Published online 20 July 2020

Nucleic Acids Research, 2020, Vol. 48, No. 15 ***8675–8685***
doi: 10.1093/nar/gkaa610

Base-intercalated and base-wedged stacking elements in 3D-structure of RNA and RNA–protein complexes

Eugene Baulin[1], Valeriy Metelev[2] and Alexey Bogdanov[2,3,*]

[1]Laboratory of Applied Mathematics, Institute of Mathematical Problems of Biology RAS - the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russia, [2]Department of Chemistry, Lomonosov Moscow State University, Moscow 119991, Russia and [3]A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 119991, Russia

Received February 06, 2020; Revised July 05, 2020; Editorial Decision July 08, 2020; Accepted July 15, 2020

ABSTRACT

Along with nucleobase pairing, base-base stacking interactions are one of the two main types of strong non-covalent interactions that define the unique secondary and tertiary structure of RNA. In this paper we studied two subfamilies of nucleobase-inserted stacking structures: (i) with any base intercalated between neighboring nucleotide residues (base-intercalated element, BIE, *i* + 1); (ii) with any base wedged into a hydrophobic cavity formed by heterocyclic bases of two nucleotides which are one nucleotide apart in sequence (base-wedged element, BWE, *i* + 2). We have exploited the growing database of natively folded RNA structures in Protein Data Bank to analyze the distribution and structural role of these motifs in RNA. We found that these structural elements initially found in yeast tRNA^{Phe} are quite widespread among the tertiary structures of various RNAs. These motifs perform diverse roles in RNA 3D structure formation and its maintenance. They contribute to the folding of RNA bulges and loops and participate in long-range interactions of single-stranded stretches within RNA macromolecules. Furthermore, both base-intercalated and base-wedged motifs participate directly or indirectly in the formation of RNA functional centers, which interact with various ligands, antibiotics and proteins.

INTRODUCTION

The unique macromolecular structure of RNA is determined by several types of contacts that include: (a) stacking interactions between the bases of nucleotide residues along with nucleobase hydrogen bonding, (b) 'magnesium bridges' between phosphate groups and (c) RNA–protein interactions (1,2). Stacking interactions are usually established between the consecutive bases. However, as determined by Holbrook *et al*. (3,4) about forty years ago, a base of a single-stranded RNA can intercalate or wedge between the two adjacent residues of another element of the same chain. The object of their study was yeast phenylalanyl tRNA (tRNA^{Phe}), the only RNA with the crystal structure established at that time. The authors described the geometry of these RNA elements in detail, emphasized their important role in folding the loop segments of the secondary structure of tRNA^{Phe} and suggested that such base-intercalated structures may occur in other RNAs. We believe that the results presented in our work fully support this hypothesis. However, despite the fact that the unusual elements in the structure of tRNA were discussed in the well-known monograph of Saenger (5), they were neglected for many years. Only recently the observation of an ability of a base of one RNA segment to intercalate between the bases of adjacent nucleotide residues of the same or another RNA was made in the studies of ribosomal function (6,7). For example, at a certain stage of the translocation process, the movement of mRNA along a small subunit of the ribosome (SSU) may be inhibited by the insertion of one or two nucleobases of rRNA of that small subunit between the neighboring nucleotides at certain positions of the mRNA chain (6). Another well-documented example is the intercalation of a 3′- terminal nucleotide residue of the E-site tRNA between the two universally conserved bases of a large ribosomal subunit (LSU) rRNA which assists the deacylated tRNA in performing a role of translational frame maintenance (7). Inter- and, more commonly, intramolecular contacts of this type were also observed within the 3D-structures of other natural and synthetic RNAs (8–17). Therefore, these base-intercalated or base-wedged RNA elements could be considered as motifs, i.e. commonly occurring 'well-defined geometric arrangements of interacting nucleotides within the 3D-structure of RNA' (15). However, the analysis of current literature available to us suggested that the occurrence of these motifs have not been thoroughly studied. Moreover, the patterns specific for these motifs have not been systematically searched for among the

*To whom correspondence should be addressed. Tel: +7 495 9393143; Fax: +7 495 9393181; Email: bogdanov@belozersky.msu.ru

currently available tertiary RNA structures and their classification has not been carried out.

In this paper, we demonstrate that base-intercalated and base-wedged motifs are quite widespread as could be deduced from the analysis of high-resolution sets of non-redundant RNA crystal structures. These motifs are postulated to play an important role in the formation of compact and functionally significant RNA tertiary structures.

MATERIALS AND METHODS

Terminology

Base-Intercalated or Base-Wedged Elements (BIE or BWE) are sets of three nucleotides N_i, N_j, N_k such that

1. N_i and N_k belong to the same RNA chain;
2. $k = i + 1$ (BIE) or $k = i + 2$ and $j \neq i + 1$(BWE);
3. There are base stacking (N_i, N_j) and (N_j, N_k).

BIE or BWE are called intramolecular if N_i and N_j belong to the same RNA chain and intermolecular otherwise.

Description of RNA secondary structure elements, i.e. stems and loops of different types, follows the generalization of the Nearest Neighbor Model (NNM, (18)) described in (19). Here, we declare and below we use the following informal definitions:

A stem is a sequence of at least two consecutive Watson–Crick or Wobble base-pairs. A loop is a set of non-paired regions confined by a stem. Each loop is assigned to one of the following types: hairpin (H), bulge (B), internal loop (I), or multiple junction (J). In addition, each loop is classified in relation to pseudoknots. A loop is called pseudoknotted (P) if it is involved in a pseudoknot, isolated (I) if it is adjacent to a pseudoknot and classical (C) otherwise. In total, we identified 12 different kinds of loops.

For formal definitions please see (19) or http://urs.lpm.org.ru/struct.py?where=3#def.

Each pair of nucleotides in a BIE or BWE is classified by its relative location in terms of RNA secondary structure (unpublished work) for three classes:

1. Same (SM) – the nucleotides belong to the same stem or loop;
2. Local (LC) – the nucleotides belong to a stem and its adjacent loop;
3. Long-range (LR) – the nucleotides belong to two distant secondary structure elements.

Data

We used the URSDB database (19) as a source of experimentally determined RNA structures from the Protein Data Bank (PDB) (20), with RNA secondary structure annotations. A search for BIEs and BWEs was conducted among 4551 PDB-entries (the whole URSDB set as of October 2019). Base-stacking interactions were annotated with the DSSR program from the X3DNA toolkit (21). The search resulted in about 55 000 BIEs and BWEs. The representative set of RNA structures (22) (version 3.96 with resolution cutoff 3.5 Å) has been used for the analysis. Full repeats (e.g. among different models of an NMR entry) and repeats among identical copies of an RNA molecule within a PDB-entry have been removed. Within the process of full repeats removal, the motifs of the better resolution were chosen.

Each BIE and BWE in the datasets was described by the following features (see also column descriptions in readme tab of Supplementary Table S1):

1. PDB-entry (e.g. 6hcj);
2. k ($i + 1$ or $i + 2$);
3. intramolecular/intermolecular;
4. N_i, N_j, N_k identifiers in DSSR format ('C.B.N' where C = chain, B = base, N = index within the RNA chain);
5. N_i, N_j, N_k bases (e.g. A–U–A);
6. N_i, N_j, N_k secondary structure elements (e.g. 'S-IC-IC' for 'stem'-'classical internal loop'-'classical internal loop');
7. N_i, N_j, N_k pairwise relative locations ij, jk, ik (e.g. 'LC-SM-LC' for 'local-same-local');
8. N_i RNA molecule type and organism source;
9. (N_i, N_j) and (N_j, N_k) base-stacking interactions parameters (21):
   a) base-stacking area (in squared angstroms) of the overlapped polygon defined by the two bases (both including and excluding exocyclic atoms);
   b) the relative direction of the corresponding bases (each base is assigned with the plus (p) and minus (m) faces so there can be four possible categories - pm, mp, pp, mm, see (21));
   c) the minimum distance between atoms of the corresponding bases;
   d) the angle between the planes of the corresponding bases.
10. Types of base-phosphate interactions that involve N_j (23) (if any).
11. Type of O4′-base stacking interaction that involves N_j (24) (if any).

The final dataset (nrRNA 3.5 dataset, see Supplementary Table S1) included only elements with minimum distances $\leq$4.0 Å and angles $\leq$30° between stacked bases and consisted of 1466 base-intercalated and base-wedged elements from 241 PDB-entries. For some analyses, we also used the dataset with the resolution cutoff 3.0 Å containing 822 base-intercalated and base-wedged elements (nrRNA 3.0 dataset).

RESULTS AND DISCUSSION

Types and some features of BIEs and BWEs

We studied two subfamilies of nucleobase-inserted structures: (i) with any base intercalated between neighboring nucleotide residues (base-intercalated element, BIE, $i + 1$); (ii) with any base wedged into a hydrophobic cavity formed by heterocyclic bases of two nucleotides which are one nucleotide apart in sequence (base-wedged element, BWE, $i + 2$). Quite frequently these structural elements are combined and may be attributed to zipper-like structures (BIE-ZIP or BWE-ZIP) within 3D-structure of RNA. Collectively they constitute a rather diverse family of motifs, formed as a result of a common phenomenon: i.e. the insertion of one

nucleobase between two others which are neighbors in any RNA polynucleotide chain.

The diversity of the major types of BIE and BWE is shown schematically in Figure 1B.

The BIE subfamily accounts for 76% of all elements found in the database. The remaining 24% are represented by BWEs. About half of all structural elements described in this work consist only of purine nucleotides, and the intercalator most often is adenine (Figure 2). Modified nucleotides in both subfamilies account for ∼2%.

As stated above, BIEs are located only in various RNA loops. However, the mechanism of BIE formation is quite comparable to the mechanism of the intercalation of aromatic dyes and side chains of amino acids into double-helical DNA and RNA (see (5), for review). It is accompanied by separation of neighboring base planes to allow the intercalation of the incoming base together with a corresponding change in the conformation of the sugar-phosphate backbone. In the case of BWEs, the integrating base expels the central nucleotide residue from an RNA trinucleotide segment. Such 'extruded' nucleotide may further engage in interactions with proteins or in tertiary contacts with other RNA regions. Even if a BWE was located on or close to the RNA molecule surface, the extruded nucleotide was shielded by either proteins or RNA–RNA interactions.

The formation of both BIE and BWE leads to a much closer contact between the RNA sites distant in sequence than in the case of other motifs performing the same function, namely, A-minor interactions, kissing loops and pseudoknots.

Also, it is important that the intercalating base retains the ability to participate in various interactions. In particular, appropriate donor groups of intercalated or wedged bases form hydrogen bonds with oxygen atoms of phosphate groups (see columns U-Z, AA-AC in Supplementary Table S1 and Figure 3A). Based on results of the comprehensive analysis of base-phosphate interactions in RNA undertaken by Zirbel *et al*. (23) we suggest that in this case the substantial contribution to the BIE and BWE stability is made by interaction of exocyclic amino groups of intercalated or wedged nucleobases with phosphate groups of neighboring nucleotides.

Also, about a tenth of the BIEs and BWEs included in the nrRNA 3.5 dataset are stabilized by lone pair-π (lp-π) interactions between O4′ of the ribose of an intercalated or wedged residue and the nucleobase ring of flanking or neighboring nucleotides (Supplementary Table S2; Figure 3B; see also Chalwa *et al*. (24) for a detailed description of this type of interactions).

In addition to stacking interactions between the intercalator and the flanking nucleobases the stability of BIE and BWE is usually maintained by stacking interactions of the latter with the adjacent nucleotide residues (see, for example, stacked C185 and A184, Figure 3A, right). In some cases, the flanking bases are paired with short trinucleotide segments of the same RNA and form so-called 'helical' BIE and BWE. In most but not all considered cases the values of B-factor of intercalator and flanking nucleotides reflecting their mobility in the analyzed structures are almost identical.

Base-intercalated elements formed by bis-(3′–5′)-cyclic dimeric guanosine monophosphate (c-di-GMP, secondary messenger controlling multiple cellular processes) forms with c-di-GMP-specific riboswitches are radically different from all the BIEs considered in this paper. In this particular case the flanking nucleotides belong to c-di-GMP, although in the c-di-GMP binding pocket αG- and βG-bases are also fixed by stacking interactions with the bases of the target RNA (25) (Supplementary Figure S1).

Selected examples of research

Short-range and long-range BIE and BWE. BIEs and BWEs taken altogether may be roughly divided into two groups: (i) formed by interaction of closely located nucleotide residues (short-range BIE and BWE); (ii) formed by interaction of residues that are not adjacent in the secondary structure of RNA (long-range BIE and BWE). The formation of quite common short-range BIEs with an intercalating base originating from the immediate neighbor of one of the BIE flanking residues allows this particular single-stranded stretch to fold into a rather compact structure or to participate in base-base paring with other RNA segments (Supplementary Figure S2).

Specifically, Figure 4 shows participation of two BIEs combined into a BIE-ZIP in compact folding of the three-way junction in the 3D structure of bacterial 5S ribosomal RNA. This BIE-ZIP is formed by two nucleotide residues of the helix II, one of which protrudes from this helix, and by two other nucleotide residues located within 3–4 nucleotide-long single-stranded elements of loop A, which connects helices II and V. Loop A of 5S rRNA inside the ribosome is coiled into a very complex 3D structure and BIE-ZIP takes part in this folding (26).

A representative example showing the participation of both short-range and long-range BIEs in the folding of the 3′-major domain of the *E. coli* 16S rRNA is shown in Figure 5. The BIEs A1318-A978-A1319 and A1362-G976-A1363 (*E. coli* base numbering) deserve particular attention. They participate in the folding of three remote parts of the 16S rRNA, i.e. two hairpin loops and one inner loop. At the same time, since the intercalating nucleotides in these motifs are separated by only one inner loop residue, all three loops are combined in one single ensemble. Ribosomal protein S14 is located in the immediate vicinity of this ensemble and together with several other SSU proteins and 16S rRNA segments forms a functionally important part of the head of the 30S subunit (28,29). These contacts, as well as many other 3D interactions in ribosomes, are conservative (see, for example, Figure 1 in (30)). Moreover, it was recently found that the tertiary structure of the 16S rRNA loop 1315–1322 shown in Figure 5 is almost identical to that of the BIE-containing segment of RNA of sweet clover necrotic mosaic virus (31).

In search of the ligands that control the function of FMN riboswitches, Howe *et al*. (32) devised an RNA aptamer consisting of two polynucleotide segments. We determined that these polynucleotides are connected via three BIEs and one BWE alongside with numerous classic Watson-Crick pairs (Supplementary Figure S3). In this case, two inter-

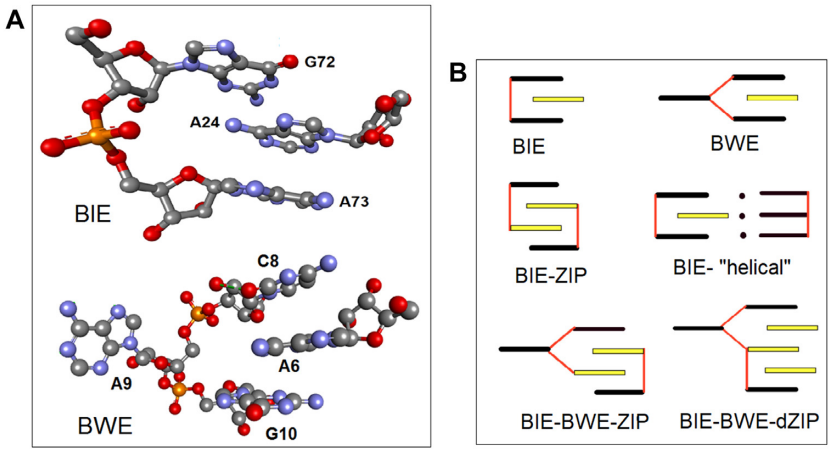

Figure 1. Overview of base-intercalated and base-wedged element structure. (**A**) The tertiary structure of representatives of BIE (PDB ID: 1Y27) and BWE (PDB ID: 7MSF) subfamilies. (**B**) Schematic representation of major types of BIEs and BWEs in RNA 3D structure. Yellow sticks represent intercalated or wedged bases; black sticks represent flanking and expelled bases; ribose-phosphate backbones are shown with thin red lines.

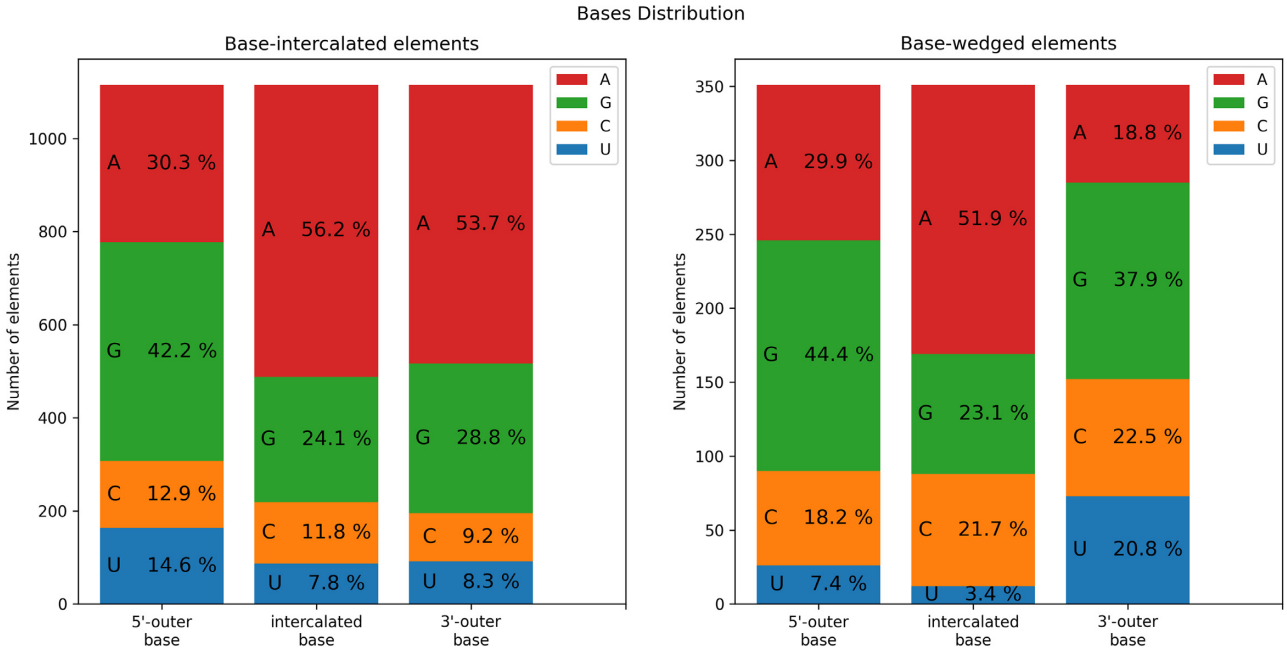

Figure 2. Percentage of nucleobases that form flanking and intercalated (wedged) residues in BIE and BWE.

calating bases belong to one polynucleotide chain and the other two originate from another chain.

Formation of BIE and BWE in RNA–protein complexes. The vast majority of motifs analyzed in this work is relevant to three-dimensional structures of RNA-protein complexes rather than free RNAs. For tRNA, there is a unique opportunity to trace the influence of proteins on the formation of BIE and BWE in the 3D RNA structure. Indeed, of all the available data in the nrRNA 3.5 dataset only in the case of tRNA the 3D structures of free RNA and its complexes with proteins can be directly compared. Our analysis shows that base-intercalated and base-wedged structural elements from the canonical set found in the pioneering work by Holbrook *et al.* (Supplementary Figure S4) (3,4) exist in many tRNAs in whole regardless of their specificity and source (Supplementary Table S1). However, in tertiary structures of many complexes of tRNAs with aminoacyl-tRNA synthetases and tRNA modifying enzymes this set is not fully represented (Supplementary Table S1). This is caused by changes in the conformation of certain segments of tRNA molecule, which are induced by these proteins. A notable exception is the complex of yeast $tRNA^{Arg}$ with specific aminoacyl-tRNA synthetase (33). In this complex, in addition to the canonical set of BIEs and BWE (circles A and B in Supplementary Figure S5), the protein-induced BIE C31-A37-C32 forms in tRNA anticodon loop (PDB ID: 1F7U) (circle C, Supplementary Figure S5). The conformation of this protein-induced BIE changes to the canonical form when the tRNA anticodon loop is involved

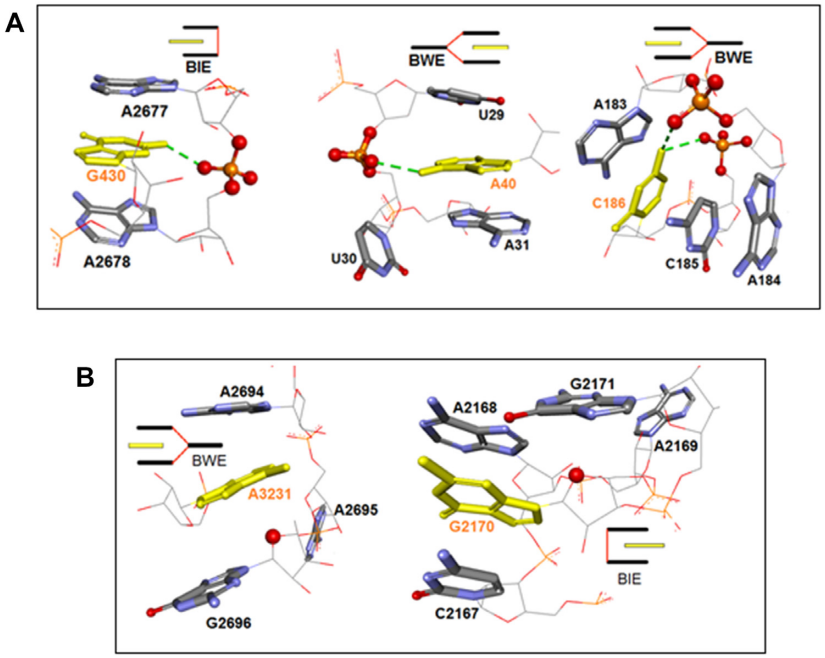

Figure 3. (**A**) Nucleobase-phosphate interactions that stabilize BIE (left; PDB ID: 3J79) and BWE (middle; PDB ID: 4M4O; right; PDB ID: 1MFQ) structure. (**B**) Stabilization of the BIE and BWE structures by lone pair-π ribose-base stacking interactions: (left) The O4′ atom of the ribose of the flanking nucleotide interacts with the base plane of the wedged nucleotide (PDB ID: 3J79), (right) Stacking interactions between the ribose residue of the intercalator and the base of neighboring nucleotide (PDB ID: 6AZ1).

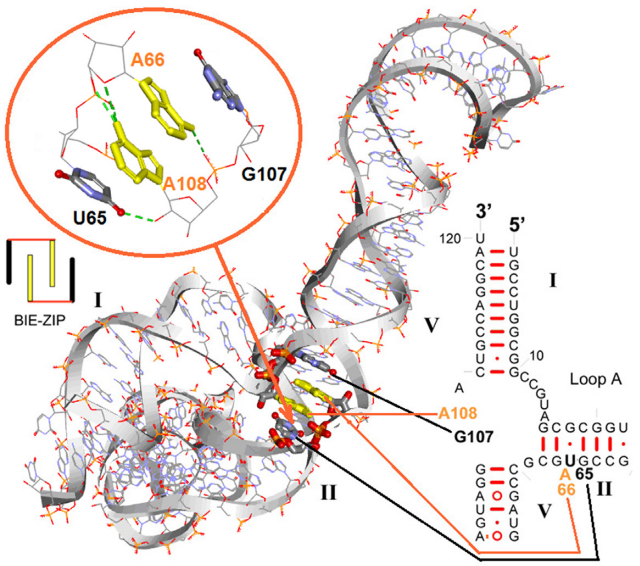

Figure 4. Crystal structure of *E. coli* 5S rRNA (PDB ID: 5MDV) and secondary structure of its loop A containing segment. Nucleobases and sugar-phosphate backbones of BIE-ZIP are colored as in Figure 1B.

in codon-anticodon interactions (Figure 6A, right). Collectively, these results point to the plasticity of both BIEs and BWEs as well as the structural elements of RNA in which they are embedded.

In the case of tRNAs, we have not found the evidence of direct contacts between proteins and RNA segments containing BIEs and BWEs. At the same time, we noticed that base-intercalated and base-wedged elements of many other RNAs directly interact with proteins. Figure 6B illustrates an example of such interaction. In addition to stabilization of the folding of its loops, BWE A5-A18-A7 binds directly to the surface of thrombin in a short synthetic RNA aptamer. In this case, amino acid residues of the protein interact with both extruded and flanking nucleotides of the element.

Direct and indirect participation of BIE and BWE in the organization of RNA functional sites. The nucleotide residues that form one of the two conservative base-intercalated elements in tRNA structure participate in the maintenance of a functionally important tRNA-elbow structural motif. In yeast $tRNA^{Phe}$ this motif consists of two BIEs (G18-G57-G19 and G57-G18-m^1A58), which are combined into the BIE-ZIP element (Supplementary Figure S4). The intercalation of guanine base 18 between the bases G57 and m^1A58 is followed by functionally important hydrogen-bonding of G18 with Ψ55 (U55 in some other tRNAs). In addition, this BIE-ZIP stabilizes the G19–C56 base pair (the key element of tRNA elbow) located at its tip. Collectively, these elbow structural motifs enable the direct interaction of tRNA with the ribosome, as well as binding with some aminoacyl-tRNA synthetases and with numerous tRNA-maturating enzymes (see (11), for review).

There are 15 different ribozymes in the nrRNA 3.5 dataset containing from one to eight BIEs in the 3D structure, 117 in total. Most of these elements are structure-forming. However, in few cases, they take part in the organization of catalytic centers of RNA enzymes. As an example, the fragment of the *Oceanobacillus iheyensis* group

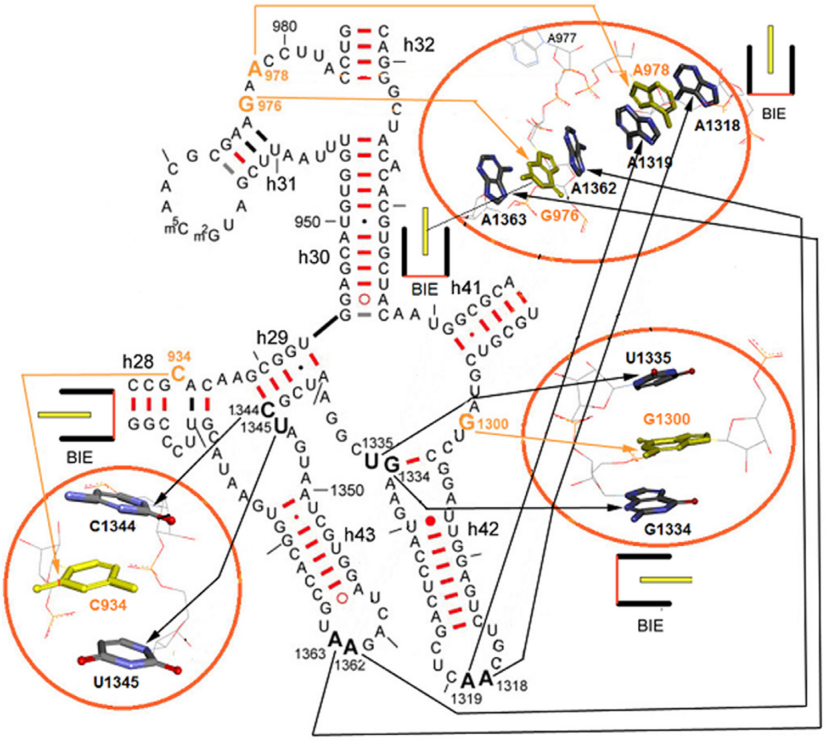

Figure 5. Location of four BIEs in the secondary structure of the segment of the 3′-major domain of the *E. coli* 16S rRNA (PDB ID: 5J7L). For simplicity, a canonical secondary structure (27) was used in this Figure instead of a tertiary rRNA structure. Nucleobases and sugar-phosphate backbones in BIEs of RNA structures are colored as in Figure 1B.

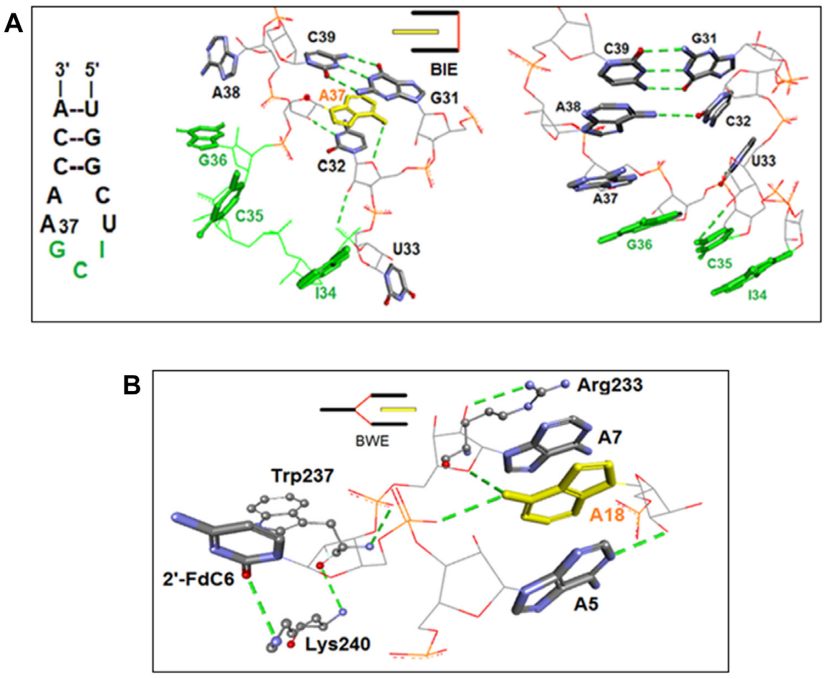

Figure 6. (**A**) Secondary (left) and tertiary (middle) structure of the anticodon loop of yeast $tRNA^{Arg}$ in the complex with yeast arginyl-tRNA synthetase (PDB ID:1F7U) and with the P-site of yeast ribosomes (mRNA codon is not shown) (PDB ID: 6T4G; right). (**B**) BWE interactions with amino acid residues located on the surface of thrombin in the complex with specific RNA aptamer (PDB ID: 3DD2).

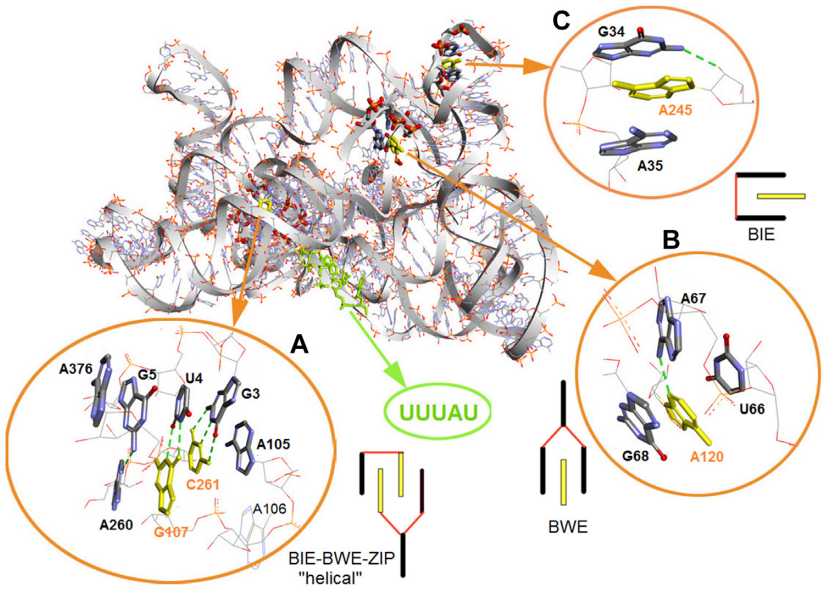

Figure 7. BIEs and BWEs of Oi5eD1–5 ribozyme and their location in the ribozyme crystal structure (PDB ID: 4FAW). Note that RNA crystals were obtained in the presence of potassium and magnesium ions. Nucleobases and sugar-phosphate backbones in BIEs and BWE of RNA structures are colored as in Figure 1B; oligonucleotide fragment hydrolyzed with ribozyme is colored in green.

II intron is a ribozyme with a high-resolution 3D structure studied in detail in (34–36). We determined that out of five BIEs identified in this RNA (Figure 7), two are located close to the ribozyme active center although they do not come into direct contact with it. The helical-like element BIE-BWE-ZIP A105-C261-G107-A260 (circle A, Figure 7) is of particular interest because it is included in the 'Z-anchor' loop enabling a complex multipoint contact of the *I*(*i*) part of the DI domain of the intron with its 5′-terminal region (34). In particular, the intercalated base of residue G107 forms a wobble pair with the intron's residue U4, while residues A260 and C261 form base pairs with residues G5 and G3, respectively. In addition, the guanine of residue G5 is stacked with A376, i.e. the key nucleotide residue in the catalytic center of the ribozyme (36).

Just as in the case of ribozymes, the majority of 141 BIEs and BWEs found in riboswitches are involved in the formation of the structure of single-stranded segments of RNA, as well as in the long-range interactions between their chains (37). Nevertheless, we identified several important scenarios these elements being involved in the organization of ligand-binding riboswitch-specific sites. For example, guanine-sensing 68-nucleotide long riboswitch from *Bacillus subtillis* (38) contains two BIEs (PDB ID: 1Y27). The BIE C61-A35-G62 plays a purely structural role and participates directly in hairpin loops kissing, and consequently, in the folding of the RNA macromolecule (Figure 8A). In this case, the intercalating nucleotide A35 forms a *cis*-Hoogsteen/Sugar Edge base pair with A64 (according to non-Watson-Crick base pair nomenclature (39)). The second BIE, G72-A24-A73 participates in the formation of the guanine-binding pocket of the riboswitch: its 3′-end flanking residue A73 is stacked with C74, which forms the base pair with the ligand. Thus, this BIE plays an important role in the riboswitch function.

Structural study of another purine-sensing riboswitch aptamer domain, i.e. the *add* adenine riboswitch from *Vibrio vulnificus* (40) allowed to trace the dynamic conformational switching in RNA upon ligand binding. The ligand-free riboswitch exists in two states shown in Figure 8B.

The ligand non-binding apo1 state (PDB ID: 5SWD; Chain B) is distinguished from the competent adenine-binding state apo2 (PDB ID: 5SWD; Chain A) by the presence of the BIE U48-A23-U49. The intercalating nucleotide A23 is mobile, and this BIE is eliminated during the course of apo1 → apo2 transition. Further, after binding a ligand the structure of the aptamer domain apo2 undergoes subsequent changes and transforms into the final guanine-bound state through an intermediate state. Importantly, BIE G72-A24-A73, which is located in the *add* adenine riboswitch (and adjacent to the BIE discussed above), does not change during these transitions. This very conservative BIE participates indirectly in the formation of guanine binding pocket of purine-sensing riboswitches (41). Here, we encounter a remarkable example of the pronounced conformational mobility of a nucleobase intercalated between two nucleotide residues which is involved in RNA function.

Base-wedged elements play key roles in organizing the binding sites of fluorophores with fluorogenic RNA aptamers. In the case of the homo dimeric Corn aptamer (42), the BWEs A10-G13-G12 and A21-G23-G22 are included in quadruplexes and occupy their opposite 'angles' with their flanking nucleotides directly interacting with the fluorophore (Figure 9, see also Figure 1C in (42)). In another fluorogenic RNA aptamer, iSpinach (43), the BIE G14-U47-G15 is also involved in the formation of the quadruplex structure of the fluorophore binding site but does not interact with the fluorogenic dye directly.

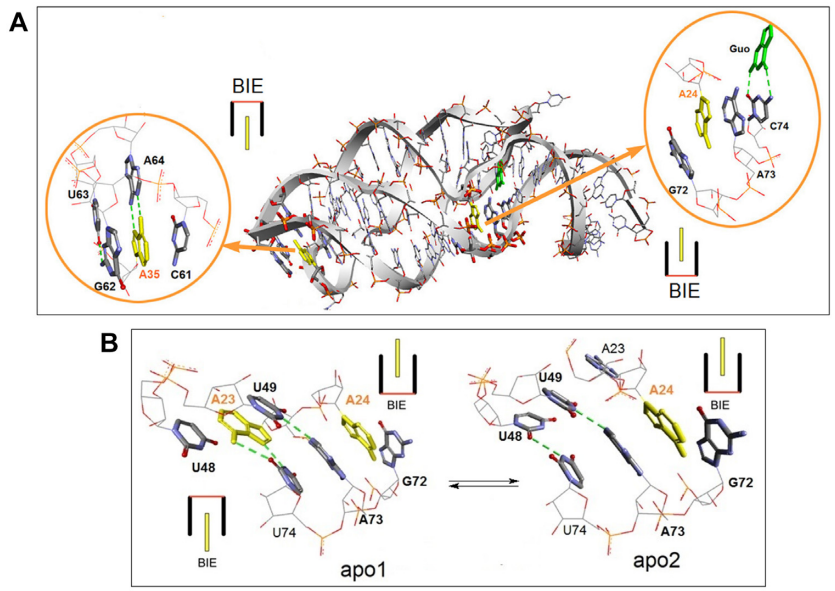

Figure 8. BIEs in purine-sensing riboswitches. (**A**) Participation of BIE G72-A24-A73 in the organization of the ligand-binding site of guanine - sensing riboswitch and location of BIEs in the crystal structure of riboswitch fragment (PDB ID: 1Y27). (**B**) Structure of the segments of *add* adenine riboswitch in apo1 and apo2 states. Structures were extracted from the crystal structure of the 71-nucleotide long aptamer domain of this RNA (PDB ID: 5SWD). Nucleobases and sugar-phosphate backbones in BIEs of RNA structures are colored as in Figure 1B.

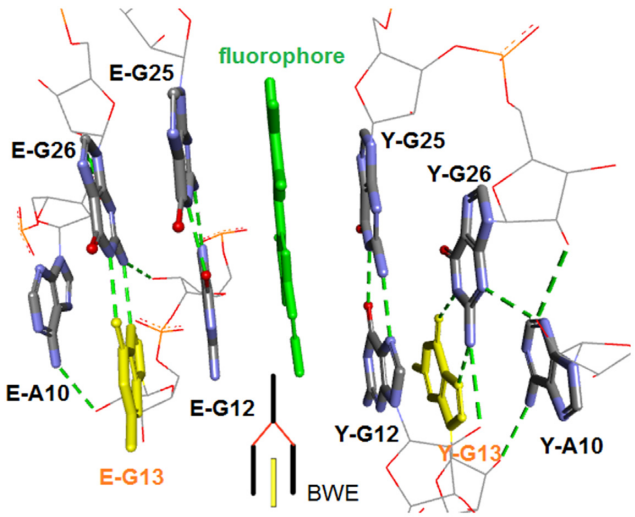

Figure 9. BWEs in two identical subunits of fluorogenic RNA (PDB ID: 5BJO). The 3′- end flanking and intercalating nucleotides of BWEs together with the G-residues with which they are paired are part of G-quartets. Two other BWEs (not shown) are located in the opposite corners of the G-quartets (see text for details). Fluorophore (3,5-difluoro-4-hydroxybenzylidene-imidazolinone-2-oxime, DFHO) is colored green.

Base-intercalated and base-wedged elements and ribosome functioning. It is known that ribosomal RNA is not only the most important structural component, but also the key functional component of the ribosome. The largest number of BIEs and BWEs which we identified by database mining are contained within the RNA of small and large ribosomal subunits. In accordance with the pronounced conservatism of the primary and secondary structures of rRNA, most of BIEs and BWEs are also conserved in the rRNA tertiary

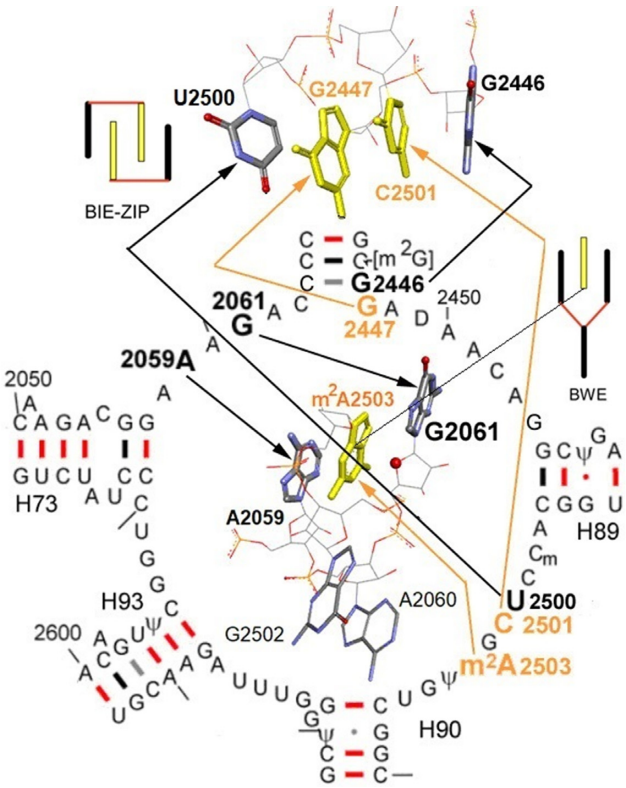

Figure 10. Location of functionally important BIE-ZIP and BWE in so-called 'peptidyl transferase circle' (46) of *E. coli* 23S rRNA (PDB ID: 5AFI). Note that nucleotide residue A2060 expelled from A2059–G2061 segment with m^2A2503 is stacked with G2502. The O4′ atom of ribose of G2061 which is involved in lone pair-π interactions with m^2A2503 is shown as red circle.

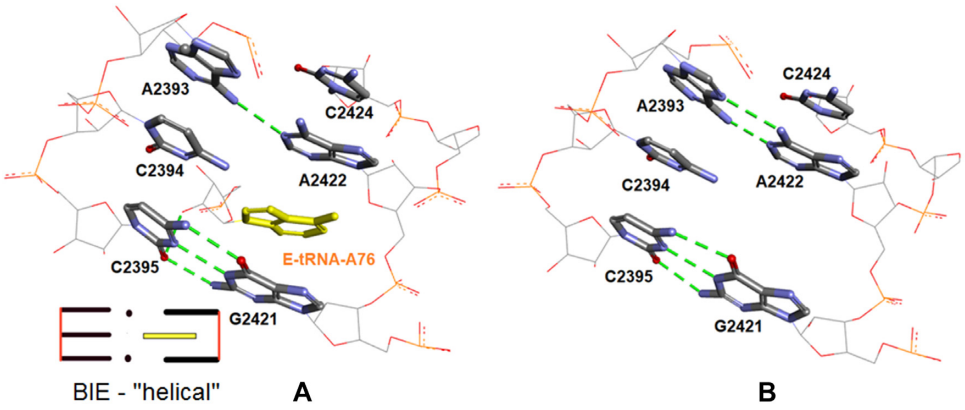

Figure 11. Structure of the 'pocket' in 23S rRNA for temporary binding the 3′-end residue of E-site tRNA in tRNA bound (**A**) and empty state (**B**), (PDB IDs: 4V90 and 4Y4O, correspondingly).

structures. Variable presence or absence of certain BIEs or BWEs in rRNA isolated from various organisms may be explained with differences in the rRNA size and composition as well as crystallization conditions. BIEs and BWEs are irregularly distributed within the secondary structure of rRNA. Almost half of these motifs are located in the 3′-terminal major and minor domains of small ribosomal subunit rRNAs. In large ribosomal subunit rRNAs BIEs and BWEs are converged mainly in Domain I, at the 5′-end of Domain II, and in Domain V. In the latter case most of them are located close to the peptidyl transferase center (PTC) and the nascent peptide exit tunnel (NPET) of the ribosome.

It is important that the formation and conformational retention of a number of BIEs and BWEs in rRNA is directly related to antibiotic resistance gained by bacteria as a result of rRNA nucleobase mutations or modifications. For example, the replacement of the 5′-end flanking C residue in the C1052-C1200-G1053 BIE by G in the *E. coli* 16S rRNA confers resistance to negamycine (44). Mutations of G2447, i.e. the key component of 23S rRNA BIEs (Figure 10) cause resistance to anisomycin and several other antibiotics (45–47). These mutations lead to noticeable alterations in rRNA conformation both in and out of the PTC (46).

The adenine base of the conservative nucleotide m^2A2503 of 23S-like rRNA intercalates between residues A2059 and G2061 (Figure 10) and participates in the formation of NPET walls of the ribosome. Hydrophobic C2- methyl group is likely to substantially increase the stability of this BIE. N6-methyl group of m^6A2030, which intercalates between G570 and U572 in 23S rRNA of *E. coli* ribosomes, also increases the stability of the corresponding element (48,49). It is worth noting, however, that residues m^2A2503 and G2061 form a stable contact not only due to nucleobase stacking but also due to previously described lone pair-π interactions between O4′ of the ribose of G2061 residue and the aromatic ring of m^2A2503**.** Interestingly, the nucleotide A2060, which is 'extruded' from the combination A2059-m^2A2053-G2061 ends up stacked with nucleotide G2502 and consequently plays an additional role in the organization of the inner surface of ribosomal NPET. The structure of this pentanucleotide ensemble is very conservative (Supplementary Figure S6). Residue m^2A2503 acquires its C2-methyl group through constitutive methylation during the assembly of bacterial ribosomes. In addition to C2-methylation, this residue may undergo selective methylation at the C8-position, which is catalyzed by Cfr methyltransferase (50). The unique chemical modification of adenine residue confers bacterial resistance to a wide range of antibiotics from various groups that bind to PTC and adjoin in the upper part of NPET (51). However, the functional role of m^2A2503, which intercalates between A2059 and G2061, is not limited to direct and indirect interaction with antibiotics: m^2A2503 is a member of the allosteric signal transmission pathway linking the inner regions of NPET with PTC (52,53). Therefore, mutations of m^2A2503 itself (54) and of flanking nucleotide residues (55) in this BWE drastically affect the ribosome functions.

As mentioned above, the LSU rRNA forms a pocket in the E-site of the ribosome into which an adenine base of the 3′-terminal residue of the deacylated tRNA intercalates (7) (Figure 11). This pocket is organized by two neighboring nucleotide residues and stabilized both by hydrogen bonds with a trinucleotide segment of the same rRNA (i.e. this base-intercalating element belongs to the BIE-'helical' type) and by stacking interactions with neighboring nucleotides. Intercalator leaves the pocket when the elongation factor II modulates a strong conformational change of the LSU and dissociation of tRNA from the ribosome. On the other hand, the C2394G mutation in this ensemble prevents intercalation of the 3′-terminal tRNA residue to the pocket and allosterically affects the activity of the elongation factor II which is located 70–100 angstroms away (56). It is remarkable that rRNA pocket structure remains unchanged and can be detected in tRNA-free ribosomes (compare e.g. PDB IDs 4V90 and 4Y4O). This particular example is unique among the cases that we analyzed so far. However, it is likely that other RNAs, which specifically bind ligands containing nucleotide residues, also carry pre-formed pockets for intercalators.

The functional importance of BIE formation can be discerned even more clearly in the case of 16S rRNA nucleobases intercalating into the mRNA polynucleotide chain (6,57). We found that the *Thermus thermophilus* 16S rRNA

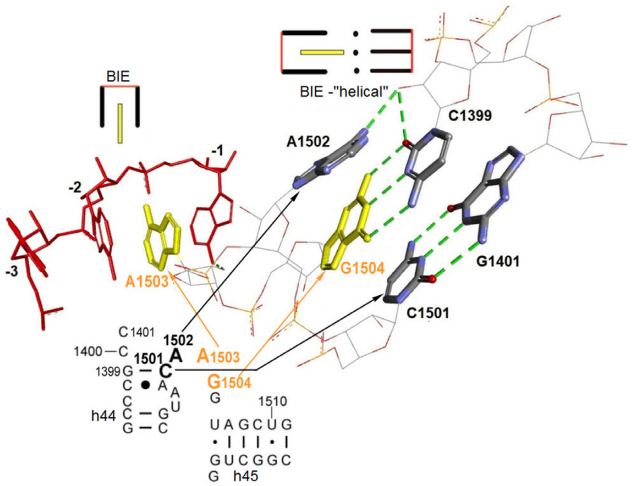

Figure 12. Interaction of mRNA with 16S rRNA nucleotide A1503 in pe*/E state of translating ribosome (PDB ID: 4V9L). The segment of the secondary structure of 16S rRNA and the connection of two BIEs are shown. mRNA chain is colored red, nucleobases and sugar-phosphate backbones in BIEs of RNA structures are colored as in Figure 1B.

nucleotide residue A1503 adopts a conformation optimal for intercalation between the –1 and –2 residues of a synthetic mRNA in association with C1501-G1504-A1502 BIE (Figure 12)

In order to allow the movement of mRNA from the intermediate state to the classical E-state, the intercalator must leave the BIE. Comparing the conformation of A1503 in the classical and intermediate states, we can conclude that this occurs due to changes in the conformation of the sugar-phosphate group that connects A1503 with G1504. This ensemble is conservative (Supplementary Figure S7), which indicates the uniqueness of this stage of the mRNA movement mechanism previously described only for *Thermus thermophilus* (6).

Thus, the search for BIEs and BWEs in ribosomal RNAs allows to detect almost all types and functional features of these elements found in other RNA. Moreover, it shows that the participation of BIEs and BWEs in the ribosome functioning is directly related to their mobility. We are confident that the rapid development of high-resolution cryogenic electronic microscopy of various RNA-containing molecules, supramolecular structures and organelles will greatly expand repertoire as well as structural and functional role of base-intercalated and base-wedged elements.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to A.V. Nazarenko and Dr A.A. Bogdanov, Jr. for help with preparation of this paper for publication. The authors thank the reviewers for careful reading of the manuscript and constructive suggestions.
Author contributions: E.B. performed the computational work and conformational analysis. V.M. analyzed data and prepared all figures. A.B. designed the project. All authors contributed significantly to the writing of the manuscript.

FUNDING

E.B. wishes to thank Institute of Mathematical Problems of Biology RAS - the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences; V.M. and A.B. acknowledge support from Lomonosov Moscow State University. Funding for open access charge: Authors' personal income.
Conflict of interest statement. None declared.

REFERENCES

1. Sponer,J., Riley,K.E. and Hobza,P. (2008) Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys. Chem. Chem. Phys.*, **10**, 2595–2610.
2. Butcher,S.E. and Pyle,A.M. (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.*, **44**, 1302–1311.
3. Holbrook,S.R., Sussman,J.L., Warrant,R.W. and Kim,S.H. (1978) Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J. Mol. Biol.*, **123**, 631–660.
4. Holbrook,S.R. and Kim,S.H. (1979) Intercalation conformations in single- and double-stranded nucleic acids. *Int. J. Biol. Macromol.*, **1**, 233–240.
5. Saenger,W. (1984) In: *Principles of Nucleic Acid Structure*. Chapter 16, Springer-Verlag, NY - Berlin – Heidelberg - Tokyo.
6. Zhou,J., Lancaster,L., Donohue,J.P. and Noller,H.F. (2013) Crystal structures of EF-G-ribosome complexes crapped in intermediate states of translocation. *Science*, **340**, 1236086.
7. Selmer,M., Dunham,C.M., Murphy,F.V. 4th, 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.
8. Bullock,L., Luke,D., Sherlin,L.D. and Perona,J.J. (2000) Tertiary core rearrangements in a tight binding transfer RNA aptamer. *Nat. Struct. Biol.*, **7**, 497–504.
9. Tereshko,V., Skripkin,E. and Patel,D.J. (2003) Encapsulating streptomycin within a small 40-mer RNA. *Chem. Biol.*, **10**, 175–187.
10. Krasilnikov,A.S. and Mondragon,A. (2003) On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA*, **9**, 640–643.
11. Krasilnikov,A.S., Yang,X., Pan,T. and Mondragon,A. (2003) Crystal structure of the specificity domain of ribonuclease P. *Nature*, **421**, 760–764.
12. Torres-Larios,A., Swinger,K.K., Pan,T. and Mondragon,A. (2006) Structure of ribonuclease P — a universal ribozyme. *Curr. Opin. Struct. Biol.*, **16**, 327–335.
13. Toor,N., Keating,K.S., Taylor,S.D. and Pyle,A.M. (2008) Crystal structure of a self-spliced group II intron. *Science*, **320**, 77–82.
14. Petrov,A.I., Zirbel,C.L. and Leontis,N.B. (2013) Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA*, **19**, 1327–1340.
15. Meyera,M., Nielsenb,H., Oliéricc,V., Roblinde,P., Johansenf,S.D., Westhof,E. and Masquidaa,B. (2014) Speciation of a group I intron into a lariat capping ribozyme. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 7659–7664.
16. Zhang,J. and Ferré-D'Amaré,A.R. (2016) The tRNA elbow in structure, recognition and evolution. *Life (Basel)*, **6**, E3.
17. Spasic,A., Kennedy,S.D., Needham,L., Manoharan,M., Kierzek,R., Turner,D.H. and Mathews,D. (2018) Molecular dynamics correctly models the unusual major conformation of the GAGU RNA internal loop and with NMR reveals an unusual minor conformation, *RNA*, **24**, 656–672.
18. Zuker,M., Mathews,D.H. and Turner,D.H. (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: Barciszewski,J. and Clark,B.F.C. (eds). *RNA Biochemistry and Biotechnology*. NATO ASI Series, Springer, Dordrecht, Vol. **70**, pp. 11–43.

19. Baulin,E., Yacovlev,V., Khachko,D., Spirin,S. and Roytberg,M. (2016) URS DataBase: universe of RNA structures and their motifs. *Database*, **2016**, baw085.
20. Rose,P.W., Prlić,A., Altunkaya,A., Bi,C., Bradley,A.R., Christie,C.H., Costanzo,L.D., Duarte,J.M., Dutta,S., Feng,Z. *et al.* (2017) The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.*, **45**, D271–D281.
21. Lu,X.J., Bussemaker,H.J. and Olson,W.K. (2015) DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.*, **43**, e142.
22. Leontis,N.B. and Zirbel,C.L. (2012) Nonredundant 3D structure data sets for RNA knowledge extraction and benchmarking. In: Leontis,N.B. and Westhof,E. (eds). *RNA 3D Structure Analysis and Prediction*. Springer, Berlin, Heidelberg, pp. 281–298.
23. Zirbel,C.L., Šponer,J.E., Šponer,J., Stombaugh,J. and Leontis,N.B. (2009) Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res.*, **37**, 4898–4918.
24. Chawla,M., Chermak,E., Zhang,O., Bujnicki,J.M, Oliva,R. and Cavallo,L. (2017) Occurrence and stability of lone pair–π stacking interactions between ribose and nucleobases in functional RNAs. *Nucleic Acids Res.*, **45**, 11019–11032.
25. Smith,K.D., Shanahan,C.A., Moore,E.L., Simon,A.C. and Strobel,S.A. (2011) Structural basis of differential ligand recognition by two classes of bis-(3′-5′)-cyclic dimeric guanosine monophosphate-binding riboswitches. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 7757–7762.
26. Szyman,M., Barciszewska,M.Z., Erdmann,V.A. and Barciszewski,J. (2003) 5S rRNA: structure and interactions. *Biochem. J.*, **371**, 641–651.
27. 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.
28. Wimberly,B.T., Brodersen,D.E., Clemons,W.M. Jr, Morgan-Warren,R.J., Carter,A.P., Vonrhein,C., Hartschk,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
29. Hoang,L., Fredrick,K. and Noller,H.F. (2004) Creating ribosomes with an all-RNA 30S subunit P site. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 12439–12443.
30. Hosseini,V., Roy,P., Sissler,V., Zirbel,E., Westhof,E. and Leontis,N. (2018) How to fold and protect mitochondrial ribosomal RNA with fewer guanines. *Nucleic Acids Res.*, **46**, 10946–10968.
31. Steckelberg,A.-L., Akiyama,B.M., Costantino,D.A., Sit,T.L., Nix,J.C. and Kieft,J.S. (2018) A folded viral noncoding RNA blocks host cell exoribonucleases through a conformationally dynamic RNA structure. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 6404–6409.
32. Howe,J.A., Wang,H., Fischmann,T.O., Balibar,C.J., Xiao,L., Galgoci,A.M, Malinverni,J.C., Mayhood,T., Villafania,A., Nahvi,A. *et al.* (2015) Selective small-molecule inhibition of an RNA structural element. *Nature*, **526**, 672–677.
33. Delagoutte,B., Moras,D. and Cavarelli,J. (2000) tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. *EMBO J.*, **19**, 5599–5610.
34. Toor,N., Keating,K.S., Taylor,S.D. and Pyle,A.M. (2008) Crystal structure of a self-spliced group II intron. *Science*, **320**, 77–82.
35. Keating,K.S., Toor,N., Perlman,P.S. and Pyle,A.M. (2010) A structural analysis of the group II intron active site and implications for the spliceosome. *RNA*, **16**, 1–9.
36. Marcia,M. and Pyle,A.M. (2012) Visualizing group II intron catalysis through the stages of splicing. *Cell*, **151**, 497–507.
37. Batey,R.T. (2012) Structure and mechanism of purine-binding riboswitches. *Q. Rev. Biophys.*, **45**, 345–381.
38. Serganov,A., Yuan,Y.-R., Pikovskaya,O., Polonskaia,O., 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.
39. Leontis,N.B., Stombaugh,J. and Westhof,E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
40. Stagno,J.R., Liu,Y., Bhandari,Y.R., Conrad,C.E., Panja,S., Swain,M., Fan,L., Nelson,G., Li,C., Wendel,D.R. *et al.* (2017) Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature*, **541**, 242–246.
41. Stoddard,C.D., Widmann,J., Trausch,J.J., Marcano-Velazquez,J.G., Knight,R. and Batey,R.T. (2013) Nucleotides adjacent to the ligand-binding pocket are linked to activity tuning in the purine riboswitch. *J. Mol. Biol.*, **425**, 1596–1611.
42. Warner,K.D., Sjekloća,L., Song,W., Filonov,G.S., Jaffrey,. R. and Ferré-D'Amaré,A.R. (2017) A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. *RNA*, **23**, 1788–1795.
43. Fernandez-Millan,P., Autour,A., Ennifar,E., Westhof,E. and Ryckelynck,M. (2017) Crystal structure and fluorescence properties of the iSpinach aptamer in complex with DFHBI. *Nat. Chem. Biol.*, **13**, 1195–1201.
44. Cocozaki,A.I., Altman,R.B., Huang,J., Buurman,E.T., Kazmirski,S.L., Doig,P., Prince,D.B., Blanchard,S.C., Cate,J.H. and Ferguson,A.D. (2016) Resistance mutations generate divergent antibiotic susceptibility profiles against translation inhibitors. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 8188–8193.
45. Blaha,G., Gurel,G., Schroeder,S.J., Moore,P.B. and Steitz,T.A. (2008) Mutations outside the anisomycin-binding site can make ribosomes drug-resistant. *J. Mol. Biol.*, **379**, 505–519.
46. Kloss,P., Xiong,L., Shinabarger,D.L. and Mankin,A.S. (1999) Resistance mutations in 23 S rRNA identify the site of action of the protein synthesis inhibitor linezolid in the ribosomal peptidyl transferase center. *J. Mol. Biol.*, **294**, 93–101.
47. Xiong,L., Kloss,P., Douthwaite,S., Andersen,N.M., Swaney,S., Shinabarger,D.L. and Mankin,A.S. (2000) Oxazolidinone resistance mutations in 23S rRNA of Escherichia coli reveal the central region of domain V as the primary site of drug action. *J. Bacteriol.*, **182**, 5325–5331.
48. Kierzek,E. and Kierzek,R. (2003) The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. *Nucleic Acids Res.*, **31**, 4472–4480.
49. Sergiev,P.V., Golovina,A.Y., Osterman,I.A., Nesterchuk,M.V., Sergeeva,O.V., Chugunova,A.A., Evfratov,S.A., Andreianova,E.S., Pletnev,P.I., Laptev,I.G. *et al.* (2016) N6-Methylated adenosine in RNA: from bacteria to humans. *J. Mol. Biol.*, **428**, 2134–2145.
50. Giessing,A.M., Jensen,S.S., Rasmussen,A., Hansen,L.H., Gondela,A., Long,K., Vester,B. and Kirpekar,F. (2009) Identification of 8-methyladenosine as the modification catalyzed by the radical SAM methyltransferase Cfr that confers antibiotic resistance in bacteria. *RNA*, **15**, 327–336.
51. Smith,L. and Mankin,A. (2008) Transcriptional and translational control of the *mlr* operon, which confers resistance to seven classes of protein synthesis inhibitors. *Antimicrob. Agents Chemother.*, **52**, 1703–1712.
52. Vázquez-Laslop,N., Ramu,H., Klepacki,D., Kannan,K. and Mankin,A.S. (2010) The key function of a conserved and modified rRNA residue in the ribosomal response to the nascent peptide. *EMBO J.*, **29**, 3108–3117.
53. Makarov,G.I., Golovin,A.V., Sumbatyan,N.V. and Bogdanov,A.A. (2015) Molecular dynamics investigation of a mechanism of allosteric signal transmission in ribosomes. *Biochemistry (Moscow)*, **80**, 1047–1056.
54. Koch,M., Willi,J., Pradère,U., Hall,J. and Polacek,N. (2017) Critical 23S rRNA interactions for macrolide-dependent ribosome stalling on the ErmCL nascent peptide chain. *Nucleic Acids Res.*, **45**, 6717–6728.
55. Sergiev,P.V., Lesnyak,D.V., Burakovsky,D.E., Svetlov,M., Kolb,V.A., Serebryakova,M.V., Demina,I.A., Govorun,V.M., Dontsova,O.A. and Bogdanov,A.A. (2012) Non-stressful death of 23S rRNA mutant G2061C defective in puromycin reaction. *J. Mol. Biol.*, **416**, 656–67.
56. Sergiev,P.V., Lesnyak,D.V., Kiparisov,S.V., Burakovsky,D.E., Leonov,A.A., Bogdanov,A.A., Brimacombe,R. and Dontsova,O.A. (2005) Function of the ribosomal E-site: a mutagenesis study. *Nucleic Acids Res.*, **33**, 6048–6056.
57. Hong,S., Sunita,S., Maehigashi,T., Hoffer,E.D., Dunkle,J.A. and Dunham,C.M. (2018) Mechanism of tRNA-mediated +1 ribosomal frameshifting. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 11226–11231.