RNA seeing double: close-packing of helices in RNA tertiary structure

Scott A. Strobel and Jennifer A. Doudna

Structured RNA molecules play essential roles in RNA processing, chromosome maintenance and protein biosynthesis. RNA necessarily uses different strategies than proteins for folding and assembly of complex architectures. The RNA-folding problem is largely an issue of helical packing: how does RNA organize and pack short, double-helical segments to produce active sites and recognition motifs for proteins? Non-canonical base pairs, metal ions and 2'-hydroxyl groups are key elements in RNA higher-order structure formation.

RNA AND PROTEIN can both adopt complex globular structures and act as biocatalysts. Likewise, RNA and protein fold into structures that are characterized by secondary and tertiary structural elements. However, what is achieved by secondary structure formation for these two classes of macromolecules is fundamentally different. In proteins, the α-helix or β-sheet places the chemical groups of the amino acid sidechains on the surface of the secondary structure, optimally positioned for tertiary structural interaction. In contrast, the double-stranded secondary structure of RNA places the unique chemical groups of the nucleotides at the interior of the A-form duplex in an environment largely inaccessible for tertiary structure formation.

A continuous A-form RNA duplex presents a particularly challenging problem for helix packing. The major groove, which displays the diversity of the base functional groups, is deep and narrow. This makes contact with the base edges difficult. The minor groove is broad, shallow and more amenable to recognition, but the functional groups presented in the minor groove by the Watson-Crick base pairs lack the diversity necessary to identify them uniquely. Despite this apparent structural handicap, RNA can adopt precise conformations in which the A-form helices pack together to produce unique structures.

In this article, we review our current understanding of the molecular basis of RNA-duplex recognition within a globular RNA structure. While we draw pertinent examples from several categories of RNA, we will focus primarily on the Tetrahymena group I intron, an autocatalytic RNA that can excise itself from a primary transcript without the need for protein cofactors. We have chosen to focus on this RNA because it has been extensively characterized biochemically and structurally; it is of sufficient size (414 nucleotides) that there are many examples of helical packing within its structure; and the principles of RNA folding observed in this eukaryotic intron are likely to be common to any RNA.

Helical stacking

One strategy for packing RNA helices into a tertiary structure is helical stacking. Coaxial alignment of one RNA helix upon another was first observed in the three-dimensional structure of tRNA, and is a strategy employed by many RNAs. The secondary structure of tRNA consists of four short helices that radiate from a central origin in the shape of a clover leaf. In the three-dimensional structure, two pairs of helices coaxially stack. A perpendicular alignment of these two pairs of helices leads to the L-shape that characterizes tRNA tertiary structure.

Helical stacking is fundamental to the Tetrahymena group I intron and many of its coaxial elements have been identified (Fig. 1). Results from UV photo-crosslinking and chemical protection experiments suggest that helix P2 coaxially stacks upon P2.1 (Ref. 8). The P2-P2.1 helix aligns the adjoining P1 helix in the ribozyme active site. Coaxial stacking of P4 upon P6 was demonstrated biochemically by circular permutation of a 160-nucleotide RNA comprising helices P4 through P6 (Ref. 10). The recent crystal structure of the P4-P6 domain demonstrates that the helix is further extended by stacking of P5 on P4, and that a second pair of helices, P5a and P5b, are coaxially stacked in the P5abc extension (Fig. 1). The overall structure of the P4-P6 domain is characterized by a sharp 150° bend at the junction between the two sets of coaxially stacked helices, leading to extensive tertiary interactions between the two helical subdomains (see below).

While helical stacking is a central feature of RNA tertiary structure formation, it is an extension of, rather than a solution to, the question of how an A-form helix is accommodated in globular RNA structures. Stacking of short helices simply produces a longer helix that must be structurally organized.

2'-Hydroxyls as molecular handles

The 2'-hydroxyl is the signature chemical group that distinguishes RNA from DNA. Unlike the chemical groups of the nucleotide bases, which are located on the interior of the helix, the ubiquitous 2'-hydroxyl lies on the outer edge of the helix, extending toward the minor groove.

Coaxial stacking of helices P4 and P6 is highlighted by the 2'-O's red color. The dashed lines indicate the direction of helical stacking in these examples. For clarity, we have only shown one 2'-O on each helix.

**Figure 1**

Schematic diagram of the Tetrahymena group I intron showing the 5'-splice site and the helical regions discussed in the text. The 3'-splice site is not shown. Copyright © 1997. Elsevier Science Ltd. All rights reserved 0968-0004/97/$17.00 PII: S0968-0004(97)01056-6
of the minor groove, where they can serve as either hydrogen-bond donors or acceptors in tertiary interactions.

Helix packing mediated by 2'-hydroxyl groups is frequently observed in the crystalline structures of RNA duplexes. For example, the UUCG [{\text{GGACUUGGGUC}]} and UUG [{\text{GGACUUGGGUC}]} dodecamer duplexes both use 2'-hydroxyls for helix packing despite having drastically different alignments on the helical units. Both duplexes form extended arrays of coaxially stacked helices. The UUCG helical arrays are aligned parallel to each other within the crystal lattice, and form four tertiary hydrogen bonds per dodecamer duplex. Each hydrogen bond involves pairs of 2'-hydroxyls, where a hydroxyl on one helix serves as a hydrogen-bond donor to a hydroxyl on the second helix. The UUG dodecamer duplex has a perpendicular alignment of the helices, but again, all the direct helical interactions involve at least one 2'-hydroxyl.

Docking of the P1 helix into the active site of the group I intron has served as a prototype for RNA-helix packing. The P1 helix comprises the last few nucleotides of the 5'-exon and a complementary internal guide sequence within the intron (Fig. 1). Docking is mediated primarily by 2'-hydroxyls. Following duplex formation, the P1 helix is aligned in the active site of the intron where the 5'-exon is cleaved during the first step of splicing. Chemical substitution of individual 2'-hydroxyls within the P1 duplex demonstrates that three hydroxyls located on both strands of the helix mediate docking into the RNA active site. The hydroxyl-bonding partner for one of the hydroxyl groups has been identified as a phylogenetically conserved A at position 302 within the J8/J7 single-stranded region of the intron. The hydroxyl likely donates a hydrogen bond to the N1 position of the A. Similar tertiary interactions have been proposed for the other two hydroxyls.

While hydrogen bonding to 2'-hydroxyl groups is important for helical packing, the hydroxyls alone are insufficient to define a specific binding register for a helix. For example, the P1 helix can dock in the active site in alternate helical registers using commensurate sets of 2'-hydroxyls. Mis docking causes the intron to improperly define the 5'-exon/intron boundary. This implies that sequence-specific tertiary interactions with the nucleotide bases must also be made to properly define the binding register of RNA helices.

**Noncanonical base pairs**

Long, fully complementary helices are quite rare in structured RNAs. Instead, the helices are short (ten base pairs or less) and usually include unpaired nucleotides and noncanonical base pairs (base pairs other than G-C and A-U). In many cases these helical perturbations are phylogenetically conserved, implying that they are important for RNA function. Noncanonical base pairs (A-A, G-A, G-G, C-U, etc.) present chemical groups in either the major or minor groove of the helix that can be used for unique tertiary interactions. They also enhance major groove accessibility at the ends of continuous helices.

The general importance of these noncanonical base pairs in RNA duplex recognition has been indirectly demonstrated by combinatorial selection. Ribozymes were selected for their ability to bind a short oligoribonucleotide substrate and to ligate it onto its 5'-termini. Instead of using the fully complementary sequence that was experimentally designed into the RNA population, all of the active ribozymes use a different oligoribonucleotide-binding site within the variable region. These binding sites include several noncanonical base pairs. This implies that an RNA duplex containing helical imperfections is more readily packed into an RNA tertiary structure than a duplex with complete sequence complementarity.

Noncanonical base pairs have been directly implicated in P1 duplex recognition. The P1 helix includes a phylogenetically conserved (99%) G-U wobble...
The GAAA tetraloop docks into the minor groove of the tetraloop receptor. The three A's of the loop (in gold) stack on bases on one side of the receptor helix (green). Stacking is mediated by the presence of an adenine platform in the receptor helix (see text). Each adenosine of the loop also makes specific hydrogen bonds to receptor nucleotides.

Figure 3
The GAAA tetraloop docks into the minor groove of the tetraloop receptor. The three A's of the loop (in gold) stack on bases on one side of the receptor helix (green). Stacking is mediated by the presence of an adenine platform in the receptor helix (see text). Each adenosine of the loop also makes specific hydrogen bonds to receptor nucleotides.

The GAAA tetraloop docks into the minor groove of the tetraloop receptor. The three A's of the loop (in gold) stack on bases on one side of the receptor helix (green). Stacking is mediated by the presence of an adenine platform in the receptor helix (see text). Each adenosine of the loop also makes specific hydrogen bonds to receptor nucleotides.
adjacent adenosines in the receptor internal loop that lie side by side to form an adenosine platform. This unusual configuration of nucleotides results in a kink in the ribose backbone that opens the minor groove of the tetraloop receptor. Although the tetraloop receptor is an asymmetric internal loop, its structure provides nearly coaxial alignment of the flanking helices.

In addition to stacking, each adenosine in the GAAA loop makes specific hydrogen bonds to the tetraloop receptor, revealing why the loop-receptor interaction is sequence-specific. The first A of the GAAA tetraloop forms an A-U base pair with an A-U reverse Hoogsteen base pair in the receptor. The second A (GAAA) stacks on top of this triple and makes three ribose-mediated contacts to the U-A pair and to the C-G base pair above it in the tetraloop receptor. The third A (GAAA) forms an extensive hydrogen-bonding interface with the C-G base pair predicted to be important for domain folding.

Related modes of tetraloop docking can occur with a subset of the nucleotides found in the 11-nucleotide receptor motif. In a hammerhead ribozyme crystal structure, for example, the second and third As of the GAAA loop of one molecule are hydrogen-bonded to consecutive C-G base pairs in an adjacent molecule in the crystal lattice. Sequence co-variations in group I introns suggest that an analogous interaction may occur between GGUG tetraloops and the minor groove of dinucleotide pairs. Tetraloop-receptor binding is likely to be a common mechanism of helical docking in large globular RNAs.

Recognition of the P4 helix by the A-rich bulge

A second example of loop-minor groove docking occurs in the P4-P6 domain between the P4 helix and an internal loop, called the A-rich bulge (Fig. 1). The backbone of the A-rich bulge forms a corkscrew turn with the phosphates at the interior of the structure and the bases flipped out (Fig. 4). Two of the four adenosines of the bulge are positioned at the junction of three helices (P5a, P5b and P5c), forming an intricate network of stacking interactions, metal coordination and hydrogen bonding. The other two adenosines are hydrogen-bonded to base pairs in the minor groove of helix P4 (Fig. 4). The unusually close proximity of the phosphates in the backbone requires the direct coordination of several magnesium ions to stabilize the structure.

The backbone of the P4 helix and the A-rich bulge interact through formation of a short 'ribose zipper', characterized by shared hydrogen bonds between the 2'-hydroxyl and purine N3 (or pyrimidine O2) of one base and the 2'-hydroxyl of its partner (Fig. 5). A similar interaction occurs between the backbones of the tetraloop and its receptor. The ribose zipper may be a common way of stabilizing the close approach of phosphate backbones in regions of helical packing.

Other potential motifs

It is likely that additional unidentified structural motifs are employed by RNA to package A-form helices into tertiary structures. Two potential examples that have not yet been observed in a tertiary RNA are G-quartets and the extended triple helix.

The G-quartet, a hydrogen-bonded cyclic array of four G nucleotides, is a potential motif for zipping together two RNA duplexes (Fig. 6). Consecutive G-quartets stack to form quadruple four-stranded helices. An idealized RNA of the sequence UGGGU forms U-quartets and G-quartets in a structure that is unusually stable. G-quartets have been implicated in HIV genomic RNA dimerization in vivo, although additional examples of G-quartets in intramolecular tertiary structures are scarce.

The triple helix is another potential motif for packaging RNA helices. There are two categories of DNA triple helices: the purine and the pyrimidine motifs, although only the pyrimidine form of the triplex is stable in RNA model systems. This motif involves Hoogsteen base pairing.
hydrogen bonding of a parallel pyrimidine third-strand to form U-A-U and C-G-C triplets (Fig. 5). The C-G-C triplet requires protonation at the N3 position of C. This makes the complex highly pH-sensitive. Biochemical characterization of model RNA duplexes has shown that RNA can adopt the pyrimidine motif over an extended sequence length (18 nucleotides) at neutral to acidic pH (Ref. 49). While there are examples of individual base triplets from both triplex motifs, an extended triple helix has not yet been observed in a naturally occurring RNA, possibly owing to the pH sensitivity of the complex.

Concluding remarks

How an RNA 'sees' an RNA duplex is an intriguing structural problem with numerous solutions. RNA-packing motifs share common elements, including base stacking, hydrogen bonding to 2'-hydroxyls, coordination of metal ions and helical distortion by noncanonical base pairs. These elements are used in a diverse collection of structural motifs that define the chemical basis of RNA function.

Acknowledgements

We thank members of our laboratories for helpful discussions and review of the manuscript. This work was supported by NIH grant GM48393-01 and an ACS FRA (S. A. S.); NIH grant GM22778-21, Lucille P. Markey Charitable Trust Scholars Program, the Donaghue Medical Research Foundation, the Searle Scholars Program and the Beckman Young Investigators Program (J. A. D.).

References

11. Cate, J. H. et al. (1996) Science 273, 1578-1585
35. Cate, J. H. and Doudna, J. A. (1996) Science 272, 1221-1229
38. Cate, J. H. et al. (1996) Science 273, 1696-1699

Letters to TIBS

TIBS welcomes letters on any topic of interest. Please note, however, that previously unpublished data and criticisms of work published elsewhere cannot be accepted by this journal.

Letters should be sent to:

Jo McEntyre,
Trends in Biochemical Sciences
Elsevier Trends Journals
68 Hills Road
Cambridge, UK CB2 1LA