RNA tectonics: towards RNA design

Eric Westhof, Benoît Masquida and Luc Jaeger

Our understanding of the structural, folding and catalytic properties of RNA molecules has increased enormously in recent years. The discovery of catalytic RNA molecules by Sidney Altman and Tom Cech, the development of *in vitro* selection procedures, and the recent crystallizations of hammerhead ribozymes and of a large domain of an autocatalytic group I intron are some of the milestones that have contributed to the explosion of the RNA field. The availability of a three-dimensional model for the catalytic core of group I introns contributed also a heuristic drive toward the development of new techniques and approaches for unravelling RNA architecture, folding and stability. Here, we emphasize the mosaic structure of RNA and review some of the recent literature pertinent to this working framework. In the long run, RNA tectonics aims at constructing combinatorial libraries, using RNA mosaic units for creating molecules with dedicated shapes and properties.

Address: Institut de Biologie Moléculaire et Cellulaire du CNRS - UPR 9002, 15 rue René Descartes, F-67084 Strasbourg Cedex, France.

Author for correspondence: Eric Westhof
e-mail: westhof@ibmc.u-strasbg.fr

Electronic identifier: 1359-0278-001-R0078

Folding & Design 01 Aug 1996, 1:R78–R88

 ISSN 1359-0278

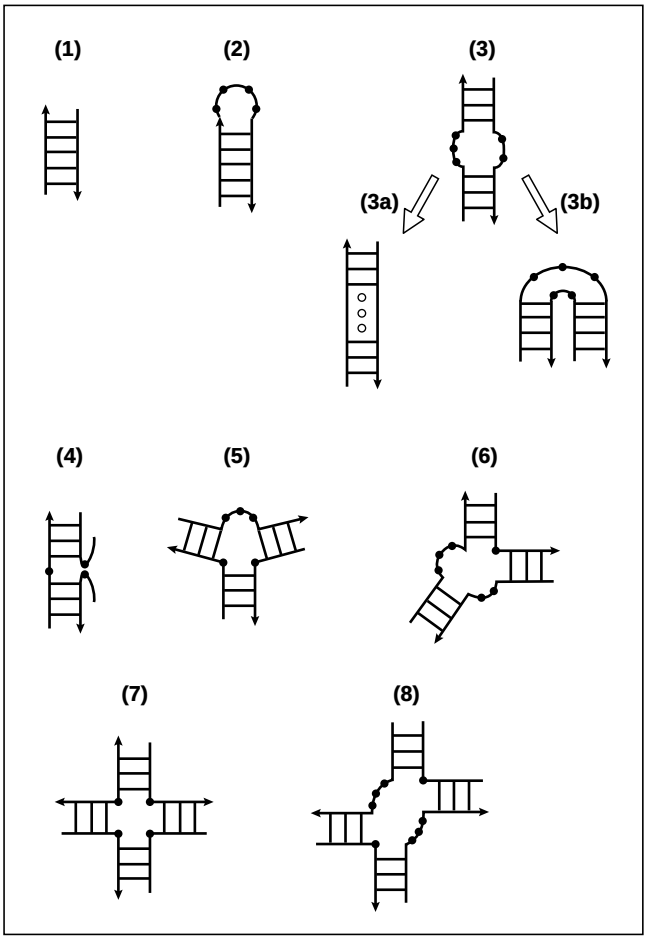
Introduction

According to the Concise Oxford Dictionary, 'tectonics' is the "art of producing useful and beautiful buildings" (from the Greek word meaning 'carpenter'). The word is used in several fields of science, especially in geology where plate tectonics refers to the movements of elements of the earth crust. The knowledge of the building blocks and of the rules governing the assembly of biological macromolecules into complex objects forms the basis for modelling and designing new objects. Although the building blocks of RNA appear, at first sight, to be well defined, the interacting and assembly rules between them are only just beginning to be delineated. The RNA field is thus not quite mature enough to allow the intelligent design of new molecules with specific properties. The most successful and efficient method for obtaining molecules with defined properties relies on random combinatorial chemistry followed by stringent selection procedures [1–4]. In the near future, one could hope to improve RNA design by exploiting our increased knowledge of RNA tectonics for the choice of the starting elements within and between which the random sequences are inserted. In this review, we select recent papers on RNA architecture and folding that broaden and increase our understanding of RNA logic. We insert them into a working framework, though far from fully established, with the hope that we will provoke and entice experimentalists and theoreticians from the protein to the RNA world.

Mosaic structure of RNA

The naturally occurring complex objects of biology, be they the haemoglobin protein, the ribosome, or the human eye, did not evolve as integral entities. Instead, biological evolution proceeds by assembling modular units, modifiable independently, sometimes redundant and often previously selected for other functions. The presence of RNA molecules at every single crucial and decisive step in the life of a cell attests to the ancestral glory and roles of RNA in the origin and evolution of life. While some RNA molecules have transient lifetimes, others are structurally central in intricate life machineries such as the ribosomes and spliceosomes. But, as for other biological objects, the complexity of the structures formed by RNA molecules can be resolved into separable pieces or modules as in a three-dimensional mosaic. Large RNA molecules are well known for being able to work in *trans*, i.e. functioning upon re-association of cut-out parts. The dissociation of complex RNA systems, such as autocatalytic introns, into parts, or modules, is facilitated by a property of nucleic acids: Watson–Crick pairing between complementary bases. The complementary Watson–Crick base pairs, with *cis* glycosyl bonds, form the only set of pairs that are isosteric in antiparallel helices. Thus, they allow formation of helices with regular, or quasi-regular, sugar-phosphate backbones. In single-stranded molecules, stacking and base pairing drive the folding of the chain on itself through the formation of helical regions linked by nonhelical elements, hairpin loops, internal bulges, and multiple junctions (see Fig. 1). The folded structure is usually discussed first in terms of a secondary structure which schematically represents the base-paired segments in a planar drawing. As in the protein field, the word 'folding' is often taken to mean alternatively the kinetic process whereby a non-native state of an RNA molecule converts to the native state or the overall RNA architecture itself described in terms of the underlying 3D motifs. For clarity, we will use the terms 'architecture' or '3D structure' when meaning or describing a native 'folded' state and the term 'folding' when discussing how the RNA molecule actually folds kinetically. The words 'assembly' or 'self-assembly' will also be used for characterizing static structures.

Figure 1



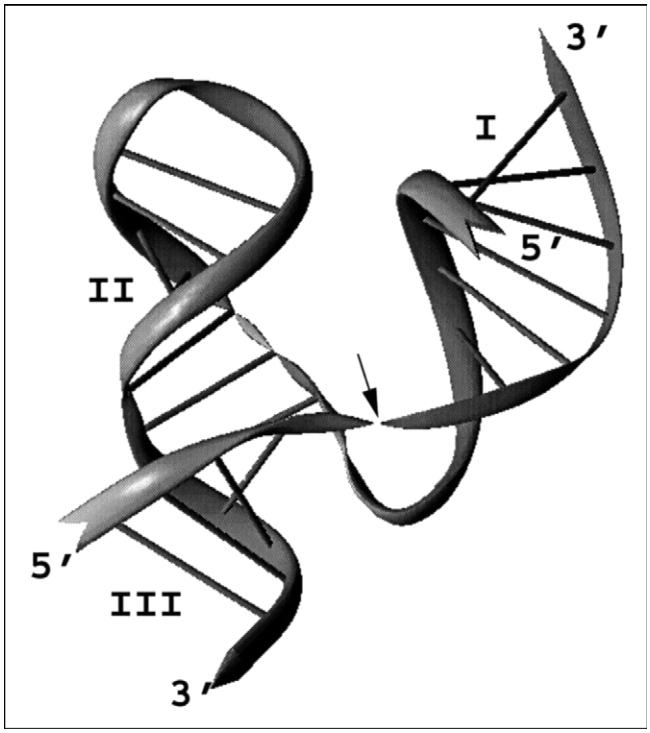
Schematic drawings of some typical RNA secondary structure elements. The arrows indicate the 5′→3′ direction. Bars denote Watson–Crick base pairs and dots phosphodiesters linking helices. (1) An RNA helix. (2) An RNA hairpin with four unpaired residues. (3) An asymmetric internal bulge between two helices. (4) Two stacked helices. (5) A three-way junction with one single-stranded region. (6) A three-way junction with two single-stranded regions. (7) A four-way junction. (8) A four-way junction with two single-stranded regions.

As noticed several years ago by Paul Sigler [5], and in contrast to proteins, RNA secondary structure is defined by hydrogen bonds (H-bonds) between sidechains and not by H-bonds between backbone atoms, as for example in α-helices. Thus, in RNA molecules, secondary structure determination is best achieved through phylogenetic sequence comparisons when more than one sequence is available (for reviews, see [6,7]) or by chemical probing of the sites involved in Watson–Crick pairing with the help of computer prediction [8,9]. Both approaches are complementary and both contain information potentially extending beyond secondary structures.

Small assemblies of secondary structure elements are known to form autonomous and functional entities. The hammerhead ribozyme [10] (Fig. 2), for example, for which two crystal structures are available, is made of a triad of helices [11,12]. At a higher hierarchical level, larger autonomously folded subdomains have been identified, such as the thoroughly studied P4–P5–P6 domain in the *Tetrahymena* ribozyme (Fig. 3) [13–16].

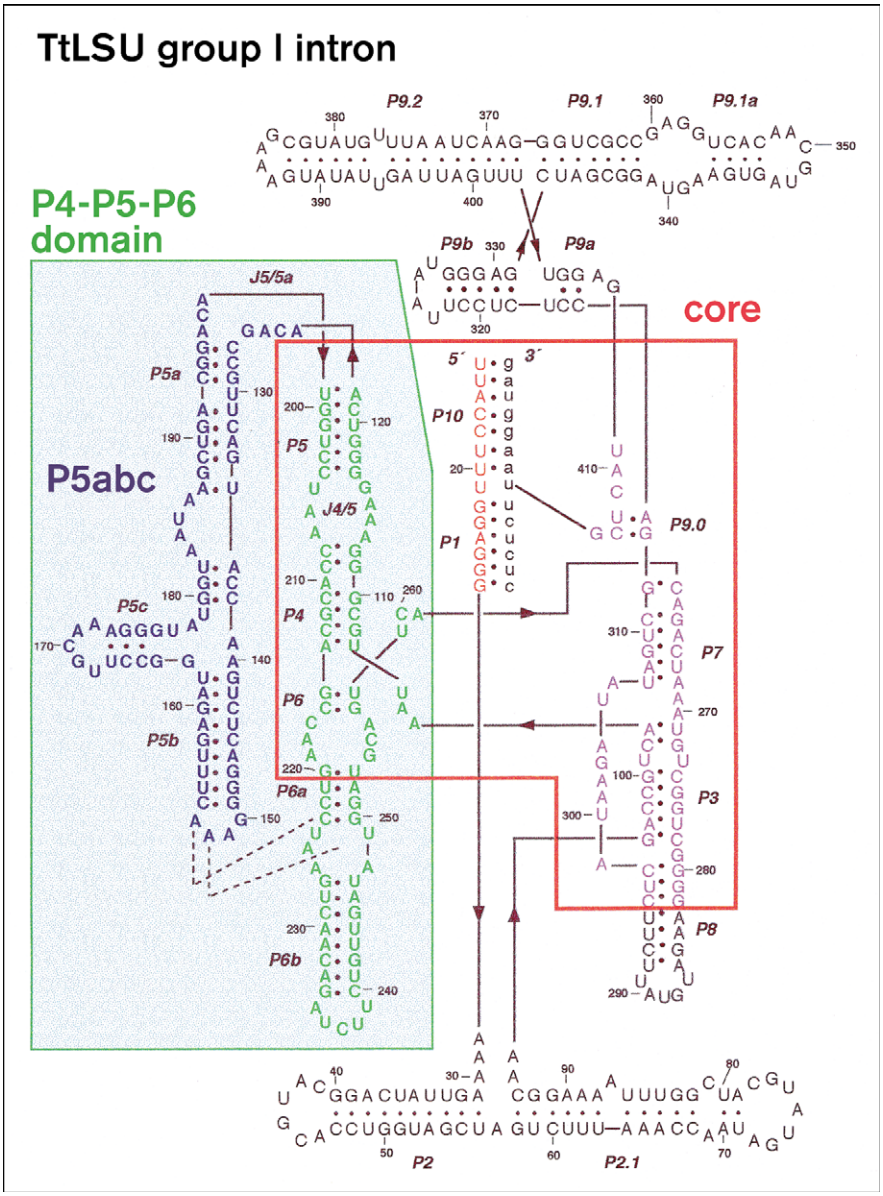
In the next level of organization, the tertiary structure, the secondary structure elements are associated through numerous van der Waals' contacts, specific H-bonds via the formation of a small number of additional Watson–Crick pairs and/or unusual pairs involving hairpin loops or internal bulges. The parsing of energy levels between secondary and tertiary structures is reasonable in large RNAs considering the relative energies and the clear identification of the secondary structure elements. In some cases it is even possible to cut RNA molecules into pieces that can re-associate only through tertiary contacts [17–19]. For example, the loss of activity in the *Tetrahymena* ribozyme (Fig. 3) concomitant with the removal of helices P5a, P5b, and P5c can be relieved by the subsequent addition of P5abc [19]. The association between P5abc and the domain P4–P6, into which P5abc docks, is essentially mediated by tertiary contacts, one of them

Figure 2



Ribbon drawing of the hammerhead ribozyme after the crystal structure of [11]. The cleavable bond is indicated by an arrow. Helices I, II, and III are indicated. In this structure, the substrate (DNA for cleavage inhibition) pairs to the ribozyme core, forming from its 5′- to its 3′-ends first helix III and then helix I.

Figure 3



Drawing of the secondary structure of the group I intron in the large subunit of rRNA in *Tetrahymena thermophyla*. The step shown is prior to the second transesterification reaction which will connect the 5′- and 3′-exons together (lower-case letters). The guide sequence, forming helices P1 and P10 with, respectively, the 5′- and 3′-exons is shown in brown. The core is enclosed within red lines. The P8–P3–P7–P9.0 domain is in violet. The P4–P5–P6 domain is enclosed within green lines and coloured in light blue. The dark blue domain P5abc is shown together with the tertiary contact between L5b and P6a (dotted lines). Adapted from [97].

resulting from the interaction of the -GAAA- loop of P5b with the internal loop of P6a [15,20].

In summary, an ensemble of observations [21–23] bears out a view of RNA folding whereby 3D architecture results from the cooperative compaction of separate, preformed, and stable sub-structures, which might undergo only minor and local rearrangement during the process (Fig. 4). Several recurrent 3D motifs that control and direct RNA–RNA recognition and RNA assembly have now been identified. We first discuss the building blocks and some assembly rules recently uncovered. We will not discuss the selecting and stabilizing roles of proteins in the RNA folding process (for discussion, see [24,25]).

Mosaic modelling of RNA

Levinthal's paradox applies to RNA as well as to proteins [26,27]. The number of possible conformers is even more astronomical in RNA than in proteins. Indeed, while there are three variable torsion angles in the polypeptide chain (and between zero and five in the sidechains), there are six variable torsion angles in polynucleotides (with one for the sidechain). Pauling and colleagues [28] reduced the number to two in polypeptide chains, and for nucleic acids, Sundaralingam [29] reduced severely two of them and contained the others to preferred domains. Thus, the torsion angles about the C–O bonds stick around 180°, while the sugar rings adopt either the C3′-*endo* or the C2′-*endo* puckers and the base is either in an *anti* or a *syn* orien-

Figure 4

An illustration of the way in which large RNA molecules are assembled. First, sub-domains are built (from left to right, clockwise): two stacked but interrupted helices, a triple helix, a pseudoknot, and a loop–loop interaction. In the middle, a complete group I intron, the sunY intron from bacteriophage T4, is shown [21].

tation with respect to the sugar. The two polymers adopt opposing strategies for inter-residue flexibility. In proteins, the torsion angles on either side of the peptide bond constitute the main flexible links, while in nucleic acids the phosphodiester bonds themselves direct chain re-orientations. Despite those conformational restrictions, the number of possible conformers for an RNA of a biologically relevant size is still intractable.

The introduction of modular units, hierarchically organized and folded, circumvents most, if not all, of the numerical nightmares inherent to mathematical modelling of macromolecules. 3D modelling based on a modular and hierarchical approach to RNA assembly has now been applied to several structured RNAs (see Fig. 4) [21,30–32]. The hierarchical framework implies that secondary structure elements form first, with the folding process moving progressively from small-scale elements to the larger scale sub-domains and domains. As such, this scenario does not imply that there is an interdependence between the sub-domains, i.e. that some secondary elements should form before others. In mosaic modelling of RNA, 3D motifs are first built and then assembled synchronically. Although this assembly process is often performed without due consideration to its actual time dependence, it is more than often pertinent to follow and exploit folding hypotheses as an insightful guide during modelling (for a review on folding kinetics, see [33]). Fundamentally, the hierarchical folding of RNA assumes that tertiary structure forms once all the secondary structure elements are present. This distinction is feasible because the secondary structure is easily identifiable and is energetically the main component of an RNA architecture while tertiary structure contributes only minimally to the stability of the native state in terms of free energy. One consequence is that RNA molecules can be observed with stable secondary structures in the absence of tertiary interactions. Such intermediate states, akin to the molten globule state of proteins, have been observed, under some experimental conditions, in a group I intron [21] and in the RNase P ribozyme of *Escherichia coli* [34].

In proteins, on the other hand, recent reports stress the importance of the solution environment [35] and of the tertiary context [36] on the formation of a given secondary structure element. Thus, an identical sequence, depending on the physico-chemical parameters of a solution or on its position in a protein, would adopt either the α-helical or the β-strand conformation. Such 'switch peptides' or 'chameleon sequences' are implicated in diseases such as Alzheimer's and prion-related dementia [37,38].

Similarly, in several cellular processes involving RNAs, regulatory strategies exploit RNA secondary structures by modulating through binding of another macromolecule (protein or RNA) the occurrence of alternative structures between which RNA molecules can often easily switch. Examples occur in the control of translation [39,40] and in transcription attenuation [41,42] in prokaryotes, as well as in the choice of pre-mRNA splicing sites in eukaryotes.

Assembly rules of RNA mosaic units

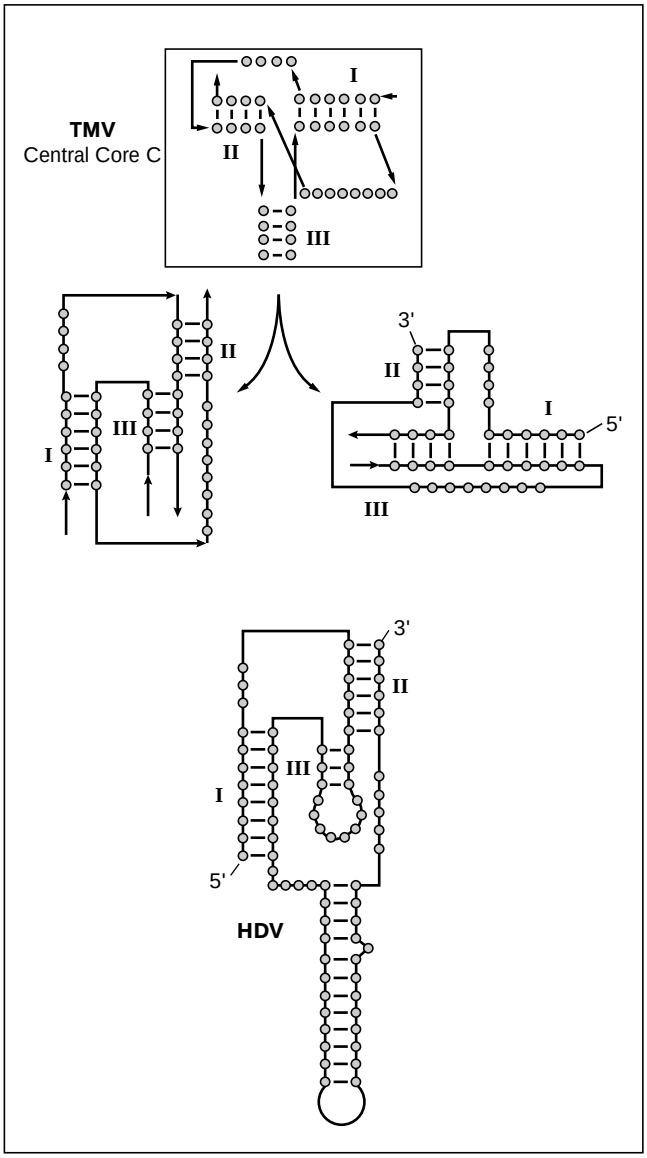
The very advantage of RNA, its conspicuous 2D structure, can be also the Achilles heel of the structuralist.

Indeed, the drawing of RNA secondary structures can be misleading. A long time ago, Struther Arnott [43] remarked that the cloverleaf structure of tRNA impeded early attempts at assembling a 3D model. Besides, in some instances, a case especially blatant for pseudoknots, the representation is arbitrary. Figure 5 shows three different drawings of a pseudoknot in the central core the tRNA-like domain present at the 3′-end of tobacco mosaic virus (TMV) [44]. One view reveals the hidden similarity in pairing schemes between the central core of TMV and the catalytic domain of hepatitis delta virus (HDV) [45]. The biological relevance of such similarities is difficult to evaluate, but the structural insight they bring along makes them valuable. The 3′-end of the RNA genome of some plant viruses, which folds as a tRNA-like domain due to the presence of an additional pseudoknot preceding the -CCA terminal part, has itself a still unknown function and much debated origin [46].

A recurring theme in the highly structured hairpins and internal bulges is that stacking between bases is maximized, thereby promoting H-bonding between bases or between bases and sugar-phosphate backbone. The same forces are at play at junctions between helices where, at the same time, they face the opposing electrostatic forces of bringing close together negatively charged backbones and the steric repulsions due to the bulkiness of RNA helices. As in tRNA (see Fig. 6), two helices with at least one contiguous strand will tend to stack co-axially [47,48], a conformation that maximizes stacking and H-bonding while minimizing electrostatic and steric repulsion. Such interplay between opposing forces rationalizes the fact that junctions with three helices do not occur in structured RNAs without at least one single-stranded stretch, the archetype being the three-way junction of 5S rRNAs. A three-way junction with two single-stranded stretches occurs in the hammerhead ribozyme. In junctions where four helices meet, on the other hand, it is common to observe continuity in base pairing [49,50]. In tRNAs, there are two single-stranded stretches linking the two sets of co-axial helices that stack on top of each other. In one stack, the two helices form a quasi-continuous helix with minor distortions (the amino acid and thymine helices). In the second stack, the loop–loop interaction between the thymine and the dihydrouridine loops imposes, at the junction between the two helices, a much larger twist angle than within a standard RNA helix (around 45° instead of 33°) [51].

Any complex secondary structure of a structured RNA, such as the 16S rRNA molecule, contains several examples of converging helices. Could one use the mosaic view of RNA structure to dissect a complex RNA molecule? Idealized junctions are shown in Figure 1: the 5S rRNA three-way junction (drawing 5 in Fig. 1 [31,52]); the hammerhead three-way junction (drawing 6 [11,12]); the Holliday-like four-way junction (drawing 7 [49,50]); and the tRNA four-way junction (drawing 8 [53]). Although it is often easy to recognize the type of junction, it is less easy to orient the helices with respect to each other according to a chosen module when it is

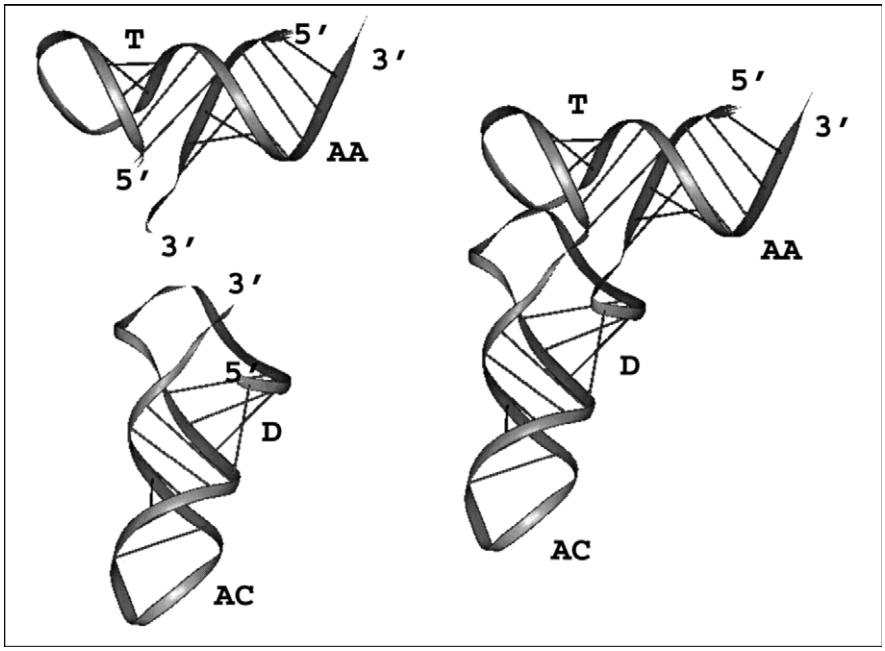
Figure 5



From tobacco mosaic virus (TMV) to hepatitis delta virus (HDV): the 3′-end of the TMV RNA genome contains a tRNA-like domain in which a pseudoknotted three-way junction is present (top); two other drawings of the corresponding secondary structure are shown below; the drawing at the left reveals the hidden similarity with the proposed secondary structure of the catalytic core of HDV shown at the bottom. In HDV, the cleaved phosphodiester is at the 5′-end of helix I, while the loop of hairpin III and the junction between helices I and II are crucial for catalysis. It is presently unknown whether this similarity is purely formal or reflects common biological function and/or origins. Adapted from [44].

Figure 6

Ribbon diagram of the tRNA structure, here the $tRNA^{asp}$ from yeast [51], on the right and its decomposition into two domains made of co-axially stacked helices on the left. Notice how the 3′-end of the anticodon (AC) helix runs over the deep groove of the dihydrouridine (D) helix. As mentioned in the text, the rotation angle between the AC and D helices is large, while the one between the acceptor (AA) helix and the thymine (T) helix is usual for RNA helices.



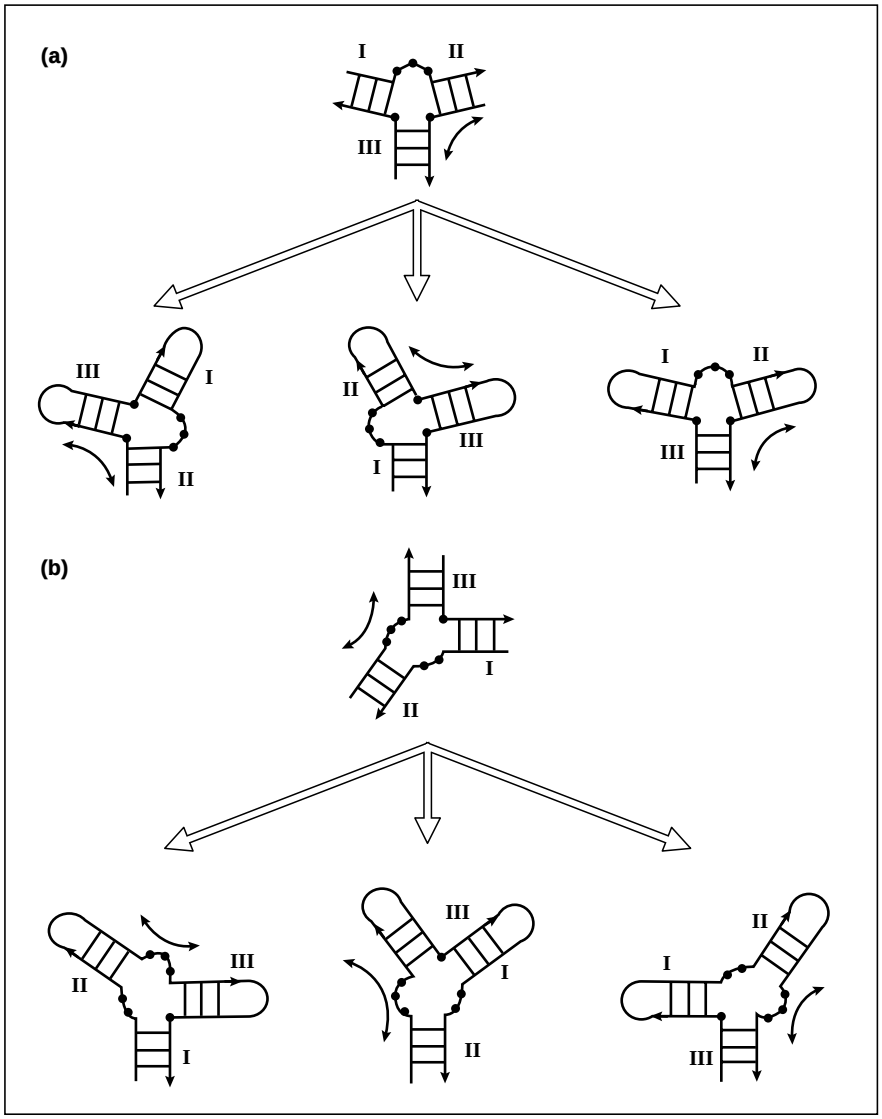
inserted within a complex structure. It is useful, though, to be able to make structural hypotheses on the relative arrangements of helices in a large RNA when discussing data or suggesting new experiments. The following discussion illustrates this point.

Right-handed RNA helices are highly asymmetric objects made of two strands of opposite polarity, each constituting chiral units. Single-stranded ends introduce a new element of asymmetry. Because of the right-handed rotation between nucleotides in helices, single-stranded nucleotides tend to leave or enter helices in a right-handed fashion. Thus, the 5′-strand of a helix would face the shallow groove and the 3′-strand the deep groove of a adjacent and stacked helix. Consider the 5S rRNA molecule for which there is evidence that helices II and III (Fig. 7) are roughly co-axial (indicated by the double arrowed arc in Fig. 7) with the single strand facing the shallow groove of helix III. Depending on the relative positions of the hairpin loops with respect to the single-stranded stretch, co-axial stacking of helices I and III (at the left) or of helices I and II (at the right) can be derived.

As remarked by Dan Herschlag [24], the high thermodynamic stability of RNA helices is coupled with a low information content of RNA secondary structure helices. These properties lead in RNA molecules to a tendency to fall into kinetic traps during folding and simultaneously to a difficulty in specifying a thermodynamically favoured tertiary structure. For a complex RNA molecule, the dependence of the 3D architecture on the presence of the extended and correct secondary structure might therefore be a necessity to guarantee the absence of kinetically trapped misfoldings. Some 3D RNA motifs are especially important both for guaranteeing the absence of kinetic misfoldings and for specifying a precise 3D structure. One of them, the pseudoknot [54–56], is governed by Watson–Crick base pairing between a hairpin loop and a single-stranded stretch or between two single-stranded stretches. Consequently, a pseudoknot can be considered either secondary structure (it constitutes an RNA helix with standard Watson–Crick pairs) or tertiary structure (a prerequisite for its formation being the presence of at least one hairpin). We prefer to consider a pseudoknot as a special 3D motif involving Watson–Crick pairs. Recent evidence shows that in a group I intron, the core of which is centred around a pseudoknot (helices P3 and P7), one helix forms after the other one, i.e. helix P7 follows helix P3 [57]. In this example, helix P7 constitutes a criterion for the validity of the correctness of the previously formed secondary structure elements. Thus, the origin of the frequent occurrence of pseudoknots in large structured RNAs, such as the autocatalytic group I intron [58] or RNase P ribozymes [59], might follow from the stringent constraints they impose and the control they exercise on the RNA folding pathway.

The interaction between two hairpin loops, called a loop–loop motif, also basically constitutes a pseudoknot. In a large RNA structure, loop–loop interactions can occur only once the two interacting hairpins are formed and properly positioned with respect to each other so that

Figure 7



(a) Analysis of the three-way junctions with one single-stranded region. The starting archetype is the 5S rRNA. One assumes that the co-axial stacking between helices II and III and present in the proposed models of 5S rRNAs is valid (indicated by the arc with double arrows). In a large RNA, three different situations could occur (from left to right). Symmetry and rotations perpendicular to the plane yield another six drawings. (b) Analysis of the three-way junctions with two single-stranded regions. The starting archetype is the hammerhead ribozyme where helices II and III stack on top of each other with noncanonical base pairings between residues of the joining single-stranded regions. One assumes that the co-axial stacking (indicated by the arc with double arrows) is maintained in a large RNA. Three different situations could occur (from left to right). Again, symmetry and rotations perpendicular to the plane yield another six drawings.

pairing can occur (see the 'kissing' complex [60,61]). Such interactions finalize the 3D structure, such as -GNRA- loops interacting in the shallow groove of helices, and are thus better considered as RNA–RNA anchors.

RNA–RNA anchors and RNA recognition motifs

The secondary structure indicates the positions of base-paired helices. These are linked by single-stranded regions that can form hairpins, internal bulges within helices, or link helices (see Fig. 1). The complexity and design variability of such structures is stunning and rivals those present in proteins.

Particularly stable hairpins (e.g. -UNCG- loops) [62] could form nucleation points for folding. Hairpins are often sites of protein binding (e.g. loop B of U1 snRNA) [63,64], but in the architecture of structured RNAs, hairpins play a crucial role as RNA–RNA anchors (see Fig. 8). Two hairpins can interact with each other, either solely through Watson–Crick base pairing (loop–loop motif) [65,66] or with a mixture of base pairing and intercalation as in the intramolecular association between the thymine and dihydrouridine loops of tRNAs (see Fig. 6). A hairpin can base pair with a single-stranded stretch forming an additional helix, leading to a pseudoknot. And finally, a family of tetraloop hairpins, the -GNRA- family, binds specifically to the shallow groove of another helix [15,67,68], while -GAAA- tetraloops recognize specifically an internal bulge [20]. A motif in which two -GNRA- loops recognize and interact with a single helix has been proposed for *E. coli* RNase P RNA [69]. The occurrence of the -GNRA-/shallow groove RNA–RNA anchor in

intermolecular contacts in the crystal structures of the hammerhead ribozyme [12,68] leads to its potential use when designing molecules for crystallogenesis.

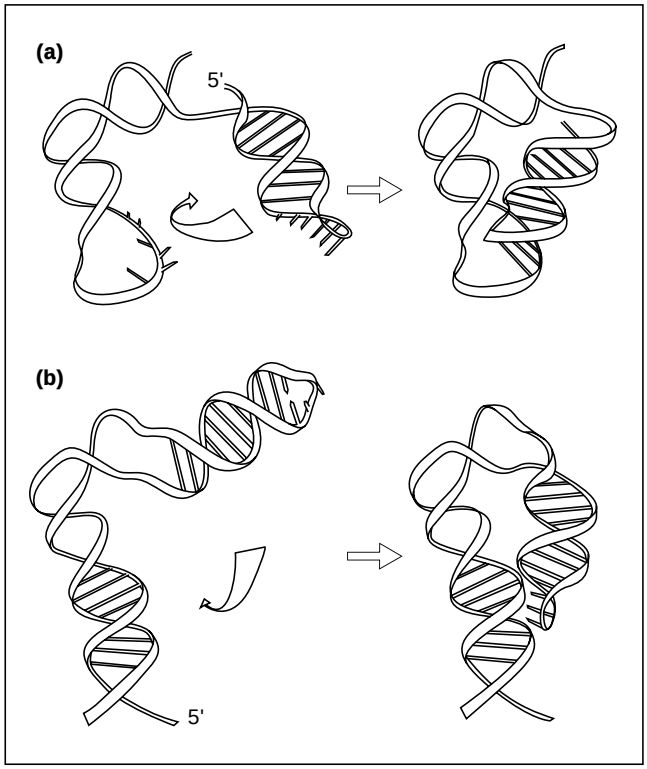
Internal bulges are frequently protein-binding sites (e.g. for Rev protein [70] and S8 ribosomal protein [71]). Structurally, they present variable flexibility [72,73], some of them being almost helical because of unusual base pairing (e.g. loop E of 5S rRNA [31,74]; see Fig. 1 (3a)) while others lead to a 180° turn [13,75] (e.g. J5/5a in *Tetrahymena*; see Fig. 1 (3b)). Internal bulges can also serve as RNA anchor points. In the *Tetrahymena* ribozyme, for example, the J4/5 internal loop recognizes specifically the invariant G–U base pair at the cleavage site [32,76]. How the base sequence governs the conformation of a given internal bulge and thus its potential structural and biological role is unknown.

Very recently, NMR structures of RNA aptamers obtained by *in vitro* selection have appeared. The structures are those of the ligand-bound aptamers. In each case, the recognition motif is embedded within a complex internal loop that displays a mixture of non-Watson–Crick interactions and of bulged-out bases. In one study [77], an aptamer against one amino acid, citrulline, was evolved against a related one, arginine (C=O to N-H change). The NMR analysis [78] indicated that the recognition scaffold remained invariant but three residues changed, one of which does not contact the recognized amino acid. For the aptamer evolved against flavin mononucleotide [79], the NMR structure [80] has the isoalloxazine both intercalating between a G–G mismatch and a G–U–A base triple while H-bonding to the Hoogsteen sites of an adenine base, which is located at the same level as the isoalloxazine ring.

Recent work on selenocysteine insertion in prokaryotes [81] and eukaryotes [82] illustrates the diversity of mechanisms that an RNA hairpin can subtend. In both cells, the stop codon UGA is recognized by the specific $tRNA^{Sec}$ and selenocysteine insertion requires the presence of an RNA hairpin. But the hairpins in the two kingdoms are very different in sequence conservation and structure. Besides, in prokaryotes the UGA codon is embedded in the hairpin whereas in eukaryotes the distance between the UGA codon and the hairpin can vary between 110 and 2700 nucleotides. The information content and recognition potential of RNA hairpins and internal bulges are clearly only beginning to be comprehended.

RNA recognition motifs can also be extracted from complex structures. Purohit and Stern [83], for example, have shown that a 45-ribonucleotide analogue of the decoding region of the 16S rRNA presents the same footprint with antibiotics and RNA ligands as the full 1542-nucleotide in the 30S subunit.

Figure 8



Schematic drawings illustrating two RNA–RNA anchors. Both motifs introduce new chiralities. (a) The single-stranded residues of a hairpin loop form Watson–Crick base pairs with an unpaired complementary sequence in another structural domain, thereby moulding a pseudoknot. As in standard RNA helices, the base pairing occurs with antiparallel polarities of the strands. (b) A -GNRA- tetraloop in a hairpin binds to the shallow groove of a helix belonging to another domain so that an A of the tetraloop interacts with either a G–C or a A–U pair and a G with only a A–U pair. The purines in the loop and in the helix are oriented with a parallel polarity.

Mosaic evolution of RNA

Since August Weismann's work at the end of the last century, biologists refer to 'mosaic evolution' as describing the evolution of (phenotypic or genotypic) constitutive units at differing rates and times. A system in which its constituent parts are so intricately correlated and integrated that it cannot be dissociated into modules cannot evolve without full redesign of the whole structure. From an evolutionary point of view, perfection is a dead-end path. Obviously, however, active molecules should be robust, i.e. must be able to withstand mutations, especially of conserved residues or residues involved in the active site. Thus, several residues, and not only those determinant for recognition or specific to catalysis, are bound to be redundant to some level. This phenomenon has been documented for proteic enzymes, for example serine proteases [84] or thymidylate synthase [85]. A well studied RNA system consists of catalytic group I introns, the cat-

alytic core of which comprises about 120 nucleotides. Among them, only seven nucleotides are strictly conserved with few exceptions [32,86], which allows for a great variety of possible sequences. The conserved residues are critical for recognition of the substrate, of the cofactor and catalytic activity. The next level in sequence order, covariations between residues distant in sequence, is rare in proteins but frequent in structured RNAs. Such covariations disclose closeness in space and conserved interactions within the RNA promoting tertiary architecture. Mutations in conserved or covarying residues can at times be rescued by increasing the concentration of magnesium ions, confirming their role in folding (e.g. see [87,88]). Some proteins, such as the much studied CYT-18 of *Neurospora crassa* mitochondria, can also suppress mutations at all conserved nucleotides in group I introns, except at a single position critical for catalysis [89,90]. Similarly, several mutations in the RNA component of ribonuclease P from *E. coli* could be compensated in the holoenzyme in the presence of the protein component [91]. In short, catalytic RNA molecules have evolved to become quite tolerant to base mutations.

Large RNA molecules have the amazing capability of functioning when their dissociated parts are put together if the dissection of the whole molecule is performed appropriately. Did the different modules of a complex RNA molecule with a defined biological function evolve simultaneously? Or did the constituent modules first evolve separately before being assembled into a complex structure? If the latter, did they evolve for the same function as in the final large structure? Or did they evolve for another unrelated function and undergo a functional shift? Are some motifs so stable or favourable that they reappeared several times? What is the lower limit in motif size below which adaptive biological convergence is meaningless? Indeed, non-orthologous similarities are common in molecular sequences and similarities in structure, with the resulting functional analogies, might be entirely due to chemical coincidences and historical contingencies bound to be present in macromolecules made from a small number of possible units. Surprising similarities have been observed. For example, the antibiotics that inhibit self-splicing of group I introns all belong to the class of antibiotics that perturb the decoding function of the small subunit of ribosomes [92]. This observation might imply similarities between recognition of the RNA duplex substrate by the core of group I intron and of the tRNAs and mRNA by the decoding region of the 16S rRNA [93]. Recently, similarities between the binding of *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase CYT-18 to $tRNA^{Tyr}$ and to several mitochondrial group I introns have been uncovered [94,95].

The tools of molecular evolution *in vitro* could be used for tackling such evolutionary issues. A better understanding of RNA evolution and a clearer description of the RNA landscape in sequence space should lead to a more focussed and, hopefully, efficient RNA design for specific functions. Is there a protean RNA motif, akin to the antibody β-barrel, adaptable to various binding properties? Hairpins and pseudoknots are frequently observed among aptamers selected from random sequences, but are they the most efficient and appropriate? Could one design an RNA scaffold for catalysis of numerous reactions like the proteic abzymes? Tentative designs, based on the P4–P5–P6 domain of the *Tetrahymena* ribozyme, are presently being developed [96].

References

1. Ellington, A.D. & Szostak, J.W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822.
2. Robertson, D.L. & Joyce, G.F. (1990). Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467–468.
3. Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510.
4. Breaker, R.R. & Joyce, G.F. (1994). Inventing and improving ribozyme function: rational design versus iterative selection methods. *Trends Biotechnol.* **12**, 268–275.
5. Sigler, P.B. (1975). An analysis of the structure of tRNA. *Annu. Rev. Biophys. Bioeng.* **4**, 477–527.
6. Westhof, E. & Michel, F. (1994). Prediction and experimental investigation of RNA secondary and tertiary foldings. In *RNA–Protein Interactions: Frontiers in Molecular Biology*. (Nagai, K. & Mattaj, I.W., eds), pp. 25–51, IRL Press at Oxford University Press, Oxford.
7. Woese, C.R. & Pace, N.R. (1993). Probing RNA structure, function, and history by comparative analysis. In *The RNA World.* (Atkins, R.F. & Gesteland, J.F., eds), pp. 91–117, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
8. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.P. & Ehresmann, B. (1987). Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**, 53–71.
9. Krol, A. & Carbon, P. (1989). A guide for probing native small nuclear RNA and ribonucleoprotein structures. *Methods Enzymol.* **180**, 212–227.
10. Symons, R.H. (1992). Small catalytic RNA. *Annu. Rev. Biochem.* **61**, 641–671.
11. Pley, H.W., Flaherty, K.M. & McKay, D.B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68–74.
12. Scott, W.G., Finch, J.T. & Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991–1002.
13. Murphy, F.L. & Cech, T.R. (1993). An independently folding domain of RNA tertiary structure within the *Tetrahymena* ribozyme. *Biochemistry* **32**, 5291–5300.
14. Murphy, F.L., Wang, Y., Griffith, J.D. & Cech, T.R. (1994). Coaxially stacked RNA helices in the catalytic center of the *Tetrahymena* ribozyme. *Science* **265**, 1709–1712.
15. Murphy, F.L. & Cech, T.R. (1994). GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J. Mol. Biol.* **236**, 49–63.
16. Doudna, J.A., Grosshans, C., Gooding, A. & Kundrot, C.E. (1993). Crystallization of ribozymes and small RNA motifs by a sparse matrix approach. *Proc. Natl. Acad. Sci. USA* **90**, 7829–7833.
17. Pan, T. (1995). Higher order folding and domain analysis of the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry* **34**, 902–909.
18. Doudna, J.A. & Cech, T.R. (1995). Self-assembly of a group I intron active site from its component tertiary structural domains. *RNA* **1**, 36–45.
19. Van der Horst, G., Christian, A. & Inoue, T. (1991). Reconstitution of group I intron self-splicing reaction with an activator RNA. *Proc. Natl. Acad. Sci. USA* **88**, 184–188.
20. Costa, M. & Michel, F. (1995). Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276–1285.
21. Jaeger, L., Westhof, E. & Michel, F. (1993). Monitoring of the coopera-

tive unfolding of the sunY group I intron of the bacteriophage T4. *J. Mol. Biol.* 234, 331–346.
22. Tanner, M.A. & Cech, T.R. (1996). Activity and thermostability of the small self-splicing group I intron in the pre-tRNAIle of the purple bacterium *Azoarcus*. *RNA* 2, 74–83.
23. Zarrinkar, P.P. & Williamson, J.R. (1994). Kinetic intermediates in RNA folding. *Science* 265, 918–924.
24. Herschlag, D. (1995). RNA chaperones and the RNA folding problem. *J. Biol. Chem.* 270, 20871–20874.
25. Weeks, K.M. & Cech, T.R. (1996). Assembly of a ribonucleoprotein catalyst by tertiary structure capture. *Science* 271, 345–348.
26. Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Phys.* 65, 44–47.
27. Zwanzig, R., Szabo, A. & Bagchi, B. (1992). Levinthal's paradox. *Proc. Natl. Acad. Sci. USA* 89, 20–22.
28. Pauling, L., Corey, R.B. & Branson, H.R. (1951). The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* 37, 205–211.
29. Sundaralingam, M. (1973). The concept of a conformationally 'rigid' nucleotide and its significance in polynucleotide conformational analysis. In *Conformation of Biological Molecules and Polymers.* (Bergmann, V.E.D. & Pullman, B., eds), pp. 417–456, Jerusalem.
30. Tanner, N.K., Schaff, S., Thill, G., Petit-Koskas, E., Crain-Denoyelle, A. & Westhof, E. (1994). A three-dimensional model of hepatitis delta virus ribozyme based on biochemical and mutational analysis. *Curr. Biol.* 4, 488–498.
31. Brunel, C., Romby, P., Westhof, E., Ehresmann, C. & Ehresmann, B. (1991). Three-dimensional model of *E. coli* ribosomal 5S RNA as deduced from structure probing in solution and computer modeling. *J. Mol. Biol.* 221, 293–308.
32. Michel, F. & Westhof, E. (1990). Modelling of the three-dimensional architecture of group-I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* 216, 585.
33. Draper, D.E. (1996). Parallel worlds. *Nat. Struct. Biol.* 3, 397–400.
34. Westhof, E., Wesolowski, D. & Altman, S. (1996). Mapping in three dimensions of regions in a catalytic RNA protected from attack by an Fe(II)-EDTA reagent. *J. Mol. Biol.* 258, 600–613.
35. Cerpa, R., Cohen, F.E. & Kuntz, I.D. (1996). Conformational switching in designed peptides: the helix/sheet transition. *Folding & Design* 1, 91–101; 1359-0278-001-00091.
36. Minor, D.L. & Kim, P.S. (1996). Context-dependent secondary structure formation of a designed protein sequence. *Nature* 380, 730–734.
37. Nguyen, J., Baldwin, M.A., Cohen, F.E. & Prusiner, S.B. (1995). Prion protein peptides induces alpha-helix to beta-sheet conformational transitions. *Biochemistry* 34, 4186–4192.
38. Zagorski, M.G. & Barrow, C.J. (1992). NMR studies of amyloid beta-peptides: proton assignments, secondary structure, and mechanism of an alpha helix-beta sheet conversion for a homologous, 28-residue N-terminal fragment. *Biochemistry* 31, 5621–5631.
39. Gold, L. (1988). Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Annu. Rev. Biochem.* 57, 199–223.
40. Philippe, C., Bénard, L., Portier, C., Westhof, E., Ehresmann, B. & Ehresmann, C. (1995). Molecular dissection of the pseudoknot governing the translation of *Escherichia coli* ribosomal protein S15. *Nucleic Acids Res.* 23, 18–28.
41. Balvay, L., Libri, D. & Fiszman, M.Y. (1993). Pre-mRNA secondary structure and the regulation of splicing. *Bioessays* 15, 165–169.
42. Yanovsky, C. (1988). Transcription attenuation. *J. Biol. Chem.* 2, 609–612.
43. Arnott, S. (1971). The structure of transfer RNA. In *Progress in Biophysics and Molecular Biology.* (Butler, J.A.V. & Noble, D., ed.), pp. 181–213, Pergamon Press, Oxford.
44. Felden, B., Florentz, C., Giegé, R. & Westhof, E. (1996). A central pseudoknotted three-way junction imposes tRNA-like mimicry and the orientation of three 5′ upstream pseudoknots in the 3′ terminus of tobacco mosaic virus RNA. *RNA* 2, 201–212.
45. Perrotta, A.T. & Been, M.D. (1991). A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. *Nature* 350, 434–436.
46. Felden, B., Florentz, C., Giegé, R. & Westhof, E. (1994). Solution structure of the 3′-end of brome mosaic virus genomic RNAs. *J. Mol. Biol.* 235, 508–531.
47. Kim, S.-H. & Cech, T.R. (1987). Three-dimensional model of the active site of the *Tetrahymena* ribozyme. *Proc. Natl. Acad. Sci. USA* 84, 8788–8792.
48. Walter, A.E., *et al.,* & Zuker, M. (1994). Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. USA* 91, 9218–9222.
49. Krol, A., Westhof, E., Bach, M., Lührmann, R., Ebel, J.-P. & Carbon, P. (1990). Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. *Nucleic Acids Res.* 18, 3803–3811.
50. Duckett, D.R., Murchie, A.I.H. & Lilley, D.M.J. (1995). The global folding of four-way helical junctions in RNA, including that in U1 snRNA. *Cell* 83, 1027–1036.
51. Westhof, E., Dumas, P. & Moras, D. (1985). Crystallographic refinement of yeast aspartic acid transfer RNA. *J. Mol. Biol.* 184, 119–145.
52. Westhof, E., Romby, P., Romaniuk, P.J., Ebel, J.P., Ehresmann, C. & Ehresmann, B. (1989). Computer modelling from solution data of spinach chloroplast and of *Xenopus laevis* somatic and oocyte 5S rRNAs. *J. Mol. Biol.* 207, 417–431.
53. Saenger, W. (1984). tRNA – a treasury of stereochemical information. In *Principles of Nucleic Acid Structure.* pp. 331–349, Springer Verlag, New York.
54. Pleij, C.W.A. (1990). Pseudoknots: a new motif in the RNA game. *Trends Biochem. Sci.* 15, 143–147.
55. Westhof, E. & Jaeger, L. (1992). RNA pseudoknots: structural and functional aspects. *Curr. Opin. Struct. Biol.* 2, 327–333.
56. Shen, L.X. & Tinoco, I. (1995). The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol.* 247, 963–978.
57. Zarrinkar, P.P. & Williamson, J.R. (1996). The kinetic folding pathway of the *Tetrahymena* ribozyme reveals possible similarities between RNA and protein folding. *Nat. Struct. Biol.* 3, 432–438.
58. Michel, F., Jaquier, A. & Dujon, B. (1982). Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* 64, 867–881.
59. James, B.D., Olsen, G.J., Liu, N.R. & Pace, N.R. (1988). The secondary structure of ribonuclease P RNA, the catalytic element of a ribonucleoprotein enzyme. *Cell* 52, 19–26.
60. Persson, C., Wagner, E.G.H. & Nordström, K. (1990). Control of replication of plasmid R1: structures and sequences of the antisense RNA, CopA, required for its binding to the target RNA, CopT. *EMBO J.* 9, 3767–3775.
61. Tomizawa, J. (1990). Control of ColE1 plasmid replication: binding of RNAI to RNA II and inhibition of primer formation. *J. Mol. Biol.* 212, 683–694.
62. Tuerk, C., *et al.,* & Gold, L. (1988). CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. USA* 85, 1364–1367.
63. Oubridge, C., Ito, N., Evans, P.R., Teo, C.-H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 372, 432–438.
64. Allain, F.H.T., Gubser, C.C., Howe, P.W.A., Nagai, K., Neuhaus, D. & Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature* 380, 646–650.
65. Gregorian, R.S.J. & Crothers, M.D. (1995). Determinants of RNA hairpin loop–loop complex stability. *J. Mol. Biol.* 248, 968–984.
66. Marino, J.P., Gregorian, R.S.J., Csankovszki, G. & Crothers, M.D. (1995). Bent helix formation between RNA hairpins with complementary loops. *Science* 268, 1448–1454.
67. Jaeger, L., Michel, F. & Westhof, E. (1994). Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.* 236, 1271–1276.
68. Pley, H.W., Flaherty, K.M. & McKay, D.B. (1994). Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* 372, 111–113.
69. Brown, J.W., Nolan, J.M., Haas, E.S., Rubio, M.A.T., Major, F. & Pace, N.R. (1996). Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial populations reveals tertiary structural elements. *Proc. Natl. Acad. Sci. USA* 93, 3001–3006.
70. Iwai, S., Pritchard, C., Mann, D.A., Karn, J. & Gait, M.J. (1992). Recognition of the high affinity binding site in rev-responsive element RNA by the human immunodeficiency virus type-1 rev protein. *Nucleic Acids Res.* 24, 6465–6472.
71. Mougel, M., Allmang, C., Eyermann, F., Cachia, C., Ehresmann, B. & Ehresmann, C. (1993). Minimal 16S rRNA binding site and role of conserved nucleotides in *Escherichia coli* ribosomal protein S8 recognition. *Eur. J. Biochem.* 215, 787–792.
72. Bhattacharyya, A., Murchie, A.I.H. & Lilley, D.M. (1990). RNA bulges and the helical periodicity of double-stranded RNA. *Nature* 343, 484–487.
73. Zacharias, M. & Hagerman, P.J. (1996). The influence of symmetric

internal loops on the flexibility of RNA. *J. Mol. Biol.* 257, 276–289.
74. Wimberly, B., Varani, G. & Tinoco, I., Jr. (1993). The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry* 32, 1078–1087.
75. Wang, Y., Murphy, F.L., Cech, T.R. & Griffith, J.D. (1994). Visualization of tertiary structural domain of the *Tetrahymena* group I intron by electron microscopy. *J. Mol. Biol.* 236, 64–71.
76. Wang, J.F., Downs, W.D. & Cech, T.R. (1993). Movement of the guide sequence during RNA catalysis by a group I ribozyme. *Science* 260, 504–508.
77. Famulok, M. (1994). Molecular recognition of amino acids by RNA-aptamers: an L-citrulline binding RNA motif and its evolution into an L-arginine binder. *J. Am. Chem. Soc.* 116, 1698–1706.
78. Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. & Famulok, M. (1996). Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science* 272, 1343–1347.
79. Burgstaller, P. & Famulok, M. (1994). Isolation of RNA aptamers for biological cofactors by *in vitro* selection. *Angew. Chem.* 33, 1084–1087.
80. Fan, P., Suri, A.K., Fiala, R., Live, D. & Patel, D.J. (1996). Molecular recognition in the FMN–RNA aptamer complex. *J. Mol. Biol.* 258, 480–500.
81. Hüttenhofer, A., Westhof, E. & Böck, A. (1996). Solution structure of mRNA hairpins promoting selenocysteine incorporation in *Escherichia coli* and their base-specific interaction with special elongation factor SELB. *RNA* 2, 354–366.
82. Walczak, R., Westhof, E., Carbon, P. & Krol, A. (1996). A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA* 2, 354–366.
83. Purohit, P. & Stern, S. (1994). Interactions of a small RNA with antibiotic and RNA ligands of the 30S subunit [see comments]. *Nature* 370, 659–662.
84. Carter, P. & Wells, J.A. (1988). Dissecting the catalytic triad of a serine protease. *Nature* 332, 564–568.
85. Carreras, C.W. & Santi, D.V. (1995). The catalytic mechanism and structure of thymidilate synthase. *Annu. Rev. Biochem.* 64, 721–762.
86. Jaeger, L., Michel, F. & Westhof, E. (1996). The structure of group I ribozymes. In *Nucleic Acids and Molecular Biology.* (Eckstein, F. & Lilley, D.M.J., eds), pp. 33–51, Springer Verlag, Berlin.
87. Beaudry, A.A. & Joyce, G.F. (1990). Minimum secondary structure requirements for catalytic activity of self-splicing group I intron. *Biochemistry* 29, 6534–6539.
88. Jaeger, L., Westhof, E. & Michel, F. (1991). Function of P11, a tertiary base pairing in self-splicing introns of subgroup IA. *J. Mol. Biol.* 221, 1153–1164.
89. Mohr, G., Zhang, A., Gianelos, J.A., Belfort, M. & Lambowitz, A.M. (1992). The *Neurospora* CYT-18 protein suppresses defects in the phage T4 td intron by stabilizing the catalytically active structure of the intron core. *Cell* 69, 483–494.
90. Myers, C.A., *et al.,* & Lambowitz, A.M. (1996). A tyrosyl-tRNA synthetase suppresses structural defects in the two major helical domains of the group I intron catalytic core. *J. Mol. Biol*. in press.
91. Lumelski, N. & Altman, S. (1988). Selection and characterization of randomly produced mutants in the gene coding for M1 RNA. *J. Mol. Biol.* 202, 443–454.
92. Wank, H., Rogers, J., Davies, J. & Schroeder, R. (1994). Peptide antibiotics of the tuberactinomycin family as inhibitors of group I intron RNA splicing. *J. Mol. Biol.* 236, 1001–1010.
93. Schröeder, R., Streicher, B. & Wank, H. (1993). Splice-site selection and decoding: are they related? *Science* 260, 1443–1444.
94. Guo, Q. & Lambowitz, A.M. (1992). A tyrosyl-tRNA synthetase binds specifically to the group I intron catalytic core. *Genes Dev.* 6, 1357–1372.
95. Caprara, M., Mohr, G. & Lambowitz, A.M. (1996). A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core. *J. Mol. Biol.* 257, 512–531.
96. Wright, M.C., Jaeger, L. & Joyce, G.F. (1995). A new class of molecules for exploring the catalytical potential of RNA. *J. Cell. Biochem. Abstract Suppl.* 19A, 228.
97. Cech, T.R., Damberger, S.H. & Gutell, R.R. (1994). Representation of the secondary and tertiary structure of group I introns. *Nat. Struct. Biol.* 1, 273–280.