Université de Montréal

RNA recurrent motifs: identification and characterization

par
Yury Butorin

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

Avril, 2010

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

RNA recurrent motifs: identification and characterization

présentée par :
Yury Butorin

A été évaluée par un jury composé des personnes suivantes :

Dr Pascal Chartrand, président-rapporteur
Dr Serguei Chteinberg, directeur de recherche
Dr Joëlle Pelletier, membre du jury
Dr Roscoe Klinck, examinateur externe
Dr. François Major, représentant du doyen

Résumé

La détermination de la structure tertiaire du ribosome fut une étape importante dans la compréhension du mécanisme de la synthèse des protéines. Par contre, l'élucidation de la structure du ribosome comme tel ne permet pas une compréhension de sa fonction. Pour mieux comprendre la nature des relations entre la structure et la fonction du ribosome, sa structure doit être étudiée de manière systématique. Au cours des dernières années, nous avons entrepris une démarche systématique afin d'identifier et de caractériser de nouveaux motifs structuraux qui existent dans la structure du ribosome et d'autres molécules contenant de l'ARN.

L'analyse de plusieurs exemples d'empaquetage de deux hélices d'ARN dans la structure du ribosome nous a permis d'identifier un nouveau motif structural, nommé « G-ribo ». Dans ce motif, l'interaction d'une guanosine dans une hélice avec le ribose d'un nucléotide d'une autre hélice donne naissance à un réseau d'interactions complexes entre les nucléotides voisins. Le motif G-ribo est retrouvé à 8 endroits dans la structure du ribosome. La structure du G-ribo possède certaines particularités qui lui permettent de favoriser la formation d'un certain type de pseudo-nœuds dans le ribosome.

L'analyse systématique de la structure du ribosome et de la ARNase P a permis d'identifier un autre motif structural, nommé « DTJ » ou « Double-Twist Joint motif ». Ce motif est formé de trois courtes hélices qui s'empilent l'une sur l'autre. Dans la zone de contact entre chaque paire d'hélices, deux paires de bases consécutives sont surenroulées par rapport à deux paires de bases consécutives retrouvées dans l'ARN de forme A. Un nucléotide d'une paire de bases est toujours connecté directement à un nucléotide de la paire de bases surenroulée, tandis que les nucléotides opposés sont connectés par un ou plusieurs nucléotides non appariés. L'introduction d'un surenroulement entre deux paires de bases consécutives brise l'empilement entre les nucléotides et déstabilise l'hélice d'ARN. Dans le motif DTJ, les nucléotides non appariés qui lient les deux paires de bases surenroulées interagissent avec une des trois hélices qui forment le motif, offrant ainsi une stratégie élégante de stabilisation de l'arrangement.

Pour déterminer les contraintes de séquences imposées sur la structure tertiaire d'un motif récurrent dans le ribosome, nous avons développé une nouvelle approche expérimentale. Nous avons introduit des librairies combinatoires de certains nucléotides retrouvés dans des motifs particuliers du ribosome. Suite à l'analyse des séquences alternatives sélectionnées *in vivo* pour différents représentants d'un motif, nous avons été en mesure d'identifier les contraintes responsables de l'intégrité d'un motif et celles responsables d'interactions avec les éléments qui forment le contexte structural du motif.

Les résultats présentés dans cette thèse élargissent considérablement notre compréhension des principes de formation de la structure d'ARN et apportent une nouvelle façon d'identifier et de caractériser de nouveaux motifs structuraux d'ARN.

Mots-clés: Motif recurrent, structure d'ARN, G-ribo, ribosome, selection *in vivo*

Abstract

Although determination of the ribosome tertiary structure has been an outstanding step towards elucidation of the mechanism of protein synthesis, the complexity of this structure does not provide an easy answer of how this large molecular complex works. In order to understand the nature of structure-function relationships in the ribosome, the ribosome structure itself should be subjected to thorough analysis. In the last years, we undertook systematic efforts toward identification and characterization of all recurrent structural motifs existing in the ribosomal RNA and in other RNA-containing molecules.

The analysis of many instances of helix-helix packing in the ribosome structure allowed us to identify a new structural motif which we called “G-ribo”. In this motif, an interaction of the sugar edge of a guanosine in one helix with the ribose of a nucleotide from another helix was found to be at the origin of a complex network of concomitant inter-nucleotide interactions. In total, the G-ribo motif was found at eight locations within the ribosomal RNA. A surprising feature of this motif consists in its ability to favor the formation of pseudoknots of a particular type. In the ribosome structure, there are four pseudoknots whose formation is mediated by the G-ribo motif.

Systematic analysis of the ribosome as well as the RNAseP crystal structures allowed for the identification of a new RNA motif, which we called “DTJ”, or Double-Twist Joint motif. This motif is made of three short RNA double helices, which stack one on top of another. In the contact zone of each pair of helices two consecutive base pairs are over-twisted compared to the regular helical twist of 32° of A-RNA. One nucleotide of the base pair is always directly connected to the one nucleotide of the over-twisted base pair, while the opposite nucleotides of these base pairs are connected with one or several unpaired nucleotides. Introduction of the helical over-twist between two consecutive base pairs breaks the inter-nucleotide stacking and destabilizes the RNA double helix. In the DTJ, the unpaired nucleotides that connect the two over-twisted base pairs interact with one of the three motif-forming helices, providing an elegant strategy for the stabilization of the whole arrangement.

To determine the nucleotide sequence constraints imposed on the structure of recurrent RNA motifs in the functional ribosome we developed a new approach consisting in the selection of functional ribosomes from a combinatorial gene library in which certain nucleotides of the rRNA gene corresponding to a particular motif were randomized. Comparison of the constraints determined for different examples of the same motif allowed us to distinguish between constraints responsible for the integrity of the motif and for its interaction with surrounding elements, including ribosomal proteins.

The work significantly improves our understanding of the principles of RNA structure formation and opens a new way to identify and characterize RNA motifs.

Keywords: Recurrent motif, RNA structure, G-ribo, ribosome, *in vivo* selection

Table of contents

List of Tables

Chapter 6 : *In vivo* study of the G-ribo motif

List of Figures

Chapter 1: Introduction

Chapter 2: G-ribo: a new structural motif in ribosomal RNA

Chapter 3: G-ribo motif favors the formation of pseudoknots in ribosomal RNA

Chapter 4: Double twist-joint – new recurrent RNA motifs

Chapter 6: *In vivo* study of the G-ribo motif

Chapter 7: Discussion

Abbreviations list

3D	Three dimensional
A	Adenosine
Å	Angstrom (10^{-10} meters)
AGPM	Along-Groove packing motif
Amp	Ampicillin
ASD	Anti-Sinde-Dalgarno sequence
C	Cytidine
CAT	Chloramphenicol acetyl-transferase
Cryo-EM	Cryo-electron microscopy
DNA	Deoxyribonucleic acid
E.coli	*Escherichia coli*
EF-G	Elongation Factor G
G	Guanosine
GFP	Green Fluorescent Protein
H-bond	Hydrogen bond
HCV	Hepatitis C virus
HDV	Hepatitis delta virus
HG	Hoogsteen
H_i	Helix number i
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
IRES	Internal ribosome entry site
Kan	Kanamycin
LB	Luria-Bertani
MD	Molecular Dynamics
miRNA	Micro RNA
MMTV	Mouse mammary tumor virus
mRNA	Messenger RNA

PABP	poly(A) tail-binding protein
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
ps	Picosecond (10^{-12} of a second)
PTC	Peptidyl Transferase Centre
Pu	Purine
Py	Pyrimidine
RNA	Ribonucleic acid
RNAse P	Ribonuclease P
RRF	Ribosomal release factor
rRNA	Ribosomal RNA
S	Svedberg, sedimentation coefficient
SD	Shine-Dalgarno sequence
SE	Sugar Edge
siRNA	Small interfering RNA
Spc	Spectinomycin
T	Thymidine
tRNA	Transfer RNA
U	Uridine
UTR	Untranslated Region
WC	Watson Crick
WT	Wild Type

To my wife, my mother and my deceased grand mother
for their constant encouragements
and unconditional help

Acknowledgements

First, I would like to thank my scientific adviser, Prof. Sergey Steinberg, who guided me through the complex process of the scientific education. His unusual approach to tackling scientific problems allowed me to develop a unique insight, which will be a valuable source of inspiration for the rest of my life. He taught me to ask unusual questions and to find even more unusual answers. With his help I learned what the structure of RNA is. Due to the constant discussions with Sergey about the structure I developed a special skill that allows me to load PDB file directly to my mind and the structural analysis does not require me to have constant access to the computer.

I would like also to thank Matthieu Gagnon, with whom I spent long hours in the wet-lab, while trying to find ways to solve numerous experimental problems. His extremely high level of responsibility made him an invaluable member of our lab for everyone who was dealing with him. Special thanks for Matthieu for teaching me French and correcting my French texts.

I would like to thank Tetsu Ishii, who is able to find a positive aspect of absolutely any event, even totally negative. Special thanks for Tetsu for being my English language teacher. His willingness to help and extreme level of respect he showed to each and every person saved me in the most difficult times of my PhD course.

I would like to thank all members of our lab: François Boulay, Konstantine Bokov and Natalia Kotlova for their help and support.

I am also very grateful to the members of my jury, who took their precious time for reading my Thesis.

I would like also to thank my family: my mother, Marina, who brought me up and who made me who I am; my wife, Liudmila, who crossed the ocean and came to Canada to always stay by me and who helped me to survive immigration; my deceased grand-mother, Zoya, who was my second mother and my best friend; my uncle, Alexander, who was always the model person for me.

Chapter 1

Introduction

1. Introduction

1.1 RNA is a versatile biopolymer

RNA is a high molecular-weight bio-polymer, which consists of a long chain of nucleotide units. Each nucleotide consists of a five-carbon sugar, a phosphate group and one of the four bases: adenine, cytosine, guanine and uracil. RNA is an extremely versatile molecule, which plays diverse functional roles in the cells of all living organisms on Earth. For example, mRNA, tRNA and rRNA are involved in the process of translation (Berg et al., 2002). mRNA participates in the transfer of genetic information from DNA to the ribosome, where this information is converted into protein sequence. tRNA delivers amino acids to the ribosome, while rRNA arranges the sequence-specific catalysis of the reaction of trans-peptidation, resulting in the synthesis of polypeptide chains. miRNA and siRNA participate in the post-transcriptional gene regulation in eukaryotes through the complex mechanism of RNA interference (Carthew & Sontheimer, 2009). Splicing is the process of removal of introns from pre-mRNA. It can proceed either with help of spliceosomes, which contain several small nuclear RNA (snRNAs), or by introns themselves, which can work as ribozymes and catalyze their own excision (Steitz & Steitz, 1993).

I will briefly discuss three major abilities of RNA that explain its functional importance: the ability to transfer genetic information, to form complex 3D structures and to catalyze chemical reactions.

1.1.1 RNA is a carrier of genetic information

As long as both DNA and RNA use the same four-letter genetic code, the primary sequence of RNA can be used for storage and transmission of genetic information. Compared to RNA, DNA is more stable; hence it is used for long-term storage. RNA, on the contrary, is extremely vulnerable and as a result, can be used only for short-time storage of genetic information and for its transmission. The fact that RNA can easily and rapidly degrade is used by cells for post-transcriptional gene regulation (Pillai et al., 2007). Representatives of the three major kingdoms of life – eubacteria, archaea and eukarya, use RNA as an intermediate carrier of genetic information in the form of

mRNA. In non-eukaryotic cells, the newly transcribed mRNA is essentially ready for translation. In contrast, in eukaryotes, a DNA encoded sequence is first transcribed into pre-mRNA, which later undergoes three steps of chemical modifications. These steps are referred to as 5' capping, 3' polyadenylation and RNA splicing. The processed strand of mRNA is then translated by ribosomes into the protein sequence (Wilson & Hunt, 2002). In certain viruses, RNA serves as a principal genetic information carrier. For example, the genome of the influenza virus is kept in the RNA form during the whole life-cycle of the virus (Bouvier & Palese, 2008). Thus, in spite of the fact that RNA is much less chemically stable compared to DNA, it is actively used as an intermediate carrier of genetic information or as the principal medium for some viruses.

1.1.2 RNA can adopt various 3D structures

Analysis of structures of various non-coding RNAs showed that in addition to double helices, the RNA chain is capable of forming diverse three dimensional forms and arrangements. The ability to fully exploit the hydrogen-bonding and base-stacking potentials of bases and riboses allows RNA to form a vast variety of structures. Various RNA molecules achieve their active states by adopting specific functional three-dimensional structures. One of the best examples of RNA molecule with a distinct 3D structure is transfer RNA (tRNA) (Figure 1A, B). tRNA is a small RNA molecule of 74-95 nucleotides that delivers a specific amino acid to the growing polypeptide chain on the ribosome. All tRNAs have a similar L-shaped 3D structure that allows them to fit into the P and A sites on the ribosome (Kim et al., 1974; Robertus et al., 1974). Non-canonical base pairs play a critical role in the formation of the specific L-shape of tRNA. For example, the reverse-Hoogsteen base pair A54-U58 was shown to be an indispensable element for the correct interaction between the D- and T- loops of the tRNA, which, in turn, is critical for the formation of the tRNA L-shape (Zagryadskaya et al., 2003).

Analysis of the principles of RNA structure formation is extremely important for better understanding of RNA function. Various aspects of RNA 3D structure will be explored in detail in the text of this manuscript. The ribosome structure will be presented later in the following sections of the Introduction, while a schematic representation of the secondary and 3D structures of the bacterial RNase P is given in the Figure 1 C,D.

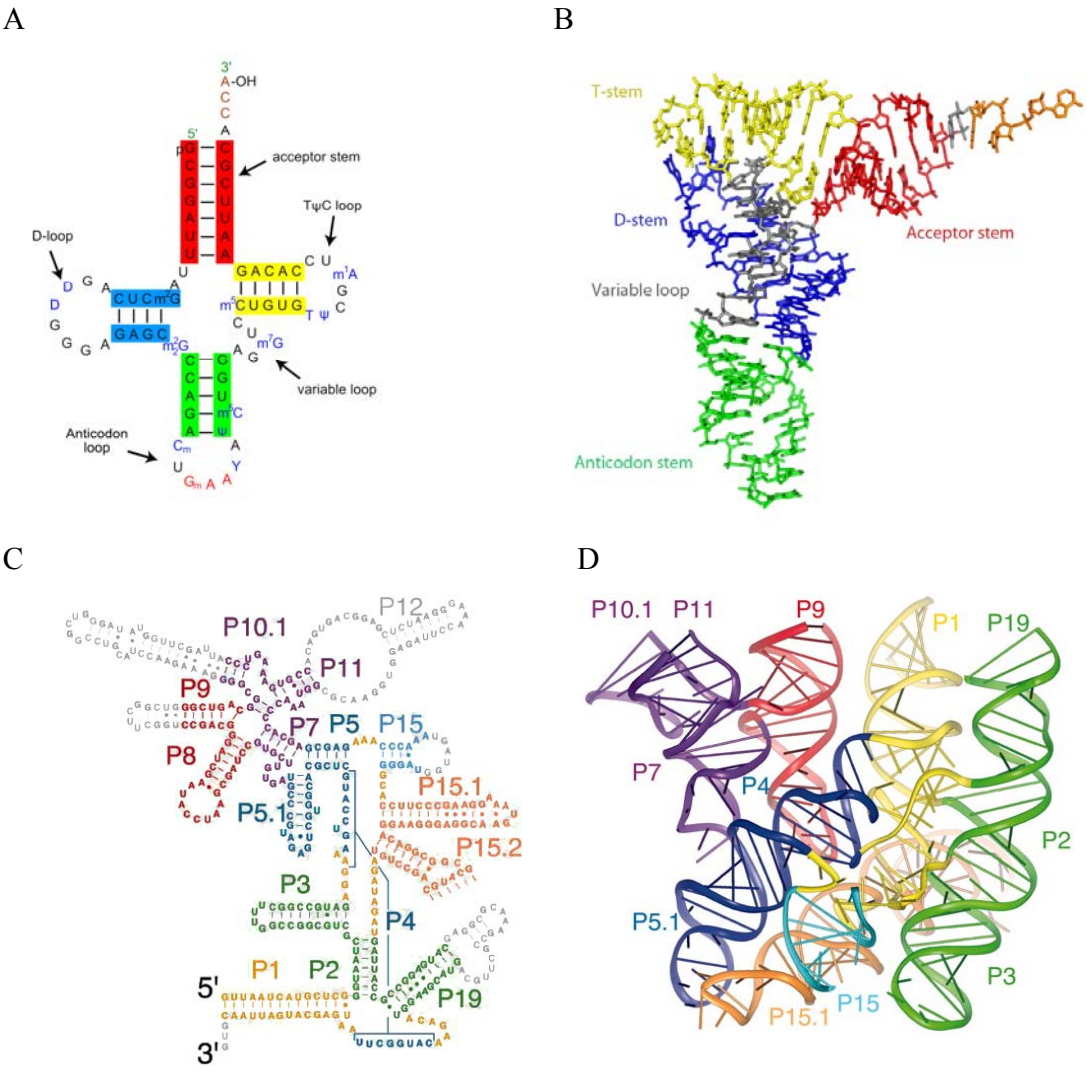


Figure 1. Secondary and tertiary structures of $tRNA^{Phe}$ and bacterial RNAseP.

(A) The canonical cloverleaf structure of the $tRNA^{Phe}$. Modified nucleotides are shown as follows: m^2G => 2-metyhl-guanosine; D => 5,6-Dihydrouridine; $m^2{}_2G$ =>: N2-dimethylguanosine; C_m => O2'-methyl-cytdine; Gm => O2'-methyl-guanosine; T => 5-Methyluridine (Ribothymidine); Y => wybutosine (Y-base); Ψ => pseudouridine; m^5C => 5-methyl-cytidine; m^7G => 5-methyl-guanosine; m^1A => 1-methyl-adenosine.

(B) 3D structure of the $tRNA^{Phe}$ forming the canonical L-shape. PDB access code of X-ray structure – 1ehz (Shi & Moore, 2000).

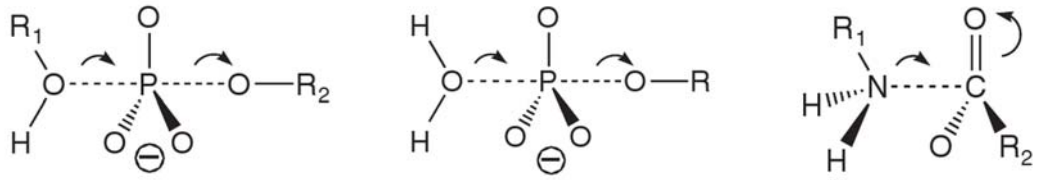
(C,D) Schematic representation of the secondary and 3D structures of the RNA component of RNase P from *Bacillus stearothermophilus* (Kazantsev et al., 2005). PDB access code of X-ray structure –2A64.

The elements of phylogenetically refined secondary structure are colored according to the coaxially stacked helical domains in the ribbon representations of the structure. The nucleotides colored in gray in the secondary structure (parts of P9, P10.1, P12, L15, and P19) could not be modeled because of disorder in the crystal.

1.1.3 RNA as an enzyme

There are two major factors that allow RNA to catalyze chemical reactions. As in the case of proteins, RNA possesses a well-defined tertiary structure. On top of that, in an RNA chain, each ribose has a 2' hydroxyl group, which can act as a nucleophilic center. This group makes RNA less stable compared to DNA because it can stimulate self-hydrolysis of the RNA phosphodiester bond. RNA molecules, which are capable of chemical catalysis, are called ribozymes (from ribonucleic acid enzyme) (Kruger et al., 1982). All known natural ribozymes catalyze one of three types of reactions: transesterification, hydrolysis and peptidyl transfer (Figure 2A). The biggest number of ribozymes performs transesterification reaction and is divided in two classes: nucleolytic ribozymes and the self-splicing introns. Both, nucleotilytic ribozymes as well as self-splicing introns perform different kinds of phosphoryltransfer reactions, which result in breakage of the RNA backbone. In nucleolytic ribozymes the phosphodiester bond is attacked by an adjacent 2'-oxygen atom, whereas in group I (Cech, 1990) and II (Lehmann & Schmidt, 2003) intron ribozymes it is attacked by a remote 3'- and 2'-oxygen atom respectively (Figure 2B, reviewed in (Lilley, 2003)). The class of nucleolytic ribozymes includes several ribozymes with unrelated structure, yet following the same hydrolytic mechanism, such as seen in the following examples: hammerhead and hairpin ribozymes, mostly found in plant viruses, the Varkud satellite (VS) ribozyme in fungal mitochondria and the hepatitis delta virus (HDV) ribozyme, which is present in a human pathogen.

A



B

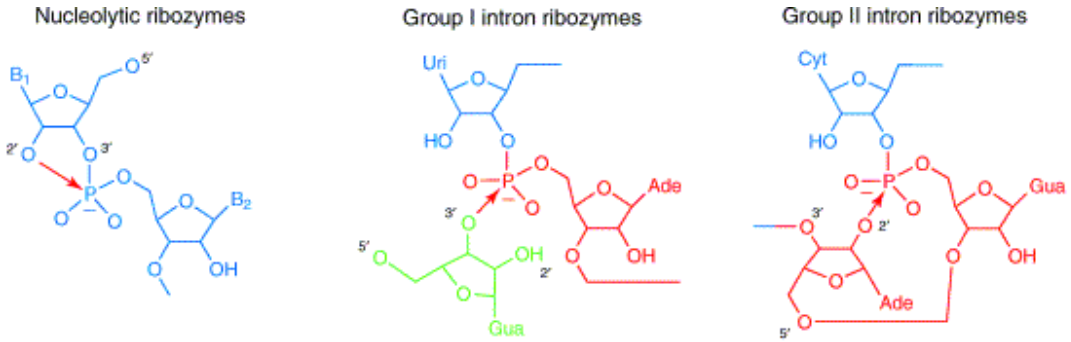


Figure 2. Reactions catalyzed by the known natural ribozymes

A. Three types of chemical reactions, catalyzed by the natural ribozymes.

B. Three classess of rybozymes that caralyse reaction of transesterification: nucleolytic, group I and group II introns.

In the nucleolytic ribozymes the phosphorus is attacked by the adjacent 2′-hydroxyl with departure of the 5′-oxygen, generating a cyclic 2′,3′ phosphate.

In the group I intron, the attack comes from the 3′-hydroxyl of an exogenous guanosine that is bound non-covalently by the ribozyme.

In the first step of the group II intron reaction, the scissile phosphorus is attacked by the 2′-hydroxyl of an adenosine that is located in the middle of the intron to be excised.

Introns are shown in red, exons in blue and exogenous guanosine in green.

The figure is taken from (Lilley, 2003) and reproduced with permission from Elsevier Limited, license number 2504960060655.

23S rRNA is a ribozyme as well because it catalyzes the peptidyl-transferase reaction, in which the α-amino group of the aminoacyl-tRNA nucleophillically attacks the ester carbon of the peptidyl-tRNA to form a new peptide bond. An initial proposal for a general acid/base catalytic mechanism involving N3 of A2451—a nucleotide of 23S rRNA in very close proximity to substrate analogues (Muth et al., 2000; Nissen et al.,

2000) - was disproved by the dispensability of A2451 for the peptidyl-transferase reaction (Polacek et al., 2001; Thompson et al., 2001; Katunin et al., 2002; Beringer et al., 2003; Youngman et al., 2004). It had been proposed that binding and orienting of substrates accounts for most of the ribosomal rate enhancement (Nierhaus, 1980). A comparison of the rate of peptide-bond formation by the ribosome and by a ribosome free model system suggested that the ribosome accelerated the reaction (almost 10^5 fold) solely by entropic effects, which may include substrate positioning, shielding the reaction from bulk solvent, or organization of the active site (Sievers et al., 2004; Trobro & Aqvist, 2005).

The first ribozymes to be discovered were the *Tetrahymena thermophila* self-splicing intron (Kruger et al., 1982) and the RNA component of RNase P (Guerrier-Takada & Altman, 1984). The discovery of RNA catalytic properties led to the proposition of the hypothesis of the RNA world by Walter Gilbert in 1986 (Gilbert, 1986). The RNA world hypothesis is based on the assumption that in the pre-biotic pre-DNA world, RNA could function as both the storage of genetic information and as the enzyme which catalyzed various chemical reactions required for pre-cellular life.

1.1.4 RNA in post-transcriptional gene regulation

The recent discovery of RNA interference (RNAi) resulted in a significant increase of attention to messenger RNA as a therapeutic target (Bumcrot et al., 2006; de Fougerolles et al., 2007; Pecot et al., 2011). RNAi is a system within living cells, which allows post-transcriptional regulation of the synthesis of protein (Fire et al., 1998). The RNAi pathway is found in many eukaryotes including humans. Some eukaryotic protozoa such as Leishmania major and Trypanosoma cruzi completely lack the RNAi pathway (Robinson & Beverley, 2003; DaRocha et al., 2004). Most or all of the components are also missing in some fungi, most notably the model organism *Saccharomyces cerevisiae* (Aravind et al., 2000). A recent study however reveals the presence of RNAi in other budding yeast species such as *Saccharomyces castellii* and *Candida albicans* (Drinnenberg et al., 2009). There are two types of molecules that are central for RNAi – microRNA (miRNA) and small interfering RNA (siRNA).

Initially, miRNAs and siRNAs appeared to be distinguished in two primary ways. First, miRNAs were viewed as endogenous and purposefully expressed products of an organism's own genome, whereas siRNAs were thought to be primarily exogenous in origin, derived directly from the virus, transposon, or transgene trigger. Second, miRNAs appeared to be processed from stem-loop precursors with incomplete double-stranded character, whereas siRNAs were found to be excised from long, fully complementary double-stranded RNAs (dsRNAs) (Tomari & Zamore, 2005). Despite these differences, the size similarities and sequence-specific inhibitory functions of miRNAs and siRNAs immediately suggested relatedness in biogenesis and mechanism. Both classes of small RNAs were quickly revealed to depend upon the same two families of proteins: Dicer enzymes to excise them from their precursors, and Ago proteins to support their silencing effector functions (Meister & Tuschl, 2004; Tomari & Zamore, 2005). Thus, these three sets of macromolecules—Dicers, Agos, and 21–23 nt duplex-derived RNAs—became recognized as the signature components of RNA silencing (reviewed in (Carthew & Sontheimer, 2009)).

siRNA are 20-25 nts long, linear, perfectly basepaired dsRNA, introduced directly into the cytoplasm or taken up from the environment (Mello & Conte, 2004). These dsRNAs are processed by Dicer into the siRNAs that direct silencing (Meister & Tuschl, 2004; Tomari & Zamore, 2005). siRNAs were originally observed during transgene- and virus-induced silencing in plants (Mello & Conte, 2004). In 2002 and 2003, centromeres, transposons, and other repetitive sequences were uncovered as another wellspring of siRNAs (Lippman & Martienssen, 2004). Shortly thereafter, functional studies in plants led to the discovery of trans-acting siRNAs (ta-siRNAs) that are diced from specific genomic transcripts and regulate discrete sets of target genes (Vazquez et al., 2004; Allen et al., 2005). More recently, other sources of endogenous siRNAs (endo-siRNAs) have been identified (Golden et al., 2008). These include convergent mRNA transcripts and other natural sense-antisense pairs, duplexes involving pseudogene-derived antisense transcripts and the sense mRNAs from their cognate genes, and hairpin RNAs (hpRNAs). Thus, it has become clear that siRNAs are not solely the products of foreign nucleic acid but arise from endogenous genomic loci as well. During the canonical RNAi pathway, the siRNA guide strand directs RISC to perfectly complementary RNA targets, which are

then degraded via cleaving activity of the Argonaute protein (Carthew & Sontheimer, 2009). One of the two RNA strands of the siRNA molecule must be perfectly complementary to the RNA target in order to promote the cleavage (Carthew & Sontheimer, 2009).

miRNAs are 20-25 nucleotides long single-stranded endogenously expressed RNA molecules that bind to the 3'UTR of target mRNAs, usually resulting in gene silencing (Bartel, 2004). The human genome may encode over 1000 miRNAs (Bentwich et al., 2005), which may target up to 60% of mammalian genes (Friedman et al., 2009), because each miRNA may repress hundreds of different mRNAs (Brennecke et al., 2005; Lim et al., 2005). Most animal miRNAs bind mRNA with mismatches, although the core region of binding (seed region) must include 2-8 Watson Crick (WC) base pairs. In contrast, most plant miRNAs bind to their target sites with near-perfect complementarity. The extent of miRNA-mRNA complementarity has been considered a key factor of the regulatory mechanism. Perfect complementarity leads to cleavage of the mRNA strand, whereas central mismatches inhibit mRNA cleavage and promote repression of mRNA translation (Matranga et al., 2005).

One of the key differences between miRNAs and most siRNAs is in the precision of their ends. Most species of a miRNA have highly exact ends, although there is a little variation. In contrast, siRNAs tend to be much more heterogeneous in end composition. It is this feature of miRNAs that has probably allowed them to interact with greater specificity on substrate mRNAs without a need for stringent complementarity or large overlap. Consequently, the processing machinery is constructed to produce miRNA duplexes with highly exact ends (reviewed in (Carthew & Sontheimer, 2009). The principal differences between miRNAs and siRNAs are listed in the Table 1.

	miRNA	siRNA (Short interfering RNA)
Occurrence	Occur naturally in plants and animals	Occur naturally in plants and lower animals. Whether or not they occur naturally in mammals is an unsettled question
Configuration	Single stranded	Double stranded
Length	19–25 nt	21–22 nt
Complementarity to target mRNA	Not exact, and therefore a single miRNA may target up to hundreds of mRNAs	100% perfect match, and therefore siRNAs knock down specific genes, with minor off-target exceptions
Action	Inhibit translation of mRNA	Cleave mRNA
Function	Regulators (inhibitors) of genes (mRNAs)	Act as gene-silencing guardians in plants and animals that do not have antibody-or cell-mediated immunity
Clinical uses	Possible therapeutic uses either as drug targets or as drug agents themselves. Expression levels of miRNAs can be used as potential diagnostic and biomarker tools	siRNAs are valuable laboratory tools used in nearly every molecular biology laboratory to knock down genes. Several siRNAs are in clinical trials as possible therapeutic agents

Table 1. The key aspects that distinguish siRNA and miRNA.

1.2 The basics of RNA structure

In this section I will give a brief overview of the essentials of RNA structure, which will be important for the better understanding of the following text.

Though RNA and DNA molecules are structurally very similar, there are two structural aspects that make them different. First, while RNA is made of ribonucleotides, DNA is made of deoxyribonucleotides (as a result there is no hydroxyl group attached to the C2' atom of the deoxyribose in DNA). Second, the complementary base to adenine in RNA is uracyl, while in DNA it is replaced by thymine, which differs by the presence of a methyl group in position 5 of the pyrimidine ring (Figure 3).

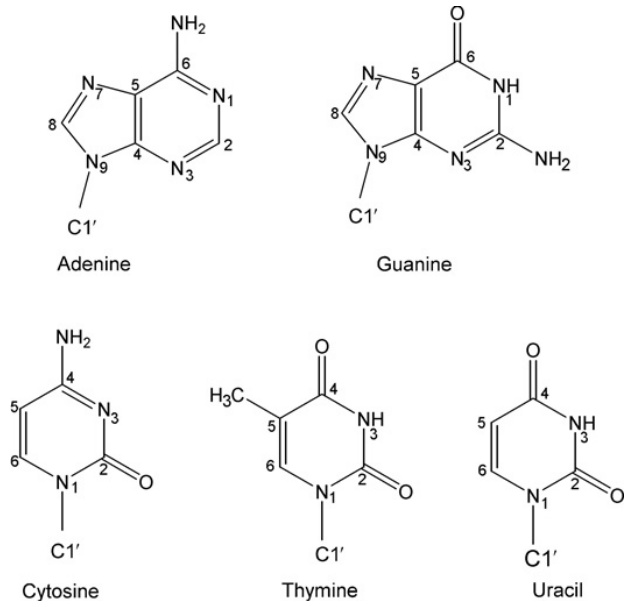


Figure 3. The five bases of DNA and RNA.

1.2.1 A- and B- helical form of nucleic acids

The RNA may form double helices through the formation of the canonical Watson-Crick (WC) base pairing between the complementary bases of the polynucleotide chain. While DNA double helices can adopt both, A- and B- helical forms, the only possible conformation of the RNA double helix is A-form (Figure 4). In DNA, both the major and the minor grooves are well accessible. The wide major groove of DNA is often used for protein binding. A-RNA double helix has a narrow and deep major groove while the minor groove is exposed for the inter-molecular interactions with other substrates and is often called the shallow groove (Figure 4). Due to the fact that the hydroxyl group is attached to the C2' atom, the ribose is locked into a 3'-endo chair conformation, thus eliminating the possibility of forming the B-helical form. If dsRNA were to form B-form duplexes, the 2'-OH group would inevitably clash with C8 (for a purine) or C6 (for a pyrimidine) as well as O5' and OP2 of the attached base (Figure 5).

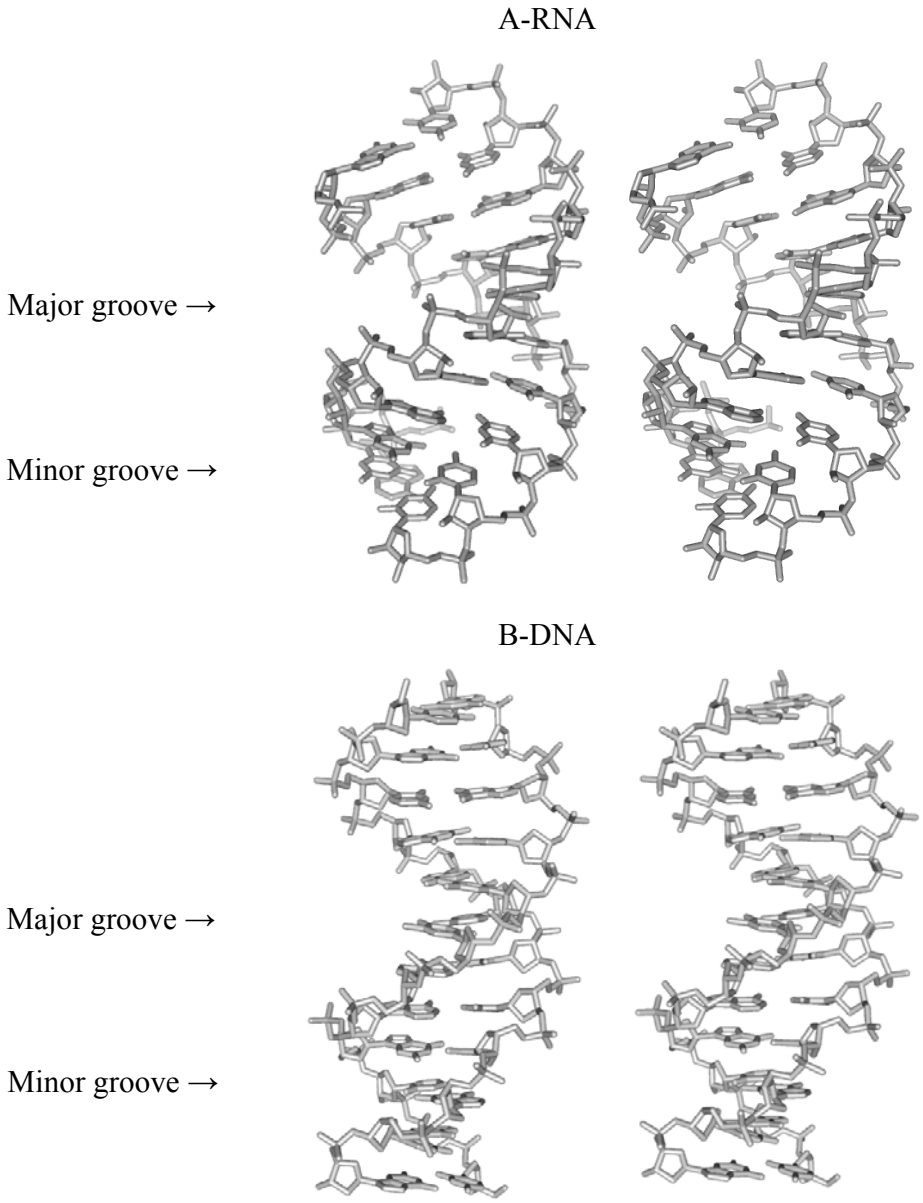


Figure 4. A-RNA and B-DNA structures.

Stereoview of 3D structure of A-RNA and B-DNA. A-RNA has a deep and narrow major groove and the minor groove is significantly exposed. The B-DNA wide major groove is almost identical in depth to the much narrower minor groove

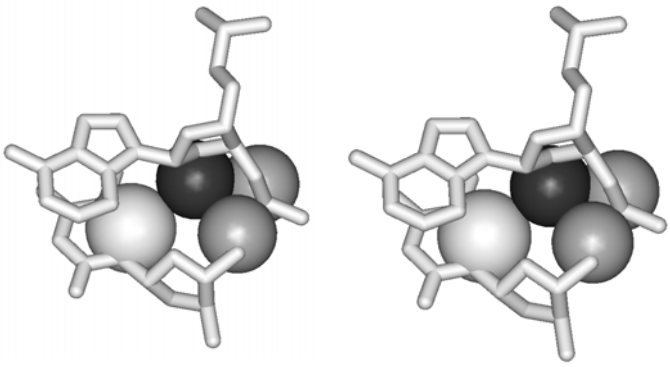


Figure 5. Potential collision of 2'OH of RNA in the B-form (stereo view)
An imaginary situation is presented, where an RNA residue is in the B-form. 2'O (black atom) collides with C6 of pyrimidine, as well as with O5' and OP2 of sugar-phosphate backbone of the neighboring nucleotide (atoms shown in gray).

1.2.2 Non-canonical base pairing

Although stacking is the most important driving force for the folding of nucleic acids, the edge-to-edge interactions between polarized atoms provide directionality and specificity. In contrast to helical B-DNA structure, RNA exploits the purine and pyrimidine base pairing not only through the Watson-Crick (WC) edge but also through Hoogsteen (HG) and Sugar edge (SE) of the nucleotide (Figure 6), which results in non-canonical base pair formation (Leontis & Westhof, 2001; Leontis et al., 2002; Lee & Gutell, 2004). Analysis of the available RNA structures shows that 60% of bases in structured RNAs participate in WC base pairing while the rest are involved in some other kind of edge-to-edge interaction (Leontis & Westhof, 2001) or even edge-to-sugar-phosphate backbone interaction.

Several attempts have been undertaken to classify the non-canonical base pairs identified in RNA structure (Jeffrey & Saenger, 1991; Tinoco, 1993; G. Dirheimer, 1995). More recent classification by Leontis and Westhof (Leontis et al., 2002) provides more comprehensive information on the non-canonical base pairing in RNA. Moreover, Leontis and Westhof proposed a convenient method of base pair annotation, which will be used throughout the text (Table 2).

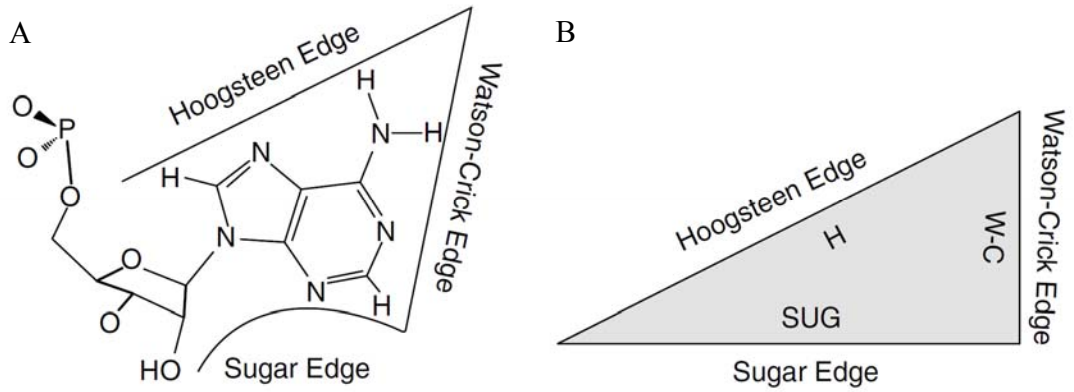


Figure 6. Definition of nucleotide edges

(A) Chemical structure of a purine nucleotide illustrating the three edges available for base-to-base interaction. (B) Representation of an RNA base as a triangle, with Hoogsteen (HG), Watson-Crick (WC) and Sugar (SE) edges labeled.

The two-dimensional diagram annotation, proposed by Leontis and Westhof (Leontis & Westhof, 2001), significantly facilitates the secondary structure representation of complex three-dimensional arrangements. As a result, this annotation is widely used in the RNA field for depicting the complex network of inter-nucleotide base pairing.

Given that either the WC, HG or Sugar edge of a nucleotide may interact with one of three available edges of another nucleotide, 6 combinations of pair-wise edge-to-edge interactions are possible. Two nucleotides can interact in *cis* or *trans*, depending on the mutual orientation of their glycosidic bonds in respect to an imaginary line drawn along the established hydrogen bonds (Figure 7). As a result all base pairs involving two or more edge-to-edge hydrogen bonds belong to one of 12 geometric families (Table 2).

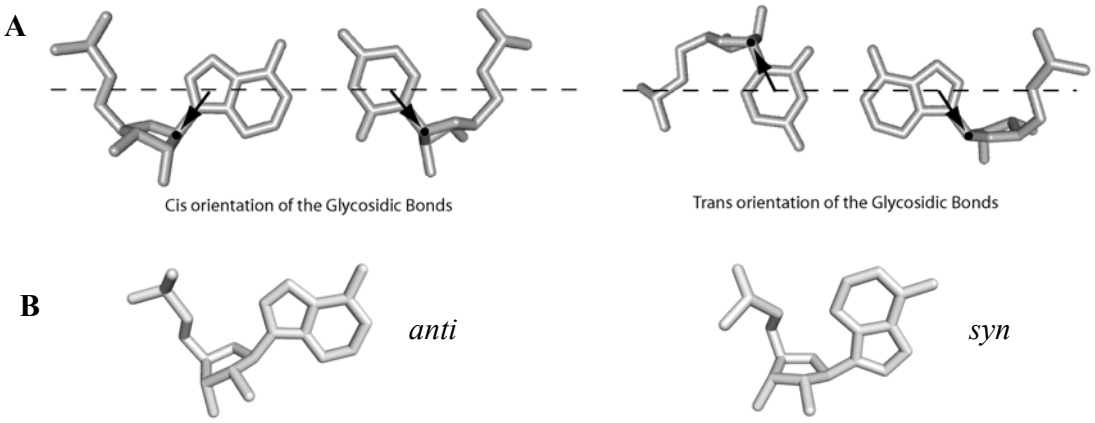


Figure 7. Glycosidic bond orientations.

(A) An example of *cis* (left panel) WC base pair versus *trans* (right panel) WC base pair. The *cis* and *trans* orientations are defined relative to a line drawn parallel to the base-to-base hydrogen bonds.

(B) Examples of *anti* (left panel) and *syn* (right panel) conformations of a ribonucleotide. The *anti* conformer has the smaller H-6 (pyrimidine) or H-8 (purine) atom above the sugar ring, while the *syn* conformer has the larger O-2 (pyrimidine) or N-3 (purine) in that position.

No.	Glycosidic bond orientation	Interacting edges	Symbol	Default local strand orientation
1	*cis*	Watson–Crick/Watson–Crick	●–	Anti-parallel
2	*trans*	Watson–Crick/Watson–Crick	○–	Parallel
3	*cis*	Watson–Crick/Hoogsteen	●■	Parallel
4	*trans*	Watson–Crick/Hoogsteen	○□	Anti-parallel
5	*cis*	Watson–Crick/Sugar edge	●▶	Anti-parallel
6	*trans*	Watson–Crick/Sugar edge	○▷	Parallel
7	*cis*	Hoogsteen/Hoogsteen	■–	Anti-parallel
8	*trans*	Hoogsteen/Hoogsteen	□–	Parallel
9	*cis*	Hoogsteen/Sugar edge	■▶	Parallel
10	*trans*	Hoogsteen/Sugar edge	□▷	Anti-parallel
11	*cis*	Sugar edge/Sugar edge	▶–	Anti-parallel
12	*trans*	Sugar edge/Sugar edge	▷–	Parallel

Table 2. The 12 geometric families of nucleic acid base pairs with symbols for annotating secondary structure diagrams (Leontis & Westhof, 2001).

The local strand orientation is given in the last column, assuming that all bases are in the default *anti* conformation; a *syn* orientation (Figure 7B) would imply a reversal of orientation.

1.3 Primary, secondary and tertiary motifs

The title of my thesis "Recurrent RNA motifs: identification and characterization" requires clear understanding of the term "motif". When speaking about "motifs" one can mean absolutely different subjects. According to the Collins English dictionary, a motif is *"a recurring form or shape in a design or pattern*" or "*a single added piece of decoration*" (Crozier, 2006). In molecular biology we can find several types of motifs. They can be either primary sequence motifs, secondary or tertiary structure motifs. We will discuss each type of these motifs in the following paragraphs.

1.3.1 Primary structure and primary sequence motifs

The primary structure of RNA is in fact the sequence of nucleotides of the RNA strand. The RNA sequence is the code which determines the secondary and tertiary structure and subsequently the function of any RNA molecule. The primary sequence motif represents a sequence of nucleotides which has a particular biological meaning. Primary sequence motifs are recognized by other nucleic acid sequences or proteins. For example, a G-N combination of a single-stranded RNA is recognized by RNAse T1, while a Py-A combination is recognized by RNAse A, which results in the strand cleavage in the middle of the dinucleotide sequence motif (D'Alessio & Riordan, 1997). The Shine-Dalgarno (SD) sequence is a conserved mRNA sequence motif, which in non-eukaryotic cells allows the correct positioning of the small ribosomal subunit on mRNA via formation of the duplex with the anti-SD sequence of 16S rRNA (Shine & Dalgarno, 1975).

1.3.2 Secondary structure and secondary structure motifs

It is generally known that the primary sequence of homologous RNA molecules may significantly vary. Yet, the three-dimensional structure varies much less. It has been demonstrated in tRNA and rRNA that conservation of the sequences involved in the formation of the double helices is less than for the single-stranded regions (Pace, 1999). The secondary structure can be determined by the alignment of multiple homologous sequences from a large number of organisms (Woese et al., 1980; Noller et al., 1981;

Gutell et al., 1992). Single nucleotide mutations in an RNA strand can occur spontaneously. If this occurs in one strand of RNA double helix, the overall stability of the helix is reduced. The destabilizing effect of a single mutation can be neutralized by a compensatory mutation of the opposite strand of the RNA double helix. As a result, the canonical base pairing is restored and the double helix is re-formed. Therefore, by looking for compensatory base changes in the alignment, it is generally possible to deduce areas forming double helices (Gutell et al., 2002).

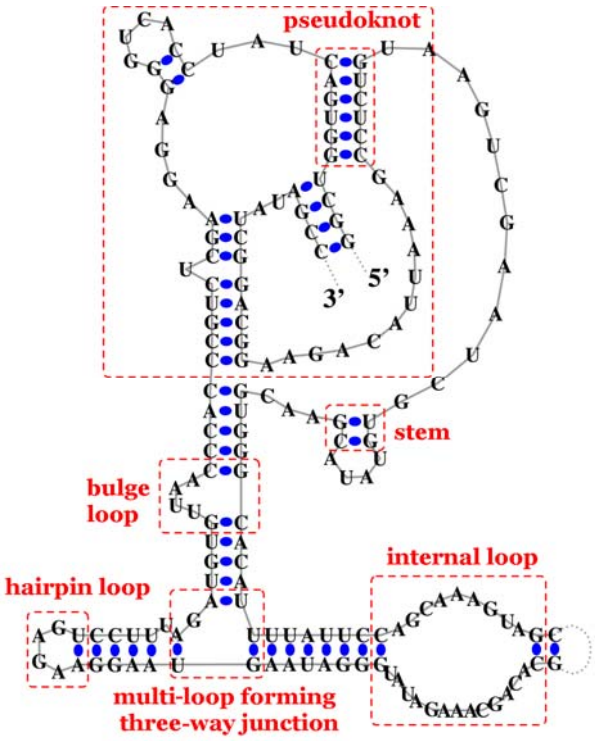


Figure 8. RNA secondary structure example**.**

The secondary structure of the RNase P RNA molecule of *Methanococcus marapaludis* from the RNase P Database is shown (Brown, 1999). Solid grey lines represent the ribose-phosphate backbone. Dotted grey lines represent missing nucleotides. Solid circles mark canonical WC base pairs. A number of secondary structure motifs can be identified here: A *stem* is composed of one or more consecutive base pairs; a *hairpin loop* contains one closing base pair, and all the bases between the paired bases are unpaired; an *internal loop* is a loop with two closing base pairs, and all bases between them are unpaired; a *bulge loop* can be seen as a variant of an internal loop in which there are no unpaired bases on one side; a *multi-loop* is a loop that has at least three closing base pairs; stems originating from these base pairs form a *multi-way junction;* a *pseudoknot* contains at least two stem-loop structures in which half of one stem is intercalated between the two halves of another stem.

The secondary structure provides information on the double helices and unpaired regions. The RNA double helix is the most common secondary structure element. The unpaired regions give rise to many other secondary structure elements: pseudoknots, bulges, internal loops, bulge loops, hairpin loops, multi-loops etc (Figure 8). Conserved secondary structures, which are identified by co-variation are often called "motifs" (Gast, 2003). A number of computational methods to predict the secondary structure have been proposed (Zuker, 1989; Gutell, 1995; Mathews, 1998; Rivas & Eddy, 1999; Hofacker, 2003) although information on the tertiary interactions or non-canonical base pairing cannot be obtained by these methods.

RNA secondary structures have been listed in various databases, such as comprehensive database of RNA families, Rfam (Griffiths-Jones et al., 2003), RNA specific databases for ribosomal RNAs (Wuyts et al., 2004), RNAseP (Brown, 1999), tmRNA (Zwieb et al., 2003), SRP (Rosenblad et al., 2003) and others. Various biochemical and biophysical experimental methods can also be used to infer secondary structure (Ehresmann et al., 1987) or in the case of NMR (Furtig et al., 2003) and X-ray crystallographic methods (Holbrook & Kim, 1997) describe secondary and tertiary structure in detail.

Once 3D structure of an RNA molecule is obtained, the secondary structure can be easily incurred by visual analysis. In this case the secondary structure can be useful for 2D representation of a complex 3D shape.

1.3.3 3D structure of RNA and tertiary motifs

The RNA tertiary structure describes the overall three dimensional conformation of a single molecule. The exact structure of RNA is usually determined by crystallography or NMR, although some biochemical methods such as probing and hydroxyl radical cleavage can be used for these studies too (Fox, 1997; Wilson, 2002; Bockelmann, 2004; Tullius & Greenbaum, 2005). The information on the crystal structures of different RNAs can be found in both the Nucleic Acid Database (Berman et al., 1992) and the RCSB Protein Data Bank (Berman et al., 2000).

RNA crystallization has always been difficult due to the problems of preparation of the adequate samples and inherent RNA flexibility. Over the past few years the RNA

crystallography and NMR techniques advanced significantly, which allowed collection of large amounts of information on 3D RNA structure.

Folding of RNA significantly differs from that of proteins. First, there are only four types of monomers used for building of an RNA molecule. The RNA backbone has six degrees of freedom for each residue while the polypeptide chain has only two. RNA structure is not nucleated by a hydrophobic core as in the case of most proteins. In contrast to this, RNA folding is driven by two major forces: hydrogen bonding and base stacking. The extreme flexibility allows single-stranded RNA regions to adopt a wide range of conformations. In spite of the fact that the 3D RNA structures are diverse and sometimes extremely complex, they can be decomposed into smaller size building blocks.

Analysis of the RNA tertiary structure has suggested that RNA molecules are made of conserved structural building blocks or motifs (Leontis & Westhof, 2003). Ability of the ribonucleotides to form a huge variety of non-canonical base pairs allows for the formation of a vast range of different structural arrangements. Generally speaking, RNA structure can be roughly split into the set of repetitive elements and those elements which occur only once and have no analogs (or these analogs have not yet been identified). Our approach to the studies of the 3D structure of RNA consists in identification of the recurrent arrangements and their detailed analysis. There are at least two reasons, which explain our interest in the RNA repetitive elements. The fact that a particular recurrent motif has been selected as a building block for different domains of RNA molecules means that such an arrangement possesses certain important properties which are beneficial for the correct folding of the whole molecule. Thus, by studying repetitive arrangements, we can concentrate on the most important elements of a given RNA structure. Another reason why the systematic study of recurrent RNA motifs is so important is the fact that they are able to fold into similar shapes in spite of different primary sequences. Identification of the rules that allow for the formation of identical 3D arrangements from unrelated sequences represents an important step for better understanding of the principles of RNA structure folding.

In the following sections a few examples of known recurrent RNA motifs will be presented. Introduction of these motifs will help in the better understanding of the matter which is presented in Chapters 2-6 of the thesis.

1.3.4 A-minor motif

The A-minor motif is one of the most simple and at the same time the most abundant tertiary structural motif in large RNA molecules. An A-minor interaction consists in the hydrogen bonding between an unpaired adenosine with the minor groove of the RNA double helix (Doherty et al., 2001; Nissen et al., 2001). Four sub-types of the A-minor interaction are known, depending on the mutual orientation of the 2' OH group of adenosine and two 2'OH groups of the receptor base pair (Figure 9). The A-minor interactions are present in practically all known structures of large RNA molecules (Cate et al., 1996; Golden et al., 1998; Ban et al., 2000; Wimberly et al., 2000; Adams et al., 2004; Kazantsev et al., 2005; Torres-Larios et al., 2005). A single A-minor interaction is relatively weak while formation of two or three consecutive A-minor interactions significantly increases the stability of the complex. Indeed, in the RNA structure we can often observe formation of consecutive stacks of two or three unpaired adenosines making A-minor interactions with closely or distantly located double helices. Within the stack, adenosines can come from the same or different RNA strands. A-minor motif mediates formation of many other tertiary structural motifs. For example, interaction of the hairpin loops with their receptor helices is mediated by the A-minor interactions of the unpaired adenosines of the loop with the minor groove of the double helix (Nagaswamy & Fox, 2002; Lee et al., 2003). Some other tertiary motifs, such as particularly folded internal loops, also rely on the A-minor interactions for the correct folding (Battle & Doudna, 2002).

A-minor motif forms many important structural contacts and on top of this it plays several important functional roles in the ribosome. For instance, the 3'-terminal adenosines of both the A- and P-site tRNA are positioned in the peptidyl transferase site via formation of A-minor interactions with 23S rRNA (Nissen et al., 2000). Another important functional aspect – monitoring of the correct codon-anticodon base pairing – is

mediated by the adenosines A1492 and A1493 of 16S rRNA with help of the A-minor motif formation (Ogle et al., 2001).

In Chapter 2 a new recurrent motif will be discussed which we identified in the ribosome and named "G-ribo". An A-minor interaction is found in the centre of all G-ribo motifs, where it helps to support a particular shape of this complex arrangement.

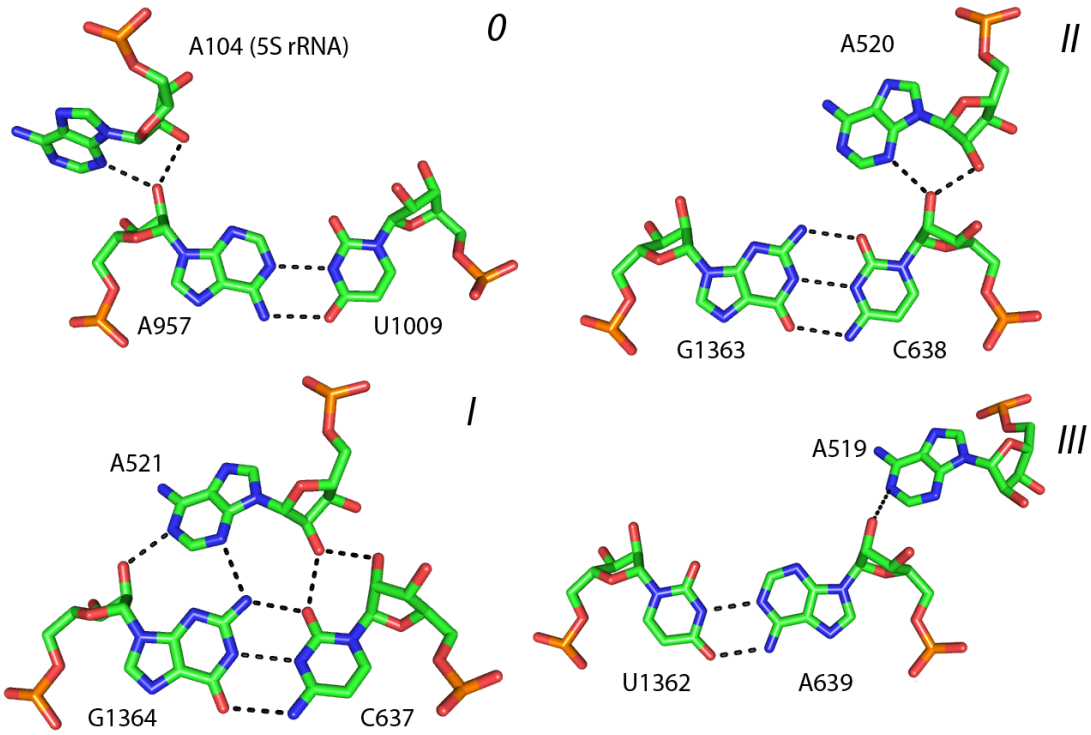


Figure 9. Four sub-types of A-minor interaction.

Four sub-types of A-minor interactions are identified in the crystal structure of 50S ribosomal subunit of *H. marismortui* (PDB: 1S72): 0, I, II and III. Each type is defined by the position of 2'-OH group of the adenosine in respect to the position of the two 2'-OH groups of the receptor base pair.

1.3.5 RNA Bulge

Incorporation of a single or several nucleotides in the otherwise regular (helical) region results in the formation of the bulge motif (Figure 10A). Although in the secondary structure the bulge motif is represented as an unpaired nucleotide which is flanked with two WC base pairs, on the level of the 3D structure the presence of the bulged nucleotide may result in different structural consequences for the RNA double

helix. (1) The bulged nucleotide may introduce a bend in an RNA double helix, if it intercalates between the 5' and 3' flanking nucleotides of the polynucleotide chain (Figure 10B). (2) Being stacked on the 3' flanking nucleotide, the bulged nucleotide may interact with the minor groove of 5' flanking base pair, thus stimulating formation of the over-twist between two WC base pairs (Figure 10C). (3) Being stacked on the 5' flanking nucleotide, the bulging nucleotide may interact with the major groove of the 3' flanking base pair, thus stimulating the formation of the over-twist between the two WC base pairs (Figure 10D). (4) The bulged nucleotide can be involved in the long-range interactions with other parts of the RNA molecule, stimulating formation of the helical over-twist between the two flanking base pairs (Figure 10E). (5) The presence of the bulging nucleotide may have no effect on the over-all shape of the RNA double helix (Figure 10F). Consequently, the bulge motif is an important structural element that can modulate the shape and direction of the RNA double helix. Different types of the bulge motif are discussed here in order to facilitate understanding of the more complex arrangements that will be presented in Chapter 4 and are in fact relevant to the structures presented here.

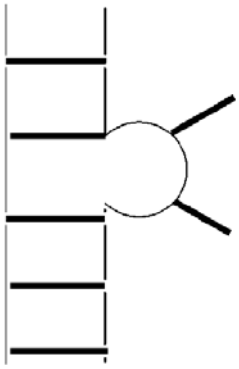


Figure 10A

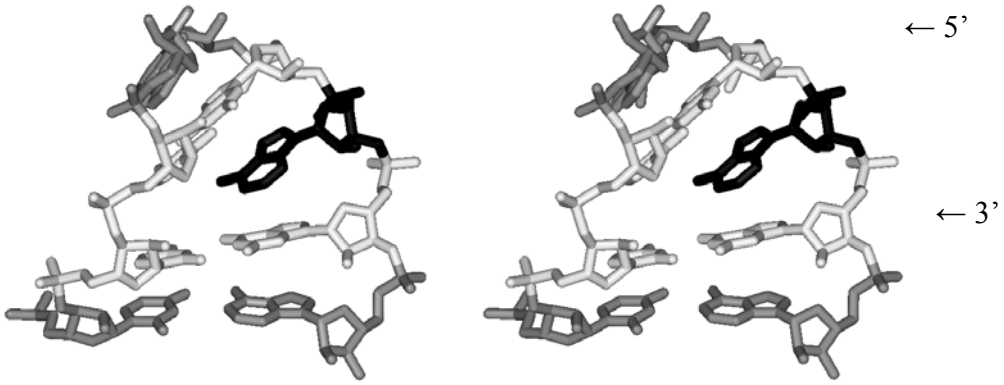


Figure 10B

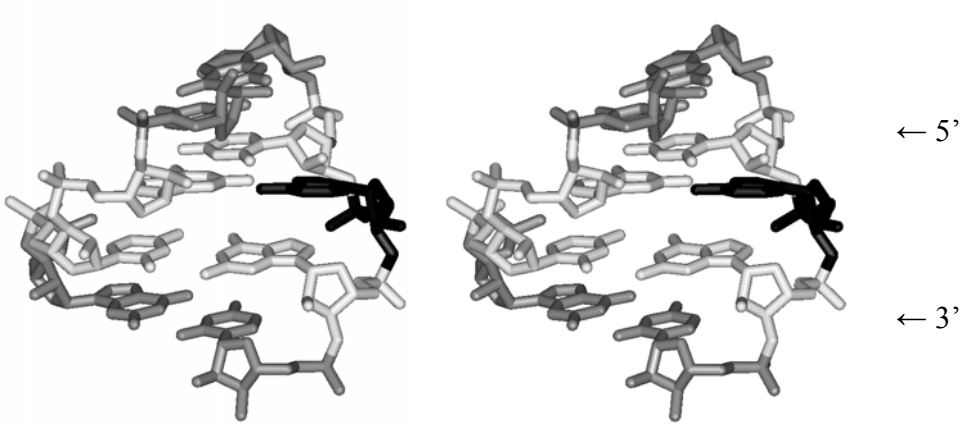


Figure 10C

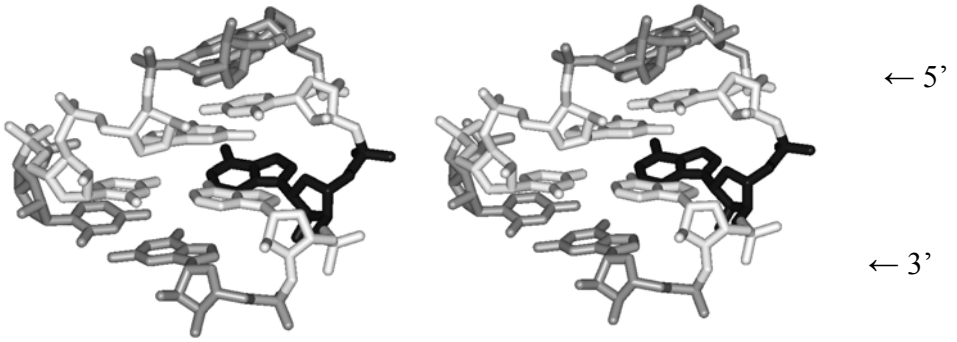


Figure 10D

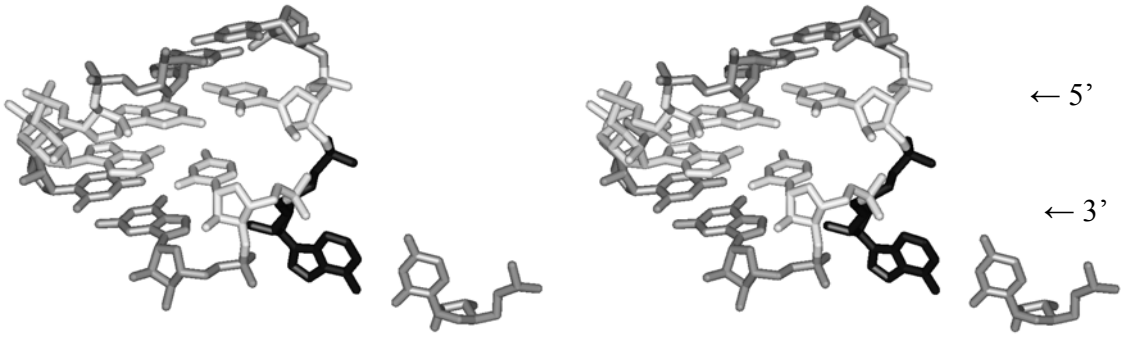


Figure 10E

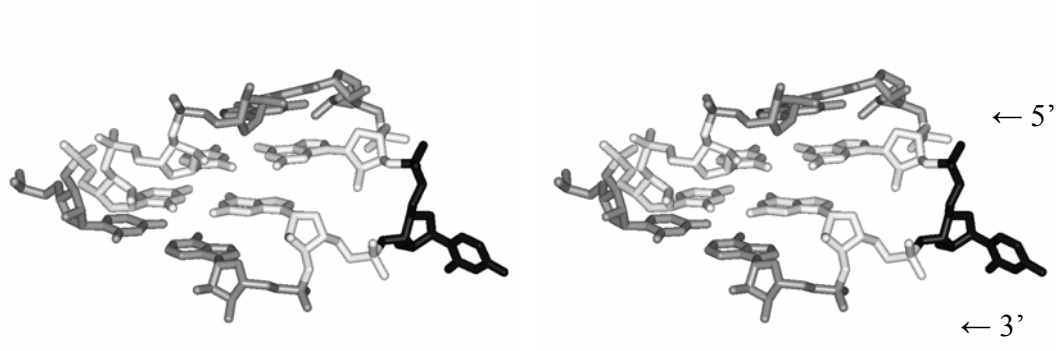


Figure 10 F

Figure 10. 3D structures of the RNA bulge.

(A) A simplified representation of the bulge. The thick lines represent base pairs in duplex regions and unpaired bases in the loops. (B-F) Stereo view of the 3D structure of different bulges. Bulged nucleotide is colored black. Two flanking base pairs are colored white. Other nucleotides of the double helix and nucleotides, which do not belong to the helix, are colored gray. “ 5’ “and “ 3’ “ indicate the 5’ and 3’ flanking base pairs of the bulge respectively. (B) Bulged nucleotide intercalates between two flanking nucleotides in the double helix and results in the helix bend. In C, D and E introduction of the bulge initiates formation of the helical over-twist between two flanking base pairs. (C) The bulged nucleotide interacts with minor groove of one of the 5’ flanking base pair. (D) The bulged nucleotide interacts with major groove of 3’ flanking base pair. (E) The bulged nucleotide does not interact with the flanking base pairs and participates in the long-range interactions with nucleotides, which do not belong to the helix. (F) The bulged nucleotide does not interact with the flanking base pairs. No helical over-twist occurs between two flanking base pairs of the bulge.

1.3.6 Internal loops

An internal loop represents a stretch of unpaired nucleotides in both strands of the RNA double helix which are flanked with regular WC base pairs on each side (Figure 11). These unpaired nucleotides may participate in non-canonical base pairing or can be bulged out of the double helix. Depending on the presence of equal or unequal number of nucleotides in each of the two composing strands, internal loops are referred to as symmetric or asymmetric (Figure 11). Loop-E motif pairing (Wimberly et al., 1993; Correll et al., 1997), C-loop (Lescoute et al., 2005) and UAA/GAN (Lee et al., 2006) are the most common examples of the internal loop-based three dimensional motifs.

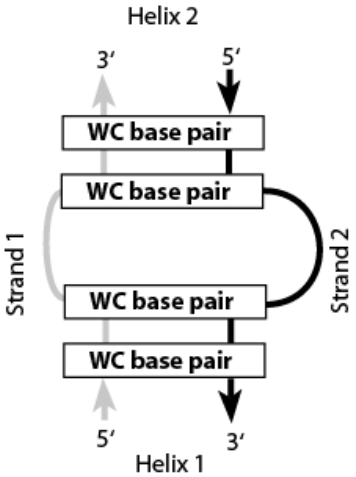


Figure 11. A simplified scheme of an internal loop.

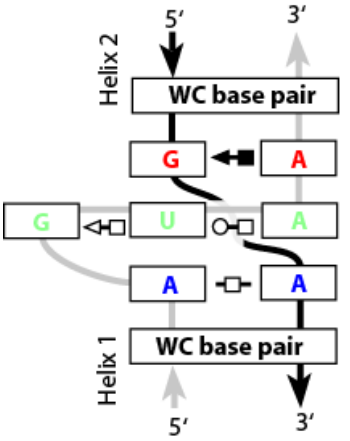
Long horizontal rectangles represent WC base pairs of the RNA double helix. The unpaired nucleotides of the internal loops belonging to the strand 1 and 2 are shown in gray and black respectively. If the number of nucleotides in both strands, 1 and 2, is equal, the internal loop is called symmetric. Otherwise it is called asymmetric.

1.3.6.1 Loop-E motif

Loop-E motif is an asymmetric internal loop that includes seven highly conserved nucleotides forming two non-WC base pairs and one base triple (Wimberly et al., 1993; Correll et al., 1997). The first base pair is trans-HG-SE (sheared) AG base pair (Figure 12, shown in red), followed by a trans-WC-HG (reverse Hoogsteen) UA base pair (shown in green), a bulged G, and finally a trans-HG AA base pair (shown in blue on the Figure 12). The three-adenosine stack represents a sticky surface that can easily bind to the minor groove of an RNA helix via formation of the A-minor interaction. In fact, most of

the Loop-E motifs located in the ribosome, indeed participate in long-range interactions with different RNA double helices. Thus, the loop-E motif is an important structural element, which participates in the organization of the three dimensional shape of RNA molecules. The loop-E motif is found in different RNA-containing molecules and has important functional implications. In particular, this motif was found in the sarcin-ricin loop of the 23S rRNA, which is known to mediate the binding of the elongation factors EF-Tu and EF-G to the ribosome. In the *E. coli* 5S rRNA, loop E represents a binding site for the ribosomal protein L23 (Leontis & Westhof, 1998). Therefore, the loop-E motif seems to be important for formation of the ribosome structure as well as for its function.

A



B

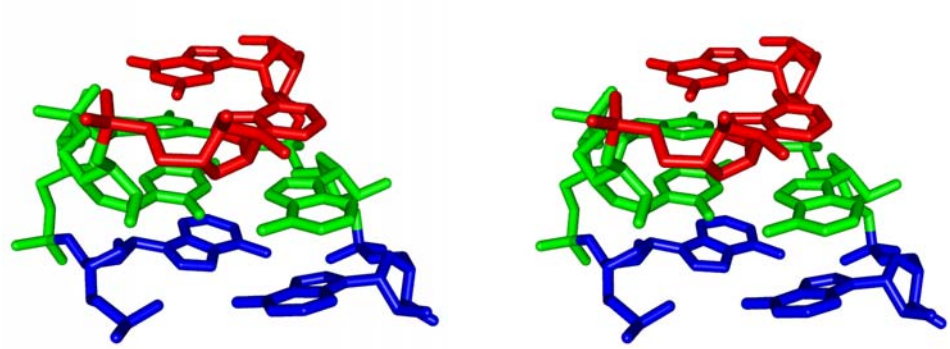


Figure 12. E-loop motif.

(A) Secondary structure of the E-loop motif. The E-loop motif represents an asymmetric internal loop. The non-canonical base pairing between pairs of nucleotides are demonstrated using symbols, suggested by Leontis and Westhof (Leontis & Westhof, 2002). (B) Stereo-view of the 3D structure of the E-loop motif. The same color code is used as in (A). The two flanking WC base pairs are not shown for clarity.

1.3.6.2 C-loop motif

The C-loop motif is another example of an asymmetric internal loop (Lescoute et al., 2005). It consists of two helices (H1 and H2) connected by an internal loop and arranged in the way that the first base pair of the second helix is partially stacked to the last base pair of the first helix. Within the C-loop motif, these two base pairs are over-twisted one in respect to the other compared to the juxtaposition of the two neighboring base pairs in the standard A-RNA conformation. The name "C-loop" is given to the whole arrangement because of the first nucleotide of the longer strand (C865 on the Figure 13A), which interacts with the major groove of the last base pair of the H1 and is always cytosine. The other nucleotides of the longer strand make hydrogen bonds with the minor groove of the first and the second base pair of the H2 (Figure 13). Nucleotides of the shorter strand bulge out and participate in long-range interactions with distant segments of the RNA molecule. The C-loop motif significantly increases the helical twist of the stem between the two WC base pairs flanking the internal loop. Together with other internal-loop based motifs, the C-loop motif represents a tool for modulation of the structure of the RNA double helix.

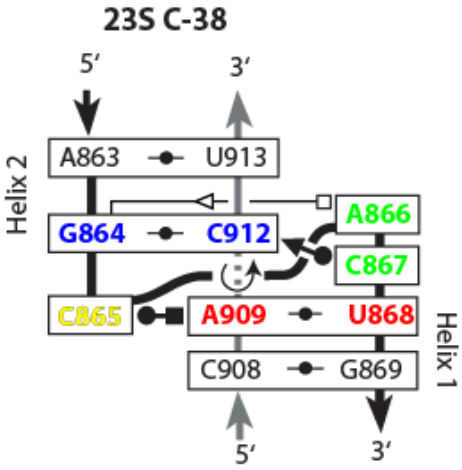


Figure 13A (See legends on the next page)

B

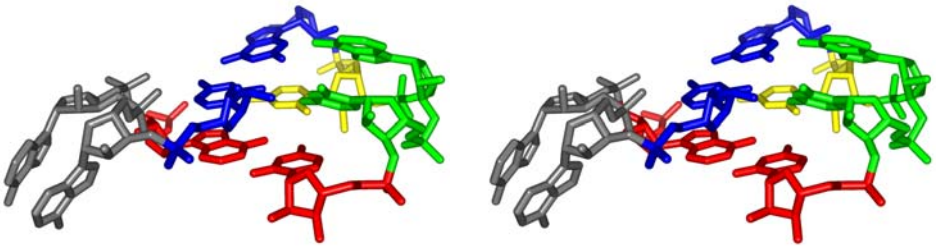


Figure 13. C-loop motif.

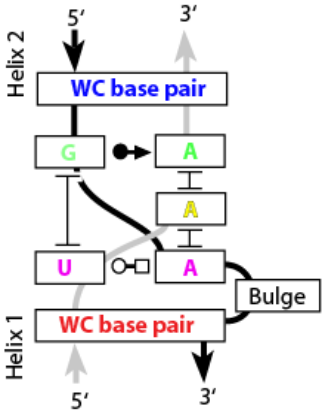
Secondary and 3D structures of the C-loop motif 23S C-38 (Lescoute et al., 2005). (A) Secondary structure of the motif. Nucleotides of the longer strand make non-canonical base pairs with minor groove of Helix 2 (C867 and A866, shown in green) and major groove of helix one (C865, shown in yellow). The name C-loop is given due to the conserved interaction of nucleotide C865 with the major groove of AU base pair [A909; U868], shown in red. Nucleotides of the short strand are not shown because they do not interact with other elements of the motif. (B) Stereo view of the motif. The same color code as in (A) is used.

1.3.6.3 UAA/GAN motif

UAA/GAN motif represents an internal loop that includes the sequence 5'-UAA on one strand (Figure 14, shown in gray) and 5'-GAN on the other strand (shown in black) forming a particular three-dimensional shape (Lee et al., 2006). Two non-canonical base pairs represent a characteristic feature of the motif. One of them is trans-WC-HG UA base pair (shown in magenta) and another is cis-WC-SE (sheared) GA base pair (shown in green). Guanosine from GA base pair and uridine from UA base pair stack to each other. Two adenosines from one strand (yellow and green) and one adenosine from another strand (magenta) make a three-nucleotide stack. The middle adenosine of the stack (shown in yellow) has no base-pairing partner, which results in the 40°−60° bend of the helix. The helix bend increases the exposure of the stacked adenines for the tertiary contact formation by expansion of the minor groove. In all UAA/GAN motifs, found in the 23S rRNA, the three-adenosine stack forms A-minor interactions with receptor helix from the distant region of the RNA molecule. Together, these motifs play an important role in the shaping of the global architecture of the 50S ribosomal subunit. Another feature of the motif, which makes it an important structure-forming arrangement, is the last nucleotide of the GAN strand (shown in black). The bulged nucleotide is also

available for the tertiary contacts with the rest of the molecule and thus is important for the correct folding of the RNA molecule.

A



B

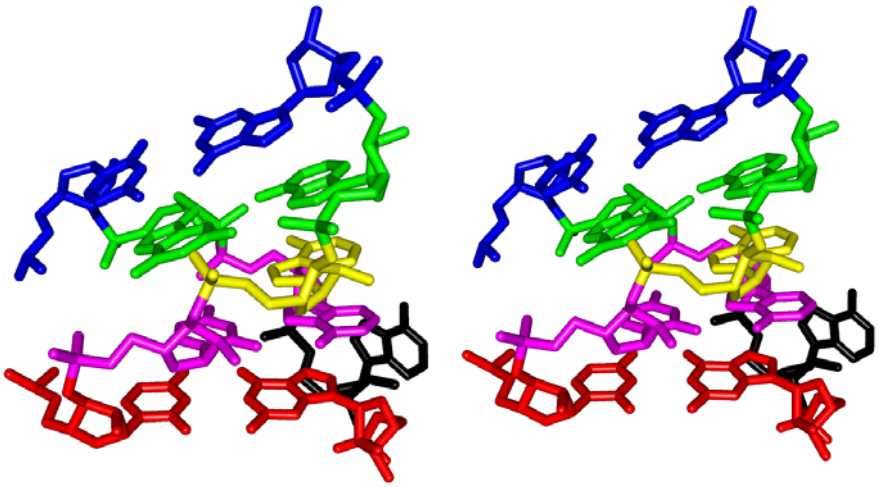


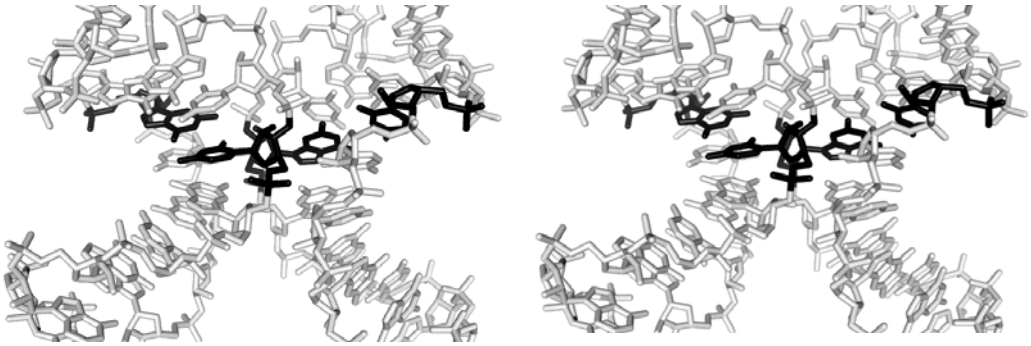
Figure 14. UAA/GAN motif.

(A) Secondary structure of the UAA/GAN motif. The motif consists of two strands, shown in gray and black, flanked by two helices. The base pairs of two helices 1 and 2 are shown in red and blue respectively. Unpaired nucleotides of the internal loop make non-canonical base pairs. The “bulge” may include one or several nucleotides, which flip out of the helix and open for the tertiary contacts. (B) Stereo view of the UAA/GAN motif. Example of the motif E23S-999 from *E. coli* 23S rRNA is shown (Lee et al., 2006). The same color code as in (A) is used

1.3.6.4 Along Groove Packing motif

Analysis of the ribosome structure shows that multiple helical regions of rRNA are interacting either with non-structured elements (such as single strand regions) or with other helical regions. Helix-helix contacts thus represent a universal type of rRNA packing. Recently, a new type of helix-helix interaction was discovered in the ribosome structure and named along-groove packing motif or AGPM (Gagnon & Steinberg, 2002). AGPM consists of two double helices closely packed via their minor grooves. In AGPM, a characteristic GU base pair of one helix is required to pack with a WC base pair of the other helix. The presence of the positively charged amino-group of the GU base pair significantly stabilizes the inter-helix contact and partially neutralizes repulsion of the negatively charged phosphates of the helices' backbones. GU base pair initiates the formation of a specific network of five hydrogen bonds between the central base pairs of both helices. The along-groove packing motif has been found in four places in the 16S rRNA and in eight places in the 23S rRNA. Two more cases of this motif are involved in the fixation of the tRNAs in the ribosomal P and E sites.

A



B

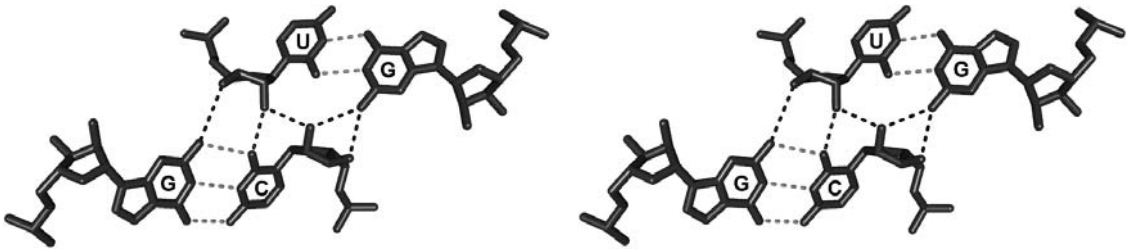


Figure 15. Along-groove packing motif

A. 3D structure of the motif, which represents close packing of two helices via minor grooves.

B. The central base pairs of two motif-forming helices, GU and GC, establish five hydrogen bonds, which are shown with black dotted lines. Hydrogen bonds between WC-edges of nucleotides are shown with gray dotted lines.

1.4 Overview of the basics of protein synthesis

Our studies focus primarily on the ribosome. Although our laboratory is primarily interested in the principles of ribosome structure formation, our research is tightly linked to ribosome function. In other words, we study the structure-functional relationships within the ribosome. A more detailed overview of the ribosome structure will be given later in the text, while here I will briefly describe a few important functional aspects of protein synthesis carried out by the ribosome.

The ribosome is an organelle that performs the protein synthesis in the cells of all living organisms. According to the canonical scheme, the genetic information is stored in DNA, from where it is transcribed into mRNA, which in its turn is translated into proteins by the ribosome (Crick, 1970). The bacterial ribosome (70S) consists of two subunits, large (50S) and small (30S), with the total molecular weight of approximately 2.5 MDa. The 30S subunit consists of 21 proteins and 16S rRNA (1540 nucleotides long), while the 50S subunit consists of about 34 proteins and two RNA molecules – 23S rRNA (2900 nucleotides long) and 5S rRNA (120 nucleotides long).

Ribosomes were first observed by electron microscopy in the mid-1950s by George Palade and were characterized as dense particles or granules (Palade, 1955). George Palade together with Albert Claude and Christian de Duve were awarded the Nobel Prize in Physiology or Medicine in 1974, for the discovery of ribosomes. The term “Ribosome” was first proposed by Richard B. Roberts in (Biophysical Society (1st symposium : 1958 : Cambridge Mass.) et al., 1958).

The large ribosomal subunit contains the peptidyl-transferase centre (PTC), which catalyzes peptide bond formation. The transfer of the mRNA message into protein sequence is mediated by the small ribosomal subunit, which assures the correct interaction between the codons of mRNA and the anticodons of tRNA (Green & Noller, 1997). Both subunits contain three binding sites for tRNA molecules, the A (aminoacyl) site, the P (peptidyl) site and the E (exit) site. The ribosomal A site binds the aminoacyl-tRNA, the P site binds the peptidyl-tRNA and the E site is occupied by the deacylated tRNA before it leaves the ribosome.

1.4.1 Initiation of protein synthesis

In prokaryotes, initiation of translation on mRNA occurs concurrently with transcription. Translation initiation is the rate-limiting and most highly regulated phase in protein synthesis. The translation initiation starts with binding of 30S ribosomal subunit by initiation factors IF1 and IF3. IF3 binding to the 30S subunit is coupled to ribosome recycling and stimulates 70S dissociation into subunits (Petrelli et al., 2001). Initiation factor IF1 stimulates activity of IF3 and as a result accelerates the dissociation of the ribosomal subunits (Gualerzi & Pon, 1990). Its main role is to specifically bind to the A-site of the 30S ribosomal subunit blocking an aminoacyl-tRNA from entering into this site and thus directing the initiator tRNA to the ribosomal P-site (Dahlquist & Puglisi, 2000; Carter et al., 2001). When ribosomal subunits dissociate, IF2, mRNA, and fMet-$tRNA^{fMet}$ associate with the 30S ribosomal subunit in an unknown and possibly random order. mRNA binds to the 30S subunit, where its so-called Shine-Dalgarno (SD) sequence forms a mini-duplex with the complementary sequence of the 3' end of the 16S rRNA, named anti-Shine-Dalgarno sequence (anti-SD) (Shine & Dalgarno, 1974). IF2 binds to an initiator tRNA and controls its entry into the P-site of the 30S ribosomal subunit (Tomsic et al., 2000). Binding of the fMet-$tRNA^{fMet}$ to the ribosomal P-site is stabilized by IF3, which controls accuracy of selection of the initiator tRNA (Gualerzi et al., 1977). When 30S initiation complex is formed, initiation factors IF1 and IF3 are ejected, and association of the 50S ribosomal subunit is stimulated by IF2. IF2 is a GTPase, which hydrolyses bound GTP to GDP and phosphate (Pi). When initiator tRNA is adjusted to the correct position in the P-site, IF2 is released from the pre-initiation complex and is accompanied by hydrolysis of one GTP molecule (reviewed in (Laursen et al., 2005). Ribosomes are now ready for the elongation step of translation. When 70S initiation complex is formed, fMet-$tRNA^{fMet}$ is positioned in the P site, while the A-site is vacant and ready to accept the ternary complex of amino-acyl-tRNA with EF-Tu and GTP.

1.4.2 Elongation

Binding of the fMet-tRNA^{fMet} in the P site results in a conformational change that opens the A site for the new aminoacyl-tRNA to bind. Delivery of the correct aminoacyl-tRNA to the ribosomal A site is mediated by EF-Tu elongation factor and involves at least two steps. The anti-codon of a new-coming aminoacyl-tRNA interacts with the codon of mRNA, located in the 30S A-site. Analysis of the crystal structures of the small ribosomal subunit in complex with mRNA and a cognate anticodon stem-loop in the A-site (Ogle et al., 2001) as well as the structure of the 70S ribosome with tRNAs in all three sites (Selmer et al., 2006), allowed for understanding the structural principles of selection of the cognate tRNAs during elongation. In these structures three conserved nucleotides of the 16S rRNA (A1492, A1493 and G530) change their conformations and interact with the minor groove of the codon-anticodon mini-helix, which is formed between mRNA and the cognate tRNA in the A-site of the 30S subunit (Ogle et al., 2001). The first position of the cognate codon-anticodon duplex is monitored by nucleotide A1493, which makes type I A-minor interactions with the minor groove of the base pair (Figure 16A). Adenosine A1492 makes type II A-minor interaction with the minor groove of the second position of the codon-anticodon duplex (Figure 16B). Nucleotide G530 undergoes conformational rearrangement of the base from *syn* to *anti*, and makes a series of hydrogen bonds with A1492 as well as with the sugar-phosphate backbone of tRNA nucleotide positions 36 and 35, which are involved in base pairing with the first and the second positions of mRNA codon (Figure 16B,C). Interestingly, the third position of the codon-anticodon mini-duplex is not monitored as precisely as in the case of the first two positions, which is consistent with the observation that wobble base pairs can occupy this position (for example, GU base pair) (Crick, 1966).

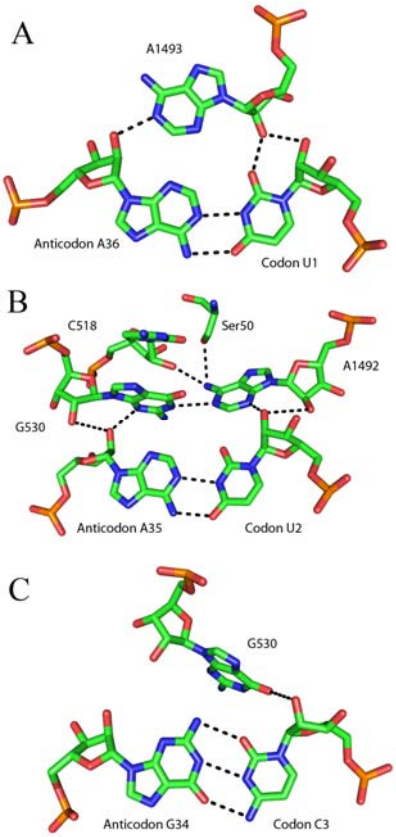


Figure 16. tRNA decoding by the ribosome

A. Nucleotide A1493 from the 16S rRNA monitors geometry of the base pairing between the first codon position and A36 of tRNA;

B. The geometry of base pairing between U2 in second codon position and A35 of tRNA is monitored by A1492 and G530;

C. The geometry of the base pair between the third codon position and A36 from tRNA is less stringently monitored with slight involvement of G530.

Delivery of a correct aminoacyl-tRNA in the A-site is communicated to the GTPase centre located in the 50S subunit through a series of conformational changes in the small ribosomal subunit (Ogle et al., 2002). As a result, GTP bound to EF-Tu is hydrolyzed, which causes the release of EF-Tu from tRNA. As soon as the aminoacyl end of the tRNA is accommodated in the Peptidyl Transferase Centre (PTC) on the 50S subunit, the new peptide bond is rapidly formed (Moazed & Noller, 1989; Pape et al., 1998). Now the ribosomal A site has a peptidyl-tRNA, with tRNA and elongated by one amino acid peptide, while the P site is occupied with a deacylated tRNA.

1.4.3 Translocation

After the reaction of the peptidyl-transfer is over, the 3' end of the tRNAs, located in the P and A site of the 50S subunit move to the E and P site, respectively (Figure 17*f*). This movement of the CCA ends of tRNAs occurs spontaneously and results

in the formation of hybrid states (Moazed & Noller, 1989). Aminoacylated tRNA with polypeptide, attached at its 3' end, is now in the A/P state, while deacylated tRNA is found in the P/E hybrid state. Formation of the hybrid state was first revealed by biochemical methods (Moazed & Noller, 1989) and then visualized by cryo-electron microscopy (Cryo-EM) studies (Valle et al., 2003). The next step is translocation of the anticodon stem-loops of both tRNAs on the 30S subunit. Binding of EF-G*GTP complex to the ribosome results in the rotation of the 30S subunit in respect to the 50S subunit, known as ratchet-like movement (Frank & Agrawal, 2000). This rotation allows the advancement of mRNA through the 30S subunit, which is required for the translocation on the 30S subunit. Hydrolysis of GTP molecule by GTPase activity of EF-G (Rodnina et al., 1997) is followed by translocation of tRNAs together with mRNA on the 30S subunit (Valle et al., 2003). Deacylated and polypeptide bound tRNAs are now in the classical E and P/P states (Figure 17*i*). The same series of events will occur many times, until one of the stop codons appears in the ribosomal A-site, which results in termination of translation.

1.4.4 Termination

Three nucleotide triplets within mRNA serve as signals for the translation termination: UAG ("amber"), UAA ("ochre") and UGA ("opal"). Termination of translation in bacteria is mediated by three release factors: RF1, RF2 and RF3 (Weaver, 2005). RF1 recognizes UAA and UAG while RF2 recognizes UAA and UGA stop codons. RF3 is a GTP-binding protein that binds to the complex of RF1/2 with the ribosome (Zavialov et al., 2001). The binding of RF1 or RF2 to one of the stop codons in the A-site results in the hydrolysis and release of the ester-linked polypeptide on the P-site tRNA. When termination is completed, RF3 promotes dissociation of RF1 or RF2 from the ribosome. After peptide release, the ribosome still contains mRNA and tRNA bound in the P site. Dissociation of the complex of 70S, mRNA and tRNA into individual components is mediated by the cooperative action of RRF (Ribosomal Release Factor) and EF-G (Hirokawa et al., 2005; Peske et al., 2005; Zavialov et al., 2005).

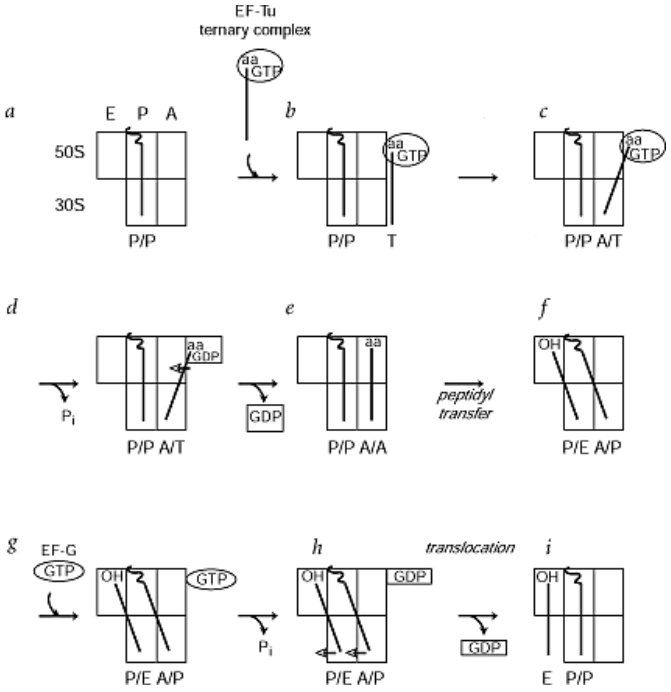


Figure 17. Model for the ribosomal hybrid states

The model of hybrid states was first proposed by Moazed and Noller (Moazed & Noller, 1989). A and P sites on the 30S and 50S as well as E site on the 50S subunit are indicated schematically.

a. P site is occupied with peptidyl tRNA while A site is empty; *b*. Ternary complex of EF-Tu-GTP with aminoacyl-tRNA binds to the ribosome; *c*. tRNA anticodon interacts with the mRNA codon in the A-site of 30S subunit; *d*. If a cognate tRNA is delivered to the A-site, EF-Tu hydrolyses GTP and subsequently dissociates from the ribosome; *e*. tRNA in the A-site is accommodated and is ready for the peptide transfer; *f*. tRNA accommodation is rapidly followed by the reaction of transpeptidation, which is followed by the spontaneous move of the acceptor end of P-site tRNA to the E site (P/E state) and the A-site tRNA to the P-site on the 50S subunit (A/P state); *g*. Elongation factor G (EF-G) in a complex with GTP binds to the ribosome; *h*. GDP hydrolysis unlocks the ribosome for the movement of the anticodons of both tRNAs in the P- and A-sites of 30S subunit, bound to the mRNA; *i*. tRNAs move to the E and P-sites, while the A-site is empty (Puglisi et al., 2000).

1.5 Ribosomal structures

In this section I will give a brief chronological overview of the structural studies of the ribosome. The first crystals of the ribosome subunits which diffracted below 3Å resolution were obtained by Ada Yonath and colleagues in the early and mid-80s (Yonath et al., 1980; Shevack et al., 1985; von Bohlen et al., 1991). Even though the crystals diffracted quite well, the major obstacle on the way to deciphering the atomic structure of the ribosome was the phase problem. (Light detectors, such as photographic plates or CCDs, measure only the intensity of the light that hits them. This measurement is incomplete because a light wave has not only an amplitude (related to the intensity), but also a phase, which is systematically lost in a measurement. In x-ray crystallography, the diffraction data when properly assembled gives the amplitude of the 3D Fourier transform of the molecule's electron density in the unit cell. If the phases are known, the electron density can be simply obtained by Fourier synthesis. This Fourier transform relation also holds for two-dimensional far-field diffraction patterns giving rise to a similar type of phase problem). Steitz and co-workers were the first to overcome the phase problem of the 50S subunit structure of *H. marismortui*, which resulted in refinement of a low resolution structure in 1998 (Ban et al., 1998). To determine the phase angles, Steitz and co-workers used a cryo-EM reconstruction of the ribosomal 50S subunit from J. Frank (Frank et al., 1995) together with multiple isomorphous replacement and anomalous scattering techniques. Shortly after this work, the crystal structure of the 50S subunit of *H. marismortui* at 5Å resolution was reported by the same group (Ban et al., 1999). The same year Ramakrishnan and co-workers reported a 5.5Å structure of the 30S subunit of *T. thermophilus* (Clemons et al., 1999), followed by a 4.5Å structure of the 30S subunit of the same organism from Yonath group (Tocilj et al., 1999). Although none of these structures allowed building of the atomic models of the ribosome, they represented an important step on the way to the high-resolution structures. Eventually the first crystal structure of the 50S subunit from *H. marismortui* at atomic resolution (2.4Å) appeared in 2000 from Steitz group (Ban et al., 2000; Nissen et al., 2000), while the same year marked the appearance of the structure of the 30S subunit from *T. thermophilus* at 3.0Å by the Ramakrishnan group (Wimberly et al., 2000) and

30S subunit from the same organism at 3.3Å made by Yonath and co-workers (Schluenzen et al., 2000). Appearance of the high-resolution structure of the 50S subunit addressed one of the long-lasting controversies about the principles of the protein-synthesis catalysis. It became clear that the ribosome is indeed a ribozyme, because no peptide chain was observed closer than 18Å to the peptidyl-transferase centre on the 50S subunit (Nissen et al., 2000).

The first crystal structure of the whole bacterial ribosome from *T. thermophilus* at 5.5Å resolution was reported in 2001 by Noller and co-workers (Yusupov et al., 2001), followed by the high-resolution structure of the 70S from *E. coli* at 3.5Å, obtained by Cate's group in 2005 (Schuwirth et al., 2005). Crystallization of the complex of the 30S subunit with a short anticodon stem-loop in the A site and a short mRNA shed light on the mechanism of the selection of the cognate tRNAs during elongation (Ogle et al., 2001). Comparison of the structure of the 30S subunit with and without an anticodon stem-loop (ASL) in the A site showed that binding of the cognate ASL resulted in the switch from an open to a closed conformation of the 30S ribosomal subunit. It was also observed that binding of the cognate ASL to the A site was accompanied by flipping of adenosines A1492 and A1493 from the internal loop of helix 44 of 16S rRNA and a *syn* to *anti* switch of nucleotide G530. It was concluded that these three nucleotides thus monitor the correct base pairing in the first and second positions of the codon (Ogle et al., 2001).

A number of crystal structures with different release factors provided insight into the mechanism of stop codon recognition by the protein factors as well as mechanism of the peptide release in the PTC upon interaction with the universally conserved GGQ-loop of the release factor. The first high-resolution structure (3.1Å) of the 70S ribosome of the *T. thermophilus* with RF1 was reported from Noller's group (Laurberg et al., 2008). It was soon followed by two other structures of the *T. thermophilus* ribosome in complex with RF2 by Ramakrishnan's (Weixlbaumer et al., 2008) and Noller's (Korostelev et al., 2008) groups.

Apart from structural information obtained through high-resolution X-ray crystallography, the ribosome was intensively studied by cryo-electron microscopy method (cryo-EM). In spite of the fact that the maximum resolution of the cryo-EM

ribosome reconstruction still does not overcome 6.7Å (Villa et al., 2009), this method provides valuable information on the various functional states of the ribosome. Thus, with help of the cryo-EM it was first shown that the 30S subunit rotates around the 50S subunit, in a ratchet-like manner upon binding of the elongation factor G (EF-G) and subsequent hydrolysis of GTP (Frank & Agrawal, 2000). Another publication from the same group has shown deacylated tRNA in the P/E hybrid state (Valle et al., 2003). Finally, both aminoacylated and deacylated tRNAs have been visualized in the A/P and P/E hybrid states with help of the cryo-EM (Agirrezabala et al., 2008; Julian et al., 2008).

1.6 An overview of the rRNA structure

Our laboratory's main scientific interest is the analysis of RNA structure in general, and subsequently our major efforts are concentrated on the RNA component of the ribosome, or rRNA. In the following sub-chapter I will briefly describe some important structural features of rRNA. The main building block of RNA structure, the double helix, significantly restricts the ability of RNA to form globular structure because of its rigidity and limited geometry. The overall dimensions of the ribosome (approximately 250Å in diameter) would theoretically allow formation of helices of up to 80 base pairs. Surprisingly, almost all RNA double helices in the ribosome are not longer than 7 contiguous WC base pairs (Figure 18A, B). A general strategy exploited by the ribosome is to connect the short helices by bulge loops or internal loops, which introduce helical bends, twists and various degrees of structural curvature. Nucleotides of the internal loops participate in non-canonical base pairs as well as base-phosphate and base-ribose interactions, which define the exact geometries of the individual bends. Before the structure of the ribosome was known, it was thought that ribosomal proteins played an indispensable role in compensation of the repulsion between negatively charged phosphate groups of RNA. The appearance of the ribosome structures showed that extended regions of RNA helices are packed together with no protein assistance. Indeed, ribosomal proteins are relatively equally spread on the outer surface of the ribosome, while only some of them have long unstructured tails penetrating deeper in to the ribosome (Brodersen et al., 2002; Klein et al., 2004b). One of the factors helping to

stabilize the RNA packing is the presence of mono- and divalent cations within the body of the ribosome (Klein et al., 2004a). Such elements as the ribose-zipper (Cate et al., 1996) and Along-Groove Packing motif (AGPM) (Gagnon & Steinberg, 2002; Mokdad et al., 2006) present alternative strategies for the close RNA packing (reviewed in (Noller, 2005)). In this thesis we discuss in detail various recurrent RNA arrangements. By studying RNA motifs, we are trying to improve our understanding of the principles of RNA folding and rRNA in particular.

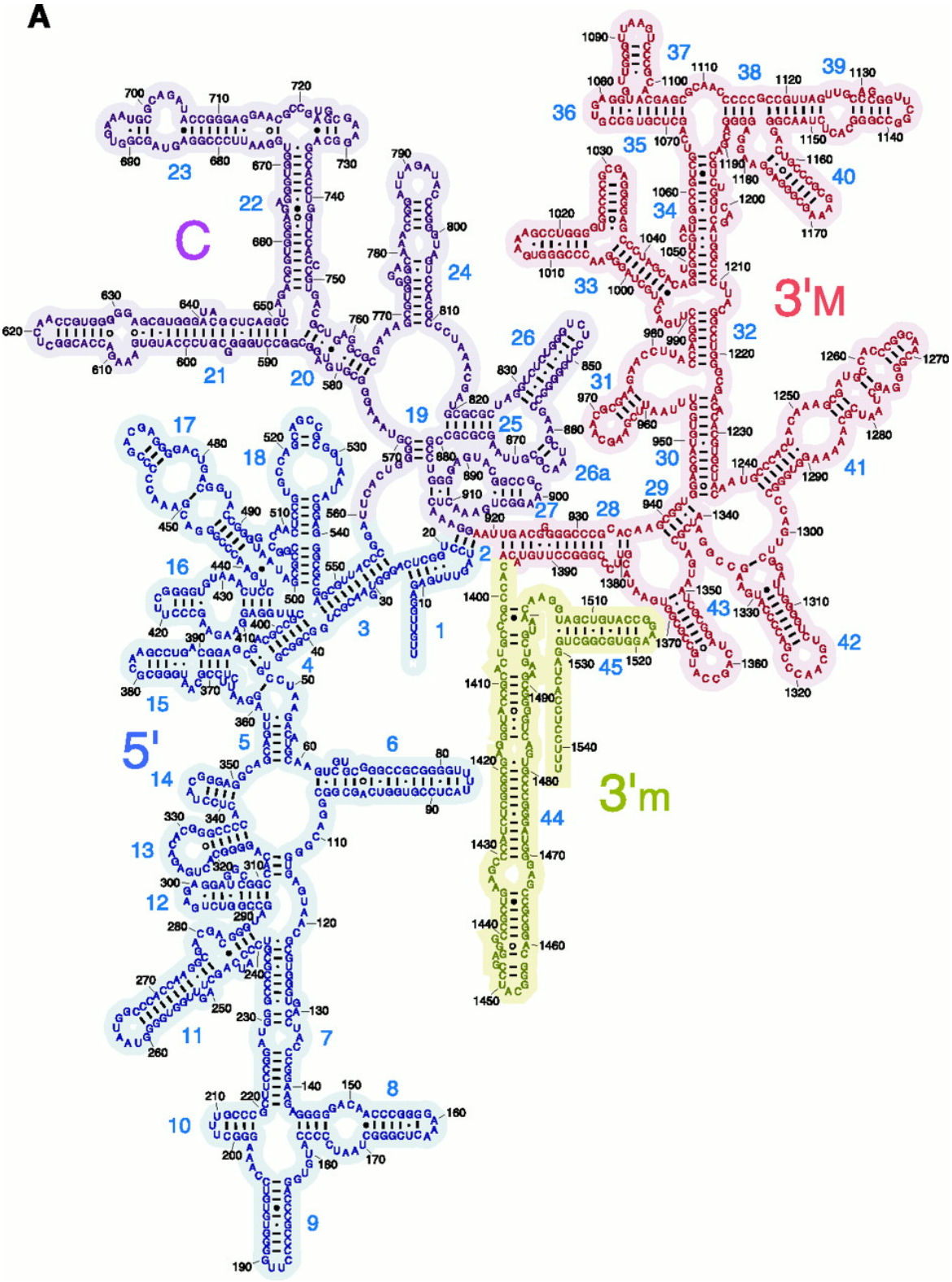


Figure 18A. See legend on page 42

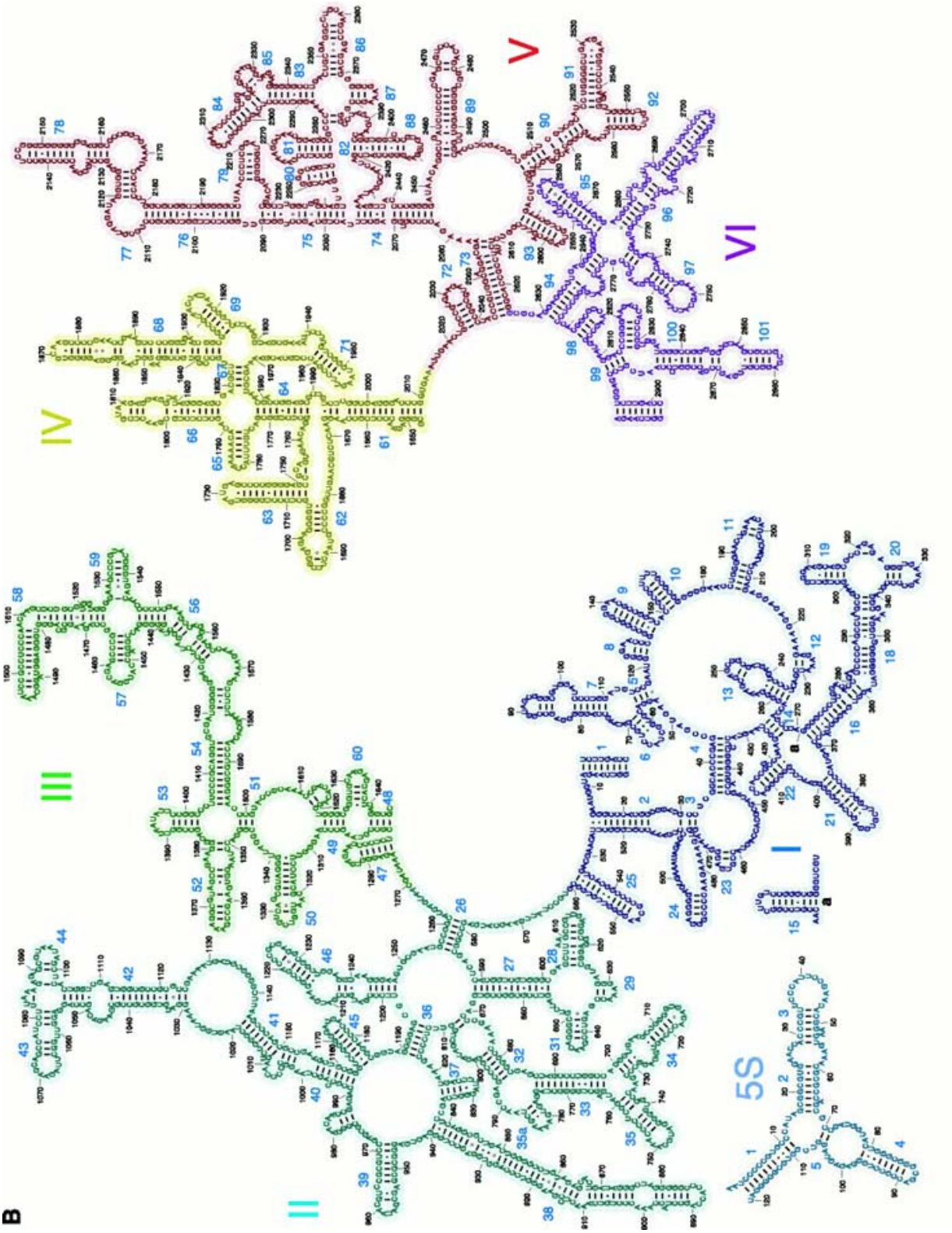


Figure 18B. See legend on page 42

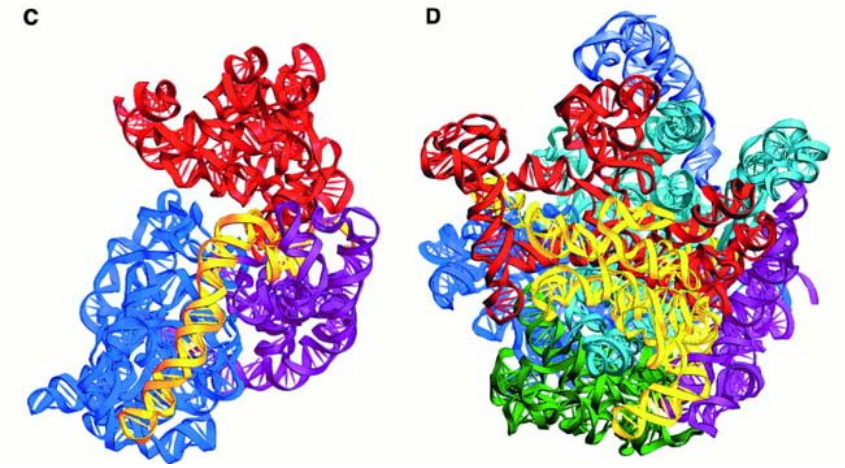


Figure 18**.** (pages 40-42) Secondary and tertiary structures of 16*S*, 23*S*, and 5*S* rRNAs.

(**A**) Secondary structure of *T. thermophilus* 16*S* rRNA, with its 5', central, 3'-major, and 3'-minor domains shaded in blue, magenta, red, and yellow, respectively.

(**B**) Secondary structures of *T. thermophilus* 23*S* and 5*S* rRNAs, indicating domains I (blue), II (cyan), III (green), IV (yellow), V (red), and VI (magenta) of 23*S* rRNA. The rRNAs are numbered according to the standard numbering of *E. coli* (Brosius et al., 1980).

(**C**) Three-dimensional fold of 16*S* rRNA in 70*S* ribosomes, with its domains colored as in (A).

(**D**) Three-dimensional folds of 23*S* and 5*S* rRNAs, with their domains colored as in (B).

The figure is taken from (Yusupov et al., 2001), and reproduced with permission from The American Association for the Advancement of Science, license number 2477450773485.

1.7 RNA pseudoknots

The pseudoknot is an RNA secondary structure motif, in which a part of the loop of the stem-loop structure participates in complementary base pairing elsewhere in the RNA chain (Figure 19). In chapter 3 of the thesis we will discuss a newly identified motif termed the "G-ribo", which stimulates formation of the pseudoknot structures in rRNA. Here will briefly discuss some biological roles of pseudoknots.

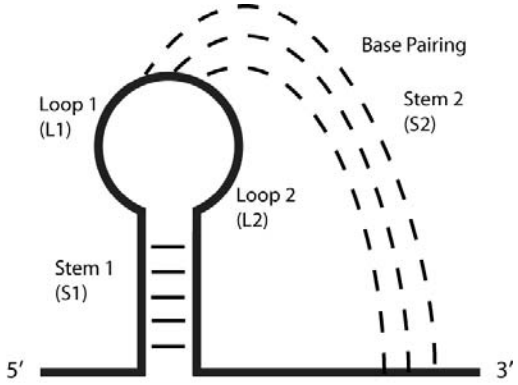


Figure 19. A simplified scheme representing a hairpin type (H-type) pseudoknot.

The pseudoknot structure contains two double stranded regions, S1 and S2, connected by single-stranded loops L1 and L2.

The first pseudoknot was discovered in the structure of RNA of the turnip yellow mosaic virus (Rietveld et al., 1982). It was shown that the 3' ends of some plant viruses can mimic some functions of tRNA while lacking the canonical clover-leaf structure. The short H-type pseudoknot at the 3'-end of the virus RNA mimics the acceptor arm of tRNA that can be aminoacylated with Val (Pinck et al., 1970). Later the same group of authors defined the general rules of the pseudoknot formation and provided evidence for the existence of pseudoknots in other RNAs – the central pseudoknot of 16S rRNA of *E. coli* and a pseudoknot from the group I intron (Pleij et al., 1985).

RNA pseudoknots play a variety of diverse biological roles. For example, the core of the hepatitis delta virus (HDV) ribozyme is made of two pseudoknots (Figure 20A) (Ferre-D'Amare et al., 1998). Five helical segments of the HDV ribozyme form two coaxial stacks: stems P2/P3 and P1/P1.1/P4. One pseudoknot is formed by stems P1/P2 while another by stems P3/P1.1. The catalytic cores of various ribozymes (Rastogi et al.,

1996; Ke et al., 2004) as well as self-splicing introns (Adams et al., 2004) also contain pseudoknots.

Telomerase is the ribonucleoprotein complex, which is responsible for the maintenance of the telomere ends of the chromosomes (McEachern et al., 2000). Human telomerase consists of 451 nucleotides of RNA, a reverse transcriptase and several proteins (Kelleher et al., 2002; Cohen et al., 2007). The 5' end of the telomerase RNA forms a pseudoknot structure, which constitues the core of the telomerase and is required for its activity (Figure 20B). Diseases as autosomal dyskeratosis congenita as well as aplastic anemia have been linked to mutations in the telomerase pseudoknot (Marciniak et al., 2000; Vulliamy et al., 2002).

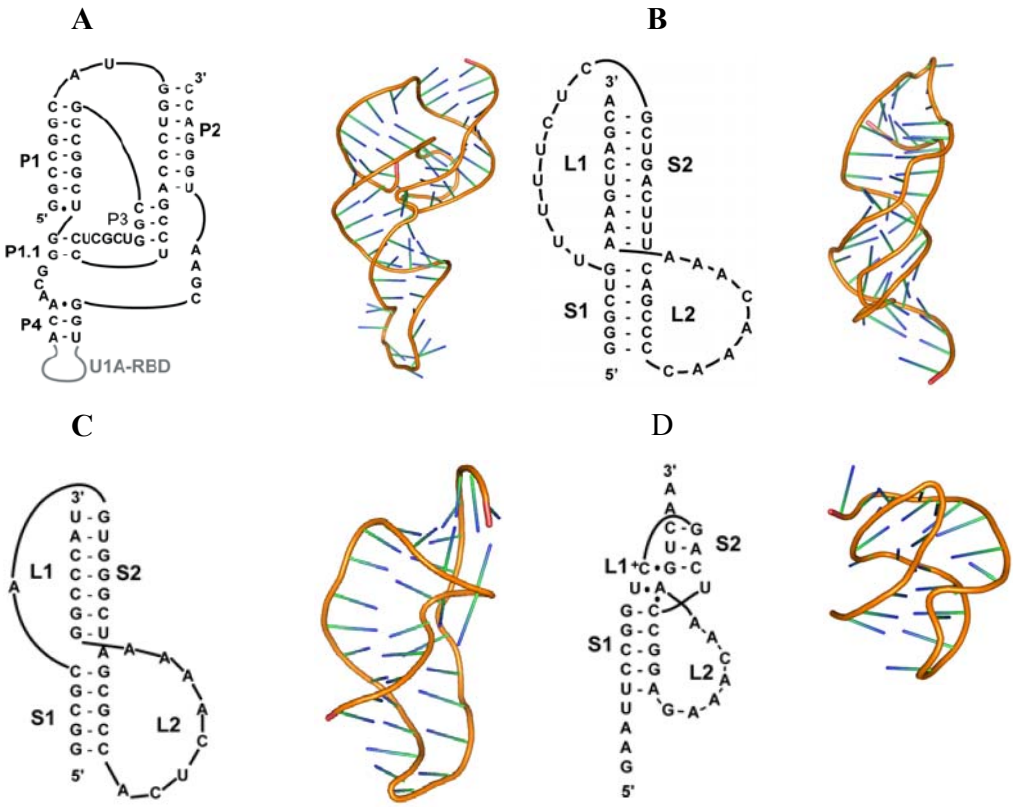


Figure 20. Secondary and tertiary structures of various pseudoknots.

A. Pseudoknot from the Precursor HDV ribozyme. Numbering of stems reflects standard nomenclature for HDV (PDB access code – 1SJ3). The solvent-exposed P4 stem was engineered to contain a high-affinity binding site for the small, basic RNA-binding domain of the U1A spliceosomal protein.

B. Human telomerase RNA pseudoknot (PDB access code – 2K95).

C. Mouse mammary tumor virus (MMTV) frameshift-inducing pseudoknot (PDB access code – 1RNK).

D. Pseudoknot from pea enation mosaic virus RNA1 (PEMV-1) (PDB access code – 1KPZ).

1.7.1 Role of pseudoknots in -1 frameshift

Normally, ribosomes translate mRNA without shifting the reading frame (Farabaugh & Bjork, 1999). However, in some situations the ribosome can shift from the correct reading frame by one nucleotide back or forward (Gesteland & Atkins, 1996). Replication and proliferation of many retroviruses depends on efficient -1 ribosomal frameshift. Thus, -1 frameshift is required for the regulation of a proportion of the products of the *gag* and *pol* genes or of the genes *gag*, *pro* and *pol* essential for viral development (Brierley, 1995). The -1 frameshift requires a slippery heptanucleotide sequence and a downstream RNA structure, which is usually a pseudoknot (ten Dam et al., 1990). The consensus slippery sequence consists of heptanucleotide 5′-X XXY YYZ-3′, in which X is any nucleotide, Y is either A or U, Z in eukaryotes is not G, and the spaces indicate the initial reading frame. There have been a number of different models for -1 frameshift (Jacks et al., 1988; Weiss et al., 1989; Plant et al., 2003). All models agree that a downstream secondary structure slows down the normal passage of mRNA through the ribosome, thus facilitating tRNA to slide one codon back on the slippery sequence, providing for the -1 frameshift. After the -1 frameshift occurs, the ribosome continues the translation in the −1 reading frame.

The first structure of the frameshift-inducing pseudoknot was obtained from the NMR studies of the mouse mammary tumor virus (MMTV) pseudoknot (Shen & Tinoco, 1995). In this structure two stems that make a pseudoknot are tilted by approximately 60° due to the presence of an unpaired adenine that intercalates between the helices and may act as a hinge (Figure 20C). The structural and functional studies of several variants of the MMTV pseudoknot showed that the intercalated nucleotide and the extent of the bend between two stems are required for an efficient frameshifting (Chen et al., 1996). The crystal structures of other frameshift-inducing pseudoknots from viruses have been solved. Thus, the structure of pseudoknots from the beet western yellow virus and pea enation mosaic virus, revealed compact H-type pseudoknots (Figure 19)with extensive loop-stem interactions (Figure 20D) (Egli et al., 2002; Nixon et al., 2002).

1.7.2 Pseudoknots in translation

Translation of most eukaryotic mRNAs starts when the 5'-end cap structure is bound by several initiation factors that start together with 40S subunit to scan along the mRNA until an AUG start codon is reached (Marintchev & Wagner, 2004). The 5' cap structure is also required for mRNA circularization, which is mediated by the interaction of the poly(A) tail-binding protein (PABP) with the initiation factor 4G (eIF4G), bound to the 5' cap via eIF4E (Wells et al., 1998). However, several animal and plant viruses as well as certain cellular mRNAs initiate protein translation from a so-called Internal Ribosome Entry Site (IRES), which allows them to bypass a cap-dependent initiation of translation (Dreher & Miller, 2006; Jang, 2006; Mokrejs et al., 2006).

The IRES is a highly structured region of mRNA, which is usually located at the 5' untranslated region (UTR) of the molecule and is often followed by the start codon. Many IRES structures contain pseudoknots, and one of the best-studied examples is Hepatitis C virus (HCV) IRES. The 40S subunit of mammalian ribosome can associate directly with HCV IRES without translation initiation factors (Fraser & Doudna, 2007). The HCV IRES contains two hairpins (domains 2 and 3) and an essential pseudoknot structure (Wang et al., 1995) (domain 3e/f) (Figure 21). Domains 2 and 3 are required for IRES function while the initial binding of the 40S subunit is mediated by the lower part of the domain 3 with a modest participation of the pseudoknot (Kieft et al., 2001; Otto & Puglisi, 2004). The exact role of the IRES pseudoknot is not yet clear, but it might be related to the ability of ribosomal protein S5 binding (Laletina et al., 2006). This protein is located close to the mRNA exit channel and also associates with domain 2 of IRES (Fukushi et al., 2001; Boehringer et al., 2005). It was proposed that a series of conformational changes of the four-way junction (domains 3a and 3c, Figure 21), is transmitted to the pseudoknot, which moves towards mRNA exit channel (Boehringer et al., 2005). The latter allows correct positioning of the AUG codon into the ribosomal P site and subsequent formation of the 80S ribosomes.

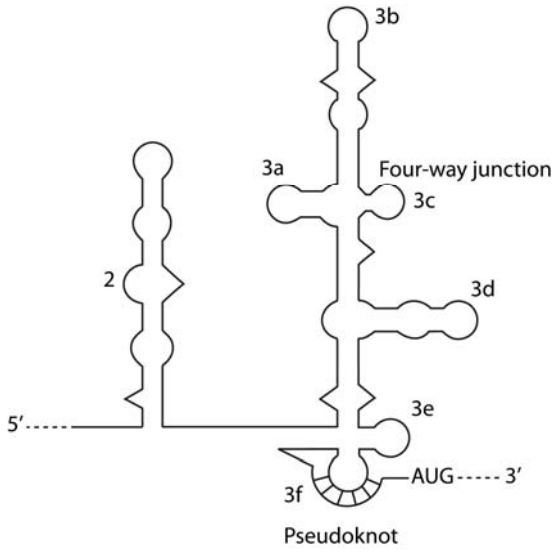


Figure 21. Secondary structure scheme of the hepatitis C virus internal ribosome entry site.
General representation of the principal secondary structure elements of the hepatitis C virus internal ribosome entry site: hairpin shaped domains 2, 3a-3e and pseudoknot domain 3f.

The RNA pseudoknot is a versatile structural motif, which plays various functionally important roles in different RNA molecules. The RNA pseudoknot represents a very rigid structural entity that can be used to direct the global folding of an RNA sequence. Due to its high rigidity the pseudoknot structures are often used as a stable element of the secondary structure, such as a stimulatory factor of the ribosome frameshifting and readthrough (ten Dam et al., 1990). Pseudoknots are important building blocks of large RNA molecules. Thus, both 16S rRNA and 23S rRNA together contain 10 pseudoknots, which mediate long-range interactions between distantly located parts of rRNA (Gutell et al., 1994). One of the ribosomal pseudoknots is located in the conserved region of loop 530, which plays a key role in the process of selecting cognate tRNA in the ribosomal A site (Powers & Noller, 1991; Ogle et al., 2001). We will come back to discuss several pseudoknots found in the rRNA later in the text of the thesis.

1.8 Principles of molecular dynamics

Molecular dynamics simulations are useful in studies of a variety of systems. Biological molecules, polymers or catalytic materials can be studied in a variety of states that include crystals, aqueous solution or in the gas phase (Brooks et al., 1988; Wang et al., 2001; Wang et al., 2008). Molecular dynamics simulations can be considered as a "virtual microscope" because it allows the hypothetical visualization of atomic motion of the molecule with temporal and spatial resolution. There is no physical method yet available which would allow access to all the time scales of motion with atomic resolution (Karplus & McCammon, 2002).

Molecular dynamics solves Newton's second law or the equation of motion for a system of N atoms interacting according to a potential energy force field, $F_i=m_ia_i(t)$, where F is the force, applied on the atom *i*, m_i is its mass and a_i is the acceleration of atom *I* (revieved in (McCammon & Harvey, 1987)). Knowing the force, applied on each atom, it is possible to determine its acceleration. The force acting on atom *i* can be computed from the gradient of the potential energy *V* with respect to the coordinates $\mathbf{r}_i$ of the atom. This is the force that is a result of interaction with all other atoms of the system.

$$F_i = -\nabla_i V(r_i)$$

Combination of two equations gives the following equation

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$

Here, *V* is the potential energy of the system. Subsequently, derivative of the potential energy relates to the changes in position as a function of time.

A simple application of the Newton's Second Law of motion gives

$$F = m \cdot a = m \cdot \frac{dv}{dt} = m \cdot \frac{d^2 x}{dt^2}$$

For the constant acceleration: $a = \frac{dv}{dt}$.

Integration of this equation gives an expression for the velocity: $v = at + v_0$

As long as $v = \frac{dx}{dt}$, integration gives an expression for the value of x: $x = vt + x_0$

As a result we can get relation between value of x at time t as a function of the acceleration, the initial position and the initial velocity, a, x_0, and v_0 respectively:

$$x = a \cdot t^2 + v_0 \cdot t + x_0$$

Expression of the acceleration can be obtained through the derivative of the potential energy with respect to the position r:

$$a = -\frac{1}{m}\frac{dE}{dr}$$

Thus, in order to get a trajectory for the group of atoms, one needs to know their initial positions and velocities while their acceleration is determined by the gradient of the potential energy function. The initial positions of the atoms are usually taken from the available X-ray or NMR structures of a molecule. The initial velocities are taken from a random distribution with the magnitudes conforming to the required temperature and corrected so there is no overall momentum, i.e.,

$$P = \sum_{i=1}^{N} m_i v_i = 0$$

The velocities, v_i, are often taken randomly from a Maxwell-Boltzmann or Gaussian distribution at a given temperature, which gives the probability that an atom i has a velocity v_x in the x direction at a temperature T.

$$p(v_{ix}) = \left(\frac{m_i}{2\pi\kappa_B T} \right)^{1/2} \exp\left[-\frac{1}{2} \frac{m_i v_{ix}^2}{\kappa_B T} \right]$$

Where κ_B is the Boltzmann constant.

The temperature is calculated from velocities with the relation

$$T = \frac{1}{(3N)} \sum_{i=1}^{N} \frac{|p_i|}{2m_i}$$

where N is the number of the atoms in the system.

Potential energy E is a sum of internal, or bonded potential energy, and external or non-bonded potential energy (MacKerell et al., 1995).

$$V(R) = E_{bonded} + E_{non-bonded}$$

The first term, E_{bonded}, describes the bonds, angles and rotations of the groups along the covalent bonds. The second term, $E_{non\text{-}bonded}$, describes the long-range electrostatic and van der Waals interactions between non-covalently bonded atoms.

$$E_{bonded} = E_{bond\text{-}stretch} + E_{bond\text{-}bend} + E_{rotate\text{-}along\text{-}bond}$$

The first term, $E_{bond\text{-}stretch}$, is a harmonic potential representing the interaction between two covalently connected atoms.

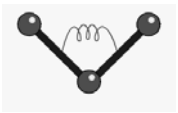
$$E_{bond-stretch} = \sum_{1,2\,pairs} K_b (b - b_0)^2$$

This is the approximation to the energy of a bond as a function of displacement from the ideal bond length, b_0. The strength of the bond is determined by the force constant, K_b. Both, b_0 and K_b are specific for each pair of bound atoms. Force constant values are often calculated based on the experimental data from such methods as infrared stretching frequencies or from quantum mechanics calculations. The bond length values

can be taken from high-resolution crystal structures or microwave spectroscopy data (reviewed in (MacKerell et al., 1998)).

The second term, $E_{bond\text{-}bend}$, is related to alteration of bond angles θ from ideal values θ_0, which is also approximated by a harmonic potential.

$$E_{bond-bend} = \sum_{angles} K_\theta (\theta - \theta_0)^2$$



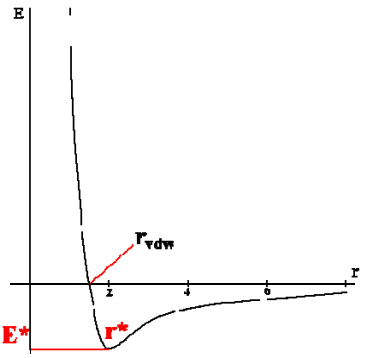
K_θ is the bending force constant that is determined empirically. θ is the actual bond angle in the molecule and θ_0 is the "natural" bond angle. Similarly, values of θ_0 and K_θ are specific for a particular combination of atoms constituting the angle θ.

The third term, $E_{rotate\text{-}along\text{-}bond}$, represents potential function of the torsion angle, which models the presence of steric barriers between atoms separated by 3 covalent bonds (1,4 pairs). Here, K_ϕ is the barrier to free rotation for the "natural" bond, n is the periodicity of the rotation (number of cycles in 360°; $n=1,2,3$), and ϕ is the torsion angle. K_ϕ can be obtained from studies of small model compounds and comparisons to the geometry and vibration spectra in the gas phase (IR and Raman spectroscopy), supplemented with *ab initio* quantum calculations (Barron et al., 1973; Barron & Buckingham, 1975).

$$E_{rotate-along-bond} = \sum_{\phi} K_\phi (1 + \cos(n\phi$$

$E_{non\text{-}bonded}=E_{electrostatic}+E_{van\text{-}der\text{-}Waals}$

The balance between repulsive and attractive forces of two interacting atoms gives rise to the van der Waals interactions. When two atoms come too close to each other, the repulsive force arises due to the electron-electron interactions. The fluctuation in the electron distribution on one atom induces an instantaneous dipole, which subsequently induces a dipole in a second atom, giving rise to an attractive interaction. Each of these two effects is equal to zero when atoms are infinitely separated and become significant as the distance decreases. The attractive interaction is longer range than the repulsion but as the distance become short, the repulsive interaction becomes dominant. This gives rise to a minimum in the energy. Positioning of the atoms at the optimal distances stabilizes the system. Both value of energy at the minimum E^* and the optimal separation of atoms r^* (which is roughly equal to the sum of van der Waals radii of the atoms) depend on chemical type of these atoms.



The Lennard-Jones 6-12 potential is most often used for modeling the van der Waals interactions.

$$E_{van-der-Waals} = \sum_{\substack{nonbonded \\ pairs}} \left(\frac{A_{ik}}{r_{ik}^{12}} - \frac{C_{ik}}{r_{ik}^{6}} \right)$$

In this equation the interaction energy is calculated using the atom-type dependent constants A and C. Values of A and C may be determined by a variety of methods, like non-bonding distances in crystals and gas-phase scattering measurements.

The van der Waals interactions play an important role for the stability of the biological macromolecules.

Coulomb potential is used for representation of the electrostatic interaction between a pair of atoms; D is the effective dielectric function for the medium and r_{ik} is the distance between two atoms with charges q_i and q_k.

$$E_{electrostatic} = \sum_{\substack{nonbonded \\ pairs}} \frac{q_i q_k}{D r_{ik}}$$

The empirical potential energy function is differentiable with respect to the atomic coordinates; this gives the value and the direction of the force acting on an atom and thus it can be used in a molecular dynamics simulation.

For more information and review of the principles of the molecular dynamics methods one can read the following articles and books (McCammon & Harvey, 1987; Ha et al., 1988).

1.9 Combinatorial approaches in biology

In the past 20 years combinatorial approaches have been intensively developed and applied in various biological fields. For instance, screening of combinatorial libraries of chemical compounds for biological activity is one of the ways of selection for new drug candidates (Carell et al., 1995; Schreiber, 2000). Search of therapeutic aptamers from the random oligonucleotide sequences proved to be a fruitful source of biologically active molecules (Keefe et al., 2011). Aptamers are single stranded oligonucleotides that form three-dimensional structures and can bind with high affinity and specificity to various molecular targets. Aptamers are discovered using systematic evolution of ligands by exponential enrichment (SELEX). This process was first described by Tuerk and Gold in 1990 (Tuerk & Gold, 1990). The principle stages of the SELEX process are demonstrated in Figure 22 and their brief description is given here. First, the pool of single-stranded oligonucleotides of random sequence is generated. The 3'- and 5'-terminal sequences that flank the region of 30-40 random nucleotides are fixed in order to

allow PCR amplification at a later step. The starting library containing approximately 10^{14} unique sequences is subjected to the binding assay with the target molecule. The mixture of oligonucleotides bound to the target molecule is separated from non-bound oligos with help of chromatography or other partitioning methods. The bound oligonucleotides are dissociated from a target and amplified, thus enriching the pool with sequences that bind to the target molecule. The selection process is repeated several times resulting in recovery of aptamers that have high affinity and specificity towards the target. The SELEX technique has been successfully used to select aptamers of extremely high affinity to a variety of ligands such as ATP (Dieckmann et al., 1996), adenosine (Huizenga & Szostak, 1995), theophylline (Zimmermann et al., 2000), prions (Mercey et al., 2006) and vascular endothelial growth factor (VEGF) (Ulrich et al., 2006). In fact, SELEX has been applied to numerous research programs in the molecular-biological fields and given rise to the application of combinatorial methods in many projects, including ours.

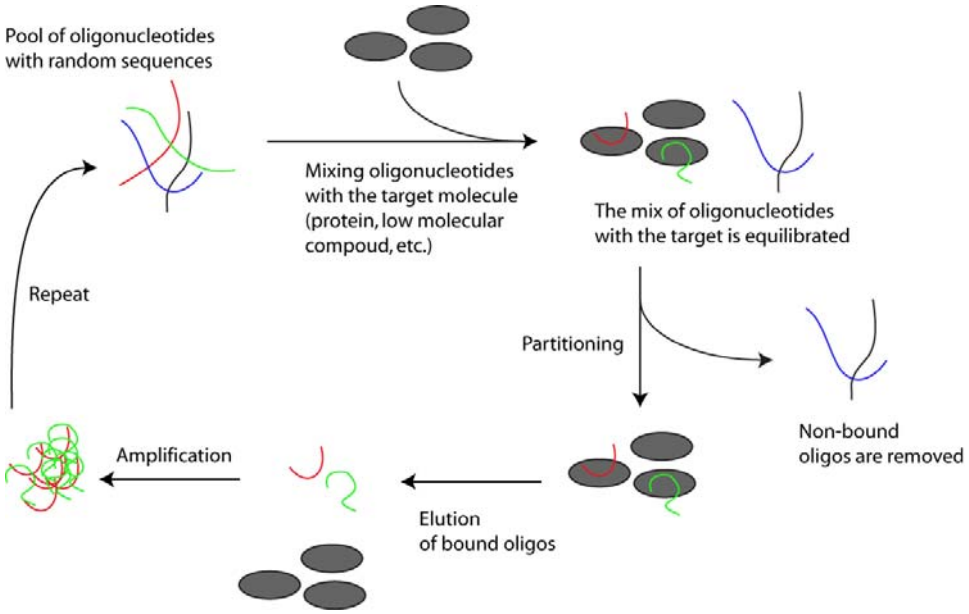


Figure 22. Principle steps of the aptamer selection.

Description of each step is given in the text.

1.10 The main outlines of the project

1.10.1 Identification of new motifs

Essential part of my PhD project was dedicated to identification and analysis of new RNA motifs. The search of a motif started from the visual observation of RNA structure. While looking at the structure, we were trying to identify new types of mutually juxtaposed nucleotides that have never been documented before. Identification of a new combination of two or more particularly juxtaposed nucleotides was followed by application of an automated search algorithm. The most common way of searching similarly juxtaposed groups of nucleotides is FR3D software developed by Leontis and co-workers (Sarver et al., 2008). This program can search for up to 20 particularly oriented nucleotides in the set of the available RNA structures. This is so-called geometric search, which requires a PDB identifier of the molecule that contains the group of nucleotides of interest (i.e. query motif) as well as the numbers of these nucleotides. The search algorithm guarantees identification of all candidates whose geometric discrepancy with the query motif is less than the selected cutoff. Usually we choose a combination of 2 to 10 nucleotides as a query motif. Including more nucleotides may result in identification of numerous false-positives, the fact that significantly complicates the search of new motif. Each identified candidate is analyzed individually within its structural context and those instances that share the highest number of common elements are then selected for the further analysis. Analysis of the selected instances may result in identification of some additional structural elements that were not included in the search, yet are associated with the original query motif. Refinement of the search parameters such as the level of discrepancy as well as varying the composition of the query motif allows identification of the limited number of arrangements with highly similar structural features. On most occasions, the set of arrangement, termed “motif” represents a group of relatively heterogeneous structures. Usually the motif’s core (represented by 3-10 nucleotides) is highly similar in all instances of the motif, while the rest of the motif’s structure can deviate within the group. A careful structural analysis is needed for identification of those elements that are critical for the motif formation and those that play a secondary role. For this reason the role of verbal description and accurate

categorization of the structural features of the motif plays extremely important role in definition of the term “motif” in each particular case.

1.10.2 In vivo studies of motifs

Apart from the theoretical characterization of the principles of recurrent motif formation, we conducted *in vivo* studies of the motifs. As the main experimental tool for the study the principles of motif structure formation in the ribosome, we used the instant evolution approach applied to the large and small subunit of the *E. coli* ribosome. The instant evolution approach engages simultaneous modification of a number of nucleotide positions combined with functional selection of those sequence variants that are able to support the functional state of the molecular complex. Introduction of the combinatorial libraries and subsequent *in vivo* selection of functional clones allows us to obtain sequence variants of ribosomes, which can lead to the understanding of the sequence constraints for the studied RNA region. Usually, analysis of the results from the selection leads to identification of unexpected patterns and/or co-variations in the structure of RNA, which can be difficult to understand on the spot. Our goal is to decipher the set of rules that govern the formation of the three-dimensional structure of the motif through analysis of the patterns of nucleotides, obtained with help of *in vivo* selection.

The key to our approach stems from one important assumption: the functionally selected variants have the same overall three-dimensional structure as in the wild-type, and that the local regions we modify are also, therefore, forming in a similar way to that of the wild-type. For example, we can study a hypothetical RNA motif which contains a dozen particularly juxtaposed nucleotides. We may be interested in modifying only a few of the centrally located nucleotide positions of this relatively large and complex arrangement. In order to facilitate our analysis we can postulate that the replacement of several central positions within the motif does not affect the over-all shape of the whole arrangement, if a functional clone is selected. Subsequently, we analyze the impact of the nucleotide replacements in the center of the motif based on the native crystal structure of the arrangement.

Such an approach is justified only if the number of modified positions does not overcome a certain reasonable number. For example, simultaneous modification of 10

nucleotides in a motif that is made of 15 nucleotides does not make any sense because the analysis of the group of sequence variants selected will be extremely difficult due to the high complexity of the library. On top of this, modification of numerous nucleotides at the same time would significantly increase the chances for the alteration of the native structure beyond recognition, which would render the structural interpretation meaningless.

1.10.2.1 In vivo systems for studies of rRNA mutagenesis

Currently, two major systems for *in vivo* studies on rRNA are exploited by different research groups. One of them is the so-called "knock-out" system and is based on an *E. coli* strain that is devoid of all seven ribosomal operons. The strain is kept alive with a multi-copy plasmid expressing the rRNA from a single ribosomal operon (Asai et al., 1999a). Thus, using this system, *Asai et al*. (Asai et al., 1999b) studied the effects of the complete replacement of the rRNA of *E. coli* with that of *Salmonella typhimurium* or *Proteus vulgaris.* Hence, the knock-out system can be used to study the global effects of the rRNA mutagenesis on the various aspects of *E. coli* functionality. The knock-out system allows introduction of mutations in both 30S and 50S ribosomal subunits. Although the knock-out system is an extremely powerful tool for the studies of rRNA mutagenesis, modifying the ribosome in the knock-out system can be sometimes difficult because rRNA modifications can significantly affect the synthesis of the cellular proteome.

Another largely used approach for rRNA mutagenesis *in vivo* is the so-called specialized ribosome system (Hui & de Boer, 1987; Lee et al., 1996; Abdi & Fredrick, 2005; Rackham & Chin, 2005; Gagnon et al., 2006). The specialized translation system capitalizes on the fact that for translation, mRNAs make a Watson-Crick mini helix between their Shine-Dargarno (SD) sequence with the anti-SD sequence of the ribosome. Thus, modification of the anti-SD sequence on the ribosome that is expressed from a plasmid, allows for the specific translation of mRNAs that are special in the fact that they have a SD sequence that is complementary to this modified anti-SD sequence. In this system, clones are selected by the ability to survive in the presence of chloramphenicol, which is the result of the proper synthesis of the protein chloramphenicol acetyl-

transferase (CAT) gene containing the modified SD sequence. The specialized ribosome system may allow getting more variation in the selected sequences, which otherwise would be lethal in the knock-out system, since modification to the specialized ribosome only affects the expression of one gene. Unfortunately, the specialized ribosome system is limited to the studies of the 30S ribosomal subunit only.

1.10.3 In silico modeling

In order to better characterize the newly identified motifs as well as for building structural models based on results obtained from *in vivo* selection we have used molecular dynamics (MD) simulations. Using MD simulations we were trying to understand the global influence of modification of certain nucleotide positions on the overall structure of the motifs under consideration. Although MD is a powerful tool for 3D structure analysis, one has to avoid the overinterpretation of the data obtained with this method. The data obtained from MD simulations should not be considered as an ultimate source of information for how a nucleotide behaves. Instead the data should be used as a complementary tool to other methods of characterization of a given structural arrangement.

Chapter 2

Article 1

G-ribo: A new structural motif in ribosomal RNA

2. G-ribo: A new structural motif in ribosomal RNA

Sergey V. Steinberg and Yury I. Boutorine

Département de Biochimie, Université de Montréal, Montréal, QC, Canada.

RNA, 2007, vol. 13, no. 4, pp. 549-54

Running title: G-ribo motif in ribosomal RNA

Contribution of each author:
Yury Boutorine: Searched the ribosomal subunit structures for the inter-helix contacts, 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 the inter-helix contacts, collected samples and participated in data analysis, preparation of manuscript and figures.

2.1 Abstract

Analysis of the available crystal structures of the ribosome and of its subunits has revealed a new RNA motif that we call G-ribo. The motif consists of two double helices positioned side-by-side and connected by an unpaired region. The juxtaposition of the two helices is kept by a complex system of tertiary interactions spread over several layers of stacked nucleotides. In the center of this arrangement, the ribose of a nucleotide from one helix is specifically packed with the ribose and the minor-groove edge of a guanosine from the other helix. In total, we found eight G-ribo motifs in both ribosomal subunits. The location of these motifs suggests that at least some of them play an important role in the formation of the ribosome structure and/or in its function.
Key words: loop 530 / ribosomal RNA / ribosome / RNA motif / RNA structure

2.2 Introduction

An essential part of our knowledge on RNA structure has accumulated in the form of recurrent structural motifs. Recurrent motifs appear in the same or different molecules and have the same or very similar conformation (for review, see (Batey et al. 1999; Moore 1999; Noller 2005). Here we present a new RNA structural motif named G-ribo, which has been found in eight different places within the ribosome. This motif represents a specific side-by-side arrangement of two double helices connected by an unpaired region. The juxtaposition of the helices is stabilized through a complex system of specific interactions, which spread over several layers of stacked nucleotides. The location of the identified cases of this motif within the ribosome suggests that at least some of them play an important role in the formation of the ribosome tertiary structure and/or in its function.

2.3 Definition of the G-ribo motif

The G-ribo motif refers to a particular side-by-side arrangement of two double helices connected by a short unpaired region. To describe this arrangement, we use the following language. The above-mentioned connector region is called the “internal

connector" (Figure 1A). The helices flanking this connector on the 5′ and 3′ sides are referred to as Helices 1 and 2, respectively. The unpaired regions on the 5′ side of Helix 1 and on the 3′ side of Helix 2 are called Connectors 1 and 2, respectively. In some cases, there is a third helix between Connectors 1 and 2. In other arrangements, such a helix does not exist, so that Connectors 1 and 2 become one. The four strands composing the two helices are named "P," "Q," "R," and "S." For the nucleotides (nt) in these strands that compose the top base pairs of both helices, the number zero (0) is assigned. For other layers within the helices, the numbering propagates in the negative direction. For the nucleotides of connector regions that stack on top of the nucleotides of the zero layer, the letters are taken from the corresponding 0 nt, while the numbers are positive.

The key element of the G-ribo motif pertains to a specific juxtaposition of top base-pairs [0P;0Q] and [0R;0S] of both helices, which form together a so-called zero layer (Figure 1B). Within this layer, the ribose of nucleotide 0R interacts with the minor-groove edge of nucleotide 0P. Nucleotide 0P should be G, which would allow it to provide two chemical groups O2′ and NH2 for formation of the hydrogen bonds with O2′ and O4′ of nucleotide 0R, respectively. The described interaction between G in one helix and the ribose of a nucleotide in the other helix has given the name "G-ribo" to the whole arrangement.

2.4 Identification of the G-ribo motif in the ribosome structure

Within the ribosome structure, we found eight cases of the G-ribo motif, three in 16S rRNA and five in 23S rRNA (Figure 2 and 3).). All these motifs are clearly seen in all available high-resolution crystal structures of the whole ribosome and of its subunits regardless of the organism (Ban et al. 2000; Schluenzen et al. 2000; Wimberly et al. 2000; Harms et al. 2001; Schuwirth et al. 2005; Korostelev et al. 2006; Selmer et al. 2006). At the same time, the detection of the motif was found to be sensitive to the resolution of the crystal structure because in the structures with a lower resolution (Schluenzen et al. 2000; Harms et al. 2001; Korostelev et al. 2006), almost none of the G-ribo motifs can be easily recognized. No example of the G-ribo motif was found in any RNA-containing entry of the PDB database not related to the ribosome. In all found cases

of the motif collected from all structures, the arrangement of two, nucleotides 0P and 0R, is virtually identical. Their superposition provides for a RMSD of 0.5 Å (Figure 2A), which also leads to a very similar juxtaposition of Helices 1 and 2. The angle between the axes of the helices is ~60°, which makes the arrangement of the two base pairs at the zero layer essentially nonplanar.

Figure 3 demonstrates the nucleotide sequences of the identified cases of the G-ribo motif in the *Escherichia coli* ribosome. The cases are named "S" or "L" depending on the subunit in which they are found, followed by the number of the nucleotide occupying position 0P in 16S or 23S rRNA. In all these cases, base-pair [0P;0Q] is GC. It is also GC in all G-ribo motifs identified in the other structures except motif L2323 from the *Thermus thermophilus* ribosome, where it is GU (data not shown; (Selmer et al. 2006)). As to base-pair [0R;0S], in all cases it is either Watson–Crick (WC) or GU. Neither the replacement of GC as base-pair [0P;0Q] by GU, nor the mentioned variations in base-pair [0R;0S] affect the ability of the two base pairs to form the arrangement seen in Figure 1B.

Additional analysis of the available nucleotide sequences of 16S and 23S rRNA (Wuyts et al. 2004) showed a rather high level of conservation of base-pairs [0P;0Q] and [0R;0S] in different G-ribo motifs (Supplemental Table 1). In most identified G-ribo motifs, the presence of GC in base-pair [0P;0Q] exceeds 96%, and the most frequent alternative to the GC base pair is GU. Together, combinations GC and GU account for >98% of the cases of base-pair [0P;0Q] in all motifs but two. The exceptions pertain to motif S861 in eubacteria and to motif L1024 in archeabacteria. In these cases, combinations GC and GU account together for 94.2% and 91.9% of all cases of base-pair [0P;0Q], respectively. A relatively low presence of GC and GU in these cases could be due to problems with the alignment of the corresponding regions among different sequences of rRNA. Base-pair [0R;0S] is also very conserved. For all identified G-ribo motifs, either combination WC or GU was found in this position in >98% of the nucleotide sequences. Due to such a high level of conservation, the correct juxtaposition of base-pairs [0P;0Q] and [0R;0S] is possible in the overwhelming majority of the cases in all identified G-ribo motifs both in eubacteria and in archeabacteria..

2.5 Tetranucleotide arrangement at the +1 layer

In addition to the arrangement at the zero layer, all G-ribo motifs have common elements at layers +1 and −1. The central role in the +1 arrangement is played by nucleotide +1R, which belongs to the internal connector and stacks on top of nucleotide 0R. Compared to the position that +1R would have occupied in a regular A-conformation of Helix 2, this nucleotide is reoriented and displaced for several angstroms farther from Helix 1, seemingly because of a potential collision with base-pair [0P;0Q] (Supplemental Figure 1). Such a displacement makes the distance between the C1′ atoms of +1R and 0R too long for a direct connection of the two nucleotides and necessitates the existence of at least one intervening nucleotide between them. In all identified cases of the G-ribo motif, nucleotide +1R is adenosine. Two more nucleotides of the +1 layer, +1P and +1S, are also characterized by predominant identities, although for each of them, there are exceptional G-ribo motifs in which these predominant identities are not conserved. Nucleotide +1P is adenosine in all motifs except L1024. Nucleotide +1S is uridine in all motifs except L1642.

The particular positions and identities of nucleotides +1R, +1P, and +1S allow them to form a specific arrangement seen in Figure 2C and in Supplemental Figure 2. Normally, these nucleotides form a base triple consisting of two base pairs, [+1P;+1R] and [+1R;+1S]. Most often, adenosines +1P and +1R form a head-to-head base pair using their WC edges. However, in different motifs, any of the two adenosines, or even both at the same time, can flip around their glycosidic bonds, providing the Hoogsteen edges for the interaction with the other adenosine. Such a movement does not change the number of hydrogen bonds between the two bases, although it requires some rearrangement in the backbone conformation for the adjustment to a shorter distance between the glycosidic bonds. For different orientations of +1R, its base pair with uridine +1S would be either reverse Hoogsteen or reverse WC. A conversion from one base pair to the other would also need an adjustment of the positions of the glycosidic bonds. The +1 arrangement can also include nucleotide +1Q, which exists only in five G-ribo motifs and does not have a distinct identity. When +1Q is a purine, it forms a sheared base pair with adenosine +1P. If it is a pyrimidine, it provides atom O2 for a hydrogen bond with the amino group of the

adenosine +1P. Regardless of the identity, +1Q always stays in about the same position with respect to adenosine +1P (Figure 2C).

Analysis of the available nucleotide sequences of ribosomal RNA shows that the +1 arrangement described above is highly conserved both in eubacteria and archeabacteria. Thus, the adenosine identity of position +1R is preserved in all motifs at the level of 98.5% or higher (Supplemental Table 1). Also, in all G-ribo motifs except L1024, adenosine in position +1P is conserved at the level of 96.6% or higher. In all motifs except L1642 and S1047 (only in archeabacteria), the uridine identity of +1S is conserved at the level of 94.9% or higher.

In the exceptional motif L1024, position +1P is always occupied by uridine, which is involved in interactions not observed in any other G-ribo motif (Supplemental Figure 2). However, nucleotides +1R and +1S stay at about the same places and form a base pair as in other G-ribo motifs. In motif L1642, the nucleotide in position +1S has neither a distinct identity nor a particular position within the 50S tertiary structure. At the same time, the other three nucleotides of the +1 layer—+1R, +1P, and +1Q—are positioned as in other G-ribo motifs. The same is probably true for motif S1047 in archeabacteria, in which position +1S does not have a distinct identity, while positions +1R and +1P are conserved as adenosines (Supplemental Table 1).

Despite some differences in the structure of the +1 layer, the conformation of the backbone and the positions of the glycosidic bonds of three nucleotides, +1R, +1P, and +1Q, in different G-ribo motifs are rather close (Figure 2C). The position of the fourth nucleotide, +1S, varies more widely, which indicates its sensitivity to the variations in the position of adenosine +1R. The similarities between the structures at the +1 layer in different G-ribo motifs are based on the high conservation of nucleotides +1R, +1P, and +1S, which allows us to qualify these structures as variations of the same type of arrangement.

2.6 A-minor interaction at the -1 layer of Helix 1

Although nucleotides +1R and 0R stack on each other, they, as mentioned above, cannot be directly connected, which requires the presence of at least 1 nt between them. The part of the polynucleotide chain enclosed between +1R and 0R forms a bulge, which in most cases consists of only 1 nt, but can also contain 2 nt, as happens in S521 and L1309 (Figure 3). For identification of nucleotides of this bulge, we use "T," and the bulge is referred to as the T-bulge (Figure 1A). The last nucleotide of the T-bulge, which at the same time is the last nucleotide of the internal connector, universally interacts with the −1 layer of Helix 1 and is thus identified as −1T.

The base of −1T interacts with the minor groove of base-pair [−1P;−1Q] (Figure 2B, Supplemental Figure 3, 4). In the known crystal structures, nucleotide −1T is almost exclusively adenosine, which allows it to make the A-minor interaction with base-pair [−1P;−1Q]. Also, there is always a hydrogen bond between O2′ of nucleotide −1Q and either N1 or N6 of adenosine −1T. Statistical analysis shows that for all motifs except S1047, both in eubacteria and archeabacteria, adenosine is the predominant identity of nucleotide −1T. However, depending on the motif, the level of conservation varies between 54% and 100% (Supplemental Table 1). Analysis of the structure of motif S1047 from *E. coli*, in which position −1T is occupied by cytidine, shows that even when −1T is not adenosine, its interaction with base-pair [−1P;−1Q] is similar to that of adenosine. In general, although position −1T displays a clear preference for adenosine, the integrity of the G-ribo motif does not seem to be critically dependent on the identity of −1T.

2.7 Participation of riboses in the stabilization of the core of the G-ribo motif

Within the structure of the G-ribo motif, there are areas in which interactions involving backbone and riboses become especially important. One such interaction between guanosine 0P and the ribose of 0R has defined the G-ribo motif. Another area in which riboses play an important role encompasses nucleotides 0Q, +1Q, +1R, and −1T (Figure 2B, Supplemental Figure 4). In all motifs containing nucleotide +1Q, the base of +1R stacks to the ribose of +1Q. Also, in all cases, the base of −1T stacks to the ribose of

0Q. Interestingly, in motif L2323 from the *T. thermophilus* ribosome (Selmer et al. 2006), base-pair [0P;0Q] is GU. As mentioned above, this is the only case of this kind for which the structure is known. Compared to its position in the GC base pair, nucleotide 0Q in this GU base pair is shifted for several angstroms in the direction of the major groove. This displacement, however, is accompanied by the corresponding movement of nucleotide −1T in the same direction, which preserves the interaction between −1T and 0Q (data not shown). The importance of this interaction can explain the limited variability of base-pair [0P;0Q], discussed above, in which GU is the only acceptable exception from the standard GC pattern. Any other nucleotide combination in position [0P;0Q] would break either the interaction between 0P and 0R or that between 0Q and −1T.

In the same area of contact between nucleotides 0Q, +1Q, +1R, and −1T, several polar groups, mostly the O2′ groups of the riboses, become close to each other and can form hydrogen bonds. Although these hydrogen bonds do not follow a common pattern, the fact that they exist in all cases of the G-ribo motif can indicate their importance for the stability of the arrangement. While stacked to the ribose of 0Q, the −1T base effectively screens these hydrogen bonds from the solution. Such a screening would protect the donors and acceptors of these bonds from interaction with the solvent, thus additionally contributing to the stability of the region.

2.8 The consensus pattern of the G-ribo motif

Based on the analysis of the tertiary structures of different G-ribo motifs and of the available nucleotide sequences of ribosomal RNA, we can suggest a consensus sequence pattern that fits almost all cases (Figure 4). This pattern includes two base pairs, four positions of individual nucleotides, and two connectors. It underlines the critical importance of the GC or GU identity for base-pair [0P;0Q] and of the WC or GU identity for base-pair [0R;0S]. Also, nucleotide +1R must be adenosine. Three more positions—+1P, +1S, and −1T—are also occupied by nucleotides of predominant identities, although in each of these positions exceptions are numerous. The preferred length of the internal connector is 4 nt, and, if the presence of the third helix is considered to be equivalent to

the presence of two additional nucleotides, the preferred length of the region connecting positions 0S and 0P is 5 nt. This consensus pattern can be used to search for new candidates for G-ribo motifs in RNAs with unknown tertiary structure.

2.9 Discussion

In this paper, we describe a new structural motif, called G-ribo, which has been found in eight places in the ribosomal RNA. The motif represents a particular juxtaposition of two double helices connected by an unpaired region. The G-ribo motif is characterized by a certain level of rigidity, as can be deduced from the superposability of all identified examples of the motif. The fixation of the juxtaposition of the two helices is achieved via formation of a complex network of contacts, which spread over three layers of stacked nucleotides. The central element of this system consists of two specifically arranged nucleotides, 0P and 0R, from the top base pairs of both helices. The ribose and the base of 0P interact with the ribose of 0R. 0P must be guanosine, while the identity of 0R may not be important. The particular juxtaposition of these nucleotides has been given the name “G-ribo” to the whole arrangement and has served for identification of all cases of the motif. It has been surprising, however, that the structures identified with use of a relatively simple definition of the motif demonstrate a more complex similarity that spreads over three layers of stacked nucleotides. The fact that all identified structures contain the same or very similar elements that have not been included in the definition of the motif signifies the importance of these elements for the motif's integrity and/or function.

The analysis of the identified cases of the G-ribo motif allows us to suggest a chain of cause–effect relationships between different elements of the structure, which can explain how the formation of the G-ribo juxtaposition of the two zero base pairs has determined the formation of other common interactions at the levels between −1 and +1. Everything seems to originate from the fact that the G-ribo juxtaposition of the base pairs at the zero layer does not allow Helix 2 to continue in the regular A-conformation to the +1 level without a collision with Helix 1 (Supplemental Figure 1). As a result, +1R becomes displaced to a very special position in which it still stacks to 0R but cannot be

directly connected to it. This particular position of +1R together with the conserved identities of +1R, +1P, and +1S help the four nucleotides of the +1 level to form a specific arrangement that would stabilize the given juxtaposition of the two helices. Then, the inability of +1R to be directly connected to 0R necessitates the appearance of at least one unpaired nucleotide, −1T, between 0R and +1R. Nucleotide −1T interacts simultaneously with the ribose of nucleotide 0Q and with base-pair [−1P;−1Q] of Helix 1, providing an additional stabilizing effect for the whole arrangement. To make these interactions possible, the ribose of 0Q should be positioned in a very particular way, which explains the almost universal GC identity of base-pair [0P;0Q]. The only acceptable exception from the standard GC pattern is GU, in which 0Q is still able to maintain the interaction with −1T.

All eight identified G-ribo motifs are highly conserved in the ribosomal RNA of eubacteria and archeabacteria, which indicates their importance for the ribosome structure and function. For some motifs, a particular functional role has already been known or can be suggested based on their position in the ribosome. For example, motif S521 caps the so-called loop 530. This loop is known to form a part of the decoding center, being involved in the accuracy control (O'Connor et al. 1997; Ogle et al. 2001). It also participates in the binding of the initiation factor IF1 (Moazed et al. 1995; Carter et al. 2001). Similarly, helix 34, which stays as Helix 1 in S1047, is known to form a part of the decoding center (O'Connor et al. 1997; Ogle et al. 2001). Therefore, the involvement of the G-ribo motif S1047 in the proper positioning of the proximal part of this helix could also be essential for the decoding process. Another motif, L2323, is involved in the formation of the central protuberance. Helix 84 of 23S rRNA, which stays as the third helix of L2323, interacts with protein L5, which, in turn, interacts with protein S13 from the small ribosomal subunit. Both L5 and S13 contact the P-site tRNA (Yusupov et al. 2001). They are also known to be important for the subunit association during the initiation of translation (Correll et al. 1999; Cukras and Green 2005) and for the ratchet-like movement of the ribosomal subunits during the translocation (Valle et al. 2003). Thus, motif L2323, due to its closeness to proteins L5 and S13, may be involved both in the subunit association and in the translocation.

It is probable that other G-ribo motifs also play particular, yet unknown, functional roles. The elucidation of these roles and the understanding of how the details of the structure of the G-ribo motif determine its functionality in each particular case will be a matter of further analysis.

2.10 Acknowledgements

The authors are grateful to Drs. Lea Brakier-Gingras and Pascal Legaut for important discussions. SVS acknowledges a grant from CIHR and fellowships from CIHR and FRSQ.

2.11 References

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

Batey, R.T., Rambo, R.P., and Doudna, J.A. 1999. Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed. Engl.* **38:** 2326-2343.

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

Correll, C.C., Wool, I.G., and 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.

Cukras, A.R. and Green, R. 2005. Multiple effects of S13 in modulating the strength of intersubunit interactions in the ribosome during translation. *J. Mol. Biol.* **349:** 47-59.

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

Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H.F. 2006. Crystal Structure of a 70S Ribosome-tRNA Complex Reveals Functional Interactions and Rearrangements. *Cell* **126:** 1065-1077.

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

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

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

O'Connor, M., Thomas, C.L., Zimmermann, R.A., and Dahlberg, A.E. 1997. Decoding fidelity at the ribosomal A and P sites: influence of mutations in three different regions of the decoding domain in 16S rRNA. *Nucleic Acids Res.* **25:** 1185-1193.

Ogle, J.M., Brodersen, D.E., Clemons, W.M., Jr., Tarry, M.J., Carter, A.P., and Ramakrishnan, V. 2001. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292:** 897-902.

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

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

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

Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., and Frank, J. 2003. Locking and unlocking of ribosomal motions. *Cell* **114:** 123-134.

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. Structure of the 30S ribosomal subunit. *Nature* **407:** 327-339.

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

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

2.12 Figures

Figure 1

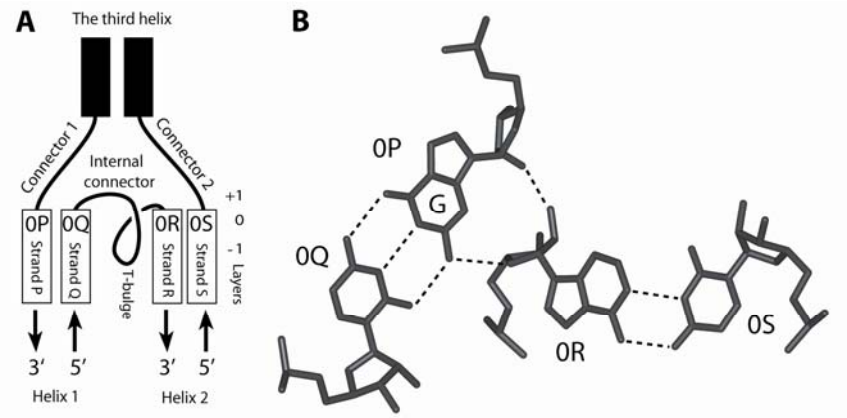


Figure 1. (*A*) The definition of different elements of the G-ribo motif. (Rectangles) Helical strands; (curves) unpaired regions. Helix 1 consists of strands P and Q, while Helix 2 consists of strands R and S. The third helix exists only, in some cases, of the G-ribo motif. In all cases of the motif, the 3′ part of the internal connector makes a loop that we call the T-bulge (see the text and Figure 3). The top base-pairs [0P;0Q] and [0R;0S] of Helices 1 and 2, respectively, form the zero layer. The positions of layers −1, 0, and +1 are shown on the *right*. (*B*) The juxtaposition of the two zero base pairs, [0P;0Q] and [0R;0S]. (Dashed lines) Hydrogen bonds within and between the base pairs. In this juxtaposition, the ribose of 0R interacts with the ribose and the base of 0P. To make this interaction possible, 0P should be guanosine.

Figure 2

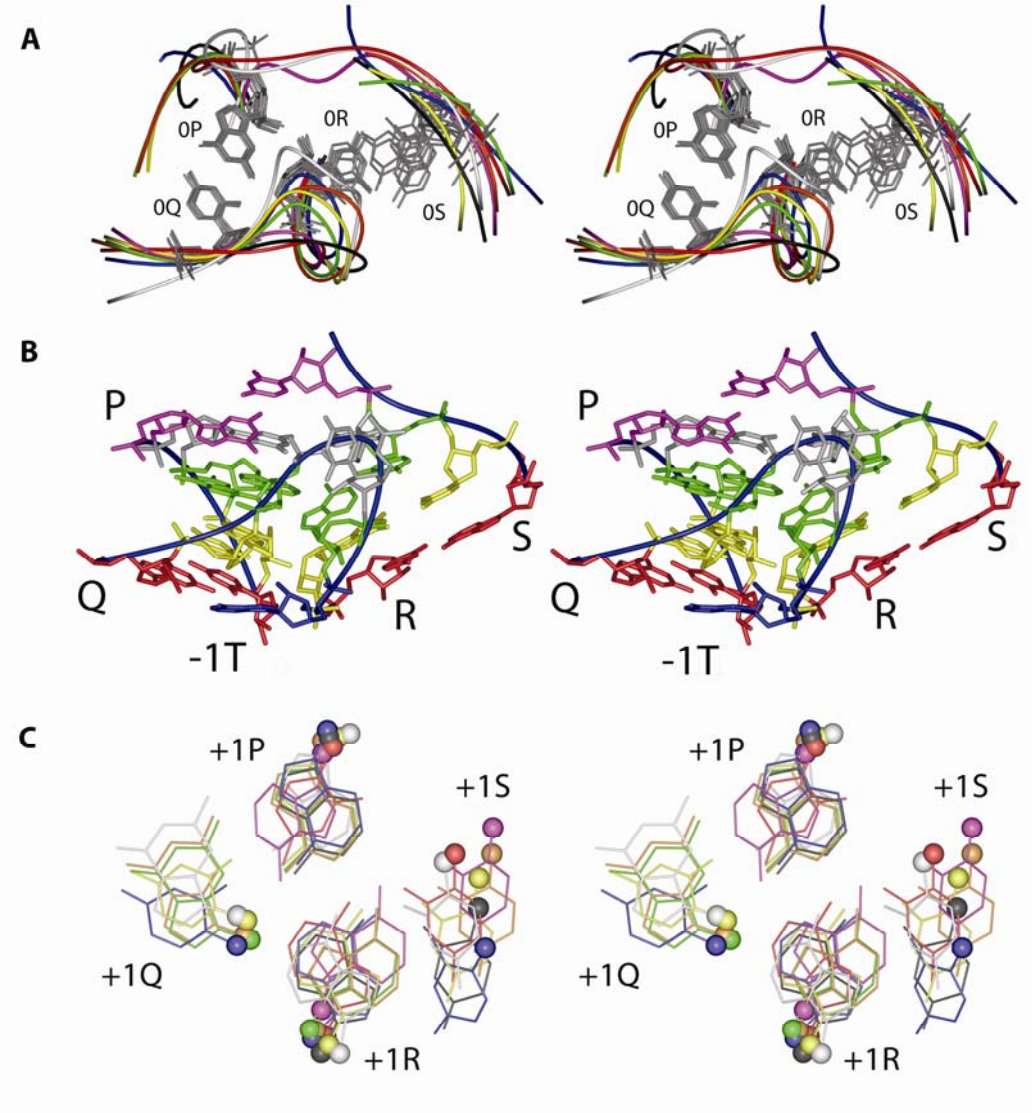


Figure 2. The tertiary structure of the G-ribo motifs in the *E. coli* ribosome (Ban et al. 2000; Schluenzen et al. 2000; Wimberly et al. 2000; Harms et al. 2001; Schuwirth et al. 2005; Korostelev et al. 2006; Selmer et al. 2006). (*A*) Superposition of all cases of the G-ribo motif. Helices 1 and 2 are on the *left* and *right*, respectively. For each case of the motif, the backbone is shown by the curve of a particular color. Only the nucleotides of the zero layer are shown explicitly. For three motifs S521 (white), L1024 (red), and L2383 (magenta), which do not have the third helix, the backbone continues from position 0S to 0P. For cases S861 (orange), A1047 (blue), L1309 (yellow), L1642 (green), and L2323 (black), which contain the third helix, the latter is not shown. The positions of 0P, 0Q, and 0R are well superimposed in all cases (for 0R, due to the variability of its identity, the superposition deals only with the backbone). The position of 0S is more flexible than those of 0P, 0Q, and 0R. (*B*) The tertiary structure of motif S861. The backbone (blue curve). The four strands P, Q, R, and S are indicated. Layers: −1 (red), 0 (yellow), and +1 (green).

Nucleotide −1T (blue). Unpaired nucleotides of the levels above +1 (gray). Noncanonical base-pair G858–U828 (magenta) on top of the arrangement forms the first base pair of the third helix. (*C*) The superposition of the arrangements in the +1 layer in different G-ribo motifs. For each motif, the same color is used as in *A*. The C1′ atom of each nucleotide is shown as a ball. Nucleotides +1S in L1642 and +1P in L1024, which do not follow the common pattern, are not shown (see the text). In motifs L1024, L2323, and L2383, nucleotide +1Q does not exist (see Figure 3). Despite the variations in the structure of different +1 arrangements, the positions of the glycosidic bonds of the equivalent nucleotides in different G-ribo motifs are rather close.

Figure 3

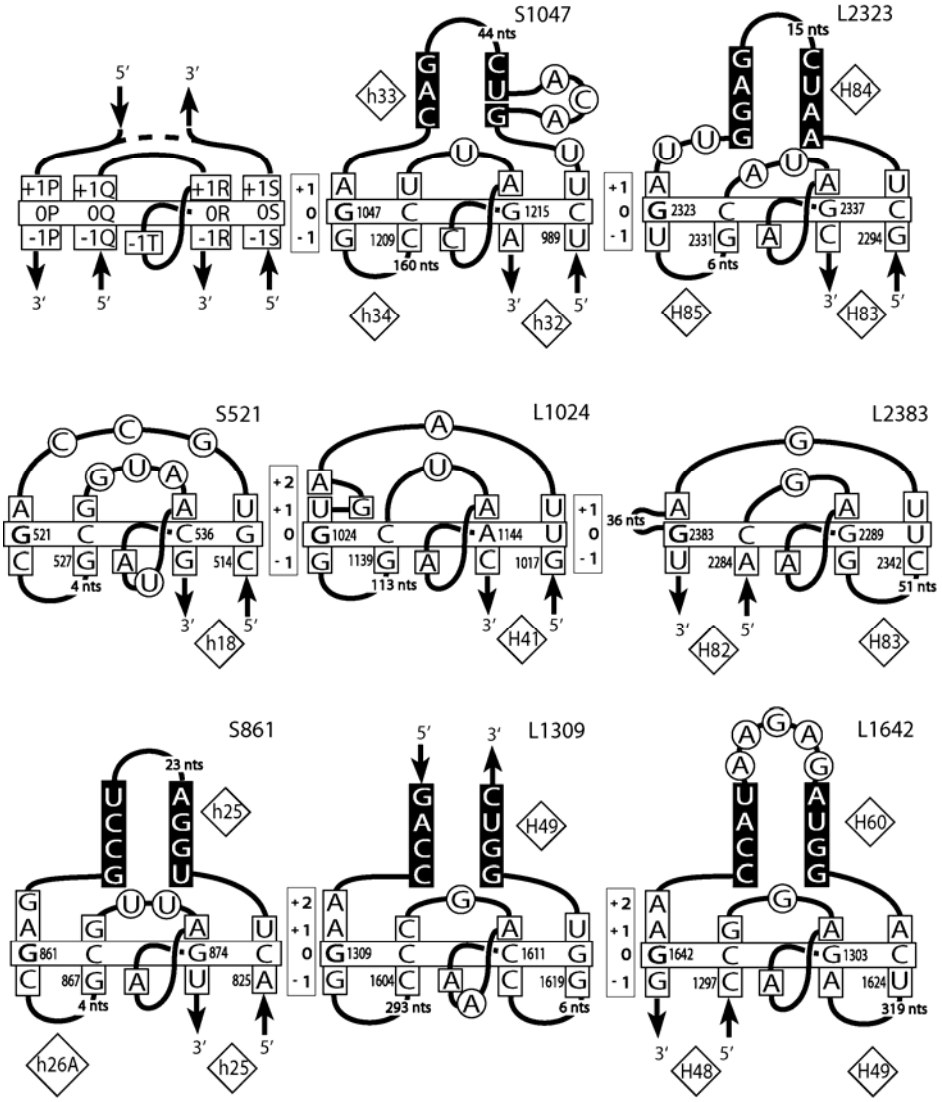


Figure 3. Secondary structures of the G-ribo motifs identified in the *E. coli* ribosome. (*Upper left* corner) A template with the named nucleotide positions of the layers from −1 to +1 is shown. (Dashed line) The possibility for a motif to have or not to have the third helix. In the name of a motif, "S" or "L" stands for the ribosomal subunit, small or large, in which the motif was found. The number in the name corresponds to that of nucleotide 0P in the standard *E. coli* numeration of rRNA. The double helices are arranged as in Figure 1A: the third helix is on *top*; Helices 1 and 2 are at *bottom left* and *bottom right*, respectively. (Vertically oriented rectangles) The positions of the layers. (Horizontally oriented rectangle) The base pairs at the zero layer. G in position 0P is bold. The nucleotides that stack to those of the zero layer are squared. The nucleotides of the third helix involved in base-pairing are shown on the black background. Other nucleotides are circled. The numbers of the helices in the standard 16S and 23S rRNA secondary structures are shown in diamonds.

Figure 4

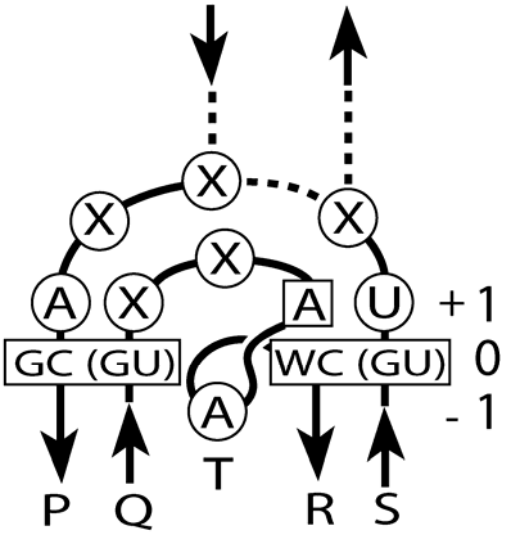


Figure 4. The consensus structure of the G-ribo motif. Helices 1 and 2, as well as connector regions, are arranged as in Figures 1A and 3. (Enclosed in rectangles) Base-pairs [0P;0Q] and [0R;0S] as well as nucleotide +1R, whose identities are very restricted in the known cases of the G-ribo motif. (In circles) The preferable identities for nucleotides +1P, +1S, and −1T. (X) An unrestrained nucleotide identity. The preferred length of the internal connector is 4 nt. (Dashed lines) The alternative possibilities for a G-ribo motif to form the third helix or not. The length of the region connecting positions 0S and 0P is calculated as the sum of the lengths of Connectors 1 and 2 plus 2 nt, if the third helix exists. The preferred length of this region is 5 nt.

2.13 Supplemental Table

Supplemental Table 1. The frequency of the occurrence of particular identities for base pairs and individual nucleotides in different G-ribo motifs

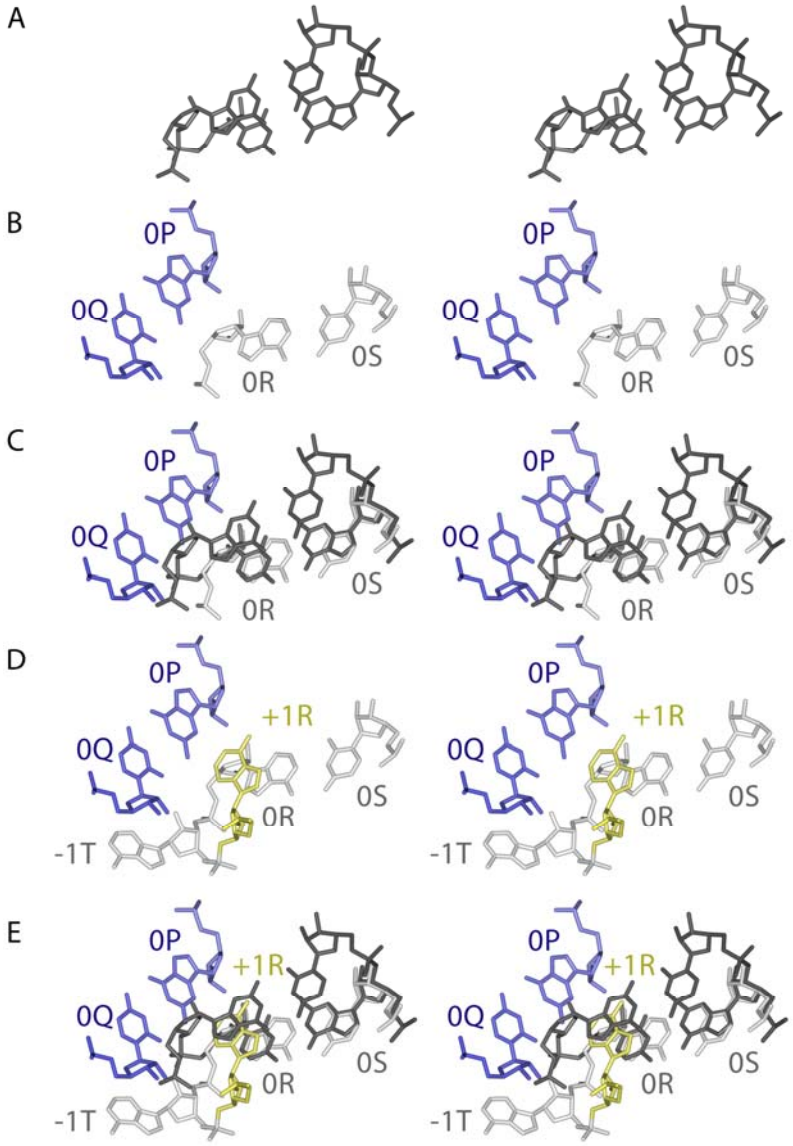
		L1642	L1309	L2383	L2323	L1024	S1047	S521	S861
[0P;0Q]	GC	98.0 (100)	96.5 (100)	99.8 (100)	86.4 (100)	98.5 (91.9)	93.6 (98.3)	99.3 (98.1)	88.7 (73.2)
	GU		3.5		13.4	0.3	5.9 (0.6)	0.1	5.5 (26.7)
[0R;0S]	WC	95.7 (100)	99.7 (100)	19.6 (78.4)	98.0 (100)	98.6 (100)	86.7 (88.9)	99.9 (99.3)	99.5 (99.5)
	GU	2.5		80.4 (21.6)	2.0	1.1	12.6 (10.0)		0.2 (0.2)
+1R	A	99.8 (100)	98.5 (100)	100 (100)	100 (100)	100 (100)	99.8 (99.6)	99.9 (99.8)	99.7 (99.2)
+1P	A	99.8 (100)	100 (100)	100 (100)	97.2 (100)		99.9 (96.6)	99.9 (100)	96.8 (99.8)
	U				1.0	100 (100)	0.1 (0.2)		0.2
+1S	U	5.5	99.8 (100)	100 (100)	100 (100)	99.0 (100)	99.7 (0.8)	99.9 (99.8)	94.9 (99.4)
	C	1.3 (84.9)				0.1	0.1 (78.6)	0.1	5.1 (0.2)
	A	51.5 (12.1)	0.2				0.1 (18.3)		(0.2)
	G	41.7 (3.0)					0.1 (2.2)	(0.2)	(0.2)
+1Q	U	0.5 (13.5)	1.5 (13.5)				99.4 (3.5)		2.5 (0.5)
	C	0.2	42.3 (13.5)				0.1	0.1 (0.2)	0.7
	A	0.2	6.1 (2.7)				0.4 (0.2)		59.3 (1.9)
	G	99.1 (86.5)	50.1 (70.3)				0.1 (96.3)	99.9 (99.8)	37.5 (97.6)
-1T	U	24.6 (8.1)	4.5 (18.9)	12.8 (8.1)	5.8		61.4 (72.2)	0.05	0.1
	C	(8.1)		1.2	(5.4)		38.4 (0.4)	0.05 (0.3)	
	A	75.4 (54.0)	95.5 (81.1)	86.0 (91.9)	94.2 (94.6)	100 (100)	0.1 (27.4)	99.8 (99.7)	99.8 (100)
	G	(29.8)					0.1	0.1	0.1

Supplemental Table 1. The frequency of the occurrence of particular identities for base pairs and individual nucleotides in different G-ribo motifs.

The numbers were calculated as a percentage of all analyzed nucleotide sequences of 16S rRNA (for motifs S521, S861 and S1047) or 23S rRNA (for motifs L1024, L1309, L1642, L2323 and L2383). The numbers without and in parentheses pertain to Eubacteria and Archaebacteria, respectively. The nucleotide sequences were taken from the European ribosomal RNA database (Wuyts *et al.*, 2004). The analyzed sequences were current as of May 2007.

2.14 Supplemental Figures

Supplemental Figure 1



Supplemental Figure 1. The special position of nucleotide +1R.

(**A**) The arrangement of two consecutive WC base pairs in the A-RNA conformation (piece A)

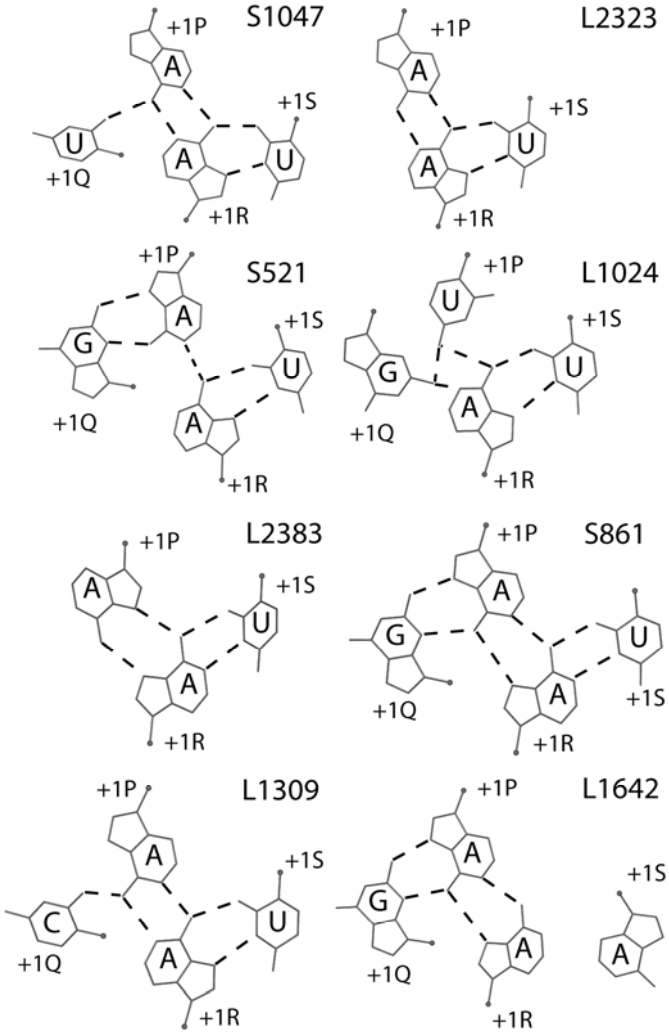
(**B**) The arrangement of the four nucleotide of the zero-layer in a G-ribo motif (piece B)

(**C**) Pieces A and B are superposed in the way that base pair [0R;0S] of piece B is superposed with the lower base pair of piece A. In this superposition, the nucleotide of piece A preceding the nucleotide superposed with nucleotide 0R of piece B collides with nucleotide 0Q. This *in silico* experiment demonstrates that the extension of strand R to layer +1 in the regular A-RNA conformation is impossible because of the collision with nucleotide 0Q of Helix 1.

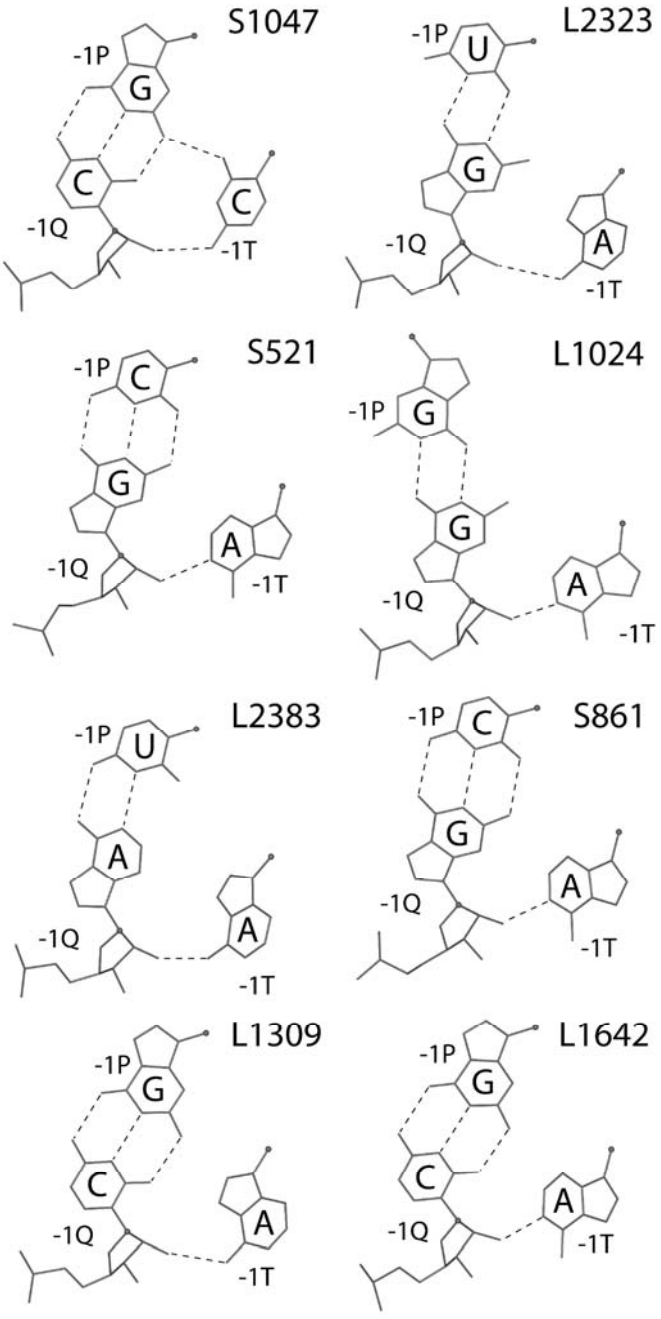
(**D**) The same arrangement as in (B) with nucleotides -1T and +1R added. There is no collision between +1R and Helix 1. In this position, +1R stacks to 0R. However, the distance between the C1´ atoms of +1R and 0R becomes too long to allow a direct connection of the two nucleotides, which necessitates the existence of at least one bulged nucleotide between them (see the text).

(**E**) The same superposition as in (C) with nucleotides -1T and +1R added. Compared to the position that +1R would have occupied in a regular A-conformation of Helix 2, this nucleotide is reoriented and displaced for several angstroms farther from Helix 1.

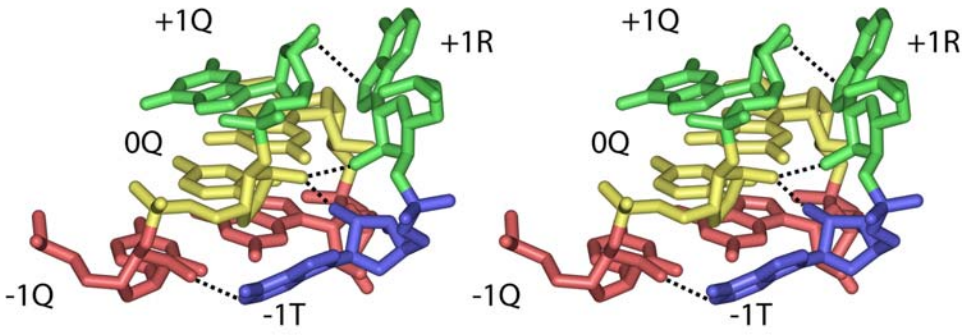
Supplemental Figure 2



Supplemental Figure 2. The tetranucleotide arrangements at the +1 layer in different cases of the G-ribo motif from the *E.coli* ribosome (Ban et al. 2000; Schluenzen et al. 2000; Wimberly et al. 2000; Harms et al. 2001; Schuwirth et al. 2005; Korostelev et al. 2006; Selmer et al. 2006). Dashed lines stand for hydrogen bonds.



Supplemental Figure 3. Interaction of nucleotide -1T with base pair [-1P;-1Q] in different cases of the G-ribo motif from the *E.coli* ribosome (Ban et al. 2000; Schluenzen et al. 2000; Wimberly et al. 2000; Harms et al. 2001; Schuwirth et al. 2005; Korostelev et al. 2006; Selmer et al. 2006). Dashed lines stand for hydrogen bonds.



Supplemental Figure 4. The structure of the region of motif L1642 encompassing nucleotides -1Q, 0Q, +1Q, +1R and -1T. The nucleotides are colored as in Figure 2B. The base of -1T forms the A-minor interaction with base pair [-1P;-1Q]. Also, the bases of -1T and +1R stack to the riboses of 0Q and +Q1, respectively. In addition, the hydrogen bonds between the ribose of +1Q and the base of +1R, between the riboses of +1R, 0Q and -1T, as well as between the ribose of -1Q and the base of -1T are shown as dashed lines. These hydrogen bonds are not universal and can be replaced by other hydrogen bonds in other G-ribo motifs or in another structure of the same motif.

Chapter 3

Article 2

G-ribo motif favors the formation of pseudoknots in ribosomal RNA

3. G-ribo motif favors the formation of pseudoknots in ribosomal RNA

Sergey V. Steinberg and Yury I. Boutorine

Département de Biochimie, Université de Montréal, Montréal, QC, Canada.

RNA, 2007, vol. 13, no. 7, pp. 1036-42

Running title: Pseudoknots in Ribosomal RNA

Contribution of each author:
Yury Boutorine: participated in data analysis, preparation of manuscript and figures.
Sergey V. Steinberg: participated in data analysis, preparation of manuscript and figures.

3.1 Abstract

Analysis of the pseudoknots existing in the ribosomal RNA showed that four of them are formed with the help of G-ribo, a recently identified RNA recurrent motif. The analysis of these pseudoknots revealed two major aspects in the G-ribo motif structure, which together provide the structural context favoring the formation of two different types of pseudoknots. The first aspect pertains to a particular side-by-side juxtaposition of two double helices that facilitates switches of the polynucleotide chain between different strands. The second aspect deals with the presence of an adenosine at a specific place where it can stabilize a particular arrangement of two quasicoaxial helices required for the pseudoknot formation. Additional analysis shows that the latter aspect is also present in other pseudoknots not related to the G-ribo motif or the ribosome, and thus represents a general structural element favoring the formation of pseudoknots.

3.2 Introduction

RNA pseudoknots can be defined as secondary structure arrangements in which a region within a hairpin loop forms a double helix with a region external to the hairpin proper (Gutell et al. 1994) (Supplemental Fig. 1). Pseudoknots are found in different RNA molecules and play a variety of different roles, which include the formation of the catalytic core of various ribozymes (Rastogi et al. 1996; Ke et al. 2004), self-splicing introns (Adams et al. 2004), and telomerase (Theimer et al. 2005). Pseudoknots also play critical roles in altering gene expression by inducing ribosomal frameshift and readthrough in different viruses (ten Dam et al. 1990; Shen and Tinoco 1995; Michiels et al. 2001; Egli et al. 2002; Nixon et al. 2002). Several pseudoknots are found in ribosomal RNA (Gutell et al. 1994), where they play potentially important structural and functional roles. The conditions for the formation of pseudoknots include the complementarity between the regions that are presumed to form the double helical stems as well as the proper lengths of the connector regions (Pleij et al. 1985). If the connectors are involved in specific interactions, their nucleotide sequences could also be important. However, to

which extent a broader structural context can affect the formation of a particular type of pseudoknot is mainly unknown.

In this paper, we show that G-ribo, a recently described RNA structural motif that pertains to a particular side-by-side arrangement of two double helices (Steinberg and Boutorine 2007), mediates the formation of four different pseudoknots in the ribosomal RNA. Analysis of these cases reveals two aspects of the G-ribo motif that together determine the structural context favoring the pseudoknot formation. Both aspects facilitate the switches of the polynucleotide chain between different strands at the core of the pseudoknot structure. One aspect pertains to the specific juxtaposition of two double helices within the G-ribo motif, while the other one deals with the presence of an adenosine that can stabilize the particular arrangement of two quasicoaxial double helices. This adenosine stacks on top of one helix and forms an A-minor interaction with the last base pair of the other helix. Because almost identical arrangement is found in other pseudoknots not related to the G-ribo motif, its presence is thus considered an important factor promoting the formation of different pseudoknots.

3.3 Background: the G-ribo motif

The G-ribo motif represents a particular side-by-side arrangement of two double helices (Helices 1 and 2) connected by an unpaired region with at least three nucleotides (Steinberg and Boutorine 2007). The definitions of the helices, strands, layers, and particular nucleotides within the G-ribo motif are given in Fig. 1A. At the zero layer, the top base pairs [0P;0Q] and [0R;0S] of both helices are juxtaposed as shown in Fig. 1B. At the center of this juxtaposition, guanosine 0P forms two hydrogen bonds with the ribose of nucleotide 0R. This interaction has given the name G-ribo to the whole arrangement. Other elements shared by all G-ribo motifs are the bulge between positions 0R and + 1R (T-bulge) and the nucleotide arrangements at layers + 1 (not shown) and − 1 (Fig. 1C). Together, these elements are able to fix the particular juxtaposition of Helices 1 and 2. Within the ribosome, the G-ribo motif has been found in eight different places: three in 16S rRNA and five in 23S rRNA..

One of the elements of the G-ribo motif that play an important role in the formation of the structures discussed in this paper is nucleotide − 1T, which is the last and, in most cases, the only nucleotide of the T-bulge (Fig. 1A). Nucleotide − 1T is predominantly adenosine, and it forms an A-minor interaction (Doherty et al. 2001; Nissen et al. 2001) with base pair [− 1P;− 1Q] (Fig. 1A). One side of the − 1T base stacks to the ribose of 0Q, while the other side is open for interaction with a nucleotide from layer − 2. As we argue here, such a special position of nucleotide − 1T is responsible for the ability of the G-ribo motif to promote the formation of pseudoknots.

3.4 Chain break in strand P

As mentioned above, the interactions that keep the particular juxtaposition of Helices 1 and 2 within the G-ribo motif are spread over three layers between + 1 and − 1. Below layer − 1, the helices become completely separated, which presumes that no elements specific to the G-ribo motif exist in this zone. However, at layer − 2, most G-ribo motifs share a new unusual element. In particular, we noticed that in five motifs there is a break in strand P between positions − 1P and − 2P, so that the two positions become separated in the polynucleotide chain by at least a dozen nucleotides, and in some cases, by hundreds of nucleotides (Fig. 2; Supplemental Fig. 2). Unlike strand P, strand Q in all G-ribo motifs continues between layers − 1 and − 2 without interruption. The presence of a noninterrupted Q strand allows Helix 1 to continue below layer − 1, while the break of the chain in strand P provides new opportunities for connections between different strands of the G-ribo motif. Further analysis showed that in all five cases of the G-ribo motif where positions − 1P and − 2P are distant from each other in the polynucleotide chain, − 1P is connected to a nucleotide of a lower layer of strand Q, while − 2P is connected to a nucleotide of either strand Q or S. Each of the two types of the − 2P connection corresponds to a particular type of pseudoknot.

3.5 Wrench pseudoknot in L1024

We will first consider motif L1024, whose secondary structure is shown in Figure 2B. Also, for convenience, the representation of the same motif as it is presented in the standard secondary structure of 23S rRNA can be seen in Supplemental Figure 2A. In L1024, nucleotide − 1P (G1025) is connected to − 4Q (G1136) with the help of a long connector region. Most of this region has no contacts with other parts of the L1024 G-ribo motif and is irrelevant for its structure, while for a proper connection between − 1P and − 4Q, one nucleotide is sufficient. Also, in L1024, − 2P (C1005) is directly connected to − 7S (U1004), while − 4P (C1007) is connected to − 6S (G1011) with the help of a 3-nucleotide (nt) connector. In the secondary structure of this arrangement, region 1005–1010, encompassing 3 nt of strand P followed by three unpaired adenosines, is a bulge of strand S between positions − 6S and − 7S. The formation of the three base pairs between nucleotides 1005–1007 and 1136–1138 creates a pseudoknot. For this reason, region 1005–1010 bulging of strand S is henceforth called the pseudoknot bulge. Due to the visible similarity to a wrench, the whole arrangement encompassing Helices 1 and 2 and the connector regions between them excluding the long connector between − 1P and − 4Q will be called the G-ribo wrench.

Analysis of the L1024 G-ribo wrench revealed additional details of its tertiary structure relevant for the pseudoknot formation. Thus, compared to the standard A-RNA conformation, base pair [− 2P; − 2Q] is substantially displaced with respect to base pair [− 1P; − 1Q] (Fig. 3A,B). The displacement can be represented as a combination of a rotation around atom O3′ of nucleotide − 2Q for about 50° and of an additional translation for about 4 Å. For − 2Q, the total displacement is about 5 Å, which allows this nucleotide to maintain its stacking to − 1Q. Nucleotide − 2P, however, has moved for > 10 Å and has lost all its contacts with − 1P. Such a movement of base pair [− 2P; − 2Q] brings it closer to Helix 2, thus allowing the two consecutive nucleotides − 7S (U1004) and − 2P (C1005) to be simultaneously involved in base pairing within the two different Helices 1 and 2. Needless to say, this movement has become possible due to the absence of the covalent link between − 1P and − 2P.

Additional interactions within the G-ribo wrench can stabilize this arrangement. First, an unpaired uridine − 5Sa (U1012), which is bulged of strand S between nucleotides − 6S and − 5S (Fig. 2A,B), forms a Hoogsteen base pair with adenosine − 1T. Within this base pair, − 1T and − 5Sa tightly bind nucleotide − 2P: while − 1T stacks to the base of − 2P, − 5Sa interacts with its ribose (Fig. 3B). This interaction effectively stabilizes the displaced position of nucleotide − 2P. Also, the last two adenosines A1009–A1010 of the pseudoknot bulge are involved in the A-minor interactions (Doherty et al. 2001; Nissen et al. 2001) with base pairs C1153–G1002 and C1152–G1003 occupying, respectively, layers − 9 and − 8 of Helix 2 (Fig. 2B).

On the level of the tertiary structure, the L1024 G-ribo wrench represents a compact globular domain with dimensions 26 Å × 34 Å × 38 Å attached to the rest of Helix 2 (Supplemental Fig. 3A). The compactness of the G-ribo wrench and its saturation with secondary and tertiary interactions suggests that it is stable on its own and can form independently of the rest of the ribosome structure.

3.6 Wrench pseudoknot in S861

Another wrench pseudoknot with a very similar structure is associated with motif S861 (Fig. 2C; Supplemental Fig. 2B). The most essential difference of S861 from L1024 pertains to the length of the pseudoknot bulge, which contains 5 nt instead of six. The superposition of the S861 structure with that of L1024 identifies the missing nucleotide as − 4P, which leaves the duplex between strand Q and the pseudoknot bulge with only two base pairs. Other differences of S861 from L1024 deal with the absence of the long insertion between positions − 1P and − 3Q and with the presence of long insertions between + 1S and + 1P and between − 6S and − 5S. None of these insertions is relevant to the structure of the S861 pseudoknot. Nucleotide U820 in S861, the last one in the insertion between − 6S and − 5S, is equivalent to − 5Sa in L1024: it occupies the same position just before − 5S and has the same uridine identity. It also forms a Hoogsteen base pair with adenosine − 1T and stacks to the ribose of − 2P (Fig. 3C).

The superposition of the structures of the L1024 and S861 pseudoknots demonstrates a very strong similarity over 11 layers from + 1 to − 9, i.e., until the end of Helix 2 in both motifs (Supplemental Fig. 3B).

.

3.7 Wrench pseudoknot in S521

Within the ribosome structure, there is one more wrench pseudoknot that associates with motif S521 (Fig. 2D; Supplemental Fig. 2C). In this pseudoknot, like in L1024, the pseudoknot bulge contains 6 nt and forms three base pairs with strand Q. Unlike L1024 and S861, S521 does not contain any long insertion and is thus built of only one piece of 16S rRNA between nucleotides 502 and 543.

Some aspects of the structure of the S521 pseudoknot make it different from those of L1024 and S861. In particular, the pseudoknot bulge in S521 is inserted between positions − 4S and − 5S of the S strand, rather than between − 6S and − 7S. Also, the last two adenosines of the pseudoknot bulge form the A-minor interactions with base pairs U543–A502 and G542–C503, which occupy, respectively, layers − 7 and − 6 of Helix 2, and not − 9 and − 8. In other words, in S521, compared to L1024 and S861, the position of the pseudoknot bulge in strand S as well as of the A-minor interactions between the pseudoknot bulge and Helix 2 is shifted for two layers up. Analysis shows that due to this shift, Helices 1 and 2 in S521 become displaced with respect to each other for a few angstroms compared to their juxtaposition in L1024 and S861 (not shown). As a result, pseudoknot S521 is superposable with neither L1024 nor S861, although it is still very close to both these structures.

Another difference between the S521 pseudoknot and those of L1024 and S861, which seems to be linked to the displacement of the pseudoknot bulge, consists in the absence of the bulged nucleotide − 5Sa. Indeed, due to the new position of the pseudoknot bulge, its interaction with strand Q would have interfered with the formation of base pair [− 5Sa; − 1T]. Interestingly, the S521 pseudoknot contains another unusual element, a second nucleotide in the T-bulge immediately before − 1T (nucleotide − 1Ta [U534]; Fig. 2D), the presence of which could compensate for the absence of uridine −

5Sa. Like − 5Sa, nucleotide − 1Ta stacks to the ribose of − 2P, and thus would stabilize the position of the latter in the situation when − 5Sa does not exist due to the displacement of the pseudoknot bulge (Fig. 3D).

Analysis of the available nucleotide sequences of 16S rRNA (Wuyts et al. 2004) showed that in archaea, unlike in bacteria, the pseudoknot bulge of S521 is integrated into the S-stem between positions − 3S and − 4S and contains either 7 or 8 nt (Supplemental Fig. 4). However, our preliminary data show that such modifications are local and do not affect the global pseudoknot conformation.

3.8 Ring pseudoknot

In the ribosomal RNA, there is another G-ribo-based pseudoknot whose structure, however, is essentially different from that of the wrench pseudoknots. This new pseudoknot is built of fragment 2283–2389 of 23S rRNA and is based on a specific arrangement of three consecutive double helices 82, 83, and 85 of the 23S rRNA secondary structure. The juxtaposition of helices 82 and 83 and of helices 83 and 85 is mediated by G-ribo motifs L2383 and L2323, respectively (Fig. 2E; Supplemental Fig. 2D). The whole arrangement is stabilized by the kissing interaction between the unpaired region adjacent to Helix 82 and the loop-closing Helix 85. This kissing interaction constitutes a pseudoknot. The four double helical regions, i.e., Helices 82, 83, 85, and the kissing helix, together form a circular almost ideally symmetric structure, which we call the G-ribo ring. In three dimensions, this structure looks like a compact disk of ~ 40 Å in diameter and 20 Å in width (Supplemental Fig. 3C).

In the secondary structure of the G-ribo ring, the P and R stems of L2323 become, respectively, the Q and S stems of L2383 and vice versa. In both motifs, as in those forming the wrench pseudoknots, the P strand is interrupted between positions − 1P and − 2P. However, unlike in G-ribo wrenches, in both motifs L2323 and L2383, nucleotides − 1P and − 2P are directly connected to nucleotides − 8Q and − 9Q, respectively. Due to the symmetry of the arrangement, positions − 1P and − 2P in one motif become, respectively, positions − 9Q and − 8Q in the other motif. Correspondingly, the connection between − 1P and − 8Q in L2323 becomes the connection between − 2P and

− 9Q in L2383 and vice versa. For the kissing double helix, each G-ribo motif donates its Q strand, which would form a double helix with the P strand provided by the other G-ribo motif. If one determines the length of each helix as the number of layers from the zero layer in one G-ribo motif to the zero layer in the other motif, Helices 1 and 2 contain 11 and 7 layers, respectively.

The ability of the G-ribo ring to form the kissing helix depends on the complementarity of the kissing regions and on their proper juxtaposition. The latter, in turn, depends on the structure of all elements between the kissing regions all along the G-ribo ring, which would together guarantee that these regions become proximal to each other in the orientation prone for the formation of a double helix. Analysis shows that within the G-ribo ring, there are four elements whose structure has been tuned in order to reach the proper juxtaposition of the kissing regions. Because both motifs L2323 and L2383 are rigidly attached to Helix 2, their juxtaposition strongly depends on its length. A deletion or an insertion of a base pair in Helix 2 would displace one kissing region with respect to the other for ~ 15 Å, which makes manipulations with the length of Helix 2 an effective but rather rough exercise, and may require an additional smoother tuning in other parts of the G-ribo ring. Such a tuning is provided through the displacement between base pairs [− 2P; − 2Q] and [− 1P; − 1Q] in both motifs L2323 and L2383 (Fig. 3E,F). In these motifs, like in those forming the wrench arrangements, the chain break between positions − 1P and − 2P allows a displacement of base pair [− 2P; − 2Q] with respect to [− 1P; − 1Q] to the extent that − 2P becomes stacked to − 1T. This displacement facilitates the formation of the kissing helix between the − 1 layers of both G-ribo motifs. Compared to the wrench arrangements, the displacement of base pair [− 2P; − 2Q] in motifs L2323 and L2383 is smaller, which indicates the existence of some flexibility in this region. This flexibility allows the juxtaposition of base pairs [− 1P; − 1Q] and [− 2P; − 2Q] to be tuned to the requirements imposed by the particular type of pseudoknot.

The structure of the G-ribo ring demonstrates that with a proper choice of the length of Helix 2 and of the juxtapositions of base pairs [− 2P; − 2Q] and [− 1P; − 1Q] in both motifs L2323 and L2383 the two kissing regions can be brought close to each other. It may, however, be more difficult to achieve an arrangement in which Helix 85, the

kissing helix, and Helix 82 form together a coaxially stacked domain. As one can see in Figure 2E, there is mismatch A2327–A2388 in the middle of the kissing double helix. In the tertiary structure, the two adenosines, instead of forming a base pair, occupy neighboring layers and stack to each other. The presence of an unpaired region in the middle of the kissing helix provides an additional flexibility to the latter, which allows the formation of a bend for about 65° between A2388 and base pair A2326–U2389 (Supplemental Fig. 3C). Without such a bend, the closure of the G-ribo ring seems to be essentially more difficult if possible at all, even though the kissing loops remain proximal to each other due to the proper choice of the other aspects discussed above. The bend in the kissing helix is stabilized by the A-minor interactions (Doherty et al. 2001; Nissen et al. 2001) of the two adenosines A2327–A2388 with Helix 81 of 23S rRNA (not shown). This interaction constitutes the only contact made by an element of the G-ribo ring with the rest of 23S rRNA. As in the case of the G-ribo wrenches, the compactness of the G-ribo ring and its saturation with secondary and tertiary interactions suggest that it can be stable independently of other parts of the ribosome.

3.9 Evolutionary conservation of the G-ribo-based pseudoknots

All identified G-ribo-based pseudoknots are present in the ribosomes of all prokaryotic organisms. This conclusion is based on the following observations. First, all four pseudoknots discussed here had been predicted as universal secondary structure elements based on comparative analysis of the nucleotide sequences of bacterial and archaeal ribosomal RNA long before the elucidation of the ribosome tertiary structure (Gutell et al. 1994). We now know that these pseudoknots are formed with help of the G-ribo motif, all examples of which within the ribosome were recently shown to be highly conserved in all prokaryotes (Steinberg and Boutorine 2007).

The structures of the pseudoknots are also highly conserved. Analysis of the available nucleotide sequences of ribosomal RNA (Wuyts et al. 2004) shows that except for the above-mentioned variation in the S521 pseudoknot between bacteria (Fig. 2D, Supplemental Fig. 2C) and archaea (Supplemental Fig. 4) the secondary structures of all other pseudoknots are the same in all prokaryotic organisms. Also, all G-ribo based

pseudoknots have virtually identical conformations in all available high resolution structures of the ribosome and of its subunits (Schluenzen et al. 2000; Wimberly et al. 2000; Harms et al. 2001; Nissen et al. 2001; Schuwirth et al. 2005; Korostelev et al. 2006; Selmer et al. 2006), including that of the archaeal 50S subunit (Nissen et al. 2001). The level of conservation of some tertiary elements of the pseudoknot structure also deserves mentioning (Supplemental Table 1). Thus, in all G-ribo motifs, the predominant identity of nucleotide -1T is adenosine (Wuyts et al. 2004). However, for different G-ribo motifs, the level of conservation varies between 54% and 100%. Only in the three motifs corresponding to the G-ribo wrenches this level is approaching or equal to 100%. So a high level of conservation is understandable in view of the special role played by − 1T in providing a strong displacement of base pair [− 2P; − 2Q] with respect to [− 1P; − 1Q] observed in the wrench pseudoknots.

Uridine in position − 5Sa exists in almost 100% of all prokaryotic sequences of S861, while in L1024, it is conserved only in archaea. In bacteria, on the contrary, position − 5Sa in L1024 is occupied by uridine only in 87.5% of the cases, while in 9% of the sequences nucleotide − 5Sa is cytidine. The cytidine identity of − 5Sa allows this nucleotide to form the Hoogsteen base pair with adenosine − 1T, which would be rather similar to the Hoogsteen UA base pair formed if − 5Sa is uridine. Compared to the uridine − 5Sa in L1024 and S861, the uridine − 1Ta in S521 is less conserved, which reflects the fact that this nucleotide, unlike − 5Sa, is not involved in formation of specific hydrogen bonds with − 1T, so that the constraints imposed on its identity can be less restrictive.

In the G-ribo motifs composing the G-ribo ring, the identity of − 1T is less restricted to adenosine than in the G-ribo wrenches, seemingly, because of a smaller displacement of base pair [− 2P; − 2Q], which does not require that the interaction between − 2P and − 1T be very stable, and also because of the absence of specific interactions equivalent to that between − 1T and − 5Sa in L1024 and S861. Another element of the structure of the G-ribo ring built of adenosines A2327–A2388 is conserved at the level of 100% both in bacteria and archaea, reflecting its pivotal role in the whole arrangement.

To conclude, not only all G-ribo-based pseudoknots are highly conserved, but also conserved are the structural elements important for their formation. Deviations from the predominant identities are allowed only for those elements whose variation does not compromise the integrity of the arrangements.

3.10 Discussion

The results presented here clearly position the G-ribo motif as pseudoknot prone. Indeed, out of eight G-ribo motifs identified in the ribosome, five are involved in the pseudoknot formation. Also, within the ribosomal RNA, there are 10 pseudoknots in which the characteristic secondary structure elements contain at least two base pairs (Gutell et al. 1994), and the G-ribo-based pseudoknots account for almost a half of them. An important feature of the G-ribo motif related to its proneness toward pseudoknots is the side-by-side juxtaposition of Helices 1 and 2. In the G-ribo wrenches, this juxtaposition facilitates the covalent reconnections between strands P and S. In the G-ribo ring, the side-by-side arrangement of Helices 82 and 83 as well as of Helices 83 and 85 mediated by two G-ribo motifs L2383 and L2323 brings the two kissing regions close to each other.

The particular juxtaposition of Helices 1 and 2 within the G-ribo motif is not, however, the only feature of its structure that may be relevant for the pseudoknot formation: all G-ribo-based pseudoknots contain a chain break between positions − 1P and − 2P. In the G-ribo ring, even the second chain break between positions − 8Q and − 9Q is, in fact, one between − 1P and − 2P, if the second G-ribo motif is taken as a reference point. Such a universal position of this chain break is rather surprising, given that the wrench and ring pseudoknots in many aspects are very different.

An important aspect of the chain break between − 1P and − 2P is that it allows an additional tuning in the position of the lower part of Helix 1, thus facilitating the other reconnections of the polynucleotide chain between the strands. Although this tuning is important for pseudoknot formation, it cannot explain the universal position of the break between − 1P and − 2P, because a break between 0P and − 1P or between − 2P and − 3P would allow similar movements of Helix 1 toward Helix 2. The uniqueness of the break

between − 1P and − 2P seems to be linked to the presence of adenosine − 1T, the interaction with which would favor the displacement of nucleotide − 2P. Thus, the particular location of nucleotide − 1T and its openness for interaction with a nucleotide of layer − 2 would favor the formation of arrangements with a chain break between − 1P and − 2P and with a displacement of nucleotide − 2P toward Helix 2.

While nucleotide − 2P stacks to − 1T, its position still remains flexible and can adapt to the requirements of the particular type of pseudoknot. However, the stacking with − 1T per se suffices the promotion of only a relatively mild displacement of − 2P like that present in the G-ribo ring. For a stronger displacement observed in the G-ribo wrenches, an additional stabilization of the − 2P position through the contact of its ribose with either − 5Sa or − 1Ta is required.

Finally, we would like to compare the G-ribo-based pseudoknots with other known pseudoknots. Our inspection of different RNA structures showed that the particular arrangement of base pairs [− 1P; − 1Q] and [− 2P; − 2Q] stabilized by adenosine − 1T, which is at the core of all four pseudoknots discussed in this paper, exists also in other pseudoknots, unrelated to the G-ribo motif, both in and outside the ribosome (Supplemental Fig. 5). We can thus conclude that this arrangement, which facilitates the particular type of chain switch between two quasicoaxial double helices, is an essential part of the structural contexts favoring the formation of different kinds of pseudoknots. From this point of view, the G-ribo motif represents a particular way of the polynucleotide chain arrangement around the central part that would provide for a compact and stable structure suited to a particular function

3.11 Acknowledgements

The authors are grateful to Dr. Lea Brakier-Gingras for important discussions. SVS acknowledges a grant from CIHR and fellowships from CIHR and FRSQ.

3.12 References

Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J., and Strobel, S.A. 2004. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430:** 45-50.

Doherty, E.A., Batey, R.T., Masquida, B., and Doudna, J.A. 2001. A universal mode of helix packing in RNA. *Nat Struct Biol* **8:** 339-343.

Egli, M., Minasov, G., Su, L., and Rich, A. 2002. Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution. *Proc Natl Acad Sci U S A* **99:** 4302-4307.

Gutell, R.R., Larsen, N., and Woese, C.R. 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., and Yonath, A. 2001. High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107:** 679-688.

Ke, A., Zhou, K., Ding, F., Cate, J.H., and Doudna, J.A. 2004. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* **429:** 201-205.

Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H.F. 2006. Crystal Structure of a 70S Ribosome-tRNA Complex Reveals Functional Interactions and Rearrangements. *Cell* **126:** 1065-1077.

Michiels, P.J., Versleijen, A.A., Verlaan, P.W., Pleij, C.W., Hilbers, C.W., and Heus, H.A. 2001. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J. Mol. Biol.* **310:** 1109-1123.

Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B., and Steitz, T.A. 2001. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci.* **98:** 4899-4903.

Nixon, P.L., Rangan, A., Kim, Y.G., Rich, A., Hoffman, D.W., Hennig, M., and Giedroc, D.P. 2002. Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J. Mol. Biol.* **322:** 621-633.

Pleij, C.W., Rietveld, K., and Bosch, L. 1985. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.* **13:** 1717-1731.

Rastogi, T., Beattie, T.L., Olive, J.E., and Collins, R.A. 1996. A long-range pseudoknot is required for activity of the Neurospora VS ribozyme. *EMBO J.* **15:** 2820-2825.

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

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

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

Shen, L.X. and Tinoco, I., Jr. 1995. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol.* **247:** 963-978.

Steinberg, S.V. and Boutorine, Y.I. 2007. G-ribo: A new structural motif in ribosomal RNA. *RNA* **13:** 549-554.

ten Dam, E.B., Pleij, C.W., and Bosch, L. 1990. RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. *Virus Genes* **4:** 121-136.

Theimer, C.A., Blois, C.A., and Feigon, J. 2005. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell* **17:** 671-682.

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. Structure of the 30S ribosomal subunit. *Nature* **407:** 327-339.

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

3.13 Figures

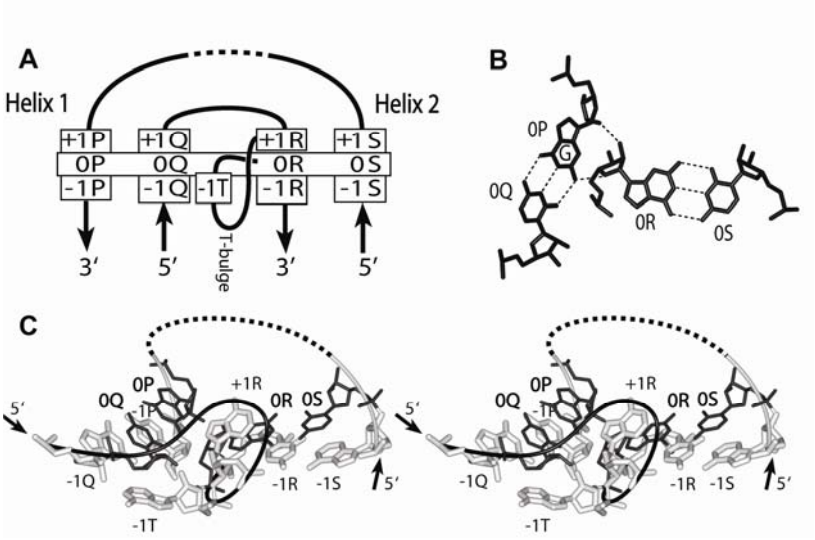


Figure. 1. The general description of the G-ribo motif. (A) The secondary structure of the G-ribo motif. Helices 1 and 2 are, respectively, on the left and on the right. They are composed of, respectively, strands P and Q and strands R and S. The dashed element indicates a possibility of a longer connector region between strands S and P, which can include additional secondary structure elements. The name of each nucleotide consists of the number of the layer and of the letter that indicates the strand to which it belongs. The top base pairs of both helices [0P;0Q] and [0R;0S] form the zero-layer. For other layers within the helices, the numbering propagates in the negative direction. For the nucleotides of the connector regions that stack on top of the nucleotides of the 0-layer, the numbers are positive. There is a bulge between nucleotides 0R and +1R (so-called T-bulge), which in most cases consists of only one nucleotide occupying position -1T. (B). The juxtaposition of base pairs [0P;0Q] and [0R;0S]. Dashed lines stand for hydrogen bonds within and between the base pairs. In this juxtaposition, the ribose of 0R interacts with the ribose and the base of 0P. To make this interaction possible, 0P should be guanosine. (C). The tertiary structure of the G-ribo motif S861 from the *E. coli* ribosome (Schuwirth et al. 2005). The nucleotides forming the zero-layer are black. Other nucleotides are white. At layer +1, only nucleotide +1R is shown. Nucleotide -1T, which is predominantly adenosine, forms the A-minor interaction (Doherty et al. 2001; Nissen et al. 2001) with base pair [-1P;-1Q].

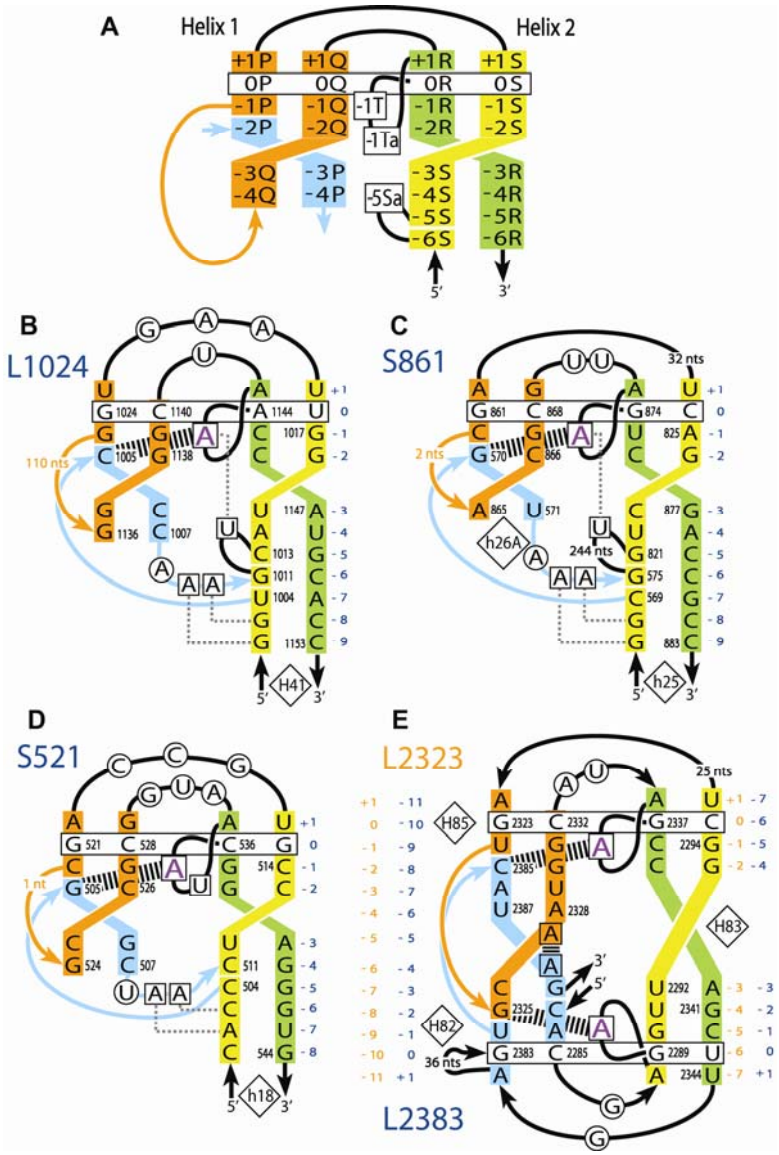


Figure. 2. The secondary structures of the four G-ribo-based pseudoknots in the *E.coli* rRNA. (A). The template used for depicting the secondary structures. The base pairs at the zero-layer are enclosed in the horizontally oriented rectangle. In all pseudoknots, there is a break of the polynucleotide chain in strand P between positions -1P and -2P. Nucleotide -1P is connected to a lower layer of strand Q, thus forming a stem-and-loop structure (orange) in Helix 1. Depending on the type of pseudoknot, region [-2P;-4P] (cyan) is connected to different parts of the structure. Strands R and S of Helix 2 are, respectively, green and yellow. Nucleotide -1Ta exists only in S521, while nucleotide -5Sa exists only in L1024 and S861. (B-E). In all structures, the first letter in the name of a motif (S or L) stands for the ribosomal subunit, small or large, in which the motif has been found. The number in the name corresponds to that of nucleotide 0P in the standard *E.coli* numeration of rRNA. The numbers of the helices in the standard 16S and 23S rRNA

secondary structures are shown in diamonds. The unpaired nucleotides that are important for the integrity of the pseudoknots are squared. Adenosine -1T is shown in magenta. All other nucleotides outside the double helical regions are circled. The zebra bands connecting positions -2P and -1T as well as adenosines A2327 and A2386 (in (E)) indicate the stacking interactions essential for the integrity of the arrangement.
(B-D). Different G-ribo wrenches. In these structures, nucleotides -2P, -3P and -4P (-4P exists only in (B) and (D)) belong to the pseudoknot bulge (cyan). The pseudoknot bulge is inserted into strand S either between positions -7S and -6S (L1024 and S861) or between positions -5S and -4S (S521). The numbers of layers in both Helices 1 and 2 are shown in blue. The dashed lines connecting the two last adenosines of the pseudoknot bulge with the base pairs in Helix 2 indicate the A-minor interactions (Doherty et al. 2001; Nissen et al. 2001). The dashed line connecting adenosine -1T and uridine -5Sa (only in L1024 and S861) stands for the Hoogsteen base pair.
(E). The G-ribo ring. The structure is formed by two symmetrically positioned G-ribo motifs L2323 and L2383. The numbers of layers are provided individually for each helix and for each motif. With respect to motifs L2323 and L2383, the numbers are, respectively, blue and orange. Strand P in L2323 (orange in the upper part, cyan in the lower part) becomes strand Q in L2383. Strand Q in L2323 (cyan in the upper part, orange in the lower part) becomes strand P in L2383. In L2383, the cyan stem-and-loop plays the same role as the orange stem-and-loop in L2323. The two loops orange and cyan form the kissing interactions within Helix 1. Within the kissing helix, adenosines A2327 and A2386 are not involved in base pairing and stack to each other. The green and yellow strands are, respectively, strands R and S in L2323 and strands S and R in L2383.

Figure 3

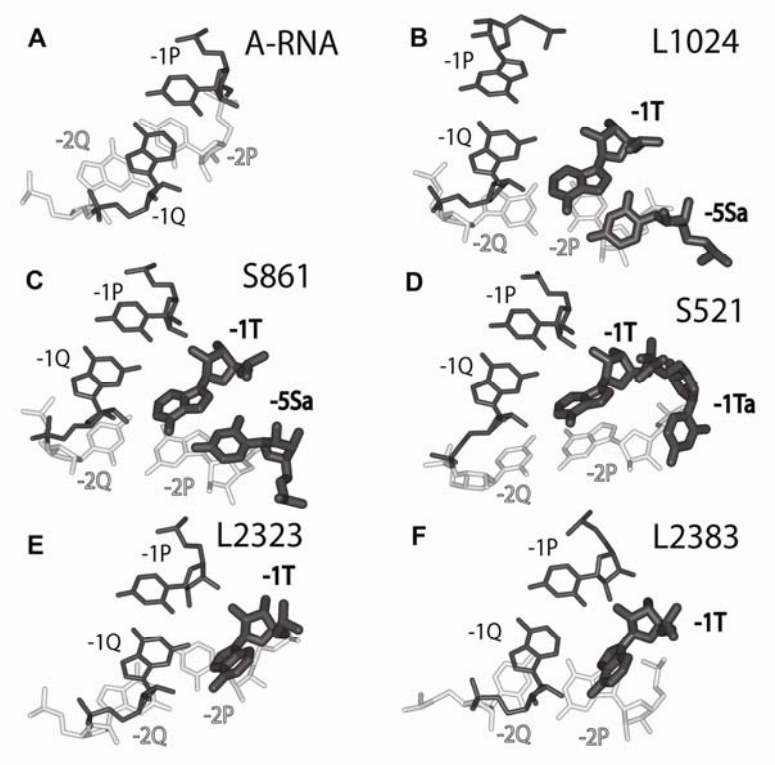


Figure. 3. The displacement of base pair [-2P;-2Q] with respect to [-1P;-1Q] in the G-ribo-based pseudoknots. All structures are aligned by the position of nucleotide -1Q. Nucleotides -2P and -2Q are white, while all other nucleotides are black. Nucleotides -1T, -5Sa (only in L1023 and S861) and -1Ta (only in S521), which are involved in the stabilisation of the position of nucleotide -2P, are shown thick. Compared to the juxtaposition of two consecutive base pairs in a regular RNA double helix (A), base pair [-2P;-2Q] in the G-ribo wrenches (B-D) is rotated for about 50º around atom O3´ of -2Q in the direction of the minor groove and is additionally shifted for about 4Å. In motifs L2323 and L2383 forming the G-ribo ring (E,F), the displacement of base pair [-2P;-2Q] is smaller than in the G-ribo wrenches, and in L2323 (E), it is smaller than in L2383 (F). Due to this displacement, -2P in all pseudoknots has lost its interaction with -1P and becomes stacked to -1T. Additional stabilization of the -2P position in the G-ribo wrenches is provided through the interaction of its ribose with either -5Sa or -1Ta. Uridine -5Sa, when exists, forms a Hoogsteen base pair with adenosine -1T (B,C).

3.14 Supplemental Table

Supplemental Table 1

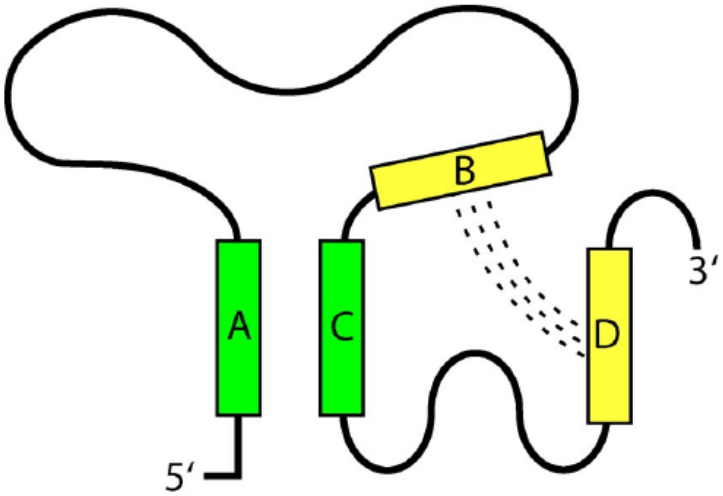
		L1642	L1309	L2383	L2323	L1024	S1047	S521	S861
[0P;0Q]	GC	98.0 100	96.5 100	99.8 100	86.4 100	98.5 91.9	93.6 98.3	99.3 98.1	88.7 73.2
	GU		3.5		13.4	0.3	5.9 0.6	0.1	5.5 26.7
[0R;0S]	WC	95.7 100	99.7 100	19.6 78.4	98.0 100	98.6 100	86.7 88.9	99.9 99.3	99.5 99.5
	GU	2.5		80.4 21.6	2.0	1.1	12.6 10.0		0.2 0.2
+1R	A	99.8 100	98.5 100	100 100	100 100	100 100	99.8 99.6	99.9 99.8	99.7 99.2
+1P	A	99.8 100	100 100	100 100	97.2 100		99.9 96.6	99.9 100	96.8 99.8
	U				1.0	100 100	0.1 0.2		0.2
+1S	U	5.5	99.8 100	100 100	100 100	99.0 100	99.7 0.8	99.9 99.8	94.9 99.4
	C	1.3 84.9				0.1	0.1 78.6	0.1	5.1 0.2
	A	51.5 12.1	0.2				0.1 18.3		0.2
	G	41.7 3.0					0.1 2.2	0.2	0.2
+1Q	U	0.5 13.5	1.5 13.5				99.4 3.5		2.5 0.5
	C	0.2	42.3 13.5				0.1	0.1 0.2	0.7
	A	0.2	6.1 2.7				0.4 0.2		59.3 1.9
	G	99.1 86.5	50.1 70.3				0.1 96.3	99.9 99.8	37.5 97.6
-1T	U	24.6 8.1	4.5 18.9	12.8 8.1	5.8		61.4 72.2	0.05	0.1
	C	8.1		1.2	5.4		38.4 0.4	0.05 0.3	
	A	75.4 54.0	95.5 81.1	86.0 91.9	94.2 94.6	100 100	0.1 27.4	99.8 99.7	99.8 100
	G	29.8					0.1	0.1	0.1
-1Ta	U							93.6 58.8	
	C							1.9 33.7	
	A							0.8 7.2	
	G							3.7 0.3	
-5Sa	U					87.5 100			99.9 99.0
	C					9.0			0.5
	A					0.8			0.1
	G					2.7			0.5
-5Q	A				100 100				
-6P	A				100 100				

Supplemental Table 1. The frequency of the occurrence of particular identities for base pairs and individual nucleotides in different G-ribo motifs.

The numbers were calculated as a percentage of all analyzed nucleotide sequences of 16S rRNA (for motifs S521, S861 and S1047) or 23S rRNA (for motifs L1024, L1309, L1642, L2323 and L2383). The numbers without and in parentheses pertain to Eubacteria and Archaebacteria, respectively. The nucleotide sequences were taken from the European ribosomal RNA database (Wuyts *et al.*, 2004). The analyzed sequences were current as of May 2007.

3.15 Supplemental figures

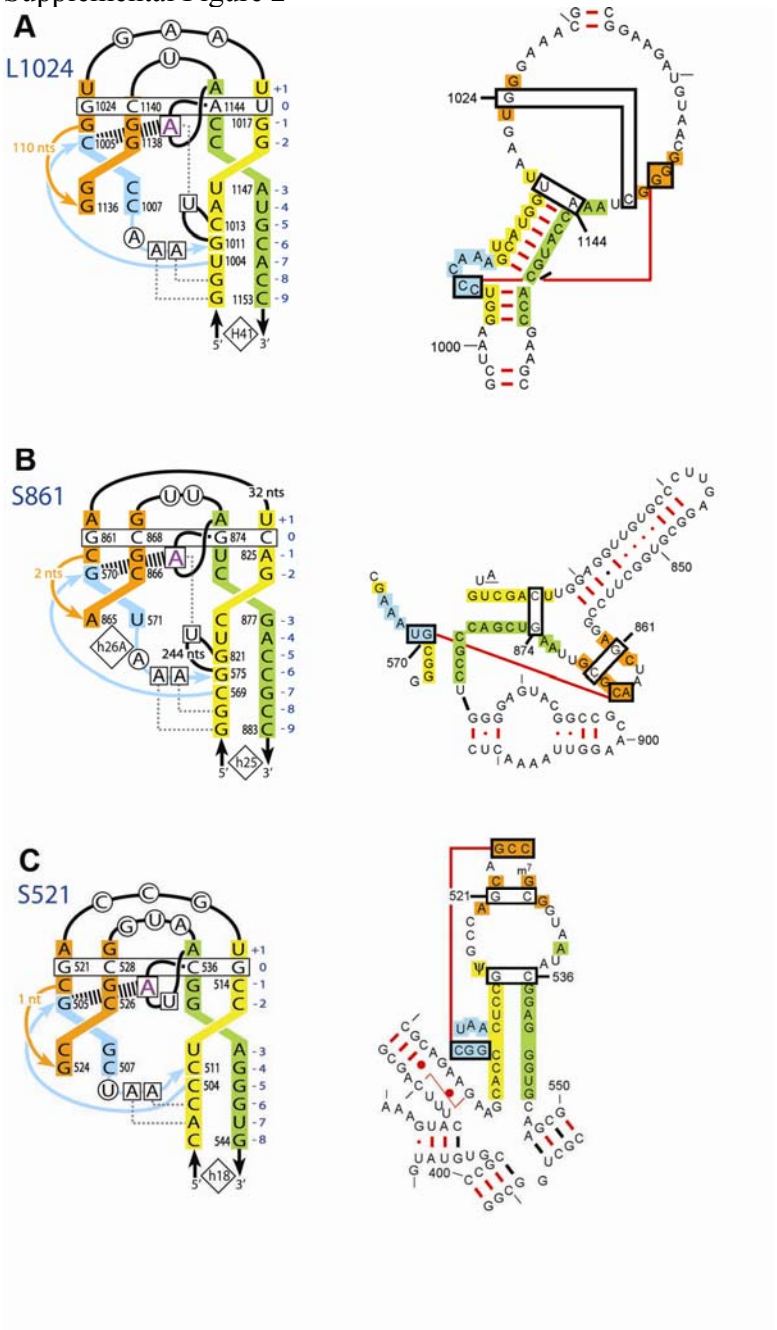
Supplemental Figure 1

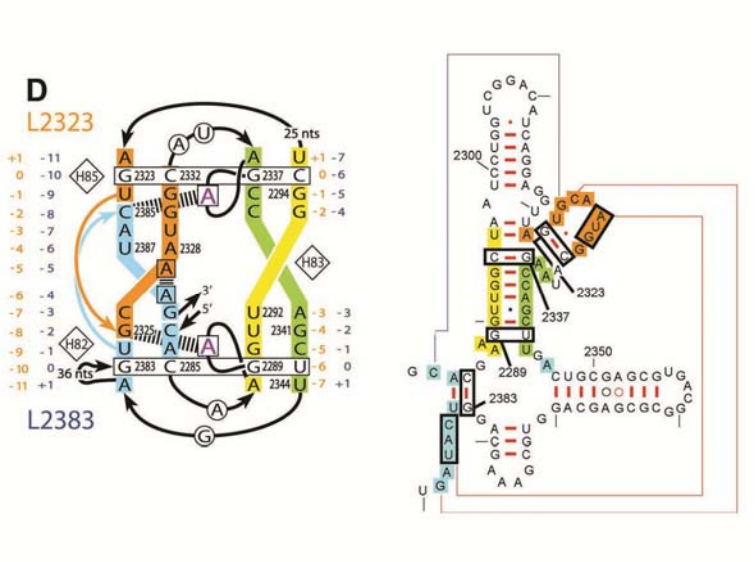


Supplemental Figure 1. The pseudoknot definition.

In this figure, regions A and C form a double helix.The whole stretch encompassing A, C and the region between them is a hairpin. Region B is a part of the hairpin loop, while Region D is outside the hairpin proper. If regions B and D also form a double helix, the whole arrangement becomes pseudoknot regardless of the structure and of interactions of other parts of the polynucleotide chain not involved in fragments A, B, C and D. All pseudoknots discussed in this paper can be arranged in the way shown in this figure.

Supplemental Figure 2

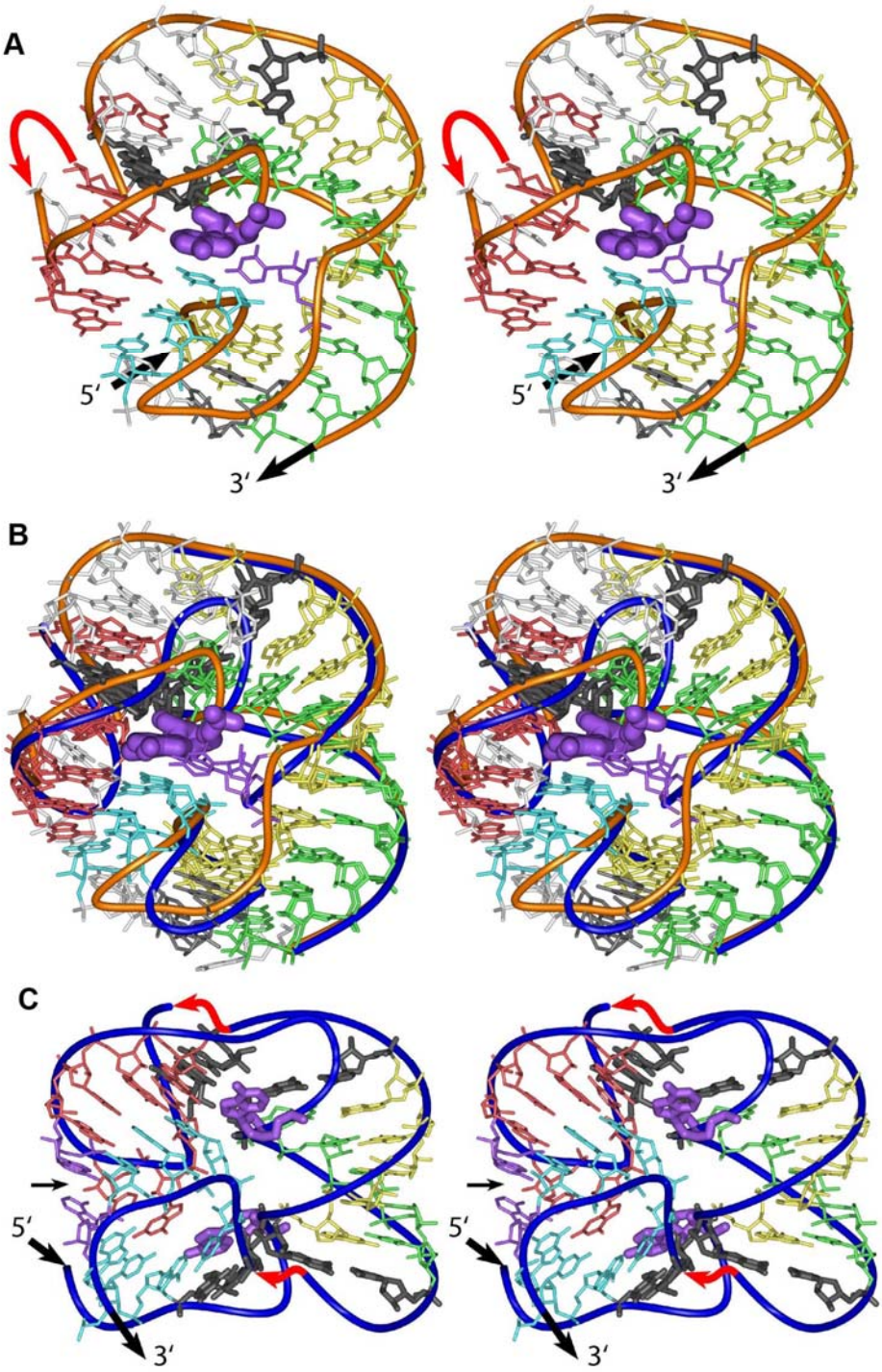




Supplemental Figure 2.The secondary structures of the G-ribo-wrenches (A, B, C) and of the G-ribo ring (D) found in the *E.coli* ribosome.

On the left: the secondary structures of the G-ribobased pseudoknots as they are shown in Fig. 2 of the paper. On the right: the same region shown as a part of the standard secondary structure of either 16S or 23S rRNA. The same nucleotides are coloured in the same way both on the left and on the right. The boxed uncoloured nucleotides form the base pairs at the 0-layer. The boxed coloured nucleotides (only on the right) form the double helices that constitute the pseudoknots.

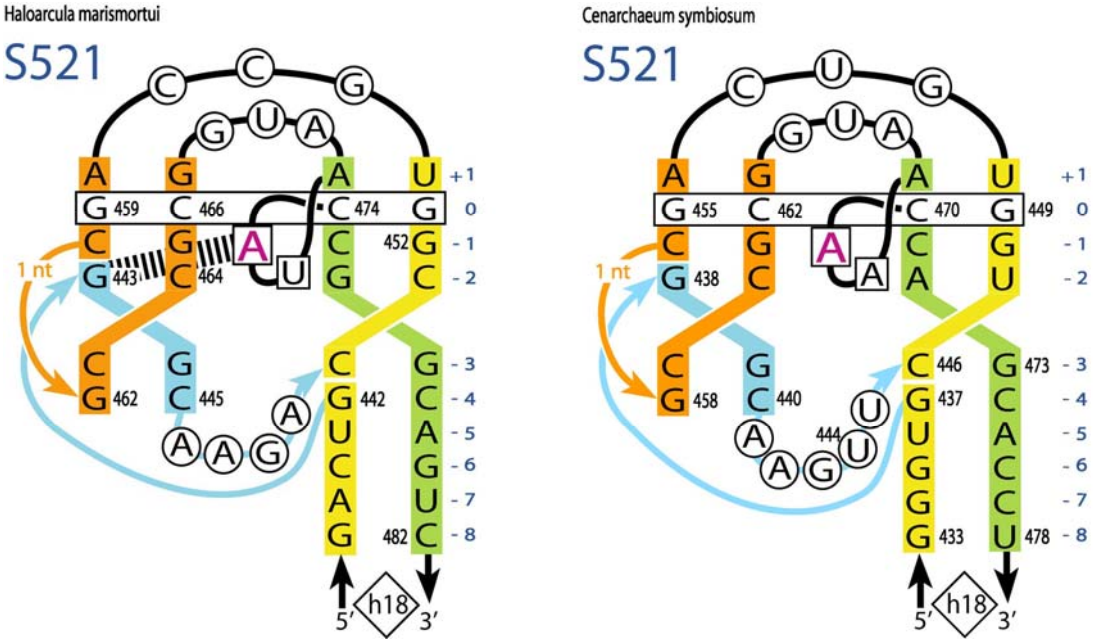
Supplemental Figure 3



Supplemental Figure 3. The tertiary structure of the G-ribo wrench L1024 (A), the superposition of the structures of the G-ribo wrenches L1024 (brown ribbon) and S861 (blue ribbon) (B) and the tertiary structure of the G-ribo ring (C).

All structures are shown in the same orientation and with the same colors as in Fig. 2 of the paper. Adenosine -1T is shown in magenta with thicker covalent bonds. The insertions that are not relevant to the pseudoknot structures are red. (A,B): Uridine -5Sa is in magenta. The nucleotides circled in Fig. 2B and 2C are grey. In L1024, insertion 1026-1135 is not shown. (B): In S861, neither insertion 576-819 nor 828-859 is shown. The superposition demonstrates a very strong similarity between the two structures over eleven layers from +1 to -9, i.e. until the end of Helix 2 in both motifs. The major differences between the two structures deal with the absence of base pair [-4P;-4Q] in S861 and with a longer region between 0Q and +1R in S861 compared to L1024. (C): For clarity, the nucleotides of positive layers in both motifs L2323 and L2383 are not shown. The small black arrow points to adenosines A2327 and A2386, which are shown in magenta. These adenosines stack to each other and to the upper part of Helix 1. At the same time, they are almost perpendicular to the lower part of this helix.

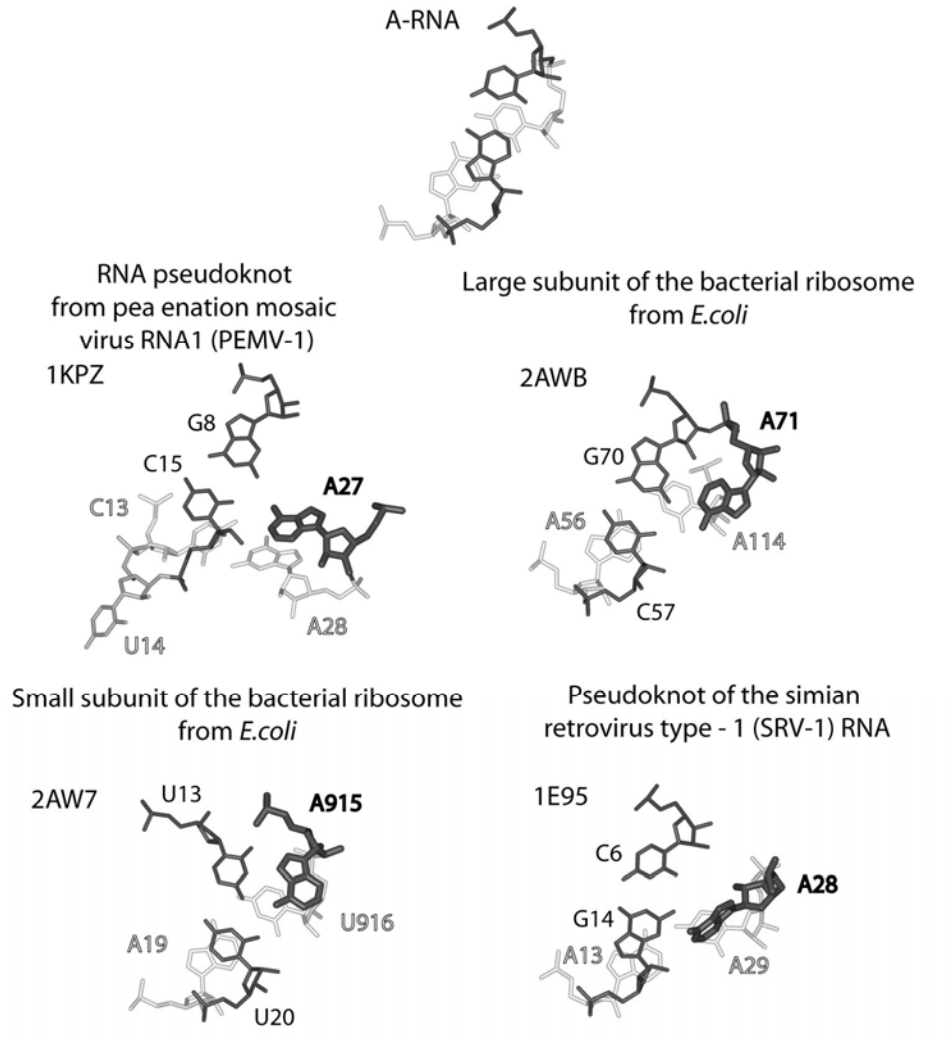
Supplemental Figure 4



Supplemental Figure 4. Possible structural variations of the G-ribo wrench S521 in different archaeal organisms.

The secondary structure of the S521 pseudoknot in archeabacteria *Haloarcula marismortui*. In this structure, like in other archaeal cases of pseudoknot S521, the pseudoknot bulge is inserted into strand S between position -4S and -3S, rather than between -5S and -4S, which happens in all eubacterial cases. In addition, the pseudoknot bulge in this structure, as well as in other archaeal cases, is longer than in eubacteria.

Supplemental Figure 5



Supplemental Figure 5. The arrangements resembling that of base pairs [-1P;-1Q] and [-2P;-2Q] and nucleotide -1T found in different pseudoknots not related to the G-ribo motif.

The arrangements are shown in the same way as in Fig. 3. For each arrangement, the name of the molecule, the numbers of the nucleotides as well as the pdb-identifier are provided. In all examples presented here, unlike in the G-ribo-based pseudoknots, adenosine -1T is immediately linked either to -1P or -2P.

Chapter 4

Article 3

Double twist-joints – new recurrent RNA motifs

4. Double twist-joints – new recurrent RNA motifs

Yury I. Boutorine and Sergey V. Steinberg

Département de Biochimie, Université de Montréal, Montréal, QC, Canada.

Manuscript in preparation

Running title: DTJ – a new RNA motif

Contribution of each authors:
Yury I. Butorin: developed the general idea of the motif, searched for the cases of the motifs in the RNA structure, participated in data analysis, preparation of manuscript and figures.
Sergey V. Steinberg: participated in data analysis, preparation of manuscript and figures.

4.1 Abstract

Analysis of the available crystal structures of the RNA molecules has allowed identification of a new RNA motif that we call Double-Twist Joint, or DTJ motif. The motif is composed of an RNA double helix with two bulges, which are incorporated into two strands of the helix and separated by two or three layers of stacked nucleotides. In DTJ, incorporation of a bulge results in the formation of a helical over-twist between two WC base pairs flanking the bulge. The unpaired nucleotides of each of the two bulges form a complex network of stacking and non-canonical base pairs with nucleotides that form the helical over-twist. In total we identified three types of different DTJ motifs, each representing a group of four related structural arrangements. Two DTJ motifs are located in the important functional centers of ribosome and RNAse P.

Keywords: RNA structure / RNA motif / nucleotide bulge / helical over-twist

4.2 Introduction

Recurrent RNA motifs are important elements of RNA structure. There are at least two reasons why studying RNA motifs is important. First, motifs can form independently (at least partly) of their structural context. Such independence would allow us to approach the problem of RNA structure formation not in most general terms, but on a limited scale, thus making the problem much simpler. Second, RNA motifs possess a unique characteristic – they are able to form similar shapes from seemingly unrelated primary sequences. The latter requires detailed analysis of the rules that govern the principles of the motif's folding. Subsequently, identification and characterization of new RNA motifs is important for understanding the general principles of RNA structure formation.

Analysis of 3D structure or RNA shows that the regular double helix is the most common structural element observed in the large RNA molecules such as ribosomal RNA, RNAse P, tRNA etc. Interestingly, uninterrupted RNA double helix rarely exceeds 7 consecutive layers of canonical base pairs. Globally, individual double helices are connected to each other with stretches of unpaired nucleotides giving rise to the various elements of the secondary structure such as nucleotide bulges, internal loops, and three-way junctions. In this article we describe a new RNA motif, which represents a combination of three short double helices connected with a loop.

4.2.1 The twist-joint structures

In the available RNA structures, one can find many situations when one double helix stacks on top of another double helix in such a way that, compared to the standard A-RNA conformation, the two contacting base pairs become over-twisted with respect to each other (Figure 1). Henceforth, we will call such arrangements twist-joints or TJs. Usually in a TJ, two strands in both helices are directly connected, while the other two strands are separated by an unpaired region or belong to different parts of the polynucleotide chain. The level of the over-twist varies between 20˚ and 70˚ and in most cases is about 30-50˚.

For the description of TJs, we will use the nomenclature introduced in Figure 2. The two helices composing a TJ are named H1 and H2. Strand P is common for both helices: its 5'-half belongs to H2, while its 3'-half is a part of H1. In helix H1, strand P forms a duplex with strand R, while in H2, strand P forms a duplex with strand Q. The two base pairs in both helices that form the inter-helix contact are [0P; 0Q] and [+1R; +1P].

Due to the over-twist between the two contacting base pairs, only one nucleotide in each base pair forms a contact with a nucleotide of the other helix (nucleotides 0P and +1P in Figure 1), while the other two nucleotides of these base pairs (nucleotides +1R and 0Q) are clearly separated from each other. On most occasions, nucleotides +1R and 0Q are connected by a relatively short loop L1 containing three nucleotides or less.

Compared to the standard A-RNA conformation, the over-twist between helices H1 and H2 corresponds to the mutual displacement of the contacting base pairs [+1P; +1R] and [0P; 0Q] in the direction of the major and minor grooves, respectively. Such displacement would weaken the interaction between the two base pairs and would thus require additional stabilization. Indeed, inspection of the TJs existing in the available RNA structures shows that practically in all of them, the juxtaposition of the two helices is stabilized through interaction with other elements of the same RNA molecule. One can identify two different types of stabilization, which we call here direct and indirect.

4.2.2 Direct stabilization of TJ

The direct stabilization is the most common type of stabilization of TJ arrangements. It takes place when an additional nucleotide is positioned at the junction point between the two helices, where it simultaneously stacks to the top of one helix and forms H-bonds with a nucleotide of the other helix. The general scheme of the direct stabilization is shown in Figure 2. Two nucleotides 0R and +1Q are involved in the direct stabilization of the over-twist between base pairs [0P; 0Q] and [+1R; +1P]. Nucleotide 0R binds to the major groove of helix H2, while nucleotide +1Q does it in the minor groove of helix H1. Additional nucleotides -1R and +2Q can stack to 0R and +1Q, respectively, and form hydrogen bonds with the corresponding base pairs in helices H2 and H1. The presence of these nucleotides would provide further stabilization of the whole arrangement.

An example of the direct TJ stabilization can be seen in the well-known C-loop motif (Lescoute et al., 2005), a typical case of which is shown in Figure 3A. Nucleotide 0R (C372) forms a trans-WC-HG base pair with 0P (A389) (Figure 3A,B), +1Q (A374) forms a cis-HG-SE base pair with +1P (U390), while +2Q (A373) forms a cis-HG-SE base pair with +1R (A371) (Figure 3A,B).

In different types of TJs, the particular patterns of direct stabilization can differ by the presence or absence of each of the four nucleotides -1R, 0R, +1Q and +2Q, by the position of these nucleotides in the RNA chain, and by the mode of their interaction with helices H1 and H2.

4.2.3 Indirect stabilization of TJ

The indirect TJ stabilization takes place when the positions of nucleotides 0Q and +1R are immobilized through immediate covalent connection to nucleotides that are firmly integrated with other parts of the same RNA structure. An example of indirect TJ stabilization represents arrangement 2AVY-44 found in the crystal structure of the 30S ribosomal subunit from *E. coli* (2AVY), Figure 4. In this arrangement, adenosine A397, which intervenes between nucleotides 0Q (U398) and +1R (C396), forms a WC base pair with uridine U37 coming from another part of the 16S rRNA structure. The fixation of the position of adenosine A397 through base paring with uridine U37 would indirectly stabilize the positions of both flanking nucleotides 0Q and +1R, thus solidifying the over-twist between base pairs [0P; 0Q] (base pair [A44; U398]), and [+1P; +1R] ([G45; C396]). Although indirect stabilization happens more rarely that the direct one, in the available RNA structures there are many TJs stabilized in this way.

4.2.4 Double twist-joints

A special type of arrangement forms when two TJs co-exist in the same duplex, being separated from each other by only a few base pairs. We call such arrangements Double Twist-Joints or DTJs. In the available RNA structures we found a dozen of cases of this kind. As demonstrated below, the essential feature of DTJs that does not reduce them to a simple merger of two individual TJs is that due to the closeness of the two TJs, the strategies of their stabilization influence each other. This proximity provides a common strategy for DTJ stabilization. In the identified DTJs we found three different types of over-twist stabilization, which we have named A, B and C. All three types fit to the same pattern of secondary structure (Figure 5), which will be used for their description. The pattern consists of three consecutive double helices H1, H2 and H3, forming two twist-joints, TJ1 and TJ2 and having two loops, L1 and L2. In spite of the visible symmetry of this pattern, the real DTJs are essentially asymmetric. One of the two TJs, which we name TJ1, is always stabilized directly in the major groove and sometimes in the minor groove. The other twist-joint, TJ2, can have no direct stabilization; sometimes, it can be stabilized directly in the major groove and is never stabilized in the

minor groove. To number layers in the whole DTJ structure, we use the numbering system shown in Figure 2, by linking it to the TJ1.

4.3 A-DTJs

The A-DTJs are defined as those DTJs in which helix H2 consists of three base pairs (Figure 6A). In the available RNA structures, we found three A-DTJ arrangements, all in the 16S rRNA. The A-DTJs existing in the *E. coli* ribosome are shown in Figure 6 (B-D). All these arrangements contain nucleotides 0R and +1Q, which are, respectively, the first nucleotide of loop L2 and the last nucleotide of loop L1. Both nucleotides 0R and +1Q play important role in stabilization of the TJ1.

Nucleotide +1Q forms a hydrogen bond with purine +1P. Although in the A-DTJs shown in Figure 6 (B-D), +1Q has different identities G, A and U, its interaction with nucleotide +1P seems to be specific. The identities of +1Q and +1P co-vary such that G and U in +1Q correspond to G in +1P, while A in +1Q corresponds to A in +1P. Such co-variation always allows the formation of a hydrogen bond between the two nucleotides (Supplemental Figure 1). Nucleotide 0R, in its turn, stacks to pyrimidine +1R and forms three hydrogen bonds with the major groove of base pair [0P; 0Q]. This interaction provides for a specific base triple [0P; 0Q; 0R] shown in Figure 7. The universality of this triple is guaranteed by the fact that in all three A-DTJs, nucleotide 0R is guanosine, while base pair [0P; 0Q] is either GC or GA.

The formation of three hydrogen bonds between 0R and base pair [0P; 0Q] firmly attaches 0R to the whole helix H2 and fixes the juxtaposition of nucleotides 0R and -2Q. The latter nucleotides are directly connected to, nucleotides -3P and -3Q, respectively, which form the first base pair of helix H3. As a result, the position of base pair [-3P;-3Q] becomes fixed with respect to the whole helix H2. We can thus say that in the A-DTJ structure, nucleotide 0R plays a dual role: the direct stabilization of the TJ1 and the indirect stabilization of the TJ2.

In all three A-DTJs the TJ2 is stabilized indirectly. In addition, in motif 2AVY-1057, one can also see a direct stabilization of TJ2 by nucleotide -2R (C1200), which stacks to -3P and interacts with the major groove of base pair [-2P; -2Q]. Interestingly,

motif 2AVY-1057 contains also another unique feature, a deformed base pair [-2P; -2Q] (base pair A1055-U1205 in the 16S rRNA), in which the two bases are linked by only one hydrogen bond (Supplemental Figure 2). We think that such deformation of base pair [-2P; -2Q] would compromise the ability of triple [0R; 0P; 0Q] to indirectly stabilize TJ2. The introduction of nucleotide -2R, which provides the direct stabilization, is thus seen as a compensatory measure that would guarantee the proper juxtaposition of helices H2 and H3 in the situation when the standard means of stabilization becomes unreliable.

Finally, our analysis showed that the formation of triple [0R; 0P; 0Q] requires that helix H2 contain strictly three base pairs. If alternatively, helix H2 contained either two or four base pairs, the orientation of the last base pair of H2 ([-2P; -2Q]) would not have allowed the proper attachment of helix H3. This aspect explains why the above-mentioned triple is found only in A-DTJs, i.e. when and only when helix H2 contains three base pairs.

4.3.1 A quasi-A-DTJ arrangement

In addition to the three A-DTJ motifs discussed above, our screening of the available RNA structures revealed another arrangement, whose structure in many aspects is similar to the A-DTJs, but helix H3 is replaced by a tetraloop (arrangement 2AVY-1142 in the *E. coli* 23S rRNA, Figure 8). The absence of helix H3 leaves no place for TJ2 and does not allow us to qualify 2AVY-1142 as a double-twisted arrangement. In a following section we argue that even though in the available structures of the 30S ribosomal subunit, arrangement 2AVY-1142 does not form an A-DTJ motif, it may do so in other organisms.

4.3.2 The conglomerate of 2AVY-68 and 2AVY-97

In the tertiary structure of 16S rRNA, the two A-DTJs, 2AVY-68 and 2AVY-97, are located so closely to each other that they form a common arrangement (Figure 9). In this arrangement, the two motifs are positioned head-to-head, so that helix H1 becomes common. Also, base pair [+2R; +2P] in 2AVY-97 serves as [+1P; +1R] in 2AVY-68 and *vice versa*. Three nucleotides +1Q (A72) and +2Q (A71) of 2AVY-97 as well as +1Q (G100) of 2AVY-68 form a stack that interacts with the minor groove of the common

helix H1. This stack is flanked by nucleotides 0Q of both motifs, thus reinforcing the direct stabilization of the two adjacent over-twists.

4.3.3 Evolutionary conservation of A-DTJs

In the available nucleotide sequences of prokaryotic 16S rRNA, the pattern of conservation of each of the three identified A-DTJs and of the closely related to them arrangement 2AVY-1142 is different. In particular, motif 2AVY-1057 is highly conserved among prokaryotes. In this motif, each of the five base pairs located between layers -3 and +1 is either WC or GU in more than 96% of the sequences (see Table 1). Due to the fact that +1R is uridine in ~99%, while +1Q is either uridine or guanosine in more than 99.8% of sequences, the cis-WC-SE base pair [+1P; +1Q] can be formed in practically all prokaryotes. Also, because in 99.6% of sequences 0R and 0P are guanosines and 0Q is either cytosine or adenosine, the formation of cis-WC-HG base pair [0R; 0P] and as a result of triple [0P;0Q;0R] is practically always possible. We thus can conclude that in motif 2AVY-1057, both TJ1 and TJ2 exist in virtually all prokaryotes. Such a high level of conservation of this motif does not seem surprising, given that it is located at the decoding region of the 30S ribosomal subunit. More specifically, the last nucleotide of loop L2, a very conserved C1054, forms a part of the A-site and is known to be essential for the ribosome function (Abdi & Fredrick, 2005).

The same analysis performed for motif 2AVY-68 shows, however, a different picture. In this motif, only TJ2 demonstrates a strong conservation among prokaryotes, which is guaranteed by the preservation of three base pairs [0P;0Q], [-1P;-1Q], [-2P;-2Q] and [-3P;-3Q], the GC or GA identity of [0P;0Q] and G identity of 0R (Table 1).

In particular, base pairs [-1P;-1Q], [-2P;-2Q] and [-3P;-3Q] are conserved as WC in 98.6, 98.3 and 99.1% of bacterial sequences and in 99.3, 82.5 and 93.9% of archeal sequences respectively (Table 1). Base pair [0P; 0Q] is conserved as GA in 98.2% of bacterial sequences while in archea it is conserved as GA and GC in 14.0% and 78.0% respectively. Combined with the high conservation of guanosine in position 0R in 97.7% of eubacterial and 69.3% of archeal sequences we can predict the formation of the triple [0P;0Q;0R], which in turn stimulates formation of TJ2 in rRNA of most bacteria.

However, the TJ1 in motif 2AVY-68 is present only in a fraction of all nucleotide sequences. Thus, in more than 75% of bacterial and more than 54% of archaeal sequences, the dinucleotide combination corresponding to base pair [+1R; +1P] is neither WC nor GU. Also, the existence of the cis-WC(HG)-SE base pair [+1Q;+1R] is not predicted in more than 50% of bacterial and 60% of archaeal rRNA sequences. All these data strongly suggest that TJ1 is not formed in most bacterial rRNA. Indeed, homologous arrangement 2J02-68, identified in the crystal structure of the *T. thermophilus* 30S subunit (Selmer et al., 2006) demonstrates existence of TJ2 while TJ1 is eliminated due to the merging of the two helices, H1 and H2, into one.

As for motif 2AVY-97, which forms a conglomerate with motif 2AVY-68, it is even less conserved than the latter: each of the two TJs of 2AVY-97 is conserved in less than 30% of the available 16S rRNA sequences both in bacteria and archaea. Thus, in the whole conglomerate 2AVY-68-2AVY-97, only TJ2 of motif 2AVY-68 is conserved in all prokaryotes, while the other three TJs exist only in a fraction of organisms. An example of alternative arrangement for conglomerate 2AVY-68- 2AVY-97 can be seen in the *T. thermophilus* ribosome, where TJ2 of motif 2AVY-68 is formed correctly, while the other three TJs are replaced by a long double helix (Supplemental Figure 3). The poor conservation of these TJs could be explained by the fact that this part of 16S rRNA is located on the surface of the 30S subunit, is not associated with any important functional centre of the ribosome and does not form contacts with proteins (Schuwirth et al., 2005; Selmer et al., 2006).

Finally, we analyzed the conservation of arrangement 2AVY-1142, whose structure in *E. coli*, as we know, contains only TJ1 and does not contain TJ2. Although the same is true for most eubacteria, in most archaea and in a few hundred bacterial species, the dinucleotide loop C1136-C1137 existing in the *E. coli* arrangement is replaced by a stem-loop structure containing between two and four base pairs. An example of such arrangement found in the 16S rRNA from *Haloarcula Marismortui* (Wuyts et al., 2004) is shown in Figure 10. The addition of this stem-loop to the structure of 2AVY-1142 effectively restores helix H3, thus fitting the whole arrangement to the standard A-DTJ pattern. This, in turn, allows us to consider 2AVY-1142 as a normal A-DTJ motif. The conservation pattern of this motif looks similar to that of 2AVY-68: in

both cases, only one of the two TJs, TJ1 in 2AVY-1142 and TJ2 in 2AVY-68, is conserved, while the other TJ is not. As for the case of 2AVY-68, the low conservation of TJ1 in 2AVY-1142 is directly linked to the fact that the corresponding region of the 16S rRNA is located on the surface of the 30S subunit. There is, however, an essential difference between the two cases: while in 2AVY-68, the absent TJ1 is replaced by a normal double helix, in 2AVY-1142, TJ2 is deleted without any evident compensation.

Thus, in the identified A-DTJs, one can see all four possible patterns of conservation, in which each of the two TJs can be conserved or not. The absence of conservation is always associated with the position of the corresponding TJ on the surface of the 30S subunit. The possibility for such partial conservation indicates that although the two TJs within an A-DTJ arrangement cooperate with each other, the level of inter-dependence is not critical, so that, in principle, each of the two TJs can appear individually, without the other TJ.

4.4 B-DTJs

In addition to the A-DTJs described above, we found DTJ arrangements that did not fit to the A-group. Because A-DTJs were originally defined as those in which helix H2 contained three base pairs, in the alternative DTJs helix H2 should be either longer or shorter than that. In fact, in all such cases, helix H2 contains only two base pairs. The shortening of helix H2 makes impossible the type of stabilization observed in A-DTJs. Correspondingly, the new DTJs use alternative strategies for stabilization of both TJs. Based on the particular scheme of stabilization, we divide all these arrangements in two groups B and C.

The B-group is composed of those DTJs, in which helix H2 contains two base pairs, while 0R belongs to loop L2 (Figure 11). In the available RNA structures, we identified four B-DTJs, the secondary structures of which are shown in Figure 12 (B-E). Motif 2AW4-1902 is found in the 23S rRNA, while the other three motifs are present in the RNase P. Motifs 1U9S-228 and 1U9S-161 exist in the Type-A (Krasilnikov et al., 2004), while motif 1NBS-232 is found in the Type-B of the RNA part of the RNase P (Krasilnikov et al., 2003).

Like in the A-DTJs, nucleotide 0R in the B-DTJs is the first nucleotide of loop L2. Its interaction with base pair [0P;0Q] would also directly stabilize TJ1 and indirectly, TJ2 (Supplemental Figure 4). However, due to the shorter helix H2, nucleotide 0R and base pair [0P;0Q] in B-DTJs cannot be juxtaposed as in the A-DTJs (see Figure 7). In the juxtaposition of 0R and [0P;0Q] that is observed in B-DTJs, they can form only one H-bond between the sugar-edge of 0R and the major groove of [0P;0Q] (Supplemental Figure 4). Although this H-bond is present in all B-DTJs, it does not seem to be stable enough to fix the juxtaposition of helices H1 and H2, as happens in the A-DTJs. It is not surprising, therefore, that the B-DTJs are characterized by the presence of an additional mode of stabilization not observed in the A-DTJs. All B-DTJs contain nucleotide -1R, which is the first nucleotide of loop L1 (Figure 12, B-E). The base of nucleotide -1R interacts with the major groove of base pair [-1P;-1Q], as it is shown in Supplemental Figure 5. At the same time, nucleotide -1R intercalates between nucleotides 0R and -2P and forms an H-bond with the ribose (or phosphate) of nucleotide 0R (Supplemental Figure 6). Thus, the H-bonding with base pair [-1P;-1Q] and stacking to nucleotide -2P enables nucleotide -1R to directly stabilize TJ2, while its interaction with 0R and covalent connection to +1R provides for indirect stabilization of TJ1. The introduction of nucleotide -1R can, therefore, be considered as a compensation of the loss of the stable nucleotide triple formed by nucleotide 0R and base pair [0P;0Q] in the A-DTJs.

Like in A-DTJs, nucleotide +1Q in some B-DTJs can form an H-bond with the minor groove of base pair [+1R;+1P] (Supplemental Figure 7). However, in motifs 1U9S-228 and 1NBS-232 such H-bond cannot be formed. Based on the existence of such cases we can suggest that the stabilization of the TJ1 in the minor groove is not critically important for integrity of B-DTJs and is certainly less important than for integrity of A-DTJs.

4.4.1. Evolutionary conservation of B-DTJs

Out of the four identified B-DTJ motifs, three motifs 2AW4-1902, 1U9S-228 and 1NBS-232 are very conserved among prokaryotes (Table 2). Thus, in all three motifs, base pairs [-2P; -2Q] and [+1R; +1P] have a WC or GU identity in at least 97% of nucleotide sequences, while the level of conservation of adenosine in position 0R varies between 87% and 100%. In motifs 2AW4-1902 and 1NBS-232, position -1R is occupied by cytidine in 100% of the sequences. In motif 1U9S-228, cytidine in this position occurs in only 15% of the sequences. However, in the remaining 85%, this position is occupied by adenosine, which is also suitable for the given structural context (see Supplemental Figure 5). For the forth motif 1U9S-161, we cannot make any valuable conclusion due to the fact that in the compilation of the nucleotide sequences of the Type-A RNAse P (Brown, 1999), the corresponding region of the RNA chain between nucleotides 137 and 161 is poorly aligned.

4.5 C-DTJs

The last group of DTJs, which we call C-DTJs, harbors four motifs (Figure 12). The secondary structure of these motifs is practically identical to that of the B-DTJs. Like in B-DTJs, helix H2 in C-DTJs consists of two base pairs, while the last nucleotide of loop L1 also occupies position +1Q. The major difference between the two groups is that while in B-DTJs, -1R and 0R belong, respectively, to loops L1 and L2, in C-DTJs the situation is opposite: nucleotide 0R is a part of loop L1, and -1R belongs to loop L2. More specifically, 0R in C-DTJs is the first nucleotide of loop L1, while -1R is the first nucleotide of loop L2. A C-DTJ motif can be represented as a B-DTJ in which nucleotides 0R and -1R exchange in their positions (compare Figures 11 (A) and 12 (A)).

Two C-DTJ motifs exist in the prokaryotic ribosome, one in the 16S rRNA, and the other one in the 23S rRNA (Figure 12 B, C) (Schuwirth et al., 2005). One more motif is found in the structure of the mRNA of operon *spc* complexed with the ribosomal protein S8 (Figure 12 D) (Merianos et al., 2004). The last motif is found in the crystal structure of the stem-loop II motif (s2m) of the RNA element in the genome of SARS virus (Figure 12 E) (Robertson et al., 2005).

Unlike in B-DTJs, nucleotides 0R and +1R in C-DTJs are neighbours in the polynucleotide chain and stack to each other. Nucleotide 0R also interacts with the major groove of base pair [0P;0Q] (Supplemental Figure 8). Although this interaction varies from one C-DTJ to another, there is always at least one H-bond between the two moieties. Nucleotides -2P, -1R and -1P, in their turn, are also consecutive in the polynucleotide chain. -1R stacks on -2P and makes a side-by-side contact with -1P (Figure 12). In three C-DTJs, this contact includes the formation of a hydrogen bond, which does not, however, exist in motif 2AW4-2460 (Supplemental Figure 9).

Like in B-DTJs, nucleotides 0R and -1R in C-DTJs form a di-nucleotide stack, which in the case of C-DTJs is additionally stabilized by two hydrogen bonds between the base of 0R and the ribose of -1R and between the base of -1R and the ribose of 0R (Supplemental Figure 10). Similarly to B-DTJs, the intercalation of stack [0R;-1R] between nucleotides +1R and -2P and its interaction with the major groove of helix H2 provide direct and indirect stabilization for both twist-joints TJ1 and TJ2.

In motifs 2AVY-597 and 1S03-37, additional direct stabilization of TJ1 in the minor groove is provided by nucleotide +1Q (Figure 12). In motif L2460 this interaction is somewhat deformed, because the nucleotide that precedes 0Q in the polynucleotide chain and is presumed to play the role of +1Q does not stack to 0Q. Even a stronger deformation happens in motif 1XJR-37, where not only does +1Q not form contacts with helix H1, but also 0Q does not form a base pair with 0P. Despite this deformation, all other nucleotides of motif 1XJR-37 between layers -2 and +1 have about the same positions as in other C-DTJs. These exceptions demonstrate that neither the direct stabilization of TJ1 by nucleotide +1Q, nor the presence of a stable base pair at the zero-level is required for the integrity of a C-DTJ arrangement.

4.5.1. Evolutionary conservation of C-DTJs

Out of four identified C-DTJ motifs, two motifs that are located in the ribosome, 2AVY-597 and 2AW4-2460, are conserved among prokaryotes (Table 3). Thus, in these two motifs, base pairs [+1R;+1P] and [-1P;-1Q] are conserved as WC in more then 95% of nucleotide sequences, while base pair [-2P;-2Q] is mostly conserved as UG or WC in more then 90% of the sequences (Table 3). Base pair [0P;0Q] is conserved as GC and UU in more then 93.7% and 99.8% of sequences of motif 2AVY-597 and 2AW4-2460 respectively. Motif 1XJR-40 is located in the so-called stem-loop II motif (s2m), a 32-nucleotide element immediately preceding the 3' poly-A tail in human astroviruses (Jonassen et al., 1998). The s2m element is the most highly conserved RNA element within the coronaviruses and astroviruses that contain it (Robertson et al., 2005). Analysis of the available genomic sequences of coronaviruses and astroviruses showed that positions, corresponding to nucleotides [+1R;+1P], [-1P;-1Q], [-2P;-2Q], 0R,-1R and 0P of motif 1XJR-40, are 100% conserved (Robertson et al., 2005), which speaks for the high importance of this motif for the virus function.

4.6 Discussion

Analysis of the available RNA tertiary structures showed that on many occasions two double helices stack co-axially one on another in such a way that the first base pair of the second helix is over-twisted in respect to the last base pair of the first helix. We call this type of arrangement a Twist Joint or TJ. Here, we present a new set of RNA motifs consisting of a combination of two closely located TJ arrangements and thus called Double Twist Joint motifs (DTJ). Although the identified structures share some common features, there are distinct differences that allowed us to group them in three motifs: A-, B- and C-DTJ. On the level of secondary structure, all DTJ arrangements can be represented as a double helix containing two bulges that belong to two different strands and are separated from each other by either two (B-DTJ and C-DTJ) or three (A-DTJ) base pairs. Although in WC double helices, bulges are generally considered as a source of local flexibility, their presence in DTJs is a necessary condition for formation of both over-twists. On top of this, the specific interactions in which the bulged nucleotides are involved are essential for stabilization of the particular conformations of the motifs. The stabilization can be direct and indirect. Direct stabilization is mostly made by nitrogen bases, while indirect stabilization is mediated by the RNA backbone. Interestingly, the same bulged nucleotide can provide a direct stabilization of one over-twist and an indirect stabilization of the other over-twist in the same DTJ.

Analysis of different DTJ arrangements allowed us to identify two distinct strategies of DTJ stabilization. The first strategy, which is found in the A-DTJs, implies the direct stabilization of TJ1 through the specific interactions of nucleotides 0R and +1Q with base pairs [0P;0Q] and [+1R;+1P], as well as the indirect stabilization of TJ2 through the covalent attachment of base pair [-3P;-3Q] to nucleotide 0R. Thus, the particular interactions made by nucleotides 0R and +1Q play the critical role in the integrity of the whole arrangement. Because the specificity of the interaction between nucleotide 0R and base pair [0P;0Q] requires that helix H2 contain strictly three base pairs, this strategy is observed only in the A-DTJs.

The second strategy, which is exemplified by the B- and C-DTJs, consists in the stabilization of the DTJ through formation of the four-nucleotide stack [+1R;0R;-1R;-2P]. The existence of such a stack will tune the positions of nucleotides +1R and -2P, thus solidifying the conformations of both over-twists. In this situation, additional interactions of bulged nucleotides with the major and minor grooves of helices H1 and H2, present in all A-DTJs, would become auxiliary. Also, given that the four-nucleotide stack can be formed only when helix H2 contains two base pairs, the specific interaction equivalent to that between 0R and [0P;0Q] in the A-DTJs seems to be impossible.

4.6.1 Different levels of cooperation between TJs in DTJ motifs

Originally, we were interested in studying DTJs because the close position of two TJs could provide a common strategy for stabilization of both of them. To a great extent, our expectations proved true. In each of the three DTJ motifs, we identified nucleotides that are simultaneously involved in stabilization of both TJs. We should admit however, that in the three identified motifs the level of cooperation between the two TJs is different. Thus, in the A-DTJ the inter-dependence between both over-twists is much less pronounced compared to what happens in the B- and C-DTJs. Indeed, in A-DTJ, each of the two TJs can exist regardless of the existence of the other TJ, which was demonstrated by the evolutionary conservation patterns of motifs 2AVY-1142 and 2AVY-68 (2J02-68).

In the B- and C-DTJs, however, the situation is substantially different. Here, the particular mode of stabilization of a DTJ, which consists in the freezing of the juxtaposition of base pairs [+1R;+1P] and [-2P;-2Q] through the formation of the four-nucleotide stack [+1R;0R;-1R;-2P], make TJ1 and TJ2 inseparable from each other. Correspondingly, in all B- and C-DTJs, the evolutionary conservation of which we were able to analyze, both TJs were always present.

4.6.2 Role of DTJs in the long-range interactions

Analysis of the interactions formed by DTJs with their surroundings in the known RNA structures allowed us to make general suggestions concerning the roles of DTJs in

RNA architecture and function. Each of the suggested roles is based on the particular detail of a DTJ conformation.

First, the presence of two consecutive over-twists has a substantial effect on the overall geometry of the regular double helix. The introduction of an over-twist between two base pairs results in the rotation of a half of the double helix around the axis that is about perpendicular to both base pairs. Because in a regular A-RNA conformation, base pairs are inclined with respect to the helix axis by approximately ~20°, the axis of the over-twist rotation will not be co-linear to the axis of the helix. As a result, the presence of an over-twist would result in a bent of the double helix (Supplemental Figure 11). In DTJs, the presence of two closely located over-twists makes the situation more complex. The global effect of the two over-twists on the helix geometry will strongly depend on the distance between the two over-twists, or, more specifically, on the length of helix H2. In particular, in A-DTJs, where helix H2 contains three base pairs, helix H1 inclines with respect to helix H3 for approximately 15° (Supplemental Figure 11A). In B- and C-DTJs, where helix H2 has two base pairs, the angle between H1 and H3 is about twice as big, i.e. is approximately 30° (Supplemental Figure 11B,C). We thus can suggest another possible role of DTJ in the global RNA architecture: the presence of a particular type of DTJ would specifically shape the conformation of the double helix to allow it fit to the given structural context.

Second, the existence of an over-twist between two base pairs opens their surfaces for harboring other nucleotides, like 0R and +1Q. These nucleotides can initiate the formation of longer stacks of nucleotides coming from different parts of the polynucleotide chain. Thus, in motifs 2AVY-68, 2AVY-97, 2AVY-1142, 2AW4-1902 and 1U9S-161 nucleotide +1Q initiates formation of a long stack of nucleotides, which also form hydrogen bonds with the minor groove of H1 (Figure 6B,C; 8; 11B,E). In such a situation, the over-twist would play a role of a nucleus for folding of domains and formation of inter-domain interactions. Thus, the DTJ structure provides opportunities for formation of the long-range interactions between distantly located regions of RNA molecule, consequently playing an important role in the overall RNA folding.

The role of the motif 2AW4-1902 in initiation of the nucleotide stack in the minor groove of H1 is of a particular interest. Here, two adenosines, A1927 and A1928, which

form a tetraloop in the lower part of the Helix 69 are stacked on top of nucleotide +1Q (G1839). In fact, interaction of two adenosines from H69 with the upper part of the motif 2AW4-1902 is the only interaction that this functionally important helix forms with 23S rRNA. Thus, motif 2AW4-1902 alone is mediating the correct positioning of the H69 on the surface of the ribosome.

Finally, analysis of the known DTJs revealed one more role that they can play in the general RNA architecture. As it was discussed above, in different DTJs, loop L2 can contain either one or two nucleotides. In those cases where the second nucleotide exists, it is never involved in interaction with other parts of the motif, but instead it forms contacts with another structural element. For example, in motif 2AW4-1902, adenosine A1900 forms the A-minor interaction with helix 66 of the 23S rRNA, thus contributing to the proper juxtaposition of helices 68 and 66. Even more interesting could be the cases where the element interacting with the second nucleotide of L2 belongs to a different molecule. Thus, in motif 2AVY-1057, found in the 16S rRNA, the second nucleotide of L2 (C1054) coordinates the codon-anticodon mini-duplex in the ribosomal A-site (Jenner et al., 2010). An almost identical situation happens in motif 1NBS-232, where nucleotide A230 of the RNase P, occupying the second position of L2, forms an A-minor interaction with the T-stem of the tRNA. The latter interaction thus mediates the recognition and correct positioning of the pre-tRNA substrate in the active site of the RNase P (Kazantsev et al., 2005).

4.7 Conclusions

DTJ is a new RNA motif that plays several important roles such as shaping of RNA tertiary structure, mediation of the long-range interactions in complex RNA structures as well as formation of the active centers of the ribosome and RNAseP. DTJ-motif represents a combination of two particularly over-twisted base pairs, yet many other types of over-twists can be seen in the available RNA structures, which require further analysis and classification.

4.8 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.

Brown JW. 1999. The Ribonuclease P Database. *Nucleic Acids Res 27*:314.

Jenner LB, Demeshkina N, Yusupova G, Yusupov M. 2010. Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat Struct Mol Biol*.

Jonassen CM, Jonassen TO, Grinde B. 1998. A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J Gen Virol 79 (Pt 4)*:715-718.

Kazantsev AV, Krivenko AA, Harrington DJ, Holbrook SR, Adams PD, Pace NR. 2005. Crystal structure of a bacterial ribonuclease P RNA. *Proc Natl Acad Sci U S A 102*:13392-13397.

Krasilnikov AS, Xiao Y, Pan T, Mondragon A. 2004. Basis for structural diversity in homologous RNAs. *Science 306*:104-107.

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

Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA 7*:499-512.

Lescoute A, Leontis NB, Massire C, Westhof E. 2005. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res 33*:2395-2409.

Merianos HJ, Wang J, Moore PB. 2004. The structure of a ribosomal protein S8/spc operon mRNA complex. *RNA 10*:954-964.

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

Robertson MP, Igel H, Baertsch R, Haussler D, Ares M, Jr., Scott WG. 2005. The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol 3*:e5.

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

Selmer M, Dunham CM, Murphy FVt, 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.

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

4.9 Tables

Table 1. The frequency of the occurrence of particular identities for base pairs and individual nucleotides in different A-DTJ motifs

	Positions	2AVY-1057		2AVY-1142		2AVY-68		2AVY-97	
+1 Layer	[+1R; +1P]	UG UU	99.6(98.0) 0.1(1.6)	CG UA UG	82.7(72.9) 12.7(7.8) 3.7(10.9)	CG UG UA AG	36.7(20.2) 18.5 17.5(8.2) 5.1(3.4)	UA CG AU AG	25.0 15.1(12.4) 13.3 10.2
	+1Q	U G	92.3(0) 7.6(99.8)	G U A C	66.7(66.4) 20.4(8.5) 8.1(8.1) 4.8(17.1)	G U C A	52.7(15.0) 32.3(32.2) 10.6(33.9) 4.4(18.9)	A G U C	46.6(2.2) 24.9(27.4) 16.1(21.0) 12.4(49.5)
0 Layer	[0P; 0Q]	GC GU	98.7(17.9) 0.9(81.9)	GC AU GG	98.6(80) 0.4(12.0) 0(1.6)	GA GC	98.2(14.0) 0.6(78.0)	GC GA CG AA	29.8(27.0) 15.8 12.3(17.1) 7.2
	0R	G	99.7(99.8)	G U	98.2(82.0) 0.7(4.2)	G A	97.7(69.3) 1.8(23.0)	G U A C	58.5(71.7) 21.4(26.1) 11.7(2.2) 8.5
-1 Layer	[-1P; -1Q]	UA CA CG	97.0(99.6) 2.6(0) 0.3(0.2)	UA CG GC AU	75.3(40.0) 22.4(20.0) 0.5(21.4) 0(6.3)	CG GC UA AU	94.5(51.3) 4.1(15.0) 0(20.8) 0(12.2)	UA AG GG AA	22.4 16.4 10.6(32.9) 9.5
-2 Layer	[-2P; -2Q]	AU AC	96.1(96.9) 3.8(2.5)	GC CG UU UA	31.4(14.3) 25.2(34.8) 23.7(4.9) 10.9(13.2)	AU GC GU	57.8(59.6) 40.5(32.9) 0.7	UC GC CG UU	18.9 16.9(61.9) 14.4 6.9
	-2R	C A G	99.7(99.0) 0.2 0(1.0)	N/A		N/A		N/A	
-3 Layer	[-3P; -3Q]	UG UA	99.6(97.8) 0(1.6)	GU AU UU GC	40.9(55.6) 21.2 17.8(9.8) 3.8(16.0)	CG AU UA	96.9(79.8) 1.3 0.9(12.8)	GC UG AU CG	23.4 18.8(6.8) 12.4 12.0(79.9)

Footnote for the Table 1

The numbers represent the frequency of the occurrence of particular nucleotide identity or base pair in a given position of motif among analyzed sequences of 16S rRNA. The numbers were calculated as a percentage of all analysed nucleotide sequences of 16S rRNA. The numbers without and in parentheses pertain to Eubacteria and Archaebacteria, respectively. The nucleotide sequences were taken from the European ribosomal RNA database (Wuyts et al., 2004). The analyzed sequences were current as of May 2007.

Table 2. The frequency of the occurrence of particular identities for base pairs and individual nucleotides in different B-DTJ motifs

	Positions		**1U9S-228**		**1NBS-232**			**2AW4-1902**	
+1 Layer	[+1R;+1P]	GC AU AG UA	86.4 12.9 0.3 0.3		CG UA	85 15		CG	100(100)
	+1Q	A	100		N/A			G A	87(0) 13(100)
0 Layer	[0P; 0Q]	CG UG UA	98.8 0.6 0.6		CG	100		CG UA GC	95.2(56.8) 4.8(0) 0(43.2)
	0R	A C G U	98.5 0.6 0.6 0.3		A	100		A G C	86.7(37.8) 13.3(43.3) 0(13.5)
-1 Layer	[-1P; -1Q]	AU CG GC	92.1 4.7 2.6		AU	100		AU GU	99.8(100) 0.2(0)
	-1R	A C	85.1 14.3		C	100		C	100(100)
-2 Layer	[-2P; -2Q]	UG CG AG	71.2 26.2 2.6		UG UA	97 3		UG CG	100(89.2) 0(10.8)

Footnote for the Table 2

The numbers represent the frequency of the occurrence of particular nucleotide identity or base pair in a given position of motif among analyzed sequences of 23S rRNA and RNA component of RNAseP. The analyzed sequences of motifs 1U9S-228 and 1NBS-232 were taken from the RNAseP Database (Brown, 1999) and were current as of April 2001. For motif 2AW4-1902, the numbers without and in parentheses pertain to 23S rRNA of Eubacteria and Archaebacteria, respectively. The nucleotide sequences of 23S rRNA were taken from the European ribosomal RNA database (Wuyts et al., 2004) and were current as of May 2007.

Table 3. The frequency of the occurrence of particular identities for base pairs and individual nucleotides in different C-DTJ motifs

	Positions	**2AW4-2460**		**2AVY-597**	
+1 Layer	[+1R; +1P]	GC UA CG UG AU	48.6(100) 36.2 6.8 5.1 3.3	AU GC UA GU	83.3(98.1) 6.8 5.2 4.5(1.2)
	+1Q	U C	100(46.0) 0(54.0)	A	99.9(99.7)
0 Layer	[0P; 0Q]	UU GU	99.8 (100) 0.2	GC AC	99.6(93.7) 0(5.8)
	0R	G	100(100)	A U G C	52.0(23.2) 39.8(71.2) 8.2(5.0) 0(0.7)
-1 Layer	[-1P; -1Q]	AU GC GU	98.7 1.3 0(100)	AU CG GC AG UU	79.4(91.1) 17.9 0.8(6.3) 0.7 0(0.9)
	-1R	G U A C	64.4(8.1) 23.1(13.5) 11.3 1.2 (78.4)	A G C	80.3(90.0) 15.6(1.9) 3.1(7.8)
-2 Layer	[-2P; -2Q]	UG AG	99.2**(**100**)** 0.8	UG CG UA AA CA AU	57.7(49.0) 29.5(21.8) 5.7(18.2) 2.5(2.4) 1.3(5.3) 1.1(1.7)

Footnote for the Table 3

The numbers represent the frequency of the occurrence of particular nucleotide identity or base pair in a given position of motif among analyzed sequences of 16S rRNA and 23S rRNA. The numbers without and in parentheses pertain to Eubacteria and Archaebacteria, respectively. The nucleotide sequences were taken from the European ribosomal RNA database (Wuyts et al., 2004) and were current as of May 2007.

4.10 Figures

Figure 1

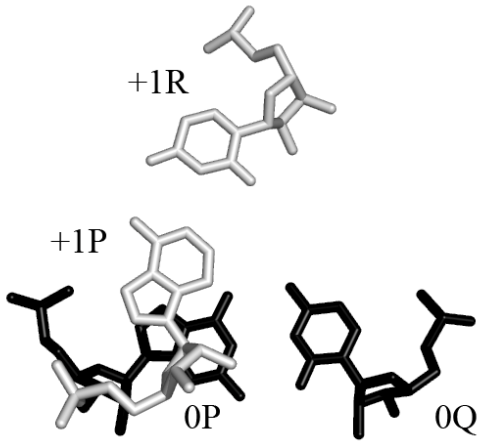


Figure 1. Example of the helical over-twist between two base pairs.

A typical juxtaposition of the contacting base pairs [+1P; +1R] and [0P; 0Q] in a TJ. Nucleotide +1P is partially stacked to 0P and follows it in the polynucleotide chain. Nucleotides 0Q and +1R are not directly connected and are separated from each other.

Figure 2

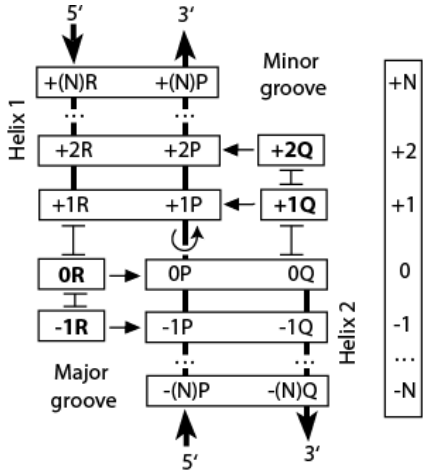


Figure 2. Schematic representation of a TJ arrangement.

Base pairs [+1R; +1P] and [+2R; +2P] make Helix 1 (H1) while base pairs [0P; 0Q] and [-1P; -1Q] make helix 2 (H2). Nucleotide +1Q stacks to 0Q and forms H-bonds with the minor groove of base pair [+1P; +1R]. Nucleotide +2Q, in its turn, stacks to +1Q and interacts with the minor groove of either the [+1P; +1R] or [+2P; +2R] base pair. Nucleotide 0R stacks to +1R and forms H-bonds with the major groove of [0P; 0Q]. Nucleotide -1R stacks on 0R and interacts with the major groove of base pair at a lower level of H2. All individual nucleotides (short horizontally oriented rectangles) and base pairs (long rectangles) are positioned in layers, each layer carrying its own number.

Figure 3

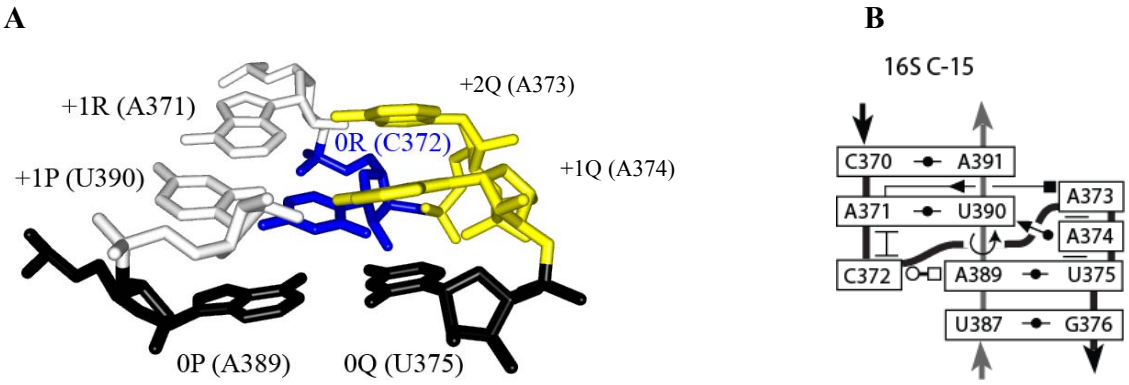


Figure 3. The C-loop motif as an example of the direct stabilization of TJ.

The tertiary (A) and secondary (B) structures of motif 16S C-15 (Lescoute et al., 2005) from the *E. coli* 16S rRNA (PDB: 2AW4). Nucleotide 0R(C372) stacks to +1R(A371) and forms H-bonds with the major groove of base pair [0P;0Q] (A389-U375). Nucleotides +1Q (A374) and +2Q (A375) stack to each other and to 0Q. They form H-bonds in the minor groove of Helix 1 with +1P (U390) and +1R (A371), respectively. Standard annotation of the non-canonical base pairs proposed by Leontis and Westhof (Leontis & Westhof, 2001) is used to demonstrate the inter-nucleotide H-bonding.

"⌶" symbol or a short horizontal line is used to demonstrate the stacking interaction between nucleotides.

Figure 4

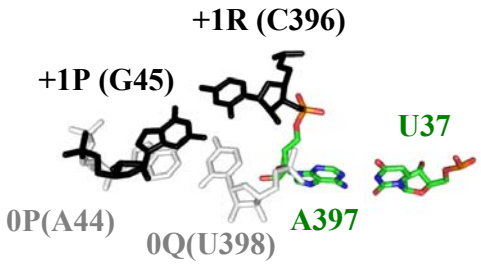


Figure 4. An example of the indirect stabilization of TJ.

An example of the indirect stabilization of TJ found in the *E. coli* 16S rRNA (PDB: 2AVY), named 2AVY-44 according to the pdb identifier of the crystal structure followed by the number of the nucleotide in position 0P. The contacting base pairs that form the over-twist are shown as in Figure 1. The fixation of position of A397 due to the formation base pair with U37 located in another part of the 16S rRNA structure would stabilize the positions of the flanking nucleotides +1R (C396) and 0Q (U398), thus indirectly stabilizing the over-twist.

Figure 5

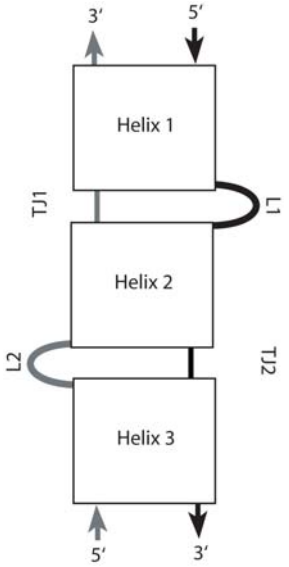


Figure 5. The general scheme of the DTJ arrangement.

A DTJ arrangement is composed of three double helices. To be qualified as such, each of the two Helices 1 and 3 should contain at least one WC or GU base pair. The central part of a DTJ arrangement consists of a double helix (Helix 2) made of either 2 or 3 WC base pairs. At both ends, this double helix is flanked by two over-twists, which occur in the opposite strands. In all DTJs, one over-twist has higher amplitude than the other one, which allows us to name them TJ1 and TJ2, respectively.

Figure 6

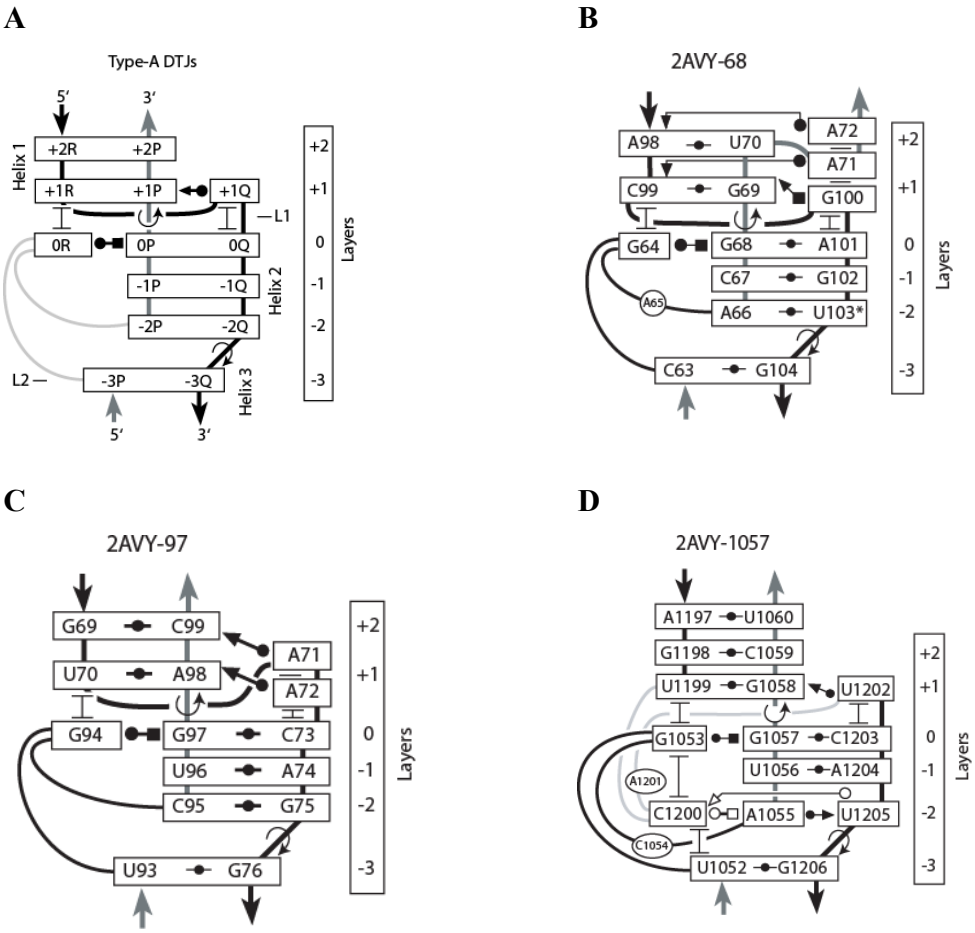


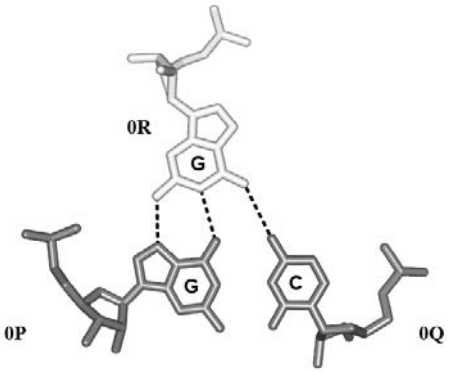
Figure 6. Secondary structure of A-DTJ motifs.

(A). The general scheme of the A-DTJ motif. TJ1 and TJ2 occur between base pairs [0P; 0Q] and [+1R; +1P] and between base pairs [-2Q; -2P] and [-3Q; -3P], respectively.

(B-D). The secondary structures of three A-DTJ motifs, found in the crystal structure of the *E. coli* 16S rRNA (Schuwirth et al., 2005). Each arrangement is named by the pdb identifier of the crystal structure followed by the number of the nucleotide in position 0P.

Figure 7

A



B

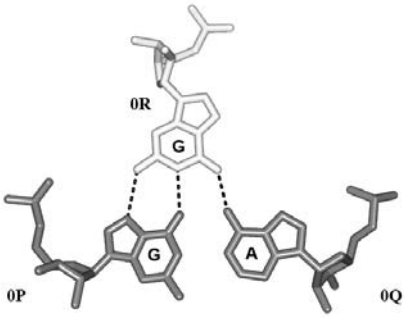


Figure 7. The structure of nucleotide triple [0R; 0P; 0Q] in A-DTJs.
In all three A-DTJs, guanosine 0R interacts with the major groove of base pair [0P; 0Q]. In motifs 2AVY-97 and 2AVY-1057, base pair [0P; 0Q] is GC (**A**), while in motif 2AVY-68 it is GA (**B**). The hydrogen bonds formed by guanosine 0R are shown by dotted lines.

Figure 8

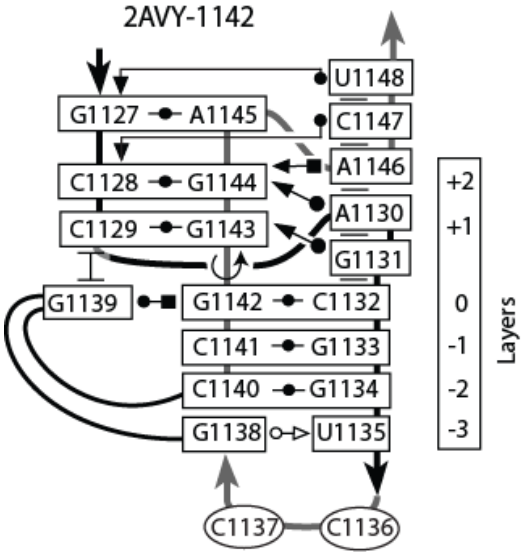


Figure 8. Quasi-A-DTJ motif 2AVY-1142.

Motif 2AVY-1142 has the same structure as all other A-DTJs except it does not contain Helix 3. Instead of Helix 3, this motif contains a four-nucleotide loop 1135-1138, the first and last nucleotides of which form a trans-WC-SE base pair [-3P; -3Q].

Figure 9

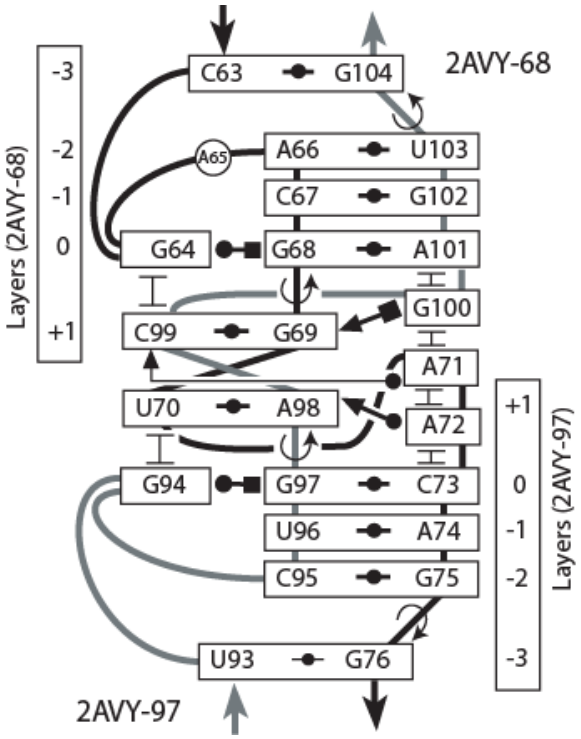


Figure 9. The common arrangement of A-DTJs 2AVY-68 and 2AVY-97.

Two motifs 2AVY-68 and 2AVY-97 are located close to each other and form a common arrangement. Within this arrangement, the two motifs are oriented head-to-head. The double helix composed of base pairs G69-C99 and U70-A98 is Helix 2 for each of the two motifs. Nucleotides A72, A71 and G100 form a continuous stack. This stack interacts with the minor groove of Helix 2, thus stabilizing the whole arrangement.

Figure 10

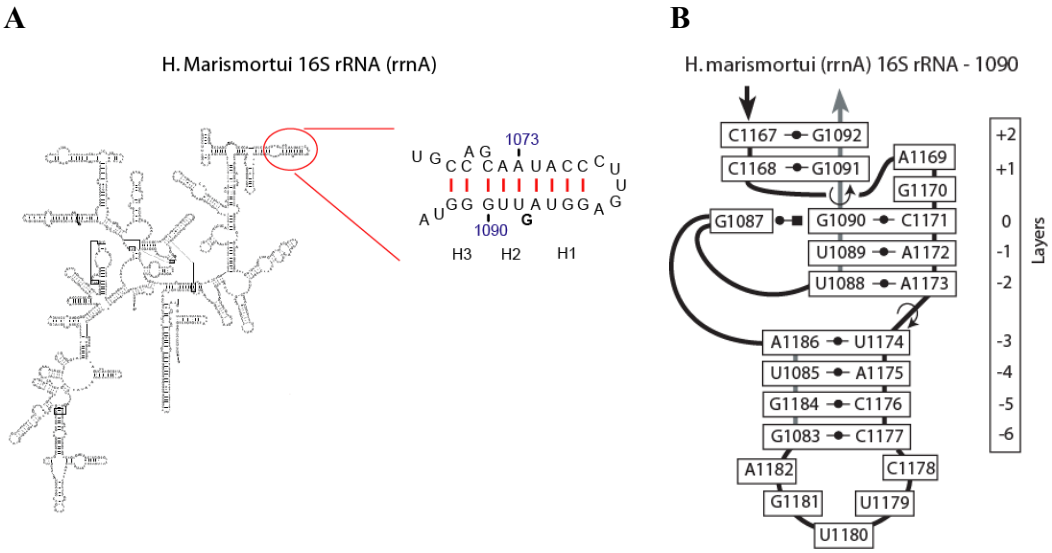


Figure 10. The reconstruction of the complete A-DTJ motif 2AVY-1142 in an archaeal nucleotide sequence of the *H. marismortui* 16S rRNA.

(A) The region of the *H. marismortui* 16S rRNA (sequence of rrnA operon) corresponding to arrangement 2AVY-1142 (Wuyts et al., 2004). (B) The secondary structure of the same region folded in the way fitting to the pattern observed in the A-DTJs. In our standard nomenclature of naming DTJ arrangements, we refer to this homolog of 2AVY-1142 as "*H. marismortui* (rrnA) 16S rRNA – 1090".

Figure 11

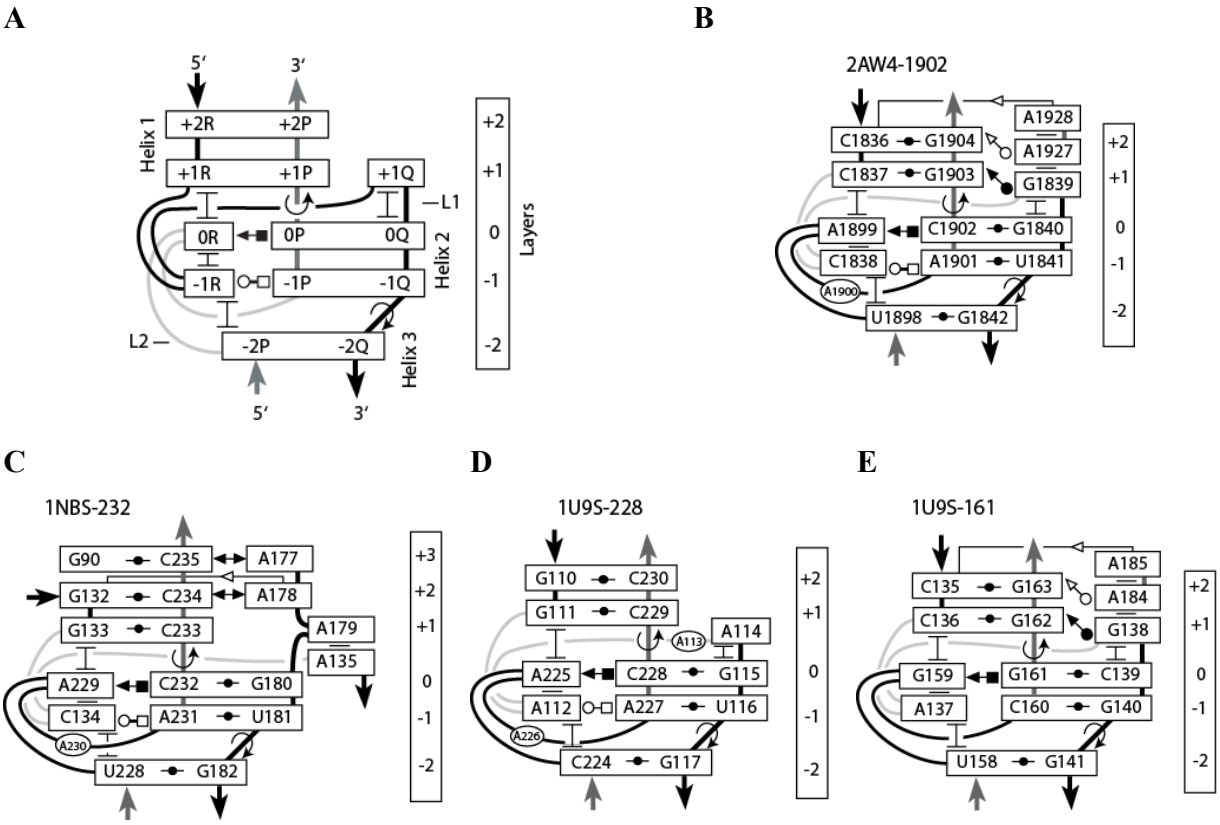


Figure 11. The secondary structures of B-DTJ motifs.

(A) The general scheme of the B-DTJ motif. Nucleotides [+2P; +2R], [+1P; +1R], [0P; 0Q], [-1P; -1Q] form WC base pairs while [-2P; -2Q] can be either WC or GU base pair. Nucleotides 0R and 0P form cis-SE-HG base pair. Nucleotide +1Q, if present, forms either cis-HG-SE or cis-WC-SE base pair with nucleotide +1P. Nucleotide -1R may form trans-HG-WC base pair with -1P. The TJ1 is formed between [0P; 0Q] and [+1R; +1P] base pairs, while TJ2 is formed between [-1Q; -1P] and [-2Q; -2P] base pairs.

(B-E) Each arrangement is named by the pdb identifier of the crystal structure followed by the number of the nucleotide in position 0P. Motifs 1U9S-161 and 1U9S-228 exist in the crystal structure of RNAseP of *Thermus thermophilus* (Krasilnikov et al., 2004), motif 1NBS-232 exists in the crystal structure of RNAseP of *Bacillus Subtilis* (Krasilnikov et al., 2003), while motifs 2AW4-1902 is found in 50S ribosomal subunit of *Escherichia coli* (Schuwirth et al., 2005).

Figure 12

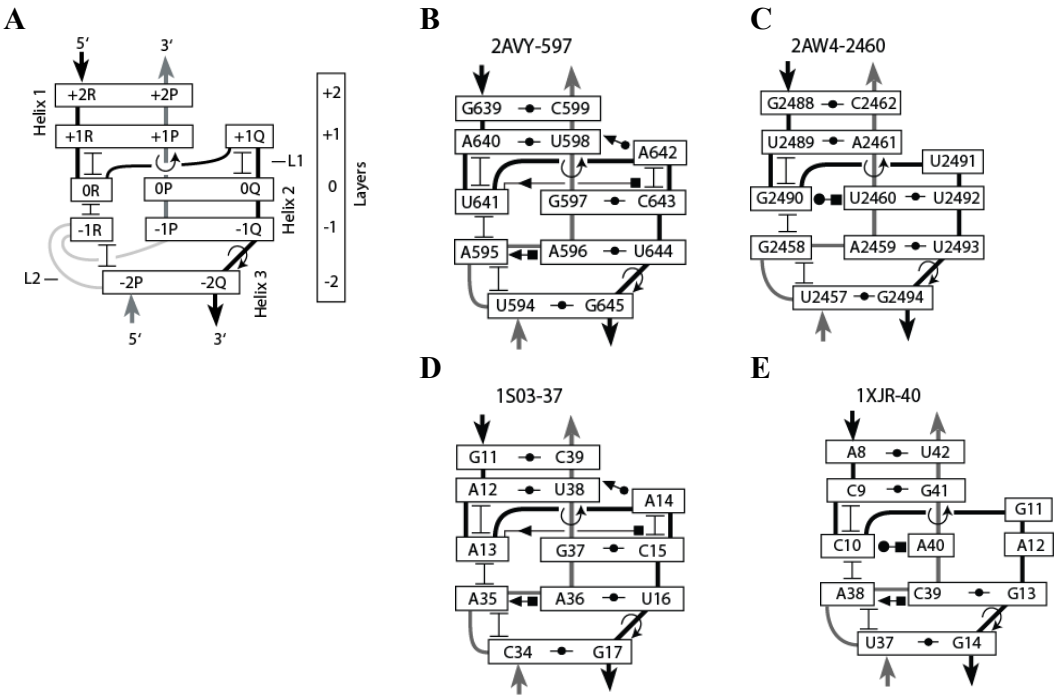


Figure 12. The secondary structures of C-DTJ motifs.

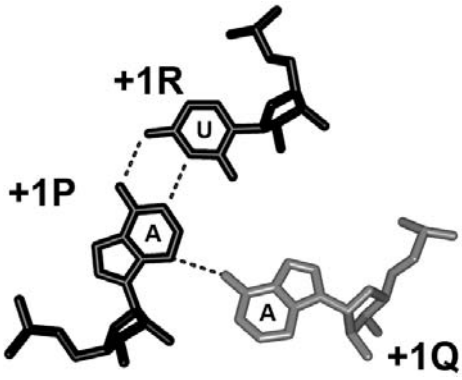
(A) The general scheme of the C-DTJ motif. Nucleotides [+2P; +2R], [+1P; +1R], [-1P; -1Q] form WC base pairs while [-2P; -2Q] can be either WC or GU base pair. Nucleotides 0R hydrogen bonds with either 0P or 0Q. Nucleotide +1Q, if present, forms cis-WC-SE base pair with nucleotide +1P. Nucleotide -1R may form trans-HG-WC base pair with -1P. Analogously to B-DTJs, TJ1 is formed between [0P; 0Q] and [+1R; +1P] base pairs, while TJ2 is formed between [-1Q; -1P] and [-2Q; -2P] base pairs.

(B-E) Each arrangement is named by the pdb identifier of the crystal structure followed by the number of the nucleotide in position 0P. Motifs 2AVY-597 and 2AW4-2460 are identified in 30S and 50S ribosomal subunits of *Escherichia coli* respectively (Schuwirth et al., 2005), motif 1S03-37 is found in the crystal structure of the spc operon mRNA of *Escherichia coli* (Merianos et al., 2004), while 1XJR-40 exists in the crystal structure of the stem-loop II motif (s2m) RNA element of the SARS virus genome (Robertson et al., 2005).

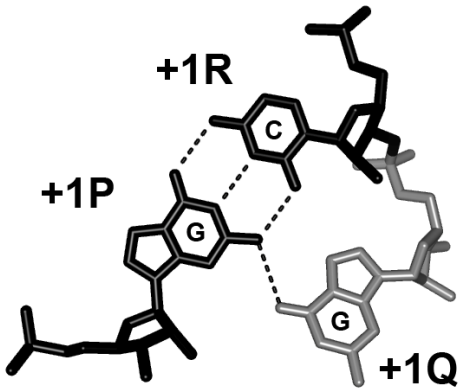
4.11 Supplemental figures

Supplemental Figure 1.

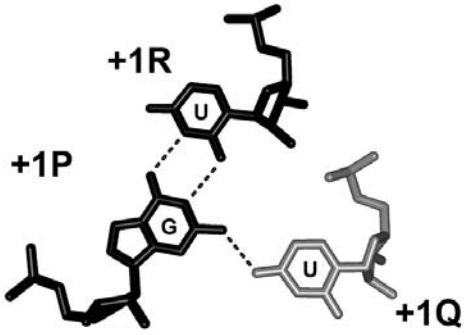
2AVY-97



2AVY-68



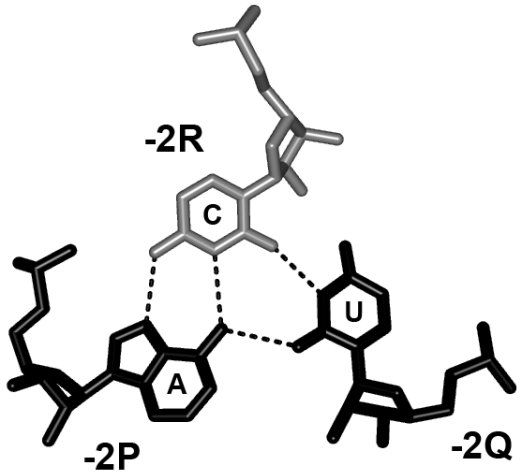
2AVY-1057



Supplemental figure 1. The interaction of nucleotides +1Q and +1P in A-DTJs.

In all cases of A-DTJ motif, +1Q interacts with the minor groove of base pair [+1P;+1R], forming either *cis* WC/SE (2AVY-97) or *cis* HG/SE (2AVY-68 and 2AVY-1057) base pair with +1P.

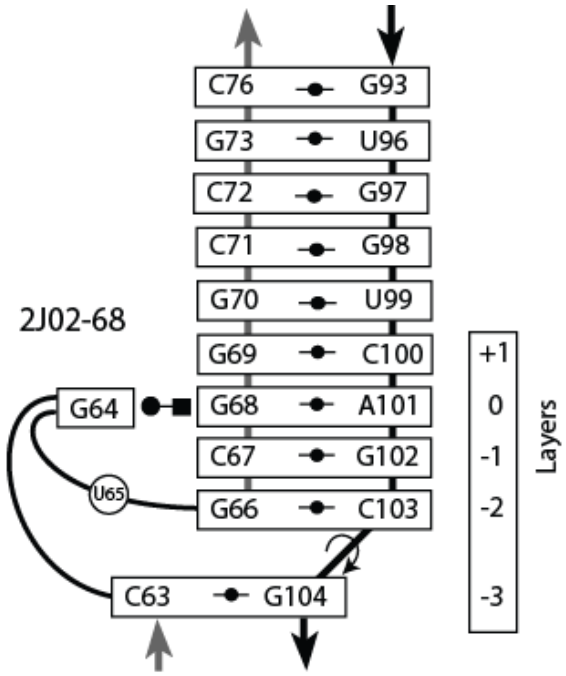
Supplemental Figure 2.



Supplemental Figure 2. Deformed base pair [-2P; -2Q] in motif 2AVY-1057.

Nucleotide -2Q (uridine) does not form a canonical WC base pair with -2P (adenosine), instead, the presence of the nucleotide -2R (C) allows formation of a triple [-2P;-2Q;-2R], where -2Q is significantly shifted in the major groove.

Supplemental Figure 3.



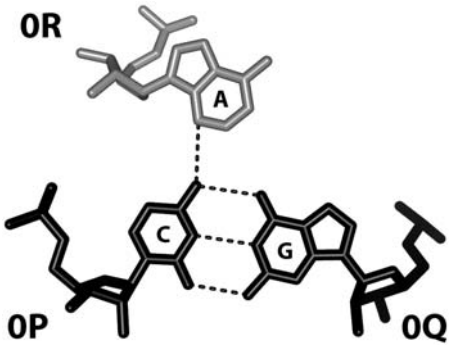
Supplemental Figure 3. Secondary structure of homologous arrangement 2AVY-68 - 2AVY-97 from the *T. thermophilus* 30S subunit.

In the 16S rRNA of *T. thermophilus* out of four TJ arrangements of the DTJ conglomerate 2AVY-68-2AVY-97 only TJ1 of 2AVY-68 (2J02-68) is formed.

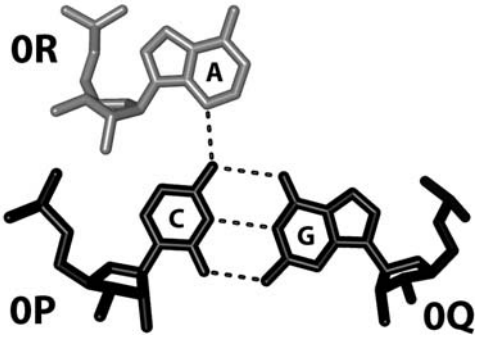
Both TJs of 2AVY-97 and TJ1 of 2AVY-68 are replaced with a long regular double helix. In *T. thermophilus*, formation of the same number of consecutively stacked WC (GU) base pairs in between positions [G93;C76] and [G68;C101] (which exactly correspond to positions [U93; G76] and [G68;C101] in *E. coli*) is due to the elimination of two nucleotides from each strand, resulting in the absence of unpaired regions L1 and L2 in 2AVY-97 and L1 in 2AVY-68, observed in *E. coli*.

Supplemental Figure 4.

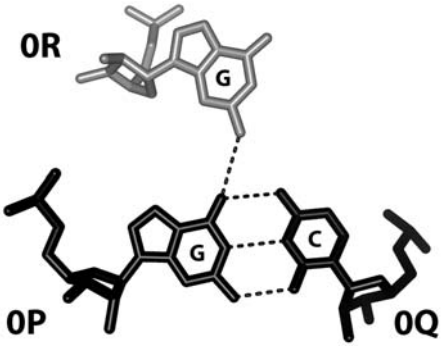
2AW4-1902 (1NBS-232)



1U9S-228



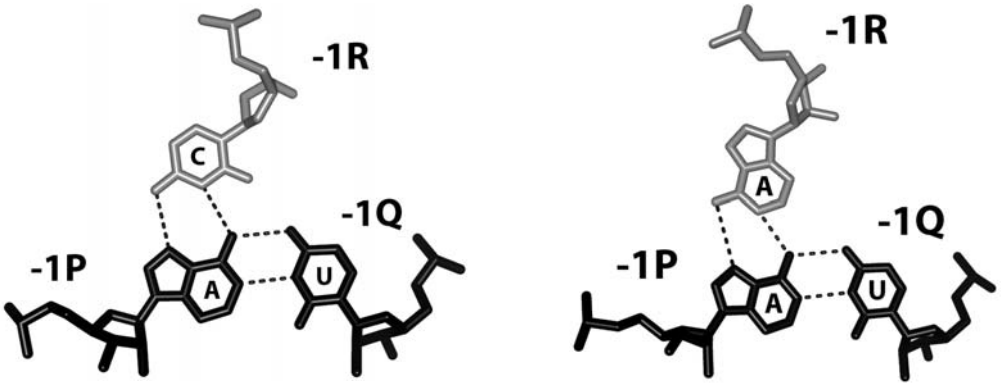
1U9S-161



Supplemental Figure 4. Interaction of nucleotide 0R with major groove of base pair [0P; 0Q] in B-DTJs. Base pairing of nucleotides at the 0 layer of the motifs is shown in case of 2AW4-1902 (1NBS-232), 1U9S-228 and 1U9S-161. The hydrogen bonds involving 0R and 0P are shown by dotted lines.

Supplemental Figure 5.

2AW4-1902 (1NBS-232) 1U9S-228

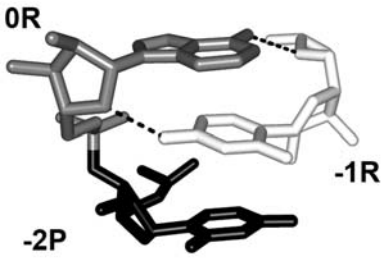


Supplemental Figure 5. Interaction of nucleotide -1R with -1P in B-DTJs.

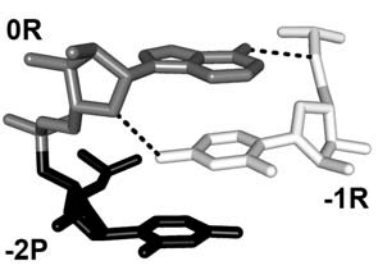
In all B-DTJs, nucleotide -1R is either adenosine or cytosine. Base pairing of nucleotides at the -1 layer of the motifs is shown in case of 2AW4-1902 (1NBS-232) and 1U9S-228. The hydrogen bonds involving -1R are shown by dotted lines.

Supplemental Figure 6.

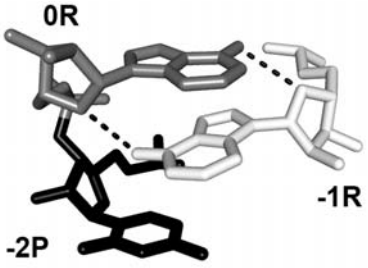
2AW4-1902



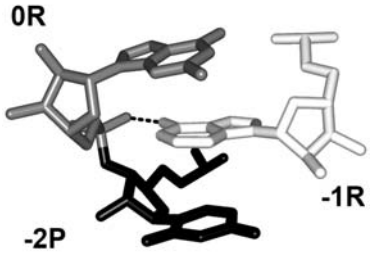
1NBS-232



1U9S-228



1U9S-161

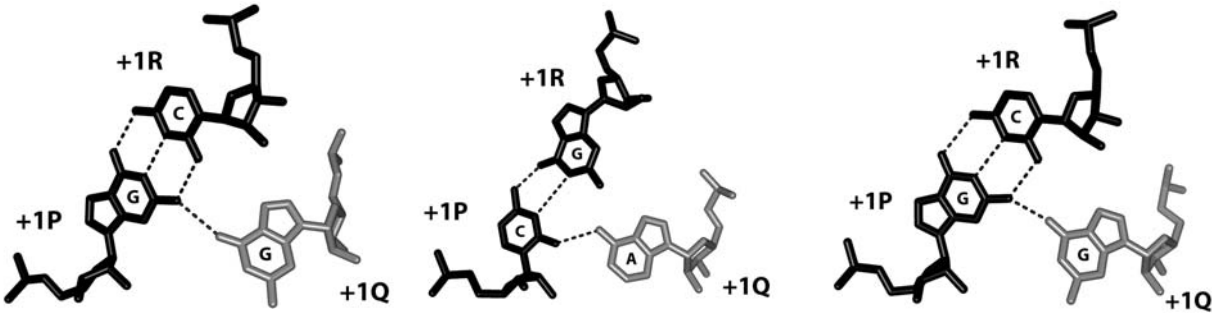


Supplemental Figure 6. Additional base-backbone hydrogen bonds found in the dinucleotide stack [0R; -1R] of the B-DTJ motifs.

The H-bond between the base of -1R and the backbone of 0R exists in all available B-DTJ structures, while the opposite bond between the base of 0R and the backbone -1R is present in all structures except 1U9S-161.

Supplemental Figure 7.

2AW4-1902 **2A2E-200 (1U9S-228)** **1U9S-161**

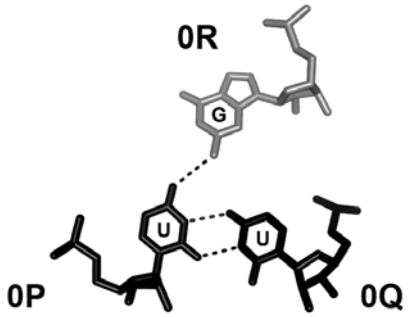


Supplemental figure 7. The interaction of nucleotides +1Q and +1P in B-DTJs.

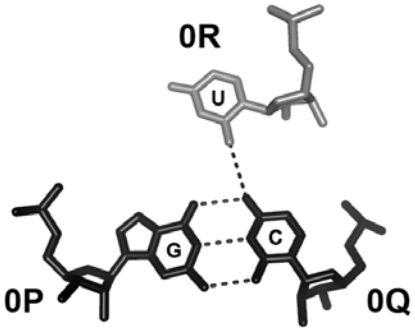
In most of the identified B-DTJ motifs, +1Q interacts with the minor groove of base pair [+1P;+1R], forming *cis* WC/SE base pair with +1P.

Supplemental Figure 8.

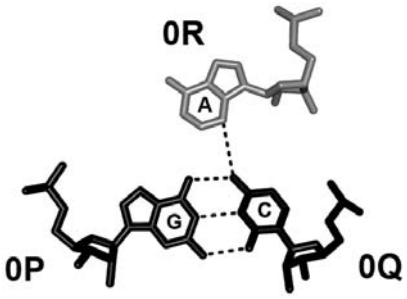
2AW4-2460



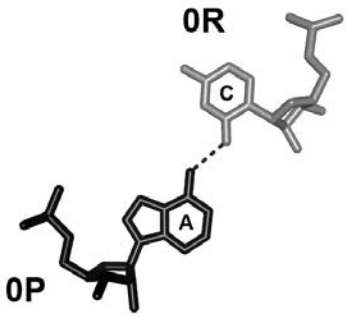
2AVY-597



1S03-37



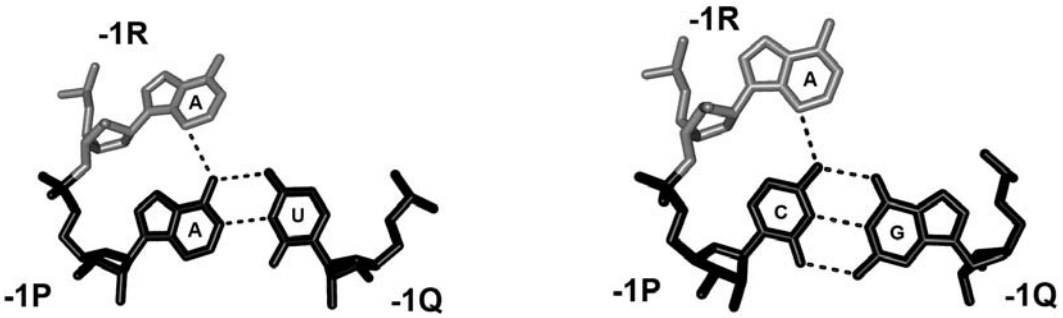
1XJR-40



Supplemental figure 8. Interaction of nucleotide 0R with major groove of base pair [0P; 0Q] in C-DTJs. Nucleotide 0R makes cis-SE-WC base pair with 0P in motifs 2AW4-2460 and 1XJR-40; 0R makes cis-SE-HG base pair with nucleotide 0Q in motifs 2AVY-597 and 1S03-37.

Supplemental Figure 9.

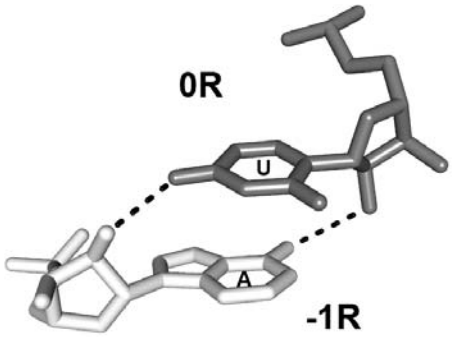
2AVY-597 (1S03-37) **1XJR-40**



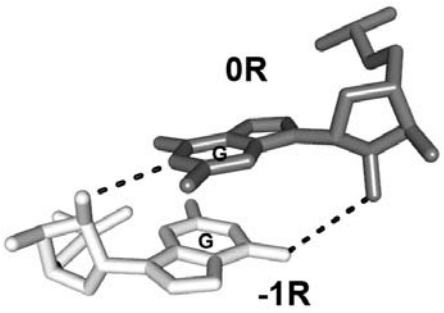
Supplemental figure 9. Interaction of nucleotide -1R with major groove of base pair [-1P; -1Q] in C-DTJs. cis-SE-HG base pair is established between -1R and -1P nucleotides in all C-DTJ motifs except 2AW4-2460.

Supplemental Figure 10

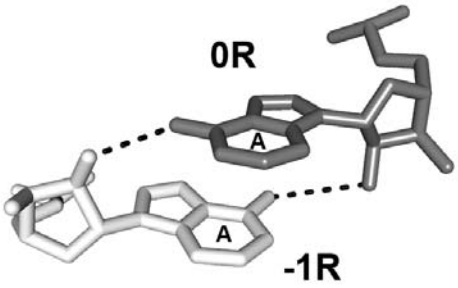
2AVY-597



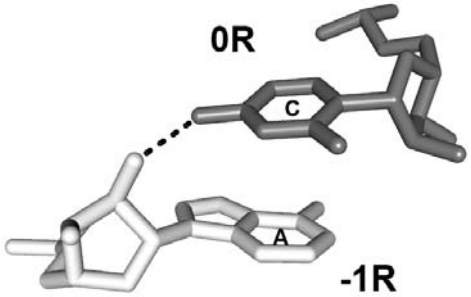
2AW4-2460



1S03-37



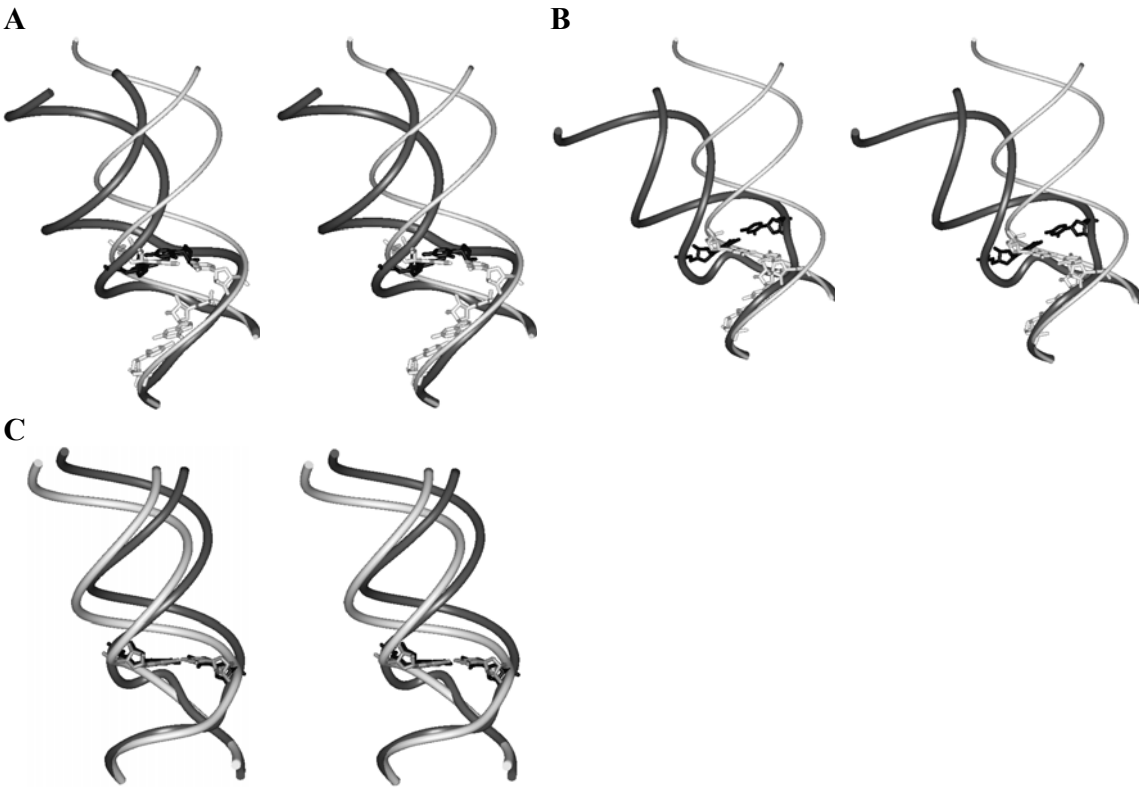
1XJR-40



Supplemental Figure 10. Additional base-backbone hydrogen bonds found in the dinucleotide stack [0R;-1R] of the C-DTJ motifs.

The H-bond between the base of 0R and the backbone of -1R exists in all available C-DTJ structures, while the opposite bond between the base of -1R and the backbone 0R is present in all structures except 1XJR-40.

Supplemental Figure 11



Supplemental Figure 11. Bending of the RNA double helix due to the presence of a DTJ motif.

(A) Comparison of an A-DTJ motif (2AVY-97, shown in black ribbon) with a regular A-form RNA double helix (shown in white ribbon). Only nucleotides at the +1 layer are shown for A-DTJ (in black) and nucleotides that correspond to those at the -3 and +1 layers are shown in white in A-RNA helix. Two structures are superposed using nucleotides of -3 layer of A-DTJ and corresponding layer in the A-form helix. Helices H1 and H3 of A-DTJ motifs are inclined by approximately 15°.

(B) Comparison of a B-DTJ motif (2AW4-1902, shown in black ribbon) with a regular A-form RNA double helix (shown in white ribbon). Only nucleotides at the +1 layer are shown for B-DTJ (in black) and nucleotides that correspond to those at the -2 and +1 layers are shown in white in A-RNA helix. Two structures are superposed using nucleotides of -2 layer of B-DTJ and corresponding layer in the A-form helix. The average angle between helices H1 and H3 in B-DTJ motifs is approximately 30°.

(C) Comparison of a B-DTJ motif (2AW4-1902, shown in black ribbon) with a C-DTJ motif (2AVY-597, shown in white ribbon). **O**nly nucleotides at the +1 layer of both motifs are shown (in black and gray for 2AW4-1902 and 2AVY-597 respectively). Two structures are superposed using nucleotides of -2 layer of both motifs. The presence of both, B- or C-DTJs results in practically similar shaping of the RNA double helix.

Chapter 5

Article 4

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

5. 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, QC, Canada.

Nucleic Acids Res. 2010 Published online on February 5th.

[1] Both authors contributed equally to this work

Running title: RNA-protein interactions in the ribosome

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

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

5.2 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 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 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 & Steinberg, 2002; Mokdad et al., 2006).

5.3 Materials and methods

5.3.1 Bacterial strains and media

For all 30S subunit related procedures, we used the *Escherichia 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 & 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).

5.3.2 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.

5.3.3 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 Supplementary Methods and Supplementary Table S3. Transformation of the plasmids harboring the combinatorial 23S rRNA gene libraries into the SQ380 cells was performed by electroporation.

5.3.4 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.

5.3.5 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 the log-plot, from which the doubling time was deduced by a linear approximation.

5.3.6 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.

5.3.7 Molecular dynamics simulations

Molecular dynamics (MD) simulations were carried out on four different constructs, each composed of two double helices forming together the AGPM (Supplementary Figure S1). 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 for the central base pairs, which 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 consisted in the fixation of the positions of the nucleotides forming the tetraloops. 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 (Supplementary Figure S1). 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, in which hydrogen bonds remained unrestrained (Supplementary Figure S1a). The constraints were introduced as penalty $K \times (R-3.3)^2$ 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.

5.4 Results

5.4.1 Background: general description of AGPM

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* (Figure 1) (Gagnon & Steinberg, 2002). Due to the frequent occurrence, 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 compact 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 AGPM in different structural environments is essential for understanding how the ribosome structure forms and functions.

Within 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 (Figure 1) (Gagnon & 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 & 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. In this arrangement, the internal and external nucleotides are responsible, respectively, for about 70% and 30% of all inter-helix atom-atom 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 (Figure 2b). This case seems more of an exception, because in most organisms, including *E. 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.

5.4.2 The motifs studied

In this study, we consider three AGPMs: S296, L639 and L657 (Figure 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 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). Unlike in S296, in motifs L639 and L657 the two interacting double 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 (Figure 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 motif L657 with L4 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 (Figure 4b). Again, it is not the bases, but the sugar-phosphate backbones of nucleotides 650-651 that form contacts with 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.

5.4.3 Cloning and selection of 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 & de Boer, 1987; Lee et al., 1996; Belanger et al., 2004; Abdi & Fredrick, 2005; Rackham & 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 protein chloramphenicol acetyl-transferase (CAT). The quantification of the efficiency of the selected clones was made through the measurement of the activity of the green fluorescence protein (GFP). Both proteins, CAT and GFP, were synthesized from mRNAs containing the modified Shine-Dalgarno sequence (Gagnon et al., 2006). For selection of functional variants of motifs L639 and L657 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. The efficiency of 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 1. For convenience, the names of the selected variants of motifs S296, L639 and L657 start with letters A, B and C, respectively.

5.4.4 Analysis of the selected clones: the minimal requirement for the integrity of 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 1). 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 1). 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 1, the majority of selected clones contained such non-standard nucleotide combinations as UU, CU, UC, CC, UG, CA, AC, GG, GA and AG. Although the A-clones harboring these 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 the description of the cloning and selection). Similarly, even though the doubling time of the selected B- and C-clones containing non-standard nucleotide combinations was generally somewhat longer than that of the wild-type (Table 1), 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 any of the 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, almost all 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 (Figure 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 (Figure 2c). As a result, all nucleotides of the AGPM will stay at their standard positions except the external central nucleotide of the second helix. The latter nucleotide could accommodate to this structure through the formation of an alternative base pair with the opposite nucleotide or, if the accommodation is impossible, it will always be able to bulge out. We thus conclude that the presence of a WC or GU base pair in one of the two helices will always provide the possibility for all nucleotides of both helices except the external central nucleotide of the second helix to stay in their standard positions. A potential loss of the contacts formed by the latter nucleotide will thus

constitute the maximal possible destabilizing effect associated with the presence of an alternative dinucleotide combination in one of the two helices.

Based on the fact that in all three libraries almost all selected clones share the same ability to form at least one central base pair, we suggest that the existence of a WC or GU base pair, and, correspondingly, the possibility to form a nucleotide triple, represents a minimally acceptable condition for the integrity of AGPM. Henceforth, the central base pair that is able to form a nucleotide triple with the internal nucleotide of the opposite helix will be called structure-forming base pair. The only two exceptional clones A11 and B16 that do not contain such a base pair (Table 1) will be discussed later.

5.4.5 Alternative dinucleotide combinations

Among the dinucleotide combinations that can play the role of a structure-forming base pairs, only for GU, the inverted combination UG cannot serve this function. The difference between GU and UG becomes obvious if one compares the dinucleotide combinations that have been co-selected with each of them (Table 1). While UG has been selected together only with WC and GU, for GU, in addition to these two, on can find combinations CC, UU, UC, GA, AG, CA and UG. We can conclude that UG imposes essentially tighter restrictions on the identity of the opposite central base pair than GU. This difference is understandable if one assumes that GU is a structure-forming base pair, while UG is not and, therefore, requires that the opposite base pair be such.

To explain the asymmetry between the GU and UG, one should take into account that in both base pairs, compared to WC, U and G are displaced in the major and minor grooves, respectively. While in GU, such displacement provides for the close packing with the opposite helix (Figure 2a), in UG, the direction of the nucleotide displacement is opposite to that required for the comfortable interaction of the two helices. The formation of UG would thus be detrimental for the helix packing. Whether this base pair still exists in AGPM in spite of its potentially destabilizing effect on the interaction with the opposite helix is unknown. However, it is clear that if the benefits provided by the existence of UG in AGPM do not exceed the energy cost associated with its maintenance, the base pair will not form. The absence of this base pair would leave the internal G in its optimal position for formation of the nucleotide triple and will allow the external U to

bulge out. The bulging of U thus corresponds to the maximal possible energy cost associated with the accommodation of UG to the contact with the opposite helix within AGPM.

Other alternative dinucleotide combinations that are found in the selected clones include CU, UC, CC, UU, CA, AG, GA and GG. At least some of these combinations can form base pairs within a double helix. However, the fact that these combinations are selected together with WC or GU strongly suggests that even if they form a base pair, its stabilizing impact will be insufficient to guarantee the proper folding and the proper arrangement of the two helices. In other words, these dinucleotide combinations will be unable to serve as structure-forming base pairs and thus will require that such a base pair be present in the opposite helix. Even if an alternative base pair can fit to the double helical geometry, its accommodation to the inter-helix interaction could face problems. However, like in the discussed above case of UG, there will always be a possibility for the external nucleotide of this combination to bulge out, thus allowing the internal nucleotide to fit to its optimal position. Given that in about 75% of all alternative dinucleotide combinations found in the selected clones the external nucleotide is a pyrimidine (Table 1), the energy cost associated with the existence of such a bulge would usually be relatively modest.

5.4.6 Molecular dynamics simulations

To test the ability of the nucleotide triple to stabilize the structure of AGPM, we performed molecular dynamics (MD) simulations on specially modeled AGPM constructs (Supplementary Figure S1). 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.

In the first part of the study, we tested the behavior of four complexes, in which the central base pairs were GU-CG, GU-UG, GC-UG and UG-UG. In all these 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 Figure 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 5 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 (Figure 5a). In the cases of combinations GU-UG and GC-UG, the break took about 500 and 300 ps, respectively, (Figure 5b and c), while for combination UG-UG the break occurred within the first 10 ps (Figure 5d). 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 Figure 2). 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.

In the second part of the study, we tested the behavior of the GU-UG complex in which both dinucleotide combinations GU and UG were initially arranged as base pairs. In total, we made three simulations. In the first of them, the UG base pair broke within the first 100 ps of the simulation, after which the complex behaved similarly to the GU-UG complex in the previous simulations (Figure 5b). In the second simulation, the UG-containing helix bent over its axis, which made the UG base pair detached from the opposite helix (Supplementary Figure S2). After about 500 ps of staying in such conformation, the UG-containing helix returned to its initial shape. In the third case, the UG base pair soon after the beginning of the simulation shifted as a whole out of the inter-helix contact zone, yielding its place to the next base pair C_{12}-G_3 (Supplementary Figures S2a and S3). This shift made the two helices closely packed; such arrangement remained stable until the end of the simulation. The results of all these simulations demonstrate that the presence of UG in one of the two helices destabilizes the AGPM structure, pushing for the exclusion of UG from the inter-helix contact zone. The exclusion can be achieved through breaking of the UG base pair (the first simulation), deformation of the UG-containing helix (the second simulation) or displacement of one helix with respect to the other (the third simulation). We thus can conclude that the requirements for incorporation of a non-standard base pair into a double helix can be different depending on whether the helix stays alone or makes a part of AGPM. For an isolated double helix, there is no difference between GU and UG, while for a helix within AGPM, GU is clearly more favorable than UG. The embedment of AGPM in the ribosome structure is expected to provide additional constraints on the motif's conformation. Due to the involvement of both helices of AGPM in multiple interactions with different parts of the ribosome structure, bending of the helices or their displacement with respect to each other, which were observed in the second and third simulations, seem to be less probable than the bulging of a single nucleotide.

5.4.7 The symmetry of the central base pairs in motif S296

In the A-clones following the GU-WC pattern, helices h3 and h12 harbor base pairs GU and WC with the same frequency (Table 1). 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 AGPM.

5.4.8 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 1). 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 (Figure 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 1).

Among B-clones there are two exceptions B3 and B6, in which the structure-forming base pair, is found in helix H31 instead of 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 close to that existing in a WC base pair (Figure 6a). Similar arrangements of A and G has also been observed on other occasions (Webster et al., 1990). The formation of such GA base pair could 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.

5.4.9 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 almost always GU or WC, while alternative dinucleotide combinations are found exclusively in helix H28 (Table 1). The only exceptional clone C84 will be discussed later. 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 (Figure 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: 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 1).

In the following analysis we argue 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, nucleotide 600, whose ribose forms a direct contact with the ribosomal protein L4, belongs to a central base pair. Presuming that the interaction between the ribose-600 and protein L4 is critical for the ribosome function, we would expect that in all C-clones this ribose occupies about the same place. For analysis, we divide all C-clones in two groups, I and II, as shown in Table 1. Group I harbors all clones following the GU-WC pattern, while all other clones fall into Group II. Group II thus contains only five clones and in all of them, base pair 600-657 is GU.

At the first step, we checked if the position of the ribose-600 is insensitive to the GU⇔WC exchange. For this, we superposed the structures of motif L657 existing in the *E. coli* (Schuwirth et al., 2005) and *H. marismortui* (Ban et al., 2000) ribosomes. Compared to *E. coli*, in the *H. marismortui* ribosome the GU and WC base pairs have exchanged in their positions (Figure 3b). The superposition of the two structures (Figure 7b) demonstrates that after such exchange, 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. The same result was obtained when the structure of the *E. coli* motif L657 was superposed with its own image rotated for 180° (not shown). This *in silico* experiment confirms that, indeed, in all Group I clones the ribose-600 maintains the same position regardless of which of the two helices harbor the GU and WC base pairs.

The insensitivity of the ribose-600 position to the GU⇔WC exchange appears to be caused by the interaction of helix H27 with the opposite helix H28. If helix H28 did not exist, the GU⇔WC replacement in helix H27 would have resulted in a substantially larger movement of the ribose-600, as seen in Figure 7a. Such movement would have included the rotation of the ribose-600 by about 15º, leading to the displacement of its atoms by at least 1 Å. However, within the AGPM, nucleotides 600 and 657 of helix H27 can be displaced only as far as it does not interfere with the position of the opposite base pair 623-605 in helix H28. The interaction with helix H28 thus limits the scope of possible rearrangements in base pair 600-657, virtually freezing the position of the ribose-600.

If helix H28 harbors an alternative dinucleotide combination, as happens in the Group II clones, its ability to resist the rearrangements in helix H27 caused by the GU⇔WC replacement will be compromised. Indeed, as we argued earlier, an alternative dinucleotide combination 623-605 is expected to weaken the interaction between nucleotides 623 and 605 and may even result in the bulging of nucleotide 623. In the absence of the strong interaction between nucleotides 623 and 605, their positions will no longer be rigidly fixed, which, in turn, will hamper their ability to influence the position of base pair 600-657. As a result, the position of the ribose-600 will become solely dependent on the GU/WC identity of base pair 600-657. Now, only one of the two

identities of this base pair (GU) will allow the ribose-600 to form the normal contact with protein L4, while the other identity (WC) will render the ribosome non-functional. This would explain the above-mentioned fact that in all Group II clones, base pair 600-657 is always GU while clones with a WC base pair 600-657 and an alternative dinucleotide combination 623-605 have never been observed in our experiments.

5.4.10 Exceptional clones A11, B16 and C84

Among all 48 selected clones, only three, A11, B16 and C84, do not fit to the pattern followed by all other clones of the given AGPM. Thus, clones A11 and B16 do not have a structure-forming base pair, while clone C84 neither follows the GU-WC pattern nor contains base pair G600-U657. The viability of these exceptional clones strongly suggest that in each of them, the combination of the four selected nucleotides has somehow been able to arrange in the way that would provide for the integrity of the AGPM and of its interaction with the corresponding ribosomal protein (the latter requirement pertains to clones B16 and C84 only). Interestingly, in all three clones one of the two helices harbors either combination CA (clones A11 and B16) or AC (C84). In the past years, different types of A-C arrangements have been reported (BPS: database of RNA Base-Pair Structures; http://bps.rutgers.edu/bps). One of these arrangements (Figure 6b) has been found on many occasions and is thus established more firmly than others. In this arrangement, A and C are juxtaposed as G and U in the GU base pair. Such juxtaposition of A and C presumes the formation of the hydrogen bond between N6 of adenine and N3 of cytosine. In addition, two acceptors of an H-bond, N1 of adenine and O2 of cytosine become close to each other. To be stable, this arrangement thus requires that either A or C harbor proton and thus become positively charged. Such nucleotide forms are facilitated by acidic pH, but can also occur at the neutral pH if the whole structure benefits from the particular juxtaposition of the two nucleotides (see the legend to Figure 6b). The formation in clone C84 of the AC base pair shown in Figure 6b would fit this clone to the same pattern with other Group I C-clones. For clones A11 and B16, however, the formation of such base pair will not be helpful. Indeed, in these cases, the same juxtaposition of adenine and cytosine will form a base pair equivalent to UG, which was shown to be not nearly as effective as GU. To understand how nucleotides A and C are arranged in clones A11 and B16 and how their arrangement makes the two clones functional will thus require further analysis.

5.5 Discussion

5.5.1 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. Additional aspects of the usefulness of the naturally selected rRNA sequences for elucidation of particular aspects of the rRNA structure are discussed in the Supplementary Data. Compared to approaches that are 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 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.

5.5.2 New findings about 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. 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 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 provided new information on different aspects of the AGPM structure. First, it has allowed us to formulate a minimal requirement for the AGPM formation consisting 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.

Another observation pertains to the analysis of the C-clones, which showed that the GU⇔WC exchange at the center of the inter-helix contact mostly leads to the displacement of the riboses of the internal nucleotides, while the external riboses remain virtually unmovable. This conclusion is based on the superposition of the structures of motif L657 in the *E. coli* and *H. marismortui* ribosomes (Figure 7b) and is supported by the fact that such replacement does not affect the *E. coli* ribosome function even though the external nucleotide 600, which forms a direct contact with protein L4, becomes involved in a WC base pair instead of GU. Because in an isolated double helix, the GU⇔WC replacement causes the movement of both riboses, we argued that the above-mentioned immobility of the external riboses is due to the specific interaction between the two helices within AGPM that allows one helix to influence the conformation of the other. This phenomenon would thus represent another example of cross-talk between the two helices within AGPM.

Finally, we observed the asymmetry between GU and WC among the C-clones, according to which, in the case of an alternative dinucleotide combination 623-605, only the GU and not WC base pair 600-657 would make the ribosome functional. We argued that the presence of an alternative dinucleotide combination 623-605 introduces flexibility into the structure of helix H28, thus breaking the pipe-line of the inter-helix cross-talk. As a result, the position of the ribose-600 can no longer be influenced by helix H28 and becomes solely dependent on the identity of base pair 600-657. Based on the fact that in all Group II clones base pair 600-657 is GU and not WC, we suggest that in the normal AGPM structure, the cross-talk between the two helices mostly modifies the conformation of the WC-containing helix, placing the ribose of its external nucleotide in the same position as in the GU base pair, and not the other way around.

5.5.3 New findings about AGPM: principles of RNA-protein interaction

Within the complexes of motifs L639 and L657 with, respectively, proteins L35 and L4, the positions of the structural elements 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 these 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. We thus suggest that the formation of the particular conformation of the strand precedes its interaction with the protein and is a prerequisite condition for this interaction.

The specificity of the RNA-protein interaction in both motifs does not originate from contacts with unique parts of nucleotides, but instead, is based on 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.

5.5.4 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.

5.6. Funding

This work was supported by operating grant from Canadian Institutes of Health Research to S.V.S.. 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.

5.7 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 advice and discussions.

Conflict of interest statement. None declared.

5.8 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.

Arnez JG, Steitz TA. 1996. Crystal structures of three misacylating mutants of Escherichia coli glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP. *Biochemistry 35*:14725-14733.

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.

Bullock TL, Uter N, Nissan TA, Perona JJ. 2003. Amino acid discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants. *J Mol Biol 328*:395-408.

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.

Jang SB, Hung LW, Chi YI, Holbrook EL, Carter RJ, Holbrook SR. 1998. Structure of an RNA internal loop consisting of tandem C-A+ base pairs. *Biochemistry 37*:11726-11731.

Krasilnikov AS, Xiao Y, Pan T, Mondragon A. 2004. Basis for structural diversity in homologous RNAs. *Science 306*:104-107.

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.

Moras D, Comarmond MB, Fischer J, Weiss R, Thierry JC, Ebel JP, Giege R. 1980. Crystal structure of yeast $tRNA^{Asp}$. *Nature 288*:669-674.

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

Pan B, Mitra SN, Sundaralingam M. 1998. Structure of a 16-mer RNA duplex r(GCAGACUUAAAUCUGC)2 with wobble C.A+ mismatches. *J Mol Biol 283*:977-984.

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.

Rath VL, Silvian LF, Beijer B, Sproat BS, Steitz TA. 1998. How glutaminyl-tRNA synthetase selects glutamine. *Structure 6*:439-449.

Rould MA, Perona JJ, Steitz TA. 1991. Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature 352*:213-218.

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.

Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ. 2004. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol 11*:1729-1741.

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.

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

Webster GD, Sanderson MR, Skelly JV, Neidle S, Swann PF, Li BF, Tickle IJ. 1990. Crystal structure and sequence-dependent conformation of the A.G mispaired oligonucleotide d(CGCAAGCTGGCG). *Proc Natl Acad Sci USA 87*:6693-6697.

Westhof E, Dumas P, Moras D. 1985. Crystallographic refinement of yeast aspartic acid transfer RNA. *J Mol Biol 184*:119-145.

Wild K, Sinning I, Cusack S. 2001. Crystal structure of an early protein-RNA assembly complex of the signal recognition particle. *Science 294*:598-601.

Wild K, Weichenrieder O, Leonard GA, Cusack S. 1999. The 2 Å structure of helix 6 of the human signal recognition particle RNA. *Structure 7*:1345-1352.

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

5.9 Table

Table 1. Nucleotide sequences of the selected clones

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

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

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

The nucleotide sequences following the GU-WC pattern are identified by a continuous underline that includes both central base pairs. The individually underlined base pairs are structure-forming. The internal and external strands of both helices are marked by italic letters *i* and *e*, respectively.

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

5.10 Figures

Figure 1

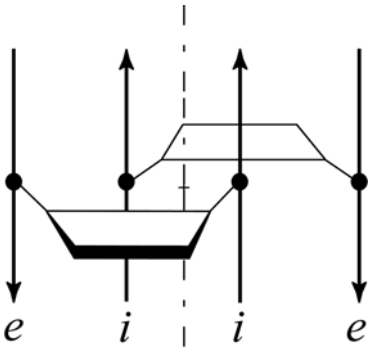


Figure 1. Schematic representation of AGPM. Trapezoids stand for base pairs opened toward the minor grooves. Arrows represent backbones directed 5′→3′. The internal and external strands of both helices are marked by italic letters *i* and *e*, respectively. The internal strand of each helix is packed 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

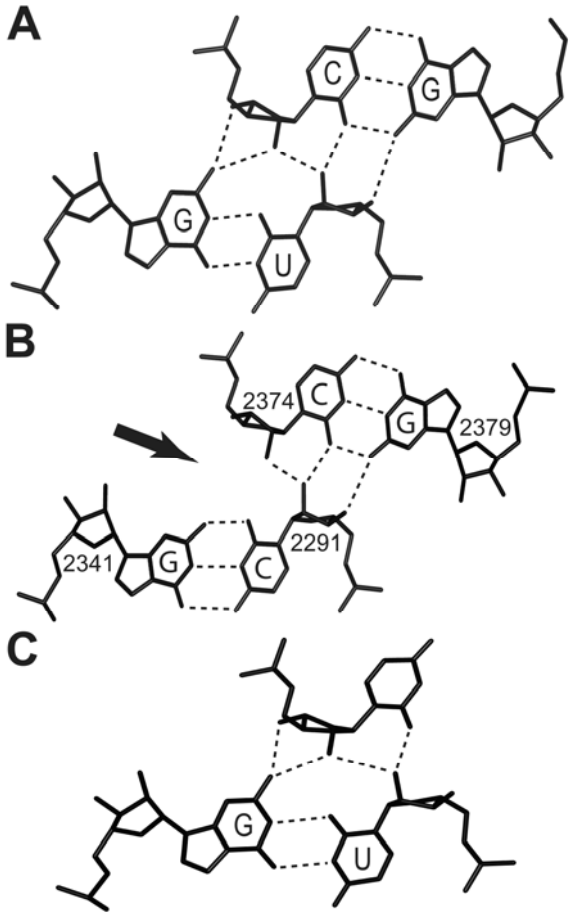


Figure 2. Different arrangements of the central base pairs in AGPM.

(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 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

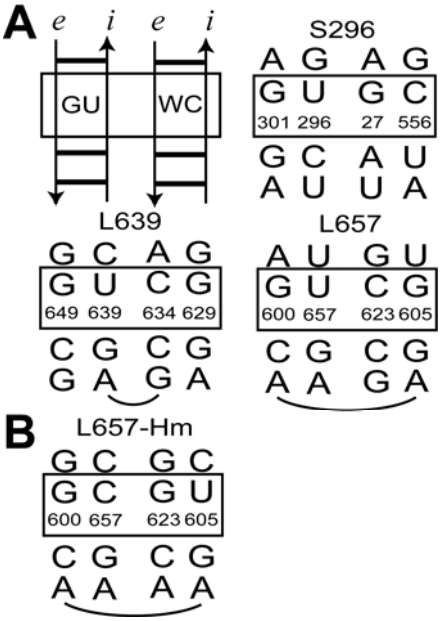


Figure 3. Nucleotide sequences of the three cases of AGPM considered in this study. In each case, the *E. coli* numbering is used.

(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 & 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 & 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

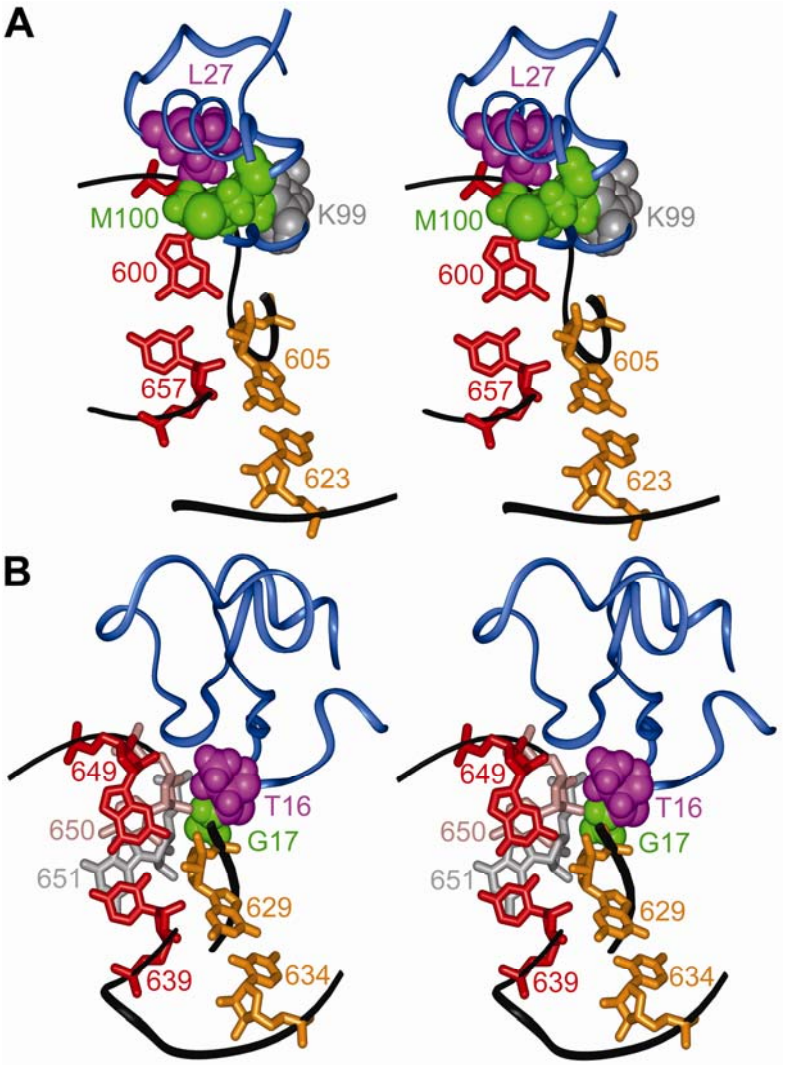


Figure 4. The structural contexts of motifs L657 (A) and L639 (B) taken from the *E. coli* ribosome.

(A) In motif L657, residues L27 (magenta), K99 (grey) and M100 (green) of the ribosomal protein L4 (blue ribbon) tightly interact with the ribose and with 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 the 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

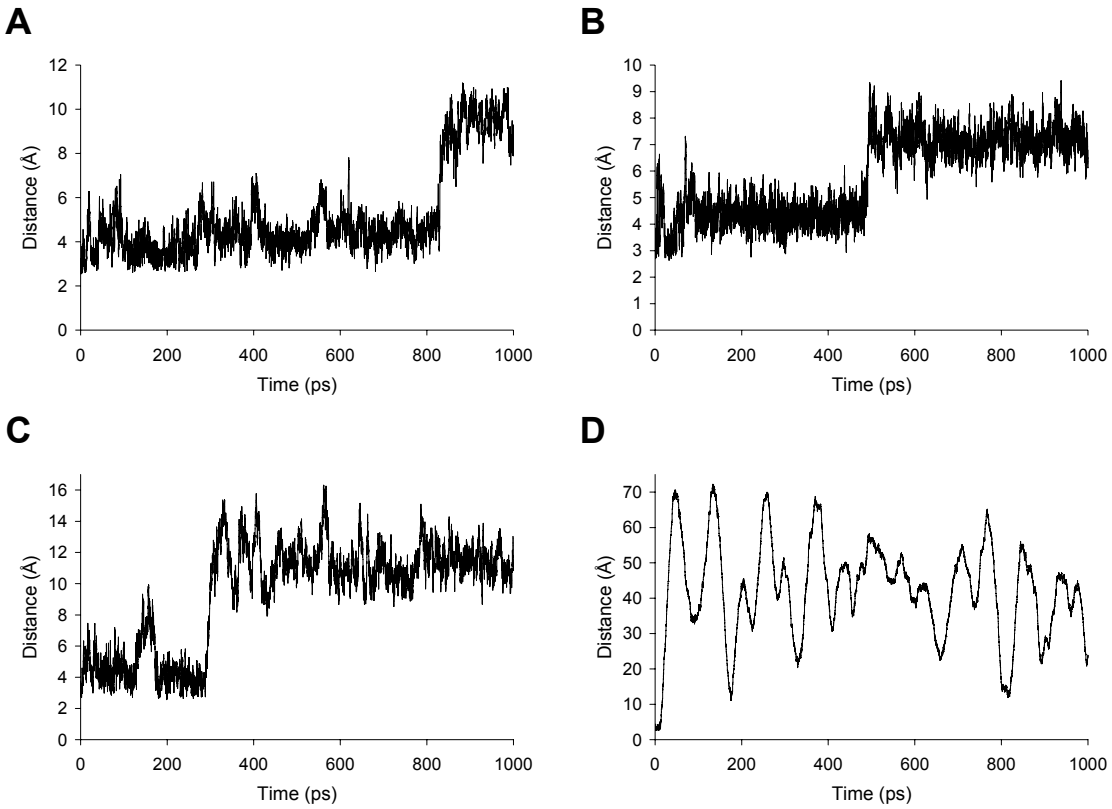


Figure 5. Molecular dynamics simulations of the AGPM structure containing different nucleotide triples. In the four complexes tested, the dinucleotide combinations occupying the central base pairs were GU-CG (A), GU-UG (B), GC-UG (C) and UG-UG (D). The CG, GC and GU combinations were initially arranged as normal base pairs, while in the UG combination, the internal guanosine was initially put in the position corresponding to that 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 6

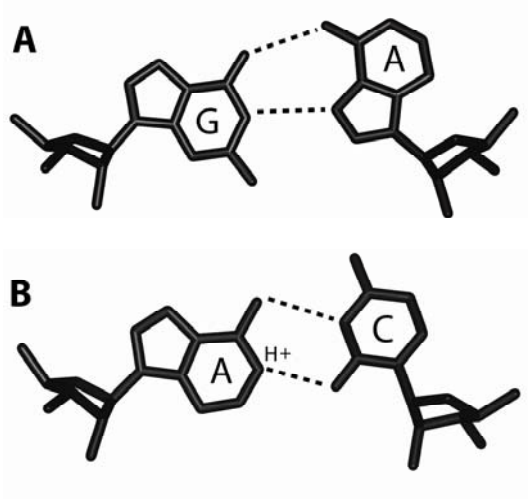


Figure 6. Juxtapositions of the bases in the GA and AC base pairs

(A) The proposed WC-like arrangement for the GA base pair in clones B3 and B6. The formation of this base pair requires that the adenosine have the unusual *syn*-conformation. Such arrangement is observed in base pairs G_9-A_{16} and G_{21}-A_4 of the crystal structure of a DNA oligonucleotide duplex (pdb entry code 1dnm) (Webster et al., 1990).

(B) The GU-like juxtaposition observed in base pairs A_{105}-C_{112} and A_{113}-C_{104} (pdb entry code 402d) (Jang et al., 1998), A_{11}-C_{22} and A_{27}-C_6 (pdb entry code 405d) (Pan et al., 1998), A_5-C_{24} and A_{23}-C_6 (pdb entry code 1d4r) (Wild et al., 1999), A_{139}-C_{158} and A_{157}-C_{140} (pdb entry code 1jid) (Wild et al., 2001), A_{26}-C_{44} (pdb entry codes 1o0b, 1qtq, 1gtr, 1qru) (Rould et al., 1991; Arnez & Steitz, 1996; Rath et al., 1998; Bullock et al., 2003), A_{1500}-C_{1402} (pdb entry code 1j5e) (Wimberly et al., 2000), A_{192}-C_{178} (pdb entry code 1u9s) (Krasilnikov et al., 2004) and A_{44}-C_{26} (pdb entry code 1y27) (Serganov et al., 2004). To be stable, this juxtaposition requires that either A or C exists in the (+)-ionized form. Such forms are favored by the acidic pH, but can also occur at neutral conditions. For example, in the yeast $tRNA^{Asp}$ (pdb entry code 3tra) (Moras et al., 1980; Westhof et al., 1985) the tertiary base pair between A46 and G22 presumes the presence of an extra proton at the inter-base contact (not shown).

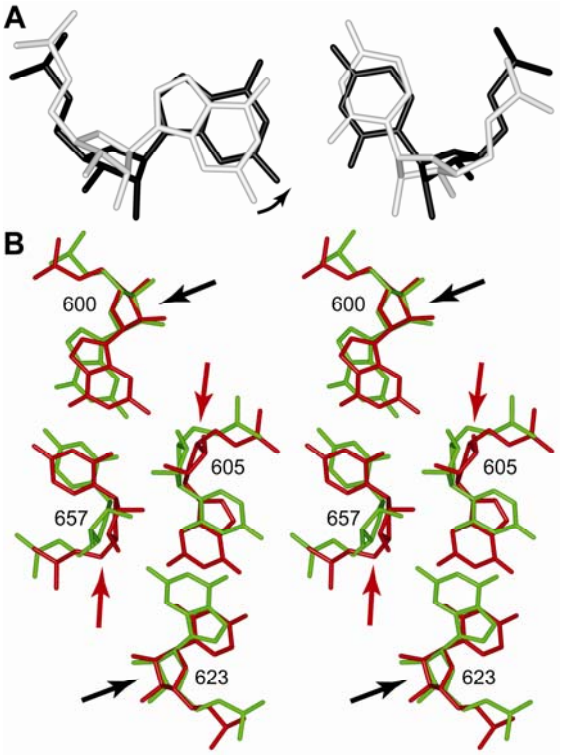


Figure 7. Conformational rearrangements associated with the GU⇔WC exchange of the central base pairs in AGPM.

(A) The superposition of base pairs G_{71}-C_2 (black) and G_4-U_{69} (white) in the structure of the yeast $tRNA^{Phe}$ (pdb entry code 1ehz) (Shi & Moore, 2000). The superposition was obtained by superposing the flanking WC base pairs of 4-69 (base pairs 5-68 and 3-70) with those of 71-2 (pairs 1-72 and 3-70) (not shown). The replacement of the GU base pair (white) by GC (black) rotates the base of the guanosine by about 15° toward the major groove (black arrow). However small this displacement is, it can be large enough to damage the interaction between the external nucleotide 600 and protein L4 and thus to make the ribosome non-functional.

(B) 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 allows the visualization of the local conformational changes in 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 AGPM, the GU⇔WC exchange of the central base pairs affects 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. For both structures, the *E. coli* nucleotide numbering is used.

5.11 Supplementary data

5.11.1 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.

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 Supplementary Table S1, in all three AGPMs (S296, L639 and L657), the GU-WC pattern is maintained at the level of 97% or higher. 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 (Table 1). 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.

In almost all alternative sequences of the naturally selected 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 (Supplementary Table S1). The extension of the analysis to all thirteen cases of AGPM found in the ribosome structure (Gagnon & 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 (Supplementary Table S2). 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 AGPM. Moreover, due to such a small number of exceptions, they may constitute artifacts originated from sequencing errors or misalignment.

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 1 and Supplementary Table S1). However, the total number of 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.

The possibility that the details of RNA-protein interactions could vary in different organisms highlights the problematic of using the abnormal sequences to deduce subtleties of the interaction of AGPMs with a particular ribosomal protein. 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. Also, as mentioned above, it is 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.

5.11.2 Potential limitations on the usage of GA as a central base pair in AGPM

Even if a GA combination forms as a WC-like base pair, we doubt whether it is able to function as a structure-forming base pair, i.e. to facilitate the formation of the second helix and of the whole 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.

5.12 Supplementary Methods

5.12.1 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 Supplementary Table S3.

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.

5.13 Supplementary Tables

Supplementary Table S1. 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 1 of the article.

Supplementary Table S2. 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**

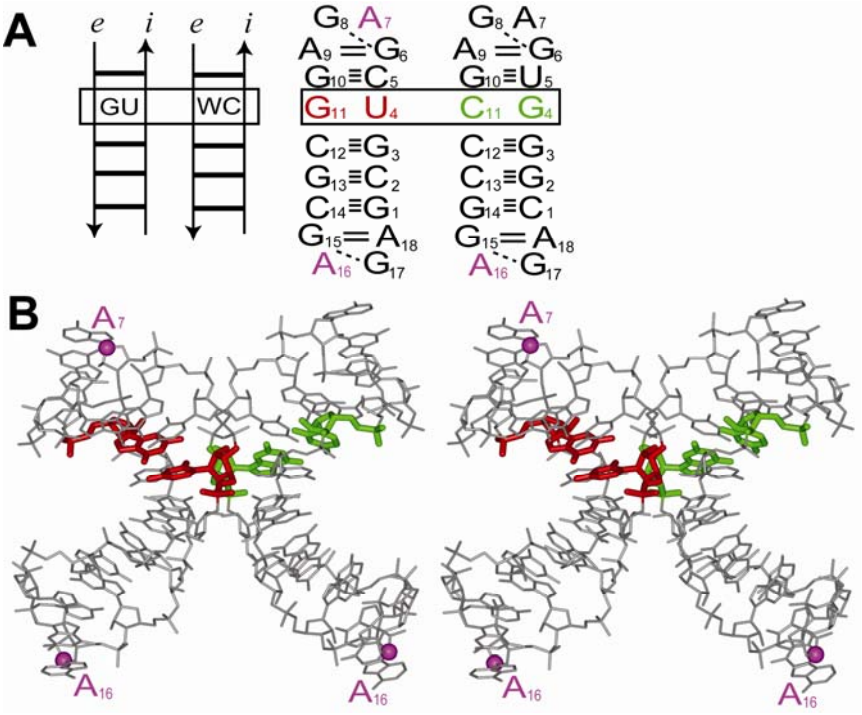
The data were obtained based on the available rRNA alignments (Wuyts et al., 2004). Cases of AGPM considered are: S62, S296, S549, S757, L554, L639, L657, L839, L1864, L2291, L2687, L2698 and L2847 (2,3). For the statistics, only those cases have been considered where the identities of all four nucleotides 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.

Supplementary Table S3. 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.

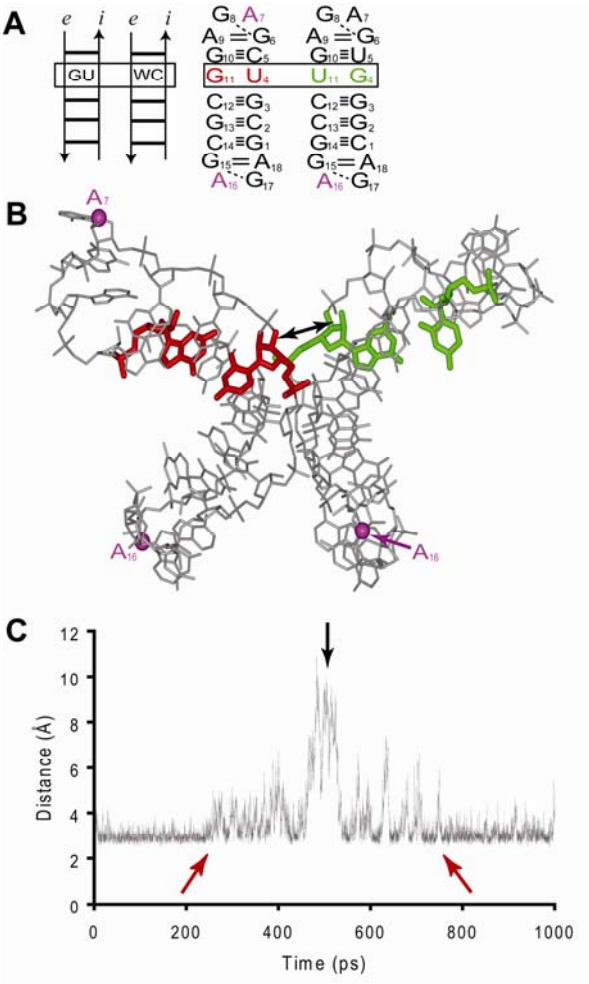
5.14 Supplementary Figures



Supplementary Figure S1. Modeled complex of AGPM used in molecular dynamics simulations.

(A) The 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 dashed line stands for a constraint (H-bond) applied for the distance between the base and the ribose. Details of the constraints are given in the Materials 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 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. These atoms are located within the GAGA tetraloops at the extremities of the helices. The fixation of these atoms helps to avoid uncontrolled deterioration of the construct at the regions outside the inter-helix contact. Because of this fixation, the helix that contains the green central base pair would dissociate from the helix that contains the red one. When the conditions of the simulations are controlled in this way, it becomes possible to estimate the stability of the whole arrangement.

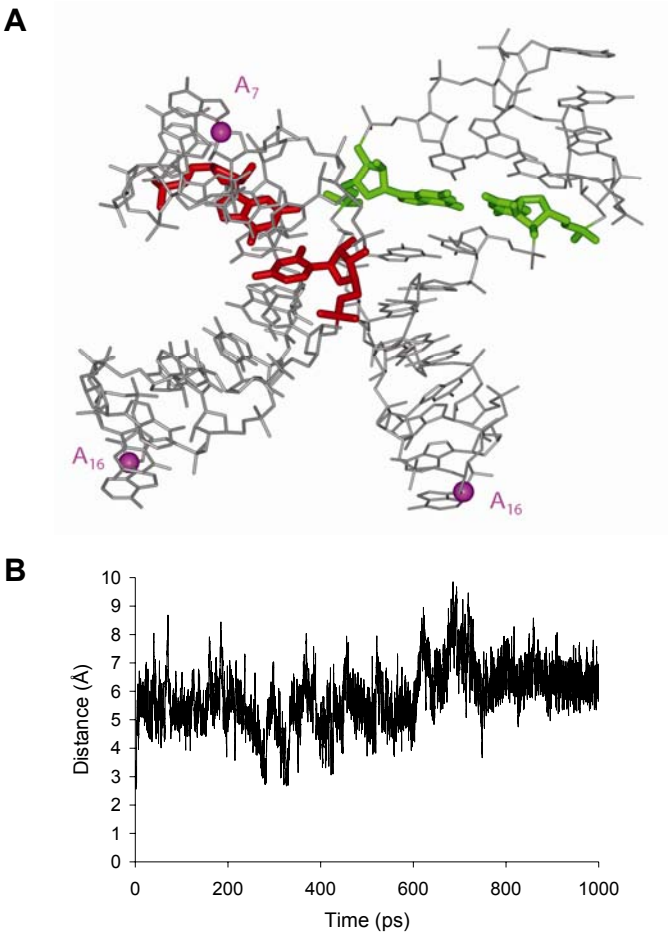


Supplementary Figure S2. The second molecular dynamics simulation of the AGPM construct containing the central GU-UG dinucleotide juxtaposition.

(A) Secondary structure of the AGPM construct containing GU-UG drawn according to the schematic representation on the left (*i* and *e* stand for the internal and external strands, respectively). The GU and UG combinations were initially arranged as normal base pairs. Other details of the figure are described in Supplementary Figure S1a.

(B) The snapshot of the AGPM construct taken at 500 ps of simulation (corresponds to the black arrow in C). The three C1′ atoms whose positions were fixed during MD simulations are shown as magenta spheres. The UG-containing helix (with the green base pair) bends over its axis, making the U of the GU base pair (red) and the G of the UG base pair (green) distant from each other (double-headed black arrow). The monitored distance between the O2′ atoms of U_4 and G_4 in this structure exceeds 9 Å.

(C) Evolution of the contact between the internal riboses monitored by following the distance between their O2′ atoms and plotted against the time. The black arrow indicates the moment corresponding to the structure shown in B. A similar conformation of the UG-containing helix is observed for the whole 500 ps interval between the red arrows.



Supplementary Figure S3. The third molecular dynamics simulation of the AGPM construct containing the central GU-UG dinucleotide juxtaposition.

The secondary structure of the construct is the same as shown in Supplementary Figure S2a. In this simulation, the UG base pair shifted as a whole out of the inter-helix contact zone soon after the beginning of the simulation, yielding its place to the next base pair C_{12}-G_3.

(A) The snapshot of the AGPM construct taken at the end of the simulation. At this moment, the UG-containing helix has already shifted as a whole up from its original position. As a result of this shift, the UG base pair (green) has left the area of the close contact with the opposite helix, yielding its place to the next base pair C_{12}-G_3. This shifted juxtaposition of the two helices maintained until the end of the simulation.

(B) Evolution of the contact between the internal riboses of the O2´ atoms of **U_4** and **G_4** plotted against the time. The shift of the GU-containing helix up, which happened at the very beginning of the simulation, resulted in the increase of the measured distance from 3 Å to about 6 Å.

5.15. Supplementary 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.

Chapter 6

Article 5

In vivo study of the G-ribo motif

6. *In vivo* analysis of the general constraints imposed on the structure of G-ribo motifs in the functional ribosome

Yury I. Boutorine and Sergey Steinberg

Département de Biochimie, Université de Montréal, Montréal, QC, Canada.

Manuscript in preparation

Running title: *In vivo* study of the G-ribo motif

Contribution of each author:
Yury I. Butorin: Developed the detailed experimental scheme of the library design of motifs S1047 and S521, performed the selection and characterization of the functional clones of both motifs, participated in data analysis and preparation of manuscript and figures.
Sergey V. Steinberg: participated in data analysis, preparation of manuscript and figures.

6.1 Abstract

G-ribo is a recurrent RNA motif recently identified in the ribosome structure. It consists of two double helices juxtaposed in a particular way and linked together by a short connector region. The arrangement of the double helices is stabilized by a network of specific contacts, which include hydrogen bonds, van der Waals and stacking interactions. Here, we analyze the importance of some of these contacts for the integrity of two G-ribo motifs S521 and S1047 located, respectively, in Domains I and III of the 16S rRNA. For each of the two G-ribo motifs studied, we designed a combinatorial library of the 16S rRNA in which several key nucleotide positions of the motif were randomized. The selection of functional 16S rRNA variants and analysis of their nucleotide sequences allowed us to formulate the general requirements to which the nucleotide sequence of a G-ribo motif should fit in order to provide for the proper tertiary structure.

Keywords: ribosomal RNA/ RNA motif/combinatorial library/specialized translation system /G-ribo

6.2 Introduction

The G-ribo motif represents a particular side-by-side arrangement of two double helices connected by an unpaired region containing at least three nucleotides (Steinberg & Boutorine, 2007b). The definitions of helices, strands, layers, and particular nucleotides of the G-ribo motif are provided in Figure 1A. At the zero layer, the top base pairs [0P; 0Q] and [0R; 0S] of both helices are juxtaposed as shown in Figure 1B. At the center of this juxtaposition, guanosine 0P forms two hydrogen bonds with the ribose of nucleotide 0R (Figure 1B). This interaction, which we call G-ribo, has also given the name to the whole arrangement. One of the two hydrogen bonds between 0P and 0R involves the amino group of 0P and thus requires that this nucleotide be guanosine. The other hydrogen bond is formed by the two riboses and is thus independent of the identities of the two nucleotides. In the ribosome structure, the G-ribo arrangement has been found in eight different places: three in 16S rRNA and five in 23S rRNA.

In addition to the G-ribo interaction, all G-ribo motifs share nucleotide arrangements at layers +1 and -1. At the -1 layer, nucleotide -1T forms the A-minor interaction with base pair [-1P; -1Q] (Figure 1A). Nucleotide -1T belongs to the so-called T-bulge, which loops out of strand R between nucleotides 0R and +1R. Although in different G-ribo motifs, the T-bulge contains between one and three nucleotides (Figure 1C), nucleotide -1T always occupies the last position. The fact that nucleotide -1T forms the A-minor interaction with base pair [-1P; -1Q] explains its predominant adenosine identity.

Analysis of the available G-ribo cases shows that the most important role in the proper juxtaposition of helices H1 and H2 and in the integrity of the whole motif is played by two base pairs [0P; 0Q] and [-1P; -1Q] of helix H1 and by nucleotide -1T from the T-bulge. The identity of base pair [0P; 0Q] is responsible for the G-ribo interaction at the core of the motif, while the interaction between base pair [-1P; -1Q] and nucleotide -1T provides the second point of fixation for the juxtaposition of the two helices.

An interesting feature of the G-ribo motif consists in its ability to promote the formation of a particular type of pseudoknot (Steinberg & Boutorine, 2007a): five out of eight identified G-ribo motifs are associated with the formation of specific pseudoknot

structures. In all these cases, strand P is interrupted between positions -1P and -2P. The part of this strand that includes nucleotide -1P becomes connected to strand Q, while the part containing nucleotide -2P becomes connected to either strand R or S (Figure 1D). Such a fragmented arrangement of strand P constitutes a pseudoknot.

Analysis of the G-ribo motif structure revealed elements responsible for the tendency of the motif to form pseudoknots (Steinberg & Boutorine, 2007a). We found that due to the A-minor interaction between base pair [-1P;-1Q] and adenosine -1T, base pair [-2P;-2Q] is able to over-twist with respect to [-1P;-1Q] such that nucleotide -2P, instead of stacking on -1P, becomes stacked on -1T. Such an over-twist, to varying extents, exists in almost all known G-ribo motifs. However, in only five identified cases, the over-twist is accompanied by the reconnection of different fragments within the polynucleotide chain that leads to a pseudoknot.

In this paper we analyzed the importance of different elements of the G-ribo motif for its structural integrity. As examples of the motif, we took motifs S1047 and S521 existing in the 16S rRNA (Figure 2). These two motifs have fundamentally different characteristics. In S521, there is a strong over-twist between base pairs [-1P;-1Q] and [-2P;-2Q], which is stabilized by adenosine -1T and is accompanied by the formation of a pseudoknot (Figure 3A). In S1047 the over-twist between these base pairs is much smaller, and position -1T is occupied by a pyrimidine (Figure 3B). It is thus not surprising that S1047 does not form a pseudoknot.

Other differences between S1047 and S521 relate to the way they interact with ribosomal proteins. In S1047, the sugar-phosphate backbone of nucleotides 0P and -1P forms contacts with protein S14. Because these contacts do not involve nitrogen bases of either 0P or -1P, they are not expected to be strongly dependent on the identities of nucleotides 0P and -1P. In S521, on the contrary, the major groove of helix H1 at level 0 and -1 tightly interacts with amino acid residues R49, N45 and D88 of protein S12. One can suggest that the identities of base pairs [0P; 0Q] and [-1P; -1Q] are important for the interaction with S12. Thus, it would be interesting to understand whether the differences in the structure of the two motifs and in the way they interact with the ribosomal proteins are reflected in the nucleotide sequence constraints imposed on each motif.

Motif S521 is especially important for the function of the ribosome because its internal connector (nucleotides 530-532) participates in tRNA recognition in the A-site of the 30S ribosomal subunit (Moazed & Noller, 1989) and is often referred to as loop 530 (Santer et al., 1995). The presence of the pseudoknot significantly stabilizes the core of the motif S521 and makes a solid structural basis for flexible and functionally important loop 530. The core of motif S1047 (nucleotide positions demonstrated on the Figure 1C) is not directly involved into formation of a functionally important centre of the ribosome, yet the lower part of the helix 1 of this motif (position C1054) participates in the formation of the A-site of the 30S subunit (Jenner et al., 2010). As long as motif S1047 is not stabilized with pseudoknot, the inter-nucleotide interactions in the core of the motif play an important role for the overall stability of the whole arrangement and the neighbouring regions. Thus, motif S1047 represents a good model for studying the impact of the nucleotide replacement in the core of the motif on the structure of the adjacent regions and the ribosome in general.

For analysis of the motifs, we generated 16S rRNA mutants in which the key nucleotides in positions 0P, 0Q, -1P, -1Q and -1T were allowed to vary. For this, we used the specialized translation system, which was based on the expression of a modified 16S rRNA having an alternative anti-Shine–Dalgarno sequence (Hui & de Boer, 1987; Lee et al., 1996; Abdi & Fredrick, 2005; Rackham & 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 protein chloramphenicol acetyl-transferase (CAT). To determine the constraints imposed on the identity of a nucleotide or of a set of nucleotides in the functional 16S rRNA, we designed combinatorial gene libraries in which the nucleotides in question were randomized. The expression of such combinatorial libraries in bacteria was expected to allow us to obtain functional clones of the 16S rRNA different from the WT. Analysis of the nucleotide sequences of these clones would allow us to formulate the general constraints imposed on the nucleotide sequence of the motif in a functional ribosome.

6.3 Results

6.3.1 Analysis of the variant sequences of motif S1047

The randomization of five nucleotides 0P, 0Q, -1P, -1Q and -1T of motif S1047 and *in vivo* expression of the combinatorial library yielded 36 functional clones different from the WT. The nucleotide sequences and activities of these clones are shown in Table 1. The activities ranged from almost 100% of the WT to a few percentage points, which was still higher than the background level.

For analysis of the importance of different elements of the motif, we grouped all variants presented in Table 1 in several subsets presented in Tables 2-9. Within each subset, variants differed from each other by a simple structural change, which allowed us to evaluate the importance of the corresponding structural element for the structure and stability of the motif.

Comparison of clones A2 and A6 (Table 2) as well as of clones A3 and A5 (Table 3) shows that the replacement of cytidine -1T by uridine results in a slight decrease of the ribosome activity. However, when cytidine -1T is replaced by a purine, as it happens between clones A14 and A15 (Table 4) and between clones A7 and A17 (Table 5), the drop of the activity becomes significant.

A replacement of the GC base pair [0P;0Q] by GU does not seem to affect the activity of the ribosome (clones A1 and A3 in Table 6). A similar situation is observed at the -1 level as well: clone A1, in which the base pair [-1P;-1Q] is GC, and clone A2, where this base pair is GU, have also similar activities. Although each of the three replacements mentioned above C(-1T)→U, GC([0P;0Q])→GU and GC([-1P;-1Q])→GU led to only a slight decrease of the ribosome activity, their co-occurrence in variant A16, resulted in a significant drop of the activity to 62% (Table 6).

A replacement of G in position 0P by any other nucleotide eliminates one of the two hydrogen bonds between nucleotides 0P and 0R that constitute the G-ribo interaction (Figure 1B). The fact that such replacement takes place in clones A9, A11, A13, A12, A22, A27 and A19 (Table 7) demonstrates that the absence of this hydrogen bond does not result in a complete loss of the ribosomal function. In spite of the absence of the canonical G-ribo interaction in the clones listed in the Table 7, their activity stays in the range of 40-70%.

In clones A13, A12, A19, and A27 (Table 7), not only was the G-ribo interaction compromised, but also nucleotides 0P and 0Q did not form a WC combination. Non-canonical dinucleotide combinations can also be observed for base pair [-1P;-1Q] in all clones presented in Table 8. Interestingly, in all such cases, a non-canonical base pair occurs either at layer 0 or -1, but not at both layers simultaneously. The other base pair has always been WC or GU. The formation of only one canonical base pair at either layer 0 or -1 in the context of a pyrimidine in position -1T provides for the activity of the clone in the range of 40-80%. If in a variant, a non-canonical base pair at either layer 0 or -1 co-occurs with a purine in position -1T, the activity drops below 20% (compare clones A33 and A34 in Table 9).

6.3.2 Analysis of the variant sequences of motif S521

Contrary to what one might have expected to observe based on our experience with motif S1047, the randomization of the same five nucleotide positions 0P, 0Q, -1P, -1Q and -1T in motif S521 and the expression of this library in the specialized translation system provided only one functional clone different from the WT. In this clone, the identity of only one nucleotide was different from the WT: the adenosine in position -1T was replaced by uridine. The activity of this clone was 85% of that of the WT.

Inspection of the structure of the 30S ribosomal subunit showed that in motif S521, the major groove of helix H1 at level 0 and -1 tightly interacts with amino acid residues R49, N45 and D88 of the ribosomal protein S12. It is thus reasonable to suggest that any modification of base pairs [0P; 0Q] and [-1P; -1Q] will strongly affect their interaction with S12, thus seriously inhibiting ribosome function. Unlike the latter base pairs, nucleotide -1T is not involved in any RNA-protein interaction, which explains why in our library search only this nucleotide has been modifiable.

6.4 Discussion

In this paper we study nucleotide sequence constraints imposed on G-ribo motifs in the functional ribosome. We analyzed two motifs, S1047 and S521, existing in the 16S rRNA. Analysis of the sequence requirements for both motifs was performed with use of

the combinatorial library approach. For each motif, we generated a combinatorial gene library, in which five positions 0P, 0Q, -1P, -1Q and -1T were randomized. Both these motifs have special features that significantly affected the results of our *in vivo* selection of sequence variants. While motif S521 is associated with the pseudoknot structure formation, motif S1047 is not. Existence of the pseudoknot solidifies the local structure of H1 of S521 and also results in the formation of the over-twist between base pairs [-1P; -1Q] and [-2P; -2Q]. Both motifs form contacts with near-by ribosomal proteins: motif S1047 interacts with protein S14, while motif S521 interacts with protein S12. The particular mode of interaction for the two motifs is, however, essentially different. While in motif S1047, RNA-protein contacts are formed by phosphate groups, in motif S521 nitrogen bases directly interact with the protein. Based on this difference, we expected that the set of allowed nucleotide sequences for motif S521 would be more restricted than for motif S1047.

6.4.1 G-ribo motif S1047 can exist without the G-ribo interaction

Analysis of the sequence variants of motif S1047 shows that even when the G-ribo interaction cannot be formed due to the absence of guanosine in position 0P, the ribosome activity remains relatively high (40% or higher) (Table 7). Also, the presence of a non-canonical base pair [0P; 0Q] does not abolish the ribosomal activity (Table 7). Based on these observations we suggest that even if the G-ribo interaction is not present, the particular juxtaposition of helices H1 and H2, specific for the G-ribo motif, is supported due to the presence of other structural elements. The interactions that nucleotide -1T forms with H1 can serve as such element. Bulged from Helix 2 (H2), nucleotide -1T stacks below the ribose of nucleotide 0Q and forms hydrogen bonds with the minor groove of base pair [-1P; -1Q]. We suggest that -1T can serve as a structural bridge between the two helices, thus stabilizing their particular juxtaposition. In order to establish an A-minor-like contact between nucleotide -1T and helix H1, nucleotides -1P and -1Q should form a stable base pair. Surprisingly, in all sequences of clones, shown in Table 7, base pair [-1P; -1Q] is always WC. As a result, in all these clones nucleotide -1T can form a hydrogen bond with the minor groove of base pair [-1P; -1Q]. We thus conclude that the presence of a stable base pair [-1P; -1Q] is a minimal requirement for

the motif formation if base pair is [0P; 0Q] non-WC, while the G-ribo interaction does not exist.

6.4.2 G-ribo motif S1047 can exist without a stable base pair at -1 layer of H1

As one can judge from the analysis of the clones shown in Table 8, the -1 layer of H1 (base pair [-1P; -1Q]) can tolerate a non-canonical base pair. It is worth mentioning that even though in all sequences shown in table 8, the -1 base pair is occupied by a non-canonical dinucleotide combination, the 0 layer is always occupied by a WC base pair. We thus conclude that if nucleotides 0P and 0Q form a stable base pair, then the stability of base pair [-1P; -1Q] is no longer critical for the integrity of the motif. Further damage of the motif will result in an essential loss of the ribosome function, as illustrated by clones A31 and A32 (Table 9), which lack stable base pairs at both layers 0 and -1 of H1. Thus, our results allow us to propose the minimal requirement that would guarantee the formation of the motif. This requirement consists in the presence of a stable base pair at either layer 0 or -1 of H1.

Interestingly, the selection of sequence variants for base pairs [0P;0Q] and [-1P; -1Q] has become possible only for motif S1047, while in all selected clones of motif S521 the identities of these positions stayed unchanged. One could suggest that the pseudoknot associated with motif S521 contributes to the stability of H1, so that more variability would be expected for the randomized positions in H1 (or at least a co-variation in H1). In contrast, nucleotide identities have been highly conserved, which can be explained by the presence of essential contacts between S521 and protein S12.

6.4.3 The role of identity of nucleotide -1T in G-ribo motifs S1047 and S521

Comparison of selected variants showed that in most active clones of the motif S1047 (clones A1-A13), position -1T was occupied by either cytosine or uridine. Also, clones having a purine in position -1T were shown to have a lower activity (clones A14 and A15 in Table 4 and clone A17 in Table 5). These results are consistent with conservation data, which suggests a strong preference for a pyrimidine over a purine in position -1T of motif S1047 (Chapter 2, Supplemental Table 2). Analysis of the nucleotide conservation of position -1T of motif S1047 shows that 99.8% and 72.6% of all bacterial and archaeal sequences respectively are occupied with pyrimidine. At the same time, in the case of motif S521, in spite of the high conservation of adenosine in position -1T (more than 99.7% of all prokaryotic sequences have adenosine in position -1T of S521, Supplemental Table 2 of Chapter 2), the only selected sequence variant, which was almost as active as WT (85% of that of the WT), had U in this position. We can conclude that based on our *in vivo* selection, a pyrimidine in position -1T of both motifs is universally accepted. One can wonder, why in motif S1047, both the *in vivo* selection and the phylogenetic analysis (Steinberg & Boutorine, 2007b) demonstrate the high preference for a pyrimidine in position -1T, while in motif S521, this position prefers to be adenosine. We hypothesize that in motif S521 the preference for a particular type of nucleotide is linked to the local structure of the motif, and more precisely, to the extent of the over-twist between base pairs [-1P; -1Q] and [-2P; -2Q].

In motif S521, which is associated with the formation of a G-ribo pseudoknot, base pairs [-1P; -1Q] and [-2P; -2Q] form a large helical over-twist. The over-twist is stabilized by adenosine -1T, which stacks to nucleotide -2P and makes an A-minor contact with base pair [-1P; -1Q] (Figure 3A). One could suggest that a replacement of this adenosine by a smaller pyrimidine would not interfere with formation of the large helical over-twist between base pairs [-1P; -1Q] and [-2P; -2Q]. This suggestion correlates with the fact that the A→U replacement in position -1T of S521 results in an only ~15% drop of the ribosome activity compared to that of the WT.

A smaller pyrimidine in position -1T could be beneficial, if the formation of a wide over-twist between base pairs [-1P; -1Q] and [-2P; -2Q] has to be avoided. Indeed, in motif S1047 a small over-twist between base pairs [-1P; -1Q] and [-2P; -2Q] can be

observed (Figure 3B). An introduction of a purine -1T instead of pyrimidine would stimulate the formation of a bigger over-twist between base pairs [-1P; -1Q] and [-2P; -2Q], which could promote mis-folding or formation of alternative structures.

6.5 Conclusions

The data presented here allow us to make suggestions concerning the minimal sequence requirements for the formation of the G-ribo motif. We argue that for the G-ribo motif to form, a Watson-Crick or GU base pair at either the 0 or -1 layer of Helix 1 must be present. The results of the selection of sequence variants demonstrate the importance of nucleotide -1T for the structural stability of the motif. We argue that nucleotide -1T plays a substantial role in supporting a particular juxtaposition of the two helices, being directly linked to nucleotide 0R, stacked to the ribose 0Q and involved in making A-minor-like contact with base pair [-1P; -1Q]. A pyrimidine in position -1T seems to be acceptable in any G-ribo motif: even though some motifs show a clear preference for adenosine -1T, the presence of a pyrimidine does not jeopardize the overall geometry of the motif. This conclusion is also supported by the fact that in motifs L2323, L2383, L1309 and L1642, from 4 to 25% of nucleotide sequences contain a pyrimidine in position -1T (Supplemental Table 1 Chapter 2).

6.6 Experimental procedures

6.6.1 Bacterial Strains and Media

For cloning and selection, we used the *E. coli* strain DH5α. Cultures were grown in the LB medium (Luria & Burrous, 1957) or in the LB medium with 100 μg/ml ampicillin (Sigma-Aldrich).

6.6.2 Plasmids

For cloning of combinatorial 16S rRNA gene libraries and for selection of functional clones, plasmid pAMMG carrying the specialized ribosome system was used (Gagnon et al., 2006). This plasmid is analogous to the ones described elsewhere (Lee et al., 1997; Morosyuk et al., 2000; Belanger et al., 2004).

6.6.3 Combinatorial gene libraries: primers and cloning

To randomize the five nucleotides comprising the base pairs [0P; 0Q] and [-1P; -1Q] as well as position -1T of S521 and S1047 motifs (Figure 2), we used an overlapping extension PCR procedure (Ho et al., 1989). In this way, the entire region comprising the motifs S521 and S1047 was amplified by consecutive multistep PCR. Oligonucleotide primers were synthesized with random nucleotides at the designated positions (Montreal Biotech Inc.). The list of used oligonicleotides is shown in Table 10. The first step involved the PCR amplification of three individual fragments using the 16S rRNA gene as a template. For the library of motif S1047, respective sets of the primers were used as follows: (A) Bgl II_forw and 1047_RP1; (B) 1047_FP1 and 1047_RP2; (C) 1047_FP2 and Xba_I_Rev. On the last step of the PCR amplification, three PCR products (A, B and C), were mixed together with two flanking primers Bgl II_forw and Xba_I_Rev. For the library of motif S521, the first step of the multistep PCR reaction consisted of amplification of two individual fragments using the 16S rRNA gene as a template and the following set of primers: (A) Kpn_I_forw and 521_RP; (B) 521_FP and Bgl II_rev. The last step of PCR amplification included two PCR products of the previous step (A and B), with use of two flanking primers Kpn_I_forw and Bgl II_rev. The PCR products were cloned in plasmid pAMMG using restriction sites *BglII* and *XbaI* for the S1047 library and *KpnI* and *BglII* for the S521 library.

Finally, plasmids harboring the combinatorial 16S rRNA gene libraries were transformed into the DH5α cells by electroporation. All PCR reagents, Vent DNA polymerase, restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. 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 (Bioshop Canada Inc.) to a final concentration of 1 mM. After incubation for 3.5 h, the library was plated on selection plates containing 200 μg/ml chloramphenicol and 1 mM isopropyl-1-thio-β-D-galactopyranoside. Out of a pool of $2x10^5$ ($1x10^5$) transformants, approximately 200 (30) colonies were obtained on chloramphenicol-containing plates for the library of motif S1047 (S521), and subsequently taken for further analysis.

6.6.4 Measurement of the ribosome efficiency

The efficiency of the selected clones was assessed with the GFPuv3 assay, as described previously (Gagnon et al., 2006).

6.6.5 Sequencing

Sequencing of the selected clones was performed on Applied Biosystems 3730 DNA analyzer at the Genomics Facility of the Institute for Research in Immunology and Cancer (IRIC), at Université de Montréal. Primers Bgl II_forw and 1047_FP1 were used for reading positions (1047-1048), and positions (1209-1210 and 1214), respectively, of motif S1047. For reading randomized position 521-522, 527-528 and 535 of motif S521, the primer Kpn_I_forw was used. In no case did mutations affect non-randomized nucleotides.

6.7 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.

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.

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

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 U S A 84*:4762-4766.

Jenner LB, Demeshkina N, Yusupova G, Yusupov M. 2010. Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat Struct Mol Biol*.

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.

Lee K, Varma S, SantaLucia J, Jr., Cunningham PR. 1997. In vivo determination of RNA structure-function relationships: analysis of the 790 loop in ribosomal RNA. *J Mol Biol 269*:732-743.

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

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

Morosyuk SV, Lee K, SantaLucia J, Jr., Cunningham PR. 2000. Structure and function of the conserved 690 hairpin in Escherichia coli 16 S ribosomal RNA: analysis of the stem nucleotides. *J Mol Biol 300*:113-126.

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

Santer UV, Cekleniak J, Kansil S, Santer M, O'Connor M, Dahlberg AE. 1995. A mutation at the universally conserved position 529 in Escherichia coli 16S rRNA creates a functional but highly error prone ribosome. *RNA 1*:89-94.

Steinberg SV, Boutorine YI. 2007a. G-ribo motif favors the formation of pseudoknots in ribosomal RNA. *RNA 13*:1036-1042.

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

6.8 Tables

Table 1. Nucleotide sequences and activities of the selected variants of motif S1047

Clone	**0P**	**0Q**	**-1P**	**-1Q**	**-1T**	GFP Activity (%)
A1 (WT)	G	C	G	C	C	100
A2	G	C	G	U	C	100±4
A3	G	U	G	C	C	100±6
A4	G	C	U	G	C	94±5
A5	G	U	G	C	U	93±6
A6	G	C	G	U	U	80±7
A7	G	C	A	U	U	79±6
A8	G	C	A	C	C	78±8
A9	C	G	U	A	U	73±6
A10	G	U	U	A	U	71±5
A11	A	U	A	U	U	68±5
A12	A	C	A	U	U	66±4
A13	A	A	C	G	U	65±7
A14	G	C	G	U	A	64±5
A15	G	C	G	U	G	63±7
A16	G	U	G	U	U	62±5
A17	G	C	A	U	G	61±4
A18	G	C	U	A	G	60±6
A19	A	C	U	A	C	59±5
A20	C	G	U	C	U	59±3
A21	C	G	C	A	U	57±4
A22	U	U	C	G	U	57±8
A23	U	A	C	A	U	53±6
A24	C	A	U	U	U	48±3
A25	A	U	U	U	G	43±5
A26	G	C	G	G	C	40±4
A27	C	A	A	U	U	39±6
A28	G	A	G	C	G	38±2
A29	C	G	G	U	G	36±4
A30	U	G	A	U	G	27±3
A31	A	C	A	C	G	23±4
A32	C	C	U	C	G	21±5
A33	C	A	G	C	G	20±3
A34	U	A	C	A	A	12±2
A35	A	C	U	G	G	6±3
A36	C	G	C	A	G	6±1

Table of identities of nucleotide positions 0P, 0Q, -1P, -1Q and -1T of the selected clones of the library of motif S1047. Clones are ordered in respect to the decrease of their activity. The ribosome activity (GFP) was measured as the mean of six independent experiments. The percentage of each clone's activity in respect to that of WT is presented in the last column "Activity". Clone A1 corresponds to the WT sequence.

Tables 2-9 contain information on particular clone sequences, extracted from the Table 1.

Table 2. Selection of sequences of clones with GC base pair in position [0P; 0Q] and GU base pair [-1P; -1Q] and either uridine or cytosine -1T.

Clone	0P	0Q	-1P	-1Q	**-1T**	Activity	
A1	G	C	G	C	C	**100**	WT
A2	G	C	**G**	**U**	**C**	**100**±4	
A6	G	C	G	U	**U**	**80**±7	

Table 3. Selection of sequences of clones with GC base pair in position [-1P; -1Q] and GU base pair [0P; 0Q] and either uridine or cytosine -1T.

Clone	0P	0Q	-1P	-1Q	**-1T**	Activity	
A1	G	C	G	C	C	**100**	WT
A3	G	U	G	C	**C**	**100**±6	
A5	G	U	G	C	**U**	**93**±6	

Table 4. Selection of sequences of clones with GC base pair in position [0P; 0Q] and GU base pair [-1P; -1Q] and either pyrimidine or purine in position -1T.

Clone	0P	0Q	-1P	-1Q	**-1T**	Activity	
A1	G	C	G	C	C	**100**	WT
A2	G	C	G	U	**C**	**100**±4	
A14	G	C	G	U	**A**	**64**±5	
A15	G	C	G	U	**G**	**63**±7	

Table 5. Selection of sequences of clones with GC base pair in position [0P; 0Q] and AU base pair [-1P; -1Q] and either uridine or guanine in position -1T.

Clone	0P	0Q	-1P	-1Q	**-1T**	Activity	
A1	G	C	G	C	C	**100**	WT
A7	G	C	A	U	**U**	**79**±6	
A17	G	C	A	U	**G**	**61**±4	

Table 6. Selection of sequences of clones with either GC or GU base pair in positions [0P; 0Q] and [-1P; -1Q] and pyrimidine in position -1T.

Clone	0P	0Q	-1P	-1Q	-1T	Activity
A1	G	C	G	C	C	100
A2	G	C	**G**	**U**	C	100±4
A3	**G**	**U**	G	C	C	100±6
A5	**G**	**U**	G	C	**U**	93±6
A16	**G**	**U**	**G**	**U**	**U**	62±5

Table 7. Selection of sequences of clones that do not have guanosine 0P, possess a WC base pair [-1P; -1Q] and pyrimidine in position -1T.

Clone	0P	0Q	-1P	-1Q	-1T	Activity
A9	C	G	U	A	U	73±6
A11	A	U	A	U	U	68±5
A13	A	A	C	G	U	65±7
A12	A	C	A	U	U	66±4
A19	A	C	U	A	C	59±5
A22	U	U	C	G	U	57±8
A27	C	A	A	U	U	39±6

Table 8. Selection of sequences of clones that have a WC base pair [0P; 0Q], a non-canonical base pair [-1P; -1Q] and pyrimidine in position -1T.

Clone	0P	0Q	-1P	-1Q	-1T	Activity
A8	G	C	A	C	C	78±8
A20	C	G	U	C	U	59±3
A21	C	G	C	A	U	57±4
A23	U	A	C	A	U	53±6

Table 9. Selection of sequences of clones that have a non-WC combination in either [0P; 0Q] or [-1P; -1Q] positions and purine in position -1T.

A31	A	C	A	C	G	23±4
A32	C	C	U	C	G	21±5
A33	C	A	G	C	G	20±3
A34	U	A	C	A	A	12±2
A35	A	C	U	G	G	6±3
A36	C	G	C	A	G	6±1

Table 10. Sequences of the oligonucleotides that were used in this study

Name	**Sequence**
Kpn_I_forw	5’- ggggttctcctgagaactccgg-3’
Bgl II_forw	5’-gggtagaattccaggtgtagcggtg-3’
Xba_I_Rev	5’-ggggattcgaacccctgttac-3’
Bgl II_rev	5’- gggggccgccttcgccaccgg-3’
1047_RP1	5'-tctcacggttcccgaaggcac-3’
1047_FP1	5'-gtgccttcgggaaccgtgagaca**NN**tgctgcatggctgtcgtc-3’
1047_RP2	5'-ccatgatgacttgacgtcatcc-3’
1047_FP2	5'-gggatgacgtcaagtcatcatggc**NN**tta**N**gaccagggctacacacg-3’
521_FP	5'-gaagcaccggctaactccgtgcca**NN**agcc**NN**ggtaat**N**cggagggtgcaagc-3’
521_RP	5'-tggcacggagttagccggtgc-3’

Bold N = A, G, T or C.

6.9 Figures

Figure 1

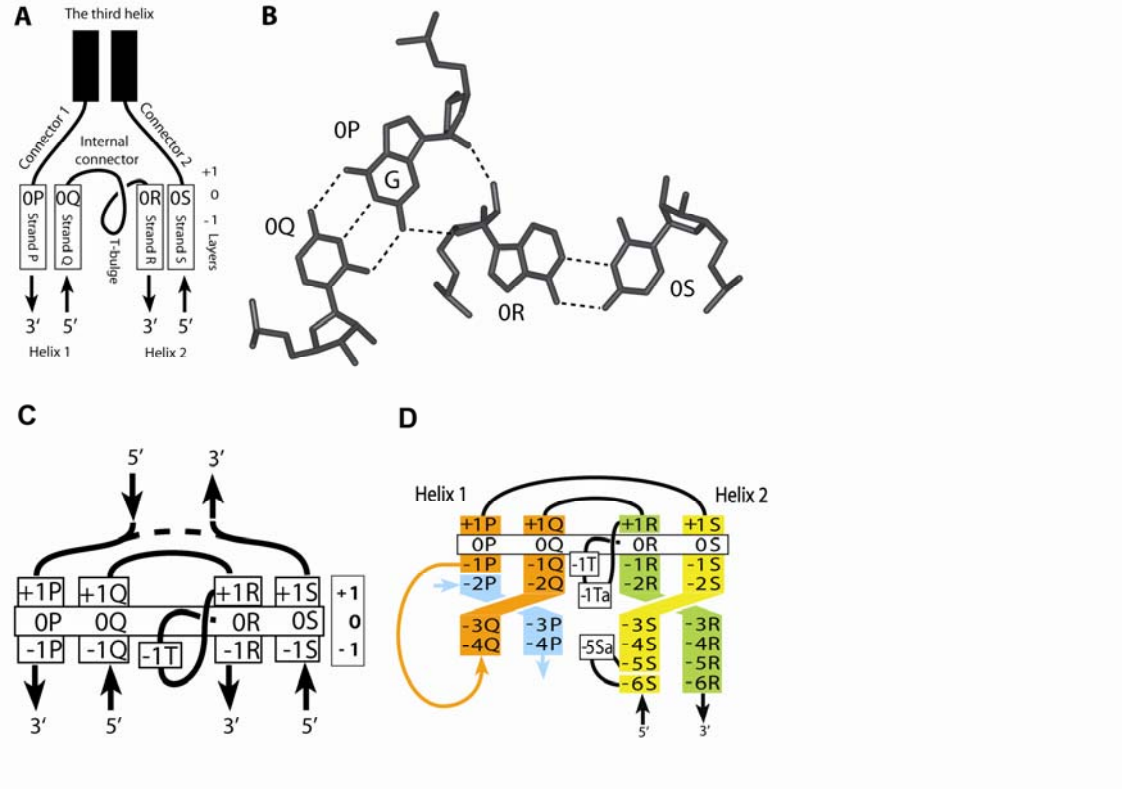


Figure 1. Definition of the G-ribo motif.

(A) The definition of different elements of the G-ribo motif. Rectangles stand for helical strands, while unpaired regions are shown by curves. Helix 1 consists of strands P and Q, while Helix 2 consists of strands R and S. The third helix exists only in some cases of the G-ribo motif. In all cases of the motif, the 3′ part of the internal connector makes a loop that we call the T-bulge. The top base pairs [0P; 0Q] and [0R; 0S] of, respectively, Helices 1 and 2 form the zero-layer. The positions of layers -1, 0 and +1 are shown on the right. (B) The juxtaposition of the two zero base pairs [0P;0Q] and [0R;0S]. Dashed lines stand for hydrogen bonds within and between the base pairs. In this juxtaposition, the ribose of 0R interacts with the ribose and the base of 0P. To make this interaction possible, 0P should be guanosine. (C) A template of the secondary structures of the G-ribo motif with the named nucleotide positions of the layers from -1 to +1. (D) The template used for depicting the secondary structures of the G-ribo-based pseudoknots. The base pairs at the zero-layer are enclosed in the horizontally oriented rectangle. In all pseudoknots, there is a break of the polynucleotide chain in strand P between positions -1P and -2P. Nucleotide -1P is connected to a lower layer of strand Q, thus forming a stem-and-loop structure (orange) in Helix 1 Strands R and S of Helix 2 are, respectively, green and yellow.

Figure 2

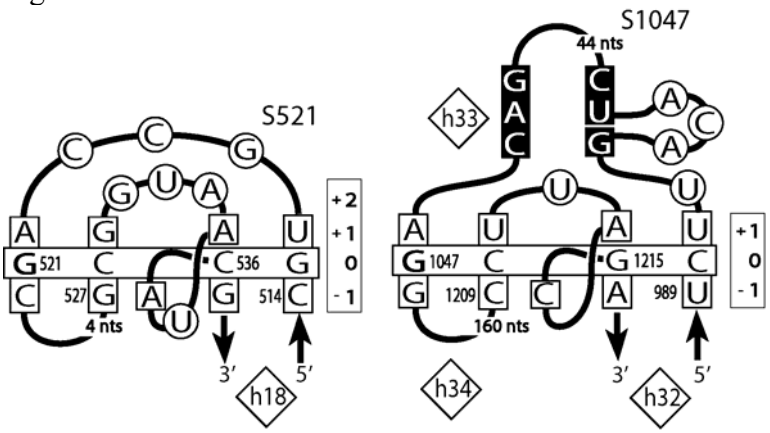
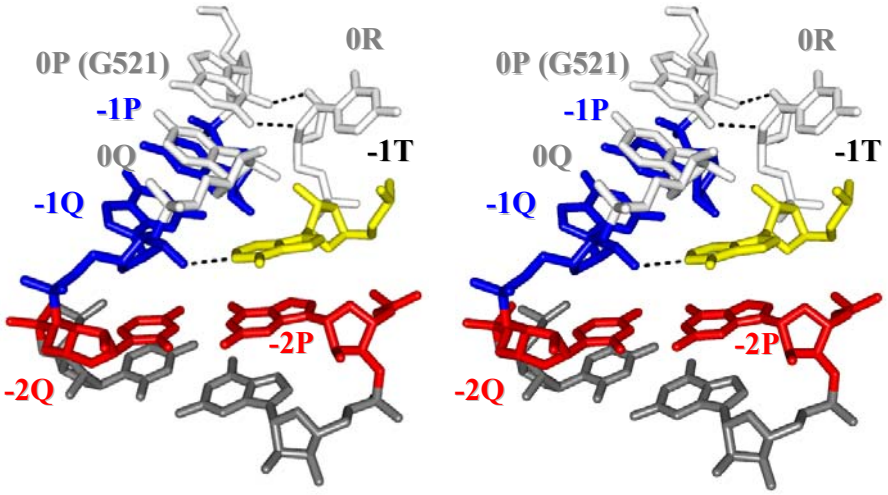


Figure 2. Secondary structures of G-ribo motifs S521 and S1047.

In the name of a motif, the first letter (S or L) stands for the ribosomal subunit, small or large, in which the motif was found. The number in the name corresponds to that of nucleotide 0P in the standard *E.coli* numeration of rRNA. The positions of the layers are shown in the vertically oriented rectangles. The base pairs at the zero-layer are enclosed in the horizontally oriented rectangle. G in position 0P is bold. The nucleotides that stack to those of the zero-layer are squared. The nucleotides of the third helix involved in base pairing are shown on the black background. Other nucleotides are circled. The numbers of the helices in the standard 16S and 23S rRNA secondary structures are shown in diamonds.

Figure 3

A



B

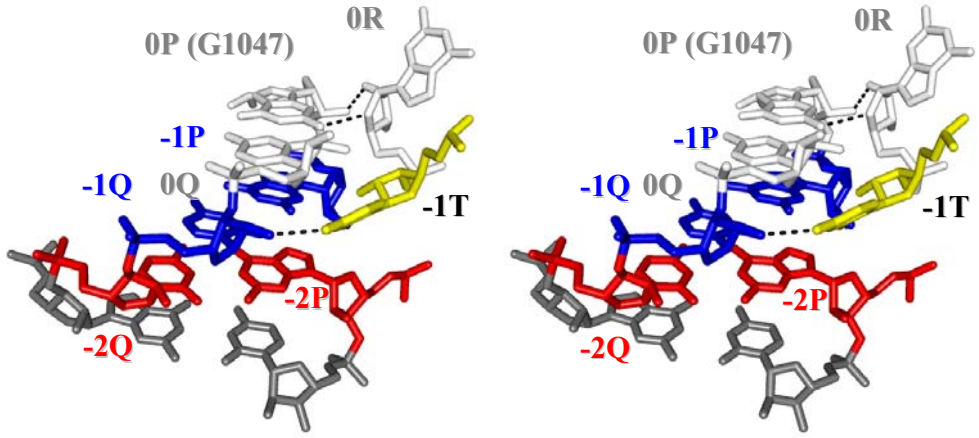


Figure 3. The over-twist between base pairs [-1P;-1Q] and [-2P;-2Q] in motifs S521 (A) and S1047 (B).

A. In S521, the over-twist is substantial, reaching approximately 60˚. It is stabilized by the simultaneous involvement of adenosine -1T (yellow) in A-minor interaction with base pair [-1P;-1Q] and stacking to base pair [-2P;-2Q].

B. In S1047, the over-twist is small and does not exceed 10˚. Pyrimidine -1T (yellow) stacks on the ribose of nucleotide 0Q and due to the small over-twist between base pairs [-1P;-1Q] and [-2P;-2Q] it can not stack to nucleotide -2P, as it was observed in motif S521.

In both structures, the two hydrogen bonds constituting the G-ribo interaction between 0P and 0R are shown with a dotted line.

Chapter 7

Discussion

7. Discussion

Significant improvement in crystallographic and NMR techniques has allowed collection of substantial amounts of information on the RNA structure. Analysis of the available RNA structures shows that apart from the regular A-form duplexes, the polynucleotide chain of RNA is able to form diverse 3D shapes. In contrast to DNA, which mostly forms Watson-Crick base pairs, RNA typically forms non-canonical base pairs via SE and HG edges of ribonucleotides (discussed in section 1.2.2 of Chapter 1). The ability to fully exploit the hydrogen-bonding and base-stacking potentials of bases and ribose moieties, allows RNA to form a large variety of complex structures.

RNA structure can be roughly split in two: one set of RNA structures represented by recurrent arrangements, or motifs and another set made of unique elements, which either occur only once or have not been shown yet to have analogs. Most of the RNA structure is composed of repeating structural elements. Even though the most common recurrent structural element is a WC base pair, which is a building block of the regular A-form double helix, the term "recurrent motif" is mostly used for the designation of the structural arrangements that are built of non-WC or combinations of WC and non-WC base pairs. Several definitions for RNA motif can be found in the literature. Leontis and Westhof define it as "*directed and ordered stacked arrays of non-Watson–Crick base pairs forming distinctive foldings of the phosphodiester backbones of the interacting RNA strands*" (Leontis & Westhof, 2003). Another definition is given by Moore, who considers RNA motif as "*a discrete sequence or combination of base juxtapositions found in naturally occurring RNAs in unexpectedly high abundance*" (Moore, 1999).

The search for new RNA motifs and the analysis of existing motifs provide for better understanding of the principles that govern RNA structure formation and function. The fact that a particular structural arrangement has been selected through evolution for building different segments of a particular RNA molecule speaks for the high importance of this motif and thus can be a target for further detailed analysis. One important feature of RNA motifs is that its overall 3D structure forms independently from its context. The ability to form similar 3D structures by seemingly unrelated RNA sequences suggests the existence of hidden rules which govern the formation of an RNA motif. Identification of

such rules would contribute to a deeper understanding of RNA structure formation. During my PhD studies, my major scientific efforts have been concentrated on the identification and characterization of new RNA motifs.

Our research is primarily based on the ribosome structure. There are a number of reasons for this. First of all, the ribosome is one of the largest RNA-containing molecules. Due to its immense size, it is highly likely to include most known RNA structural motifs (Nagaswamy & Fox, 2002; Krasilnikov & Mondragon, 2003); moreover some recurrent RNA motifs are exclusively found in the ribosome and not in other smaller RNA molecules (Klein et al., 2001; Gagnon & Steinberg, 2002; Lee et al., 2006).

Release of the ribosome structure was a fateful event for many researchers. Previously, gaining insight into the ribosome structure was mainly done through different chemical methods, such as chemical probing, cross-linking and free radical labeling (Fox, 1997; Wilson, 2002; Bockelmann, 2004; Tullius & Greenbaum, 2005). Substantial scientific effort was concentrated on the acquisition of data using such "indirect" techniques. As a result, the moment when the release of the high-resolution structures of the ribosome was made, "indirect" approaches had become less significant for determination of the 3D structure. In contrast, the role of the methods for the RNA structural analysis gained substantial amount of attention.

The 2009 Nobel Prize in Chemistry was awarded to Venkatraman Ramakrishnan, Thomas A. Steitz, Ada E. Yonath, for the achievement of refining the ribosome structure. The ribosome structure represents an indispensable source of information for the understanding of how the ribosome functions. Even though the structure is now known, we are still far from having the detailed picture of the main steps in translation. Currently, a major scientific challenge is to understand the principles that drive this complex machine. Studies of RNA motifs can help decipher the principles of rRNA structure formation and at the same time contribute to the understanding of ribosome function.

Although we were mostly interested in the analysis of the ribosomal structure, some of the motifs, identified in the course of my PhD project, have also been found in other RNA-containing molecules. In order to identify all instances of a particular structural arrangement we have used the FR3D algorithm, which allows the searching of similar arrangements throughout all available RNA structures (Sarver et al., 2008). Prior

to the release of FR3D software in 2008, we used our own algorithm which searched for similarly juxtaposed nucleotides in all available PDB files. This software was developed with help of a former PhD student from our lab, Jiang Hong Chen.

Close to 7 years spent on my Ph.D. project have resulted in the identification of two new recurrent motifs, which are named “G-ribo” and “DTJ” motifs.

7.1 Identification and characterization of new motifs

7.1.1 G-ribo motif

The ribosome structure contains enormous amounts of nucleotides and amino acids, which significantly complicates the analysis of known arrangements and search for new motifs. When the ribosome structures first appeared, it seemed almost impossible to get through the maze of this “structural jungle” of the 16S and 23S rRNA and therefore the analysis of these structures presented a serious challenge for researchers. When looking at the structure and searching for new arrangements, one has to have at least an approximate idea of what to search for. For example when we identified the G-ribo motif, we had a preconceived interest in the identification of alternative ways of packing two helices via the minor groove of one helix and either the minor groove or the sugar-phosphate backbone of another helix. Such an idea came from the analysis of the Along-groove packing motif, which represents a particular strategy of closely packing two helices together via their minor grooves (Gagnon & Steinberg, 2002).

Visual analysis of the high-resolution structures of the 30S and 50S ribosomal subunits allowed identification of a new type of the helix packing. Within this new arrangement the backbone of one helix is packed into the minor groove of another helix. The helix packing is mediated by the hydrogen bonding between the SE of guanosine from one helix and sugar moiety of a nucleotide from another helix, which was named G-ribo interaction. In order to identify all cases of the G-ribo interaction we used a previously developed algorithm that searched for similarly juxtaposed pairs of nucleotides through the whole structure of the ribosome. As a result, we identified 8 cases of the G-ribo interaction in both 30S and 50S ribosomal subunits.

The G-ribo interaction is always accompanied by additional structural elements that are summarized as follows based on the description given in Chapter 2. The G-ribo

interaction is established between two nucleotides, each of which is involved in formation of a WC (or GU in rare cases) base pair. These base pairs are located at the so-called zero layer of the motif-forming helices, H1 and H2. H1 is composed of strands P and Q, while H2 is composed of strands R and S. While the -1 layer of H1 and H2 is made of the WC base pairs, +1 layer is occupied with a non-canonical base pair or a single nucleotide. Positions +1R and 0R are connected with a one- or two-nucleotide bulge, which is referred to as T-bulge, and the last nucleotide of the T-bulge as -1T. In all G-ribo motifs, nucleotide -1T makes A-minor interaction with base pair [-1P; -1Q].

The particular geometry of the G-ribo motif is supported with help of a set of inter-helix contacts at +1, 0 and -1 layers. At the +1 layer of most G-ribo motifs, positions +1P from H1 and +1R from H2 are occupied with adenosine and form either trans-HG-HG, or trans-WC-WC or trans-HG-WC base pair (except for motif L1024 where nucleotides +1P, +1Q and +1R form a triple UGA, supplemental Figure 2 in Chapter 2). At the 0 level of the G-ribo motif the key interaction between guanosine 0P and the ribose 0R is formed, which stabilizes the particular orientation of two motif-forming helices. At the -1 layer, nucleotide -1T, which bulges of the H2, interacts with the minor groove of [-1P; -1Q] base pair. Apart from hydrogen bonding with the minor groove of [-1P; -1Q] base pair nucleotide -1T stacks to ribose of nucleotide 0Q. This interaction additionally stabilizes position of -1T in respect to the H1.

Through the analysis of the G-ribo motif an interesting phenomenon was observed, which has not been discussed in the current literature on RNA structure. The base of nucleotide -1T was found to stack on the sugar of nucleotide 0Q in all cases of the G-ribo motif. Indeed, the stacking interaction is the most efficient between two molecules that possess pi-electron clouds. Although nucleotide bases possess pi-electron clouds, which allow them to efficiently stack on each other, the ribose of a nucleotide does not have any pi-electrons. Yet, analysis of interactions between nucleotide positions -1T and 0Q shows that in all cases of the G-ribo motif the base of -1T is inclined towards the ribose of 0Q in such a way that van der Waals radii of ribose 0Q and base of -1T are closely packed. Interestingly, this phenomenon is widely observed in various structures of RNA molecules and in the ribosome structure in particular, although to date, it has not been reported in literature.

Comparative analysis of eight G-ribo motifs resulted in the identification of several unexpected structural features not included in the simple definition of the motif described previously. Four rRNA pseudoknots in both large and small subunits of the ribosome exploit the G-ribo motif for their formation. Our analysis showed that the structure of the G-ribo motif contains certain elements which stimulate the formation of pseudoknots. For instance, in all G-ribo motifs that are involved in pseudoknots, the polynucleotide chain P is interrupted between positions -1P and -2P, which opens a possibility for the formation of an over-twist arrangement between base pairs at -1 and -2 layers of H1. Such an over-twist appears to be instigated by the presence of bulged nucleotide -1T. The presence of nucleotide -1T favors the displacement of base pair [-2P; -2Q] towards the minor groove of helix 1 in respect to base pair [-1P; -1Q] (Figure 3 of Chapter 3). This is helped by the fact that nucleotide -1T stacks on -2P and since -1T is involved in an A-minor interaction with [-1P; -1Q], it stabilizes the over-twist formation between base-pairs [-1P; -1Q] and [-2P; -2Q]. Thus, nucleotide -1T is important, not only for stabilizing the particular juxtaposition of two motif-forming helices (explained before), but also because it can stimulate the formation of G-ribo-associated pseudoknots depending on its context.

The study of G-ribo motifs has led to the elucidation of the roles that it plays in the functional ribosome. For instance, G-ribo motif S521 appears to play a particularly important role in the ribosome function. The internal connector of this motif is also called and is better recognized as loop 530, which is known to play a critical role in the recognition of cognate tRNA in the A-site of the 30S subunit (Moazed & Noller, 1990). When cognate tRNA enters the A-site of the 30S subunit, nucleotide G530 switches from *anti*- to *syn*- conformation (Ogle et al., 2001). Comparison of structures of the loop 530 in the presence and absence of the cognate tRNA shows that not only G530 changes its conformation but other nucleotides of this flexible loop, U531 and A532 are rearranged as well (Schuwirth et al., 2005; Selmer et al., 2006). The three-nucleotide loop 530 is flanked by G529 on the 5' end and an A533 at the 3' end, each of which is involved in a non-canonical base pair in the +1 layer of motif S521. The conformation of these two nucleotides stays the same in the presence or absence of cognate tRNA in the A-site. Thus, from a structural point of view, the G-ribo motif represents a solid foundation for

the flexible loop 530, which can easily adopt different functionally important conformations. G-ribo motif S521 is also associated with a pseudoknot that additionally solidify the juxtaposition of helices H1 and H2 of motif S521.Thus, the solid nature of this G-ribo motif allows for loop 530 to be accommodated at the A-site of the 30S subunit in a relatively flexible manner while the remaining structure outside of loop 530 remains an unchanged monolith which we think is key to its ability to perform its critical function. In fact, G-ribo-based pseudoknots represent very rigid and compact structures, which fold independently of the rest of the molecule and help direct the folding of the whole ribosome.

G-ribo interaction is not limited to the eight cases of the G-ribo motif identified in the ribosome, as it was described in Chapter 2. Five more cases of the G-ribo interaction were identified in the structure of the ribosome with help of an automated search (Sarver et al., 2008). We refer to these cases as alternative G-ribo motifs. The alternative G-ribo motifs are different from the canonical ones (Figure 1). For example, in 2 out of 5 cases of non-canonical G-ribo motifs the H2 on the 0 and +1 level is reduced to the strand R only. Besides this, the short connector between H1 and H2, which is present in all canonical G-ribo motifs, does not exist in non-canonical G-ribo motifs. Although on the level of the secondary structure these alternative G-ribo motifs differ from canonical ones, some similarities still exist at the tertiary level. Besides the existence of the G-ribo interaction, which is established between nucleotides 0P and 0R, nucleotide -1T also exists. In all these alternative G-ribo motifs, nucleotide -1T precedes nucleotide 0R and stacks to the ribose of 0Q (or +1Q) and in some cases interacts with the minor groove of base pair [-1P; -1Q]. Analysis of these alternative G-ribo motifs suggests that nucleotide -1T is a universal feature for G-ribo interactions in all structural contexts. The presence of -1T can be suggested to be required for the stabilization of a particular superposition of the two G-ribo-interaction-forming nucleotides.

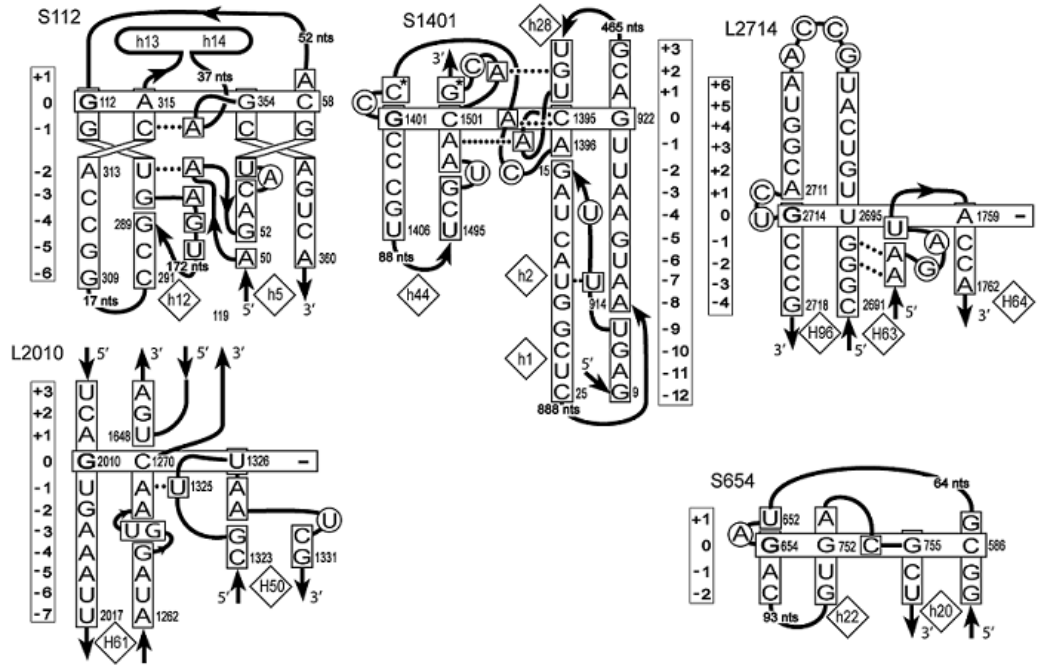


Figure 1. Secondary structures of the alternative G-ribo motifs identified in the *E. coli* ribosome.
The same nomenclature is used as for the secondary structures of the G-ribo motifs in Figure 3 of Chapter 2.

The G-ribo project served as a starting point for another important publication from our lab. Analysis of the G-ribo pseudoknots led to the so-called mushroom theory that provides rationale for how the ribosome evolved. The idea of this theory was that the outer layer of the ribosome consists of structural blocks (or "mushrooms") that can be easily removed as a single piece without major perturbation of the rest of the molecule. Refinement of this idea led to the evolutionary hypothesis, which posits that the ribosome evolved by random insertions of RNA blocks (Bokov & Steinberg, 2009).

This example shows how the meditation on a structural idea, such as the G-ribo pseudoknot, can lead to asking other structure-based questions. The step-wise refinement of the initial relatively simple theoretical idea resulted in the formulation of a major principle for the evolution of the ribosome structure. This is a vivid example of how a theoretical idea in molecular biology can develop.

7.1.2 DTJ-motif

Musing on the structure of the G-ribo motif allowed me to initiate another project, which culminated in the identification of a new recurrent motif. In some G-ribo motifs, the continuous stacking between two consecutive base pairs [-1P; -1Q] and [-2P; -2Q] is lost due to the incorporation of an unpaired region in the strand P. As a result, formation of the helical over-twist between these two consecutive base pairs can be observed (Figure 2). The presence of the over-twisted base pair destabilizes the local structure of the RNA double helix mainly due to the loss of the continuous inter-nucleotide stacking (Figure 2). Therefore, one can expect to find some additional reinforcement measures provided by other elements that are in close vicinity of the helical over-twists. Indeed, in the case of the G-ribo motif, such a stabilizing element is nucleotide -1T, which interacts with the minor groove of base pair [-1P; -1Q] and stacks on the open surface of a nucleotide -2P at the -2 layer (Figure 2).

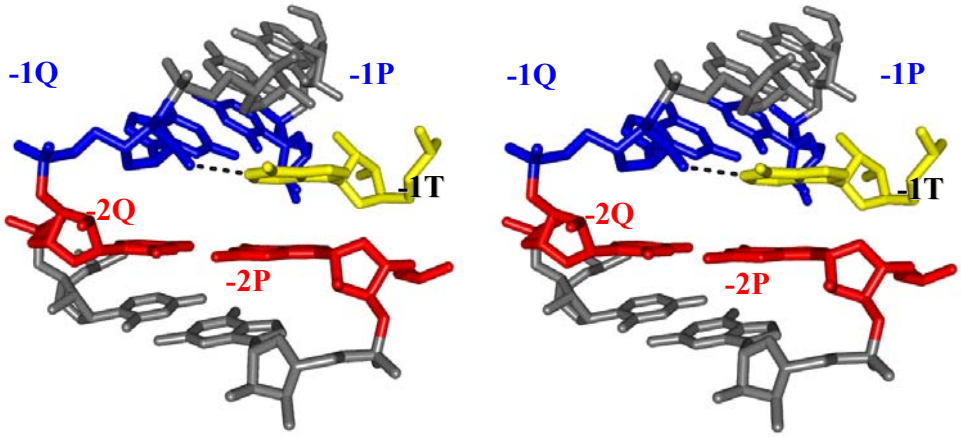


Figure 2. A typical example of the helical over-twist.
Base pairs [-2P; -2Q] and [-1P; -1Q] form a helical over-twist (shown in red and blue respectively). The stacking between nucleotides -2P and -1P is lost. The yellow nucleotide -1T stacks on the red nucleotide -2P and through interaction with the minor groove of the blue base [-1P; -1Q] stabilizes the over-twist between two base pairs in red. G-ribo motif S521 is shown as an example (Steinberg & Boutorine, 2007).

Such a phenomenon where a single nucleotide positions itself in a minor groove to stabilize an over-twisted helix intrigued me. So, I decided to systematically study the principles behind such helical over-twist stabilizations that are independent from the association with the G-ribo motif. I collected all available examples of over-twists in

rRNA and tried to identify the principles for their stabilization. As a result, I selected a few dozen cases of different over-twists that exploited several distinct stabilization strategies. The further analysis of the selected over-twisted base pairs resulted in identification of a group of structural arrangements that consists in two closely located over-twists and can be thought of as a new motif. Basically, one can think of this motif as a combination of three consecutive co-axially stacked double helices, each of which is connected to the neighbouring helix with one or more nucleotide(s) bulged out (Figure 5 in Chapter 4). This definition led to the term "Double Twist Joint", or DTJ to describe such arrangements. Here, a "Twist Joint" refers to the co-axial stacking of two double helices, which are joined through the formation of an over-twist between the last base pair of one helix and the first base pair of another helix.

In total we identified three types of DTJ arrangements, which we call A-, B- and C-DTJs. In the A-DTJ, the central double helix consists of three base pairs while in B- and C-DTJs, it only includes two. Detailed analysis of these three types of DTJ arrangements showed that two different strategies of the over-twist stabilization are exploited. In the A-DTJ, the two over-twists are stabilized by the presence of either one or two nucleotides that make a direct contact with the major groove of the central helix. The role of the major-groove interacting nucleotide can be thought to provide either a stacking platform for the over-twisted base pair (so-called "direct stabilization"), or to tether the neighboring nucleotide that also participates in the over-twist formation in such a way to limit its conformational flexibility (so-called "indirect stabilization", discussed in Chapter 4). In contrast, in the B- and the C-DTJ, both over-twists are mainly stabilized through the formation of a dinucleotide stack that positions itself in the major groove of the central helix (discussed in Chapter 4).

The bending and twisting of a double helix is unavoidable for establishing long-range interactions in the large RNA molecules, such as rRNA. Such modulation of the helical domain is achieved by many ways as seen in the examples that have a mono- or multi-nucleotide bulge, C-loop, UAA/GAN and many other motifs (discussed in sections 1.3.5 and 1.3.6 of Chapter 1). The introduction of such motifs can modify the geometrical properties of the RNA double helix and can allow fine-tuning of the local RNA context to fit to the global needs of the structure. For instance, the insertion of a single or several

nucleotides in one of the strands of the RNA double helix results in the formation of a bulge (Figure 10A in Chapter 1). In some situations, the bulge may facilitate the formation of a helical over-twist between two consecutive base pairs that can be twisted by up to ~60° (Figure 10C, D, E in Chapter 1). In the C-loop motif the over-twist between two consecutive base pairs may reach up to ~90° (Lescoute et al., 2005). Three types of the DTJ arrangements have different effects on the structure of the double helix. All three types of DTJ motifs speed up rotation of the double helix by ~70-80°. The presence of the DTJ motif results in the helix bend of ~15° for A-type motifs and ~30° bend for B- and C-type motifs. (Supplemental Figure 11 in chapter 4). Thus, DTJ motifs are important elements that mediate the correct RNA structure formation.

Apart from the purely structural role that DTJs play in RNA folding, two DTJ motifs are involved in the formation of the functional centers of the ribosome and RNAseP. In motif 2AVY-1057 an unpaired nucleotide (C1054) from the bulge loop L2 participates in the coordination of the codon-anticodon mini-duplex formed between the tRNA and the mRNA in the ribosomal A-site (Jenner et al., 2010). In motif 1NBS-232 an unpaired nucleotide (A230) of bulge loop L2 participates in the stabilization of the complex between RNAseP and the pre-tRNA substrate (Loria & Pan, 1997; Krasilnikov et al., 2003; Kazantsev et al., 2005).

Preliminary analysis shows that numerous examples of the helical over-twist can be identified in the structure of the ribosome as well as other RNA molecules. Practically all of these over-twists are stabilized in one of the ways mentioned above. Subsequent studies that apply these principles for the over-twist stabilization, which were developed based on the analysis of DTJ arrangements, should prove to be useful for further classification of other RNA helical over-twists seen in RNA structure.

7.2 *In vivo* studies of the motifs

Identification of motifs in the crystal structure of RNA molecules and their comparative sequence analysis is the first step for understanding the principles of RNA structure formation. The application of experimental systems that test mutant RNA molecules for functional efficiency can serve as a powerful way to study the RNA motifs.

Modification of a single nucleotide in a given motif at a time (rational design) however, does not guarantee selection of the functional clones and as a result can prove to be an unproductive exercise. However, the simultaneous randomization of several nucleotide positions (combinatorial library) that form the motif and then selection of functional variants *in vivo* can be more fruitful. The major advantage of the combinatorial approach is that the selection itself is not influenced by the researcher's biases and therefore can in most instances select unexpected nucleotide combinations. This can serve as a source of new observations that can help decipher the principles for motif formation. This combinatorial approach is also known as "instant evolution". In order to study structure-function relationships in the ribosome we randomized key nucleotide positions of a given RNA motif. The selection of active clones and the analysis of their sequences allowed us to identify the sequence constraints imposed on the structure of the motifs for *in vivo* functioning molecules.

7.2.1 Functional characterization of the motifs

The use of combinatorial gene libraries is a powerful method of studying RNA motifs in their natural environment. Even if a motif under consideration is not involved in the formation of a particular functional centre of the ribosome, we assume that its proper assembly is indispensable for correct folding of the ribosome. As a result we presume that the selection of functional clones guarantees the formation of the near-Wild-Type type geometry in the motif in spite of the fact that certain nucleotide positions in the motif structure have non-Wild-Type identities. The latter assumption is extremely important for further sequence analysis of the motif and for molecular modeling of the selected sequences.

Structure-functional relationships for AGPM and the G-ribo motif were studied using the bacterial ribosome knock-out strain and also the specialized ribosome system. We introduced randomized nucleotides into key positions in these motifs and selected for functional clones. Each selected clone had an alternative sequence of the rRNA to that of the Wild-Type. Analysis of the sequences of the selected clones (alternative sequences, or sequence variants) entailed structural interpretation, accompanied with the determination of relative functional activity of each selected clone.

Before even approaching an understanding of the structural enigmas, an enormous amount of experimental work is required. One of the most discouraging and difficult parts of the experimental work was the numerous failed attempts at cloning the rRNA libraries in the knock-out strain. While dealing with this strain we faced numerous problems including the genetic recombination and rejection of newly introduced modifications.

7.2.2 AGPM

Along-Groove packing motif (AGPM) is a new RNA structural motif, which was recently discovered by our lab (Gagnon & Steinberg, 2002). A brief description of the motif is given in section 1.3.6.4 of Chapter 1. The motif consists of two double helices that are closely packed via their minor grooves. For most AGPMs, the central base pairs of the two motif-forming helices are occupied with a GU base pair in one helix and a WC base pair (mostly GC base pair) in the other (Gagnon & Steinberg, 2002). Previous *in vivo* studies utilizing combinatorial libraries of the motif SU296 from the 30S subunit demonstrated that the central base pairs GU and GC are inter-changeable, thus providing for a new type of nucleotide co-variation (Gagnon et al., 2006). Studies of the central nucleotide positions of AGPM with the combinatorial method proved to be successful for structure analysis and thus we decided to enlarge the scope of study in these motifs.

Analysis of a single representative of the recurrent motif can give a lot of useful information about the principles for motif formation. As long as each instance of the motif is located in its unique structural context, analysis of the isolated case can be significantly biased due to the local RNA-RNA or RNA-protein contacts. A systematic application of the combinatorial-library approach to different examples of a recurrent motif will provide for each example a set of acceptable nucleotide sequences (sequence variants). Analysis of these nucleotide sequences will result in elucidation of the variability patterns for each case of the motif. Comparison of such patterns for different cases of the motif allows identification of the genuine structural requirements for the formation of the motif and of the additional constraints imposed by the context within which each particular example of the motif exists in the whole RNA structure.

Following this logic, we randomized the central positions of three Along-Groove Packing motifs and obtained three sets of sequence variants. Analysis of these sets showed that the central base pairs of the AGPM can not be modified beyond a certain extent without major consequences to the ribosome structure (and subsequently, function). Identification of the limits of acceptable modification of the central base pairs constitutes the minimal structural requirements for the AGPM formation. Our results show that the minimal requirement for the AGPM to form is the interaction of the central base pair of one helix with the internal nucleotide of the opposite base pair. For such an interaction to happen the central nucleotide position of at least one helix must form a canonical WC base pair (or wobble GU), while the central positions of the opposite helix can be occupied with nucleotides that form a non-canonical base pair. Thus the minimal requirement for motif formation is the presence of a base triple, which is formed between the internal nucleotide of one helix and the minor groove of a stable (canonical) base pair in the central position of the opposite helix. Interestingly, results of all three libraries showed the same pattern for the minimal nucleotide requirements as described, thus characterizing the intrinsic properties of the motif itself.

Although the existence of GU and WC base pairs at the 0 level is important for integrity of the AGPM, the complete contact zone of the two helices includes nucleotides located at levels +1, -1 and -2. In fact, a more detailed analysis of the AGPM structure showed that the two helices touch each other at three spots, which are represented by the pair-wise packing of riboses of nucleotides +1Q and +1S, -2P and -1S, -1Q and -2R. The first nucleotide in each pair comes from the GU-helix and the second nucleotide comes from the WC-helix (Figure 3). These interactions involve mostly the hydrophobic surfaces of the ribose. The presence of the ribose-ribose interactions alone seems to be able to support the integrity of the motif since variants that lack the proper central base pairs as in the wild-type GU-WC combination were still selected *in vivo*. Even though the ribose-ribose interactions lack specificity, we presume that they are responsible for the correct positioning of one motif-forming helix in respect to the other.

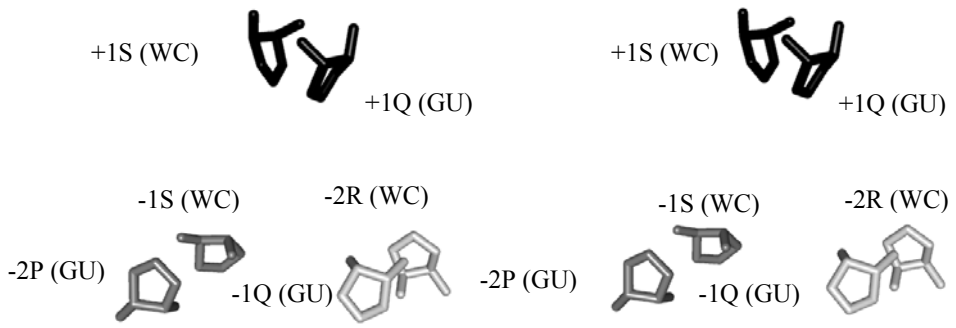


Figure 3. Ribose-ribose contacts in AGPM.
Inter-ribose contacts of nucleotide +1Q and +1S are black, -2P and -1S are gray, -1Q and -2R are white. Nucleotide positions that belong to GU and WC-containing helices are annotated in brackets respectively.

Having in mind the idea of the minimal requirement for the central base pairs for the general formation of AGPM helped us to understand the role the local context plays in each instance of AGPM in the ribosome. For example, analysis of the sequences from the selected variants that were randomized in motif S296 showed that a non-canonical base pair can be accepted in the central position of any of the two motif-forming helices. In contrast to this, motif L639 clearly shows that a non-canonical base pair occurs with relatively high frequency in the central position of only one of the two motif-forming helices. To explain such a variation, further analysis of the structure showed that in case of S296, the central base pairs of the helices are not involved in any RNA-protein or RNA-RNA interactions, while in motif L639, one of the motif-forming helices makes contact with a ribosomal protein. As a result, this context dependent protein interaction imposed certain limitations which translate into requirement for the existence of a stable base pair in one of the helices. Thus, analysis of the sequence variants of three AGPMs, which are located in different parts of the ribosome, allowed us to discriminate between the minimal requirements for the motif formation and the requirements imposed by the structural context of each motif.

Although existence of the ribosomal 3D structures can help in better understanding of structure-functional relationships within the ribosome, I believe that *in vivo* combinatorial mutagenesis provides important information on the principles of the structure formation and function that can not be obtained otherwise. Despite modern

success in crystal structure determinations, it only represents a snapshot of a single functional state of a molecule. If molecules, as we understand, make multiple conformational changes during its functional cycle, dynamic formation and dissociation of particular domains in the molecule is extremely hard, if possible at all, to visualize with the currently available methods. Thus, even if in the crystal structure the studied region is not involved in any substantial RNA-RNA or RNA-protein interactions, formation of such contacts can potentially occur in the course of the functional cycle. Indeed, results from the selection of combinatorial gene libraries are dependent on all intermediate functional states and even folding intermediates, thus reflecting the true requirements for an *in vivo* functioning molecule.

As shown above, *in vivo* studies of the AGPM demonstrate the power of the instant evolution approach. This method allows for the selection of never before seen functional variants of the ribosome which would have been inevitably eliminated in the course of evolution. Therefore, the identification of the minimal requirements for a particular motif would not be ever possible based on the analysis of the collection of conserved nucleotides from the existing phylogenetic data. Thus the instant evolution approach can overcome such an obstacle and allows for the selection of functional ribosomes with a wide range in sequence variability in comparison with what we can see from the existing phylogenetic data.

7.2.3 In vivo study of G-ribo motif

A similar *in vivo* approach as mentioned above was applied to the study of the G-ribo motif. In order to make a comparative sequence analysis, we randomized the central nucleotide positions of two G-ribo motifs in the small ribosomal subunit. Analysis of the sequence variants allowed us to make some important conclusions on the principles of the G-ribo motif formation. As in AGPM, we identified the minimal requirements for the central nucleotide positions of the G-ribo motif. We showed that the existence of a canonical WC (or wobble-GU) base pair at either 0 or -1 layer of helix 1 constitutes the minimal requirement for the motif to form. Another interesting observation is that the absence of the G-ribo interaction does not completely inactivate the ribosome. Although

G-ribo interaction is an important element of the motif, substantial number of clones that lack the G-ribo interaction were selected.

Previously, in the case of AGPM, we discussed the minimal requirements for motif formation where two central base-pairs were described to be important. However, our library results show that the introduction of a non-canonical base pair is tolerated in one of the two central positions in the motif-forming helices. Thus, one can suggest the existence of additional structural elements, which allow for the formation of the structure despite having only one stable base pair in the central position in one of the two helices. Indeed, in AGPM the ribose-ribose contacts of nucleotides at +1, -1 and -2 layers of both motif-forming helices were also found to play an important role in the formation of the motif. Our results from the G-ribo motif study, in a similar way to the above AGPM example, show that despite the absence of the G-ribo interaction, the motif can still be formed. Based on the analysis of the variants of the G-ribo motif we showed that in order for the ribosome to be functional, either [0P; 0Q] or [-1P; -1Q] base pair of the G-ribo motif must make a canonical base pair. We suggest that even if G-ribo interaction is not established, as long as the interaction of nucleotide -1T through H-bonds with the minor groove of [-1P; -1Q] and stacking to ribose 0Q is maintained, the ribosome keeps its function. Thus, the interaction of -1T with H1 appears to be sufficient for G-ribo motif formation in the ribosome. As in the case of the AGPM (where ribose-ribose interactions seem to support the geometry of the motif even in the absence of the stable base pair in one of the central positions of the motif-forming helices), the stacking interaction of ribose-to-nucleotide in the core of the G-ribo motif seem to be an indispensable element that stabilizes the wild-type geometry of the motif even in the absence of the G-ribo interaction.

Another important observation about the G-ribo motif pertains to the identity of -1T nucleotide position. We showed that a pyrimidine in the -1T position is accepted in those G-ribo motifs which normally have A, suggesting that a pyrimidine in this position can be tolerated in all G-ribo instances. Indeed, in motif S1047, -1T position is occupied by a pyrimidine in 99.8% of the existing cases (Supplemental Table 1 in Chapter 2). In other motifs (L2323, L2383, L1309 and L1642) however, position -1T is occupied by a pyrimidine in 4 to 25% of all available nucleotide sequences. Analysis of the G-ribo

motif structure shows that the principal role of nucleotide -1T is to stack on the ribose of nucleotide 0Q and make an A-minor like interaction with base pair [-1P; -1Q]. In fact, uridine, cytosine as well as adenosine can comply with these simple requirements. The question arises however, why in most of the G-ribo motifs adenosine is preferred over a pyrimidine? Our results show that the identity of nucleotide -1T is mainly determined by what happens below level -1 via the participation of helix 1 of the motif. In all pseudoknot-associated G-ribo motifs, the nucleotide position -2P is significantly shifted towards the minor groove of Helix 1 due to the formation of the large helical over-twist between base pairs [-1P; -1Q] and [-2P; -2Q]. The presence of an adenosine and not pyrimidine in position -1T would help stabilize such an over-twist. The bigger surface of the purine in -1T would be favorable since it can stack better with -2P. This could explain why in the G-ribo motifs involved in pseudoknot formation, the identity of nucleotide -1T is largely conserved as adenosine (Supplemental Table 1 in Chapter 2). In contrast, those motifs which do not have a large over-twist between the -1 and -2 layers of helix 1, the presence of the large adenosine can be disadvantageous. Indeed, motif S1047 demonstrates that a pyrimidine is highly welcomed in -1T, while forcing an adenosine in this position results in decreased ribosomal activities (discussed in Chapter 6)

In fact, stacking of base of -1T on ribose of 0Q fixes the position of nucleotide -1T in respect to H1, which subsequently stabilizes position of H2 in respect to H1, as long as -1T is covalently attached to 0R. Thus, base-to-ribose interaction is very important for stabilization of a particular geometry of the G-ribo motif. Interestingly, in both AGPM and G-ribo motifs, ribose-ribose and ribose-to-base interactions play an important structure-forming role. Compared to the inter-nucleotide contacts, interactions that involve ribose lack sequence-specificity. Yet, examples of AGPM and G-ribo motifs show that ribose-mediated contacts can serve as an important structure-forming element.

At least in some other known RNA structural motifs, ribose-to-base interactions play an important role. Thus, analysis of all A-minor interactions identified in the ribosome showed that most of them have base-to-ribose stacking (Figure 4A). Stacking of adenosine to ribose provides for extra stability of the A-minor motif due to formation of additional hydrogen bonds between adenosine and SE of the receptor nucleotide. A special role of the base-to-ribose stacking can be seen in the case of multiple adenosine

stacks. This type of interaction was first detected in the crystal of the hammerhead ribozyme, where three stacked adenosines from a tetraloop belonging to one molecule were packed into the minor groove of a double helix belonging to another molecule (acceptor helix) (Figure 4B) (Pley et al., 1994a, b). This type of interaction was named tetraloop receptor and later, numerous instance of this motif were identified in various RNA structures (Cate et al., 1996). We now can say that the tetraloop receptor is only one type of the arrangement of a poly-adenine stack in the minor groove of a double helix. In the known RNA structures, there are many examples of two, three and even four consecutively stacked adenines docked into the minor groove of double helices, which also can be considered as a stack of consecutive A-minor motifs (Bokov & Steinberg, 2009). In fact, a detailed analysis of several A-minor stacks shows that ribose-to-nucleotide stacking can serve as a "stopper" that prevents sliding of the adenosines along the minor groove of the helix and thus can be considered as an important component of the motif. Another potential structural role of the ribose-to-adenosine stacking would consist in shaping of the RNA structure. Stacking of adenosine to ribose fixes the whole multi-adenosine stack with respect to the minor groove of the acceptor helix, thus contributing to the folding of the whole domain. In Figure 4C, an example of such structure is shown. Here, a stack of two adenosines located in the internal loop of helix 81 of 23S rRNA is packed into the minor groove of helix 39. The inter-domain interaction between helices 81 and 39 is shaped by the stacking of adenosine A2273 present in the donor helix to the ribose of C965 located in the acceptor helix.

Another motif whose integrity depends on the ribose-to-base stacking is the UNCG tetraloop (Figure 4D). In this motif, the ribose of the second last nucleotide of the loop stacks to the base of the last nucleotide of the tetra-loop, thus making allowing the formation of a relatively short connector between the two strands of the double helix.

A ribose-ribose contact can also be seen in the kink-turn motif. Here, two double helices are positioned closely to each other, so that their backbones get in touch with each other (Figure 4E). Interestingly, the interaction of the backbones is mediated by ribose-ribose packing, which is similar to that observed in the AGPM.

Visual analysis of RNA structure shows that ribose-ribose, ribose- base and even ribose- amino acid contacts are widely represented in RNA molecules, where they

mediate numerous RNA-RNA and RNA-protein contacts (Figure 4F, G). Stacking of a base to a ribose seems to be an important factor that directs RNA folding and contributes to the overall stability of RNA molecules. Further analysis of the ribose-mediated interactions is needed to establish the particular rules that govern the formation of these interactions.

My PhD thesis was dedicated to the identification and characterization of new recurrent RNA motifs. RNA is forming large and complex structures that can be hard to analyze at large. Analysis of the constituent building blocks (or motifs) of RNA can be considered as an easier way of analysis of complex molecules. Identification and systematic analysis of new recurrent motifs is thus important for understanding of the principles of RNA structure formation. Development of new methods of RNA motif search and characterization constitutes a necessary step on the way to deciphering enigmas of the RNA folding. Because 3D structure of biological molecules determines its function, detailed analysis of the RNA motifs is also important for better understanding of the principles of the RNA molecule function. I hope that the results of my work have contributed to the general knowledge on the RNA structure and provided a new vision on the principles of identification and characterization of recurrent RNA motifs.

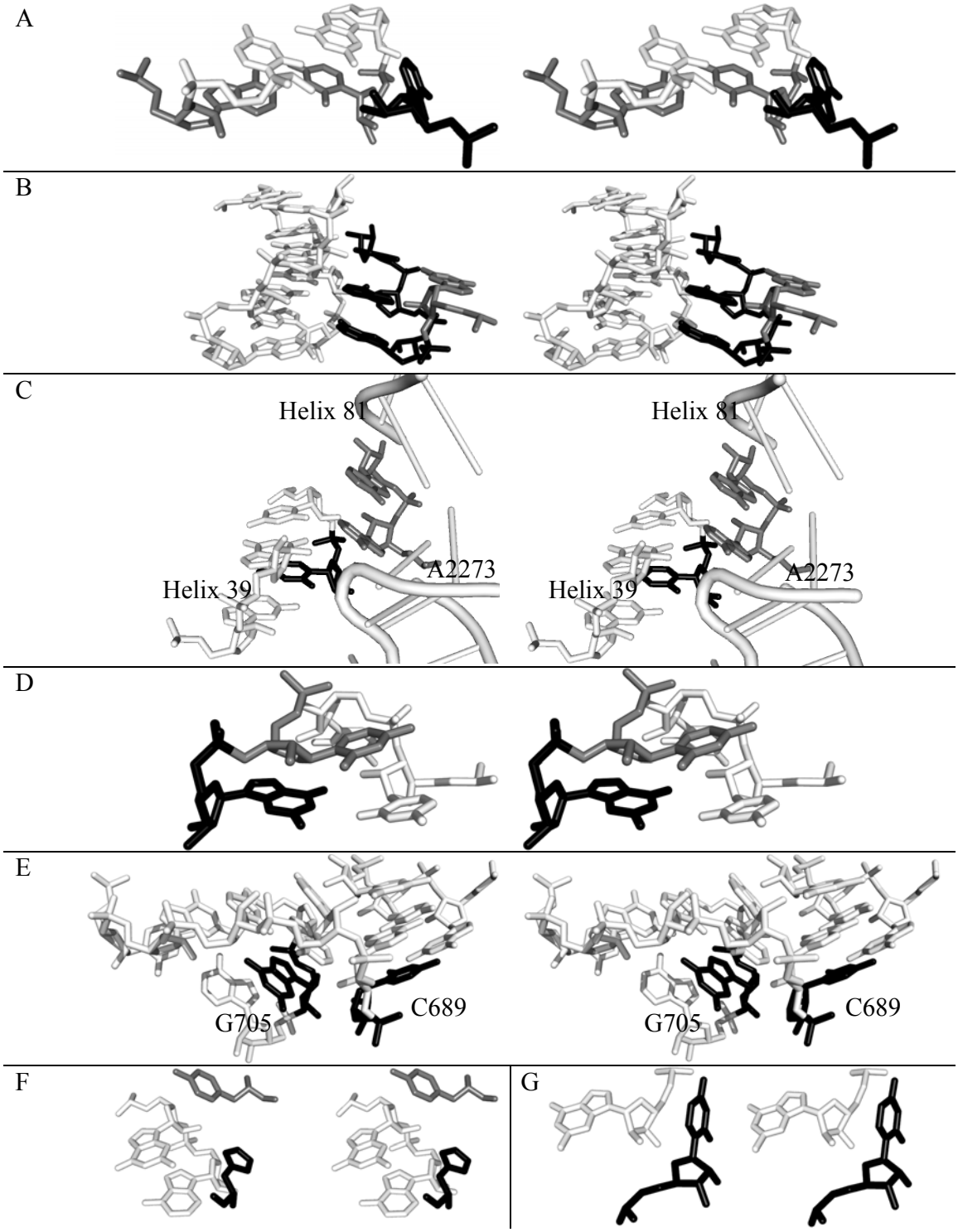


Figure 4. See legend on the next page

Figure 4. Variety of interactions that involve ribose of nucleotide

A. An example of the A-minor motif, where adenosine is closely packed with ribose of the neighboring nucleotide. Adenosine A1101 (black) interacts with the minor groove of base pair G1047-U1083 (white) and stacks on the ribose of nucleotide U1075 (gray). A fragment of the 30S ribosomal subunit is presented (PDB access code - 2AVY).

B. GAAA tetraloop-receptor motif from the crystal of the hammerhead ribozyme (Pley et al., 1994b) (PDB access code - 1HMH). A stack of three consecutive adenosines (shown in black) interacts with the minor groove of an RNA double helix (shown in white), forming three consecutive A-minor motifs.

C. An example of the ribose-to-adenine stacking that plays important role for coordination of the inter-domain folding within rRNA. The base of adenosine A2273 from helix 81 of 23S rRNA (shown in gray) stacks to the ribose of cytosine C965 from helix 39 of 23S rRNA (shown in black). Adenosine A2273 initiates a stack of two adenosines (A2273, A2274, both shown in gray) that shapes the whole inter-domain contact zone. All nucleotides of helix 81 except for two adenosines are shown in bars and ribbons for simplicity. A fragment of the 50S ribosomal subunit is presented (PDB access code - 2AW4).

D. An example of internal stabilization of the UNCG tetraloop by stacking of the base of one nucleotide (G156, shown in black) on ribose of the previous nucleotide (C155, shown in gray) in the strand. A fragment of the specificity domain of Ribonuclease P RNA structure is presented (PDB access code – 1NBS).

E. An example of the kink-turn motif. The two motif-forming helices closely pack together in such a way that riboses of nucleotides G705 and C689 (shown in black) are stacked to each other. A fragment of the 30S ribosomal subunit is presented (PDB access code - 2AVY).

F. Example of RNA-protein contact, where aromatic amino acid residues interact with ribose of nucleotide. Tyr 69 (gray) stacks to ribose of G574 while His 117 (black) stacks to ribose of nucleotide A675. A fragment of the 30S ribosomal subunit is presented (PDB access code - 2AVY).

G. An example of the close packing of uridine and ribose of a nucleotide. U451 (black) stacks to ribose of G36 (white). A fragment of the 50S ribosomal subunit is presented (PDB access code - 2AW4).

Sections C and D demonstrate not yet classified examples of the ribose-to-nucleotide and ribose-to-amino acid residue types of interactions, which are wide spread in the available structures of RNA molecules.

Chapter 8

Conclusions

8. Conclusions

- Two new RNA motifs, G-ribo and Double-Twist Joint (DTJ), have been identified in the ribosome and several other RNA-containing molecules. Both of these recurrent RNA arrangements are important elements mediating RNA structure formation. Examples of both motifs have been identified in functionally important centers of the ribosome and RNAseP.
- It was demonstrated that the G-ribo motif possesses certain structural features, which stimulate formation of the pseudoknot structures in both, 16S and 23S rRNA.
- A new *in vivo* approach for analysis of recurrent RNA motifs was developed. Application of this approach allowed us to identify the minimal structural requirements for both, AGPM and G-ribo motifs and to determine how the local context, including the influence of local proteins, affects the requirements for proper RNA motif formation.
- The work done provides new knowledge of the principles of RNA structure formation and opens new perspectives for identification and characterization of novel RNA motifs.

9. References (Introduction and Discussion)

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.

Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA. 2004. Crystal structure of a self-splicing group I intron with both exons. *Nature 430*:45-50.

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.

Allen E, Xie Z, Gustafson AM, Carrington JC. 2005. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell 121*:207-221.

Aravind L, Watanabe H, Lipman DJ, Koonin EV. 2000. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc Natl Acad Sci U S A 97*:11319-11324.

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 U S A 96*:1971-1976.

Ban N, Freeborn B, Nissen P, Penczek P, Grassucci RA, Sweet R, Frank J, Moore PB, Steitz TA. 1998. A 9 A resolution X-ray crystallographic map of the large ribosomal subunit. *Cell 93*:1105-1115.

Ban N, Nissen P, Hansen J, Capel M, Moore PB, Steitz TA. 1999. Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit. *Nature 400*:841-847.

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

Barron LD, Bogaard MP, Buckingham AD. 1973. Raman scattering of circularly polarized light by optically active molecules. *J Am Chem Soc 95*:603–605.

Barron LD, Buckingham AD. 1975. Rayleigh and Raman Optical Activity. *Annu Rev Phys Chem 26*:381-396

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell 116*:281-297.

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

Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, Sharon E, Spector Y, Bentwich Z. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet 37*:766-770.

Berg JM, Tymoczko JL, Stryer L, National Center for Biotechnology Information (U.S.). 2002. *Biochemistry*. New York

Bethesda, MD: W.H. Freeman ;

NCBI.

Beringer M, Adio S, Wintermeyer W, Rodnina M. 2003. The G2447A mutation does not affect ionization of a ribosomal group taking part in peptide bond formation. *RNA 9*:919-922.

Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B. 1992. The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys J 63*:751-759.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Res 28*:235-242.

Biophysical Society (1st symposium : 1958 : Cambridge Mass.), Roberts RB, Washington Academy of Sciences. 1958. *Microsomal particles and protein synthesis; papers presented at the First Symposium of the Biophysical Society, at the Massachusetts Institute of Technology, Cambridge, February 5, 6, and 8, 1958*. New York: Published on behalf of the Washington Academy of Sciences Washington D. C. by Pergamon Press.

Bockelmann U. 2004. Single-molecule manipulation of nucleic acids. *Curr Opin Struct Biol 14*:368-373.

Boehringer D, Thermann R, Ostareck-Lederer A, Lewis JD, Stark H. 2005. Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure 13*:1695-1706.

Bokov K, Steinberg SV. 2009. A hierarchical model for evolution of 23S ribosomal RNA. *Nature 457*:977-980.

Bouvier NM, Palese P. 2008. The biology of influenza viruses. *Vaccine 26 Suppl 4*:D49-53.

Brennecke J, Stark A, Russell RB, Cohen SM. 2005. Principles of microRNA-target recognition. *PLoS Biol 3*:e85.

Brierley I. 1995. Ribosomal frameshifting viral RNAs. *J Gen Virol 76 (Pt 8)*:1885-1892.

Brodersen DE, Clemons WM, Jr., Carter AP, Wimberly BT, Ramakrishnan V. 2002. Crystal structure of the 30 S ribosomal subunit from Thermus thermophilus: structure of the proteins and their interactions with 16 S RNA. *J Mol Biol 316*:725-768.

Brooks CL, Karplus M, Pettitt BM. 1988. *Proteins : a theoretical perspective of dynamics, structure, and thermodynamics*. New York: J. Wiley.

Brosius J, Dull TJ, Noller HF. 1980. Complete nucleotide sequence of a 23S ribosomal RNA gene from Escherichia coli. *Proc Natl Acad Sci U S A 77*:201-204.

Brown JW. 1999. The Ribonuclease P Database. *Nucleic Acids Res 27*:314.

Bumcrot D, Manoharan M, Koteliansky V, Sah DW. 2006. RNAi therapeutics: a potential new class of pharmaceutical drugs. *Nat Chem Biol 2*:711-719.

Carell T, Wintner EA, Sutherland AJ, Rebek J, Jr., Dunayevskiy YM, Vouros P. 1995. New promise in combinatorial chemistry: synthesis, characterization, and screening of small-molecule libraries in solution. *Chem Biol 2*:171-183.

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.

Carthew RW, Sontheimer EJ. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell 136*:642-655.

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.

Cech TR. 1990. Self-splicing of group I introns. *Annu Rev Biochem 59*:543-568.

Chen X, Kang H, Shen LX, Chamorro M, Varmus HE, Tinoco I, Jr. 1996. A characteristic bent conformation of RNA pseudoknots promotes -1 frameshifting during translation of retroviral RNA. *J Mol Biol 260*:479-483.

Clemons WM, Jr., May JL, Wimberly BT, McCutcheon JP, Capel MS, Ramakrishnan V. 1999. Structure of a bacterial 30S ribosomal subunit at 5.5 A resolution. *Nature 400*:833-840.

Cohen SB, Graham ME, Lovrecz GO, Bache N, Robinson PJ, Reddel RR. 2007. Protein composition of catalytically active human telomerase from immortal cells. *Science 315*:1850-1853.

Correll CC, Freeborn B, Moore PB, Steitz TA. 1997. Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell 91*:705-712.

Crick F. 1970. Central dogma of molecular biology. *Nature 227*:561-563.

Crick FH. 1966. Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol 19*:548-555.

Crozier J. 2006. *Collins English dictionary*. Glasgow: HarperCollins.

D'Alessio G, Riordan JF. 1997. *Ribonucleases : structures and functions*. San Diego: Academic Press.

Dahlquist KD, Puglisi JD. 2000. Interaction of translation initiation factor IF1 with the E. coli ribosomal A site. *J Mol Biol 299*:1-15.

DaRocha WD, Otsu K, Teixeira SM, Donelson JE. 2004. Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in Trypanosoma cruzi. *Mol Biochem Parasitol 133*:175-186.

de Fougerolles A, Vornlocher HP, Maraganore J, Lieberman J. 2007. Interfering with disease: a progress report on siRNA-based therapeutics. *Nat Rev Drug Discov 6*:443-453.

Dieckmann T, Suzuki E, Nakamura GK, Feigon J. 1996. Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA 2*:628-640.

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

Dreher TW, Miller WA. 2006. Translational control in positive strand RNA plant viruses. *Virology 344*:185-197.

Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP. 2009. RNAi in budding yeast. *Science 326*:544-550.

Egli M, Minasov G, Su L, Rich A. 2002. Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution. *Proc Natl Acad Sci U S A 99*:4302-4307.

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res 15*:9109-9128.

Farabaugh PJ, Bjork GR. 1999. How translational accuracy influences reading frame maintenance. *EMBO J 18*:1427-1434.

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

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature 391*:806-811.

Fox KR. 1997. *Drug-DNA interaction protocols*. Totowa, NJ: Humana Press.

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

Frank J, Zhu J, Penczek P, Li Y, Srivastava S, Verschoor A, Radermacher M, Grassucci R, Lata RK, Agrawal RK. 1995. A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome. *Nature 376*:441-444.

Fraser CS, Doudna JA. 2007. Structural and mechanistic insights into hepatitis C viral translation initiation. *Nat Rev Microbiol 5*:29-38.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res 19*:92-105.

Fukushi S, Okada M, Stahl J, Kageyama T, Hoshino FB, Katayama K. 2001. Ribosomal protein S5 interacts with the internal ribosomal entry site of hepatitis C virus. *J Biol Chem 276*:20824-20826.

Furtig B, Richter C, Wohnert J, Schwalbe H. 2003. NMR spectroscopy of RNA. *Chembiochem 4*:936-962.

G. Dirheimer GK, P. Dumas, E. Westhof. . 1995. In: *tRNA: structure, biosynthesis, and function.* (eds. Soll D, RajBhandary U.), pp. 93-126. Washington, D.C.: ASM Press.

Gagnon MG, Mukhopadhyay A, Steinberg SV. 2006. Close packing of helices 3 and 12 of 16 S 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.

Gast FU. 2003. A new structural motif in the left terminal domain of large viroids identified by covariation analysis. *Virus Genes 26*:19-23.

Gesteland RF, Atkins JF. 1996. Recoding: dynamic reprogramming of translation. *Annu Rev Biochem 65*:741-768.

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

Golden BL, Gooding AR, Podell ER, Cech TR. 1998. A preorganized active site in the crystal structure of the Tetrahymena ribozyme. *Science 282*:259-264.

Golden DE, Gerbasi VR, Sontheimer EJ. 2008. An inside job for siRNAs. *Mol Cell 31*:309-312.

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

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res 31*:439-441.

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.

Gualerzi CO, Pon CL. 1990. Initiation of mRNA translation in prokaryotes. *Biochemistry 29*:5881-5889.

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

Gutell RR. 1995. Comparative sequence analysis and the structure of 16S and 23S rRNA. *In Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein*

Biosynthesis (eds R A Zimmerman & A E Dahlberg), pp 109–126 Boca Raton, Florida: CRC Press.

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.

Gutell RR, Lee JC, Cannone JJ. 2002. The accuracy of ribosomal RNA comparative structure models. *Curr Opin Struct Biol 12*:301-310.

Gutell RR, Power A, Hertz GZ, Putz EJ, Stormo GD. 1992. Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Res 20*:5785-5795.

Ha SN, Giammona A, Field M, Brady JW. 1988. A revised potential-energy surface for molecular mechanics studies of carbohydrates. *Carbohydr Res 180*:207-221.

Hirokawa G, Nijman RM, Raj VS, Kaji H, Igarashi K, Kaji A. 2005. The role of ribosome recycling factor in dissociation of 70S ribosomes into subunits. *RNA 11*:1317-1328.

Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res 31*:3429-3431.

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

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 U S A 84*:4762-4766.

Huizenga DE, Szostak JW. 1995. A DNA aptamer that binds adenosine and ATP. *Biochemistry 34*:656-665.

Jacks T, Madhani HD, Masiarz FR, Varmus HE. 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell 55*:447-458.

Jang SK. 2006. Internal initiation: IRES elements of picornaviruses and hepatitis c virus. *Virus Res 119*:2-15.

Jeffrey GA, Saenger W. 1991. *Hydrogen bonding in biological structures*. Berlin ; New York: Springer-Verlag.

Jenner LB, Demeshkina N, Yusupova G, Yusupov M. 2010. Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat Struct Mol Biol*.

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 U S A 105*:16924-16927.

Karplus M, McCammon JA. 2002. Molecular dynamics simulations of biomolecules. *Nat Struct Biol 9*:646-652.

Katunin VI, Muth GW, Strobel SA, Wintermeyer W, Rodnina MV. 2002. Important contribution to catalysis of peptide bond formation by a single ionizing group within the ribosome. *Mol Cell 10*:339-346.

Kazantsev AV, Krivenko AA, Harrington DJ, Holbrook SR, Adams PD, Pace NR. 2005. Crystal structure of a bacterial ribonuclease P RNA. *Proc Natl Acad Sci U S A 102*:13392-13397.

Ke A, Zhou K, Ding F, Cate JH, Doudna JA. 2004. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature 429*:201-205.

Keefe AD, Pai S, Ellington A. 2011. Aptamers as therapeutics. *Nat Rev Drug Discov 9*:537-550.

Kelleher C, Teixeira MT, Forstemann K, Lingner J. 2002. Telomerase: biochemical considerations for enzyme and substrate. *Trends Biochem Sci 27*:572-579.

Kieft JS, Zhou K, Jubin R, Doudna JA. 2001. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA 7*:194-206.

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, Moore PB, Steitz TA. 2004a. The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA 10*:1366-1379.

Klein DJ, Moore PB, Steitz TA. 2004b. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J Mol Biol 340*:141-177.

Korostelev A, Asahara H, Lancaster L, Laurberg M, Hirschi A, Zhu J, Trakhanov S, Scott WG, Noller HF. 2008. Crystal structure of a translation termination complex formed with release factor RF2. *Proc Natl Acad Sci U S A 105*:19684-19689.

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.

Laletina E, Graifer D, Malygin A, Ivanov A, Shatsky I, Karpova G. 2006. Proteins surrounding hairpin IIIe of the hepatitis C virus internal ribosome entry site on the human 40S ribosomal subunit. *Nucleic Acids Res 34*:2027-2036.

Laurberg M, Asahara H, Korostelev A, Zhu J, Trakhanov S, Noller HF. 2008. Structural basis for translation termination on the 70S ribosome. *Nature 454*:852-857.

Laursen BS, Sorensen HP, Mortensen KK, Sperling-Petersen HU. 2005. Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev 69*:101-123.

Lee JC, Cannone JJ, Gutell RR. 2003. The lonepair triloop: a new motif in RNA structure. *J Mol Biol 325*:65-83.

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.

Lee JC, Gutell RR, Russell R. 2006. The UAA/GAN internal loop motif: a new RNA structural element that forms a cross-strand AAA stack and long-range tertiary interactions. *J Mol Biol 360*:978-988.

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.

Lehmann K, Schmidt U. 2003. Group II introns: structure and catalytic versatility of large natural ribozymes. *Crit Rev Biochem Mol Biol 38*:249-303.

Leontis NB, Stombaugh J, Westhof E. 2002. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res 30*:3497-3531.

Leontis NB, Westhof E. 1998. A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J Mol Biol 283*:571-583.

Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA 7*:499-512.

Leontis NB, Westhof E. 2002. The annotation of RNA motifs. *Comp Funct Genomics 3*:518-524.

Leontis NB, Westhof E. 2003. Analysis of RNA motifs. *Curr Opin Struct Biol 13*:300-308.

Lescoute A, Leontis NB, Massire C, Westhof E. 2005. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res 33*:2395-2409.

Lilley DM. 2003. The origins of RNA catalysis in ribozymes. *Trends Biochem Sci 28*:495-501.

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature 433*:769-773.

Lippman Z, Martienssen R. 2004. The role of RNA interference in heterochromatic silencing. *Nature 431*:364-370.

Loria A, Pan T. 1997. Recognition of the T stem-loop of a pre-tRNA substrate by the ribozyme from Bacillus subtilis ribonuclease P. *Biochemistry 36*:6317-6325.

MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B, 102*:3586–3616.

MacKerell AD, Wiorkiewicz-Kuczera J, Karplus M. 1995. An all-atom empirical energy function for the simulation of nucleic acids. *J Am Chem Soc 117*:11946–11975.

Marciniak RA, Johnson FB, Guarente L. 2000. Dyskeratosis congenita, telomeres and human ageing. *Trends Genet 16*:193-195.

Marintchev A, Wagner G. 2004. Translation initiation: structures, mechanisms and evolution. *Q Rev Biophys 37*:197-284.

Mathews DH, Andre, T.C., Kim, J., Turner, D.H., & Zuker, M. 1998. An updated recursive algorithm for RNA secondary structure prediction with improved free energy parameters. *American Chemical Society Symposium Series 682*:246–257.

Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD. 2005. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell 123*:607-620.

McCammon JA, Harvey SC. 1987. *Dynamics of proteins and nucleic acids*. Cambridge Cambridgeshire ; New York: Cambridge University Press.

McEachern MJ, Krauskopf A, Blackburn EH. 2000. Telomeres and their control. *Annu Rev Genet 34*:331-358.

Meister G, Tuschl T. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature 431*:343-349.

Mello CC, Conte D, Jr. 2004. Revealing the world of RNA interference. *Nature 431*:338-342.

Mercey R, Lantier I, Maurel MC, Grosclaude J, Lantier F, Marc D. 2006. Fast, reversible interaction of prion protein with RNA aptamers containing specific sequence patterns. *Arch Virol 151*:2197-2214.

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

Moazed D, Noller HF. 1990. Binding of tRNA to the ribosomal A and P sites protects two distinct sets of nucleotides in 16 S rRNA. *J Mol Biol 211*:135-145.

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.

Mokrejs M, Vopalensky V, Kolenaty O, Masek T, Feketova Z, Sekyrova P, Skaloudova B, Kriz V, Pospisek M. 2006. IRESite: the database of experimentally verified IRES structures (www.iresite.org). *Nucleic Acids Res 34*:D125-130.

Muth GW, Ortoleva-Donnelly L, Strobel SA. 2000. A single adenosine with a neutral pKa in the ribosomal peptidyl transferase center. *Science 289*:947-950.

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

Nierhaus KH, Schulze H. and Cooperman B.S. 1980. Molecular mechanisms of the ribosomal peptidyltransferase center. *Biochem Int 1*:185–192.

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science 289*:920-930.

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 U S A 98*:4899-4903.

Nixon PL, Rangan A, Kim YG, Rich A, Hoffman DW, Hennig M, Giedroc DP. 2002. Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J Mol Biol 322*:621-633.

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

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.

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, 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.

Otto GA, Puglisi JD. 2004. The pathway of HCV IRES-mediated translation initiation. *Cell 119*:369-380.

Pace NR, Thomas, B.C. and Woese, C.R. 1999. Probing RNA structure, function and history by comparative analysis. In *The RNA World, 2nd Edn*. (eds. R.F. Gesteland, T.R. Cech, and J.F. Atkins), pp. 113-141. Cold Spring Harbor Laboratory Press.

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.

Pecot CV, Calin GA, Coleman RL, Lopez-Berestein G, Sood AK. 2011. RNA interference in the clinic: challenges and future directions. *Nat Rev Cancer 11*:59-67.

Peske F, Rodnina MV, Wintermeyer W. 2005. Sequence of steps in ribosome recycling as defined by kinetic analysis. *Mol Cell 18*:403-412.

Petrelli D, LaTeana A, Garofalo C, Spurio R, Pon CL, Gualerzi CO. 2001. Translation initiation factor IF3: two domains, five functions, one mechanism? *EMBO J 20*:4560-4569.

Pillai RS, Bhattacharyya SN, Filipowicz W. 2007. Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol 17*:118-126.

Pinck M, Yot P, Chapeville F, Duranton HM. 1970. Enzymatic binding of valine to the 3' end of TYMV-RNA. *Nature 226*:954-956.

Plant EP, Jacobs KL, Harger JW, Meskauskas A, Jacobs JL, Baxter JL, Petrov AN, Dinman JD. 2003. The 9-A solution: how mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. *RNA 9*:168-174.

Pleij CW, Rietveld K, Bosch L. 1985. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res 13*:1717-1731.

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.

Polacek N, Gaynor M, Yassin A, Mankin AS. 2001. Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide. *Nature 411*:498-501.

Powers T, Noller HF. 1991. A functional pseudoknot in 16S ribosomal RNA. *EMBO J 10*:2203-2214.

Puglisi JD, Blanchard SC, Green R. 2000. Approaching translation at atomic resolution. *Nat Struct Biol 7*:855-861.

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

Rastogi T, Beattie TL, Olive JE, Collins RA. 1996. A long-range pseudoknot is required for activity of the Neurospora VS ribozyme. *EMBO J 15*:2820-2825.

Rietveld K, Van Poelgeest R, Pleij CW, Van Boom JH, Bosch L. 1982. The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. *Nucleic Acids Res 10*:1929-1946.

Rivas E, Eddy SR. 1999. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J Mol Biol 285*:2053-2068.

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.

Robinson KA, Beverley SM. 2003. Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite Leishmania. *Mol Biochem Parasitol 128*:217-228.

Rodnina MV, Savelsbergh A, Katunin VI, Wintermeyer W. 1997. Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature 385*:37-41.

Rosenblad MA, Gorodkin J, Knudsen B, Zwieb C, Samuelsson T. 2003. SRPDB: Signal Recognition Particle Database. *Nucleic Acids Res 31*:363-364.

Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. 2008. FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J Math Biol 56*:215-252.

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 angstroms resolution. *Cell 102*:615-623.

Schreiber SL. 2000. Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science 287*:1964-1969.

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

Selmer M, Dunham CM, Murphy FVt, 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.

Shen LX, Tinoco I, Jr. 1995. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J Mol Biol 247*:963-978.

Shevack A, Gewitz HS, Hennemann B, Yonath A, HG. aW. 1985. Characterization and crystallization of ribosomal particles from Halobacterium marismortui. *FEBS Lett 184*:68-71.

Shi H, Moore PB. 2000. The crystal structure of yeast phenylalanine tRNA at 1.93 A 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 U S A 71*:1342-1346.

Shine J, Dalgarno L. 1975. Determinant of cistron specificity in bacterial ribosomes. *Nature 254*:34-38.

Sievers A, Beringer M, Rodnina MV, Wolfenden R. 2004. The ribosome as an entropy trap. *Proc Natl Acad Sci U S A 101*:7897-7901.

Steitz TA, Steitz JA. 1993. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A 90*:6498-6502.

ten Dam EB, Pleij CW, Bosch L. 1990. RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. *Virus Genes 4*:121-136.

Thompson J, Kim DF, O'Connor M, Lieberman KR, Bayfield MA, Gregory ST, Green R, Noller HF, Dahlberg AE. 2001. Analysis of mutations at residues A2451 and G2447 of 23S rRNA in the peptidyltransferase active site of the 50S ribosomal subunit. *Proc Natl Acad Sci U S A 98*:9002-9007.

Tinoco J. 1993. In appendix 1 of: The RNA world : the nature of modern RNA suggests a prebiotic RNA world. (eds Gesteland RF, Atkins JF.), Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Tocilj A, Schlunzen F, Janell D, Gluhmann M, Hansen HA, Harms J, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A. 1999. The small ribosomal subunit from Thermus thermophilus at 4.5 A resolution: pattern fittings and the identification of a functional site. *Proc Natl Acad Sci U S A 96*:14252-14257.

Tomari Y, Zamore PD. 2005. Perspective: machines for RNAi. *Genes Dev 19*:517-529.

Tomsic J, Vitali LA, Daviter T, Savelsbergh A, Spurio R, Striebeck P, Wintermeyer W, Rodnina MV, Gualerzi CO. 2000. Late events of translation initiation in bacteria: a kinetic analysis. *EMBO J 19*:2127-2136.

Torres-Larios A, Swinger KK, Krasilnikov AS, Pan T, Mondragon A. 2005. Crystal structure of the RNA component of bacterial ribonuclease P. *Nature 437*:584-587.

Trobro S, Aqvist J. 2005. Mechanism of peptide bond synthesis on the ribosome. *Proc Natl Acad Sci U S A 102*:12395-12400.

Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science 249*:505-510.

Tullius TD, Greenbaum JA. 2005. Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr Opin Chem Biol 9*:127-134.

Ulrich H, Trujillo CA, Nery AA, Alves JM, Majumder P, Resende RR, Martins AH. 2006. DNA and RNA aptamers: from tools for basic research towards therapeutic applications. *Comb Chem High Throughput Screen 9*:619-632.

Valle M, Zavialov A, Sengupta J, Rawat U, Ehrenberg M, Frank J. 2003. Locking and unlocking of ribosomal motions. *Cell 114*:123-134.

Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, Mallory AC, Hilbert JL, Bartel DP, Crete P. 2004. Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol Cell 16*:69-79.

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 U S A 106*:1063-1068.

von Bohlen K, Makowski I, Hansen HA, Bartels H, Berkovitch-Yellin Z, Zaytzev-Bashan A, Meyer S, Paulke C, Franceschi F, Yonath A. 1991. Characterization and preliminary attempts for derivatization of crystals of large ribosomal subunits from Haloarcula marismortui diffracting to 3 A resolution. *J Mol Biol 222*:11-15.

Vulliamy T, Marrone A, Dokal I, Mason PJ. 2002. Association between aplastic anaemia and mutations in telomerase RNA. *Lancet 359*:2168-2170.

Wang C, Le SY, Ali N, Siddiqui A. 1995. An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5' noncoding region. *RNA 1*:526-537.

Wang W, Donini O, Reyes CM, Kollman PA. 2001. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu Rev Biophys Biomol Struct 30*:211-243.

Wang X, Han Y, Lin L, Fuji M, Endo T, Watanabe H, Takahashi M. 2008. A molecular dynamics study on aqueous solutions for preparation of hollow CaCO3 particles. *Modelling Simul Mater Sci Eng 16*.

Weaver RF. 2005. *Molecular biology*. Boston, Mass.: McGraw-Hill.

Weiss RB, Dunn DM, Shuh M, Atkins JF, Gesteland RF. 1989. E. coli ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biol 1*:159-169.

Weixlbaumer A, Jin H, Neubauer C, Voorhees RM, Petry S, Kelley AC, Ramakrishnan V. 2008. Insights into translational termination from the structure of RF2 bound to the ribosome. *Science 322*:953-956.

Wells SE, Hillner PE, Vale RD, Sachs AB. 1998. Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell 2*:135-140.

Wilson J, Hunt T. 2002. *Molecular biology of the cell : a problems approach*. New York: Garland Science.

Wilson WD. 2002. Tech.Sight. Analyzing biomolecular interactions. *Science 295*:2103-2105.

Wimberly B, Varani G, Tinoco I, Jr. 1993. The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry 32*:1078-1087.

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, 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.

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

Yonath A, Mussig J, Tesche B, Lorenz S, Erdmann VA, Wittmann HG. 1980. Crystallization of the large ribosomal subunits from Bacillus stearothermophilus. *Biochem Internat 1*:428-435.

Youngman EM, Brunelle JL, Kochaniak AB, Green R. 2004. The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. *Cell 117*:589-599.

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

Zagryadskaya EI, Doyon FR, Steinberg SV. 2003. Importance of the reverse Hoogsteen base pair 54-58 for tRNA function. *Nucleic Acids Res 31*:3946-3953.

Zavialov AV, Buckingham RH, Ehrenberg M. 2001. A posttermination ribosomal complex is the guanine nucleotide exchange factor for peptide release factor RF3. *Cell 107*:115-124.

Zavialov AV, Hauryliuk VV, Ehrenberg M. 2005. Splitting of the posttermination ribosome into subunits by the concerted action of RRF and EF-G. *Mol Cell 18*:675-686.

Zimmermann GR, Wick CL, Shields TP, Jenison RD, Pardi A. 2000. Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *RNA 6*:659-667.

Zuker M. 1989. Computer prediction of RNA structure. *Methods Enzymol 180*:262-288.

Zwieb C, Gorodkin J, Knudsen B, Burks J, Wower J. 2003. tmRDB (tmRNA database). *Nucleic Acids Res 31*:446-447.