Published in final edited form as:

Wiley Interdiscip Rev RNA. 2013 ; 4(2): . doi:10.1002/wrna.1153.

The GA-minor submotif as a case study of RNA modularity, prediction, and design

Wade W. Grabow¹, Zhuoyun Zhuang¹, Joan-Emma Shea^{1,2}, and Luc Jaeger^{1,3,*}

¹Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA, USA

²Department of Physics, University of California, Santa Barbara, CA, USA

³Bio-Molecular Science and Engineering Program, University of California, Santa Barbara, CA, USA

Abstract

Complex natural RNAs such as the ribosome, group I and group II introns, and RNase P exemplify the fact that three-dimensional (3D) RNA structures are highly modular and hierarchical in nature. Tertiary RNA folding typically takes advantage of a rather limited set of recurrent structural motifs that are responsible for controlling bends or stacks between adjacent helices. Herein, the GA minor and related structural motifs are presented as a case study to highlight several structural and folding principles, to gain further insight into the structural evolution of naturally occurring RNAs, as well as to assist the rational design of artificial RNAs.

INTRODUCTION

A considerable amount of work has been invested in understanding the principles of RNA folding and structural composition.^{1–6} It is generally agreed that, in the presence of metal ions, RNA folding occurs rapidly by forming secondary structures that collapse into compact conformers before undergoing a slower process during which structural rearrangement and conformational search occur to reach the final native state.^{7–9} While the former is driven by the formation of RNA helices and charge neutralization, the latter stage is governed by the presence of recurrent RNA structural motifs that control how helical elements are able to bend, stack, and pack.^{2,10}

It is somewhat surprising to find that elaborate three-dimensional (3D) RNA architectures are often constructed from a limited number of recurrent structural motifs defined by prevalent sequence patterns that are dictated by a select set of conserved and semi-conserved nucleotides. Thus far, a significant number of structural motifs (partially reviewed in Refs 4 and 11 and Table 1) have been independently identified and described. They include small submotifs (such as the U-turn,¹² the A-minor interaction,^{13,14} the UA_handle,⁶ and the ribose zipper¹⁵), terminal and internal loops (such as triloops,^{16,17} tetraloops,^{18–20} sarcin loop,^{21,22} and T-loop^{6,23,24}), turns and junctions (such as kink-turns,^{25,26} hook turns,^{27,28} G-ribo and UA_h turns,^{6,29} and multihelix junctions^{6,30,31}), and long-range interactions and pseudoknots (such as kissing loops,^{32,33} along-groove packing motifs,^{34,35} GNRA/tetraloop interactions,^{36,37} and pseudoknot domains^{6,38}). However, very few attempts have been made to thoroughly categorize and possibly relate these (sub)motifs to one another.⁶ Such studies are valuable because they highlight the fundamentally modular nature of the RNA. Identifying the inherent levels of hierarchy and modularity that exist between distinct

structural motifs offers the potential to understand evolutionary relationships between naturally occurring RNAs as well as to gain insight into the sequence space signature of a given RNA structural component.

Determining the relationship among various motifs is useful for predicting RNA 3D structure and assisting with the rational design of novel RNA molecules.^{65–67} In this regard, the RNA architectonics strategy, which aims at using RNA structural motifs as building blocks for controlling the self-assembly of RNA nanoarchitectures, offers a valuable experimental approach toward addressing the characterization of RNA motifs and their role in the RNA folding (e.g., Refs 53, 54, and 68–70). The syntax network of RNA motifs described herein—by showing the modular and hierarchical nature of RNA structural motifs—offers an abstracted view of the principles of RNA architectonics strategy in order to provide a causative bridge between information encoded within the sequence space of a given molecule and the particular 3D fold that an individual RNA naturally prefers.⁸ Rather than simply relying upon reverse engineering to unravel the sequence/tertiary structure relationship of RNA, this higher level understanding of RNA structures and their interrelatedness offers the traditional RNA architectonics strategy a more predictive posture for designing and constructing new RNA architectures.

RNA BASE PAIRS AND SMALL PREVALENT STRUCTURAL MOTIFS

RNA nucleotides have the ability to form at least two hydrogen bonds between three different edges of the base: the Watson–Crick (WC), Hoogsteen (HG), and shallow groove (SG) edges. Given that there are two possible orientations of the nucleotides with respect to one another (*cis* and *trans*) there are 12 possible types of base pairs (bps),⁷¹ among which 11 are called noncanonical bps. From this categorization, several structural 3D patterns, also called RNA structure motifs, have been identified in stable RNAs such as the ribosome, ribozymes, and riboswitches.^{4,5,11,42,72} These recurrent 3D patterns can be defined by sequence space signatures that correspond to sets of conserved or semiconserved nucleotides that are able to fold into specific structural conformers (with well-defined conformation of the ribose-phosphate backbone). The simplest and most recognizable RNA motif is the A-form helix which results from the assembly of complementary nucleotide strands through canonical WC:WC *cis* bps - the most abundant type of bps in natural RNAs. Nevertheless, some noncanonical bps are also highly prevalent in RNAs^{63,73}: these include the U:A (WC:HG) *trans*, the G:A (HG:SG) *trans*, and the A:G (SG:SG) *trans* noncanonical bps. As such, small recurrent structural motifs such as the GA-sheared,⁴⁴ bulged-G,²¹ or A-minor submotifs^{13,14,41} often result from a combination of a limited number of both canonical and noncanonical bps.

Within the same category of noncanonical bps, bps can often be classified further according to their tilt, propeller twist angle and hydrogen-bonding patterns (Figure 1(a)). These structural variations are often indicative of the larger structural context to which these bps belong and, as such, can be used as markers of a particular RNA structural motif. For example, no less than four different types of A:G (SG:SG) *trans* bps can be distinguished (Figure 1(a) and (b)). Interestingly, type I-P (planar) and type I-T (tilted) A:G (SG:SG) *trans* bps are mostly identified in A-minor submotifs, the most abundant class of tertiary contacts identified in the ribosome and other stable RNAs. Typically, A-minors involve the insertion of the shallow groove edges of adenines into the minor shallow groove of neighboring helices.^{14,37,41} On the basis of the positioning of the interacting adenine with respect to the interacting WC bp, it has been possible to distinguish four types of A-minor (types 0, I, II, and III described in Ref 14). Of these, type I-P and type I-T A-minors are unique in that the adenine fits into the minor groove of a classic G:C WC:WC *cis* bp by forming an A:G SG:SG *trans* bp in such a way that maximizes the number of hydrogen bonds.¹⁴ In contrast

to type I-P and type I-T G:A (SG:SG) *trans* bps [Figure 1(b) (right two panels)], the twisted (type I-Tw) and supertwisted (type I-STw) G:A (SG:SG) *trans* bps [Figure 1(b) (left two panels)] have propeller twist angle of 90° and 35–45°, respectively. Although type I-Tw can form the same H-bond pattern as type I-P and type I-T bps, type I-STw A:G SG:SG *trans* bps cannot.³⁷ Remarkably, both type I-Tw and type I-STw bps are characterized by C2 endo sugar pucker at the level of the interacting adenine (Figure 1(b)). Note that this sugar pucker is seldom (if ever) observed for classic A-minor interactions of type I-P or I-T that typically have their interacting adenosine in C3' endo.^{14,37,41} Interestingly, type I-Tw and type I-STw A:G SG:SG *trans* bps are usually linked to a regular WC:WC *cis* bp (Figure 1(c) and (d)) rather than resulting from an A-minor interaction typically formed between a third strand intercalating an existing helix. As such, they are particular structural features in their own right of another motif that we called the GA-minor submotif (or G:A SG:SG *trans* bp at helix ends).

THE GA-MINOR MOTIF

Previously, A:A and G:A bps at the ends of RNA helices (GA@helix.ends motifs) were reported to involve either HG:SG *trans* bps (sheared bp) or WC:WC bps.⁴⁴ While not as abundant as these two noncanonical bps, G:A SG:SG *trans* bps at helix ends (or GA-minor motifs) are nevertheless prevalent components in structured RNAs. An exhaustive search for GA-minor motifs using FR3D⁷⁵ led to the identification of 39 occurrences of this motif within known X-ray structures (see Table 2). The GA-minor motif consists of a conserved set of four to five nucleotides that form a A(1):G(4) SG:SG *trans* bp directly linked to a classic X(2):X(3) WC bp (Figure 1(c) and (d)). As discussed above, two types of GA minors can be distinguished according to the presence of a type I-STw or type I-Tw A(1):G(4) SG:SG *trans* bp [Figure 1(b) (left two panels)]. While both types of GA minors are characterized by an H-bond between atom N1 of A(1) and the 2' OH of G(4), the base of G(4) is perpendicular to A(1) in type I-STw (far left panel) and cannot form additional H-bond between its N2 position and atom N3 of A(1) (Figure 1(b) and (c)). Instead, N2 of G(4) interacts with the 2' OH of A(1). For type I-Tw GA minors, the stacking interactions between the base of A(1) and the ribose (or base) of N(5) can provide additional stabilization (Figure 1(d)). Because the G:A SG:SG *trans* bp is linked to a WC:WC bp, the sugar puckering for A(1) is forced to be C2' endo. Consequently, the local backbone of A(1) in GA minors is almost oriented parallel to the one of G(4). This local reversal of strand directionality can therefore promote the formation of turns (Figures 2 and 3). Note that one or more nucleotides can be inserted between position A(1) and X(2) in the type I-Tw GA minor, resulting in the GA-minor bulge submotif (Figure 4). However, because these variants are seldom observed in stable RNAs, they were not systematically searched among known X-ray structures.

THE HIERARCHICAL 'GA-MINOR SYNTAX NETWORK'

The GA minor constitutes a new category of submotif, which can, like several other submotif or small RNA motifs such as the UA_handle,⁶ the G/RA,⁴⁴ U-turn,^{12,43} and A-minor,^{12–14,41} enter into the composition of a variety of other motifs to generate patterned RNAs possessing greater structural complexity.^{6,78} Careful examination of the various occurrences of this motif in known RNA structures demonstrates the way in which the GA minor can be combined with other structural elements to generate important structural components of more complex RNA motifs (Table 2). While the type I-STw is a special case of GA minors that is only found in the RA motif (Figure 2), the more abundant type I-Tw GA minor is present in a greater variety of larger structural motifs including the right angle (RA), kink-turn, and two-helix stack (2h_stack) motifs (Figures 2, 3, and 5).

The RA Motif (or RA Turn)

The RA motif was originally identified in the ribosome.^{42,63,64} It was first described as a central structural component in programmable self-assembling tectoRNA nanostructures.^{53,54} The RA motif is characterized by 13 nt positions (with the 13th position being the least conserved) specifying for a 90° bend between two adjacent helices (H5 and H3) that are separated by two conserved nucleotides^{53,54} (Figures 2(a)–(c) and 3(a)). Although the RA motif was originally described as the combination of a ‘hook-turn’ motif^{27,28} and an ‘along-groove packing’ motif,^{34,35} it is best described as a combination of two GA-minor motifs stabilized by the along-groove packing interaction⁷¹ (Figure 2(a), (d), (e), and (f)). The RA motif is always formed through a type I-STw GA minor localized in 5 of a type I-Tw GA minor. Nucleotide conservations (Figure 2(a)) can be justified by the fact that they contribute to the specific formation of the two conserved G:A SG:SG *trans* bps (shown in blue, Figure 2(c)) that are stacked onto one another and constrain the formation of a sharp turn at the level of the ribose-phosphate backbone connecting the stems H5 and H3 (Figure 2(b) and (c)). Because type I-STw and type I-Tw G:A bps are not perfectly identical, with A(1):R(6) on the top of H5 being significantly more twisted than A(7):g(12) on the top of H3, position N(13) of the RA motif can be in perfect helical continuity with H3 with its ribose in stacking interaction with the base of A(7) (Figure 2).

On the basis of the X-ray structures of the rRNAs (see Table 2), the H5 and H3 stems are arranged similar to the corner of a log cabin (Figure 2(b)), with the two helical stems packed along their shallow grooves through ribose-zipper interactions. The ‘along-groove packing’ (shown in rose, Figure 2(a)–(c)) typically involves the formation of a total of 11 interhelical H-bonds between three classic G:C WC bps and one G:U wobble bp: two of the G:C bps interact in a symmetrical fashion (Figure 2(e)) and the other G:C is in quasi-symmetrical interaction with the G:U wobble bp (Figure 2(f)). Because of the quasi-symmetry of the packing interaction, the G:U can be found either in H3 or H5 without affecting the overall geometry of the RA motif.

The Kink-Turn Motif

In addition to the RA motif, the GA minor of type I-Tw enters into the formation of the kink-turn motif (also referred as k-turn in the literature), another widespread motif found in numerous RNA molecules including rRNAs,^{26,55,56,72,79} riboswitches,⁸⁰ and RNP complexes^{55,81} (see Figure 3(b) and Table 2). In the case of the kink-turn, the GA minor is associated with an A-rich motif, typically a GA-sheared (GAA:ggA or GA:gA) or loop E motifs, to form an A-minor packing interaction between one of the conserved adenines from the A-rich motif and positions X(2):X(3) from the GA minor^{82,83} (Figure 3(b)). The SG:SG *trans* bp of the GA minor from the kink-turn is typically a G:A, but other nucleotides with the same bp geometry might be occasionally found.⁷² Interestingly, the X(2):X(3) WC bp of the GA minor is usually a G:C bp in both RA and kink-turn motifs: the same sequence conservation is required for the formation of the optimal along-groove packing and A-minor packing interactions in the RA and kink-turn, respectively (Figure 2(a)).

In contrast with the RA motif, the kink-turn utilizes no more than five H-bonds at the level of its A-minor packing interaction. While the GA minor found in the kink-turn can occasionally have the base of A(1) stacked with the ribose of N(5) like in the RA motif, A(1) is usually stacked with the base of N(5) instead (Figure 3(b)). As a result of these interactions, the kink-turn approximately forms a 45° angle between two helices (Figure 3(b)). Because kink-turn involves significantly less H-bond contacts than the RA motif (5 for the kink-turn vs 11 for the RA), it is expected to be less rigid. As a matter of fact, kink-turns adopt slightly different conformations depending on their structural context within

larger RNA molecules or upon binding of proteins,^{84–86} suggesting that they might be involved in RNA motion and dynamics.

The GA-Minor 2h_stack Motif

While the GA minor is most commonly associated with helical bends, it has also been identified with a small subset of two helix stacks (2h_stack) (Figure 3(c)). Adjacent contiguous A-form RNA helices can be stacked onto one another to form 2h_stack motifs that are often stabilized by additional tertiary bp contacts involving either the 5' or 3' leaving strands.^{57,87–90} 2h_stack motifs can be stabilized by various triple bp contacts that involve either the minor groove of H3' or the major groove of H5' to form A-minor, GA minor, or major 2h_stacks (Figure 4)—among which the most prevalent are A-minor 2h_stacks (t1 and t0).¹³ Although less prevalent than the A-minor 2h_stack, at least six instances of the GA-minor 2h_stack have been identified in known X-ray structures of the ribosome (Table 2). In this motif, a regular A-form helix (H3') is joined to the 3' side of a GA-minor type Tw (Figure 3(c)). The C2' endo sugar pucker of position A(1) favors a parallel orientation of the 5' leaving strand with respect to the connecting strand (Figures 3(c) and 4(a)). Meanwhile, the highly conserved G in position 2 is typically involved in a WC:WC *cis* bp that constitutes the first W:C bp of H3'.

Motifs of Greater Complexity Derived from the GA-Minor Motif

RNA structures are highly modular; meaning that they are constructed from different combinations of shared components. In this regard, RNA structures can be categorically related to one another based on the make up of their fundamental structural components (i.e., recurrent RNA motifs and/or submotifs). By identifying the sequence space signatures of recurrent motifs and categorically relating them to other structural motifs, an RNA motif lexicon can be generated and used to build structural motif syntax networks. Such networks provide the foundation of the RNA syntax. In this regard, the GA-minor structural motif syntax network (Figure 5) shows how the GA-minor submotif can be combined to other prevalent motifs like the along-groove packing, the 2h_stack, the A-minor, and G/RA motif to generate the RA turn, kink-turn, and GA-minor 2h_stack. All three of these resulting motifs can be expanded to generate motifs of increasing size and complexity largely based on the GA minor's ability to program the formation of bends or stacks between adjacent helices (Figure 6). For instance, while both the GA-minor h_stack and A-minor 2h_stack are stabilized through SG:SG *trans* bps between the 5' strand and the shallow groove of H3', the GA-minor 2h_stack uniquely allows a complete change of orientation of the 5' strand that better accommodates turns because of its twist angle and positioning of the nucleotide in A(1). In this regard, it presents an ideal structural element for extending into a GA/A-minor 2h_stack (two instances) (Table 2). This motif of greater complexity combines a GA-minor 2h_stack with an A-minor 2h_stack (Figure 6) and can also be seen as a variant of the A-minor type t1 2h_stack.

The GA minor takes advantage of the modular nature of RNA as well as its own ability to direct bends or stacks to create increasingly more complex motifs in several additional cases. For example, the GA minor found in the RA motif can be combined with the A-minor/G-major 2h_stack motifs to generate the expanded RA turn present in domain H29-H30 of archaeal 23S rRNAs (Figure 6). Similarly, the GA-minor submotif present in kink-turns, when combined with the G-major 2h_stack, forms the expanded kink-turn found in the TPP riboswitch (Figure 6). Furthermore, just as the GA-minor submotif can combine with the 2h_stack to form the GA-minor 2h_stack motif, the GA minor responsible for the RA motif can combine with the 2h_stack in the same way to form the expanded RA 2h_stack motif found at the P2.1_P3_P8 junction of group IC and ID introns.³⁹ In the case of the 16S ribosomal RNA, the GA minor in the RA is largely responsible for the makeup of

the H3-H4-H18 junction (Figure 6). Finally, the GA minor found in the kink-turn allows the expansion of junctions of three (three-way junction, H4.1-H5-H10 found in 16S rRNA) and five helices (five-way junction, H12-26-27-36-46 junction found in 23S rRNA) (Figure 6). Clearly, like the P2.1_P3_P8 junction of group IC and ID introns (which has not been crystallized yet), additional complex motifs that take advantage of the GA minor are expected to exist in nature even though they have yet to be observed in the present set of X-ray RNA structures.

PRINCIPLES OF RNA TERTIARY STRUCTURES

Principles of Structural Equivalence

The ‘GA-minor syntax network’ presented here not only extends the previous ‘UA_h motif syntax network’⁶ and ‘A-minor junction syntax network’¹³ but also reinforces the remarkable modularity and hierarchical build up of RNA molecules by elucidating three equivalence principles at the foundation of RNA modularity (see also Refs 6 and 66). These principles relating to three distinct aspects of modularity involve: (1) isosteric equivalence, (2) structural equivalence, and (3) functional/topological equivalence (Figure 7). At their core, all three principles arise from the observation that the small prevalent building blocks of RNA, resulting from a particular set of conserved and semiconserved nucleotides, are often interchangeable. Furthermore, we will see how each principle presented below can be seen as representing a specific subclass of the next.

- *Isosteric equivalence*: it corresponds to structures with different nucleotide sequences that are almost perfectly superimposable. In the most basic sense, classic WC bps are considered to be isosteric as they are perfectly superimposable to one another. This nature of WC bps constitutes a key characteristic of RNA modularity in that it allows structural elements to be parsed into simpler forms, so that specific portions of the individual structural components when combined retain the ability to isosterically overlap with other motifs (Figure 7(a)). For example, the GA minor and along-groove motifs in the case of the RA can recombined to generate motifs having greater complexity because of the single WC bp overlap that exists between the two submotifs (Figures 2(a) and 7(a)).
- *Structural equivalence*: it is the guiding principle behind the characterization of RNA structural motifs. RNA motifs obeying a given sequence signature, but whose structures are mostly superimposable (i.e., not necessarily superimposable at every nucleotide position) are said to be structurally equivalent. For example, kink-turn motifs that define a characteristic 45–60° turn stabilized by an A-minor interaction can take advantage of either a loop E or a G-shared loop—two types of internal loops with different patterns of bps (Figures 3(b) and 7(b)).⁹¹
- *Functional/topological equivalence*: it refers to motifs with different sequence signatures and hence different structures that can nonetheless serve the same function in the context of a larger structural context.⁹² These motifs are generally not superimposable. For example, the RA motif used in the formation of square-shaped tetrameric particles has been assembled using three different RNA motifs: a five-way tRNA junction, a three-way junction, and the two-helix bend found in the RA⁶⁵ (Figure 7(c)). Because all three motifs share the common feature of possessing a 90° angle between two RNA helices they can be considered interchangeable. However, this alone is not enough to ensure topological equivalence in all structural contexts. In other words, for two structurally distinct motifs to be considered topologically equivalent, it is necessary to verify that they are interchangeable within a particular structural context by also serving the same function (albeit even when function and structure are one in the same).

Principle of Scaffolding

The 'GA-minor syntax network' presented here emphasizes a fourth principle concerning the modularity of RNA—the principle of structural scaffolding. The principle of scaffolding involves the addition or substitution of modular components to create motifs having expanded sequence signatures. Structural expansion often occurs at the level of bulging nucleotides or single-stranded turn regions (see section on motifs of increased complexity and Figure 6). The GA-minor motif provides ideal anchoring points for the grafting or appending of additional structural components. While the principle of structural scaffolding is often coupled to the principle of topological/functional equivalence, it is not limited to motifs sharing any of the three primary principles of equivalence. Identical smaller components found within topologically nonequivalent motifs can also be interexchanged with identical larger components. For example, despite not being topologically equivalent, the RA turn (90°) and kink-turn (45°) can both be expanded into larger motifs by the addition of an extra helix because both contain the GA-minor type I-Tw motif, which can be exchanged with the larger A-minor 2h_stack motif (Figure 6).

Scaffolds provide the means for a RNA structure to grow and increase in complexity. In the case of bacteria, all 23S rRNAs have two RA motifs localized in domain H27-H30 (Figures 6 and 8). One of the two conserved RA motifs is used as a scaffold for structural expansion in archaea. Replacement of the GA-minor type I-Tw by the A-minor/G-major 2h_stack allows insertions of up to 30 nucleotides between nucleotide positions 7 and 8. This structural interchangeability reflects the similar modularity associated with otherwise distinct categories of structural motifs. In the context of the archaea 23S rRNA, the insertion of an extra helix within the RA turn H29-H30 increases the structural complexity from a four-way junction to a five-way junction (Figure 8). As expected, the kink-turn can be expanded in a similar fashion (Figure 6) by substituting its GA-minor type I-Tw by an A-minor 2h_stack module. Furthermore, the expansion of a GA-minor 2h_stack motif into a RA-2h_stack motif, as demonstrated in the P2.1-P3-P8 junction in class C and D group I introns, shows that the addition of structural elements can enhance the stability of a motif involved in scaffolding.³⁹ This may be due to the fact that certain expansions require only slight variations in the sequence signature of the motif associated with the expansion.

RNA STRUCTURAL EVOLUTION

RNA Motifs as Molecular Signatures for RNA Structural Evolution

Given the variety of possible structural evolutionary pathways for the ribosomal RNA (for examples, see Refs 93–95), some pathways may be deemed more plausible than others based on a minimal set of criteria required to aid the expansion of an RNA structure. For example, among all the RNA motifs identified, it is apparent that some are particularly suited for structural expansions. Furthermore, the insertion of additional RNA segments should occur in such a way that minimizes structural interference and maintains the function of the pre-existing RNA domains acting as scaffoldings. Finally, as the case of the RA found in the group I intron suggests, accommodations made by the motif involved in scaffolding are tolerated if they are compensated for when placed within the new structural context.

The general considerations highlighted above can provide a framework for understanding how small RNA domains might grow and increase in structural complexity through evolutionary processes. For instance, careful analysis of the RA motifs involved in expansions can be used as molecular signatures for verifying or/and unraveling possible evolutionary pathways leading to the emergence of functional molecular machineries as complex as the ribosome. While the probability that a RA motif can be found within a random pool of RNAs is quite high, RA turns are only found twice in bacterial 23S rRNAs. Furthermore, because the two RA turns are adjacent to one another at the level of the 4W-

junction peripheral domain defined by helices H27-H30, they are likely to have originated during the course of ribosome evolution through an expansion mechanism involving a local duplication. It is possible that local duplications of the GA-minor module (H28) of an ancestral RA motif (H27-H28) might have given rise to a four-way junction domain formed by two RA motifs (H27-H28 and H29-H30) before the major divergence of bacteria, archaea, and eukaryotes (Figure 8(a)). While this resulting four-way RA junction is well preserved in all bacterial 23S rRNA, it has undergone additional structural changes in archaea. The four-way junction in archaea is expanded into a five-way junction through insertion of an additional helix within the H29-H30 RA motif. The extra helix (helix 30a) is inserted such that it does not disrupt the major tertiary contacts of the original RA motif (Figure 8(b)). In fact, the four-way junction in bacteria is superimposable with the five-way junction in archaea (Figure 8(c)). This suggests that the duplication of the RA motif happened before the major divergence between bacteria and archaea, where the bacterial 23S rRNA is structurally closer to the 'proto 23S' rRNA of the last universal common ancestor (LUCA)^{96–99} than the equivalent rRNAs of archaea. The H27-H30 domain from eukaryotes, especially helices H29-H30, is even more divergent from the bacterial prototype than the one from archaea but is clearly related to the archaea version. Knowledge of these structural expansions in the context of the RA motif itself suggests that the emergence of H29-H30 is likely due to a duplication of H27-28 and not the other way around. The RA motif, as detailed previously, is comprised of two distinct types of GA-minor interactions (type I-STw and type I-Tw). Our investigation of the GA-minor motif has shown that the type I-Tw provides a better platform for expansion than the type I-STw (all known examples to date and documented herein involve expansion through a GA-minor Type I-Tw). Furthermore, local insertion sites within the bacterial ribosome, as determined by genetic insertions of small RNA fragments within *Escherichia coli* rRNAs,¹⁰⁰ tend to coincide with where the expansion segments are found in eukaryotic rRNAs, implying that eukaryotic rRNAs might have evolved from a prokaryotic-like rRNAs undergoing genetic duplication, insertions, and selections.

The general theory that H29-H30 arose from the duplication of H27-28 is further supported by the fact that across all phylogenetic domains of life, the RA motif in H29-H30 is recognized by protein L15. On the other hand, the RA motif in H27-H28 is only recognized by protein L18e in archaea and eukaryotes (Figure 9). L15 and L18e are related homologous proteins over 80 amino acids.^{25,101} The mode of recognition of these RA motifs by two different proteins L15 and L18e is strikingly similar (Figure 9(b)).²⁵ For instance, superimposition of the C-terminal domains of L15 and L18e shows the striking similarity of the RA motif recognition domain of L15 and L18—suggesting that they have a common evolutionary origin (Figure 9(c)). This is corroborated by phylogenetic analysis and thus has been proposed that L15 and L18e are the products of a gene duplication event that occurred before the divergence of the three primary lineages.¹⁰¹ According to this scenario, the ancestral proto-rRNA of the large subunit probably contained two similar RA motifs, with L15 recognizing RA H29-H30. After the divergence of the three primary lineages, the domain was allowed to expand in archaea and eukaryotes. The RA H27-H28 in the ancestral large rRNA of LUCA might have been used as a secondary landing site for an ancestral form of L18e derived from L15. Archaea and eukaryotes might have retained this protein for stabilization purposes, whereas the bacteria did not rely on it because of the strong selection pressure for minimal genome size.

While the RA can act as a scaffold—as in the case of the four-way junction domain formed by two RA motifs (H27-H28 and H29-H30) detailed above—the presence of an RA motif does not always require it to have functioned as one. Previously existing motifs can also act as scaffolds to incorporate the RA motif. For example, it is apparent that expansion in the case of P2.1-P3-P8 found in the group I intron probably occurred through the 2h_stack

(Figure 6). The RA motif likely followed with the formation of the peripheral long-range interaction of P13 (IC1) or P16 (ID), as P3 appears to be part of an ancient core of group I introns. This could explain why the along-groove motif of the RA motif does not take advantage of a classic GC at the level of P3. Likewise, the RA found in the H3-H4-H18 junction of 16S rRNA is also likely to have been added to a pre-existing structural element, preceded by the formation of H3-H4, which is stabilized by a long-range UA WC *trans* bp. This hypothesis is in good agreement with the fact that the RA motif signature at the level of H3-H18 is not the most canonical or most stable when characterized in isolation.³⁹ In fact, H18 is thought to be a later addition in the structural and functional growth of the core of the 16S rRNA.

Structural Motifs and Evolutionary Convergence

While some RNA motifs might spread by local duplication (as in the case of the RA motifs of domains 27–30 of the 23S rRNA), many RNA motifs are building blocks that occur independently in unrelated natural or artificial RNA molecules. This phenomenon of evolutionary convergence is grounded in the ‘bottom-up’ biophysical–biochemical properties whereby particular RNA sequences are predisposed to adopt similar RNA conformations. For example, the RA is found in a broad array of naturally occurring RNAs including the 16S, 23S, and group I intron (see Table 2). As illustrated by our structural motif syntax networks, modular domains of more than 20 nucleotides can easily emerge independently through either artificial or/and natural selection. For instance, while the A-minor/G-major kinked_PK is found in the Doughnut PK_domain H5-H7 from 23S rRNA, it is also part of the ‘B12’ aptamer obtained by SELEX^{102,103} (Figure 6). This observation highlights the modularity of RNA and how certain motifs can be recreated and reused under different functional contexts. Over the last 2 decades, *in vitro* evolution techniques have isolated an ever-growing number of RNA functions, suggesting the immense functional potential of RNA. Our present structural investigation suggests that these functionalities can be achieved by combining a rather limited set of recurrent structural motifs. In other words, any RNA shape could arise from modular combination of a limited set of structural motifs.

RNA 3D STRUCTURE PREDICTION AND RATIONAL DESIGN

Although it is clear that the hierarchical structural nature of the ‘GA-minor syntax network’ represents a small portion of all the motifs identified by X-ray structure of RNA, the various principles derived from our studies are applicable to all RNA motifs and offer a comprehensive framework for the rational design, prediction of RNA 3D structures, and study of RNA structure in general. The ‘GA-minor syntax network’ deepens our understanding of the RNA folding syntax—the relationship existing between the information encoded at the level of an RNA sequence and the resulting 3D shape. By looking at RNA structural motifs as part of an extended network that relates the sequence space of RNA with conformational space, we have started to establish some of the rules and principles that control the folding and assembly of RNA into complex 3D structures. Therefore, sequence constraints and rules derived from this study can be working hypotheses for identifying putative RNA motifs, which in turn can be used for modeling prior to refinement or interpretation after the refinement of X-ray structures. Our present motif syntax has already proven to be valuable for identifying and predicting the tertiary structures of natural riboswitches⁶ and ribozymes.³⁹ For example, we have identified and experimentally corroborated the presence of a RA motif in class C and D group I introns.³⁹ On the basis of our structural analysis and molecular dynamics simulations, we proposed a detailed atomic 3D model for the P2.1-P3-P8 junction of the *Tetrahymena* group I intron that is fully compatible with the formation of a peripheral belt for stabilizing its catalytic core.³⁹ Furthermore, our previous prediction on part of the FMN riboswitch structure was subsequently corroborated by X-ray crystallographic studies.^{104,105} Despite recent progress

in crystallizing RNA molecules, it is unlikely to get atomic X-ray structures for all newly discovered RNA. We anticipate that our RNA syntax network will facilitate the development of bioinformatics tools for the prediction of the tertiary structure of natural RNA molecules.^{104,106,107}

The principles highlighted by the ‘GA-minor syntax network’ defined throughout this article can also be applied to the rational design of self-assembling RNA architectures and scaffolds for synthetic biology and nanomedicine applications.^{54,66} In fact, the RA motif was already proven to be a remarkable RNA building block for nanoconstruction.^{53,54,66} Using the ‘GA-minor syntax network’ as a guide for viewing the RA motif, it is possible to understand how the tectosquare could assemble in a programmable and directional fashion. The fact that the RA can be combined with the GA 2h_stack provides a rational explanation for the particular directionality that the tail–tail interactions of the tectosquares adopt and the resulting ability of tectosquares to form extensive two-dimensional RNA nanoarrays.^{53,54,66} Indeed, the 3 tails of two monomer units can assemble with one another to form a similar type of RA-2h_stack motif (Figure 10). It is worth mentioning that the RA-2h_stack motif was first rationally engineered for tectosquare assembly^{53,54} before being discovered as a natural motif in group IC1 and ID introns.³⁹

In many ways, the structural modularity and interchangeability of RNA motifs can be used to produce both divergent and convergent RNA structures. For example, although the GA minor is found in both the RA and kink-turn motifs, the two motifs can be used to create structures with very different bend angles. For example, while the RA motif was used to form RNA tectosquares,^{53,54,66} the kink-turn has been used to generate triangular-shaped nanoparticles.¹⁰⁸ Likewise, although the kink-turn and UA_handle motifs are not related through structural hierarchy, it is likely that the two, based on their apparent topological equivalence, could be substituted for one another within a variety of given contexts.⁶

CONCLUSION

The structural syntax network presented extends the already published RNA syntax networks associated with the UA_h motif and A-minor junction.^{6,13} We estimate that the resulting combined syntax networks represent about 70% of all the structural motifs identified in known X-ray structure of RNA. Like other recurrent motifs (e.g., Refs 3–6, 14, 21, 24, 26, 29, 42, 55, and 80, see Table 1 for an exhaustive list of additional references), the motifs characterized here are considered as ‘modules’ that can be combined to form motifs of higher complexity—in many ways illustrating a type of hierarchy akin to that which is found within the basic structure of a primitive language.^{6,109,110} The principles that govern the formation of complex RNA structures can be mapped onto the grammatical rules found in linguistics: syllables (small submotifs) combine to form words (motifs), which further combine to form sentences (higher order motifs or domains). Ultimately, it should be possible to use the complete RNA structural syntax network to design and synthesize any arbitrary shapes or assembly made of RNA. Most of our rational design and predictions thus far have been based on a case-by-case study of few RNA motifs. A long-term goal is therefore to use the full potential of RNA modularity and syntax to generate an RNA compiler for computational design of novel RNA structures in a fully automated fashion.

Acknowledgments

This work was funded by the National Institutes of Health (R01-GM079604) to LJ and the NSF (MCB-1158577) and the David and Lucile Packard Foundation to JS. LJ wishes to dedicate this article to saint Mary MacKillop, a great missionary teacher.

References

- Holbrook SR. RNA structure: the long and the short of it. *Curr Opin Struct Biol.* 2005; 15:302–308. [PubMed: 15963891]
- Holbrook SR. Structural principles from large RNAs. *Annu Rev Biophys.* 2008; 37:445–464. [PubMed: 18573090]
- Noller HF. RNA structure: reading the ribosome. *Science.* 2005; 309:1508–1514. [PubMed: 16141058]
- Leontis NB, Lescoute A, Westhof E. The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol.* 2006; 16:279–287. [PubMed: 16713707]
- Leontis NB, Westhof E. Analysis of RNA motifs. *Curr Opin Struct Biol.* 2003; 13:300–308. [PubMed: 12831880]
- Jaeger L, Verzemnieks EJ, Geary C. The UA_handle: a versatile submotif in stable RNA architectures. *Nucleic Acids Res.* 2009; 37:215–230. [PubMed: 19036788]
- Gesteland, RF.; Cech, TR.; Atkins, JF. *The RNA World.* 3. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2005.
- Tinoco I Jr, Bustamante C. How RNA folds. *J Mol Biol.* 1999; 293:271–281. [PubMed: 10550208]
- Woodson SA. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr Opin Chem Biol.* 2005; 9:104–109. [PubMed: 15811793]
- Laederach A, Chan JM, Schwartzman A, Willgoos E, Altman RB. Coplanar and coaxial orientations of RNA bases and helices. *RNA.* 2007; 13:643–650. [PubMed: 17339576]
- Hendrix DK, Brenner SE, Holbrook SR. RNA structural motifs: building blocks of a modular biomolecule. *Q Rev Biophys.* 2005; 38:221–243. [PubMed: 16817983]
- Gutell RR, Cannone JJ, Konings D, Gautheret D. Predicting U-turns in ribosomal RNA with comparative sequence analysis. *J Mol Biol.* 2000; 300:791–803. [PubMed: 10891269]
- Geary C, Chworos A, Jaeger L. Promoting RNA helical stacking via A-minor junctions. *Nucleic Acids Res.* 2011; 39:1066–1080. [PubMed: 20876687]
- Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci U S A.* 2001; 98:4899–4903. [PubMed: 11296253]
- Tamura M, Holbrook SR. Sequence and structural conservation in RNA ribose zippers. *J Mol Biol.* 2002; 320:455–474. [PubMed: 12096903]
- Lee JC, Cannone JJ, Gutell RR. The lonepair triloop: a new motif in RNA structure. *J Mol Biol.* 2003; 325:65–83. [PubMed: 12473452]
- Lisi V, Major F. A comparative analysis of the triloops in all high-resolution RNA structures reveals sequence structure relationships. *RNA.* 2007; 13:1537–1545. [PubMed: 17652406]
- Woese CR, Winker S, Gutell RR. Architecture of ribosomal RNA: constraints on the sequence of “tetra-loops”. *Proc Natl Acad Sci U S A.* 1990; 87:8467–8471. [PubMed: 2236056]
- Correll CC, Swinger K. Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 Å resolution. *RNA.* 2003; 9:355–363. [PubMed: 12592009]
- Hsiao C, Mohan S, Hershkovitz E, Tannenbaum A, Williams LD. Single nucleotide RNA choreography. *Nucleic Acids Res.* 2006; 34:1481–1491. [PubMed: 16531589]
- Correll CC, Beneken J, Plantinga MJ, Lubbers M, Chan YL. The common and the distinctive features of the bulged-G motif based on a 1.04 Å resolution RNA structure. *Nucleic Acids Res.* 2003; 31:6806–6818. [PubMed: 14627814]
- Leontis NB, Stombaugh J, Westhof E. Motif prediction in ribosomal RNAs Lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie.* 2002; 84:961–973. [PubMed: 12458088]
- Krasilnikov AS, Mondragon A. On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA.* 2003; 9:640–643. [PubMed: 12756321]
- Nagaswamy U, Fox GE. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA.* 2002; 8:1112–1119. [PubMed: 12358430]

25. Klein DJ, Moore PB, Steitz TA. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J Mol Biol.* 2004; 340:141–177. [PubMed: 15184028]
26. Razga F, Zacharias M, Reblova K, Koca J, Sponer J. RNA kink-turns as molecular elbows: hydration, cation binding, and large-scale dynamics. *Structure.* 2006; 14:825–835. [PubMed: 16698544]
27. Szep S, Wang J, Moore PB. The crystal structure of a 26-nucleotide RNA containing a hook-turn. *RNA.* 2003; 9:44–51. [PubMed: 12554875]
28. Wadley LM, Pyle AM. The identification of novel RNA structural motifs using COMPADRES: an automated approach to structural discovery. *Nucleic Acids Res.* 2004; 32:6650–6659. [PubMed: 15608296]
29. Steinberg SV, Boutorine YI. G-ribo: a new structural motif in ribosomal RNA. *RNA.* 2007; 13:549–554. [PubMed: 17283211]
30. Lescoute A, Westhof E. Topology of three-way junctions in folded RNAs. *RNA.* 2006; 12:83–93. [PubMed: 16373494]
31. Laing C, Schlick T. Analysis of four-way junctions in RNA structures. *J Mol Biol.* 2009; 390:547–559. [PubMed: 19445952]
32. Ennifar E, Walter P, Ehresmann B, Ehresmann C, Dumas P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol.* 2001; 8:1064–1068. [PubMed: 11702070]
33. Brunel C, Marquet R, Romby P, Ehresmann C. RNA loop-loop interactions as dynamic functional motifs. *Biochimie.* 2002; 84:925–944. [PubMed: 12458085]
34. Gagnon MG, Steinberg SV. GU receptors of double helices mediate tRNA movement in the ribosome. *RNA.* 2002; 8:873–877. [PubMed: 12166642]
35. Mokdad A, Krasovska MV, Sponer J, Leontis NB. Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.* 2006; 34:1326–1341. [PubMed: 16522645]
36. Costa M, Michel F. Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: comparison with in vivo evolution. *EMBO J.* 1997; 16:3289–3302. [PubMed: 9214644]
37. Geary C, Baudrey S, Jaeger L. Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. *Nucleic Acids Res.* 2008; 36:1138–1152. [PubMed: 18158305]
38. Steinberg SV, Boutorine YI. G-ribo motif favors the formation of pseudoknots in ribosomal RNA. *RNA.* 2007; 13:1036–1042. [PubMed: 17507660]
39. Grabow WW, Zhuang Z, Swank ZN, Shea JE, Jaeger L. The right angle (RA) motif: a prevalent ribosomal RNA structural pattern found in group I introns. *J Mol Biol.* 2012; 424:54–67. [PubMed: 22999957]
40. Zirbel CL, Sponer JE, Sponer J, Stombaugh J, Leontis NB. Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res.* 2009; 37:4898–4918. [PubMed: 19528080]
41. Doherty EA, Batey RT, Masquida B, Doudna JA. A universal mode of helix packing in RNA. *Nat Struct Biol.* 2001; 8:339–343. [PubMed: 11276255]
42. Gagnon MG, Steinberg SV. The adenosine wedge: a new structural motif in ribosomal RNA. *RNA.* 2010; 16:375–381. [PubMed: 20038632]
43. Jucker FM, Pardi A. GNRA tetraloops make a U-turn. *RNA.* 1995; 1:219–222. [PubMed: 7585251]
44. Elgavish T, Cannone JJ, Lee JC, Harvey SC, Gutell RR. AA.AG@helix ends: A:A and A:G base-pairs at the ends of 16 S and 23 S rRNA helices. *J Mol Biol.* 2001; 310:735–753. [PubMed: 11453684]
45. Lehnert V, Jaeger L, Michel F, Westhof E. New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the *Tetrahymena thermophila* ribozyme. *Chem Biol.* 1996; 3:993–1009. [PubMed: 9000010]
46. Leontis NB, Westhof E. The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA.* 1998; 4:1134–1153. [PubMed: 9740131]

47. Lee JC, Gutell RR, Russell R. The UAA/GAN internal loop motif: a new RNA structural element that forms a cross-strand AAA stack and long-range tertiary interactions. *J Mol Biol.* 2006; 360:978–988. [PubMed: 16828489]
48. Costa M, Michel F. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* 1995; 14:1276–1285. [PubMed: 7720718]
49. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Szewczak AA, Kundrot CE, Cech TR, Doudna JA. RNA tertiary structure mediated by adenosine platforms. *Science.* 1996; 273:1696–1699. [PubMed: 8781229]
50. Leontis NB, Westhof E. A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J Mol Biol.* 1998; 283:571–583. [PubMed: 9784367]
51. St-Onge K, Thibault P, Hamel S, Major F. Modeling RNA tertiary structure motifs by graph-grammars. *Nucleic Acids Res.* 2007; 35:1726–1736. [PubMed: 17317683]
52. Robertson MP, Igel H, Baertsch R, Haussler D, Ares M Jr, Scott WG. The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol.* 2005; 3:e5. [PubMed: 15630477]
53. Chworos A, Severcan I, Koyfman AY, Weinkam P, Oroudjev E, Hansma HG, Jaeger L. Building programmable jigsaw puzzles with RNA. *Science.* 2004; 306:2068–2072. [PubMed: 15604402]
54. Severcan I, Geary C, Verzemnieks E, Chworos A, Jaeger L. Square-shaped RNA particles from different RNA folds. *Nano Lett.* 2009; 9:1270–1277. [PubMed: 19239258]
55. Klein DJ, Schmeing TM, Moore PB, Steitz TA. The kink-turn: a new RNA secondary structure motif. *EMBO J.* 2001; 20:4214–4221. [PubMed: 11483524]
56. Razga F, Spackova N, Reblova K, Koca J, Leontis NB, Sponer J. Ribosomal RNA kink-turn motif—a flexible molecular hinge. *J Biomol Struct Dyn.* 2004; 22:183–194. [PubMed: 15317479]
57. Michel F, Ellington AD, Couture S, Szostak JW. Phylogenetic and genetic evidence for base-triples in the catalytic domain of group I introns. *Nature.* 1990; 347:578–580. [PubMed: 2215683]
58. Vicens Q, Cech TR. Atomic level architecture of group I introns revealed. *Trends Biochem Sci.* 2006; 31:41–51. [PubMed: 16356725]
59. Michel F, Westhof E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol.* 1990; 216:585–610. [PubMed: 2258934]
60. Michel F, Hanna M, Green R, Bartel DP, Szostak JW. The guanosine binding site of the *Tetrahymena* ribozyme. *Nature.* 1989; 342:391–395. [PubMed: 2685606]
61. Nowakowski J, Tinoco I Jr. Conformation of an RNA molecule that models the P4/P6 junction for group I introns. *Biochemistry.* 1996; 35:2577–2585. [PubMed: 8611561]
62. Gilbert SD, Rambo RP, Van Tyne D, Batey RT. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat Struct Mol Biol.* 2008; 15:177–182. [PubMed: 18204466]
63. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science.* 2000; 289:905–920. [PubMed: 10937989]
64. Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonnrhein C, Hartsch T, Ramakrishnan V. Structure of the 30S ribosomal subunit. *Nature.* 2000; 407:327–339. [PubMed: 11014182]
65. Severcan, I.; Geary, C.; Jaeger, L.; Bindewald, E.; Kasprzak, W.; Shapiro, BA. Computational and experimental RNA nanoparticle design. In: Alterovitz, G.; Ramoni, M.; Mary Benson, R., editors. *Automation in Genomics and Proteomics: An Engineering Case Based Approach*. New York: John Wiley & Sons; 2009. p. 193–220.
66. Jaeger L, Chworos A. The architectonics of programmable RNA and DNA nanostructures. *Curr Opin Struct Biol.* 2006; 16:531–543. [PubMed: 16843653]
67. Chworos, A.; Jaeger, L. Nucleic Acid Foldamers: Design, Engineering and Selection of Programmable Biomaterials with Recognition, Catalytic and Self-assembly Properties. In: Hecht, S.; Huc, I., editors. *Foldamers: Structure, Properties, and Applications*. Weinheim: Wiley-VCH; 2007. p. 291–329.
68. Jaeger L, Leontis NB. Tecto-RNA: one-dimensional self-assembly through tertiary interactions. *Angew Chem Int Ed Engl.* 2000; 39:2521–2524. [PubMed: 10941124]

69. Nasalean L, Baudrey S, Leontis NB, Jaeger L. Controlling RNA self-assembly to form filaments. *Nucleic Acids Res.* 2006; 34:1381–1392. [PubMed: 16522648]
70. Grabow WW, Zakrevsky P, Afonin KA, Chworos A, Shapiro BA, Jaeger L. Self-assembling RNA nanorings based on RNAI/II inverse kissing complexes. *Nano Lett.* 2011; 11:878–887. [PubMed: 21229999]
71. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. *RNA.* 2001; 7:499–512. [PubMed: 11345429]
72. Lescoute A, Leontis NB, Massire C, Westhof E. Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res.* 2005; 33:2395–2409. [PubMed: 15860776]
73. Gutell RR, Lee JC, Cannone JJ. The accuracy of ribo-somal RNA comparative structure models. *Curr Opin Struct Biol.* 2002; 12:301–310. [PubMed: 12127448]
74. Saenger, W. *Principles of Nucleic Acid Structure.* New York: Springer-Verlag; 1984.
75. Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J Math Biol.* 2008; 56:215–252. [PubMed: 17694311]
76. Stombaugh J, Zirbel CL, Westhof E, Leontis NB. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.* 2009; 37:2294–2312. [PubMed: 19240142]
77. Jaeger L, Westhof E, Leontis NB. TectoRNA: modular assembly units for the construction of RNA nano-objects. *Nucleic Acids Res.* 2001; 29:455–463. [PubMed: 11139616]
78. Zhuang Z, Jaeger L, Shea JE. Probing the structural hierarchy and energy landscape of an RNA T-loop hairpin. *Nucleic Acids Res.* 2007; 35:6995–7002. [PubMed: 17940098]
79. Razga F, Koca J, Sponer J, Leontis NB. Hinge-like motions in RNA kink-turns: the role of the second A-minor motif and nominally unpaired bases. *Biophys J.* 2005; 88:3466–3485. [PubMed: 15722438]
80. Blouin S, Lafontaine DA. A loop loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA.* 2007; 13:1256–1267. [PubMed: 17585050]
81. Tiedge H. K-turn motifs in spatial RNA coding. *RNA Biol.* 2006; 3:133–139. [PubMed: 17172877]
82. Antonioli AH, Cochrane JC, Lipchick SV, Strobel SA. Plasticity of the RNA kink turn structural motif. *RNA.* 2010; 16:762–768. [PubMed: 20145044]
83. Schroeder KT, McPhee SA, Ouellet J, Lilley DM. A structural database for k-turn motifs in RNA. *RNA.* 2010; 16:1463–1468. [PubMed: 20562215]
84. Li H, Dalal S, Kohler J, Vilardell J, White SA. Characterization of the pre-mRNA binding site for yeast ribosomal protein L32: the importance of a purine-rich internal loop. *J Mol Biol.* 1995; 250:447–459. [PubMed: 7616567]
85. Mao H, White SA, Williamson JR. A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nat Struct Biol.* 1999; 6:1139–1147. [PubMed: 10581556]
86. Chao JA, Williamson JR. Joint X-ray and NMR refinement of the yeast L30e-mRNA complex. *Structure.* 2004; 12:1165–1176. [PubMed: 15242593]
87. Chastain M, Tinoco I Jr. A base-triple structural domain in RNA. *Biochemistry.* 1992; 31:12733–12741. [PubMed: 1463745]
88. Chastain M, Tinoco I Jr. Nucleoside triples from the group I intron. *Biochemistry.* 1993; 32:14220–14228. [PubMed: 8260508]
89. Westhof E, Masquida B, Jaeger L. RNA tectonics: towards RNA design. *Fold Des.* 1996; 1:R78–R88. [PubMed: 9079386]
90. Michiels PJ, Versleijen AA, Verlaan PW, Pleij CW, Hilbers CW, Heus HA. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J Mol Biol.* 2001; 310:1109–1123. [PubMed: 11501999]
91. Meyer M, Westhof E, Masquida B. A structural module in RNase P expands the variety of RNA kinks. *RNA Biol.* 2012; 9:254–260. [PubMed: 22336704]

92. Auletta G, Ellis GF, Jaeger L. Top-down causation by information control: from a philosophical problem to a scientific research programme. *J R Soc Interface*. 2008; 5:1159–1172. [PubMed: 18319208]
93. Agmon I. The dimeric proto-ribosome: structural details and possible implications on the origin of life. *Int J Mol Sci*. 2009; 10:2921–2934. [PubMed: 19742176]
94. Agmon I, Bashan A, Zarivach R, Yonath A. Symmetry at the active site of the ribosome: structural and functional implications. *Biol Chem*. 2005; 386:833–844. [PubMed: 16164408]
95. Bokov K, Steinberg SV. A hierarchical model for evolution of 23S ribosomal RNA. *Nature*. 2009; 457:977–980. [PubMed: 19225518]
96. Fitch WM, Upper K. The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. *Cold Spring Harb Symp Quant Biol*. 1987; 52:759–767. [PubMed: 3454288]
97. Kyrpides N, Overbeek R, Ouzounis C. Universal protein families and the functional content of the last universal common ancestor. *J Mol Evol*. 1999; 49:413–423. [PubMed: 10485999]
98. Lazcano A, Forterre P. The molecular search for the last common ancestor. *J Mol Evol*. 1999; 49:411–412. [PubMed: 10485998]
99. Woese C. The universal ancestor. *Proc Natl Acad Sci U S A*. 1998; 95:6854–6859. [PubMed: 9618502]
100. Yokoyama T, Suzuki T. Ribosomal RNAs are tolerant toward genetic insertions: evolutionary origin of the expansion segments. *Nucleic Acids Res*. 2008; 36:3539–3551. [PubMed: 18456707]
101. Roberts E, Sethi A, Montoya J, Woese CR, Luthey-Schulten Z. Molecular signatures of ribosomal evolution. *Proc Natl Acad Sci U S A*. 2008; 105:13953–13958. [PubMed: 18768810]
102. Sussman D, Nix JC, Wilson C. The structural basis for molecular recognition by the vitamin B 12 RNA aptamer. *Nat Struct Biol*. 2000; 7:53–57. [PubMed: 10625428]
103. Lorsch JR, Szostak JW. In vitro selection of RNA aptamers specific for cyanocobalamin. *Biochemistry*. 1994; 33:973–982. [PubMed: 7508262]
104. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA*. 2009; 15:189–199. [PubMed: 19144906]
105. Serganov A, Huang L, Patel DJ. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature*. 2009; 458:233–237. [PubMed: 19169240]
106. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*. 2008; 452:51–55. [PubMed: 18322526]
107. Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, et al. RNA-puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*. 2012; 18:610–625. [PubMed: 22361291]
108. Ohno H, Kobayashi T, Kabata R, Endo K, Iwasa T, Yoshimura SH, Takeyasu K, Inoue T, Saito H. Synthetic RNA-protein complex shaped like an equilateral triangle. *Nat Nanotechnol*. 2011; 6:116–120. [PubMed: 21240283]
109. Dahl, V.; Jimenez-Lopez, MD.; Perriquet, O. Poetic RNA: Adapting RNA Design Methods to the Analysis of Poetry. In: Demazeau, Y.; Dignum, F.; Corchado, J.; Bajo, J.; Corchuelo, R.; Corchado, E.; Fernández-Riverola, F.; Julián, V.; Pawlewski, P.; Campbell, A., editors. *Trends in Practical Applications of Agents and Multiagent Systems*. Vol. 71. Berlin: Springer; 2010. p. 403–410.
110. Rivas E, Eddy SR. The language of RNA: a formal grammar that includes pseudoknots. *Bioinformatics*. 2000; 16:334–340. [PubMed: 10869031]

FIGURE 1.

The GA-minor motif. (a) Comparison of A:G SG:SG *trans* bps of type I-STw (magenta) and Tw (green) with the more classic A-minor interaction of type I-P planar (violet).³⁷ (Left panel) Diagram that illustrates the two types of angles (tilt and propeller twist) that define the orientation of the planar, aromatic rings of two base-paired nucleotides with respect to one another.⁷⁴ (Right panel) The G positions of the A:G-minor motifs listed in Table 2 were all superimposed with LSQMAN. The adenine position can adopt various propeller twist angles with respect to the G position. (b) 3D views of the four different categories of A-minor interaction of type I: supertwisted (STw), twisted (Tw), planar (P), and tilted (T). The typical tilt and propeller twist angles are also indicated for each type of interaction (they can vary by $\pm 15^\circ$). They all involve two H-bonds. A-minor type I-P and type I-T are most of the time associated with A-minor type II interactions.³⁷ The sugar pucker of the nucleotide in position A(1) is typically C3' endo. By contrast, A-minor type I-Tw and type I-STw are seldom or never associated with A-minor type II interactions. The sugar pucker of the nucleotide position A(1) is typically C2' endo. (c) 2D diagram and 3D view of the GA minor of type I-STw: the G:A SG:SG *trans* bp of type I^{14,37} is supertwisted (STw) so that atom N1 of A interacts with the 2' OH of G and atom N2 of G interacts with the 2' OH of A. (d) 2D diagram and 3D view of the GA minor of type I-Tw: the G:A SG:SG *trans* bp is twisted

(Tw). Atom N3 of A interacts with atom N2 of G. Atom N1 of A interacts with 2' OH of G. In addition, the adenine is stacked on the ribose of the nucleotide in 3' of the G. While A(1) is highly conserved, the guanine position g(4) is less conserved and thus depicted as a lowercase 'g' on the 2D diagrams. Open triangles indicate SG:SG *trans* bps and open rectangles indicate stacking interactions. Circled nts in red have C2' endo sugar pucker.

FIGURE 2.

Definition and structural characteristics of the RA motif (or RA turn). (a) Nomenclature and generic sequence signature based on the structural analysis of RA motifs from known X-ray structures.³⁹ Tertiary interactions and noncanonical base pairs (bps) are indicated on the 2D diagram where open triangles represent SG:SG *trans* bp. SG:SG *trans* bps can sometimes be of type 1 twisted (indicated by T) or supertwisted (indicated by ST). SG:SG *trans* interactions in the along-groove motif can be either symmetrical (indicated by =) or quasi-symmetrical (indicated by ~). Circled nts in red have C2' endo sugar pucker. Capital letters indicate that the nucleotide position is conserved in more than 90% of the cases; small letters indicate that the nucleotide position is conserved in more than 75% of the cases. Open rectangles indicate stacking interactions; N, any nucleotide (A, U, C, or G); R or r, purine; Y or y, pyrimidine; 5' and the arrow symbol indicate 5' and 3' ends, respectively. Additional nomenclature is defined in the legends of Figures 3 and 5. The regions colored in blue and rose highlight the 'GA-minor' and 'along-groove' components of the RA motif. The region in violet corresponds to the overlap of these two motifs. (b) Topological characteristics of the RA motif. The two adjacent helices H5' and H3' are oriented by 90° similarly to the corners of a log cabin. Nucleotide at position 13, at the 3' end of the motif, is in perfect helical continuity with H3', allowing an additional helix to be stacked in continuity of this helix. (c) Three-dimensional stereo image of the RA motif using the same color code as in (a) (from Ec_23S_RA.2). Note the quasi-symmetrical arrangement of helices H5' and H3' and the perfect helical continuity existing between N13 and H3'. (d) Stereo image of the G:A SG:SG *trans* bps formed between A1 and R6 (type Stw) and between A7 and g12 (type Tw). Note the stacking interactions between the bases of A1, A7 and ribose of N13 (from Tt_23S_RA.1). (e and f) The along-groove packing interactions with associated H-bonds networks (from Hm_23S_RA.1): (e) symmetrical SG:SG bp contact between G2:C5 WC bp and G8:C11 WC bp (highlighted in yellow) and (f) quasi-symmetrical SG:SG bp contact (highlighted in yellow) between Y3:R4 WC bp (G:U shown) and Y9:R10 WC bp. While R10 and R4 are often Gs, the position Y9 or Y4 can be either a U or a C. Typically, when Y3 is a C, Y4 is a U, or vice versa.

FIGURE 3.

GA-minor submotif as building block for larger recurrent RNA motifs. Two-dimensional sequence signatures, topological cartoons, and three-dimensional stereo views of (a) RA motif (or RA turn), (b) kink-turns, and (c) GA-minor 2h_stack all containing the GA-minor type I-Tw submotif (highlighted in blue). The most significant structural characteristics of the motifs are indicated on their respective sequence signatures according to the annotation of Leontis and coworkers.^{6,77} For bps symbols: Watson–Crick, Hoogsteen, and shallow groove edges are indicated by circle, square, and triangle, respectively. For example, the HG:SG *trans* bp is symbolized by an open square associated to an open triangle. All SG:SG *trans* bp interactions between nts are represented by open triangles. Capital letters indicate that the nucleotide position is conserved in more than 90% of the cases; small letters indicate that the nucleotide position is conserved in more than 75% of the cases. N_m , a sequence of n nucleotides; 5' and the arrow symbol indicate 5' and 3' ends, respectively. Note that the A-minor 2h_stack is the former A-minor triple motif mentioned in Ref 6.

FIGURE 4.

Variants of the combined GA-minor and 2h_stack motifs: (a) local extension of the GA-minor motif displayed in Figure 2 with GA-minor bulge motifs. Three-dimensional stereo views of (b) a GA-minor bulge motif [A50-G52-U359-A360 from Tt_16S rRNA (PDB_ID: 1J5E)] and (c) a GA-minor bulge 2h_stack motif [A58-C60-G87-G88 from GlmS ribozyme (PDB_ID: 2GCV)]. For annotation, see legends of Figures 2 and 3.

FIGURE 5. The hierarchical syntax network of RNA structural motifs related to the GA-minor submotif. Motifs are minimally sized recurrent set of nucleotides with conserved conformation. They are generally organized from the left to the right according to their increase in structural complexity. All motifs that comprise the GA-minor motif as a building block are circled in blue. The most significant structural characteristics of the motifs are indicated on their respective sequence signatures according to the annotation of Leontis and coworkers.^{6,77} Visual of the three-dimensional structures of some of these motifs can be found in the figures indicated in blue below the motif name. Points of connection to the previous ‘UA_h motif syntax network’⁶ are indicated by red stars. For bps and other symbols, see legend in inset: WC, Watson–Crick edge (circle); HG, Hoogsteen edge (square); SG, shallow groove

edge (triangle). For example, the HG:SG *trans* bp is symbolized by an open square associated to an open triangle. The SG:SG *cis* bp is represented by a plain triangle. N_n , a sequence of n nucleotides; H and P stand for pairings; 5' and the arrow symbol indicate 5' and 3' ends, respectively. See also legends of Figures 2 and 3. Note that the A-minor 2h_stack is the former A-minor triple motif mentioned in Ref 6.

FIGURE 6. The generation of motifs of increased complexity from the expansion of the GA minor. The RA turn, kink-turn, and GA-minor 2h-stack motifs show their ability to accommodate additional motifs to increase structural complexity as exemplified in by the hydrogen-bonding patterns found in several naturally occurring RNAs. While the kink-turn can have insertions of additional sequence and helical elements within some of its modular components,⁷² kink-turn motifs expanded through GA-minor 2h_stack have not yet been identified. However, it is anticipated that a helix could be inserted within a kink-turn and positioned in perfect helical continuity of H3 like in the RA motif [see for example the remarkable structure of the kink-turn in domain L1 of the Tt_23S rRNA (L1:H76-78 from Tt_23S_rRNA [PDB_ID:1MZP]) that would allow this type of structural arrangement]. For bps and other symbols, see the legend in inset.

FIGURE 7.

RNA modularity and the three principles of equivalence. (a) The principle of isosteric equivalence exemplified by the RA motif constructed from its elemental submotifs. Watson–Crick (WC) base pairing is the basis for isosteric equivalence. The closing base pair (bp) on the GA minor is superimposable to the WC bps of the along-groove motif (overlaid with their electron densities in purple). The secondary structure and the stereo view of the RA motif are shown on the right. (b) The principle of structural equivalence as exemplified by the kink-turn motif. The kink-turn can take advantage of two different types of internal loops with different patterns of bps [GAA:gga (top) and loop E (bottom)]. The two types of kink-turns (GA-shared darker shade and loop E lighter) are shown overlaid in stereo (right). (c) The principle of topological/functional equivalence exemplified by three different RNA tectosquares. The individual motifs, the RA, the 3WJ, and the tRNA motif, while exhibiting

completely different secondary structures and hydrogen-bonding patterns provide similar 90° bends that form the corner of the tectosquares.

FIGURE 8.

The RA motif as signature for ribosomal evolution: a hypothetical scenario for the structural evolution of the H27-H30 peripheral domain of the large rRNA of the large ribosomal subunit. (a) Before the major divergence of bacteria, archaea, and eukaryotes, local duplications of one of the GA-minor elements H28 of an ancestral RA motif (H27-H28) might have given rise to a four-way junction domain formed by two RA motifs (H27-H28 and H29-H30). (b) While the ancestral four-way RA junction is preserved in bacteria, it has undergone several additional structural changes in archaea, with the four-way junction serving as scaffolding for further structural expansion. Insertion of additional sequences between positions 7 and 8 within the H29-H30 RA in archaea might have transformed the

four-way junction into a four-way junction. (c) Superposition of domains H27-30 from bacteria [Ec_23S rRNA (PDB_ID: 2AW4)] and archaea [Hm_23S rRNA (PDB_ID: 1JJ2)].

FIGURE 9.

RA motifs and proteins. (a) In the 23S ribosomal RNA of *E. coli* (i), the interhelical G-A connector of the H29-H30 RA turn (magentas) is recognized by the ribosome-binding protein L15 (green) (ii). (iii) Recognition by L15 involves the N-H group from the peptide backbone belonging to Thr128, Lys129, and Gly130, the positively charged side chain of Lys109 as well as the negatively charged side chain of Glu76. (b) In the 23S ribosomal RNA of *Haloarcula marismortui* (i), the interhelical A-A connector of the H29-H30 RA turn (magentas) is recognized by the ribosome-binding protein L15 (green), whereas the H27-H28 RA turn (ruby) is recognized by the ribosome-binding protein L18e (cyan) (ii). (iii) Recognition by L18e involves the N-H group from peptide backbone belonging to Ser82, Gly83, and Thr84, the positively charged side chain of Lys63 and the negatively charged side chain of Glu42. The mode of recognition of these RA motifs by two different proteins L15 and L18e is strikingly similar.²⁵ (c) (i) Superimposition of the C-terminal domains of L15 and L18e. (ii and iii) Closer views. The striking similarity of the RA motif recognition domain of L15 and L18e suggests that they have a common evolutionary origin. This is corroborated by phylogenetic analysis.¹⁰¹

FIGURE 10.

Structural and functional equivalency of the RA-2h_stack motif from the *Tetrahymena* group IC1 intron and the tail–tail interaction used in RNA tectosquares. The 2D (top) and 3D (bottom) diagrams of the RA-2h_stack motif, previously reported in RNA tectosquares⁵⁴ and later identified in the P2.1-P3-P8 junction of group IC1 and ID introns,³⁹ demonstrate the inherent modularity associated with independently generated structures. On the left is shown a revised 3D model of the *Tetrahymena* group I intron³⁹ with the P2.1-P3-P8 junction corresponding to the RA-2h_stack (in blue). The directionality of the tectosquare assemblies is dictated by the pairing that takes place between the two exiting 3' tails (seen on the right). The RA motif residing in individual tectosquare monomers assembling through their respective 3' tails results in the formation of two RA-2h_stack motifs. The GA-minor motif present in the RA motif, in both cases, allows the 3' tail to exit the RA in such a way that promotes the exiting strands ability to stack nascent helices.

TABLE 1

List of RNA Structural Motifs Mentioned in This Article and in Refs 6 and 39

Complexity	Name	Min. Size (nt)	Figure (This Article)	References
Submotifs				
<i>3° nt contact</i>				
	G-clamp	2		6, 40
	Ribose zipper	2		15
	G-ribose	2		6, 29
	Intercalating interaction	2–3		6, 11
<i>3° bp contact</i>				
	A-minor (type I, type II)	2–3		14, 41
	A-minor (type IT, type IP)	2–3	1, 5	This Article, 6, 37
	A-minor (type I.Tw, type I.STw)	2–3	1, 5	This Article, 42
<i>ss conformer</i>				
	U turn	3		3, 12, 43
<i>2° bps component</i>				
	G/RA	3	5	6, 44
	Dinucleotide platform	4		37, 45
	GU/GC	4	5	35
	GA/UA	4	5	13
	XG/AX	4	5	44
Motifs				
<i>Small motifs at helix ends</i>				
	GA-minor (type I.STw, type I.Tw)	4–5	1–5	This Article, 42
	XG/RAX	5	5	44
	UA_h	5–7	5	6
	GH/SG_trans_bp bulge	5–8		13
	GA-minor bulge	6	4	This Article
<i>Terminal loops</i>				
	Lonpair triloop	5		16, 17
	GNRA	4–6		18, 19, 37
	UNCG	4–6		18
	T-loop	5–8		6, 23, 24
<i>Internal loops</i>				
	Loop E	6–8	3, 5, 7	22, 46
	GAA:ggA loop	6–8	3, 5, 7	22
	UGA/gAN	10–11		6, 47
	UAA/GAN	10–11		6, 47
	11-nt receptor	11		37, 48, 49
	Sarcin loop	11–13		21, 50, 51
	UA_h A-minor twist interaction	13–14		6
	UAA/GAN A-minor interaction	14–15		6, 47
<i>Turns</i>				

Complexity	Name	Min. Size (nt)	Figure (This Article)	References
	Hook-turn	8	5	27, 28
	2xG-ribose turn (S2m turn)	12		52
	UA_h_turn	12	5	6, 29
	Right angle (RA turn)	13	2(a), 3(a), 5, 6, 7(a), 8	This Article, 39, 53, 54
	Kink_turn	13	3(b), 5, 6, 7(b)	This Article, 6, 55, 56
<i>Stacks</i>				
	2-Helix stack (2h_stack)	4–8	4, 5, 6	This Article, 13
	GA-minor 2h_stack	7–9	4, 5, 6	This Article, 39
	GA-minor bulge 2h_stack	8–10	4	This Article
	Major 2h_stack	10		57, 58
	GA/A-minor 2h_stack	11	5, 6	This Article
	A-minor/G-major 2h_stack	11	5, 6	This Article
	G-major 2h_stack (G-binding site)	11–13		58–60
	A-minor 2h_stack (t1 and t0)	11–13	5, 6	This Article, 6, 61
	GA-minor/major 2h_stack	13	5, 6	This Article
	A-minor/major 2h_stack or 'P4-P6 junction'	17		57, 58
<i>Junctions</i>				
	UA_h_3WJ	13–15		6, 30, 54
	A-minor junction	~16		6, 13
	GA/A-minor junction	~16		This Article, 13
	UA_h A-minor junction	16		6
<i>3° interactions</i>				
	A-minor (type I/IIT, type I/IIP)	6	1, 5	This Article, 37
	A-minor twist	6–8		62
	Along-groove packing	8	2, 3	This Article, 34, 35
	GAAA/helix	10		37
	GAAA/11-nt interaction	17		37, 48, 49
	Nested double T-loop	20		6
Domains				
<i>Stack + turn</i>				
	RA-2h_stack (P2.1-P3-P8 junction)	15	6	This Article, 39, 45
	Expanded Kink-turn (3WJ)	~20	5	This Article, 13
	Expanded RA turn	22	6, 8	This Article, 39
<i>PK</i>				
	Kinked_PK	19–26	6	This Article, 6
	A-minor/G-major kinked_PK	24–26	6	This Article
	T-loop_PK	40–44		6
	Alpha_PK	42–46		6, 38
	Doughnut_PK domain	48–56		6, 38

The various structural motifs and modules mentioned in this table are listed according to their increasing structural complexity. References correspond to articles that have described earlier some of these structural motifs or that have described RNA structures containing structural motifs without explicitly mentioning them as being structural motifs. Figures corresponding to each motif are indicated. When a motif's figure is not available in this article, see the corresponding reference. Motif names are those used in this article. Most of these motifs were initially identified

within the original 50S ribosomal subunit of *Haloarcula marismortui*⁶³ and the 30S ribosomal subunit of *Thermus thermophilus*.⁶⁴ Submotifs typically define small set of conserved nucleotides that usually adopt well-defined three-dimensional conformations but are unlikely to be found as stand-alone motifs as they are usually part of more complex RNA motifs. Note that for small motifs of size greater than 4 nt but smaller than 8 nt, the distinction between motif and submotif might not always be clear cut because some of them require a particular structural context to form. This is the case of small motifs at helix ends, small bulges, the UA_h motif, and tertiary (3°) interactions such as A-minor (type I/II), A-minor twist, and along-groove motifs. In this table, these motifs have been listed as full fledge motifs but might sometimes be better described as submotifs. Tertiary contacts between 2 and 3 nucleotides are indicated as 3° nt contacts (when they do not involve more than one base position). Small recurrent conformers that correspond to a single strand (ss) are indicated as ss conformers. Submotifs that correspond to one or two set of base pairs (classic or noncanonical) identified within RNA secondary (2°) structures are indicated as 2° bps components.

TABLE 2

List of GA-Minor Submotifs Identified in X-Ray Atomic Structures of RNAs

Type	RNA Molecule	PDB ID	Location A(5')-X1:X2-R(3')	Modularity	Twist (°)	Figures
STw	Ec_16S_rRNA	2AVY	A499-G500-C545-A546	RA		
			A1067-G1068-C1107-G1108	—		
	Tl_16S_rRNA	1J5E	A499-G500-C545-G546	RA		
			A1067-G1068-C1107-G1108	—		
	Ec_23S_rRNA	2AW4	A627-G628-C635-G636	RA		
			A655-G656-C601-A602	RA		
	Tl_23S_rRNA	2J01	A627-G628-C635-G636	RA		1(a)
			A614C-G615-C612-G613	RA		
	Hm_23S_rRNA	1JJ2	A746-G747-C659-A660	RA		6(b)
			A688-G689-C696-G697	RA expanded		3, 9
Tw	Ec_16S_rRNA	2AVY	A547-G548-C36-U37	RA		
			A60-G61-C106-G107	RA_open		
	Ec_23S_rRNA	2AW4	A197-G198-U219-G220	GA/A-minor	48	3
			A246-G247-C277-G278	2h_stack		
	Tl_16S_rRNA	1J5E	A687-G688-C699-G700	Kink-turn Kink-turn		2(b), 3
			A84-G85-C97-G98	Kink-turn		
	Ec_23S_rRNA	1J5E	A603-G604-C624-G625	RA		
			A637-G638-C650-G651	RA		6(a)
	Tl_16S_rRNA	1J5E	A547-G548-C36-U37	RA		
			A60-G61-C106-G107	RA_open		
Tw	Ec_16S_rRNA	2AVY	A197-G198-C219-G220	GA/A-minor	48	
			A246-G247-C277-G278	2h_stack		
	Ec_23S_rRNA	2AW4	A687-G688-C699-G700	Kink-turn		
			A1101-A1102-U1073-G1074	Kink-turn GA-minor/major 2h_stack	51	3
	Tl_23S_rRNA	2J01	A84-G85-C97-G98	Kink-turn		
			A637-G638-C650-G651	RA		
	Ec_23S_rRNA	2J01	A1237-G1238-C1208-G1209	Kink-turn		
	Tl_23S_rRNA	2J01				

Type	RNA Molecule	PDB ID	Location A(5')-X1:X2-R(3')	Modularity	Twist (°)	Figures
Hm_23S_rRNA		1JJ2	A1275-A1276-U1294-C1295	GA/A-minor 2h_stack	45	2(b), 3
			A80-G81-C93-G94	Kink-turn		
			A660-G661-C685-A686	RA		
			A939-G940-C1026-G1027	Kink-turn		
			A1381-G1382-C1400-G1401	GA-minor		
L1:H76-78 23S_rRNA		1M2P	A2681-C2682-G2712-G2713	2h_stack GA-minor 2h_stack	40	3
Twort gpl intron		1YOQ	A22-G23-C32-G33	Kink-turn	44	
			A148-G149-C230-G231	Loop-loop interaction		
			A44-G45-C28-A29	Kink-turn		
			A12-G13-C22-G23	Kink-turn		
			A20-G21-C31-G32	Kink-turn		
			A15-G16-C8-G9	GA-minor 2h_stack		
U4 snRNA		1E7K, 2OZB				
L1 mRNA		1ZHO, 2HW8				
SAM 1 riboswitch		2GIS				
Bulge-helix-bulge substrate for RNA splicing endonuclease		2GIW				

Query of the GA-minor motif in existing .pdb files using FR3D had three basic requirements: (1) a reference structure (including the conserved set of nucleotides from a given .pdb file); (2) a cutoff discrepancy value based on the similarity between two structures⁷⁵; and (3) a list of .pdb files for FR3D to search. The same list of nonredundant high-resolution crystal structures used for our right angle search previously reported in Ref 39, which is an updated version by Stombaugh et al.,⁷⁶ was used for our search of the GA-minor motif. A rather generous cutoff of 0.6—which often lead to false positives—was used in order to minimize the chance that a true candidate would be missed. The more similar the two structures are, the lower the discrepancy value. All structures with discrepancy values (when compared with the reference) less than the cutoff were identified as potential matches. Visual inspection was used to manually discard false positives. Using the above criteria, the GA-minor submotifs were classified either as STw or Tw, according to the nature of their A-minor type I interaction. They are named according to their nucleotide numbers for positions [A(5')-X1:X2-R(3')] within their respective X-ray structure. A total of 39 motifs have been identified with the program FR3D using a list of nonredundant X-ray structures from the Protein Data Bank as defined in reference. Type STw and Tw GA-minor submotifs are respectively found at 4 and 20 independent locations within nonhomologous molecules. Type STw GA minor is typically associated with a type Tw GA minor to form a RA motif. On the basis of our search, we have identified four different RA motifs, nine kink-turns, and six GA-minor 2h_stacks (comprising two GA/A-minor 2h_stacks and one GA-minor/major 2h_stack).