RNA Pseudoknots

R.M.W. Mans[1,2] and C.W.A. Pleij[1]

1 Introduction

Our view on RNA has changed radically over the last decade. In addition to its function as carrier of genetic information, RNA was shown to have several catalytic properties, culminating recently in the stunning fact that probably 23S ribosomal RNA alone participates in peptide bond formation (Noller et al. 1992).

The many biological activities require specific three-dimensional foldings of the RNA, but it is in just this aspect of natural RNA molecules that our understanding is limited or even lacking. Only for some tRNAs, encompassing some 75 nucleotides only, has a 3-D structure been resolved to atomic resolution.

Our insight in the folding of most RNAs is limited to the level of the secondary structure, i.e. the two-dimensional map of (Watson-Crick) base-pairing interactions. Such intramolecular double-helix formation leads to an assembly of structural elements like stems (double-helical regions) and a number of single-stranded regions or loops like hairpin, bulge, internal and multibranched (or multiple) loops or junctions (see Fig. 1; for a detailed discussion, Tinoco et al. 1990; Chastain and Tinoco 1991). Usually such secondary structures are proposed on the basis of computer prediction or covariation search and are at best supported by chemical modification, enzymatic digestion, or mutagenesis data.

The functionally relevant spatial folding is attained, however, by tertiary interactions like triple base formation, extra H-bonding between sugar and phosphate moieties, base intercalation, etc. as we have learned so well from the tRNA structure.

An important feature of this spatial folding of RNA has been shown in the last decade to be the formation of the RNA pseudoknot. An RNA pseudoknot is formed when, in the secondary structure of a single-stranded RNA molecule, nucleotides of a loop region base-pairs with a complementary region outside that loop. Although the possibility of this type of base-pairing

was first mentioned by Richards (1969) and Ninio (1971), it was Studnicka et al. (1978) who proposed the term pseudoknot in order to avoid confusion with the related, but topologically different, "real knot". The concept of pseudoknotting in RNA attracted more attention when Rietveld et al. (1982) proposed a model for the structure of the 3' terminus of turnip yellow mosaic virus (TYMV) RNA in which a pseudoknot structure played a crucial role in understanding its tRNA-like behavior. Since then, so many other RNA pseudoknots have been described in all kinds of RNAs, with often interesting structural and functional properties, that a review on this subject is justified (for other reviews, see Schimmel 1989; Pleij 1990; Puglisi et al. 1991; Westhof and Jaeger 1992).

   We will summarize here some basic aspects of the structure and function of RNA pseudoknots with emphasis on the structure of the so-called H-type pseudoknot and on problems related to classification and nomenclature.

2 Structural Aspects

Since a pseudoknot is defined as the base pairing of a loop with a sequence outside that loop, it is easy to see that in principle 14 different classes of pseudoknots can be discerned, if interloop interactions are included (see Fig. 1; Pleij 1990). Examples of almost all possibilities have been proposed in the literature. Apart from their function this has raised questions about their structural relationship and whether each of these 14 pseudoknots can be classified as separate species. This problem will be discussed in Section 3. There is, however, one type of pseudoknot which is not only widespread in
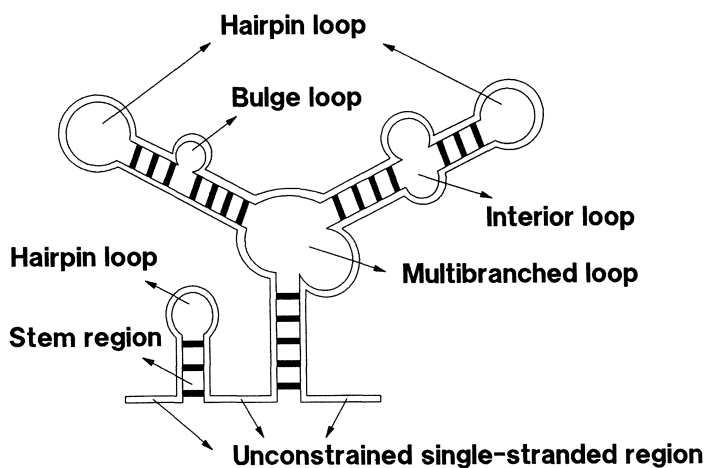

Fig. 1. Schematic presentation of the structural motifs in an orthodox secondary structure of RNA. The *black bars* indicate base pairs

nature but which has also provided most of our current understanding of the structure of RNA pseudoknots. This is the "classic" (Westhof and Jaeger 1992) or H-type pseudoknot (Pleij and Bosch 1989) of which the structural details will be described below.

2.1 The H-Type Pseudoknot

2.1.1 General Structural Features

The simplest way to create a loop in an RNA chain is by folding a hairpin. Consequently, base pairing of this loop with either the upstream or downstream single-stranded region gives rise to the most elementary form of a pseudoknot, which we called the H-type pseudoknot (hairpin) (Pleij and Bosch 1989).

The generalized form of this H-type pseudoknot is shown in Fig. 2A. The structure is characterized by two stems (S1 and S2) and three loop regions (L1–L3), but note that it can be represented in two equivalent ways, depending on whether one first draws S1 (the 5′ hairpin) or S2 (the 3′ hairpin). The presence of these three loops makes it difficult to predict the orientation of the two stems. This situation changes when one of the three loops is absent. In that case the base-pairing loop region becomes contiguous to one of the stems (S2 in Fig. 2B, C or S1 in Fig. 2D) and therefore coaxial stacking of S1 and S2 can be envisioned. This is schematically illustrated on the right-hand side of Fig. 2. Evidence for such a coaxial stacking was obtained for the first time from a study on the tRNA-like structure at the 3′ end of TYMV RNA (Rietveld et al. 1982). In order to explain its resemblance to elongator tRNAs a pseudoknot with coaxially stacked stems of the type in Fig. 2B was proposed on the basis of structure mapping techniques (Fig. 3A).

Coaxial stacking in an H-type pseudoknot can be obtained in three different ways, depending on which one of the three connecting loops is missing (Fig. 2B–D; Abrahams et al. 1990; Westhof and Jaeger 1992). They differ in the way the stem segments are spanned by the two remaining connecting loops or linkers. Due to the polarity in an A-form RNA helix, L_S bridges the shallow groove and L_D the deep groove of the formed, quasi-continuous double helix (see Pleij et al. 1985 for a more thorough discussion).

Note that in Fig. 2C,D one loop (L_P) is present which extends outside the helical cylinder and, in fact, not only crosses the grooves but also the sugar-phosphate backbone. On the basis of the geometry of a regular A-type helix and assuming a perfect quasi-continuous double helix, we have explored the minimal length requirements for spanning the grooves by L_S and L_D, respectively (Pleij et al. 1985). It is evident that this is dependent on the number of base pairs in both stem segments S1 or S2, but the surprising outcome was that the linker spanning the deep groove over 5–7
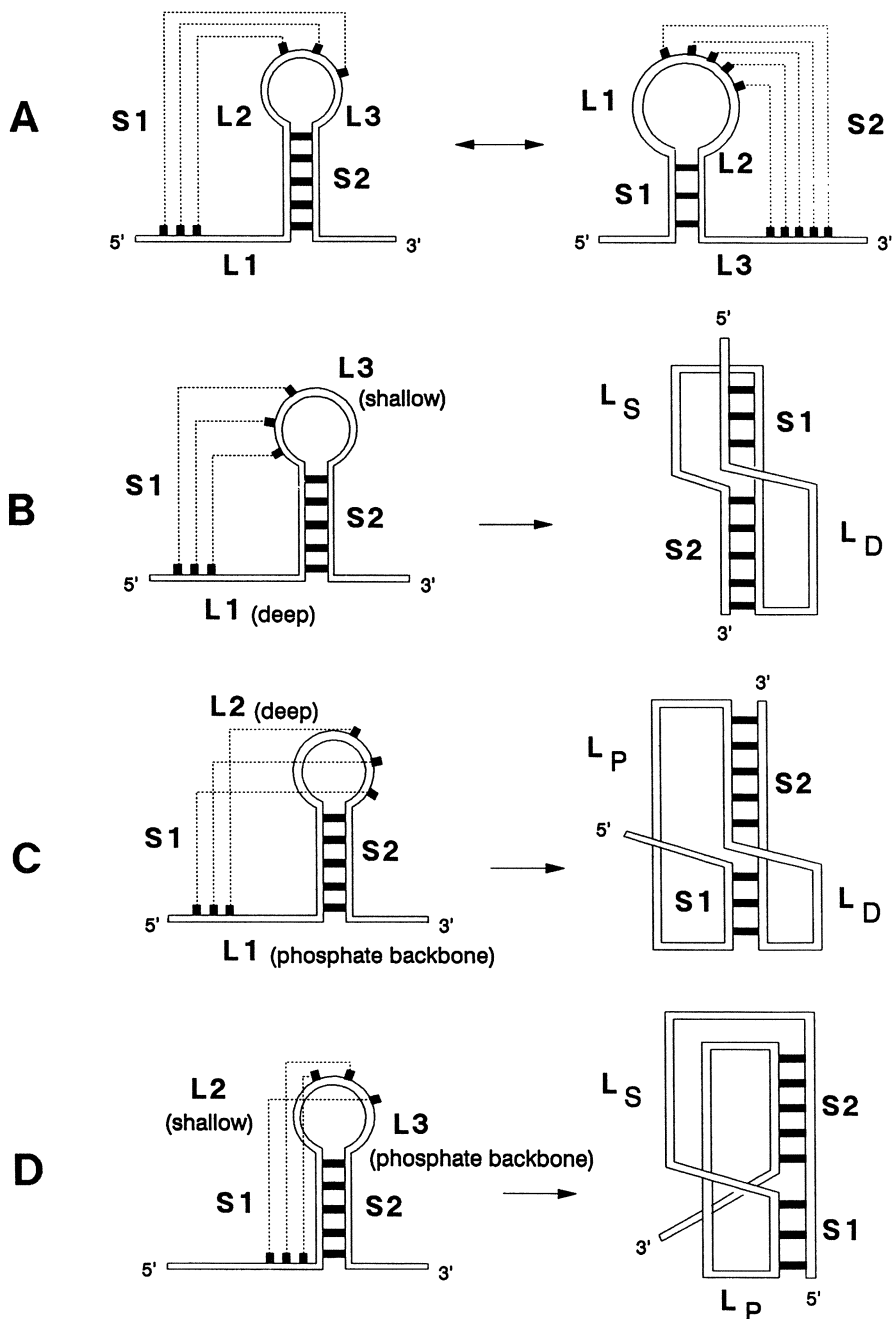
Fig. 2A–D. The H-type pseudoknot. **A** Two alternative, secondary structure presentations of the generalized H-pseudoknot. **B–D** Three classes of the H-type pseudoknot when connecting loop *L2* is zero (**B**), *L3* is zero (**C**), or *L1* is zero (**D**). The structures on the *right-hand side* give the coaxial stacking of *S1* and *S2*. Pseudoknotting is indicated by *dashed lines*. L_D is the loop crossing the deep groove, L_S the loop crossing the shallow groove, and L_P the loop spanning the sugar-phosphate backbone
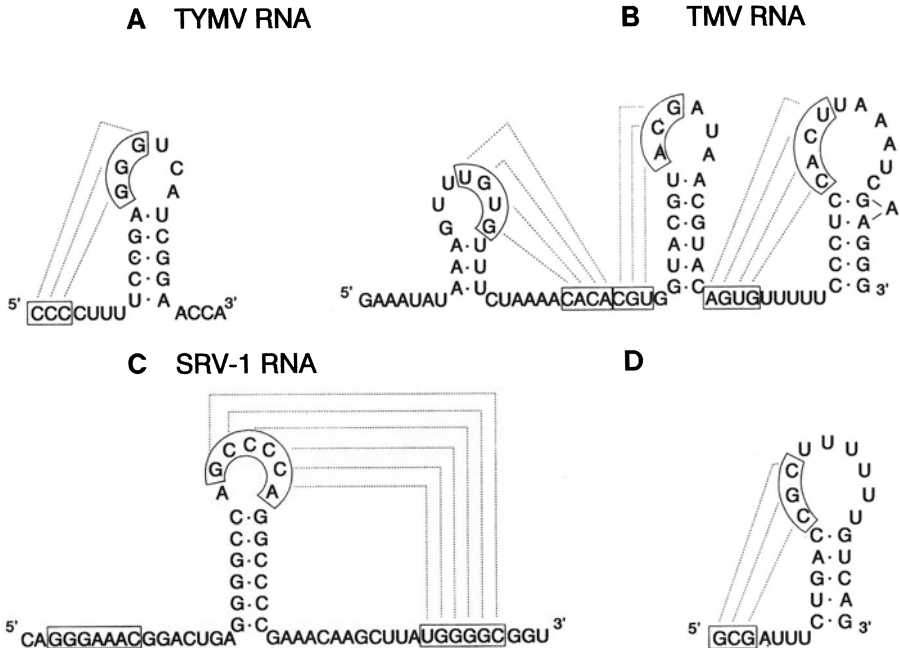
Fig. 3A–D. H-type pseudoknots. **A** The pseudoknot in the tRNA-like structure at the 3′ end of TYMV RNA. **B** Three consecutive pseudoknots located in the 3′ noncoding region of TMV (L strain) RNA between the coding region and the tRNA-like structure. **C** The pseudoknot in the *gag-pol* overlap region of simian retrovirus-1 (SRV-1) RNA. The *boxed* sequence is the shifty heptanucleotide. **D** Pseudoknot-containing RNA oligonucleotide studied with NMR by Tinoco and coworkers (see text)

base pairs can be as short as two nucleotides thanks to the close proximity of the two opposing strands across the deep groove (10–12 Å). It is even very likely that H-type pseudoknots may have loops as short as one nucleotide, as suggested by later structure mapping studies and sequence comparisons of various RNAs (Pleij et al. 1987). Two examples are shown in Fig. 3B,C, where a single A or G residue seems to be sufficient for crossing S2.

The shortest distance to cross the shallow groove was found to be 18 Å, which needs three nucleotides to be bridged. In Fig. 4 this is illustrated in a 3-D model of the pseudoknot in the tRNA-like structure of TYMV RNA. In that case three base pairs are present in S1. The fact that L_S, on average, should be larger than L_D is reflected in many natural H-type pseudoknots (e.g. Fig. 3B,C). In this respect the four nucleotides CUUU in TYMV RNA (Fig. 3A) would not be needed to span the deep groove over five base pairs. This was supported by sequence comparisons among related tymoviral RNAs, indicating a connecting loop of three or even two nucleotides at this position (Van Belkum et al. 1988). Site-directed mutagenesis demonstrated that this loop can even be shortened to one nucleotide without affecting the amino-

acylation efficiency of the tRNA-like structure of TYMV RNA (Mans et al. 1992).

It is worth mentioning here that model building also suggests that the loop length can be reduced to two and one nucleotide for L_S and L_D, respectively (Dumas et al. 1987; Major et al. 1991). The H-type pseudoknot in Fig. 2B, as first described in a number of plant viral RNAs, has appeared to be the more abundant and better studied one of these three types of "classic" pseudoknots (Westhof and Jaeger 1992), for which even detailed models exist (Dumas et al. 1987; Puglisi et al. 1990; Major et al. 1991). The characteristics and length requirements of loop L_P (Fig. 2C,D) have not been determined as yet and only a few examples of these H-type pseudoknots have been reported in the literature (Kössel et al. 1990; Michel et al. 1989; Pilipenko et al. 1992). Clearly, further experiments are needed to prove the existence of the latter, if they exist at all.

2.1.2 NMR

So far, no X-ray crystallographic studies have been reported for RNA pseudoknots, but other biophysical techniques like NMR have provided structural information about the H-type pseudoknot. NMR spectroscopy has
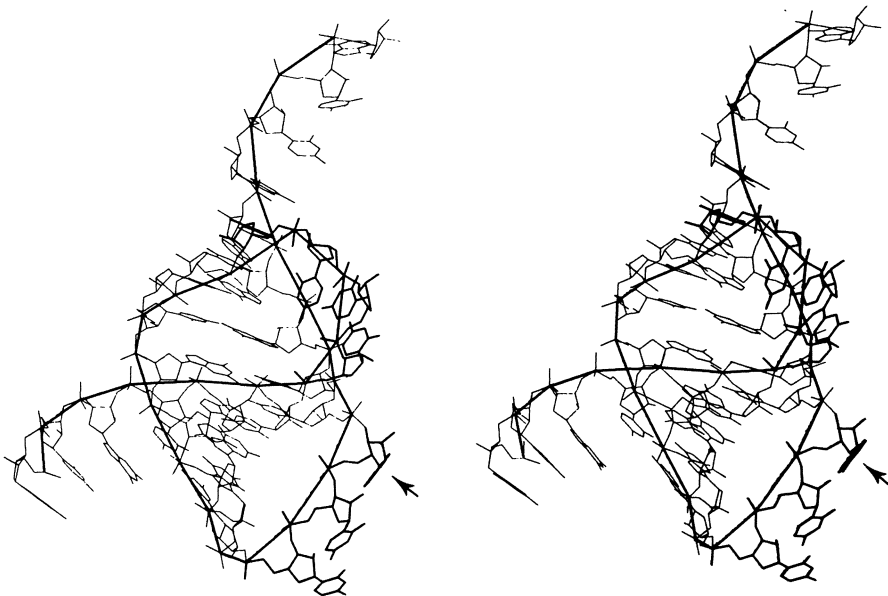

Fig. 4. Stereoscopic view of the pseudoknot in the aminoacyl acceptor arm of the tRNA-like structure of TYMV RNA as obtained by computer graphics (Dumas et al. 1987). The *arrows* indicate the A residue of the three-membered connecting loop crossing the shallow groove; see also Fig. 3A

been used successfully in the elucidation of the solution structure of small
RNA fragments, usually hairpins, in the range of 10–20 nucleotides (Heus
and Hilbers, this Vol.). Pseudoknot-containing fragments necessarily have
to be longer because they contain at least two stem regions, which makes it
difficult to solve the structure at an atomic level. Tinoco and coworkers,
using proton NMR and 2-D techniques, nevertheless succeeded to obtain
interesting results with a model H-type pseudoknot (Fig. 3D; Puglisi et al.
1990; Wyatt et al. 1990). The structure has two stem regions of three and
five base pairs, respectively, and as such resembles that of TYMV RNA
(Fig. 3A). Pseudoknot formation at 5 mM Mg^{2+} and room temperature
was confirmed by the observation of the eight expected nonexchangeable
iminoprotons. A nuclear Overhauser effect (NOE) between the imino-
protons of G3-C13 and C12-G22 (Fig. 3D) confirmed the coaxial stacking of
the two stems, although some irregularities at the junction were apparent at
the side of the quasi-continuous helix where the two connecting loops
emerge. No strong NOEs for the iminoprotons of G3 and G22 were observed,
while G22 was found to have mainly a 2'-endo pucker instead of the 3'-endo
pucker characteristic of an A-form helix. This may help to avoid the un-
favorable close proximity of the two loops at this site of the quasi-continuous
helix. When the pseudoknot was unfolded by raising the temperature or by
lowering the Mg^{2+} concentration, the 5' hairpin was the remaining structure,
in contrast to that expected from the number of base pairs in S1. This
illustrates the importance of the conformation of hairpin loops and con-
necting loops involved in pseudoknot formation (see below). No detailed
information about the loop nucleotides was available, except that U16 was
found to be stacked on residue C15.

These results are in agreement with the original proposal for the H-type
pseudoknot in TYMV RNA (Rietveld et al. 1982; Pleij et al. 1985), but it
remains to be seen whether all H-type pseudoknots from natural RNAs
behave in a similar way, although it would not be surprising if coaxial
stacking were the rule. The influence of stem size and, consequently, loop
size on the structure, however, needs to be examined further. Another
feature deserving closer analysis is whether the coaxial stacking is still
preserved when the third connecting loop is present, as indicated in Fig. 2A.
It will be interesting to see how many nucleotides can be inserted between
the junction of the two stems before coaxial stacking and quasi-continuous
helix are disrupted. This question is related to that of the conformation of
bulged loops in otherwise regular double helices.

2.1.3 Thermodynamic Stability

It is obvious that the folding of pseudoknots is mainly driven by the decrease
in enthalpy due to the formation of the base pairs in the second stem. This
contribution can be calculated from known nearest neighbor parameters,

thereby also assuming that the base pairs are regularly stacked at the junction of the two stems (Mans et al. 1992). The thermodynamic stability of the H-type pseudoknot, however, must also take into account the postive contribution due to the formation of the two linkers. The values for the latter are simply not known, but are a prerequisite for the prediction of pseudoknots with computer programs.

Thermodynamic parameters have been determined by Tinoco and co-workers for a set of pseudoknot structures based on the model oligonucleotide depicted in Fig. 3D (Puglisi et al. 1990; Wyatt et al. 1990). From optical melting studies it was concluded that this pseudoknot in its folded state was only about 2 kcal/mol stabler at 37 °C than either the 5' or 3' hairpin (unfolded state, cf. Fig. 2A with Fig. 3D). The measured enthalpy of pseudoknot formation was lower than expected, when compared to that of either the 5' or 3' hairpin. Apart from possible deviations of the stacking in the helical regions, we assume that especially the conformation of the connecting loop spanning the shallow groove may be largely responsible for this effect. Model building of the pseudoknot in the tRNA-like structure of TYMV RNA has shown that the three nucleotides in this connecting loop are pointing outwards into solvent and have a largely destacked conformation (Dumas et al. 1987). The situation for the other linker is probably enthalpic more favorable because it can position its residues in the major groove, even enabling extra interactions of the bases with the stem. Other factors which may be relevant to the relatively low stability of the H-type pseudoknot may be the sharp turns needed in relatively short linkers or the close proximity of the two linkers near the junction of stems S1 and S2. Anyhow, the length and base composition of these loops will be of utmost importance for the pseudoknot stability. This is confirmed by the findings of Wyatt et al. (1991) who studied the influence of loop size and loop composition on stability using the same two stem regions S1 and S2 as shown in Fig. 3D. Under their experimental conditions the pseudoknot was the only conformation present when the loop spanning the deep groove was reduced to three nucleotides or when the loop crossing the shallow groove was shortened to four nucleotides. Further shortening gave rise to the appearance of one of the two hairpin conformations. The equilibria between pseudoknot form and the alternate hairpins appeared to be dependent on base composition in a very subtle way. Note, for instance, that changing the connecting loops of the pseudoknot necessarily changes that of the hairpin loops at the same time. For a more detailed discussion, the reader is referred to Puglisi et al. (1991) and Pleij et al. (1992).

3 Classification and Nomenclature

RNA pseudoknots involve Watson-Crick base pairing of loop regions. With four different loop regions, the H-type pseudoknot is only 1 out of 14

possibilities (Fig. 1). Other types of loops may base pair with single-stranded regions or complementary regions of two loops may form double-stranded helices. In order to be able to name all these different types of pseudoknots we have introduced, in analogy to the H-type pseudoknot, the terms B- (bulge), I- (internal) and M-pseudoknot (multiple). Pseudoknots involving two loops are then designated by H-H, B-M, and so on. We have tried to develop a classification for RNA pseudoknots assuming that all pseudoknots are derived from the most elementary one, the H-type pseudoknot (Fig. 2A; Mans 1991). This pseudoknot is defined by two stems and maximally three loops. All other pseudoknots have one or more additional stems. This is illustrated for two simple cases in Fig. 5, where a bulge or internal loop base-pairs with a single-stranded region outside the loop. The relation with the H-type pseudoknot is easily seen when one of the three loop regions in Fig. 2, L1–3, is partially or entirely replaced by a hairpin. In Fig. 5A this insertion has taken place in L3 and in Fig. 5B in L2.

In this way any other pseudoknot can be constructed by introducing an increasing number of secondary structure elements in the original three linkers L1–3 of the generalized H-type pseudoknot. The most complicated type would be an M-M pseudoknot which is defined by at least seven stems (Mans 1991).
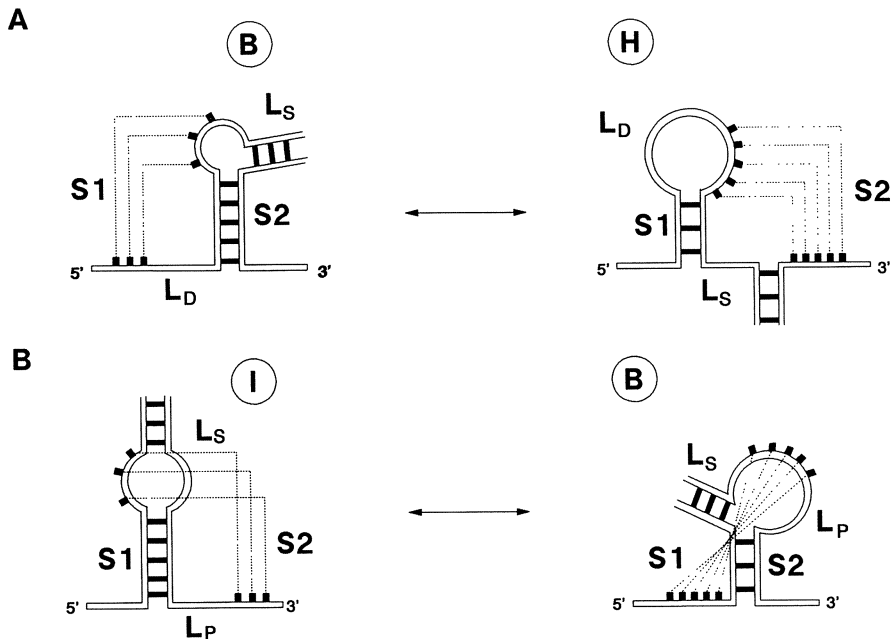

Fig. 5A,B. Alternative secondary structure presentations for **A** a B-type pseudoknot and **B** an I-type pseudoknot. For further details, see Fig. 2

Our approach to name and classify RNA pseudoknots led, however, to some problems. The introduction of a nomenclature based on the various loops gives rise to ambiguities in the orthodox structure presentation exemplified in Fig. 5. The B-type pseudoknot can also be described as an H-type pseudoknot upon redrawing the secondary structure. In that case the loop spanning the shallow groove simply harbors a hairpin. Similar reasoning is possible for the I-type pseudoknot (Fig. 5B). Both alternative representations, in fact, tells you only that a pseudoknot is formed but cannot tell you a priori which stem segment is folded first, let alone what the 3-D structure will look like. As illustrated in Fig. 6, this problem can even become worse for more complex pseudoknots (Mans 1991). As Westhof and Jaeger (1992) say: "Where there is no short loop over either groove (of the H-type pseudoknot), however, the notion of structural motif is obviously lost". These authors therefore consider the three types of H-pseudoknots in which stacking of the stem regions take place (Fig. 2) as the only meaningful structural motif, whereas all other pseudoknotted structures are designated as tertiary interactions. We believe, however, that the use of the word "tertiary interaction" for a pseudoknot-forming stem can lead to some ambiguities as well in that it is not always clear which stem has to be considered as such. For instance, is stem S1 or stem S2 the tertiary interaction in Fig. 2A?

In conclusion, the introduction of a classification of RNA pseudoknots does not seem to be very useful at the moment, the main objection being that it does not contribute to our understanding of the spatial folding of RNA.
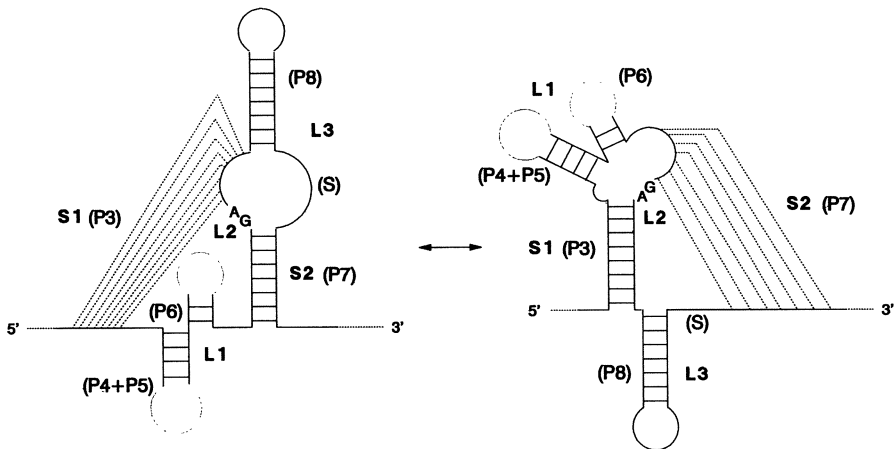

Fig. 6. Schematic presentation of part of the secondary structure of a group I intron. The structure at the *left* illustrates the I-type pseudoknot formed by stem *P3*. The structure at the *right* is the same secondary structure in which the pseudoknot is depicted as M-type formed by *P7*

It remains still worthwhile, however, to explore a few simple cases more closely, either experimentally or by model building, to see how they are (potentially) folded and whether some rules can be derived. The third stem segment formed upon basepairing of a bulge loop can stack on either of the two other stems. We noticed this problem of alternative coaxial stackings already in the case of a B-type pseudoknot when we proposed a model for the tRNA-like structure of tobacco mosaic virus (TMV) RNA (Rietveld et al. 1984). Similar stereochemical questions can be raised for the pseudo-knot in the self-cleavage site of hepatitis delta virus RNA (Perrotta and Been 1991) or the pseudoknot involving a hairpin loop and a bulge loop in the so-called 530 region of the 16S rRNA of *E. coli* (Woese and Gutell 1989).

4 Real Knots

A recurrent theme related to the concept of RNA pseudoknots is the possibility of creating topologically real knots in RNA. Many authors hold that real knots in RNA are formed when each of the two stem regions in a pseudoknot contains a full turn of the A-form double helix, i.e., 11 base pairs or more (Studnicka et al. 1978; Westhof and Jaeger 1992). This more theoretical aspect of RNA pseudoknot formation has become relevant since pseudoknots have been proposed in retroviral and coronaviral RNAs, playing a role in ribosomal frameshifting, which have more than one full turn of a helix (Ten Dam et al. 1990). An H-type pseudoknot will be converted to a real knot if both the 3′ and the 5′ end of the RNA pass through the two connecting loops. The resulting structure is a simple "overhand knot", upon denaturation (Puglisi et al. 1991), but looks like a "figure eight knot" as long as basepairing is maintained.

The probability that real knots are formed in the cell is in our view extremely low (Pleij et al. 1992). Not only two long stem regions are required, but also rather long loops. We estimate that the loop of the hairpin which is folded first must be composed of 25–30 nucleotides or even more, to allow the dangling strand to pass through the hole left by the connecting loop and the newly formed stem. Moreover, this free dangling end should be rather short, because it is highly unlikely that a chain segment of 10 or even 100 nucleotides will move freely in its entirety through this hole. The formation of a real knot in vivo will be the more unlikely because the bulky RNA polymerase will also prevent the growing RNA chain to pass through the hole between linker and second stem. In other words, real knot formation is only feasible at the ends of an RNA chain and when large hairpin loops come into play, conditions which are very rare in nature or are easily avoided. It remains to be seen, however, whether in vitro, conditions can be found which allow real knot formation.

5 Prediction of RNA Pseudoknots

There are a variety of methods at hand to trace pseudoknots in an RNA molecule. These methods are basically the same as those used in studying any other double helical region in RNA and include enzymatic digestion, chemical modification, covariation search, and mutational analysis (see also Pleij and Bosch 1989). Computer programs for the prediction of orthodox RNA secondary structures, mainly based on the prediction of the lowest free energy, have proven to be very useful (Zuker 1989). The dynamic programming methods used do not allow one to include the prediction of pseudoknots. The algorithms would become very complex, also because the final result is no longer a simple sum of contributions of separate structural elements.

Nevertheless, two programs have been published recently which are able to predict pseudoknots in an otherwise classic secondary structure (Abrahams et al. 1990; Gultyaev 1991). Both programs are based on a simulation of RNA folding in which the next most stable stem is added stepwise to the growing structure. At each step the possibility of pseudoknot formation is considered. The final structure obtained does not necessarily have the lowest free energy and suboptimal structures have not produced so far. On the other hand, because of their heuristic character, these programs need less computer time and can be run on personal computers. A major obstacle in the development of any computer-aided prediction of pseudoknots is the lack of energy parameters for the connecting loop elements. Abrahams et al. (1990) used a single value of 4.2 kcal/mol for all connecting loops, irrespective of size or base composition. Despite this oversimplification, they were reasonably successful in predicting H-type pseudoknots. Gultyaev (1991) has improved the predicting power further by introducing the principle of sequential folding, i.e., the addition of stretches of RNA to the growing structure starting from the 5′ end, thereby mimicking the direction of RNA synthesis in the cell. It is clear that further progress in this field will mainly depend on the availability of more reliable energy parameters. The study of Wyatt et al. (1990), in which the influence of loop size on pseudoknot stability was investigated, is certainly a big step towards this aim.

6 Functional Aspects

Over the last few years, RNA pseudoknots have been found in almost every class of RNA (Pleij 1990; Westhof and Jaeger 1992). In many cases these pseudoknots seem to be a suitable way of bringing sequences together into a compact functional tertiary RNA structure, a role which sometimes is played equally well by a standard secondary structure motif (Haas et al. 1991). There are a few examples, however, in which the pseudoknot structure itself

has a special functional role, e.g., by binding a specific protein due its unique stereochemical properties. This functional significance of RNA pseudoknots will be summarized below.

6.1 Messenger RNAs

Translational regulation is occasionally mediated by pseudoknots occurring near the ribosome binding site of prokaryotic messengers. In that case, the binding of proteins to pseudoknotted structures in their cognate mRNAs autoregulates their expression. A very interesting example is the mRNA for gene 32 of bacteriophage T4 (McPheeters et al. 1988). Gene 32 protein binds to single-stranded nucleic acids in a cooperative manner. The leader of its own mRNA is largely single-stranded, but the gene 32 protein was shown to start binding at low concentrations at the very 5' end where a H-type pseudoknot is located. When the concentration of gene 32 protein increases, it covers the leader downstream of the pseudoknot up to the ribosome binding site, thereby repressing the translation of its own mRNA. The pseudoknot, whose existence was supported by covariations in related phages, has a single A residue crossing the deep groove and five residues crossing the shallow groove. We have suggested earlier that the initial, strong binding of gene 32 protein to this pseudoknot may be related to the peculiar properties of the connecting loop spanning the shallow groove (Pleij 1990). As explained above, this loop may have a largely exposed and destacked character. In this way this loop may have a preformed conformation which enables binding of the gene 32 protein with high affinity.

Autogenous regulation is a recurrent mechanism among prokaryotic messengers. It has been shown for two different messengers from E. coli, coding for ribosomal proteins, that a pseudoknot structure plays a crucial role in this process. The ribosome binding site of the messenger for protein S15 is located within a pseudoknot structure and it is proposed that binding of S15 shifts the equilibrium between the pseudoknot structure and an alternative hairpin by stabilizing the pseudoknot structure, thereby preventing the formation of the initiation complex (Philippe et al. 1990; Portier et al. 1990).

Protein S4 is one of the four proteins encoded by the α-operon of E. coli. It is involved in translational repression by binding to a stretch of more than 100 nucleotides of the leader, including the initiation codon and a short stretch of the coding region. This region harbors a complex pseudoknot (a "double" pseudoknot) whose structure is supported by site-directed muta-genesis and structure mapping experiments (Deckman and Draper 1987). It was postulated that the ribosome and the protein S4 recognize different domains of the pseudoknot and that repressor binding induces an allosteric conformational change in the ribosome binding site (Tang and Draper 1990).

Pseudoknots have been shown to occur in coding regions as well. In this case they function in ribosomal frameshifting and translational read-through (for a review, see Atkins et al. 1990). In some viruses the presence of a pseudoknot downstream of a "slippery" heptanucleotide sequence or downstream of a leaky stop codon is essential for efficient translational suppression. In retroviruses the *pol* gene, encoding three essential proteins, is expressed as one large fusion protein together with the core or *gag* proteins, encoded upstream. The majority of the retroviruses shift the reading frame in the -1 direction at a specific heptanucleotide sequence in the overlap region. The sequence motif is XXXYYYN, where X is any base and YYYN is AAAC, UUUA or AAAU. Jacks et al. (1988) proposed a simultaneous slippage of the two tRNAs in the A- and P-site on the ribosome, respectively. Besides this slippery sequence, a second frameshift signal (or stimulator; Atkins et al. 1990) is needed. Brierley et al. (1989) obtained strong evidence for the involvement of a pseudoknot structure six nucleotides downstream of the heptanucleotide sequence in the orf 1A-orf 1B overlap of the coronaviral IBV RNA. Mutational analysis demonstrated that the two stems forming the H-type pseudoknot were both needed to obtain efficient frameshifting. In a later study these authors showed that the base pairs in the stems could be changed in others without influencing frameshifting. Moreover, the downstream pseudoknot could not be replaced by a hairpin of the same length (Brierley et al. 1991). This indicates that some specific structural feature of the pseudoknot is responsible for this phenomenon. Pseudoknotted structures could be folded downstream of a (potential) slippery sequence in the majority of the retroviral RNAs, but also in coronaviral, luteoviral, and yeast viral RNAs (Brierley et al. 1989; Ten Dam et al. 1990). Pseudoknot-dependent frameshifting has now been proven in a few cases (e.g. Dinman et al. 1991; Chamorro et al. 1992; Morikawa and Bishop 1992). Figure 3C shows the pseudoknot in the retroviral SRV-1 RNA, which is needed for efficient frameshifting (E. Ten Dam and CWAP, unpubl. observ.).

In type C retroviruses the *gag* and *pol* genes are in frame, but separated by a single UAG stop codon. The downstream *pol* reading frame is expressed by suppression of the amber stop codon. Ten Dam et al. (1990) proposed on the basis of computer prediction that also a pseudoknot structure is present downstream of the leaky amber codon. An extensive mutational analysis by Wills et al. (1991) provided evidence that the presence of an H-type pseudoknot eight nucleotides downstream of the UAG codon is essential for readthrough. To date it is not known which feature of the pseudoknot is responsible for both the frameshifting and the readthrough process, but the current idea is that it helps the ribosome in stalling at the slippery heptanucleotide sequence or leaky stop codon (Brierley et al. 1991; Tsuchihashi 1991; Wills et al. 1991). It may well be that the pseudoknot structure is more resistant to a helicase or unwinding activity than a simple hairpin.

6.2 Ribosomal RNAs

There are at least three different pseudoknot structures present in the small subunit ribosomal RNA (Fig. 7). They are supported by phylogenetic comparisons and site-directed mutagenesis (Stern et al. 1988; Woese and Gutell 1989; Powers and Noller 1991). The first one plays a literally central role in the tertiary structure of 16S rRNA by connecting the hairpin at the 5' end to a region some 900 nucleotides downstream. This pseudoknot served as the starting point for modeling the tertiary structure and brings three major domains of the secondary structure together (Stern et al. 1988). Two groups proposed this central pseudoknot to function as a conformational switch during or after the initiation step of protein synthesis (Kössel et al. 1990; Leclerc and Brakier-Gingras 1991). In both models the loop of the 5' hairpin is involved in alternate pairings: either with another complementary sequence near the anti-Shine-Dalgarno sequence at the 3' terminus (Kössel et al. 1990) or with a sequence just 5' of the original base-pairing stretch in the 900 region (Leclerc and Brakier-Gingras 1991). Both models, however, lack sufficient support from phylogenetic comparisons and seem to be
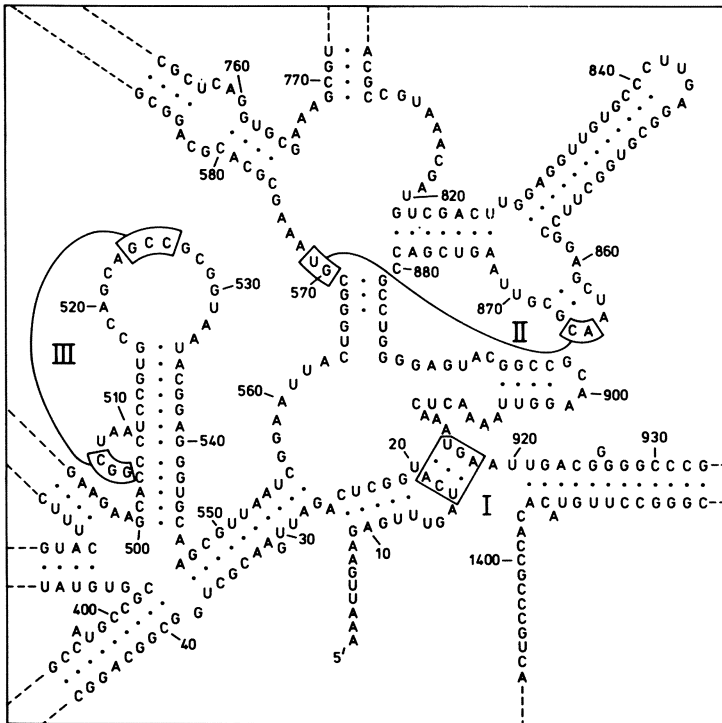

Fig. 7. The central part of the 16S ribosomal RNA of *E. coli*. The Roman numbers *I–III* indicate the pseudoknots. The numbering is the same as for the entire RNA

mutually exclusive. Although there is some evidence for a conformational equilibrium in the 900 region (Allen and Noller 1989), both proposals have to be considered as tentative and certainly needs further verification.

A second pseudoknot is formed by the interaction of G570 and U571 with C866 and A865 (*E. coli* numbering). These two base pairs connect a hairpin loop with a multiple loop (Stern et al. 1988).

The third one involves the strongly conserved 530-hairpin loop, which base pairs with three nucleotides of the bulge loop in the same hairpin (Woese and Gutell 1989; Fig. 7). It is the only example to date of an H-B-type pseudoknot (see Sect. 3). Powers and Noller (1991) have performed a mutational analysis of this pseudoknot and have tested the altered ribosomal RNAs in vivo and in vitro. Their experiments strongly support both the existence and the functional importance of this interaction. Moreover, the behavior of some mutants, having a G-U or A-U base pair instead of a G-C pair in the wild type, suggested that this pseudoknot is involved in the binding of the antibiotic streptomycin to the 16S rRNA, though in an indirect way.

Pseudoknot structures have been proposed in the large ribosomal subunit as well (see Gutell and Woese 1990; Neefs and De Wachter 1990).

6.3 Viral RNAs

As discussed above, coding regions of viral RNAs contain pseudoknots which stimulate ribosomal frameshifting and translational readthrough (see Sect. 6.1). Besides these pseudoknots, many others have been detected in the noncoding regions. First, they appear to be essential for the folding of tRNA-like structures occurring at the 3′ end of various plant viral RNAs. It was the analysis of the 3′ terminus of TYMV RNA which led to the discovery of the H-type pseudoknot as described in Section 2.1 (Figs. 3A and 4). Plant viral tRNA-like structures and their aminoacylation are fully dependent on the pseudoknot structure (Dreher and Hall 1988; Mans et al. 1992), but why just pseudoknots are used to fold these viral 3′ ends remains obscure as yet. We have recently reviewed this subject extensively elsewhere (Mans et al. 1991). Transfer RNA-like structures have been reported recently to occur in picornaviral RNAs downstream of the poly (A) tail (Pilipenko et al. 1992).

Many nonpolyadenylated plant virus RNAs have one or more pseudo-knotted structures, often of the H-type, just downstream of the coding region (Pleij et al. 1987). In tobacco mosaic virus (TMV) RNA there is a set of three consecutive pseudoknots located between the tRNA-like structure and the coding region (Fig. 3B; Van Belkum et al. 1985). A similar extended quasi-continuous double helix or stalk has been found in various other plant viral RNAs (Pleij et al. 1987). Interestingly, when this domain is linked to messenger RNAs, it can fulfill the same role as a poly (A) tail. It protects a

messenger against breakdown and increases the efficiency of translation by two orders of magnitude (Gallie and Walbot 1990). It is likely that these pseudoknots interact with the 5' leader during initiation either directly or mediated by one or more proteins. It may also demonstrate that 3' terminal pseudoknots must not necessarily play a role in the replication of viral RNA (Westhof and Jaeger 1992).

The 5' noncoding region of some animal virus RNAs was shown to contain consecutive pseudoknots as well. In picornaviral RNAs, like that of foot and mouth disease (FMDV) virus, hepatitis A virus, or encephalomyocarditis (EMCV) virus, the pseudoknots appear to be associated with a single-stranded region or poly (C) stretch, as described above for the gene 32 protein messenger (Clarke et al. 1987; Brown et al. 1991; Duke et al. 1992). The function of these pseudoknots is not known however.

6.4 Catalytic RNAs

Pseudoknots are key structural elements in folding ribozymes in their active conformation. Two pseudoknots have been proposed in the catalytic RNA of RNase P (James et al. 1988; Haas et al. 1991), supported, as usual, by mutational analyses and phylogenetic comparisons. A long-range interaction between a multiple loop and an internal loop creates a stem of eight base pairs, including a bulged U residue and brings a number of conserved residues into proximity. Although the catalytic activity diminishes appreciably when sequences involved in the pseudoknot formation are deleted, its exact role in the folding of the substrate binding pocket remains to be established.

The second pseudoknot is formed by the interaction of a bulge loop and a hairpin loop, though not in all eubacteria. This pseudoknot, occurring in E. coli, is replaced in B. subtilis by an additional hairpin which probably plays an equivalent architectural role (Haas et al. 1991). It illustrates that a pseudoknot is not a unique way to solve a folding problem in RNA.

Another, well-known example of pseudoknotting in the core of a catalytic RNA is found in the self-splicing group I introns. The interaction involves base pairing of one strand of an internal loop, located between stem P7 and P8 and a single-stranded region 5' of this hairpin (Fig. 6). The resulting stem region (P3) was modeled as a coaxial stack with P7 and P8 (Kim and Cech 1987). As such it contributes to building the compact, functionally important core structure, where the guanosine cofactor and the splicing substrate are bound. A detailed model of the three-dimensional structure was proposed recently by Michel and Westhof (1990). The latter group has also reported a second, peripheral pseudoknot in a subclass of the group I introns. It connects a hairpin loop with an internal loop (Jaeger et al. 1991).

Self-cleavage reactions occur in viral or virus-related RNAs. Both the genomic and the anti-genomic RNAs of hepatitis delta virus contain a self-cleavage site, comprising some 85 nucleotides (Perrotta and Been 1991).

The pseudoknot in both cases is in fact of the B-type. As discussed above (Sect. 3), this gives rise to two mutually exclusive stacking possibilities. Site-directed mutagenesis studies indicated that the pseudoknot stabilizes the tertiary structure but is in fact not essential for its formation (Perrotta and Been 1991).

In the case of the self-cleaving hammerhead structure occurring in the satellite RNA associated with barley yellow dwarf virus (sBYDV), a pseudoknot has been described, connecting one of the hairpin loops of the hammerhead structure with an internal loop outside this "core". Interestingly, disruption of the pseudoknot by site-directed mutagenesis appears to enhance the cleavage rate rather than to diminish it (Miller and Silver 1991).

7 Concluding Remarks

An RNA pseudoknot is nowadays a common structural motif in almost all classes of RNA. One may say that this type of base pairing is simply equivalent to all other stem regions in an RNA structure. Problems in visualizing pseudoknotted structures may be due to the practice of presenting the structure of RNA in two dimensions, which has led to the notion of pseudoknots as an extra or peculiar folding principle. The nonunique character is exemplified in the case of one of the pseudoknots in the catalytic RNA of RNase P, where a pseudoknot is replaced by a hairpin in another organism.

There are a few cases, however, where the structural properties of pseudoknots are both unique and directly related to a function. It may well be that a short connecting loop crossing the shallow groove of the H-type pseudoknot is one of the few ways to expose a stretch of (hydrophobic) base residues. The latter may be the basis for the high affinity of the single strand-specific gene 32 protein for an H-type pseudoknot and it could explain the apparently low stability of H-type pseudoknots. This low or even marginal stability makes the possibility that pseudoknots function as conformational switches very attractive; this seems to be realized in ribosomal RNAs and in some prokaryotic messengers.

References

Abrahams JP, van der Berg M, van Batenburg E, Pleij CWA (1990) Prediction of RNA secondary structure, including pseudoknotting, by computer simulation. Nucleic Acids Res 18:3035–3044
Allen PN, Noller HF (1989) Mutations in ribosomal proteins S4 and S12 influence the higher order structure of 16S ribosomal RNA. J Mol Biol 208:457–468
Atkins JF, Weiss RB, Gesteland RF (1990) Ribosome gymnastics – degree of difficulty 9.5, style 10.0. Cell 62:413–423
Brierley I, Digard P, Inglis SC (1989) Characterization of an efficient coronavirus frame-shifting signal: requirement for an RNA pseudoknot. Cell 57:537–547

Brierley I, Rolley NJ, Jenner AJ, Inglis SC (1991) Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. J Mol Biol 229:889–902

Brown EA, Day SP, Jansen RW, Lemon SM (1991) THe 5′ nontranslated region of hepatitis A virus RNA: secondary structure and elements required for translation in vitro. J Virol 65:5828–5838

Chamorro M, Parkin N, Varmus HE (1992) An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. Proc Natl Acad Sci USA 89:713–717

Chastain N, Tinoco I Jr (1991) Structural elements in RNA. Prog Nucleic Acids Res Mol Biol 41:131–177

Clarke BE, Brown AL, Currey KM, Newton SE, Rowlands DJ, Carroll AR (1987) Potential secondary and tertiary structure in the genomic RNA of foot and mouth disease virus. Nucleic Acids Res 15:7067–7078

Deckman IC, Draper DE (1987) S4-αmRNA translation regulation complex II. Secondary structures of the RNA regulatory site in the presence and absence of S4. J Mol Biol 196:323–332

Dinman JD, Icho T, Wickner RB (1991) A −1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. Proc Natl Acad Sci USA 88:174–178

Dreher TW, Hall TC (1988) Mutational analysis of the tRNA mimicry of brome mosaic virus RNA. The sequence and structural requirements for aminoacylation and 3′-adenylation. J Mol Biol 201:41–55

Duke GM, Hoffman M, Palmenberg AC (1992) Sequence and structural elements that contribute to efficient encephalomyocarditis viral RNA translation. J Virol 66:1602–1609

Dumas P, Moras D, Florentz C, Giege R, Verlaan P, van Belkum A, Pleij CWA (1987) 3D-graphic modeling of the tRNA-like 3′ end of turnip yellow mosaic virus RNA: structural and functional implications. J Biomol Struct Dynam 4:707–728

Gallie DR, Walbot V (1990) RNA pseudoknot domain of tobacco mosaic virus can functionally substitute for a poly(A) tail in plant and animal cells. Genes Dev 4:1149–1157

Gultyaev AP (1991) The computer simulation of RNA folding involving pseudoknot formation. Nucleic Acids Res 19:2489–2494

Gutell RR, Woese CR (1990) Higher order structural elements in ribosomal RNAs: pseudoknots and the use of noncanonical pairs. Proc Natl Acad Sci USA 87:663–667

Haas ES, Morse DP, Brown JW, Schmidt FJ, Pace NR (1991) Long-range structure in ribonuclease P RNA. Science 254:853–856

Jacks T, Madhani HD, Masiarz FR, Varmus HE (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. Cell 55:447–458

Jaeger L, Westhof E, Michel F (1991) Function of P11, a tertiary base pairing in self-splicing introns of subgroup 1A. J Mol Biol 221:1153–1164

James BD, Olsen GJ, Liu J, Pace NR (1988) The secondary structure of ribonuclease P RNA, the catalytic element of a ribonucleoprotein enzyme. Cell 52:19–26

Kim S-H, Cech TR (1987) Three-dimensional model of the active site of the self-splicing rRNA precursor of Tetrahymena. Proc Natl Acad Sci USA 84:8788–8792

Kössel H, Hoch B, Zeltz P (1990) Alternative base pairing between 5′- and 3′-terminal sequences of small subunit RNA may provide the basis of a conformational switch of the small ribosomal subunit. Nucleic Acids Res 18:4083–4088

Leclerc D, Brakier-Gingras L (1991) A conformational switch involving the 915 region of Escherichia coli 16S ribosomal RNA. FEBS Lett 279:171–174

Major F, Turcotte M, Gautheret D, Lapalme G, Fillion E, Cedergren R (1991) The combination of symbolic and numerical computation for three-dimensional modeling of RNA. Science 253:1255–1260

Mans RMW (1991) Functional properties of plant viral RNA pseudoknots. Thesis, University of Leiden, The Netherlands

Mans RMW, Pleij CWA, Bosch L (1991) Transfer RNA-like structures: structure, function and evolutionary significance. Eur J Biochem 201:303–324

Mans R, Van Steeg MH, Verlaan PWG, Pleij CWA, Bosch L (1992) Mutational analysis of the pseudoknot in the tRNA-like structure of turnip yellow mosaic virus RNA. Aminoacylation efficiency and RNA pseudoknot stability. J Mol Biol 223:221–232

McPheeters DS, Stormo GD, Gold L (1988) Autogenous regulatory site on the bacteriophage T4 gene 32 messenger RNA. J Mol Biol 201:517–535

Michel F, Westhof E (1990) Modelling of the three-dimensdional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol 216:585–610

Michel F, Umesono K, Ozeki H (1989) Comparative and functional anatomy of group II catalytic introns – a review. Gene 82:5–30

Miller WA, Silver SL (1991) Alternative tertiary structure attenuates self-cleavage of the ribozyme in the satellite RNA of barley yellow dwarf virus. Nucleic Acids Res 19:5313–5320

Morikawa S, Bishop DHL (1992) Identification and analysis of the gag-pol ribosomal frameshift site of feline immunodeficiency virus. Virology 186:389–397

Neefs J-M, De Wachter R (1990) A proposal for the secondary structure of a variable area of eukaryotic small ribosomal subunit RNA involving the existence of a pseudoknot. Nucleic Acids Res 18:5695–5704

Ninio J (1971) Properties of nucleic acid representations I. Topology. Biochimie 53:485–494

Noller HF, Hoffarth V, Zimniak L (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. Science 256:1416–1419

Perrotta AT, Been MD (1991) A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. Nature 350:434–436

Philippe C, Portier C, Mougel M, Grunberg-Manago M, Ebel JP, Ehresmann B, Ehresmann C (1990) Target site of Escherichia coli ribosomal protein S15 on its messenger RNA. J Mol Biol 211:415–426

Pilipenko EV, Maslova SV, Sinyakov AN, Agol VI (1992) Towards identification of cis-acting elements involved in the replication of enterovirus and rhinovirus RNAs: a proposal for the existence of tRNA-like terminal structures. Nucleic Acids Res 20:1739–1745

Pleij CWA (1990) Pseudoknots: a new motif in the RNA game. Trends Biochem Sci 15:143–147

Pleij CWA, Bosch L (1989) RNA pseudoknots: structure, detection and prediction. Methods Enzymol 180a:289–303

Pleij CWA, Rietveld K, Bosch L (1985) A new principle of RNA folding based on pseudoknotting. Nucleic Acids Res 13:1717–1731

Pleij CWA, Abrahams JP, Van Belkum A, Rietveld K, Bosch L (1987) The spatial folding of the 3′ noncoding region of aminoacylatable plant viral RNAs. In: Brinton MA, Rueckert R (eds) Positive strand RNA viruses. UCLA Symp on Mol Cell Biol, Ser 54. Liss, New York, pp 299–316

Pleij CWA, Mans RMW, Verlaan P, Bosch L (1992) H-type RNA pseudoknots: structure, mutational analysis and prediction. In: Sarma RH, Sarma MH (eds) Structure and function, vol 1. Nucleic acids. Adenine Press, pp 261–275

Portier C, Philippe C, Dondon L, Grunberg-Manago M, Ebel JP, Ehresmann B, Ehresmann C (1990) Translational control of ribosomal protein S15. Biochim Biophys Acta 1050:328–336

Powers T, Noller HF (1991) A functional pseudoknot in 16S ribosomal RNA. EMBO J 10:2203–2214

Puglisi JD, Wyatt JR, Tinoco I Jr (1990) Conformation of an RNA pseudoknot. J Mol Biol 214:437–453

Puglisi JD, Wyatt JR, Tinoco I Jr (1991) RNA pseudoknots. Acc Chem Res 24:152–158

Richards EG (1969) 5S RNA. An analysis of possible base pairing schemes. Eur J Biochem 10:36–42

Rietveld K, van Poelgeest R, Pleij CWA, van Boom JH, Bosch L (1982) The tRNA-like structure at the 3′ terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. Nucleic Acids Res 10:1929–1946

Rietveld K, Linscooten K, Pleij CWA, Bosch L (1984) The three-dimensional folding of the tRNA-like structure of tobacco mosaic virus RNA. A new building principle applied twice. EMBO J 3:2613–2619

Schimmel P (1989) RNA pseudoknots that interact with components of the translation apparatus. Cell 58:9–12

Stern S, Weiser B, Noller HF (1988) Model for the three-dimensional folding of 16S ribosomal RNA. J Mol Biol 204:447–481

Studnicka GM, Rahn GM, Cummings IW, Salser WA (1978) Computer method for predicting the secondary structure of single-stranded RNA. Nucleic Acids Res 5:3365–3387

Tang CK, Draper DE (1990) Evidence for allosteric coupling between the ribosome and repressor binding sites of a translationally regulated mRNA. Biochemistry 29:4434–4439

Ten Dam EB, Pleij CWA, Bosch L (1990) RNA pseudoknots and translational frame-shifting on retroviral, coronaviral and luteoviral RNAs. Virus Genes 4:121–136

Tinoco I Jr, Puglisi JD, Wyatt JR (1990) RNA folding. In: Eckstein F, Lilley DMJ (eds) Nucleic Acids and Molecular Biology, vol 4. Springer, Berlin Heidelberg New York, pp 205–226

Tsuchihashi Z (1991) Translational frameshifting in the *Escherichia coli* dnaX gene in vitro. Nucleic Acids Res 19:2457–2462

Van Belkum A, Abrahams JP, Pleij CWA, Bosch L (1985) Five pseudoknots are present at the 204 nucleotides long 3' noncoding region of tobacco mosaic virus RNA. Nucleic Acids Res 13:7673–7686

Van Belkum A, Bingkun J, Rietveld K, Pleij CWA, Bosch L (1987) Structural similarities among valine-accepting tRNA-like structures in tymoviral and elongator tRNAs. Biochemistry 26:1144–1151

Westhof E, Jaeger L (1992) RNA pseudoknots. Curr Opinion Struct Biol 2:327–333

Wills N, Gesteland RF, Atkins JF (1991) Evidence that a downstream pseudoknot is required for translational readthrough of the Moloney murine leukemia virus *gag* stop codon. Proc Natl Acad Sci USA 88:6991–6995

Woese CR, Gutell RR (1989) Evidence for several higher order structural elements in ribosomal RNA. Proc Natl Acad Sci USA 86:3119–3122

Wyatt JR, Puglisi JD, Tinoco I Jr (1990) RNA pseudoknots, stability and loop size requirements. J Mol Biol 214:455–470

Zuker M (1989) On finding all suboptimal foldings of an RNA molecule. Science 244:48–52