# 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? Noncanonical 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[1]. 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. By 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[2].

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[3]. 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[1]. 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 molecule are likely to be common to any RNA.

**S. A. Strobel** and **J. A. Doudna** are in the Department of Molecular Biophysics and Biochemistry, 260 Whitney Avenue, Yale University, New Haven, CT 06520, USA.

## 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[4–7]. 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[9]. 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[11] (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 in the interior of the helix, the ubiquitous 2′-hydroxyls line the outer edge

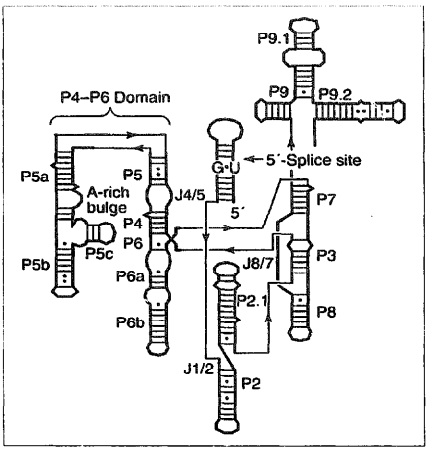


**Figure 1**
Schematic diagram of the *Tetrahymena* group I intron showing the 5′-splice site and the helical regions discussed in the text. The 3′-splice site is not shown.



 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 crystal structures of RNA duplexes. For example, the UUCG [(rGGACUUCGGUCC)$_2$] and UUUG [(rGGACUUUGGUCC)$_2$] dodecamer duplexes both use 2′-hydroxyls for helix packing despite having drastically different alignments of the helical units[12,13]. 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[12]. 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 UUUG dodecamer duplex has a perpendicular alignment of the helices, but again, all the direct helical interactions involve at least one 2′-hydroxyl[13].

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[14–16]. 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[14–16]. The hydrogen-bonding partner for the hydroxyl at the −3 position of the exon has been identified as a phylogenetically conserved A at position 302 within the J8/7 single-stranded region of the intron[17]. 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[18].

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[19,20]. Misdocking 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)[21]. 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[22].

The general importance of these noncanonical base pairs in RNA-duplex recognition has been indirectly demonstrated by combinatorial selection[23]. Ribozymes were selected for their ability to bind a short oligoribonucleotide substrate and to ligate it onto their 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 oligonucleotide-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.

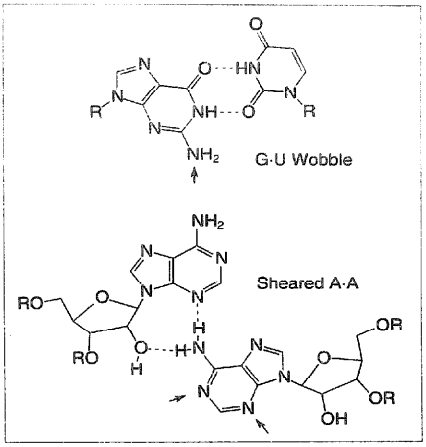


**Figure 2**
Noncanonical base pairs important for helix packing in the group I intron. Both noncanonical pairs mediate P1 docking into the intron active site. Arrows indicate the chemical groups likely to be involved in tertiary hydrogen-bonding.

Noncanonical base pairs have been directly implicated in P1 duplex recognition[24–27]. The P1 helix includes a phylogenetically conserved (>99%) G·U wobble

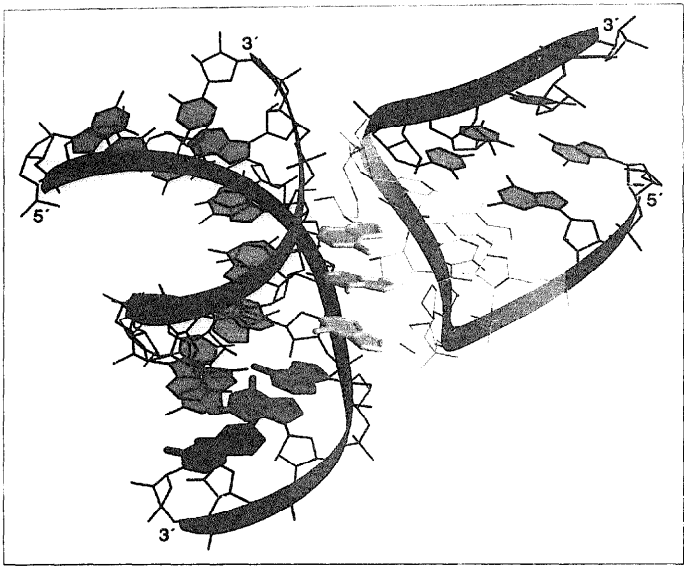


**Figure 3**

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

pair that defines the 5′-splice site[28] (Fig. 2). The splice site can be efficiently transferred to other positions by changing the location of the G·U pair within the helix[29]. This strongly implies that the noncanonical pair is important for defining the binding register of the P1 helix in the active site. Characterization of ribozymes containing nucleotide analog substitutions of the G·U pair demonstrates that the N2 amino group is a fundamental determinant of P1 helix recognition[27]. Like the 2′-hydroxyls discussed above, the exocyclic amine is located in the helical minor groove, suggesting that the minor groove is the primary surface used in packing the P1 helix into the active site. A similar set of determinants is used by $tRNA^{Ala}$ synthetase to bind the aminoacyl acceptor stem of its cognate tRNA[30]. This suggests a strong similarity between RNA–protein and RNA–RNA interactions.

The J4/5 region provides a second example of noncanonical pairs employed in RNA duplex recognition. J4/5 is a phylogenetically conserved, asymmetric internal loop that connects the coaxially stacked P4 and P5 helices. Photocrosslinking experiments have shown that J4/5 is the receptor for the P1 helix when it docks into the active site[31]. In the P4–P6 crystal structure, noncomplementary A nucleotides in the J4/5 loop base-pair to form two, consecutively stacked sheared A·A pairs[11] (Fig. 2). Each noncanonical A·A pair orients the Watson–Crick hydrogen-bonding face of one A on the minor groove surface where it could form tertiary interactions with determinants on the P1 helix. Results from methylation interference experiments support the proposal that the hydrogen bonding face of these As is important for intron function[32]. Thus, noncanonical pairs within both the P1 and J4/5 regions serve as docking interfaces for helix packing.

### Metal ions in helix packing

All RNAs require divalent cations for folding and catalysis. This derives, in part, from the requirement that the negatively charged phosphate backbone is neutralized in order to facilitate close packing of RNA helices. Beyond this, however, metals bind to specific sites within the RNA where they provide an ionic scaffold for tertiary structure formation[11].

Much of what is currently known about metal-mediated interhelical packing in RNA has been inferred from phosphorothioate interference experiments[33]. In the *Tetrahymena* intron, splicing is disrupted if phosphorothioate substitutions are introduced at positions that line potential helical interfaces within the catalytic core[34]. It is likely that magnesium ions bind along these helical surfaces and serve as an ionic bridge for close-packing of the negatively charged phosphate backbones.

Two different classes of metal binding sites are seen in the P4–P6 crystal structure. There are two cases where a metal ion binds in the major groove of tandem G·U base pairs[35]. In these locations, outer sphere coordination to base keto groups gives rise to binding affinity. A second mode of magnesium ion binding involves inner sphere coordination to either $R_p$ phosphate oxygens or the O6 keto group of G[11]. A cluster of these sites within the A-rich bulge of P5a plays a key role in helical packing within the domain (J. H. Cate *et al.*, unpublished).

### Higher-order structural motifs

While each of the elements discussed in the previous sections plays an important role in RNA helical packing, the crystal structure of the P4–P6 domain provides examples of helical-packing motifs that use complex combinations of these elements[11]. Within the P4–P6 domain, the P5a–P5b helix is docked against the P4–P5–P6 helix. This is achieved primarily by interaction at two helical interfaces. The GAAA tetraloop at the end of the P5b packs into the minor groove of J6a/6b, and the A-rich bulge of P5a packs into the minor groove of P4.

### Binding of the GAAA loop into the tetraloop receptor

Many structured RNAs contain hairpin loops of sequence GNRA (N, any nucleotide; R, purine nucleotide)[36]. Costa and Michel noticed that RNAs containing the GAAA subclass of these loops often contain an 11-nucleotide motif, consisting of two adjacent C-G base pairs, a five-nucleotide internal loop, and a G·U base pair[37]. They demonstrated that the two motifs form tertiary contacts in self-splicing introns.

The P4–P6 crystal structure provides an atomic resolution view of the interaction between the GAAA tetraloop and its canonical 11-nucleotide tetraloop-receptor duplex[11]. The three adenines in the tetraloop are stacked on the bases to the 5′ side of the tetraloop-receptor helix (Fig. 3). Stacking is facilitated by

adjacent adenosines in the receptor internal loop that lie side by side to form an adenosine platform[38]. 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·A·U triple 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[39].

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[40]. Sequence co-variations in group I introns suggest that an analogous interaction may occur between GUGA tetraloops and the minor groove of $\frac{5'CU3'}{3'GA5'}$ dinucleotide pairs[41,42]. Tetraloop-receptor binding is likely to be a common mechanism of helical packing in large globular RNAs.

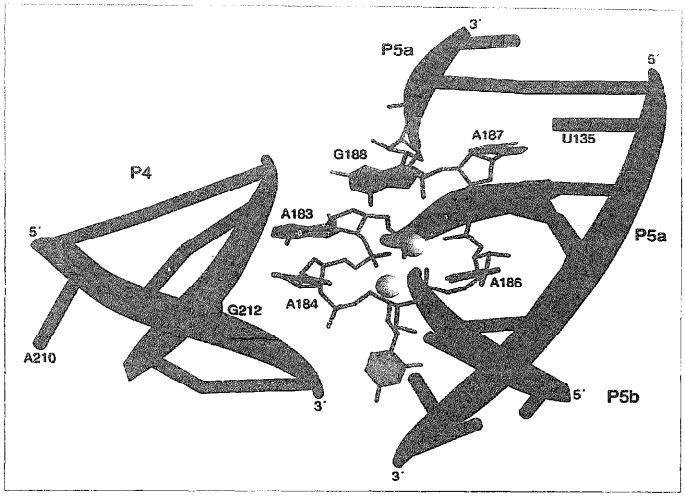


**Figure 4**
Adenosines in the A-rich bulge dock into the P4 helix minor groove. The A-rich bulge internal loop (orange) forms a corkscrew structure stabilized by magnesium ions (gold spheres). Two of the As in the bulge are hydrogen-bonded to the minor groove face of base pairs in helix P4 (blue).

### 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[11] (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-ion 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 backbones 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[11] (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[43]. G-quartets have been implicated in HIV genomic RNA dimerization *in vitro*[44], although additional examples of G-quartets in intramolecular tertiary structures are scarce.

**The triple helix** is another potential motif for packing RNA helices. There are two categories of DNA triple helices: the purine and the pyrimidine motifs[45,46], although only the pyrimidine form of the triplex is stable in RNA model systems[47,48]. This motif involves Hoogsteen

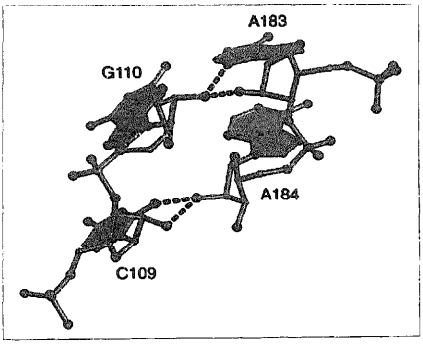


**Figure 5**
Ribose zippers occur between the backbones of the A-rich bulge and the P4 helix, and between the backbones of the GAAA tetraloop and the tetraloop receptor. Color scheme is the same as in Fig. 4.

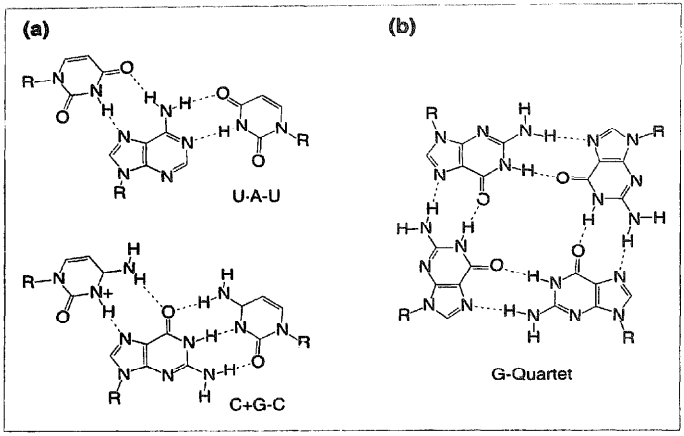


**Figure 6**
Two additional motifs that could potentially mediate helix packing in RNA tertiary structure. **(a)** The U·A-U and C+G-C pyrimidine triple helix. **(b)** The four-stranded G-quartet.

hydrogen bonding of a parallel pyrimidine third-strand to form U·A-U and C+G-C triplets[45] (Fig. 6). 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 triples 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 GM54839-01 and an ACS JFRA (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

1 Cech, T. R. and Bass, B. L. (1986) *Annu. Rev. Biochem.* 55, 599–629
2 Alden, C. J. and Kim, S-H. (1979) *J. Mol. Biol.* 132, 411–434
3 Seeman, N. C., Rosenberg, J. M. and Rich, A. (1976) *Proc. Natl. Acad. Sci. U. S. A.* 73, 804–808
4 Kim, S-H. *et al.* (1974) *Science* 185, 435–440
5 Robertus, J. D. *et al.* (1974) *Nature* 250, 546–551
6 Pley, H. W., Flaherty, K. M. and McKay, D. B. (1994) *Nature* 372, 68–74
7 Scott, W. G., Finch, J. T. and Klug, A. (1995) *Cell* 81, 991–1002
8 Downs, W. D. and Cech, T. R. (1990) *Biochemistry* 29, 5605–5613
9 Downs, W. D. and Cech, T. R. (1994) *Genes Dev.* 8, 1198–1211
10 Murphy, F. L. *et al.* (1994) *Science* 265, 1709–1712
11 Cate, J. H. *et al.* (1996) *Science* 273, 1678–1685
12 Holbrook, S. R. *et al.* (1991) *Nature* 353, 579–581
13 Baeyens, K. J., Bondt, H. L. D. and Holbrook, S. R. (1995) *Nat. Struct. Biol.* 2, 56–62
14 Pyle, A. M. and Cech, T. R. (1991) *Nature* 350, 628–631
15 Bevilacqua, P. C. and Turner, D. H. (1991) *Biochemistry* 30, 10632–10640
16 Strobel, S. A. and Cech, T. R. (1993) *Biochemistry* 32, 13593–13604
17 Pyle, A. M., Murphy, F. L. and Cech, T. R. (1992) *Nature* 358, 123–128
18 Michel, F. and Westhof, E. (1994) *Nat. Struct. Biol.* 1, 5–7
19 Herschlag, D. (1992) *Biochemistry* 31, 1386–1399
20 Strobel, S. A. and Cech, T. R. (1994) *Nat. Struct. Biol.* 1, 13–17
21 Gutell, R. R. (1993) *Curr. Opin. Struc. Biol.* 3, 313–322
22 Weeks, K. M. and Crothers, D. M. (1993) *Science* 261, 1574–1577
23 Ekland, E. H., Szostak, J. W. and Bartel, D. P. (1995) *Science* 269, 364–370
24 Barfod, E. T. and Cech, T. R. (1989) *Mol. Cell. Biol.* 9, 3657–3666
25 Pyle, A. M. *et al.* (1994) *Biochemistry* 33, 13856–13863
26 Knitt, D. S., Narlikar, G. J. and Herschlag, D. (1994) *Biochemistry* 33, 13864–13879
27 Strobel, S. A. and Cech, T. R. (1995) *Science* 267, 675–679
28 Waring, R. B. *et al.* (1986) *Nature* 321, 133–139
29 Doudna, J. A., Cormack, B. P. and Szostak, J. W. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86, 7402–7406
30 Musier-Forsyth, K. and Schimmel, P. (1992) *Nature* 357, 513–515
31 Wang, J-F., Downs, W. D. and Cech, T. R. (1993) *Science* 260, 504–508
32 von Ahsen, U. and Noller, H. F. (1993) *EMBO J.* 12, 4747–4754
33 Gish, G. and Eckstein, F. (1988) *Science* 240, 1520–1522
34 Christian, E. L. and Yarus, M. (1993) *Biochemistry* 32, 4475–4480
35 Cate, J. H. and Doudna, J. A. (1996) *Structure* 4, 1221–1229
36 Woese, C. R., Winker, S. and Gutell, R. R. (1990) *Proc. Natl. Acad. Sci. U. S. A.* 87, 8467
37 Costa, M. and Michel, F. (1995) *EMBO J.* 14, 1276
38 Cate, J. H. *et al.* (1996) *Science* 273, 1696–1699
39 Murphy, F. L. and Cech, T. R. (1994) *J. Mol. Biol.* 1994, 49–63
40 Pley, H. W., Flaherty, K. M. and McKay, D. B. (1994) *Nature* 372, 111–113
41 Michel, F. and Westhof, E. (1990) *J. Mol. Biol.* 216, 585–610
42 Jaeger, L., Michel, F. and Westhof, E. (1994) *J. Mol. Biol.* 236, 1271–1276
43 Cheong, C. and Moore, P. B. (1992) *Biochemistry* 31, 8406–8414
44 Sundquist, W. I. and Heaphy, S. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 3393–3397
45 Moser, H. E. and Dervan, P. B. (1987) *Science* 238, 645–650
46 Beal, P. B. and Dervan, P. B. (1991) *Science* 251, 1360–1363
47 Roberts, R. W. and Crothers, D. M. (1992) *Science* 258, 1463–1466
48 Semerad, C. L. and Maher, L. J. (1994) *Nucleic Acids Res.* 22, 5321–5325
49 Han, H. and Dervan, P. B. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 3806–3810