Université de Montréal

Le motif d'empaquetage le long du sillon: une nouvelle entité structurale récurrente dans les ARN ribosomiques

par

Matthieu Gagnon

Département de biochimie

Faculté de médecine

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de Ph.D.

en biochimie

Décembre, 2008

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Le motif d'empaquetage le long du sillon: une nouvelle entité structurale récurrente dans les ARN ribosomiques

présentée par :
Matthieu Gagnon

a été évaluée par un jury composé des personnes suivantes :

Dr James G. Omichinski, président-rapporteur
Dr Serguei Chteinberg, directeur de recherche
Dr Joëlle Pelletier, membre du jury
Dr Roscoe Klinck, examinateur externe
Dr Gertraud Burger, représentante du doyen

Résumé

La plupart des molécules d'ARN doivent se replier en structure tertiaire complexe afin d'accomplir leurs fonctions biologiques. Cependant, les déterminants d'une chaîne de polynucléotides qui sont nécessaires à son repliement et à ses interactions avec d'autres éléments sont essentiellement inconnus. L'établissement des relations structure-fonction dans les grandes molécules d'ARN passe inévitablement par l'analyse de chaque élément de leur structure de façon individuelle et en contexte avec d'autres éléments. À l'image d'une construction d'immeuble, une structure d'ARN est composée d'unités répétitives assemblées de façon spécifique. Les motifs récurrents d'ARN sont des arrangements de nucléotides retrouvés à différents endroits d'une structure tertiaire et possèdent des conformations identiques ou très similaires. Ainsi, une des étapes nécessaires à la compréhension de la structure et de la fonction des molécules d'ARN consiste à identifier de façon systématique les motifs récurrents et d'en effectuer une analyse comparative afin d'établir la séquence consensus.

L'analyse de tous les cas d'empaquetage de doubles hélices dans la structure du ribosome a permis l'identification d'un nouvel arrangement nommé motif d'empaquetage le long du sillon (AGPM) (*along-groove packing motif*). Ce motif est retrouvé à 14 endroits dans la structure du ribosome de même qu'entre l'ARN ribosomique 23S et les molécules d'ARN de transfert liées aux sites ribosomaux P et E. Le motif se forme par l'empaquetage de deux doubles hélices *via* leur sillon mineur. Le squelette sucre-phosphate d'une hélice voyage le long du sillon mineur de l'autre hélice et *vice versa*. Dans chacune des hélices, la région de contact comprend quatre paires de bases. L'empaquetage le plus serré est retrouvé au centre de l'arrangement où l'on retrouve souvent une paire de bases GU dans une hélice interagissant avec une paire de bases Watson-Crick (WC) dans l'autre hélice. Même si la présence des paires de bases centrales GU versus WC au centre du motif augmente sa stabilité, d'autres alternatives existent pour différents représentants du motif. L'analyse comparative de trois librairies combinatoires de gènes d'AGPM, où les paires de bases centrales ont été variées de manière complètement aléatoire, a montré que le contexte structural influence l'étendue de la variabilité des séquences de nucléotides formant les paires de bases centrales.

Le fait que l'identité des paires de bases centrales puisse varier suggérait la présence d'autres déterminants responsables au maintien de l'intégrité du motif. L'analyse de tous les contacts entre les hélices a révélé qu'en dehors du centre du motif, les interactions entre les squelettes sucre-phosphate s'effectuent *via* trois contacts ribose-ribose. Pour chacun de ces contacts, les riboses des nucléotides qui interagissent ensemble doivent adopter des positions particulières afin d'éviter qu'ils entrent en collision. Nous montrons que la position de ces riboses est modulée par des conformations spécifiques des paires de bases auxquelles ils appartiennent.

Finalement, un autre motif récurrent identifié à l'intérieur même de la structure de trois cas d'AGPM a été nommé « *adenosine-wedge* ». Son analyse a révélé que ce dernier est lui-même composé d'un autre arrangement, nommé motif triangle-NAG (*NAG-triangle*). Nous montrons que le motif « *adenosine-wedge* » représente un arrangement complexe d'ARN composé de quatre éléments répétitifs, c'est-à-dire des motifs AGPM, « *hook-turn* », « *A-minor* » et triangle-NAG. Ceci illustre clairement l'arrangement hiérarchique des structures d'ARN qui peut aussi être observé pour d'autres motifs d'ARN.

D'un point de vue plus global, mes résultats enrichissent notre compréhension générale du rôle des différents types d'interactions tertiaires dans la formation des molécules d'ARN complexes.

Mots-clés: Motif récurrent, structure d'ARN, ribosome, sélection *in vivo*, ARN ribosomique

Abstract

Most RNA molecules have to adopt a complex tertiary structure to accomplish their biological functions. However, the important determinants of a polynucleotide chain that are required for its proper folding and its interactions with other elements are essentially unknown. The establishment of structure-function relationships in large RNA molecules goes inevitably through the analysis of each element of their structure separately and in context with other elements. Like a building, an RNA structure is built of repetitive pieces that are glued together in a specific way. These repetitive elements, instead of being bricks, are recurrent motifs. Recurrent RNA motifs are arrangements of nucleotides found in different parts of a tertiary structure and have identical or very similar conformations. Thus, a necessary step toward the understanding of RNA structure and function consists in the systematic identification of recurrent motifs, followed by their comparative analysis and establishment of their sequence consensus.

The analysis of all instances of helical packing within the ribosome structure led to the identification of a new structural arrangement, named the along-groove packing motif (AGPM), which is found in 14 places of the ribosome structure as well as between the 23S ribosomal RNA and the transfer RNA molecules bound to the P and E sites. The motif is formed by the packing of two double helices *via* their minor grooves. The sugar-phosphate backbone of one helix goes along the minor groove of the other helix and *vice versa*. In each helix, the contact region includes four base pairs. The closest packing occurs in the center where one can often see a GU base pair packed against a WC base pair. While the presence of the central base pairs GU versus WC in the core of the motif enhances its stability, other alternatives are also present among available structures of the motif. A comparative analysis of three different combinatorial gene libraries of AGPM, in which the central base pairs were fully randomized, shows that the structural context influences the scope of nucleotide sequence variability of the central base pairs.

The fact that the identity of the central base pairs can vary suggested that there are other determinants responsible of the motif's integrity. Analysis of all other inter-helix contacts has shown that outside the center of the motif the interactions between backbones are made *via* three ribose-ribose contacts. Within each of these contacts, the

riboses of the nucleotides that are in touch adopt particular positions in order to provide for collision-free interactions between them. We show that the position of these riboses is modulated by the specific base pair conformation in which it belongs.

Finally, another recurrent arrangement that occurs within the structure of three cases of AGPM was identified and called the adenosine-wedge. Analysis has shown that the latter motif is itself composed of a smaller arrangement, called the NAG-triangle motif. We show that the adenosine-wedge motif represents a complex RNA arrangement composed of four repetitive elements, AGPM, the hook-turn, the A-minor and the NAG-triangle, which clearly illustrates the hierarchical organisation of the structure that could also occur in other RNA motifs as well.

Altogether, my results enrich our general understanding of the role of different types of tertiary interactions in the formation of large RNA molecules.

Keywords: Recurrent motif, RNA structure, ribosome, *in vivo* selection, ribosomal RNA

Table des matières

Résumé iii
Abstract v
Table des matières vii
Liste des tableaux xi
Liste des figures xiv
Liste des abréviations xvii
Remerciements xx
1. Introduction 2
1.1 De la séquence, à la structure, à la fonction 2
1.2 L'importance de l'ARN: Considérations biologiques 3
1.3 Types d'ARN 4
1.4 Les éléments structuraux des ARN 6
1.4.1 Structure primaire 6
1.4.2 Structure secondaire 7
1.4.2.1 Représentation de la structure secondaire 8
1.4.2.2 Forme A versus forme B des doubles hélices 10
1.4.3 Structure tertiaire 15
1.5 Le ribosome 16
1.5.1 L'organisation moléculaire du ribosome 17
1.5.2 Survol de la synthèse protéique 17
1.5.2.1 L'initiation 17
1.5.2.2 L'élongation 18
1.5.2.3 Terminaison 21
1.6 Structure tertiaire du ribosome 21
1.6.1 Structures à haute résolution du ribosome 22
1.6.2 Analyse systématique des structures d'ARN 25
1.7 Les motifs tertiaires des structures d'ARN 26
1.7.1 Le motif "A-minor" 26
1.7.2 Le motif "ribose zipper" 28
1.7.3 Le motif de la boucle T 30
1.7.4 Le motif "kink-turn" 30
1.7.5 Le motif "hook-turn" 31
1.8 Objectif 33
1.9 La révélation qui a conduit à la découverte du motif AGPM 33
2. GU receptors of double helices mediate tRNA movement in the ribosome 36
2.1 Abstract 37
2.2 Introduction 37
2.3 Along-groove packing motif 38
2.4 Role of the GU base pair 38
2.5 Involvement in the tRNA association with the ribosome 39
2.6 Possible role in translocation 40
2.7 Concluding remarks 42
2.8 Acknowledgments 42
2.9 References 42

2.10 Figures 46
3. Close packing of helices 3 and 12 of 16S rRNA is required for the normal ribosome function 52
3.1 Abstract 53
3.2 Introduction 53
3.3 Experimental procedures 55
3.4 Results 58
3.4.1 Experimental System and Library Design 58
3.4.2 Cloning and Selection of Functional Clones 58
3.4.3 Characterization of the Selected Clones 59
3.4.4 Potentially Negative Role of Uridine 1192 60
3.5 Analysis of the selected clones 60
3.5.1 Close Packing in Group I Clones 61
3.5.2 Preferences for the WC Base Pair 63
3.5.3 A Suboptimal Close Packing Arrangement for Clone A2 64
3.6 Discussion 66
3.7 Acknowledgments 68
3.8 References 68
3.9 Tables 71
3.10 Figures 73
3.11 Supplemental methods 78
3.12 Supplemental tables 81
3.13 Supplemental references 85
4. Recurrent RNA motifs as probes for studying RNA-protein interactions in the ribosome 87
4.1 Abstract 88
4.2 Introduction 88
4.3 Results 89
4.3.1 Background: general description of the AGPM 89
4.3.2 The motifs studied 91
4.3.3 Cloning and selection of the functional clones 92
4.3.4 Analysis of the selected clones: the minimal requirement for the integrity of the AGPM 93
4.3.5 Molecular dynamics simulations 95
4.3.6 The symmetry of the central base pairs in motif S296 97
4.3.7 Interaction of motif L639 with ribosomal protein L35 97
4.3.8 Interaction of motif L657 with ribosomal protein L4 98
4.3.9 Instant evolution versus natural evolution 100
4.4 Discussion 101
4.4.1 The power of the approach 101
4.4.2 New findings about the AGPM: principles of RNA structure formation 102
4.4.3 New findings about the AGPM: principles of RNA-protein interaction 105
4.4.4 The sensitivity of the approach 106
4.5 Materials and Methods 107
4.6 Acknowledgments 110
4.7 References 110
4.8 Tables 113

4.9 Figures 114
4.10 Supplemental Data 121
4.10.1 Instant evolution versus natural evolution 121
4.11 Supplemental Methods 122
4.11.1 Combinatorial gene libraries: primers and cloning 122
4.12 Supplemental Tables 124
4.13 Supplemental Figure 127
4.14 Supplemental References 128
5. Structural rules for the formation of backbone-backbone interactions between closely packed RNA double helices 130
5.1 Abstract 131
5.2 Introduction 132
5.3 Results 133
5.3.1 Background: the general description of the AGPM 133
5.3.2 Nomenclature of different elements of the AGPM 134
5.3.3 Collection of the set of the AGPM 135
5.3.4 Principles of helix packing within the AGPM 136
5.3.5 The inter-helix interactions in contact zones QR and PS 137
5.3.5.1 The central role of the -1-base pairs 137
5.3.5.2 The position of WC -2-base pairs 138
5.3.5.3 The position of non-WC -2-base pairs 139
5.3.5.4 Statistical analysis of the identities of the -2-base pairs 139
5.3.6 The nature of the asymmetry between the two helices within the AGPM 140
5.3.6.1 Potential collisions of the helices at the 0- and +1-levels 140
5.3.6.2 The adjustment of base pair [0P;0Q] to the potential collision with nucleotide 0S 140
5.3.6.3 The GU-WC pattern for the central base pairs 141
5.3.6.4 The asymmetry between the -1-base pairs as a consequence of the asymmetry between the 0-base pairs 142
5.3.7 Ribose-ribose interaction in contact zone QS 143
5.3.7.1 Introduction of a non-WC base pair 143
5.3.7.2 The WC-WC pattern at the +1-level 144
5.3.8 The consensus secondary structure of the AGPM 145
5.3.9 tRNA fixation in the P and E sites of the 50S ribosomal subunit 145
5.4 Discussion 147
5.5 Methods 150
5.6 Acknowledgments 150
5.7 References 150
5.8 Tables 154
5.9 Figures 159
5.10 Supplemental tables 166
5.11 Supplemental references 184
6. The adenosine wedge: A new structural motif in ribosomal RNA 186
6.1 Abstract 187
6.2 Description of the adenosine-wedge motif 187
6.3 Asymmetry of the adenosine-wedge motif 189
6.4 Displacement of the adenosine stack 189

6.5 Inclination of the adenosine stack.....191
6.6 Role of nucleotide -3P 192
6.7 Variations in the structure of the adenosine-wedge motif.....192
6.8 Conservation of the elements forming the adenosine-wedge motif 193
6.9 NAG-triangle at the core of the adenosine-wedge motif.....195
6.10 Concluding remarks 196
6.11 Methods.....197
6.12 Acknowledgments 197
6.13 References 197
6.14 Figures.....199
6.15 Supplemental tables 206
6.16 Supplemental figures 210
7. Discussion.....214
7.1 Découverte du motif d'empaquetage le long du sillon (AGPM) (*along-groove packing motif*).....214
7.2 Expression de librairies combinatoires de gènes du motif AGPM.....216
7.3 Règles structurales liées à la formation des interactions squelette-squelette dans le motif AGPM 218
7.4 Organisation hiérarchique des structures tertiaires d'ARN.....220
8. Conclusion.....223
9. Bibliographie (Introduction et Discussion).....224

Liste des tableaux

Chapitre 1: Introduction

Tableau I. Liste des structures à haute résolution du ribosome et de ses sous-unités……...…..25

Chapitre 3: Close packing of helices 3 and 12 of 16S rRNA is required for the normal ribosome function

Table I. Nucleotide sequences and characteristics of the selected clones……71
Table II. Statistical spectrum of the identities of the central base pairs in AGPM SU296……...72

Supplemental Table I. Efficiency of different miscoding events for all selected clones……81
Supplemental Table II. Statistical spectrum of the central base pairs fitting the GU-WC pattern in the known AGPMs……82
Supplemental Table III. The constructs used for measuring the efficiency of stop codons readthrough and frameshifting events……..83
Supplemental Table IV. The constructs used for measuring the efficiency of mis-incorporation events…….83
Supplemental Table V. GFPuv3 activity of the six clones with the reverse mutation U1192C in the 16S rRNA…….84

Chapitre 4: Recurrent RNA motifs as probes for studying RNA-protein interactions in the ribosome

Table I. Nucleotide sequences of the selected clones…….113

Supplemental Table I. Presence of different tetra-nucleotide combinations as the central base pairs of AGPMs S296, L639 and L657 in the prokaryotic 16S and 23S rRNAs……..124

Supplemental Table II. Presence of different tetra-nucleotide combinations as the central base pairs of all AGPMs in the prokaryotic 16S and 23S rRNAs ...…..125
Supplemental Table III. Sequences of the oligonucleotides used in this study..126

Chapitre 5: Structural rules for the formation of backbone-backbone interactions between closely packed RNA double helices

Table I. Structure of the contact zones……...154
Table II. Occurrence of different combinations of the -1-base pairs in AGPMs existing in prokaryotic rRNA…….155
Table III. Occurrence of different identities of the -2-base pairs in AGPMs existing in prokaryotic rRNA…….156
Table IV. Occurrence of different combinations of the 0-base pairs in AGPMs existing in prokaryotic rRNA…….157
Table V. Occurrence of different combinations of the +1-base pairs in AGPMs existing in prokaryotic rRNA…….158

Supplemental Table I. Statistical data of the identities of the -1 base pairs for all known ribosomal AGPMs…….....166
Supplemental Table II. Statistical data of the identities of the -2 base pairs for all known ribosomal AGPMs…….....171
Supplemental Table III. Statistical data of the identities of the 0-base pairs for all known AGPMs……...177
Supplemental Table IV. Statistical data of the identities of the +1 base pairs for all known AGPMs……..181

Chapitre 6: The adenosine wedge: A new structural motif in ribosomal RNA

Supplemental Table I. List of hydrogen bonds formed in the adenosine-wedge motif……...206

Supplemental Table II. Statistical data of the identities of nucleotides -2P, -2R, -2Q and -2S in AGPMs S549, L639 and L657……..207
Supplemental Table III. List of all cases of the NAG-triangle motif found in the *E. coli* ribosome structure ……..208

Liste des figures

Chapitre 1: Introduction

Figure 1. Différents types de motifs de structures secondaires ……8
Figure 2. Structures secondaire et tertiaire de l'ARNt…….10
Figure 3A. Structure secondaire de l'ARNr 16S de *E. coli*……..12
Figure 3B. Structure secondaire de l'ARNr 23S de *E. coli*……..13
Figure 4. Structures des doubles hélices d'ARN et d'ADN..……...14
Figure 5. Le cycle d'élongation……19
Figure 6. Structures des sous-unités ribosomiques individuelles vues de l'interface et du ribosome 70S...……...23
Figure 7. Différents types d'interactions du motif « *A-minor* »……...27
Figure 8. Le motif « *ribose zipper* »…….29
Figure 9. Stéréogramme de la structure du motif de la boucle T tirée de l'ARNt^{Phe} de la levure …….29
Figure 10. Stéréogramme du motif « *kink-turn* » ……32
Figure 11. Stéréogramme du motif « *hook-turn* »……32

Chapitre 2: GU receptors of double helices mediate tRNA movement in the ribosome

Figure 1. Along-groove packing of double helices……..46
Figure 2. Nucleotide sequences of the along-groove packing motifs……..47
Figure 3. Juxtaposition of the central base pairs……..48
Figure 4. Superimposition of the rRNA-tRNA intermolecular along-groove packing motifs……..49
Figure 5. Positions of the binding sites in the acceptor and D stems of tRNAs in different pure and hybrid states with respect to the D and AC receptors in helices 69 and 68 of 23S rRNA.....50

Chapitre 3: Close packing of helices 3 and 12 of 16S rRNA is required for the normal ribosome function

Figure 1. Schematic representation of the along-groove packing motif (AGPM).....73
Figure 2. Juxtaposition of different central base pairs within AGPM……..74
Figure 3. AGPM SU296 in *E. coli* 16S rRNA……75
Figure 4. The juxtaposition of the central base pairs in clone A1 superposed with the same juxtaposition rotated for 180° around the center of pseudo-symmetry……..76
Figure 5. A model of close packing of two RY base pairs with a deformed geometry of hydrogen bonds……77

Chapitre 4: Recurrent RNA motifs as probes for studying RNA-protein interactions in the ribosome

Figure 1. Schematic representation of the AGPM……..114
Figure 2. Different arrangements of the central base pairs in the AGPM …….115
Figure 3. Nucleotide sequences of the three cases of AGPM considered in this study ……..116
Figure 4. Stereo view of the structural contexts of motifs L657 and L639 taken from the *E. coli* ribosome……...117
Figure 5. Juxtapositions of the bases in the AC and GA base pairs ……..118
Figure 6. Molecular dynamics simulations of the AGPM structure containing different nucleotide triples ……119
Figure 7. Conformational rearrangements associated with the GU⇔WC exchange of the central base pairs in the AGPM ……..120

Supplemental Figure 1. Modeled complex of the AGPM used in molecular dynamics simulations ……127

Chapitre 5: Structural rules for the formation of backbone-backbone interactions between closely packed RNA double helices

Figure 1. The along-groove packing motif…….159
Figure 2. Nucleotide sequences of all known AGPMs identified within ribosomal RNA, between two hammerhead ribozymes, and between 23S rRNA and a tRNA bound at the P-site and at the E-site……...160
Figure 3. The arrangement of the three contact zones within the AGPM……..161
Figure 4. Ribose-ribose interactions within contact zones PS and QR……..162
Figure 5. The potential collision of the internal nucleotides at the 0- and +1-levels and its consequences for the structure and position of base pair [0P;0Q]……...163
Figure 6. Stereo view of the asymmetry between base pairs [-1P;-1Q] and [-1R; -1S] caused by the displacement of nucleotide 0Q……164
Figure 7. Stereo view of the deformation of the double helical geometry at level +1 when both +1-base pairs are WC……..165
Figure 8. The consensus secondary structure for most known AGPMs ……...165

Chapitre 6: The adenosine wedge: A new structural motif in ribosomal RNA

Figure 1. The along-groove packing motif (AGPM)……..199
Figure 2. Simplified stereo view representation of the structure of the adenosine-wedge motif ……...201
Figure 3. Stereo view of the structure of the adenosine-wedge motif……202
Figure 4. Variations from the structure of the adenosine-wedge motif……..204

Supplemental Figure 1. Stereo view of the along-groove packing motif (AGPM)…….....210
Supplemental Figure 2. Stereo view of the structure of the adenosine-wedge motif within the context of the AGPM L639 from the *E. coli* ribosome ……..211
Supplemental Figure 3. The NAG-triangle motif……..212

Liste des abréviations

A	Adénine
Å	Angström (1×10^{-10} m)
ADN	Acide Désoxyribonucléique
AGPM	*Along-Groove Packing Motif*
Amp	Ampicilline
ARN	Acide ribonucléique
ARNi	Interférence d'ARN
ARNr	ARN ribosomique
ARNt	ARN de transfert
ASD	Séquence anti-Shine-Dalgarno
ATP	Adenosine triphosphate
C	Cytidine
CAT	Chloramphénicol acétyl-transférase
Cm	Chloramphénicol
Cryo-EM	*Cryo-Electron Microscopy*
EF-G	Facteur d'élongation G (*Elongation Factor-G*)
EF-Tu	Facteur d'élongation Tu (*Elongation Factor-Tu*)
EM	*Electron Microscopy*
fMet	formyl-Méthionine
G	Guanosine
GDP	Guanosine diphosphate
GFP	*Green Fluorescent Protein*
GTP	Guanosine triphosphate
h	Hélice de l'ARNr 16S
H	Hélice de l'ARNr 23S
HDV	*Hepatitis Delta Virus*
IF	Facteur d'initiation (*Initiation Factor*)
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
Kan	Kanamycine

L	Utilisé en préfixe: protéine ribosomique de la grande sous-unité du ribosome
LB	Luria-Bertani
mM	millimolaire
MD	*Molecular Dynamics*
MDa	Méga-Dalton
MIC	*Minimal Inhibitory Concentration*
miARN	Micro-ARN
PCR	*Polymerase Chain Reaction*
PDB	*Protein Data Bank*
PTC	*Peptidyl-Transferase Center*
ps	picoseconde
R	Purine
RF	Facteur de terminaison (*Release Factor*)
RNase P	ribonucléase P
RRF	Facteur de recyclage du ribosome (*Ribosome Recycling Factor*)
S	Utilisé en préfixe: protéine ribosomique de la petite sous-unité du ribosome Utilisé en suffixe: unité du coefficient de sédimentation, le Svedgerg
SD	Séquence Shine-Dalgarno
snRNA	Petit ARN nucléaire
snoRNA	Petit ARN nucléolaire
Spc	Spectinomycine
T	Thymine
Tet	Tétracycline
U	Uracile
µM	micromolaire
UTR	*Untranslated region*
Y	Pyrimidine
WC	Watson-Crick
WT	*Wild-Type*

À papa, maman, mon frère et ma défunte grand-maman
pour leurs encouragements constants
au cours de mes études

Remerciements

Je tiens tout d'abord à remercier mon directeur de recherche, Dr Serguei Chteinberg, de m'avoir accepté dans son laboratoire au mois de juin 2000. Tout au cours de ces années, il m'a incité à développer une vision structurale pour tous les problèmes qu'un scientifique peut se poser et à y adopter une approche d'analyse systématique. Il possède un esprit analytique et critique impressionnant. Sa compréhension de la structure des macromolécules est absolument remarquable. Sans aucun doute, il a été et restera toujours un exemple à suivre des plus inspirant. Je veux également remercier Dr Léa Brakier-Gingras de m'avoir incité à poursuivre mes études supérieures en biochimie et d'avoir toujours été une source de conseils judicieux.

Je remercie également tous les membres de mon comité de thèse d'avoir accepté d'évaluer mon travail.

Je remercie spécialement Dr Alka Mukhopadhyay pour son aide très précieuse dans la construction du système de ribosome spécialisé et la production ainsi que la caractérisation fonctionnelle des mutants de l'ARNr 16S. Merci également à Yury Boutourine qui est une source de stratégies expérimentales ingénieuses et qui prend aussi le temps d'avoir du plaisir tout en travaillant rigoureusement. Je tiens à remercier Jianhong Chen pour son aide précieuse avec le système UNIX et pour avoir élaboré les outils informatiques qui ont servi à nos analyses statistiques des séquences ribosomiques de même qu'à la recherche de nouveaux motifs dans les structures contenant de l'ARN. Je remercie tous les autres membres du laboratoire, Tetsu Ishii, François Boulay, Konstantine Bokov et Natalia Kotlova pour leur aide et leur soutien.

Finallement, je remercie tous les membres de ma famille, mon père Jean-Pierre, ma mère Céline et mon frère Jonathan pour leurs encouragements continus tout au cours de mes études.

Chapitre 1

Introduction

1. Introduction

1.1 De la séquence, à la structure, à la fonction

Depuis le début de ce siècle, une quantité énorme de structures d'ARN à haute résolution ont été déterminées. D'une certaine manière, ces progrès ont donné l'impression que nous commencions à comprendre ce qu'était vraiment la structure de l'ARN. Cependant, la disponibilité *per se* de plusieurs structures d'ARN complexes n'est en aucun point informatif sur la façon utilisée par différents éléments de structure pour s'assembler et se maintenir ensemble.

Puisque la fonction d'une macromolécule est intimement reliée à sa structure, l'établissement des relations structure-fonction dans des molécules d'ARN complexes passe inévitablement par l'analyse de chaque élément de leur structure individuellement et en contexte avec d'autres éléments. Une possibilité est d'identifier, de manière systématique, les « blocs » de construction récurrents (ci-après nommés motifs), de les comparer et d'établir leurs éléments structuraux communs essentiels à leur formation. À ce jour, nous en sommes toujours à l'étape d'identification et de collection des motifs individuels qui composent les molécules d'ARN et commençons seulement à comprendre comment ils permettent à différentes régions d'interagir ensemble au cours du processus de repliement de l'ARN en structures tridimensionnelles. Le jour où tous les motifs et leurs variations auront été identifiés et classifiés, il sera possible d'intégrer ces connaissances aux algorithmes de repliement de l'ARN.

Même si le repliement d'une molécule d'ARN est souvent assisté par des protéines, il est accepté que la séquence primaire contienne toute l'information nécessaire à la formation des structures secondaires et tertiaires. En effet, au moment où les structures secondaires sont formées (les doubles hélices dans l'ARN) en présence d'ions divalents [revu par (Draper, 2004; Draper et al., 2005; Woodson, 2005)], elles interagissent ensemble et forment une structure tertiaire compacte. Le but primaire de ma thèse consistait à élucider les règles qui régissent l'empaquetage des doubles hélices d'ARN d'une structure secondaire en une structure tridimensionnelle compacte, incluant la détermination des contraintes imposées sur la séquence des nucléotides.

1.2 L'importance de l'ARN: Considérations biologiques

Les molécules d'ARN sont présentes dans toutes les cellules vivantes et possèdent des fonctions variées centrales au maintien de la vie. L'ARN joue un rôle important dans une multitude de processus cellulaires, spécialement dans la synthèse des protéines.

Les premières étapes dans la compréhension de l'importance de l'ARN sont survenues dans les années 1950, lorsque Francis Crick formulait l'hypothèse des adaptateurs (*adaptor hypothesis*) stipulant que la traduction s'effectue par des molécules adaptatrices d'ARN de transfert (ARNt) (Crick, 1966). Il était déjà connu à ce moment que les protéines sont assemblées séquentiellement à partir d'acides aminés par les ribosomes. En 1956, Elliot « Ken » Volkin et Lazarus Astrachan ont découvert l'ARN messager au Laboratoire National de Oak Ridge, TN, USA. Cependant, à ce moment, ils référaient à cette molécule en tant que « *DNA-like RNA* »; plus tard, officiellement nommé ARN messager (ARNm) par François Jacob, Matthew Meselson and Sydney Brenner, une molécule d'ARN instable qui sert à transporter l'information à partir des gènes jusqu'au ribosome [revu par (Judson, 2001)]. Quelques années plus tard, James Watson établit que ces molécules d'ARN (ARNt et ARNm) étaient directement impliquées dans le processus de la synthèse protéique. Finalement, en 1965, Francis Crick a déchiffré le code génétique avec l'ARN en tant que messager de l'information génétique. Jusqu'à ce moment, toute autre classe d'ARN était soit considérée en tant qu'ARNm ou tout simplement peu importante, communément appelée « *junk RNA* ».

Quelques années plus tard, Carl Woese était à l'origine de l'idée d'un monde où la vie dépendait totalement de l'ARN (Woese, 1967). Par la suite, au début des années 1980, les découvertes des ARN auto-épisseurs (T. Cech) et de la ribonucléase P (RNase P) (S. Altman), ou plus généralement des molécules d'ARN possédant une activité catalytique, ont stimulé l'idée qu'à un certain moment, l'évolution était dominée par l'ARN (Kruger et al., 1982; Guerrier-Takada and Altman, 1984). Ces molécules catalytiques d'ARN ont été par la suite nommées ribozymes. Cependant, Walter Gilbert a été le premier à prononcer les mots « *The RNA World* » (Gilbert, 1986). Ce modèle se basait principalement sur le fait que certaines molécules d'ARN possèdent une activité

catalytique. Puisque l'ARN sert également au transport de l'information, il semblait raisonnable de suggérer que d'anciennes molécules d'ARN puissent avoir agi comme point de départ à l'origine de la vie. Cette découverte a été à l'origine de la notion du nouveau « Monde à ARN » ainsi qu'à la base du changement de croyance concernant l'évolution de l'ADN, l'ARN et des protéines. À ce point, il était évident que l'ARN n'agissait pas seulement en tant que capacité d'emmagasinage de l'information (comme l'ADN), mais possédait aussi des fonctions catalytiques (comme les protéines). En 1989, Sidney Altman et Thomas Cech se sont partagé le Prix Nobel de Chimie pour leurs travaux sur les propriétés catalytiques de l'ARN.

Plus tard, d'autres molécules d'ARN possédant des fonctions catalytiques ont été identifiées et dans les années 1990, le rôle actif de l'ARN dans le centre catalytique du ribosome a été confirmé. Nos connaissances de la biologie de l'ARN continuent leur expension, telle que révélée par l'attribution du Prix Nobel de Médecine en 2006 à Andrew Fire et Craig Mello pour leur découverte fondamentale du mécanisme de l'interférence de l'ARN (ARNi), un processus qui permet le contrôle de l'expression génétique.

1.3 Types d'ARN

Nombreuses sont les molécules d'ARN impliquées dans différents processus cellulaires. Tel que mentionné ci-dessus, le médiateur entre l'ADN et les protéines est l'ARNm qui contient l'information nécessaire à la spécification de l'ordre des acides aminés qui composent les protéines. L'ARNm agit comme matrice à l'assemblage des molécules de polypeptides. Cependant, plusieurs molécules d'ARN ne codent pas pour une protéine. Les exemples les plus importants parmi les ARN non-codants sont les ARN ribosomiques (ARNr) et les ARN de transferts (ARNt), toutes deux impliquées dans le processus de la traduction. Suivant l'enlèvement des introns par un mécanisme d'auto-épissage (Cannone et al., 2002), les ARNr participent à la lecture du message porté par l'ARNm et catalysent la formation des liens peptidiques entre les acides aminés lors de la production d'une protéine (Noller et al., 1992). En revanche, les molécules d'ARNt agissent comme adaptateurs entre la matrice d'ARNm et la chaîne

polypeptidique en extension. Parmi les autres molécules d'ARN non-codantes, la RNase P est responsable de la maturation de l'extrémité 5' des molécules d'ARNt.

Les ARNm des procaryotes possèdent des domaines hautement structurés connus comme riborégulateurs (*riboswitches*) localisés dans la région 5' non-traduite (5'UTR) de certains ARNm. Ces éléments régulateurs en *cis* lient sélectivement des métabolites et modulent l'expression de certains gènes en réponse aux changements de concentration d'un ligand [revu par (Vitreschak et al., 2004; Batey, 2006)].

Chez les eucaryotes, plusieurs autres processus biologiques dépendent de différents types d'ARN. En effet, les micro-ARNs (miARNs) peuvent réduire l'expression génétique de certains ARNm en s'appariant à une partie du gène. Ces miARNs agissent *via* un mécanisme appelé interférence de l'ARN (ARNi). Dans certains cas, les miARNs peuvent empêcher la traduction de l'ARNm et dans d'autres situations, l'ARNm est simplement dégradé. Un autre processus nécessitant l'intervention de molécules d'ARN est celui de la modification d'autres ARNs. Les complexes appelés splicéosomes, lesquels sont composés de plusieurs petits ARN nucléaires (snRNAs), ont comme fonction de catalyser l'excision des introns des précurseurs d'ARNm et la suture des exons au cours de l'étape appelée processus d'épissage. Les nucléotides de certains ARNs peuvent également être modifiés. Ce processus est effectué par les petits ARN nucléolaires (snoRNAs). Les snoRNAs s'associent avec des enzymes qui les guident à l'endroit approprié sur une molécule d'ARN *via* des appariements de bases. Ces enzymes effectuent ensuite la modification de certains nucléotides.

Une autre molécule d'ARN d'intérêt est le ribozyme à la tête en forme de marteau (*hammerhead ribozyme*). Il s'agit d'un ARN auto-épisseur impliqué dans la réplication de l'ARN circulaire de certains viroïdes qui infectent les plantes. Un aspect commun entre toutes ces molécules d'ARN est qu'elles doivent adopter un repliement tridimensionnel défini afin d'accomplir leur fonction respective.

1.4 Les éléments structuraux des ARN

Dans cette section, je fais une révision des composantes structurales de l'ARN qui se combinent pour former des arrangements tridimensionnels complexes. Par analogie aux protéines, les ARN biologiques peuvent être décomposés en structure primaire, secondaire et tertiaire.

1.4.1 Structure primaire

Le nucléotide est l'unité fondamentale des acides nucléiques. Il est composé de trois parties : un sucre, un phosphate et une base organique. La base peut soit être une purine soit une pyrimidine. Dans le cas de l'ARN, le sucre est un ribose, les purines sont adénine (A) et guanine (G), et les pyrimidines sont cytosine (C) et uracile (U). La structure primaire d'une chaîne de polynucléotides consiste en un enchaînement ordonné des nucléotides des extrémités 5' vers 3'. La séquence des nucléotides d'une molécule d'ARN reflète directement la séquence d'ADN du gène qui l'encode. L'ARNm est un bon exemple d'une molécule d'ARN pour laquelle sa fonction principale, i.e. sa traduction par le ribosome, s'accomplit seulement sous sa forme simple brin. Le ribosome lit le message génétique porté par l'ARNm et le traduit en protéine qui exprimera éventuellement le phénotype. Même si l'ARNm représente seulement une succession de nucléotides attachés ensemble, à quelques endroits de la séquence, la chaîne de polynucléotides peut se replier en structures secondaires définies et peut même créer localement des arrangements structuraux tertiaires.

À plus grande échelle, les ARNt et ARNr doivent former des structures secondaires et tertiaires afin d'accomplir leur fonction biologique. Dans chacun de ces cas, la séquence des nucléotides contient toute l'information requise à la formation locale et globale de la structure tridimensionnelle.

1.4.2 Structure secondaire

Le niveau supérieur d'organisation de la structure de l'ARN consiste aux paires de bases, ou structure secondaire, qui permet de distinguer les régions formant des paires de bases canoniques (doubles hélices) de celles formant des régions non-appariées (boucles). Les motifs de structure secondaire incluent, entre autres, les régions simples brins, les doubles hélices, les « *bulges* » ou les boucles en « *bulge* », les mésappariements, les boucles en forme d'épingle à cheveux (*hairpin loops*), les tétraboucles, les boucles internes et les jonctions (Figure 1). Ils peuvent également être appelés blocs de structure élémentaire. Même si l'ARN peut exister sous forme simple brin, les groupements hautement hydrophiles des bases azotées ne peuvent pas être complètement solvatés par les molécules d'eau environnantes. Pour cette raison, à l'image de la situation avec l'ADN, le seul moyen de solvater complètement les groupements hydrophiles est de créer des doubles hélices dans lesquelles chaque nucléotide interagit avec un autre nucléotide du brin opposé en formant des ponts hydrogènes. Au même moment, les bases azotées relativement hydrophobes sont protégées du solvant (eau) par la formation des paires de bases et l'empilement de ces dernières, tandis que les groupements phosphates hautement polaires forment des interactions électrostatiques favorables avec l'eau et les cations. En effet, il a été démontré que les paires de bases et leur empilement sont des déterminants importants de la stabilité de l'ARN (Turner et al., 1986).

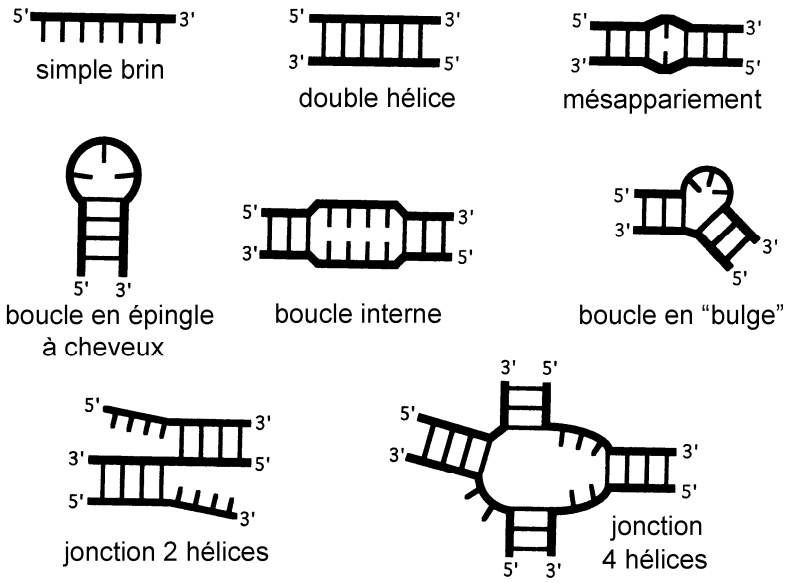


Figure 1. Différents types de motifs de structures secondaires. Les lignes épaisses représentent la chaîne de polynucléotides et les lignes fines représentent soit les bases azotées soit les paires de bases.

1.4.2.1 Représentation de la structure secondaire

La structure secondaire est habituellement utilisée pour définir quelles bases sont appariées dans une molécule. Cependant, contrairement à ce qui est observé dans l'ADN, les paires de bases Watson-Crick (WC) traditionnelles AU et GC ne sont pas les seules permises dans l'ARN. Par exemple, plusieurs paires de bases non-canoniques sont retrouvées dans les structures d'ARN (Lee and Gutell, 2004). Parmi celles-ci, les paires de bases GU sont largement répandues et, dans quelques situations, jouent un rôle fonctionnel important. Un exemple bien connu est la paire de bases GU à la position 3-70 de la tige acceptrice de l'ARNt^{Ala}, où elle sert de déterminant majeur pour son aminoacylation par l'alanine-ARNt-synthétase dans *Escherichia coli* (Hou and Schimmel, 1988; McClain et al., 1988; McClain and Foss, 1988; Francklyn and Schimmel, 1989).

À ce jour, plus d'un millier de séquences d'ARNt sont disponibles (Sprinzl et al., 1998) et sont habituellement représentées en utilisant la structure en feuille de trèfle (*cloverleaf structure*) (Figure 2A). Brièvement, cette façon de représenter la structure secondaire permet de montrer explicitement les paires de bases dans chacune des hélices

de même que la séquence de nucléotides des boucles D et T ainsi que de la tige anticodon. Une approche similaire est utilisée pour la visualisation des séquences d'ARN ribosomiques (ARNr). Les hélices et les paires de bases sont numérotées de sorte qu'il est facile de référer à une région particulière de la structure secondaire. Les structures secondaires des ARNr ont été principalement obtenues à partir d'analyses comparatives de séquences en utilisant au début, un nombre limité de séquences provenant de différents organismes (Woese et al., 1980; Noller et al., 1981; Noller and Woese, 1981). Par la suite, ces structures ont été raffinées en utilisant un nombre grandissant de séquences d'ARNr de la petite et de la grande sous-unité (Gutell et al., 1994). Ces analyses ont permis d'établir un large éventail d'éléments de structures secondaires qui sont toujours utilisés en tant que référence dans tous les travaux reliés au ribosome (Figure 3A, B).

Les gènes d'ARNr sont les plus conservés dans toutes les cellules vivantes. Pour cette raison, les gènes qui encodent les ARNr (ADNr) sont séquencés afin d'identifier le groupe taxonomique d'un organisme, calculer les groupes reliés et estimer les vitesses de divergence entre les espèces. L'alignement des séquences connues d'ARNr sont conservées dans des bases de données comme la « *European ribosomal RNA database* » (Wuyts et al., 2004).

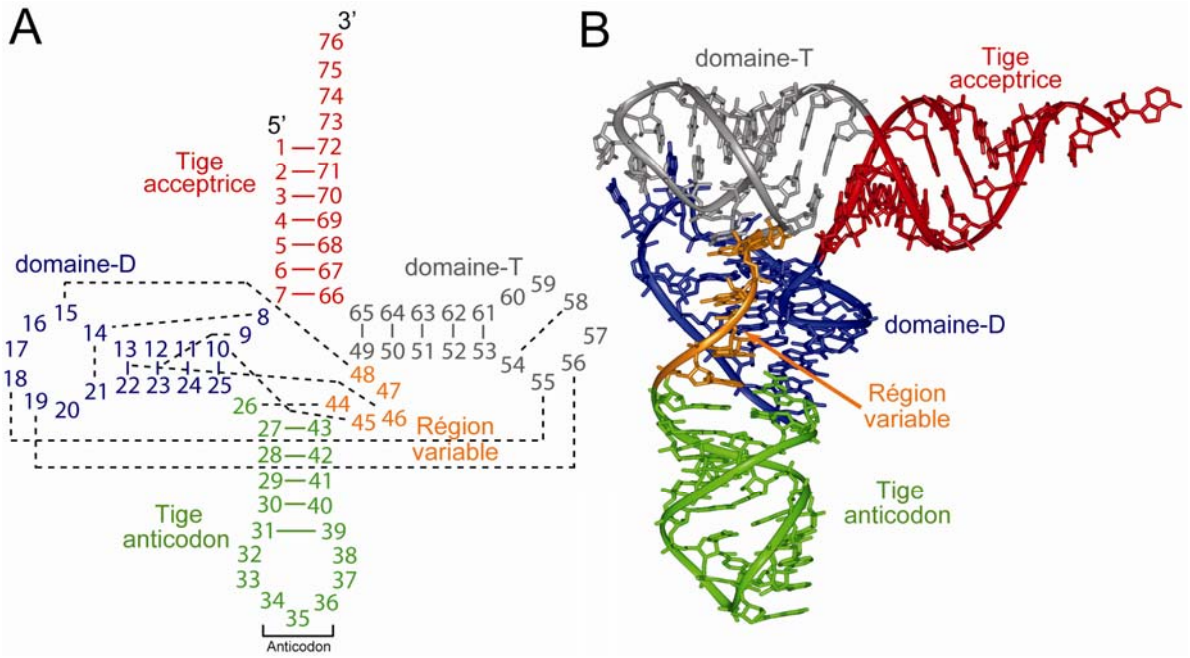


Figure 2. Structures secondaire et tertiaire de l'ARNt. (A) Représentation conventionnelle de la structure secondaire en feuille de trèfle (*cloverleaf structure*). Les paires de bases et les interactions tertiaires sont représentées respectivement par les lignes pleines et pointillées. (B) Structure tertiaire en forme de "L" de l'ARNt^{Phe} de la levure (code pdb 1ehz) (Shi and Moore, 2000). Les domaines sont colorés comme dans le panneau A.

1.4.2.2 Forme A versus forme B des doubles hélices

Dans les années 1970, la première visualisation de doubles hélices d'ARN à haute résolution a été possible suite à la détermination de plusieurs structures dimériques d'ARN (GpC et ApU) (Rosenberg et al., 1976; Seeman et al., 1976). Tel qu'anticipé, ces structures formaient des duplexes antiparallèles de pas droit avec une conformation de forme-A. En effet, une double hélice d'ARN comportant des paires de bases WC possède essentiellement une géométrie de forme-A (Figure 4A). Dans les acides nucléiques, chaque groupe phosphate est chargé négativement, de sorte que la molécule représente un polyanion fortement chargé. Par conséquent, la répulsion entre les phosphates favorise une conformation rectiligne des hélices (Williams and Maher, 2000; Tan and Chen, 2008). La forme-A idéale de l'ARN est une hélice de pas droit de onze paires de bases par tour dans laquelle les paires de bases sont déplacées d'environ 4 Å de

l'axe de l'hélice (voir Figure 4A : vue de haut) et ont une inclinaison de 18° par rapport à l'axe de l'hélice (Nowakowski and Tinoco, 1999). La forme-A d'une double hélice d'ARN diffère de la forme-B de l'ADN où l'axe de l'hélice passe directement au travers des dix paires de bases par tour (Figure 4A : vue de haut). Le déplacement accru des paires de bases dans l'ARN produit un sillon majeur étroit et profond, tandis que le sillon mineur est exposé en surface (Figure 4A : vue de côté). La présence des groupes O2'-H dans les doubles hélices d'ARN figent presqu'exclusivement les riboses dans une conformation C3'-endo; ce qui a pour effet d'éliminer la possibilité de former une hélice stable ayant une géométrie de forme-B. Par opposition, dans l'ADN, les sucres désoxyriboses peuvent alterner entre les conformations C2'-endo et C3'-endo, permettant à l'ADN d'adopter les conformations de formes-B et -A selon la nécessité.

La différence essentielle entre les conformations hélicoïdales de l'ADN et l'ARN repose sur l'accessibilité relative des sillons mineurs et majeurs. Tandis que dans l'ADN les deux sillons ont essentiellement la même accessibilité, dans l'ARN, le sillon mineur est beaucoup plus exposé; ce qui permet des interactions additionnelles dans les molécules d'ARN complexes (Figure 4B).

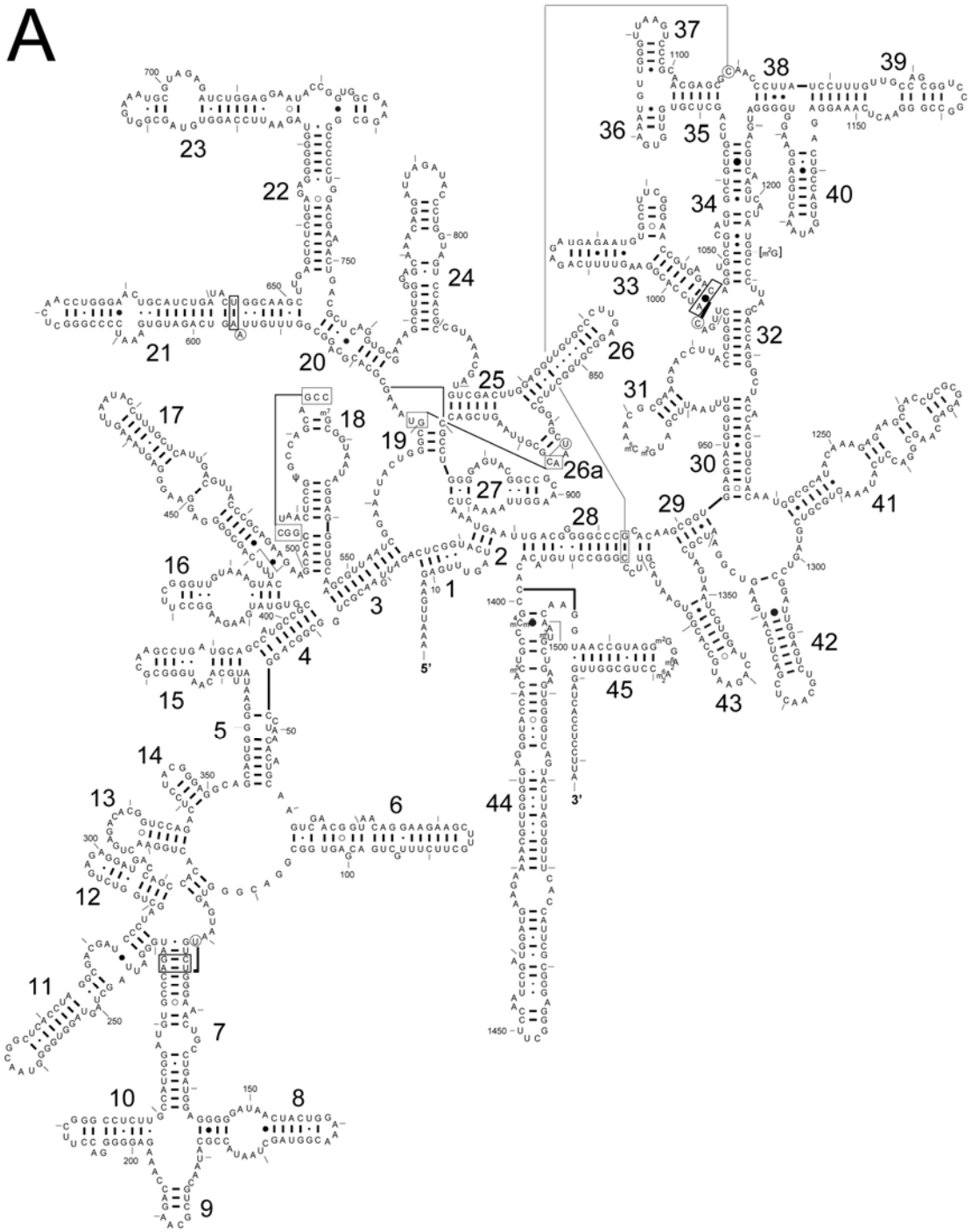


Figure 3. Voir légende à la page 14

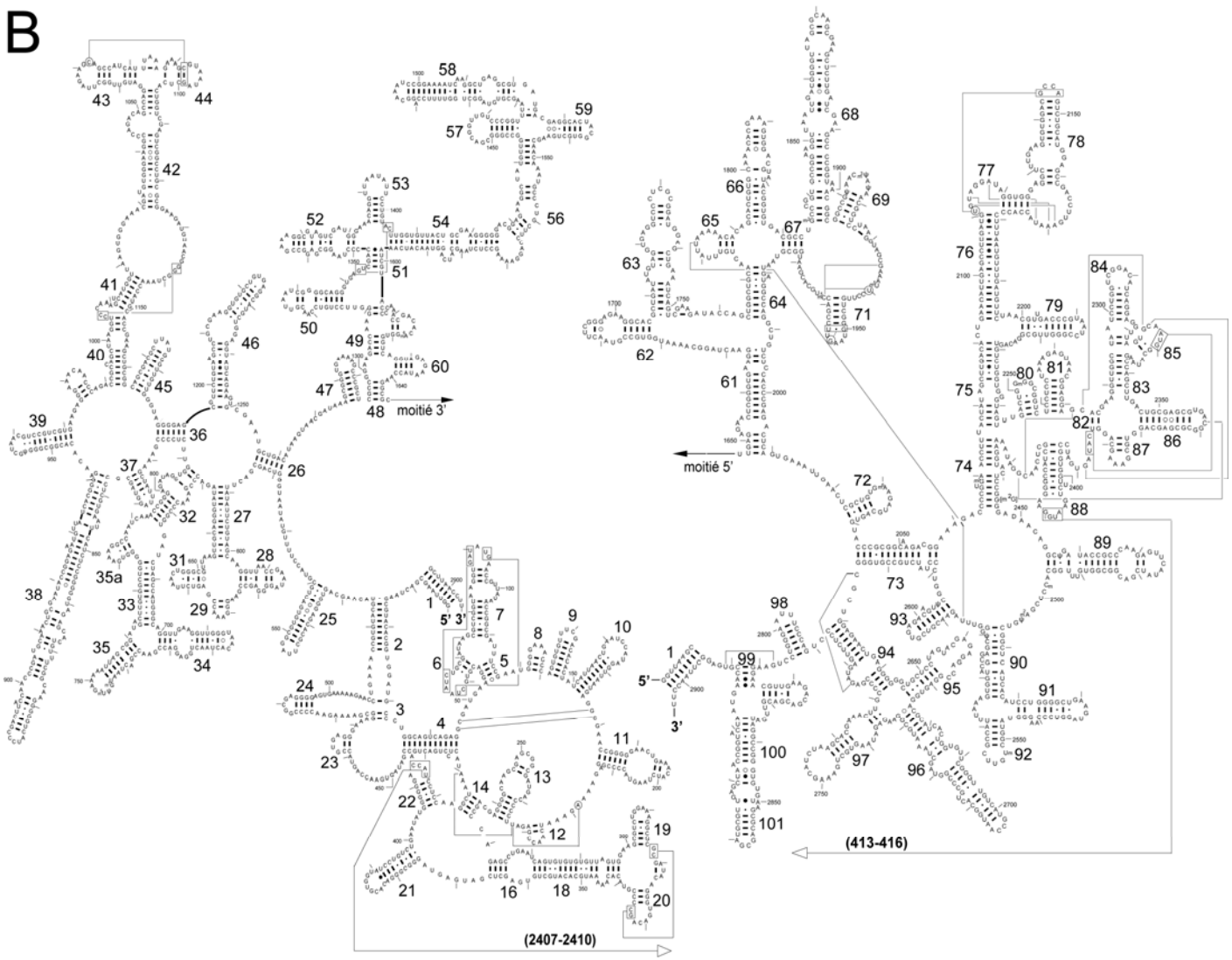


Figure 3. Voir légende page 14

Figure 3 (pages 12 et 13). Structures secondaires des ARNr 16S (A) et 23S (B) de *E. coli* avec la numérotation conventionnelle des hélices tirée de (Yusupov et al., 2001). Ces deux diagrammes ont été adaptés avec la permission de (Cannone et al., 2002).

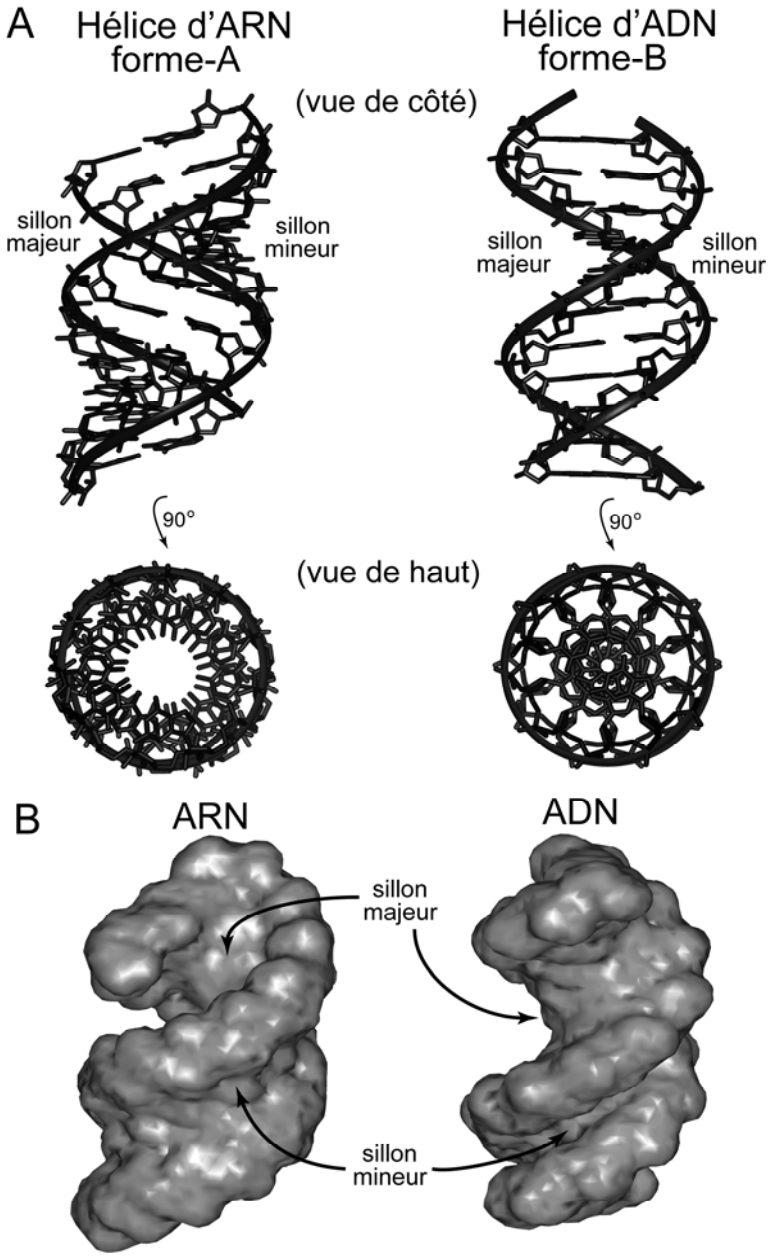


Figure 4. Structures des doubles hélices d'ARN et d'ADN. (A) Vues de côté et de haut des doubles hélices. Dans l'ARN, la présence des groupes O2'-H figent presqu'exclusivement les riboses dans une conformation permettant seulement une forme-A. (B) Comparée à la forme-B, habituellement associée aux doubles hélices d'ADN, les doubles hélices de forme-A possèdent un sillon majeur étroit et profond, tandis que le sillon mineur est exposé en surface. Ce dernier aspect permet des interactions additionnelles dans les molécules d'ARN complexes.

1.4.3 Structure tertiaire

L'arrangement tertiaire de l'ARN décrit la façon dont les éléments de structure secondaire interagissent les uns avec les autres en se repliant en structure tridimensionnelle. En guise d'exemple, on peut nommer la formation de brins triples, de quadruplexes, de pseudonoeuds, d'interactions entre boucles (*kissing loops*), entre tétraboucles et leurs récepteurs de même qu'entre doubles hélices.

Suite à la formation des doubles hélices, l'avantage énergétique de former des empilements de paires de bases en solution pousse les hélices d'ARN à former des empilements coaxiaux, un niveau supérieur d'organisation de la structure [revu par (Ferre-D'Amare and Doudna, 1999)]. Ce dernier type d'organisation domine l'arrangement de fragments en de longues hélices quasi-continues dans les cristaux (Holbrook and Kim, 1997). En 1974, l'ARNt^{Phe} de la levure a été le premier ARN naturel à voir sa structure déterminée par cristallographie aux rayons-X (Kim et al., 1974; Robertus et al., 1974) et illustre bien le dernier concept. Cette molécule d'ARN se replie en forme de « L » dans laquelle deux empilements d'hélices coaxiales se rencontrent à angle droit. Les tiges T et acceptrice s'empilent ensemble, tandis que les tiges D et anticodon forment une deuxième hélice continue (Figure 2B). Ces deux domaines hélicoïdaux interagissent *via* leurs boucles terminales T et D, de même que *via* les nucléotides des régions connectrices qui lient les quatre hélices. La majorité de ces interactions tertiaires aboutissent à la formation de courts triplex de paires de bases.

Tel que mentionné ci-dessus, l'ARN est un polyanion fortement chargé. Pour cette raison, un aspect important au problème du repliement de l'ARN est de comprendre comment les fortes répulsions électrostatiques entre les groupements phosphates empaquetés de façon serrée sont compensées dans les structures d'ARN compactes. La présence d'ions, comme le magnésium, joue un rôle important à réduire ces répulsions et à stabiliser les structures tertiaires de la plupart des molécules d'ARN (Misra and Draper, 1998, 2002; Woodson, 2005). En fait, il n'est pas possible de décrire les premières étapes du repliement de l'ARN sans invoquer le rôle des contres-ions [revu par (Thirumalai et al., 2001)].

Dans la structure en forme de « L » des ARNt, des études précédentes ont montré que des ions divalents se liaient préférablement à la région centrale de l'ARNt, où l'on retrouve un grand potentiel électrostatique induit par la forte concentration de groupements phosphates (Gueron and Leroy, 1982; Ferre-D'Amare and Doudna, 1999). Dans les molécules d'ARN complexes, en plus des empilements coaxiaux des hélices, les hélices voisines s'approchent les unes des autres et forment un repliement tridimensionnel compact. La présence de contres-ions donne plus de flexibilité aux doubles hélices d'ARN de sorte qu'elles peuvent se courber et se replier sur elles-mêmes. Dû au fait que le sillon mineur des hélices d'ARN de forme-A est plus accessible (Figure 4B), les doubles hélices auront une tendence d'interagir ensemble principalement *via* leurs sillons mineurs. Tandis que l'empilement coaxial des hélices, qui est favorisé par les surfaces essentiellement planes des paires de bases terminales, laisse peu de possibilité dans la configuration des assemblages, l'empaquetage des hélices côte à côte se présente de plusieurs façons. En effet, les interactions hélice-hélice dans les molécules d'ARN complexes représentent un aspect fondamental dans la formation d'une structure tridimensionnelle finale. Tel que nous le verrons dans les prochaines sections, la structure cristallographique du ribosome a révélé une grande concentration de régions qui forment des doubles hélices, faisant du ribosome l'objet biologique le plus approprié pour étudier l'empaquetage des doubles hélices d'ARN.

1.5 Le ribosome

Les ribosomes, qui jouent un rôle central dans le mécanisme de la synthèse des protéines en convertissant le message porté par les ARNm en chaînes polypeptidiques, ont fait l'objet d'études structurales et biochimiques pendant plus de 50 ans. Composé de deux sous-unités individuelles, le ribosome procaryote possède un poids moléculaire de 2.5 MDa, duquel 60% est de l'ARN et 40% des protéines.

1.5.1 L'organisation moléculaire du ribosome

Les premières images par microscopie électronique des ribosomes dans une cellule ont été obtenues par George Palade, et ont été initialement nommées particules de « Palade » (Palade, 1955). Dans les années 1960, des études biochimiques chez *E. coli* ont montré que la plus large des sous-unités de ce complexe protéine-ARN catalysait la formation du lien peptidique, tandis que les interactions entre l'anticodon de l'ARNt et l'ARNm lié à la petite sous-unité effectuaient la traduction du message. Chez les procaryotes, la plus large des deux sous-unités possède un coefficient de sédimentation de 50S, un poids moléculaire d'environ 1.5 MDa, et contient environ 3000 nucléotides d'ARNr et 34 protéines. La plus petite sous-unité a un coefficient de sédimentation de 30S, un poids moléculaire d'environ 0.8 MDa, et contient environ 1500 nucléotides d'ARNr et 21 protéines [revu par (Steitz, 2008)]. Nous savons maintenant que les deux sous-unités contiennent trois sites de liaison pour les molécules d'ARNt qui sont dans trois états fonctionnels différents. Le site A (*Aminoacyl*) lie l'ARNt aminoacylé qui est sur le point d'incorporer son acide aminé à la chaîne polypeptidique en élongation, le site P (*Peptidyl*) accommode l'ARNt portant le peptide et le site E (*Exit*) est occupé par tous les ARNt désaminoacylés avant leur dissociation du ribosome (Figure 6C). De nos jours, le processus de traduction est décrit comme étant composé de trois étapes principales: l'initiation, l'élongation et la terminaison de la synthèse protéique.

1.5.2 Survol de la synthèse protéique

Dans cette sous-section, j'effectue un survol de chacune des trois étapes impliquées durant la synthèse protéique chez les procaryotes.

1.5.2.1 L'initiation

L'initiation chez les bactéries implique une interaction entre la sous-unité 30S et la séquence Shine-Dalgarno (SD) de l'ARNm, un segment riche en purines et

complémentaire à la séquence anti-Shine-Dalgarno (ASD) localisée à l'extrémité 3' de l'ARNr 16S, laquelle est composée principalement de pyrimidines (Shine and Dalgarno, 1974). La séquence SD est localisée de 4 à 8 nucléotides en amont du codon initiateur; ce qui positionne le codon AUG au site P de la 30S. Le processus implique également trois facteurs d'initiation (*initiation factors*), IF1, IF2 et IF3. IF3 se lie fortement à la sous-unité 30S et empêche son association avec la sous-unité 50S (Kaempfer, 1972). Il aide aussi à la sélection de l'ARNt initiateur (fMet-ARNt^{fMet}) en déstabilisant la liaison d'autres ARNt au site P du ribosome (Gualerzi et al., 1977). IF2 est une GTPase qui lie de façon préférentielle fMet-ARNt^{fMet} et son affinité pour le ribosome est augmentée par le facteur IF1. Basé sur des données biochimiques précédentes (Moazed et al., 1995) et maintenant sur la structure cristallographique du complexe 30S-IF1 [(Carter et al., 2001) et revu dans la prochaine section], IF1 se lie au site A de la sous-unité ribosomique 30S et par le fait même, empêche la liaison d'ARNt au site A. La fin du processus d'initiation laisse un ARNt initiateur aminoacylé dans le site P du ribosome et un site A vacant, lequel sert à démarrer le cycle d'élongation. Suite à l'association des sous-unités 30S et 50S, un ARNt aminoacylé est amené au site A sous forme de complexe ternaire avec EF-Tu·GTP (Figure 5A). Les interactions codon-anticodon entraînent des changements conformationnels dans le ribosome qui stabilisent la liaison de l'ARNt dans le site A et stimulent l'hydrolyse du GTP par EF-Tu. Ceci conduit à la relâche de l'extrémité aminoacylée de l'ARNt lié au site A par EF-Tu, permettant à cette dernière de se lier au site peptidyle-transfert (PTC pour *Peptidyl-Transferase Center*) de la sous-unité 50S dans un processus nommé accommodation (Figure 5B). Durant ce processus, également appelé « *proofreading* », l'ARNt qui occupe maintenant le site A dans les deux sous-unités 30S et 50S (A/A) peut toujours être rejeté.

1.5.2.2 L'élongation

Au moment où l'extrémité aminoacylée de l'ARNt lié au site A entre dans le centre peptidyle-transfert (PTC), le lien peptidique se forme rapidement et spontanément (Cooperman, 1977; Moazed and Noller, 1989; Pape et al., 1998). Les évidences biochimiques pour un rôle de l'ARNr 23S dans la réaction peptidyle-transférase se sont

accumulées [revu par (Green and Noller, 1997)] et ultimement, ont permis l'établissement d'un rôle catalytique définitif de l'ARNr. Suite au transfert du peptide, l'ARNt lié au site P est maintenant désaminoacylé et l'ARNt au site A possède une chaîne peptidique avec un résidu additionnel. Afin de préparer le ribosome à un nouveau cycle d'élongation du peptide, les ARNt doivent être déplacés: l'ARNt désaminoacylé doit être déplacé du site P vers le site E et éventuellement éjecté du ribosome, tandis que l'ARNt portant le peptide doit être déplacé du site A vers le site P.

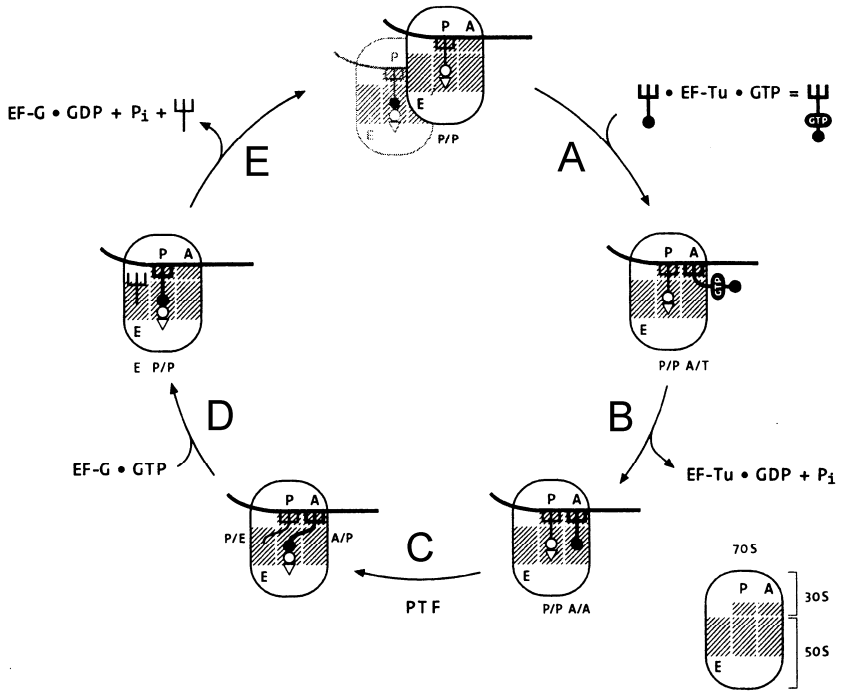


Figure 5. Le cycle d'élongation. (A) Suite à l'association des sous-unités 30S et 50S, un ARNt aminoacylé est apporté dans le site A en tant que complexe ternaire avec EF-Tu·GTP. (B) Accomodation de l'ARNt aminoacylé dans le site A. Suite à l'hydrolyse du GTP par EF-Tu, l'extrémité aminoacylée de l'ARNt lié au site A est relâchée de EF-Tu, permettant à cette dernière de se lier au centre peptidyle-transfert (PTC) de la sous-unité 50S dans un processus nommé accommodation. (C) Réaction de trans-peptidation spontanée. Suite au transfert du peptide, l'ARNt lié au site P est maintenant désaminoacylé et l'ARNt au site A porte maintenant la chaîne polypeptidique plus longue d'un résidu; ceci permet aux ARNt de prendre les positions hybrides A/P et P/E. (D) Translocation. EF-G·GTP provoque la translocation du complexe anticodon-ARNm; ce qui se traduit par un site A vacant et les ARNt peptidyle et désaminoacylé maintenant retrouvés dans les états P/P et E/E, respectivement. (E) Dissociation de EF-G·GDP et de l'ARNt désaminoacylé du ribosome.

Vers la fin des années 1960, il a été proposé que la translocation des ARNt implique un mouvement relatif des deux sous-unités et que le mouvement d'une sous-unité puisse avoir lieu à un moment différent de l'autre sous-unité (Bretscher, 1968). Ceci entraînerait le positionnement des molécules d'ARNt dans des états hybrides, permettant ainsi à un ARNt d'occuper le site A dans la 30S et le site P dans la 50S (A/P). Ce modèle permettait une rationalisation de l'existence universelle de deux sous-unités dans toutes les espèces vivantes. L'évidence expérimentale sur la nature du mouvement des ARNt durant la translocation a été établie plus tard en utilisant des techniques d'empreintes (*footprint*) sur le ribosome. Les empreintes caractéristiques des ARNt dans chaque site ont été utilisées afin de suivre le mouvement des ARNt au cours du cycle d'élongation (Moazed and Noller, 1989). Suite à sa dissociation de EF-Tu, la tige acceptrice de l'ARNt aminoacylé entre dans le site PTC de la 50S; ce qui se traduit par une empreinte de l'ARNt dans un état classique A/A (Figure 5B). Cependant, suite au transfert du peptide, il a été montré que de façon spontanée, les ARNt occupent les sites respectifs A et P dans la sous-unité 30S, mais que leurs tiges acceptrices ont été déplacées aux sites P et E de la sous-unité 50S, en accord avec les états hybrides A/P et P/E des ARNt (Figure 5C). L'addition du facteur EF-G·GTP provoque la translocation rapide du complexe anticodon-ARNm; ceci se traduit par un site A vacant tandis que les ARNt peptidyle et désaminoacylé sont maintenant retrouvés dans les états P/P et E/E, respectivement (Figure 5D), d'où l'ARNt désaminoacylé peut se dissocier du ribosome (Figure 5E).

Puisque l'article de Bretscher (Bretscher, 1968) était complètement théorique, il n'y avait aucune évidence d'un mouvement d'une ou l'autre des sous-unités 30S et 50S par rapport à l'autre. Récemment, un mouvement relatif des sous-unités a été observé lors d'études par microscopie électronique (cryo-EM) (*cryo-Electron Microscopy*) suite à la comparaison du ribosome sous forme libre et lié au facteur EF-G (Frank and Agrawal, 2000). Ceci constitue le plus grand changement conformationel dans le ribosome, connu sous le nom de rotation (*ratchet-like movement*) de la petite sous-unité par rapport à la grande sous-unité. La rotation de la petite sous-unité pousse l'ARNm dans la direction requise pour la translocation dans la sous-unité 30S. Quelques annnées plus tard, une autre étude de cryo-EM permettait de visualiser l'ARNt désaminoacylé en

état hybride P/E (Valle et al., 2003). Récemment, deux études indépendantes de cryo-EM ont finalement permis de visualiser les ARNt aminoacylé et désaminoacylé respectivement dans les états hybrides A/P et P/E (Agirrezabala et al., 2008; Julian et al., 2008). *In vivo*, les ribosomes procaryotes ajoutent entre 15 et 20 acides aminés par seconde à la chaîne polypeptidique en croissance tout en y incorporant en moyenne une erreur tous les 1000 à 10 000 résidus.

1.5.2.3 Terminaison

La traduction prend fin lorsqu'un codon de terminaison sur l'ARNm est retrouvé au site A. Il existe trois facteurs de terminaison (*release factors*): RF1, RF2 et RF3. RF1 et RF2 reconnaissent le codon UAA, tandis que UAG est reconnu par RF1 et UGA par RF2. La liaison de RF1/2 à un ribosome ayant un codon de terminaison approprié au site A stimule l'hydrolyse et la relâche de la chaîne peptidique de l'ARNt lié au site P. RF3, qui est une GTPase, favorise la dissociation rapide de RF1 et RF2 [revu par (Ramakrishnan, 2002)]. Suite au relâchement de la chaîne peptidique, l'ARNm et l'ARNt désaminoacylé dans le site P sont toujours liés au ribosome. Ce complexe doit être désassemblé afin de préparer le ribosome à un nouveau cycle d'élongation. Le facteur de recyclage du ribosome (RRF) (*Ribosome Recycling Factor*), accompagné du facteur EF-G, est nécessaire à ce processus [(revu par (Janosi et al., 1996)]. Pour une revue récente du cycle de la traduction, voir (Marshall et al., 2008).

1.6 Structure tertiaire du ribosome

Les premières observations de la structure du ribosome ont été obtenues par microscopie électronique (EM) (*Electron Microscopy*) et identifiaient une particule subdivisée en deux sous-unités de grosseurs différentes (Huxley and Zubay, 1960). La première détermination de la forme des sous-unités a été obtenue au courant des années 1970 (Lake, 1976). Aujourd'hui, la résolution des structures ribosomiques obtenues par cryo-EM a augmenté à 6.7 Å pour les meilleures reconstitutions (Villa et al., 2009). Les structures à haute résolution des ribosomes peuvent cependant être obtenues par

cristallographie rayons-X. Néanmoins, ce niveau de résolution pour le ribosome est demeuré innacessible durant plusieurs années. Au début de ce siècle, en raison des progrès et du travail acharné dans le domaine de la cristallographie rayons-X, plusieurs structures à haute résolution de chacune des sous-unités ribosomiques de même que du ribosome entier sont apparues.

1.6.1 Structures à haute résolution du ribosome

Vingt ans suite à la publication de la première structure d'ARNt en 1974, la structure cristallographique du ribozyme à la tête en forme de marteau (*hammerhead ribozyme*) est apparue et représentait la deuxième molécule d'ARN de grandeur comparable à l'ARNt à être disponible (Pley et al., 1994b). En 1996, la structure cristallographique du domaine P4-P6 de l'intron auto-épisseur du groupe I de *Tetrahymena thermophila* a révélé plusieurs aspects reliés à la structure secondaire et tertiaire de l'ARN (Cate et al., 1996). Ces avancées ne représentaient que le début de la complexité des molécules d'ARN déterminées à haute résolution. En 2000, les premières structures à haute résolution de l'ARNr de la grande sous-unité de *Haloarcula marismortui* (incluant les ARNr 5S, 23S et les protéines) (Ban et al., 2000) et de la petite sous-unité de *Thermus thermophilus* (incluant l'ARNr 16S et les protéines) (Schluenzen et al., 2000; Wimberly et al., 2000), ont été publiées (Figure 6A,B). Au courant de l'année qui a suivi, la structure cristallographique de la sous-unité 50S de la bactérie *Deinococcus radioduarans* a été rapportée (Harms et al., 2001). Les structures des sous-unités ont permis d'obtenir les détails de la liaison au ribosome de quelques antibiotiques (Brodersen et al., 2000; Carter et al., 2000; Brodersen et al., 2001; Pioletti et al., 2001; Schlunzen et al., 2001) et ont également permis l'étude à haute résolution des interactions entre des ligands et facteurs fonctionnels et les sous-unités 30S et 50S (Carter et al., 2001; Ogle et al., 2001; Pioletti et al., 2001). De plus, la disponibilité des structures à haute résolution des deux sous-unités a facilité la reconstruction d'un modèle des squelettes de l'ARNr et des protéines pour le ribosome 70S de *Thermus thermophilus* à une résolution de 5.5 Å (Yusupov et al., 2001). Cette structure de 70S inclut un ARNm et des molécules d'ARNt liées aux sites A, P et E du ribosome, de sorte

que les interactions entre le ribosome et ces ligands, de même que les interactions entre les deux sous-unités, pouvaient être interprétées d'un point de vue moléculaire.

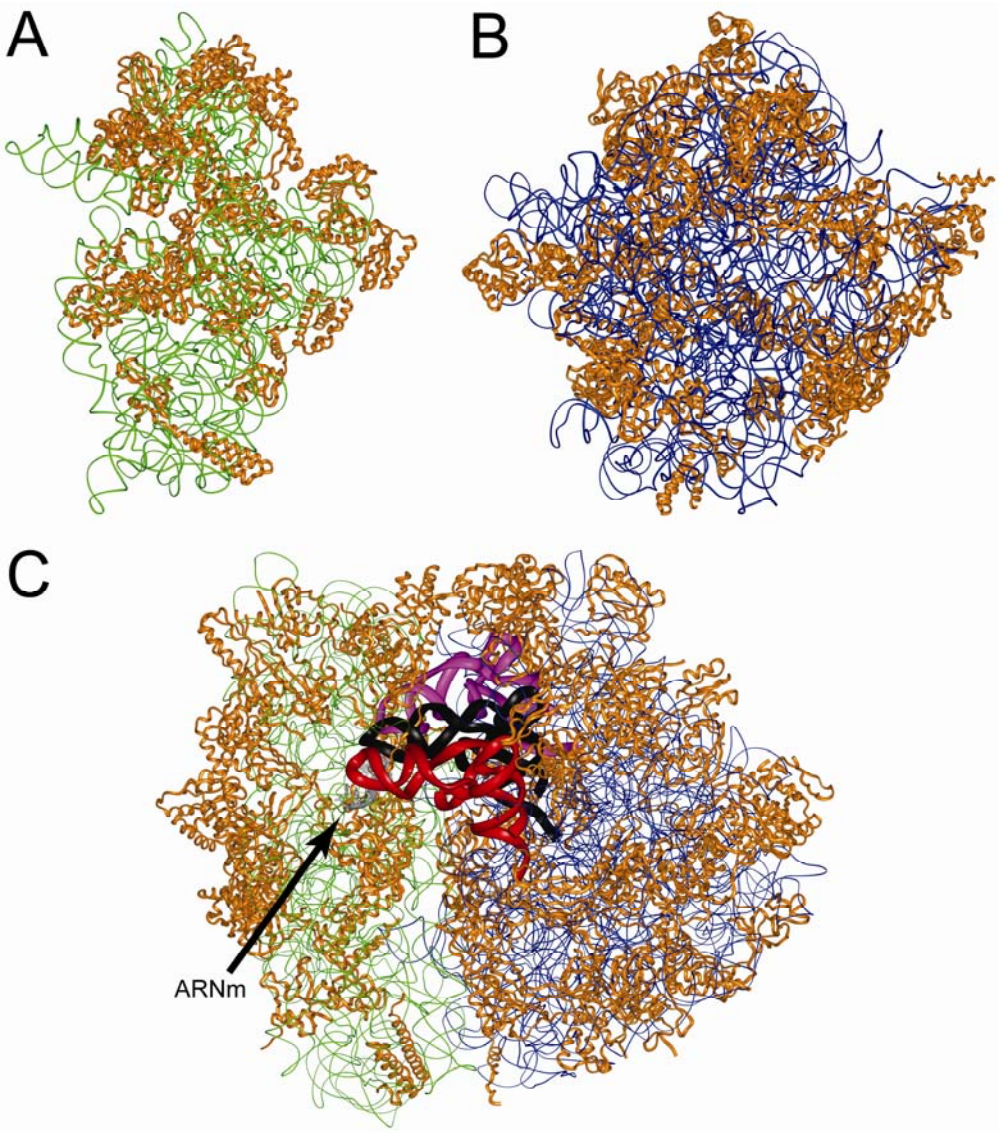


Figure 6. Structures des sous-unités ribosomiques individuelles vues de l'interface (A, B) et du ribosome 70S (C). (A) Structure de la sous-unité 30S de *T. thermophilus* (code pdb 1j5e) (Wimberly et al., 2000). L'ARNr 16S et les protéines sont colorés en vert et orange, respectivement. (B) Structure de la sous-unité 50S de *H. marismortui* (code pdb 1s72) (Ban et al., 2000). L'ARNr 23S et les protéines sont colorés en bleu et orange, respectivement. (C) Structure du ribosome 70S de *T. thermophilus* (codes pdb 2j00-2j01) (Selmer et al., 2006). Les ARNr 16S et 23S sont colorés comme dans les panneaux A et B, respectivement. Les ARNt localisés aux sites A, P et E sont colorés rouge, noir et magenta, respectivement. L'ARNm est coloré en gris.

Au cours des trois dernières années, trois groupes ont déterminé la structure du ribosome 70S à haute résolution; ce qui a permis de révéler les détails à l'échelle atomique des interactions entre les deux sous-unités ribosomiques (Schuwirth et al., 2005; Korostelev et al., 2006; Selmer et al., 2006). En effet, ces structures ont permis de visualiser les ponts entre les sous-unités qui se forment par des contacts ARNr-ARNr, ARNr-protéine et protéine-protéine. De plus, la structure du ribosome 70S de *Thermus thermophilus* à une résolution de 2.8 Å en complexe avec des molécules d'ARNt aux sites A, P et E, a révélé les détails de leurs interactions avec la sous-unité 50S (Selmer et al., 2006) (Figure 6C). Une liste des structures à haute résolution du ribosome et de ses sous-unités est présentée dans le Tableau I (page suivante).

Tableau I. Liste des structures à haute résolution du ribosome et de ses sous-unités

code PDB	Description	Résolution (Å)	Organisme	Référence
1j5e	Sous-unité 30S	3.0	*T. thermophilus*	(Wimberly et al., 2000)
1hr0	Sous-unité 30S liée au facteur d'initiation-1 (IF1)	3.2	*T. thermophilus*	(Carter et al., 2001)
1fka	Sous-unité 30S	3.3	*D. radiodurans*	(Schluenzen et al., 2000)
1s72	Sous-unité 50S	2.4	*H. marismortui*	(Ban et al., 2000)
1nkw	Sous-unité 50S	3.1	*D. radiodurans*	(Harms et al., 2001)
1gix, 1giy	Ribosome 70S en complexe avec un ARNm et des ARNt aux sites A, P et E	5.5	*T. thermophilus*	(Yusupov et al., 2001)
2avy, 2aw4	Ribosome 70S	3.5	*E. coli*	(Schuwirth et al., 2005)
2j00, 2j01	Ribosome 70S en complexe avec un ARNm et des ARNt aux sites A, P et E	2.8	*T. thermophilus*	(Selmer et al., 2006)
2ow8, 1vsa	Ribosome 70S en complexe avec un ARNm et des ARNt aux sites P et E	3.7	*T. thermophilus*	(Korostelev et al., 2006)

1.6.2 Analyse systématique des structures d'ARN

Afin de comprendre comment l'ARN se replie et accomplit sa fonction, il est nécessaire d'identifier tous les arrangements structuraux récurrents, nommés motifs, qui sont capables de se replier en structures similaires ou identiques et ce, dans différents contextes structuraux [revu par (Batey et al., 1999; Moore, 1999; Noller, 2005)]. Une comparaison de tous les représentants d'un motif récurrent permet d'identifier les aspects importants de sa structure nécessaires à sa formation et de les distinguer de ceux

qui peuvent être sacrifiés. C'est seulement à ce moment qu'il devient possible d'établir les prérequis de séquences responsables à la formation d'un motif récurrent, lesquels peuvent être ensuite vérifiés expérimentalement. L'analyse des sous-unités ribosomiques a révélé quelques arrangements qui avaient déjà été observés dans d'autres structures, comme les motifs « *A-minor* » et de la boucle T (*T-loop*), de même que de nouveaux motifs (revu dans la prochaine section). Parmi les nouveaux arrangements qui ont été identifiés pour la première fois dans la structure du ribosome, nous retrouvons le motif d'empaquetage le long du sillon (AGPM) (*along-groove packing motif*), qui constitue le sujet de cette thèse.

1.7 Les motifs tertiaires des structures d'ARN

Dans cette section, je fais une révision de quelques motifs d'ARN communs qui ont été identifiés avant même l'apparition des structures du ribosome et par la suite retrouvés dans les ARNr, ou identifiés pour la première fois suite à l'analyse de l'architecture du ribosome.

1.7.1 Le motif "A-minor"

Cet arrangement consiste en un empilement d'adénosines non-appariées habituellement faisant partie d'une double hélice ou d'une tétraboucle qui empaquettent leurs sillons mineurs contre le sillon mineur d'une hélice réceptrice. Ce type d'empaquetage a été observé pour la première fois en tant que contact intermoléculaire retrouvé entre deux molécules de ribozyme à la tête en forme de marteau (*hammerhead ribozyme*) à l'intérieur de la même unité asymétrique du cristal (Pley et al., 1994a). Deux ans plus tard, la structure du domaine P4-P6 de l'intron du groupe I de *Tetrahymena thermophila* a été la première à illustrer ce motif en tant que contact tertiaire capable de stabiliser deux hélices (Cate et al., 1996). Au moment où les structures des sous-unités ribosomiques sont devenues disponibles, Nissen et ses collaborateurs ont établi que c'était de loin l'élément structural le plus communément retrouvé dans la grande sous-unité ribosomique et par extension, dans les molécules d'ARN complexes en général;

ceci leur a permis de nommer cet arrangement motif « *A-minor* » (Nissen et al., 2001). Basé sur la position du groupement O2′-H de l'adénosine par rapport aux groupements O2′-H de la paire de bases réceptrice, ces interactions ont été divisées en trois types principaux (Figure 7). Des études de repliement de l'intron du groupe I de *Tetrahymena thermophila* ont révélé l'importance de la présence de paires de bases canoniques dans l'hélice réceptrice permettant aux adénosines de reconnaître spécifiquement la géométrie WC du sillon mineur (Doherty et al., 2001; Battle and Doudna, 2002).

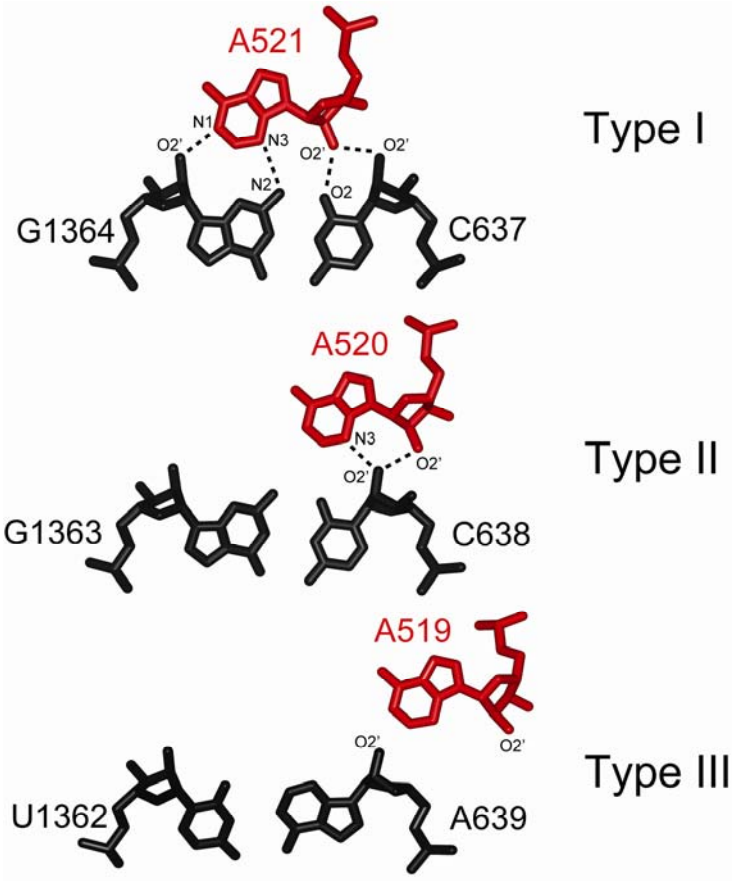


Figure 7. Différents types d'interactions du motif « *A-minor* » tirés de la structure de la sous-unité ribosomique 50S de *H. marismortui* (code pdb 1s72) (Ban et al., 2000). La numérotation des nucléotides est la même que celle utilisée lors de la publication originale de la structure. L'adénosine qui s'empaquette contre le sillon mineur d'une double hélice est rouge. Chaque type de « *A-minor* » est défini par la position du groupe O2′-H de l'adénosine par rapport aux positions des deux groupes O2′-H de la paire de bases réceptrice. Les lignes pointillées indiquent les ponts hydrogènes.

Dans la structure cristallographique de la sous-unité 30S en complexe avec différents antibiotiques (Carter et al., 2000), la molécule paromomycine se lie à la boucle interne située au haut de l'hélice h44 et favorise des réarrangements structuraux spécifiques. Dans cette structure, les résidus A1492 et A1493 sont retrouvés complètement en dehors de l'hélice de sorte qu'ils sont dans une position leur permettant d'interagir directement avec le sillon mineur de l'hélice formée par l'interaction codon-anticodon dans le site A. La structure cristallographique de la sous-unité 30S en complexe avec l'ARNm et un ARNt dans le site A montre également ce changement conformationnel (Ogle et al., 2001; Ogle et al., 2002). En effet, les nucléotides A1492 et A1493 conservés de façon universelle interagissent avec le sillon mineur des deux premières paires de bases de l'hélice codon-anticodon, leur permettant ainsi de scruter la géométrie des appariements WC et de discriminer les ARNt qui forment de mauvais appariements. Ces observations ont expliqué le mécanisme d'action de l'antibiotique paromomycine. Cette molécule permet aux ARNt formant des mauvais appariements avec un codon de l'ARNm de passer avec succès l'étape du « *proofreading* » du ribosome; ce qui favorise l'acceptation de ces « mauvais » ARNt par le ribosome. Ceci a comme résultat l'incorporation de mauvais acides aminés dans les protéines; ce qui conduit finalement à la mort cellulaire [revu par (Ogle et al., 2003)].

1.7.2 Le motif "ribose zipper"

Le motif « *ribose zipper* » est un arrangement structural récurrent qui comprend un réseau spécifique de ponts hydrogènes. Le groupement O2'-H du nucléotide (*a*) donne un pont hydrogène au groupement O2'-H d'un deuxième nucléotide (*b*), qui à son tour donne un pont hydrogène à la base azotée du nucléotide (*a*) (Figure 8). Ce motif a été observé dans les structures cristallographiques du domaine P4-P6 de l'intron du groupe I de *Tetrahymena* (Cate et al., 1996), dans le ribozyme de HDV (*hepatitis delta virus*) (Ferre-D'Amare et al., 1998), et entre deux molécules cristallisées du ribozyme à la tête en forme de marteau (*hammerhead ribozyme*) (Pley et al., 1994a, b). Plus tard, ce motif a été retrouvé à près d'une centaine d'endroits dans la structure du ribosome (Tamura and Holbrook, 2002).

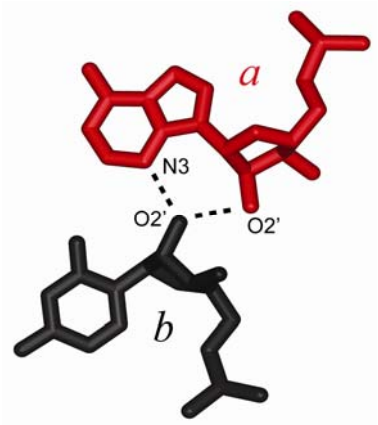


Figure 8. Le motif « *ribose zipper* ». Ce motif représente une interaction largement retrouvée dans les structures d'ARN. Le groupe O2′-H d'un nucléotide (*a*) donne un pont hydrogène au groupe O2′-H du deuxième nucléotide (*b*), qui à son tour donne un pont hydrogène à la base azotée du nucléotide (*a*). Les lignes pointillées indiquent les ponts hydrogènes.

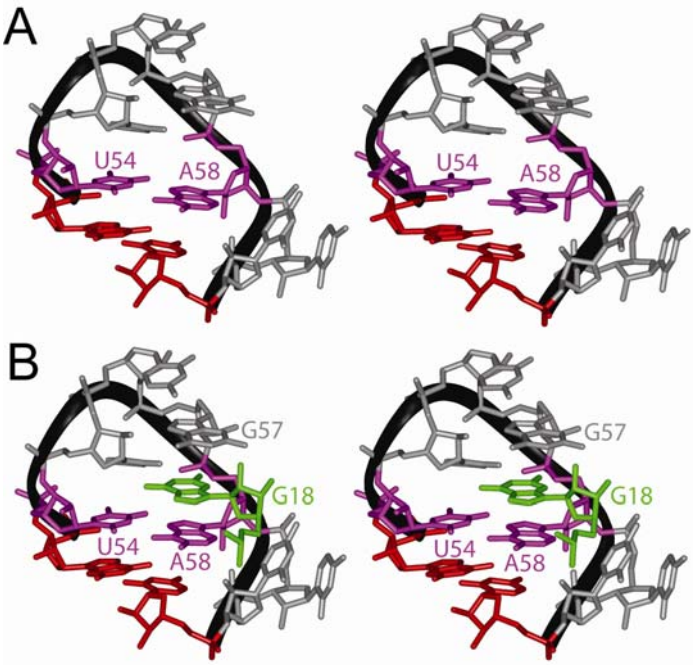


Figure 9. Stéréogramme de la structure du motif de la boucle T tirée de l'ARNt^{Phe} de la levure (code pdb 1ehz) (Shi and Moore, 2000). (A) La dernière paire de bases de la tige T est rouge. La paire de bases « *reverse-Hoogsteen* » U54-A58 est magenta. (B) Le double espacement entre les nucléotides G57 et A58 de la boucle T est utilisé pour emprisonner et fixer le nucléotide G18 (vert) provenant de la boucle D.

1.7.3 Le motif de la boucle T

Le motif de la boucle T (*T-loop*) a été observé pour la première fois dans la région « *elbow* » de la structure de l'ARNt^{Phe} de la levure (Kim et al., 1974; Robertus et al., 1974). Ce motif à cinq nucléotides contient un virage en « U » flanqué par une paire de bases non-canonique et possède une structure tridimensionnelle distincte (Figure 9A). Par la suite, ce même arrangement a été identifié à neuf endroits dans la structure du ribosome (Nagaswamy and Fox, 2002). Dans la structure du domaine de spécificité de la RNase P de *Bacillus subtilis*, il a été noté que le motif de la boucle T pouvait exister dans différents contextes qui n'incluaient pas nécessairement la même paire de bases non-canonique observée précédemment (Krasilnikov et al., 2003). En permettant un consensus de séquence plus large, huit cas additionnels ont été identifiés dans la structure du ribosome (Krasilnikov and Mondragon, 2003). Des analyses supplémentaires ont montré que ce motif est utilisé pour emprisonner un autre nucléotide appartenant à une autre portion de la structure. En effet, dans les ARNt cytosoliques, le double espacement entre les nucléotides G57 et A58 de la boucle T emprisonnent et fixent le nucléotide G18 provenant de la boucle D (Figure 9B), un aspect qui a été étudié *in vivo* avec l'aide de librairies combinatoires de gènes (Doyon et al., 2004).

1.7.4 Le motif "kink-turn"

Le motif « *kink-turn* », aussi appelé « *k-turn* », est un arrangement hélice–boucle interne–hélice qui contient environ 15 nucléotides (Klein et al., 2001). La première hélice se termine avec deux paires de bases WC suivie d'une boucle interne, tandis que la deuxième hélice qui suit la boucle interne débute avec deux paires de bases non-WC, typiquement des paires « *sheared* » GA. De plus, la structure possède un « *kink* » dans le squelette sucre-phosphate qui produit un tournant prononcé d'environ 120° dans l'hélice d'ARN, lui donnant une structure en forme de « V » (Figure 10). Retrouvé à huit endroits dans la structure du ribosome, l'analyse du motif « *k-turn* » a montré que même si différents cas du motif varient au niveau de leur séquence, chacun possède

essentiellement la même structure tridimensionnelle. Ces motifs sont également des sites importants de reconnaissance par des protéines, interagissant ainsi avec 11 protéines ribosomiques. L'identification du motif « *k-turn* » dans la structure du ribosome a permis à Klein et ses collaborateurs (Klein et al., 2001) d'identifier le même motif dans d'autres complexes contenant de l'ARN. Plus tard, des expériences de dynamiques moléculaires ont montré que quelques motifs « *k-turn* » dans l'ARNr pouvaient agir comme charnière moléculaire et de ce fait, seraient capables de transmettre des ajustements conformationnels entre des segments d'ARN éloignés les uns des autres durant le cycle de la synthèse protéique (Razga et al., 2004; Razga et al., 2005; Razga et al., 2006).

1.7.5 Le motif "hook-turn"

La structure cristallographique d'un ARN contenant 26 nucléotides a révélé un nouveau type de conformation du squelette sucre-phosphate et nommé motif « *hook-turn* » (Szép et al., 2003). Le motif consiste en une hélice de forme-A qui se sépare en deux brins suivant une paire de bases « *sheared* » AG. Le squelette du brin qui contient le G de la paire AG effectue un tournant d'environ 180° dans l'espace de deux nucléotides et interagit ensuite avec le sillon mineur de l'hélice de laquelle il provient (Figure 11). Une recherche dans les structures des ARNr 16S et 23S a révélé quatre cas pour qui le squelette sucre-phosphate possède une trajectoire très similaire au motif « *hook-turn* » original.

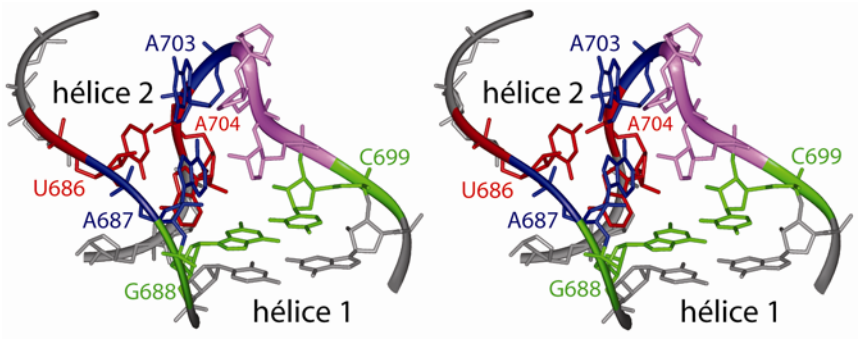


Figure 10. Stéréogramme du motif « *kink-turn* » tiré de la structure de la sous-unité ribosomique 30S de *E. coli* (code pdb 2avy) (Schuwirth et al., 2005). La dernière paire de bases WC de la première hélice (hélice 1) est verte, tandis que la deuxième hélice (hélice 2) suivant la boucle interne (magenta) débute avec deux paires de bases non-WC (bleue et rouge). La structure possède un « *kink* » dans la chaîne de polynucléotides qui occasionne un tournant prononcé d'environ 120° dans l'hélice d'ARN; ce qui lui donne une structure en forme de « V ». La numérotation des nucléotides correspond à celle utilisée dans le ribosome de *E. coli*.

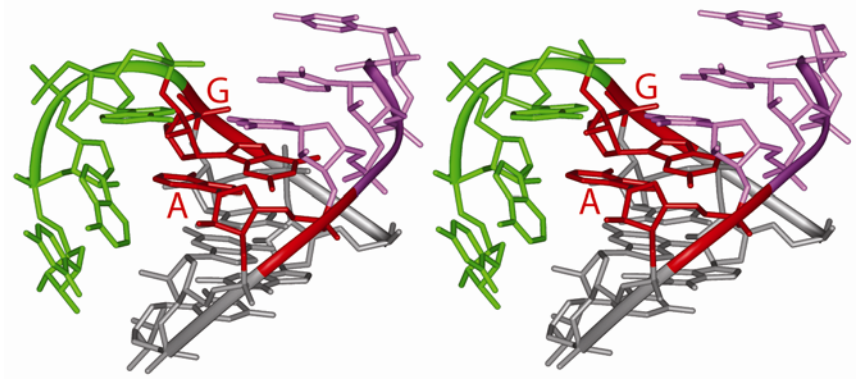


Figure 11. Stéréogramme du motif « *hook-turn* » tiré de la structure cristallographique d'un ARN contenant 26 nucléotides (code pdb 1mhk) (Szép et al., 2003). Le motif est formé par une hélice de forme-A (grise) qui se sépare en deux brins (magenta et vert) suite à une paire de bases « *sheared* » AG (rouge). Le squelette du brin qui contient le G de la paire AG effectue un tournant d'environ 180° dans l'espace de deux nucléotides.

1.8 Objectif

En résumé, l'arrivée de plusieurs structures cristallographiques du ribosome procaryote a permis des études systématiques de son architecture. Il devenait alors possible d'étudier les déterminants structuraux utilisés par la chaîne de polynucléotides pour se replier sur elle-même et interagir avec d'autres éléments de la structure. Au cours des dernières années, ces analyses ont mené à la découverte de nouveaux types d'arrangements structuraux; ce qui augmente notre compréhension de la structure de l'ARN. L'objectif principal de ma thèse est d'élucider les règles structurales qui régissent l'empaquetage entre deux doubles hélices d'ARN.

1.9 La révélation qui a conduit à la découverte du motif AGPM

L'élément déclencheur qui a lancé mon projet de doctorat s'est produit à l'automne 2001 lorsque j'assistais au cours « Protéines et acides nucléiques », dans le cadre de ma maîtrise, enseigné par mon directeur de recherche le Dr Serguei Chteinberg. Le cours portait sur la diversité et la complexité des structures tertiaires d'ARN. Le Dr Chteinberg présentait la structure nouvellement publiée de la sous-unité ribosomique 50S de *H. marismortui* par le groupe du Dr Thomas Steitz (Ban et al., 2000). À un certain moment, le Dr Chteinberg a discuté du haut niveau d'empaquetage retrouvé dans cette structure entre différentes régions de l'ARNr contenant des doubles hélices. Tel que prévu, les hélices interagissaient entre elles principalement *via* leurs sillons mineurs pour la simple raison que dans la forme-A de l'ARN, le sillon mineur est exposé en surface. Le Dr Chteinberg a aussi mentionné qu'une des façons utilisées par les hélices pour former des empaquetages est par l'insertion du squelette sucre-phosphate d'une hélice dans le sillon mineur de l'autre hélice, et *vice versa*. Au moment où il mentionnait que la présence de guanosines dans les doubles hélices formant des empaquetages aiderait nécessairement à la stabilisation de ces arrangements en neutralisant les charges négatives des groupements phosphates, j'ai pensé que des paires de bases GU pourraient aider davantage. En effet, en raison de leur géométrie caractéristique, les paires de bases GU permettent au groupement amine de la guanosine une exposition accentuée dans le

sillon mineur; ce qui pouvait faciliter la neutralisation des charges négatives. À la fin du cours, j'ai discuté avec le Dr Chteinberg et je lui ai dévoilé mon idée. Il m'a dit: "*It could be possible, I don't know. Go through the ribosome structure and look for all instances where two double helices pack between themselves. Then, we will see*".

La journée suivante, j'ai débuté l'analyse de la structure du ribosome en décomposant la chaîne de polynucléotides en plusieurs fragments pour ensuite être en mesure de vérifier systématiquement l'interaction de chaque région avec toutes les autres. Après environ une semaine, j'avais en main quatre représentants d'un empaquetage de deux doubles hélices, et dans chaque cas, une paire de bases GU était présente dans une seule des deux hélices. J'ai rapporté mes résultats à mon directeur de recherche et je lui ai montré les cas identifiés et la façon dont ils se superposaient. À ce moment, il m'a dit que quatre représentatifs étaient déjà suffisants pour que cet arrangement soit qualifié de nouveau motif. Au courant des quelques semaines qui ont suivi, nous avons scruté la structure du ribosome afin d'identifier tous les cas représentant ce nouvel arrangement. À ce moment, chacun de nous avions trouvé indépendamment quatre cas dans la sous-unité 30S et huit autres dans la sous-unité 50S; ce qui nous permettait de croire que nous les avions tous identifiés. Le nom motif d'empaquetage le long du sillon (AGPM) (*along-groove packing motif*) est apparu. Un peu plus tard, suite à la publication de la structure du ribosome 70S de Yusupov et ses collaborateurs (Yusupov et al., 2001), deux autres cas du motif impliqués dans la liaison des molécules d'ARNt aux sites P et E de la sous-unité 50S ont été identifiés. La découverte du motif AGPM constitue le sujet de l'article présenté dans le chapitre suivant (Chapitre 2). D'autres développements ont été effectués expérimentalement qui avaient pour but l'élucidation des prérequis de séquences pour différents représentatifs du motif localisés dans différents contextes structuraux du ribosome (Chapitres 3 et 4). Des études *in silico* ont permis la détermination des prérequis de séquences des quatre paires de bases dans chacune des deux hélices qui interagissent entre elles pour former le motif AGPM (Chapitre 5). Finalement, une analyse de tous les cas identifiés du motif AGPM a révélé que trois d'entre eux possèdent un autre élément récurrent à l'intérieur de leur structure que nous avons nommé motif « *adenosine-wedge* » (Chapitre 6).

Chapitre 2

Article 1

GU receptors of double helices mediate tRNA movement in the ribosome

2. GU receptors of double helices mediate tRNA movement in the ribosome

Matthieu G. Gagnon and Sergey V. Steinberg

Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

RNA, 2002, vol. 8, no. 7, pp. 873-877

Running title: GU base pair in packing of double helices

Contribution of each author:

Matthieu G. Gagnon: Searched the ribosomal subunit structures for close helix packing, collected samples and participated in data analysis, preparation of manuscript and figures.

Sergey V. Steinberg: Developed the general approach for the search, searched the ribosomal subunit structures for close helix packing, collected samples and participated in data analysis, preparation of manuscript and figures.

2.1 Abstract

A new RNA structural motif consisting of two double helices closely packed *via* minor grooves is found in many places in the ribosome structure. The packing requires that a GU base pair in one helix be packed against a Watson–Crick pair in the other helix. Two such motifs mediate the interaction of the P- and E-tRNA with the large ribosomal subunit. Analysis of the particular positions of these two motifs in view of the available data on occupancy of tRNA-binding sites and structural changes in the ribosome during the elongation cycle suggests a distinct role for each motif in tRNA translocation.

Keywords: double helix packing; GU base pair; ribosomal RNA; ribosomal translocation; ribosome structure

2.2 Introduction

Recent achievements in the X-ray crystallography of the whole ribosome (Yusupov et al., 2001) and its subunits (Ban et al., 2000; Schluenzen et al., 2000; Wimberly et al., 2000; Harms et al., 2001) are invaluable both for elucidating the mechanisms of the protein synthesis and for providing insight into how nucleotide sequence shapes RNA tertiary structure and how the latter determines the function. A necessary step toward these goals is a systematic analysis of the ribosome conformation, which has already succeeded with identification of new RNA structural motifs (Doherty et al., 2001; Klein et al., 2001; Nissen et al., 2001). One of the most common elements of the ribosome structure is the interaction of RNA double helices *via* minor grooves, which was acknowledged by all authors who determined X-ray structures of the ribosome or its subunits. This element has also been found in the structures of other RNA molecules (Pley et al., 1994; Strobel and Cech, 1995; Cate et al., 1996; Strobel et al., 1998) and thus should be considered as an important block of the RNA architecture in general. The presented analysis of the minor groove interactions existing in the ribosome structure shows that the close packing of two double helices imposes certain constraints on their nucleotide sequences, providing for a specific and stable complex. For two such complexes that mediate the interaction of the P- and E-tRNA with 23S rRNA, we suggest a distinct role in ribosome translocation.

2.3 Along-groove packing motif

In the heterogeneous population of RNA helix–helix contacts existing in the crystal structures of both ribosomal subunits (Ban et al., 2000; Wimberly et al., 2000), we were looking for those where the minor grooves of two helices closely packed with each other. Because the minor groove in A-RNA has a slightly concave shape, we expected to find structures like that shown in Figure 1A, where the sugar–phosphate backbone of each helix packs along the minor groove of the other in the so-called along-groove packing. Analysis of the X-ray conformations of both ribosomal subunits reveals 12 such cases, 4 in 30S and 8 in 50S, having well superimposed structures with root mean square deviation of 0.84 Å (Fig. 1B). In most cases, four base pairs from each helix are involved in contact with the other helix. The average contact area is about 150 $Å^2$ and includes more than 60 non-hydrogen atoms accountable for about 60 interhelix atom–atom interactions. Of course, because of the spiral character of the helices, it is impossible to keep the same pattern of base pair juxtaposition all along the helices. However, in all cases, one can identify two so-called "central" base pairs that stay close to the center of the contacting region and juxtapose in a manner resembling that seen in Figure 1A. Of the four strands forming the two double helices, two stay closer to the center (internal strands), whereas the other two are at the periphery of the structure (external strands). The arrangement is characterized by an axial symmetry shown in Figure 1A.

2.4 Role of the GU base pair

In view of this symmetry, it was therefore surprising to see that in the region of contact, the two helices are in fact asymmetric. Thus, we noticed among the identified cases a strong tendency to have one central base pair GU, which was observed in 10 out of 12 cases (Fig. 2A), whereas the other central base pair was always of the Watson–Crick (WC) type. Moreover, in all GU base pairs, G occupied exclusively the external position. This nonrandomness suggests that GU serves a unique role in the along-groove packing. Analysis of the juxtaposition of the central base pairs in the identified cases suggests an explanation for this asymmetry. As one can see in Figure 3, the close packing of the central base pairs GU and GC (the first letter in each base pair

corresponds to the external position) is stabilized by a unique interhelix network of five H-bonds, which is a variation of the well-known ribose zipper motif (Schindelin et al., 1995; Shah and Brunger, 1999; Doherty et al., 2001; Nissen et al., 2001). The existence of this network reflects a perfect complementarity between the two interacting surfaces, which includes the correspondence of the shapes as well as of the donors and acceptors of H-bonds. Analysis shows that only the asymmetric combination of GU versus WC would allow this packing. Any other combination including two WC base pairs would create a crack between the helices clearly seen in the two exceptional cases, SC549C501 and LC2291C2374. Such a crack destabilizes the packing by eliminating several interhelix atom–atom contacts in the middle of the contact region, which, in some cases, may become critical. As to the WC base pair, it would prefer GC or CG because G in both cases can make a direct H-bond with the nearby O3′ atom, which in AU and UA has to be replaced by a water bridge. On average, GU versus WC as two central base pairs is observed in 85% of the corresponding regions of rRNAs from other organisms (De Rijk and De Wachter, 1993; Wuyts et al., 2001; Wuyts et al., 2002). In view of the asymmetry between the two helices, we can consider the GU-containing helices as receptors of WC helices able to bind them almost indiscriminately.

2.5 Involvement in the tRNA association with the ribosome

The analysis of the 70S ribosome (Yusupov et al., 2001) revealed two more cases of the motif. Their identification, however, was not as straightforward as in the previous cases due to the 5.5 Å resolution of the structure with only phosphorus positions given for all ribosomal RNAs. Nevertheless, the knowledge of the exact conformations of the motifs already identified allowed us to superimpose them with the conformations of the candidates and thus select the structures in which the chains were arranged closely enough to those in the known cases to guarantee the same type of helix–helix arrangement. Also, the knowledge of the sequence requirements for the formation of the motif, including the position and orientation of the central GU base pairs, served as an additional criterion for selection. The two new motifs dealt with intermolecular complexes formed between elements of helices 69 and 68 of 23S rRNA and the tRNA molecules bound to the P- and E-sites, respectively (Figs. 2B and 4). In the tRNAs, the

central base pairs were, respectively, 12–23 in the D stem and 2–71 in the acceptor stem, which allowed us to refer to the corresponding tRNA-binding elements as the D and AC receptors.

Although tRNAs at any ribosomal site make different contacts with the ribosome (Yusupov et al., 2001), their interactions with the D and AC receptors are overwhelmingly more stable than all other tRNA–ribosome interactions outside the peptidyl transferase center. This can be deduced from comparison of the potential number of van der Waals contacts and H-bonds as well as of the estimated sizes of the contact areas. Analysis shows that these interactions occur in all types of organisms. First, in both cases, the GU-containing helix is a part of rRNA. In the absence of conservative GU base pairs in tRNA, only GU from rRNA can guarantee that most tRNAs fit these interactions. Both GU base pairs are very conservative: We have found only one (D receptor) and no (AC receptor) exception in the 595 available nucleotide sequences of the large subunit rRNA from all three major branches of evolution (De Rijk and De Wachter, 1993; Wuyts et al., 2001). Finally, the WC character of base pairs 12–23 and 2–71 is observed in 98% and 97% of cytosolic tRNAs, respectively (Sprinzl et al., 1998).

2.6 Possible role in translocation

Because the D and AC receptors bind tRNAs, they have to change partners each elongation cycle and, therefore, must be somehow involved in the ribosomal translocation. Analysis of the particular positions of both receptors within the ribosome in view of the available experimental data on occupancy of the tRNA-binding sites and structural changes in the ribosome during the elongation cycle suggests for each of them a very distinct and active role in this process.

There have already been indications of the involvement of the AC receptor in translocation. Indeed, modifications of the 2′-hydroxyl group of ribose 71 that block the tRNA interaction with this receptor severely affect the translocation (Feinberg and Joseph, 2001). The fact that the AC receptor binds to the very end of the E-tRNA acceptor stem at the side farthest from the P-site (Fig. 5) allows this complex to form even when the tRNA is in the P/E hybrid state. This state was postulated by Moazed and

Noller (1989a, b) as one of two intermediate states (together with A/P) during translocation (Rodnina et al., 2000). In fact, the end of the acceptor stem together with the four 3′-terminal nucleotides seems to be the only part of the deacylated tRNA that can touch the E-site in the P/E hybrid state. In view of the expected high stability of the complex between the acceptor stem and the AC receptor, we can identify the latter with the hypothetical E-based tRNA-binding site for the deacylated tRNA that has long been thought to provide the thermodynamic driving force for the first spontaneous step of translocation (Bergemann and Nierhaus, 1983; Spirin, 1985; Noller et al., 2000; Rodnina et al., 2000; Noller et al., 2002).

As to the D receptor, several observations can give clues to its functional role. First, the position of helix 69 and, therefore, of the D receptor is rather flexible, which was acknowledged previously (Yusupov et al., 2001). This would allow a tRNA to assume the P/E hybrid state without dissociation from the D receptor (Fig. 5). Second, we found that the D stem of a tRNA in either the A/P or P/E hybrid position is notably closer to the D receptor than in the pure A and E positions. This would make easier the exchange of the D receptor between the two tRNAs at some moment while they are in the hybrid positions. In other words, the hybrid state of the tRNAs, due to a particular arrangement of the two tRNA-binding receptors on the ribosome, facilitates both the formation of the complex of the AC receptor with the P/E tRNA and the redirection of the D receptor interaction from the P/E to A/P tRNA. This will force the translocation to take a pathway *via* the [A/P–P/E] intermediate. It will also make the D receptor bound to a tRNA for most of the time, and not only when the tRNA is in the pure P-site. The latter aspect becomes especially important in view of the fact that loop 1915, which closes helix 69, forms bridge B2a with helix 44 in 30S subunit (Moazed and Noller, 1989a; Mitchell et al., 1992; Yusupov et al., 2001). The simultaneous binding of helix 69 and its closing loop to tRNAs and to helix 44 of the 30S subunit, respectively, mechanically couples the position of tRNAs to a particular arrangement of helix 44 within the whole ribosome. We suggest that at the first spontaneous step of translocation, the changes in the tRNA position induce rearrangements in the 30S subunit. At further steps, the same communication line can be used to transmit a signal from this subunit to the tRNAs to accomplish the translocation (VanLoock et al., 2000). When this communication line is

affected by mutations in loop 1915, the tRNAs are no longer able to move properly, which results in frameshift (O'Connor and Dahlberg, 1995). The detailed mechanism of this communication is, however, a matter of future analysis.

2.7 Concluding remarks

This is only an example of how the along-groove packing of double helices can affect the ribosome function. The other cases of this motif can also play important structural or functional roles, which, however, are still to be discovered.

2.8 Acknowledgments

We thank Drs. L. Brakier-Gingras, S.W. Michnick, and H.F. Noller for stimulating discussions and critical reading of the manuscript. S.V.S. acknowledges grants from Canadian Institutes of Health Research and Human Frontiers Science Program and a fellowship from Fonds de la Recherche en Santé du Québec.

2.9 References

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**:905-920.

Bergemann K, Nierhaus KH. 1983. Spontaneous, elongation factor G independent translocation of *Escherichia coli* ribosomes. *J. Biol. Chem.* **258**:15105-15113.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**:1678-1685.

De Rijk P, De Wachter R. 1993. DCSE, an interactive tool for sequence alignment and secondary structure research. *Comput. Appl. Biosci.* **9**:735-740.

Doherty EA, Batey RT, Masquida B, Doudna JA. 2001. A universal mode of helix packing in RNA. *Nat. Struct. Biol.* **8**:339-343.

Feinberg JS, Joseph S. 2001. Identification of molecular interactions between P-site tRNA and the ribosome essential for translocation. *Proc. Natl. Acad. Sci. USA* **98**:11120-11125.

Harms J, Schluenzen F, Zarivach R, Bashan A, Gat S, Agmon I, Bartels H, Franceschi F, Yonath A. 2001. High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107**:679-688.

Klein DJ, Schmeing TM, Moore PB, Steitz TA. 2001. The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**:4214-4221.

Mitchell P, Osswald M, Brimacombe R. 1992. Identification of intermolecular RNA cross-links at the subunit interface of the *Escherichia coli* ribosome. *Biochemistry* **31**:3004-3011.

Moazed D, Noller HF. 1989a. Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell* **57**:585-597.

Moazed D, Noller HF. 1989b. Intermediate states in the movement of transfer RNA in the ribosome. *Nature* **342**:142-148.

Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. 2001. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. USA* **98**:4899-4903.

Noller HF, Cate J, Dallas A, Culver G, Earnest TN, Green R, Holmcomb L, Joseph S, Lancaster L, Lieberman K, Merryman C, Newcomb L, Samaha R, Von Ahsen U, Yusupov M, Yusupova G, Wilson K. 2000. Studies on the structure and function of ribosomes by combined use of chemical probing and X-ray crystallography. In: Garrett RA, Douthwaite, S. R., Liljas, A., Matheson, A. T., Moore, P. B., Noller, H. F., ed. *The ribosome: Structure, function, antibiotics, and cellular interactions.* Washington, DC: American Society for Microbiology Press. pp 129-150.

Noller HF, Yusupov MM, Yusupova GZ, Baucom A, Cate JH. 2002. Translocation of tRNA during protein synthesis. *FEBS Lett.* **514**:11-16.

O'Connor M, Dahlberg AE. 1995. The involvement of two distinct regions of 23 S ribosomal RNA in tRNA selection. *J. Mol. Biol.* **254**:838-847.

Pley HW, Flaherty KM, McKay DB. 1994. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**:111-113.

Rodnina MV, Pape T, Savelsbergh A, Mohr D, Matassova NB, Wintermeyer W. 2000. Mechanisms of partial reactions of the elongation cycle catalyzed by elongation factors Tu and G. In: Garrett RA, Douthwaite, S. R., Liljas, A., Matheson, A. T., Moore, P. B., Noller, H. F., ed. *The ribosome: Structure, function, antibiotics, and cellular interactions.* Washington, DC: American Society for Microbiology Press. pp 301-317.

Schindelin H, Zhang M, Bald R, Furste JP, Erdmann VA, Heinemann U. 1995. Crystal structure of an RNA dodecamer containing the *Escherichia coli* Shine-Dalgarno sequence. *J. Mol. Biol.* **249**:595-603.

Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A. 2000. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**:615-623.

Shah SA, Brunger AT. 1999. The 1.8 Å crystal structure of a statically disordered 17 base-pair RNA duplex: principles of RNA crystal packing and its effect on nucleic acid structure. *J. Mol. Biol.* **285**:1577-1588.

Spirin AS. 1985. Ribosomal translocation: facts and models. *Prog. Nucleic Acid Res. Mol. Biol.* **32**:75-114.

Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S. 1998. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26**:148-153.

Strobel SA, Cech TR. 1995. Minor groove recognition of the conserved G•U pair at the Tetrahymena ribozyme reaction site. *Science* **267**:675-679.

Strobel SA, Ortoleva-Donnelly L, Ryder SP, Cate JH, Moncoeur E. 1998. Complementary sets of noncanonical base pairs mediate RNA helix packing in the group I intron active site. *Nat. Struct. Biol.* **5**:60-66.

VanLoock MS, Agrawal RK, Gabashvili IS, Qi L, Frank J, Harvey SC. 2000. Movement of the decoding region of the 16S ribosomal RNA accompanies tRNA translocation. *J. Mol. Biol.* **304**:507-515.

Wimberly BT, Brodersen DE, Clemons WM, Jr., Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. 2000. Structure of the 30S ribosomal subunit. *Nature* **407**:327-339.

Wuyts J, De Rijk P, Van de Peer Y, Winkelmans T, De Wachter R. 2001. The European Large Subunit Ribosomal RNA Database. *Nucleic Acids Res.* **29**:175-177.

Wuyts J, Van de Peer Y, Winkelmans T, De Wachter R. 2002. The European database on small subunit ribosomal RNA. *Nucleic Acids Res.* **30**:183-185.

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF. 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**:883-896.

2.10 Figures

Figure 1

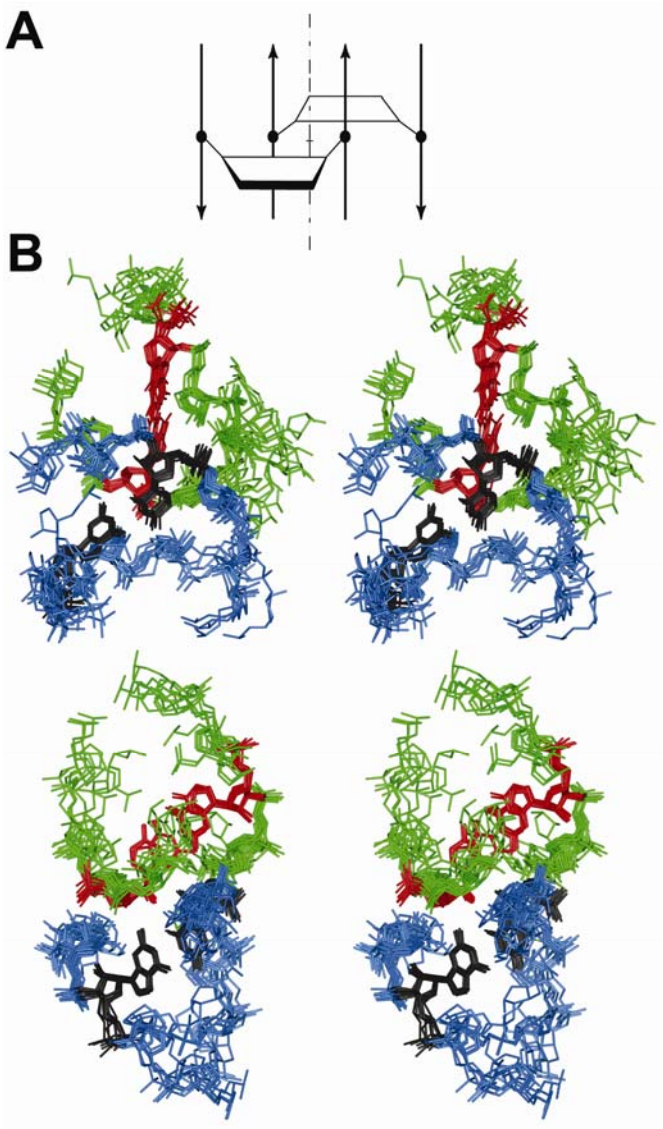


Figure 1. Along-groove packing of double helices.

A: Schematic representation. Trapezoids represent base pairs opened toward the minor grooves. Arrows represent backbones directed 5′→3′. The internal strand of each helix interacts with the minor groove of the other helix. Rotation of a helix 180° around the symmetry axis (dash-dotted line) superimposes it with the other helix.

B: Two different orientations of the superimposition of 10 motifs having a GU central base pair. For clarity, only backbones and central base pairs are shown. The GU and WC central base pairs are black and red, respectively. Their helices are blue and green.

Figure 2

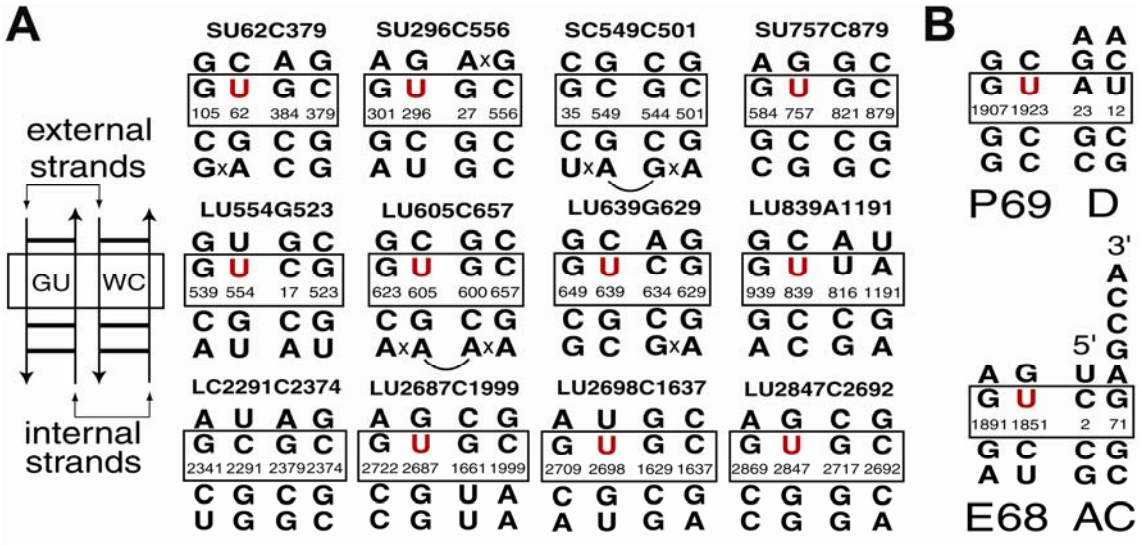


Figure 2. Nucleotide sequences of the along-groove packing motifs identified within ribosomal subunits (**A**) and between 23S rRNA and tRNA (**B**). The positions and orientations of the GU- and WC-containing helices correspond to those in the schematic representation on the left. Central base pairs are boxed. U in a central internal position is red. × designates the absence of nucleotide–nucleotide interactions. The nucleotide numbering in rRNA and tRNA is taken from (Yusupov et al., 2001) and (Sprinzl et al., 1998), respectively.

A: The name of each motif starts with letter S or L, reflecting the small or large subunit in which it is found, followed by the identity and the number of the internal central nucleotides.

B: Helices 69 and 68 of 23S rRNA pack, respectively, with the D and acceptor (AC) stems of the tRNAs in the P- and E-sites.

Figure 3

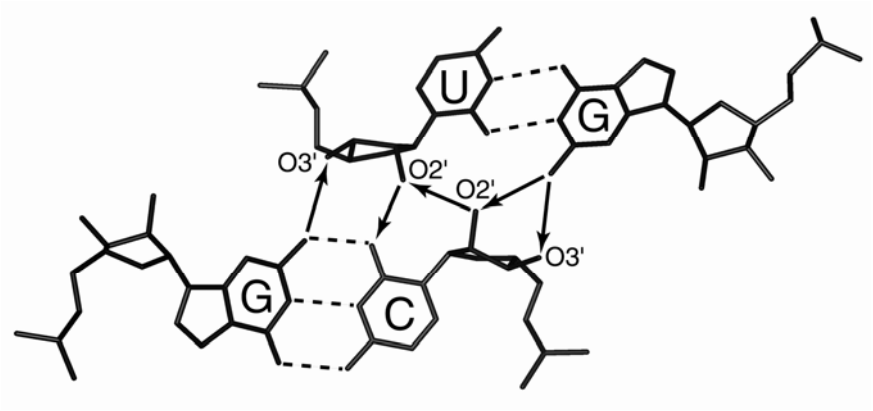


Figure 3. Juxtaposition of the central base pairs.

Arrows designate interhelix H-bonds directed from the donor to the acceptor atom. The presence of a central base pair GU in one helix makes the close packing possible and the arrangement asymmetrical.

Figure 4

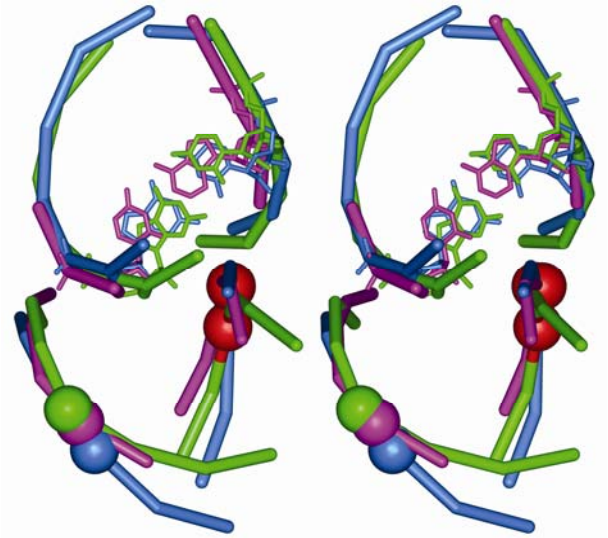


Figure 4. Superimposition of the rRNA-tRNA intermolecular along-groove packing motifs [D receptor–P-tRNA] (magenta) and [AC receptor–E-tRNA] (green) with motif LU554G523 (blue). In all structures, the WC central base pair is shown explicitly, and the rest of each complex is represented by lines connecting the consecutive phosphorus atoms. The phosphorus atoms of the nucleotides comprising all GU central base pair are shown as spheres. For uridines, these spheres are red. In both rRNA–tRNA complexes, the WC central base pair belongs to a tRNA, and the GU pair is a part of 23S rRNA. Even with the 5.5 Å resolution of the 70S ribosome conformation, the superimposition of the rRNA–tRNA complexes with LU554G523 is high enough to tell with certainty that they exemplify the close along-groove packing motif.

Figure 5

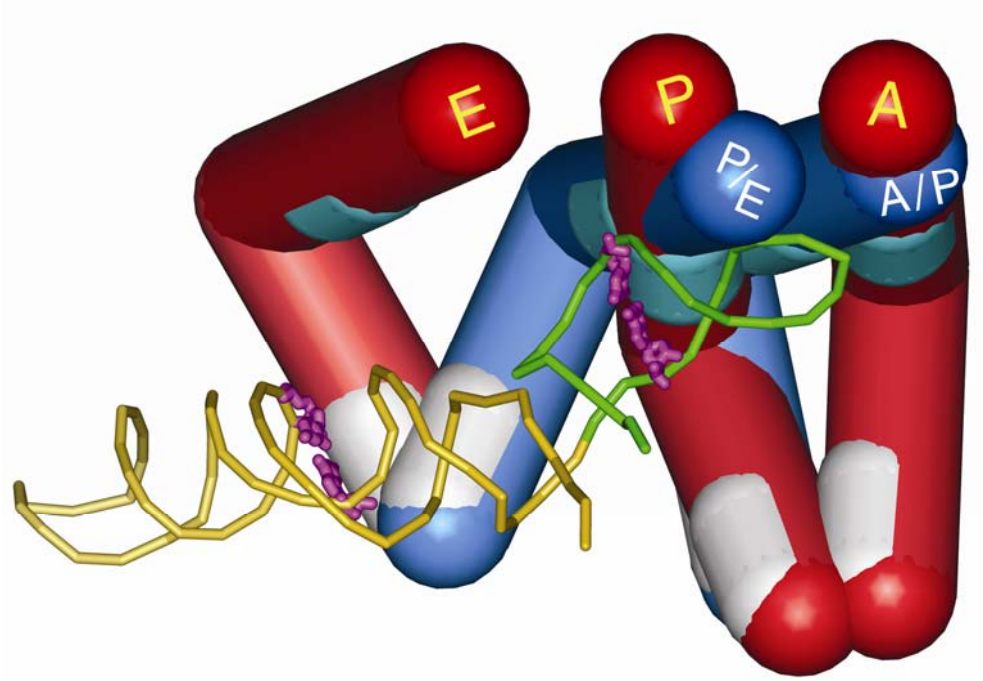


Figure 5. Positions of the binding sites in the acceptor (white patches) and D stems (light blue patches) of tRNAs in different pure (red) and hybrid (dark blue) states with respect to the D and AC receptors in helices 69 (green) and 68 (yellow) of 23S rRNA. tRNAs are shown as L-shapes with anticodons marked by the name of the state. The hybrid tRNAs are positioned according to (Moazed and Noller, 1989b). Other elements are positioned as in (Yusupov et al., 2001). The central GU base pairs of both receptors are purple. The D and AC receptors are close to their binding sites in the P/E-tRNA, and the D receptor is also close to its binding site in the A/P-tRNA. Loop 1915, which closed helix 69, interacts with helix 44 of the 30S subunit forming bridge B2a (not shown).

Chapitre 3

Article 2

Close packing of helices 3 and 12 of 16S rRNA is required for the normal ribosome function

3. Close packing of helices 3 and 12 of 16S rRNA is required for the normal ribosome function

Matthieu G. Gagnon[1], Alka Mukhopadhyay[1] and Sergey V. Steinberg

Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Journal of Biological Chemistry, 2006, vol. 281, no. 51, pp. 39349-39357

[1] Both authors contributed equally to this work

Running title: Along-groove packing motif in 16S rRNA

Contribution of each author:
Matthieu G. Gagnon: Developed the detailed experimental scheme, planned the design of the vector carrying the specialized ribosomal operon, participated in the construction of the specialized plasmid, incorporation of the 16S rRNA gene library, *in vivo* characterization of the functional clones, data analysis, preparation of manuscript and figures.
Alka Mukhopadhyay: Developed the detailed experimental scheme, planned the design of the vector carrying the specialized ribosomal operon, participated in the construction of the specialized plasmid, incorporation of the 16S rRNA gene library, *in vivo* characterization of the functional clones, data analysis, preparation of manuscript and figures.
Sergey V. Steinberg: Developed the general library design and participated in data analysis, preparation of manuscript and figures.

3.1 Abstract

The along-groove packing motif is a quasi-reciprocal arrangement of two RNA double helices in which a backbone of each helix is closely packed within the minor groove of the other helix. At the center of the inter-helix contact, a GU base pair in one helix packs against a Watson-Crick base pair in the other helix. Here, based on *in vivo* selection from a combinatorial gene library of 16S rRNA and on functional characterization of the selected clones, we demonstrate that the normal ribosome performance requires that helices 3 and 12 be closely packed. In some clones the Watson-Crick and GU base pairs exchange in their positions between the two helices, which affects neither the quality of the helix packing, nor the ribosome function. On the other hand, perturbations in the close packing usually lead to a substantial drop in the ribosome activity. The functionality of the clones containing such perturbations may depend on the presence of particular elements in the vicinity of the area of contact between helices 3 and 12. Such cases do not exist in natural 16S rRNA and their selection enriches our knowledge of the constraints imposed on the structure of ribosomal RNA in functional ribosomes.

3.2 Introduction

Ribosomes are RNA-protein complexes that perform protein synthesis in all living organisms. In eubacteria, the ribosome consists of three RNA chains and of more than 50 different proteins. Although the major steps of protein synthesis have been known for a long time, the detailed picture of how the ribosome structure forms and how it functions is still to be drawn. The recent achievements in X-ray crystallography of individual ribosomal subunits (1-4) and of the whole ribosome (5, 6) have opened a new era in the study of the mechanisms of protein synthesis. Now, structure-function relationships can be analyzed for any region of the ribosome taken within its natural context. Of special interest are those elements of the ribosome structure that have very similar, almost identical conformations and thus constitute recurrent motifs. Due to the similarity, recurrent motifs could play similar roles in the formation of the ribosome structure and, potentially, in the ribosome function.

Recently, we described a new, so-called along-groove packing motif (AGPM) found in a dozen places in the RNA of both ribosomal subunits (7). AGPM consists of two double helices closely packed *via* their minor grooves in the way that the sugar-phosphate backbone of one chain of each helix goes along the minor groove of the other helix (Fig. 1). In each helix, the chain that is packed in the minor groove of the opposite helix is positioned closer to the center of the arrangement and is thus called internal, contrary to the other chain that is called external. The interaction between the two base pairs at the center of the inter-helix contact zone is responsible for about one-half of the inter-helix atom-atom contacts, including a framework of five H-bonds (Fig. 2A and B). Henceforth, we call these base pairs central. The whole arrangement is characterized by a quasi-reciprocity, which is limited by the fact that the close packing of the two helices requires that one of the central base pairs be GU, while the other one be Watson-Crick (WC). The GU base pair should be oriented in the way that G and U belong to the external and internal strand, respectively (henceforth, the first nucleotide in each base pair corresponds to the external position). A violation of this requirement would result in a loss of the close packing of the helices and may ruin the whole arrangement (Fig. 2C).

An essential feature of AGPM is that it is able to bring together elements distant from each other in the secondary structure. This ability and the fact that AGPM has been found in many parts of the ribosome structure make this motif an essential element of the ribosome architecture. Therefore, the elucidation of the general requirements for the formation of AGPM is important for understanding how the whole ribosome structure forms and how it functions.

In this report, we explored a possibility of alternative nucleotide arrangements in AGPM SU296 (Fig. 3), which is formed between helices 3 and 12 of 16S rRNA. Within this motif, the central base pairs occupy positions G301-U296 and G27-C556**.** SU296 is positioned close to the root of the central pseudoknot at the region where three domains of the 30S subunit the shoulder, the platform and the neck come together. Each of two fragments 28-295 and 302-555 of 16S rRNA starts at SU296 and after making a large loop comes back to SU296 as a part of the other helix. Therefore, the formation of SU296 would strongly affect the set of possible conformations for each of the two fragments. Given that regions 28-295 and 302-555 form together the body and the

shoulder of 30S, SU296 is expected to play a major role in the formation of the tertiary structure of these two domains of the small ribosomal subunit. Thus, the closeness to several major domains and a probable involvement in the formation of their structures would make SU296 crucial for the integrity of the whole 30S subunit.

Unlike most AGPMs, SU296 is not immediately involved in any interaction either with another part of rRNA or with a ribosomal protein (3, 6), which makes it a perfect candidate for studying the properties of AGPM within its natural structural context without a necessity of taking into consideration a potential role of neighboring regions. Using *in vivo* instant evolution approach applied to a combinatorial gene library of 16S rRNA, we selected fourteen functional variants in which the identities of the nucleotides composing the two central base pairs of motif SU296 were different from those found in the wild-type (wt) *Escherichia coli* 16S rRNA. Only in a few variants the efficiency was close to wt. Analysis of the nucleotide sequences of these variants helped us to identify the limits within which the wt sequence can be modified in order to preserve the ribosome function. In two variants, the GU and WC base pairs stayed as in wt, which made the close helix packing undisturbed. In two more variants, GU was replaced by WC, while WC was replaced by GU. Individually, each of two replacements GU→WC and WC→GU would result in a loss of the close packing between the helices. However, when both replacements occur simultaneously, the quality of the helix packing is not affected. In other clones, the nucleotide sequences do not allow the close packing of the two helices; correspondingly, the efficiency of such ribosomes is severely damaged. Based on these results, we can conclude that the close packing of the helices within SU296 is a major aspect of the ribosome functionality.

3.3 Experimental procedures

Bacterial Strains and Media - For cloning and selection, we used the *E. coli* strain DH5α. Cultures were grown in the LB medium (8) or in the LB medium with 100 μg/mL ampicillin (LB-Amp-100) and with 50 μg/mL kanamycin, when appropriate (Sigma-Aldrich).

Plasmids - For cloning of combinatorial 16S rRNA gene libraries and for selection of functional clones, plasmid pAMMG carrying the specialized ribosome

system was constructed. This plasmid is analogous to the ones described elsewhere (9-17). Briefly, it contains a copy of the *E. coli rrn*B operon (from pKK1192U) (18) under the transcriptional regulation of the *lacUV*5-inducible promoter. It also contains two genes with modified Shine-Dalgarno (SD) sequences, of a chloramphenicol acetyl-transferase (CAT) and of a green fluorescent protein (GFP), which were used, respectively, for selection and for quantifying the efficiency of the selected clones (supplemental Methods). To measure the level of miscoding of the selected clones, seven variants of plasmid pLuc were constructed (supplemental Methods).

Construction of the Combinatorial Gene Library and Selection of Functional Clones - To randomize the four nucleotides comprising the two central base pairs of the SU296 motif (Fig. 3C), we used an overlapping extension PCR procedure (19). In this way, the entire region comprising the SU296 motif (902 bp) was amplified by consecutive multistep PCR. The sequences of the primers used for different steps of PCR are given in supplemental Methods. Prior to selection, the transformants were grown for 1 h in the LB medium. The synthesis of the plasmid-encoded ribosomes was induced by addition of isopropyl-1-thio-β-D-galactopyranoside (IPTG, Bioshop Canada Inc.) to a final concentration of 1 mM. After incubation for 3.5 h, the library was plated on the selection plates containing 200 μg/mL chloramphenicol and 1 mM IPTG. The minimal inhibitory concentration (MIC) of chloramphenicol for the cells expressing pAMMG-encoded rRNA was ~500 μg/mL. Out of a pool of 1.4×10^5 transformants, 50 survivors were obtained as chloramphenicol resistant and taken for further analysis.

GFPuv3 Assay - For quantifying the efficiency of the selected clones, the GFPuv3 variant (20) of the GFPuv gene (BD Biosciences) was incorporated into pAMMG under the control of a synthetic lactococcal strong constitutive CP25 promoter (21). Compared to GFPuv, the GFPuv3 variant provided for a higher sensitivity of measuring the fluorescence. Freshly transformed colonies of the selected clones were inoculated into 1 mL of the LB-Amp-100 medium containing 1 mM IPTG and incubated at 37°C with shaking for 16 h. After the incubation, the cells were harvested, washed twice in 1 mL of the HN buffer (20 mM Hepes, 0.85% NaCl, pH 7.4) and re-suspended in 1 mL of the HN buffer. The cell density (A_{600}) and the fluorescence (excitation = 497 nm, emission = 511 nm) of each clone was determined using a Packard Fusion α-FP

plate reader. For each clone, the fluorescence was divided by A_{600} and presented as a percentage of wt.

Firefly Luciferase Assay - Each of the seven miscoding variants of plasmid pLuc (supplemental Tables S3 and S4) was co-transformed together with pAMMG-wt or with a mutant derivative of it into DH5α cells and plated on the LB agar containing 100 μg/mL ampicillin and 50 μg/mL kanamycin. To induce the specialized ribosome system, we proceeded as described above. To measure the level of miscoding, the luciferase assay was performed as described in supplemental Methods**.**

Subunit Association Assay - Cells containing pAMMG or a mutant derivative of it were induced and grown in 500 mL of the LB-Amp-100 media. The ribosomal particles were isolated from the cellular lysates prepared by freezing and thawing method (22). The lysates were applied onto a 15-40% (w/v) linear sucrose gradient in 20 mM Tris–HCl (pH 7.8), 5 mM $MgCl_2$, 100 mM NH_4Cl and 6 mM β-mercaptoethanol, followed by centrifugation at 20,000 rpm for 20 h in a Beckman rotor SW28 at 4°C. Fractions containing the 30S subunits and the 70S ribosomes were collected using an ISCO gradient fractionator. RNA was extracted from the appropriate fractions with phenol. The proportion of the plasmid-encoded and host-encoded 16S rRNA was determined by the primer extension method (12, 23) (supplemental Methods).

Sequencing - Sequencing of the selected clones was performed on the LI-COR DNA sequencing system (Département de Biochimie, Université de Montréal) using primer 5'-AATTTATCAGACAATCTGTGTGGGCACTCG-3' for reading positions 27, 296 and 301 and primer 5'-GTAGGTGGGGTAACGGCTCACCTAGGCGAC-3' for reading position 556. These primers were labeled at the 5' end with IRDye-800 (LI-COR Biosciences).

Computer Modeling - The models were based on the crystal structure of the *E. coli* 30S subunit (6) (PDB code 2AVY). Fragments 26-30 and 553-558 of helix 3 and 293-304 of helix 12 were taken from the X-ray structure and were used as a starting conformation into which the nucleotide replacements corresponding to a particular selected clone were introduced. Modeling was done interactively, using the InsightII/Discover package (version 2000; Molecular Simulations Inc., San Diego, CA). Each model was submitted to a restrained energy minimization using the AMBER force

field (http://amber.ch.ic.ac.uk/amber). During the minimization, base pairs 30-553 and 26-558 of helix 3, base pair 293-304 of helix 12, as well as nucleotides 298 and 299 of the tetraloop capping helix 12 were immobilized. Visualization was done on a Silicon Graphics Fuel computer.

3.4 Results

3.4.1 Experimental System and Library Design

To study structure-function characteristics of SU296 within the 30S subunit, we used the instant evolution approach that consisted in selection of functionally active ribosomes from a combinatorial library in which several nucleotides of 16S rRNA were randomized. The selection was done *in vivo* with use of a specialized ribosome system pAMMG analogous to the ones previously described (9-17). The system contained genes of rRNA and of two reporter proteins, the chloramphenicol acetyl-transferase (CAT) and the green fluorescent protein (GFPuv3) (20). The rRNA gene was needed for incorporation and expression of the designed combinatorial library, while CAT and GFPuv3 were used, respectively, for selection of active clones in the presence of chloramphenicol and for measurement of the ribosome activity. In the 16S rRNA gene, the anti-Shine-Dalgarno (ASD) sequence was modified as suggested (14, 24), and the corresponding modification of the Shine-Dalgarno (SD) sequence was introduced into the genes of both reporter proteins. In the cell, this system coexists without interference with the normal cellular protein-synthesis machinery (14). In the combinatorial library, all four positions composing the two central base pairs (Figs. 1 and 2A) were fully randomized (Fig. 3), which provided for 256 possible variants.

3.4.2 Cloning and Selection of Functional Clones

After cloning the library, a pool of ~1.4 x 10^5 transformants was obtained. A random sequencing of 12 transformants showed that all four positions of the central base pairs in the SU296 motif were reasonably randomized, except that there was some bias towards cytosines and against adenines, accounting for 38% and 7% of the identities of randomized nucleotides, respectively (data not shown). Out of 50 clones that were selected as survivors in the presence of 200 μg/mL chloramphenicol, 25 had the wt

sequence, most probably, due to an incomplete digestion of plasmid pAMMG during the cloning of the library. The other 25 clones contained 14 unique sequences (Table I), which demonstrated a mild level of redundancy. In no case did mutations touch non-randomized nucleotides.

3.4.3 Characterization of the Selected Clones

For the selected clones, the GFPuv3 activity varied between 2 and 100% of that of wt (Table I, column GFPuv3). The selected clones were further characterized as to their capacity for the normal protein synthesis. First, for each clone, we determined the proportion of the plasmid-encoded 16S rRNA in the population of 30S subunits (column 30S). This proportion mostly depends on the relation between the rates of formation and degradation of the specialized 30S subunit and indirectly characterizes the stability of its structure. Second, we determined the proportion of the specialized ribosomes within the total 70S ribosome population (column 70S). This proportion depends on several factors, including some not related to the mutated ribosomes, like the availability of mRNAs carrying the specialized Shine-Dalgarno sequence. However, if the general conditions are maintained and if the number in column 30S is comparable to that of wt, a low number in column 70S will suggest a damaged initiation of translation.

Next, for each clone, the accuracy control was performed. Using previously reported techniques (12, 25, 26), we measured the efficiency of seven different miscoding events, of the *opal, ochre* and *amber* stop codon read-through, of the -1 and +1 frameshift, as well as of two types of amino acid misincorporation. Each time, the particular miscoding event was required for the expression of the reporting protein firefly luciferase from the specially constructed derivative of its gene, and the activity of this protein was taken as a measure of the efficiency of the miscoding. For each clone, the average and the highest level of the seven measured miscoding events are given in Table I (three last columns). The complete set of data on translational fidelity of the mutated ribosomes is presented in the supplemental Table I.

3.4.4 Potentially Negative Role of Uridine 1192

When this work was in progress, Rodriguez-Correa and Dahlberg (27) demonstrated that mutation C1192U, which is also present in the 16S rRNA encoded in pAMMG, can show synthetic lethality with some other mutations in 16S rRNA. To test a potentially damaging role of U1192 in the selected clones, we created six additional mutants through introduction of the reverse mutation U1192C into six clones with varying activities. The activities of all new mutants are presented in the supplemental Table V. They were within the same ranges of activities as of their U1192-containing analogs, being, on average, 7% higher. These results demonstrate that the observed phenotypes are essentially independent of the C1192U mutation. These results also show that mutation C1192U provides a mild negative effect on the ribosome function, which is independent of the packing between helices 3 and 12.

3.5 Analysis of the selected clones

Although the activity of all selected clones was sufficient for survival in the presence of chloramphenicol, the efficiency of the protein synthesis varied substantially among the clones. Based on the characteristics studied, we divided all clones in three groups seen in Table I. Group I consists of clone A1 with the wt sequence and of three more clones whose functional characteristics are comparable to those of wt. Group II consists of only one clone, A2, which in all aspects is as good as the Group I clones, except for a slightly elevated miscoding level. All other clones form Group III; in each clone of this group, some functional aspects are essentially damaged compared with those in Group I and II clones.

Analysis of the selected clones demonstrates that the ribosome efficiency strongly depends on the identity of the central base pairs of SU296. Thus, we have noticed that in all Group I and Group II clones, both central base pairs are either GU or WC. On the other hand, all clones in which at least one central base pair is different from GU and WC belong to Group III. One can say that the presence of any kind of base pair different from GU and WC results in a serious damage to the ribosome function. Also, in most Group III clones, the fraction of the specialized 30S subunits is substantially lower than in wt, which indicates that the formation of the 30S subunit is

compromised. This observation is in agreement with the discussed earlier suggestion that the integrity of SU296 is critical for the integrity of the structure of the whole 30S subunit.

3.5.1 Close Packing in Group I Clones

From Table I, one can also learn that in all Group I clones and only in them, one central base pair is GU, while the other one is WC. Therefore, the coexistence of a GU and WC central base pair in SU296 should be considered as a major factor of the normal ribosome functioning. Two clones A1 and A8 differ only by the identity of base pair 27-556, which in these clones is GC and CG, respectively. As one can see in Fig. 2 (A and B), replacement GC→CG preserves both the close helix packing and the inter-helix H-bonds. Interestingly, in the other two Group I clones, A5 and A7, GU has moved to the other helix, being replaced by either GC (A5) or CG (A7). Such an interchange of the central base pairs is not expected, however, to affect the quality of the helix packing. Indeed, due to the quasi-reciprocity of the arrangement, none of the two helices has a preference of harboring GU or WC. Therefore, if the close packing of the two helices exists in clones A1 and A8, it should be preserved in clones A5 and A7 as well.

To elucidate additional details related to the GU↔WC interchange, we superposed the juxtaposition of the central base pairs in clone A1 with the same juxtaposition rotated for 180° around the center of pseudo-symmetry of the arrangement (Fig. 4). This superposition would simulate the GU↔GC interchange that occurs in clone A5 compared to A1. The superposition shows that after such an interchange, none of the four nucleotides would remain in its position. The displacement of each nucleotide can be well approximated by a rotation around a particular point. The external purines rotate for ~12° around their C1' atoms, while the internal pyrimidines rotate for ~17° around their N1 atoms. The direction of the rotation is the same for nucleotides of the same base pair and opposite for nucleotides of different base pairs. In each case, the direction of the rotation is such that it would allow the bases to rearrange H-bonds within the newly formed base pair GC instead of GU and GU instead of GC.

Although the rotation of each nucleotide is driven by the changes of the hydrogen bond scheme within each base pair and is thus independent of the

rearrangements that occur in the other double helix, the simultaneous movements of all four nucleotides complement each other in the sense that together they allow the maintenance of the close packing between the two helices. In particular, these movements preserve the distance, and therefore, the H-bond between the O2' atoms of the internal nucleotides as well as between the N2 atom of G and the O3' atom of the opposite internal nucleotide (for the reference, see Fig. 2).

A different situation would occur if only one of the two replacements GU→WC or WC→GU takes place, while the other base pair remains unchanged. In this case, the nucleotide displacements in the mutated base pair will not be matched by the corresponding displacements in the other base pair, which will result in the loss of the close helix packing. For example, replacement GU→WC would displace the external nucleotide of this base pair for several angstroms farther from the opposite helix, thus creating a crack in the arrangement clearly seen in the known exceptional cases (Fig. 2C). From this point of view, the GU↔WC interchange constitutes a structural compensation in which a negative effect of one base pair replacement is compensated by the other base pair replacement. Because GU and WC base pairs coexist in all Group I clones, and only in them, we can conclude that the close packing of helices 3 and 12 is critical for the normal ribosome function.

As the GU↔WC interchange does not affect the packing of the helices, it has a chance to occur in evolution and thus could be detected through comparison of available nucleotide sequences of rRNA. Indeed, as recently pointed out by Mokdad and co-workers (28), the GU↔WC co-variation of the central base pairs exists in some AGPMs of both ribosomal subunits. Interestingly, SU296 is not included in these AGPMs, and, on the contrary, demonstrates a strong preference of GU for helix 12 and of WC base pair for helix 3 (Table II and supplemental Table II). In fact, the latter observation was one of the reasons that determined our choice of motif SU296 for experimental studies (M.G.G., A.M., Y. Butorin, J. Chen & S.V.S., A new type of structural compensation in ribosomal RNA, in RNA 2005, Tenth Annual Meeting of the RNA Society, Banff, Alberta, Canada, May 24-29, 2005, p. 380). The results presented here show that a possibility for the GU↔WC interchange is a general property of AGPM, which goes beyond the natural variability of rRNA.

On a more detailed level, however, one can see differences in the performance of the Group I clones before and after the GU↔WC interchange. In particular, comparison of clone A5 to A1 and A7 to A8 shows that a displacement of the GU base pair from helix 12 to helix 3 results in a drop of the overall activity and of the 70S level by ~15%. Although these changes are not dramatic, they could be sufficient for driving the evolution towards the wt sequence. The differences in the performance of the Group I clones demonstrate that the GU↔WC interchange is not, in fact, completely adequate, even though it preserves the close helix packing. A probable explanation of this phenomenon deals with the discussed earlier fact that the GU↔WC interchange forces a displacement of all four nucleotides of the central base pairs. This displacement could affect the positions of the neighboring nucleotides and, indirectly, the conformations of more distant regions of the 30S subunit. A potentially negative effect of such conformational changes could explain why the GU↔GC interchange does not occur naturally in SU296.

3.5.2 Preferences for the WC Base Pair

Further analysis shows that the identity of the WC base pair is also important for the ribosome function. Thus, one can see that in none of the Group I clones the WC base pair is either AU or UA. Although at least partly this fact can be explained by the mentioned above bias of the original library against A, it also fits well to what is known about AGPM in general. Indeed, the GC/CG base pairs have an obvious advantage compared to AU/UA because of the H-bond between the amino group of G and O3' of U from the opposite GU base pair (Fig. 2A and B), which does not exist in the AU/UA case. This additional H-bond provides a stabilizing effect on the AGPM, which could be critical for the normal ribosome function. The preference towards GC/CG observed in our selection goes along with the fact that in the wt rRNA sequences, GC/CG as central base pairs of AGPM occur overwhelmingly more often than AU/UA. For most AGPMs, the content of AU/UA is lower than 10% and only in one case exceeds 20% (29) (supplemental Table II).

Another observation deals with the preference of GC over CG. In particular, comparison of clone A8 to A1 and A7 to A5 shows that the replacement of GC by CG in

any of the two helices results in a drop of the overall activity by ~20%. As one can see in Fig. 2B, nucleotide C in the external position, unlike G, does not form any interaction with the opposite base pair. The absence of such an interaction will make the arrangement less cooperative and consequently, less stable, which can explain the preference of GC *versus* CG. It is important that this preference is not specific to SU296 and is also observed in other cases of AGPM. For example, in eubacteria, GC as a central base pair is, on average, about three times as frequent as CG (28, 29) (supplemental Table II). In the case of SU296, the local context can also contribute to the GC preference over CG in helix 3. Indeed, the presence of C in position 27 would allow an alternative base pairing of this nucleotide with the unpaired G in position 557 (Fig. 3C). We may speculate that a potential formation of such an alternative base pair in clone A8 results in the drop of the 30S level by more than 40% compared to A1 (Table I).

3.5.3 A Suboptimal Close Packing Arrangement for Clone A2

Clone A2, the only one composing Group II, represents a special case. In this clone, both central base pairs are GC, which, according to our findings, does not allow the close packing of the two helices. While all other clones without a close helix packing are functionally damaged, A2 functions almost as well as the Group I clones. This ability of A2 is not shared by other clones with two WC central base pairs. Indeed, a replacement of a central GC base pair in A2 by either CG or UA, which happens in clones A12 and A10, respectively, results in a substantial drop of the ribosome activity. For A12, which contains CG compared to GC in A2, the drop of activity is much stronger than in the previously discussed pairs of clones A1-A8 and A5-A7, experiencing the same kind of base pair replacement. Interestingly, the GC-GC combination can occur in different AGPM naturally, without any alarming consequence for the survival of the organism. In particular, for SU296 in eubacteria, combination GC-GC is one of two most popular alternatives to GU-GC (29) (Table II). Based on our results and on the mentioned above phylogenetic data, one can consider the GC-GC combination as a possible exception of the GU-WC rule able to make the ribosome functional.

Further analysis revealed an unusual feature of the GC-GC combination that may relate to its ability to maintain the ribosome function. We found that if the geometry of H-bonds within the two central base pairs is allowed to be somewhat deformed, a closely packed suboptimal arrangement will still be possible (Fig. 5). Despite the deformed hydrogen bonds, this arrangement corresponds to an energy minimum and is thus characterized by some level of stability. Essentially, this arrangement can be built only if both internal positions are occupied by pyrimidines, *i.e.* for GC-GC and not for GC-CG, CG-GC or CG-CG. We can thus speculate that the ability of the GC-GC combination to form such a closely packed arrangement is responsible for the strong performance of clone A2. Because for the GC-GC combination, the suggested arrangement would compete with the arrangement seen in Fig. 2C, some level of conformational flexibility for AGPM containing this combination seems to be unavoidable. This flexibility could be responsible for the increased level of miscoding observed for clone A2 compared to Group I clones.

Interestingly, a similar close packing arrangement with deformed H-bonds can also be formed for any other combination of two RY base pairs, where R and Y stand for a purine and a pyrimidine, respectively. Fig. 5 demonstrates how such an arrangement is formed for two GC base pairs. Other RY dinucleotide combinations like GU, AU and AC could also fit to this arrangement, although the two latter combinations would make it weaker due to the absence of an amino group in the minor groove. In connection with this, it is worth mentioning that in the case of SU296, combinations RY-RY are accountable for most exceptions from the standard GU-WC pattern (Table II).

In addition to A2, there are two more selected clones A12 and A10 in which both central base pairs are WC. The efficiency of these clones, however, was notably lower than that of A2, which was the reason for us to put them in Group III. A poorer performance of these clones correlates with the fact that neither in A12 nor in A10 the central pairs fit to the RY-RY pattern, and, therefore, in none of these clones a close helix packing is reachable even as a suboptimal arrangement. The most probable non-close arrangement of two WC base pairs is seen in Fig. 2D. Compared to the close packing, in this arrangement, one inter-helix hydrogen bond will be lost, which can explain a lower activity of clones A12 and A10 relative to A2. Also, the presence of a

UA base pair instead of GC or CG would deprive the helix-helix contact of another inter-helix hydrogen bond, which could be critical for the function and would explain the low efficiency of A10 compared not only to A12, but also to most Group III clones. In all other Group III clones, at least one central base pair is different from GU and WC, which suggests stronger rearrangements in the area of contact between helices 3 and 12. Correspondingly, none of these clones functions properly. Moreover, these clones demonstrate a variety of problems related to all aspects of the ribosome function from the formation of the 30S subunit structure and the subunit association to low elongation efficiency and a compromised accuracy. Altogether, these results clearly show that a normal function of the ribosome requires helices 3 and 12 be closely packed accordingly to the standard GU-WC pattern, or, at least, to the RY-RY pattern.

3.6 Discussion

In this study, we used *in vivo* selection of functional clones of 16S rRNA for analysis of general constraints imposed on AGPM SU296 in working ribosomes. The selection was based on the ability of the mutated ribosomes to synthesize CAT that would allow the cells to survive in the presence of chloramphenicol. It is important, however, that the fact of survival *per se* has not been able to guarantee a high efficiency of protein synthesis. Indeed, only an extensive functional characterization of the selected clones, which included the measurement of the activity of protein synthesis, of the proportion of the specialized 30S subunits and of specialized 70S ribosomes within the whole population of, respectively, 30S subunits and 70S ribosomes, and, finally, a systematic control of the accuracy of the specialized ribosomes has allowed us to distinguish the clones with efficiency approaching wt from those clones in which some aspects of functionality were severely damaged. A systematic comparison of the nucleotide sequences of more and less effective clones helped us to reveal some important characteristics of AGPM SU296.

Most characteristics of SU296 elucidated in this study deal with close packing and stability of the inter-helix arrangement and as such are expected to be shared by all or at least most AGPMs. The close packing of the helices can be achieved if one of the central base pairs is GU, while the other one is WC. To make the arrangement

sufficiently stable, the WC base pair should be either GC or CG. The Group I clones satisfying both requirements demonstrate the activity comparable to wt. The GC base pair is expected to provide a more stable arrangement than CG, and, correspondingly, the clones containing combinations GU-GC and GC-GU demonstrate a higher level of activity than the clones containing GU-CG and CG-GU. On average, these four combinations occupy the central base pairs in different AGPMs more often than any other combination, thus reflecting the general importance of stable helix-helix packing for the ribosome structure and function (supplemental Table II). Although one successful clone A2 does not follow the GU-WC sequence pattern, even here, a close helix packing scheme can be suggested as a suboptimal arrangement. Even a replacement of a GC or CG base pair by UA, which deprives the helix-helix interaction of an important hydrogen bond, is critical for the ribosome function. The latter conclusion is made based on comparison of clone A10 with other clones and is independent of the fact that, as mentioned above, there was some bias in the randomized positions towards C and against A.

The importance of the close packing between helices 3 and 12 becomes even more evident if one compares clones of the first two groups with those of Group III. In no clone of the latter group, the standard packing of helices 3 and 12 is reachable; correspondingly, none of these clones works properly. It would be interesting to know which aspects of the ribosome functional cycle are affected the most if helices 3 and 12 are not closely packed. Based on the results presented here, we can say that the absence of close packing would affect both the integrity of the tertiary structure of the 30S subunit and its function. Indeed, as it is seen in Table I, for most Group III clones the proportion of the specialized 30S subunits in the population of 30S subunits is substantially lower than for wt. However, even when this proportion is comparable to that of wt, which happens in clones A12, A3 and A9, the level of 70S ribosomes is still severely affected. On top of this, in most Group III clones, the ribosomes are essentially less accurate than in wt. In some clones, the miscoding level for specific events exceeds the corresponding level for the wt ribosomes more that 20 times. Altogether, these data demonstrate that AGPM SU296 is positioned at the intersection of different important processes related to the ribosome function, and, depending on the nature of the

interaction between helices 3 and 12 within this motif, different aspects of the ribosome function can be affected differently. Additional studies are required to elucidate the particular mechanism of how imperfections in packing of helices 3 and 12 affect the 30S and 70S formation, as well as the efficiency and accuracy of translation.

A specific characteristic of SU296 deals with the fact that for this motif, unlike for several other AGPMs, the GU↔WC interchange has never been observed among natural rRNA. The conservation of the positions of the GU and WC base pairs demonstrates the asymmetry between the two helices, which has been confirmed by a somewhat lower activity of the clones having GU in helix 3 compared to those having this base pair in helix 12. The existence of such an asymmetry shows that the stability and the close helix packing are not the only aspects of SU296 that are important for the ribosome function. It also indicates the importance of SU296 relationships with its structural context, which can be affected by the move of GU from one helix to the other. The nature of these relationships will be a matter of further analysis.

3.7 Acknowledgments

We thank Dr. Harry F. Noller for providing plasmid pKK1192U and Dr. Léa Brakier-Gingras for critical reading of the manuscript. This work was supported by operating grant from the Canadian Institutes of Health Research. S.V.S. acknowledges fellowships from the Canadian Institutes of Health Research and from the Fonds de la Recherche en Santé du Québec. M.G.G. held scholarships from the Natural Sciences and Engineering Research Council of Canada and from the Fonds de la Recherche en Santé du Québec.

3.8 References

1. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science* **289**, 905-920
2. Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000) *Cell* **102**, 615-623

3. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) *Nature* **407**, 327-339
4. Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001) *Cell* **107**, 679-688
5. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) *Science* **292**, 883-896
6. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. (2005) *Science* **310**, 827-834
7. Gagnon, M. G., and Steinberg, S. V. (2002) *RNA* **8**, 873-877
8. Luria, S. E., and Burrous, J. W. (1957) *J. Bacteriol.* **74**, 461-476
9. Lee, K., Varma, S., SantaLucia, J., Jr., and Cunningham, P. R. (1997) *J. Mol. Biol.* **269**, 732-743
10. Morosyuk, S. V., Lee, K., SantaLucia, J., Jr., and Cunningham, P. R. (2000) *J. Mol. Biol.* **300**, 113-126
11. Lee, K., Holland-Staley, C. A., and Cunningham, P. R. (2001) *J. Nutr.* **131**, 2994S-3004S
12. Belanger, F., Gagnon, M. G., Steinberg, S. V., Cunningham, P. R., and Brakier-Gingras, L. (2004) *J. Mol. Biol.* **338**, 683-693
13. Morosyuk, S. V., SantaLucia, J., Jr., and Cunningham, P. R. (2001) *J. Mol. Biol.* **307**, 213-228
14. Rackham, O., and Chin, J. W. (2005) *Nat. Chem. Biol.* **1**, 159-166
15. Abdi, N. M., and Fredrick, K. (2005) *RNA* **11**, 1624-1632
16. Hui, A., and de Boer, H. A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4762-4766
17. Hui, A., Jhurani, P., and de Boer, H. A. (1987) *Methods Enzymol.* **153**, 432-452
18. Brosius, J., Ullrich, A., Raker, M. A., Gray, A., Dull, T. J., Gutell, R. R., and Noller, H. F. (1981) *Plasmid* **6**, 112-118
19. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene* **77**, 51-59
20. Ito, Y., Suzuki, M., and Husimi, Y. (1999) *Biochem. Biophys. Res. Commun.* **264**, 556-560

21. Jensen, P. R., and Hammer, K. (1998) *Appl. Environ. Microbiol.* **64**, 82-87
22. Ron, E. Z., Kohler, R. E., and Davis, B. D. (1966) *Science* **153**, 1119-1120
23. Sigmund, C. D., Ettayebi, M., Borden, A., and Morgan, E. A. (1988) *Methods Enzymol.* **164**, 673-690
24. Lee, K., Holland-Staley, C. A., and Cunningham, P. R. (1996) *RNA* **2**, 1270-1285
25. Dyer, B. W., Ferrer, F. A., Klinedinst, D. K., and Rodriguez, R. (2000) *Anal. Biochem.* **282**, 158-161
26. Viviani, V. R., Uchida, A., Viviani, W., and Ohmiya, Y. (2002) *Photochem. Photobiol.* **76**, 538-544
27. Rodriguez-Correa, D., and Dahlberg, A. E. (2004) *RNA* **10**, 28-33
28. Mokdad, A., Krasovska, M. V., Sponer, J., and Leontis, N. B. (2006) *Nucleic Acids Res.* **34**, 1326-1341
29. Wuyts, J., Perriere, G., and Van De Peer, Y. (2004) *Nucleic Acids Res.* **32**, D101-103
30. Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R. (2002) *BMC Bioinformatics* **3**, 2

ABBREVIATIONS

The abbreviations used are: AGPM, Along-Groove Packing Motif; WC, Watson-Crick; wt, wild-type; SD, Shine-Dalgarno; CAT, chloramphenicol acetyl-transferase; GFP, green fluorescent protein; PCR, polymerase chain reaction; IPTG, isopropyl-1-thio-β-D-galactopyranoside; MIC, minimal inhibitory concentration; ASD, anti-Shine-Dalgarno.

3.9 Tables

Table I. Nucleotide sequences and characteristics of the selected clones

Clone	*N* [a]	Randomized positions				GFPuv3 [b] (%)	Plasmid-encoded 16S rRNA [c] (%)		Ribosome miscoding [d]		
		301	296	27	556		30S	70S	Average	Highest	Miscoding event
Group I											
A1 (wt)	25	G	U	G	C	100	37±2	64±2	1.0	1.0	-
A8	3	G	U	C	G	81±5	21±2	61±3	1.0	1.3	Opal
A5	3	G	C	G	U	85±8	39±3	54±4	1.2	1.7	Ochre
A7	1	C	G	G	U	70±4	44±2	51±1	1.1	1.7	Amber
Group II											
A2	7	G	C	G	C	79±6	35±1	60±2	1.4	2.3	S→R
Group III											
A12	2	C	G	G	C	49±3	32±2	17±1	0.8	1.8	S→R
A3	1	C	G	A	G	26±3	26±1	10±1	5.8	28.3	G→R
A9	1	G	C	C	U	11±2	28±2	7±1	8.6	19.2	-1 fs
A11	1	U	G	C	A	22±3	10±2	29±3	1.6	2.7	G→R
A4	1	C	C	G	U	78±3	14±4	17±2	1.3	2.6	Amber
A6	1	C	C	G	C	10±2	11±1	11±2	3.0	3.9	Amber
A10	1	G	C	U	A	4±1	10±2	11±3	7.6	25.6	-1 fs
A13	1	C	C	U	C	2±1	8±3	4±1	6.8	21.8	-1 fs
A15	1	C	U	G	G	2±1	3±1	ND [e]	7.6	17.5	-1 fs
A14	1	G	U	G	G	2±1	2±1	ND	ND	ND	

a. *N* is the number of times each clone has been isolated.

b. The ribosome activity was calculated as the mean ± standard deviation of five to eight independent experiments.

c. The proportion of the plasmid-encoded 16S rRNA in the 30S and 70S fractions was calculated as the mean ± standard deviation of experiments performed with three independent RNA preparations. The determination of the level of specialized 30S subunits within the 70S ribosomes of 2% or less is not-detectable.

d. Effect of the mutations in AGPM SU296 on the level of ribosome miscoding. A value of 1.0 was arbitrarily ascribed to the wt 16S rRNA. For other clones, the level of miscoding was calculated as the mean ± standard deviation of three independent experiments. The extended set of data is present in supplemental Table I. Here, only the average level of efficiency of seven different miscoding events as well as the highest level of miscoding efficiency and the corresponding miscoding event are provided (see the text). Different miscoding events: *ochre*, *amber*, *opal*: suppression of the corresponding stop codon; -1 fs, +1 fs: the -1 and +1 frameshifts; S→R, G→R: insertion of arginine in response to a serine or glycine codon, respectively. For clone A14, due to a very low activity, none of the levels of miscoding was reliably reproducible.

e. ND, not detectable.

Table II. Statistical spectrum of the identities of the central base pairs in AGPM SU296

Eubacteria					Archaebacteria				
A. GU-WC type				Number	A. GU-WC type				Number
301	296	27	556	of cases	301	296	27	556	of cases
G	U	G	C	5968	G	U	G	C	286
G	U	A	U	3	G	U	C	G	61
					G	U	U	A	6
					G	U	A	U	6
		TOTAL		5971			TOTAL		359
B. RY-RY type					B. RY-RY type				
G	C	G	C	26	G	U	G	U	31
G	U	A	C	26	A	U	G	C	4
A	U	G	C	25					
G	U	G	U	4					
A	C	G	C	1					
		TOTAL		82			TOTAL		35
C. Others					C. Others				
G	U	U	C	18	G	U	U	G	1
G	U	C	C	14	A	U	C	G	1
U	U	G	C	3	G	U	C	U	1
C	U	G	C	2	G	U	A	G	1
G	U	G	A	1					
C	U	U	C	1					
A	U	C	C	1					
		TOTAL		40			TOTAL		4

The data were obtained from the available rRNA alignments (29). Only those cases were taken into consideration where the identities of all four nucleotides were known. For SU296 both in eubacteria and in archaebacteria, combinations of the RY-RY type dominate among non-GU-WC cases.

3.10 Figures

Figure 1

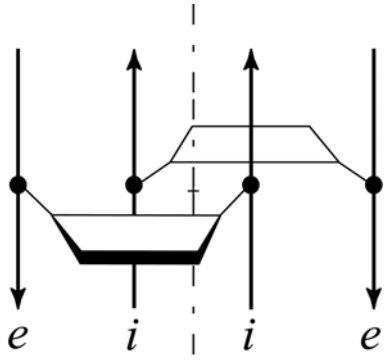


Figure 1. Schematic representation of the along-groove packing motif (AGPM).
Trapezoids represent base pairs opened toward the minor grooves. Arrows represent backbones directed 5'→3'. The internal and external strands of both helices are marked by italic letters *i* and *e*, respectively. The internal strand of each helix is packed within the minor groove of the other helix. The whole arrangement is characterized by a quasi-reciprocity, which is limited by the fact that for close helix packing, one of the helices should contain a GU base pair at the inter-helix contact zone, while the other helix should contain a Watson-Crick base pair (see text).

Figure 2

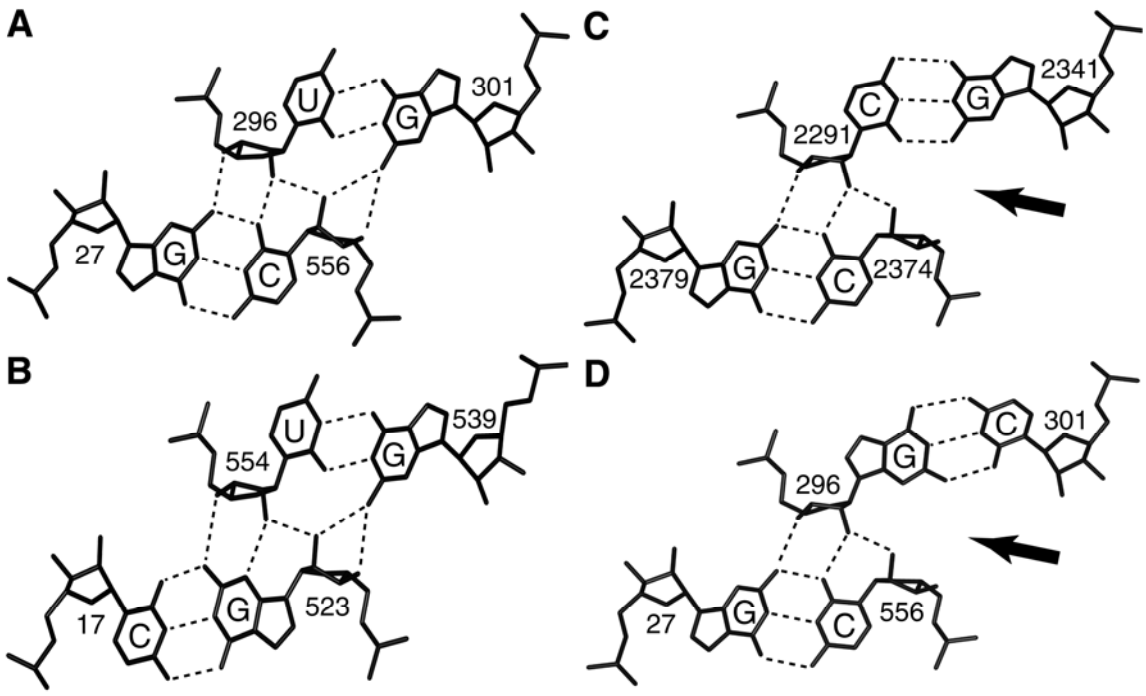


Figure 2. Juxtaposition of different central base pairs within AGPM.

Dashed lines stand for inter-nucleotide H-bonds. The coexistence of GU and a WC base pair in motifs SU296 (A) and LU554 (B) introduces an asymmetry between the helices that allows their close packing. The deviation from the GU-WC pattern in motif LC2291 results in a crack between the two helices marked by the arrow (C). The crack will remain if a central GC base pair is replaced by CG (D), which happens in clone A12. Motif SU296 was extracted from 16S rRNA (PDB code 2AVY) (6), while motifs LU554 and LC2291 were extracted from 23S rRNA (PDB code 1JJ2) (1, 7). For all juxtapositions, the *E. coli* nucleotide numbering is used.

Figure 3

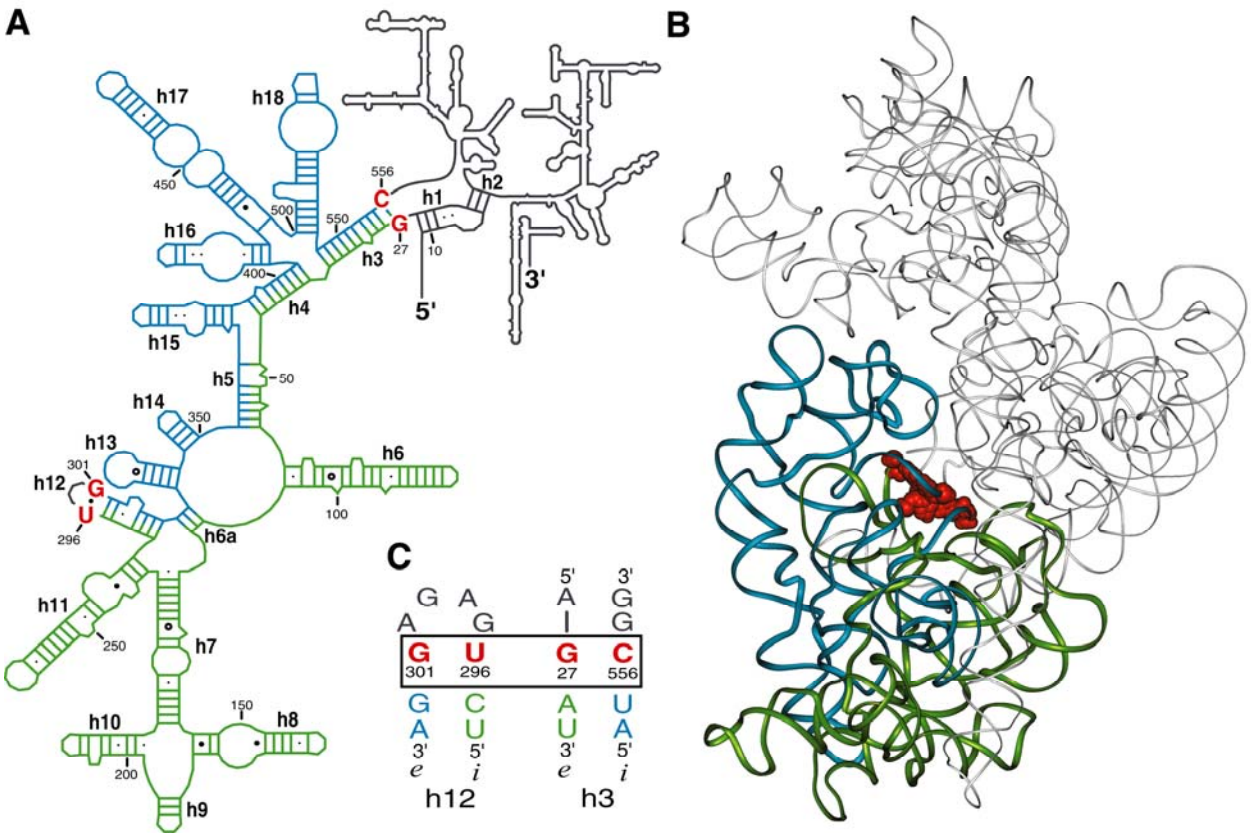


Figure 3. AGPM SU296 in *E. coli* 16S rRNA.

The central base pairs 301-296 and 27-556 are red. Fragments 28-295 (green) and 302-555 (blue) connect helices 3 and 12 (see text).

(A) Secondary structure of the *E. coli* 16S rRNA, modified with permission (30).

(B) Tertiary structure of the *E.coli* 16S rRNA. The nucleotides of the central base pairs are shown explicitly in a space-filling representation. Fragments 28-295 and 302-555 are shown with the colored ribbons. Together, these two fragments form the shoulder and most of the body of the 30S structure.

(C) Arrangement of helices 12 and 3 within AGPM SU296. Helix 12 is capped by a tetraloop, while helix 3 is formed by two strands distant from each other in the polynucleotide chain. In both helices, the central base pairs are boxed. The internal and external strands of both helices are marked by italic letters *i* and *e*.

Figure 4

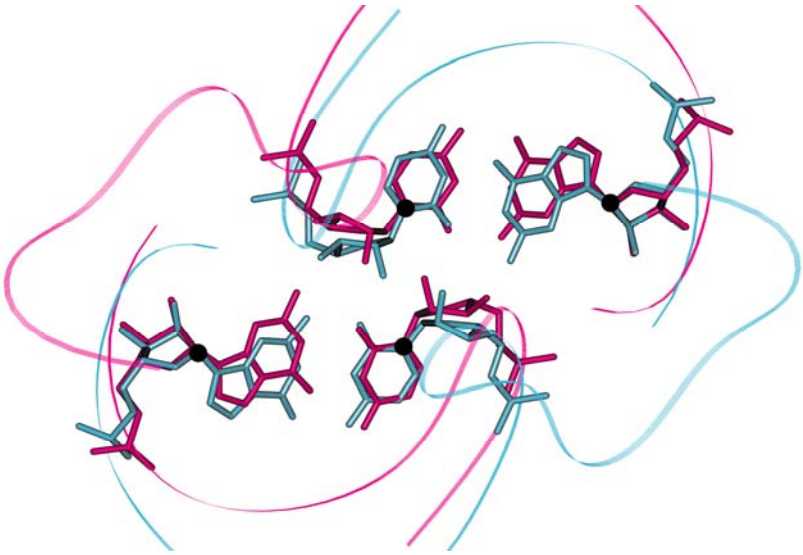


Figure 4. The juxtaposition of the central base pairs in clone A1 (blue) superposed with the same juxtaposition rotated for 180° around the center of pseudo-symmetry (magenta).

This superposition simulates the GU↔GC interchange observed in clone A5 with respect to A1. After such an interchange, none of the nucleotides composing the two central base pairs remains in its position. The external purines rotate for ~12° around their C1' atoms, while the internal pyrimidines rotate for ~17° around their N1 atoms. The centers of rotation are shown as black dots. For each nucleotide, the direction of rotation corresponds to the conversion of the nucleotide position in clone A1 to that in clone A5. The upper and lower nucleotides rotate, respectively, clock-wise and counter-clock-wise.

Figure 5

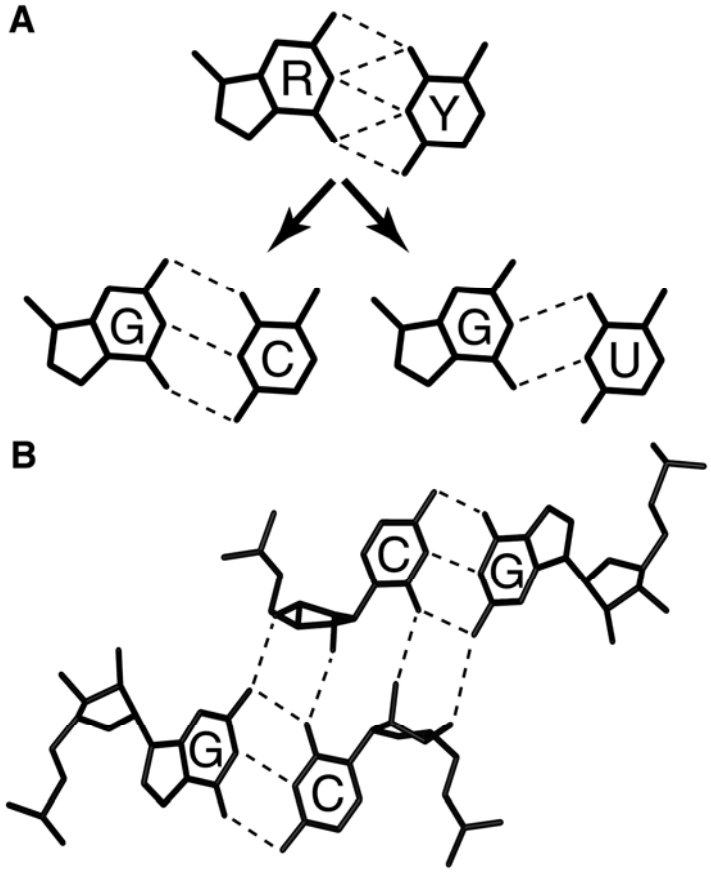


Figure 5. A model of close packing of two RY base pairs with a deformed geometry of hydrogen bonds. For such a packing, in each base pair, the pyrimidine should occupy a position intermediate between those in the GC and GU base pairs (A). Depending on the identity of this pyrimidine, different types of inter-base hydrogen bonds would be formed. Analysis shows that two such base pairs can be packed closely with formation of several inter-helix hydrogen bonds. (B) Example of such a close packing of two GC base pairs. Within this arrangement, GC base pairs are replaceable by GU base pairs and, potentially, by any other RY combination. In spite of some imperfections in geometry of the hydrogen bonds within each base pair, for GC-GC and GU-GU combinations, this arrangement corresponds to an energy minimum and is thus characterized by a certain level of stability.

3.11 Supplemental methods

Plasmid (pAMMG) - For expression of combinatorial libraries of rRNA, a specialized ribosome system was arranged. The plasmid carrying this system was named pAMMG. Because the ribosome binding site (RBS) of both CAT and GFP messengers has been changed to 5'-AUCCC (1, 2) while the messenger binding site (MBS) of the 16S rRNA has been changed to 5'-GGGAU, the reporter mRNAs in the cells harboring this plasmid are translated only by the plasmid-encoded ribosomes. The CAT gene is under the control of a constitutive *trp* promoter (3, 4) while the GFP gene is under the control of a synthetic lactococcal strong constitutive CP25 promoter (5). The CAT protein is used for selection of the functional mutants while the GFP is used to determine their functional activity by measuring the fluorescence of the cell culture.

Plasmids pLuc construction - The gene of the *Photinus* (Firefly) luciferase enzyme was taken from plasmid pGEM-*luc* (Promega) and cloned into pACYC177 (New England Biolabs) under the control of a synthetic lactococcal strong constitutive CP25 promoter (5). In this gene, RBS was changed to 5'-AUCCC. For cloning the miscoding cassettes, a *Sac*I restriction site was created at the beginning of the luciferase coding sequence, between the first and the second codon. This restriction site was used together with *Bam*HI for insertion of stop codon and frameshift cassettes, providing for constructs pLuc-WT, pLuc-UGA, pLuc-UAA, pLuc-UAG, pLuc-(-1) and pLuc-(+1). Arg-218 is known to be involved in the active center of the firefly luciferase (6). For the mis-incorporation constructs, pLuc-R218G (the arginine codon in position 218 is replaced by a glycine codon) and pLuc-R218S (here, the arginine codon is replaced by a serine codon) were made; both replacements would inactivate the enzyme (6, 7), so that any detected luciferase activity should be attributed to the mis-incorporation of arginine. The nucleotide sequences of the incorporated cassettes are presented in supplemental Tables III and IV. The sequences of these constructs were confirmed by sequencing of the luciferase gene and were used to measure the level of miscoding of the selected clones.

Combinatorial gene library primers and cloning - Oligonucleotide primers were synthesized with random nucleotides at the desired positions (Invitrogen Canada Inc.).

The first step involved PCR amplification of 4 individual products, the 229 bp, 291 bp, 275 bp and 168 bp, using rRNA gene as a template. For this, respective set of the four primers, *L1-A-Kpn I*: 5'-CACACAGGGGTACCAGAAAAAGCGAAGCGGCACTG-3' and *L1-B*: 5'-GTTAGGCCTGCCGCCAGCGTTCAATNTGAGCCATG-3'; *L1-C*: 5'-TGAACGCTGGCGGCAGGCCTAACACATGCAAGTCG-3' and *L1-D*: 5'-TCCAGTGTGGCTGGTCATCNTCTCNGACCAGCTAG-3'; *L1-E*: 5'-GATGACCAGCCACACTGGAACTGAGACACGGTCCAG-3' and *L1-F*: 5'-GCTTTACGCCCAGTAATTCCNATTAACGCTTGCAC-3'; and *L1-G*: 5'-GAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTG-3' and *L1-H-Bgl II*: 5'-CCACCGGTATTCCTCCAGATCTCTACGCATTTCAC-3', was used (bold N = A, G, T or C). At the next PCRs, these four products were joined together through annealing of the overlapping regions. The PCR products 275 bp and 168 bp were used to amplify a 424 bp PCR product using the *L1-E* and *L1-H-Bgl II* set of primers. Subsequently, the above 424 bp product and the 291 bp product were used as templates, using the *L1-C* and *L1-H-Bgl II* primers, that amplified a 696 bp product. Finally, using this 696 bp product and the 229 bp product as templates and using the flanking primers *L1-A-Kpn I* & *L1-H-Bgl II*, the entire 902 bp region of the SU296 motif was amplified and purified. This 902 bp PCR product harbored the library that contained the randomized nucleotides of the motif. This PCR product (902 bp) was cloned into the pAMMG plasmid using the unique *Kpn I* and *Bgl II* restriction sites. All PCR reagents, Vent DNA polymerase, restriction enzymes and T4 DNA ligase were from New England Biolabs.

Primer extension - Comparing to wt rRNA, the rRNA gene of pAMMG harbors mutation C to U at position 1192 of 16S rRNA (8). For detection of the wt 16S rRNA and of the plasmid-encoded 16S rRNA, we used a ^{32}P-labelled primer, 5'-CGTAAGGGCCATGATGACTTGA-3', which is complementary to region 1194-1215. The reaction was carried out using the M-MuLV reverse transcriptase enzyme (New England Biolabs) and ddGTP nucleotide (Amersham Biosciences). The resulting DNA fragments were resolved on a 20% (w/v) polyacrylamide denaturing gel. The relative band intensities from the directly exposed K-screen (Kodak) were scanned and quantified using Quantity One software version 4.6 (Bio-Rad) of MultiImager FX.

Firefly luciferase assay - The luciferase assay was performed as previously described (7, 9) with some modifications. Briefly, 45 μL of cultured cells were mixed with 5 μL of buffer K (1 M K_2HPO_4 (pH 7.8), 20 mM EDTA, 1 mM benzamidine), followed by one cycle of freeze-thaw. Fifty microliters of lysozyme solution (5 mg/mL of lysozyme in 0.1× buffer K) and 100 μL of lysis solution (50 mM Tris–phosphate (pH 7.8), 4 mM DTT, 4 mM EDTA, 20% (v/v) glycerol, 2% (v/v) Triton X-100) were then added to each sample, followed by incubation at room temperature for about ten minutes. Luciferase activities (luminescence) were measured with a Packard Fusion α-FP plate reader using 10 μL cell lysate added to 50 μL of a noncommercial luciferase assay reagent (9) containing 75 μM D-luciferin (BD Biosciences). The luciferase activities were normalized with the cell density of the cultures. The readthrough, frameshift and mis-incorporation level was calculated by dividing the luciferase activity of the stop codon, frameshift or mis-incorporation construct by that of the wild-type luciferase construct.

3.12 Supplemental tables

SUPPLEMENTAL TABLE I. Efficiency of different miscoding events for all selected clones

Clone	Level of miscoding events						
	UGA (Opal)	UAA (Ochre)	UAG (Amber)	-1 frameshift	+1 frameshift	G218R	S218R
Group I							
A1 (wt)	1.0	1.0	1.0	1.0	1.0	1.0	1.0
A8	1.3 ± 0.4	1.0 ± 0.1	0.7 ± 0.3	0.7 ± 0.1	1.1 ± 0.3	1.0 ± 0.2	1.0 ± 0.3
A5	1.3 ± 0.4	1.7 ± 0.1	1.1 ± 0.1	0.8 ± 0.2	1.6 ± 0.1	0.8 ± 0.1	1.2 ± 0.1
A7	0.9 ± 0.1	1.2 ± 0.1	1.7 ± 0.4	1.5 ± 0.1	1.4 ± 0.4	0.2 ± 0.1	1.0 ± 0.3
Group II							
A2	1.4 ± 0.5	1.7 ± 0.1	1.1 ± 0.4	1.3 ± 0.8	1.5 ± 0.1	0.7 ± 0.1	2.3 ± 0.5
Group III							
A12	0.8 ± 0.1	0.4 ± 0.2	0.5 ± 0.1	0.5 ± 0.2	0.7 ± 0.2	0.8 ± 0.2	1.8 ± 0.5
A3	0.6 ± 0.1	1.2 ± 0.4	2.6 ± 0.2	2.45 ± 0.03	2.6 ± 0.3	28.3 ± 4.3	2.8 ± 0.4
A9	ND	4.1 ± 1.9	7.2 ± 0.8	19.2 ± 0.9	7.2 ± 1.1	ND	5.1 ± 0.5
A11	0.19 ± 0.01	1.7 ± 0.3	1.3 ± 0.2	2.2 ± 0.5	2.2 ± 0.6	2.7 ± 0.8	0.94 ± 0.03
A4	1.1 ± 0.2	0.9 ± 0.1	2.6 ± 0.8	1.3 ± 0.2	1.1 ± 0.4	1.7 ± 0.3	0.7 ± 0.1
A6	2.5 ± 0.1	2.2 ± 0.5	3.9 ± 1.2	3.4 ± 0.2	3.3 ± 0.8	ND	2.8 ± 0.5
A10	3.4 ± 0.6	5.4 ± 2.1	1.8 ± 0.4	25.6 ± 0.4	9.1 ± 3.2	ND	0.5 ± 0.2
A13	2.1 ± 0.6	8.1 ± 2.5	0.32 ± 0.04	21.8 ± 5.1	5.3 ± 1.9	ND	3.1 ± 0.5
A15	ND	2.9 ± 0.9	0.3 ± 0.1	17.5 ± 4.4	9.5 ± 3.5	ND	ND
A14	ND	ND	ND	ND	ND	ND	ND

For each clone, the miscoding event corresponding to the highest efficiency is shown in red (Table I, column Highest). The levels of the readthrough of stop codons, of the -1 and +1 frameshifting and of amino acid mis-incorporations (G218R and S218R) were calculated as a ratio of the luciferase activity obtained with the specific constructs and with the wild-type luciferase construct. All miscoding levels were normalized by the corresponding level obtained for the wt 16S rRNA (experimental values (%) were, respectively, 0.053(±0.009), 0.66(±0.12) and 6.5(±2.5) for UGA, UAA and UAG stop codons readthrough, 0.24(±0.11) and 0.27(±0.07) for -1 and +1 frameshifting, 0.016(±0.006) and 0.18(±0.06) for G218R and S218R mis-incorporations). Results were calculated as the mean ± standard deviation of three independent experiments. The levels of the UGA readthrough and of the G→R mis-incorporation are not-detectable if the GFPuv3 activity is 10% or less. For clone A14, due to a very low activity, none of the levels of miscoding was reliably reproducible. ND: Not-detectable.

SUPPLEMENTAL TABLE II. Statistical spectrum of the central base pairs fitting the GU-WC pattern in the known AGPMs

AGPM (11)	Eubacteria Central base pairs				%	Archaebacteria Central base pairs				%
	e	*i*	*e*	*i*		*e*	*i*	*e*	*i*	
SU62	G	U	G	C	74	G	U	G	C	53
	G	U	C	G	25	G	U	C	G	37
	G	U	U	A	0.2	G	U	U	A	7
	G	C	G	U	0.1	C	G	G	U	0.4
	G	U	A	U	0.1					
SU296	G	U	G	C	98	G	U	G	C	72
	G	U	A	U	0.1	G	U	C	G	15
						G	U	U	A	2
						G	U	A	U	2
SC549	G	C	G	U	3	G	U	C	G	42
	G	U	G	C	0.2	G	U	U	A	17
						G	U	G	C	12
						G	U	A	U	6
SU757	G	U	G	C	99	G	U	G	C	95
	G	U	C	G	0.4	G	U	A	U	1
	G	U	A	U	0.1	A	U	G	U	0.2
LU554	G	U	G	C	97	G	U	C	G	100
	G	C	G	U	2					
LU605	C	G	G	U	62	G	C	G	U	43
	G	C	G	U	33	C	G	G	U	22
	U	A	G	U	4	G	U	G	C	11
						G	U	A	U	8
						U	A	G	U	5
LU639	G	U	G	C	62	G	U	C	G	97
	G	U	C	G	26	G	U	G	C	3
	G	U	U	A	9					
	G	C	G	U	2					
LU839	G	U	C	G	67	G	U	C	G	47
	G	U	U	A	16	G	U	U	A	11
	G	U	G	C	14					
LU1864 (12)	G	U	G	C	99	Replaced by GNRA (12)				
LC2291	G	U	G	C	96	G	U	G	C	62
	G	U	C	G	2	G	U	C	G	24
	G	U	U	A	1					
LU2687	G	U	G	C	82	G	U	G	C	62
	G	U	U	A	10	G	U	C	G	30
	G	U	C	G	7	G	U	U	A	5
LU2698	G	U	G	C	98	G	U	G	C	100
LU2847	G	U	G	C	83	G	U	G	C	100
	G	U	C	G	17					
	G	U	U	A	0.3					

e and *i* stand for external and internal nucleotides, respectively. The data has been obtained based on the available rRNA alignments (10). Nucleotides in red gave the name to each motif. For all cases of AGPM, the *E. coli* numbering is used. For the statistics, only those cases where the identities of all four nucleotides are known have been considered.

SUPPLEMENTAL TABLE III. The constructs used for measuring the efficiency of stop codons readthrough and frameshifting events

Name of plasmid	Sequence of cassette (in bold) cloned into the luciferase gene	Type of mutation incorporated
pLuc-A	atg **GAG CTC TAT CTT GTA AGA TCT** gaa gac gcc…	Wild-type
pLuc-B	atg **GAG CTC TAT CTT TGA AGA TCT** gaa gac gcc…	UGA stop (Opal)
pLuc-C	atg **GAG CTC TAT CTT TAA AGA TCT** gaa gac gcc…	UAA stop (Ochre)
pLuc-D	atg **GAG CTC TAT CTT TAG AGA TCT** gaa gac gcc…	UAG stop (Amber)
pLuc-E	atg **GAG CTC TTT TT* TTA TGG AGA TCT** gaa gac gcc…	(-1) frameshifting
pLuc-F	atg **GAG CTC TTT TTTT TTA TGG AGA TCT** gaa gac gcc…	(+1) frameshifting

All miscoding pLuc-plasmids contained the firefly luciferase gene in which a specific cassette (bold) harboring a mutation (shown in red) and the *SacI* restriction site (underlined) were incorporated between the ATG start codon and the second codon. The place in the -1 frameshifting cassette where the -1 frameshift was expected to occur is marked by asterisk. In the same way, the nucleotide "T" shown in red within the +1 frameshifting cassette shows the place where +1 frameshift was expected to occur to produce a functional luciferase protein.

SUPPLEMENTAL TABLE IV. The constructs used for measuring the efficiency of mis-incorporation events

Name of plasmid	The mutation within the luciferase gene arranged for measuring the efficiency of amino acid mis-incorporation
pLuc-G	R218G (AGA→GGA)
pLuc-H	R218S (AGA→AGU)

For these two mis-incorporation constructs, the wild-type cassette was cloned between the first and the second codon.

SUPPLEMENTAL TABLE V. GFPuv3 activity of the six clones with the reverse mutation U1192C in the 16S rRNA

Clone	GFPuv3 activity with the reverse U1192C mutation in the 16S rRNA (%)
A1	100
A7	80 ± 3
A2	75 ± 6
A12	62 ± 8
A10	8 ± 2
A13	5 ± 2

The ribosome activity was calculated as the mean ± standard deviation of five to eight independent experiments.

3.13 Supplemental references

1. Lee, K., Holland-Staley, C. A., and Cunningham, P. R. (1996) *RNA* **2**, 1270-1285
2. Rackham, O., and Chin, J. W. (2005) *Nat. Chem. Biol.* **1**, 159-166
3. Chandler, L. R., and Lane, A. N. (1988) *Biochem. J.* **250**, 925-928
4. Bennett, G. N., and Yanofsky, C. (1978) *J. Mol. Biol.* **121**, 179-192
5. Jensen, P. R., and Hammer, K. (1998) *Appl. Environ. Microbiol.* **64**, 82-87
6. Viviani, V. R., Uchida, A., Viviani, W., and Ohmiya, Y. (2002) *Photochem. Photobiol.* **76**, 538-544
7. Belanger, F., Gagnon, M. G., Steinberg, S. V., Cunningham, P. R., and Brakier-Gingras, L. (2004) *J. Mol. Biol.* **338**, 683-693
8. Sigmund, C. D., Ettayebi, M., Borden, A., and Morgan, E. A. (1988) *Methods Enzymol.* **164**, 673-690
9. Dyer, B. W., Ferrer, F. A., Klinedinst, D. K., and Rodriguez, R. (2000) *Anal. Biochem.* **282**, 158-161
10. Wuyts, J., Perriere, G., and Van De Peer, Y. (2004) *Nucleic Acids Res.* **32**, D101-103
11. Gagnon, M. G., and Steinberg, S. V. (2002) *RNA* **8**, 873-877
12. Mokdad, A., Krasovska, M. V., Sponer, J., and Leontis, N. B. (2006) *Nucleic Acids Res.* **34**, 1326-1341

Chapitre 4

Article 3

Recurrent RNA motifs as probes for studying RNA-protein interactions in the ribosome

4. Recurrent RNA motifs as probes for studying RNA-protein interactions in the ribosome

Matthieu G. Gagnon[1], Yury I. Boutorine[1] and Sergey V. Steinberg

Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Manuscript submitted.

[1] Both authors contributed equally to this work

Running title: RNA-protein interactions in the ribosome

Contribution of each author:

Matthieu G. Gagnon: Developed the detailed experimental scheme of the library design of motif L657, performed the selection and characterization of the functional clones of motif L657, the molecular dynamic simulations, the statistical analysis and participated in data analysis and preparation of manuscript and figures.

Yury I. Boutorine: Developed the detailed experimental scheme of the library design of motif L639, performed the selection and characterization of the functional clones of motif L639, and participated in data analysis and preparation of manuscript and figures.

Sergey V. Steinberg: Developed the general library design and participated in data analysis, preparation of manuscript and figures.

4.1 Abstract

To understand how the nucleotide sequence of ribosomal RNA determines its tertiary structure, we developed a new approach for identification of those features of rRNA sequence that are responsible for formation of different short-range and long-range interactions. The approach is based on the co-analysis of several examples of a particular recurrent RNA motif. For different cases of the motif, we design combinatorial gene libraries in which equivalent nucleotide positions are randomized. Through *in vivo* expression of the designed libraries we select those variants that provide for functional ribosomes. Then, analysis of the nucleotide sequences of the selected clones would allow us to determine the sequence constraints imposed on each case of the motif. The constraints shared by all cases are interpreted as providing for the integrity of the motif, while those ones specific for individual cases would enable the motif to fit into the particular structural context. Here we demonstrate the validity of this approach for three examples of the so-called along-groove packing motif found in different parts of ribosomal RNA.

Keywords: ribosomal RNA / ribosome structure / recurrent motif / combinatorial library / selection *in vivo* / ribosomal protein

4.2 Introduction

The ribosome is a large ribonucleoprotein complex that performs protein synthesis in all living organisms. It consists of three RNA chains, 23S, 16S and 5S and of several dozens proteins [reviewed in (Steitz 2008)]. The tertiary structure of the ribosome is defined by the nucleotide and amino acid sequences of its components, although the code of correspondence between the sequences and the tertiary structure is not simple. For each element of the ribosome tertiary structure, its nucleotide or amino acid sequence plays a dual role: not only does it determine the particular conformation of the element, but also the way this element interacts with other structural elements. Therefore, understanding how the ribosome structure forms would require the

elucidation of the sequence constraints that enable the sequence of each element to play both roles.

In this paper, we suggest a new approach to study different types of interactions existing in the ribosome, which would allow us to distinguish between the nucleotide sequence requirements associated with the integrity of a local rRNA arrangement and those associated with the interactions of this arrangement with other structural elements, RNA or proteins. The approach is based on co-analysis of several examples of a particular recurrent RNA motif, which are positioned in different parts of the ribosome structure and have identical or very similar conformations [reviewed in (Batey et al. 1999; Moore 1999; Noller 2005)]. For different cases of the same motif, we design combinatorial gene libraries through randomization of equivalent nucleotide positions and select those variants that provide for functional ribosomes. Then, for each case of the motif, we determine the limits of nucleotide variability and compare them with the analogous limits for the other cases of the same recurrent motif. Such comparison allows us to identify the aspects of the nucleotide sequences that are common for all cases and to distinguish them from those that are unique to a particular case. The common aspects would thus be interpreted as those responsible for the integrity of the motif, while the unique ones would characterize the interaction of each particular case of the motif with its own structural context. Here we demonstrate the validity of this approach for the so-called along-groove packing motif (AGPM), which is found in more than a dozen places of the ribosome structure (Gagnon and Steinberg 2002; Mokdad et al. 2006).

4.3 Results

4.3.1 Background: general description of the AGPM

The AGPM represents the arrangement of two double helices closely packed *via* their minor grooves in the way that a sugar-phosphate backbone of one helix packs along the minor groove of the other helix and *vice versa* (Fig. 1) (Gagnon and Steinberg 2002). Due to the frequent occurrence, the AGPM constitutes an important element of the ribosome structure. Its major role consists in bringing two elements of the rRNA

secondary structure together into a tight specific arrangement. In addition, the tRNA molecules located in the P- and E-sites are bound to 23S rRNA with help of two AGPMs. Therefore, the elucidation of the rules that govern the formation of the AGPM in different structural environments is essential for understanding how the ribosome structure forms and functions.

Within the AGPM, one of the two chains of each helix is packed in the minor groove of the opposite helix. This chain is positioned closer to the center of the arrangement and is thus called internal. The other chain of each helix stays at the periphery of the arrangement and is called external (Fig. 1) (Gagnon and Steinberg 2002). Although in each helix, the area of the inter-helix contacts spreads over four base pairs, the most extensive inter-helix interactions occur at the center of the contact area between two base pairs, which we call central. The close packing of the helices requires that one of the two central base pairs be Watson-Crick (WC), while the other one be GU (Gagnon and Steinberg 2002) (henceforth, in the two-letter identity of each base pair, the first and second letter stand for the external and internal nucleotide, respectively). The arrangement of the central base pairs shown in Figure 2a allows the formation of the network of five inter-helix hydrogen bonds. Within this arrangement, the internal and external nucleotides are responsible, respectively, for about 70% and 30% of all inter-helix contacts formed by each central base pair. The exchange of the WC and GU base pairs between the two helices does not disturb their close packing (Gagnon et al. 2006; Mokdad et al. 2006). Henceforth, the combination of GU and a WC as central base pairs will be referred to as the GU-WC pattern.

Although most cases of AGPM follow the GU-WC pattern, there are also a few cases in which this pattern is not observed. In particular, in motif L2291 from *Haloarcula marismortui* (Ban et al. 2000), both central base pairs are WC, which provides a crack between the two helices (Fig. 2b). This case seems more of an exception, because in most organisms, including *Escherichia coli*, motif L2291 follows the GU-WC pattern (Wuyts et al. 2004). At the same time, this case shows that the absence of the close packing is not necessarily critical for the integrity of the motif. The existence of arrangements alternative to GU-WC raises the question of how much the AGPM structure can differ from the standard pattern without being destroyed altogether.

It is also possible that the scope of the allowed variations of the central base pairs depends on the structural context in which each AGPM case appears and thus is not necessarily the same for different representatives of the motif. To explore these possibilities, we constructed combinatorial gene libraries for three AGPMs located in different places of the ribosome structure. In each library, all four nucleotides composing the central base pairs were fully randomized and the variants providing for a functional ribosome were selected. The co-analysis of the nucleotide sequences of all selected clones allowed us to elucidate the constraints imposed on the structure of each motif and to connect these constraints to the particular interaction of the motif with its surroundings.

4.3.2 The motifs studied

In this study, we consider three AGPMs: S296, L639 and L657 (Fig. 3a). In the available X-ray structures, all three motifs follow the GU-WC pattern. The structural contexts in which they appear within the ribosome are, however, different. Motif S296 is located at the center of the small ribosomal subunit and is formed by helices h3 and h12, which are distant from each other in the 16S rRNA secondary structure. An unusual feature of S296 is that it does not directly interact with any other part of rRNA or with a ribosomal protein. This aspect determined our initial choice of this motif as a context-free model system to study the general rules that govern the formation of the AGPM (Gagnon et al. 2006).

The other two motifs, L639 and L657, are located on the solvent side of the 50S subunit far from all functional centers of the ribosome. They are formed by helices H29-H31 (motif L639) and H27-H28 (motif L657). In each of the latter motifs, unlike in S296, both helices are neighbors in the 23S rRNA secondary structure. Also unlike S296, both motifs L639 and L657 participate in interactions with ribosomal proteins. In L657, nucleotide 600 of helix H27, which occupies the external position of a central base pair, forms a tight contact with residues L27, K99 and M100 of protein L4 (Fig. 4a). All three residues interact only with the ribose of nucleotide 600, and not with the base. Based on the available experimental data, one can suggest that the interaction of protein L4 with motif L657 is critical for the association of this protein with the 23S

rRNA (Li et al. 1996). In motif L639, nucleotides of the central base pairs are not directly involved in interactions with other parts of the 50S structure. However, nucleotides 650 and 651, which are proximate to the external nucleotide 649 of a central base pair, directly contact residues T16 and G17 of protein L35 (Fig. 4b). Again, it is not the bases, but the sugar-phosphate backbone of nucleotides 650-651 that forms contacts with protein L35.

In this paper, we demonstrate how the above-mentioned differences in the structural contexts of the three chosen motifs affect the variability of the central base pairs.

4.3.3 Cloning and selection of the functional clones

As mentioned above, in each of the three AGPMs all four nucleotide positions forming the two central base pairs were fully randomized. As a result, each combinatorial gene library provided 4^4=256 possible variants, of which only some were expected to make the ribosome functional. For selection of functional variants of motif S296, we used the specialized translation system, which is based on the expression of a modified 16S rRNA having an alternative anti-Shine-Dalgarno sequence (Hui and de Boer 1987; Lee et al. 1996; Belanger et al. 2004; Abdi and Fredrick 2005; Rackham and Chin 2005; Gagnon et al. 2006). In this system, clones were selected by the ability to survive in the presence of chloramphenicol due to the synthesis of the chloramphenicol acetyl-transferase protein (CAT). The quantification of the efficiency of the selected clones was made through the measurement of the green fluorescence protein (GFP) activity. Both proteins, the CAT and the GFP, were synthesized from mRNAs containing the modified Shine-Dalgarno sequence (Gagnon et al. 2006). For selection of functional variants of the L639 and L657 motifs located in the 23S rRNA, we used the ribosome knock-out strain SQ380 (Asai et al. 1999a; Asai et al. 1999b). In this experimental system, clones were selected based on the ability of a plasmid-based rRNA to maintain life in the absence of other sources of ribosomal RNA. Sequencing of the PCR products before cloning showed that all four positions were fully randomized. The efficiency of the clones was evaluated by measuring the doubling time of the cells (see Materials and Methods). The complete list of the selected clones from all three libraries

is shown in Table I. For convenience, the names of the selected variants of motifs S296, L639 and L657 start with letters A, B and C, respectively.

4.3.4 Analysis of the selected clones: the minimal requirement for the integrity of the AGPM

As expected, in all three selections we have found clones following the GU-WC pattern (clones A5, A7, A8, B1, B8, B14, B18, C7, C13, C55, C64, C78 and C85 in Table I). We believe that in all these clones, the coexistence of the GU and WC central base pairs reflects the close packing of the two helices. In clones A5, A7, B14, B18, C13, C55 and C85, compared to the wild-type *E. coli* ribosome, the GU and WC base pairs have exchanged between the helices, which, however, does not affect the packing (Gagnon et al. 2006; Mokdad et al. 2006). For variants of motif S296, due to the usage of the specialized translation system, the efficiency of the ribosomes could be accurately measured. Correspondingly, among the variants of this motif, those that followed the GU-WC pattern had generally a high activity (Table I). These data demonstrate that the structural integrity of motif S296 is important for the ribosome function.

Surprisingly, in all three libraries, the majority of selected clones did not follow the standard GU-WC pattern. Moreover, as one can see in Table I, the majority of selected clones contained non-standard nucleotide combinations as central base pairs. Among such combinations, we found UU, CU, UC, UG, CA, AC, CC, GG, GA and AG. Although the A-clones that contained such combinations were generally characterized by a reduced activity, this activity was still sufficient to allow the cells to survive under an elevated concentration of chloramphenicol (see Materials and Methods). Similarly, even though the doubling time of the selected B- and C-clones containing non-standard nucleotide combinations as central base pairs was generally somewhat longer than that of the wild-type (Table I), all such clones were perfectly viable. These findings allow us to conclude that in all three AGPMs tested, the close packing between the helices, which is manifested by the maintenance of the GU-WC pattern, is not a prerequisite of the ribosome function: the ribosome can function, although, generally, with a reduced efficiency, even in the absence of the close helix packing.

Based on the fact that most selected clones contained abnormal dinucleotide combinations, one could suggest that none of the three tested AGPM arrangements is essential for the basic ribosome function. This would mean that there are no rigid constraints imposed on the structure of the central base pairs in all three motifs, so that the ribosome would maintain residual activity regardless of the quality of their inter-helix contacts. Further analysis, however, showed that such a simple suggestion was incorrect. Even though many selected clones did not fit to the standard pattern, most of them shared another feature: regardless of the particular motif, a non-standard base pair was present only in one of the two helices, while the opposite central base pair in almost all clones was either WC or GU (as we defined above, in the GU base pair nucleotides G and U belonged, respectively to the external and internal strand, as in the standard GU-WC pattern).

The presence of a WC or GU base pair even in only one of the two helices could play a critical role in the AGPM's integrity. An obvious effect of such a base pair would be the stabilization of the corresponding double helix. Then, a stable double helix would be able to work as a scaffold for folding and proper positioning of the second helix. In particular, it will enable one of the two nucleotides forming a non-standard combination in the second helix to keep the same position and to form all inter-helix interactions exactly as it does in the standard AGPM structure (Fig. 2c). Because, as mentioned above, the internal nucleotide is responsible for most inter-helix contacts, the preservation of its position will provide a notably higher stabilizing effect on the whole arrangement compared to the situation when instead, the external nucleotide stayed at its place. Together with the opposite central base pair, the internal nucleotide will form a nucleotide triple (Fig. 2c). As a result, all nucleotides of the AGPM will stay at their standard positions except the external central nucleotide of the second helix. Depending on its identity, the latter nucleotide could still form a base pair with its internal partner or, alternatively, it would bulge out. However, given that in about 75% of all abnormal dinucleotide combinations found in the selected clones the external nucleotide is a pyrimidine (Table I), the instability associated with the existence of such bulge would usually be relatively modest.

Among the total of 48 selected clones, there are only two exceptions (A11 and B16) that contain neither a WC nor GU base pair (Table I). Interestingly, both clones have a CA base pair combination in one of the helices. Analysis of the AC/CA dinucleotide combinations existing in the ribosome structure (for this purpose, we took the structure of the 50S subunit from *H. marismortui*, (Ban et al. 2000)) showed that on different occasions, these combinations form as GU-like or WC-like arrangements (Fig. 5a, b). The arrangement of the CA dinucleotide combination as a WC-like base pair would fit clones A11 and B16 to the common pattern with all other clones.

Based on the fact that in all three libraries all selected clones share the same ability to form at least one central base pair, and, consequently, a nucleotide triple, one can strongly suggest that those variants that did not fit to this requirement were unable to provide for a functional ribosome. From this point of view, the formation of a WC, GU or AC/CA base pair in one of the two helices would constitute the minimal requirement for the integrity of the AGPM. Henceforth, the central base pair that forms a nucleotide triple with the internal nucleotide of the opposite helix will be called structure-forming base pair.

4.3.5 Molecular dynamics simulations

To test the ability of the nucleotide triple to stabilize the structure of the AGPM, we performed molecular dynamics (MD) simulations on specially modeled AGPM constructs (Supplemental Fig. 1a, b). The modeling of the constructs and the particular conditions of the MD simulations are explained in Materials and Methods. In all constructs, the identities of all nucleotides were the same except those four nucleotides that composed the central base pairs (Supplemental Fig. 1a, b).

In the four tested complexes, the central base pairs were GU-CG, GU-UG, GC-UG and UG-UG. In all simulations, the CG, GC and GU dinucleotide combinations were initially arranged as normal base pairs. In the UG combinations, however, the location of the guanosine corresponded to the position of the internal nucleotide in the standard AGPM structure, while the uridine was bulged out. Thus, the GU-CG combination corresponded to the standard AGPM structure, the GU-UG and GC-UG combinations contained nucleotide triples with, respectively, GU and GC as structure-forming base

pairs, while the UG-UG combination did not contain a structure-forming base pair and, correspondingly, did not contain a nucleotide triple. For the latter combination, the initial arrangement consisted of two guanosines occupying the internal positions, while the two uridines were bulged. During the simulations, the integrity of the inter-helix contact was monitored by measuring the distance between the O2′ atoms of the two internal nucleotides, which were initially connected by a hydrogen bond (see Fig. 2). The stability of the inter-helix arrangement was thus evaluated by the time required to break the contact between the riboses of the two internal nucleotides. For each complex, the simulations were performed four times, and Figure 6 shows the typical results for each case.

In the MD simulations performed for the GU-CG combination, the break between the two internal riboses occurred after 800 ps of simulation (Fig. 6a). In the cases of combinations GU-UG and GC-UG, the break took about 500 and 300 ps, respectively, (Fig. 6b, c), while for combination UG-UG the break occurred within the first 10 ps (Fig. 6d). Based on the results of these simulations, one can conclude that although the arrangements of the two double helices mediated by a nucleotide triple are generally less stable than the arrangement following the GU-WC pattern, they are overwhelmingly more stable than the arrangement characterized by the absence of a nucleotide triple.

Interestingly, in the performed simulations, the GU-UG construct had a notably longer life-time than the GC-UG construct. Such a higher stability of the GU-based construct correlates with the fact that, compared to the construct in which the structure-forming base pair was WC, this one contained an additional inter-helix hydrogen bond between the amino group of the uridine-paired guanosine and the O2′-H group of the opposite internal guanosine of the UG base pair (for reference, see Fig. 2a, b). Taken together, these simulations clearly demonstrate that the presence of a structure-forming base pair in one of the two helices is critical for the stability of the whole arrangement and explains the fact that in our library selection all clones contained such base pair in at least one of the two helices.

4.3.6 The symmetry of the central base pairs in motif S296

In the A-clones that follow the GU-WC pattern, helices h3 and h12 harbor base pairs GU and WC with the same frequency (Table I). Also, among the A-clones in which the minimal requirement related to the formation of the nucleotide triple is respected, the structure-forming base pair appears in each of the two helices with comparable frequency. Finally, abnormal dinucleotide combinations that do not provide for a structure-forming base pair are found in both helices in almost the same number of A-clones. Based on these facts, we can conclude that the ribosome function does not depend on the type of base pair that appears in each of the two helices h3 and h12, as long as the arrangement of the two base pairs follows a particular pattern. Such symmetry between the S296 variants fits well to the fact that none of helices h3 and h12 interacts with any other element of the ribosome structure. In this sense, motif S296 represents an unbiased context-free case of the AGPM.

4.3.7 Interaction of motif L639 with ribosomal protein L35

Compared to the A-clones, B-clones demonstrate a clear asymmetry between helices H29 and H31. In particular, in almost all B-clones, the structure-forming base pair is located within helix H31 (Table I). The only two exceptions are clones B3 and B6, where the structure-forming base pair is located in helix H29. Such asymmetry between helices H29 and H31 correlates well with the fact that in motif L639, unlike in motif S296, nucleotides 650 and 651, which belong to the external strand of helix H31, interact with ribosomal protein L35 (Fig. 4b). Although this interaction does not directly include nucleotide 649 of the central base pair, the fact that the neighboring nucleotides 650 and 651 form a tight contact with L35 would limit the mobility of nucleotide 649. Such reduced mobility, in turn, would limit the set of acceptable dinucleotide combinations for the central base pair in helix H31, making only WC and GU base pairs acceptable. Unlike H31, the opposite helix H29 does not form contacts with any other element of the ribosome structure. Correspondingly, the central base pair located in helix H29 harbors different dinucleotide combinations (Table I).

In two clones B3 and B6, the structure-forming base pair is located in the opposite helix H29. Interestingly, in both clones, the dinucleotide combination located in helix

H31 is GA. Our modeling experiment demonstrates that if the internal adenosine adopts a *syn* conformation, the position of the external guanosine within the GA base pair would be rather close to that existing in a WC base pair (Fig. 5c). The formation of such a GA base pair can thus be considered as an alternative way of fixing the position of the external nucleotide when the structure-forming base pair belongs to the opposite helix.

4.3.8 Interaction of motif L657 with ribosomal protein L4

Similarly to the previous case, analysis of the variants of motif L657 demonstrates a clear asymmetry between the two helices. Indeed, in the C-clones, the central base pair belonging to helix H27 is always GU, WC or AC, while alternative dinucleotide combinations are found exclusively in helix H28 (Table I). Again, like in motif L639, the conservative location of the structure-forming base pair in helix H27 correlates with the involvement of the external strand of this helix in a tight interaction with the ribosomal protein L4 (Fig. 4a).

However, a more detailed analysis reveals a substantial difference between the B- and C-clones. In the B-clones, the GU and WC base pairs seemed to be completely interchangeable. In particular, both GU and WC were able to function as structure-forming base pairs when the opposite helix harbored an alternative dinucleotide combination. In the C-clones, however, only GU plays such role, while a WC base pair appears exclusively in the clones following the GU-WC pattern (Table I). The only exception pertains to clone C84, in which CG in one helix coexists with AC in the other helix. This clone, however, will also follow the same GU-WC pattern if base pair A600-C657 forms in the GU-like manner (Fig. 5a).

Further analysis shows that the asymmetry between the GU and WC base pairs observed in the C-clones originates from the fact that in motif L657, unlike in L639, the ribosomal protein L4 forms a direct contact with nucleotide 600 of the central base pair. Below we demonstrate that the particular pattern of variability of the central base pairs in the C-clones will always keep the ribose of nucleotide 600 at the same position, thus allowing it to interact normally with protein L4. For the purpose of analysis, we divide all C-clones in Groups I and II, as shown in (Table I). Group I harbors all clones following the GU-WC pattern, including the above-mentioned clone C84, while all other

clones fall into Group II. Group II thus contains only four clones and in all of them, base pair 600-657 is GU.

A replacement of the GU base pair by WC in any Group II clone would lead to a displacement of nucleotide 600 from its position, which includes the rotation of the base by about 15º toward the major groove (Fig. 7a). As a result, different atoms of the ribose-600 would become displaced by about 1 Å. However small this displacement is, it is expected to be large enough to affect the interaction of nucleotide 600 with protein L4 and to make the ribosome non-functional. Such potentially damaging effect of the G600-U657→WC replacement can explain why Group II clones contain exclusively the G600-U657 base pair.

While non-existent in Group II clones, a WC base pair 600-657 can be found in Group I clones, i.e. when the opposite central base pair 623-605 is GU. The functionality of such clones suggests that the presence of base pair G623-U605 is somehow able to rescue the ribosome function which would have otherwise been damaged by the replacement of the GU base pair 600-657 by WC. To elucidate the particular mechanism of the rescue of the ribosome function in this case, we took the structure of motif L657 existing in the ribosome of *H. marismortui* (Ban et al. 2000). Compared to *E. coli*, in *H. marismortui* the WC and GU base pairs of this motif are exchanged in their positions (Fig. 3b). The superposition of the structures of motif L657 in these two organisms would thus visualize the local conformational changes associated with the GU⇔WC replacement (Fig. 7b). This superposition demonstrates that within the GU-WC pattern, after the exchange of the GU and WC base pairs, the atoms of the internal riboses become displaced by >1 Å, while the equivalent atoms of the external nucleotides remain within 0.2 Å of their original positions. Such immovability of the external riboses will preserve the interaction of nucleotide 600 with protein L4 in the situations when motif L657 follows the GU-WC pattern.

4.3.9 Instant evolution versus natural evolution

An important aspect of our approach consists in the elucidation of the limits of variability for a set of available nucleotide sequences. The sequences are, in turn, collected using the so-called instant evolution, when the regions of interest in rRNA are randomized and viable clones are selected. An alternative way of collecting the information on nucleotide sequence variability could consist in gathering all available nucleotide sequences of naturally selected rRNA. It thus would be interesting to know whether the particular aspects of the rRNA structure and of its interaction with ribosomal proteins revealed through the analysis of the instantly selected clones could be elucidated solely based on the naturally selected sequences. Here we discuss the major results of this analysis, while all details are given in the Supplemental Data.

A major feature of the nucleotide sequences of the AGPMs existing in naturally selected rRNA consists in the overwhelming presence of the GU-WC pattern. In fact, for each of the three motifs studied, this pattern is followed in more than 97% of all naturally selected sequences (Supplemental Table I). Such a strong bias toward the GU-WC pattern differs the naturally selected sequences from the instantly selected ones, where this pattern is observed in only 16 clones out of 48. This difference is understandable, given that naturally selected rRNA sequences have evolved in a tough competition with similar sequences and are thus expected to have been optimized for the ribosome efficiency, while for a successful selection of a rRNA variant through the instant evolution a modest level of cell viability would be sufficient.

A relatively low number of the naturally selected sequences that do not follow the GU-WC pattern would question their validity as exceptions, because some of them could originate from sequencing errors. Also, an attempt to use the abnormal sequences to deduce subtleties of the interaction of AGPMs with a particular ribosomal protein would be problematic because details of such interaction could vary in different organisms. In general, analysis of naturally selected nucleotide sequences of rRNA can provide additional information on integrity of AGPMs and/or on their interaction with ribosomal proteins. However, due to the above-mentioned reasons, the value of this information would be essentially lower than that of the information based on the analysis of instantly selected sequences.

4.4 Discussion

4.4.1 The power of the approach

We present here a new approach for analysis of structure-function relationships in the ribosome, which consists in randomization of core nucleotides in different examples of the same recurrent RNA motif, selection of viable clones, and analysis of their nucleotide sequences. This approach allows us to identify those features of the rRNA nucleotide sequence that provide for the integrity of a particular arrangement and to distinguish them from the features responsible for the interaction of this arrangement with elements of its immediate structural context.

An important aspect of our approach consists in the usage of combinatorial rRNA gene libraries, which allows the exploration of a large array of nucleotide sequence possibilities based on a single act of cloning. The variations of nucleotide and base pair identities revealed through the library expression often exceed the variations observed in the naturally selected rRNA sequences, thus providing new otherwise inaccessible information on the nature of different short-range and long-range interactions within the ribosome. Compared to approaches based on direct mutagenesis of rRNA, the usage of combinatorial libraries does not require any preliminary hypotheses on the nature of the interactions in which the particular region is involved. As a result, the set of nucleotide sequences obtained through selection from a combinatorial library would characterize the studied RNA arrangement more objectively than a set of premeditated constructs. Analysis of selected clones allows us to determine the limits of nucleotide variability in a given set of clones.

Another important feature of our approach pertains to the usage of recurrent RNA motifs and to the fact that in all tested cases, the randomized nucleotides occupy equivalent positions. These aspects make possible a systematic comparison of the limits of variability related to different examples of the same motif. Based on such comparison, we can determine common features valid for all examples of the motif and distinguish them from features specific to particular cases. The common features, which are deduced from the limits of variability of all selected clones in all studied cases of a motif, would constitute the minimum requirement for the motif formation. The specific

features, in their turn, are determined as a difference between the limits of variability related to the particular case and the limits of variability obtained for all tested cases; they are attributed to the interaction of the given case of the motif with its surroundings.

4.4.2 New findings about the AGPM: principles of RNA structure formation

As a proof of principle, we used the AGPM, a recurrent RNA arrangement frequently found in the ribosome structure (Gagnon and Steinberg 2002; Mokdad et al. 2006). In this motif, the optimal interaction between the two double helices is achieved when at the core of the arrangement a WC base pair in one helix is packed against a GU base pair in the other helix. At the same time, the coexistence of the WC and GU as the central base pairs is not a prerequisite for the AGPM formation, so that deviations from the optimal helix packing are known among naturally occurring rRNA sequences. Such softness of the requirement for the GU-WC pattern makes the nucleotides forming the central base pairs a useful object for randomization and selection in our approach. On one hand, the absence of rigid sequence requirements would facilitate the selection of alternative variants. On the other hand, a clear dependence of the stability of the AGPM on the identity of the central base pairs would limit the scope of acceptable variants, thus making the selection a sensible procedure. For the analysis, we chose three representatives of the AGPM from both ribosomal subunits for which the central base pairs had different levels of interaction with other structural elements of the surrounding, varying from the complete absence of interaction (S296) to the presence of indirect (L639) and tight direct interaction (L657) with ribosomal proteins.

Analysis of the selected clones has provided plenty of new information on different aspects of the AGPM structure. First, it has allowed us to formulate the minimal requirement for the AGPM formation, which consists in the presence of either WC or GU as a structure-forming base pair in only one of the two helices. The validity of such requirement infers the existence of a cross-talk between the helices, so that the introduction of instability in one helix can be partly neutralized by the remaining solidity of the other helix. We argued that the requirement for the presence of a structure-forming base pair in one helix pertains to the ability of such a base pair to accommodate the internal nucleotide of the opposite helix, so that the position of only one external

nucleotide would be changed compared to that observed in the optimal helix packing. Our MD simulations showed that bulging of the external nucleotide in only one central base pair does not dramatically reduce the motif's stability, thus providing additional support for the suggested minimal requirement. Also, the existence of one of the two central base pairs would enable the corresponding double helix to work as a scaffold for the folding of the second helix, thus facilitating the formation of the whole arrangement.

Our analysis also shows that AC/CA combinations can form as WC (AC and CA) or GU (only AC) base pairs, which would allow them to play the role of a structure-forming base pair. Although the particular scheme of hydrogen bonds in these base pairs is obscure, the appearance of the AC/CA dinucleotide combinations in the ribosome structure makes such base pairs valuable alternatives to the more common WC and GU base pairs. A potential drawback of the AC/CA base pairs consists in their inability to fix the juxtaposition of the two bases, as they can make both WC-like and GU-like base pairs. Such uncertainty in the base juxtaposition is mostly attributed to the AC combination, which in both cases, whether it is formed as WC or as GU, would be able to play the role of the structure-forming base pair. The opposite combination CA can play such role only if it forms as WC and not as the UG base pair, which would eliminate the uncertainty related to its structure. Interestingly, in the five selected clones containing an AC/CA combination (clones A11, B16, B21, C52 and C84, Table I), the AC combination is present only in C84. Because in this clone, the ribosome function requires that base pair 600-657 have a GU or GU-like structure, the AC combination would have the distinct nucleotide juxtaposition unreachable for CA. In all other cases, where the formation of the base pair as either WC or GU would make no difference for the ribosome functionality, we invariably see the CA combination. Therefore, in all selected clones containing AC/CA combinations, the uncertainty of the base pair structure is avoided through the particular choice between AC and CA. Whether this pattern of the appearance of the AC/CA combinations is only a coincidence or indeed reflects the fundamental limitation on the usage of such base pairs, would need further analysis.

Another interesting aspect pertains to the GA combination, which is found only in clones B3 and B6. Because in both cases, GA occurs in helix H31, which interacts with

protein L35, we expect this dinucleotide combination to have either a WC-like or GU-like structure. Indeed, if the internal adenosine 639 takes the unusual *syn*-conformation, the juxtaposition of the glycosidic bonds in such GA combination would be close to that observed in a WC base pair (Fig. 5c). On the other hand, we doubt whether a GA combination, even if it forms as a WC-like base pair, is able to function as a structure-forming base pair, i.e. to facilitate the formation of the AGPM. Indeed, the *syn*-conformation is rather unusual in the RNA world and in most cases it forms as an adaptation to a particular structural context. From this point of view, the GA base pair with the adenosine having the *syn*-conformation would most probably emerge as a response to the interaction with the opposite helix and not the other way around. It is not surprising, therefore, that in both clones B3 and B6 the GA combination in helix H31 coexists with either the GU (clone B3) or AU (clone B6) combination in helix H29. We thus suggest that helix H29, which in both clones B3 and B6 contains the structure-forming base pair, folds first and then assists the folding of helix H31. Only if the base pair in the latter helix can form as either WC or GU, the interaction of motif L639 with protein L35 will not be disturbed, which automatically provides for the particular orientation of adenosine 639. Whether indeed, GA is unable to serve as a structure-forming base pair despite its ability to accept a WC-like geometry, would need further analysis.

Finally, analysis of the C-clones allowed us to make two important observations related to the AGPM structure. First, we showed that the GU⇔WC replacement at the center of the inter-helix contact mostly leads to the displacement of the riboses of internal nucleotides, while the riboses of the external nucleotides remain virtually unmovable. This conclusion is based on the superposition of the structures of motif L657 in the *E. coli* and *H. marismortui* ribosomes (Fig. 7b) and is supported by the fact that such replacement does not affect the *E. coli* ribosome functionality even though the external nucleotide 600, which forms a direct contact with protein L4, becomes involved in a WC base pair instead of GU.

The second observation pertains to the asymmetry between GU and WC, according to which the presence of the G600-U657 base pair is sufficient for the ribosome function, while a WC base pair in this position can make the ribosome functional only if

the opposite base pair is GU. A thorough analysis of our experimental data in view of the known AGPM structures led us to the conclusion that this asymmetry relates to the fact that the ribose of nucleotide 600 in the WC base pair 600-657 has different positions depending on whether the opposite base pair is GU or not. In other words, the structure of helix H28 could influence the structure of helix H27 by allowing the ribose of nucleotide 600 to preserve its position when base pair 600-657 is WC. Such influence would represent another example of a cross-talk between the two helices within the AGPM.

4.4.3 New findings about the AGPM: principles of RNA-protein interaction

Within the complexes of motifs L639 and L657 with, respectively, proteins L35 and L4, the positions of the external strands that directly interact with the proteins are fixed. Generally, there are two possibilities for this fixation to take place either before or upon the formation of the rRNA-protein contacts. Our results, however, support only one of the two possibilities. The fact that in the selected B- and C-clones, the structure-forming base pair systematically belongs to the helix interacting with the protein, while combinations like UU or UC, which do not provide for a solid conformation of the external strand, occur exclusively in the opposite helix, clearly demonstrates that for the ribosome to be functional, the position of the strand interacting with the protein must be fixed by the means of RNA alone. Therefore, the formation of the particular conformation of the strand precedes its interaction with the protein and is a prerequisite condition for it.

The specificity of the RNA-protein interaction in both motifs originates not from contacts with unique parts of nucleotides, but instead, from the particular arrangement in space of such sequence-independent elements as riboses and the backbone. The proper positioning of these elements, however, is achieved with an active participation of bases, mainly through the particular type of base pairing, and is thus sequence-specific. We can say that the uniqueness of RNA contacts with both proteins L35 and L4 is achieved through the specific arrangement of non-specific RNA elements. It seems probable that the same principle is valid for rRNA interaction with many other ribosomal proteins. Moreover, based on the fact that similar phenomena have also been observed in the

interaction of tRNA with aminoacyl-tRNA synthetases (McClain et al. 1998), the same principle can be essential for RNA-protein interactions at large.

4.4.4 The sensitivity of the approach

In the cases of AGPM analyzed here, the positions of the external nucleotides of the central base pairs have demonstrated different levels of flexibility, which can be divided in three categories:

A. Unrestrained. The external nucleotide can be involved in a base pair with its internal counterpart, but can also be bulged out. The strand to which this nucleotide belongs does not form long-range interactions. In different clones, the position of this nucleotide can vary within 6-8 Å. This level of flexibility is attributed to both external nucleotides 27 and 301 of motif S296, as well as to nucleotide 634 of motif L639 and to nucleotide 623 of motif L657.

B. Restrained. The external strand to which this nucleotide belongs forms long-range interactions, which, however, do not touch the given nucleotide. The position of the nucleotide can vary within about 2 Å. This level of flexibility is attributed to nucleotide 649 of motif L639.

C. Fixed. The ribose is directly involved in a long-range interaction. The allowed variation in the position of the ribose atoms is about 0.2 Å. This level of flexibility is attributed to nucleotide 600 of motif L657.

Each category of the nucleotide flexibility corresponds to the particular pattern of variability of the central base pairs, and our approach has been sensitive enough to clearly distinguish between all three possibilities. Thus, the approach described here represents a powerful tool to study different types of short-range and long-range interactions in the ribosome and, potentially, in other RNA-protein complexes.

4.5 Materials and Methods

Bacterial strains and media

For all 30S subunit related procedures, we used the *E. coli* strain DH5α. For all 50S related procedures, we used the *E. coli* Δ7 prrn strain SQ380 (Δ*rrnE* Δ*rrnB* Δ*rrnA* Δ*rrnH* Δ*rrnG::lacZ* Δ*rrnC::cat* Δ*rrnD::cat* Δ*recA*/ptRNA67-Spc^R) carrying the rRNA-coding plasmid pHK*rrnC-sacB*-Kan^R (Asai et al. 1999a; Asai et al. 1999b). As a host for plasmids with the λP_L promoter, we used the *E. coli* strain POP2136 (*F¯ glnV44 hsdR17 endA1 thi-1 aroB mal¯ cI857 lambdaPR tet^R*). This strain contains the chromosomal cI857 allele coding for the thermo-sensitive repressor of the λP_L promoter (Pinard et al. 1993). Cultures were grown in the Luria–Bertani (LB) medium (Luria and Burrous 1957) or in the LB medium supplemented with appropriate antibiotics, 100 μg/mL ampicillin (Amp), 50 μg/mL kanamycin (Kan) and 40 μg/mL spectinomycin (Spc) (Sigma-Aldrich Canada).

Plasmids

The combinatorial 16S rRNA gene library of motif S296 was obtained previously using the specialized ribosome system cloned in plasmid pAMMG (Gagnon et al. 2006). For expression of wild-type and mutant 23S rRNA, we used plasmids pKK1192U-Amp^R (Brosius et al. 1981) and pLΔH1192U-Amp^R (Pinard et al. 1993). These plasmids contain an intact wild-type *rrnB* operon with the Spc-resistance marker mutation C1192U in the 16S rRNA. In plasmid pLΔH1192U, the transcription of the *rrnB* operon is controlled by the thermo-inducible λP_L promoter. In cells POP2136 at 30ºC, this promoter is repressed due to the presence of the temperature sensitive cI857 repressor encoded by the host chromosome.

Design of the combinatorial gene libraries

The four nucleotides comprising the two central base pairs of motifs S296, L639 and L657 were fully randomized using the overlapping extension PCR procedure (Ho et al. 1989). In this way, the entire regions comprising motifs S296 (902 bp), L639 and L657 (1541 bp or 2238 bp) were amplified by a multi-step-PCR. All PCR steps,

oligonucleotide sequences and restriction enzymes used for cloning are described in Supplemental Methods and Supplemental Table III. Transformation of the plasmids harboring the combinatorial 23S rRNA gene libraries into the SQ380 cells was performed by electroporation.

Plasmid replacement and selection of functional clones

The exchange of the resident wild-type pHK*rrnC*-*sacB*-Kan^R plasmid with the pKK1192U or pLΔH1192U plasmid carrying mutant 23S rRNA was performed as previously described with some modifications (Asai et al. 1999a; Asai et al. 1999b). First, the cell culture was grown for 1 hour at 37°C without antibiotic. Then, to facilitate the plasmid replacement, the growth continued for 3 more hours at 42°C in the presence of ampicillin. The increase of the temperature was required to inhibit the replication of the resident thermo-sensitive pHK*rrnC*-*sacB*-Kan^R plasmid thus promoting the effective displacement of the resident plasmid. Finally, the cultures were plated onto LB-Amp-Spc-agar plates (without NaCl) containing 3% sucrose and incubated for 16 hours at 30°C for efficient expression of the *sacB* gene conferring sucrose sensitivity (Gay et al. 1985; Blomfield et al. 1991). A total of ~1 x 10^5 transformants were obtained for both motifs L639 and L657, out of which several hundred grew after selection. For each library, 50 selected clones were checked on LB-Kan-agar plates to confirm the loss of the resident pHK*rrnC*-*sacB*-Kan^R plasmid followed by the sequencing of the 23S rRNA gene in the pKK1192U or pLΔH1192U plasmid.

Measurement of the ribosome efficiency and of the growth rates

The GFP activity of each A-clone was measured previously (Gagnon et al. 2006). For the B- and C-clones, the growth rates were measured with use of a Packard Fusion α-FP plate reader. The measurements were performed at 37 °C in the LB-Amp medium, starting with the 1:100 dilution of overnight cultures. For each measurement, we took five to eight colonies. The A_{600} data corresponding to the mid-log phase was used to construct a log-plot, from which the doubling time was deduced by a linear approximation.

Sequencing

Sequencing of the selected clones was performed on the LI-COR DNA sequencing system (Département de Biochimie, Université de Montréal) using primer 5′-actgaccgatagtgaaccagtaccgtgagg-3′ for reading positions 629, 634, 639 and 649 of motif L639 and positions 600, 605, 623 and 657 of motif L657. This primer was labeled with IRDye-800 (LI-COR Biosciences) at the 5′ end. In no case did mutations affect non-randomized nucleotides.

Molecular dynamics simulations

Molecular dynamics (MD) simulations were carried out on four different constructs, each composed of two double helices forming together the AGPM (Supplemental Fig. 1). To increase the stability of the helices during the simulations, each helix was capped on both ends by GAGA tetraloops. All complexes were based on the conformation of motif L657 in the crystal structure of the *E. coli* ribosome (pdb entry code 2aw4, (Schuwirth et al. 2005)) and had identical nucleotide sequences, except in the center, where the central base pairs were modified to obtain different starting nucleotide arrangements. The modification was done with use of the Insight II software (version 2000; Accelrys Inc., San Diego, CA). In the first construct, the central base pairs were GU and CG. Two other constructs contained combinations GU-UG and GC-UG, in which the GU and GC combinations formed normal base pairs. In each UG combination, the internal guanosine formed a triple with the opposite base pair, while the external uridine was bulged. Finally, in the UG-UG construct, both external uridines were bulged.

Each construct was subjected to an unrestrained energy minimization in the AMBER force field (http://ambermd.org) (300 steps of the steepest descent algorithm), followed by a restrained minimization using the conjugate gradient algorithm until a convergence was obtained. The restraints pertained to the positions of the nucleotides involved in the tetraloops, which were fixed during the minimization. Each MD simulation was done in the AMBER force field with the implicit solvent at 300 K. During the MD simulations, we fixed the positions of the C1′ atoms in nucleotides A_{16} of both helices and in nucleotide A_7 of the helix in which the central base pair is red (Supplemental Fig. 1a, b).

To maintain the integrity of the helices, minor distance constraints were imposed on the lengths of the hydrogen bonds in all base pairs, except the central ones, which did not contain any constraint (Supplemental Fig. 1a). These constraints were introduced as penalty $K \times (R\text{-}3.3)^2$ added to the energy function when the distance R between the two electro-negative atoms involved in the formation of a corresponding hydrogen bond exceeded 3.3 Å. The value of K was chosen to be 5 kcal / (mol × Å^2). Finally, after 1 ns simulation, the MD trajectories were analyzed using the Insight II/Analysis package and visualized on a Silicon Graphics Fuel computer.

4.6 Acknowledgments

We thank Drs C. L. Squires and S. Quan (Tufts University, Boston, MA) for providing the SQ380 strain and experimental advice for the use of the knock-out system. We are also grateful to Drs H.F. Noller (University of California, Santa Cruz, CA) for providing the pKK1192U plasmid, A.E. Dalhberg (Brown University, Providence, RI) for providing the POP2136 strain and the pLΔH1192U plasmid and Léa Brakier-Gingras for important discussions and experimental advice. The work was supported by an operating grant from Canadian Institutes of Health Research. M.G.G. held scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the Fonds de la Recherche en Santé du Québec (FRSQ).

4.7 References

Abdi NM, 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.

Asai T, Condon C, Voulgaris J, Zaporojets D, Shen B, Al-Omar M, Squires C, Squires CL. 1999a. Construction and initial characterization of *Escherichia coli* strains with few or no intact chromosomal rRNA operons. *J. Bacteriol.* **181**:3803-3809.

Asai T, Zaporojets D, Squires C, Squires CL. 1999b. An *Escherichia coli* strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria. *Proc. Natl. Acad. Sci. USA* **96**:1971-1976.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**:905-920.

Batey RT, Rambo RP, Doudna JA. 1999. Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed. Engl.* **38**:2326-2343.

Belanger F, Gagnon MG, Steinberg SV, Cunningham PR, Brakier-Gingras L. 2004. Study of the functional interaction of the 900 Tetraloop of 16S ribosomal RNA with helix 24 within the bacterial ribosome. *J. Mol. Biol.* **338**:683-693.

Blomfield IC, Vaughn V, Rest RF, Eisenstein BI. 1991. Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. *Mol. Microbiol.* **5**:1447-1457.

Brosius J, Ullrich A, Raker MA, Gray A, Dull TJ, Gutell RR, Noller HF. 1981. Construction and fine mapping of recombinant plasmids containing the *rrn*B ribosomal RNA operon of *E. coli*. *Plasmid* **6**:112-118.

Gagnon MG, Mukhopadhyay A, Steinberg SV. 2006. Close packing of helices 3 and 12 of 16S rRNA is required for the normal ribosome function. *J. Biol. Chem.* **281**:39349-39357.

Gagnon MG, Steinberg SV. 2002. GU receptors of double helices mediate tRNA movement in the ribosome. *RNA* **8**:873-877.

Gay P, Le Coq D, Steinmetz M, Berkelman T, Kado CI. 1985. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J. Bacteriol.* **164**:918-921.

Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**:51-59.

Hui A, de Boer HA. 1987. Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **84**:4762-4766.

Lee K, Holland-Staley CA, Cunningham PR. 1996. Genetic analysis of the Shine-Dalgarno interaction: selection of alternative functional mRNA-rRNA combinations. *RNA* **2**:1270-1285.

Li X, Lindahl L, Zengel JM. 1996. Ribosomal protein L4 from *Escherichia coli* utilizes nonidentical determinants for its structural and regulatory functions. *RNA* **2**:24-37.

Luria SE, Burrous JW. 1957. Hybridization between *Escherichia coli* and *Shigella*. *J. Bacteriol.* **74**:461-476.

McClain WH, Schneider J, Bhattacharya S, Gabriel K. 1998. The importance of tRNA backbone-mediated interactions with synthetase for aminoacylation. *Proc. Natl. Acad. Sci. USA* **95**:460-465.

Mokdad A, Krasovska MV, Sponer J, Leontis NB. 2006. Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.* **34**:1326-1341.

Moore PB. 1999. Structural motifs in RNA. *Annu. Rev. Biochem.* **68**:287-300.

Noller HF. 2005. RNA structure: reading the ribosome. *Science* **309**:1508-1514.

Pinard R, Payant C, Melancon P, Brakier-Gingras L. 1993. The 5' proximal helix of 16S rRNA is involved in the binding of streptomycin to the ribosome. *FASEB J* **7**:173-176.

Rackham O, Chin JW. 2005. A network of orthogonal ribosome x mRNA pairs. *Nat. Chem. Biol.* **1**:159-166.

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH. 2005. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**:827-834.

Steitz TA. 2008. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* **9**:242-253.

Wuyts J, Perriere G, Van De Peer Y. 2004. The European ribosomal RNA database. *Nucleic Acids Res.* **32**:D101-103.

4.8 Tables

Table I. Nucleotide sequences of the selected clones

Clone	Randomized positions				GFP [a] (%)
	Helix 12		Helix 3		
	301	296	27	556	
Motif S296					
Wild-type	G	U	G	C	100
A8	G	U	C	G	81±5
A5	G	C	G	U	85±8
A7	C	G	G	U	70±4
A2	G	C	G	C	79±6
A12	C	G	G	C	49±3
A3	C	G	A	G	26±3
A9	G	C	C	U	11±2
A11	U	G	C	A	22±3
A4	C	C	G	U	78±3
A6	C	C	G	C	10±2
A10	G	C	U	A	4±1

Motif L657	Randomized positions				Doubling time [a] (min)
	Helix 27		Helix 28		
	600	657	623	605	
Group I					
Wild-type	G	U	C	G	71±4
C64	G	U	A	U	76±6
C78	G	U	U	A	82±4
C7	G	U	G	C	75±5
C84	A	C	C	G	78±7
C13	A	U	G	U	75±8
C55	C	G	G	U	79±9
C85	G	C	G	U	75±5
C52	G	U	C	A	82±5
Group II					
C1	G	U	U	C	75±7
C50	G	U	U	G	72±1
C4	G	U	G	U	78±4
C97	G	U	U	U	82±2

Clone	Randomized positions				Doubling time [a] (min)
	Helix 31		Helix 29		
	649	639	634	629	
Motif L639					
Wild-type	G	U	C	G	71±4
B8	G	U	A	U	77±6
B1	G	U	G	C	71±6
B14	G	C	G	U	62±6
B18	A	U	G	U	65±6
B22	U	A	U	A	88±8
B20	A	U	A	U	67±3
B19	G	C	C	G	76±5
B5	G	C	A	U	89±7
B2	G	C	U	U	74±7
B3	G	A	G	U	82±7
B4	G	U	U	U	64±4
B6	G	A	A	U	75±4
B7	G	U	A	G	91±3
B10	C	G	U	G	78±6
B11	G	C	U	G	80±6
B12	A	U	U	U	67±6
B13	G	U	G	U	74±7
B15	G	U	U	G	83±4
B17	C	G	C	U	74±5
B9	G	C	G	G	84±6
B21	G	C	C	A	69±7
B16	C	A	C	U	69±6

In the sequences of clones, a continuous underline that includes both central base pairs indicates a close packing. The individually underlined base pairs are structure-forming.

[a]: The ribosome activity (GFP) and the growth rate (doubling time) were calculated as the mean ± standard deviation of five to eight independent experiments.

4.9 Figures

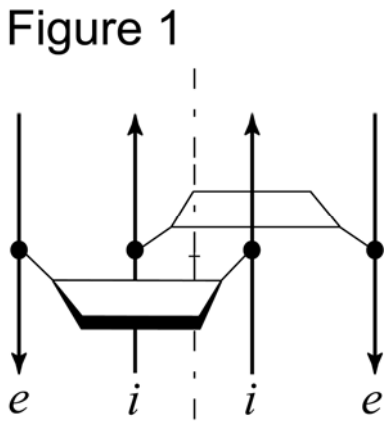


Figure 1. Schematic representation of the AGPM. Trapezoids stand for base pairs opened toward the minor grooves. Arrows represent backbones directed 5′→3′. The internal and external strands of both helices are marked by italic letters *i* and *e*, respectively. The internal strand of each helix is packed along the minor groove of the other helix. Rotation of one helix for 180° around the symmetry axis (dash-dotted line) superposes it with the other helix.

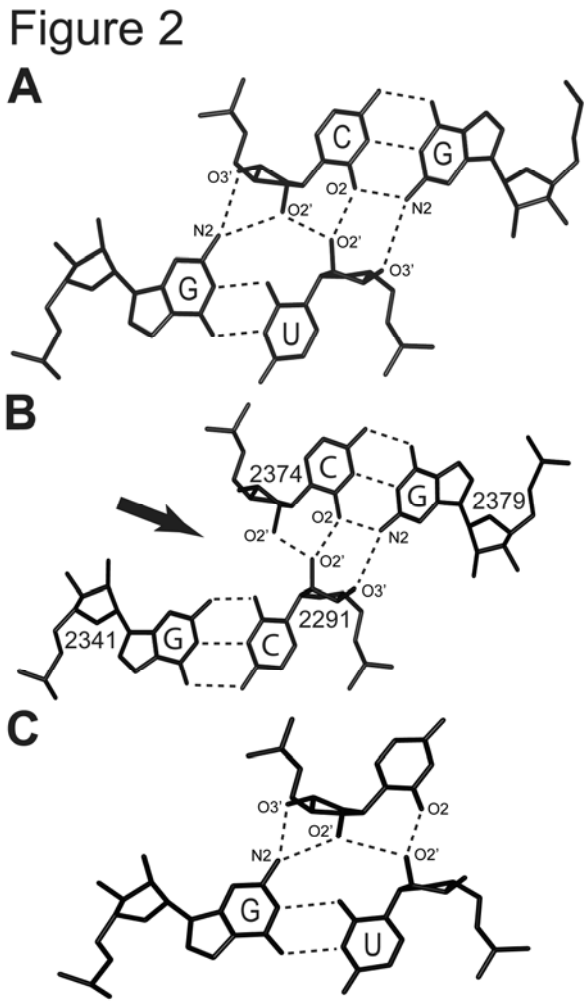


Figure 2. Different arrangements of the central base pairs in the AGPM.

(A) The canonical GU-WC arrangement. The presence of the GU base pair allows the close packing between the helices with the formation of the inter-helical network of five hydrogen bonds.

(B) The GC-GC juxtaposition taken from motif L2291 (*E. coli* numbering) in the *H. marismortui* 23S rRNA (pdb entry code 1s72.pdb, (Ban et al. 2000)). The absence of a GU base pair provides a crack between the two helices, which is indicated by the arrow.

(C) A model of a nucleotide triple at the center of the AGPM. The existence of a structure-forming base pair will stabilize the helix in which it appears and indirectly, will assist the folding and the proper positioning of the other helix.

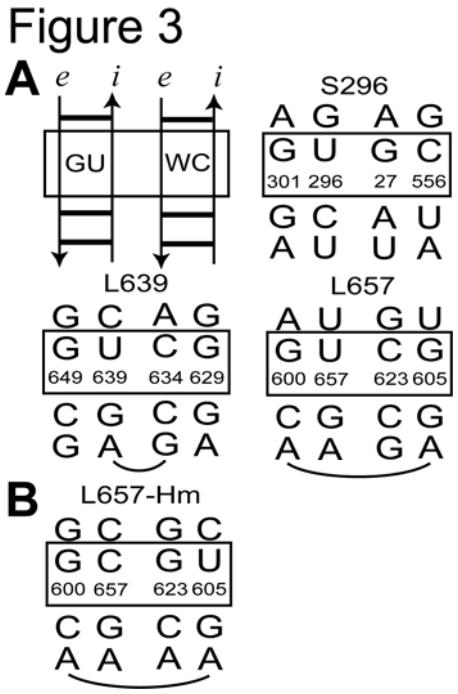


Figure 3. Nucleotide sequences of the three cases of AGPM considered in this study. In each case, the *E. coli* numbering is used.

(A) Nucleotide sequences of motifs S296, L639 and L657 from the *E. coli* ribosome (pdb entry codes 2avy-2aw4, (Schuwirth et al. 2005)). The secondary structures are drawn accordingly to the scheme shown in the upper left corner, in which the internal and external strands of both helices are marked by italic letters *i* and *e*, respectively. Boxed are the two central base pairs. In most known cases of AGPM, one central base pair is GU, while the other one is WC (Gagnon and Steinberg 2002). The name of each motif starts with either letter "S" or "L", depending on the ribosomal subunit, small or large, in which it is found, followed by the number of the internal nucleotide of the GU central base pair in the rRNA polynucleotide chain of the *E. coli* ribosome (Gagnon and Steinberg 2002).

(B) Nucleotide sequence of motif L657 (L657-Hm) from the *H. marismortui* ribosome (pdb entry code 1s72, (Ban et al. 2000)). Compared to the same motif in *E. coli* (panel A), the GU and WC base pairs in *H. marismortui* are exchanged in their positions.

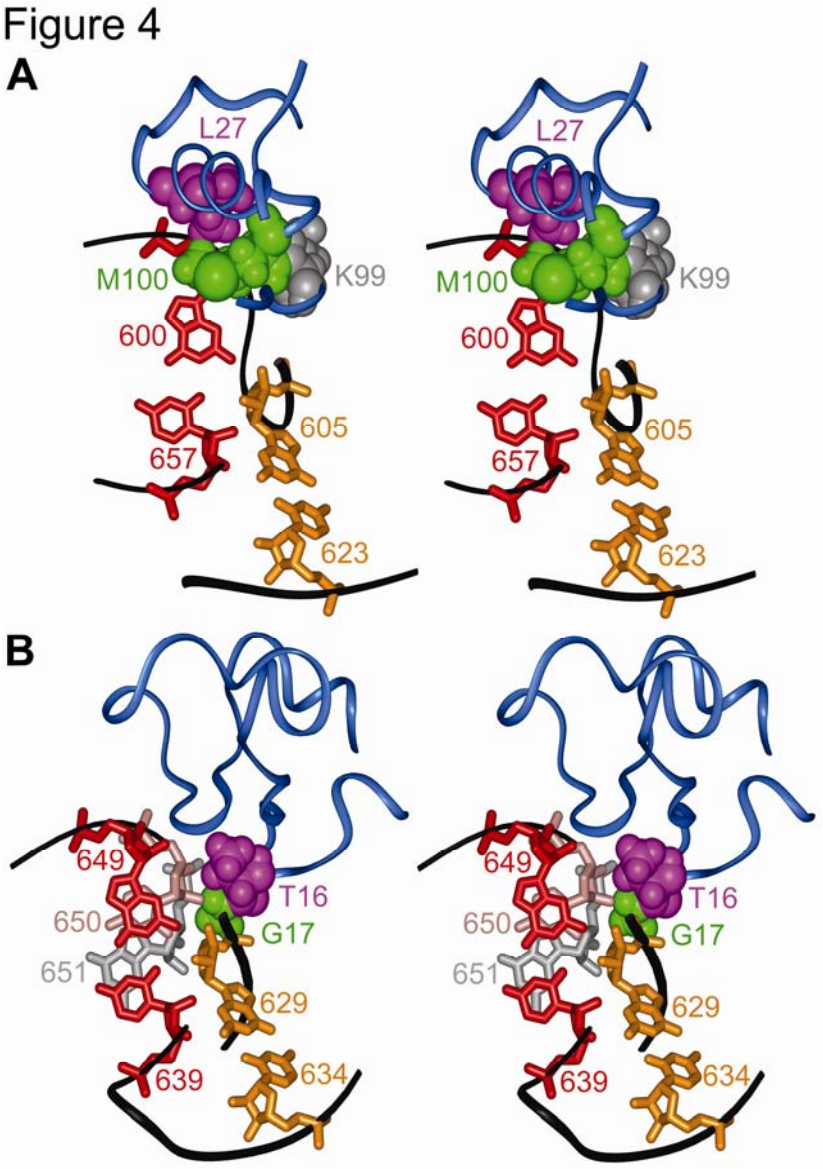


Figure 4. Stereo view of the structural contexts of motifs L657 (A) and L639 (B) taken from the *E. coli* ribosome.

(A) In motif L657, residues L27 (magenta), K99 (grey) and M100 (green) of ribosomal protein L4 (blue ribbon) tightly interact with the ribose and the backbone of nucleotide 600, which occupies the external position of a central base pair (red).

(B) In motif L639, residues T16 (magenta) and G17 (green) of ribosomal protein L35 (blue ribbon) interact with the sugar-phosphate backbone of nucleotides 650 (pink) and 651 (grey) proximal to the external nucleotide 649 of a central base pair (red).

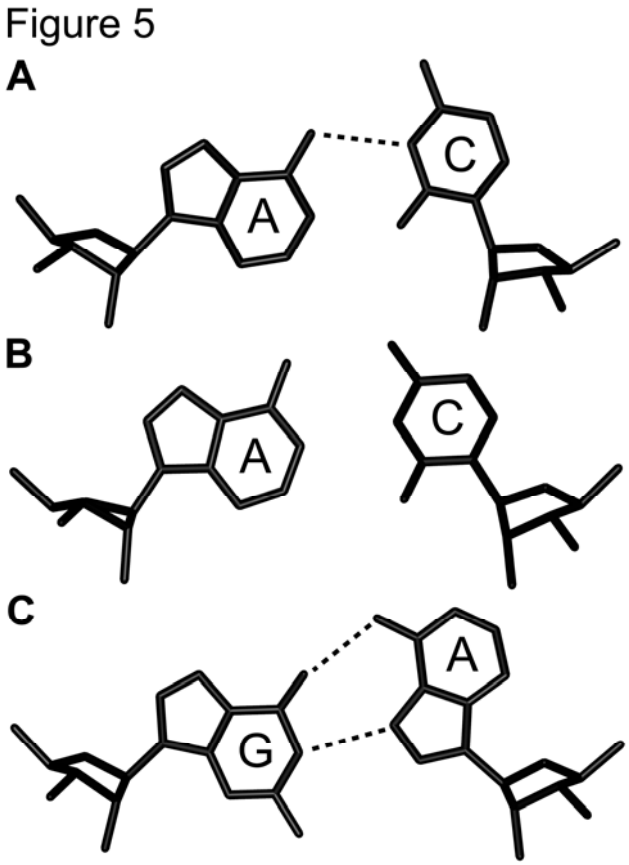


Figure 5. Juxtapositions of the bases in the AC and GA base pairs.

(A) The GU-like base pair A_{1005}-C_{963} found in the 50S subunit structure of *H. marismortui* ribosome (pdb entry code 1s72.pdb, (Ban et al. 2000)).

(B) The WC-like base pair A_{288}-C_{364} taken from the same structure as in (A). The same juxtaposition of A and C is observed in base pairs A_{597}-C_{560} and A_{2488}-C_{2534}.

(C) The proposed WC-like arrangement for the GA base pair in clones B3 and B6. The formation of such a GA base pair requires that the adenosine have the unusual *syn*-conformation.

FIGURE 6

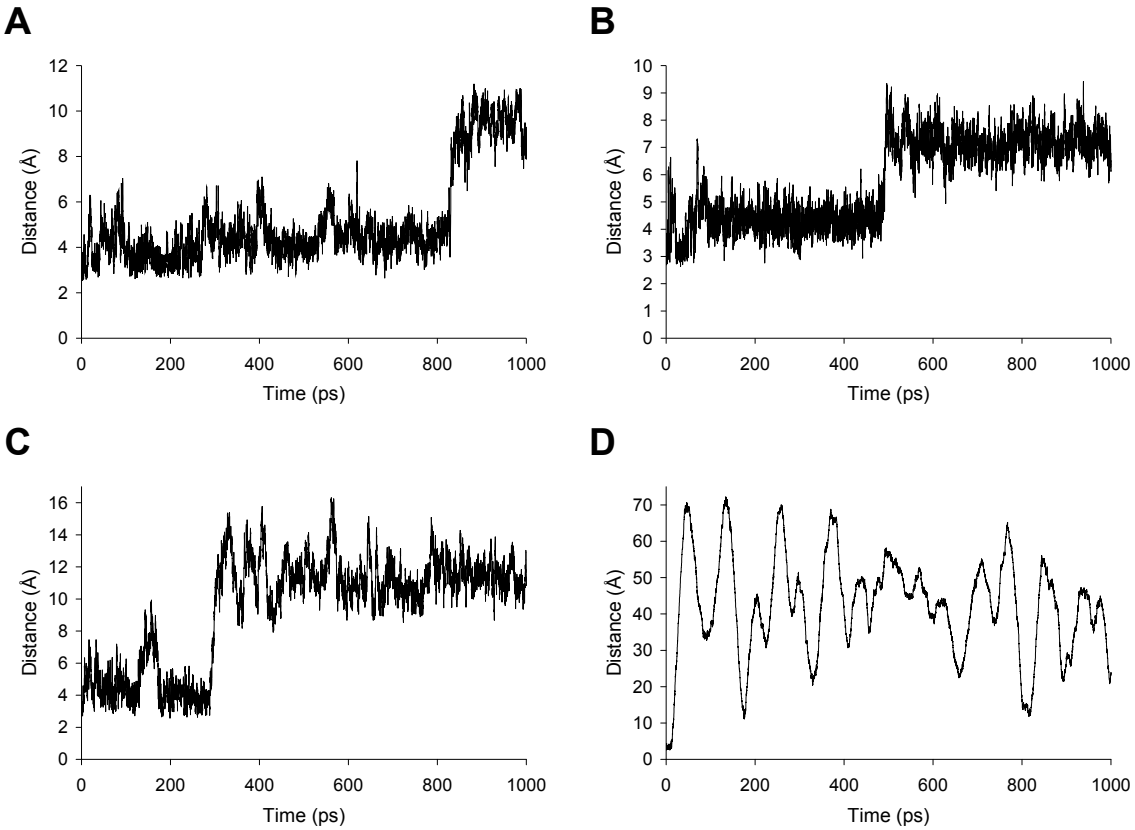


Figure 6. Molecular dynamics simulations of the AGPM structure containing different nucleotide triples. In the four complexes tested, the dinucleotide combinations occupying the central base pairs were GU-CG (A), GU-UG (B), GC-UG (C) and UG-UG (D). The CG, GC and GU combinations were initially arranged as normal base pairs, while in the UG combination, the internal guanosine initially occupied the position as in the standard AGPM structure, while the external uridine was bulged out. For each simulation, the contact between the internal riboses was monitored by following the distance between their O2´ atoms, which was plotted against the time.

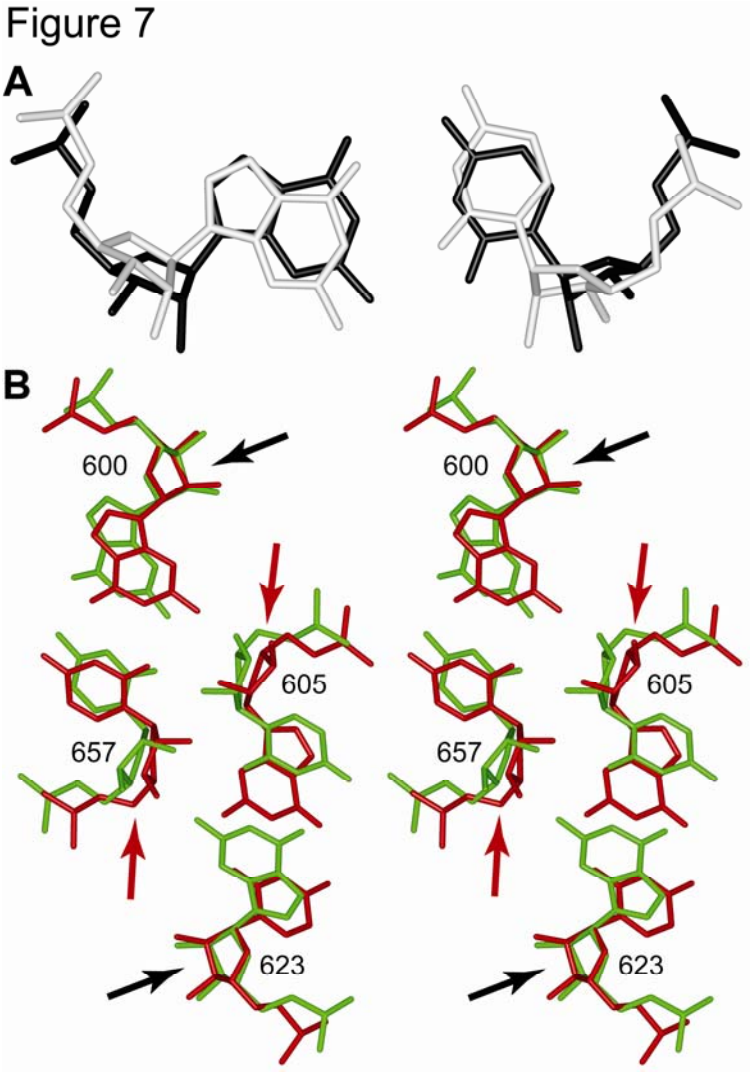


Figure 7. Conformational rearrangements associated with the GU⇔WC exchange of the central base pairs in the AGPM.

(A) The replacement of the GU base pair (white) by GC (black) rotates the base of the guanosine for about 15° toward the major groove. However small this displacement is, it can be large enough to damage the interaction between external nucleotide 600 and protein L4 and thus make the ribosome non-functional.

(B) Stereo view of the superposition of the two versions of motif L657 found in the structures of the *E. coli* (red, pdb entry code 2aw4, (Schuwirth et al. 2005)) and *H. marismortui* (green, pdb entry code 1s72, (Ban et al. 2000)) ribosomes. This superposition allows the visualization of the local conformational changes in the AGPM associated with the GU⇔WC replacement. The superposition was performed for base pairs 601-656 and 624-604 in both structures (not shown); it demonstrates that within the AGPM, the GU⇔WC exchange of the central base pairs affect the positions of the external riboses only slightly (black arrows), while the internal riboses become displaced substantially (red arrows). The immovability of the external riboses will thus preserve the interaction of the ribose-600 with protein L4 if motif L657 follows the GU-WC pattern. The *E. coli* nucleotide numbering is used for both structures.

4.10 Supplemental Data

4.10.1 Instant evolution versus natural evolution

To find out whether the specific characteristics of the particular AGPMs determined in the analysis of the selected clones can also be deduced from analysis of the naturally selected molecules, we checked for the presence of these characteristics in the available nucleotide sequences of prokaryotic rRNA (Wuyts et al. 2004). As one can see in the Supplemental Table I, in all three AGPMs (S296, L639 and L657), the GU-WC pattern is maintained at the level of 97% or higher. In almost all alternative sequences, one of the central base pairs is either WC or GU, which allows for the formation of the nucleotide triple. Among all 7 359 available nucleotide sequences of the three motifs, the base triple does not exist in only two. The extension of the analysis to all thirteen cases of AGPM found in the ribosome structure (Gagnon and Steinberg 2002; Mokdad et al. 2006) provides a total of 43 610 nucleotide sequences, of which the base triple cannot be formed in only 6 cases (Supplemental Table II). The fact that in almost all cases, one of the central base pairs is either WC or GU strongly supports the hypothesis that the formation of the nucleotide triple is indeed, a minimal requirement for the formation of the AGPM. Moreover, due to such a small number of exceptions, they may constitute artifacts originated from sequencing errors, misalignment or posttranscriptional modifications.

Further analysis showed that in the available nucleotide sequences of rRNA, both the L639 and L657 motifs were characterized by the same type of asymmetry between the two helices as was observed in the selected clones (Table I and Supplemental Table I). However, the total number of the sequences demonstrating this asymmetry was very modest. In particular, in the case of motif L639, helices H31 and H29 harbored the structure-forming base pair only six and zero times, respectively. In motif L657, helices H27 and H28 harbored such base pair only two and zero times, respectively. Interestingly, in three archaeal organisms, both central base pairs of motif L657 were GC. The coexistence of two central WC base pairs in motif L657 of the *E. coli* ribosome would lead to the displacement of nucleotide 600 from its preferred position, thus compromising its interaction with protein L4. The fact that such combinations of base

pairs are found in some archaea may indicate that in these organisms, the mode of interaction between motif L657 and protein L4 is different from other species.

In general, the naturally selected sequences of all three motifs follow the same structural rules that were determined through the analysis of the selected clones. However, due to the very strong evolutionary pressure toward the GU-WC pattern, the alternative cases account for only a very small fraction of all sequences. Moreover, whether such alternatives are artifacts remains unclear. It is also possible that in different organisms, the mode of interaction of a given motif with its surroundings is either slightly or considerably different. All these aspects argue against using naturally selected nucleotide sequences of rRNA instead of those obtained through library selection as a primary source of information for elucidation of structure-function relationships in the ribosome.

4.11 Supplemental Methods

4.11.1 Combinatorial gene libraries: primers and cloning

Oligonucleotide primers were synthesized with random nucleotides at the desired positions (Montreal Biotech Inc.). The first step involved PCR amplification of two individual fragments using the 16S or 23S rRNA gene as a template. The production of the 16S rRNA gene library for motif S296 was described previously (Gagnon et al. 2006). For the 23S rRNA gene libraries of motifs L639 and L657, respective sets of the primers were used as follow: pUC-1 & Reverse_U639; L1_A_HpaI_U639 & pUC-4 (library of motif L639 to be cloned in plasmid pKK1192U) and pUC-1 & L1_B_U605; L1_C_U605 & pUC-4 (library of motif L657 to be cloned in pKK1192U). Using the flanking primers pUC-1 and pUC-4, the entire 1541 bp region of motifs L639 and L657 was amplified and purified. This 1541 bp PCR product harbored the library that contained the randomized nucleotides of the motif. This PCR product (1541 bp) was cloned into the pKK1192U plasmid using the unique *SacI* and *BbvCI* restriction sites. For cloning the 23S rRNA gene libraries of the same motifs in the pLΔH1192U plasmid, we used the following set of primers: 23S-I & Reverse_U639; L1_A_HpaI_U639 & I-CeuI_REV (motif L639) and 23S-I & L1_B_U605; L1_C_U605 & I-CeuI_REV (motif

L657). Using the flanking primers 23S-I and I-CeuI_REV, the entire 2238 bp region of motifs L639 and L657 was amplified and purified. This 2238 bp PCR product was cloned into the pLΔH1192U plasmid using the unique *XbaI* and *I-CeuI* restriction sites. The sequences of the oligonucleotides that were used are shown in the Supplemental Table III.

For the purpose of this study, we used a modified version of plasmid pLΔH1192U that lacks a 466-long DNA fragment of the 23S rRNA gene operon between the restriction sites *SnaBI* and *BmgBI*. Due to this deletion, the gene becomes non-functional. The generation of this construct was possible due to the fact that the *rrnB* operon is transcribed from the λP_L promoter. The latter aspect allowed efficient repression of the *rrnB* operon in the POP2136 cells at 30ºC. In this way, we prevented wild-type 23S rRNA sequences from being present as contamination from the left-over undigested plasmids. Finally, plasmids harboring the combinatorial 23S rRNA gene libraries were transformed in the SQ380 cells by electroporation. All PCR reagents, Vent DNA polymerase, restriction enzymes and T4 DNA ligase were from New England Biolabs.

4.12 Supplemental Tables

Supplemental Table I. Presence of different tetra-nucleotide combinations as the central base pairs of AGPMs S296, L639 and L657 in the prokaryotic 16S and 23S rRNAs

Central base pairs		Number of sequences	%
Motif S296			
	GU-GC	6 254	96.33
	GU-CG	61	0.94
	GU-AU	9	0.14
	GU-UA	6	0.09
	GU-AC	27	0.42
	WC-WC	56	0.86
	GU-GU	35	0.54
	AC-GC	1	0.01
	BP-no	37	0.57
	no-BP	5	0.08
	no-no	1	0.02
	TOTAL	6 492	100
Motif L639			
	GU-GC	248	57.27
	GU-CG	137	31.64
	GU-AU	0	0
	GU-UA	34	7.85
	GC-GU	8	1.85
	BP-no	6	1.39
	no-BP	0	0
	no-no	0	0
	TOTAL	433	100
Motif L657			
	GU-GC	148	34.10
	GU-CG	256	58.99
	GU-AU	0	0
	GU-UA	17	3.92
	GC-GU	4	0.92
	AU-GU	3	0.69
	WC-WC	3	0.69
	BP-no	2	0.46
	no-BP	0	0
	no-no	1	0.23
	TOTAL	434	100

The data were obtained based on the available rRNA alignments (Wuyts et al. 2004). For the statistics, only those cases have been considered where the identities of all four nucleotides are known. BP stands for the structure-forming base pairs GU or WC. “no” stands for any other dinucleotide combination except GU, WC and AC/CA. For each base pair, the nucleotides appear in the same order as in Table I of the article.

Supplemental Table II. Presence of different tetra-nucleotide combinations as the central base pairs of all AGPMs in the prokaryotic 16S and 23S rRNAs

Central base pairs	Number of sequences	%
Closely packed (fitting to the GU-WC pattern)		
GU-GC	29 726	68.163
GU-CG	3 959	9.078
GU-AU	73	0.167
GU-UA	307	0.704
GU-AC	46	0.105
GU-CA	4	0.009
AC-WC	30	0.069
Sub-total	34 145	78.295
At least one structure-forming base pair		
WC-WC	8 977	20.584
GU-no	209	0.479
WC-no	207	0.475
GU-GU	66	0.151
Sub-total	9 459	21.689
Absence of a structure-forming base pair		
GG-CC	1	0.002
UU-CC	1	0.002
UU-GA	1	0.002
GG-GG	1	0.002
CU-UC	1	0.002
GA-UU	1	0.002
Sub-total	6	0.012
	TOTAL: 43 610	**100**

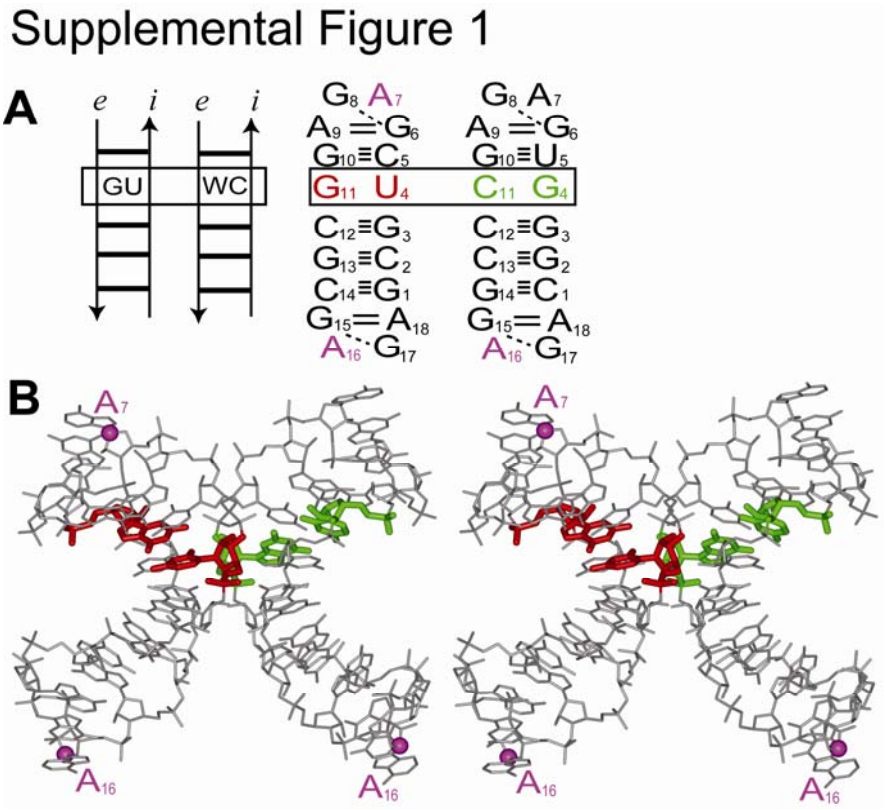
The data were obtained based on the available rRNA alignments (Wuyts et al. 2004). Cases of AGPM considered are: S62, S296, S549, S757, L554, L639, L657, L839, L1864, L2291, L2687, L2698 and L2847 (Gagnon and Steinberg 2002; Mokdad et al. 2006). For the statistics, only those cases have been considered where the identities of all four nucleotides are known. “no” stands for any dinucleotide combination except GU, WC and AC/CA. For the combinations shown in this table, the order of the two base pairs is not respected, such that GU always takes the first position regardless whether in a real nucleotide sequence it stays in the other position.

Supplemental Table III. Sequences of the oligonucleotides used in this study

Name	Sequence
23S-I	5'TCGTGTCCCCTTCGTCTAGAGGCCCAGG3'
pUC-1	5'AGCGTCTGGAAAGGCGCGCGATACAGGGTGACAGC3'
pUC-4	5'CCGGGCAGGCGTCACACCGTATACGTCCACTTTCGTG3'
I-CeuI_REV	5'GCAGGTCGGAACTTACCCGACAAGG3'
L1_B_U605	5'TCGGTTTCCCTTCG**N**CTCCCCTATTCGGTTAA**N**CTTG**N**TACAGAA TATAAGTCGCTGACC3'
L1_C_U605	5'G**N**TTAACCGAATAGGGGAG**N**CGAAGGGAAACCGAGTCTTAACT GGGCGTTAAG**N**TGCAGG3'
L1_A_HpaI_U639	5'GTAGCAAGGTTAACCGAATAGGGGAGCCGAAG**N**GAAA**N**CGAG**N** CTTAACTGG**N**CGTTAAG3'
Reverse_U639	5'CTTCGGCTCCCCTATTCGGTTAACCTTGCTAC3'

Bold N = A, G, T or C.

4.13 Supplemental Figure



Supplemental Figure 1. Modeled complex of the AGPM used in molecular dynamics simulations.

(A) Secondary structure of the modeled AGPM construct drawn according to the schematic representation on the left (*i* and *e* stand for the internal and external strands, respectively). The construct is based on the structure of motif L657 in the *E. coli* ribosome (pdb entry code 2aw4, (Schuwirth et al. 2005)) with some modifications. The GU (red) and CG (green) central base pairs are boxed. Each helix consists of 18 nucleotides and is capped by a GAGA tetraloop on both ends. Within base pairs, each solid line represents a distance constraint (H-bond), while a dash line stands for a distance constraint (H-bond) applied between a base and a ribose. Details of the constraints used are given in the Material and Methods. C1′ atoms of the magenta nucleotides were fixed during the simulations. The other three constructs, GU-UG, GC-UG and UG-UG were based on the same modeled complex of the AGPM (not shown).

(B) Stereo-drawing of the tertiary structure of the modeled AGPM construct shown in panel A. The colors are the same as in panel A. The three C1′ atoms whose positions were fixed during MD simulations are shown as magenta spheres. They are located within the GAGA tetraloops at the extremities of the helices. Their fixation helps to avoid uncontrolled deterioration of the construct at the regions outside the inter-helix contact. Due to this fixation, the helix that contains the green central base pair would dissociate from the helix that contains the red central base pair. When the conditions of the simulations are controlled in this way, it becomes possible to estimate the stability of the whole arrangement.

4.14 Supplemental References

Gagnon MG, Mukhopadhyay A, Steinberg SV. 2006. Close packing of helices 3 and 12 of 16S rRNA is required for the normal ribosome function. *J. Biol. Chem.* **281**:39349-39357.

Gagnon MG, Steinberg SV. 2002. GU receptors of double helices mediate tRNA movement in the ribosome. *RNA* **8**:873-877.

Mokdad A, Krasovska MV, Sponer J, Leontis NB. 2006. Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.* **34**:1326-1341.

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH. 2005. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**:827-834.

Wuyts J, Perriere G, Van De Peer Y. 2004. The European ribosomal RNA database. *Nucleic Acids Res.* **32**:D101-103.

Chapitre 5

Article 4

Structural rules for the formation of backbone-backbone interactions between closely packed RNA double helices

5. Structural rules for the formation of backbone-backbone interactions between closely packed RNA double helices

Matthieu G. Gagnon, Jianhong Chen and Sergey V. Steinberg

Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Manuscript in preparation.

Running title: Packing of RNA double helices

Contribution of each author:

Matthieu G. Gagnon: Participated in the *in silico* and data analysis, performed the statistical analysis and participated in the preparation of the manuscript and figures.

Jianhong Chen: Performed the search for additional cases of AGPM within the RNA-containing structure database.

Sergey V. Steinberg: Participated in the *in silico* and data analysis as well as in the preparation of the manuscript and figures.

5.1 Abstract

Although backbone-backbone interactions play an important role in the tertiary structure of large RNA molecules, the structural constraints imposed on their formation and the rules determining their integrity remain basically unknown. One RNA structural element for which the backbone-backbone interactions are essential is the along-groove packing motif. This motif is found in numerous locations within the ribosome structure; it consists of two double helices arranged such that the backbone of one helix is packed in the minor groove of the other helix and *vice versa*. Here we analyze the internal structure of the along-groove packing motif and the dependence of this structure's stability on the nucleotide sequence. We show that within the motif, the contact area between the two helices is mostly formed by riboses and in total involves twelve nucleotides. The proper positioning of the riboses that allows them to form these contacts is achieved through the strict choice of the identities of the base pairs involved. For different base pairs participating in the inter-helix contacts, the optimal identities can be Watson-Crick, GC/CG, or certain non-Watson-Crick base pairs. The proper choice of the base pairs provides for the stable inter-helix interaction. In some cases of the motif, certain base pairs are not optimal, which may reflect the fact that these motifs break and form during the ribosome function.

Keywords: RNA structure; along-groove packing motif; AGPM; ribosome; recurrent motif

5.2 Introduction

An essential part of the knowledge on RNA structure is accumulated in the form of recurrent structural motifs, which appear in the same or different molecules and have identical or very similar conformation[1-11]. The fact that such motifs can form in different structural contexts demonstrates a certain level of autonomy to their folding. Due to this autonomy, recurrent motifs play essential roles in the formation of tertiary structures of large RNA molecules. Therefore, analysis of the aspects that govern the formation of RNA recurrent motifs is important for understanding how RNA molecules fold and function. In most cases, analysis of the conditions for the formation of RNA motifs has been focused on specific interactions that involve nitrogen bases, while the role of the sugar-phosphate backbone has been largely ignored. However, when the backbone participates in the formation of the core of the arrangement, the role of the interactions formed by the backbone can no longer be ignored.

One such case is the so-called along-groove packing motif (AGPM)[12], which consists of two double helices closely packed *via* minor grooves in the way that a sugar-phosphate backbone of one helix interacts with the minor groove of the other helix and *vice versa* [Figure 1(a)]. The AGPM has been found in more than a dozen places in ribosomal RNA (rRNA), which makes this motif an important element of the ribosome architecture. Two more cases of AGPM are involved in the association of the P- and E-site tRNAs with the 50S subunit. The recurrence of the AGPM and the active involvement of the sugar-phosphate backbone in its formation make this motif an excellent model for studying the general role of the backbone in RNA structure formation. Although in recent years, several studies concerning different aspects of the AGPM formation have been reported[12-15], in none of them has the role of the backbone-backbone interactions been analyzed.

In this paper, we undertake a systematic analysis of the factors governing the interaction between the two double helices within the AGPM. This analysis is based on the available X-ray conformations of the AGPM, as well as on the collected data for more than sixty-five thousand available nucleotide sequences of the AGPM. The analysis demonstrates a very active role of backbone-backbone interactions in the shaping of the motif. We show here that in different parts of the motif, the nucleotide

identities are specifically tuned to provide for a stable collision-free interaction between the backbones of both helices. Because backbone-backbone interactions play an essential role in the formation of different RNA complexes, including the ribosome, the rules we discuss here are expected to be of general importance for RNA structure formation.

5.3 Results

5.3.1 Background: the general description of the AGPM

As mentioned above, the AGPM consists of two double helices arranged such that a sugar-phosphate backbone of one helix is packed along the minor groove of the other helix and *vice versa* [Figure 1(a)][12]. In each helix, the strand that interacts with the minor groove of the opposite helix is positioned close to the center of the arrangement and is thus called internal. The other strand in each helix stays at the periphery of the arrangement and is called external. The internal strands of both helices go in the same direction, opposite to that of the external strands. The conventional representation of the motif given in Figure 1a demonstrates the existence of symmetry between the two helices[12]. Indeed, within the AGPM, the position of each helix can be roughly determined based on the position of the other helix through the rotation for 180º around the common symmetry axis.

At the center of the contact area, there are two base pairs, called central, that pack with each other most closely. Despite the general symmetry of the AGPM, the packing of the central base pairs is essentially asymmetric. In most identified cases, one of these base pairs is Watson-Crick (WC), while the other one is GU. In the GU base pair, G and U stay, respectively, in the external and internal strand[12]. The arrangement of these base pairs shown in Figure 1(b) provides for a close contact between the helices and allows the formation of a network of several inter-helix hydrogen bonds. Such GU-WC arrangement at the center of the contact area is found in most known cases of the AGPM. Moreover, as recently pointed out by Mokdad *et al.*[15], the GU base pairs involved in the AGPM are among the most conserved GU base pairs in rRNA. Based on these findings, the coexistence of a WC and GU base pair in the middle of a helix-helix contact has been considered as a signature of the AGPM that could facilitate the

identification of new examples of the motif. In particular, this pattern enabled us to identify AGPMs formed by two tRNAs and the 23S rRNA[12].

Despite the almost universal presence of the GU-WC combination, some cases of the AGPM do not follow this pattern, which indicates the existence of other factors within the structure of the motif that are important for its integrity. The presence of such factors can also be deduced from the fact that in the AGPM, the area in each helix which is in contact with the other helix is not limited to one base pair but spreads over several consecutive base pairs. These additional contacts are largely from interactions between the backbones of both helices. In this paper, we perform a systematic analysis of these interactions in order to understand their role in the formation of the AGPM. This analysis is based on the comparison of the known conformations of the AGPM and the statistical analysis for the occurrence of particular nucleotides in different positions of the AGPM. The latter data were collected from the set of available nucleotide sequences of rRNA, which included 12 107 bacterial and 590 archaeal sequences of 16S rRNA as well as 399 bacterial and 37 archaeal sequences of 23S rRNA[16]. For the tRNA sequence analysis, the database containing 819 bacterial and 220 archaeal tRNA sequences was used[17].

5.3.2 Nomenclature of different elements of the AGPM

To facilitate the discussion of the inter-helix interactions within the AGPM, we will use the following nomenclature. For the four strands of these helices, capital letters P, Q, R and S are assigned as shown in Figure 2 (upper left corner). One helix is formed by strands P and Q, while the other one is formed by strands R and S. Strands P and R are external, while strands Q and S are internal. The helix containing GU as the central base pair in the *Escherichia coli* rRNA is composed of strands P and Q and is called the GU-helix (Figure 2). The opposite helix is called the WC-helix. For each base pair within each helix, a number is assigned, so that the central base pairs carry number zero, and the positive propagation of the numbering corresponds to the 5'→3' direction of the internal chains. In the identity of a base pair, the first and last letter will correspond to the external and internal position.

5.3.3 Collection of the set of the AGPM

The nucleotide sequences of all identified motifs are shown in Figure 2. The original set of 12 motifs[12] shown in Figure 2 (a-l) exists in all ribosome structures. In addition, Mokdad *et al.*[15] showed that motif L1864 [Figure 2(m)] exists in the bacterial 50S subunits of *Deinococcus radiodurans* (pdb entry codes 1kpj-1lnr[18]) and *E. coli* (pdb entry codes 2aw4-2awb[19]) but not in the archaeal 50S subunit of *Haloarcula marismortui* (pdb entry codes 1jj2-1s72[20]). For this work, we undertook an additional analysis of all ribosome-related crystal structures (see Methods) and found one more motif, S911 [Figure 2(n)], which is formed between helices h27 and h44 of the 16S rRNA as a result of a conformational rearrangement in the 30S subunit caused through its association with the initiation factor IF1 (pdb entry code 1hr0[21]).

Two more motifs, named L1923-P and L1851-E, are formed between the D-stem of the P-site tRNA and helix H69 of the 23S rRNA [Figure 2(p)] as well as between the acceptor stem of the E-site tRNA and helix H68 of the 23S rRNA [Figure 2(q)][12]. These motifs exist in two low-resolution structures of the *Thermus thermophilus* 70S ribosome (pdb entry codes 1gix-1giy[22] and 2ow8-1vsa[23], respectively). The high resolution ribosome structure from the same organism (pdb entry codes 2j00-2j01 and 2j02-2j03[24]) confirmed the presence of motif L1923-P, while the E-site tRNA in this structure was positioned differently.

Finally, a systematic analysis of all 1,131 RNA-containing structures in the PDB database[25] revealed a case of AGPM, named HH, which was formed by two hammerhead ribozyme molecules within the same asymmetric unit of the crystal [Figure 2(o)] (pdb entry code 1hmh[26]). So far, this case of AGPM has been the only one identified outside the ribosome. Recently, Jang *et al.*[27] announced that an additional case of AGPM was formed by two RNA octamer duplexes (pdb entry code 2g3s). However, visual inspection of the corresponding crystal structure did not support this conclusion, because the arrangement in question was found to be substantially different from all other known AGPMs.

5.3.4 Principles of helix packing within the AGPM

The attractiveness of the AGPM as an element of RNA architecture relates to its ability to partly neutralize the repulsion between the negatively charged phosphates of the two interacting helices. Indeed, when the internal strand of one helix packs in the minor groove of the other helix, its negatively charged phosphate groups could interact with the positively charged groups of the second helix that are exposed to the minor groove. Due to the general reciprocity of the AGPM, the same type of interaction would also occur between the phosphates of the second helix and the positively charged groups positioned in the minor groove of the first helix. Because positive charges in RNA are mostly associated with amino groups and also because the only standard base in which the amino group is exposed to the minor groove is guanine, the double helices involved in the formation of the AGPM are expected to have a high proportion of G-containing base pairs. Indeed, in all motifs shown in Figure 2 both central base pairs are GU, GC or CG. An example of the interaction between the central base pairs is shown in Figure 1(b), where the amino group of the guanosines forms a hydrogen bond with the O3′ atom of the nucleotide from the internal strand of the opposite helix.

Regardless of how effective this interaction is, a major limitation is the fact that it cannot spread for more than one base pair in each helix. Indeed, because of the spiral character of both helices, the juxtaposition of the base pairs at each level is different. Only at the 0-level, the arrangement of the base pairs is such as it is shown in Figure 1(b), while even at the neighboring +1- and -1-layers, it is so different that it is no longer possible to describe it as a packing of a backbone of one helix in a groove of the other helix. It does not mean, however, that outside the 0-level the two helices do not interact. On the contrary, analysis of the available AGPM conformations shows that the inter-helix contacts spread over four layers between -2 and +1. While at the 0-level, these contacts include bases, at the other levels contacts are mainly formed by elements of the backbones. Most of the backbone contacts are formed by riboses and are thus mainly hydrophobic.

Analysis of the AGPM structure shows that the whole contact area can be divided in three zones, depending on the particular strands involved in the inter-helix contacts. The first zone, named QS, corresponds to the interaction between chains Q and

S. The Q- and S-moieties of this zone are mainly formed by the riboses of nucleotides +1Q and +1S, but also include some atoms of the neighboring nucleotides 0Q and 0S (Figure 3 and Table I). The second zone QR, formed by chains Q and R, mainly consists of the contact between riboses -1Q and -2R, but also includes some atoms of 0Q and -1R. The third zone PS is symmetrical to zone QR with strands P and S being equivalent to R and Q, respectively. A complete list of the atoms participating in the formation of the three contact zones is given in Table I. On the surface of each helix, these zones form a triangle with the vertices positioned at the riboses of the internal +1- and -1-nucleotides as well as of the external -2-nucleotide [Figure 3(b)]. The interaction of the two helices can thus be seen as superimposition of the triangle in one helix on the equivalent triangle in the other helix.

Comparison of the known AGPMs shows that the three contact zones are preserved in all cases regardless of the presence of other features and are thus considered important for the integrity of the motif. The existence of the three contact zones depends on the particular positions of the riboses involved and is thus sensitive to the structures of the base pairs of which these riboses are a part. In the following sections, we will show how the system of backbone-backbone contacts shapes the AGPM and how it restricts the identities of the essential base pairs.

5.3.5 The inter-helix interactions in contact zones QR and PS

5.3.5.1 The central role of the -1-base pairs

We start the analysis of the inter-helix contacts with those that are formed within the QR and PS contact zones. As one can see from Figure 3 and Table I, among all base pairs participating in the formation of the AGPM, only in the -1-base pairs are both nucleotides involved in inter-helix backbone-backbone contacts. In particular, the internal nucleotide -1Q forms a major part of the Q-moiety of zone QR, while its external base pair partner -1P participates in the formation of the P-moiety of zone PS. The same is true for nucleotides -1S and -1R with respect to zones PS and QR. The necessity for proper fitting of all four -1-nucleotides to the inter-helix contacts should impose strong restrictions on the structure of the -1-base pairs, and indeed, these base

pairs are WC in all known conformations of the AGPM. Moreover, statistical analysis of the available nucleotide sequences from prokaryotes of AGPMs existing in the ribosome shows that the WC identities of the -1-base pairs are preserved in all motifs at the average level of 98% (Table II)[16].

5.3.5.2 The position of WC -2-base pairs

Unlike the -1-base pairs, the -2-base pairs internal nucleotides -2Q and -2S are not involved in inter-helix contacts (Table I). Therefore, it is not surprising that the -2-base pairs harbor a variety of different structural forms, in contrast with the -1-base pairs. In particular, in the 17 cases of AGPM shown in Figure 2, there are 34 potential -2-base pairs, which include 16 WC base pairs, 8 non-WC base pairs, and 10 cases where the two nucleotides do not form a base pair. For convenience, we will first analyze motifs S296, S757, L554 and L2698, in which both -2-base pairs are WC. In these motifs, the two -1- and two -2-base pairs of both helices form a four-base pair arrangement seen in Figure 4 (a,b). All cases of this arrangement are superposable with r.m.s.d. 0.56 Å. Also, each arrangement is reciprocal in the sense that it can be superposed with its own image rotated for 180º around the symmetry axis. At the center of contact zone QR, the five-member rings of two riboses -1Q and -2R are closely packed with each other. In total, between the Q and R strands there are about 15 atom-atom van der Waals contacts in which two non-hydrogen atoms are positioned within 4.2 Å of each other. Due to the symmetry of the arrangement, analogous interactions occur in the PS contact zone between nucleotides -2P and -1S.

In three additional ribosome-related motifs S62, L2291 and L2687, as well as in motif HH, only one of the two -2-base pairs is WC. The superposition of these motifs with those in which both -2-base pairs are WC shows that the WC base pair of the -2-level always stays at the same position regardless of the structure or the existence of the other -2-base pair. This positioning of the WC base pair allows for the formation of the same interaction between its external nucleotide (-2P or -2R) and the internal nucleotide of the opposite -1-base pair (respectively, -1S or -1Q).

5.3.5.3 The position of non-WC -2-base pairs

As mentioned above, in most motifs shown in Figure 2, the -2-base pairs are either non-WC or do not exist at all. The non-WC base pair can be UG (3 times), GA (3), AA (1) or UU (1). Analysis shows that despite the variety of structural forms accepted by two nucleotides at the -2-level, the external nucleotide always stays at about the same position regardless of the nature of the interaction, if any, it forms with the internal nucleotide [Figure 4(c)]. This position allows the ribose of the external -2-nucleotide to preserve the close interaction with the ribose of the internal -1-nucleotide from the opposite helix. Interestingly, compared to the WC base pairs, the external -2-nucleotide in the non-WC combinations is slightly over-twisted, so that the WC edge of its base becomes more open to the major groove [Figure 4(c)]. This displacement of the nucleotide brings its O2′ atom closer to the O2′ atom of the internal -1-nucleotide from the opposite helix for the formation of the inter-ribose hydrogen bond. This hydrogen bond does not exist when the -2-base pair is WC, and it is found in most cases when this base pair is either non-WC or non-existent. Analysis of the known -2 non-WC base pairs shows that they cooperate with the formation of this hydrogen bond. As mentioned above, the formation of this hydrogen bond requires a displacement of the external -2-nucleotide toward the major groove. Correspondingly, the structures of all four non-WC base pairs found at the -2-level are such that they support such displacement.

5.3.5.4 Statistical analysis of the identities of the -2-base pairs

Statistical analysis of the nucleotide sequences of prokaryotic rRNA[16] shows that for those -2-dinucleotide combinations that correspond to a base pair in the ribosome tertiary structure, WC combinations as well as combinations GA, UG, AA and UU occur in 99% of all cases (Table III). More specifically, base pair GA is almost 600 times more frequent than AG, while base pair UG is almost six times more frequent than GU. All predominant non-WC combinations can support the displacement of the external -2-nucleotide in the direction of the major groove. These statistical data demonstrate that the nucleotide sequences of rRNA are specifically tuned toward the formation of the particular type of inter-helix contacts, which includes ribose-ribose interactions and specific hydrogen bonds. The existence of such tuning supports the importance of these

contacts for the integrity of the AGPM. The symmetry and superposability of the arrangement of the -1- and -2-base pairs in different cases of AGPM allows us to look at this arrangement as stable and rigid.

5.3.6 The nature of the asymmetry between the two helices within the AGPM

5.3.6.1 Potential collisions of the helices at the 0- and +1-levels

Based on the juxtaposition of the two double helices fixed by the inter-helix interactions in contact zones QR and PS, one can foresee the type of inter-helix interaction in zone QS (Figure 3 and Table I). A simple *in silico* experiment would consist in the extension of both helices in the A-RNA conformation from layer -1 to layers 0 and +1. The results of this experiment show that the two double helices arranged in this way would collide with each other at both layers 0 and +1 [Figure 5 (a,b)]. At the 0-level, the two nucleotides 0Q and 0S are positioned closer to each other by about 1 Å over the distance considered acceptable for tight packing. The collision occurs between the ribose of each internal nucleotide and the ribose and the base of the opposite internal nucleotide [Figure 5(a)]. At the +1-level, a similar situation happens with riboses +1Q and +1S, which penetrate each other by about 1.5 Å [Figure 5(b)]. Below we show how the potential collision between the two internal strands within contact zone QS is avoided through the tuning of the nucleotide sequences of both helices.

5.3.6.2 The adjustment of base pair [0P;0Q] to the potential collision with nucleotide 0S

The fact that the extension of both helices from layer -1 to 0 in the standard A-RNA conformation leads to their collision strongly suggests that at least in one of the two helices the conformation is not standard. A deviation from the standard conformation would displace nucleotides 0Q and 0S farther from each other, thus allowing them to avoid the collision. To check whether such deviations really take place, we compared the predicted juxtaposition of the two helices with the real ones existing in those motifs in which both 0-base pairs are WC (Figure 2). The latter motifs include S549 in all known structures of the 30S subunit (Table IV), L2291 in the structure of the

50S subunit from *H. marismortui* and motif HH. Analysis shows that while base pair [0R;0S] in the real and predicted structures occupies about the same location, the position of base pair [0P;0Q] in the predicted structure is notably different from that in the real ones. Compared to the prediction, nucleotide 0Q in the real structures is over-twisted by about 10º and is displaced toward the major groove by about 1 Å [Figure 5(c)], which is sufficient for the elimination of its collision with 0S.

For the maintenance of base pair [0P;0Q], the position of 0P should be adjusted to the displacement of 0Q. Indeed, the position of 0P also becomes displaced in the direction of the major groove by about 1 Å in the real structures [Figure 5(c)]. Such displacement creates a crack in the arrangement of the two central base pairs between nucleotides 0P and 0S clearly seen in all known structures of the AGPM in which both 0-base pairs are WC.

5.3.6.3 The GU-WC pattern for the central base pairs

The displacement of base pair [0P;0Q] described above, which eliminates the collision between 0Q and 0S, would make the geometry of the GU-helix no longer optimal, causing tension in both strands. While in strand Q, due to the potential collision with strand S, this tension seems to be unavoidable, in strand P it could be relaxed through the introduction of a non-WC base pair [0P;0Q]. Indeed, in most cases of AGPM presented in Figure 2, base pair [0P;0Q] is GU, which allows nucleotide 0P to move back to the minor groove without movement of 0Q [Figure 5(d)]. Thus, the introduction of the GU base pair represents a way to relax the conformational tension in strand P that emerged from the displacement of 0P when its position was adjusted to the displacement of 0Q. Statistical analysis of the available nucleotide sequences of rRNA[16] shows that the GU base pairs involved in the AGPM are conserved, on average, at the level of 97% (Table IV), which positions them among the most conserved GU pairs existing in rRNA[15].

The new position of base pair [0P;0Q] is also stabilized by five inter-pair hydrogen bonds [Figure 1(b)]. Two of these bonds are formed by the amino group of 0P and depend on the GU identity of base pair [0P;0Q]. One hydrogen bond is formed by the amino group of the guanine from base pair [0R;0S]. This bond can exist only if the

latter base pair is either GC or CG, which is the case in, respectively, 69% and 24% of all nucleotide sequences of AGPM found in rRNA (Figure 2 and Table IV). Two more hydrogen bonds are formed by atom O2′ of nucleotide 0Q. These latter bonds do not depend on the nucleotide identities and are present in all cases of AGPM.

5.3.6.4 The asymmetry between the -1-base pairs as a consequence of the asymmetry between the 0-base pairs

While at levels -1 and -2 the two helices are arranged in a symmetrical way, the displacement of 0Q caused by the necessity to avoid the collision with 0S breaks this symmetry. Interestingly, the resulting asymmetry of the two double helices touches also the -1-level. We found that in all the AGPM structures, an additional hydrogen bond is formed between atom O4′ of ribose 0S and the amino group of the guanine of base pair [-1P;-1Q] (Figure 6). The symmetrical bond between atom O4′ of ribose 0Q and the amino group of the guanine of base pair [-1R;-1S] does not exist due to the previously discussed displacement of 0Q farther from 0S. Because the hydrogen bond between ribose 0S and base pair [-1P;-1Q] involves guanine, its formation requires that the latter base pair be either GC or CG. At the same time, because the opposite hydrogen bond does not form, the identity of base pair [-1R;-1S] would remain unrestrained and could accept all possible WC combinations. Analysis of the motifs presented in Figure 2 shows that both base pairs [-1P;-1Q] and [-1R;-1S] follow the suggested pattern. Indeed, in all motifs base pair [-1P;-1Q] is always either GC or CG, while base pair [-1R;-1S] has the GC/CG identities only in eleven cases, and in the other six cases it is either AU or UA. Statistical analysis of the available nucleotide sequences of rRNA[16] shows that although for both -1-base pairs the GC/CG identities are predominant, the level of their conservation is different. In particular, while for base pair [-1P;-1Q] the GC/CG identities account for 98.5% of the cases containing two WC base pairs at the -1-level, for base pair [-1R;-1S] this number reaches only 73% with the remaining 27% taking the AU/UA identities (Table II). These data demonstrate the importance of the hydrogen bond between base pair [-1P;-1Q] and the ribose of 0S for the integrity of the whole arrangement. They also show that the asymmetry between the two double helices is not

limited to the 0-base pairs, but spreads to the -1-level as well, even though the two -1-base pairs are arranged symmetrically to each other.

5.3.7 Ribose-ribose interaction in contact zone QS

As mentioned above, the extension of both double helices from -1 to the upper levels in the standard A-RNA conformation leads to the collision between the internal nucleotides not only at level 0, but also at level +1. At this level, the collision occurs between the riboses of nucleotides +1Q and +1S. In order to avoid this collision, at least one of the two nucleotides should move farther from the opposite nucleotide, i.e. in the direction of the major groove. Like at the 0-level, such displacement can be facilitated by the introduction of a non-WC base pair.

5.3.7.1 Introduction of a non-WC base pair

In nine out of seventeen AGPMs shown in Figure 2, one +1-base pair is WC, while the other one is not. These nine motifs form the so-called WC-non-WC group. Among the non-WC base pairs, one can find AG (6 times), AC (2) and GU (1). All AG base pairs represent sheared arrangements, the GU base pair forms in the standard way, while the AC base pairs are arranged such that the amino group of the adenosine forms a hydrogen bond with atom O2 of the cytidine (not shown). In comparison with the WC geometry, the internal nucleotide is displaced toward the major groove in all these base pairs. Such displacement allows the two internal nucleotides (+1Q and +1S) to avoid the collision and to form a comfortable ribose-ribose contact. Among the nine motifs of the WC-non-WC group that are present in Figure 2, eight are completely nested in rRNA. Analysis of the available nucleotide sequences of these motifs revealed a very strong bias toward those non-WC base pairs in which the internal nucleotide is displaced in the major groove. In particular, while base pairs AG and GU were detected in, respectively, 22 726 and 558 cases, their mirror combinations GA and UG were virtually non-existent, being found in, respectively, 0 and 1 time. The same sequence analysis revealed a substantial number of examples in which the non-WC base pair at the +1-level was AA or AU. The AA combination can form a sheared base pair similar to AG, while the AU combination can be arranged as the AC base pair discussed above. Such arrangements fit

these nucleotide combinations to the common pattern. In total, the favorable inter-helix interactions within contact zone QS were maintained in all WC-non-WC motifs in prokaryotic rRNA at the level of 98.7% (Table V).

Two more motifs, S296 and L1864, show a clear similarity to motifs of the WC-non-WC group. In both motifs, one +1-base pair exists while the other one does not (Figure 2). In S296, base pair [+1P;+1Q] is AG, while the opposite base pair [+1R;+1S] does not exist. Despite the absence of this base pair, the unpaired nucleotide +1S occupies the same position that it would have occupied if a WC base pair [+1R;+1S] existed. In L1864, the situation is the opposite. Here, the existing base pair [+1R;+1S] is WC, while the unpaired nucleotide +1Q is displaced in the direction of the major groove as if it formed a GU or AG base pair [+1P;+1Q]. As a result, in both motifs, the interaction between the internal nucleotides at the +1-level is preserved as in the motifs of the WC-non-WC group.

5.3.7.2 The WC-WC pattern at the +1-level

Among AGPMs shown in Figure 2, there are six motifs in which both +1-base pairs are WC. These motifs form the so-called WC-WC group. Analysis of the motifs of this group shows that the collision between the +1-base pairs is avoided through a deviation of one helix from the standard A-RNA geometry. The deviation involves the displacement of the internal nucleotide in the direction of the major groove and the appropriate position adjustment of the corresponding external nucleotide (Figure 7). Although such displacement is expected to create tensions in the conformation of the double helix, the fact that it occurs in about a third of all known AGPMs demonstrates that the level of conformational tension in this case should be acceptable.

However strong these conformational tensions really are, in many organisms they can be essentially relaxed. In particular, out of the six WC-WC motifs shown in Figure 2, four are found within rRNA. Analysis of the nucleotide sequences of these motifs shows that in three of them, S911, L839 and L2698, a substantial number of the nucleotide sequences contain at the +1-level a combination of the GU and WC base pairs (Table V), which will effectively relax the conformational tensions associated with the WC-WC situation. The only obvious exception of this pattern pertains to motif S549.

Not only does it contain WC-GU cases, but also, in archaea, more than 80% of its sequences have the combination WC-UG (Table V). The presence of a UG base pair, in which the internal nucleotide is expected to move in the direction of the minor groove, can only aggravate the conformational tension within the motif, and is not observed at the +1-level in any other AGPM. Although, as our modeling experiment show, such a base pair could still be arranged within the AGPM, its stability is expected to be relatively low. A possible functional implication of this feature of motif S549 is discussed later.

To conclude this part, we can say that the contact between nucleotides +1Q and +1S exists in all cases of AGPM and is thus considered important for the integrity of the motif. In most cases, the identities of the +1-base pairs are specially tuned to each other in order to provide for a stable inter-helix interaction within contact zone QS. In other cases, the formation of this interaction can cause some conformational tensions, which would make the motif less stable.

5.3.8 The consensus secondary structure of the AGPM

Based on the analysis of different interactions within the AGPM we can suggest a consensus secondary structure to which most known motifs fit (Figure 8). In this consensus structure, the identities of base pairs are divided in three categories of those that are satisfied in practically all nucleotide sequences, those for which a strong preference exists in the analyzed AGPM structures, and those that are generally acceptable and appear at a notable level. This consensus structure can be used for the search of new cases of AGPM in RNA molecules for which the tertiary structure is yet unknown.

5.3.9 tRNA fixation in the P and E sites of the 50S ribosomal subunit

As mentioned above, two AGPMs are formed between 23S rRNA and two tRNAs positioned in the P and E ribosomal sites (Figure 2)[12]. In the P-site, the D-stem of the tRNA is closely packed with helix H69[22,24], while in the E-site, the tRNA acceptor stem is packed with helix H68[22]. Analysis shows that these motifs are accountable for most van der Waals contacts and hydrogen bonds formed between each of the two

tRNAs and 23S rRNA outside the peptidyl-transferase center. The knowledge of the nucleotide sequence requirements for the AGPM formation can provide the information on universality and stability of these interactions.

Inspection of the structure of the two tRNA-rRNA motifs shows that in both of them, the GU-helix belongs to rRNA, while the WC-helices come from the tRNA molecules. Such arrangement of AGPMs guarantees that this interaction can be formed by most tRNAs. Analysis of the available nucleotide sequences of 23S rRNA[16] shows that in both helices H68 and H69 the GU base pair is extremely conserved among all prokaryotes. In fact, in all 399 bacterial and 37 archaeal sequences checked, there have been no exception when a GU base pair was not found at the appropriate place in either helix H68 or H69. As to the corresponding central tRNA-based base pairs 23-12 and 2-71, among the 1039 tRNA sequences tested[17], they are conserved as WC at the level of 98.2% and 97.9%, respectively.

At the other levels, most nucleotide sequences also fit to the general pattern, although the presence of particular elements in these two motifs would allow us to suggest that both L1923-P and L1851-E are somewhat less stable than most previously discussed motifs. In particular, in both helices of motif L1923-P, most +1-base pairs are WC, which indicates the existence of some level of conformational tension within the motif. The -1-base pairs in all cases are almost exclusively conserved as WC, except in the WC helix of motif L1851-E, where the level of conservation is below 90% mostly due to $tRNA^{Ala}$ in which the universally present base pair G3-U70 serves as a major identity element for the cognate aminoacyl-tRNA synthetase[28,29]. Interestingly, although the proportion of AU/UA base pairs in the -1 positions of the GU-helices is lower than in the WC-helices, it is much higher than in the motifs completely nested in rRNA. As mentioned above, the presence of an AU/UA base pair in position [-1P;-1Q] will not allow the formation of the hydrogen bond between this base pair and atom O4′ of nucleotide 0S, which existed in all previously discussed motifs. The presence of such weaknesses in the structures of both rRNA-tRNA motifs could reflect their transient status in the sense that both motifs must break and form *de novo* each elongation cycle. Because of such status, a too high stability of these motifs could be harmful for the ribosome function.

5.4 Discussion

In this paper, we analyze the principles of helix-helix interaction within the AGPM, which constitutes an important structural element of the ribosome architecture. Until now, the nucleotide sequence requirements for the AGPM formation have been thought to be limited to the co-existence of the GU and WC central base pairs. The fact that the presence of two central WC base pairs in some cases of AGPM does not interfere with the formation of the motif has been largely ignored. For the first time, we demonstrate that the AGPM cannot be reduced to the interaction of the central base pairs. The contact area spreads over four base pairs in both helices and includes 12 nucleotides in total. The additional interactions are very important: they are primarily responsible for maintening the juxtaposition of the helices and are able to keep the integrity of the motif even when the central base pairs do not follow the GU-WC pattern.

A specific characteristic of the AGPM is the fact that in this motif, most inter-helix interactions occur between the backbones of two helices. Based on the particular strands involved in the inter-helix contacts, we identified three contact zones QS, QR and PS, and in each zone, the contacts are mainly made by two riboses, respectively, [+1Q;+1S], [-1Q;-2R] and [-2P;-1S]. These interactions are mainly hydrophobic, although hydrogen bonds could also be formed. The ribose-ribose interactions exist in all cases of AGPM and thus constitute an essential aspect of the motif. In each helix, the three riboses form a triangle, which is superimposed on the corresponding triangle from the opposite helix. The interaction of the two triangles guides the formation of the whole arrangement and provides rigidity.

The fact that most inter-helix contacts within the AGPM are made by the backbones does not mean that the identities of the bases are not important. On the contrary, in order to allow the simultaneous inter-helix interaction within all contact zones, the shapes of both helices must be tuned to each other. This tuning proceeds through the precise selection of the identities of all base pairs participating in the inter-helix contacts, which would enable the particular positioning of the riboses. At the -1-level, all four nucleotides are involved in the inter-helix contacts; the possibility for formation of these interactions is guaranteed by the high conservation of the WC

identities of both -1-base pairs in examples of this motif. At levels +1 and -2, only one nucleotide in each base pair is involved in the inter-helix contact. Correspondingly, the identities of the +1- and -2-base pairs are such that they are able to provide the proper position of this nucleotide with respect to the nucleotide from the opposite helix with which it interacts.

An important result of this paper consists in the understanding that layers -1 and -2 and layers 0 and +1 play essentially different roles in the integrity of the AGPM. The formation of the inter-helix interactions at layers -1 and -2 provides the scaffold to which the interactions at levels 0 and +1 should fit. Indeed, the geometry of the inter-helix packing of layers -1 and -2 is close to optimal. When all -1- and -2-base pairs are WC, the two helices have the standard A-RNA conformation and form extensive interaction with each other within contact zones PS and QR. When one or both -2-base pairs are non-WC, the external nucleotides of these base pairs are still able to maintain the interactions with the opposite helix. The introduction of non-WC base pairs at the -2-level is thus not dictated by the internal logic of the AGPM and seems to be required for a better accommodation of the motif to its immediate surroundings. However, at the levels 0 and +1 the situation is different. The extension of both helices up from layer -1 in the regular A-RNA conformation leads to their collision at levels 0 and +1. A way to avoid this collision would consist in a deformation of the optimal geometry in at least one helix, which can be relaxed through the introduction of non-WC base pairs at both levels. Thus, for levels 0 and +1, unlike for level -2, the presence of non-WC base pairs is mainly determined by the internal logic of the AGPM and only to a lesser extent by the interaction with surrounding regions.

The importance of the ribose-ribose interactions in maintaining the integrity of the AGPM downplays the role of the central base pairs. We show here that the presence of the GU-WC pattern at the zero level is neither necessary nor sufficient for the formation of the AGPM. In fact, the role of the GU base pair in position [0P;0Q] to a great extent is limited to providing a way for relaxation of the conformational tensions caused by the displacement of nucleotide 0Q from its regular position due to the potential collision with 0S. Even without this relaxation the motif can still form, although in this case, it is expected to be less stable. The collision between 0Q and 0S

originates from the optimal helix-helix interactions at levels -1 and -2 and is strongly dependent on the WC identity of the -1-base pairs. In other words, the importance of the central base pairs seems to be secondary compared to that of the -1-base pairs.

WC-WC combinations rarely occur at the 0-level, while at the +1-level they are found in about a third of all motifs and are thus more acceptable than at the 0-level. However, even here, the negative effect of such a combination is obvious, as one can judge from the fact that in all such cases the optimal geometry of one of the helices is distorted between layers 0 and +1. Most probably, the presence of the distortion results in a lower stability of the motifs, which is not necessarily harmful to the ribosome function. One can expect that the ribosome function requires that some motifs break at particular moments of the functional cycle and that a lower stability of the motif would be helpful for effective breakage. Interestingly, two of the five motifs in Figure 2 that are associated with rRNA and contain a WC-WC combination at the +1-level have already been known to break during ribosome function. In particular, motif S911 is expected to break and form *de novo* at the initiation of translation[21], while motif L1923-P breaks and forms during the ribosome translocation. A lower stability of these motifs could facilitate the dissociation of the helices when it is functionally required. Another motif of this group, S549, has different kinds of abnormalities. In bacterial rRNA, it has WC-WC combinations at both levels 0 and +1. In archaeal rRNA, it contains a UG base pair at the +1-level. The presence of such abnormalities in S549 indicates that it could play a specific role during ribosome function. Whether this is true or not would require additional experimental support.

The phenomenon of shaping the backbone through formation of particular base pairs has a high probability to be observed not only in the AGPM, but in other RNA arrangements as well. Given that in the ribosome and in other RNA-containing complexes there are many occasions when the RNA backbone is involved in essential inter- or intra-molecular interactions, the situations when the shape of the backbone and its ability to participate in these interactions are regulated by the formation of particular base pairs are expected to be very common[30]. The understanding of the fundamentals of RNA structure and folding will require a systematic consideration of all such cases.

5.5 Methods

The alignments of nucleotide sequences of ribosomal RNA were taken from the European ribosomal RNA database[16]. The statistical analysis was performed using specially designed software, which allowed the simultaneous following of four nucleotide positions and counted the number of times a particular four-nucleotide combination was found. Only the sequences for which the identities of all four nucleotides are known were taken for analysis.

For identification of tertiary arrangements similar to those existing in a prototype structure, we took two nucleotides of the structure that played a potentially important role in its integrity. The position of each of the two nucleotides was represented as a coordinate system rigidly linked to the base of the nucleotide. Then, we calculated three rotational and three translational parameters that would allow the transition from one coordinate system to the other. Then, in known RNA structures, we calculated the same set of six parameters for all dinucleotide combinations in which the distance between the two nucleotides did not exceed 18 Å. Those cases in which the values of all parameters were close enough to the reference values were taken for further visual inspection. For identification of new cases of AGPM, we used the juxtaposition of the internal nucleotides 0Q and 0S of the two central base pairs.

5.6 Acknowledgments

The work was supported by an operating grant from Canadian Institutes of Health Research. M.G.G. held scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the Fonds de la Recherche en Santé du Québec (FRSQ).

5.7 References

1. Kim, S. H. & Sussman, J. L. (1976). pi turn is a conformational pattern in RNA loops and bends. *Nature* **260**, 645-646.

2. Quigley, G. J. & Rich, A. (1976). Structural domains of transfer RNA molecules. *Science* **194**, 796-806.

3. Holbrook, S. R., Sussman, J. L., Warrant, R. W. & Kim, S. H. (1978). Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J. Mol. Biol.* **123**, 631-660.

4. Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. (2001). The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214-4221.

5. Szep, S., Wang, J. & Moore, P. B. (2003). The crystal structure of a 26-nucleotide RNA containing a hook-turn. *RNA* **9**, 44-51.

6. Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. (1997). Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91**, 705-712.

7. Correll, C. C., Wool, I. G. & Munishkin, A. (1999). The two faces of the *Escherichia coli* 23 S rRNA sarcin/ricin domain: the structure at 1.11 Å resolution. *J. Mol. Biol.* **292**, 275-287.

8. Lescoute, A., Leontis, N. B., Massire, C. & Westhof, E. (2005). Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res.* **33**, 2395-2409.

9. Hendrix, D. K., Brenner, S. E. & Holbrook, S. R. (2005). RNA structural motifs: building blocks of a modular biomolecule. *Q. Rev. Biophys.* **38**, 221-243.

10. Leontis, N. B., Lescoute, A. & Westhof, E. (2006). The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* **16**, 279-287.

11. Noller, H. F. (2005). RNA structure: reading the ribosome. *Science* **309**, 1508-1514.

12. Gagnon, M. G. & Steinberg, S. V. (2002). GU receptors of double helices mediate tRNA movement in the ribosome. *RNA* **8**, 873-877.

13. Sponer, J. E., Reblova, K., Mokdad, A., Sychrovsky, V., Leszczynski, J. & Sponer, J. (2007). Leading RNA Tertiary Interactions: Structures, Energies, and Water Insertion of A-Minor and P-Interactions. A Quantum Chemical View. *J. Phys. Chem. B* **111**, 9153-9164.

14. Gagnon, M. G., Mukhopadhyay, A. & Steinberg, S. V. (2006). Close packing of helices 3 and 12 of 16S rRNA is required for the normal ribosome function. *J. Biol. Chem.* **281**, 39349-39357.

15. Mokdad, A., Krasovska, M. V., Sponer, J. & Leontis, N. B. (2006). Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.* **34**, 1326-1341.

16. Wuyts, J., Perriere, G. & Van De Peer, Y. (2004). The European ribosomal RNA database. *Nucleic Acids Res.* **32**, D101-103.

17. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26**, 148-153.

18. Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F. & Yonath, A. (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107**, 679-688.

19. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M. & Cate, J. H. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827-834.

20. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905-920.

21. Carter, A. P., Clemons, W. M., Jr., Brodersen, D. E., Morgan-Warren, R. J., Hartsch, T., Wimberly, B. T. & Ramakrishnan, V. (2001). Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* **291**, 498-501.

22. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. & Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883-896.

23. Korostelev, A., Trakhanov, S., Laurberg, M. & Noller, H. F. (2006). Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* **126**, 1065-1077.

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

25. Murthy, V. L. & Rose, G. D. (2003). RNABase: an annotated database of RNA structures. *Nucleic Acids Res.* **31**, 502-504.

26. Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68-74.

27. Jang, S. B., Hung, L. W., Jeong, M. S., Holbrook, E. L., Chen, X., Turner, D. H. & Holbrook, S. R. (2006). The crystal structure at 1.5 Å resolution of an RNA octamer duplex containing tandem G.U basepairs. *Biophys. J.* **90**, 4530-4537.

28. Hou, Y. M. & Schimmel, P. (1988). A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* **333**, 140-145.

29. McClain, W. H. & Foss, K. (1988). Changing the identity of a tRNA by introducing a G-U wobble pair near the 3' acceptor end. *Science* **240**, 793-796.

30. McClain, W. H., Schneider, J., Bhattacharya, S. & Gabriel, K. (1998). The importance of tRNA backbone-mediated interactions with synthetase for aminoacylation. *Proc. Natl. Acad. Sci. USA* **95**, 460-465.

5.8 Tables

Table I. Structure of the contact zones

Zones	Nucleotides	Atoms
QS	+1Q	O4′-C4′-C5′-O2′-O3′
	0Q	C1′-O2′
	+1S	O4′-C4′-C1′-O2′-C5′
	0S	C1′-O2′
QR	-1Q	C4′-C3′-O2′-O3′-C5′
	0Q	O1P-O5′-C5′
	-2R	C1′-O4′-C4′-C2′-O2′-C5′
	-1R	C1′-C2′-O2′
PS	-2P	C1′-O4′-C4′-C2′-O2′-C5′
	-1P	C1′-C2′-O2′
	-1S	C4′-C3′-O2′-O3′-C5′
	0S	O1P-O5′-C5′

A non-hydrogen atom was considered a part of a contact zone if its distance from a non-hydrogen atom of the opposite helix was equal or less than 4.2 Å. See also Figure 3.

Table II. Occurrence of different combinations of the -1-base pairs in AGPMs existing in prokaryotic rRNA

-1-base pairs	Number of sequences	%
GC/CG-GC/CG	38250	70.4
GC/CG-AU/UA	14176	26.1
AU/UA-GC/CG	438	0.8
AU/UA-AU/UA	380	0.7
Total WC-WC	53244	98.0
Total	54315	100

For Tables II-V, the data were obtained from the available rRNA alignments[16]. In all these tables, "Total" refers to the total number of nucleotide sequences for which the identities of all nucleotides in question are known. The extended statistics are shown in the Supplemental Tables I-IV.

Table III. Occurrence of different identities of the -2-base pairs in AGPMs existing in prokaryotic rRNA

-2-base pairs	Number of base pairs	% [b]
WC and closely related conformations [a]		
WC	54815	82.1
UG	1669	2.5
GU	293	0.4
UU	286	0.4
GA	42	0.06
AA	19	0.03
AG	14	0.02
Sheared conformations [a]		
GA	8507	12.7
AA	856	1.3
UA	320	0.5
UG	27	0.04
AG	1	0.001
Sub-total	66849	100
Others [c]	42031	
Total [d]	108880	

a: The base pair conformations were deduced from the available crystal structures of the AGPM.

b: Percentages were calculated with respect to the total number of the -2-dinucleotide combinations that form a base pair.

c: This number relates to those -2-dinucleotide combinations in motifs S62, S549, S911, L639, L657, L839 and L2847 that do not form a base pair.

d: The total number of -2-base pairs is based on 54440 sequences for which the identities of all four nucleotides are known.

See also the footnote to Table II.

Table IV. Occurrence of different combinations of the 0-base pairs in AGPMs existing in prokaryotic rRNA

0-base pairs	Number of sequences	%
All motifs except S549		
GU-GC	29524	69.3
GU-CG	10108	23.7
GU-AU	38	0.1
GU-UA	1818	4.2
Total GU-WC	41488	97.3
WC-WC	549	1.3
Others	590	1.4
Total	42627	100
Motif S549		
Bacteria		
WC-WC	8811	95.9
GU-WC	250	2.7
Others	129	1.4
Total	9190	100
Archaea		
GU-WC	381	77.1
WC-WC	12	2.4
Others	101	20.5
Total	494	100

For the combinations shown in this table, the normal order of two base pairs is not respected. GU always takes the first position regardless of whether in a real nucleotide sequence it stays at position [0P;0Q] or [0R;0S].

See the footnote to Table II.

Table V. Occurrence of different combinations of the +1-base pairs in AGPMs existing in prokaryotic rRNA

+1-base pairs	Number of sequences	%
Group [WC; non-WC] [a]		
WC-AG	22543	86.2
WC-AC[b]	2291	8.7
WC-AA	459	1.8
WC-GU	428	1.6
WC-AU[b]	112	0.4
WC-WC	42	0.2
Others	284	1.1
Total	26159	100
Group [WC-WC] [a]		
Motifs S911, L839 and L2698		
WC-WC	8184	89.9
WC-GU	840	9.2
Others	78	0.9
Total	9102	100
Motif S549		
Bacteria		
WC-WC	8962	98.3
WC-GU	32	0.4
WC-UG	28	0.3
Others	92	1.0
Total	9114	100
Archaea		
WC-UG	404	81.6
WC-WC	86	17.4
WC-GU	0	0
Others	5	1.0
Total	495	100

For the combinations shown in this table, the normal order of the two base pairs is not respected. WC always takes the first position regardless of whether in a real nucleotide sequence it stays at position [+1P;+1Q] or [+1R;+1S].

a: Group [WC; non-WC] includes motifs S62, S757, L554, L639, L657, L2291, L2687 and L2847. Group [WC-WC] includes motifs S549, S911, L839 and L2698 (Figure 2).

b: In the given structural context, all AC and AU base pairs are arranged in the way that the amino group of the adenosine forms a hydrogen bond with atom O2 of the pyrimidine. Such arrangement makes these base pairs similar to the sheared base pairs AG and AA.

See also the footnote to Table II.

5.9 Figures

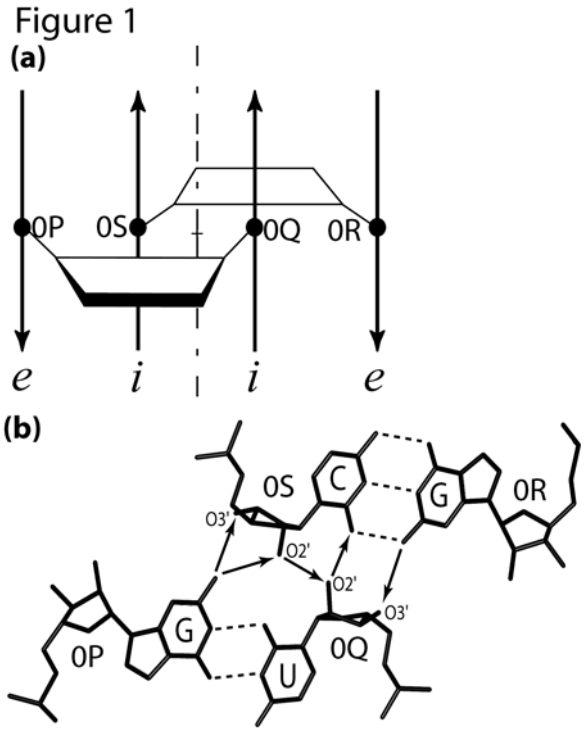


Figure 1. The along-groove packing motif.

(a): Schematic representation of the AGPM. Trapezoids represent base pairs opened toward the minor grooves. Arrows represent backbones directed 5'→3'. The internal and external strands of each helix are marked by italic letters *i* and *e*, respectively. The internal strand of each helix interacts with the minor groove of the other helix. Rotation of one helix for 180° around the symmetry axis (dash-dotted line) superposes it with the other helix. The positions of nucleotides 0P, 0Q, 0R and 0S, which form the central base pairs, are marked.

(b): Juxtaposition of the central base pairs in the AGPM. Arrows designate inter-helix hydrogen bonds directed from the donor to the acceptor atom. The characteristic geometry of the GU base pair allows one helix to closely pack against the other helix.

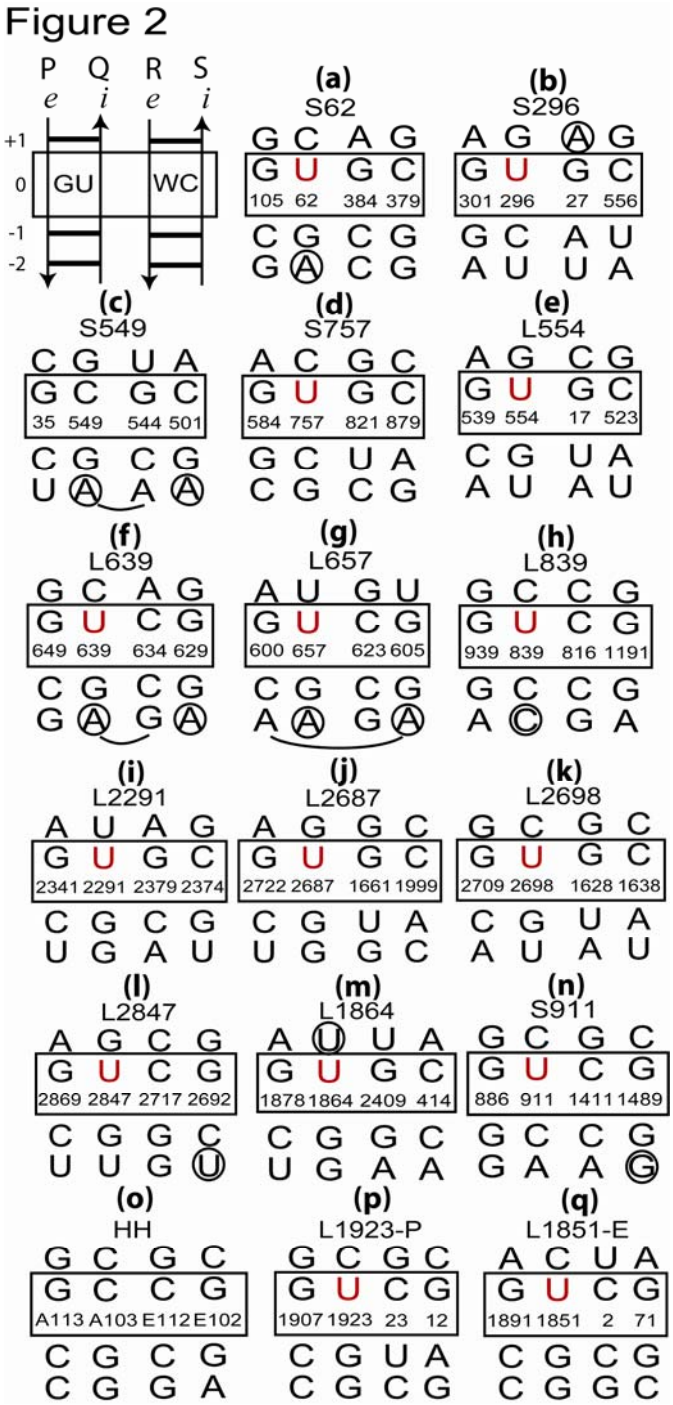


Figure 2. Nucleotide sequences of all known AGPMs identified within ribosomal RNA (a-n), between two hammerhead ribozymes (o), and between 23S rRNA and a tRNA bound at the P-site (p) and at the E-site (q). Motifs (a-m) are taken from the structure of the *E. coli* ribosome (pdb entry codes 2avy-2aw4[19]); motif (n) is taken from the complex of the *T. thermophilus* 30S subunit and the initiation factor IF1 (pdb entry code 1hr0[21]); motif (o) is taken from the structure of the Hammerhead ribozyme (pdb entry code 1hmh[26]); motifs (p, q) are taken from the structures of the *T. thermophilus* ribosome (pdb entry codes

2j00-2j01[24] and 1gix-1giy[22]). The positions and orientations of the GU- and WC-containing helices correspond to those shown at the upper left corner. Central base pairs are boxed. U in position 0Q is red. The *E. coli* nucleotide numbering is used for all cases found within the ribosome. The name of each motif starts with letter 'S' or 'L', which reflects the small or large subunit in which it is found, followed by the number in the standard *E. coli* nomenclature of the nucleotide occupying position 0Q in 16S or 23S rRNA. Black circles indicate those nucleotides that do not form a base pair and do not stack on the following nucleotide.

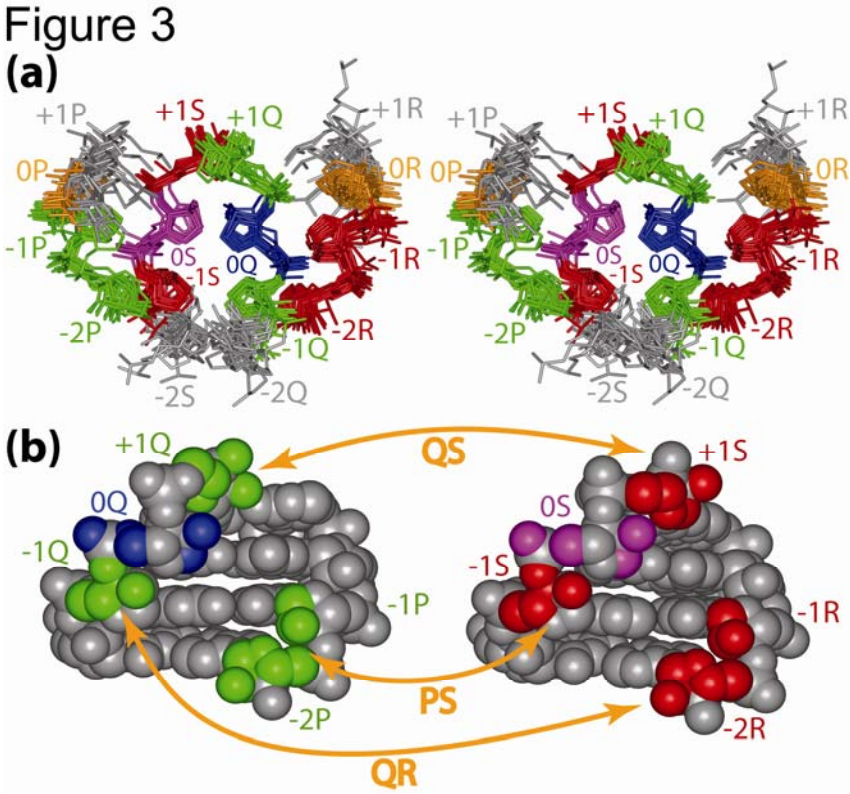


Figure 3. The arrangement of the three contact zones within the AGPM.

(a): Stereo view of all known AGPM structures [Figure 2(a-p)] superposed based on the positions of the C4′ atoms (rmsd = 0.87 Å). The high resolution structure of the *T. thermophilus* ribosome (pdb entry codes 2j00-2j01 and 2j02-2j03[24]) allowed including motif L1923-P [Figure 2(p)] to this superposition, while the E-site tRNA was positioned differently. For clarity, the bases are not shown. The nucleotides involved in the formation of the contact zones are green in the GU-helix and red in the WC-helix. Also, 0Q is blue, 0S is magenta, while 0P and 0R are brown. The external +1-nucleotides +1P and +1Q as well as internal -2-nucleotides -2Q and -2S are grey. Grey nucleotides are not involved in inter-helix contacts, which can explain the flexibility of their positions.

(b): Location of the atoms forming the three contact zones in the GU-helix (left) and in the WC-helix (right). For the atoms of the contact zones, the same colors are used as in panel (a). Brown two-headed arrows indicate the correspondence of the areas within each contact zone. The complete list of the atoms involved in the contact zones is given in Table I.

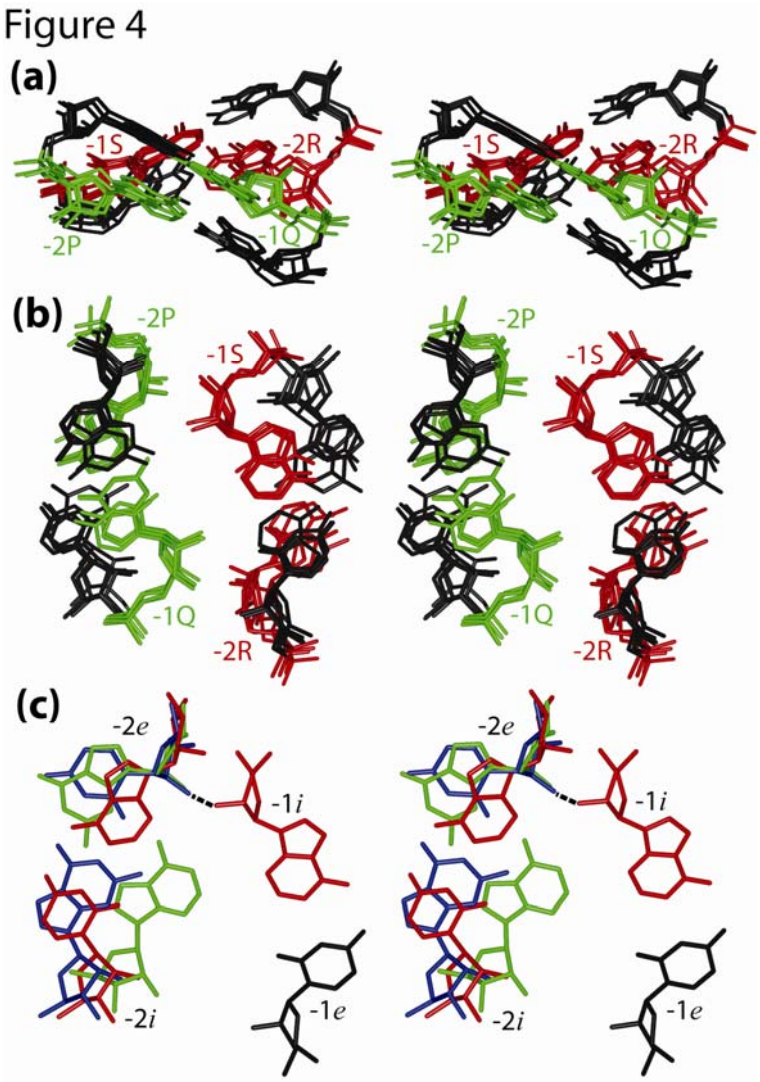
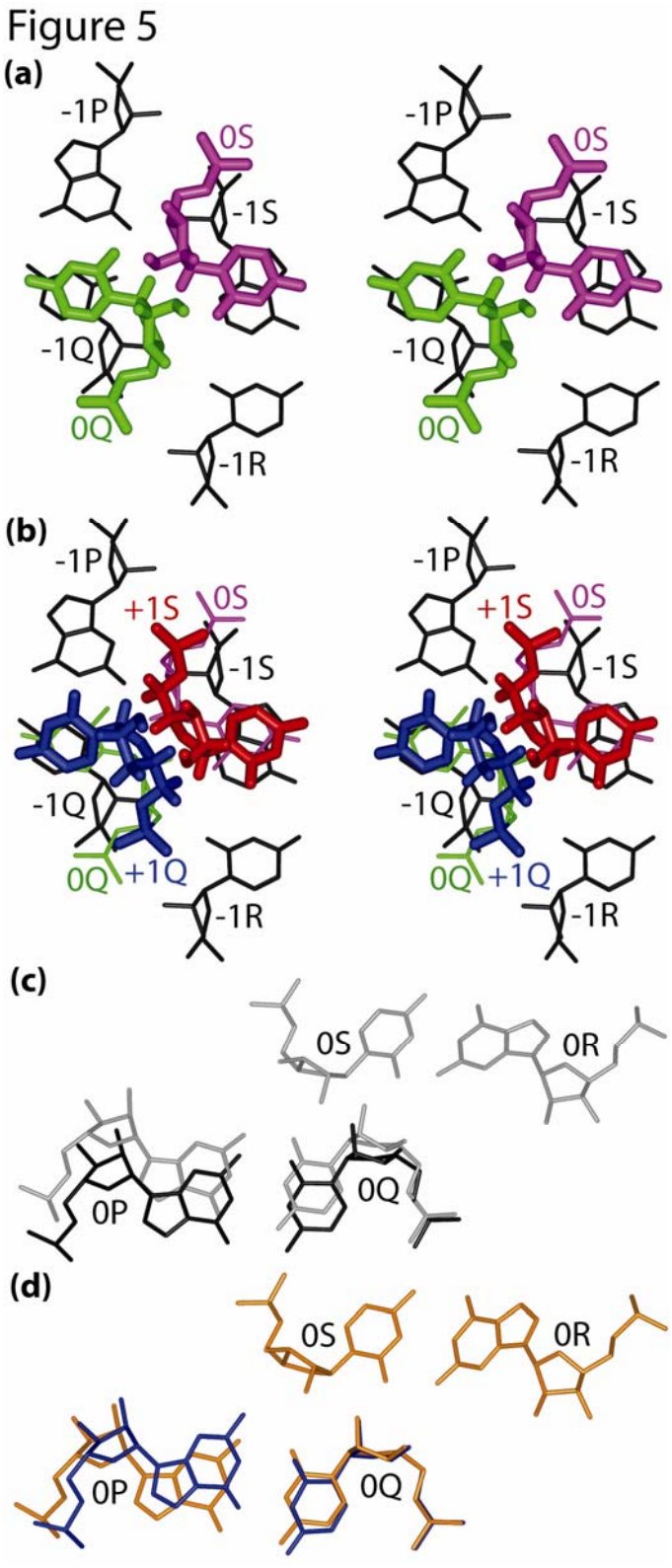


Figure 4. Ribose-ribose interactions within contact zones PS and QR.

(a-b): Two different stereo views of the same superposition of motifs S296, S757, L554 and L2698, in which both -2-base pairs are WC (rmsd = 0.56 Å). Nucleotides -2P and -1S form a major part of contact zone PS, while nucleotides -1Q and -2R form a major part of zone QR. These nucleotides are colored as in Figure 3.

(c): Stereo view of the interaction between the riboses of the external -2-nucleotide (-2*e*) and of the internal -1-nucleotide (-1*i*) for different -2-base pairs. For the -2-base pair, AU is red, UG is blue, and GA is green. In all -2-base pairs, the ribose of the external nucleotide -2*e* is positioned closely enough to the ribose of the internal nucleotide -1*i* of the -1-base pair to form a ribose-ribose contact. Compared to the -2-base pair AU, in both base pairs UG and GA, the external nucleotide -2*e* is slightly over-twisted, which allows the formation of a hydrogen bond between the O2′ atoms of the two riboses (dashed line). The internal nucleotide -2*i* of the -2-base pair shows a strong variability in its position, which is not surprising given that this nucleotide is not involved in inter-helix interactions.



See the legend on next page

Figure 5. The potential collision of the internal nucleotides at the 0- and +1-levels and its consequences for the structure and position of base pair [0P;0Q].

(a,b): Stereo view representations. The extension of both double helices from the -1-level (black) to levels 0 (a) and +1 (b) leads to the collision of the internal nucleotides (green and magenta in (a), blue and red in (b)).

(c): the displacement of the WC base pair [0P;0Q] towards the major groove allows nucleotide 0Q to avoid the collision with 0S. Grey: the juxtaposition of the central base pairs in the same theoretically obtained structure as in (a) and (b). Black: the adjustment of the position of a WC base pair [0P;0Q] that allows it to avoid the collision with base pair [0R;0S].

(d): the subsequent adjustment of the position of nucleotide 0P when a WC base pair [0P;0Q] (brown) is replaced by GU (blue).

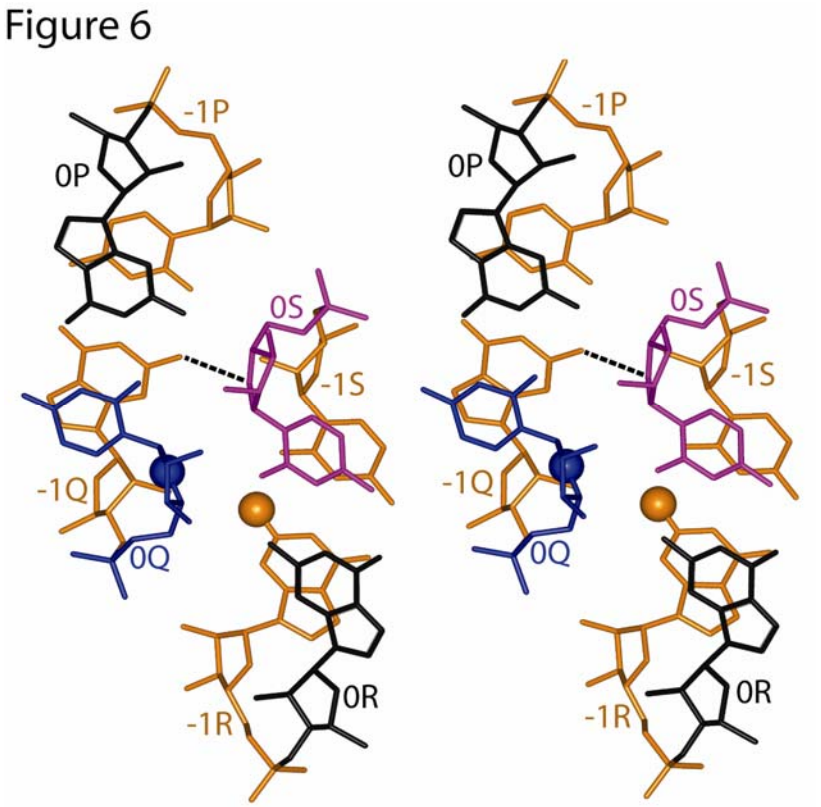


Figure 6. Stereo view of the asymmetry between base pairs [-1P;-1Q] and [-1R;-1S] caused by the displacement of nucleotide 0Q. In all cases of the AGPM, there is a hydrogen bond between atom O4′ of ribose 0S and the amino group of the guanine of base pair [-1P;-1Q] (dashed line). The analogous bond between atom O4′ of ribose 0Q (blue ball) and the amino group of the guanine of base pair [-1R;-1S] (brown ball) does not exist due to the asymmetric displacement of nucleotide 0Q.

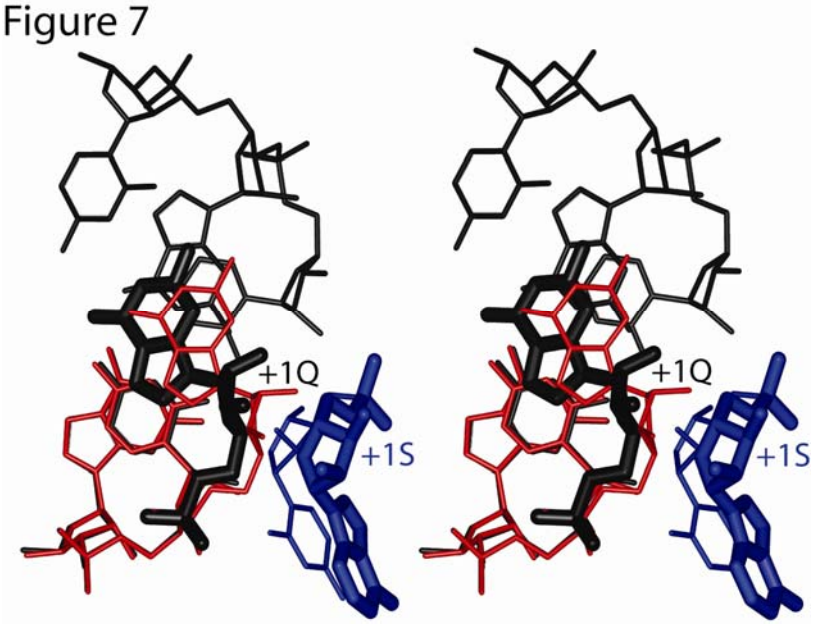


Figure 7. Stereo view of the deformation of the double helical geometry at level +1 when both +1-base pairs are WC. Red and blue are the theoretically obtained conformations of the internal strands as in Figure 5. The collision between nucleotides +1Q (red) and +1S (blue bold) is avoided through the displacement of +1Q toward the major groove for 1.5 Å (black bold) and the corresponding adjustment in the position of nucleotide +1P.

Figure 8

+1	WC,*AG*,*AC*	WC,*AG*,*AC*
0	GU,*GC*,*CG*	GC/CG,UA
-1	**GC/CG**	**WC**
-2	WC,*GA*,*UG*	WC,*GA*,*UG*

Figure 8. The consensus secondary structure for most known AGPMs. The base pair identities found in practically all nucleotide sequences, in most sequences, and in a notable number of sequences are shown, respectively, in bold, with the regular and comic fonts. The central base pairs are boxed.

5.10 Supplemental tables

Supplemental Table I. Statistical data of the identities of the -1-base pairs for all known ribosomal AGPMs

AGPM	base pairs [-1P;-1Q]-[-1R;-1S]	Eubacteria		base pairs [-1P;-1Q]-[-1R;-1S]	Archaebacteria			AGPM	base pairs [-1P;-1Q]-[-1R;-1S]	Eubacteria		base pairs [-1P;-1Q]-[-1R;-1S]	Archaebacteria	
		N [b]	%		N [b]	%				N [b]	%		N [b]	%
S62	CG-CG	8900	82.301	CG-CG	281	53.626		L657	CG-CG	393	98.992	CG-CG	37	100
	CG-UA	1676	15.498	CG-UA	218	41.603			CG-GG	2	0.504	TOTAL [a]	37	
	CG-GC	55	0.509	CG-AU	10	1.908			CG-UA	1	0.252			
	CG-AU	30	0.277	CG-GC	6	1.145			CA-CG	1	0.252			
	UA-CG	28	0.259	CU-CG	4	0.763			TOTAL [a]	397				
	CU-CG	18	0.166	CA-CG	2	0.382								
	CC-CG	16	0.148	CG-UC	1	0.191								
	CG-UG	12	0.111	UG-UA	1	0.191								
	UG-CG	10	0.092	GG-CG	1	0.191								
	GG-CG	8	0.074	TOTAL [a]	524									
	CG-CA	7	0.065											
	CG-GG	6	0.055											
	CG-CC	6	0.055											
	CA-CG	6	0.055											
	CG-GU	5	0.046											
	GG-UA	5	0.046											
	CG-CU	3	0.028											
	CG-UC	3	0.028											
	UG-UA	2	0.018											
	AG-CG	2	0.018											
	CG-AG	2	0.018											
	CU-UA	2	0.018											
	AG-UA	2	0.018											
	UA-UA	2	0.018											
	UA-AG	1	0.009											
	CC-GC	1	0.009											
	GG-UG	1	0.009											
	GG-AU	1	0.009											
	UG-CA	1	0.009											
	CA-UA	1	0.009											
	CC-UA	1	0.009											
	UG-CC	1	0.009											
	TOTAL [a]	10814												

S296	GC-GC	3533	44.790	CG-GC	242	58.596		L839	GC-CG	264	66.667	GC-CG	36	100
	GC-AU	3521	44.637	CG-CG	47	11.380			CG-CG	131	33.081	TOTAL [a]	36	
	GU-AU	309	3.917	UA-GC	39	9.443			AG-CG	1	0.253			
	AU-AU	126	1.597	CG-GU	25	6.053			TOTAL [a]	396				
	AU-GC	123	1.559	GC-GC	18	4.358								
	CG-AU	75	0.951	GC-CG	15	3.632								
	GC-GU	52	0.659	AU-GU	10	2.421								
	GC-CG	37	0.469	AU-GC	7	1.695								
	CG-GC	28	0.355	GC-GU	3	0.726								
	GC-AC	21	0.266	CC-CG	1	0.242								
	GU-GC	16	0.203	CG-CC	1	0.242								
	GC-UU	12	0.152	CG-AU	1	0.242								
	GC-UC	10	0.127	UG-AU	1	0.242								
	GU-GU	3	0.038	AG-GC	1	0.242								
	UA-GU	3	0.038	UA-CC	1	0.242								
	GG-AU	2	0.025	UA-AU	1	0.242								
	CG-CC	2	0.025	TOTAL [a]	413									
	GC-CC	2	0.025											
	GC-CU	2	0.025											
	CG-UC	1	0.013											
	CG-AC	1	0.013											
	AC-AU	1	0.013											
	AU-AC	1	0.013											
	GC-GA	1	0.013											
	CC-GC	1	0.013											
	GG-GC	1	0.013											
	AC-GC	1	0.013											
	GC-AA	1	0.013											
	GC-AG	1	0.013											
	UA-GC	1	0.013											
	TOTAL [a]	7888												
S549	CG-CG	8888	95.416	CG-CG	501	98.622		L1864	CG-GC	332	84.051	Replaced by GNRA[1]		
	UA-CG	185	1.986	GC-CG	3	0.591			GC-GC	61	15.443			
	CG-UA	71	0.762	GG-CG	1	0.197			GG-GC	1	0.253			
	UG-CG	35	0.376	UG-CG	1	0.197			CG-AC	1	0.253			
	CC-CG	24	0.258	CA-CG	1	0.197			TOTAL [a]	395				
	CA-CG	24	0.258	CU-CG	1	0.197								
	GG-CG	23	0.247	TOTAL [a]	508									
	CG-CC	13	0.140											
	AG-CG	12	0.129											
	CG-CA	9	0.097											
	CG-UG	6	0.064											
	CG-CU	5	0.054											
	CU-CG	4	0.043											
	CG-AU	3	0.032											
	UC-CG	2	0.021											

	CG-GG	2	0.021											
	CC-CA	1	0.011											
	UA-CU	1	0.011											
	CA-CC	1	0.011											
	CA-CA	1	0.011											
	GG-CU	1	0.011											
	CG-AG	1	0.011											
	CC-CC	1	0.011											
	UA-GG	1	0.011											
	AA-CG	1	0.011											
	TOTAL [a]	9315												
S757	GC-AU	4200	35.247	CG-CG	454	78.141		L2291	CG-UA	271	68.608	CG-CG	24	68.571
	GC-GC	3532	29.641	CG-UA	120	20.654			CG-CG	91	23.038	CG-GC	5	14.286
	GC-UA	3208	26.922	CG-AU	3	0.516			GC-UA	17	4.304	UA-CG	3	8.571
	GC-CG	551	4.624	UG-CG	1	0.172			CG-AU	10	2.532	GC-CG	2	5.714
	AU-AU	170	1.427	GC-CG	1	0.172			UA-CG	4	1.013	CG-UA	1	2.857
	AU-UA	66	0.554	CG-GC	1	0.172			GC-CG	2	0.506	TOTAL [a]	35	
	GC-GU	22	0.185	CG-CA	1	0.172			TOTAL [a]	395				
	AU-GC	17	0.143	TOTAL [a]	581									
	CG-GC	13	0.109											
	GC-AC	13	0.109											
	GU-AU	9	0.076											
	CG-UA	7	0.059											
	CC-UA	7	0.059											
	GC-CA	7	0.059											
	GC-UU	7	0.059											
	GC-CC	6	0.050											
	GC-UC	6	0.050											
	CG-AU	5	0.042											
	UC-GC	5	0.042											
	GC-AA	5	0.042											
	AC-GC	5	0.042											
	GC-UG	5	0.042											
	GA-GC	4	0.034											
	GU-GC	4	0.034											
	GA-UA	4	0.034											
	GC-CU	4	0.034											
	GC-GA	3	0.025											
	GG-UA	3	0.025											
	GU-UA	3	0.025											
	GC-GG	3	0.025											
	AC-AU	3	0.025											
	UU-CG	3	0.025											
	GG-AU	2	0.017											
	CC-AU	2	0.017											
	AC-UA	2	0.017											

	GC-AG	2	0.017											
	UC-UA	1	0.008											
	GA-CG	1	0.008											
	CG-CG	1	0.008											
	CC-GC	1	0.008											
	GG-GC	1	0.008											
	GA-AU	1	0.008											
	UC-AU	1	0.008											
	UA-CG	1	0.008											
	TOTAL [a]	11916												
S911	GC-CG	8259	98.146	GC-CG	385	99.741		L2687	CG-UA	287	87.234	CG-CG	26	70.270
	GC-UA	57	0.677	GC-GC	1	0.259			CG-AU	23	6.991	CG-UA	11	29.730
	GC-CA	18	0.214	TOTAL [a]	386				CG-CG	18	5.471	TOTAL [a]	37	
	GC-CU	12	0.143						CG-UG	1	0.304			
	AC-CG	11	0.131						TOTAL [a]	329				
	GC-CC	11	0.131											
	GU-CG	9	0.107											
	GC-GC	7	0.083											
	GC-GG	7	0.083											
	GC-UG	6	0.071											
	CC-CG	5	0.059											
	UC-CG	3	0.036											
	GG-CG	3	0.036											
	GA-CG	2	0.024											
	GC-AG	2	0.024											
	GC-AU	1	0.012											
	AC-CA	1	0.012											
	GC-UC	1	0.012											
	TOTAL [a]	8415												
L554	CG-CG	204	59.649	CG-CG	31	93.939		L2698	CG-UA	155	46.970	CG-GC	26	70.270
	CG-UA	110	32.164	CG-UA	2	6.061			CG-GC	75	22.727	CG-CG	8	21.622
	GC-CG	15	4.386	TOTAL [a]	33				CG-AU	62	18.788	UA-CG	3	8.108
	AU-CG	10	2.924						CG-GU	20	6.061	TOTAL [a]	37	
	UG-UA	1	0.292						UA-AU	7	2.121			
	AU-UG	1	0.292						UA-UA	6	1.818			
	UA-UA	1	0.292						UA-GC	2	0.606			
	TOTAL [a]	342							CG-CG	2	0.606			
									CG-GA	1	0.303			
									TOTAL [a]	330				
L639	CG-CG	385	97.716	CG-CG	37	100		L2847	CG-GC	235	80.205	CG-GC	21	65.625
	CG-CA	5	1.269	TOTAL [a]	37				GC-GC	20	6.826	UA-GC	9	28.125
	CG-GG	2	0.508						CG-AU	18	6.143	CG-AU	2	6.250
	CU-GG	1	0.254						CG-CG	10	3.413	TOTAL [a]	32	
	CG-CC	1	0.254						UA-GC	6	2.048			
	TOTAL [a]	394							CG-GU	2	0.683			
									UG-CG	1	0.341			

									UA-AU TOTAL [a]	1 293	0.341			

The data were obtained from the available rRNA alignments[2]. For all cases of AGPM, the *E. coli* numbering is used.
a: For the statistics, only those cases where the identities of all four nucleotides are known have been considered.
b: N is the number of sequences which have the corresponding -1-base pairs combination.

Supplemental Table II. Statistical data of the identities of the -2-base pairs for all known ribosomal AGPMs

AGPM	base pairs [-2P;-2Q]-[-2R;-2S]	Eubacteria		base pairs [-2P;-2Q]-[-2R;-2S]	Archaebacteria			AGPM	base pairs [-2P;-2Q]-[-2R;-2S]	Eubacteria		base pairs [-2P;-2Q]-[-2R;-2S]	Archaebacteria	
		N [b]	%		N [b]	%				N [b]	%		N [b]	%
S62	GA-CG	9555	88.154	GA-GC	320	60.952		L657	AA-GA	251	64.359	GA-GA	23	62.162
	AA-CG	943	8.700	AA-GC	95	18.095			GA-GA	121	31.026	GA-AA	9	24.324
	GA-UA	73	0.673	GA-GU	56	10.667			AA-AA	13	3.333	AA-AA	3	8.108
	GU-UA	46	0.424	AA-AU	19	3.619			GU-GU	1	0.256	AA-GA	2	5.405
	GA-UG	45	0.415	AA-GU	8	1.524			AG-GA	1	0.256	TOTAL [a]	37	
	GA-GC	37	0.341	GA-CG	8	1.524			CA-AA	1	0.256			
	GU-CG	35	0.323	GA-AU	4	0.762			GU-GA	1	0.256			
	GA-GG	18	0.166	AA-UG	4	0.762			UA-GA	1	0.256			
	CA-CG	14	0.129	UA-GC	2	0.381			TOTAL [a]	390				
	GG-CG	10	0.092	AA-CC	2	0.381								
	AA-GC	7	0.065	UU-AU	1	0.190								
	AU-UA	7	0.065	GA-CC	1	0.190								
	AU-CG	6	0.055	CA-GU	1	0.190								
	GC-CG	5	0.046	AA-UC	1	0.190								
	GA-GU	5	0.046	GG-GC	1	0.190								
	AA-UG	5	0.046	UA-AU	1	0.190								
	AA-GG	4	0.037	GA-CU	1	0.190								
	AA-UA	4	0.037	TOTAL [a]	525									
	UA-CG	3	0.028											
	GU-GC	3	0.028											
	GA-CU	3	0.028											
	GA-CC	3	0.028											
	GU-UG	2	0.018											
	GA-AG	2	0.018											
	AA-AU	1	0.009											
	GA-UU	1	0.009											
	GA-UC	1	0.009											
	AC-CG	1	0.009											
	TOTAL [a]	10839												
S296	GC-AU	3102	37.959	AU-CG	162	34.764		L839	AC-GA	371	94.885	AC-GA	35	97.222
	AU-UA	1857	22.724	GC-CG	126	27.039			AU-GA	19	4.859	UC-GA	1	2.778
	AU-AU	1725	21.109	AU-UA	98	21.030			AG-GA	1	0.256	TOTAL [a]	36	
	AU-GC	633	7.746	AU-UG	31	6.652			TOTAL [a]	391				
	UA-AU	262	3.206	CG-CG	13	2.790								
	GC-GC	170	2.080	GC-UA	9	1.931								
	CG-AU	114	1.395	CG-UA	6	1.288								
	UA-UA	97	1.187	AU-GC	6	1.288								
	GU-AU	26	0.318	GC-AU	4	0.858								
	UG-AU	25	0.306	GC-GC	3	0.644								
	AU-GU	21	0.257	AC-CG	1	0.215								

	AU-UU	17	0.208	CG-UG	1	0.215								
	GC-UA	14	0.171	CG-GC	1	0.215								
	GC-UU	13	0.159	GU-UA	1	0.215								
	AU-GA	12	0.147	UA-AU	1	0.215								
	AU-AA	11	0.135	AU-GG	1	0.215								
	GC-CU	10	0.122	UA-CG	1	0.215								
	AU-AC	9	0.110	AU-AU	1	0.215								
	GC-GU	8	0.098	TOTAL [a]	466									
	AU-UC	5	0.061											
	CG-GC	5	0.061											
	UC-AU	4	0.049											
	AU-CA	4	0.049											
	GC-AC	4	0.049											
	UG-UA	3	0.037											
	AC-AU	3	0.037											
	UA-UU	2	0.024											
	UU-UA	2	0.024											
	CA-AU	1	0.012											
	UU-AU	1	0.012											
	AU-AG	1	0.012											
	CG-GU	1	0.012											
	AU-CU	1	0.012											
	AU-UG	1	0.012											
	GC-AG	1	0.012											
	GA-AU	1	0.012											
	AA-AU	1	0.012											
	CG-CU	1	0.012											
	AG-AU	1	0.012											
	AG-UA	1	0.012											
	GG-UU	1	0.012											
	GU-UU	1	0.012											
	TOTAL [a]	8172												
S549	UA-AA	4707	51.725	UA-GA	335	66.075		L1864	UG-AA	244	61.461	Replaced by GNRA[1]		
	UA-GA	2816	30.945	UA-AA	129	25.444			AU-AA	75	18.892			
	UA-UA	1391	15.286	UA-UA	38	7.495			UA-AA	20	5.038			
	UG-GA	68	0.747	UG-AA	1	0.197			GC-AA	14	3.526			
	UA-CA	33	0.363	CA-AA	1	0.197			CG-AA	10	2.519			
	CA-AA	17	0.187	GA-GA	1	0.197			AU-AC	6	1.511			
	UG-AA	15	0.165	UA-UG	1	0.197			UA-GA	6	1.511			
	GA-AA	10	0.110	CA-GA	1	0.197			UU-AA	6	1.511			
	GA-GA	5	0.055	TOTAL [a]	507				GU-AA	5	1.259			
	UA-AG	4	0.044						UG-GA	3	0.756			
	AA-AA	4	0.044						AU-GA	3	0.756			
	UC-AA	3	0.033						UG-AU	2	0.504			
	UC-GA	3	0.033						UA-AC	1	0.252			
	CA-UA	3	0.033						CA-AA	1	0.252			

	AA-UA	2	0.022						AU-AG	1	0.252			
	UG-UA	2	0.022						TOTAL [a]	397				
	CG-GA	2	0.022											
	UA-UG	2	0.022											
	UG-CA	2	0.022											
	UU-AA	2	0.022											
	UA-GG	2	0.022											
	CA-GA	1	0.011											
	GG-GA	1	0.011											
	UU-CU	1	0.011											
	UA-AC	1	0.011											
	AA-CA	1	0.011											
	UA-AU	1	0.011											
	UU-UA	1	0.011											
	TOTAL [a]	9100												
S757	CG-GC	3896	32.803	CG-GC	244	42.583		L2291	UG-AU	182	46.193	CG-GC	22	59.459
	CG-AU	2820	23.743	CG-UA	79	13.787			UG-CG	126	31.980	CG-CG	6	16.216
	CG-CG	2138	18.001	CG-AU	56	9.773			CG-CG	74	18.782	UG-GC	4	10.811
	UA-CG	683	5.751	CG-CG	46	8.028			UG-UA	7	1.777	UG-AU	4	10.811
	CG-UA	639	5.380	UA-CG	34	5.934			CG-GC	2	0.508	CG-CC	1	2.703
	UA-GC	508	4.277	UG-CG	27	4.712			UG-GC	1	0.254	TOTAL [a]	37	
	UG-GC	471	3.966	UA-UA	23	4.014			UA-CG	1	0.254			
	UA-AU	406	3.418	UA-GU	19	3.316			UA-AU	1	0.254			
	UA-UA	105	0.884	UA-GC	18	3.141			TOTAL [a]	394				
	UG-CG	37	0.312	CG-AC	10	1.745								
	UG-UA	16	0.135	UG-GC	6	1.047								
	CG-GU	15	0.126	UA-AU	5	0.873								
	CG-AC	13	0.109	CG-GU	3	0.524								
	CA-AU	8	0.067	CA-UA	1	0.175								
	CG-CC	8	0.067	GC-AU	1	0.175								
	CG-UU	8	0.067	CC-GC	1	0.175								
	UG-AU	8	0.067	TOTAL [a]	573									
	GC-AU	7	0.059											
	CC-AU	7	0.059											
	CG-AG	6	0.051											
	CG-AA	6	0.051											
	CA-CG	5	0.042											
	CA-GC	5	0.042											
	CU-GC	5	0.042											
	CC-GC	5	0.042											
	CG-UG	5	0.042											
	GG-GC	3	0.025											
	GG-CG	3	0.025											
	CU-AU	3	0.025											
	CG-UC	3	0.025											
	CG-CA	3	0.025											

	CG-GG	2	0.017											
	GA-UU	2	0.017											
	AG-CG	2	0.017											
	GC-GC	2	0.017											
	AU-CG	2	0.017											
	CU-CG	1	0.008											
	UG-GG	1	0.008											
	CA-UG	1	0.008											
	AA-AU	1	0.008											
	UC-GC	1	0.008											
	UC-CG	1	0.008											
	UA-GU	1	0.008											
	UA-CA	1	0.008											
	UA-UG	1	0.008											
	UA-AC	1	0.008											
	CG-CU	1	0.008											
	GG-AU	1	0.008											
	CC-CG	1	0.008											
	CA-GU	1	0.008											
	GA-AU	1	0.008											
	AU-GC	1	0.008											
	UG-UG	1	0.008											
	CC-UA	1	0.008											
	UA-UU	1	0.008											
	UG-CU	1	0.008											
	GC-UA	1	0.008											
	CG-GA	1	0.008											
	TOTAL [a]	11877												
S911	GA-AG	7828	92.42	AA-AG	224	60.054		L2687	UG-UA	151	46.037	CA-UA	9	24.324
	UA-AG	316	3.731	GA-AG	137	36.729			UG-GC	132	40.244	UG-CG	7	18.919
	AA-AG	243	2.869	AA-AA	5	1.340			CG-UA	9	2.744	UG-GC	6	16.216
	GA-CG	30	0.354	UA-AG	3	0.804			UA-UA	9	2.744	CG-UA	5	13.514
	GA-AC	10	0.118	GA-UA	3	0.804			UG-CG	8	2.439	UG-AU	4	10.811
	GA-AA	9	0.106	AA-AU	1	0.268			CG-CG	5	1.524	UG-UA	2	5.405
	GA-AU	9	0.106	TOTAL [a]	373				UG-AU	3	0.915	CA-GC	2	5.405
	GU-AG	6	0.071						CG-UG	3	0.915	CA-CG	1	2.703
	AA-AA	6	0.071						UG-UG	2	0.610	CG-CG	1	2.703
	GG-AG	4	0.047						AU-UG	2	0.610	TOTAL [a]	37	
	GA-UA	3	0.035						AU-UA	2	0.610			
	GC-AG	2	0.024						CG-GC	1	0.305			
	GA-CU	1	0.012						UG-GA	1	0.305			
	GA-UG	1	0.012						TOTAL [a]	328				
	UA-AC	1	0.012											
	CA-AG	1	0.012											
	TOTAL [a]	8470												

L554	CG-GU	76	22.419	AU-AU	7	21.212		L2698	AU-AU	249	75.684	AU-UG	27	72.973
	AU-AU	67	19.764	CG-GC	7	21.212			AU-UU	49	14.894	AU-GA	5	13.514
	CG-AU	54	15.929	AU-GC	4	12.121			AU-CG	10	3.040	AU-CU	2	5.405
	UA-AU	32	9.440	CG-AU	4	12.121			AU-GC	8	2.432	AU-AA	2	5.405
	CG-GC	18	5.310	UG-UU	3	9.091			GC-AU	5	1.520	AU-UU	1	2.703
	GU-AU	18	5.310	UG-AU	3	9.091			AU-UG	4	1.216	TOTAL [a]	37	
	CG-UA	17	5.015	UA-GC	2	6.061			AU-UA	2	0.608			
	GC-GU	13	3.835	CG-GU	1	3.030			AU-GU	1	0.304			
	AU-GU	10	2.950	UA-AU	1	3.030			AU-CU	1	0.304			
	GC-AU	6	1.770	AU-UA	1	3.030			TOTAL [a]	329				
	CG-CG	6	1.770	TOTAL [a]	33									
	GC-GC	4	1.180											
	AU-UU	3	0.885											
	UU-AU	3	0.885											
	UA-GC	2	0.590											
	GC-CG	2	0.590											
	UA-UA	2	0.590											
	AU-UA	2	0.590											
	UG-GU	2	0.590											
	UU-GC	1	0.295											
	GA-AU	1	0.295											
	TOTAL [a]	339												
L639	GA-GA	347	88.747	CG-GA	24	64.865		L2847	UU-GA	74	25.256	CG-GA	32	100
	GA-AA	25	6.394	GC-GA	6	16.216			UU-GU	69	23.549	TOTAL [a]	32	
	GA-UA	14	3.581	CG-AA	4	10.811			UG-GA	50	17.065			
	UA-GA	2	0.512	GC-AA	3	8.108			GA-GA	20	6.826			
	AA-GA	2	0.512	TOTAL [a]	37				UU-UG	12	4.096			
	GG-AA	1	0.256						GC-AA	10	3.413			
	TOTAL [a]	391							UU-GG	7	2.389			
									CG-GA	7	2.389			
									UU-AA	6	2.048			
									UG-GG	6	2.048			
									UG-UG	5	1.706			
									UG-GU	5	1.706			
									UA-GA	4	1.365			
									GA-AA	3	1.024			
									UU-UA	3	1.024			
									UA-UG	3	1.024			
									CG-UG	3	1.024			
									GC-GA	1	0.341			
									CG-GC	1	0.341			
									UC-GA	1	0.341			
									UC-UG	1	0.341			
									UU-AC	1	0.341			
									UA-GG	1	0.341			
									TOTAL [a]	293				

The data were obtained from the available rRNA alignments[2]. For all cases of AGPM, the *E. coli* numbering is used.
a: For the statistics, only those cases where the identities of all four nucleotides are known have been considered.
b: N is the number of sequences which have the corresponding -2-base pairs combination.

Supplemental Table III. Statistical data of the identities of the 0-base pairs for all known AGPMs

AGPM	Central base pairs	Eubacteria		Central base pairs	Archaebacteria		AGPM	Central base pairs	Eubacteria		Central base pairs	Archaebacteria	
		N [b]	%		N [b]	%			N [b]	%		N [b]	%
S62	**GU-GC**	8036	73.779	**GU-GC**	280	53.030	L657	**GU-CG**	248	62.469	**GU-GC**	16	43.243
	GU-CG	2676	24.568	**GU-CG**	195	36.932		**GU-GC**	132	33.249	**GU-CG**	8	21.622
	CU-GC	25	0.230	**GU-UA**	37	7.008		**GU-UA**	15	3.778	**GC-GU**	4	10.811
	GU-CC	22	0.202	GU-CC	6	1.136		GU-CC	1	0.252	**AU-GU**	3	8.108
	GU-UA	21	0.193	**CG-GU**	2	0.379		UU-GA	1	0.252	GC-GC	3	8.108
	GC-GU	13	0.119	GC-CG	2	0.379		TOTAL [a]	397		**GU-UA**	2	5.405
	GC-GC	11	0.101	GA-UA	1	0.189					GU-UG	1	2.703
	UU-GC	11	0.101	AU-GC	1	0.189					TOTAL [a]	37	
	GU-GG	10	0.092	CU-GC	1	0.189							
	AU-GC	10	0.092	GC-GC	1	0.189							
	GU-AU	10	0.092	GU-GU	1	0.189							
	GU-GU	6	0.055	CG-GC	1	0.189							
	GU-AC	6	0.055	TOTAL [a]	528								
	GA-GC	6	0.055										
	CU-CG	4	0.037										
	GU-UC	4	0.037										
	GU-CA	3	0.028										
	GG-GC	3	0.028										
	GC-CG	3	0.028										
	GU-AG	3	0.028										
	UG-GU	3	0.028										
	UU-CG	3	0.028										
	GG-CC	1	0.009										
	GU-GA	1	0.009										
	GU-UG	1	0.009										
	TOTAL [a]	10892											
S296	**GU-GC**	5968	97.932	**GU-GC**	286	71.859	L839	**GU-CG**	265	67.089	**GU-CG**	17	47.222
	GU-AC	27	0.443	**GU-CG**	61	15.327		**GU-UA**	63	15.949	GC-CG	10	27.778
	GC-GC	26	0.427	GU-GU	31	7.789		**GU-GC**	56	14.177	**GU-UA**	4	11.111
	AU-GC	25	0.410	**GU-UA**	6	1.508		AU-UA	4	1.013	AU-CG	2	5.556
	GU-UC	18	0.295	**GU-AU**	6	1.508		CG-CG	3	0.759	CG-CG	2	5.556
	GU-CC	14	0.230	AU-GC	4	1.005		GU-UG	2	0.506	UA-UG	1	2.778
	GU-GU	4	0.066	GU-UG	1	0.251		GU-GG	1	0.253	TOTAL [a]	36	
	GU-AU	3	0.049	AU-CG	1	0.251		UA-CG	1	0.253			
	UU-GC	3	0.049	GU-CU	1	0.251		TOTAL [a]	395				
	CU-GC	2	0.033	GU-AG	1	0.251							
	GU-GA	1	0.016	TOTAL [a]	398								
	AC-GC	1	0.016										
	CU-UC	1	0.016										
	AU-CC	1	0.016										

	TOTAL [a]	6094												
S549	GC-GC	8032	87.399	**GU-CG**	206	41.7		L1864	**GU-GC**	390	98.734	Replaced by GNRA[1]		
	GC-CG	668	7.269	GU-UG	96	19.433			AU-GC	3	0.759			
	GC-GU	230	2.503	**GU-UA**	85	17.206			CU-GC	1	0.253			
	GC-AU	95	1.034	**GU-GC**	59	11.943			GU-GA	1	0.253			
	CC-GC	35	0.381	**GU-AU**	31	6.275			TOTAL [a]	395				
	UC-GC	21	0.229	GC-GC	7	1.417								
	GU-GC	15	0.163	CG-CG	4	0.81								
	AC-GC	15	0.163	CG-UG	3	0.607								
	GC-UA	14	0.152	CU-CG	1	0.202								
	GC-CC	11	0.120	GC-UA	1	0.202								
	GC-GG	10	0.109	GU-GU	1	0.202								
	GC-UG	9	0.098	TOTAL [a]	494									
	CC-CG	5	0.054											
	GC-UC	4	0.044											
	GU-AU	4	0.044											
	CC-GU	4	0.044											
	GC-AC	2	0.022											
	GG-GC	2	0.022											
	AC-GU	2	0.022											
	GC-AG	1	0.011											
	GA-GC	1	0.011											
	GA-UU	1	0.011											
	GU-CG	1	0.011											
	GG-GU	1	0.011											
	GG-GG	1	0.011											
	GC-GA	1	0.011											
	GG-CG	1	0.011											
	CG-GC	1	0.011											
	CG-CG	1	0.011											
	CC-AU	1	0.011											
	UC-GU	1	0.011											
	TOTAL [a]	9190												
S757	**GU-GC**	11737	98.755	**GU-GC**	555	95.197		L2291	**GU-GC**	382	96.465	**GU-GC**	23	62.162
	GU-CG	43	0.362	GC-AC	12	2.058			**GU-CG**	8	2.020	**GU-CG**	9	24.324
	AU-GC	26	0.219	GU-GU	9	1.544			**GU-UA**	3	0.758	GC-GC	3	8.108
	CU-GC	18	0.151	**GU-AU**	5	0.858			AU-GC	2	0.505	AU-GC	2	5.405
	GU-GU	13	0.109	GG-GC	1	0.172			CU-GC	1	0.253	TOTAL [a]	37	
	UU-GC	11	0.093	**AU-GU**	1	0.172			TOTAL [a]	396				
	GU-AU	10	0.084	TOTAL [a]	583									
	GU-AC	6	0.050											
	GA-GC	5	0.042											

	GC-GC GU-UC GG-GC GU-CC GU-GA UU-CC TOTAL [a]	5 4 2 2 2 1 11885	0.042 0.034 0.017 0.017 0.017 0.008											
S911	**GU-CG** **GU-UA** AU-CG AU-UA GU-CA GU-UG CU-CG GU-CC UA-CG GU-GG GU-CU UU-CG **GU-GC** GC-CG GU-GA AU-GC GU-AG UU-GC GA-CG AU-GG GU-UC TOTAL [a]	6125 1585 273 104 89 78 14 11 7 7 7 5 4 3 2 2 1 1 1 1 1 8321	73.609 19.048 3.281 1.250 1.070 0.937 0.168 0.132 0.084 0.084 0.084 0.060 0.048 0.036 0.024 0.024 0.012 0.012 0.012 0.012 0.012	**GU-CG** **GU-GC** GU-UG **GU-UA** AU-CG GU-CA GC-CG TOTAL [a]	231 98 33 11 5 1 1 380	60.789 25.789 8.684 2.895 1.316 0.263 0.263		L2687	**GU-GC** **GU-UA** **GU-CG** GU-GU GU-CA UU-GC TOTAL [a]	269 34 23 1 1 1 329	81.763 10.334 6.991 0.304 0.304 0.304	**GU-GC** **GU-CG** **GU-UA** GU-UC TOTAL [a]	23 11 2 1 37	62.162 29.730 5.405 2.703
L554	**GU-GC** **GC-GU** GC-GC AU-GC TOTAL [a]	332 6 2 1 341	97.361 1.760 0.587 0.293	**GU-GC** TOTAL [a]	33 33	100		L2698	**GU-GC** GU-AC TOTAL [a]	313 5 318	98.428 1.572	**GU-GC** TOTAL [a]	37 37	100
L639	**GU-GC** **GU-CG** **GU-UA** **GC-GU** GU-GG GU-CC TOTAL [a]	247 101 34 8 5 1 396	62.374 25.505 8.586 2.020 1.263 0.253	**GU-CG** **GU-GC** TOTAL [a]	36 1 37	97.297 2.703		L2847	**GU-GC** **GU-CG** **GU-UA** TOTAL [a]	243 49 1 293	82.935 16.724 0.341	**GU-GC** TOTAL [a]	32 32	100

The data were obtained from the available rRNA alignments[2]. For all cases of AGPM, the *E. coli* numbering is used.
a: For the statistics, only those cases where the identities of all four nucleotides are known have been considered.

b: N is the number of sequences which have the corresponding 0-base pairs combination.
Central base pairs in bold are those following the GU-WC pattern. Bold red indicates a GU↔WC; WC↔GU replacement.

Supplemental Table IV. Statistical data of the identities of the +1-base pairs for all known AGPMs

AGPM	base pairs [+1P;+1Q]-[+1R;+1S]	Eubacteria		base pairs [+1P;+1Q]-[+1R;+1S]	Archaebacteria			AGPM	base pairs [+1P;+1Q]-[+1R;+1S]	Eubacteria		base pairs [+1P;+1Q]-[+1R;+1S]	Archaebacteria	
		N [b]	%		N [b]	%				N [b]	%		N [b]	%
S62	GC-AG	10565	96.299	GC-AG	417	79.127		L657	AU-GU	191	48.111	GC-GC	22	59.459
	UA-AG	143	1.303	AU-AG	68	12.903			UA-GU	99	24.937	GC-GU	15	40.541
	AU-AG	94	0.857	GU-AG	36	6.831			GC-GU	89	22.418	TOTAL [a]	37	
	GC-AC	38	0.346	UC-AG	1	0.190			CG-GU	12	3.023			
	GU-AG	36	0.328	GC-CU	1	0.190			GC-GC	2	0.504			
	CG-AG	20	0.182	UU-AG	1	0.190			AU-GC	1	0.252			
	GG-AG	15	0.137	GC-CG	1	0.190			UU-GA	1	0.252			
	GC-CG	10	0.091	GC-AA	1	0.190			GC-UA	1	0.252			
	GC-AA	8	0.073	GG-AG	1	0.190			UA-AU	1	0.252			
	GC-GC	6	0.055	TOTAL [a]	527				TOTAL [a]	397				
	UC-AG	6	0.055											
	GC-GG	5	0.046											
	CC-AG	4	0.036											
	GC-AU	4	0.036											
	AC-AG	4	0.036											
	GA-AG	3	0.027											
	GC-UG	2	0.018											
	GC-CA	2	0.018											
	AA-AG	1	0.009											
	GC-CU	1	0.009											
	CA-AG	1	0.009											
	UA-GG	1	0.009											
	UA-AC	1	0.009											
	AU-AU	1	0.009											
	TOTAL [a]	10971												
S296	AG-AG	5794	97.690	AG-AG	250	62.814		L839	GC-CG	194	49.114	GC-GC	9	25.000
	AU-AG	40	0.674	AG-AA	124	31.156			GC-UA	90	22.785	GC-CG	8	22.222
	AA-AG	23	0.388	AG-CG	9	2.261			AU-CG	59	14.937	GC-AU	7	19.444
	AG-CG	15	0.253	AG-CA	6	1.508			AU-UA	17	4.304	CG-CG	7	19.444
	AG-AA	15	0.253	AG-UA	4	1.005			GU-CG	13	3.291	UG-CG	3	8.333
	AG-UG	15	0.253	AG-GA	1	0.251			GU-UA	7	1.772	GU-CG	1	2.778
	AG-GG	10	0.169	GG-AA	1	0.251			AU-UG	3	0.759	CG-GC	1	2.778
	AG-AC	6	0.101	CU-CA	1	0.251			UA-GC	3	0.759	TOTAL [a]	36	
	CU-AG	6	0.101	AG-AU	1	0.251			AU-GC	2	0.506			
	GU-UA	3	0.051	AG-UG	1	0.251			UU-CG	2	0.506			
	AG-AU	2	0.034	TOTAL [a]	398				GC-GC	1	0.253			
	GG-AG	1	0.017						GA-CG	1	0.253			
	AC-AG	1	0.017						UG-UA	1	0.253			
	TOTAL [a]	5931							UA-UA	1	0.253			
									GC-AU	1	0.253			

									TOTAL [a]	395				
S549	CG-UA	4505	49.429	UG-CG	372	75.152		L1864	AU-UA	231	75.987	Replaced by GNRA[1]		
	CG-GC	3228	35.418	CG-CG	80	16.162			AC-UA	57	18.750			
	CG-CG	930	10.204	UG-UA	32	6.465			AA-UA	8	2.632			
	CG-AU	276	3.028	GC-UA	5	1.010			CC-UA	3	0.987			
	CG-GU	32	0.351	GG-CG	1	0.202			AG-UA	2	0.658			
	GG-GC	20	0.219	UA-CG	1	0.202			CU-UA	2	0.658			
	UA-UA	13	0.143	UG-GG	1	0.202			GU-UA	1	0.329			
	CG-UC	10	0.110	UG-CC	1	0.202			TOTAL [a]	304				
	UG-GC	10	0.110	UC-UA	1	0.202								
	CG-UG	8	0.088	UG-UG	1	0.202								
	CG-GA	7	0.077	TOTAL [a]	495									
	UA-CG	7	0.077											
	CG-CA	7	0.077											
	UG-CG	6	0.066											
	CC-UA	5	0.055											
	GG-UA	5	0.055											
	CG-AC	4	0.044											
	AG-UA	4	0.044											
	UG-UA	4	0.044											
	CC-GC	4	0.044											
	AG-GC	3	0.033											
	CG-GG	3	0.033											
	CG-AA	3	0.033											
	CU-UA	3	0.033											
	CG-UU	2	0.022											
	CG-CC	2	0.022											
	UA-GC	2	0.022											
	GG-CG	2	0.022											
	GG-GA	1	0.011											
	CU-AU	1	0.011											
	CU-GC	1	0.011											
	CU-CC	1	0.011											
	CC-AU	1	0.011											
	UG-GA	1	0.011											
	UG-CU	1	0.011											
	CA-GC	1	0.011											
	UA-AU	1	0.011											
	TOTAL [a]	9114												
S757	AG-GC	8952	76.324	AG-GC	526	91.638		L2291	AU-AG	251	63.544	GC-AG	31	83.784
	AC-GC	2237	19.072	AG-AU	18	3.136			UA-AG	73	18.481	AU-AG	5	13.514
	AA-GC	378	3.223	AA-GC	16	2.787			CG-AG	46	11.646	GC-GG	1	2.703
	AU-GC	59	0.503	AC-AU	4	0.697			GC-AG	19	4.810	TOTAL [a]	37	
	AU-AU	33	0.281	AA-AU	3	0.523			UA-CG	4	1.013			
	AG-GU	11	0.094	AC-GC	3	0.523			GU-AG	2	0.506			
	GG-GC	8	0.068	AG-AC	2	0.348			TOTAL [a]	395				

	AG-AC	7	0.060	GG-GC	1	0.174								
	AC-GU	5	0.043	AG-CC	1	0.174								
	CG-GC	4	0.034	TOTAL [a]	574									
	AG-GA	4	0.034											
	GU-GC	4	0.034											
	AG-GG	4	0.034											
	UG-GC	3	0.026											
	AA-AU	3	0.026											
	AC-CC	3	0.026											
	CC-GC	2	0.017											
	AC-AC	2	0.017											
	GC-GC	2	0.017											
	AG-AU	1	0.009											
	UC-GC	1	0.009											
	AC-GG	1	0.009											
	AA-GA	1	0.009											
	AC-GA	1	0.009											
	AG-UC	1	0.009											
	AG-CC	1	0.009											
	GG-GU	1	0.009											
	TOTAL [a]	11729												
S911	GC-AU	5554	69.897	GU-AU	197	54.722		L2687	AG-GC	134	40.854	AU-GC	11	29.730
	GC-GC	1958	24.641	GU-UA	98	27.222			AG-CG	118	35.976	AC-GC	8	21.622
	GC-GU	192	2.416	GU-GC	53	14.722			GG-GC	21	6.402	AG-GC	7	18.919
	GU-AU	181	2.278	GU-CG	4	1.111			AG-GU	16	4.878	AG-CG	7	18.919
	GC-AC	11	0.138	GU-GU	2	0.556			AG-AU	13	3.963	GU-GC	3	8.108
	GC-AG	8	0.101	GC-UA	2	0.556			AG-UA	11	3.354	AU-CG	1	2.703
	GC-GG	6	0.076	GU-GG	1	0.278			AA-GC	8	2.439	TOTAL [a]	37	
	GC-UC	5	0.063	GU-UG	1	0.278			UG-GC	3	0.915			
	GC-AA	5	0.063	GU-AA	1	0.278			AU-CG	1	0.305			
	GU-UA	3	0.038	GU-AG	1	0.278			GG-CG	1	0.305			
	GG-AU	3	0.038	TOTAL [a]	360				AG-UG	1	0.305			
	GC-CG	3	0.038						AU-UA	1	0.305			
	GU-GU	2	0.025						TOTAL [a]	328				
	UC-AU	2	0.025											
	GC-CC	2	0.025											
	GC-GA	2	0.025											
	CC-AU	2	0.025											
	AC-AU	1	0.013											
	GC-UA	1	0.013											
	GU-GC	1	0.013											
	AC-GU	1	0.013											
	GC-UU	1	0.013											
	GU-AC	1	0.013											
	GA-GC	1	0.013											
	TOTAL [a]	7946												

L554	AG-GC	129	38.279	GC-GU	10	30.303		L2698	GC-GC	107	32.622	GC-GC	18	48.649
	AG-CG	77	22.849	GC-GC	8	24.242			AU-AU	64	19.512	GC-AU	8	21.622
	AG-AU	47	13.947	GU-GC	6	18.182			GC-AU	33	10.061	GC-GU	6	16.216
	AG-UA	28	8.309	GC-CG	4	12.121			GU-AU	31	9.451	AU-GC	4	10.811
	AG-GU	21	6.231	AU-GC	2	6.061			GC-GU	28	8.537	GC-CG	1	2.703
	AA-UA	14	4.154	GU-CG	1	3.030			CG-GC	20	6.098	TOTAL [a]	37	
	AA-CG	14	4.154	AU-GU	1	3.030			AU-GU	11	3.354			
	AA-GC	2	0.593	GC-UA	1	3.030			UA-AU	9	2.744			
	AA-GU	2	0.593	TOTAL [a]	33				CG-GU	8	2.439			
	AU-UA	1	0.297						UG-GC	4	1.220			
	UA-UA	1	0.297						GU-GC	4	1.220			
	GU-GU	1	0.297						UG-AU	3	0.915			
	TOTAL [a]	337							UA-GU	2	0.610			
									AU-GC	2	0.610			
									AU-UU	1	0.305			
									GU-GU	1	0.305			
									TOTAL [a]	328				
L639	GC-AG	352	89.114	GC-AG	26	70.270		L2847	AG-UA	158	53.925	AG-CG	32	100
	AU-AG	21	5.316	CG-AG	10	27.027			AG-CG	105	35.836	TOTAL [a]	32	
	CG-AG	11	2.785	UA-AG	1	2.703			AG-AU	27	9.215			
	GC-AA	8	2.025	TOTAL [a]	37				AA-UA	3	1.024			
	GU-AG	3	0.759						TOTAL [a]	293				
	TOTAL [a]	395												

The data were obtained from the available rRNA alignments[2]. For all cases of AGPM, the *E. coli* numbering is used.
a: For the statistics, only those cases where the identities of all four nucleotides are known have been considered.
b: N is the number of sequences which have the corresponding +1-base pairs combination.

5.11 Supplemental references

1. Mokdad, A., Krasovska, M. V., Sponer, J. & Leontis, N. B. (2006). Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res* **34**, 1326-41.
2. Wuyts, J., Perriere, G. & Van De Peer, Y. (2004). The European ribosomal RNA database. *Nucleic Acids Res* **32**, D101-3.

Chapitre 6

Article 5

The adenosine wedge: A new structural motif in ribosomal RNA

6. The adenosine wedge: A new structural motif in ribosomal RNA

Matthieu G. Gagnon and Sergey V. Steinberg

Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Manuscript submitted.

Running title: Adenosine-wedge motif

Contribution of each author:

Matthieu G. Gagnon: Participated in the *in silico*, statistical and data analysis and in the preparation of the manuscript and figures; also performed the search within the ribosome structure for other similar arrangements.

Sergey V. Steinberg: Participated in the *in silico* and data analysis, preparation of the manuscript and figures.

6.1 Abstract

We present here a new recurrent RNA arrangement, the so-called adenosine-wedge motif, which is found in three places in the ribosomal RNA of both ribosomal subunits. The arrangement has a hierarchical structure, consisting of elements previously described as recurrent motifs, namely, the along-groove packing motif, the A-minor and the hook-turn. Within the adenosine wedge, these elements are involved in different types of cause-effect relationships, providing together for the unique tertiary structure of the motif.

Keywords: RNA structure; RNA motif; ribosomal RNA; A-minor; Along-Groove Packing Motif

6.2 Description of the adenosine-wedge motif

An essential part of the knowledge on RNA structure is collected in the form of recurrent motifs. Recurrent motifs are found in different molecules or in different parts of the same molecule and have identical or very similar tertiary structure (for review, see Batey et al. 1999; Moore 1999; Noller 2005). Analysis of recurrent motifs is helpful in establishing the major principles governing the formation of RNA tertiary structure. Here we present a new motif, the so-called adenosine wedge, which is found in three locations in the ribosomal RNA and can be considered as a derivative of three previously identified recurrent motifs, the along-groove packing motif (AGPM) (Gagnon and Steinberg 2002; Mokdad et al. 2006), the A-minor motif (Doherty et al. 2001; Nissen et al. 2001) and the hook-turn motif (Szép et al. 2003).

The main body of the adenosine-wedge motif is formed by the AGPM, which refers to the arrangement of two closely packed double helices positioned such that the backbone of one helix interacts with the minor groove of the other helix and *vice versa* (Fig. 1A and Supplemental Fig. 1). The AGPM has been found in more than a dozen locations in ribosomal RNA (Gagnon and Steinberg 2002; Mokdad et al. 2006). The nomenclature of the helices, strands, levels and individual nucleotides of the AGPM is

given in Figure 1B. The structure of the AGPM is characterized by the axial symmetry shown in Figure 1A. This symmetry is usually disturbed at the zero level, where the two helices interact most closely and where a WC base pair of one helix is often packed against a GU base pair of the other helix (Fig. 1C and Supplemental Fig. 1). The co-existence of the GU and WC base pairs at the zero level makes the shapes of the two helices complementary to each other and extends the area of their close packing. In each helix, this area spreads over four layers between -2 and +1.

In the ribosome structure, the adenosine-wedge motif is associated with three cases of AGPM, S549 (30S subunit), L639 and L657 (both in the 50S subunit) (Fig. 1D). The essential element of the adenosine wedge that makes it different from other cases of AGPM consists in the particular position of nucleotides -2Q and -2S, which are both adenosines in all three cases of the motif existing in the *Escherichia coli* ribosome (Schuwirth et al. 2005). Instead of being involved in base pairing with, respectively, -2P and -2R, these adenosines stack in the area between the two double helices roughly perpendicular to nucleotides of both helices (Figs. 2A, 3A). Such an arrangement could be seen as a wedge represented by the adenosine stack [-2Q;-2S] that is packed in the crack between the two helices.

Within the adenosine-wedge motif, the two adenosines -2Q and -2S form contacts with nucleotides -2P and -2R, respectively. Both contacts resemble the A-minor motif and include inter-base van der Waals interactions and hydrogen bonds (Fig. 3B, C). In particular, two symmetrical hydrogen bonds are observed between atoms N1 of both adenosines -2Q and -2S and the ribose O2′-H groups of -2P and -2R, respectively (Fig. 3B, C). These hydrogen bonds are specific to the Type-I A-minor interaction (Nissen et al. 2001). Thus, a unique feature of the adenosine-wedge motif consists in the fact that although the two adenosines -2Q and -2S stack on each other, they form A-minor interactions with two different double helices. Other interactions that involve adenosines [-2Q;-2S] are discussed in the following sections (see also Supplemental Table I).

6.3 Asymmetry of the adenosine-wedge motif

Based on the given description of the adenosine-wedge motif, one could suggest that the adenosine stack [-2Q;-2S] is arranged symmetrically with respect to both double helices. However, this arrangement is characterized by the presence of an essential asymmetrical component. Above, we have already mentioned the asymmetry between the two double helices within the AGPM that relates to the presence of the GU versus WC base pairs at the zero level (Fig. 1C and Supplemental Fig. 1) (Gagnon and Steinberg 2002). This type of asymmetry is observed in most cases of AGPM, where it guarantees the close packing of the two helices. However, the asymmetry related to the structure of the adenosine-wedge motif has a different origin; it is attributed only to those cases of AGPM that form the adenosine wedge and it is completely independent of the GU-versus-WC asymmetry at the zero level.

First, we noticed a difference in the structure of the two presumed symmetrical external strands P and R. While nucleotide -2P and the following nucleotide -3P in all three cases of the adenosine-wedge motif stack together, nucleotide -2R, which is symmetrical to -2P, never interacts with the following nucleotide. In all three cases of the motif existing in the *E. coli* ribosome (Schuwirth et al. 2005), the nucleotide following -2R occupies position -2Q, which makes the two strands R and Q directly connected (Fig. 1D). In fact, this asymmetry between the external strands P and R was used for the initial definition of which of the two helices should be named Helix 1 and 2 (Fig. 1B). Below, we demonstrate that the presence of nucleotide -3P and the absence of nucleotide -3R are essential for the structure of the adenosine-wedge arrangement. Moreover, this asymmetry leads for other asymmetrical aspects specific to the adenosine-wedge structure.

6.4 Displacement of the adenosine stack

Further analysis showed that the universal absence of one of the two external nucleotides at the -3 level (nucleotide -3R) is critical for the integrity of the adenosine wedge arrangement. As one can see in Figures 2A, 3A, adenosine stack [-2Q;-2S] is located in

the area between the two external strands P and R. One can imagine a situation when in both strands, the nucleotide at the -2 level (-2P and -2R) and the following nucleotide (-3P and -3R) stacked on each other. In such a case, the space between the riboses of -3P and -3R would have been too narrow to accommodate the adenosine stack [-2Q;-2S] (Fig. 2B). However, the absence of one of the two external nucleotides at the -3 level (nucleotide -3R; black in Fig. 2B) opens the space for the adenosine stack [-2Q;-2S] to be displaced by about 2 Å from the symmetrical position toward strand R and to take the space originally designated to the missing nucleotide -3R. Due to this displacement, the adenosine stack avoids the collision with the remaining external nucleotide at the -3 level (nucleotide -3P).

The described displacement of adenosines [-2Q;-2S] toward strand R leads to other asymmetric aspects of the adenosine-wedge motif. In particular, it brings nucleotide -2Q close enough to -2R to allow the covalent connection between them. This connection makes Helices 1 and 2 immediate neighbors in the rRNA secondary structure. Simultaneously, the distance between the symmetrically positioned nucleotides -2P and -2S becomes longer, which makes the direct connection between -2P and -2S impossible. The existence of the covalent link between nucleotides -2R and -2Q provides a strong stabilizing effect for the whole arrangement. Structurally, this connection represents a bent of the polynucleotide chain by almost 180°, which can be described in terms of the recently identified hook-turn motif (Szép et al. 2003) (see Fig. 2 and Supplemental Fig. 2). This conformation of the connector allows the formation of the asymmetrical hydrogen bond between the phosphate group of -2Q, which is positioned between bases -2R and -2Q, and the amino group of adenosine -2S (see Supplemental Table I).

Another consequence of the adenosine stack displacement consists in the shortening of the distance between the ribose of -2S and nucleotide -2R and the simultaneous extension of the distance between the symmetrically positioned ribose -2Q and nucleotide -2P. This type of asymmetry is reflected in the different patterns of H-bonding between the ribose of -2Q and nucleotide -2P and between the ribose of -2S and nucleotide -2R (the corresponding hydrogen bonds are listed in Supplemental Table I and shown in Figs. 3B,C and 4B,C).

6.5 Inclination of the adenosine stack

Even though adenosines -2Q and -2S have been displaced from their symmetrical positions toward strand R, they still maintain their interactions with the symmetrically positioned external nucleotides -2P and -2R. To make these interactions possible, the bases of both adenosines become inclined by about 35°-40° in the direction opposite to that in which these nucleotides were originally displaced (Fig. 2 and Supplemental Fig. 2). Although such inclination of the adenosine bases [-2Q;-2S] restores their contacts with nucleotides -2P and -2R, it introduces the asymmetry to these interactions.

In Helix 1, this inclination opens the angle between the bases of -2Q and -2P, thus facilitating the H-bonding between them. In motifs L639 and L657, in addition to the discussed above hydrogen bond between atom N1 of -2Q and the O2′-H group of -2P, the inclination of the bases allows for the formation of another hydrogen bond between atom N3 of adenosine -2Q and the amino group of guanosine -2P (Fig. 3B). The two hydrogen bonds correspond to the previously described Type-I A-minor interaction pattern (Nissen et al. 2001). In motif S549, position -2P is occupied by uridine, which is unable to form a hydrogen bond with adenosine -2Q equivalent to that observed in motifs L639 and L657. However, because this uridine is involved in the reverse-WC base pair with adenosine A397 of 16S rRNA, the amino group of the latter adenosine forms the equivalent hydrogen bond with atom N3 of adenosine -2Q (Fig. 3C).

While opening the angle between the bases of -2Q and -2P, the above-mentioned inclination of adenosines [-2Q;-2S] simultaneously sharpens the angle between the symmetrically positioned bases of -2S and -2R, thus preventing the formation of the mirror hydrogen bond between them. Still, regardless of the angle between these bases, a van der Waals contact between them is still possible. Consequently, of the three cases of the adenosine-wedge motif existing in the *E. coli* ribosome (Schuwirth et al. 2005), only in L639 is position -2R occupied by guanosine, which forms a rather weak hydrogen bond with adenosine -2S (Fig. 4B). In motifs S549 and L657, nucleotide -2R is adenosine, which makes a van der Waals contact with the base of -2S without H-bonding in both cases.

6.6 Role of nucleotide -3P

The inclined position of the adenosine stack [-2Q;-2S] is also stabilized by the interaction of the base of -2Q with the ribose of -3P (Fig. 2 and Supplemental Fig. 2). As mentioned above, nucleotide -3P exists in all three cases of the adenosine-wedge motif found in the *E. coli* ribosome (Schuwirth et al. 2005), and it always stacks to the previous nucleotide -2P. This position of nucleotide -3P orients its ribose toward the base of -2Q, thus promoting the interaction between the two. This interaction also stabilizes the inclined position of adenosine -2Q as well as of its partner adenosine -2S with respect to the rest of the adenosine-wedge motif. Thus, not only the universal absence of nucleotide -3R, but also the universal presence of its symmetrical analog -3P is important for the integrity of the adenosine-wedge motif.

6.7 Variations in the structure of the adenosine-wedge motif

In the structure of motif L639 existing in the 50S subunit of *Haloarcula marismortui* (Ban et al. 2000), there is a 31-nucleotide insertion 699-729 between positions -2Q and -1Q (Fig. 4A). This insertion virtually eliminates the hook-turn connection between strands R and Q, thus making Helices 1 and 2 no longer neighbors in the 23S rRNA secondary structure. All other elements of this arrangement stay at the same places as in the standard structure of the adenosine-wedge motif (Fig. 4B). Analysis of the available nucleotide sequences of 23S rRNA shows that the same insertion exists in all archaeal ribosomes (Wuyts et al. 2004). The existence of such an insertion shows that the hook-turn, which is formed if strands R and Q are immediately connected, is not a necessary element of the adenosine-wedge arrangement. In the particular case of motif L639, the insertion forms a hairpin, so that the two helices constituting the motif are still neighbors in the secondary structure.

Interestingly, in the same case of motif L639 from *H. marismortui*, the last nucleotide of the 31-nucleotide insertion is cytidine C729, which forms a WC base pair with guanosine G742 occupying position -2P. As a result, nucleotide -2P (G742) forms hydrogen bonds simultaneously with two nucleotides A698 and C729, which approach

-2P from different sides. For this reason, both nucleotides A698 and C729 are named in Figure 4B as nucleotides -2Q.

Among cases of AGPM not associated with the adenosine wedge, there is one, S62, which contains some wedge elements (Fig. 4A). In particular, in motif S62 taken from the *E. coli* ribosome (Schuwirth et al. 2005), adenosine -2Q (A60 in 16S rRNA) interacts with guanosines -2P (G107) and -3P (G108) exactly as it does in the standard adenosine-wedge pattern (Fig. 4C). Nucleotide -2S, however, does not stack to -2Q, but instead, forms a WC base pair with -2R. This particular arrangement can thus be seen as a half of the whole adenosine-wedge arrangement. It demonstrates that to maintain the specific position of adenosine -2Q, its interaction with nucleotides -2P and -3P is sufficient, while the interactions with nucleotides of Helix 2 are not required.

6.8 Conservation of the elements forming the adenosine-wedge motif

Analysis of the 12 697 and 436 available prokaryotic nucleotide sequences of, respectively, 16S and 23S rRNA (Wuyts et al. 2004) shows that the structural elements responsible for the integrity of the adenosine-wedge motif are highly conserved among prokaryotes. Thus, for each of the three motifs S549, L639 and L657, each of the two adenosines -2Q and -2S is present in more than 99% of the sequences (Supplemental Table II).

Also, in most nucleotide sequences, the identity of nucleotide -2P is such that it would allow the proper fixation of adenosine -2Q. In particular, in motifs L639 and L657, guanosine in position -2P is present in 93.2% of all 23S rRNA sequences, and the exceptions almost exclusively contain cytidine (3.3%) or adenosine (3.3%) (Supplemental Table II). The presence of guanosine in position -2P would allow the maintenance of its hydrogen bond with adenosine -2Q. All instances of cytidine in position -2P are found exclusively in motif L639 from archaea. As mentioned above, in all nucleotide sequences of archaeal 23S rRNA, there is an insertion between positions -2Q and -1Q, and the last nucleotide of this insertion (nucleotide 729) forms a base pair with nucleotide 742 in position -2P. Analysis shows that base pair 742-729 is WC in all 37 available nucleotide sequences of archaeal 23S rRNA. In 9 cases, this base pair is

G742-C729, while in the other 28 cases it is C742-G729. Therefore, in all cases when position -2P is occupied by a cytidine, this cytidine forms a CG base pair with guanosine G729. The latter guanosine will be able to form a hydrogen bond with atom N3 of adenosine -2Q in the same way as guanosine in position -2P (Fig. 4B). In the remaining 3.3% of the sequences when nucleotide -2P is adenosine, it forms a van der Waals contact with adenosine -2Q, as in motif L657 of the 50S subunit from *H. marismortui* (pdb entry code 1s72) (Ban et al. 2000) (not shown). Finally, in motif S549 the UA combination between uridine in position -2P and adenosine A397 (Fig. 3C) is present in more than 99% of the sequences, and this allows the maintenance of the hydrogen bond between adenosines -2Q and A397 (Supplemental Table II).

As mentioned above, the inclination of the adenosine stack [-2Q;-2S] toward nucleotide -2P would weaken the hydrogen bond between adenosine -2S and guanosine -2R. Correspondingly, the presence of guanosine in position -2R is relatively low, reaching 62.7% in motifs L639 and L657 and only 33.7% in S549. At the same time, the percentage of adenosine in position -2R becomes relatively high, reaching 35.4% in L639 and L657 and almost 51% in S549 (Supplemental Table II). Moreover, in 15% of the sequences, position -2R in S549 is occupied by a pyrimidine, which virtually eliminates the contact of its base with -2S. All these data indicate a relatively low conservation of the contact between -2R and -2S.

To conclude, the variations in the identities of the key nucleotides in the adenosine-wedge arrangement would guarantee the integrity of all cases of the motif. Among all key nucleotides, adenosines -2Q and -2S are the most conserved. A somewhat lower conservation is observed for position -2P, although even here, in the overwhelming majority of the sequences, the identity of this nucleotide or of the nucleotide with which -2P forms a base pair still allows the formation of the specific hydrogen bond with atom N3 of adenosine -2Q. Although in most rRNA sequences, nucleotide -2R is able to form an interaction with adenosine -2S, the level of conservation of -2R is notably lower than that of all other nucleotides at the -2 level. We suggest that such a low conservation of -2R reflects the fact that due to the particular position at the outskirt of the structure, the base of this nucleotide cannot contribute

substantially to the stability of the arrangement and thus plays a relatively minor role in the integrity of the motif.

6.9 NAG-triangle at the core of the adenosine-wedge motif

Based on the analysis of the structures of the standard adenosine wedge as well as of its modifications, including the one observed in the AGPM S62, we suggest that the arrangement encompassing three nucleotides -3P, -2Q and -2P plays the central role in the formation of the motif. Indeed, this arrangement represents the only part of the structure that is present in all variations of the adenosine-wedge motif. When nucleotide -2P is guanosine, which happens in most cases, all three nucleotides of the [-3P;-2Q;-2P] arrangement become involved in tight pair-wise interactions with the other two nucleotides: -3P stacks to -2P and donates the ribose for stacking with -2Q, while -2Q and -2P form two hydrogen bonds and several van der Waals contacts. To underline the importance of these pair-wise interactions and to reflect the identities of all three participating nucleotides, we will call such a three-nucleotide arrangement the NAG-triangle (N – any nucleotide in -3P, A – adenosine in -2Q, G – guanosine in -2P).

In different cases of the adenosine-wedge motif, all modifications of the structure of the NAG-triangle are restricted to variations in the identity of the nucleotide at position -2P. When guanosine -2P is replaced by adenosine, the triangle structure would remain virtually the same, although somewhat less stable. A lower stability explains the low frequency of adenosines in position -2P. However, when -2P becomes a pyrimidine, the damaging effect of such a replacement would become more profound. First, the area of contact between pyrimidine -2P and nucleotide -2Q will be reduced to a virtual non-existence. Also, it is known that a pyrimidine stacks poorly to the following nucleotide (Kotlova et al. 2007), which would severely compromise the interaction between -2P and -3P. However, the formation of either a WC (motif L639 in archaea) or reverse-WC (motif S549 in all prokaryotes) base pair between -2P and another nucleotide of rRNA in the case when nucleotide -2P is a pyrimidine can effectively solve both problems. Indeed, such additional nucleotide will be able to glue all three nucleotides -2P, -3P and

-2Q together through the base pairing with -2P, the stacking with -3P and the H-bonding with -2Q (Figs. 3C, 5B).

Thus, in all cases of the adenosine-wedge motif, the NAG-triangle and its modifications NAG(G→A) and NAG(G→Y=R base pair) represents a compact arrangement in which each element tightly interacts with the other two elements. We could expect that the importance of the NAG-triangle goes beyond the adenosine-wedge motif, so that these arrangements would be found in other structural contexts as well. To test this suggestion, we performed a search of all arrangements existing in the *E. coli* ribosome structure (pdb entry codes 2avy-2aw4) (Schuwirth et al. 2005) in which three nucleotides were juxtaposed as nucleotides -2Q, -2P and -3P in the adenosine-wedge motif (see Methods). Within the ribosome, in addition to the cases discussed above, we found twenty-three new arrangements of this kind unrelated to the AGPM (see Supplemental Table III and Supplemental Figure 3). In all these arrangements, the nucleotide corresponding to -2Q was invariably adenosine, while the nucleotide corresponding to -2P was guanosine, adenosine and cytidine in 13, 4 and 6 cases, respectively. Interestingly, in five out of six cases of cytidine in position -2P, this nucleotide was involved in a WC base pair. In all these cases, the corresponding guanosine provided its amino group for the hydrogen bond with the adenosine in position -2Q, thus forming a structure analogous to that found in the archaeal motif L639. The other nucleotides that arranged around the NAG-triangle did not, however, show a common pattern, which demonstrates that the NAG-triangle could be formed in different structural contexts, and not only in those related to the AGPM. The latter aspect would allow us to qualify the NAG-triangle itself as a new recurrent RNA motif.

6.10 Concluding remarks

The adenosine-wedge motif represents an RNA arrangement of four repetitive elements, the AGPM, the A-minor, the hook-turn and the NAG-triangle, which are involved in complex cause-effect relationships between themselves. The scaffold for the adenosine-wedge motif consists of two symmetrically arranged double helices forming the AGPM. However, the fitting of the adenosine stack [-2Q;-2S] into the space between these

helices faces a steric problem and thus requires that only one of the two external nucleotides at the -3 level (-3P) be stacked to the previous nucleotide -2P. The presence of nucleotide -3P and the absence of the corresponding nucleotide in position -3R unleash, in their turn, a chain of cause-effect relationships that eventually leads to an essential asymmetry of the motif. The core element of the adenosine-wedge motif, the NAG-triangle, has a completely asymmetric structure.

Each of the four elements that constitute the adenosine-wedge arrangement can also be found in other unrelated structural contexts. It seems obvious that such hierarchical structure is not specific to the adenosine-wedge motif and can be observed in other RNA arrangements as well. In most of these cases, different elements are expected to be involved in complex cause-effect relationships with other parts of the structure. Elucidation of the nature of these relationships is a necessary step toward understanding the principles of RNA folding.

6.11 Methods

The collection of the statistical data on the identity of nucleotides involved in the adenosine-wedge motif and the search for similar tertiary arrangements in the *E. coli* ribosome were performed as described in the Methods of Chapter 5.

6.12 Acknowledgments

The work was supported by an operating grant from Canadian Institutes of Health Research.

6.13 References

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**:905-920.

Batey RT, Rambo RP, Doudna JA. 1999. Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed. Engl.* **38**:2326-2343.

Doherty EA, Batey RT, Masquida B, Doudna JA. 2001. A universal mode of helix packing in RNA. *Nat. Struct. Biol.* **8**:339-343.

Gagnon MG, Steinberg SV. 2002. GU receptors of double helices mediate tRNA movement in the ribosome. *RNA* **8**:873-877.

Klein DJ, Schmeing TM, Moore PB, Steitz TA. 2001. The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**:4214-4221.

Kotlova N, Ishii TM, Zagryadskaya EI, Steinberg SV. 2007. Active suppressor tRNAs with a double helix between the D- and T-loops. *J. Mol. Biol.* **373**:462-475.

Mokdad A, Krasovska MV, Sponer J, Leontis NB. 2006. Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.* **34**:1326-1341.

Moore PB. 1999. Structural motifs in RNA. *Annu. Rev. Biochem.* **68**:287-300.

Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. 2001. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. USA* **98**:4899-4903.

Noller HF. 2005. RNA structure: reading the ribosome. *Science* **309**:1508-1514.

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH. 2005. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**:827-834.

Steinberg SV, Boutorine YI. 2007. G-ribo: a new structural motif in ribosomal RNA. *RNA* **13**:549-554.

Szép S, Wang J, Moore PB. 2003. The crystal structure of a 26-nucleotide RNA containing a hook-turn. *RNA* **9**:44-51.

Wuyts J, Perriere G, Van De Peer Y. 2004. The European ribosomal RNA database. *Nucleic Acids Res.* **32**:D101-103.

6.14 Figures

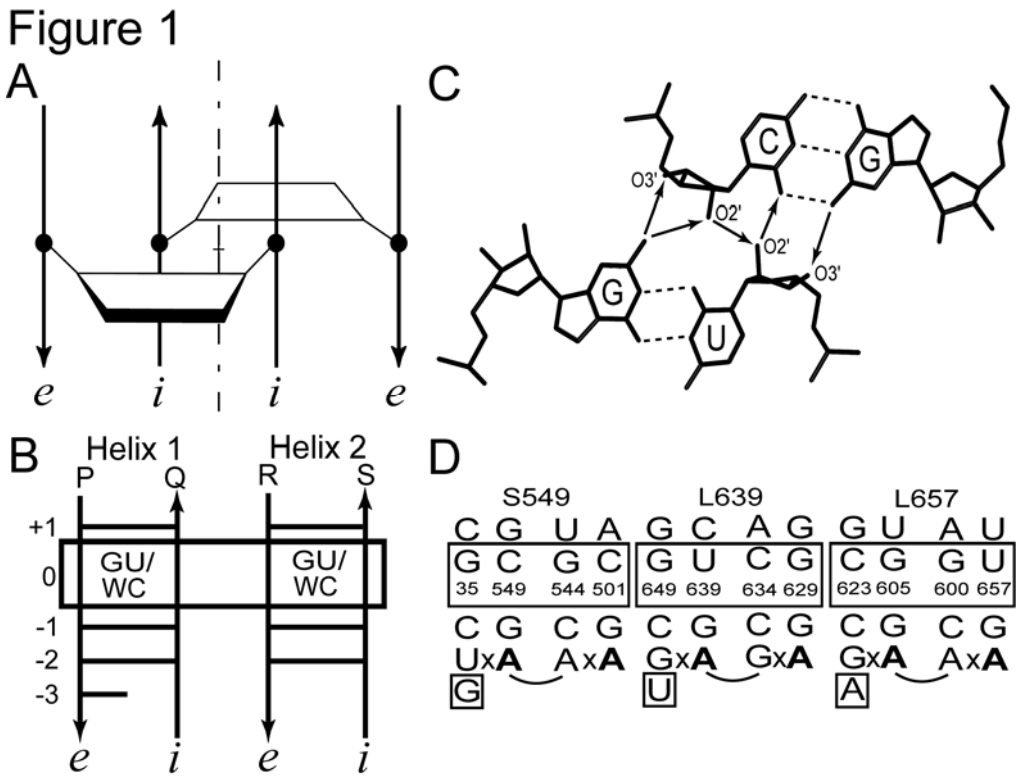


Figure 1. The along-groove packing motif (AGPM).

(A) Schematic representation of the AGPM. Trapezoids represent base pairs opened toward the minor grooves. Arrows represent backbones directed 5′→3′. The internal and external strands of each helix are marked by italic letters *i* and *e*, respectively. The internal strand of each helix interacts with the minor groove of the other helix. Rotation of one helix for 180° around the symmetry axis (dash-dotted line) superposes it with the other helix.

(B) Secondary structure representation of the AGPM. Boxed are the two central base pairs, which are located at the center of the arrangement, where the two helices are packed most closely. For each base pair within each helix, a number is assigned, so that the central base pairs carry number zero, and the positive propagation of the numbering corresponds to the 5′→3′ direction of the internal chains. For the four strands, capital letters P, Q, R and S are assigned. The helix in which the external nucleotide at the -3 level (nucleotide -3P) stack to the previous nucleotide (-2P) is named Helix 1. The opposite helix, in which the equivalent interaction does not exist (nucleotides -2R and -3R do not stack), is named Helix 2.

(C) Juxtaposition of the central base pairs in the AGPM. Arrows stand for inter-helix hydrogen bonds directed from the donor to the acceptor atom. In most known cases of AGPM, one of the central base pairs is WC, while the other one is GU. The characteristic geometry of the GU base pair allows the formation of a complex network of several inter-helix hydrogen bonds.

(D) Nucleotide sequences of the three cases of AGPM that have the adenosine-wedge motif within their structure. The name of each motif starts with letter S or L, depending on subunit, small or large, in which it was found followed by the number of the internal nucleotide in the GU or equivalent base pair at the 0-level (Gagnon and Steinberg 2002). The secondary structure of each case is drawn according to the secondary structure scheme shown in panel B. In each case, the rRNA sequence and the nucleotide numbers are taken from the structure of the *E. coli* ribosome (pdb entry codes 2avy-2aw4) (Schuwirth et al. 2005). Adenosines -2Q and -2S, which compose the adenosine-wedge motif, are shown in bold, while nucleotide -3P, whose ribose interacts with the base of adenosine -2Q, is boxed (see Fig. 2). Sign × stands for the absence of a base pair.

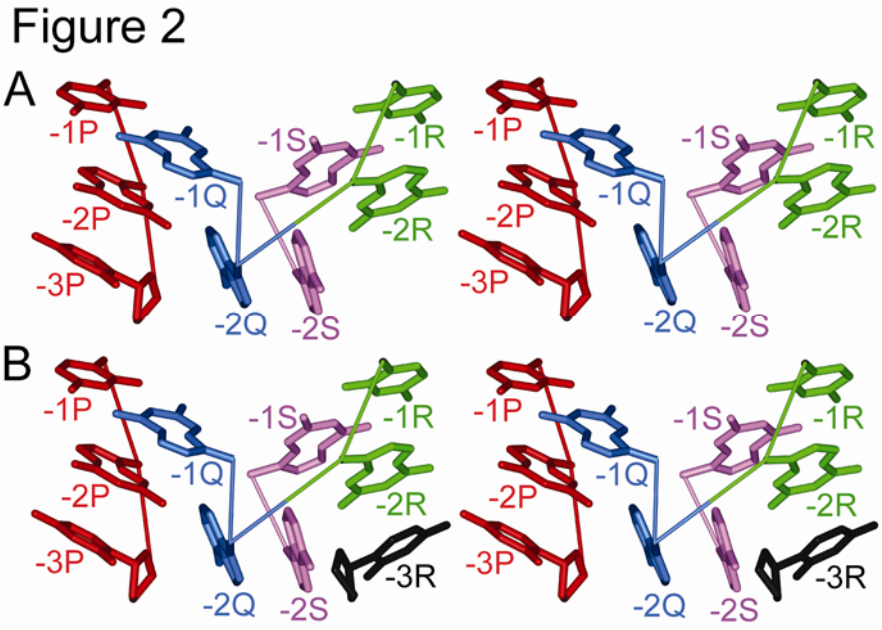
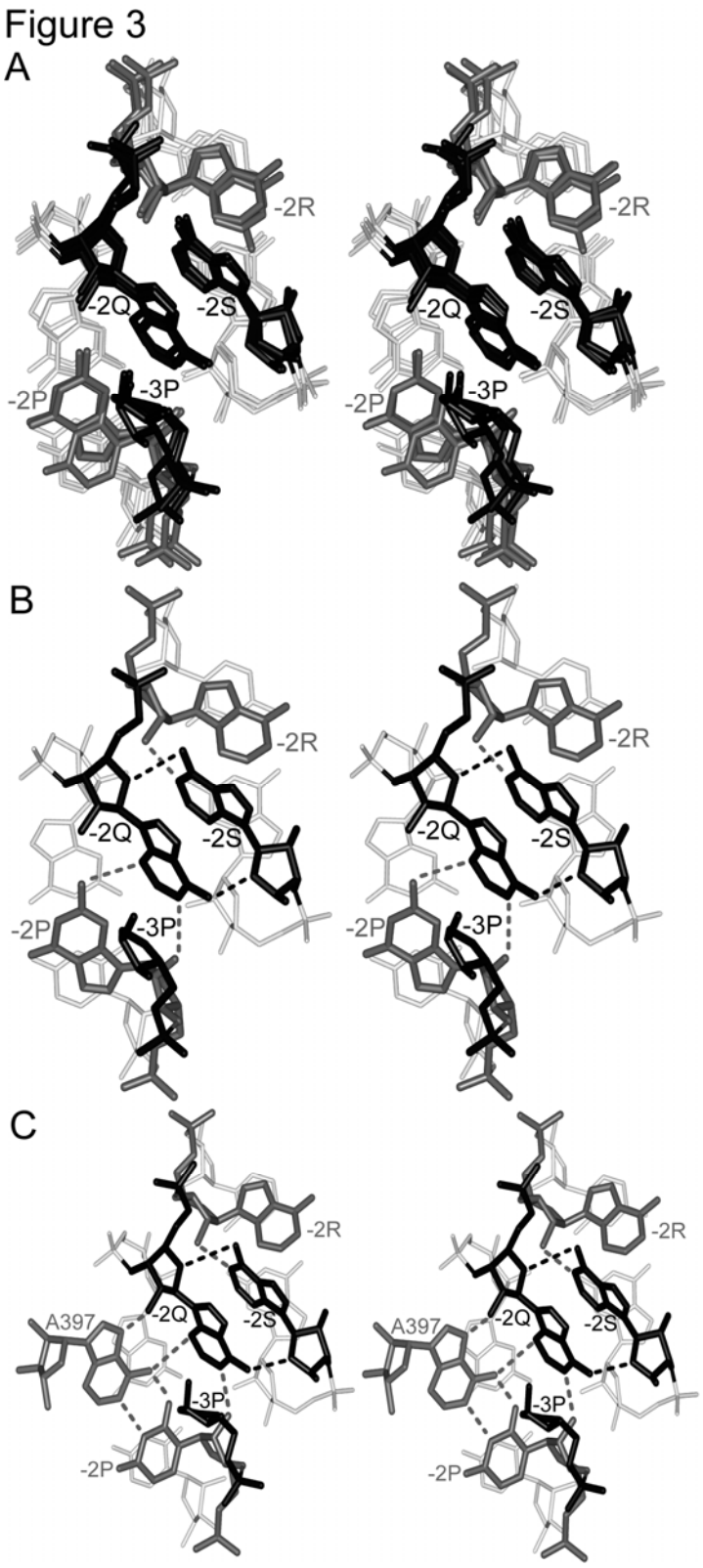


Figure 2. Simplified stereo view representation of the structure of the adenosine-wedge motif.

(A) Adenosines -2Q (blue) and -2S (magenta) are packed in the niche formed by nucleotides -2P (red) and -2R (green) and are oriented about perpendicular to base pairs of the two helices. Such an arrangement could be seen as a wedge formed by the stack of adenosines [-2Q;-2S] and packed in the crack between the two helices. The covalent connection between strands R and Q is possible due to the displacement of the adenosine stack [-2Q;-2S] by about 2 Å from the symmetrical position toward strand R. For the adenosine stack [-2Q;-2S] to preserve its interactions with nucleotides -2P, -3P and -2R, it needs to incline by about 35°-40° from a symmetrical position in the direction opposite to that in which it was originally displaced. In all adenosine-wedge arrangements, the adenosine stack [-2Q;-2S] closely interacts with the ribose of -3P, while the nucleotide position that would follow -2R in the stacked conformation is empty. The complete structure of the adenosine-wedge motif within the context of the AGPM is shown in Supplemental Figure 2.

(B) The position of -3R (black) was obtained through rotation by 180° of the whole arrangement around the symmetry axis. In such a situation, nucleotides -3R and -2S would strongly collide with each other. The color code used is the same as in panel A.

Figure 3



See the legend on next page

Figure 3. Stereo view of the structure of the adenosine-wedge motif.

(A) Superposition of three adenosine-wedge motifs existing in AGPMs S549, L639 and L657 of the *E. coli* ribosome (pdb entry codes 2avy-2aw4) (Schuwirth et al. 2005). The structures are superposed based on the positions of the C1′ atoms of nucleotides -2P, -3P, -2Q and -2S (r.m.s.d. = 0.33 Å). Adenosines -2Q and -2S as well as the ribose of nucleotide -3P are black. Nucleotides -2P and -2R are gray. The base pairs of the -1-layers are white.

(B) The adenosine wedge arrangement existing in the AGPM L657. Two symmetrical hydrogen bonds form between atoms N1 of both adenosines -2Q and -2S and the ribose O2′-H groups of -2P and -2R, respectively. Two more symmetrical hydrogen bonds are formed between the amino group of each of the two adenosines and the O4′ atom of the other adenosine. These hydrogen bonds exist in all known cases of the adenosine wedge. The exo-cyclic amino group of guanosine -2P makes a hydrogen bond with atom N3 of adenosine -2Q. This bond exists in motifs L639 and L657.

(C) The adenosine wedge existing in the AGPM S549. In this case, position -2P is occupied by U37, which forms a reverse-WC base pair with A397. The exo-cyclic amino group of A397 forms a hydrogen bond with atom N3 of adenosine -2Q in the same way as G in position -2P does in the other two cases of the adenosine-wedge motif. Additional interaction of A397 includes the hydrogen bond between its atom N7 and the O2′-H group of adenosine -2Q. For a list of all hydrogen bonds that form in the adenosine-wedge motif, see the Supplemental Table I.

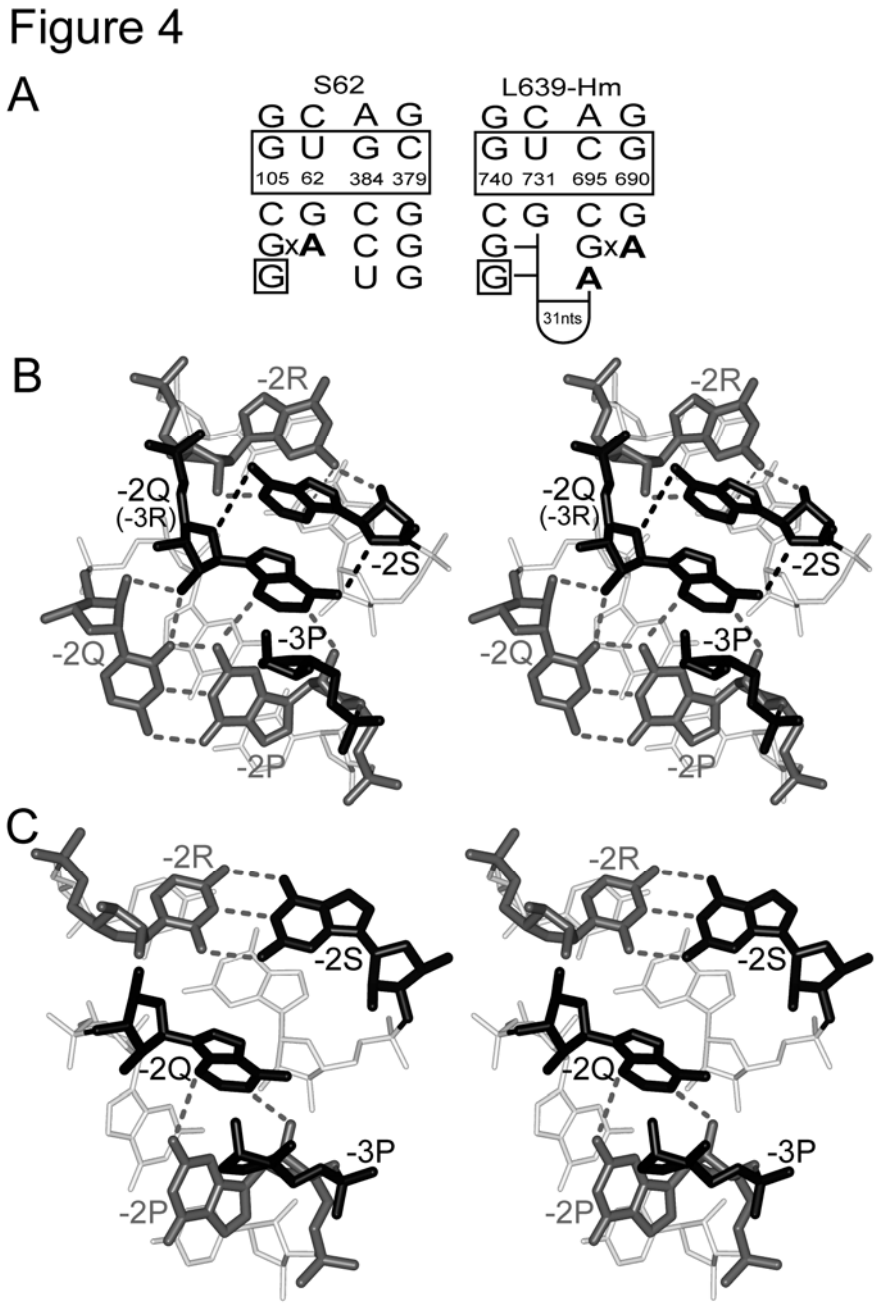


Figure 4. Variations from the structure of the adenosine-wedge motif.

(A) Nucleotide sequences of AGPMs S62 from the *E. coli* 16S rRNA (Schuwirth et al. 2005) and L639 from the *H. marismortui* 23S rRNA (L639-Hm) (Ban et al. 2000). For each structure, the nucleotide numbering corresponds to that used in the original publication of the subunit structure. The secondary structure of each AGPM is drawn accordingly to those shown in Figure 1D.

(B) Stereo view of the adenosine-wedge motif within the AGPM L639 taken from the *H. marismortui* 23S rRNA (pdb entry code 1s72) (Ban et al. 2000). Nucleotides A698 and C729, which occupy respectively positions -3R and -2Q in the polynucleotide chain, approach the base of nucleotide -2P (G742) from two

different sides and form hydrogen bonds with it. Thus, in the tertiary structure, both nucleotides A698 and C729 can be considered as playing the role of nucleotide -2Q. The 31-nucleotide hairpin insertion between nucleotides G730 (-1Q) and A698 (-2Q) is specific to the AGPM L639 in archaea, which allows cytidine C729 to form a WC base pair with guanosine G742 in position -2P. The closeness of cytidine C729 and adenosine -2Q allows the formation of hydrogen bonds between atom O2 of C729 and the O2′-H group of adenosine -2Q and between the O2′-H groups of both C729 and adenosine -2Q. When nucleotide -2R is guanosine, its amino group would form a hydrogen bond with atom O2′of -2S. All hydrogen bonds are listed in the Supplemental Table I.

(C) Stereo view of the structure of the "half of the adenosine wedge" found in the AGPM S62 from *E. coli* (pdb entry code 2avy). Nucleotides -2P, -3P and -2Q in this structure stay practically at the same position as the corresponding nucleotides in the standard adenosine wedge arrangement: the positions of these nucleotides in AGPMs S62 and L639 can be superposed with the r.m.s.d. value of 0.59 Å (not shown). Unlike in the standard adenosine wedge structure, nucleotide -2S (thick black) does not stack to -2Q, but instead, forms a WC base pair with nucleotide -2R.

6.15 Supplemental tables

Supplemental Table I. List of hydrogen bonds formed in the adenosine-wedge motif

Hydrogen bonds	Symmetric / Asymmetric	Origin
N6 (-2S) – O5′ (-2Q)	Asymmetric	Presence of a covalent connection between strands Q and R.
N1 (-2Q) – O2′ (-2P)	Symmetric	Internal symmetry of both adenosines -2Q and -2S.
N1 (-2S) – O2′ (-2R)		
N2 (-2R if G) – O2′ (-2S)	Asymmetric	Displacement of the adenosine wedge [-2Q;-2S] in direction of strand R. The other H-bond between atoms N2 (-2P if G) – O2′ (-2Q) does not form.
N2 (-2P if G) – N3 (-2Q)	Symmetric	Inclination of the adenosine wedge [-2Q;-2S] for about 35°-40° toward strand P.
N2 (-2R if G) – N3 (-2S)		
N6 (-2Q) – O4′ (-2S)	Symmetric	Internal symmetry of both adenosines -2Q and -2S.
N6 (-2S) – O4′ (-2Q)		
N7 (A397 in S549) – O2′ (-2Q)	Equivalent H-bonds	Proximity of A397 and C729 to the ribose of -2Q.
O2 (C729) – O2′ (-2Q) (L639 in archaebacteria)		
O2′ (C729) – O2′ (-2Q) (L639 in archaebacteria)	No analog	Closeness of the position of the riboses of both interacting helices.

Supplemental Table II. Statistical data of the identities of nucleotides -2P, -2R, -2Q and -2S in AGPMs S549, L639 and L657

AGPMs	Nucleotide position	Nucleotide identity	Number of sequences	%
S549	-2P	G	17	0.18
		A	7	0.07
		U	9558	99.49
		C	25	0.26
	-2R	G	3235	33.67
		A	4895	50.95
		U	1440	14.99
		C	37	0.39
	397	G	6	0.06
		A	9590	99.82
		U	3	0.03
		C	8	0.08
			TOTAL: 9607	**100**
L639 and L657	-2P	G	797	93.22
		A	28	3.27
		U	2	0.23
		C	28	3.27
	-2R	G	536	62.69
		A	303	35.44
		U	15	1.75
		C	1	0.12
			TOTAL: 855	**100**
S549	-2Q	A	9506	98.95
	-2S	A	9595	99.88
			TOTAL: 9607	**100**
L639	-2Q	A	427	99.77
	-2S	A	428	100
			TOTAL: 428	**100**
L657	-2Q	A	426	99.77
	-2S	A	424	99.30
			TOTAL: 427	**100**

The data were obtained from the available prokaryotic rRNA sequence alignments of the 16S and 23S rRNAs (Wuyts et al. 2004).

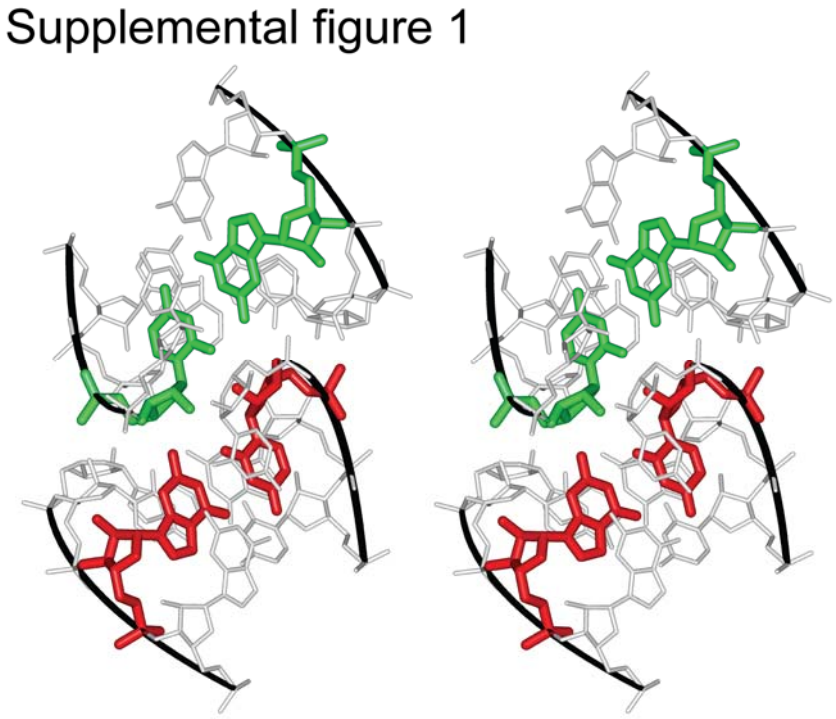
Supplemental Table III. List of all cases of the NAG-triangle motif found in the *E. coli* ribosome structure

30S subunit				
	Nucleotide position and identity			
Name	**-2P**	**-3P**	**-2Q**	**Notes and references**
S50	G360	G361	A50	
S197	G220	C221	A197	
S246	G278	A279	A246	Part of a K-turn motif (Klein et al. 2001)
S498	G404	U405	A498	
S510	C503	C504	A510	Nucleotide -2P (C503) is paired with G542
S535	G527	C528	A535	Part of a G-ribo motif (Steinberg and Boutorine 2007)
S572	A19	U20	A572	
S687	G700	U701	A687	Part of a K-turn motif (Klein et al. 2001)
S913	G22	C23	A913	
S1067	G1108	C1109	A1067	Involved in the formation of the inter-subunit bridge B7
S1394	A1500	C1501	A1394	
S1502	C924	G925	A1502	Nucleotide -2P (C924) is paired with G1392
50S subunit				
L84	G98	U99	A84	Part of a K-turn motif (Klein et al. 2001)
L223	G407	G408	A223	
L526	G2625	C2626	A526	
L1111	C1044	C1045	A1111	Part of a K-turn motif (Klein et al. 2001); Nucleotide -2P (C1044) is unpaired
L1156	A975	G976	A1156	
L1253	C564	C565	A1253	Nucleotide -2P (C564) is paired with G577
L1302	C1297	C1298	A1302	Part of a G-ribo motif (Steinberg and Boutorine 2007); Nucleotide -2P (C1297) is paired with G1643
L1912	C1407	A1408	A1912	Involved in the formation of the inter-subunit bridge B2a; Nucleotides -2P and -3P are from helix h44 of 16S rRNA, while nucleotide -2Q is from helix H69 of 23S rRNA; Nucleotide -2P (C1407) is paired with G1494 in helix h44 of the 16S rRNA
L2288	A2284	C2285	A2288	Part of a G-ribo motif (Steinberg and Boutorine 2007)
L2388	G2280	A2281	A2388	
L2781	G2627	C2628	A2781	

See the footnote on next page

In the name of each arrangement, “S” or “L” stands for the small or large ribosomal subunit in which it is found, followed by the number of adenosine -2Q in the *E. coli* ribosomal subunits (pdb entry codes 2avy-2aw4) (Schuwirth et al. 2005).

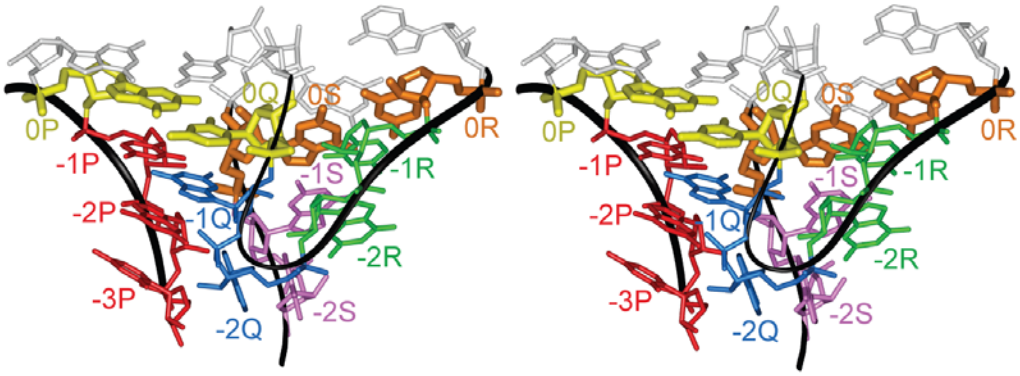
6.16 Supplemental figures



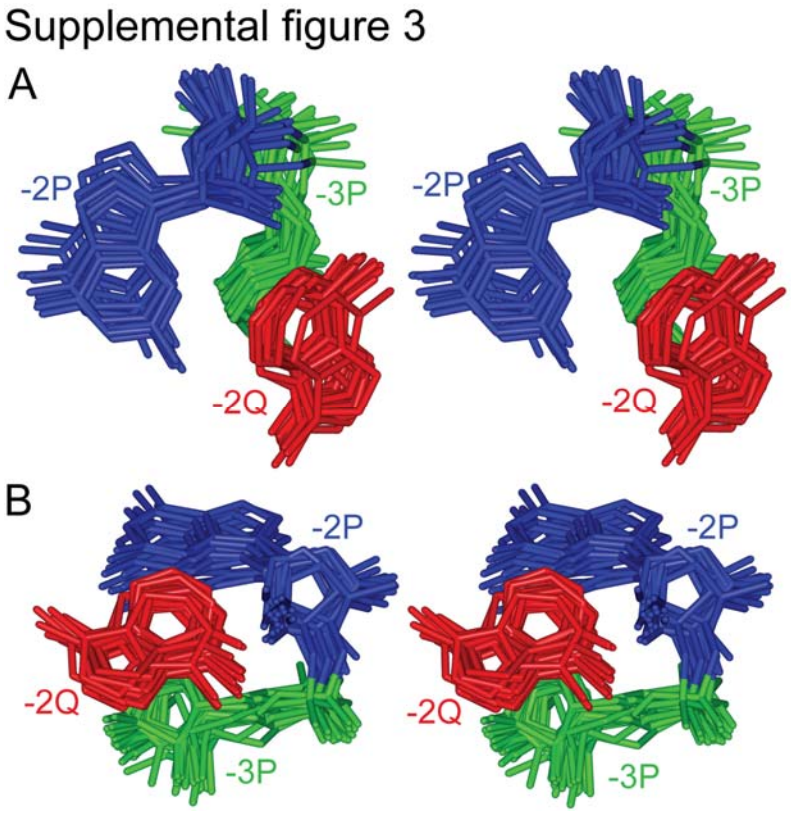
Supplemental figure 1. Stereo view of the along-groove packing motif (AGPM).

The AGPM represents a close packing of two double helices positioned such that the backbone of one helix interacts with the minor groove of the other helix and *vice versa*. In each helix, the contact area with the opposite helix spreads over four layers, from -2 to +1. At the zero-layer, the two helices pack most closely; often one of these base pairs is GU, while the other one is WC. The packing of the GU (red) and WC (green) base pairs within the AGPM allows the formation of a network of hydrogen bonds (see Fig. 1C).

Supplemental figure 2



Supplemental figure 2. Stereo view of the structure of the adenosine-wedge motif within the context of the AGPM L639 from the *E. coli* ribosome (pdb entry code 2aw4) (Schuwirth et al. 2005). Nucleotides are colored as in Figure 2. At the 0-level, the GU and WC base pairs are, respectively, yellow and brown. The conformation of the connection, which represents a bend of the polynucleotide chain by almost 180°, was recently described in terms of the recently identified hook-turn motif (Szép et al. 2003).



Supplemental figure 3. The NAG-triangle motif.

The top (A) and side (B) stereo views of the superposition of 23 examples of the NAG-triangle motif found in the *E. coli* ribosome (pdb entry codes 2avy-2aw4) (Schuwirth et al. 2005). It consists in a three-nucleotide arrangement composed of nucleotides occupying the equivalent positions -2Q (red), -2P (blue) and -3P (green) in the adenosine-wedge motif. All cases of the motif can be superposed based on the positions of the C1′ atoms of nucleotides -2Q, -2P and -3P with the r.m.s.d. value of 0.51 Å. All cases are listed in the Supplemental Table III.

Chapitre 7

Discussion

7. Discussion

7.1 Découverte du motif d'empaquetage le long du sillon (AGPM) (*along-groove packing motif*)

Les interactions entre les doubles hélices dans les molécules d'ARN complexes représentent un aspect fondamental dans la formation de la structure tertiaire. Même si l'empaquetage entres hélices *via* leurs sillons mineurs a été rapporté par les auteurs qui ont publié les premières structures à haute résolution des sous-unités ribosomiques 30S et 50S (Ban et al., 2000; Wimberly et al., 2000), les règles qui régissent les interactions entre deux doubles hélices d'ARN n'ont jamais été explorées. Évidemment, les groupements phosphates chargés négativement doivent être neutralisés afin que deux doubles hélices puissent s'approcher l'une de l'autre. Mis à part la présence de cations (voir introduction), les régions de l'ARN formant des doubles hélices destinées à former des interactions hélice-hélice bénéficieraient grandement de la présence de paires de bases GC/CG. En effet, dans ces paires de bases, le groupe amine exo-cyclique de la guanosine est localisé dans le sillon mineur, favorisant ainsi la neutralisation des groupements phosphates. De plus, la présence de paires de bases GU a été proposée comme étant capable de favoriser les interactions hélice-hélice en raison de leur conformation caractéristique qui expose davantage le groupe amine exo-cyclique de la guanosine dans le sillon mineur comparé aux paires de bases GC/CG.

Les publications des structures cristallographiques à haute résolution des sous-unités ribosomiques, de même que du ribosome entier, ont révélé que la densité des doubles hélices est si élevée qu'à plusieurs endroits, deux hélices se rencontrent et interagissent ensemble. D'une perspective plus générale, ces structures ont également montré que les interactions hélice-hélice sont cruciales pour la formation de la structure tertiaire du ribosome. Cependant, un simple examen visuel de la structure du ribosome n'apportait aucune information à propos des règles utilisées par les doubles hélices d'ARN pour interagir ensemble. Il était donc nécessaire d'entreprendre une analyse systématique, i.e. vérifier tous les cas d'empaquetage hélice-hélice dans ces structures nouvellement disponibles.

Tous les cas d'empaquetage entre les doubles hélices dans la structure du ribosome ont été collectés systématiquement. Suite à la superposition de chaque arrangement avec les autres et une classification de tous ces cas en différents sous-groupes, nous avons noté qu'un de ces sous-groupes avait essentiellement la même structure et qu'il s'agissait en fait d'un nouveau motif d'ARN récurrent, que nous avons nommé le motif d'empaquetage le long du sillon (AGPM) (*along-groove packing motif*). Au total, cinq ont été identifiés dans la sous-unité 30S, tandis que neuf sont localisés dans la sous-unité 50S. Deux cas supplémentaires sont impliqués dans la fixation des molécules d'ARNt aux sites P et E de la sous-unité 50S. À ce jour, seulement un cas a été identifié en dehors de la structure du ribosome et est observé entre deux molécules de ribozyme à la tête en forme de marteau (*hammerhead ribozyme*) situées à l'intérieur de la même unité asymétrique du cristal. En partant du fait que les interactions entre les ARNt et le ribosome doivent se briser et se reformer *de novo* à chaque cycle d'élongation, nous avons suggéré une transition énergétiquement favorable des molécules d'ARNt au cours de leur translocation dans le ribosome. Au même moment que le motif AGPM entre l'ARNt désaminoacylé lié au site P et le ribosome se brise, le même ARNt commence à former un deuxième AGPM dans le site E. Basé sur la position de chaque AGPM intermoléculaire, une telle transition en douceur forcerait les molécules d'ARNt à emprunter une voie *via* les états hybrides [A/P-P/E] (Figure 5, Chapitre 2, page 50).

Dans AGPM, les deux hélices se rencontrent avec un angle d'environ 90° de sorte que le squelette sucre-phosphate d'une hélice s'empaquette le long du sillon mineur de l'autre hélice et *vice versa*. Au centre du motif, on retrouve deux paires de bases qui s'empaquettent de manière très serrée et par conséquent, ont été nommées paires de bases centrales. Dans la plupart des cas identifiés et décrits dans le chapitre 2, une des paires de bases centrales est GU, tandis que l'autre est WC. Parmi tous les cas connus d'AGPM montrés dans la Figure 2, Chapitre 5 (page 160) qui ont les paires de bases centrales GU et WC, l'analyse de plus de 42 000 séquences procaryotiques d'ARNr 16S et 23S disponibles a révélé que GU versus WC est conservé dans 97.3% de ces séquences (Tableau IV, Chapitre 5, page 157). Par conséquent, pour souligner l'importance de GU versus WC dans le cœur d'AGPM, cette interaction fût référée en

tant que modèle GU-WC (*GU-WC pattern*). L'importance de la paire GU dans AGPM fût additionnellement supportée par les deux cas intermoléculaires retrouvés entre l'ARNr 23S et les molécules d'ARNt liées aux sites P et E du ribosome. En effet, pour chacun de ces cas, la paire de bases GU est toujours localisée dans l'ARNr 23S; ce qui garantit que toutes les molécules d'ARNt puissent former cette interaction. En même temps, cependant, l'existence de combinaisons alternatives a montré que des variations n'affectant pas drastiquement la stabilité du motif étaient possibles. Dans la structure du ribosome, il existe deux cas d'AGPM qui possèdent une juxtaposition de paires de bases centrales différente. Les motifs S549 dans toutes les structures connues de 30S (Schluenzen et al., 2000; Wimberly et al., 2000; Schuwirth et al., 2005; Selmer et al., 2006) et L2291 dans la structure de la 50S de *H. marismortui* (Ban et al., 2000) possèdent les paires de bases centrales GC-GC. Un fait intéressant est que leurs structures se superposent très bien avec tous les autres cas d'AGPM possédant la juxtaposition modèle GU-WC; ce qui démontre que le motif peut toujours se former même en l'absence d'une paire de bases centrale GU. D'autres anomalies affectant le motif S549 sont discutées plus loin. À partir de ces observations, nous étions intéressés à connaître les limites dans lesquelles les paires de bases centrales pourraient être modifiées.

7.2 Expression de librairies combinatoires de gènes du motif AGPM

Une nouvelle approche permettant l'étude des interactions protéines-ARN dans les molécules d'ARN complexes a été présentée et consiste à déterminer l'étendue de la variabilité des séquences de nucléotides (*scope of nucleotide variability*) pour différents représentants du même motif récurrent retrouvé dans différents contextes structuraux dans la structure du ribosome. Afin de valider la stratégie, l'utilisation du motif AGPM était possible dû au fait que les nucléotides dans le cœur de cet arrangement récurrent peuvent varier et l'effet se reflète sur la stabilité du motif (Chapitre 3). Le premier aspect permet l'apparition d'autres séquences que le type sauvage (WT) (*Wild-Type*), tandis que le second permet une sélection positive des ribosomes fonctionnels. Au chapitre 3, nous avons présenté une librairie combinatoire des paires de bases centrales du motif

S296, un cas totalement libre d'autres interactions et de ce fait, représentait un bon candidat pour l'étude des propriétés du motif. La conclusion principale de cet article est que l'empaquetage serré des hélices 3 et 12 de l'ARNr 16S est requis afin d'assurer une fonction normale au ribosome. À ce moment, le prérequis minimum pour la formation du motif n'avait pas encore été identifié. C'est seulement suite à l'incorporation d'autres librairies combinatoires des paires de bases centrales pour deux autres représentants du motif, L639 et L657, localisés dans la sous-unité 50S, que le prérequis minimum du motif AGPM a été identifié (Chapitre 4).

Nous avons commencé par comparer l'étendue de la variabilité des séquences de nucléotides (*scope of nucleotide variability*) des paires de bases centrales pour les séquences sélectionnées des trois cas du motif AGPM, S296, L639 et L657, localisés dans différents contextes structuraux. Pour que le motif puisse se former, au moins une paire de bases centrale doit adopter une structure définie (*structure-forming base pair*) comme GU, WC ou AC/CA, tandis que l'identité de la paire de bases centrale retrouvée dans l'hélice opposée devient peu importante. Ceci permet à trois nucléotides sur quatre d'être positionnés comme dans le modèle standard GU-WC. Ainsi, le prérequis minimum des paires de bases centrales dans AGPM consiste en la formation d'une triade de nucléotides. Même si l'hélice qui ne contient pas de paire de bases formant une structure définie peut être moins stable, le fait que l'autre hélice contienne une paire de bases GU, WC ou AC/CA formant une structure définie peut stabiliser la première hélice. En d'autres termes, l'hélice qui contient la paire de bases formant une structure définie agit comme chaperon moléculaire en assistant le repliement de l'autre hélice.

En l'absence d'autres interactions impliquant les paires de bases centrales, la paire de bases qui forme une structure définie peut se retrouver dans n'importe laquelle des deux hélices. De manière intéressante, la présence de protéines ribosomiques aux environs du motif peut introduire une asymétrie dans le positionnement de la paire de bases ayant une structure définie. Lorsque l'interaction entre la protéine et les paires de bases centrales dans AGPM est indirecte, c'est-à-dire la protéine touche aux nucléotides localisés aux alentours de ceux impliqués dans la formation des paires centrales, nous avons observé une forte polarisation dans le positionnement de la paire de bases ayant une structure définie. Cependant, un contact serré impliquant directement un nucléotide

d'une paire de bases centrale impose des règles strictes en ce qui a trait à la position exacte de ce dernier; l'identité des nucléotides des paires de bases centrales s'y reflète. Notre analyse nous a permis de suggérer que les motifs L639 et L657 font partie des sites de reconnaissance primaire des protéines ribosomiques L35 et L4, respectivement.

Jusqu'à tout récemment, GU versus WC était considéré comme une signature caratéristique du motif AGPM [Chapitre 2 et (Mokdad et al., 2006)]. Cependant, nous savons maintenant que dû au prérequis minimum pour la formation du motif, les paires de bases centrales peuvent être réduites à la présence d'une paire de bases formant une structure définie comme GU, WC, AC ou CA, tandis que la paire de bases opposée ne requiert aucunement une complémentarité WC. Le fait qu'il est possible d'intégrer des faiblesses au centre de l'arrangement suggérait fortement l'existence d'autres interactions entre les deux hélices responsables au maintien de l'intégrité du motif AGPM.

7.3 Règles structurales liées à la formation des interactions squelette-squelette dans le motif AGPM

Tel que mentionné au chapitre 2, quatre paires de bases dans chaque hélice du motif AGPM sont impliquées dans les contacts inter-hélices. Pour tous les cas connus du motif AGPM, les interactions entre squelettes s'effectuent *via* des contacts ribose-ribose; ce qui est plus favorable que des interactions entre groupements phosphates (Figure 3, Chapitre 5, page 161). De plus, chaque ribose impliqué dans les contacts inter-hélices est positionné de manière à éviter une collision avec l'hélice opposée tout en favorisant son interaction avec le ribose opposé. Au chapitre 5, nous avons montré que la conformation du squelette dans AGPM est modulée par un choix judicieux de l'identité des paires de bases à chaque niveau d'interaction entre les hélices; ce qui permet des contacts squelette-squelette sans collision.

Cette analyse a apporté une explication rationnelle de la présence de la paire de bases centrale GU dans une des deux hélices. En effet, si les deux hélices dans AGPM avaient une conformation standard de forme-A, il y aurait eu collision entre les riboses centraux internes (Figure 5, Chapitre 5, page 163). La présence de la paire de bases GU

a résolu ce problème en déplaçant l'uridine interne en direction du sillon majeur, permettant ainsi d'éviter une collision avec le ribose interne opposé. Une situation similaire est retrouvée au niveau +1 dans AGPM, où les deux riboses internes seraient entrés en collision (Figure 5, Chapitre 5, page 163). Ainsi, n'importe quelle paire de bases non-WC permettant au ribose interne d'une hélice d'être dévié en direction du sillon majeur est acceptable. Cependant, l'analyse a montré que la présence de deux paires de bases non-WC au niveau +1 aurait déplacé les deux riboses internes en direction des sillons majeurs, créant ainsi un espace vide entre les hélices, une situation qui aurait déstabilisé le motif.

Dans quelques exemples d'AGPM, deux paires de bases WC sont retrouvées au niveau +1. Pour que les riboses internes forment des interactions sans collision, la conformation de chaque hélice est modulée de sorte qu'elle dévie légèrement de la forme-A standard (Figure 7, Chapitre 5, page 165); il en résulte une situation non idéale au niveau +1. Nous avons suggéré que la tension accumulée dans chaque hélice pourrait jouer un rôle fonctionnel. Dans les motifs S911 et L1923-P, deux paires de bases WC conservées sont retrouvées au niveau +1 (Tableau V, Chapitre 5, page 158). Dans le premier cas, le motif se forme lors de la liaison du facteur d'initiation-1 (IF1) à la sous-unité ribosomique 30S (Carter et al., 2001), tandis que dans le deuxième cas, le motif se forme entre l'hélice H69 de l'ARNr 23S et la tige D d'un ARNt lié au site P du ribosome. D'un point de vue fonctionnel, ces deux motifs doivent se former et se briser à chaque cycle d'élongation. La présence d'une situation non idéale au niveau +1 pourrait favoriser la rupture de ces motifs lorsque nécessaire. Le motif S549 possède deux anomalies. Chez les eubactéries, une juxtaposition WC-WC est retrouvée aux niveaux 0 et +1 (Tableaux IV et V, Chapitre 5, pages 157 et 158). Chez les archaebactéries, ce même motif contient une paire de bases UG au niveau +1 (Tableau V, Chapitre 5, page 158); ce qui crée encore plus de tension en déplaçant le ribose interne du G en direction du ribose +1 opposé. La présence de ces anomalies dans le motif S549 indique qu'elles pourraient jouer un rôle spécifique dans la fonction du ribosome qui, cependant, nécessitera des analyses supplémentaires.

Deux autres contacts symétriques ribose-ribose sont retrouvés entre les niveaux -2 et -1 de chaque hélice. Chaque nucléotide formant les paires de bases -1 est impliqué

dans des interactions inter-hélices; ce qui inclut également les bases azotées. Dû aux contraintes imposées sur ces paires de bases, elles sont restreintes à une identité WC et ont une conformation standard de forme-A. Ainsi, seulement les paires de bases -2 peuvent s'accommoder à la position des riboses du niveau -1. À l'image du niveau +1, le niveau -2 peut accommoder des paires de bases WC et non-WC mais ici, la conformation acceptable des paires de bases non-WC doit permettre le déplacement du ribose externe en direction du sillon majeur.

Un choix judicieux de l'identité des paires de bases à chaque niveau d'interaction entre les hélices permet la formation de contacts ribose-ribose optimaux en dehors des paires de bases centrales dans AGPM; ce qui permet une compensation efficace des faiblesses d'interactions retrouvées entre les paires de bases centrales.

7.4 Organisation hiérarchique des structures tertiaires d'ARN

Une comparaison de la structure de tous les cas connus du motif AGPM a révélé un arrangement spécifique qui se forme au niveau -2 dans trois AGPM. Dans ces cas, les nucléotides du niveau -2 ne forment pas de paires de bases, mais plutôt un nucléotide (adénosine) de chaque paire de bases -2 est positionné presque verticalement similaire à une cale (*wedge*) entre les deux hélices. Ces trois arrangements se superposent très bien et ainsi constituent un autre motif récurrent, que nous avons nommé « *adenosine-wedge* » (Chapitre 6). Le motif « *adenosine-wedge* » inclut également le motif « *A-minor* » et quelquefois, le motif « *hook-turn* » (pour les définitions, voir introduction). Le cœur de l'arrangement du motif « *adenosine-wedge* » consiste en trois nucléotides qui forment ensemble un triangle. Une recherche de cet arrangement dans la structure du ribosome a révélé 23 cas non reliés au motif AGPM. Ainsi, cet arrangement de trois nucléotides à l'intérieur du motif « *adenosine-wedge* » représente par lui-même un autre motif récurrent, nommé triangle-NAG (*NAG-triangle*). Dans ce dernier arrangement, une face de l'adénosine est stabilisée par son interaction avec deux nucléotides qui font partie d'une conformation hélicoïdale, tandis que l'autre face peut initier un empilement de bases pouvant former une autre hélice.

Il est étonnant de voir à quel point la structure tertiaire de l'ARN peut être hiérarchique. En effet, à l'intérieur même du motif AGPM, il est possible de retrouver le motif « *adenosine-wedge* » qui contient lui-même les motifs « *A-minor* », triangle-NAG « *NAG-triangle* », et quelquefois « *hook-turn* », qui sont tous impliqués dans des relations de causes à effets complexes. Une telle hiérarchie a de fortes chances d'être retrouvée dans d'autres motifs d'ARN récurrents. Ainsi, une compréhension complète de la formation de la structure de l'ARN passe inévitablement par une collecte des motifs récurrents suivie d'une étude systématique de leurs structures visant à déterminer le rôle de chaque nucléotide dans la formation de l'arrangement en question. De meilleures prédictions du repliement des molécules d'ARN complexes deviendront possibles seulement lorsque nous aurons un tableau complet de tous les « blocs » utilisés dans la formation de ces structures de même que comprendrons la façon dont ils sont inter-reliés. Afin d'atteindre un tel niveau de compréhension, l'identification et l'analyse systématique des motifs récurrents sont des étapes nécessaires à la résolution du problème général du repliement de l'ARN.

Chapitre 8

Conclusion

8. Conclusion

Le motif récurrent d'empaquetage de long du sillon (AGPM) (*along-groove packing motif*) a été identifié dans la structure du ribosome et caractérisé de façon extensive *in vivo* et *in silico*. Une analyse systématique de sa structure et une comparaison de l'étendue de la variabilité des séquences de nucléotides (*scope of nucleotide variability*) obtenues expérimentalement à partir de librairies combinatoires des gènes d'ARNr 16S et 23S, ont permis d'établir:

1) La présence et l'importance des paires de bases GU versus WC dans le centre de l'arrangement;
2) Le prérequis minimum des paires de bases centrales responsable à la formation du motif qui consiste en une triade de nucléotides dans le cœur de l'arrangement;
3) Les trois contacts ribose-ribose caractéristiques entre les hélices et la modulation de leur position *via* un choix judicieux du type de paire de bases dans laquelle chaque ribose est impliqué;
4) La présence d'un autre arrangement récurrent qui est retrouvé dans la structure de trois cas d'AGPM, nommé motif « *adenosine-wedge* », lui-même composé d'autres motifs récurrents, tel « *A-minor* », triangle-NAG (*NAG-triangle*) et quelquefois, « *hook-turn* ».

La croissance du nombre de motifs structuraux identifiés jusqu'à ce jour, de même que l'étude des prérequis de séquences nécessaires à leur formation, augmentent notre compréhension générale sur la façon dont une chaîne de polynucléotides se replie et forme des interactions stables avec d'autres régions de la structure. De plus, la connaissance des prérequis de séquences pour différents motifs permet d'établir des consensus de séquences pour ces derniers; ce qui ouvre la possibilité de les identifier dans des molécules d'ARN dont la structure est toujours inconnue.

Maintenant que le motif AGPM a été identifié et caractérisé, il serait intéressant d'étudier la dynamique de formation et de dissociation des deux hélices qui composent ce motif. En effet, une compréhension de la façon dont AGPM se forme et se brise pourrait apporter des indications importantes relatives à la dynamique des ARNt lors de leur translocation dans le ribosome.

9. Bibliographie (Introduction et Discussion)

Agirrezabala X, Lei J, Brunelle JL, Ortiz-Meoz RF, Green R, Frank J. 2008. Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol. Cell* **32**:190-197.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**:905-920.

Batey RT. 2006. Structures of regulatory elements in mRNAs. *Curr. Opin. Struct. Biol.* **16**:299-306.

Batey RT, Rambo RP, Doudna JA. 1999. Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed. Engl.* **38**:2326-2343.

Battle DJ, Doudna JA. 2002. Specificity of RNA-RNA helix recognition. *Proc. Natl. Acad. Sci. USA* **99**:11676-11681.

Bretscher MS. 1968. Translocation in protein synthesis: a hybrid structure model. *Nature* **218**:675-677.

Brodersen DE, Carter AP, Clemons WM, Jr., Morgan-Warren RJ, Murphy FV, Ogle JM, Tarry MJ, Wimberly BT, Ramakrishnan V. 2001. Atomic structures of the 30S subunit and its complexes with ligands and antibiotics. *Cold Spring Harb. Symp. Quant. Biol.* **66**:17-32.

Brodersen DE, Clemons WM, Jr., Carter AP, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V. 2000. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell* **103**:1143-1154.

Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, Pande N, Shang Z, Yu N, Gutell RR. 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**:2.

Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V. 2000. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **407**:340-348.

Carter AP, Clemons WM, Jr., Brodersen DE, Morgan-Warren RJ, Hartsch T, Wimberly BT, Ramakrishnan V. 2001. Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* **291**:498-501.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**:1678-1685.

Cooperman BS. 1977. Identification of binding sites on the *E. coli* ribosome by affinity labeling. *Adv. Exp. Med. Biol.* **86A**:595-609.

Crick FHC. 1966. The genetic code - yesterday, today, and tomorrow. *Cold Spring Harb. Symp. Quant. Biol.* **31**:1-9.

Doherty EA, Batey RT, Masquida B, Doudna JA. 2001. A universal mode of helix packing in RNA. *Nat. Struct. Biol.* **8**:339-343.

Doyon FR, Zagryadskaya EI, Chen J, Steinberg SV. 2004. Specific and non-specific purine trap in the T-loop of normal and suppressor tRNAs. *J. Mol. Biol.* **343**:55-69.

Draper DE. 2004. A guide to ions and RNA structure. *RNA* **10**:335-343.

Draper DE, Grilley D, Soto AM. 2005. Ions and RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **34**:221-243.

Ferre-D'Amare AR, Doudna JA. 1999. RNA folds: insights from recent crystal structures. *Annu. Rev. Biophys. Biomol. Struct.* **28**:57-73.

Ferre-D'Amare AR, Zhou K, Doudna JA. 1998. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**:567-574.

Francklyn C, Schimmel P. 1989. Aminoacylation of RNA minihelices with alanine. *Nature* **337**:478-481.

Frank J, Agrawal RK. 2000. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406**:318-322.

Gilbert W. 1986. Origin of life: The RNA world. *Nature* **319**:618.

Green R, Noller HF. 1997. Ribosomes and translation. *Annu. Rev. Biochem.* **66**:679-716.

Gualerzi C, Risuleo G, Pon CL. 1977. Initial rate kinetic analysis of the mechanism of initiation complex formation and the role of initiation factor IF-3. *Biochemistry* **16**:1684-1689.

Gueron M, Leroy JL. 1982. Significance and mechanism of divalent-ion binding to transfer RNA. *Biophys. J.* **38**:231-236.

Guerrier-Takada C, Altman S. 1984. Catalytic activity of an RNA molecule prepared by transcription in vitro. *Science* **223**:285-286.

Gutell RR, Larsen N, Woese CR. 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* **58**:10-26.

Harms J, Schluenzen F, Zarivach R, Bashan A, Gat S, Agmon I, Bartels H, Franceschi F, Yonath A. 2001. High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107**:679-688.

Holbrook SR, Kim SH. 1997. RNA crystallography. *Biopolymers* **44**:3-21.

Hou YM, Schimmel P. 1988. A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* **333**:140-145.

Huxley HE, Zubay G. 1960. Electron microscope observations on the structure of microsomal particles from *Escherichia coli*. *J. Mol. Biol.* **2**:10-18.

Janosi L, Hara H, Zhang S, Kaji A. 1996. Ribosome recycling by ribosome recycling factor (RRF)--an important but overlooked step of protein biosynthesis. *Adv. Biophys.* **32**:121-201.

Judson HF. 2001. Talking about the genome. *Nature* **409**:769.

Julian P, Konevega AL, Scheres SH, Lazaro M, Gil D, Wintermeyer W, Rodnina MV, Valle M. 2008. Structure of ratcheted ribosomes with tRNAs in hybrid states. *Proc. Natl. Acad. Sci. USA* **105**:16924-16927.

Kaempfer R. 1972. Initiation factor IF-3: a specific inhibitor of ribosomal subunit association. *J. Mol. Biol.* **71**:583-598.

Kim SH, Suddath FL, Quigley GJ, McPherson A, Sussman JL, Wang AH, Seeman NC, Rich A. 1974. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* **185**:435-440.

Klein DJ, Schmeing TM, Moore PB, Steitz TA. 2001. The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**:4214-4221.

Korostelev A, Trakhanov S, Laurberg M, Noller HF. 2006. Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* **126**:1065-1077.

Krasilnikov AS, Mondragon A. 2003. On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA* **9**:640-643.

Krasilnikov AS, Yang, X, Pan, T, Mondragon A. 2003. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**:760-764.

Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**:147-157.

Lake JA. 1976. Ribosome structure determined by electron microscopy of *Escherichia coli* small subunits, large subunits and monomeric ribosomes. *J. Mol. Biol.* **105**:131-139.

Lee JC, Gutell RR. 2004. Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs. *J. Mol. Biol.* **344**:1225-1249.

Marshall RA, Aitken CE, Dorywalska M, Puglisi JD. 2008. Translation at the Single-Molecule Level. *Annu. Rev. Biochem.* **77**:177-203.

McClain WH, Chen YM, Foss K, Schneider J. 1988. Association of transfer RNA acceptor identity with a helical irregularity. *Science* **242**:1681-1684.

McClain WH, Foss K. 1988. Changing the identity of a tRNA by introducing a G-U wobble pair near the 3' acceptor end. *Science* **240**:793-796.

Misra VK, Draper DE. 1998. On the role of magnesium ions in RNA stability. *Biopolymers* **48**:113-135.

Misra VK, Draper DE. 2002. The linkage between magnesium binding and RNA folding. *J. Mol. Biol.* **317**:507-521.

Moazed D, Noller HF. 1989. Intermediate states in the movement of transfer RNA in the ribosome. *Nature* **342**:142-148.

Moazed D, Samaha RR, Gualerzi C, Noller HF. 1995. Specific protection of 16 S rRNA by translational initiation factors. *J. Mol. Biol.* **248**:207-210.

Mokdad A, Krasovska MV, Sponer J, Leontis NB. 2006. Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.* **34**:1326-1341.

Moore PB. 1999. Structural motifs in RNA. *Annu. Rev. Biochem.* **68**:287-300.

Nagaswamy U, Fox GE. 2002. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA* **8**:1112-1119.

Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. 2001. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. USA* **98**:4899-4903.

Noller HF. 2005. RNA structure: reading the ribosome. *Science* **309**:1508-1514.

Noller HF, Hoffarth V, Zimniak L. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**:1416-1419.

Noller HF, Kop J, Wheaton V, Brosius J, Gutell RR, Kopylov AM, Dohme F, Herr W, Stahl DA, Gupta R, Waese CR. 1981. Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res.* **9**:6167-6189.

Noller HF, Woese CR. 1981. Secondary structure of 16S ribosomal RNA. *Science* **212**:403-411.

Nowakowski J, Tinoco I. 1999. RNA structure in solution. In: Neidle S, ed. *Oxford Handbook of Nucleic Acid Structure*. New York, NY: Oxford University Press. pp 567-602.

Ogle JM, Brodersen DE, Clemons WM, Jr., Tarry MJ, Carter AP, Ramakrishnan V. 2001. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**:897-902.

Ogle JM, Carter AP, Ramakrishnan V. 2003. Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem. Sci.* **28**:259-266.

Ogle JM, Murphy FV, Tarry MJ, Ramakrishnan V. 2002. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **111**:721-732.

Palade GE. 1955. A small particulate component of the cytoplasm. *J. Biophys. Biochem. Cytol.* **1**:59-68.

Pape T, Wintermeyer W, Rodnina MV. 1998. Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the *E. coli* ribosome. *EMBO J.* **17**:7490-7497.

Pioletti M, Schlunzen F, Harms J, Zarivach R, Gluhmann M, Avila H, Bashan A, Bartels H, Auerbach T, Jacobi C, Hartsch T, Yonath A, Franceschi F. 2001. Crystal

structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *EMBO J.* **20**:1829-1839.

Pley HW, Flaherty KM, McKay DB. 1994a. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**:111-113.

Pley HW, Flaherty KM, McKay DB. 1994b. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**:68-74.

Ramakrishnan V. 2002. Ribosome structure and the mechanism of translation. *Cell* **108**:557-572.

Razga F, Koca J, Sponer J, Leontis NB. 2005. Hinge-like motions in RNA kink-turns: the role of the second a-minor motif and nominally unpaired bases. *Biophys. J.* **88**:3466-3485.

Razga F, Spackova N, Reblova K, Koca J, Leontis NB, Sponer J. 2004. Ribosomal RNA kink-turn motif--a flexible molecular hinge. *J. Biomol. Struct. Dyn.* **22**:183-194.

Razga F, Zacharias M, Reblova K, Koca J, Sponer J. 2006. RNA kink-turns as molecular elbows: hydration, cation binding, and large-scale dynamics. *Structure* **14**:825-835.

Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BF, Klug A. 1974. Structure of yeast phenylalanine tRNA at 3 A resolution. *Nature* **250**:546-551.

Rosenberg JM, Seeman NC, Day RO, Rich A. 1976. RNA double-helical fragments at atomic resolution. II. The crystal structure of sodium guanylyl-3',5'-cytidine nonahydrate. *J. Mol. Biol.* **104**:145-167.

Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A. 2000. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**:615-623.

Schlunzen F, Zarivach R, Harms J, Bashan A, Tocilj A, Albrecht R, Yonath A, Franceschi F. 2001. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413**:814-821.

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH. 2005. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**:827-834.

Seeman NC, Rosenberg JM, Suddath FL, Kim JJ, Rich A. 1976. RNA double-helical fragments at atomic resolution. I. The crystal and molecular structure of sodium adenylyl-3',5'-uridine hexahydrate. *J. Mol. Biol.* **104**:109-144.

Selmer M, Dunham CM, Murphy FV, Weixlbaumer A, Petry S, Kelley AC, Weir JR, Ramakrishnan V. 2006. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**:1935-1942.

Shi H, Moore PB. 2000. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **6**:1091-1105.

Shine J, 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. USA* **71**:1342-1346.

Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S. 1998. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26**:148-153.

Steitz TA. 2008. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* **9**:242-253.

Szép S, Wang J, Moore PB. 2003. The crystal structure of a 26-nucleotide RNA containing a hook-turn. *RNA* **9**:44-51.

Tamura M, Holbrook SR. 2002. Sequence and structural conservation in RNA ribose zippers. *J. Mol. Biol.* **320**:455-474.

Tan ZJ, Chen SJ. 2008. Electrostatic free energy landscapes for DNA helix bending. *Biophys. J.* **94**:3137-3149.

Thirumalai D, Lee N, Woodson SA, Klimov D. 2001. Early events in RNA folding. *Annu. Rev. Phys. Chem.* **52**:751-762.

Turner DH, Freier SM, Sugimoto N, Hickey DR, Jaeger JA, Sinclair A, Alkema D, Neilson T, Caruthers MH, Kierzek R. 1986. Improved parameters for prediction of RNA secondary structure and insights into why RNA forms double helices. In: van Knippenberg PH, Hilbers CW, eds. *Structure and Dynamics of RNA*. New York, NY: Plenum Press. pp 1-13.

Valle M, Zavialov A, Sengupta J, Rawat U, Ehrenberg M, Frank J. 2003. Locking and unlocking of ribosomal motions. *Cell* **114**:123-134.

Villa E, Sengupta J, Trabuco LG, LeBarron J, Baxter WT, Shaikh TR, Grassucci RA, Nissen P, Ehrenberg M, Schulten K, Frank J. 2009. Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc. Natl. Acad. Sci. USA* **106**:1063-1068.

Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS. 2004. Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet.* **20**:44-50.

Williams LD, Maher LJ, 3rd. 2000. Electrostatic mechanisms of DNA deformation. *Annu. Rev. Biophys. Biomol. Struct.* **29**:497-521.

Wimberly BT, Brodersen DE, Clemons WM, Jr., Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. 2000. Structure of the 30S ribosomal subunit. *Nature* **407**:327-339.

Woese CR. 1967. *The Genetic Code*. New York: Harper and Row.

Woese CR, Magrum LJ, Gupta R, Siegel RB, Stahl DA, Kop J, Crawford N, Brosius J, Gutell R, Hogan JJ, Noller HF. 1980. Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res.* **8**:2275-2293.

Woodson SA. 2005. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.* **9**:104-109.

Wuyts J, Perriere G, Van De Peer Y. 2004. The European ribosomal RNA database. *Nucleic Acids Res.* **32**:D101-103.

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF. 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**:883-896.