Etudes structurales de différents processus biologiques impliquant les ARN de transfert

Pierre Barraud

▶ **To cite this version:**

Pierre Barraud. Etudes structurales de différents processus biologiques impliquant les ARN de transfert. Biochimie [q-bio.BM]. Université René Descartes - Paris V, 2007. Français. NNT : . tel-00364789

HAL Id: tel-00364789
https://theses.hal.science/tel-00364789

Submitted on 27 Feb 2009

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

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

Université Paris Descartes

Thèse

présentée et soutenue publiquement le 17 juillet 2008 par

Pierre Barraud

pour l'obtention du grade de
Docteur de l'Université Paris Descartes,
spécialité : Biologie Structurale
au titre de l'École Doctorale du Médicament

Etudes structurales de différents processus biologiques impliquant les ARN de transfert

Directeurs de thèse :	Carine	Tisné
	Frédéric	Dardel
Après avis de :	Jean-Pierre	Simorre
	Marat	Yusupov

Devant la commission d'examen formée de :

Jean-Pierre	Simorre	(rapporteur)
Marat	Yusupov	(rapporteur)
Louis	Droogmans	(examinateur)
Pascale	Romby	(examinateur)
Carine	Tisné	(directeur de thèse)
Frédéric	Dardel	(directeur de thèse)

Préparée au Laboratoire de Cristallographie et RMN Biologiques
Université Paris Descartes, CNRS UMR 8015
4, avenue de l'Observatoire, 75006 Paris

Remerciements

Cette thèse a été réalisée dans l'équipe « RMN, ARN et Interactions », au sein du Laboratoire de Cristallographie et RMN Biologiques, une unité mixte de recherche du CNRS et de l'Université Paris Descartes.

Merci à Carine Tisné, pour la disponibilité dont elle a fait preuve tout au long de cette thèse. Tout en me laissant une grande autonomie, elle a toujours su me guider et m'aider pour faire avancer les projets. J'ai beaucoup apprécié travailler avec elle, notamment lors de la rédaction des articles. Son indéfectible bonne humeur m'a de plus permis de passer trois agréables années qui n'auraient jamais été les mêmes sans elle.

Merci à Frédéric Dardel, qui m'a permis de réaliser ce travail au sein de son équipe, pour sa confiance et son soutien.

Merci à Jean-Pierre Simorre et Marat Yusupov, pour avoir accepté d'être rapporteur de ce travail et à Pascale Romby et Louis Droogmans, qui m'ont également fait l'honneur de prendre part au jury.

Merci à Béatrice Golinelli, qui m'a initié à la cristallographie, et qui m'a accueilli dans son équipe pendant plus d'un mois pour « les expériences radioactives ». Merci au passage à Sylvie Auxilien et Hélène Walbott pour leurs conseils lors de ce passage au LEBS.

Merci à Emmanuelle Schmitt et Yves Mechulam, pour les enseignements de qualité qu'ils donnent aux étudiants de M2, pour avoir rattrapé mes bourdes sans broncher lors de ces TPs, et pour notre collaboration enrichissante pendant ma thèse sur le « fMet ».

Merci à Cyril Gaudin, pour sa sympathie, son accueil et ses précieux conseils lors de mes passages à l'ICSN.

Merci à nos collaborateurs : Jean-Christophe Paillart de l'Institut de Biologie Moléculaire et Cellulaire à Strasbourg ; Jérôme Boisbouvier de l'Institut de Biologie Structurale à Grenoble ; Cédric Atmanene, Sarah Sanglier et Alain Van Dorsselaer du Laboratoire de Spectrométrie de Masse Bio-organique à Strasbourg.

Merci à l'ensemble des membres du laboratoire que j'ai côtoyé de près ou de loin durant ces trois années. Tout d'abord, un Grand Merci aux ingénieurs et techniciens du laboratoire, sans qui pas grand chose ne tournerait rond : Valéry Larue (pour la RMN), Philippe Bénas (pour la Cristallo et le BAG), Franck Brachet (pour les robots-gouttes), Nathalie Ulryck (pour la BM, et pour toute l'énergie mise au service de l'organisation du labo), Bili Seijo (pour le stage de clonage et pour tout le reste), ainsi que Cathel Tourmente et Patricia Schmidt pour l'administration. Merci ensuite à Arnaud Ducruix, son directeur, de m'avoir accueilli dans cette unité où se mélangent RMN-istes et Cristallographes. Merci également à tous ceux qui m'ont aidé au cours de cette thèse : Gilles Phan (Merci pour l'aide au labo, mais aussi pour toutes les bonnes soirées en dehors du labo), Stéphane Réty, Frédérique Maurice, Luc Ponchon, Isabelle Broutin, Marie-Bernard Lascombe, Thierry Prangé, et tous les autres membres du laboratoire, sans exception pour leur sympathie et leur disponibilité. Merci aussi aux anciens : Sylvestre Grizot, Angela Lopez-Giral et Daniel Auguin.

Merci au Poil O'Brass Band, sans qui bien évidemment rien de tout ça n'aurait été possible.

Merci à Mathilde et Melaine, mes deux colocataires de la bruyante N20.

Merci aux anciens lyonnais, thésards de la rive gauche ou profs de banlieue : Joaquim, Hélène, Samuel, Colas, Elise, Euzeb et les autres pour les bières, les frites–bières, les cinés–bières et les bières–bières.

Merci à Mario, mon partenaire d'errances parisiennes.

Merci à Damien, Yvon, Marlène, Nicolas, Tristan, Clément, Céline, Antoine et Raphaële, mes différents points de chutes provinciaux pour les bons week-end passés loin de la capitale, et notamment lors de mes séjours à l'ESRF.

Merci à ma famille.

Merci à Marie-Paule et François, mes parents.

Merci à ma grande sœur Hélène et merci à ma petite sœur Marie.

Sommaire

Remerciements **iii**

Avant propos **ix**

Introduction 1

1 Les acides ribonucléiques de transfert **3**
1.1 Structure des ARNt 3
Structure primaire 3
Structure secondaire 5
Structure tridimensionnelle 7
1.2 Fonction des ARNt 9
Fonction des ARNt dans la traduction 9
Fonctions alternatives des ARNt 12
1.3 Biosynthèse des ARNt 13
1.3.1 Maturation des extrémités 13
Maturation de l'extrémité 5′ 13
Maturation de l'extrémité 3′ 14
1.3.2 Modifications post-transcriptionnelles 16
Diversité des nucléotides modifiés dans les ARNt 16
Localisation des nucléotides modifiés dans les ARNt . . 18
Rôle des nucléotides modifiés dans les ARNt 19

Organisation du manuscrit et présentation des résultats **25**

I Formation du complexe d'initiation de la transcription inverse du VIH-1 29

2 Contexte biologique **31**
2.1 Les acteurs 31
2.1.1 La transcriptase inverse (RT) 31
Organisation structurale 31
La RT, cible des traitements antirétroviraux 32
2.1.2 L'ARN viral 33
2.1.3 L'amorce $ARNt_3^{Lys}$ 34

2.1.4 La protéine de nucléocapside (NC) 36
Généralités 36
Propriétés structurales 37
Propriétés de liaison aux acides nucléiques 39
2.2 Déroulement de la transcription inverse 40
Initiation et synthèse de l'ADN$_c$(−) *strong-stop* 40
Premier transfert de brin 41
Synthèse du brin d'ADNc(−) 41
Synthèse du brin d'ADNc(+) 41
Second transfert de brin 42
Terminaison de la transcription inverse 43
2.3 Formation de l'hybride ARNt$_3^{Lys}$/ARN viral 43
2.3.1 Hybridation de l'ARNt$_3^{Lys}$ avec le PBS 44
2.3.2 Interactions additionnelles entre l'ARNt$_3^{Lys}$ et l'ARN viral 47
Domaine PBS des virus de sous-type A et B 48
La boucle riche en A 48
La séquence *Primer Activation Signal* 50

3 Rôle de la NC dans la formation de l'hybride ARNt$_3^{Lys}$/PBS 53
3.1 Problématique et résumé de l'Article 1 53
3.2 Article 1 | Barraud *et al.* Biochimie (2007) 54

4 Etude par RMN du complexe NC/bras D de l'ARNt$_3^{Lys}$ 63
4.1 Première approche 63
4.1.1 Spectre HSQC et signature de la protéine 63
4.1.2 Optimisation de la longueur du bras D 65
4.1.3 La NC dans le complexe avec le bras D 67
4.2 Amélioration de la sensibilité des expériences HSQC 68
4.2.1 Problématique et résumé de l'Article 2 68
4.2.2 Article 2 | Barraud *et al.* C.R. Chimie (2008) 71
4.3 Poursuite de l'étude du complexe NC/bras D 78
4.3.1 Le complexe vu du côté de la protéine 78
Attribution à partir des expériences de titration 78
Attribution à partir d'expériences triple résonances . . . 79
4.3.2 Le complexe vu du côté de l'ARN 81
Obtention de l'ARN $^{13}C/^{15}N$ par transcription *in vitro* . 81
Comportement RMN de l'ARN au sein du complexe . . 82

5 Conclusion et Perspectives 85

II Etude d'une méthyltransférase des ARN de transfert – la m^{1}A$_{58}$ ARNt méthyltransférase 89

6 Contexte biologique 91
6.1 Les méthyltransférases 91
6.1.1 La S-adénosyl-L-méthionine 91

La SAM, un donneur de méthyle 92
La SAM, un cofacteur versatile 94
6.1.2 Classification des méthyltransférases 94
MTases de Classe I 95
MTases des Classes II à V 96
6.1.3 Motifs conservés parmi les méthyltransférases de Classe I 99
6.2 Les méthyltransférases d'ARNt 101
Diversité des méthylations rencontrées dans les ARNt . 101
Rôle des méthylations rencontrées dans les ARNt 101
Repliement des MTases d'ARNt 102
Atomes cibles et mécanisme 102
6.3 Les m^1A méthytransférases d'ARNt 103
6.3.1 Le nucléotide m^1A et sa localisation au sein des ARNt . 103
6.3.2 Les m^1A_{58} MTases 104
La famille des m^1A_{58} MTases 104
Particularités biochimiques des m^1A_{58} MTases 106
Fonctions avérées de la modification m^1A_{58} 107

7 Etude de la m^1A_{58} ARNt MTase de *T. thermophilus* 109
7.1 Importance de l'étude structurale des enzymes de modification 109
7.2 Projet concernant la m^1A_{58} ARNt MTase 111
7.3 Problématique et résumé de l'Article 3 113
7.4 Article 3 | Barraud *et al.* J. Mol. Biol. (2008) 114
7.5 Discussion concernant le mécanisme 131

8 Perspectives 135

III Structure Cristallographique de l'$ARNt_f^{Met}$ initiateur d'*E. coli* 139

9 Contexte biologique 141
9.1 L'initiation de la traduction chez les bactéries 141
9.2 Particularité des ARNt initiateurs 143
9.2.1 Déterminants au sein des ARNt initiateurs bactériens . 143
9.2.2 Particularité des anticodons des ARNt initiateurs 145
9.3 L'ARNt initiateur au sein du site P du ribosome 148
9.3.1 Les interactions au niveau du site P 148
Avec l'ARNr 16S 148
Avec les protéines S9 et S13 150
9.3.2 Une préférence pour les $ARNt_i^{Met}$ au sein du site P . . . 150
9.4 La sélection IF3-dépendante de l'ARNt initiateur 151

10 Une structure unique de l'anticodon de l'$ARNt_f^{Met}$ d'*E. coli* 155
10.1 Problématique et résumé de l'Article 4 155
10.2 Article 4 | Barraud *et al.* (2008) 156

11 Conclusion et Perspectives **167**

Appendices 169

Listes **171**
Liste des figures 171
Liste des tables 173
Liste des abréviations 175
Liste des publications 177

Bibliographie **181**

Annexes **213**
Classification des paires de bases dans l'ARN 215
Obtention des échantillons pour la RMN et la Cristallographie 219
Surproduction et purification de la NC_{1-55} 219
Surproduction et purification de TrmI 221
Surproduction et purification de l'$ARNt_f^{Met}$ de *E. coli* 223

Résumé & Abstract 226

Avant propos

Dans la cellule, les acides ribonucléiques de transfert ou ARNt sont les molécules clé du décodage de l'information génétique portée par l'acide désoxyribonucléique (ADN). Le processus de décodage se déroule en deux temps. La première phase est appelée Transcription. Elle correspond à la synthèse d'ARN messager (ARNm) à partir de l'ADN génomique. La transcription permet de segmenter l'information génétique en divers fragments, dont les quantités produites et la durée de vie — et donc le nombre de copies à l'intérieur de la cellule — peuvent être finement contrôlés, ce qui permet une régulation précise et efficace de l'expression du génome [1]. La deuxième phase, appelée Traduction, ou synthèse des protéines, consiste à passer de l'alphabet à quatre lettres (A, U, G, C) de l'ARNm, à un nouvel alphabet composé de 20 lettres — les 20 acides aminés qui entrent dans la composition des protéines. Le code génétique permet de passer d'un langage à l'autre. Dans la cellule, la clé du code génétique est matérialisée par les ARNt.

Les ARNt sont de petites molécules qui assurent le lien physique entre l'ARNm et la protéine naissante, lors du mécanisme de traduction. Ils portent en effet à une de leur extrémité un anticodon, formé de l'enchaînement de trois nucléotides, qui va s'associer avec les codons présents sur l'ARNm par complémentarité. Ils portent aussi, à leur autre extrémité, un acide aminé dont la nature dépend de l'anticodon présent de l'autre côté. Cet acide aminé sera incorporé à la protéine naissante après appariement correct entre le codon de l'ARNm et l'anticodon de l'ARNt. La fidélité de la traduction est conditionnée par la fixation du bon acide aminé sur l'ARNt qui lui correspond. Cette fonction est assurée par une famille d'enzyme, les aminoacyl-ARNt synthétases (aaRS). Le tandem aaRS–ARNt constitue un système de décodage fidèle de l'information génétique. En résumé, les ARNt correctement aminoacylés assurent le transfert de l'information génétique, depuis les ARN messagers — qui la tiennent directement de l'ADN par le mécanisme de transcription — vers les protéines, lors de la traduction.

Dans les cellules, la traduction a lieu au niveau des ribosomes. Les ribosomes sont des macromolécules complexes, formées d'ARN et de protéines, qui catalysent les étapes clés de la synthèse des protéines. Le ribosome est secondé dans cette tâche par tout un arsenal de facteurs de traduction. Comme toute réaction de polymérisation, la traduction peut être divisée en trois phases : l'initiation, l'élongation et la terminaison. La phase d'initiation est détermi-

nante car elle « choisit » en quelque sorte le codon de démarrage sur l'ARNm et par conséquent conditionne le cadre de lecture de celui-ci. Il s'agit donc d'une étape majeure dans la biologie de la cellule, de laquelle va dépendre la fidélité de l'expression des gènes. De manière universelle, la biosynthèse des protéines débute chez tous les êtres vivants par l'incorporation d'un acide aminé méthionine. Tout l'enjeu de l'étape d'initiation consiste donc à reconnaître le codon AUG codant pour la méthionine initiatrice sur l'ARNm. Les êtres vivants ont répondu à cet enjeu par l'utilisation d'un ARNt spécialisé, l'ARNt initiateur méthionylé (l'$ARNt_i^{Met}$). Cet ARNt particulier est recruté spécifiquement par le ribosome et les facteurs d'initiation parmi l'ensemble des ARNt présents dans la cellule, et plus particulièrement, il est distingué des ARNt élongateurs méthionylés qui possèdent le même anticodon. Une des questions importantes concernant l'initiation de la traduction est de savoir comment cet ARNt initiateur est distingué des ARNt élongateurs. De façon plus ou moins fortuite, le travail réalisé pendant ma thèse m'a amené à résoudre la structure de l'ARNt initiateur d'*E. coli* par cristallographie. Cette structure originale nous a permis de proposer un modèle pour la discrimination entre les ARNt initiateurs et élongateurs. Ce travail sera présenté dans la Partie III.

Pour être fonctionnels, les ARNt doivent subir de façon universelle — c'est-à-dire aussi bien pour les ARNt initiateurs que pour les ARNt élongateurs — des étapes de maturation post-transcriptionnelles. En particulier, certains de ses nucléotides subissent des modifications chimiques catalysées par des enzymes dites de modification. Tous les types d'acides nucléiques sont concernés par les modifications chimiques. Plus d'une centaine de nucléotides modifiés ont été identifiés, dont 91 pour les seuls ARNt [2]. C'est dans la famille des ARNt que l'on rencontre ainsi la plus grande diversité de modifications, mais aussi la plus grande proportion de nucléotides modifiés puisque la séquence de certains ARNt peut en comprendre jusqu'à un quart. Parmi toutes les modifications chimiques rencontrées dans les ARNt, les méthylations sont les plus nombreuses. Elles sont catalysées par des enzymes appelées méthyltransférases (MTases). Une partie de mon travail de thèse a consisté en l'étude biochimique et structurale d'une méthyltransférase des ARNt : la m^1A_{58} ARNt méthyltransférase de *T. thermophilus*. Ce travail sera présenté dans la Partie II.

Outre cette implication des ARNt dans la traduction, qui constitue leur rôle premier, ceux-ci prennent part à des processus biologiques variés, en dehors du phénomène de traduction. Ainsi certains ARNt servent d'amorce de polymérisation lors de l'initiation de la transcription inverse des rétrovirus et de certains rétrotransposons [3]. Dans le cas du Virus de l'Immunodéficience Humaine de type 1 (VIH-1), l'$ARNt_3^{Lys}$ est ainsi utilisé comme première amorce lors de la transcription inverse. La formation du complexe d'initiation de la transcription inverse du VIH-1 est un phénomène complexe nécessitant l'intervention de facteurs protéiques viraux, comme la protéine de nucléocapside (NC). Une partie de mon travail de thèse a consisté à étudier le rôle de la protéine de nucléocapside dans la formation du complexe d'initiation de la

transcription inverse du VIH-1. Cette étude sera présentée dans la PARTIE I.

Ainsi, même si les différentes parties de mon travail de thèse se rapportent à des sujets biologiques bien différents, elles s'articulent toutes autour de ces petites molécules que sont les ARN de transfert. Aussi avant d'entrer dans le vif du sujet propre à chacune des parties, nous avons cru bon d'effectuer en premier lieu une présentation générale des ARN de transfert, dans laquelle nous aborderons leur structure, leur fonction et leur biosynthèse. Cette présentation fera l'objet du CHAPITRE 1.

Introduction

Chapitre 1

Les acides ribonucléiques de transfert

Dès 1955, Francis Crick prédisait l'existence de molécules adaptatrices [4] faisant le lien entre le langage des acides nucléiques, porté par les ARNm, et la chaîne polypeptidique naissante au niveau du ribosome. Découverts seulement trois ans plus tard en 1958 par Mahlon Hoagland [5], les ARNt ont fait l'objet d'études acharnées et passionnées [6] jusqu'à la résolution au début des années 70 de la structure tridimensionnelle de l'ARNt$^{\text{Phe}}$ de levure par les équipes d'Aaron Klug et d'Alexander Rich, alors en intense compétition [7–10]. Actuellement, les ARNt font toujours l'objet d'études structurales et fonctionnelles, et le nombre croissant des séquences d'ARNt disponibles [1] [11] a de plus créé un engouement particulier pour l'étude phylogénétique de ces molécules [12–14].

1.1 Structure des ARNt

Structure primaire

Les ARN sont des polymères de ribonucléotides. Les ribonucléotides sont constitués de trois éléments reliés entre eux :

- un groupement phosphate lié à
- un sucre, le ribose, lui-même lié à
- une base azotée

Les nucléotides sont reliés entre eux par une liaison phosphodiester impliquant l'oxygène 3′ du ribose du nucléotide n et le phosphate en 5′ du nucléotide suivant n+1 (Figure 1.1). Les quatres bases canoniques constitutives des ARN sont deux purines, adénine (A) et guanine (G), et deux pyrimidines, cytosine (C) et uracile (U) (Figure 1.2).

1. La base de données « *tRNA Compilation* »[11] maintient les différents gènes et séquences d'ARNt connus, la dernière mise à jour datant tout de même de 2001, et permet d'effectuer des recherches croisées sur plusieurs critères tels que l'organisme et l'acide aminé correspondant à l'ARNt. Cette base de données est accessible à l'adresse suivante : `http://www.staff.uni-bayreuth.de/~btc914/search/`

Les ARNt sont des molécules composées de l'enchaînement de 74 à 98 nucléotides, avec une moyenne de 76. L'enchaînement des nucléotides est spécifique à chaque ARNt et définit sa structure primaire. C'est en 1965 que le premier ARNt a été entièrement séquencé par l'équipe de Robert Holley, ce qui lui a valu le Prix Nobel de médecine en 1968. Il s'agissait de l'ARNt alanine de levure [15].

Depuis une dizaine d'années, un système conventionnel est utilisé pour numéroter les nucléotides dans la séquence de l'ARNt — du numéro 1 pour celui situé à l'extrémité 5′ jusqu'au numéro 76 pour celui situé à l'extrémité 3′ [16]. Avec cette numérotation, les nucléotides très conservés parmi les différentes séquences portent le même numéro. Les trois nucléotides de l'anticodon portent ainsi les numéros 34, 35 et 36.

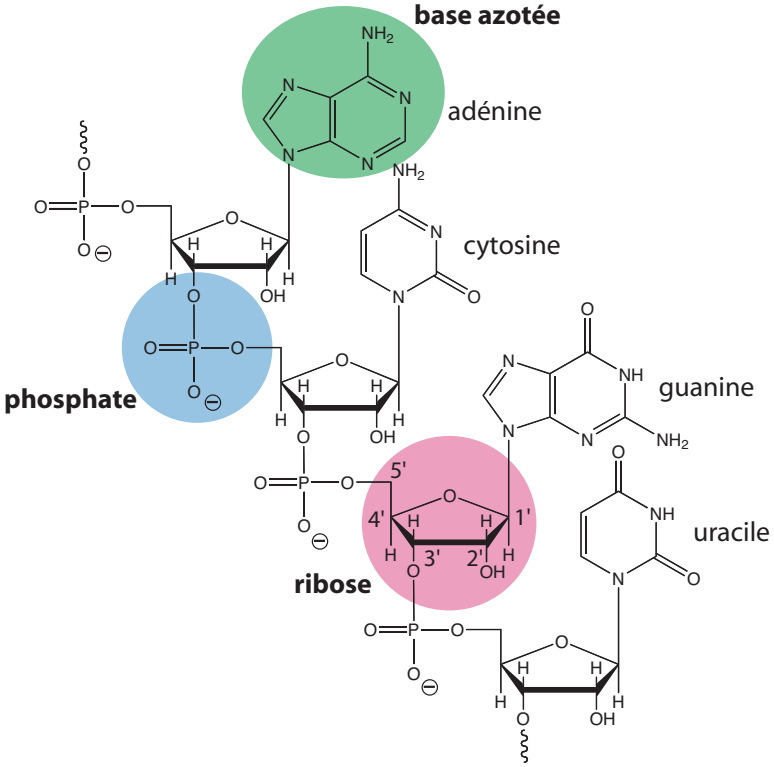


Figure 1.1 – Structure primaire des ARN
Les trois éléments constitutifs des nucléotides sont représentés sur fond de couleur. Le groupement phosphate (*en bleu*), le sucre ribose (*en rouge*) avec la numérotation des carbones du cycle (de 1′ à 5′), et une base azotée (*en vert*) reliée par une liaison N-glycosidique au ribose.

En plus des quatre bases canoniques, les ARNt sont également constitués de bases non canoniques. On parle de bases modifiées. Les modifications chimiques peuvent être très variées. Il peut s'agir par exemple d'une simple méthylation du ribose ou de modifications plus importantes, comme l'ajout d'un acide aminé sur une position réactive d'une base azotée. Les premières bases modifiées dans les ARN ont été détectées en 1957 [17], et identifiées par la

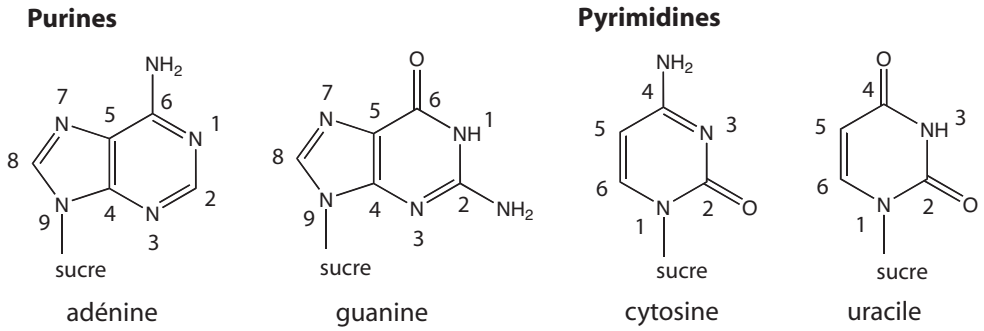


Figure 1.2 – Nomenclature des cycles puriques et pyrimidiques
Les bases adénine, guanine, cytosine et uracile sont rangées par famille (Purines et Pyrimidines).

suite [18, 19]. Il s'agissait entre autre de la pseudouridine (Ψ) et de la ribothymidine (T) des nucléotides rencontrés très fréquemment dans les ARNt. Ces modifications chimiques sont quasiment toujours réalisées par des enzymes de modification sur des nucléotides canoniques de la séquence de l'ARNt. Ils ne sont dans tous les cas pas incorporés directement lors de la synthèse de l'ARNt par transcription. On parle pour cela de modifications post-transcriptionnelles. La formation de ces nucléotides modifiés sera abordé dans le Paragraphe 1.3.

Structure secondaire

La chaîne nucléotidique de l'ARNt se replie et adopte une structure secondaire caractéristique, définie par une série d'appariements canoniques de type Watson-Crick entre bases complémentaires (A–U et G≡C), mais aussi non canoniques de type bancal ou « wobble » (par exemple G∘U) [2]. Cette structure secondaire très bien conservée parmi les ARNt est dite en « feuille de trèfle » (Figure 1.3). Cinq régions sont ainsi formées par cette structure secondaire et sont communément appelées : tige-boucle D, ou bras D, dénommé ainsi car elle contient souvent une ou plusieurs dihydrourines (D) — tige-boucle anticodon — région variable — tige-boucle T, ou bras T, dénommé ainsi car elle contient souvent une thymidine (T) en position 54 — et tige acceptrice, au bout de laquelle sont fixés les acides aminés.

La tige acceptrice se termine à l'extrémité 3′ par la séquence CCA, et c'est l'adénosine 76 qui porte l'acide aminé spécifique de l'ARNt, si celui-ci a été aminoacylé par son aminoacyl-ARNt synthétase (aaRS) (pour des articles de revue voir [21, 22]). L'ARNt et l'acide aminé sont liés par une fonction ester entre un groupement alcool du ribose de l'adénosine 76 et la fonction acide carboxylique portée par l'acide aminé. Selon la classe de l'aaRS (Classe I ou II), l'acide aminé est incorporé sur le groupement alcool 2′OH ou 3′OH de l'adénosine 76 [23, 24].

2. Tout au long du texte, les notations employées pour symboliser les appariements des bases, canoniques ou non, cherchent tant que faire se peu à satisfaire les conventions définies dans Leontis & Westhof [20]. Ces conventions sont présentées en annexe à la page 215.

La boucle anticodon et la boucle T ont une longueur constante de 7 nucléotides, alors que celle de la boucle D peut varier de 7 à 11 nucléotides. La région variable est celle qui présente la plus grande variation de longueur parmi les différents ARNt. Elle est généralement constituée de 4 ou 5 nucléotides [3]. Cependant, dans quelques cas — $ARNt^{Ser}$, $ARNt^{Leu}$ et $ARNt^{Tyr}$ de *Bacteria* — elle comprend entre 10 et 24 nucléotides [25].

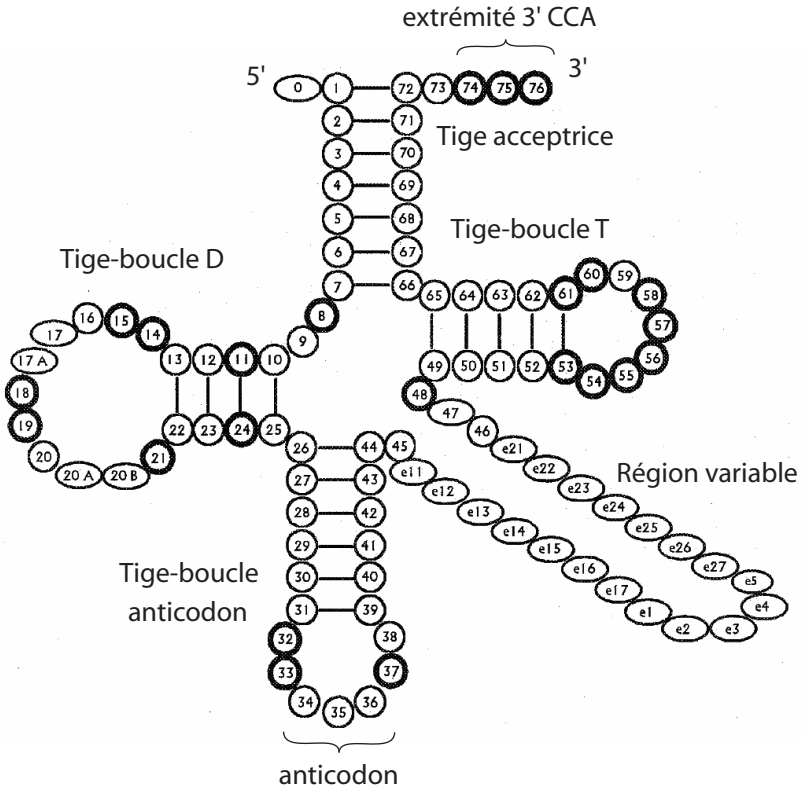


FIGURE 1.3 – Structure secondaire et numérotation des ARNt
Les 5 régions des ARNt sont nommées sur la FIGURE. Les cercles représentent des nucléotides toujours présents, et les ovales des nucléotides qui ne sont pas toujours présents dans les ARNt. Les positions auxquelles on retrouve des nucléotides invariants sont indiquées par des cercles à traits épais. D'après Sprinzl *et al.* [16].

La comparaison des séquences a permis l'identification de résidus conservés (U_8, A_{14}, G_{18}, G_{19}, A_{21}, U_{33}, G_{53}, T_{54}, Ψ_{55}, C_{56}, A_{58}, C_{61}, C_{74}, C_{75} et A_{76}) et semi-conservés [4] (Y_{11}, R_{15}, R_{24}, Y_{32}, R_{37}, Y_{48}, R_{57} et Y_{60}) dans presque tous les ARNt. Ceux-ci sont situés tout au long de la séquence, mais avec une forte concentration dans les boucles T et D (FIGURE 1.3). La structure tridimensionnelle de l'$ARNt^{Phe}$, a montré que la plupart de ces nucléotides conservés étaient impliqués dans des interactions tertiaires et participaient ainsi au repliement

3. La région variable peut en fait compter 3 ou 4 nucléotides si l'on n'y inclut pas la base 44, une purine très souvent impliquée dans une paire non canonique de type Watson-Crick avec une autre purine en position 26 (typiquement G_{26}–•–A_{44}) que l'on inclut de ce fait quelques fois dans la tige anticodon.

4. Y pour pyrimidine et R pour purine

caractéristique des ARNt [9, 10]. Il est intéressant de noter que certaines de ces interactions tertiaires avaient été proposées avant même la résolution de la structure de l'ARNt^{Phe}, comme la paire dite de Levitt G_{15}–○–C_{48} ou la paire *Trans*-Hoogsteen U_8○–□A_{14} [26].

Structure tridimensionnelle

La première structure d'ARNt, l'ARNt^{Phe} de levure, a été résolue par cristallographie dans les années 1973–1974 [7–10]. Cette structure est présentée à la FIGURE 1.4. Par la suite, l'ensemble des études cristallographiques menées sur les ARNt cytoplasmiques a montré que cette structure tridimensionnelle était très conservée.

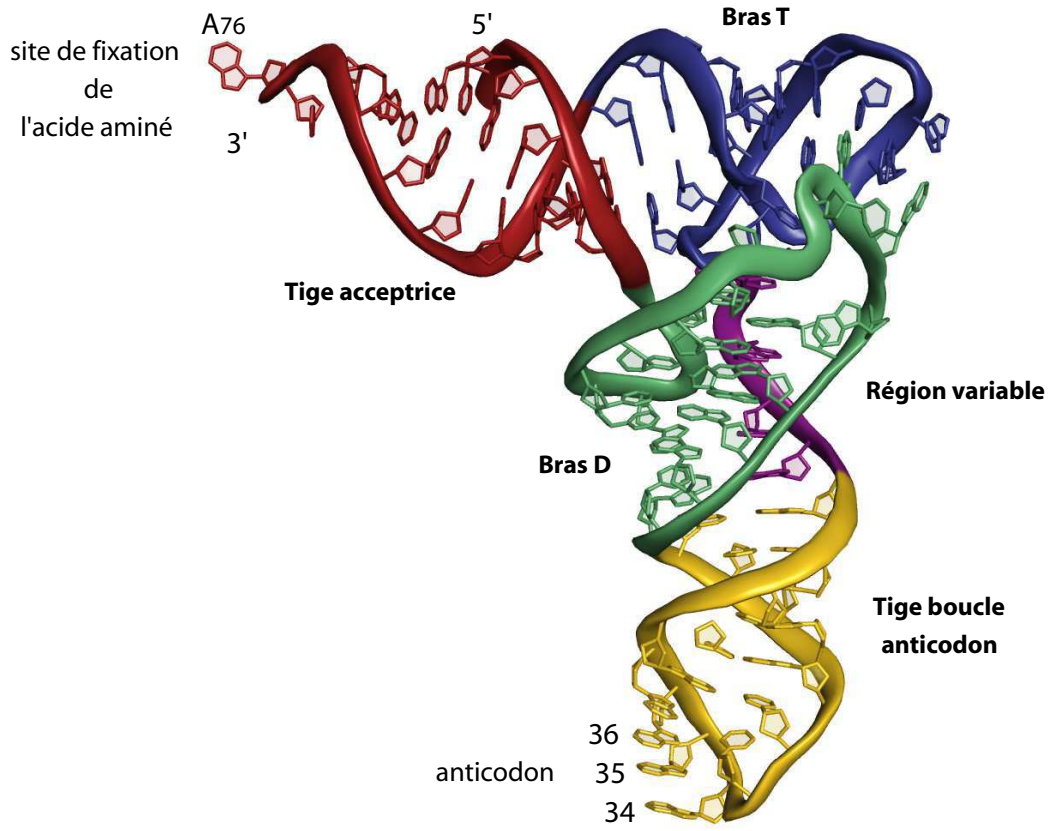


FIGURE 1.4 – Structure tridimensionnelle de l'ARNt^{Phe} de levure
Structure cristallographique de l'ARNt^{Phe} de levure (PDB code 1EHZ) [27]. Les 5 régions de l'ARNt sont indiquées avec des couleurs différentes. La tige acceptrice est *en rouge*, le bras D *en vert*, la tige-boucle anticodon *en jaune*, la région variable *en mauve* et le bras T *en bleu*.

La structure tridimensionnelle des ARNt présente une forme caractéristique dite en forme de « L » [7]. Cette forme en « L » résulte de l'empilement de la tige acceptrice et de la tige T — qui forment l'une des branches du « L » — l'autre branche étant formée par l'empilement de la tige D et de la tige

anticodon (FIGURE 1.5). Les deux hélices d'ARN ainsi formées sont orientées quasi perpendiculairement. L'angle du « L » est constitué par les boucles T et D, et plus particulièrement par la paire de base Watson-Crick G_{19}=C_{56}. Ainsi l'anticodon (positions 34, 35 et 36) et le site de fixation de l'acide aminé (A_{76}) se trouvent aux deux extrémités de la molécule (FIGURE 1.4).

Cette structure est maintenue principalement par des interactions de type liaisons hydrogènes entre différents nucléotides plus ou moins éloignés dans la séquence, et par des interactions d'empilement. L'ensemble de ces interactions est regroupé sous le terme d'interactions tertiaires [28, 29]. Les différentes interactions tertiaires par liaison hydrogène sont représentées à la FIGURE 1.5. Elles ont lieu principalement entre les nucléotides du bras D d'une part et les nucléotides de la région variable et de la boucle T d'autre part. On trouve ainsi la paire dite de Levitt G_{15}–○–C_{48}, la paire à une liaison hydrogène G_{18}○–▷Ψ_{55}, la paire Watson-Crick G_{19}=C_{56}, la paire de type Watson-Crick G_{26}–●–A_{44}, ainsi que les bases triples G_{45} ▶–■m^2G_{10}=C_{25}, m^7G_{46}○–□G_{22}=C_{13} et A_9–□–A_{23}–U_{12}. On trouve également les deux paires très conservées de type *Trans*-Hoogsteen T_{54}○–□m^1A_{58} et U_8○–□A_{14}.

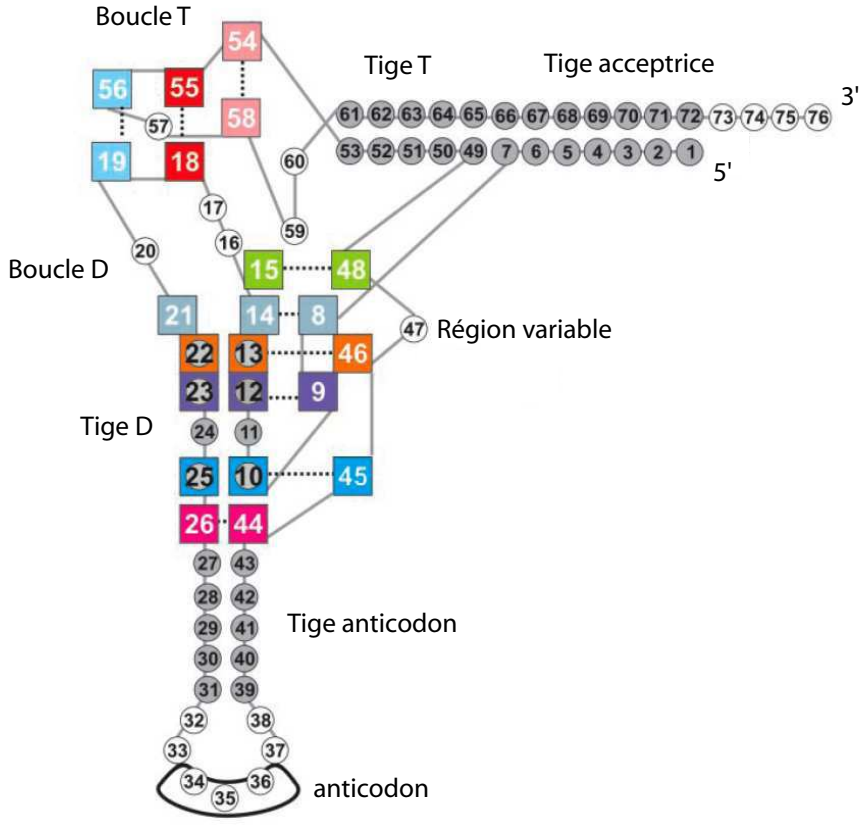


FIGURE 1.5 – Représentation schématique de la structure 3D d'un ARNt
Les nucléotides impliqués dans des interactions tertiaires sont dans des cadres colorés. Les interactions tertiaires par liaisons hydrogènes sont représentées par des pointillés. D'après Oliva *et al.* [29].

De plus, comme pour la plupart des ARN structurés [30, 31], le repliement des ARNt nécessite la présence d'ions positifs monovalents comme le sodium (Na^+) et divalents comme le magnésium (Mg^{2+}) [32–34]. Ainsi une structure récente à haute résolution (1.9 Å) de l'$ARNt^{Phe}$ [27] a révélé la présence de 9 sites de fixation de cations divalents au sein de la structure de l'ARNt. Ceux-ci participent au repliement de l'ARNt en servant d'écran aux charges négatives des groupements phosphates du squelette.

Comme nous l'avons déjà mentionné, les interactions tertiaires font intervenir des nucléotides invariants ou semi-invariants parmi les séquences d'ARNt (FIGURES 1.3 et 1.5). De plus, d'autres structures d'ARNt ont été résolues par cristallographie, à la fois sous forme libre — comme pour l'$ARNt^{Asp}$ de levure [35], l'$ARNt_i^{Met}$ initiateur de levure [36] et l'$ARNt_3^{Lys}$ humain [37] — et sous forme de complexes, comme pour l'$ARNt_f^{Met}$ initiateur de *E. coli* en complexe avec la formylase [38], les nombreuses structures des couples ARNt–synthétase (par exemple [39–45]), et les ARNt au sein des structures de ribosome [46, 47]. Ces structures ont toutes montré le même repliement global, avec néanmoins quelques petites variations essentiellement au niveau de la boucle D et de la région variable qui ne comportent pas toujours le même nombre de nucléotides. Ainsi, tous les ARNt adoptent une structure tridimensionnelle similaire en forme de « L ».

Cette structure globale conservée trouve bien évidemment son origine dans le fait que les différents ARNt de la cellule interagissent avec les mêmes partenaires, comme le ribosome ou les facteurs de traduction. Les petites variations locales, de structure ou de séquence, expliquent quant-à-elles les interactions spécifiques que chaque ARNt peut entretenir avec certains partenaires cellulaires, notamment son aminoacyl-ARNt synthétase. Les diverses fonctions, communes ou spécifiques, assurées par les ARNt, seront abordées dans le PARAGRAPHE suivant.

1.2 Fonction des ARNt

Fonction des ARNt dans la traduction

Au cours de la traduction, les ARNt assurent la lecture des codons de l'ARNm par l'intermédiaire de leur anticodon au niveau des sites A et P de la petite sous-unité du ribosome, et apportent l'acide aminé fixé à leur extrémité 3′ au niveau du centre peptidyl-transférase, situé dans la grande sous-unité du ribosome, où a lieu la synthèse de la chaîne polypeptidique (FIGURES 1.6 et 1.7) [48, 49].

Parmi les ARNt cellulaires, on distingue les ARNt dits élongateurs — qui vont apporter chaque nouvel acide aminé au niveau du centre peptidyl-transférase, en venant se lier au site A du ribosome pendant la phase d'élongation — et l'ARNt initiateur, qui va apporter la méthionine N-terminale de la protéine synthétisée, en venant se lier de façon très spécifique au site P du ribosome pendant la phase d'initiation de la traduction. Les mécanismes de

discrimination des ARNt initiateurs et élongateurs sont cruciaux pour assurer la fidélité de la traduction, notamment car cette discrimination conditionne « le choix » du codon initiateur, et donc du cadre de lecture. Ces mécanismes de discrimination, essentiels à l'étape d'initiation de la traduction seront abordés dans la PARTIE III.

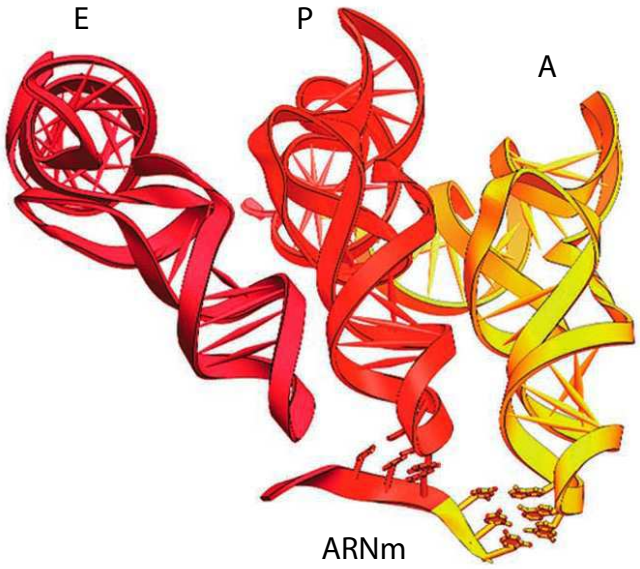


FIGURE 1.6 – ARNt fixés aux sites A, P et E du ribosome
L'ARNt au site A est *en jaune*, celui du site P *en orange* et l'ARNt du site E est *en rouge*. La figure montre l'orientation relative des trois ARNt et la courbure de l'ARNm entre les sites A et P. D'après Yusupov *et al.* [46].

Les ARNt sont ainsi des molécules adaptatrices faisant le lien physique entre l'information génétique codée sur l'ARNm par 61 codons sens et 3 codons stop, et une séquence protéique écrite avec les 20 acides aminés naturels. Les 64 codons sont le résultat des différentes combinaisons de mots de trois lettres écrits avec les quatre lettres A, U, G et C des acides nucléiques. Le code génétique donne la correspondance entre le codon de l'ARNm et l'acide aminé incorporé. Il est dit redondant, car plusieurs codons codent pour le même acide aminé (FIGURE 1.13). Il existe ainsi des ARNt avec des anticodons différents, mais qui portent le même acide aminé. On parle d'ARNt isoaccepteurs. De façon surprenante, le nombre d'ARNt avec des anticodons différents varie entre 30 et 46 selon les organismes, ce qui semble insuffisant pour reconnaître la totalité des 61 codons sens. La reconnaissance de l'ensemble des codons s'effectue en effet dans tous les domaines du vivant par une stratégie d'économie d'anticodons [12]. La reconnaissance de la dernière position du triplet du codon — par la première base de l'anticodon — peut en effet ne pas s'effectuer par l'intermédiaire d'appariements de type Watson-Crick, mais par des appariements non canoniques appelés appariements de type « wobble », car ils constituent en quelque sorte une irrégularité dans la règle classique d'appariement des bases [50, 51].

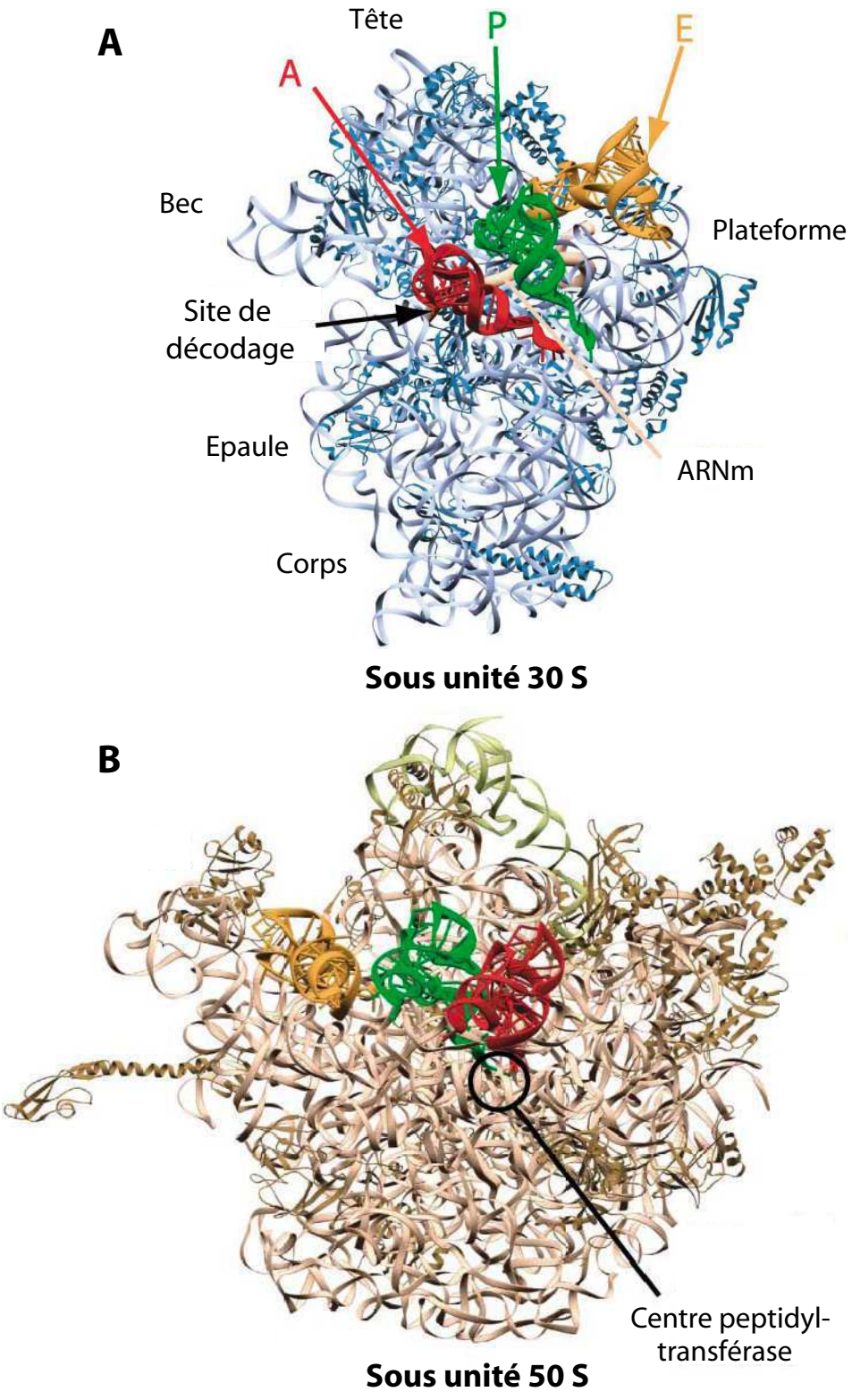


FIGURE 1.7 – Vue générale du ribosome bactérien
Représentation de la structure du ribosome bactérien 70S complexé à l'ARNm et à trois ARNt situés aux sites A, P et E [46]. Pour la clarté de l'image, les deux sous-unités ont été artificiellement séparées. **A** : Petite sous-unité du ribosome bactérien (30S). **B** : Grande sous-unité du ribosome bactérien (50S). La face visible de chaque sous-unité correspond à la face qui interagit avec l'autre sous-unité au sein du ribosome 70S. D'après Ramakrishnan [49].

Outre cette interaction codon–anticodon, les données cristallographiques récentes [46, 47] ont montré qu'environ 30 % des nucléotides des ARNt sont impliqués dans des contacts avec les ARNr ou les protéines ribosomiques. De plus, comme les ARNt doivent à la fois s'apparier avec leur anticodon au niveau du site de décodage et présenter leur tige acceptrice aminoacylée au centre peptidyl-transférase, la structure en « L » très conservée des ARNt est primordiale pour qu'ils assurent leur fonction principale de molécules adaptatrices lors de la traduction.

Outre cette implication majeure dans la traduction, certains ARNt interviennent également, de façon moins universelle, dans un certain nombre de mécanismes biologiques. Quelques uns d'entre eux seront évoqués dans le Paragraphe suivant.

Fonctions alternatives des ARNt

Les ARNt aminoacylés (aa–ARNt) peuvent avoir des rôles biologiques en dehors de la traduction. Ainsi l'acide glutamique chargé sur l'ARNt$^{\mathrm{Glu}}$ est modifié dans les chloroplastes, par la glutamyl-ARNt réductase, en 1-semialdéhyde glutamate, un précurseur de la voie de biosynthèse de diverses porphyrines [52], puis par transamination en acide 5-aminolevulinique, un composé intervenant dans la biosynthèse de la chlorophylle [53]. Certains aa–ARNt sont également impliqués dans la synthèse du peptidoglycane des bactéries à Gram positif [54].

Les ARNt non aminoacylés assurent également des fonction particulières au sein des cellules. Un ARNt$^{\mathrm{Arg}}$ tronqué des 9 nucléotides de son extrémité 3′ a par exemple été décrit comme guide permettant de cibler l'activité d'une ribonucléase au niveau de l'ARNr 18S murin [55]. Mais le rôle le plus connu des ARNt en dehors de la traduction est de servir de première amorce lors de l'initiation de la transcription inverse des rétrovirus et de certains rétrotransposons [3]. En effet, comme beaucoup d'ADN polymérases, les transcriptases inverses ont besoin d'une amorce venant s'apparier avec l'ARN matrice afin d'initier l'incorporation des désoxyribonucléotides complémentaires à cette matrice. Les rétroéléments détournent ainsi un ARNt cellulaire de sa fonction première afin de l'utiliser comme amorce de polymérisation. Ces rétroéléments présentent ainsi un site de liaison (PBS pour *Primer Binding Site*), complémentaire par interactions Watson-Crick avec une portion de l'ARNt amorce. Chez les rétrovirus il s'agit de l'extrémité 3′ de l'ARNt, alors que chez les rétrotransposons, il peut s'agir soit de l'extrémité 3′, soit d'une région interne de l'ARNt. Les ARNt les plus largement utilisés sont l'ARNt$^{\mathrm{Trp}}$, l'ARNt$^{\mathrm{Pro}}$, ARNt$_{\mathrm{i}}^{\mathrm{Met}}$, les ARNt$_{1,2}^{\mathrm{Lys}}$ et l'ARNt$_{3}^{\mathrm{Lys}}$. Ce dernier est utilisé comme amorce par la transcriptase inverse du virus de l'immunodéficience humaine de type 1 (VIH-1). Cette utilisation détournée de l'ARNt$_{3}^{\mathrm{Lys}}$ par le VIH-1 sera détaillée dans la Partie I.

En résumé, les ARNt ont un rôle central dans la traduction de l'information génétique. Ils assurent également d'autres fonctions, moins universelles, dans la biologie de la cellule. La séquence, la structure tridimensionnelle et

les modifications post-transcriptionnelles portées par les ARNt sont dans tous les cas essentielles à leur fonction. Elles résultent des étapes successives de la biosynthèse des ARNt.

1.3 Biosynthèse des ARNt

La biosynthèse des ARNt est un processus complexe qui se déroule en plusieurs étapes (pour des revues [56–58]). La première étape est la transcription du gène de l'ARNt par une ARN polymérase qui incorpore les quatre nucléotides canoniques (A, U, G et C) sur la matrice d'ADN. Pour transcrire l'ensemble de leur gènes, les *Bacteria* et les *Archaea* n'utilisent qu'une seule ARN polymérase, alors que chez les *Eukarya*, il existe trois polymérases différentes, chacune spécialisée dans l'expression de gènes qui lui sont spécifiques [5]. C'est l'ARN polymérase III qui transcrit l'ensemble des gènes d'ARNt eucaryotes [60], sous forme de précurseurs possédant des extensions 5′ et 3′ (FIGURE 1.8). Les ARNt précurseurs subissent alors une maturation de leurs extrémités, ce qui leur donne leur longueur standard. De nombreuses modifications sont également incorporées sur ce transcrit primaire afin de former l'ARNt mature. Celui-ci peut alors être aminoacylé par son aminoacyl-ARNt synthétase. Ces différentes étapes de maturation (FIGURE 1.8) seront détaillées dans les PARAGRAPHES suivants.

1.3.1 Maturation des extrémités

Maturation de l'extrémité 5′

L'excision de l'extension située à l'extrémité 5′ des ARNt est réalisée par une endonucléase, la ribonucléase P (RNase P) [61, 62]. Chez les *Bacteria* elle est composé d'une petite sous-unité protéique d'environ 14 kDa (RnpA) et d'une sous-unité ARN d'environ 350–400 nucléotides (RnpB). Bien que les deux sous-unités soient essentielles *in vivo*, une augmentation de la force ionique permet la catalyse *in vitro* par la sous-unité ARN seule [63]. La RNase P bactérienne est ainsi une enzyme ARN ou « ribozyme ». La sous-unité ARN est constituée d'un domaine de spécificité aux ARNt et d'un domaine catalytique [64]. Chez les *Eukarya* et les *Archaea*, la RNase P comprend un grand nombre de protéines, et la sous-unité ARN ne possède pas d'activité catalytique par elle-même. La RNase P est ainsi constituée de 4 à 10 sous-unités protéiques indispensables à la catalyse. En l'occurrence, les 9 protéines de la RNase P de levure sont essentielles à la survie de la cellule [65]. Même si les rôles des différentes sous-unités protéiques ne sont pas encore bien connus, elles interviennent probablement dans la stabilisation et la maturation de la sous-unité ARN, ainsi que dans la formation du site actif [66]. Le mécanisme catalytique de la coupure par la RNase P n'est pas encore élucidé.

5. Récemment une autre classe d'ARN polymérases, synthétisant les petits ARN interférents chez les plantes, a été identifiée et dénommée ARN polymérase IV [59].

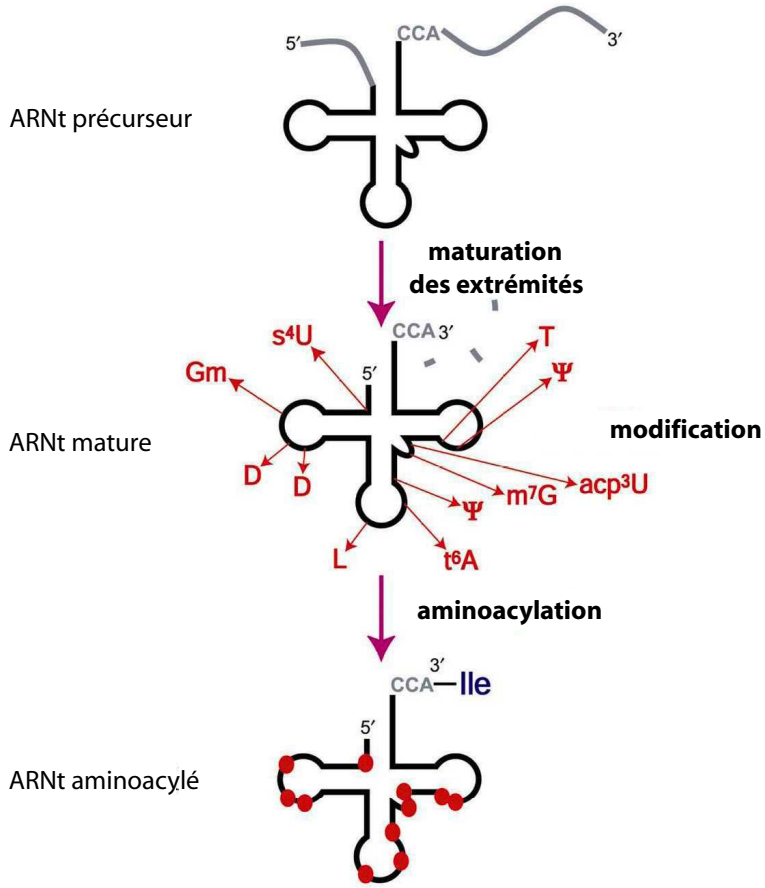


Figure 1.8 – Les étapes de la maturation des ARNt
La maturation des ARNt se décompose en deux étapes. L'étape de *maturation des extrémités* par diverses nucléases, qui donne la longueur native à l'ARNt. Et l'étape de *modification post-transcriptionnelle* des nucléotides, qui consiste en la modification chimique de nombreux nucléotides de la séquence. L'ARNt mature est ensuite aminoacylé par son aaRS. D'après Nakanishi & Nureki [57].

Maturation de l'extrémité 3′

Le clivage de l'extension en 3′ fait intervenir un grand nombre de nucléases agissant de concert [67, 68]. Des endonucléases comme la RNase E [69, 70] et la RNase III coupent l'extension 3′ en son milieu, puis des exonucléases comme la RNase II, la RNase BN′ et la RNase PH [71] réduisent la portion restante de l'extension 3′. Si les exonucléases raccourcissent l'extension 3′ au delà de la séquence CCA d'attache de l'acide aminé, une ARN polymérase matrice indépendante, appelée enzyme d'ajout du CCA (*CCA-adding enzyme*), permet de réparer la séquence CCA terminale. Il existe également une endonucléase, la RNase Z, qui clive l'extension 3′ juste au delà de la séquence CCA [72, 73]. La restauration de celle-ci se fait également grâce à l'enzyme d'ajout du CCA. La Figure 1.9 récapitule les différentes enzymes — endonucléases, exonucléases et polymérase — intervenant dans la maturation des extrémités des ARNt.

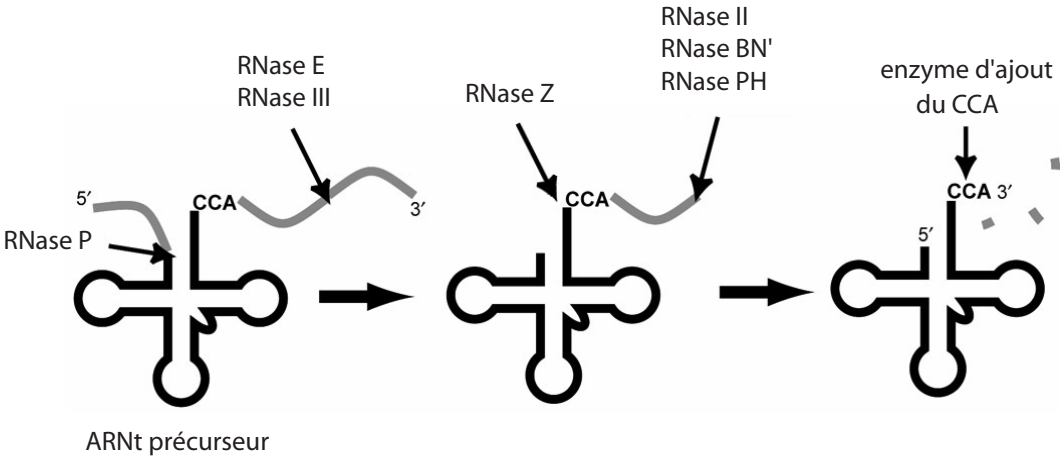


Figure 1.9 – Maturation des extrémités 5′ et 3′ des ARNt
L'endonucléase RNase P clive l'extension 5′. L'extension 3′ est excisée par l'action d'endonucléases (RNase E, RNase III et RNase Z) et d'exonucléases (RNase II, RNase BN′ et RNase PH). L'extrémité 3′ est éventuellement réparée par l'enzyme d'ajout du CCA. D'après Nakanishi & Nureki [57].

L'enzyme d'ajout du CCA est une nucléotidyltransférase unique, car elle étend une amorce spécifique (un ARNt) en y ajoutant une séquence bien précise (CCA), sans avoir besoin de matrice. Elle est aussi capable de contrôler l'état dans lequel se trouve l'extrémité CCA et de la réparer si besoin. Cette enzyme lie ses substrats NTP (ATP et CTP) dans la même poche de fixation [74, 75]. La polymérisation matrice indépendante semble s'expliquer par des changements de conformation de l'extrémité flexible CCA de l'ARNt et par l'utilisation d'une « matrice protéique », à la place de la matrice nucléotidique classique [76, 77].

Remarque : Chez les *Eukarya* et les *Archaea*, la maturation de certains précurseurs d'ARNt fait de plus intervenir l'épissage d'introns. Les introns sont localisés en 3′ de l'anticodon, entre les nucléotides 37 et 38. Leur excision fait intervenir une endonucléase, une ligase et une 2′ phosphatase [78].

Ainsi, de nombreuses enzymes interviennent dans la maturation des extrémités des ARNt. Elles produisent un ARNt de longueur contrôlée et possédant une extrémité 3′ de séquence bien définie. Dans le même temps, les nucléotides sont le siège de modifications chimiques post-transcriptionnelles. Ce dernier aspect de la biosynthèse des ARNt sera détaillé dans le Paragraphe suivant.

1.3.2 Modifications post-transcriptionnelles

Diversité des nucléotides modifiés dans les ARNt

A ce jour, 107 nucléotides modifiés ont été identifiés dans les différents type d'ARN [2, 79, 80]. C'est dans la famille des ARNt que l'on rencontre la plus grande diversité de modifications, avec 91 nucléotides modifiés différents [6] (Figure 1.10). Quelques rares modifications se retrouvent parmi tous les types d'ARN, mais la plupart sont spécifiques d'une famille – et notamment un grand nombre est spécifique des ARNt (Figure 1.10). Tous les ARNt contiennent des modifications chimiques, mais à des taux très variables. La diversité et les proportions de nucléotides modifiés rencontrés dans les différents type d'ARN sont regroupés dans la Table 1.1. On trouve ainsi en moyenne 10 nucléotides modifiés le long de la séquence des ARNt, ce qui correspond à environ 13 % de nucléotides modifiés. Mais cette proportion peut varier de 3 à 26 % selon l'ARNt considéré. Les ARNt forment ainsi la famille d'ARN la plus modifiée (Table 1.1).

Type d'ARN	ARNt 74–98 nt	sn(o)ARN 70–1200 nt	ARNr 1600–4000 nt	ARNm 100–3000 nt
Nombre de nt modifiés par ARN	3–20 nt (3–26 %)	0–13 nt (0–10 %)	65–200 nt (1–3 %)	0–5 nt ($<$ 1 %)
Diversité	91	11	31	13

Table 1.1 – Diversité et proportion de nucléotides modifiés pour les différents types d'ARN
D'après Rozenski *et al.* [2].

Si l'on regarde la distribution au sein des trois domaines du vivant des nucléotides modifiés des ARNt, on remarque que la plupart des modifications qu'ils portent sont spécifiques d'un règne (*Bacteria*, *Eukarya* et *Archaea*). Cependant, un petit nombre est au contraire commun aux trois domaines du vivant (Figure 1.11). Ces modifications communes remplissent certainement une fonction importante au sein des ARNt, pour avoir ainsi été conservées au cours de l'évolution.

Les modifications rencontrées dans les ARNt sont de nature extrêmement variée. Elles affectent les purines et les pyrimidines à la fois sur des positions endocycliques — en touchant ainsi un atome de carbone ou d'azote — et sur des positions exocycliques. Elle peuvent aussi affecter le ribose, et dans ce cas elles se trouvent toujours sur la position 2′OH [81].

6. La base de données « *The RNA Modification Database* »[79] maintient à jour la liste des différentes modifications connues, et permet d'effectuer des recherches croisées sur plusieurs critères, tels que le type d'ARN, le type de nucléotides, les domaines du vivant, etc. La base de données est accessible à l'adresse internet suivante : `http://library.med.utah.edu/RNAmods/`

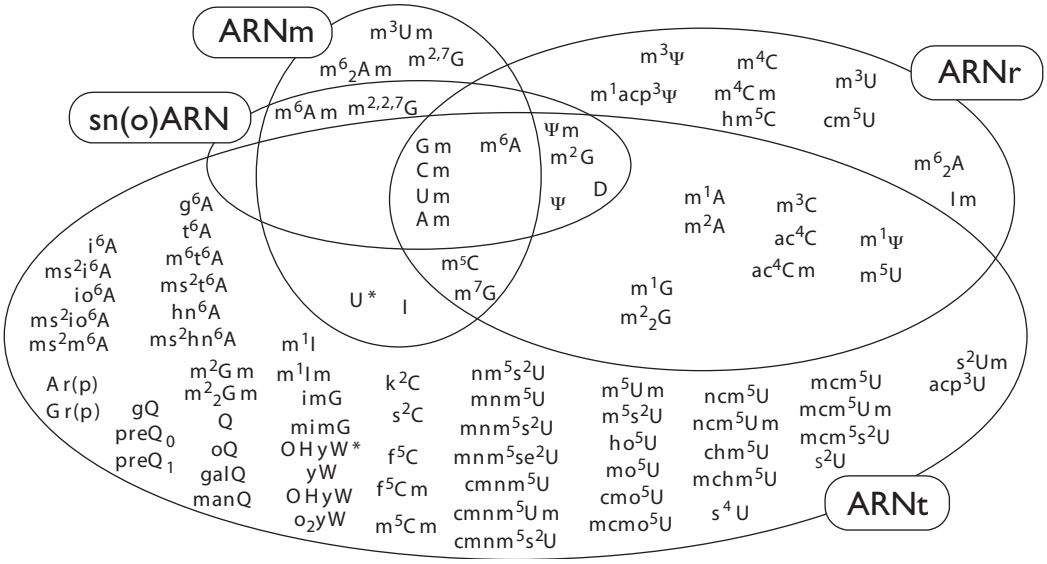


Figure 1.10 – Les nucléotides modifiés dans les différents type d'ARN
D'après Motorin & Grosjean [81].

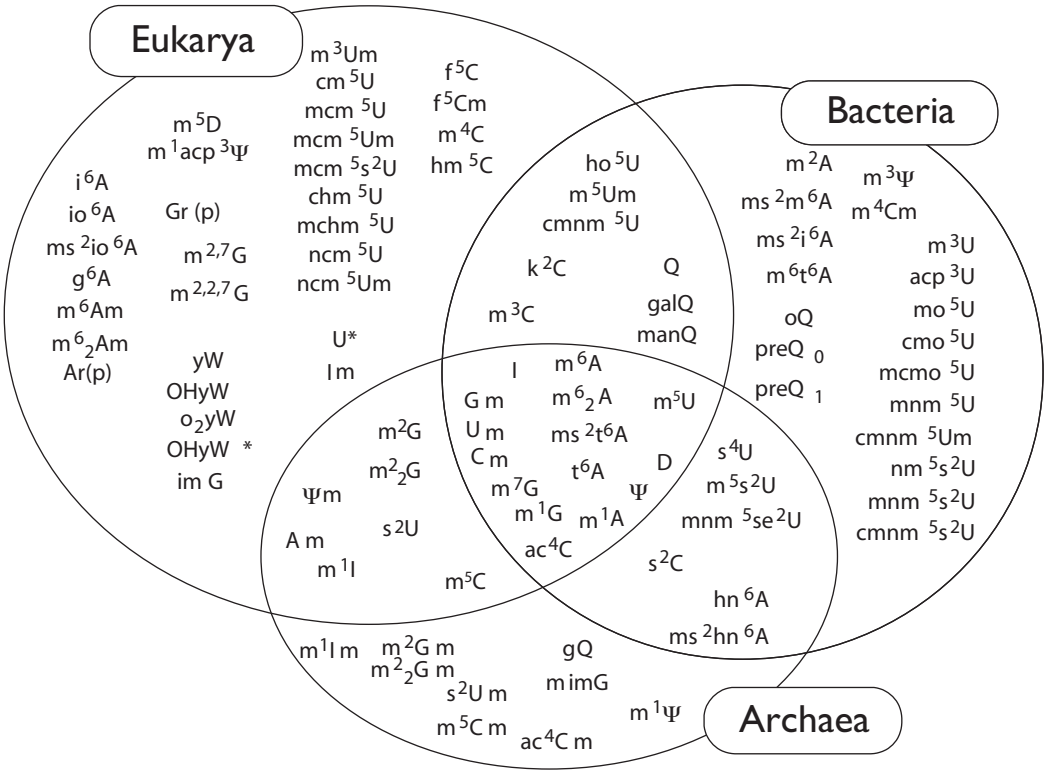


Figure 1.11 – Distribution phylogénétique des nucléotides modifiés dans les ARNt des trois domaines du vivant
U* correspond à une uridine formée par désamination enzymatique d'une cytidine.
D'après Motorin & Grosjean [81].

Ces modifications peuvent toutefois être regroupées en deux catégories suivant la complexité de la modification et notamment le nombre d'enzyme mis en jeu lors de sa synthèse [82]. On distingue d'une part les modifications dites simples, car elles ne requièrent qu'une seule réaction enzymatique. Ce sont les modifications les plus nombreuses. Il peut s'agir :

- d'une simple méthylation — par exemple la 5-méthyl-uridine m^5U ou T retrouvée en position 54 de la plupart des ARNt
- d'une isomérisation — comme pour la pseudouridine Ψ, fréquemment rencontrée en position 55
- d'une thiolation — par exemple la 4-thio-uridine s^4U, fréquemment rencontrée en position 8 chez les *Bacteria* et les *Archaea*
- d'une réduction — comme pour la dihydrouridine D, que l'on retrouve à différentes positions au sein de la boucle D
- etc.

On distingue d'autre part les modifications dites complexes, constituées de plusieurs modifications simples ayant lieu sur la même position ou sur des positions différentes d'un même nucléotide. Elles nécessitent l'intervention de plusieurs enzymes. On rencontre ainsi des groupements aussi complexes que :

- l'isopentényl (i)
- le thréonylcarbamoyl (t)
- l'acétyl (ac)
- le méthoxycarbonylméthyl (mcm)
- le carboxyméthylaminométhyl (cmnm)
- etc.

Ces modifications se retrouvent dans toutes les parties des ARNt, mais les différentes positions au sein de la séquence ne sont pas touchées dans les mêmes proportions. Ainsi certaines positions sont plus modifiées que d'autres, on parle de positions hypermodifiées [83].

Localisation des nucléotides modifiés dans les ARNt

La proportion de nucléotides modifiés rencontrés à chaque position de la séquence des ARNt est présentée sur la Figure 1.12. Les positions les plus modifiées se situent dans les boucles T et D, dans la région variable — positions 46, 47 & 48 — et dans la boucle anticodon — position 32, 34, 37 & 39. Il est intéressant de noter, en comparant les Figures 1.3 & 1.12, que beaucoup de positions hypermodifiées sont des positions auxquelles on retrouve des nucléotides invariants — par exemple les positions 32 et 37 de la boucle anticodon, 48 de la région variable et 54, 55 et 58 de la boucle T. C'est au niveau des positions 34 — position dite « wobble » — et 37 — celle qui suit directement la dernière base de l'anticodon — que l'on retrouve la plus grande diversité des modifications. De plus, c'est à ces positions que sont formées les modifications les plus complexes [84].

La conservation de ces modifications à des positions particulières est bien entendu à relier aux rôles que celles-ci jouent dans la structure et/ou la fonction des ARNt. Les principales fonctions connues des nucléotides modifiés des ARNt seront abordées dans le Paragraphe suivant.

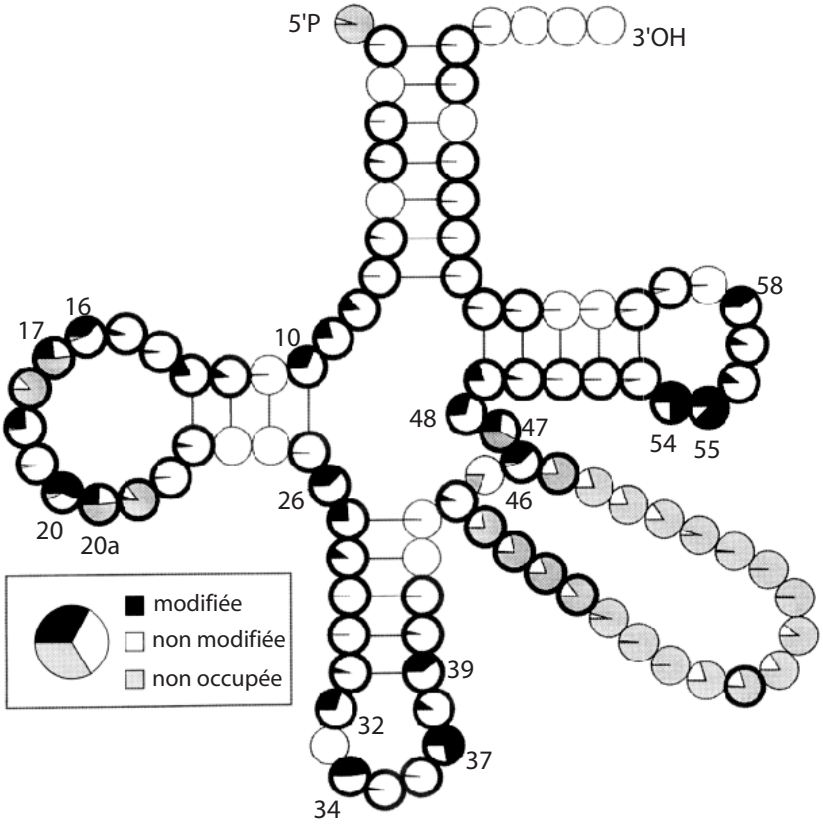


Figure 1.12 – Localisation des nucléotides modifiés dans les ARNt
Statistiques à partir des 546 séquences présentent dans la base de données de séquences d'ARNt en 1998 [16]. Les cercles en gras symbolisent les positions où l'on rencontre des nucléotides modifiés. Les positions numérotées correspondent à celle où l'on trouve plus de 25 % de nucléotides modifiés. On les considère arbitrairement comme des positions hypermodifiées. D'après Auffinger & Westhof [83].

Rôle des nucléotides modifiés dans les ARNt

Tous les ARNt séquencés contiennent des nucléotides modifiés, et certaines de ces modifications ont été conservées par l'évolution au sein des trois domaines du vivant, ce qui laisse à penser que celles-ci jouent des rôles essentiels vis-à-vis des diverses fonctions des ARNt [84]. Cependant, en dehors des modifications présentes dans la région de l'anticodon, les modifications post-transcriptionnelles ne semblent pas essentielles à la survie cellulaire, et l'échec de mise en évidence de phénotypes pour beaucoup de mutants de ces enzymes empêche la poursuite d'études fonctionnelles [56, 85]. Ainsi, les modifications post-transcriptionnelles semblent plutôt contribuer au réglage fin et synergique de l'activité des ARNt, qu'être les acteurs d'une fonction bien précise [86]. Cependant, le rôle de certaines d'entre-elles est maintenant bien établi et quelques uns d'entre eux seront succinctement abordés dans la suite du texte.

Nucléotides modifiés et interaction codon–anticodon Les positions 34 et 37 sont très fréquemment modifiées (FIGURE 1.12). Les positions 35 et 36, responsables de la lecture des deux premières lettre du codon ne sont quant-à-elle quasiment jamais modifiées.

Position 34 Le fait que le code génétique soit dégénéré, et que les deux première lettres du codon suffisent à définir l'acide aminé dans 8 cases sur 16 (FIGURE 1.13), a très tôt permis de proposer l'hypothèse d'une position « wobble » en position 34 de l'anticodon [50]. A cette position, les règles classiques d'appariement se relâchent et par exemple, les ARNt de *Bacteria* qui possèdent une uridine non modifiée en position 34 peuvent lire les codons se terminant par n'importe lequel des 4 nucléotides [12]. Ceci pose donc un problème lorsque la case du code génétique définie par les deux premières lettres du codon contient plusieurs acides aminés — ce qui arrive pour les 8 autres cases (FIGURE 1.13). Ce problème est résolu par l'utilisation adéquate de modifications à la position « wobble ». Ces modifications permettent de moduler les propriétés de reconnaissance de ce nucléotide, et ainsi de modifier les propriétés d'association codon–anticodon. Ces observations ont permis de réactualiser les règles d'appariement codon–anticodon [87–90]. On peut noter par exemple que :

- Les modifications des nucléotides U_{34} réduisent en général ses capacités de décodage aux seules purines [89], et permet ainsi le décodage fidèle des cases du code génétique codant pour deux acides aminés distincts, comme la case AAN [7] (Asn & Lys), où les uridines modifiées vont décoder les purines A et G et donc reconnaître les codons Lys (FIGURE 1.13). On trouve ainsi uniquement des uridines non modifiées en position 34 dans les anticodons d'ARNt codant pour des acides aminés seuls dans leur case.
- La présence d'une queuosine à la place d'une guanosine en position 34 favorise l'appariement avec une uridine plutôt qu'avec une cytosine [91].
- La désamination de l'adénosine 34 en inosine (I_{34}) est très fréquemment rencontrée [92]. Elle étend la capacité de décodage de l'adénine (principalement U) en formant des paires de bases avec les nucléotides U, C et A [51, 93].
- La lysidine C_{34} (k^2C_{34}) reconnaît une adénosine au lieu d'une guanosine, et facilite ainsi la bonne lecture du codon AUA (Ile) par l'$ARNt^{Ile}_{CAU}$ mineur de *E. coli* tout en empêchant la lecture erronée du codon AUG (Met) dans la case AUN (Ile & Met) du code génétique (FIGURE 1.13).

Position 37 Les nucléotides modifiés en position 37, bien que n'appartenant pas à l'anticodon, stabilisent l'appariement du nucléotide en position 36 avec le premier nucléotide du codon, et augmentent ainsi l'efficacité et la fidélité de la traduction [94, 95]. De plus, comme certaines modifications peuvent

7. où N signifie : « n'importe lequel des 4 nucléotides. »

empêcher l'appariement de cette position avec l'ARNm, elles contribuent au maintient de la bonne phase de lecture — en diminuant les risques de décalage du cadre de lecture, ou *frameshifting*. Ceci a notamment été montré pour le nucléotide modifié m^1G_{37} [96].

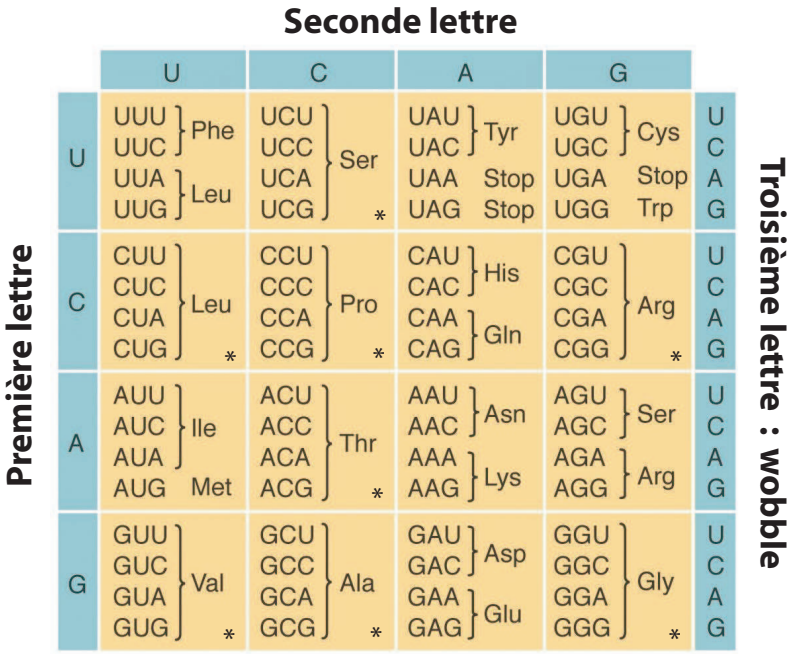


Figure 1.13 – Le code génétique
Les 61 codons sens et les 3 codons stop sont représentés dans un tableau dont chaque case est définie par la première et la deuxième lettre seulement. La troisième lettre définit les 4 triplets à l'intérieur de chaque case. Les 8 cases contenant un seul acide aminé sont estampillées d'une étoile *.

Nucléotides modifiés et structure des ARNt Différentes techniques ont été utilisées pour étudier le rôle des nucléotides modifiés dans la structure des ARNt, comme la dénaturation thermique [32, 97, 98], le marquage chimique et enzymatique [99–101], la cristallographie [102] et la RMN [103–106]. Ces études ont montré que les ARNt « nus », c'est-à-dire dépourvus de modifications post-transcriptionnelles, adoptent une structure proche des ARNt naturels, mais présentent des différences significatives, notamment au niveau local — en ce qui concerne les interactions avec les molécules du solvant, cations et molécules d'eau, et les conformations particulières que peuvent induire ces modifications — mais aussi au niveau des interactions tertiaires participant au repliement global des ARNt.

Solvant Les nucléotides modifiés participent à la fixation du magnésium au sein de l'ARNt et affectent l'organisation des molécules d'eau de la première couche de solvatation. En effet, comme nous l'avons déjà évoqué au Paragraphe 1.1, les ARNt ont besoin d'ions positifs — notamment Mg^{2+}

— pour adopter un repliement correct. Des études RMN ont montré que des transcrits nus adoptent une structure différente des ARNt naturels, en présence de faibles concentrations de Mg^{2+}, et que celle-ci était restaurée à de plus fortes concentrations en Mg^{2+} [103, 104]. Ainsi les nucléotides modifiés sont impliqués dans la fixation spécifique d'ions Mg^{2+}, ce qui abaisse la concentration minimale de cations nécessaire au bon repliement des ARNt [107, 108].

Les nucléotides modifiés peuvent également interagir différemment avec les molécules d'eau. Ceci a notamment été beaucoup étudié pour la pseudouridine Ψ. La fixation spécifique d'une molécule d'eau grâce à la fonction imino supplémentaire — comparé à l'uridine — permettrait de rigidifier localement la structure des ARNt [102, 109, 110].

Effets conformationnels et thermodynamiques Certaines modifications post-transcriptionnelles changent radicalement les conformations privilégiées et certaines propriétés physico-chimiques des nucléotides. Il s'agit principalement d'effets stéréoélectroniques jouant sur la conformation du ribose, d'effets d'empilement des bases et d'effets électrostatiques [110]. On peut noter par exemple que :

- La méthylation du ribose sur le groupement 2′OH favorise une conformation C3′-endo, à cause d'un effet de répulsion stérique entre le groupe méthyle et la base. Cette conformation confère une stabilité thermodynamique supplémentaire aux nucléotides présentant cette modification. La grande occurrence de cette modification chez les organismes hyperthermophiles indique combien celle-ci peut être importante pour la stabilité globale des ARNt.
- Les thiolations en position 2 des uridines (s^2U) ont globalement les mêmes effets que les méthylations du ribose, en cela qu'elles favorisent également une conformation C3′-endo.
- Les méthylations en position 5 des pyrimidines (m^5U et m^5C) augmentent la stabilité des ARNt en favorisant les interactions d'empilement des bases. Cette modification ne semble pas modifier la conformation privilégiée du ribose, et ainsi la stabilité accrue est directement imputable aux effets d'empilement.
- La réduction de l'uridine en dihydrouridine (D) abolit la planarité du cycle ce qui lui procure de très mauvaises propriétés d'empilement. De plus, cette modification favorise une conformation C2′-endo. Cette modification semble ainsi apporter une flexibilité supplémentaire. Ceci est à mettre en relation avec le fait que cette modification est retrouvée plus fréquemment chez les organismes psychrophiles, vivant à basse température.
- Certaines modifications (s^2U, s^4U, m^5C, etc.) peuvent avoir une influence sur le pK_a de certaines positions des bases — notamment le N_3 des pyrimidines — et ainsi influer sur la stabilité de formes protonées impliquées dans des appariements non-canoniques.

– Certaines méthylations (m^1A m^3C et m^7G) introduisent une charge positive au niveau de la base et peuvent ainsi induire des effets électrostatiques forts. En effet, dans le contexte d'un ARN replié, la répulsion électrostatique des groupements phosphate du squelette ne favorise pas l'obtention d'une structure compacte. Le rapprochement de différentes charges négatives peut s'effectuer grâce à l'écrantage de celles-ci par des ions positifs du solvant, comme nous l'avons déjà évoqué, mais également par les rares charges positives présentes au sein de la structure de l'ARNt. Il a par exemple été montré qu'une m^1A pouvait jouer un rôle dans la croissance à haute température de certains organismes thermophiles [111].

Incidence sur le repliement global Les nombreuses modifications rencontrées au niveau des bras T et D ainsi qu'au niveau de la région variable augmentent considérablement la stabilité des interactions tertiaires assurant le repliement en « L » des ARNt [101]. Certaines de ces modifications sont même très conservées, comme la m^5U_{54}, la Ψ_{55} et la m^1A_{58}. Certains nucléotides méthylés — comme les m^5C, m^5U, $m^1\Psi$, m^1G, m^7G, etc. — contribuent à augmenter les interactions d'empilement, et donc à stabiliser le cœur de la structure de l'ARNt [108, 110].

Certaines modifications peuvent induire des changements conformationnels importants, en cela qu'elles peuvent par exemple bloquer la face d'interaction Watson-Crick d'un nucléotide, et empêcher ainsi qu'il prenne part à des appariements canoniques. Certaines modifications assurent donc au sein des ARNt une structure secondaire correcte en « feuille de trèfle », et participent donc à l'intégrité de leur structure native. Ceci a notamment été montré pour l'$ARNt^{Asn}$ humain et l'$ARNt^{Lys}$ mitochondrial humain ainsi que pour diverses tiges-boucles anticodon [112].

Organisation du manuscrit et présentation des résultats

Après cette introduction générale sur les ARN de transfert, qui a permis de présenter leur structure, leur fonction et les étapes successives de leur biosynthèse, nous allons pouvoir présenter les sujets qui ont constitué le travail de ma thèse. Ceux-ci ont tous un rapport avec les ARNt, mais concernent des sujets biologiques éloignés et seront pour cela présentés dans trois parties indépendantes.

Partie I

La Partie I abordera la formation du complexe d'initiation de la transcription inverse du VIH-1, dans lequel intervient l'$ARNt_3^{Lys}$ humain. La formation du complexe d'initiation de la transcription inverse fait aussi intervenir une protéine virale : la protéine de nucléocapside (NC). Le laboratoire s'intéresse depuis de nombreuses années à la détermination du mécanisme de formation du complexe d'initiation et notamment au rôle joué par la protéine de nucléocapside [113–117]. Au cours de cette thèse, j'ai affiné le modèle de mécanisme proposé par le laboratoire quant à la formation du complexe d'initiation. J'ai en effet tout d'abord identifié, par des techniques de fluorescence, un site de fixation fort et spécifique de la NC sur le bras D de l'$ARNt_3^{Lys}$. Dans le contexte de la formation du complexe d'initiation, ceci explique l'implication précise de la NC quant à l'ouverture de la structure tridimensionnelle de l'$ARNt_3^{Lys}$. Cette étude a donné lieu à l'Article 1 présenté à la page 53.

J'ai par la suite entrepris de résoudre la structure du complexe NC/bras D de l'$ARNt_3^{Lys}$ par RMN. J'ai cependant été confronté à un problème inhérent à l'étude de complexes par RMN en solution, le phénomène d'échange chimique. Le phénomène d'échange rencontré dans le cadre de l'étude de ce complexe s'est trouvé être dans la gamme de temps la plus dommageable pour réaliser une étude structurale. Le complexe NC/bras D se trouve en effet en échange intermédiaire à l'échelle de temps spectrale avec les formes non associées NC libre et bras D libre. Ceci conduit à une perte de signal très importante et à un élargissement des raies tel que celles-ci peuvent devenir inobservables. Nous n'avons, pour cette raison, pas pu déterminer la structure du complexe entre la protéine de nucléocapside et le bras D de l'$ARNt_3^{Lys}$. Cependant, l'amélioration de la qualité des spectres ainsi que l'étude approfondie d'expériences de titration par RMN, m'a permis de mettre en évidence que le mode de recon-

naissance du bras D de l'$ARNt_3^{Lys}$ par la NC devait être identique à celui des signaux d'encapsidation du génome viral SL3 et SL2, mode de fixation révélé par la structure des complexes NC/SL3 et NC/SL2 [118, 119]. Ceci a donné lieu à l'Article 2 présenté à la page 68.

Partie II

La Partie II consiste en l'étude d'une enzyme de modification des ARNt, la m^1A_{58} méthyltransférase de *T. thermophilus* ou TrmI. Cette enzyme ajoute un groupement méthyle sur la position 1 du cycle purine de l'adénosine 58 située dans la boucle TΨC des ARNt. Au cours de cette thèse, j'ai entrepris de résoudre la structure de cette enzyme en complexe avec son substrat ARNt par cristallographie. Cependant, malgré de nombreux essais de co-cristallisation, avec des partenaires ARN variés, je n'ai pas obtenu de co-cristaux et n'ai donc pas résolu la structure du complexe. J'ai cependant obtenu des cristaux de la protéine TrmI seule. Ceci m'a permis de résoudre la structure de la protéine par remplacement moléculaire, à partir de l'homologue de TrmI chez *M. tuberculosis*, et de l'affiner jusqu'à 1.7 Å de résolution. J'ai également identifié des résidus de l'enzyme potentiellement impliqués dans la catalyse de transfert de méthyle, et j'ai vérifié cette prédiction en comparant les paramètres catalytiques de variants de l'enzyme affectés au niveau de ces résidus à ceux de l'enzyme TrmI sauvage. Nous avons aussi montré que la protéine tétramérique TrmI, qui présente quatre sites actifs distincts, ne peut fixer que deux substrats ARNt en même temps et nous avons pu construire manuellement un modèle d'amarrage moléculaire de l'ARNt qui rend compte de ce résultat. Cette étude a donné lieu à l'Article 3 présenté à la page 113.

Partie III

La Partie III traite quant-à-elle de la discrimination de l'$ARNt_i^{Met}$ initiateur vis-à-vis des ARNt élongateurs lors de l'initiation de la traduction. L'origine moléculaire de cette discrimination a été étudiée depuis longtemps. Il ressort de ces études que la capacité des ARNt initiateurs à initier la synthèse protéique en venant s'apparier au site P du ribosome est corrélée à deux particularités de leur tige-boucle anticodon :

- La tige-boucle anticodon des ARNt initiateurs adopte une conformation particulière [120], révélée par des études de clivage aux endonucléases et de sondes chimiques en solution [121].
- Cette conformation particulière est liée à l'existence de trois paires de bases GC dans la tige anticodon, qui sont caractéristiques des ARNt initiateurs [122, 123].

Cependant, jusqu'à maintenant, aucune donnée structurale ne venait appuyer ses observations.

Lors des essais de co-cristallisation de TrmI avec l'ARNt$_f^{Met}$ initiateur de *E. coli*, j'ai obtenu des cristaux de l'ARNt$_f^{Met}$ seul, qui m'ont permis de résoudre la structure de cet ARNt et de l'affiner jusqu'à une résolution de 3.1 Å. Cette structure a révélé une conformation unique de la tige-boucle anticodon. Cette conformation unique implique les trois paires de bases GC de la tige et est parfaitement en accord avec les études réalisées par sonde chimique en solution. Dans cette structure, la base A_{37}, située juste après le triplet de l'anticodon, ne vient pas s'empiler entre les bases 36 et 38, comme dans toutes les structures d'ARNt élongateurs résolues à ce jour, mais est appariée en base triple dans le grand sillon de la tige anticodon avec la paire G_{29}=C_{41}. Cette structure unique permet d'expliquer au niveau structural la discrimination entre les ARNt initiateurs et élongateurs. Elle a donné lieu à l'Article 4 qui est présenté à la page 155.

Première partie

Formation du complexe d'initiation de la transcription inverse du VIH-1

Chapitre 2

Contexte biologique

Les rétrovirus tels que le virus de l'immunodéficience humaine de type 1 (VIH-1) portent leur information génétique sur deux molécules homologues d'ARN simple brin. Ces molécules d'ARN sont rétro-transcrites en ADN proviral double brin lors de l'infection d'une nouvelle cellule par le virus. La transcription inverse fait partie de la phase pré-intégrative du cycle viral qui se termine éventuellement par l'intégration de l'ADN proviral dans l'ADN de la cellule hôte. Bien que certaines protéines auxiliaires puissent affecter la transcription inverse [124–129], celle-ci fait intervenir quatre acteurs principaux dont nous soulignerons quelques caractéristiques biochimiques et structurales utiles à la compréhension de la transcription inverse.

2.1 Les acteurs

2.1.1 La transcriptase inverse (RT)

La RT du VIH-1 est comme toute les transcriptases inverses, une polymérase capable de catalyser la polymérisation d'ADN à la fois sur une matrice d'ARN et sur une matrice d'ADN. Elle possède de plus un domaine RNase H lui permettant de dégrader le brin d'ARN venant de lui servir de matrice lors de la synthèse d'un hybride ADN/ARN. Comme de nombreuses polymérases, elle a besoin d'une amorce pour commencer la polymérisation des désoxyribonucléotides sur une extrémité 3′ OH libre. Celle-ci, dans le cas du VIH-1, est l'extrémité 3′ OH libre d'un ARN de transfert cellulaire, l'ARNt$_3^{\mathrm{Lys}}$. L'extrémité 3′ de cet ARNt est en effet complémentaire sur 18 nucléotides d'une région de l'ARN viral que l'on nomme PBS pour *Primer Binding Site*.

Organisation structurale

La transcriptase inverse est un hétérodimère formé d'une sous-unité de 560 résidus (p66) et d'une sous-unité de 440 résidus (p51) [131]. Ces deux sous-unités proviennent toutes deux de la polyprotéine Pol. Chaque sous-unité contient un domaine polymérase composé de quatre sous-domaines appelés « doigts », « paume », « pouce » et sous-domaine de connexion. La sous-unité

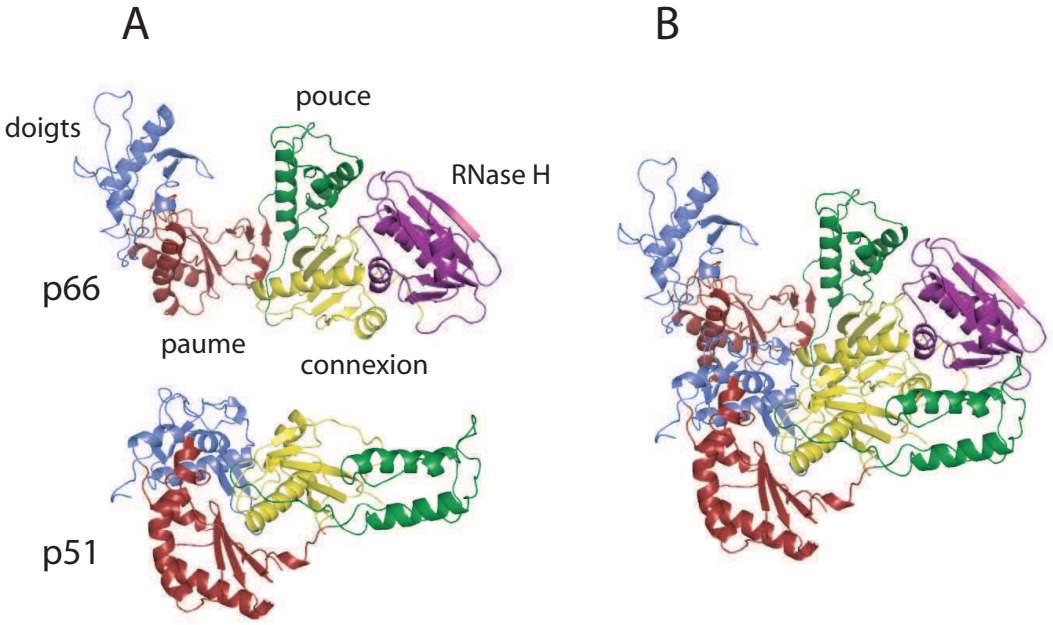


FIGURE 2.1 – Organisation des sous-domaines de la RT du VIH-1
A : Les deux sous-unités p66 et p51 ont été artificiellement séparées pour plus de clarté. Les sous-domaines de p66 s'organisent comme une main droite dont on regarderait la paume. *En bleu*, le sous-domaine « doigts ». *En rouge*, le sous-domaine « paume ». *En vert*, le sous-domaine « pouce ». *En jaune*, le sous-domaine de connexion. *En violet*, le domaine RNase H. Celui-ci est absent de la sous-unité p51. **B** : Structure complète de la RT, organisation relative des deux sous-unités p66 et p51. Cette structure, dans laquelle la main est ouverte, correspond à une structure en complexe avec un ADN double brin (PDB code 1RTD) [130].

p66 possède en plus un domaine RNase H (FIGURE 2.1). La sous-unité p51 résulte du clivage de ce domaine RNase H par la protéase rétrovirale [132]. De nombreuses structures de la RT libre [133, 134], complexée à des inhibiteurs [135] ou à de l'ADN double brin [130, 136] ont été résolues par cristallographie. Ces études structurales combinées à des études biochimiques donnent une compréhension approfondie du fonctionnement de la RT [137, 138], des mécanismes d'action des inhibiteurs [134, 139, 140], et de l'apparition des résistances vis-à-vis de ces inhibiteurs [130, 139].

La RT, cible des traitements antirétroviraux

La transcriptase inverse du VIH-1 est actuellement une des cibles principales des antirétroviraux ayant eu une autorisation de mise sur le marché. Les antirétroviraux ciblant la RT appartiennent à deux classes distinctes : les inhibiteurs nucléotidiques et nucléosidiques, qui sont incorporés dans le brin d'ADN nouvellement formé, mais qui empêchent l'incorporation du nucléotide suivant, et les inhibiteurs non nucléosidiques, qui inhibent la RT sans être incorporés dans l'ADN nouvellement formé. Dans le cadre de la tri-thérapie,

deux inhibiteurs nucléosidiques sont typiquement associés soit à un inhibiteur non-nucléosidique, soit à un inhibiteur de la protéase.

2.1.2 L'ARN viral

L'ARN viral, qui constitue le génome du virus, est constitué de deux molécules d'ARN de polarité positive [141], c'est-à-dire de même polarité que l'ARN messager traduit. Le génome du VIH-1 est dit complexe puisqu'il code pour les protéines structurales, enzymatiques et de l'enveloppe au travers des gènes *gag*, *pol* et *env*, comme pour tous les rétrovirus, auxquelles viennent s'ajouter des protéines dites régulatrices (Tat, Rev, Vpu, Nef, Vif et Vpr) [142]. La FIGURE 2.2 présente l'organisation des gènes sur l'ARN génomique.

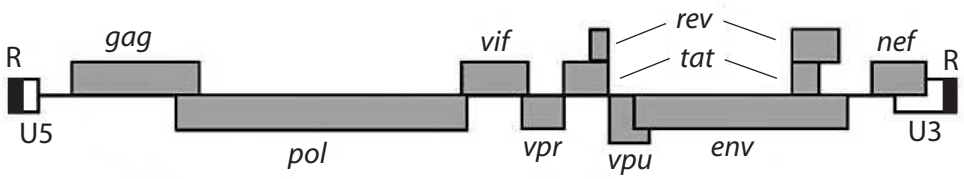


FIGURE 2.2 – Organisation des différents gènes portés par l'ARN génomique du VIH-1
Localisation des trois gènes principaux, *gag*, *pol* et *env* ainsi que des gènes codant pour les protéines dites régulatrices sur l'ARN viral.

Dans cette partie, seules les régions de l'ARN viral intervenant dans le processus de la transcription inverse seront décrites. Celles-ci sont représentées sur la FIGURE 2.3.

La première de ces régions est la région R (pour *Redundant*) qui s'étend sur une centaine de nucléotides. Elle est présente en deux copies identiques de même polarité aux extrémités 5′ et 3′ de l'ARN viral. Cette redondance permet de réaliser le premier saut de brin de la transcription inverse (voir PARAGRAPHE 2.2). Cette région R contient deux structures secondaires en tige-boucle (FIGURE 2.4) [143]. La séquence TAR (pour *Tat Responsive Element*), une séquence de 57 nucléotides constituant l'extrémité 5′ de l'ARN viral qui intervient dans la régulation de l'expression de l'ADN proviral par l'intermédiaire de la protéine Tat [144]. Et le signal de polyadénylation AAUAAA, uniquement fonctionnel dans la région R de l'extrémité 3′.

La région U5, est située directement en aval de R, du côté 5′ de l'ARN viral et directement en amont du site d'initiation de la transcription inverse (PBS) . De part sa position, U5 est la première partie du génome à être rétro-transcrite. Cette région contient aussi la séquence *att* intervenant dans l'intégration de l'ADN proviral dans l'ADN de la cellule hôte.

Figure 2.3 – Organisation schématique de l'ARN génomique du VIH-1
Organisation générale des séquences intervenant au cours du cycle réplicatif du VIH-1. R : séquence redondante, U5 : séquence unique en 5′, PBS : *Primer Binding Site*, 3′ PPT : PolyPurine Tract localisée en amont de la séquence U3, cPPT : central PolyPurine Tract, U3 : séquence unique en 3′.

Vient ensuite la séquence PBS, longue de 18 nucléotides. Cette séquence est strictement complémentaire des 18 nucléotides de l'extrémité 3′ de l'ARNt$_3^{Lys}$ (G_{59} à A_{76}) qui sert de première amorce à la transcriptase inverse [145].

Les séquences PPT (PolyPurine Tracts : 3′ PPT et cPPT) sont des séquences riches en purines, en partie résistantes à l'activité RNase H de la RT. Elles servent d'amorces à la synthèse du brin d'ADNc(+) (Paragraphe 2.2).

La région U3, est située directement en amont de R, du côté 3′ de l'ARN viral. De même que la séquence U5, elle contient la séquence *att* du site d'intégration. Elle contient aussi des éléments de régulation de la transcription par des facteurs cellulaires.

La partie 5′ de l'ARN viral est une région très structurée qui contient de nombreux éléments nécessaires au cycle du virus [146]. Ces différents éléments de structure sont représentés Figure 2.4. Les tiges-boucles TAR et Poly(A) font partie de la région redondante R. Le PBS quant à lui termine la région U5. Il est précédé d'un signal d'activation de la transcription inverse, la séquence PAS (*Primer Activation Signal*). Ceci sera abordé lors du Paragraphe 2.3. Le PBS est suivi de quatre tiges-boucles (stem-loop) : SL1, qui contient le site d'initiation de la dimérisation de l'ARN viral [147–149] ; SL2 qui contient le site donneur d'épissage (SD) ; SL3 qui est le déterminant majeur du site d'encapsidation de l'ARN viral (Ψ) [148, 149] et SL4 qui contient le codon AUG d'initiation de la traduction.

2.1.3 L'amorce ARNt$_3^{Lys}$

Les cellules de mammifères contiennent trois types majeurs d'ARNt^{Lys} isoaccepteurs : l'ARNt$_1^{Lys}$, l'ARNt$_2^{Lys}$ et l'ARNt$_3^{Lys}$ [150]. Alors que les ARNt$_1^{Lys}$ et ARNt$_2^{Lys}$ ne diffèrent l'un de l'autre que par un nucléotide dans la tige anticodon, la séquence de l'ARNt$_3^{Lys}$ et les modifications post-transcriptionnelles qu'il porte le rendent bien différent des deux autres. Néanmoins, malgré ces différences, les trois types d'ARNt^{Lys} sont incorporés dans les virions [151, 152], avec une prédominance d'ARNt$_3^{Lys}$ [152]. Ceci s'explique par le fait que les ARNt^{Lys} isoaccepteurs sont encapsidés sélectivement dans les virions par l'intermédiaire de leur aminoacyl-ARNt synthétase (LysRS) [153–155] celle-ci

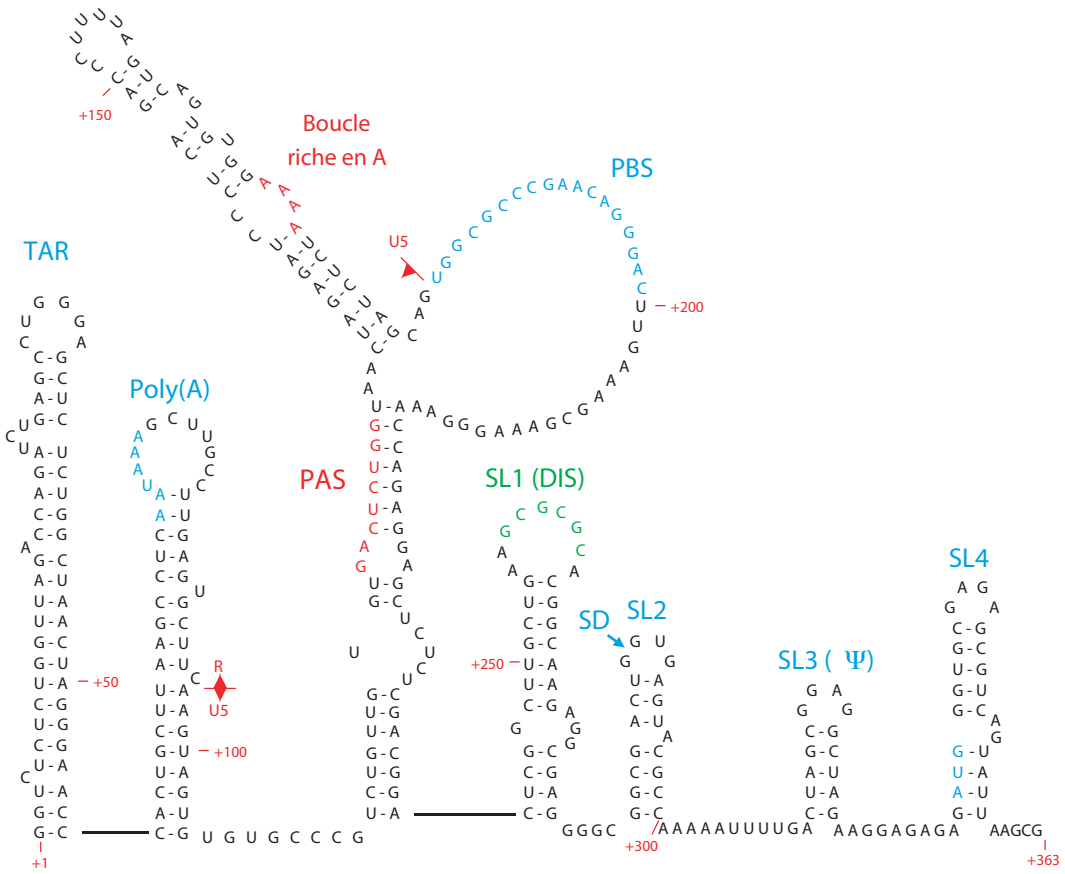


Figure 2.4 – Structure secondaire de la partie 5' de l'ARN viral
Représentation d'un modèle des éléments de structure secondaire pour la partie 5′ non traduite (5′ UTR) de l'ARN viral du VIH-1. Il s'agit de l'élément TAR, de la tige-boucle Poly(A), du PBS et des quatre tige-boucles SL1-4. La séquence correspond à l'isolat HXB2.

étant recrutée par la polyprotéine Gag (pour un article de revue [145]). Il est à noter que l'ARNt amorce est encapsidé dans les virions, et placé sur le PBS avant l'infection d'une nouvelle cellule [156].

L'ARNt$_3^{Lys}$ possède treize modifications post-transcriptionnelles (voir Figure 2.5). Certaines d'entre-elles jouent un rôle important dans différentes étapes du cycle réplicatif du virus. Le rôle de la méthylation en position 1 de la base A_{58} sera notamment discutée au Paragraphe 2.2.

La structure des ARNt a été longuement présentée au Chapitre 1. Ici, nous voulons seulement rappeler que le repliement caractéristique des ARNt lui conférant une organisation générale en forme de « L » est le résultat d'un grand nombre d'interactions dites tertiaires (liaisons hydrogène et interactions d'empilement). Toutes ces interactions procurent à l'ARNt une très grande stabilité vis à vis de la dénaturation thermique [157, 158]. Les interactions par liaison hydrogène sont représentées sur la Figure 2.5 en trait continu.

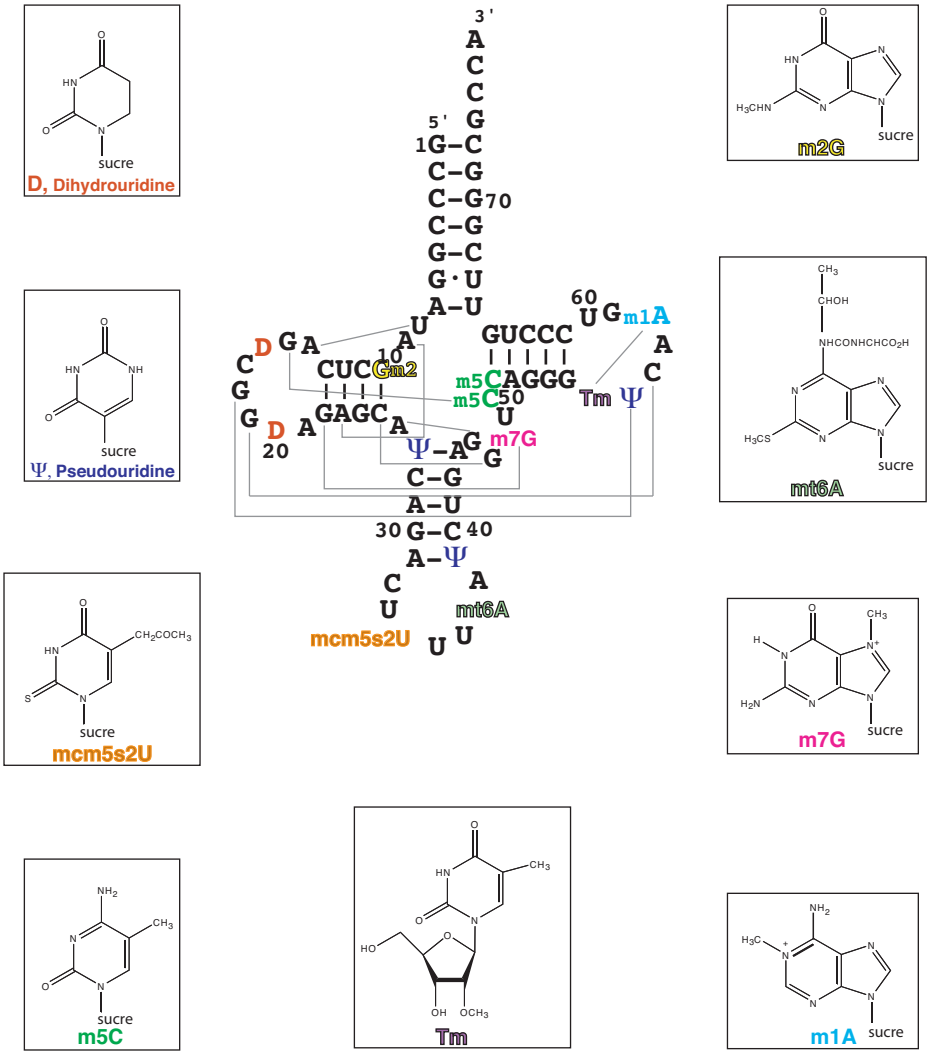


FIGURE 2.5 – Structure secondaire de l'ARNt$_3^{Lys}$
Les treize modifications post-transcriptionnelles portées par les nucléotides de l'ARNt$_3^{Lys}$ humain sont reportés en couleur sur la séquence et leur structure chimique est représentée dans les encarts. Les interactions tertiaires qui assurent l'intégrité de la structure de l'ARNt sont représentées par des lignes continues.

2.1.4 La protéine de nucléocapside (NC)

Généralités

La protéine de nucléocapside (NC) est une petite protéine de 55 résidus, très basique (pI = 9.9), et qui contient deux domaines, chacun structuré autour d'un ion Zn^{2+}, dits domaines en doigts de zinc, avec la séquence consensus suivante : $CX_2CX_4HX_4C$ [159] (voir FIGURE 2.6). On parle de doigt de zinc de type CCHC [160]. La NC est un composant très abondant des virions de VIH-1 dans lesquels elle est associée à l'ARN génomique. Elle est tout d'abord synthétisée comme une partie de la polyprotéine Gag et est ensuite clivée par la protéase virale jusqu'à sa forme mature de 55 résidus [161, 162].

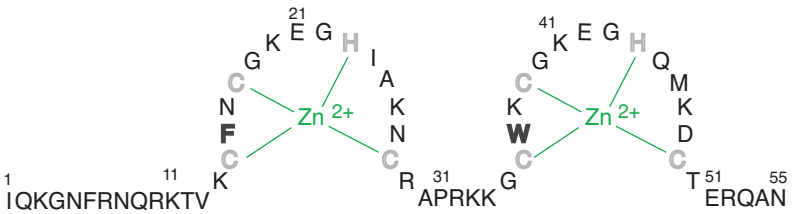


FIGURE 2.6 – Séquence de la NC du VIH-1
La séquence est celle de l'isolat NL4-3. Les cystéines et les histidines qui coordinent les ions Zn^{2+} sont représentés *en gris*. Les résidus aromatiques des doigts de zinc, *i.e.* F_{16} du doigt N-terminal et W_{37} du doigt C-terminal sont représentés en lettres grasses.

La NC est une protéine de liaison aux acides nucléiques qui joue un rôle dans quasiment toutes les étapes du cycle réplicatif du virus, de l'encapsidation du génome et l'assemblage des virions jusqu'à la transcription inverse et l'intégration de l'ADN proviral. Ainsi, la protéine NC (ou le domaine NC du précurseur Gag) est impliquée dans la dimérisation de l'ARN génomique [163, 164], dans son encapsidation [165–167], dans l'hybridation de l'ARNt amorce au PBS [116, 168–170] (voir également PARAGRAPHE 2.3), et dans les étapes de transfert de brins de la transcription inverse [171–173].

Nombreuses de ces fonctions consistent en un remaniement de structure d'acides nucléiques. La NC catalyse ces remaniements pour aboutir à la structure la plus stable d'un point de vue thermodynamique. Ceci en fait une protéine chaperonne des acides nucléiques [174, 175].

Pour comprendre le rôle de la NC dans le processus de transcription inverse, il est important de s'attarder sur les propriétés structurales et sur les propriétés de liaisons aux acides nucléiques de la NC. La plupart des informations structurales disponibles, quant aux propriétés de liaisons aux acides nucléiques, réfèrent aux interactions de la NC avec le signal d'encapsidation Ψ. Bien que ces études ne soient pas directement reliées à la transcription inverse, nous en détaillerons quelques points essentiels.

Propriétés structurales

La structure des doigts de zinc individuels [176, 177] ainsi que celle de la protéine de pleine taille [178, 179] ont été déterminées en solution par RMN (voir FIGURE 2.7 **A**). Les structures montrent que le repliement global de chaque doigt de zinc est très similaire. Cependant, le doigt de zinc C-terminal semble être plus labile que le doigt de zinc N-terminal [176]. L'existence de faibles nOe (nuclear Overhauser effect) entre des résidus des différents doigts de zinc a tout d'abord permis de penser que les deux domaines étaient relativement proches l'un de l'autre [179]. Une étude postérieure a confirmé les faibles nOe inter-doigts, mais a surtout révélé une structure très dynamique,

avec des temps de corrélation de rotation différents pour chacun des doigts de zinc, montrant ainsi que ces interactions inter-doigts sont plutôt transitoires et que les deux domaines évoluent indépendamment l'un de l'autre [180]. Dans cette étude, les extensions N- et C-terminales apparaissent aussi comme très flexibles, le tout semblant procurer à la NC une grande adaptabilité de liaison à différents partenaires de séquences et de structures variées.

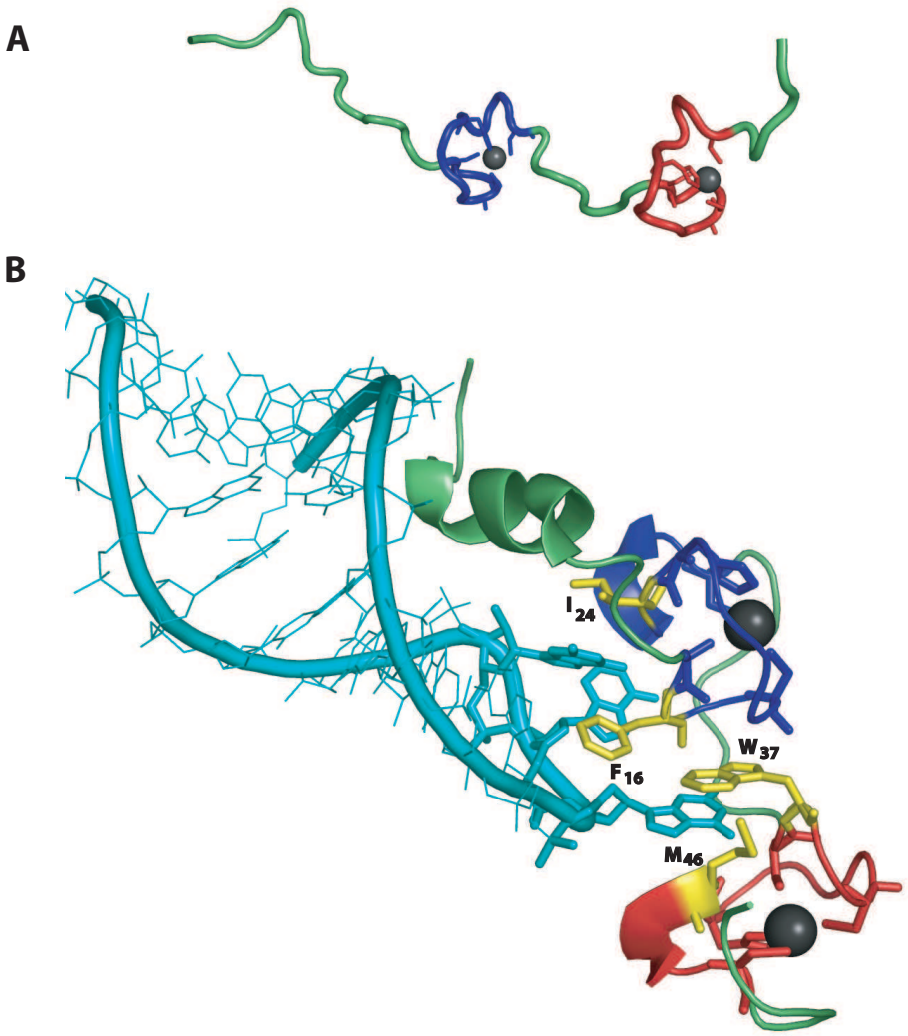


Figure 2.7 – Structures RMN de la $NC_{1\text{-}55}$ du VIH-1
A : Structure RMN de la NC libre (PDB code 1AAF) [178]. **B** : Structure RMN de la $NC_{1\text{-}55}$ en complexe avec le signal d'encapsidation SL3 (PDB code 1A1T) [118]. Le doigt de zinc N-terminal est représenté *en bleu*, le doigt de zinc C-terminal *en rouge*. Les extensions N- et C-terminale sont en *vert clair*. La tige-boucle SL3 est représentée en *bleu cyan*. Les résidus majeurs pour la reconnaissance sont représentés *en jaune*.

Propriétés de liaison aux acides nucléiques

La NC possède 15 résidus basiques, Lysine ou Arginine, chargés positivement au pH physiologique (séquence de l'isolat NL4-3 - FIGURE 2.6). Ces charges lui assurent une interaction électrostatique, le plus souvent non spécifique, avec des acides nucléiques variés, tant ARN que ADN [181–185]. A l'intérieur du virion, elle recouvre ainsi entièrement l'ARN génomique, à raison d'une protéine de NC tous les 5 à 8 nucléotides [162, 181, 186–188]. La NC protège de cette façon l'ARN viral d'une dégradation par des nucléases. Ces interactions électrostatiques représentent une part importante des interactions entre la NC et les acides nucléiques. La fixation de la NC, tant à l'ADN qu'à l'ARN, est en effet fortement dépendante de la force ionique [181–185]. Ces propriétés ressemblent fortement aux interactions non spécifiques que peuvent établir des ligands polycationiques comme les polyamines ou la poly-Lysine.

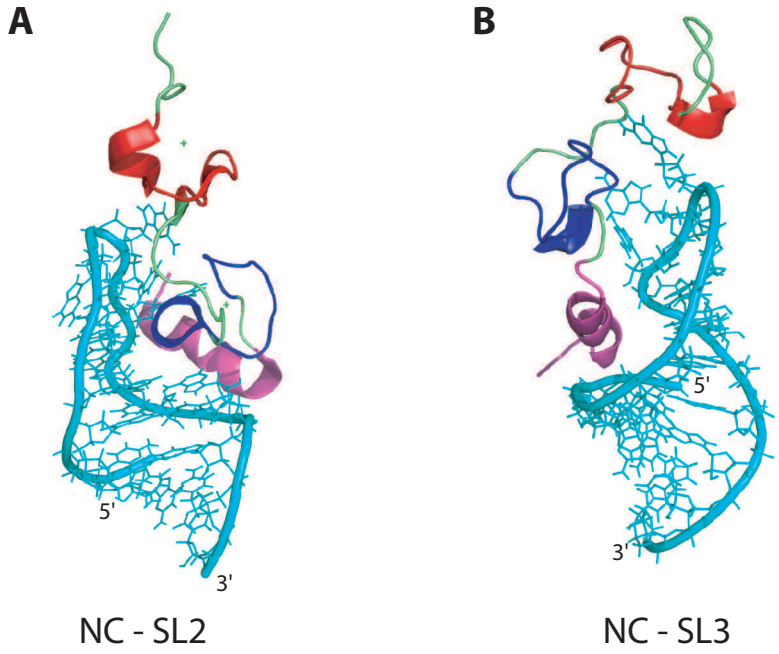


FIGURE 2.8 – Comparaison des structures $NC_{1\text{-}55}$-SL2 et $NC_{1\text{-}55}$-SL3
A : Structure RMN du complexe $NC_{1\text{-}55}$-SL2 (PDB code 1F6U) [119]. **B** : Structure RMN du complexe $NC_{1\text{-}55}$-SL3 (PDB code 1A1T) [118]. Le doigt de zinc N-terminal est représenté *en bleu*, le doigt de zinc C-terminal *en rouge*. L'hélice 3_{10} N-terminale est représentée *en mauve*. Les tiges-boucles SL2 et SL3 sont représentées en *bleu cyan*.

Néanmoins, certaines interactions de la NC avec des acides nucléiques sont plus spécifiques. Les nombreuses études relatives à la fixation spécifique de la NC à des séquences particulières d'acides nucléiques réfèrent essentiellement au signal d'encapsidation Ψ. Celui-ci est constitué de quatre tiges-boucles (SL1, SL2, SL3 et SL4) situées en aval du PBS (FIGURE 2.4). La fixation de la NC aux tiges-boucles SL2 et SL3 est très forte (K_d = 20 – 30 nM pour une force ionique physiologique), et plus importante que pour les tiges-boucles SL1 et

SL4 (K_d = 100 – 320 nM) [185]. Les structures de la NC liée aux tiges-boucles SL2 et SL3 ont été déterminées par RMN en solution [118, 119]. La structure du complexe NC-SL3 est représentée FIGURE 2.7 **B**. Dans cette structure, il est intéressant de noter que chaque doigt de zinc réalise une poche hydrophobe venant reconnaître une base guanine de la boucle d'ARN. Les résidus majeurs formant ces poches hydrophobes sont les résidus aromatiques présents à la deuxième position de chaque doigt de zinc (*i.e.* F_{16} et W_{37} - FIGURE 2.6) ainsi que des résidus à chaîne latérale hydrophobe (*i.e.* I_{24} et M_{46}). La partie N-terminale, quant à elle, se structure en hélice 3_{10} lors de fixation à la tige boucle et vient se loger dans le grand sillon de la tige d'ARN.

Même si le mode de reconnaissance de SL2 est très comparable, en ce qui concerne chaque doigt de zinc, l'organisation globale du complexe NC-SL2 est différente. En effet, les doigts de zinc N- et C-terminaux adoptent une orientation relative différente, et l'extension N-terminale, bien que se structurant de même en hélice 3_{10}, vient s'enchâsser dans le petit sillon de la tige [119]. La comparaison des structures des deux complexes est présentée FIGURE 2.8. Ces différences appuient le fait que la NC puisse reconnaître des séquences d'ARN variées avec des affinités comparables — la flexibilité de la partie N-terminale et du lien reliant les deux doigts de zinc lui conférant une grande adaptabilité.

Ces études structurales, en accord avec de nombreuses études par spectroscopie de fluorescence [181, 182], montrent que ces doigts de zinc de type CCHC, qui contiennent des résidus hydrophobes à des positions adéquates forment une surface idéale de reconnaissance pour les résidus guanine des boucles d'ARN. Il s'agit là d'un point important pour la compréhension des RÉSULTATS des CHAPITRES 3 et 4.

Après avoir présenté les différents acteurs prenant part au processus de la transcription inverse, nous pouvons maintenant aborder la description de son déroulement, de l'initiation à la terminaison.

2.2 Déroulement de la transcription inverse

Initiation et synthèse de l'ADN$_c$(–) *strong-stop*

La polymérisation des désoxy-nucléotides est initiée à partir d'un complexe dit de pré-initiation : l'hybride ARNt$_3^{Lys}$/ARN viral, l'ARNt$_3^{Lys}$ jouant le rôle d'amorce. L'élongation de l'amorce s'arrête à l'extrémité 5′ de l'ARN viral, donnant ainsi naissance à l'ADN$_c$(–) *strong-stop* (FIGURE 2.9 ① et ②). La formation de l'hybride ARNt$_3^{Lys}$/ARN viral est un phénomène complexe et fera l'objet du PARAGRAPHE 2.3.

Au cours de la polymérisation, la RT dégrade partiellement la matrice d'ARN viral par son activité RNase H. Les fragments d'ARN encore hybridés sont éliminés en fin de synthèse de l'ADN$_c$(–) *strong-stop* par la RT. En effet, l'activité RNasique de la RT a lieu en deux temps qui correspondent à des activités *endo-* et *exo*-ribonucléasique, encore appelée polymérisation-dépendante et polymérisation-indépendante.

Le démarrage de la synthèse du brin (−) se fait en deux étapes : initiation et élongation [189]. Ces deux phases se distinguent par une processivité différente de la RT. La synthèse d'ADN est dite processive lorsque la RT incorpore plusieurs nucléotides sans se dissocier du complexe amorce/matrice. Elle est dite distributive lorsque la RT se dissocie du complexe amorce/matrice après l'addition de chaque nucléotide. Une vitesse de polymérisation (k_{pol}) très supérieure à la vitesse de dissociation (k_{off}) de la RT du complexe amorce/matrice traduit donc une polymérisation de type processif. Par opposition, une polymérisation de type distributif se traduit par un rapport k_{pol}/k_{off} inférieur à 1. L'étape d'initiation est ainsi distributive ($k_{pol}/k_{off} < 1$), alors que l'étape d'élongation est hautement processive ($k_{pol}/k_{off} \gg 1$) [190, 191].

Transfert de l'ADN$_c$(−) *strong-stop* : premier transfert de brin

L'hypothèse du transfert de brin d'ADN$_c$(−) *strong-stop* vers l'extrémité 3′ de l'ARNv génomique a été proposée, puis démontrée chez le virus de la leucémie murine de Moloney [192], ceci après la découverte de formes circulaires d'ADN proviral et d'une séquence répétée, la séquence R, en 5′ et en 3′ de l'ARNv [193]. Bien que le premier transfert de brin ait uniquement besoin de la RT comme facteur protéique viral, du moins *in vitro*, la NC semble augmenter l'efficacité de ce transfert de brin [171].

Etant donné la nature dimérique de l'ARN génomique viral, le transfert de l'ADN$_c$(−) *strong-stop* à l'extrémité 3′ de l'ADN génomique peut être intra- ou inter-moléculaire. Même si les premières études ont été contradictoires, ce transfert de brin semble bien être autant intramoléculaire qu'intermoléculaire [194, 195].

Synthèse du brin d'ADNc(−)

Selon que le premier transfert de brin est intra- ou inter-moléculaire, la synthèse de l'ADNc(−) se poursuit jusqu'au PBS ou jusqu'à l'extrémité 5′ de l'ARN viral. Aussi, l'ARN génomique est dégradé dans son ensemble par l'activité RNase H de la RT. Néanmoins deux séquences polypuriques, ou *PolyPurines Tracts* (PPT), localisées respectivement en amont de la séquence U3 (3′ PPT) et au centre de l'ARN viral (central PPT, cPPT) sont résistantes à l'activité RNAse de la RT et subsistent donc à la fin de la synthèse du brin d'ADNc(−) (Figure 2.9 ③ et ④).

Synthèse du brin d'ADNc(+)

Ces fragments PPT non dégradés servent d'amorces à la synthèse du brin d'ADNc(+) [196–198]. La séquence 3′ PPT donnant naissance au brin U(+), et la séquence cPPT, qui est la réplique exacte de la séquence 3′ PPT, donnant naissance au brin D(+) (Figure 2.9 ⑤).

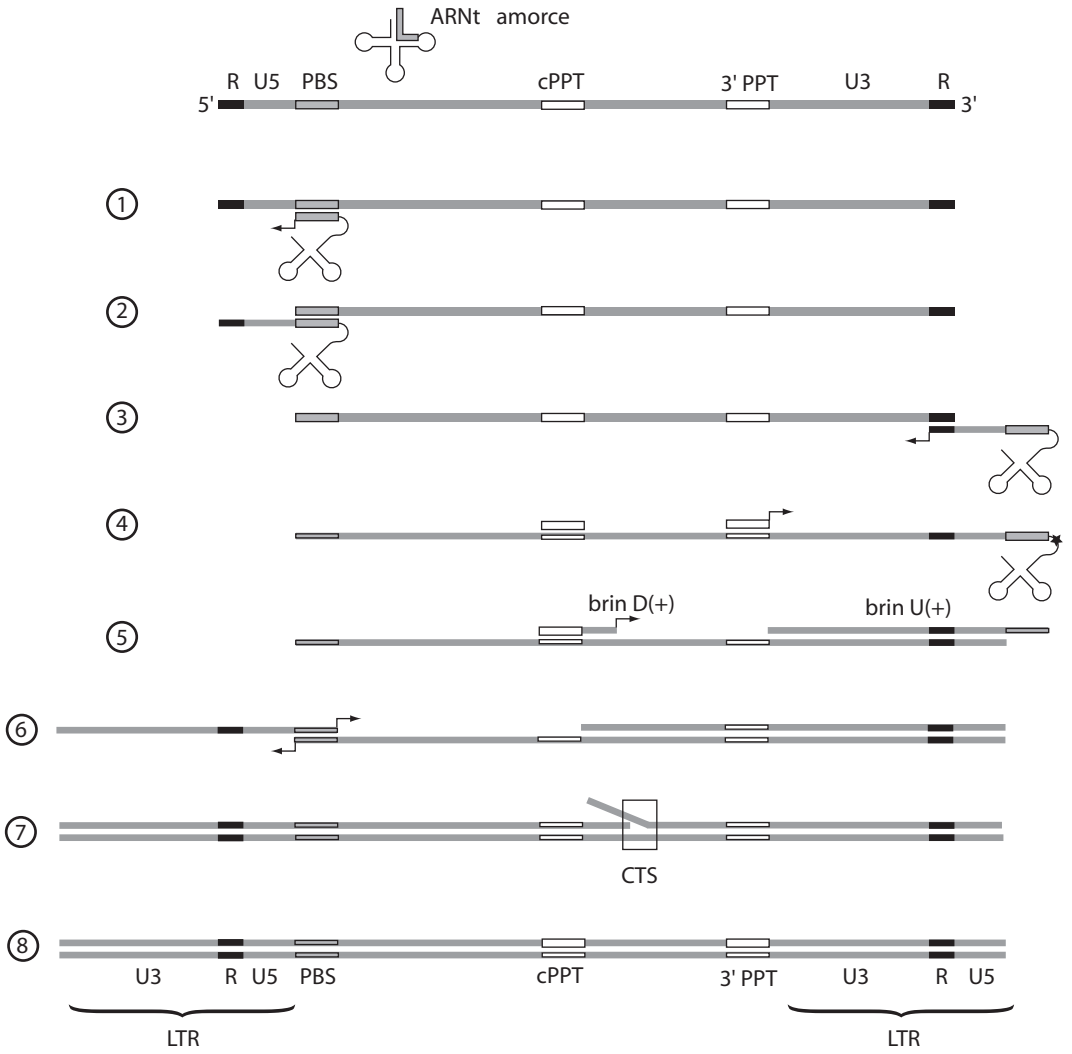


Figure 2.9 – Déroulement de la transcription inverse du VIH-1
① Initiation de la synthèse du brin (−). ② Synthèse de l'ADN$_c$(−) *strong-stop*. ③ Premier transfert de brin, fin de la synthèse du brin d'ADNc(−), dégradation de l'ARN viral par la RNase H. ④ Initiation de la synthèse du brin U(+) à partir du 3′ PPT. ⑤ Synthèse de l'ADN$_c$(+) *strong-stop* jusqu'à la modification m^1A_{58} sur l'ARNt$_3^{Lys}$ ($\star$). Coupure de l'ARNt$_3^{Lys}$ par la RNase H. Initiation et synthèse du brin D(+) à partir du cPPT. ⑥ Deuxième transfert de brin. Fin de la synthèse du brin d'ADNc(−). ⑦ Terminaison de la synthèse du brin U(+) au niveau du CTS. ⑧ Fin de la transcription inverse.

Second transfert de brin

L'ARNt$_3^{Lys}$ amorce est rétrotranscrit jusqu'à la base modifiée m^1A_{58} [172, 199, 200]. La modification en position 1 de l'adénine empêche celle-ci de réaliser un appariement canonique de type Watson-Crick. Celle-ci termine donc la synthèse du brin d'ADNc(+) par une séquence complémentaire au PBS du brin (−). Le domaine RNase H de la RT clive ensuite l'ARNt$_3^{Lys}$ hybridé au PBS, entre le nucléotide A_{76} et le nucléotide C_{75} [201]. Le brin U(+) porte alors une

extrémité 3′ simple brin, complémentaire de l'extrémité 3′ du brin d'ADNc(−), elle même complémentaire de la séquence PBS présente initialement sur l'ARN viral. Un second transfert de brin peut donc avoir lieu entre ces deux séquences complémentaires (FIGURE 2.9 ⑥). Même si, comme pour le premier transfert de brin, la RT est capable de catalyser à elle seule le deuxième transfert de brin, il semblerait que la NC augmente l'efficacité de celui-ci [172].

Terminaison de la transcription inverse

L'élongation du brin (+) a lieu sur la matrice d'ADNc(−) jusqu'à la séquence cPPT, où elle se poursuit alors par un mécanisme appelé déplacement de brin, jusqu'à ce que la RT atteigne la séquence de terminaison centrale (CTS) situées 80 à 100 nucléotides en aval du cPPT (FIGURE 2.9 ⑦). Ce motif est très efficace pour interrompre l'élongation d'ADN par la RT du VIH-1.

A ce stade, la RT a synthétisé un ADN proviral double-brin flanqué de deux séquences répétées (*Long Terminal Repeats* - LTR), et qui contient un brin (+) discontinu en son centre (« le flap »). Une endonucléase cellulaire coupe le brin discontinu, et une ADN ligase permet d'obtenir un ADN double brin complet sur toute sa longueur (FIGURE 2.9 ⑧) [202]. L'ADN proviral est ensuite éventuellement intégré dans le génôme par l'Intégrase (IN) du virus.

Tout ce mécanisme de réplication est conditionné par l'utilisation d'une amorce initiale adéquate. Elle doit comme nous l'avons déjà vu être complémentaire de la région du PBS de l'ARN viral, et posséder la modification m^1A_{58} dans sa boucle TΨC. La formation de l'hybride $ARNt_3^{Lys}$/ARN viral est un phénomène relativement complexe qui nécessite des facteurs protéiques viraux et qui est régulée par d'autres séquences de l'ARN viral. Ces différents aspects seront détaillés dans le PARAGRAPHE suivant.

2.3 Formation de l'hybride $ARNt_3^{Lys}$/ARN viral

La formation de l'hybride $ARNt_3^{Lys}$/ARN viral consiste en l'appariement des 18 nucléotides du PBS à la séquence complémentaire formant l'extrémité 3′ de l'ARNt amorce. Ceci implique deux réarrangements majeurs de l'$ARNt_3^{Lys}$:

- Un échange de brin durant lequel les hélices intramoléculaires formant la tige acceptrice et la tige du bras T sont converties en une hélice intermoléculaire ARNt/PBS.
- L'ouverture de la structure tertiaire de l'ARNt, et en particulier la fusion des nombreuses interactions entre les boucles TΨC et D.

Ce phénomène d'ouverture de la structure de l'ARNt, suivi d'un réappariement des bases de celui-ci avec le PBS (FIGURE 2.10), même s'il conduit à une structure globalement plus stable d'un point de vue thermodynamique – puisque viennent au moins s'ajouter les interactions au niveau de l'extrémité 3′ GCCA et au niveau des bases G_{59} et U_{60} de la boucle TΨC – possède une barrière d'activation élevée [203].

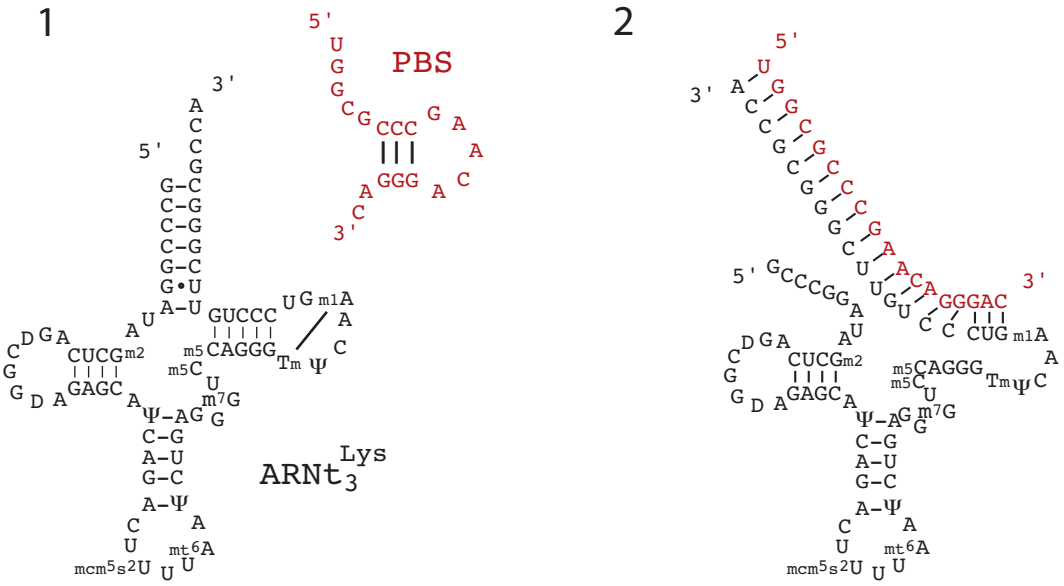


FIGURE 2.10 – Structure de l'hybride entre l'ARNt$_3^{Lys}$ et le PBS
Paires de bases entre l'ARNt$_3^{Lys}$ (*en noir*) et le PBS (*en rouge*). **1** : Avant hybridation. **2** : Après hybridation.

Cette haute barrière d'activation fait que, *in vitro*, en présence seulement d'ARN viral, l'ARNt$_3^{Lys}$ ne s'hybride pas au PBS, du moins à température physiologique [116]. Il a été montré que la formation de l'hybride était stimulée par le domaine NC de Gag *in vivo* [169, 204] et que la protéine NC mature présentait aussi cette activité de chaperonne *in vitro* [116, 168, 170, 203, 205]. Celle-ci augmenterait la vitesse d'hybridation de l'ARNt$_3^{Lys}$ au PBS d'un facteur 10^5 [203]. Il est aussi intéressant de noter que les complexes formés par la NC ou par une hybridation à la chaleur semblent équivalent d'un point de vue fonctionnel [206, 207].

Même si l'intervention de la NC (ou du domaine NC de Gag) dans ce processus d'hybridation a toujours été un point d'accord de nombreuses équipes, l'implication des différents sous-domaines de la protéine (doigts de zinc et extensions basiques) dans ce mécanisme de remodelage d'ARN, a quant à elle été longtemps une source de divergence. En effet, les différents groupes ont utilisé des NC de différentes longueurs, provenant de différents isolats viraux, et souvent des amorces ARNt$_3^{Lys}$ produites par transcription, qui sont de ce fait dépourvues de modifications post-transcriptionnelles, et sont donc intrinsèquement moins stables. Un consensus sur le rôle de la NC dans la formation de l'hybride ARNt$_3^{Lys}$/ARN viral semble toutefois émerger [117].

2.3.1 Hybridation de l'ARNt$_3^{Lys}$ avec le PBS

Etant donné les facultés de la NC à permettre l'hybridation d'acides nucléiques hautement structurés, il a tout d'abord été envisagé que la NC soit capable de faire fondre les interactions secondaires et tertiaires de ces molé-

cules [208]. Cependant, aucun effet de désenroulement des hélices de l'$ARNt_3^{Lys}$ n'a été observé en présence de NC seulement [114, 209, 210].

La délétion de courtes séquences contenant les résidus basiques de l'extension N-terminale conduit à une perte de l'activité de la NC [168]. Ainsi les extensions N- et C-terminale ont tout d'abord été décrites comme suffisantes à l'activité d'hybridation *in vitro* [168, 205]. En parallèle, un variant plus court de la NC, la NC_{12-53}, à laquelle il manque justement ces extensions, a aussi été décrit comme permettant la formation de l'hybride [211]. De plus, des mutations dans les doigts de zinc abolissent l'activité de la NC sur le processus d'hybridation *in vitro*, suggérant ainsi un rôle direct des doigts de zinc dans celui-ci [211–213]. Toutefois, d'autres études ont montré que la structure des doigts de zinc n'était pas essentielle pour la formation de l'hybride amorce/matrice [170]. Et des études de mutations *in vivo*, ont quant à elles montré que la présence intacte des doigts de zinc était requise pour la plupart des fonctions remplies par la NC [167, 212].

Malgré toutes ces divergences, un consensus semble émerger des études récentes. Il est maintenant admis que les doigts de zinc sont nécessaires pour une activité chaperonne optimale de la NC [116, 214, 215].

En l'absence d'une séquence complémentaire, la fixation de la NC sur un ARN n'entraîne que de petites variations de structure sur celui-ci. Néanmoins, les zones de fixations préférentielles de la NC sur l'$ARNt_3^{Lys}$ peuvent se révéler utiles à la compréhension du mode d'action de la NC. Il a été montré par RMN que la NC_{12-53} se fixait préférentiellement à la tige acceptrice et au bras D de l'$ARNt_3^{Lys}$, ainsi qu'au niveau des interactions tertiaires entre les boucles TΨC et D [114]. L'extension basique N-terminale, ne modifie pas l'empreinte de fixation de la NC, mais semble renforcer la fixation [115].

En présence d'une séquence complémentaire, l'effet déstabilisateur de la NC sur les acides nucléiques entraîne cette fois-ci des changements radicaux dans leur structure. Deux études récentes ont apporté des résultats complémentaires qui ont permis d'expliquer bon nombre d'observations passées sur le mécanisme d'hybridation entre l'$ARNt_3^{Lys}$ et l'ARN viral. La première, menée dans l'équipe de Karin Musier-Forsyth, a utilisé des expériences de clivage au terbium et des études cinétiques [170, 203]. L'autre a été menée dans notre laboratoire par spectroscopie RMN [114–116]. Il a ainsi été établi que l'hybridation suit une loi cinétique bimoléculaire, ce qui implique que la formation des premières paires de bases intermoléculaires – ou nucléation – soit l'étape cinétiquement déterminante du processus [203]. La nucléation est précédée de la fusion de quelques paires de bases au sein de l'$ARNt_3^{Lys}$ (3–4 pb) et est suivi d'un échange de brin rapide avec le PBS, dont on peut envisager le fonctionnement en imaginant celui d'une « fermeture éclair ». L'étude RMN montre que la nucléation ne nécessite pas la présence de la NC, même si celle-ci facilite fortement son apparition. La NC est dans tous les cas requise pour atteindre l'hybridation complète avec le PBS [116, 203].

Un modèle de mécanisme, en accord avec les études cinétiques [203], a ainsi pu être proposé par notre laboratoire [116]. Celui-ci est décrit à la Figure 2.11.

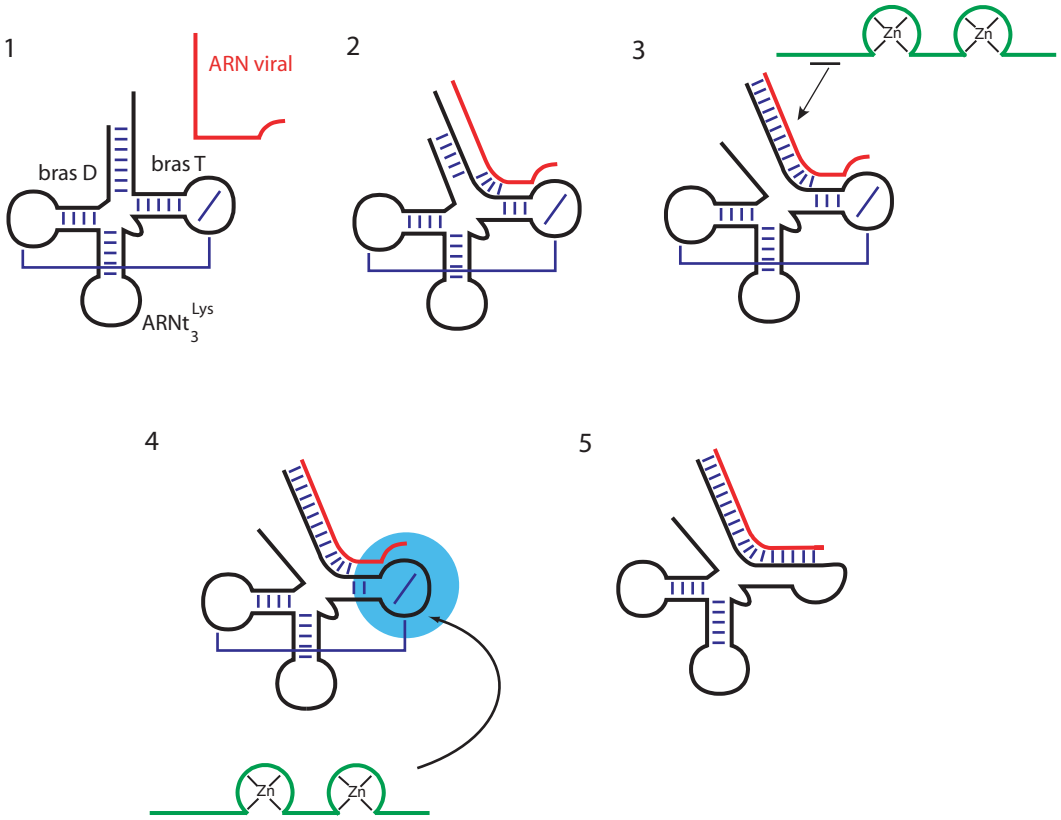


FIGURE 2.11 – Mécanisme de l'hybridation entre l'ARNt$_3^{Lys}$ et le PBS **1** : L'ARNt$_3^{Lys}$ est *en noir*, le PBS *en rouge* et la NC *en vert*. **2** : L'invasion de la structure de l'ARNt amorce par l'ARN viral débute au pied de la tige acceptrice et peut avoir lieu sans l'intervention de la NC. **3** : L'échange de brin se propage ensuite le long de la tige acceptrice puis aux premières paires de bases du bras T. La NC accélère ce réappariement, probablement par le biais d'une interaction électrostatique non spécifique. **4** : A ce stade, le cœur de la structure tertiaire de l'ARNt persiste et constitue un verrou structural stable. La NC parvient à faire sauter ce verrou structural, par l'intervention spécifique de ses doigts de zinc. **5** : Hybridation complète. Le mécanisme proposé est basé sur une étude RMN hétéronucléaire avec l'ARNt$_3^{Lys}$ marqué ^{15}N et un PBS non marqué [116].

L'invasion de la structure de l'ARNt amorce par l'ARN viral débute au pied de la tige acceptrice et peut avoir lieu sans l'intervention de la NC. L'échange de brin se propage ensuite le long de la tige acceptrice puis aux premières paires de bases du bras T. La NC accélère ce réappariement, probablement par le biais d'interactions électrostatiques non spécifiques, car cette étape ne nécessite pas l'intervention stricte de la NC et peut tout aussi bien être réalisée grâce à un polycation tel que la poly-Lysine [116]. A ce stade, le cœur de la structure tertiaire de l'ARNt persiste et constitue un verrou structural stable. La NC parvient à faire sauter ce verrou structural, grâce à l'intervention spécifique de ces doigts de zinc, car cette étape ne peut être réalisée avec la poly-Lysine. La NC intervient vraisemblablement en interagissant au niveau

des interactions tertiaires entre les boucles TΨC et D qui « ferment » la structure de l'ARNt. L'intervention spécifique de la NC permet d'atteindre l'état d'hybridation complète entre l'$ARNt_3^{Lys}$ et le PBS.

La seule différence entre les deux études [116, 203], réside dans le point de nucléation entre l'amorce et l'ARN viral. L'équipe de Karin Musier-Forsyth [203] propose en effet que la nucléation ait lieu au niveau de l'extrémité 3′ CCA de l'ARNt. Cette partie de l'$ARNt_3^{Lys}$ ne peut pas être observée par RMN dans les expériences qui ont mené à la proposition du mécanisme précédent. En effet, dans les expériences RMN, la formation des paires de bases est suivie par l'intermédiaires des protons iminos de l'$ARNt_3^{Lys}$ marqué ^{15}N. Les protons iminos sont portés par les bases G et U. L'extrémité 3′ CCA n'est donc pas observable dans ce type d'expérience.

Des expériences complémentaires, dans lesquelles cette fois-ci le PBS a été marqué ^{15}N, ont été réalisées au laboratoire pour tenter de résoudre l'ambiguïté relative au site de nucléation. Cette étude a fait partie d'une publication conjointe à mon travail de thèse et sera pour cela présentée parmi les Résultats, au Chapitre 3.

Même si l'intervention de la NC ou du domaine NC de Gag a permis d'envisager des mécanismes expliquant l'hybridation entre l'$ARNt_3^{Lys}$ et le PBS [116, 203], certains points ne peuvent s'expliquer si l'on considère que la formation de l'hybride $ARNt_3^{Lys}$/ARN viral ne fait intervenir que les 18 nucléotides du PBS. En effet, il existe une stricte spécificité d'usage de l'amorce $ARNt_3^{Lys}$ par le VIH-1. Il n'a ainsi jamais été observé de mutant qui ait adopté un autre ARNt comme amorce. De plus, les études cherchant à créer de tels mutants en modifiant simplement la séquence du PBS pour la rendre complémentaire d'un autre ARNt, ont toujours échouées [216–218]. Les virions produits par des virus mutés dans le PBS présentent rapidement un génotype de révertant vers la séquence du PBS sauvage (PBS_3^{Lys}) [216]. Ceci peut paraître surprenant, mais il est possible de former de tels virions en seulement un cycle de transcription inverse, si l'amorce $ARNt_3^{Lys}$ naturelle est utilisée face au PBS muté. Ainsi, même si les deux types des virions sont tout d'abord produits (sauvage et muté), comme la transcription inverse est beaucoup plus efficace avec l'amorce naturelle, les virions produits finissent rapidement par retrouver la séquence naturelle PBS_3^{Lys}. Ceci montre clairement que le PBS n'est pas le seul déterminant de spécificité de l'ARNt amorce.

2.3.2 Interactions additionnelles entre l'$ARNt_3^{Lys}$ et l'ARN viral

Afin d'améliorer la compréhension du mécanisme de sélection exclusive de l'amorce $ARNt_3^{Lys}$ par le VIH-1, la recherche des déterminants de sélection spécifique de l'ARNt, autres que le PBS, a fait l'objet de nombreuses études. Ces déterminants secondaires ont tout naturellement été envisagés comme des contacts additionnels entre l'$ARNt_3^{Lys}$ et l'ARN viral.

Les modèles d'interactions additionnelles proposées par les différents groupes ont là-encore été longtemps en relative contradiction. Ceci est venu essentiellement du fait de l'utilisation de séquences d'ARN viral provenant de différents isolats. Ainsi les premières études de structures du complexe $ARNt_3^{Lys}$/ARN viral ont été effectuées dans l'équipe de Roland Marquet sur l'isolat MAL du VIH-1 [219–221]. Ces études ont mis en évidence une interaction entre une boucle riche en A de l'ARN viral et la boucle riche en U de l'anticodon de l'$ARNt_3^{Lys}$. D'autres études ont été menées sur les isolats de sous-type B du VIH-1, qui prédomine largement en Europe occidentale et en Amérique du nord, tels que HXB2 et NL4-3. Certaines de ces études, menées dans l'équipe de Casey Morrow, ont semblé confirmer l'importance de la boucle riche en A dans les isolats HXB2/NL4-3 [222–227]. Alors que d'autres études, menées dans l'équipe de Ben Berkhout, ne proposent aucun rôle notoire à la boucle riche en A, mais ont en revanche identifié une séquence en amont du PBS, qui aurait la propriété d'activer l'initiation de la transcription inverse et qui a pour cela été dénommée *Primer Activation Signal* ou PAS [228–232].

Les contradictions qui sont apparues entre les différentes équipes cherchant à mettre en évidence les interactions additionnelles entre l'$ARNt_3^{Lys}$ et l'ARN viral proviennent vraisemblablement des différences de structure de la partie PBS de l'ARN viral dans les différents isolats utilisés.

Domaine PBS des virus de sous-type A et B du VIH-1

Les sous-types A et B sont les plus utilisés dans la littérature. L'isolat MAL est issu de recombinaisons entre virus de sous types A, D et I, tandis que les isolats HXB2 et NL4-3 appartiennent au sous-type B. La région PBS de l'isolat MAL correspond à celle des isolats de sous-type A.

Les différents isolats présentent de fortes homologies de séquence. Cependant de nombreuses mutations sont observées de part et d'autre de la région PBS entre HXB2 et NL4-3 d'une part et MAL d'autre part. L'isolat NL4-3 présente 4 substitutions en aval du PBS par rapport à l'isolat HXB2. L'isolat MAL diffère des isolats HXB2/NL4-3 par la présence d'une insertion de 23 nucléotides en aval du PBS, en plus de mutations ponctuelles. Cette insertion résulte de la duplication d'une séquence comprenant une partie du PBS. Des études de cartographie en solution, menées sur les ARN viraux des isolats MAL [143, 220] et HXB2 [233–235], ont permis d'établir des modèles de structure secondaire du domaine PBS (Figure 2.12). Les différences observées au niveau des séquences virales conduisent à des différences structurales importantes du domaine PBS. La différence majeure entre ces deux modèles est la localisation de la séquence riche en A, qui se situe dans une boucle apicale dans le cas de l'isolat MAL et dans une boucle interne pour les isolats de sous-type B.

La boucle riche en A

L'interaction possible entre une boucle riche en A en aval du PBS (Figure 2.12) et la boucle riche en U de l'$ARNt_3^{Lys}$ a été analysée en détail pour l'isolat MAL du VIH-1. Cette interaction ARNt/ARNv a été proposée par modélisa-

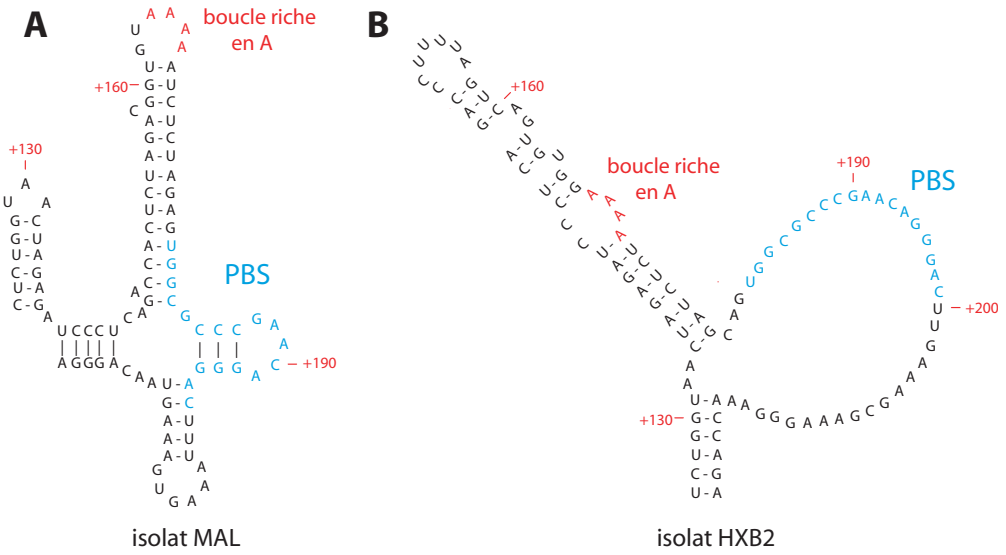


FIGURE 2.12 – Modèles de structure secondaire de l'ARN viral des isolats MAL et HXB2 du VIH-1
A : Modèle de structure secondaire de l'ARN viral de l'isolat MAL du VIH-1 (nts +123 à +218). La séquence de l'ARNv est en noir, sauf pour le PBS (*en bleu*) et pour la boucle riche en A (*en rouge*). **B** : Modèle de structure secondaire de l'ARN viral de l'isolat HXB2 du VIH-1 (nts +125 à +223). Le code couleur est identique à celui de l'isolat MAL. La différence majeure entre ces deux modèles est la localisation de la séquence conservée riche en A, qui se situe dans une boucle apicale dans le cas de l'isolat MAL et dans une boucle interne pour l'isolat HXB2.

tion de structures d'ARN sur la base de cartographie en solution par sonde chimiques et enzymatiques sur l'isolat MAL du VIH-1 [219, 221]. De plus, cette interaction entre la boucle riche en A et l'anticodon de l'ARNt$_3^{Lys}$ semble stabilisée par les modifications post-transcriptionnelles de l'ARNt [219, 220]. Le rôle possible de cette interaction dans la transcription inverse a été grandement discuté tant pour les isolats MAL que pour les isolats de sous-type B (HXB2/NL4-3) [219, 222, 236–240]. Il a entre autres été montré que la délétion de la boucle riche en A en amont du PBS affectait grandement les phases d'initiation et d'élongation de la transcription inverse [189, 241], que cette délétion affectait la réplication des virus et que la séquence riche en A était restaurée par les virus mutés lors de cultures de longue durée [241].

Le rôle de la boucle riche en A a aussi été grandement étudiée dans l'équipe de Casey MORROW sur des isolats de sous-type B [222–227]. L'adaptation simultanée du PBS et de la boucle riche en A, pour forcer l'utilisation d'un autre ARNt amorce dans la transcription inverse, a été examinée *in vivo*. De tels double-mutants présentent une capacité de réplication très réduite. Mais des variants utilisant les amorces ARNt^{His}, ARNt^{Glu} et ARNt$_{1,2}^{Lys}$ ont pu être sélectionnés sur des cultures de longue durée. Ceci venant bien évidemment très fortement renforcer l'idée d'une interaction additionnelle entre la « boucle riche en A » et l'anticodon de l'ARNt amorce, qui serait indispensable à l'initiation

de la transcription inverse.

Cependant, des études plus récentes d'interactions par sondes chimiques ont été réalisées *in vitro* et *ex vivo* sur les isolats HXB2/NL4-3. Ces études révèlent que l'interaction entre l'ARNt$_3^{Lys}$ et la boucle riche en A n'a pas lieu dans ces isolats [242, 243]. Ainsi dans les double-mutants PBS/boucle riche en A de l'équipe de Casey Morrow, l'interaction additionnelle entre l'anticodon et la « boucle riche en A » doit plutôt compenser une contribution négative introduite par l'utilisation d'une amorce non naturelle, que refléter une interaction présente dans le virus sauvage.

De plus, cette boucle riche en A n'est pas présente dans la région du PBS chez certains *Lentivirus* qui utilisent l'ARNt$_3^{Lys}$ comme amorce. Tout ceci suggère que cette interaction entre la séquence riche en A et l'amorce ARNt n'est pas conservée parmi les rétrovirus et qu'elle n'a lieu que dans certains sous-type du VIH-1 dont la structure de la région du PBS correspond a celle de l'isolat MAL, soit 14 % des isolats du VIH-1 [242]. Cette interaction semble en tout cas très faible, du moins *in vitro*, puisque les paires de bases A-U entre la boucle riche en A et l'anticodon de l'ARNt$_3^{Lys}$ n'ont pu être observées par RMN [116].

La séquence *Primer Activation Signal*

Une autre interaction additionnelle entre l'ARNt$_3^{Lys}$ et l'ARN viral a été proposé par l'équipe de Ben Berkhout. Celle-ci implique la séquence +123 à +130 de l'ARN viral, appelée PAS (*Primer Activation Signal*) et le bras TΨC (nucléotides 48 à 55) de l'ARNt$_3^{Lys}$, appelée anti-PAS (Figure 2.13) [228]. Cette séquence a tout d'abord été identifiée par l'analyse des capacités de réplication de virus mutés dans les régions flanquant le PBS. Cette séquence n'intervient pas sur l'hybridation de l'amorce au PBS, mais active l'initiation de la transcription inverse, puisqu'il a été montré qu'elle facilitait l'incorporation du premier nucléotide en 3′ de l'ARNt [228]. La séquence PAS est en quelque sorte « cachée » dans une hélice intramoléculaire de l'ARN viral (Figure 2.13). Des mutations sur l'autre brin de l'hélice, empêchant la formation de celle-ci, augmente l'efficacité de l'initiation de la transcription inverse, probablement parce qu'elles favorisent l'accessibilité à la séquence PAS. Cette séquence PAS jouerait donc un rôle régulateur à la fois positif et négatif sur la transcription inverse.

Deux interactions semblables ont été proposées. Elles impliquent respectivement le bras TΨC de l'ARNt^{Trp} et l'ARN viral du virus du sarcome de Rous (RSV) [244], et le bras TΨC de l'ARNt$_3^{Lys}$ et l'ARN viral du VIH-2 [245]. De plus des études phylogénétiques semblent indiquer que cette interaction soit conservée chez tous les rétrovirus [246, 247].

Un point important venant conforter l'idée selon laquelle la séquence PAS jouerait un rôle majeur dans l'initiation de la transcription inverse, est le fait que l'on puisse changer la spécificité d'usage de l'amorce ARNt en combinant les mutations des séquences PBS et PAS. Le changement d'amorce, pour les ARNt$_{1,2}^{Lys}$, a ainsi été possible *in vitro* et *in vivo* sur des mutants d'ARN viral présentant un changement conjoint des séquences PBS et PAS [228–230, 246].

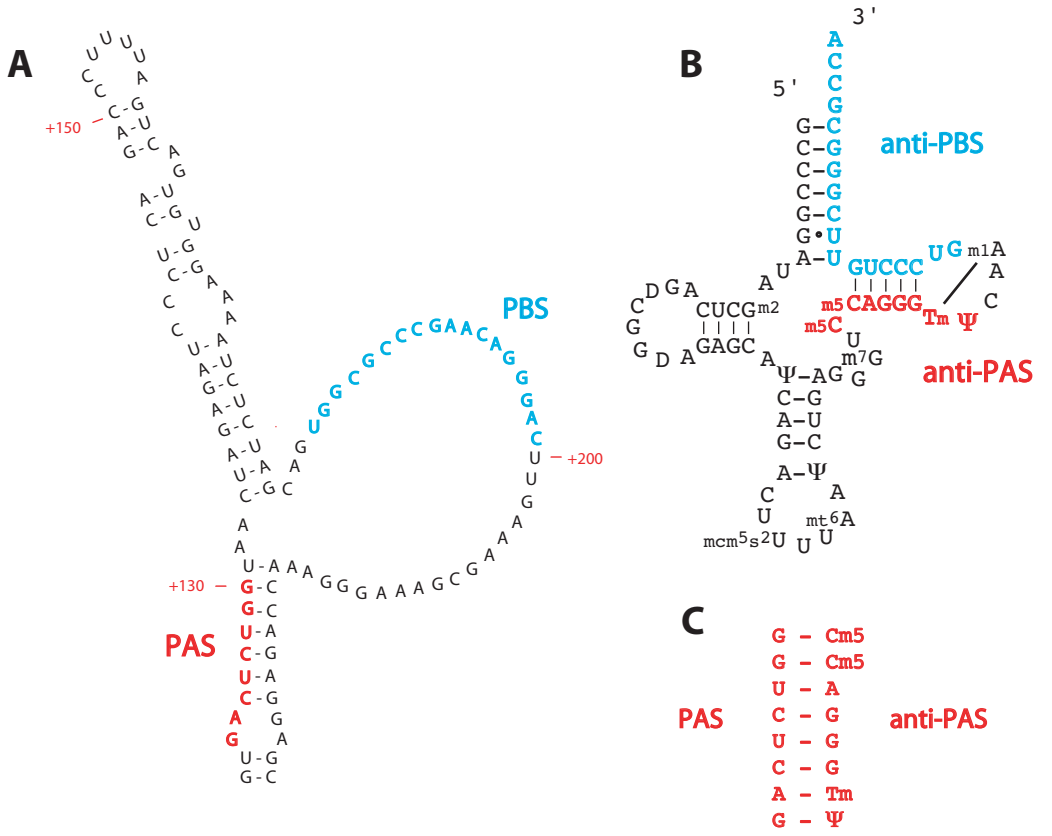


FIGURE 2.13 – Modèle de l'interaction PAS/anti-PAS
A : Structure secondaire autour du PBS de l'ARN viral de HXB2 (nts +121 à +127). Mise en évidence des séquences PBS *en bleu* et PAS *en rouge* (nts +123 à +130). **B** : Structure secondaire de l'ARNt$_3^{Lys}$. Mise en évidence des séquences anti-PBS *en bleu* et anti-PAS *en rouge* (nts +48 à +55). **C** : Appariements proposés pour l'interaction PAS/anti-PAS.

Il ressort des études de l'équipe de Ben BERKHOUT, que la séquence PAS est une séquence régulatrice de l'initiation de la transcription inverse, qui peut être activée ou inhibée par des mutations de la séquence PAS, selon que celles-ci renforcent ou affaiblissent l'interaction avec l'amorce ARNt. De plus, la transcription inverse du VIH-1 sauvage semble être inhibée par la localisation de la séquence PAS dans une structure secondaire stable. L'activation de l'amorce nécessiterait ainsi un réarrangement structural de l'hybride ARNt$_3^{Lys}$/ARN viral afin d'établir l'interaction PAS/anti-PAS. Ce changement conformationnel pourrait impliquer l'activité chaperonne de la NC et jouer un rôle dans le bon déroulement temporel de la transcription inverse [232].

Bien que de nombreux résultats viennent conforter l'importance de la séquence PAS sur l'initiation de la transcription inverse [248, 249], l'interaction directe PAS anti-PAS n'a pas été mise en évidence par des techniques de sondes chimiques, tant *in vitro* qu'*ex vivo* [235, 243]. De ce fait, l'importance de la séquence PAS sur l'initiation de la transcription inverse reste toujours contestée.

Chapitre 3

Rôle de la NC dans la formation de l'hybride $ARNt_3^{Lys}$/PBS

3.1 Problématique et résumé de l'Article 1

Les travaux antérieurs réalisés au laboratoire par Carine Tisné [114–116] ont permis de proposer un mécanisme détaillé du processus d'hybridation de l'$ARNt_3^{Lys}$ avec le PBS (Figure 2.11). Cette étude laissait ouverte deux questions non résolues — à savoir d'une part l'ambiguïté persistante quant au site de nucléation entre le PBS et l'ARNt amorce — et d'autre part l'implication précise de la protéine de nucléocapside dans l'étape cruciale d'ouverture de la structure de l'$ARNt_3^{Lys}$.

En ce qui concerne le site de nucléation, une apparente contradiction existait entre les travaux menés au laboratoire par RMN et ceux réalisés par l'équipe de Karin Musier-Forsyth [170, 203]. L'équipe de Karin Musier-Forsyth propose en effet que la nucléation ait lieu au niveau de l'extrémité 3′ CCA de l'ARNt. Alors que les travaux de Carine Tisné suggèrent que la nucléation ait lieu au pied de la tige acceptrice.

Le point de nucléation a été observé sur la base d'expériences RMN hétéronucléaires dans lesquelles un seul des partenaires (*i.e.* l'$ARNt_3^{Lys}$) est marqué ^{15}N, l'autre (*i.e.* le PBS) n'étant pas marqué. Etant donné que seuls les protons iminos des bases G et U appartenant au partenaire marqué sont observables dans ce type d'expérience, si le point de nucléation a bien lieu au niveau de l'extrémité 3′ CCA de l'ARNt, ceci ne pouvait être observé. Afin de réconcilier les deux études, l'expérience réciproque dans laquelle cette fois le PBS est marqué ^{15}N et l'$ARNt_3^{Lys}$ ne l'est pas a été réalisée et a permis de montrer que la nucléation avait lieu à la fois au niveau de l'extrémité 3′ CCA et au niveau du bras TΨC, c'est à dire aux deux extrémités de la séquence venant s'apparier avec le PBS.

Même si l'implication de la protéine de nucléocapside dans l'ouverture de la structure de l'ARNt amorce ne fait aucun doute, les bases moléculaires de son intervention restaient inconnues. En effet, l'interaction entre la NC et

l'$ARNt_3^{Lys}$ a été décrite comme non spécifique, puisqu'il existe plusieurs sites de fixation de la NC sur l'$ARNt_3^{Lys}$ et que les paramètres de liaison sont fort peu dépendants de l'ARNt choisi [183]. Cependant la structure de l'ARNt amorce est fortement modifiée au cours du processus d'hybridation. Par exemple, les interactions tertiaires entre les boucles TΨC et D sont rompues dans l'hybride final. Ainsi l'étude des propriétés de fixation de la NC sur différents fragments de l'$ARNt_3^{Lys}$ pourrait mettre en évidence des sites de fixation forts et spécifiques, mais dont les motifs de reconnaissance sont cachés dans la structure native de l'ARNt et qui n'apparaissent qu'au cours de la formation de l'hybride.

Nous avons donc étudié par spectroscopie de fluorescence et par RMN l'interaction entre la NC et diverses tiges-boucles d'ARN mimant le bras TΨC ou le bras D de l'$ARNt_3^{Lys}$. Nous avons ainsi mis en évidence un site de fixation fort et spécifique de la NC au niveau du bras D de l'$ARNt_3^{Lys}$.[1] La mise en évidence de cette interaction permet d'envisager l'implication précise de la NC dans la formation de l'hybride $ARNt_3^{Lys}$/PBS. Elle facilite le transfert de brin grâce à son extension N-terminale basique. Elle permet ensuite l'ouverture de la structure tertiaire de l'ARNt en venant piéger le bras D de l'$ARNt_3^{Lys}$. Cette interaction ne risque pas d'interférer avec le processus d'hybridation proprement dit puisque le bras D n'y participe pas. On peut aussi raisonnablement penser que cette interaction déplace l'équilibre vers la formation de l'hybride en venant contrebalancer la perte des interactions tertiaires qu'effectuaient les bases de la boucle D dans l'ARNt natif.

3.2 Article 1 | Barraud *et al.* Biochimie (2007)

1. K_d = 30 nM à 100 mM NaCl. De plus, celui-ci est insensible à la force ionique entre 25 mM et 100 mM NaCl

Available online at www.sciencedirect.com

ScienceDirect

Biochimie 89 (2007) 1204–1210

BIOCHIMIE

www.elsevier.com/locate/biochi

New insights into the formation of HIV-1 reverse transcription initiation complex

Pierre Barraud [a], Cyril Gaudin [b], Frédéric Dardel [a], Carine Tisné [a,*]

[a] *Laboratoire de Cristallographie et RMN biologiques, Université Paris Descartes, CNRS, UMR 8015, 4 avenue de l'Observatoire, 75006 Paris, France*
[b] *Laboratoire de Chimie et de Biologie Structurale, ICSN, CNRS, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette, France*

Received 11 December 2006; accepted 29 January 2007
Available online 20 February 2007

Abstract

HIV-1 reverse transcriptase uses the host $tRNA_3^{Lys}$ as a primer for the synthesis of the minus DNA strand. The first event in viral replication is thus the annealing of tRNA to the primer binding site (PBS) in the 5′ UTR of the viral RNA. This event requires a major RNA rearrangement which is chaperoned by the viral NC protein. The binding of NC to nucleic acids is essentially non-specific, however, NC is known to bind selectively to hairpins located in the 5′ region of the viral RNA. In a previous study, using an NMR approach in which the reaction is slowed down by controlling temperature, we were able to follow details in this RNA unfolding/refolding process and to uncover an intermediate state. We showed that annealing initiates at the junction between the acceptor and the TΨC stems, and that, at physiological temperature, complete annealing is reached only in the presence of NC, probably when the zinc fingers contact the TΨC/D loops.

In the present work, we have refined our model of the formation of the $tRNA_3^{Lys}$/PBS duplex. First, we show that annealing can initiate both from the single-stranded CCA 3′-end bases of the acceptor stem and from the bases in the TΨC stem. Secondly, by NMR and fluorescence spectroscopy, we have studied the complex between the NC protein and RNA hairpins that mimic the D and T arms of the $tRNA_3^{Lys}$. Interestingly, the NC protein shows strong and specific binding to the D arm of $tRNA_3^{Lys}$, which could explain the overall annealing mechanism.
© 2007 Elsevier Masson SAS. All rights reserved.

Keywords: HIV; Reverse transcription; $tRNA_3^{Lys}$; Nucleocapsid; NMR

1. Introduction

In retroviruses, initiation of reverse transcription is primed by a cellular tRNA that is encapsidated in viral particles. $tRNA_3^{Lys}$ is the natural primer of all immunodeficiency viruses, including the type 1 human immunodeficiency virus (HIV-1) [1–3]. The HIV-1 nucleocapsid protein (NC) is a short, basic, nucleic acid binding protein with two zinc finger domains, each containing the invariant CCHC metal-ion binding motif. The mature protein (55 amino acid residues) is produced by proteolytic cleavage of the Gag precursor and is found in the interior of the virus particle, where it is tightly associated with genomic RNA. NC or the NC domain in Gag is involved in multiple functions during the virus replication cycle, including genomic RNA packaging, primer placement on viral RNA, reverse transcription, and integration. Many of these functions rely on the nucleic acid chaperone activity of NC, i.e. its ability to catalyse nucleic acid conformational rearrangements that lead to the most thermodynamically stable structure (for recent reviews [4–6]).

In a previous study [7], we were able to observe the progressive formation of the HIV-1 reverse transcription initiation complex using heteronuclear NMR of RNA–RNA–protein complexes under controlled temperature conditions. In particular, we have identified a nucleation site, at the end of the acceptor stem, where the viral RNA starts invading the $tRNA_3^{Lys}$ structure. In addition, we were also able to characterize the different roles of the nucleocapsid protein during the formation of the initiation

Abbreviations: BME, 2-mercaptoethanol; HIV, human immunodeficiency virus; HSQC, heteronuclear single quantum correlation; NC, nucleocapsid; PBS, primer binding site; TROSY, transverse relaxation optimized spectroscopy.

* Corresponding author. Tel.: +33 1 53 73 15 72; fax: +33 1 53 73 99 25.
E-mail address: carine.tisne@univ-paris5.fr (C. Tisné).

0300-9084/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.biochi.2007.01.016

complex. In our model, the viral NC protein plays a key role by ''unlocking'' stable 3D structure of the primer tRNA through specific interaction with the D/TΨC loops. Therefore, at the end of this study, the question of the involvement of the 3′-end unpaired CCA bases of $tRNA_3^{Lys}$ primer in the annealing process was still open. In addition, the mechanism of the specific 3D structure opening was still not clearly understood.

In the present report, we show that the annealing can proceed from two starting points in the $tRNA_3^{Lys}$ primer, one at the beginning of the acceptor stem and the second at the end of the T stem. Both sites can be used by the viral RNA to begin its annealing with the 18 complementary bases of $tRNA_3^{Lys}$ acceptor and T stems. The presence of NC protein is not necessary at this step, at least in vitro, but is essential to open the 3D structure of tRNA. The melting of the interaction at the level of the D/TΨC loops that locks the structure of the tRNA is possible through a specific interaction between the D loop and the NC protein.

2. Materials and methods

2.1. Samples

The PBS (18 nucleotides) was ^{15}N-labelled on its guanine nucleotides by in vitro transcription from an oligonucleotide template containing a 2′-*O*-methyl-G in position 2 (DNA template: 5′-G(2′OmeG)TCCCTGTTCGGGCGCCACTATAGTG AGTCGTATT-3′) [8]. Labelled nucleotides were purchased from Spectra Gases Inc. The 20 nucleotides RNA obtained was then purified by electrophoresis. After electroelution and ethanol precipitation, the RNA was resuspended, microdialysed against a first buffer (10 mM sodium phosphate, pH 6.4) and then against a second buffer (2 mM sodium phosphate, pH 6.4) and freeze dried. The NMR sample (0.2 mM) was finally prepared by dissolving in 300 μL H_2O containing 10% ^2H_2O

The recombinant 55-residue NC protein (HIV-1 strain NL4-3) was overexpressed from the bacterial expression vector pRD2, which contains the NC coding region from HIV-1 strain NL4-3 subcloned into pET-3a (Novagen, WI). This plasmid was kindly given by M.F. Summers group [9,10]. pRD2 was transformed into *Escherichia coli* strain BL21(DE3) pLysE and the overexpressed protein was purified as previously described [9,10].

The small RNAs oligonucleotides (Fig. 2) were purchased from Dharmacon Research with 2′-*O*-bis (acetoxyethoxy)-methyl (ACE) protection. The samples were deprotected by following manufacturer recommendations and dialysed several times against deionized water. The pH was adjusted between 6 and 6.5 before freeze-drying.

Human $tRNA_3^{Lys}$ was expressed in vivo from a recombinant plasmid and purified as previously described [11].

2.2. Fluorescence titrations

Fluorescence measurements were performed at 30.0 °C on a JASCO spectrofluorimeter. Excitation and emission wavelengths were 295 nm and 345 nm, respectively. The excitation and emission bandwidths were 5 nm.

Fluorescence titrations experiments were performed by adding increasing concentrations of nucleic acid to a fixed amount of NC protein (1 μM) in buffers with different ionic strength (25 mM Na-acetate pH 6.5, 0.1 mM BME, 0.1 mM $ZnCl_2$ for two salt concentrations, 25 mM and 100 mM NaCl, respectively). Fluorescence intensities were corrected for dilution and were fitted using Eq. (1) assuming that the n binding sites on the protein are totally independent and have the same dissociation constant K_d. Confidence limits on the K_d were estimated by Monte-Carlo sampling using the MC-Fit program [12].

$$I = I_0 - \frac{I_0 - I_\infty}{2nN_t}\left(K_d + L_t + nN_t - \sqrt{(K_d + L_t + nN_t)^2 - 4L_t nN_t}\right) \quad (1)$$

where I_0, fluorescence intensity without RNA; I, fluorescence intensity at a given concentration of RNA; I_∞, fluorescence intensity at the plateau; n, number of RNA binding sites on the protein; L_t, total concentration of RNA, N_t, total concentration of protein.

2.3. NMR Spectroscopy

NMR data were recorded on a Bruker Avance DRX600 spectrometer equipped with a standard triple resonance probe and z-axis gradients, processed with NMRPipe/NMRDraw [13] and analyzed with SPARKY (T.D. Goddard and D.G. Kneller, SPARKY 3, University of California, San Francisco).

The formation of the $tRNA_3^{Lys}$/PBS duplex was followed in a $^1H-^{15}N$ TROSY experiment [14] carried out with 256 t1 increments and 1024 t2 data points. The spectral widths were 4194 Hz and 700 Hz in the proton and the nitrogen dimensions, respectively. Assignments of imino groups of the $tRNA_3^{Lys}$/PBS duplex were previously performed [7].

Footprint experiments of small RNAs on the uniformly ^{15}N-labelled NC protein were performed using echo-anti echo HSQC experiments [15] at 30 °C, at a concentration of 0.5 mM of each species in NMR buffer (25 mM Na-d_3-acetate pH 6.5, 25 mM NaCl, 0.1 mM BME, 0.1 mM $ZnCl_2$, 10% ^2H_2O). Each partner was prepared at a concentration of about 2 mM in NMR buffer, and the required volumes of each sample were mixed together, immediately vortexed and incubated at 30 °C for 30 min before NMR data acquisition.

Assignment of NC amide cross-peaks was determined using standard ^{15}N-based strategy.

3. Results and discussion

3.1. Where are the nucleation sites for $tRNA_3^{Lys}$/PBS duplex formation?

In a previous study [7], we showed that the melting process starts at the junction between the acceptor and T stems, at the

level of weak AU and GU base pairs which can be easily disrupted and hence allow subsequent invasion of the $tRNA_3^{Lys}$ structure by the viral RNA. However, our experimental approach, which relied on the NMR observation of imino protons, did not allow observation of PBS binding to the unpaired 3′-end of the tRNA, since the imino protons of the first three base pairs are carried by the viral RNA and only $tRNA_3^{Lys}$ was ^{15}N-labelled. To further describe the very first event in the $tRNA_3^{Lys}$/PBS annealing, we have thus performed the reciprocal experiment where, this time, the PBS is ^{15}N-labelled and the tRNA is unlabelled. The annealing process is then followed by monitoring the imino groups of the PBS (Fig. 1) for which the assignment of proton NMR signals was already performed [7]. Fig. 1b shows the first base pairs that are stable enough to be observed in TROSY experiment, namely G183 and G194. The imino group signals of the free PBS are still mostly present, and only a small part of the PBS has begun to anneal to $tRNA_3^{Lys}$. In addition, after 2 h at 25 °C, in one-dimensional ^1H spectrum where we can further observe non-labelled signals and more labile protons, the signals of U64, U66, G180, G192 and G193 in the duplex can be observed (Fig. 2b) and are rather broad, explaining that G192 and G193 are not yet observed in the TROSY spectrum (Fig. 1b). In addition, in NOESY experiment carried out at 25 °C, the NOE connections between U64 and G192, and between G192 and G193 are unambiguously observed (data not shown). After heating to 70 °C and cooling back to 15 °C (Fig. 1c), all guanine signals corresponding to the duplex are present in TROSY experiment and their chemical shifts are identical to those previously assigned through a NOESY/TROSY strategy [7].

This shows that two distinct parts of $tRNA_3^{Lys}$ are used by the viral RNA to grip its reverse transcription primer: G183 and G194, respectively, contact C72 and C61 on $tRNA_3^{Lys}$, at either end of the acceptor stem. This result reconciles apparently conflicting data, ours [6] and those of Hargittai et al. [16], which suggested two different initiation sites for annealing, either the centre of the cloverleaf structure or the 3′-end unpaired bases of the primer tRNA. Our current data indicate that apparently both sites are being used for annealing. It remains to be established whether the same viral RNA molecule can simultaneously invade the tRNA acceptor stem from both ends or whether some initiate at center of the tRNA while others ''zip-in'' from the 3′-end, in a statistical manner. Interestingly, within the $tRNA_3^{Lys}$ tertiary structure [17], C72 and C61 are distant from each other by 11 nucleotides, i.e. exactly one A-helix turn apart and are thus located on the same side of the tRNA acceptor–TΨC helix, i.e. in a favourable orientation for simultaneous invasion.

3.2. NC protein binds specifically to the D arm of $tRNA_3^{Lys}$

We have previously shown that the presence of NC protein is necessary to promote the complete melting of the TΨC/D loop–loop interaction that is required to obtain a full annealing [7] of $tRNA_3^{Lys}$ on the PBS. Indeed, the NMR signals of the T54 and Ψ55 N3H that make hydrogen bonds with A58 base and phosphate, respectively, are conserved until the later stages of the annealing process. Heating to 80 °C in the absence of NC or NC-mediated annealing at 37 °C is required to destroy these interactions. This activity is nucleocapsid-specific and cannot be reproduced using poly-L-lysine, suggesting that the zinc knuckles could play an important part in this essential step of the viral cycle. This also strongly suggested that there existed an NC-specific binding site in the TΨC/D loop region of the primer $tRNA_3^{Lys}$. To investigate this hypothesis, we have studied the interaction of the NC protein with RNA hairpins mimicking the D and the T arms (Fig. 3) by NMR and fluorescence spectroscopy. Indeed, the multiplicity of NC binding sites on $tRNA_3^{Lys}$ rendered analyses using the full length molecule intractable, even when stringency was raised by increasing salt concentration (data not shown). This is a result of the binding dynamics, as we previously showed that the NC protein slides from one site to another [18]. Reducing the target RNA length prevents this process. Binding assays to the various RNA using the intrinsic fluorescence of NC tryptophan 37 are summarized in Table 1. As a reference, the same experiments were also carried out on the SL3 HIV-1 genomic packaging signal that makes specific interaction with NC [9]. As previously described for packaging domain loops of HIV-1 RNA [19–21], tightly bound RNA quenched nearly all the fluorescence of NC. The NMR and fluorescence analyses were initially performed at 25 mM NaCl, i.e. conditions similar to those used by Summers and co-workers to resolve the NMR structure of the SL3/NC complex [9]. Ionic strength was then increased to investigate the binding specificity. TΨC hairpin bound to NC, however, increasing salt concentration strongly destabilised this interaction (Fig. 4d and Table 1). We also tested TΨC hairpins of alternate lengths (15 and 21 nucleotides) to investigate whether the number of NC binding sites or the stability of the T stem could play a role. The dissociation constant increased by a factor of two when the hairpin size was shortened from 21 nucleotides to 15 (Table 1). Longer stems thus improve the affinity and also increase the apparent number of NC binding sites. One can notice that the stoechiometry at low salt concentration for TΨC2 (Table 1) is quite unusual, but it was previously showed that non-specific interactions at low ionic strength between highly charged RNA molecule and NC can bias the number of binding site [21].

In contrast to what is observed with the TΨC hairpins, NC binding to the D arm was tighter and much less sensitive to the salt concentration (Fig. 4d). This behaviour is in line with what was observed for the specific binding of NC to the SL3 hairpin, for which a similar affinity and salt independence are seen (Table 1). All these results show that the NC specifically binds to the D arm, whereas binding to the T arm is weaker and dominated by electrostatic effects, as evidenced by the salt dependence.

This different behaviour in NC binding of these two hairpins is observable in the corresponding NMR footprints on

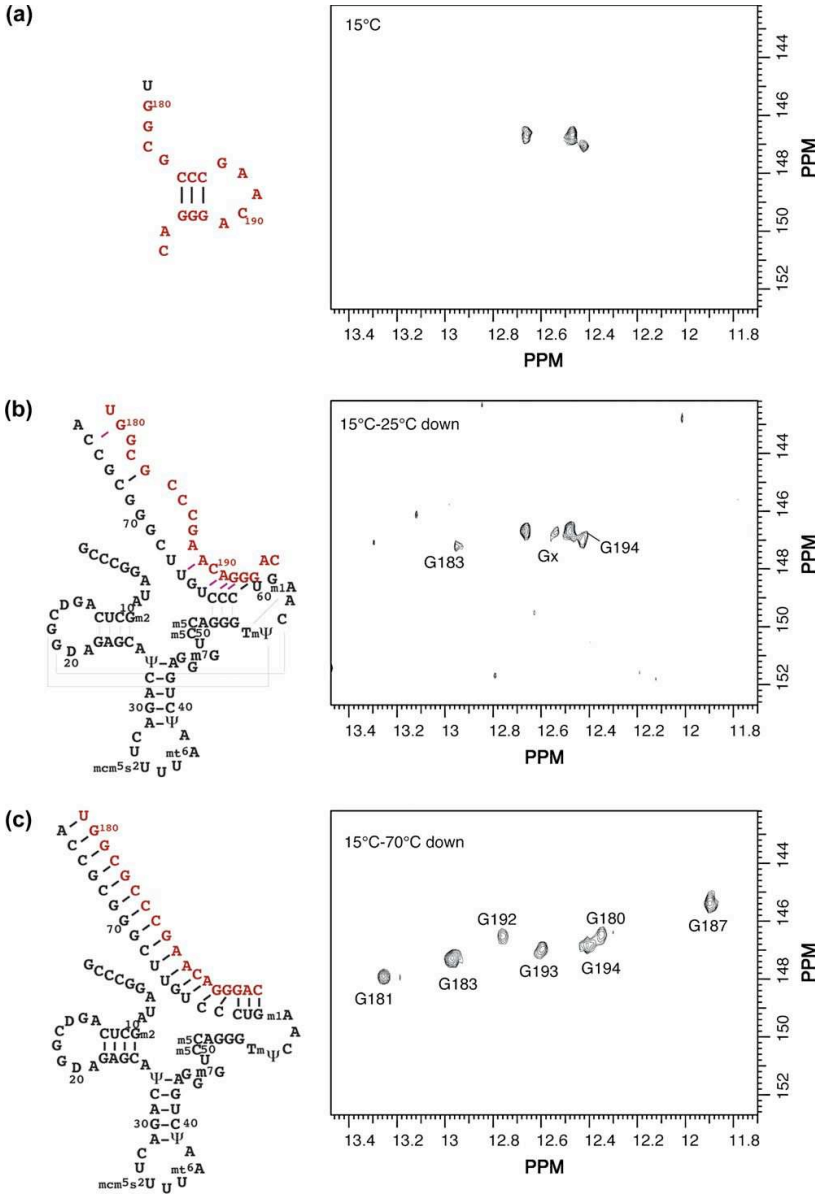


Fig. 1. (a) Secondary structure of the PBS (left), TROSY experiment of the ^{15}N-PBS alone at 15 °C (right). (b) Secondary structure of tRNA$_3^{Lys}$/PBS complex (left): in black, base pairs for which the imino groups is observed in TROSY experiment, in pink, base pairs for which the imino proton is observed in the 1D spectrum, TROSY experiment of a 1/1 tRNA$_3^{Lys}$/^{15}N-PBS at 15 °C after heating the sample to 25 °C (right). Gx cannot be assigned and probably belong to an intermediate structure. (c) Secondary structure of tRNA$_3^{Lys}$/PBS complex (left), TROSY experiment at 15 °C after heating the sample to 70 °C (right).

^{15}N-labelled NC, even at 25 mM NaCl. At this ionic strength, the dissociation constant for the NC binding on T arm only differ from that on the D arm by a factor two. On the NMR spectra (Fig. 4a and c), only half of amide peaks are observed for the T footprint. This is typical of chemical exchange in the μs–ms time scale, where NMR signals can disappear due to severe line broadening [22]. This is not surprising as it is known that the binding of NC on nucleic acids is very dynamic and NC is mobile on them especially when the binding is not specific. Contrary to the T arm, approximately 40 amide cross-peaks are still observed for the D arm. That means, chemical exchange reaches almost the slow exchange regime, which is characteristic of tight binding. As previously described for the interaction with the packaging sequence [9,23], the NC

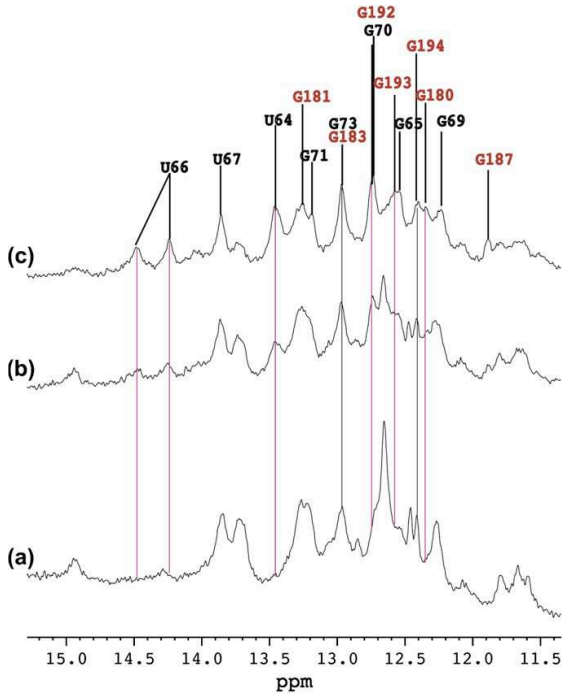


Fig. 2. One-dimensional proton NMR spectra at 15 °C: (a) just after the mixture between $tRNA_3^{Lys}$ and the PBS, (b) after heating the sample for 2 h at 25 °C and cooling it back to 15 °C to recorder the spectrum, (c) after heating the sample to 70 °C and cooling it back to 15 °C.

structure is significantly modified upon RNA binding and the NMR spectrum of the complex needs to be completely re-assigned. Taken together, these fluorescence and NMR observations indicate that there exists a specific binding site

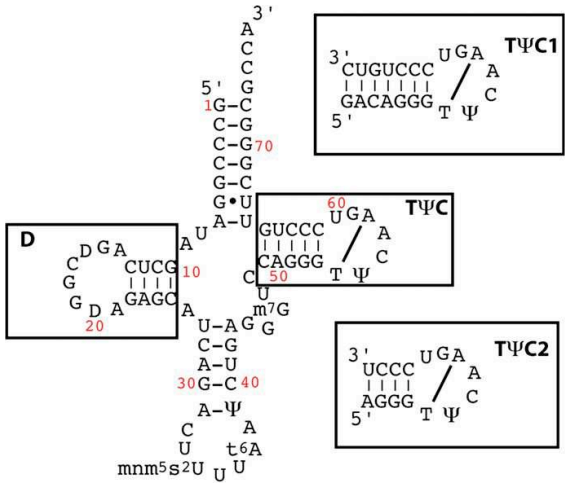


Fig. 3. Secondary structures of $tRNA_3^{Lys}$ and of hairpins that were designed to mimic $tRNA_3^{Lys}$ D and T arms. The hairpins that mimic $tRNA_3^{Lys}$ T arm carry the T and Ψ base modifications whereas the D hairpin is not modified.

of NC on $tRNA_3^{Lys}$ in the TΨC/D core and it is located in the D hairpin.

Based on both fluorescence and NMR structural studies performed to date [9,19–21,23], it appears that the CCHC-type zinc knuckle domains of NC contain specifically positioned hydrophobic residues that form an ideal binding surface for exposed G residues within nucleic acid sequences. More accurately, stacking interactions between Phe16 in the N-terminal finger and Trp37 in the C-terminal zinc finger and nucleic acid bases have been detected by fluorescence spectroscopy and have been proposed to be a major driving force for NC-nucleic acid interactions [19,24]. Indeed, NMR structures of SL2/NC and SL3/NC complexes [9,23] show close interactions between Phe16 and Trp37 and guanine residues in single-stranded regions of SL2 and SL3 RNAs. In addition to the preferential binding of NC's zinc fingers to the single-stranded hairpin loop regions, the cationic N-terminal domain of NC binds to the double-stranded stem of the hairpin. From these data, the NC's zinc fingers are likely to interact with the guanine residues in the D loop whereas the N-terminal residues would interact with the D stem. The D loop contains three guanine nucleotides (G15, G18 and G19) whereas the T loop contains only one (Fig. 3). G15 makes a base pair with C48 and is rather buried within the 3D structure of $tRNA_3^{Lys}$ [17]. G19 makes a rather weak base pair with C56, as two alternate conformations are observed in the crystal structure and, as a consequence of this mobility, could be readily accessible to interactions with NC. G18 makes a hydrogen bond with Ψ55 (O6) and interestingly, the presence of NC in the annealing process is required to induce disappearance of the Ψ55 (N3H) signal. Moreover, bases 16, 17 and 20 that surround G18 and G19 are completely exposed to the solvent that render nucleotide in the D loop easily accessible. Therefore, within the D loop, G18 and G19 appear as likely candidates as being part of the specific NC binding site. Moreover, in the context of the annealing process between the PBS and the primer $tRNA_3^{Lys}$, it makes more sense to bind to the D loop rather than the TΨC loop, where the annealing to the PBS is taking place.

4. Conclusion

In this paper, the use of selective labelling of the primer binding site allowed us to characterize the nucleation points of the annealing process between the PBS and the $tRNA_3^{Lys}$. Moreover, we used small hairpins to mimic the D and T arms of $tRNA_3^{Lys}$ to determine the specific role of NC in the formation of $tRNA_3^{Lys}$/PBS duplex recognized by the HIV reverse transcriptase. Studying separate hairpins cannot obviously mimic the entire tRNA, in particular at the level of the TΨC/D loops interaction. However, in the annealing process, there is a step where the TΨC/D loops interaction are broken and where the NC protein plays a crucial role. NC will thus pull the equilibrium toward the "open state" by stabilising it. In this context, the study of separate hairpins makes perfect sense, since they are likely to be an intermediate state in the unfolding pathway. We previously demonstrated that unless

Table 1
Binding Parameters of NC_{55} to RNAs stem-loop measured by fluorescence spectroscopy titration

Ionic strength		25 mM NaCl			100 mM NaCl		
RNA	RNA length	K_d[a] (nM)	n[b]	Q_{max}[c] (%)	K_d[a] (nM)	n[b]	Q_{max}[c] (%)
D	16	33 (25–37)	1.1 (±0.1)	98 (±1)	28 (28–35)	1.0 (±0.1)	96 (±1)
TΨC	17	61 (55–68)	1.2 (±0.1)	88 (±1)	301 (272–319)	1.0 (±0.1)	90 (±1)
SL3	20	20 (20–27)	1.0 (±0.1)	99 (±1)	12 (11–15)	1.0 (±0.1)	97 (±1)
TΨC1	21	19 (11–22)	2.0 (±0.1)	92 (±2)	87 (85–88)	1.6 (±0.1)	92 (±1)
TΨC2	15	46 (44–53)	0.6 (±0.1)	92 (±1)	379 (335–389)	0.9 (±0.1)	90 (±1)

[a] The dissociation constant K_d is indicated with the 96% confidence interval in parenthesis.
[b] Number of sites for NC_{55} on the RNA stem-loop.
[c] Maximum extent of fluorescence quenching.

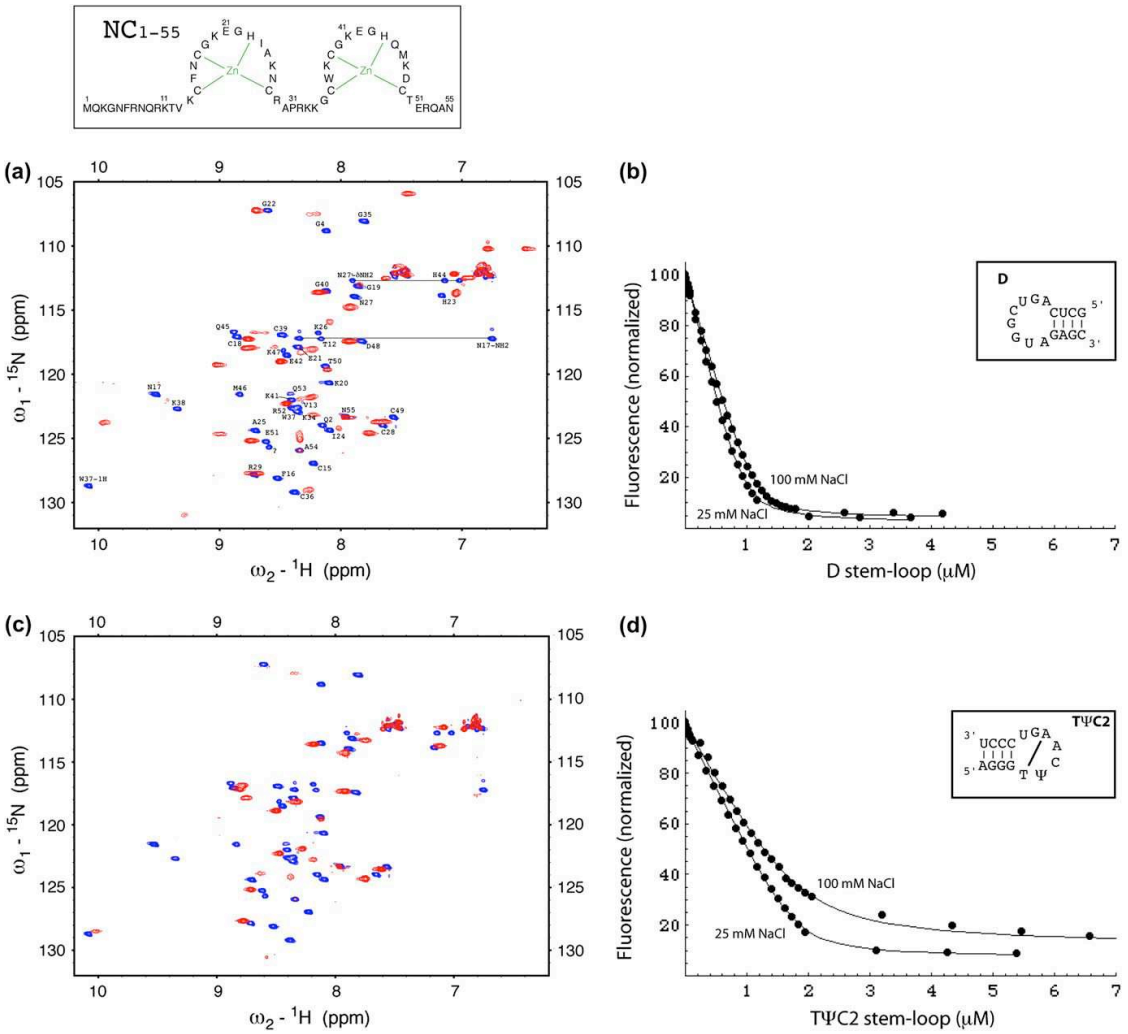


Fig. 4. HSQC experiments at 30 °C of ^{15}N-labelled NC with the D (a) and TΨC2 (c) hairpins, in blue the reference spectrum and in red the spectrum of the equimolar 1/1 mixture. Fluorescence titrations of NC protein/D hairpin (b) and NC protein/TΨC2 hairpin (d) obtained by adding increasing concentrations of nucleic acid to a fixed amount of NC protein (1 μM) in buffers with two ionic strength: 25 mM and 100 mM NaCl.

NC is added to the primer−PBS mixture, complete melting of $tRNA_3^{Lys}$ cannot be obtained. The specific interaction between NC and the D loop that we observed here could be the key factor driving the unfolding process, thereby allowing complete annealing of $tRNA_3^{Lys}$ to the PBS. The NMR structure determination of the NC protein/D hairpin complex is under investigation and should provide useful information on the NC-mediated annealing process of $tRNA_3^{Lys}$ and on the design of anti-NC or anti-annealing drugs.

Acknowledgements

P. Barraud is supported by a studentship from Ministère de la Recherche. This work was supported by the French AIDS national Agency (ANRS) and 'Ensemble contre le SIDA' (Sidaction).

References

[1] S. Wain-Hobson, P. Sonigo, O. Danos, S. Cole, M. Alizon, Nucleotide sequence of the AIDS virus, LAV, Cell 40 (1985) 9−17.
[2] L. Kleiman, S. Caudry, F. Boulerice, M.A. Wainberg, M.A. Parniak, Incorporation of tRNA into normal and mutant HIV-1, Biochem. Biophys. Res. Commun. 174 (1991) 1272−1280.
[3] R. Marquet, C. Isel, C. Ehresmann, B. Ehresmann, tRNAs as primer of reverse transcriptase, Biochimie 77 (1995) 113−124.
[4] L. Kleiman, R. Halwani, H. Javanbakht, The selective packaging and annealing of primer $tRNA_3^{Lys}$ in HIV-1, Curr. HIV Res. 2 (2004) 163−175.
[5] J.G. Levin, J. Guo, I. Rouzina, K. Musier-Forsyth, Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism, Prog. Nucleic Acid Res. Mol. Biol. 80 (2005) 217−286.
[6] C. Tisné, Structural bases of the annealing of primer tRNA(3Lys) to the HIV-1 viral RNA, Curr. HIV Res. 3 (2005) 147−156.
[7] C. Tisné, B.P. Roques, F. Dardel, The annealing mechanism of HIV-1 reverse transcription primer onto the viral genome, J. Biol. Chem. 279 (2004) 3588−3595.
[8] C. Kao, M. Zheng, S. Rudisser, A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase, RNA 5 (1999) 1268−1272.
[9] R.N. de Guzman, Z.R. Wu, C.C. Stalling, L. Pappalardo, P.N. Borer, M.F. Summers, Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element, Science 279 (1998) 384−388.
[10] B.M. Lee, R.N. de Guzman, B.G. Turner, N. Tjandra, M.F. Summers, Dynamical behaviour of the HIV-1 nucleocapsid protein, J. Mol. Biol. 279 (1998) 633−649.
[11] C. Tisné, M. Rigourd, R. Marquet, C. Ehresmann, F. Dardel, NMR and biochemical characterisation of recombinant human $tRNA_3^{Lys}$ expressed in *E. coli*: identification of post-transcriptional nucleotide modifications required for efficient initiation of HIV-1 reverse transcription, RNA 6 (2000) 1403−1412.
[12] F. Dardel, MC-Fit: using Monte-Carlo methods to get accurate confidence limits on enzyme parameters, Comput. Appl. Biosci. 10 (1994) 273−275.
[13] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J. Biomol. NMR 6 (1995) 277−293.
[14] J. Weigelt, Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments, J. Am. Chem. Soc. 120 (1998) 10778−10779.
[15] J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky, S.J. Glaser, O.W. Sorensen, C. Griesinger, A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients, J. Biomol. NMR 4 (1994) 301−306.
[16] M.R. Hargittai, R.J. Gorelick, I. Rouzina, K. Musier-Forsyth, Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site, J. Mol. Biol. 337 (2004) 951−968.
[17] P. Benas, G. Bec, G. Keith, R. Marquet, C. Ehresmann, B. Ehresmann, P. Dumas, The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop, RNA 6 (2000) 1347−1355.
[18] C. Tisné, B.P. Roques, F. Dardel, Heteronuclear NMR studies of the interaction of tRNALys3 with HIV-1 nucleocapsid protein, J. Mol. Biol. 306 (2001) 443−454.
[19] C. Vuilleumier, E. Bombarda, N. Morellet, D. Gerard, B.P. Roques, Y. Mely, Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study, Biochem. 38 (1999) 16816−16825.
[20] M.F. Shubsda, A.C. Paoletti, B.S. Hudson, P.N. Borer, Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein, Biochem. 41 (2002) 5276−5282.
[21] A.C. Paoletti, M.F. Shubsda, B.S. Hudson, P.N. Borer, Affinities of the nucleocapsid protein for variants of SL3 RNA in HIV-1, Biochem. 41 (2002) 15423−15428.
[22] A.G. Palmer 3rd, C.D. Kroenke, J.P. Loria, Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules, Methods Enzymol. 339 (2001) 204−238.
[23] G.K. Amarasinghe, R.N. De Guzman, R.B. Turner, K.J. Chancellor, Z.R. Wu, M.F. Summers, NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition, J. Mol. Biol. 301 (2000) 491−511.
[24] Y. Mely, H. de Rocquigny, E. Piemont, H. Demene, N. Jullian, M.C. Fournie-Zaluski, B. Roques, D. Gerard, Influence of the N- and C-terminal chains on the zinc-binding and conformational properties of the central zinc-finger structure of Moloney murine leukaemia virus nucleocapsid protein: a steady-state and time-resolved fluorescence study, Biochim. Biophys. Acta 1161 (1993) 6−18.

Chapitre 4

Etude par RMN du complexe NC/bras D de l'ARNt$_3^{Lys}$

Le travail décrit dans le CHAPITRE 3 nous a permis de mettre en évidence une interaction forte et spécifique entre la NC du VIH-1 et le bras D de l'ARNt$_3^{Lys}$. Afin d'appréhender finement les bases moléculaires de cette interaction, à l'échelle atomique, nous avons entrepris de résoudre la structure du complexe NC/bras D par RMN en solution.

L'étude précédente nous a aussi montré que les paramètres de liaison de la NC pour les tiges-boucles d'ARN étaient assez fortement dépendantes de la longueur de la tige d'ARN. Ceci avait été observé pour le bras T de l'ARNt$_3^{Lys}$, mais cette étude n'avait pas été menée pour le bras D. Nous avons donc ensuite entrepris une étude plus approfondie de l'influence de la longueur du bras D afin d'optimiser celle-ci en vue de l'étude du complexe par RMN. Des expériences HSQC ^{1}H–^{15}N de la NC marquée ^{15}N en complexe avec des bras D de différentes longueurs ont ainsi été enregistrées car elles constituent une bonne signature du « comportement RMN » de le protéine au sein du complexe.

4.1 Première approche

4.1.1 Spectre HSQC et signature de la protéine

Le spectre HSQC ^{1}H–^{15}N d'une protéine est une bonne signature de son état d'organisation tridimensionnelle. Dans une expérience HSQC ^{1}H–^{15}N, chaque acide aminé de la protéine donne un pic de corrélation correspondant à la fonction amide de la chaîne principale. Quelques acides aminés (Q, N et W) donnent naissance à des signaux supplémentaires issus des chaînes latérales. Ceux-ci sont généralement facilement identifiables, soit par leur déplacement chimique (H indole des tryptophanes), soit par la forme des pics (amides terminaux des glutamines et asparagines). De plus le degré de dispersion des déplacements chimiques des protons amides est un bon indicateur de l'état de repliement de la protéine. Un spectre HSQC constitue donc en quelque sorte « l'empreinte digitale » ou la signature de la protéine.

L'attribution de la NC libre a été réalisée à l'aide des expériences à 3 dimensions NOESY–HSQC ^{1}H–^{15}N et TOCSY–HSQC ^{1}H–^{15}N, à 10 °C selon la méthode classique d'attribution séquentielle des protéines simplement marquées ^{15}N [250, 251], simple adaptation à 3 dimensions de la méthode d'attribution des protéines non marquées [252]. Celle-ci a été effectuée à 10 °C car à plus haute température (25–30 °C), les signaux correspondant à la partie N-terminale (résidus 1 à 13) deviennent très peu intenses voir inobservables sur les spectres. Le bilan d'attribution de la NC est présenté à la FIGURE 4.1.

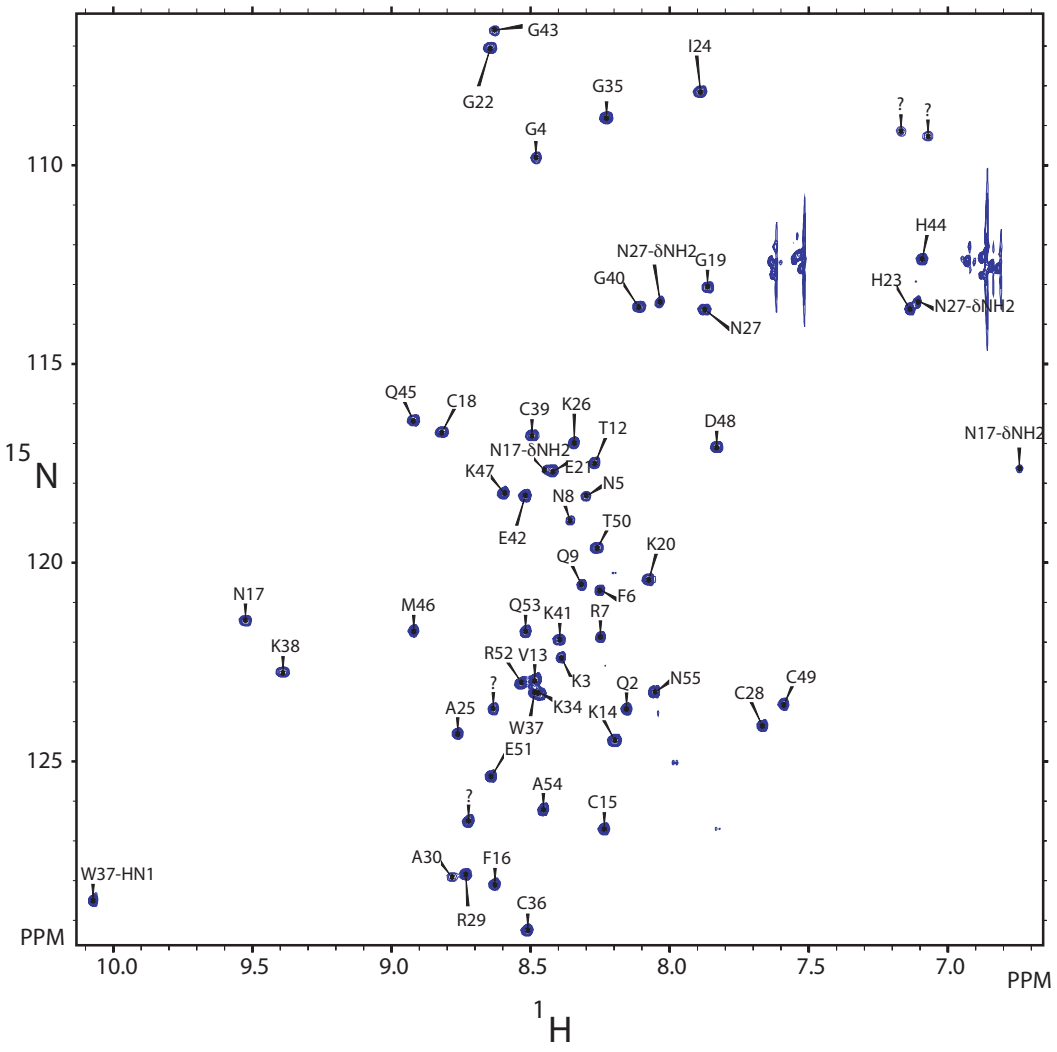


FIGURE 4.1 – Bilan d'attribution de la NC$_{1\text{-}55}$

Bilan d'attribution des protons amides sur le spectre HSQC de la NC$_{1\text{-}55}$ marquée ^{15}N. L'attribution a été réalisée à 10 °C à partir des expériences à 3 dimensions NOESY–HSQC ^{1}H–^{15}N et TOCSY–HSQC ^{1}H–^{15}N. Les « *?* » désignent des signaux non identifiés.

4.1.2 Optimisation de la longueur du bras D

Les paramètres de liaison de la NC sur différentes tiges-boucles mimant le bras D de l'ARNt_3^{Lys} ont été mesurés par spectroscopie de fluorescence, en suivant le même protocole expérimental que celui présenté dans l'Article 1. Trois longueurs de tige ont été choisies. La tige la plus courte, qui comprend 4 paires de bases correspond au bras D proprement dit. Des tiges plus longues (6 et 8 paires de bases) ont été choisies afin de favoriser l'interaction de la partie N-terminale de la NC avec la tige d'ARN. Elles ont aussi permis de réaliser les expériences RMN à plus haute température, sans risquer de rompre les appariements de la tige. Les courbes de titration ainsi que les séquences de ces tiges-boucles sont présentées à la FIGURE 4.2. Les valeurs numériques relatives à ces titrations sont quant-à-elles présentées dans la TABLE 4.1.

force ionique	ARN	K_d [a] (nM)	n[b]	Q_{max} [c] (%)
25 mM NaCl	Bras D16	33 (25–37)	1.1 (±0.03)	98 (±1)
	Bras D20	20 (14–24)	1.1 (±0.02)	98 (±1)
	Bras D24	21 (18–27)	1.0 (±0.04)	97 (±1)
100 mM NaCl	Bras D16	28 (28–35)	1.0 (±0.04)	96 (±1)
	Bras D20	22 (17–25)	1.0 (±0.03)	95 (±1)
	Bras D24	25 (20–36)	1.0 (±0.01)	96 (±1)

TABLE 4.1 – Paramètres de liaison de la NC avec des tiges-boucles Bras D de longueurs variées
[a] La constante de dissociation K_d est indiquée avec un intervalle de confiance à 96 % entre parenthèses. [b] Nombres de sites pour la NC sur les tiges-boucles d'ARN. [c] Extinction maximale de fluorescence.

Conjointement, des expériences HSQC ^{1}H–^{15}N ont été enregistrées pour les différents complexes NC/bras D, afin d'évaluer l'influence de la longueur de la tige sur la qualité des spectres (FIGURE 4.2). Même si les constantes de dissociations évoluent peu lorsque l'on augmente la longueur de la tige (20 nM $< K_d <$ 33 nM à 25 mM NaCl), la qualité des spectres RMN est quant à elle très nettement améliorée lorsque l'on double la longueur de celle-ci. On observe ainsi de nombreux signaux supplémentaires avec la tige la plus longue — la tige-boucle Bras D24. Les expériences RMN présentées dans la suite ont ainsi été réalisées avec la séquence Bras D24, longue de 24 nucléotides, à 25 mM NaCl et à pH 6.5.

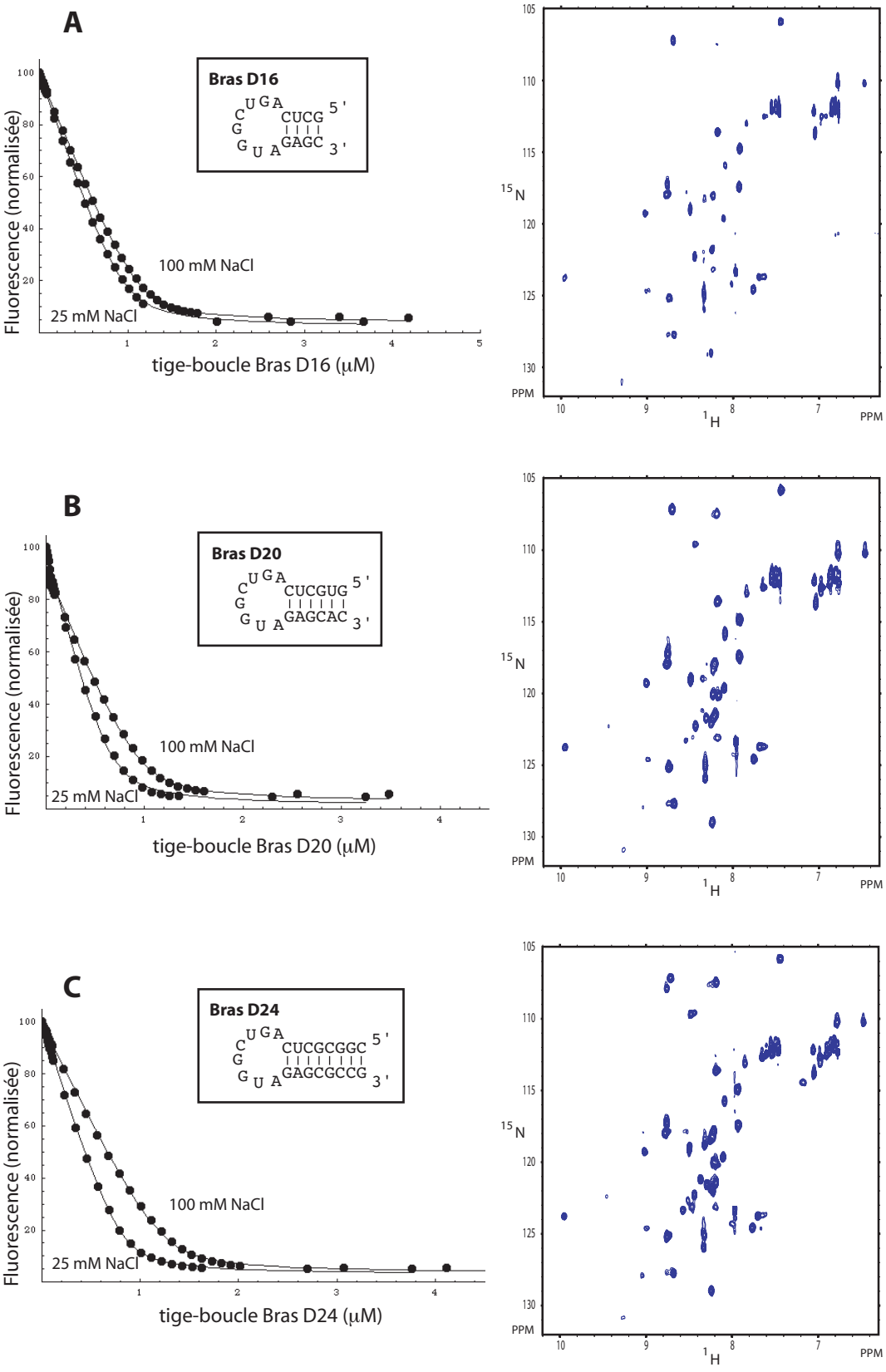


FIGURE 4.2 – Optimisation de la longueur du bras D pour l'étude du complexe NC/bras D par RMN

Les expériences de titration suivies par fluorescence ainsi que les spectres HSQC de la NC ^{15}N en complexe avec 1 équivalent de bras D sont présentés pour trois longueurs de tige. **A** : Bras D16, longueur 16 nt. **B** : Bras D20, longueur 20 nt. **C** : Bras D24, longueur 24 nt.

4.1.3 La structure de la NC est très fortement modifiée dans le complexe avec le bras D

Comparaison des spectres HSQC de la NC libre et liée

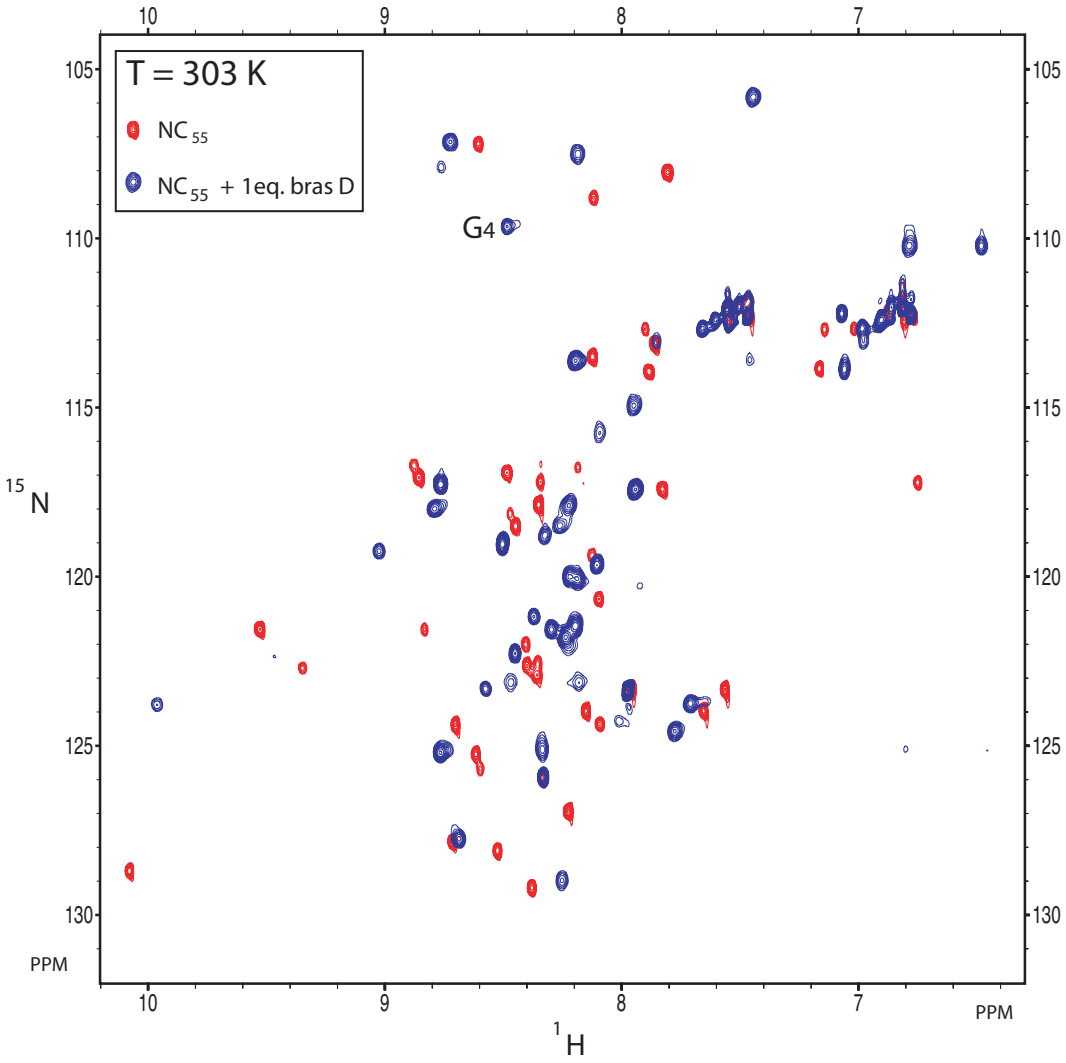


FIGURE 4.3 – Spectres HSQC de la NC ^{15}N libre et en complexe avec le Bras D24 de l'ARNt$_3^{Lys}$
Le spectre HSQC de la NC libre est *en rouge*. Celui de la NC en présence d'un équivalent de Bras D24 est *en bleu*. Les spectres ont été enregistrés à 303 K. Le signal du proton amide de la glycine 4 est reporté sur le spectre.

Lorsque l'on compare les spectres HSQC de la NC libre et de la NC au sein du complexe NC/bras D (FIGURE 4.3), on remarque que pratiquement tous les signaux de la protéine libre voient leur déplacement chimique modifié, puisque l'on observe quasiment aucun des signaux de la protéine libre superposés à ceux de la protéine liée. Cela signifie que l'environnement chimique de l'ensemble des résidus est modifié, et donc que la structure de la protéine est fortement

altérée dans le complexe. Ceci n'est pas très surprenant, puisque la NC est connue pour être une protéine très flexible, tant dans sa partie N-terminale, qu'au niveau du lien séparant les deux doigts de zinc — ceci lui conférant une grande adaptabilité dans la liaison aux acides nucléiques [180].

Outre cette observation qui vient corroborer l'existence d'une association intermoléculaire entre la NC et le bras D de l'ARNt$_3^{Lys}$, deux points importants méritent d'être relevés :

- On observe moins de pics dans le spectre HSQC du complexe. De plus certains pics sont très larges et sont quasiment écrasés au niveau du bruit de l'expérience. Il faut pour obtenir un spectre d'une sensibilité raisonnable enregistrer huit fois plus d'accumulations que pour la NC seule [1]. Ceci est dû à un échange chimique entre la forme libre et la forme liée de la protéine. L'optimisation de la sensibilité des expériences HSQC du complexe NC/bras D a fait l'objet de l'Article 2 et sera détaillé dans le PARAGRAPHE suivant.
- Cependant, il est tout de même possible de voir que certains éléments de la NC doivent être plus structurés dans le complexe que dans la protéine libre. En effet, les spectres sont réalisés à 30 °C et le pic correspondant à la glycine G_4 est visible sur le spectre du complexe, alors qu'elle ne l'est plus à cette température sur le spectre de la protéine seule. Ainsi, de même que dans les complexes NC-SL2 et NC-SL3 [118, 119], il est possible que la partie N-terminale de la NC vienne aussi se structurer dans un des sillons de la tige du bras D (voir PARAGRAPHE 2.1.4).

4.2 Amélioration de la sensibilité des expériences HSQC

4.2.1 Problématique et résumé de l'Article 2

L'étude de complexes par RMN en solution est toujours intimement liée à un phénomène inhérent à l'étude des associations de molécules, le phénomène d'échange. En effet, considérons deux molécules A et B s'associant pour former un complexe A•B selon

$$A + B \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} A\bullet B$$

Un spin S de la molécule A peut se trouver dans deux environnements chimiques très différents selon qu'il se trouve dans la molécule libre A ou dans le complexe A•B, et donc posséder deux déplacements chimiques distincts dans la forme libre et la forme liée. Or les formes libres et liées sont en équilibre, et selon la fréquence à laquelle a lieu l'échange, son incidence sur le signal RMN peut être très différent. La fréquence de l'échange, notée k_{ex}, est reliée aux

1. Le spectre de la NC seule à 0,5 mM a été enregistré avec 16 scans et 256 incréments. Alors que le spectre de la NC en complexe avec le bras D comprend 128 scans et 256 incréments.

constantes cinétiques d'association et de dissociation (k_{on} et k_{off}). Le phénomène d'échange dans son ensemble provient du passage de la forme libre A à la forme liée A•B ainsi que du passage de la forme liée A•B à la forme libre A. Ainsi pour un spin S de la molécule A, la fréquence de l'échange correspond donc à la somme des deux pseudo-constantes d'ordre 1 selon $k_{ex} = k_{on}\,[\mathrm{B}] + k_{off}$.

Lorsque les constantes d'échange appartiennent à la gamme de temps spectrale[2]— ou gamme de temps des déplacements chimiques — le phénomène d'échange introduit une modification importante de l'allure des spectres. L'influence principale de l'échange chimique sur le signal RMN est la contribution additionnelle de l'échange à la relaxation transversale. L'échange chimique induit une perte de cohérence des spins pendant les périodes d'évolution. En effet, le même spin S est porté par l'ensemble des molécules A de l'échantillon, et chaque spin S, sur chacune des molécules A subit des changements de fréquence qui n'ont aucune raison d'être corrélés entre eux. On note R_{ex} la contribution de l'échange à la relaxation transversale.

D'autre part, il est assez intuitif de voir que la mesure du déplacement chimique d'un spin impliqué dans un phénomène d'échange est dépendant de la fréquence d'échange. Ainsi si k_{ex} est telle que le spin S ne subit pas d'échange pendant le temps d'acquisition, on doit pouvoir mesurer le déplacement chimique de chacune des deux formes et l'intensité du signal sera pondérée de la population relative des deux états. Par contre, si k_{ex} est telle que le spin S change « quelques fois » d'environnement chimique au cours de l'acquisition, on doit s'attendre à quelques difficultés pour mesurer ses déplacements chimiques. L'appréciation quantitative de ce « quelques fois » nécessiterait des développements mathématiques sur la théorie de l'échange chimique et de son influence sur le signal RMN. Celle-ci est très bien documentée (pour des revues récentes [253–255]), et nous n'en reporterons ici que les résultats principaux.

Quelques paramètres supplémentaires doivent être introduits pour appréhender le comportement de systèmes en échange chimique. Il s'agit des populations relatives des espèces en équilibre (p_{A} et $p_{\mathrm{A\bullet B}}$), et de la différence de déplacement chimique entre les deux états ($\Delta\omega = \omega_{\mathrm{A}} - \omega_{\mathrm{A\bullet B}}$).

La comparaison de la constante d'échange k_{ex} à la différence de déplacement chimique $\Delta\omega$ définit les régimes d'échange à l'échelle de temps des dé-

2. La gamme de temps spectrale est reliée à la différence de fréquence de résonance de deux spins au sein du spectre. On a la relation suivante entre la différence de fréquence $\Delta\omega_{Hz}$ (en Hz) et la différence de déplacement chimique $\Delta\omega_{ppm}$ (en ppm) :

$$\Delta\omega_{Hz} = \frac{\Delta\omega_{ppm} \times 10^{-6} \times \gamma \mathrm{B}_0}{2\pi}$$

Avec γ le rapport gyromagnétique du noyau considéré et B_0 l'intensité du champ magnétique statique. Ainsi, pour un champ B_0 de 14,1 T, deux protons séparés de 10 ppm correspondent à une échelle de temps spectrale d'environ 150 µs, et deux protons séparés de 0,2 ppm correspondent à une échelle de temps spectrale d'environ 8 ms. L'échelle de temps spectrale correspond donc à des fréquences de l'ordre de la centaine de µs à la dizaine de ms.

placements chimiques [3]. On distingue communément trois régimes d'échange :

- le régime d'échange rapide, pour $k_{ex} > \Delta\omega$
- le régime d'échange intermédiaire, pour $k_{ex} \sim \Delta\omega$
- le régime d'échange lent, pour $k_{ex} < \Delta\omega$

Dans le régime d'échange rapide, on observe une seule résonance pour le spin S, à une fréquence intermédiaire entre ω_A et $\omega_{A\bullet B}$. On observe la résonance du spin S à la fréquence moyenne pondérée des deux états libre et lié de la molécule A, soit $\omega = p_A\omega_A + p_{A\bullet B}\omega_{A\bullet B}$. D'autre part, dans le régime limite d'échange rapide ($k_{ex} \gg \Delta\omega$), la contribution de l'échange à la relaxation transverse dépend de p_A et $p_{A\bullet B}$, de $\Delta\omega$ et de k_{ex} selon

$$R_{ex} = \frac{p_A p_{A\bullet B}\Delta\omega^2}{k_{ex}}$$

Dans le régime d'échange lent, on observe deux résonances pour le spin S, correspondant aux deux états libre et lié de la molécule A. Dans le régime limite d'échange lent ($k_{ex} \ll \Delta\omega$), on observe les deux résonances aux fréquences propres de chaque état, soit à ω_A et $\omega_{A\bullet B}$, et la contribution de l'échange à la relaxation transverse ne dépend pas de la différence de déplacement chimique $\Delta\omega$, mais seulement des populations p_A et $p_{A\bullet B}$ et de la constante d'échange k_{ex}. Ainsi pour la fréquence ω_A, $R_{ex,A} = p_{A\bullet B}k_{ex}$ et pour la fréquence $\omega_{A\bullet B}$, $R_{ex,A\bullet B} = p_A k_{ex}$. Il ressort de ces formules que la composante la moins peuplée est celle qui est la plus élargie par l'échange.

Le régime d'échange intermédiaire correspond à une gamme de constante d'échange k_{ex} dont la valeur est du même ordre de grandeur que la différence de déplacement chimique $\Delta\omega$. C'est dans ce régime intermédiaire que la contribution de l'échange à la relaxation transversale est maximale. On parle de phénomène de coalescence, en référence aux deux raies distinctes caractéristiques du régime d'échange lent qui se rejoignent progressivement lorsque k_{ex} augmente, et fusionnent pour $k_{ex} \simeq \Delta\omega$. Cependant, les raies ne font pas que se rapprocher lorsque l'on passe d'un régime d'échange lent à un régime d'échange intermédiaire. La largeur des raies est considérablement augmentée, et en pratique les signaux deviennent inobservables, car perdus dans le bruit.

Etant donné que la différence de déplacement chimique $\Delta\omega$ peut être très différente pour deux spins distincts qui ne subissent pas le même changement d'environnement chimique entre l'état libre et l'état lié — par exemple un spin situé au cœur de la molécule et un spin prenant directement part à l'interaction avec l'autre partenaire — il peut y avoir des spins, au sein de la même molécule, dont les régimes d'échange sont différents.

Le complexe entre la NC du VIH-1 et le bras D de l'ARNt$_3^{Lys}$ s'est trouvé être entre les régimes d'échange rapide et intermédiaire. C'est-à-dire en régime

3. Dans toute la suite du PARAGRAPHE, les notations ω et $\Delta\omega$ référeront respectivement à un déplacement chimique et à une différence de déplacement chimique exprimée en Hz.

d'échange rapide, mais proche de la coalescence, notamment pour les spins qui subissent la plus grande variation de déplacement chimique entre la forme libre et la forme liée. En effet, si l'on considère le modèle d'échange à deux états décrit précédemment, pour une même constante d'échange k_{ex}, les spins qui subissent une faible variation de déplacement chimique $\Delta\omega$ se trouvent en régime d'échange rapide ($\Delta\omega < k_{ex}$), et ceux qui subissent une variation plus importante se rapprochent de la coalescence ($\Delta\omega \sim k_{ex}$). Ainsi, certains résidus sont difficilement observables dans le complexe, et ceci est d'autant plus importun qu'il s'agit des résidus directement impliqués dans l'interaction avec le partenaire.

Dans cet article, nous avons donc utilisé et quelque peu modifié une séquence de type HSQC premièrement décrite par Frans Mulder *et al.* [256]. Cette séquence introduit des trains d'impulsions pendant les périodes de transfert de cohérence INEPT, au lieu de l'unique impulsion centrale π, ce qui permet d'améliorer le rapport signal/bruit en minimisant la perte de cohérence due à l'échange pendant les transferts INEPT.

L'utilisation de cette séquence nous a permis d'améliorer la sensibilité des expériences HSQC réalisées sur le complexe NC/bras D. De plus le comportement différent de certaines résonances nous a permis d'identifier les résidus clés de la NC impliqués dans l'interaction avec le bras D de l'ARNt$_3^{Lys}$. Il s'agit des résidus entourant les aromatiques en position 2 de chaque doigt de zinc (*i.e.* F_{16} et W_{37}), ainsi que des résidus à chaîne latérale hydrophobe (*i.e.* I_{24}, A_{25} et M_{46}) — voir FIGURE 4.4. Ces observations suggèrent un mode de reconnaissance identique du bras D de l'ARNt$_3^{Lys}$ et des signaux d'encapsidation SL2 et SL3. En effet, les structures des complexes NC-SL2 [119] et NC-SL3 [118] ont montré que chaque doigt de zinc de la NC formait une poche hydrophobe, par l'intermédiaire des résidus cités plus haut, et que cette poche reconnaissait alors de façon spécifique une guanine de la boucle d'ARN, par l'intermédiaire des groupements amides et carbonyles de la chaîne principale. Ainsi, les guanines G_{18} et G_{19} de la boucle D de l'ARNt$_3^{Lys}$, qui réalisent des interactions tertiaires avec la boucle TΨC, pourraient être reconnues et séquestrées par la NC, ce qui aboutirait à l'ouverture complète de l'ARNt amorce.

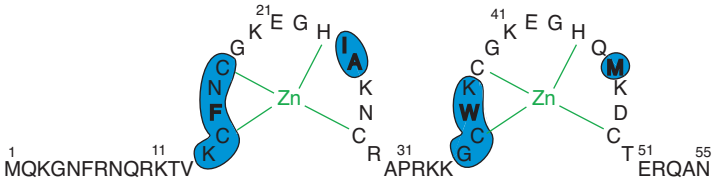


FIGURE 4.4 – Résidus de la NC impliqués dans l'interaction avec le bras D
Les résidus clés de l'interaction sont entourés *en bleu* sur la séquence. Les résidus à chaîne latérale hydrophobe sont en gras.

4.2.2 Article 2 | Barraud *et al.* C.R. Chimie (2008)

Available online at www.sciencedirect.com

ScienceDirect

C. R. Chimie 11 (2008) 474–479

COMPTES RENDUS CHIMIE

http://france.elsevier.com/direct/CRAS2C/

Full paper / Mémoire

Optimizing HSQC experiment for the observation of exchange broadened signals in RNA–protein complexes

Pierre Barraud, Frédéric Dardel, Carine Tisné*

Laboratoire de cristallographie et RMN biologiques, Université Paris Descartes, CNRS UMR 8015, 4, avenue de l'Observatoire, 75006 Paris, France

Received 6 April 2007; accepted after revision 27 June 2007
Available online 30 October 2007

Abstract

Sites for interaction in protein–RNA complexes are often regions of conformational exchange. Although the study of exchange processes could bring valuable information about the recognition mode between protein and RNA, chemical exchange can be detrimental to the NMR spectra quality, resulting in broad, very weak or even unobservable signals. In the present report, we used CPMG-like experiments to improve HSQC spectra of an RNA–protein complex in fast exchange on the chemical shift time scale. The use of such improvement will allow us to handle the resolution of the three-dimensional structure of the complex by NMR. ***To cite this article: P. Barraud et al., C. R. Chimie X 11 (2008).***
© 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Résumé

Les sites d'interaction des complexes ARN–protéine sont souvent des régions d'échange conformationnel. Même si l'étude de ces échanges peut apporter des informations importantes sur le mode de reconnaissance ARN–protéine, l'échange chimique est souvent une cause de mauvaise qualité des spectres RMN en donnant lieu à des signaux larges, de faible intensité et parfois non observables. Dans ce travail, nous avons utilisé des blocs CPMG pour améliorer l'intensité des signaux des spectres HSQC d'un complexe ARN–protéine en équilibre intermédiaire–rapide à l'échelle de temps spectrale. Cette amélioration nous permet maintenant d'envisager de résoudre la structure de ce complexe par RMN. ***Pour citer cet article : P. Barraud et al., C. R. Chimie X 11 (2008).***
© 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Keywords: NMR; Chemical exchange; CPMG; HIV; Reverse transcription; Nucleocapsid

Mots-clés : RMN ; échange chimique ; CPMG ; VIH ; transcription inverse ; nucléocapside

1. Introduction

In the absence of radiofrequency radiation, the contribution of chemical exchange to relaxation is dependent on the rate of the exchange process, the chemical shift differences between involved sites and the population

Abbreviations: CPMG, Carr–Purcell–Meiboom–Gill; HIV, human immunodeficiency virus; HSQC, heteronuclear single quantum coherence; NC, nucleocapsid; PBS, primer binding site; SE, sensitivity enhanced; tRNA, transfer RNA.

* Corresponding author.
E-mail address: carine.tisne@univ-paris5.fr (C. Tisné).

1631-0748/$ - see front matter © 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.crci.2007.06.017

of each site. Near coalescence, the rate of exchange is of the order of the chemical shift difference, and the contribution of chemical exchange to relaxation is therefore maximal. Regions of conformational exchange often correspond to some of the functionally most important sites in proteins, for example, sites for interactions with RNA. Although the study of exchange processes could bring valuable information about the recognition process between protein and RNA, chemical exchange is often detrimental to the NMR spectra quality resulting in broad, very weak or even unobservable signals. Several methods have been developed to characterize chemical exchange processes (for reviews, see Refs. [1–3]) and to improve NMR signal intensities in the presence of such exchange, i.e. to preserve spin coherence [4–6].

In the present report, we used previously described CPMG-like experiment [5,6] to improve HSQC spectrum of an RNA–protein complex. In this kind of experiment, the intensity of the signals involved in chemical exchange process is increased by refocusing spins during coherence transfer period. The HSQC experiment modified with a CPMG pulse train during INEPT period allowed us to improve the overall signal-to-noise ratio and to tear some signals off the noise. These NMR experiments were carried out on a complex between the HIV-1 nucleocapsid protein (NC) and an RNA hairpin that mimicks the D-arm of $tRNA_3^{Lys}$. The HIV-1 nucleocapsid protein (NC) is a short, basic, nucleic acid binding protein with two zinc finger domains. NC is a versatile protein which is involved in multiple functions during the virus replication cycle, many of these functions relying on the nucleic acid chaperone activity of NC (for recent reviews see Refs. [7–9]). In previous studies [10], we have highlighted a strong and specific interaction between the NC protein and the D-arm of $tRNA_3^{Lys}$. This interaction could be the key factor driving the unfolding process of $tRNA_3^{Lys}$, thereby allowing complete annealing of $tRNA_3^{Lys}$ to the primer binding site (PBS), the first event in the HIV-1 reverse transcription. However, the NMR study of the complex between the NC protein and the D-arm of $tRNA_3^{Lys}$ is rendered tricky by chemical exchange process in the microsecond to millisecond time scale, leading to unobservable or broadened signals.

2. Materials and methods

2.1. NMR sample

The recombinant NC_{55} protein was overexpressed from a recombinant plasmid (kind gift of M.F. Summers) and purified as previously described [11,12].

The RNA oligonucleotide mimicking the $tRNA_3^{Lys}$ D-arm (24 nucleotides length) was purchased from Dharmacon Research with 2′-*o*-bis (acetoxyethoxy)-methyl (ACE) protection. The sample was deprotected by following the manufacturer's recommendations and dialyzed several times against deionised water. The pH was adjusted between 6 and 6.5 before freeze-drying.

The ^{15}N-labelled NC_{55} and the unlabelled RNA D-arm were prepared at about 2 mM and 10 mM, respectively, in an NMR buffer (25 mM Na-d_3-acetate pH 6.5, 25 mM NaCl, 0.1 mM BME, 0.1 mM $ZnCl_2$, 10% ^2H_2O). The titration experiment is achieved by adding increased amount of RNA D-arm in a 0.5 mM sample of ^{15}N-labelled NC_{55}. After RNA addition, the sample is immediately vortexed and kept at 30 °C during 30 min before NMR data acquisition.

2.2. NMR experiments

NMR experiments were carried out on a Bruker Avance DRX spectrometer operating at 600 MHz and equipped with a 5-mm triple-resonance gradient probe with *z*-axis gradients. All sequences used sine-bell shaped pulsed field gradients (PFG) along the *z*-axis. Acquisitions of data were performed at 303 K. All experiments were recorded with 512 points in the ^1H dimension with a spectral width of 4.6 ppm and 256 points in the ^{15}N dimension with a spectral width of 30.0 ppm. NMR data were processed with NMRPipe/NMRDraw software [13]. A pure squared sine bell and a pure squared cosine bell were applied along t_2 and t_1, respectively. The data were zero-filled to 1024 points along t_1 and to 4096 along t_2 prior to Fourier transform. Spectra were analyzed with the software SPARKY (T.D. Goddard and D.G. Kneller, SPARKY 3, University of California, San Francisco).

The SE-HSQC experiment is a standard sensitivity-enhanced 2D $^1H/^{15}N$ correlation via double INEPT transfer [14]. Frequency discrimination with retention of absorption mode lineshape is achieved by the echo anti-echo method [15]. Sensitivity enhancement is optimized for amide cross peaks by setting both reverse transfer periods to ($1/2J$). Heteronuclear decoupling during acquisition was achieved by GARP sequence.

The SE-CPMG-HSQC experiment (Fig. 1) is adapted from the SE-HSQC experiment in which the INEPT and reverse-INEPT periods are modified by the inclusion of a phase-modulated pulse train. This experiment is based on previously described HSQC experiments for the observation of exchange broadened signals [6]. The sequence was modified to optimize sensitivity enhancement for amide cross peaks by setting both reverse

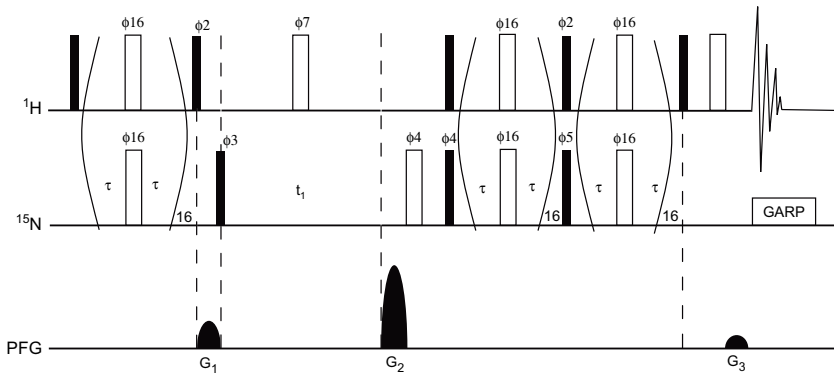


Fig. 1. Pulse scheme of the ^{15}N–^{1}H SE-CPMG-HSQC experiment. Narrow (filled) and wide (open) rectangles represent 90° and 180° pulses, respectively. Unless indicated, the pulses are applied along the *x*-axis. The delay τ was set to 140 μs such that the total length of the spin-echo pulse train equalled 5.55 ms, i.e. $1/2J$ with $J = 90$ Hz for N–H ^{1}J coupling. ^{15}N pulses were achieved using a field strength $\gamma B1 = 47$ kHz. The phase modulation scheme for ϕ_{16} is *x*, *y*, *x*, *y*, *y*, *x*, *y*, *x*, $-x$, $-y$, $-x$, $-y$, $-y$, $-x$, $-y$, $-x$. The following phase cycling was used: $\phi_2 = y$; $\phi_3 = x$, $-x$; $\phi_7 = 2(x)$, $2(-x)$; $\phi_4 = 2(x)$, $2(-x)$; $\phi_5 = 2(y)$, $2(-y)$; $\phi_{rec} = x$, $-x$, $-x$, *x*. Heteronuclear decoupling during acquisition was achieved by GARP. For each t_1 increment, axial peaks were shifted to the sides of the spectrum by inversion of ϕ_3 in concert with the receiver phase, similar as in States-TPPI [23]. The sequence uses sine-bell shaped pulsed field gradients (PFG) along the *z*-axis. To yield absorption mode spectra P- and N-type coherence selection is achieved by inversion of G_2 in concert with inversion of ϕ_5. The duration of all the gradients was 1 ms, followed by a 400 μs recovery delay. Gradient relative strengths were $G_1 = 50\%$; $G_2 = 80\%$; $G_3 = 8.1\%$.

transfer periods to $(1/2J)$. See caption of Fig. 1 for more details on NMR pulse sequence.

3. Results and discussion

3.1. Biological context

The HIV-1 virus uses a cellular tRNA, the human tRNA$_3^{Lys}$ as a primer to initiate the reverse transcription of the viral genome. In the virus, the 3′ end of the primer tRNA is strongly bound by 18 Watson–Crick base pairs to a complementary sequence of the viral RNA called the primer binding site (PBS). The first event in viral replication is thus the annealing of tRNA$_3^{Lys}$ to the PBS, this step being promoted by the viral NC. Most of the known functions of the HIV-1 NC involve interactions with nucleic acids (for recent reviews, see Refs. [7–9]). The structure of the free NC consists in two zinc knuckle domains, connected by a flexible linker attached to flexible N- and C-terminal tails [12,16,17]. By NMR, it was shown that NC has an inherent conformational flexibility that could be crucial to allow its binding to different nucleic acid sequences and structures along the HIV-1 replication cycle [12,18]. In the currently available structures [11,19,20] of NC with nucleic acids, most of the specific contacts between the protein and the nucleic acids occur via the zinc knuckles, notably residues F_{16} and W_{37}, interacting with the guanine-rich loop of RNA hairpin. We have shown by NMR of ^{15}N-labelled tRNA$_3^{Lys}$ that NC interacts preferentially with the acceptor and D stems and with the nucleotides involved in tertiary interactions between the D and T loops [21]. We also performed the reciprocal experiment, i.e. the NMR footprint of tRNA$_3^{Lys}$ binding on ^{15}N-labelled NC (data not shown). The signals are so broadened that roughly only the C-teminal amide group and the amino groups of side chains remain observable. Indeed, the multiplicity of NC binding sites on tRNA$_3^{Lys}$ rendered analyses using the full-length tRNA intractable, even when stringency was raised by increasing salt concentration (data not shown). Reducing the target RNA length could decrease the number of non-specific sites and still have a biological interest. Indeed, studying separate hairpins cannot obviously mimic the entire tRNA, in particular at the level of the TΨC/D loops interaction. However, in the annealing process, there is a step where the TΨC/D loops interaction are broken and where the NC protein plays a crucial role [22]. Therefore, we have previously demonstrated that NC makes a specific interaction with the D hairpin mimicking the D-arm of tRNA$_3^{Lys}$ (dissociation constant $K_d = 28$ nM at 100 mM NaCl). This interaction could be a key factor driving the unfolding process,

thereby allowing complete annealing of $tRNA_3^{Lys}$ to the PBS [10]. Fig. 2a shows the HSQC experiment of ^{15}N-labelled NC with the $tRNA_3^{Lys}$ D hairpin. The NMR signals are quite broad for the size of the complex (14.2 kDa) and may indicate that the NC in the context of the complex is involved in a chemical exchange process in the microsecond to millisecond time scale. This chemical exchange is attributable to the equilibrium between the free and bound forms of the protein.

3.2. Analysis of the exchange process and NMR optimization of HSQC spectra

In the presence of such chemical exchange, multidimensional NMR experiments suffer from dephasing of spin coherence during periods of evolution and coherence transfer by scalar coupling [2,3]. The result of chemical exchange during periods of evolution is a broadening of the resonances, while it leads to loss of magnetization during periods of coherence transfer and therewith of signal intensity. Therefore, resonances are broad, very weak or even unobservable (Fig. 2a). The loss of magnetization during the period of coherence transfer can be minimized by using a modified INEPT block, i.e. the CPMG-INEPT block first described by Mueller and co-workers [5]. The intensity of the signals involved in the chemical exchange process is increased by refocusing them during the coherence transfer period. We then compared CPMG-HSQC [6] and SE-CPMG-HSQC described in Fig. 1, since the gain of intensity from the sensitivity-enhanced HSQC could be counterbalanced by the increase of the pulse sequence length. As a result, the SE-CPMG-HSQC spectrum shows more intense signals than those from the CPMG-HSQC (data not shown).

The comparison of the performance of the SE-HSQC and SE-CPMG-HSQC experiment is given in Fig. 2. The CPMG-derived sequence shows improved sensitivity for almost all resonances, resulting in an average enhancement of about 40%. This result confirms that the exchange process lies in the microsecond to millisecond time scale. Moreover, some unobservable signals in the classic SE-HSQC experiment grow up in the SE-CPMG-HSQC and become observable (see signals 1, 2, 3 in Fig. 2). This could be of significant importance, as these signals, which show considerably higher intensity, belong to residues directly involved in RNA binding.

Fig. 3 shows the titration of ^{15}N-labelled NC by increased amounts of D hairpin until reaching the equivalence. Two kinds of behaviour are observed: (1) a progressive variation of chemical shifts as the amount

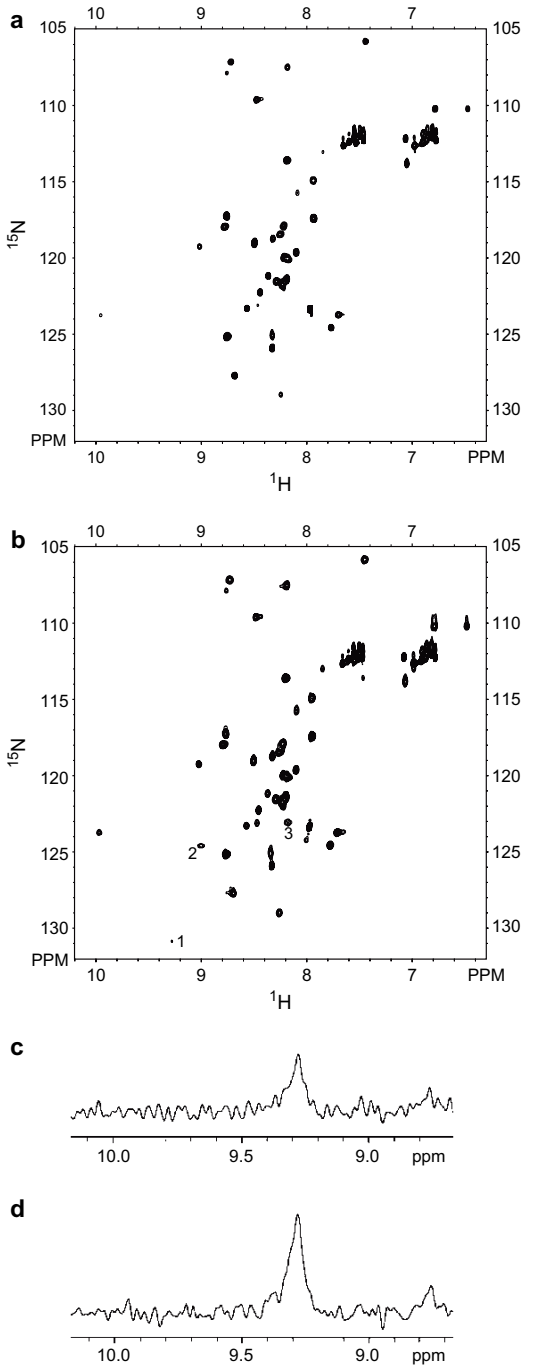


Fig. 2. Contour plot of the 2D HSQC-type spectra of the complex between uniformly ^{15}N-labelled NC and (a) unlabelled $tRNA_3^{Lys}$ D-arm (SE-HSQC experiment), (b) unlabelled $tRNA_3^{Lys}$ D-arm (SE-CPMG-HSQC experiment), (c) 1D SE-HSQC trace of signal 1 and (d) 1D SE-CPMG-HSQC trace of signal 1.

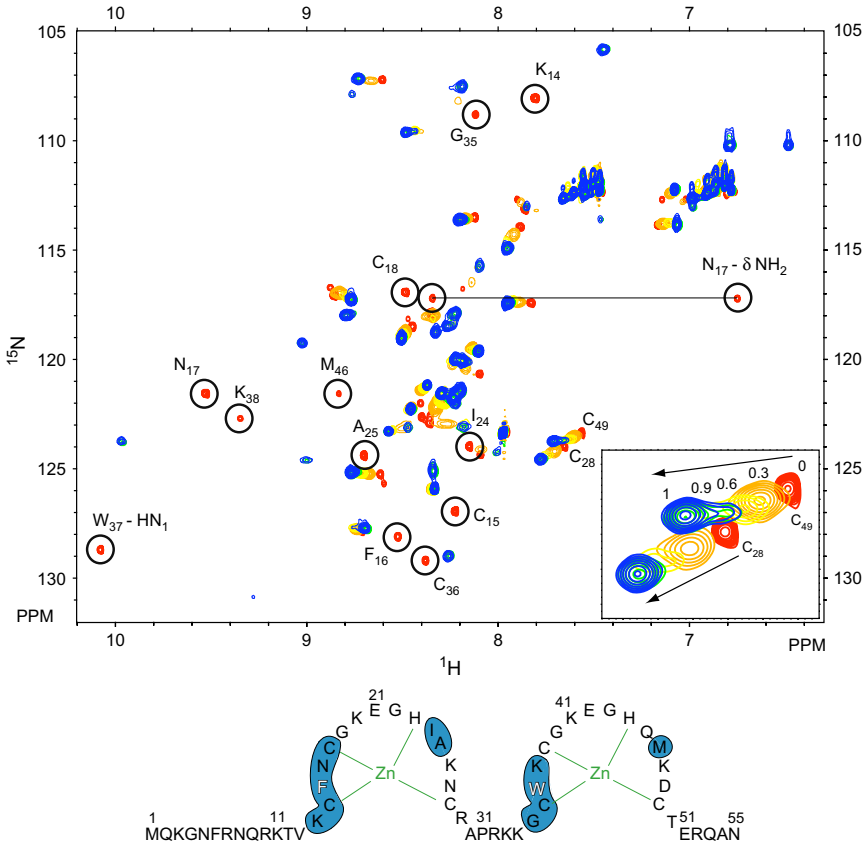


Fig. 3. Titration experiment (SE-CPMG-HSQC) of ^{15}N-labelled NC protein by unlabelled tRNA$_3^{Lys}$ D hairpin. In red: 1/0 protein/RNA ratio; in orange: 1/0.3; in yellow: 1/0.6, in green: 1/0.9, in blue: 1/1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.) The circled signals disappear during the titration. The corresponding residues are highlighted on the NC sequence. The major aromatic residue of each zinc knuckle is written in open letter.

of RNA increases (see for instance the NMR signals of C_{28} and C_{49} – Fig. 3); (2) a disappearance at the substoichiometric level of NC signals and their reappearance close to the 1:1 ratio of NC:RNA (Fig. 3). These results are consistent with an exchange relaxation constant (R_{ex}) behaviour in the fast exchange regime. Indeed, for a simple model of two-site exchange process in the fast exchange limit, the exchange relaxation constant is given by $R_{ex} = p_a p_b(\Delta\omega^2/k_{ex})$, where k_{ex} is the exchange rate constant, $\Delta\omega$ is the chemical shift difference between the two sites, and p_a and p_b the relative population of each site. For a complex with a dissociation constant more than 3 orders of magnitude smaller than the species concentrations, we can assume that the concentration of the complex in bound form is rather equal to the partner added in substoichiometric molarity. Thus, during the titration, for a fraction p of added RNA in the protein ($p \in [0,1]$), the fraction of the bound protein is p and the fraction of the free one is $(1-p)$ and so R_{ex} becomes $R_{ex} = p(1-p)(\Delta\omega^2/k_{ex})$. Even if these equations apply only to the fast exchange limit ($k_{ex} > \Delta\omega$), they provide useful qualitative information for the interpretation of the relaxation behaviour of a complex in a fast exchange regime. For instance, the R_{ex} maximum during the titration occurs for $p = 0.5$, and depends quadratically on the chemical shift difference between the free and the bound forms. Therefore, the NMR signals that experience the largest variation of chemical shifts between the free and bound forms of NC are also those that will experience the largest broadening during the titration due to the exchange regime.

We can reasonably think that these signals which present the largest differences in their chemical shifts

between the free and bound forms of NC are those involved in the interaction with the RNA partner. Actually, they belong to amino acids surrounding the major aromatic residue of each zinc knuckle, i.e. F_{16} and W_{37}, and to residues with hydrophobic side chain, i.e. I_{24}, A_{25} and M_{46} (Fig. 3). These residues have been identified to largely participate in the recognition by the NC of RNA encapsidation signals SL2 and SL3 [11,20]. Indeed, in these structures, each zinc knuckle forms a hydrophobic cleft by means of the hydrophobic side chains listed above and binds to an exposed guanine nucleobase of the RNA loop. Thus the NMR footprint of the RNA D-arm on the NC (Fig. 3) suggests that the recognition mode between the NC and the $tRNA_3^{Lys}$ D-arm is similar to that of the encapsidation signals.

4. Conclusion

In this paper, the use of modified HSQC experiment allowed us to significantly improve the signal to noise ratio in the spectra of the complex between the NC and the $tRNA_3^{Lys}$ D-arm. This experiment also allowed us to tear some signals off the noise and could be of pronounced importance in the way of the determination of the NC/D-arm complex structure by NMR. The determination of this structure is under investigation and should confirm the recognition mode similarity between the encapsidation signals and the $tRNA_3^{Lys}$ D-arm.

Acknowledgments

P. Barraud is supported by a studentship from Ministère de la Recherche. The French AIDS National Agency (ANRS) and 'Ensemble contre le sida (Sidaction) supported this work.

References

[1] M. Akke, Curr. Opin. Struct. Biol. 12 (2002) 642.
[2] A.G. Palmer, C.D. Kroenke, J.P. Loria, Part B, Nucl. Magn. Reson. Biol. Macromol. 339 (2001) 204.
[3] J.G. Kempf, J.P. Loria, Methods Mol. Biol. 278 (2004) 185.
[4] H. Desvaux, P. Berthault, N. Birlirakis, M. Goldman, M. Piotto, J. Magn. Reson. 113 (1995) 47.
[5] L. Mueller, P. Legault, A. Pardi, J. Am. Chem. Soc. 117 (1995) 11043.
[6] F.A.A. Mulder, C.A.E.M. Spronk, M. Slijper, R. Kaptein, R. Boelens, J. Biomol. NMR 8 (1996) 223.
[7] L. Kleiman, R. Halwani, H. Javanbakht, Curr. HIV Res. 2 (2004) 163.
[8] J.G. Levin, J.H. Guo, I. Rouzina, K. Musier-Forsyth, Prog. Nucleic Acid Res. Mol. Biol. 80 (80) (2005) 217.
[9] C. Tisné, Curr. HIV Res. 3 (2005) 147.
[10] P. Barraud, C. Gaudin, F. Dardel, C. Tisné, Biochimie 89 (2007) 1204.
[11] R.N. de Guzman, Z.R. Wu, C.C. Stalling, L. Pappalardo, P.N. Borer, M.F. Summers, Science 279 (1998) 384.
[12] B.M. Lee, R.N. de Guzman, B.G. Turner, N. Tjandra, M.F. Summers, J. Mol. Biol. 279 (1998) 633.
[13] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, J. Biomol. NMR 6 (1995) 277.
[14] A.G. Palmer, J. Cavanagh, P.E. Wright, M. Rance, J. Magn. Reson. 93 (1991) 151.
[15] L.E. Kay, P. Keifer, T. Saarinen, J. Am. Chem. Soc. 114 (1992) 10663.
[16] M.F. Summers, L.E. Henderson, M.R. Chance, J.W. Bess Jr., T.L. South, P.R. Blake, I. Sagi, G. Perez-Alvarado, R.C. Sowder 3rd, D.R. Hare, et al., Protein Sci. 1 (1992) 563.
[17] N. Morellet, N. Jullian, H. de Rocquigny, B. Maigret, J.-L. Darlix, B.P. Roques, EMBO J. 11 (1992) 3058.
[18] S. Ramboarina, N. Srividya, R.A. Atkinson, N. Morellet, B.P. Roques, J.F. Lefevre, Y. Mely, B. Kieffer, J. Mol. Biol. 316 (2002) 611.
[19] N. Morellet, H. Déméné, V. Teilleux, T. Huynh-Dinh, H. de Rocquigny, M.-C. Fournié-Zaluski, B.P. Roques, J. Mol. Biol. 283 (1998) 419.
[20] G.K. Amarasinghe, R.N. de Guzman, R.B. Turner, K.J. Chancellor, Z.R. Wu, M.F. Summers, J. Mol. Biol. 301 (2000) 491.
[21] C. Tisné, B.P. Roques, F. Dardel, J. Mol. Biol. 306 (2001) 443.
[22] C. Tisné, B.P. Roques, F. Dardel, J. Biol. Chem. 279 (2004) 3588.
[23] D. Marion, M. Ikura, R. Tschudin, A. Bax, J. Magn. Reson. 85 (1989) 393.

4.3 Poursuite de l'étude du complexe NC/bras D

4.3.1 Le complexe vu du côté de la protéine

Attribution à partir des expériences de titration

La moitié des résidus de la NC étant en régime d'échange rapide à l'échelle de temps des déplacements chimiques, il est possible de suivre le déplacement des résonances de ces résidus lorsque l'on passe de la forme libre à la forme liée de la protéine.

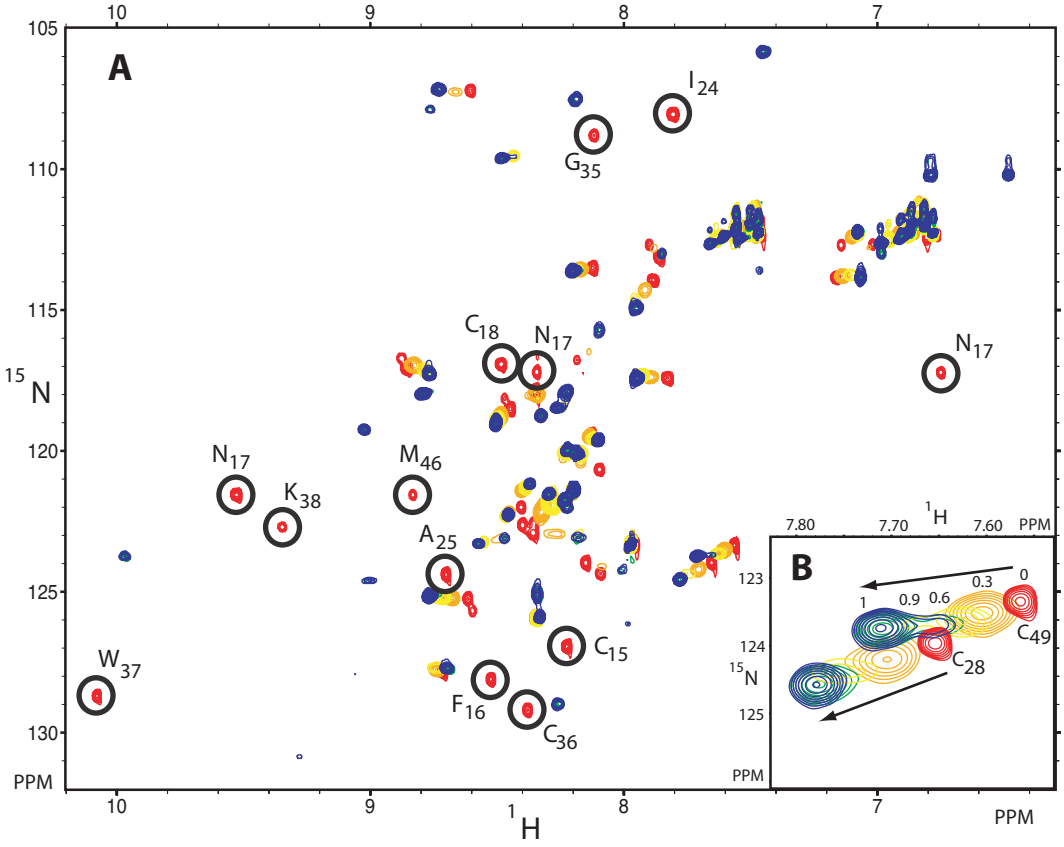


Figure 4.5 – Expérience de titration et attribution de la NC
A : Superposition des expériences CPMG–HSQC ^{1}H–^{15}N de la NC ^{15}N libre (*en rouge*) et en présence de 0.3 éq. (*en orange*), 0.6 éq. (*en jaune*), 0.9 éq. (*en vert*) et 1.0 éq. (*en bleu*) de Bras D24. Les expériences sont réalisées à 303 K. Les résidus des doigts de zinc dont les signaux n'ont pas pu être attribués à partir des ces expériences sont entourés sur le spectre. **B** : Agrandissement de la région des signaux C_{28} et C_{49}.

Les expériences de titration ont donc permis d'attribuer la majorité des résidus des doigts de zinc de la protéine. A ce stade, 25 résidus sur 55 ont pu être attribués. La partie N-terminale (résidus 1 à 13) ainsi que la partie reliant les deux doigts de zinc (résidus 30 à 34) n'ont pas pu être attribuées, car à 30 °C, les signaux correspondant sont très peu intenses voir inobservables sur le spectre de la protéine libre. Le choix de la température est en effet un peu

délicat car on a d'une part tout intérêt à accélérer la vitesse d'échange pour s'éloigner de la coalescence mais d'autre part, en chauffant on perd certains signaux de la protéine — ceux de la partie N-terminale et de l'espaceur central. Nous avons ainsi choisi de réaliser les spectres à 30 °C ce qui affine sensiblement les raies, mais ne permet plus l'observation de tous les signaux. Les expériences de titration sont présentées à la FIGURE 4.5 et le bilan d'attribution à la FIGURE 4.6.

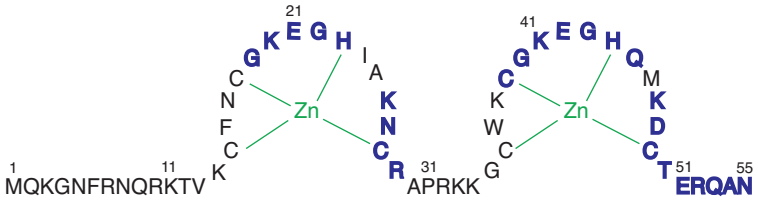


FIGURE 4.6 – Bilan d'attribution grâce aux expériences de titration
Les résidus de la NC attribués par les expériences de titration sont représentés *en bleu* sur la séquence de la protéine.

L'ensemble de l'attribution ne pouvant se déduire simplement des expériences de titration, l'approche classique d'attribution séquentielle de protéine est apparue incontournable pour obtenir une attribution plus complète. L'attribution séquentielle à l'aide des mêmes expériences que celles ayant permis l'attribution de la NC libre a tout d'abord été envisagée — à savoir les expériences à 3 dimensions NOESY–HSQC ^{1}H–^{15}N et TOCSY–HSQC ^{1}H–^{15}N avec la NC simplement marquée ^{15}N. Mais étant donné le nombre de scans à accumuler afin d'obtenir un signal suffisant dans les expériences HSQC, les temps de mélange long de ces séquences, et la faible sensibilité de l'expérience NOESY, ceci n'a pas été réalisable. Les expériences de corrélation scalaire sur le squelette peptidique de la protéine doublement marquée ^{13}C/^{15}N nous ont paru plus adéquates. Cependant, l'accumulation de 128 scans par expériences rend la réalisation de celles-ci impossible avec les séquences standard.

Aussi, nous avons entrepris une collaboration avec Jérôme BOISBOUVIER (Laboratoire de RMN, IBS, Grenoble), afin de profiter de la grande expérience du Laboratoire de RMN de l'IBS pour continuer l'étude délicate de ce complexe et notamment pour bénéficier des séquences optimisées pour l'enregistrement rapide des expériences nécessaires à l'attribution des protéines — la série d'expériences BEST [257, 258], ainsi que de spectromètres 600 MHz et 800 MHz équipés de cryosondes.

Attribution à partir d'expériences triple résonances

Les trois expériences les plus sensibles de l'arsenal des séquences triple résonances — à savoir les séquences HNCO, HNCA et HN(CO)CA [259–261] — ont été réalisées dans leur version BEST [257, 258]. La séquence HNCO corrèle le proton amide du résidu i (H(i)), l'azote de la chaîne principale du résidu i (N(i)) et le carbone du carbonyle du résidu $i-1$ (C′($i-1$)). On observe donc

un pic de corrélation par proton amide, mais aucune information séquentielle ne peut en être extraite. La séquence HNCA corrèle les noyaux H(i), N(i) avec les deux carbones alpha C$_\alpha$(i) et C$_\alpha$($i-1$). Cette expérience permet donc d'obtenir une information séquentielle, le tout étant de pouvoir distinguer le C$_\alpha$ *intra-résidu* du C$_\alpha$ *séquentiel*. Cette information est apportée par l'expérience HN(CO)CA qui corrèle le couple H(i), N(i) uniquement avec le C$_\alpha$($i-1$), à travers le carbonyle C′($i-1$).

Ces expériences ont confirmé l'attribution obtenue directement des expériences de titration et ont permis d'attribuer 7 nouveaux signaux, pour la plupart situés dans le doigt de zinc C-terminal (FIGURE 4.7). Les résidus entourant la F$_{16}$ du doigt de zinc N-terminal n'ont pu être identifiés car certains pics de corrélations dans l'expérience HNCA sont absents ce qui bloque de ce fait le chemin d'attribution. L'attribution des nouveaux résidus est présentée à la FIGURE 4.8 sur les spectres HSQC des expériences de titration de façon à souligner les variations très importantes de déplacements chimiques entre les formes libre et liée de la protéine [4].

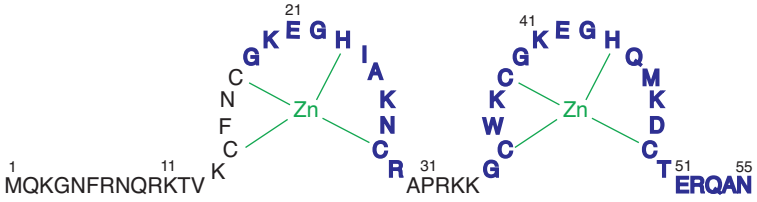


FIGURE 4.7 – Bilan d'attribution grâce aux expériences triple résonances
Les résidus de la NC attribués d'après les expériences HNCA et HN(CO)CA sont représentés *en bleu* sur la séquence de la protéine.

A ce stade de l'étude, il semble difficile de pouvoir réaliser la résolution de la structure du complexe NC/bras D par RMN. Les problèmes d'échange dans la gamme µs–ms conduisent à une perte de signal très importante et rendent quasi inobservables les résidus de la NC directement impliqués dans la reconnaissance du bras D. Nous n'avons en effet pas pu obtenir l'attribution complète de la protéine, et ce même avec des expériences relativement sensibles (HN(CO)CA et HNCA). De plus, les expériences NOESY réalisées sur un échantillon 1 mM du complexe NC/bras D, ne montrent que peu de pics de corrélation N$_H$–N$_H$ et N$_H$–H$_\alpha$. Ainsi, au regard des spectres où l'on observe la protéine, il apparaît difficile de pouvoir obtenir des informations de contraintes structurales intermoléculaires. Cependant, nous avons voulu regar-

4. On peut évaluer la variation de déplacement chimiques $\Delta\omega$ (exprimée en Hz) pour chaque résonance, en tenant compte des variations de déplacement chimique ^{1}H et ^{15}N selon la formule :

$$\Delta\omega = \sqrt{\Delta\omega_H^2 + \Delta\omega_N^2}$$

où $\Delta\omega_H$ et $\Delta\omega_N$ représentent les variations de déplacement chimique (exprimées en Hz) respectivement ^{1}H et ^{15}N. Ces variations sont en moyenne de 400 Hz pour les résidus directement impliqués dans l'interaction avec le bras D. Elles varient d'environ 80 Hz pour le signal du résidu G$_{35}$ à environ 725 Hz pour le signal du résidu W$_{37}$.

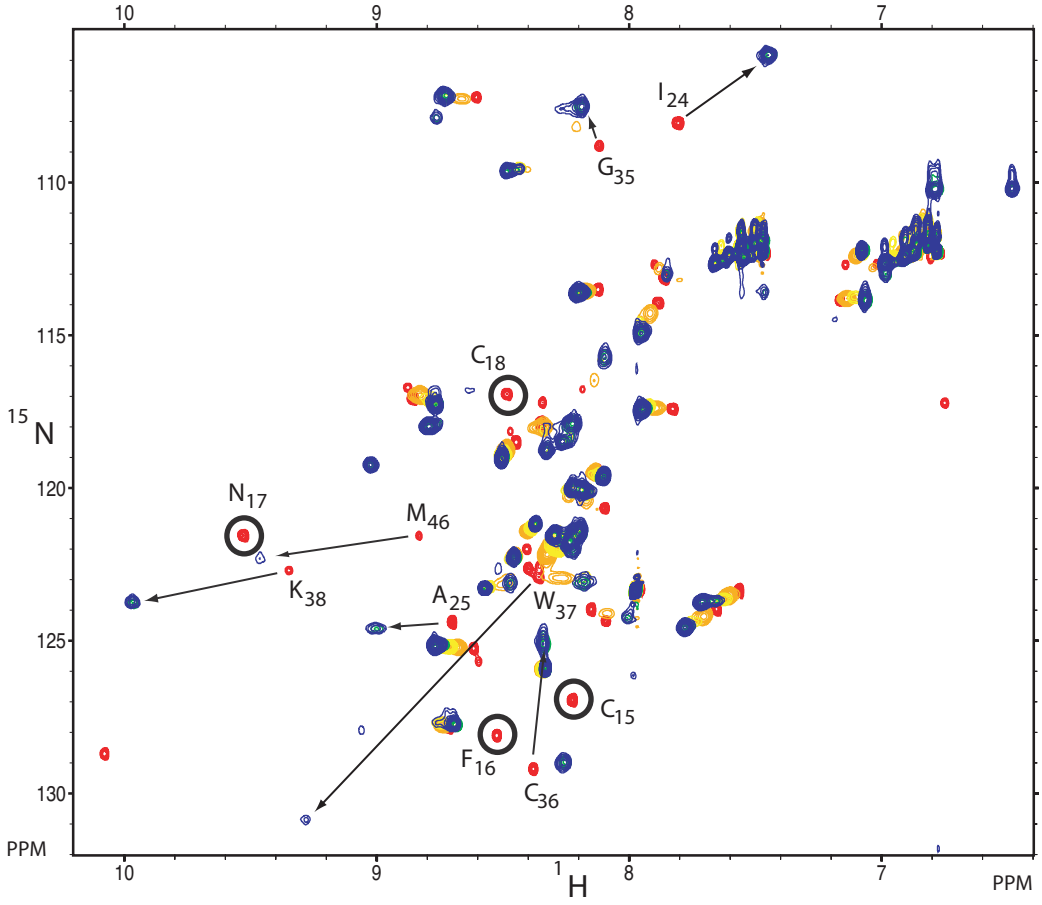


FIGURE 4.8 – Bilan d'attribution grâce aux expériences triples résonances
Les résidus nouvellement attribués grâce aux expériences HNCA et HN(CO)CA, ainsi que les variations de déplacements chimiques correspondant sont indiqués par des flèches. Les résidus du doigt de zinc N-terminal qui n'ont pas pu être identifiés sont entourés sur le spectre.

der le comportement RMN du partenaire ARN dans le complexe, pour voir si le phénomène d'échange était aussi problématique lorsque l'on observait l'ARN. Pour cela, nous avons produit un bras D doublement marqué $^{13}C/^{15}N$ par transcription *in vitro*.

4.3.2 Le complexe vu du côté de l'ARN

Obtention de l'ARN doublement marqué $^{13}C/^{15}N$ par transcription *in vitro*

Une tige boucle mimant le bras D de l'ARNt_3^{Lys} a été produite par transcription par l'ARN polymérase de T7 sur un ADN matrice présentant deux nucléotides modifiés 2′-*O*-méthyle-G en 5′. Ces modifications diminuent l'activité N+1 de la polymérase et donne ainsi des transcrits plus homogène [262]. Pour cette raison, la séquence exacte du bras D24, utilisée jusqu'à maintenant

dans les expériences RMN, n'a pas pu être reproduite telle quelle. Nous avons choisi d'en garder la séquence, mais d'enlever la première paire de base, solution simple qui permet de satisfaire la contrainte des deux C à l'extrémité 3′, complémentaires des deux 2′-*O*-méthyle-G de l'extrémité 5′. L'étude sur l'influence de la longueur du bras D sur la qualité des spectres RMN (FIGURE 4.2) avait de plus montré peu de différences entre une longueur de bras de 7 ou de 8 paires de bases. La comparaison des séquences des Bras D24 et D22 est présentée sur la FIGURE 4.9.

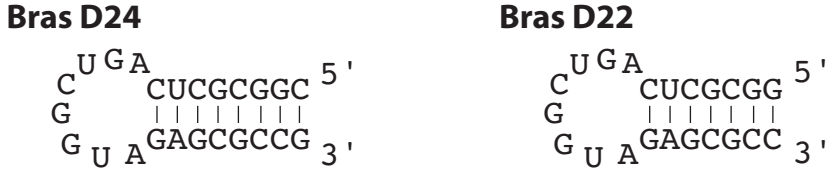


FIGURE 4.9 – Comparaison des séquences Bras D24 et Bras D22 utilisées dans les expériences RMN
La première paire de base de la séquence du Bras D24 a été supprimée dans la séquence du Bras D22 pour satisfaire la contrainte de la méthode de transcription aux 2′-*O*-méthyle-G, à savoir de posséder 2 C en 3′ de la séquence transcrite.

La transcription est tout d'abord réalisée en petit volume (100 µL), pour divers rapports [Mg^{2+}]/[NTPs], afin de contrôler l'efficacité de la transcription et d'évaluer la quantité de Mg^{2+} donnant le maximum de transcrit. L'influence de la concentration en Mg^{2+} sur l'efficacité de la transcription est présentée à la FIGURE 4.10 **A**. La réaction de transcription est ensuite réalisée en grand volume (20 mL), et le transcrit voulu est ensuite purifié par FPLC sur une colonne résolutive échangeuse d'anions (colonne MonoQ). La pureté des diverses fractions d'élution est évaluée sur gel acrylamide-urée (FIGURE 4.10 **B**). L'ARN polymérase de T7 purifiée, le protocole de transcription, et de nombreux conseils nous ont été gentiment donnés par Cyril GAUDIN, alors en stage post-doctoral à l'ICSN (CNRS, Gif-sur-Yvette).

Comportement RMN de l'ARN au sein du complexe

Les expériences RMN sur l'échantillon NC non-marquée/Bras D22 $^{13}C/^{15}N$ ont été réalisées à l'IBS à 600 et 800 MHz. Une expérience HSQC ^{1}H–^{13}C optimisée pour l'observation des protons des sucres ($HSQC_{sucres}$) a été réalisée à 600 MHz, dans une version à temps constant. Sur le même échantillon, une expérience TROSY ^{1}H–^{13}C optimisée pour l'observation des protons aromatiques ($TROSY_{aromatiques}$) a été enregistrée à 800 MHz, également dans une version à temps constant. Les spectres de ces expériences sont présentés sur la FIGURE 4.11.

Le spectre $HSQC_{sucres}$ (FIGURE 4.11 **B**) permet de dénombrer les nucléotides observables au sein du complexe. Le décompte des pics pour chacun des

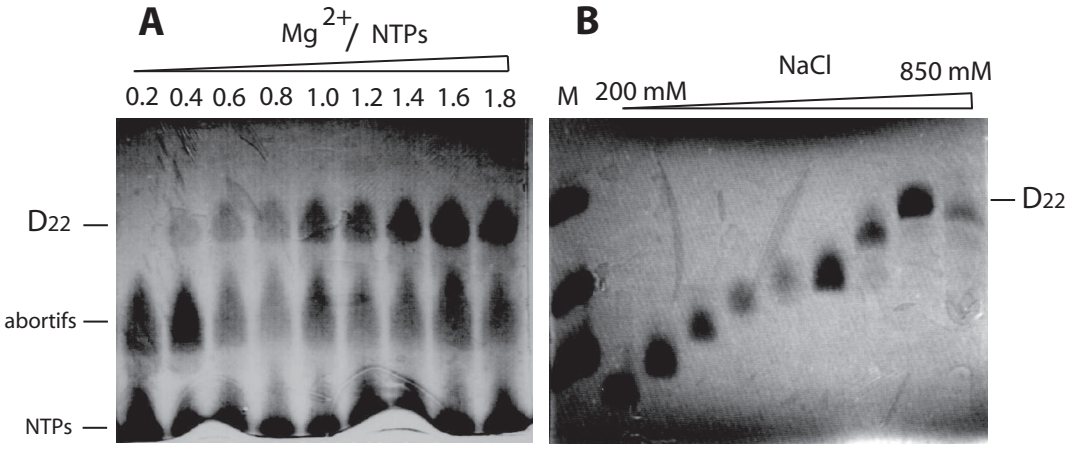


FIGURE 4.10 – Transcription *in vitro* et purification du bras D par FPLC. **A** : Influence de la concentration en Mg^{2+} sur l'efficacité de la transcription. Le rapport $[Mg^{2+}]/[NTPs]$ varie entre 0.2 et 1.8. L'obtention du maximum de transcrit de pleine longueur est évalué pour un rapport $[Mg^{2+}]/[NTPs] = 1.6$. **B** : Purification du bras D par chromatographie échangeuse d'anions (colonne MonoQ). Le bras D pleine longueur est élué pour $[NaCl] \approx 800$ mM.

quatre massifs (H_1', H_4', H_2' & H_3', H_5' & H_5'') donne entre 12 et 14 nucléotides observables, pour un ARN en comprenant 22. Le spectre $TROSY_{aromatiques}$ (FIGURE 4.11 **C**) permet de connaître la nature des nucléotides observables. En effet, dans une expérience à temps constant, les protons H_8 et H_2 des A et des G apparaissent en anti-phase des protons H_5 et H_6 des C et des U, car le carbone qui les porte n'est couplé à aucun ^{13}C, alors que les carbones portant les H_5 et les H_6 des pyrimidines sont couplés entre eux. On peut donc ainsi distinguer les H_6 des pyrimidines des H_8 des purines. De plus, le déplacement chimique particulier du H_2 des A permet de dénombrer les adénines. Ainsi sur ce spectre on dénombre 2 adénines, 5 guanines et 6 pyrimidines.

L'intégralité de la tige-boucle n'est donc pas visible sur les expériences RMN. De plus, compte tenu de la nature des nucléotides observables, à savoir $A_2G_5(C/U)_6$, on peut raisonnablement penser que l'on observe uniquement la tige du Bras D22, sans la paire de base terminale et avec une des adénines de la boucle directement adjacente à la tige. Ainsi, la boucle D ne serait pas observable, car elle serait directement en interaction avec la NC, et verrait pour cela ses signaux très diminués par les mêmes phénomènes d'échange que ceux rencontrés pour la protéine.

Les phénomènes d'échange, de même qu'ils empêchaient de voir les résidus de la NC directement impliqués dans la reconnaissance du bras D, empêchent ici également — lorsque l'on regarde le partenaire ARN — de voir l'interface d'interaction entre la NC et le bras D de l'$ARNt_3^{Lys}$. Il n'est donc pas possible d'obtenir des informations de contraintes structurales intermoléculaires dans ces conditions, et la résolution de la structure du complexe NC/bras D n'a pour cela pas pu être réalisée.

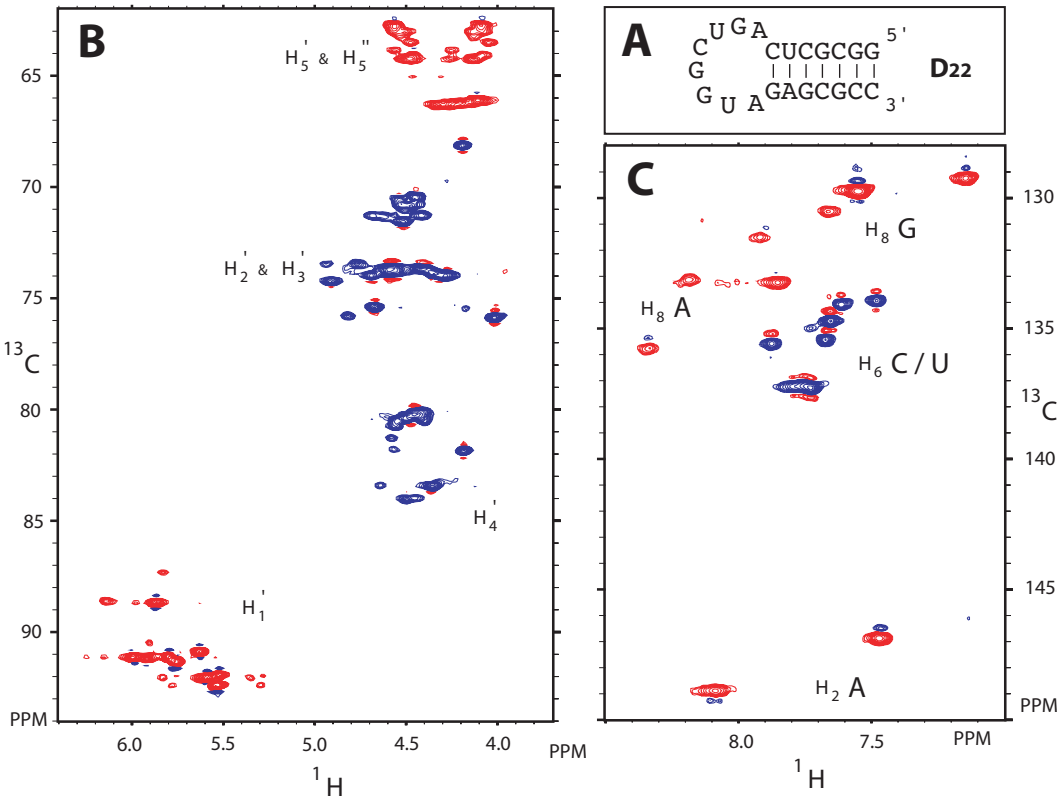


FIGURE 4.11 – Spectres de corrélation ^{1}H–^{13}C du bras D marqué ^{13}C/^{15}N au sein du complexe NC/bras D

A : Séquence du bras D produit par transcription et utilisé dans les expériences RMN. **B** : Spectre HSQC ^{1}H–^{13}C réalisé par une expérience CT-HSQC optimisée pour l'observation des protons des riboses. Les protons H_1', H_5' et H_5'' apparaissent en anti-phase des protons H_2', H_3' et H_4', car le carbone qui les porte n'est couplé qu'à un seul ^{13}C, alors que les autres sont couplés à deux ^{13}C. Dans ce spectre on dénombre entre 12 et 14 résidus. **C** : Spectre TROSY ^{1}H–^{13}C réalisé par une expérience CT-TROSY optimisée pour l'observation des protons des aromatiques. Les protons H_8 et H_2 des A et des G apparaissent en anti-phase des protons H_5 et H_6 des C et des U, car le carbone qui les porte n'est couplé à aucun ^{13}C, alors que les autres sont couplés entre eux.

Chapitre 5

Conclusion et Perspectives

Conclusion

Bien que les phénomènes d'échange nous aient empêchés de résoudre la structure du complexe entre la NC et le bras D de l'ARNt$_3^{Lys}$, nous avons obtenu un certain nombre de résultats qui, dans le contexte de la formation du complexe d'initiation de la transcription inverse du VIH-1, permettent d'affiner le modèle de mécanisme proposé pour l'hybridation entre l'amorce ARNt$_3^{Lys}$ et l'ARN viral. Le travail réalisé au cours de cette thèse a en effet permis de montrer que :

- la NC possède un site de fixation fort et spécifique sur le bras D de l'ARNt$_3^{Lys}$.
- les résidus majeurs de la NC responsables de la reconnaissance du bras D de l'ARNt$_3^{Lys}$ sont les résidus aromatiques situés en position 2 de chaque doigt de zinc de la NC (*i.e.* F_{16} et W_{37}), ainsi que les résidus à chaîne latérale hydrophobe I_{24}, A_{25} et M_{46}.

D'une part, la mise en évidence d'un site de fixation fort et spécifique pour la NC sur le bras D de l'ARNt$_3^{Lys}$ nous a permis de comprendre l'implication de la NC dans l'ouverture de la structure de l'ARNt amorce. Le mécanisme affiné est présenté à la FIGURE 5.1. Le rôle clé de la NC dans ce mécanisme est de permettre la fusion des interactions tertiaires entre les boucles T et D de l'ARNt$_3^{Lys}$. Ce résultat s'intègre parfaitement dans notre compréhension de la formation du complexe d'initiation de la transcription inverse. En effet, pour réaliser la fusion des interactions tertiaires entre les boucles T et D, la NC peut *a priori* se fixer soit sur la boucle T soit sur la boucle D de l'ARNt$_3^{Lys}$ — soit sur les deux boucles à la fois. Mais dans le contexte de l'hybridation entre le PBS et l'ARNt amorce, il est finalement plus logique que la NC vienne se fixer sur la boucle D qui n'est pas affectée par le processus d'hybridation — plutôt que sur la boucle T, où elle viendrait gêner l'appariement avec le PBS.

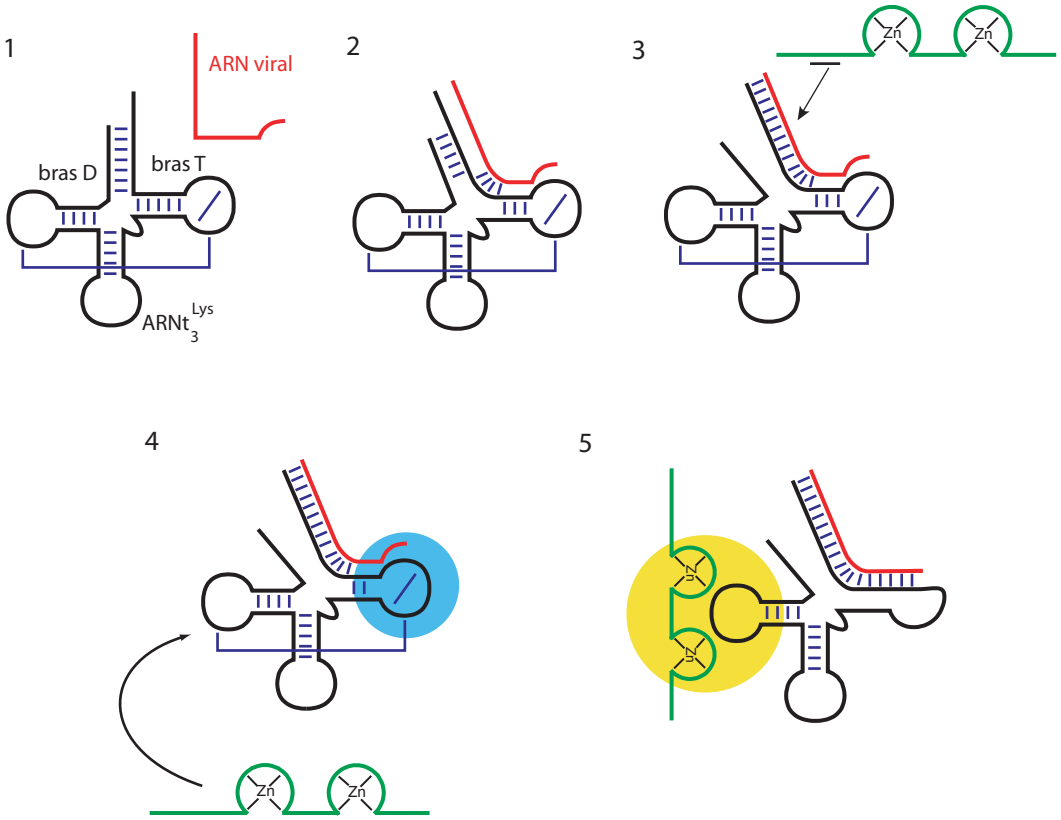


FIGURE 5.1 – Rôle de la NC dans le mécanisme d'hybridation entre l'ARNt$_3^{Lys}$ et le PBS
Cette figure fait référence à la FIGURE 2.11 que l'on peut trouver à la page 46. **1** : L'ARNt$_3^{Lys}$ est *en noir*, le PBS *en rouge* et la NC *en vert*. **2** : L'invasion de la structure de l'ARNt amorce par l'ARN viral débute au pied de la tige acceptrice et peut avoir lieu sans l'intervention de la NC. **3** : L'échange de brin se propage ensuite le long de la tige acceptrice puis aux premières paires de bases du bras T. La NC accélère ce réappariement, probablement par le biais d'une interaction électrostatique non spécifique. **4** : A ce stade, le cœur de la structure tertiaire de l'ARNt persiste et constitue un verrou structural stable. La NC parvient à faire sauter ce verrou structural, en reconnaissant de façon spécifique le bras D de l'ARNt$_3^{Lys}$ grâce à ses doigts de zinc. Cette reconnaissance implique vraisemblablement les guanines G_{18} et G_{19} de la boucle D. **5** : Hybridation complète. L'interaction forte NC/bras D déplace l'équilibre vers cet état.

D'autre part, la mise en évidence du fait que les résidus de la NC impliqués dans la fixation du bras D de l'ARNt$_3^{Lys}$ soient les mêmes que ceux impliqués dans la reconnaissance des signaux d'encapsidation SL2 et SL3 (FIGURES 2.7 et 2.8 pages 38 et 39), suggère très fortement que le mode de reconnaissance de la NC pour le bras D soit très comparable à ceux des signaux SL2 et SL3. C'est-à-dire que chaque doigt de zinc doit vraisemblablement venir séquestrer une guanine de la boucle D, et que la partie N-terminale de la NC doit venir se structurer en hélice 3_{10} dans un des sillons de la tige.

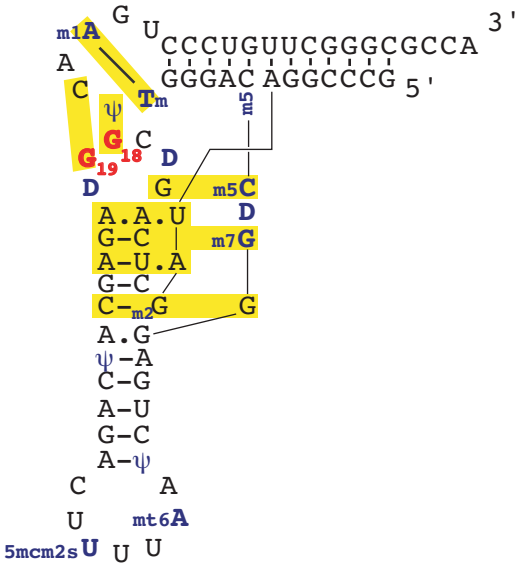


FIGURE 5.2 – Les guanines de la boucle D potentiellement reconnues par la nucléocapside
La séquence de l'ARNt$_3^{Lys}$ est représentée selon le repliement en forme de « L » des ARNt. Les interactions tertiaires par liaisons hydrogènes sont représentées *en jaune*, les nucléotides modifiés *en bleu*, et les deux guanines candidates de la boucle D pour la reconnaissance par la NC (G_{18} et G_{19}) sont *en rouge*.

Les guanines de la boucle D potentiellement reconnues par la NC correspondent aux guanines G_{18} et G_{19}. En effet, celles-ci participent aux interactions tertiaires principales avec la boucle T (FIGURE 5.2).

Perspectives

S'il l'on voulait poursuivre l'étude structurale du complexe entre la NC et le bras D plusieurs pistes pourraient être envisagées. Il peut arriver que de simples variations dans les séquences de la protéine ou dans celle de l'ARN change de façon significative le régime d'échange d'un complexe ARN/protéine. Cependant, la séquence de la boucle D de l'ARNt$_3^{Lys}$ est unique et la pertinence biologique du sujet nécessite de garder celle-ci inchangée. Nous avons déjà cherché à optimiser la longueur de la tige. On pourrait par exemple utiliser une séquence de la boucle qui possède les modifications post-transcriptionnelles, c'est-à-dire dans le cas de la boucle D de l'ARNt$_3^{Lys}$, les deux dihydrouridines D_{16} et D_{20}. Nous avons en effet utilisé dans cette étude des ARN produits par synthèse chimique ou par transcription *in vitro* qui ne possèdent pas ces modifications. Il se peut que celles-ci influent suffisamment sur l'interaction avec la NC pour que les propriétés relatives à l'échange soit complètement changées. D'autre part, on peut envisager de changer la séquence de la protéine — et dans cette optique se procurer des nucléocapsides provenant de différents isolats du VIH-1 — et d'utiliser une protéine tronquée dans sa partie N-terminale, qui ne garderait dans ce cas que les deux doigts de zinc. En effet, cette extension

N-terminale agit principalement au travers d'interactions non spécifiques et n'est peut-être pas strictement nécessaire à la reconnaissance du bras D.

D'autre part, la compréhension que l'on a actuellement du mécanisme de formation de l'hybride entre l'amorce $ARNt_3^{Lys}$ et l'ARN viral ne comprend que le domaine PBS de l'ARN viral. Mais, comme nous l'avons mentionné au Paragraphe 2.3.2 page 47, la formation du complexe d'initiation implique d'autres régions de l'ARN viral (boucle riche en A et séquence PAS). Certaines études sur les interactions additionnelles entre l'$ARNt_3^{Lys}$ et l'ARN viral sont en contradictions et il serait intéressant d'étudier le mécanisme d'hybridation dans un contexte plus large au niveau de l'ARN viral — c'est à dire en présence non seulement de la séquence PBS, mais aussi par exemple de la séquence PAS — afin de voir quels éléments pourrait apporter la RMN en ce qui concerne ces interactions additionnelles.

Deuxième partie

Etude d'une méthyltransférase des ARN de transfert – la m^1A_{58} ARNt méthyltransférase

Chapitre 6

Contexte biologique

L'addition d'un groupement méthyle à des molécules biologiquement actives, comme des hormones, des neurotransmetteurs, des lipides, des protéines, ou des acides nucléiques provoque un changement des propriétés physico-chimiques de ces molécules. Les fonctions biologiques de la méthylation sont très larges, allant du métabolisme à la maturation des acides nucléiques en passant par la transduction du signal ou la détoxification. La diversité des fonctions dans lesquelles intervient la méthylation trouve pour reflet la variété étonnante des méthyltransférases (MTases), les enzymes qui catalysent les réactions de méthylation.

6.1 Les méthyltransférases

Bien que différentes familles de MTases soient connues, la grande majorité des réactions de méthylation sont catalysées par les MTases S-adénosyl-L-méthionine dépendantes (pour des revues [263–265]). Ces enzymes utilisent la S-adénosyl-L-méthionine comme donneur de méthyle.

6.1.1 La S-adénosyl-L-méthionine

La S-adénosyl-L-méthionine — aussi connu sous les noms de SAM et d'AdoMet — est une petite molécule présentant une fonction sulfonium (FIGURE 6.1). Elle est impliquée dans un grand nombre de réactions biologiques. Il s'agit en effet du cofacteur le plus utilisé par les enzymes, juste après l'ATP [266]. La SAM résulte de la condensation du nucléotide adénosine et de l'acide aminé méthionine, deux composés biologiques aux fonctions diverses qui étaient probablement présents sur Terre dans l'environnement prébiotique [267]. La SAM est bio-synthétisée par la SAM synthétase — ou méthionine adénosyltransférase — par couplage de l'acide aminé méthionine avec une molécule d'ATP (FIGURE 6.2 ①) [268]. Cette réaction est stéréospécifique. Elle n'engendre que la configuration absolue S au niveau de l'atome de soufre asymétrique.

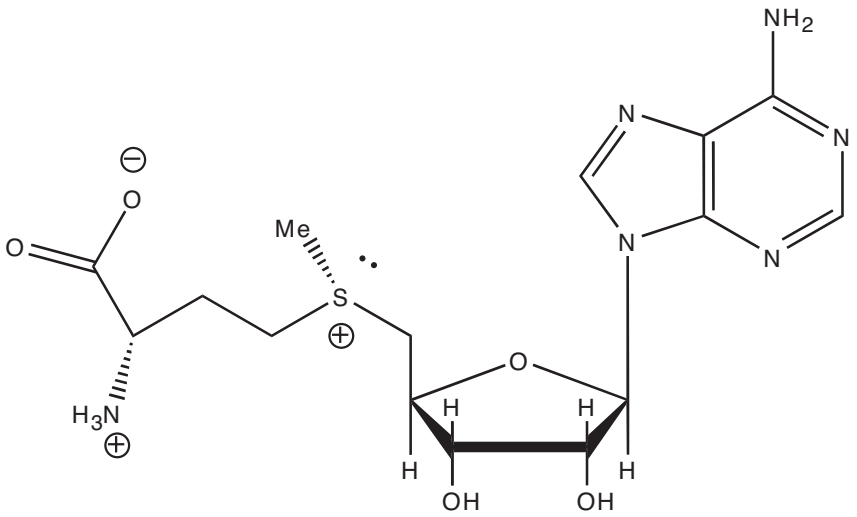


Figure 6.1 – Structure chimique de la S-adénosyl-L-méthionine
La fonction chimique centrale de la SAM est la fonction sulfonium qui rend le méthyle issu de la méthionine très électrophile. L'atome de soufre est chargé positivement. C'est un centre asymétrique et seule la molécule de configuration *S* est formée lors de la bio-synthèse de la SAM par la SAM synthétase [268].

La SAM, un donneur de méthyle

La SAM est reconnue comme étant le principal cofacteur donneur de méthyle des réactions de méthylation, et ce chez tous les êtres vivants [266, 268, 269]. Les réactions de méthylation de divers nucléophiles exploitent une propriété intrinsèque de la SAM, à savoir le très fort caractère électrophile de son groupement méthyle — conséquence de la fonction sulfonium directement adjacente. Ce fort caractère électrophile est mis à profit par les MTases qui mettent en contact la SAM avec les groupements nucléophiles de leurs substrats. Durant la réaction de transfert de méthyle, la SAM est convertie en S-adénosyl-L-homocystéine (SAH) (Figure 6.2 ②). La très forte prédominance de la SAM sur d'autres donneurs de méthyle, comme le 5-méthyl-tétrahydrofolate, reflète la thermodynamique extrêmement favorable de la réaction de transfert de méthyle faisant intervenir la SAM. L'enthalpie libre standard de réaction ($\Delta_r G°$) pour la formation de méthionine à partir d'homocystéine (SAM + Hcy $\longrightarrow$ SAH + Met) est par exemple évaluée à -71 kJ.mol^{-1} soit plus du double de celle de la réaction d'hydrolyse de l'ATP (ATP + $H_2O \longrightarrow$ ADP + P_i) (-30 kJ.mol^{-1}) [265, 266]. La méthionine est ensuite régénérée en deux étapes (Figure 6.2 ③ & ④). La SAH est tout d'abord hydrolysée en adénosine et homocystéine par la SAH hydrolase [270]. L'homocystéine peut être ensuite soit convertie en glutathion — un antioxydant cellulaire majeur — soit méthylée pour régénérer la méthionine. Cette réaction est catalysée par la méthionine synthétase. Elle utilise le 5-méthyl-tétrahydrofolate (N^5-mTHF) comme substrat donneur de méthyle, et la cobalamine (vitamine B_{12}) comme cofateur, selon un mécanisme en deux étapes dans lequel la cobalamine (B_{12}) est tout

d'abord méthylée en méthylcobalamine (MeB_{12}) par le substrat N^5-mTHF. MeB_{12} transfert alors son groupement méthyle à l'homocystéine pour former le produit méthionine et régénérer le cofacteur cobalamine [271–273]. Cette séquence de réactions (①, ②, ③ & ④) constitue un cycle faisant partie du métabolisme des dérivés des acides aminés soufrés, lui même régulé au niveau génétique par la SAM [274].

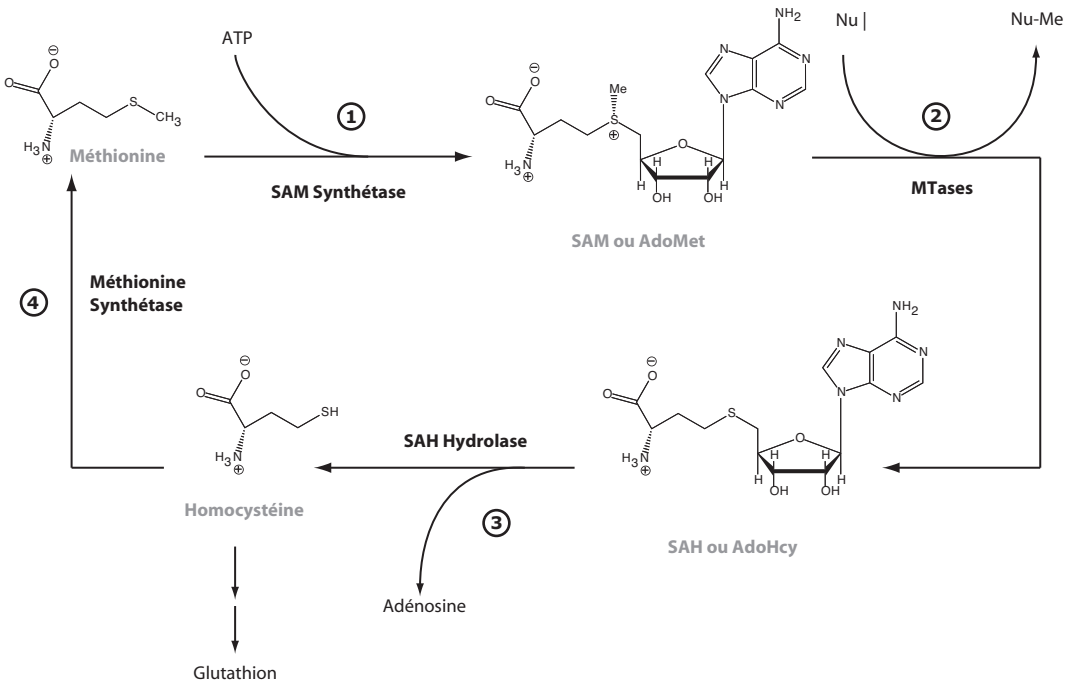


Figure 6.2 – Le cycle de la S-adénosyl-L-méthionine
Nu représente un nucléophile substrat d'une MTase. ① La SAM est formée par couplage de la méthionine avec l'ATP par la SAM synthétase. ② Les MTases catalysent le transfert du groupement méthyle de la SAM aux nucléophiles. La méthionine est régénérée, par hydrolyse de la SAH ③, et par méthylation de l'homocystéine, catalysée par la méthionine synthétase ④.

Les substrats nucléophiles méthylés par la SAM (Figure 6.2 ②) sont extrêmement divers. Il peut en effet s'agir de composés simples, tels que des ions, comme l'oxyde d'arsenic [275] ou les halogénures (chlorure, bromure et iodure) [276–279], ou de petites molécules, comme des acides aminés [280], des dérivés de la caféine [281–287], etc. Parmi ces petites molécules, les neurotransmetteurs de la famille des catécholamines (dopamine, épinéphrine et norépinéphrine) ont été très étudiés. Ils sont inactivés du côté post-synaptique, par l'ajout entre autre d'un groupement méthyle par la cathécol-*O*-méthyltransférase (COMT) [288–292]. La COMT est la cible de nombreux traitements thérapeutiques visant à réguler la disponibilité des catécholamines [293–296]. Outre ces petits composés, des molécules complexes comme les protéines et les acides nucléiques sont aussi des substrats des MTases SAM dépendantes. Environ 120 membres de la famille des MTases SAM dépendantes ont été classifiées (EC 2.1.1.x)

selon leur spécificité de substrat (petite molécule, lipide, protéine, acide nucléique, etc.) et selon l'atome cible de la méthylation (azote, oxygène, carbone, soufre, etc.).

La SAM, un cofacteur versatile

Cependant, même si l'implication majeure de la SAM dans le métabolisme implique bien les réactions de méthylation, la seule intervention de la SAM comme cofacteur donneur de méthyle ne saurait décrire l'incroyable diversité des réactions auxquelles elle prend part [297]. Cette versatilité surprenante s'explique principalement par l'électrophilie importante de tous les carbones adjacents au sulfonium. La SAM est la source de groupements méthylène pour la synthèse d'acides gras contenant une fonction cyclopropane (CFA pour *Cyclopropane Fatty Acids*) [298, 299]. Elle est aussi la source de groupements amino sur la voie de biosynthèse de la biotine [300, 301], de groupements ribosyle pour la synthèse de la queuosine, un nucléoside hypermodifié rencontré dans les ARNt [302–304], et de groupements aminoalkyle entre autres pour la synthèse des polyamines comme la spermidine [305, 306] et pour la formation de 3-(3-amino-3-carboxypropyl)uridine (acp^3U), un nucléoside modifié rencontré dans la boucle variable de certains ARNt de *Bacteria* [307] et dans les ARNr des *Archaea* [308].

Outre toutes ces réactions, dans laquelle la SAM exhibe une réactivité ionique extrêmement variée, elle intervient aussi dans des réactions radicalaires, comme donneur de radicaux 5′-désoxyadénosyle [309–311]. Des enzymes possédant un centre Fer-Soufre [4Fe–4S] réalisent la coupure homolytique de la liaison S–C(5′-désoxyadénosyle) de la SAM après la réduction à 1 électron de celle-ci, ce qui conduit à la formation de méthionine et du radical 5′-désoxyadénosyle qui peut alors être utilisé comme initiateur de réactions radicalaires.

Même si la SAM intervient dans des réactions biochimiques variées, son implication majeure reste cependant le transfert de groupement méthyle. Les MTases SAM dépendantes interviennent, comme nous l'avons vu, dans la méthylation de nombreux substrats et agissent sur des atomes cibles variés. La première structure de MTase SAM dépendante — la m^5C ADN MTase M•HhaI [312] — a été déterminée en 1993. D'autres structures ont par la suite révélé une convergence fonctionnelle pour les MTases SAM dépendantes, qui en effet présentent plusieurs repliements distincts. Cinq classes de MTases SAM dépendantes ont été définies d'après leur structure.

6.1.2 Classification des méthyltransférases

Même si actuellement 5 classes de MTases SAM dépendantes sont connues, elles ne sont pas toutes également représentées dans la PDB (*Protein Data Bank*). La Classe I est la plus représentée, avec environ 90 % des MTases SAM dépendantes [265]. L'enzyme ayant fait l'objet de mon travail de thèse fait elle aussi partie de cette classe de MTase. Pour ces raisons, nous allons présenter la Classe I en détail et nous décrirons les autres classes plus succinctement.

MTases de Classe I

Repliement canonique Le repliement des MTases de Classe I s'organise autour d'un cœur très conservé auquel vient éventuellement s'ajouter quelques éléments supplémentaires. Le cœur des MTases de Classe I (FIGURE 6.3) consiste en une alternance de brins β (β_1–β_7) et d'hélices α (α_Z et α_A–α_E). Les brins β forment un large feuillet β central à 7 brins, sur lequel viennent s'agencer les hélices α — trois de chaque côté. Les brins β sont presque tous parallèles, excepté le brin β_7 qui vient s'insérer entre les brins β_6 et β_5 de façon antiparallèle. La position relative des brins β est la suivante : $6\uparrow 7\downarrow 5\uparrow 4\uparrow 1\uparrow 2\uparrow 3\uparrow$ (FIGURE 6.3 **B**). Le premier brin β se termine généralement par le motif GxGxG — la signature du site de fixation d'un nucléotide — et vient étroitement épouser la SAM avant le début de l'hélice α_A. A part le brin β_7, ce repliement est très similaire aux domaines de liaison aux dinucléotides NAD et NADP, qui possèdent le repliement bien connu dit de Rossmann (*Rossmann fold*) [313]. Dans les deux cas, le point d'inversion topologique entre les brins β_1 et β_4 — β_1 à β_3 dans un sens, et β_4 à β_6 dans l'autre sens — forme un sillon profond dans lequel les cofacteurs SAM ou NAD(P) viennent se lier.

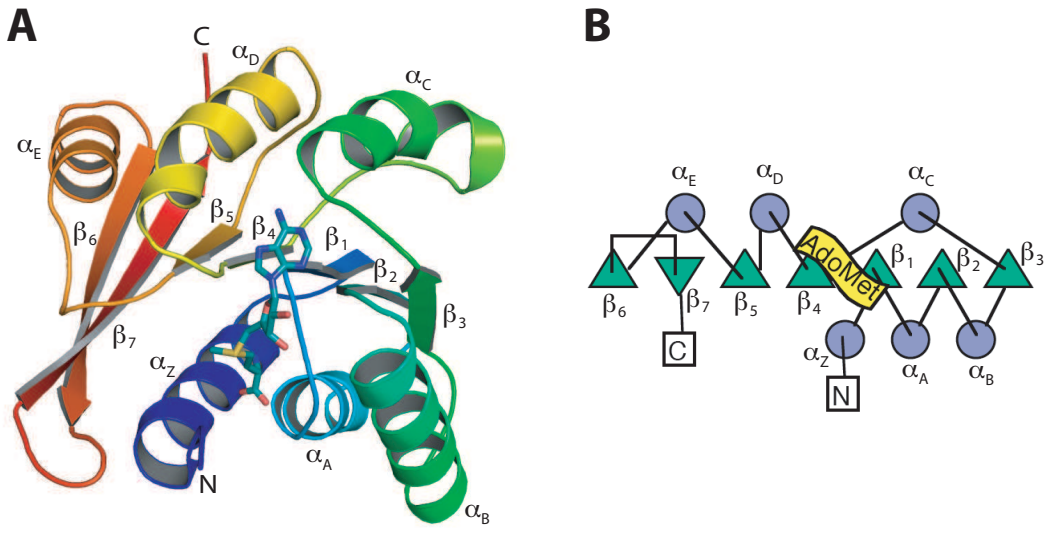


FIGURE 6.3 – Organisation tridimensionnelle des MTases de Classe I
A : Repliement caractéristique des MTases de Classe I. L'extrémité N-terminale est *en bleu foncé* et l'extrémité C-terminale *en rouge*. La SAM est représentée en bâtonnets. La structure correspond à la Catéchol-*O*-MTase de Rat (PDB code 1VID) [292]. **B** : Diagramme topologique des MTases de Classe I. Les hélices α sont *en bleu*, les brins β *en vert*, et le site de fixation de la SAM (AdoMet) *en jaune*. Diagramme d'après Schubert *et al.* [265].

Variations par rapport au repliement canonique La partie C-terminale des MTases (hélices α_D–α_E & brins β_4–β_7) est responsable de la liaison au substrat cible de la méthylation. Cette partie est donc beaucoup plus sujette à

variation, vis-à-vis du repliement canonique, que la partie N-terminale, qui lie le cofacteur commun SAM. Par exemple, les brins β_6 et β_7 sont inversés dans la séquence primaire des MTases réparatrices des isoaspartates (PIMT pour *Protein Isoaspartyl MTases*), résidus rencontrés dans les protéines endommagées [314]. Ces mêmes brins sont absents des MTases de protéines spécifiques des résidus arginines (PRMT pour *Protein R* (arginine) *MTases*) [315]. Certaines MTases disposent, en plus du domaine central, de domaines auxiliaires, en position N ou C-terminale, ou même insérés entre deux éléments prenant part au repliement canonique. Ces domaines sont interprétés comme des domaines de fixation au substrat. Ils permettent ainsi une déclinaison plus importante des substrats reconnus et méthylés par les MTases SAM dépendantes. Les petites variations rencontrées dans les structures de MTases, par rapport à l'archétype de repliement des MTases de Classe I sont très bien détaillées dans Martin *et al.* [264].

Le site de fixation de la SAM Un point important à relever en ce qui concerne le site de fixation de la SAM, est que même si la localisation de celui-ci au sein du repliement canonique est très conservé, les résidus directement impliqués dans la fixation, eux le sont moins. La SAM adopte une conformation allongée et vient se loger au point d'inversion topologique entre les brins β_1 et β_4. Elle est essentiellement liée par les 4 boucles situées entre les éléments β_1 et α_A, β_2 et α_B, β_3 et α_C et β_4 et α_D. Plus précisément, la partie aminoacide est liée par la boucle reliant le brin β_1 à l'hélice α_A — essentiellement par des interactions faisant intervenir la chaîne principale d'un motif conservé GxGxG — ainsi que par la boucle reliant le brin β_4 à l'hélice α_D. La partie ribose est quant-à-elle liée à la fois par la boucle reliant le brin β_1 à l'hélice α_A et par celle reliant le brin β_2 à l'hélice α_B. Le cycle purine de l'adénine est lié par un petit nombre de contacts qu'il effectue avec les boucles reliant le brin β_2 à l'hélice α_B et le brin β_3 à l'hélice α_C. La localisation de la SAM par rapport aux éléments de structure secondaire est ainsi très conservée, mais celle-ci ne s'accompagne pas d'une conservation des fonctions chimiques des chaînes latérales des résidus impliqués dans la fixation du cofacteur. Les résidus hydrophobes en interaction avec les parties aromatique et alkyle de la SAM sont en effet peu conservés. Les résidus F, I, V, W, P et M sont souvent retrouvés, et leur utilisation diffère selon les sous-familles d'enzymes [264]. Les seuls résidus extrêmement conservés pour la fixation de la SAM, dans la famille des MTases SAM dépendantes de Classe I, sont ceux du motif déjà évoqué GxGxG, et un résidu E/D à la fin du brin β_2, venant reconnaître les hydroxyles du ribose de la SAM [316].

MTases des Classes II à V

En 1996, soit trois ans après la première structure de MTase de Classe I, une nouvelle structure de MTase SAM dépendante montre un repliement différent de celui adopté par les MTases alors connues [317]. Cette enzyme — la méthionine synthétase cobalamine dépendante — va définir l'archétype des MTases de Classe II. Par la suite, en 1998 [318] puis en 2002 [319–325], trois nouvelles

familles de MTases ont été découvertes. Elles constituent les MTases de Classes III, IV et V. L'organisation tridimensionnelle des MTases des Classes II à V est présentée à la FIGURE 6.4 au travers de la structure d'une MTase de la famille et du diagramme topologique correspondant.

MTase de Classe II Cette famille ne comprend en fait à ce jour qu'une seule enzyme — le domaine de réactivation de la méthionine synthétase cobalamine dépendante (MetH). Cette enzyme produit de la méthionine en transférant le groupement méthyle du 5-méthyl-tétrahydrofolate à une molécule d'homocystéine, et utilise le cofacteur cobalamine (B_{12}) comme relais (FIGURE 6.2 ④ et PARAGRAPHE 6.1.1). Périodiquement, le cobalt du cofacteur B_{12} est oxydé et devient alors inactif. Sa réactivation nécessite une méthylation réductive à partir de SAM et de flavodoxine. Celle-ci est effectuée par le domaine de réactivation de MetH.

MTases de Classe III Cette famille comprend les enzymes réalisant les différentes méthylation du cycle tétrapyrrolique sur la voie de biosynthèse de la vitamine B_{12} [318, 326–329]. On parle de précorrine MTases. Il est intéressant de noter que toutes les précorrines MTases n'appartiennent pas à la Classe III des MTases SAM dépendantes. Certaines adoptent en effet le repliement caractéristique de la Classe I [330].

MTases de Classe IV Cette famille, aussi appelée SPOUT, comprend essentiellement des ARNr et ARNt MTases [331, 332]. Le repliement de cette famille n'est pas sans rappeler celui de la Classe I (FIGURE 6.3 et FIGURE 6.4 **C**). Elle comprend en effet un feuillet β central à 6 brins parallèles flanqué d'hélices α de part et d'autre. La principale différence a lieu dans la partie C-terminale, dans laquelle les protéines de la famille SPOUT se replie en formant un nœud topologique, au niveau de l'hélice représentée en magenta sur la FIGURE 6.4 **C**. Malgré cette ressemblance structurale, il n'y a pas de similarité de séquence entre les MTases des classes I & IV [333]. Cependant, même si ceci est loin d'être évident, ces deux classes pourraient provenir d'une MTase ancestrale commune [333].

MTases de Classe V Cette famille, également appelée MTases à domaine-SET, comprend des MTases transférant un groupement méthyle sur des résidus lysine de protéines nucléaires impliquées dans le fonctionnement de la chromatine et dans la régulation de la transcription [334], comme les histones [335–337], TAF10 [338], le suppresseur de tumeur p53 [339], mais aussi sur d'autres protéines comme la Rubisco [321] ou le cytochrome c [340, 341]. Les MTases de cette famille à domaine SET se replient comme les MTases SPOUT, en formant un nœud topologique, au niveau de l'hélice représentée en magenta sur la FIGURE 6.4 **D**. Il est intéressant de noter que dans les deux cas, la région du nœud intervient dans la fixation de la SAM.

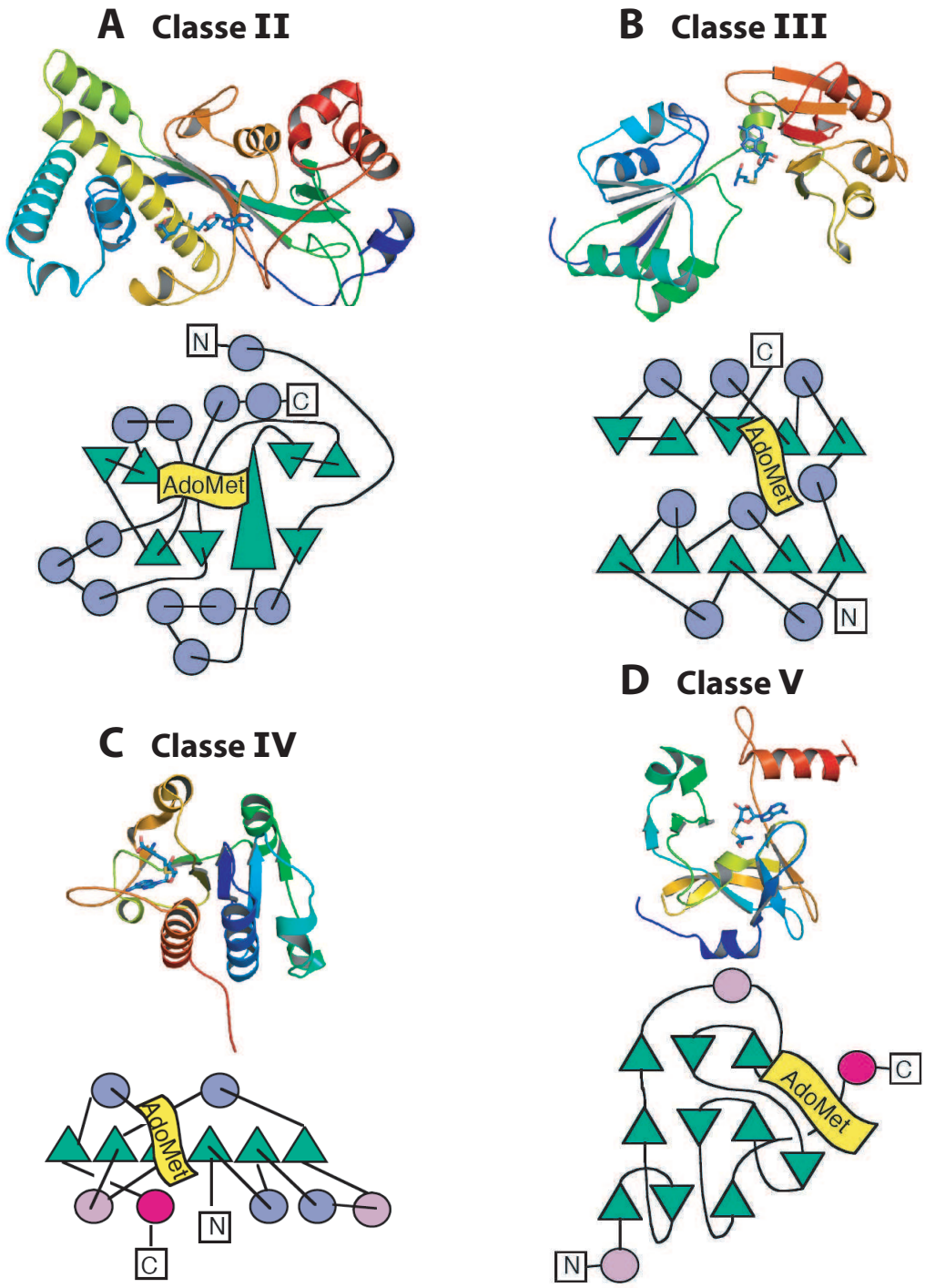


Figure 6.4 – Organisation tridimensionnelle des MTases de Classe II–V
A : MTases de Classe II. Domaine de réactivation de la méthionine synthétase (PDB code 1MSK) [317]. **B** : MTases de Classe III. Cobalt Précorrine-4 MTase (PDB code 1CBF) [318]. **C** : MTases de Classe IV, la famille SPOUT. ARNr MTase (PDB code 1MXI) [342]. **D** : MTases de Classe V, la famille SET. Lysine Histone N-MTase (PDB code 1O9S) [335]. Ces deux dernières familles présentent un nœud topologique, au niveau de l'hélice représentée *en magenta* sur le diagramme. Les hélices α sont *en bleu*, les brins β *en vert*, et le site de fixation du cofacteur AdoMet *en jaune*. Les hélices α non conservées parmi les différents membres de la famille sont *en rose*. Diagrammes topologiques d'après Schubert *et al.* [265].

Comme nous l'avons vu, la majorité des MTases SAM dépendantes se replient selon le repliement canonique des MTases de Classe I. Cependant, même si ce repliement est très conservé, les analyses comparatives de séquences de MTases ne montrent que peu de similarité [263, 333]. Des similarités de séquence existent tout de même et sont particulièrement visibles sur les régions du repliement qui participent à la liaison de la SAM. Des motifs conservés ont initialement été identifiés chez les m^5C ADN MTases [343]. Ils étaient au nombre de 10 (motifs I à X). Ils ont été numérotés dans leur ordre d'apparition sur la séquence des m^5C ADN MTases. Par la suite, il est apparu que le motif IX n'était pas conservé chez l'ensemble des DNA MTases [344]. Il est aussi apparu que l'ordre des motifs n'était pas conservé chez toutes les MTases de Classe I, mais que de façon surprenante leur localisation spatiale, elle l'était bien. Ainsi plusieurs sous-classes de DNA MTases diffèrent par des permutations circulaires de l'ordre des motifs [344, 345], qui n'affectent pas la disposition de ceux-ci par rapport aux motifs de structure secondaire. Nous allons dans le Paragraphe suivant détailler quelques points concernant ces motifs carcatéristiques des MTases.

6.1.3 Motifs conservés parmi les méthyltransférases de Classe I

L'ordre des motifs le plus fréquemment rencontré est quasi identique à celui initialement rencontré dans les m^5C ADN MTases [343]. La différence majeure réside dans la position du motif X, situé au niveau de l'hélice α_Z dans la partie N-terminale, et donc situé avant le motif I sur les séquences. L'autre différence est celle déjà évoquée de l'absence de motif IX dans la plupart des MTases. La localisation des différents motifs par rapport aux éléments de structure secondaire du repliement canonique des MTases de Classe I est présentée sur la Figure 6.5.

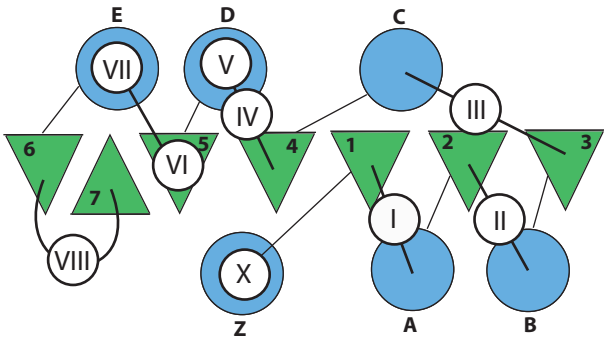


Figure 6.5 – Localisation des motifs conservés parmi les MTases de Classe I
La localisation des motifs X et I à VIII est présenté sur le diagramme topologique canonique des MTases de Classe I. Les hélices α sont *en bleu*, les feuillets β *en vert*. Pour la dénomination des hélices et des feuillets, voir le Paragraphe 6.1.2.

Les motifs I à IV prennent part à la fixation de la SAM et se retrouvent donc assez facilement parmi les différentes familles de MTases. Les motifs X, V, VI, VII et VIII sont moins bien conservés. Quelques uns de ces motifs — les motifs VI, VIII et quelques fois X — prennent part à la formation du site actif et à la catalyse [346–350]. Ils participent ainsi à la définition de la spécificité de substrat. Les motifs V et VII, situés sur les hélices α_D et α_E semblent surtout importants pour la stabilité structurale du repliement.

Avant les études structurales sur les ADN MTases, les rôles de deux motifs centraux dans le fonctionnement de ces enzymes étaient connus. Il s'agissait des motifs I et IV, dont nous dirons quelques mots dans la suite du texte. Par la suite, les études structurales combinées à des études fonctionnelles ont permis de comprendre l'implication de tous ces motifs dans la fixation des substrats et dans la catalyse.

Motif I Il comprend la séquence GxGxG, la signature du site de fixation de la SAM chez les MTases de Classe I. Les trois glycines ne sont pas toujours conservées, mais les substitutions font intervenir des résidus possédant une courte chaîne latérale [333]. Comme nous l'avons déjà mentionné, cette séquence est localisée dans la boucle reliant le brin β_1 à l'hélice α_A. Elle épouse étroitement la partie amino-carboxylate de la SAM.

Motif IV Le motif IV, situé à la charnière entre le site de fixation de la SAM et celui du substrat, se trouve à une position clé chez les MTases de Classe I. Les résidus de ce motif prennent souvent part à la catalyse du transfert de méthyle, en agissant comme base ou comme nucléophile. Ainsi, chez les m^5C ADN MTases, le motif IV contient un dipeptide invariant Pro-Cys, dont la cystéine est connue pour intervenir comme résidu catalytique nucléophile, et pour former un intermédiaire covalent avec le substrat ADN [351–355]. Chez certaines m^5C ARN MTases, ce rôle de nucléophile est joué par une cystéine conservée du motif VI [348, 349], alors que la cystéine conservée du motif IV semble intervenir dans l'étape de dissociation de l'intermédiaire covalent, en jouant le rôle de base [356, 357]. Ce dipeptide Pro-Cys n'est pas conservé chez les m^4C et m^6A ADN MTases. Chez ces enzymes, le motif IV contient la séquence conservée (D/N/S)PP(Y/F), à une position structurale analogue à celle du dipeptide Pro-Cys des m^5C ADN MTases. Le résidu D/N/S du motif IV intervient comme catalyseur basique, en acceptant le proton de l'amino exocyclique [346, 347, 358, 359]. De plus le résidu aromatique Y/F du motif IV intervient comme catalyseur cation–π [346, 347, 360–363].

Le Paragraphe 6.1 sur les Méthyltransférases nous a permis de mettre en évidence l'extraordinaire diversité de la famille des MTases SAM dépendantes, tant au niveau du type de substrats méthylés que des atomes cibles de la méthylation. Le travail de thèse présenté dans cette partie s'est intéressé à une méthyltransférase des ARN de transfert. Aussi, dans la suite du texte, allons nous nous intéresser aux méthylations rencontrées dans les ARNt.

6.2 Les méthyltransférases d'ARNt

La méthylation des nucléotides est la modification post-transcriptionnelle la plus fréquemment rencontrée dans les ARNt. Elle peut ainsi représenter entre 30 et 70 % des nucléotides modifiés selon l'ARNt considéré. Les méthylations rencontrées dans les ARNt sont très variées et agissent aussi bien au niveau de la base, qu'au niveau du ribose [364–366].

Diversité des méthylations rencontrées dans les ARNt

En ce qui concerne la méthylation des bases, les quatre nucléotides canoniques (A, U, G et C) sont sujets à méthylation. Elles ont lieu aussi bien sur les atomes de carbone que sur les atomes d'azote des cycles purine et pyrimidine, et ce sur des positions tant endocycliques qu'exocycliques. On note le nucléoside méthylé sur la base en faisant précéder la lettre correspondant à la nature du nucléoside par la lettre « m » en minuscule avec le chiffre correspondant à la position méthylée en exposant de celle-ci (pour la nomenclature des cycles purique et pyrimidique voir FIGURE 1.2 page 5). Lorsque la position exocyclique est deux fois méthylée, ceci est signalé en plaçant le chiffre 2 en indice de la lettre « m » correspondante. On obtient ainsi par exemple les nucléosides modifiés m^1A, m^7G, m^5U, m^5C, et m^2_2G (FIGURES 1.10 et 1.11 page 17).

La méthylation du ribose a quant-à-elle toujours lieu sur le groupement hydroxyle en 2′, et affecte aussi les quatre nucléotides canoniques A, U, G et C. On note le nucléotide méthylé sur l'hydroxyle en 2′ du ribose en faisant suivre la lettre correspondant à la nature du nucléotide par la lettre « m » en minuscule. On obtient ainsi par exemple les nucléotides modifiés Am, Um, Gm et Cm, et en combinant avec les modifications de la base, m^1Am ou m^5Um (FIGURES 1.10 et 1.11 page 17).

Rôle des méthylations rencontrées dans les ARNt

Le rôle des méthylations rencontrées dans les ARNt a déjà été évoqué lors du PARAGRAPHE 1.3.2. Rappelons seulement succinctement que :

- Les méthylations du ribose sur le groupement 2′OH favorisent une conformation C3′-endo. Cette conformation induit une stabilité thermodynamique supplémentaire aux nucléotides présentant cette modification, et donc aux ARNt qui en sont pourvus [110].
- De façon générale, les méthylations des bases augmente la stabilité des ARNt en favorisant les interactions d'empilement [110].
- Certaines méthylations (m^1A, m^7G et m^3C) introduisent une charge positive au niveau de la base et peuvent ainsi induire des effets électrostatiques forts en minimisant la répulsion électrostatique des groupements phosphate du squelette [108].
- Certaines modifications peuvent bloquer la face d'interaction Watson-Crick d'un nucléotide, et empêcher ainsi qu'il prenne part à des appariements canoniques. Ceci peut avoir une influence sur le repliement propre

des ARNt [112], mais aussi sur les interactions codon–anticodon et ainsi diminuer les phénomènes de décalage de cadre de lecture [96].

Repliement des MTases d'ARNt

La plupart des MTases d'ARNt appartiennent à la Classe I des MTases SAM dépendantes. Cependant, un petit nombre adopte le repliement des MTases de Classe IV (Paragraphe 6.1.2) [331, 367]. Cette classe est aussi appelée SPOUT, du nom des deux principales familles d'enzymes la constituant, SpoU et TrmD. En ce qui concerne les ARNt, les enzymes de cette classe catalysent la formation de m^1G (TrmD) [332, 368] et la méthylation sur l'hydroxyle 2′OH du ribose (SpoU, TrmH) [369, 370]. A ce jour, aucune MTase d'ARNt appartenant à une autre classe de MTases que les Classes I et IV n'a été identifiée.

Atomes cibles et mécanisme

Les atomes cibles des méthylations dans les ARNt sont aussi très variés. Et de ce fait, les mécanismes de méthylation le sont aussi. En effet, les MTases d'ARNt agissent sur des atomes d'oxygène, de carbone, d'azote ou de soufre [82]. En ce qui concerne les atomes d'oxygène, on ne trouve que des méthylations sur l'hydroxyle 2′OH du ribose. Les atomes d'azote se répartissent en deux classes distinctes : les azotes exocycliques (dans les groupements amino NH_2) et les azotes endocycliques des cycles purines (N_1 et N_7) et pyrimidines (N_3). En effet, leur nucléophilie peut être bien différente. Ainsi par exemple, les azotes exocycliques présentent toujours un doublet électronique libre, alors que les azotes endocycliques peuvent en être dépourvus, comme c'est le cas pour le N_1 des G et le N_3 des U qui doivent subir une déprotonation pour devenir nucléophiles. En ce qui concerne les atomes de carbone, il s'agit presque exclusivement de méthylation en position 5 des cycles pyrimidines C et U.

La question générale sous entendue derrière celle du mécanisme, dans le cas des MTases, est la suivante : comment l'enzyme active-t-elle la nucléophilie de l'atome cible de la méthylation ? Dans bien des cas, il semble que le substrat à méthyler soit suffisamment nucléophile pour que la simple fixation des deux substrats, nucléophile et SAM, dans une orientation adéquate, suffisent à la catalyse [82]. Dans d'autres cas, le nucléophile est en quelque sorte caché et l'enzyme doit d'abord révéler ce caractère du substrat soit par une déprotonation — comme dans le cas des m^1G MTases [332, 368] — soit par attaque nucléophile — comme dans le cas des m^5C [349] et m^5U MTases [371, 372].

Ainsi, à l'image de la diversité des méthylations rencontrées dans les nucléotides modifiés, la famille des MTases d'ARNt est extrêmement variée — tant au niveau de leur repliement et donc de leur origine phylogénétique qu'au niveau de la chimie du transfert de méthyle. Le travail retracé dans cette Partie a consisté en l'étude d'une m^1A MTase d'ARNt. Aussi dans la suite allons nous nous intéresser plus particulièrement à ce nucléotide modifié, et aux enzymes catalysant cette modification.

6.3 Les m^{1}A méthytransférases d'ARNt

6.3.1 Le nucléotide m^{1}A et sa localisation au sein des ARNt

La découverte et l'identification du nucléotide m^{1}A dans l'ARN date du début des années 60 [373, 374]. Sa structure chimique est représentée à la FIGURE 6.6 **A**. C'est l'un des rares nucléotides à posséder une charge positive (avec les nucléotides m^{7}G et m^{3}C). Comme nous l'avons déjà plusieurs fois mentionné, ceci pourrait jouer un rôle non négligeable dans la stabilité des ARNt, grâce à des interactions électrostatiques fortes [108].

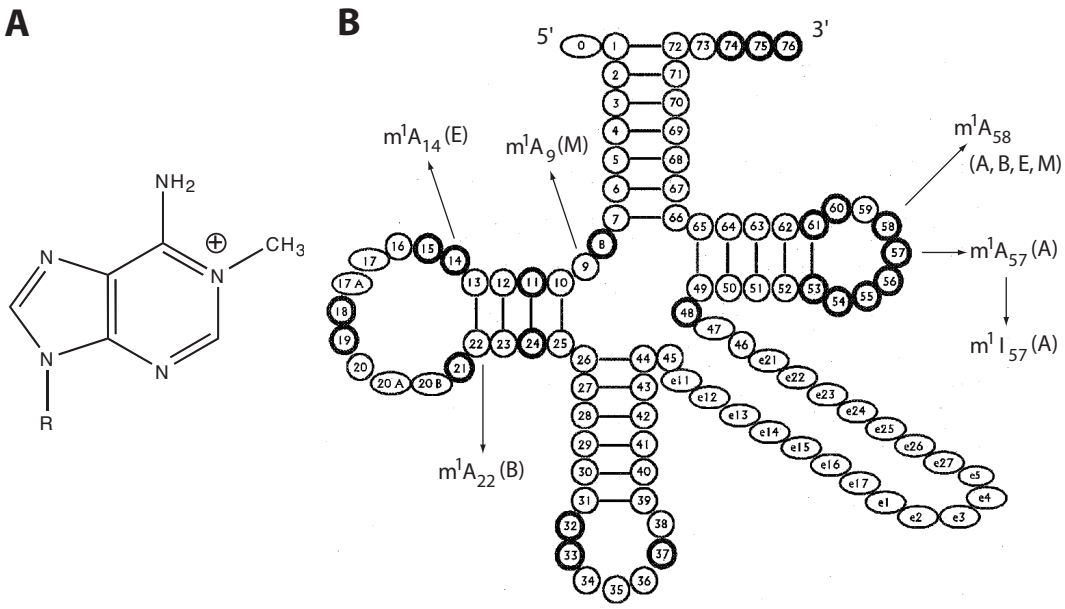


FIGURE 6.6 – Le nucléotide m^{1}A dans les ARNt
A : Structure chimique de la base m^{1}A. **B** : Positions auxquelles est rencontré le nucléotide modifié m^{1}A. Les lettres A, E et B désignent le règne du vivant dans lequel on retrouve cette modification (*Archaea*, *Eukarya* et *Bacteria*). M signifie que la modification est rencontrée dans les ARNt mitochondriaux. D'après Anderson & Droogmans [375].

Le nucléotide m^{1}A est rencontré dans les ARNt aux positions 9, 14, 22, 57 (de façon transitoire) et 58 (FIGURE 6.6 **B**). Alors que la modification aux positions 9, 14, 22 et 57 n'est caractéristique que d'un des règnes du vivant (*Archaea*, *Eukarya* ou *Bacteria*), ou des ARNt mitochondriaux, la modification m^{1}A à la position 58 est quant-à-elle rencontrée parmi les ARNt des trois règnes du vivant. Le nucléotide m^{1}A en position 57 est un intermédiaire sur la voie de biosynthèse du nucléotide modifié m^{1}I chez les *Archaea* (FIGURE 6.6 **B** et FIGURE 6.8).

Le fait que le nucléotide modifié m^1A soit très souvent rencontré en position 58 — environ 25 % des ARNt séquencés portant un A en position 58 sont modifiés en m^1A_{58} — et ce chez les trois domaines du vivant, a communément fait admettre que cette modification conservée par l'évolution jouait un rôle dans la structure et/ou la fonction des ARNt [376, 377]. Nous décrirons dans la suite ce que l'on connaît actuellement concernant la modification m^1A_{58}.

6.3.2 Les m^1A_{58} MTases

Même si le nucléotide modifié m^1A est connu depuis plus de 40 ans [373, 374], et que des études réalisées sur des extraits cellulaires d'organisme thermophiles [378, 379] ou de foie de rat [380] ont permis de mettre en évidence des activités de m^1A MTase, l'étude de cette famille d'enzyme n'a réellement pris de l'ampleur que depuis l'identification de la m^1A_{58} MTase de levure en 2000 par l'équipe de James Anderson. En effet, la purification et la caractérisation de la m^1A_{58} MTase de *S. cerevisiae* [381, 382] a permis d'identifier et de cloner les MTases correspondantes chez l'homme [383], chez *T. thermophilus* [111], et chez *P. abyssi* [384].

La famille des m^1A_{58} MTases

Le travail initial de l'équipe de James Anderson a montré que la m^1A_{58} MTase de levure est une enzyme hétérotétramérique $\alpha_2\beta_2$ codée par deux gènes essentiels, GCD10 et GCD14 (renommés par la suite TRM6 et TRM61), à l'origine respectivement des deux sous-unités de l'enzyme. La sous-unité Trm61p est responsable de la fixation de la SAM et de la catalyse mais ne peut pas catalyser seule la réaction de transfert de méthyle puisque les deux sous-unités Trm61p et Trm6p sont essentielles à la fixation du substrat ARNt [381, 382, 385]. Des homologues de la sous-unité Trm61p ont ensuite été identifiés chez les procaryotes [377], et leur activité présumée de m^1A_{58} MTase a par la suite été vérifiée chez *T. thermophilus* [111], *P. abyssi* [384] et *M. tuberculosis* [386]. De façon intéressante, la structure de cette dernière — la m^1A_{58} MTase de *M. tuberculosis* — avait été déterminée en 2001 sans que sa fonction soit alors connue [387]. Actuellement, les seules structures de m^1A_{58} MTases caractérisées en tant que telles sont celle de *M. tuberculosis* (PDB code 1I9G) [387], et celle de *T. thermophilus* (PDB code 2PWY) [350]. Contrairement aux m^1A_{58} MTases eucaryotes, les enzymes procaryotes (appelées TrmI) ne possèdent pas la sous-unité Trm6p. Ces enzymes sont ainsi des homotétramères α_4.

Un alignement de séquences de la famille des m^1A_{58} MTases est présenté à la Figure 6.7. Les séquences des protéines procaryotiques TrmI de *T. thermophilus*, *M. tuberculosis* et *P. abyssi* y sont alignés avec leurs homologues Trm61p de *S. cerevisiae*, *H. sapiens* et *M. musculus*. Cet alignement montre que ces enzymes ne partagent pas seulement le site de fixation de la SAM, mais aussi les autres motifs caractéristiques des MTases de Classe I (Voir Paragraphe 6.1.3 pour la description des motifs).

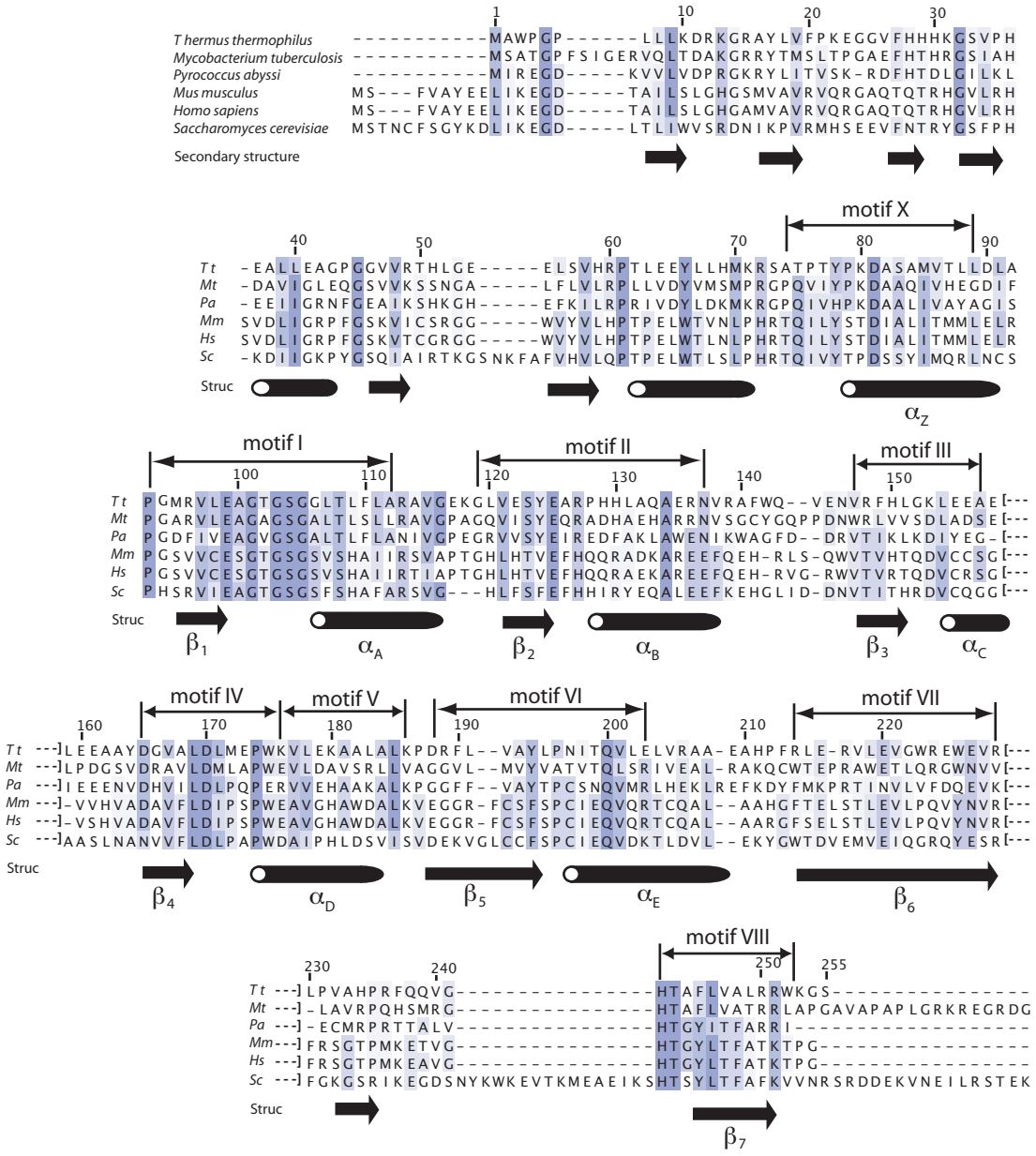


FIGURE 6.7 – Alignement de séquences de la famille m^{1}A$_{58}$ MTases
Les séquences des protéines procaryotiques TrmI de *T. thermophilus*, *M. tuberculosis* et *P. abyssi* y sont alignés avec leurs homologues Trm61p de *M. musculus*, *H. sapiens* et *S. cerevisiae*. La numérotation des résidus correspond à la séquence de *T. thermophilus*. Les résidus très conservés sont représentés sur fond bleu foncé, et les résidus présentant les mêmes propriétés physico-chimiques sont représentés sur fond bleu clair. La structure secondaire de *T. thermophilus* TrmI est représentée sous les séquences avec des flèches pour les brins β et des cylindres pour les hélices α. Les noms des éléments de structure secondaire correspondent à ceux définis à la FIGURE 6.3. Les motifs conservés parmi les MTases SAM dépendantes de Classe I sont également représentés. Leur délimitation correspond à celle proposée par Bujnicki [377].

Particularités biochimiques des m^1A_{58} MTases

Spécificité vis-à-vis du substrat Une étude réalisée sur l'enzyme TrmI de *T. thermophilus* purifiée à partir du thermophile s'est intéressée à la spécificité de cette enzyme vis-à-vis du substrat ARNt [378]. Cette étude a montré que l'enzyme était capable de transférer le groupement méthyle de la SAM à la moitié 3′ de l'ARNt initiateur de *E. coli*, suggérant que le site de reconnaissance principal doit se situer dans le bras T de l'ARNt et que l'enzyme ne reconnaît pas l'architecture générale de l'ARNt. Il a également été montré que la m^1A_{58} MTase de *Tetrahymena pyriformis*, purifiée à partir du protozoaire, pouvait méthyler un ARN de 17 nucléotides correspondant au bras T d'un ARNt [82]. De plus, l'enzyme TrmI de *P. abyssi* est capable de méthyler un fragment d'ARNt comprenant la tige acceptrice prolongée de la tige-boucle TΨC [379]. Ces résultats sont également en accord avec les résultats d'expériences réalisées par Jaunius Urbonavicius sur les enzymes TrmI de *T. thermophilus* et de *P. abyssi*, qui ont montré que ces deux enzymes étaient capables de transférer le groupement méthyle de la SAM à une tige-boucle de 17 nucléotides mimant le bras T de l'ARNt initiateur de *T. thermophilus* et portant les deux modifications conservées T_{54} et Ψ_{55} [388]. Il a également été montré que la m^1A_{58} MTase de *Xenopus laevis* a la capacité de méthyler des ARNt portants des mutations qui affectent leur structure tridimensionnelle [389]. Ainsi, la m^1A_{58} MTase de *Xenopus laevis* ne reconnaît pas, elle non plus, la structure tridimensionnelle de l'ARNt, mais seulement des éléments de structure secondaire.

Tous ces résultats suggèrent que les m^1A_{58} MTases, procaryotes et eucaryotes, possèdent leur motif principal de reconnaissance dans le bras T des ARNt.

Multispécificité de l'enzyme TrmI de *P. abyssi* La biosynthèse du nucléotide modifié m^1I_{57} chez les *Archaea* est un processus en deux étapes. Le nucléotide A_{57} est tout d'abord méthylé en m^1A_{57} puis celui-ci est désaminé pour donner le nucléotide m^1I_{57} (Figure 6.8) [390]. Cette voie diffère de celle de la synthèse du nucléotide m^1I_{37} chez les eucaryotes, où la désamination précède la méthylation [391].

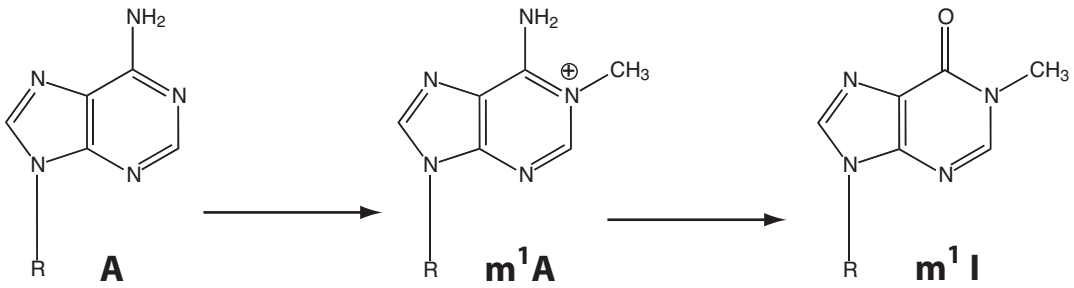


Figure 6.8 – Voie de biosynthèse du nucléotide m^1I_{57} chez les *Archaea*
Le nucléotide A_{57} est d'abord méthylé en m^1A_{57} puis désaminé en m^1I_{57}.

L'étude biochimique de l'enzyme TrmI de *P. abyssi* a révélé que celle-ci était multispécifique et qu'elle catalysait la formation du nucléotide modifié m^1A_{58}, comme les autres membres de la famille TrmI, mais aussi celle du nucléotide modifié m^1A_{57} [384]. Cette enzyme a été qualifiée de région-spécifique, car elle n'est pas spécifique d'un nucléotide donné sur la séquence, mais de la région de la boucle T des ARNt.

Fonctions avérées de la modification m^1A_{58}

Il est assez délicat de déterminer la fonction d'une modification post-transcriptionnelle particulière. En effet, la plupart des enzymes de modifications des ARNt ne sont pas essentielles à la survie cellulaire, et les mutants défectueux dans la synthèse d'une seule de ces modifications ne présentent que rarement un phénotype identifiable. L'inactivation de la m^1A_{58} MTase a cependant permis la mise en évidence de phénotypes chez *T. thermophilus* et chez *S. cerevisiae*, permettant ainsi de comprendre le rôle de cette modification chez ces deux organismes.

Chez *Thermus thermophilus* *Thermus thermophilus* est une bactérie thermophile dont la croissance se révèle optimale entre 48 et 85 °C. Les molécules de tels organismes sont plus résistantes à la dénaturation thermique que celles des organismes mésophiles. Dans le cas des ARNt, les modifications post-transcriptionnelles permettent de stabiliser leur structure à haute température [392, 393]. Certaines modifications voient même leur biosynthèse activée par une élévation de température, comme dans le cas de la thiolation s^2T_{54} [394], dont la présence au sein des ARNt influe sur leur stabilité vis-à-vis de la dénaturation thermique [395].

De façon intéressante, un mutant de *T. thermophilus*, dont le gène *trmI* a été inactivé par l'introduction d'une cassette de résistance à un antibiotique, a révélé un phénotype de sensibilité à la température [111]. Ce mutant présente en effet un défaut de croissance à 80 °C. Ceci suggère bien évidemment que la méthylation de la base A_{58} stabilise la structure des ARNt à haute température.

Chez *Saccharomyces cerevisiae* L'identification de la m^1A_{58} MTase de levure par l'équipe de James Anderson est venue de l'observation que des mutants au niveau des gènes GCD10 et GCD14 présentaient une quantité d'$ARNt_i^{Met}$ initiateur réduite ainsi qu'une absence de nucléotides modifiés m^1A dans leurs ARNt [381]. Les auteurs ont proposé que l'absence de modification m^1A_{58} donnerait une structure aberrante de l'$ARNt_i^{Met}$ initiateur ce qui conduirait à sa dégradation sélective, et que cette modification n'influerait pas sur la structure des ARNt élongateurs [381].

Cette hypothèse pourrait paraître étrange, mais elle est parfaitement en accord avec la structure de l'$ARNt_i^{Met}$ initiateur de levure [36]. En effet, l'$ARNt_i^{Met}$ possède une structure unique au niveau des boucles T et D (Figure 6.9). Cette structure locale unique implique trois adénosines conservées parmi les ARNt

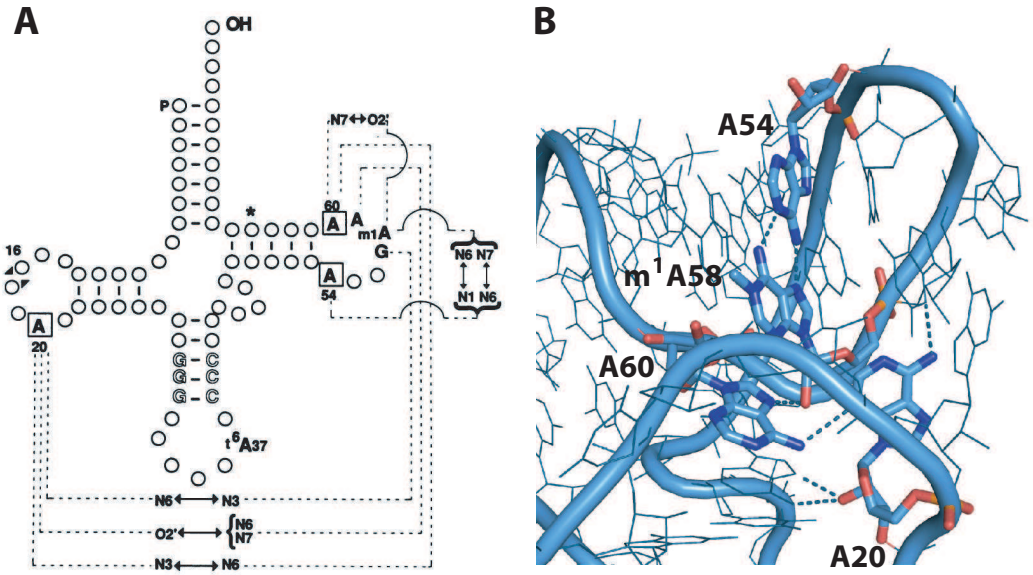


FIGURE 6.9 – Structure particulière de l'ARNt$_i^{Met}$ initiateur de levure
A : Structure secondaire de l'ARNt$_i^{Met}$ initiateur de levure montrant les interactions particulières par liaisons hydrogène entre les adénosines A_{20}, A_{60}, A_{54} et m^1A_{58}. D'après Basavappa & Sigler [36]. **B** : Agrandissement de la région des boucles T et D de l'ARNt$_i^{Met}$ initiateur de levure montrant les liaisons hydrogènes entre les adénosines A_{20}, A_{60}, A_{54} et m^1A_{58} (PDB code 1YFG) [36].

initiateurs eucaryotes : A_{20}, A_{60} et A_{54}. Un jeu de liaisons hydrogènes particulier a lieu entre ces trois adénosines par l'intermédiaire de la m^1A_{58} (FIGURE 6.9 **A** et **B**). La m^1A_{58} semble ainsi jouer un rôle important dans cette structure locale, où elle effectue une paire *Trans*-Hoogsteen avec le A_{54} (A_{54}○–□m^1A_{58}). Il n'est ainsi pas déraisonnable de penser que cette structure locale particulière puisse être modifiée par la seule perte de la modification m^1A_{58}.

Il a par la suite été montré que l'absence de la modification m^1A_{58} dans l'ARNt$_i^{Met}$ initiateur conduisait à la polyadénylation de l'extrémité 3′ de l'ARNt par la polyA-polymérase Trf4p en complexe avec l'hélicase Mtr4p et la protéine de fixation à l'ARN Air2p. Le précurseur ARNt$_i^{Met}$ ainsi polyadénylé est ensuite dégradé de 3′ en 5′ par digestion exonucléotidique par l'exosome nucléaire [396–398].

Ainsi, de façon générale, la modification m^1A_{58} retrouvée dans la boucle TΨC de nombreux ARNt des trois domaines du vivant participe de façon non négligeable à la structure et à la stabilité de ceux-ci. Dans le cas des procaryotes thermophiles, cette modification contribue à la stabilité des ARNt vis-à-vis de la dénaturation thermique, et chez les eucaryotes elle participe à une structure locale unique des ARNt initiateurs, son absence entraînant une désorganisation de cette structure et une dégradation de l'ARNt$_i^{Met}$ ainsi mal replié.

Chapitre 7

Etude de la m^1A_{58} ARNt MTase de *T. thermophilus*

7.1 Importance de l'étude structurale des enzymes de modification des ARN

La résolution de structures tridimensionnelles d'enzymes de modifications en complexe avec leur substrat ARN est un enjeu majeur pour comprendre les phénomènes de reconnaissance ARN–protéine, car celles-ci ne font pas intervenir les domaines relativement bien caractérisés de liaisons aux ARN [399–401] — tels que les domaines RRM (*RNA recognition motif*) [402], KH (*K-homology*), dsRBD (*double-stranded RNA-binding domain*) et les domaines en doigts de zinc ZnF (*Zinc finger*). A l'heure actuelle, seulement neuf structures d'enzymes de modification en complexe avec leur substrat ARN ont été résolues, et celles-ci ont à chaque fois apporté des éléments nouveaux à la compréhension des interactions ARN–protéines [403–411]. Parmi ces neuf complexes, sept impliquent les ARNt, une les ARNr et la dernière enzyme présente une double spécificité ARNt et ARNr. Les neuf structures d'enzymes en complexe avec leur substrat ARN sont :

- deux pseudouridine synthases (TruB) d'organismes différents, à l'origine de la formation du Ψ_{55} des ARNt [403, 404]
- deux guanine transglycosylases des ARNt (ArcTGT et TGT) [405, 406]
- une m^5U MTase des ARNr (RumA) [407]
- une m^5U à l'origine de la formation du T_{54} des ARNt (TrmA) [411]
- une désaminase (TadA) à l'origine de la formation d'inosine en position wobble des ARNt [408]
- une thiouridylase (MnmA) à l'origine de la formation de s^2U en position wobble des ARNt [409]
- et enfin une pseudouridine synthase (RluA) présentant la double spécificité ARNt et ARNr [410]

Parmi les structures de complexes impliquant un ARNt, deux seulement ont été résolues en présence d'un ARNt de pleine taille [405, 409]. La plupart

ont en effet été résolues avec des mini-substrats de l'enzyme, en l'occurrence des tiges-boucles mimant la région concernée par la modification — le bras T [403, 404, 411] ou la tige-boucle anticodon [406, 408, 410].

Même si les enzymes de modification ne présentent pas les domaines classiques de liaison aux ARN (RRM, KH, dsRBD, ZnF), elles présentent toutefois, et de manière plus ou moins universelle, une organisation modulaire, avec un domaine catalytique et des domaines additionnels, dont l'un des rôles supposés serait de participer à la reconnaissance de l'ARN cible. Ces domaines additionnels, bien qu'ils ne leur soient pas spécifiques, sont principalement retrouvés chez les enzymes de modification. Il s'agit par exemple des domaines PUA (pour *PseudoUridine synthase* & *Archeosine transglycosylase*) [412], THUMP (pour *THioUridine synthases, RNA MTases* & *Pseudouridine synthases*) [413] et TRAM (pour *TRm2* & *MiaB domain*) [414]. Ces domaines ont été très peu caractérisés, et même si certaines structures ont montré une participation évidente de certains d'entre eux dans la fixation de l'ARN cible (TRAM & PUA) [405, 407], tous ces domaines supposés de liaison à l'ARN n'ont pas d'affinité propre pour leur substrat ARN *in vitro* [415, 416]. Il a ainsi été proposé que ces domaines additionnels participeraient à la stabilisation de l'ARN dans une conformation active, nécessaire à la catalyse, mais ne seraient pas les premiers acteurs de la reconnaissance.

Ainsi, de nouvelles structures de complexes ARN–enzyme de modification, apporteraient à coup sûr des éléments nouveaux et passionnants quant aux interactions ARN–protéines. Notamment, toutes les structures d'enzymes de modification en complexe avec leur substrat ARN [403–411] ont montré des déformations plus ou moins importantes de l'ARN — allant d'un simple retournement de la base à modifier, si celle-ci est située dans une boucle accessible comme la boucle anticodon des ARNt [406, 408–410] — à une déformation globale de l'ARNt, quittant alors son repliement classique en « L » pour adopter une structure en « λ » dans laquelle les boucles T et D deviennent accessibles (Figure 7.1) [405]. Il serait ainsi intéressant de voir si cette forme « λ » de l'ARNt est une forme reconnue par différentes enzymes de modifications des boucles T et D, où si d'autres formes non canoniques, dans lesquelles le cœur de l'ARNt est accessible, peuvent être stabilisées par ces enzymes.

Les études structurales ont aussi permis dans quelques cas d'élucider les mécanismes d'action de ces enzymes de modification — de façon très élégante pour la pseudouridine synthase TruB, la guanine transglycosylase TGT et les m^5U d'ARNr et d'ARNt [404, 406, 407, 411]. En effet, l'emploi de certains substrats modifiés ou d'enzymes mutées au niveau de résidus catalytiques peut bloquer la réaction à une étape précise du mécanisme peut ainsi permettre de piéger des intermédiaires covalents et donc d'identifier les résidus clés de l'interaction, ainsi que les différentes étapes du mécanisme. Ces intermédiaires covalents peuvent également faciliter l'obtention de co-cristaux.

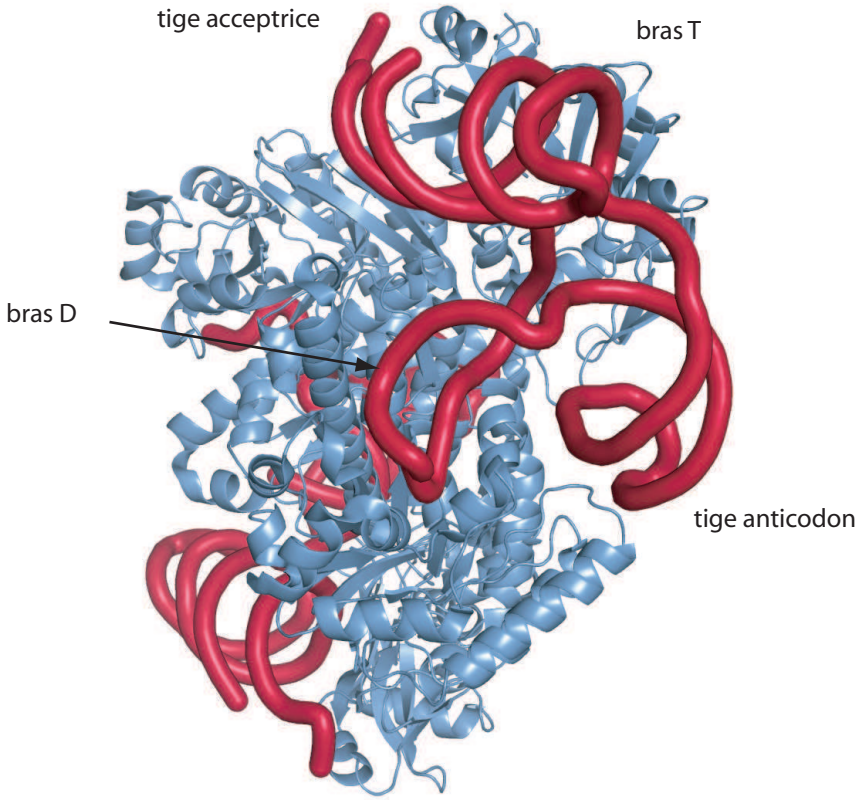


FIGURE 7.1 – Forme « λ » de l'ARNt
Structure cristalline du complexe ArcTGT/ARNt (PDB code 1J2B) [405] dans laquelle la protéine ArcTGT stabilise une forme dite « λ » de l'ARNt. Le bras D n'est plus empilé dans l'axe de la tige anticodon et les boucles T et D ne sont plus intimement liées et sont ainsi accessibles.

La biologie structurale des enzymes de modification des ARNt est ainsi actuellement en plein essor, et c'est dans ce contexte que nous avons entrepris l'étude structurale d'une enzyme de modification des ARNt — la m^1A_{58} ARNt MTase de *T. thermophilus* ou TrmI, une méthyltransférase qui agit sur la position 1 de la base A_{58} de la boucle T des ARNt (pour la nomenclature des positions des bases voir FIGURE 1.2 page 5).

7.2 Projet concernant la m^1A_{58} ARNt MTase

Initialement, notre intérêt concernant cette enzyme n'était pas de réaliser une étude structurale, mais de l'utiliser pour introduire la modification m^1A_{58} dans l'$ARNt_3^{Lys}$ humain produit dans *E. coli*. Nous utilisons en effet au laboratoire cet ARNt comme modèle pour l'étude de la formation du complexe d'initiation de la transcription inverse du VIH-1 (voir PARTIE I). Celui-ci, bien que fonctionnel dans les tests de transcription inverse [113], est dépourvu de certaines modifications post-transcriptionnelles — notamment la m^1A_{58} car *E. coli* ne possède pas la MTase correspondante. Ceci pourrait influer sur la

fusion de l'$ARNt_3^{Lys}$ amorce et donc jouer un rôle non négligeable dans la formation de l'hybride $ARNt_3^{Lys}$/ARN viral. Le clone permettant la surproduction de la m^1A_{58} ARNt MTase de *T. thermophilus* [111] nous a été donné par l'équipe de Louis Droogmans (Laboratoire de Microbiologie, Université Libre de Bruxelles, Belgique), ce qui nous a permis de réaliser avec succès la méthylation de l'$ARNt_3^{Lys}$. Mais une influence de cette modification sur la formation de la fusion de l'$ARNt_3^{Lys}$ n'a pas pu être mise en évidence. Cette étude à été réalisée pendant mon stage de Master 2 qui a directement précédé mon travail de thèse. Elle n'a pas été poursuivie et nous n'en parlerons plus dans la suite.

Comme nous étions alors en possession d'une grande quantité de protéine TrmI purifiée, et que nous disposons au laboratoire de tout l'arsenal nécessaire à la cristallogénèse — notamment un robot permettant de tester sur la même boîte les 96 conditions des kits commerciaux — il n'a pas été difficile de réaliser les premiers essais de cristallisation. Ceux-ci ont donné de nombreuses pistes encourageantes et c'est à ce moment que nous avons en quelque sorte débuté le projet d'étude structurale de TrmI. Nous avons rapidement envisagé de réaliser la structure de TrmI en complexe avec son partenaire ARN, car l'équipe possède un grand savoir faire dans la production et la purification des ARNt. Pour cela, nous avons purifié divers ARNt pleine longueur — d'une part l'$ARNt_f^{Met}$ initiateur d'*E. coli*, car nous savions que nous pouvions l'obtenir très pur, contrairement à d'autres ARNt dont nous réalisons la production au laboratoire — et d'autre part l'$ARNt^{Phe}$ de levure, qui est commercial, et dont la cristallisation est abondamment documentée. Nous avons aussi réalisé les essais de co-cristallisation de TrmI avec divers fragments d'ARN — essentiellement des tiges-boucles mimant le bras T d'un ARNt — car nous savions grâce à des expériences réalisées par Jaunius Urbonavicius, alors en stage post-doctoral dans l'équipe d'Henri Grosjean et de Béatrice Golinelli (Laboratoire d'Enzymologie et de Biochimie Structurale, Gif-sur-Yvette), que ces tiges-boucles étaient substrat de notre enzyme.

Cependant, que ce soit pour les ARNt pleine longueur ou les petits fragments, nous n'avons pas obtenu de co-cristaux et n'avons donc pas résolu la structure du complexe. On peut tenter de proposer quelques raisons à cela :

- D'une part, la protéine TrmI a révélé une forte propension à cristalliser seule — ce qu'ont montré tous les essais de cristallisation — et cette cristallisation rapide l'exclurait d'une co-cristallisation plus lente.
- D'autre part, étant donné le mécanisme de la réaction de transfert de méthyle — celui-ci est présenté dans l'Article 3, et sera plus amplement discuté dans la suite de ce Chapitre — il n'est pas envisageable de réaliser un complexe covalent enzyme–ARN, ce qui a dans de nombreux cas aidé à la co-cristallisation [404, 406, 407].
- De plus, l'association TrmI–ARN est assez fortement dépendante de la force ionique (voir Article 3) et de nombreuses conditions de cristallisation contiennent des sels à forte concentration.

Par exemple, et de façon assez surprenante, nous avons obtenu dans la même solution de cristallisation, à savoir 2.0 M de sulfate d'ammonium, des cristaux de la protéine TrmI seule, ainsi que des cristaux du partenaire ARNt seul. Il s'agissait en l'occurrence de l'ARNt$_f^{Met}$ initiateur d'*E. coli*. La structure de cet ARNt avait été résolue en 1980 dans l'équipe d'Alexander Rich [417]. Cependant, les coordonnées n'étaient pas disponibles dans la *Protein Data Bank* (PDB) et nous avons pensé qu'il pourrait être intéressant d'affiner de nouveau cette structure avec les outils dont nous disposons actuellement. Nous avons donc réalisé l'affinement de la structure de cet ARNt$_f^{Met}$. Cette structure s'est révélée être particulièrement intéressante car elle a montré une organisation unique de l'anticodon qui pourrait être à l'origine de la discrimination entre les ARNt initiateurs et élongateurs. Ce travail est présenté dans la Partie III.

Ainsi, même si nous n'avons pas pu obtenir de cristaux du complexe TrmI–ARNt, nous avons obtenu des cristaux de chacun des deux partenaires. Les cristaux de TrmI nous ont permis de résoudre la structure de l'enzyme par remplacement moléculaire [1] et de l'affiner jusqu'à 1.7 Å de résolution. Nous avons par ailleurs caractérisé son interaction avec son partenaire ARNt par des techniques biochimiques et biophysiques. Ceci a fait l'objet de l'Article 3.

7.3 Problématique et résumé de l'Article 3

Les questions que nous nous posions alors et auxquelles nous avons donné une réponse dans cet article concernaient deux points distincts :

La première de ces questions nous intéressait tout particulièrement, d'un point de vue pratique, pour les essais de co-cristallisation TrmI/ARN que nous avons toujours poursuivis. Il s'agissait de la détermination de la stœchiométrie d'interaction entre la protéine TrmI qui est tétramérique et son substrat ARNt. Nous avons répondu à cette question grâce à des expériences de Spectroscopie de Masse réalisées en conditions non-dénaturantes. Celles-ci ont été réalisées par Cédric Atmanene, étudiant en thèse sous la direction de Sarah Sanglier, dans l'équipe d'Alain Van Dorsselaer (Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC, Université Louis Pasteur, Strasbourg). Ces expériences ont montré que la protéine TrmI reste sous forme de tétramère lorsqu'elle fixe son substrat ARNt, et qu'elle est capable de fixer jusqu'à deux ARNt. Les quatre sites actifs ne peuvent ainsi pas être occupés par quatre ARNt en même temps. Ceci a été confirmé par des expériences de gel retard en conditions natives. Cette stœchiométrie d'interaction de deux ARNt par tétramère est en parfait accord avec le fait que les protéines TrmI eucaryotes aient évoluées depuis un homotétramère α_4 vers un hétérotétramère $\alpha_2\beta_2$, ou chaque hétérodimère $\alpha\beta$ pourrait interagir avec un ARNt.

1. Le remplacement moléculaire a été effectué en utilisant la structure de la protéine TrmI homologue de *Mycobacterium tuberculosis* comme modèle (PDB code 1I9G) [387].

L'autre question concernait le mécanisme d'action de cette enzyme. Contrairement aux m^6A MTases d'ADN et d'ARN, le mécanisme des m^1A MTases n'avait jamais été étudié. Nous avons cherché à appréhender celui-ci par l'intermédiaire de variants de l'enzyme. En effet, la structure ainsi que les alignements des séquences des protéines de la famille de TrmI, nous ont permis d'identifier trois résidus potentiellement impliqués dans la catalyse. Ces trois résidus — à savoir les résidus aromatiques Y_{194} et Y_{78}, et le résidu D_{170} — sont soit conservés comme résidus aromatiques (Y_{194} et Y_{78}), soit universellement conservé (D_{170}) parmi les représentants de la famille TrmI (FIGURE 6.7). Ils sont de plus situés à proximité directe du méthyle activé de la SAM, et pourraient ainsi intervenir dans la catalyse du transfert de méthyle. Nous avons réalisé des expériences de cinétique enzymatique sur des variants de TrmI affectés au niveau de ces résidus (D170A, Y78A et Y194A) qui ont confirmé leur implication directe dans la catalyse. Ces expériences, nécessitant l'utilisation de SAM portant un groupement méthyle marqué radioactivement [méthyl–^{3}H]SAM, ont été réalisées dans l'équipe de Béatrice GOLINELLI dans le cadre d'une collaboration, car d'une part nous ne possédons par d'agrément pour la manipulation de radioactivité au laboratoire et que d'autre part son équipe possède un grand savoir faire dans la réalisation de ces cinétiques enzymatiques (Voir Matériel Supplémentaire de l'Article 3 FIGURE 7.2 page 131). Nous avons ensuite pu utiliser ces résultats sur les variants de TrmI comme des informations solides pour réaliser l'amarrage moléculaire (*docking*) d'une adénine dans le site actif de TrmI. Cette modélisation a été réalisée à l'aide du programme HADDOCK [2] (*High Ambiguity Driven biomolecular DOCKing*) [418, 419] qui permet d'utiliser des contraintes ambiguës — c'est-à-dire entre groupe d'atomes, et pas entre un couple unique d'atomes. C'est ce type de contraintes ambiguës que l'on obtient typiquement avec des expériences de mutagénèse. Cette modélisation a donné une solution unique qui nous a permis de proposer des rôles individuels pour chacun des résidus identifiés.

7.4 Article 3 | Barraud *et al.* J. Mol. Biol. (2008)

2. Le programme HADDOCK, ainsi qu'une documentation détaillée sur ses nombreuses applications et leurs mises en œuvre sont accessibles à l'adresse suivante : `http://www.nmr.chem.uu.nl/haddock/`

doi:10.1016/j.jmb.2008.01.041 *J. Mol. Biol.* (2008) **377**, 535–550

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structure of *Thermus thermophilus* tRNA m^1A_{58} Methyltransferase and Biophysical Characterization of Its Interaction with tRNA

Pierre Barraud[1], Béatrice Golinelli-Pimpaneau[2], Cédric Atmanene[3], Sarah Sanglier[3], Alain Van Dorsselaer[3], Louis Droogmans[4], Frédéric Dardel[1] and Carine Tisné[1]*

[1]*Laboratoire de Cristallographie et RMN biologiques, Université Paris-Descartes, CNRS UMR 8015, 4 avenue de l'Observatoire, 75006 Paris, France*

[2]*Laboratoire d'Enzymologie et Biochimie Structurales, CNRS Bâtiment 34, 1 avenue de la Terrasse, 91190 Gif-sur-Yvette, France*

[3]*Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC-DSA, ULP, CNRS UMR 7178, 25 rue Becquerel, 67087 Strasbourg, France*

[4]*Laboratoire de Microbiologie, Institut de Recherches Microbiologiques Wiame, Université Libre de Bruxelles, avenue E. Gryson 1, B-1070 Bruxelles, Belgium*

Received 26 November 2007; received in revised form 11 January 2008; accepted 16 January 2008
Available online 26 January 2008

Edited by J. Doudna

Methyltransferases from the m^1A_{58} tRNA methyltransferase (TrmI) family catalyze the *S*-adenosyl-L-methionine-dependent N_1-methylation of tRNA adenosine 58. The crystal structure of *Thermus thermophilus* TrmI, in complex with *S*-adenosyl-L-homocysteine, was determined at 1.7 Å resolution. This structure is closely related to that of *Mycobacterium tuberculosis* TrmI, and their comparison enabled us to enlighten two grooves in the TrmI structure that are large enough and electrostatically compatible to accommodate one tRNA per face of TrmI tetramer. We have then conducted a biophysical study based on electrospray ionization mass spectrometry, site-directed mutagenesis, and molecular docking. First, we confirmed the tetrameric oligomerization state of TrmI, and we showed that this protein remains tetrameric upon tRNA binding, with formation of complexes involving one to two molecules of tRNA per TrmI tetramer. Second, three key residues for the methylation reaction were identified: the universally conserved D170 and two conserved aromatic residues Y78 and Y194. We then used molecular docking to position a N_9-methyladenine in the active site of TrmI. The N_9-methyladenine snugly fits into the catalytic cleft, where the side chain of D170 acts as a bidentate ligand binding the amino moiety of *S*-adenosyl-L-methionine and the exocyclic amino group of the adenosine. Y194 interacts with the N_9-methyladenine ring, whereas Y78 can stabilize the sugar ring. From our results, we propose that the conserved residues that form the catalytic cavity (D170, Y78, and Y194) are essential for fashioning an optimized shape of the catalytic pocket.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: TrmI; m^1A_{58} methyltransferase; X-ray structure; noncovalent mass spectrometry; protein–RNA interactions

Introduction

Functional tRNAs carry a number of chemically modified nucleosides that are formed enzymatically after transcription, during the tRNA maturation process. To date, >90 different modifications have been identified in tRNAs from various organisms.[1] Among these posttranscriptional nucleoside modifications, N_1-methyladenosine (m^1A) is found at a highly conserved A_{58} position in the TΨC loop of

*Corresponding author. E-mail address: carine.tisne@univ-paris5.fr.

Abbreviations used: TrmI, m^1A_{58} tRNA methyltransferase; m^1A, N_1-methyladenosine; AdoMet, *S*-adenosyl-L-methionine; MTase, methyltransferase; *myco*TrmI, *Mycobacterium tuberculosis* TrmI; *th*TrmI, *Thermus thermophilus* TrmI; AdoHcy, *S*-adenosyl-L-homocysteine; ESI-MS, electrospray ionization mass spectrometry; PDB, Protein Data Bank; AIR, ambiguous interaction restraint.

0022-2836/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.

many tRNAs in the three domains of life (Bacteria, Archaea, and Eukarya). This modification occurs infrequently in bacteria and, for instance, is absent in *Escherichia coli* tRNAs. On the contrary, it is common in the tRNAs of most eukaryotes and archaea, and this modification seems to be essential for some organisms. Actually, in the yeast *Saccharomyces cerevisiae*, mutants that are defective in N_1-methylation of A_{58} are nonviable.[2] Likewise, in *Thermus thermophilus*, gene disruption studies have shown that m^1A_{58} modification is required for the growth of this bacterium at high temperatures.[3]

S-Adenosyl-L-methionine (AdoMet)-dependent methyltransferases (MTases) are involved in a wide variety of biological processes involving methylation of nucleic acids, proteins, phospholipids, and small molecules using the ubiquitous methyl donor AdoMet. Recently, two essential genes of *S. cerevisiae*, GCD10 and GCD14 (renamed TRM6 and TRM61), were identified to encode the two types of subunits of the yeast m^1A_{58} MTase.[4] The recombinant enzyme behaves as an $\alpha_2\beta_2$ heterotetramer, with Trm61p being responsible for AdoMet binding and presumably for catalysis of the methyl transfer reaction,[2] whereas both types of subunits are essential for tRNA binding.[5] Indeed, the purified recombinant Trm61p subunit, which binds AdoMet, cannot bind tRNA *in vitro* in the absence of Trm6p.[2] Interestingly, Trm61p was found to be closely related to a group of prokaryotic proteins that share not only the same AdoMet-binding site but also other highly conserved motifs. Therefore, the corresponding prokaryotic proteins were presumed to act as m^1A_{58} MTases.[6] This hypothesis was demonstrated experimentally for the bacterial orthologs in *T. thermophilus* and *Mycobacterium tuberculosis*, and for the archaeal ortholog in *Pyrococcus abyssi*.[3,7–9] These proteins are α_4 homotetramers of Trm61p-like proteins, hereinafter called TrmI (m^1A_{58} tRNA methyltransferase). Interestingly, orthologs of the Trm6p protein could only be found in eukaryotes. It was shown that Trm6p, despite the absence of the characteristic MTase motifs, is structurally and evolutionary related to Trm61p, suggesting that the eukaryotic m^1A_{58} MTases evolved by gene duplication and speciation to form a heteromultimeric protein, whereas their prokaryotic orthologs remained homomultimers.[6]

Despite the low level of sequence similarity between the various families of AdoMet MTases, most of them contain a structurally highly conserved catalytic AdoMet-binding domain organized in a Rossman-like fold (for a review, see Schubert *et al.*[10]). Only one crystal structure of an TrmI has been reported to date—that of the *M. tuberculosis* Rv2118c protein (*myco*TrmI), in complex with AdoMet.[7] It shows that the catalytic C-terminal domain is very similar to that of other AdoMet-dependent MTases, whereas the N-terminal domain, which is mainly composed of β-sheets, is not found in other MTases of known structure.

In the present work, we have first solved the crystal structure of *T. thermophilus* TrmI (*th*TrmI) at 1.7 Å resolution, in complex with *S*-adenosyl-L-homocysteine (AdoHcy), the product of methyl transfer. We then focused our attention on the tRNA-binding properties of *th*TrmI using several biophysical techniques, including electrospray ionization mass spectrometry (ESI-MS), site-directed mutagenesis, steady-state kinetic assays, fluorescence spectroscopy, and molecular docking. ESI-MS has demonstrated its particular suitability for the investigation of noncovalent complexes.[11–13] More particularly, the so-called noncovalent ESI-MS has been extensively used to characterize supramolecular assemblies involving oligonucleotides in interaction with drugs and proteins (for reviews, see Veenstra,[14] Beck *et al.*,[15] Hofstadler and Griffey,[16] and Rusconi *et al.*[17]). As protein–RNA systems are partly driven by electrostatics-based interactions that are strongly enhanced in the gas phase, ESI-MS is well suited for their characterization.[18,19] Indeed, numerous studies of protein–RNA complexes have revealed a strong agreement between mass spectrometric gas-phase measurements and results obtained by solution-phase techniques.[20–24] In our work, we used noncovalent ESI-MS: (i) to unambiguously assess the oligomerization state of *th*TrmI; (ii) to monitor the *th*TrmI oligomerization state upon tRNA binding; and (iii) to determine the stoichiometry of the *th*TrmI–tRNA complexes. In addition, a mutagenesis study was performed to identify residues that are potentially crucial for the methylation reaction. We thus showed that the universally conserved D170 residue, together with two conserved aromatic residues Y78 and Y194 that line the catalytic pocket, is a key residue for the enzymatic catalysis. Actually, D170A and Y78A variants are severely altered in their catalytic efficiency. Finally, we have used molecular docking to position a N_9-methyladenine in the active site of TrmI to further investigate the possible role of these residues.

Results and Discussion

Structure determination of *th*TrmI

Production and purification of *th*TrmI have been described by Droogmans *et al.*[3] However, the recombinant His_6-tagged protein exhibited low solubility (<0.5 mg/mL), unless high salt concentrations were added (0.2 M imidazole–HCl and 0.5 M KCl). This was a serious issue for crystallization assays, in which precipitating solutions must be added to concentrated protein. In order to improve solubility, the N-terminal His_6-tag of *th*TrmI was cleaved with thrombin. Indeed, after His_6-tag removal, *th*TrmI became highly soluble (>10 mg/mL), even at moderate ionic strengths (150 mM). This type of behavior has already been reported for some His-tagged proteins.[25] We then carried out crystallization assays on the complex between *th*TrmI, AdoHcy, and *E. coli* $tRNA_i^{Met}$. The latter tRNA was selected because it is devoid of modification on A_{58} and is a substrate of the *T. thermophilus*

enzyme (see Steady-State Kinetic Assay). Although the tRNA was present in all crystallization conditions, crystals only contained the free *th*TrmI protein in complex with AdoHcy. Table 1 summarizes the data collection and refinement statistics. The structure of the *th*TrmI protein was then solved by molecular replacement with the structure of TrmI from *M. tuberculosis* TrmI [Protein Data Bank (PDB) entry code 1I9G][7] using PHASER[26] (*Z* score = 23.7). Actually, *th*TrmI shares 39% identical and 13% similar residues with *myco*TrmI. The model was then refined to 1.7 Å, to a final R_{free} factor of 20.9% (Table 1). The last four residues at the N-terminus were too poorly ordered to be included; therefore, the final model consisted of 251 residues per monomer. The asymmetric unit contained a dimer of *th*TrmI. However, the symmetry-related subunits interact extensively, leading to a tetrameric organization of *th*TrmI (Fig. 1b).

Overall structure of *th*TrmI

The *th*TrmI monomer is composed of two distinct domains connected by an extended linker containing a short α-helix (helix 2; Fig. 1a). The large C-terminal domain (residues 63–255) adopts a typical class I MTase fold, which consists of a central seven-stranded β-sheet (β_1–β_6 and β_8; Fig. 1a) flanked by α-helices on both sides (α_A–α_E; Fig. 1a).[10] The first five strands of the β-sheet (β_1–β_5; Fig. 1a) are parallel, whereas the last two (β_6–β_8) are antiparallel. The C-terminal domain contains the AdoMet-binding site and the catalytic pocket. The smaller N-terminal domain (residues 5–62) is largely a β structure composed of six β strands (β_A–β_F; Fig. 1a) and one small α-helix (α_1).

Table 1. Data collection and refinement statistics for the determination of TrmI structure

Data collection statistics	
Wavelength (Å)	0.9340
Space group	$C222_1$
Unit cell parameters (Å)	
a	90.4406
b	96.7946
c	140.5837
Number of measured reflections	499,190 (71,441)
Number of unique reflections	67,981 (9830)
Multiplicity	7.3 (7.3)
Resolution limits (Å)	48.39–1.70 (1.79–1.70)
Data completeness (%)	100.0 (100.0)
Mean $I/\sigma(I)$	13.6 (2.9)
R_{merge}	10.6 (67.6)
Refinement statistics	
Resolution limits (Å)	33.67–1.70 (1.74–1.70)
Number of reflections	64,480 (4709)
R-factor (%)	18.0 (24.7)
R_{free} (%)	20.9 (27.9)
Number of atoms	4311
Protein	3823
Ligands	67
Water molecules	421
Mean *B*-factors (Å^2)	15.92
RMSD (bounds) (Å)	0.009
RMSD (angles) (°)	1.265

Values in parenthesis are for the outer resolution shell.

The *th*TrmI tetramer is organized as a dimer of tight dimers. Two monomers interact via their extended β_6 strands, which form a cross-subunit antiparallel β-sheet (Fig. 1b, pale and dark blue). This dimer is further stabilized by ionic interactions between helices α_2 and α_3 from each subunit (Fig. 1b). A pair of β_6 strands from each dimer packs on the related β_6 strands of the other dimer to form a central antiparallel β-sheet structure. The tetramer interaction is limited to this β-sheet region and stabilized by both ionic and hydrophobic interactions, mainly a salt bridge between E220 and R224, and hydrophobic interactions between W226 and F245, and between H234 and F237 from one dimer to H242 from the other dimer. The salt bridge and the hydrophobic interactions between W226 and F245 have been also observed in the *myco*TrmI structure[7] and involve conserved residues.

In the tight dimer found in the asymmetric unit, the two cofactor-binding sites are occupied by AdoHcy molecules in identical conformations, similar to that of AdoMet in the *M. tuberculosis* TrmI structure. The binding cleft is formed by residues belonging to conserved motifs, named I, II, III, and IV (Figs. 1c and d, and 7a).[28] The adenine ring of the AdoHcy moiety is bound by motifs II, III, and IV via water-mediated hydrogen bonds between the exocyclic amino group and residues E155 and E173, and between the N_7 position of the adenine and the backbone amide of E173 (Fig. 1c). Adenine binding is also reinforced by hydrophobic contacts between its aromatic ring and residues L171, V177, K153, L154, Y124, A126, and A100. The ribose moiety of AdoHcy is bound by motif II, essentially by hydrogen bonds involving the 2′ and 3′ hydroxyl groups and residues E125 and H130. The amino-acid part of the cofactor is mainly bound by motif I. Interestingly, an aspartate side chain that could be involved in the methylation reaction, the carboxylate chain of D170 from motif IV, participates in a direct polar interaction with the ammonium group of AdoHcy. The GXGXGG pattern of motif I binds to the amine and the carboxylate through a tight network of water-mediated hydrogen bonds involving the backbone (Fig. 1c).

Comparison with *myco*TrmI structure and tRNA recognition

Globally, the overall structure of *th*TrmI closely resembles that of *myco*TrmI. Superimposition of *T. thermophilus* and *M. tuberculosis* TrmI structures gives a root mean square deviation (RMSD) of 4.47 Å for C^α, with the largest variation occurring in the β strands (β_6, β_7, and β_8) that protrude out of the structure. When comparing the active sites more accurately, that of *myco*TrmI is narrower by about 1 Å than that of *th*TrmI. Moreover, the protruding R249 in *myco*TrmI, which makes a supplementary step at the exit of the pocket that does not exist in *th*TrmI as R249, is replaced by a shorter side-chain

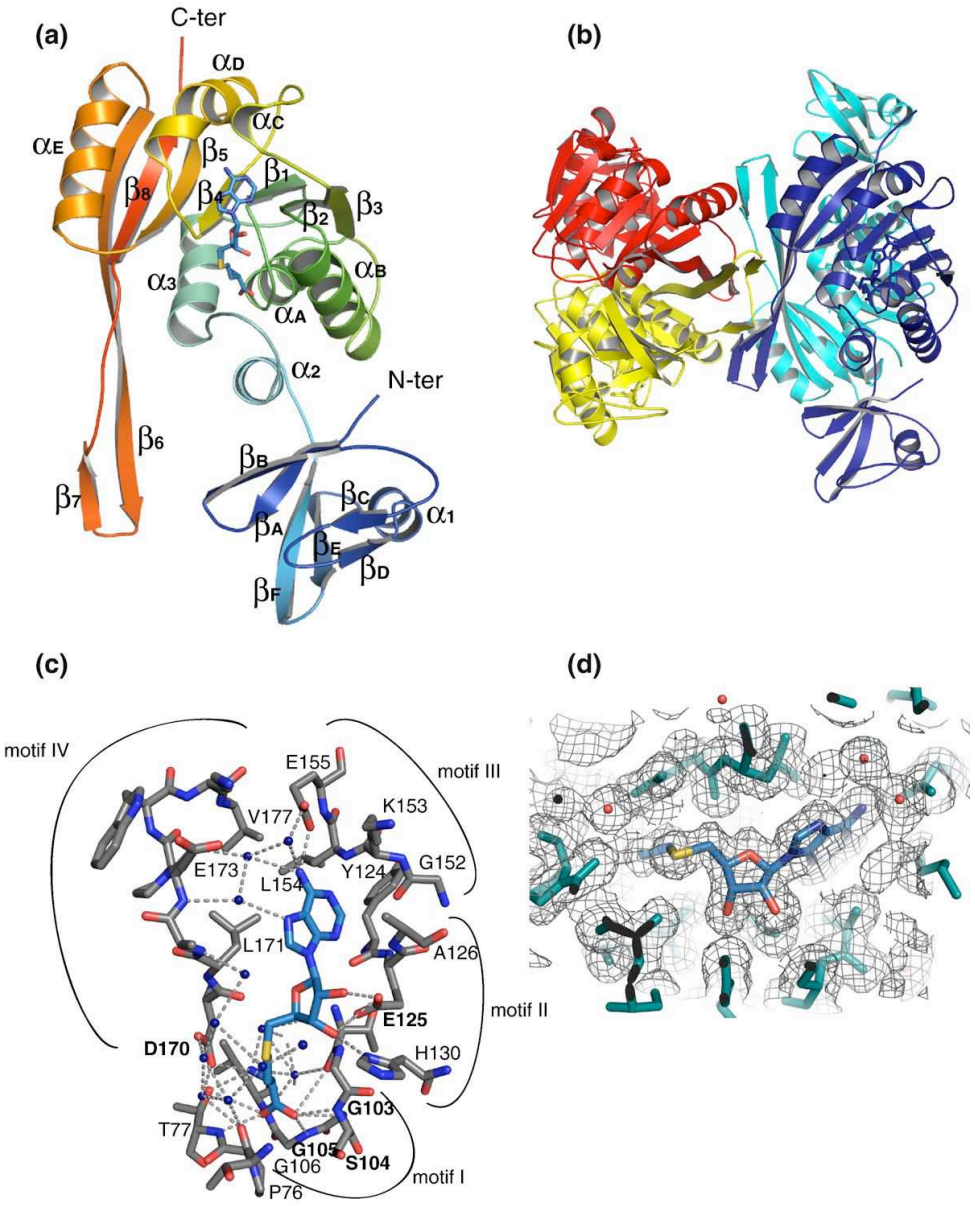


Fig. 1. Crystal structure of *th*TrmI. (a) Overall structure of *th*TrmI monomer. The secondary structure elements are labeled following the nomenclature defined in Schluckebier *et al.*[27] The cofactor AdoHcy is shown in blue sticks. (b) Organization of *th*TrmI as a tetramer. Each monomer is drawn with a different color. The two tight dimers are, respectively, in red and yellow, and in pale and dark blue. The tetramerization interface involved strands β_6 and β_7. (c) The AdoHcy-binding site. The ligand (blue) and the enzyme residues (gray) are shown as sticks. Water molecules are drawn as blue spheres. Polar contacts are represented with dashed lines. Strictly conserved residues among TrmIs are labeled with bold letters. Characteristic motifs from MTases are highlighted. (d) Density map ($2F_o - F_c$ at 1σ) around the cofactor. The AdoHcy ligand (blue) and the enzyme residues (green) are shown as sticks. Water molecules are represented as red spheres.

residue (V240). As a consequence, the accessible surface of Y78 is twice larger in *th*TrmI than that of Y84 in *myco*TrmI. All these features seem to make the methyl donor less easily accessible in *myco*TrmI. Overall, in both structures, the adenine should enter deeply in the binding cleft to have access to the

methyl donor, therefore implicating a major deformation of the phosphodiester backbone around A_{58}. It is also highly interesting to compare the electrostatic potentials of their molecular surfaces (Fig. 2). These calculations uncover that both proteins have two linked grooves of positive electrostatic potentials indicated by large blue surfaces in Fig. 2 that are large enough to accommodate double-helical RNA. The first groove that covers the β-sheet region forming the tetrameric architecture is less positively charged in *th*TrmI than in *myco*TrmI, whereas the second groove that encompasses the N-terminal domain of one TrmI subunit and that looks like an open hand is more positively charged in *th*TrmI than in *myco*TrmI. Several residues with basic side chains are conserved among the TrmI protein family, namely, K13 (R in some sequences), R15 (K), R72, R127, R217, and R229 (K), together with the histidines involved in the formation of the tetramers H234 and H242. R15 and R217 are only conserved in bacterial TrmI proteins. These grooves have the dimensions to accommodate the acceptor arm (groove 1: 25 Å large and 45 Å long) and the anticodon arm (groove 2: 25 Å large and roughly 30 Å long) without steric clashes, enabling us to position the adenosine 58 near the catalytic pocket (manual docking; data not shown). The gray line in Fig. 2a indicates a possible position of the helix axis of tRNA. Numerous clashes occur within the T-arm, which undoubtedly undergoes huge conformational changes that probably involve the entire T-arm and not only a simple flip of the adenosine 58. These changes are difficult to predict and, for this reason, we could not pursue this docking. In conclusion, the conservation of two grooves of positively charged surfaces supports the idea that TrmI binds tRNA as a tetramer

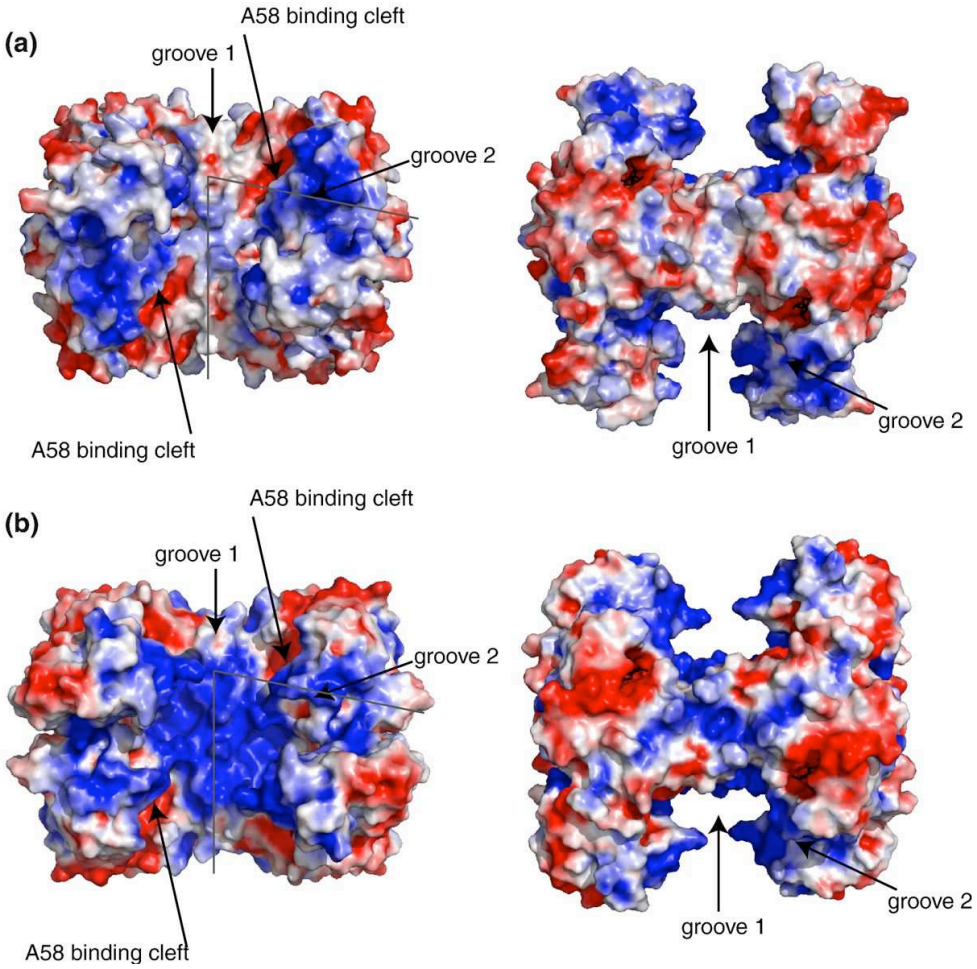


Fig. 2. Electrostatic surface potential of (a) *th*TrmI and (b) *M. tuberculosis* TrmI. Blue indicates positive charge, and red indicates negative charge, with the maximum color saturation corresponding to −3 kT (red) and +3 kT (blue). The figure was prepared using the APBS[29] PyMOL plug-in.

and that two molecules of tRNA can interact simultaneously with the TrmI tetramer. The fact that, in eukaryotes, the enzyme has evolved from an α_4 homotetramer to an $\alpha_2\beta_2$ heterotetramer is compatible with a stoichiometry of two tRNAs per TrmI tetramer, with one tRNA being expected to interact with an $\alpha\beta$ subsystem.

*th*TrmI binds up to two tRNA molecules as a tetramer

Close inspection of the crystal structures of *th*TrmI and *myco*TrmI[7] revealed that the tetramers are formed by two pairs of extensively interacting subunits stabilized by relatively small numbers of con-

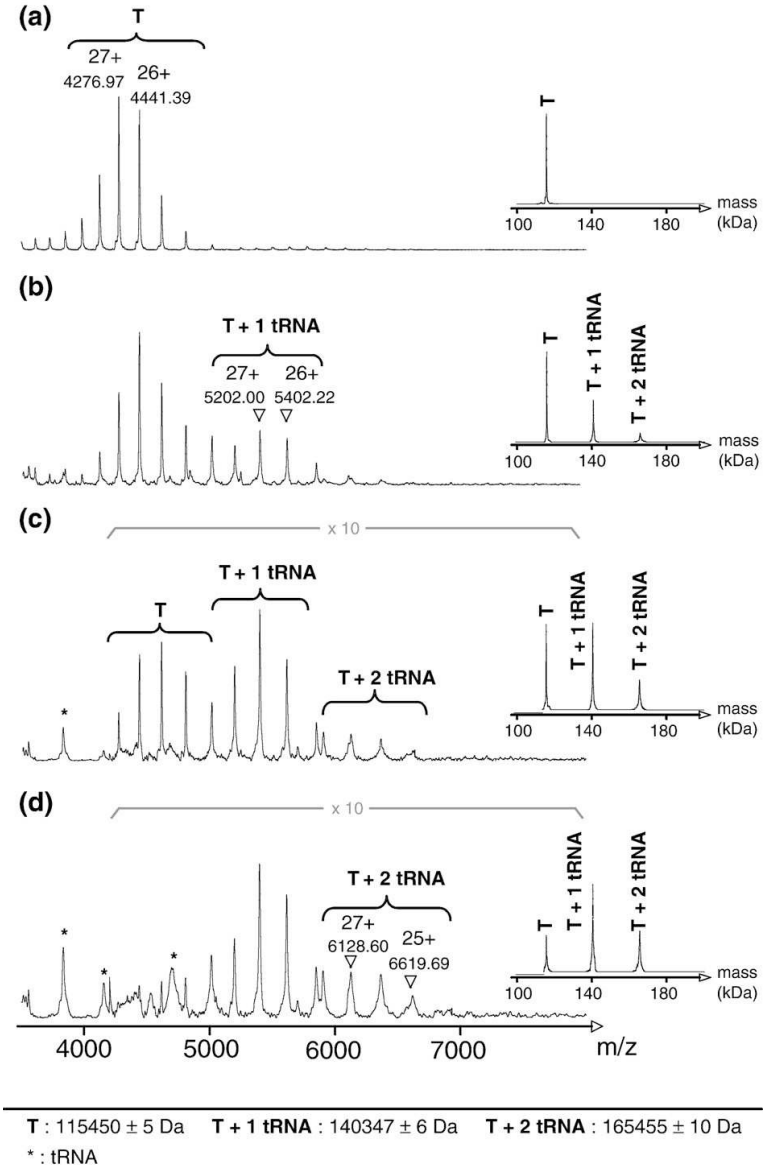


Fig. 3. Noncovalent ESI-MS analyses of *th*TrmI and *th*TrmI–tRNA assemblies. ESI mass spectra obtained under nondenaturing conditions (200 mM ammonium bicarbonate buffer, pH 8.0; $V_c = 120$ V; $P_i = 6.1$ mbar) for (a) *th*TrmI tetramer (20 μM) alone and in the presence of (b) 30 μM, (c) 100 μM, and (d) 200 μM *E. coli* $tRNA_i^{Met}$. Insets correspond to deconvoluted spectra showing molecular weights on the *x*-scale. T ($M_r = 115{,}450 \pm 5$) corresponds to $(th\text{TrmI})_4$; T + 1tRNA ($M_r = 140{,}347 \pm 6$) corresponds to the $(th\text{TrmI})_4$:$(tRNA)_1$ complex; and T + 2tRNA ($M_r = 165{,}455 \pm 10$) corresponds to the $(th\text{TrmI})_4$:$(tRNA)_2$ complex. *tRNA signals.

tacts between the two dimers. Given the architecture of TrmI, it was of interest to investigate the oligomerization state of *th*TrmI alone and upon tRNA binding, and to determine the binding stoichiometry of the *th*TrmI–tRNA complexes. Previously reported gel filtration experiments have indicated that both *myco*TrmI and *th*TrmI should form a tetramer in solution.[3,9] However, in the case of *th*TrmI, due to an inherent limited precision of this technique, it is difficult to unambiguously state whether the protein is tetrameric or pentameric.[3] In this study, we used ESI-MS, which has been proven to be a valuable technique for the determination of oligomerization states of noncovalent assemblies,[30–33] to first confirm the oligomerization state of *th*TrmI. Figure 3a presents the ESI mass spectrum obtained for *th*TrmI under nondenaturing conditions and carefully controlled operating conditions (see Materials and Methods): a single ion distribution is observed in the m/z range of 4000–5000. With a measured molecular weight (M_r) of 115,450±5, this distribution can be assigned to the +23 to +30 charge states of a *th*TrmI tetramer (theoretical M_r = 115,453). Thus, ESI-MS allowed us to unambiguously assess the tetrameric nature of the protein in agreement with our crystallographic results.

The tRNA binding stoichiometry was first investigated by gel retardation assay. As shown in Fig. 4, the free tRNA band progressively disappears upon addition of increasing amounts of *th*TrmI, thus evidencing the existence of *th*TrmI–tRNA interactions. Moreover, the free tRNA band completely disappears at 1:2 *th*TrmI:tRNA ratio, substantiating the formation of complexes involving two molar equivalents of *th*TrmI monomer per tRNA. To further confirm gel retardation experiments, we also used the potentialities of noncovalent ESI-MS to investigate the oligomerization state of *th*TrmI upon tRNA binding and to determine the tRNA binding stoichiometry. Titration experiments, monitored by noncovalent ESI-MS, involving a fixed concentration of *th*TrmI and increasing amounts of $tRNA_i^{Met}$ revealed the presence

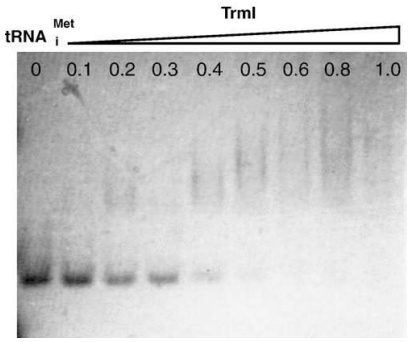


Fig. 4. Nondenaturing agarose gel electrophoresis. Constant amount of *E. coli* $tRNA_i^{Met}$ (4 μg) with increasing amounts of *th*TrmI protein (from 0 to 1 molar equivalent of *th*TrmI tetramer). Bands were visualized by UV shadowing.

of three ion distributions (Fig. 3b–d). The first one in the m/z region of 4000–5000, with a molecular mass of 115,450±5 Da, is related to the *th*TrmI tetramer. For the second ion distribution in the m/z region of 5000–5900, the molecular mass of 140,347±6 Da corresponds to the $(thTrmI)_4$:$(tRNA)_1$ complex. Finally, a third ion distribution within the m/z range of 5900–6800, with a molecular mass of 165,455±10 Da, refers to the $(thTrmI)_4$:$(tRNA)_2$ complex. Even a 10-fold molar excess of tRNA per *th*TrmI tetramer does not lead to the detection of complexes with higher tRNA binding stoichiometries. Interestingly, upon an increase in the tRNA concentration from 1.5 to 10 molar equivalents per *th*TrmI tetramer, the relative abundances of $(thTrmI)_4$:$(tRNA)_1$ and $(thTrmI)_4$:$(tRNA)_2$ complexes statistically increase, favoring noncooperative tRNA-binding system and further suggesting the presence of two independent tRNA-binding sites.[34] The same experiments were also carried out in the presence of AdoHcy (data not shown), and no effect either on the binding stoichiometry of the complex or on the binding affinity of the tRNA for *th*TrmI was observed.

As ESI-MS detects species in the gas phase of the mass spectrometer, control experiments are always needed to ensure that mass spectra faithfully reflect the behavior in solution.[35] Thus, since electrostatic interactions are sensitive to the ionic strength of the medium,[36,37] experiments were carried out at different ammonium bicarbonate concentrations (Fig. 5). Decreasing the ammonium bicarbonate concentration from 200 mM to 100 mM displaces the equilibrium towards the formation of $(thTrmI)_4$:$(tRNA)_1$ and $(TrmI)_4$:$(tRNA)_2$ complexes (Fig. 5a and b). However, the signal intensities were dramatically lower when 100 mM ammonium bicarbonate was used rather than 200 mM buffer, mainly because of protein precipitation. Therefore, no ESI-MS analysis could be performed at buffer concentrations lower than 100 mM. While both $(thTrmI)_4$:$(tRNA)_1$ and $(thTrmI)_4$:$(tRNA)_2$ complexes are favored at low ionic strengths, the use of higher salt concentrations destabilizes these assemblies (Fig. 5c and d), leading even to a complete dissociation of *th*TrmI:tRNA complexes into *th*TrmI tetramer at 1250 mM ammonium bicarbonate (Fig. 5d). The fact that mass spectra do reflect expected changes induced by modification of the solution conditions is further evidence that allows us to definitely rule out the possibility that the TrmI–tRNA complexes result from an artifact of the technique. Moreover, as described in other publications,[22,38] an additional ESI-MS control experiment was performed in strictly identical experimental and operating conditions, with a nonsubstrate RNA as negative control. The absence of any *th*TrmI–control RNA complex in ESI mass spectra even in the presence of a 5-fold molar excess of control RNA (Fig. 6) indicates that the gas-phase detection of the TrmI–tRNA complexes arises from a specific recognition in solution and not from any gas-phase artifact.

In conclusion, noncovalent ESI-MS results clearly support that *th*TrmI remains tetrameric upon tRNA

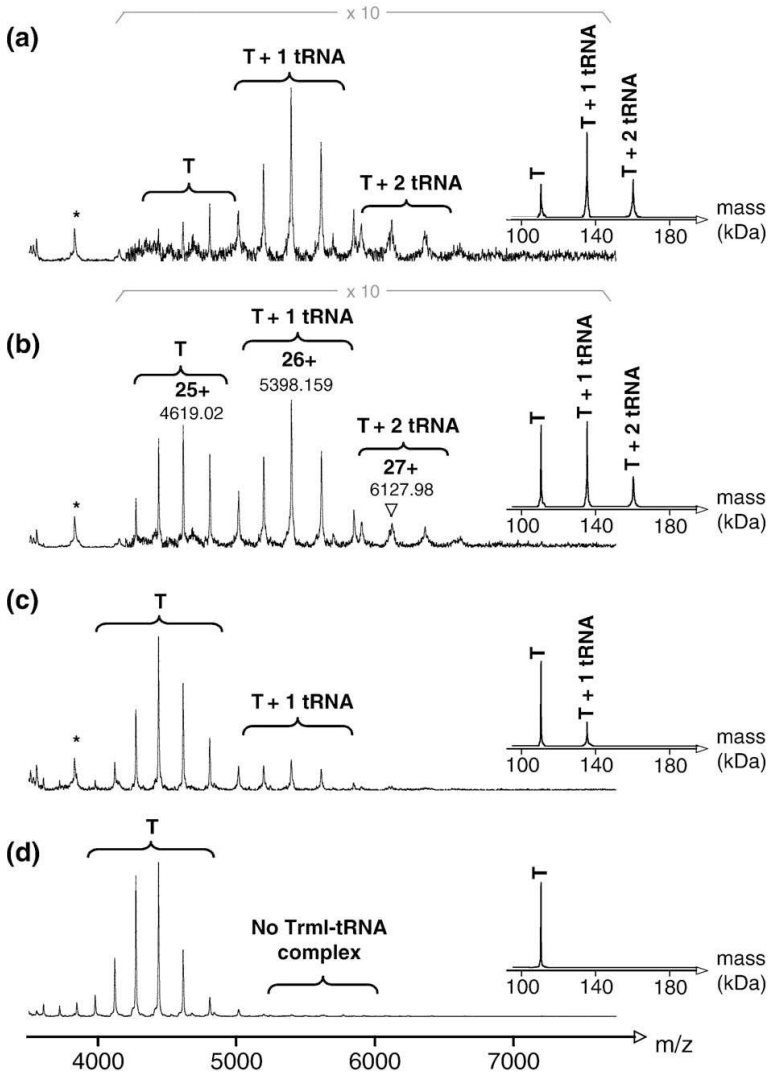


Fig. 5. Influence of the buffer ionic strength on *th*TrmI–tRNA complex stabilities. ESI mass spectra obtained under nondenaturing conditions (ammonium bicarbonate, pH 8.0; $V_c = 120$ V; $P_i = 6.1$ mbar) for *th*TrmI tetramer (20 μM) in the presence of 100 μM *E. coli* $tRNA_i^{Met}$. Ammonium bicarbonate concentration was set to (a) 100 mM, (b) 200 mM, (c) 400 mM, and (d) 1250 mM. Insets correspond to deconvoluted spectra showing molecular weights on the *x*-scale. T corresponds to $(thTrmI)_4$; T + 1tRNA corresponds to the $(thTrmI)_4$:$(tRNA)_1$ complex; and T + 2tRNA corresponds to the $(thTrmI)_4$:$(tRNA)_2$ complex. *tRNA signals.

binding and that *th*TrmI binds up to two molecules of tRNA.

The D170A and Y78A variants of *th*TrmI are severely altered in their catalytic efficiency

Based on multiple sequence alignments of the TrmI family members (Fig. 7a) and on the comparison of the crystallographic structures of TrmI from *M. tuberculosis*[7] and *T. thermophilus*, we chose to mutate three conserved residues, D170, Y194, and Y78, in order to study their involvement in adenosine binding and/or in catalysis. Figure 7b presents the *th*TrmI residues that are conserved or semiconserved in the catalytic pocket, except the conserved residues that bind to AdoMet, since their role in the catalysis mechanism is known. Yet, the universally conserved AdoMet-binding D170 was selected as it forms the back of the catalytic pocket and may also bind the adenine 58 ring of the tRNA substrate (Fig. 7b). Y194 and Y78 were mutated because they respectively form the left side and the floor of the adenosine-

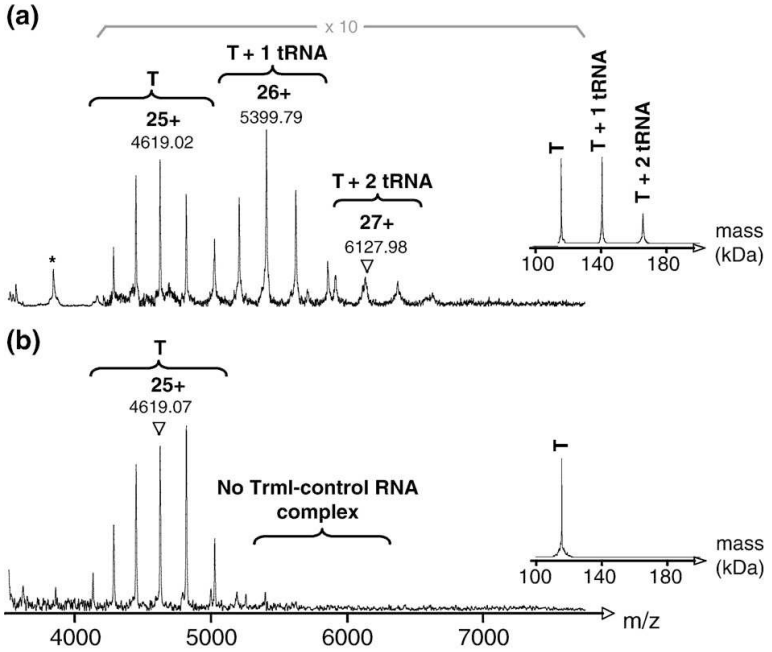


Fig. 6. Specificity of *th*TrmI–tRNA complexes formation. ESI mass spectra obtained under nondenaturing conditions (200 mM ammonium bicarbonate buffer, pH 8.0; $V_c = 120$ V; $P_i = 6.1$ mbar) for *th*TrmI tetramer (20 μM) in the presence of (a) 100 μM *E. coli* $tRNA_i^{Met}$ and (b) 100 μM nonsubstrate control RNA. Insets correspond to deconvoluted spectra showing molecular weights on the *x*-scale. T corresponds to $(thTrmI)_4$; T + 1tRNA corresponds to the $(thTrmI)_4$:$(tRNA)_1$ complex; and T + 2tRNA corresponds to the $(thTrmI)_4$:$(tRNA)_2$ complex. *RNA signal.

binding pocket in the 3D structure (Fig. 7b) and are conserved as aromatic residues among TrmIs, belonging, respectively, to motifs V and X (Fig. 7a). Therefore, these aromatic residues could be involved in the stabilization of the target adenine that needs to be flipped out of the tRNA structure to be methylated. P196, which is involved in making the shape of the catalytic cavity, was not selected for mutation because, in *M. tuberculosis*, it is naturally replaced by an alanine. Therefore, we individually mutated Y78, D170, and Y194 to alanines. We expressed and purified the variant proteins as described in Materials and Methods. We then determined the kinetic and RNA-binding parameters for wild-type and variant proteins (Table 2).

First, we investigated whether the mutations altered tRNA binding by the enzyme. Interaction with tRNA induces a quenching of the intrinsic tryptophan fluorescence of the protein by about 40%. We therefore used fluorimetric titrations to determine the apparent dissociation constant between *th*TrmI variants and $tRNA_i^{Met}$ (Table 2). All mutants showed quite unchanged dissociation constants (K_d of about 15±3 nM), thereby indicating that they retained a native folding and similar RNA-binding abilities compared to the wild-type enzyme. We then analyzed their catalytic properties by determining their kinetic parameters k_{cat} and K_M for both AdoMet and tRNA substrates. The overall results are summarized in Table 2. Interestingly, all mutants retain some catalytic activity, but are altered to various extents, essentially in their catalytic turnover k_{cat}. The D170A mutant retains some activity, but its k_{cat} value is severely reduced by a factor of about 300. D170A also shows an increased K_M for AdoMet substrate, in agreement with the crystal structure. It seems, therefore, to be the most important residue for the catalysis of the reaction. The Y78A mutant leads to a roughly 20-fold decrease in the catalytic turnover constant. Its binding constant for the methyl donor is unchanged compared to the wild-type protein. The Y194A mutant shows a smaller decrease in k_{cat}, by a factor of 3 (Table 2). The K_M value for AdoMet is increased, suggesting that this residue is involved in cofactor binding. Yet, in the crystal structure, it is not involved in a direct interaction with AdoHcy, in contrast to D170. The mutation of Y194 to an alanine probably slightly destabilizes the catalytic pocket or indirectly weakens AdoMet binding. In conclusion, D170A is the variant that is most largely altered in its catalytic efficiency, followed by Y78A and then Y194A. Therefore, D170, Y78, and Y194 are key residues in the catalytic mechanism.

The substrate adenine snugly fits into the active site pocket of *th*TrmI

To investigate the positioning of the target A58 into the catalytic pocket of *T. thermophilus* and *M. tuberculosis* TrmI, we decided to locate a N_9-

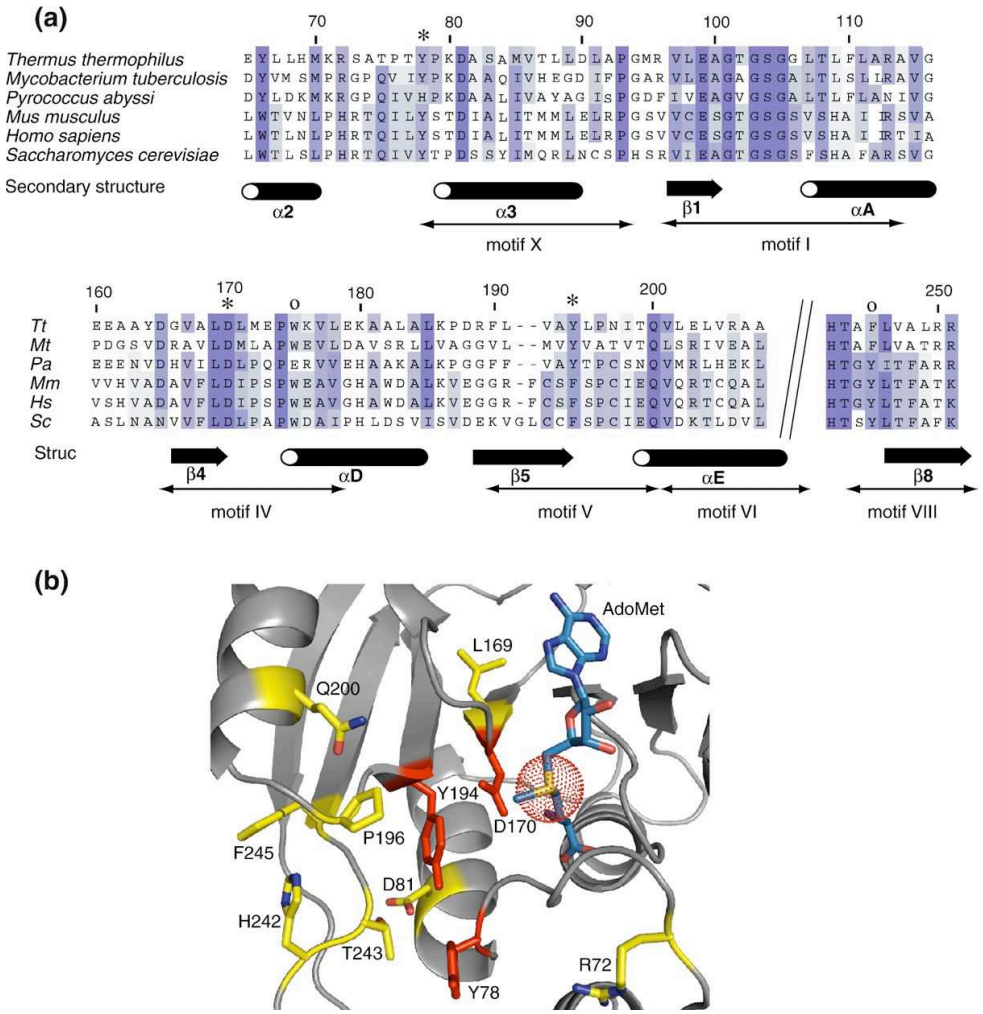


Fig. 7. Sequence alignments of TrmI proteins and conserved residues in *th*TrmI catalytic pocket. (a) Part of the sequence alignment of *T. thermophilus*, *M. tuberculosis*, *P. abyssi*, *Mus musculus*, *Homo sapiens*, and *S. cerevisiae* TrmIs. Residue numbers are those of the *T. thermophilus* sequence. Highly conserved residues are shown against a dark blue background, and residues with a similar physicochemical character are shown against a pale blue background. The secondary structure of the *th*TrmI is shown below the sequences as arrows for β strands and as cylinder for α-helices. The names of the secondary structure elements refer to Fig. 1a. The sequence motifs typical of the AdoMet-dependent MTase family are also indicated. Residues marked with a star (★) (i.e., Y78, D170, and Y194) are involved in the catalytic pocket formation and are studied by mutagenesis in the present work. Residues marked with a circle (○) (i.e., W175 and F245) are conserved aromatic residues in motifs IV and VIII, belonging to the catalytic pocket in m^6A DNA MTases. (b) Conserved and semiconserved residues around the catalytic pocket of *th*TrmI. The strictly conserved residues among the TrmI family are shown as sticks on the *T. thermophilus* structure, except the conserved residues that bind AdoMet. AdoMet was modeled from AdoHcy by adding a methyl group on the sulfur atom; the methyl is positioned by sulfonium chirality. Y78, Y194, and D170 are indicated as red sticks.

methyladenine in the catalytic pocket of these proteins by molecular docking. We docked a N_9-methyladenine and not an adenosine or an AMP molecule, since it is impossible to predict the conformation of the sugar of this flipped-out nucleotide. Moreover, Y78, which makes the exit of the catalytic pocket and which may interact with the ribose of the adenosine 58, can also flip to prevent steric clashes with the A58 ribose. As the molecules of AdoHcy present in the active sites of *th*TrmI have identical conformations to the AdoMet cofactor present in *myco*TrmI, we modeled an AdoMet molecule in the catalytic pocket of *th*TrmI by superimposing the AdoMet from *myco*TrmI with the AdoHcy in *th*TrmI.

Table 2. Kinetic and thermodynamic parameters of wild-type and mutant TrmI enzymes

Mutants	AdoMet		$tRNA_i^{Met}$			Relative k_{cat}
	k_{cat} ($s^{-1} \times 10^{-3}$)	K_M (μM)	k_{cat} ($s^{-1} \times 10^{-3}$)	K_M (μM)	K_d (nM)	
Wild type	37 (30–44)	2.1 (1.4–3.0)	36 (28–45)	2.7 (1.7–4.8)	13 (5–24)	100.0
Y194A	6 (5–11)	5.7 (3.9–8.7)	10 (9–13)	2.2 (1.9–3.3)	18 (10–37)	27.4
Y78A	1.4 (1.1–1.7)	1.9 (1.2–2.7)	1.6 (1.5–1.8)	0.16 (0.11–0.23)	16 (7–33)	4.1
D170A	0.13 (0.11–0.14)	5.4 (3.7–8.0)	0.11 (0.10–0.12)	0.17 (0.07–0.39)	14 (6–25)	0.3

The parameters are indicated with a confidence interval of 90%.

We used the program HADDOCK,[39] which can take into account mutagenesis results by introducing data as ambiguous interaction restraints (AIRs) to drive the docking process. An AIR is defined as an ambiguous distance between all residues shown to be involved in the interaction. Ambiguous intermolecular distance restraints were applied between D170, Y194, Y78, and the AdoMet cofactor, on one hand, and the N_9-methyladenine, on the other hand. A distance restraint between the N_1 atom of N_9-methyladenine and the methyl of AdoMet was also added. The model resulting from the docking with *th*TrmI is presented in Fig. 8a and shows that the N_9-methyladenine fits snugly into the cavity where methylation takes place. The same result was obtained for *myco*TrmI (data not shown). Moreover, the methyl group introduced to substitute the ribose points above Y78 (Fig. 8b), enabling us to add a ribose without steric hindrance. Therefore, Y78 is likely to strongly interact with the ribose of the adenosine 58, probably participating in the stabilization of the flipped-out conformation. The distance between the N_9-methyladenine N_1 and the AdoMet methyl group is 2.3 Å. This corresponds to a 4.1-Å distance between the N_9-methyladenine N_1 and the AdoMet sulfur atom. This is in keeping with distances observed in cocrystal structures of other endocyclic purine N-MTases (distances between 3.6 Å and 4.5 Å).[40] The docked model of the N_9-methyladenine ring suggests a binding mode involving essentially van der Waals interactions with Y194, and a polar contact between the exocyclic amino group of the purine and the carboxylate of D170 (Fig. 8b). Therefore, the side chain of D170 could act as a bidentate ligand binding the amino terminal moiety of AdoMet and the N_6 exocyclic

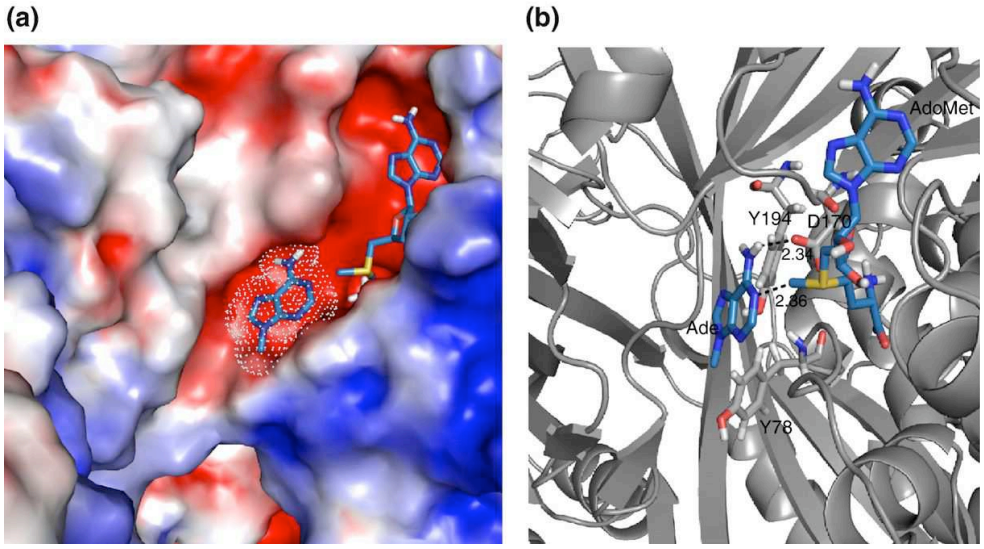


Fig. 8. Modeling a N_9-methyladenine in the catalytic pocket of *th*TrmI. (a) Average docked structure of the N_9-methyladenine ring inside the catalytic pocket of *th*TrmI. The enzyme surface is colored according to the electrostatic surface potential. Positive potential is shown in blue, and negative potential is shown in red. The AdoMet ligand and the docked N_9-methyladenine ring are shown as sticks (carbon atoms in light blue, nitrogen atoms in dark blue, oxygen atoms in red, sulfur atoms in yellow, and hydrogen atoms in white). The van der Waals surface of the N_9-methyladenine ring is shown as white dots. (b) Distances between the average structure of the docked N_9-methyladenine ring and the key moiety of the enzyme. The enzyme is shown in gray with the cartoon representation. The AdoMet ligand, the docked N_9-methyladenine ring, and residues D170, Y78, and Y194 are shown as sticks. Distances are indicated with dashed lines (in Å).

amino group of the adenine ring substrate. Thus, the sulfur–methyl bond of AdoMet is almost coplanar with the N_9-methyladenine ring and points to the same direction as the N_1 lone-pair orbital, which can be of significant importance for the chemical mechanism. We thus propose that D170 could minimize the transition state energy by positioning both the AdoMet and the target base in an optimal orientation for the methylation reaction. D170 could also deprotonate the N_6 group to activate the N_1 lone-pair orbital in order to attack the methyl group of AdoMet and to generate the exocyclic imino tautomer of m^1A as the initial product. The nucleophilic attack leads to the m^1A basic form, which has to be subsequently protonated to give the m^1A cationic form that is stable at physiological pH ($pK_a = 9.3$).[41] As the deprotonation of N_6 of m^1A is easier than the deprotonation of N_6 of the nonmethylated adenine ring ($pK_a = 16.7$),[42] it seems more reasonable that the deprotonation and the nucleophilic attack occur in the same elementary step. The model perfectly situates D170 to serve that role, which is most definitely consistent with the kinetic results.

The chemical mechanism of the adenine N_1-methylation of RNA has not yet been investigated, in contrast to the N_1-methylation of guanine.[43] In the case of m^1G RNA MTase (TrmD),[43] an aspartate residue was shown to be involved in the deprotonation step of N_1 of guanine. The two processes are likely to be different, as, under physiological conditions, the N_1 of guanine is protonated but that of adenine is not. DNA MTases have been extensively studied, but the m^1A modification is not encountered in DNA. The methylation of the N_6 position of adenine has been investigated in DNA (for structural studies, see Labahn *et al.*,[44] Tran *et al.*,[45] and Goedecke *et al.*[46]) and, more recently, in rRNA.[47,48] Interestingly, the main catalytic residue in m^6A RNA or DNA MTases has been clearly identified by mutagenesis studies to be the D/N/S residue of the conserved pattern (D/N/S)PP(Y/F/W), which belongs to MTase motif IV (for reviews on DNA MTases, see Bheemanaik *et al.*[49] and Jeltsch[50]). In known structures of m^6A DNA MTases (e.g., M•*Taq*I; PDB code 2ADM) or m^6A RNA MTases (ErmC′; PDB code 1QAN) in complex with cofactor (AdoMet or AdoHcy), the position of the D/N/S residue of motif IV relative to the cofactor is closely similar to that of D170 residue in the *th*TrmI crystal structure. Additionally, two other residues have been identified to be important in the catalysis in m^6A DNA MTases. These residues are aromatic and stabilize the flipped-out target base and/or act as cation–π catalysts. The first one is the aromatic Y/F/W residue of motif IV, which is proposed to stabilize the cationic transition state by cation–π interaction.[51,52] The second one is the aromatic residue Y/F/W of motif VIII, which stabilizes the flipped base outside the DNA helix.[51,52] These aromatic residues are conserved in motifs IV and VIII in m^1A tRNA MTases (Fig. 7a, symbol ○), but they are not part of the catalytic pocket in the TrmI structure. Strikingly, two other aromatic residues (i.e., Y78 and Y194) are located in the active site cavity of *th*TrmI (Fig. 7b) at relative positions with respect to the AdoMet cofactor, similar to those of the aromatic residues listed above for DNA MTases. These residues are also conserved as aromatic residues among TrmIs and belong, respectively, to motifs X and V (Fig. 7a). The noticeable structural similarity between the catalytic pocket of m^1A tRNA MTases and the catalytic pocket of m^6A DNA MTases suggests a similar role in catalysis for these residues. Y78 is thus the most probable candidate for the stabilization of the flipped-out adenosine given its position in the catalytic pocket and the dramatic decrease in k_{cat} of the Y78A mutant. Then, Y194 could act as a cation–π catalyst to stabilize a cationic transition state. These roles are also in agreement with our docking results.

Materials and Methods

Expression and purification of *th*TrmI for structural studies

Recombinant *th*TrmI protein (255 residues) was overexpressed and purified as previously described.[3] The N-terminal His_6-tag was removed by thrombin cleavage (25 U thrombin/mg protein) performed overnight at ambient temperature. The *th*TrmI protein was then further purified by gel filtration on a HiLoad 26/60 Superdex 200 prepgrade chromatography column (Amersham Biosciences) equilibrated in 50 mM Tris–HCl buffer (pH 8.0) and 100 mM KCl. Fractions containing pure recombinant protein were pooled, concentrated to 8–10 mg/mL using Amicon® Ultra (Millipore), and stored at 4 °C.

Crystallization and crystal structure determination

Crystallization was performed at 19 °C using the sitting-drop vapor-diffusion method. Protein samples were prepared at 3 mg/mL in 50 mM Tris–HCl buffer (pH 8.0) containing 100 mM KCl, 2 mM AdoHcy, and 1 Eq of $tRNA_i^{Met}$. Drops of 1 μL were prepared by using a Cybi-Disk robot system that mixes equal volumes of protein and reservoir solutions. Reservoir volumes of 100 μL were used. Crystals of *th*TrmI were obtained in 2 M ammonium sulfate and 5% isopropanol (vol/vol) (condition no. 5 of the Hampton Research Crystal Screen II kit). The crystals were harvested, soaked in a cryoprotectant solution [2.2 M ammonium sulfate, isopropanol 5% (vol/vol), and glycerol 20% (vol/vol)], and flash-frozen in liquid nitrogen before data collection. Diffraction data were collected at beamline ID14-1 of the European Synchrotron Radiation Facility (Grenoble, France).

All crystallographic calculations were performed using the CCP4 suite version 6[53] as implemented in the graphical user interface.[54] X-ray diffraction data were processed using MOSFLM[55] and scaled with SCALA.[56] The structure of *th*TrmI was solved by molecular replacement using the program PHASER[26] and the structure of TrmI from *M. tuberculosis* (PDB entry code 1I9G) as a model. In the model, nonconserved residues were truncated to alanine. Model building of *th*TrmI was first performed with ARP/wARP[57] using the warpNtrace automated procedure. Restrained refinements of the structure were performed with the program REFMAC5.[58] Model and map visualizations for manual reconstruction

were performed with the program Coot.[59] Solvent molecules were automatically added using the ARP waters module implemented in REFMAC5. In the last stages of refinement, TLS parameters were refined[60] using one group for each domain of the protein (i.e., the large C-terminal domain and the small N-terminal domain).

We computed the electrostatic surface potential using APBS Tools—an interface for performing APBS electrostatics calculation[29] from within PyMOL†.

Expression and purification of $tRNA_i^{Met}$

The *E. coli* $tRNA_i^{Met}$ was overexpressed from plasmid pBSTtRNA$_i^{Met}$ in *E. coli* JM101TR using a protocol derived from that of Meinnel and Blanquet.[61] Briefly, after phenol extraction of RNAs from bacteria, total tRNA was separated by gel filtration on a HiLoad 26/60 Superdex 75 prepgrade chromatography column (Amersham Biosciences) equilibrated in 25 mM Tris–HCl (pH 7.0). The overexpressed $tRNA_i^{Met}$ was then separated from other tRNAs by an anion exchange step (Resource Q column) equilibrated in 25 mM Tris–HCl (pH 7.0). tRNAs were eluted using a 350-mM to 550-mM NaCl gradient in the same buffer. The fractions containing the purified $tRNA_i^{Met}$ were pooled together, dialyzed against 50 mM Tris–HCl and 100 mM KCl (pH 8.0), concentrated using Amicon® Ultra (Millipore), and stored at −20 °C.

ESI-MS

Noncovalent mass spectrometry experiments were performed on an electrospray time-of-flight mass spectrometer (LCT; Waters). Samples were continuously infused into the mass spectrometer with a syringe pump (Harvard Apparatus) at a flow rate of 6 μl/min. Mass spectra were acquired in the positive ion mode in the m/z mass range of 2000–8000 after calibration with the multiply charged ions produced by horse heart myoglobin diluted to 2 μM in a 1:1 water/acetonitrile mixture (vol/vol) acidified with 1% (vol/vol) formic acid. Deconvoluted mass spectra were obtained using the Transform algorithm of MassLynx 4.0 software. Instrumental parameters were optimized to get the best compromise between ion desolvation, ion transmission, and preservation of noncovalent complexes during the ionization and desorption processes. This optimization concerned especially the pressure in the interface (P_i) and the accelerating voltage (V_c), which were set to 6.1 mbar and 120 V, respectively.

The purity and homogeneity of *th*TrmI were checked in denaturing conditions by diluting the protein to 5 μM in a 1:1 water/acetonitrile mixture acidified with 1% (vol/vol) formic acid. A good agreement was found between the measured molecular mass (28,863.0 ± 0.3 Da) and the mass calculated from the amino acid sequence (28,863.3 Da). For experiments performed in nondenaturing conditions, different ammonium bicarbonate concentrations ranging from 100 mM to 1250 mM were tested. A concentration of 200 mM ammonium bicarbonate was chosen as the best compromise to provide good-quality ESI mass spectra without extensive dissociation of *th*TrmI:tRNA complexes (which takes place at high ionic strengths) or protein aggregation (occurring at low ionic strengths). Thus, this buffer concentration was used to study both *th*TrmI oligomerization state and its tRNA-binding properties. All experiments were carried out at 20 μM *th*TrmI (tetramer concentration), whereas *E. coli* $tRNA_i^{Met}$ concentration (molecular weight M_r = 24,888) was increased up to 200 μM for titration experiments. To study the influence of AdoHcy on the *th*TrmI–tRNA complexation equilibrium, *th*TrmI and AdoHcy, both at 20 μM (tetramer concentration), were incubated in the ammonium bicarbonate buffer in the presence or in the absence of tRNA (30 μM). Finally, a nonsubstrate control RNA of 59 nucleotides (M_r = 19,255) was also used in a 5-fold molar excess relative to the *th*TrmI tetramer concentration in order to investigate the specificity of the *th*TrmI–tRNA interactions. All solutions were incubated for 15 min at room temperature before analysis under nondenaturing conditions.

Nondenaturing agarose gel electrophoresis

Samples were prepared by mixing a constant amount of *E. coli* $tRNA_i^{Met}$ (4 μg) with increasing amounts of *th*TrmI protein (from 0 to 1 molar equivalent of *th*TrmI tetramer) in loading buffer [Tris–HCl (pH 8.0), 50 mM, KCl 100 mM, and glycerol 18% (vol/vol)]. Samples were kept at 4 °C for 12 h and loaded on a 2% (wt/vol) agarose gel. Migration was performed for 1 h in Tris–acetate buffer (pH 8.0) at 4 °C under an electric field of 100 V. Bands were visualized by UV shadowing.

Site-directed mutagenesis

Alanine site-directed mutagenesis was carried out using the Quickchange® kit and protocol (Stratagene). All oligonucleotides for mutagenesis were designed to have appropriate pairing stability (fusion temperature near 80 °C) and were between 27 nt and 32 nt long. All mutant genes were fully sequenced and contained only the desired substitutions.

Expression and purification of wild-type and mutant *th*TrmI proteins for kinetic studies

The wild-type and mutant proteins were overexpressed and purified as for structural studies, except that the thrombin cleavage was omitted. Protein concentrations were evaluated with the Bradford method. The proteins were then diluted in 50 mM Tris–HCl (pH 8.0) to a final concentration of 0.25 mg/mL and stored at 4 °C.

Fluorescence titrations of *th*TrmI by $tRNA_i^{Met}$

Fluorescence measurements were performed at 20 °C with a JASCO spectrofluorimeter. Excitation and emission wavelengths were 280 nm and 343 nm, respectively. The excitation and emission bandwidths were 5 nm and 10 nm, respectively.

Fluorescence titration experiments were performed by adding increasing concentrations of nucleic acid to a fixed amount of *th*TrmI protein (15 nM) in 50 mM Tris–HCl, 200 mM KCl, and 0.5% (wt/vol) polyethylene glycol 8000 (pH 8.0). Fluorescence intensities were corrected for dilution and were fitted using Eq. (1), assuming two equivalent tRNA-binding sites per *th*TrmI tetramer (see Fig. 3):

$$I = I_0 - \frac{I_0 - I_\infty}{2nN_t}\left(K_d + L_t + nN_t - \sqrt{(K_d + L_t + nN_t)^2 - 4L_t nN_t}\right) \quad (1)$$

† http://www.pymol.org

where I_0 is the fluorescence intensity without tRNA; I is the fluorescence intensity at a given concentration of tRNA; I_8 is the fluorescence intensity at the plateau; n is the number of tRNA-binding sites on the protein; N_t is the total concentration of protein; and L_t is the total concentration of tRNA. Confidence limits on the parameters were estimated by Monte Carlo sampling.[62]

Steady-state kinetic assay

[Methyl-^{3}H]AdoMet (555 GBq/mmol, 20.5 MBq/mL; MP Biomedicals) was mixed with nonradioactive AdoMet (Sigma) to achieve a specific radioactivity of about 500 cpm/pmol (16.7 Bq/pmol). The methylation kinetic assays were performed in 50 mM Tris–HCl (pH 8.0) at 60 °C. The tRNA and [methyl-^{3}H]AdoMet were thermally equilibrated in 1.25× concentrated reaction buffer (96 μL) at 60 °C for 5 min, and the reaction was started by adding the enzyme (24 μL). Aliquots (25 μL) were removed after different incubation times and transferred into 800 μL of 5% (wt/vol) trichloroacetic acid at 0 °C for 30 min in order to quench the reaction and to precipitate the tRNA. The precipitates were collected by filtration through GF/C filters (Whatman). The filters were washed with cold 5% trichloroacetic acid and dried, and the radioactivity was measured by liquid scintillation counting for 2 min, resulting in a counting error below 4%. Data were corrected by subtracting the background radioactivity determined from a control without enzyme.

For the determination of AdoMet K_M and k_{cat}, the reaction mixtures contained MTase (wild type, 25 nM; Y194A, 50 nM; Y78A, 100 nM; D170A, 500 nM), $tRNA_i^{Met}$ (20 μM), and [^{3}H]AdoMet (wild type, 0.5–8 μM; Y194A, 0.5–8 μM; Y78A, 0.5–8 μM; D170A, 5–50 μM). For $tRNA_i^{Met}$ K_M and k_{cat}, the reaction mixtures contained MTase (wild type, 25 nM; Y194A, 50 nM; Y78A, 100 nM; D170A, 500 nM), $tRNA_i^{Met}$ (wild type, 0.5–8 μM; Y194A, 0.5–8 μM; Y78A, 0.25–4 μM; D170A, 0.5–8 μM), and [^{3}H]AdoMet (wild type, 20 μM; Y194A, 50 μM; Y78A, 20 μM; D170A, 50 μM). Initial rates (v_i) for each substrate concentration ($tRNA_i^{Met}$) were determined from the slope of linear fits of time-course data points (wild type, 9 min; Y194A, 20 min; Y78A, 40 min; D170A, 40 min). Enzyme parameters were obtained by nonlinear least squares fitting using Eq. (2) of Michaelis–Menten kinetics (wild-type and mutant Y194A) or Eq. (3) in cases where enzyme concentration was not negligible compared to substrate concentration (mutants Y78A and D170A). k_{cat} values were calculated for two catalytic sites per TrmI tetramer. Confidence limits on the parameters were estimated by Monte Carlo sampling using the MC-Fit program:[62]

$$v_i = \frac{V_{max} S_0}{K_M + S_0} \quad (2)$$

$$v_i = \frac{V_{max}}{2E_0}\left(K_M + S_0 + E_0 - \sqrt{(K_M + S_0 + E_0)^2 - 4S_0E_0}\right) \quad (3)$$

Molecular docking

Docking of the N_9-methyladenine ring into the active site of TrmI was performed using HADDOCK 1.3‡,[39] which makes use of CNS[63] as structure calculation software. HADDOCK allows dealing with ambiguous constraints (i.e., restraints can be applied between residues and not only between atoms). The protein structure used for the docking was the tetrameric *th*TrmI crystal structure in which the AdoHcy ligand was replaced with the AdoMet methyl donor. A 2-Å distance was used to define AIRs that were applied between the adenine and a set of neighbors composed of AdoMet cofactor and *th*TrmI residues identified by mutagenesis (i.e., Y78, D170, and Y194). One unambiguous restraint set to 2 Å was applied between the atoms that react in the methyl transfer reaction (i.e., the N_1 position of the N_9-methyladenine ring and the C^ε atom of the AdoMet ligand). During the rigid-body energy minimization, 2000 structures were calculated, and the 1000 best solutions based on the intermolecular energy were used for the semiflexible refinement. The docking converged to a single solution. The same protocol was also applied to the *M. tuberculosis* TrmI.

‡ http://www.nmr.chem.uu.nl/haddock/

Accession numbers

Coordinates and structure factors for *th*TrmI protein, in complex with AdoHcy, have been deposited at the PDB with accession code 2PWY.

Acknowledgements

The authors would like to thank Franck Brachet for his skills in the setup of the pipetting robot. We acknowledge Dr. Thibaut Crépin for his kind help with synchrotron data collection at beamline ID14-1 of the European Synchrotron Radiation Facility. We thank Nathalie Ulryck for help in site-directed mutagenesis. Pierre Barraud and Cédric Atmanene were supported by a studentship from the "Ministère de la recherche." We thank the CNRS for financial support. Louis Droogmans was supported by grants from the Fonds pour la Recherche Fondamentale Collective and the Institut Interuniversitaire des Sciences Nucléaires.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.01.041

References

1. Rozenski, J., Crain, P. F. & McCloskey, J. A. (1999). The RNA Modification Database: 1999 update. *Nucleic Acids Res.* **27**, 196–197.
2. Anderson, J., Phan, L. & Hinnebusch, A. G. (2000). The Gcd10p/Gcd14p complex is the essential two-subunit tRNA(1-methyladenosine) methyltransferase of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **97**, 5173–5178.
3. Droogmans, L., Roovers, M., Bujnicki, J. M., Tricot, C., Hartsch, T., Stalon, V. & Grosjean, H. (2003). Cloning

and characterization of tRNA (m1A58) methyltransferase (TrmI) from *Thermus thermophilus* HB27, a protein required for cell growth at extreme temperatures. *Nucleic Acids Res.* **31**, 2148–2156.
4. Anderson, J., Phan, L., Cuesta, R., Carlson, B. A., Pak, M., Asano, K. *et al.* (1998). The essential Gcd10p–Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes Dev.* **12**, 3650–3662.
5. Ozanick, S. G., Bujnicki, J. M., Sem, D. S. & Anderson, J. T. (2007). Conserved amino acids in each subunit of the heteroligomeric tRNA m1A58 MTase from *Saccharomyces cerevisiae* contribute to tRNA binding. *Nucleic Acids Res.* **35**, 6808–6819.
6. Bujnicki, J. M. (2001). In silico analysis of the tRNA: m1A58 methyltransferase family: homology-based fold prediction and identification of new members from Eubacteria and Archaea. *FEBS Lett.* **507**, 123–127.
7. Gupta, A., Kumar, P. H., Dineshkumar, T. K., Varshney, U. & Subramanya, H. S. (2001). Crystal structure of Rv2118c: an AdoMet-dependent methyltransferase from *Mycobacterium tuberculosis* H37Rv. *J. Mol. Biol.* **312**, 381–391.
8. Roovers, M., Wouters, J., Bujnicki, J. M., Tricot, C., Stalon, V., Grosjean, H. & Droogmans, L. (2004). A primordial RNA modification enzyme: the case of tRNA (m1A) methyltransferase. *Nucleic Acids Res.* **32**, 465–476.
9. Varshney, U., Ramesh, V., Madabushi, A., Gaur, R., Subramanya, H. S. & RajBhandary, U. L. (2004). *Mycobacterium tuberculosis* Rv2118c codes for a single-component homotetrameric m^1A_{58} tRNA methyltransferase. *Nucleic Acids Res.* **32**, 1018–1027.
10. Schubert, H. L., Blumenthal, R. M. & Cheng, X. (2003). Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* **28**, 329–335.
11. Loo, J. A. (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16**, 1–23.
12. Loo, J. A. (2000). Electrospray ionization mass spectrometry: a technology for studying noncovalent macromolecular complexes. *Int. J. Mass Spectrom.* **200**, 175–186.
13. Heck, A. J. & Van Den Heuvel, R. H. (2004). Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* **23**, 368–389.
14. Veenstra, T. D. (1999). Electrospray ionization mass spectrometry: a promising new technique in the study of protein/DNA noncovalent complexes. *Biochem. Biophys. Res. Commun.* **257**, 1–5.
15. Beck, J. L., Colgrave, M. L., Ralph, S. F. & Sheil, M. M. (2001). Electrospray ionization mass spectrometry of oligonucleotide complexes with drugs, metals, and proteins. *Mass Spectrom. Rev.* **20**, 61–87.
16. Hofstadler, S. A. & Griffey, R. H. (2001). Analysis of noncovalent complexes of DNA and RNA by mass spectrometry. *Chem. Rev.* **101**, 377–390.
17. Rusconi, F., Guillonneau, F. & Praseuth, D. (2002). Contributions of mass spectrometry in the study of nucleic acid-binding proteins and of nucleic acid–protein interactions. *Mass Spectrom. Rev.* **21**, 305–348.
18. Robinson, C. V., Chung, E. W., Kragelund, B. B., Knudsen, J., Aplin, R. T., Poulsen, F. M. & Dobson, C. M. (1996). Probing the nature of noncovalent interactions by mass spectrometry. A study of protein–CoA ligand binding and assembly. *J. Am. Chem. Soc.* **118**, 8646–8653.
19. Rogniaux, H., Van Dorsselaer, A., Barth, P., Biellmann, J. F., Barbanton, J., van Zandt, M. *et al.* (1999). Binding of aldose reductase inhibitors: correlation of crystallographic and mass spectrometric studies. *J. Am. Soc. Mass Spectrom.* **10**, 635–647.
20. Mei, H. Y., Mack, D. P., Galan, A. A., Halim, N. S., Heldsinger, A., Loo, J. A. *et al.* (1997). Discovery of selective, small-molecule inhibitors of RNA complexes: I. The Tat protein/TAR RNA complexes required for HIV-1 transcription. *Bioorg. Med. Chem.* **5**, 1173–1184.
21. Liu, C., Tolic, L. P., Hofstadler, S. A., Harms, A. C., Smith, R. D., Kang, C. & Sinha, N. (1998). Probing RegA/RNA interactions using electrospray ionization-Fourier transform ion cyclotron resonance-mass spectrometry. *Anal. Biochem.* **262**, 67–76.
22. Potier, N., Donald, L. J., Chernushevich, I., Ayed, A., Ens, W., Arrowsmith, C. H. *et al.* (1998). Study of a noncovalent trp repressor: DNA operator complex by electrospray ionization time-of-flight mass spectrometry. *Protein Sci.* **7**, 1388–1395.
23. Hagan, N. & Fabris, D. (2003). Direct mass spectrometric determination of the stoichiometry and binding affinity of the complexes between nucleocapsid protein and RNA stem–loop hairpins of the HIV-1 Psi-recognition element. *Biochemistry*, **42**, 10736–10745.
24. Kilic, T., Sanglier, S., Van Dorsselaer, A. & Suck, D. (2006). Oligomerization behavior of the archaeal Sm2-type protein from *Archaeoglobus fulgidus*. *Protein Sci.* **15**, 2310–2317.
25. Hamilton, S., Odili, J., Pacifico, M. D., Wilson, G. D. & Kupsch, J. M. (2003). Effect of imidazole on the solubility of a His-tagged antibody fragment. *Hybrid. Hybridomics*, **22**, 347–355.
26. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr. Sect. D*, **61**, 458–464.
27. Schluckebier, G., O'Gara, M., Saenger, W. & Cheng, X. (1995). Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247**, 16–20.
28. Fauman, E. B., Blumenthal, R. M. & Cheng, X. (1999). Structure and evolution of AdoMet-dependent methyltransferases. Cheng, X. & Blumenthal, R. M. pp. 1–38, World Scientific Publishing, Singapore.
29. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
30. Rostom, A. A. & Robinson, C. V. (1999). Detection of the intact GroEL chaperonin assembly by mass spectrometry. *J. Am. Chem. Soc.* **121**, 4718–4719.
31. van Berkel, W. J., van den Heuvel, R. H., Versluis, C. & Heck, A. J. (2000). Detection of intact megadalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry. *Protein Sci.* **9**, 435–439.
32. Sanglier, S., Ramstrom, H., Haiech, J., Leize, E. & Van Dorsselaer, A. (2002). Electrospray ionization mass spectrometry analysis revealed a similar to 310 kDa noncovalent hexamer of HPr kinase/phosphatase from *Bacillus subtilis*. *Int. J. Mass Spectrom.* **219**, 681–696.
33. Sanglier, S., Leize, E., Van Dorsselaer, A. & Zal, F. (2003). Comparative ESI-MS study of approximately 2.2 MDa native hemocyanins from deep-sea and shore crabs: from protein oligomeric state to biotope. *J. Am. Soc. Mass Spectrom.* **14**, 419–429.
34. Rogniaux, H., Sanglier, S., Strupat, K., Azza, S., Roitel, O., Ball, V. *et al.* (2001). Mass spectrometry as a novel approach to probe cooperativity in multimeric enzymatic systems. *Anal. Biochem.* **291**, 48–61.

35. Smith, R. D. & Light-Wahl, K. J. (1993). The observation of non-covalent interactions in solution by electrospray ionization mass spectrometry: promise, pitfalls and prognosis. *Biol. Mass Spectrom.* **22**, 493–501.
36. Kapur, A., Beck, J. L., Brown, S. E., Dixon, N. E. & Sheil, M. M. (2002). Use of electrospray ionization mass spectrometry to study binding interactions between a replication terminator protein and DNA. *Protein Sci.* **11**, 147–157.
37. Gupta, R., Hamdan, S. M., Dixon, N. E., Sheil, M. M. & Beck, J. L. (2004). Application of electrospray ionization mass spectrometry to study the hydrophobic interaction between the epsilon and theta subunits of DNA polymerase III. *Protein Sci.* **13**, 2878–2887.
38. Cassiday, L. A., Lebruska, L. L., Benson, L. M., Naylor, S., Owen, W. G. & Maher, L. J., III (2002). Binding stoichiometry of an RNA aptamer and its transcription factor target. *Anal. Biochem.* **306**, 290–297.
39. Dominguez, C., Boelens, R. & Bonvin, A. M. (2003). HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737.
40. McCarthy, A. A. & McCarthy, J. G. (2007). The structure of two *N*-methyltransferases from the caffeine biosynthetic pathway. *Plant Physiol.* **144**, 879–889.
41. Kettani, A., Gueron, M. & Leroy, J. L. (1997). Amino proton exchange processes in mononucleosides. *J. Am. Chem. Soc.* **119**, 1108–1115.
42. Stewart, R. & Harris, M. G. (1977). Amino group acidity in nucleotide bases. *Can. J. Chem.* **55**, 3807–3814.
43. Elkins, P. A., Watts, J. M., Zalacain, M., van Thiel, A., Vitazka, P. R., Redlak, M. *et al.* (2003). Insights into catalysis by a knotted TrmD tRNA methyltransferase. *J. Mol. Biol.* **333**, 931–949.
44. Labahn, J., Granzin, J., Schluckebier, G., Robinson, D. P., Jack, W. E., Schildkraut, I. & Saenger, W. (1994). Three-dimensional structure of the adenine-specific DNA methyltransferase M•*Taq* I in complex with the cofactor *S*-adenosylmethionine. *Proc. Natl Acad. Sci. USA*, **91**, 10957–10961.
45. Tran, P. H., Korszun, Z. R., Cerritelli, S., Springhorn, S. S. & Lacks, S. A. (1998). Crystal structure of the DpnM DNA adenine methyltransferase from the DpnII restriction system of *Streptococcus pneumoniae* bound to *S*-adenosylmethionine. *Structure*, **6**, 1563–1575.
46. Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J. & Weinhold, E. (2001). Structure of the N_6-adenine DNA methyltransferase M•*Taq*I in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **8**, 121–125.
47. Schluckebier, G., Zhong, P., Stewart, K. D., Kavanaugh, T. J. & Abad-Zapatero, C. (1999). The 2.2 A structure of the rRNA methyltransferase ErmC′ and its complexes with cofactor and cofactor analogs: implications for the reaction mechanism. *J. Mol. Biol.* **289**, 277–291.
48. Maravic, G., Feder, M., Pongor, S., Flogel, M. & Bujnicki, J. M. (2003). Mutational analysis defines the roles of conserved amino acid residues in the predicted catalytic pocket of the rRNA:m6A methyltransferase ErmC′. *J. Mol. Biol.* **332**, 99–109.
49. Bheemanaik, S., Reddy, Y. V. & Rao, D. N. (2006). Structure, function and mechanism of exocyclic DNA methyltransferases. *Biochem. J.* **399**, 177–190.
50. Jeltsch, A. (2002). Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *ChemBioChem*, **3**, 274–293.
51. Pues, H., Bleimling, N., Holz, B., Wolcke, J. & Weinhold, E. (1999). Functional roles of the conserved aromatic amino acid residues at position 108 (motif IV) and position 196 (motif VIII) in base flipping and catalysis by the N_6-adenine DNA methyltransferase from *Thermus aquaticus*. *Biochemistry*, **38**, 1426–1434.
52. Roth, M., Helm-Kruse, S., Friedrich, T. & Jeltsch, A. (1998). Functional roles of conserved amino acid residues in DNA methyltransferases investigated by site-directed mutagenesis of the EcoRV adenine-N_6-methyltransferase. *J. Biol. Chem.* **273**, 17333–17342.
53. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D*, **50**, 760–763.
54. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr. Sect. D*, **59**, 1131–1137.
55. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. In Joint CCP4+ ESF-AMCB Newsletter on Protien Crystallography (Science and Engineering Research Council, ed), Daresbury Laboratory, Warrington, UK.
56. Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*. Science and Engineering Research Council, Daresbury Laboratory, Warrington, UK.
57. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463.
58. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D*, **53**, 240–255.
59. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.
60. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. Sect. D*, **57**, 122–133.
61. Meinnel, T. & Blanquet, S. (1995). Maturation of pre-tRNA(fMet) by *Escherichia coli* RNase P is specified by a guanosine of the 5′-flanking sequence. *J. Biol. Chem.* **270**, 15908–15914.
62. Dardel, F. (1994). MC-Fit: using Monte-Carlo methods to get accurate confidence limits on enzyme parameters. *Comput. Appl. Biosci.* **10**, 273–275.
63. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D*, **54**, 905–921.

Matériel Supplémentaire de l'Article 3

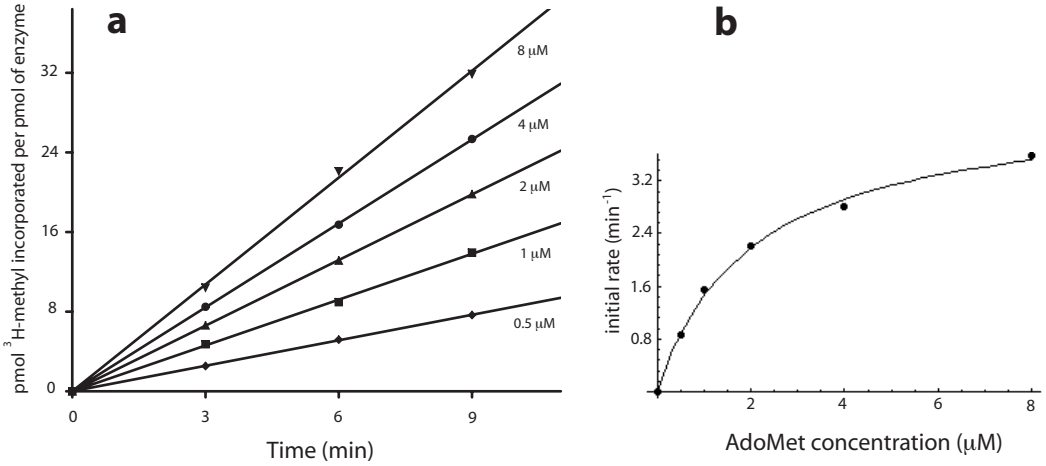


Figure 7.2 – Détermination des paramètres cinétiques k_{cat} et K_M
Exemple de l'enzyme TrmI sauvage. **a** : Cinétique d'incorporation du ^{3}H–méthyle dans l'ARNt$_f^{Met}$ pour différentes concentrations de SAM (0.5, 1, 2, 4, 8 μM). **b** : Fit des vitesses initiales d'incorporation à l'équation de Michaelis et Menten :

$$V_i = \frac{k_{cat}[E]_0[S]}{K_M + [S]}$$

7.5 Discussion concernant le mécanisme

Dans cet article, nous avons proposé des rôles individuels pour les résidus catalytiques que nous avons identifiés. Rappelons les ici pour les discuter plus en détail.

Les résidus aromatiques Y_{194} et Y_{78} pourraient globalement stabiliser le retournement du nucléotide A_{58}. Plus précisément le résidu Y_{78} pourrait interagir avec le cycle ribose du A_{58}, et le résidu Y_{194} avec le cycle aromatique. Ce dernier pourrait également participer à la catalyse en tant que catalyseur cation–π, comme proposé pour le résidu aromatique conservé du motif IV des DNA MTases [346] (Paragraphe 6.1.3), qui occupe une position équivalente, vis-à-vis du cofacteur SAM, à celle du résidu Y_{194} dans notre structure.

Le résidu D_{170}, étant donné sa position charnière, pourrait servir à orienter correctement les deux substrats SAM et A_{58}. En effet, la structure montre que celui-ci lie la fonction amino-terminale de la SAM, et le résultat de la modélisation suggère qu'il puisse également lier la position 6 amino-exocyclique de l'adénine cible. Il contribuerait ainsi à l'orientation adéquate des orbitales réactives des deux substrats — à savoir l'orbitale anti-liante S–$C^{\star}_{Me}$ de la SAM et l'orbitale non liante du N_1 du A_{58} — et ainsi abaisserait considérablement

l'enthalpie libre d'activation de la réaction ($\Delta_r G^{\ddagger}$). D'autre part, le fait que dans le modèle d'amarrage la fonction carboxylate de la chaîne latérale de l'aspartate 170 réalise une liaison hydrogène avec la fonction amino-exocyclique de l'adénine cible, suggère aussi que ce résidu puisse intervenir comme catalyseur basique. En effet, la déprotonation de la fonction amino de l'adénine, couplée au transfert de méthyle aboutirait à la formation de la forme imino tautomérique déprotonée du nucléotide m^1A (FIGURE 7.3 **A**).

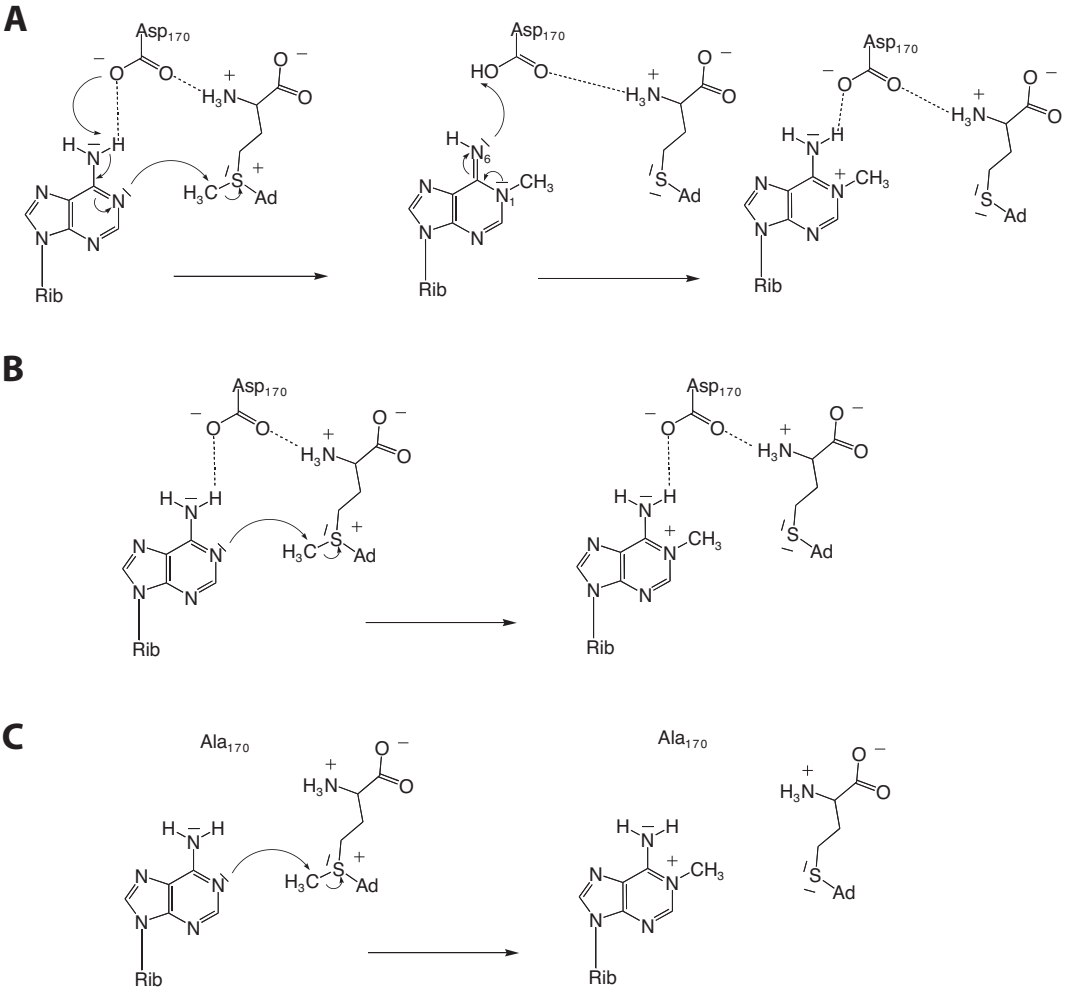


FIGURE 7.3 – Mécanismes possibles pour la formation du m^1A
A et **B** : Les deux mécanismes possibles pour la formation du m^1A dans l'enzyme sauvage. **A** : Avec déprotonation de la fonction amino exoclyclique de l'adénine et passage par l'intermédiaire imino tautomère du m^1A. **B** : Transfert direct du méthyle, sans intermédiaire. **C** : Mécanisme pour la formation du m^1A dans le variant D170A de TrmI, privé de la chaîne latérale du D170.

Cette déprotonation peut sembler délicate, étant donné la valeur des pK_a des couples en présence — à savoir $pK_a \simeq 4,6$ pour la fonction carboxylate de l'aspartate et $pK_a \simeq 16,7$ pour la fonction amino de l'adénine [420]. Si elle a lieu, celle-ci est selon toute vraisemblance couplée au transfert de méthyle. En

effet, le pK_a de la fonction amino du nucléotide m^{1}A est fortement abaissé par rapport à celui de l'adénine non méthylée (p$K_a \simeq 9,3$) [421]. Ainsi le transfert de méthyle aiderait à la déprotonation, et il semble déraisonnable d'envisager un mécanisme en deux étapes distinctes — déprotonation puis transfert de méthyle — qui verrait la formation d'un intermédiaire très peu stable. Pour les mêmes raisons, c'est-à-dire étant donné les valeurs des pK_a des couples, il semble plus raisonnable de penser que la déprotonation, si elle a lieu, arrive tard sur le chemin réactionnel, le transfert de méthyle se produisant alors plus précocement. C'est ce qui est proposé pour la méthylation en position 6 des adénines par les m^{6}A ADN MTases [422]. La formation de la forme acide du nuléotide m^{1}A, stable à pH physiologique (p$K_a \simeq 9,3$), nécessite alors la protonation de la fonction imino exocyclique. Le donneur de protons pourrait être soit le D_{170}, soit une molécule de solvant (FIGURE 7.3 **A**).

Si l'on admet que la déprotonation se produit tard sur le chemin réactionnel, il n'y a alors qu'un pas pour admettre que celle-ci ne se produit pas. On peut en effet très bien envisager que le N_1 de l'adénine soit suffisamment nucléophile pour attaquer de lui même — si la proximité et l'orientation des orbitales le permet — le groupement méthyle extrêmement réactif de la SAM (FIGURE 7.3 **B**). Rappelons ici pour mémoire, la thermodynamique extrêmement favorable de la réaction de transfert de méthyle faisant intervenir la SAM. L'enthalpie libre standard de réaction ($\Delta_r G°$) pour la formation de méthionine à partir d'homocystéine (SAM + Hcy $\longrightarrow$ SAH + Met) est par exemple évaluée à −71 kJ.mol^{-1} soit plus du double de celle de la réaction d'hydrolyse de l'ATP (ATP + $H_2O \longrightarrow$ ADP + P_i) (−30 kJ.mol^{-1}) [265, 266]. Ce mécanisme sans déprotonation (FIGURE 7.3 **B**) est donc parfaitement envisageable. Dans ce mécanisme, il n'y a pas d'intermédiaire et le transfert de méthyle donne directement la forme acide du nucléotide m^{1}A, stable à pH physiologique.

Un point intéressant vient du fait que le variant D170A de TrmI est très affecté dans ses paramètres cinétiques, mais est toujours actif (son k_{cat} est diminué d'un facteur 300 par rapport à celui de l'enzyme sauvage). Dans ce variant, le catalyseur basique n'étant plus présent, le mécanisme **A** n'est plus possible. La méthylation doit donc se produire selon un mécanisme analogue au mécanisme **B**. En l'occurrence il doit s'agir du mécanisme **C** présenté à la FIGURE 7.3 **C**. Le mécanisme **C** est donc 300 fois moins efficace qui celui ayant lieu dans l'enzyme sauvage. Mais on ne peut pas conclure quant à la nature de celui-ci (mécanisme **A** ou **B**). Ces deux mécanismes sont en effet parfaitement en accord avec nos résultats de cinétique, et la différence d'activité entre l'enzyme sauvage et le variant D170A peut signifier deux choses :

– Le mécanisme qui se produit dans l'enzyme sauvage est le mécanisme **A** et la différence d'activité des enzymes est due à un défaut de déprotonation du substrat qui empêche de passer par la forme imino tautomérique du m^{1}A.

– Le mécanisme qui se produit dans l'enzyme sauvage est le mécanisme **B** et la différence d'activité des enzymes est due à une mauvaise orientation des orbitales réactives, qui est procurée par la chaîne latérale du D_{170} dans l'enzyme sauvage.

Ainsi, même si les expériences de cinétique ne nous ont pas permis d'affirmer catégoriquement le mécanisme de la méthylation en position 1 des adénines, nous avons identifié trois résidus importants dans la catalyse — les résidus Y_{194}, Y_{78} et D_{170} — dont un résidu clé dans l'étape de transfert du groupement méthyle, le résidu D_{170}.

Chapitre 8

Perspectives

Lors de ce travail sur la m^1A_{58} MTase de *T. thermophilus*, nous avons mis en évidence deux sillons de taille et de charge adéquate pour accueillir les deux hélices d'un ARNt — l'hélice acceptrice–bras T et l'hélice anticodon–bras D — tout en positionnant la boucle T à proximité du site de fixation du cofacteur SAM (FIGURE 8.1 et FIGURE 2 de l'Article 3). En effet :

- Ces deux sillons sont chargés positivement.
- Ils se rencontrent au niveau du site de fixation du cofacteur SAM.
- Ils s'agencent approximativement à angle droit.
- Ils ont une largeur de 25 Å environ, ce qui permet d'accommoder parfaitement une hélice d'ARN.

Le premier sillon est le sillon le plus confiné. Il est situé au niveau de l'interface de tétramérisation de la protéine (au niveau des brins β_6 et β_7. La face interne du domaine N-terminal participe à la formation de celui-ci.

Le deuxième sillon est un sillon semi-ouvert, qui n'est en fait fermé que vers l'arrière par l'hélice α_B du domaine C-terminal, et vers le bas par le petit domaine N-terminal. Celui-ci constitue en quelque sorte une main ouverte sur laquelle pourrait venir se poser le substrat ARNt.

Après avoir identifié ces deux sillons, nous avons réalisé un modèle d'amarrage manuel de l'ARNt entier sur le tétramère de TrmI. Nous avons pu positionner les hélices acceptrice et anticodon sans rencontrer de gène stérique importante avec la protéine. Ce modèle est présenté à la FIGURE 8.2. Dans celui-ci, la tige acceptrice ainsi que le bras T occupe le premier sillon, celui situé à l'interface de tétramérisation. Le bras D et la tige anticodon occupent quant à eux le deuxième sillon, et reposent sur la main formée par le petit domaine N-terminal.

Ce modèle place la boucle T à proximité du site actif de TrmI. Cependant, dans ce modèle, de nombreuses gènes stériques ont lieu au niveau du bras T de l'ARNt. Celui-ci doit donc vraisemblablement subir une déformation importante lors de sa fixation à la protéine, ce qui semble de toute façon

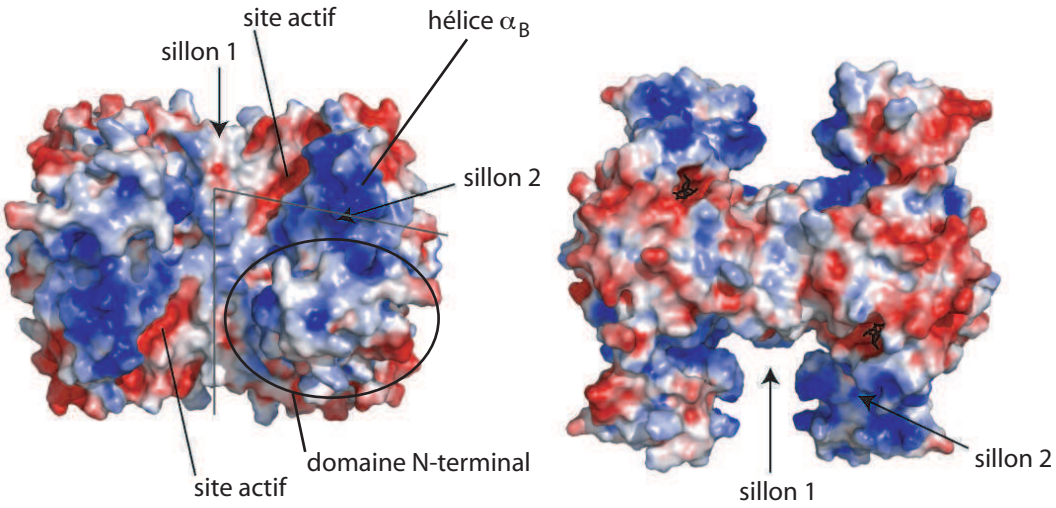


FIGURE 8.1 – Les deux sillons pour la fixation de l'ARNt
Mise en évidence des sillons grâce au potentiel électrostatique de surface. Les charges positives sont en bleu et les charges négatives en rouge, la couleur la plus intense correspondant à $-3\ kT$ (*rouge*) et $+3\ kT$ (*bleu*). La vue de gauche correspond à une « vue de face », et la vue de droite à une « vue de dessus ».

nécessaire afin de rendre l'adénosine 58 accessible et de l'amener au site actif.

Nous n'avons pas poursuivi plus avant ce modèle d'amarrage manuel car il n'est pas possible de prévoir avec certitude les déformations rencontrées au niveau du bras T.

Dans ce modèle, le domaine N-terminal participe à la fixation :

- d'une part, de la tige anticodon de l'ARNt au niveau de sa face supérieure qui forme la main ouverte du sillon 2
- et d'autre part à celle de la tige acceptrice par l'intermédiaire de sa face interne qui pointe vers le sillon 1

Nous proposons donc que ce domaine additionnel au domaine catalytique C-terminal soit un domaine de fixation à l'ARN. Ceci pourrait être vérifié en produisant une protéine TrmI tronquée au niveau de ce domaine et en mesurant ses paramètres thermodynamiques et cinétiques de liaison à l'ARNt (K_d et K_M).

D'autre part, les essais de co-cristallisation entre TrmI et divers ARN doivent être poursuivis en cherchant à garder une force ionique modérée dans la solution de cristallisation, et donc en privilégiant les agents précipitants non ioniques.

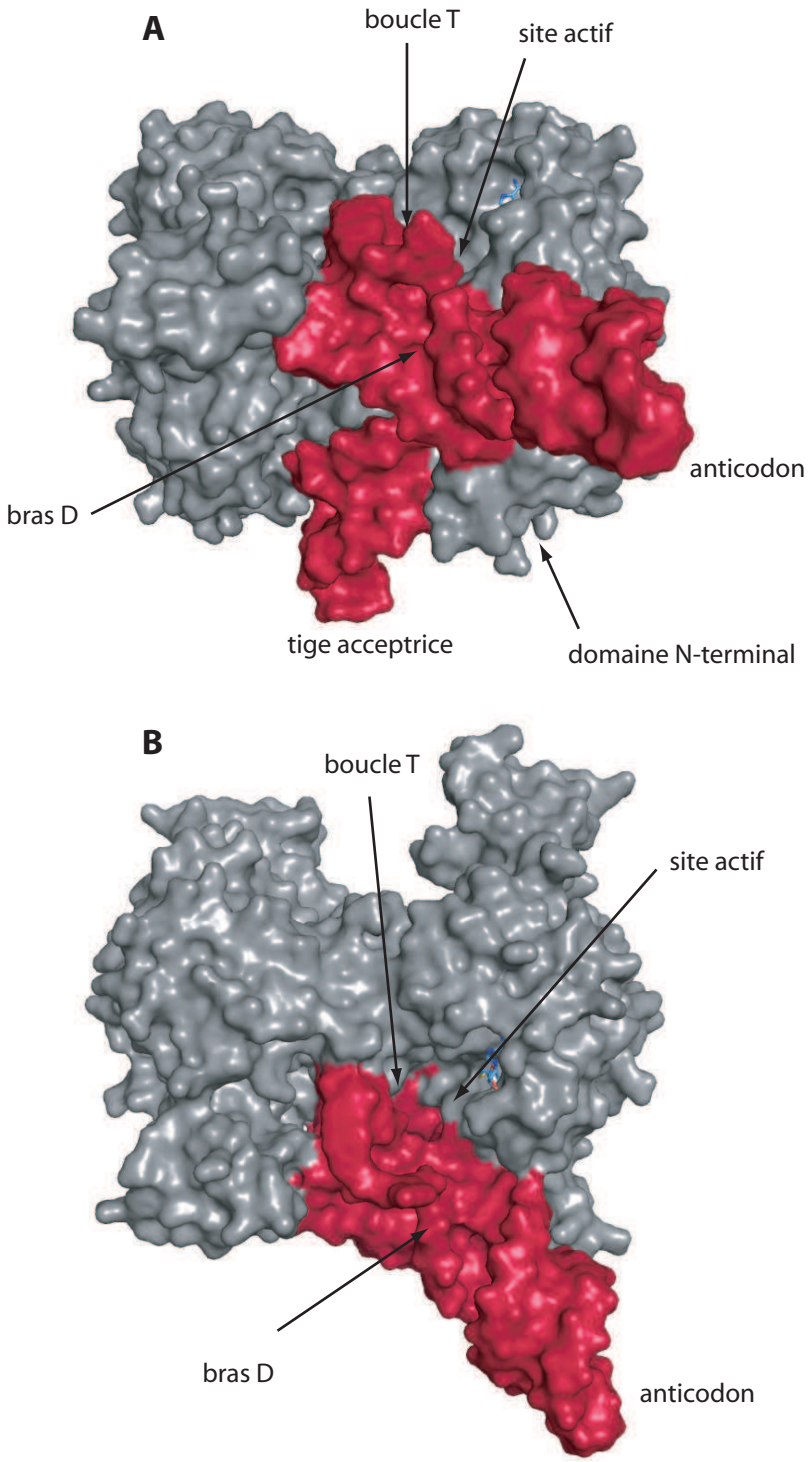


FIGURE 8.2 – Modèle d'amarrage manuel d'un ARNt entier
A et **B** : La surface de la protéine TrmI tétramérique et de l'ARNt sont représentées respectivement *en gris* et *en rouge*. **A** : Vue de face. **B** : Vue de dessus. L'ARNt correspond à l'$ARNt_f^{Met}$ initiateur d'*E. coli* (PDB code 3CW5).

Troisième partie

Structure Cristallographique de l'ARNt$_f^{Met}$ initiateur d'*E. coli*

Chapitre 9

Contexte biologique

L'initiation de la traduction est un processus complexe et hautement régulé. Chez les bactéries, seulement trois facteurs protéiques additionnels viennent assister le ribosome. Il s'agit des facteurs d'initiation IF1, IF2 et IF3. Ces dernières années, les nombreux résultats apportés par la cristallographie et la cryo-microscopie électronique sur les assemblages macromoléculaires que sont les ribosomes, ont permis des avancées spectaculaires quant à la compréhension des étapes successives de l'initiation traductionnelle (pour des articles de revues [49, 423, 424]). L'initiation de la traduction couvre toutes les étapes entre la dissociation des deux sous-unités du ribosome — correspondant à la terminaison du cycle précédent — et l'assemblage sur un codon initiateur d'ARNm d'un ribosome prêt pour l'étape d'élongation. Ces différentes étapes seront décrites succinctement dans le PARAGRAPHE suivant, pour le système de traduction bactérien.

9.1 L'initiation de la traduction chez les bactéries

Le commencement d'un nouveau cycle de traduction nécessite la dissociation préalable du complexe 70S en sous-unités séparées. Cette dissociation est déclenchée par le facteur de recyclage RRF (*Ribosome Recycling Factor*) et le facteur d'élongation EF-G. Une fois les deux sous-unités dissociées, l'assemblage prématuré 30S–50S est empêché grâce au rôle anti-associateur de IF3 secondé dans cette tâche par IF1 — les deux facteurs venant se fixer sur la sous-unité 30S (FIGURE 9.1 ①).

Après dissociation du ribosome, le fMet-ARNt$_f^{Met}$ initiateur, l'ARNm et le facteur IF2 se fixent à la sous-unité 30S dans un ordre qui reste à déterminer, et qui est peut-être aléatoire. L'ARNm est prépositionné sur le ribosome grâce à la séquence dite de Shine-Dalgarno (SD) — située en amont de la séquence à traduire — qui interagit avec une séquence qui lui est complémentaire sur l'ARNr 16S (Anti Shine-Dalgarno) [425, 426]. Le positionnement précis de l'ARNm de façon à ce que le codon de démarrage se trouve au niveau du site

P du ribosome semble nécessiter la présence des trois facteurs d'initiation, et spécialement celle de IF3 [427]. Le fMet-ARNt$_f^{Met}$ initiateur est de même positionné au niveau du site P de la sous-unité 30S selon un mode de liaison qui est tout d'abord indépendant de l'interaction codon–anticodon [428]. Cette étape nécessite le facteur IF2 qui interagit avec l'ARNt initiateur sur le ribosome. De plus, le facteur IF1, qui se fixe au niveau du site A de la sous-unité 30S du ribosome, empêche le positionnement de l'ARNt initiateur au niveau de celui-ci [429, 430]. IF3 participe également à la stabilisation du fMet-ARNt$_f^{Met}$ au niveau du site P du ribosome [431] (Figure 9.1 ②).

Le complexe de préinitiation correspond ainsi à la sous-unité 30S sur laquelle sont fixés les trois facteurs d'initiation, l'ARNm dans une position d'attente, et le fMet-ARNt$_f^{Met}$ — ancré sur la 30S de façon indépendante du codon initiateur. Ce complexe relativement instable est alors le siège d'un changement de conformation qui permet l'interaction codon–anticodon, et donne ainsi un complexe d'initiation 30S stable [431, 432]. Dans cette étape, la validité de l'appariement codon–anticodon est vérifiée selon un mécanisme dit de décodage, notamment par le facteur IF3, qui est capable de dissocier les complexes présentant un mésappariement entre le codon de l'ARNm et l'anticodon de l'ARNt$_f^{Met}$ [123] (Figure 9.1 ③).

Après le relargage des facteurs IF1 et IF3 (Figure 9.1 ④), la sous-unité 50S vient s'associer au complexe d'initiation 30S pour former le complexe d'initiation 70S. Cette association est tout d'abord stimulée par le facteur IF2, puis celui-ci est également relargué — cette étape, étant induite par l'hydrolyse du GTP porté par IF2 en GDP + Pi. Le fMet-ARNt$_f^{Met}$ initiateur est alors ajusté de façon fine dans le site P ce qui positionne sa tige acceptrice au niveau du centre peptidyl-transférase de la 50S. Le complexe d'initiation 70S est alors prêt pour la phase d'élongation de la traduction (Figure 9.1 ⑤).

Une étape cruciale dans ce cycle, correspond donc à l'étape de décodage, étape pendant laquelle l'intégrité de l'ARNt$_f^{Met}$ initiateur est vérifiée ainsi que l'appariement correct codon–anticodon. Cette étape est conditionnée par des interactions directes entre le ribosome et l'ARNt initiateur au niveau du site P et par l'intervention des facteurs d'initiation — notamment IF3. La sélection de l'ARNt initiateur ne dépend ainsi pas seulement de l'interaction codon–anticodon, mais d'autres éléments de discrimination qui confèrent à l'ARNt$_f^{Met}$ sa fonction d'ARNt initiateur. Dans la suite, nous présenterons tout d'abord les particularités qui caractérisent les ARNt initiateurs, puis nous détaillerons comment certaines d'entre elles sont directement reconnues au niveau du site P du ribosome. Enfin, nous terminerons en présentant l'étape de décodage IF3-dépendante et notamment les zones d'ombres restant à éclairer.

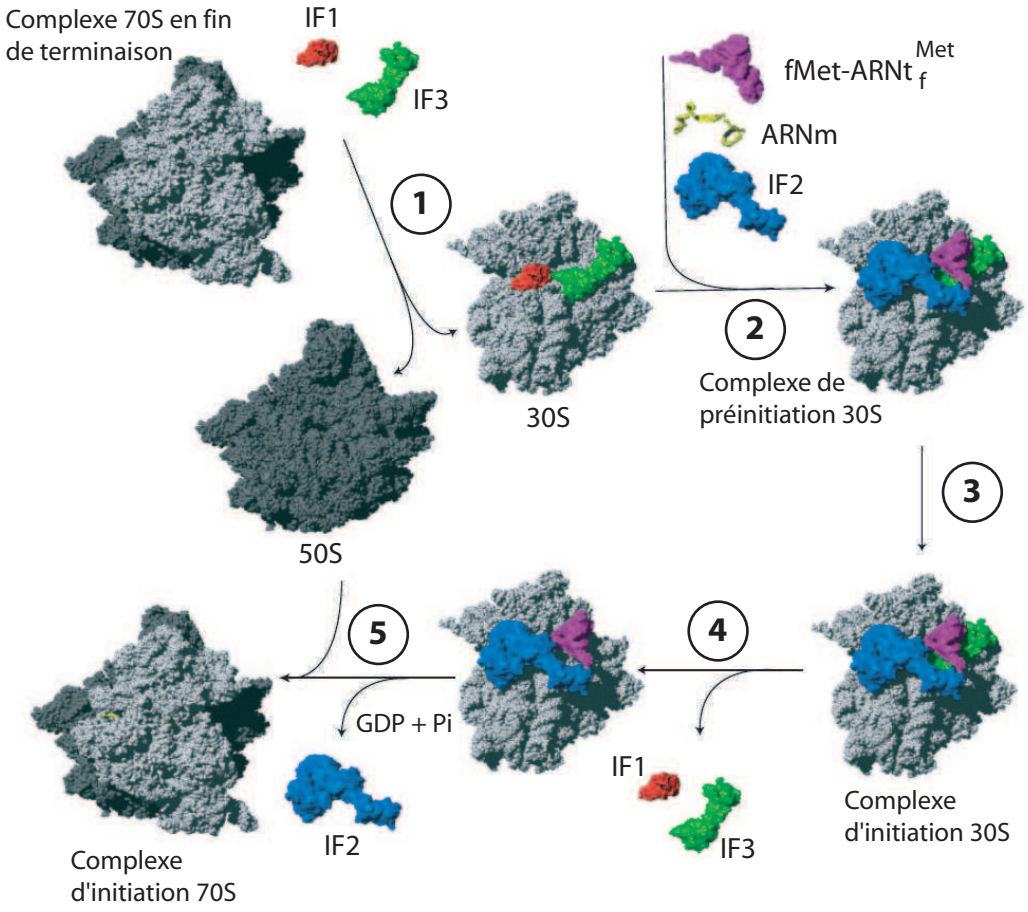


FIGURE 9.1 – L'initiation de la traduction chez les bactéries
① : Dissociation des sous-unités ribosomales par RRF/EF-G. Fixation de IF3 et IF1 sur la sous-unité 30S. Activité anti-associatrice de IF3 et IF1. ② : Fixation de l'ARNm dans une position d'attente. Fixation du fMet-ARNt$_f^{Met}$ de façon indépendante du codon. ③ : Interaction codon–anticodon. Changement conformationnel. ④ : Relargage des facteurs IF1 et IF3. ⑤ : Association de la sous-unité 50S. Relargage de IF2 et hydrolyse du GTP. Ajustement du fMet-ARNt$_f^{Met}$ au site P. D'après Laursen *et al.* [424].

9.2 Particularité des ARNt initiateurs

9.2.1 Déterminants de séquence des ARNt initiateurs de bactéries

Les ARNt initiateurs bactériens présentent quatre particularités de séquence qui les distinguent des ARNt élongateurs et qui participent ainsi à leur reconnaissance spécifique par les divers facteurs avec lesquels ils interagissent au cours de la traduction [433, 434]. Ces caractéristiques sont présentées à la FIGURE 9.2.

Ces quatre caractéristiques sont :

le groupement formyle porté par la méthionine. Ce groupement est ajouté sur le Met-ARNt$_f^{Met}$ par la méthionyl-ARNt transformylase (MTF). Ce groupement formyle est le déterminant majeur de la reconnaissance du fMet-ARNt$_f^{Met}$ par le facteur d'initiation IF2 [435]. De plus celle-ci empêche la sélection de l'ARNt initiateur par le facteur d'élongation EF1A et donc empêche son utilisation en tant qu'ARNt élongateur [436–438]. La nature de l'acide aminé formylé semble moins importante que le groupement formyle lui-même [439].

la paire de base bases 1–72. Cette paire de base est non appariée dans le cas de l'ARNt initiateur. Cette particularité est un des éléments essentiels pour sa prise en charge spécifique par la MTF [440]. La structure du complexe fMet-ARNt$_f^{Met}$/MFT apporte une réponse claire quant à cette spécificité. En effet, cette paire de base faible permet d'avoir une extrémité 3′ mono-brin suffisamment longue pour atteindre le site actif de la formylase, ce qui n'est pas possible avec un ARNt élongateur portant une paire forte en position 1–72 [38]. De la même manière, l'absence d'appariement pour les positions 1–72 est également un déterminant majeur qui empêche l'interaction du fMet-ARNt$_f^{Met}$ avec la peptidyl-ARNt hydrolase (PTH) — l'enzyme qui recycle les ARNt N-bloqués, comme les peptidyl-ARNt qui se sont décrochés du site P du ribosome en cours de synthèse et qui portent ainsi une chaîne polypeptidique à leur extrémité 3′, ou le fMet-ARNt$_f^{Met}$ dont l'acide aminé méthionine est N-formylé. Le fMet-ARNt$_f^{Met}$ échappe à l'action de la PTH, car la présence d'une paire de base parfaitement appariée en position 1–72 est cruciale pour son activité [441].

la paire purine-11–pyrimidine-24. Chez les ARNt élongateurs, on trouve en effet une paire pyrimidine-11–purine-24. Même si ce déterminant est loin d'être aussi crucial que la paire 1–72, il affecte également l'activité de la MTF [442]. Là encore, la structure du complexe fMet-ARNt$_f^{Met}$/MFT éclaire parfaitement les données biochimiques, car le nucléotide U_{24} y est spécifiquement reconnu par la MTF [38].

les trois paires G=C conservées dans la tige anticodon. Ces trois paires G=C ne sont pas seulement retrouvées chez les ARNt initiateurs de bactéries, mais sont une caractéristique commune à tous les ARNt initiateurs cytoplasmiques [12]. Ces paires G=C sont nécessaires au recrutement IF3-dépendant de l'ARNt$_f^{Met}$ au site P [122, 123, 443] et sont à l'origine d'une conformation particulière de la région de l'anticodon des ARNt initiateurs [120, 122, 444]. Nous allons, dans le Paragraphe suivant, décrire plus précisément les particularités de la région de l'anticodon des ARNt initiateurs qui sont en relation avec ces trois paires de bases G=C.

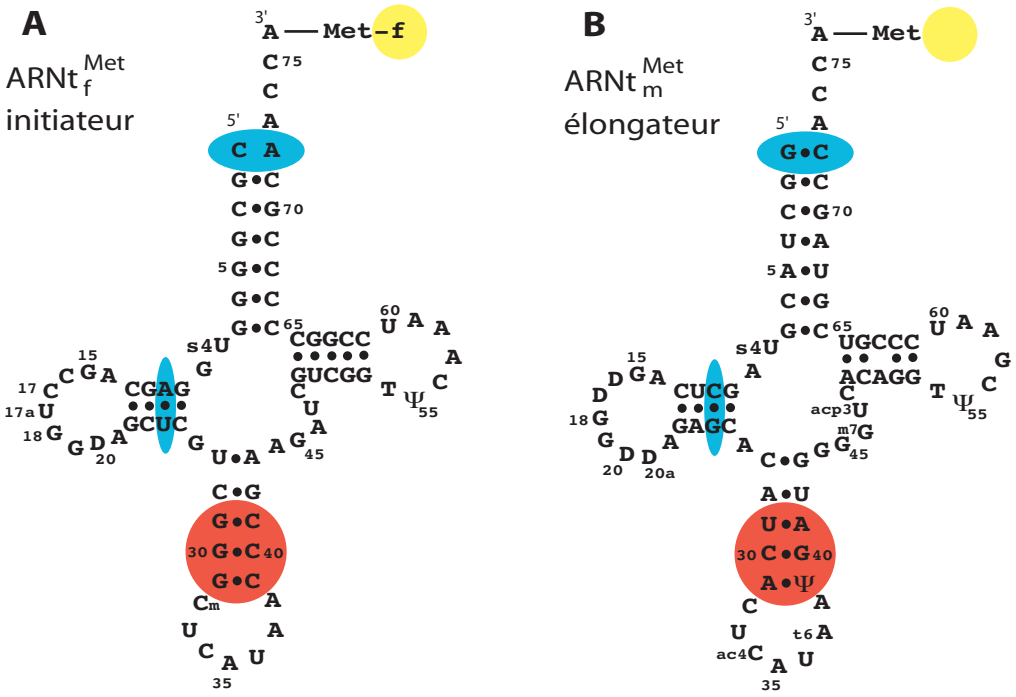


FIGURE 9.2 – Caractéristiques des ARNt initiateurs de bactéries
A : fMet-ARNt$_f^{Met}$ initiateur de *E. coli*. **B** : Met-ARNt$_m^{Met}$ élongateur de *E. coli*. Les régions importantes pour la MTF sont *en bleu*. Les bases C_1 A_{72} sont exclues par la PTH. Le groupement formyle responsable de la spécificité de reconnaissance du fMet-ARNt$_f^{Met}$ par IF2 est *en jaune*. Les trois paires G=C nécessaires au recrutement IF3-dépendant de l'ARNt$_f^{Met}$ au site P sont *en rouge*.

9.2.2 Particularité des anticodons des ARNt initiateurs

De nombreuses études biochimiques ont montré que la boucle anticodon des ARNt initiateurs devait adopter une conformation particulière. La plus ancienne de ces études a été réalisée par des tests de clivage enzymatique à la nucléase S1 [120]. Cette étude a montré que la boucle anticodon des ARNt initiateurs était plus résistante à l'action de la nucléase S1 que ne l'est celle des ARNt élongateurs.[1] En effet, tous les ARNt initiateurs sont clivés après les nucléotides C_{34} et A_{35} (FIGURE 9.3 **A**, **C** et **E** – colonne de gauche), alors que les ARNt élongateurs sont clivés après les nucléotides U_{33}, C_{34}, A_{35} et U_{36} (FIGURE 9.3 **B**, **D** et **F** – colonne de droite). Les auteurs ont de ce fait conclu que la boucle anticodon des ARNt initiateurs adopte une conformation unique, différente de celle des ARNt élongateurs.

1. Les ARNt initiateurs testés dans cette étude sont l'ARNt$_f^{Met}$ d'*E. coli*, l'ARNt$_i^{Met}$ de levure et l'ARNt$_i^{Met}$ de mammifères, purifié à partir de placenta humain. Les ARNt élongateurs testés dans cette étude sont les ARNt$_m^{Met}$ d'*E. coli* et de levure, les ARNt^{Phe} de levure et de *S. pombe* ainsi que l'ARNt$_2^{Glu}$ et l'ARNt^{Trp} d'*E. coli*.

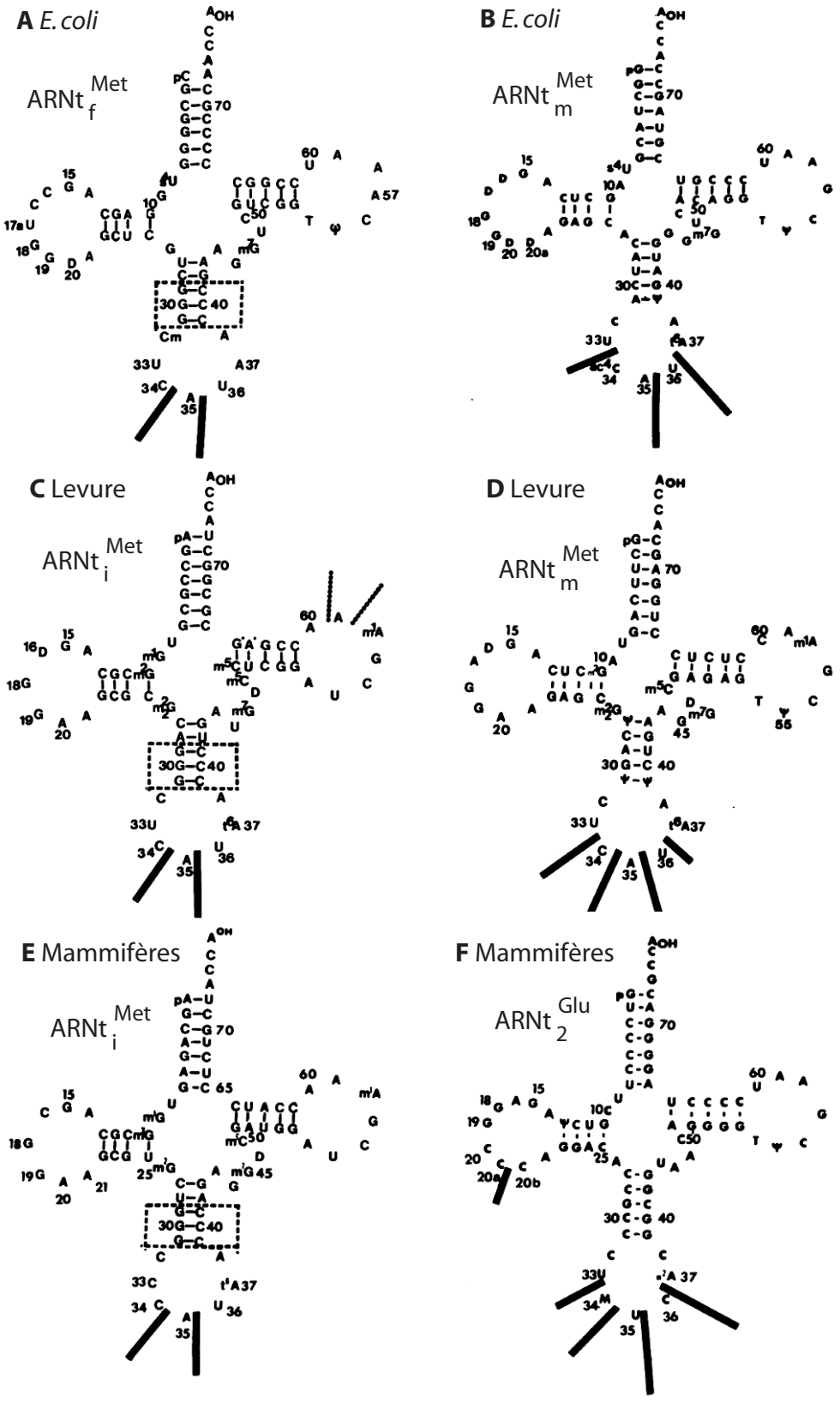


Figure 9.3 – Sites de clivage à la nucléase S1
Les sites de clivages comparés des ARNt initiateurs et élongateurs sont indiqués pour *E. coli* (**A** et **B**), la Levure (**C** et **D**) et les Mammifères (**E** et **F**). Les longueurs des barres noires reflètent l'importance relative des clivages. D'après Wrede *et al.* [120].

De plus, le fait que la séquence de la boucle anticodon soit identique pour l'ARNt$_i^{Met}$ initiateur de levure et pour l'ARNt$_m^{Met}$ élongateur de levure (FIGURE 9.3 **C** et **D**) a fait formuler l'hypothèse que les trois paires G=C caractéristiques de la tige anticodon des ARNt initiateurs étaient à l'origine de cette conformation particulière [120].

Par la suite, la stabilité de cette conformation particulière a été étudiée — également par des tests de clivage enzymatiques à la nucléase S1 — et il a été montré que celle-ci était stable sur une gamme de pH comprise entre 4.5 et 7.4, qu'elle n'était pas sensible à la force ionique, et qu'elle n'était pas modifiée en présence de Mg^{2+} [444, 445]. De plus ces sites de clivages à la nucléase S1 ont été depuis reproduits par de nombreuses équipes [122, 123, 445, 446].

La conformation particulière de la boucle anticodon a aussi été étudiée en solution par sonde chimique [121]. Dans cette étude, le fait que les positions N7 des trois guanines consécutives de la tige anticodon ne soient pas réactives vis-à-vis du diméthyle sulfate, et spécialement le N7 du G_{30}, a amené les auteurs à conclure que la tige anticodon des ARNt initiateurs était plus régulière et plus rigide que celle des ARNt élongateurs. En effet, quelques soient les bases qui l'entourent, le N7 du G_{30} est toujours réactif chez les ARNt élongateurs possédant une guanine en position 30 [447–449].

Ainsi la conformation particulière de la région de l'anticodon des ARNt initiateurs semble affecter aussi bien la boucle que la tige anticodon. Et les trois paires G=C de la tige, caractéristiques des ARNt initiateurs, semblent à la fois y participer, et en être la cause. Cependant, à ce jour aucune étude structurale n'est venue confirmer ces expériences biochimiques. En effet, deux structures d'ARNt initiateurs ont été résolues par cristallographie. Il s'agit de l'ARNt$_f^{Met}$ d'*E. coli* [417], et de l'ARNt$_i^{Met}$ de levure [36].

En ce qui concerne la structure de l'ARNt$_f^{Met}$ d'*E. coli*, une conformation particulière de l'anticodon a été décrite par les auteurs [417]. Cependant, cette conformation particulière — connue sous le nom de « uridine pivotante » (*uridine swivel hypothesis*) — n'est en accord, ni avec les résultats d'expériences réalisées par sonde chimique en solution [121], ni avec la structure cristalline de l'ARNt$_f^{Met}$ d'*E. coli* en complexe avec la MTF [38], qui montrent que l'ARNt$_f^{Met}$ — comme tous les ARNt élongateurs — possède un motif en « U-turn » au niveau de l'uridine 33.

Pour ce qui est de la structure de l'ARNt$_i^{Met}$ de levure, la région de l'anticodon est mal définie et elle n'a pas permis de discuter d'une conformation particulière au niveau de la boucle anticodon [36].

D'autre part, une étude RMN réalisée sur des tiges-boucles mimant les anticodons de l'ARNt$_f^{Met}$ et de l'ARN$_m^{Met}$ d'*E. coli* n'a pas mis en évidence de différence notoire dans la conformation des deux tiges-boucles [445]. Les auteurs proposent en fait que la différence de sensibilité des deux tiges-boucles à la nucléase S1 soit due à une simple différence de flexibilité — la tige-boucle initiatrice possédant un caractère plus rigide.

Ainsi, d'un point de vue structural, la conformation particulière de la tige-boucle anticodon des ARNt initiateurs — révélée par des tests enzymatiques de clivage à la nucléase S1 et par sondes chimiques en solution — reste dans une certaine mesure inexpliquée.

9.3 L'ARNt initiateur au sein du site P du ribosome

La sélection du fMet-ARNt$_f^{Met}$ en tant qu'ARNt initiateur s'explique en partie par des interactions spécifiques entre le site P de la sous-unité 30S et les particularités de séquence des ARNt initiateurs que nous venons de présenter. Nous allons ainsi détailler les différents contacts entre l'ARNt initiateur au sein du site P du ribosome et la sous-unité 30S — essentiellement l'ARN 16S, mais aussi quelques protéines de la petite sous-unité (pour des revues abordant la structure du site P [450, 451]).

9.3.1 Les interactions au niveau du site P

Avec l'ARNr 16S

Lorsque la structure cristalline à 7.8 Å du ribosome 70S de *T. thermophilus* a été résolue, en complexe avec une tige-boucle anticodon au sein du site P de la sous-unité 30S [452], les contacts entre le ribosome et la tige-boucle anticodon de l'ARNt ont été vus pour la première fois. Trois zones de contacts ont été observées entre l'ARNt$_f^{Met}$ et l'ARNr 16S auxquelles viennent s'ajouter d'autres contacts additionnels au niveau de la zone d'interaction codon–anticodon. Depuis, la résolution de structures de ribosomes — en complexe avec l'ARNm, des ARNt entiers ou des tiges-boucles anticodon — à des résolutions sans cesse améliorées [46, 47, 453–455], a permis d'identifier ces contacts et d'en donner une représentation plus précise. Ceux-ci sont présentés à la FIGURE 9.4.

Les trois zones de contact entre l'ARNt$_f^{Met}$ et l'ARNr 16S sont les suivantes :
- Les bases G1338 et A1339 qui forment des interactions dans le petit sillon de l'ARNt$_f^{Met}$ avec respectivement les paires de bases G_{29}=C_{41} et G_{30}=C_{40}.
- La chaîne ribose-phosphate du nucléotide A790 qui interagit avec la tige anticodon de l'ARNt$_f^{Met}$ au niveau des nucléotides A_{38} et C_{39}.
- La chaîne ribose-phosphate des nucléotides A1229 et C1230, qui interagit avec la chaîne ribose-phosphate de l'ARNt$_f^{Met}$ au niveau des G_{30} et G_{31}. [2]

Les contacts additionnels au niveau de l'interaction codon–anticodon impliquent six nucléotides dont quatre d'entre eux présentent des modifications post-transcriptionnelles. Ces modifications sont toutes des méthylations qui

2. Dans un souci de clarté, les nucléotides A1229 et C1230 n'ont pas été représentés sur la FIGURE 9.4.

ont lieu sur la base ou le ribose. Les nucléotides qui interviennent au niveau de la zone d'interaction codon–anticodon sont les suivants :

- Les nucléotides m^5C1400 et m_2^2G966, sur lesquels viennent s'empiler respectivement la base et le ribose du C_{34} de l'anticodon.
- Les nucléotides $m^4Cm1402$ et C1403, qui interagissent par l'intermédiaire de leur groupement amino avec le groupement phosphate du nucléotide G3 de l'ARNm.
- Le nucléotide m^3U1498, qui vient en contact avec la chaîne ribose-phosphate des nucléotides A1 et U2 du codon de l'ARNm.
- Le nucléotide G926, dont les positions N_1 et N_2 sont en liaison hydrogène avec le phosphate du nucléotide A1 du codon de l'ARNm. [3]

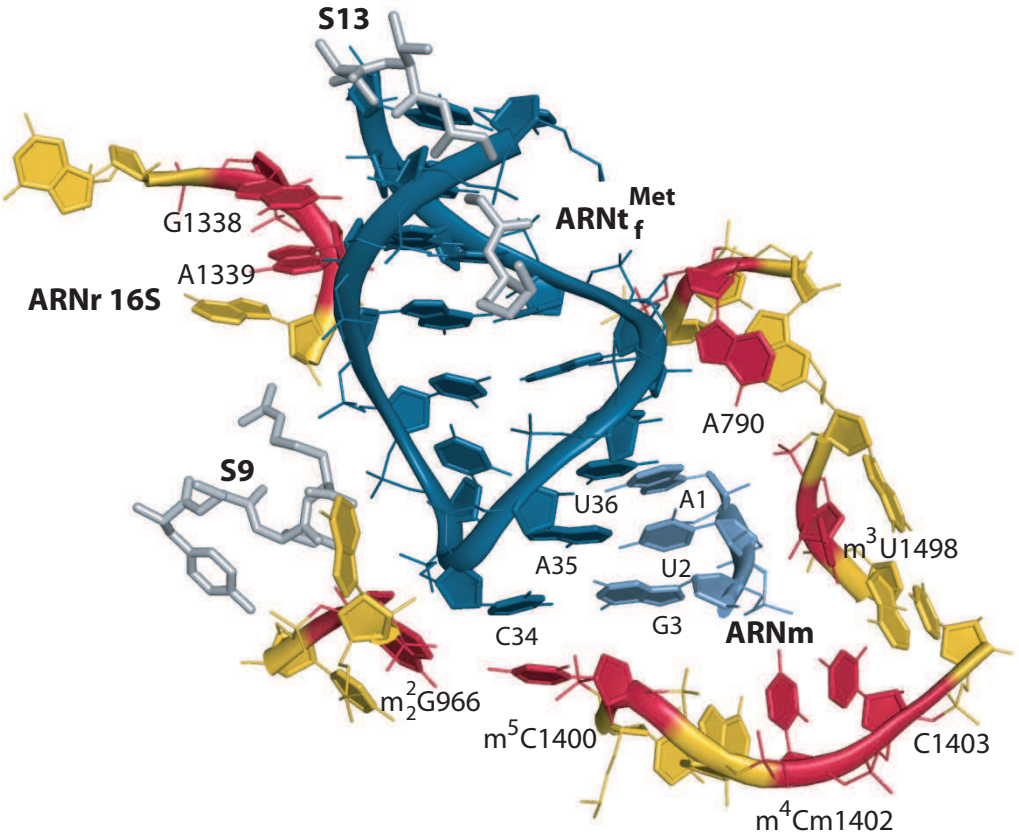


FIGURE 9.4 – Interactions au sein du site P du ribosome
La tige-boucle anticodon de l'$ARNt_f^{Met}$ (residus 28–42) est *en bleu foncé*. L'ARNr 16S est *en jaune*, et les nucléotides de l'ARNr 16S directement en contact avec l'$ARNt_f^{Met}$ ou avec l'ARNm sont *en rouge*. Le codon de l'ARNm ($A_1U_2G_3$) est en *bleu clair*. Les protéines ribosomales S9 et S13 en contact avec l'$ARNt_f^{Met}$ sont en *gris*. Pour ne pas surcharger la figure, les nucléotides A1229, C1230 et G926 de l'ARNr 16S n'ont pas été représentés (se reporter au texte pour une description précise des contacts qu'ils effectuent). Les coordonnées du site P et de l'$ARNt_f^{Met}$ correspondent à celle du ribosome 70S de *T. thermophilus* à 2.8 Å de résolution (PDB code 2J00) [47].

3. Dans un souci de clarté, le nucléotide G966 n'a pas été représenté sur la FIGURE 9.4.

Avec les protéines S9 et S13

Deux protéines de la petite sous-unité ribosomique sont retrouvées à proximité de l'ARNt fixé au site P. Il s'agit des protéines S9 et S13, et plus particulièrement de leur extension C-terminale. Ces deux protéines ont toutes les deux des longues extensions C-terminales qui viennent se loger au niveau du site P de la sous-unité 30S. Ces extensions contiennent des résidus basiques lysines et arginines qui peuvent interagir avec les phosphates de l'ARNt.

Cependant, le rôle de ces protéines dans le fonctionnement du site P semble secondaire, puisque lorsque les extensions C-terminales des deux protéines S9 et S13 sont clivées, les ribosomes restent fonctionnels [456]. Ces extensions C-terminales aident cependant dans une certaine mesure à la fixation des ARNt au site P, puisque leur délétion affecte la vitesse de croissance des mutants correspondants.

9.3.2 Une préférence pour les ARNt initiateurs au sein du site P

La sélection du fMet-ARNt$_f^{Met}$ en tant qu'ARNt initiateur s'explique en partie par les interactions au sein du site P entre les nucléotides G1338 et A1339 et les paires G_{29}=C_{41} et G_{30}=C_{40} de la tige anticodon. En effet, les nucléotides G1338 et A1339 effectuent des interactions dites A-mineur (*A-minor interaction*) [457] dans le petit sillon de la tige anticodon. Plus précisement, le nucléotide A1339 effectue une interaction A-mineur de type I avec la paire G_{30}=C_{40} (FIGURE 9.5 **A**) et le nucléotide G1338, une interaction A-mineur de type II avec la paire G_{29}=C_{41} (FIGURE 9.5 **B**).

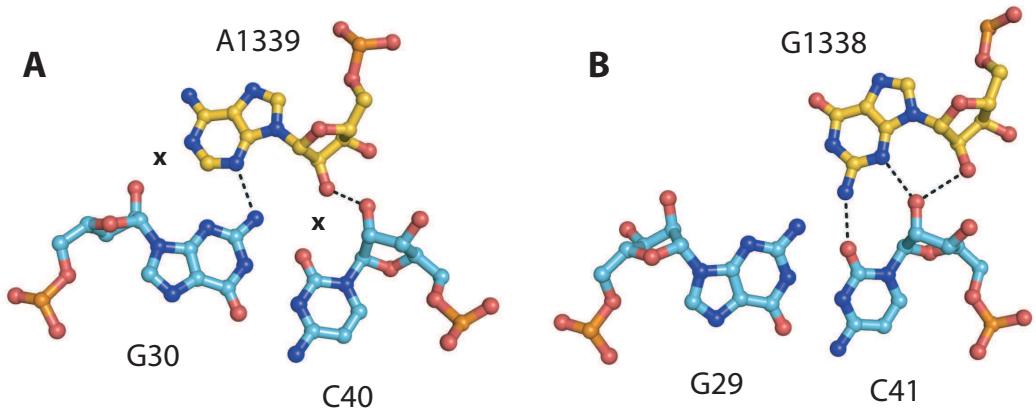


FIGURE 9.5 – Interaction de type A-mineur avec la tige anticodon de l'ARNt$_f^{Met}$ au sein du site P
A : Interaction A-mineur de type I. Les croix **x** indiquent que les distances sont trop grandes pour former les liaisons hydrogènes trouvées typiquement dans ce motif. **B** : Interaction A-mineur de type II. Les bases correspondent à la structure du ribosome 70S de *T. thermophilus* (PDB code 2J00) [47].

Bien que ces interactions de type A-mineur soient possibles avec n'importe quelle paire de base, elles sont plus fortes avec les paires G=C qu'avec les paires A–U [458]. Il a ainsi été proposé que ces interactions de type A-mineur entre l'ARNr 16S du site P et la tige anticodon des ARNt puissent participer à la discrimination des ARNt initiateurs [459] — qui présentent trois paires G=C caractéristiques dans leur tige anticodon (FIGURE 9.2). Cette hypothèse a ensuite été confirmée grâce à des expériences de mutagénèse [460]. Cependant ces seules interactions A-mineur ne semblent pas suffire à définir la spécificité de sélection IF3-dépendante de l'ARNt$_f^{Met}$ en tant qu'ARNt initiateur [460].

En effet, à l'heure actuelle des zones d'ombres persistent quant à l'étape de décodage IF3-dépendante au sein du site P du ribosome. Nous tâcherons dans le PARAGRAPHE suivant de présenter les données communément admises concernant l'étape de décodage IF3-dépendante ainsi que les points restant à élucider.

9.4 La sélection IF3-dépendante de l'ARNt initiateur

Comme nous l'avons vu au PARAGRAPHE 9.1, une des fonctions cruciales du site P de la petite sous-unité du ribosome est de fixer sélectivement l'ARNt initiateur sur le codon de démarrage de l'ARNm. Cette fixation sélective conditionne la fidélité de la synthèse protéique. Chez les bactéries, le facteur d'initiation IF3 induit la sélection des ARNt au site P au travers d'interactions avec la petit sous-unité du ribosome. Ce facteur entraîne en effet l'éjection des ARNt élongateurs présents au niveau du site P, et renforce au contraire la fixation de l'ARNt$_f^{Met}$ initiateur [461, 462]. Même si des changements conformationnels, induits par IF3, ont été décrits pour l'ARNm, l'ARNt$_f^{Met}$ fixé au site P et la sous-unité 30S du ribosome [427, 463, 464], le rôle de ceux-ci dans la sélection IF3-dépendante de l'ARNt initiateur n'a jusqu'alors pas été montré de façon satisfaisante.

Un des points clés du mécanisme de sélection IF3-dépendante de l'ARNt$_f^{Met}$ est la présence des trois paires G=C consécutives, retrouvées dans la tige anticodon de tous les ARNt initiateurs [12, 433, 434]. Cette caractéristique de séquence est nécessaire, en présence des facteurs d'initiation, à la fixation de l'ARNt$_f^{Met}$ au site P de la sous-unité 30S *in vitro* [122]. Elle est aussi indispensable pour l'initiation de la traduction *in vivo* [446]. Il a aussi été montré par des expériences de mutagénèse, que la sélection IF3-dépendante de tiges-boucles anticodons ne nécessitait pas seulement ces paires G=C caractéristiques des initiateurs, mais également la séquence inchangée de la boucle anticodon [123]. De façon très intéressante, ces mêmes expériences de mutagénèse ont montré que la sélection IF3-dépendante d'un ARNt en tant qu'ARNt initiateur était corrélée à sa moindre sensibilité vis-à-vis de la nucléase S1 [123]. En effet, des mutations dans la tige ou dans la boucle anticodon de l'ARNt$_f^{Met}$

abolissent de façon concomitante sa capacité à initier la synthèse protéique et son profil particulier de clivage à la nucléase S1 [122, 123] (voir FIGURE 9.3 pour les profils de clivages à la nucléase S1). Ceci suggère très fortement que la conformation particulière de la tige-boucle anticodon des ARNt initiateurs (PARAGRAPHE 9.2.2) soit nécessaire à leur fonction initiatrice, et cela au travers d'une sélection IF3-dépendante de cette conformation au sein du site P.

Cependant, une étude menée par l'équipe de Uttam RAJBHANDARY semble en contradiction avec cette affirmation. En effet, ils ont montré qu'un ARNt suppresseur pouvait servir d'ARNt initiateur sur le codon de terminaison UAG [465]. Et bien que les trois paires G=C de la tige soient indispensables à son activité initiatrice, la boucle anticodon de cet ARNt ne présente pas de résistance particulière à la nucléase S1. De plus, son profil de clivage est inchangé lorsque des mutations qui abolissent sa sélection en tant qu'ARNt initiateur sont introduites dans la tige [446]. Ceci suggère donc que la conformation particulière de la tige-boucle anticodon — qui trouve vraisemblablement son origine dans les trois paires G=C de la tige et dans la séquence particulière de la boucle — ne soit pas un déterminant de sélection des ARNt initiateurs, mais que seule la présence des trois paires G=C soit requise et que celles-ci soient directement reconnues au niveau du site P du ribosome — soit par le ribosome lui-même, soit par le facteur IF3. Cependant, comme le font remarquer les auteurs, le fait que l'ARNt suppresseur construit présente un profil de clivage à la nucléase S1 différent de celui des ARNt initiateurs, mais qu'il puisse tout de même initier la synthèse protéique pourrait également s'expliquer si l'on considère que celui-ci est impliqué dans un équilibre conformationnel, entre la conformation de l'$\mathrm{ARNt_f^{Met}}$ — qui lui confère par là son activité initiatrice — et celle des ARNt élongateurs — qui lui donne ainsi ce profil de digestion à la nucléase S1 [446]. En effet, si l'on admet une interconversion possible entre la conformation « initiateur » et la conformation « élongateur » au sein des tiges boucles anticodons, cet équilibre conformationnel pourrait être déplacé, dans le cas de l'ARNt suppresseur de cette étude, en faveur de celle des ARNt élongateurs par une simple modification de la séquence de la boucle. Mais la modification de séquence étant tout de même assez minime, une certaine proportion d'ARNt suppresseur pourrait présenter la conformation « initiateur » de l'anticodon et ainsi permettre l'initiation de la synthèse protéique. De façon intéressante, un équilibre conformationnel de la tige-boucle anticodon de l'$\mathrm{ARNt_f^{Met}}$ a été observé par RMN, mais sa nature exacte n'a pas été déterminée [445].

En résumé, il n'est ainsi pas absolument certain que la conformation particulière des tiges-boucles anticodon des ARNt initiateurs soit nécessaire à leur fonction initiatrice. Ainsi, les trois paires G=C requises pour la fonction initiatrice des ARNt doivent être directement reconnues au niveau du site P du ribosome — soit par le ribosome lui-même, soit par le facteur IF3. Les structures des complexes 70S–$\mathrm{ARNt_f^{Met}}$ [46, 47, 455] ont donné une explication structurale quant au rôle de deux des paires G=C de la tige. En effet, deux d'entre elles sont impliquées dans des interactions de type A-mineur avec les

nucléotides G1338 et A1339 de l'ARNr 16S (Paragraphe 9.3.2 et Figure 9.5). L'implication de ces nucléotides du site P dans la sélection IF3 dépendante de l'ARNt initiateur a de plus été montré par des expériences de mutagénèse [460].

Cependant, le fait que la sélection de l'ARNt initiateur ne puisse pas s'effectuer en l'absence du facteur d'initiation IF3 a amené à formuler l'hypothèse que sous l'effet d'un changement conformationnel induit par IF3, les bases G1338 et A1339 venaient vérifier l'identité de l'ARNt présent au site P grâce à ces interactions A-mineur [459]. Le fait que l'interaction A-mineur entre la base A1339 et la paire G_{30}=C_{40} ne soit pas optimale dans la structure du ribosome 70S (Figure 9.5) va également dans ce sens, puisque durant l'initiation, de petits mouvements de la « tête » de la petite sous-unité pourraient permettre d'obtenir une géométrie optimale afin d'inspecter le petit sillon de la tige anticodon de l'ARNt présent au site P [47].

Il aurait également été possible que le facteur IF3 intervienne en reconnaissant directement les paires G=C de la tige anticodon de l'ARNt initiateur, mais au regard de la connaissance que l'on a de la position de IF3 sur la petite sous-unité [459, 463], celui-ci semble trop éloigné de cette région de l'anticodon pour pouvoir réaliser la vérification par lui-même.

Toutes ses données semblent bien expliquer le rôle des paires G=C, celui des nucléotides G1338 et A1339, et celui du facteur IF3. Cependant, quelques zones d'ombres subsistent concernant ce mécanisme de décodage IF3-dépendant. En effet, si seules les deux interactions A-mineur avec les paires G_{29}=C_{41} et G_{30}=C_{40} étaient requises pour la sélection de l'ARNt initiateur, certains ARNt présentant deux paires G=C (G_{29}=C_{41} et G_{30}=C_{40}) dans leur tige anticodon devraient être plus résistants à l'éjection du site P par IF3 [460]. De plus, ce modèle n'explique pas le rôle joué par la troisième paire G=C (G_{31}=C_{39}). D'autre part, si seules les deux interactions A-mineur étaient requises, des mutants de la sous-unité 30S aux positions 1338 et 1339 devraient, en présence du facteur IF3, éjecter tous les ARNt du site P, y compris l'$ARNt_f^{Met}$. Ceci a été observé à de fortes concentrations d'IF3, mais le comportement du site P à de faibles concentrations d'IF3 suggère qu'il existe une interaction additionnelle, spécifique de la région de l'anticodon de l'$ARNt_f^{Met}$, qui favorise sa fixation au sein du site P [460]. La nature de cette interaction additionnelle n'a cependant toujours pas été identifiée.

Chapitre 10

Une structure unique de l'anticodon de l'ARNt$_f^{Met}$ d'*E. coli*

10.1 Problématique et résumé de l'Article 4

D'un point de vue structural, la conformation particulière de la tige-boucle anticodon des ARNt initiateurs — révélée par des tests enzymatiques de clivage à la nucléase S1 et par sondes chimiques en solution — reste à l'heure actuelle inexpliquée. Lors d'essais de co-cristallisation de la m^{1}A$_{58}$ MTase de *T. thermophilus* avec l'ARNt$_f^{Met}$ d'*E. coli*, nous avons obtenu des cristaux de l'ARNt$_f^{Met}$ seul, qui nous ont permis de résoudre la structure de cet ARNt initiateur. La diffraction des cristaux d'ARNt$_f^{Met}$ s'est révélée être très anisotrope. Cependant, le traitement attentif de cette anisotropie nous a permis d'obtenir des cartes de densité très bien définies, et ce également pour la région de l'anticodon. L'ARNt$_f^{Met}$ adopte la structure classique en forme de « L » des ARNt. Cependant, la région de l'anticodon révèle une structure unique impliquant directement deux des trois paires G=C de l'anticodon. Cette structure permet d'expliquer de nombreuses données biochimiques de la littérature. Elle peut être décrite par trois éléments particuliers :

- Une paire de base Cm$_{32}$–•–A$_{38}$ de type wobble qui étend la tige anticodon d'une paire supplémentaire
- Une base triple entre la base A$_{37}$ située juste après le triplet de l'anticodon et la paire G$_{29}$=C$_{41}$ dans le grand sillon de la tige anticodon
- Un empilement particulier des bases de la boucle anticodon, où la base A$_{38}$ est directement empilée sur la dernière base de l'anticodon (U$_{36}$), alors qu'elle est empilée sur la base 37 dans toutes les structures d'ARNt élongateurs

Cette structure particulière de la région de l'anticodon de l'ARNt$_f^{Met}$ est représentée en vue stéréoscopique à la FIGURE 10.1.

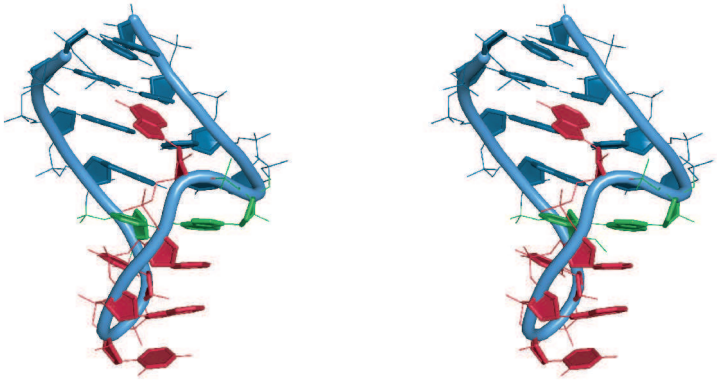


FIGURE 10.1 – Structure unique de l'anticodon de l'ARNt$_f^{Met}$
Vue stéréoscopique de l'anticodon de l'ARNt$_f^{Met}$. Les nucléotides de la tige sont en *bleu foncé*, la paire de base Cm_{32}–•–A_{38} de type wobble est *en vert*, et les nucléotides de la boucle sont *en rouge*. La base A_{37} vient interagir avec la paire G_{29}=C_{41} dans le grand sillon de la tige anticodon.

Cette structure unique pourrait permettre une interaction additionnelle discriminante entre la région de l'anticodon de l'ARNt initiateur et la machinerie de traduction au niveau du site P du ribosome — soit de manière directe, soit par l'intermédiaire d'IF3, et ainsi être un élément crucial — en plus des paires G=C impliquées dans les interactions A-mineur — permettant la sélection de l'ARNt initiateur lors de l'initiation de la traduction.

10.2 Article 4 | Barraud *et al.* (2008)

4894–4901 *Nucleic Acids Research, 2008, Vol. 36, No. 15* *Published online 24 July 2008*
doi:10.1093/nar/gkn462

A unique conformation of the anticodon stem-loop is associated with the capacity of $tRNA_f^{Met}$ to initiate protein synthesis

Pierre Barraud[1], Emmanuelle Schmitt[2], Yves Mechulam[2], Frédéric Dardel[1] and Carine Tisné[1,*]

[1]Laboratoire de Cristallographie et RMN Biologiques, Université Paris Descartes, CNRS, 4 avenue de l'Observatoire, 75006 Paris and [2]Laboratoire de Biochimie, Ecole Polytechnique, CNRS, 91128 Palaiseau, France

Received May 26, 2008; Revised June 30, 2008; Accepted July 2, 2008

ABSTRACT

In all organisms, translational initiation takes place on the small ribosomal subunit and two classes of methionine tRNA are present. The initiator is used exclusively for initiation of protein synthesis while the elongator is used for inserting methionine internally in the nascent polypeptide chain. The crystal structure of *Escherichia coli* initiator $tRNA_f^{Met}$ has been solved at 3.1 Å resolution. The anticodon region is well-defined and reveals a unique structure, which has not been described in any other tRNA. It encompasses a Cm32•A38 base pair with a peculiar geometry extending the anticodon helix, a base triple between A37 and the G29-C41 pair in the major groove of the anticodon stem and a modified stacking organization of the anticodon loop. This conformation is associated with the three GC basepairs in the anticodon stem, characteristic of initiator tRNAs and suggests a mechanism by which the translation initiation machinery could discriminate the initiator tRNA from all other tRNAs.

INTRODUCTION

During translation initiation, the start codon of the message must be identified and aligned in the P-site of the 30S subunit, so that it basepairs with the anticodon of initiator tRNA. In *Escherichia coli*, this step involves initiation factor IF3 and requires the presence of three consecutive GC base pairs closing the anticodon loop. Interestingly, these GC base pairs are an exclusive hallmark of cytoplasmic initiator tRNAs from all living organisms (1). It has been shown for a long time that three consecutive GC pairs in the anticodon stem are a crucial feature that distinguishes initiator tRNA from elongator tRNAs during translational initiation (2–5). The presence of this specific sequence in the anticodon stem is associated with a modified reactivity towards nucleases and chemical reagents, suggesting that it induces structural differences related to the translation initiation process (2–4). However, currently available structures of initiator tRNA have so far provided little clues as to what this specific anticodon conformation might be, probably because this stem-loop is often less ordered in crystal structures (6,7). An NMR study comparing initiator and elongator stem-loop structures showed that they have closely similar solution structures (8). However, in the same study, NMR data collected at low temperature revealed a possible second conformation of the initiator stem-loop that could not be further analyzed for technical reasons. In addition, despite the large amount of structural data accumulated on the ribosome (9–11), it is not yet fully understood how the initiator tRNA is selected among all other tRNAs during initiation. Recent crystal structures of 70S ribosome functional complexes (9–11) provide a detailed description of how the ribosome interacts with its mRNA and tRNA partners [for a recent review, (12)]. Actually, the ribosome makes a number of contacts with the anticodon arm of tRNA in the P-site. Among these, the universally conserved nucleotides G1338 and A1339 which are located in the head of the small subunit form type II and type I A-minor interactions (13) with the GC base pairs 29:41 and 30:40 of initiator tRNA. By mutational studies (14,15), these interactions were demonstrated to help in differentiating between the initiator and the elongator tRNAs. However, displacement by IF3 of initiator-like tRNAs that contain these two GC pairs indicates that an additional feature allow to distinguish them from authentic initiator tRNAs (14,15). Additionally, although the three GC pairs are required for discrimination of initiation tRNA by IF3, they are not sufficient to specify the identity of initiator tRNA (4,16). These results are in favor with the formation of an additional interaction possibly specific to the anticodon

*To whom correspondence should be addressed. Tel: +33 1 53 73 15 72; Fax: +33 1 53 73 99 25; Email: carine.tisne@univ-paris5.fr

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

Nucleic Acids Research, 2008, Vol. 36, No. 15 **4895**

region of the initiator tRNA that would favor its binding to the P-site in the presence of IF3.

We have determined the crystal structure of *E. coli* initiator tRNA$_f^{Met}$ at 3.1 Å resolution. Overall, tRNA$_f^{Met}$ has the classical L-shape geometry, with usual tertiary interactions. However, its structure reveals a unique conformation of the anticodon stem-loop, which explains previous biochemical data and provides information of its specific function in translation initiation.

MATERIALS AND METHODS

tRNA purification

The *E. coli* tRNA$_f^{Met}$ was overexpressed from plasmid pBSTtRNA$_f^{Met}$ in *E. coli* JM101TR using a protocol derived from that of Meinnel *et al.* (17). Briefly, after phenol extraction of RNAs from bacteria, total tRNA was separated by gel filtration on a HiLoad 26/60 Superdex 75 prepgrade chromatography column (Amersham Biosciences) equilibrated in 25 mM Tris–HCl pH 7.0. The overexpressed tRNA$_f^{Met}$ was then separated from other tRNAs by an anion exchange step (Resource Q column) equilibrated in 25 mM Tris–HCl pH 7.0. tRNAs were eluted using a 350–550 mM NaCl gradient in the same buffer. The tRNA$_f^{Met}$ was further purified by a hydrophobic interaction step (Phenyl Superose column) equilibrated in 10 mM ammonium acetate pH 6.5, 1.7 M ammonium sulfate. tRNA$_f^{Met}$ was eluted using a 1.7–1.0 M ammonium sulfate reverse gradient. The fractions containing the purified tRNA$_f^{Met}$ were pooled, dialyzed against 50 mM Tris–HCl pH 8.0, 100 mM KCl, concentrated using Amicon® Ultra (Millipore) and stored at –20°C.

Crystallization and crystal characterization

tRNA$_f^{Met}$ was prepared at a initial concentration of 3 mg/ml in 50 mM Tris–HCl pH 8.0, 100 mM KCl and 10 mM $MgCl_2$ and crystallized by the vapor diffusion method at 19°C in several conditions. All conditions contained at least 2.0 M ammonium sulfate as a precipitating agent. Suitable crystals for diffraction data collection, were grown up at two different pH conditions: pH 4.6 (2.0 M ammonium sulfate, 100 mM AcOH/AcONa pH 4.6) and 8.0 (2.0 M ammonium sulfate, 50 mM Tris–HCl pH 8.0). Crystals grew up to 100 µm within 1 week. The hexagonal crystals belonged to space group $P6_422$ with one tRNA$_f^{Met}$ per asymmetric unit and with a particularly high solvent content of 79% (Table 1 for statistics).

Table 1. Data collection and refinement statistics for the determination of tRNA$_f^{Met}$ structure

	pH 4.6 (PDB ID 3CW5)	pH 8.0 (PDB ID 3CW6)
Data collection statistics		
Beamline (ESRF)	ID14–3	ID29
Wavelength (Å)	0.931	0.918
No. of frames ($\Delta\phi = 1°$)	90	90
Space group	$P6_422$	$P6_422$
Unit-cell parameters (Å)		
a	93.79	93.80
c	219.36	218.59
No. of measured reflections	79 471 (10 063)[a] {96 111 (13 300)}[b]	68 263 (4286) {82 534 (11 149)}
No. of unique reflections	10 063 (899) {10 810 (1475)}	8490 (875) {9240 (1318)}
Resolution limits (Å)	40.00–3.10 (3.27–3.10)	40.00–3.30 (3.48–3.30)
R_{merge}[c] (%)	9.9 (47.7) {11.7 (81.0)}	10.3 (67.4) {12.4 (137.2)}
Mean I/σ(I)	15.7 (2.7) {15.3 (3.5)}	13.9 (2.2) {13.5 (1.9)}
R_{pim}[d] (%)	3.7 (27.4) {4.2 (27.7)}	3.8 (32.8) {4.4 (49.2)}
Data completeness (%)	91.4 (59.8) {97.9 (95.1)}	91.4 (67.4) {99.5 (100.0)}
Multiplicity	7.9 (3.2) {8.9 (9.0)}	8.0 (4.9) {8.9 (8.5)}
Wilson *B*-value (Å^2)	31.3 {65.1}	36.7 {56.7}
Refinement statistics		
Resolution range	40.00–3.10 (3.18–3.10)	40.00–3.30 (3.39–3.30)
Number of reflexions	9197 (468)	7123 (268)
R-factor (%)	24.1 (32.8)	23.1 (28.1)
R_{free} (%)	27.0 (43.4)	27.4 (39.5)
Number of tRNA atoms	1645	1645
RMSD (bonds) (Å)	0.009	0.009
RMSD (angles) (°)	1.735	1.946
Mean residual *B*-value (Å^2)[e]	19.5	82.3
Solvent content (%)	79	79

[a]Values for the outermost shell are given in parenthesis.
[b]Values before anisotropic treatment are given in brackets. Statistics for anisotropic data sets, used in structure refinement, were assessed on data sets including only reflections with $I/\sigma(I) > 2.0$. The lower completeness at the high-resolution limit reflects the diffraction anisotropy.
[c]$R_{merge} = 100 \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$, where the outer sum ($h$) is over the unique reflections, and the inner sum (i) is over the set of independent observations of each unique reflection.
[d]$R_{pim} = 100 \sum_h \sqrt{1/N-1} \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$, where the outer sum ($h$) is over the unique reflections, the inner sum (i) is over the set of independent observations of each unique reflection, and N the redundancy of the measured reflections.
[e]As TLS parameters were refined, B-value include only residual B-value.

4896 *Nucleic Acids Research, 2008, Vol. 36, No. 15*

Data collection, structure solution and refinement

The crystals were harvested, soaked in a cryoprotectant solution [2.2–2.4 M ammonium sulfate, $MgCl_2$ 10 mM, glycerol 20% (v/v), and 50 mM AcOH/AcONa pH 4.6 or 50 mM Tris–HCl pH 8.0 depending the pH at which the crystals were grown up] for about 1–5 min and flash-frozen in liquid nitrogen before data collection. Diffraction data were collected at beam lines ID14–3 and ID29 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). All crystallographic calculations were performed using the CCP4 suite version 6 (18) as implemented in the graphical user interface (19). X-ray diffraction data were processed using MOSFLM (20) and scaled with SCALA (21). The structure of $tRNA_f^{Met}$ was solved by molecular replacement using the program PHASER (22) and the coordinates of $tRNA_f^{Met}$ from the *E. coli* $tRNA_f^{Met}$/formylase complex (PDB entry code 2FMT) as a model. Model and map visualizations for manual reconstruction were performed with the program COOT (23). In the last stages of refinement, TLS parameters (24) were refined using one group for the entire tRNA molecule. The acidic structure was refined as described in the following paragraph, and the model was then used to refined the basic structure.

Coordinates and structure factors of *E. coli* $tRNA_f^{Met}$ have been deposited at the Protein Data Bank with the accession code 3CW5 and 3CW6.

RESULTS AND DISCUSSION

Structure determination

We have solved the crystal structure of the *E. coli* initiator $tRNA_f^{Met}$ at two pH conditions (i.e. 4.6 and 8.0) using careful treatment of data anisotropy. Indeed, the diffraction patterns of the two crystal forms were severely anisotropic, with the acidic condition giving however higher resolution limits (Table 1). For instance, in the acidic condition, diffraction limits are 2.8 Å along the a^* and b^* directions, and 3.4 Å along the c^* direction. We first attempted to refine the structure using the anisotropic scaling procedure applied by REFMAC (25). However, the obtained electron density maps were relatively featureless, especially around the anticodon stem-loop. To improve map quality, we then tried the procedure described by David Eisenberg's group (26) to deal with highly anisotropic data (http://www.doe-mbi.ucla.edu/~sawaya/anisoscale/). Briefly, data falling outside an ellipse centered at the reciprocal lattice origin and having vertices at 1/2.8, 1/2.8 and 1/3.4 $Å^{-1}$ along a^*, b^* and c^*, respectively, were removed. After this ellipsoidal truncation, anisotropic scale factors were applied. Lastly, a negative isotropic B factor was used to restore the magnitude of the high-resolution reflections diminished by anisotropic scaling. The statistics of the anisotropically scaled dataset are shown in Table 1. The structure was then refined to a resolution of 3.1 Å against the anisotropically scaled data using the program REFMAC (25). This yielded maps of very good quality including for the anticodon region (Figure 1 and Supplementary Figure 1). The final model shows a *R*-factor of 24.1% and a R_{free} factor of 27.0% (Table 1). Data from crystals grown under basic conditions were processed similarly. The corresponding structure was refined to a resolution of 3.3 Å (R = 23.1%; R_{free} = 27.4%). Otherwise stated, we will describe the $tRNA_f^{Met}$ structure obtained in the acidic conditions that is solved at a higher resolution.

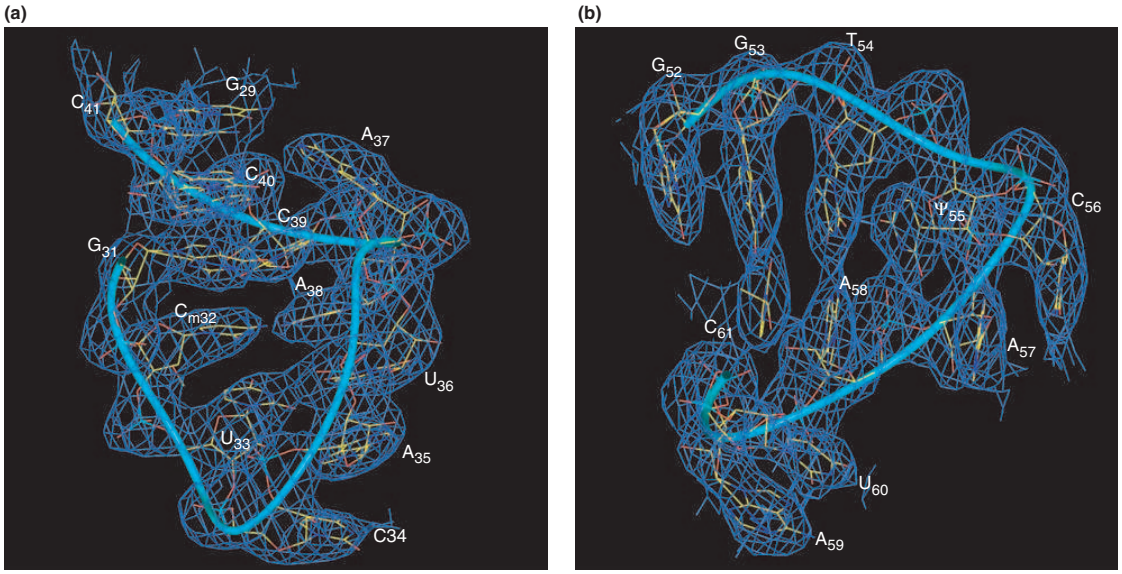


Figure 1. Electron density maps around anticodon and TΨC arms of $tRNA_f^{Met}$ structure in the acidic conditions. (**a**) The anticodon arm showing the A37 base triple interaction with G29-C41. (**b**) The TΨC arm showing well-defined base and phosphate position. The electron density map corresponds to a 2Fo–Fc contoured at 1σ.

Overall model and crystal packing

The overall tertiary structure of tRNA$_f^{Met}$ matches the well-known L-shape structure first revealed in yeast tRNA^{Phe} (27) with usual tertiary interactions (Figure 2): the reverse-Hoogsteen base pairs s^{4}U8•A14 and T54•A58, the *trans* Watson–Crick G15•C48 ('Levitt pair'), the one bond imino-4-carbonyl G18•Ψ55, the Watson–Crick base pair G19–C56 forming the corner of the L, the Watson–Crick like G26•A44, the 'U-turn' at U33 in the anticodon loop and at Ψ55 in the T loop, and finally the base triples A46•(G22–C13) and G45•(G10–C25). The base triple equivalent to the A9•(A23–U12) found in tRNA^{Phe}, i.e. G9•(C23–G12) was not formed in tRNA$_f^{Met}$ where N7-G9 is at 5.2 Å away from N4-C23.

The crystal packing is quite remarkable. First, the acceptor arm of the tRNA molecule interacts with the major groove of the acceptor stem of a symmetry-related tRNA molecule, thereby forming a triple helix (Figure 3a and b). Namely, nt 72–76 of one tRNA molecule are base-paired in an antiparallel manner to G2-G6 of the acceptor stem of another tRNA molecule. The C1 nucleotide is very mobile and is not visible in electron density maps. However the triple helix formation necessarily rejects C1 in the solvent with G2 rather stacks on A72. Therefore, the always mismatched 1 and 72 bases, characteristic of bacterial initiator tRNAs, are not face-to-face. Another packing region involves anticodon–anticodon base pairing (Figure 3c and d). Anticodon loops interact through pairing between $C_{34}A_{35}U_{36}A_{38}$ and the four same bases in a symmetry-related tRNA molecule, thereby forming two Watson–Crick AU pairs and two non-canonical CA pairs. The same packing interaction at the level of the anticodon was previously observed in the *E. coli* tRNA$_f^{Met}$/formylase complex (PDB entry code 2FMT) and in the yeast initiator tRNA crystal structure (PDB entry code 1YFG). In tRNA^{Asp}, the crystal packing also involves anticodon–anticodon interactions (28). We cannot exclude that the crystal packing at the level of the anticodon loop, in particular contacts between C34 and A38, could contribute to the stabilization of the anticodon loop conformation, in particular stabilisation of the non-canonical Cm32•A38 base pair described in the following paragraph.

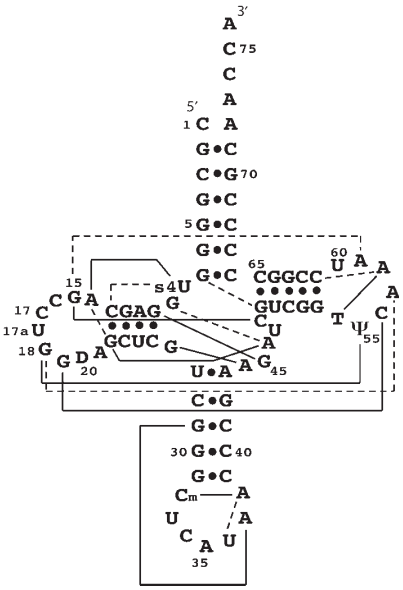


Figure 2. Tertiary interactions on the secondary structure of *E. coli* tRNA$_f^{Met}$. Solid lines represent base–base tertiary interaction through hydrogen bonds, and dashed lines represent base stacking interactions.

Although the overall structure globally resembles that of elongator tRNAs, there are however striking features at the level of the anticodon arm that deserve discussion.

The anticodon stem-loop of initiator tRNA adopts a non-canonical conformation

The structure of the anticodon stem-loop region can be described by three specific features (Figure 4): (i) an additional Cm32•A38 unusual base pair extending the anticodon stem (Figure 4c). (ii) Formation of a base triple between A37 in the anticodon loop and G29-C41 in the stem (Figure 4d). This induces a large turn in the phosphate backbone just after the anticodon triplet. (iii) An unusual stacking pattern for the anticodon loop: A38 is directly stacked onto the last base of the anticodon (U36), whereas it is stacked on nt 37 in all other tRNA structures (Figure 4a and b). Overall, this induces a more compact and constrained geometry, stabilized by additional specific interactions involving the first and the last initiator-characteristic GC pairs. The Cm32•A38 wobble-like base pair, not encountered in other tRNA structures (29), is stabilized by stacking onto G31-C39 and by a crystal contact between A38 and C34. It involves a polar contact between N1(A38) and O2(Cm32) in addition to the N3(Cm32)–N6(A38) hydrogen bond and suggests that the N1 group of A38 is protonated (Figure 4c). The particular Cm32•A38 wobble-like base pair is probably formed because of the acidic pH of the crystallization conditions. Unfortunately, the lower resolution of the model refined from crystals grown at pH 8.0 did not allow us to confirm this proposal, since the prediction on the protonation state from the structure depends on small variations in base pair orientation that could not be strongly reliable at this resolution. In the structure of the *E. coli* tRNA$_f^{Met}$/formylase complex (30), which was determined using crystals grown at pH 6.6, the Cm32•A38 base pair is less tight than in the present structure. However, the two bases are also facing each other. Such an (A+)•C base pair was already shown to occur, with a pKa of about 6.0, in the NMR structures of hypomodified tRNA$_3^{Lys}$ anticodon stem-loop (31). In tRNA$_f^{Met}$, as a consequence of the (A+)•C pairing, the anticodon stem extends to Cm32•A38.

Additionally, A37 participates in a base triple within the major groove of the anticodon stem. The Watson–Crick edge of A37 interacts with the Hoogsteen edge of G29 and is stabilized by an hydrogen bond between A37 O4′ and C39 N4 (Figure 2d). This triple interaction is probably strengthened by protonation of A37 N1. The A37•(G29-C41) interaction has never been described and

4898 *Nucleic Acids Research, 2008, Vol. 36, No. 15*

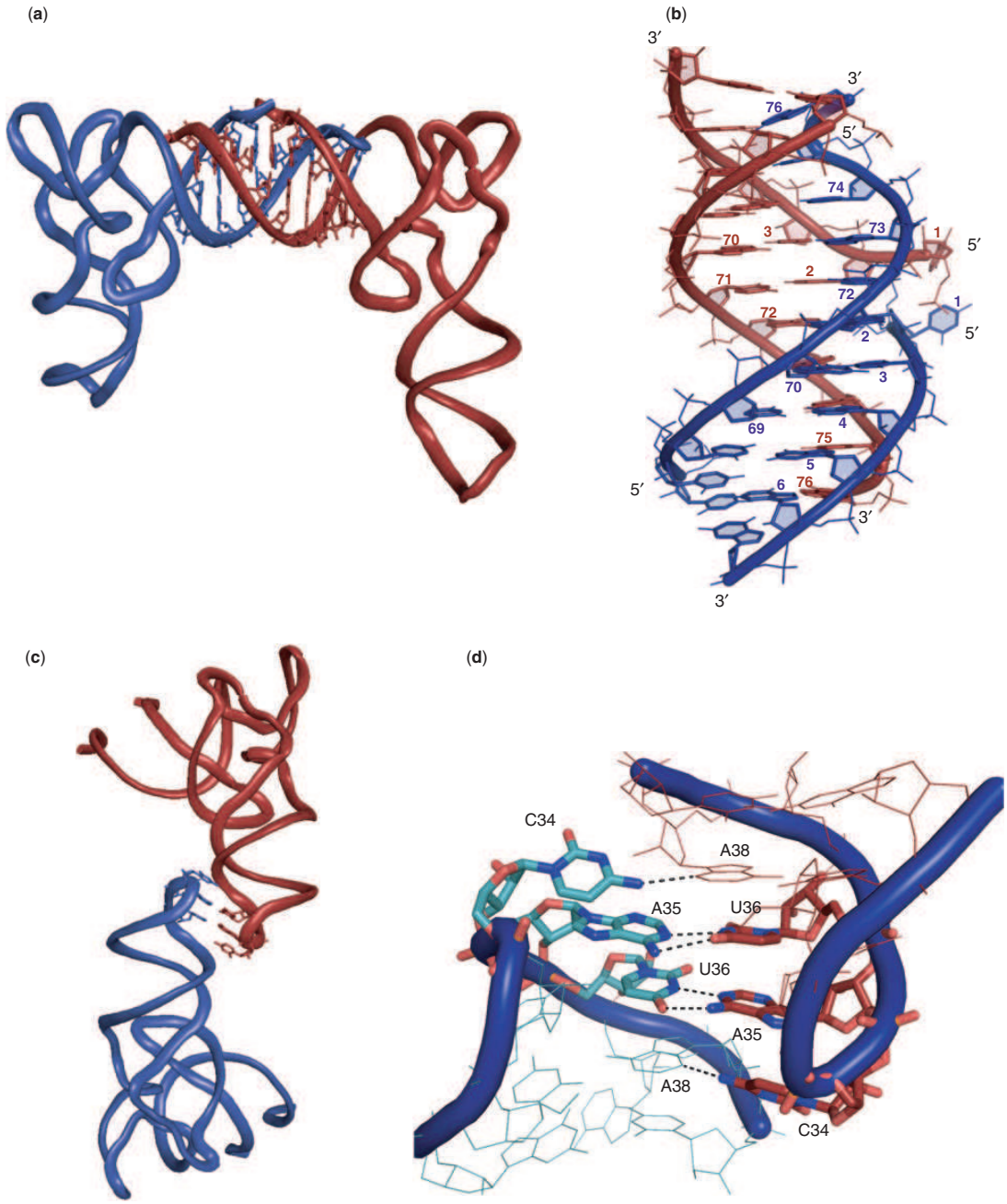


Figure 3. Packing interaction in the crystal structure of $tRNA_f^{Met}$. (**a**) and (**b**) in the acceptor stem, (**c**) and (**d**) in the anticodon loop.

is clearly observed in the present structures obtained from crystals at two pH conditions (see Supplementary Figure 2 for the basic structure). As a consequence, the major groove of the anticodon stem is severely obstructed at the level of the three GC pairs. This could explain the previously reported unreactivity of the N7 groups of G29, G30, G31 in the major groove of *E. coli* initiator tRNA anticodon stem (32). In particular, the N7 group of G30 which is reactive in all elongator tRNAs does not react with dimethyl-sulfate in $tRNA_f^{Met}$. In conclusion, two features are undoubtedly conserved in the $tRNA_f^{Met}$ structures solved at two pH conditions: the peculiar stacking of bases in the anticodon loop and the base triple made by A37.

Nucleic Acids Research, 2008, Vol. 36, No. 15 **4899**

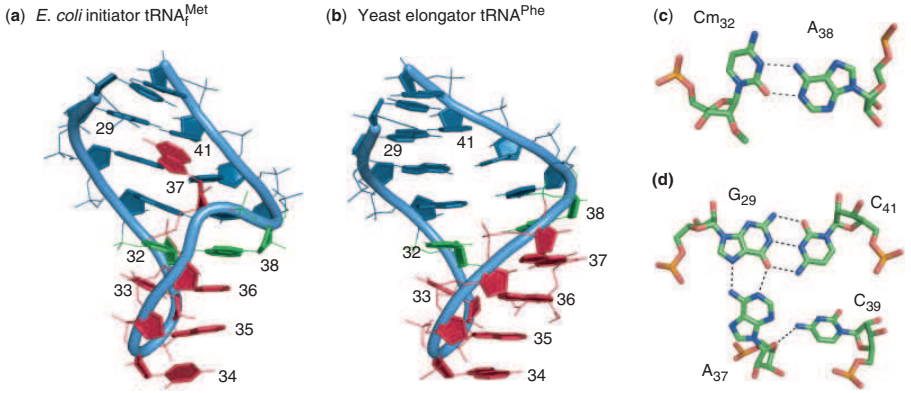


Figure 4. Unique conformation of the anticodon arm of the *E. coli* initiator tRNA$_f^{Met}$: (**a**) the anticodon arm of tRNA$_f^{Met}$, (**b**) for comparison tRNA^{Phe} anticodon arm (PDB ID 1EHZ), (**c**) the Cm32•A38 wobble-like base pair observed in crystal structure of tRNA$_f^{Met}$ and (**d**) the A37• (G29-C41) base triple observed in crystal structure of tRNA$_f^{Met}$.

E. coli initiator tRNA was previously crystallized in complex with methionyl–tRNA$_f^{Met}$ transformylase (30). The anticodon stem-loop region does not interact with the protein. In this structure, A37, although not involved in a base-triple with G29-C41, also fails to stack between U36 and A38 (Figure 3a and b), in contrast to what is observed in all available elongator tRNA structures (27,33–38). Differences observed for the position of A37 in the present structure and in the tRNA$_f^{Met}$/formylase complex structure are in favor of its mobility. In this context, it should be noted that in the structure of tRNA$_f^{Met}$/formylase complex the position of the adenine ring of base 37 was only tentatively modelled (30). Increased mobility of A37 in the tRNA$_f^{Met}$/formylase complex is also supported by an increased RNase sensitivity at phosphate G29, G30 upon binding of tRNA$_f^{Met}$ to the formylase (39). Finally, although in yeast initiator tRNA structure the anticodon region is poorly defined (6), a similar organization of the anticodon bases was observed, with U36 stacked on A38 and A37 rejected in the solvent. Therefore, in available X-ray structures, initiator tRNAs appear consistently different from elongator tRNAs at the level of the base stacking of the anticodon loop. This corroborates their unique S1 nuclease cleavage pattern (2,3,40). Indeed, all initiator tRNAs (*E. coli*, yeast and mammalian) are cleaved after C34 and A35 (2), while the anticodon loops of elongator tRNAs, including elongator tRNA^{Met}, are much more sensitive, being cleaved after U33, C34, A35 and U36 (2,40). Interestingly, in yeast, the anticodon loop sequences of initiator and elongator tRNA^{Met} are identical, including for the post-transcriptional base modification. These observations led to the proposal that the three conserved GC base pairs of the initiator anticodon stem were the key determinants for an unusual conformation of the anticodon loop (2–4). This hypothesis is in full agreement with the structure reported here.

Relevance for initiator tRNA discrimination

Mutagenesis of the anticodon or stem-loop showed that the S1 cleavage pattern correlates with the initiator function and selection by IF3 (3,4). This strongly suggests that the peculiar conformation of the anticodon loop of initiator tRNAs is required for their initiator function. This raises the question of the detailed underlying mechanism. In the structure of the *Thermus thermophilus* 70S ribosome, tRNA$_f^{Met}$ is paired with the initiation codon in the P-site (9), and the canonical conformation of the anticodon loop is restored, with A37 stacked between U36 and A38 (Figure 3c). Such a conformation was also seen in the structure of *E. coli* ribosome with a tRNA$_f^{Met}$ anticodon stem-loop paired to mRNA in the P-site (41). This conformation is stabilized by a number of contacts within the P-site. Among these, universally conserved bases G1338 and A1339 make types I and II A-minor interactions with G29-C41 and G30-C40, respectively (9–11). Although the GC pairs are crucial for IF3-dependent tRNA$_f^{Met}$ discrimination, they are too far from its 30S subunit binding site to be directly inspected by IF3, according to the current understanding of the position of IF3 (42). This led these authors to propose that under the influence of IF3 binding, G1338 and A1339 might check the identity of tRNA$_f^{Met}$ through their minor groove interactions with these conserved GC. By mutational studies, these interactions were actually shown to play a role in the IF3-dependent tRNA$_f^{Met}$ discrimination (14,15). However, tRNAs that contain two consecutive GC pairs at positions 29–41 and 30–40 are not resistant to rejection by IF3, indicating that discrimination must involve recognition of at least one additional feature of the tRNA$_f^{Met}$ anticodon stem-loop (15). The singular structure of the anticodon stem-loop that we report here strongly supports this idea.

Relevance for translation initiation

Clearly, the '37-unstacked' conformation observed in the anticodon loop of free initiator tRNA is markedly different from the '37-stacked' observed in P-site-bound tRNA$_f^{Met}$. Therefore, the available data suggests that an initiator tRNA must switch between the two conformations during the initiation process. An attractive

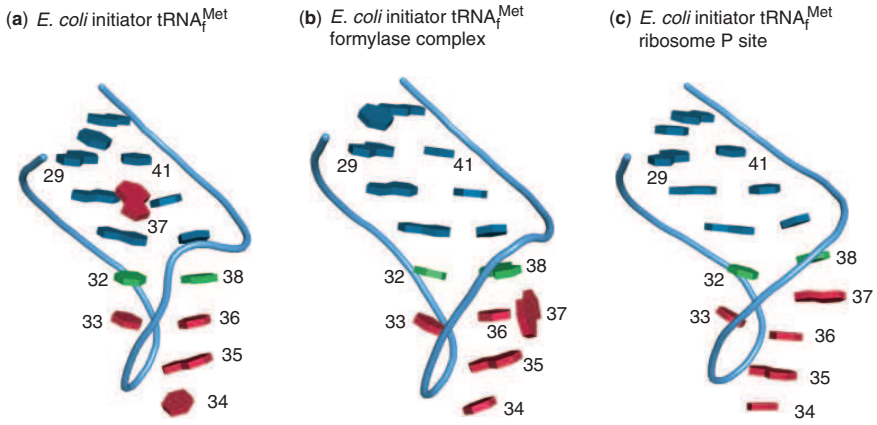


Figure 5. Comparison of the anticodon loop base stacking. (**a**) *E. coli* tRNA$_f^{Met}$, (**b**) *E. coli* tRNA$_f^{Met}$ in the complex with formylase (PDB ID 2FMT), (**c**) in the P-site (PDB ID 2J00). The '37-unstacked' conformation refers to (a) and (b) structures and the '37-stacked' to the (c) one.

hypothesis would be that a '37-unstacked' conformation is required at the first stages of initiator tRNA accommodation by the ribosomal P-site and IF3. Recognizing a '37-unstacked' conformation would help to reject elongator tRNAs (Figure 5). In a second step, stabilization of the '37-stacked' conformation would require correct pairing with the AUG initiation codon. Such a conformational transition could be used as a 'sensor' that a correct initiation codon has been found. Interestingly, in the present structure as well as in that of the formylase/tRNA$_f^{Met}$ complex and in that of yeast initiator tRNA, crystal packing involves only two Watson–Crick base pairs of the anticodon (A35 and U36; see Supplementary Figure 2). In the context of the above hypothesis, such a structure may mimic pairing with an incorrect start codon, unable to trigger a '37-stacked' conformation of the anticodon loop.

SUPPLEMENTARY DATA

Supplementary Data are available on NAR Online.

ACKNOWLEDGEMENTS

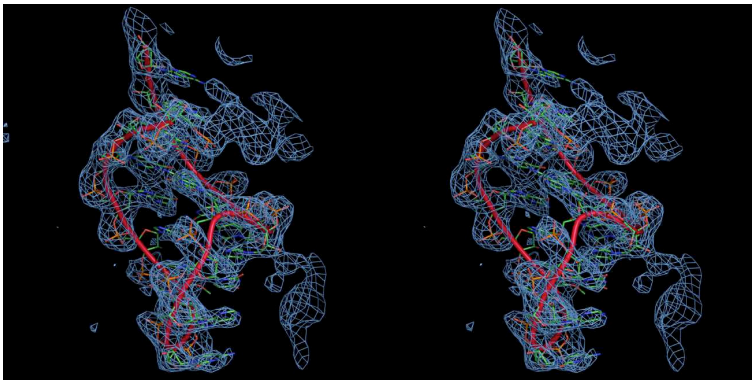
The authors acknowledge the European Synchrotron Radiation Facility for providing the synchrotron radiation facilities on beamlines ID14-3 and ID29. P.B. was supported by a studentship from the 'Ministère de la recherche'. We thank the CNRS for financial support. Funding to pay the Open Access publication charges for this article was provided by CNRS, French National Agency of Research.

Conflict of interest statement. None declared.

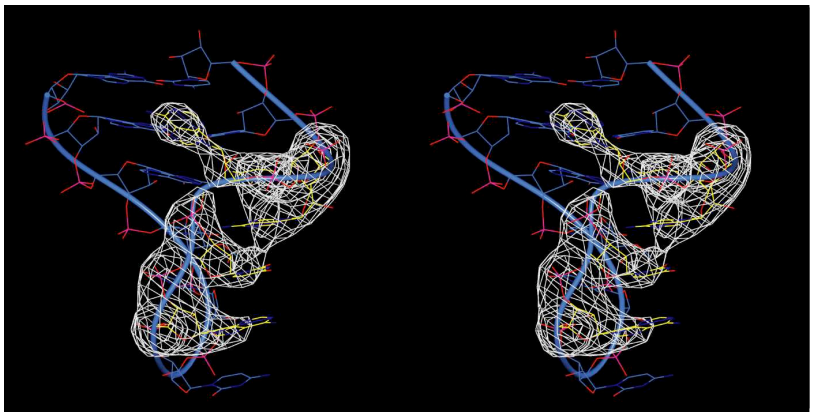
REFERENCES

1. Marck,C. and Grosjean,H. (2002) tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA*, **8**, 1189–1232.
2. Wrede,P., Woo,N.H. and Rich,A. (1979) Initiator tRNAs have a unique anticodon loop conformation. *Proc. Natl Acad. Sci. USA*, **76**, 3289–3293.
3. Seong,B.L. and RajBhandary,U.L. (1987) Escherichia coli formylmethionine tRNA: mutations in GGGCCC sequence conserved in anticodon stem of initiator tRNAs affect initiation of protein synthesis and conformation of anticodon loop. *Proc. Natl Acad. Sci. USA*, **84**, 334–338.
4. Hartz,D., Binkley,J., Hollingsworth,T. and Gold,L. (1990) Domains of initiator tRNA and initiation codon crucial for initiator tRNA selection by Escherichia coli IF3. *Genes Dev.*, **4**, 1790–1800.
5. Varshney,U., Lee,C.P. and RajBhandary,U.L. (1993) From elongator tRNA to initiator tRNA. *Proc. Natl Acad. Sci. USA*, **90**, 2305–2309.
6. Basavappa,R. and Sigler,P.B. (1991) The 3 A crystal structure of yeast initiator tRNA: functional implications in initiator/elongator discrimination. *EMBO J.*, **10**, 3105–3111.
7. Woo,N.H., Roe,B.A. and Rich,A. (1980) Three-dimensional structure of *Escherichia coli* initiator tRNAfMet. *Nature*, **286**, 346–351.
8. Schweisguth,D.C. and Moore,P.B. (1997) On the conformation of the anticodon loops of initiator and elongator methionine tRNAs. *J. Mol. Biol.*, **267**, 505–519.
9. Selmer,M., Dunham,C.M., Murphy,F.V.t., Weixlbaumer,A., Petry,S., Kelley,A.C., Weir,J.R. and Ramakrishnan,V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.
10. Korostelev,A., Trakhanov,S., Laurberg,M. and Noller,H.F. (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell*, **126**, 1065–1077.
11. Yusupov,M.M., Yusupova,G.Z., Baucom,A., Lieberman,K., Earnest,T.N., Cate,J.H. and Noller,H.F. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science*, **292**, 883–896.
12. Korostelev,A. and Noller,H.F. (2007) The ribosome in focus: new structures bring new insights. *Trends Biochem. Sci.*, **32**, 434–441.
13. Nissen,P., Ippolito,J.A., Ban,N., Moore,P.B. and Steitz,T.A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA*, **98**, 4899–4903.
14. Abdi,N.M. and Fredrick,K. (2005) Contribution of 16S rRNA nucleotides forming the 30S subunit A and P sites to translation in Escherichia coli. *RNA*, **11**, 1624–1632.
15. Lancaster,L. and Noller,H.F. (2005) Involvement of 16S rRNA nucleotides G1338 and A1339 in discrimination of initiator tRNA. *Mol. Cell*, **20**, 623–632.
16. Mangroo,D., Limbach,P.A., McCloskey,J.A. and RajBhandary,U.L. (1995) An anticodon sequence mutant of Escherichia coli initiator tRNA: possible importance of a newly

acquired base modification next to the anticodon on its activity in initiation. *J. Bacteriol.*, **177**, 2858–2862.
17. Meinnel,T., Mechulam,Y. and Fayat,G. (1988) Fast purification of a functional elongator tRNAMet expressed from a synthetic gene in vivo. *Nucleic Acids Res.*, **16**, 8095–8096.
18. Collaborative Computational Project Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
19. Potterton,E., Briggs,P., Turkenburg,M. and Dodson,E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D*, **59**, 1131–1137.
20. Leslie,A.G. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr. D*, **62**, 48–57.
21. Evans,P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D*, **62**, 72–82.
22. McCoy,A.J., Grosse-Kunstleve,R.W., Storoni,L.C. and Read,R.J. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr.*, **61**, 458–464.
23. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*, **60**, 2126–2132.
24. Winn,M.D., Isupov,M.N. and Murshudov,G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D*, **57**, 122–133.
25. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D*, **53**, 240–255.
26. Strong,M., Sawaya,M.R., Wang,S., Phillips,M., Cascio,D. and Eisenberg,D. (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl Acad. Sci. USA*, **103**, 8060–8065.
27. Kim,S.H., Quigley,G.J., Suddath,F.L., McPherson,A., Sneden,D., Kim,J.J., Weinzierl,J. and Rich,A. (1973) Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science*, **179**, 285–288.
28. Moras,D., Dock,A.C., Dumas,P., Westhof,E., Romby,P., Ebel,J.P. and Giege,R. (1986) Anticodon-anticodon interaction induces conformational changes in tRNA: yeast tRNAAsp, a model for tRNA-mRNA recognition. *Proc. Natl Acad. Sci. USA*, **83**, 932–936.
29. Auffinger,P. and Westhof,E. (1999) Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes. *J. Mol. Biol.*, **292**, 467–483.
30. Schmitt,E., Panvert,M., Blanquet,S. and Mechulam,Y. (1998) Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet. *EMBO J.*, **17**, 6819–6826.
31. Durant,P.C. and Davis,D.R. (1999) Stabilization of the anticodon stem-loop of tRNALys,3 by an A+-C base-pair and by pseudouridine. *J. Mol. Biol.*, **285**, 115–131.
32. Wakao,H., Romby,P., Westhof,E., Laalami,S., Grunberg-Manago,M., Ebel,J.P., Ehresmann,C. and Ehresmann,B. (1989) The solution structure of the Escherichia coli initiator tRNA and its interactions with initiation factor 2 and the ribosomal 30 S subunit. *J. Biol. Chem.*, **264**, 20363–20371.
33. Robertus,J.D., Ladner,J.E., Finch,J.T., Rhodes,D., Brown,R.S., Clark,B.F. and Klug,A. (1974) Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature*, **250**, 546–551.
34. Suddath,F.L., Quigley,G.J., McPherson,A., Sneden,D., Kim,J.J., Kim,S.H. and Rich,A. (1974) Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0angstroms resolution. *Nature*, **248**, 20–24.
35. Jovine,L., Djordjevic,S. and Rhodes,D. (2000) The crystal structure of yeast phenylalanine tRNA at 2.0 A resolution: cleavage by Mg(2+) in 15-year old crystals. *J. Mol. Biol.*, **301**, 401–414.
36. Shi,H. and Moore,P.B. (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA*, **6**, 1091–1105.
37. Westhof,E., Dumas,P. and Moras,D. (1985) Crystallographic refinement of yeast aspartic acid transfer RNA. *J. Mol. Biol.*, **184**, 119–145.
38. Benas,P., Bec,G., Keith,G., Marquet,R., Ehresmann,C., Ehresmann,B. and Dumas,P. (2000) The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop. *RNA*, **6**, 1347–1355.
39. Ramesh,V., Mayer,C., Dyson,M.R., Gite,S. and RajBhandary,U.L. (1999) Induced fit of a peptide loop of methionyl-tRNA formyltransferase triggered by the initiator tRNA substrate. *Proc. Natl Acad. Sci. USA*, **96**, 875–880.
40. Wrede,P. and Rich,A. (1979) Stability of the unique anticodon loop conformation of E.coli tRNAfMet. *Nucleic Acids Res.*, **7**, 1457–1467.
41. Berk,V., Zhang,W., Pai,R.D. and Cate,J.H. (2006) Structural basis for mRNA and tRNA positioning on the ribosome. *Proc. Natl Acad. Sci. USA*, **103**, 15830–15834.
42. Dallas,A. and Noller,H.F. (2001) Interaction of translation initiation factor 3 with the 30S ribosomal subunit. *Mol. Cell*, **8**, 855–864.



Supplementary Figure 1: Electron density map around the anticodon region of the $tRNA_f^{Met}$ structure obtained at pH 4.6 showing unambigously the A37•(G29-C41) base triple. The map is a simulated annealing composite omit map ($2F_O$-F_C) contoured at 1 σ calculated with CNS.



Supplementary Figure 2: Electron density map of the anticodon region of the $tRNA_f^{Met}$ structure obtained at pH 8.0. Molecular replacement for the basic structure was performed using the acidic refined structure in which the coordinates around base A37 were omitted (*i.e.* A35 to C39) in yellow on the figure. After refinement of the Molecular replacement solution, the F_O-F_C map contoured at 3 σ is displayed around the anticodon stem-loop. It shows unambiguously the same stacking pattern and the same position of A37 as for the acidic structure.

Chapitre 11

Conclusion et Perspectives

La conformation particulière de la tige-boucle anticodon de l'ARNt$_f^{Met}$ que nous avons mise en évidence dans cette étude, est vraisemblablement impliquée dans la sélection IF3-dépendante de l'ARNt$_f^{Met}$ au sein du site P. Afin de valider cette hypothèse, il faudrait pouvoir obtenir la structure à haute résolution de la sous-unité 30S en phase de préinitiation, avec l'ARNm, l'ARNt$_f^{Met}$ au site P et le facteur IF3. Malheureusement, ce complexe est relativement instable, et il semble ainsi délicat d'en obtenir la structure. Aussi, l'obtention de la structure de la sous-unité 30S en présence seulement d'IF3, pourrait permettre — par une approche d'amarrage moléculaire — de comprendre comment l'ARNt$_f^{Met}$ dans sa conformation « 37–non-empilée » pourrait venir interagir au niveau du site P de la sous-unité 30S, soit avec le ribosome lui-même soit avec le facteur d'initiation IF3. Un tel modèle pourrait permettre de comprendre les bases moléculaires de la discrimination IF3-dépendante de l'ARNt initiateur au sein du site P du ribosome.

Appendices

Liste des figures

1.1 Structure primaire des ARN 4
1.2 Nomenclature des cycles puriques et pyrimidiques 5
1.3 Structure secondaire et numérotation des ARNt 6
1.4 Structure tridimensionnelle de l'ARNt^{Phe} de levure 7
1.5 Représentation schématique de la structure 3D d'un ARNt 8
1.6 ARNt fixés aux sites A, P et E du ribosome 10
1.7 Vue générale du ribosome bactérien 11
1.8 Les étapes de la maturation des ARNt 14
1.9 Maturation des extrémités 5′ et 3′ des ARNt 15
1.10 Les nucléotides modifiés dans les différents type d'ARN 17
1.11 Distribution phylogénétique des nucléotides modifiés dans les ARNt 17
1.12 Localisation des nucléotides modifiés dans les ARNt 19
1.13 Le code génétique 21

2.1 Organisation des sous-domaines de la RT du VIH-1 32
2.2 Organisation des différents gènes du génome du VIH-1 33
2.3 Organisation schématique de l'ARN génomique du VIH-1 34
2.4 Structure secondaire de la partie 5' de l'ARN viral 35
2.5 Structure secondaire de l'ARNt$_3^{Lys}$ 36
2.6 Séquence de la NC du VIH-1 37
2.7 Structures RMN de la NC$_{1\text{-}55}$ du VIH-1 38
2.8 Comparaison des structures NC$_{1\text{-}55}$-SL2 et NC$_{1\text{-}55}$-SL3 39
2.9 Déroulement de la transcription inverse du VIH-1 42
2.10 Structure de l'hybride entre l'ARNt$_3^{Lys}$ et le PBS 44
2.11 Mécanisme de l'hybridation entre l'ARNt$_3^{Lys}$ et le PBS 46
2.12 Modèles de structure secondaire des isolats MAL et HXB2 49
2.13 Modèle de l'interaction PAS/anti-PAS 51

4.1 Bilan d'attribution de la NC$_{1\text{-}55}$ 64
4.2 Optimisation de la longueur du bras D pour l'étude RMN 66
4.3 Spectres HSQC de la NC ^{15}N libre et en complexe avec le Bras D24 67
4.4 Résidus de la NC impliqués dans l'interaction avec le bras D 71
4.5 Expérience de titration et attribution de la NC 78
4.6 Bilan d'attribution grâce aux expériences de titration 79
4.7 Bilan d'attribution grâce aux expériences triple résonances 80
4.8 Bilan d'attribution grâce aux expériences triples résonances 81
4.9 Comparaison des séquences Bras D24 et Bras D22 82

4.10 Transcription *in vitro* et purification du bras D par FPLC. 83
4.11 Spectres de corrélation ^{1}H–^{13}C du bras D marqué ^{13}C/^{15}N 84

5.1 Rôle de la NC dans l'hybridation entre l'ARNt$_3^{Lys}$ et le PBS 86
5.2 Guanines de la boucle D potentiellement reconnues par la NC . . . 87

6.1 Structure chimique de la S-adénosyl-L-méthionine 92
6.2 Le cycle de la S-adénosyl-L-méthionine 93
6.3 Organisation tridimensionnelle des MTases de Classe I 95
6.4 Organisation tridimensionnelle des MTases de Classe II–V 98
6.5 Localisation des motifs conservés parmi les MTases de Classe I . . . 99
6.6 Le nucléotide m^{1}A dans les ARNt 103
6.7 Alignement de séquences de la famille m^{1}A$_{58}$ MTases 105
6.8 Voie de biosynthèse du nucléotide m^{1}I$_{57}$ chez les *Archaea* 106
6.9 Structure particulière de l'ARNt$_i^{Met}$ initiateur de levure 108

7.1 Forme « λ » de l'ARNt 111
7.2 Détermination des paramètres cinétiques k_{cat} et K_M 131
7.3 Mécanismes possibles pour la formation du m^{1}A 132

8.1 Les deux sillons pour la fixation de l'ARNt 136
8.2 Modèle d'amarrage manuel d'un ARNt entier 137

9.1 L'initiation de la traduction chez les bactéries 143
9.2 Caractéristiques des ARNt initiateurs de bactéries 145
9.3 Sites de clivage à la nucléase S1 146
9.4 Interactions au sein du site P du ribosome 149
9.5 Interaction de type A-mineur avec la tige anticodon de l'ARNt$_f^{Met}$. 150

10.1 Structure unique de l'anticodon de l'ARNt$_f^{Met}$ 156

A.1 Les faces d'interactions des nucléotides 216
A.2 Les 12 familles de paires de bases et leur notation symbolique 217
B.1 Etapes de purification de la NC$_{1-55}$ 221
B.2 Etapes de purification de TrmI 222

Liste des tables

1.1 Diversité et proportion de nucléotides modifiés dans les ARN 16

4.1 Paramètres de liaison de la NC avec diverses tiges-boucles Bras D . 65

Liste des abréviations

aa	acide aminé
aaRS	aminoacyl-ARNt synthétase
AdoMet	S-adénosyl-L-méthionine (ou SAM)
ADN	acide désoxyribonucléique
ARN	acide ribonucléique
ARNm	ARN messager
ARNr	ARN ribosomal
ARNt	ARN de transfert
$ARNt_3^{Lys}$	ARNt lysine isoaccepteur 3
$ARNt_f^{Met}$	ARNt initiateur de *Bacteria*
$ARNt_i^{Met}$	ARNt initiateur d'*Eukarya* ou d'*Archaea*
ATP/ADP	adénosine tri/diphosphate
βME	2-mercaptoéthanol
CPMG	Séquence de Carr, Purcell, Meiboom et Gill
CT	constant time
Da	Dalton
DO	densité optique
$\Delta\omega$	différence de déplacement chimique
DTT	dithiothréitol
EDTA	acide éthylène-diamine tétra-acétique
EF1A	facteur d'élongation bactérien 1A
FPLC	Fast Protein Liquid Chromatography
GTP/GDP	guanosine tri/diphosphate
HSQC	Heteronuclear Single Quantum Coherence
INEPT	Insensitive Nucleus Enhancement by Polarisation Transfer
K_d	constante de dissociation
k_{ex}	constante d'échange
LB	Luria Bertani
LTR	Long Terminal Repeat
IF1, IF2, IF3	facteurs d'initiation de la traduction bactérienne
IPTG	isopropyl-β-D-thiogalactoside
m^1A	1-méthyladénosine
MTase	méthyltransférase

NC	nucléocapside
nOe	nuclear Overhauser effect
NOESY	Nuclear Overhauser Effect SpectroscopY
nt	nucléotides
PBS	Primer Binding Site
PAS	Primer Activation Signal
PDB	Protein Data Bank
Pi	phosphate inorganique
K_a	constante d'acidité
PMSF	phenylmethylsulfonyl fluoride
ppm	partie par million
Ψ	pseudouridine ou signal d'encapsidation, selon contexte
RMN	Résonance Magnétique Nucléaire
RT	Transcriptase Inverse
SAH	S-adénosyl-L-homocystéine
SAM	S-adénosyl-L-méthionine (ou AdoMet)
SDS	sodium dodécyl sulfate
Tris	Tris(hydroxyméthyl)aminométhane
TROSY	Transverse Relaxation Optimized SpectroscopY
u.a.	unité d'abosorbance
30S	petite sous-unité du ribosome bactérien
50S	grande sous-unité du ribosome bactérien
70S	ribosome bactérien
16S	ARNr de la petite sous-unité du ribosome bactérien

Liste des publications

New insights into the formation of HIV-1 reverse transcription initiation complex 54
Pierre Barraud, Cyril Gaudin, Frédéric Dardel & Carine Tisné
Biochimie 89 (**2007**) 1204–1210

Abstract

HIV-1 reverse transcriptase uses the host $tRNA_3^{Lys}$ as a primer for the synthesis of the minus DNA strand. The first event in viral replication is thus the annealing of tRNA to the primer binding site (PBS) in the 5' UTR of the viral RNA. This event requires a major RNA rearrangement which is chaperoned by the viral NC protein. The binding of NC to nucleic acids is essentially non-specific, however, NC is known to bind selectively to hairpins located in the 5' region of the viral RNA. In a previous study, using an NMR approach in which the reaction is slowed down by controlling temperature, we were able to follow details in this RNA unfolding/refolding process and to uncover an intermediate state. We showed that annealing initiates at the junction between the acceptor and the TΨC stems, and that, at physiological temperature, complete annealing is reached only in the presence of NC, probably when the zinc fingers contact the TΨC/D loops. In the present work, we have refined our model of the formation of the $tRNA_3^{Lys}$/PBS duplex. First, we show that annealing can initiate both from the single-stranded CCA 3'-end bases of the acceptor stem and from the bases in the TΨC stem. Secondly, by NMR and fluorescence spectroscopy, we have studied the complex between the NC protein and RNA hairpins that mimic the D and T arms of the $tRNA_3^{Lys}$. Interestingly, the NC protein shows strong and specific binding to the D arm of $tRNA_3^{Lys}$, which could explain the overall annealing mechanism.

Optimizing HSQC experiment for the observation of exchange broadened signals in RNA–protein complexes 71
Pierre Barraud, Frédéric Dardel & Carine Tisné
Comptes Rendus Chimie 11 (**2008**) 474–479

Abstract

Sites for interaction in protein–RNA complexes are often regions of conformational exchange. Although the study of exchange processes could bring valuable information about the recognition mode between protein and RNA, chemical exchange can be detrimental to the NMR spectra quality, resulting in broad, very weak or even unobservable signals. In the present report, we used CPMG-like experiments to improve HSQC spectra of an RNA–protein complex in fast exchange on the chemical shift time scale. The use of such improvement will allow us to handle the resolution of the three-dimensional structure of the complex by NMR.

Crystal structure of *Thermus thermophilus* tRNA m^1A_{58} methyltransferase and biophysical characterization of its interaction with tRNA 114
Pierre Barraud, Béatrice Golinelli-Pimpaneau, Cédric Atmanene, Sarah Sanglier, Alain Van Dorsselaer, Louis Droogmans, Frédéric Dardel & Carine Tisné
Journal of Molecular Biology 377 (**2008**) 535–550

Abstract

Methyltransferases from the m^1A_{58} tRNA methyltransferase (TrmI) family catalyze the S-adenosyl-L-methionine-dependent N_1-methylation of tRNA adenosine 58. The crystal structure of *Thermus thermophilus* TrmI, in complex with S-adenosyl-L-homocysteine, was determined at 1.7 Å resolution. This structure is closely related to that of *Mycobacterium tuberculosis* TrmI, and their comparison enabled us to enlighten two grooves in the TrmI structure that are large enough and electrostatically compatible to accommodate one tRNA per face of TrmI tetramer. We have then conducted a biophysical study based on electrospray ionization mass spectrometry, site-directed mutagenesis, and molecular docking. First, we confirmed the tetrameric oligomerization state of TrmI, and we showed that this protein remains tetrameric upon tRNA binding, with formation of complexes involving one to two molecules of tRNA per TrmI tetramer. Second, three key residues for the methylation reaction were identified : the universally conserved D170 and two conserved aromatic residues Y78 and Y194. We then used molecular docking to position a N_9-methyladenine in the active site of TrmI. The N_9-methyladenine snugly fits into the catalytic cleft, where the side chain of D170 acts as a bidentate ligand binding the

amino moiety of S-adenosyl-L-methionine and the exocyclic amino group of the adenosine. Y194 interacts with the N_9-methyladenine ring, whereas Y78 can stabilize the sugar ring. From our results, we propose that the conserved residues that form the catalytic cavity (D170, Y78, and Y194) are essential for fashioning an optimized shape of the catalytic pocket.

A unique conformation of the anticodon-stem loop is associated with the capacity of $tRNA_f^{Met}$ to initiate protein synthesis 156
Pierre Barraud, Emmanuelle Schmitt, Yves Mechulam, Frédéric Dardel & Carine Tisné
Nucleic Acids Research 36 (**2008**) 4894–4901

Abstract

In all organisms, translational initiation takes place on the small ribosomal subunit and two classes of methionine tRNA are present. The initiator is used exclusively for initiation of protein synthesis while the elongator is used for inserting methionine internally in the nascent polypeptide chain. The crystal structure of Escherichia coli initiator $tRNA_f^{Met}$ has been solved at 3.1 Å resolution. The anticodon region is well-defined and reveals a unique structure, which has not been described in any other tRNA. It encompasses a Cm32•A38 base pair with a peculiar geometry extending the anticodon helix, a base triple between A37 and the G29-C41 pair in the major groove of the anticodon stem and a modified stacking organization of the anticodon loop. This conformation is associated with the three GC basepairs in the anticodon stem, characteristic of initiator tRNAs and suggests a mechanism by which the translation initiation machinery could discriminate the initiator tRNA from all other tRNAs.

Bibliographie

[1] Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**, 318–356.

[2] Rozenski, J., Crain, P. F., and McCloskey, J. A. (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res* **27(1)**, 196–197.

[3] Marquet, R., Isel, C., Ehresmann, C., and Ehresmann, B. (1995) tRNAs as primer of reverse transcriptases. *Biochimie* **77(1-2)**, 113–124.

[4] Crick, F. (1955) On degenerate templates and the adaptor hypothesis *unpublished communication to the "RNA Tie Club", cited by MB Hoagland (1960) in The Nucleic Acids. New York: Academic Press. pp 349.*

[5] Hoagland, M., Stephenson, M., Scott, J., Hecht, L., and Zamecnik, P. (1958) A soluble ribonucleic acid intermediate in protein synthesis. *J Biol Chem* **231(1)**, 241–257.

[6] Clark, B. F. (2001) The crystallization and structural determination of tRNA. *Trends Biochem Sci* **26(8)**, 511–514.

[7] Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J., and Rich, A. (1973) Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science* **179(70)**, 285–288.

[8] Suddath, F. L., Quigley, G. J., McPherson, A., Sneden, D., Kim, J. J., Kim, S. H., and Rich, A. (1974) Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0 angstroms resolution. *Nature* **248(443)**, 20–24.

[9] Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F., and Klug, A. (1974) Structure of yeast phenylalanine tRNA at 3 A resolution. *Nature* **250(467)**, 546–551.

[10] Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H., Seeman, N. C., and Rich, A. (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* **185(149)**, 435–440.

[11] Sprinzl, M. and Vassilenko, K. S. (2005) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **33(Database issue)**, D139–40.

[12] Marck, C. and Grosjean, H. (2002) tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **8(10)**, 1189–1232.

[13] Xue, H., Tong, K.-L., Marck, C., Grosjean, H., and Wong, J. T.-F. (2003) Transfer RNA paralogs: evidence for genetic code-amino acid biosynthesis coevolution and an archaeal root of life. *Gene* **310**, 59–66.

[14] Di Giulio, M. (2006) The non-monophyletic origin of the tRNA molecule and the origin of genes only after the evolutionary stage of the last universal common ancestor (LUCA). *J Theor Biol* **240(3)**, 343–352.

[15] Holley, R. W., Everett, G. A., Madison, J. T., and Zamir, A. (1965) Nucleotide sequences in the Yeast alanine transfer ribonucleic acid *J Biol Chem* **240**, 2122–2128.

[16] Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **26(1)**, 148–153.

[17] Davis, F. F. and Allen, F. W. (1957) Ribonucleic acids from yeast which contain a fifth nucleotide. *J Biol Chem* **227(2)**, 907–915.

[18] Littlefield, J. W. and Dunn, D. B. (1958) The occurrence and distribution of thymine and three methylated-adenine bases in ribonucleic acids from several sources. *Biochem J* **70(4)**, 642–651.

[19] Cohn, W. E. (1960) Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *J Biol Chem* **235**, 1488–1498.

[20] Leontis, N. B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* **7(4)**, 499–512.

[21] Cusack, S. (1997) Aminoacyl-tRNA synthetases. *Curr Opin Struct Biol* **7(6)**, 881–889.

[22] Woese, C. R., Olsen, G. J., Ibba, M., and Soll, D. (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev* **64(1)**, 202–236.

[23] Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347(6289)**, 203–206.

[24] Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N., and Leberman, R. (1990) A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 A. *Nature* **347(6290)**, 249–255.

[25] Rich, A. and RajBhandary, U. L. (1976) Transfer RNA: molecular structure, sequence, and properties. *Annu Rev Biochem* **45**, 805–860.

[26] Levitt, M. (1969) Detailed molecular model for transfer ribonucleic acid. *Nature* **224(5221)**, 759–763.

[27] Shi, H. and Moore, P. B. (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA* **6(8)**, 1091–1105.

[28] Saenger, W. (1984) Principles of nucleic acid structure, Springer New York, .

[29] Oliva, R., Cavallo, L., and Tramontano, A. (2006) Accurate energies of hydrogen bonded nucleic acid base pairs and triplets in tRNA tertiary interactions. *Nucleic Acids Res* **34(3)**, 865–879.

[30] Draper, D. E. (1996) Strategies for RNA folding. *Trends Biochem Sci* **21(4)**, 145–149.

[31] Brion, P. and Westhof, E. (1997) Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct* **26**, 113–137.

[32] Maglott, E. J., Deo, S. S., Przykorska, A., and Glick, G. D. (1998) Conformational transitions of an unmodified tRNA: implications for RNA folding. *Biochemistry* **37(46)**, 16349–16359.

[33] Serebrov, V., Clarke, R. J., Gross, H. J., and Kisselev, L. (2001) Mg2+-induced tRNA folding. *Biochemistry* **40(22)**, 6688–6698.

[34] Shelton, V. M., Sosnick, T. R., and Pan, T. (2001) Altering the intermediate in the equilibrium folding of unmodified yeast tRNAPhe with monovalent and divalent cations. *Biochemistry* **40(12)**, 3629–3638.

[35] Westhof, E., Dumas, P., and Moras, D. (1988) Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals. *Acta Crystallogr A* **44 (Pt 2)**, 112–123.

[36] Basavappa, R. and Sigler, P. B. (1991) The 3 A crystal structure of yeast initiator tRNA: functional implications in initiator/elongator discrimination. *EMBO J* **10(10)**, 3105–3111.

[37] Benas, P., Bec, G., Keith, G., Marquet, R., Ehresmann, C., Ehresmann, B., and Dumas, P. (2000) The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop. *RNA* **6(10)**, 1347–1355.

[38] Schmitt, E., Panvert, M., Blanquet, S., and Mechulam, Y. (1998) Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet. *EMBO J* **17(23)**, 6819–6826.

[39] Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989) Structure of E. coli glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 A resolution. *Science* **246(4934)**, 1135–1142.

[40] Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., and Moras, D. (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science* **252(5013)**, 1682–1689.

[41] Biou, V., Yaremchuk, A., Tukalo, M., and Cusack, S. (1994) The 2.9 A crystal structure of T. thermophilus seryl-tRNA synthetase complexed with tRNA(Ser). *Science* **263(5152)**, 1404–1410.

[42] Sankaranarayanan, R., Dock-Bregeon, A. C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B., and Moras, D. (1999) The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* **97(3)**, 371–381.

[43] Cusack, S., Yaremchuk, A., and Tukalo, M. (2000) The 2 A crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *EMBO J* **19(10)**, 2351–2361.

[44] Yaremchuk, A., Tukalo, M., Grotli, M., and Cusack, S. (2001) A succession of substrate induced conformational changes ensures the amino acid specificity of Thermus thermophilus prolyl-tRNA synthetase: comparison with histidyl-tRNA synthetase. *J Mol Biol* **309(4)**, 989–1002.

[45] Bullock, T. L., Uter, N., Nissan, T. A., and Perona, J. J. (2003) Amino acid discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants. *J Mol Biol* **328(2)**, 395–408.

[46] Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science* **292(5518)**, 883–896.

[47] Selmer, M., Dunham, C. M., Murphy, F. V. t., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R., and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313(5795)**, 1935–1942.

[48] Haselkorn, R. and Rothman-Denes, L. B. (1973) Protein synthesis. *Annu Rev Biochem* **42**, 397–438.

[49] Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation. *Cell* **108(4)**, 557–572.

[50] Crick, F. H. (1966) Codon–anticodon pairing: the wobble hypothesis. *J Mol Biol* **19(2)**, 548–555.

[51] Murphy, F. V. t. and Ramakrishnan, V. (2004) Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat Struct Mol Biol* **11(12)**, 1251–1252.

[52] Moser, J., Lorenz, S., Hubschwerlen, C., Rompf, A., and Jahn, D. (1999) Methanopyrus kandleri glutamyl-tRNA reductase. *J Biol Chem* **274(43)**, 30679–30685.

[53] Schon, A., Krupp, G., Gough, S., Berry-Lowe, S., Kannangara, C. G., and Soll, D. (1986) The RNA required in the first step of chlorophyll biosynthesis is a chloroplast glutamate tRNA. *Nature* **322(6076)**, 281–284.

[54] Roberts, W. S., Petit, J. F., and Strominger, J. L. (1968) Biosynthesis of the peptidoglycan of bacterial cell walls. 8. Specificity in the utilization of L-alanyl transfer ribonucleic acid for interpeptide bridge synthesis in Arthrobacter crystallopoietes. *J Biol Chem* **243(4)**, 768–772.

[55] Nashimoto, M. (1993) 3' truncated tRNAArg is essential for in vitro specific cleavage of partially synthesized mouse 18S rRNA. *Nucleic Acids Res* **21(20)**, 4696–4702.

[56] Hopper, A. K. and Phizicky, E. M. (2003) tRNA transfers to the limelight. *Genes Dev* **17(2)**, 162–180.

[57] Nakanishi, K. and Nureki, O. (2005) Recent progress of structural biology of tRNA processing and modification. *Mol Cells* **19(2)**, 157–166.

[58] Li, H. (2007) Complexes of tRNA and maturation enzymes: shaping up for translation. *Curr Opin Struct Biol* **17(3)**, 293–301.

[59] Herr, A. J., Jensen, M. B., Dalmay, T., and Baulcombe, D. C. (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* **308(5718)**, 118–120.

[60] Geiduschek, E. P. and Tocchini-Valentini, G. P. (1988) Transcription by RNA polymerase III. *Annu Rev Biochem* **57**, 873–914.

[61] Evans, D., Marquez, S. M., and Pace, N. R. (2006) RNase P: interface of the RNA and protein worlds. *Trends Biochem Sci* **31(6)**, 333–341.

[62] Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragon, A. (2006) Structure of ribonuclease P–a universal ribozyme. *Curr Opin Struct Biol* **16(3)**, 327–335.

[63] Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35(3 Pt 2)**, 849–857.

[64] Krasilnikov, A. S., Yang, X., Pan, T., and Mondragon, A. (2003) Crystal structure of the specificity domain of ribonuclease P. *Nature* **421(6924)**, 760–764.

[65] Chamberlain, J. R., Lee, Y., Lane, W. S., and Engelke, D. R. (1998) Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP. *Genes Dev* **12(11)**, 1678–1690.

[66] True, H. L. and Celander, D. W. (1998) Protein components contribute to active site architecture for eukaryotic ribonuclease P. *J Biol Chem* **273(13)**, 7193–7196.

[67] Li, Z. and Deutscher, M. P. (1994) The role of individual exoribonucleases in processing at the 3' end of Escherichia coli tRNA precursors. *J Biol Chem* **269(8)**, 6064–6071.

[68] Li, Z. and Deutscher, M. P. (1996) Maturation pathways for E. coli tRNA precursors: a random multienzyme process in vivo. *Cell* **86(3)**, 503–512.

[69] Li, Z. and Deutscher, M. P. (2002) RNase E plays an essential role in the maturation of Escherichia coli tRNA precursors. *RNA* **8(1)**, 97–109.

[70] Ow, M. C. and Kushner, S. R. (2002) Initiation of tRNA maturation by RNase E is essential for cell viability in E. coli. *Genes Dev* **16(9)**, 1102–1115.

[71] Ishii, R., Nureki, O., and Yokoyama, S. (2003) Crystal structure of the tRNA processing enzyme RNase PH from Aquifex aeolicus. *J Biol Chem* **278(34)**, 32397–32404.

[72] de laSierra-Gallay, I. L., Pellegrini, O., and Condon, C. (2005) Structural basis for substrate binding, cleavage and allostery in the tRNA maturase RNase Z. *Nature* **433(7026)**, 657–661.

[73] Ishii, R., Minagawa, A., Takaku, H., Takagi, M., Nashimoto, M., and Yokoyama, S. (2005) Crystal structure of the tRNA 3' processing endoribonuclease tRNase Z from Thermotoga maritima. *J Biol Chem* **280(14)**, 14138–14144.

[74] Li, F., Xiong, Y., Wang, J., Cho, H. D., Tomita, K., Weiner, A. M., and Steitz, T. A. (2002) Crystal structures of the Bacillus stearothermophilus CCA-adding enzyme and its complexes with ATP or CTP. *Cell* **111(6)**, 815–824.

[75] Okabe, M., Tomita, K., Ishitani, R., Ishii, R., Takeuchi, N., Arisaka, F., Nureki, O., and Yokoyama, S. (2003) Divergent evolutions of trinucleotide polymerization revealed by an archaeal CCA-adding enzyme structure. *EMBO J* **22(21)**, 5918–5927.

[76] Tomita, K., Fukai, S., Ishitani, R., Ueda, T., Takeuchi, N., Vassylyev, D. G., and Nureki, O. (2004) Structural basis for template-independent RNA polymerization. *Nature* **430(7000)**, 700–704.

[77] Xiong, Y. and Steitz, T. A. (2004) Mechanism of transfer RNA maturation by CCA-adding enzyme without using an oligonucleotide template. *Nature* **430(7000)**, 640–645.

[78] Abelson, J., Trotta, C. R., and Li, H. (1998) tRNA splicing. *J Biol Chem* **273(21)**, 12685–12688.

[79] Limbach, P. A., Crain, P. F., and McCloskey, J. A. (1994) Summary: the modified nucleosides of RNA. *Nucleic Acids Res* **22(12)**, 2183–2196.

[80] Dunin-Horkawicz, S., Czerwoniec, A., Gajda, M. J., Feder, M., Grosjean, H., and Bujnicki, J. M. (2006) MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res* **34(Database issue)**, D145–9.

[81] Motorin, Y. and Grosjean, H. (1998) Appendix 1: Chemical structures and Classification of posttranscriptionally Modified Nucleosides in RNA In H. Grosjean and R. Benne, (ed.), Modification and Editing of RNA, pp. 543–549 ASM Press Washington, DC.

[82] Garcia, G. and Goodenough-Lashua, D. (1998) Mechanisms of RNA-Modifying and Editing Enzymes In H. Grosjean and R. Benne, (ed.), Modification and Editing of RNA, pp. 135–168 ASM Press Washington, DC.

[83] Auffinger, P. and Westhof, E. (1998) Appendix 5: Localisation and Distribution of Modified Nucleotides in tRNA In H. Grosjean and R. Benne, (ed.), Modification and Editing of RNA, pp. 569–576 ASM Press Washington, DC.

[84] Grosjean, H., Sprinzl, M., and Steinberg, S. (1995) Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. *Biochimie* **77(1-2)**, 139–141.

[85] Nishimura, S. and Watanabe, K. (2006) The discovery of modified nucleosides from the early days to the present: a personal perspective. *J Biosci* **31(4)**, 465–475.

[86] Grosjean, H. (2005) Fine-tuning of RNA Functions by Modification and Editing, Springer-Verlag, Berlin.

[87] Agris, P. F. (1991) Wobble position modified nucleosides evolved to select transfer RNA codon recognition: a modified-wobble hypothesis. *Biochimie* **73(11)**, 1345–1349.

[88] Lim, V. I. and Curran, J. F. (2001) Analysis of codon:anticodon interactions within the ribosome provides new insights into codon reading and the genetic code structure. *RNA* **7(7)**, 942–957.

[89] Agris, P. F. (2004) Decoding the genome: a modified view. *Nucleic Acids Res* **32(1)**, 223–238.

[90] Agris, P. F., Vendeix, F. A. P., and Graham, W. D. (2007) tRNA's wobble decoding of the genome: 40 years of modification. *J Mol Biol* **366(1)**, 1–13.

[91] Meier, F., Suter, B., Grosjean, H., Keith, G., and Kubli, E. (1985) Queuosine modification of the wobble base in tRNAHis influences 'in vivo' decoding properties. *EMBO J* **4(3)**, 823–827.

[92] 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(4)**, 437–458.

[93] Gerber, A. P. and Keller, W. (1999) An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **286(5442)**, 1146–1149.

[94] Bjork, G. R., Durand, J. M., Hagervall, T. G., Leipuviene, R., Lundgren, H. K., Nilsson, K., Chen, P., Qian, Q., and Urbonavicius, J. (1999) Transfer RNA modification: influence on translational frameshifting and metabolism. *FEBS Lett* **452(1-2)**, 47–51.

[95] Curran, J. (1998) Modified nucleosides in Translation In H. Grosjean and R. Benne, (ed.), Modification and Editing of RNA, pp. 493–516 ASM Press Washington, DC.

[96] Bjork, G. R., Wikstrom, P. M., and Bystrom, A. S. (1989) Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. *Science* **244(4907)**, 986–989.

[97] Sampson, J. R. and Uhlenbeck, O. C. (1988) Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proc Natl Acad Sci U S A* **85(4)**, 1033–1037.

[98] Derrick, W. B. and 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 Res* **21(21)**, 4948–4953.

[99] Perret, V., Garcia, A., Puglisi, J., Grosjean, H., Ebel, J. P., Florentz, C., and Giege, R. (1990) Conformation in solution of yeast tRNA(Asp) transcripts deprived of modified nucleotides. *Biochimie* **72(10)**, 735–743.

[100] Perret, V., Garcia, A., Grosjean, H., Ebel, J. P., Florentz, C., and Giege, R. (1990) Relaxation of a transfer RNA specificity by removal of modified nucleotides. *Nature* **344(6268)**, 787–789.

[101] Nobles, K. N., Yarian, C. S., Liu, G., Guenther, R. H., and Agris, P. F. (2002) Highly conserved modified nucleosides influence Mg2+-dependent tRNA folding. *Nucleic Acids Res* **30(21)**, 4751–4760.

[102] 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(24)**, 7560–7567.

[103] Hall, K. B., Sampson, J. R., Uhlenbeck, O. C., and Redfield, A. G. (1989) Structure of an unmodified tRNA molecule. *Biochemistry* **28(14)**, 5794–5801.

[104] Yue, D., Kintanar, A., and Horowitz, J. (1994) Nucleoside modifications stabilize Mg2+ binding in Escherichia coli tRNA(Val): an imino proton NMR investigation. *Biochemistry* **33(30)**, 8905–8911.

[105] Durant, P. C., Bajji, A. C., Sundaram, M., Kumar, R. K., and Davis, D. R. (2005) Structural effects of hypermodified nucleosides in the Escherichia coli and human tRNALys anticodon loop: the effect of nucleosides s2U, mcm5U, mcm5s2U, mnm5s2U, t6A, and ms2t6A. *Biochemistry* **44(22)**, 8078–8089.

[106] Vermeulen, A., McCallum, S. A., and Pardi, A. (2005) Comparison of the global structure and dynamics of native and unmodified tRNAval. *Biochemistry* **44(16)**, 6024–6033.

[107] Chen, Y., Sierzputowska-Gracz, H., Guenther, R., Everett, K., and Agris, P. F. (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(38)**, 10249–10253.

[108] Agris, P. F. (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.

[109] Auffinger, P. and Westhof, E. (1998) Effects of Pseudouridylation on tRNA Hydratation and Dynamics: a Theoretical Approach In H. Grosjean and R. Benne, (ed.), Modification and Editing of RNA, pp. 103–112 ASM Press Washington, DC.

[110] Davis, D. (1998) Biophysical and Conformational Properties of Modified Nucleosides in RNA (Nuclear Magnetic Resonance Studies) In H. Grosjean and R. Benne, (ed.), Modification and Editing of RNA, pp. 85–102 ASM Press Washington, DC.

[111] Droogmans, L., Roovers, M., Bujnicki, J. M., Tricot, C., Hartsch, T., Stalon, V., and Grosjean, H. (2003) Cloning and characterization of tRNA (m1A58) methyltransferase (TrmI) from Thermus thermophilus HB27, a protein required for cell growth at extreme temperatures. *Nucleic Acids Res* **31(8)**, 2148–2156.

[112] Helm, M. (2006) Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res* **34(2)**, 721–733.

[113] Tisne, C., Rigourd, M., Marquet, R., Ehresmann, C., and Dardel, F. (2000) NMR and biochemical characterization of recombinant human tRNA(Lys)3 expressed in Escherichia coli: identification of posttranscriptional nucleotide modifications required for efficient initiation of HIV-1 reverse transcription. *RNA* **6(10)**, 1403–1412.

[114] Tisne, C., Roques, B. P., and Dardel, F. (2001) Heteronuclear NMR studies of the interaction of tRNA(Lys)3 with HIV-1 nucleocapsid protein. *J Mol Biol* **306(3)**, 443–454.

[115] Tisne, C., Roques, B. P., and Dardel, F. (2003) Specific recognition of primer tRNA Lys 3 by HIV-1 nucleocapsid protein: involvement of the zinc fingers and the N-terminal basic extension. *Biochimie* **85(5)**, 557–561.

[116] Tisne, C., Roques, B. P., and Dardel, F. (2004) The annealing mechanism of HIV-1 reverse transcription primer onto the viral genome. *J Biol Chem* **279(5)**, 3588–3595.

[117] Tisne, C. (2005) Structural bases of the annealing of primer tRNA(3Lys) to the HIV-1 viral RNA. *Curr HIV Res* **3(2)**, 147–156.

[118] De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **279(5349)**, 384–388.

[119] Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., and Summers, M. F. (2000) NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J Mol Biol* **301(2)**, 491–511.

[120] Wrede, P., Woo, N. H., and Rich, A. (1979) Initiator tRNAs have a unique anticodon loop conformation. *Proc Natl Acad Sci U S A* **76(7)**, 3289–3293.

[121] Wakao, H., Romby, P., Westhof, E., Laalami, S., Grunberg-Manago, M., Ebel, J. P., Ehresmann, C., and Ehresmann, B. (1989) The solution structure of the Escherichia coli initiator tRNA and its interactions with initiation factor 2 and the ribosomal 30 S subunit. *J Biol Chem* **264(34)**, 20363–20371.

[122] Seong, B. L. and RajBhandary, U. L. (1987) Escherichia coli formylmethionine tRNA: mutations in GGGCCC sequence conserved in anticodon stem of initiator tRNAs affect initiation of protein synthesis and conformation of anticodon loop. *Proc Natl Acad Sci U S A* **84(2)**, 334–338.

[123] Hartz, D., Binkley, J., Hollingsworth, T., and Gold, L. (1990) Domains of initiator tRNA and initiation codon crucial for initiator tRNA selection by Escherichia coli IF3. *Genes Dev* **4(10)**, 1790–1800.

[124] Schwartz, O., Marechal, V., Danos, O., and Heard, J. M. (1995) Human immunodeficiency virus type 1 Nef increases the efficiency of reverse transcription in the infected cell. *J Virol* **69(7)**, 4053–4059.

[125] Harrich, D., Ulich, C., Garcia-Martinez, L. F., and Gaynor, R. B. (1997) Tat is required for efficient HIV-1 reverse transcription. *EMBO J* **16(6)**, 1224–1235.

[126] Apolloni, A., Meredith, L. W., Suhrbier, A., Kiernan, R., and Harrich, D. (2007) The HIV-1 Tat protein stimulates reverse transcription in vitro. *Curr HIV Res* **5(5)**, 473–483.

[127] Goncalves, J., Korin, Y., Zack, J., and Gabuzda, D. (1996) Role of Vif in human immunodeficiency virus type 1 reverse transcription. *J Virol* **70(12)**, 8701–8709.

[128] Henriet, S., Sinck, L., Bec, G., Gorelick, R. J., Marquet, R., and Paillart, J.-C. (2007) Vif is a RNA chaperone that could temporally regulate RNA dimerization and the early steps of HIV-1 reverse transcription. *Nucleic Acids Res* **35(15)**, 5141–5153.

[129] Stark, L. A. and Hay, R. T. (1998) Human immunodeficiency virus type 1 (HIV-1) viral protein R (Vpr) interacts with Lys-tRNA synthetase: implications for priming of HIV-1 reverse transcription. *J Virol* **72(4)**, 3037–3044.

[130] Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282(5394)**, 1669–1675.

[131] Goff, S. P. (1990) Retroviral reverse transcriptase: synthesis, structure, and function. *J Acquir Immune Defic Syndr* **3(8)**, 817–831.

[132] Skalka, A. M. and Goff, S. P. (1993) Reverse transcriptase, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

[133] Rodgers, D. W., Gamblin, S. J., Harris, B. A., Ray, S., Culp, J. S., Hellmig, B., Woolf, D. J., Debouck, C., and Harrison, S. C. (1995) The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* **92(4)**, 1222–1226.

[134] Esnouf, R., Ren, J., Ross, C., Jones, Y., Stammers, D., and Stuart, D. (1995) Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat Struct Biol* **2(4)**, 303–308.

[135] Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1992) Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256(5065)**, 1783–1790.

[136] Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D. J., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., and Clark, P. (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc Natl Acad Sci U S A* **90(13)**, 6320–6324.

[137] Kati, W. M., Johnson, K. A., Jerva, L. F., and Anderson, K. S. (1992) Mechanism and fidelity of HIV reverse transcriptase. *J Biol Chem* **267(36)**, 25988–25997.

[138] Hsiou, Y., Ding, J., Das, K., Clark, A. D. J., Hughes, S. H., and Arnold, E. (1996) Structure of unliganded HIV-1 reverse transcriptase at 2.7 A resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure* **4(7)**, 853–860.

[139] Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R. G., Boyer, P. L., Hughes, S. H., Pauwels, R., Andries, K., Janssen, P. A., and Arnold, E. (1994) Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase. Implications for mechanisms of drug inhibition and resistance. *J Mol Biol* **243(3)**, 369–387.

[140] Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., and Stuart, D. (1995) High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat Struct Biol* **2(4)**, 293–302.

[141] Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R., and Mak, J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol* **2(6)**, 461–472.

[142] Frankel, A. D. and Young, J. A. (1998) HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* **67**, 1–25.

[143] Baudin, F., Marquet, R., Isel, C., Darlix, J. L., Ehresmann, B., and Ehresmann, C. (1993) Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J Mol Biol* **229(2)**, 382–397.

[144] Peterlin, B. M., Luciw, P. A., Barr, P. J., and Walker, M. D. (1986) Elevated levels of mRNA can account for the trans-activation of human immunodeficiency virus. *Proc Natl Acad Sci U S A* **83(24)**, 9734–9738.

[145] Kleiman, L., Halwani, R., and Javanbakht, H. (2004) The selective packaging and annealing of primer tRNALys3 in HIV-1. *Curr HIV Res* **2(2)**, 163–175.

[146] Berkhout, B. (1996) Structure and function of the human immunodeficiency virus leader RNA. *Prog Nucleic Acid Res Mol Biol* **54**, 1–34.

[147] Paillart, J. C., Skripkin, E., Ehresmann, B., Ehresmann, C., and Marquet, R. (1996) A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci U S A* **93(11)**, 5572–5577.

[148] Clever, J. L. and Parslow, T. G. (1997) Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol* **71(5)**, 3407–3414.

[149] Laughrea, M., Jette, L., Mak, J., Kleiman, L., Liang, C., and Wainberg, M. A. (1997) Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *J Virol* **71(5)**, 3397–3406.

[150] Raba, M., Limburg, K., Burghagen, M., Katze, J. R., Simsek, M., Heckman, J. E., Rajbhandary, U. L., and Gross, H. J. (1979) Nucleotide sequence of three isoaccepting lysine tRNAs from rabbit liver and SV40-transformed mouse fibroblasts. *Eur J Biochem* **97(1)**, 305–318.

[151] Jiang, M., Mak, J., Ladha, A., Cohen, E., Klein, M., Rovinski, B., and Kleiman, L. (1993) Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type 1. *J Virol* **67(6)**, 3246–3253.

[152] Li, Z., Shalom, A., Huang, Y., Mak, J., Arts, E., Wainberg, M. A., and Kleiman, L. (1996) Multiple forms of tRNA(Lys3) in HIV-1. *Biochem Biophys Res Commun* **227(2)**, 530–540.

[153] Guo, F., Cen, S., Niu, M., Javanbakht, H., and Kleiman, L. (2003) Specific inhibition of the synthesis of human lysyl-tRNA synthetase results in decreases in tRNA(Lys) incorporation, tRNA(3)(Lys) annealing to viral RNA, and viral infectivity in human immunodeficiency virus type 1. *J Virol* **77(18)**, 9817–9822.

[154] Halwani, R., Cen, S., Javanbakht, H., Saadatmand, J., Kim, S., Shiba, K., and Kleiman, L. (2004) Cellular distribution of Lysyl-tRNA synthetase and its interaction with Gag during human immunodeficiency virus type 1 assembly. *J Virol* **78(14)**, 7553–7564.

[155] Cen, S., Javanbakht, H., Niu, M., and Kleiman, L. (2004) Ability of wild-type and mutant lysyl-tRNA synthetase to facilitate tRNA(Lys) incorporation into human immunodeficiency virus type 1. *J Virol* **78(3)**, 1595–1601.

[156] Huang, Y., Wang, J., Shalom, A., Li, Z., Khorchid, A., Wainberg, M. A., and Kleiman, L. (1997) Primer tRNA3Lys on the viral genome exists in unextended and two-base extended forms within mature human immunodeficiency virus type 1. *J Virol* **71(1)**, 726–728.

[157] Hilbers, C. W., Robillard, G. T., Shulamn, R. G., Blake, R. D., Webb, P. K., Fresco, R., and Riesner, D. (1976) Thermal unfolding of yeast glycine transfer RNA. *Biochemistry* **15(9)**, 1874–1882.

[158] Hinz, H. J., Filimonov, V. V., and Privalov, P. L. (1977) Calorimetric studies on melting of tRNA Phe (yeast). *Eur J Biochem* **72(1)**, 79–86.

[159] Green, L. M. and Berg, J. M. (1990) Retroviral nucleocapsid protein-metal ion interactions: folding and sequence variants. *Proc Natl Acad Sci U S A* **87(16)**, 6403–6407.

[160] Krishna, S. S., Majumdar, I., and Grishin, N. V. (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res* **31(2)**, 532–550.

[161] Wondrak, E. M., Louis, J. M., deRocquigny, H., Chermann, J. C., and Roques, B. P. (1993) The gag precursor contains a specific HIV-1 protease cleavage site between the NC (P7) and P1 proteins. *FEBS Lett* **333(1-2)**, 21–24.

[162] Henderson, L. E., Bowers, M. A., Sowder, R. C. n., Serabyn, S. A., Johnson, D. G., Bess, J. W. J., Arthur, L. O., Bryant, D. K., and Fenselau, C. (1992) Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings, and complete amino acid sequences. *J Virol* **66(4)**, 1856–1865.

[163] Darlix, J. L., Gabus, C., Nugeyre, M. T., Clavel, F., and Barre-Sinoussi, F. (1990) Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J Mol Biol* **216(3)**, 689–699.

[164] Muriaux, D., De Rocquigny, H., Roques, B. P., and Paoletti, J. (1996) NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J Biol Chem* **271(52)**, 33686–33692.

[165] Gorelick, R. J., Henderson, L. E., Hanser, J. P., and Rein, A. (1988) Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a "zinc finger-like" protein sequence. *Proc Natl Acad Sci U S A* **85(22)**, 8420–8424.

[166] Gorelick, R. J., Nigida, S. M. J., Bess, J. W. J., Arthur, L. O., Henderson, L. E., and Rein, A. (1990) Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J Virol* **64(7)**, 3207–3211.

[167] Gorelick, R. J., Chabot, D. J., Rein, A., Henderson, L. E., and Arthur, L. O. (1993) The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent. *J Virol* **67(7)**, 4027–4036.

[168] De Rocquigny, H., Gabus, C., Vincent, A., Fournie-Zaluski, M. C., Roques, B., and Darlix, J. L. (1992) Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc Natl Acad Sci U S A* **89(14)**, 6472–6476.

[169] Feng, Y. X., Campbell, S., Harvin, D., Ehresmann, B., Ehresmann, C., and Rein, A. (1999) The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site. *J Virol* **73(5)**, 4251–4256.

[170] Hargittai, M. R., Mangla, A. T., Gorelick, R. J., and Musier-Forsyth, K. (2001) HIV-1 nucleocapsid protein zinc finger structures induce tRNA(Lys,3) structural changes but are not critical for primer/template annealing. *J Mol Biol* **312(5)**, 985–997.

[171] Peliska, J. A., Balasubramanian, S., Giedroc, D. P., and Benkovic, S. J. (1994) Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity. *Biochemistry* **33(46)**, 13817–13823.

[172] Ben-Artzi, H., Shemesh, J., Zeelon, E., Amit, B., Kleiman, L., Gorecki, M., and Panet, A. (1996) Molecular analysis of the second template switch during reverse transcription of the HIV RNA template. *Biochemistry* **35(32)**, 10549–10557.

[173] Johnson, P. E., Turner, R. B., Wu, Z. R., Hairston, L., Guo, J., Levin, J. G., and Summers, M. F. (2000) A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry* **39(31)**, 9084–9091.

[174] Rein, A., Henderson, L. E., and Levin, J. G. (1998) Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci* **23(8)**, 297–301.

[175] Levin, J. G., Guo, J., Rouzina, I., and Musier-Forsyth, K. (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog Nucleic Acid Res Mol Biol* **80**, 217–286.

[176] South, T. L., Blake, P. R., Hare, D. R., and Summers, M. F. (1991) C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain. *Biochemistry* **30(25)**, 6342–6349.

[177] Summers, M. F., South, T. L., Kim, B., and Hare, D. R. (1990) High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach. *Biochemistry* **29(2)**, 329–340.

[178] Summers, M. F., Henderson, L. E., Chance, M. R., Bess, J. W. J., South, T. L., Blake, P. R., Sagi, I., Perez-Alvarado, G., Sowder, R. C. r., and Hare, D. R. (1992) Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein Sci* **1(5)**, 563–574.

[179] Morellet, N., Jullian, N., De Rocquigny, H., Maigret, B., Darlix, J. L., and Roques, B. P. (1992) Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR. *EMBO J* **11(8)**, 3059–3065.

[180] Lee, B. M., De Guzman, R. N., Turner, B. G., Tjandra, N., and Summers, M. F. (1998) Dynamical behavior of the HIV-1 nucleocapsid protein. *J Mol Biol* **279(3)**, 633–649.

[181] Fisher, R. J., Rein, A., Fivash, M., Urbaneja, M. A., Casas-Finet, J. R., Medaglia, M., and Henderson, L. E. (1998) Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J Virol* **72(3)**, 1902–1909.

[182] Vuilleumier, C., Bombarda, E., Morellet, N., Gerard, D., Roques, B. P., and Mely, Y. (1999) Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry* **38(51)**, 16816–16825.

[183] Mely, Y., deRocquigny, H., Sorinas-Jimeno, M., Keith, G., Roques, B. P., Marquet, R., and Gerard, D. (1995) Binding of the HIV-1 nucleocapsid protein to the primer tRNA(3Lys), in vitro, is essentially not specific. *J Biol Chem* **270(4)**, 1650–1656.

[184] Urbaneja, M. A., Kane, B. P., Johnson, D. G., Gorelick, R. J., Henderson, L. E., and Casas-Finet, J. R. (1999) Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: elucidation of the structural determinants for function. *J Mol Biol* **287(1)**, 59–75.

[185] Shubsda, M. F., Paoletti, A. C., Hudson, B. S., and Borer, P. N. (2002) Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. *Biochemistry* **41(16)**, 5276–5282.

[186] You, J. C. and McHenry, C. S. (1993) HIV nucleocapsid protein. Expression in Escherichia coli, purification, and characterization. *J Biol Chem* **268(22)**, 16519–16527.

[187] Khan, R. and Giedroc, D. P. (1994) Nucleic acid binding properties of recombinant Zn2 HIV-1 nucleocapsid protein are modulated by COOH-terminal processing. *J Biol Chem* **269(36)**, 22538–22546.

[188] Karpel, R. L., Henderson, L. E., and Oroszlan, S. (1987) Interactions of retroviral structural proteins with single-stranded nucleic acids. *J Biol Chem* **262(11)**, 4961–4967.

[189] Isel, C., Lanchy, J. M., Le Grice, S. F., Ehresmann, C., Ehresmann, B., and Marquet, R. (1996) Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer tRNA3Lys. *EMBO J* **15(4)**, 917–924.

[190] Lanchy, J. M., Keith, G., Le Grice, S. F., Ehresmann, B., Ehresmann, C., and Marquet, R. (1998) Contacts between reverse transcriptase and the primer strand govern the transition from initiation to elongation of HIV-1 reverse transcription. *J Biol Chem* **273(38)**, 24425–24432.

[191] Lanchy, J. M., Ehresmann, C., Le Grice, S. F., Ehresmann, B., and Marquet, R. (1996) Binding and kinetic properties of HIV-1 reverse transcriptase markedly differ during initiation and elongation of reverse transcription. *EMBO J* **15(24)**, 7178–7187.

[192] Gilboa, E., Goff, S., Shields, A., Yoshimura, F., Mitra, S., and Baltimore, D. (1979) In vitro synthesis of a 9 kbp terminally redundant DNA carrying the infectivity of Moloney murine leukemia virus. *Cell* **16(4)**, 863–874.

[193] Stoll, E., Billeter, M. A., Palmenberg, A., and Weissmann, C. (1977) Avian myeloblastosis virus RNA is terminally redundant: implications for the mechanism of retrovirus replication. *Cell* **12(1)**, 57–72.

[194] Hu, W. S. and Temin, H. M. (1990) Retroviral recombination and reverse transcription. *Science* **250(4985)**, 1227–1233.

[195] vanWamel, J. L. and Berkhout, B. (1998) The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly. *Virology* **244(2)**, 245–251.

[196] Mitra, S. W., Chow, M., Champoux, J., and Baltimore, D. (1982) Synthesis of murine leukemia virus plus strong stop DNA initiates at a unique site. *J Biol Chem* **257(11)**, 5983–5986.

[197] Pullen, K. A., Rattray, A. J., and Champoux, J. J. (1993) The sequence features important for plus strand priming by human immunodeficiency virus type 1 reverse transcriptase. *J Biol Chem* **268(9)**, 6221–6227.

[198] Charneau, P. and Clavel, F. (1991) A single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. *J Virol* **65(5)**, 2415–2421.

[199] Auxilien, S., Keith, G., Le Grice, S. F., and Darlix, J. L. (1999) Role of post-transcriptional modifications of primer tRNALys,3 in the fidelity and efficacy of plus strand DNA transfer during HIV-1 reverse transcription. *J Biol Chem* **274(7)**, 4412–4420.

[200] Burnett, B. P. and McHenry, C. S. (1997) Posttranscriptional modification of retroviral primers is required for late stages of DNA replication. *Proc Natl Acad Sci U S A* **94(14)**, 7210–7215.

[201] Smith, J. S. and Roth, M. J. (1992) Specificity of human immunodeficiency virus-1 reverse transcriptase-associated ribonuclease H in removal of the minus-strand primer, tRNA(Lys3). *J Biol Chem* **267(21)**, 15071–15079.

[202] Rumbaugh, J. A., Fuentes, G. M., and Bambara, R. A. (1998) Processing of an HIV replication intermediate by the human DNA replication enzyme FEN1. *J Biol Chem* **273(44)**, 28740–28745.

[203] Hargittai, M. R. S., Gorelick, R. J., Rouzina, I., and Musier-Forsyth, K. (2004) Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J Mol Biol* **337(4)**, 951–968.

[204] Cen, S., Huang, Y., Khorchid, A., Darlix, J. L., Wainberg, M. A., and Kleiman, L. (1999) The role of Pr55(gag) in the annealing of tRNA3Lys to human immunodeficiency virus type 1 genomic RNA. *J Virol* **73(5)**, 4485–4488.

[205] Lapadat-Tapolsky, M., Pernelle, C., Borie, C., and Darlix, J. L. (1995) Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1. *Nucleic Acids Res* **23(13)**, 2434–2441.

[206] Brule, F., Marquet, R., Rong, L., Wainberg, M. A., Roques, B. P., Le Grice, S. F. J., Ehresmann, B., and Ehresmann, C. (2002) Structural and functional properties of the HIV-1 RNA-tRNA(Lys)3 primer complex annealed by the nucleocapsid protein: comparison with the heat-annealed complex. *RNA* **8(1)**, 8–15.

[207] Iwatani, Y., Rosen, A. E., Guo, J., Musier-Forsyth, K., and Levin, J. G. (2003) Efficient initiation of HIV-1 reverse transcription in vitro. Requirement for RNA sequences downstream of the primer binding site abrogated by nucleocapsid protein-dependent primer-template interactions. *J Biol Chem* **278(16)**, 14185–14195.

[208] Khan, R. and Giedroc, D. P. (1992) Recombinant human immunodeficiency virus type 1 nucleocapsid (NCp7) protein unwinds tRNA. *J Biol Chem* **267(10)**, 6689–6695.

[209] Gregoire, C. J., Gautheret, D., and Loret, E. P. (1997) No tRNA3Lys unwinding in a complex with HIV NCp7. *J Biol Chem* **272(40)**, 25143–25148.

[210] Chan, B., Weidemaier, K., Yip, W. T., Barbara, P. F., and Musier-Forsyth, K. (1999) Intra-tRNA distance measurements for nucleocapsid proteindependent tRNA unwinding during priming of HIV reverse transcription. *Proc Natl Acad Sci U S A* **96(2)**, 459–464.

[211] Remy, E., deRocquigny, H., Petitjean, P., Muriaux, D., Theilleux, V., Paoletti, J., and Roques, B. P. (1998) The annealing of tRNA3Lys to human immunodeficiency virus type 1 primer binding site is critically dependent on the NCp7 zinc fingers structure. *J Biol Chem* **273(9)**, 4819–4822.

[212] Demene, H., Dong, C. Z., Ottmann, M., Rouyez, M. C., Jullian, N., Morellet, N., Mely, Y., Darlix, J. L., Fournie-Zaluski, M. C., and Saragosti, S. (1994) 1H NMR structure and biological studies of the His23–>Cys mutant nucleocapsid protein of HIV-1 indicate that the conformation of the first zinc finger is critical for virus infectivity. *Biochemistry* **33(39)**, 11707–11716.

[213] Guo, J., Wu, T., Kane, B. F., Johnson, D. G., Henderson, L. E., Gorelick, R. J., and Levin, J. G. (2002) Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* **76(9)**, 4370–4378.

[214] Rong, L., Liang, C., Hsu, M., Guo, X., Roques, B. P., and Wainberg, M. A. (2001) HIV-1 nucleocapsid protein and the secondary structure of the binary complex formed between tRNA(Lys.3) and viral RNA template play different roles during initiation of (-) strand DNA reverse transcription. *J Biol Chem* **276(50)**, 47725–47732.

[215] Williams, M. C., Gorelick, R. J., and Musier-Forsyth, K. (2002) Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Proc Natl Acad Sci U S A* **99(13)**, 8614–8619.

[216] Das, A. T., Klaver, B., and Berkhout, B. (1995) Reduced replication of human immunodeficiency virus type 1 mutants that use reverse transcription primers other than the natural tRNA(3Lys). *J Virol* **69(5)**, 3090–3097.

[217] Li, X., Mak, J., Arts, E. J., Gu, Z., Kleiman, L., Wainberg, M. A., and Parniak, M. A. (1994) Effects of alterations of primer-binding site sequences on human immunodeficiency virus type 1 replication. *J Virol* **68(10)**, 6198–6206.

[218] Wakefield, J. K., Wolf, A. G., and Morrow, C. D. (1995) Human immunodeficiency virus type 1 can use different tRNAs as primers for reverse transcription but selectively maintains a primer binding site complementary to tRNA(3Lys). *J Virol* **69(10)**, 6021–6029.

[219] Isel, C., Ehresmann, C., Keith, G., Ehresmann, B., and Marquet, R. (1995) Initiation of reverse transcription of HIV-1: secondary structure of the HIV-1 RNA/tRNA(3Lys) (template/primer). *J Mol Biol* **247(2)**, 236–250.

[220] Isel, C., Marquet, R., Keith, G., Ehresmann, C., and Ehresmann, B. (1993) Modified nucleotides of tRNA(3Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription. *J Biol Chem* **268(34)**, 25269–25272.

[221] Isel, C., Westhof, E., Massire, C., Le Grice, S. F., Ehresmann, B., Ehresmann, C., and Marquet, R. (1999) Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO J* **18(4)**, 1038–1048.

[222] Wakefield, J. K., Kang, S. M., and Morrow, C. D. (1996) Construction of a type 1 human immunodeficiency virus that maintains a primer binding site complementary to tRNA(His). *J Virol* **70(2)**, 966–975.

[223] Zhang, Z., Kang, S. M., LeBlanc, A., Hajduk, S. L., and Morrow, C. D. (1996) Nucleotide sequences within the U5 region of the viral RNA genome are the major determinants for an human immunodeficiency virus type 1 to maintain a primer binding site complementary to tRNA(His). *Virology* **226(2)**, 306–317.

[224] Zhang, Z., Kang, S. M., Li, Y., and Morrow, C. D. (1998) Genetic analysis of the U5-PBS of a novel HIV-1 reveals multiple interactions between the tRNA and RNA genome required for initiation of reverse transcription. *RNA* **4(4)**, 394–406.

[225] Kang, S. M., Zhang, Z., and Morrow, C. D. (1997) Identification of a sequence within U5 required for human immunodeficiency virus type 1 to stably maintain a primer binding site complementary to tRNA(Met). *J Virol* **71(1)**, 207–217.

[226] Kang, S. M., Zhang, Z., and Morrow, C. D. (1999) Identification of a human immunodeficiency virus type 1 that stably uses tRNALys1,2 rather than tRNALys,3 for initiation of reverse transcription. *Virology* **257(1)**, 95–105.

[227] Dupuy, L. C., Kelly, N. J., Elgavish, T. E., Harvey, S. C., and Morrow, C. D. (2003) Probing the importance of tRNA anticodon: human immunodeficiency virus type 1 (HIV-1) RNA genome complementarity with an HIV-1 that selects tRNA(Glu) for replication. *J Virol* **77(16)**, 8756–8764.

[228] Beerens, N., Groot, F., and Berkhout, B. (2001) Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. *J Biol Chem* **276(33)**, 31247–31256.

[229] Beerens, N. and Berkhout, B. (2002) The tRNA primer activation signal in the human immunodeficiency virus type 1 genome is important for initiation and processive elongation of reverse transcription. *J Virol* **76(5)**, 2329–2339.

[230] Abbink, T. E. M., Beerens, N., and Berkhout, B. (2004) Forced selection of a human immunodeficiency virus type 1 variant that uses a non-self tRNA primer for reverse transcription: involvement of viral RNA sequences and the reverse transcriptase enzyme. *J Virol* **78(19)**, 10706–10714.

[231] Ooms, M., Cupac, D., Abbink, T. E. M., Huthoff, H., and Berkhout, B. (2007) The availability of the primer activation signal (PAS) affects the efficiency of HIV-1 reverse transcription initiation. *Nucleic Acids Res* **35(5)**, 1649–1659.

[232] Abbink, T. E. M. and Berkhout, B. (2007) HIV-1 reverse transcription: close encounters between the viral genome and a cellular tRNA. *Adv Pharmacol* **55**, 99–135.

[233] Abbink, T. E. M. and Berkhout, B. (2003) A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J Biol Chem* **278(13)**, 11601–11611.

[234] Damgaard, C. K., Andersen, E. S., Knudsen, B., Gorodkin, J., and Kjems, J. (2004) RNA interactions in the 5' region of the HIV-1 genome. *J Mol Biol* **336(2)**, 369–379.

[235] Goldschmidt, V., Ehresmann, C., Ehresmann, B., and Marquet, R. (2003) Does the HIV-1 primer activation signal interact with tRNA3(Lys) during the initiation of reverse transcription ? *Nucleic Acids Res* **31(3)**, 850–859.

[236] Arts, E. J., Ghosh, M., Jacques, P. S., Ehresmann, B., and Le Grice, S. F. (1996) Restoration of tRNA3Lys-primed(-)-strand DNA synthesis to an HIV-1 reverse transcriptase mutant with extended tRNAs. Implications for retroviral replication. *J Biol Chem* **271(15)**, 9054–9061.

[237] Arts, E. J., Stetor, S. R., Li, X., Rausch, J. W., Howard, K. J., Ehresmann, B., North, T. W., Wohrl, B. M., Goody, R. S., Wainberg, M. A., and Grice, S. F. (1996) Initiation of (-) strand DNA synthesis from tRNA(3Lys) on lentiviral RNAs: implications of specific HIV-1 RNA-tRNA(3Lys) interactions inhibiting primer utilization by retroviral reverse transcriptases. *Proc Natl Acad Sci U S A* **93(19)**, 10063–10068.

[238] Isel, C., Keith, G., Ehresmann, B., Ehresmann, C., and Marquet, R. (1998) Mutational analysis of the tRNA3Lys/HIV-1 RNA (primer/template) complex. *Nucleic Acids Res* **26(5)**, 1198–1204.

[239] Liang, C., Li, X., Quan, Y., Laughrea, M., Kleiman, L., Hiscott, J., and Wainberg, M. A. (1997) Sequence elements downstream of the human immunodeficiency virus type 1 long terminal repeat are required for efficient viral gene transcription. *J Mol Biol* **272(2)**, 167–177.

[240] Wakefield, J. K. and Morrow, C. D. (1996) Mutations within the primer binding site of the human immunodeficiency virus type 1 define sequence requirements essential for reverse transcription. *Virology* **220(2)**, 290–298.

[241] Liang, C., Li, X., Rong, L., Inouye, P., Quan, Y., Kleiman, L., and Wainberg, M. A. (1997) The importance of the A-rich loop in human immunodeficiency virus type 1 reverse transcription and infectivity. *J Virol* **71(8)**, 5750–5757.

[242] Goldschmidt, V., Paillart, J.-C., Rigourd, M., Ehresmann, B., Aubertin, A.-M., Ehresmann, C., and Marquet, R. (2004) Structural variability of the initiation complex of HIV-1 reverse transcription. *J Biol Chem* **279(34)**, 35923–35931.

[243] Paillart, J.-C., Dettenhofer, M., Yu, X.-F., Ehresmann, C., Ehresmann, B., and Marquet, R. (2004) First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J Biol Chem* **279(46)**, 48397–48403.

[244] Aiyar, A., Cobrinik, D., Ge, Z., Kung, H. J., and Leis, J. (1992) Interaction between retroviral U5 RNA and the T psi C loop of the tRNA(Trp) primer is required for efficient initiation of reverse transcription. *J Virol* **66(4)**, 2464–2472.

[245] Freund, F., Boulme, F., Litvak, S., and Tarrago-Litvak, L. (2001) Initiation of HIV-2 reverse transcription: a secondary structure model of the RNA-tRNA(Lys3) duplex. *Nucleic Acids Res* **29(13)**, 2757–2765.

[246] Beerens, N. and Berkhout, B. (2002) Switching the in vitro tRNA usage of HIV-1 by simultaneous adaptation of the PBS and PAS. *RNA* **8(3)**, 357–369.

[247] Leis, J., Aiyar, A., and Cobrinik, D. (1993) Regulation of initiation of reverse transcription of retroviruses. In A M Skalka and S P Goff, (ed.), Reverse Transcriptase, pp. 33–48 Cold Spring Harbor Laboratory Press Cold Spring Harbor, New York.

[248] Voronin, Y. A. and Pathak, V. K. (2004) Frequent dual initiation in human immunodeficiency virus-based vectors containing two primer-binding sites: a quantitative in vivo assay for function of initiation complexes. *J Virol* **78(10)**, 5402–5413.

[249] Yuste, E., Borderia, A. V., Domingo, E., and Lopez-Galindez, C. (2005) Few mutations in the 5' leader region mediate fitness recovery of debilitated human immunodeficiency type 1 viruses. *J Virol* **79(9)**, 5421–5427.

[250] Fesik, S. and Zuiderweg, E. (1988) Heteronuclear three-dimensional NMR spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra *J. Magn. Reson* **78**, 588–593.

[251] Bax, A. and Grzesiek, S. (1993) Methodological advances in protein NMR *Acc. Chem. Res.* **26**, 131–138.

[252] Wüthrich, K. (1986) NMR of proteins and nucleic acids, John Wiley & Sons, New York.

[253] Palmer, A. G. r., Kroenke, C. D., and Loria, J. P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* **339**, 204–238.

[254] Kempf, J. and Loria, J. (2003) Protein dynamics from solution NMR: theory and applications. *Cell Biochem Biophys* **37(3)**, 187–211.

[255] Kempf, J. G. and Loria, J. P. (2004) Measurement of intermediate exchange phenomena. *Methods Mol Biol* **278**, 185–231.

[256] Mulder, F., Spronk, C., Slijper, M., Kaptein, R., and Boelens, R. (1996) Improved HSQC experiments for the observation of exchange broadened signals *Journal of Biomolecular NMR* **8(2)**, 223–228.

[257] Schanda, P., Van Melckebeke, H., and Brutscher, B. (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. *J Am Chem Soc* **128(28)**, 9042–9043.

[258] Lescop, E., Schanda, P., and Brutscher, B. (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson* **187(1)**, 163–169.

[259] Ikura, M., Kay, L., and Bax, A. (1990) A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29(19)**, 4659–67.

[260] Kay, L., Ikura, M., Tschudin, R., and Bax, A. (1990) Three-Dimensional Triple-Resonance NMR Spectroscopy of isotopically enriched proteins *J. Magn. Reson* **89(3)**, 496–514.

[261] Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients *Progress in Nuclear Magnetic Resonance Spectroscopy* **34(2)**, 93–158.

[262] Kao, C., Zheng, M., and Rudisser, S. (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5(9)**, 1268–1272.

[263] Cheng, X. and Blumenthal, R. (1999) S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions, World Scientific Publishing, Singapore.

[264] Martin, J. L. and McMillan, F. M. (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr Opin Struct Biol* **12(6)**, 783–793.

[265] Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci* **28(6)**, 329–335.

[266] Cantoni, G. L. (1975) Biological methylation: selected aspects. *Annu Rev Biochem* **44**, 435–451.

[267] Waddell, T. G., Eilders, L. L., Patel, B. P., and Sims, M. (2000) Prebiotic methylation and the evolution of methyl transfer reactions in living cells. *Orig Life Evol Biosph* **30(6)**, 539–548.

[268] Markham, G. D., Hafner, E. W., Tabor, C. W., and Tabor, H. (1980) S-Adenosylmethionine synthetase from Escherichia coli. *J Biol Chem* **255(19)**, 9082–9092.

[269] Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P., Pardhasaradhi, K., and McCann, P. P. (1996) S-Adenosylmethionine and methylation. *FASEB J* **10(4)**, 471–480.

[270] Palmer, J. L. and Abeles, R. H. (1979) The mechanism of action of S-adenosylhomocysteinase. *J Biol Chem* **254(4)**, 1217–1226.

[271] Banerjee, R. V. and Matthews, R. G. (1990) Cobalamin-dependent methionine synthase. *FASEB J* **4(5)**, 1450–1459.

[272] Drennan, C. L., Matthews, R. G., and Ludwig, M. L. (1994) Cobalamin-dependent methionine synthase: the structure of a methylcobalamin-binding fragment and implications for other B12-dependent enzymes. *Curr Opin Struct Biol* **4(6)**, 919–929.

[273] Ludwig, M. L. and Matthews, R. G. (1997) Structure-based perspectives on B12-dependent enzymes. *Annu Rev Biochem* **66**, 269–313.

[274] Epshtein, V., Mironov, A. S., and Nudler, E. (2003) The riboswitch-mediated control of sulfur metabolism in bacteria. *Proc Natl Acad Sci U S A* **100(9)**, 5052–5056.

[275] Thomas, D. J., Waters, S. B., and Styblo, M. (2004) Elucidating the pathway for arsenic methylation. *Toxicol Appl Pharmacol* **198(3)**, 319–326.

[276] Wuosmaa, A. M. and Hager, L. P. (1990) Methyl chloride transferase: a carbocation route for biosynthesis of halometabolites. *Science* **249(4965)**, 160–162.

[277] Attieh, J. M., Hanson, A. D., and Saini, H. S. (1995) Purification and characterization of a novel methyltransferase responsible for biosynthesis of halomethanes and methanethiol in Brassica oleracea. *J Biol Chem* **270(16)**, 9250–9257.

[278] Saxena, D., Aouad, S., Attieh, J., and Saini, H. (1998) Biochemical characterization of chloromethane emission from the wood-rotting fungus phellinus pomaceus *Appl Environ Microbiol* **64(8)**, 2831–2835.

[279] Ohsawa, N., Tsujita, M., Morikawa, S., and Itoh, N. (2001) Purification and characterization of a monohalomethane-producing enzyme S-adenosyl-L-methionine: halide ion methyltransferase from a marine microalga, Pavlova pinguis. *Biosci Biotechnol Biochem* **65(11)**, 2397–2404.

[280] Fu, Z., Hu, Y., Konishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M., and Takusagawa, F. (1996) Crystal structure of glycine N-methyltransferase from rat liver. *Biochemistry* **35(37)**, 11985–11993.

[281] Suzuki, T. and Takahashi, E. (1975) Biosynthesis of caffeine by tea-leaf extracts. Enzymic formation of theobromine from 7-methylxanthine and of caffeine from theobromine. *Biochem J* **146(1)**, 87–96.

[282] Kato, M., Mizuno, K., Fujimura, T., Iwama, M., Irie, M., Crozier, A., and Ashihara, H. (1999) Purification and characterization of caffeine synthase from tea leaves. *Plant Physiol* **120(2)**, 579–586.

[283] Ogawa, M., Herai, Y., Koizumi, N., Kusano, T., and Sano, H. (2001) 7-Methylxanthine methyltransferase of coffee plants. Gene isolation and enzymatic properties. *J Biol Chem* **276(11)**, 8213–8218.

[284] Uefuji, H., Ogita, S., Yamaguchi, Y., Koizumi, N., and Sano, H. (2003) Molecular cloning and functional characterization of three distinct N-methyltransferases involved in the caffeine biosynthetic pathway in coffee plants. *Plant Physiol* **132(1)**, 372–380.

[285] Misako, K. and Kouichi, M. (2004) Caffeine synthase and related methyltransferases in plants. *Front Biosci* **9**, 1833–1842.

[286] McCarthy, A. A. and McCarthy, J. G. (2007) The structure of two N-methyltransferases from the caffeine biosynthetic pathway. *Plant Physiol* **144(2)**, 879–889.

[287] Ashihara, H., Sano, H., and Crozier, A. (2008) Caffeine and related purine alkaloids: Biosynthesis, catabolism, function and genetic engineering. *Phytochemistry* **69(4)**, 841–856.

[288] Axelrod, J. and Tomchick, R. (1958) Enzymatic O-methylation of epinephrine and other catechols. *J Biol Chem* **233(3)**, 702–705.

[289] Molinoff, P. B. and Axelrod, J. (1971) Biochemistry of catecholamines. *Annu Rev Biochem* **40**, 465–500.

[290] Guldberg, H. C. and Marsden, C. A. (1975) Catechol-O-methyl transferase: pharmacological aspects and physiological role. *Pharmacol Rev* **27(2)**, 135–206.

[291] Huh, M. M. and Friedhoff, A. J. (1979) Multiple molecular forms of catechol-O-methyltransferase. Evidence for two distinct forms, and their purification and physical characterization. *J Biol Chem* **254(2)**, 299–308.

[292] Vidgren, J., Svensson, L. A., and Liljas, A. (1994) Crystal structure of catechol O-methyltransferase. *Nature* **368(6469)**, 354–358.

[293] Mannisto, P. T. and Kaakkola, S. (1989) New selective COMT inhibitors: useful adjuncts for Parkinson's disease ? *Trends Pharmacol Sci* **10(2)**, 54–56.

[294] Kaakkola, S., Gordin, A., and Mannisto, P. T. (1994) General properties and clinical possibilities of new selective inhibitors of catechol O-methyltransferase. *Gen Pharmacol* **25(5)**, 813–824.

[295] Mannisto, P. T. and Kaakkola, S. (1999) Catechol-O-methyltransferase (COMT): biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective COMT inhibitors. *Pharmacol Rev* **51(4)**, 593–628.

[296] Kaakkola, S. (2000) Clinical pharmacology, therapeutic use and potential of COMT inhibitors in Parkinson's disease. *Drugs* **59(6)**, 1233–1250.

[297] Fontecave, M., Atta, M., and Mulliez, E. (2004) S-adenosylmethionine: nothing goes to waste. *Trends Biochem Sci* **29(5)**, 243–249.

[298] Taylor, F. R. and Cronan, J. E. J. (1979) Cyclopropane fatty acid synthase of Escherichia coli. Stabilization, purification, and interaction with phospholipid vesicles. *Biochemistry* **18(15)**, 3292–3300.

[299] Grogan, D. W. and Cronan, J. E. J. (1997) Cyclopropane ring formation in membrane lipids of bacteria. *Microbiol Mol Biol Rev* **61(4)**, 429–441.

[300] Stoner, G. L. and Eisenberg, M. A. (1975) Purification and properties of 7, 8-diaminopelargonic acid aminotransferase. *J Biol Chem* **250(11)**, 4029–4036.

[301] Stoner, G. L. and Eisenberg, M. A. (1975) Biosynthesis of 7, 8-diaminopelargonic acid from 7-keto-8-aminopelargonic acid and S-adenosyl-L-methionine. The kinetics of the reaction. *J Biol Chem* **250(11)**, 4037–4043.

[302] Iwata-Reuyl, D. (2003) Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA. *Bioorg Chem* **31(1)**, 24–43.

[303] Van Lanen, S. G., Kinzie, S. D., Matthieu, S., Link, T., Culp, J., and Iwata-Reuyl, D. (2003) tRNA modification by S-adenosylmethionine:tRNA ribosyltransferase-isomerase. Assay development and characterization of the recombinant enzyme. *J Biol Chem* **278(12)**, 10491–10499.

[304] Van Lanen, S. G. and Iwata-Reuyl, D. (2003) Kinetic mechanism of the tRNA-modifying enzyme S-adenosylmethionine:tRNA ribosyltransferase-isomerase (QueA). *Biochemistry* **42(18)**, 5312–5320.

[305] Bowman, W. H., Tabor, C. W., and Tabor, H. (1973) Spermidine biosynthesis. Purification and properties of propylamine transferase from Escherichia coli. *J Biol Chem* **248(7)**, 2480–2486.

[306] Zappia, V., Cacciapuoti, G., Pontoni, G., and Oliva, A. (1980) Mechanism of propylamine-transfer reactions. Kinetic and inhibition studies on spermidine synthase from Escherichia coli. *J Biol Chem* **255(15)**, 7276–7280.

[307] Nishimura, S., Taya, Y., Kuchino, Y., and Oashi, Z. (1974) Enzymatic synthesis of 3-(3-amino-3-carboxypropyl)uridine in Escherichia coli phenylalanine transfer RNA: transfer of the 3-amino-acid-3-carboxypropyl group from S-adenosylmethionine. *Biochem Biophys Res Commun* **57(3)**, 702–708.

[308] Kowalak, J. A., Bruenger, E., Crain, P. F., and McCloskey, J. A. (2000) Identities and phylogenetic comparisons of posttranscriptional modifications in 16 S ribosomal RNA from Haloferax volcanii. *J Biol Chem* **275(32)**, 24484–24489.

[309] Cheek, J. and Broderick, J. (2001) Adenosylmethionine-dependent iron-sulfur enzymes: versatile clusters in a radical new role *Journal of Biological Inorganic Chemistry* **6(3)**, 209–226.

[310] Frey, P. and Magnusson, O. (2003) S-Adenosylmethionine: a wolf in sheep's clothing, or a rich man's adenosylcobalamin ? *Chem Rev* **103(6)**, 2129–48.

[311] Jarrett, J. T. (2003) The generation of 5'-deoxyadenosyl radicals by adenosylmethionine-dependent radical enzymes. *Curr Opin Chem Biol* **7(2)**, 174–182.

[312] Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W., and Roberts, R. J. (1993) Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell* **74(2)**, 299–307.

[313] Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature* **250(463)**, 194–199.

[314] Skinner, M. M., Puvathingal, J. M., Walter, R. L., and Friedman, A. M. (2000) Crystal structure of protein isoaspartyl methyltransferase: a catalyst for protein repair. *Structure* **8(11)**, 1189–1201.

[315] Zhang, X., Zhou, L., and Cheng, X. (2000) Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *EMBO J* **19(14)**, 3509–3519.

[316] Cheng, X. (1995) Structure and function of DNA methyltransferases. *Annu Rev Biophys Biomol Struct* **24**, 293–318.

[317] Dixon, M. M., Huang, S., Matthews, R. G., and Ludwig, M. (1996) The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12. *Structure* **4(11)**, 1263–1275.

[318] Schubert, H. L., Wilson, K. S., Raux, E., Woodcock, S. C., and Warren, M. J. (1998) The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase. *Nat Struct Biol* **5(7)**, 585–592.

[319] Wilson, J. R., Jing, C., Walker, P. A., Martin, S. R., Howell, S. A., Blackburn, G. M., Gamblin, S. J., and Xiao, B. (2002) Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell* **111(1)**, 105–115.

[320] Zhang, X., Tamaru, H., Khan, S. I., Horton, J. R., Keefe, L. J., Selker, E. U., and Cheng, X. (2002) Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111(1)**, 117–127.

[321] Trievel, R. C., Beach, B. M., Dirk, L. M. A., Houtz, R. L., and Hurley, J. H. (2002) Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* **111(1)**, 91–103.

[322] Jacobs, S. A., Harp, J. M., Devarakonda, S., Kim, Y., Rastinejad, F., and Khorasanizadeh, S. (2002) The active site of the SET domain is constructed on a knot. *Nat Struct Biol* **9(11)**, 833–838.

[323] Min, J., Zhang, X., Cheng, X., Grewal, S. I. S., and Xu, R.-M. (2002) Structure of the SET domain histone lysine methyltransferase Clr4. *Nat Struct Biol* **9(11)**, 828–832.

[324] Michel, G., Sauve, V., Larocque, R., Li, Y., Matte, A., and Cygler, M. (2002) The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot. *Structure* **10(10)**, 1303–1315.

[325] Nureki, O., Shirouzu, M., Hashimoto, K., Ishitani, R., Terada, T., Tamakoshi, M., Oshima, T., Chijimatsu, M., Takio, K., Vassylyev, D. G., Shibata, T., Inoue, Y., Kuramitsu, S., and Yokoyama, S. (2002) An enzyme with a deep trefoil knot for the active-site architecture. *Acta Crystallogr D Biol Crystallogr* **58(Pt 7)**, 1129–1137.

[326] Stroupe, M. E., Leech, H. K., Daniels, D. S., Warren, M. J., and Getzoff, E. D. (2003) CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis. *Nat Struct Biol* **10(12)**, 1064–1073.

[327] Vevodova, J., Graham, R. M., Raux, E., Schubert, H. L., Roper, D. I., Brindley, A. A., Ian Scott, A., Roessner, C. A., Stamford, N. P. J., Elizabeth Stroupe, M., Getzoff, E. D., Warren, M. J., and Wilson, K. S. (2004) Structure/function studies on a S-adenosyl-L-methionine-dependent uroporphyrinogen III C methyltransferase (SUMT), a key regulatory enzyme of tetrapyrrole biosynthesis. *J Mol Biol* **344(2)**, 419–433.

[328] Frank, S., Deery, E., Brindley, A. A., Leech, H. K., Lawrence, A., Heathcote, P., Schubert, H. L., Brocklehurst, K., Rigby, S. E. J., Warren, M. J., and Pickersgill, R. W. (2007) Elucidation of substrate specificity in the cobalamin (vitamin B12) biosynthetic methyltransferases. Structure and function of the C20 methyltransferase (CbiL) from Methanothermobacter thermautotrophicus. *J Biol Chem* **282(33)**, 23957–23969.

[329] Wada, K., Harada, J., Yaeda, Y., Tamiaki, H., Oh-Oka, H., and Fukuyama, K. (2007) Crystal structures of CbiL, a methyltransferase involved in anaerobic vitamin B biosynthesis, and CbiL in complex with S-adenosylhomocysteine–implications for the reaction mechanism. *FEBS J* **274(2)**, 563–573.

[330] Keller, J. P., Smith, P. M., Benach, J., Christendat, D., deTitta, G. T., and Hunt, J. F. (2002) The crystal structure of MT0146/CbiT suggests that the putative precorrin-8w decarboxylase is a methyltransferase. *Structure* **10(11)**, 1475–1487.

[331] Anantharaman, V., Koonin, E. V., and Aravind, L. (2002) SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J Mol Microbiol Biotechnol* **4(1)**, 71–75.

[332] Ahn, H. J., Kim, H.-W., Yoon, H.-J., Lee, B. I., Suh, S. W., and Yang, J. K. (2003) Crystal structure of tRNA(m1G37)methyltransferase: insights into tRNA recognition. *EMBO J* **22(11)**, 2593–2603.

[333] Kozbial, P. Z. and Mushegian, A. R. (2005) Natural history of S-adenosylmethionine-binding proteins. *BMC Struct Biol* **5**, 19.

[334] Yeates, T. O. (2002) Structures of SET domain proteins: protein lysine methyltransferases make their mark. *Cell* **111(1)**, 5–7.

[335] Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, I. A., Blackburn, G. M., and Gamblin, S. J. (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421(6923)**, 652–656.

[336] Manzur, K. L., Farooq, A., Zeng, L., Plotnikova, O., Koch, A. W., and Zhou, M.-M. (2003) A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. *Nat Struct Biol* **10(3)**, 187–196.

[337] Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D. J., Kitabayashi, I., Herr, W., and Cleary, M. L. (2004) Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol* **24(13)**, 5639–5649.

[338] Kouskouti, A., Scheer, E., Staub, A., Tora, L., and Talianidis, I. (2004) Gene-specific modulation of TAF10 function by SET9-mediated methylation. *Mol Cell* **14(2)**, 175–182.

[339] Chuikov, S., Kurash, J. K., Wilson, J. R., Xiao, B., Justin, N., Ivanov, G. S., McKinney, K., Tempst, P., Prives, C., Gamblin, S. J., Barlev, N. A., and Reinberg, D. (2004) Regulation of p53 activity through lysine methylation. *Nature* **432(7015)**, 353–360.

[340] Durban, E., Nochumson, S., Kim, S., Paik, W. K., and Chan, S. K. (1978) Cytochrome c-specific protein-lysine methyltransferase from Neurospora crassa. Purification, characterization, and substrate requirements. *J Biol Chem* **253(5)**, 1427–1435.

[341] Polevoda, B., Martzen, M. R., Das, B., Phizicky, E. M., and Sherman, F. (2000) Cytochrome c methyltransferase, Ctm1p, of yeast. *J Biol Chem* **275(27)**, 20508–20513.

[342] Lim, K., Zhang, H., Tempczyk, A., Krajewski, W., Bonander, N., Toedt, J., Howard, A., Eisenstein, E., and Herzberg, O. (2003) Structure of the YibK methyltransferase from Haemophilus influenzae (HI0766): a cofactor bound at a site formed by a knot. *Proteins* **51(1)**, 56–67.

[343] Posfai, J., Bhagwat, A. S., Posfai, G., and Roberts, R. J. (1989) Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res* **17(7)**, 2421–2435.

[344] Malone, T., Blumenthal, R. M., and Cheng, X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J Mol Biol* **253(4)**, 618–632.

[345] Bujnicki, J. M. (2002) Sequence permutations in the molecular evolution of DNA methyltransferases. *BMC Evol Biol* **2**, 3.

[346] Pues, H., Bleimling, N., Holz, B., Wolcke, J., and Weinhold, E. (1999) Functional roles of the conserved aromatic amino acid residues at position 108 (motif IV) and position 196 (motif VIII) in base flipping and catalysis by the N6-adenine DNA methyltransferase from Thermus aquaticus. *Biochemistry* **38(5)**, 1426–1434.

[347] Roth, M., Helm-Kruse, S., Friedrich, T., and Jeltsch, A. (1998) Functional roles of conserved amino acid residues in DNA methyltransferases investigated by site-directed mutagenesis of the EcoRV adenine-N6-methyltransferase. *J Biol Chem* **273(28)**, 17333–17342.

[348] Liu, Y. and Santi, D. V. (2000) m5C RNA and m5C DNA methyl transferases use different cysteine residues as catalysts. *Proc Natl Acad Sci U S A* **97(15)**, 8263–8265.

[349] Walbott, H., Husson, C., Auxilien, S., and Golinelli-Pimpaneau, B. (2007) Cysteine of sequence motif VI is essential for nucleophilic catalysis by yeast tRNA m5C methyltransferase. *RNA* **13(7)**, 967–973.

[350] Barraud, P., Golinelli-Pimpaneau, B., Atmanene, C., Sanglier, S., Van Dorsselaer, A., Droogmans, L., Dardel, F., and Tisne, C. (2008) Crystal structure of Thermus thermophilus tRNA m1A58 methyltransferase and biophysical characterization of its interaction with tRNA. *J Mol Biol* **377(2)**, 535–550.

[351] Santi, D. V., Garrett, C. E., and Barr, P. J. (1983) On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell* **33(1)**, 9–10.

[352] Santi, D. V., Norment, A., and Garrett, C. E. (1984) Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. *Proc Natl Acad Sci U S A* **81(22)**, 6993–6997.

[353] Wu, J. C. and Santi, D. V. (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. *J Biol Chem* **262(10)**, 4778–4786.

[354] Chen, L., MacMillan, A. M., Chang, W., Ezaz-Nikpay, K., Lane, W. S., and Verdine, G. L. (1991) Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry* **30(46)**, 11018–11025.

[355] Friedman, S. and Ansari, N. (1992) Binding of the EcoRII methyltransferase to 5-fluorocytosine-containing DNA. Isolation of a bound peptide. *Nucleic Acids Res* **20(12)**, 3241–3248.

[356] King, M., Ton, D., and Redman, K. L. (1999) A conserved motif in the yeast nucleolar protein Nop2p contains an essential cysteine residue. *Biochem J* **337 (Pt 1)**, 29–35.

[357] King, M. Y. and Redman, K. L. (2002) RNA methyltransferases utilize two cysteine residues in the formation of 5-methylcytosine. *Biochemistry* **41(37)**, 11218–11225.

[358] Gong, W., O'Gara, M., Blumenthal, R. M., and Cheng, X. (1997) Structure of pvu II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Res* **25(14)**, 2702–2715.

[359] Scavetta, R. D., Thomas, C. B., Walsh, M. A., Szegedi, S., Joachimiak, A., Gumport, R. I., and Churchill, M. E. (2000) Structure of RsrI methyltransferase, a member of the N6-adenine beta class of DNA methyltransferases. *Nucleic Acids Res* **28(20)**, 3950–3961.

[360] Willcock, D. F., Dryden, D. T., and Murray, N. E. (1994) A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J* **13(16)**, 3902–3908.

[361] Kong, H. and Smith, C. L. (1997) Substrate DNA and cofactor regulate the activities of a multi-functional restriction-modification enzyme, BcgI. *Nucleic Acids Res* **25(18)**, 3687–3692.

[362] Friedrich, T., Roth, M., Helm-Kruse, S., and Jeltsch, A. (1998) Functional mapping of the EcoRV DNA methyltransferase by random mutagenesis and screening for catalytically inactive mutants. *Biol Chem* **379(4-5)**, 475–480.

[363] Roth, M. and Jeltsch, A. (2001) Changing the target base specificity of the EcoRV DNA methyltransferase by rational de novo protein-design. *Nucleic Acids Res* **29(15)**, 3137–3144.

[364] Pintard, L., Lecointe, F., Bujnicki, J. M., Bonnerot, C., Grosjean, H., and Lapeyre, B. (2002) Trm7p catalyses the formation of two 2'-O-methylriboses in yeast tRNA anticodon loop. *EMBO J* **21(7)**, 1811–1820.

[365] Watanabe, K., Nureki, O., Fukai, S., Ishii, R., Okamoto, H., Yokoyama, S., Endo, Y., and Hori, H. (2005) Roles of conserved amino acid sequence motifs in the SpoU (TrmH) RNA methyltransferase family. *J Biol Chem* **280(11)**, 10368–10377.

[366] Purta, E., vanVliet, F., Tkaczuk, K. L., Dunin-Horkawicz, S., Mori, H., Droogmans, L., and Bujnicki, J. M. (2006) The yfhQ gene of Escherichia coli encodes a tRNA:Cm32/Um32 methyltransferase. *BMC Mol Biol* **7**, 23.

[367] Tkaczuk, K. L., Dunin-Horkawicz, S., Purta, E., and Bujnicki, J. M. (2007) Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinformatics* **8**, 73.

[368] Elkins, P. A., Watts, J. M., Zalacain, M., vanThiel, A., Vitazka, P. R., Redlak, M., Andraos-Selim, C., Rastinejad, F., and Holmes, W. M. (2003) Insights into catalysis by a knotted TrmD tRNA methyltransferase. *J Mol Biol* **333(5)**, 931–949.

[369] Cavaille, J., Chetouani, F., and Bachellerie, J. P. (1999) The yeast Saccharomyces cerevisiae YDL112w ORF encodes the putative 2'-O-ribose methyltransferase catalyzing the formation of Gm18 in tRNAs. *RNA* **5(1)**, 66–81.

[370] Nureki, O., Watanabe, K., Fukai, S., Ishii, R., Endo, Y., Hori, H., and Yokoyama, S. (2004) Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme. *Structure* **12(4)**, 593–602.

[371] Santi, D. V. and Hardy, L. W. (1987) Catalytic mechanism and inhibition of tRNA (uracil-5-)methyltransferase: evidence for covalent catalysis. *Biochemistry* **26(26)**, 8599–8606.

[372] Kealey, J. T. and Santi, D. V. (1991) Identification of the catalytic nucleophile of tRNA (m5U54)methyltransferase. *Biochemistry* **30(40)**, 9724–9728.

[373] Dunn, D. B. (1961) The occurrence of 1-methyladenine in ribonucleic acid. *Biochim Biophys Acta* **46**, 198–200.

[374] Dunn, D. (1963) The isolation of 1-methyladenylic acid and 7-methylguanylic acid from ribonucleic acid *Biochem J* **86**, 14P–15P.

[375] Anderson, J. and Droogmans, L. (2005) Biosynthesis and function of 1-methyladenosine in transfer RNA In H. Grosjean, (ed.), Fine-tuning of RNA Functions by Modification and Editing, pp. 121–139 Springer-Verlag.

[376] Bjork, G. R., Ericson, J. U., Gustafsson, C. E., Hagervall, T. G., Jonsson, Y. H., and Wikstrom, P. M. (1987) Transfer RNA modification. *Annu Rev Biochem* **56**, 263–287.

[377] Bujnicki, J. M. (2001) In silico analysis of the tRNA:m1A58 methyltransferase family: homology-based fold prediction and identification of new members from Eubacteria and Archaea. *FEBS Lett* **507(2)**, 123–127.

[378] Yamazaki, N., Hori, H., Ozawa, K., Nakanishi, S., Ueda, T., Kumagai, I., Watanabe, K., and Nishikawa, K. (1992) Purification and characterization of tRNA(adenosine-1-)-methyltransferase from Thermus thermophilus HB27. *Nucleic Acids Symp Ser* **27**, 141–142.

[379] 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(5)**, 1308–1315.

[380] Glick, J. M. and Leboy, P. S. (1977) Purification and properties of tRNA(adenine-1)-methyltransferase from rat liver. *J Biol Chem* **252(14)**, 4790–4795.

[381] Anderson, J., Phan, L., Cuesta, R., Carlson, B. A., Pak, M., Asano, K., Bjork, G. R., Tamame, M., and Hinnebusch, A. G. (1998) The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes Dev* **12(23)**, 3650–3662.

[382] Anderson, J., Phan, L., and Hinnebusch, A. G. (2000) The Gcd10p/Gcd14p complex is the essential two-subunit tRNA(1-methyladenosine) methyltransferase of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **97(10)**, 5173–5178.

[383] Ozanick, S., Krecic, A., Andersland, J., and Anderson, J. T. (2005) The bipartite structure of the tRNA m1A58 methyltransferase from S. cerevisiae is conserved in humans. *RNA* **11(8)**, 1281–1290.

[384] Roovers, M., Wouters, J., Bujnicki, J. M., Tricot, C., Stalon, V., Grosjean, H., and Droogmans, L. (2004) A primordial RNA modification enzyme: the case of tRNA (m1A) methyltransferase. *Nucleic Acids Res* **32(2)**, 465–476.

[385] Ozanick, S. G., Bujnicki, J. M., Sem, D. S., and Anderson, J. T. (2007) Conserved amino acids in each subunit of the heteroligomeric tRNA m1A58 Mtase from Saccharomyces cerevisiae contribute to tRNA binding. *Nucleic Acids Res* **35(20)**, 6808–6819.

[386] Varshney, U., Ramesh, V., Madabushi, A., Gaur, R., Subramanya, H. S., and RajBhandary, U. L. (2004) Mycobacterium tuberculosis Rv2118c codes for a single-component homotetrameric m1A58 tRNA methyltransferase. *Nucleic Acids Res* **32(3)**, 1018–1027.

[387] Gupta, A., Kumar, P. H., Dineshkumar, T. K., Varshney, U., and Subramanya, H. S. (2001) Crystal structure of Rv2118c: an AdoMet-dependent methyltransferase from Mycobacterium tuberculosis H37Rv. *J Mol Biol* **312(2)**, 381–391.

[388] Li, X., Martin, C., Urbonavicius, J., Roovers, M., Wouters, J., Barraud, P., Tisne, C., Droogmans, L., and Golinelli-Pimpaneau, B. (2008) Crystal structure of the region-specific P. abyssi tRNA m1A58 methyltransferase. *Manuscript in preparation.*

[389] Grosjean, H., Edqvist, J., Straby, K. B., and Giege, R. (1996) Enzymatic formation of modified nucleosides in tRNA: dependence on tRNA architecture. *J Mol Biol* **255(1)**, 67–85.

[390] 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(21)**, 4312–4319.

[391] Grosjean, H., Auxilien, S., Constantinesco, F., Simon, C., Corda, Y., Becker, H. F., Foiret, D., Morin, A., Jin, Y. X., Fournier, M., and Fourrey, J. L. (1996) Enzymatic conversion of adenosine to inosine and to N1-methylinosine in transfer RNAs: a review. *Biochimie* **78(6)**, 488–501.

[392] Yokoyama, S., Watanabe, K., and Miyazawa, T. (1987) Dynamic structures and functions of transfer ribonucleic acids from extreme thermophiles. *Adv Biophys* **23**, 115–147.

[393] Kowalak, J. A., Dalluge, J. J., McCloskey, J. A., and Stetter, K. O. (1994) The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles. *Biochemistry* **33(25)**, 7869–7876.

[394] Watanabe, K., Shinma, M., Oshima, T., and Nishimura, S. (1976) Heat-induced stability of tRNA from an extreme thermophile, Thermus thermophilus. *Biochem Biophys Res Commun* **72(3)**, 1137–1144.

[395] Horie, N., Hara-Yokoyama, M., Yokoyama, S., Watanabe, K., Kuchino, Y., Nishimura, S., and Miyazawa, T. (1985) Two tRNAIle1 species from an extreme thermophile, Thermus thermophilus HB8: effect of 2-thiolation of ribothymidine on the thermostability of tRNA. *Biochemistry* **24(21)**, 5711–5715.

[396] Kadaba, S., Krueger, A., Trice, T., Krecic, A. M., Hinnebusch, A. G., and Anderson, J. (2004) Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae. *Genes Dev* **18(11)**, 1227–1240.

[397] Anderson, J. T. (2005) RNA turnover: unexpected consequences of being tailed. *Curr Biol* **15(16)**, R635–8.

[398] Wang, X., Jia, H., Jankowsky, E., and Anderson, J. T. (2008) Degradation of hypomodified tRNA(iMet) in vivo involves RNA-dependent ATPase activity of the DExH helicase Mtr4p. *RNA* **14(1)**, 107–116.

[399] Draper, D. E. (1995) Protein-RNA recognition. *Annu Rev Biochem* **64**, 593–620.

[400] Nagai, K. (1996) RNA-protein complexes. *Curr Opin Struct Biol* **6(1)**, 53–61.

[401] Lunde, B. M., Moore, C., and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* **8(6)**, 479–490.

[402] Maris, C., Dominguez, C., and Allain, F. H.-T. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* **272(9)**, 2118–2131.

[403] Hoang, C. and Ferre-D'Amare, A. R. (2001) Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **107(7)**, 929–939.

[404] Pan, H., Agarwalla, S., Moustakas, D. T., Finer-Moore, J., and Stroud, R. M. (2003) Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proc Natl Acad Sci U S A* **100(22)**, 12648–12653.

[405] Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S., and Yokoyama, S. (2003) Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell* **113(3)**, 383–394.

[406] Xie, W., Liu, X., and Huang, R. H. (2003) Chemical trapping and crystal structure of a catalytic tRNA guanine transglycosylase covalent intermediate. *Nat Struct Biol* **10(10)**, 781–788.

[407] Lee, T. T., Agarwalla, S., and Stroud, R. M. (2005) A unique RNA Fold in the RumA-RNA-cofactor ternary complex contributes to substrate selectivity and enzymatic function. *Cell* **120(5)**, 599–611.

[408] Losey, H. C., Ruthenburg, A. J., and Verdine, G. L. (2006) Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nat Struct Mol Biol* **13(2)**, 153–159.

[409] Numata, T., Ikeuchi, Y., Fukai, S., Suzuki, T., and Nureki, O. (2006) Snapshots of tRNA sulphuration via an adenylated intermediate. *Nature* **442(7101)**, 419–424.

[410] Hoang, C., Chen, J., Vizthum, C. A., Kandel, J. M., Hamilton, C. S., Mueller, E. G., and Ferre-D'Amare, A. R. (2006) Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure. *Mol Cell* **24(4)**, 535–545.

[411] Alian, A., Lee, T. T., Griner, S. L., Stroud, R. M., and Finer-Moore, J. (2008) Structure of a TrmA-RNA complex: A consensus RNA fold contributes to substrate selectivity and catalysis in m5U methyltransferases. *Proc Natl Acad Sci U S A* **105(19)**, 6876–6881.

[412] Aravind, L. and Koonin, E. V. (1999) Novel predicted RNA-binding domains associated with the translation machinery. *J Mol Evol* **48(3)**, 291–302.

[413] Aravind, L. and Koonin, E. V. (2001) THUMP–a predicted RNA-binding domain shared by 4-thiouridine, pseudouridine synthases and RNA methylases. *Trends Biochem Sci* **26(4)**, 215–217.

[414] Anantharaman, V., Koonin, E. V., and Aravind, L. (2001) TRAM, a predicted RNA-binding domain, common to tRNA uracil methylation and adenine thiolation enzymes. *FEMS Microbiol Lett* **197(2)**, 215–221.

[415] Gabant, G., Auxilien, S., Tuszynska, I., Locard, M., Gajda, M. J., Chaussinand, G., Fernandez, B., Dedieu, A., Grosjean, H., Golinelli-Pimpaneau, B., Bujnicki, J. M., and Armengaud, J. (2006) THUMP from archaeal tRNA:m22G10 methyltransferase, a genuine autonomously folding domain. *Nucleic Acids Res* **34(9)**, 2483–2494.

[416] Sabina, J. and Soll, D. (2006) The RNA-binding PUA domain of archaeal tRNA-guanine transglycosylase is not required for archaeosine formation. *J Biol Chem* **281(11)**, 6993–7001.

[417] Woo, N. H., Roe, B. A., and Rich, A. (1980) Three-dimensional structure of Escherichia coli initiator tRNAfMet. *Nature* **286(5771)**, 346–351.

[418] Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125(7)**, 1731–1737.

[419] deVries, S. J., vanDijk, A. D. J., Krzeminski, M., vanDijk, M., Thureau, A., Hsu, V., Wassenaar, T., and Bonvin, A. M. J. J. (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* **69(4)**, 726–733.

[420] Stewart, R. and Harris, M. (1977) Amino group acidity in nucleotide bases *Canadian Journal of Chemistry* **55(21)**, 3807–3814.

[421] Kettani, A., Gueron, M., and Leroy, J. (1997) Amino Proton Exchange Processes in Mononucleosides *Journal of the American Chemical Society* **119(5)**, 1108–1115.

[422] Jeltsch, A. (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem* **3(4)**, 274–293.

[423] Marintchev, A. and Wagner, G. (2004) Translation initiation: structures, mechanisms and evolution. *Q Rev Biophys* **37(3-4)**, 197–284.

[424] Laursen, B. S., Sorensen, H. P., Mortensen, K. K., and Sperling-Petersen, H. U. (2005) Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* **69(1)**, 101–123.

[425] Shine, J. and Dalgarno, L. (1974) The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A* **71(4)**, 1342–1346.

[426] Yusupova, G. Z., Yusupov, M. M., Cate, J. H., and Noller, H. F. (2001) The path of messenger RNA through the ribosome. *Cell* **106(2)**, 233–241.

[427] La Teana, A., Gualerzi, C. O., and Brimacombe, R. (1995) From stand-by to decoding site. Adjustment of the mRNA on the 30S ribosomal subunit under the influence of the initiation factors. *RNA* **1(8)**, 772–782.

[428] Tomsic, J., Vitali, L. A., Daviter, T., Savelsbergh, A., Spurio, R., Striebeck, P., Wintermeyer, W., Rodnina, M. V., and Gualerzi, C. O. (2000) Late events of translation initiation in bacteria: a kinetic analysis. *EMBO J* **19(9)**, 2127–2136.

[429] Dahlquist, K. D. and Puglisi, J. D. (2000) Interaction of translation initiation factor IF1 with the E. coli ribosomal A site. *J Mol Biol* **299(1)**, 1–15.

[430] Carter, A. P., Clemons, W. M. J., Brodersen, D. E., Morgan-Warren, R. J., Hartsch, T., Wimberly, B. T., and Ramakrishnan, V. (2001) Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* **291(5503)**, 498–501.

[431] Gualerzi, C., Risuleo, G., and Pon, C. L. (1977) Initial rate kinetic analysis of the mechanism of initiation complex formation and the role of initiation factor IF-3. *Biochemistry* **16(8)**, 1684–1689.

[432] Pon, C. L. and Gualerzi, C. O. (1984) Mechanism of protein biosynthesis in prokaryotic cells. Effect of initiation factor IF1 on the initial rate of 30 S initiation complex formation. *FEBS Lett* **175(2)**, 203–207.

[433] RajBhandary, U. L. (1994) Initiator transfer RNAs. *J Bacteriol* **176(3)**, 547–552.

[434] RajBhandary, U. and Chow, C. (1995) Initiator tRNAs and initiation of protein synthesis In D. Söll and U.L. RajBhandary, (ed.), tRNA: Structure, Biosynthesis, and Function, pp. 511–528 American Society for Microbiology Washington, DC.

[435] Sundari, R. M., Stringer, E. A., Schulman, L. H., and Maitra, U. (1976) Interaction of bacterial initiation factor 2 with initiator tRNA. *J Biol Chem* **251(11)**, 3338–3345.

[436] Seong, B. L. and RajBhandary, U. L. (1987) Mutants of Escherichia coli formylmethionine tRNA: a single base change enables initiator tRNA to act as an elongator in vitro. *Proc Natl Acad Sci U S A* **84(24)**, 8859–8863.

[437] Guillon, J. M., Mechulam, Y., Blanquet, S., and Fayat, G. (1993) Importance of formylability and anticodon stem sequence to give a tRNA(Met) an initiator identity in Escherichia coli. *J Bacteriol* **175(14)**, 4507–4514.

[438] Guillon, J. M., Heiss, S., Soutourina, J., Mechulam, Y., Laalami, S., Grunberg-Manago, M., and Blanquet, S. (1996) Interplay of methionine tRNAs with translation elongation factor Tu and translation initiation factor 2 in Escherichia coli. *J Biol Chem* **271(37)**, 22321–22325.

[439] Wu, X. Q. and RajBhandary, U. L. (1997) Effect of the amino acid attached to Escherichia coli initiator tRNA on its affinity for the initiation factor IF2 and on the IF2 dependence of its binding to the ribosome. *J Biol Chem* **272(3)**, 1891–1895.

[440] Guillon, J. M., Meinnel, T., Mechulam, Y., Lazennec, C., Blanquet, S., and Fayat, G. (1992) Nucleotides of tRNA governing the specificity of Escherichia coli methionyl-tRNA(fMet) formyltransferase. *J Mol Biol* **224(2)**, 359–367.

[441] Dutka, S., Meinnel, T., Lazennec, C., Mechulam, Y., and Blanquet, S. (1993) Role of the 1-72 base pair in tRNAs for the activity of Escherichia coli peptidyl-tRNA hydrolase. *Nucleic Acids Res* **21(17)**, 4025–4030.

[442] Lee, C. P., Seong, B. L., and RajBhandary, U. L. (1991) Structural and sequence elements important for recognition of Escherichia coli formylmethionine tRNA by methionyl-tRNA transformylase are clustered in the acceptor stem. *J Biol Chem* **266(27)**, 18012–18017.

[443] Varshney, U., Lee, C. P., and RajBhandary, U. L. (1993) From elongator tRNA to initiator tRNA. *Proc Natl Acad Sci U S A* **90(6)**, 2305–2309.

[444] Wrede, P. and Rich, A. (1979) Stability of the unique anticodon loop conformation of E.coli tRNAfMet. *Nucleic Acids Res* **7(6)**, 1457–1467.

[445] Schweisguth, D. C. and Moore, P. B. (1997) On the conformation of the anticodon loops of initiator and elongator methionine tRNAs. *J Mol Biol* **267(3)**, 505–519.

[446] Mandal, N., Mangroo, D., Dalluge, J. J., McCloskey, J. A., and Rajbhandary, U. L. (1996) Role of the three consecutive G:C base pairs conserved in the anticodon stem of initiator tRNAs in initiation of protein synthesis in Escherichia coli. *RNA* **2(5)**, 473–482.

[447] Peattie, D. A. and Gilbert, W. (1980) Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci U S A* **77(8)**, 4679–4682.

[448] Garret, M., Labouesse, B., Litvak, S., Romby, P., Ebel, J. P., and Giege, R. (1984) Tertiary structure of animal tRNATrp in solution and interaction of tRNATrp with tryptophanyl-tRNA synthetase. *Eur J Biochem* **138(1)**, 67–75.

[449] Romby, P., Moras, D., Dumas, P., Ebel, J. P., and Giege, R. (1987) Comparison of the tertiary structure of yeast tRNA(Asp) and tRNA(Phe) in solution. Chemical modification study of the bases. *J Mol Biol* **195(1)**, 193–204.

[450] Noller, H. F., Hoang, L., and Fredrick, K. (2005) The 30S ribosomal P site: a function of 16S rRNA. *FEBS Lett* **579(4)**, 855–858.

[451] Korostelev, A. and Noller, H. F. (2007) The ribosome in focus: new structures bring new insights. *Trends Biochem Sci* **32(9)**, 434–441.

[452] Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. F. (1999) X-ray crystal structures of 70S ribosome functional complexes. *Science* **285(5436)**, 2095–2104.

[453] Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H. F. (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* **126(6)**, 1065–1077.

[454] Yusupova, G., Jenner, L., Rees, B., Moras, D., and Yusupov, M. (2006) Structural basis for messenger RNA movement on the ribosome. *Nature* **444(7117)**, 391–394.

[455] Berk, V., Zhang, W., Pai, R. D., and Cate, J. H. D. (2006) Structural basis for mRNA and tRNA positioning on the ribosome. *Proc Natl Acad Sci U S A* **103(43)**, 15830–15834.

[456] Hoang, L., Fredrick, K., and Noller, H. F. (2004) Creating ribosomes with an all-RNA 30S subunit P site. *Proc Natl Acad Sci U S A* **101(34)**, 12439–12443.

[457] Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci U S A* **98(9)**, 4899–4903.

[458] Battle, D. J. and Doudna, J. A. (2002) Specificity of RNA-RNA helix recognition. *Proc Natl Acad Sci U S A* **99(18)**, 11676–11681.

[459] Dallas, A. and Noller, H. F. (2001) Interaction of translation initiation factor 3 with the 30S ribosomal subunit. *Mol Cell* **8(4)**, 855–864.

[460] Lancaster, L. and Noller, H. F. (2005) Involvement of 16S rRNA nucleotides G1338 and A1339 in discrimination of initiator tRNA. *Mol Cell* **20(4)**, 623–632.

[461] Pon, C. L. and Gualerzi, C. (1974) Effect of initiation factor 3 binding on the 30S ribosomal subunits of Escherichia coli. *Proc Natl Acad Sci U S A* **71(12)**, 4950–4954.

[462] Hartz, D., McPheeters, D. S., and Gold, L. (1989) Selection of the initiator tRNA by Escherichia coli initiation factors. *Genes Dev* **3(12A)**, 1899–1912.

[463] McCutcheon, J. P., Agrawal, R. K., Philips, S. M., Grassucci, R. A., Gerchman, S. E., Clemons, W. M. J., Ramakrishnan, V., and Frank, J. (1999) Location of translational initiation factor IF3 on the small ribosomal subunit. *Proc Natl Acad Sci U S A* **96(8)**, 4301–4306.

[464] Shapkina, T. G., Dolan, M. A., Babin, P., and Wollenzien, P. (2000) Initiation factor 3-induced structural changes in the 30 S ribosomal subunit and in complexes containing tRNA(f)(Met) and mRNA. *J Mol Biol* **299(3)**, 615–628.

[465] Varshney, U. and RajBhandary, U. L. (1990) Initiation of protein synthesis from a termination codon. *Proc Natl Acad Sci U S A* **87(4)**, 1586–1590.

[466] Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H., and Westhof, E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res* **31(13)**, 3450–3460.

[467] Ji, X., Klarmann, G. J., and Preston, B. D. (1996) Effect of human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein on HIV-1 reverse transcriptase activity in vitro. *Biochemistry* **35(1)**, 132–143.

[468] Zubay, G. (1962) The isolation and fractionation of soluble ribonucleic acid *J Mol Biol* **4**, 347–356.

[469] Meinnel, T., Mechulam, Y., and Fayat, G. (1988) Fast purification of a functional elongator tRNAmet expressed from a synthetic gene in vivo. *Nucleic Acids Res* **16(16)**, 8095–8096.

[470] Meinnel, T. and Blanquet, S. (1995) Maturation of pre-tRNA(fMet) by Escherichia coli RNase P is specified by a guanosine of the 5'-flanking sequence. *J Biol Chem* **270(26)**, 15908–15914.

Annexes

Classification des paires de bases dans l'ARN

Les faces d'interactions

Dans les ARN, outre les interactions canoniques G=C et A–U, où les nucléotides interagissent avec leur face Watson-Crick, on trouve de nombreuses interactions non-canoniques, faisant intervenir d'autres faces d'interaction que l'incontournable face Watson-Crick. Ces faces d'interactions sont au nombre de trois et sont présentées à la FIGURE A.1 **a**. Les trois faces d'interactions sont :

- La face Watson-Crick (*Watson-Crick Edge*) — définie en ce qui concerne les purines par les positions 1, 2 et 6 — et par les positions 2, 3 et 4 pour les pyrimidines.
- La face Hoogsteen (*Hoogsteen Edge*) — définie en ce qui concerne les purines par les positions 6 et 7 — et par les positions 4 et 5 pour les pyrimidines.
- La face Sucre (*Sugar Edge*) — définie en ce qui concerne les purines par les positions 2 et 3 — et par la position 2 pour les pyrimidines — auxquelles viennent s'ajouter la position 2′OH du ribose, tant pour les purines que pour les pyrimidines.

Les douze familles et les symboles associés

Etant donné ces trois faces d'interactions, on peut définir 12 familles de paires de bases (FIGURE A.2), en choisissant les faces d'interactions de chacun des nucléotides ainsi que l'orientation relative de leurs liaisons glycosidiques — *Cis* ou *Trans* (FIGURE A.1 **b**). Des symboles ont de plus été définis pour représenter ces paires de bases [1] [20]. Dans le même article, Leontis & Westhof proposent de noter les paires canoniques Watson-Crick G=C par deux traits fins, les paires canoniques Watson-Crick A–U par un trait fin et les paires « wobble » G∘U par un rond blanc. Etant donné la difficulté à insérer les symboles de la FIGURE A.2 dans un texte, les paires de bases non canoniques y sont souvent représentées par un simple rond noir « • ».

1. Ces notations sont utilisées par certains logiciels d'analyse de structure d'ARN, comme RNAView [466]. Ce programme est accessible à l'adresse suivante : `http://ndbserver.rutgers.edu/services/download/`

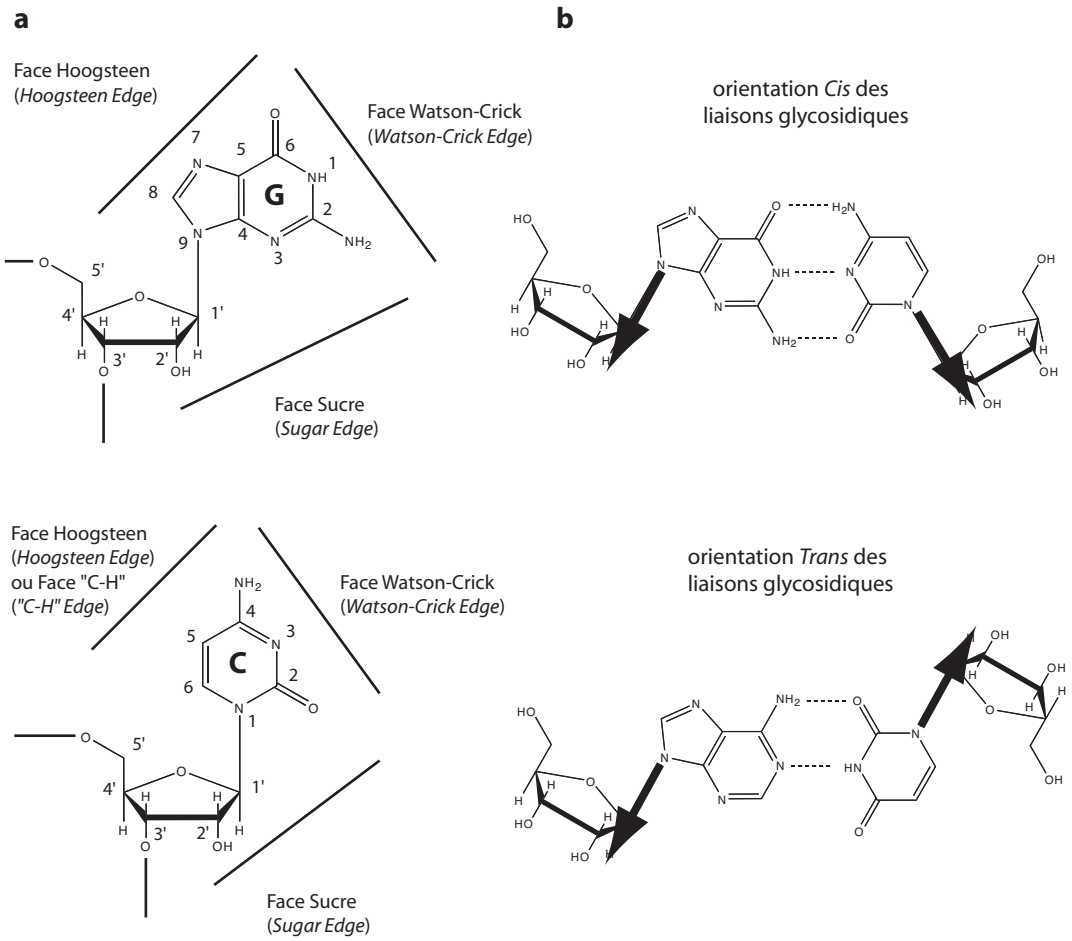


Figure A.1 – Les faces d'interactions des nucléotides

a : Les trois faces d'interactions des nucléotides sont présentées sur une guanosine (purine) et une cytosine (pyrimidine). Les trois faces sont la face Waston-Crick, la face Hoogsteen et la face Sucre. Chez les pyrimidines, la face Sucre est aussi appelée face « C–H ». **b** : Orientation *Cis* et *Trans* des liaisons glycosidiques.

No.	Orientation des liaisons glycosidiques	Faces d'interactions	Symbole
1	*Cis*	Watson-Crick / Watson-Crick	
2	*Trans*	Watson-Crick / Watson-Crick	
3	*Cis*	Watson-Crick / Hoogsteen	
4	*Trans*	Watson-Crick / Hoogsteen	
5	*Cis*	Watson-Crick / Sugar Edge	
6	*Trans*	Watson-Crick / Sugar Edge	
7	*Cis*	Hoogsteen / Hoogsteen	
8	*Trans*	Hoogsteen / Hoogsteen	
9	*Cis*	Hoogsteen / Sugar Edge	
10	*Trans*	Hoogsteen / Sugar Edge	
11	*Cis*	Sugar Edge / Sugar Edge	
12	*Trans*	Sugar Edge / Sugar Edge	

Figure A.2 – Les 12 principales familles de paires de bases et leur notation symbolique
Les numéros et les symboles correspondent à ceux définis dans Leontis & Westhof [20]. Chaque famille est définie par les deux faces en interaction ainsi que par l'orientation relative des liaisons glycosidiques.

Obtention des échantillons pour la RMN et la Cristallographie

Surproduction et purification de la protéine de nucléocapside NC_{1-55}

Le protocole de surproduction et de purification de la NC_{1-55} est adapté de protocoles existants [180, 186, 467]

Surproduction de la NC_{1-55}

Souches bactériennes et vecteur plasmidique

Le système pET de Novagen est utilisé pour la surproduction de la NC du VIH-1. Le gène d'intérêt *NCp7* est cloné dans le vecteur pET3a. Son expression est sous le contrôle du promoteur T7 et est induite par l'IPTG. Le plasmide pET3a–Ncp7 nous a été donné par l'équipe de Michael Summers (University of Maryland, Baltimore). Les cellules (*E. coli* BL21(DE3)pLysE) sont transformées avec le plasmide par la technique d'électroporation. Après le choc électrique, 1 mL de milieu LB est ajouté et les cellules sont incubées 1 h à 37 °C. 100 μL de cette suspension sont ensuite étalés sur boîte ZB agar (10 g/L Tryptone, 5 g/L NaCl, 15 g/L agar). La sélection des cellules se fait par l'ajout d'antibiotiques *ad hoc* : chloramphénicol (34 μg/mL) dont le gène de résistance est porté par le plasmide pLysE, et ampicilline (100 μg/mL) dont le gène de résistance est porté par le plasmide pET3a.

Test d'induction

Des cultures de 5 mL de M9ZB (10 g/L Tryptone, 5 g/L NaCl, 1 g/L NH_4Cl, 3 g/L KH_2PO_4, 6 g/L Na_2HPO_4, 0,4 % glucose, 1 mM $MgSO_4$) en présence de $ZnCl_2$ (0,1 mM), de chloramphénicol (34 μg/mL) et d'ampicilline (100 μg/mL) sont ensemencées à partir de six clones choisis sur la boîte ZB agar. L'induction se fait par l'ajout d'IPTG (1 mM final), lorsque l'absorbance à 600 nm atteint 0,5 u.a. Des aliquotes sont prélevées avant l'induction et 1, 2, 3 et 4 h après l'ajout d'IPTG pour contrôler la surproduction de la NC par les clones choisis et pour déterminer la durée optimale d'induction. Un clone dont la qualité de surproduction est jugé satisfaisante est choisi et sera utilisé lors des cultures postérieures.

Culture cellulaire

Les cellules provenant du clone sélectionné sont cultivées une nuit en milieu ZB en présence de $ZnCl_2$ (0,1 mM), de chloramphénicol (34 μg/mL) et d'ampicilline (100 μg/mL). 2 L de M9ZB sont ensemencés avec 200 mL de cette préculture. L'induction se fait par l'ajout d'IPTG (1 mM finale), lorsque la DO à 600 nm atteint 0,5 u.a. La croissance se poursuit alors pour une période de 3 h. La culture est centrifugée à 9000 g pendant 20 min à 4 ℃ (rotor JLA 8100, Beckman). Les culots sont rassemblés et resuspendus dans un volume de tampon de lyse équivalent à 1/50 du volume de culture (Tris-HCl pH 8,0 50 mM, glycérol 10 %, NaCl 0,1 M, $ZnCl_2$ 0,1 mM, DTT 5 mM, EDTA 2 mM). Les culots resuspendus sont ensuite placés à – 70 ℃.

Purification de la NC_{1-55}

Préparation de l'extrait protéique

Les cellules sont lentement décongelées sur la glace puis des inhibiteurs de protéases (86 μL de PMSF 10 mM et 15 μL de Pepstatine A 1 mg/mL) et 1 mL de désoxycholate de sodium 1 % (w/v) sont ajoutés par litre de culture. Les cellules sont soniquées à 4 ℃ pendant 5 min par cycles de 20 s entrecoupés de 40 s de pause. L'extrait obtenu est placé dans un erlenmeyer et 5 mL de polyethylène imine (4 % (v/v) dans le tampon de lyse, pH ajusté à 8,0 par ajout de HCl concentré) est ajouté goutte à goutte sous vive agitation. L'agitation est poursuivie à 4 ℃ pendant 45 min. Les acides nucléiques ainsi précipités et les débris cellulaires sont éliminés par une centrifugation à 48000 g pendant 30 min à 4 ℃ (rotor JA 20, Beckman). Le surnageant est filtré sur une membrane de 0,22 μm. Les étapes suivantes de purification par chromatographie sont réalisées à 4 ℃ sur un appareil FPLC ÄKTA Prime (Pharmacia Biotech).

Chromatographies échangeuses d'ions

Lors de la première étape de purification, deux colonnes échangeuses d'ions sont montées en série : une colonne Q–sépharose (Pharmacia Biotech, 50 mL) suivie d'une colonne SP-sépharose (Pharmacia Biotech, 50 mL). Les colonnes sont préalablement équilibrées avec le tampon A (Tris-HCl pH 8,0 50 mM, glycérol 10 %, NaCl 0,1 M, $ZnCl_2$ 0,1 mM, βME 10 mM). L'extrait brut est déposé sur la colonne Q–sépharose à raison de 2 mL/min. Après le passage de 150 mL de tampon A, la colonne Q–sépharose est détachée et un gradient de 600 mL est réalisé sur la colonne SP–sépharose entre 20 % et 50 % de tampon B (Tris-HCl pH 8,0 50 mM, glycérol 10 %, NaCl 1 M, $ZnCl_2$ 0,1 mM, βME 10 mM) à raison de 3 mL/min. La protéine de nucléocapside est éluée à 350 mM de NaCl.

Tamis moléculaire

La deuxième étape de purification consiste en un tamis moléculaire (Superdex 75, Pharmacia Biotech, 320 mL). La colonne est préalablement équilibrée

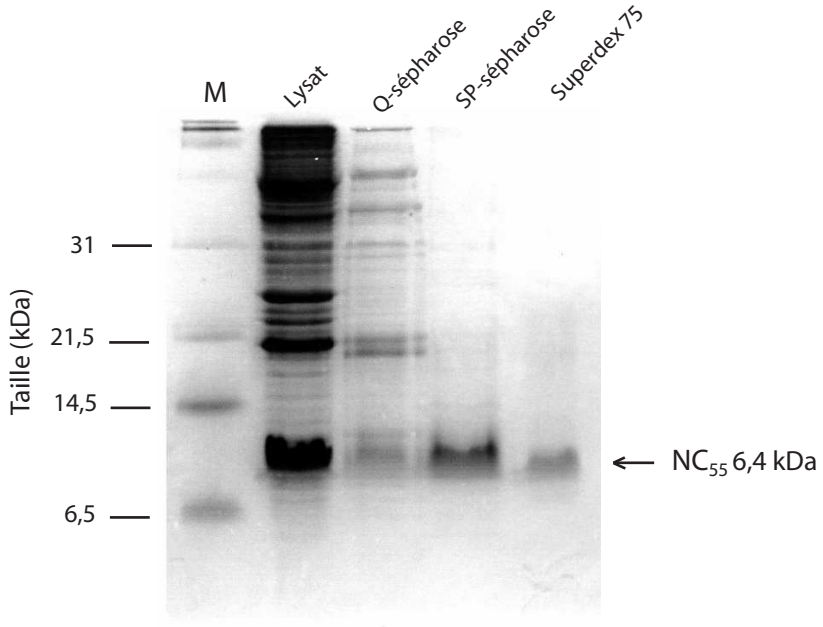


Figure B.1 – Etapes de purification de la NC_{1-55}

avec le tampon d'élution C (Tris-HCl pH 7,0 50 mM, glycérol 10 %, NaCl 0,1 M, $ZnCl_2$ 0,1 mM, βME 10 mM). La NC est éluée après 220 mL de tampon C. Les différentes étapes de purification sont présentées sur un gel acrylamide à la Figure B.1. 5 mg de NC purifiée sont généralement obtenus par litre de culture.

Surproduction et purification de TrmI

Le protocole de surproduction et de purification de TrmI est adapté de Droogmans *et al.* [111].

Surproduction de TrmI

La protéine TrmI a été surproduite dans la souche BL21(DE3)Rosetta (Novagen) de *E. coli* transformée par un plasmide pET15b portant le gène *trmI* ainsi qu'un gène de résistance à l'ampicilline. Le plasmide pET15b-*trmI* a été donné au laboratoire par l'équipe de Louis Droogmans [111] (Université Libre de Bruxelles, Belgique). L'expression de *trmI* est sous le contrôle du promoteur de T7 et est inductible par l'IPTG. La protéine produite porte une extension N-terminale avec une étiquette de 6 histidines (His_6). La souche BL21(DE3)Rosetta porte 6 gènes d'ARNt rares de *E. coli* sur un plasmide conférant également la résistance au chloramphénicol. Ceci permet une meilleure expression de la protéine TrmI dont le gène *trmI* provient d'un organisme thermophile (*T. thermophilus*) ne présentant pas la même utilisation des divers codons dégénérés du code génétique que *E. coli*.

Les cellules sont cultivées à 37 ℃ dans un milieu LB en présence de chloramphénicol (25 μg/mL) et d'ampicilline (100 μg/mL). L'induction se fait par l'ajout d'IPTG (1 mM final), lorsque la DO à 600 nm atteint 0,5 u.a. La croissance se poursuit alors pour une période de 3 h. La culture est centrifugée à 9000 g pendant 20 min à 4 ℃ (rotor JLA 8100, Beckman). Les culots sont rassemblés et suspendus dans un volume de tampon de lyse équivalent à 1/50 du volume de culture (Tris-HCl pH 8,5 50 mM, KCl 500 mM). Les culots suspendus sont ensuite placés à – 70 ℃.

Purification de TrmI

Les cellules sont décongelées sur la glace puis soniquées en présence de 0,1 % de désoxycholate de sodium à 4 ℃ pendant 5 min par cycles de 30 s entrecoupés de 30 s de pause. Les débris cellulaires sont éliminés par une centrifugation à 48000 g pendant 30 min à 4 ℃ (rotor JA 20, Beckman). Le surnageant est filtré sur une membrane de 0,22 μm. Les étapes suivantes de purification par chromatographie sont réalisées à température ambiante sur un appareil FPLC ÄKTA Prime (Pharmacia Biotech).

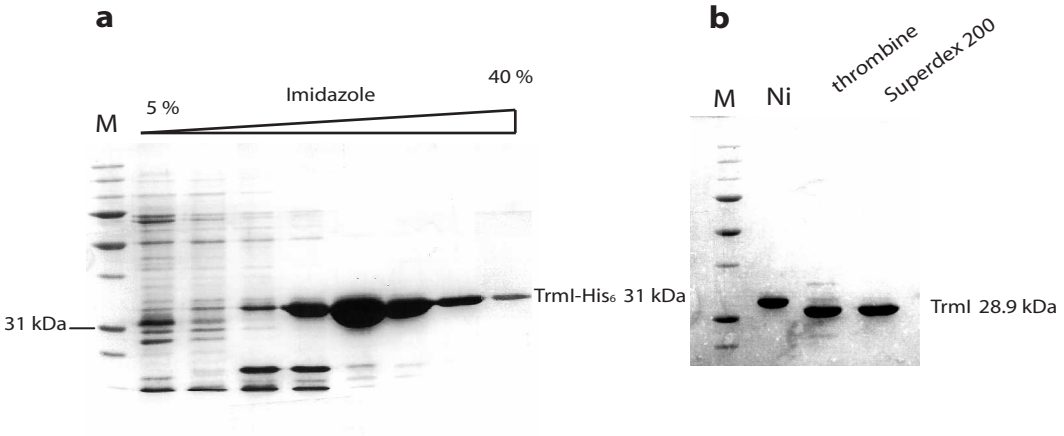


Figure B.2 – Etapes de purification de TrmI
a : Chromatographie IMAC, colonne nickel, élution de TrmI-His_6 par un gradient d'imidazole. **b** : clivage de l'étiquette N-terminale de 6 histidines (His_6) par la thrombine. Purification de TrmI après clivage par tamis moléculaire (Superdex 200).

La première étape de purification consiste en une colonne d'affinité pour les ions métalliques (IMAC), en l'occurrence ici une colonne d'affinité pour le nickel. La colonne Ni-NTA (Pharmacia Biotech, 5 mL) est équilibrée avec le tampon A (Tris-HCl pH 8,5 50 mM, KCl 500 mM). L'extrait brut est déposé sur la colonne Q–sépharose à raison de 1 mL/min. Après le passage de 50 mL de tampon A, un gradient de 100 mL est réalisé entre 0 et 100 % de tampon B (Tris-HCl pH 8,5 50 mM, KCl 500 mM, Imidazole 1 M) à raison de 1 mL/min. La protéine TrmI est éluée entre 250 et 350 mM d'imidazole (Figure B.2 **a**).

L'étiquette N-terminale est ensuite clivée par coupure enzymatique à la thrombine (25 U de thrombine par mg de TrmI). La coupure est réalisée à température ambiante sur la nuit (FIGURE B.2 **b**).

La protéine tronquée est ensuite purifiée par une étape de tamis moléculaire (Superdex 200, Pharmacia Biotech, 320 mL). La colonne est préalablement équilibrée avec le tampon de cristallisation (Tris-HCl pH 8,0 50 mM, KCl 100 mM). Les fractions contenant la protéine TrmI purifiée sont rassemblées et concentrées à l'aide d'Amicon® Ultra jusqu'à 8–10 mg/mL et conservées à 4 °C. 20 mg de TrmI purifiée sont généralement obtenus pour un litre de culture.

Surproduction et purification de l'ARNt$_f^{Met}$ de *E. coli*

Le protocole de surproduction et de purification de l'ARNt$_f^{Met}$ est adapté du protocole d'extraction d'ARNt de Zubay [468] et du protocole de surproduction et de purification de Meinnel *et al.* [469, 470]

Surproduction d'ARNt$_f^{Met}$

L'ARNt a été surproduit dans la souche JM101TR de *E. coli* transformée par le plasmide pBSTtRNA$_f^{Met}$ portant le gène de l'ARNt$_f^{Met}$ ainsi qu'un gène de résistance à l'ampicilline. Les cellules sont cultivées pendant une nuit à 37 °C en milieu LB en présence d'ampicilline (50 μg/mL), jusqu'en phase stationnaire. Les cellules sont récoltées par centrifugation à 9000 g pendant 30 min à 4 °C (rotor JLA 8100, Beckman), puis reprises dans 5 mL de tampon (Acétate de magnésium 10 mM, Tris-HCl pH 7,4 1 mM).

Purification de l'ARNt$_f^{Met}$

Préparation de l'ARNt total

La lyse cellulaire et l'extraction sont menées en une seule étape par l'ajout d'un volume de phénol saturé en eau. Une agitation douce est appliquée durant une heure. Après décantation par centrifugation à 9000 g pendant 30 min à 15 °C (rotor JA 12, Beckman), les acides nucléiques contenus dans la phase aqueuse sont précipités par l'ajout de 2 volumes d'éthanol absolu et de 0,1 volume de chlorure de sodium 5 M. La suspension d'acides nucléiques est centrifugée à 9000 g pendant 30 min à 4 °C et le culot est repris dans 5 mL de chlorure de sodium 1 M. Seuls les petits ARN sont solubles à cette force ionique. Ainsi, après une nouvelle centrifugation, le surnageant qui contient l'ARNt total est précipité par l'ajout de 2,5 volumes d'éthanol absolu et centrifugé à nouveau (9000 g, 30 min, 4 °C). Le culot d'ARNt total est alors repris dans 2 mL de Tris-HCl pH 8,0 2 M et incubé pendant deux heures à 37 °C. Cela permet la déacylation de l'extrémité 3′ des ARNt. Après une dernière précipitation à l'éthanol, les ARNt sont repris dans 5 mL d'eau.

Chromatographies

La première étape de purification consiste en un tamis moléculaire (Superdex 75, Pharmacia Biotech, 320 mL). La colonne est préalablement équilibrée avec le tampon d'élution (Tris-HCl pH 7,0 25 mM). Les fractions correspondant à un volume d'élution compris entre 130 et 180 mL sont rassemblées. La deuxième étape de la purification consiste en une colonne échangeuse d'anions (Pharmacia Biotech, Resource–Q, 50 mL). La colonne est préalablement équilibrée avec le tampon A (Tris-HCl pH 7,0 50 mM), puis l'échantillon est déposé à raison de 2 mL/min. Un gradient de 500 mL est réalisé entre 30 % et 60 % de tampon B (Tris-HCl pH 7,0 50 mM, NaCl 1 M) à raison de 2 mL/min. L'ARNt$_f^{Met}$ est élué entre 45 % et 50 % de tampon B. La troisième étape de la purification consiste en une autre colonne échangeuse d'anions plus résolutive (Pharmacia Biotech, Mono–Q, 8 mL). Cette étape est identique à la précédente, toute proportion gardée en ce qui concerne les volumes de colonnes. La dernière étape de purification consiste en une colonne en phase inverse (Pharmacia Biotech, Phenyl–Superose, 8 mL). La colonne est préalablement équilibrée avec le tampon C (acétate d'ammonium pH 6,5 10 mM, sulfate d'ammonium 1,7 M). puis l'échantillon est déposé à raison de 1 mL/min. Un gradient inverse est réalisé entre 0 et 100 % de tampon B (acétate d'ammonium pH 6,5 10 mM, sulfate d'ammonium 1,0 M). L'ARNt$_f^{Met}$ est élué à environ 1,5 M de sulfate d'ammonium. Les fractions contenant l'ARNt$_f^{Met}$ purifié sont rassemblées, dialysées contre le tampon de cristallisation (Tris-HCl pH 8,0 50 mM, KCl 100 mM), concentrées à l'aide d'Amicon® Ultra et conservées à -20 ℃.

Résumé

Le travail retranscrit dans cette thèse regroupe l'étude de différents processus biologiques impliquant les ARN de transfert. Premièrement, dans le cadre de l'étude du rôle de la protéine de nucléocapside (NC) dans la formation du complexe d'initiation de la transcription inverse du VIH-1, un site de fixation fort et spécifique de la NC a été identifié sur le bras D de l'ARNt$_3^{Lys}$ — l'ARNt servant d'amorce à la transcriptase inverse lors de la synthèse du brin d'ADNc(−). Cette fixation permet d'une part la fusion des interactions tertiaires entre les bras T et D de l'ARNt$_3^{Lys}$ et pourrait d'autre part être l'un des facteurs déplaçant l'équilibre vers la formation de l'hybride ARNt$_3^{Lys}$/ARN viral. L'étude structurale par RMN du complexe NC/bras D s'est heurtée à des problèmes d'échange chimique dans la gamme intermédiaire–rapide de l'échelle de temps spectrale. L'utilisation de séquences de type CPMG–HSQC a permis d'améliorer le rapport signal/bruit des expériences RMN, et particulièrement celui des signaux des résidus de l'interface. Ceci nous a permis d'identifier les résidus de la NC impliqués dans la reconnaissance du bras D. Deuxièmement, la structure de la m^{1}A$_{58}$ méthyltransférase de *T. thermophilus* (TrmI) — une enzyme de modification des ARNt — a été résolue par cristallographie et affinée à 1.7 Å. Ceci a permis d'identifier trois résidus potentiellement impliqués dans la catalyse et/ou la reconnaissance du substrat ARNt (D$_{170}$, Y$_{78}$ et Y$_{194}$). La production de variants de TrmI au niveau de ces résidus ainsi que des études d'amarrage moléculaire d'une adénine au site actif ont permis de confirmer cette hypothèse et de proposer un rôle catalytique pour chacun d'eux. Parallèlement, des études par gel natif et par spectrométrie de masse en conditions non dénaturantes ont montré une stœchiométrie 1/2 pour le complexe entre l'enzyme TrmI tétramérique et le substrat ARNt. Troisièmement, la résolution de la structure cristallographique de l'ARNt$_f^{Met}$ initiateur d'*E. coli* a révélé une conformation unique de la région de l'anticodon. Cette conformation unique est associée au paires GC de la tige anticodon, caractéristiques des ARNt initiateurs. Cette conformation particulière — dans laquelle la base A$_{37}$ ne vient pas s'empiler entre les bases 36 et 38, comme dans tous les ARNt élongateurs — permet de mettre en lumière de nombreux résultats biochimiques de la littérature et suggère un mécanisme par lequel la machinerie de l'initiation de la traduction pourrait discriminer l'ARNt initiateur de tous les ARNt cellulaires.

Mots clés

ARNt – Cristallographie – RMN – transcription inverse du VIH-1 – nucléocapside – ARNt$_3^{Lys}$ – méthyltransférase – TrmI – 1-méthyladénosine – m^{1}A$_{58}$ – initiation de la traduction – ARNt$_f^{Met}$ – anticodon

---

Abstract

During this thesis, we have studied several biological processes involving transfert RNA. First, the study of the role of the nucleocapsid protein (NC) in the formation of the HIV-1 reverse transcription initiation complex enabled us to identify a strong and specific binding site for the NC onto the D-arm of tRNA$_3^{Lys}$ — the tRNA used as a primer by the reverse transcriptase for the cDNA(−) synthesis. This strong binding allows the melting of tertiary interactions between the T and D-arms of tRNA$_3^{Lys}$ and could also be a key factor driving the tRNA$_3^{Lys}$/viral RNA hybrid formation. The NMR structural study of the NC/D-arm complex encountered chemical exchange problems in the range intermediate–fast of the chemical shift time-scale. The use of CPMG-like sequences improved signal/noise ratio, particularly for signals belonging to the residues involved in the interface. This allowed us to identify NC residues involved in the D-arm recognition. Secondly, the structure the m^{1}A$_{58}$ methyltransferase of *T. thermophilus* (TrmI) — a tRNA modification enzyme — was solved by cristallography and refined to 1.7 Å. Three residues potentially involved into catalysis and/or tRNA substrat recognition were identified (D$_{170}$, Y$_{78}$ et Y$_{194}$). The production of TrmI variants affected at these positions together with the molecular docking of an adenine into the TrmI active site enabled us to confirm this hypothesis and to propose a catalytic role for each mutated residue. In addition, non-denaturating gel electrophoresis and non covalent mass-spectrometry revealed the 1/2 stoichiometry of the TrmI tetramer/tRNA complex. Thirdly, the resolution of the *E. coli* initiator tRNA$_f^{Met}$ structure uncovered a unique conformation of the anticodon region. This unique conformation is associated with the GC pairs within the anticodon stem, which are characteristic of initiator tRNA. This peculiar structure — in which the A$_{37}$ base is not stacked between bases 36 and 38, as in all elongator tRNA structures — put in light many biochemical results of the literature and suggests a mechanism by which the translation initiation machinery could discriminate the initiator tRNA from all other tRNA.

Key words

tRNA – Crystallography – NMR – HIV-1 reverse transcription – nucleocapsid – tRNA$_3^{Lys}$ – methyltransferase – TrmI – 1-methyladenosine – m^{1}A$_{58}$ – translation initiation – tRNA$_f^{Met}$ – anticodon