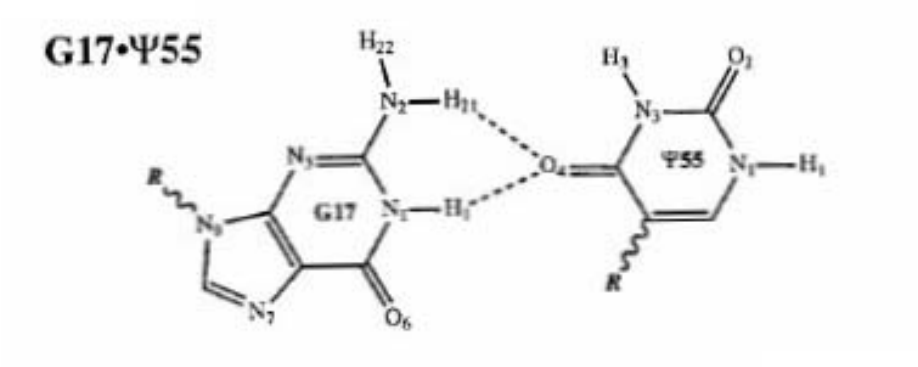


Figure 22 : Représentation schématique de l'interaction existant entre les résidus G17 et Ψ55 de l'ARNt^{Asp} de *S. cerevisiae*. D'après (Auffinger, 1999).

Ainsi, les résidus Ψ sont impliqués dans de nombreuses fonctions des ARNt, essentiellement en stabilisant des motifs structuraux spécifiques liés à la capacité intrinsèque de ce résidu à former un plus grand nombre de liaisons hydrogène que le résidu U avec d'autres nucléotides de l'ARNt par l'intermédiaire d'une molécule d'eau..

Ainsi les résidus Ψ sont impliqués dans de nombreuses fonctions des ARNt et contribuent également à préserver la fidélité du processus de traduction.

B. Rôles dans les ARN ribosomiques

En comparaison avec les ARNt, les rôles des résidus Ψ dans les ARNr sont encore mal compris même si toutes les études menées s'accordent sur le fait que leur localisation dans des régions fonctionnelles de l'ARNr suggère leur importance (Branlant et al., 1981a; Branlant et al., 1981b); Ofengand, 2002).

Une première approche pour aborder l'étude de la fonction des modifications des ARNr a consisté à analyser les caractéristiques de ribosomes assemblés *in vitro* à partir d'ARNr non modifiés ou partiellement modifiés. Il s'est avéré que les grandes sous-unités de trois espèces bactériennes, reconstituées à partir de transcrits dépourvus de modifications, catalysent encore la synthèse de liaisons peptidiques. Cependant, ces ribosomes présentent une activité moindre en comparaison des sous-unités reconstituées avec des ARNr naturels (Green & Noller, 1996 ; Khaitovich et al., 1999). En dépit du nombre parfois élevé de résidus Ψ présents dans un ARNr, il semble qu'ils ne soient pas indispensables (Ofengand et al., 2001a). Par conséquent, les résidus Ψ confèreraient un avantage sélectif, mais ils ne seraient pas essentiels. Cette hypothèse est en accord avec le fait que certaines espèces présentent peu de résidus Ψ, principalement des espèces bactériennes et d'archaea, et que ces positions ne sont pas strictement conservées dans les trois domaines du vivant.

Chez *E. coli*, l'approche visant à montrer une importance fonctionnelle des résidus Ψ a consisté dans un premier temps à interrompre individuellement les gènes codant les ARN:Ψ-synthases responsables de la formation de résidus Ψ dans les ARNr, et à mesurer d'éventuels

effets sur la croissance. Une série d'études analogues par délétions des snoRNA à boîtes H/ACA de *S. cerevisiae* a également été réalisée. La délétion de 6 des 7 ARNr:Ψ-synthases présentes chez *E. coli* n'a aucun effet détectable sur la croissance cellulaire, ceci pour des tests effectués à plusieurs températures (Conrad et al., 1998 ; Conrad et al., 1999 ; Raychaudhuri et al., 1999 ; Del Campo et al., 2001). Seule la perte de l'ARN:Ψ-synthase RluD, responsable de la formation des résidus Ψ1911, Ψ1915 et Ψ1917 dans l'ARNr 23S, provoque un défaut de la croissance (Raychaudhuri et al., 1998 ; Gutgsell et al., 2001). Cependant, dans ce cas, il a été prouvé que ce défaut de croissance peut être corrigé par l'apport d'un plasmide portant le gène codant RluD muté au niveau de l'acide aspartique catalytique. Ceci suggère que la protéine RluD est essentielle pour une croissance normale, mais pas son activité ARN:Ψ-synthase (Gutgsell et al., 2001). Des études complémentaires ont montré que le défaut de croissance observé en absence de RluD est une conséquence directe d'un mauvais assemblage de la sous-unité 50S du ribosome, suggérant une implication de RluD dans le processus d'assemblage du ribosome, peut-être en tant que protéine chaperonne (Gutgsell et al., 2001).

Des délétions simultanées de plusieurs gènes ont été réalisées pour empêcher la formation de résidus Ψ appartenant à une même région. Ainsi, la délétion des gènes codant les protéines RluB, RluC, RluE et RluF qui abolit la formation de 6 résidus Ψ situés à proximité du centre peptidyl transfèrase (PTC) conduit à l'observation de certains défauts métaboliques (ralentissement de la croissance, variation du profil des polysomes) (pour revue, (Ofengand, 2002). Il a donc été suggéré que les résidus Ψ n'ont pas d'effet indépendant mais contribuent chacun, partiellement et de manière coopérative, à la stabilisation d'une conformation particulière de l'ARNr. De même, chez *S. cerevisiae*, la délétion individuelle de 5 snoRNA à boîtes H/ACA guide de pseudourydilation (snR10, snR34, snR37, snR42 et snR46) responsables de la formation de 6 résidus Ψ du centre PTC n'a pas d'effet sur la croissance (Fig. 23) (Ofengand & Fournier, 1998), alors que la délétion combinée des 5 guides entraîne une réduction de la vitesse de croissance et du taux de synthèse protéique, une hypersensibilité aux antibiotiques et des anomalies dans la structuration de l'ARNr de la grande sous-unité (King et al., 2003). Cependant, les analyses de l'effet de la perte d'expression de l'ARN snR10, responsable de la modification de la formation du résidu Ψ en position 2919, sont compliquées car ce snoRNA est également impliqué dans la maturation de l'ARNr de la petite sous-unité (Tollervey, 1987). Pour distinguer ces deux effets, un nucléotide de l'élément antisens 3' de snR10 a été muté pour inhiber l'appariement du guide à sa cible (snR10ΔC) (King et al., 2003). Il a ainsi été observé que l'absence de formation du résidu Ψ2919 aurait pour conséquences un profil de polysomes altéré et un défaut de synthèse protéique sans pour autant altérer la vitesse de croissance cellulaire alors que l'absence d'expression de snR10 entraîne une diminution plus forte de la synthèse protéique. Ainsi, la diminution de la croissance et l'hypersensibilité aux antibiotiques sont associées à la perte de snR10 et non à la perte du résidu Ψ2919. Il est à noter que, dans la souche pour laquelle les 5 ARN guides ne

sont pas exprimés, les défauts de synthèse protéique sont encore plus marqués, ce qui suggère une influence de la présence d'autres snoRNP dans la formation des ribosomes actifs.

L'élimination globale des résidus Ψ des ARNr de *S. cerevisiae* a été réalisée par des mutations de la protéine CBF5 qui est la protéine qui porte l'activité catalytique dans le système de pseudourydilation utilisant des guides (Zebarjadian et al., 1999). Cependant, l'interprétation des résultats est rendue difficile par l'existence d'un grand nombre d'interactions entre CBF5 et des protéines participant à la ségrégation des chromosomes, des kinases spécifiques de la méiose, des facteurs transcriptionnels de l'ARN pol. I, hTR le composant ARN de la télomérase et des snoRNA à boîtes H/ACA impliqués dans la maturation des ARNr (Charette & Gray, 2000). Il n'est donc pas surprenant d'observer que l'absence d'expression de CBF5 est létale pour la cellule (Jiang et al., 1993 ; Lafontaine et al., 1998a). Des souches produisant une protéine CBF5 mutée au niveau de l'aspartate catalytique, ce qui abolit la formation de l'ensemble des résidus Ψ dans les ARNr, survivent mais présentent une sévère inhibition de croissance (Zebarjadian et al., 1999). Un rôle possible des résidus Ψ dans la structure du site peptidyl transferase (PTC) a été analysé par modification chimique au DMS de la région de l'ARNr 23S de *S. cerevisiae* contenue dans ce site et provenant de souches n'exprimant pas les 5 snoRNA responsables des 6 résidus Ψ situés dans cette région (Fig. 23). Des sites additionnels de réactivité au DMS apparaissent dans la souche mutante démontrant ainsi que les résidus Ψ du PTC influencent la conformation de cette région fonctionnelle de l'ARNr 23S (King et al., 2003). L'étude de la structure 3D de la grande sous-unité ribosomique (Ban et al., 2000) révèle qu'aucun de ces 6 résidus Ψ n'est localisé à moins de 13 Å du site de formation de la liaison peptidique et donc à priori aucun de ces résidus ne peut avoir une influence directe sur la formation de cette liaison (Ban et al., 2000).

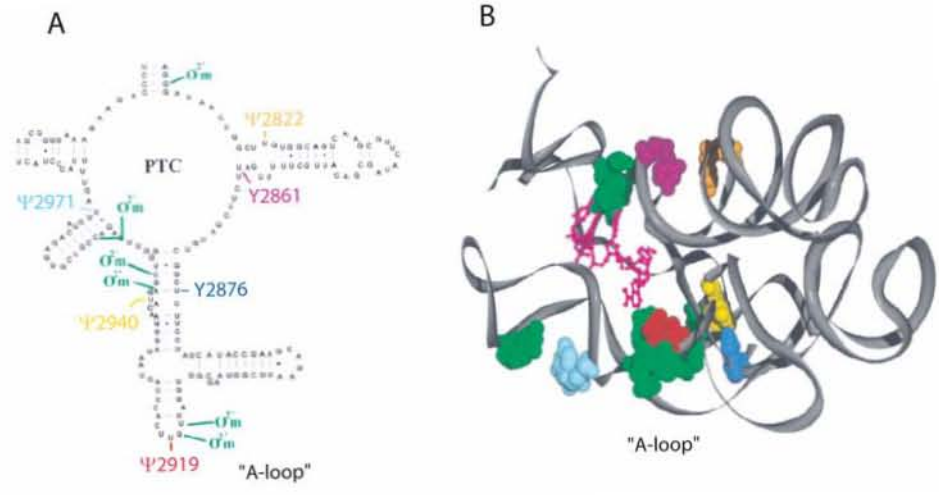


Figure 23 : Le centre peptidyl transférase (PTC) de l'ARNr 23S de *S. cerevisiae* contient 6 résidus pseudouridine. (King et al., 2003).
A. Structure secondaire de la région centrale du PTC de l'ARN 23S de *S. cerevisiae*. Les résidus 2'-O-méthylés sont représentés en vert. Les résidus pseudouridines sont indiqués par des couleurs différentes.

B. Prédiction de la localisation des nucléotides modifiés au sein de la structure 3D du PTC. Les modifications post-transcriptionnelles observées chez *S. cerevisiae* ont été replacées sur la structure du PTC de *H. marismortui*, complexé à de la CCdA-p-puromycine (Yarus analog, en magenta) (Nissen et al., 2000). Le code couleur est le même qu'en **A.**

Cependant, du fait de leur localisation à proximité des sites A et P, les résidus Ψ pourraient agir sur la fixation des ARNt, leur sélection, leur translocation et leur sortie du ribosome. Il a été suggéré, d'après la structure 3D du ribosome, qu'une interaction entre le résidu Ψ2919 et le trinucléotide CCA à l'extrémité de l'ARNt ait lieu (Fig. 24A) (King et al., 2003). Les résidus adjacents en 5', G2918 et U2920 (2588 et 2590 chez *H. marismortui*) forment des appariements de type Watson-Crick avec le résidu C75 et C74 du CCA de l'ARNt fixé au site A (Kim & Green, 1999 ; Ban et al., 2000). Le résidu Ψ2919 pourrait stabiliser ces appariements en augmentant le potentiel de formation de liaisons hydrogène. Par ailleurs, le résidu Ψ2861, situé à proximité du site P, pourrait de la même manière intervenir lors de la fixation de l'ARNt à ce site (Fig. 24B).

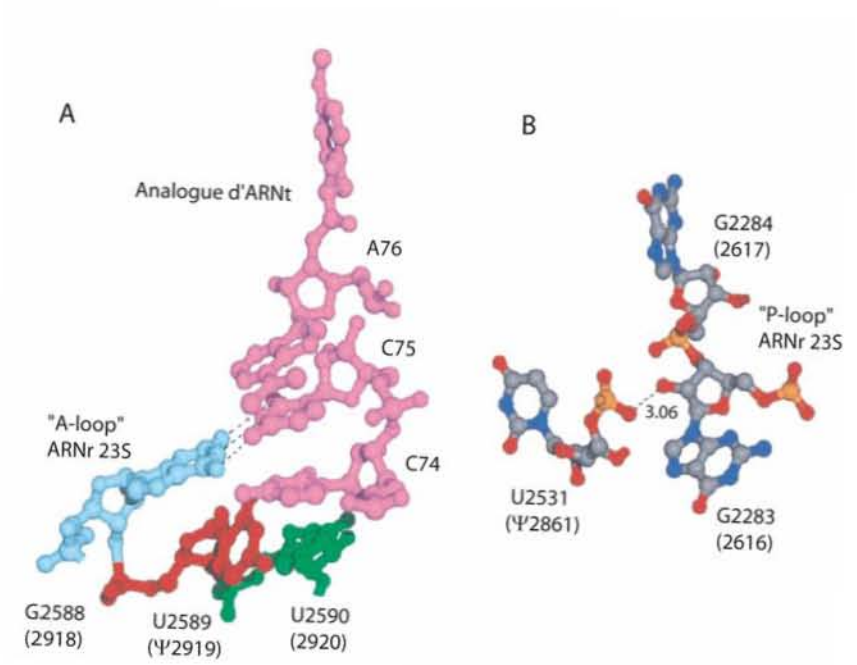


Figure 24: Deux résidus Ψ du site PTC sont localisés à proximité des sites A et P, sites d'interaction avec les ARNt (King et al., 2003).
A. Le résidu Ψ2919 pourrait être impliqué dans la stabilisation de l'interaction entre les ARNt et le site A du ribosome. Les nucléotides de l'ARNr 23S qui interagissent avec l'extrémité 3' de l'ARNt fixé au site A sont représentés en utilisant la structure et la numérotation du ribosome de *H. marismortui*. La numérotation de *S. cerevisiae* est indiquée entre parenthèses. Le résidu U2589 (en rouge) correspond au résidu Ψ2919. L'analogue de l'ARNt fixé au site A est représenté en magenta.
B. Le résidu Ψ2861 de *S. cerevisiae* pourrait être impliqué dans la stabilisation de l'interaction entre les ARNt et le site P du ribosome. L'oxygène du phosphate du résidu Ψ2861 pourrait former une liaison hydrogène avec le ribose du résidu G2283 appartenant au site P (Nissen et al., 2000).

De même, il est envisageable que le résidu Ψ2492 (numérotation de *H. marismortui*) (Ban et al., 2000) conservé dans la plupart des espèces étudiées et localisé 6 nt en aval du résidu A2486 qui serait lui-même le nucléotide agissant directement dans la formation de la liaison peptidique, pourrait jouer un rôle dans la stabilisation de la conformation la plus efficace de ce dernier résidu (Nissen et al., 2000; Ofengand et al., 2001b). Une autre illustration est la présence de deux résidus Ψ localisés dans l'hélice 69 de l'ARN de la grande sous-unité ribosomale et conservés entre espèces éloignées. Ces deux résidus modifiés ne sont pas essentiels mais confèrent un avantage dans des expériences de compétition (Badis et al., 2003). Ils seraient importants pour stabiliser une conformation spécifique de la boucle de l'hélice 69 qui interagit avec des résidus de la petite sous-unité et des résidus de l'ARNt fixé au site A (Meroueh et al., 2000 ; Sumita et al., 2005). Cette région de l'ARN de la grande sous-unité comporte les résidus Ψ les mieux conservés dans les trois domaines du vivant.

Des études récentes menées chez l'homme ont également permis de mettre en évidence l'implication des résidus Ψ dans le développement d'une maladie génétique rare appelée dyskératose congénitale qui se traduit par une pigmentation anormale de la peau, une dystrophie des ongles et une leucoplasie des muqueuses (apparition de plaques blanchâtres au niveau des muqueuses buccales ou génitales) (Zinesser, 1906 ; Engmann, 1926; Cole et al., 1930). Les tissus touchés sont ceux dont les cellules se renouvellent rapidement (peau, ongles, cheveux, intestin, moelle osseuse). La principale anomalie non-cutanée est une aplasie médullaire progressive (défaillance au niveau de la moelle épinière), qui se produit chez 80% des patients et représente la source majeure de mort prématurée (Dokal, 1996 ; Knight et al., 1998). La transmission de la dyskératose congénitale peut se faire de façon autosomale (récessive ou dominante) ou peut être liée au chromosome X (forme la plus fréquente). Cette dernière est causée par des mutations du gène DKC1, qui code la dyskérine. Il a été démontré qu'il existe 2 facteurs principaux qui expliquent cette maladie. D'un côté une dérégulation des fonctions du ribosome comprenant une réduction de 10 à 40% du taux de pseudouridylation des ARNr, l'accumulation d'intermédiaires anormaux de maturation de l'ARNr (45S, 41S et 32S) et une hypersensibilité aux inhibiteurs traductionnels (Lee et al., 1998; Rudolph et al., 1999; Ruggero et al., 2003). De l'autre, un défaut d'activité de la télomérase qui est un complexe ribonucléoprotéique, actif dans les cellules germinales ou à fort taux de renouvellement, permet de maintenir la longueur des séquences télomériques situées à l'extrémité des chromosomes (Mitchell et al., 1999; Vulliamy et al., 2001).

La restriction des résidus Ψ au niveau des régions fonctionnelles et le fait que deux systèmes de pseudouridylation aient été maintenus au cours de l'évolution suggèrent que les résidus Ψ jouent un rôle important dans le fonctionnement des ribosomes, certainement en optimisant chaque aspect de la traduction. La fonction principale des résidus Ψ présents dans les ARNr pourrait être de jouer le rôle d'une « colle moléculaire » qui fortifie une conformation pas suffisamment rigide (Ofengand, 2002).

C. Rôle dans les UsnRNA

Comme nous l'avons mentionné précédemment, les résidus Ψ sont principalement localisés dans les régions fonctionnelles des UsnRNA et stabilisent les interactions ARN-ARN et/ou ARN/protéines.

1. Interactions UsnRNA/ARN pré-messager

1.1. Cas du snRNA U1

La présence de 2 résidus Ψ en positions 5 et 6 du snRNA U1 est quasiment universelle (Massenet et al., 1998). Ils ont été détectés chez les vertébrés, la drosophile et chez *S. cerevisiae*. Les résidus en positions 5 et 6 sont aussi fréquemment modifiés chez les plantes (Ψm5Um6 chez l'algue *Chlorella saccharophila)* (Kiss et al., 1988). Chez *S. pombe*, un seul résidu Ψ a été détecté en position 3, mais il n'est pas sûr que la recherche ait été exhaustive (Gu et al., 1996). Par contre, aucun résidu n'est détecté dans le snRNA U11 de vertébrés, dont la séquence est différente et qui s'apparie à des séquences 5' différentes (Massenet, 1999, pour revue, Massenet, 1998).

Le snRNA U1 contient donc 2 résidus Ψ qui font partie d'un segment impliqué dans une interaction de bases avec le site 5' d'épissage de l'ARN pré-messager (Fig. 25). Une étude structurale par RMN a démontré que l'appariement entre le snRNA U1 humain et la séquence des sites 5' d'épissage est stabilisé par la présence des résidus Ψ en positions 5 et 6 (Hall, 1991).

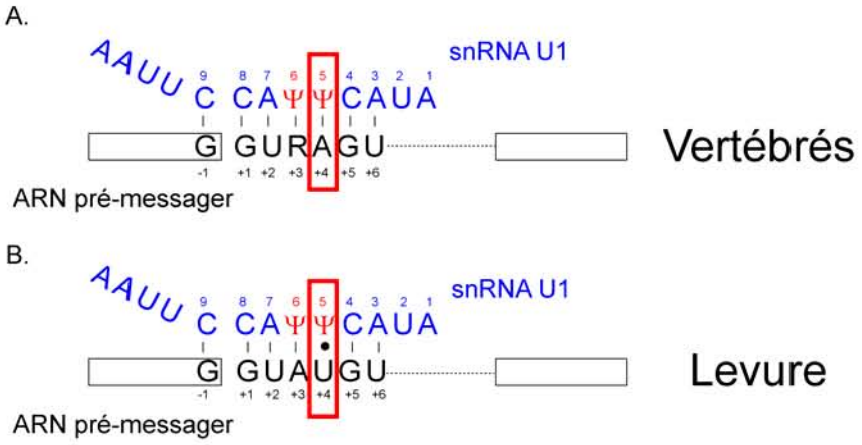


Figure 25 : Représentation de l'interaction du site 5' d'épissage de l'ARN pré-messager avec le snRNA U1.

A. Cas de l'ARN pré-messager de vertébrés. (D'après les revues, Madhani, 1994 et Wu, 1999). B. Cas de l'ARN pré-messager de la levure *S. cerevisiae*. (Massenet et al., 1999).

La comparaison de l'appariement snRNA U1-ARN pré-messager entre les vertébrés et la levure *S. cerevisiae* révèle que la plupart des introns de levure présentent en 5' une séquence (GUAUGU) qui n'est que partiellement complémentaire à la région 5' du snRNA U1. Ainsi, lors de la première étape de la réaction d'épissage, le résidu Ψ5 du snRNA U1 est positionné

en face du résidu U+4 de l'intron (Massenet et al., 1999) (Fig. 25). Il a été suggéré que la présence de ce mésappariement induit la stabilisation de l'interaction U1-ARN pré-messager par le biais d'une protéine capable de reconnaître la structure de l'ARN contenant le mésappariement Ψ5-U+4 (Libri et al., 2002).

La sélection d'un appariement imparfait snRNP U1-site 5' d'épissage chez *S. cerevisiae* favorise sans doute la transition de l'interaction snRNP U1-site 5' vers l'interaction snRNP U6-site5' lors de la réaction d'épissage. Il a été récemment démontré que le fait d'augmenter artificiellement la longueur de l'appariement entre le snRNA U1 et le site 5' d'épissage inhibe la réaction d'épissage à basse température (Staley & Guthrie, 1999). Il est admis actuellement que l'interaction de la snRNP U1 avec le site 5' d'épissage est stabilisée par une protéine de la snRNP U1. L'activité de cette protéine serait modulée par phosphorylation ou par l'action d'autres protéines, au cours de la réaction d'épissage, ceci afin de favoriser soit l'association, soit la dissociation de la snRNP U1 au niveau du site 5' d'épissage (Libri et al., 2002). L'ensemble de cette régulation reposerait donc sur la présence du résidu Ψ5 dans le snRNA U1. Chez les vertébrés, la séquence des sites 5' d'épissage est beaucoup moins conservée que chez *S. cerevisiae*, la position des mésappariements de bases est variable et les mécanismes de stabilisation de l'interaction snRNP U1-site 5' d'épissage reposent sur l'intervention des protéines SR, qui ne sont pas présentes chez *S. cerevisiae*.

1.2. Cas du snRNA U2

En dépit du grand nombre de résidus Ψ que l'on retrouve chez les vertébrés, seuls 2 de ces résidus sont conservés dans tous les snRNA U2 étudiés (et snRNA U12, là encore, nécessité intro spliceosome mineur dans partie 2) : la position 34 (35 chez *S. cerevisiae*) qui est localisée au niveau de la paire de bases qui jouxte le résidu A responsable de la première attaque nucléophile et la position 43 (44 chez *S. cerevisiae*) qui se situe en amont de la boîte de branchement (Fig. 26).

L'importance des résidus Ψ a pu être mise en évidence par des expériences de reconstitution de spliceosome et analyse de l'efficacité d'épissage *in vitro*. Il a en effet été observé que les modifications post-transcriptionnelles sont essentielles à la reconstitution d'un spliceosome actif, à partir d'un extrait nucléaire de cellules HeLa où le snRNA U2 endogène a été éliminé (Segault, 1995; Yu, 1998; Zhao, 2004).

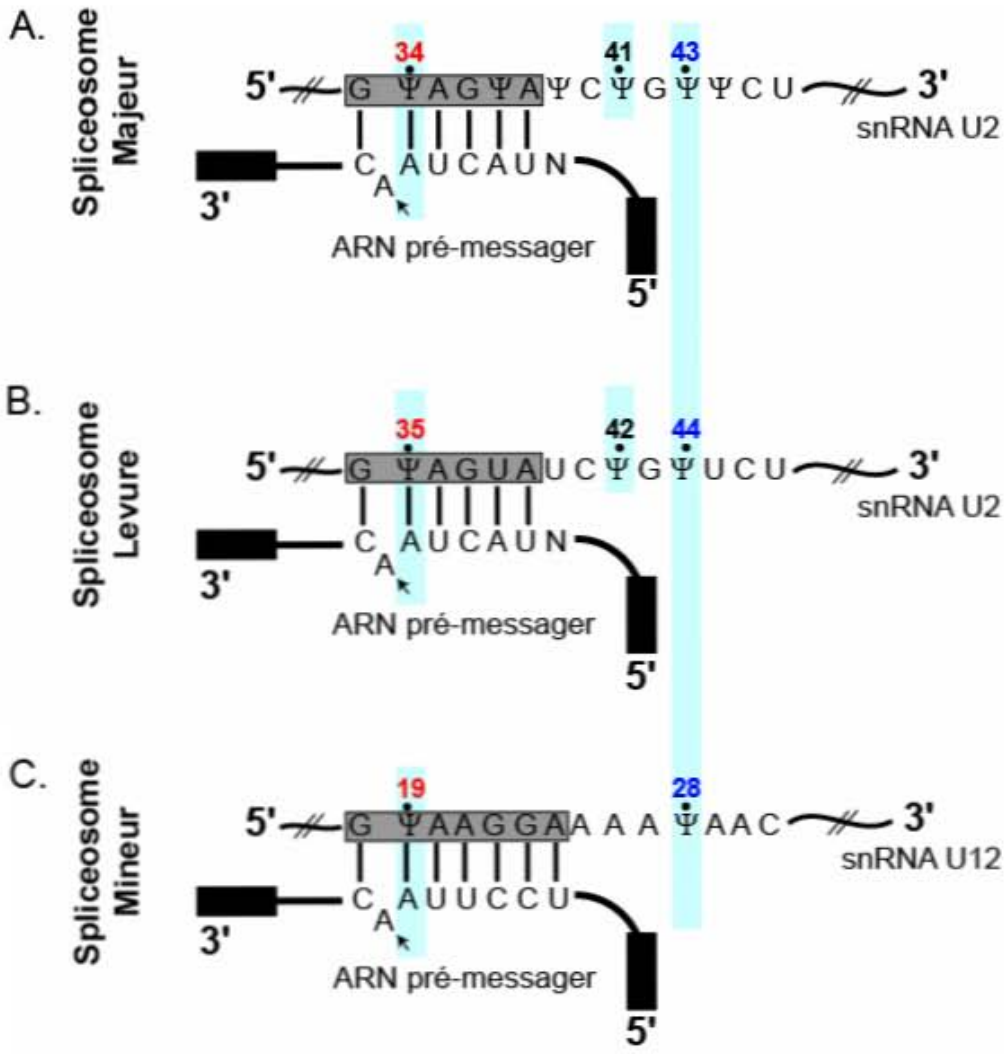


Figure 26 : Interactions entre les snRNA U2 de vertébrés (A), de levure (B) ou le snRNA U12 (C) et la boîte de branchement des ARN pré-messagers lors de l'épissage.
Tous les résidus Ψ impliqués dans des interactions avec l'ARN pré-messager sont indiqués. Les résidus Ψ équivalents sont mentionnés par des flèches tandis que les résidus Ψ conservés dans les trois systèmes sont représentés en rouge et en bleu. La boîte de branchement de l'ARN pré-messager est entourée en gris. (Massenet et al., 1998).

L'étude par RMN de la structure 3D à haute résolution d'un duplex formé par interaction de deux oligoribonucléotides, mimant respectivement l'un la boîte de branchement et l'autre, la région du snRNA U2 interagissant avec cette dernière a permis de démontrer le résidu Ψ impliqué dans la paire de base qui jouxte l'adénosine du point de branchement joue un rôle majeur dans son positionnement (Newby & Greenbaum, 2001, 2002). En effet, il permet son maintien dans une conformation extrahélicale qui doit favoriser la première réaction de transestérification de la réaction d'épissage (Fig. 27).

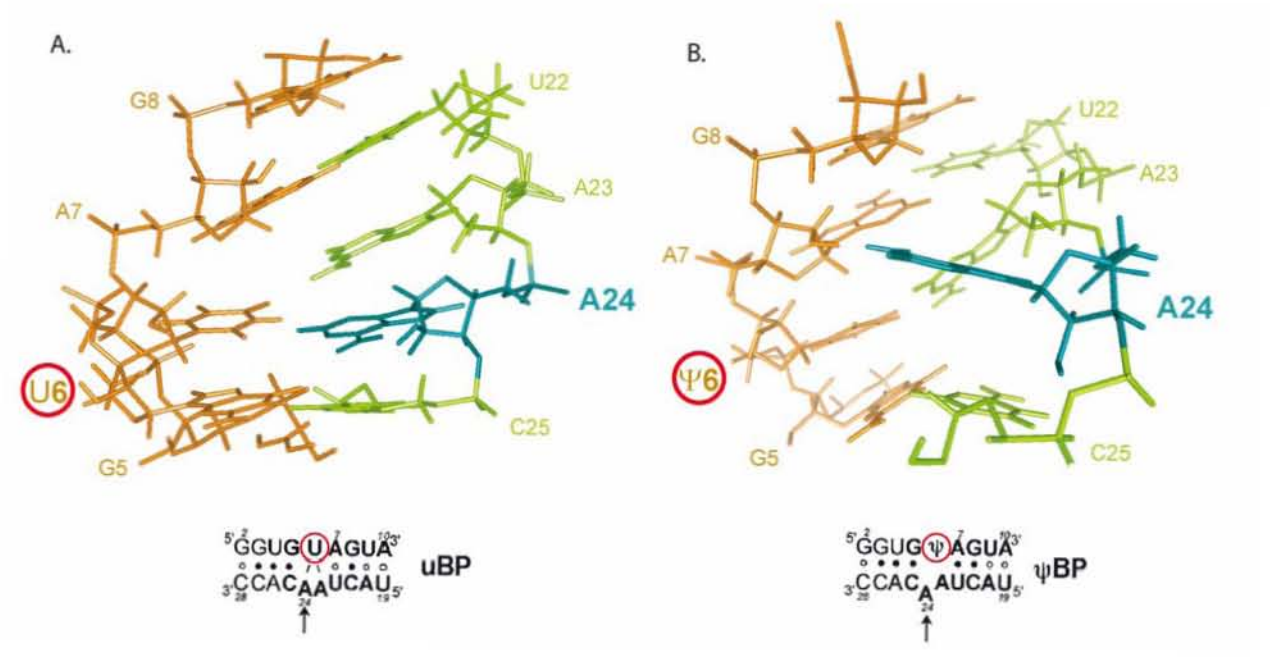


Figure 27 : Structure 3D établie par RMN pour l'hybride formé entre un oligoribonucléotide mimant la région du point de branchement et un oligoribonucléotide mimant le snRNA U2.
Dans le panneau A, l'oligoribonucléotide mimant le snRNA U2 contient un résidu U en position 6, tandis que dans le panneau B cet oligoribonucléotide contient un résidu Ψ6.
Le brin correspondant au snRNA U2 est représenté en marron, celui correspondant à la boîte de branchement est représenté en vert, l'adénosine de la boîte de branchement est représentée en bleu et indiquée par une flèche. (Newby & Greenbaum, 2002).

L'analyse détaillée de la structure 3D formée en présence d'un résidu Ψ6 révèle une distance de 3,6 +/- 0,1 Å entre le groupement imino apporté par le résidu Ψ et l'oxygène du phosphate en 5' de ce même résidu, cette distance est tout à fait compatible avec la formation d'une liaison hydrogène médiée par une molécule d'eau (Newby & Greenbaum, 2002). Par ailleurs, l'analyse des spectres RMN indique que le proton du groupement imino est protégé des échanges avec le solvant et donc sûrement impliqué dans la formation d'une liaison hydrogène (Newby & Greenbaum, 2002). Il apparaît donc qu'au moins une partie de la différence conformationnelle observée entre UBP et ΨBP est due à des interactions polaires supplémentaires provenant du résidu Ψ (Newby & Greenbaum, 2002).

Une étude récente a permis de résoudre la structure 3D par radiocristallographie de plusieurs duplexes ARN mimant l'appariement entre le snRNA U2 et l'ARN pré-messager au niveau de la boîte de branchement. Ainsi, il a pu être démontré que l'adénosine responsable de l'attaque nucléophile lors de la réaction d'épissage possède bien une confirmation extrahélicale lorsque la séquence de la boîte de branchement est de type levure raccourcie (BPS2 = ACUAAC) ou de type mammifère (BPS3 = UACUGAC) mais pas de type levure canonique (BPS1 = UACUAAC) (Fig. 28A). De manière intéressante, il apparaît que le résidu Ψ présent dans le snRNA U2 favorise l'interaction entre celui-ci et le résidu localisé en amont

par empilement de base et par interaction de type simple (BPS2 et BPS3) ou double water-bridge (BPS1) ce qui augmente la stabilité du duplexe (Fig. 28B).

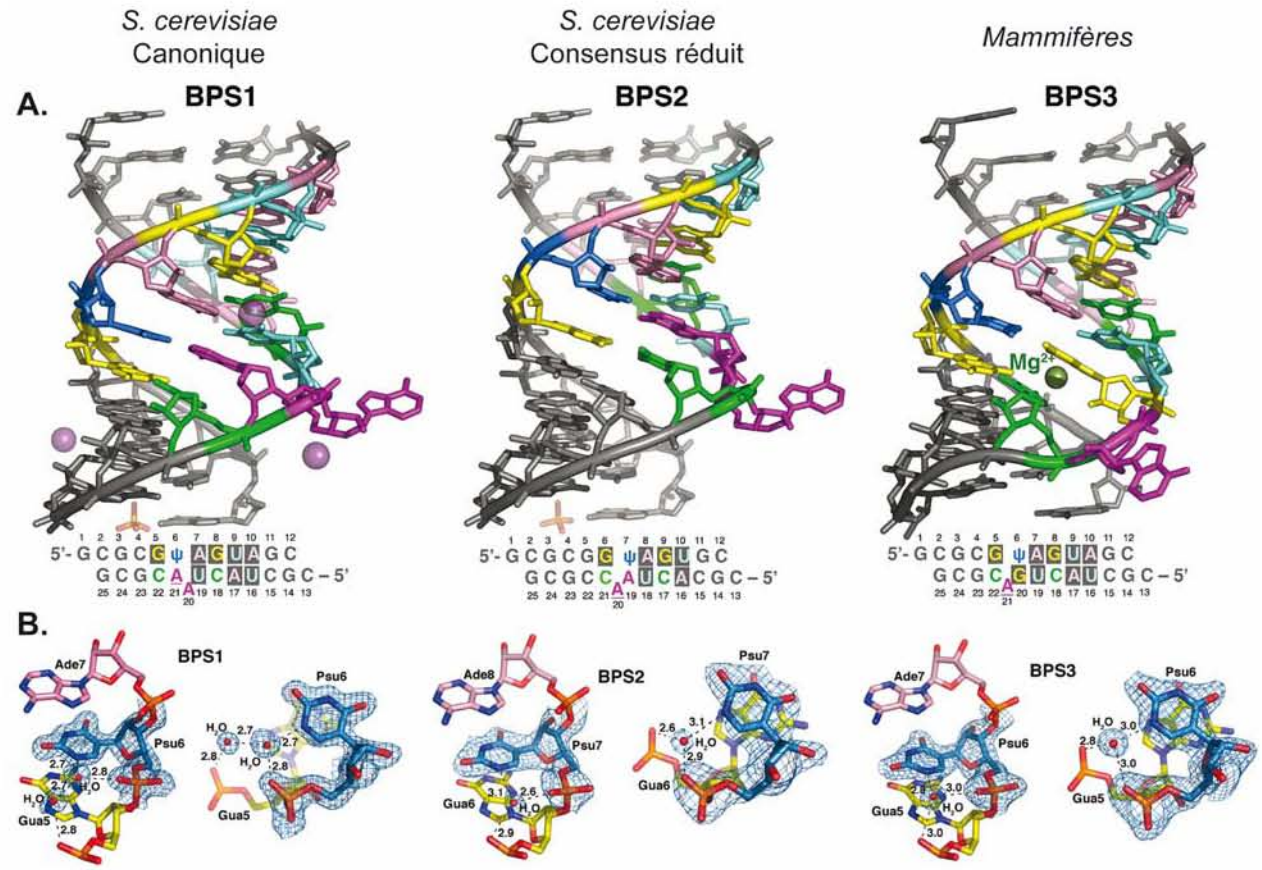


Figure 28 : Structures 3D établies par RX de 2 oligoribonucléotides mimant l'interaction entre la région du point de branchement et le snRNA U2. (Lin & Kielkopf, 2008).

A. Structure 3D des duplexes entre le snRNA U2 et les différentes boîte de branchement : canonique chez la levue (BPS1), consensus réduit chez la levure (BPS2) ou de type mammifères (BPS3). La séquence du dessus représente le snRNA U2, celle du dessous la boîte de branchement. Les nucléotides conservés dans chaque duplexe sont colorés : G en jaune, C en vert, A en magenta, U en cyan et Ψ en bleu.

B. Interactions du type empilement de base et hydratation liées à la présence du résidu Ψ. La carte de densité électronique autour du résidu Ψ et des molécules d'eau est représentée en bleue. La vue de droite correspond à une rotation de 90° autour le l'axe des x par rapport à la vue de gauche.

Bien que la délétion du gène PUS7 codant l'enzyme responsable de la formation du résidu Ψ35 dans le snRNA U2 de *S. cerevisiae* n'entraîne aucun phénotype de croissance, il a été démontré que ce résidu était important pour l'épissage des ARN pré-messagers dans certaines conditions. En effet, si l'on cumule la délétion du gène PUS7 avec la mutation U40G ou la délétion U40Δ du snRNA U2, on observe une baisse de l'épissage de certains ARN pré-messagers et un ralentissement de la croissance de la levure à 37°C (Yang et al., 2005).

Des travaux de l'équipe de Manley montrent également un rôle du résidu Ψ35 dans l'activation de l'adénosine du point de branchement (Valadkhan & Manley, 2001). Cette équipe a pu montrer qu'en incubant un petit ARN contenant la séquence du point de branchement (ARN Br) avec le complexe U2/U6, il y avait activation du groupement 2' OH de l'adénosine du point de branchement et attaque du squelette phosphodiester du snRNA U6 au niveau de la triade AGC (Valadkhan & Manley, 2001). Cette réaction est très similaire à la première transestérification de la réaction d'épissage et conduit à la production d'un ARN, appelé ARN X, dans lequel une liaison phosphodiester est formée entre le snRNA U6 et le point de branchement. L'expérience réalisée avec un snRNA U2 synthétique contenant ou ne contenant pas le résidu Ψ35 met en évidence l'importance de ce résidu (Fig. 29).

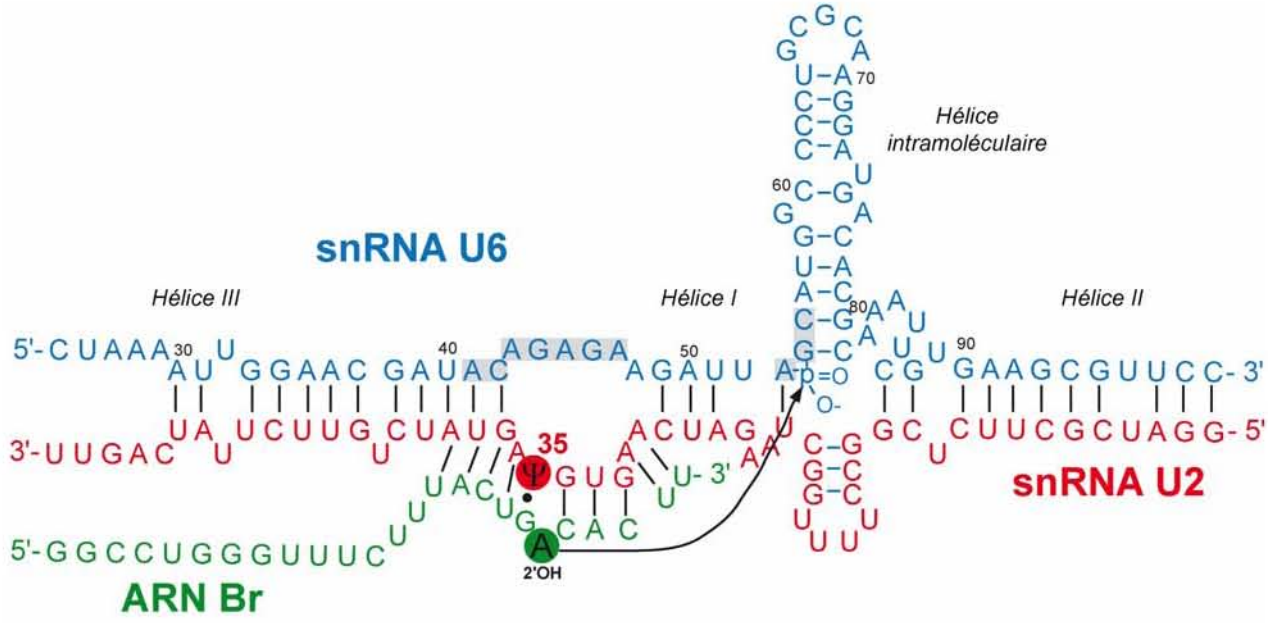


Figure 29 : Représentation de l'interaction entre le complexe U2/U6 et l'ARN Br. (Valadkhan & Manley, 2003).
Les régions grisées correspondent aux résidus invariants du snRNA U6. Les hélices intermoléculaires I, II et III formées entre le snRNA U2 et le snRNA U6, ainsi que l'hélice intramoléculaire du snRNA U6 sont indiquées. Les groupements chimiques impliqués dans la formation de l'ARN X sont montrés. La flèche noire indique la réaction responsable de la formation de l'ARN X. Le résidu Ψ est indiqué en rouge (Madhani & Guthrie, 1994).

Dans toutes les espèces, le snRNA U2 peut adopter deux conformations alternatives, A et B (Fig. 30). La forme B, qui comporte une structure en pseudonoeud, est celle adoptée par le snRNA U2 lorsqu'il s'associe à l'ARN pré-messager pour constituer les spliceosomes (Zavanelli et al., 1994). Trois résidus Ψ et deux résidus 2'-O-méthylés sont présents dans les segments du snRNA U2 humain impliqués dans la formation de cette structure en pseudonœud (Massenet et al., 1998). La présence de ces résidus modifiés pourrait ainsi faciliter la transition conformationnelle de la forme A (inactive) à la forme B (active).

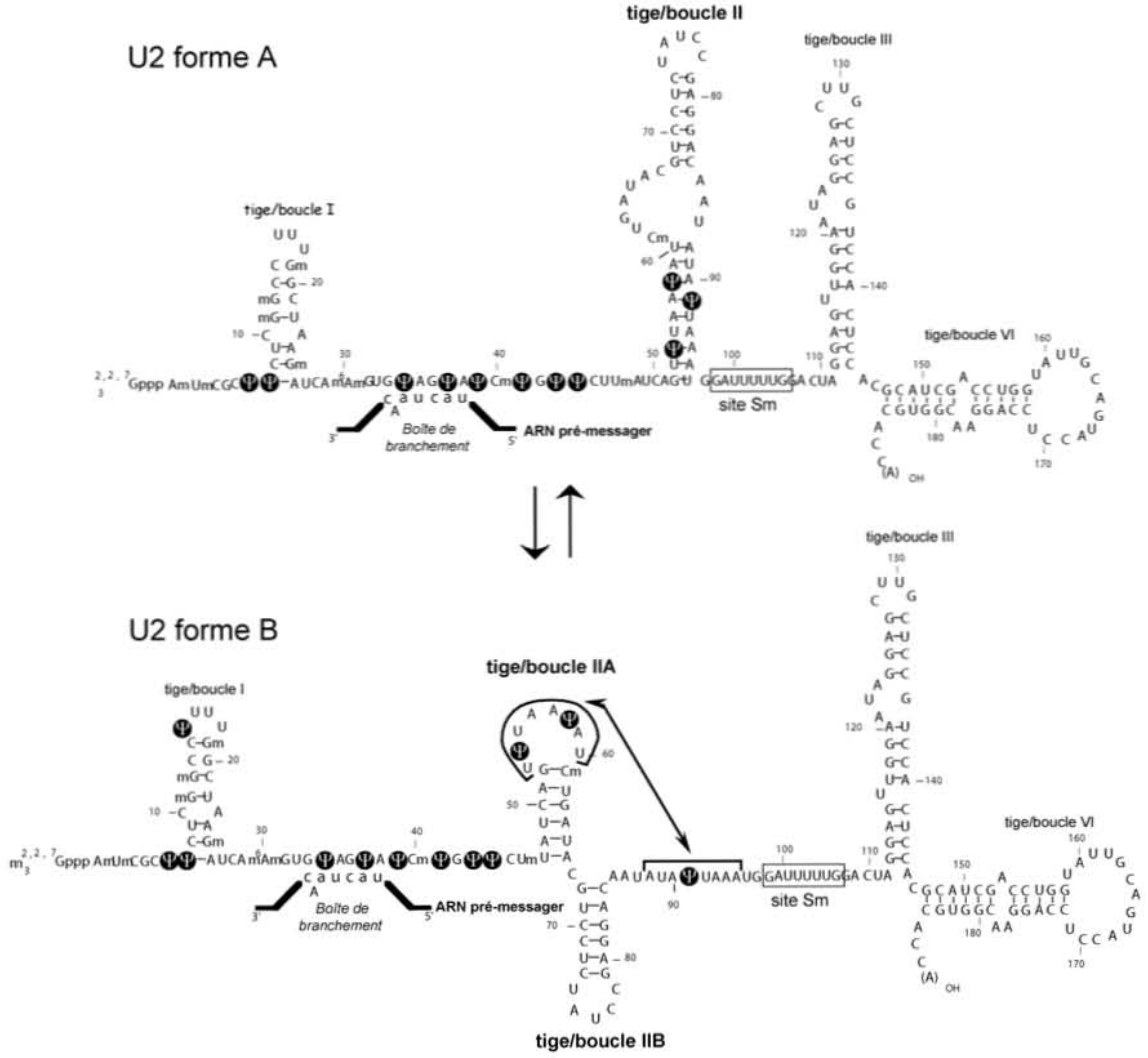


Figure 30 : Les deux structures alternatives du snRNA U2 de cellules HeLa.
La structure de type pseudonœud est représentée par la double flèche (Branlant et al., 1982; Ares & Igel, 1990; Zavanelli & Ares, 1991).

1.3. Cas du snRNA U5

Il a été démontré que la boucle terminale du snRNA U5 interagit dans un premier temps avec l'extrémité du 1er exon, puis dans un deuxième temps avec les extrémités des deux exons, ceci, afin de faciliter leur ligation (Fig. 31A et B) (Madhani & Guthrie, 1994; Umen & Guthrie, 1995). Une interaction entre les snRNA U1 et U5 humains a également pu être mise en évidence (Ast & Weiner, 1997). Dans ce dernier cas, c'est le résidu Ψ conservé dans la boucle terminale du snRNA U5 qui est directement impliqué (Fig. 31C). Il a été proposé que la présence des modifications post-transcriptionnelles dans la boucle terminale du snRNA U5 permet une structuration particulière de cette boucle et facilite ainsi la réaction d'épissage.

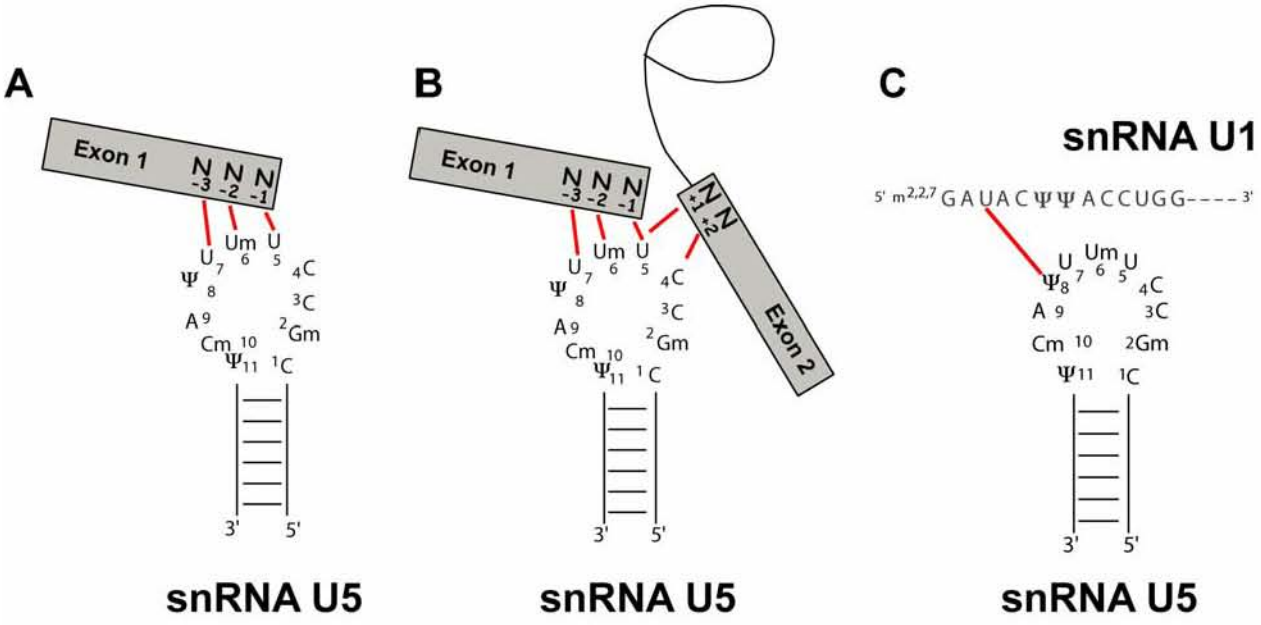


Figure 31 : Mise en évidence, par pontage covalent, des interactions hétérologues entre le snRNA U5 et les deux exons et entre les snRNA U1 et U5 au cours de la réaction d'épissage.
Les Panneaux A et B correspondent à des interactions détectées entre le snRNA U5 et les exons avant la première réaction de transestérification (A) ou avant la deuxième réaction de transestérification (B). Le Panneau C correspond à l'interaction détectée entre le snRNA U1 et le snRNA U5. Les résidus pontés sont reliés par un trait rouge.

2. Interactions UsnRNA/UsnRNA

Comme nous l'avons décrit dans la partie précédente, de nombreux résidus Ψ sont présents dans des régions d'interaction entre UsnRNA. Ainsi, le snRNA U2 possède des résidus Ψ qui sont impliqués dans les hélices II et III formées par l'association entre les snRNA U2 et snRNA U6 (Fig. 11). De même, on notera que chez les vertébrés, certains résidus Ψ présents dans le snRNA U4 sont impliqués dans l'interaction avec le snRNA U6, et ce à la fois dans les spliceosome majeurs et mineurs (Fig. 12). La présence de ces nombreux nucléotides modifiés, et notamment des résidus Ψ, est encore mal connue mais il a été proposé que ces résidus pourraient stabiliser ou faciliter la formation de ces différentes hélices ou bien encore être impliquée dans la catalyse (Massenet et al., 1998).

Il apparaît donc que beaucoup des résidus Ψ présents au sein des UsnRNA ont vraisemblablement pour fonction, soit de stabiliser des éléments de structures secondaires, soit de favoriser des conformations particulières des ARN ou des interactions de types hétérologues.

4. Caractérisation, classification, mécanismes enzymatiques et structure 3D des ARN:Ψ-synthases

A. Caractérisation des ARN:Ψ-synthases

Les ARN:Ψ-synthases connues à ce jour ont été caractérisées par 3 approches distinctes : soit par voie biochimique, soit par homologie de séquence avec les ARN:Ψ-synthases préalablement identifiées suivie de leur caractérisation expérimentale, soit par criblage de produits de gènes par des techniques à haut débit (pour revue, Ansmant et al., 2001).

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 qui consiste à utiliser comme substrat des ARN ayant incorporé des uraciles marqués au tritium (^{3}H) en position 5 (Cortese et al., 1974). En effet, la réaction d'isomérisation entraîne l'élimination de tritium à cette position qui peut être dosé. C'est ainsi que l'enzyme TruA, a été caractérisée chez *Salmonella typhimurium* (Arena et al., 1978) et chez *E. coli* (Kammen et al., 1988). D'autres ARN:Ψ-synthases d'*E. coli* ont ensuite été identifiées par la même approche (Tableau 3). De nombreuses ARN:Ψ-synthases d'*E. coli* et d'autres organismes ont été détectées par homologie de séquence avec les ARN:Ψ-synthases identifiées biochimiquement.

Lorsqu'un gène est supposé coder une ARN:Ψ-synthase, deux approches expérimentales sont utilisées pour le prouver. D'une part, lorsque les outils génétiques sont disponibles chez l'organisme étudié et que le gène est non essentiel, il est possible de le déléter et de repérer l'absence de résidus Ψ dans les différents substrats possibles. Cette méthode implique l'extraction des ARN totaux, leur modification chimique par un réactif capable de modifier spécifiquement les résidus Ψ (CMCT) et leur séquençage par rétrotranscription. La modification de ces positions peut ensuite être rétablie en complémentant la souche délétée par apport du gène par un plasmide. Dans le cas de gènes essentiels, la région codante du gène n'est pas délétée mais placée sous contrôle d'un promoteur régulable. Les substrats potentiels sont analysés après extinction de l'expression du gène pendant un temps supérieur à la demi-vie des ARN étudiés.

La seconde approche consiste à cloner le gène d'intérêt dans un vecteur d'expression qui permet la surproduction de la protéine chez *E. coli*. Cette protéine sera ensuite purifiée et sa capacité à modifier des substrats ARN produits par transcription *in vitro* sera testée. Ces différentes approches sont complémentaires et permettent la caractérisation précise d'une ARN:Ψ-synthase (pour revue, (Ansmant et al., 2001).

Plus récemment, une nouvelle approche a été développée. Celle-ci est basée sur le criblage de produits de l'expression de gènes à haut débit en utilisant une banque d'expression de l'ensemble des ORF de *S. cerevisiae* permettant de produire les protéines correspondantes fusionnées à la GST. Cette technique nécessite donc de connaître la position du résidu Ψ et permet de retrouver l'enzyme qui est responsable de sa formation. C'est en utilisant cette

stratégie que l'équipe de Yu a mis en évidence l'ARN:Ψ-synthase Pus7p de *S. cerevisiae* qui a la particularité de ne présenter que peu d'homologie de séquence avec les autres ARN:Ψ-synthases connues (Ma et al., 2003). Cette équipe cherchait à caractériser les protéines responsables de la pseudouridylation en position 35 et 42 du snRNA U2. La stratégie consiste à tester l'activité, sur un snRNA U2 produit *in vitro*, de différents ensembles de protéines de fusion purifiées (Fig. 32). Le snRNA U2 utilisé comme substrat contient un résidu U35 marqué au ^{32}P à l'extrémité 5'. Par recoupement entre plusieurs ensembles portant l'activité, et test de l'activité de sous-ensembles de clones, les auteurs ont caractérisé le gène Yor243c comme codant l'ARN:Ψ-synthase responsable de la modification de la position 35 du snRNA U2. La démonstration définitive de l'activité de la protéine Yor243p a été obtenue par production et purification de cette protéine chez *E. coli* et test *in vitro* de son activité sur un transcrit du snRNA U2. Une dernière démonstration de son activité en position 35 du snRNA U2 a été obtenue par comparaison de l'activité sur un snRNA U2 produit *in vitro*, d'un extrait de levure sauvage et d'un extrait de levure correspondant à une souche dans laquelle le gène YOR243 avait été délété. Parmi les trois résidus Ψ présents dans le snRNA U2 seul le résidu Ψ35 est formé par l'enzyme Pus7p (Ma et al., 2003).

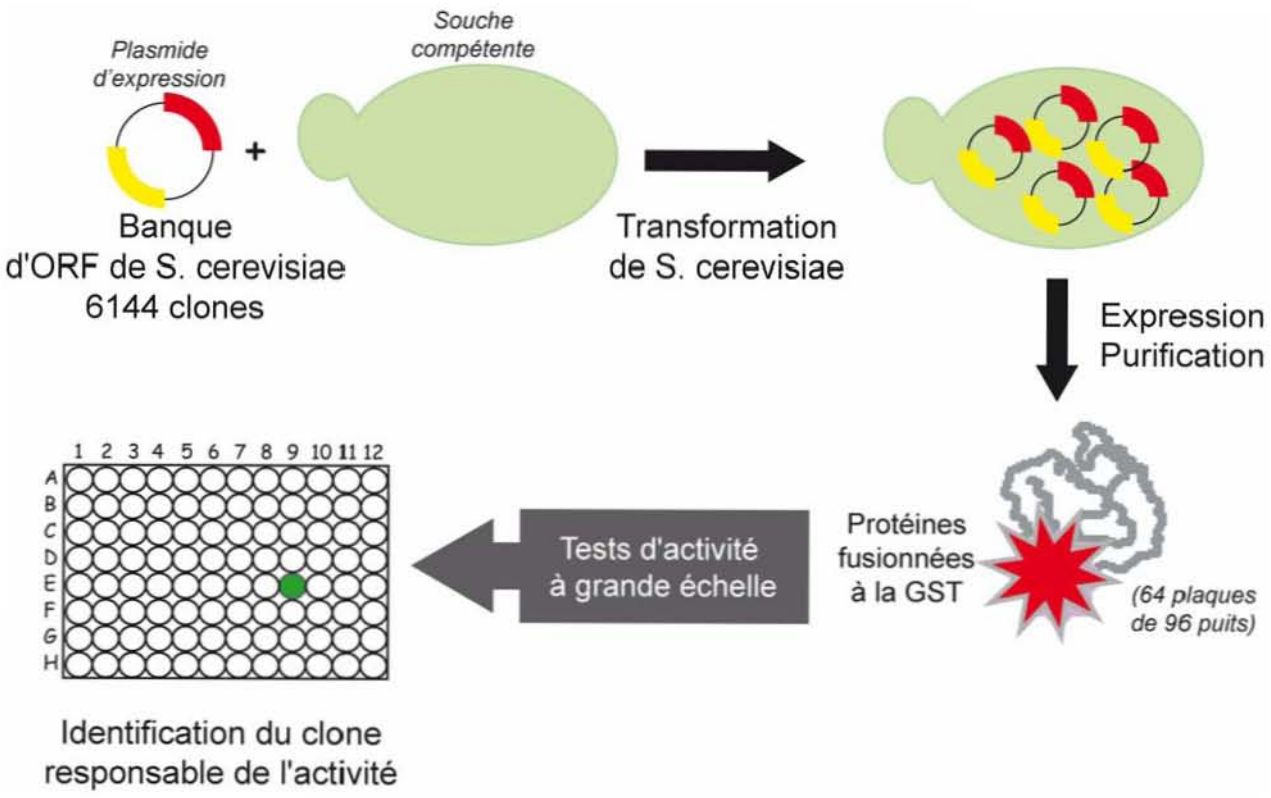


Figure 32 : Principe de la stratégie d'identification d'une ARN:Ψ-synthase par criblage à haut débit.

B. Classification des ARN:Ψ-synthases

La caractérisation de la première ARN:Ψ-synthase par voie biochimique il y a plus de 20 ans chez *E. coli* a permis, avec le développement de la bioinformatique, d'accélérer de manière considérable l'identification de ces enzymes. En effet, nous savons aujourd'hui que toutes les ARN:Ψ-synthases sont réparties en 5 familles distinctes sur la base de leurs similarités de séquence : les familles TruA, TruB, RluA, RsuA et TruD (Tableau 3) (pour revue, (Hamma & Ferre-D'Amare, 2006). Une sixième famille pourrait être représentée par les enzymes Pus10/PsuX identifiées chez certains eucaryotes et archaea mais qui seraient absentes chez les bactéries (Watanabe & Gray, 2000 ; Roovers et al., 2006).

Toutes ces familles présentent un ou plusieurs membres qui ont en commun un certain nombre de motifs conservés que nous détaillerons par la suite. Ainsi, les familles RluA, RsuA, TruB, TruA et TruD comportent 5 motifs communs qui sont assez conservés, dont 3 (motifs I, II et III) ont été décrits très tôt à partir des alignements de séquence de l'ensemble des ARN:Ψ-synthases et les motifs IIa et IIIa qui ont été défini plus tardivement sur la base de la connaissance de plusieurs structures 3D de membres de cette famille d'enzymes (Fig. 33) (Del Campo et al., 2004). Ces motifs IIa et IIIa sont situés en aval des motifs II et III, et correspondent respectivement aux enchaînements « +xYx### » et « Lxxxx# » où + correspond à un acide aminé basique, x, à n'importe quel résidu et #, à un résidu hydrophobe. Bien que peu d'information soit disponible sur la famille Pus10, il semble qu'elle présente au moins le motif II. Le motif II, commun aux différentes familles d'ARN:Ψ-synthases,s comporte deux acides aminés très conservés, il s'agit d'un acide aspartique strictement conservé et d'une glycine. Notons que ce motif est également présent chez les déoxycytidine triphosphate déaminases (Dcd) et deoxyuridine triphosphatases (Dut) qui lient également spécifiquement des résidus U (Koonin, 1996). Par conséquent, le motif II des ARN:Ψ-synthases a été proposé comme étant le site de fixation de l'uridine et il appartiendrait donc au site catalytique (Koonin, 1996). En effet, l'implication de l'acide aspartique conservé du motif II dans la catalyse a été démontré pour l'enzyme TruA (Huang et al., 1998a), puis pour les enzymes RluA (Raychaudhuri et al., 1998) ; (Ramamurthy et al., 1999), RsuA (Conrad et al., 1999), TruB (Ramamurthy et al., 1999) et RluC d'*E. coli* (Conrad et al., 1999).

On peut remarquer qu'il existe chez *S. cerevisiae* des équivalents pour 5 des 6 familles d'ARN:Ψ-synthases identifiées à partir d'*E. coli* puisque aucun membre de la famille RsuA n'est présent chez la levure. Il faut noter que la famille TruB regroupe les protéines Pus4 et Cbf5 qui, bien que présentant une forte homologie de séquence, possèdent une spécificité de substrat très différente.

Des équivalents de ces 6 familles d'ARN:Ψ-synthases ont également été recherchés par similarités de séquence chez les Archaea (Watanabe & Gray, 2000). Il semble que seuls les euryotes présentent un membre de la famille TruA. Les deux règnes comportent un

	Organisme	Mode d'identification	Spécificité de substrat	Références
Famille 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
Famille 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
Famille RluA				
RluA	*E. coli*	Biochimique	Ψ_{746} ARNr 23S + Ψ_{32} ARNt	Wrzesinski *et al.*, 1995a
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
Famille RsuA				
RsuA	*E. coli*	Biochimique	Ψ_{516} ARNr 16S	Wrzesinski *et al.*, 1995b
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
Famille 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
			Ψ_{50} ARNr 5S	Decatur et Schnare, 2008
Famille Pus10				
Pus10 / PsuX	*P. furiosus*	Homologie de séquence	Ψ_{55} ARNt	Watanabe and Gray 2000 Roovers *et al.*, 2006

Tableau 3: 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 : biochimique en orange, par homologie de séquence en rouge et par utilisation d'une banque d'ORF fusionnées à la GST en bleu.

représentant pour la famille TruB qui a été annoté comme étant aCBF5 chez les Archaea et CBF5 chez les eucaryotes. Les familles TruD et Pus10 sont également représentées chez les Archaea. Cependant, aucun représentant des familles RsuA et RluA n'y a été observé.

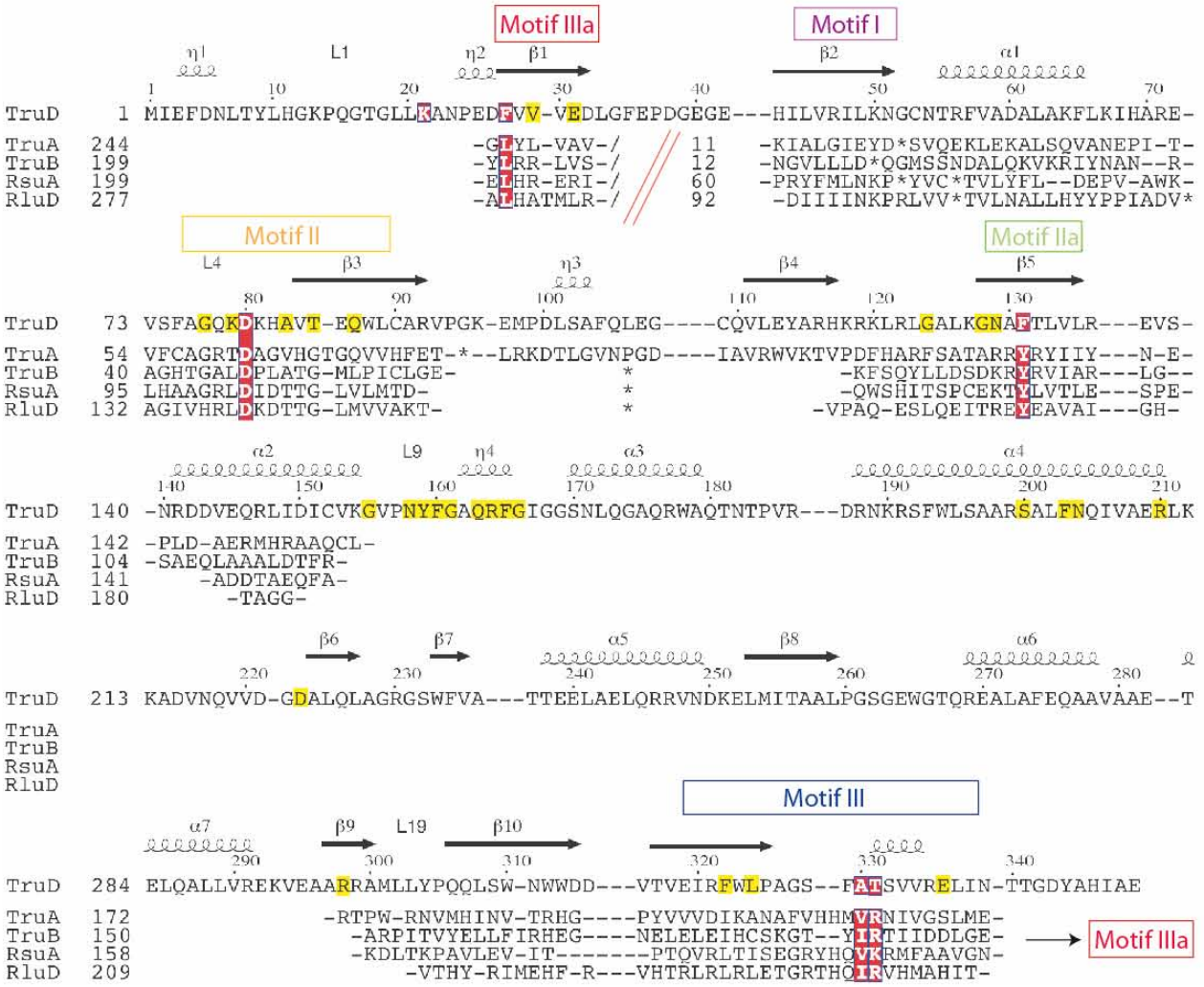


Figure 33: Alignement multiple de séquence des ARN:Ψ-synthases appartenant aux 5 familles structuralement caractérisées à ce jour.

Les résidus présentant plus de 80% de similarité de séquence sont représentés en jaune. Les astérisques impliquent la présence de plus de deux acides aminés non alignés. Le motif IIIa présent en N-ter de TruD est localisé en C-ter dans les autres familles. Les 5 résidus appartenant à 4 des 5 motifs conservés entre les 5 familles et jouant un rôle dans le site catalytique sont encadrés en rouge (Kaya & Ofengand, 2003).

C. Mécanisme enzymatique

Deux mécanismes catalytiques différents impliquant tous deux l'acide aspartique conservé ont été proposés (Fig. 34) (Huang et al., 1998a). Le premier mécanisme catalytique repose sur l'attaque nucléophile de la position C6 de l'uridine par le carboxylate de l'acide aspartique conservé de l'enzyme (Fig. 34A). Une attaque nucléophile de ce type a déjà été proposée pour

la méthyltransférase responsable de la méthylation de la base U en position 54 des ARNt (RUMT) mais dans ce cas l'attaque nucléophile est réalisée par une cystéine (pour revue, (Garcia & Goodenough-Lashua, 1998). Cependant, des résultats récents ne sont pas en cohérence avec le premier mécanisme. En effet, il a été observé que pour TruB liée à un substrat contenant du 5FU, il est possible de récupérer un produit d'hydrolyse du 5FU (5-fluoro-6-hydroxypseudouridine) et qu'aucun complexe covalent entre l'acide aspartique et l'ARN n'est formé (Hoang & Ferre-D'Amare, 2001 ; Spedaliere & Mueller, 2004).

Le second mécanisme catalytique possible (Fig. 34B) repose sur l'attaque nucléophile du carbone C1' du ribose de l'uridine par le carboxylate de l'acide aspartique catalytique. Cette réaction a été suggérée sur la base du mécanisme enzymatique proposé pour la première étape de la formation de la queuine à partir de la guanine par l'ARNt guanine transglycosylase (pour revue, (Romier et al., 1998). Il est à noter que ce mécanisme est retrouvé chez les ADN-glycosylases (Labahn et al., 1996).

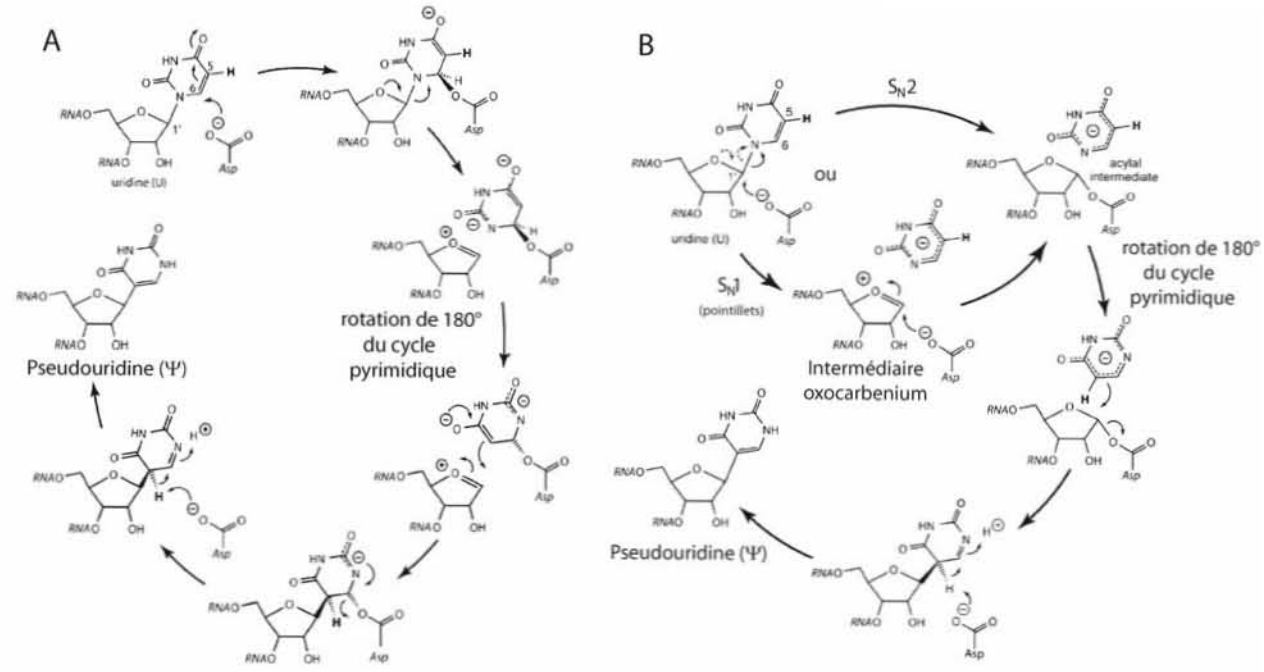


Figure 34 : Mécanismes proposés pour la réaction de pseudouridylation (d'après (Hamilton et al., 2006). Deux mécanismes différents reposent l'un sur l'attaque nucléophile en position C6 de l'uridine (**A**), l'autre sur l'attaque nucléophile de la position C1' du ribose de l'uridine (**B**).

Bien qu'il soit difficile de distinguer lequel de ces mécanismes est utilisé, de nombreux arguments sont en faveur du premier. L'utilisation d'un ARNt uniformément marqué par incorporation de 5-fluorouracile (5FU) a montré pour TruA d'*E. coli* que le complexe formé entre TruA et l'ARNt marqué est un complexe covalent qui, après dénaturation de TruA, libère

le 5FU complexé à TruA et que la séparation du complexe TruA-5FU-tRNA en ses composants ARN et protéine donne lieu à la libération de 6-hydroxy-5,6-dihydro-fluorouridine (Gu et al., 1999). Ces résultats ne peuvent s'expliquer qu'avec le premier mécanisme reposant sur l'attaque nucléophile en position C6 de la base et non sur le ribose (Gu et al., 1999 ; Hamilton et al., 2006). Cependant, aucune conclusion claire ne peut être donnée et il est fort possible que certaines ARN:Ψ-synthases utilisent le premier mécanisme tandis que d'autres utilisent le second, la répartition s'effectuant très probablement en fonction de la famille à laquelle elles appartiennent. Le rôle des autres résidus intervenant dans la catalyse a été en grande partie clarifié après étude structurale de certains représentants de ces enzymes comme nous allons le voir dans le chapitre suivant.

D. Structure tridimensionnelle des ARN:Ψ-synthases

L'étude de structure 3D des ARN:Ψ-synthases a permis de mettre en évidence un certain nombre de caractéristiques communes entre ces enzymes. Actuellement, au moins un membre de chaque famille a été cristallisé seul ou, comme c'est le cas pour les enzymes TruB, RluA, TruA (Hur & Stroud, 2007) et aCBF5, sous forme de complexes avec un ARN. Les ARN substrats utilisés miment respectivement la tige-boucle du bras TΨC, la tige-boucle anticodon, un ARNt entier et enfin un sRNA à boîtes H/ACA (Tableau 4).

Famille	Enzyme	Organisme	Substrat	Auteurs
TRUA	TruA	*E.coli*		Foster et al., 2000
			ARNt entier	Hur et Stroud, 2007
		T.thermophilus		Dong et al., 2006
TRUB	TruB	*E.coli*	ARNt T-SL	Hoang et Ferre-D'Amare, 2001
		T.maritima	ARNt T-SL	Pan et al., 2003
	CBF5	*P. abyssi*		Hamma *et al.*, 2005
			Particule	Li et Ye, 2006
				Manival et al., 2006
				Rashid et al., 2006
RLUA	RluA	*E.coli*	ARNt AC-SL	Hoang et al., 2006
	RluD	*E.coli*		Del Campo *et al.*, 2004
				Mizutani et al., 2004
				Sivaraman et al., 2004
	RluC	*E.coli*		Mizutani *et al.*, 2004
RSUA	RsuA	*E.coli*	U et UMP	Sivaraman *et al.*, 2002
		H. influenzae		Matte et al., 2005
	RluE	*E.coli*		Pan *et al.*, 2007
	RluF	*E.coli*		Sunita *et al.*, 2006
TRUD	TruD	*E.coli*		Ericsson *et al.*, 2004
				Hoang et Ferre-D'Amare, 2004
				Kaya et al., 2004
		M. mazei		NWSGC, unpublished
PUS10	PUS10	*H. sapiens*		McCleverty *et al.*, 2007

Tableau 4 : Bilan des ARN:Ψ-synthases dont la structure tridimensionnelle a été déterminée.

La co-cristallisation en présence d'un substrat ou analogue de substrat est indiquée en jaune. Les substrats T-SL et AC-SL correspondent à des fragments d'ARN ne contenant que le bras TΨC ou le bras anticodon, respectivement.

On constate qu'en dépit des faibles similarités de séquence, il est possible de distinguer un domaine structural commun aux 5 familles d'ARN:Ψ-synthases qui renferme les 5 motifs caractéristiques que nous avons évoqués précédemment (Fig. 35) (Del Campo et al., 2004). Ce repliement comporte au moins 8 brins β flanqués d'hélices et de boucles, qui sont organisés en deux feuillets, le plus souvent anti-parallèles centrés autour du site catalytique (Fig. 35A). L'aspartate catalytique est systématiquement au centre d'une crevasse dont la surface est couverte de résidus basiques. Il est orienté de telle façon que sa chaîne latérale soit dirigée vers le solvant.

La superposition des sites actifs des 5 enzymes révèle cinq acides aminés conservés : l'acide aspartique catalytique strictement conservé du motif II, la tyrosine (ou phénylalanine pour TruD) du motif IIa, et la leucine (ou phénylalanine pour TruD) du motif IIIa, et deux autres résidus dont les propriétés chimiques sont conservées : arginine ou lysine et isoleucine ou valine du motif III (Fig. 35B) (Del Campo et al., 2004). Une exception est le remplacement de ces résidus par des résidus alanine et thréonine dans la famille TruD (Fig. 35B).

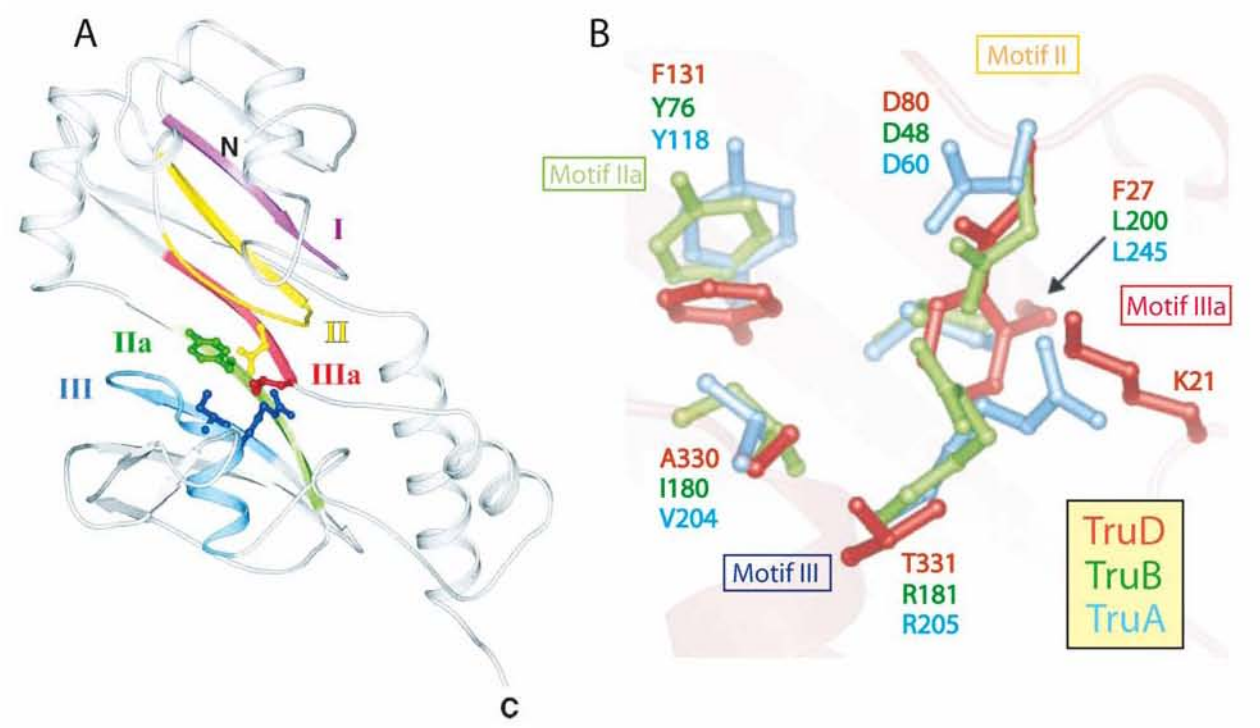


Figure 35 : Caractéristiques structurales communes à l'ensemble des ARN:Ψ-synthases.

A. Structure 3D de l'enzyme RluD (famille RluA) mettant en évidence la conservation structurale du core-enzyme. B. Superposition des chaînes latérales des 5 résidus jouant un rôle dans le site catalytique. Le code couleur pour chacune des trois familles est le suivant : en rouge TruD, en bleu TruA, et en vert

TruB. L'action supplémentaire du résidu K21 de TruD pourrait se substituer à l'action du résidu R du motif III conservé chez les 4 autres familles qui est absent chez TruD.

Il semblerait que le résidu basique du motif III permette la formation d'un pont salin avec l'acide aspartique catalytique ce qui pourrait accentuer le caractère nucléophile du résidu catalytique en favorisant sa forme déprotonnée (Hoang & Ferre-D'Amare, 2001). L'absence de résidu basique à cette position dans la famille TruD serait compensée par la présence d'une lysine en position 21 (Fig. 35B) (Kaya et al., 2004). La tyrosine du motif IIa pourrait jouer un rôle soit dans le transfert de protons, soit dans le positionnement de la base de l'uridine cible (Hoang & Ferre-D'Amare, 2001 ; Phannachet et al., 2005). Le Tableau 5 résume le rôle, dans le site actif, des acides aminés conservés entre TruA, TruB, RsuA, RluD, TruD et aCBF5.

Motif	RluD	RsuA	TruA	TruD	TruB	aCBF5	Rôle
II	-	-	-	-	His43	His77	Stabilisation de l'extrusion de bases
	Arg137	Arg100	Arg58	-	-	-	Stabilisation de l'extrusion de bases
	Asp139	**Asp102**	**Asp60**	**Asp80**	**Asp48**	**Asp82**	**Acide aspartique catalytique**
	Gly144	Gly107	Gly65	Thr85	Gly53	Gly87	?
IIa	Arg169	Lys130	Arg116	Asn129	Lys74	Lys108	?
	Tyr171	**Tyr132**	**Tyr118**	**Phe131**	**Tyr76**	**Tyr110**	**Stacking avec la base de l'uridine cible**
III	His234	His184	His202	-	-		?
	-	-	-	-	Tyr 179	Tyr179	Stacking avec le ribose de l'uridine cible
	Ile236	**Val186**	**Val204**	**Ala330**	**Ile180**	**Ile180**	**?**
	Arg237	**Lys187**	**Arg205**	**Thr331***	**Arg181**	**Arg181**	**Formation d'un pont salin avec l'acide aspartique catalytique**
IIIa	**Leu279**	**Leu200**	**Leu245**	**Phe27**	**Leu200**	**Leu200**	**?**
	-	-	-	**Lys21**	-		**Formation d'un pont salin avec l'acide aspartique catalytique**

Tableau 5 : Localisation et rôle suggéré des résidus appartenant aux cinq motifs conservés entre les cinq familles d'ARN:Ψ-synthases.
Les résidus représentés en rouge sur fond jaune correspondent aux 5 résidus les plus conservés. * : le résidu K21 de TruD compenserait l'absence de résidu basique à la position T331. La protéine d'archaea aCBF5 est un représentant de la famille TruB. Ses résidus sont indiqués en suivant la numérotation de l'équivalent provenant de *P. abyssi*. La protéine RluD est un représentant de la famille RluA. Les numérotations des protéines RluD, TruA, TruB, TruD et RsuA correspondent aux numérotations des équivalents d'*E. coli*.

Par ailleurs, il a également été démontré, par l'étude des enzymes TruB et RluA, que les résidus lysine et proline assez fortement conservés au sein du motif I ont une importance fonctionnelle (Spedaliere et al., 2000). En fait, leurs mutations n'ont que peu d'effet sur

l'efficacité de la catalyse mais provoquent une déstabilisation des protéines. Ces résidus joueraient donc plus un rôle structural que catalytique, et il a été proposé qu'ils pourraient être couplés aux acides aminés du site actif et assurer le mouvement de celui-ci (Hoang et al., 2005).

Bien que le core-enzyme de l'ensemble des ARN:Ψ-synthases soit très similaire, il existe une large variété de domaines qui se replient de manière indépendante aux extrémités N et C terminales (Fig. 36A) entre les membres des différentes familles. Parmi ces repliements particuliers, on note le domaine N-ter de la famille RsuA qui présente des homologies avec celui de la protéine ribosomale S4. La partie N-ter de la famille TruD forme un domaine spécifique à cette famille (domaine TruD). Une partie de la région C-ter des protéines de la famille RluA forme un domaine caractéristique de cette famille. Le domaine C-ter de TruB est appelé PUA car il est également présent dans d'autres enzymes de modification (PseudoUridine synthase and Archaeosine transglycosylases).

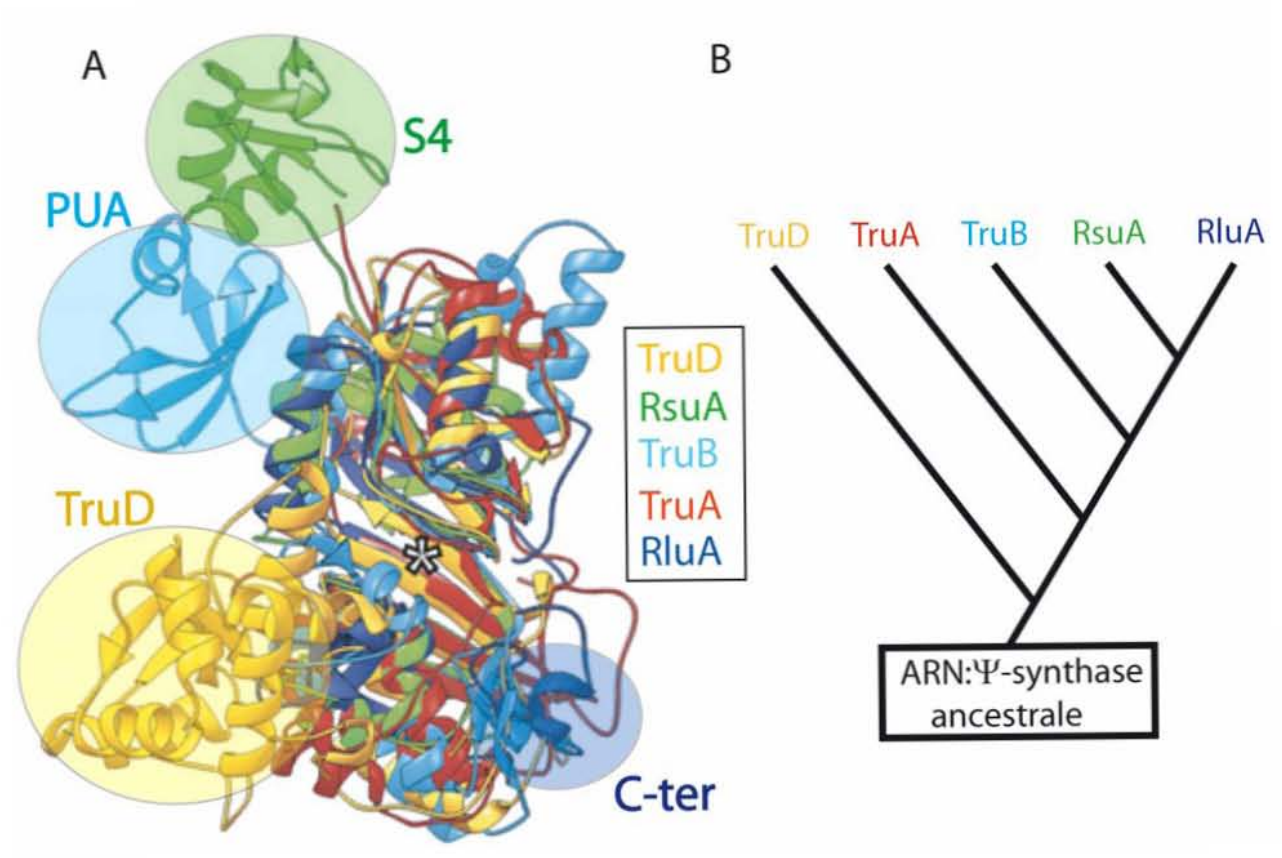


Figure 36 : Comparaison des structures 3D de membres de chacune des 5 familles d'ARN :Ψ-synthases (d'après (Hamma & Ferre-D'Amare, 2006) ; (Kaya et al., 2004).
A. Superposition des modèles en ruban des structures 3D de TruD (jaune), TruA (rouge), RsuA (verte), RluA (bleu foncé) et TruB (bleu clair). Le domaine central est très similaire et permet la mise en exergue des domaines périphériques TruD, PUA, S4 et sous-domaine C-terminal uniques aux familles indiquées. Le site catalytique est marqué par un astérisque. **B.** Evolution hypothétique des relations entre les 5 familles d'ARN:Ψ-synthases, basée sur la comparaison de leur séquence et de leur structure.

Il est probable que l'ensemble des familles actuelles d'ARN:Ψ-synthases dérivent d'un seul et même ancêtre commun (Fig. 36B). Ainsi, les enzymes de la famille TruD présentent le motif IIIa à l'extrémité N-ter ce qui semble indiquer que cette famille a divergé plus tôt au cours de l'évolution. L'enchaînement différent des motifs observés dans la famille TruD pourrait résulter d'une permutation circulaire (Hamma & Ferre-D'Amare, 2006). La conservation des acides aminés du site actif et de son voisinage met en évidence que les enzymes des familles TruA sont les plus éloignées par rapport aux autres familles. De plus, les représentants de la famille TruA sont les seuls connus pour agir en tant que dimères. Les familles RluA et RsuA sont les deux familles qui partagent le plus de similarités avec en particulier la conservations des motifs I, II et III). Finalement, bien que ne possédant pas le motif III, les enzymes de la famille TruB présentent de fortes similarités de séquence avec celles des familles RluA et RsuA.

Il ressort de la comparaison de la structure tridimensionnelle des ARN:Ψ-synthases que ces enzymes partagent de nombreuses caractéristiques communes parmi lesquelles une importante conservation structurale du cœur catalytique mais aussi des acides aminés impliqués dans la catalyse. Du côté des différences, on notera principalement la présence de motifs additionnels toujours situés en périphérie et qui pourraient expliquer, de part leur nature, les différences de spécificité entre ces enzymes.

E. La spécificité de reconnaissance des ARN:Ψ-synthases

Les ARN:Ψ-synthases ont la capacité de modifier spécifiquement un ou plusieurs résidus dans une séquence ARN. Cette spécificité fait intervenir des mécanismes précis et conservés au cours de l'évolution qui permettent la reconnaissance du résidus cible que ce soit dans les ARNr, les ARNt ou les UsnRNA.

Il existe actuellement 2 systèmes majeurs de modification des résidus Ψ qui sont très différents. Le premier repose sur une interaction directe entre l'ARN:Ψ-synthase et son substrat sans nécessiter de cofacteur. Le second fait appel à un ARN guide de type snoRNA ou sRNA à boîtes H/ACA qui va s'apparier avec l'ARN cible et définir la position qui sera modifiée par un complexe enzymatique spécifique que nous décrirons par la suite.

1. ARN:Ψ-synthases ne nécessitant pas de guide ARN

1.1. Classification des ARN:Ψ-synthases

1.1.1. Classification générale des ARN:Ψ-synthases

Comme nous l'avons vu précédemment, il est possible de classer les ARN:Ψ-synthases en 5 familles sur la base de la conservation de certains motifs structuraux. Cependant, si on réalise une analyse détaillée de la spécificité d'action de ces enzymes, on peut remarquer qu'il existe

de grandes différences entre les membres d'une même famille. En effet, une ARN:Ψ-synthase peut reconnaître une ou plusieurs positions dans un ou plusieurs types d'ARN. Sur la base des connaissances actuelles, nous pouvons distinguer trois catégories d'ARN:Ψ-synthases (Fig. 37) :

1) Les enzymes dites « site-spécifiques » ne modifient qu'un seul résidu U dans une classe d'ARN voire dans un seul ARN donné.
2) Les enzymes dites « région-spécifiques » modifient plusieurs positions proches au sein d'un même ARN
3) Les enzymes à « spécificité multiple » modifient plusieurs résidus U dans des contextes différents au sein d'un ou plusieurs types d'ARN. Il est également possible de séparer, parmi les enzymes à spécificité multiple, 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.

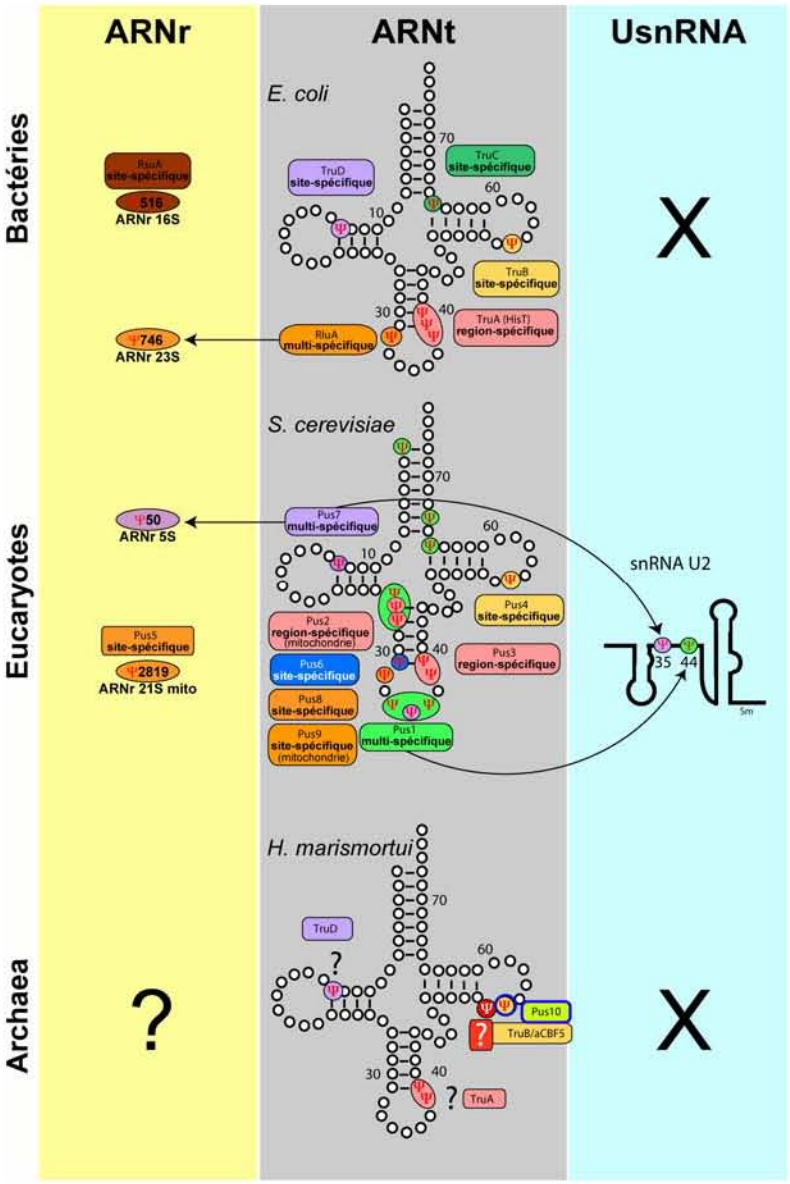


Figure 37 : Classification des ARN:Ψ-synthases selon leur spécificité de substrat.

Toutes les ARN:Ψ-synthases actuellement caractérisées se classent dans ces 3 catégories qui peuvent donc regrouper des enzymes très différentes en fonction des organismes. Nous détaillerons dans la partie suivante les points communs et les divergences qui existent pour chaque catégorie d'enzymes dans les 3 domaines du vivant.

1.1.2. ARN:Ψ-synthases agissant sur les ARNt et UsnRNA

Comme nous l'avons évoqué précédemment, les ARNt possèdent un grand nombre de résidus Ψ, parmi lesquels certains sont communs à plusieurs molécules dans un organisme donné. Il apparaît donc que les ARN:Ψ-synthases impliquées dans ces modifications sont adaptées à la reconnaissance d'une région cible au sein d'un ensemble d'ARNt pouvant présenter des variations de séquence et de structure.

Ainsi, chez *E. coli*, cinq ARNt:Ψ-synthases différentes sont impliquées dans la modification de 7 positions différentes (Fig. 37) (Del Campo et al., 2001). TruA est « région-spécifique » comme évoqué précédemment. RluA est à « spécificité multiple » car elle modifie également un résidu uridine appartenant à la grande sous-unité ribosomale. Les trois autres enzymes ne modifient qu'une seule position et sont donc « site-spécifiques ».

Chez *S. cerevisiae*, six ARN:Ψ-synthases différentes sont impliquées dans la modification de 7 positions différentes des ARNt (Fig. 37). Trois d'entre elles sont « site-spécifiques », Pus4p l'équivalent de TruB et Pus8p et Pus9p les équivalents de RluA. Néanmoins, contrairement à leur homologue bactérien, ces dernières ne seraient pas impliquées dans la modification des ARNr. Bien que Pus7p soit l'équivalent de TruD qui est « site-spécifique », elle est une enzyme à spécificité multiple car impliquée dans la modification des positions 13 de certains ARNt et 35 du pré-ARNt^{Tyr} (Behm-Ansmant et al., 2003), de la position 35 du snRNA U2 (Ma et al., 2003) et de la position 50 dans l'ARNr 5S (Decatur & Schnare, 2008). Pus3p est l'équivalent de TruA qui est « région-spécifique ». Enfin, Pus1p qui n'a pas d'équivalent bactérien, est à « spécificité multiple » car impliquée dans la modification de plusieurs positions des ARNt (Motorin et al., 1998) et de la position 44 du snRNA U2 (Massenet et al., 1999). Il faut aussi remarquer que, contrairement aux eucaryotes supérieurs, aucune pseudouridine n'est détectée en position 54 des ARNt de *S. cerevisiae*.

Chez les Archaea, et particulièrement pour *H. marismortui* dont les modifications des ARNt ont été analysées, les attributions de positions modifiées aux enzymes connues restent hypothétiques (Fig. 37). Récemment, une étude a montré que TruB/aCBF5 et aPus10 de *P. furiosus* sont toutes deux capables de modifier la position 55 des ARNt (Roovers et al., 2006; Muller et al., 2007). L'attribution de la modification des positions 13 et 38/39 aux équivalents archaea des enzymes de TruD et TruA, respectivement, n'est qu'une prédiction logique réalisée par comparaison avec les situations chez les Bactéries et Eucaryotes et doit être vérifiée expérimentalement. Enfin, il est à noter que l'ARN:Ψ-synthase responsable de la pseudouridylation de la position 54 de tous les ARNt élongateurs n'a pas encore été identifiée.

Cette position est également modifiée chez la plupart des eucaryotes, or aucune enzyme n'a encore été proposée comme responsable de sa modification.

Les ARNt constituent un substrat de choix pour l'étude de la spécificité des ARN:Ψ-synthases. En effet, les substrats ARNt se produisent avec de bons rendements par transcription *in vitro* et leur structure tridimensionnelle peut être modifiée facilement par mutagenèse dirigée. Ainsi, Edqvist *et al*. (Edqvist et al., 1993) et Grosjean *et al.* (Grosjean et al., 1996) ont étudié les éléments structuraux de la molécule d'ARNt nécessaires à la formation des différents nucléotides modifiés. Pour cela, des variants d'ARNt de levure ont été microinjectés dans des oocytes de *X. laevis* et les modifications post-transcriptionnelles formées ont été analysées. Les résultats obtenus ont amené Grosjean *et al*. (1996) à proposer une classification des enzymes de modification des ARNt en deux groupes, en fonction de leur sensibilité aux perturbations de structure de la molécule d'ARNt. L'existence de deux groupes distincts a été, par la suite, confirmée par l'étude spécifique de certaines de ces enzymes. On trouve des ARN:Ψ-synthases dans chacun de ces groupes d'enzymes de modification. Les enzymes dites du groupe I reconnaissent un élément structural local de l'ARNt (Fig. 38). C'est le cas des enzymes de la famille TruB (Becker et al., 1997a) ;(Gu et al., 1998). La formation du résidu Ψ55 dépend de la structure de la boucle TΨC et en particulier du maintien de la paire de bases reverse Hoogsteen entre la position 54 et la position 58. La reconnaissance d'éléments structuraux localisés dans le bras TΨC est également une nécessité pour qu'il y ait formation d'autres nucléotides modifiés, comme le résidu m^5U54 (T54) chez *E. coli* et la levure (Santi & Hardy, 1987; Becker et al., 1997b). De la même façon, la méthylation sur le ribose du nucléotide Gm34 dans les ARNt de *X. laevis* (Droogmans et al., 1986), la formation du résidu queuine à partir de la guanine en position 34 chez *E. coli* (Curnow et al., 1993 ; Mueller & Slany, 1995) et la formation du résidu Ψ32 chez *X. laevis* (Grosjean et al., 1996), ne dépendent que de la présence de la structure tige-boucle anticodon.

Les enzymes dites du groupe II sont sensibles à la structure tridimensionnelle globale de l'ARNt (Fig. 38). Les enzymes TruA et Pus3p catalysant la formation des résidus Ψ en position 38, 39 et 40 chez *E. coli* et chez *S. cerevisiae*, respectivement, appartiennent au groupe II (Edqvist et al., 1993 ; Grosjean et al., 1996 ; Lecointe et al., 1998). Cependant, Pus3p semble présenter des différences subtiles de reconnaissance de l'ARNt par rapport à TruA puisqu'elle modifie *in vitro* et *in vivo* les positions 38 et 39, mais pas la position 40 (Lecointe et al., 1998). Les enzymes catalysant la formation notamment des nucléotides modifiés $m^2{}_2G26$ (Edqvist et al., 1992; Edqvist et al., 1994), m^1G37 (Holmes et al., 1992 ; Redlak et al., 1997 ; Brule et al., 1998), I34 (Auxilien et al., 1996), i^6A37 (Motorin et al., 1997) et t^6A34 (Morin et al., 1998) font également partie du groupe II (Fig. 37). L'hypothèse émise dans le cas de la formation des résidus I34 et i^6A37 est que la reconnaissance de la cible requiert la reconnaissance de l'ARNt en globalité et que sa fixation induit une torsion de la structure tige-boucle de l'anticodon

permettant d'optimiser un contact avec le site actif de l'enzyme nécessaire à la formation de ces nucléotides modifiés (Auxilien et al., 1996 ; Motorin et al., 1997).

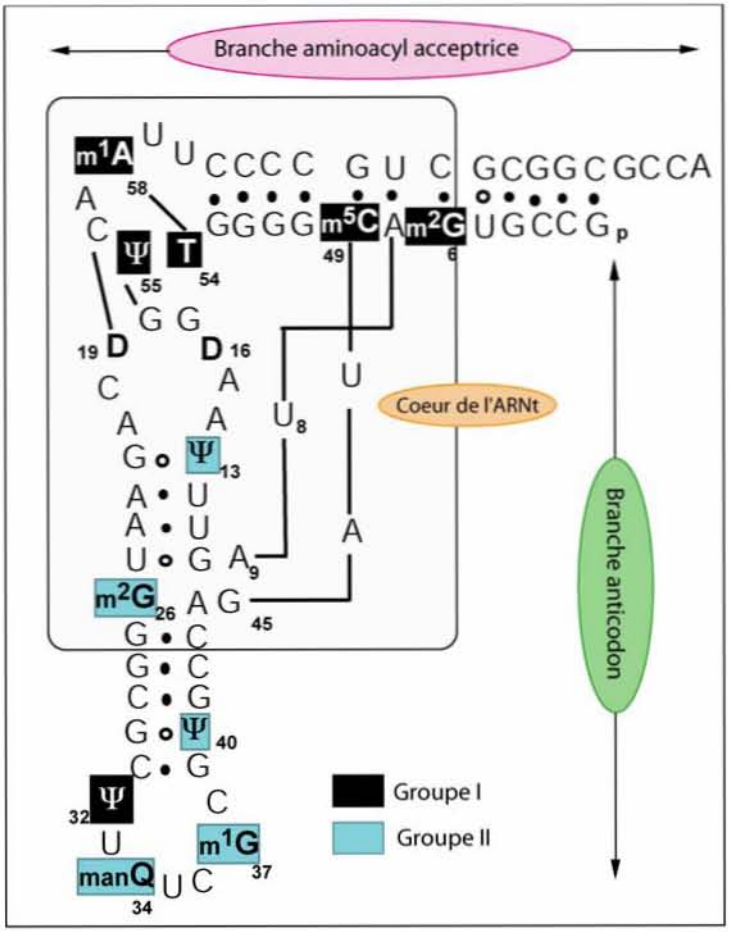


Figure 38 : Classification des enzymes de modification des ARNt en deux groupes selon l'influence de la structure tertiaire de l'ARNt (adapté de (Grosjean et al., 1996).
Les enzymes du groupe I ne sont pas affectées par des perturbations de l'architecture de l'ARNt, tandis que les enzymes du groupe II sont dépendantes de la structure tridimensionnelle de l'ARN. La structure 3D de l'ARNt^{Asp} de *S. cerevisiae* et les modifications formées dans cet ARNt (ou ses variants) lorsque ces ARNt sont microinjectés dans le cytoplasme d'oocytes de *X. laevis* sont représentées ici.

Les ARNt cytosoliques matures, d'une longueur moyenne de 75 à 80 nts sont produits chez les eucaryotes et les archaebactériens à partir de précurseurs plus longs (pré-ARNt) synthétisés par l'ARN polymérase III dans le nucléoplasme. Plusieurs études *in vivo* et *in vitro* ont pu démontrer que certaines enzymes de modification des ARNt, parmi lesquelles les ARN:Ψ-synthases, sont sensibles à la présence d'un intron (Jiang et al., 1997) ; pour revue, (Motorin et al., 1997). Plus précisement, il est possible de distinguer 3 types d'enzymes: les enzymes insensibles à la présence d'un intron, celles dont l'activité est intron-dépendante et celles qui nécessitent l'absence de l'intron (Fig. 39). Il faut remarquer que la majorité des enzymes de modification des ARNt sont insensibles à la présence d'un intron. Le groupe des enzymes de

modification intron-dépendantes est constitué uniquement d'enzymes modifiant des nucléotides appartenant au bras de l'anticodon, mais toutes les modifications situées dans cette région des ARNt ne sont pas intron-dépendantes. Par exemple, il a été démontré que la synthèse du résidu pseudouridine en position 35 de l'ARNt^{Tyr}(GΨA) de plantes dépend probablement de la taille de l'intron plutôt que de sa séquence ou de sa structure (Pienkowska et al., 2002). Enfin, en ce qui concerne le dernier groupe, l'effet inhibiteur de l'intron sur la formation de la modification serait du à un changement de structure local et global qui empêcherait la reconnaissance et/ou la modification de la cible par les enzymes appartenant à ce groupe.

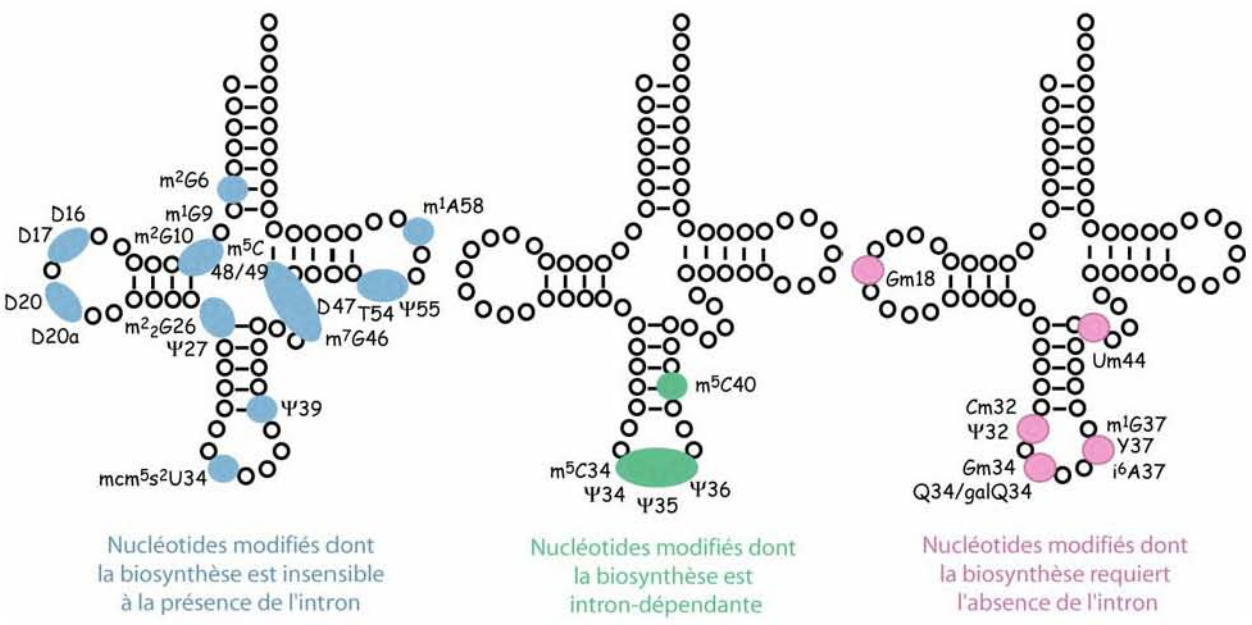


Figure 39: Classification des enzymes de modification des ARNt en trois groupes selon l'influence de la présence d'un intron au sein de l'ARNt (adapté de (Grosjean et al., 1997).

1.1.3. ARN:Ψ-synthases agissant sur les ARNr

Comme nous le verrons par la suite, les 2 systèmes de formation des résidus Ψ, c'est à dire avec ou sans guide de pseudourydilation, sont présents dans les ARNr avec une importance plus ou moins marquée en fonction du domaine du vivant auquel on se rapporte.

Chez les bactéries, tous les résidus Ψ sont formés par des ARN:Ψ-synthases qui ne nécessitent pas d'ARN guide. Les 10 résidus Ψ identifiés chez *E. coli* résultent de l'action de 6 ARN:Ψ-synthases (Fig. 40) (Del Campo et al., 2001). Trois d'entre elles sont « site-spécifiques » : RluB, RluE et RluF. L'enzyme RluA, comme précisé ci-avant, modifie également les ARNt et a donc une spécificité multiple. RluD modifie trois positions dans l'hélice 69 du domaine IV de l'ARN de la grande sous-unité ribosomale et est donc « région-spécifique ». L'hélice 69 pourrait être l'élément reconnu sachant que chacun des 3 résidus U (situés en position 1911, 1915 et 1917) est situé à une distance compatible avec le site actif de

l'enzume RluD pour être isomérisé en pseudouridine (Raychaudhuri et al., 1998 ; Huang et al., 1998b). Il est à noter également que chacun des résidus U modifiés par RluD est suivi par un résidu A en 3'. Enfin, RluC reconnaît trois positions (955, 2504 et 2580) dans des contextes apparemment différents et a donc une spécificité multiple. Ces positions sont éloignées les unes des autres dans la séquence primaire mais sont proches dans la structure tertiaire de l'ARNr 23S (Conrad et al., 1998 ; Huang et al., 1998b) ce qui a pu être démontré par des expériences de pontage (Dokudovskaya et al., 1996). RluC pourrait donc reconnaître un élément structural dans l'ARNr 23S et modifier dans cette région des résidus U qui, par ailleurs, sont toujours suivis par un résidu G en 3' (Conrad et al., 1998). Par conséquent, RluC et RluD semblent sensibles à la nature du nucléotide en 3' du site à modifier. Il est intéressant de remarquer que les enzymes RluB et RluF d'*E. coli* reconnaissent des cibles très proches soit, respectivement les positions 2605 et 2604 de l'ARNr 23S (Fig. 40) (Del Campo et al., 2001). Ces deux protéines reconnaissent donc la même région de l'ARNr 23S, mais dirigent la conversion spécifique d'un seul résidu U.

De la même manière, la formation du seul résidu Ψ détecté dans l'ARNr 16S d'*E. coli* est catalysée par l'enzyme RsuA.

Chez *S. cerevisiae*, la situation est beaucoup plus simple puisque seule la formation du résidu Ψ2819 présent dans les ARNr mitochondriaux est catalysée par une ARN:Ψ-synthase (Pus5) (Ansmant et al., 2000) Les autres résidus Ψ présents dans les ARNr cytoplasmiques étant formés par un système associant l'enzyme Cbf5 à des ARN guides.

Chez les Archaea, aucune ARN:Ψ-synthase n'a été identifiée comme directement responsable de la formation de résidus Ψ dans les ARNr.

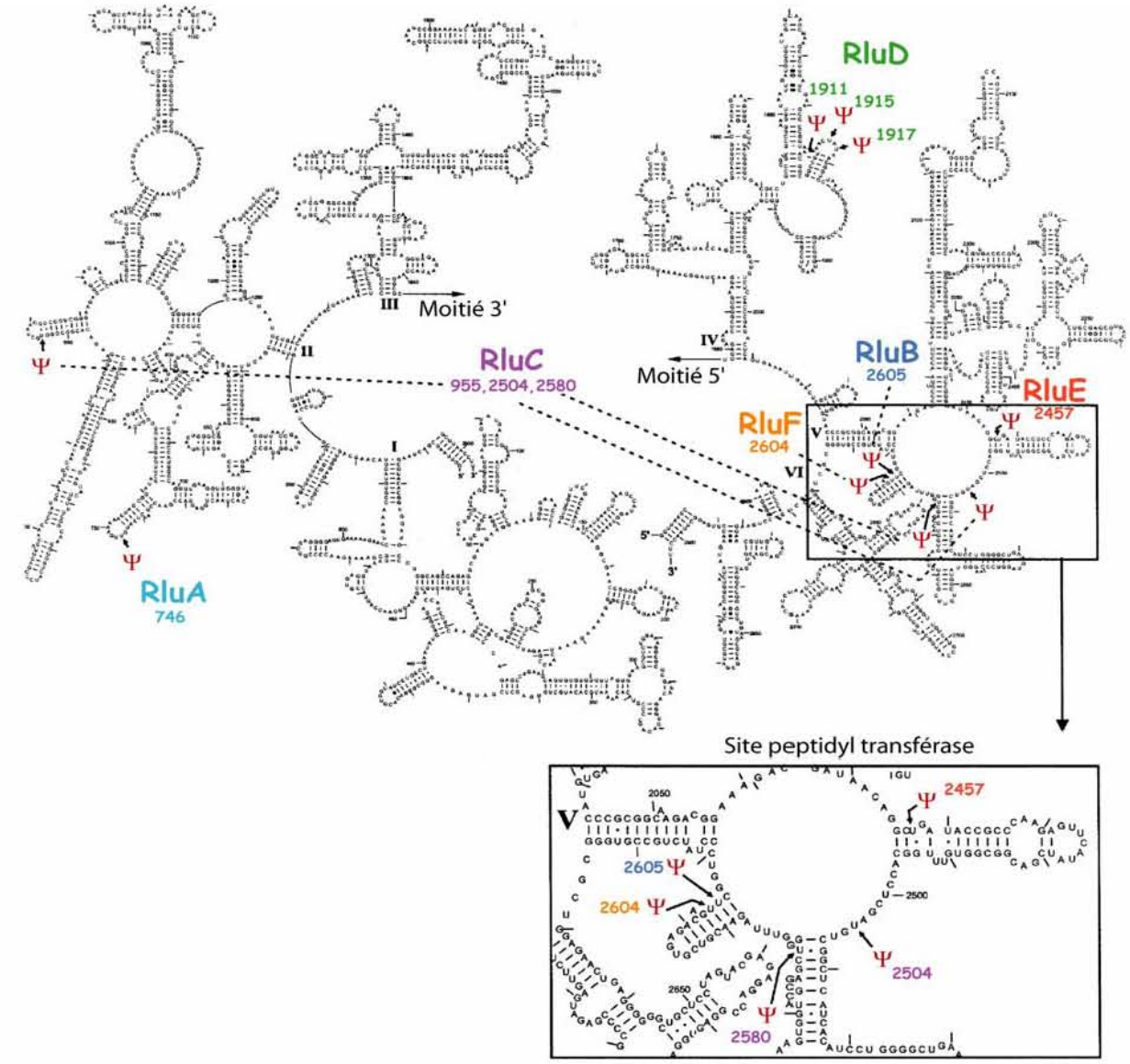


Figure 40 : Les cibles des ARNr:Ψ-synthases d'*E. coli* au sein de l'ARNr 23S (d'après (Del Campo et al., 2001)**.**

1.2. Mode de reconnaissance du substrat ARN

1.2.1. Site-spécificité de TruB

L'ARN:Ψ-synthase TruB d'*E. coli* est une enzyme dite « site-spécifique » de 314 acides aminés, qui modifie la position 55 de tous les ARNt à l'exception des ARNt initiateurs qui comportent un résidu A à la place d'un U en position 54. Des études de spécificité des enzymes TruB d'*E. coli* (Gu et al., 1998) et de son homologue Pus4p chez *S. cerevisiae* ont été réalisées par mutagenèse dirigée des bras accepteur et TΨC (Becker et al., 1997b). Ces études ont montré que TruB et Pus4 sont capables de modifier des fragments d'ARNt ne comportant que le bras TΨC. Ceci signifie d'une part que la structure 3D de l'ARNt n'est pas requise et d'autre part que la présence de la tige acceptrice n'est pas essentielle pour que la position U55 soit pseudouridylée. Il a été démontré également, pour TruB, que la longueur de la tige TΨC doit

être au moins de 2 paires de base et que la paire 53-61 doit être impérativement une paire G-C (Fig. 41E) (Gu et al., 1998). Une boucle TΨC de 7 nucléotides est optimale pour la modification mais une boucle de 6 ou 8 nucléotides est également reconnue par TruB. Au sein de cette boucle TΨC, la nature de certains nucléotides, U54, C56, une purine en position 57 ainsi que A58 semble essentielle à la modification de la position U55. La spécificité de Pus4 est très similaire à celle de TruB à l'exception d'une plus grande contrainte au niveau de la longueur de l'hélice qui doit être d'au moins 4 pbs G-C (Becker et al., 1997b).

La structure de la protéine TruB complexée à un ARN synthétique mimant le bras TΨC d'un ARNt a permis de détailler les interactions entre TruB et son ARN cible (Hoang & Ferre-D'Amare, 2001). L'ARN cible TSL (pour T Stem-Loop) correspond à un fragment d'ARN d'une taille de 22 nucléotides dans lequel le résidu U cible (situé en position équivalente à la 55) est remplacé par un résidu 5-fluorouridine a priori non hydrolysable permettant la formation d'un complexe convalent entre enzyme et substrat. On remarque que le domaine PUA C-terminal correspondant aux acides aminés 250 à 314 n'interagit qu'avec le squelette ribose-phosphate de certains résidus de la tige acceptrice de l'ARNt substrat (Fig. 41). La tige-boucle TΨC, plus précisément sa partie distale, est imbriquée dans la protéine TruB au niveau d'une crevasse principalement constituée de l'association du domaine « pouce » et de résidus provenant à la fois du motif II et de la région en aval du motif II (résidus 40 à 67). Le domaine « pouce » est constitué de résidus provenant de l'élément dit d'insertion Ins2 (résidus 135 à 150) car absents des autres ARN:Ψ-synthases de la même famille et de quelques résidus en amont de celui-ci (résidus 128 à 135). Les bases U55, C56, G57 sont les seules qui sont en interaction spécifique avec, respectivement les résidus des motif II/III, Ins1 (résidus 90 à 105) et Ins2. Les résidus du motif III jouent principalement un rôle dans l'environnement immédiat du U55. Un résidu tyrosine (Y179) uniquement conservé dans la famille TruB fait partie du motif III et jouerait un rôle dans le positionnement du ribose du résidu U55. De plus ses orbitales Π riches en électrons pourraient compenser la charge positive de l'intermédiaire réactionnel oxocarbenium si celui-ci est bien un intermédiaire de la réaction (Fig. 34).

Par conséquent, les résidus de TruB responsables de la reconnaissance du substrat ARN sont principalement localisés dans le motif II et quelques nucléotides situés en C-ter de celui-ci, et dans les éléments Ins1 et Ins2. Le domaine PUA semble jouer un rôle mineur. Les nucléotides de la tige-boucle TΨC qui sont reconnus sont donc principalement G53, U55, C56, G57, C61 et la paire U54-A58. Ces derniers résultats sont en accord avec les données expérimentales de l'étude de spécificité de TruB et de Pus4 (Becker et al., 1997b; Gu et al., 1998)

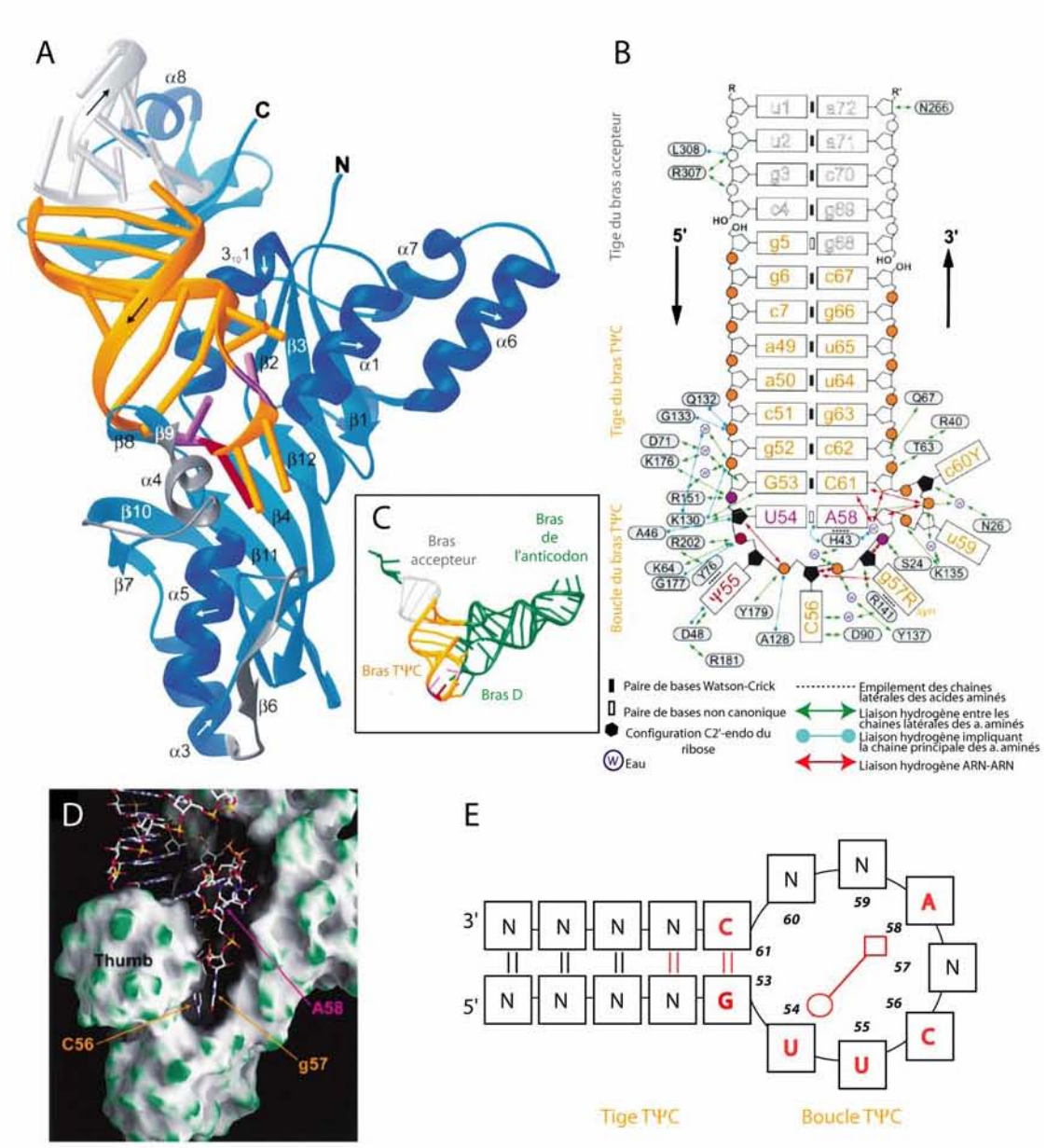


Figure 41: Structure de l'ARNt:Ψ55-synthase TruB d'*E. coli* complexée à un ARN synthétique mimant le bras TΨC de l'ARNt (d'après Hoang & Ferre-D'Amare, 2001).
A. Représentation de la structure 3D du complexe. Les hélices de la protéine sont représentées en bleu foncé, les brins et les boucles en bleu clair. Les deux segments protéiques (Ins1 et Ins2) caractéristiques des protéines du représentant TruB d'*E. coli* sont indiqués en gris. La tige-boucle TΨC de l'ARNt est représentée en jaune, à l'exception de l'uridine cible (U55) en rouge et des nucléotides U54 et A58 qui forment une paire de bases reverse Hoogsteen en rose. **B.** Schématisation de l'ensemble des interactions protéine-ARN et ARN-ARN. Les résidus conservés au sein de l'ARN sont écrits en majuscules et R et Y désignent respectivement la conservation d'un résidu purine ou pyrimidine. **C.** Représentation de la structure 3D de l'ARNt^{Phe} de *S. cerevisiae* (Shi & Moore, 2000). Les bras T et accepteur sont représentés selon le code couleur utilisé dans le panneau A, le reste de la molécule est en vert. **D.** Bases moléculaires de la fixation de la tige-boucle TΨC 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 de l'ARNt substrat dans une crevasse profonde formée notamment par le domaine « pouce ». La protéine TruB est ici représentée par sa surface moléculaire colorée selon l'incurvation (les zones vertes correspondant à des surfaces convexes et les zones grises à des surfaces concaves). **E.** Séquence consensus reconnue par TruB obtenue grâce à l'étude de mutants 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.

Il faut noter également que, dans un contexte naturel, TruB doit accommoder les modifications présentes aux positions 54 et 58 qui apparaissent précocement lors de la biogenèse des ARNt (Grosjean et al., 1996). Par exemple, le ribose du nucléotide U54 peut être 2'-O-méthylé (Um54) et le C5 de la base peut aussi être méthylé (T54). Il existe en effet un espace libre à proximité du nucléotide 54 dans TruB qui autoriserait la présence de ces groupements méthyles (Hoang & Ferre-D'Amare, 2001).

1.2.2. Double spécificité de substrat de RluA

Pour rappel, RluA appartient à la famille des enzymes à spécificité multiple puisque responsable de la formation du résidu Ψ32 dans les ARNt cytoplasmiques et de la position 746 dans l'ARNr 23S d'*E. coli*. Seul le mécanisme de reconnaissance de RluA a été étudié expérimentalement par le biais de l'analyse structurale d'un cocristal correspondant à RluA associé à un fragment d'ARN qui mime le bras anticodon d'un ARNt (Fig. 42) (Hoang et al., 2006). La famille RluA présente une boucle qui joue le même rôle que le domaine « pouce » de TruB mais aussi un second domaine additionnel appelé domaine « forefinger » ou « index » en interaction avec le petit sillon de l'ARN substrat (Hoang et al., 2006). Comme l'étude de TruB, celle-ci montre l'implication dans les interactions avec l'ARN cible de résidus appartenant aux motifs II, III, domaine « pouce » ainsi que, et contrairement à TruB, des résidus appartenants au motif I, en particulier de résidus constituant le domaine « index ». Cette étude montre également que seule la partie distale de la tige-boucle, i.e. les bases des deux dernières paires de bases, ainsi que les bases de la paire de bases reverse-Hoogsteen U33-A36 et des nucléotides extrudés de la boucle, à savoir le site de modification U32, ainsi que G34, A35 et A37 mais aussi A38 sont en interaction spécifique avec des résidus de l'enzyme. Le résidu U746 de l'ARNr de la grande sous-unité ribosomale est situé dans un contexte analogue. Bien que la boucle comportant la position U746 de l'ARNr 23S compte un nucléotide de plus, la formation d'une paire de bases reverse Hoogsteen U-A est possible. Il n'est actuellement pas possible de prédire si cette modification dans l'ARNr est la marque de l'évolution et de la divergence de RluA ou si cette modification présente une fonction au sein de l'ARNr.

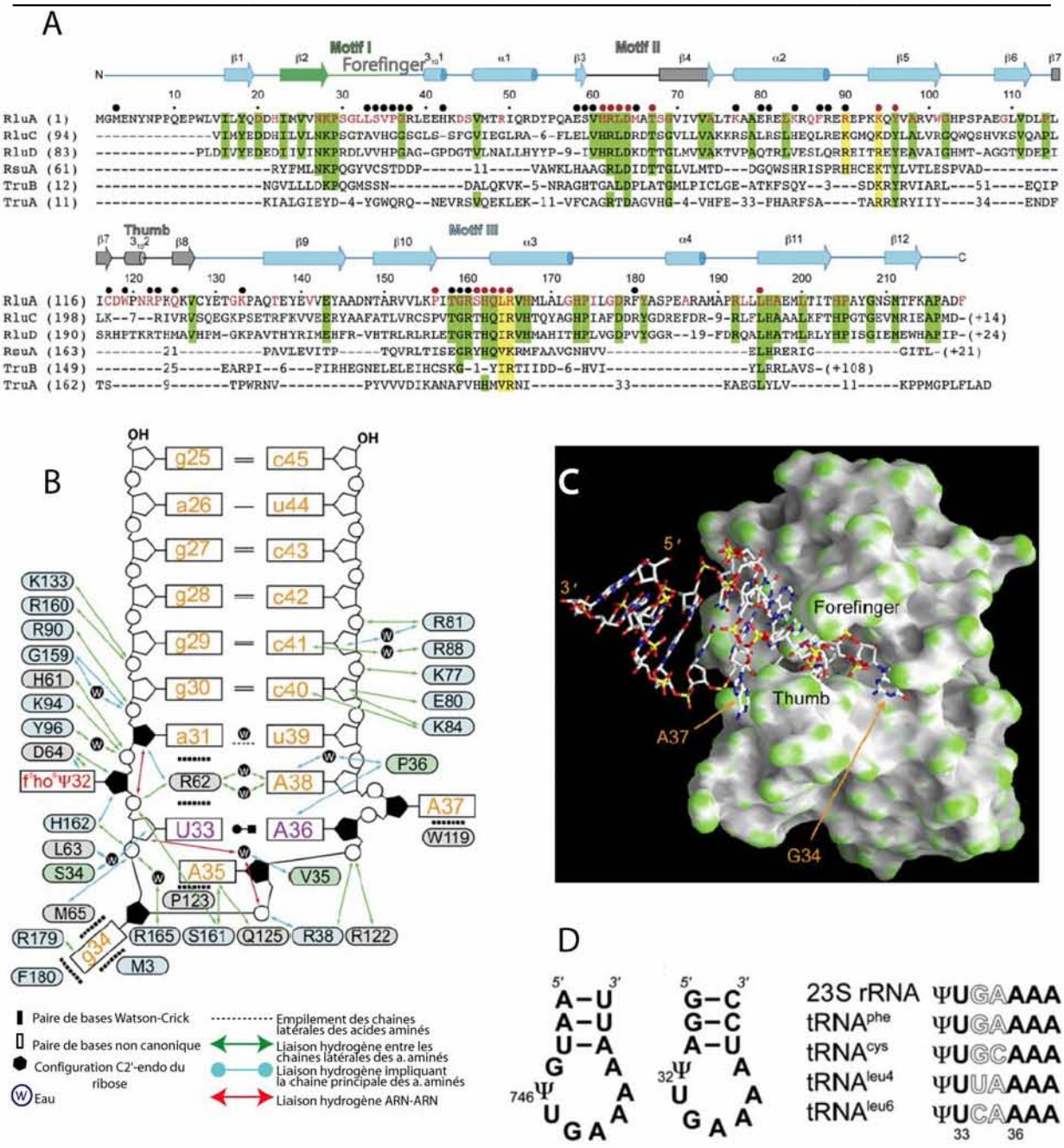


Figure 42 : Structure de l'ARNt:Ψ32-synthase RluA d'E. coli complexée à un ARN synthétique mimant le bras anti-codon de l'ARNt (d'après Hoang et al., 2006).

A. L'alignement structural de 6 ARN:Ψ-synthases a été réalisé à l'aide du programme DALI (Holm & Sander, 1993), en utilisant RluA comme référence. Les cercles noirs et rouges placés au dessus de l'alignement indiquent une perte de plus de 20% de la surface accessible au solvant lors de la fixation de l'ARN. Les cercles rouges représentent les résidus qui délimitent la cavité du site actif. Les résidus représentés en rouge dans la séquence correspondent à des résidus invariants entre orthologues de RluA. **B.** Schémas de l'ensemble des interactions existant entre l'ARN et l'enzyme. Les liaisons hydrogène représentées sont de taille inférieure à 3.5 Å. **C.** Représentation 3D de la surface accessible au solvant de RluA colorée selon l'incurvation (les zones vertes correspondant à des surfaces convexes et les zones grises à des surfaces concaves). **D.** Séquences cibles de RluA dans l'ARNr ou dans l'ARNt^{Phe} de *E. coli*, *et al*ignement, selon la séquence, des sites de pseudouridylation de RluA.

1.2.3. Le mécanisme d'extrusion de base

La résolution du co-cristal de TruB associé à une partie mimant le bras TΨC de l'ARNt a permis de mettre en évidence un mécanisme d'extrusion de bases (ou « flipping out») (Hoang & Ferre-D'Amare, 2001) ; pour revue, (Hamma & Ferre-D'Amare, 2006). La reconnaissance par TruB d'une tige-boucle TΨC induit un changement de conformation de trois nucléotides de la boucle TΨC. Les résidus U55, uridine cible, ainsi que C56 et G57 sont rejetés vers l'extérieur de l'empilement hélicoïdal auquel ils appartiennent au sein de la boucle TΨC. Ainsi, des liaisons hydrogène peuvent se former entre ces bases et des résidus de TruB (Fig. 43). Cette conformation particulière de la boucle TΨC semble stabilisée par un résidu histidine (H43 chez *E. coli*) qui est conservé chez tous les représentants de cette famille. Ce résidu, placé 5 résidus en amont du résidu catalytique, interagit avec la paire reverse Hoogsteen U54-A58, (i.e. interaction en *trans* de la face W-C du U avec la face Hoogsteen du A) caractéristique des boucles TΨC et rend le résidu U55 accessible à l'aspartate catalytique (Fig. 43).

Ce mécanisme d'extrusion de base n'est pas unique à TruB. En effet, l'étude structurale du co-cristal de RluA associé à un segment d'ARN mimant la tige-boucle anticodon montre que cette association impose une nouvelle structure à la boucle anticodon (Hoang et al., 2006). La fixation de RluA entraîne la formation d'une nouvelle paire de bases reverse Hoogsteen U33-A36 ainsi que l'extrusion de trois bases, U32 le site de modification, ainsi que G34 et A37 au sein de la boucle (Fig. 43). Un résidu arginine présent deux résidus en amont du résidu catalytique et conservé chez tous les membres des familles RluA, RsuA et TruA stabiliserait cette conformation. Le rôle d'intercalant de cette arginine est donc analogue au rôle de l'histidine de TruB évoqué ci avant. Le mécanisme d'extrusion de bases serait donc indispensable à la catalyse réalisée par chacune des familles d'ARN:Ψ-synthases (Hoang & Ferre-D'Amare, 2001 ; Sivaraman et al., 2002). Par ailleurs ce mécanisme d'extrusion est fréquemment observé pour des enzymes qui requièrent un accès à une base de l'ADN ou de l'ARN lors d'évènements de modification ou de réparation (pour revue, Cheng & Roberts, 2001).

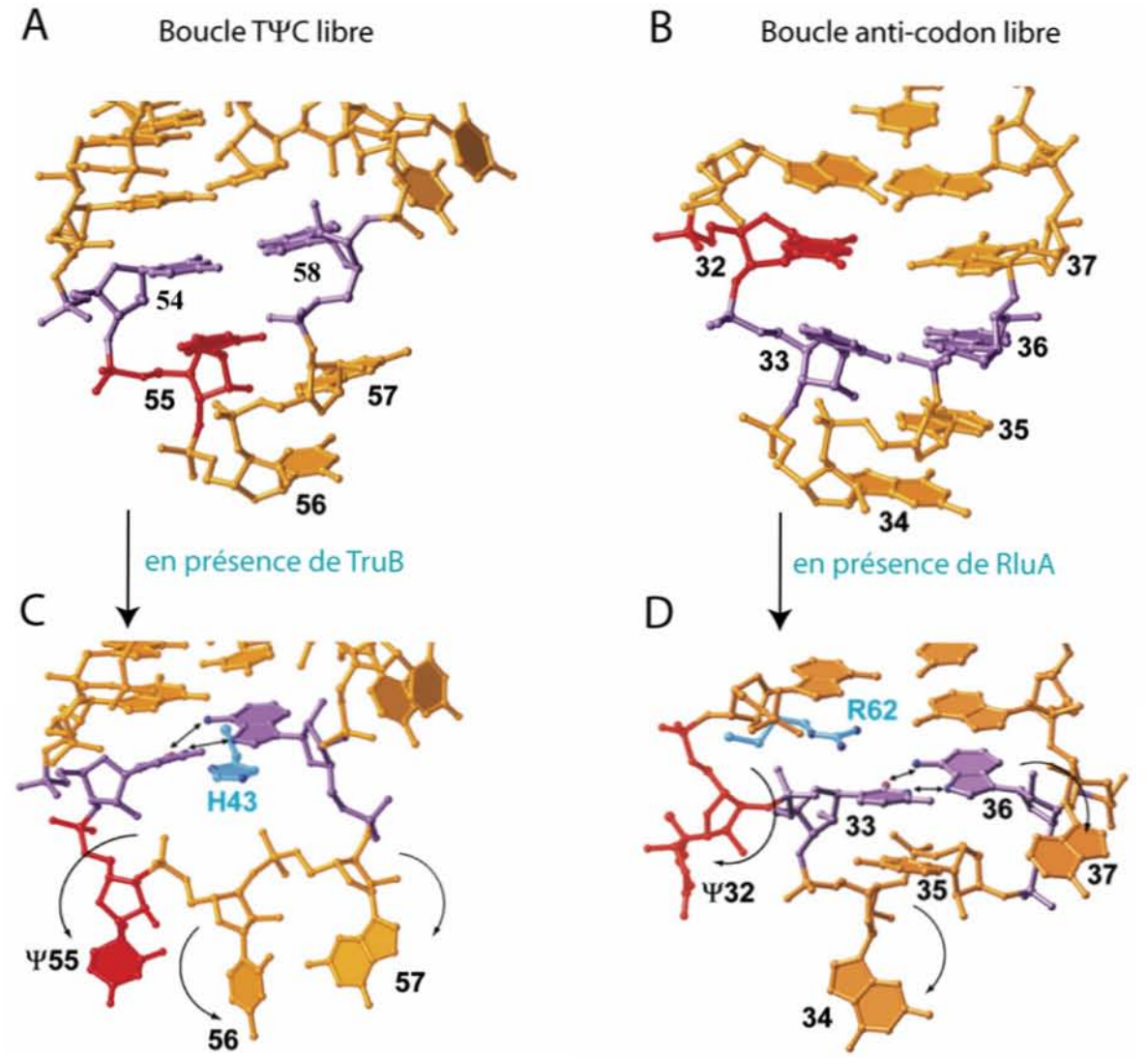


Figure 43 : Mécanisme d'extrusion de bases mis en évidence par l'étude de la reconnaissance de l'ARN par TruB et RluA (d'après Hamma & Ferre-D'Amare, 2006).
A. et B. Structure 3D des boucles TΨC et anti-codon libre provenant de la structure 3D établie pour l'ARNt^{Phe} *d'E. coli* (Shi & Moore, 2000). Les résidus U ciblés par TruB et RluA sont représentés en rouge. **C.** Structure 3D de la boucle TΨC liée à TruB (Hoang & Ferre-D'Amare, 2001). Le résidu H43 indiqué en bleu provient de TruB. La formation de paires reverse-Hoogsteen entre les résidus U et A représentés est indiquée en violet. **D.** Structure 3D de la boucle anti-codon liée à RluA (Hoang et al., 2006). Le résidu R62 indiqué en bleu provient de RluA. La formation de paires reverse-Hoogsteen entre les résidus U et A représentés est indiquée en violet.

On constate donc que le nombre d'interactions entre l'enzyme et l'ARN cible est relativement élevé après le réarrangement induit par la fixation de l'enzyme ce qui suggère que la reconnaissance de la position à modifier repose avant tout sur la capacité de la séquence ARN à adopter cette structure caractéristique. La possibilité de former la paire de bases reverse-Hoogsteen U-A serait donc un pré-requis pour qu'il y ait reconnaissance de la cible.

1.2.4. Région-spécificité de TruA

Le mode d'extrusion de bases exploité par l'enzyme TruA semble légèrement différent à celui de TruB que nous venons de présenter car il ne repose pas sur la présence d'une paire reverse-Hoogsteen (Hur & Stroud, 2007) et aucun motif spécifiquement reconnu par toutes les ARN:Ψ-synthases n'a encore été déterminé. Contrairement à TruB qui est site-spécifique, et cible une seule position dans le bras TΨC, l'enzyme TruA modifie trois positions différentes : 38, 39 et 40 dans l'hélice du bras anti-codon. Elle est donc qualifiée de région-spécifique. Il est à noter également que TruA modifie plusieurs ARNt (17 chez *E. coli*) comportant d'importantes divergences dans le bras anti-codon. Il était intéressant de déterminer comment TruA, avec un site actif unique, peut cibler des positions pourtant distantes de 15 Å. Le co-cristal du dimère de TruA associé à deux ARNt différents a été résolu récemment (Hur & Stroud, 2007). C'est la première fois qu'un ARNt entier a été utilisé comme substrat pour l'étude structurale d'une ARNt:Ψ-synthase. Ceci était justifié par le fait que des expériences antérieures avaient démontré que le bras anti-codon seul n'est pas substrat de TruA (Huang et al., 1998b). La structure 3D du complexe révèle qu'il existe 3 conformations différentes de l'ARNt qui miment trois intermédiaires réactionnels différents ce qui a conduit les auteurs à proposer un modèle de reconnaissance d'un ARNt par TruA dans lequel ni TruA ni l'ARNt ne subissent de changement conformationnel majeur lors de l'association (Fig. 44) (Hur & Stroud, 2007).

Ainsi, TruA interagit dans un premier temps, avec le coude (jonction entre les boucles D et TΨC) et l'hélice D de l'ARNt par le biais d'interactions électrostatiques et par reconnaissance de la forme de la chaîne ribose phosphate. Ces interactions orientent l'ARNt et positionnent le bras anti-codon à proximité du site actif (Fig. 44A). Puis, en passant d'une conformation dite ouverte vers une conformation fermée, le résidu Arg58 vient interagir avec un des trois nucléotides : 38, 39 ou 40 et provoque son extrusion vers le résidu Asp60 catalytique. 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, il a été démontré, qu'en stabilisant ou déstabilisant cette hélice, le taux de modification de chacun des trois nucléotides varie. La présence d'un résidu Ψ modifie la stabilité de l'hélice du bras anti-codon et affecte la lecture du codon par l'anti-codon (ce rôle est discuté ci-après). 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 l'hélice (Hur & Stroud, 2007). Au final, les enzymes de la classe d'enzyme TruA ont évolué vers une reconnaissance de la flexibilité de l'hélice anticodon. Ce mode de reconnaissance permet de distinguer les hélices stables, leur évitant une stabilisation excessive, des hélices peu stables pour lesquelles il est possible 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 & Stroud, 2007).

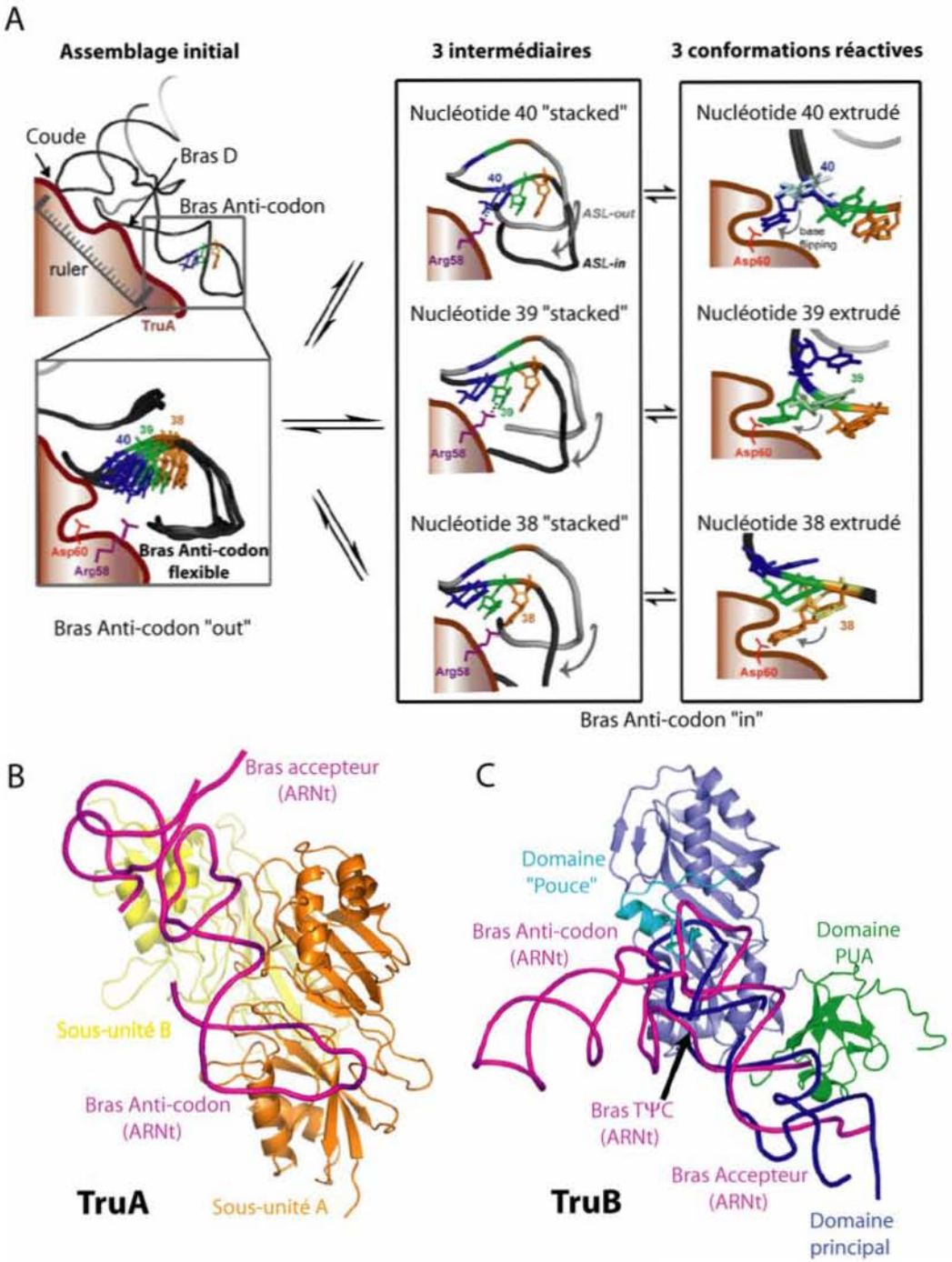


Figure 44: Modèle proposé pour la reconnaissance du nucléotide cible par TruA et comparaison avec le mécanisme de reconnaissance de TruB (d'après (Hur & Stroud, 2007). **A.** Modèle de reconnaissance par TruA. Compte tenu de la flexibilité du bras anti-codon lors de l'assemblage initial, trois intermédiaires peuvent être formés selon l'identité du nucléotide adjacent au résidu Arg58, ce qui conduit à trois conformations réactives. **B.** et **C.** Comparaison des structures de dimères de TruA avec TruB dans la même orientation. L'ARNt est représenté en rose dans les deux cas. En **C.** il s'agit d'une modélisation du positionnement de l'ARNt, le fragment utilisé pour la cristallisation apparaît en bleu.

En comparant TruA avec TruB (Fig. 44 B et C) on remarque que l'ARNt interagit avec la même crevasse du domaine principal de l'enzyme mais ceci avec des orientations différentes. TruA utilise la dimérisation pour reconnaître le coude de l'ARNt alors que TruB utilise son domaine PUA pour la reconnaissance du bras accepteur. La dynamique de l'assemblage de TruA sur l'ARNt repose principalement sur la reconnaissance du coude alors que celle de TruB repose sur le positionnement du bras TΨC dans le site actif grâce au domaine « pouce », le mécanisme d'extrusion de bases et les nombreuses interactions spécifiques avec les bases de la boucle TΨC. Au final, il existe donc un fort contraste entre ces deux enzymes ayant évolué indépendamment pour accommoder des substrats différents. TruB distingue le nucléotide substrat en employant des domaines périphériques et additionnels ainsi que des changements induits à la fois dans sa structure et celle de l'ARNt substrat tandis qu'un dimère de TruA distingue le coude des ARNt et positionne le bras anti-codon à proximité du site actif autorisant ainsi la modification de trois nucléotides dynamiquement accessibles grâce à la flexibilité du bras anti-codon.

1.2.5. Rôle des domaines additionnels N et C terminaux

Le domaine additionnel « pouce » présent chez TruB subit un important changement de conformation lors de l'interaction avec le grand sillon de la molécule ARN substrat et joue donc un rôle important dans la reconnaissance de la cible (Pan et al., 2003 ; Chaudhuri et al., 2004 ; Phannachet & Huang, 2004). Ainsi, l'association avec l'ARNt provoque un changement de conformation des deux éléments Ins1 et Ins2 qui n'interagissent plus entre eux mais avec la boucle TΨC (Phannachet & Huang, 2004). Par ailleurs, il a été démontré que des perturbations du site catalytique se propagent vers le domaine « pouce » (Hoang et al., 2005).

Dans les ARN:Ψ-synthases de la famille RluA, les domaines « pouce » et « index » saisissent l'ARN substrat comme pourrait le faire une pince (Fig. 45). Notons que ce domaine « index » bien que présent au sein des familles RsuA, RluA et TruA, est absent des familles TruB et TruD (Tableau 6). Le domaine « pouce », quant à lui, est absent des représentants de la famille RsuA. Les membres de la famille TruD présentent un domaine « pouce » matérialisé par le domaine « TruD » qui est particulier puisque représentant 40% de la protéine, en contraste donc avec les petits domaines observés pour TruB ou RluA. Il est très probable que des résidus du domaine TruD jouent un rôle essentiel qui reste à élucider.

Certaines familles d'ARN:Ψ-synthases présentent également des domaines périphériques. Ainsi, le domaine PUA localisé en C-ter de TruB est responsable de contacts non spécifiques avec la tige acceptrice de l'ARNt (Hoang & Ferre-D'Amare, 2001). Les sous-domaines C-ter de RluB ou RluD ainsi que le domaine N-ter S4 de RsuA et RluD sont également impliqués à un degré qui n'est pas encore déterminé dans le mécanisme de reconnaissance de la cible.

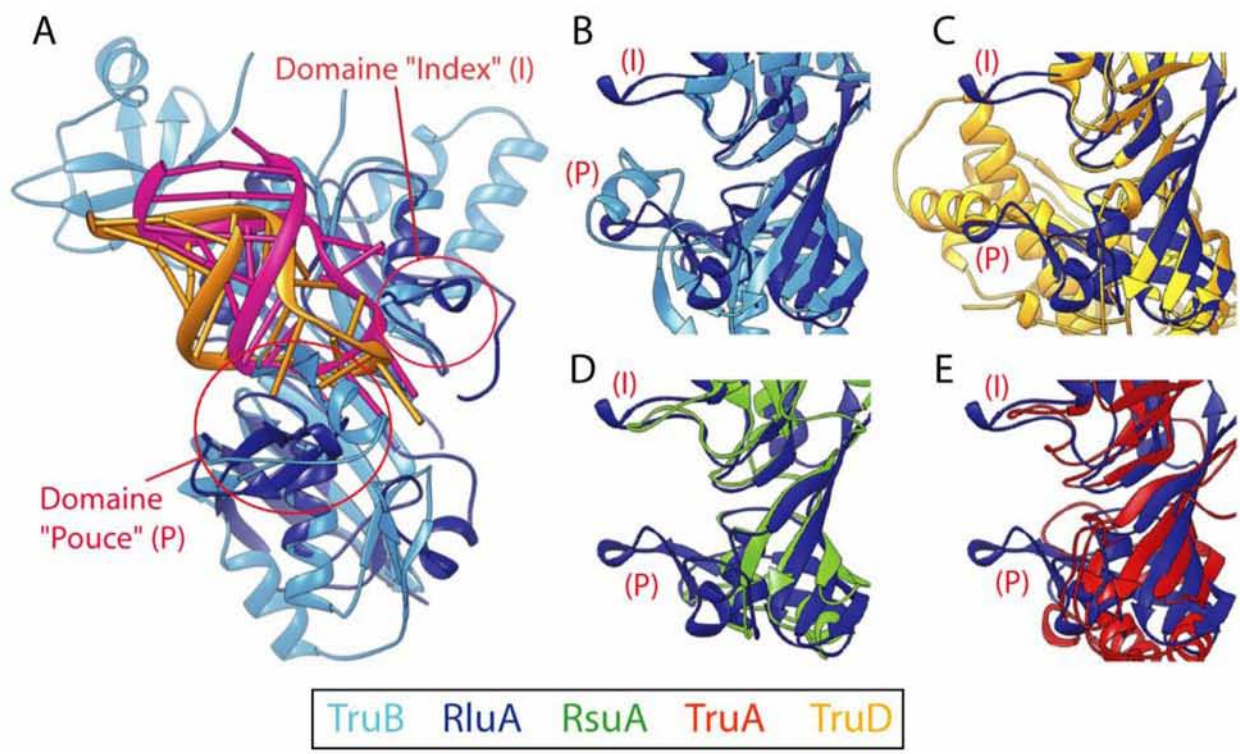


Figure 45 : Comparaison des domaines additionnels « pouce » et « index » par superposition des structures 3D des 5 familles d'ARN:Ψ-synthases (d'après (Hamma & Ferre-D'Amare, 2006)**.**
A. Superposition des structures 3D des enzymes RluA (bleu foncé) et TruB (bleu clair) liées à leur substrat boucle anti-codon (orange) et boucle TΨC (rose). Les domaines « pouce » et « index » sont encerclés en rouge. **B.** Superposition de TruB (bleu clair) et RluA (bleu foncé). **C.** Superposition de TruD (jaune) et RluA (bleu foncé). **D.** Superposition de RsuA (vert) et RluA (bleu foncé). **E.** Superposition de TruA (rouge) et RluA (bleu foncé).

Familles	Représentants	Extrémité N-ter	Domaine "index"	Séquence consensus du site actif	Domaine "pouce"	Extrémité C-ter
TruD	TruD	-	-	XAGXK**D**	TruD	-
TruA	TruA	-	+	XXX**RTD**	+	-
TruB	TruB	-	-	**H**XGX**LD**	+	Domaine PUA
RsuA	RsuA/RluE	Domaine S4	+	XXG**RLD**	-	-
	RluB/RluF	Domaine S4	+	XXG**RLD**	-	Domaine C-ter
RluA	RluA	-	+	XXH**RLD**	+	-
	RluC/RluD	Domaine S4	+	XXH**RLD**	+	Sous-domaine C-ter

Tableau 6 : Eléments caractéristiques des ARN:Ψ-synthases (d'après (Hamma & Ferre-D'Amare, 2006).

Il semble donc que différentes stratégies associées aux combinaisons de domaines « pouce », domaines « index », domaines périphériques et substitutions dans la séquence consensus du site catalytique confèrent une remarquable spécificité d'action aux ARN:Ψ-

synthases en dépit des nombreuses caractéristiques communes de leur domaine catalytique (Tableau 6).

2. ARN:Ψ-synthases associées à des guides ARN

Chez les eucaryotes, 3 des 4 ARN des ribosomes cytoplasmiques (5,8S, 18S et 26/28S) sont produits dans le nucléole à partir d'un transcrit unique (le pré-ARNr 35S chez la levure et 45S chez les mammifères) (Fig. 46). Ce précurseur est maturé par un processus complexe impliquant une cascade de clivages endonucléolytiques et de digestions par des exonucléases (pour revues, Morrissey & Tollervey, 1995; Venema & Tollervey, 1995; Tollervey, 1996). La formation des résidus Ψ et 2'-O-méthylés a lieu dans le pré-ARNr (Maden & Salim, 1974); Brand et al., 1979; Maden, 1990; Maden & Hughes, 1997) (Fig. 46). La synthèse de ces ARNr cytoplasmiques dans les cellules eucaryotes implique l'action d'un grand nombre de petits ARN stables, appelés snoRNA pour "small nucleolar RNA". Ils peuvent être regroupés en trois familles en fonction des séquences conservées qu'ils contiennent. On distingue les snoRNA à boîtes C/D, les snoRNA à boîtes H/ACA et l'ARN de la MRP, un complexe ARN/protéine de type ribozyme qui constitue à lui seul la troisième famille (pour revues, Morrissey & Tollervey, 1995; Tollervey & Kiss, 1997). Nous ne décrirons en détails que les snoRNA à boîtes H/ACA car seule cette famille a été abordée dans notre travail.

Les snoRNA à boîtes H/ACA comportent des ARN participant à la formation de résidus de résidus Ψ, mais aussi des ARN intervenant dans les étapes de clivage de l'ARNr précurseur, tels que les snoRNA à boîtes H/ACA snR10 et snR30 (chez la levure).

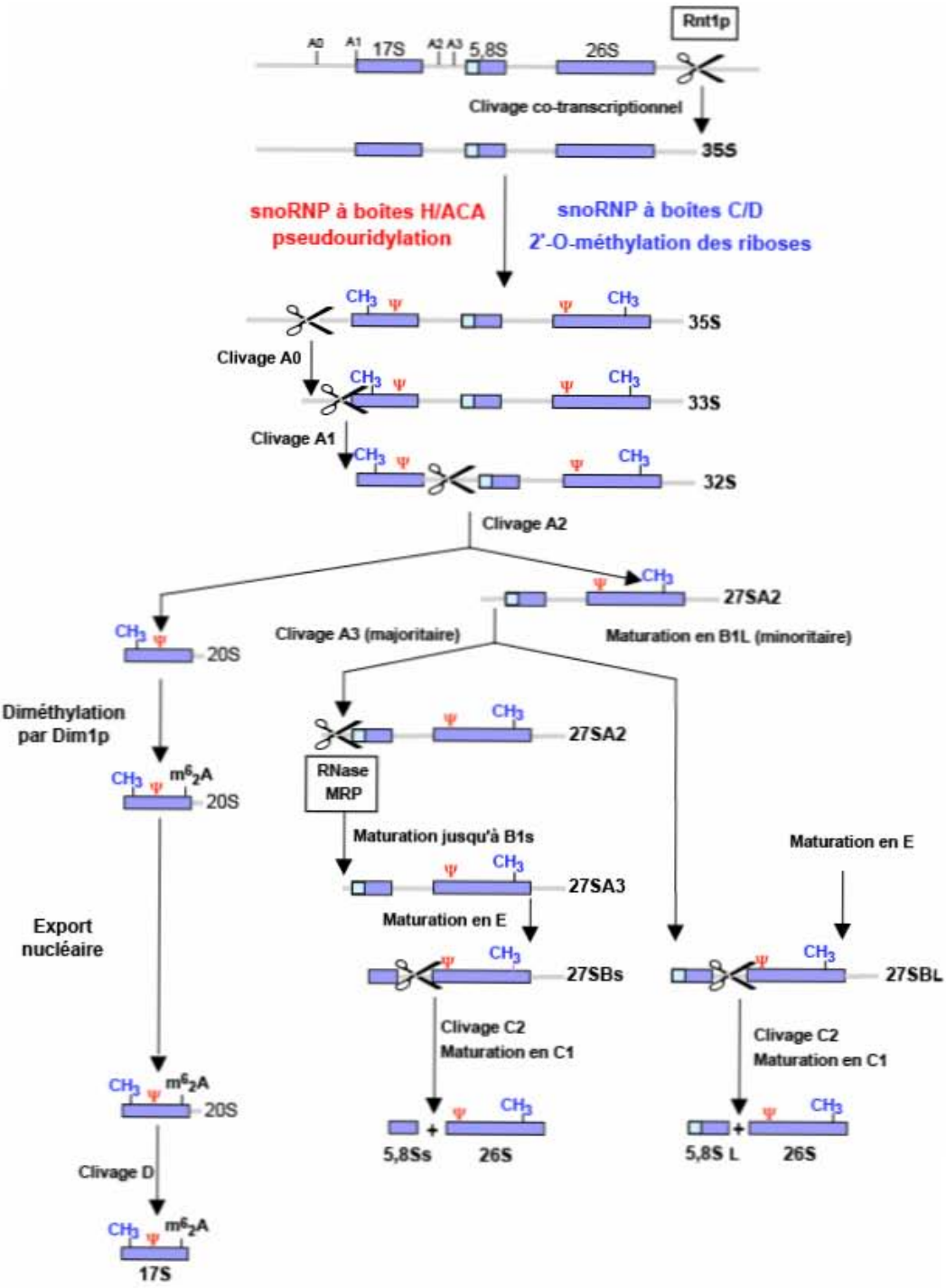


Figure 46 : Etapes de maturation du précurseur 35S des ARN ribosomiques de la levure *S. cerevisiae*.

Les étapes de modification du pré-ARNr sont indiquées. D'après (Lafontaine et al., 1998b), pour revue, (Venema & Tollervey, 1999).

Il est important de noter dans ce cas le snoRNA snR10 intervient à la fois dans les étapes de clivage et de modification du pré-ARNr (Li et al., 1990; Liang & Fournier, 1995; Kiss-Laszlo et al., 1996; Ni et al., 1997), alors que le snoRNA snR30 n'est impliqué que dans les étapes de clivages (pour revue, Tollervey & Kiss, 1997). Les données actuelles sont résumées dans plusieurs revues (Bachellerie & Cavaille, 1997; Bachellerie & Cavaille, 1998; Lafontaine &

Tollervey, 1998; Ofengand & Fournier, 1998; Kiss, 2001). On constate de plus, que sur la base de leur biosynthèse, les snoRNA se répartissent aussi en deux classes distinctes : ceux produits à partir d'un gène classique et ceux produits par maturation d'introns contenus dans des ARN pré-messagers (pour revues, Maxwell & Fournier, 1995; Tollervey & Kiss, 1997).

Des études ont montré que bien que les snoRNA guides soient impliqués dans la formation de la majorité des résidus Ψ et 2'-O-méthylés des ARN des ribosomes cytoplasmiques eucaryotes, il subsiste toujours un système enzymatique de nature purement protéique responsable de la formation d'un petit nombre de ces résidus. Ainsi, si l'on s'intéresse à la levure *S. cerevisiae,* il a été démontré que deux systèmes sont requis pour la méthylation du ribose en position 2918 de l'ARNr 23S: à la fois le système guidé, par l'intermédiaire du snoRNA C/D snR52, et un système enzymatique reposant sur la protéine Spb1 (Bonnerot et al., 2003). L'absence de formation du résidu 2'-O-méthylé en position 2918 n'est observée que lorsque les deux systèmes de modification sont simultanément déplétés (Bonnerot et al., 2003). De la même manière, des résultats récents ont mis en évidence qu'il existe 2 systèmes conjoints pour la formation des résidus Ψ dans les ARNr chez *S. cerevisiae*. Ainsi, des snoRNA à boîtes H/ACA ont été identifiés pour 45 des 46 sites présents dans les ARNr (Samarsky & Fournier, 1999; Badis et al., 2003, Torchet et al., 2005) et une seule autre enzyme impliquée dans la pseudouridylation des ARNr, autre que la protéine Cbf5/dyskérine associée aux snoRNA H/ACA, a été caractérisée (Decatur & Schnare, 2008). Il s'agit de la protéine Pus7p sur laquelle nous reviendrons de manière plus approfondie dans la partie « Résultats ».

2.1. Présentation générale

Il était connu qu'une classe d'ARN:Ψ-synthases appartenant à la famille TruB et nommée Cbf5p chez *S. cerevisiae*, Dyskérine chez l'homme, NAP57 chez les rongeurs (Meier & Blobel, 1994 ; Lafontaine et al., 1998a ; Watanabe & Gray, 2000; Wang & Meier, 2004) assure la pseudouridylation de nombreuses positions grâce à l'utilisation des ARN guides appelés snoRNA à boîtes H/ACA. Il a été démontré expérimentalement que l'homologue aCBF5 d'archaea joue le même rôle (Charpentier et al., 2007). La position à modifier par cette classe d'ARN:Ψ-synthases est déterminée par la complémentarité entre la séquence de la boucle interne de la tige-boucle du sno/sRNA à boîtes H/ACA et de la séquence ciblée.

En plus des guides à boîtes H/ACA, l'enzyme CBF5 et ses homologues nécessitent l'intervention de trois autres protéines dites auxiliaires appelées Nop10p, Gar1p et Nhp2p chez la levure, NOP10, GAR1 et NHP2 chez l'Homme et aNOP10, aGAR1 et L7Ae chez les Archaea (Henras et al., 1998 ; Watkins et al., 1998a ; Watanabe & Gray, 2000).

La partie qui suit vise à décrire les caractéristiques des snoRNA guides de pseudouridylation et celles des 4 protéines impliquées. Des études structurales très récentes,

réalisées chez les archaea à partir de différentes combinaisons des 4 protéines provenant de trois espèces euryotes : *P. abyssi*, *P. furiosus* et *M. jannaschii*, avec ou sans un fragment ARN mimant un sRNA à boîtes H/ACA et toujours sans une molécule d'ARN ciblée, ont permis de définir :

- les acides aminés de aCBF5, L7Ae et aNOP10, mais pas aGAR1, interagissant avec les sRNA à boîtes H/ACA
- les nucléotides du sRNA à boîtes H/ACA impliqués dans les contacts avec ces 3 protéines
- les résidus importants pour les 3 interactions protéines-protéines entre aNOP10 et L7Ae, aNOP10 et aCBF5, aCBF5 et aGAR1 (Charron et al., 2004 ; Hamma & Ferre-D'Amare, 2004 ; Hamma et al., 2005 ; Li & Ye, 2006; Manival et al., 2006 ; Rashid et al., 2006).

Très récemment, l'équipe de Ye a réussi à cristalliser un complexe comportant un fragment d'ARN mimant une tige-boucle d'un sRNA à boîtes H/ACA associé aux protéines L7Ae, aCBF5, aNOP10 et aGAR1 provenant de *P. furiosus* (Li & Ye, 2006). Cette structure, bien qu'elle n'ait pas été réalisée en présence d'un fragment ARN substrat, est actuellement la plus aboutie et sera utilisée comme référence pour discuter des différentes interactions qui se produisent entre ces 4 protéines et le sRNA à boîtes H/ACA.

2.2. Structure des snoRNA H/ACA

2.2.1. Les snoRNA H/ACA eucaryotes

Ce groupe de snoRNA, comme cela a été dit précédemment, sert de guide pour l'isomérisation de résidus uridines en pseudouridines. Cette famille d'ARN a été difficile à caractériser, du fait du faible degré de conservation de séquence de l'ARN et de la complexité de la structure formée entre l'ARN et sa cible. En effet, les snoRNA H/ACA partagent une structure commune qui consiste en deux structures tige-boucle, séparées par une région simple brin (Fig. 47) (Balakin et al., 1996; Ganot et al., 1997a). Cette région simple brin, contient un élément de séquence conservée, la boîte H (consensus ANANNA) (Ganot et al., 1997a). Une seconde séquence conservée est retrouvée en 3' des snoRNA H/ACA, il s'agit de la boîte ACA (AAA ou AUA dans environ 10% des cas), qui est située exactement trois nucléotides en amont de l'extrémité 3' de l'ARN (Balakin et al., 1996; Ganot et al., 1997a).

Les ARN cibles s'apparient dans des boucles internes, appelées « poches de pseudouridylation », des deux structures tige-boucle, ce qui permet de définir le résidu de l'ARNr qui est converti en résidu Ψ. Au sein de la poche de pseudouridylation, l'ARN cible forme deux petites hélices (3 à 10 pbs) qui sont séparées par 2 nt non appariés (Ganot et al., 1997b; Ni et al., 1997). Le premier nucléotide non apparié est le résidu U converti en résidu Ψ. Les snoRNA H/ACA peuvent ainsi diriger la pseudouridylation vers deux sites distincts au sein des ARNr (Ganot et al., 1997b) (Fig. 47). Les boîtes H et ACA sont respectivement

localisées, en moyenne, 14 à 16 nts en aval des appariements hétérologues. Une augmentation de longueur de cette distance diminue l'efficacité de pseudouridylation (Ni et al., 1997). La réaction dépend à la fois de la sélection du résidu à modifier et de l'action de protéines fixées sur le snoRNA H/ACA, les boîtes conservées devant intervenir dans cette fixation (Ganot et al., 1997b; Ni et al., 1997). Une hypothèse admise est que la boîte H joue le même rôle que la boîte ACA, mais pour la sélection du résidu à modifier dans la poche de pseudouridylation 5' du snoRNA (Bortolin et al., 1999). De plus, les tige-boucles 3' et 5' semblent agir en synergie, car la délétion de l'un des domaines empêche la pseudouridylation du pré-ARNr dans l'autre domaine (Bortolin et al., 1999).

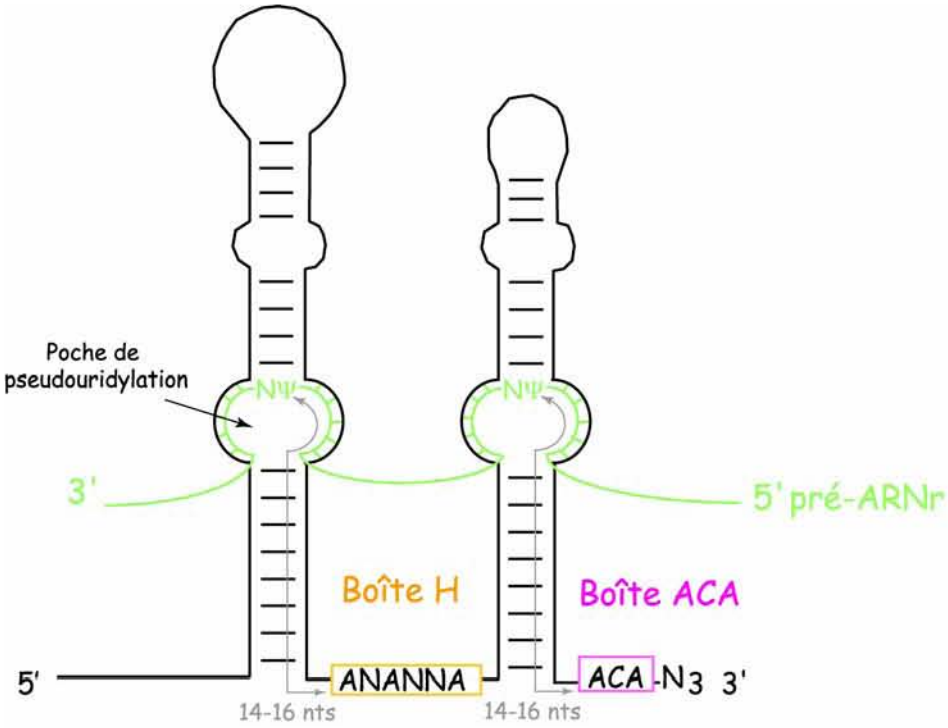


Figure 47 : Représentation schématique des complexes formés entre le pré-ARNr et les snoRNA à boîtes H/ACA.
Dans cette structure, l'uridine substrat est représentée par « Ψ ». La distance entre les boîtes H ou ACA et l'uridine cible varie entre 14 et 16 nucléotides (flèches grises). Les séquences consensus des différentes boîtes sont encadrées (pour revue, (Filipowicz & Pogacic, 2002)).

Les boîtes H et ACA sont absolument requises pour la stabilité des snoRNA, que ce soit chez les mammifères ou chez la levure, de même que les deux hélices à la base des deux structures tige-boucles (Balakin et al., 1996; Ganot et al., 1997a; Bortolin et al., 1999). L'altération de la distance entre la boîte ACA et l'hélice terminale en 3' des snoRNA ou de la boîte H par rapport à l'hélice 5' conduit à une diminution de la concentration cellulaire du snoRNA (Balakin et al., 1996; Bortolin et al., 1999). Ceci est en accord avec l'idée que la boîte H avec l'hélice 5' et la boîte ACA avec l'hélice 3', constituent les signaux de reconnaissance des protéines des snoRNP. En s'associant, ces protéines doivent protéger les

snoRNA contre l'action des endo et exonucléases et, de cette façon, elles interviendraient dans la définition des extrémités des snoRNA H/ACA lors des processus de biogenèse à partir des introns (Bortolin et al., 1999).

La caractérisation exhaustive des snoRNA à boîtes H/ACA de la levure *S. cerevisiae* a été réalisée récemment par deux approches, une approche bioinformatique et une approche expérimentale. L'approche bioinformatique est basée sur la description des snoRNA à boîtes H/ACA existants chez les levures ainsi que sur les positions précises des résidus Ψ détectés dans les ARNr de *S. cerevisiae* (Schattner et al., 2004 ; Schattner et al., 2005). L'approche expérimentale a consisté à produire les protéines communes Gar1p et Nhp2p sous forme fusionnée avec une étiquette TAP (tandem-affinity purification). Cette fusion a permis de réaliser une purification en tandem des deux protéines et des ARN associés. Les ARN associés ont été marqués puis hybridés sur une puce à ADN contenant des fragments recouvrant l'ensemble de l'ADN génomique de *S. cerevisiae* sous forme de fragments (Torchet et al., 2005). Ces études complémentaires ont révélé que 45 des 46 résidus pseudouridines présents dans les ARNr de *S. cerevisiae* sont guidés par 29 snoRNA à boîtes H/ACA (Tableau 7). Il s'agit de la recherche la plus aboutie pour un génome eucaryote.

2.2.2. Les sRNA H/ACA d'archaea

A première vue, le nombre de résidus Ψ détecté au sein des ARNr d'archaea est très similaire à celui trouvé chez les bactéries (4 résidus Ψ chez *H. halobium*, au moins 7 chez *S. acidocaldarius*, et 14 chez *P. abyssi*, contre 9 résidus Ψ chez *E. coli*) (pour revue, (Ofengand, 2002)). Cependant de tels ARN, appelés sRNA H/ACA existent chez les archaea.

Les premiers sRNA de type H/ACA ont été caractérisés en 2002, chez *Archaeglobus fulgidus*, lors d'une recherche systématique de petits ARN non-messagers (nmsRNAs) (Tang et al., 2002). Ainsi, 4 sRNA guidant la pseudouridylation de 6 positions distinctes au sein des ARNr 16S et 23S ont été décrits (Tang et al., 2002). Depuis, des sRNA ont été identifiés chez les 3 espèces du genre *Pyrococcus* (Rozhdestvensky et al., 2003). S. Muller et F. Leclerc, au sein de notre laboratoire, ont également mis en évidence 7 sRNA dans les trois espèces du genre *Pyrococcus*. L'expression de ces ARN a été testée par Northern blot et les cibles prédites ont été vérifiées expérimentalement (Muller et al., 2008).

Contrairement aux snoRNA H/ACA eucaryotes, qui comportent toujours 2 structures tige-boucles, les sRNA H/ACA peuvent être structurés en 1, 2 ou 3 tige-boucles, d'où l'existence de sRNA comportant soit une boîte ACA, soit une boîte H et une boîte ACA, soit encore une boîte H et deux boîtes ACA (Fig. 48). L'existence d'une boîte H n'est pas obligatoire, très souvent, elle est remplacée par une boîte ACA. Les sRNA H/ACA d'archaea ont également la particularité de posséder un motif K-turn, motif généralement retrouvé uniquement dans les ARN guides de type C/D (Fig. 48) (Rozhdestvensky et al., 2003). Ce

Position modifiée	snoRNA	Prédiction Bioinformatique	Vérification expérimentale
SSU-106	snR44	A	B, G
SSU-120	snR49		B, G
SSU-211	snR49		B, G
SSU-302	snR49		B, G
SSU-466	snR189	A	B, G
SSU-632	snR161		B, G
SSU-759	snR80		G
SSU-766	snR161		B, G
SSU-1000	snR31		C
SSU-1179	snR85	B	G
SSU-1185	snR36		A
SSU-1189	snR35	A	nr
SSU-1289	RUF3/snR83		B, G
SSU-1414	RUF3/snR83		G
LSU-775	snR80		B, G
LSU-959	snR8		C
LSU-965	snR43	D	G
LSU-985	snR8		C
LSU-989	snR49	A	B, G
LSU-1003	snR5		A
LSU-1041	snR33		C
LSU-1051	snR81		B
LSU-1055	snR44	A	B, G
LSU-1109	RUF2/snR82		G
LSU-1123	snR5		A
LSU-2128	snR3		B, G
LSU-2132	snR3	C	B, G
LSU-2190	snR32		C
LSU-2257	snR191		E
LSU-2259	snR191		E
LSU-2263	snR3		C
LSU-2265	RUF1/snR84		B, G
LSU-2313	snR86		G
LSU-2339	snR9		D
LSU-2348	RUF2/snR82	F	B, G
LSU-2350	RUF2/snR82		B, G
LSU-2415	snR11		D, G
LSU-2730	snR189		D, G
LSU-2822	snR34		D, G
LSU-2861	snR46		C
LSU-2876	snR34		C
LSU-2919	snR10		C
LSU-2940	snR37		A
LSU-2971	snR42		C
5,8S LSU-73	snR43		H

Tableau 7 : Les 28 snoRNA à boîtes H/ACA identifiés chez *S. cerevisiae.*

L'identité de l'ARNr cible, la position pseudouridylée et le nom des équipes ayant contribué à sa découverte sont indiqués : A=Ganot et al. (1997); B=Schattner et al. (2004), C=Ni et al. (1997); D= Ofengand et Fournier (1998); E= Badis et al. (2003); F=McCutcheon et Eddy (2003), G= Torchet et al. (2005), H= Decatur et Schnare (2008).

motif K-turn n'est pas trouvé au niveau des boîtes conservées H/ACA, mais est présent dans l'hélice située au-dessus des poches de pseudouridylation (Fig. 48). Nous allons voir par la suite, que la nature des protéines associées à ces ARN peut expliquer la présence de ce motif K-turn.

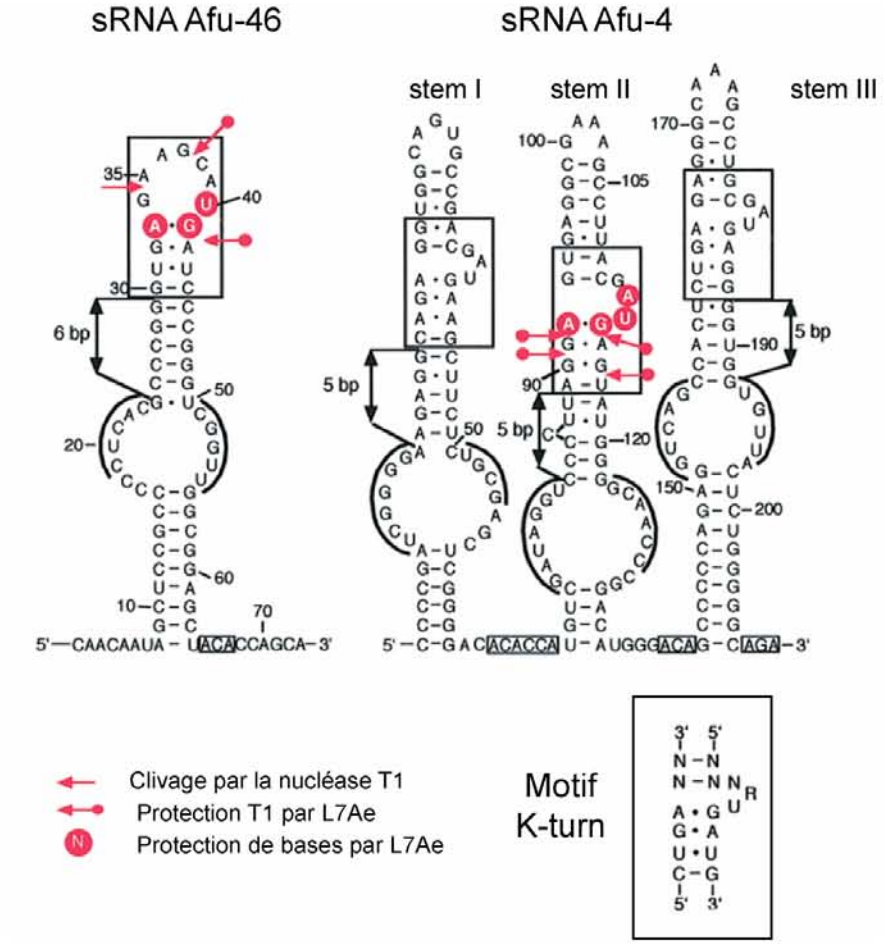


Figure 48 : Structure secondaire des sRNA Afu-46 et Afu-4 d'*Archaeglobus fulgidus*.
Les boîtes H et ACA sont encadrées et les séquences complémentaires à l'ARN cible sont indiquées. Le motif K-turn est encadré, et le consensus de ce motif au sein des sRNA C/D est indiqué pour comparaison. Les protections observées lors de l'association avec la protéine L7Ae sont également représentées. D'après (Rozhdestvensky et al., 2003).

2.3. Caractéristiques des 4 protéines des sno/sRNP à boîtes H/ACA

2.3.1. CBF5 (Cbf5p / aCBF5 / NAP57 / Dyskérine)

L'ARN:Ψ-synthase CBF5 a une taille comprise entre 40 kD (archaea) et 65 kD (levure). La représentante archaea présente environ 35 % d'identité de séquence avec TruB et adopte une structure très similaire à celle de TruB, d'*E. coli* (Fig. 49) (Hoang & Ferre-D'Amare, 2001 ; Hamma et al., 2005 ; Manival et al., 2006 ; Rashid et al., 2006). Le domaine catalytique de aCBF5 est composé, comme TruB de deux sous-domaines (D1 et D2) qui se rejoignent au

niveau d'un sillon comportant le site catalytique. Le site actif de aCBF5 se superpose bien avec celui de TruB et le résidu His responsable de l'extrusion de bases est bien conservé. Le domaine « pouce » de TruB, constitué par les deux éléments d'insertion Ins1 et Ins2, est absent ou raccourci dans les enzymes CBF5 d'archaea et d'eucaryotes. En effet, la boucle β7 de aCBF5 peut remplacer en partie le domaine d'insertion 2 de TruB. Le domaine PUA de CBF5 est plus grand que celui de TruB. Par comparaison avec TruB, une trentaine de résidus supplémentaires situés en N-ter entourent le domaine PUA, et quelques résidus supplémentaires sont aussi présents à l'extrémité C-ter (Fig. 46) (Manival et al., 2006). Ces extensions terminales sont importantes pour la stabilité et/ou le repliement de aCBF5. Il est à noter également que de nombreuses mutations responsables de la dyskératose congénitale sont localisées dans le domaine PUA localisé à l'extrémité C-ter ainsi que dans l'extrémité N-ter de aCBF5 qui entoure ce dernier. Le domaine PUA flanque une crevasse située du côté opposé au site actif et qui est impliquée dans l'interaction avec le sRNA.

Chez la levure, Cbf5p contient un domaine KKE/D dans sa région C-terminale, qui n'est pas retrouvé dans les protéines homologues de certains eucaryotes et archaea (Winkler et al., 1998; Watkins et al., 1998a). Le rôle de ce domaine n'est pas élucidé et sa délétion n'a pas d'effet sur la viabilité des cellules de *K. lactis* (Winkler et al., 1998). Néanmoins, il faut noter que le domaine KKE/D est retrouvé dans un certain nombre d'autres protéines associées à des petits ARN (Nop56p et Nop58p associées aux snoRNA à boîtes C/D, et Prp31p associée au snRNA U4 spliceosomal). Cependant, il ne semble pas impliqué dans des interactions avec des ARN.

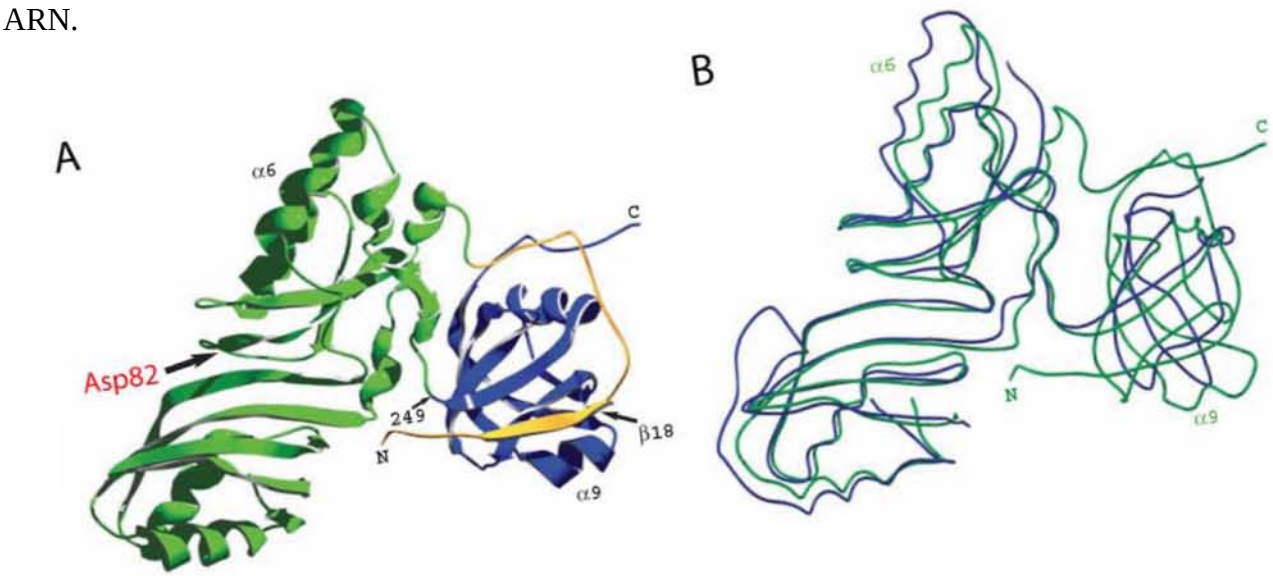


Figure 49 : Structure 3D de aCBF5 établie à partir d'un cristal du complexe aCBF5-aNOP10 de *P. abyssi* et comparaison avec la structure 3D de TruB de *Thermotoga maritima* (d'après (Manival et al., 2006).

A. Représentation de la structure 3D de aCBF5. L'extension en N-ter est représentée en jaune. **B.** Superposition de la structure 3D de aCBF5 (en vert) avec la structure 3D de TruB libre (en bleu).

2.3.2. NOP10 (Nop10p / aNOP10)

Il s'agit d'une petite protéine (entre 6 et 10 kDa) essentielle aux clivages A1 et A2 des pré-ARNr et à l'activité de pseudouridylation des sno/sRNA H/ACA (Henras et al., 1998 ; Charpentier et al., 2005) ; nos résultats). Elle a été identifiée avec Nhp2p par purification des snoRNP à boîtes H/ACA de la levure grâce à une version étiquetée de Gar1 (Gar1-ZZ) (Henras et al., 1998). Cette protéine, conservée chez tous les archaea et eucaryotes, ne présente aucun homologue en dehors des protéines NOP10. La comparaison des séquences de NOP10 d'eucaryotes et d'archaea révèle la conservation de la partie C-ter composée d'une hélice α mais révèle aussi de nombreuses différences dans le domaine N-ter. En effet, aNOP10 comporte un doigt à Zinc du côté N-ter chez les archaea qui est remplacé par une boucle chez les eucaryotes (Hamma et al., 2005; Khanna et al., 2006). Les deux domaines sont reliés par une région « linker » qui paraît désordonnée (Fig. 50) (Hamma et al., 2005 ; Manival et al., 2006).

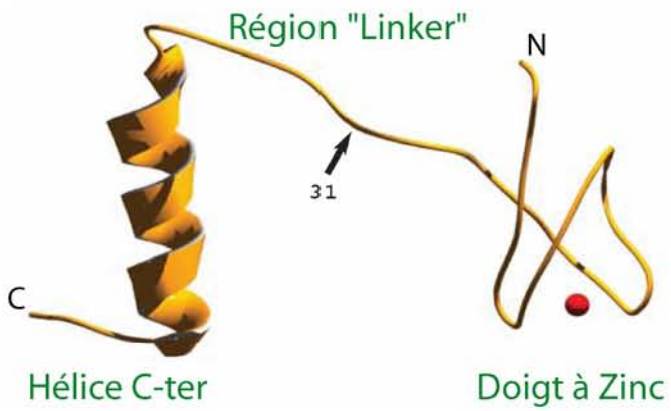


Figure 50 : Structure 3D de la protéine aNOP10 établie à partir d'un cristal du complexe aCBF5-aNOP10 de *P. abyssi* (d'après (Manival et al., 2006)**.** L'ion Zinc est représenté en rouge. La flèche indique la séparation entre les domaines N et C-ter.

2.3.3. Nhp2p / L7Ae

Les protéines Nhp2p et L7Ae présentes respectivement chez les eucaryotes et les archaea, appartiennent, comme les protéines Snu13p de levure et 15,5 kD humaine, à la famille L7Ae/L30. Leur taille varie entre 7 kD (archaea) et 22 kD (eucaryotes). La protéine L7Ae de *M. jannaschii* présente 45% d'identité de séquence avec la protéine NHP2 humaine (Henras et al., 1998). Bien que Nhp2p ait des propriétés physico-chimiques semblables à celles des protéines HMG (High Mobility Group), elle ne présente aucune similarité de séquence significative avec ces protéines (Kolodrubetz & Burgum, 1991). La protéine Nhp2p est associée à tous les snoRNA à boîtes H/ACA et elle est essentielle à ces complexes puisque son absence conduit à la déstabilisation des snoRNP et à l'absence d'activité (Henras et al., 1998 ; Watkins et al., 1998a). Contrairement aux protéines L7Ae, 15,5 kD et Snu13p qui s'associent aux motifs K-Turn, Nhp2p ne reconnaît pas spécifiquement ce motif. En effet, il semble que la

présence d'une tige-boucle irrégulière est nécessaire et suffisante pour la fixation de Nhp2p (Henras et al., 2001). Les sites d'interaction entre Nhp2p et le snoRNA ne sont pas encore déterminés. La spécificité de la liaison de Nhp2p pourrait résulter de l'interaction avec d'autres facteurs protéiques de la snoRNP. Ainsi, il est proposé qu'*in vivo*, Nhp2p se fixe à la boîte H après interaction avec CBF5 et NOP10 (Wang & Meier, 2004).

2.3.4. GAR1 (Gar1p / aGAR1)

Il s'agit d'une protéine nucléolaire de 25 kD chez les eucaryotes, essentielle à la viabilité cellulaire, mais non nécessaire à la stabilité *in vivo* des snoRNA H/ACA. Cette protéine contient deux domaines renfermant des répétitions de séquences glycine/arginine : les domaines GAR, localisés à chacune de ses extrémités. Ces domaines GAR, présents dans de nombreuses protéines nucléolaires comme la nucléoline ou la fibrillarine (pour revue, Ginisty *et al*., 1999), participent à la localisation nucléolaire de ces protéines et ont la capacité de fixer aspécifiquement l'ARN. Cependant, dans le cas de Gar1p, ces domaines ne sont nécessaires ni à la localisation nucléolaire, ni à la fixation sur les snoRNA (Girard et al., 1994 ; Bagni & Lapeyre, 1998). La protéine aGAR1 d'archaea d'une taille d'environ 11 kD ne présente pas ces domaines GAR mais partage avec les homologues eucaryotes un domaine central qui adopte une structure en tonneau à 6 brins β. Ce dernier ressemble à des domaines de fixation à l'ARN présents dans des facteurs bactériens d'initiation et d'élongation de la traduction. Chez les eucaryotes, il a été proposé que Gar1p se fixe au snoRNA (Bagni & Lapeyre, 1998 ; Wang & Meier, 2004) alors que chez les archaea aGAR1 n'interagit pas directement avec l'ARN (Li & Ye, 2006). La différence pourrait provenir de l'absence de domaine GAR.

2.4. Structure des sno/sRNP H/ACA

2.4.1. Les snoRNP H/ACA eucaryotes

Plusieurs approches de purification biochimiques ont permis de caractériser les constituants protéiques des snoRNP H/ACA. Ainsi, dès 1995, Lübben et collaborateurs ont pu purifier par immunoaffinité la snoRNP snR30 de *S. cerevisiae* et caractériser, après plusieurs étapes de purification sur résine MonoQ, quatre protéines de 10, 23, 25 et 65 kDa, constituant la partie stable de la snoRNP (Lubben et al., 1995). La protéine de 25 kDa s'est avérée être une protéine nucléolaire déjà connue, Gar1 (Girard et al., 1992) et les protéines de 65 kDa et 23 kDa correspondent respectivement aux protéines Cbf5 et Nhp2 (Watkins et al., 1998a). L'équipe de R. Lührmann a pu visualiser des particules purifiées par cryomicroscopie électronique et faire une modélisation 3D de ces particules à partir de ces données (Fig. 51). Ces particules présentent une symétrie bilatérale, ce qui suggère que chacun des domaines symétriques corresponde à l'assemblage des protéines autour d'une des structures tige-

boucles, d'où vraisemblablement une stoechiométrie d'ordre 2 des protéines des snoRNP H/ACA (Fig. 51B) (Watkins et al., 1998a).

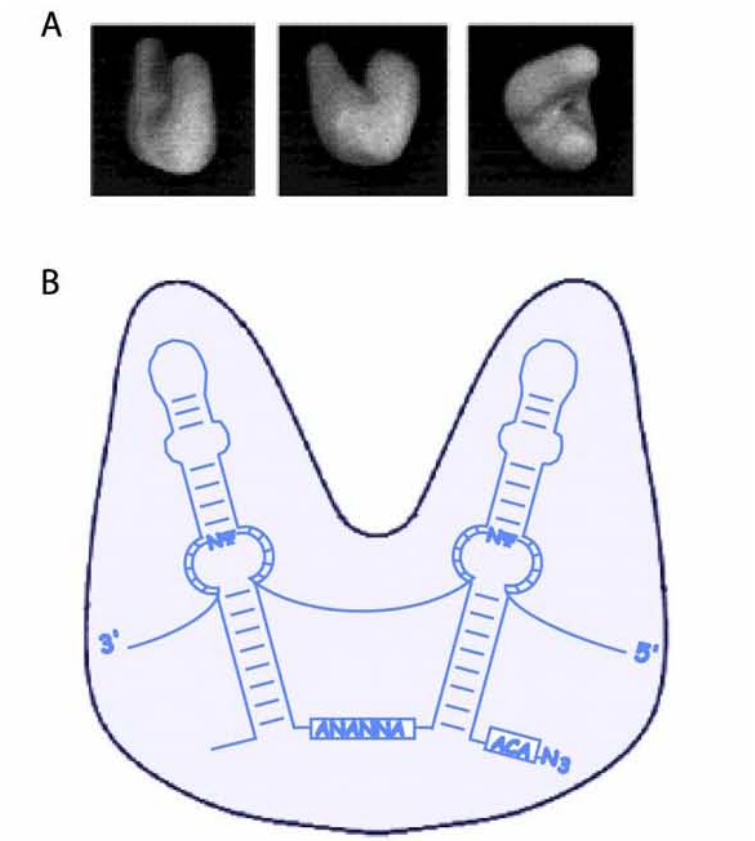


Figure 51 : Modélisation d'une snoRNP H/ACA réalisée à partir de résultats de cryomicroscopie électronique obtenus pour les snoRNP snR30 et snR42. D'après (Watkins et al., 1998a).

A. Trois vues différentes du modèle proposé.
B. Interprétation du volume observé en tenant compte de la structure typique d'un snoRNA H/ACA.

Comme nous l'avons dit précédemment, les snoRNA eucaryotes diffèrent des sRNA d'archaea par l'absence de motifs en K-turn/K-loop, la présence d'une boîte H et leur architecture, le plus fréquemment, en deux structures tige-boucles. Les snoRNA eucaryotes sont capables de se lier à deux molécules de CBF5 (Watkins et al., 1998a)). Or, la mutation de la boîte ACA terminale des snoRNA inhibe complètement la fixation de CBF5 (Baker et al., 2005). Ceci suggère un couplage de la fixation de CBF5 sur chacune des structures tige-boucles et, certainement, des autres protéines associées à ces structures organisées en deux tige-boucles.

L'absence de reconnaissance spécifique des ARN par NHP2, l'association stable de NHP2 avec le complexe CBF5-NOP10 ainsi que la fixation directe de GAR1 sur le snoRNA sont trois autres différences déjà évoquées ci avant qui ont conduit l'équipe de U. Meier à proposer

un mécanisme d'association des particules snoRNP à boîtes H/ACA différent de celui qui est pressenti pour les archaea (Fig. 52) (Wang & Meier, 2004) ; (Darzacq et al., 2006).

En effet, chez les mammifères, la formation du dimère NAP57-NOP10 est requise pour le recrutement de NHP2 sur le snoRNA. Il a également été démontré chez les eucaryotes que GAR1 est en compétition avec une autre protéine, NAF1, qui comporte d'importantes similarités structurales avec GAR1 (Leulliot *et al.*, 2007). NAF1 peut aussi s'associer au dimère NAP57-NOP10. Elle s'associerait précocement au cours de la biosynthèse des particules et maintiendrait les particules sous forme inactives dans les premières étapes de biogenèse et de leur transport dans les corps de Cajal (Darzacq et al., 2006 ; Leulliot et al., 2007). Par conséquent, un complexe protéique contenant NOP10, NAP57, NHP2 et NAF1 formé dans le cytoplasme pourrait être importé dans le noyau par l'intermédiaire de la protéine NAF1 avant que GAR1 remplace NAF1 (Fig. 52). La protéine SHQ1 est comme NAF1 essentielle à la biogenèse des particules snoRNP à boîtes H/ACA de levure mais ne fait pas partie de la particule mature (Dez et al., 2002 ; Fatica et al., 2002 ; Yang et al., 2002). L'interaction de NAF1 avec le domaine CTD de l'ARN pol II suggère que ces deux protéines pourraient recruter les protéines CBF5 et NOP10 sur un transcrit snoRNA naissant (Fig. 52). Une interaction possible de SHQ1 avec RNT1, RNase double-brin homologue de la RNase III bactérienne qui clive en amont et en aval des pré-snoRNA (Chanfreau et al., 1998), suggère aussi une implication de SHQ1 dans le couplage maturation/assemblage des snoRNP H/ACA. Par ailleurs, les interactions entre GAR1 et RNT-1 sont importantes pour la localisation nucléolaire de GAR1 (Tremblay et al., 2002). La protéine SMN (Survival of Motor Neuron) absente de *S. cerevisiae* mais présente chez de nombreux eucaryotes interagit avec GAR1 via ses répétitions GAR et pourrait également être impliquée dans l'assemblage des particules snoRNP H/ACA en tant que protéine chaperonne (Pellizzoni et al., 2001 ; Terns & Terns, 2001).

Enfin, les études menées par E. Bertrand et T. Kiss suggèrent que la biogenèse et le transit dans différentes régions du noyau fait intervenir un grand nombre de protéines autres que celles contenues dans les particules matures. Malgré ces différences, la conservation entre archaea et eucaryotes de nombreux résidus importants pour les interactions présentées ci avant suggère un type d'association de CBF5-NOP10-GAR1 similaire au sein des particules actives des deux domaines du vivant.

Récemment, deux analyses réalisées indépendamment par RMN des deux composants eucaryotes solubles, le snoRNA U65hp humain et la protéine NOP10 de *S. cerevisiae*, ainsi que des expériences de cartographie par modification chimique des interactions entre ces deux composants ont suggéré l'interaction directe de la protéine Nop10 de levure avec la région constituée de la jonction entre la tige basale (P1) et l'élément antisens 3' de la poche de pseudouridylation du snoRNA (Khanna et al., 2006). La protéine Nop10p interagirait par son extrémité N-ter qui, contrairement à aNOP10 d'archaea ne présente aucun doigt à Zinc. Les

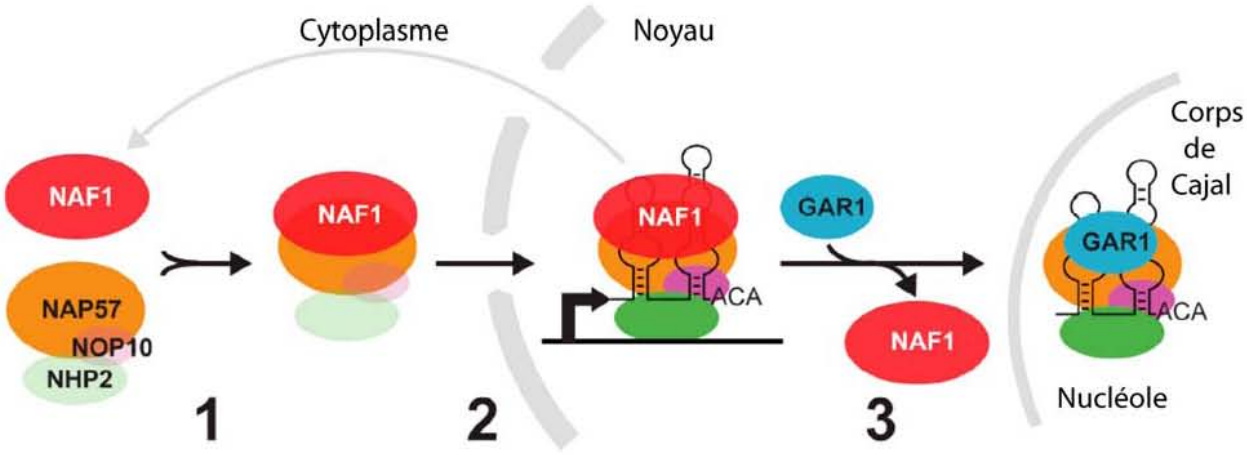


Figure 52 : Modèles d'assemblage des snoRNP d'eucaryote (d'après (Darzacq et al., 2006)**.**
Le modèle a été proposé pour l'assemblage des snoRNP H/ACA chez l'homme. **1.** La protéine NAF1 s'associe dans le cytoplasme avec la protéine NAP57/Dyskérine/CBF5 déjà liée aux protéines NOP10 et NHP2. **2.** Le complexe est recruté sur l'ARN en cours de transcription. **3.** La protéine NAF1 serait remplacée par la protéine GAR1 pour former la snoRNP H/ACA mature dans les corps de Cajal.

auteurs proposent que cette interaction permette une ouverture de cette poche facilitant l'hybridation de l'ARN cible. Cependant, il reste envisageable que les protéines NOP10, CBF5 et NHP2 se lient au snoRNA sous forme de complexes préformés. Une première analyse de l'association de CBF5 sur le snoRNA a été réalisée grâce à une protéine Cbf5p de levure tronquée (Normand et al., 2006). En effet, le domaine KKE/D en C-ter (138 aa des 483 aa) dont la fonction inconnue n'est pas essentielle pour la survie cellulaire a été retiré ce qui rend soluble la protéine recombinante Cbf5p phrase très lourde est difficile à suivre. Cette étude a révélé que CBF5 s'associe au snoRNA d'une manière similaire à ses homologues d'archaea, ceci en interagissant principalement avec des résidus des boîtes H et ACA et de l'hélice basale (P1).

2.4.2. Les snRNP H/ACA d'archae

La structure 3D des snRNP d'archaea est mieux connue et il a été démontré que le complexe ressemble à un triangle équilatéral dont les côtés mesurent environ 80-90 Å de longueur et 20 Å d'épaisseur (Fig. 53). Le site catalytique de aCBF5 se trouve au centre et les trois angles sont représentés par le domaine PUA, L7Ae et aGAR1 (Li & Ye, 2006). La protéine aNOP10 est prise en « sandwich » entre L7Ae et aCBF5. La tige-boucle de l'ARN s'étend sur 80 Å et se trouve encerclée par L7Ae, aNOP10 et aCBF5 au niveau de l'hélice apicale, et par le domaine PUA de aCBF5 au niveau de l'hélice basale. Leur fixation provoque également une courbure de 50° de l'axe de la tige-boucle. La poche de pseudouridylation se positionne au niveau du site catalytique de aCBF5 de telle sorte que l'élément antisens 5' se situe loin de la surface de aCBF5 alors que l'élément antisens 3' se situe à proximité de

aCBF5. Les 4 protéines s'associent sur une seule face du sRNA autorisant ainsi un accès à l'ARN substrat qui peut entrer et sortir du complexe sans que les 4 protéines ne se dissocient du sRNA. L'ensemble des interactions relevées entre les 4 protéines et l'ARN est détaillé dans les figures 53 et 54.

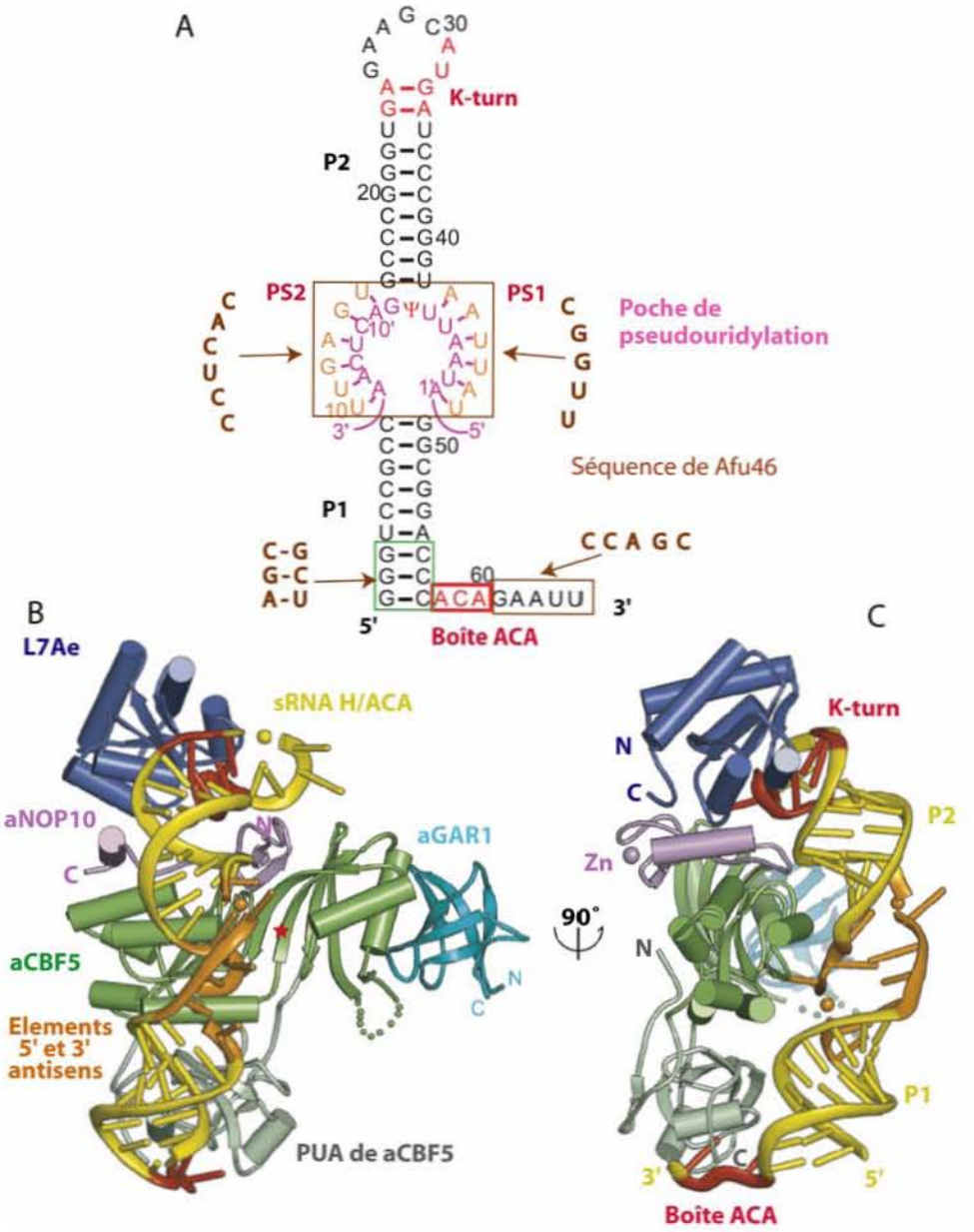


Figure 53 : Structure 3D à haute résolution d'un complexe comportant un sRNA à boîtes H/ACA, aCBF5, aNOP10, aGAR1 et L7Ae établie par radiocristallographie (adapté de (Li & Ye, 2006). **A.** Structure secondaire du sRNA à boîtes H/ACA utilisé pour cette étude. Il s'agit du sRNA Afu46 provenant de *A. fulgidus* auquel des modifications ont été apportées : les séquences de la poche de pseudouridylation, de la base de l'hélice basale (P1) et de la séquence terminale en 3' ont été modifiées (encadrées en marron). Les hélices formées lors de l'appariement de l'ARN substrat (rose) avec les éléments antisens (orange) sont appelées PS1 et PS2. Le motif K-turn et la boîte ACA sont représentés

en rouge. **B.** et **C.** Structure 3D du complexe présentée selon deux angles de vue différents. L'étoile indique l'aspartate catalytique. Différentes couleurs sont utilisées pour le sRNA (jaune), le domaine principal de aCBF5 (vert), son domaine PUA (gris), L7Ae (bleu foncé), aNOP10 (rose) et aGAR1 (bleu clair). L'ARN substrat présenté en A n'est pas présent dans la structure 3D.

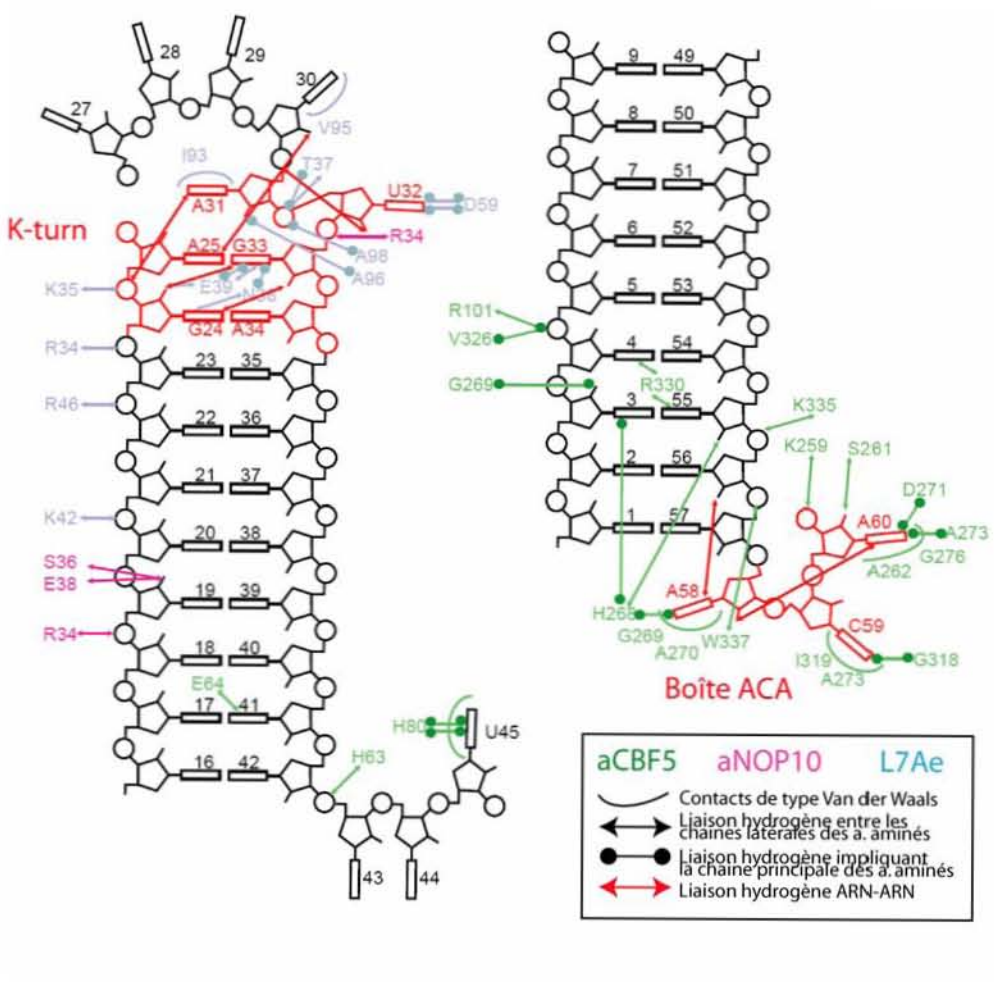


Figure 54 : Représentation schématique des interactions de du complexe sRNA-aCBF5-aNOP10-L7Ae-aGAR1 (d'après (Li & Ye, 2006).
Les nucléotides du sRNA qui ne contactent pas de résidus des protéines et qui sont désordonnés ne sont pas représentés.

L'ensemble des études structurales concernant les protéines des snRNP d'archaea, qu'il s'agisse des protéines seules, en duplexe ou encore de la particule entière, ont apporté de nombreuses informations concernant les interactions présentes entre les différents partenaires du complexe. Ainsi, concernant l'association de la protéine L7Ae sur le motif en K-Turn de l'ARN guide, il a pu être mis en évidence que le squelette polypeptidique de la protéine L7Ae libre de *P. abyssi* est très similaire à celui de la protéine humaine 15,5kD associée au snRNA U4 et à celui de la protéine L7Ae de *H. marismortui* dans la sous unité 50S (Fig. 56). Seuls de faibles réarrangements de chaînes latérales doivent avoir lieu lors de la fixation de l'ARN. En effet, la poche de fixation du nucléotide U proéminant dans la structure en «K-turn» est préformée dans

la protéine libre. Il apparaît que l'ARN subit de forts changements de conformation lors de sa fixation à la protéine, contrairement à la protéine qui ne subit que très peu de modifications. C'est donc l'ARN qui adapterait sa conformation à celle de la protéine (Charron et al., 2004).

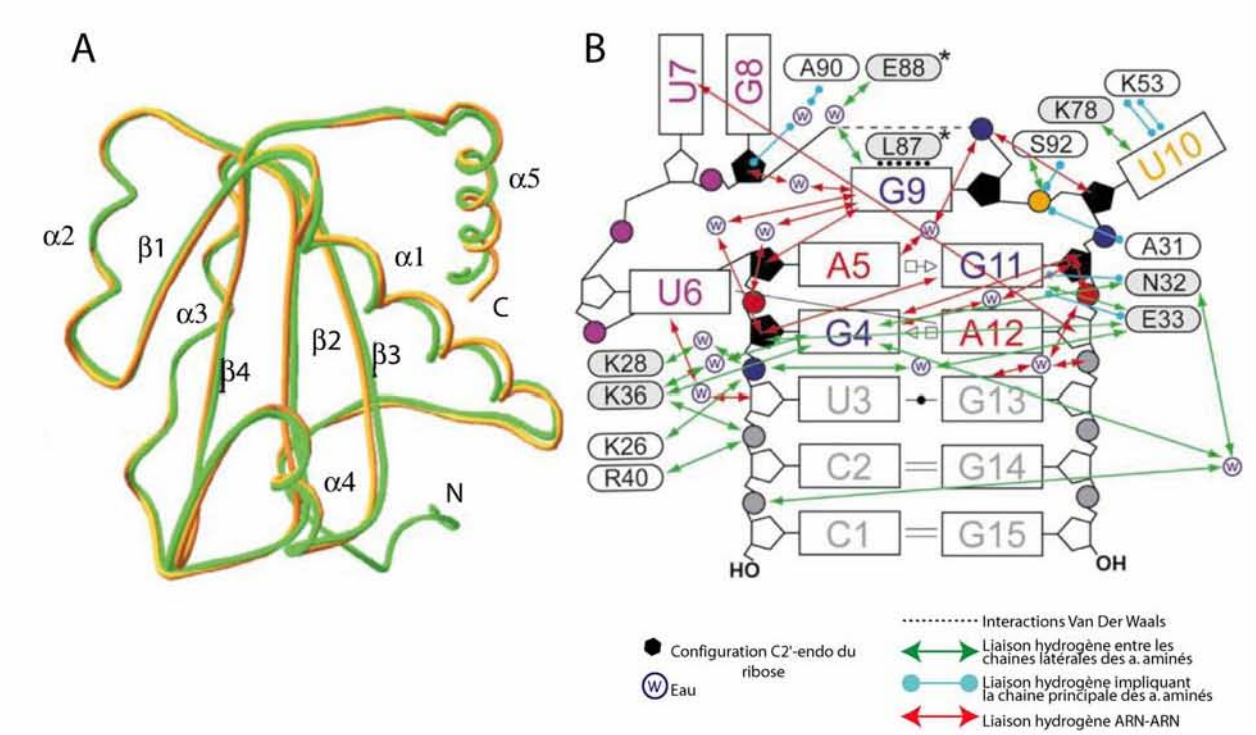


Figure 56 : Structures des protéines L7Ae libre ou en interaction avec un motif K-turn (d'après (Charron et al., 2004) ; (Hamma & Ferre-D'Amare, 2004).
A. Superposition des structures 3D des protéines L7Ae libre de *P. abyssi* (vert) et L7Ae lié au K-turn de *H. marismortui* (orange) (PDB : 1JJ2). **B.** Schématisation des interactions protéine-ARN et ARN-ARN mis en place lors de l'interaction de L7Ae de *M. jannaschii* et du motif K-turn de Afu190 (Tang et al., 2002). Les astérisques indiquent des résidus qui ne sont conservés qu'au sein des archaea.

D'autre part, il a été démontré que :

- aNop10 adopte une structure allongée et interagit sur toute une face avec le domaine catalytique de aCBF5 (Hamma et al., 2005).
- aGAR1 n'interagit qu'avec aCBF5 et pourrait jouer un rôle dans la modulation de la reconnaissance de la cible (Reichow et al., 2007).
- aNOP10 n'interagit pas seulement avec aCBF5 mais aussi avec L7Ae et le guide sRNA (Li & Ye, 2006).
- le domaine PUA de aCBF5 interagit avec le trinucléotide ACA ainsi qu'avec le sillon mineur de l'hélice basale (P1) du sRNA (Fig. 57).

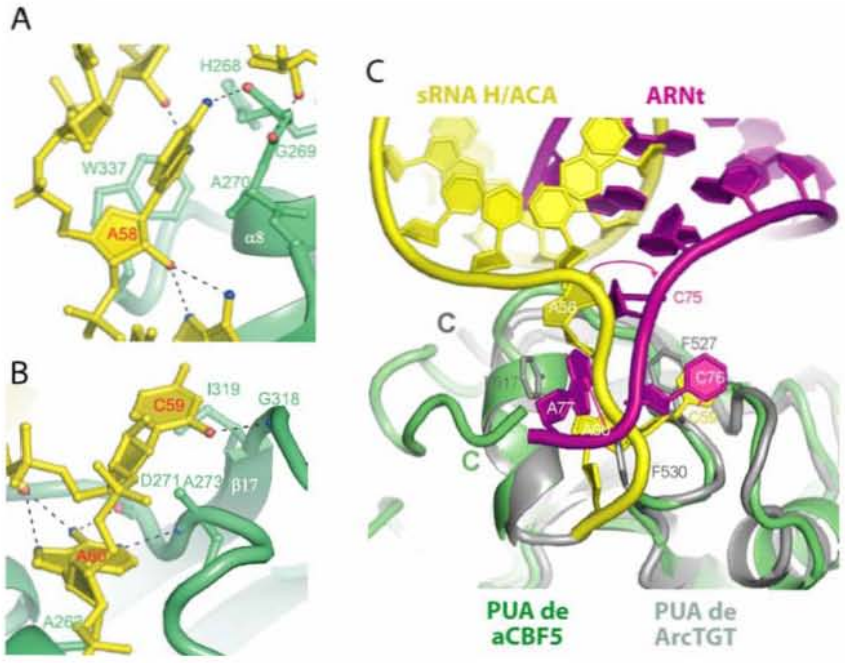


Figure 57 : Reconnaissance de la boîte ACA par le domaine PUA de aCBF5, comparaison à la reconnaissance du triplet CCA de la tige acceptrice d'un ARNt par le domaine PUA de Arc-TGT (Archaeosine tRNA-guanine transglycosylase) (d'après (Li & Ye, 2006).
A. Interaction du résidu A58 (1er résidu de la boîte ACA) avec les résidus du domaine PUA. **B.** Interaction des résidus C59 et A60 (2^{ième} et 3^{ième} résidus de la boîte ACA) avec les résidus du domaine PUA. **C.** Comparaison des interactions entre la boîte ACA d'un sRNA à boîtes H/ACA (en jaune) ou le triplet CCA terminal d'un ARNt (en violet) avec, respectivement, le domaine PUA de aCBF5 (en vert) ou le domaine PUA de ArcTGT (en gris).

- 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 (Fig. 58) (Jin et al., 2007; Wu & Feigon, 2007).

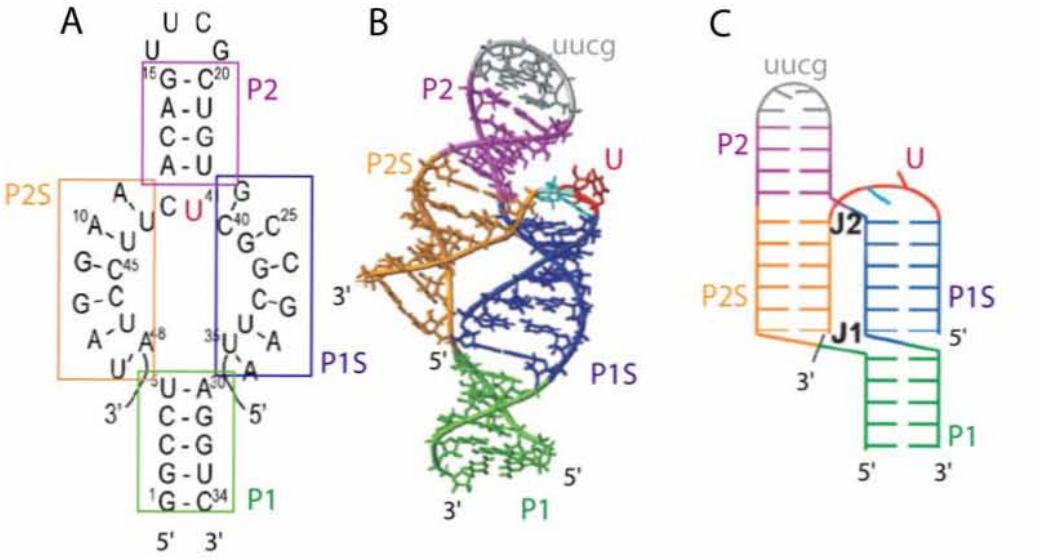


Figure 58 : Etude par RMN de l'association du sRNA et de l'ARN substrat (d'après (Wu & Feigon, 2007).
A. Structure secondaire du snoRNA U65hp associé à un fragment d'ARNr substrat (partie 4423-4435 de l'ARNr 28S humain). Le snoRNA U65hp correspond à la poche de pseudouridylation de la

deuxième tige-boucle du snoRNA U65 comportant une partie des hélices basales (P1) et apicales (P2). Une boucle uucg a été placée au sommet de l'hélice apicale. **B.** Structure en solution du complexe U65hp associé au substrat ARN. Le résidu U substrat est représenté en rouge, le nucléotide en 3' du résidu U substrat est non apparié et représenté en bleu clair. Les hélices P1, P1S, P2S et P2 sont représentées respectivement en vert, bleu foncé, orange et rose. **C.** Schéma représentant l'empilement des hélices et la localisation des jonctions.

2.5. Attribution des sno/sRNA à boîtes H/ACA aux résidus Ψ présent dans les différentes classes d'ARN

2.5.1. Situation dans les UsnRNA

L'équipe de T. Kiss a mis en évidence l'existence d'une nouvelle classe de petits ARN portant des signatures d'ARN guide de modifications post-transcriptionnelles portées par les UsnRNA spliceosomaux transcrits par l'ARN polymérase II. En hybridation *in situ*, ces ARN sont exclusivement détectables dans les corpuscules de Cajal, d'où leur nom : scaRNA (small Cajal protéines) (Darzacq et al., 2002; Kiss et al., 2002). Les scaRNA sont soit des ARN guides à boîtes C/D ou H/ACA, soit des ARN composites possédant à la fois un domaine C/D et un domaine H/ACA. Le scaRNA U93 a même la particularité de contenir deux motifs H/ACA (Kiss et al., 2002). Il est à noter qu'aucun ARN composite de la sorte n'a été mis en évidence chez la levure. Le premier ARN guide caractérisé capable de modifier un snRNA, est l'ARN U85 (Jady & Kiss, 2001). Cet ARN guide à la fois, la méthylation du résidu 46 et la pseudouridylation du résidu 45 du snRNA U5 (Jady & Kiss, 2001). Il a été montré en 2002 que cet ARN était exclusivement localisé dans les corpuscules de Cajal (Darzacq et al., 2002). Le Tableau 8 présente les scaRNA caractérisés à ce jour. La taille de ces ARN, les types de boîtes présentes, ainsi que les positions modifiées au sein des UsnRNA sont indiquées. L'implication des scaRNA dans la modification des UsnRNA n'a pas systématiquement été vérifiée expérimentalement.

scaRNA	Taille (nts)	Type d'ARN	Modifications guidées	Références
U85	330	C/D et H/ACA	Cm45 dans le snRNA U5 Ψ46 dans le snRNA U5	Jady et Kiss, 2001
U87	277	C/D et H/ACA	Am65 dans le snRNA U4 Um41 dans le snRNA U5	Darzacq et al., 2002
U88	266	C/D et H/ACA	Um41 dans le snRNA U5	Darzacq et al., 2002
U89	270	C/D et H/ACA	Ψ46 dans le snRNA U5	Darzacq et al., 2002
U90	330	C/D	Am70 dans le snRNA U1	Darzacq et al., 2002
U91	83	C/D	Cm8 dans le snRNA U4	Darzacq et al., 2002
U92	131	H/ACA	Ψ34 dans le snRNA U2 Ψ44 dans le snRNA U2	Darzacq et al., 2002
U93	275	double H/ACA	Ψ54 dans le snRNA U2	Kiss et al., 2002

Tableau 8 : Les scaRNA mis en évidence chez l'homme.

Les scaRNA contiennent les boîtes C/D et/ou H/ACA retrouvées également chez les snoRNA. Ces boîtes ont été décrites comme étant essentielles pour la localisation nucléolaire des ces ARN.

Des ARN guides murins potentiels, guidant spécifiquement des modifications post-transcriptionnelles des UsnRNA, ont également été détectés (Huttenhofer et al., 2001). La localisation cellulaire de ces ARN n'ayant pas été faite, il n'est pas pour l'instant possible de les classer dans la catégorie des scaRNA. Cependant, certains de ces ARN sont des homologues des scaRNA décrits précédemment, on peut donc supposer qu'ils appartiennent aussi à la classe des scaRNA. Il s'agit des ARN MBI 57 (Ψ34 et Ψ44 dans le snRNA U2), MBI 125 (Ψ91 dans le snRNA U2), MBII 19 (Cm40 dans le snRNA U2), MBII 119 (Cm8 dans le snRNA U4) et MBII 382 (Gm11 et Cm61 dans le snRNA U2) (Huttenhofer et al., 2001). Les ARN MBI 57 et MBII 119 sont respectivement les homologues des scaRNA U92 et U91.

Concernant la levure *S. cerevisiae*, le snRNA U2 est le seul ARN qui soit modifié par un système de guide en dehors des ARNr. En effet, cet ARN possède 3 résidus Ψ : le résidu Ψ42 dont la formation est guidée par le snoRNA H/ACA snR81 (Ma et al., 2003) et les résidus Ψ35 et Ψ44 formés par les enzymes Pus7 et Pus1, respectivement (Massenet et al., 1999; Ma et al., 2003).

2.5.2. Situation dans les ARNt

Contrairement aux sRNA à boîtes C/D d'archaea, aucun sno/sRNA à boîtes H/ACA d'eucaryotes ou d'archaea modifiant un ARNt n'avait été mis en évidence. Cependant, nos travaux, présentés dans la partie « Résultats » démontrent qu'une famille de sRNA à boîtes H/ACA présent chez certains archaea (crénotes) est responsable de la modification de la position 35 des $ARNt^{Tyr}$.

2.5.3. Situation dans les ARNr

Comme nous l'avons décrit précédemment, les études réalisées dans le domaine ont permis la caractérisation de l'ensemble des snoRNA H/ACA responsable de la formation des résidus Ψ dans l'ARNr 28S et l'ARNr 5,8S de la grande sous unité et l'ARN 18S de la petite sous unité du ribosome chez la levure *S. cerevisiae*.

La situation dans le domaine archaea n'est pas aussi claire puisque aucune recherche systématique, expérimentale ou bioinformatique, de gènes de sRNA à boîtes H/ACA dans les espèces archaea n'a été réalisée.

5. Objectifs de la thèse

Comme nous l'avons décrit dans l'introduction, la plupart des ARN métaboliquement stables renferment des nucléotides modifiés. Les plus fréquents étant les résidus pseudouridines (Ψ) et 2'-O méthylés. Les résidus Ψ peuvent intervenir de façon cruciale dans de nombreux processus cellulaires et sont essentiellement trouvés dans des positions où ils contribuent, grâce à leur structure chimique particulière, à la stabilisation de motifs structuraux spécifiques. Il apparaît que la formation des résidus Ψ est assurée par 2 systèmes distincts qui sont bien conservés dans les 3 domaines du vivant. Le premier regroupe une série d'enzymes capable d'agir sur une grande variété de substrats ARN (ARNt, UsnRNA et ARNr) sans nécessiter de cofacteur. Le second fait appel à des guides sno/sRNA H/ACA qui sont associés à plusieurs protéines dont l'une est responsable de l'activité et qui assurent la formation de la majorité des résidus Ψ dans les ARNr.

Le premier volet de ma thèse correspond à la caractérisation des activités ARNt:Ψ-synthases de l'enzyme Pus7p de *S. cerevisiae*. En effet, les travaux menés au laboratoire ces dernières années on conduit à la caractérisation de la majorité des ARN:Ψ-synthases chez la levure *S. cerevisiae* ceci en utilisant la présence de motifs signatures qui avaient été établis à partir des connaissances sur les ARN:Ψ-synthases d'*E. coli*. Cependant, sur la base de ces expériences, aucune des 9 enzymes identifiées ne catalyse la formation des résidus Ψ présents dans les segments de haute importance fonctionnelle des UsnRNA spliceosomaux (positions 5 et 6 dans le snRNA U1, 35 dans le snRNA U2 et 99 dans le snRNA U5), dans les ARNt (positions 13 dans plusieurs ARNt cytoplasmiques et 35 dans le pré-ARNt^{Tyr}) et dans les ARNr 5S (position 50) et 5,8S (position 73) de la grande sous unité du ribosome. En 2003, l'équipe de Yu (Rochester) a caractérisé l'ARN:Ψ-synthase Pus7 de *S. cerevisiae*, qui est responsable de la formation du résidu Ψ35 dans le snRNA U2 (Ma et al., 2003) ceci en utilisant une banque de plasmides recombinants contenant l'ensemble des ORF de la levure. Comme il avait été montré au laboratoire qu'en plus de son action à 8 positions des ARNt (Motorin et al., 1998), l'enzyme Pus1p modifie le snRNA U2 de *S. cerevisiae* en position 44 (Massenet et al., 1999), nous avons testé, en collaboration avec l'équipe de Yu, si Pus7p agit aussi sur les ARNt. Nous avons ainsi montré, avec I. Behm-Ansmant, qu'il s'agit d'une enzyme ayant la capacité de modifier le snRNA U2 en position 35, les ARNt cytoplasmiques en position 13 et le pré-ARNt^{Tyr} en position 35.

De manière intéressante, la comparaison des différents substrats de Pus7p nous avait permis de mettre en évidence l'existence d'une séquence de 7 nucléotides (R G/c U N Ψ A Y) conservée dans tous les ARN au sein desquels la modification est générée. Par contre, aucune conservation de structure secondaire n'était observée. La seconde partie de mon travail de thèse correspond donc à la recherche des déterminants des requis en séquence et en structure pour qu'un ARN soit substrat de Pus7p. Afin de vérifier si cette séquence est nécéssaire pour la reconnaissance et/ou pour son activité catalytique, nous avons testé, *in*

vitro, l'activité de l'enzyme Pus7p recombinante sur une série de mutants de l'ARNt^{Asp}(GUC), du pré-ARNt^{Tyr}(GΨA) et du snRNA U2. Nous avons démontré pour la première fois la dépendance stricte d'une ARN:Ψ-synthase vis-à-vis de la séquence de son substrat ARN. Il avait été démontré que la spécificité de reconnaissance des enzymes de modification déjà caractérisées dépend soit de la structure 3D globale, soit uniquement de la structure locale de l'ARN modifié. On pouvait donc se demander si en plus de la séquence conservée, Pus7p était dépendante de paramètres de structure secondaire. Pour cela, nous avons réalisé plusieurs mutants structuraux des l'ARNt^{Asp}, du pré-ARNt^{Tyr} et du snRNA U2 afin de mettre en évidence que l'enzyme Pus7p possède à la fois un requis au niveau séquence et au niveau structure secondaire. Connaissant les paramètres nécessaires à la formation des résidus Ψ dans les ARN modifiés par Pus7p, notre objectif a été d'identifier d'autres substrats potentiels de cette enzyme. En effet, le snRNA U2 étant très grand, la recherche des résidus Ψ dans cette ARN n'avait été réalisée que dans la partie 5' très conservées au cours de l'évolution.

La troisième partie de mon travail de thèse correspond à l'étude structurale de Pus7p et de ses variants. Vue la particularité des paramètres requis pour que les ARN soient efficacement modifiés par l'enzyme Pus7p, il était important de comprendre les bases moléculaires de cette reconnaissance et donc d'établir la structure 3D de cette enzyme, libre et complexée avec un substrat. Comme nous l'avons mentionné dans l'introduction, l'enzyme homologue à Pus7p chez *E. coli* (TruD) qui ne semble catalyser que la conversion du résidu U13 de l'ARNt^{Glu}(UUC) en Ψ13 est de taille plus réduite. La comparaison des séquences de Pus7p et TruD nous a permis de mettre en évidence 5 domaines additionnels dans Pus7p. Nous les avons délétés individuellement par mutagenèse dirigée et les propriétés des enzymes mutantes ont été étudiées. Deux des domaines se sont avérés nécessaires à l'activité, ils sont localisés à proximité de la crevasse qui a été définie comme importante dans la fixation du substrat pour l'enzyme TruD. Afin d'essayer de comprendre la fonction de ces domaines, nous avons produit l'enzyme Pus7p et un des mutants actif (Pus7Δ5) en grande quantité. Un travail de cristallogenèse a été entrepris en collaboration avec C. Charron, cristallographe au laboratoire, sur le mutant Pus7Δ5 et des cristaux de cette protéine ont été obtenus.

Puisque les systèmes enzymatiques responsables de la formation des trois résidus Ψ du snRNA U2 (Ψ35, Ψ42 et Ψ44) de *S. cerevisiae* avaient été identifiés, nous avons débuté une étude de leur importance pour la réaction d'épissage. En effet, le résidu Ψ35 est situé dans la région du snRNA U2 s'appariant avec la boîte de branchement de l'ARN pré-messager et les 2 autres résidus Ψ appartiennent au segment d'ARN jouxtant cet appariement. Des études structurales par RMN suggéraient déjà un rôle du résidu Ψ35 consistant à favoriser la conformation extra-hélicale du résidu A impliqué dans l'attaque nucléophile de la première étape de la réaction (Newby & Greenbaum, 2001). L'importance

fonctionnelle de ce résidu Ψ35 à été également montrée *in vivo*, en cumulant la délétion du gène de l'enzyme Pusp7 (responsable de sa formation) et des mutations dans le snRNA U2 à des positions proches de la position 35. Dans ces conditions, un ralentissement de l'efficacité d'épissage de certains ARNm a été observée (Yang et al., 2005). Ainsi, nous avons construit plusieurs souches portant des délétions individuelles ou multiples des gènes *PUS1*, *PUS7* codant les enzymes Pus1p et Pus7p responsables de la formation des résidus Ψ44 et Ψ35 respectivement, mais aussi du gène du snoRNA H/ACA *snR81* guidant la modification à la position 42. De façon surprenante, l'accumulation des délétions n'affecte pas la viabilité de *S. cerevisiae* à différentes températures de croissance et sur des milieux différents. Malgré l'absence de phénotype de croissance, il était possible que des variations du taux d'épissage de certains pré-ARNm aient lieu dans les souches mutées. Pour le tester, nous avons décidé de réaliser une étude globale de l'épissage dans ces souches mutées, ceci en utilisant la technologie d'analyse globale des taux d'épissage des ARNm par l'approche puces à ADN, que l'équipe de J. Beggs (Edinburgh University) avait mise au point pour *S. cerevisiae*.

Une ARN:Ψ-synthase homologue à Pus7p existe aussi chez les archaea, mais l'activité de cette enzyme n'a jamais été étudiée. Par analogie avec les eucaryotes, on pouvait supposer que, chez les archaea, Pus7p catalyse la formation du résidu Ψ13 et, éventuellement, du résidu Ψ35. Pourtant, S. Muller, lors de sa recherche de sRNA H/ACA chez ces organismes, a trouvé chez une des espèces d'archaea, *Sulfolobus solfataricus*, un sRNA H/ACA capable de guider la formation d'un résidu Ψ en position 35 des ARNt. Aucun sRNA H/ACA d'archaea ou d'autres organismes, agissant sur des ARN autres que les ARNr ou UsnRNA n'avait été décrit jusque-là. En examinant la séquence du gène de la protéine Pus7 de cet organisme, S. Muller a observé que la protéine codée par ce gène est mutée au niveau d'acides aminés requis pour l'activité en particulier, un résidu basique du site catalytique impliqué dans la formation d'un paire d'ions avec l'Asp catalytique. En collaboration avec S. Muller, nous avons alors entrepris de produire la protéine Pus7 de cet organisme, ainsi que celle de *P. abyssi* utilisée comme référence. Nous avons pu montrer que l'enzyme Pus7 de *P. abyssi*, contrairement à celle de *S. solfataricus*, possède l'activité pour la position 35 des ARNt. Ayant produit l'enzyme de *P. abyssi*, nous avons réalisé une étude comparative de la spécificité des ARN:Ψ-synthases de la famille Pus7 dans les 3 domaines du vivant.

De plus, il était intéressant de comparer sur le plan structural l'enzyme Pus7 active de *P. abyssi* à celle inactive de *S. solfataricus*. Pour cela, nous avons produit chacune de ces enzymes en grandes quantités et C. Charron a réalisé des tests de cristallisation. Des cristaux de Pus7 de *P. abyssi* ont été obtenus et une première mesure complète de données de diffraction de rayons X a été réalisée au laboratoire à 2,4 Å de résolution. Des cristaux de la protéine sélénométhionylée sont en cours d'obtention afin de résoudre la structure par la méthode MAD.

Matériel & Méthodes

1. Bioinformatique

A. Matériel informatique

Tous les alignements, les recherches de motifs et les prédictions de structure ont été réalisés à partir d'une station PC utilisant le système d'exploitation Windows XP Professionnel.

B. Logiciels utilisés

BLASTp (Basic Local Alignement Search Tool) est un algorithme basé sur l'interprétation de comparaison de deux séquences protéiques (Altschul et al., 1997).

FASTA est un algorithme qui, comme Blast, est basé sur l'interprétation de deux séquences mais celui-ci aboutit à un alignement global de deux séquences (Pearson & Lipman, 1988). La version utilisée est Fasta 2.0.

CLUSTAL-w (version 1.81) est un programme qui permet un alignement global de plusieurs séquences (Thompson et al., 1994).

pHD est un programme permettant la prédiction de structures à partir d'un ensemble de séquences obtenues par alignement multiple (Rost et al., 1994).

MFOLD (version 3.1) est un programme de prédiction de structures secondaires (Zuker & Jacobson, 1998; Mathews et al., 1999).

2. Biologie Moléculaire

A. Matériel

1. Les souches utilisées

1.1. 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, *lacI*^{q}, *lac*ZΔM15, Tn10(Tet^{r})]), **TOP 10 F'** (Invitrogen) (F', *end*A1, *rec*A1, *hsd*R17 (r_k^-, m_k^+), *sup*E44, *thi*-1, *gyr*A96, *rel*A1, Φ80*lac*ZΔM15, Δ(*lac*ZY A- *arg*F), U169 λ-) et **TG1** (*sup*E, *hsd*Δ5, *thi*, Δ(*lac-pro*AB), F' [*tra*D36, *pro*AB+, *lac*I^{q}, *lac*ZΔM15]) 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($r_{B-}m_{B-}$), *dcm*+, Tet^r, *gal* λ(DE3), *end*A, Hte[*arg*U *ile*Y *leu*W Cam^r]) et **B834 (DE3) RIL** d'*E. coli* (F$^-$, *omp*T, *hsd*S($r_{B-}m_{B-}$), *dcm*+, Tet^r, *gal* λ(DE3), *end*A, Hte[*arg*U *ile*Y *leu*W Cam^r] *met* ont été utilisées pour la production de protéines recombinantes sauvage et sélénomethionylées, respectivement. Ces souches sont mutées au niveau de 2 protéases. De plus, elles surexpriment les gènes des

ARNt permettant la transcription des codons rares (Ile, Leu, Arg). Elle possède une copie du gène de l'ARN polymérase du bactériophage T7, placée sous le contrôle du promoteur lac UV5 inductible par l'IPTG, et intégrée dans le chromosome via le prophage λ(DE3). Ces souches sont utilisées pour l'expression d'ORF clonées dans un plasmide sous le contrôle du promoteur du bactériophage T7.

1.2. Souches de levure

Toutes les souches de *S. cerevisiae* utilisée proviennent de la firme Research Genetics.

S. cerevisiae Hansen BY4742 : MATα, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0.

S. cerevisiae Hansen BY4742 ΔYOR243c: MATα, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, YOR243c::kanMX4.

S. cerevisiae Hansen BY4741 ΔYOR243c: MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, YOR243c::kanMX4.

S. cerevisiae Hansen BY4742 ΔYPL242c: MATα, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, YOR243c::kanMX4.

1.3. Souches d'archaea

Les souches *Pyrococcus abyssi GE* et *Pyrococcus furiosus DSM 3638* correspondent aux souches dont le génome est disponible. Des culots cellulaires des souches de *P. abyssi* et *S. solfataricus* cultivés en milieu VSM nous ont été aimablement transmis par F. Matsunaga et A. Hecker appartenant au laboratoire LBMGE de P. Forterre à Orsay.

2. Les plasmides utilisés

Plasmide pCR2.1 (Invitrogen) **:** Le clonage de fragments d'ADN obtenus par amplification par la technique PCR a été réalisé dans ce qui 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 sous cloner certaines protéines utilisée lors de notre étude.

Plasmide pUC18 : ce plasmide a été utilisé pour cloner les fragments ADN amplifiés par PCR et phosphorylés de l'ensemble des ARN qui ont servi à notre travail. Ce plasmide a égalementété utilisé pour sous cloner certaines protéines utilisée lors de notre étude.

Plasmide p413GalS : ce plasmide a été utilisé pour produire la protéine Pus7p chez la levure sous contrôle d'un promoteur inductible au galactose.

Plasmide pET28b : ce plasmide a été utilisé pour cloner les fragments correspondant aux protéines Pus7 sauvage et mutantes de *S. cerevisiae* et de tout ses homologues en vue d'une production *in vitro* avec une étiquette His et d'une purification sur résine de Nickel.

Plasmide pMAL-c2X (Biolabs) : ce plasmide a été utilisé pour la production de protéines recombinantes fusionnées à l'étiquette MBP en vue d'une purification sur billes de maltose.

3. ADN génomique de *P. abyssi* et *S. solfataricus*

Des lots d'ADN génomique de ces deux espèces nous ont été transmis par F. Matsunaga du laboratoire de P. Forterre à Orsay (LBMGE).

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

Les solutions et tampons d'usage courant sont répertoriés dans le tableau 9.

Conditions d'utilisation	Dénomination	Composition
Préparation de cellules compétentes	Tampon TfbI	AcK 30mM, MnCl2 50mM ; KCl 100mM ; CaCl2 10mM ; Glycérol 15%
	Tampon TfbII	NaMOPS 10mM pH7 ; CaCl2 75mM ; KCl 10mM ; Glycérol 15%
Régénération des cellules	SOC	2% Bactotryptone ; 0,5% Extrait de levure ; NaCl 10mM ; KCl 2,5mM ; MgCl2 10mM ; MgSO4 10mM, Glucose 10mM
Electrophorèse en gel	TBE	Acide borique 89mM ; EDTA 2mM ; Tris Borate 89mM pH 8,3
	Tampon de dépôt sur gel Agarose	bleu de Bromophénol 0,09% ; bleu de xylène cyanol 0,09% ; glycérol 60% ; EDTA 60mM
	Tampon de dépôt sur gel de protéines dénaturant	EDTA 20mM ; bleu de Bromophénol 0,05% ; bleu de xylène cyanol 0,05% dans de la formamide désionisée
Electrophorèse selon Laemmli	Tampon d'électrode	25 mM Tris (Base), 192 mM Glycine, 0,1 % SDS
	Tampon pour le gel de séparation	1,5 M Tris-HCl pH 8,8
	Tampon pour le gel de concentration	0,5 M Tris-HCl pH 6,8
	Solution d'Acrylamide/bisacrylamide 40 % (29:1)	Acrylamide 38,7% ; Bisacrylamide 1,33%
	Tampon d'échantillon (SBL)	62,5 mM Tris-HCl pH 6,8 ; 2 % SDS ; 10 % Glycérol ; 100 mM 2-mercaptoéthanol ; 0,005-0,01 % Bleu de bromophénol
Coloration et Décoloration des gels de protéines	Solution de coloration	Ethanol 20% ; Acide Acétique 10% ; 2g/l de Bleu de Coomassie G 250
	Solution de décoloration	Ethanol 20% ; Acide Acétique 10%
Séquençage	TE	Tris-HCl 10mM pH 8 : EDTA 1mM
	Tampon d'hybridation	Tris HCl 200mM pH 7,5 ; MgCl2 50mM ; NaCl 250mM
Purification des protéines sur IMAC	Tampon A start	Tris HCl 50mM pH7 ; 300mM NaCl
	Tampon d'élution	Tris HCl 50mM pH7 ; 300mM NaCl ; Imidazole 1M
	Tampon de régénération de la colonne	Tris HCl 50mM pH7 ; 300mM NaCl ; EDTA 50mM
Sonication	Tampon de lyse des cellules	50mM Tris-HCl pH 7,5 ; 300 mM NaCl; 10 mM Imidazole ; 10 mM 2-Mercaptoéthanol
Purification des protéines recombinantes sur colonne Ni-NTA	Tampon de lavage	NaCl 300mM ; Imidazole 20mM ; Tris-HCl 50mM pH 7,5
	Tampon d'élution	NaCl 300mM ; Imidazole 150mM ; Tris-HCl 50mM pH 7,5
Tampons divers	Tampon d'élution	Tris-HCl 10 mM, pH 7,5 ; NaCl 300 mM ; EDTA 1 mM ; SDS 1%
	Tampon D 10x	HEPES-KOH 20 mM, pH 7,9 ; KCl 150 mM ; MgCl2 1,5 mM ; EDTA 0,2 mM
	Tampon Psi 1x	Tris-HCl 100 mM, pH 8 ; AcNH4 100 mM ; MgCl2 5 mM ; DTT 2 mM ; EDTA 0,1 mM
	Tampon de migration de gel retard	Glycérol 87% 115 ml ; TBE 20x 50 ml ; H2O qsp 2 l
	Tampon T2 2x	Acétate d'ammonium 100 mM, pH 4,6
	Tampon P1 2x	Acétate d'ammonium 100 mM, pH 5,2
	Tampon de migration N1 pour TLC	Acide Isobutyrique 25% 330 ml ; Ammoniac 25% 5 ml ; EDTA 0,5 M pH 8 ; H2O qsp 500 ml
	Tampon de migration N2 pour TLC	Isopropanol 70 ml ; HCl 37% 15 ml ; H2O qsp 100 ml
	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

Tableau 9 : Les solutions et tampons d'usage courant.

Les milieux que nous avons utilisé lors de cette étude sont :

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. Pour le milieu solide l'agar est ajouté à raison de 17 g par litre de milieu de culture 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 YPD : Bactopeptone10 g/l, Extrait de levure 10 g/l, Glucose 20 g/l. Le milieu de culture est ensuite autoclavé 20 min à 121°C sous une pression de 1 bar. Pour le milieu solide, l'agar est ajouté à raison de 20 g par litre de milieu de culture.

Milieu YPD : Yeast Nitrogen Base (Difco) 6,7 g/l, Glucose 20 g/l. Le milieu de culture est ensuite autoclavé 20 min à 121°C sous une pression de 1 bar. Pour le milieu solide, l'agar est ajouté à raison de 20 g par litre de milieu de culture.

B. Méthodes

1. Méthodes concernant l'ADN

1.1. Extraction d'ADN génomique de la levure *S. cerevisiae*

Une colonie de la souche de levure est incubée dans 2-3 ml de milieu riche YPD toute la nuit à 30°C. Les cellules sont centrifugées 1 min à 12000 rpm. Le culot est repris dans 200 µl de tampon d'extraction (Tris-HCl 10 mM pH 8,0 ; NaCl 100 mM ; EDTA 1 mM ; SDS 1% ; Triton X-100 2%). On ajoute 200 µl de billes de verre siliconnées et 200 µl d'un mélange phénol-chloroforme (vol/vol). Le tout est agité 5 min puis centrifugé 5 min à 12000 rpm. L'ADN contenu dans le surnageant est précipité à l'éthanol à 96 %. Le culot obtenu est séché, repris dans 30 µl de tampon TE contenant 0,33 mg/ml de RNase A. Après 5 min d'incubation à 37°C, une précipitation à l'$AcNH_4$ 2 M est réalisée. L'ADN génomique obtenu est repris dans 50 ou 100 µl d'H_2O bidistillée selon la quantité d'ADN présente.

1.2. Réaction de polymérisation en chaîne (PCR)

Nous avons utilisé cette technique pour amplifier l'ORF YOR243c à partir de l'ADN génomique de *S. cerevisiae*. Pour cela, 1 µl d'ADN génomique est mis en présence du mélange réactionnel décrit ci-dessous. Les 2 oligodéoxynucléotides utilisés lors de la réaction ont la particularité d'être complémentaires de chaque brin de l'ADN matrice tout en permettant l'insertion des sites de clivage *Nhe* I et *BamH* I, respectivement en amont et aval du gène.

- Amplification d'ADN par la méthode PCR

Matrice (ADN génomique)	5 ng
Oligodéoxynucléotide 5'(100 ng/µl)	1 µl
Oligodéoxynucléotide 3'(100 ng/µl)	1 µl
dNTP (1,25 mM de chaque nucléotide)	8 µl
Tampon Pfu 10x (Tris-HCl 200 mM; KCl 100 mM ; $(NH_4)_2SO_4$ 100 mM ; $MgSO_4$ 20 mM ; Triton X-100 1 % ; BSA1 mg/m pH 8,8l)	5 µl
ADN polymérase Pfu (Stratagene 2,5 U/µl)	1 µl
H_2O	qsp 50 µl

L'amplification est réalisée dans un appareil Perkin-Elmer ou Hybaid.
Les cycles suivants sont réalisés :

cycle	fonction	température	temps	Nombre de cycles
1	dénaturation	94°C	5 min	1
2-31	dénaturation	94°C	1 min	30
	hybridation	55°C	1 min	
	élongation	74°C	2 min	
32	élongation	74°C	10 min	1
33	stockage	20°C		

- Amplification d'ADN à partir d'une colonie bactérienne par la méthode PCR

Oligodéoxynucléotide 5'(100 ng/µl)	1 µl
Oligodéoxynucléotide 3'(100 ng/µl)	1 µl
Tampon 10x (Tris-HCl 100 mM pH 8,8 ; $MgCl_2$ 15 mM ; KCl 150 mM ; Triton X-100 1 %)	2 µl
dNTP (1,25 mM de chaque nucléotide)	3,2 µl
DyNAzyme™ II DNA polymérase (FINNZYMES, 2 U/µl)	0,3 µl
H_2O	qsp 20 µl

La colonie est ensuite ajoutée dans ce mélange réactionnel. Les cycles suivants sont ensuite réalisés :

Cycle	fonction	température	temps	Nombre de cycles
1	dénaturation	94°C	5 min	1
2-31	dénaturation	94°C	30 s	30
	hybridation	42°C	30 s	
	élongation	74°C	30 s	
32	élongation	74°C	10 min	1
33	stockage	20°C		

1.3. Clonage de gènes

1.3.1. Conditions d'hydrolyse de l'ADN

La quantité d'enzyme utilisée est de 5 U/µg d'ADN. Les enzymes sont fournies par MBI Fermentas et Biolabs et sont utilisées dans les conditions et avec les tampons préconisés par le fournisseur. Dans tous les cas, la concentration en ADN est de 1 µg pour 10 µl de mélange d'hydrolyse.

1.3.2. Purification des fragments ADN

Les fragments d'ADN hydrolysés par les différentes enzymes de restriction sont séparés par électrophorèse sur gel d'agarose 0,8 % et purifiés en utilisant le kit " Nucleospin " (Macherey-Nagel).

1.3.3. Conditions de phosphorylation/déphosphorylation

Après purification, l'insert est phosphorylé par 1 unité de T4 polynucléotide kinase (Boehringer) dans son tampon 1x ($MgCl_2$ 10 mM ; DTT 5 mM ; Tris-HCl 70 mM pH 7,5) pendant 45 mn à 37°C en présence de 1 mM d'ATP. Finalement, l'insert est incubé 10 mn à 65°C pour inactiver la kinase puis extrait avec un mélange phénol-chloroforme et précipité à l'éthanol 96% en présence d'AcNa 0,3M.
Le plasmide hydrolysé est traité par la phosphatase alcaline bovine (Roche Molecular Biochemicals) directement dans le mélange de restriction à raison d'1 U/µg de plasmide. Les enzymes sont inactivées par incubation à 70°C pendant 10 min. Le vecteur hydrolysé subit une extraction au phénol-chloroforme (vol/vol) et est ensuite précipité avec de l'éthanol à 96 % en présence d'acétate de sodium 0,3 M.

1.3.4. Conditions de ligation

Le rapport des concentrations fragment/vecteur utilisé est de 1 à 5. En général 100 ng de vecteur sont utilisés. L'ADN ligase du bactériophage T4 (MBI Fermentas) est utilisée à 2 concentrations différentes en fonction de la nature des extrémités ADN à liguer. Pour la ligation bouts francs, 2 µl de PEG 4000 sont ajoutés et la concentration de la ligase est de 0,2 U/µl de mélange réactionnel, alors que dans le cas d'extrémités cohésives la concentration en ligase est 2 fois inférieure et le milieu ne contient pas de PEG 4000. La ligation se déroule dans un volume total de 20 µl pendant une nuit à 22°C dans du tampon de ligation 1x (Tris-HCl 40 mM pH 7,8 ; $MgCl_2$ 10 mM ; DTT 10 mM ; ATP 0,5 mM).

1.4. Transformation d'*E. coli.*

Les bactéries sont rendues compétentes par un traitement au chlorure de manganèse (Cohen *et al.*, 1972) avant d'être incubées 30 min à 4°C avec la totalité du produit de ligation décrit ci dessus. Après un choc thermique de 30 s à 42°C, les cellules renfermant un plasmide

sont sélectionnées sur un milieu LB + ampicilline/kanamycine, selon la résistance portée par le plasmide avec éventuellement en plus, un substrat produisant une coloration lorsqu'il est métabolisé.

Ainsi les cellules d'*E. coli* transformées par un plasmide pUC19 contenant une insertion de l'ORF YOR243c dans son site de clonage ne peuvent avoir l'activité β-galactosidase, car la séquence insérée interrompt la séquence codant la protéine du gène *lac*Z. L'utilisation d'un analogue incolore, le 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal), entraîne après hydrolyse par la β-galactosidase, la production d'un catabolite de couleur bleue. On reconnaît donc les colonies bactériennes possédant une insertion dans le site de clonage à leur couleur blanche, tandis que les bactéries contenant seulement le plasmide, sans insertion dans ce site, forment des colonies bleues.

1.5. Analyse des clones recombinants

1.5.1. PCR sur colonie

Pour vérifier l'insertion d'un fragment dans un vecteur, une amplification de l'ADN par PCR est réalisée directement sur les colonies obtenues après transformation des bactéries par le mélange de ligation et mises en suspension dans le mélange réactionnel dont la composition est donnée dans la Figure X. Différents couples d'oligodéoxynucléotides s'hybridant sur les différents plasmides sont utilisés. Les produits d'amplification sont fractionnés par électrophorèse en gel d'agarose à 0,8%, puis visualisés en présence de bromure d'éthidium (0,4 µg/ml) sous UV (254 nm).

1.5.2. Analyse rapide du contenu plasmidique

Pour vérifier l'insertion d'un fragment dans un vecteur, nous avons également réalisé une comparaison entre la taille du vecteur vide et celle du vecteur censé contenir l'insert. Une fraction de la colonie à tester est cultivée dans 200 µl de milieu LB contenant l'antibiotique approprié, pendant 4 à 5h à 37°C. Au bout de cette période, 100 µl de culture sont centrifugés pendant 1 min à 13000 rpm. Le culot est ensuite repris dans 50 µl de tampon de lyse (5 mM EDTA, 10 % sucrose, 0,25 % SDS, 10 mM NaOH, 60 mM KCl, 0,05 % bleu de bromophénol), après 5 min d'incubation à 37°C suivies de 5 min d'incubation à 4°C, les cellules sont de nouveau centrifugées à 13000 rpm pendant 5 min et le matériel contenu dans 10 µl de surnageant est fractionné par électrophorèse en gel d'agarose à 0,7%, puis visualisé en présence de bromure d'éthidium (0,4 µg/ml) sous UV (254 nm).

1.6. Préparation d'ADN plasmidique

L'ADN plasmidique est extrait par la méthode de lyse alcaline (Birnboïm et Doly, 1979), modifiée par Ish-Horowicz (1981). Les préparations d'ADN plasmidique à plus grande échelle

sont réalisées d'après le même principe, avec le kit "Quick Flow Midi and Maxi Kit Protocol" (QIAGEN).

1.7. Séquençage d'ADN

L'ensemble des constructions produites a été vérifié par séquençage du segment inséré. Pour cela, l'ADN plasmidique double brin est dissocié en simple brin par dénaturation alcaline et le séquençage est réalisé selon la méthode de Tabor et Richardson (1987). Un oligodéoxynucléotide de synthèse, hybridé en 3' du fragment à séquencer, sert d'amorce à la synthèse d'un ADN complémentaire par l'ADN polymérase du phage T7. Cette synthèse est effectuée en présence de déoxynucléotides (dNTP), de dATP [α-^{33}P] et de didéoxynucléotides (ddNTP). L'incorporation de ddNTP arrête la synthèse statistiquement du fait de la concentration relative des dNTP/ddNTP.

1.8. Mutagenèse par PCR

La mutagenèse dirigée est réalisée directement sur un plasmide double brin superenroulé et nécessite l'utilisation de 2 oligodésoxyribonucléotides de mutagenèse portant chacun la mutation à introduire. La méthode utilise l'ADN polymérase thermostable *Pfu* (6 fois plus fidèle dans la réplication de l'ADN que la Taq polymérase) pour répliquer les 2 brins du plasmide avec une grande fidélité et sans déplacer les oligodésoxyribonucléotides de mutagenèse. Les oligodésoxyribonucléotides, chacun complémentaire d'un brin du plasmide, vont servir d'amorce pour la synthèse de l'ADN lors des cycles d'élongation de la PCR, ce qui conduit à un plasmide muté avec des extrémités non religuées. Le mélange réactionnel est ensuite traité par l'endonucléase *Dpn I* (séquence cible 5'-G^{m6}ATC-3') pour digérer l'ADN parental dam méthylé issu de la souche d'*E. coli* XL1 Blue et ainsi sélectionner l'ADN néosynthétisé contenant la mutation. Finalement, une fraction du mélange de digestion contenant l'ADN db muté est utilisée pour transformer des cellules compétentes, au sein desquelles les extrémités du plasmide muté sont religuées. Le principe de la méthode est illustré dans la figure 59, le protocole est décrit ci-dessous :

1.Amplification d'ADN par la méthode PCR

Matrice (plasmide)	5 ng
Oligodéoxynucléotide 5'(100 ng/µl)	1 µl
Oligodéoxynucléotide 3'(100 ng/µl)	1 µl
dNTP (1,25 mM de chaque)	8 µl
Tampon *P*fu10x (Tris-HCl 200 mM pH 8,8 ; KCl 100 mM $(NH_4)_2SO_4$100 mM ; $MgSO_4$ 20 mM ; Triton X-100 ; BSA 1 mg/ml)	5 µl
ADN polymérase Pfu (Stratagene 2,5 U/µl)	1 µl
H_2O	qsp 50 µl

L'amplification est réalisée dans un appareil Perkin-Elmer ou Hybaid.
Les cycles suivants sont réalisés :

cycle	fonction	température	temps	Nombre de cycles
1	dénaturation	94°C	5 min	1
2-21	dénaturation	94°C	3 min	20
	hybridation	50°C	1 min	
	élongation	74°C	12 min	
22	élongation	74°C	10 min	1
23	stockage	20°C		

2. Digestion du produit de la réaction PCR

Produit de la réaction	50 µl
Dpn I	10 U

Le mélange de digestion est ensuite incubé 1 heure à 37°C

3. Transformation des bactéries compétentes

La transformation est réalisée avec 1 et 10 µl du mélange de mutagenèse et 100 µl de bactéries compétentes XL1 Blue.
Les cellules sont étalées sur boîte de LB avec l'antibiotique adéquat.
Les clones sont séquencés pour vérifier la présence de la mutation introduite et l'absence de toute autre mutation dans la région codante.

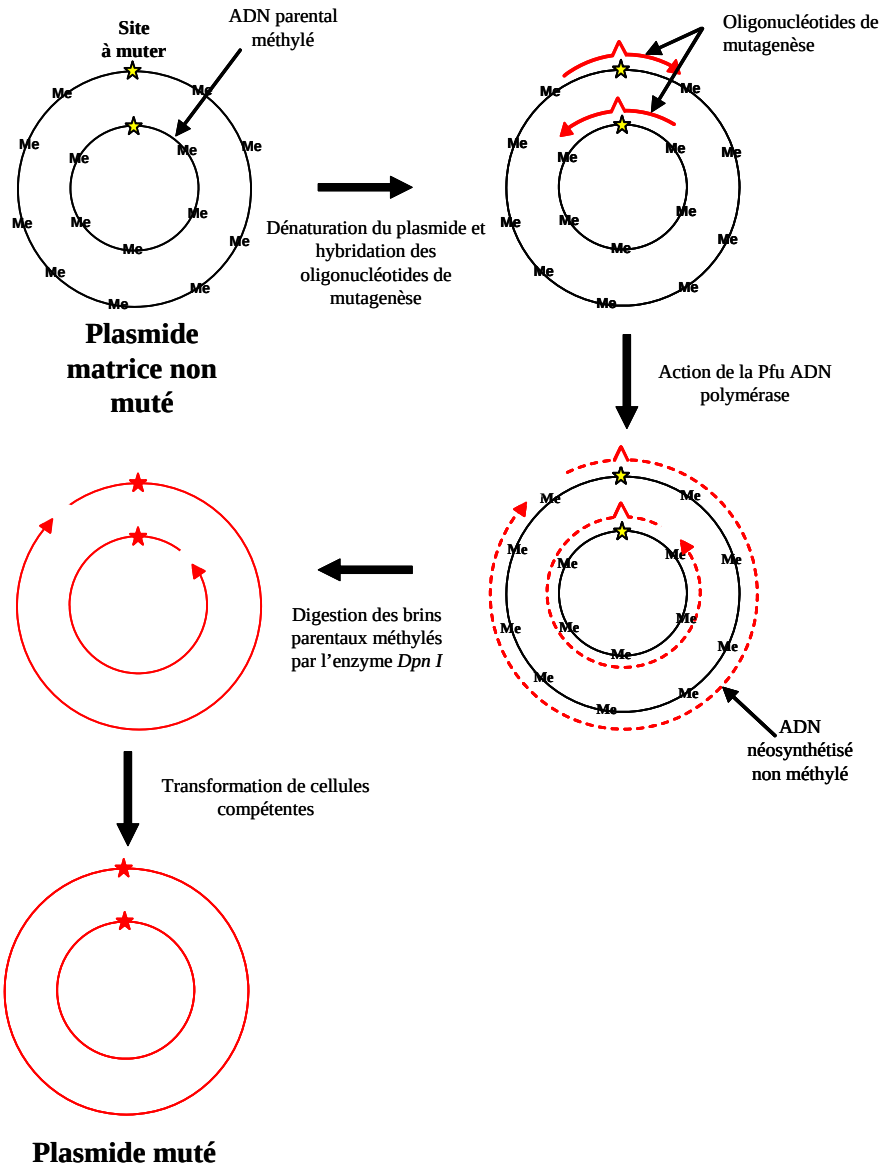


Figure 59 : Principe de la mutagenèse par PCR

1.9. Transformation de *S. cerevisiae.*

Une culture de S. cerevisiae est ensemencée à une DO_{600}nm de 0,1 (50 ml en milieu YPD) et arrêtée lorsque la DO_{600}nm est de 0,6. Les cellules sont ensuite centrifugées 10 min à 2500 rpm. Le culot est repris dans 10 ml d'eau stérile et les cellules sont à nouveau centrifugées. Le culot cellulaire est repris dans 500 µl de tampon LiTE 1x (LiAc 100 mM, EDTA 1 mM, Tris-HCl 10 mM pH7,5). Pour la transformation, 50 ng de plamide, 4,5 µl d'ADN de poisson (10 µg/µl), 50 µl de levures compétentes et 300 µl de PEG-LiTE sont incubés 30 min à 30°C, puis soumis à un choc thermique de 15 min à 42°C après ajout de 40 µl de DMSO (10 % final). Les cellules sont ensuite étalées directement sur milieu sélectif afin de sélectionner les souches devenues capables de se développer en absence d'histidine et les boîtes sont placées à 30°C durant 48h.

1.10. Sporulation de levure, dissection des spores et vérification par répliques au velours

Les levures diploïdes sporulent spontanément dans des milieux pauvres. Après sporulation on obtient des asques (la levure est un Ascomycète) contenant 4 spores haploïdes qui sont les produits directs de la méiose (tétrades). Pour disséquer des tétrades, il est nécessaire de les libérer par digestion partielle de la paroi des asques à l'aide de la zymolyase. La zymolyase digère les liaisons β 1-3 glucane des protéines glycosylées de la paroi. On effectue une digestion partielle de façon à éviter la lyse des spores et pour que les 4 spores d'une tétrade restent ensemble et ne se dispersent pas sur le milieu.
Afin de vérifier les spores, on effectue des répliques au velours. Ces répliques permettent de transférer des cellules à partir des colonies d'une boîte sur d'autres boîtes vierges. Ceci s'effectue dans différents buts : 1) étudier la croissance de ces colonies; 2) obtenir des colonies fraîches; 3) vérifier la ségrégation des marqueurs sur différents milieux sélectifs afin de sélectionner les spores possédant les marqueurs d'intérêt; 4) vérifier le signe sexuel des spores.

Nous avons utilisé le protocole de sporulation suivant :
1. Répartir à la pipette 2 ml de milieu YPD dans des tubes de culture.
2. Inoculer un clone à partir d'une boîte YPD fraîche (ou ayant séjourné moins d'une semaine à 4°C) dans les tubes de culture.
3. Incuber une nuit à 30°C sous agitation (sur une roue).
4. Transférer 1 ml de la culture dans un tube Eppendorf stérile et centrifuger rapidement.
5. Eliminer le surnageant.
6. Resuspendre le culot de cellules dans 1 ml de milieu YPA (Bacto peptone Difco 10 g/l, Bacto yeast extract Difco 10 g/l, Acétate de Potassium 20 g/l) et transférer dans un tube de culture.
7. Incuber au moins 6 h à 30°C avec agitation (sur une roue).
8. Transférer les cellules dans un tube Eppendorf stérile et centrifuger rapidement.
9. Éliminer le surnageant et resuspendre le culot de cellules dans 0,5 ml de milieu A (Acétate de Potassium 20 g/l) puis transférer dans un tube de culture.
10. Incuber à 30°C pendant au moins 4 jours sur roue motorisée.

La dissection des tétrades a été réalisée selon le protocole suivant :
1. Observer au microscope entre lame et lamelle la formation des tétrades. Si le pourcentage de tétrades est correcte, prélever une anse de cellules et les resuspendre dans 50 µl de la solution de dissection (Zymolyase à 1 mg/ml dans du sorbitol 1 M).
2. Incuber 10 min à 37°C puis mettre dans la glace.
3. Rajouter de l'eau distillée stérile et faire une strie sur une boîte de milieu YPD.
4. Utiliser le micro manipulateur pour disséquer les tétrades.
5. Les ascospores provenant d'une même tétrade sont placées à intervalles régulières par groupe de 4.

6. Les boîtes sont placées à 30°C jusqu'à l'obtention de colonies bien visibles qui seront restriées sur boîtes de milieu riche.

2. Méthodes concernant les ARN

2.1. Extraction d'ARN de *S. cerevisiae*

L'extraction a été réalisée à partir de cultures liquides des différentes souches de *S. cerevisiae*.

25 ml de milieu YPD sont ensemencés à une DO_{600}nm de 0,3 à partir d'une préculture de la souche de *S. cerevisiae* étudiée. La culture est arrêtée à une DO_{600nm} de 0,8.

Les cellules sont centrifugées 10 min à 2500 rpm et le culot est repris dans 200 µl de tampon d'extraction (NaCl 100 mM, EDTA 10 mM, Tris-HCl 50 mM pH 7,5). On ajoute 300 µl de billes de verre siliconées (0,45 à 0,5 mm de diamètre). Les tubes sont ensuite agités à 2 reprises pendant 1 min, en les laissant 2 min dans la glace après chaque agitation. On ajoute ensuite 200 µl de tampon d'extraction, 50 µl de SDS 10% et 400 µl d'un mélange phénol-chloroforme (vol/vol). Le tout est agité 1 min et incubé 10 min à 65 °C. Une centrifugation de 5 min à 13000rpm ensuite réalisée et on procède à une extraction phénol-chloroforme classique (vol/vol). Les ARN, contenus dans le surnageant, sont précipités à l'éthanol, en présence d'AcNa 0,1M. Le culot est lavé, séché et repris dans 50 µl d'eau stérile. La quantité d'ARN est estimée par spectrophotométrie à 260 nm.

2.2. Extraction d'ARN de bactéries et d'archaea

L'extraction des ARN totaux de bactérie et d'archaea a été réalisée selon la méthode décrite par Motorin et al. (Motorin et al., 2007).

2.3. Analyse des résidus Ψ dans les ARN par la méthode au CMCT

2.3.1. Modification des ARN par le CMCT

Cette technique repose sur la différence de stabilité en milieu alcalin (pH 10,4) du produit d'addition du CMCT en position N-3 sur les résidus Ψ et les résidus U (Bakin & Ofengand, 1993). Le CMCT s'additionne également, mais plus faiblement, en position N-1 des résidus G. La présence d'une modification en position N-3 des résidus U et Ψ conduit à un arrêt de la transcriptase inverse, lors d'une réaction d'élongation d'amorce. En effet, la position N-3 est impliquée dans la formation de l'appariement Watson et Crick entre les nucléotides U et A. Un traitement alcalin des ARN modifiés par le CMCT, conduit à l'élimination du groupement CMCT porté par les résidus U, mais pas de celui porté par les résidus Ψ. Les résidus Ψ seront donc mis en évidence par ce traitement du fait de la persistance de la modification après hydrolyse alcaline.

Chaque réaction se fait dans des conditions statistiques de manière à avoir, en moyenne, une modification par molécule d'ARN. Le protocole est le suivant :

Environ 10 µg d'ARN totaux sont ajoutés dans 4 tubes : T, C_1, C_{10} et C_{20}.
30µl de CMCT 0,17 M dans un tampon EDTA 4 mM, Urée 7 M, Bicine 50 mM, pH 8,3 sont ajoutés dans les tubes C_1, C_{10} et C_{20}.
30 µl de ce même tampon, en absence de CMCT, sont ajoutés dans le tube T.

Les tubes sont ensuite placés à 37°C pendant différents temps : 1 min (tube C_1), 10 min (tube C_{10}) ou 20 min (tubes C_{20} et T). L'ARN est précipité 2 fois à l'éthanol en présence d'AcNa 0,3 M. Le culot d'ARN est alors repris dans 40 µl de Na_2CO_3 50 mM, pH 10,4 et incubé 3 h à 37° C (pour les tubes T, C_{10} et C_{20}, le tube C_1 est stocké à -80°C).
Les ARN sont ensuite précipités à l'éthanol en présence d'AcNa 0,3 M, séchés sous-vide et repris dans 1 µl d'H_2O stérile afin d'être analysés par extension d'amorce par la transcriptase inverse.

2.3.2. Analyse par extension d'amorce avec la transcriptase inverse

2.3.2.1. Marquage des oligodéoxynucléotides

La réaction est catalysée par la T4 polynucléotide kinase en présence d'ATP [γ-^{32}P]. Cette enzyme catalyse le transfert du phosphate en position γ de l'ATP sur l'extrémité 5' OH d'une molécule d'ADN ou d'ARN. Le mélange réactionnel utilisé est le suivant:

ATP [γ^{32}P] (10 mCi/ml, 3000 Ci/mmol, Amersham)	1µl(10µCi)
Oligodéoxynucléotide	100 ng
Tampon 10X	1 µl
(Tris-HCl 700 mM; 200 mM KCl pH 7,6)	
T4 polynucléotide kinase (10 U/µl) (Biolabs)	1 µl
H_2O stérile	qsp 10 µl

Incubation 45 min à 37°C.
Les oligonucléotides marqués sont purifiés par électrophorèse en gel de polyacrylamide dénaturant à 10 %. Après migration à 17 V/cm, durant 1 heure, les bandes de gel contenant l'oligodéoxynucléotide marqué sont révélées par autoradiographie. Ces bandes sont ensuite découpées et placées dans 150 µl de tampon d'élution. Les oligodéoxynucléotides sont élués à 37°C pendant une nuit puis incubés 30 min à 65°C. Les oligodéoxynucléotides sont précipités à l'éthanol en présence d'acétate de sodium 0,3 M, séchés et repris dans un volume d'eau bidistillée tel que 1 µl corresponde à 100 cps.

2.3.2.2. Elongation par la transcriptase inverse

Un oligodéoxynucléotide marqué en 5' au ^{32}P est hybridé à une région donnée de l'ARN à analyser. L'extrémité 3' de cet oligodéoxynucléotide sert d'amorce à la transcriptase inverse (AMV reverse transcriptase, Life Sciences), qui synthétise en présence des 4 dNTP, un brin d'ADN complémentaire à l'ARN. Au niveau des pseudouridines, la transcriptase inverse est stoppée, ce qui se traduit par des ADNc raccourcis. Le mélange réactionnel pour l'hybridation est :

ARN modifié	1 µl
Tampon d'hybridation 5x	0,5 µl
(250 mM Tris HCl; 60mM $MgCl_2$; 400 mM KCl pH 8,3)	
Oligodéoxynucléotide marqué au ^{32}P	5 ng (1 µl)

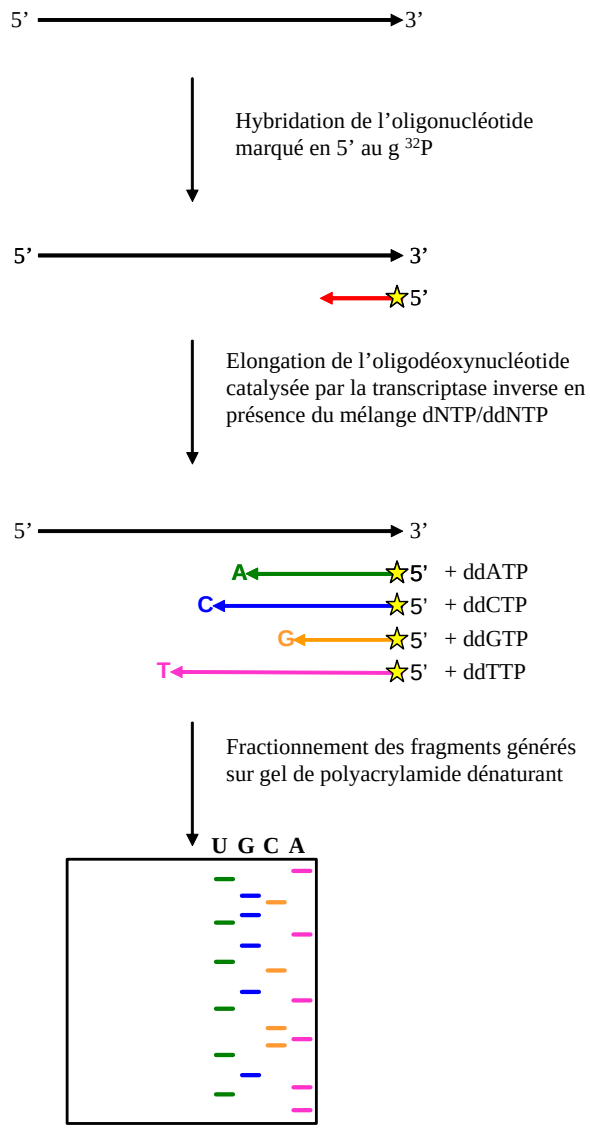
Incubation 10 min à 65°C, puis 10 min dans la glace.

Ajout de 2,5 µl de mélange de synthèse dont la composition est indiquée ci-dessous.

Mélange de dNTP (5 mM de chaque nucléotide)	0,1 µl
Tampon RT 10x	0,25 µl
(500 mM Tris HCl; 60mM $MgCl_2$; 400 mM KCl pH 8,3)	
Transcriptase inverse AMV (20 U/µl) (Life Sciences)	0,025 µl

Incubation 45 min à 42°C

Afin d'identifier les positions converties en pseudouridines, un séquençage sur l'ARN non modifié est réalisé en parallèle. Le principe est le même mis à part que des ddNTP sont ajoutés au milieu réactionnel en concentration telle que la synthèse du brin d'ADN soit arrêtée statistiquement. Le principe de cette analyse est présenté dans la Figure 60.



<u>Figure 60</u> : Principe du séquençage de l’ARN par élongation d’une amorce par la transcriptase inverse (Qu et al., 1983).

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

1. Hybridation de l’amorce

Oligonucléotide marqué au ^{32}P	1 µl (10 µCi)
ARN total	10 µg
Tampon d’hybridation 10X	1 µl
(600 mM NaCl, 100 mM DTT, 500 mM Tris HCl pH 8,3)	
H_2O stérile	qsp 10 µl

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

2. Synthèse du brin d’ADN complémentaire

+2,5 µl de mélange de synthèse

+0,5 µl du ddNTP voulu (A, C, G ou T)
Incubation 45 min à 42°C.

2.4. Production de transcrits *in vitro*

La matrice ADN comporte le promoteur de l'ARN polymérase T7 (5'-TAATACGACTCACTATAGGG-3') et est obtenue par trois méthodes différentes. Une première méthode, utilisée pour la production de matrices utilisées pour la transcription de petits ARN substrats, consiste à apparier deux oligonucléotides de séquences complémentaires en quatre étapes d'incubation : 2 min à 96°C , 15 min à 65°C, 15 min à 37°C et 15 min à 20°C. L'ADN double brin formé est purifié par électrophorèse en gel de polyacrylamide à 10% non dénaturant puis élué. Une deuxième méthode également utilisée pour la production de matrices utilisées pour la transcription d'ARN substrats consiste à associer deux oligonucléotides partiellement complémentaires et à réaliser une élongation de ceux-ci par PCR. Une troisième méthode, utilisée pour la production de matrices utilisées pour la transcription des sRNA à boîtes H/ACA et des tRNA, consiste à amplifier par PCR une partie de l'ADN génomique grâce à un couple d'oligonucléotides qui permet également l'incorporation de la séquence du promoteur T7. Les fragments amplifiés par PCR sont ensuite clonés dans un plasmide pUC18 ou pCR2.1. La nature de ces fragments est vérifiée par un séquençage réalisé au laboratoire. Ces plasmides sont utilisés comme matrice après linéarisation. Les protocoles de transcription utilisés sont les suivants :

- Transcription non radioactive

Matrice ADN (produits PCR ou plasmide linéarisé 1 µg/µl)	1 µl
Tampon de transcription 10X (100 m*M* $MgCl_2$, 100 m*M* Tris-HCl pH 7.5)	2 µl
rNTP (10 mM de chaque rATP, rCTP, rGTP, UTP)	1 µl
RNasin (RNA guard Pharmacia Biotech, 35,7 U/ml)	1 µl
T7 RNA polymérase (Ambion) (20 U/µl)	1,5 µl
H_2O	qsp 20 µl

Le mélange est incubé 3 heure à 37°C, puis 1 µl de DNase I (Amersham Pharmacia, 7,5 U/µl) est ajouté et le mélange est à nouveau incubé 30 min à 37°C. Les transcrits sont purifiés sur gel de polyacrylamide dénaturant 8%. Les ARN sont visualisés par ombrage aux UV (254 nm), découpés puis élués une nuit à 4°C dans du tampon d'élution à forte concentration saline puis précipités avec de l'éthanol 96% en présence d'acétate d'ammonium. Les ARN sont ensuite repris dans de l'eau stérile.

- Transcription radioactive uniforme au [α^{32}P]-ATP

Matrice ADN (produit PCR ou plasmide linéarisé 1 µg/µl)	1 µl
Tampon de transcription 10X (100 m*M* $MgCl_2$, 100 m*M* Tris-HCl pH 7.5)	2 µl
rNTP (10 mM de chaque rCTP, rGTP, UTP)	1 µl
[α^{32}P]-ATP (Amersham, 3000 Ci/mmol ; 10 µCi/µl)	2 µl
ATP 50 µM	1 µl
RNasin (RNA guard Pharmacia Biotech, 35,7 U/ml)	1 µl
T7 RNA polymérase (Ambion) (20 U/µl)	1,5 µl
H_2O	qsp 20 µl

Le mélange est incubé 3 heures à 37°C, puis 1 µl de DNase I (Amersham Pharmacia, 7,5 U/µl) est ajouté et le mélange est à nouveau incubé 30 min à 37°C. Les ARN sont purifiés sur gel de polyacrylamide dénaturant 8% puis visualisés par autoradiographie. Les étapes qui suivent sont identiques à celles décrites pour la transcription non radioactive. Les transcrits sont repris dans de l'eau stérile à 100 000 cpm/ml.

2.5. Evaluation de l'activité de pseudouridylation

Ce test d'activité repose sur la méthode du plus proche voisin (Fig. M.2). L'ARN substrat uniformément marqué en présence de l'un des 4 radio-isotopes est incubé à 65°C (ou autre température signalée) en présence d'une ARN:Ψ-synthase ou d'un complexe sRNP (Charpentier et al., 2005; Charpentier et al., 2007).

Dans le cas d'un complexe sRNP, environ 10 pmoles soit 200 ng de sRNA à boîtes H/ACA sont mis en présence de 150 fmoles d'ARN substrat marqué (10 000 cpm) dans du tampon D1x à température ambiante. Un mélange de protéines (200 nM chacune) est ajouté dans les mêmes conditions que pour les expériences de gel retard. Le volume final de 4,5 µl est incubé à 65°C (ou autre température signalée). Lors des expériences de cinétique, un aliquot de 4,5 µl est retiré d'un mélange comportant 45 µl.

Dans le cas d'une ARN:Ψ-synthase agissant sans sRNA à boîtes H/ACA, l'ARN substrat uniformément marqué est incubé dans du tampon Psi 1x pendant 5 min à 80°C. Lorsque la température est lentement redescendue à 37°C, 0,1 à 1 µg de l'ARN:Ψ-synthase est ajouté pour déclencher la réaction. L'incubation se fait à une température en accord avec l'origine de l'ARN:Ψ-synthase.

Dans les deux cas, la réaction est arrêtée en plaçant le mélange dans la glace et par ajout de 150 µl d'acétate de sodium 0,3 M, 1 µl d'ARNt de levure à 2 µg/µl et 150 µl de phénol-chloroforme (1:1) (v/v). L'ARN présent dans la phase aqueuse est précipité à l'éthanol à 96%.

L'ARN substrat est ensuite totalement digéré par une RNase afin de recueillir des mononucléotides marqués en 3' qui sont fractionnés en Chromatographie sur Couche Mince

(CCM ou TLC) soit en une dimension, soit en deux dimensions (Fig. M.2). Deux RNases différentes peuvent être utilisées. La RNase T2 coupe l'ARN en libérant une extrémité 3' phosphate. Il s'agit donc de placer un phosphate marqué dans la liaison phosphodiester en 3' de l'uridine qui sera modifiée en résidu Ψ. L'hydrolyse totale par la RNase T2 est réalisée dans du tampon T2 1x avec 0,5U de RNase T2 (Invitrogen) dans 8 µl final. L'incubation s'effectue toute la nuit à 37°C. Chaque échantillon est ensuite déposé sur une plaque de cellulose en plastique ou en verre pour chromatographie en couche mince (TLC). Notons que les migrations N1/N2 ne peuvent être effectuées sur des plaques en plastique. La migration est effectuée pendant 4 h dans une cuve et à saturation dans le solvant approprié pour la 1ère dimension (Tampon N1). Si une 2^{nde} dimension est réalisée, la plaque est retournée de 90°, séchée, puis placée en présence du solvant pour la 2^{nde} dimension (Tampon N2 ou RII). Les plaques sont séchées puis exposées au PhosphoImager. Les profils de séparation des taches correspondant aux différentes modifications sont donnés en figure 61.

La RNase P1 est une enzyme qui clive l'ARN en libérant des nucléotides 5' monophosphate. Dans ce cas, un marquage uniforme doit être effectué au [α^{32}P]-UTP et permet de visualiser directement le résidu Ψ formé après digestion à la RNase P1. La digestion est réalisée dans du tampon P1 1x avec 0,3U de RNase P1 dans 5 µl final. L'incubation est réalisée la nuit à 37°C. Les échantillons sont déposés sur plaque TLC, la séparation se fait en une seule dimension et les taches sont quantifiées de la même façon que lors de la digestion par la RNase T2.

La radioactivité contenue dans chaque tache correspondant à chacun des nucléotides 3' monophosphate (Keith, 1995) est quantifiée à l'aide du logiciel « ImageQuant Software Version 5.2 » (Molecular Dynamics). Puisque la composition de la séquence ARN substrat en nucléotide est connue, il est possible de mesurer le taux de résidus U convertis en Ψ (exprimé en mole de résidu Ψ par mole de substrat) selon le tableau 10.

Les données obtenues par les études ont été traitées à l'aide du logiciel « SigmaPlot 2000 version 6.1 » (SPSS Inc.). L'équation utilisée, y = a(1-e-bx) est la fonction la plus simple pour ajuster au mieux les données expérimentales à un modèle mathématique. Les paramètres entrés correspondent à : x = temps en minutes, y = mole de résidu Ψ formées par mole de substrat, a = nombre maximal de moles de résidu Ψ formées par mole de substrat et correspond à l'amplitude, notée A ; b = vitesse apparente du nombre de moles de résidu Ψ formées en min^{-1} et représente une constante de vitesse apparente, notée k_{obs}.

	A	B	C	D	E	F
	Nucléotides	Nombre de nucléotides marqués en 3'	Valeur de quantification (Image Quant)	Quantité relative	Comptage relatif (Ratio)	Rendement de la formation de pseudouridine
1	**Blanc**		C1			
2	**A**	B2	C2	D2=(C2-C1)/B2	E2=D2/D6	F2=(C6-C1)/D2
3	**C**	B3	C3	D3=(C3-C1)/B3	E3=D3/D6	F3=(C6-C1)/D3
4	**G**	B4	C4	D4=(C4-C1)/B4	E4=D4/D6	F4=(C6-C1)/D4
5	**U**	B5	C5	D5=(C5+C6-2C1)/B5	E5=D4/D6	F5=(C6-C1)/D5
6	**Psi**		C6	D6=(D2+D3+D4+D5)/4		F6=(F2+F3+F4+F5)/4

Tableau 10 : Evaluation du taux de résidus Ψ formé à partir des quantifications des taches de CCM. Dans la colonne B sont déterminés les nombre de nucléotides A, C, G ou U dont le phosphate est marqué en 3'. La colonne C rassemble les quantifications provenant du logiciel Image Quant. Les ratios de la colonne E doivent être idéalement proches de 1. Une valeur qui s'écarte trop de 1 ne permet pas une quantification du taux de pseudouridylation au final car traduit un défaut dans la composition de l'ARN substrat marqué. La valeur F6 traduit le taux de modification de U en Ψ exprimé en moles U/ moles de substrat.

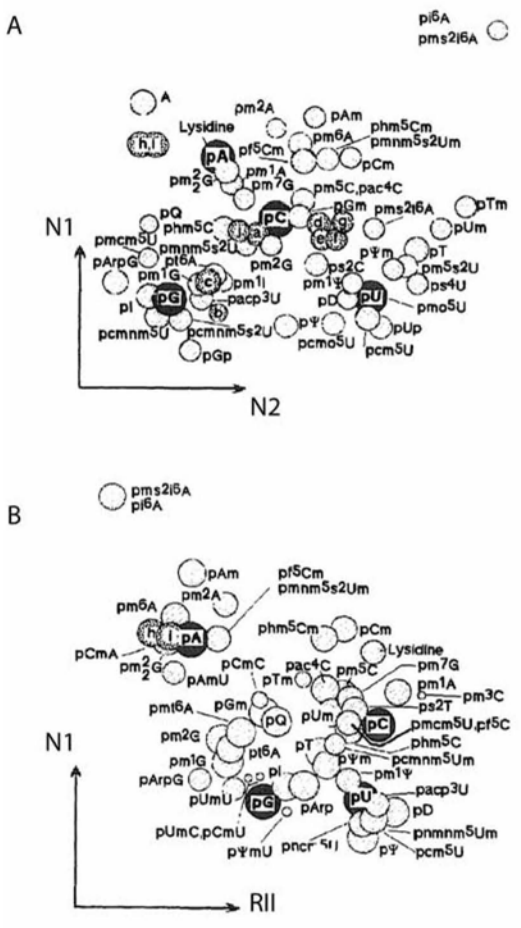


Figure 61 : Positionnement relatif en chromatographie en deux dimensions sur couche mince des taches correspondant aux nucléotides majeurs et nucléotides modifiés de l'ARN (Keith, 1995). Deux distributions différentes sont observées selon l'utilisation de tampon N2 (**A.**) ou RII (**B.**) pour la seconde

dimension de séparation par chromatographie. La distribution est ici celle de mononucléotides portant un phosphate marqué en 5' mais ne diffère pas significativement de la distribution de mononucléotides portant le phosphate marqué en 3'.

3. Méthodes concernant les protéines

3.1. Préparation d'extraits cellulaires

3.1.1. Préparation d'extraits cellulaires de levure

Le protocole expérimental pour la préparation d'extrait de levure est identique à celui utilisé pour l'extraction d'ARN de la première étape à la fin de l'agitation avec les billes de verres. En effet, après cette étape, les billes de verre sont centrifugées pendant quelques secondes à 13000 rpm. Le surnageant est prélevé et soumis à une nouvelle centrifugation à 4°C pendant 15 min à 13000 rpm. Le nouveau surnageant est aliquoté en fractions de 50 µl qui sont ensuite congelées dans de l'azote liquide et stockées à -80 °C.

3.1.2. Préparation d'extraits cellulaires de *P. abyssi* et *S. solfataricus*

Les extraits cellulaires sont obtenus par sonication de 10^{10} cellules de ces espèces dont la culture cellulaire a été stoppée en fin de phase exponentielle. Les culots cellulaires sont repris avec 200 ml de tampon d'extraction : Tris-HCl 25 mM, pH 7,5 ; KCl 25 mM ; $MgCl_2$ 5 mM ; Glycérol 10 % ; DTT 2 mM ; solution soumise à l'autoclave avant l'ajout de DTT. La sonication se fait par 5 séries de 20 sec en maintenant l'échantillon à froid. L'échantillon est centrifugé à 10 000 g puis le surnageant (S 10) est récupéré. La concentration en glycérol est ajustée à 20% en vue d'une conservation à -20°C.

3.2. Production des protéines recombinantes

Les stratégies de production et de purification sont identiques pour les protéines Pus7 sauvage et mutée au niveau du site actif et aboutissent à l'obtention d'une protéine recombinante portant l'étiquette His6x à son extrémité N-terminale.
La culture des bactéries BL21 DE3 Codon Plus transformées est réalisée en milieu LBK dans un volume de 500 ml. Lorsque l'absorbance de la culture à 600 nm est comprise entre à 0,6 et 0,9 U/ml (milieu de phase exponentielle de croissance), la production de la protéine est induite par l'addition de 1 mM d'IPTG et la culture poursuivie pendant 15 heures à 20°C.

3.3. Purification des protéines recombinantes

3.3.1. Sonication

En fin de culture, les cellules sont centrifugées à 4000 rpm pendant 25 min.

Les culots cellulaires sont repris dans 8 ml de tampon de lyse (Tableau 3). Les cellules sont traitées aux ultrasons, sous une puissance de 60 W, pendant 2 min par périodes de 2 s suivies de 2 s de repos. Après sonication, le lysat bactérien contenant les protéines, les acides nucléiques et les débris cellulaires est centrifugé pendant 25 min à 15 000 rpm à 4°C. Le surnageant contient la protéine taggée Pus7-His6x soluble tandis que les débris cellulaires sont éliminés dans le précipité.

3.3.2. Purification des protéines portant l'étiquette « His-6x » : chromatographie d'affinité sur ions métalliques immobilisés (IMAC)

La méthode de purification par chromatographie d'affinité sur complexe Ni^{2+} utilisée pour la purification de la protéine d'intérêt est basée sur l'introduction d'une séquence supplémentaire contenant 6 résidus histidine à son extrémité N-terminale. En effet, le cycle imidazole de l'histidine est capable de former des complexes avec les ions métalliques divalents tels que le Ni^{2+}. La résine Ni-NTA Agarose (QIAGEN) est composée de Sepharose CL-6B couplée à des groupements nitriloacétiques qui possèdent également un pouvoir chélatant vis-à-vis de ces ions, ce qui leurs donnent la propriété de retenir sélectivement des protéines présentant des résidus histidine accessibles.

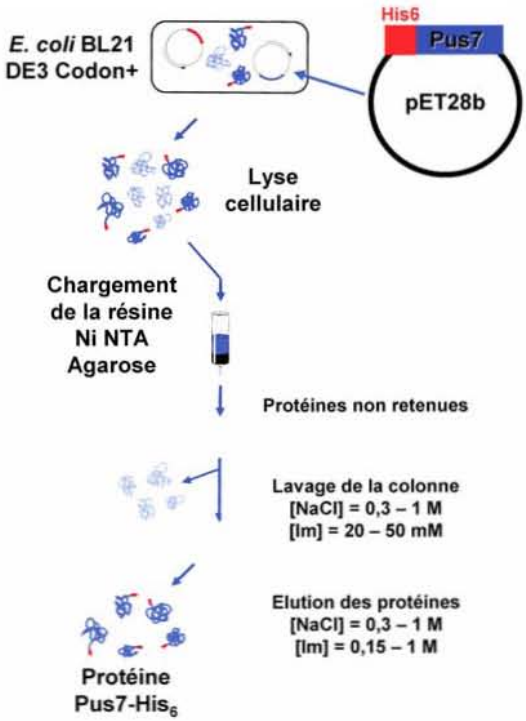


Figure 62 : Représentation schématique des principales étapes du protocole de production et purification des protéines recombinantes utilisées pour notre étude.

L'utilisation de ce support chromatographique a été réalisée à pression atmosphérique dans de petites colonnes en plastique d'un volume de 10 ml, obturables à chaque extrémité, dans lesquelles ont été introduit 500 µl de résine permettant la fixation d'approximativement 5 mg de

protéine. Avant d'introduire le surnageant de centrifugation dans la colonne, la résine a été préalablement rincée avec 5 ml d'eau stérile puis avec le même volume de tampon de lyse contenant 300 mM de NaCl qui limite les interactions ioniques non spécifiques des protéines avec la matrice de la colonne. Afin d'assurer l'accessibilité optimale des ions Ni^{2+} aux chaînes latérales des résidus histidine, la résine est homogénéisée sur une roue motorisée à 4°C pendant 30 min.

A la fin de cette étape, la résine est lavée avec 5 ml de tampon de lavage, puis les protéines sont éluées avec 3 ml de tampon d'élution dont la composition figure dans le Tableau 9 (Figure 62).

3.4. Caractérisation des protéines

3.4.1. Analyse par électrophorèse en gel de polyacrylamide en conditions dénaturantes

Nous avons utilisé cette technique d'électrophorèse selon Laemmli (Laemmli, 1970) afin d'analyser la qualité des échantillons protéiques (concentration, contamination) que nous avons utilisé lors de notre étude. L'appareillage utilisé est de référence Mini Protean II et III® (Biorad) et la composition des gels est détaillée dans le protocole suivant :

1. Préparation du gel

Laver les plaques, espaceurs et peigne au savon. Rincer à l'eau puis à l'alcool technique. Faire le montage quand les plaques sont parfaitement sèches.

Gel de séparation (resolving gel) Préparer le mélange suivant : 3.3 ml de solution d'acrylamide/bisacrylamide, 2.5 ml de Tris pH 8.8, 0.1 ml de SDS 10%, 0.1 ml d'APS, qsp 10 ml H2O. Ajouter 10 µl de TEMED et couler rapidement le mélange entre les deux plaques jusqu'à 1.5 cm du bas du peigne (faire une marque au préalable). Déposer délicatement 1 ml d'eau à la surface du gel.

Après polymérisation (généralement 5 à 15 min), éliminer l'eau.

Gel de concentration (stacking gel) 1.5 ml de solution d'acrylamide/bisacrylamide, 2.5 ml de Tris pH 6.8, 0.1 ml de SDS 10%, 0.1 ml d'APS, qsp 10 ml H2O puis ajouter 10 µl de TEMED et couler. Positionner rapidement le peigne en évitant la formation de bulles. Après polymérisation, enlever délicatement le peigne et rincer les poches à l'eau puis à l'alcool.

2. Préparation des échantillons et électrophorèse

Dans un tube ependorf, préparer les dépôts en mélangeant l'échantillon protéique avec 5 µl de bleu de dépôt SBL dans un volume final de 20µl.

Dénaturer les échantillons 3 min à 96°C au bain sec. Laisser refroidir puis centrifuger brièvement à 13000 rpm.

Charger les échantillons après avoir rempli les deux réservoirs avec du tampon de migration.

La migration s'effectue à courant constant 200V pendant 1H.

Lorsque le front de migration (bleu de bromophénol) est en bas du gel, couper le générateur et démouler le gel.

3. Coloration du gel
Eliminer le gel de concentration et colorer au bleu de Coomassie sous agitation douce. Pour cela, immerger le gel pendant 3 x 10 min dans 100 ml d'eau déminéralisée puis au moins 60 minutes dans 20 ml de la solution de coloration. Laver le gel à l'eau (5 min).

3.4.2. Analyse par Western-Blot

Nous avons utilisé cette technique afin de vérifier la qualité des protéines produites avec une étiquette poly-histidine. Le protocole est le suivant :

1**.** Electrophorèse dénaturante (SDS-PAGE) en cuve Mini Protean BioRad dans du tampon d'électrophorèse (Tris-HCl 25 mM, Glycine 192 mM, SDS 0.1% pH 8.3)

2. Transfert liquide des protéines du gel sur membrane de nitrocellulose en cuve Mini-Trans Blot BioRad dans du tampon de transfert (25 mM, glycine 192 mM, SDS 0.01%, 20% éthanol absolu)
Montage de la cassette de transfert :
Orientation de la cassette dans la cuve transfert face noire vers le face claire vers le positif.
Humidifier les différentes couches de la «galette» dans le tampon de transfert.
Les protéines migrent du vers le positif, donc sur la face noire de la cassette entasser :
- Scotch brite blanc.
- 2 à 3 épaisseurs de papier Whatman 3mm.
- Le gel SDS-PAGE.
- La membrane de nitrocellulose.
- 2 à 3 épaisseurs de papier Whatman 3 mm.
- Scotch brite blanc.

Pour un bon transfert, bien enlever toutes les bulles entre les différentes couches.
Migration à 100V constants, 1 heure.

3. Vérification de la qualité du transfert par coloration transitoire.
Incuber la membrane pendant1 min dans une solution de rouge ponceau à 0,1g dans 100ml d'acide acétique 1%.
Décolorer à l'eau sous agitation légère (la décoloration des protéines est très rapide mais néanmoins plus rapide là où il n'y a pas de protéines...).

4. Hybridation de la membrane, sous agitation sur table orbitale.
Saturation de la membrane dans 10 ml de tampon TBST (Tris 25 mM, pH 7.6, 137 mM NaCl, 27 mM KCl, Tween X 20 0.1%) contenant 5% lait en poudre

Incubation1 heure à T°C ambiante ou la nuit, 4°C.
Lavages de la membrane en utilisant le tampon TBST : 1x15 min + 2x5 min à T° ambiante.

1er anticorps : Anticorps anti-His6x monoclonal dilué au1/1000ème dans 5 ml de tampon TBST contenant 2.5% de lait en poudre.
Incubation de 1 heure, T° ambiante ou la nuit, 4°C.
Lavages de la membrane en utilisant le tampon TBST : 1x15 min + 2x5 min à T° ambiante.
2ème anticorps : antisourisHRP (His) dilué au 1/1000ème dans 5 ml de tampon TBST.
Incubation 1 heure, T° ambiante.
Lavages de la membrane au TBST: 1x15 min + 2x 5 min, T° ambiante.

5. Révélation du complexe protéine/anticorps par autoradiographie avec le kit de Chimiluminescence (ECL) utilisé selon les recommandations du constructeur.
Au moment de l'emploi mélanger 1 ml de chacun des 2 réactifs du kit (1 membrane)
Incubation de la membrane pendant 1 min dans ce mélange.
Mise en contact avec un film autoradiographique.

3.4.3. Analyse de la formation de complexes ARN/protéine

Les expériences de gel retard sont réalisées comme détaillé dans la publication N°2 présentée dans la partie « Résultats ». Un mélange comprenant 50 fmol d'ARN uniformément marqué (10 000 cpm) et potentiellement 0,5 µg d'ARNt de levure (Roche Molecular Biochemicals) et l'ARN substrat froid (100 ng, 15 pmol) est incubé dans 3,5 µl de tampon D 1x, 5 min à 65°C, puis 10 min à température ambiante. Après cette étape de renaturation de l'ARN, la formation des complexes est initiée par ajout d'1 µl du mélange de protéines (concentration finale : 200 nM chacune). Les échantillons sont incubés 10 min à 65°C. Puis 4 µl de bleu CSB sont ajoutés aux échantillons qui doivent être rapidement déposés sur gel. Les complexes ARN-protéines sont fractionnés par électrophorèse en gel polyacylamide-glycérol 6% non dénaturant. L'électrophorèse se fait à 100 V pendant 3 h à température ambiante.

3. Biologie structurale

A. Matériel

1. Souches bactériennes

Nous avons testé différentes souches pour la surproduction des bactéries dans l'objectif d'études structurales qui présentent de nombreux avantages parmi lesquels une croissance rapide sur des milieux bon marché et à haute densité (en fermenteur), une déficience en protéases

(lonB, OmpT), une surexpression des ARNt rares (arginine, leucine et isoleucine) et la possibilité d'utiliser un grand nombre de systèmes d'expression compatible. Ces souches sont :

- **BL21 (DE3)** (F^-, *omp*T, lon, *hsd*S($r_{B-}m_{B-}$), *dcm*+, Tet^r, *gal* λ(DE3), *end*A.
- **BL21 CodonsPlus (DE3) RIL** (F^-, *omp*T, lon, *hsd*S($r_{B-}m_{B-}$), *dcm*+, Tet^r, *gal* λ(DE3), *end*A, Hte[*arg*U *ile*Y *leu*W Cam^r]).
- **Rosetta (DE3) pLysS** F^- ompT $hsdS_B(R_B^- m_B^-)$ gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysSRARE (Cam^R).
- **C41/C43 (DE3)** (F^-, *omp*T, lon, *hsd*S($r_{B-}m_{B-}$), *dcm*+, Tet^r, *gal* λ(DE3), *end*A + mutations inconnues.

2. Milieux de culture

La composition des différents milieux est décrite dans le tableau 11.

Conditions d'utilisation	Dénomination	Composition
Production des protéines (Studier, 2005)	ZYM 5052	Tryptone 1%, extrait de levure 0,5%, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4
	Gly-Lac-Glu	25% Glycérol, 10% Lactose, 2,5% Glucose
	$MgS0_4$	1M $MgSO_4$
	Mix Métaux I	50 mM $FeCl_3$
	Mix Métaux II	20 mM $CaCl_2$, 10 mM $MnCl_2$, 10 mM $ZnSO_4$, 2 mM $CoCl_2$, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2MoO_4, 2 mM Na_2SeO_3, 2 mM H_3BO_3
Marquage des protéines (Se-MET)	SeMET-MM	1g NH_4Cl, 3g KH_2PO4, 4g $Na_2HPO_4.2H_2O$ H_2O qsp 1 L pH=7,4
	Solution Glc-Thiamine	20g D-Glucose, 0,3g $MgSO_4$, 10 mg $Fe_2(SO4)_3$, 10 mg Thiamine (Vitamine B1), H_2O qsp 100 mL

Tableau 11 : Milieux de culture utilisés pour la biologie structurale.

B. Méthodes

1. Production des protéines

1.1. Système de production

Les exigences de la biologie structurale orientent fortement le choix des systèmes de production de protéines limitant d'emblée l'utilisation des systèmes peu productifs (cellules de mammifère), ceux qui provoquent des modifications de protéines (glycosylation chez les levures par exemple) ou ne permettant pas les marquages (absence d'auxotrophe). Une analyse rapide de la littérature indique que la majorité des structures tridimensionnelles de protéines dérivent de molécules recombinantes produites dans *E. coli*. En effet, la croissance des cristaux de protéines (pour la cristallographie) ou la préparation de solutions concentrées de protéines (pour la RMN) ne peuvent se faire que si les protéines sont produites et purifiées en quantité et avec une qualité suffisante (Tableau 12).

PARAMETRES	Cristallographie R-X	RMN
Quantité totale	>10mg	0,5 ml à 1 mM (soit environ 10 mg pour une protéine de 25 kDa)
Concentration	Environ 10 mg/ml	1 mM
Pureté	>98%	>98%
Homogénéité	Absence dans la plupart des cas de modifications post-traductionnelles hétérogènes (glycosylation), clivage interne, hétérogénéité des extrémités, régions flexibles/instables pouvant compromettre la croissance des cristaux	Idem mais tolérance supérieure pour les régions flexibles/instables (non -visible)
Oligomérisation	Pas de limite de taille des complexes mais il doivent être homogènes (complexes stables)	10 kDa en absence de marquage 15 kDa marquage 25-30 kDa double ou triple marquage 15N+13C+2H
Marquage	Marquage par sélénométhionine souvent requis (MAD)	15N/13C/2H

Tableau 12 : Paramètres qualitatifs et quantitatifs des échantillons protéiques utilisés en biologie structurale.

Nous avons donc utilisé plusieurs souches bactériennes pour la production de protéines recombinantes qui sont toutes dérivées de la souche *E. coli* BL21 qui sont décrites dans la partie matériels de ce chapitre.

1.2. Système d'expression

Le taux de production d'une protéine peut être modulé en contrôlant le nombre de copies du gène codant pour la protéine (d'une copie unique si le gène est inséré dans le chromosome bactérien à plusieurs centaines pour certains plasmides) ou en utilisant des promoteurs régulables. Pour notre étude, nous avons travaillé avec l'un de ces systèmes d'expression les plus efficace et utilisant un plasmide de la famille pET (pET28b Novagen) permettant la production de protéines portant l'étiquette His6x à l'extrémité N- ou C-terminale. Ce système d'expression est basé sur le promoteur tardif du phage T7 placé dans un vecteur à nombre de copies moyen. La régulation du promoteur est réalisée, en trans, par la T7 ARN polymérase, elle-même sous le contrôle du promoteur lacUV5. Plusieurs améliorations substantielles de ce système ont été réalisées, notamment pour limiter les problèmes de toxicité causés par le fort niveau de transcription provoquant l'instabilité des plasmides et du niveau d'expression. Une amélioration remarquable est issue de la sélection empirique de souches de *E. coli* (C43DE3, C41DE3) capables de tolérer l'expression de protéines solubles ou membranaires avec le système de la T7 ARN polymérase (Miroux & Walker, 1996).

1.3. Conditions de culture

Les conditions de culture peuvent avoir un effet important sur le taux de production de protéine recombinante et leur solubilité. Nous présenterons donc les 2 systèmes qui nous ont permis d'obtenir les protéines nécessaires au travail de cristallogenèse.

1.3.1. Culture en batch

Lorsque nous avons débuté au laboratoire la production de protéines en vue d'étude structurale au laboratoire, nous ne disposions d'aucun équipement de fermentation. Nous avons donc réalisé les premières cultures à grande échelle (8L) en erlen en utilisant soit le milieu LB dans les conditions décrites dans la partie 3.2. soit un milieu auto-inductible ZYM5052 (Studier, 2005) dont la composition est détaillée dans la partie matériel de ce paragraphe. Le protocole que nous avons utilisé est le suivant :

1. Préculture

Ensemencer ~ 100 à 150 mL de LBK avec quelques paillettes de cellules issues d'un stock de cellules de souche BL21 Codon+ transformées par le plasmide d'intérêt (Congel Cell) et stockées à -80°C dans 30% glycérol.

Laisser la culture O/N à 37°C afin d'atteindre la saturation A= 4-5 UDO_{600nm}/mL

2. Culture

Réaliser une culture de 8L divisée 8x1L en erlen de 2L en suivant les étapes du protocole :

La veille préparer le milieu ZYM 5052.

En milieu d'après-midi (vers 16H), ensemmencer chaque erlen à 0,1 UDO_{600nm}/mL à partir de la préculture après avoir ajouter les sucres (Gly-Lac-Glu), le $MgSO_4$, les mélanges de métaux I et II puis l'antibiotique.

Laisser sous agitation à 200 rpm à 37°C pendant 16H minimum.

Contrôle de la culture bactérienne

Prélever un aliquot de culture, faire une dilution au 1/100 et mesurer la DO à 600 nm.

Vérifier que la DO finale est convenable : $4 < DO < 8$ UDO_{600nm}/mL si ce n'est pas le cas, remettre la culture sous agitation en réalisant un suivi de Do toutes les 2 heures jusqu'à ce que la densité cellulaire soit suffisante.

Récolter les cellules par centrifugation à 4000rpm pendant 20 min dans les pots de 1L (J6)

<u>Rq</u> : La culture dégage souvent une forte odeur de « fermentation » due l'utilisation du lactose par la bactérie.

Reprise des culots dans le tampon de lyse bactérien

En principe, les cellules sont ressuspendues de manière à avoir une densité de $100DO_{600nm}$/mL, cependant pour limiter le volume de sonicat nous diminuerons la quantité de tampon de lyse.

Reprendre les cellules dans le tampon de lyse (20 mM TRIS-HCl pH 7,5, 1 M NaCl filtré à 0,22 µm et dégazé) suivant à raison de 16 mL/L de culture :

Bien homogénéiser afin que ne subsistent aucuns amas de cellules, au besoin utiliser le vortex (attention car l'utilisation du vortex génère une mousse importante) ou des aller-retour à la pipette P5000.

3. Sonication

Pour limiter la protéolyse il est conseillé de travailler le plus souvent possible dans la glace (4°C), température à laquelle la protéine est stable (test effectué sur 1 mois au frigo).
Penser à récupérer des aliquots de 100 µl de surnageant afin d'effectuer une analyse qualitative et quantitive de notre protéine lors de chaque étape de purification.

Cette étape sera effectuée sur le sonicateur BRANSON en utilisant la sonde de taille normale et une cupule en plastique d'une contenance ~ 15 mL (falcon de 50 mL redecoupé).
Durée de la sonication → 1 x 4 min 50% Duty / Output = 6

Rq : la qualité de la lyse est validée par l'apparition d'un sonicat visqueux (ADN génomique). Attention, veillez à allonger le temps de sonication si la lyse n'est pas optimale car l'efficacité de la lyse est moins importante dans le plastique.

Collecte des débris cellulaires
Centrifuger des échantillons dans les tubes J21 à 20000 rpm pendant 30 min à 4°C.
Récupérer le surnageant dans 2 falcons de 50 mL (vol env. 100 mL pour 8 L de culture).
Prélevez un aliquot de 100 µl de surnageant pour analyse ultérieure sur gel SDS-PAGE.

4. Clarification du surnageant par choc thermique à 75°C (protéines thermostables)

Cette étape a pour objectif d'améliorer la qualité de la purification en éliminant une partie des protéines bactériennes non thermostables.

Plonger les erlens dans un bain marie à 75°C pendant 30 min en vortexant après 15 min d'incubation.
Rq : L'efficacité de cette opération est validée par l'apparition d'un précipité blanchâtre à l'intérieur du falcon.
Transvaser le précipité dans des tubes J21 puis centrifuger 30 min à 20000 rpm.
Récupérer le surnageant dans 2 falcons de 50 mL.
Prélevez un aliquot de 100 µl de surnageant pour analyse ultérieure sur gel SDS-PAGE.

5. Précipitation des acides nucléiques au PEI 0,1% final

Cette étape a pour objectif d'éliminer les acides nucléiques (ADN, ARN) liés plus ou moins spécifiquement à notre protéine (basique) dans l'objectif d'augmenter la pureté de l'échantillon et de favoriser la cristallisation.

Attention : Procéder avec méthode car le PEI est un agent précipitant des protéines.

<u>Rq</u> : Une solution mère à ~ 10%-20% sera réalisée par dialyse dans des boudins de type Spectra Pore 7 (gros diamètre) sur la nuit à T°C ambiante à partir d'une solution de PEI 50% SIGMA. Procéder par pesage car la solution stock est très visqueuse et impossible à pipeter. Le tampon utilisé pourra être adapté à celui de purification mais l'utilisation d'une solution de PEI dans de l'eau stérile est possible. 10 mL de solution de PEI à ~ 10% sont suffisant pour une vingtaine de purifications. Cette solution pourra être conservée durant 6 mois à 4°C.

A partir d'une solution de PEI à ~ 10% dans de l'eau (ou dans le tampon d'intérêt), calculer le volume nécessaire pour atteindre la concentration finale dans l'échantillon de 0,1% (vol à ajouter inférieur à 1 mL)
Bien vortexer, la solution stock car le PEI 10% est visqueux, prélever le vol à la pipette puis introduire quelques gouttes dans le falcon et vortexer immédiatement pour éviter la précipitation protéique à l'interface PEI/solution.
Recommencer l'opération jusqu'à l'ajout complet du PEI.

<u>Rq</u> : La réaction de précipitation des acides nucléiques est quasi-immédiate à température ambiante, elle se traduit par l'apparition d'un précipité blanc.

Laisser agiter à T°C ambiante sur roue motorisée pendant 5 min.
Transvaser le précipité dans des tubes J21 puis centrifuger 30 min à 20000 rpm.
Récupérer le surnageant dans des falcons de 50 mL.
Prélevez un aliquot de 100 µl de surnageant pour analyse ultérieure sur gel SDS-PAGE.

6. Conservation des échantillons avant purification dans la glace à 4°C

En absence d'équipement de fermentation, nous avons remarqué que les deux paramètres qui ont le plus d'influence sur le niveau de production de nos protéines solubles produites dans des cultures en « batch » (de 50 ml jusqu'à 2 L) sont la densité de cellules à laquelle le déclenchement de la production est réalisé (induction à une densité optique à 600nm comprise entre 0.4 et 1 unité) et surtout la température (une baisse de la température, jusqu'à 15°C, limite la formation de corps d'inclusion). Ce travail d'optimisation des conditions de production de la protéine Pus8 de *S. cerevisiae* à fait l'objet d'une publication dans *Journal of Chromatography B*, dont je suis premier auteur.

Publication n°1

Optimisation of expression and purification of the recombinant Yol066 (Rib2) protein from *Saccharomyces cerevisiae*.

A. Urban, I. Ansmant and Y. Motorin

J Chromatogr B Analyt Technol Biomed Life Sci., **786(1-2)**,187-195, 2003

Publication n°1

Available online at www.sciencedirect.com
SCIENCE DIRECT®
JOURNAL OF CHROMATOGRAPHY B
ELSEVIER

c
C

8

m

©

|

C

C

1 1

C C

©

eukaryotes [7,9]. Several prokaryotic members of tRNA- and rRNA:C -synthases families have been expressed in *E. coli* and have been well-characterised biochemically [10–14]. However, the expression in *E. coli* of eukaryotic RNA:C -synthases and, in general, eukaryotic RNA:modification enzymes, is often complicated by problems related to inoptimal codon usage, limited level of expression and insolubility of obtained recombinant proteins. Among eukaryotic RNA:C -synthases only yeast Pus1 (together with its mouse homologue mPus1) and Pus4 proteins were successfully expressed in *E. coli* with a high yield and in a soluble form [15–18]. The other reported attempts of heterologous expression (yeast Pus5 and Pus6 proteins expressed in *E. coli*) were not successful, due to accumulation of the recombinant product in inclusion bodies ([8,19] and unpublished observations). Detailed studies of yeast tRNA:C -synthase Pus1 have already pointed out the importance of cell culture media composition in achieving the highest yield and optimal solubility of the product [15]. Another complication in such heterologous expression–purification procedures was the strong tendency of the protein to aggregate, which limits the final recovery level. Eukaryotic expression systems like methylotrophic yeast *P. pastoris* may be attractive alternatives to *E. coli*, but their practical use is much more laborious at the level of cloning and not as straightforward as prokaryotic expression in *E. coli*. Taking together these considerations, we decided to optimise the expression and purification protocols developed for putative RNA:C -synthase/deaminase Yol066. The final aim was to produce sufficient amounts of protein required for detailed functional and physicochemical characterisation and, ultimately, crystallisation studies.

The efficiency of recombinant protein expression in bacterial host cells like *E. coli* depends on numerous parameters. Some of them are directly defined by the expression vector used for gene cloning (such as promoter and terminator signals, ribosome binding sequence efficiency and possibility of regulation by repressors, for review see [20,21]). Other important factors are linked to nucleotide and amino acid sequence of the expressed protein (appropriate codon usage, mRNA and protein stability). In most cases these parameters are predefined by the choice of the expression vector and protein sequence to be expressed. Thus, optimisation of protein expression may be performed by varying parameters affecting *E. coli* cell growth (media composition, temperature) and protein solubility (inducer's concentration, induction time). The other area of optimisation relates to developing of a rapid and efficient purification protocol to minimise the losses during purification.

Numerous bacterial expression systems are described in the literature and are commonly used in laboratory practice. Most of them rely on the use of a strong bacterial promoter (such as *lac*, *trp* or synthetic *tac* promoters) and the induction by IPTG (see [20] and Refs. therein). These systems are rather efficient for protein production but their regulation is not as tight as two-step induction systems based on phage T7 late promoters. Protein expression in such host-vector couples (*E. coli* BL21 (DE3) strains and pET-family plasmid) depends on a transcription unit which supplies the T7 RNA polymerase. In addition to the tight regulation and high level of protein expression achieved by pET-plasmid/*E. coli* BL21 (DE3), the availability of numerous versions of pET-derived expression plasmids facilitates the design of expression construct and allows the use of affinity tags for recombinant protein purification [22].

Another limitation in recombinant protein expression is related to differential usage of synonymous codons between *E. coli* and heterologous nucleotide sequences [23,24]. Indeed, arginine codons AGA and AGG are rare codons in *E. coli* while they are frequently used in many eukaryotic organisms [25]. This problem in recombinant protein expression is frequently solved by co-transformation of the producing strain by a plasmid bearing the corresponding rare tRNA gene or even by integration of such tRNA at the genome level [26,27].

Here we report the optimised protocol for expression in *E. coli* and purification of recombinant N-terminally His_6-tagged yeast protein Yol066. The expression was performed using pET28 plasmid as the expression vector and *E. coli* BL21 (DE3)RIL as the host strain. The optimisation was achieved by adaptation of media composition, temperature and inducer's concentration to increase the yield of soluble protein. The purification protocol based on a nickel–nitrilotriacetic acid (Ni^{21} –NTA) agarose af-

finity column and ion-exchange chromatography steps is also described.

2. Experimental

2.1. Chemicals

Ni^{21}–NTA agarose was from Qiagen (France), gel filtration molecular mass standards from BioRad (USA), restriction enzymes (*Nhe*I, *Sma*I, *Bam*HI), kanamycin and IPTG from Euromedex (France). All other chemicals were from Merck Biochemicals (Germany). Luria Broth (LB) and minimal M9 media were prepared according to standard procedures [28]. M9 medium was systematically complemented by 2% of casamino acids (Difco, UK).

2.2. Equipment

The following prepacked chromatographic columns and equipment were used: MonoQ HR 5/5, Resource Q 1 ml, Superose 12 HR 10/30, DEAE TSK 5 PW (83 75 mm) and the fast protein liquid chromatography (FPLC) medium-pressure chromatographic system (all from Pharmacia, Sweden). Protein gel electrophoresis was performed using Mini-Protean II and protein bands were quantified using a GelDoc 1000 CCD camera system and Molecular Analyst software package (BioRad, USA).

2.3. Strains and plasmids

Escherichia coli strain BL21-CodonPlus(DE3) RIL (*E. coli* B F^2 *ompThsdS*(r_B^2 m_B^2) *dcm*^{1} Tet^r gal l (DE3) *endA* Hte [*argUileYleuW* Cam^r] (Stratagene, USA) was used for expression of the recombinant Yol066 protein.

2.3.1. Preparation of pET28b–Yol066 expression vector

Gene encoding yeast protein Yol066 was amplified by polymerase chain reaction (PCR) using yeast genomic DNA and two DNA oligonucleotides, corresponding to the 59-extremity (ATG-start codon) and to 39-UTR, respectively, of the gene. The 59-oligonucleotide had an *Nhe*I restriction site in frame with starting ATG of YOL066. Amplified blunt-ended PCR product was directly inserted in pUC18 cloning vector cut by *Sma*I restriction endonuclease. The desired orientation of the insert was found by restriction analysis and the complete insert sequence was confirmed by a dideoxy sequencing approach. This intermediate pUC18–Yol066 plasmid was used for further manipulation to construct the pET28–Yol066 expression vector. The pUC18–Yol066 plasmid was cut by *Nhe*I–*Bam*HI restriction enzymes (*Bam*HI site is located downstream to YOL066 stop codon) and the purified fragment was inserted to *Nhe*I/*Bam*HI-digested pET28b expression vector. The correct fragment integration was verified by restriction analysis and partial DNA sequencing.

2.4. Protein expression and cell disruption

E. coli cells bearing pET28b–Yol066 plasmid were grown in rich (LB) and in a minimal (M9) media supplemented by 100 mg/l of kanamycin (Euromedex, France). Preculture in 20 ml of LB medium was inoculated directly from glycerol stock at 2 80 8C and grown overnight at 37 8C. Then, this preculture was used to inoculate 600 ml LB or M9 media (about 1/40 of the final volume) in order to obtain an initial AU_{600} of 0.1. Cultures were grown at 37 8C with vigorous shaking (200 rpm) to AU_{600} of 0.7 (| 23 10^8 cells/ml). Yol066 expression at three different temperatures (18 8C, 28 8C and 37 8C) and three different IPTG concentrations (50 m*M*, 300 m*M* and 1000 m*M*) was studied in 100 ml aliquots. After 3 or 15 h of induction, cells were harvested by centrifugation at 4 8C for 25 min at 4000 *g*. The cells were washed with 5 vol of sterile water and once more pelleted in the same conditions.

Lysis of cells was done in 5 vol of lysis buffer (50 m*M* $NaPO_4$ buffer, pH 7.7, 300 m*M* NaCl, 10 m*M* imidazole) by 1 min sonication (30 cycles of 2 s and 2 s on ice between each pulse). Soluble (S10) and insoluble (P10) fractions were separated by centrifugation at 4 8C for 25 min at 16 000 *g*.

2.5. Monitoring the expression level by SDS–PAGE analysis

Amount of soluble Yol066 protein produced upon expression was estimated by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–

PAGE). Electrophoresis was performed in the presence of SDS according to the method of Laemmli [29]. The discontinuous gel consisted of a 5% stacking gel and a 10% separating gel which was run on a vertical electrophoresis unit (Mini-Protean II, BioRad) at 200 V for 45 min. Aliquots of S10 extract (10 ml, about 50 mg of total protein) were loaded on the SDS–PAGE and protein bands were stained with Coomassie Blue G-250. The intensity of protein bands on SDS–PAGE was measured by densitometry using the GelDoc system and Molecular Analyst software package (BioRad). The final yield of Yol066 was calculated in arbitrary units taking into account the amount of soluble Yol066 in S10 and the cell culture density after induction. Although the volume of soluble S10 extract loaded on the gel was the same, the amount of protein varies depending on growth conditions. Therefore, we normalise the intensity of the Yol066 band to the signals of two other constitutive *E. coli* proteins. Cell culture density was monitored photometrically by reading the optical density at 600 nm.

2.6. Data analysis

SYSTAT software v. 10 was used for surface analysis of the experimental data. The quality of the fit to the non-linear model equation was expressed by the coefficient of correlation R^2 and the significance of regression coefficient was tested by a *t*-test.

2.7. Chromatographic procedures

2.7.1. Immobilised metal affinity chromatography (Ni^{21} –NTA agarose)

For preparative purification of the recombinant Yol066 the *E. coli* culture was grown using the optimised expression conditions (LB medium, 18 8C, 50 m*M* IPTG, 15 h of induction). Cells harvested from 3 l of culture (about 15 g of cells) were resuspended in 50 ml of the extraction buffer and lysed by sonication as described above. The soluble protein fraction (S10 cell extract) was batch-loaded onto 2 ml of Ni^{21} –NTA agarose (Qiagen). The resin was previously washed with 3 vol of sterile water and equilibrated with 3 vol of the lysis buffer.

Adsorption was done for 30 min at 4 8C with constant mixing on a motorised wheel. After incubation the resin was transferred into a small column (103 100 mm) or left in batch for following wash and elution. The resin was washed with 5 vol of wash buffer (50 m*M* $NaPO_4$, pH 7.7, 300 m*M* NaCl, 20 m*M* imidazole). Elution of bound protein was performed using 3 ml of elution buffer containing imidazole at different concentrations (50 m*M* $NaPO_4$, pH 7.7, 300 m*M* NaCl, 100, 150, 200 or 250 m*M* imidazole).

Before performing the second purification step, concentrated protein fractions were pooled and subjected to overnight dialysis against 25 m*M* Tris–HCl, pH 7.5, 100 m*M* NaCl, 100 m*M* imidazole and 1 m*M* 2-mercaptoethanol to remove the excess of NaCl present in the elution buffer.

2.7.2. Ion-exchange chromatography

Three different chromatographic supports were tested for further purification of the recombinant Yol066 protein. Two columns had strong tertiary aminoethyl groups (MonoQ and Resource Q) and the third one was a weak anion-exchanger DEAE TSK. Dialysed protein fraction from Ni^{21} –NTA column (200 ml, about 1 mg of Yol066 protein) was loaded onto a column equilibrated with buffer A (25 m*M* Tris–HCl, pH 7.5, 1 m*M* 2-mercaptoethanol) and chromatographied at a flow-rate of 1 ml/min using the Pharmacia FPLC system. The column was washed with 3 column vol of the same buffer, bound protein fraction was eluted with a linear NaCl gradient (0 –1 *M*) in 30 ml of buffer A (80 ml for DEAE–TSK column). The fractions containing Yol066 protein were pooled and dialysed against 25 m*M* Tris–HCl, pH 7.5, 100 m*M* NaCl, 2 m*M* DTT, 50% of glycerol and stored at 2 20 8C. The presence of Yol066 protein in eluted protein fractions was controlled by SDS–PAGE. Protein concentration in samples was determined by the Bradford dye binding method [30] using bovine serum albumin as the protein standard.

2.7.3. Gel filtration on a Superose 12 column

Native molecular mass determination was performed by gel filtration using a Superose 12 FPLC column. The column was equilibrated with buffer containing 25 m*M* Tris–HCl, pH 7.5, 150 m*M* NaCl and 1 m*M* 2-mercaptoethanol and elution was per-

formed at a flow-rate of 0.2 ml/min after injection of 1 mg of Ni^{21} –NTA purified Yol066. For calibration, the following molecular mass standards (BioRad) were used: vitamin B12 (1350), horse myoglobin (17 500), chicken ovalbumin (44 000), beef a-globulin (158 000), and thyroglobulin (670 000). The void and total volumes of the column, 6.9 ml and 18.2 ml, respectively, were determined with blue dextran dye and potassium dichromate to enable calculation of the distribution coefficient K_e.

3. Results and discussion

3.1. Optimisation of expression

3.1.1. Strategy of optimisation

Preliminary tests of the Yol066 expression in *E. coli* strain BL21-CodonPlus(DE3) RIL indicated that the recombinant protein is well expressed during IPTG induction, but the majority (up to 60%) of the product is present in an insoluble form in cell extract (data not shown). Thus we first attempted the optimisation of protein production to gain in yield and solubility. The following parameters were chosen for optimisation: media composition (rich versus minimal media), inducer's (IPTG) concentration, temperature and expression time. Besides the choice of the expression system itself, these parameters generally have the most important impact on the yield of soluble protein upon heterologous expression. Earlier observations suggested that medium composition and growth temperature are important factors since they affect the ratio of soluble and insoluble enzyme forms upon expression (see for examples Refs. [31,32]). Most likely the influence of medium composition and temperature is related to modulation of the chaperones activity or some other accessory proteins involved in protein folding. The inducer's concentration and induction time also play a crucial role in protein production since low concentration may not be sufficient for full induction, whereas high concentration of the inducer may affect the growth of the cell, especially in the case of production of toxic proteins [33]. Moreover, excessive induction time may conduct to pronounced degradation of the expressed recombinant protein.

For optimal exploration of parameter's importance we decided to apply the incomplete factorial and response surface design. This approach is used when two or more factors are varied simultaneously and allows reducing very considerably the total number of experiments in comparison to full factorial design. Four parameters which might affect Yol066 yield (media composition, inducer's concentration, temperature and induction time) were chosen and tested at two or three levels using the matrix generated by the interactive software SAmBA [34]. This software helps to set up an optimal experimental design with a minimal number of tests required to explore all possible combinations of parameters. Out of 36 possible combinations 12 representative sets of parameters were tested. Selected parameter sets and results of total and soluble protein expression are presented in Table 1. The total and soluble protein expression was estimated by quantification of protein SDS gel electrophoresis pattern of whole *E. coli* cells and soluble S10 extract (Fig. 1, see also Experimental section). The yield of soluble protein was calculated in arbitrary units taking into account the cell culture density after expression and amount of Yol066 in S10 supernatant.

3.1.2. Global expression vs. solubility

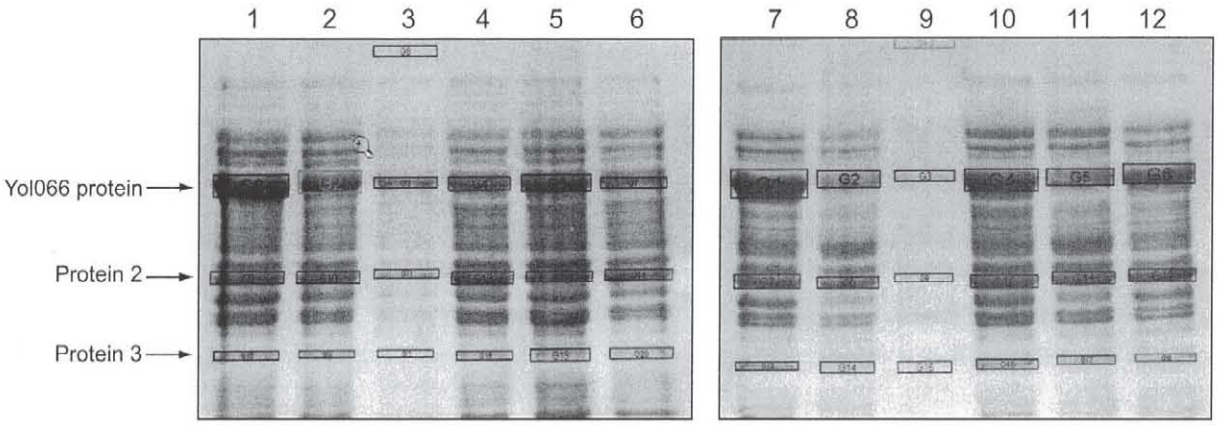
The results presented in Fig. 1 and Table 1, show that the intracellular concentration of soluble Yol066 depends mainly on induction temperature and IPTG concentration. Slowing down the expression rate (low temperature and IPTG concentration, 18 8C and 50 m*M* IPTG, respectively) favours the accumulation of the soluble product, while the use of high IPTG concentration (1 m*M*) and high temperatures (37 8C) diminishes the yield considerably. This is most probably related to saturation of endogenous *E. coli* chaperones during the massive expression of the foreign polypeptide. On the other hand, the limited solubility of the expressed product may be in part compensated by the accumulation of cell biomass. Taken together, these two factors determine optimal expression conditions. As shown in Table 1, the maximal expression of soluble Yol066 is achieved during long (15 h) induction at 18 8C, however, optimal concentration of IPTG varied in function of media used. The induction in minimal medium required relatively high IPTG concentration (1000

8

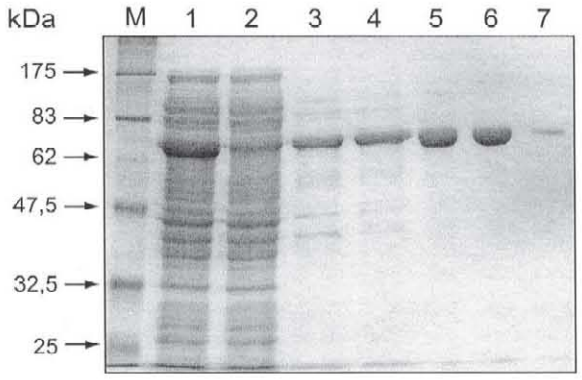
m

m m

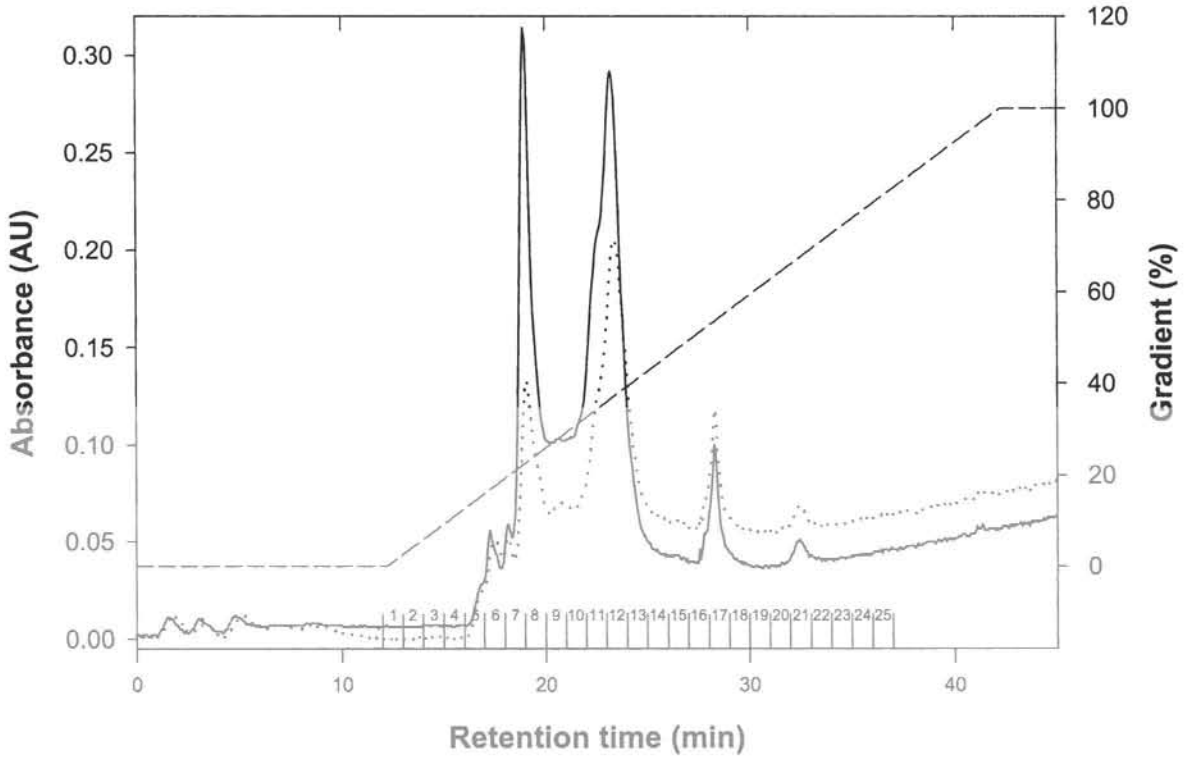
m 8


1
2
3
4
5
6
7
8
9
10
11
12
Yol066 protein
Protein 2
Protein 3
G2
G3
G5

| m


kDa
M
1
2
3
4
5
6
7
175
83
62
47,5
32,5
25

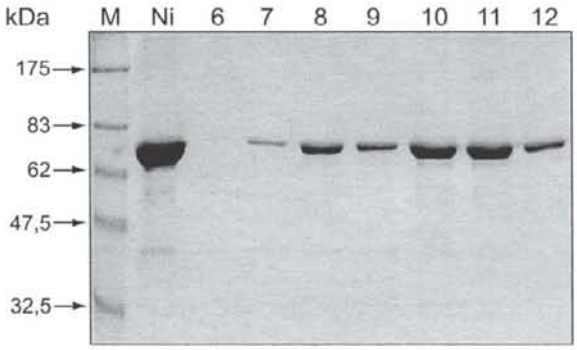
A


Absorbance (AU)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
120
100
80
60
40
20
0
Gradient (%)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
0
10
20
30
40
Retention time (min)

1

1

B


kDa
M
Ni
6
7
8
9
10
11
12
175
83
62
47,5
32,5

C

C

does not aggregate in solution even in the absence of RNA.

4. Nomenclature

C	Pseudouridine
PCR	Polymerase chain reaction
IPTG	Isopropyl-b-D-thiogalactopyranoside
PAAG	Polyacrylamide gel
PAGE	Polyacrylamide gel electrophoresis

Acknowledgements

This work was supported by laboratory funds from the CNRS and ''Région Lorraine''. I. Ansmant is a pre-doctoral fellow supported by a fellowship from the ''Ministère de la Recherche et de l'Enseignement Supérieur''. We thank C. Branlant (MAEM, Nancy) for fruitful discussions and comments on the manuscript and S. Azza (MAEM, Nancy) for help in conducting chromatographic experiments.

References

[1] E.M. Logvinenko, G.M. Shavlovskii, A.E. Zakal'skii, E.Z. Seniuta, Biokhimiia 45 (1980) 1284.
[2] G.M. Shavlovskii, A.A. Sibirnyi, B.V. Kshanovskaia, L.V. Koltun, E.M. Logvinenko, Genetika 15 (1979) 1561.
[3] G.M. Shavlovskii, A.A. Sibirnyi, D.V. Fedorovich, E.Z. Seniuta, Mikrobiologiia 51 (1982) 96.
[4] G.M. Shavlovskii, E.M. Logvinenko, D. Schlee, L.V. Koltun, Biochim. Biophys. Acta 428 (1976) 611.
[5] M. Tzermia, C. Katsoulou, D. Alexandraki, Yeast 13 (1997) 583.
[6] G. Richter, M. Fischer, C. Krieger, S. Eberhardt, H. Luttgen, I. Gerstenschlager, A. Bacher, J. Bacteriol. 179 (1997) 2022.
[7] E.V. Koonin, Nucleic Acids Res. 24 (1996) 2411.
[8] I. Ansmant, Y. Motorin, S. Massenet, H. Grosjean, C. Branlant, J. Biol. Chem. 276 (2001) 34934.
[9] I. Ansmant, Y. Motorin, Mol. Biol. (Moscow) 35 (2001) 206.
[10] J. Conrad, D. Sun, N. Englund, J. Ofengand, J. Biol. Chem. 273 (1998) 18562.
[11] K. Nurse, J. Wrzesinski, A. Bakin, B.G. Lane, J. Ofengand, RNA 1 (1995) 102.
[12] S. Raychaudhuri, L. Niu, J. Conrad, B.G. Lane, J. Ofengand, J. Biol. Chem. 274 (1999) 18880.
[13] J. Wrzesinski, K. Nurse, A. Bakin, B.G. Lane, J. Ofengand, RNA 1 (1995) 437.
[14] J. Wrzesinski, A. Bakin, K. Nurse, B.G. Lane, J. Ofengand, Biochemistry 34 (1995) 8904.
[15] V. Arluison, C. Hountondji, B. Robert, H. Grosjean, Biochemistry 37 (1998) 7268.
[16] Y. Motorin, G. Keith, C. Simon, D. Foiret, G. Simos, E. Hurt, H. Grosjean, RNA 4 (1998) 856.
[17] J. Chen, J.R. Patton, RNA 5 (1999) 409.
[18] H.F. Becker, Y. Motorin, R.J. Planta, H. Grosjean, Nucleic Acids Res. 25 (1997) 4493.
[19] I. Ansmant, S. Massenet, H. Grosjean, Y. Motorin, C. Branlant, Nucleic Acids Res. 28 (2000) 1941.
[20] S.C. Makrides, Microbiol. Rev. 60 (1996) 512.
[21] M.J. Weickert, D.H. Doherty, E.A. Best, P.O. Olins, Curr. Opin. Biotechnol. 7 (1996) 494.
[22] Novagen, pET System manual, 2000.
[23] T. Ikemura, J. Mol. Biol. 146 (1981) 1.
[24] H. Dong, L. Nilsson, C.G. Kurland, J. Mol. Biol. 260 (1996) 649.
[25] K. Zahn, J. Bacteriol. 178 (1996) 2926.
[26] R. Novy, D. Drott, K. Yaeger, R. Mierendorf, InNovations 1 (2001).
[27] P.M. Schenk, S. Baumann, R. Mattes, H.H. Steinbiss, Biotechniques 19 (1995) 196.
[28] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989.
[29] U.K. Laemmli, Nature 227 (1970) 680.
[30] M.M. Bradford, Anal. Biochem. 72 (1976) 248.
[31] M. Piatak, J.A. Lane, W. Laird, M.J. Bjorn, A. Wang, M. Williams, J. Biol. Chem. 263 (1988) 4837.
[32] S. Cabilly, Gene 85 (1989) 553.
[33] J.A. Vasina, F. Baneyx, Protein Expr. Purif. 9 (1997) 211.
[34] S. Audic, F. Lopez, J.M. Claverie, O. Poirot, C. Abergel, Proteins 29 (1997) 252.
[35] R. Schwartz, C.S. Ting, J. King, Genome Res. 11 (2001) 703.

1.3.2. Culture en fermenteur

Après avoir constaté que nos protéines d'intérêt se produisaient sous forme soluble et présentaient une bonne qualité, nous avons chercher à augmenter les rendements de production par utilisation d'un fermenteur BIOSTAT® Bplus (Sartorius). Pour ce faire, nous avons réalisé des cultures en milieu LB et ZYM5052 dans des conditions identiques à celle utilisée en batch mais en réalisant un contrôle très précis des paramètres affectant la productivité. Les conditions expérimentales que nous avons utilisé sont:

- Volume de milieu : 5L.
- Vitesse de rotation de la turbine d'agitation : 200 rpm.
- pH du milieu : 7,00 régulé par des pompes d'acide et de base.
- Température de croissance : 37°C puis 20°C pendant l'induction.
- Concentration d'oxygène dissous : 50% minimum régulé par la vitesse de rotation de la turbine.

1.3.3. Production de protéines sélénométhionylées (Se-Met)

La résolution d'un nombre croissant de structure cristallographique de protéines est réalisée par la méthode Multiwavelengh Anomalous Dispersion (MAD) grâce au rayonnement synchrotron. La recherche de dérivés lourds, empirique, fastidieuse et aléatoire, est avantageusement remplacée par l'incorporation de sélénométhionine aux positions occupées par des méthionines dans la protéine. Cet échange par un acide aminé non naturel est souvent, mais pas toujours, sans conséquences sur les propriétés des protéines (solubilité, fonction). Le marquage est réalisé par croissance d'un mutant auxotrophe de *E. coli* **B834 (DE3) RIL** (F^-, *omp*T, *hsd*S($r_{B-}m_{B-}$), *dcm*+, Tet^r, *gal* λ(DE3), *end*A, Hte[*arg*U *ile*Y *leu*W Cam^r] *met* (Novagen) exprimant la protéine d'intérêt dans un milieu contenant de la sélénométhionine à la place de la méthionine suivant le protocole décrit ci-dessous. Le taux de production de la protéine est très souvent réduit d'un facteur 2 à 4.

Protocole de marquage :

1. Préculture n°1

En fin de journée (17h) : Ensemencer 100 mL (erlen 500L) en milieu LB à partir d'une colonnie fraîche sur boîte LB+Antibiotique 100µg/ml (KC pour pET28b dans souche RIL ou à partir du stock Congel-Cell à -80°C).

Laisser agiter O/N à 37°C.

2. Préculture n°2

En fin de journée (17h) : Ensemencer 500 mL (erlen 2L) en milieu LB+Antibiotique à partir de la préculture 1.

Laisser agiter O/N à 37°C.

3. Lavage de la préculture
Prendre la DO de la préculture (normalement à saturation donc DO >= 5).
Centrifuger la préculture à 4000 rpm (2 falcon 50mL stérile) pendant 15 min.
Eliminer le surnageant et laver les culots avec de l'eau mQ stérile (bien ressuspendre au vortex) sous la flamme.
Centrifuger une seconde fois la préculture à 4000 rpm (2 falcon 50mL stérile) pendant 15 min.
Eliminer le surnageant et reprendre le culot dans 40 mL d'eau mQ stérile (2x20mL) sous la flamme.

4. Culture SE-Met → Phase 1 à 37°C (3-5h)
La veille, préparer 4L de milieu minimum Se-MetMM reparti en 8x500mL (erlen de 2L).
Le jour même, ajouter 100 mL de la solution Glc-Thiamine par litre de milieu.
Préparer la solution contenant la Se-Met en pesant 200mg de L-Se-Met (port du masque) et les dissoudre dans 20 mL d'eau stérile (falcon stérile 50 mL) en vortexant ou en laissant agiter 15 min sur roue à température ambiante.
Distribuer 2,5 mL (25 mg) de la solution de Se-Met dans chaque erlen contenant le milieu Se-MetMM préchauffé à 37°C.
Ajouter les antibiotiques nécessaires (1/1000).
Ensemmencer chaque erlen avec la pré-culture a une DO initiale de 0,3 U/mL.
Laisser agiter à 37°C et faire un suivi de DO jusqu'à obtenir une DO de 0,8 U/mL (ne pas dépasser cette valeur !).

5. Culture SE-Met → Phase 2 à 20°C (16-20h min)
Lorsque la densité optique du milieu a atteint 0,8 U/mL, ajouter 500 µl d'IPTG 1M (1mM final) dans chaque erlen et placer les erlens à 37°C sous agitation (180-200 rpm) pendant 16-20h.
Faire un prélèvement et vérifier la DO finale qui doit être >= à 2 U/mL. Si ce n'est pas le cas, laisser induire jusqu'à obtention de cette DO.

Centrifuger la préculture à 4000 rpm (4 pots à centrifugation de 1L) pendant 15 min.
Eliminer le surnageant.

2. Purification des protéines

La purification s'effectue en 2 étapes de chromatographie successives suivies par le déssallage et la concentration de l'échantillon protéique (Fig X). La première étape consiste en une chromatographie d'affinité réalisée sur une colonne de résine Ni-NTA agarose (Qiagen) à l'aide d'un système AKTA Prime (Amersham Biosciences). La seconde étape consiste en une gelfiltration préparative dite de « polissage » réalisée sur une colonne de type Superdex 26/60 HR en utilisant un dispositif AKTA FPLC (Fast Protein Liquid Chromatography, Amersham

Biosciences). Les différentes fractions de la purification (protéines non fixées : filtrat, lavages, éluat) récupérées sont analysées par SDS-PAGE, afin de vérifier la qualité de la purification puis quantifiées par test colorimètrique. Le protocole utilisé est le suivant :

- Purification sur colonne Ni-NTA Agarose en système semi-batch :

1. Prélever 15 ml de résine et les introduire dans l'enceinte de chromatographie Pharmacia 26/30.
2. Packer la colonne puis rincer la résine avec 15 volumes colonne
3. Equilibrer la résine avec 10 vol de tampon de casse (20 mM Tris-HCl pH7,5, 1M NaCl).

Rq : Pour la purification des protéines sélénométhionylées, il faut ajouter 1mM DTT à toutes les solutions.

4. Dépacker la colonne et transférer la résine dans un falcon de 50 mL
5. Ajouter le surnageant de sonication contenant les protéines et mélanger 10s au vortex.
6. Laisser sous agitation pendant 30 min sur roue motorisée.
7. Introduire le mélange dans l'enceinte de chromatographie et repacker la colonne en récupérant le surnageant de sonication contenant les protéines non fixées.
8. Procéder à la purification de la protéine en utilisant un débit constant de 5ml/min en respectant les étapes suivantes :

- Réinjecter l'échantillon contenant les protéines non fixées par la pompe externe. Cette étape est réalisée afin de s'assurer de la fixation de l'intégralité des protéines sur la colonne.
- Déclencher la collecte de fraction de 5 ml et suivre la densité optique à 280 nm.
- Laver la résine en utilisant le tampon de lavage (20 mM Tris-HCl pH7,5, 1M NaCl, 50 mM Imidazole) jusqu'à ce que la DO280nm soit revenue à la ligne de base assurant le bon lavage de la résine.
- Eluer les protéines d'intérêt en utilisant le tampon d'élution (20 mM Tris-HCl pH7,5, 1M NaCl, 50 mM Imidazole).

9. Laver la colonne avec 20 vol d'eau mQ, dépacker la résine et stocker à 4°C dans un falcon de 50 ml.

- Gel filtration préparative sur colonne Superdex HR26/60 :

1. Equilibrer de la colonne avec 1,5 vol de tampon GF (10 mM Tris-HCl pH7,5, 500mM NaCl).
2. Injecter la protéine par la boucle de 5 ml.
3. Déclencher la collecte de fractions de 5 ml.
4. Réaliser l'élution de la protéine en conditions isocratiques à un débit de 2 ml/min.

Rq : On constate parfois que la protéine est retenue non spécifiquement sur la résine (la protéine est co-éluée avec les sels) ce qui entraîne un retard d'élution. En conséquence, l'imidazole présent dans la solution n'est pas éliminé mais on peut tout de même considérer que le polissage de l'échantillon est bon car on élimine tous les contaminants protéiques dont le PM est supérieur à celui de notre protéine d'intérêt.

5. Laver la colonne avec 1,5 vol d'eau mQ.

3. Concentration et déssalage des protéines

Ces étapes sont réalisées dans des cellules de concentration Vivaspin YM-10/30 d'un volume de 15 mL (Vivascience) dont la taille d'exclusion est de 10 ou 30 kDA en fonction de la taille de la protéine à concentrer. Au préalable, le dispositif est lavé par passage de 3x15 mL d'eau mQ, puis équilibré par passage de 3x15 mL de tampon GF. L'échantillon est alors concentré dans un volume de 1 ml par centrifugation à 3000 rpm puis transféré dans un falcon de 50 ml. On réalise alors une dilution par ajout de 49 ml de tampon de déssalage (10 mM Tris-HCl pH7,5) puis le mélange est homogénéisé au vortex pendant 30s. L'échantillon est alors concentré une nouvelle fois jusqu'à l'obtention d'une solution protéique à la concentration souhaitée.

4. Caractérisation des protéines

4.1. Dosage des protéines par la méthode de Bradford

La méthode de Bradford est un dosage colorimétrique, basé sur le changement de longueur d'onde d'absorbance, se manifestant par le changement de la couleur du bleu de Coomassie après liaison (complexation) avec les acides aminées aromatiques (tryptophane, tyrosine et phénylalanine) et les résidus hydrophobes des acides aminés présent dans la ou les protéines.
La forme anionique (liée) du colorant est bleue, et possède un spectre d'absorption maximal estimé historiquement à 595 nm. Les formes cationiques (libres) du colorant sont rouges et vertes, absorbant à 465-470 nm. Le changement d'absorbance est proportionel à la quantité de colorant lié, indiquant donc la concentration en protéines dans l'échantillon.
Toutes les mesures ont été réalisées avec le kit Bradford Dye Assay (Biorad) selon le protocole du fabricant en réalisant régulièrement une courbe d'étalonnage par la BSA.

4.2. Spectrométrie de masse

4.2.1. Principe

Les échantillons protéiques purifiés sont soumis à une étape de caractérisation par spectrométrie de masse par ionisation électrospray (ESI : Electrospray Ionisation) réalisée au Laboratoire de Spectrométrie de Masse Bio-Organique (LSMBO) à l'Université Louis Pasteur (Strasbourg) par Jean-Michel SALIOU. Cette méthode permet d'analyser des particules ionisées dans des champs électriques et d'obtenir leur poids moléculaire à partir de leur rapport masse/charge. L'échantillon à analyser doit être dans un premier temps ionisé au moyen de la source d'ionisation, dans notre cas ESI. L'échantillon est dissout dans un mélange aqueux (eau/acétonitrile- acide volatile) puis introduit par infusion dans un capillaire porté à un fort potentiel. Le champ électrique créé entre ce capillaire et une contre-électrode dans un flux d'azote chaud permet la formation d'un fin brouillard de gouttelettes chargées. Ces gouttelettes soumises à une atmosphère d'azote chaud vont, par évaporation voir leur taille diminuer et provoquer ainsi leur éclatement par «explosions coulombiennes». Ce phénomène se répète plusieurs fois jusqu'à obtention de l'analyte chargé en phase gazeuse. Les ions formés sont

ensuite dirigés *via* l'interface au moyen de gradients de tension et de pression vers l'analyseur de masse où règne un vide poussé et qui permet la séparation des ions en fonction de leur rapport masse/charge (*m/z*). Le spectromètre de masse utilisé est de géométrie Q-TOF (Quadrupole - Time Of Flight), le Q-TOF Micro (Waters, Manchester, UK). Au cours de nos expériences l'analyseur de type TOF permet de séparer les ions en fonction de leur vitesse, elle-même dépendante de leur masse à énergie cinétique constante pour tous les ions. Les analyseurs TOF mesurent le temps que mettent les ions pour franchir une distance fixe.

4.2.2. Contrôle qualité par spectrométrie de masse ESI dénaturante

Afin de vérifier si notre protéine purifiée possède bien le poids moléculaire attendu, les échantillons sont concentrés en utilisant des cellules de concentration YM-10 (Vivaspin) jusqu'à une concentration d'au moins 30 µM. L'analyse par spectrométrie de masse électrospray peut être réalisée après dessalage par chromatographie liquide en phase inverse sur une colonne C4 ou C8 (Pharmacia). L'échantillon est injecté sur la cartouche, il est dessalé par une solution H2O, 0,1% acide tri-fluoroacétique, puis la protéine est éluée par une solution à 60% CH3CN/H2O (9/1 volume/volume), 0,1% acide tri-fluoroacétique. Dans les deux cas, l'échantillon est injecté à un débit de 5 µL/min. Les paramètres de la source ESI du spectromètre de masse sont : une tension de capillaire de 3 kV et les voltages du cône d'échantillonnage et du cône d'extraction sont respectivement de 55 V et 2 V. Les spectres de masse sont enregistrés dans une gamme de 500-2100 en rapport *m/z*.
Les différentes données sont enregistrées et traitées par le logiciel Masslynx (Waters).

5. Cristallographie R-X

5.1. Introduction à la cristallographie des rayons X

L'observation de l'arrangement des atomes constituant les macromolécules que nous cherchons à étudier, implique l'utilisation d'ondes électromagnétiques dont la longueur d'onde est de l'ordre de grandeur de l'objet à observer. Dans le cas de la cristallographie aux rayons X, particulièrement adaptée à la visualisation de ces objets, on utilise le rayonnement X dont la longueur d'onde est de l'ordre de l'Angström (10^{-10} m). Lorsqu'une molécule est soumise au rayonnement X, ses électrons diffusent des rayons X dans toutes les directions. Le signal produit par une molécule isolée est trop faible pour être mesuré. Par contre, dans un cristal, l'arrangement tridimensionnel périodique des molécules induit des interférences entre les rayonnements diffusés par chacune d'elles. Dans certaines directions bien précises (dictées par la géométrie du cristal), il y a amplification du signal par interférences constructives : c'est le phénomène de diffraction. Une expérience de diffraction consiste à faire tourner un cristal dans le faisceau incident de manière à amener tous les plans qui le composent en position de diffraction et à mesurer l'intensité de chacune des réflexions. A partir de leur intensité et après

reconstitution de leur phase, le contenu de la maille élémentaire constituant le cristal va pouvoir être calculé sous forme d'une carte de densité électronique. La structure de la macromolécule constituant le cristal est alors reconstruite à l'intérieur de cette carte.

5.2. Cristallogenèse

5.2.1. Principe

La solubilité d'une macromolécule est fonction de nombreux paramètres tels que la concentration de la macromolécule, le pH, la température, la force ionique, l'effet d'additifs etc. La cristallisation est un compromis entre les facteurs thermodynamiques (solubilité) et cinétiques (nucléation et croissance cristalline).

La méthode la plus souvent utilisée pour la cristallisation des protéines est la diffusion de vapeur en goutte suspendue (McPherson, 1990). Au sein d'un système clos, une goutte constituée d'un mélange de l'échantillon de macromolécule et d'une solution de cristallisation, est mise en présence d'un réservoir constitué de cette même solution de cristallisation (Figure 63A). Initialement, la concentration en agents cristallisant dans le réservoir est donc supérieure à celle dans la goutte de cristallisation. L'équilibre s'établit dans le temps entre la goutte et le réservoir par diffusion des espèces volatiles jusqu'à ce que la tension de vapeur soit la même dans la goutte et dans le réservoir. La concentration de la protéine et de l'agent précipitant (sels non chaotropiques, solvants organiques ou polymères) dans la goutte augmente donc progressivement. Ainsi, la solubilité de la macromolécule diminue jusqu'à atteindre l'état de sursaturation favorable à l'obtention de cristaux, sans toutefois tomber dans la zone de précipitation (Figure 63B).

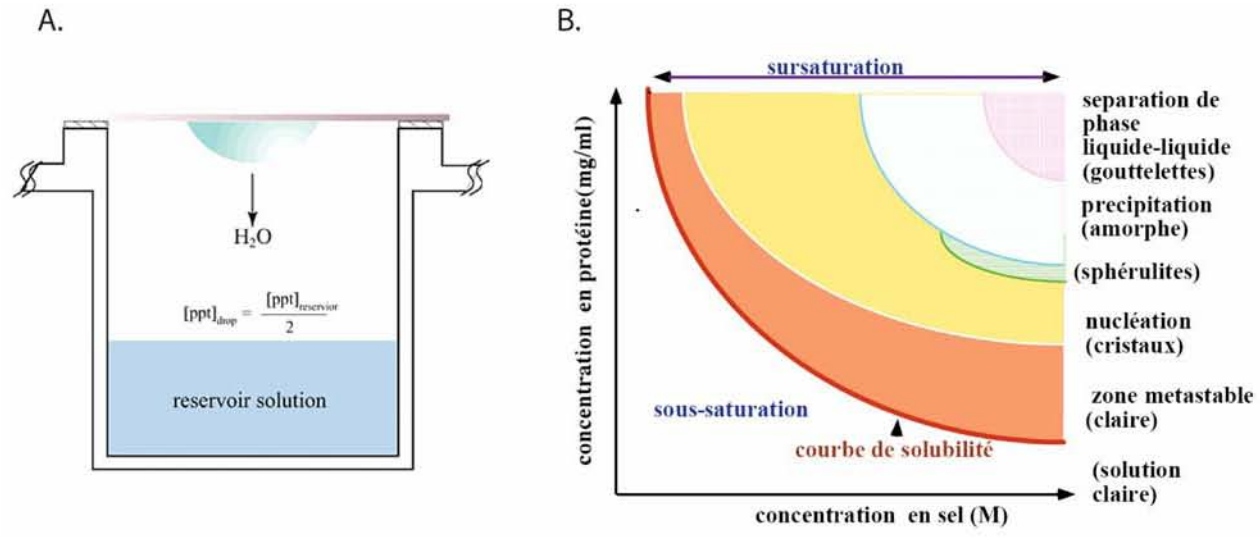


Figure 63 : Diagramme du processus de cristallogenèse.

5.2.2. Recherche des conditions de nucléation

La recherche des conditions initiales de cristallisation s'est effectuée en utilisant les kits Crystal Screen I & II, PEG/Ion Screen, SaltRX (Hampton Research), Strucure Screen I & II, Clear Strategy Screen I & II (Molecular Dimensions Ltd), Wizard I & II (Emerald Biostructure) et JBScreen 1-10 (Jena Biosciences). Ainsi, environ 1000 conditions de cristallisation différentes peuvent être testées lors d'une première étape.

Les essais ont été effectués en boîtes de 96 puits comprenant trois gouttes par puits en utilisant un robot de pipetage Mosquito (TTPLabtech). Pour chaque condition de cristallisation, 100 ul de solution d'agent précipitant est placé dans le réservoir tandis que les trois gouttes sont consituées de 100 nl de protéine et 100nl de solution de cristallisation, 200 nl de protéine et 100 nl de solution, 200 nl de protéine et 100 nl de solution de cristallisation.

5.2.3. Optimisation des conditions initiales pour cristalliser des monocristaux pour la collecte de données de diffraction des R-X

Lorsque des conditions menant à une forme solide (cristaux, aiguilles, plaquettes, sphérulite, précipité microcristallin) sont décelées, ces conditions de cristallisation sont ensuite optimisées pour l'obtention de monocristaux de taille et de qualité de diffraction requises pour l'enregistrement radiocristallographique en faisant varier les différents paramètres initiaux de concentration en protéines, agents cristallisants, température, pH, volume de goutte.

5.3. Collecte de données de diffraction

Les cristaux ainsi formés sont soumis à un rayonnement X provenant soit d'un générateur de laboratoire, soit d'un synchrotron. Le rayonnement provoque la formation de radicaux libres à l'intérieur du cristal pouvant endommager les molécules. On effectue donc une étape préliminaire de congélation de ces cristaux pour éviter les dommages radiatifs. Pour ce faire, on utilise des solutions dîtes cryoprotectrices qui empêchent l'eau de cristalliser et d'entraîner une forte dégradation du cristal lors de l'expérience de radiation.

Des mesures de diffraction des rayons X sont réalisées au laboratoire en utilisant un diffractomètre X8-proteum (BrukerAXS) comportant notamment un générateur de rayons X microfocus à anode tournante (longueur d'onde du faisceau incident de 1,54 Å) et un détecteur de type CCD de 135 mm. Les données obtenues sont traitées avec le programme Proteum^{plus} (BrukerAXS).

Des mesures de données de diffraction X sont aussi réalisées au synchrotron de Grenoble, afin d'obtenir des mesures à hautes résolutions et aussi pour les mesures sur des cristaux de protéines comportant des sélénométhionines à la place des méthionines, dans le cadre de collecte de type MAD (diffusion anomale multiple) nécessitant de modifier la longueur d'onde incidente.

4. Transcriptomique (Splicing Array)

A. Matériel

1. Définition et type de puces utilisées

Une puce à ADN, aujourd'hui communément appelée « DNA microarray » en anglais (de « array » = rang ordonné), est basé sur la technique d'hybridation. Immobilisés sur un support solide (matrice), des oligonucléotides (simples ou double brins) spécifiques de différents gènes ou ADNc connus constituent les sondes dont le rôle est de détecter des cibles marquées complémentaires, présentes dans le mélange complexe à analyser (ARNm extraits de cellules, tissus ou organismes entiers et convertis en ADNc). Les sondes sont soit greffées sur le support, soit synthétisées in situ (unité d'hybridation = plot). Les signaux d'hybridation sont détectés selon le type de marquage, radioactivité ou fluorescence, par mesure radiographique ou par fluorescence, et quantifiés.

Lors de nos expériences nous avons utilisés des puces « spottées » qui nous ont été fournies par le laboratoire de J. Beggs au sein duquel nous avons réalisé cette étude. Ces puces appelées « Splicing array » sont conçues pour l'étude de l'épissage chez la levure *S. cerevisiae* puisque tous les gènes contenant des introns sont analysés grâce à l'utilisation de 4 sondes spécifiques (E, I, M et V) qui seront détaillées dans la partie résultats.

2. Milieux de culture, tampons et solutions spécifiques

Pour ces expériences, nous avons utilisé, en plus des consommables et réactifs d'usage courant, les produits listés dans le tableau ci-dessous :

Random primers	Invitrogene	Ref : 48190-011
Superscript III Reverse Transciptase	Invitrogene	Ref : 18080-044
OligodT 25 µg (0,5 µg/µl)	Invitrogene	Ref : 18418-012
Colonne de purification Qiaquick	Qiagen	Ref : 28106
Cy3 MONO 5-PACK	Amersham	Ref : PA53022
Cy5 MONO 5-PACK	Amersham	Ref : PA55022
96% Ethanol for UV/Spectrophotometry	Fluka	
aadUTP	Sigma	Ref : A.0410
Hydroxylamine (NH2OH-HCl)	Sigma	Ref : H.2391

Tableau 13 : Matériel spécifique utilisé lors des expériences de Splicing Array.

B. Méthodes

1. Optimisation des conditions expérimentales

L'expression globale des gènes est très sensible aux variations des conditions expérimentales. Etant donné que l'on considère les variations transcriptionnelles de l'ensemble des gènes d'un organisme, il se peut que les effets observés rendent compte, non pas de la réponse cellulaire étudiée, mais d'artéfacts techniques ou biologiques mal maîtrisés. La reproductibilité des résultats peut en être aussi grandement affectée et il est particulièrement difficile dans ces circonstances de comparer des données d'expériences qui n'ont pas été réalisées au même moment dans les mêmes conditions, et *a fortiori* des données issues de deux laboratoires différents. Lors de notre analyse, nous avons pris en compte différents paramètres :

- Les fonds génétiques des souches que nous avons comparé présentent de grandes similarités aussi bien génotypiques que phénotypiques. Les souches mutantes analysées sont isogéniques à la souche sauvage de contrôle et leur différence ne porte que sur la mutation étudiée. De plus, les marqueurs d'auxotrophies ont été homogénéisé « au mieux » entre les souches à comparer, ainsi que le type sexuel s'il s'agit de souches haploïdes.
- Les conditions de culture ont été contrôlées afin d'obtenir des populations homogènes de cellules. L'expression génique peut être affectée par le type de milieu utilisé (même par des différences entre des lots d'un même milieu), le matériel, les conditions de température et d'aération, etc. Il convient aussi d'éviter les situations conduisant à une altération incontrôlée du niveau d'expression, telles que le passage en phase stationnaire induisant une transition à l'état diauxique, en condition de létalité cellulaire, ou encore en condition de stress métabolique lié à des changements de milieu. Les cultures comparées doivent être aussi dans le même état de croissance et si possible synchronisées au niveau du cycle cellulaire.
- Nous avons réduit les stress métaboliques et autres effets indésirables (choc thermique, osmotique…) lors de la préparation des échantillons en prélevant les cellules le plus rapidement possible, sans étapes de lavage superflues, et en congelant immédiatement les cellules dans de l'azote liquide afin de limiter la dégradation des ARNm à annalyser.

2. Condition de culture des différentes souches de levure

Afin d'augmenter la reproductibilité de nos résultats, nous avons réalisé un suivi de croissance très précis de façon à arrêter toutes les cultures à la même valeur d'absorbance finale et donc avec une biomasse très similaire. En effet, il s'avère que les variations d'épissage entre 2 souches peuvent être très faible et donc sensibles aux changements des conditions expérimentales. Le protocole de culture utilisé est le suivant :

La culture est réalisée dans un erlen de 250 ml contenant 25 ml de milieu YPD préchauffée à 30°C et ensemencés à une DO_{600nm} de 0,3 à partir d'une préculture de la souche de *S. cerevisiae* étudiée. La culture est réalisée à 30°C sous agitation à 200 rpm et la croissance est suivie par

spectrophotométrie à 600 nm. La culture est arrêtée en phase exponentielle de croissance à une DO_{600nm} de 0,8 ± 0.05. Si la culture dépasse cette densité, il est alors nécessaire de relancer une culture de 25 ml dans les mêmes conditions en utilisant la moitié de la culture en question complétée par du milieu préchauffé à 30°C. Les cellules sont transférées dans un falcon de 50 mL contenant 25 mL d'éthanol 96% préalablement refroidi à -80°C puis l'ensemble est mélangé par inversion. Les cellules sont alors centrifugées 3 min à 3000 rpm à 4°C et le culot est repris dans 1 ml de milieu YPD puis transférer dans un tube ependorf. Finalement, les cellules sont culotées par centrifugation rapide (Short-Spin) à 13000rpm puis stockées à -70°C ou extraites immédiatement.

3. Préparation des cibles et hybridation des puces

3.1. Extraction des ARN totaux

L'extraction des ARN a été réalisée dans des conditions de particulière de stérilité dites « RNases free » nécessitant le port de gant, le maintient des échantillons au froid et l'utilisation de solutions aqueuses préparées avec de l'eau stérile préalablement traitée par le DEPC ($1/1000_e$) qui inhibe les RNases. Enfin nous utilisons, à la place de l'éthanol absolu, de l'éthanol à 95% réservé à la spectrométrie UV, qui ne contient pas de résidu fluorescent. Nous avons remarqué qu'une mauvaise purification peut être la cause d'une synthèse inefficace des cibles marquées et d'une augmentation des bruits de fond sur la lame et que la dégradation des ARN pouvait induire l'obtention de données biaisées en altérant la proportion des espèces présentes initialement dans l'échantillon étudié.
Nous avons réalisé des extraits d'ARN totaux issues des différentes souches de levure selon un protocole adapté d'une méthode classique d'extraction en utilisant un mélange phénolchloroforme/thiocyanate de guanidinium acide à chaud. Chez la levure, il est important de noter que les parois cellulaires sont difficiles à lyser et doivent être broyées par une forte agitation au vortex en présence de billes de zirconium.

1. Dans un ependorf contenant une quantité de cellules équivalent à 10 DO600nm ajouter rapidement :
100µl d'un mélange GTC/phénol 50 :50
100µl de billes de Zirconium
Agiter 5 min sur un vortex à tubes multiples à 4°C.

2. Centrifuger 1s à 13000 rpm.
Ajouter 700 µl d'un mélange GTC/phénol 50 :50
Vortexer 1 min et incuber 5min à 65°C

3. Centrifuger 1s à 13000 rpm.
Ajouter 350 µl de chloroforme

Ajouter 120 µl de Acétate de Sodium 3M à pH 5,2
Vortexer 20s et centrifuger 5 min à température ambiante

4. Transférer 550µl de la phase supérieure dans un nouveau tube ependorf
Ajouter 500 µl d'un mélange phénol/chloroforme 50 :50
Vortexer 20s et centrifuger 5 min à température ambiante

5. Transférer 550µl de la phase supérieure dans un nouveau tube ependorf
Ajouter 500 µl de chloroforme
Vortexer 20s et centrifuger 5 min à température ambiante

6. Transférer 450µl de la phase supérieure dans un nouveau tube ependorf
Ajouter 1 ml d'éthanol 100%
Incuber 10 min dans la glace
Centrifuger 10 min à 20000rpm à 4°C.

7. Laver le culot à l'éthanol 70% puis ressuspendre dans 50µl

3.2. Mesure de la concentration et contrôle de la qualité des ARN

L'analyse de la qualité des ARN totaux extraits a été réalisée par un spectrophotomètre NanoDrop® ND-1000. Cet appareil permet la réalisation d'un spectre de fluorescence UV entre 220 et 350 nm en utilisant une quantité très faible d'échantillon (1µl). L'analyse du résultat permet d'apprécier la contamination de l'échantillon due à des problèmes d'extraction (contamination par des protéines, des sels, de l'ADN ou du phénol) ainsi que la concentration de la préparation ARN (Figure 64).

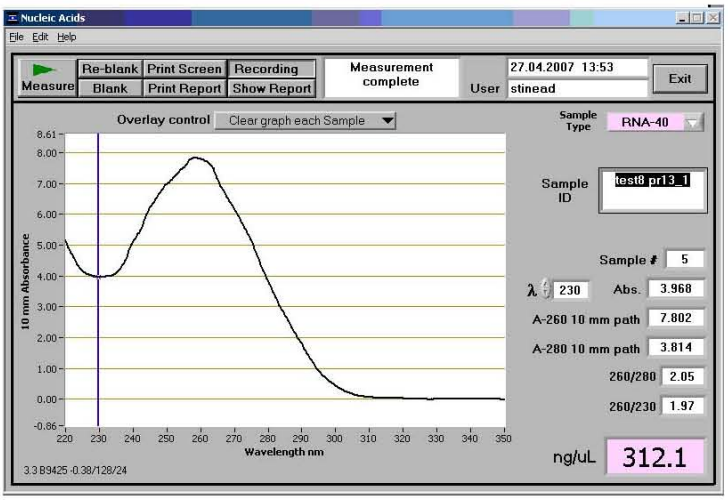


Figure 64 : Analyse d'un échantillon ARN par le spectrophotomètre NanoDrop® ND-1000.

Les rendements attendus avec les protocoles précédents sont de 500-1000 µg d'ARN totaux à partir de 1.10^8 cellules de levure en phase exponentielle. On estime que le rapport A_{260}/A_{280} doit être supérieur à 1,8 si l'extrait d'ARN est correctement purifié.
Nous vérifions également le profil des ARN totaux visualisé après électrophorèse sur un gel d'agarose et coloration au bromure d'éthydium (Figure 65). Chez la levure, nous observons deux bandes d'intensité égale à 3,8 kb (ARNr 28S) et 1,7 kb (ARNr 18S), ainsi que plusieurs bandes de faible taille autour de 0,1 kb (ARNr 5S, ARNt…). Les ARNm ne représentent qu'environ 1% de l'ensemble et forment une trace étalée (« smear ») autour de 1 kb. Une dégradation partielle se traduit par l'apparition de fragments de faible taille moléculaire. Dans ce cas, nous recommençons l'extraction car une telle dégradation visible à l'oeil conduit à l'obtention de données biaisées.

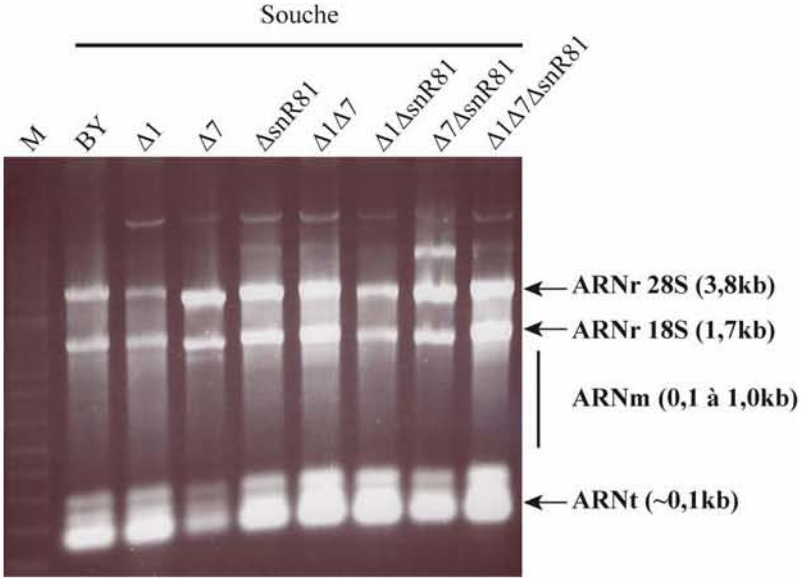


Figure 65 : Vérification de la qualité des ARN totaux par électrophorèse en gel d'agarose 1%.

3.3. Transcription inverse et incorporation des marqueurs fluorescents

Les cibles sont synthétisées par une transcription inverse des ARN totaux en présence d'amorces oligo(dT)$_{15}$ et d'amorces aléatoires (hexamers Pd(N)$_6$) afin d'obtenir des brins d'ADNc marqués en effectuant une incorporation « indirecte » des colorants fluorescents. Dans cette méthode, des nucléotides portant un groupement amine réactif, l'amino-allyl dUTP (aa-dUTP), sont incorporés lors de la transcription inverse. Ceux-ci sont couplés dans un deuxième temps à des N-hydroxysuccinimidyl esters (NHSesters) de Cy3 ou de Cy5, selon la réaction décrite par Randolph et Waggoner (Randolph & Waggoner, 1997). L'efficacité d'incorporation des aa-dUTP est nettement supérieure à celle des nucléotides directement couplés aux colorants, et la méthode permet d'éviter les biais d'incorporation dus aux spécificités de chaque fluorochrome.

Les cibles marquées sont ensuite purifiées pour éliminer l'excédent de colorants, qui risque d'augmenter le bruit de fond en utilisant le kit de purification de fragments d'ADN QIAquick PCR purification kit de QIAGEN.
Le protocole que nous utilisons est décrit ci-dessous :

1. Marquage de l'ARN total :
Toute manipulation des ARN doit être réalisé sur la glace en utilisant les tubes RNAse free. Réduire si nécessaire le volume des échantillons d'ARN par séchage au speed-vac.

Mélanger :
2-3 µg ARN total
5 µg de random hexamers
2 µg d'oligo dT
Qsp 23 µl d'eau
Faire une réaction par échantillon d'ARN. Placer les tubes dans l'appareil PCR et démarrer le programme suivant :

70 °C	10 minutes (dénaturation de l'ARN)
4 °C	5 minutes
Pause à 4 °C	ajouter dans chaque tube le mélange de RT et de Cy
23 °C	10 minutes (annealing)
42 °C	2 heures (élongation)
Pause à 4 °C	
37 °C 15 minutes ou 70 °C 10 minutes (hydrolyse de l'ARN)	
Pause à 4 °C	

Pendant que les tubes sont incubés à 70 °C, préparer le mélange suivant :
Note : Ceci est la quantité pour un échantillon. Multiplier chaque composant en fonction du nombre d'échantillons.

5x Superscript III buffer (Invitrogen)	8 µl
0,1 M DTT (Invitrogen)	4 µl
aadNTP's*	2 µl
H20	1 µl
Superscript III (Invitrogen, 200 U/ml)	2 µl
TOTAL	17 µl

*Préparation des aadNTPs : Mélanger 25 µl de dA, dC et dGTP à 100 mM chacun, 10 µl de dTTP et 15 µl d'aadUTP avec 900 µl d'eau Rnase/Dnase free. Concentration finale : 2,5 mM dA, dC et dGTP, 1 mM dTTP et 1,5 mM de aadUTP.
Lorsque les échantillons sont a 4 °C, après l'incubation de 70 °C, ajouter les 17 µl du mélange contenant les aadNTP dans chaque tube et agiter en faisant un aller et retour avec la pipette.

2. Hydrolyse des ARN :
Après la reverse transcription (incubation à 42 °), il est nécessaire d'enlever les traces d'ARN par hydrolyse alcaline.
Ajouter : 15 µl de NaOH à 0,1 N pour dégrader l'ARN.
Incuber les échantillons dans l'appareil à PCR à 70 °C pendant 10 minutes.
Neutraliser par addition de 15 µl de HCL à 0,1 M.

3. Purification de l'échantillon :
Après le traitement à la RNase, il est nécessaire d'enlever les enzymes, les dNTPs et les fluorochromes non incorporés. Une précipitation à l'éthanol améliore la purification.
Préchauffer l'eau à 37 °C.
Transférer si nécessaire l'échantillon dans un tube eppendorf stérile de 1,5 ml pour pouvoir le centrifuger dans une centrifugeuse réfrigérée.
Pour une réaction de 40 µl ajouter :
4 µl d'acétate de sodium 3 M, pH 5.2 (1/10ème de volume).
100 µl d'éthanol à 96 % (2,5 volumes).
Note : Attention, pour l'hydrolyse chimique, on a 70 µl par réaction.
Vortexer et précipiter à –80 °C pendant 30 minutes ou à –20 °C pendant au minimum 12 heures (sur la nuit ou le week-end).
Centrifuger 30 minutes à 13000 rpm.
Enlever l'éthanol en retournant délicatement le tube et en le tapotant doucement sur un kleenex. Le culot ne doit pas sécher complètement.

4. Purification des cibles (colonnes QuiaQuick de Qiagen) :
Ajouter 40 µl d'eau par tube.
4 µl d'Acétate de Sodium 3 M, pH 5.2.
200 ml du tampon PB de Qiagen, puis vortexer brièvement.
Si le culot ne se dissout pas, chauffer à 37 °C 15 minutes et vortexer. Si ce n'est pas suffisant, chauffer à 95 °C pendant 30 secondes et vortexer.
Déposer l'ensemble de ce mélange sur une colonne Qiaquick PCR clean up (Il est possible de regrouper 2 réactions/colonne).
Centrifuger 1 minute à 13500 rcf.
Enlever le filtrat.

Ajouter 600 µl de tampon PE, puis centrifuger 1 minute à 13500 rcf.
Enlever le filtrat.
Centrifuger 2 minutes à 13500 rcf pour sécher le filtre.
Enlever le filtrat.
Ajouter 30 µl d'eau, pH 8.0, préchauffée à 37 °C au centre de la colonne.
Attendre 1 minute.
Centrifuger 1 minute à 13500 rcf. Si la colonne reste colorée, répéter l'étape d'élution.
L'ADNc peut-être conservé indéfiniment à –20 °C.

5. Couplage du NHS-ester Cy3/5 avec l'échantillon :
Ajouter 3 µl de tampon bicarbonate de Na 0,5 M pH 9.0 dans les 30 µl d'échantillon pour une concentration finale de 0,05 M de Bicarbonate. de Na.
Laisser à température ambiante pendant 10 à 15 minutes.
Ajouter la sonde ADNc à une aliquote de fluorochrome NHS-ester Cy3 ou Cy5.
Incuber 1 heure à température ambiante à l'obscurité.

6. Préparation des aliquotes de fluorochromes :
Reprendre un tube de Cy3 ou Cy5 (NHS-ester) dans 32 µl de DMSO (Sigma-D.5879).
Aliquoter 2 µl x 16 tubes et sécher immédiatement au speed-vac (ou sans sécher pour une utilisation immédiate). Stocker à –20 °C.

7. Blocage et purification :
Note : avant de mélanger les échantillons couplés au Cy3 et au Cy5, il faut bloquer les liaisons NHS-ester libres des fluorochromes pour éviter les réactions croisées.

Ajouter 6 µl d'hydroxylamine 4 M.
Incuber 15 minutes à température ambiante dans l'obscurité.
Mélanger les 2 sondes de Cy3 et de Cy5.
Purifier avec une colonne Qiaquick PCR (Qiagen) :
Ajouter 500 µl de tampon PB dans les échantillons.
Laisser 5 minutes à température ambiante.
Transférer sur une colonne Qiaquick.
Centrifuger 1 minute à 13500 rcf.
Laver avec 750 µl de tampon PE.
Centrifuger 1 minute à 13500 rcf.
Centrifuger à vide 2 minutes à 13500 rcf.
Eluer avec 30 µl H2O à 37 °C.
Laisser 2 minutes à température ambiante avant de centrifuger 1 minute à 13500 rcf pour récupérer l'éluat. La sonde est prête à être hybridée.

4. Hybridation et lavage

Les méthodes d'hybridation sont dérivées de méthodes classiquement utilisées dans les études de Southern et Northern blots. La cible d'ADNc sont incubées dans un tampon de force ionique élevée (ex. : SSC) afin de réduire les répulsions électrostatiques et faciliter l'appariement des brins complémentaires. Des détergents (ex. : SDS) et des agents bloquants permettant de réduire l'hybridation non spécifiques (ex. : poly(A), ADN C0t-1…) sont ajoutés pour réduire le bruit de fond. Nous utilisons le protocole suivant :

1. Pré-hybridiation des lames :
Cette étape de pré-hybridation ne doit pas être faite plus de deux heures avant l'hybridation. Et il ne faut pas conserver les lames pré-hybridées plus de 3 heures.
Exposer l'échantillon aux UV en utilisant une énergie de 360J à l'aide du crosslinker (UVC 500 GE healthcare).
Préparer la solution de blocage suivante : SSC 5X, SDS 0,1%, BSA 1%
Préchauffer la solution de blocage à 42°C.
Incuber les puces à ADN dans cette solution à 42°C pendant 45 minutes. Toutes les incubations et les lavages peuvent être fait dans un tube Falcon 50ml.
Rincer la lame cinq fois dans de l'eau distillée (effectuer 10 retournement du tube Falcon pour chaque étape de rinçage).
Rincez la lame dans l'isopropanol (effectuer 10 retournement du tube Falcon pour chaque étape de rinçage) et laissez la sécher verticalement sur un papier Whatman 3M propre.
Ranger la lame dans une boîte à l'abri de la poussière avant hybridation.

2. Hybridation des puces à ADN :
Après la purification Quiaquick : mélanger les 30µl d'ADNc purifiés avec 30µl de tampon d'hybridation 2X (formamide 50%, SSC 10X, SDS 0,2%, préchauffer à 70°C avant utilisation).
Déposer la puce à ADN dans une chambre à hybridation (Camlab) avec 15µl d'eau dans chaque puit.
Préchauffer l'ADNc purifié à 70°C pendant 2-5 minutes (centrifuger pour éliminer les bulles).
Déposer immédiatement la solution d'ADNc sur la puce à ADN (il est préférable de faire plusieurs petits dépôts plutôt qu'un seul gros dépôt au centre) et immédiatement mettre une lamelle de 22X60mm du bas de la lame (gauche) vers le haut (droite). Éviter les bulles autant que possible quand vous déposez le mélange d'ADNc. Fermer hermétiquement la chambre et mettez là dans un bain-marie à 42°C sur la nuit.

3. Lavage des lames :
Rincer les lames avec :
SSC 1X, SDS 0,2% préchauffé à 42°C
SSC 0,1X, SDS 0,2% à température de la pièce

SSC 0,1X à température de la pièce

Procéder de la même façon que pour l'étape de lavage de la pré-hybridation : effectuez 10 retournements du tube Falcon de 50ml pour chaque étape de rinçage. Faites attention à la lamelle durant le premier lavage, une fois qu'elle est tombée dans le liquide, retourner la lame pour éviter que la lamelle ne se retrouve en contact avec les spots de la puce.

Après la dernière étape de lavage, sécher la lame par centrifugation (500g pendant 3 minutes). Mettre un papier absorbant en bas du portoir de lames pour absorber le liquide durant la centrifugation. Procéder rapidement pour empêcher que le tampon ne sèche sur la lame.

5. Acquisition et analyse des données

5.1. Acquisition des données brutes

5.1.1. Lecture sur le scanner

L'acquisition des images est réalisée par lecture des puces sur un scanner arrayWoRx (Applied Precision) installés au Wellcome Trust Centre for Cell Biology d'Edimbourg (Figure 66). Le scanner est muni de deux lasers (excitations à 532 nm et 635 nm) qui permettent l'acquisition simultanée des signaux émis par les fluorochromes Cy3 et Cy5. Il permet de lire une lame complète du génome de la levure en moins de 5 mn avec une résolution de 10 µm/pixel et une sensibilité maximale de 0,1 molécule de fluorochrome/µm^2.

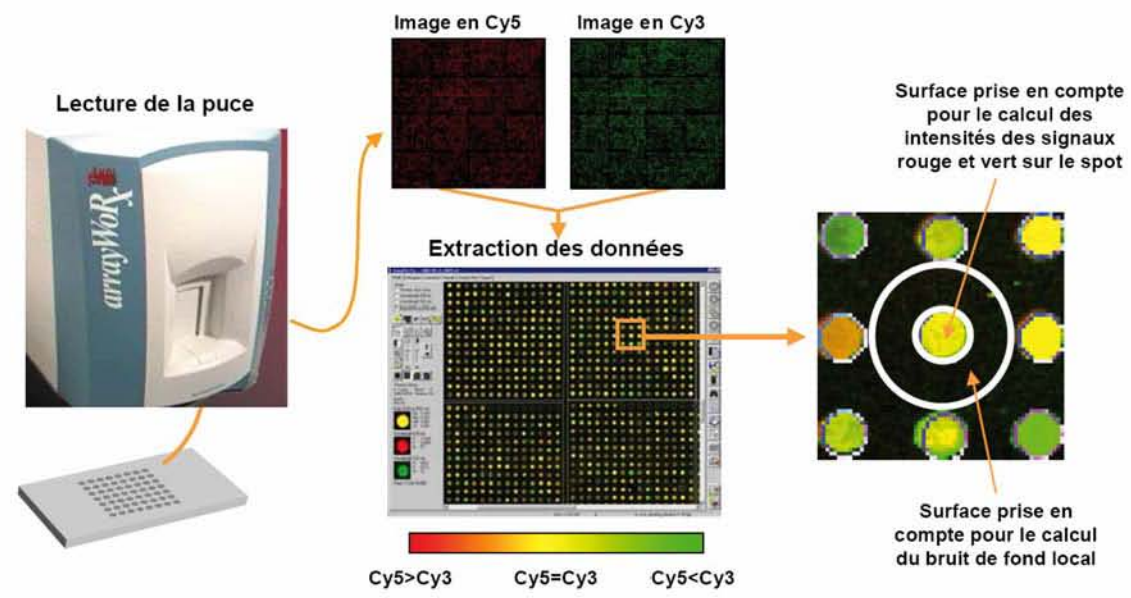


Figure 66 : Acquisition des données sur le scanner arrayWoRx.

Les images obtenues pour chaque canal Cy3 et Cy5 sont enregistrées au format TIFF 16 bits en 65535 niveaux de gris (enregistrement de 2_{16}-1 niveaux maximum par pixel). La gamme d'intensité détectée est donc comprise entre 0 (noir) et 65535 (blanc). La puissance de

l'excitation laser doit être réglée manuellement (sur une échelle de 0 à 1000) de manière à ce que le signal sur les spots soit détectable au-dessus du bruit de fond mais ne sature pas (dans la pratique, nous réglons la puissance de manière à voir apparaître juste 2-3 spots saturés en blanc). Il est admis qu'à l'intérieur de cette gamme, l'intensité du signal augmente de manière linéaire avec la quantité de molécules de fluorochrome sur le spot. Il faut noter que si le marquage est trop faible et requiert un niveau excessivement élevé de la puissance du laser, l'intensité du bruit de fond augmente plus vite que celle du signal sur les spots. Le rapport signal/bruit de fond devrait être supérieur à 1,5-2 pour que le signal soit quantifiable.

5.1.2. Extraction des données numériques

Les images sont analysées grâce au logiciel Array-ProAnalyser (MediaCibernetics) afin d'extraire les données numériques correspondant à chaque spot (Figure 66). Les images sont colorées artificiellement – celle du canal Cy3 en vert et celle de Cy5 en rouge – et superposées pour leur visualisation. Ainsi, un spot de couleur verte indique un gène dont le niveau d'expression est plus élevé dans l'échantillon marqué avec le Cy3 que celui marqué avec le Cy5, et inversement pour un spot de couleur rouge. Le spot apparaît jaune lorsque le gène est exprimé de manière identique dans les deux échantillons comparés.
Le logiciel permet de définir une grille sur l'image afin d'identifier chaque spot en lui assignant des coordonnées uniques (adressage des spots, ou « gridding ») et de délimiter la surface du spot par rapport au reste de la lame (« segmentation » du « foreground » contenant le spot par rapport au « background » définissant le bruit de fond). Le logiciel intègre un algorithme performant de placement automatique des grilles, ce qui accélère considérablement l'analyse Le logiciel génère les données numériques correspondant aux valeurs moyennes et médianes du signal émis par les pixels de chaque spot et du bruit de fond local (délimités par la grille comme illustré dans la Figure 66) et divers autres paramètres (l'écart-type associé aux intensités des pixels d'un spot, le rapport signal/bruit de fond, etc.).

5.2 Transformation des données

5.2.1. Filtrage primaire des données brutes

Le logiciel d'extraction permet de repérer visuellement les spots non exploitables (zones de la lame couvertes de bruit de fond, spot absent…) par un système de balisage (dépôt de drapeaux ou « flags » en anglais) qui assigne un code numérique selon la qualité du spot considéré, afin de faciliter le filtrage des données non significatives. Notamment, nous choisissons en général d'éliminer de la suite de l'analyse les spots dont la somme des intensités médianes rouge et verte est inférieure à 300, et les spots dont le rapport signal/bruit de fond est inférieur à 1,5.

5.2.2. Le calcul des ratios d'expression

Le niveau d'expression relatif de chaque gène est estimé par la valeur du ratio d'expression calculé à partir des intensités des signaux en rouge et en vert : intensité médiane du spot i en rouge – médiane du bruit de fond local en rouge intensité médiane du spot i en vert – médiane du bruit de fond local en vert. Ainsi le ratio est >1 si le gène i est plus exprimé dans l'échantillon marqué en rouge que celui marqué en vert et vice-versa. Il existe plusieurs méthodes de calcul du ratio d'expression (ratio des intensités moyennes, médianes, régression linéaire…). Nous avons choisi d'utiliser le ratio des intensités médianes car celles-ci sont moins sensibles que les intensités moyennes aux biais provoqués par les valeurs extrêmes (par exemple, un pixel émettant un signal particulièrement intense à cause d'une poussière déposée dessus). Pour manipuler et analyser les données, les ratios d'expression sont usuellement transformés dans une échelle logarithmique : $\mathbf{\log_2(R_{r/v})}$.

Le logarithme permet de transformer le ratio d'expression en une fonction linéaire, symétrique de – l'infini à + l'infini, centrée sur 0 et additive, ce qui facilite les analyses statistiques (notamment, on ne peut calculer les moyennes arithmétiques des ratios de plusieurs expériences que dans un espace logarithmique). On a ainsi $\log(R_{r/v}) = -\log(1/ R_{r/v})$. N'importe quelle base de logarithme peut être utilisée de manière équivalente. Nous utilisons le logarithme de base 2, qui est commode pour se rendre compte facilement de la vraie valeur du ratio ($\log_2(2) = 1$; $\log_2(4) = 2$; $\log_2(8) = 3$; etc).

5.2.3. La normalisation par rapport à la moyenne globale des intensités

La normalisation consiste à ajuster l'intensité globale des images acquises sur chacun des deux canaux rouge et vert, de manière à corriger des biais techniques systématiques qui tendent à déséquilibrer le signal de l'un des canaux par rapport à l'autre. Ces biais sont dus en particulier aux différences de caractéristiques des deux fluorochromes Cy3 et Cy5, qui ne possèdent pas le même coefficient d'extinction molaire (à incorporation égale, Cy5 émet un signal plus fort que Cy3), aux différences d'incorporation des marqueurs lors de la synthèse des cibles, et aux paramètres de lecture au scanner (réglages de la puissance des lasers…).

On suppose que la majorité des gènes sont exprimés de la même manière dans les deux échantillons comparés, et donc que la majorité des spots émet un signal d'intensité égale en rouge et en vert. Dans ces conditions, la moyenne arithmétique des ratios d'expression $R_{r/v}$ de tous les spots devrait être égale à 1. Si tel n'est pas le cas, nous corrigeons les ratios d'expressions par un facteur de normalisation N, tel que :

Pour chaque gène : $R_{r/v}$ normalisé = N x $R_{r/v}$ où N = 1/exp(moyenne de $\log_2(R_{r/v})$ de tous les spots de la puce)

Afin d'éviter les biais, les spots non exploitables (éliminés par les filtres) et les spots ayant des intensités rouges et vertes très différentes ($R_{r/v}$ >10 ou < 0,1) ne sont pas pris en compte dans le calcul du facteur de normalisation. On peut aussi utiliser la médiane des ratios à la place de la moyenne.

5.3. Analyse et interprétation des résultats

Les puces à ADN permettent de mesurer le niveau d'expression relatif de chaque gène dans un échantillon cellulaire comparé à un contrôle de référence. Plus le ratio d'expression $R_{r/v}$ est élevé, plus la quantité d'ARNm du gène est importante dans l'échantillon considéré.
Classiquement, on considère que le niveau d'expression du gène est significativement plus élevé dans l'échantillon par rapport à la référence si $R_{r/v} \geq 2$ (soit $\log_2(R_{r/v})=1$), et inversement qu'il est significativement réduit si $R_{r/v} \leq 0,5$ (soit $\log_2(R_{r/v})=-1$). Ce seuil a été originellement établi en supposant que les ratios d'expression ont une distribution aléatoire suivant la loi normale, auquel cas l'intervalle de confiance à 95% de la valeur du ratio est de $\pm$ 1,96 (soit environ $\pm$ 2).
Dans la pratique, des seuils de confiance sont fixés de manière arbitraire pour déterminer si le ratio indique une variation d'expression significative. Certains laboratoires recommandent de calculer un seuil différent à chaque expérience en fonction de la moyenne et de l'écart-type de l'ensemble des ratios dans l'expérience considérée. Cependant, cette méthode n'est applicable que si le nombre de gènes dont l'expression varie fortement est relativement faible par rapport à l'ensemble de la population permettant d'établir la variabilité aléatoire des ratios. Enfin, il est souvent recommandé de reproduire l'expérience plusieurs fois avec des échantillons indépendants afin d'estimer la variabilité technique et biologique pour distinguer l'effet réel étudié des fluctuations aléatoires. Il serait sans doute possible de considérer des seuils plus proches de 1 avec un nombre suffisant d'expériences répétées indépendamment (par exemple, un gène montrant systématiquement un ratio de 1,5 aurait plus de chances d'être transcrit préférentiellement qu'un autre dont le ratio varierait aléatoirement entre 2 et 0,5). Il est toutefois difficile de réaliser des études de reproductibilité à grande échelle en raison notamment des limitations matérielles et financières.

Résultats & Discussion

PARTIE 1

CARACTERISATION DES ACTIVITES ARNT:ψ-SYNTHASES DE L'ENZYME PUS7P DE *S. CEREVISIAE*

1. Contexte

Comme nous l'avons expliqué dans l'introduction du manuscrit, il existe plusieurs éléments de séquence bien conservés au sein des ARN:Ψ-synthases. Ces éléments avaient permis de mettre en évidence 9 gènes pouvant coder des ARN:Ψ-synthases chez *S. cerevisiae*. Des études biochimiques ont ensuite permis d'identifier les substrats pour plusieurs de ces enzymes. Lorsque j'ai débuté ce travail, toutes les ARN:Ψ-synthases de la levure n'étaient pas caractérisées.

Par ailleurs, plusieurs ARN métaboliquement stables de *S. cerevisiae*, contenaient des résidus Ψ pour lesquels aucune ARN:Ψ-synthase n'avait été identifiée (Tableau 14).

ARN		Position	ARN:Ψ-synthase responsable de la formation du résidu	
ARNr	28S LSU	30 positions	Cbf5 + snoRNA H/ACA	
	18S SSU	14 positions		
	5,8S LSU	Ψ73	**Inconnue**	
	5S LSU	Ψ50	**Inconnue**	
snRNA U1		Ψ5	**Inconnue**	
		Ψ6	**Inconnue**	
snRNA U2		Ψ35	Pus7p	
		Ψ42	Cbf5 + snoRNA H/ACA	
		Ψ44	Pus1p	
snRNA U5		Ψ99	**Inconnue**	
ARNt cytoplasmiques			**Cytoplasmiques**	**Mitochondriales**
		Ψ1	Pus1p	/
		Ψ13	**Inconnue**	/
		Ψ26	Pus1p	/
		Ψ27	Pus1p	Pus2p
		Ψ28	Pus1p	Pus2p
		Ψ31	Pus6p	Pus6p
		Ψ32	Pus8p	Pus9p
		Ψ34	Pus1p	/
		Ψ35	**Inconnue**	/
		Ψ36	Pus1p	/
		Ψ38	Pus3p	
		Ψ39	Pus3p	
		Ψ55	Pus4p	
		Ψ65	Pus1p	/
		Ψ67	Pus1p	/
		Ψ72	/	Pus1p

Tableau 14 : Les sites de pseudouridylation connus des différents ARN cellulaires de la levure *S. cerevisiae* et les ARN:Ψ–synthases connues pour être responsables de ces modifications au début de ma thèse.

Toutes les positions modifiées dans les ARNt cytoplasmiques ne le sont pas dans les ARNt mitochondriaux, dans ce cas le signe « / » est utilisé.

Bien que l'ARN:Ψ-synthase Pus1p puisse catalyser *in vitro* la formation du résidu Ψ35 au sein des ARNt cytoplasmiques, la délétion du gène *PUS1* n'affecte pas *in vivo* la formation de ce résidu, d'où l'hypothèse de l'existence d'une seconde ARN:Ψ-synthase capable de modifier cette position.

Les modifications identifiées dans les ARNr, UsnRNA et ARNt, ainsi que les enzymes responsables de leur formation sont indiquées.

Comme la délétion de l'ensemble des gènes codant des ARN:Ψ-synthases potentielles n'abolissait pas la formation de certains des résidus Ψ trouvés dans les ARNt cytoplasmiques, en particulier en positions 13 et 35, cela suggérait, soit que plusieurs enzymes ont des activités redondantes et que de ce fait leur délétion ne conduisait pas à une perte de l'activité de modification *in vivo*, soit que des ARN:Ψ-synthases ne portant pas les signatures caractéristiques identifiées à partir des enzymes d'*E. coli* existaient chez *S. cerevisiae*. Cette dernière hypothèse s'est révélée être exacte, puisque l'équipe de Yu (Ma et al., 2003) a isolé le gène d'une nouvelle ARN:Ψ-synthase par un criblage systématique de l'activité ARN:Ψ-synthase agissant en position 35 du snRNA U2 de *S. cerevisiae*. Le criblage avait été réalisé sur l'ensemble des protéines codées par les différentes ORF de cet organisme. Ce gène *PUS7* code une protéine qui ne possède pas les signatures classiques des ARN:Ψ-synthases connues jusque-là. Etant donné que l'équipe de Yu n'avait testé l'activité de cette enzyme que sur le snRNA U2, il nous a paru intéressant de tester si cette enzyme pouvait en plus agir sur d'autres ARN tels que les snRNA U1 et U5 et les ARNt cytoplasmiques et mitochondriaux.

2. Démarche expérimentale

Un moyen d'obtenir des informations sur l'activité d'une ARN:Ψ-synthase encore non caractérisée est d'analyser les défauts de modifications des ARN extraits d'une souche dans laquelle le gène codant cette enzyme a été inactivé. La vérification que les effets observés sont bien directement liés à la substitution générée peut alors être réalisée par complémentation de la souche portant la délétion par un plasmide exprimant l'ORF de cette enzyme. Par ailleurs, comme toutes les ARN:Ψ-synthases contiennent un acide aspartique essentiel au sein de leur site actif. Une autre confirmation de l'action de l'enzyme sur le ou les substrats identifiés peut être apportée par la transformation de la souche mutée par un plasmide exprimant l'enzyme dans laquelle l'acide aspartique supposé essentiel a été muté. Dans ce cas, aucune complémentation de la souche n'est attendue (Fig. 67). Ces études peuvent alors être complétées par des tests d'activité *in vitro* réalisés avec l'enzyme produite sous forme recombinante.

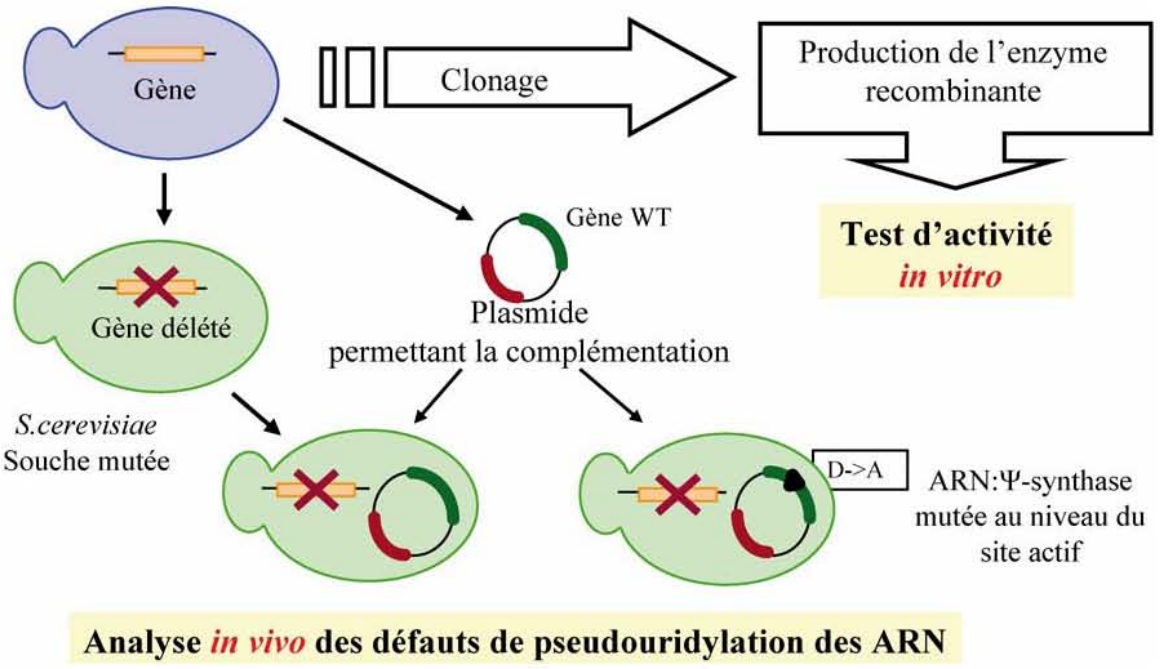


Figure 67 : Représentation schématique de la stratégie utilisée pour caractériser une ARN:Ψ-synthase.

L'analyse des résidus Ψ, présents dans les différents ARN cellulaires a été réalisée à l'aide de la méthode de modification chimique par le CMCT (Bakin & Ofengand, 1993) (Fig. 68). Cette technique repose sur l'addition d'un groupement carbodiimide en position 3 des résidus U et des résidus Ψ. Ce composé réagit aussi mais à un faible niveau sur l'azote en position 1 des résidus G. La réaction est réversible en milieu alcalin pour les résidus U, mais ne l'est pas pour les résidus Ψ. Après traitement à pH 10,4, la présence de modifications sur les résidus U peut alors être détectée par la méthode d'extension d'amorce. Le principe de cette analyse repose sur l'arrêt de l'élongation d'amorce au niveau des résidus modifiés en position Watson Crick, à savoir les résidus U et Ψ, après traitement par le CMCT sans traitement alcalin, et seulement au niveau des résidus Ψ après traitement alcalin. Il est à noter qu'en présence d'un résidu modifié par le CMCT, l'extension de l'amorce par la transcriptase inverse s'arrête au niveau du nucléotide précédant la position modifiée. L'utilisation de la transcriptase inverse pour l'analyse a l'avantage de permettre l'analyse de grandes molécules d'ARN, telles que l'ARN de la grande sous-unité ribosomique.

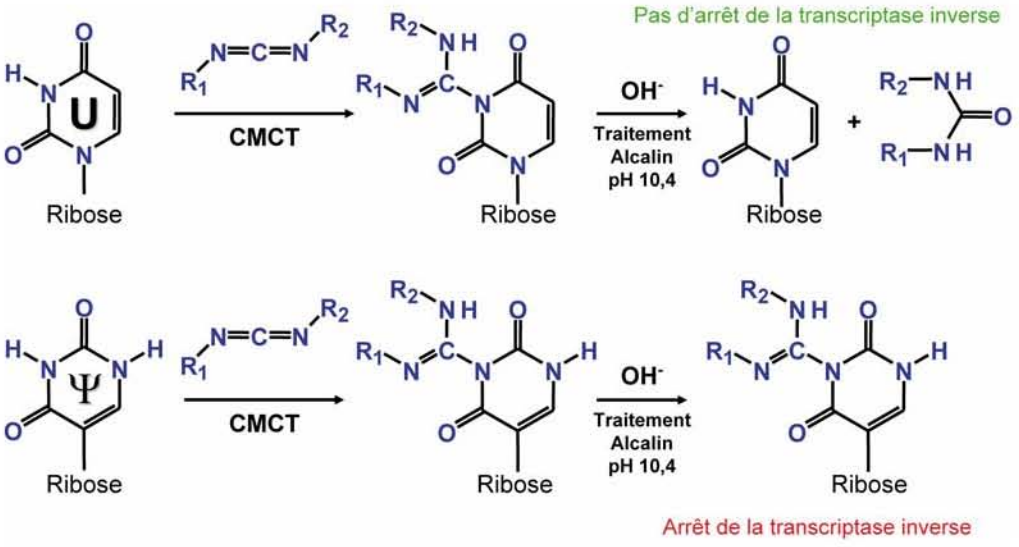


Figure 68 : Principe de la détection des résidus Ψ par modification chimique par le CMCT.

Cette approche permet donc d'analyser rapidement des défauts de modification ayant lieu *in vivo*, quelle que soit la taille de la molécule d'ARN. De plus, elle permet la localisation du résidu modifié par référence aux produits d'un séquençage de l'ARN étudié. Le séquençage est réalisé avec le même oligonucléotide que celui utilisé pour l'extension d'amorce. Cette approche a cependant l'inconvénient de ne pas permettre une quantification précise du taux de modification. Elle ne permet donc pas d'estimer le rendement de formation des résidus Ψ. Elle n'est donc pas adaptée à l'étude *in vitro* de l'efficacité enzymatique des ARN:Ψ-synthases contenues dans des extraits ou produites sous forme recombinante. Dans ce cas, les études doivent être réalisées par la méthode dite du plus proche voisin.

Le principe de cette technique est le suivant (Fig. 69) : l'ARN substrat, utilisé pour la recherche d'une activité ARN:Ψ-synthase donnée, est produit par transcription *in vitro* avec marquage interne par incorporation d'un rNTP radiomarqué, de l'[α-^{32}P] ATP dans l'exemple illustré dans la Figure 69. Cet ARN substrat est ensuite incubé, soit avec des extraits de levure, soit avec une enzyme purifiée, dans des conditions permettant la pseudouridylation de l'ARN. L'ARN est alors hydrolysé par la RNase T2, ce qui permet la libération de nucléotides 3'-monophosphates (3'NMP). Ainsi, seuls les nucléotides situés en 5' d'un nucléotide radiomarqué sont radioactifs et peuvent être détectés par autoradiographie après chromatographie bidimensionnelle sur couche mince de cellulose. Si l'activité ARN:Ψ-synthase recherchée est présente dans l'extrait, le résidu Ψ formé peut être identifié du fait de sa mobilité particulière en chromatographie. Des cartes de références du fractionnement des

nucléotides canoniques et modifiés ont été établies (Nishimura, 1979; Keith, 1995) (Fig. 61). La séquence de l'ARN substrat étant connue, il est alors possible d'estimer le rendement de pseudouridylation obtenu en se basant sur les quantités molaires de résidus A, C, G et U attendues. Il faut tout de même noter que seules de petites molécules d'ARN peuvent être étudiées par cette approche, puisque pour des tailles supérieures à 150 nts le ratio U/Ψ devient trop important et il devient souvent difficile, si ce n'est impossible, de mesurer avec précision le taux de résidus Ψ. Par ailleurs, cette approche ne permet pas de localiser le résidu Ψ formé.

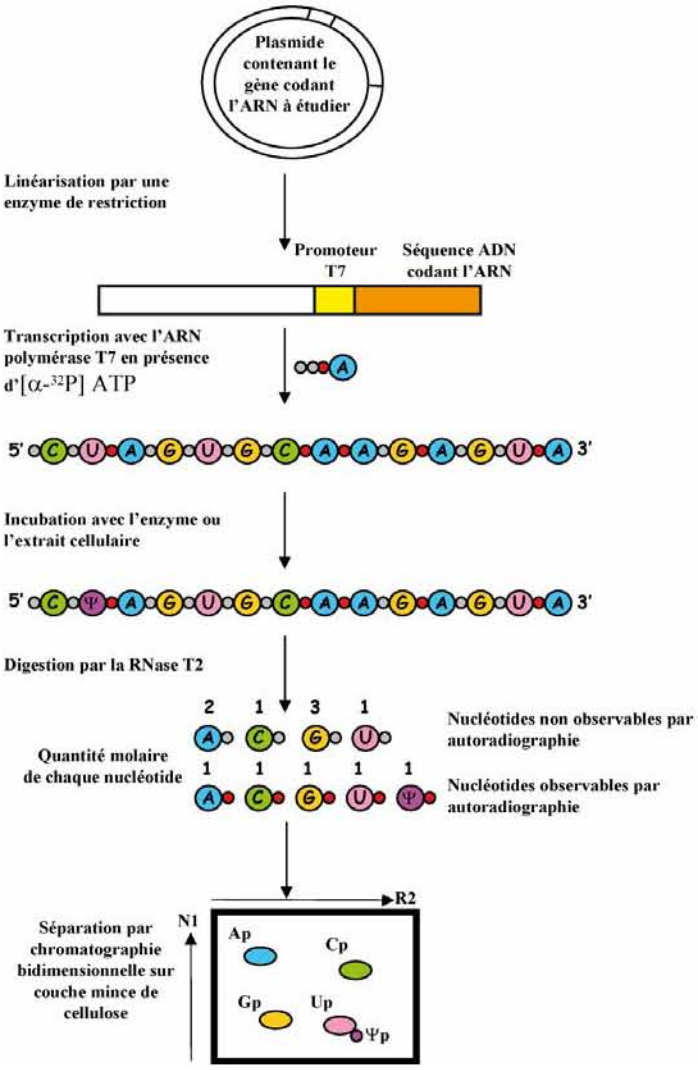


Figure 69 : Principe de la recherche de nucléotides modifiés par la technique du plus proche voisin (d'après Silberklang et al., 1979).

Les expériences CMCT-RT ont été réalisées sur des ARN totaux extraits de la souche haploïde commerciale BY4742 dans laquelle l'ORF YOR243c était substituée par le gène *kanMX4* (EUROSCARF). Différents oligonucléotides amorces, décrits dans la publication n°2, ont été définis de façon à analyser les régions des snRNA U1, U2 et U5 renfermant des nucléotides modifiés et les segments d'ARNt cytoplasmiques et mitochondriaux renfermant des résidus Ψ pour lesquels l'ARN:Ψ-synthase catalysant la modification n'avait pas été caractérisée. Les résultats obtenus par ces expériences nous ont permis d'identifier de nouveaux substrats pour Pus7p. Nous avons ensuite validé ces résultats en analysant l'activité de l'enzyme Pus7p sur ces ARN produits par transcription *in vitro* en utilisant soit des extraits cellulaires de levure, soit la protéine recombinante produite et purifiée *in vitro*.

3. Résultats (travail réalisé en collaboration avec I. Behm-Ansmant)

Pour cela, nous avons cloné l'ORF de l'enzyme Pus7p dans 2 types de vecteurs. Un premier vecteur navette *E. coli*/*S. cerevisiae* p416GalS (Mumberg et al., 1995) a été utilisé pour complémenter la souche mutante par un gène actif. Nous avons produit 2 versions de ce plasmide : une version permettant l'expression de la protéine sauvage et une autre, permettant l'expression de la protéine mutée au niveau de l'acide aspartique essentiel du site actif. Dans ce plasmide, l'ORF est placée sous contrôle du promoteur régulable GalS et de ce fait l'enzyme Pus7p n'est produite que lorsque les cellules sont placées dans un milieu de culture contenant du galactose. La substitution du résidu acide aspartique catalytique en résidu alanine a été réalisée par mutagenèse dirigée par la méthode PCR.

Le second vecteur de clonage que nous avons utilisé est un plasmide de la série pET (pET28b). Il permet la production de l'enzyme sous forme recombinante chez *E. coli*. L'activité de l'enzyme recombinante a ensuite été testée, soit sur des ARN substrats produits par transcription *in vitro*, soit sur des ARN totaux issus de la souche mutée pour le gène *PUS7*. Afin de contrôler que les activités obtenues ne correspondaient pas à des activités contaminantes d'*E.coli*, nous avons produit par la même approche l'enzyme Pus7p dont le résidu Asp catalytique était muté.

Comme mentionné ci-dessus, l'équipe de Yu venait de démontrer que l'enzyme Pus7p modifiait la position 35 du snRNA U2 de *S. cerevisiae* (Ma et al., 2003). De manière intéressante, une recherche dans des banques de séquences génomiques révélait l'existence d'homologues de Pus7p dans les 3 règnes du vivant (archaea, bactéries et eucaryotes). Comme il n'y pas de snRNA chez les archaea et les bactéries, ces homologues doivent agir sur d'autres substrats ARN (comme les ARNt). Il était donc possible que Pus7p puisse agir, de façon similaire, sur d'autres types d'ARN que le snRNA U2.

Afin de tester cette hypothèse, nous avons, dans un premier temps, établi une collaboration avec l'équipe de Yu qui nous a fourni de l'enzyme purifiée. Ensuite nous avons produit Pus7p au laboratoire et nous avons utilisé une souche de *S. cerevisiae* où le gène

PUS7 était inactivé. En utilisant cette souche, nous avons pu montrer par emploi de la technique CMCT-RT qu'en ce qui concerne les UsnRNA, seul le résidu en position 35 du snRNA U2 est converti en résidu Ψ. En effet, la souche ΔPus7 ne présente aucun autre défaut de pseudouridylation dans le snRNA U2 et dans les snRNA U1 et U5. Par contre, un défaut de conversion du résidu U13 en résidu Ψ a été observé dans certains ARNt cytoplasmiques (10 ARNt cytoplasmiques portent cette modification). Un défaut de formation du résidu Ψ35 dans l'ARNt^{Tyr} cytoplasmique (seul cet ARNt porte cette modification) a aussi été observé et cette modification n'a lieu que dans le pré-ARNt^{Tyr} contenant un intron (Johnson & Abelson, 1983; Choffat et al., 1988; van Tol & Beier, 1988; Zerfass & Beier, 1992). La formation du résidu Ψ35 dans le pré-ARNt^{Tyr}, comme celle en position 35 du snRNA U2, doit avoir lieu dans le noyau.

Par production de l'enzyme recombinante et en utilisant des extraits des cellules sauvages ou mutées de *S. cerevisiae*, nous avons pu confirmer que Pus7p, comme Pus1p, est capable d'agir sur 2 types de substrats : le snRNA U2 et les ARNt cytoplasmiques. Un alignement de séquences réalisé à partir de l'ensemble des substrats naturels de Pus7p montre que le résidu Ψ est toujours présent au sein d'une séquence conservée de 7 nucléotides : Pu(G/C)UNΨAPu (Pu= purine, N= nucléotide quelconque. D'où l'étude qui a suivi concernant le rôle de cette séquence dans la spécificité de reconnaissance de l'ARN à modifier par Pus7p (Partie Résultats - Chapitre 2).

La présence d'un homologue de Pus7p chez les bactéries a été établie par l'équipe d'Ofengand qui a isolé et caractérisé l'enzyme TruD d'*E. coli* (Kaya & Ofengand, 2003). Sur la base de l'alignement de la séquence de Pus7p de *S. cerevisiae* avec celle de tous ces homologues potentiels chez les autres espèces dont le génome avait été séquencé, plusieurs motifs très conservés ont pu être mis en évidence, en particulier, un motif renfermant un acide aspartique. J'ai alors testé par remplacement de cet acide aminé par un résidu alanine, si celui-ci pouvait être le résidu catalytique. L'observation d'une perte totale d'activité, aussi bien dans des expériences réalisées *in vitro* que dans des expériences réalisées *in vivo*, suggérait fortement que ce résidu était le résidu catalytique. Ces travaux ont donné lieu à l'article inséré ci-dessous, dont je suis second auteur et qui a été publié dans la revue RNA.

Publication n°2

The *Saccharomyces cerevisiae* U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Ψ-synthase also acting on tRNAs

Behm-Ansmant I., Urban A., Ma X., Yu YT., Motorin, Y. and Branlant C.
RNA, **9(11)**, 1371-82, 2004.

Publication n°2

The Saccharomyces cerevisiae U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite–multisubstrate RNA: Ψ-synthase also acting on tRNAs

ISABELLE BEHM-ANSMANT, [1] ALAN URBAN, [1] XIAOJU MA, [2] YI-TAO YU, [2] YURI MOTORIN, [1] and CHRISTIANE BRANLANT [1]

[1]Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP Nancy I, Faculté des Sciences, BP 239, 54506 Vandoeuvre-les-Nancy Cedex, France

[2]Department of Biochemistry and Biophysics, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, New York 14642, USA

ABSTRACT

The Saccharomyces cerevisiae Pus7 protein was recently characterized as a novel RNA:pseudouridine (Ψ)-synthase acting at position 35 in U2 snRNA. However, U2 snRNA was the only potential substrate tested for this enzyme. In this work, we demonstrated that although Pus7p is responsible for the formation of only one of the six Ψ residues present in yeast UsnRNAs, it catalyzes U to Ψ conversion at position 13 in cytoplasmic tRNAs and at position 35 in pre-tRNA ^{Tyr}. Sites of RNA modification by Pus7p were identified by analysis of the in vivo RNA modification defects resulting from the absence of active Pus7p production and by in vitro tests using extracts from WT and genetically modified yeast cells. For demonstration of the direct implication of Pus7p in RNA modification, the activity of the WT and mutated Pus7p recombinant proteins was tested on in vitro produced tRNA and pre-tRNA transcripts. Mutation of an aspartic acid residue (D256) that is conserved in all Pus7 homologs abolishes the enzymatic activity both in vivo and in vitro. This suggests the direct involvement of D256 in catalysis. Target sites of Pus7p in RNAs share a common sequence Pu(G/C)UN ΨAPu (Pu = purine, N = any nucleotide), which is expected to be important for substrate recognition. Modification of tRNAs by Pus7p explains the presence of Pus7p homologs in archaea and some bacteria species, which do not have U2 snRNA, and in vertebrates, where Ψ34 (equivalent to Ψ35 in yeast) formation in U2 snRNA is an H/ACA snoRNA guided process. Our results increase the number of known RNA modification enzymes acting on different types of cellular RNAs.

Keywords: Yeast; pseudouridine synthase; tRNAs; pre-tRNA ^{Tyr}; UsnRNAs; intron-dependent

INTRODUCTION

Most of the metabolically stable RNAs (tRNAs, rRNAs, and UsnRNAs) contain post-transcriptionally modified nucleotides, the most frequent ones are Ψ and 2′-O-methylated residues. These modified residues may be generated in RNA by two distinct mechanisms. The first one is based on a single protein (RNA: Ψ-synthase), which carries both the RNA recognition capacity and the catalytic activity (for review, see Charette and Gray 2000; Ansmant and Motorin 2001; Ofengand 2002), while in the second one, the modification is performed by an RNP complex that carries the two activities. In this case, a guide RNA (H/ACA sno- or sca-RNA) ensures the recognition of targeted RNA, and one of the proteins bound to the guide RNA has the catalytic activity (Ganot et al. 1997; Ni et al. 1997; and for review, see Lafontaine and Tollervey 1998; Filipowicz et al. 1999; Kiss 2001, 2002). The advantage of this second system is that a single RNA: Ψ-synthase bound to several guide RNAs can catalyze Ψ formation at several positions in a given RNA and also in different RNAs.

Some of the post-transcriptional modifications in tRNAs, rRNAs, and UsnRNAs are highly phylogenetically conserved (Branlant et al. 1981; Veldman et al. 1981; and for review, see Massenet et al. 1998; Charette and Gray 2000;

Reprint requests to: Yuri Motorin, Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP Nancy I, Faculté des Sciences, BP 239, 54506 Vandoeuvre-les-Nancy Cedex, France; e-mail Yuri.Motorin@maem.uhp-nancy.fr; fax: 33.3.83.68.43.07.

Abbreviations: A, adenosine; NMP, monophosphate nucleotide; Ψ, pseudouridine; PCR, polymerase chain reaction; CMCT, (1-cyclohexyl-3-[2-morpholinoethyl] carbodiimide metho-p-toluenesulfonate); RT, reverse transcriptase; TLC, thin layer chromatography; U, uridine; WT, wild type.

Article and publication are at http://www.rnajournal.org/cgi/doi/10.1261/rna.5520403.

Decatur and Fournier 2002; Ofengand 2002). This is the case for several pseudouridylation sites in tRNAs (positions 13, 32, 38, 39, and 55; for review, see Sprinzl et al. 1998) and also for six pseudouridylation sites in the spliceosomal UsnRNAs; three of them are located in U2 snRNA (positions 35, 42, 44 in yeast and 34, 41, 43 in human or Xenopus laevis; for review, see Massenet et al. 1998). However, such highly phylogenetically conserved modified residues in RNA can be generated by different mechanisms, depending on the organism. This was first demonstrated for Ψ residues in ribosomal RNAs. Their formation is catalyzed by several specific enzymes in bacteria (for review, see Ofengand 2002) and by snoRNPs in eukaryotes (Ganot et al. 1997; for review, see Lafontaine and Tollervey 1998; Kiss 2001, 2002). Recently, it turned out to be the case for two of the conserved Ψ residues in U2 snRNA. Indeed, modifications at positions 34 and 43 (Ψ45 in Caenorhabditis elegans) in vertebrates are likely dependent upon H/ACA guide RNAs (Huttenhofer et al. 2001; Zhao et al. 2002; Patton and Padgett 2003), whereas the corresponding Ψ residues in the S. cerevisiae U2 snRNA (Ψ35 and Ψ44) are generated by the snoRNA-independent enzymes Pus7p and Pus1p, respectively (Massenet et al. 1999; Ma et al. 2003). Interestingly, Pus1p is a multisubstrate enzyme, which also catalyzes the formation of Ψ residues at eight distinct positions in tRNAs (Simos et al. 1996; Motorin et al. 1998).

From the RNA recognition point of view, RNA: Ψ-synthases, that catalyze uridine (U) to Ψ conversion without assistance of guide RNAs, can be classified into three groups: (1) site-specific enzymes acting on a unique site in a given RNA (like RsuA, Conrad et al. 1999; TruB, Nurse et al. 1995; and Pus4, Becker et al. 1997); (2) region-specific enzymes capable to modify several neighboring positions in a given molecule (e.g., TruA, Kammen et al. 1988; Pus3, Lecointe et al. 1998; RluC, Conrad et al. 1998; and RluD, Huang et al. 1998); and (3) multisite and multisubstrate specific enzymes that modify distinct positions in different classes of RNAs (like Escherichia coli RluA, Wrzesinski et al. 1995). The yeast Pus1p enzyme mentioned above belongs to this last group (Motorin et al. 1998; Massenet et al. 1999). Enzymes with broad and dual specificities may be more frequent than currently imagine, because an RNA:methyltransferase with dual specificity was also described (Gu et al. 1994). Together with the guide RNA system, such multisite-specific proteins probably contribute to the generation of a large number of modifications in RNAs with only a limited number of enzymes. Hence, we tested whether the S. cerevisiae Pus7 enzyme, which, like Pus1p, modifies U2 snRNA, can also modify tRNAs. Indeed, despite the almost complete characterization of the S. cerevisiae genes encoding proteins with RNA: Ψ-synthase signatures (Simos et al. 1996; Lecointe et al. 1998; Becker et al. 1997; Ansmant et al. 2000, 2001), the enzymes catalyzing the formation of the frequent Ψ13 residue in cytoplasmic tRNAs (7 tRNAs), Ψ1 residue in cytoplasmic tRNA^{Arg} and tRNA^{Lys}, and Ψ72 residue in mitochondrial initiator tRNA^{Met} are not discovered yet. Furthermore, the situation concerning Ψ35 formation, which is restricted to tRNA^{Tyr} in all eukaryotes and takes place on the intron-containing tRNA precursor (Johnson and Abelson 1983), is not clear. Indeed, despite the fact that the yeast multisite–multisubstrate-specific RNA: Ψ-synthase Pus1p modifies this position in pre-tRNA^{Tyr} transcripts in vitro, disruption of the PUS1 gene in S. cerevisiae does not abolish Ψ35 formation in tRNA^{Tyr}, either in cellular extract or in vivo (Motorin et al. 1998). These data indicated that an additional, yet uncharacterized, yeast RNA: Ψ-synthase is responsible for U35 modification in tRNA^{Tyr}.

In this work, we tested whether the recently identified U2 snRNA: Ψ-synthase Pus7p may display multisubstrate specificity and modify one of the three orphan pseudouridylation sites in cytoplasmic tRNAs (positions 1, 13, and 35), and/or one of the four orphan pseudouridylation sites in yeast UsnRNAs (Ψ5 and Ψ6 in U1 snRNA, Ψ42 in U2 snRNA and Ψ99 in U5 snRNA) (Massenet et al. 1999). The results reveal that, like Pus1p, Pus7p modifies both U2 snRNA and tRNAs, and that tRNA modification occurs at two distinct positions (13 and 35; Fig. 1). Interestingly, Pus7p does not belong to any of the four identified families of RNA: Ψ-synthases (Ma et al. 2003), explaining why it was not previously identified by computer search based on sequence homology approach (Koonin 1996).

RESULTS

Ψ formation in S. cerevisiae U1 and U5 snRNAs is not dependent upon Pus7p

Ma et al. (2003) showed that disruption of the PUS7 gene in S. cerevisiae abolishes U to Ψ conversion at position 35 in U2 snRNA (Fig. 1A), without altering Ψ42 and Ψ44 formation. Because the S. cerevisiae enzymes responsible for Ψ formation at positions 5 and 6 in U1 snRNA and at position 99 in U5 snRNA are not identified yet, we used the CMCT/RT approach (Bakin and Ofengand 1993) to test for a possible activity of Pus7p at these positions of UsnRNAs. To this end, we analyzed U1 and U5 snRNAs, extracted from the wild-type (WT) and two ΔPUS7 S. cerevisiae BY4742 strains (a and α mating types). The reverse transcriptase stops, indicating the presence of Ψ residues, were identical for RNAs from the WT and both ΔPUS7 strains. Only the results for the mat-a ΔPUS7 strain are shown (Fig. 2). Thus, PUS7 gene disruption does not affect U to Ψ conversion in both U1 and U5 snRNAs. These data strongly indicate that Pus7p forms only one of the six Ψ residues detected in S. cerevisiae UsnRNAs.

Disruption of the PUS7 gene abolishes Ψ13 formation in tRNAs in vivo

As mentioned in the Introduction, the yeast enzymes responsible for formation of Ψ1, Ψ13, and Ψ35 in cytoplas-

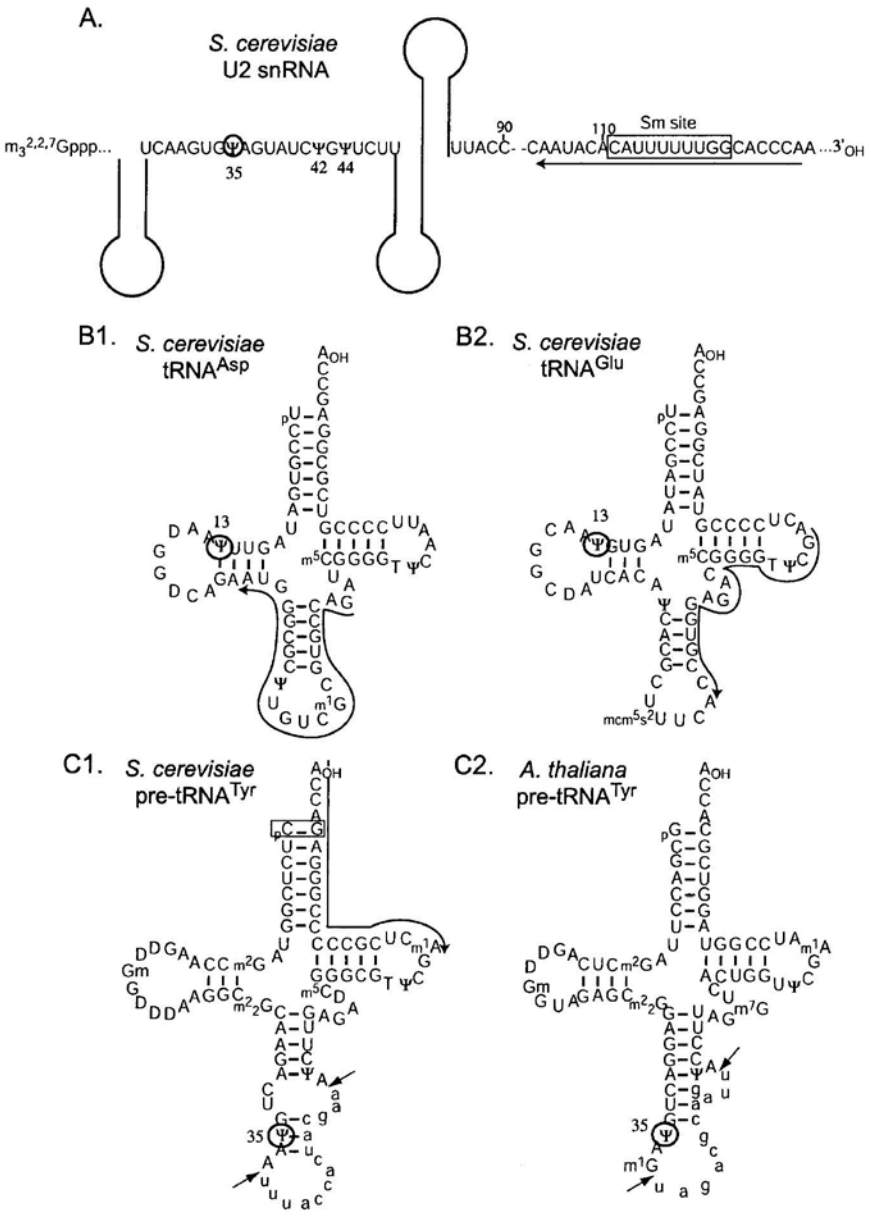


FIGURE 1. Sequences and secondary structures of the Pus7p RNA substrates studied in this work. (A) The 5 -terminal part of S. cerevisiae U2 snRNA containing the three identified residues and the Sm site (Massenet et al. 1999). (B) The S. cerevisiae tRNA ^{Asp} (B1) and tRNA ^{Glu} (B2). (C) The S. cerevisiae (C1) and A. thaliana (C2) pre-tRNA ^{Tyr} are drawn with all their identified post-transcriptional modifications (tRNA database; Sprinzl et al. 1998; see also Internet site http://www.uni-bayreuth.de/departments/biochemie/sprinzl/trna/). The intronic sequences in pre-tRNAs are shown in small characters and arrows indicate exon–intron borders. The residues found to be formed by Pus7p are circled. Oligonucleotides used for primer extension analysis are indicated. In the yeast pre-tRNA ^{Tyr} transcript the C1–G72 base pair (boxed) was converted into a G1–C72 base pair to increase transcription efficiency.

mic tRNAs and of 72 in mitochondrial tRNA ^{Met} have not been identified. The CMCT/RT approach used above for UsnRNAs could also be used to test for the presence of 13 in cytoplasmic tRNAs. Formation of residues 1 and 72 could not be analyzed by this approach due to their location at the 5 extremity and very close to the 3 extremity of tRNAs, respectively. Use of the CMCT/RT approach for analysis of 35 in tRNA ^{Tyr} also turned out to be impossible because of a pause of the reverse transcriptase at residue A36, that was found to be independent from tRNA modification. Thus, the CMCT/RT approach was only used to test for the presence of residue 13 in tRNAs and two cytoplasmic tRNAs, tRNA ^{Asp}(GUC), and tRNA ^{Glu}(UUC) (Fig. 1B), which both contain a residue at position 13 were used in the assays. The results obtained for the PUS7 mat-a strain are shown in Figure 3. They clearly demonstrate the absence of 13 formation in tRNA ^{Asp} and tRNA ^{Glu} extracted from the PUS7 strain (disappearance of pauses in lanes 5,6 in Fig. 3A,B). These data suggested the implication of Pus7p in 13 formation in cytoplasmic tRNAs.

A cell-free extract from the PUS7 strain is defective in 13 and 35 formation in tRNAs

Incubation of in vitro produced tRNA and pre-tRNA transcripts with yeast cell extracts was previously shown to generate specific U to conversions at almost all positions known to be modified in vivo (Jiang et al. 1997; Motorin et al. 1998). Thus, to complete the analysis of RNA-modification defects in the PUS7 strain, we compared the capacities of WT and PUS7 cell extracts to modify different yeast tRNA transcripts at positions 1 and 13 and an in vitro produced pre-tRNA ^{Tyr} at position 35. Formation of residues at the expected positions was tested by the nearest neighbor approach. The tRNA ^{Asp}(GUC) and tRNA-^{His}(GUG) transcripts used in the assays were labeled by incorporation of [-^{32}P]ATP, because among all residues described for these two tRNAs, only residue 13 is followed by an adenosine (A) residue (Sprinzl et al. 1998; see Fig. 1B). Based on the hypothesis that yeast extract reproduces the tRNA modification pattern found in vivo, among the 3 monophosphate nucleotides (3 NMPs) released by T2 RNase digestion of the modified transcripts, only residue 13 was expected to be labeled and thus detected by chromatography on thin layer plates (TLC). In accordance with 13 formation with the extract from the WT strain, a U to conversion occurred

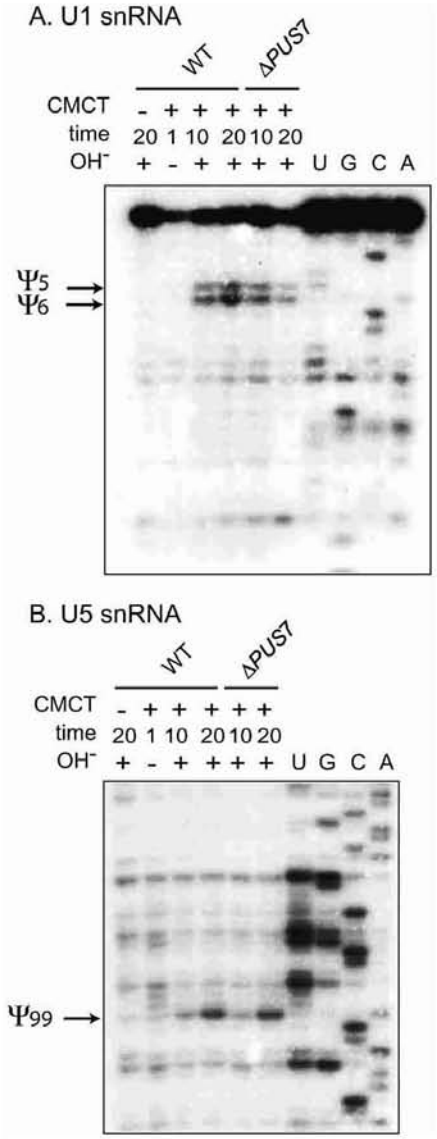


FIGURE 2. Disruption of the PUS7 gene does not alter formation of residue in U1 (A) and U5 snRNAs (B). Total RNA was extracted from the WT and mat-a PUS7 S. cerevisiae BY4742 strains and modified by CMCT, for 1, 10, and 20 min with (+) or without (–) subsequent alkaline treatment (OH^-). A control experiment was performed in the absence of CMCT treatment. Lanes U, G, C, and A correspond to the sequencing ladders obtained with the same oligonucleotide. The reverse transcription stops, corresponding to residues 5 and 6 (A) and to 99 (B) are indicated by arrows.

5 to an adenosine residue of tRNA^{Asp} (Fig. 4A) and tRNA^{His} (data not shown) upon incubation in this extract. Such conversion was abolished in the extract from the PUS7 strain (Fig. 4A). Similarly, the activity toward U35 was tested on an in vitro produced precursor of the S. cerevisiae tRNA^{Tyr} (Fig. 1C1), that was labeled by incorporation of [-^{32}P]ATP. Labeling with this nucleotide was justified by the fact that residue 35 is the only residue in tRNA^{Tyr} followed by a A residue. Here again, in accordance with the presence of an RNA: -synthase activity in the WT extract, T2 RNase digestion followed by 2D TLC demonstrated the formation of a residue 5 to a A residue upon incubation with this extract (Fig. 4B). No labeled 3 MP was detected in the same conditions with the extract from the PUS7 strain (Fig. 4B). Moreover, the same result was obtained when the Arabidopsis thaliana pre-tRNA^{Tyr} (Fig. 1C2), was used as the substrate. This strongly suggested that Pus7p is responsible for 35 formation in the pre-tRNA^{Tyr}. As a control, we confirmed that deletion of the PUS1 gene in the BY4742 strain did not alter 35 formation in both S. cerevisiae and A. thaliana pre-tRNA^{Tyr}.

The tRNA: 1-synthase activity of the extracts was tested on a transcript of the S. cerevisiae cytoplasmic tRNA^{Arg}(ACG). Labeling was done with [-^{32}P]UTP, because among all the residues present in tRNA^{Arg}, only residue 1 is followed by an U residue. As evidenced by TLC, both the WT and the PUS7 cell extracts were capable to convert the 5 residue of a UU sequence into a residue in tRNA^{Arg} (Fig. 4C).

Taking into account the previously demonstrated specificity of yeast extracts, we concluded that Pus7p is most probably involved in the formation of residues 13 and 35, but not of residue 1 of the S. cerevisiae cytoplasmic tRNAs. For a more direct demonstration of the implication of Pus7p in 13 and 35 formation, we complemented the PUS7 strain with the p413GalS–PUS7 plasmid, bearing a WT copy of the PUS7 gene under the control of the inducible GalS promoter. As shown by the CMCT/RT analysis of RNAs extracted from the complemented strain (Fig. 3A,B), expression of the PUS7 gene from plasmid p413GalS restored 13 formation in tRNA^{Asp} and tRNA^{Glu} (lanes 7,8 in Fig. 3A,B, respectively). Restoration of U to conversion was also observed when extracts from the PUS7 and PUS7 complemented strains were used to test for the activities at position 13 of tRNA^{Asp} and 35 of pre-tRNA^{Tyr} (Fig. 4A,B).

Mutation of the conserved aspartic acid residue (D256) in Pus7p abolishes RNA: -synthase activity

Sequence alignment of bacterial, archaeal, and eukaryotic Pus7p homologs revealed the universal conservation of several aspartic acid residues. However, one of them is present in a conserved GTKD sequence showing some similarity with the conserved sequences at the active site of RNA: -synthases from other families. Thus, D256 could be the catalytic residue in Pus7p. To test for this possibility, we introduced a mutation in the PUS7 gene that generated a D256A substitution in Pus7p, and we tested the effect of this mutation on the Pus7p RNA: -synthase activity in vivo and in vitro. As shown by the CMCT/RT approach (Fig. 3A,B, lanes 9,10), the mutated Pus7 enzyme was no longer capable to form residue 13 in tRNA^{Asp} and tRNA^{Glu} in vivo. As a confirmation, residue 13 was not formed in tRNA^{Asp} upon incubation with an extract of the PUS7

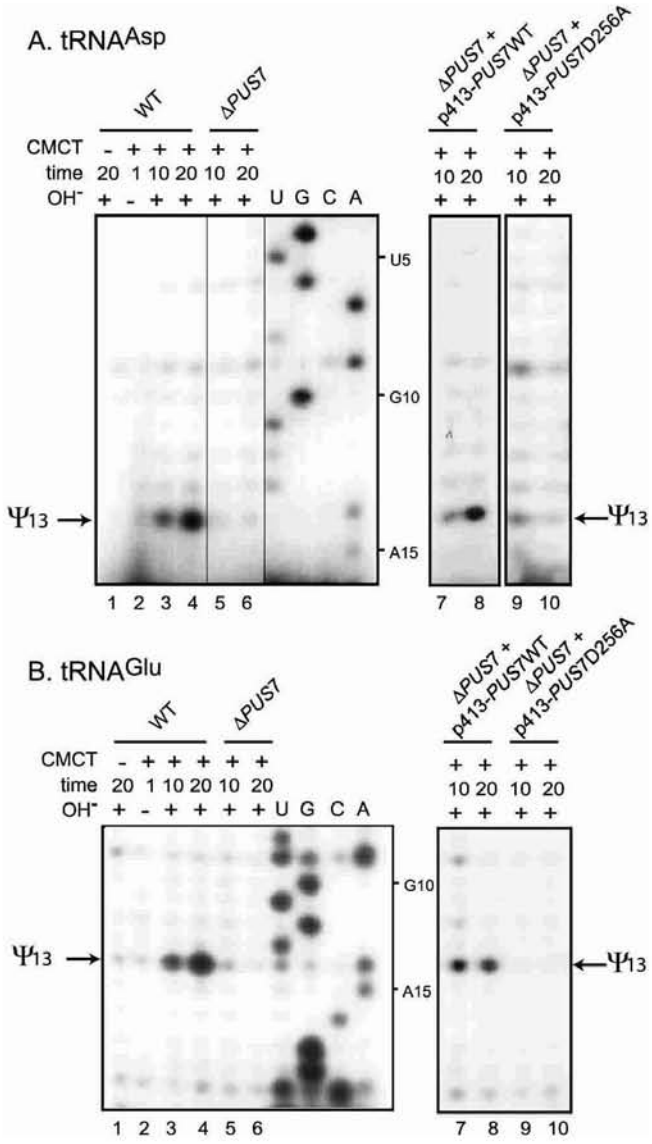


FIGURE 3. 13 formation in cytoplasmic tRNA^{Asp} (A) and tRNA^{Glu} (B) is abolished upon PUS7 gene disruption and restored by complementation with plasmid p413GalS-PUS7. Total RNA was extracted from the WT and mat-a PUS7 strains and modified by CMCT, for 1, 10, and 20 min with (+) or without (–) subsequent alkaline treatment (OH$^-$). A control experiment was performed in the absence of CMCT treatment. Lanes U, G, C, and A correspond to the sequencing ladders obtained with the same oligonucleotide. The reverse transcription stops, corresponding to residues 13, are indicated by arrows. To verify the direct involvement of Pus7p in the U to conversion at position 13 of tRNAs, the PUS7 strain was transformed with plasmid p413GalS-PUS7, bearing the wild-type (WT) or mutated (D256A) PUS7 gene. Total RNA was extracted from these two transformed strains, and tRNA^{Asp} (A) and tRNA^{Glu} (B) were analyzed by the CMCT/RT approach (shown in lanes 7, 8, 9, and 10 in A and B).

strain transformed with a plasmid carrying the D256A mutated PUS7 gene (Fig. 4A).

Similarly, no RNA: -synthase activity was detected, when the pre-tRNA^{Tyr} labeled with [-^{32}P]ATP was incubated with this extract (Fig. 4B). Altogether, these data strongly suggested that D256 is the catalytic residue of Pus7p.

Recombinant Pus7p has tRNA^{Asp}: 13- and pre-tRNA^{Tyr}: 35-synthase activities

To confirm the direct implication of Pus7p in 13 and 35 formation in cytoplasmic tRNAs, we produced the recombinant WT His$_6$–Pus7 and His$_6$–Pus7D256A proteins in E. coli. The RNA: -synthase activities of the WT and mutant His$_6$–Pus7 proteins towards the uridine at position 13 in tRNA^{Asp} were tested by both the CMCT/RT and the nearest neighbor approaches. As expected, using the CMCT/RT approach (Fig. 5A), a reverse transcriptase stop corresponding to 13 formation was detected in the transcript incubated with the WT (lanes 5,6), but not with the variant recombinant enzyme (data not shown). In addition, no other stop indicative of U to conversion at other positions in tRNA^{Asp} was detected. Similarly, UA to A conversion in tRNA^{Asp} labeled with [-^{32}P]ATP was detected for the WT, but not for the variant recombinant enzyme (Fig. 5B). To confirm that the U to conversion detected occurred exclusively at position 13, a U13C base substitution was generated in tRNA^{Asp}. As expected, no UA to A conversion was detected in this variant tRNA^{Asp} (Fig. 5C). Having shown by this experiment the specificity of the U to conversion detected by TLC, we then used this method for a time-course study of the modification reaction (Fig. 5D). The kinetic curve showed that, in the experimental conditions used, a plateau level of about 0.9 mole of residue per mole of tRNA^{Asp} was reached after 40 min of incubation. Taken together, these data unambiguously map the site of action of the recombinant Pus7p enzyme at position 13 in tRNA^{Asp} and exclude its possible action at other U residues in this tRNA.

Similarly, the capacity of the recombinant His$_6$–Pus7 protein to catalyze UA to A conversion in both yeast and A. thaliana pre-tRNAs^{Tyr} labeled with [-^{32}P]ATP was demonstrated by TLC analysis (Fig. 6A, and data not shown). No UA to A conversion was detected when the experiment was performed with the D256A variant protein (Fig. 6A). As several U residues in yeast pre-tRNA^{Tyr} are present in a UA sequence context and that there was no evidence that the purified enzyme will be as much specific as were the yeast extracts, a precise mapping of the formed residue was made by combining T1 and T2 or P1 RNase digestions, gel electrophoresis, and TLC. According to T1 RNase specificity, complete digestion of the pre-tRNA^{Tyr} by this enzyme was expected to release several short oligonucleotides ranging in size from 1 to 8 nt, and a longer fragment of 15 nt corresponding to the tRNA anticodon loop and the intron (Fig. 6B1). This fragment contains U35 expected to be the target of the recombinant His$_6$–Pus7 enzyme. Thus, we analyzed the presence of residue 35 in this T1 RNase product. First, the T1 RNase digestion prod-

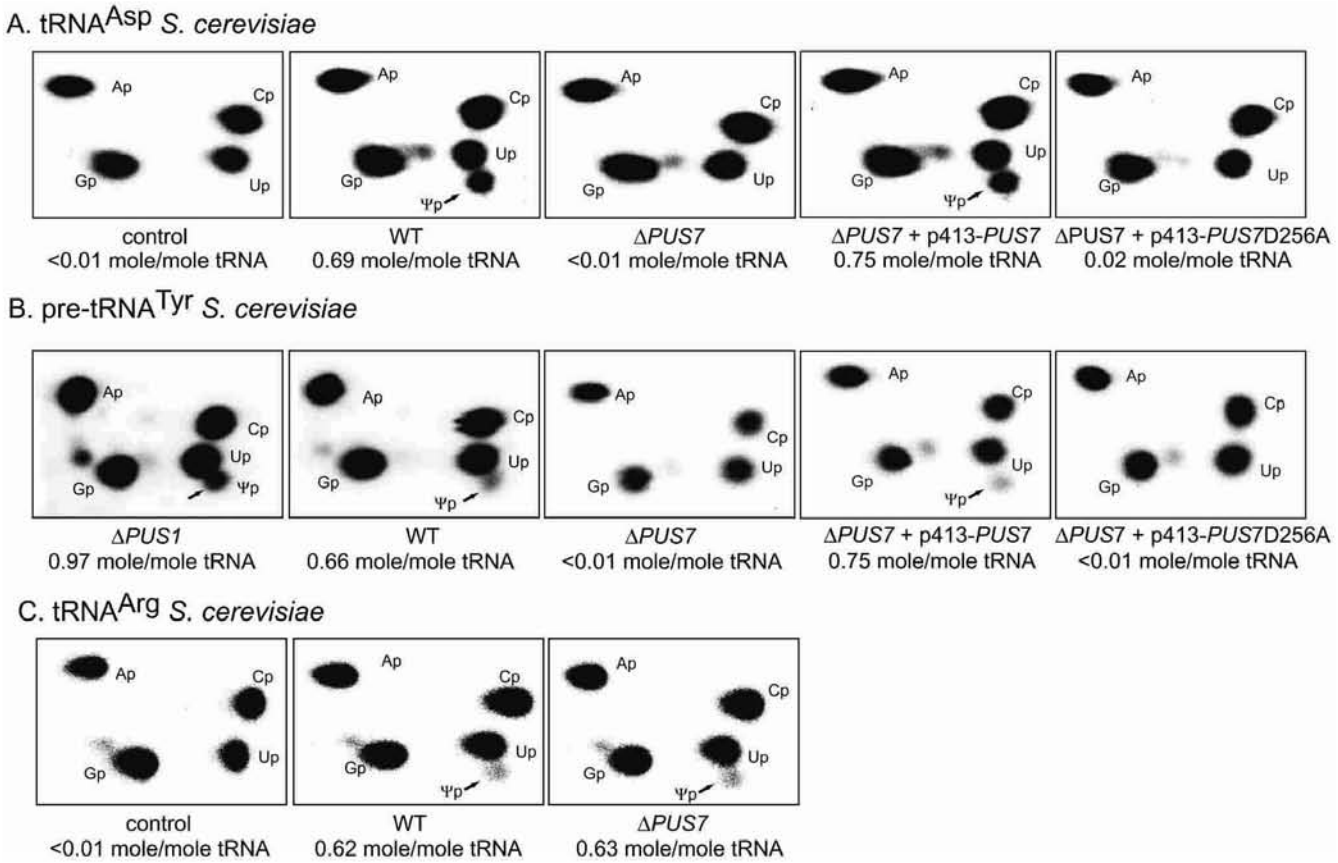


FIGURE 4. Tests of the tRNA: 1, 13 and 35-synthase activities in different yeast S10 extracts. In vitro transcribed RNA substrates labeled by incorporation of [-^{32}P]ATP (yeast tRNA^{Asp}, A, yeast pre-tRNA^{Tyr}, B) or [-^{32}P]UTP (yeast tRNA^{Arg}, C) were incubated with different S10 extracts in the conditions described in Materials and Methods. Extracts were prepared from cells of the WT BY4742 strain (WT), the isogenic PUS7 strain, and the PUS7 strain complemented with the WT or D256A variant PUS7 gene. The activity on yeast pre-tRNA^{Tyr} of an extract from the PUS1 BY4742 strain was also tested. The tRNA^{Asp} and tRNA^{Arg} transcripts incubated in the same conditions but in the absence of S10 extract, were used as controls. After incubation, the transcripts were digested with T2 RNase and 3 NMPs were fractionated on TLC as described in Materials and Methods. The autoradiograms of the TLC plates are shown. Positions of the NMPs (Ap, Cp, Up, Gp) and MP nucleotides were identified according to Keith (1995). Quantification of the residue formation was done by measuring the radioactivity in each spot with a PhosphoImager and the ImageQuant software.

ucts from a pre-tRNA^{Tyr} labeled with [-^{32}P]ATP and incubated in the presence or absence of the recombinant His$_6$–Pus7 enzyme were purified by electrophoresis on a 20% polyacrylamide gel (Fig. 6B2) and the 15-nt fragment was further analyzed by T2 RNAse digestion followed by TLC (Fig. 6B3). As expected, formation of a residue in a A context was detected in the 15-nt fragment, when the pre-tRNA^{Tyr} was incubated with His$_6$–Pus7p (Fig. 6B3). However, because the 15-nt T1 RNase fragment contained three UA dinucleotides, a second experiment was performed with a pre-tRNA^{Tyr} labeled with [-^{32}P]UTP and incubated in the presence or absence of His$_6$–Pus7p. Again, the 15-nt T1 RNase digestion product was purified by gel electrophoresis. In this case, digestion was performed with P1 nuclease before TLC analysis. P1 nuclease releases 5 NMPs; thus, all the released U residues of the 15-nt T1 digestion product contained a labeled 5 phosphate, except residue U35 located at the 5 extremity of the T1 digestion product (T1 RNase releases products with 5 OH at the 5 extremity). Hence, if the U to conversion in pre-tRNA^{Tyr} was occuring at position 35, no labeled residue was expected to be detected upon digestion of the 15-nt fragment with P1 nuclease. This was indeed the case, demonstrating the specific action of the His$_6$–Pus7 recombinant enzyme at position 35 in the 15-nt fragment (Fig. 6B3). Finally, a time course analysis of formation in pre-tRNA^{Tyr} labeled with [-^{32}P]ATP (Fig. 6C) revealed a plateau, that was reached after about 50 min in the conditions used, and corresponded to a ratio of 1 mole of residue per mole of pre-tRNA^{Tyr}. Altogether, these data demonstrated the specific action of the recombinant His$_6$–Pus7 enzyme at position 35 of pre-tRNA^{Tyr}.

In conclusion, the recombinant His$_6$–Pus7 enzyme has the same specificity in vitro as the Pus7 enzyme in vivo: it modifies specifically U2 snRNA at position 35 (Ma et al. 2003), cytoplasmic tRNAs at position 13, and the pre-tRNA^{Tyr} at position 35. Furthermore, the D256A substitution abolishes all three activities (the verification was also

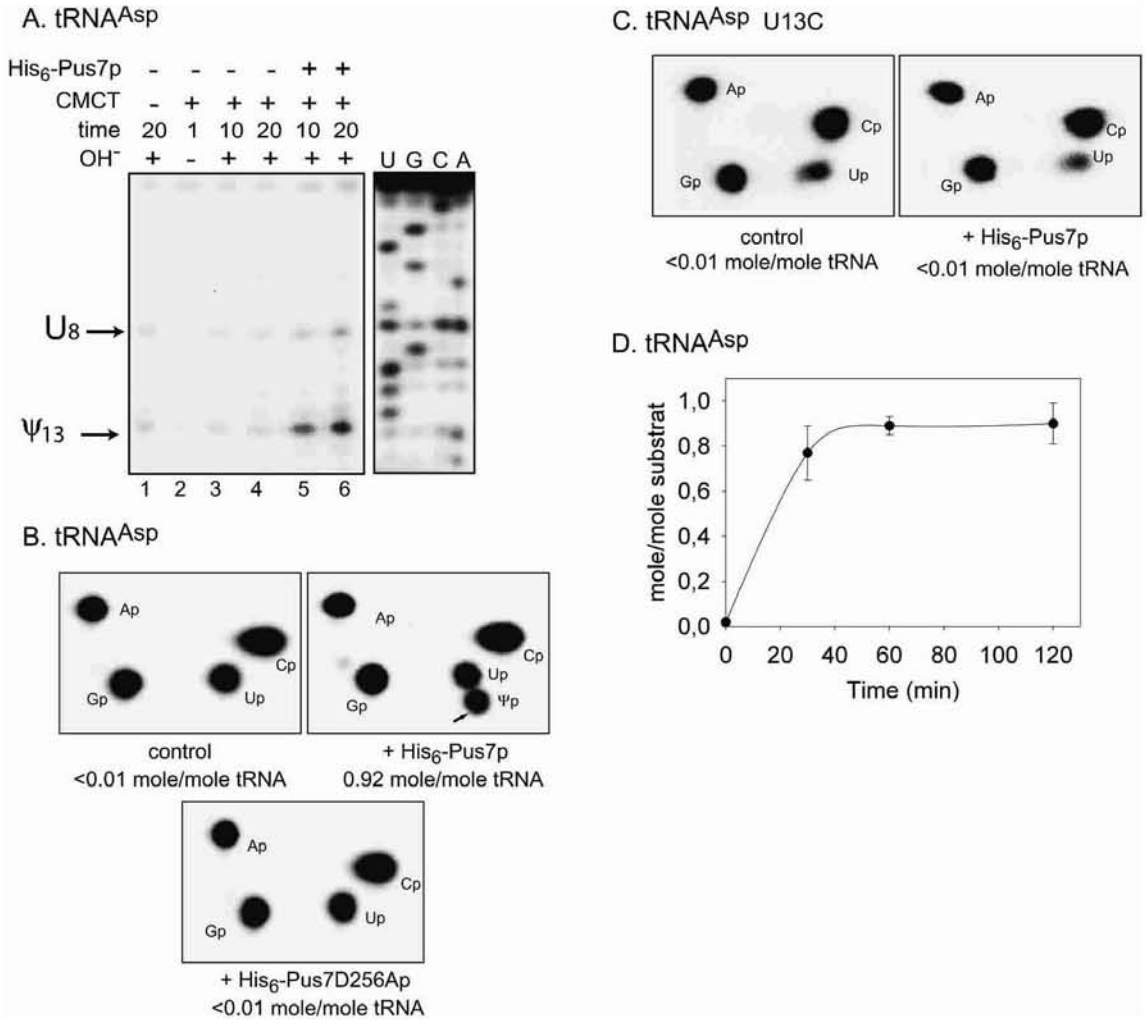


FIGURE 5. Recombinant His$_6$–Pus7p catalyzes specifically U to conversion at position 13 in yeast cytoplasmic tRNA^{Asp}. (A) CMCT/RT mapping of residues formed upon incubation of tRNA^{Asp} with His$_6$–Pus7p. Cold in vitro produced tRNA^{Asp} was incubated in the presence (+) or absence(–) of His$_6$–Pus7p in the conditions described in Materials and Methods. The modified tRNA^{Asp} was analyzed by the CMCT/RT approach (same legends as in Fig. 3). (B) Analysis of 13 formation in tRNA^{Asp} by the nearest neighbor approach. tRNA^{Asp} transcript labeled with [-^{32}P]ATP was incubated with the recombinant His$_6$–Pus7p or the His$_6$–Pus7D256Ap mutant in the conditions described in Materials and Methods. A control incubation was performed in the absence of the recombinant protein. The incubated tRNAs^{Asp} were digested with T2 RNase and the released 3 NMPs were fractionated by TLC, as described in Materials and Methods. The autoradiograms of the TLC plates are shown. The molar ratio of residues formed in tRNAs, as deduced from quantification of the radioactivity of the spots, is given at the bottom of the panels. (C) Same experiment as in (B) with the U13C variant tRNA^{Asp}. (D) Time-course analysis of tRNA^{Asp} modification by the recombinant His$_6$–Pus7p enzyme. tRNA^{Asp} transcript (50–100 fmoles) was incubated with 750 fmoles of His$_6$–Pus7p in the conditions described in Materials and Methods. Aliquot fractions were collected at the indicated times after the beginning of the incubation. For each fraction, RNA was digested with T2 RNase and the released products were analyzed by 2D TLC. Confidence intervals are calculated taking into account the relative radioactivity measured for the MP spot and the Ap, Cp, Gp, and Up spots on 2D TLC.

done for U2 snRNA; data not shown). This observation reinforces the idea that D256 is the catalytic amino acid involved in modification of the three types of substrates.

DISCUSSION

Pus7 is a novel -synthase with multisite and multisubstrate specificity

Here, we demonstrate that like Pus1p, the S. cerevisiae RNA: -synthase Pus7p modifies both U2 snRNA, tRNAs, and pre-tRNA^{Tyr}, and modification in mature tRNAs takes place at position 13 and in the pre-tRNA^{Tyr} at position 35. This observation increases the number of known RNA modification enzymes that act on several RNA substrates, which reinforces the idea that this strategy is rather frequently used to reduce the number of enzymes required for complete post-transcriptional modification of cellular RNAs.

One important question is to know whether the various target sites of Pus7p share some common features (secondary structure or sequence). Clearly, RNA secondary structures around the various modification sites are different (see

Fig. 1). U13 to Ψ13 conversion takes place in a terminal loop of tRNAs, whereas U35 to Ψ35 conversion in U2 snRNA occurs in a single-stranded region linking two stem-loop structures, and the target site in pre-tRNA^{Tyr} is located in a short helix. This diversity suggests that RNA secondary structure is probably not an important parameter in target site recognition by Pus7p. Sequence alignment of all the natural substrates of Pus7p (7 tRNAs, the pre-tRNA^{Tyr}, and U2 snRNA) shows that modification occurs in a highly conserved 7-nt long sequence: Pu(G/C)UNΨAPu (Pu = purine, N = any nucleotide). Importance of this conserved sequence for Pus7p recognition and/or modification is supported by the fact that S. cerevisiae tRNAs with unmodified U13 have this residue located in a sequence that does not fit exactly the consensus sequence. For example, the yeast tRNA^{Ala}(ICG) and tRNA^{Arg} (mcm^{5}UCU), that both have a U to C substitution at position 11 in the conserved sequence, have an unmodified U13. Thus, the presence of a U residue two nucleotides upstream of the modified site seems to be crucial for recognition and/or modification by Pus7p. Inspection of the sequences of all tRNAs from any kind of species, which are known to contain a Ψ13 residue (tRNA database Sprinzl et al. 1998; see also Internet site http://www.unibayreuth.de/departments/biochemie/sprinzl/trna/) revealed a strong evolutionary conservation of the $Pu_9(G/C)_{10}U_{11}N_{12}$ $_{13}A_{14}Pu_{15}$ sequence in tRNAs from all organisms, except archaea. Thus, the tRNA:Ψ13-synthase recognition sequence seems to be highly phylogenetically conserved. Interestingly, the formation of Ψ35 in the A. thaliana pre-tRNA^{Tyr} by plant nuclear extract was also found to require an $U_{33}N_{34}U_{35}A_{36}Pu_{37}$ sequence (Choffat et al. 1988; Szweykowska-Kulinska and Beier 1992; Szweykowska-Kulinska et al. 1995). In this case, apart from this sequence, only the length of the intron seems to be important for the U to Ψ conversion (Choffat et al. 1988; Szweykowska-Kulinska and Beier 1992; Szweykowska-Kulinska et al. 1995). In this work we confirmed that the recombinant Pus1p enzyme is also capable to modify yeast pre-tRNA^{Tyr} at position 35 (data not shown), as previously reported (Motorin et al. 1998). However, in yeast cells, only PUS7 but not PUS1 gene disruption leads to the loss of Ψ35-formation activity in pre-tRNA^{Tyr}. This demonstrates that only Pus7p is implicated in U35 modification in pre-tRNA^{Tyr} in vivo. Therefore, the modification of yeast pre-tRNA^{Tyr} transcript by the recombinant Pus1 enzyme probably represents an in vitro artifact.

It should to be noted that yeast U2 snRNA modification at position 35 seems to depend not only on the consensus sequence described above, but also on a structural motif located downstream from the modification site (Ma et al. 2003).

Another interesting question, which remains to be

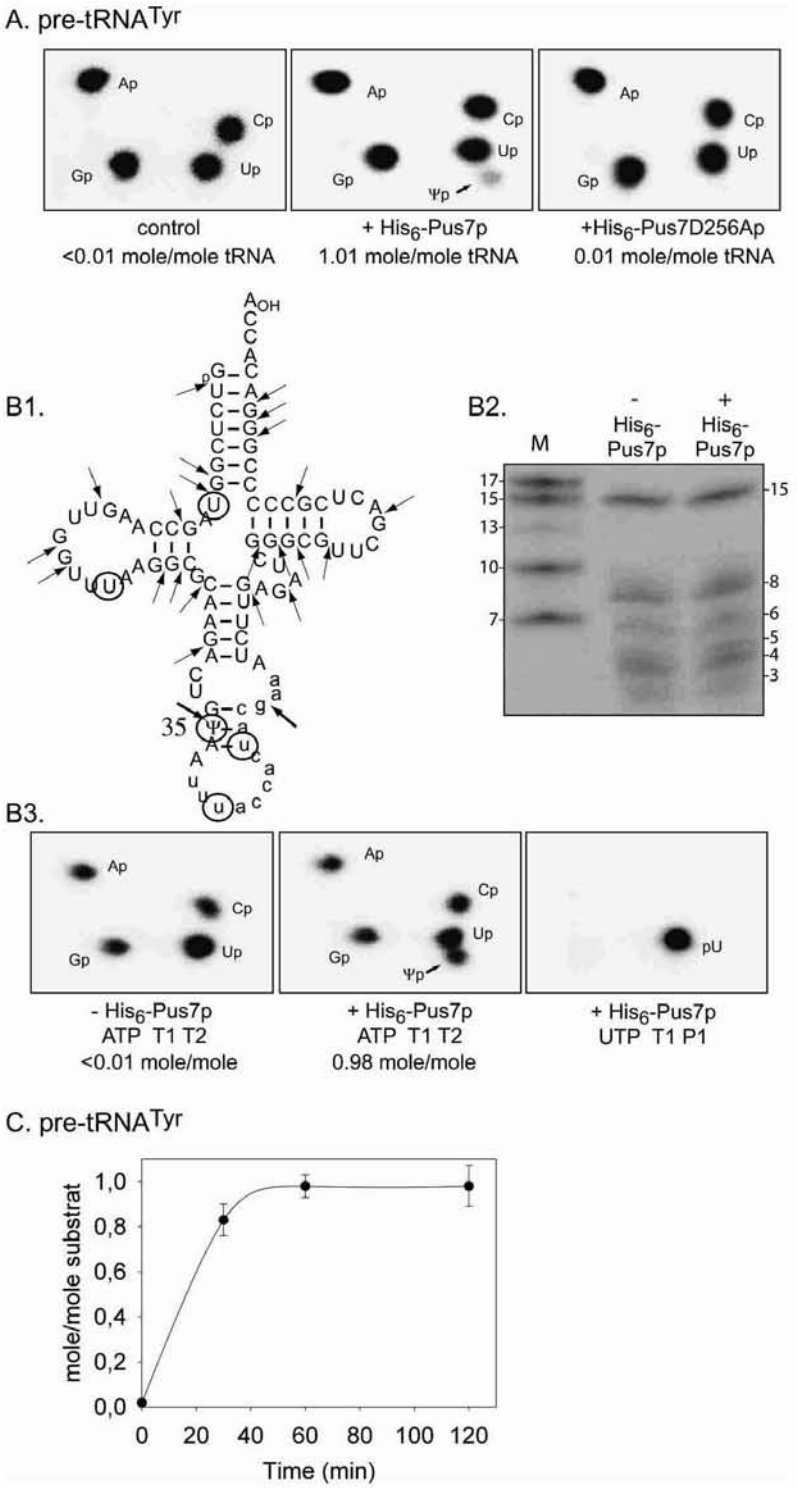


FIGURE 6. (Legend on next page)

solved, is the site(s) of action of Pus7p in terms of cell compartments. The internal modifications in U2 snRNA, as well as, in pre-tRNA^{Tyr}, are expected to occur in the nucleus. In contrast, Ψ13 formation in cytoplasmic tRNAs, which is not intron-dependent, may take place either in the nucleus or in the cytoplasm.

Homologs of the Pus7 RNA: Ψ-synthase are present in all three life kingdoms

The multisite-specific RNA: Ψ-synthase Pus7p belongs to a novel, until recently uncharacterized family of RNA: Ψ-synthases. Based on BLAST search, Pus7p has one homolog protein in all Archaea and Eukarya, whose genome has been completely sequenced. In contrast, Pus7 homologs are present only in 17 out of 77 sequenced bacterial genomes. Archaea and bacteria do not have UsnRNAs; nevertheless, the implication of Pus7p in tRNA modification at position 13 likely explains the presence of a Pus7p homolog in these organisms. Indeed, Ψ13 was detected in one tRNA of E. coli and Thermus thermophilus and several tRNAs of Haloferax volcanii (tRNA database; Sprinzl et al. 1998; see also Internet site http://www.uni-bayreuth.de/departments/biochemie/sprinzl/trna/). Only very recently, the enzyme responsible for Ψ13 formation in E. coli was identified (Kaya and Ofengand 2003). Interestingly, the N-terminal domain of this enzyme presents 45% of similarity with the central region of Pus7p.

In contrast to residue Ψ13, which is rather frequent in living organisms, Ψ35 in tRNA^{Tyr} seems to be an exclusively eukaryotic feature. Implication of Pus7p in pre-tRNA modification in eukaryotes suggests a gain of function of Pus7p in the course of evolution. However, we cannot rule out the possibility that Pus7p homologs also act on other RNA substrates in bacteria and archaea.

The role of the vertebrate homologs of Pus7p will have to be defined. The most probable target sites are positions 13 and 35 in tRNAs, because Ψ34 formation in U2 snRNA (equivalent to Ψ35 in yeast) is dependent upon an H/ACA guide RNA. However, according to recent observation (Zhao et al. 2002), the X. laevis Pus7p homolog may have conserved its capacity to modify U2 snRNA at position 34, because, after depletion of the pugU2–34/44 H/ACA snoRNA, a slow recovery of Ψ34 formation was observed in X. laevis oocytes. In agreement with this observation, we noticed that the U34 pseudouridylation site in X. laevis U2 snRNA fits the consensus sequence required for modification by the Pus7p RNA: Ψ-synthase. The existence of two redundant modification systems acting at position 34 in vertebrate U2 snRNA may be related to the high importance of this pseudouridylation site (Newby and Greenbaum 2001; Newby and Greenbaum 2002). Residue Ψ34 is located in the sequence that base pairs with the intron branch-site sequence. More precisely, Ψ34 participates to the formation of the base pair that precedes the A residue involved in the first step of the splicing reaction. Residue Ψ34 in U2 snRNA was shown to be responsible for a peculiar spatial arrangement of the U2 snRNA/branch-point sequence interaction, that places the reactive A residue in a conformation favorable for the first nucleophilic attack of the splicing reaction (Massenet et al. 1998; Newby and Greenbaum 2001; Newby and Greenbaum 2002).

MATERIALS AND METHODS

Yeast strains

The haploid S. cerevisiae BY4742 strain (denoted WT in this article) and the two isogenic haploid strains carrying a disruption of the PUS7 gene (mat-a and mat-α) were obtained from EUROSCARF collection (Germany). As RNA modification were found to be identical in the two mutant strains, only the WT and mat-a ΔPUS7 strains were used for further experiments. Transformation of yeast S. cerevisiae strains was done with the standard lithium acetate procedure (Adams et al. 1997).

Construction of plasmids for yeast cells complementation

For construction of the p413GalS–PUS7 shuttle vector, the YOR243 (PUS7) ORF was PCR amplified from the genomic DNA

FIGURE 6. Recombinant His$_6$–Pus7p catalyzes specifically the U to Ψ conversion at position 35 in yeast pre-tRNA^{Tyr}. (A) Analysis of Ψ35 formation in pre-tRNA^{Tyr} by the nearest neighbor approach. Yeast pre-tRNA^{Tyr} uniformly labeled with [α-^{32}P]ATP was incubated with the recombinant His$_6$–Pus7p or His$_6$–Pus7D256Ap variant in the conditions described in Materials and Methods. A control experiment was performed in the absence of recombinant protein. Modified and unmodified pre-tRNAs^{Tyr} were digested with T2 RNase and the 3′NMPs were fractionated by TLC. The autoradiograms of the TLC plates are shown. The molar ratio of Ψ residues formed in tRNAs, as deduced from quantification of the radioactivity of the spots, is given at the bottom of the panels. (B) Mapping of the Ψ residue formed in the pre-tRNA^{Tyr} by recombinant His$_6$–Pus7p enzyme. (B1) Cloverleaf representation of the pre-tRNA^{Tyr} with indication of the T1 RNase cleavage sites. The U residues located at the 5′ position of an A residue are circled. (B2) Fractionation by gel electrophoresis of the T1 RNase products of the pre-tRNA^{Tyr} labeled by [α-^{32}P]ATP incorporation. Electrophoresis conditions are given in Materials and Methods. (B3) The 15-nt fragment obtained for a pre-tRNA^{Tyr} incubated in the absence(– His$_6$–Pus7p) or presence(+ His$_6$–Pus7p) of the recombinant enzyme were digested with T2 RNase, and the resulting 3′NMPs were fractionated by 2D TLC . A similar experiment was performed with a pre-tRNA^{Tyr} labeled by incorporation of [α-^{32}P]UTP. In this case, the 15-nt T1 RNase digestion product was hydrolyzed with P1 nuclease and the resulting 5′NMPs were fractionated by 2D TLC. (C) Analysis of time-course formation of Ψ residue in the pre-tRNA^{Tyr} upon incubation with the His$_6$–Pus7p enzyme. pre-tRNA^{Tyr} labeled by incorporation of [α-^{32}P]ATP was incubated with His$_6$–Pus7p enzyme in the conditions described in Materials and Methods. Aliquot fractions were collected at the indicated time after the beginning of the incubation. The pre-tRNA^{Tyr} of each aliquot was digested with T2 RNase. The released 3′NMPs were fractionated by 2D TLC and the radioactivity of each 3′NMP was estimated by measurement with a PhosphorImager. (C) Represents the deduced ratio of Ψ residue moles formed per pre-tRNA^{Tyr} moles as a function of the incubation times. Error bars were calculated as described in Materials and Methods.

of strain BY4742 using an oligonucleotide corresponding to the 5′ end of the ORF (AAACTAGCTAGCatgtctgactcctca) and an oligonucleotide complementary to the 3′ end of the ORF (AAACGC GGATCCttagatattctcctt). Sense and antisense sequences of the YOR243 ORF are shown in lower case. The restriction sites (NheI and BamHI) generated by the 5′- and 3′-oligonucleotides are underlined. The amplified PCR product was subcloned into the SmaI site of the pUC18 cloning vector and entirely sequenced (pUC18–PUS7 plasmid). To build the p413GalS–PUS7 recombinant plasmid the pUC18–PUS7 was cleaved by NheI and BamHI and the fragment containing the YOR243 ORF was inserted, downstream from the GalS promoter, between the XbaI and BamHI sites of the E. coli/S. cerevisiae shuttle vector p413GalS (Mumberg et al. 1995). A point mutation (GAT →GCT) generating the D256A substitution in the Pus7p enzyme was introduced in the cloned ORF by PCR-mediated site-directed mutagenesis using the Quick Change Kit (Stratagene).

Preparation of DNA matrices for tRNA production

The template for in vitro transcription of the S. cerevisiae pre-tRNA^{Tyr} was generated by PCR amplification of the yeast genomic DNA. The sense primer (AAAAATAATACGACTCACTATAgtctc ggtagccaagttt) contained the sequence required for addition of a T7 RNA polymerase promoter upstream of the coding sequence (underlined) and the 17 nucleotides at the 5′ end of the pre-tRNA^{Tyr} (lower case). The antisense primer (tggtgtcccgggggcga) was complementary to the 17 nucleotides at the 3′ end of the pre-tRNA^{Tyr}. The nucleotides corresponding to the first base pair of the pre-tRNA^{Tyr} acceptor stem (C1–G72) were converted into a G1–C72 base pair to increase the yield of the transcription reaction. The amplified PCR product was cloned into the SmaI site of plasmid pUC19. The same approach was used for amplification and cloning of DNA encoding the WT cytoplasmic tRNA^{Arg}(ACG). The plasmids, pTFM–Asp and pTFM–His, used for in vitro transcription of tRNA^{Asp} and tRNA^{His} were generously provided by C. Florentz (IBMC). The U13C variant of the pTFM–Asp plasmid was obtained by PCR-mediated site-directed mutagenesis using the Quick Change Kit (Stratagene). The DNA template for in vitro transcription of the A. thaliana pre-tRNA^{Tyr} was constructed with synthetic oligonucleotides, by matrix-independent filling-in reaction catalyzed by the Pfu DNA polymerase, followed by PCR amplification. First, the central region of the pre-tRNA was produced by the filling-in reaction using partially overlapping oligonucleotides. Then, a second step of amplification was performed using this amplified central region as the matrix and two primers: one bearing a T7 RNA polymerase promoter and the 5′ extremity of the pre-tRNA and an antisense primer complementary to the 3′ extremity of the pre-tRNA^{Tyr}. The resulting PCR product was cloned into the SmaI site of plasmid pUC19. The identity of all the generated recombinant plasmids was verified by DNA sequencing.

Construction of plasmids for protein production and protein production

To generate the pET28b–PUS7 recombinant plasmid the pUC18–PUS7 plasmid was cleaved by NheI and BamHI, and the fragment containing the PUS7 ORF was inserted, downstream from the T7 promoter and the His_6 sequence, between the NheI and BamHI sites of the E. coli expression vector pET28b (Novagen). A point mutation (GAT →GCT) generating the D256A substitution in the Pus7p enzyme was introduced in the ORF of the recombinant pET28 plasmid as described above for the p314GalS–PUS7 plasmid.

The recombinant His_6–Pus1p, WT, and D256A His_6–Pus7p enzymes were produced in E. coli and purified to homogeneity using Ni-NTA agarose column in the previously described conditions (Simos et al. 1996).

In vitro transcription

Prior to in vitro transcription, the pUC19–pre-tRNAs^{Tyr}, pUC19–tRNA^{Arg}, pTFM–Asp, and the variant construct U13C of the pTFM–Asp plasmid were cleaved with the BstNI nuclease; the pTFM–His plasmid was linearized with the BamHI nuclease. In vitro T7 RNA-polymerase transcription with or without incorporation of the appropriate [α-^{32}P]-ribonucleotide triphosphate and purification of the resulting RNA transcripts by electrophoresis on denaturing gel were performed as previously described (Jiang et al. 1997).

Prior to activity tests, in vitro transcribed RNAs were subjected to the following renaturation procedure: 50–100 fmoles (5–10 nM) of ^{32}P-labeled tRNA or pre-tRNA substrate for nearest neighbor approach or 4 pmoles of cold transcript for CMCT/RT analysis were dissolved in 10 µL of 100 mM Tris-HCl, pH 8.0, 100 mM ammonium acetate, 5 mM $MgCl_2$, 2 mM DTT, 0.1 mM EDTA, and incubated 5 min at 80°C, followed by a 20-min incubation at room temperature for transcript renaturation.

Analysis of Ψ residues by the CMCT/RT approach

Preparation of total RNA from yeast strains and CMCT-RT mapping of Ψ residues were performed as described previously (Bakin and Ofengand 1993; Massenet et al. 1999; Ansmant et al. 2000). This CMCT/RT method is based on the alkaline-resistant modification of Ψ residues by the water-soluble carbodiimide CMCT, followed by detection of modification positions by primer extension analysis. The oligonucleotides, used for analysis of yeast UsnRNAs, were described previously (Massenet et al. 1999). The presence of Ψ13 residue in cytoplasmic tRNA^{Asp} (anticodon GUC) and tRNA^{Glu} (anticodon UUC) and of Ψ35 residue in cytoplasmic tRNA^{Tyr} (anticodon GΨA) was tested with oligonucleotides complementary to residues 24–45 of tRNA^{Asp}, 37–58 of tRNA^{Glu} and 58–76 of the tRNA^{Tyr}, respectively.

For activity tests of the recombinant His_6–Pus7 enzyme, incubation of 4 pmoles of transcript was performed for 2 h in the presence of 1.25 pmoles of the recombinant Pus7p in 100 mM Tris-HCl, pH 8.0, 100 mM ammonium acetate, 5 mM $MgCl_2$, 2 mM DTT, 0.1 mM EDTA. After incubation, RNA was extracted by phenol-chloroform and used for Ψ localization using the CMCT/RT approach.

Analysis of Ψ residues by the nearest neighbor approach

The RNA: Ψ-synthase activity at position 13 in yeast cytoplasmic tRNA^{Asp} (anticodon GUC) and tRNA^{His} (anticodon GUG) and at

position 35 in the yeast pre-tRNA^{Tyr} (anticodon GUA, containing the intron) and the A. thaliana pre-tRNA^{Tyr} (anticodon GUA, containing the intron) were tested by the nearest neighbor approach. Transcripts corresponding to these tRNAs were produced by T7 RNA polymerase in the presence [α-^{32}P]ATP. The tRNA: Ψ1-synthase activity was tested on a transcript of the yeast cytoplasmic tRNA^{Arg} (anticodon ACG) produced in the presence of [α-^{32}P]UTP. Cell-free S10 extracts from yeast were prepared as previously (Auxilien et al. 1996). After transcripts renaturation, the reaction was initiated by addition of 10 µL of S10 yeast extract (final concentration of about 0.5–1.0 mg of total protein/mL) or 10 µL of buffer containing 750 fmoles of WT or D256A His$_6$–Pus7 protein. After 1 h incubation at 37°C, the modified RNAs were phenol-extracted, ethanol-precipitated, and digested overnight with T2 RNase (0.01 U/µL), in 50 mM ammonium acetate buffer pH 4.6. The resulting 3′NMPs were fractionated by 2D TLC on cellulose plates (Polygram CEL 400, Macherey-Nagel). An isobutyric acid: ammonia 25%:water (66:1:33/v:v:v) mixture was used for the first dimension, while the second dimension was done in 2-propanol:HCl 37%:water (68:17.6:14.4/v:v:v). Assignment of 3′NMPs was based on previously published maps (Keith 1995). Radioactive spots were quantified on a PhosphoImager instrument (Molecular Dynamics) using the ImageQuant software. Calculations of molar amount of the Ψ per mole of tRNA substrate were done taking into account the nucleotide composition of tRNA transcript. The experimental error of quantification was estimated to be of ±0.05 mole of Ψ residue/mole of tRNA. Error bars on time-course studies of Ψ formation were calculated using the Student's factor for probability 95% and the estimated radioactivities of the ΨMP and NMPs as deduced from Phosphorimager measurement using the ImageQuant software.

Identification of Ψ residues in T1 pre-tRNA^{Tyr} digestion products

Complete cleavage of pre-tRNA^{Tyr} was achieved by incubation with 1 U of T1 RNase for 12 h at 37°C. The resulting oligonucleotides were separated by 20% urea-PAGE, and their position was localized by autoradiography. The band corresponding to the U35-containing oligonucleotide (15-nt) was cut from the gel and was recovered by passive elution. The eluted oligonucleotide was then digested by either T2 RNase or P1 nuclease, and the resulting NMPs were analyzed by 2D TLC as described above. For P1 nuclease, overnight digestion at 37°C was carried out in 50 mM ammonium acetate buffer pH 5.3 in the presence of 0.3 U of enzyme.

ACKNOWLEDGMENTS

We thank C. Florentz (CNRS, Strasbourg, France) for providing the plasmids containing the yeast tRNA^{Asp} and tRNA^{His} genes, and J. Ugolini and C. Mathieu for technical assistance. This work was supported by laboratory funds from the CNRS and the "Ministère de la jeunesse, de l'éducation nationale et de la recherche." I. Behm-Ansmant is a predoctoral fellow from the "Ministère de la jeunesse, de l'éducation nationale et de la recherche."

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Received March 14, 2003; accepted August 15, 2003.

REFERENCES

Adams, A., Gottschling, D.E., Kaiser, C.A., and Stearns, T. 1997. Methods in yeast genetics A Cold Spring Harbor Laboratory course manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Ansmant, I. and Motorin, Y. 2001. Identification of RNA modification enzymes using sequence homology. Mol. Biol. (Mosk.) 35: 248–267.

Ansmant, I., Massenet, S., Grosjean, H., Motorin, Y., and Branlant, C. 2000. Identification of the Saccharomyces cerevisiae RNA: pseudouridine synthase responsible for formation of Ψ(2819) in 21S mitochondrial ribosomal RNA. Nucleic Acids Res. 28: 1941–1946.

Ansmant, I., Motorin, Y., Massenet, S., Grosjean, H., and Branlant, C. 2001. Identification and characterization of the tRNA: Ψ31-synthase (Pus6p) of Saccharomyces cerevisiae J. Biol. Chem. 276: 34934–34940.

Auxilien, S., Crain, P.F., Trewyn, R.W., and Grosjean, H. 1996. Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. J. Mol. Biol. 262: 437–458.

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.

Becker, H.F., Motorin, Y., Planta, R.J., and Grosjean, H. 1997. The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of Ψ55 in both mitochondrial and cytoplasmic tRNAs. Nucleic Acids Res. 25: 4493–4499.

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.

Charette, M. and Gray, M.W. 2000. Pseudouridine in RNA: What, where, how, and why. IUBMB Life 49: 341–351.

Choffat, Y., Suter, B., Behra, R., and Kubli, E. 1988. Pseudouridine modification in the tRNA^{Tyr} anticodon is dependent on the presence, but independent of the size and sequence, of the intron in eukaryotic tRNA^{Tyr} genes. Mol. Cell. Biol. 8: 3332–3337.

Conrad, J., Sun, D., Englund, N., and Ofengand, J. 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 23S ribosomal RNA. J. Biol. Chem. 273: 18562–18566.

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: 751–763.

Decatur, W.A. and Fournier, M.J. 2002. rRNA modifications and ribosome function. Trends Biochem. Sci. 27: 344–351.

Filipowicz, W., Pelczar, P., Pogacic, V., and Dragon, F. 1999. Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification. Acta Biochim. Pol. 46: 377–389.

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.

Gu, X., Ofengand, J., and Santi, D. V. 1994. In vitro methylation of Escherichia coli 16S rRNA by tRNA (m5U54)-methyltransferase. Biochemistry 33: 2255–2261.

Huang, L., Ku, J., Pookanjanatavip, M., Gu, X., Wang, D., Greene, P.J., and Santi, D.V. 1998. Identification of two Escherichia coli pseudouridine synthases that show multisite specificity for 23S RNA. Biochemistry 37: 15951–15957.

Huttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach,

H., Bachellerie, J.P., and Brosius, J. 2001. RNomics: An experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. EMBO J. 20: 2943–2953.

Jiang, H.-Q., Motorin, Y.A., Jin, Y.-X., and Grosjean, H. 1997. Pleiotropic effects of intron removal on base modifications pattern of yeast tRNA ^{Phe}: An in vitro study. Nucleic Acids Res. 25: 2694–2701.

Johnson, P.F. and Abelson, J. 1983. The yeast tRNA ^{Tyr} gene intron is essential for correct modification of its tRNA product. Nature 302: 681–687.

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.

Kaya, Y. and Ofengand, J. 2003. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. RNA 9: 711–721.

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

Kiss, T. 2001. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. EMBO J. 20: 3617–3622.

———. 2002. Small nucleolar RNAs: An abundant group of noncoding RNAs with diverse cellular functions. Cell 109: 145–148.

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.

Lafontaine, D.L., and Tollervey, D. 1998. Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. Trends Biochem. Sci. 23: 383–388.

Lecointe, F., Simos, G., Sauer, A., Hurt, E.C., Motorin, Y., and Grosjean, H. 1998. Characterization of yeast protein Deg1 as pseudouridine synthase (Pus3) catalyzing the formation of Ψ38 and Ψ39 in tRNA anticodon loop. J. Biol. Chem. 273: 1316–1323.

Ma, X., Zhao, X., and Yu, Y.-T. 2003. Pseudouridylation (Psi) of U2 snRNA in S. cerevisiae is catalyzed by an RNA-independent mechanism. EMBO J. 22: 1889–1897.

Massenet, S., Mougin, A., and Branlant, C. 1998. Posttranscriptional modifications in the U Small Nuclear RNAs. In The modification and editing of RNA (eds. H. Grosjean and R. Benne), pp. 201–228. ASM Press, Washington, DC.

Massenet, S., Motorin, Y., Lafontaine, D.L., Hurt, E.C., Grosjean, H., and Branlant, C. 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: 2142–2154.

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.

Mumberg, D., Muller, R., and Funk, M. 1995. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene 156: 119–122.

Newby, M.I., and Greenbaum, N.L. 2001. A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. RNA 7: 833–845.

———. 2002. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. Nat. Struct. Biol. 9: 958–965.

Ni, J., Tien, A.L., and Fournier, MJ. 1997. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA Cell 89: 565–573.

Nurse, K., Wrzesinski, J., Bakin, A., Lane, B.G., and Ofengand, J. 1995. Purification, cloning, and properties of the tRNA Ψ55 synthase from Escherichia coli. RNA 1: 102–112.

Ofengand, J. 2002. Ribosomal RNA pseudouridines and pseudouridine synthases. FEBS Lett. 514: 17–25.

Patton, J.R., and Padgett, R.W. 2003. Caenorhabditis elegans pseudouridine synthase 1 activity in vivo: tRNA is a substrate but not U2 small nuclear RNA. Biochem. J. 372: 595–602.

Simos, G., Tekotte, H., Grosjean, H., Segref, A., Sharma, K., Tollervey, D., and Hurt, E.C. 1996. Nuclear pore proteins are involved in the biogenesis of functional tRNA. EMBO J. 15: 2270–2284.

Sprinzl, S., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. 1998. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 26: 148–153.

Szweykowska-Kulinska, Z. and Beier, H. 1992. Sequence and structure requirements for the biosynthesis of pseudouridine (Ψ35) in plant pre-tRNA ^{Tyr}. EMBO J. 11: 1907–1912.

Szweykowska-Kulinska, Z., Krajewski, J., and Wypijewski, K. 1995. Mutations of Arabidopsis thaliana pre-tRNA ^{Tyr} affecting pseudouridylation of U35. Biochim. Biophys. Acta 1264: 87–92.

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.

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 tRNAPhe. RNA 1: 437–448.

Zhao, X., Li, Z.H., Terns, R.M., Terns, M.P., and Yu, Y.T. 2002. An H/ACA guide RNA directs U2 pseudouridylation at two different sites in the branchpoint recognition region in Xenopus oocytes. RNA 8: 1515–1525.

PARTIE 2

ETUDE DES REQUIS EN SEQUENCE ET STRUCTURE POUR QU'UN ARN SOIT SUBSTRAT DE L'ENZYME PUS7P DE *S. CEREVISIAE*

1. Contexte

Comme nous l'avons mentionné dans le chapitre d'introduction, un moyen de générer un grand nombre de modifications à des positions bien définies dans les ARN est l'utilisation des systèmes de modification dépendants d'ARN guides. Une autre possibilité qui est utilisée aussi bien chez les bactéries que chez la levure *S. cerevisiae*, est l'existence d'enzymes ayant une spécificité de substrat large comme c'est le cas pour Pus1p et Pus7p.

Il était donc intéressant de définir quels étaient les requis en séquence et en structure pour qu'un ARN soit substrat de Pus7p. Dans le cas de Pus1p déjà étudiée, la reconnaissance du substrat semble dépendre à la fois de l'identité du résidu situé en 3' du nucléotide à modifier, avec une préférence pour un résidu G, ainsi que de paramètres de structure 2D (Motorin et al., Massenet et al., résultats non publiés).

Comme nous l'avons vu dans l'article précédent, Pus7p a la particularité de former des résidus Ψ dans une séquence qui est conservée dans tous ses substrats. Nous avons donc étudié l'importance de cette séquence conservée dans 3 types de substrat (l'ARNt^{Asp}, le pré-ARN^{Tyr} et le snRNA U2). Pour chaque substrat, nous avons analysé l'effet de la substitution de chacun des résidus de la séquence conservée par 1 à 3 autres résidus. La séquence conservée comporte 7 résidus, nous avons produit au total 17 mutants ponctuels pour chaque substrat. Nous avons étudié la capacité de chacun de ces ARN mutés a être modifié par la protéine Pus7p recombinante. Pour cela, nous avons utilisé la méthode du plus proche voisin que nous avons décrite dans le chapitre précédent.

Bien que les domaines modifiés dans les ARN substrats de Pus7 ne présentent aucun motif commun au niveau de leur structure secondaire (Fig. X), nous ne pouvions pas exclure un effet possible de la structure secondaire de l'ARN sur l'activité de Pus7p. En effet, certaines ARN:Ψ-synthases nécessitent l'intégrité de la structure 3D pour être active (groupe II), tandis que d'autres reconnaissent un élément de structure local de l'ARN (groupe I) (Grosjean et al., 1996) Nous avons donc introduit un certain nombre de mutations dans l'ARNt^{Asp}, le pré-ARN^{Tyr} et le snRNA U2 en vue de modifier leur structure 2D ou 3D. Nous avons aussi produit des fragments de ces ARN et nous avons testé l'activité de Pus7p sur ces différents ARN mutés et tronqués.

2. Démarche et Résultats

Dans un premier temps, nous nous sommes donc intéressés à la séquence conservée (Pu G/c U N Ψ A Pu) au sein de laquelle la modification a lieu. Pour cela, comme mentionné ci-dessus, nous avons produit 17 mutants pour chacun des types d'ARN étudiés. Nous avons utilisé l'ARNt^{Asp}(GUC) portant un résidu Ψ13 comme représentant des ARNt cytoplasmiques. Le pré-ARN^{Tyr}(GΨA) a été utilisé dans son intégralité pour la production de mutants. Par contre, vu sa taille, seul le domaine 5' terminal du snRNA U2 de *S. cerevisiae*

qui contient les séquences fonctionnelles (139 nucléotides 5' terminaux) a été utilisé pour la réalisation des mutagenèses. L'enzyme recombinante produite chez *E. coli* a été utilisée pour les tests d'activité, dans les conditions déjà établies dans l'article 2.

L'efficacité de modification des ARN mutés a été testée après 90 min d'incubation avec l'enzyme recombinante. Le taux de résidu Ψ formé dans les ARN a été estimé par la méthode du plus proche voisin, en utilisant des ARN radiomarqués par incorporation d'[α-^{32}P]ATP , ce choix s'imposait puisque le résidu en 3' du nucléotide modifié était toujours un A.

Les données obtenues montrent, que dans tous les substrats, l'identité des résidus en position -2 et +1 par rapport au nucléotide modifié a un rôle majeur dans l'efficacité de la réaction. Les mutations aux autres positions conservées ont des effets nettement plus modérés. D'après nos données, la position ayant le plus d'effet parmi ces 4 autres positions est la position +2, et ceci quelque soit le substrat. L'importance de l'identité des résidus en positions -3 et -4 semble dépendre de l'identité du substrat. Enfin, dans tous les substrats, l'identité du résidu en position -1 a peu d'importance.

Le fait que nos données révèlent une importance cruciale de la présence d'un résidu U en position -2 pouvait expliquer l'absence de résidu Ψ13 dans l'ARNt^{Ala}(ICG) cytoplasmique de *S. cerevisiae*. En effet, cet ARNt possède une séquence, autour du résidu U13, qui est conforme à la séquence consensus des substrats de Pus7p à l'exception de la substitution du résidu U-2 par un C. En convertissant ce résidu C en un U dans un l'ARNt^{Ala}(ICG) produit par transcription *in vitro*, nous avons pu montrer qu'un transcrit portant la substitution C11U est modifié par Pus7p alors qu'un transcrit sauvage ne l'est pas. C'était une démonstration supplémentaire de l'importance de l'identité de ce résidu.

La séquence autour du résidu U13 de certains ARNt mitochondriaux est conforme à la séquence conservée dans les substrats de Pus7p. Néanmoins, ils ne sont pas modifiés à cette position. En utilisant un transcrit de l'ARNt^{Trp}(UCA) mitochondrial répondant au consensus, nous avons montré que Pus7p le modifie *in vitro*. L'absence de modification *in cellulo* est donc sans doute liée à la compartimentation cellulaire de Pus7p. Elle ne doit pas entrer dans les mitochondries.

Dans un second temps, nous avons abordé l'étude de l'influence des structures secondaires et tertiaires des ARN substrats de Pus7p sur le taux de modification. Nous avons, là encore, utilisé l'ARNt^{Asp}(GUC) comme modèle d'ARNt comportant un résidu Ψ13. Nous avons successivement éliminé chaque structure tige boucle ou hélice de cet ARN et nous avons testé l'effet produit sur le taux de résidu Ψ formé. Les effets de ces délétions se sont avérés modérés. L'effet le plus marqué est celui observé pour la délétion du bras TΨC.

Nous avons ensuite testé l'effet de la disruption de l'hélice de la boucle D dans laquelle a lieu la modification. Dans ce cas, un effet négatif plus marqué sur le taux de résidu Ψ est observé. L'augmentation de la stabilité de l'hélice de la tige boucle de l'anticodon

entraîne une légère augmentation du taux de résidu Ψ formé. Globalement l'idée qui ressort de cette étude est qu'un renforcement de la stabilité de la structure secondaire de l'ARNt conduit à une légère augmentation du taux de modification. Néanmoins, la structure 3D de l'ARNt n'est pas requise pour la modification. Seule, la présence d'éléments de structure secondaire semble requise pour que les ARN substrats de Pus7p soient modifiés efficacement.

Nous avons ensuite essayé de définir quelle était la taille minimale pour qu'un ARN soit substrat de Pus7p. Pour l'ARNt^{Asp}(GUC) et le pré-ARN^{Tyr}(GΨA), les structures tige boucle D et de l'anticodon contenant l'intron, sont respectivement suffisantes pour obtenir un taux de conversion de l'ordre de 35% par rapport à celui obtenu pour les ARNt entiers.
Comme le résidu Ψ35 dans l'ARN^{Tyr} mature est aussi présent dans une séquence en accord avec la séquence conservée dans tous les substrats de Pus7p, à l'exception d'une purine en position +3, nous avons testé si l'enzyme recombinante pouvait modifier cet ARNt. Cela s'est avéré être le cas. Néanmoins, lorsque nous avons comparé l'efficacité de modification de Pus7p sur l'ARN^{Tyr} mature et sur le pré-ARN^{Tyr} *in vitro*, nous avons constaté que la quantité d'enzyme requise pour la modification de l'ARN^{Tyr} mature est 10 fois plus élevée que celle requise pour la modification de l'ARN^{Tyr} précurseur. Ces données confirment qu'en plus de l'identité de certains résidus, la structuration de la séquence reconnue par Pus7p favoriserait la modification de l'ARNt.

Enfin, en ce qui concerne le snRNA U2, nous avons pu montrer que bien que la conversion du résidu U35 en résidu Ψ ait lieu dans une séquence en simple brin au sein de cet ARNt, la présence d'au moins une structure tige boucle est nécessaire pour qu'il y ait modification. La séquence en simple brin seule, n'est pas modifiée.

Au travers ce travail, qui est présenté sous la forme d'un manuscrit à soumettre à JBC, nous avons donc montré l'existence de requis à la fois en séquence et en structure 2D pour qu'un ARN soit substrat de l'enzyme Pus7p de *S. cerevisiae*.

Publication n°3

RNA sequence and 2D structure features required for efficient activity of the *Saccharomyces cerevisiae* RNA:Ψ-synthase Pus7p

Alan Urban, Christiane Branlant and Yuri Motorin

Préparé pour soumission à JBC

Publication n°3

Version #33, 08/09/2008 Prepared for JBC

RNA sequence and 2D structure features required for efficient activity of the *Saccharomyces cerevisiae* RNA:Ψ-synthase Pus7p

Alan Urban, Isabelle Behm-Ansmant, Christiane Branlant and Yuri Motorin*

[1]Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP Nancy I, Faculté des Sciences, BP 239, 54506 Vandoeuvre-les-Nancy Cedex, France

Running title: RNA specificity of the yeast Pus7p

Keywords: Pseudouridine, Pus7p, tRNA, pre-tRNA^{Tyr}, U2 snRNA, RNA:Ψ-synthase

*Corresponding author :
Yuri MOTORIN
Laboratoire Maturation des ARN et Enzymologie Moléculaire
UMR 7567 CNRS-UHP Nancy I
Faculté des Sciences, BP 239
54506 Vandoeuvre-les-Nancy Cedex, FRANCE
Tel. 33.3.83.68.43.16
FAX. 33.3.83.68.43.07

ABSTRACT

The Pus7p RNA:pseudouridine (Ψ) synthase Pus7p of *Saccharomyces cerevisiae* is a multisite-specific enzyme that is able to modify U13 in several yeast tRNAs, U35 in the pre-tRNA^{Tyr} (GΨA), U35 in U2 snRNA and U50 in 5S rRNA. Pus7p belongs to the universally conserved TruD-like family of RNA:Ψ-synthases. While several RNA substrates for Pus7p have been identified, their specificities of recognition and modification have not been studied. However, conservation of a 8 nt–long sequence surrounding the modified U residue in all natural Pus7p substrates suggests importance of these nucleotides for Pus7p recognition and/or catalysis. Using site-directed mutagenesis we designed a set of RNA variants derived from the yeast tRNA^{Asp}(GUC), pre-tRNA^{Tyr}(GΨA) and U2 snRNA and tested their ability to be modified by Pus7p *in vitro*. We demonstrated that the highly conserved U-2 and A+1 residues are crucial identity elements for efficient modification by Pus7p. Nucleotide substitutions at other surrounding positions (-4, -3, +2, +3) have more moderate effects. Surprisingly, the identity of the nucleotide 5' to the pseudourydilated residue (position -1) is not important for modification. Alteration of tRNA 3D structure had no detectable effect on Pus7 activity at position 13. However, our results suggest that at least the presence of one stem-loop structure including or aside the nucleotide to be modified is required to get modification.

INTRODUCTION

RNA:pseudouridine(Ψ)-synthases catalyze the post-transcriptional U to Ψ conversion in RNAs. Extensive studies during the last ten years allowed the discovery and characterization of almost the complete set of bacterial and yeast enzymes responsible for this modification in tRNAs, rRNAs and snRNAs.

All known RNA:Ψ-synthases share a set of conserved amino acid sequence motifs and can be grouped in 6 distinct families based on the degree of amino acid sequence homology : families TruA, TruB, RluA, RsuA, TruD and PsuX, respectively (Koonin, 1996; Kaya & Ofengand, 2003; Roovers et al., 2006). The *S. cerevisiae* RNA:Ψ-synthase Pus7p belongs to the recently described TruD-related family of proteins whose members are conserved in all kingdoms of life, but have no apparent sequence homology with the other RNA:Ψ-synthases. However, strong homologies are found at the level of the 3D structure of the catalytic site (Kaya & Ofengand, 2003). Pus7p was initially fished out in a high-throughput screening aiming to characterize U2 snRNA-specific pseudouridylation activity. By this approach, Ψ35-forming activity in U2 snRNA was attributed to the YOR243c ORF, designed as Pus7p (Ma et al., 2003). Later on, by analysis of *in vivo* RNA modification defects resulting from deletion of the *PUS7* gene and by *in vitro* tests using the recombinant protein, we identified other targets of the Pus7p enzyme (U13 in several yeast cytoplasmic tRNAs and U35 in the intron containing pre-tRNA^{Tyr}(GΨA)) (Behm-Ansmant et al., 2003). Recently, the activity of Pus7p at position 50 in 5S rRNA was also demonstrated (Decatur & Schnare, 2008).

The *E. coli* homologue of the yeast Pus7p (TruD) (Kaya & Ofengand, 2003) was also identified. As the tRNA^{Tyr}(GΨA) is produced without intron in bacteria and that no snRNAs are present in these organisms, the activity of the bacterial TruD enzyme seems to be

restricted to U13 conversion in tRNAs (Kaya & Ofengand, 2003). TruD-like proteins are also present in archaea. However, their activity and specificity have not been studied up to now.

The rules governing RNA substrate recognition by RNA:Ψ-synthases and RNA:modification enzymes in general, have only been elucidated in a few cases (for review, (Hur et al., 2006). Some of the characterized RNA:Ψ-synthases have a rather strict substrate specificity and are only able to modify one position in only one type of cellular RNA, as for instance tRNAs or even in one cellular RNA, like 5S RNA. This is the case for several of the charaterized yeast RNA:Ψ-synthases: Pus5p modifies a unique position in the mitochondrial 21S rRNA (Ansmant et al., 2000), Pus6p modifies only position 31 in both cytoplasmic and mitochondrial tRNAs (Ansmant et al., 2001), Pus8p and Pus9p modify position 32 in the cytoplasmic and mitochondrial tRNAs, respectively. Finally, one well studied example of RNA:Ψ-synthase acting at a unique position in tRNAs, is Pus4p, that converts the universally conserved U55 residue into a Ψ residue in all elongator tRNAs (Becker et al., 1997a). It was demonstrated that this strict substrate specificity depends on the universally conserved G_{53}UUCNANNC_{60} sequence of the TΨC loop and the peculiar 3D structure of this loop, which is stabilized by a reverse Hoogsteen interaction between residues U54 and A58. Mutations, introduced at any of the conserved positions, completely abolish or considerably reduce the modification efficiency (Becker et al., 1997b). Very similar results were also obtained for its bacterial TruB counterpart. No deep analysis on the basis for RNA specificity of the Pus5p, Pus6p, Pus8p and Pus9p enzymes was performed, so that the sequence and structure requierement for activity of these enzymes are not known.

In contrast to the yeast RNA:Ψ-synthase Pus4p, Pus5p, Pus6p, Pus8p and Pus9p, both Pus1p and Pus7p have the pecularity to act on different types of substrates: 8 positions in tRNAs and position 44 in U2 snRNA for Pus1p (Motorin et al., 1998; Massenet et al., 1999; Behm-Ansmant et al., 2006) and position 13 in cytoplasmic tRNAs, 35 in the pre-

ARNt^{Tyr}(GΨA), 35 in U2 snRNA and 50 in 5S rRNA for Pus7p (Behm-Ansmant et al., 2003; Ma et al., 2003; Decatur & Schnare, 2008). One can ask the question of how an RNA:Ψ-synthase can act with a high degree of specificity in such a series of different substrates. These RNAs have different size, different 2D and 3D structures. The local 2D structure of the modified regions share no obvious commun features.

Previous unpublished studies on the Pus1p RNA specificity ((Motorin et al., 1998; Massenet et al., 1999); Massenet et al., unpublished data ; Motorin et al., unpublished data) revealed the importance of the identity of the residue 3' to the targeted U residue with a preference for a purine residue, more precisely a G residue. The presence of the targeted U residue in a double-stranded region and in the close vicinity of a flexible segment was found to increase Pus1p activity.

RNA regions modified by Pus7p show no obvious commun features at the level of the 2D structure, U13 in cytoplasmic tRNAs, U35 in the pre-ARNt^{Tyr}(GΨA) and 50 in 5S rRNA are located in an helical region according to the 2D structures proposed for these RNAs (Fig. 1) (Swerdlow & Guthrie, 1984) and the degree of stability of these structures are not high. However, U35 in U2 snRNA is located in a single-stranded RNA region (Branlant et al., 1982; Ares, 1986) (Fig. 1A). Inspection of the Pus7p RNA targets in *S. cerevisiae* revealed a high level of conservation of an 8-nt long sequence containing the targeted U residue (Behm-Ansmant et al., 2003). The newly identified 5S rRNA target also contains this conserved sequence (Fig. 1B).

Site directed mutagenesis of the *A. thaliana* pre-ARNt^{Tyr}(GΨA), which also contain an intron and is pseudourydilated at position 35 had revealed the importance in this RNA of the $U_{33}N_{34}U_{35}A_{36}Pu_{37}$ sequence for efficient Ψ35 formation in a lupin seed extract (Szweykowska-Kulinska & Beier, 1992). To test whether this sequence requierement is a general feature of the eukaryal Pus7-like enzyme and if the conserved residues have an equal

importance in different types of RNA substrates, we produced variants of each of the yeast Pus7p substrates (tRNA^{Asp}(GUC), pre-tRNA^{Tyr}(GΨA) and U2 snRNA) with base substitutions at position -4, -3, -2, -1, +1, +2 or +3 and the activity of the recombinant Pus7p protein on these variant RNAs were compared. As the cytoplasmic tRNA^{Ala}(ICG), which contains the conserved 8-nt long sequence around position 13, except for a C to U substitution at position -2, is not pseudouridylated at position 13 *in vivo*, we tested whether it can be modified *in vitro* after C12U substitution. Several mitochondrial tRNAs also have a sequence around the U13 residue which fits to the consensus sequence of the Pus7p substrates. Therefore, we also tested the *in vitro* activity of recombinant Pus7p on these tRNAs. Finally, we produced several variants of the tRNA^{Asp}(GUC), pre-tRNA^{Tyr}(GΨA) and U2 snRNA substrates of Pus7p that were designed in order to modify their 2D and 3D structures. Then, we tested the activity of the recombinant Pus7p enzyme on these RNA variants. Altogether, the data presented here demonstrate the strong importance of two residues of the highly conserved sequence containing the targeted U residue for 3 types of Pus7 RNA substrate tested and a modulator effect on the activity of the other conserved or semi-conserved residues that depend on the identity of the RNA. Furthermore, the presence of some stable RNA 2D structure elements was found to reinforce the activity, while a highly compact 3D structure may block the action of Pus7p.

RESULTS

Test for the importance of the conserved sequence found in Pus7p substrates on Pus7p activity

The sequence alignment of the identified *S. cerevisiae* Pus7p substrates (Fig. 1B) revealed the strict conservation of a U residue at positions -2 and a A residue at position +1 as refered to the modified residue, a G or a C residue is found at position -3, a purine residue at both positions -4 and +2 and a pyrimidine residue at position +3. In contrast, all possible residues are found at position -1, suggesting that the identity of this residue is not important for activity. To test for the influence of these conserved residues on Pus7p activity, a set of tRNAs^{Asp}(GUC), pre-tRNA^{Tyr}(GΨA) and 5'-terminal regions of U2 snRNA variants were produced. The 5' terminal region of U2 snRNA (positions 1 to 139), which contains all the functional elements of U2 snRNA, was previously shown to be sufficient for Pus7p activity (Ma et al., 2003). Versions of these 3 RNAs with any nucleotides at positions -4, -3, -2, +1 and +2 were produced. Variant RNAs with a pyrimidine to purine substitution at position +3 were also generated. In addition, one variant with a base substitution at position -1 was used as a control.

The recombinant Pus7 purified protein was produced as previously described and its activity on the variant and WT RNAs were tested in the presence of a large excess of RNA substrate (1 µM of RNA for 1 nM of enzyme) as previously described (Behm-Ansmant et al., 2003)and in conditions where the enzyme was in large excess (about 1 fM of labeled RNA alone for 1 nM of enzyme). The level of RNA modification by Pus7p was mesured by the nearest neighbourg approach after a long incubation period (90 min). This long incubation time was selected based on the observation that complete modification of each of the WT RNAs was obtained after 30 min of incubation in the same conditions. If some of the conserved residues in the RNA substrate were important for Pus7p activity, after their

mutation we expected to detect a decreased level of Ψ formation after 90 min of incubation, as previously found for Ψ35 formation in the *A. thaliana* pre-tRNA^{Tyr}(GΨA) incubated in a wheat germ extract (Szweykowska-Kulinska & Beier, 1992). For each variant RNA, the identity of the [α-^{32}P]NTP used for transcription was defined according to the identity of the residue at position +1 as refered to the modified residue. After incubation with the enzyme, the RNAs were digested with T2 RNase and the release 3'-phosphate mononucleotides (3'-NMPs) were fractionnated by thin layer chromatography. The yield of U to Ψ conversion was evaluated by measurement of the radioactivity in the fractionated 3'-NMPs, as previously described (Behm-Ansmant et al., 2003). Quantification of Ψ residue could not be as accurate in U2 snRNA as in tRNA^{Asp}(GUC) and pre-tRNA^{Tyr}(GΨA), because of the high level of UA dinucleotides in this RNA. Therefore only Ψ/U ratios were calculated. They gave a rough estimate of the modification efficiency. To verify that the observed Ψ residue formed in the WT RNAs corresponded to the expected U to Ψ conversions, control experiments were performed with variant U13C tRNA^{Asp}(GUC), U35C pre-tRNA^{Tyr}(GΨA) variant and U35C snRNA U2 5'-terminal region transcripts. As expected, no significant Ψ residue formation was detected after a 90-min long incubation of these RNAs with Pus7p. These results confirmed the strict specificity of Pus7p for a single position within each of these substrates and we tried to define the needed determinants for this specificity.

Residues at positions -2 and +1 play a crucial role in the yield of U to Ψ conversion by Pus7p

Three independent series of *in vitro* modifications of the WT and variants tRNA^{Asp}(GUC) and pre-tRNA^{Tyr}(GΨA) by recombinant Pus7p were made using an excess of RNA as compared to Pus7p. For each of the tested substrates, the mean value of the levels of

U to Ψ conversion found in the three experiments is given in Figure 2A, together with the standard deviation. The autoradiograms of the 2D thin layer chromatographies obtained in one of these three series of experiments are shown as exemples in Figure 1 in supplementary data. In parallel, three other independent series of *in vitro* modifications of the WT and variants $tRNA^{Asp}$(GUC) and pre-$tRNA^{Tyr}$(GΨA) were performed, using an excess of enzyme compared to RNA. For each of the tested substrates, the mean value of the levels of U to Ψ conversion found in the three experiments is given in Figure 2B. A major observation is that the yield of Ψ formation strongly depended on the positions of the base substitutions. Secondly, nearly identical results were obtained when using an excess of substrate or an excess of enzyme (Fig. 2). This might reflect the occurrence of structural heterogeneity of some of the variant RNAs with only one part of the molecules being able to be modified. As base substitutions in the $tRNA^{Asp}$(GUC) were expected to modify the tRNA 2D structure, we tested whether the presence of a compensatory mutations at position 24 in the D-stem could limit the strong decrease of Ψ formation found for base substitutions at position 11 (position -2 in the Pus7p conserved sequence). As shown in Figure 2A, no increase was detected upon restoration of a N11-N24 Watson-Crick base pair in D-stem. Therefore modification of the sequence is responsible of the effect. As the levels of modification after base substitutions at both positions -2 and +1 in the two substrates were nearly identical to those formed after U13C substitution in $tRNA^{Asp}$(GUC) and U35C substitution in the pre-$tRNA^{Tyr}$(GΨA), we concluded that base substitutions at positions -2 an +1 in both $tRNA^{Asp}$(GUC) and pre-$tRNA^{Tyr}$(GΨA) almost completely abolished U to Ψ conversion. The same strong negative effect of base substitutions at these two positions was found in U2 snRNA (Fig. 3 in supplementary data).

Hence, residues U-2 and A+1 are probably required to achieve one step of the overall catalytic process in any of the Pus7p substrate. These data are in perfect agreement with the

strict conservation U-2 and A+1 in all Pus7p substrates (Fig. 1B), and with previous data obtained upon incubation of the *A. thaliania* pre-tRNA^{Tyr}(GΨA) with a wheat germ extract (Szweykowska-Kulinska & Beier, 1992).

Base substitutions at the four other positions, which are less strictly conserved (R_{-4}, G/c_{-3}, R_{+2}, Y_{+3}), had less dramatic effects on the yield of U to Ψ conversion in both the three substrates tested, (Fig. 2A and Fig. 1A, B and C in supplementary data). Mutations at positions -3 and -4 had differential effects in tRNA^{Asp}(GUC) and pre-tRNA^{Tyr}(GΨA) and this may be linked to the RNA 2D structure. Indeed, base substitutions at positions -3 and -4 in tRNA^{Asp}(GUC) that do not affect the RNA 2D structure, were found to have a limited effect on the yield of Ψ13 formation. In constrast, base substitutions at position -4 in the pre-tRNA^{Tyr}(GΨA), which is expected to destabilize the RNA 2D structure (Swerdlow & Guthrie, 1984) (Fig. 1A), had a stronger negative effect on the level of modification. Interestingly, substitution of residue A+2 had a marked negative effect in both the tRNA^{Asp}(GUC) and pre-tRNA^{Tyr}(GΨA). In addition, in both substrates the A to U substitution had the lowest negative effect while, the A to C substitution had the greatest negative effect (Fig. 2A). Surprisingly, both A or G residues are found at this position in Pus7p substrates. As expected mutations at position -1, where any of the four possible nucleotides is found in Pus7p RNA substrates, had no effect on the yield of modification in both the tRNA^{Asp}(GUC) and pre-tRNA^{Tyr}(GΨA) (Fig. 2A). This was also the case for the pyrimidine to purine substitution at position +3 (Fig. 2A).

Presence of a U residue at position -2 is required for Pus7p activity

Only three of the yeast cytoplasmic tRNAs, which have a U residue at position 13 are not substrates of Pus7p (Fig. 3A). Interestingly, these three tRNAs (tRNA^{Ala}(IGC) and two isoacceptors tRNA^{Arg}(mcm^{5}UCU)) carry a C instead of a U residue at position 11 (position -2 in the consensus sequence of Pus7p substrates). Based on the above data, we

postulated that a C11U mutation in these tRNAs will convert them into pus7p substrates. To test this hypothesis, we used the $tRNA^{Ala}$(IGC) as a model. The activity of the recombinant Pus7p was tested both on a WT $tRNA^{Ala}$(IGC) transcript and on a transcript with the C11U substitution. As illustrated in Figure 3B, whereas an unsignificant level of Ψ formation was detected in the WT RNA, the C11U variant was almost completely modified (U to Ψ conversion of 95±3%). These data demonstrated the requierement of a U residue at position -2 for Pus7p activity.

Recombinant Pus7p can modify mitochondrial tRNAs with U and A residues at positions 11 and 14, respectively.

Interestingly, in spite that no Ψ13 residue was detected in mitochondrial tRNAs (Sprinzl & Vassilenko, 2005), the sequence found for several of them around residue U13 fits to the consensus sequence of Pus7p RNA substrates (Fig. 3A). We also noticed that the mitochondrial $tRNA^{Tyr}$(GΨA) contain a sequence around position 35 that fits to this consensus (Fig. 3A). Absence of U13 and U35 pseudouridylation in mitochondria could therefore be due, either to the inability of Pus7p to act on mitochondrial tRNAs, because of differences in 2D structures as compared to cytoplasmic tRNAs, or simply due to the absence of Pus7p in mitochondria. To clarify this point, we tested the *in vitro* activity of Pus7p on one of the mitochondrial tRNAs having an U at position 13 surrounded by the sequence which is conserved in Pus7p substrates, namely, $tRNA^{Trp}$(UCA). A significant level of *in vitro* modification of this tRNA by Pus7p was obtained (77±2%), showing that the absence of Ψ13 modification in mitochondrial tRNAs is probably due to the inability of Pus7p to be imported into mitochondria. This hypothesis is in agreement with the very low probability for a mitochondrial localization of Pus7p, which is found when using various computer programs

(PSORT I and II, SubLoc, ESLPred and YPLS) developed for the prediction of protein subcellular localization.

Efficiency of Ψ13 formation in tRNA^{Asp}(GUC) is decreased upon alteration of the tRNA 2D structure

In the tRNA^{Asp}(GUC) residues -4 (A9), +1 (A14) and +2 (A15) as refered to U13 are involved in the establishment of the RNA 3D structure, A9 is involved in the A9-A23-U12 triple-base formation, A14 forms the conserved U8-A14 interaction, and A15 forms the semi-conserved purine-pyrimidine Levitt base pair A15-U48) (Quigley & Rich, 1976). However, as we shown above that the same requierement for the presence of an A+1 residue (A14 in the tRNA^{Asp}(GUC)) is found for the tRNA^{Asp}(GUC), pre-tRNA^{Tyr}(GΨA) and U2 snRNA, we postulated that the observed negatif effect of the A14 substitution in tRNA^{Asp}(GUC) on Pus7p activity, was not due to the disruption of the tRNA 3D structure. An additional argument for the limited requierement of the tRNA 3D structure is our observation that substitution of residue A9 in tRNA^{Asp}(GUC) (position -4 in the Pus7p consensus sequence), had a very limited effect on Pus7p activity. To verify that tRNA 3D structure is not needed for Pus7p activity and to test for a possible effect of the tRNA 2D structure on this activity, we produced a large series of tRNA^{Asp}(GUC) variants.

The 3D structure was disrupted in most of these variants, especially, in the variant ΔTΨ-SL (Fig. 4A), the TΨ-stem loops which interacts with the D-loop was deleted. Variants ΔAC-SL and ΔAA-SL (Fig. 4A) were missing the anticodon stem-loop and acceptor stem, respectively. In MC3 and MC4 the anticodon was either slightly stabilized (MC3) or slightly destabilized (MC4). For each of these RNAs, 3 independent series of *in vitro* modification experiments were performed in the presence of an excess of RNA as refered to the enzyme (1000/1) (Fig. 4A).

We already showed above, that point mutations that punctual destabilizations of the D-stem have limited effect on the Pus7p activity (Fig. 4A). To go one step further in the destabilization of this stem, the sequence of its 3'-strand was completely mutated in variant OD-SL (Fig. 4A) (Note that, an alternative helix may be formed but it does not include the targeted U13 residue). Finally, two minisubstrates corresponding to the stem-loop structure D of $tRNA^{Asp}$(GUC) and to the anticodon stem-loop containing the intron of the pre-$tRNA^{Tyr}$(GΨA), respectively were produced (Fig. 4B).

Altogether, the results of three independent series of experiements on the 8 RNA variants produced (Fig. 4) confirmed that the tRNA 3D structure is not strictly required for Pus7p activity. Whereas the elimination of the acceptor stem (ΔAA-SL) had a limited effect on the activity, elimination of one of the stem-loop structures other than the D stem-loop (ΔAC-SL, ΔTΨ-SL) or the destabilization of the anticodon stem loop (MC4) slightly decreased the yield of modification. However, the effect was not dramatic. A stronger negative effect was observed when the conformation of D stem-loop where modification occurs was altered (variant level of Ψ residue formed divided by a factor of 2) and OD-SL a quite lower level of modification was obtained for the two mini-substrates (28 and 32%) (Fig. 4B). We noticed a slight increased of Ψ formation upon stabilization of the anticodon loop. These data suggested that at least one stem-loop structure, containing the residue to be modified, is requiered for Pus7p activity. In addition, the presence of other stable stem-loops, which may facilitate the overall folding of the tRNA, probably also favor Pus7p activity.

Ψ35 can be formed in the mature $tRNA^{Tyr}$(GΨA) *in vitro*

Alignment of the sequence containing Ψ35 in the mature *S. cerevisiae* $tRNA^{Tyr}$(GΨA) with the sequences containing Ψ13 in Pus7 tRNA substrates, reveals that $tRNA^{Tyr}$(GΨA) almost fits perfectly to the consensus sequence of Pus7p RNA substrates. The

only difference is the substitution of the pyrimidine at position +3 by an A residue. As shown above, substitution of this pyrimidine at position +3 by an A residue in both $tRNA^{Asp}$(GUC) and pre-$tRNA^{Tyr}$(GΨA) had no marked influence on the yield of U to Ψ conversion by Pus7p *in vitro*. Therefore, we postulated that the recombinant Pus7p should be able to modify the mature $tRNA^{Tyr}$ (GΨA) *in vitro*. To verify this hypothesis, modification experiments were performed in parallel on the precursor and the mature $tRNA^{Tyr}$(GΨA), using an excess of RNA substrate. The yields of modification of the two RNA transcripts were tested after incubation with either 1 nM, 10 nM or 1µM of recombinant enzyme. Interestingly, whereas 81% of U35 was converted into Ψ35 in the pre-$tRNA^{Tyr}$(GΨA) in the presence of 10 nM of protein, no significant Ψ35 formation was detected in the mature RNA in these conditions (Fig. 5A). However, when the protein concentration was increased to 1 µM, 63% of the mature tRNA was modified. The modification did occur at position 35, since it was not observed after U35C substitution (Fig. 5B).

At this stage, we could ask the question why $tRNA^{Asp}$(GUC) with an A residue at position +3, as referred to U13, was modified at a yield of 76% in the presence of 10 nM of enzyme, whereas, $tRNA^{Tyr}$(GΨA) with the same sequence surrounding U35 was not modified. One possible explanation could be the highly constraint 3D structure of the anticodon loop containing Ψ35 $tRNA^{Tyr}$(GΨA). Indeed, residue U33 at position -2 in this loop is known to be involved in a peculiar ribose-phosphate backbone folding (U33-turn) (Quigley & Rich, 1976). The stability of this U33 turn has been attributed to the formation of three interactions involving U33: one hydrogen bond formed by the 2'-OH of residue U33 and the N-7 of residue A35, a second hydrogen bound formed between the N3-H of residue U33 and the phosphate of residue A36, and the stacking of residue U33 on residue A35. Therefore, residue U33 which is the important U-2 residue in the conserved sequence of Pus7p substrates might have an inappropriate conformation in the anticodon loop of the

mature $tRNA^{Tyr}$ (GΨA). The other residue essential for Pus7 activity (position +1 in the Pus7 conserved sequence, A36 in mature $tRNA^{Tyr}$ (GΨA)) might also be poorly accessible in the anticodon loop. If this was the case, insertions of residues between positions 31 and 32 in the loops were expected to abolish the 3D structure constraints and therefore to reinforce Pus7p activity. We indeed observed 61% and 58% of U35Ψ conversion in the presence of 10 nM of Pus7p, when we respectively inserted one or two A between positions 31 and 32 (Fig. 5C). In contrast, the A38U substitution in the loop that restored the complete Pus7p consensus sequence did not increased the yield of modification (Fig. 5B). These data strongly suggested that the 3D structure of the mature $tRNA^{Tyr}$(GΨA) anticodon loop strongly limits its modification by Pus7p.

Presence of 2D structure motifs are required for efficient modification of U2 snRNA *in vitro*

In the 5'-terminal region of yeast U2 snRNA, residue U35 is flanked by two stem-loop structures I an IIa. Stem-loop IIa is immediately followed by a third stem-loop structure IIb and a pseudoknot structure can be formed between the terminal stem-loop IIa and the single-stranded segment located downstream from stem-loop IIb (Fig. 6 upper part). To test for the importance of the U2 snRNA 2D structure on Pus7p activity, we first eliminated the sequence downstream from position 85 in order to disrupt the pseudoknot structure (variant U2:1-85). Then, the stem-loop structures I and IIb were individually deleted (variants U2:1-85 ΔI and U2:1-85 ΔIIb) or deleted together (variants U2:1-85 ΔIΔIIb) and all stem-loop structures were eliminated in the RNA variant U2:28-47. Elimination of the pseudoknot structure slightly decreased the yield of Ψ35 formation. In agreement with previous data (Ma et al., 2003), elimination of stem-loop structure I had a limited effect on the yield of Ψ35 formation, whereas deletion of stem-loop IIb alone or together with stem-loop I decreased the

level of modification by a factor of about 2. In addition, the RNA variant U2:28-47 without any double-stranded region was not modified by Pus7p.

Therefore, altogether the data obtained on both U2 snRNA and the $tRNA^{Asp}$(GUC) strongly suggested that the presence of stable 2D structure motifs increases the level of *in vitro* activity of Pus7p.

DISCUSSION

Here we made a deep characterization of the RNA sequence and structure requierements needed for modification of three distinct RNA substrates of the yeast RNA:Ψ-synthase Pus7p. The data obtained are as much important as Pus7p is presently the RNA:Ψ-synthase with the largest number of identified RNA substrates (tRNA, pre-tRNA^{Tyr}(GΨA), U2 snRNA and 5S rRNA). This large RNA specificity is likely explained by our observation of a strong importance of a limited number of punctual RNA determinants and only a modulator effect of the RNA 2D structure.

Punctual sequence determinants are required for Pus7p activity on any of its substrates

Sequence alignments of all yeast Pus7p substrates reveals that the targeted U residue is located in an 8-long nt sequence with three strictly conserved residues (the targeted U residue, U at position -2 and A at position +1), the other residues being semi- or partially conserved, except nucleotide at position -1. Here, we demonstrate that the three strictly conserved residues are absolutely required for Pus7p activity on three of its substrates (tRNA^{Asp}(GUC), pre-tRNA^{Tyr}(GΨA) and U2 snRNA). Both purine-purine or pyrimidine-pyrimidine replacements at positions -2 and +1, respectively abolish the activity. Therefore we show that the $U_{33}N_{34}U_{35}A_{36}$ sequence previously found to be required for Pus7p activity in plant pre-tRNA^{Tyr}(GΨA) is needed ifor all substrates. These residues may be necessary for stable anchoring of the RNA onto the enzyme to achieve some structural transition, or for catalytic activity. Residue at position +2 in the consensus sequence is important for the yield of modification, but is not essential as are residues U-2 and A+1. The identities of the other partially conserved residues of the 8-nt sequence only have moderate effects on the modification yield and these effects vary from one substrate to the other, probably according

to their RNA 2D structure. Surprisingly, when nucleotides at these positions (-4, -3 or +2) were mutated in $tRNA^{Asp}$(GUC) or in pre-$tRNA^{Tyr}$(GΨA), identical yields of Ψ formation were obtained for a given variant when using an excess of RNA or an excess of enzyme. This suggests that the observed defects were due to an heterogeneous state of the substrate RNA, with only one portion of the molecules being able to bind the enzyme or to undergo the needed conformational changes during U to Ψ conversion.

Remodeling of the tRNA structure is expected to occur for Ψ13 formation by Pus7p

In the L-shaped canonical 3D structure of tRNAs, the essential U-2 and A+1 residues of Pus7p RNA substrates are not accessible. Both of them are involved in base-pair interactions: U11 base pairs with A24 in the D stem, while A14 form a tertiary interaction with U8. Furthermore, they are buried into the molecule due to the tRNA 3D structure. The 3D structure of the complex formed by a tRNA and another RNA modification enzyme acting on a residue of the D loop has been solved, namely, the complex formed by the archaeal tRNA-guanine transglycosylase (TGT) which converts G at the position 15 into archaeosine in many archaeal tRNAs and one of its tRNA substrate (Ishitani et al., 2003). This tRNA carries a U11, U13 and A14 residues like Pus7p substrates. In the 3D structure, it fits into the enzyme in an open conformation called the lambda-tRNA (λ-tRNA) conformation (Ishitani et al., 2003). This conformation is characterized by the opening of the D-stem loop and the interaction of its 3'-strand with the variable loop (Ishitani et al., 2003). By inspection of yeast tRNA sequences we noticed that three non contiguous up to 5 contiguous base pairs can allways be formed between the 3'-strand of the D-stem and the variable loop, so that all Pus7p tRNA substrate may be able to adopt λ-shape structure. In this conformation, the 5' strand of the D-stem containing U11 and U13 would be free of base-pair interaction and these two essential residues as well as residue A14 would become highly accessible as compared to

their position in the L-shaped tRNA structure (Fig. 7). Therefore, one can guess that by interaction with its tRNA substrate, Pus7p like the archaeal TGT enzyme, converts the L-shaped tRNA into a λ-like shaped tRNA so that the three identity elements (U11, U13 and A14) become available allowing U to Ψ13 conversion by the catalytic Asp256 residue.

The intron in pre-tRNA^{Tyr}(GΨA) may avoid 3D structural constraint limiting Pus7p activity in the mature tRNA^{Tyr}(GΨA)

The presence on an intron in pre-tRNA^{Tyr}(GΨA) is a general feature in eukarya as well as the presence of a Ψ35 residue in mature tRNA^{Tyr}(GΨA). Previous work of the Abelson group's (Johnson & Abelson, 1983) demonstrated that the absence of the intervening sequence in the yeast RNA^{Tyr} ochre suppressor gene (SUP6) leads to the absence of Ψ35 modification in the mature tRNA^{Tyr}(GΨA) *in vivo*. Later on, studies performed by RNA microinjection in *X. laevis* oocytes showed that Ψ35 could be formed in both *X. laevis* and *Drosophila* tRNA^{Tyr}(GΨA), in spite of the strong sequence and length differences of the introns of these two pre-tRNAs (Choffat et al., 1988). The conclusion of this study was that the tRNA:Ψ35-synthase needs the intron for tRNA^{Tyr}(GΨA) modification, but that Ψ formation is not dependent on the size and sequence of the intervening sequence. Similar observations were made for the plant pre-tRNA^{Tyr}(GΨA) (Pienkowska et al., 2002): the presence of an intron is required to get modification and only the identity of residues U33, A36 and A37 have a strong influence on the activity (Szweykowska-Kulinska & Beier, 1992; Pienkowska et al., 2002).

Here we show that the recombinant Pus7p can only form Ψ35 residue in the mature tRNA^{Tyr}(GΨA) when used at high concentration (10 nM), which is probably superior to that found *in cellulo*. Indeed, experimental measurements have indicated that Pus7p is moderately

expressed in *S. cerevisiae*. its estimated amount is about 4000 molecules/cell (see http://yeastgfp.ucsf.edu/). We could markedly increase the activity of Pus7p on the mature $tRNA^{Tyr}$(GΨA) by insertion of one or two A residues in the anticodon stem-loop, suggesting that the low activity of Pus7p on the mature $tRNA^{Tyr}$(GΨA) is probably due to the inability of Pus7p to disrupt the 3D structure of the $tRNA^{Tyr}$(GΨA) anticodon loop. The presence of an intron in this anticodon loop in all eukaryal pre-$tRNA^{Tyr}$(GΨA) probably releaves the 3D structural constraints and therefore allows Pus7p activity on this tRNA. Therefore, Pus7p has probably the capability to open the L-shaped structure of the tRNAs, but is not able to open the 3D structure of the anticodon loop of mature $tRNA^{Tyr}$(GΨA). In this case, the solution retained is the course of evolution for modification at position 35 was the insertion of an intron in the pre-$tRNA^{Tyr}$(GΨA).

Stable double-stranded RNA regions are required for Pus7p activity

Whereas tight 3D structure may be an handicap for Pus7p activity as seen for $tRNA^{Tyr}$(GΨA), a substrate without any double-stranded region is not modified by Pus7p as evidenced for the single-stranded U2 snRNA segment 28 to 47 containing the conserved 8-nt long sequence. More generally, all the mutations that we performed in $tRNA^{Asp}$(GUC) or 5'-terminal region of U2 snRNA, that were decreasing the number of stem-loop structures, namely variants ΔTΨ-SL, ΔAC-SL and ΔAA-SL in $tRNA^{Asp}$(GUC) and variants U2:1-85ΔI, U2:1-85 ΔIIb, U2:1-85 ΔIΔIIb and U2:28-47 in U2 snRNA were all modified to a lower extent compared to the WT RNAs. Similarly, minisubstrates corresponding to the D stem-loop of $tRNA^{Asp}$(GUC) or the anticodon stem-loop of the pre-$tRNA^{Tyr}$(GΨA) were modified at a low level. In the case of U2 snRNA, detection of one stem-loop structure IIb had a particularly strong effect on RNA modification. Mutation at position +4 in the pre-$tRNA^{Tyr}$(GΨA) which disrupts the intron structure also had a strong negative effect as well as

disruption of the D stem-loop structure in $tRNA^{Asp}$(GUC). Hence, either Pus7p has numerous specific binding sites for recognition of each of its substrates in their integrity or more likely, it contains binding sites able to recognize one stem-loop structure including or very close to the target residue and site recognizing with a looser specificity. The specific recognition of substrate may be ensured by interaction with the three sequence punctual determinants and one stem-loop structure. The presence of poor specific recognition sites might reinforce the stability of the RNA-protein complex and therefore the yield of the reaction.

Numerous Pus7p orthologs have been identified in all kingdoms of life (Kaya & Ofengand, 2003). Comparison of these protein reveals an increase of size from bacteria to archaea and a greater one from archaea to eukarya. This size increase may be linked to the presence of additional domains needed to accommodate a larger number of RNA substrates in eukarya compared to bacteria (pre-$tRNA^{Tyr}$(GΨA), 5S rRNA and U2 snRNA). Inspection of the aminoacid sequence of Pus7p-like enzymes does not reveal the presence of domain belonging to the double-stranded RNA binding domain family (DSRBD), that might explain a reinforced binding in the presence of several stem-loop structure. We did not detected either strong sequence homologies between the Pus7p-like proteins outside of the catalytic domain. Further structural analysis of Pus7p are required to understand how this enzyme recruits a large variety of substrates and accommodates them into its active site.

Several mitochondrial yeast tRNAs carry the conserved sequence of Pus7p substrates without being modified *in vivo*

Interestingly, 6 out of 17 yeast mitochondrial tRNAs carry a U13 residue with a surrounding sequence that fits perfectly to the consensus sequence of Pus7p substrates. Although we show that Pus7p can modify them *in vitro*, they do not carry a Ψ13 residue *in vivo*. This is probably explained by the absence of import of Pus7p in the mitochondria.

Indeed, expression of GFP-Pus7p fusions revealed that its location is mainly nuclear (Huh et al., 2003). In this cellular compartment it should modify both the pre-tRNA^{Tyr}(GΨA), 5S rRNA and probably some of the mature tRNAs, since only a small amount of GFP-Pus7p protein was detected in the cytoplasm (Huh et al., 2003). Presence of the consensus sequence for Pus7p activity in several mitochondrial tRNAs may be explained by the fact that this sequence preexisted in tRNAs before the appearance of the Pus7 RNA:Ψ-synthase and that it may have additionnal functions. We do not know why cytoplasmic but not mitochondrial tRNAs need Ψ13 residues. Indeed, several cytoplasmic tRNAs carry the Ψ13 modification (10 out of 34 yeast cytoplasmic tRNAs). The functional role of Ψ13 in tRNA may to stabilize the D-stem or to favor interaction with partner proteins. In contrast, stabilization of the antocodon stem-loop by Ψ35 in pre-tRNA^{Tyr}(GΨA) was found to increase the translation efficiency (Auffinger & Westhof, 1998, 2001). A role in the first step of the splicing reaction was also found for residue Ψ35 in U2 snRNA (Yu et al., 1995; Massenet et al., 1999). No role has up to now be identified for Ψ50 in 5S rRNA, which was only recently shown to be the substrate of Pus7p. Hence, there are strong arguments explaining the need for a cellular Pus7p activity (production for functionnal U2 snRNA and pre-tRNA^{Tyr}(GΨA)). Its activity at position 13 in cytoplasmic tRNAs implies, according to our data, that it acquieres the capability to open the tRNA 3D structure. It is therefore unlikely that Pus7p modifies some of the cytoplasmic tRNAs at position 13, only because of the presence of its recognition sequence in these molecules.

MATERIALS AND METHODS

Strains and plasmids

The haploid yeast strain BY4742 *MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0* from the EUROSCARF collection was used as a source of genomic DNA for PCR amplification.

The plasmids used for *in vitro* transcription of $tRNA^{Asp}$(GUC) and its mutants ΔTΨ-SL, MC3 and MC4 were kindly provided by C. Florentz (IBMC, Strasbourg, France). The construct bearing the sequence of *S. cerevisiae* pre-$tRNA^{Tyr}$(GΨA) under the control of T7 promoter was described previously (Behm-Ansmant et al., 2003). Plasmid $pT7U2_{Sc}$, kindly provided by P. Fabrizio, was used for *in vitro* transcription of *S. cerevisiae* U2 snRNA fragment 1-139 (Massenet et al., 1999). Other sequences of tRNA genes used in the study were PCR-amplified and inserted into SmaI site of pUC19, along with T7 promoter sequence at 5' and BstNI (MvaI) restriction site at 3'. Deletion of the 14-nt intron from pre-$tRNA^{Tyr}$(GΨA) construct was performed by PCR, using the following primers: 5'-ATCTTGAGAT CGGGCGTTCGACTCGCCCCCG-3' and 5'-TTACAGTCTTGCGCCTTAAACCA ACTTGG-3'. Amplified linear DNA was self-ligated and used to transform competent *E. coli* cells. DNA matrices for *in vitro* transcription of minisubstrates and U2 snRNA truncated variants were produced by PCR or annealing of complementary DNA oligonucleotides. The forward primers used for PCR amplification generated the T7 RNA polymerase promoter. All variants of $tRNA^{Asp}$ and $tRNA^{Tyr}$ with point mutations were made by site-directed mutagenesis using Quick Change Kit (Stratagene, USA). The sequences of all the generated recombinant plasmids were verified by DNA sequencing.

In vitro T7 RNA polymerase transcription with or without incorporation of the appropriate [α-^{32}P]-NTP and purification of the resulting RNA transcripts by electrophoresis on

denaturing gel were performed as previously described (Jiang et al., 1997; Behm-Ansmant et al., 2003).

Cloning, expression and purification of the recombinant His6-Pus7p

Preparation of the pET28-Pus7 plasmid used for expression of the recombinant *S. cerevisiae* Pus7p was described previously (Behm-Ansmant et al., 2003). N-terminally His6-tagged Pus7p protein was expressed in *E. coli* strain BL21(DE3)-RIL (Stratagene, USA). The transformed cells were grown during 5 h in auto-induction ZYM-5052 media (Studier, 2005) at 37°C followed by 20 h at 20°C. Cells where then harvested and disrupted in buffer A (20 mM Tris–HCl pH 7.5, 1 M NaCl) by a 2 x 2 min sonication at 4°C using a Branson-250 sonicator (50% duty cycle, Output 6). The lysate was cleared by centrifugation (20000g for 30 min) and was directly applied to a column of Ni-Sepharose (Amersham Bioscience, France) charged with Ni^{2+}. The column was first washed with buffer A and than 50 mM imidazole in Buffer A, the protein is eluted using 500 mM imidazole in the same buffer. Fractions containing the Pus7p were pooled, concentrated to 5 mg/ml in a buffer containing 10 mM Tris–HCl pH 7.5, 300 mM NaCl using Vivaspin YM-30 (Vivascience, France) concentration cell and stored at -80°C in 50% glycerol.

In vitro activity tests using nearest neighbour approach

The enzymatic activity of the recombinant Pus7p was assayed in 100 mM Tris-HCl, pH 8.0, 100 mM ammonium acetate, 5 mM $MgCl_2$, 2 mM DTT, 0.1 mM EDTA. All RNA substrates, diluted in 10 µl of the reaction buffer, were renatured by incubation for 5 min at 80°C, followed by a 20 min incubation at room temperature. Two different ratios of Pus7p/RNA were used, in the first series of tests (excess of Pus7p) we used only [^{32}P]-labeled RNA substrate (50-100 fmoles, 1 nM final concentration), in the second series of tests (excess of RNA) we used the same amount of [^{32}P]-labeled RNA but complemented by 100

pmol of non-radiolabeled RNA transcript (final concentration of 1 µM). The reaction was initiated by addition of purified Pus7p diluted in the reaction buffer: 1 µM (conditions of Pus7p excess) or 10 nM (excess of RNA substrate).

After 90 min of incubation at 30°C, the modified RNAs were phenol-extracted, ethanol-precipitated and digested overnight by RNase T2 (0.01 U/µl) in 50 mM ammonium acetate buffer pH 4.6. The resulting 3'-nucleotide monophosphates were fractionated by two-dimensional chromatography on thin layer cellulose plates (Polygram CEL 400, Macherey-Nagel, Germany). An isobutyric acid: ammonia 25%: water (66:1:33/v:v:v) mixture was used for the first dimension, while the second dimension was done in 2-propanol: HCl 37%: water (68:17.6:14.4/v:v:v). Assignment of nucleotides was based on previously published maps (Keith, 1995). Radioactive spots were quantified on a PhosphoImager instrument (Typhoon 9410, Amersham Bioscience) using the ImageQuant software.

ACKNOWLEDGEMENTS

We thank C. Florentz and R. Giegé (CNRS, Strasbourg, France) for providing plasmids containing the yeast $tRNA^{Asp}$ and its variants, P. Fabrizio for providing plasmid $pT7U2_{Sc}$. This work was supported by laboratory funds from the CNRS and the "Ministère de la jeunesse, de l'éducation nationale et de la recherche". A. URBAN was a pre-doctoral fellow from the "Ministère de la jeunesse, de l'éducation nationale et de la recherche".

REFERENCES

Ansmant I, Massenet S, Grosjean H, Motorin Y, Branlant C. 2000. Identification of the Saccharomyces cerevisiae RNA:pseudouridine synthase responsible for formation of psi(2819) in 21S mitochondrial ribosomal RNA. *Nucleic acids research 28*:1941-1946.

Ansmant I, Motorin Y, Massenet S, Grosjean H, Branlant C. 2001. Identification and characterization of the tRNA:Psi 31-synthase (Pus6p) of Saccharomyces cerevisiae. *The Journal of biological chemistry 276*:34934-34940.

Ares M, Jr. 1986. U2 RNA from yeast is unexpectedly large and contains homology to vertebrate U4, U5, and U6 small nuclear RNAs. *Cell 47*:49-59.

Auffinger P, Westhof E. 1998. Hydration of RNA base pairs. *J Biomol Struct Dyn 16*:693-707.

Auffinger P, Westhof E. 2001. An extended structural signature for the tRNA anticodon loop. *RNA (New York, NY 7*:334-341.

Becker HF, Motorin Y, Planta RJ, Grosjean H. 1997a. The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs. *Nucleic acids research 25*:4493-4499.

Becker HF, Motorin Y, Sissler M, Florentz C, Grosjean H. 1997b. Major identity determinants for enzymatic formation of ribothymidine and pseudouridine in the T psi-loop of yeast tRNAs. *J Mol Biol 274*:505-518.

Behm-Ansmant I, Massenet S, Immel F, Patton JR, Motorin Y, Branlant C. 2006. A previously unidentified activity of yeast and mouse RNA:pseudouridine synthases 1 (Pus1p) on tRNAs. *RNA (New York, NY 12*:1583-1593.

Behm-Ansmant I, Urban A, Ma X, Yu YT, Motorin Y, Branlant C. 2003. The Saccharomyces cerevisiae U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. *RNA (New York, NY 9*:1371-1382.

Branlant C, Krol A, Ebel JP, Lazar E, Haendler B, Jacob M. 1982. U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *Embo J 1*:1259-1265.

Choffat Y, Suter B, Behra R, Kubli E. 1988. Pseudouridine modification in the tRNA(Tyr) anticodon is dependent on the presence, but independent of the size and sequence, of the intron in eucaryotic tRNA(Tyr) genes. *Mol Cell Biol 8*:3332-3337.

Decatur WA, Schnare MN. 2008. Different Mechanisms for Pseudouridine Formation in Yeast 5S and 5.8S Ribosomal RNAs. *Mol Cell Biol*.

Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. 2003. Global analysis of protein localization in budding yeast. *Nature 425*:686-691.

Hur S, Stroud RM, Finer-Moore J. 2006. Substrate recognition by RNA 5-methyluridine methyltransferases and pseudouridine synthases: a structural perspective. *The Journal of biological chemistry 281*:38969-38973.

Ishitani R, Nureki O, Nameki N, Okada N, Nishimura S, Yokoyama S. 2003. Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell 113*:383-394.

Jiang HQ, Motorin Y, Jin YX, Grosjean H. 1997. Pleiotropic effects of intron removal on base modification pattern of yeast tRNAPhe: an in vitro study. *Nucleic acids research 25*:2694-2701.

Johnson PF, Abelson J. 1983. The yeast tRNATyr gene intron is essential for correct modification of its tRNA product. *Nature 302*:681-687.

Kaya Y, Ofengand J. 2003. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA (New York, NY 9*:711-721.

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

Koonin EV. 1996. Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic acids research 24*:2411-2415.

Ma X, Zhao X, Yu YT. 2003. Pseudouridylation (Psi) of U2 snRNA in S. cerevisiae is catalyzed by an RNA-independent mechanism. *Embo J 22*:1889-1897.

Massenet S, Motorin Y, Lafontaine DL, Hurt EC, Grosjean H, Branlant C. 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*:2142-2154.

Motorin Y, Keith G, Simon C, Foiret D, Simos G, Hurt E, Grosjean H. 1998. The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA (New York, NY 4*:856-869.

Pienkowska J, Michalowski D, Krzyzosiak WJ, Szweykowska-Kulinska Z. 2002. Pseudouridylation of U(35) in the anticodon of Arabidopsis thaliana pre-tRNA(Tyr) depends on length rather than structure of an intron. *Biochim Biophys Acta 1574*:137-144.

Quigley GJ, Rich A. 1976. Structural domains of transfer RNA molecules. *Science 194*:796-806.

Roovers M, Hale C, Tricot C, Terns MP, Terns RM, Grosjean H, Droogmans L. 2006. Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic acids research 34*:4293-4301.

Sprinzl M, Vassilenko KS. 2005. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic acids research 33*:D139-140.

Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif 41*:207-234.

Swerdlow H, Guthrie C. 1984. Structure of intron-containing tRNA precursors. Analysis of solution conformation using chemical and enzymatic probes. *The Journal of biological chemistry 259*:5197-5207.

Szweykowska-Kulinska Z, Beier H. 1992. Sequence and structure requierements for the biosynthesis of pseudouridine (psi 35) in plant pre-tRNA(Tyr). *Embo J 11*:1907-1912.

Yu YT, Maroney PA, Darzynkiwicz E, Nilsen TW. 1995. U6 snRNA function in nuclear pre-mRNA splicing: a phosphorothioate interference analysis of the U6 phosphate backbone. *RNA (New York, NY 1*:46-54.

FIGURE LEGENDS

Figure 1

Panel A: Sequences and secondary structures of Pus7p RNA substrates. The *S. cerevisiae* $tRNA^{Asp}$(GUC), pre-$tRNA^{Tyr}$(GΨA), the 5'-terminal part of *S. cerevisiae* U2 snRNA and 5S rRNA are drawn according to their proposed 2D structures (Quigley & Rich, 1976; Branlant et al., 1982; Swerdlow & Guthrie, 1984). Only the Ψ residue formed by Pus7p (Massenet et al., 1999; Behm-Ansmant et al., 2003; Decatur & Schnare, 2008) are shown in dark circles, the other post-transcriptional modifications of these RNAs are not represented. The intronic sequence in pre-$tRNA^{Tyr}$(GΨA) is in small characters. The residues at positions -4 to -1 and +1 to +3 as referred to the Ψ residue formed by Pus7p are indicated. In the yeast $tRNA^{Asp}$ and pre-$tRNA^{Tyr}$ transcripts, the A1-U72 and C1-G72 base pairs were both converted into a G1-C72 in the RNA transcripts, in order to increase transcription efficiency.

Panel B: Multiple sequence alignment of the *S. cerevisiae* Pus7p substrates. The alignment of the sequences from positions 22 to 59 in U2 snRNA, 37 to 74 in 5S rRNA, 22 to 45 in pre-$tRNA^{Tyr}$(GΨA), 1 to 34 in cytoplasmic tRNAs is shown. The name of the RNAs are given in the first column with the anticodon sequence shown fot tRNAs (3=mcm5s2U, .= unknown nucleotide, N=unknown modified uridine, I=inosine, 1=5-methoxycarbonylmethyluridine, != 5-carboxymethylaminomethyluridine). The conserved U_{-2} and A_{+1} residues are indicated by black boxes, the Ψ residues formed by Pus7p are in black circles. The 8-nt long conserved sequence present in all Pus7p RNA substrates is shown at the bottom of the panel (R=purine, Y=pyrimidine, N=any nucleotide)

Figure 2 : Effects of point mutations in the $tRNA^{Asp}$(GUC) and pre-$tRNA^{Tyr}$(GΨA) transcripts on their modification by the recombinant Pus7p enzyme in the presence of an excess of substrate (RNA concentration of 1 µM, Pus7p concentration of 10 nM) (Panel A) or when a large excess of enzyme is used compared to the RNA (RNA concentration of 1 pM, Pus7p concentration of 1 µM) (Panel B). After renaturation in the reaction buffer, the RNAs were incubated for 90 min at 30°C in the presence of Pus7p (as described in Materials and Methods). The amount of Ψ residue formed in the transcripts was measured by digestion with the RNA by RNase T2, followed by fractionation of 3'P-mononucleotides by thin layer chromatography as described in Materials and Methods (Fig. 1 in supplementary data). The values given in Panel A and B are the mean values of the mol Ψ amount / mole of $tRNA^{Asp}$(GUC) or pre-$tRNA^{Tyr}$(GΨA) obtained in three independent experiments. The standard deviation values are also given.

For each position of the conserved 8-nt long sequence (noted -4, -3, -2, -1, Ψ, +1, +2, +3) one or three variant RNA were produced. In panel A, the identity of each mutation is indicated on the left of the measured ratio (mol Ψ/ mol RNA). In panel B, the data are represented by dendograms for an easier observation of the strong negative effects of base substitutions at positions -2 and +1. Here also, the number of moles of Ψ residue formed per mole of RNA are the mean value from three independent experiments and the standard deviations are also given.

Figure 3: An U residue at position-2 and an A residue at position +1, as refered to the targeted U13 residue, allow Ψ13 formation.

Panel A: Alignments of cytoplasmic and mitochondrial tRNAs that contain a U residue at position 13 which is not conserted into a Ψ residue. The name of the RNAs are given in the first column and the anticodon sequence is shown (3=mcm5s2U, .= unknown

nucleotide, N=unknown modified uridine, I=inosine, 1=5-methoxycarbonylmethyluridine, != 5-carboxymethylaminomethyluridine). Residues U-2 and A+1, as refered to position 13, are black boxed. The 8-nt long sequence conserved in Pus7p RNA substrates is shown at the bottom of the panel (R=purine, Y=pyrimidine, N=any nucleotide).

Panel B and C: Test for the tRNA:Ψ13-synthase activity of recombinant Pus7p on the WT and C11U variant of the cytoplasmic $tRNA^{Ala}$(IGC) and on the mitochondrial $tRNA^{Trp}$(UCA). Uniformly [α-^{32}P]ATP-labelled RNA transcripts of the WT and mutated yeast cytoplasmic $tRNA^{Ala}$(IGC) (Panel B) and the mitochondrial $tRNA^{Trp}$(UCA) (Panel C) were incubated with 10 ng of recombinant Pus7p for 90 min in the conditions described in Materials and Methods. Control incubations were performed in the absence of recombinant protein. After incubation, the RNA substrates were digested with RNase T2 and the 3'P-MNPs were fractionated by thin layer chromatography. The autoradiograms of the TLC plates are shown. Expected molar ratios of 3'P- MNPs obtained after digestion are indicated on the two control plates. The mean value of the molar yield of Ψ residue formed per mole of tRNA obtained in three distincts experiments is indicated at the bottom of each 2D plate with the standard deviation.

Figure 4 : The yeast $tRNA^{Asp}$(GUC) variants used to study the effect of alterations of 3D and 2D structures of $tRNA^{Asp}$(GUC) on the yield of Ψ13 formation by Pus7p (Panel A). Mini $tRNA^{Asp}$(GUC) and mini pre-$tRNA^{Tyr}$(GΨA) were also used to test Pus7p activity (Panel B). Three distinct series of modification experiments were performed using an excess of RNA and 10 nM of Pus7p. The mean values of the molar yields of Ψ residue formed per mole of RNA and the standard deviations are given below each structure.

Figure 5 : Comparison of the Pus7p activity on the precursor and mature $tRNA^{Tyr}$(GΨA).

Panel A: The sequence around position 35 in the mature $tRNA^{Tyr}$(GΨA) fits to the consensus sequence established for Pus7p substrates.

Panel B: Modification of the $tRNA^{Tyr}$(GΨA) precursor and mature $tRNA^{Tyr}$(GΨA) using different concentration of the recombinant Pus7p for the incubation (1 nM, 10 nM and 1µM). The modified transcripts were analysed as described in the legend to Figure 2. The molar amounts of Ψ residue formed per mole of tRNA in three distinct experiments are indicated at the bottom of the 2D plates. A control assay was done with 10 ng of Pus7p and the $tRNA^{Tyr}$ U35C variant (2D TLC autoradiogram shown on the right of the Panel).

Panel C: Modification of $tRNA^{Tyr}$(GΨA) variants with an extended anticodon loop or a A38U point mutation. One (+A) or two (+AA) adenosine residues were inserted between nucleotide 31 and 32 (indicated by arrow). Nucleotide A38 mutated into a U residue is indicated by an arrow. The results of *in vitro* modification in the presence of 10 nM of Pus7p are given on the bottom of the autoradiograms.

Figure 6: Presence of stem loop structures IIa and IIb are essential for U2 snRNA modification by Pus7p. Upper part: Schematic drawing of the 2D structure of the yeast U2 snRNA fragment (U2:1-139) according to . The stem-loops I, IIa and IIb are indicated and the double arrow shows the interaction responsible for formation of the pseudoknot structure. The molar amounts of Ψ residue formed per mole of RNA after Pus7p-catalyzed modification in RNA U2:1-139 and its truncated variants is indicated at the bottom of 2D plates.

Figure 7

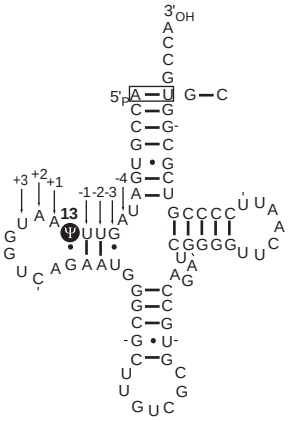
Panel A: Residue Ψ13 and A14 are exposed at the surface of the λ-shaped tRNA observed in the crystal structure of arcTGT-tRNA complex (right structure) as compared to the canonical L-shaped tRNA structure (left structure) (Ishitani et al., 2003). Ribbon 3D structure models and the 2D secondary structure of the λ-shape $tRNA^{Val}$(UAC) are shown. The region corresponding to the 8-nt conserved sequence found in all Pus7p tRNA substrates is coloured in red. Positions of U11, Ψ13 and A14 are shown.

Supplementary Data 1

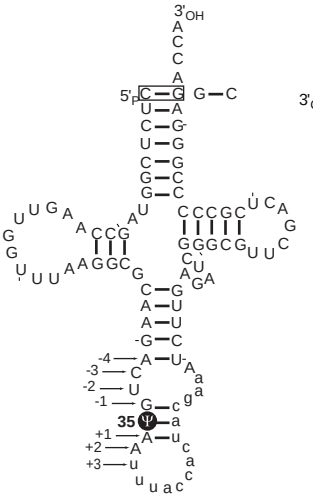
Panel A, B and C: *In vitro* U to Ψ modification of $tRNA^{Asp}$(GUC), pre-$tRNA^{Tyr}$(GΨA) and U2 snRNA and their variants by recombinant Pus7p. Uniformly [α-^{32}P]NTP-labelled RNA transcripts of each WT and variants RNAs were incubated with 10 nM of the recombinant Pus7p in the conditions described in Materials and Methods. Control incubations were performed in the absence of the recombinant protein. The incubated RNA substrates were digested with RNase T2 and the 3'P-MNPs were fractionated by thin layer chromatography. The autoradiograms of the TLC plates are shown. Expected molar ratios of 3'P-MNPs are indicated on the control plates.

Figure 1

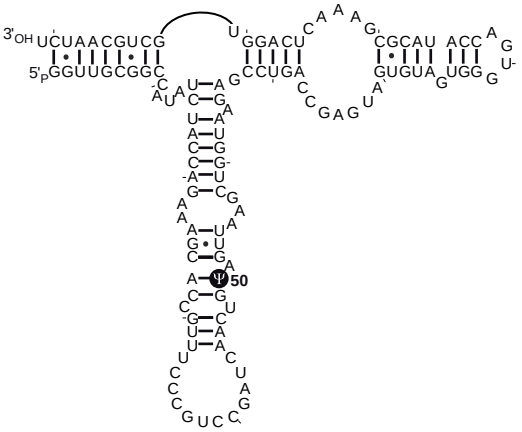
A.



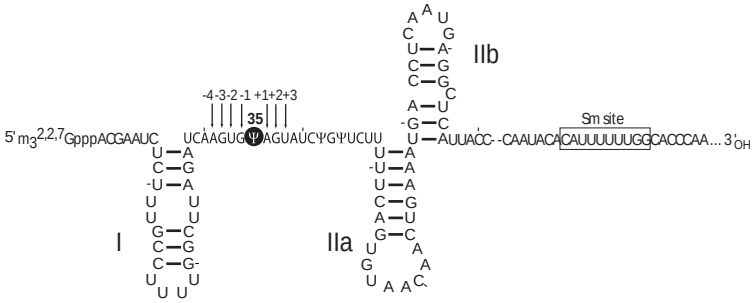
S. cerevisiae
tRNA^{Asp}(GUC)



S. cerevisiae
pre-tRNA^{Tyr}(GΨA)



S. cerevisiae
5S rRNA



S. cerevisiae
U2 snRNA:1-139

B.

U2 snRNA	CUUAGAUCAAGUUGΨAGUAUCPGPUCUUUUCAGUGUAAC
5S rRNA	GUCCGAUCAACUGΨAGUUAAGCUGGUAAGAGCCUGACC
pre-tRNA Tyr(GΨA)	AGGCGCAAGAACUGΨAAuuuaccacuacgaaAPCUUGAG
tRNA Asp(GUC)	-UCCGUGAUAGUUΨAAU--GGUC-AGAAUGGGCGCUUG
tRNA Glu(3UC)	-UCCGAUAUAGUGΨAAC--GGCU-AUCACAU-CACGCU
tRNA Gly(GCC)	-GCGCAAGUGGUUΨAGU--GGU--AAAAUCCAACGUUG
tRNA Gly(.CC)	-GGGCGGUUAGUGΨAGU--GGUU-AUCAUCCCACCCUC
tRNA His(GUG)	GGCCAUCUUAGUAΨAGU--GGUU-AGUACACAACAUUG
tRNA His(GUG)	GGCCAUCUUAGUAΨAGU--GGUU-AGUACACAUCGUUG
tRNA Pro(NGG)	-GGGCGUGUGGUCΨAGU--GGU--AUGAUUCUCGCUUN
tRNA Pro(NGG)	-GGGCGUGUGGUCΨAGU--GGU--AUGAUUCUCGCUUN
tRNA Val(IAC)	-GGUUUCGUGGUCΨAGUC-GGUU-AUGGCAUCUGCUUI
tRNA Val(CAC)	-GUUCCAAUAGUGΨAGC--GGCU-AUCACLUUGCCUUC

Conserved sequence: R G/c U N Ψ A R Y
-4 -3 -2 -1 +1 +2 +3

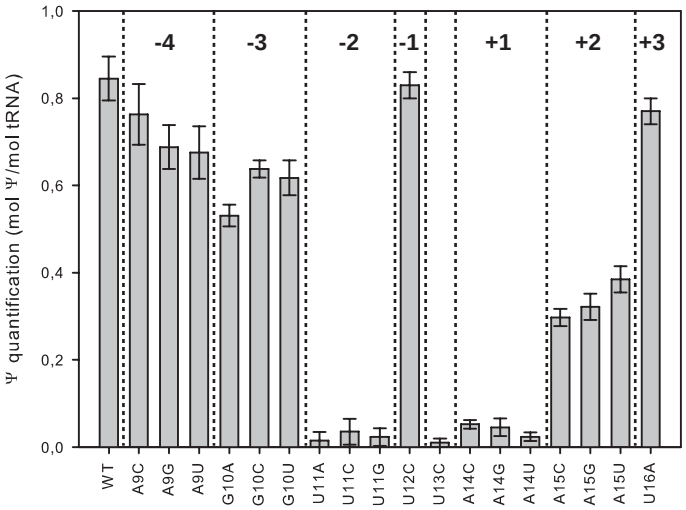
Figure 2

A.

Position	$tRNA^{Asp}$(GUC) Ψ quantification (mol Ψ/mol tRNA)		$pre\text{-}tRNA^{Tyr}$(GΨA) Ψ quantification (mol Ψ/mol pre-tRNA)	
	WT	**0.84 ±0.03**	**WT**	**0.83 ±0.02**
-4	A9C	0.74 ±0.06	A31C	0.46 ±0.03
	A9G	0.65 ±0.05	A31G	0.41 ±0.03
	A9U	0.70 ±0.06	A31U	0.48 ±0.02
-3	G10A	0.56 ±0.03	C32A	0.84 ±0.04
	G10C	0.71 ±0.03	C32G	0.69 ±0.02
	G10U	0.64 ±0.03	C32U	0.74 ±0.03
-2	U11A	0.01 ±0.05	U33A	0.03 ±0.01
	U11C	0.04 ±0.01		
	U11G	0.02 ±0.02	U33C	0.04 ±0.01
	U11A/A24U	0.02 ±0.03		
	U11C/A24G	0.01 ±0.02	U33G	0.06 ±0.01
	U11G/A24C	0.01 ±0.03		
-1	U12C	0.81 ±0.03	U34C	0.81 ±0.03
Ψ	U13C	0.00 ±0.01	U35C	0.03 ±0.01
+1	A14C	0.05 ±0.01	A36C	0.09 ±0.05
	A14G	0.05 ±0.02	A36G	0.03 ±0.01
	A14U	0.02 ±0.01	A36U	0.00 ±0.01
+2	A15C	0.24 ±0.02	A37C	0.22 ±0.01
	A15G	0.38 ±0.03	A37G	0.31 ±0.02
	A15U	0.46 ±0.03	A37U	0.39 ±0.02
+3	U16A	0.79 ±0.03	U38A	0.76 ±0.02

B.

S. cerevisiae $tRNA^{Asp}$(GUC)



S. cerevisiae $pre\text{-}tRNA^{Tyr}$(GΨA)

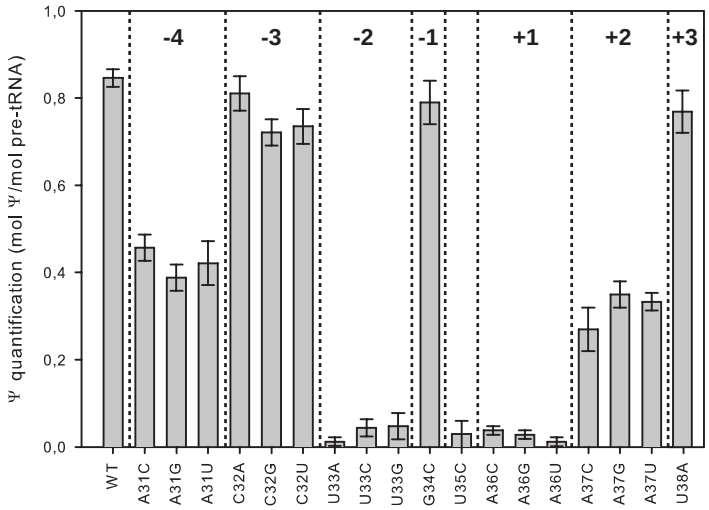


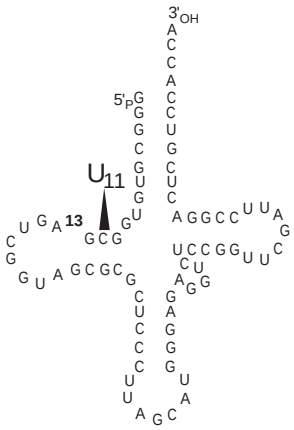
Figure 3

A.

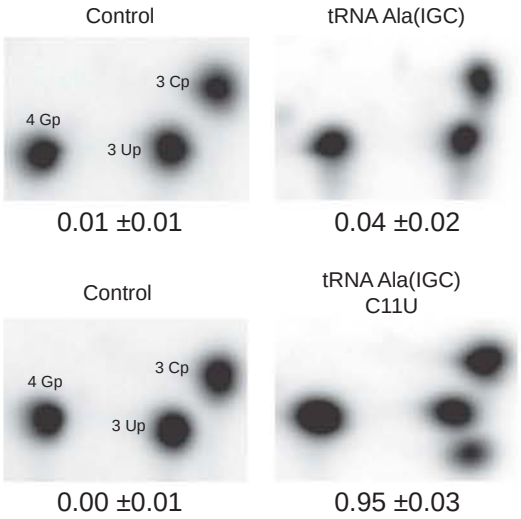
		0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 17a 18 19 20 20a 20b 21 22 23 24 25 26 27 28 29 30 31 32 33 34
Cytoplasmic	tRNA Ala(IGC)	-GGGCGUGUGGCG U GUC-GG U--AGCGCGCUCCCUUI
	tRNA Arg(1CU)	-GCUCGCGUGGCG U AU--GGC--AACGCGUCUGACUC
	tRNA Arg(1CU)	-GCUUGCGUGGCG U AU--GGC--AACGCGUCUGACUC
	tRNA Phe(GAA)	-GCUUUUAUAGCU U GU--GGU--AAAGCGAUAAAUUG
	tRNA Gly(NCC)	-AUAGAUAUAAGU U AUU-GGU--AAACUGGAUGUCUN
Mitochondrial	tRNA Lys(NUU)	-GAGAAUAUUG U U AU--GGU--AAAAC-AGUUGUCU
	tRNA Met(CAU)	-GCUUGUAUAG U U AUU-GG UU-AAAACAUUUGUCUN
	tRNA Arg(ACG)	-AUAUCUUUAA U U AU--GGU--AAAAUAUUAGAAUA
	tRNA Arg(NCU)	-GCUCUCUUAGCU U AU--GGDU-AAAGCAUAAUACUN
	tRNA Thr(UAG)	-GUAAAUAUAA U U AU--GGU--AAAAUGUAUGUUUA
	tRNA Trp(!CA)	-AAGGAUAUAG U U AU--GGU--AAAAC-AGUUGAUU
	tRNA Ini(CAU)	-UGCAAUAUGA G U AUU-GGUU-AACAUUUUAGGGUC

	R	G/c	U	N	Ψ	A	R	Y
Conserved sequence	**-4**	**-3**	**-2**	**-1**		**+1**	**+2**	**+3**

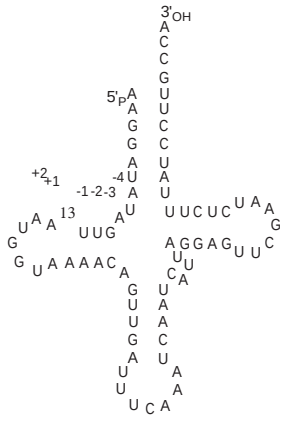
B.



S. cerevisiae
tRNA Ala (IGC)



C.



S. cerevisiae
tRNA Trp (UCA)
mitochondrial

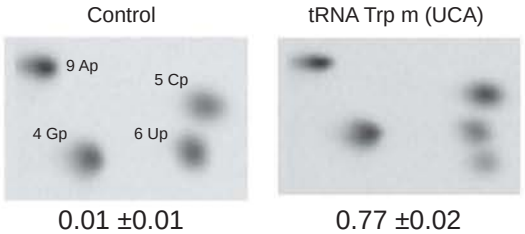
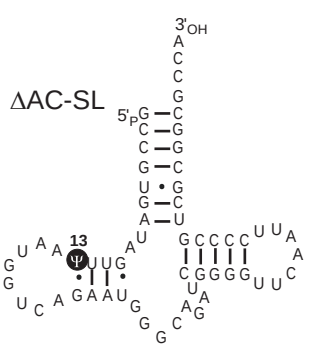
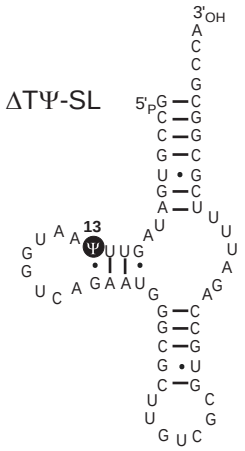


Figure 4

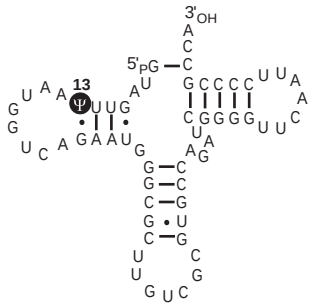
A.



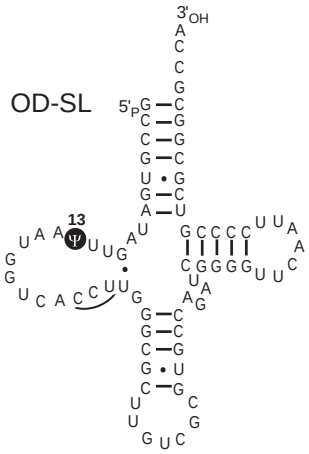
0.73±0.02



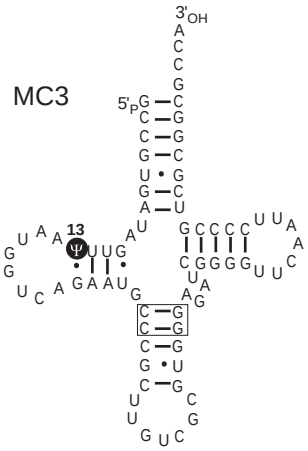
0.65±0.03



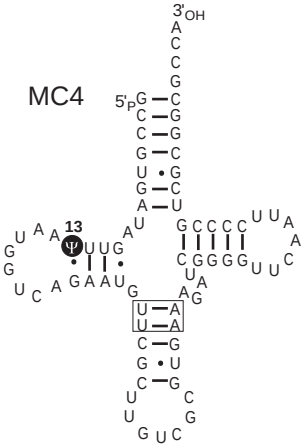
0.81±0.01



0.41±0.02

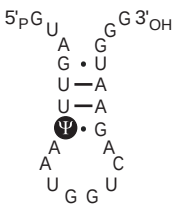


0.91±0.03



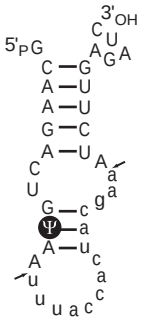
0.75±0.01

B.



S. cerevisiae
mini tRNA^{Asp}(GUC)

0.28±0.02

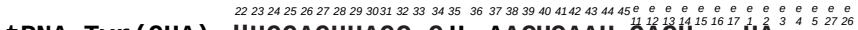


S. cerevisiae
mini pre-tRNA^{Tyr}(GΨA)

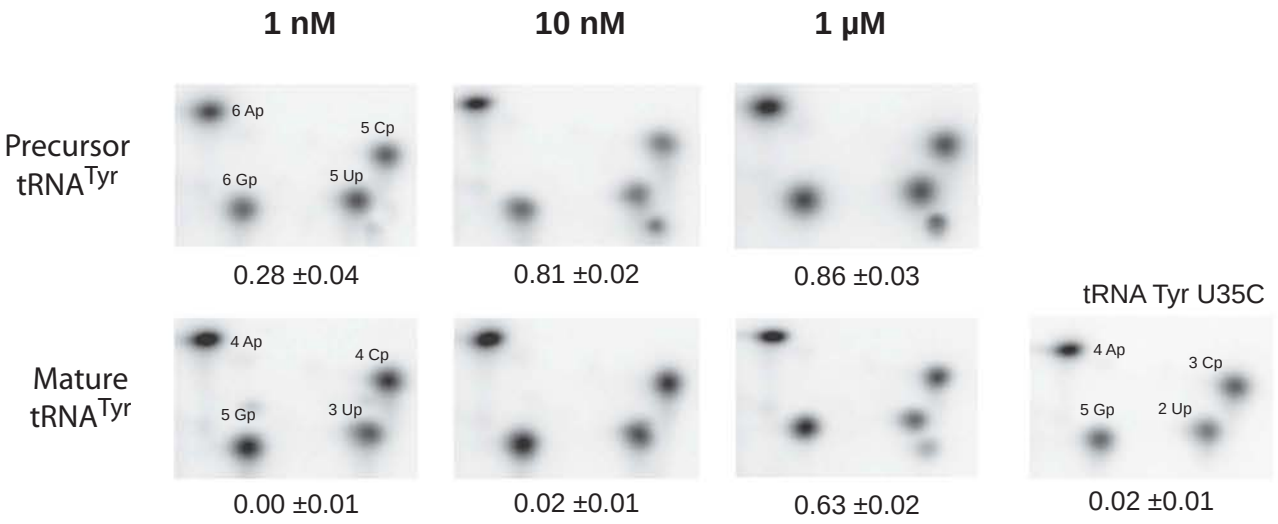
0.32±0.02

Figure 5

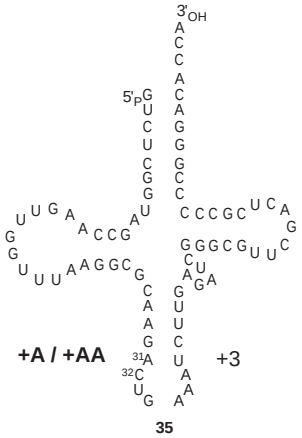
A.



B.



C.



S. cerevisiae tRNA^{Tyr}(GΨA)

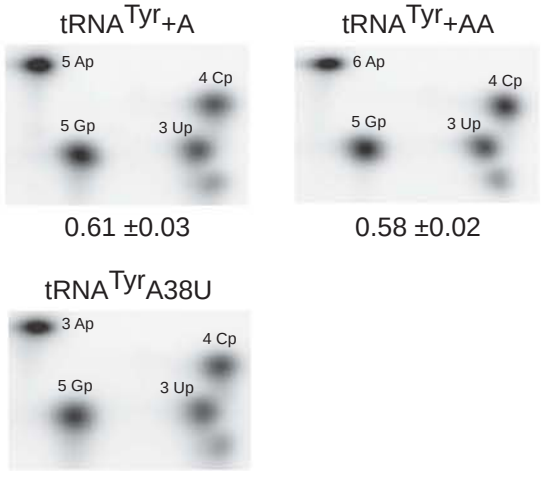
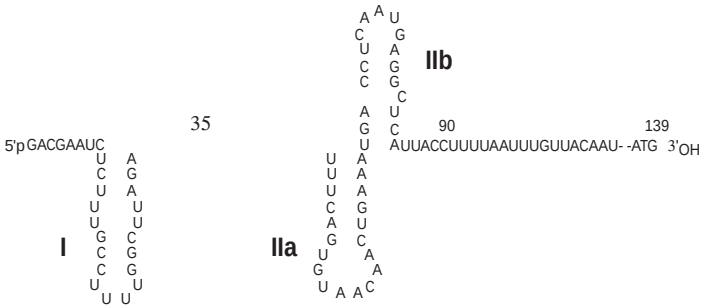
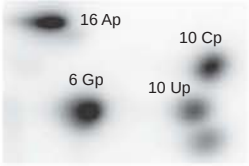


Figure 6

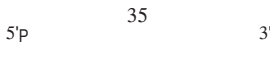
U2:1-139

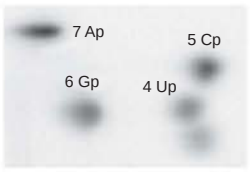




0.84 ±0.00

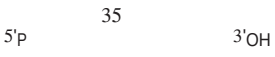
U2:1-85

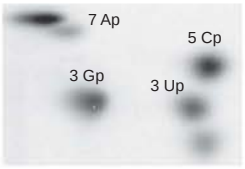




0.63 ±0.00

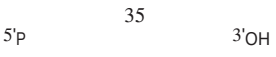
U2:1-85 ΔI

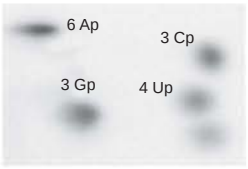




0.65 ±0.00

U2:1-85 ΔIIb

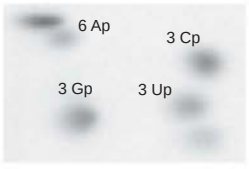




0.36 ±0.00

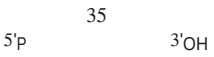
U2:1-85 ΔIΔIIb

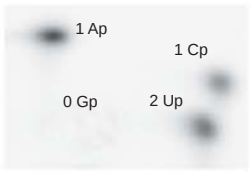




0.39 ±0.00

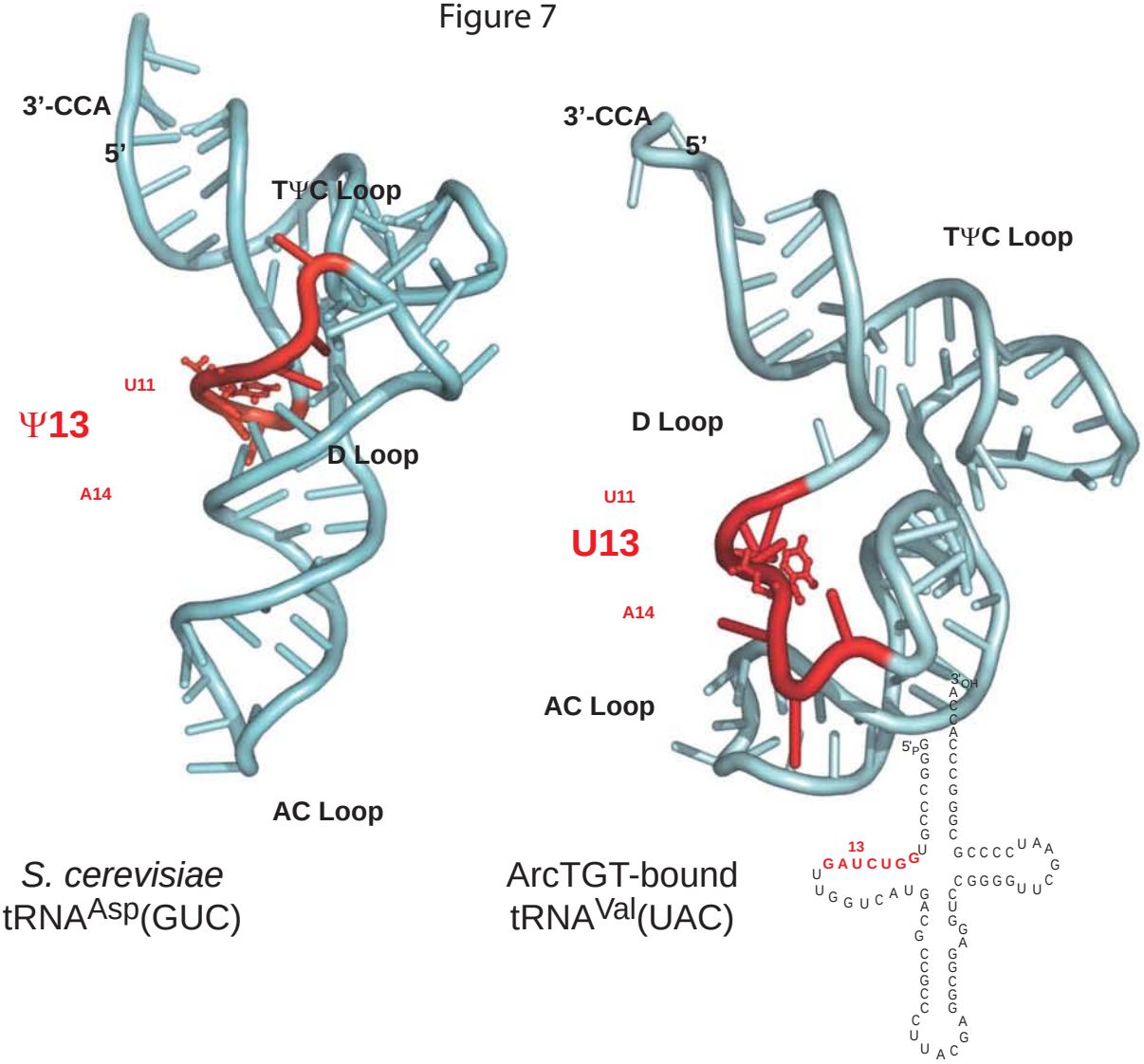
U2:28-47





0.00 ±0.00

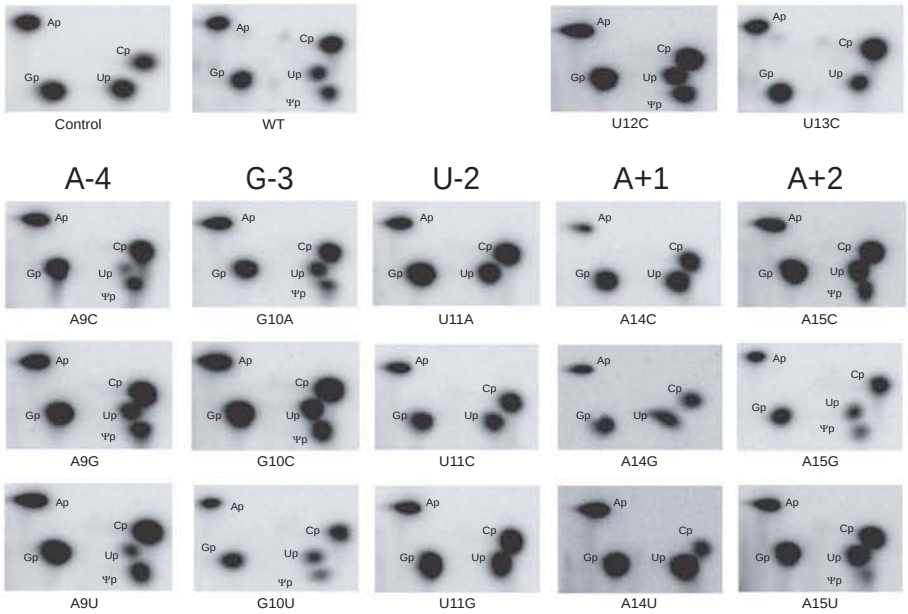
Figure 7



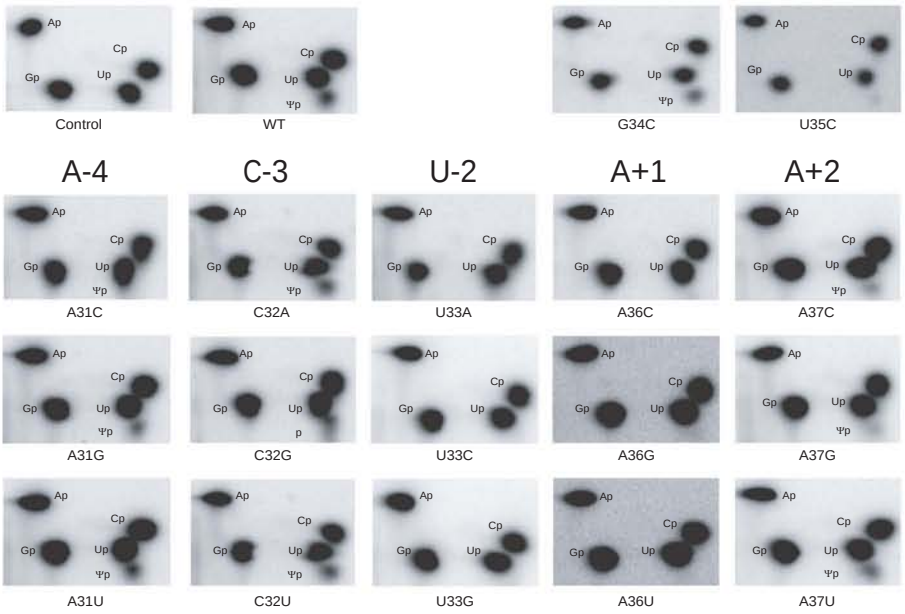
S. cerevisiae tRNA^{Asp}(GUC)

ArcTGT-bound tRNA^{Val}(UAC)

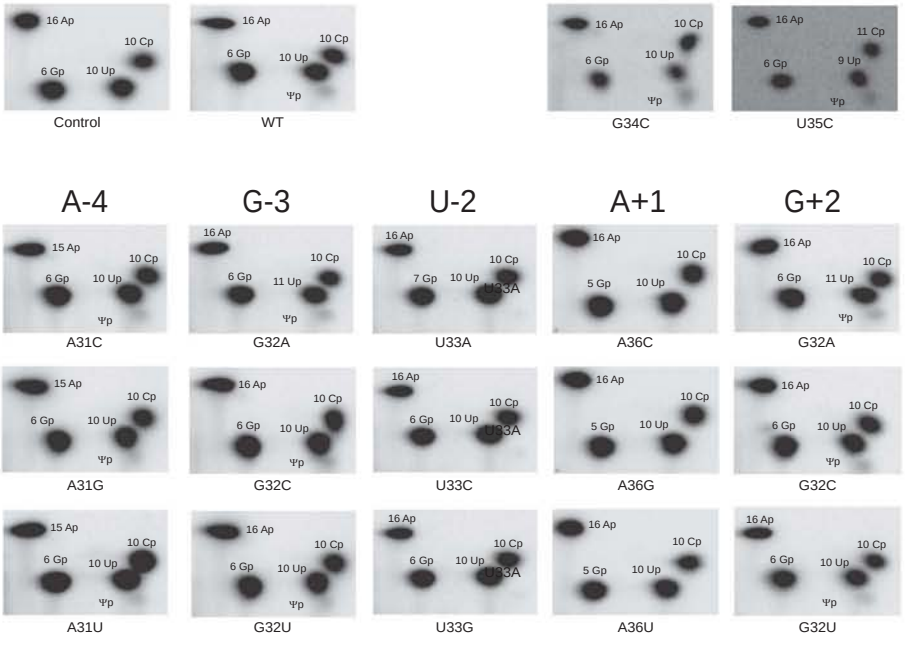
tRNA^{Asp}(GUC)



pre-tRNA^{Tyr}(GΨA)



snRNA U2



3. Résultats complémentaires

Comme nous l'avons mentionné dans l'Introduction, plusieurs méthodes ont été utilisées par d'autres auteurs pour caractériser les propriétés catalytiques des ARN:Ψ-synthases. La détermination des paramètres catalytiques Km et kcat est difficile pour ces enzymes, qui ne nécessitent pas de cofacteur. La technique basée sur la libération de ^{3}H lors de la réaction d'isomérisation (Cortese et al., 1974), qui est citée dans l'Introduction, est très lourde. Certains auteurs ont aussi utilisé la méthode des plus proches voisins pour la détermination précise des valeurs de Km et kcat par réalisation de cinétiques enzymatiques. Cependant, cette approche est aussi très lourde. La détermination des paramètres catalytiques pour l'ensemble des mutants d'ARNt que nous avons étudié dans l'article précédent aurait impliqué la réalisation d'au moins 1000 chromatographies sur plaque. Pour obtenir une estimation de l'efficacité de l'enzyme Pus7p vis-à-vis de ses différents substrats ARNt, nous avons donc choisi de réaliser une série de tests cinétiques simples basés sur la technique des plus proches voisins, ceci sur les 4 mutants ponctuels de l'ARNt^{Asp}(GUC) et du pré-ARN^{Tyr}(GΨA) pour lesquels les taux de modifications après 90 min d'incubation avec l'enzyme étaient les plus diminués, à savoir les variants G10U, U11A, A14C et A15U de l'ARNt^{Asp}(GUC) et A31C, U33A, A36G et A37C du pré-ARN^{Tyr}(GΨA). Nous avons comparé les courbes obtenues avec celles des deux ARN sauvages.

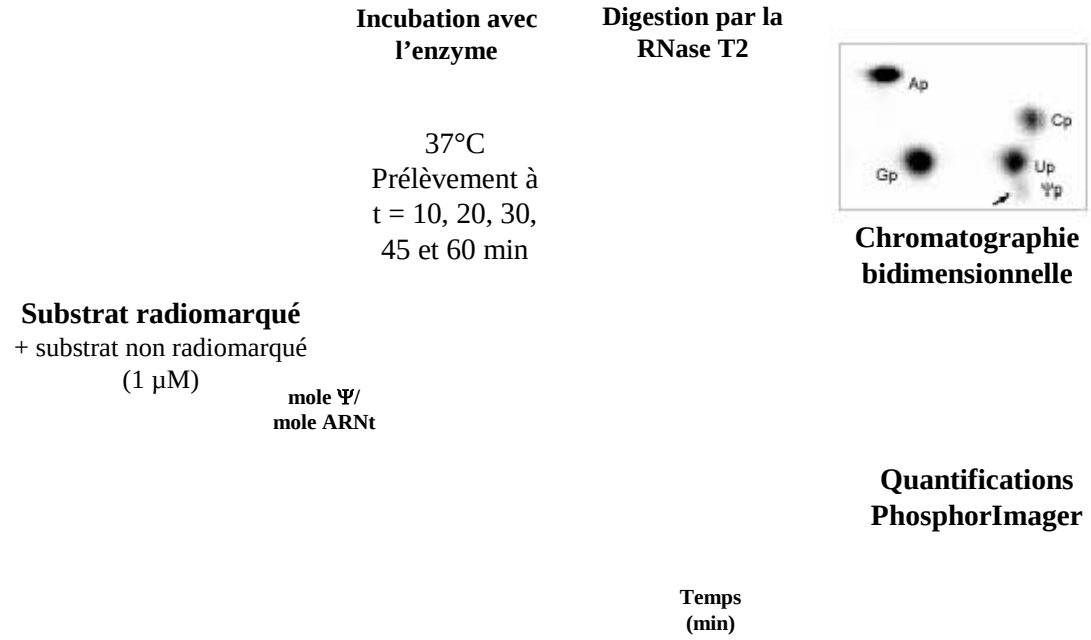


Figure 70: Représentation schématique de la stratégie utilisée pour comparer les cinétiques réactionnelles de Pus7p vis-à-vis des ARNt^{Asp}(GUC) et du pré-ARN^{Tyr}(GΨA) sauvages et mutés.

La modification de chaque ARN substrat a été réalisée en utilisant une concentration fixe d'enzyme (1 nM) et le taux de résidu Ψ formé après différents temps d'incubation (10, 20, 30, 45 et 60 min) a été quantifié après fractionnement par chromatographie sur couche mince de cellulose (Fig. 70). Les transcrits ont été produits par transcription *in vitro* en présence d'[α-^{32}P]ATP sauf pour le variant A14C où le marquage a été réalisé en présence d'[α-^{32}P]CTP. Afin de réaliser les essais à une même concentration de substrat, le transcrit uniformément marqué a été mélangé à un transcrit non radioactif (1 µM) ceci afin que la concentration totale en substrat soit supérieure au Km estimé pour d'autres ARNt:Ψ-synthases (50-150 nM) (Arluison et al., 1999). Comme pour les travaux présentés dans l'article 2, les ARNt ont d'abord été renaturés dans du tampon $AcNH_4$ 100 mM, $MgCl_2$ 5 mM, DTT 2 mM, EDTA 0,1 mM, Tris-HCl 100 mM pH 8,0 ensuite utilisé pour la réaction enzymatique. La renaturation a été réalisé par diminution progressive de la température jusqu'à 30°C puis les ARNt ont été incubés en présence de l'enzyme recombinante Pus7p et des prélèvements ont été réalisés après 10, 20, 30, 45 et 60 min. Les ARN modifiés ont ensuite été hydrolysés toute une nuit par la RNase T2, et les produits d'hydrolyse ont été fractionnés par chromatographie bidimensionnelle sur plaques de cellulose. Le nombre de moles de résidu Ψ formées par mole d'ARNt a été calculé par mesure de la radioactivité présente dans le spot correspondant à ΨMP et aux autres mononucléotides. Le taux de résidu Ψ formé a ensuite été calculé en se basant sur les quantités attendues de résidus AMP, CMP, GMP et UMP après un marquage de l'ARN substrat par incorporation d'[α-^{32}P]ATP. A partir des données obtenues, on peut alors tracer une courbe représentant le nombre de moles de résidu Ψ formées par mole d'ARNt en fonction du temps (Fig. 71).

L'examen des cinétiques de modification obtenues conforte les conclusions que nous avons formulées sur la base des tests d'activité décrits dans l'article 3. En effet, on constate une absence totale d'activité de Pus7p vis-à-vis de l'ARNt^{Asp}(GUC) possédant les substitutions U11A et A14C ainsi que du pré-ARN^{Tyr}(GΨA) possédant les substitutions U33A et A36G ce qui démontre bien une importance cruciale de l'identité des résidus aux positions -2 et +1.

On observe également que les substitutions de bases A15U dans l'ARNt^{Asp}(GUC) et A37C dans le pré-ARN^{Tyr}(GΨA) (positions +2) sont celles qui ont les effets négatifs les plus forts sur l'efficacité de la modification par Pus7p, en particulier dans le pré-ARN^{Tyr}(GΨA). Les substitutions aux positions -3 dans ARNt^{Asp}(GUC) et -4 dans le pré-ARN^{Tyr}(GΨA) entraînent seulement une diminution d'un facteur 2 du taux de modification

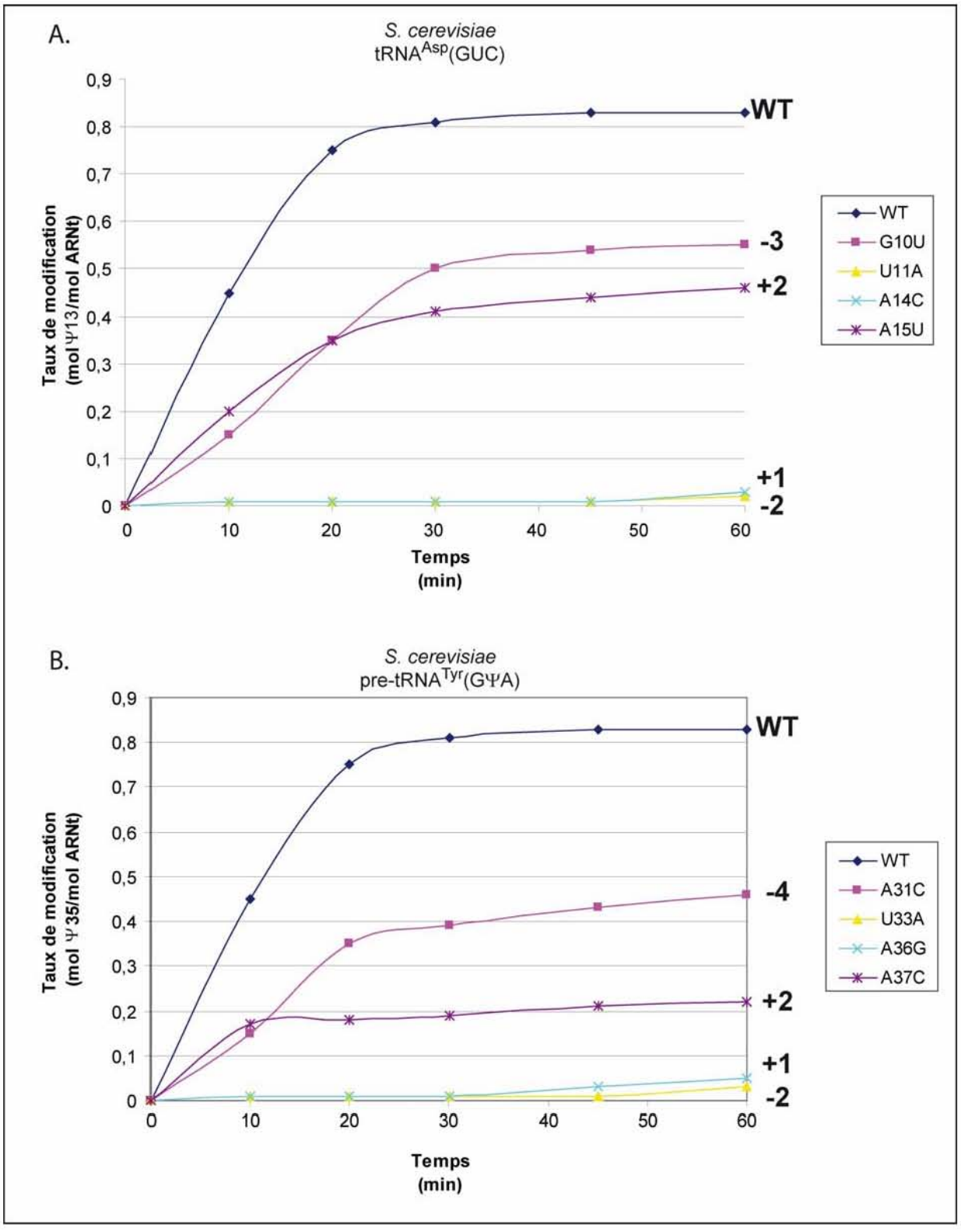


Figure 71 : Cinétiques de modification de l'ARNt^{Asp}(GUC) et de 4 de ses variants (G10U, U11A, A14C et A15U) (Panneau A) et du pré-ARN^{Tyr}(GΨA) et de 4 de ses variants A31C, U33A, A36G et A37C (Panneau B) par Pus7p.

Le nombre de moles de résidu Ψ formées par mole de substrat est représenté en fonction du temps. Le code couleur permettant l'identification des ARNt substrats est indiqué à droite des courbes.

Nous notons que le faible taux de modification que nous avions observé après 90 min d'incubation pour les mutants A15U et G10U de l'ARNt^{Asp}(GUC) et A31C du pré-ARN^{Tyr}(GΨA) est déjà observé après 30 min d'incubation et même après 10 min de réaction pour le variant A37C du pré-ARN^{Tyr}(GΨA) dont le taux de modification par Pus7p est le plus affecté.

Après 10 min d'incubation, les taux de modifications obtenus pour ces 4 variants sont très similaires et correspondent à environ 30% du taux de modification obtenu pour l'ARN sauvage. Ensuite, suivant le variant, un plateau plus ou moins haut entre 20 et 50% est atteint.

Les mêmes taux de modification sont obtenus après 90 min d'incubation que l'ARN ou que l'enzyme soient en excès. Plusieurs hypothèses peuvent être émises pour expliquer cette observation : soit les substrats sont au départ hétérogènes si bien qu'une partie des molécules n'est jamais active, soit ils donnent lieu à la formation de deux types de complexes avec l'enzyme, un complexe conduisant à la formation d'un intermédiaire actif et l'autre restant fixé à l'enzyme sans que l'activité ait lieu. Il faudrait alors imaginer que les taux de complexes actifs et non actifs soient différents suivant la mutation réalisée dans l'ARN. Une autre explication possible serait la présence d'ARN d'*E. coli* compétiteurs fixés sur l'enzyme, seuls les ARN ayant une meilleure affinité pour l'enzyme pouvant être modifiés à 100%. Enfin, nous ne pouvons exclure une dégradation de l'enzyme au bout d'un certain temps d'incubation qui pourrait être plus rapide si elle a peu d'affinité pour l'ARN substrat. Cette hypothèse nous parait néanmoins peu vraisemblable car nous avons déjà constaté que l'enzyme reste active après 90 min d'incubation à 30°C. Nous privilégions donc les deux premières hypothèses, mais des expériences complémentaires devront être réalisées pour apporter des informations invalidant ou consolidant ces explications possibles.

Il sera important de refaire les cinétiques en présence de différentes quantités d'enzymes et aussi de rajouter des ARN après un certain temps d'incubation pour vérifier l'activité de l'enzyme. Notons que dans les expériences qui avaient été réalisées par l'équipe de Z. Szweykowska-Kulinska avec des variants du pré-ARN^{Tyr}(GΨA) d' *A. thaliana*, le même type de cinétique était observé lors des incubations en extrait cellulaire de germe de blé (Szweykowska-Kulinska & Beier, 1992).

PARTIE 3

ETUDE DE L'IMPORTANCE FONCTIONNELLE DES DOMAINES PROTEIQUES QUI SONTADDITIONNELS DANS L'ARN:Ψ-SYNTHASE PUS7P PAR RAPPORT A L'ENZYME TRUD

1. Identification des domaines additionnels de Pus7p par rapport à TruD

1.1. Stratégie utilisée

Après avoir défini les éléments de séquence et de structure requis pour qu'un ARN soit substrat de l'enzyme Pus7p de *S. cerevisiae*, nous avons abordé l'étude des domaines de la protéine qui pourraient être importants pour son activité ou sa spécificité de substrat. En effet, lorsque l'on compare Pus7p à son homologue TruD d'*E. coli*, on constate que TruD est beaucoup plus petite que Pus7p. Comme TruD ne modifie qu'un seul ARN, l'ARNt^{Glu}(UUC), alors que Pus7p modifie 4 types de substrats, nous avons formulé l'hypothèse que des domaines additionnels de Pus7p par rapport à TruD pourraient intervenir dans la plus large spécificité de substrat.

La structure 3D de TruD était déjà établie. Par ailleurs, nous disposions de banques de données de protéines potentielles établies à partir des séquences complètes de génome bactériens, d'archaea et d'eucaryotes. Nous avons ainsi pu rechercher des homologues de TruD et Pus7p dans ces banques par des approches informatiques basées sur des techniques d'alignement multiples et de prédiction de structures secondaires de protéines (Fig.72).

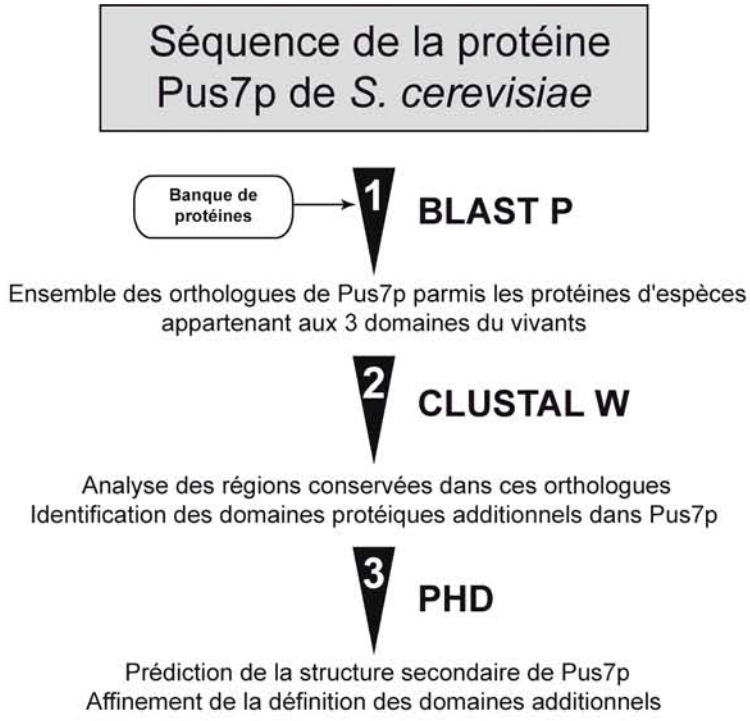


Figure 72 : Stratégie utilisée pour identifier les domaines additionnels de Pus7p de *S. cerevisiae* par rapport à TruD d'*E. coli*.

Les 3 étapes basées sur l'emploi de 3 logiciels distincts (BLASTp, CLUSTALw et pHD) et leurs objectifs sont indiqués.

Ainsi nous avons pu, comme nous allons le décrire, délimiter des domaines protéiques de Pus7p supplémentaires par rapport à ceux connus de TruD.

1.2. Résultats obtenus par alignements multiples de séquence

Dans un premier temps, nous avons recherché des homologues de Pus7p dans les banques de séquence protéiques (NCBI), ceci en utilisant l'algorithme BLASTp (Basic Local Alignement Search Tool) (Altschul et al., 1997). Comme cela avait déjà été observé (Kaya & Ofengand, 2003), nous avons trouvé des homologues de Pus7p dans les 3 domaines du vivant. De manière générale, nous avons constaté que les protéines bactériennes sont de petite taille (env. 35 kDa), les protéines d'archaea ont une taille un peu plus grande (env. 46 kDa) et les protéines eucaryotes sont beaucoup plus grandes (de l'ordre de 70 kDa). *S. cerevisiae* n'est donc pas un cas atypique parmi les eucaryotes. Ceci renforçait l'idée que la multiplicité des substrats chez les eucaryotes pouvait avoir un lien avec cette augmentation de taille des enzymes Pus7p eucaryotes. A moins que cette augmentation reflète d'autres fonctions non identifiées de cette protéine chez les eucaryotes.

Afin de définir avec précision les régions additionnelles présentes dans Pus7p de *S. cerevisiae* par rapport à ses homologues bactériens et d'archaea, nous avons réalisé un alignement multiple de séquences à l'aide du logiciel CLUSTALw (Thompson et al., 1994) en utilisant 12 séquences protéiques représentatives des 3 domaines du vivant, 3 de bactéries (*A. aeolicus*, *E. sakazakii* et *E. coli*), 3 d'archaea appartenant au règne des Euryotes (*M. burtonii*, *A. fulgidus* et *M. mazei*), 3 de levure (*S. cerevisiae*, *K. lactis*, *P. guilliermondii*), 3 de vertébrés (*X. laevis*, *H. sapiens* et *M. musculus*) et celle de la drosophile (*D. melanogaster*). Nous avions constaté un fort degré d'homologie de l'ensemble de ces protéines lors des alignements par logiciel BLASTp (E-value faible) (Tableau 15).

Règne	Organisme	Taille (aa)	E-value	% Identité	% Gap
EUCARYOTES	**Saccharomyces cerevisiae**	**676**	**0**	**100**	**0**
	Kluyveromyces lactis	636	0	81	0
	Pichia guilliermondii	668	1,00E-179	68	4
	Xenopus laevis	638	7,00E-99	54	16
	Homo sapiens	634	2,00E-98	53	14
	Mus musculus	640	4,00E-97	53	16
	Drosophila melanogaster	680	3,00E-80	49	13
ARCHAEA	Methanococcoides burtonii	439	1,00E-36	49	14
	Archaeoglobus fulgidus	411	2,00E-36	46	19
	Methanosarcina mazei	438	3,00E-29	45	20
BACTERIES	Aquifex aeolicus	385	5,00E-22	54	12
	Enterobacter sakazakii	349	3,00E-10	50	6
	Escherichia coli	349	3,00E-08	46	6

Tableau 15 : Résultats obtenus en alignant par le logiciel BLASTp les séquences de protéines homologues à Pus7p appartenant à des espèces des 3 domaines du vivant.

La séquence de la protéine de référence est indiquée par un rectangle gris. La taille des protéines, leur pourcentage d'identité et de séquences additionnelles ou manquantes (gaps) par rapport à Pus7p sont

indiquées ainsi que les valeurs de E-value que le logiciel calcule et qui donnent une idée du degré d'homologie par rapport à la séquence de référence.

L'alignement multiple que nous avons réalisé (Figure 73) est en accord avec le fait que les ARN:Ψ-synthases de la famille TruD possèdent bien les 5 motifs conservés qui servent de signature aux ARN:Ψ-synthase en général (motif I, II, IIa, III et IIIa) (Del Campo et al., 2001). Comme déjà décrit, le motif IIIa trouvé en aval du motif II dans les enzymes des 4 autres familles d'ARN:Ψ-synthases est, dans le cas de la famille TruD, trouvé en amont du motif I. Cette permutation circulaire n'est pas incompatible avec la formation d'un site catalytique canonique comme déjà décrit (Kaya et al., 2004).

L'alignement obtenu permet aussi de retrouver dans l'ensemble des enzymes comparées les 5 acides aminés du site actif, qui dans toutes les ARN:Ψ-synthases étudiées jusque-là se sont avérés avoir un rôle dans la catalyse : l'acide aspartique catalytique (motif II), le résidu basique (Lys) (boucle L1) qui forme un pont salin avec l'acide aspartique catalytique, le résidu phénylalanine (motif IIa) qui oriente la base de l'uridine cible par empilement des cycles aromatiques et les deux résidus hydrophobes (Ala du motif III et Phe du motif IIIa) (Kaya & Ofengand, 2003). Par ailleurs, l'alignement de séquence que nous avons réalisé montre que l'augmentation en taille des ARN:Ψ-synthases eucaryotes par rapport aux enzymes d'archaea ou de bactéries ne correspond pas à une insertion unique mais à un ensemble d'insertions de tailles et de séquences variables (Figure 73).

Ainsi, si l'on se réfère aux homologues bactériens de Pus7p dont la séquence est la plus courte, on constate que les protéines eucaryotes possèdent 5 insertions (Ins 1, 2, 3, 4 et 5) toujours situées aux mêmes positions ainsi qu'une grande extension (env. 100 aa) dans la séquence N-Terminale (Fig. 73 et 74B). Les protéines d'archaea ne possèdent que 3 insertions (Ins 2, 3 et 4) par rapport aux enzymes bactériennes. Là encore trouvées aux mêmes postions dans les 3 espèces étudiées. L'insertion des eucaryotes est de grande taille (env. 150 aa) et sa séquence est fortement conservée de la levure jusqu'à l'homme ce qui suggère un rôle fonctionnel. La taille et la séquence de l'insertion 3 sont très conservées entre archaea et eucaryotes. L'insertion 5, de taille réduite, a une séquence conservée chez les vertébrés. Par contre, les insertions 2 et 4 sont variables en taille et en séquence.

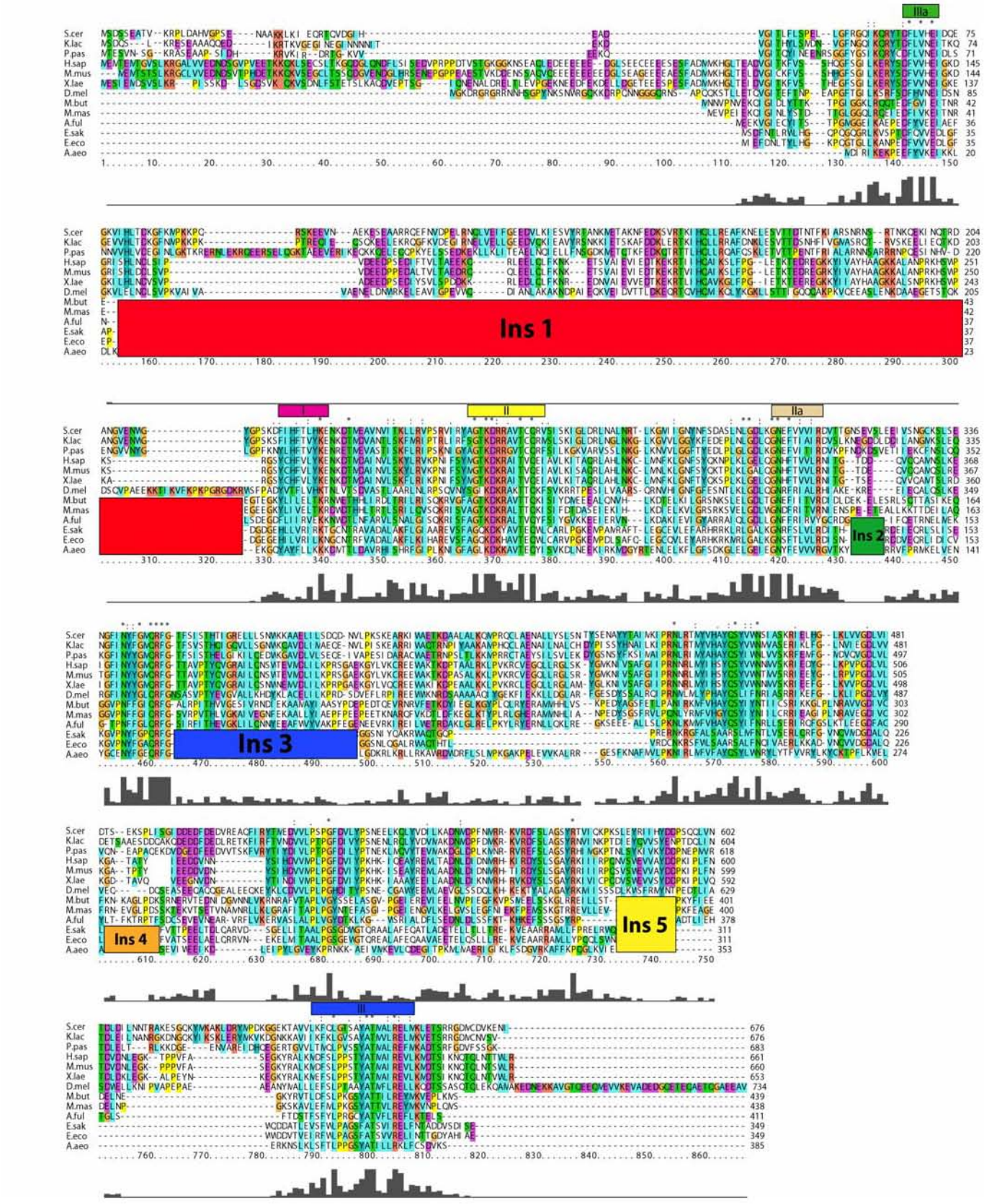


Figure 73 : Alignement multiple des séquences des enzymes Pus7p de *S. cerevisiae* et de 12 protéines homologues identifiées à partir de banques de protéines potentielles de 12 espèces représentatives de chacun des domaines du vivant.

Les séquences sont numérotées de 10 en 10 aa, sur la base de la séquence de Pus7p de *S. cerevisiae*. Les signes au dessus de l'alignement indiquent le degré de conservation des acides aminés : *=100% de conservation, :=80% de conservation, .=50% de conservation. Les 5 insertions (Ins) sont représentées par des rectangles de couleurs différentes et sont numérotées. Les motifs conservés dans toutes les ARN:Ψ-synthases (I, II, IIa, III et IIIa) sont indiqués au dessus de l'alignement de séquence.

2. Délétions des domaines additionnels en vue de tester leur importance fonctionnelle

2.1. Raffinement des alignements de séquence par la prise en compte de la structure 2D des protéines

Pour cela, nous avons utilisé le programme pHD (Rost et al., 1994) qui permet d'optimiser les alignements de séquence réalisés par ClustalW lorsque la structure 3D de l'une des protéines dont la séquence est alignée est connue. Dans notre cas, 2 structures 3D étaient disponibles (TruD d' *E. coli* et de *M. mazei*). L'algorithme pHD permet de prédire la structure 2D des protéines dont la séquence a été alignée par ClustalW en tenant compte des protéines dont la structure 3D a été établie. Le logiciel permet aussi d'évaluer le niveau de fiabilité des structures 2D proposées. Un haut niveau de fiabilité est donné pour celle établie pour Pus7p (valeur pHDsec > 8). La structure 2D proposée pour Pus7p et l'alignement des séquences homologues de Pus7p et TruD sont représentés dans la figure 74A. Dans la structure prédite pour le domaine catalytique de Pus7p, on retrouve bien les 8 brins β (1, 2, 3, 4, 5, 6, 10, 11) flanqués de 2 hélices α (1 et 2) caractéristiques du site catalytique de TruD d'*E. coli* (Kaya et al., 2004) (Fig. 78C).

Les domaines insérés sont respectivement situés entre les brins β1 et β2 pour l'insertion 1, entre le brin β5 et l'hélice α2 pour l'insertion 2, entre les hélices α3 et α4 pour l'insertion 3, entre le brin β6 et l'hélice α5 pour l'insertion 4 et entre les brins β10 et β11 pour l'insertion 5 (Fig. 74A). De plus, on constate que les insertions 1 et 5 sont localisées dans le domaine catalytique, tandis que les insertions 3 et 4 sont situées dans le domaine TruD qui interviendrait dans la reconnaissance de l'ARN. L'insertion 2, de petite taille, est présente dans une région faisant le lien entre les 2 domaines dans l'enzyme TruD.

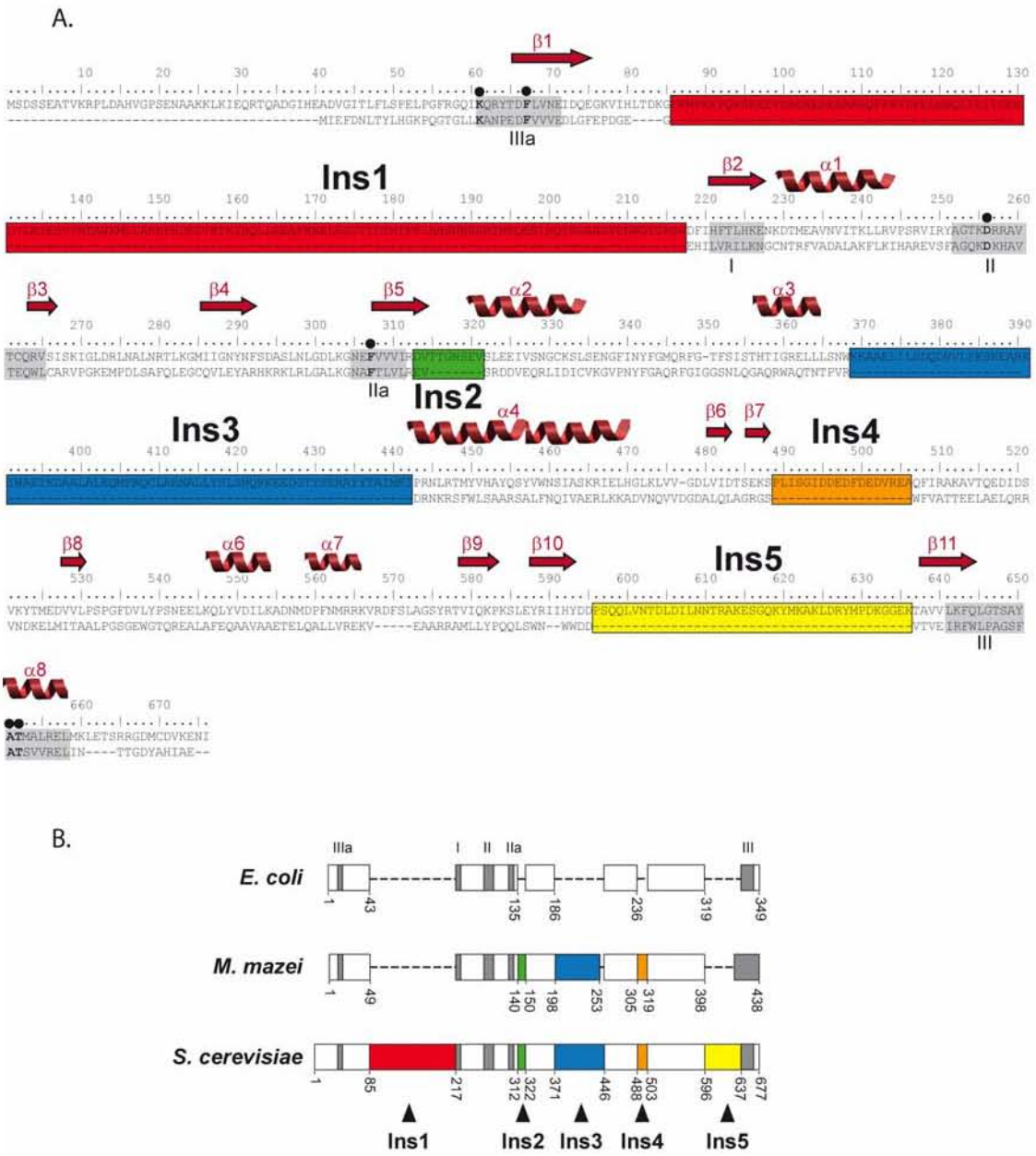


Figure 74 : Comparaison des architectures des ARN:Ψ-synthases Pus7p de *S. cerevisiae* et de ses homologues d'*E. coli* et de *M. mazei*.

A. Alignement des séquences des enzymes Pus7p de *S. cerevisiae* et TruD d'*E. coli* réalisé en tenant compte de la conservation des motifs de structure 2D. Ces motifs sont représentés sous forme de flèches (Brin β) et d'hélices (Hélice α). Ils sont numérotés en utilisant la nomenclature définie pour la protéine TruD d' *E. coli*. Les 5 insertions sont représentées par des rectangles de couleur. Les motifs conservés (I, II, IIa, III et IIIa) dans toutes les ARN:Ψ-synthases sont représentés par des rectangles gris et les acides aminés important du site actif sont indiqués en gras, surmontés par un point noir.

B. Représentation schématique des domaines conservés et additionnels de Pus7p de *S. cerevisiae* et ses homologues de bactérie (*E. coli*) et d'archaea (*M. mazei*). Les domaines additionnels sont indiqués par

des rectangles de couleurs différentes et sont numérotés selon la séquence de chaque protéine. Les motifs conservés dans toutes les ARN:Ψ-synthases sont représentés par des rectangles gris.

Dans le but de comprendre si ces domaines additionnels sont impliqués dans la reconnaissance des ARNs substrats et s'ils peuvent expliquer la plus large spécificité de substrat de l'enzyme Pus7p de *S. cerevisiae* par rapport à l'enzyme TruD d'*E. coli*, nous avons éliminé indépendamment les insertions 1, 3, 4 et 5. Nous avons choisi de ne pas éliminer l'insertion 2 car la prédiction de structure secondaire semble indiquer que cette insertion, dont la taille varie beaucoup suivant les espèces,, forme une courte boucle (10 aa chez *S. cerevisiae*) joignant le brin β5 et l'hélice α2. Sa suppression risquait de perturber la structure 3D du mutant concerné.

2.2. Production de gènes codant des protéines ou chaque insertion est individuellement éliminée

Pour réaliser la délétion des régions codants les insertions 1, 3, 4 et 5, nous avons utilisé une méthode qui avait été développée par Y. Imai et al. (1991). Dans cette approche, le plasmide contenant le gène où la délétion est réalisée a été amplifié par PCR à l'exception de la séquence à déléter (Fig. 75). Nous avons utilisé le plasmide pET28b-Pus7WT permettant l'expression de Pus7p chez *E. coli* (voir publication n°2) et des couples d'oligonucléotides amorces qui s'hybrident de part et d'autres des régions à éliminer. Le produit de la réaction a été ensuite circularisé après hydrolyse de l'ADN parental par l'enzyme DpnI, qui ne clive que les ADN méthylés préparés à partir d'*E. coli*. Les produits de ligation ont alors été utilisés pour transformer la souche XL1 d'*E. coli*. Après sélection des transformants, les plasmides contenus dans 3 à 5 clones ont été amplifiés directement sur colonie. Une amplification réalisée avec le plasmide pET28b de départ servait de contrôle. L'analyse des fragments amplifiés par électrophorèse en gel d'agarose a permis de vérifier que nous avions bien obtenus des plasmides contenant les 4 délétions (Fig. 76). L'identité des inserts contenus dans les plasmides sélectionnés permettant la production des protéines His6-Pus7Δ1, His6-Pus7Δ3, His6-Pus7Δ4 et His6-Pus7Δ5 a été vérifiée par séquençage. Les différents oligonucléotides utilisés pour ces expériences de délétion sont répertoriés dans le Tableau 16 et le protocole expérimental utilisé est décrit dans le chapitre Matériel et Méthodes.

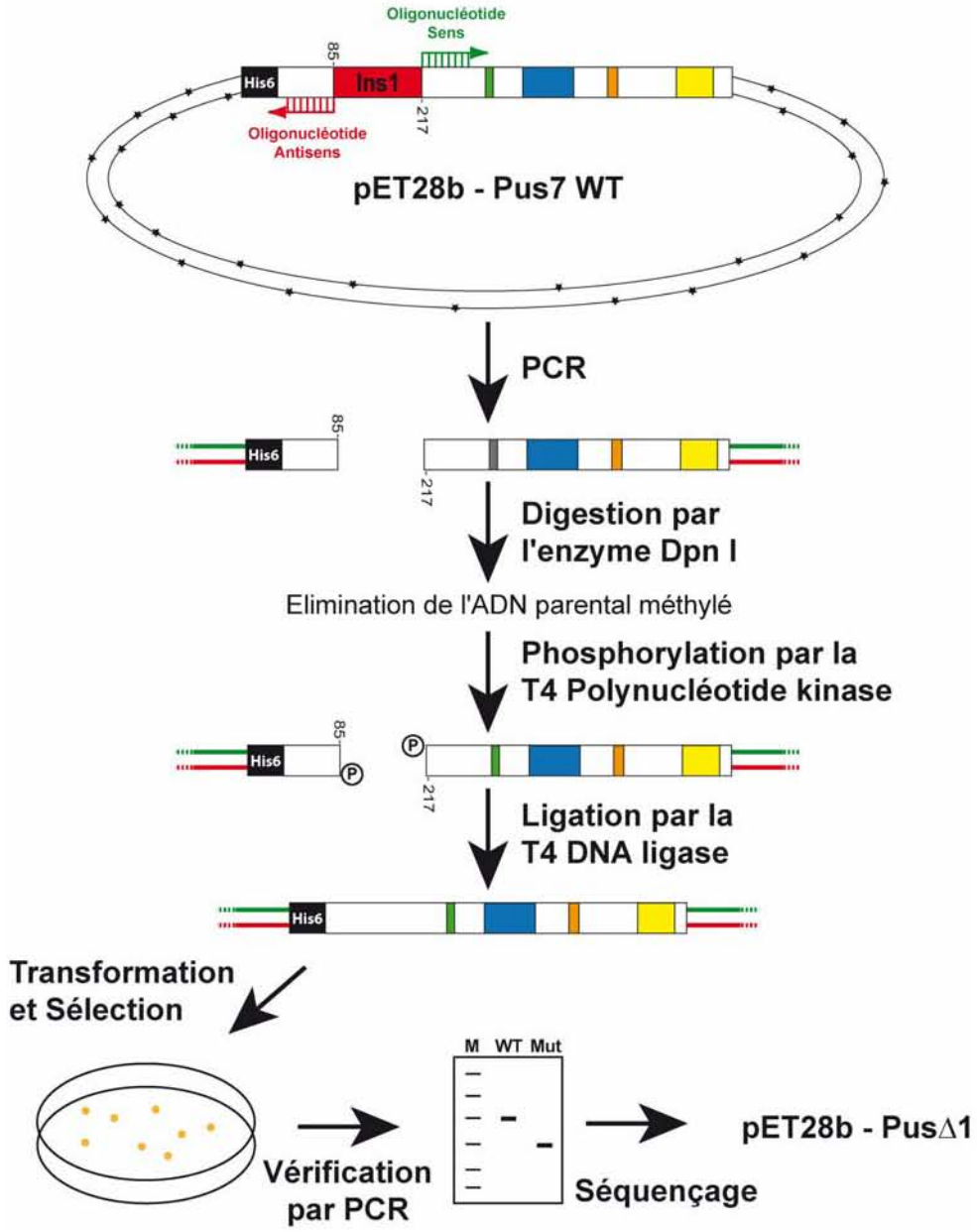


Figure 75 : Stratégie utilisée pour la délétion des régions additionnelles de la protéine Pus7p de *S. cerevisiae*. D'après Y. Imai et al., 1991. (ref endnote)

Nom	Séquence 5' vers 3'	Utilisation
4442	GATTTTATTCACTTCACTCTACATAAAGAAAATAAGGAT	Oligo sens de la délétion de la région 85-217 de Pus7 S.cerevisiae
4469	TTTATCAGTTAAGTGAATGACCTTACCTTCTTG	Oligo antisens de la délétion de la région 85-217 Pus7 S.cerevisiae
4444	TTGAGAACTATGTATGTACATGCCTACCAAAGT	Oligo sens de la délétion de la région 371-446 Pus7 S.cerevisiae
4470	AGCAGCCTTCTTCCAGTTTGAAAGGAG	Oligo antisens de la délétion de la région 371-446 Pus7 S.cerevisiae
4446	CGTGAAGCTCAGTTCATTCGTGCAAAA	Oligo sens de la délétion de la région 488-503 Pus7 S.cerevisiae
4447	TGACTTTTCACTAGTATCAATGACAAGGTC	Oligo antisens de la délétion de la région 488-503 Pus7 S.cerevisiae
4448	ACTGCGGTGGTGTTAAAATTTCAATTAGGA	Oligo sens de la délétion de la région 596-636 Pus7 S.cerevisiae
4471	ATCATCATAGTGAATTATTCTATACTCTAAACTCTTTGG	Oligo antisens de la délétion de la région 596-636 Pus7 S.cerevisiae
67	GTTTTCCCAGTCACGAC	Oligo sens de séquençage des plasmides de la série pUC
122	CAGGAAACAGCTATGAC	Oligo reverse de séquençage des plasmides de la série pUC
2923	AAACTAGCTAGCATGTCTGACTCCTCA	Oligo sens utilisé pour la vérification de la taille des mutants de Pus7p
967	CTAATACGACTCACTATAGGG	Oligo sens de séquençage des plasmides de la série pET
1431	GCTCAGCGGTGGCAGCAGCC	Oligo reverse séquençage des plasmides de la série pET

Tableau 16 : Oligonucléotides utilisés pour réaliser les délétions des régions additionnelles de Pus7p de *S. cerevisiae* par rapport à TruD et pour vérifier ces délétions.

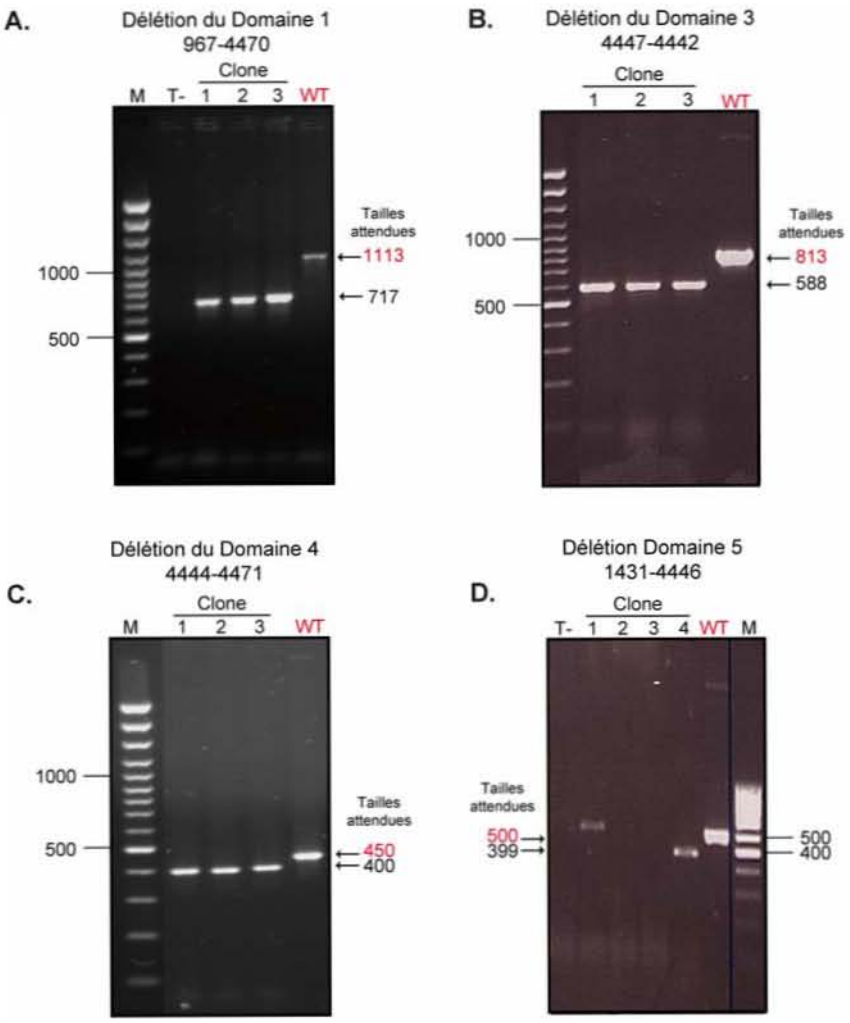


Figure 76 : Vérification par PCR sur colonie de la présence de la délétion escomptée dans les plasmides pET28b sélectionnés après mutagenèse dirigée.

La délétion de la région concernée a été vérifiée par amplification d'ADN sur colonies sur au moins 3 clones, ceci en employant un couple d'amorces spécifique pour chaque mutant. Une amplification contrôle a également été réalisée avec le plasmide sauvage (WT) (Panneaux A, B, C et D). Les tailles des fragments PCR attendus pour chaque construction sont indiquées. La piste T- correspond à une réaction effectuée en absence de colonie et donc d'ADN. Un marqueur de taille (M) a été déposé dans chaque électrophorèse.

3. Production et purification des mutants Δ1, Δ3, Δ4 et Δ5 de l'enzyme Pus7p de *S. cerevisiae*

La production et la purification des protéines recombinantes dans les bactéries sont souvent des étapes limitantes. En particulier, lorsque l'on exprime une protéine eucaryote dans une cellule hôte procaryote, les différentes étapes de maturation (modification post-traductionnelle, clivage de peptides, etc..) n'ont pas lieu.

Le plasmide pET28b utilisé ci-dessus pour les mutagenèses, permettait de produire aussi bien la protéine sauvage que les 4 mutants avec une étiquette polyhistidine à l'extrémité N-terminale. Nous avons également réalisé une construction, dans le même plasmide, permettant la production de la protéine avec une étiquette polyhistidine à l'extrémité C-terminale. Les protéines ont ainsi pu être produites en une étape unique de purification sur colonne d'affinité selon le protocole décrit dans la partie Matériels et Méthodes et dont les principes généraux sont rappelés dans la partie 1 du chapitre Résultats. Les fractions d'élution recueillies ont été analysées par électrophorèse en gel dénaturant (SDS-PAGE) coloré au bleu de Coomassie (Fig. 77A).

Toutes les productions ont été réalisées dans des conditions très similaires à savoir : culture en milieu riche LB, temps d'induction de 15 h à 20°C. Une première constatation est que le niveau d'expression et de solubilité des différents mutants de la protéine Pus7p est variable (Fig. 77A). Comme décrit dans la publication n°2, la protéine Pus7p sauvage possédant une étiquette His_6 à son extrémité N-terminale peut être produite en grande quantité (50 mg/l de culture) et ainsi être obtenue à une concentration élevée après purification (10 mg/mL en sortie de colonne d'affinité). Cependant, l'analyse par SDS-PAGE (Fig. 77A piste 1) révèle la présence de contaminants de plus faible poids moléculaire pouvant correspondre à des produits de dégradation et/ou des produits de traduction abortive. Les mêmes observations sont faîtes pour les protéines Pus7-His6CTer, Pus7-Δ3 et Pus7-Δ4 (Figure 77A pistes 2, 4 et 5 respectivement) Nous constatons aussi une très faible solubilité du mutant Pus7-Δ1 suggérant que la délétion de grande taille (132 aa) perturbe le repliement de la protéine et ainsi diminue sa solubilité. De manière intéressante, nous observons que la protéine Pus7-Δ5 peut être obtenue à un plus haut degré de pureté, puisque seul un contaminant d'environ 35 kDa est observable dans les fractions d'élutions à l'issue de la purification.

Nous avons aussi utilisé la détection des protéines par Western-Blot en utilisant un anticorps dirigé contre l'étiquette His_6 (Fig. 77B). Nos résultats démontrent que la plupart des contaminants correspondent bien à des produits de dégradation et/ou des produits de traduction abortive et ce malgré l'utilisation d'une souche bactérienne *E. coli* BL21-Codon Plus (Stratagene) dont les gènes de plusieurs protéases sont mutés et qui surexprime les ARNt rares d'*E. coli* correspondant à des codons plus fréquemment trouvés chez les eucaryotes (Arg, Ile et Leu). En particulier, on note la présence de 4 doublets arginine (RR) dans la

séquence de Pus7p (Fig. 77C) pouvant conduire à un arrêt de la traduction et donc à la production de protéines dont la taille plus réduite (de 14,4 kDa à 77,9 kDa) correspond à celle de certains sous fragments de Pus7p observés sur gel (Fig. 77B).

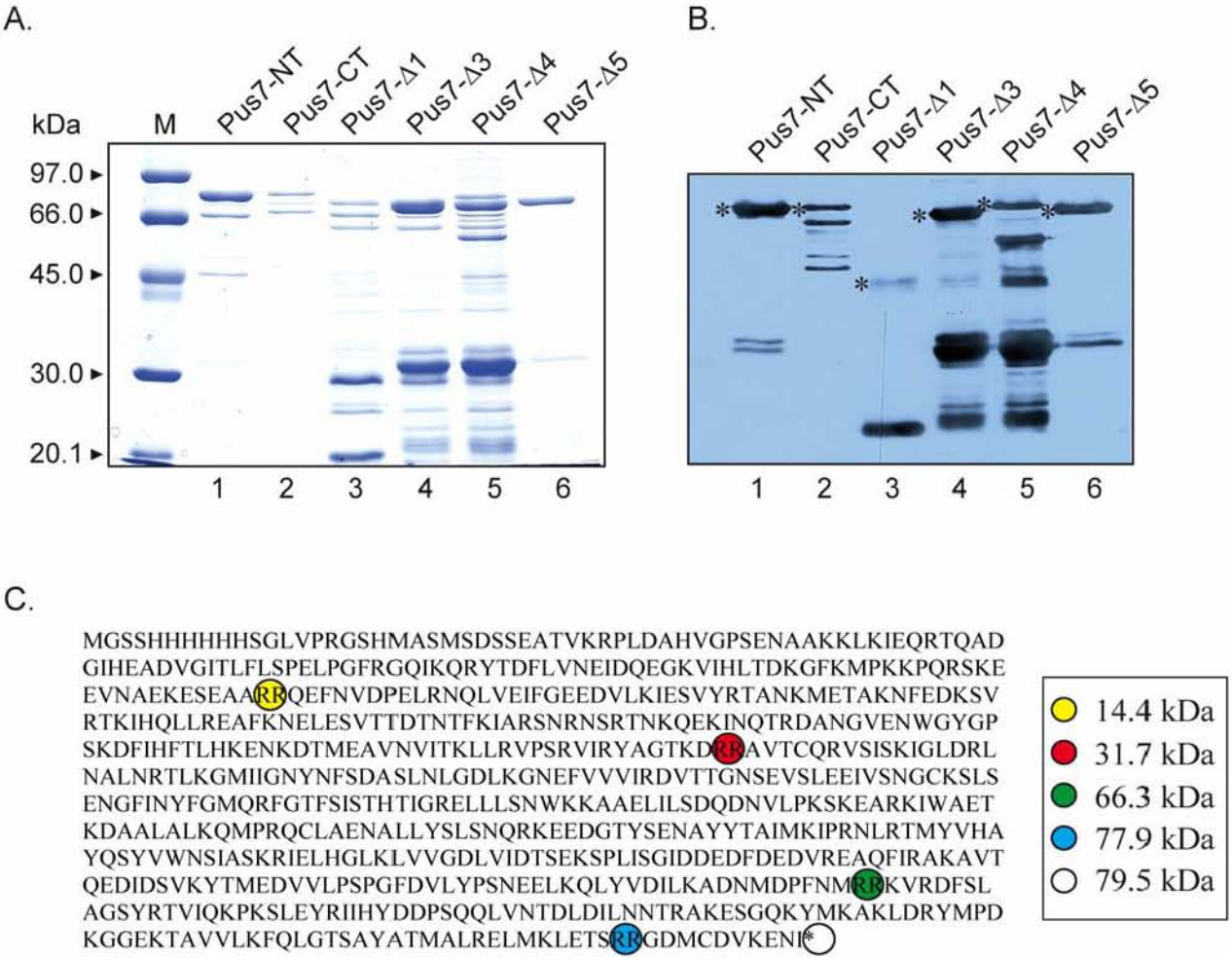


Figure 77 : Analyse du contenu en protéines des fractions collectées au cours de la purification sur colonne de Ni-NTA Agarose des différents mutants de Pus7p-His$_6$.

A. Electrophorèse en gel de polyacrylamide dénaturant à 12,5%. Les bactéries ayant produit les protéines possédant une étiquette His$_6$ ont été lysées par sonication, centrifugées puis le surnageant a été purifié sur colonne Ni-NTA Agarose. Deux µl de la fraction d'élution la plus concentrée ont été déposés pour chaque protéine (pistes 2 à 6). Le gel a été coloré au bleu de Coomassie. La piste M contient un marqueur précoloré contenant plusieurs protéines dont le poids moléculaire est indiqué à gauche de la figure.

B. Analyse par Western-Blot en utilisant un anticorps spécifique de l'étiquette polyhistidine, His$_6$. Le protocole utilisé est décrit dans la partie Matériels et Méthodes. Les protéines marquées par une étoile correspondent aux protéines : His$_6$-Pus7 (79,6 kDa), Pus7-His$_6$ (79,5 kDa), His$_6$-Pus7-Δ1 (59,5 kDa), His$_6$-Pus7-Δ3 (68,7 kDa), His$_6$-Pus7-Δ4 (75,3 kDa) et His$_6$-Pus7-Δ5 (72,5 kDa).

C. Analyse de la présence de doublet arginine dans la séquence de la protéine Pus7 de *S. cerevisiae* pouvant conduire à des événements de traduction abortive. Les doublets arginine sont encadrés par des cercles de différentes couleurs et la masse molaire des protéines tronquées en cas de traduction abortive est indiquée sur la droite.

4. Test de l'activité des mutants Δ1, Δ3, Δ4 et Δ5 de l'enzyme Pus7p de *S. cerevisiae*

Nous avons malgré tout testé l'activité des 4 protéines mutées après l'étape de purification sur colonne de Ni-NTA Agarose sur l'ARNt^{Asp}(GUC) et le pré-ARNt^{Tyr}(GΨA). Les tests d'activité ont été réalisés en présence d'un excès d'ARNt. Les autoradiogrammes des chromatographies en couche mince sont illustrés dans la Fig. 78B. L'efficacité de modification pour chaque substrat a été calculée grâce au logiciel ImageQuant.

Nous avons constaté que les délétions de chacun des domaines additionnels (Fig. 78A) avaient des effets similaires que l'un ou l'autre des 2 substrats soit utilisé. Le taux de modification des ARNt^{Asp}(GUC) et pré-ARNt^{Tyr}(GΨA) par les mutants His$_6$-Pus7Δ4 et His$_6$-Pus7Δ5 n'est que très légèrement inférieur à celui obtenu avec la protéine sauvage. A l'inverse, le mutant His$_6$-Pus7Δ3 présente une activité réduite de l'ordre d'un facteur compris entre 2,5 et 3 par rapport à la protéine sauvage. Enfin, aucune activité n'était détectée pour le mutant His$_6$-Pus7Δ1, mais comme nous l'avons mentionné plus haut très peu de protéine entière devait être présente dans l'échantillon du fait de sa faible solubilité.

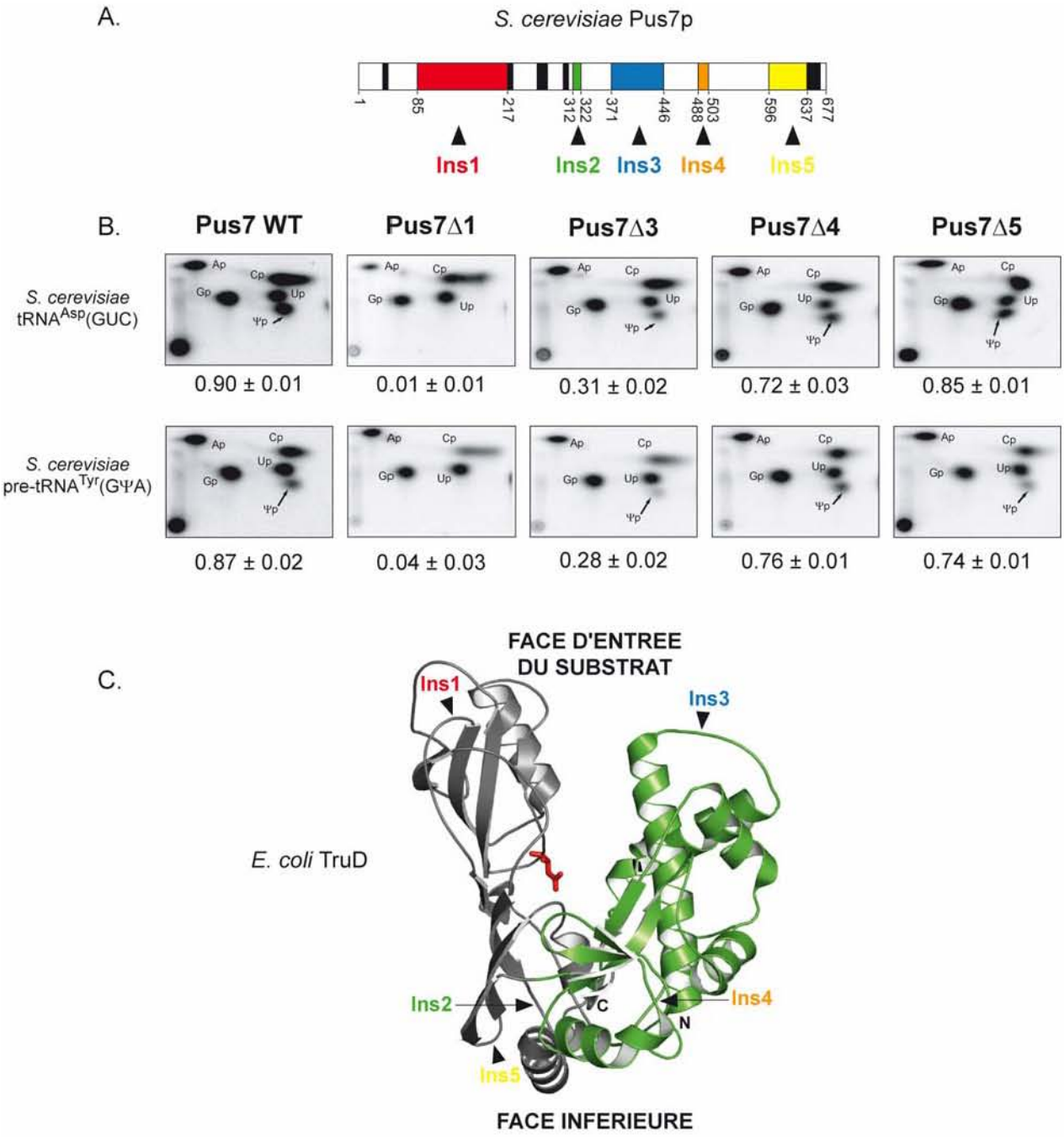


Figure 78 : Analyse du taux de modification des ARNt^{Asp}(GUC) et du pré-ARNt^{Tyr}(GΨA) après incubation avec la protéine His$_6$-Pus7 et avec les variants His$_6$-Pus7Δ1, His$_6$-Pus7Δ3, His$_6$-Pus7Δ4 et His$_6$-Pus7Δ5.

A. Architecture de Pus7p et position des différentes insertions. Le code couleur utilisé pour les domaines de Pus7p est celui utilisé dans la structure 3D établie pour TruD dans le panneau C.

B. Chromatographies obtenues lors de l'analyse de l'activité des protéines Pus7p sauvages et mutantes sur les $ARNt^{Asp}$(GUC) et pré-$ARNt^{Tyr}$(GΨA). Le système de migration utilisé est le système N/R. L'identification des spots a été faîte sur la base des cartes établies précédemment par Keith (Keith, 1995).
C. Structure 3D de l'enzyme TruD de *E. coli* mettant en évidence la localisation des insertions présentes dans Pus7p. Les domaines catalytiques et TruD sont représentés sous forme de rubans gris et verts, respectivement. L'acide aminé catalytique est représenté sous forme d'un bâtonnet rouge. Les insertions sont numérotées et leurs positions sont représentées par des triangles noirs.

Ces données suggèrent que le domaine 3 et peut être le domaine 1 pourraient avoir une importance fonctionnelle, peut être dans la fixation des substrats. Il est important de remarquer que par rapport à la structure 3D établie pour la protéine TruD d'*E. coli*, les insertions 1 et 3 sont localisées à proximité de la crevasse qui a été proposée comme le site d'entrée du substrat (Fig. 78C). De plus, les domaines 1 et 3 contiennent une proportion importante de résidus basiques (arginine et lysine). Ainsi, il est possible que ces domaines, par liaisons électrostatiques aspécifiques avec les ARN substrats, facilitent leur association à l'enzyme.

Pour confirmer nos hypothèses, il aurait été intéressant de disposer de la structure 3D de la protéine Pus7p de *S. cerevisiae* complexée ou non à un ARN substrat. Dans cet objectif, nous avons entrepris un travail visant à cristalliser la protéine Pus7 de levure ce qui imposait de la produire en grande quantité et avec un haut niveau de pureté.

5. Optimisation des conditions de production et purification de la protéine Pus7 de *S. cerevisiae* et du variant Pus7Δ5 en vue d'études structurales

5.1. Optimisation des conditions de production

Le premier problème à résoudre était d'obtenir une protéine Pus7 parfaitement homogène et hautement purifiée. Nous avons à la fois essayé d'optimiser les conditions de production chez *E. coli* et de purification. Les premiers tests de production ont été réalisés en utilisant les conditions de culture que nous avions mises au point pour obtenir un bon rendement de production de l'ARN:Ψ-synthase Pus8p (Article 1 dans le chapitre Matériel et Méthodes). Dans le protocole établi, la culture bactérienne est d'abord réalisée en milieu riche LB à 37°C, puis la production d'enzyme est induite par ajout d'IPTG à une concentration de 1 mM lorsque l'absorbance du milieu à 600 nm atteint 0.8 U/mL. La culture est alors poursuivie à 20°C pendant 15 h. Comme nous pouvons le constater dans la figure 79, l'analyse par SDS-PAGE de l'extrait cellulaire obtenu après 15 h d'induction, ainsi que celle des fractions

d'élutions issues de la purification sur colonne de Ni-NTA Agarose (pistes 5 et 7) montre que le taux de protéine Pus7 soluble pouvait être satisfaisant dans l'objectif de réaliser des travaux de cristallogenèse. Néanmoins, la présence de nombreux contaminants de poids moléculaire plus faible était très peu favorable pour ce type d'expérience.

Nous avons testé si l'utilisation d'un milieu autoinductible contenant du glucose, du glycérol et du lactose (Studier, 2005) dans les conditions décrites dans la partie Matériel et Méthodes pouvait permettre d'éviter la présence des sous fragments de Pus7p. En plus de ne pas nécessiter d'étape d'induction, l'utilisation d'une faible concentration de lactose dans le milieu permettait de produire les protéines recombinantes de manière plus progressive qu'en utilisant l'IPTG ce qui pouvait limiter l'existence de traduction abortive. Comme nous pouvons le constater dans la figure 79, nous n'avons pas observé d'effet drastique sur la qualité de la protéine obtenue après purification sur colonne Ni-NTA Agarose (piste 5). Le seul avantage était l'obtention d'une meilleure biomasse dans ce milieu (36 g de bactéries par litre de milieu contre 19g avec le milieu LB). Nous avons donc utilisé ce milieu autoinductible pour la suite des essais.

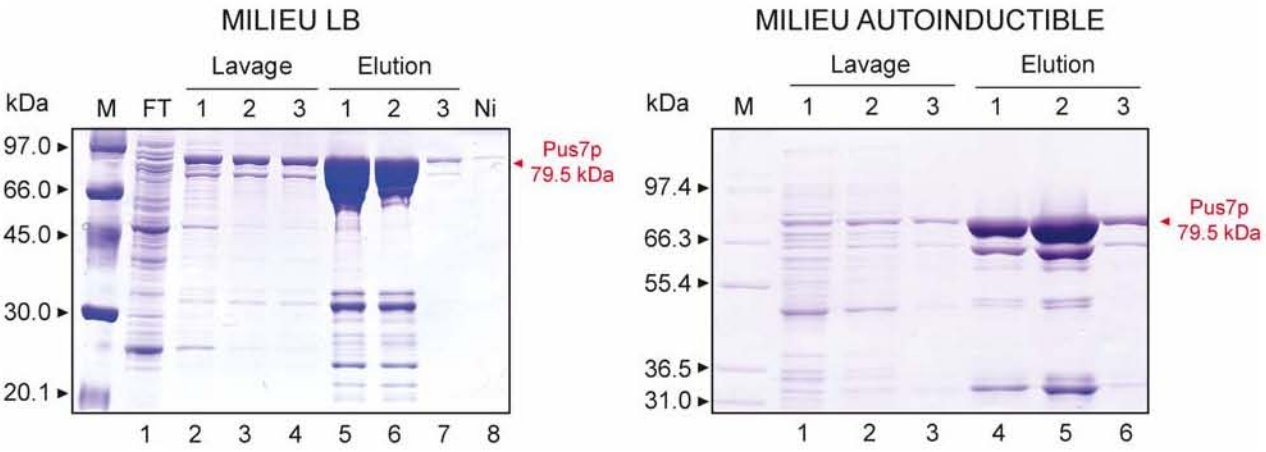


Figure 79 : Analyse du contenu en protéines des fractions collectées au cours de la purification sur colonne de Ni-NTA Agarose de l'enzyme His_6-Pus7p sauvage produite en milieu LB ou en milieu autoinductible.

A. Les bactéries transformées par le plasmide pET28b-His_6-Pus7 cultivées en milieu LB et induites par traitement à l'IPTG ou cultivées en milieu autoinductible ont été lysées par sonication. La fraction soluble du lysat (S15), obtenue après centrifugation, a été mélangée à 0,5 ml de billes de Ni-NTA. Les protéines Pus7-His_6 contenues dans la fraction non retenue sur les billes (piste FT) et les différentes fractions de lavage (pistes 1 à 3) et d'élution (pistes 4 à 6) ont été séparées par électrophorèse en gel dénaturant (SDS-PAGE) à 12,5%. La piste Ni contient un aliquote de résine de Ni-NTA. Un marqueur de taille contenant plusieurs protéines dont le poids moléculaire est indiqué à gauche de la figure a été déposé dans la piste M. 8µl de marqueur de protéines ont été déposés dans chaque piste.

Puisque le changement de milieu et des conditions de culture ne permettait pas de diminuer le nombre de sous fragments de Pus7p, j'ai essayé de produire la protéine dans d'autres souches bactériennes, comme les souches *E. coli* BL21 C41(DE3), C43(DE3) et (DE3) pLysS (Novagen). Cette stratégie n'a pas non plus permis d'améliorer la qualité de l'échantillon en sortie de colonne Ni-NTA Agarose et en plus les rendements de production étaient inférieurs à ceux obtenus avec la souche *E. coli* BL21-Codon Plus (données non présentées). Afin de limiter l'éventuelle protéolyse de la protéine d'intérêt qui intervient le plus souvent lors de la lyse cellulaire, j'ai également testé l'effet de l'ajout d'inhibiteurs de protéases (inhibiteur de la trypsine ou PMSF) dans les tampons de purification. J'ai aussi testé l'efficacité d'un cocktail anti-protéase à large spectre d'action (Sigma). Les résultats n'étaient pas satisfaisants ce qui était plus en faveur d'événements de traduction abortive. J'ai donc entrepris d'optimiser les conditions de purification en utilisant différentes chromatographies pour éliminer les sous produits contaminants.

5.2. Optimisation des conditions de purification

Dans un premier temps, j'ai essayé d'éliminer les contaminants de Pus7p sur des critères de charge nette en complétant la première étape de purification de Pus7p par une étape de chromatographie sur différentes colonnes échangeuses d'ions HiTrap Q et HiTrap SP décrites dans le Tableau 17.

Type de résine	Dénomination	Groupement	pH	Tampon
Echangeuse d'anions	Hitrap Q XL	Ammonium quaternaire	9,9	Borate
			8,9	Tris-HCl
Echangeuse de cations	Hitrap SP XL	Sulfopropyle	6,9	Phosphate
			5,9	Phosphate
Exclusion de taille	Superdex 200	-	7,5	Tris-HCl

Tableau 17 : Colonnes de chromatographie testées pour la purification de l'enzyme Pus7p et les tampons utilisés.

Etant donné que Pus7 interagit avec différents ARN, nous pouvions supposer qu'au moins certaines surfaces de cette protéine soient chargées positivement à pH neutre, ceci était en accord avec le point isoélectrique théorique de cette protéine (pI=7,9). Nous avons en effet constaté que Pus7 est fortement retenue sur colonne échangeuse d'anions et ceci pour les 2 conditions de pH utilisées (pH 8,9 et 9,9) (Fig. 80A). Elle est en effet éluée à des concentrations élevées en NaCl (780 mM à pH 8,9 et 820 mM à pH 9,9). Cependant l'analyse sur SDS-PAGE des fractions collectées montrait que ce type de chromatographie ne permettait pas d'éliminer les contaminants de Pus7p et à pH 9,9 très peu de protéine était collectée probablement du fait d'une agrégation sur la matrice de la colonne. L'utilisation d'une colonne échangeuse de cations ne s'est pas avérée apporter de meilleurs résultats (Fig.

80B). Pus7p n'était pas retenue sur la matrice à pH 6,9 en tampon phosphate (Fig. XB Panneau du haut) et bien qu'elle se fixait efficacement à pH 5,9, la protéine éluée à une concentration d'environ 670 mM en NaCl contenait toujours la bande d'environ 66 kDa. Ces données renforçaient l'idée que cette bande correspond à une protéine Pus7p tronquée, dont les propriétés physico-chimiques étaient très voisines de celles de Pus7p.

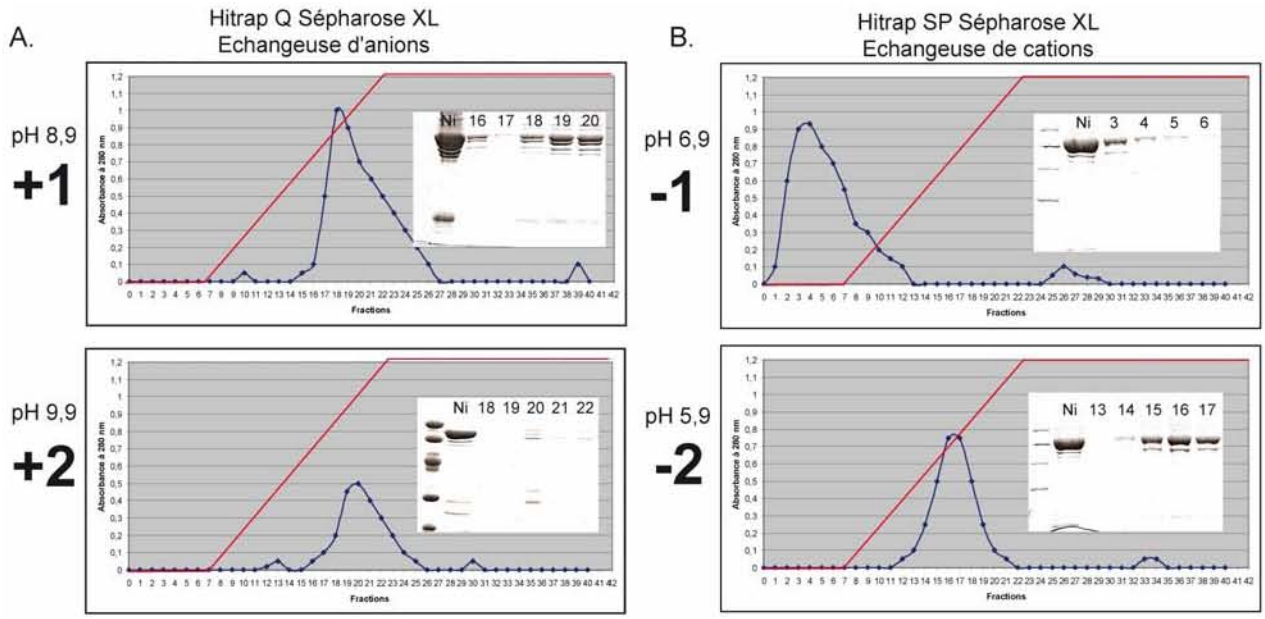


Figure 80 : Purification de la protéine Pus7p de *S. cerevisiae* sur colonnes échangeuses d'ions.

Profil d'élution de la protéine Pus7 purifiée sur colonne Hitrap Q Sépharose XL (A) ou Hitrap SP Sépharose XL (B) par utilisation d'un gradient continu de NaCl 1M dont le volume global correspondait à 10 fois le volume de la colonne. Les conditions de pH utilisées sont indiquées à gauche des profils, la valeur en gras indique l'écart en unité de pH avec le point isoélectrique prédit de la Pus7p. L'analyse des fractions a été réalisée par électrophorèse SDS-PAGE sur un gel à 12,5%. Dix µl de protéine purifiée ont été déposée dans chaque piste. La piste Ni correspond à 10 µl de protéine issue de la première étape de purification par chromatographie d'affinité sur résine Ni-NTA Agarose.

Puisque le critère de la charge nette ne pouvait être utilisé pour éliminer les contaminants co-purifiés avec la protéine Pus7p, lors de l'étape de chromatographie d'affinité, nous avons donc essayé une séparation selon le poids moléculaire (et donc la taille de la protéine), en réalisant une chromatographie d'exclusion de taille (gel filtration) sur une colonne Superdex 200 préparative (GE-Healthcare). L'analyse par gel SDS-PAGE du chromatogramme a révélé la présence de Pus7p dans un pic correspondant à un volume d'élution de 11,42 ml (Fig. 81B), ce qui correspond à une masse molaire d'environ 80 kDa si l'on se réfère à la droite établie en utilisant le kit de calibration (Fig. 81A). Cette valeur est compatible la masse molaire de 79 523 Da calculé à partir de la séquence. Ce résultat nous indiquait que la protéine n'était pas présente sous forme d'un hétéro-dimère. Cependant

l'analyse en gel SDS-PAGE des fractions correspondant au pic montrait toujours la présence des bandes contaminantes correspondant à des protéines de taille inférieure à Pus7p (Fig. 81B). Ce type de chromatographie n'était donc pas assez résolutif.

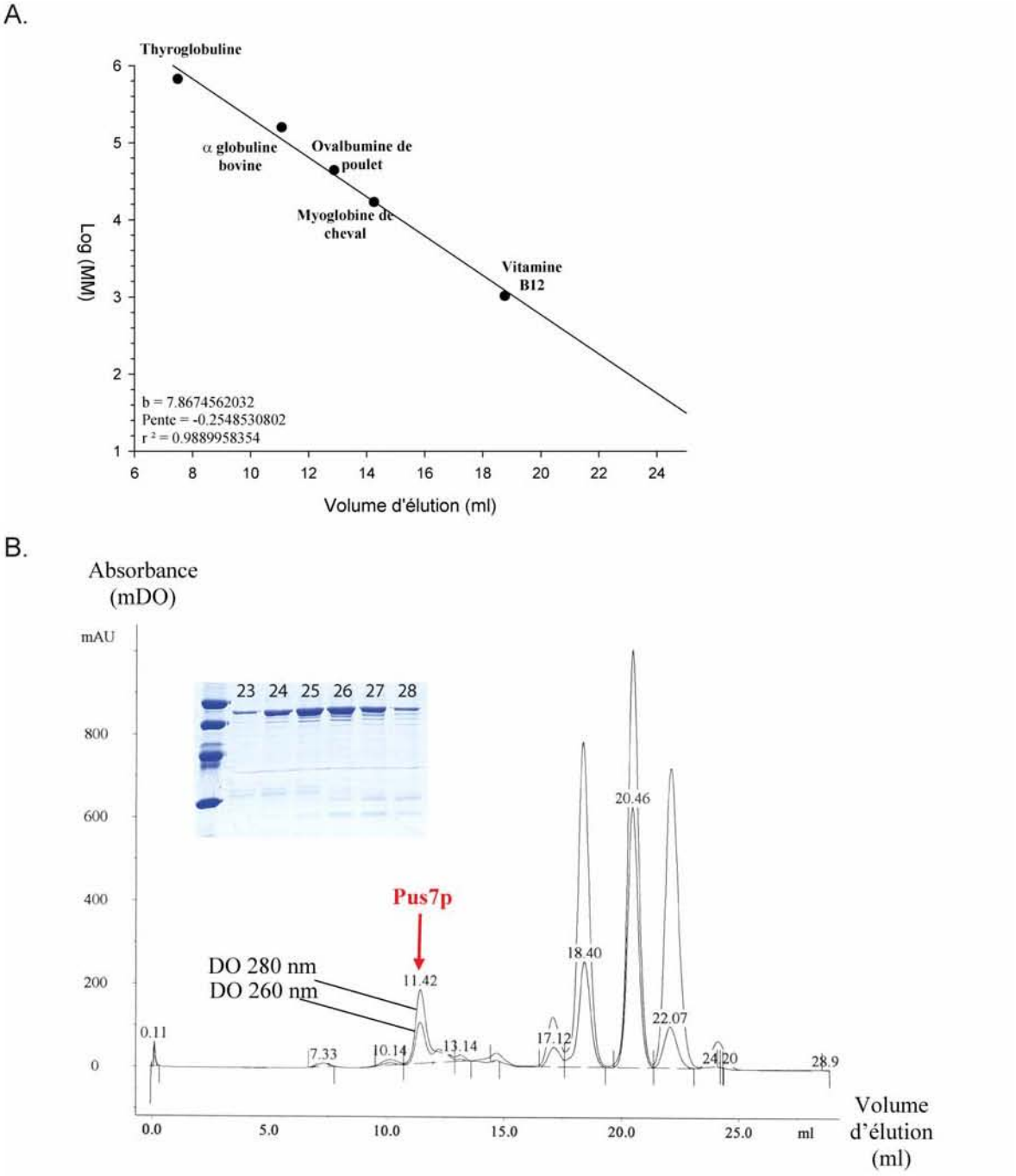


Figure 81 : Fractionnement par chromatographie d'exclusion de taille sur colonne Superdex 200 (Ge-Healthcare) de la protéine Pus7 de *S. cerevisiae* obtenue après purification sur colonne Ni-NTA Agarose.

A. Courbe d'étalonnage de la colonne en tampon 50 mM Tris-HCl pH 7,5 ; 1M NaCl. L'équation de la droite obtenue est log(MM) = -0,26 Ve + 7,87.

B. Chromatogramme enregistré en sortie de colonne. Le volume auquel est éluée Pus7p est indiqué par une flèche et les fractions collectées ont été analysées en gel SDS-PAGE à 12,5 % des fractions 23 à 28 correspondant au pic d'absorbance à 280 nm. La flèche rouge indique le volume auquel est éluée la protéine Pus7p.

Bien que la pureté de la protéine Pus7 ne soit pas optimale, nous avons réalisé une série de tests de cristallisation de cette protéine purifiée selon les 2 étapes que nous avons décrites précédemment et concentrée à environ 50 mg/ml. Cependant, ces tests n'ont pas abouti, probablement à cause du manque d'homogénéité de cette protéine. Par contre, comme nous avions observé que le mutant His_6-Pus7Δ5 conserve son activité vis-à-vis du $ARNt^{Asp}$(GUC) et du pré-$ARNt^{Tyr}$(GΨA) et qu'il semblait donner lieu à moins de sous produits après purification sur colonne de Ni-NTA Agarose (Fig. 77A), j'ai alors essayé de produire cette protéine en grande quantité afin de réaliser des tests de cristallisation.

5.3. Production, purification et cristallisation du mutant Pus7Δ5 de *S. cerevisiae*

J'ai appliqué le protocole de production (milieu autoinductible) et de purification (première étape de chromatographie d'affinité sur colonne Ni-NTA Agarose suivie d'une seconde étape de gel filtration sur colonne Superdex 200 préparative) que nous avions mis au point pour Pus7p. J'ai ainsi obtenu cette protéine très concentrée (environ 45 mg/ml) et à un fort degré de pureté (Fig. 82).

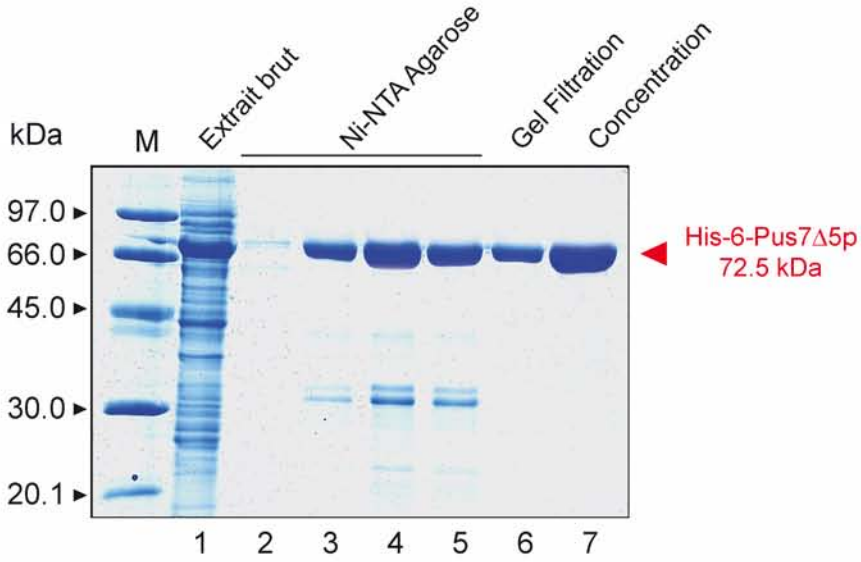


Figure 82 : Analyse du contenu en protéines des fractions collectées au cours de la purification sur colonne de Ni-NTA Agarose et sur colonne de gel filtration Superdex 200 de l'enzyme His_6-Pus7Δ5p produite en milieu autoinductible.

Electrophorèse en gel de polyacrylamide dénaturant à 12,5%. Les bactéries ayant produit les protéines possédant une étiquette His6 ont été lysées par sonication, centrifugées puis le surnageant (piste 1) a été purifié sur colonne Ni-NTA Agarose (pistes 2 à 5) puis sur colonne de gel filtration Superdex 200 (piste 6) et finalement concentré (piste 7). Cinq µl de la fraction de lavage (piste 2) et des fractions d'élution les plus concentrées ont été déposés (pistes 3 à 6) sauf dans la piste 7 où 0,5 µl de la fraction à l'issue de la concentration a été déposée. Le gel a été coloré au bleu de Coomassie. La piste M

contient un marqueur précoloré contenant plusieurs protéines dont le poids moléculaire est indiqué à gauche de la figure.

Il fallait alors essayer d'obtenir des cristaux exploitables par diffraction des rayons X. Pour cela, j'ai testé avec C. Charron un très grand nombre de conditions de cristallisation (918 au total), en utilisant 9 kits différents (Hampton Research Screens I and II, Hampton Research PEG/Ion Screen, Hampton Research SaltRX Screen, Molecular Dimensions Limited Clear Strategy Screens I and II, Molecular Dimensions Limited 3D Structure Screen I and II, deCODE Genetics Wizard Screens I and II and Jena BioSciences JBScreen 1-10). Dans chaque condition 100 nl de protéine à 40 mg/ml ont été mélangées à 100 nl de solution de cristallisation, ceci en utilisant un robot de type Mosquito (TTPLabtech). Ces nombreux tests de conditions de cristallisation nous ont permis d'obtenir des cristaux (Fig. 83) dans les conditions suivantes : 0.2 M Sulfate de Lithium monohydrate, 0.1 M Tris hydrochloride pH 8.5, 30% w/v Polyethylene glycol 4000, [Pus7Δ5p] = 43.5 mg/mL.

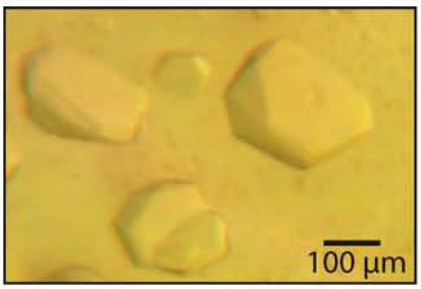


Figure 83 : Cristaux de la protéine Pus7Δ5 obtenus par diffusion de vapeur en goutte assise en présence de PEG4000.

Cependant, l'utilisation d'un rayonnement synchrotron a montré que ces cristaux diffractent seulement à basse résolution (environ 4 Å). Les tentatives d'amélioration de leur pouvoir diffractant par la modification des paramètres de cristallisation, et notamment l'ajout d'additifs de cristallisation, n'ont pas abouti. La protéine comporte à l'extrémité N-terminale une étiquette polyhistidine (contenant les 6 acides aminés histidine, un site de clivage à la thrombine et une courte séquence de liaison) qui constitue une séquence d'une vingtaine de résidus non structurée, ce qui pouvait être défavorable à un bon empilement cristallin. L'élimination de cette étiquette par protéolyse n'a pas néanmoins permis d'obtenir des cristaux de meilleure qualité. Nous avons alors entrepris la production de cette protéine sous forme fusionnée avec la Maltose Binding Protein (MBP), en espérant que cette protéine de fusion soit plus propice à la cristallisation mais ce système ne s'est pas avéré performant pour produire les quantités de protéines nécessaires à la recherche des conditions de cristallisation. D'autres pistes devront être testées en vue d'obtenir des cristaux exploitables en diffraction.

PARTIE 4

ETUDE DU ROLE FONCTIONNEL POSSIBLE DES RESIDUS Ψ DANS LE SNRNA U2 DE *S. CEREVISIAE*

1. Contexte

Comme nous l'avons vu dans l'introduction bibliographique, le snRNA U2 est un composant essentiel de la machinerie d'épissage des ARN-prémessagers nucléaires eucaryotes. En effet, il participe à la formation de la structure assurant la catalyse au sein du spliceosome, ceci par ses interactions avec le snRNA U6 et le pré-ARNm (Fig. 84). Cette structure de type ribozyme catalyse les deux étapes successives de trans-estérification de la réaction d'épissage.

La comparaison du contenu en nucléotides modifiés des snRNA U2 de plusieurs eucaryotes avait révélé la présence d'un faible nombre de résidus Ψ dans le snRNA U2 de levure par rapport aux snRNA U2 de vertébrés (14 chez l'homme et 16 chez le rat contre 3 chez *S. cerevisiae*) (Massenet et al., 1998). Point intéressant, les 3 résidus Ψ présents dans le snRNA U2 de levure sont très conservés phylogénétiquement et sont localisés dans un segment de cet ARN fonctionnellement important pour la réaction d'épissage. En effet, le résidu Ψ35 est localisé dans le segment du snRNA U2 s'appariant à la boîte de branchement de l'intron et les résidus Ψ42 et Ψ44 sont localisés juste en amont dans un segment du snRNA U2 fixant des facteurs protéiques essentiels à la réaction d'épissage. D'où l'hypothèse, formulée par l'équipe de C. Branlant, que la modification de ces 3 résidus joue un rôle important dans la réaction d'épissage. Des études structurales, réalisées par RMN, sur l'appariement formé entre le fragment du snRNA U2 s'appariant à la boîte de branchement et un segment d'intron contenant la boîte de branchement, confortent cette hypothèse, puisqu'elles établissent que la présence du résidu Ψ35 favorise la conformation extra-hélicale du résidu A de la boîte de branchement lors de l'attaque nucléophile du phosphate en 5' de l'intron par l'hydroxyle en 2' de ce résidu (première étape de la réaction) (Newby & Greenbaum, 2001; Lin & Kielkopf, 2008).

Deux de ces 3 résidus conservés, Ψ35 (Ψ34 chez les vertébrés) et Ψ44 (Ψ43 chez les vertébrés) sont respectivement formés par les ARN:Ψ-synthases Pus7p (Ma et al., 2003) et Pus1p (Massenet et al., 1999), tandis que le résidu Ψ42 (Ψ41 chez les vertébrés) est formé par l'enzyme Cbf5p couplée à l'ARN guide snR81 (Ma et al., 2005).

Bien que ces 3 résidus Ψ soient strictement conservés au cours de l'évolution, les systèmes responsables de leur formation ne sont pas identiques chez toutes les espèces. En effet, le résidu Ψ34 du snRNA U2 de *X. laevis* est formé par un système guidé comportant un scaRNA H/ACA (Zhao et al., 2002), alors que le résidu équivalent chez la levure (Ψ35) est formé par une enzyme n'utilisant pas d'ARN guide (Pus7p) (Ma et al., 2003) (Fig. 84).

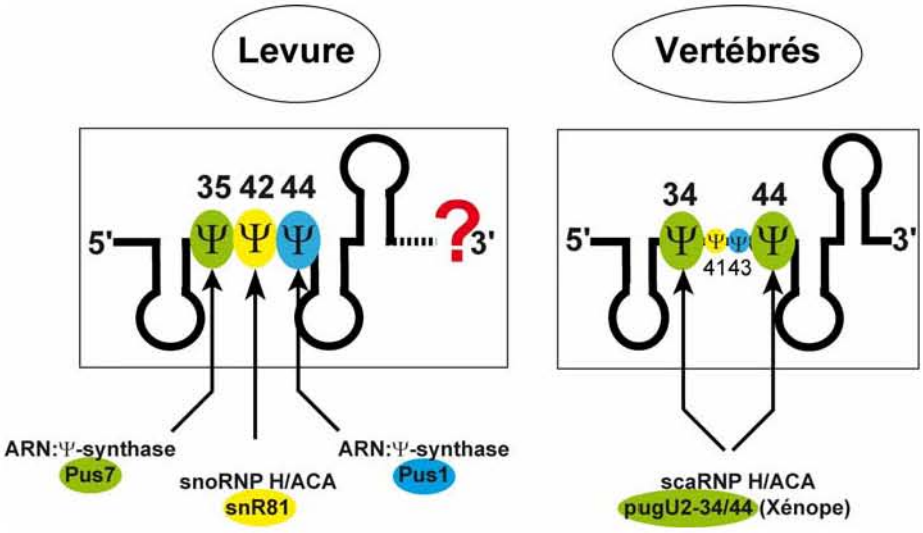


Figure 84 : Localisation respective des résidus Ψ dans le snRNA U2 de la levure *S. cerevisiae* et de vertébrés et les systèmes enzymatiques les générant.

Les systèmes enzymatiques responsable de la formation des résidus Ψ35, Ψ42 et Ψ44 chez *S. cerevisiae* ainsi que ceux responsables de la formation des résidus Ψ34 et Ψ44 chez le xénope sont indiqués.

2. Démarche expérimentale

Comme les systèmes enzymatiques responsables de la formation de trois résidus Ψ du snRNA U2 de *S. cerevisiae* avaient été identifiés, nous pouvions débuter une étude de l'importance fonctionnelle de ces 3 résidus pour la réaction d'épissage, ceci en abolissant leur formation. L'importance fonctionnelle du résidu Ψ35 avait déjà été étudiée *in vivo* chez *S. cerevisiae* en cumulant la délétion du gène de l'enzyme Pusp7 et des mutations dans le snRNA U2 à des positions proches de la position 35. Dans ces conditions, un ralentissement de l'efficacité d'épissage de certains pré-ARNm avait été observé (Yang et al., 2005). Mais la délétion des systèmes catalysant la formation des résidus Ψ du snRNA U2 n'avait pas été réalisée.

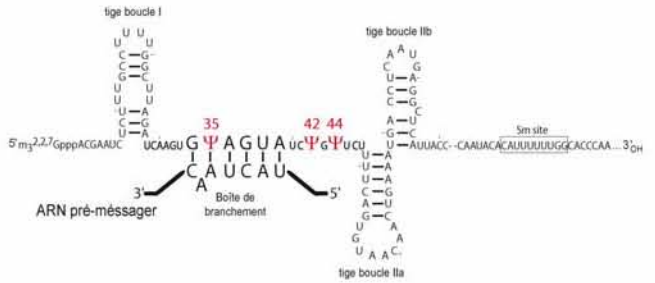
Dans l'objectif de réaliser une telle étude avec analyse globale de l'importance des 3 résidus Ψ du snRNA U2 sur l'épissage des pré-ARNm de *S. cerevisiae*, nous avons construit plusieurs souches portant des délétions individuelles ou multiples des gènes *PUS1*, *PUS7* codant les enzymes Pus1p et Pus7p, et aussi du gène du snoRNA H/ACA *snR81* guidant la modification à la position 42. Les délétions multiples ont été réalisées par transformation directe et par croisement des souches portant une simple délétion, les mutants haploïdes étant ensuite obtenus par microdissection des spores. Nous avons ainsi pu obtenir des souches présentant une délétion de deux des gènes (tous les couples ont été obtenus), ainsi qu'une souche où les 3 gènes étaient délétés.

Pour tester de manière globale dans la levure *S. cerevisiae* l'effet de l'absence de l'un, de deux ou des 3 résidus Ψ du snRNA U2 sur l'efficacité relative d'épissage des différents pré-ARNm contenant un intron épissé par le spliceosome, nous avons établi, au sein du réseau Européen EURASNET, une collaboration avec l'équipe de J. Beggs (Edinburgh University). Cette équipe avait construit une puce à ADN permettant de tester en parallèle les taux d'épissage de pré-ARNm nucléaires de cet organisme.

Contrairement à la situation observée chez les vertébrés, la séquence de la boîte de branchement des introns d'ARN pré-messagers de *S. cerevisiae* est très conservée et le nombre de gènes contenant un intron chez *S. cerevisiae* est limité (320 gènes contiennent un intron sur environ 5800 gènes au total). On observe en général un unique intron par gène, seuls 8 gènes contiennent 2 introns. Sur le faible nombre d'introns répertoriés dans les pré-ARNm de *S. cerevisiae,* seuls 19 contiennent une boîte de branchement dont la séquence dévie à une ou maximum 2 positions par rapport à la séquence consensus établie à partir de l'ensemble des introns (Fig. 85).

Nous pensions que la présence des résidus Ψ dans le snRNA U2 pouvait être plus importante pour ces introns, que pour ceux contenant une séquence de branchement canonique car des études réalisées par mutagenèse dirigée de boîte de branchement des introns de levure montraient que les mutations limitant l'appariement avec le snRNA U2 diminuent fortement l'efficacité d'épissage (Langford et al., 1984; Jacquier et al., 1985; Jacquier & Rosbash, 1986; Vijayraghavan et al., 1986; Parker et al., 1987; Seraphin et al., 1991).

A.



B.

		Séquence de la boîte de branchement						
ORF	Protéine	U	A	C	U	A	A	C
YBR189W	RPS9B	G	A	C	U	A	A	C
YBR219C	Inconnue	G	A	C	U	A	A	C
YBR230C	Inconnue	G	A	C	U	A	A	C
YDR381W	YRA1	G	A	C	U	A	A	C
YGL251C	HFM1	G	A	C	U	A	A	C
YOR235W	Inconnue	G	A	C	U	A	A	C
YEL012W	UBC8	C	A	C	U	A	A	C
YLR333C	RPS25B	C	A	C	U	A	A	C
YML067C	Inconnue	C	A	C	U	A	A	C
YPL081W	RPS9A	C	A	C	U	A	A	C
YKR057W	RPS21A	A	A	C	U	A	A	C
YER003C	PMI4	A	A	C	U	A	A	C
YBR084CA	RPL19A	U	A	A	U	A	A	C
YPL241C	cin2	U	A	A	U	A	A	C
YGR029W	ERV1	U	A	U	U	A	A	C
YIL106W	MOB1	U	A	U	U	A	A	C
YGL076C	RPL7A	U	G	C	U	A	A	C
YPL198W	RPL7B	U	G	C	U	A	A	C
YHR041C	srb2	U	G	C	U	A	A	C

Figure 85 : Lien possible entre la présence de résidus Ψ dans le snRNA U2 et la stabilité de son appariement avec la boîte de branchement des introns de *S. cerevisiae*.

A. Représentation de l'appariement formé entre le snRNA U2 et un ARN pré-messagers de la levure *S. cerevisiae*. Les résidus Ψ détectés dans le snRNA U2 sont indiqués en rouge.

B. Introns de *S. cerevisiae* possédant une boîte de branchement divergeant du consensus établi pour les boîtes de branchement des introns de levure. La séquence consensus est indiquée en haut du tableau, les déviations par rapport à la séquence consensus sont indiquées en rouge.

3. Résultats

A. Production et caractérisation des souches de *S. cerevisiae* portant des délétions individuelles ou multiples des gènes *PUS1, PUS7* et *snR81*

1. Souche ΔPus1ΔPus7

1.1. Approche utilisée pour interrompre les ORF Pus1 et Pus7 dans une même souche

Saccharomyces cerevisiae est un champignon ascomycète hétérothallique dont le cycle cellulaire est haplodiplobiontique. Ainsi, les cellules peuvent se maintenir sous forme haploïde ou diploïde par bourgeonnement. En conditions de croissance défavorables (carence en source de carbone ou d'azote), les cellules diploïdes sporulent et donnent naissance à un asque contenant quatre spores non ordonnées, dont deux de signe sexuel a et deux de signe sexuel α. Lorsque les conditions redeviennent plus favorables, ces spores germent et forment des cellules haploïdes. Deux cellules de signes opposés peuvent être croisées afin de générer une souche diploïde. Toutes ces étapes peuvent être provoquées en laboratoire.

La souche contenant la double délétion des gènes PUS1 et PUS7 a été réalisée par croisement, sporulation puis dissection de 2 souches haploïdes qui étaient disponibles au laboratoire, ceci en utilisant le protocole décrit dans le chapitre Matériels et Méthodes. En bref, les souches de signe sexuel opposé BY4742 ΔPus1 (MATα his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0; YPL212C:kanMX4) et BY4741 ΔPus7 (MATa his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YOR243c:kanMX4) ont été cultivées en milieu liquide jusqu'en phase de croissance exponentielle, puis nous avons réalisé leur croisement sur boîte en milieu riche YPD. Nous avons constaté que les zygotes de forme caractéristique apparaissent après 3 à 4 heures et sont facilement distinguables au microscope. Les cellules diploïdes en phase exponentielle de croissance ont ensuite été étalées sur milieu riche YPD et incubées 24 heures à 30°C avant d'être transférées en milieu liquide permettant la sporulation et mise sous agitation sur roue motorisée à 20°C. Les premiers asques sont apparus au bout de deux à trois jours et ont été analysés par dissection de tétrades. D'après les lois de la génétique, les tétrades disséquées devraient présentées une ségrégation de type 2 spores parentales/2 spores recombinées dont l'une correspondait à la souche sauvage et l'autre à la souche doublement délétée (Fig. 86).

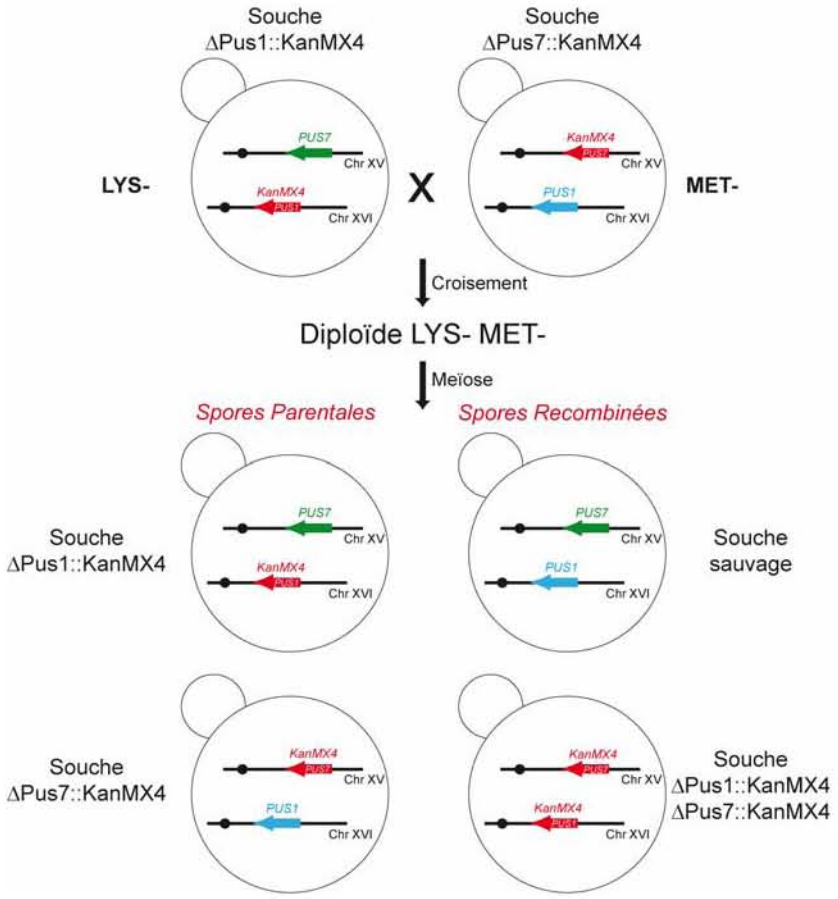


Figure 86 : Ségrégation attendue dans le croisement de la souche haploïde ΔPus1::KanMX4 et de la souche ΔPus7::KanMX4.

Cette technique permet donc l'étude qualitative mais aussi quantitative des produits de la méiose (Sherman & Hicks, 1991). La première étape consiste à observer au microscope la formation des tétrades. Lors de ces expériences, nous avons obtenu un faible pourcentage de tétrades (estimé à 10%). Il était néanmoins possible de réaliser leur dissection. Pour cela, un peu du mélange de sporulation a été mis en suspension dans de l'eau stérile à laquelle la zymolyase était ajoutée. Une fois la paroi des asques totalement hydrolysée par l'enzyme, les spores sont libérées mais restent regroupées en tétrades. Nous avons sont ensuite aligné une dizaine de tétrades sur une boîte de milieu complet, grâce à un micromanipulateur Singer de type MSM. Au bout de 48 heures à 30°C 6 tétrades, formées chacune de 4 spores, sont apparues. Ces spores ont ensuite été analysées par différentes techniques que nous présenterons dans le chapitre suivant.

1.2. Analyse phénotypique et génotypique de la souche ΔPus1ΔPus7

Nous avons alors testé le phénotype de chaque spore par leur repiquage sur différents milieux sélectifs : LYS^-, MET^-, LYS^-MET^- et sur un milieu contenant l'antibiotique généticine (G418). Comme nous pouvons l'observer dans la figure 87, l'analyse de la répartition des marqueurs d'auxotrophie et de résistance après microdissection des tétrades

montre bien une ségrégation de type 2 spores parentales / 2 spores recombinées pour les séries 2, 5 et 6. Au contraire, on constatait que les séries 1 et 3 donnaient lieu à 4 colonies capables de se développer en présence de généticine. Nous en avons conclu que les colonies observées ne correspondaient pas aux 4 spores d'une même asque, puisque lors d'une ségrégation normale, une colonie doit être de type sauvage et donc dans le cas qui nous concernait, devait être incapable de se développer sur milieu G418. De même, le génotype des spores de la série 4 mettait en évidence une répartition non attendue des marqueurs d'auxotrophie qui reflétait une ségrégation anormale. Ce crible rapide nous a donc permis d'éliminer les séries 1, 3 et 4 qui ne pouvaient pas contenir de cellules portant la double délétion ΔPus1ΔPus7. Nous avons néanmoins décidé de conserver la série 1 pour la suite des expériences en tant que contrôle négatif.

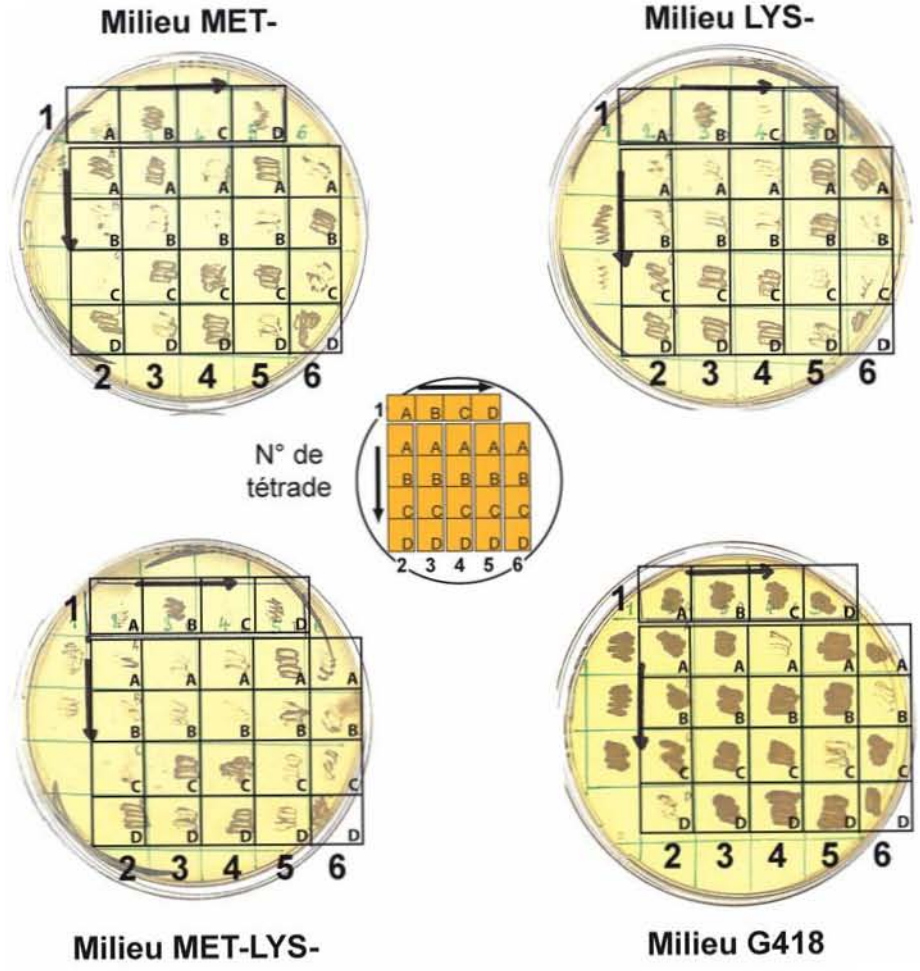
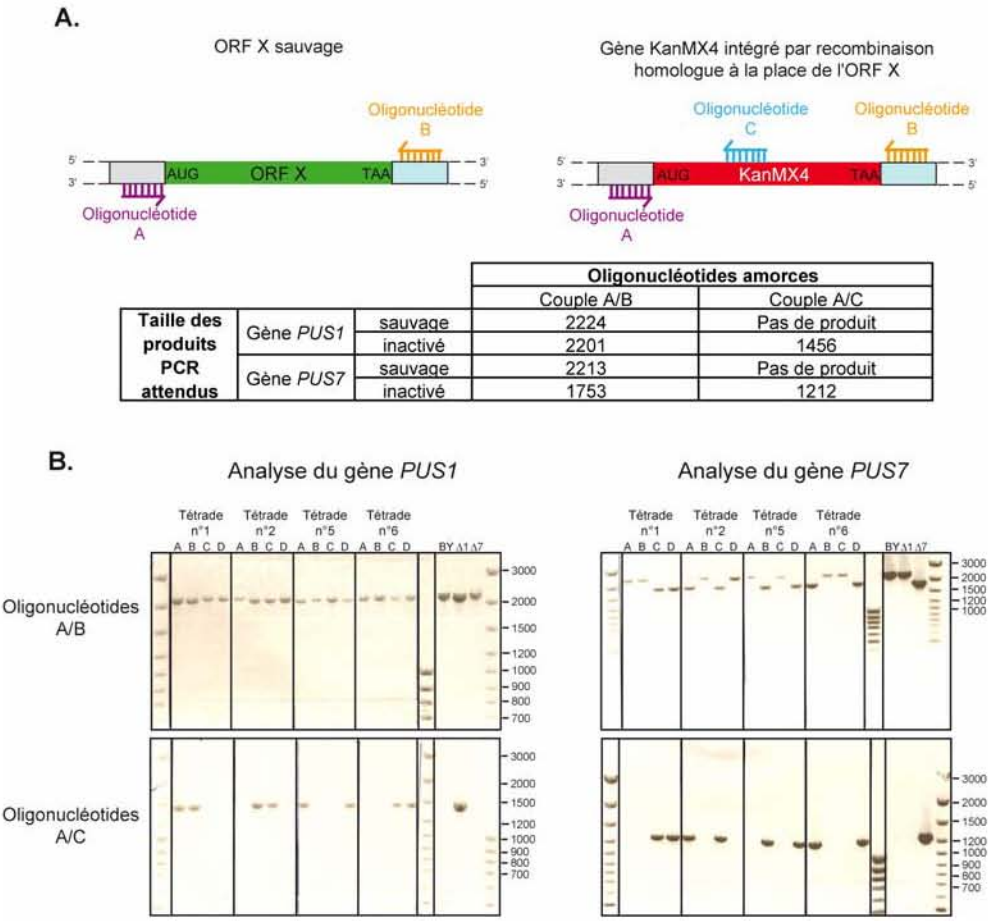


Figure 87 : Sélection de la souche doublement délétée ΔPus1::KanMX4/ΔPus7::KanMX4 de *S. cerevisiae* par étalement sur milieux sélectifs.

Analyse des phénotypes des tétrades obtenues après sporulation par étalement sur milieux sélectifs MET-, LYS-, MET-LYS- et sur milieu contenant l'antibiotique généticine (G418). Le schéma de dépôt des 6 séries de tétrade est indiqué au centre avec le numéro des tétrades (1, 2, etc…) et le n° de chaque spore (A, B, C et D).

Une autre façon de caractériser les spores obtenues consistait à analyser le génotype des tétrades sélectionnées par amplification de l'ADN génomique extrait de chacune des spores. Pour cela, nous avons utilisé 3 oligonucléotides (A, B et C) par gène à analyser, leur séquence est présentée dans le tableau 19. Ces oligonucléotides amorces permettaient non seulement de tester la présence ou l'absence de l'ORF à déléter (couple A/B), mais également de vérifier que la délétion était bien située au bon locus (couple A/C) (Fig. 88A).

Comme on peut le constater dans la Figure 88B, les tétrades 2, 5 et 6 contenaient bien 4 spores dont la ségrégation était conforme à celle attendue. En effet, les produits d'amplification obtenus en utilisant les couples d'oligonucléotides A/B et A/C, aussi bien pour le gène *PUS1* (panneau de gauche) que pour le gène *PUS7* (panneau de droite) confirmaient la présence de la substitution des ORF d'intérêt par les ORF KanMX4 dans les spores 2C, 5D et 6D. Comme attendu, aucune amplification n'est détectée avec le couple d'oligonucléotides A/C pour les spores sauvages. Nous avons donc conclu que nous avions bien obtenu 3 souches haploïdes portant la double délétion des ORF PUS1 et PUS7. Le tableau 16 fait la synthèse des résultats obtenus en rappelant le génotype et le phénotype de chaque spore analysée.



			Oligonucléotides amorces	
			Couple A/B	Couple A/C
Taille des produits PCR attendus	Gène *PUS1*	sauvage	2224	Pas de produit
		inactivé	2201	1456
	Gène *PUS7*	sauvage	2213	Pas de produit
		inactivé	1753	1212

Figure 88 : Caractérisation de la souche doublement délétée ΔPus1::KanMX4/ΔPus7::KanMX4 de *S. cerevisiae*.

A. Stratégie utilisée pour l'analyse des tétrades par PCR. Les oligonucléotides amorces et leur zone d'hybridation sont représentés sur le schéma au dessus du tableau qui indique la taille des fragments attendus.
B. Résultats de l'analyse des 4 tétrades sélectionnées. Le couple d'amorces spécifiques pour chaque gène est indiqué à droite. Les fragments PCR ont été séparés par électrophorèse en gel d'agarose à 1%, avec en parallèle un marqueur de taille et 3 pistes témoin correspondant à l'ADN génomique extrait de la souche sauvage (BY), de la souche ΔPus1::KanMX4 (Δ1) ou de la souche ΔPus7::KanMX4 (Δ7).

		Auxotrophie			Résistance	Analyse PCR		BILAN
		LYS	MET	LYS MET	G418	ΔPUS1	ΔPUS7	Génotype/phénotype de la spore
Série n°1	A	-	+	-	+	+	-	ΔPus1 Lys-
	B	-	+	-	+	+	-	ΔPus1 Lys-
	C	+	-	-	+		+	ΔPus7 Met-
	D	+	-	-	+	-	+	ΔPus7 Met-
Série n°2	A	-	+	-	+	-	+	ΔPus7 Lys-
	B	-	-	-	+	+	-	ΔPus1 Met-
	C	**+**	**-**	**-**	**+**	**+**	**+**	**ΔPus1ΔPus7 Met-**
	D	+	+	+	-	-	-	WT
Série n°3	A	-	+	-	+	nd	nd	nd
	B	-	-	-	+	nd	nd	nd
	C	+	+	+	+	nd	nd	nd
	D	+	-	-	-	nd	nd	nd
Série n°4	A	-	-	-	-	nd	nd	nd
	B	-	-	-	+	nd	nd	nd
	C	+	+	+	+	nd	nd	nd
	D	+	+	+	+	nd	nd	nd
Série n°5	A	+	+	+	+	+	-	ΔPus1
	B	+	-	-	+	-	+	ΔPus7 Met-
	C	-	+	-	-	-	-	WT Lys-
	D	**-**	**-**	**-**	**+**	**+**	**+**	**ΔPus1ΔPus7 Lys-Met-**
Série n°6	A	+	-	-	+	-	+	ΔPus7 Met-
	B	-	+	-	-	-	-	WT Lys-
	C	-	-	-	+	+		ΔPus1 Lys-Met-
	D	**+**	**+**	**+**	**+**	**+**	**+**	**ΔPus1ΔPus7**

Tableau 18 : Bilan des résultats obtenus lors de la construction de la souche ΔPus1ΔPus7.
Les séries ne présentant pas une ségrégation correcte ont été barrées en rouge.

Puisque nous disposions au laboratoire des outils requis, nous avons réalisé une vérification de contrôle des souches doublement délétées ΔPus1ΔPus7 par analyse fonctionnelle de l'activité ARN:Ψ-synthase dans plusieurs ARN.

1.3. Caractérisation de la souche ΔPus1ΔPus7 par analyse de l'activité ARN:Ψ-synthase sur les ARNt^{Asp}, ARNt^{Val} et pré-ARNt^{Tyr}

La vérification des souches dans lesquelles les gènes *PUS1* et *PUS7* ont été supposés être interrompus par le marqueur de résistance KanMX4 d'après les analyses phénotypiques et par PCR a été réalisée en testant l'absence des activités correspondantes dans les extraits cellulaires. Nous avons utilisé la méthode du plus proche voisin dans les conditions décrites dans le chapitre Matériel et Méthodes. Pour cela, nous avons produits des transcrits

uniformément marqués par incorporation d'[α-^{32}P]ATP pour les ARNt^{Asp}(GUC) et pré-ARN^{Tyr}(GΨA) et d'[α-^{32}P]CTP pour l'ARNt^{Val}(AAC) ce qui permettait d'analyser l'activité des protéines Pus7p (position 13 de l'ARNt^{Asp}(GUC) et 35 du pré-ARN^{Tyr}(GΨA)) et Pus1p (position 27 l'ARNt^{Val}(AAC)) respectivement (Fig. 89A). Point important, il faudra tenir compte pour l'analyse des résultats que, lors d'un marquage par incorporation d'[α-^{32}P]CTP, le taux de modification mesuré en extrait cellulaire correspond en fait au cumul de l'activité de 2 enzymes : Pus1p (position 27) et Pus4p (position 55) (Fig. 89A).

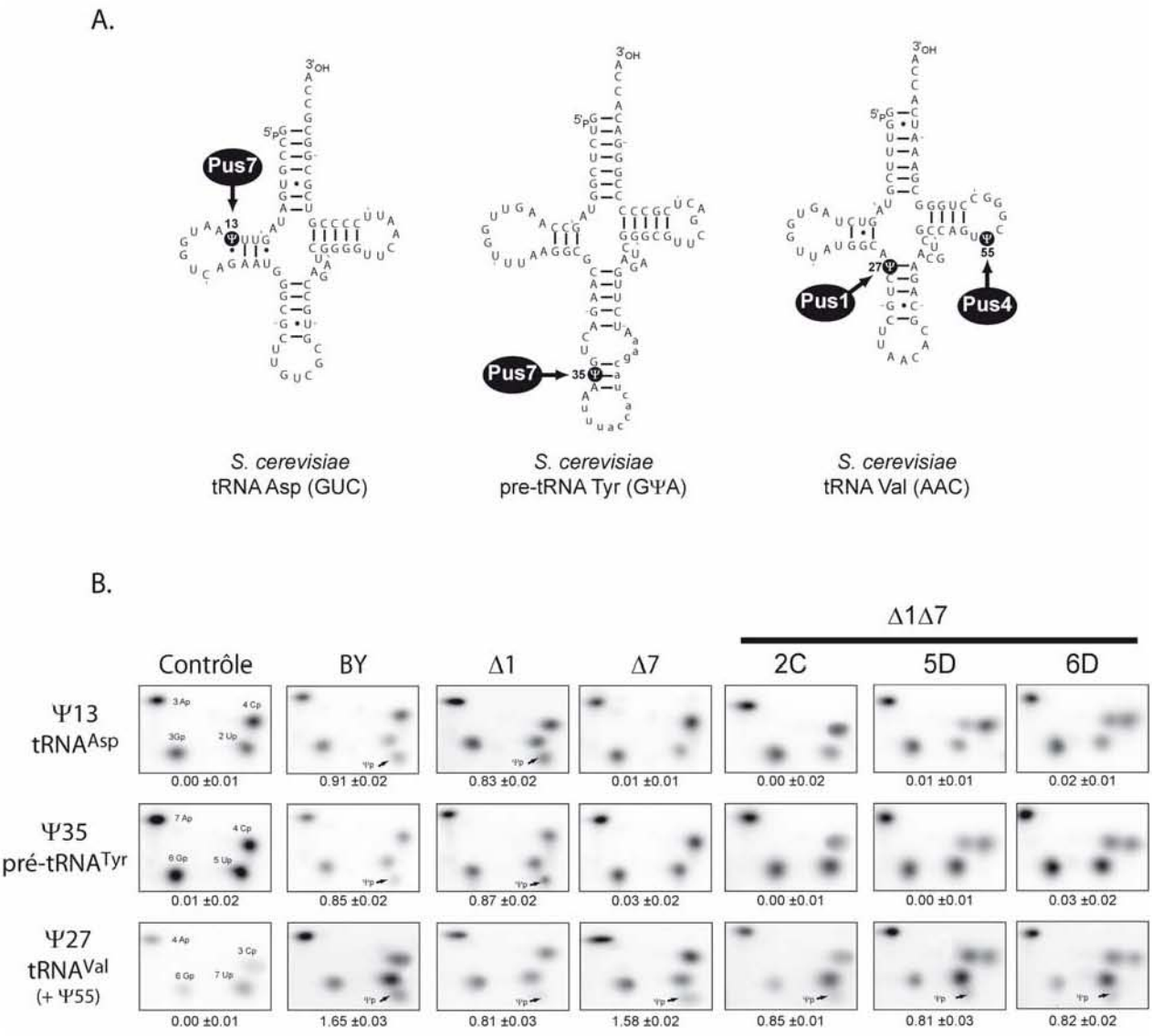


Figure 89 : Analyse de l'activité ARN:Ψ-synthase des extraits cellulaires issus des différentes souches mutantes de *S. cerevisiae* que nous avons sélectionnées.

A. Séquence et structure 2D des ARNt^{Asp}(GUC), pré-ARNt^{Tyr}(GΨA) et ARNt^{Val}(AAC) utilisé pour l'analyse de l'activité ARN:Ψ-synthase des souches sélectionnées. La position des résidus Y modifiés et des enzymes correspondantes est indiquée.

B. Fractionnement par chromatographies bidimensionnelles en couche mince des mono-nucléotides 3'-phosphates obtenus par hydrolyse à la RNase T2 des ARNt modifiés dans les extraits des souches

sauvage (WT), ΔPus1, ΔPus7 et ΔPus1ΔPus7 (2C, 5D et 6D). Le système de migration utilisé est le système N/R. L'identification des nucléotides a été faîte en se basant sur les cartes établies par Keith (Keith, 1995).

On constate dans la Figure 89B que les résidus Ψ13 de l'ARNt^{Asp}(GUC) et Ψ35 du pré-ARN^{Tyr}(GΨA) sont bien formés après incubation des transcrits radiomarqués correspondants avec des extraits issus de la souche sauvage (BY) mais pas avec des extraits issus de la souche ΔPus7 et des souches ΔPus1ΔPus7 2C, 5D et 6D. De la même manière, on constate que le taux de modification de l'ARNt^{Val}(AAC) (correspondant à la formation des résidus Ψ27 par Pus1 et Ψ55 par Pus4p) est de 1,65 mol Ψ/ mol ARN dans l'extrait de la souche BY alors qu'il est réduit à environ 0,80 mol Ψ/ mol ARN dans les souches dans lesquelles le gène PUS1 a été délété (souches ΔPus1, ΔPus1ΔPus7 2C, 5D et 6D). Ces résultats nous permettaient de conclure que les 3 souches que nous avions produites présentaient bien une double délétion des gènes PUS1 et PUS7. Nous avons sélectionné la souche ΔPus1ΔPus7 5D pour la suite de travail car elle présentait l'avantage d'être auxotrophe à la fois pour la lysine et la méthionine (Tableau 18) ce qui nous permettait d'utiliser un plus grand nombre de plasmides de complémentation.

2. Souches ΔsnR81, Δ1ΔsnR81, Δ7ΔsnR81 et Δ1Δ7ΔsnR81

2.1. Approche utilisée pour interrompre le gène codant le snoRNA H/ACA dans les différentes souches

Comme, nous l'avons mentionné précédemment, notre objectif était d'étudier les effets de l'absence d'un ou de plusieurs résidus Ψ dans le snRNA U2 sur l'épissage des ARN pré-messagers. Pour cela, il était important de disposer d'une collection complète de souches dans lesquelles chaque ORF et/ou ARN guide impliqué dans la pseudouridylation du snRNA U2 est inactivé individuellement ou simultanément. Puisque nous disposions déjà de la souche sauvage BY4742, des souches portant une simple délétion ΔPus1, ΔPus7 et de la souche portant une double délétion ΔPus1ΔPus7, l'étape suivante consistait à inactiver le gène codant le snoRNA H/ACA qui guide la pseudouridylation en position 42 du snRNA U2 dans ces 4 souches. Pour réaliser cette délétion, j'ai utilisé une stratégie différente de délétion, car il n'était pas possible de procéder par croisement/sporulation de différentes souches de levure, puisque la souche dans laquelle le snoRNA H/ACA snR81 était inactivé n'était pas disponible commercialement.

Différentes techniques existent pour abolir l'activité d'un gène chez *S. cerevisiae*, toutes font appel à la recombinaison homologue. Nous avons utilisé la méthode de remplacement en une étape décrite par Baudin *et al.* (1993). Cette technique est basée sur le fait que, chez la levure, la recombinaison homologue est très efficace avec des fragments

d'ADN linéaires. En effet, l'existence d'une homologie d'environ 40 pbs est suffisante pour avoir un taux de recombinaison homologue important. Pour tester facilement la présence de l'événement de recombinaison homologue, on peut remplacer l'ORF à éliminer par une ORF permettant de corriger une auxotrophie de la souche hôte. Nous avons utilisé l'ORF LEU2 de *S. cerevisiae*. Pour réaliser les substitutions, un fragment d'ADN dans lequel, l'ORF ou le gène de remplacement, était encadré par les séquences flanquant en 5' et en 3' l'ORF cible a été généré (Fig. 90). Ce fragment a été produit par PCR, en utilisant des oligonucléotides contenant à la fois une séquence permettant la recombinaison homologue et une séquence servant à l'amplification de l'ORF du gène de substitution. Nous avons amplifié l'ORF LEU2 à partir du plasmide pUC18-LEU2, que nous avons construit à partir du vecteur navette p413 (Mumberg et al., 1995). Les souches de *S. cerevisiae* BY4742 (MATα his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0), BY4742 ΔPus1 (MATα his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0; YPL212C:kanMX4), BY4741 ΔPus7 (MATa his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YOR243c:kanMX4) et BY4742 ΔPus1ΔPus7 (spore 5D) (MATa his3Δ1; leu2Δ0; lys2Δ0; met15Δ0; ura3Δ0; YPL212C:kanMX4 ; YOR243c:kanMX4) sous forme haploïde ont été utilisées pour réaliser les substitutions. Toutes ces souches ont la particularité d'avoir une délétion du gène *LEU2*, ce qui permet d'éviter les événements de recombinaison homologue au niveau du locus correspondant à ce gène. Après transformation avec les produits d'amplification, les levures ayant intégré dans leur génome le fragment d'ADN, ont été cultivées sur un milieu permettant de sélectionner les clones devenus prototrophes pour la leucine.

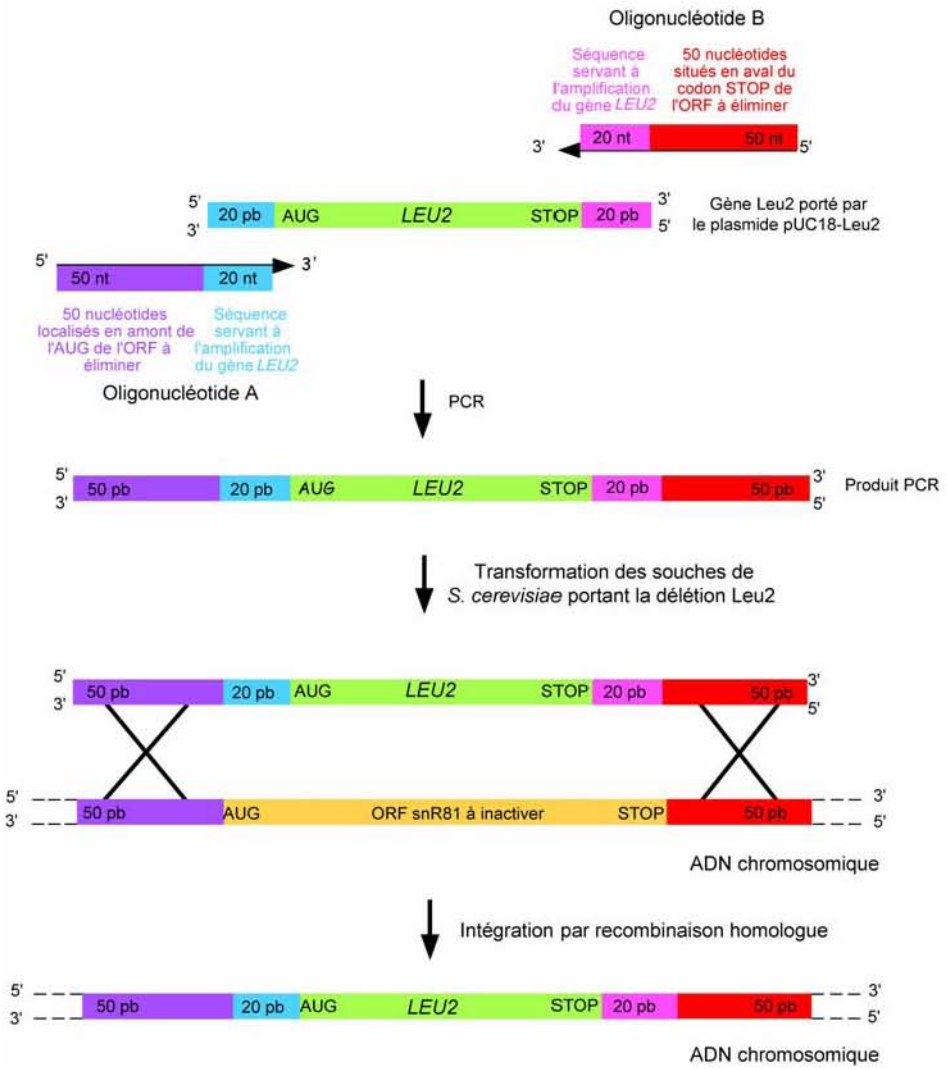


Figure 90 : Principe de la délétion du gène codant le snoRNA H/ACA snR81 réalisée par substitution de l'ORF d'intérêt par l'ORF LEU2 dans les différentes souches de *S. cerevisiae.* D'après (Baudin et al., 1993).

Les différents oligonucléotides utilisés lors des expériences de délétion sont répertoriés dans le Tableau 19 et les protocoles expérimentaux sont décrits dans la partie Matériel et Méthodes.

Numéro	Séquence 5'-->3'	Longueur	Description
3625	ATAGTCAATAGGTACAGCTCATCGATT	27	Oligo sens utilisé pour la vérfication de la délétion du gène Pus7
3626	CCTGAAGTGACACGACACGGACACAAC	27	Oligo antisens utilisé pour la vérfication de la délétion du gène Pus7
5242	CGATCCAATCACTAAACCGTC	21	Oligo sens utilisé pour la vérfication de la délétion du gène Pus1
5243	TGGTGGTGAGTAAGATGACAG	21	Oligo antisens utilisé pour la vérfication de la délétion du gène Pus1
1654	CCTCGACATCATCTGCCC	18	Oligo sens s'hybridant dans le gène KanMX4
5448	ACATGATGGAGTACCCTCATGCTTTGGCGCGTGCGCA TCTATATAGCGCGTTTCGGTGATGAC	63	Oligo sens utilisé pour la vérification de délétion du guide snR81
5449	GACTGAATCATATTACTTAGTTACTACTACTCTTATCAT GCAAGATTCCTGATGCGGTATTTTCTCCT	68	Oligo antisens utilisé pour la vérification de la délétion du guide snR81
4898	ATAAAGGACAATAAAGTGCTAGTAAATAACAATTATAAG TGATATCAAGGGCGCGTTTCGGTGATGAC	68	Oligo sens de délétion du gène Pus1 par LEU2
4899	AATGTCAATGCCTTAGAAATTAAAGTTGGTAAGAAAGA AGGAAAGGGCAATTCCTGATGCGGTATTTTCTCCT	73	Oligo antisens de délétion du gène Pus1 par LEU2
4900	ATAATCACGTTGTTCCTTCCCAAGGAAAAAAGCAATAT CTGTTGCATGCCGCGCGTTTCGGTGATGAC	68	Oligo sens de délétion du gène Pus7 par LEU2
4901	CCGACAAAAAAAGTAAACTAATATACATGAATATTGCAC CTCGTATAACCTTCCTGATGCGGTATTTTCTCCT	73	Oligo antisens de délétion du gène Pus7 par LEU2
5450	CATCAAACATGACGTTAGCG	20	Oligo sens de vérfication de la délétion du guide snR81
5451	GCACCTTCATCATCCTCATC	20	Oligo antisens de vérfication de la délétion du guide snR81
5452	GTGACCACGTTGGTCAAGAA	20	Oligo antisens de vérfication de la délétion snR81 s'hybridant dans le gene Leu2

Tableau 19: Description des oligonucléotides utilisés pour réaliser les substitutions des ORF de *S. cerevisiae,* codant les ARN:Ψ-synthases et le snoRNA H/ACA responsables de la formation des résidus Ψ dans le snRNA U2, et pour vérifier ces substitutions.

2.2. Caractérisation des souche sélectionnées après interruption du gène codant le snoRNA H/ACA snR81

Pour chacune des 4 souches BY4742, ΔPus1, ΔPus7 et ΔPus1ΔPus7 auxquelles nous avons appliqué la stratégie de délétion décrite ci-dessus, nous avons vérifié systématiquement que les clones sélectionnés sur milieu sélectif sans leucine présentaient bien la délétion souhaitée. Pour cela, nous avons analysé l'interruption du gène d'intérêt par PCR sur ADN génomique, en utilisant 2 couples d'oligonucléotides, le premier permettant de vérifier l'insertion du marqueur d'auxotrophie LEU2 (Couple A/B) et le second permettant vérifier le locus (Couple A/C) (Fig. 91A et Tableau 19).

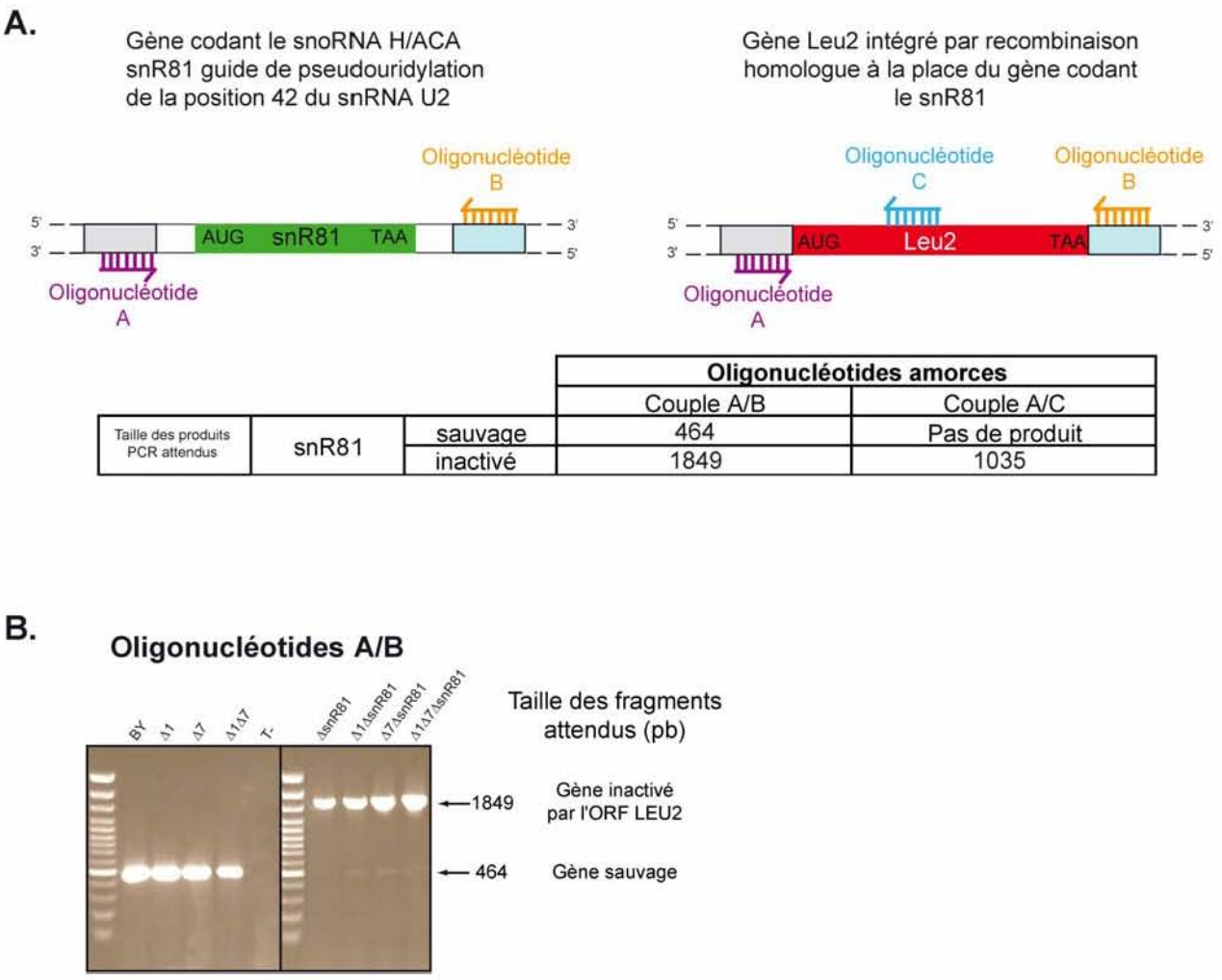


Figure 91 : Vérification par amplification par PCR, de la présence de la délétion du gène codant le snoRNA H/ACA snR81 dans les souches d'intérêt.

A. Stratégie utilisée pour l'analyse de la délétion du gène codant le snoRNA H/ACA guide snR81 dans les différentes souches utilisées. Les oligonucléotides amorces et leur zone d'hybridation sont représentés sur le schéma au dessus du tableau qui indique la taille des fragments attendus.

B. La présence des gène *LEU2* a été testée en utilisant le couple d'oligonucléotides A/B spécifique des souches sauvages BY, ΔPus1, ΔPus7, ΔPus1ΔPus7 (partie droite du gel) et des souches ΔsnR81, Δ1ΔsnR81, Δ7ΔsnR81 et Δ1Δ7ΔsnR81 (partie gauche du gel). Un témoin, correspondant à une réaction en absence d'ADN a été réalisé avec le couple d'oligonucléotides utilisé (T-). Un marqueur de taille a été déposé en parallèle. La taille des produits PCR attendus est indiquée.

Comme on peut l'observer dans la Figure 91B, dans toutes les souches haploïdes, sélectionnées (partie droite du gel), les produits d'amplification attendus pour le couple d'oligonucléotides A/B (1849 pbs) et A/C (résultats non présentés) confirment la présence de la substitution du gène codant le snoRNA H/ACA snR81 par l'ORF LEU2.

A cette étape, nous disposions donc d'un jeu complet de 8 souches (BY, ΔPus1, ΔPus7, ΔPus1ΔPus7, ΔsnR81, Δ1ΔsnR81, Δ7ΔsnR81 et Δ1Δ7ΔsnR81) permettant l'étude de l'importance des résidus Ψ du snRNA U2 de levure dans l'épissage des ARN pré-messagers.

B. Analyse du phénotype de croissance des souches mutées

Chez la levure, la preuve du rôle fonctionnel d'un gène peut être obtenue par l'analyse des changements phénotypiques présentés par une souche mutée pour ce gène dans des conditions de culture données. En particulier, un test informatif et rapide à mettre en œuvre consiste à analyser la capacité de croissance des cellules lors d'une variation de la température de croissance. Ainsi, nous avons testé la thermosensibilité à 37°C ainsi que la cryosensibilité à 18°C et 23°C de l'ensemble des souches.

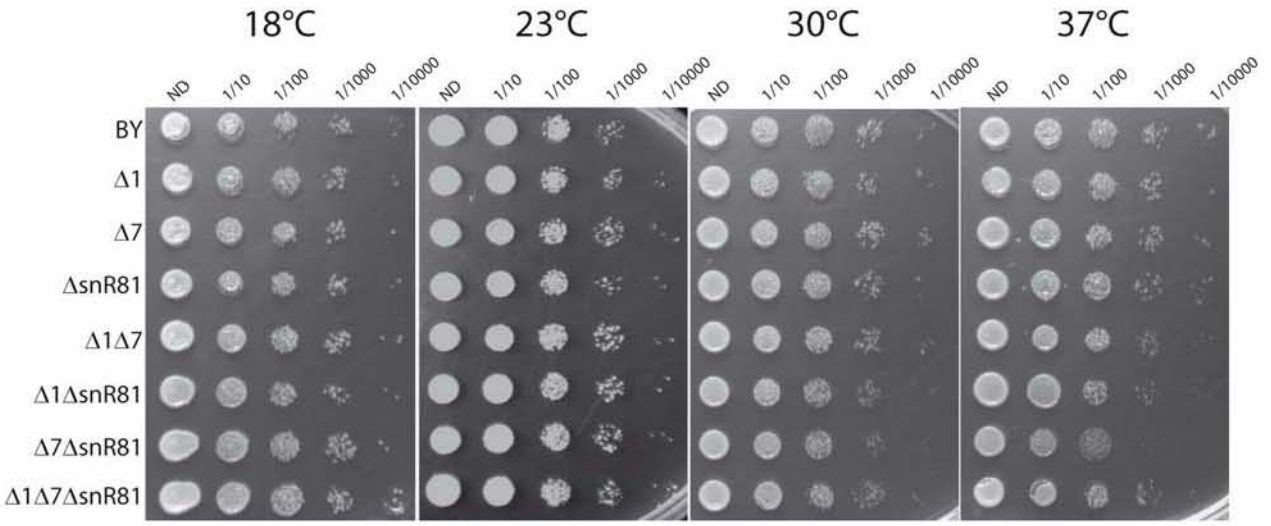


Figure 92 : Analyse du phénotype de croissance des souches utilisées dans notre étude.
Les souches ont été diluées en cascade à partir d'une culture mère, spottées sur boîte de milieu riche YPD puis incubées à différentes températures. Le facteur de dilution est précisé (ND= non dilué, 1/10 = dilution au 1/10ème, etc…) et les températures de croissance sont indiquées.

Comme l'illustre la Figure 92, aucune différence n'est observée dans les conditions normales de croissance à 30°C entre les souches dans lesquelles un ou plusieurs gènes codant une ARN:Ψ-synthase ont été inactivé et la souche sauvage. De façon surprenante, l'accumulation des délétions n'affecte pas non plus la viabilité de *S. cerevisiae* à 18°C, 23°C et 37°C sur milieu riche.
De plus, la comparaison des courbes de croissance réalisées en milieu liquide à 30°C (résultats non présentés) confirmait l'absence de phénotype de croissance. De manière

surprenante, les souches doublement mutées Δ7ΔsnR81 et triplement mutées Δ1Δ7ΔsnRN81 semblaient se développer même un peu plus rapidement que la souche sauvage (résultats non présentés). Ces résultats montraient le caractère non essentiel des ARN:Ψ-synthases Pus1p et Pus7p ainsi que du snoRNA H/ACA snR81 même lorsque les 3 ne sont pas exprimés.

Malgré l'absence de phénotype de croissance, il était néanmoins possible que des altérations de l'efficacité d'épissage de certains pré-ARNm se produisent dans les souches mutées, ce que nous avons testé par une approche globale basée sur l'utilisation de puces à ADN.

C. Analyse globale de l'effet de l'absence des résidus Ψ dans le snRNA U2 sur l'épissage des ARN pré-messagers de *S. cerevisiae* par l'approche puces à ADN

1. Analyse de l'épissage des ARN-prémessagers par Splicing Array

Le génome de la levure boulangère *Saccharomyces cerevisiae* est relativement petit. Il comporte environ 12 mégabases réparties sur 16 chromosomes. Les bases de données de référence (SGD (Cherry et al., 1998), MIPS3 (Mewes et al., 1997), YPD4 (Hodges et al., 1998) recensent aujourd'hui environ 5800 ORFs confirmés et plus de 800 ORFs dits « hypothétiques ». Près de 70% de l'ADN non ribosomal code pour des protéines et seulement 4,9% des gènes contiennent des introns dont 320 avec un intron et 8 avec 2 introns (Tableau 20).

Comme mentionné précédemment, le faible nombre d'introns présents dans les gènes de *S. cerevisiae* permet de faire, avec une seule puce à ADN, une analyse globale de l'effet d'une mutation donnée sur l'épissage de chacun des introns. Une telle puce à ADN avait été développée par le laboratoire de J. Beggs au Wellcome Trust Centre for Biology d'Edimbourg. J'ai donc préparé des ARN totaux à partir de chacune de 8 souches que j'avais produites à Nancy (BY, Δ1, Δ7, ΔsnR81, Δ1Δ7, Δ1ΔsnR81, Δ7ΔsnR81, Δ1Δ7ΔsnR81). Les cellules ont été cultivées à 30°C en milieu de culture riche (YPD) (conditions optimales de croissance) afin de comparer les effets des délétions individuelles ou simultanées d'un ou plusieurs gènes codant les ARN:Ψ-synthases Pus1p, Pus7p et le guide snoRNA H/ACA snR81 sur l'épissage de l'ensemble des ARN pré-messagers de la levure *S. cerevisiae* possèdant un intron selon le protocole décrit dans le chapitre Matériels et Méthodes. Pour cela, nous avons réalisé une transcription inverse des ARN totaux puis un marquage indirect des molécules d'ADNc synthétisées en utilisant 2 nucléotides fluorescents différents : le Cy-3 pour les ADNc issus de la souche sauvage et le Cy-5 pour les ADNc issus des souches mutées (le marquage inverse dit en « Dye Swap » à également été réalisé). Nous pouvions ainsi comparer le profil d'épissage de chacune des souches mutantes que nous avons générées avec le profil d'épissage de la souche sauvage BY4742 (Fig. 93).

N°	ORF	Protéine	N°	ORF	Protéine	N°	ORF	Protéine	N°	ORF	Protéine	N°	ORF	Protéine	N°	ORF	Protéine	N°	ORF	Protéine
1	YFL039C	act1	51	YBL026W	LSM2	101	YNL069C	RPL16B	151	YLR448W	RPL6B	201	YDR129C	sac6	251	YBR219C	Inconnue	301	YML056C	Inconnue
2	YPL129W	ANC1	52	YNL147W	LSM7	102	YKL180W	RPL17A	152	YPR187W	RPO26	202	YHR079C-A	SAE3	252	YBR230C	Inconnue	302	YML067C	Inconnue
3	YKL157W	APE2	53	YDR005C	MAF1	103	YJL177W	RPL17b	153	YDL130W	RPP1B	203	YPL218W	SAR1	253	YBR255C-A	Inconnue	303	YMR147W	Inconnue
4	YJL024C	APS3	54	YDR318W	MCM21	104	YOL120C	RPL18A	154	YDR025W	RPS11A	204	YMR079W	SEC14	254	YCL002C	Inconnue	304	YMR194C-B	Inconnue
5	YDL137W	arf2	55	YKL150W	MCR1	105	YNL301C	RPL18B	155	YBR048W	RPS11B	205	YBL050W	SEC17	255	YCL012C	Inconnue	305	YNL050C	Inconnue
6	YDL029W	ARP2	56	YOR221C	MCT1	106	YBR084C-A	RPL19A	156	YDR064W	RPS13	206	YGL137W	sec27	256	YCR033W	Inconnue	306	YNL065W	Inconnue
7	YMR033W	ARP9	57	YER044C-A	mei4	107	YBL027W	RPL19B	157	YCR031C	RPS14A	207	YKL006C-A	SFT1	257	YDL012C	Inconnue	307	YNL194C	Inconnue
8	YMR116C	ASC1	58	YGL087C	MMS2	108	YBR191W	RPL21A	158	YJL191W	RPS14B	208	YIL123W	SIM1	258	YDL115C	Inconnue	308	YNL246W	Inconnue
9	YIL004C	BET1	59	YGL183C	MND1	109	YPL079W	RPL21B	159	YJL191W	RPS14B	209	YLR275W	SMD2	259	YDL189W	Inconnue	309	YNR053C	Inconnue
10	YJL031C	BET4	60	YIL106W	MOB1	110	YLR061W	RPL22A	160	YMR143W	RPS16A	210	YAL030W	SNC1	260	YDL219W	Inconnue	310	YOL047C	Inconnue
11	YHR101C	BIG1	61	YFL034C-B	MOB2	111	YFL034C-A	RPL22B	161	YMR143W	RPS16A	211	SNR17A	snR17A	261	YDR367W	Inconnue	311	YOL048C	Inconnue
12	YDR099W	bmh2	62	YGL178W	MPT5	112	YBL087C	RPL23A	162	YDL083C	RPS16B	212	SNR17B	snR17B	262	YDR381C-A	Inconnue	312	YOR235W	Inconnue
13	YLR078C	BOS1	63	YDL079C	MRK1	113	YER117W	RPL23B	163	YML024W	RPS17A	213	YNL012W	spo1	263	YDR535C	Inconnue	313	YOR318C	Inconnue
14	YER014C-A	BUD25	64	YLR312W-A	MRPL15	114	YGL031C	RPL24A	164	YDR447C	RPS17B	214	YIL073C	SPO22	264	YEL076C-A	Inconnue	314	YPL109C	Inconnue
15	YOR074C	cdc21	65	YMR225C	MRPL44	115	YGR148C	RPL24B	165	YDR450W	RPS18A	215	YGR225W	SPO7	265	YER007C-A	Inconnue	315	YPR004C	Inconnue
16	YPL241C	cin2	66	MT-1	MT-1	116	YOL127W	RPL25	166	YML026C	rps18B	216	YPL175W	spt14	266	YER093C-A	Inconnue	316	YPR063C	Inconnue
17	YKL190W	CNB1	67	YKL186C	MTR2	117	YLR344W	RPL26A	167	YOL121C	RPS19A	217	YPL175W	spt14	267	YFR024C-A	Inconnue	317	YPR098C	Inconnue
18	YLL050C	COF1	68	YBR119W	MUD1	118	YGR034W	RPL26B	168	YNL302C	RPS19B	218	YHR041C	srb2	268	YFR045W	Inconnue	318	YPR153W	Inconnue
19	YGL187C	COX4	69	YDR397C	NCB2	119	YHR010W	RPL27A	169	YOR293W	RPS1A	219	YML034W	SRC1	269	YGL232W	Inconnue	319	YPR170W-B	Inconnue
20	YIL111W	COX5b	70	YJL205C	Nce11	120	YDR471W	RPL27B	170	YMR230W	RPS1B	220	YMR125W	sto1	270	YGR001C	Inconnue	320	YPR202W	Inconnue
21	YNL130C	CPT1	71	YJL205C-A	NCE11	121	YGL103W	RPL28	171	YKR057W	RPS21A	221	YNL066W	SUN4	271	YHL050C	Inconnue	321	YBR111W-A	SUS1
22	YPR111W	dbf2	72	YHR077C	NMD2	122	YFR032C-A	RPL29	172	YJL136C	RPS21B	222	YKL081W	TEF4	272	YHR076W	Inconnue	322	YCL005W-A	VMA9
23	YNL112W	DBP2	73	YJL041W	NSP1	123	YFR031C-A	RPL2A	173	YLR367W	RPS22B	223	YAL001C	TFC3	273	YHR097C	Inconnue	323	YCR097W	HMRA1
24	YER179W	dmc1	74	YLR093C	NYV1	124	YMR242C	RPL2A	174	YLR367W_2	RPS22B	224	YAL016W	tpd3	274	YHR199C-A	Inconnue	324	YDR424C	DYN2
25	YDR424C	DYN2	75	YGL226C-A	OST5	125	YIL018W	RPL2B	175	YGR118W	RPS23A	225	YML085C	TUB1	275	YIL156W-B	Inconnue	325	YER074W-A	YOS1
26	YBR078W	ECM33	76	YLR054C	OSW2	126	YOR312C	RPL2B	176	YPR132W	RPS23B	226	YML124C	tub3	276	YIL177C	Inconnue	326	YGL076C	RPL7A
27	YKR004C	ECM9	77	YBR186W	PCH2	127	YGL030W	RPL3	177	YER074W	RPS24A	227	YLR306W	UBC12	277	YJL037W	Inconnue	327	YLR316C	TAD3
28	YAL003W	EFB1	78	YOR153W	pdr5	128	YDL075W	RPL31A	178	YIL069C	RPS24B	228	YDR092W	UBC13	278	YJL225C	Inconnue	328	YPL198W	RPL7B
29	YHR123W	EPT1	79	YFR023W	PES4	129	YLR406C	RPL31B	179	YGR027C	RPS25A	229	YBR082C	ubc4	279	YJR005C-A	Inconnue			
30	YBL040C	ERD2	80	YOR122C	pfy1	130	YBL092W	RPL32	180	YLR333C	RPS25B	230	YDR059C	ubc5	280	YJR079W	Inconnue			
31	YGR029W	ERV1	81	YPL031C	pho85	131	YPL143W	RPL33A	181	YGL189C	RPS26A	231	YEL012W	UBC8	281	YJR112W-A	Inconnue			
32	YAL042W	FUN9	82	YER003C	PMI4	132	YOR234C	RPL33B	182	YER131W	RPS26b	232	YDL064W	UBC9	282	YKL002W	Inconnue			
33	YMR307W	GAS1	83	YBL018C	POP8	133	YER056C-A	RPL34A	183	YKL156W	RPS27A	233	YJL130C	ura2	283	YKL023W	Inconnue			
34	YPL075W	GCR1	84	YJL001W	PRE3	134	YIL052C	RPL34B	184	YHR021C	RPS27B	234	YPL230W	USV1	284	YKL133C	Inconnue			
35	YEL003W	GIM4	85	YBR237W	PRP5	135	YDL191W	RPL35A	185	YLR388W	RPS29A	235	YHR039C-A	VMA1	285	YKL158W	Inconnue			
36	YML094W	GIM5	86	YML017W	PSP2	136	YDL136W	RPL35B	186	YLR388W	RPS29A	236	YKL080W	VMA5	286	YKR005C	Inconnue			
37	YER133W	GLC7	87	YHR001W-A	QCR1	137	YMR194W	RPL36A	187	YDL061C	RPS29B	237	YHR012W	VPS29	287	YKR095W-A	Inconnue			
38	YMR292W	GOT1	88	YGR183C	qcr9	138	YPL249C-A	RPL36B	188	YLR287C-A	RPS3A	238	YPR026W	YIP2	288	YLL057C	Inconnue			
39	YNL038W	GPI15	89	YMR201C	RAD14	139	YLR185W	RPL37A	189	YOR182C	RPS3B	239	YNL044W	YIP3	289	YLR049C	Inconnue			
40	YFL031W	HAC1	90	YNL138W-A	RCP1	140	YDR500C	RPL37b	190	YJR145C	RPS4A	240	YDR381W	YRA1	290	YLR128W	Inconnue			
41	YGL251C	HFM1	91	YMR133W	REC114	141	YJL189W	RPL39	191	YHR203C	RPS4B	241	YGR296W	YRF1-3	291	YLR173W	Inconnue			
42	YDL125C	HNT1	92	YLR329W	rec12	142	YNL162W	RPL42A	192	YPL090C	RPS6A	242	YNL339C	YRF1-6	292	YLR199C	Inconnue			
43	YDR305C	HNT2	93	YJR021C	rec17	143	YHR141C	RPL42B	193	YBR181C	RPS6B	243	YPL283C	YRF1-7	293	YLR202C	Inconnue			
44	YGL033W	HOP2	94	YNL312W	RFA2	144	YPR043W	RPL43A	194	YOR096W	RPS7A	244	YHR016C	YSC84	294	YLR211C	Inconnue			
45	YBR215W	HPC2	95	YCR028C-A	rim1	145	YJR094W-A	RPL43B	195	YNL096C	RPS7B	245	YBL059C-A	Inconnu	295	YLR426W	Inconnue			
46	YNL004W	HRB1	96	YDL082W	RPL13A	146	YIL148W	RPL4A	196	YBL072C	RPS8A	246	YBL059W	Inconnu	296	YLR445W	Inconnue			
47	YNL318C	HXT14	97	YMR142C	RPL13B	147	YKR094C	RPL4B	197	YER102W	RPS8b	247	YBL091C-A	Inconnu	297	YLR464W	Inconnue			
48	YNL265C	IST1	98	YKL006W	RPL14A	148	YML073C	RPL6A	198	YPL081W	RPS9A	248	YBL111C	Inconnu	298	YML025C	Inconnue			
49	YLR096W	KIN2	99	YHL001W	RPL14B	149	YLR048W	RPSB	199	YBR189W	RPS9B	249	YBR062C	Inconnu	299	YML036W	Inconnue			
50	YDL106W	kin28	100	YIL133C	rpl16A	150	YDR139C	RUB1	200	YGR214W	RPSA	250	YBR090C	Inconnu	300	YML053C	Inconnue			

Tableau 20 : Les gènes contenant un ou plusieurs introns chez la levure *S. cerevisiae*

Le nom de l'ORF ainsi que celui de la protéine correspondant est indiqué. Les gènes coloriés en jaune contiennent 2 introns, ceux en bleu possèdent un intron dont la boîte de branchement varie du consensus UACUAAC.

Après hybridation des ADNc fluorescents puis lavage, la puce a alors été révélée par lecture sur un scanner permettant de mesurer le signal émis dans chaque spot après excitation des molécules de fluorochrome par un laser (Fig. 93).

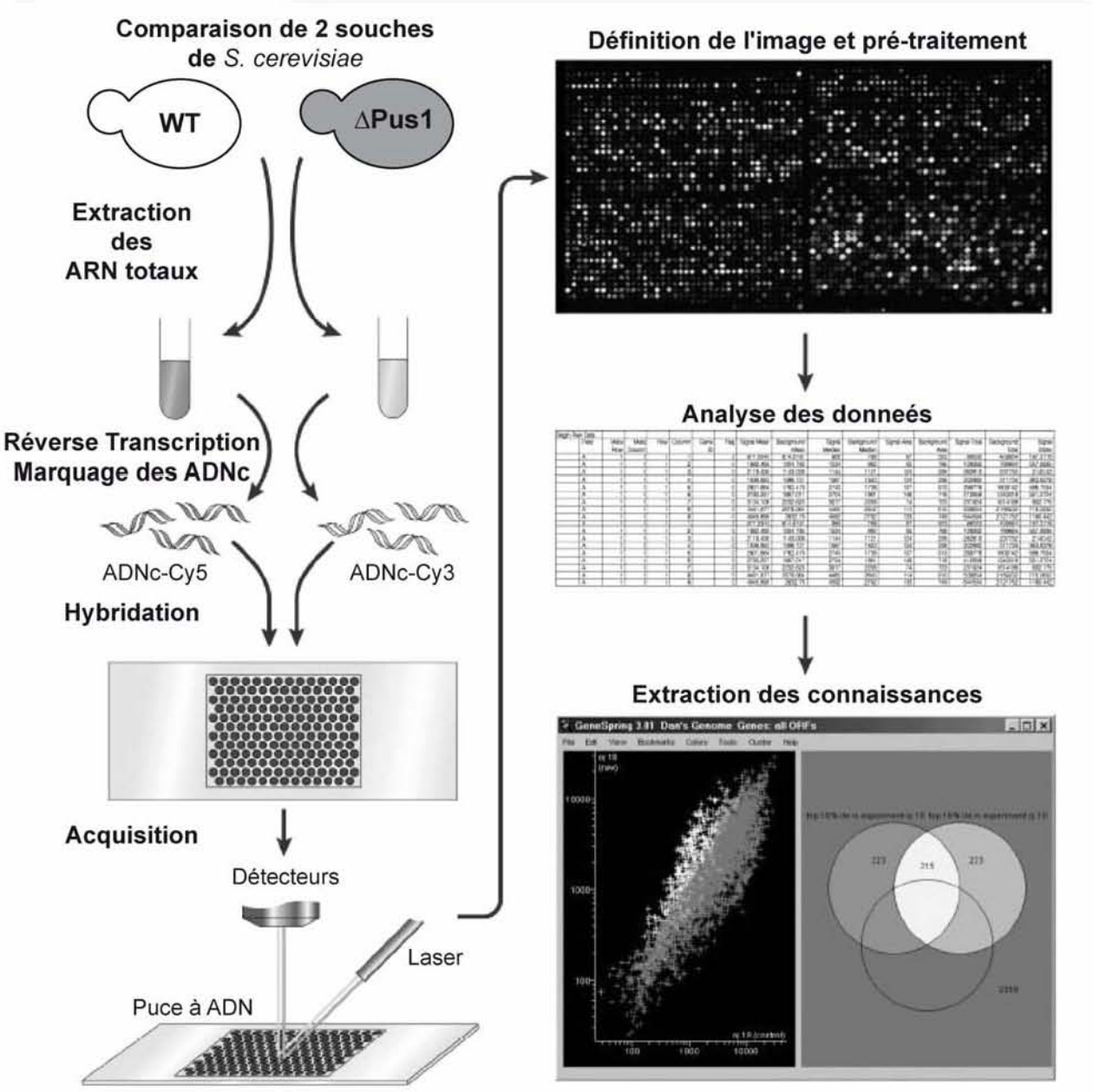


Figure 93 : Schéma de principe des expériences à base de puces à ADN.

L'analyse des taux d'épissage relatifs des différents ARN pré-messagers entre une souche sauvage et une souche dans laquelle le gène a été inactivé a été réalisée en utilisant un type particulier de puce à ADN, appelée Splicing Array (Fig. 94). Ces puces à ADN ont été conçues de façon à mesurer les taux d'ARN pré-messager (sonde V), d'intron (sonde I), et d'ARN mature (sonde M) grâce à l'utilisation de sondes spécifiques présentes en 3 copies

(triplicat expérimental) pour chacune des formes de l'ARN (Fig. 95). Une quatrième sonde ciblant l'ARNm mature (sonde E) est également utilisée pour normaliser les données des expériences. Ainsi, les rapports ARN pré-messager/ARN messager (V/M), ainsi que l'accumulation de l'intron (I/M) peuvent être évalués pour tous les gènes de *S. cerevisiae* contenant un ou plusieurs introns. Pour tenir compte de la variabilité possible du matériel biologique, l'expérience d'hybridation avec les puces à ADN a été répétée trois fois avec des lots d'ARN différents (triplicat biologique).

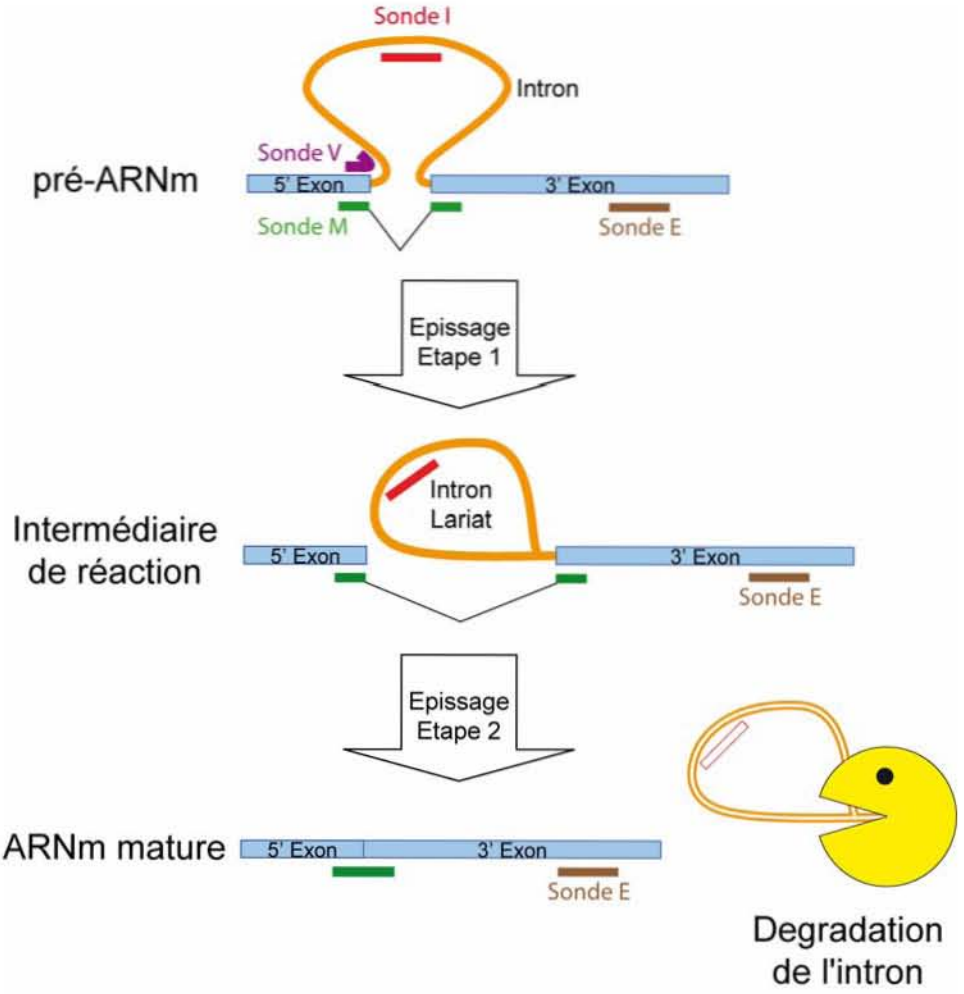


Figure 94 : Région d'hybridation des différentes sondes disposées sur les lames de type « Splicing Array » utilisées pour l'analyse des ARN totaux des souches de levure.
La sonde V permettant de détecter l'ARN-prémessager est schématisée en violet, la sonde M ciblant l'ARN mature est en vert, la sonde I utilisée pour mettre en évidence l'intron est en rouge et la sonde E utilisée pour la normalisation est en marron.

Nous ne détaillerons pas les étapes d'acquisition et de normalisation des données qui ont été en majeure partie réalisées par D. Barass à Edimbourg. En bref, les images TIFF acquises pour chaque lame de l'expérience (Fig. 95) ont été utilisées pour obtenir des tableaux de données après traitement par le logiciel Array-ProAnalyser. Comme le montre l'exemple

de la Figure 96A comparant l'épissage des ARN pré-messagers dans la souche sauvage (WT) et dans la souche dans laquelle le gène *PUS7* a été interrompu (ΔPus7).

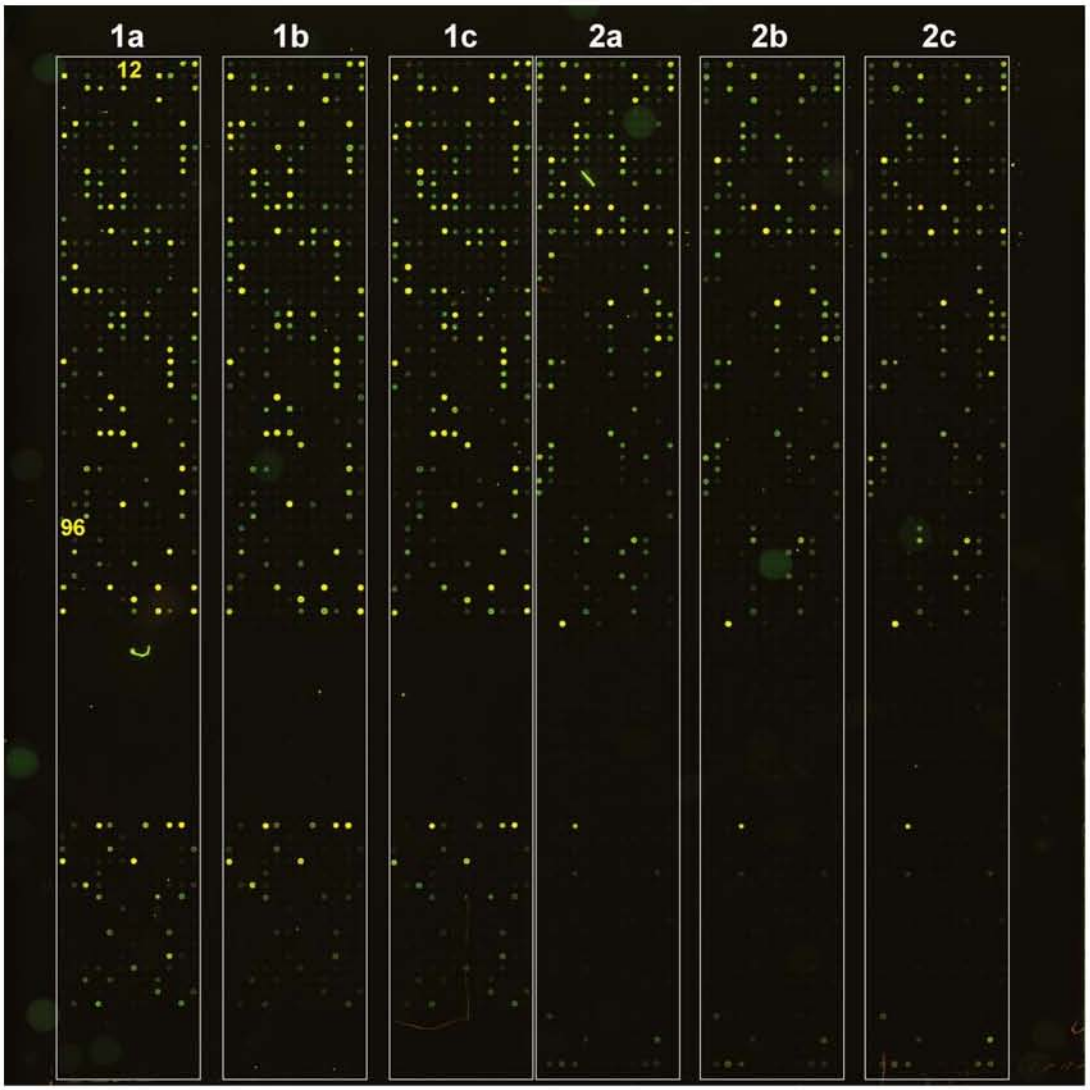


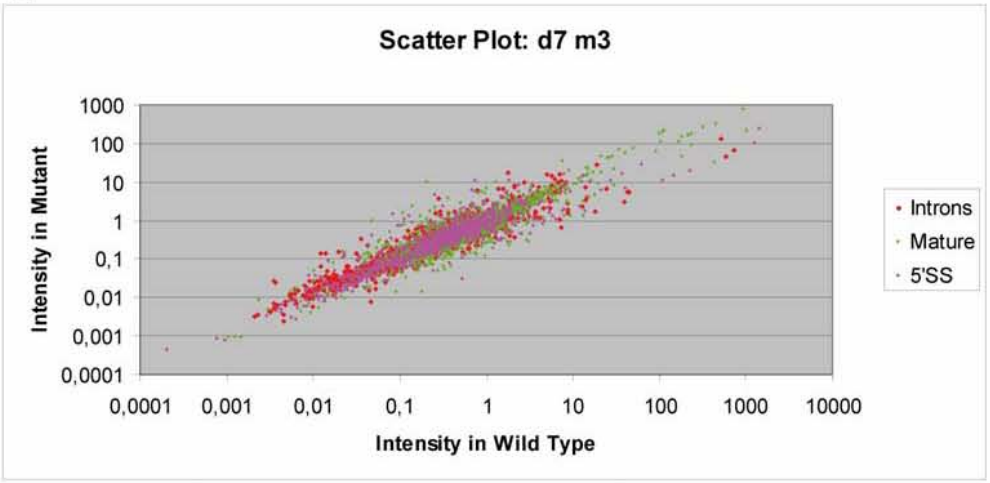
Figure 95 : Exemple de résultats obtenus pour une hybridation effectuée sur la puce et après superposition des images obtenues pour chacun des 2 fluorochromes par le logiciel Array-ProAnalyser.
Les 6 grilles comportant chacune 12x96 spots correspondant aux sondes E, I, M et V de l'ensemble des ARN-prémessagers de la levure *S. cerevisiae* sont représentées sous la forme de rectangle blancs. Les séries 1 a, b, c et 2 a, b et c correspondent à un triplicat expérimental.

La visualisation des intensités obtenues pour les différentes sondes a permis de valider la qualité des données et la méthode de normalisation (Fig. 96B). Dans le cas de l'hybridation de deux cibles marquées en Cy3 et Cy5 synthétisées à partir d'un même échantillon d'ARNm, tous les ratios d'expression devraient être proches de 1 (soit log2(Rr/v)=0).

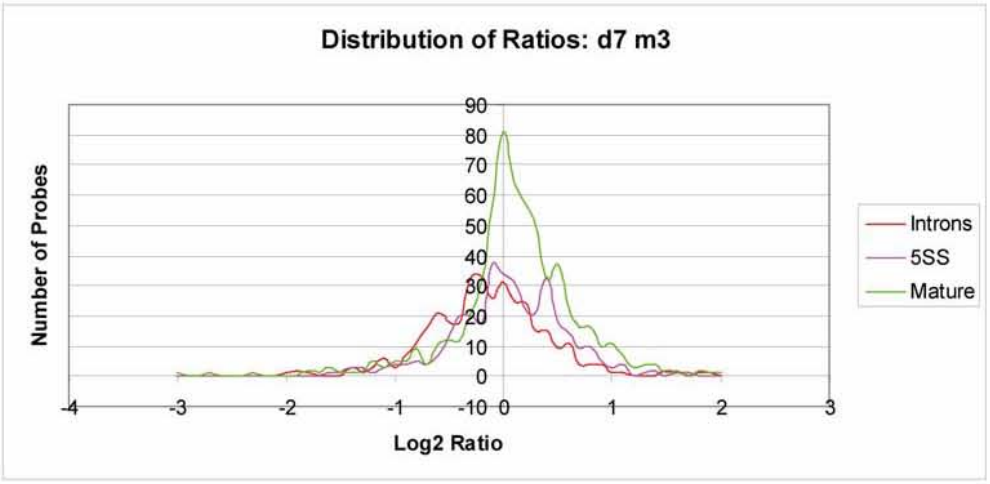
A.

SpotID	Name	Replicate	Gene	Type	BasicType	Intron	DensityMean_WT	DensityMean_mt	DensityStDev_WT	DensityStDev_mt	Background Mean_WT	Background Mean_mt	SignalMean WT	SignalMean mt	ExonSignal Mean_WT	ExonSignal Mean_mt	ExonNorm_ WT	ExonNorm_ mt	Ratio	Log2 Ratio	MedLogRatio	%NearVar
2,2,8,9	YCR097W_I1_PS1.0	1	a1	I	I	YCR097W_1	382,618304	350,945429	17,108009	35,738508	374,457627	330,367232	7,618304	19,946429	38,625	77,819196	0,19723764	0,25631759	0,7695049	-0,37799758	-0,23786239	9,25658792
6-8,9	YCR097W_I1_PS1.0	3	a1	I	I	YCR097W_1	459,341518	362,194196	29,060176	41,129634	447,59322	343,836158	9,341518	22,194196	38,625	77,819196	0,2418516	0,28520207	0,84800085	-0,23786239		
5-8,9	YCR097W_I1_PS1.0	2	a1	I	I	YCR097W_1	387,136161	351,238839	20,794547	44,537026	373,677966	324,485876	13,136161	27,238839	38,625	77,819196	0,34009478	0,35002725	0,97162374	-0,04153036		
2,2,6,10	YCR097W_I2_PS1.0	1	a1	I	I	YCR097W_2	396,299107	384,064732	26,219002	85,083366	381,084746	328,378531	14,299107	56,054732	38,625	77,819196	0,37020342	0,72044862	0,51385124	-0,96057732	-0,64216969	19,8045345
5-6,10	YCR097W_I2_PS1.0	2	a1	I	I	YCR097W_2	397,819196	381,174107	30,301998	90,272251	378,79096	322,305085	18,819196	59,174107	38,625	77,819196	0,48722838	0,76040502	0,6407485	-0,64216969		
6-6,10	YCR097W_I2_PS1.0	3	a1	I	I	YCR097W_2	632,475446	821,171875	58,865443	156,018998	607,40678	759,576271	38,475446	77,171875	38,625	77,819196	0,99612805	0,99168173	1,00448362	0,00645403		
6-6,5	YCR097W_M1_PS1.0	3	a1	M	M	YCR097W_1	384,801339	433,017857	13,562625	16,900674	380,966102	417,468927	1,801339	10,017857	38,625	77,819196	0,04663661	0,12873247	0,36227542	-1,46484116	-0,12466877	60,5026544
5-6,5	YCR097W_M1_PS1.0	2	a1	M	M	YCR097W_1	381,178571	341,178571	13,895366	16,640273	377,146893	332,621469	4,178571	9,178571	38,625	77,819196	0,10818307	0,11794739	0,9172146	-0,12466877		
2,2,6,5	YCR097W_M1_PS1.0	1	a1	M	M	YCR097W_1	396,104911	340,212054	13,792005	15,304724	392,926554	335,19774	4,104911	5,212054	38,625	77,819196	0,10627601	0,06697646	1,58676673	0,66609005		
2,2,8,10	YCR097W_M2_PS1.0	1	a1	M	M	YCR097W_2	392,901786	436,546875	22,165418	153,778139	380,248588	341,180791	10,901786	96,546875	38,625	77,819196	0,28224689	1,24065529	0,22749805	-2,13607392	-1,79463489	21,0746325
5-8,10	YCR097W_M2_PS1.0	2	a1	M	M	YCR097W_2	389,111607	456,573661	29,37067	214,000207	368,932203	316,175141	20,111607	140,573661	38,625	77,819196	0,52068885	1,80641369	0,28824452	-1,79463489		
6-8,10	YCR097W_M2_PS1.0	3	a1	M	M	YCR097W_2	512,725446	510,959821	40,130526	251,695927	483,621469	366,542373	33,725446	162,959821	38,625	77,819196	0,87315071	2,09406256	0,41696096	-1,26201573		
6-8,5	YCR097W_V1_PS1.0	3	a1	V	V	YCR097W_1	452,524554	1173,116071	81,340527	955,292173	436,960452	1065,84181	71,524554	825,116071	38,625	77,819196	1,85176839	10,6029689	0,17464589	-2,51749544	-0,95528994	50,1687017
5-8,5	YCR097W_V1_PS1.0	2	a1	V	V	YCR097W_1	387,049107	365,350446	17,742435	53,299583	377,728814	329,858757	9,049107	35,350446	38,625	77,819196	0,23428109	0,45426383	0,51573793	-0,95528994		
2,2,8,5	YCR097W_V1_PS1.0	1	a1	V	V	YCR097W_1	401,198661	351,328125	15,901073	35,580729	392,988701	330,644068	8,198661	21,328125	38,625	77,819196	0,21226307	0,2740728	0,77447695	-0,36870578		
6-2,6	YCR097W_V2_PS1.0	3	a1	V	V	YCR097W_2	412,395089	429,524554	101,136396	198,171811	371,966102	303,677966	40,395089	125,524554	38,625	77,819196	1,04582755	1,61302815	0,64836286	-0,62512964	-0,61429682	0,74785598
5-2,6	YCR097W_V2_PS1.0	2	a1	V	V	YCR097W_2	422,542411	462,040179	51,329858	164,201159	383,112994	337,932203	40,542411	125,040179	38,625	77,819196	1,04964171	1,60680379	0,65324822	-0,61429682		
2,2,2,6	YCR097W_V2_PS1.0	1	a1	V	V	YCR097W_2	417,310268	411,877232	83,874332	103,802345	392,615819	341,412429	27,310268	70,877232	38,625	77,819196	0,70706195	0,91079368	0,77631408	-0,36528765		

B.



C.



d7 m3			
Value	**Introns**	**5SS**	**Mature**
<-1	20	16	26
>1	7	12	36

Figure 96 : Etapes de traitement et de validation des résultats de Splicing Array.

L'exemple présenté correspond à une expérience comparant le taux relatif d'épissage des ARN pré-messagers d'une souche sauvage par rapport à ceux de la souche ΔPus7.

A. Tableau de sortie de données obtenues après intégration et normalisation des résultats. La puce à ADN permet d'analyser pour chaque gène le taux d'ARN pré-messager (V en violet), d'ARN mature (M en vert) et d'intron (I en rouge), grâce à l'utilisation de sondes spécifiques pour chaque forme. Ainsi, chaque sonde est présente dans 3 régions différentes de la puce, ce qui permet d'augmenter la qualité de l'analyse. La valeur retenue pour chaque sonde correspondant à la médiane de l'ensemble des résultats.
B. Analyse de la qualité de l'hybridation par dispersion du signal observé. Le graphique obtenu représente le signal obtenus pour chaque sonde (I, V, M, E) à partir des lots d'ARN issus de la souche sauvage et en comparaison avec les lots issus de la souche mutée ΔPus7.
C. Représentation graphique de la distribution des résultats obtenus avec les différents types de sondes utilisées pour l'analyse permettant de valider la qualité de l'hybridation.

D'une manière générale, nous constatons que la majorité des points sont effectivement alignés autour de la droite log2(Rr/v)=0 et compris entre les droites log2(Rr/v)=1 et log2(Rr/v)=-1. Nous remarquons également que les points de faible intensité tendent à s'écarter de la droite log2(Rr/v)=0, car ils ont tendance à montrer une plus grande variabilité à cause des fluctuations du bruit de fond.

Nous avons systématiquement éliminé les points correspondants aux spots présentant des défauts d'impression et des intensités très faibles qui étaient susceptibles de présenter une variation de ratio trop importante. Après ce filtrage, très peu de points étaient à l'extérieur des droites log2(Rr/v)=1 et log2(Rr/v)=-1 comme l'illustre le graphique présenté dans la figure 96C représentant la distribution du nombre de points en fonction du ratio. En effet, nous constatons que la distribution est relativement symétrique et centrée autour de 0 dans la majorité des expériences réalisées et ce quelque soit la souche mutante sélectionnée (résultats non présentés).

2. Analyse et interprétation des données

L'analyse des données a été réalisée en 2 étapes successives. Dans un premier temps, nous avons examiné les résultats pris individuellement en recherchant, dans chaque souche mutée, les ARN pré-messagers pour lesquels le taux d'épissage était le plus diminué par rapport à la souche sauvage. Pour cela, nous avons défini l'indice d'inhibition de l'épissage (IE) correspondant à la formule suivante : $IE = (V/M + E/I)_{mutant} / (V/M + E/I)_{sauvage}$ tenant compte à la fois de l'accumulation de l'ARN pré-messager et de de celle de l'intron. Plus la valeur de l'indice IE est élévée et plus l'épissage est diminué entre la souche sauvage et la souche mutée.

Dans ces conditions, il apparaît clairement que l'épissage de la majorité (80%) des ARN pré-messagers n'est ni activé, ni réprimé par l'interruption d'un ou plusieurs gènes

codant les ARN:Ψ-synthases responsables de la formation des résidus Ψ dans le snRNA U2 (données non présentées). En particulier, nous n'observons aucune différence notable de l'épissage des ARN possédant une boîte de branchement non canonique dans les souches mutées et en particulier dans les souches pour lesquelles le gène PUS7 est inactivé seul ou en combinaison avec les gènes PUS1 ou le gène codant le snoRNA H/ACA snR81. Ce résultat invalide donc l'hypothèse d'un lien fonctionnel entre la conservation de certains nucléotides de la boîte de branchement et celle des résidus Ψ présents dans le snRNA U2 et en particulier du résidu Ψ35. En outre, il apparaît que l'absence d'effets sur l'épissage dans les souches mutées peut être corrélée avec l'absence de phénotype de croissance dans ces mêmes souches. Cette observation est également en accord avec le caractère non essentiel des ARN:Ψ-synthases Pus1p, Pus7p et du snoRNA H/ACA snR81 délétées individuellement ou simultanément.

Nous avons cependant sélectionné, pour chacune des souches que nous avions construites, 20 candidats pour lesquels l'indice IE était le plus élevé puis nous avons comparé les résultats obtenus entre les expériences A, B et C. De manière intéressante, nous observons, pour les candidats sélectionnés, une diminution du taux d'ARN pré-messager d'un facteur compris entre 2,5 et 4. Cependant, bien que les résultats pour un lot d'ARN donné dans une souche donnée (i.e. lot A pour la souche Δ1) qui correspondent à un triplicat expérimental (puisque chaque sonde est répliqué 3 fois sur la lame) présentaient le même profil d'expression, les résultats des 3 expériences pour une même souche (i.e. lots A, B et C pour la souche Δ1) n'avaient pas toujours une bonne corrélation (résultats non présentés). Cela suggérait une variabilité de l'échantillon biologique très probablement liée au protocole d'extraction des ARN totaux et/ou de marquage des ADNc synthétisés à partir de ces ARN totaux.

Puisqu'il était difficile d'exploiter les données pour chaque souche prise individuellement, nous avons recherché les gènes dont le profil d'épissage varie de manière identique dans l'ensemble des souches pour lesquelles une ou plusieurs ARN:Ψ-synthases sont inactivées. Ainsi, en sélectionnant non plus les 20, mais les 50 candidats possédant l'indice IE le plus élevé pour chaque souche, nous avons mis en évidence 10 gènes, dont l'épissage est diminué d'un facteur compris entre 1,5 et 2,5 dans toutes les souches dans lesquelles une ou plusieurs ARN:Ψ-synthases sont délétées. Ces candidats sont présentés dans le tableau 21. Il apparaît que la moitié des gènes mis en évidence codent des protéines ribosomales, 3 codent des protéines de fonctions inconnues et les deux derniers gènes codent des protéines jouant un rôle dans l'arrêt du cycle cellulaire et la synthèse de la membrane de la cellule, respectivement. Il est intéressant de noter que la longueur des ARN pré-messagers et la taille des introns n'est pas conservée entre ces 10 gènes (Tableau 21).

Bien que nous avons observé une baisse du taux d'épissage de certains ARN-pré-messagers lorsque une ou plusieurs ARN:Ψ-synthases sont délétées nous ne pouvons pas

déterminer si ces effets sont une conséquence de l'absence de un ou plusieurs résidus Ψ dans le snRNA U2 ou de la délétion des ARN:Ψ-synthases Pus1p, Pus7p et/ou du guide de pseudouridylation snR81. En effet, la diminution de l'efficacité de la réaction d'épissage pouvant être liée à une plus faible stabilité de l'appariement entre le snRNA U2 et les ARN pré-messager que nous avons identifié ou à des effets indirects de l'absence du guide snR81 et des enzymes Pus1p et Pus7p qui pourrait affecter la machinerie d'épissage.

ORF	Protéine	Fonction	ARN pré-messager	intron
YOR096W	RPS7A	Protéine Ribosomale S7A (rp30)	974	400
YNL069C	RPL16B	Protéine Ribosomale L16B (rp23).	1046	448
YDL012C	-	Protéine de fonction inconnue	410	85
YPR043W	RPL43A	Protéine Ribosomale L43A	682	402
YJL136C	RPS21B	Protéine Ribosomale S21B	724	459
YMR194W	RPL36A	Protéine Ribosomale L36A	766	462
YIL156W-B	-	Protéine de fonction inconnue	284	61
YJR112W-A	-	Protéine de fonction inconnue	379	45
YBR186W	PCH2	Protéine requise pour l'arrêt du cycle cellulaire	1808	112
YNL038W	GPI15	Protéine impliquées dans la synthèse du GlcNAc-PI	764	73

Tableau 21 : Candidats retenus à l'issue de l'analyse globale de l'effet de la délétion des ARN:Ψ-synthases responsables de la formation des résidus Ψ dans le snRNA U2 de la levure *S. cerevisiae*.

Cependant les résultats que nous avons obtenus sont préliminaires et nécessitent d'être reproduits et surtout validés par d'autres méthodes d'analyse comme la RT-PCR quantitative ou encore l'utilisation d'un gène rapporteur *in vivo*. Nous allons également retraiter les données en tenant compte de la variabilité que nous avons observée lors de l'analyse des résultats et ceci par approche statistique en analysant l'intensité en fonction des valeur de Log2(ratio) à l'aide du logiciel VARAN (http://www.bionet.espci.fr/varan/) ou en utilisant des tests d'hypothèses.

En plus de la confirmation des résultats obtenus par emploi de techniques alternatives aux puces à ADN, nous rechercherons les facteurs responsables des diminutions de l'efficacité d'épissage que nous avons observées pour certains pré-ARNm. Ces différences peuvent être liées à la séquence du pré-ARNm avoisinant la boîte de branchement ou éventuellement à la présence d'une structuration particulière de l'ARN pré-messager. Pour faciliter l'identification des paramètres impliqués, nous développons actuellement un système basé sur un gène rapporteur. Nous avons choisi le gène *MEL1* codant l'α-galactosidase qui est absente chez *S. cerevisiae*. Les séquences contenant les introns présentant un défaut d'épissage seront fusionnées à l'ORF de l'α-galactosidase, de façon à ce qu'elle ne soit exprimée que si l'intron est éliminé. L'utilisation de cette construction nous permettra de préciser l'importance des différents facteurs, tant de séquence que de structure qui renforcent le défaut que représente l'absence de certains des résidus Ψ du snRNA U2, ceci *in vivo*.

Nous envisageons aussi d'utiliser l'approche synthétiques léthaux, en testant l'effet cumulé des défauts dans la formation des résidus Ψ dans le snRNA U2 et de mutations dans certaines des protéines spliceosomales, en particulier, celles importantes pour l'association du snRNA U2 à l'ARN pré-messager.

PARTIE 5

CARACTERISATION DES HOMOLOGUES DE L'ENZYME PUS7 DE LEVURE CHEZ DES ARCHAEA THERMOPHILES

1. Le domaine Archaea

La découverte, dans les années 1970, de « bactéries » qui semblaient inhabituelles (Kushner et al., 1964 ; Brock et al., 1972; Langworthy et al., 1972) puis, plus tard, des études phylogénétiques réalisées sur les ARN ribosomiques (Woese, 1977) ont conduit à la prise de conscience de l'existence d'un 3ème domaine du monde vivant, les archaea, entre procaryotes et eucaryotes. Les études phylogénétiques de Carl Woese montraient que tous les organismes sans noyau (procaryotes) ne partagent pas un ancêtre commun et qu'ils ne font pas tous partie d'un même groupe monophylétique (ou clade) (Figure 97). La longue liste des caractéristiques non bactériennes mais également non eucaryotes, i.e. propres aux Archaea, apporte aujourd'hui de nombreux arguments convaincants confortant cette division en trois domaines primaires (pour revue, Woese, 1987). Les archaea sont d'autant plus intéressantes à étudier qu'un grand nombre d'espèces sont adaptées à la croissance dans une ou plusieurs conditions extrêmes de température, pH, et/ou salinité.

Du fait de leur découverte récente, les Archaea ont été peu étudiées jusque là. Ce retard est d'autant plus marqué qu'en raison des conditions extrêmes de croissance de beaucoup d'entre eux, la mise au point des conditions de culture en laboratoire a été difficile. Néanmoins, l'étude de ces organismes a largement progressé au cours des dernières années. Etant donné qu'ils possèdent un certain nombre de systèmes cellulaires en commun avec les eucaryotes mais qui sont absent chez les bactéries, ils représentent des modèles plus simples permettant d'accéder à la compréhension de ces systèmes eucaryotes ainsi qu'à une meilleure compréhension de l'évolution. De plus, la stabilité des composants cellulaires des archaea facilite les études structurales de leurs protéines ou complexes ARN-protéines. Une des meilleures illustrations est l'établissement de la première structure 3D à haute résolution de sous-unité ribosomiques à partir des ribosomes d'archea (Ban et al., 2000).

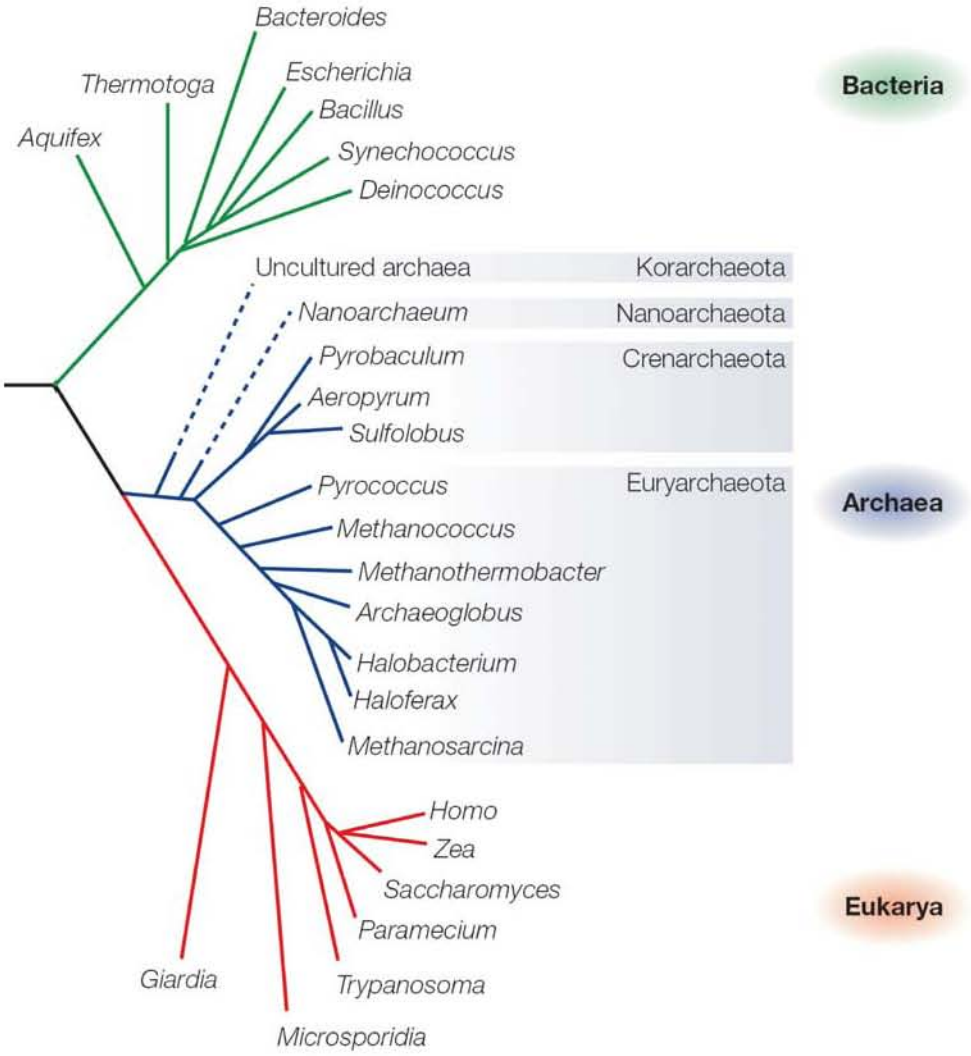


Figure 97: Arbre phylogénétique universel selon Carl Woese et collaborateurs (Allers et Mevarech, 2005).

Cet arbre est basé sur la comparaison des séquences des ARNr de la petite sous-unité ribosomale. Il décrit une division du vivant en trois domaines primaires : Bactéries, Archaea et Eucaryotes. Le domaine Archaea est subdivisé en deux domaines principaux, les euryotes (euryarchaeota) et les crénotes (crenarchaeota). L'existence de phyla supplémentaires (Korarchaeota et Nanoarchaeota) aux deux précédents n'est pas encore clairement établie.

2. Contexte des études réalisées sur les enzymes Pus7p d'archaea

Jusque là, il avait été montré que toutes les conversions de résidus U en Ψ dans les ARNt, aussi bien chez les bactéries que chez les eucaryotes, sont catalysées par des enzymes ne nécessitant pas d'ARN guide. L'existence possible de sRNA H/ACA guidant la pseudouridylation des ARNt d'archaea restait une question ouverte. En effet, peu de gènes susceptibles de coder des ARN:Ψ-synthases avaient été mis en évidence chez les archaea, puisque seuls des gènes codant les protéines aCBF5 et aPus7 (homologue de TruD/Pus7) et aPus10 avait pu être identifiés (Watanabe *et al*., 2000). Chez les euryotes, un homologue de TruA est en plus présent. L'enzyme aPus10 peut modifier les résidus U en positions 54 (Muller et al., non publié) et 55 (Roovers et al., 2006) des ARNt, et aCBF5 les résidus U en position 55 (Roovers et al., 2006). Il ne restait donc que aTruA (chez les euryotes) et aPus7 comme ARN:Ψ-synthases pouvant potentiellement former d'autres positions dans les ARNt. Comme nous l'avons montré dans les chapitres précédents, l'enzyme Pus7 de levure agit en position 13, en position 35 du pré-ARNt^{Tyr} (GUA) (Behm-Ansmant et al., 2003), en position 35 dans le snRNA U2 (Ma et al., 2003) et en position 50 de l'ARNr 5S (Decatur & Schnare, 2008). On savait de part les études réalisées chez *H. volcanii*, *H. morrhua, H. halobium, S. solfataricus, M. thermophilus* et *T. acidocaldarius* (Sprinzl et al., 1998) que de nombreux résidus Ψ sont présents dans les ARNt d'archaea. Le faible nombre d'ARNt:Ψ-synthases pouvant potentiellement modifier les ARNt chez les archaea était donc étonnant.

Cette contradiction apparente pouvait s'expliquer soit par l'existence de sRNA H/ACA guidant la formation de ces résidus ou par une très large spécificité des ARN:Ψ-synthases d'archaea par rapport à celles d'eucaryotes et de bactéries. L'idée de l'existence possible de sRNA H/ACA guidant des modifications dans les ARNt était d'autant plus plausible que des sRNA à boîtes C/D dédiés aux ARNt avaient été trouvés chez ces organismes mais pas chez les eucaryotes (Omer et al., 2003).

3. Démarche et Résultats

Sébastien Muller a recherché par l'approche bioinformatique qu'il avait développée (Muller *et al.*, 2007a) si des sRNA H/ACA capables de guider des pseudouridylations dans les ARNt existaient chez les archaea dont le génome avait été complètement séquencé. Il s'est avéré qu'un nombre limité de sRNA H/ACA présentaient les caractéristiques structurales requises pour être des guides d'ARNt. Ils étaient essentiellement trouvés dans l'espèce *A. pernix* de l'ordre des Desulfurococcales et dans les 4 espèces de l'ordre des Sulfolobales. Les prédictions informatiques montraient qu'ils pouvaient guider la formation du résidu Ψ35 dans le pré-ARNt^{Tyr}(GUA) qui contient un intron dans ces espèces. Par emploi de la méthode de reconstitution de sRNP H/ACA, S. Muller et moi-même avons démontré l'activité de ces particules sur le pré-ARNt^{Tyr}(GΨA) de *S. solfataricus*. Comme l'enzyme Pus7p catalyse la

formation de ce résidu chez les eucaryotes, S. Muller a recherché, toujours par approche bioinformatique, les gènes des enzymes homologues à Pus7p dans les archaea dont le génome était disponible puis il a comparé ces séquences. Il a alors pu mettre en évidence l'existence de mutations ponctuelles à la fois dans les motifs II et IIIa des ARN:Ψ-synthases « Pus7-like » d'archaea de l'ordre des Sulfolobales et d'*A. pernix*, où il avait découvert le sRNA H/ACA guidant la modification en position 35. Les deux motifs II et IIIa sont situés dans le site actif des ARN:Ψ-synthases, ils sont très conservés et plusieurs de leurs acides aminés jouent un rôle essentiel dans l'activité catalytique (Kaya & Ofengand, 2003). Le résidu Asp catalytique était néanmoins conservé dans les enzymes Pus7 d'archaea de l'ordre des Sulfolobales et d'*A. pernix*. S. Muller a alors produit les 2 enzymes Pab Pus7 et Sso Pus7 de *P. abyssi* et *S. solfataricus* sous forme recombinante et nous avons comparé ensemble les propriétés de ces enzymes, ceci en référence à celle de *S. cerevisiae* et d'*E. coli* (TruD).

Nous avons ainsi pu montrer que l'enzyme aPus7 de *P. abyssi* agit comme son homologue eucaryote à la fois en position 13 et 35 des ARNt. A la différence de son homologue eucaryote, elle peut agir en position 35 en absence d'intron. Par ailleurs, elle agit sans doute aussi à d'autres positions des ARNt. La génération des mutations trouvées dans les motifs II et IIIa des enzymes aPus7 de *S. solfataricus*, dans l'enzyme aPus7 de *P. abyssi* abolit son activité à 55°C. Ceci est en accord avec la très faible activité de l'enzyme aPus7 de *S. solfataricus* à cette température. Néanmoins, nous avons constaté qu'à température élevée, l'enzyme aPus7 de *S. solfataricus* agit très faiblement en position 13 mais pas en position 35 des ARNt. Ceci explique sans doute la présence d'un sRNA H/ACA dédié à la modification de la position 35 dans ces espèces, mais l'absence de détection de sRNA H/ACA dédié à la modification de la position 13.

Nous en concluons donc, qu'en cas de défaut d'une enzyme de pseudouridylation des ARNt chez les archaea, le système des sRNP H/ACA peut prendre le relais. Cette observation montre l'importance de la modification en position 35 des ARNt, puisqu'en cas de défaillance d'un système, la cellule en développe un autre pour compenser la déficience.

Ces données ont été rédigées sous la forme d'un article qui a été soumis à NAR et dont je suis co-auteur.

Publication N°4 :

Deficiency of the tRNA^{Tyr}:Ψ35-synthase aPus7 in Archaea of the Sulfolobales order is rescued by the H/ACA sRNA-guided machinery

Sébastien Muller[1], Alan Urban[1], Arnaud Hecker[2], Fabrice Leclerc[1], Christiane Branlant*,[1] and Yuri Motorin[1]

Soumis à Nucleic Acid Research

Publication n°4

Deficiency of the tRNA^{Tyr}:Ψ35-synthase aPus7 in Archaea of the Sulfolobales order is rescued by the H/ACA sRNA-guided machinery

Sébastien Muller[1], Alan Urban[1], Arnaud Hecker[2], Fabrice Leclerc[1], Christiane Branlant*,[1] and Yuri Motorin[1]

[1]Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP Nancy Université, BP 239, 54506 Vandoeuvre-les-Nancy Cedex, FRANCE

[2]Institut de Génétique et Microbiologie, Université Paris-Sud, IFR115, UMR8621-CNRS, 91405 Orsay, FRANCE

Short running title : Pseudouridine Ψ35 formation in archaeal tRNAs

*Corresponding author :

Christiane BRANLANT

Laboratoire Maturation des ARN et Enzymologie Moléculaire

UMR 7567 CNRS-UHP Nancy I

Faculté des Sciences, BP 239

54506 Vandoeuvre-les-Nancy Cedex, FRANCE

Tel. + 33 (0).3.83.68.43.03

Fax + 33 (0).3.83.68.43.07

e-mail christiane.branlant@maem.uhp-nancy.fr

Keywords : Archaea / pseudouridine / RNA:pseudouridine-synthase / tRNA / sRNA-guide

ABSTRACT

Up to now, Ψ formation in tRNAs was found to be catalysed by stand-alone enzymes. By computational analysis of archaeal genomes we detected, in four Sulfolobales species and in *A. pernix*, putative H/ACA sRNAs that might guide Ψ35 formation in pre-tRNA^{Tyr}(GUA). This modification is achieved by Pus7p in eukarya. The computational prediction was verified by *in vitro* reconstitution of H/ACA sRNPs using the identified *S. solfataricus* H/ACA sRNA. Comparison of Pus7-like enzymes encoded by archaeal genomes revealed amino acids substitutions in motifs IIIa and II in Sulfolobales and *A. pernix* Pus7-like enzymes. These conserved RNA:Ψ−synthase motifs are essential for catalysis. Comparison of the activities of recombinant enzymes demonstrated that *P. abyssi* aPus7 acts at positions 35 and 13 and other positions in tRNAs, while *S. solfataricus* aPus7 only has a poor activity at position 13. We took the advantage of the requirement of an UA dinucleotide for modification by *P. abyssi* aPus7, to demonstrate that aPus7 is responsible for Ψ35 formation in tRNAs in *P. abyssi* extracts, whereas, in the absence of this aPus7 activity, an H/ACA sRNP compensates the defect in *S. solfataricus* extracts. This is the first demonstration of an H/ACA sRNA guided Ψ formation in tRNAs.

INTRODUCTION

In all domains of life, pseudouridine residues (Ψ) are the most frequent post-transcriptionally modified residues in RNAs. They are universally found in ribosomal RNA (rRNAs) and in tRNAs (1,2). U to Ψ conversions are catalysed either by stand-alone enzymes (specific RNA:Ψ-synthases) (for review, 3) or by small ribonucleoprotein particles (H/ACA snoRNPs or H/ACA scaRNPs in eukarya, and H/ACA sRNPs in Archaea) (for review, 4). H/ACA RNPs contain a small RNA that defines the targeted U residue by base-pair interaction with the RNA substrate (5-8). Eukaryal and archaeal H/ACA RNPs contain 4 proteins: Nop10, Gar1, L7ae/Nhp2p and Cbf5/Dyskerin (9). CBF5 belongs to the TruB family of RNA:Ψ-synthases. Whereas aCBF5 alone has no activity on rRNAs (5-8) recent data showed an *in vitro* activity of aCBF5 at position 55 in tRNAs, which does not require the presence of a guide RNA (10-12).

The additional free N_1-H of the Ψ nucleobase allows the formation of an additional hydrogen bond, either in *cis*, within the RNA molecule, or in *trans*, with another RNA molecule or a protein. For instance, residue Ψ35 in the eukaryal cytoplasmic $tRNA^{Tyr}$(GUA) allows the formation of an hydrogen bond with the O2' of residue U33, which stabilizes the anticodon loop structure (13,14). Furthermore, substitution of the C-N bond between the ribose and the nucleobase by a C-C bond limits the flexibility of the ribose-phosphate backbone and favours RNA-RNA base-pair interactions (13-15).

In all organisms, pseudouridylations were found to occur at numerous positions in tRNAs (7 in *E. coli*, at least 15 in *S. cerevisiae*, and even more in higher eukaryotes) (2). Formation of residue Ψ55 in the TΨC loop is the most frequent pseudouridylation in tRNAs. It is likely involved in the recognition of elongator tRNA at the A site of ribosomes. Pseudouridylations are also frequently found at tRNA positions 13, 27, 28 and 39 (2). As it was found in all eukaryal $tRNA^{Tyr}$(GUA) and not in bacterial $tRNA^{Tyr}$(GUA), residue Ψ35 in the anticodon loop of

$tRNA^{Tyr}$(GUA) is considered as an eukaryal conserved specific modification. It plays an important role in mRNA decoding in eukarya (16,17). Identification of modified residues in archaeal RNAs, especially tRNAs, is far less advanced than in bacteria and eukarya. Studies mainly focused on archaeal tRNA specific post-transcriptional modifications, such as methylations (m^1I58, Gm18) and archaeosine (18-20). For some hyperthermophilic Archaea, global profiles of post-transcriptional modifications in tRNAs were studied by the HPLC/MS approach (1). However, the locations of the modified residues were not defined. The full set of tRNA molecules was only sequenced for the mesophilic euryarchaeote *Haloferax volcanii* (2,21) and this halophilic organism may not be representative of the whole archaeal domain, since this domain includes organisms growing in a large variety of conditions (22). However, the data obtained for the complete set of *H. volcanii* tRNAs and for a few tRNAs of other archaeal species, show that U to Ψ conversions occur at several positions in archaeal tRNAs.

Up to now, in bacteria and eukarya, all pseudouridylations in tRNAs were shown to be catalysed by stand-alone enzymes and most of the tRNA:Ψ-synthases in *E. coli* and in *S. cerevisiae* were identified (for review, 3,23). Some of them have quite narrow substrate specificities, as is the case for the bacterial TruB, TruD and TruC enzymes (modification at a single position for each of them, 55, 13 and 65, respectively) (24-26). The yeast Pus4p, Pus8p (cytoplasmic) and Pus9p (mitochondrial) enzymes also have narrow specificities (positions 55 for Pus4p and 32 for Pus8p and Pus9p) (27-29). In contrast, the TruA bacterial tRNA:Ψ-synthases acts at several sites in tRNAs (30), and RluA modifies both tRNAs and rRNAs (31). The eukaryal Pus1p and Pus7p enzymes also have very broad specificities. Pus1p acts at 8 distinct positions in tRNAs (32,33) and at position 44 in U2 snRNA (34). Pus7p converts U into Ψ in U2 snRNA (position 35) (35), several tRNAs (position 13), the pre-$tRNA^{Tyr}$(GUA) (position 35) (36) and 5S rRNA (position 50) (37).

The number of rRNA:Ψ–synthases is highly reduced in eukarya compared to bacteria, since several stand-alone specific rRNA:Ψ–synthases are replaced by the H/ACA snoRNP system. Altogether, 11 and 10 distinct RNA:Ψ-synthases were characterized in *E. coli* and in *S. cerevisiae*, respectively. Five of them act on tRNAs in *E. coli* (24,25,26,30,31), while 8 of them act on tRNAs in *S. cerevisiae* (27-29,32,36,38,39).

Studies on RNA:Ψ–synthases in archaea are far less advanced. One puzzling observation is the detection of only a very limited number of putative genes for RNA:Ψ-synthases in archaeal genomes, when taking into account the known signatures of RNA:Ψ-synthases. The only putative genes, which are found in all species, encode a TruD/Pus7 homologue, and aCBF5 (TruB homologue) and PusX/Pus10 which can both act at position 55 in tRNAs (9,10,11,12). In euryarchaeal species an additional gene encoding a member of the TruA family was also detected (9). Up to now, the activities of the TruD/Pus7 and the TruA family members have not been studied.

According to present knowledge on pseudouridylations in archaeal tRNAs, the most frequently modified positions are 13, 39, 54 and 55. *In vitro* tests made with a *Pyroccocus furiosus* cellular extract and *H. volcanii* on *S. cerevisiae* tRNA transcripts revealed the presence of tRNA:Ψ-synthase activities acting at positions 39 and 55 (40). As mentioned above, aCBF5 and/or Pus10 might be the catalyst(s) at position 55. The enzymes acting at the other positions still remain to be identified. Based on the specificities identified for the *E. coli* TruD enzyme and its yeast and plant Pus7 counterparts, archaeal Pus7-like enzymes may catalyze modification at position 13 in some tRNAs and perhaps at position 35 in $tRNA^{Tyr}$(GUA) which are expressed with an intron. Indeed, modification at this position in $tRNA^{Tyr}$(GUA) was found to be intron dependent in eukarya (41,42). Absence of an intron in the *H. volcanii* pre-$tRNA^{Tyr}$(GUA) may explain the absence of Ψ35 in mature $tRNA^{Tyr}$(GUA) (21). Finally, based on the limited number of putative RNA:Ψ–synthases found in archaea and the detection of tRNA specific C/D sRNPs in archaea

(43), one can reasonably ask the question whether there may also be some tRNA specific H/ACA sRNPs in archaea explaining why, in spite of a limited number of RNA:Ψ-synthases, archaeal tRNAs contain several Ψ residues. An alternative explanation can be a very broad specificity of tRNA:Ψ-synthases in these organisms.

Here, by a computational search in archaeal genomes, using a recently developed approach (44), we found H/ACA sRNAs which guide Ψ35 formation in the pre-tRNA^{Tyr}(GUA). However, these H/ACA sRNAs were only found in species of the Sulfolobales and Desulfurococcales orders. We demonstrated their activity by *in vitro* reconstitution of the H/ACA sRNPs and activity tests on various archaeal cellular extracts. Comparison of the amino acid sequences of Sulfolobales and Desulfurococcales aPus7 enzymes with those of other archaeal species revealed amino acid substitutions in the active site. We verified the defect of activity of the enzymes of Sulfolobales by production of recombinant aPus7 enzymes from *P. abyssi* and *S. solfataricus*. Interestingly, we found that the recombinant *P. abyssi* aPus7 enzyme has a broad specificity. It modifies tRNAs at position 13 and at other positions as well as intron-less and intron-containing tRNA^{Tyr}(GUA) at position 35.

MATERIALS AND METHODS

DNA constructs for the production of aPus7 proteins and their variants

The DNA fragments coding the WT aPus7 proteins from *P. abyssi* and *S. solfataricus* were obtained by PCR amplification using genomic DNAs from these two species. The forward and reverse primers used for amplification generated a NheI and BamHI restriction site, respectively. After digestion with NheI and BamHI the amplified DNA was inserted into plasmid pET28b digested by the same nucleases. The variants aPus7 proteins were produced by site-directed mutagenesis (QuickChange kit, Stratagene, USA). The sequences of the

oligonucleotides used for the PCR amplifications and site-directed mutagenesis are available on request.

Purification of recombinant proteins

Wild-type and variant aPus7 proteins were produced in *E. coli* as His_6-tagged protein fusions and purified by affinity chromatography on Ni-NTA agarose (Qiagen, France). Induction of protein production was done by addition of 1 mM IPTG in the LB media supplemented by 50 µg/ml of kanamycin. Cell pellets were suspended in Lysis buffer (20 mM Tris-HCl, pH 7.7; 300 mM NaCl; 1 mM DTT ;10 mM Imidazole) and sonicated. The resulting cell-free extracts were directly loaded onto Ni-NTA agarose. After a washing step with 20 mM imidazole dissolved in Lysis buffer, His_6-tagged proteins were eluted in Lysis buffer containing 250 mM of imidazole. Purified proteins were stored at room temperature in the elution buffer supplemented with 10% glycerol, or at -80°C by adjusting the glycerol concentration to 50%. The *P. abyssi* aCBF5, aNOP10, aGAR1 and L7Ae were produced as GST fusion proteins as previously described (5). They were stored at -80°C in the following buffer, 50 mM Tris-HCl pH 7.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol.

***In vitro* transcription of tRNAs, H/ACA sRNA and the sRNP target RNA**

Wild-type and mutated $tRNA^{Asp}$(GUC) and $tRNA^{Tyr}$(GUA) from *P. abyssi* and WT and mutated $tRNA^{Asp}$(GUC) and pre-$tRNA^{Tyr}$(GUA) from *S. solfataricus* as well as the Sso1 H/ACA sRNA from *S. solfataricus* were produced by *in vitro* transcription of linearized plasmids. The DNA templates coding for those tRNAs and sRNA were amplified by PCR using genomic DNAs of the *P. abyssi* strain GE5 or the *S. solfataricus* strain P2. Forward primers used for amplification bear the sequence for a T7 RNA polymerase promoter. Amplified products were cloned in plasmid pUC18 using its SmaI restriction site. Linearization was performed with the BstNI restriction enzyme. $tRNA^{Asp}$(GUC) and

$tRNA^{Tyr}$(GUA) variants were produced by site-directed mutagenesis (QuickChange kit, Stratagene, USA). All *in vitro* produced RNA transcripts were purified by electrophoresis on an 8M urea 8% polyacrylamide gel. For the pseudouridylation assays, the corresponding RNA transcripts were uniformly labeled with [α–^{32}P]ATP or other [α–^{32}P]NTPs. Conditions for transcription and labeling were described previously (45).

Cellular extracts

About 10^{10} cells from *P. abyssi* or *S. solfataricus* grown to exponential phase and centrifuged, were resuspended in the extraction buffer (25 mM Tris-HCl, pH 7,5 ; 25 mM KCl; 5 mM $MgCl_2$; 10 % Glycérol ; 2 mM DTT). The cells were sonicated for 5 min followed by centrifugation at 10 000 g. The concentration of glycerol was adjusted to 20% for storage at -20°C.

***In vitro* pseudouridylation assay**

The tRNA:Ψ35-synthase activity of purified recombinant aPus7 enzymes or H/ACA sRNP complexes was measured by the nearest-neighbor approach in the conditions described previously (5,45). The sRNA guided activity of reconstituted H/ACA sRNPs was tested by mixing of ~ 4 pmol of Sso1 sRNA with ~150 fmol of the [α-^{32}P]ATP-labeled tRNA in buffer D (20 mM HEPES, pH 7.9, 150 mM KCl; 1.5 mM MgCl2; 0.2 mM EDTA) along with aCBF5, aNOP10, aGAR1 and L7Ae proteins at a 200 mM concentration (5). The non RNA-guided reaction of aPus7 enzymes on tRNA was tested on ~50 fmol of [α-^{32}P]ATP-labeled tRNAs in buffer Psi (100 mM Tris-HCl, pH8.0; 100 mM $AcNH_4$; 5 mM $MgCl_2$; 2 mM DTT; 100 μM EDTA). The activity assays with cell-free extracts were performed with the same amount of labeled tRNA in buffer D in the presence of a similar amount of *P. abyssi* and *S. solfataricus* cellular extracts. For both the sRNA-guided and non RNA-guided reactions, samples were incubated at 55°C, 65°C or 80°C and the reaction was started by addition of the proteins (200 nM each). Activity assays with cellular extracts were carried out at 65°C. Then,

the RNAs were extracted and digested by RNase T2. The resulting 3' phosphate-mononucleotides were fractionated by 2D thin-layer cellulose chromatography (46). 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.

Mapping of pseudouridylation sites in tRNAs

Pseudouridine mapping in RNAs was done using modification by *N*-cyclohexyl-*N'*-(2-morpholinoethyl)-carbodiimide metho-*p*-toluolsulfonate (CMCT) followed by an alkaline treatment for specific modification of Ψ residues (47). Modifications were performed with 0.5 μg of *in vitro* transcribed *S. solfataricus* $tRNA^{Tyr}$. The modified residues were identified by primer extension with the reverse transcriptase using the AMV RT (QBiogene, USA) in the conditions previously described (48). The 5' labeled primer used was complementary to the 14 nt at the 3'-end of the tRNA. RNA sequencing was done with 4 μg of *in vitro* transcribed RNA.

RESULTS

Some archaeal H/ACA sRNAs are predicted to guide Ψ formation in tRNAs

By applying the computational method that we recently developed for the search of H/ACA sRNA genes in archaeal genomes (44), we found five putative H/ACA sRNA motifs, which might direct U to Ψ conversion at position 35 in pre-$tRNAs^{Tyr}$(GUA) that contain an intron (Fig. 1A). One of these unusual H/ACA motifs was detected in a species of the Desulfurococcales order, *Aeropyrum pernix* (sRNA Ape7), the four other ones were found in four species of the Sulfolobales order, namely *Sulfolobus solfataricus* (sRNA Sso1), *Sulfolobus tokodaii* (sRNA

Sto1), *Sulfolobus acidocaldarius* (sRNA Sac1) and *Metallosphaera sedula* (sRNA Mse2). These 5 putative H/ACA motifs fitted to the rules established based on our recent exhaustive analysis of the *P. abyssi* H/ACA sRNAs (44): they all contain a K-turn structure which is linked to the pseudouridylation pocket by a 5- or 6-bp long helix 2. They also contain the conserved ACA trinucleotide at the 3'-end of the stem-loop structure (Fig. 1A). Interestingly, each DNA sequences coding the Sso1, Sto1 and Sac1 H/ACA motifs is directly flanked by a DNA sequence corresponding to a C/D box motif. In *S. solfataricus* this C/D motif corresponds to the already characterized Sso-159 C/D box sRNA (49). The C/D motifs detected in *S. tokodaii* and *S. acidocaldarius* code for yet uncharacterized C/D box sRNAs that we denoted Sto-sR4 and Sac-sR39, respectively. By using a DNA probe complementary to the Sso1 H/ACA motif, we verified the presence of the Sso1 H/ACA sRNA in total RNA extracted from *S. solfataricus* cells. The northern blot analysis (Fig. 1B) revealed the presence of a long RNA that likely contains both the H/ACA and the C/D motifs, and a short RNA corresponding to the H/ACA motif alone. Therefore, the Sso1 H/ACA sRNA is expressed in *S. solfataricus.*

The predicted RNA base-pair interactions between the five putative H/ACA motifs detected and their predicted pre-tRNA^{Tyr}(GUA) targets involved both the anticodon stem-loops and the intron sequences (Fig. 1C). These putative interactions fitted to the criteria that we recently established for efficient H/ACA sRNA-target RNA interaction (44). Indeed, as found in *Pyrococcus* species or in *Thermococcus kodakarensis*, the distance between the targeted U residue and the ACA motif is of 14 nts and a stable base-pair interaction is formed between the 3'-guide sequence and the target RNA (Fig. 1C). Noticeably, the predicted interactions of the pre-tRNAs^{Tyr}(GUA) with the 3' guide sequences of the Sac1 and the Mse2 H/ACA motifs are exceptionally long. as compared to the previously characterized interactions (5,6,11,49,50,51).

Based on computational analysis using the RNAMOT program (44), no other possible target sites in rRNAs or tRNAs were found for the Mse2, Sso1, Sto1 and Sac1 H/ACA motifs.

However, the Ape7 motif might guide pseudourylation at position 2444 in 23S rRNA, as previously found for the *P. abyssi* Pab91 H/ACA sRNA (5). Therefore we concluded that the Mse2, Sso1, Sto1 and Sac1 H/ACA motifs likely guide pseudouridylation in pre-tRNA^{Tyr}(GUA) and this may also be the case for the Ape7 H/ACA motif.

The reconstituted H/ACA Sso1 sRNP catalyses Ψ35 formation in pre-tRNA^{Tyr} transcripts

For an experimental confirmation of the pseudouridylation guiding capability of the identified H/ACA motifs, we used the *in vitro* reconstitution assay established for archaeal H/ACA sRNPs (45). Taking into account the strong sequence conservation in archaea of both the H/ACA sRNP proteins and the H/ACA sRNA characteristic features, we hypothesized that an H/ACA sRNP reconstituted by using the Sso1 H/ACA motif and recombinant H/ACA sRNP proteins from *P. abyssi* produced in *E. coli* may be active. EMSA experiments, indeed confirmed the association of an Sso1 H/ACA RNA transcript with the recombinant *P. abyssi* L7Ae (L), aCBF5 (C), aNOP10 (N) and aGAR1 (G) proteins. The association patterns were similar to those previously obtained for the *P. abyssi* Pab91 sRNA (5) (data not shown). Therefore, as the target U residue in the *S. solfataricus* pre-tRNA^{Tyr}(GUA) was followed by an A residue (Fig. 2A), we tested the activity of the reconstituted particles by the nearest neighbour approach, using an [α-^{32}P]ATP-labeled *S. solfataricus* pre-tRNA^{Tyr}(GUA) transcript (45). Digestion by RNase T2 allowed the detection of Ψp residues released from ΨA dinucleotide sequences. As a control, an assay was also performed in the absence of Sso1 RNA (Fig. 2B). Fractionation by 2D thin layer chromatography of the resulting 3'-phosphate mononucleotides showed that, in spite of the presence of several UA dinucleotides in the Sso pre-tRNA^{Tyr}(GUA), the ratio of U to Ψ conversion in this RNA in the presence of the reconstituted sRNP was close to 1 mole/mole of RNA (Fig. 2B). In addition, no modification was obtained in the absence of Sso1 RNA. These

data were in agreement with the expected U to Ψ conversion at position 35 in the transcript. To complete the demonstration, we repeated the experiment on an Sso pre-tRNA^{Tyr}(GUA) with a U to C mutation at position 35. As expected, no U to Ψ conversion was detected in this RNA (Fig. 2B). We also verified that the U to Ψ conversion in the WT Sso pre-tRNATyr(GUA) occurred at position 35 by CMCT-RT analysis. This is illustrated by detection of an RT-stop at position 36 in the primer extension analysis (Fig. 2C). Therefore, we concluded that in association with *P. abyssi* H/ACA sRNP proteins, the Sso1 H/ACA motif is capable to guide U to Ψ conversion at position 35 in an *in vitro* transcribed pre-tRNA^{Tyr}(GUA).

Archaeal species with a putative tRNA specific guide sRNA show sequence variations in the aPus7-like enzyme

As mentioned in the introduction, U to Ψ conversion at position 35 in pre-tRNATyr (GUA) is catalyzed by the Pus7 stand alone enzyme in yeast (36). By genomic sequence analysis we found putative genes for aPus7-like enzymes in the archaeal species whose genomes were completely sequenced. When the amino acid sequences of the highly conserved motifs I, II and III of RNA:Ψ–synthases were aligned for all the putative aPus7-like proteins (Fig. 3A), we found a strong conservation of these motifs in the archaeal aPus-like enzymes, except for the enzymes of the 4 Sulfolobales and the *A. pernix* species that contain putative tRNA specific H/ACA sRNAs. The sequence divergences concern motifs II and III. Motif II contains the catalytic Asp residue (52) (Fig. 3B). It also contains a conserved pair of Arg-His residues (53). The catalytic Asp residue is found in motifs II of the 4 Sulfolobales species and *A. pernix*, but the Arg-His pair (R78-H79 in *P. abyssi*) is replaced by an Ala(or Thr)-Asp(or Cys) pair (A90/N91 in *S. solfataricus*) (Fig. 3B). Substitution of the Arg-His pair in motif II by other residues is also observed in a few other archaeal species that do not contain a tRNA specific H/ACA sRNA. However, in these species the highly conserved amino acids of motif IIIa are present, especially

the conserved basic residue (*N. maritimus* and *C. symbosium*) which is replaced by an isoleucine in the 4 Sulfolobales species studied (Ile in *S. solfataricus*) and in *A. pernix* (Fig. 3). Furthermore, other amino acid substitutions are also observed in motifs II and IIIa of these 5 species. Altogether, these observations suggested an altered activity or specificity of the aPus7-like enzymes in *S. solfataricus, S. tokodaii, S. acidocaldarius, M. sedula* and *A. pernix* that might be compensated by the presence of a specific H/ACA sRNP.

A pre-tRNA^{Tyr}(GUA):Ψ35–synthase activity is present in *S. solfataricus* cellular extracts

As no Ψ35 had yet been detected in archaeal tRNA^{Tyr}(GUA), the first question was to know whether such modification can occur *in vivo*. We failed to analyze directly the cellular *S. solfataricus* tRNA^{Tyr}(CUA) by CMCT treatment in total RNA, because of the difficulty to find an efficient and specific primer for the extension analysis. Therefore, we tested whether a pre-tRNA^{Tyr}(GUA) Ψ35:-synthase activity is present in *S. solfataricus* extracts. To this end, the *in vitro* transcribed WT Sso pre-tRNA^{Tyr}(GUA) and its U35C variant, labeled by incorporation of [α-^{32}P] ATP, were incubated for 90 min at 65°C in a *S. solfataricus* cellular extract. Then, by using the same approaches as above, namely T2 RNase digestion followed by 2D thin layer chromatography, we detected the formation of 0.60 mole of Ψ residue per mole of RNA in the WT RNA, whereas no modification was detected in the mutated RNA (Fig. 4). The location at position 35 of the Ψ residue formed in the WT pre-tRNA^{Tyr}(GUA) was also verified by CMCT-RT analysis (data not shown). For comparison, the same kind of experiment was performed using a *P. abyssi* cellular extract and the *in vitro* transcribed U35C variant and WT *P. abyssi* (Pab) tRNA^{Tyr}(GUA), which is naturally synthetized in an intron-less form. Interestingly, 0.77 moles of Ψ residues were formed by mole of Pab tRNA^{Asp} and this amount was reduced to 0.42 moles of Ψ per mole of RNA after U35C substitution (Fig. 4). Hence, taken together, the data demonstrated that a specific pre-tRNA^{Tyr}(GUA):Ψ35-synthase activity is present in *S.*

solfataricus, and a tRNA^{Tyr}(GUA):Ψ35-synthase activity capable to modify an intron-less tRNA^{Tyr} is present in *P. abyssi* cells as well as one or more activities able to modify this tRNA at other positions.

In contrast to Pab Pus7-like, the Sso Pus7-like enzyme fails to form Ψ35 in the Sso pre-tRNA^{Tyr} (GUA) *in vitro*

As the Sso Pus7-like enzyme contains the catalytic Asp residue in motif II, despite the amino acid substitutions found in this motif and motif IIIa, it might carry the pre-tRNA^{Tyr}(GUA):Ψ35-synthase activity that we detected in the extract. To test for this possibility, the recombinant *S. solfataricus* (Sso) Pus7-like enzyme and, as a control the *P. abyssi* (Pab) Pus7-like enzyme, were produced in *E. coli* as His$_6$-tagged fusions and purified. The activity of the two recombinant enzymes were tested on *in vitro* transcribed Sso pre-tRNA^{Tyr}(GUA) and Pab tRNA^{Tyr}(GUA) (Fig. 5A) that were labeled by incorporation of [α-^{32}P]ATP. The assays were performed at 55°C, in order to preserve the tRNA 2D and 3D structure. Incubation conditions are given in Materials and Methods. When the Sso Pus7-like enzyme was used, no Ψ formation was detected, neither in the Sso pre-tRNA^{Tyr}(GUA) (Fig. 5C) nor in the Pab tRNA^{Tyr}(GUA) (Fig. 5B). In contrast, an average amount of 1.47 and 1.91 moles of Ψ residue per mole of RNA (mean values of 3 distinct experiments) was formed by the Pab aPus7-like enzyme in the Sso pre-tRNA^{Tyr}(GUA) and Pab tRNA^{Tyr}(GUA), respectively (Fig. 5B&C). When the Sso pre-tRNA^{Tyr}(GUA) or the Pab tRNA^{Tyr}(GUA) with a U to C substitution at position 35 were used, only average values of 0.71 and 0.97 mole of Ψ per mole of RNA were obtained, respectively (Fig. 5B&C). This was a strong indication that the Pab Pus7-like enzyme has an activity at position 35 in these RNA molecules and therefore acts both on intron-containing tRNA and on intron-less tRNAs. The data also showed that at least one other position than U35 was modified by the Pab aPus7-like

enzyme in both the Sso and Pab tRNAs^{Tyr}. In contrast, we concluded that the amino acid substitutions in the Sso Pus7-like enzyme have a strong deleterious effect on its activity at position 35 in tRNAs.

The Sso Pus7-like enzyme also fails to form Ψ13 in Sso tRNA^{Asp} *in vitro*

As yeast Pus7p also catalyzes Ψ13 formation in some cytoplasmic tRNAs, in particular the tRNA^{Asp}(GUC), we tested whether the Sso and Pab Pus7-like enzymes can catalyze Ψ13 formation in the *in vitro* transcribed Sso tRNA^{Asp}(GUC) and Pab tRNA^{Asp}(GUC) that both contain a U residue at position 13 (Fig. 5A). By incubation of transcripts labeled by [α-^{32}P]ATP incorporation with the recombinant Pab Pus7-like enzyme, we found that this enzyme catalyzes the formation of an average amount of 1.15 and 1.08 moles of Ψ in UA dinucleotides per mole of Sso tRNA^{Asp} and Pab tRNA^{Asp}, respectively (Fig. 5). In contrast again, no Ψ formation was detected when using the recombinant Sso Pus7-like enzyme. U to C substitution at position 13 in the Pab tRNA^{Asp}(GUC) completely abolished Ψ formation by the Pab Pus7-like enzyme in this tRNA (Fig. 5B). Taken together, these data confirmed the absence or a very low activity of the Sso Pus7-like enzyme and revealed an activity of the Pab Pus7-like enzyme at position 13 in tRNA^{Asp}(GUC). Therefore, like yeast Pus7p, the Pab Pus7-like enzyme can act at positions 13 and 35 in tRNAs. However, the presence of an intron is not required for modification at position 35.

In Pab tRNA^{Tyr}(GUA), the Pab aPus7-like enzyme only modifies U residues followed by an A residue

According to the above data, the Pab aPus7-like protein might form one Ψ residue in addition to Ψ35 in tRNA^{Tyr}(GUA), we wondered whether this additional modification can occur at position 13 which is also occupied by a U residue in this RNA (Fig. 5A). A U13C substitution indeed

reduced the level of Ψ formation from 1.91 to 1.31 mole of Ψ/mole of RNA and only 0.37 moles of Ψ per mole of RNA were formed when both U35 and U13 were replaced by a C residue (Fig. 6A). By individual substitutions of the U residue in the two remaining UA dinucleotides of this tRNA (U8 and U22), we could show the absence of Ψ formation at position 8 and a low level of Ψ formation (about 0.2 mole Ψ/mole of RNA) at position 22. Then, we wondered whether Pab aPus7 can modify UU, UC or UG dinucleotides in Pab tRNA^{Tyr}(GUA). Utilization of Pab tRNAs^{Tyr}(GUA) labeled by [α-^{32}P]GTP, [α-^{32}P]CTP or [α-^{32}P]UTP incorporation showed that this is not the case (Fig. 6B). Similarly, no Ψ residue were detected in an [α-^{32}P]GTP-labeled Sso tRNA^{Tyr} (not shown). Therefore, we concluded that the Pab aPus7-like enzyme has a marked preference for UA dinucleotides as compared to UU, UC and UG dinucleotides.

Two distinct activities catalyze Ψ35 formation in tRNA^{Tyr}(GUA) in Sso and Pab cellular extracts

For a complete demonstration that *P. abyssi* needs Pab Pus7-like for Ψ35 formation in tRNA, whereas *S. solfataricus* needs an H/ACA sRNA-guided system, we took advantage of the requirement of a 3' A residue for U to Ψ conversion by Pab aPus7 and the absence of this requirement for U to Ψ conversion catalyzed by H/ACA sRNPs. Indeed, based on the above characterized specificity of Pab Pus7-like enzyme, we expected that an A36G substitution in the Pab tRNA^{Tyr}(GUA) transcript would abolish the activity of this enzyme at position 35 in this tRNA. In contrast, the same A36G base substitution was not expected to affect the activity of the Sso1 H/ACA sRNP at position 35 in the Sso pre-tRNA^{Tyr}(GUA) (residue A36 as well as the target U35 residue are not base-paired with the guide RNA, see Fig. 1). In order to release labeled Ψp from position 35 after RNase T2 digestion, the variant A36G Pab tRNA^{Tyr}(GUA) and A36G Sso pre-tRNA^{Tyr}(GUA) transcripts were labeled by [α-^{32}P]GTP incorporation. As expected (Fig. 7A), in spite of the presence of 4 UG dinucleotides in the Pab tRNA^{Tyr} transcript,

no Ψ formation was detected after incubation of the A36G Pab tRNA^{Tyr}(GUA) either with Pab aPus7 or with a Pab cellular extract. In contrast, formation of nearly one mole of Ψ/mole of RNA was observed when using the reconstituted Sso1 H/ACA sRNP in the presence of the Sso A36G pre-tRNA^{Tyr} (GUA) variant. In addition, similar levels of U to Ψ conversion were detected upon incubation of the [α-^{32}P]ATP labeled WT Sso pre-tRNA^{Tyr}(GUA) or the [α-^{32}P]GTP labeled Sso A36G pre-tRNA^{Tyr}(GUA) transcript in an Sso cellular extract (Fig. 7B).

Taken together these data reinforced the demonstration that in a *P. abyssi* cellular extract, Pab Pus7-like is responsible for Ψ35 formation in the Pab tRNA^{Tyr}(GUA) whereas in an *S. sulfolobus* cellular extract, the Sso H/ACA sRNP catalyses Ψ35 formation in the Sso pre-tRNA^{Tyr}(GUA). Therefore, the deficiency of aPus7 activity in *S. solfataricus* is compensated by the presence of a specific H/ACA sRNA guide in this species. This is the first observation of an H/ACA sRNA guide dedicated to tRNA modification. Moreover, we clearly demonstrated an activity of Pus7-like enzyme at positions 35 and 13 in tRNAs, that will now be denoted aPus7.

Important roles of the conserved basic residue in motif IIIa and the KR/RH pair in motif II of archaeal aPus7

We tested which of the detected amino acid substitutions in motifs II and IIIa of Sso aPus7 might explain the absence of *in vitro* activity at 55°C. In order to dissociate the effect of amino acid substitutions in motif IIIa, from amino acid substitutions in motif II, the K27I substitution found in motif IIIa of the Sso aPus7 was transferred into the Pab aPus7 enzyme by site-directed mutagenesis (variant Pab K19I). We also transferred the R90A/H91N sequence found in motif II of Sso aPus7 in Pab aPus7 (variant R78A/H79N). Finally, we produced a Pab aPus7 variant mutated in both motifs (K19I + R78A/H79N). Conversely, we tested whether we could restore some activity in Sso aPus7 by individual I27K substitution in motif IIIa, double A90R and N91H

substitutions in motif II or by combination of these 3 amino acid substitutions. Variant Pab aPus7 and Sso aPus7 were produced as recombinant proteins in *E. coli* and purified using their His_6-tag sequence. As WT Pab aPus7 modifies only position 13 in Pab $tRNA^{Asp}$, this substrate was used to test the activity of the variant proteins. As above, each test of activity was done at 55°C in triplicate and the mean values obtained are shown in Figure 8. Both the K19I and R78A/H79N substitutions in Pab aPus7 decreased the activity at position 13 in Pab $tRNA^{Asp}$ (Fig. 8A), and combination of the 3 amino acid substitutions completely abolished this activity. Restoration of an RH pair in motif II of Sso aPus7 had almost no effect on its activity, whereas restoration of the conserved basic residue in motif IIIa (I27K variant) allowed a significant increase of U to Ψ conversion (Fig. 8A). However, the combination of all 3 substitutions had no positive effect. The overall data showed that both K19 in motif IIIA and the R78/H79 pair in motif II are important for activity of Pab aPus7 at position 13 in tRNAs and that some activity can be restored in the Sso aPus7 by introduction of the basic residue in motif IIIa.

As both *S. solfataricus* and *P. abyssi* grow at elevated temperatures (80°C and 100°C, respectively), we tested the activities of the WT and mutated enzymes at 80°C (Fig. 8B). In these *in vitro* conditions, the tRNA 2D structure was expected to be considerably altered. However, at this elevated temperature, a strong activity of the WT Pab aPus7 was still observed on Pab $tRNA^{Asp}$ (Fig. 8B). The individual K19I or R78A/H79N substitutions had almost no negative effect at this temperature. However, their combination nearly abolished the activity.

Tests performed at 80°C with WT Pab aPus7, using Pab $tRNA^{Tyr}$(GUA) substrates displaying double U13C-U35C mutation or with the U8C mutant, showed that at this high temperature, the Pab aPus7 activity remained stronger at position 13 and 35 compared to position 22 and no activity was detected at position 8 (data not shown). Therefore, the aPus7p specificity seems to be conserved at high temperatures.

Interestingly, the WT Sso aPus7 showed some activity on Pab $tRNA^{Asp}$(GUA) at 80°C and the I27K substitution in this enzyme increased this activity (0.71 mole of Ψ/mole of tRNA, against 0.45 for the WT enzyme). However, here again the A78R and N79H substitutions had no positive effect on the Sso aPus7 activity.

We concluded that in any of the active site sequence context, and at any temperature tested, the presence of a basic residue in motif IIIa is important for aPus7 activity. In addition, based on the data obtained at 80°C, WT Sso aPus7 may have some activity at position 13 in tRNAs *in vivo*.

DISCUSSION

In this work, we described for the first time some H/ACA sRNAs dedicated to tRNA pseudouridylation and, in order to explain their presence, we made the first detailed analysis of the activity of archaeal Pus7-like enzymes on tRNAs.

Like the eukaryal enzymes, archaeal aPus7 enzymes are active at positions 13 and 35 in tRNAs

The archaeal aPus7 enzymes belong to the TruD family of tRNA:Ψ-synthases.Whereas *E. coli* TruD only modifies position 13 in $tRNA^{Glu}$ (24), our data reveal that, like its eukaryal counterpart (35,36), Pab aPus7 can modify tRNAs at positions 13 and 35. Our data also suggest that this enzyme modifies some other positions in archaeal tRNAs.

Consistent with our observations, 13 out of the 42 sequenced *H. volcanii* tRNAs contain a Ψ13 residue (21). Surprisingly, whereas both in plants and in yeast, the U35 to Ψ35 conversion in $tRNA^{Tyr}$(GUA) is strictly intron dependent (41,42,54), here we observed the efficient Ψ35 formation by both Pab aPus7 and a Pab cellular extract in an intron-less Pab $tRNA^{Tyr}$(GUA) transcript. This strongly suggests that this modification can also occur on an intron-less tRNA *in cellulo*. Up to now, Ψ35 formation in intron-less tRNA was only observed in the mitochondrial $tRNA^{Asn}$(GUU) from the starfish *Asteria amurensis* (55). Recent data revealed that only half of the archaeal $tRNA^{Tyr}$(GUA) genes contain an intron (56,57). Based on the present data, this absence of intron may have no marked influence on the presence of a Ψ35 residue in these tRNAs.

Modification by archaeal aPus7 preferentially occurs at UA dinucleotides

Previous comparisons of the *S. cerevisiae* Pus7p RNA substrates revealed that this enzyme modifies U residues that belong to a highly conserved sequence ($Pu_{-4}(G/c)_{-3}U_{-2}N_{-1}\Psi A_{+1}Pu_{+2}$) (36). A similar consensus sequence was also proposed to be required for Ψ35 formation in plant $tRNAs^{Tyr}(GUA)$ (42). In addition, most of the tRNA sequences surrounding Ψ35 residue in bacteria, plant and animal tRNAs fit to the above consensus. Here, in agreement with this observation, our data strongly suggest that the presence of an A residue located 3' to the targeted U residue is required for Pab aPus7 activity. In addition, at the 3 positions where we observed an efficient *in vitro* U to Ψ conversion by Pab aPus7 at 55°C, positions 13 in the Pab $tRNA^{Asp}(GUC)$ and positions 35 in the Pab $tRNA^{Tyr}(GUA)$ and Sso pre-$tRNA^{Tyr}(GUA)$, respectively, the modified sequence fits to the above consensus (Fig. 5). However, modification at position 13 and 22 in the Pab $tRNA^{Tyr}(GUA)$ which occurred at lower yields were formed in sequences which do not fit to the consensus sequence (Fig. 5A). Therefore, the strict requirement of a ($Pu_{-4}(G/c)_{-3}U_{-2}N_{-1}\Psi A_{+1}Pu_{+2}$) found for the eukaryal enzymes is not found for the archaeal enzymes. In addition, when *in vitro* assays were performed at such a high temperature that tRNA structures were melted (80°C), Pab aPus7 modified almost all UA dinucleotides at a high level (Fig. 8B). Recent results obtained for the *P. abyssi* tRNA:m^5C methyltransferase Trm4 demonstrated that its interaction with a protein called Archease modulates both its solubility and its specificity (58). Therefore, we cannot exclude the possibility that in its cellular context the aPus7 specificity is maintained at the very high growth temperature of *P. abyssi* by interaction with cellular proteins. However, the sequence alignment of archaeal tRNAs known to contain a Ψ13 residue shows that, whereas an A residue is always present at position 14, the U residue conserved at position -1 in substrates for eukaryal Pus7 enzymes is only found in about one third of the modified RNAs. Together

these observations suggest lower sequence constraints for aPus7 activity as compared to eukaryal enzymes.

Peculiar aPus7 proteins are found in some crenarchaeote species

By sequence alignment of archaeal Pus7-like proteins with yeast Pus7p and *E. coli* TruD sequences, the Sulfolobales Pus7-like enzymes were all found to display several amino acid substitutions in the highly conserved motifs IIIa and II. According to the 4 available Sulfolobales genomic sequences, this peculiar sequence of aPus7p seems to be a general feature in this archaeal order. In contrast, among the 4 complete genomic sequences available for the Desulfurococcales order, only *A. pernix* was found to carry amino-acid substitutions in the aPus7 motifs IIIa.

As expected by the location of motifs IIIa and II at the active sites of RNA:Ψ-synthases and their strong conservation in these enzymes (59,60) alterations of these motifs in the Sulfolobales aPus7 proteins strongly decrease the activity. Our data show that among the numerous amino acid substitutions found in the Sulfolobales motifs IIIA and II, replacement of the conserved basic residue in motif IIIa by Ile (Ile 27 in Sso aPus7, K19 at the corresponding position in Pab aPus7) has the one having the strongest negative effect on the activity. This strong effect can be explained by the need for a salt bridge formation between this conserved lysine residue in motif IIIa and the catalytic aspartic acid residue (D77 in Pab aPus7) (Fig. 3). This interaction was predicted to facilitate the nucleophilic attack by the COO^- group of the Asp residue at position C6 of the uridine base. The I27 residue in Sso aPus7 cannot form this salt bridge (Fig. 3). This probably explains the almost complete inactivity of Sso aPus7 at 55°C. Our observation that the Sso aPus7 activity is markedly increased when the conserved basic residue in motif IIIa is restored reinforced this hypothesis. However, we observed that Sso aPus7 conserved some capability to form Ψ13 in tRNAs at 80°C. At this high temperature, other amino acids of the catalytic site may contribute to D89

activation due to the higher flexibility of the protein backbone. In contrast, no activity of Sso aPus7 was detected at position 35, even at 80°C. This restricted specificity of Sso aPus7 compared to Pab aPus7 may be explained by a difference in the orientation of the tRNA substrate in the protein RNA binding pocket depending on the nucleotide which is to be modified. Indeed, modification at position 13 implies the interaction of the enzyme with the D-stem loop, and U35 modification implies aPus7 interaction with the anticodon stem loop. We can imagine that, in the case of Ψ13 formation, some stabilizing RNA-protein interactions outside of the catalytic site may compensate for the functional limitation of this site in Sso aPus7.

Even though Sso aPus7 activity at position 13 in tRNAs is very low, it may be sufficient to ensure a minimum level of modification at this position *in vivo*. In contrast, Sso aPus7 is probably unable to ensure sufficient modification at position 35 in $tRNA^{Tyr}$, explaining why a specific H/ACA sRNA dedicated to Ψ35 formation in $tRNA^{Tyr}$ appeared in Sulfolobales and *A. pernix* but not for formation of Ψ13. The lower negative effect of amino acid substitutions in motif II (RH/KR) on aPus7 activity compared to substitution of the basic conserved amino acid in motif IIIa, is consistent with the absence of Ψ35-specific H/ACA sRNA in Desulfurococcales species that display mutations in the aPus7 RH/KR conserved pair in motif II but not of the basic residue in motif IIIa.

The high functional importance of Ψ35 in archaeal $tRNA^{Tyr}$ may explain the requirement for an H/ACA sRNA in Sulfolobales

The need for a guide RNA for Ψ35 formation in Sulfolobales $tRNA^{Tyr}$ may also be related to a high functional importance of Ψ35 in $tRNA^{Tyr}$(GUA). As mentioned in the introduction, residue Ψ35 reinforces the stability of the anticodon stem-loop structure by interaction with

residue U33 (14). This structural stabilization may increase the aminoacylation efficiency of $tRNA^{Tyr}$(GUA) by its cognate aminoacyl-tRNA synthetase (17) and also increases the stability of the codon-anticodon interaction during mRNA translation on the ribosome. This was shown by extensive studies on suppressor $tRNA^{Tyr}$. For instance, the absence of Ψ35 in SUP6 $tRNA^{Tyr}$(UUA) was found to decrease its *in vivo* ability to read-through the ochre stop codon UAA (41). Similarly, the presence of residue Ψ35 increased the $tRNA^{Tyr}$ suppression activity on ochre (UAA) and amber (UAG) stop codons in the Tobacco Mosaic Virus RNA (16). One can imagine that the stabilization of codon-anticodon interactions is particularly important for organisms growing at very high temperatures like the hyperthermophilic Archaea. Therefore, it is reasonable to expect an important role of residue Ψ35 in the codon-anticodon interaction during mRNA decoding in these organisms.

Sulfolobales and *A. pernix* are the first organisms found to contain H/ACA RNA guides for tRNA modification

Up to now, in Bacteria as well as in Eukarya, all tRNA posttranscriptional modifications were shown to be catalyzed by enzymes that do not require a guide RNA. In contrast, an important implication of C/D box sRNAs in tRNA 2'-*O*-methylations was already demonstrated in Archaea (43). The present study shows for the first time the involvement of H/ACA sRNAs in tRNA pseudouridylation. By the computational approach we previously developed (44), we tested whether other H/ACA sRNAs dedicated to tRNA modification may exist in the 48 available complete archaeal genomic sequences. However, no other of the predicted H/ACA sRNAs were found to guide tRNA modification. Therefore, the H/ACA sRNAs present in Sulfolobales and *A. pernix* constitute a unique example of adaptive evolution: an H/ACA sRNA that was probably initially dedicated to rRNA modification was deviated from its

original target to complement the deficiency of a tRNA:Ψ35-synthase. The shift from a non-guided to a guided post-transcriptional modification machinery was already observed for 2'-*O*-methylation at position 56 in archaeal tRNAs (61). Formation of Cm56 in the tRNA TΨ-loop is catalysed by the specific tRNA: 2'-*O*-methyltransferase (aTrm56) in Archaea (61). The aTrm56 protein is missing in the crenarchaeote *Pyrobaculum aerophilum*, and the modification is achieved by an additional C/D sRNP. Together with the present data, these findings demonstrate the extreme plasticity of the RNA modification machinery in Archaea, and more generally, the adaptation potential constituted by the sRNA guides.

Links between tRNA and rRNA modifications in Sulfolobales and *A. pernix*

Interestingly, the Ape7 sRNA, in addition to guide tRNA modification, has also conserved the possibility to guide Ψ formation at position 2444 in 23S rRNA (Fig. 1). The U residue at the homologous position in *P. abyssi* 23S rRNA is modified by the Pab91 sRNA (5). The Ape7 sRNA probably gained an additional activity at position 35 of the pre-tRNA^{Tyr}(GUA) after deleterious mutations in aPus7. In contrast, based on computational analysis, the four Sulfolobales H/ACA sRNAs that guide tRNA modification are not expected to guide any modification in rRNAs. Interestingly, both their 5'- and 3'- guide sequences form unusually long interactions with the pre-tRNA^{Tyr} (GUA) (Fig. 1) and some base compensatory mutations are found between the guide and target RNA when base-pair interactions are compared in the four species. A strong link remained between these H/ACA sRNAs and rRNA maturation, since the genes encoding Sso1, Sto1 and Sac1 H/ACA sRNAs also encode C/D box sRNAs (Fig. S1, supplementary data). Our observation of a common transcript containing the H/ACA and C/D sRNAs is in agreement with the recent finding that the Sso-159 C/D sRNA is expressed as a 120-nt long RNA (49). The Sso-159 sRNA was proposed to guide 2'-O-methylation at position U2621 in 23S rRNA. The C/D motifs of the Sto1 sR4 and Sac1 sR39

C/D box sRNAs can also guide 2'-O-methylation at the same positions in 23S rRNA (data not shown). In addition, by computational approach, we could predict that these 3 C/D sRNAs may also guide 2'-O-methylation of 23S rRNA at positions 2645 and 2648 (*S. solfataricus* numbering, Fig. S1). Altogether this study shows that sRNA sequences which have a capacity of rapid evolution represent a flexible adaptive reservoir for RNA modification in archaea and this is probably also the case for eukaryal snoRNAs.

ACKNOWLEDGEMENTS

S. Muller and A.Urban were pre-doctoral fellows supported by 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. We thank P. Forterre (IGM, Orsay, France) for helpful discussions.

REFERENCES

1. McCloskey, J.A., Graham, D.E., Zhou, S., Crain, P.F., Ibba, M., Konisky, J., Soll, D. and Olsen, G.J. (2001) Post-transcriptional modification in archaeal tRNAs: identities and phylogenetic relations of nucleotides from mesophilic and hyperthermophilic Methanococcales. *Nucleic Acids Res.*, **29**, 4699-4706.
2. Rozenski, J., Crain, P.F. and McCloskey, J.A. (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res.*, **27**, 196-197.
3. Hamma, T. and Ferre-D'Amare, A.R. (2006) Pseudouridine synthases. *Chem. Biol.*, **13**, 1125-1135.
4. Ye, K. (2007) H/ACA guide RNAs, proteins and complexes. *Curr. Opin. Struct. Biol.*, **17**, 287-292.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.*, **35**, 5610-5624.
12. 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.
13. Massenet, S., Mougin, A. and Branlant, C. (1998) In Grosjean, H. and Benne, R. (eds.), *The Modification and Editing of RNA*. ASM Press, New York, pp. 201-227.
14. Auffinger, P. and Westhof, E. (1997) RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *J. Mol. Biol.*, **269**, 326-341.
15. 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.
16. Zerfass, K. and Beier, H. (1992) Pseudouridine in the anticodon G Ψ A of plant cytoplasmic $tRNA^{Tyr}$ is required for UAG and UAA suppression in the TMV-specific context. *Nucleic Acids Res.*, **20**, 5911-5918.
17. Bare, L.A. and Uhlenbeck, O.C. (1986) Specific substitution into the anticodon loop of yeast tyrosine transfer RNA. *Biochemistry*, **25**, 5825-5830.
18. Grosjean, H., Constantinesco, F., Foiret, D. and Benachenhou, N. (1995) A novel enzymatic pathway leading to 1-methylinosine modification in *Haloferax volcanii* tRNA. *Nucleic Acids Res.*, **23**, 4312-4319.
19. 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.

20. Gregson, J.M., Crain, P.F., Edmonds, C.G., Gupta, R., Hashizume, T., Phillipson, D.W. and McCloskey, J.A. (1993) Structure of the archaeal transfer RNA nucleoside G*-15 (2-amino-4,7-dihydro- 4-oxo-7-beta-D-ribofuranosyl-1H-pyrrolo[2,3-d]pyrimidine-5-carboximi dam ide (archaeosine)). *J. Biol. Chem.*, **268**, 10076-10086.
21. Gupta, R. (1984) *Halobacterium volcanii* tRNAs. Identification of 41 tRNAs covering all amino acids, and the sequences of 33 class I tRNAs. *J. Biol. Chem.*, **259**, 9461-9471.
22. Chaban, B., Ng, S.Y. and Jarrell, K.F. (2006) Archaeal habitats--from the extreme to the ordinary. *Can. J. Microbiol.*, **52**, 73-116.
23. Ansmant, I. and Motorin, I. (2001) Identification of RNA modification enzymes using sequence homology. *Mol Biol (Mosk)*, **35**, 248-267.
24. Kaya, Y. and Ofengand, J. (2003) A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA*, **9**, 711-721.
25. Del Campo, M., Kaya, Y. and Ofengand, J. (2001) Identification and site of action of the remaining four putative pseudouridine synthases in *Escherichia coli*. *RNA*, **7**, 1603-1615.
26. Nurse, K., Wrzesinski, J., Bakin, A., Lane, B.G. and Ofengand, J. (1995) Purification, cloning, and properties of the tRNA Ψ55 synthase from *Escherichia coli*. *RNA*, **1**, 102-112.
27. Becker, H.F., Motorin, Y., Planta, R.J. and Grosjean, H. (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**, 4493-4499.
28. Behm-Ansmant, I., Grosjean, H., Massenet, S., Motorin, Y. and Branlant, C. (2004) Pseudouridylation at position 32 of mitochondrial and cytoplasmic tRNAs requires two distinct enzymes in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **279**, 52998-53006.
29. Ansmant, I., Motorin, Y., Massenet, S., Grosjean, H. and Branlant, C. (2001) Identification and characterization of the tRNA:Ψ31-synthase (Pus6p) of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **276**, 34934-34940.
30. 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.
31. 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.
32. 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.
33. Simos, G., Tekotte, H., Grosjean, H., Segref, A., Sharma, K., Tollervey, D. and Hurt, E.C. (1996) Nuclear pore proteins are involved in the biogenesis of functional tRNA. *EMBO J.*, **15**, 2270-2284.
34. Massenet, S., Motorin, Y., Lafontaine, D.L., Hurt, E.C., Grosjean, H. and Branlant, C. (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**, 2142-2154.
35. Ma, X., Zhao, X. and Yu, Y.T. (2003) Pseudouridylation (Ψ) of U2 snRNA in *S.cerevisiae* is catalyzed by an RNA-independent mechanism. *EMBO J.*, **22**, 1889-1897.
36. 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.
37. Decatur, W.A. and Schnare, M.N. (2008) Different mechanisms for pseudouridine formation in yeast 5S and 5.8S rRNAs. *Mol. Cell. Biol.*, **28**, 3089-3100.
38. Behm-Ansmant, I., Branlant, C. and Motorin, Y. (2007) The *Saccharomyces cerevisiae* Pus2 protein encoded by YGL063w ORF is a mitochondrial tRNA:Ψ27/28-synthase. *RNA*, **13**, 1641-1647.
39. Lecointe, F., Simos, G., Sauer, A., Hurt, E.C., Motorin, Y. and Grosjean, H. (1998) Characterization of yeast protein Deg 1 as Pseudouridine synthase (Pus 3) catalysing the formation of Ψ_{38} and Ψ_{39} in tRNA anticodon loop. *J. Biol. Chem.*, **273**, 1316-1323.

40. Constantinesco, F., Motorin, Y. and Grosjean, H. (1999) Transfer RNA modification enzymes from *Pyrococcus furiosus*: detection of the enzymatic activities in vitro. *Nucleic Acids Res.*, **27**, 1308-1315.
41. Johnson, P.F. and Abelson, J. (1983) The yeast $tRNA^{Tyr}$ gene intron is essential for correct modification of its tRNA product. *Nature*, **302**, 681-687.
42. Szweykowska-Kulinska, Z. and Beier, H. (1992) Sequence and structure requirements for the biosynthesis of pseudouridine (Ψ35) in plant pre-$tRNA^{Tyr}$. *EMBO J.*, **11**, 1907-1912.
43. 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.
44. 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.
45. 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.
46. Keith, G. (1995) Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie*, **77**, 142-144.
47. 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.
48. 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.
49. Tang, T.H., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., Bachellerie, J.P. and Huttenhofer, A. (2005) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.*, **55**, 469-481.
50. 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.
51. Muller, S., Leclerc, F., Behm-Ansmant, I., Fourmann, J.B., Charpentier, B. and Branlant, C. (2008) Combined in silico and experimental identification of the Pyrococcus abyssi H/ACA sRNAs and their target sites in ribosomal RNAs. *Nucleic Acids Res.*
52. 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.
53. 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.
54. Nishikura, K. and De Robertis, E.M. (1981) RNA processing in microinjected Xenopus oocytes. Sequential addition of base modifications in the spliced transfer RNA. *J. Mol. Biol.*, **145**, 405-420.
55. Tomita, K., Ueda, T. and Watanabe, K. (1999) The presence of pseudouridine in the anticodon alters the genetic code: a possible mechanism for assignment of the AAA lysine codon as asparagine in echinoderm mitochondria. *Nucleic Acids Res.*, **27**, 1683-1689.
56. Sprinzl, M. and Vassilenko, K.S. (2005) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **33**, D139-140.
57. Mallick, B., Chakrabarti, J., Sahoo, S., Ghosh, Z. and Das, S. (2005) Identity elements of archaeal tRNA. *DNA Res.*, **12**, 235-246.
58. Auxilien, S., El Khadali, F., Rasmussen, A., Douthwaite, S. and Grosjean, H. (2007) Archease from *Pyrococcus abyssi* improves substrate specificity and solubility of a tRNA m^5C methyltransferase. *J. Biol. Chem.*, **282**, 18711-18721.
59. Hoang, C. and Ferre-D'Amare, A.R. (2004) Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. *RNA*, **10**, 1026-1033.
60. 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.
61. Renalier, M.H., Joseph, N., Gaspin, C., Thebault, P. and Mougin, A. (2005) The Cm56 tRNA modification in archaea is catalyzed either by a specific 2'-O-methylase, or a C/D sRNP. *RNA*, **11**, 1051-1063.

FIGURE LEGENDS

Figure 1: Putative H/ACA sRNA guides for Ψ35 modification in pre-tRNATyr(GUA) are found in Sulfolobales species and in *A. pernix.*

Panel A: Structures of the putative H/ACA sRNAs involved in Ψ35 formation in the pre-tRNA^{Tyr}(GUA) of *A. pernix, M. sedula, S. solfataricus, S .tokodaii* and *S. acidocaldarius*. Characteristic K-turns and H/ACA sequences are boxed. The identity of the C/D box sRNAs present at the 3'extremities of the Sso1 (Sso-159), Sac1 (sR4) and Sto1 (sR39) H/ACA sRNAs are indicated. Panel B: Northern blot analysis of Sso1 expression in *S. solfataricus*. Total RNA was fractionated by electrophoresis in 8M urea 8% polyacrylamide gel, then, transferred to Nylon membrane (Hybond N+, GE Healthcare, USA) and hybridized with an Sso1-specific [γ-^{32}P]-labeled oligonucleotide. Sizes of molecular weight markers are given on the left of the autoradiogram. Arrows indicate the positions of the short form RNA corresponding to the free Sso1 H/ACA sRNA (H/ACA) and of the full length transcript containing the 3' terminal C/D box sRNA (H/ACA+C/D). Panel C: Proposed base-pair interactions of the Ape7, Mse2, Sso1, Sto1 and Sac1 H/ACA sRNAs with their proposed target pre-tRNA^{Tyr}(GUA), and potential interaction between LSU RNA target and the Ape7 sRNA. An arrow indicates the site of U to Ψ conversion. The distance between the pseudouridylation pocket and the K-turn and that between the pseudouridylation pocket and the ACA box are shown.

Figure 2: Sso1 sRNA-guided Ψ35 formation in the Sso pre-tRNA^{Tyr}(GUA).

Panel A: 2D structure of the *S. solfataricus* pre-tRNA^{Tyr}(GUA). The targeted U35 residue is circled. The sequence of the intron present in the anticodon loop is represented by small letters. Panel B: *In vitro* reconstitution of the Sso1 H/ACA sRNP

and test of its activity at position Ψ35 in pre-tRNA^{Tyr}(GUA). [α-^{32}P]ATP-labeled WT pre-tRNA^{Tyr} (GUA) or its U35C variant were incubated for 90 min at 65°C in the presence of the L7Ae, CBF5, NOP10 and GAR1 recombinant proteins at a 200 nM concentration (LCNG), in the presence (Sso/LCNG) or the absence (LCNG) of the Sso1 H/ACA sRNA. Modified transcripts were then digested by RNAse T2 and the resulting 3'-NMPs were analysed by 2D TLC (46). Positions corresponding to the pseudouridine spot are circled on the autoradiograms. The 3'-NMP composition of the RNAse T2 digest, as well as the molar amount of Ψ residue formed by mole of RNA are also indicated on the 2D TLC autoradiograms.

Panel C: CMCT-RT mapping of the pseudouridine residue formed upon treatment of the *S. solfataricus* pre-tRNA^{Tyr}(GUA), in conditions allowing the reconstitution of an active Sso1 H/ACA sRNPs (Sso1/LCNG). The modified RNA was treated by CMCT, for 2, 10 and 20 min with (+) or without (-) subsequent alkaline treatment (OH-) and analyzed by primer extension. A control experiment was performed in the absence of CMCT treatment. Lanes U, G, C and A correspond to the sequencing ladders obtained with the same oligonucleotide. An arrow shows the position of the RT stop corresponding to Ψ35.

Figure 3: Presence of unusual Pus7-like enzymes in Sulfolobales species and *A. pernix*.

Panel A: Multiple sequence alignment of Pus7 proteins from various archaeal species. Only the regions of motif IIIa and of the motif II which contains the catalytic D residue are shown. Different colours are used to highlight each of the conserved residues in motifs IIIa and II. The names of the organisms are indicated on the left, red arrows show species that contain a pre-tRNA^{Tyr}-specific H/ACA sRNA.

Panel B: 3D structure modelization of the Pab aPus7 (left) and Sso aPus7 (right) active sites. Modelization was done by using the crystal structure of *M. mazei* TruD (PDB 1Z2Z). Only highly conserved amino acids in the catalytic sites of RNA:Ψ-synthases are represented in the Pab and Sso aPus7 models, the K27I, R90A and H91N substitutions in Sso aPus7 can be seen.

Figure 4: Test of the activity of Sso and Pab cellular extracts on the pre-tRNA^{Tyr}(GUA).

The [α-^{32}P]ATP-labeled WT pre-tRNA^{Tyr} (GUA) or its U35C variant were incubated for 90 min at 65°C in the presence of a Sso or Pab cellular extract. Formation of the residue Ψ35 was tested as described in figure 2B, by 2D-TLC analysis after digestion of the RNA by the T2 RNase. The molar amounts of Ψ residue formed per mole of RNA are indicated on the autoradiograms of the four 2D-TLCs. Reproducible results were obtained in two other independent experiments.

Figure 5: Test of the activity of recombinant Pab and Sso aPus7 on different tRNA and pre-tRNA substrates.

Panel A: 2D structures of the *P. abyssi* tRNA^{Tyr}(GUA) and tRNA^{Asp}(GUC) and *S. solfataricus* pre-tRNA^{Tyr}(GUA) and tRNA^{Asp}(GUC). The UA dinucleotides present in these RNAs are indicated by arrows. The potential U13 target sites in tRNA^{Asp} and U35 target sites in mature and pre-tRNA^{Tyr} are circled.

Panels B and C: Tests of the *in vitro* activity of the two recombinant Pab aPus7 and Sso aPus7 on the *P. abyssi* tRNA^{Tyr}(GUA) and tRNA^{Asp}(GUC) and their respective U35C and U13C variants (Panel B) and on the *S. solfataricus* pre-tRNA^{Tyr}(GUA) and tRNA^{Asp}(GUC) and its U35C variant (Panel C). All transcripts were [α-^{32}P]ATP-labeled and incubated for 90 min at 55°C with the recombinant proteins (Pab aPus7 or

Sso aPus7). Modified RNAs were analysed as described in Figure 2. The molar amounts of Ψ residue formed per mole of RNA in the experiments illustrated in this figure are given for each 2D-TLC autoradiogram. Reproducible results were obtained in two other independent experiments.

Figure 6: Recombinant Pab aPus7 modifies both positions 13 and 35 at a high yields in the *P. abyssi* $tRNA^{Tyr}$(GUA) and a UA dinucleotide is required for its activity.

Panel A: The activity of recombinant Pab aPus7 was tested on variant *P. abyssi* $tRNA^{Tyr}$(GUA) in which each of the UA dinucleotides were converted into a UC dinucleotides. To this end, RNA transcripts were labeled by [α-^{32}P] CTP incorporation in the course of transcription. Incubation with the enzyme was performed for 90 min at 55°C. The molar amount of Ψ residue formed in each assay was determined as described in figure 2. They are given on each of the 2D-TLC autoradiograms.

Panel B: Pab aPus7 needs a UA dinucleotide for activity. The *P. abyssi* $tRNA^{Tyr}$(GUA) was labeled by incorporation of [α-^{32}P] ATP, CTP, UTP or GTP. Formation of Ψ residue by Pab aPus7 was tested after 90 min of incubation at 55°C. No Ψ formation was detected by 2D TLC analysis.

Figure 7: Formation of Ψ35 in $tRNA^{Tyr}$(GUA) is catalyzed by Pab aPus7 in *P. abyssi* and by the Sso1 H/ACA sRNP in *S. solfataricus.*

Panel A: The activities of recombinant Pab aPus7 and a Pab cellular extract on the WT $tRNA^{Tyr}$(GUA) and its A36G variant were tested by the nearest neighbour approach after [α-^{32}P]ATP and [α-^{32}P] GTP labeling of the WT and variant RNAs, respectively. Incubations with the recombinant proteins were performed for 90 min at 55°C and with the extract for 90 min at 65°C. Additional spots visible on the 2D-TLC

autoradiograms correspond to formation of methylated nucleotides naturally present in many archaeal tRNAs (m^1G37, Cm56, m^1I57 and m^1A58). The molar amounts of Ψ residue formed per mole of RNA are given.

Panel B: The activity of the reconstituted Sso1 sRNP (Sso1/LCNG) and an Sso cellular extract are tested on the WT Sso pre-$tRNA^{Tyr}$(GUA) and its U36G variant. The Sso1 H/ACA sRNP was reconstituted and its activity was tested using the protocol described in Materials and Methods.

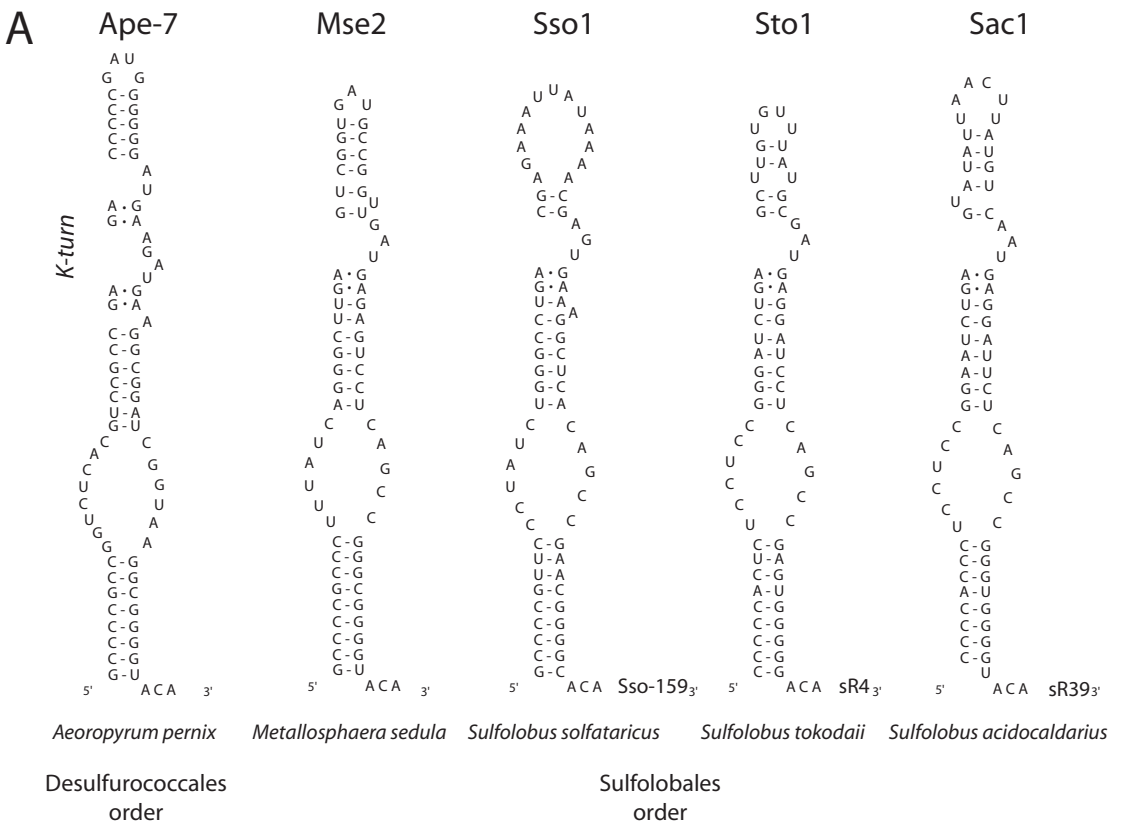
Figure 8: Test of the activity of Pab aPus7 and Sso aPus7 recombinant WT or variant proteins on the [α-^{32}P]ATP-labeled *P. abyssi* $tRNA^{Asp}$(GUC). Enzymatic tests were performed as described in the legend of Figure 5, except that we used point-mutated variants of the enzymes, Pab aPus7 K19I, R78A/H79N, and with the three R78A, H79N and K19I mutations and Sso aPus7 I127K, A90R/N91H and with the three I127K, A90R and N91H mutations. Incubations with the enzymes were performed both at 55°C (Panel A) and at 80°C (Panel B) and the molar amounts of Ψ residue formed per mole of $tRNA^{Asp}$ are given both on the 2D-TLC autoradiograms and in a histogram. Error bars correspond to the standard deviations in three independent experiments.

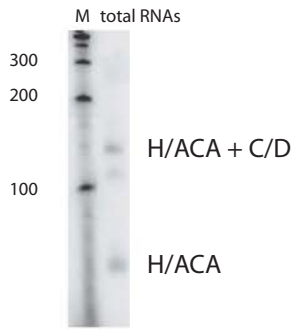
Figure S1: The Sso sRNA contains both an H/ACA and a C/D motif.

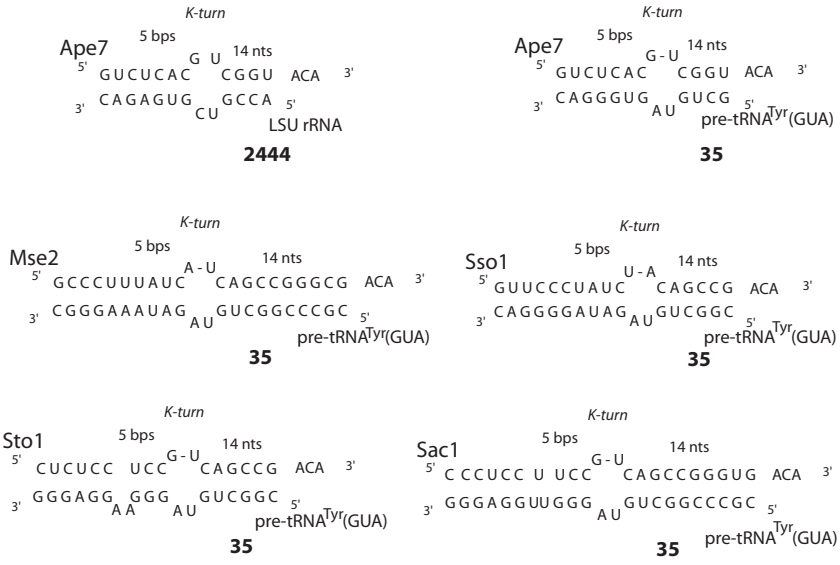
Panel A: Predicted secondary structure of the extended form of the Sso1 sRNA containing both the H/ACA and C/D motifs. The K-turn, ACA–box, as well as C/D and C'/D' boxes are indicated.

Panel B: Predicted base-pair interactions between the Sso-159 C/D box sRNA and its three possible 23S rRNA target sequences.

Figure 1


A
Ape-7
Mse2
Sso1
Sto1
Sac1
K-turn
Sso-159
sR4
sR39
Aeoropyrum pernix
Metallosphaera sedula
Sulfolobus solfataricus
Sulfolobus tokodaii
Sulfolobus acidocaldarius
Desulfurococcales order
Sulfolobales order


B
M total RNAs
300
200
100
H/ACA + C/D
H/ACA


C
K-turn
Ape7
5 bps
14 nts
LSU rRNA
2444
pre-tRNATyr(GUA)
35
Mse2
Sso1
Sto1
Sac1

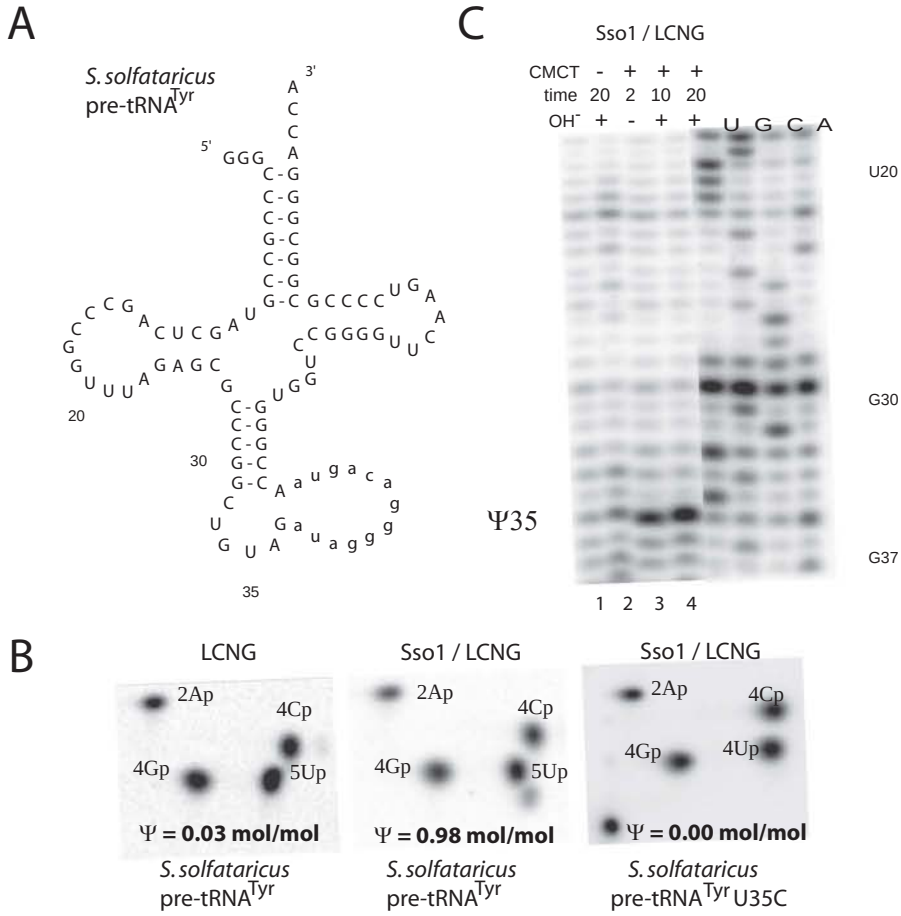

A
S. solfataricus
pre-tRNA Tyr
5'
3'
20
30
35
C
Sso1 / LCNG
CMCT - + + +
time 20 2 10 20
OH- + - + +
U G C A
U20
G30
G37
Ψ35
1 2 3 4
B
LCNG
Sso1 / LCNG
Sso1 / LCNG
2Ap
4Cp
4Gp
5Up
4Up
Ψ = 0.03 mol/mol
Ψ = 0.98 mol/mol
Ψ = 0.00 mol/mol
S. solfataricus
pre-tRNA Tyr
S. solfataricus
pre-tRNA Tyr
S. solfataricus
pre-tRNA Tyr U35C

Figure 2

Figure 3

A

		K19 F25		Pab aPus7	D77 R78/H79
		I27 F33		Sso aPus7	D89 A90/N91
M. acetivorans	24	GLGGRLRQEVEDFIVKEITN	78	SQKRISVAGTKDKRALTTQKI	
M. mazei	21	GLGGQLRQEIEDFIVKEITN	75	SQKRISVAGTKDKRALTTQKI	
H. salinarium	22	GTGGRLRDSPADFRVRELEA	82	SRERVRWAGTKDRHAVTTQLF	
A. fulgidus	16	GMGGEIKAEPEDFYVEEIAE	70	SQKRISFAGTKDKRALTVQYF	
T. acidophilum	14	RPVIRIKENPEDFTVEEVAD	68	SRKRISFAGTKDKRARKIQYF	
T. volcanium	8	EEAIEIKRDPEDFSVEEIAD	62	SRKRVTFAGTKDKRAVKLQYF	
P. furiosus	16	GIGGKIKVYPEDFIVKEVIP	70	HYSEIGFAGTKDRHAVTYQYI	
P. horikoshii	17	GIGGKIKVYPEDFIVKEVIP	71	HYSEIGFAGTKDRHAVTYQYI	
P. abyssi	13	GIGGRIKVKPEDFIVREIIP	66	HYSEIGFAGTKDRHAVTYQYI	
M. jannaschii	51	-TGGIIKKYPEDFIVEEITP	119	QRKHFGFAGNKDKYAVTTQRV	
M. thermauto.	13	GTGGRIRVHNRDFEVEEIP-	67	DRRRMGFAGMKDKRAVTRQWI	
S. solfataricus	21	NELDGTIERPNGFKVIEEID	79	LHSKVNYIGIKDA AITSQLV	
S. tokodaii	14	SPLKVFINRPDGFRVIEEIS	71	LHSKIHYIGIKDT AITEQIV	
S. acidocaldarius	22	EPLQISINRPEGFKVIEEIS	79	MKRKIHYLGIKDA AITQQII	
M. sedula	20	ASIRAEIPRPDGFRVTEEID	81	LGEKPRYLGIKDT AVTEQLI	
P. aerophilum	18	PSGGVIKRSPEDFVVEEVLL	86	SHRDVSVGGLKDTKAVTSQII	
A. pernix	25	RGPEALIMFPWGFRVVESRV	77	-CRLYSYAGLKDACSVSWQHV	
N. maritimus	20	GIGGKIRTTPEDFEVSEKIL	75	KGLRLKSLGLKDASAITEQFV	
C. symbiosum	26	GTGGKIRERHEDFEVSEVLT	79	AGLRYKALGLKDAAAITEQFV	
		Motif IIIa		Motif II	

B

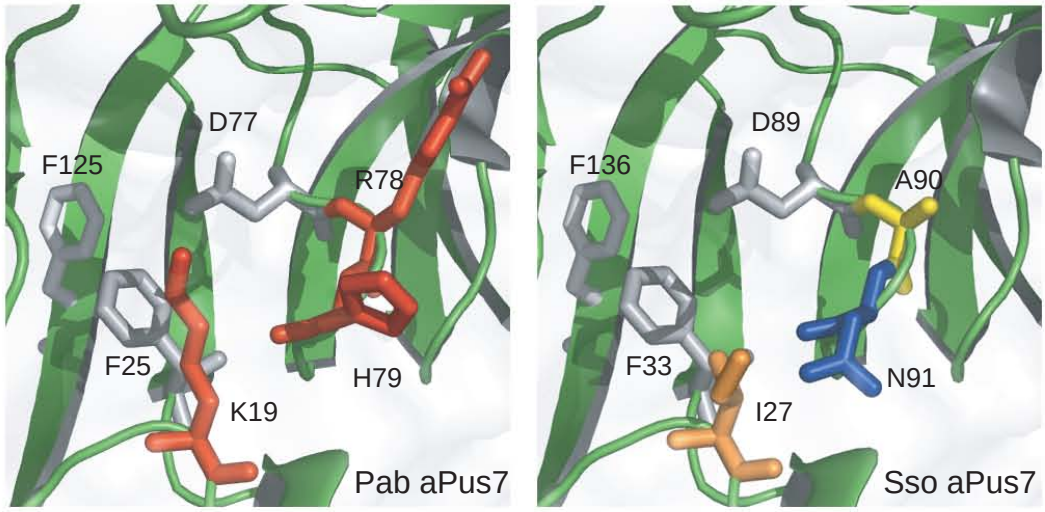


Figure 4

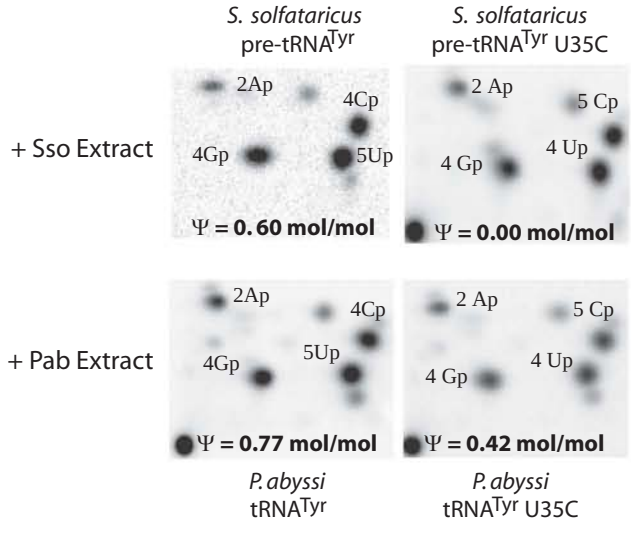
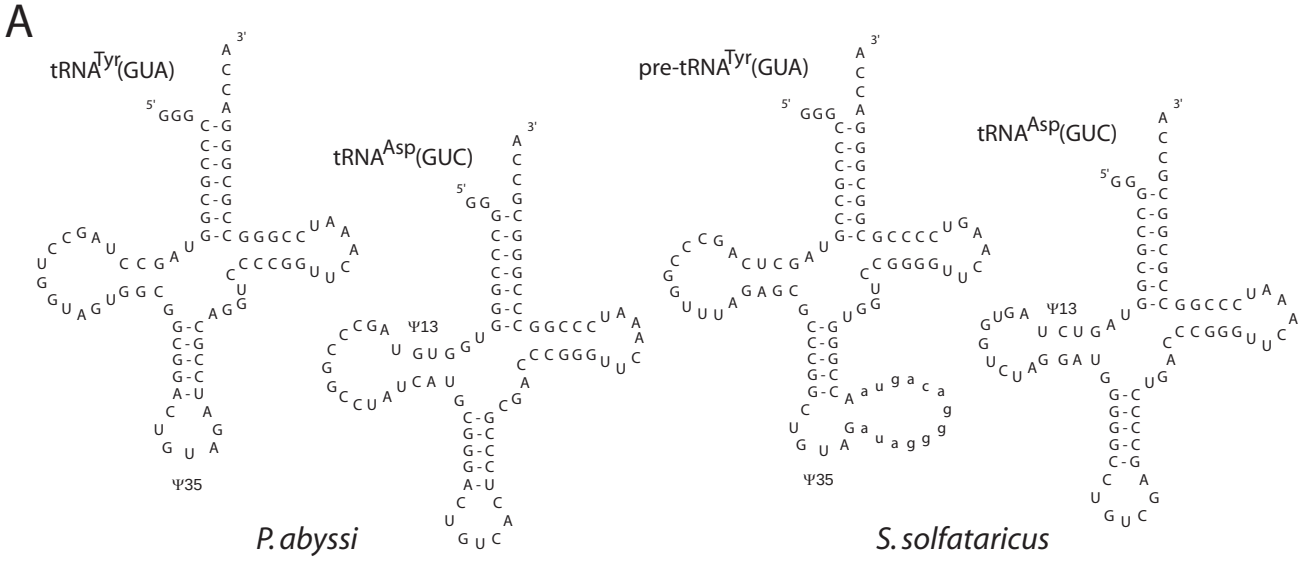
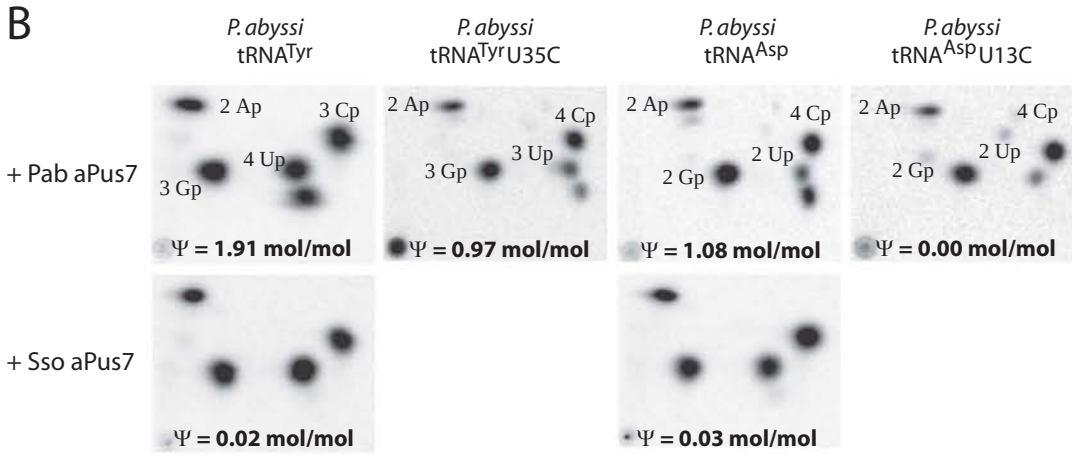

S. solfataricus pre-tRNA^Tyr
S. solfataricus pre-tRNA^Tyr U35C
+ Sso Extract
+ Pab Extract
2Ap
4Cp
4Gp
5Up
Ψ = 0. 60 mol/mol
2 Ap
5 Cp
4 Gp
4 Up
Ψ = 0.00 mol/mol
2Ap
4Cp
4Gp
5Up
Ψ = 0.77 mol/mol
2 Ap
5 Cp
4 Gp
4 Up
Ψ = 0.42 mol/mol
P. abyssi tRNA^Tyr
P. abyssi tRNA^Tyr U35C

Figure 5


A
tRNA^Tyr(GUA)
tRNA^Asp(GUC)
pre-tRNA^Tyr(GUA)
tRNA^Asp(GUC)
P. abyssi
S. solfataricus


B
P. abyssi tRNA^Tyr
P. abyssi tRNA^Tyr U35C
P. abyssi tRNA^Asp
P. abyssi tRNA^Asp U13C
+ Pab aPus7
+ Sso aPus7
Ψ = 1.91 mol/mol
Ψ = 0.97 mol/mol
Ψ = 1.08 mol/mol
Ψ = 0.00 mol/mol
Ψ = 0.02 mol/mol
Ψ = 0.03 mol/mol

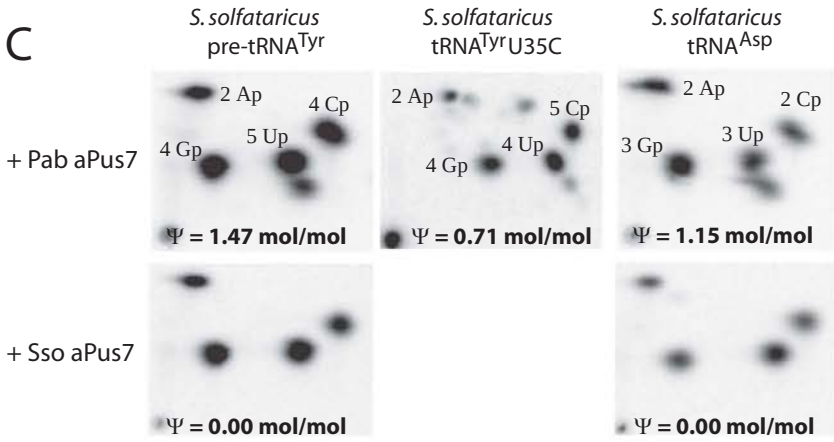

C
S. solfataricus pre-tRNA^Tyr
S. solfataricus tRNA^Tyr U35C
S. solfataricus tRNA^Asp
+ Pab aPus7
+ Sso aPus7
Ψ = 1.47 mol/mol
Ψ = 0.71 mol/mol
Ψ = 1.15 mol/mol
Ψ = 0.00 mol/mol
Ψ = 0.00 mol/mol

Figure 6

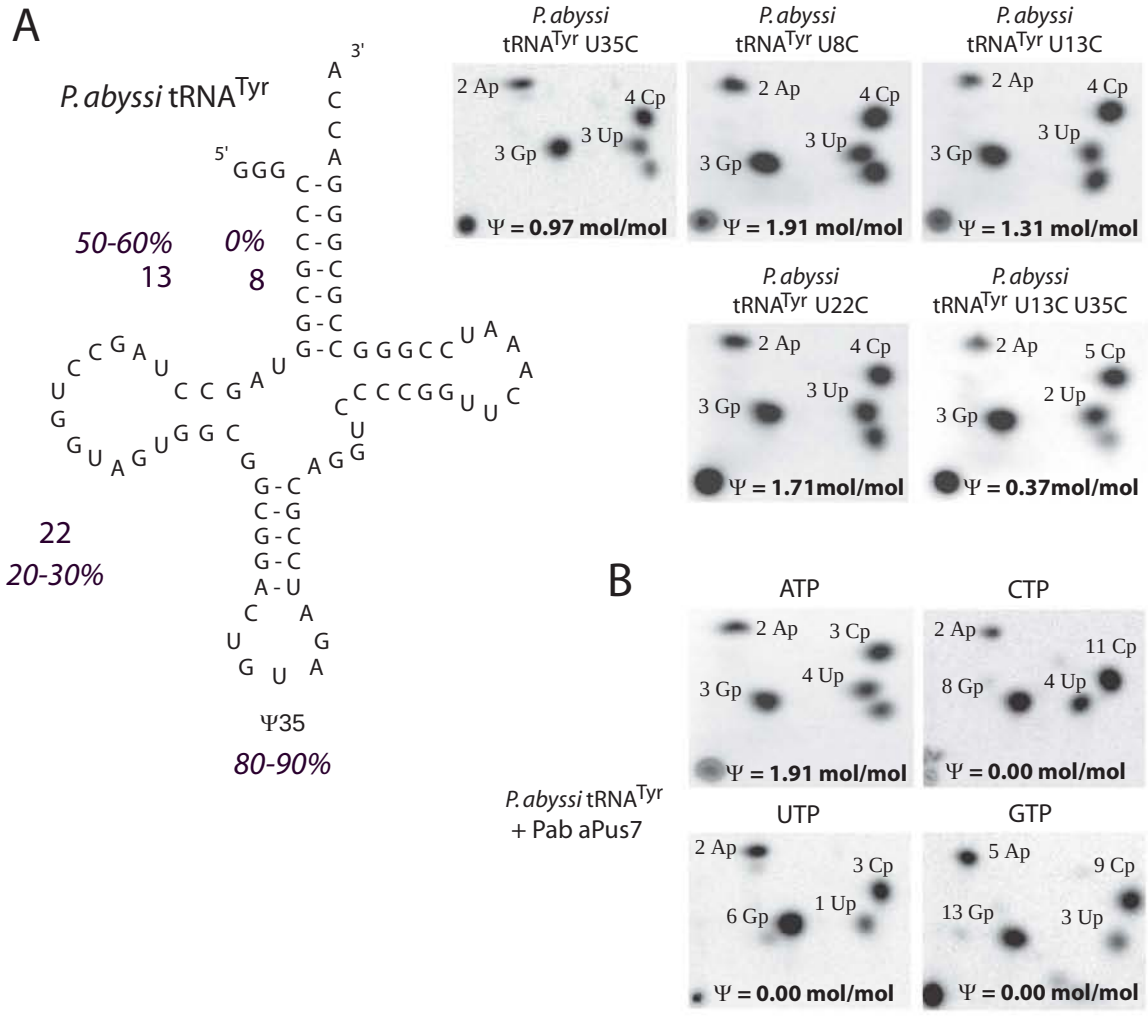

A
P. abyssi tRNATyr
50-60%
13
0%
8
22
20-30%
Ψ35
80-90%
P. abyssi tRNATyr U35C
2 Ap
4 Cp
3 Up
3 Gp
Ψ = 0.97 mol/mol
P. abyssi tRNATyr U8C
Ψ = 1.91 mol/mol
P. abyssi tRNATyr U13C
Ψ = 1.31 mol/mol
P. abyssi tRNATyr U22C
Ψ = 1.71mol/mol
P. abyssi tRNATyr U13C U35C
5 Cp
2 Up
Ψ = 0.37mol/mol
B
P. abyssi tRNATyr
+ Pab aPus7
ATP
3 Cp
Ψ = 1.91 mol/mol
CTP
11 Cp
8 Gp
Ψ = 0.00 mol/mol
UTP
6 Gp
1 Up
GTP
5 Ap
9 Cp
13 Gp

Figure 7

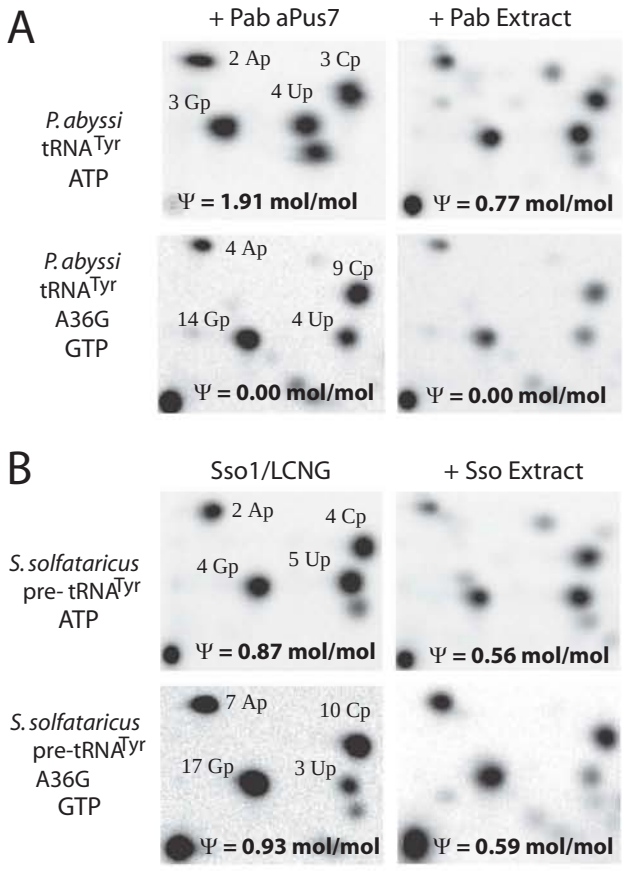

A
+ Pab aPus7
+ Pab Extract
P. abyssi tRNA Tyr ATP
2 Ap
3 Cp
4 Up
3 Gp
Ψ = 1.91 mol/mol
Ψ = 0.77 mol/mol
P. abyssi tRNA Tyr A36G GTP
4 Ap
9 Cp
14 Gp
4 Up
Ψ = 0.00 mol/mol
Ψ = 0.00 mol/mol
B
Sso1/LCNG
+ Sso Extract
S. solfataricus pre- tRNA Tyr ATP
2 Ap
4 Cp
4 Gp
5 Up
Ψ = 0.87 mol/mol
Ψ = 0.56 mol/mol
S. solfataricus pre-tRNA Tyr A36G GTP
7 Ap
10 Cp
17 Gp
3 Up
Ψ = 0.93 mol/mol
Ψ = 0.59 mol/mol

Figure 8

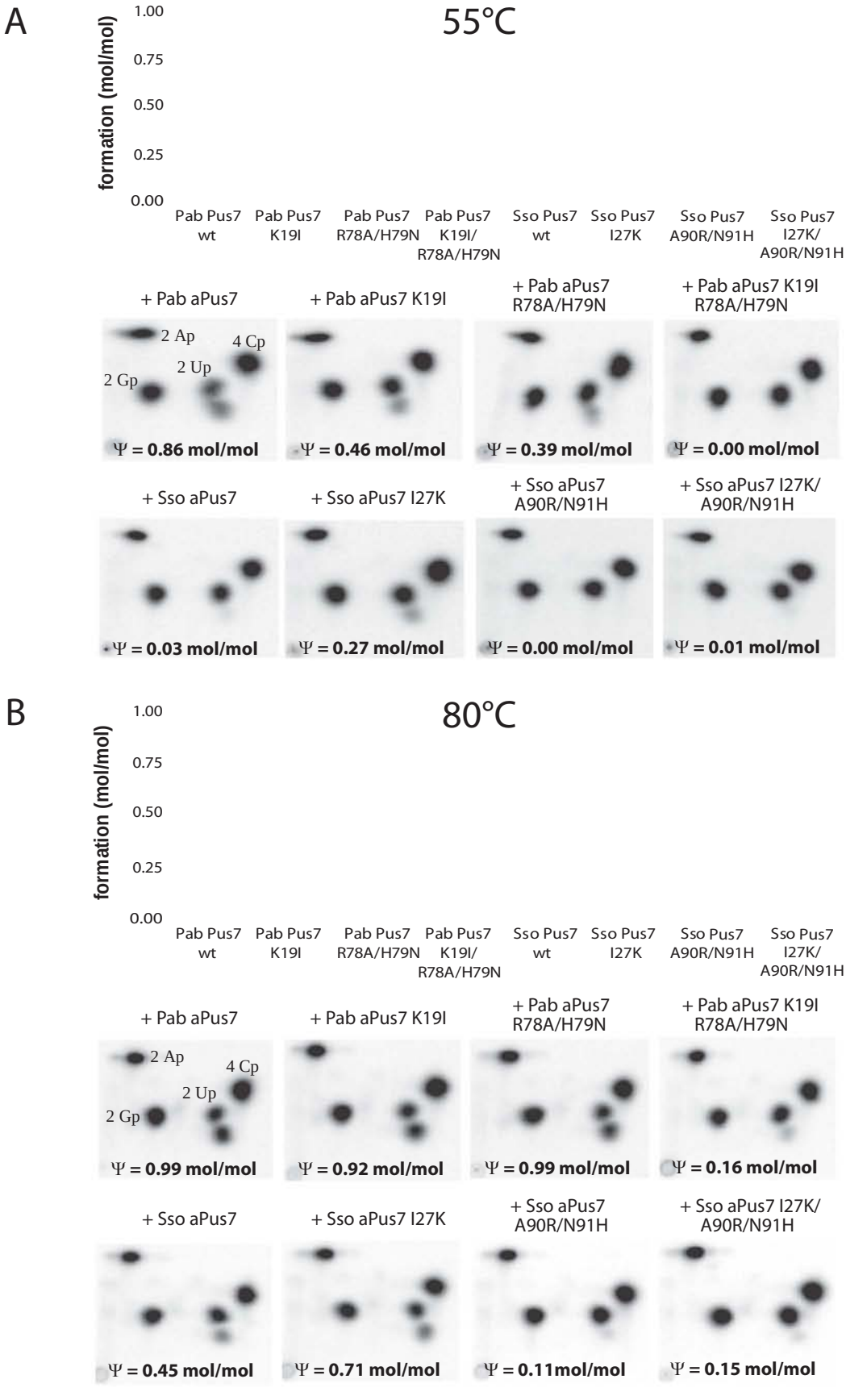
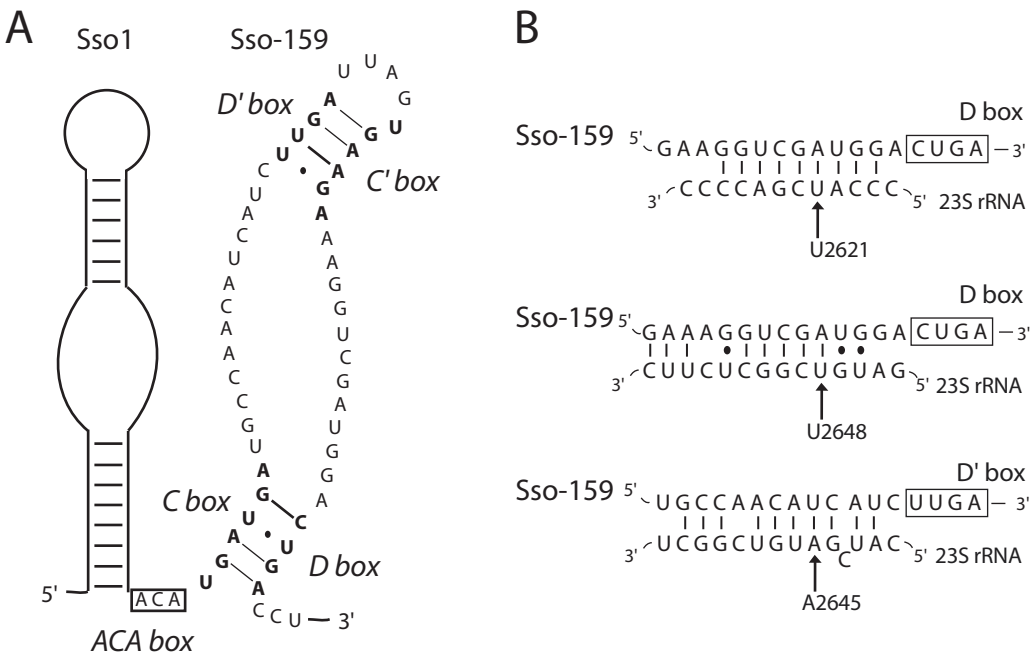

A
55°C
formation (mol/mol)
1.00
0.75
0.50
0.25
0.00
Pab Pus7 wt
Pab Pus7 K19I
Pab Pus7 R78A/H79N
Pab Pus7 K19I/ R78A/H79N
Sso Pus7 wt
Sso Pus7 I27K
Sso Pus7 A90R/N91H
Sso Pus7 I27K/ A90R/N91H
+ Pab aPus7
+ Pab aPus7 K19I
+ Pab aPus7 R78A/H79N
+ Pab aPus7 K19I R78A/H79N
2 Ap
4 Cp
2 Up
2 Gp
Ψ = 0.86 mol/mol
Ψ = 0.46 mol/mol
Ψ = 0.39 mol/mol
Ψ = 0.00 mol/mol
+ Sso aPus7
+ Sso aPus7 I27K
+ Sso aPus7 A90R/N91H
+ Sso aPus7 I27K/ A90R/N91H
Ψ = 0.03 mol/mol
Ψ = 0.27 mol/mol
Ψ = 0.00 mol/mol
Ψ = 0.01 mol/mol
B
80°C
formation (mol/mol)
1.00
0.75
0.50
0.25
0.00
Pab Pus7 wt
Pab Pus7 K19I
Pab Pus7 R78A/H79N
Pab Pus7 K19I/ R78A/H79N
Sso Pus7 wt
Sso Pus7 I27K
Sso Pus7 A90R/N91H
Sso Pus7 I27K/ A90R/N91H
+ Pab aPus7
+ Pab aPus7 K19I
+ Pab aPus7 R78A/H79N
+ Pab aPus7 K19I R78A/H79N
2 Ap
4 Cp
2 Up
2 Gp
Ψ = 0.99 mol/mol
Ψ = 0.92 mol/mol
Ψ = 0.99 mol/mol
Ψ = 0.16 mol/mol
+ Sso aPus7
+ Sso aPus7 I27K
+ Sso aPus7 A90R/N91H
+ Sso aPus7 I27K/ A90R/N91H
Ψ = 0.45 mol/mol
Ψ = 0.71 mol/mol
Ψ = 0.11mol/mol
Ψ = 0.15 mol/mol

Figure S1


A
Sso1
Sso-159
D' box
C' box
C box
D box
ACA box
5'
3'
B
Sso-159
D box
5' GAAGGUCGAUGGA CUGA 3'
3' CCCCAGCUACCC 5' 23S rRNA
U2621
Sso-159
D box
5' GAAAGGUCGAUGGA CUGA 3'
3' CUUCUCGGCUGUAG 5' 23S rRNA
U2648
Sso-159
D' box
5' UGCCAACAUC AUC UUGA 3'
3' UCGGCUGUAG UAC 5' 23S rRNA
C
A2645

PARTIE 6

ETUDE STRUCTURALE DE L'ARN:Ψ-SYNTHASE APUS7 DE P. ABYSSI ET S. SOLFATARICUS

1. Bilan des connaissances structurales sur les ARN:Ψ-synthases de la famille TruD/Pus7p

Lorsque j'ai débuté ce projet, la structure de 2 ARN:Ψ-synthases de la famille TruD/Pus7p était déjà connue : l'enzyme TruD d' *E. coli* (Kaya et al., 2004) que nous allons détailler et l'enzyme Pus7p de *M.* mazei (Northeast Structural Genomics Consortium, résultats non publiés). L'ARN:Ψ-synthase TruD est dépourvue de plusieurs insertions internes dans sa séquence et est donc beaucoup plus petite que l'enzyme Pus7p de *S. cerevisiae*. Cette protéine adopte un repliement en forme de « V », et peut être décomposée en deux domaines distincts. Le premier, représenté en gris dans les Figures 98A et 98B, est très similaire au domaine catalytique des ARN:Ψ-synthases précédemment cristallisées. Le second domaine (aa 155-303), représenté en orange dans les Figures 98A et 98B, ne présente aucune homologie ni de séquence, ni de structure avec des protéines connues, mais est conservé dans tous les homologues de TruD. Ce domaine a été appelé « domaine TruD » et pourrait correspondre à un nouveau domaine de fixation des ARN. L'agencement particulier des deux domaines crée une profonde crevasse, dont les parois sont couvertes d'acides aminés basiques (Fig. 98C) et qui semble correspondre au site de fixation de l'ARNt. L'acide aspartique catalytique (D80) est présent à la base de cette crevasse (Fig. 98A et B) (Kaya et al., 2004).

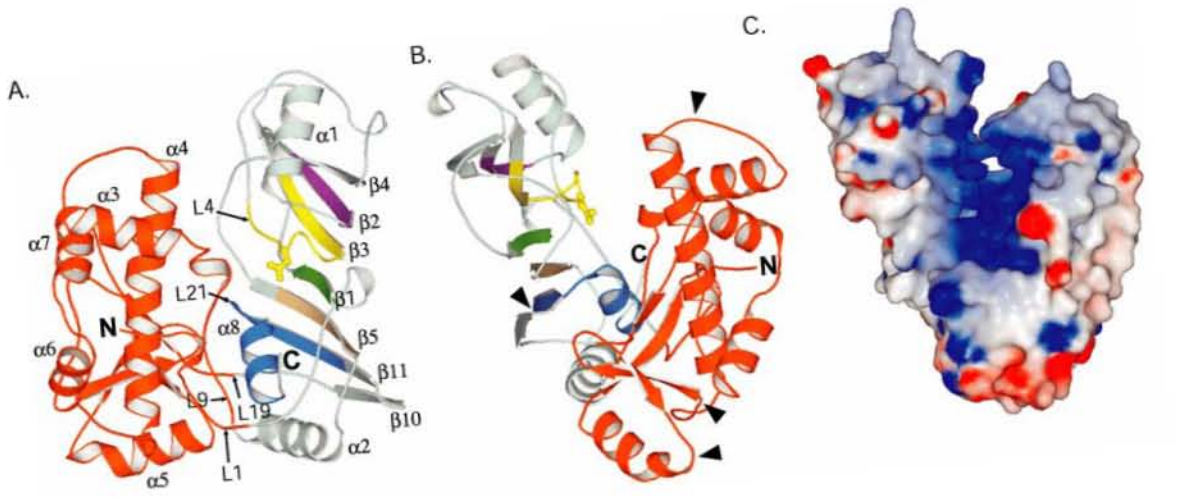


Figure 98 : Illustrations de la structure 3D (A et B) et des propriétés électrostatiques de surface (C) de l'ARN:Ψ-synthase TruD d'*E. coli* (Kaya et al., 2004).

A. Structure 3D de TruD: le domaine catalytique et le domaine TruD sont respectivement représentés en gris et en orange. Les différents motifs conservés au sein des ARN:Ψ-synthases sont indiqués: le motif I en violet, le motif II en jaune, le motif IIA en crème, le motif III en bleu et le motif IIIa en vert. L'acide aspartique catalytique (D80) est représenté sous forme ball and stick. B. Vue de la structure 3D de TruD après une rotation de 140° selon l'axe vertical par rapport à A. Les sites d'insertion de séquences additionnelles de taille supérieure à 8aa dans les homologues de TruD sont indiqués par des

flèches noires. C. Représentation de la surface moléculaire de TruD, avec le potentiel électrostatique de surface en bleu s'il est positif, en rouge s'il est négatif.

Comme cela a déjà été dit précédemment, cinq motifs sont conservés dans les enzymes appartenant aux familles d'ARN:Ψ-synthases TruA, TruB, RsuA et RluA (Del Campo et al., 2004). L'alignement de séquences basé sur les structures 3D et présenté dans la Figure 74B, montre que TruD, en dépit de l'absence d'homologie de séquences avec les autres ARN:Ψ-synthases, contient également ces 5 motifs et qu'ils sont structuralement superposables aux motifs présents dans les autres ARN:Ψ-synthases. Il est cependant intéressant de noter que dans le cas de TruD, le motif IIIa, généralement trouvé en aval du motif III, est présent en amont du motif I. Il apparaît donc que deux agencements différents des motifs au niveau de la structure primaire permettent néanmoins une disposition identique des motifs au sein de la structure 3D (Kaya et al., 2004).

Nous avons vu dans la partie Introduction de ce manuscrit, que le site actif des ARN:Ψ-synthases TruA, TruB, RsuA et RluD est constitué de 5 acides aminés provenant des motifs II, IIa, III et IIIa. (Del Campo et al., 2004). Le site actif étant constitué de l'acide aspartique catalytique (motif II), d'un résidu basique (Arg ou Lys) (motif III) qui forme un pont salin avec l'acide aspartique catalytique, d'un résidu tyrosine (motif IIa) qui va orienter la base de l'uridine cible par empilement des cycles aromatiques et de deux résidus hydrophobes (Ile ou Val du motif III et Leu du motif IIIa). Le site actif de TruD, superposé aux sites actifs de TruB et de TruA est présenté page 55 dans la Figure 35B (Partie 4 du chapitre Introduction). On remarque que dans le cas de TruD, l'acide aspartique catalytique (D80) du motif II est conservé, mais il forme un pont salin non plus avec un acide aminé basique du motif III (remplacé par un résidu thréonine), mais avec un résidu lysine de la boucle L1 (K21). De même le résidu tyrosine du motif IIa est remplacé par un résidu phénylalanine (F131) situé au sein du même motif, qui remplit exactement la même fonction que le résidu tyrosine. Le résidu leucine du motif IIIa, est remplacé par un résidu plus hydrophobe, le résidu F27 et le résidu isoleucine ou valine du motif III est remplacé par un résidu hydrophobe plus petit, le résidu A330. Il apparaît donc, que l'architecture générale du site actif de TruD est très similaire à celle du site actif des autres ARN:Ψ-synthases. Les acides aminés impliqués ne sont pas forcément identiques, mais peuvent remplir la même fonction (Kaya et al., 2004).

2. Résultats

Comme nous l'avons décrit dans l'article n°4, chez certaines espèces, en particulier, celles de l'ordre des Sulfolobales et *A. pernix* de l'ordre des Dessulfurococcales, des mutations d'acides aminés sont présentes dans les motifs II et IIIa qui font partie du site catalytique. Ces substitutions abolissent l'activité de aPus7 à la position 35 du pré-

ARNt^{Tyr}(GΨA), et seule une faible activité à la position 13 des ARNt est observée *in vitro* à température élevée. De ce fait, un sRNA H/ACA guide la modification à la position 35 du pré-ARNt^{Tyr}(GΨA) dans les espèces où Pus7p présente ce défaut d'activité. Il paraissait très intéressant de pouvoir comparer la structure 3D de l'enzyme aPus7 de *P. abyssi* qui possède les activités RNA:Ψ-synthases en position 35 du pré-ARN^{Tyr}(GΨA) et 13 de l'ARNt^{Asp}(GUC) à celle de l'enzyme de *S. solfataricus* qui ne possède qu'une faible activité en position 13. La comparaison des structures 3D de ces 2 protéines était susceptible d'apporter des informations générales sur le mode d'action des enzymes de la famille TruD. En effet, comme nous l'avons dit ci-dessus, on ne dispose actuellement que des structures 3D établies en absence d'ARNt substrat pour TruD d'*E. coli* et aPus7 de l'archaea *M. mazei*. La séquence en acide aminés de cette dernière protéine présentant une trop grande divergence par rapport à celle de *P. abyssi* pour qu'une structure 3D puisse être proposée par simple modélisation moléculaire. Nous avons donc entrepris un travail de production, de purification et de cristallisation des enzymes aPus7 de *P. abyssi* et de *S. solfataricus* en vue d'établir leur structure 3D par radiocristallographie.

Sur la base des conditions de production et de purification que nous avions établies pour la protéine Pus7p de levure (Partie 3 du chapitre Résultats) et en tenant compte des caractéristiques physico-chimiques des protéines d'archaea thermophiles, telles qu'une très grande solubilité et une thermostabilité accrue, je suis parvenu à produire et à purifier l'homologue de Pus7p chez *P. abyssi* dans des conditions permettant de réaliser des tests de cristallisation. Le travail de cristallogenèse, réalisé par C. Charron au laboratoire, a rapidement abouti à l'obtention de petits cristaux en forme de fines baguettes de 200 µm de longueur. Ces conditions de cristallisation ont par la suite été optimisées (Fig. 99) et les conditions de cristallisation suivantes qui donnaient lieu à des cristaux de qualité satisfaisante ont été retenues : 180 mM NaH_2PO_4, 18% polyéthylène glycol 3350, Concentration de Pab aPus7p = 37,5 mg/ml.

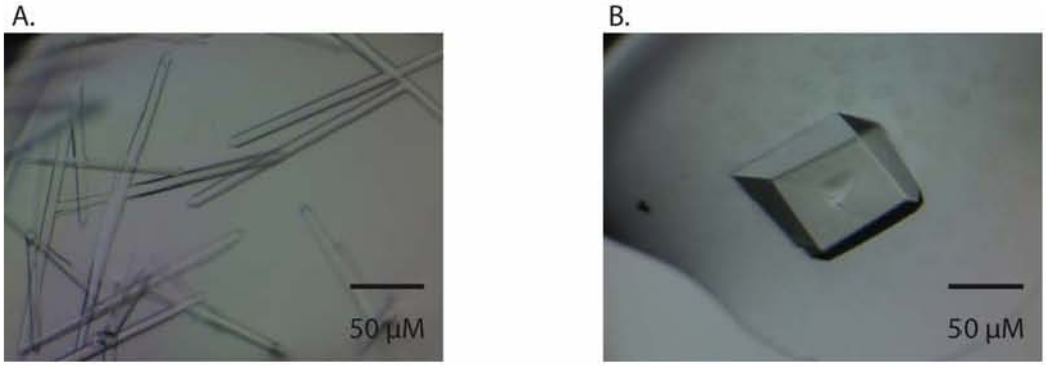


Figure 99 : Cristaux de l'ARN :Ψ-synthase aPus7 de *P. abyssi*

A. Cristaux obtenus à l'issue de la première série de test de cristallisation de l'enzyme.

B. Un des cristaux obtenus après optimisation des conditions de cristallisation. Ce cristal appartient au système orthorhombique.

Après optimisation des conditions de congélation des cristaux, nous avons réalisé une première collecte de données de diffraction des rayons X à 2,8 Å de résolution au laboratoire puis à 2,4 Å en utilisant le rayonnement synchrotron (ESRF). Ces données sont rédigées sous la forme d'un article, dont je suis premier auteur, à soumettre à Acta Crystallographica Section D.

Publication N°5 :

Expression, purification, crystallization and preliminary diffraction studies of the *Pyroccocus abyssi* tRNA:pseudouridine synthase aPus7

Alan Urban[1], Sebastien Muller[1], Jean-Michel Saliou[2], Sarah Sanglier[2], Yuri Motorin[1], Christiane Branlant[1] and Christophe Charron[1*]

Préparé pour soumission à Acta Crystallographica Section D

Publication n°5

Version #2, 13/08/2008 Prepared for Acta Crystallographica Section D

Expression, purification, crystallization and preliminary diffraction studies of the *Pyroccocus abyssi* tRNA:pseudouridine synthase aPus7

Alan Urban[1], Sebastien Muller[1], Jean-Michel Saliou[2], Sarah Sanglier[2], Yuri Motorin[1], Christiane Branlant[1] and Christophe Charron[1*]

[1]Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP Nancy I, Faculté des Sciences, BP 239, 54506 Vandoeuvre-les-Nancy Cedex, France

[2]LSMBO, Laboratoire de Spectrométrie de Masse Bio-Organique ECPM, Bâtiment R5, 67087 Strasbourg, France CNRS-UMR 7509/Université Louis-Pasteur, Strasbourg, France.

Keywords: Pseudouridine, Pab Pus7p, tRNA, pre-tRNA^{Tyr}, RNA modification

*Corresponding author :

Christophe CHARRON

Laboratoire Maturation des ARN et Enzymologie Moléculaire

UMR 7567 CNRS-UHP Nancy I

Faculté des Sciences, BP 239

54506 Vandoeuvre-les-Nancy Cedex, FRANCE

Tel. 33.3.83.68.43.09

FAX. 33.3.83.68.43.07

ABSTRACT

Pyroccocus abyssi Pus7 is a recently characterized archaeal RNA:pseudouridine (Ψ) synthase catalyzing the formation of Ψ13 in $tRNA^{Asp}$(GUC) and Ψ35 in pre-$tRNA^{Tyr}$(GΨA). Pab Pus7 belongs to the universally conserved TruD-like family of RNA:Ψ-synthases present in bacteria, archaeabacteria and eukaryotes. Recombinant Pab Pus7 expressed in *E. coli* was crystallized using the hanging-drop vapour-diffusion method. The crystal belonged to space group orthorhombic, with the unit-cell parameters a = 69.0, b = 87.7, c = 112.6 Å and two molecules in the asymetric unit. A complete native X-ray diffraction data set was collected to 2.40 Å resolution using synchrotron radiation.

1. Introduction

Pseudouridine (5-β-D-ribofuranosyluracil, Ψ) is the most abundant modified nucleoside found in RNAs. They are produced by conversion of U residues catalysed by RNA:pseudouridine synthases in rRNA (for review, (Ofengand *et al.*, 2001) , tRNA (Limbach *et al.*, 1994, Limbach *et al.*, 1995), snRNA (for review (Massenet *et al.*, 1998) and snoRNA (Kato & Harada, 1984, Reddy *et al.*, 1972). Extensive studies during the last ten years allowed the discovery and characterization of almost the complete set of bacterial and yeast RNA:Ψ-synthases. They can be divided into five families based on their degree of amino acid sequence homology : TruA, TruB, RsuA, RluA and TruD (Kaya & Ofengand, 2003, Koonin, 1996). 3D structures were established for at least one representative of each RNA:Ψ-synthases families (Del Campo *et al.*, 2004, Foster *et al.*, 2000, Hoang & Ferre-D'Amare, 2001, Kaya *et al.*, 2004, Sivaraman *et al.*, 2002). This shows that their catalytic domains display the same overall fold, suggesting a common ancestry (Hoang & Ferre-D'Amare, 2001, Kaya *et al.*, 2004, Mueller, 2002, Sivaraman *et al.*, 2002).

The *P. abyssi* (Pab) aPus7p enzyme is a member of the TruD family of RNA:Ψ-synthases. Proteins of this family show little sequence homology with RNA:Ψ-synthases of the other families. However, they share with them several conserved amino acids in their catalytic site, including the catalytic aspartate (D) residue (Kaya *et al.*, 2004). Previous studies already reported the presence of numerous Pus7p orthologs in all kingdoms of life (Kaya & Ofengand, 2003). Multiple sequence alignments of proteins of the Pus7/TruD proteins family reveals strong length variations: bacterial enzymes are the shortest ones, archael enzymes have with 3 insertions, while eukaryal orthologs are the largest enzymes with at least 5 insertions. This broad diversity and the presence of additional protein domains may be linked to different RNA-substrate specificities of these enzymes. Indeed, *E. coli* TruD is only

responsible for Ψ13 modification on given tRNA $tRNA^{Glu}$(UUC), while Pab aPus7 catalyses both Ψ13 and Ψ35 formation in $tRNA^{Asp}$(GUC) and pre-$tRNA^{Tyr}$ (GΨA) respectively. {Muller et al., unpublished}. Furthermore, the yeast Pus7p enzymes has a large variety of substrates, it catalyzes Ψ formation at positions 13 in many cytoplasmic tRNAs, 35 in the pre-$tRNA^{Tyr}$ (GΨA) (Behm-Ansmant *et al.*, 2003), 35 in U2 snRNA (Ma *et al.*, 2003) and 50 in 5S rRNA (Decatur & Schnare, 2008). Structural data obtained for *E. coli* TruD (Kaya *et al.*, 2004) reveales a V shaped structure composed of two distinct domains. One closely resembles the catalytic domain of proteins that belongs to other RNA:Ψ-synthases. The other domain, called TRUD domain, lack any similarity in either sequence or structure to known protein domains but appears to be a novel RNA binding domain (Kaya *et al.*, 2004). Some of the additionnal domains in archaeal and eukaryal Pus7 RNA:Ψ-synthases may allow the accomodation of additional substrates of these enzymes compared to *E. coli* TruD.

Here, we describe the cloning, overexpression, crystallization and prelimary crystallographic characterisation of Pus7 from *P. abyssi*.

2. Methods and results

2.1. Overexpression and purification

Transformed BL21-CodonPlus(DE3) competent cells (Stratagene) were transformed with PET28b-PabPus7 {Muller et al., unpublished} and grown in 4 l of ZYM5052 auto-inducting media (Studier, 2005) complemented with 100 µg ml^{-1} kanamycin and 50 µg ml^{-1} chloramphenicol. Cells were first grown at 310 K with vigorous shaking to an OD_{600} of 0.8 followed by 20 h of incubation at 293 K. Harvest cells were ressuspended in lysis buffer (NaCl 1M, Tris-HCl 20mM pH 7.5) and disrupted by sonication for 5 min using a Branson S250 sonicator. Cell debris were pelleted by centrifugation at 20 000 g for 15 min and the supernatant was immediately heated to 348 K for 30 min. This solution was centrifugated 30 min at 20 000*g* for 30 min. Polyethylene imine (Sigma) was added to the supernatant to the final concentration of 0.1%. Nucleic acid contaminants were removed by centrifigation for 30 min at 20 000*g*. The surpernatant was loaded onto a 25 ml column of Ni-Sepharose High Performance resin (Amersham Biosciences) pre-equilibrated with lysis buffer to remove unbound material. The column was washed with the washing buffer (NaCl 1M, Imidazole 50 mM, Tris-HCl 20mM pH 7.5). His-Tagged Pab aPus7 was eluted with the elution buffer (NaCl 1M, Imidazole 1 M, Tris-HCl 20mM pH 7.5) and the fraction containing the purified protein were pooled and applied onto a Hiload 26/60 Superdex 200 gel-filtration column (Amersham Biosciences) equilibrated in the GF buffer (NaCl 500 mM, Tris-HCl 10mM pH 7.5). The apparent molecular weight on gel-filtration was of 50 kDa, suggesting that Pab aPus7 was a functional monomer. Salt removal and protein concentration were performed at the same time, using a Vivaspim YM-10 centrifugal filter device (Vivascience). The Pab sPus7 protein was them concentrated to about 45 mg ml^{-1} in 10 mM Tris-HCl pH 7.5 buffer, as estimated by Bradford method (Biorad). The 17-residue N-terminal tag was not cleaved off

prior to crystallization trials. The purified protein exhibited a single band on SDS-PAGE (Fig. 1) analysis, by electro-spray ionisation mass-spectrometry (ESI-MS) spectrum revealed molecular ions corresponding to the expected mass of the His6-tagged wild-type Pab aPus7 (49863.33±0.71 Da) and a molecule with a mass of 50041.23±0.78 Da (178 Da larger than the expected mass) suggested the presence of a post-translational α-N-6 gluconoylation modification at the N-terminus His6-tag sequence (data not shown) (Geoghegan *et al.*, 1999).

2.2. Enzymatic assays

In vitro transcribed *P. abyssi* GE5 $tRNA^{Asp}$(GUC) was used as a substrate to test for the Pab aPus7 enzymatic activity. The gene encoding the *P. abyssi* $tRNA^{Asp}$(GUC) was amplifed by PCR using genomic DNA as the template and the following forward and reverse primers (5'-TAATACGACTCACTATAGGGCCCGGGTGGTGTAGCCCGGCC-3' and 5'-TGGCGCCCGGGCCGGGATTTGAACCCGGGTCGCGGGAGT-3', respectively. The forward primer generated a T7 polymerase promoter and the reverse primer contains an MvaI restriction site. PCR amplification was performed using the GoTaq Flexi DNA polymerase (Promega). The amplified DNA fragment was cloned into the SmaI restriction site of plasmid pUC18. The template for *in vitro* transcription was obtained by MvaI restriction of this plasmid. The Pab aPus7 enzymatic activity was tested at 338 K in 50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$ using 1-2 fmol of [α-^{32}P]ATP radiolabelled T7-runoff transcript of *P. abyssi* $tRNA^{Asp}$(GUC) as the substrate. The reaction was terminated by phenol extraction and ethanol precipitation. The tRNA was hydrolyzed by RNase T2 and the resulting 3' phosphate-mononucleotides were fractionated by 2D thin-layer cellulose chromatography (Keith, 1995). The radioactivity in the spots was quantified using the ImageQuant software after exposure of a phosphorimager screen. The amount of pseudouridine residue formed was determined taking into account the total number of U residues in the tRNA.

2.3. Crystallisation

Screening for optimal crystallization conditions was performed using the sitting-drop method of vapour diffusion at 293 K. The 300 nl drops were equilibrated over 100 µl reservoirs. A total of 918 conditions were covered with eleven screens (Hampton Research Screens I and II, Hampton Research PEG/Ion Screen, Hampton Research SaltRX Screen, Molecular Dimensions Limited Clear Strategy Screens I and II, Molecular Dimensions Limited 3D Structure Screen I and II, deCODE Genetics Wizard Screens I and II and Jena BioSciences JBScreen 1-10). In each condition, 100 nl of a 45 $mg.l^{-1}$ Pab aPus7 solution was mixed with 100 nl of the reservoir solution using a crystallization robot Mosquito (TTPLabtech).

2.4. Diffraction measurements

A first set of X-ray data was collected in our laboratory at 100 K on a X8-Proteum diffractometer (Bruker AXS). The wavelength of the incident radiation was 1.542 Å; the crystal-to-detector distance was 75 mm with 0.25° oscillation per CCD image. Data were processed using the X8-Proteum^{Plus} software (BrukerAXS).

X-ray diffraction intensity data were collected at the European Synchrotron Radiation Facility (ESRF) at Grenoble (France) on beamline ID14-4. Prior to data collection, a suitable native crystal was soaked in a cryobuffer identical to the mother liquor but with 25% glycerol. The crystal was then flash-cooled in a nitrogen-gas stream at 100 K before mounting. The wavelength of the incident radiation was 0.9385 Å and the crystal-to-detector distance was 395 mm. The diffraction spots were recorded on a Quantum Q315r ADSC X-ray detector with a 0.2° oscillation for each CCD image; a total range of 110° was covered. Data were processed using the MARXDS package (Klein, 1993; Kabsch, 1993).

2.5. Data collection and analysis

Large crystals suitable for X-ray diffraction analysis (Fig. 2) were obtained by mixing 1 μl of a 45 mg ml^{-1} protein solution in 10 mM Tris-HCl pH 7.5 and 2 μl of reservoir solution containing 180 mM sodium dihydrogen phosphate and 18% (w/v) polyethylene glycol 3350. The 3 μl drops were equilibrated over 200 μl reservoirs at 285 K.

A total of 71892 reflections in the resolution range 50-2.79 Å were collected in our laboratory. They were reduced to 16302 unique reflections. Overall, the data set had an Rsym of 5.0 % on intensities (Rsym = $\Sigma|I-\langle I\rangle|/\Sigma I$) and was 92.0 % complete (Table 1). Autoindexing routines gave solutions consistent with an orthorhombic space group, which enabled unit-cell dimensions to be refined to a = 69.0, b = 87.7, c = 112.6 Å. Systematic extinctions [$I/\sigma(I) < 3$] of h00 (with h = 2n + 1), 0k0 (with k = 2n + 1) and 00l (with l=2n + 1) reflections were observed, corresponding to the P21212 space group. Assuming two monomers (MW 44 kDa) in the asymmetric unit, the predicted packing density VM is 1.94 $Å^3.Da^{-1}$ and the solvent content is 36%. These values are in good agreement with those for other proteins (Matthews, 1968).

A second set of X-ray data was collected to a resolution limit of 2.4 Å using a synchrotron X-ray source (Table 1). A total of 106075 reflections in the resolution range 100-2.42 Å were collected and reduced to 25256 unique reflections. Overall, the data set had an Rsym of 3.3 % on intensities (Rsym = $\Sigma|I-\langle I\rangle|/\Sigma I$) and was 96.3 % complete (Table X).

Initial attempts to solve the crystal structure of Pus7 by molecular replacement with the program AmoRe (Navaza, 1994) failed. Therefore, our efforts are currently focused on the acquisition of phase information from selenomethionine-substituted protein. We have produced crystals of the selenomethionine (SeMet) substituted protein and we are currently optimizing the crystallization conditions for SeMet crystals of Pus7 for use in the MAD method.

ACKNOWLEDGEMENTS

We thank F. Matsunaga and P. Forterre (CNRS, Orsay, France) for providing genomic DNA from *P. abyssi* GE5 strain and the staff of ID14-4 (ESRF, France) for assistance with diffraction experiments. This work was supported by laboratory funds from the CNRS and the "Ministère de la jeunesse, de l'éducation nationale et de la recherche" and the PRST "Bioingénerie" of Région Lorraine. A. URBAN was a pre-doctoral fellow from the "Ministère de la jeunesse, de l'éducation nationale et de la recherche".

REFERENCES

Behm-Ansmant, I., Urban, A., Ma, X., Yu, Y. T., Motorin, Y. & Branlant, C. (2003). RNA (New York, N.Y 9, 1371-1382.

Decatur, W. A. & Schnare, M. N. (2008). Molecular and cellular biology.

Del Campo, M., Ofengand, J. & Malhotra, A. (2004). RNA (New York, N.Y 10, 231-239.

Foster, P. G., Huang, L., Santi, D. V. & Stroud, R. M. (2000). Nat Struct Biol 7, 23-27.

Geoghegan, K. F., Dixon, H. B., Rosner, P. J., Hoth, L. R., Lanzetti, A. J., Borzilleri, K. A., Marr, E. S., Pezzullo, L. H., Martin, L. B., LeMotte, P. K., McColl, A. S., Kamath, A. V. & Stroh, J. G. (1999). Analytical biochemistry 267, 169-184.

Hoang, C. & Ferre-D'Amare, A. R. (2001). Cell 107, 929-939.

Kato, N. & Harada, F. (1984). Biochim Biophys Acta 782, 127-131.

Kaya, Y., Del Campo, M., Ofengand, J. & Malhotra, A. (2004). The Journal of biological chemistry 279, 18107-18110.

Kaya, Y. & Ofengand, J. (2003). RNA (New York, N.Y 9, 711-721.

Keith, G. (1995). Biochimie 77, 142-144.

Koonin, E. V. (1996). Nucleic acids research 24, 2411-2415.

Limbach, P. A., Crain, P. F. & McCloskey, J. A. (1994). Nucleic acids research 22, 2183-2196.

Limbach, P. A., Crain, P. F., Pomerantz, S. C. & McCloskey, J. A. (1995). Biochimie 77, 135-138.

Ma, X., Zhao, X. & Yu, Y. T. (2003). The EMBO journal 22, 1889-1897.

Massenet, S., Mougin, A. & Branlant, C. (1998). Posttranscriptional modifications in the U Small Nuclear RNAs, The modification and editing of RNA, edited by H. Grosjean & R. Benne, pp. 201-228. Washington, DC: ASM Press.

Mueller, E. G. (2002). Nat Struct Biol 9, 320-322.

Ofengand, J., Malhotra, A., Remme, J., Gutgsell, N. S., Del Campo, M., Jean-Charles, S., Peil, L. & Kaya, Y. (2001). Cold Spring Harb Symp Quant Biol 66, 147-159.

Reddy, R., Ro-Choi, T. S., Henning, D., Shibata, H., Choi, Y. C. & Busch, H. (1972). The Journal of biological chemistry 247, 7245-7250.

Sivaraman, J., Sauve, V., Larocque, R., Stura, E. A., Schrag, J. D., Cygler, M. & Matte, A. (2002). Nat Struct Biol 9, 353-358.

Studier, F. W. (2005). Protein expression and purification 41, 207-234.

TABLES

Table 1 Statistics of X-ray data measurement for *P. abyssi* Pus7 crystals.

	ESRF data	'In-house' data
Resolution limit (Å)	2.42 (2.49-2.42)	2.79 (2.89-2.79)
No. unique reflections	25256	16302
Rsym (%)	3.3 (41.7)	5.0 (13.4)
Completeness (%)	96.3 (96.4)	92.0 (50.0)
$\langle I/\sigma(I) \rangle$	21.8 (3.4)	17.9 (4.5)
Multiplicity	4.2	4.4

Values in parentheses correspond to the last resolution shell.

FIGURE LEGENDS

Figure 1 Polyacrylamide gel electrophoresis under denaturing conditions of the purified *P. abyssi* Pus7. Lane 1, crude extract. Lane 2, supernatant after thermal clarification. Lane 3, supernatant after nucleic acid removal by precipitation with PEI. Lane 4, molecular-weight markers (GE Healthcare). The molecular weight are given on the left of the SDS-PAGE gel. Lanes 5-8 fractions collected from the Ni-Sepharose column. Lane 9 pic fraction collected from the gel-filtration column.

Figure 2 Crystal of *P. abyssi* Pus7 obtained (A) from condition n°X of the index screen (B) from optimised conditions. The crystal dimension is about 0.100mm.

Figure 1

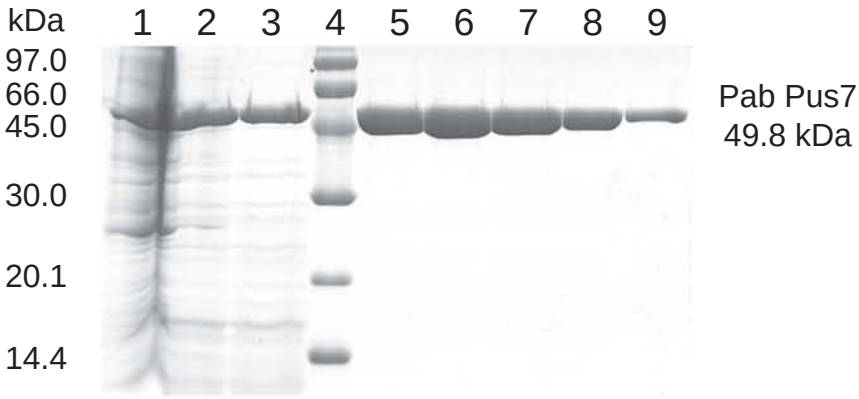
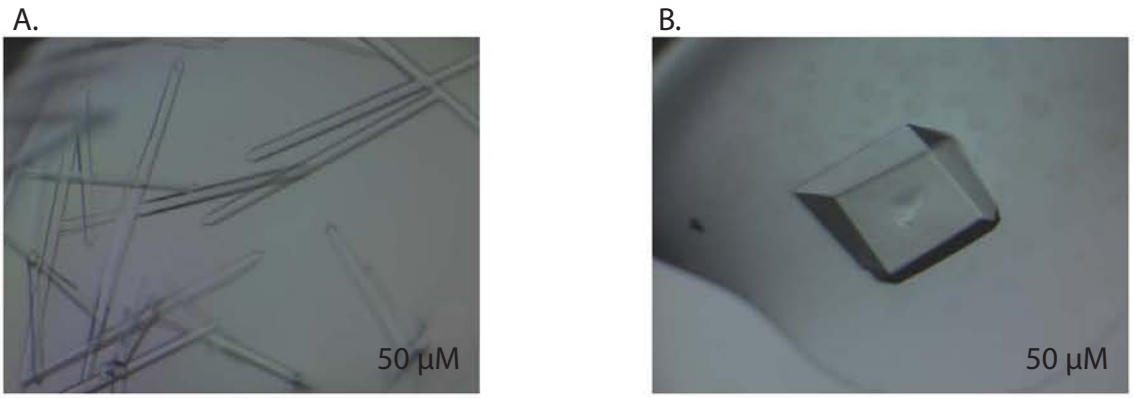


Figure 2



3. Résultats complémentaires

3.1. Production, purification et cristallisation de la protéine Pab aPus7 sélénométhionylée

Comme C. Charron n'était pas parvenu à résoudre la structure de la protéine aPus7 de *P. abyssi* par la méthode de remplacement moléculaire en utilisant comme modèle la structure de la protéine Pus7 de *Methanosarcina mazei* (PDB ID 1Z2Z ; 35 % d'identité de séquence), il était nécessaire d'obtenir des cristaux de la protéine sélénométhionylée afin de pouvoir réaliser le phasage de la protéine par la méthode MAD.

Le marquage de la protéine Pab aPus7 a été réalisé en utilisant une souche bactérienne auxotrophe pour la méthionine (*E. coli* B834 (DE3) RIL) (Stratagene). Les conditions de culture ont dues être adaptées, puisque la croissance a été réalisée dans un milieu pauvre contenant de la sélénométhionine à la place de la méthionine, suivant le protocole décrit dans la partie Matériels et Méthodes. La purification a aussi été réalisée en 2 étapes : une étape de chromatographie d'affinité sur colonne Ni-Sépharose (GE-Healthcare), suivie d'une étape de gel filtration préparative sur colonne Superdex 75 (GE-Healthcare). L'analyse des fractions récoltées au cours des différentes étapes de purification révélait que le rendement de production de la protéine Pab aPus7 sélénométhionylée était environ 2,5 fois plus faible que celui de la protéine non marquée, la pureté des échantillons était comparable et donc compatible avec des tests de cristallisation (résultats non présentés).

Afin de confirmer l'incorporation efficace de la sélenométhionine dans la protéine Pab aPus7p, nous avons fait réaliser une analyse par spectrométrie de masse (ESI-MS) par J-M. Saliou dans l'équipe A. Van Dorsselaer au Laboratoire de Spectrométrie de Masse Bio-Organique de Strasbourg. Pour cela, les échantillons on été déssalés par gel filtration sur colonne Zeba micro (Pierce) contre de l'acétate d'ammonium 300 mM pH 7,5 et dilués dans un solvant Acétonitrile/Eau/Acide formique (50/50/1) de façon à obtenir une concentration finale en protéine de 15 µM. L'analyse de la protéine en condition dénaturante a été réalisée sur un spectromètre de masse muni d'une source à ionisation nanospray (Nanomate, Advion) couplé à un analyseur à temps de vol (LCT, Waters). Les résultats des analyses démontrent que les échantillons protéiques contenaient 4 espèces majoritaires (Fig. 101) (correspondant aux séries de pics A, B, C et D) :

- A (50240.98±3.28) dont la masse moléculaire mesurée correspondant à la masse moléculaire théorique de la protéine Pab aPus7 possèdent l'étiquette polyhistidine à l'extrémité N-terminale sans la méthionine initiatrice et dont les 8 méthionines restantes sont sélénométhionylées (50241 Da).

- B (50415.93±2.94) dont la masse moléculaire mesurée correspondant à la masse moléculaire théorique de la forme gluconoylée de la protéine A (+178 Da → 50 419 Da).

- C (50282.91±4.09) dont la masse moléculaire mesurée s'approche de la masse moléculaire théorique de la protéine A + 2 adduits Na^+ (+2X21 Da → 50283 Da).

- D (15073.70±0.92 Da) correspondant à un contaminant.

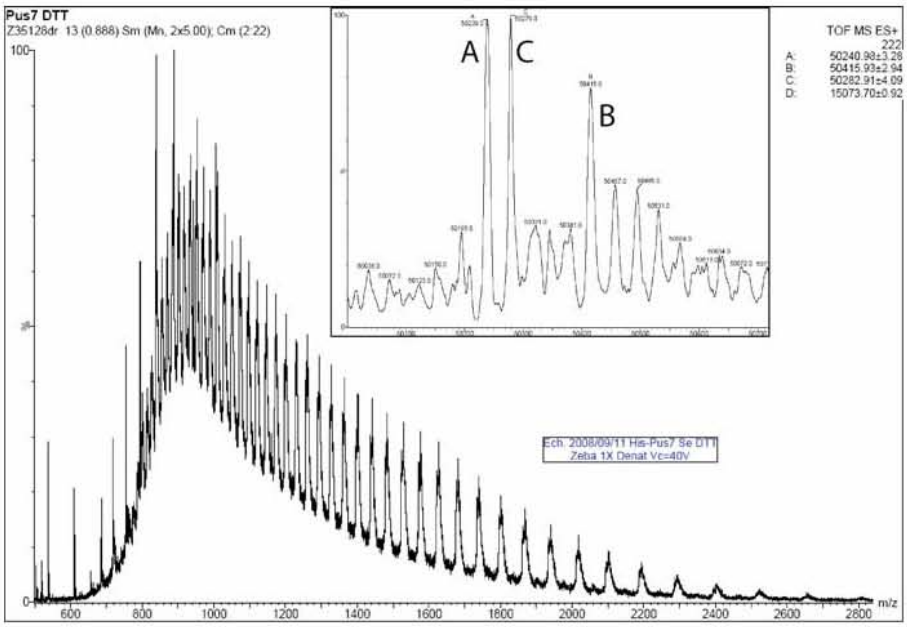


Figure 101 : Spectre obtenu lors de l'analyse de la protéine Pab aPus7 en conditions dénaturantes par spectrométrie de masse et spectre déconvolué

Afin de conforter le résultat obtenu, l'échantillon a été incubé en présence de thrombine (toute la nuit à 4°C) ce qui permettait d'éliminer l'étiquette polyhistidine et d'éviter le dédoublement de pic dû à la gluconoylation de la protéine portant cette étiquette (et probablement la phosphogluconoylation). Comme attendu, une seule espèce majoritaire était alors identifiée, dont la masse moléculaire mesurée correspondait à la masse moléculaire théorique de la protéine Pab aPus7 (sans l'étiquette) sélénométhionylée sur les 8 méthionines (48490 Da) (Figure 102).

Pour le moment, nous n'avons pas réussi à obtenir des cristaux exploitables de la protéine Pab aPus7 sélénométhionylée dans les conditions de cristallisation utilisées pour la protéine non sélénométhionylée mais C. Charron travaille actuellement à l'optimisation des conditions de cristallogenèse.

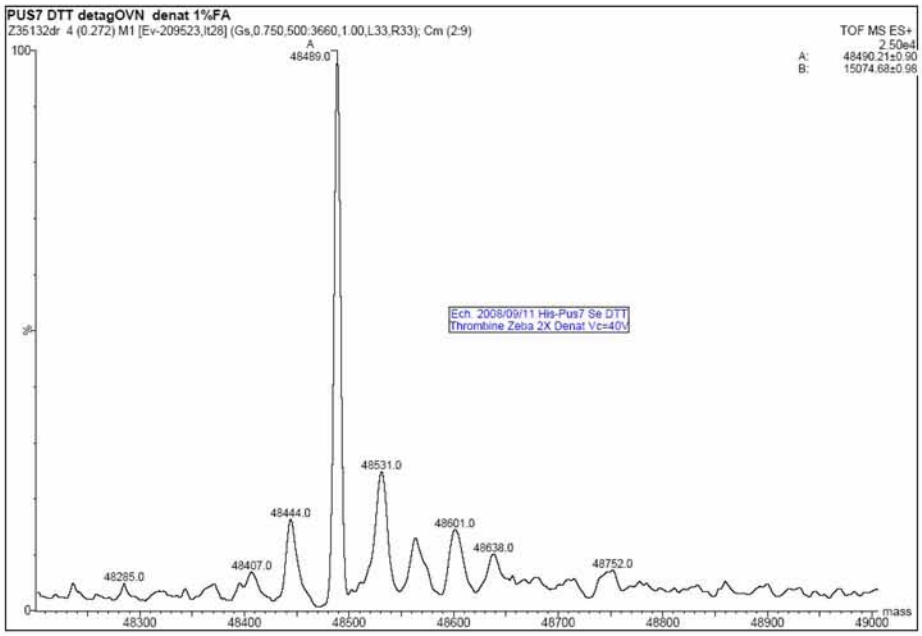


Figure 102 : Forme déconvoluée du spectre obtenu lors de l'analyse par spectrométrie de masse de la protéine Pab aPus7 après clivage de l'étiquette His_6 par la thrombine

3.1. Production, purification et cristallisation de la protéine Sso aPus7

Nous avons utilisé les conditions établies pour la production de l'enzyme Pab aPus7 pour produire et purifier l'enzyme aPus7 de *S. solfataricus*. Des essais de cristallisation ont été réalisés. Ils ont permis d'obtenir des cristaux dans plusieurs conditions de cristallisation (Fig. 102). La caractérisation de ces cristaux par diffraction des rayons X montre que leur pouvoir diffractant est faible (4,5 Å au plus bas) pour toutes les conditions de cristallisation obtenues. Ces conditions de cristallisation sont donc en cours d'optimisation afin d'améliorer le pouvoir diffractant des cristaux.

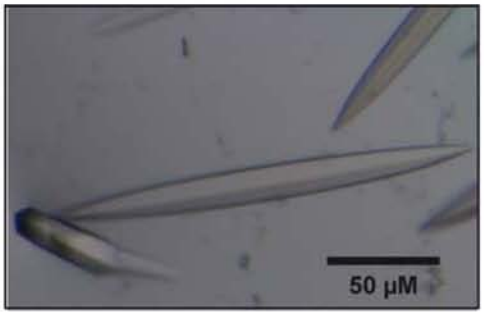


Figure 102 : Cristaux obtenus pour aPus7 de *S. solfataricus*

Les conditions de cristallisation sont les suivantes : 30% PEG 400, 0,1 M Acétate de Sodium pH 4.6, 0.1 M Chlorure de Cadmium, Concentration de l'enzyme Sso aPus7 = 38,5 mg/ml.

Perspectives

Si l'on dresse un bilan des connaissances actuelles concernant les ARN:Ψ-synthases chez *S. cerevisiae*, on constate que toutes les enzymes responsables de la formation des résidus Ψ dans les ARNt ont été caractérisées, à l'exception de celle responsable de la formation du résidu Ψ72 de l'ARNt^{Met} mitochondrial (Tableau 22). Néanmoins, les enzymes responsables de la formation des résidus Ψ dans les UsnRNA ne sont que partiellement identifiées, puisque seules les enzymes responsables de la formation de 3 des 6 résidus sont connues (Pus1p et Pus7p). Les perspectives de travail peuvent être subdivisées en trois volets: (i) la recherche des ARN:Ψ-synthases responsables de la formation des résidus 5 et 6 dans le snRNA U1, 99 dans le snRNA U5 et 72 dans l'ARNt^{Meti} mitochondrial ; (ii) l'étude approfondie de la spécificité d'action des ARN:Ψ-synthases que nous avons caractérisées ; (iii) la poursuite des études structurales sur les ARN:Ψ-synthases Pus7p de *S. cerevisiae* et d'archaea.

Nom de l'enzyme	Substrat(s) et positions modifiées
Pus1p	1, 26, 27, 28, 34, 36, 65, 67 ARNt cytoplasmiques/44 snRNA U2
Pus2p	27 et 28 ARNt mitochondriaux
Pus3p	38 et 39 ARNt cytoplasmiques et mitochondriaux
Pus4p	55 ARNt cytoplasmiques et mitochondriaux
Cbf5p	ARNr (système snoARN guides)
Pus5p (Ylr165p)	2819 ARNr 21S mitochondrial
Pus6p (Ygr169p)	31 ARNt cytoplasmiques et mitochondriaux
Pus7 (Yor243p)	**13 et 35 ARNt cytoplasmiques**/35 snRNA U2/50 5S rRNA
Pus8p (Rib2p)	32 ARNt cytoplasmiques
Pus9p (Ydl036p)	32 ARNt mitochondriaux

Tableau 22 : Les ARN:Ψ-synthases de *S. cerevisiae* actuellement caractérisées.
Le nom et les substrats connus de chaque ARN:Ψ-synthase de *S. cerevisiae* sont indiqués. L'apport en connaissance de notre travail est indiqué en rouge.

1. Recherche de nouvelles ARN:Ψ-synthases

Comme cela vient d'être mentionné, les enzymes responsables de la formation de résidus Ψ à au moins 4 sites dans les ARN de la levure *S. cerevisiae* ne sont pas connues. Par ailleurs, aucune ARN:Ψ-synthase potentielle supplémentaire n'a pu être identifiée par homologie de séquence avec les ARN:Ψ-synthases déjà caractérisées. Comme nous l'avons déjà dit précédemment, deux hypothèses sont envisageables : soit plusieurs enzymes ont des activités redondantes et de ce fait leur délétion ne conduit pas à une perte de l'activité de modification *in vivo*, soit des ARN:Ψ-synthases non homologues aux enzymes déjà identifiées existent chez *S. cerevisiae*. Afin, d'identifier les ARN:Ψ-synthases « manquantes », nous envisageons d'utiliser trois stratégies différentes : (i) la construction de souches de *S. cerevisiae* portant des délétions multiples de gènes ; (ii) des essais d'identification biochimique d'ARN:Ψ-

synthases en utilisant des substrats ARN, produits par transcription *in vitro*, contenant des résidus 5-fluoro-uridine (5FU) et (iii) une approche de criblage de gènes à haut débit similaire à celle utilisée par l'équipe de Yu (Ma et al., 2003).

La construction de souches de *S. cerevisiae* portant des délétions multiples de gènes par la méthode de remplacement de gènes par un marqueur d'auxotrophie, nécessite de disposer de souches mutantes pour de nombreux marqueurs d'auxotrophie. Afin d'éviter ce problème, nous allons utiliser un système avec lequel toutes les délétions sont réalisées avec un seul marqueur (Guldener et al., 1996). Ce système repose sur le remplacement, par recombinaison homologue, du gène à disrupter par une cassette contenant le gène *kan*^{r} encadré de séquences loxP (34pb) (Fig. 103). La souche mutante, ainsi construite, est ensuite transformée par le plasmide pSH47, qui permet l'expression de la recombinase Cre en présence de galactose. Cette recombinase a la particularité de provoquer des événements de recombinaison spécifiques des séquences loxP (Abremski & Hoess, 1984) et permet donc l'excision du gène *kan*^{r} et d'une copie loxP au niveau du locus du gène que l'on a délété. Il ne reste au niveau du locus d'intérêt qu'une copie de la séquence loxP (Guldener et al., 1996). Il est donc possible de déléter un second gène dans cette même souche en utilisant à nouveau le gène *kan*^{r}. L'avantage est aussi que le gène *kan*^{r}, qui comme nous l'avons vu peut perturber l'étude des phénotypes, peut être éliminé. Ce système est très utile pour étudier des familles d'enzymes comportant de nombreux représentants. Nous pourrons ainsi construire des souches de *S. cerevisiae* portant des délétions au niveau de plusieurs gènes codant des ARN:Ψ-synthases et analyser la pseudouridylation des UsnRNA et de l'ARNt^{Meti} mitochondrial dans ces souches. Les souches ainsi générées pourront également être utilisées lors de l'étude de l'effet de l'absence des résidus Ψ dans les ARNt ou dans les UsnRNA sur le fonctionnement des cellules.

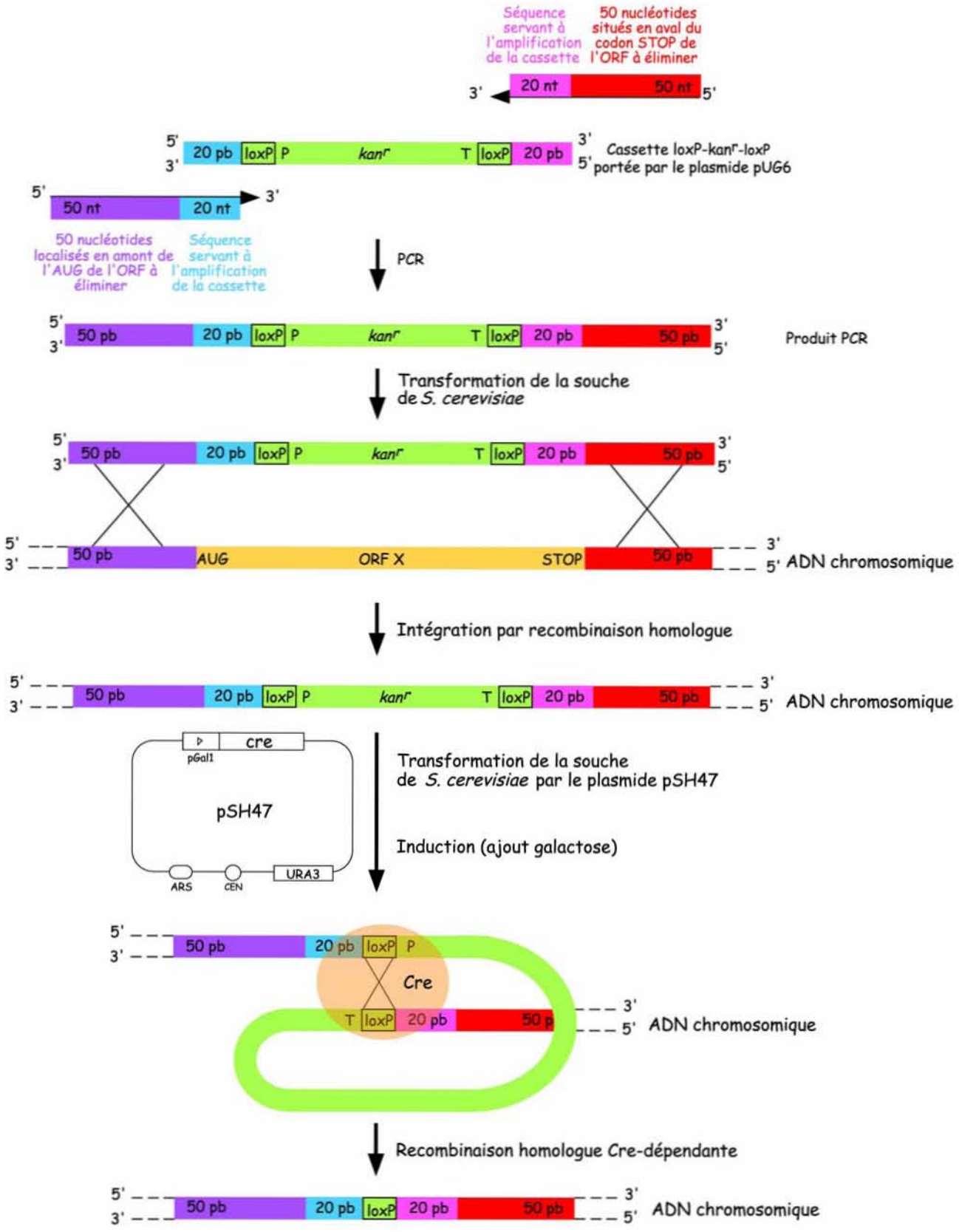


Figure 103 : Principe des délétions des ORF de *S. cerevisiae* réalisées par substitution du gène d'intérêt par la cassette loxP-*kan*^{r}-loxP et excision de cette cassette grâce à la recombinase Cre.
D'après (Guldener et al., 1996).

L'identification de nouvelles ARN:Ψ-synthases par une approche biochimique est également envisagée. Cependant, au lieu d'utiliser le test de Cortese précédemment décrit (Cortese et al., 1974), nous envisageons d'utiliser un ARN substrat contenant des résidus 5FU. En effet, comme cela a été dit précédemment, le résidu 5FU, est un inhibiteur compétitif de la réaction de pseudouridylation (Samuelsson, 1991; Patton et al., 1994; Yu et al., 1998; Gu et al., 1999 ; Zhao et al., 2002), formant un intermédiaire réactionnel stable qui ne peut pas donner lieu à l'isomérisation. Avant d'utiliser cette approche, il faudra d'abord vérifier que l'activité d'intérêt est détectable au sein d'un extrait cellulaire. Il faudra aussi s'assurer que le système de pseudouridylation recherché est bien inhibé par le 5FU. En effet, il a été démontré que, contrairement à ce qui était admis jusqu'à présent, toutes les ARN:Ψ-synthases ne sont pas sujettes à une inhibition par le 5FU (Spedaliere & Mueller, 2004). Cette propriété a déjà été utilisée pour inhiber spécifiquement des ARN:Ψ-synthases dans des extraits cellulaires. L'équipe de J. Patton avait aussi essayé de déterminer la localisation subcellulaire des UsnRNA:Ψ-synthases humaines (pour revue, Patton, 1994). Notre objectif est d'utiliser cette propriété pour isoler des enzymes. Pour cela, des ARN produits par synthèse chimique contenant des résidus 5FU seront utilisés pour essayer d'isoler, au sein d'extraits cellulaires, les systèmes de pseudouridylation agissant sur cet ARN (Fig. 104). Une fois les complexes isolés, les partenaires ARN et/ou protéine de l'ARN substrat devraient pouvoir être identifiés par marquage à l'extrémité 3' et séquençage chimique ou par spectrométrie de masse respectivement (Fig. 104). Cette approche a été utilisée avec succès par l'équipe de Yu pour identifier la machinerie de pseudouridylation agissant sur le snRNA U2 de *Xenopus laevis* (Zhao et al., 2002). Il a ainsi été montré que l'ARN guide pugU2-34/44 est responsable de la formation des résidus Ψ en positions 34 et 44 du snRNA U2 de Xénope (Zhao et al., 2002). Cette méthode présente l'avantage de pouvoir être utilisée aussi bien dans le cas d'une ARN:Ψ-synthase de nature purement protéique, que dans le cas d'une ARN:Ψ-synthase appartenant à un complexe incluant un ARN guide.

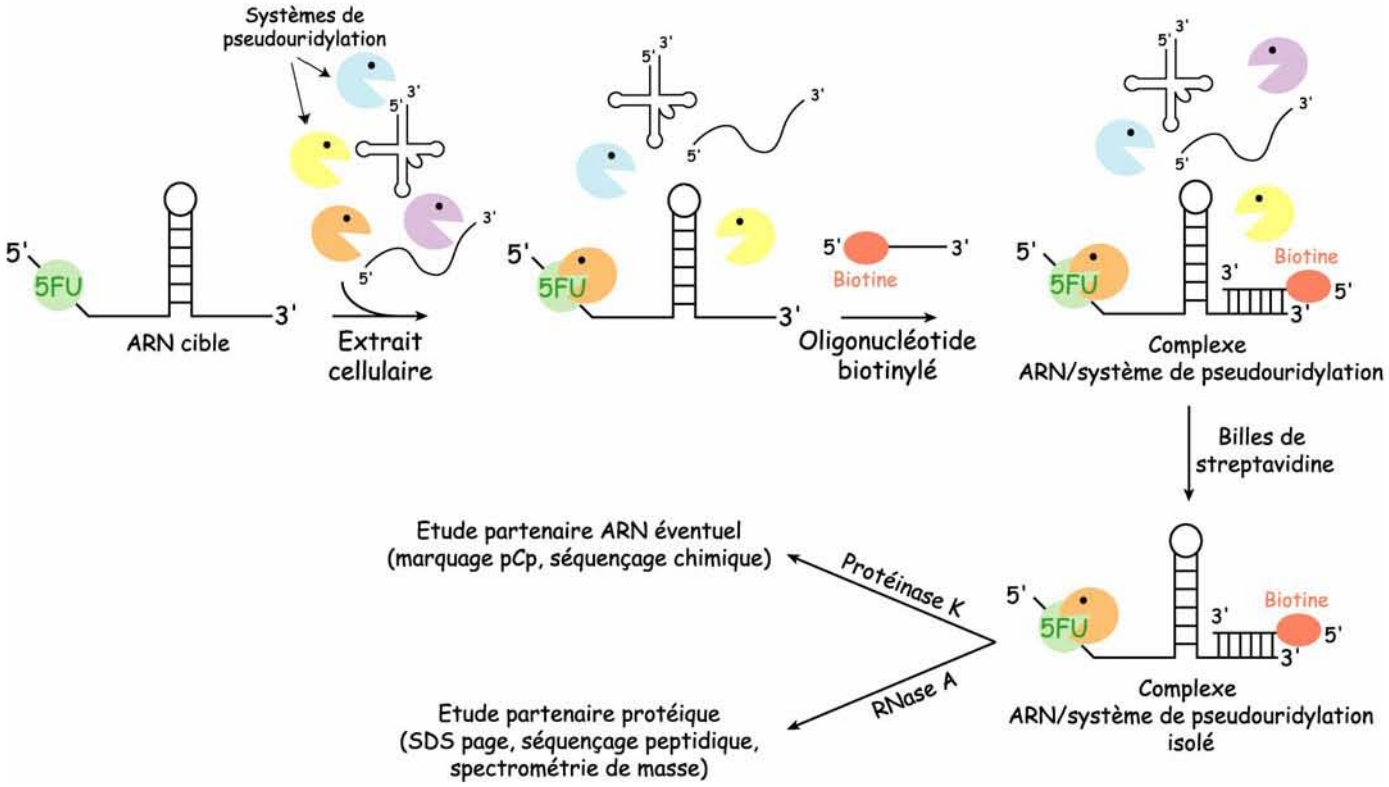


Figure 104 : Stratégie expérimentale envisagée pour isoler les systèmes de pseudouridylation associés à un ARN donné.
L'ARN cible contenant des résidus 5FU est incubé avec un extrait cellulaire. Le système de pseudouridylation agissant sur cet ARN forme un complexe stable avec l'ARN. Ce complexe est ensuite isolé sur billes de streptavidine grâce un oligonucléotide biotinylé complémentaire à l'ARN cible. La composition en ARN et protéines du complexe est ensuite analysée.

Il faut noter que dans la stratégie expérimentale présentée ci-dessus, tous les résidus uridine de l'ARN substrat sont substitués par des résidus 5FU. Dans le cas d'ARN substrats contenant naturellement plusieurs résidus Ψ, cas du snRNA U2 ou des ARNt par exemple, il pourrait être intéressant d'incorporer le résidu 5FU de façon spécifique, afin de pouvoir cibler le système de pseudouridylation recherché. Ainsi, dans le cas du système responsable de la modification de la position 72 des ARNt mitochondriaux, il est envisageable de produire, par transcription *in vitro*, un ARN correspondant aux nucléotides 1 à 71 et de liguer en 3' de cet ARN un oligonucléotide ARN biotinylé en 3' et contenant un résidu 5FU en 5'. Le même type d'approche pourra être utilisé pour la recherche du système de modification responsable de la formation des 2 résidus Ψ de la région 5'-terminale du snRNA U1. Dans ce cas, le transcrit pourra être biotinylé en 5'. Par ailleurs, il faut tenir compte, dans ce type

d'expérience, que la présence d'un résidu 5FU peut inhiber la formation d'autres résidus modifiés que les résidus Ψ, tels les résidus m5U (T) et 5,6 dihydrouridine (D).

La dernière stratégie que nous pouvons envisager pour identifier les ARN:Ψ-synthases « manquantes » est une stratégie de type « criblage de gènes à haut débit ». Nous avons vu précédemment que cette méthodologie a été utilisée avec succès par l'équipe de Yu et a permis l'identification de l'ARN:Ψ-synthase Pus7p (Ma et al., 2003). L'utilisation de ce type d'approche nécessite de disposer d'un test simple, rapide et fiable de l'activité des différents sous-ensembles de protéines, d'autant plus que l'activité détectée ne conduit pas forcément à une modification totale de l'ARN étudié. Pour cela, le résidu U cible de l'ARN substrat doit être le seul résidu marqué au sein de l'ARN substrat afin de tester son isomérisation en résidu Ψ par chromatographie monodimensionnelle (Yu et al., 1998; Yu, 2000; Ma et al., 2003).

L'activité du laboratoire étant fortement centrée sur les problèmes d'épissage, nous nous efforcerons de poursuivre la recherche des enzymes agissant sur les UsnRNA, ceci toujours en collaboration avec l'équipe de Yu. Toujours dans le cadre d'une meilleure compréhension des mécanismes assurant la modification des UsnRNA de vertébrés, nous envisageons d'étendre l'étude à ces organismes. En effet, actuellement des ARN guides (scaRNA) n'ont pas été trouvés pour toutes les modifications identifiées dans les UsnRNA de vertébrés. Il est donc possible que certaines d'entre elles soient générées par des enzymes ne nécessitant pas d'ARN guides. Par ailleurs, comme nous l'avons vu précédemment, il est possible que certaines modifications soient à la fois catalysées par un système d'ARN guide et une enzyme spécifique de nature protéique.

D'une manière générale, la grande majorité des enzymes de modification caractérisées à ce jour sont des enzymes de procaryotes ou d'eucaryotes inférieurs comme la levure *S. cerevisiae*. Les enzymes responsables des modifications des ARN de vertébrés et de mammifères (y compris de l'homme) restent encore très peu étudiées. L'avancée importante dans la caractérisation des ARN:Ψ-synthases de levure permettra d'aborder plus facilement l'étude des enzymes humaines. Les candidats potentiels pourront être recherchés par homologie de séquences. La vérification expérimentale de leurs fonctions pourra ensuite être réalisée *in vitro* et *in vivo*. Les ARN humains contiennent en général plus de modifications post-transcriptionnelles que ceux de la levure *S. cerevisiae* (pour revues, Ofengand, 2002; Massenet et al., 1998; Sprinzl et al., 1998). Bien que ces modifications puissent dans de nombreux cas être expliquées par la présence du système d'ARN guide, il est possible que de nouvelles enzymes ou des enzymes avec une spécificité de reconnaissance différente soient présentes chez l'homme.

2. Etude de la spécificité d'action des ARN:Ψ-synthases

Bien que toutes les enzymes requises pour la formation des résidus Ψ dans les ARNt aient été caractérisées, peu d'informations sont disponibles concernant leur spécificité de reconnaissance. En particulier, il serait intéressant de comparer les déterminants de spécificité des ARN:Ψ-synthases de *S. cerevisiae* qui modifient soit la même position de l'ARNt dans deux compartiments cellulaires différents (Pus8p et Pus9p), soit deux positions juxtaposées (31 et 32) dans les ARNt (Pus6p et Pus8p) (Fig. 105). Ce travail sera complémentaire de celui que nous avons réalisé sur l'enzyme Pus7p et pourrait apporter des informations concernant l'incapacité de Pus7p à agir sur les ARNt mitochondriaux *in vivo*.

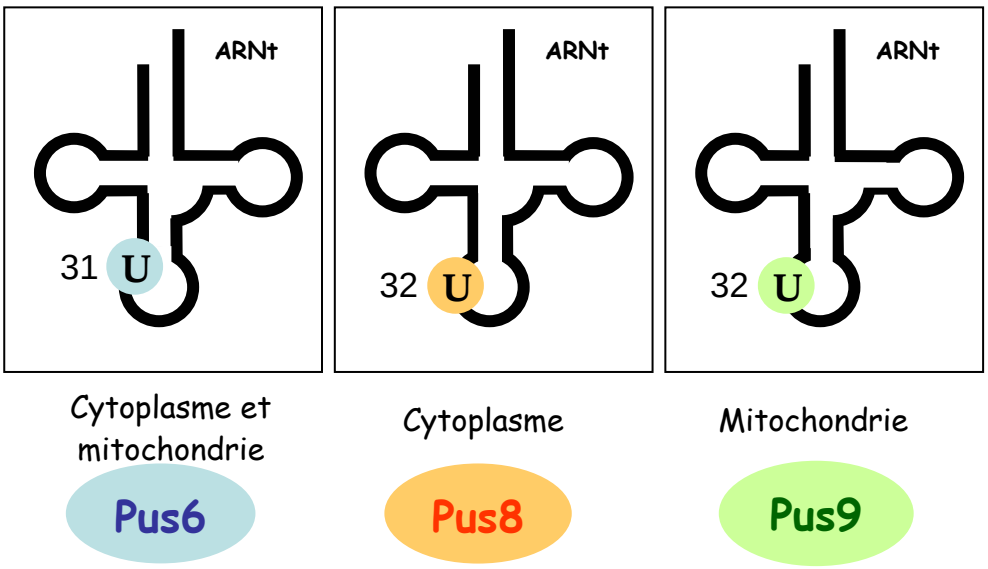


Figure 105 : Les activités ARN:Ψ-synthases des enzymes Pus6p, Pus8p et Pus9p de la levure *S. cerevisiae*.
La position modifiée, ainsi que le compartiment cellulaire où a lieu la modification sont indiqués.

Les éléments de reconnaissance entre ARN substrat et enzyme de modification sont de nature très diverse, depuis la séquence nucléotidique jusqu'à la structure 3D, en passant par les éléments de structure 2D. La distinction entre ces différentes possibilités peut être faite par une étude de l'activité sur différents mutants des substrats, telle que celle que nous avons réalisée avec Pus7p. Par ailleurs, pour les enzymes qui ne s'avéreront pas être dépendantes de la structure 3D du substrat, l'équipe va essayer de développer une méthode de type SELEX permettant de définir les requis en structure 2D et en séquence. Elle sera basée sur l'utilisation d'anticorps reconnaissant spécifiquement les résidus Ψ (Fig. 106). Un ARN substrat portant une séquence aléatoire sera modifié par l'enzyme. Les ARN renfermant un ou plusieurs

résidus modifiés seront sélectionnés par chromatographie d'affinité sur des billes de protéine G-Sepharose, où les anticorps spécifiques des résidus modifiés auront été immobilisés. Les ARN retenus seront rétrotranscrits et l'ADN obtenu sera amplifié par PCR, pour servir de matrice pour la production d'un second lot d'ARN soumis au même traitement. Après quelques cycles, les meilleurs substrats devraient être sélectionnés par cette approche et pourront être ensuite caractérisés. Les données obtenues devraient permettre des études approfondies sur les mécanismes d'action des ARN:Ψ-synthases.

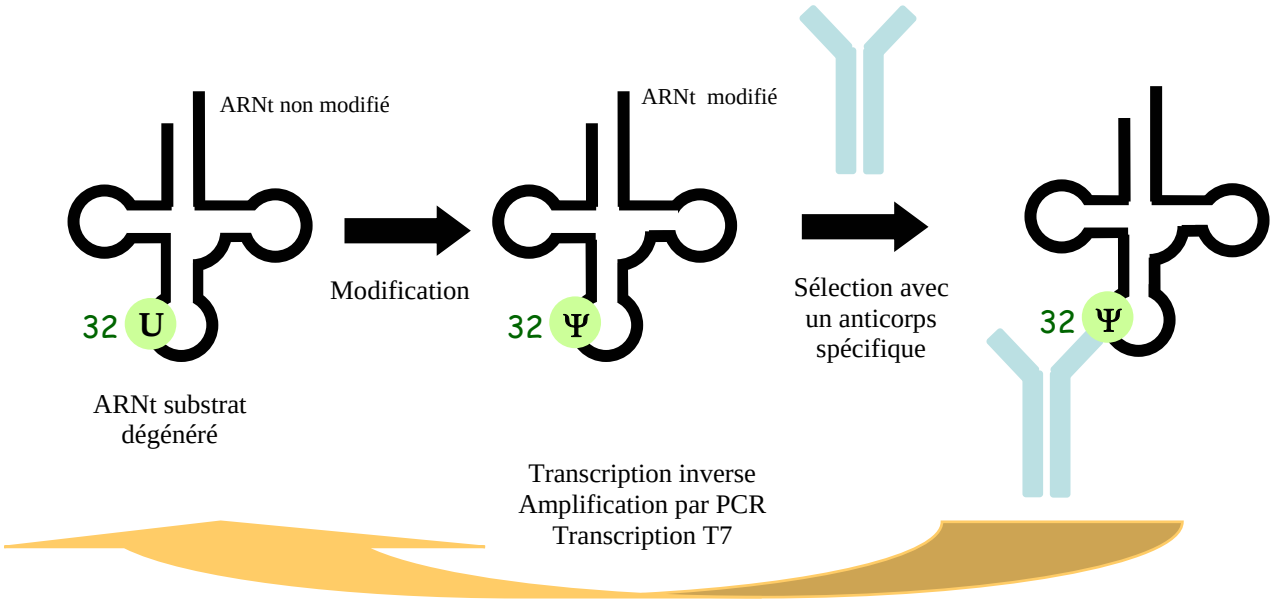


Figure 106 : Principe de la méthode SELEX que nous envisageons d'utiliser pour l'étude de la spécificité des ARN:Ψ-synthases Pus6p, Pus8p et Pus9p.

Des anticorps dirigés contre les résidus Ψ ont déjà été produits par une équipe japonaise (Itoh et al., 1992), cette approche devrait donc être réalisable. Notre équipe a, par ailleurs, déjà utilisé largement l'anticorps anti-$m^{2,2,7}G$, produit par l'équipe de R. Lührmann, pour la purification des UsnRNA et des complexes ARN-protéine qu'ils forment (Luhrmann et al., 1982; Bochnig et al., 1987; Watkins et al., 2000). Par ailleurs des anticorps spécifiquement dirigés contre les nucléotides et nucléosides modifiés m^1G, m^2G, m^7G, m^5C, m^1A, m^6A, yW ont déjà été produits (Munns & Liszewski, 1980; Vold, 1981) et se sont avérés permettre la distinction entre nucléotide modifié et non modifié (Milstone et al., 1978; Vold, 1981; Rainen & Stollar, 1978 ; Bochnig et al., 1987).

3. Poursuite des études structurales sur les ARN:Ψ-synthases Pus7p de *S. cerevisiae* et d'archaea

Comme nous l'avons mentionné dans l'introduction, l'étude de la structure 3D des ARN:Ψ-synthases a permis de mettre en évidence un certain nombre de caractéristiques communes entre ces enzymes comme la conservation de la structure du core-enzyme et la présence, pour certaines enzymes de la famille, de domaines additionnels impliqués dans la reconnaissance de l'ARN. Seule l'ARN:Ψ-synthase TruA a été cristallisée avec un substrat entier (Hur & Stroud, 2007). Les enzymes de la famille TruD que nous avons étudiées au cours de ce travail sont très différentes en séquence des autres enzymes ARN:Ψ-synthases et possèdent un domaine de liaison à l'ARN (domaine TruD) absent dans les enzymes de la famille TruA.

Dans un premier temps, il serait intéressant de déterminer la structure du complexe entre l'ARN:Ψ-synthase aPus7 de *P. abyssi* et un ARN substrat. L'ARNt^{Asp}(GUC) que nous avons utilisé lors de la caractérisation de l'enzyme Pab aPus7 (publication n°3) apparaît être un candidat de choix pour cette étude d'autant qu'il pourrait être également utilisé pour les essais de co-cristallisation avec d'autres protéines comme l'enzyme aPus7 de *S. solfataricus* ou Pus7p de *S. cerevisiae*. Réussir à cristalliser une de ces protéines avec un ARNt substrat permettrait de mieux comprendre le rôle des domaines additionnels qui sont présents chez les eucaryotes et les archaea mais pas chez les bactéries et apporterait des informations structurales concernant les bases moléculaires de la reconnaissance du substrat par les enzymes de cette famille. Pour cela, il sera nécessaire de disposer d'une quantité importante d'ARN qui soit homogène en taille. Une piste possible sera le clonage de l'ARN cible dans un plasmide permettant de produire un ARN possédant des séquences ribozymes à ses extrémités 5' et 3' terminales. Un travail de biologie moléculaire sera cependant obligatoire avant la cristallogenèse. Il s'agira de muter l'acide aminé catalytique (D) par mutagenèse dirigée afin d'inactiver ces enzymes et de permettre la formation de complexes stables entre ces enzymes et l'ARNt cible. Cette mutation ne devrait pas avoir de conséquence sur la solubilité de la protéine puisqu'elle a déjà été réalisée sans problème sur plusieurs autres enzymes de la famille.

Une autre possibilité sera d'utiliser un ARN raccourci produit par synthèse chimique dans lequel l'uridine cible pourra être remplacée par un résidu 5FU permettant la formation d'un complexe enzyme-substrat covalent. Il faudra préalablement vérifier que chaque enzyme est capable de modifier efficacement ce substrat raccourci. Puisque nous avons démontré qu'un substrat raccourci de l'ARNt^{Asp}(GUC) de levure correspondant à la tige boucle D de et ARN est un substrat pour Pus7p de *S. cerevisiae*, il serait intéressant de réaliser les expériences de co-cristallisation avec un substrat correspondant à la tige boucle D de

l'ARNt^{Asp}(GUC) de *P. abyssi* et ceci après avoir vérifié par des expériences de retard sur gel, la fixation efficace de celui-ci sur toutes les protéines.

Nous avons déjà obtenu des cristaux des enzymes His_6-Pus7Δ5 et aPus7 de *S. solfataricus* mais leur faible qualité de diffraction n'a pas permis de résoudre la structure de ces protéines. Nous envisageons de produire et de purifier ces protéines selon le protocole habituel en ajoutant une étape de clivage de l'étiquette poly-histidine présente à l'extrêmité N-terminale ce qui pourrait permettre l'amélioration de la qualité des cristaux.

Références Bibliographiques

A

Abremski K, Hoess R. 1984. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *The Journal of biological chemistry 259*:1509-1514.

Agris PF. 1996. The importance of being modified: roles of modified nucleosides and Mg2+ in RNA structure and function. *Prog Nucleic Acid Res Mol Biol 53*:79-129.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research 25*:3389-3402.

Ansmant I, Massenet S, Grosjean H, Motorin Y, Branlant C. 2000. Identification of the *Saccharomyces cerevisiae* RNA:pseudouridine synthase responsible for formation of Ψ_{2819} in 21S mitochondrial ribosomal RNA. *NAR 28*:1941-1946.

Ansmant I, Motorin Y, Massenet S, Grosjean H, Branlant C. 2001. Identification and characterization of tRNA:Ψ_{31}-synthase (Pus6p) in yeast *S. cerevisiae*. *JBC 276*:34934-34940.

Arena F, Ciliberto G, Ciampi S, Cortese R. 1978. Purification of pseudouridylate synthetase I from Salmonella typhimurium. *Nucleic acids research 5*:4523-4536.

Ares M, Jr., Igel AH. 1990. Lethal and temperature-sensitive mutations and their suppressors identify an essential structural element in U2 small nuclear RNA. *Genes & development 4*:2132-2145.

Arluison V, Buckle M, Grosjean H. 1999. Pseudouridine synthetase Pus1 of Saccharomyces cerevisiae: kinetic characterisation, tRNA structural requirement and real-time analysis of its complex with tRNA. *Journal of molecular biology 289*:491-502.

Arnez JG, Steitz TA. 1994. Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl- tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry 33*:7560-7567.

Ast G, Weiner AM. 1997. A novel U1/U5 interaction indicates proximity between U1 and U5 snRNAs during an early step of mRNA splicing. *RNA (New York, NY 3*:371-381.

Auffinger P, Louise-May S, Westhof E. 1996. Hydration of C-H groups in tRNA. *Faraday Discuss 103*:151-173.

Auffinger P, Westhof E. 1996. H-bond stability in the tRNA(Asp) anticodon hairpin: 3 ns of multiple molecular dynamics simulations. *Biophys J 71*:940-954.

Auffinger P, Westhof E. 1997. RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *Journal of molecular biology 269*:326-341.

Auffinger P, Westhof E. 1998a. Effects on pseudouridylation on tRNA hydration and dynamics: a theoretical approach. In: Grosjean H, Benne R, eds. *The modification and editing of RNA*. Washington, DC: ASM Press.

Auffinger P, Westhof E. 1998b. Hydration of RNA base pairs. *J Biomol Struct Dyn 16*:693-707.

Auffinger P, Westhof E. 1999. Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes. *Journal of molecular biology 292*:467-483.

Auffinger P, Westhof E. 2001. An extended structural signature for the tRNA anticodon loop. *RNA (New York, NY 7*:334-341.

Auxilien S, Crain PF, Trewyn RW, Grosjean H. 1996. Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. *Journal of molecular biology 262*:437-458.

B

Bachellerie JP, Cavaille J. 1997. Guiding ribose methylation of rRNA. *Trends Biochem Sci 22*:257-261.

Bachellerie JP, Cavaille J. 1998. Small nucleoloar RNAs guide the ribose methylations of eukaryotic rRNAs. In: Grosjean H, Benne R, eds. *Modification and Editing of RNA*. Washington DC: American Society for Microbiology. pp 255-272.

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

Bagni C, Lapeyre B. 1998. Gar1p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element. *The Journal of biological chemistry 273*:10868-10873.

Bajji AC, Sundaram M, Myszka DG, Davis DR. 2002. An RNA complex of the HIV-1 A-loop and tRNA(Lys,3) is stabilized by nucleoside modifications. *J Am Chem Soc 124*:14302-14303.

Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, Terns MP. 2005. RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes & development 19*:1238-1248.

Bakin A, Ofengand J. 1993. Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry 32*:9754-9762.

Balakin AG, Smith L, Fournier MJ. 1996. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell 86*:823-834.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science (New York, NY 289*:905-920.

Bare LA, Uhlenbeck OC. 1986. Specific substitution into the anticodon loop of yeast tyrosine transfer RNA. *Biochemistry 25*:5825-5830.

Baudin A, Ozier-Kalogeropoulos O, Denouel A, Lacroute F, Cullin C. 1993. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic acids research 21*:3329-3330.

Becker HF, Motorin Y, Planta RJ, Grosjean H. 1997a. The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs. *Nucleic acids research 25*:4493-4499.

Becker HF, Motorin Y, Sissler M, Florentz C, Grosjean H. 1997b. Major identity determinants for enzymatic formation of ribothymidine and pseudouridine in the T psi-loop of yeast tRNAs. *Journal of molecular biology 274*:505-518.

Behm-Ansmant I, Urban A, Ma X, Yu YT, Motorin Y, Branlant C. 2003. The Saccharomyces cerevisiae U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. *RNA (New York, NY 9*:1371-1382.

Benas P, Bec G, Keith G, Marquet R, Ehresmann C, Ehresmann B, Dumas P. 2000. The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop. *RNA (New York, NY 6*:1347-1355.

Björk GR. 1995. Biosynthesis and function of modified nucleosides. In: Söll D, RajBhandary UL, eds. *tRNA: structure, biosynthesis and function*. Washington, DC: ASM press. pp 165-205.

Bjork GR, Ericson JU, Gustafsson CE, Hagervall TG, Jonsson YH, Wikstrom PM. 1987. Transfer RNA modification. *Annu Rev Biochem 56*:263-287.

Bochnig P, Reuter R, Bringmann P, Luhrmann R. 1987. A monoclonal antibody against 2,2,7-trimethylguanosine that reacts with intact, class U, small nuclear ribonucleoproteins as well as with 7-methylguanosine-capped RNAs. *Eur J Biochem 168*:461-467.

Bonnerot C, Pintard L, Lutfalla G. 2003. Functional redundancy of Spb1p and a snR52-dependent mechanism for the 2'-O-ribose methylation of a conserved rRNA position in yeast. *Molecular cell 12*:1309-1315.

Bortolin ML, Ganot P, Kiss T. 1999. Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *The EMBO journal 18*:457-469.

Brand RC, Klootwijk J, Sibum CP, Planta RJ. 1979. Pseudouridylation of yeast ribosomal precursor RNA. *Nucleic acids research 7*:121-134.

Branlant C, Krol A, Ebel JP, Gallinaro H, Lazar E, Jacob M. 1981b. The conformation of chicken, rat and human U1A RNAs in solution. *Nucleic acids research 9*:841-858.

Branlant C, Krol A, Ebel JP, Lazar E, Gallinaro H, Jacob M, Sri-Widada J, Jeanteur P. 1980. Nucleotide sequences of nuclear U1A RNAs from chicken, rat and man. *Nucleic acids research 8*:4143-4154.

Branlant C, Krol A, Ebel JP, Lazar E, Haendler B, Jacob M. 1982. U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *The EMBO journal 1*:1259-1265.

Branlant C, Krol A, Lazar E, Haendler B, Jacob M, Galego-Dias L, Pousada C. 1983. High evolutionary conservation of the secondary structure and of certain nucleotide sequences of U5 RNA. *Nucleic acids research 11*:8359-8367.

Branlant C, Krol A, Machatt MA, Pouyet J, Ebel JP, Edwards K, Kossel H. 1981a. 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 research 9*:4303-4324.

Brock TD, Brock KM, Belly RT, Weiss RL. 1972. Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Archiv fur Mikrobiologie 84*:54-68.

Brule H, Holmes WM, Keith G, Giege R, Florentz C. 1998. Effect of a mutation in the anticodon of human mitochondrial tRNAPro on its post-transcriptional modification pattern. *Nucleic acids research 26*:537-543.

Bykhovskaya Y, Casas K, Mengesha E, Inbal A, Fischel-Ghodsian N. 2004. Missense mutation in pseudouridine synthase 1 (PUS1) causes mitochondrial myopathy and sideroblastic anemia (MLASA). *American journal of human genetics 74*:1303-1308.

C

Cabello-Villegas J, Nikonowicz EP. 2005. Solution structure of psi32-modified anticodon stem-loop of Escherichia coli tRNAPhe. *Nucleic acids research 33*:6961-6971.

Chanfreau G, Legrain P, Jacquier A. 1998. Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *Journal of molecular biology 284*:975-988.

Charette M, Gray MW. 2000. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life 49*:341-351.

Charpentier B, Fourmann JB, Branlant C. 2007. Reconstitution of archaeal H/ACA sRNPs and test of their activity. *Methods in enzymology 425*:389-405.

Charpentier B, Muller S, Branlant C. 2005. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic acids research 33*:3133-3144.

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

Chaudhuri BN, Chan S, Perry LJ, Yeates TO. 2004. Crystal structure of the apo forms of psi 55 tRNA pseudouridine synthase from Mycobacterium tuberculosis: a hinge at the base of the catalytic cleft. *The Journal of biological chemistry 279*:24585-24591.

Chen Y, Sierzputowska-Gracz H, Guenther R, Everett K, Agris PF. 1993. 5-Methylcytidine is required for cooperative binding of Mg2+ and a conformational transition at the anticodon stem-loop of yeast phenylalanine tRNA. *Biochemistry 32*:10249-10253.

Cheng X, Roberts RJ. 2001. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic acids research 29*:3784-3795.

Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T, Schroeder M, Weng S, Botstein D. 1998. SGD: Saccharomyces Genome Database. *Nucleic acids research 26*:73-79.

Choffat Y, Suter B, Behra R, Kubli E. 1988. Pseudouridine modification in the tRNA(Tyr) anticodon is dependent on the presence, but independent of the size and sequence, of the intron in eucaryotic tRNA(Tyr) genes. *Molecular and cellular biology 8*:3332-3337.

Ciampi MS, Arena F, Cortese R. 1977. Biosynthesis of pseudouridine in the in vitro transcribed tRNATyr precursor. *FEBS Lett 77*:75-82.

Claesson C, Lustig F, Boren T, Simonsson C, Barciszewska M, Lagerkvist U. 1995. Glycine codon discrimination and the nucleotide in position 32 of the anticodon loop. *J Mol Biol 247*:191-196.

Cohn WE. 1960. Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *J Biol Chem 235*:1488-1498.

Cole H, Rauschkolb J, Toomey J. 1930. Dyskeratosis congenita with pigmentation, dystrophia unguis and leukokeratosis oris. *Arch Dermatol Syphil 21*:71-95.

Comarmond M, Giégé R, Thierry JC, Moras D. 1986. Three-dimensional structure of yeast tRNAAsp. I. Structure determination. *Acta Cryst B 42*:272-280.

Conrad J, Niu L, Rudd K, Lane BG, 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 (New York, NY 5*:751-763.

Conrad J, Sun D, Englund N, Ofengand J. 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. *The Journal of biological chemistry 273*:18562-18566.

Cortes JJ, Sontheimer EJ, Seiwert SD, Steitz JA. 1993. Mutations in the conserved loop of human U5 snRNA generate use of novel cryptic 5' splice sites *in vivo*. *The EMBO journal 12*:5181-5189.

Cortese R, Kammen HO, Spengler SJ, Ames BN. 1974. Biosynthesis of pseudouridine in transfer ribonucleic acid. *J Biol Chem 249*:1103-1108.

Crick FH. 1966. The genetic code. 3. *Sci Am 215*:55-60 passim.

Curnow AW, Kung FL, Koch KA, Garcia GA. 1993. tRNA-guanine transglycosylase from Escherichia coli: gross tRNA structural requirements for recognition. *Biochemistry 32*:5239-5246.

Curran J. 1998. Modified nucleosides in translation. In: Grosjean H, Benne R, eds. *The modification and editing of RNA.* Washington, DC: ASM Press. pp 493-516.

D

Darzacq X, Jady BE, Verheggen C, Kiss AM, Bertrand E, Kiss T. 2002. Cajal body-specific small nuclear RNAs: a novel class of 2'-O- methylation and pseudouridylation guide RNAs. *The EMBO journal 21*:2746-2756.

Darzacq X, Kittur N, Roy S, Shav-Tal Y, Singer RH, Meier UT. 2006. Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *The Journal of cell biology 173*:207-218.

Datta B, Weiner AM. 1991. Genetic evidence for base pairing between U2 and U6 snRNA in mammalian mRNA splicing. *Nature 352*:821-824.

Davis DR, Veltri CA, Nielsen L. 1998. An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNALys, tRNAHis and tRNATyr. *J Biomol Struct Dyn 15*:1121-1132.

Davis FF, Allen FW. 1957. Ribonucleic acids from yeast which contain a fifth nucleotide. *J Biol Chem 227*:907-915.

Decatur WA, Fournier MJ. 2002. rRNA modifications and ribosome function. *Trends Biochem Sci 27*:344-351.

Decatur WA, Schnare MN. 2008. Different Mechanisms for Pseudouridine Formation in Yeast 5S and 5.8S Ribosomal RNAs. *Molecular and cellular biology*.

Del Campo M, Kaya Y, Ofengand J. 2001. Identification and site of action of the remaining four putative pseudouridine synthases in Escherichia coli. *RNA (New York, NY 7*:1603-1615.

Del Campo M, Ofengand J, Malhotra A. 2004. Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of Escherichia coli. *RNA (New York, NY 10*:231-239.

Derrick WB, Horowitz J. 1993. Probing structural differences between native and in vitro transcribed Escherichia coli valine transfer RNA: evidence for stable base modification-dependent conformers. *Nucleic acids research 21*:4948-4953.

Dez C, Noaillac-Depeyre J, Caizergues-Ferrer M, Henry Y. 2002. Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs. *Molecular and cellular biology 22*:7053-7065.

Dirheimer G. 1983. Chemical nature, properties, location, and physiological and pathological variations of modified nucleosides in tRNAs. *Recent Results Cancer Res 84*:15-46.

Dokal I. 1996. Dyskeratosis congenita: an inherited bone marrow failure syndrome. *Br J Haematol 92*:775-779.

Dokudovskaya S, Dontsova O, Shpanchenko O, Bogdanov A, Brimacombe R. 1996. Loop IV of 5S ribosomal RNA has contacts both to domain II and to domain V of the 23S RNA. *RNA (New York, NY 2*:146-152.

Droogmans L, Haumont E, de Henau S, Grosjean H. 1986. Enzymatic 2'-O-methylation of the wobble nucleoside of eukaryotic tRNAPhe: specificity depends on structural elements outside the anticodon loop. *The EMBO journal 5*:1105-1109.

Durant PC, Davis DR. 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*:56-57.

Durant PC, Davis DR. 1999. Stabilization of the anticodon stem-loop of tRNALys,3 by an A+-C base-pair and by pseudouridine. *Journal of molecular biology 285*:115-131.

E

Edqvist J, Blomqvist K, Straby KB. 1994. Structural elements in yeast tRNAs required for homologous modification of guanosine-26 into dimethylguanosine-26 by the yeast Trm1 tRNA- modifying enzyme. *Biochemistry 33*:9546-9551.

Edqvist J, Grosjean H, Straby KB. 1992. Identity elements for N2-dimethylation of guanosine-26 in yeast tRNAs. *Nucleic acids research 20*:6575-6581.

Edqvist J, Straby KB, Grosjean H. 1993. Pleiotrophic effects of point mutations in yeast tRNA(Asp) on the base modification pattern. *Nucleic acids research 21*:413-417.

Engmann M. 1926. A unique case of reticular pigmentation of the skin with atrophy. *Arch Dermatol Syphil 13*:685-687.

Eshete M, Marchbank MT, Deutscher SL, Sproat B, Leszczynska G, Malkiewicz A, Agris PF. 2007. Specificity of phage display selected peptides for modified anticodon stem and loop domains of tRNA. *Protein J 26*:61-73.

F

Fatica A, Dlakic M, Tollervey D. 2002. Naf1 p is a box H/ACA snoRNP assembly factor. *RNA (New York, NY 8*:1502-1514.

Filipowicz W, Pogacic V. 2002. Biogenesis of small nucleolar ribonucleoproteins. *Curr Opin Cell Biol 14*:319-327.

G

Ganot P, Bortolin ML, Kiss T. 1997b. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell 89*:799-809.

Ganot P, Caizergues-Ferrer M, Kiss T. 1997a. 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 & development 11*:941-956.

Garcia GA, Goodenough-Lashua DM. 1998. Mechanisms of RNA-Modifying and -Editing Enzymes. In: Grosjean H, Benne R, eds. *The modification and editing of RNA*. Washington, DC: ASM Press. pp 135-168.

Girard JP, Bagni C, Caizergues-Ferrer M, Amalric F, Lapeyre B. 1994. Identification of a segment of the small nucleolar ribonucleoprotein- associated protein GAR1 that is sufficient for nucleolar accumulation. *The Journal of biological chemistry 269*:18499-18506.

Girard JP, Lehtonen H, Caizergues-Ferrer M, Amalric F, Tollervey D, Lapeyre B. 1992. GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *The EMBO journal 11*:673-682.

Green CJ, Kammen HO, Penhoet EE. 1982. Purification and properties of a mammalian tRNA pseudouridine synthase. *J Biol Chem 257*:3045-3052.

Green R, Noller HF. 1996. In vitro complementation analysis localizes 23S rRNA posttranscriptional modifications that are required for Escherichia coli 50S ribosomal subunit assembly and function. *RNA (New York, NY 2*:1011-1021.

Grosjean H, Edqvist J, Straby KB, Giege R. 1996. Enzymatic formation of modified nucleosides in tRNA: dependence on tRNA architecture. *Journal of molecular biology 255*:67-85.

Grosjean H, Sprinzl M, Steinberg S. 1995. Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. *Biochimie 77*:139-141.

Grosjean H, Szweykowska-Kulinska Z, Motorin Y, Fasiolo F, Simos G. 1997. Intron-dependent enzymatic formation of modified nucleosides in eukaryotic tRNAs: a review. *Biochimie 79*:293-302.

Gu J, Patton JR, Shimba S, Reddy R. 1996. Localization of modified nucleotides in *Schizosaccharomyces pombe* spliceosomal small nuclear RNAs: modified nucleotides are clustered in functionally important regions [published erratum appears in RNA 1996 Oct;2(10):1077]. *RNA (New York, NY 2*:909-918.

Gu X, Liu Y, Santi DV. 1999. The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA. *Proceedings of the National Academy of Sciences of the United States of America 96*:14270-14275.

Gu X, Yu M, Ivanetich KM, Santi DV. 1998. Molecular recognition of tRNA by tRNA pseudouridine 55 synthase. *Biochemistry 37*:339-343.

Guldener U, Heck S, Fielder T, Beinhauer J, Hegemann JH. 1996. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic acids research 24*:2519-2524.

Gutgsell N, Englund N, Niu L, Kaya Y, Lane BG, Ofengand J. 2000. Deletion of the Escherichia coli pseudouridine synthase gene truB blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells. *RNA (New York, NY 6*:1870-1881.

Gutgsell NS, Del Campo MD, Raychaudhuri S, Ofengand J. 2001. A second function for pseudouridine synthases: A point mutant of RluD unable to form pseudouridines 1911, 1915, and 1917 in Escherichia coli 23S ribosomal RNA restores normal growth to an RluD-minus strain. *RNA (New York, NY 7*:990-998.

H

Hall SL, Padgett RA. 1994. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J Mol Biol 239*:357-365.

Hamilton CS, Greco TM, Vizthum CA, Ginter JM, Johnston MV, Mueller EG. 2006. Mechanistic investigations of the pseudouridine synthase RluA using RNA containing 5-fluorouridine. *Biochemistry 45*:12029-12038.

Hamma T, Ferre-D'Amare AR. 2004. Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution. *Structure 12*:893-903.

Hamma T, Ferre-D'Amare AR. 2006. Pseudouridine synthases. *Chemistry & biology 13*:1125-1135.

Hamma T, Reichow SL, Varani G, Ferre-D'Amare AR. 2005. The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nature structural & molecular biology 12*:1101-1107.

Harada F, Kato N, Nishimura S. 1980. The nucleotide sequence of nuclear 4.8S RNA of mouse cells. *Biochem Biophys Res Commun 95*:1332-1340.

Hausner TP, Giglio LM, Weiner AM. 1990. Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. *Genes & development 4*:2146-2156.

Henras A, Dez C, Noaillac-Depeyre J, Henry Y, Caizergues-Ferrer M. 2001. Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. *Nucleic acids research 29*:2733-2746.

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

Hoang C, Chen J, Vizthum CA, Kandel JM, Hamilton CS, Mueller EG, Ferre-D'Amare AR. 2006. Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure. *Molecular cell 24*:535-545.

Hoang C, Ferre-D'Amare AR. 2001. Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell 107*:929-939.

Hoang C, Hamilton CS, Mueller EG, Ferre-D'Amare AR. 2005. Precursor complex structure of pseudouridine synthase TruB suggests coupling of active site perturbations to an RNA-sequestering peripheral protein domain. *Protein Sci 14*:2201-2206.

Hodges PE, Payne WE, Garrels JI. 1998. The Yeast Protein Database (YPD): a curated proteome database for Saccharomyces cerevisiae. *Nucleic acids research 26*:68-72.

Holm L, Sander C. 1993. Protein structure comparison by alignment of distance matrices. *Journal of molecular biology 233*:123-138.

Holmes WM, Andraos-Selim C, Roberts I, Wahab SZ. 1992. Structural requirements for tRNA methylation. Action of Escherichia coli tRNA(guanosine-1)methyltransferase on tRNA(1Leu) structural variants. *The Journal of biological chemistry 267*:13440-13445.

Huang L, Ku J, Pookanjanatavip M, Gu X, Wang D, Greene PJ, Santi DV. 1998b. Identification of two *Escherichia coli* pseudouridine synthases that show multisite specificity for 23S RNA. *Biochemistry 37*:15951-15957.

Huang L, Pookanjanatavip M, Gu X, Santi DV. 1998a. A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry 37*:344-351.

Hur S, Stroud RM. 2007. How U38, 39, and 40 of many tRNAs become the targets for pseudouridylation by TruA. *Molecular cell 26*:189-203.

Huttenhofer A, Kiefmann M, Meier-Ewert S, O'Brien J, Lehrach H, Bachellerie JP, Brosius J. 2001. RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *The EMBO journal 20*:2943-2953.

I

Ikemura T. 1973. [Tow-dimensional polyacrylamide gel electrophoresis of RNA species (author's transl)]. *Tanpakushitsu Kakusan Koso 18*:1204-1210.

Imai Y, Matsushima Y, Sugimura T, Terada M. 1991. A simple and rapid method for generating a deletion by PCR. Nucleic acids research 19:2785.

Incorvaia R, Padgett RA. 1998. Base pairing with U6atac snRNA is required for 5' splice site activation of U12-dependent introns in vivo. *RNA (New York, NY 4*:709-718.

Isel C, Marquet R, Keith G, Ehresmann C, Ehresmann B. 1993. Modified nucleotides of tRNA(3Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription. *The Journal of biological chemistry 268*:25269-25272.

Itoh K, Ishiwata S, Ishida N, Mizugaki M. 1992. Diagnostic use of anti-modified nucleoside monoclonal antibody. *Tohoku J Exp Med 168*:329-331.

J

Jackson IJ. 1991. A reappraisal of non-consensus mRNA splice sites. *Nucleic acids research 19*:3795-3798.

Jacquier A. 1998. RNA-catalysed RNA processing. In: Krainer AR, ed. *Eukaryotic mRNA processing*. Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press.

Jacquier A, Rodriguez JR, Rosbash M. 1985. A quantitative analysis of the effects of 5' junction and TACTAAC box mutants and mutant combinations on yeast mRNA splicing. *Cell 43*:423-430.

Jacquier A, Rosbash M. 1986. Efficient trans-splicing of a yeast mitochondrial RNA group II intron implicates a strong 5' exon-intron interaction. *Science (New York, NY 234*:1099-1104.

Jady BE, Kiss T. 2001. A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *The EMBO journal 20*:541-551.

Jiang HQ, Motorin Y, Jin YX, Grosjean H. 1997. Pleiotropic effects of intron removal on base modification pattern of yeast tRNAPhe: an in vitro study. *Nucleic acids research 25*:2694-2701.

Jiang W, Middleton K, Yoon HJ, Fouquet C, Carbon J. 1993. An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules. *Molecular and cellular biology 13*:4884-4893.

Jin H, Loria JP, Moore PB. 2007. Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Molecular cell 26*:205-215.

Johnson L, Soll D. 1970. In vitro biosynthesis of pseudouridine at the polynucleotide level by an enzyme extract from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America 67*:943-950.

Johnson PF, Abelson J. 1983. The yeast tRNATyr gene intron is essential for correct modification of its tRNA product. *Nature 302*:681-687.

Jones CI, Spencer AC, Hsu JL, Spremulli LL, Martinis SA, DeRider M, Agris PF. 2006. A counterintuitive Mg2+-dependent and modification-assisted functional folding of mitochondrial tRNAs. *Journal of molecular biology 362*:771-786.

K

Kammen HO, Marvel CC, Hardy L, Penhoet EE. 1988. Purification, structure, and properties of *Escherichia coli* tRNA pseudouridine synthase I. *J Biol Chem 263*:2255-2263.

Kato N, Harada F. 1981. Nucleotide sequence of nuclear 5S RNA of mouse cells. *Biochem Biophys Res Commun 99*:1468-1476.

Kaya Y, Del Campo M, Ofengand J, Malhotra A. 2004. Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *The Journal of biological chemistry 279*:18107-18110.

Kaya Y, Ofengand J. 2003. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA (New York, NY 9*:711-721.

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

Kelley SO, Steinberg SV, Schimmel P. 2000. Functional defects of pathogenic human mitochondrial tRNAs related to structural fragility. *Nat Struct Biol 7*:862-865.

Kelmers AD, Novelli GD, Stulberg MP. 1965. Separation of transfer ribonucleic acids by reverse phase chromatography. *The Journal of biological chemistry 240*:3979-3983.

Kersten H. 1984. On the biological significance of modified nucleosides in tRNA. *Prog Nucleic Acid Res Mol Biol 31*:59-114.

Khaitovich P, Tenson T, Kloss P, Mankin AS. 1999. Reconstitution of functionally active Thermus aquaticus large ribosomal subunits with in vitro-transcribed rRNA. *Biochemistry 38*:1780-1788.

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

Kim DF, Green R. 1999. Base-pairing between 23S rRNA and tRNA in the ribosomal A site. *Molecular cell 4*:859-864.

Kim SH, Suddath FL, Quigley GJ, McPherson A, Sussman JL, Wang AH, Seeman NC, Rich A. 1974. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science (New York, NY 185*:435-440.

King TH, Liu B, McCully RR, Fournier MJ. 2003. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Molecular cell 11*:425-435.

Kirpekar F, Hansen LH, Rasmussen A, Poehlsgaard J, Vester B. 2005. The archaeon Haloarcula marismortui has few modifications in the central parts of its 23S ribosomal RNA. *Journal of molecular biology 348*:563-573.

Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. 1996. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell 85*:1077-1088.

Kiss AM, Jady BE, Darzacq X, Verheggen C, Bertrand E, Kiss T. 2002. A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucleic acids research 30*:4643-4649.

Kiss T. 2001. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *The EMBO journal 20*:3617-3622.

Kiss T, Jakab G, Antal M, Palfi Z, Hegyi H, Kis M, Solymosy F. 1988. Plant small nuclear RNAs. V. U4 RNA is present in broad bean plants in the form of sequence variants and is base-paired with U6 RNA. *Nucleic acids research 16*:5407-5426.

Knight S, Vulliamy T, Copplestone A, Gluckman E, Mason P, Dokal I. 1998. Dyskeratosis Congenita (DC) Registry: identification of new features of DC. *Br J Haematol 103*:990-996.

Koike I, Naito S, Adachi W, Tsunoda M, Ono A, Takenaka A. 2004. Structural significances of pseudouridination and methylation of uridine residues in tRNA: X-ray analyses of DNA dodecamers containing uridine and pseudouridine. *Nucleic Acids Symp Ser (Oxf)*:253-254.

Kolodrubetz D, Burgum A. 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*:79-90.

Kolossova I, Padgett RA. 1997. U11 snRNA interacts in vivo with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. *RNA (New York, NY 3*:227-233.

Koonin EV. 1996. Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic acids research 24*:2411-2415.

Kramer A. 1992. Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a presplicing complex. *Molecular and cellular biology 12*:4545-4552.

Krol A, Branlant C, Lazar E, Gallinaro H, Jacob M. 1981a. Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs. *Nucleic acids research 9*:2699-2716.

Kushner DJ, Bayley ST, Boring J, Kates M, Gibbons NE. 1964. Morphological and Chemical Properties of Cell Envelopes of the Extreme Halophile, Halobacterium Cutirubrum. *Canadian journal of microbiology 10*:483-497.

L

Labahn J, Scharer OD, Long A, Ezaz-Nikpay K, Verdine GL, Ellenberger TE. 1996. Structural basis for the excision repair of alkylation-damaged DNA. *Cell 86*:321-329.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*:680-685.

Lafontaine DL, Preiss T, Tollervey D. 1998b. Yeast 18S rRNA dimethylase Dim1p: a quality control mechanism in ribosome synthesis? *Molecular and cellular biology 18*:2360-2370.

Lafontaine DL, Tollervey D. 1998. Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem Sci 23*:383-388.

Lafontaine DLJ, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D. 1998a. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes & development 12*:527-537.

Langford CJ, Klinz FJ, Donath C, Gallwitz D. 1984. Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. *Cell 36*:645-653.

Langworthy TA, Smith PF, Mayberry WR. 1972. Lipids of Thermoplasma acidophilum. *Journal of bacteriology 112*:1193-1200.

Lecointe F, Namy O, Hatin I, Simos G, Rousset JP, Grosjean H. 2002. Lack of pseudouridine 38/39 in the anticodon arm of yeast cytoplasmic tRNA decreases in vivo recoding efficiency. *The Journal of biological chemistry 277*:30445-30453.

Lecointe F, Simos G, Sauer A, Hurt EC, Motorin Y, Grosjean H. 1998. Characterization of yeast protein Deg1 as pseudouridine synthase (Pus3) catalyzing the formation of psi 38 and psi 39 in tRNA anticodon loop. *The Journal of biological chemistry 273*:1316-1323.

Lee HW, Blasco MA, Gottlieb GJ, Horner JW, 2nd, Greider CW, DePinho RA. 1998. Essential role of mouse telomerase in highly proliferative organs. *Nature 392*:569-574.

Lesser CF, Guthrie C. 1993. Mutations in U6 snRNA that alter splice site specificity: implications for the active site [see comments]. *Science (New York, NY 262*:1982-1988.

Leulliot N, Godin KS, Hoareau-Aveilla C, Quevillon-Cheruel S, Varani G, Henry Y, Van Tilbeurgh H. 2007. The box H/ACA RNP assembly factor Naf1p contains a domain homologous to Gar1p mediating its interaction with Cbf5p. *Journal of molecular biology 371*:1338-1353.

Li HD, Zagorski J, Fournier MJ. 1990. Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in Saccharomyces cerevisiae. *Molecular and cellular biology 10*:1145-1152.

Li L, Ye K. 2006. Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature 443*:302-307.

Liang WQ, Fournier MJ. 1995. U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. *Genes & development 9*:2433-2443.

Libri D, Duconge F, Levy L, Vinauger M. 2002. A role for the Psi-U mismatch in the recognition of the 5' splice site of yeast introns by the U1 small nuclear ribonucleoprotein particle. *The Journal of biological chemistry 277*:18173-18181.

Limbach PA, Crain PF, McCloskey JA. 1994. Summary: the modified nucleosides of RNA. *Nucleic acids research 22*:2183-2196.

Lin Y, Kielkopf CL. 2008. X-ray Structures of U2 snRNA-Branchpoint Duplexes Containing Conserved Pseudouridines. *Biochemistry*.

Lubben B, Fabrizio P, Kastner B, Luhrmann R. 1995. Isolation and characterization of the small nucleolar ribonucleoprotein particle snR30 from Saccharomyces cerevisiae. *The Journal of biological chemistry 270*:11549-11554.

Luhrmann R, Appel B, Bringmann P, Rinke J, Reuter R, Rothe S, Bald R. 1982. Isolation and characterization of rabbit anti-m3 2,2,7G antibodies. *Nucleic acids research 10*:7103-7113.

Lustig F, Boren T, Claesson C, Simonsson C, Barciszewska M, Lagerkvist U. 1993. The nucleotide in position 32 of the tRNA anticodon loop determines ability of anticodon UCC to discriminate among glycine codons. *Proc Natl Acad Sci USA 90*:3343-3347.

M

Ma X, Yang C, Alexandrov A, Grayhack EJ, Behm-Ansmant I, Yu YT. 2005. Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *Embo J 24*:2403-2413.

Ma X, Zhao X, Yu YT. 2003. Pseudouridylation (Psi) of U2 snRNA in S.cerevisiae is catalyzed by an RNA-independent mechanism. *The EMBO journal 22*:1889-1897.

Maden BE. 1990. The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol 39*:241-303.

Maden BE, Hughes JM. 1997. Eukaryotic ribosomal RNA: the recent excitement in the nucleotide modification problem. *Chromosoma 105*:391-400.

Maden BE, Salim M. 1974. The methylated nucleotide sequences in HELA cell ribosomal RNA and its precursors. *Journal of molecular biology 88*:133-152.

Madhani HD, Guthrie C. 1994. Dynamic RNA-RNA interactions in the spliceosome. *Annu Rev Genet 28*:1-26.

Madore E, Florentz C, Giege R, Sekine S, Yokoyama S, Lapointe J. 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*:1128-1135.

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

Marquet R. 1998. Importance of modified nucleotides in replication of retroviruses, plant pararetroviruses, and retrotransposons. In: Grosjean H, Benne R, eds. *The modification and editing of RNA*. Washington, DC: ASM Press. pp 517-533.

Massenet S, Motorin Y, Lafontaine DL, Hurt EC, Grosjean H, Branlant C. 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*:2142-2154.

Massenet S, Mougin A, Branlant C. 1998. Posttranscriptional modifications in the U Small Nuclear RNAs. In: Grosjean H, Benne R, eds. *The modification and editing of RNA*. Washington, DC: ASM Press. pp 201-228.

Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of molecular biology 288*:911-940.

Maxwell ES, Fournier MJ. 1995. The small nucleolar RNAs. *Annu Rev Biochem 64*:897-934.

Meier UT, Blobel G. 1994. NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria. *The Journal of cell biology 127*:1505-1514.

Meroueh M, Grohar PJ, Qiu J, SantaLucia J, Jr., Scaringe SA, Chow CS. 2000. Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of Escherichia coli 23S rRNA. *Nucleic acids research 28*:2075-2083.

Merryman C, Moazed D, Daubresse G, Noller HF. 1999a. Nucleotides in 23S rRNA protected by the association of 30S and 50S ribosomal subunits. *Journal of molecular biology 285*:107-113.

Merryman C, Moazed D, McWhirter J, Noller HF. 1999b. Nucleotides in 16S rRNA protected by the association of 30S and 50S ribosomal subunits. *Journal of molecular biology 285*:97-105.

Mewes HW, Albermann K, Heumann K, Liebl S, Pfeiffer F. 1997. MIPS: a database for protein sequences, homology data and yeast genome information. *Nucleic acids research 25*:28-30.

Milstone DS, Vold BS, Glitz DG, Shutt N. 1978. Antibodies to N6-(delta2-isopentenyl) adenosine and its nucleotide: interaction with purified tRNAs and with bases, nucleosides and nucleotides of the isopentenyladenosine family. *Nucleic acids research 5*:3439-3455.

Miroux B, Walker JE. 1996. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of molecular biology 260*:289-298.

Mitchell JR, Wood E, Collins K. 1999. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature 402*:551-555.

Miyazaki M. 1974. Studies on the nucleotide sequence of pseudouridine-containing 5S RNA from Saccharomyces cerevisiae. *J Biochem 75*:1407-1410.

Moore JM, Query CC, Sharp PA. 1993. Splicing of precursors to messenger RNAs by the spliceosome. In: Gesteland R, Athuis J, eds. *The RNA world*. Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press. pp 303-358.

Morin A, Auxilien S, Senger B, Tewari R, Grosjean H. 1998. Structural requirements for enzymatic formation of threonylcarbamoyladenosine (t6A) in tRNA: an in vivo study with Xenopus laevis oocytes. *RNA (New York, NY 4*:24-37.

Morrissey JP, Tollervey D. 1995. Birth of the snoRNPs: the evolution of RNase MRP and the eukaryotic pre- rRNA-processing system. *Trends Biochem Sci 20*:78-82.

Motorin Y, Bec G, Tewari R, Grosjean H. 1997. Transfer RNA recognition by the Escherichia coli delta2-isopentenyl- pyrophosphate:tRNA delta2-isopentenyl transferase: dependence on the anticodon arm structure. *RNA (New York, NY 3*:721-733.

Motorin Y, Grosjean H. 1998. Chemical structures and classification of posttranscriptionally modified nucleosides in RNA. In: Grosjean H, Benne R, eds. *The modification and editing of RNA*. Washington, DC: ASM Press. pp 543-550.

Motorin Y, Keith G, Simon C, Foiret D, Simos G, Hurt E, Grosjean H. 1998. The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA (New York, NY 4*:856-869.

Motorin Y, Muller S, Behm-Ansmant I, Branlant C. 2007. Identification of modified residues in RNAs by reverse transcription-based methods. *Methods in enzymology 425*:21-53.

Mount SM. 1996. AT-AC introns: an ATtACk on dogma [comment]. *Science (New York, NY 271*:1690-1692.

Mount SM, Steitz JA. 1981. Sequence of U1 RNA from Drosophila melanogaster: implications for U1 secondary structure and possible involvement in splicing. *Nucleic acids research 9*:6351-6368.

Mueller SO, Slany RK. 1995. Structural analysis of the interaction of the tRNA modifying enzymes Tgt and QueA with a substrate tRNA. *FEBS Lett 361*:259-264.

Muller S, Fourmann JB, Loegler C, Charpentier B, Branlant C. 2007. Identification of determinants in the protein partners aCBF5 and aNOP10 necessary for the tRNA:Psi55-synthase and RNA-guided RNA:Psi-synthase activities. *Nucleic acids research 35*:5610-5624.

Muller S, Leclerc F, Behm-Ansmant I, Fourmann JB, Charpentier B, Branlant C. 2008. Combined in silico and experimental identification of the Pyrococcus abyssi H/ACA sRNAs and their target sites in ribosomal RNAs. *Nucleic acids research 36*:2459-2475.

Mumberg D, Muller R, Funk M. 1995. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene 156*:119-122.

Munns TW, Liszewski MK. 1980. Antibodies specific for modified nucleosides: an immunochemical approach for the isolation and characterization of nucleic acids. *Prog Nucleic Acid Res Mol Biol 24*:109-165.

Myslinski E, Branlant C, Wieben ED, Pederson T. 1984. The small nuclear RNAs of Drosophila. *J Mol Biol 180*:927-945.

N

Newby MI, Greenbaum NL. 2001. A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA (New York, NY 7*:833-845.

Newby MI, Greenbaum NL. 2002. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat Struct Biol 9*:958-965.

Newman A. 1994. Small nuclear RNAs and pre-mRNA splicing. *Curr Opin Cell Biol 6*:360-367.

Newman A, Norman C. 1991. Mutations in yeast U5 snRNA alter the specificity of 5' splice-site cleavage. *Cell 65*:115-123.

Newman AJ, Norman C. 1992. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell 68*:743-754.

Newman AJ, Teigelkamp S, Beggs JD. 1995. snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. *RNA (New York, NY 1*:968-980.

Ni J, Tien AL, Fournier MJ. 1997. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell 89*:565-573.

Nilsen TW. 1998. RNA-RNA interactions in nuclear pre-mRNA splicing. In: Simos RW, Grunberg-Manago M, eds. *RNA structure and fonction*. Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press. pp 279-308.

Nishimura H. 1979. Chromatographic mobilities of modified nucleotides. In: Schimmel P, Söll D, Abelson J, eds. *Transfer RNA. Structure, Properties and Recognition*. Cold Spring Harbor: CSHL Press. pp 547-552.

Nishimura S. 1983. Structure, biosynthesis, and function of queuosine in transfer RNA. *Prog Nucleic Acid Res Mol Biol 28*:49-73.

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science (New York, NY 289*:920-930.

Normand C, Capeyrou R, Quevillon-Cheruel S, Mougin A, Henry Y, Caizergues-Ferrer M. 2006. Analysis of the binding of the N-terminal conserved domain of yeast Cbf5p to a box H/ACA snoRNA. *RNA (New York, NY 12*:1868-1882.

Nurse K, Wrzesinski J, Bakin A, Lane BG, Ofengand J. 1995. Purification, cloning, and properties of the tRNA psi 55 synthase from *Escherichia coli*. *RNA (New York, NY 1*:102-112.

O

Ofengand J. 2002. Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett 514*:17-25.

Ofengand J, Del Campo M, Kaya Y. 2001a. Mapping pseudouridines in RNA molecules. *Methods 25*:365-373.

Ofengand J, Fournier MJ. 1998. The pseudouridine residues of rRNA: number, location, biosynthesis and function. In: Grosjean H, Benne R, eds. *The modification and editing of RNA*. Washington, DC: ASM Press. pp 229-254.

Ofengand J, Malhotra A, Remme J, Gutgsell NS, Del Campo M, Jean-Charles S, Peil L, Kaya Y. 2001b. Pseudouridines and pseudouridine synthases of the ribosome. *Cold Spring Harb Symp Quant Biol 66*:147-159.

Omer AD, Ziesche S, Decatur WA, Fournier MJ, Dennis PP. 2003. RNA-modifying machines in archaea. *Molecular microbiology 48*:617-629.

P

Pan H, Agarwalla S, Moustakas DT, Finer-Moore J, Stroud RM. 2003. Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proceedings of the National Academy of Sciences of the United States of America 100*:12648-12653.

Pan H, Ho JD, Stroud RM, Finer-Moore J. 2007. The crystal structure of E. coli rRNA pseudouridine synthase RluE. Journal of molecular biology 367:1459-1470.

Parker R, Siliciano PG, Guthrie C. 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell 49*:229-239.

Patton JR. 1994. Pseudouridine formation in small nuclear RNAs. *Biochimie 76*:1129-1132.

Patton JR, Bykhovskaya Y, Mengesha E, Bertolotto C, Fischel-Ghodsian N. 2005. Mitochondrial myopathy and sideroblastic anemia (MLASA): missense mutation in the pseudouridine synthase 1 (PUS1) gene is associated with the loss of tRNA pseudouridylation. *The Journal of biological chemistry 280*:19823-19828.

Patton JR, Jacobson MR, Pederson T. 1994. Pseudouridine formation in U2 small nuclear RNA. *Proceedings of the National Academy of Sciences of the United States of America 91*:3324-3328.

Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America 85*:2444-2448.

Pellizzoni L, Baccon J, Charroux B, Dreyfuss G. 2001. The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. *Curr Biol 11*:1079-1088.

Perret V, Garcia A, Puglisi J, Grosjean H, Ebel JP, Florentz C, Giege R. 1990a. Conformation in solution of yeast tRNA(Asp) transcripts deprived of modified nucleotides. *Biochimie 72*:735-743.

Persson BC, Bjork GR. 1993. Isolation of the gene (miaE) encoding the hydroxylase involved in the synthesis of 2-methylthio-cis-ribozeatin in tRNA of Salmonella typhimurium and characterization of mutants. *Journal of bacteriology 175*:7776-7785.

Phannachet K, Elias Y, Huang RH. 2005. Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase. *Biochemistry 44*:15488-15494.

Phannachet K, Huang RH. 2004. Conformational change of pseudouridine 55 synthase upon its association with RNA substrate. *Nucleic acids research 32*:1422-1429.

Pienkowska J, Michalowski D, Krzyzosiak WJ, Szweykowska-Kulinska Z. 2002. Pseudouridylation of U(35) in the anticodon of Arabidopsis thaliana pre- tRNA(Tyr) depends on length rather than structure of an intron. *Biochim Biophys Acta 1574*:137-144.

Q

Qu HL, Michot B, Bachellerie JP. 1983. Improved methods for structure probing in large RNAs: a rapid 'heterologous' sequencing approach is coupled to the direct mapping of nuclease accessible sites. Application to the 5' terminal domain of eukaryotic 28S rRNA. *Nucleic acids research 11*:5903-5920.

Query CC, Moore MJ, Sharp PA. 1994. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes & development 8*:587-597.

R

Rainen L, Stollar BD. 1978. Antibodies distinguishing between intact and alkali-hydrolyzed 7-methylguanosine. *Nucleic acids research 5*:3877-3889.

Ramamurthy V, Swann SL, Paulson JL, Spedaliere CJ, Mueller EG. 1999. Critical aspartic acid residues in pseudouridine synthases. *The Journal of biological chemistry 274*:22225-22230.

Randolph JB, Waggoner AS. 1997. Stability, specificity and fluorescence brightness of multiply-labeled fluorescent DNA probes. *Nucleic acids research 25*:2923-2929.

Rashid R, Liang B, Baker DL, Youssef OA, He Y, Phipps K, Terns RM, Terns MP, Li H. 2006. Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Molecular cell 21*:249-260.

Ray BK, Apirion D. 1979. Characterization of 10S RNA: a new stable rna molecule from Escherichia coli. *Mol Gen Genet 174*:25-32.

Raychaudhuri S, Conrad J, Hall BG, Ofengand J. 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 (New York, NY 4*:1407-1417.

Raychaudhuri S, Niu L, Conrad J, Lane BG, Ofengand J. 1999. Functional effect of deletion and mutation of the Escherichia coli ribosomal RNA and tRNA pseudouridine synthase RluA. *The Journal of biological chemistry 274*:18880-18886.

Reddy R. 1988. Compilation of small RNA sequences. *Nucleic acids research 16*:r71-85.

Reddy R, Henning D, Busch H. 1985. Primary and secondary structure of U8 small nuclear RNA. *The Journal of biological chemistry 260*:10930-10935.

Redlak M, Andraos-Selim C, Giege R, Florentz C, Holmes WM. 1997. Interaction of tRNA with tRNA (guanosine-1)methyltransferase: binding specificity determinants involve the dinucleotide G36pG37 and tertiary structure. *Biochemistry 36*:8699-8709.

Reichow SL, Hamma T, Ferre-D'Amare AR, Varani G. 2007. The structure and function of small nucleolar ribonucleoproteins. *Nucleic acids research 35*:1452-1464.

Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BF, Klug A. 1974. Structure of yeast phenylalanine tRNA at 3 A resolution. *Nature 250*:546-551.

Romier C, Ficner R, Suck D. 1998. Structural basis of base exchange by tRNA-guanine transglycosylases. In: Grosjean H, Benne R, eds. *The modification and editing of RNA*. Washington, DC: ASM Press. pp 169-182.

Roovers M, Hale C, Tricot C, Terns MP, Terns RM, Grosjean H, Droogmans L. 2006. Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic acids research 34*:4293-4301.

Rost B, Sander C, Schneider R. 1994. PHD--an automatic mail server for protein secondary structure prediction. *Comput Appl Biosci 10*:53-60.

Roth JR, Silbert DF, Fink GR, Voll MJ, Anton D, Hartman PE, Ames BN. 1966. Transfer RNA and the control of the histidine operon. *Cold Spring Harb Symp Quant Biol 31*:383-392.

Rozenski J, Crain PF, McCloskey JA. 1999. The RNA Modification Database: 1999 update. *Nucleic acids research 27*:196-197.

Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, 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 research 31*:869-877.

Rubin GM. 1973. The nucleotide sequence of Saccharomyces cerevisiae 5.8 S ribosomal ribonucleic acid. *The Journal of biological chemistry 248*:3860-3875.

Rudolph KL, Chang S, Lee HW, Blasco M, Gottlieb GJ, Greider C, DePinho RA. 1999. Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell 96*:701-712.

Ruggero D, Grisendi S, Piazza F, Rego E, Mari F, Rao PH, Cordon-Cardo C, Pandolfi PP. 2003. Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. *Science (New York, NY 299*:259-262.

Rymond BC, Rosbash M. 1985. Cleavage of 5' splice site and lariat formation are independent of 3' splice site in yeast mRNA splicing. *Nature 317*:735-737.

S

Samarsky DA, Fournier MJ. 1999. A comprehensive database for the small nucleolar RNAs from Saccharomyces cerevisiae. *Nucleic acids research 27*:161-164.

Samuelsson T. 1991. Interactions of transfer RNA pseudouridine synthases with RNAs substituted with fluorouracil. *Nucleic acids research 19*:6139-6144.

Samuelsson T, Olsson M. 1990. Transfer RNA pseudouridine synthases in *Saccharomyces cerevisiae*. *J Biol Chem 265*:8782-8787.

Santi DV, Hardy LW. 1987. Catalytic mechanism and inhibition of tRNA (uracil-5-)methyltransferase: evidence for covalent catalysis. *Biochemistry 26*:8599-8606.

Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic acids research 33*:W686-689.

Schattner P, Decatur WA, Davis CA, Ares M, Jr., Fournier MJ, Lowe TM. 2004. Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the Saccharomyces cerevisiae genome. *Nucleic acids research 32*:4281-4296.

Schultz DW, Yarus M. 1994a. tRNA structure and ribosomal function. I. tRNA nucleotide 27-43 mutations enhance first position wobble. *Journal of molecular biology 235*:1381-1394.

Schultz DW, Yarus M. 1994b. tRNA structure and ribosomal function. II. Interaction between anticodon helix and other tRNA mutations. *Journal of molecular biology 235*:1395-1405.

Segault V, Will CL, Sproat BS, 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. *The EMBO journal 14*:4010-4021.

Senger B, Auxilien S, Englisch U, Cramer F, Fasiolo F. 1997. The modified wobble base inosine in yeast tRNAIle is a positive determinant for aminoacylation by isoleucyl-tRNA synthetase. *Biochemistry 36*:8269-8275.

Seraphin B, Abovich N, Rosbash M. 1991. Genetic depletion indicates a late role for U5 snRNP during in vitro spliceosome assembly. *Nucleic acids research 19*:3857-3860.

Serebrov V, Vassilenko K, Kholod N, Gross HJ, Kisselev L. 1998. Mg2+ binding and structural stability of mature and in vitro synthesized unmodified Escherichia coli tRNAPhe. *Nucleic acids research 26*:2723-2728.

Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic acids research 15*:7155-7174.

Sherman F, Hicks J. 1991. Micromanipulation and dissection of asci. *Methods in enzymology 194*:21-37.

Shi H, Moore PB. 2000. The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA (New York, NY 6*:1091-1105.

Silberklang M, Gillum AM, RajBhandary UL. 1979. Use of in vitro 32P labeling in the sequence analysis of nonradioactive tRNAs. *Methods in enzymology 59*:58-109.

Simos G, Tekotte H, Grosjean H, Segref A, Sharma K, Tollervey D, Hurt EC. 1996. Nuclear pore proteins are involved in the biogenesis of functional tRNA. *The EMBO journal 15*:2270-2284.

Sivaraman J, Sauve V, Larocque R, Stura EA, Schrag JD, Cygler M, Matte A. 2002. Structure of the 16S rRNA pseudouridine synthase RsuA bound to uracil and UMP. *Nat Struct Biol 9*:353-358.

Solymosy F, Pollak T. 1993. Uridylate-Rich Small Nuclear RNAs (UsnRNAs), their genes and pseudogenes, and UsnRNPS in plants: stucture and function. A comparative approach. *Critical Reviews in Plant Sciences 12*:275-369.

Spedaliere CJ, Hamilton CS, Mueller EG. 2000. Functional importance of motif I of pseudouridine synthases: mutagenesis of aligned lysine and proline residues. *Biochemistry 39*:9459-9465.

Spedaliere CJ, Mueller EG. 2004. Not all pseudouridine synthases are potently inhibited by RNA containing 5-fluorouridine. *RNA (New York, NY 10*:192-199.

Sprinzl S, Horn C, Brown M, Ioudovitch A, Steinberg S. 1998. Compilation of tRNA sequences and sequences of tRNA genes. *NAR 26*:148-153.

Staley JP, Guthrie C. 1999. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Molecular cell 3*:55-64.

Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification 41*:207-234.

Sumita M, Desaulniers JP, Chang YC, Chui HM, Clos L, 2nd, Chow CS. 2005. Effects of nucleotide substitution and modification on the stability and structure of helix 69 from 28S rRNA. *RNA (New York, NY 11*:1420-1429.

Sun JS, Manley JL. 1995. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes & development 9*:843-854.

Szkukalek A, Myslinski E, Mougin A, Luhrmann R, Branlant C. 1995. Phylogenetic conservation of modified nucleotides in the terminal loop 1 of the spliceosomal U5 snRNA. *Biochimie 77*:16-21.

Szweykowska-Kulinska Z, Beier H. 1992. Sequence and structure requirements for the biosynthesis of pseudouridine (psi 35) in plant pre-tRNA(Tyr). *The EMBO journal 11*:1907-1912.

Szweykowska-Kulinska Z, Senger B, Keith G, Fasiolo F, Grosjean H. 1994. Intron-dependent formation of pseudouridines in the anticodon of Saccharomyces cerevisiae minor tRNA(Ile). *The EMBO journal 13*:4636-4644.

Szymanski M, Barciszewska MZ, Barciszewski J, Erdmann VA. 2000. 5S ribosomal RNA database Y2K. *Nucleic acids research 28*:166-167.

Szymanski M, Barciszewska MZ, Erdmann VA, Barciszewski J. 2002. 5S Ribosomal RNA Database. *Nucleic acids research 30*:176-178.

T

Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Huttenhofer A. 2002. Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proceedings of the National Academy of Sciences of the United States of America 99*:7536-7541.

Tarn WY, Steitz JA. 1996. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell 84*:801-811.

Tarn WY, Steitz JA. 1997. Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci 22*:132-137.

Terns MP, Terns RM. 2001. Macromolecular complexes: SMN--the master assembler. *Curr Biol 11*:R862-864.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research 22*:4673-4680.

Tollervey D. 1987. A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *The EMBO journal 6*:4169-4175.

Tollervey D. 1996. Small nucleolar RNAs guide ribosomal RNA methylation. *Science (New York, NY 273*:1056-1057.

Tollervey D, Kiss T. 1997. Function and synthesis of small nucleolar RNAs. *Curr Opin Cell Biol 9*:337-342.

Tomita K, Ueda T, Watanabe K. 1999. The presence of pseudouridine in the anticodon alters the genetic code: a possible mechanism for assignment of the AAA lysine codon as asparagine in echinoderm mitochondria. *Nucleic acids research 27*:1683-1689.

Torchet C, Badis G, Devaux F, Costanzo G, Werner M, Jacquier A. 2005. The complete set of H/ACA snoRNAs that guide rRNA pseudouridylations in Saccharomyces cerevisiae. *RNA (New York, NY 11*:928-938.

Tremblay A, Lamontagne B, Catala M, Yam Y, Larose S, Good L, Elela SA. 2002. A physical interaction between Gar1p and Rnt1pi is required for the nuclear import of H/ACA small nucleolar RNA-associated proteins. *Molecular and cellular biology 22*:4792-4802.

Tsui HC, Arps PJ, Connolly DM, Winkler ME. 1991. Absence of hisT-mediated tRNA pseudouridylation results in a uracil requirement that interferes with Escherichia coli K-12 cell division. *Journal of bacteriology 173*:7395-7400.

Tworowska I, Nikonowicz EP. 2006. Base pairing within the psi32,psi39-modified anticodon arm of Escherichia coli tRNA(Phe). *J Am Chem Soc 128*:15570-15571.

U

Umen JG, Guthrie C. 1995. A novel role for a U5 snRNP protein in 3' splice site selection. *Genes & development 9*:855-868.

Urban A, Ansmant I, Motorin Y. 2003. Optimisation of expression and purification of the recombinant Yol066 (Rib2) protein from Saccharomyces cerevisiae. *Journal of chromatography 786*:187-195.

V

Valadkhan S, Manley JL. 2001. Splicing-related catalysis by protein-free snRNAs. *Nature 413*:701-707.

Valadkhan S, Manley JL. 2003. Characterization of the catalytic activity of U2 and U6 snRNAs. *RNA (New York, NY 9*:892-904.

van Tol H, Beier H. 1988. All human tRNATyr genes contain introns as a prerequisite for pseudouridine biosynthesis in the anticodon. *Nucleic acids research 16*:1951-1966.

Venema J, Tollervey D. 1995. Processing of pre-ribosomal RNA in Saccharomyces cerevisiae. *Yeast 11*:1629-1650.

Venema J, Tollervey D. 1999. Ribosome synthesis in Saccharomyces cerevisiae. *Annu Rev Genet 33*:261-311.

Vijayraghavan U, Parker R, Tamm J, Iimura Y, Rossi J, Abelson J, Guthrie C. 1986. Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. *The EMBO journal 5*:1683-1695.

Vold BS. 1981. Preparation and specificity of antibodies directed toward the ribose methylated nucleotide, 2'-O-methylguanosine 5'-monophosphate. *Biochim Biophys Acta 655*:265-267.

Vulliamy T, Marrone A, Goldman F, Dearlove A, Bessler M, Mason PJ, Dokal I. 2001. The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature 413*:432-435.

W

Wang C, Meier UT. 2004. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *The EMBO journal 23*:1857-1867.

Watanabe Y, Gray MW. 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 research 28*:2342-2352.

Watkins NJ, Gottschalk A, Neubauer G, Kastner B, Fabrizio P, Mann M, Luhrmann R. 1998a. Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA- binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA (New York, NY 4*:1549-1568.

Watkins NJ, Segault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Luhrmann R. 2000. A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell 103*:457-466.

Westhof E, Dumas P, Moras D. 1985. Crystallographic refinement of yeast aspartic acid transfer RNA. *Journal of molecular biology 184*:119-145.

Westhof E, Dumas P, Moras D. 1988. Hydration of transfer RNA molecules: a crystallographic study. *Biochimie 70*:145-165.

Will CL, Luhrmann R. 1997. SnRNP structure and function. In: Krainer AR, ed. *Eukaryotic mRNA processing*. Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press. pp 130-173.

Will CL, Luhrmann R. 2001. Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol 13*:290-301.

Winkler AA, Bobok A, Zonneveld BJ, Steensma HY, Hooykaas PJ. 1998. The lysine-rich C-terminal repeats of the centromere-binding factor 5 (Cbf5) of Kluyveromyces lactis are not essential for function. *Yeast 14*:37-48.

Wise JA, Tollervey D, Maloney D, Swerdlow H, Dunn EJ, Guthrie C. 1983. Yeast contains small nuclear RNAs encoded by single copy genes. *Cell 35*:743-751.

Wittenberg WL, Uhlenbeck OC. 1985. Specific replacement of functional groups of uridine-33 in yeast phenylalanine transfer ribonucleic acid. *Biochemistry 24*:2705-2712.

Woese CR. 1977. A comment on methanogenic bacteria and the primitive ecology. *Journal of molecular evolution 9*:369-371.

Woese CR. 1987. Bacterial evolution. *Microbiological reviews 51*:221-271.

Wolff T, Bindereif A. 1993. Conformational changes of U6 RNA during the spliceosome cycle: an intramolecular helix is essential both for initiating the U4-U6 interaction and for the first step of slicing. *Genes & development 7*:1377-1389.

Wrzesinski J, Nurse K, Bakin A, Lane BG, Ofengand J. 1995a. 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 (New York, NY 1*:437-448.

Wu H, 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. *Proceedings of the National Academy of Sciences of the United States of America 104*:6655-6660.

Wu JA, Manley JL. 1991. Base pairing beetween U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA. *Nature 352*:818-821.

Wu Q, Krainer AR. 1999. AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. *Molecular and cellular biology 19*:3225-3236.

Y

Yang C, McPheeters DS, Yu YT. 2005. Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in Saccharomyces cerevisiae. *The Journal of biological chemistry 280*:6655-6662.

Yang PK, Rotondo G, Porras T, Legrain P, Chanfreau G. 2002. The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis. *The Journal of biological chemistry 277*:45235-45242.

Yarian CS, Basti MM, Cain RJ, Ansari G, Guenther RH, Sochacka E, Czerwinska G, Malkiewicz A, Agris PF. 1999. Structural and functional roles of the N1- and N3-protons of psi at tRNA's position 39. *Nucleic acids research 27*:3543-3549.

Yu YT. 2000. Site-specific 4-thiouridine incorporation into RNA molecules. *Methods in enzymology 318*:71-88.

Yu YT, Shu MD, Steitz JA. 1998. Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *The EMBO journal 17*:5783-5795.

Yu YT, Steitz JA. 1997. Site-specific crosslinking of mammalian U11 and u6atac to the 5' splice site of an AT-AC intron. *Proc Natl Acad Sci USA 94*:6030-6035.

Yue D, Kintanar A, Horowitz J. 1994. Nucleoside modifications stabilize Mg2+ binding in Escherichia coli tRNA(Val): an imino proton NMR investigation. *Biochemistry 33*:8905-8911.

Z

Zamore PD, Green MR. 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proceedings of the National Academy of Sciences of the United States of America 86*:9243-9247.

Zamore PD, Green MR. 1991. Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *The EMBO journal 10*:207-214.

Zavanelli MI, Ares M, Jr. 1991. Efficient association of U2 snRNPs with pre-mRNA requires an essential U2 RNA structural element. *Genes & development 5*:2521-2533.

Zavanelli MI, Britton JS, Igel AH, Ares M, Jr. 1994. Mutations in an essential U2 small nuclear RNA structure cause cold- sensitive U2 small nuclear ribonucleoprotein function by favoring competing alternative U2 RNA structures. *Mol Cell Biol 14*:1689-1697.

Zebarjadian Y, King T, Fournier MJ, Clarke L, Carbon J. 1999. Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Molecular and cellular biology 19*:7461-7472.

Zerfass K, Beier H. 1992. Pseudouridine in the anticodon G psi A of plant cytoplasmic tRNA(Tyr) is required for UAG and UAA suppression in the TMV-specific context. *Nucleic acids research 20*:5911-5918.

Zhao X, Li ZH, Terns RM, Terns MP, Yu YT. 2002. An H/ACA guide RNA directs U2 pseudouridylation at two different sites in the branchpoint recognition region in *Xenopus oocytes*. *RNA (New York, NY 8*:1515-1525.

Zhao X, Yu YT. 2004. 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 (New York, NY 10*:681-690.

Zinesser F. 1906. Atrophia cutis reticularis cum pigmentatione, dystrophia unguium et leukoplakia oris. *Ikonogr Dermatol 5*:219-223.

Zuker M, Jacobson AB. 1998. Using reliability information to annotate RNA secondary structures. *RNA (New York, NY 4*:669-679.

Résumé

Titre : Caractérisation biochimique, structurale et fonctionnelle des ARN:pseudouridine synthases Pus7 de la levure *Saccharomyces cerevisiae* et d'archaea thermophiles

La conversion des résidus U en pseudouridine (Ψ) est la modification la plus fréquente dans les ARN. Cette réaction est catalysée par des ARN:Ψ-synthases qui fonctionnent soit seules, soit au sein de particules snoRNP H/ACA composées de 4 protéines et d'un snoRNA.

Nous avons démontré que l'ARN:Ψ-synthase Pus7p de *S. cerevisiae* est capable d'agir sur les ARNt cytoplasmiques (position 13 dans 10 ARNt et position 35 dans le pré-ARNt^{Tyr} contenant un intron). En parallèle, j'ai recherché les requis en séquence et en structure pour qu'un ARN soit substrat de Pus7p. J'ai également réalisé une étude préliminaire de l'importance des domaines additionnels de Pus7p par rapport à l'enzyme TruD d'*E. coli*.

Comme les systèmes enzymatiques responsables de la formation de trois résidus Ψ du snRNA U2 de *S. cerevisiae* avaient été identifiés, nous avons pu débuter une étude de l'importance fonctionnelle de ces résidus pour la réaction d'épissage en utilisant l'approche globale par puces à ADN que l'équipe de J. Beggs (Edinburgh University) avait mise au point pour *S. cerevisiae*. Nous avons ainsi observé une baisse de l'efficacité d'épissage de certains ARN pré-messagers lorsque plusieurs gènes codant des U2 snRNA:Ψ-synthases sont inactivés simultanément.

L'ARN:Ψ-synthase Pus7p est présente dans les 3 domaines du vivant et en particulier chez les archaea où l'activité de cette enzyme n'avait jamais été étudiée. Nous avons découvert l'existence de 2 systèmes enzymatiques différents pour la modification d'une même position d'un ARNt chez 2 espèces d'archaea thermophiles : une enzyme seule chez *Pyrococcus abyssi* contre un système de guide sRNP H/ACA chez *Sulfolobus solfataricus* dans lequel la protéine Pus7 est mutée au niveau d'acides aminés requis pour l'activité. Finalement, nous avons produit chacune de ces enzymes en grandes quantités et nous avons réalisé des tests de cristallisation. Des cristaux de Pus7 de *P. abyssi* ont été obtenus et une mesure complète de données de diffraction de rayons X a été réalisée au laboratoire à 2,5 Å de résolution.

Mots clés : pseudouridine, ARN:Ψ-synthase, Pus7, tRNA, archaea, *Pyrococcus abyssi*.