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

Matthieu G. Gagnon, Yury I. Boutorine and Sergey V. Steinberg*

Département de Biochimie, Université de Montréal, Montréal, C.P. 6128, Succursale Centre-Ville, QC H3C 3J7, Canada

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- 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.

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 (1). 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 article, 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 (2–4). 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

*To whom correspondence should be addressed. Tel: +1 514 343 6320; Fax: +1 514 343 2210; Email: serguei.chteinberg@umontreal.ca
Present address:
Matthieu G. Gagnon, Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, Bass Center Room 413, P.O. Box 208114, New Haven, CT 06520-8114, USA.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

found in more than a dozen places of the ribosome structure (5,6).

MATERIALS AND METHODS

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 (*ΔrrnE ΔrrnB ΔrrnA ΔrrnH ΔrrnG::lacZ ΔrrnC::cat ΔrrnD::cat ΔrecA*/ptRNA67-Spc^R) carrying the rRNA-coding plasmid pHK*rrnC-sacB*-Kan^R (7,8). 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 (9). Cultures were grown in the Luria–Bertani (LB) medium (10) or in the LB medium supplemented with appropriate antibiotics, 100 μg/ml ampicillin (Amp), 50 μg/ml kanamycin (Kan) and 40 μg/ml spectinomycin (Spc) (Sigma-Aldrich Canada).

Plasmids

The combinatorial 16S rRNA gene library of motif S296 was obtained previously using the specialized ribosome system cloned in plasmid pAMMG (11). For expression of wild-type and mutant 23S rRNA, we used plasmids pKK1192U-Amp^R (12) and pLΔH1192U-Amp^R (9). 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 30°C, this promoter is repressed due to the presence of the temperature sensitive cI857 repressor encoded by the host chromosome.

Design of the combinatorial gene libraries

The 4 nts comprising the two central base pairs of motifs S296, L639 and L657 were fully randomized using the overlapping extension PCR procedure (13). In this way, the entire regions comprising motifs S296 (902 bp), L639 and L657 (1541 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' section and Supplementary Table S3. Transformation of the plasmids harboring the combinatorial 23S rRNA gene libraries into the SQ380 cells was performed by electroporation.

Plasmid replacement and selection of functional clones

The exchange of the resident wild-type pHK*rrnC-sacB*-Kan^R plasmid with the pKK1192U or pLΔH1192U plasmid carrying mutant 23S rRNA was performed as previously described with some modifications (7,8). First, the cell culture was grown for 1 h at 37°C without antibiotic. Then, to facilitate the plasmid replacement, the growth continued for three 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 h at 30°C for efficient expression of the *sacB* gene conferring sucrose sensitivity (14,15). A total of ~1×10^5 transformants were obtained for both motifs L639 and L657, out of which several hundred grew after selection. For each library, 50 selected clones were checked on LB-Kan-agar plates to confirm the loss of the resident pHK*rrnC-sacB*-Kan^R plasmid followed by the sequencing of the 23S rRNA gene in the pKK1192U or pLΔH1192U plasmid.

Measurement of the ribosome efficiency and of the growth rates

The GFP activity of each A-clone was measured previously (11). 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.

Sequencing

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

Molecular dynamics simulations

Molecular dynamics (MD) simulations were carried out on four different constructs, each composed of two double helices forming together the AGPM (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) (16) 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.

RESULTS

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) (6). Due to the frequent occurrence, AGPM constitutes an important element of the ribosome structure. Its major
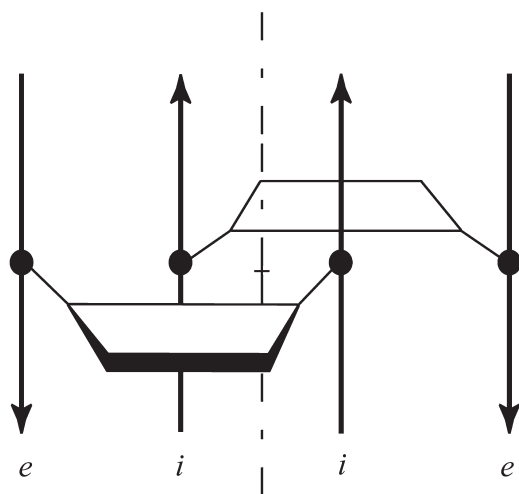

Figure 1. Schematic representation of AGPM. Trapezoids stand for base pairs opened toward the minor grooves. Arrows represent backbones directed 5′→3′. The internal and external strands of both helices are marked by italic letters *i* and *e*, respectively. The internal strand of each helix is packed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

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 the 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) (6). 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 (6) (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 ∼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 (5,11). 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* (17), 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 (18). 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 4 nts 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.
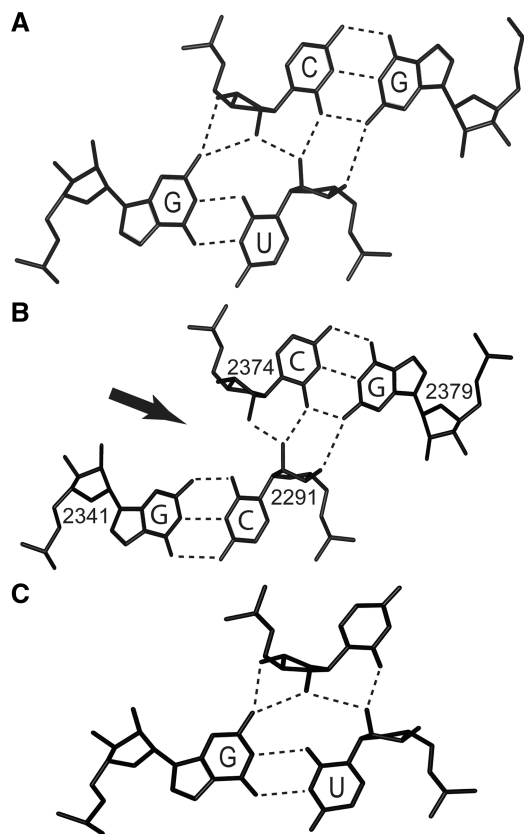
Figure 2. Different arrangements of the central base pairs in AGPM. (**A**) The canonical GU–WC arrangement. The presence of the GU base pair allows the close packing between the helices with the formation of the inter-helical network of five hydrogen bonds. (**B**) The GC–GC juxtaposition taken from motif L2291 (*E. coli* numbering) in the *H. marismortui* 23S rRNA (pdb entry code 1s72.pdb) (17). 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. 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) (16). 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 is WC (6). 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 (6). (**B**) Nucleotide sequence of motif L657 (L657-Hm) from the *H. marismortui* ribosome (pdb entry code 1s72) (17). Compared with the same motif in *E. coli* (A), the GU and WC base pairs in *H. marismortui* are exchanged in their positions.

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 (11).

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
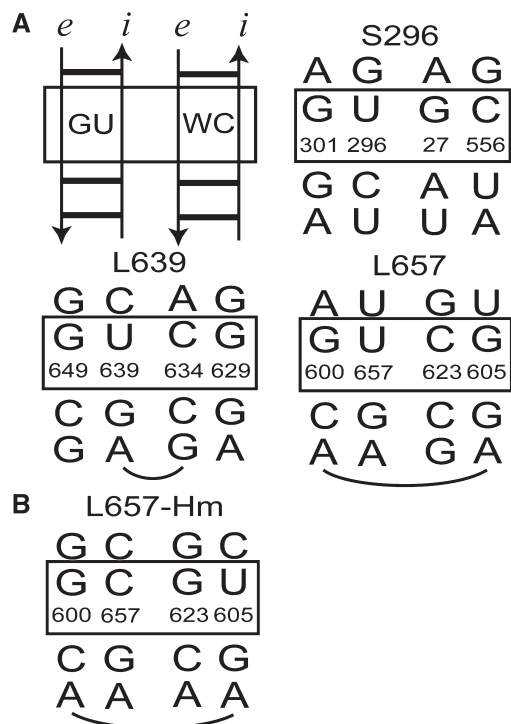
participate in interactions with ribosomal proteins. In L657, nt 600 of helix H27, which occupies the external position of a central base pair, forms a tight contact with residues L27, K99 and M100 of protein L4 (Figure 4a). All three residues interact only with the ribose of nt 600, and not with the 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 (19). In motif L639, nucleotides of the central base pairs are not directly involved in interactions with other parts of the 50S structure. However, nts 650 and 651, which are proximate to the external nt 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 nts 650–651 that form contacts with L35.

In this article, we demonstrate how the above-mentioned differences in the structural contexts of the three chosen motifs affect the variability of the central base pairs.
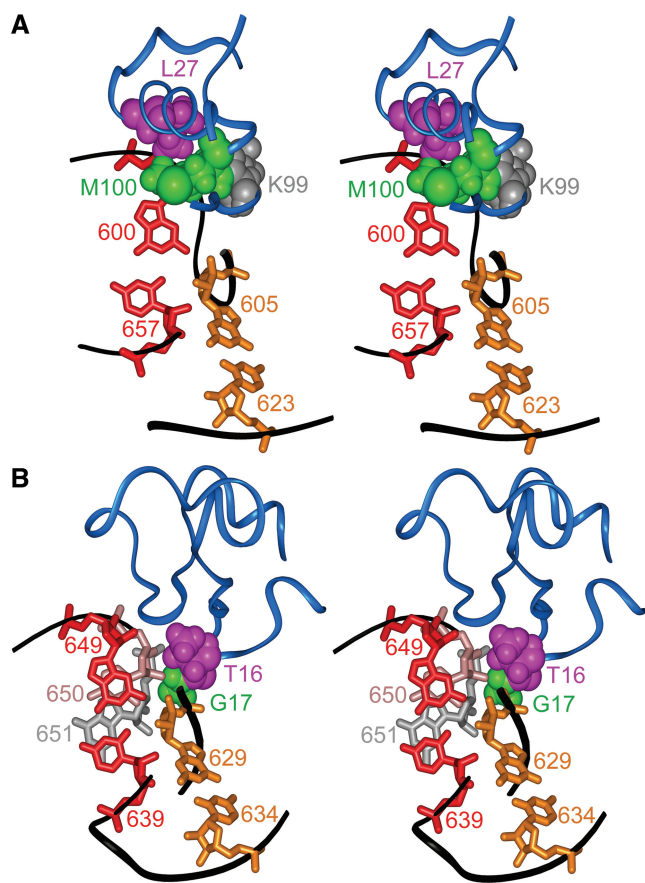
Figure 4. The structural contexts of motifs L657 (**A**) and L639 (**B**) taken from the *E. coli* ribosome. (A) In motif L657, residues L27 (magenta), K99 (grey) and M100 (green) of the ribosomal protein L4 (blue ribbon) tightly interact with the ribose and with the backbone of nt 600, which occupies 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 nts 650 (pink) and 651 (grey) proximal to the external nt-649 of a central base pair (red).

Cloning and selection of functional clones

As mentioned above, in each of the three AGPMs all four nt 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 (11,20–24). 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 (11). For selection of functional variants of motifs L639 and L657 located in the 23S rRNA, we used the ribosome knock-out strain SQ380 (7,8). 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' section). 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.

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 with the wild-type *E. coli* ribosome, the GU and WC base pairs have exchanged between the helices, which, however, does not affect the packing (5,11). 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,

Table 1. Nucleotide sequences of the selected clones

Clone	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	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

Clone	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

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.

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 2 nts 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 with 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.

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 with 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 ~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.

MD simulations

To test the ability of the nucleotide triple to stabilize the structure of AGPM, we performed 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' section. In all constructs, the identities of all nucleotides were the same except those 4 nts 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 ~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 with 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
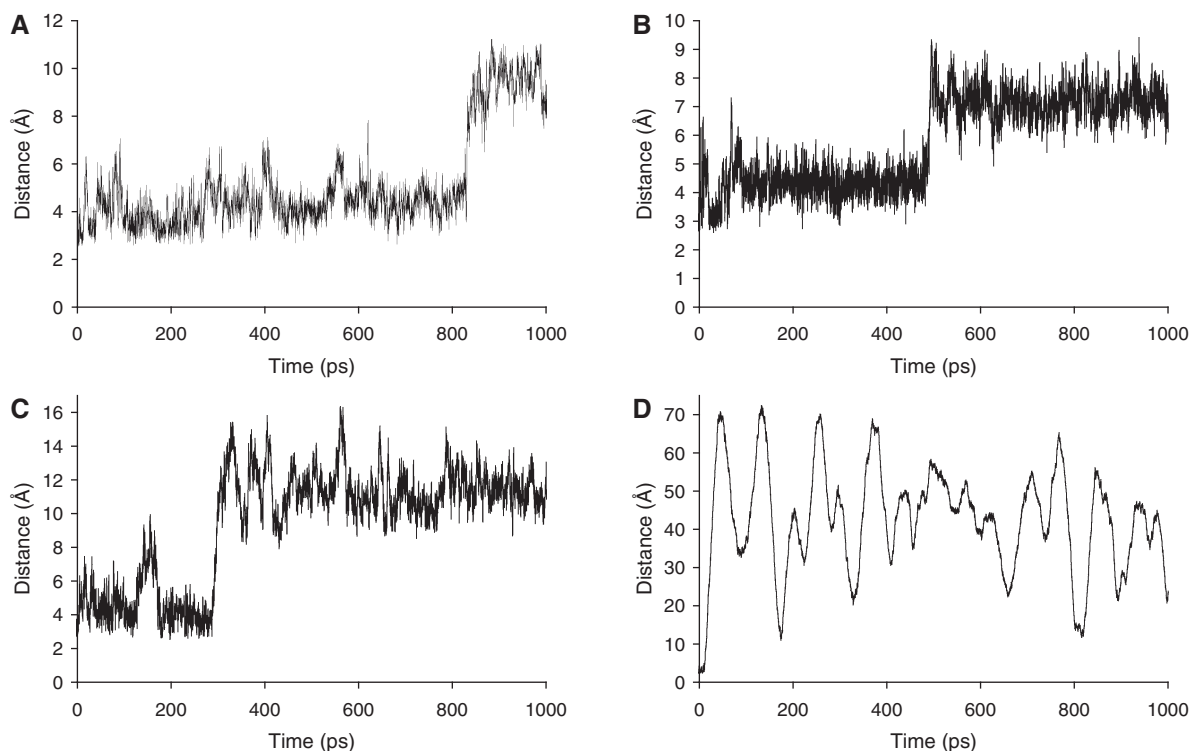
Figure 5. Molecular dynamics simulations of the AGPM structure containing different nucleotide triples. In the four complexes tested, the dinucleotide combinations occupying the central base pairs were GU–CG (**A**), GU–UG (**B**), GC–UG (**C**) and UG–UG (**D**). The CG, GC and GU combinations were 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.

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 ∼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.

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.

Interaction of motif L639 with ribosomal protein L35

Compared with 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, nts 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 nt 649 of the central base pair, the fact that the neighboring nts 650 and 651 form a tight contact with L35 would limit the mobility of nt 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 (25). 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.

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,
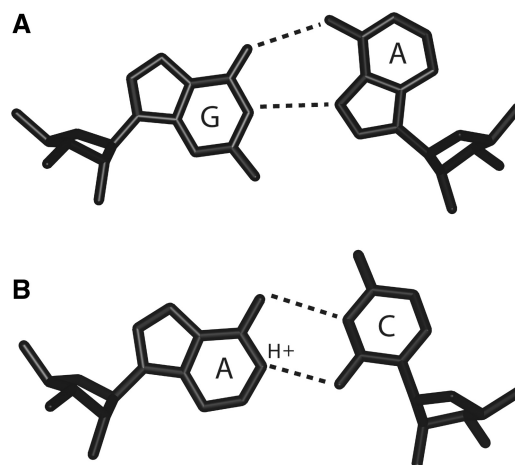

Figure 6. Juxtapositions of the bases in the GA and AC base pairs. (**A**) The proposed WC-like arrangement for the GA base pair in clones B3 and B6. The formation of this base pair requires that the adenosine have the unusual *syn*-conformation. Such arrangement is observed in base pairs G_9–A_{16} and G_{21}–A_4 of the crystal structure of a DNA oligonucleotide duplex (pdb entry code 1dnm) (25). (**B**) The GU-like juxtaposition observed in base pairs A_{105}–C_{112} and A_{113}–C_{104} (pdb entry code 402d) (27), A_{11}–C_{22} and A_{27}–C_6 (pdb entry code 405d) (28), A_5–C_{24} and A_{23}–C_6 (pdb entry code 1d4r) (29), A_{139}–C_{158} and A_{157}–C_{140} (pdb entry code 1jid) (30), A_{26}–C_{44} (pdb entry codes 1o0b, 1qtq, 1gtr, 1qru) (31–34), A_{1500}–C_{1402} (pdb entry code 1j5e) (35), A_{192}–C_{178} (pdb entry code 1u9s) (36) and A_{44}–C_{26} (pdb entry code 1y27) (37). 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) (38,39) the tertiary base pair between A_{46} and G_{22} presumes the presence of an extra proton at the inter-base contact (not shown).

unlike in L639, nt 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 four 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* (16) and *H. marismortui* (17) ribosomes. Compared with *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 ~15°, leading to the displacement of its atoms by at least 1 Å. However, within the AGPM, nts 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 nts 623 and 605 and may even result in the bulging of nt 623. In the absence of the strong interaction between nts 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.

Exceptional clones A11, B16 and C84

Among all 47 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
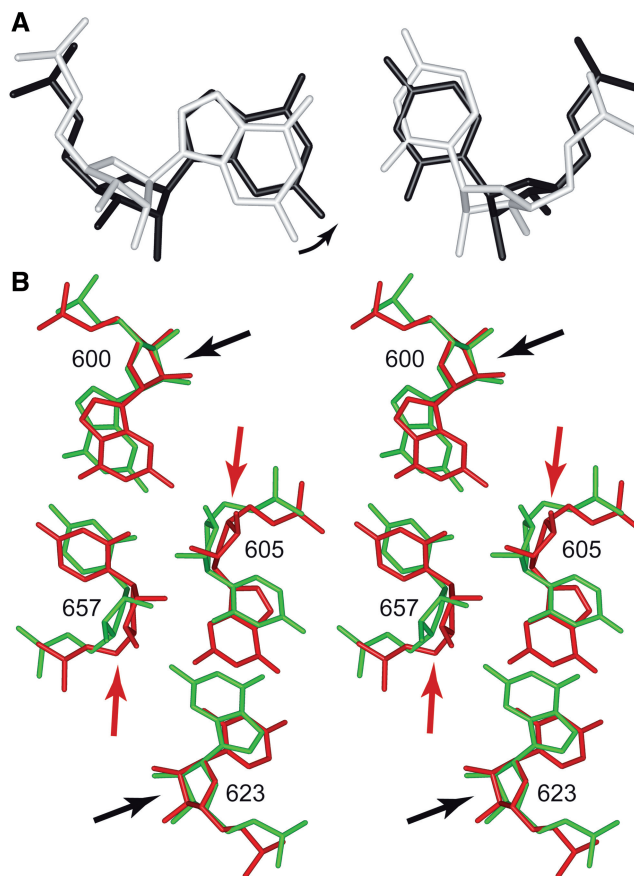
Figure 7. Conformational rearrangements associated with the GU⇔WC exchange of the central base pairs in AGPM. (**A**) The superposition of base pairs G_{71}–C_2 (black) and G_4–U_{69} (white) in the structure of the yeast tRNA^{Phe} (pdb entry code 1ehz) (40). The superposition was obtained by superposing 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 ~15° toward the major groove (black arrow). However small this displacement is, it can be large enough to damage the interaction between the external nt 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) (16) and *H. marismortui* (green, pdb entry code 1s72) (17) 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.

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 2 nts (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.

DISCUSSION

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

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 with 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.

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 (26), the same principle can be essential for RNA–protein interactions at large.

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 ~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 ~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- and long-range interactions in the ribosome and, potentially, in other RNA–protein complexes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Steitz,T.A. (2008) A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.*, **9**, 242–253.
2. Noller,H.F. (2005) RNA structure: reading the ribosome. *Science*, **309**, 1508–1514.
3. Moore,P.B. (1999) Structural motifs in RNA. *Annu. Rev. Biochem.*, **68**, 287–300.
4. 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.
5. Mokdad,A., Krasovska,M.V., Sponer,J. and Leontis,N.B. (2006) Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.*, **34**, 1326–1341.
6. Gagnon,M.G. and Steinberg,S.V. (2002) GU receptors of double helices mediate tRNA movement in the ribosome. *RNA*, **8**, 873–877.
7. Asai,T., Zaporojets,D., Squires,C. and Squires,C.L. (1999) 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.
8. Asai,T., Condon,C., Voulgaris,J., Zaporojets,D., Shen,B., Al-Omar,M., Squires,C. and Squires,C.L. (1999) Construction and initial characterization of *Escherichia coli* strains with few or no intact chromosomal rRNA operons. *J. Bacteriol.*, **181**, 3803–3809.
9. Pinard,R., Payant,C., Melancon,P. and 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.
10. Luria,S.E. and Burrous,J.W. (1957) Hybridization between *Escherichia coli* and *Shigella*. *J. Bacteriol.*, **74**, 461–476.
11. Gagnon,M.G., Mukhopadhyay,A. and Steinberg,S.V. (2006) Close packing of helices 3 and 12 of 16S rRNA is required for the normal ribosome function. *J. Biol. Chem.*, **281**, 39349–39357.
12. Brosius,J., Ullrich,A., Raker,M.A., Gray,A., Dull,T.J., Gutell,R.R. and Noller,H.F. (1981) Construction and fine mapping of recombinant plasmids containing the *rrn*B ribosomal RNA operon of *E. coli. Plasmid*, **6**, 112–118.
13. Ho,S.N., Hunt,H.D., Horton,R.M., Pullen,J.K. and Pease,L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51–59.
14. Blomfield,I.C., Vaughn,V., Rest,R.F. and Eisenstein,B.I. (1991) Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. *Mol. Microbiol.*, **5**, 1447–1457.
15. Gay,P., Le Coq,D., Steinmetz,M., Berkelman,T. and Kado,C.I. (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J. Bacteriol.*, **164**, 918–921.
16. 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.
17. 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.
18. Wuyts,J., Perriere,G. and Van De Peer,Y. (2004) The European ribosomal RNA database. *Nucleic Acids Res.*, **32**, D101–D103.
19. Li,X., Lindahl,L. and Zengel,J.M. (1996) Ribosomal protein L4 from *Escherichia coli* utilizes nonidentical determinants for its structural and regulatory functions. *RNA*, **2**, 24–37.
20. Lee,K., Holland-Staley,C.A. and Cunningham,P.R. (1996) Genetic analysis of the Shine-Dalgarno interaction: selection of alternative functional mRNA-rRNA combinations. *RNA*, **2**, 1270–1285.
21. Hui,A. and de Boer,H.A. (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.
22. Rackham,O. and Chin,J.W. (2005) A network of orthogonal ribosome x mRNA pairs. *Nat. Chem. Biol.*, **1**, 159–166.
23. Abdi,N.M. and Fredrick,K. (2005) Contribution of 16S rRNA nucleotides forming the 30S subunit A and P sites to translation in *Escherichia coli. RNA*, **11**, 1624–1632.
24. Belanger,F., Gagnon,M.G., Steinberg,S.V., Cunningham,P.R. and 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.
25. Webster,G.D., Sanderson,M.R., Skelly,J.V., Neidle,S., Swann,P.F., Li,B.F. and Tickle,I.J. (1990) Crystal structure and sequence-dependent conformation of the A.G mispaired oligonucleotide d(CGCAAGCTGGCG). *Proc. Natl Acad. Sci. USA*, **87**, 6693–6697.
26. McClain,W.H., Schneider,J., Bhattacharya,S. and Gabriel,K. (1998) The importance of tRNA backbone-mediated interactions with synthetase for aminoacylation. *Proc. Natl Acad. Sci. USA*, **95**, 460–465.
27. Jang,S.B., Hung,L.W., Chi,Y.I., Holbrook,E.L., Carter,R.J. and Holbrook,S.R. (1998) Structure of an RNA internal loop consisting of tandem C-A+ base pairs. *Biochemistry*, **37**, 11726–11731.
28. Pan,B., Mitra,S.N. and Sundaralingam,M. (1998) Structure of a 16-mer RNA duplex r(GCAGACUUAAAUCUGC)2 with wobble C.A+ mismatches. *J. Mol. Biol.*, **283**, 977–984.
29. Wild,K., Weichenrieder,O., Leonard,G.A. and Cusack,S. (1999) The 2 Å structure of helix 6 of the human signal recognition particle RNA. *Structure*, **7**, 1345–1352.
30. Wild,K., Sinning,I. and Cusack,S. (2001) Crystal structure of an early protein-RNA assembly complex of the signal recognition particle. *Science*, **294**, 598–601.
31. Arnez,J.G. and Steitz,T.A. (1996) Crystal structures of three misacylating mutants of Escherichia coli glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP. *Biochemistry*, **35**, 14725–14733.
32. Rould,M.A., Perona,J.J. and Steitz,T.A. (1991) Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature*, **352**, 213–218.
33. Rath,V.L., Silvian,L.F., Beijer,B., Sproat,B.S. and Steitz,T.A. (1998) How glutaminyl-tRNA synthetase selects glutamine. *Structure*, **6**, 439–449.
34. Bullock,T.L., Uter,N., Nissan,T.A. and Perona,J.J. (2003) Amino acid discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants. *J. Mol. Biol.*, **328**, 395–408.
35. 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.
36. Krasilnikov,A.S., Xiao,Y., Pan,T. and Mondragon,A. (2004) Basis for structural diversity in homologous RNAs. *Science*, **306**, 104–107.
37. Serganov,A., Yuan,Y.R., Pikovskaya,O., Polonskaia,A., Malinina,L., Phan,A.T., Hobartner,C., Micura,R., Breaker,R.R. and Patel,D.J. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.*, **11**, 1729–1741.
38. Westhof,E., Dumas,P. and Moras,D. (1985) Crystallographic refinement of yeast aspartic acid transfer RNA. *J. Mol. Biol.*, **184**, 119–145.
39. Moras,D., Comarmond,M.B., Fischer,J., Weiss,R., Thierry,J.C., Ebel,J.P. and Giege,R. (1980) Crystal structure of yeast tRNA^Asp. *Nature*, **288**, 669–674.
40. Shi,H. and Moore,P.B. (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA*, **6**, 1091–1105.