Biochemistry



Volume 31, Number 47 December 1, 1992

*Perspectives in Biochemistry*

# Structural and Functional Aspects of RNA Pseudoknots

Edwin ten Dam,[‡] Kees Pleij,[‡] and David Draper*,[§]

*Department of Biochemistry, Gorlaeus Laboratories, University of Leiden, P.O. Box 9502, 2300 RA Leiden, The Netherlands, and Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218-2685*

*Received July 24, 1992; Revised Manuscript Received September 9, 1992*



A full understanding of the function of an RNA molecule requires a detailed knowledge of its spatial folding. In recent years the pseudoknot fold has been added to the possible structures an RNA may adopt. Starting from a fully denatured RNA chain, the first structural element formed upon base pairing is the well-known hairpin or stem-loop structure (Figure 1A). Addition of a second stem region, which can be considered as the next step in a renaturation process, is possible in three ways (Figure 1B; Richards, 1969, Westhof & Jaeger, 1992). The first two give rise to a "classical" secondary structure with the formation of interior loops, bulge loops, and multibranched loops [see also Zuker and Stiegler (1981)]. The third possibility, in which the hairpin loop basepairs with a complementary sequence outside that loop, gives rise to what is now called a pseudoknot. Historically the term pseudoknot was first proposed, to the best of our knowledge, by Studnicka and co-workers in the late seventies upon developing algorithms for the prediction of RNA secondary structures (Studnicka et al., 1978). These authors distinguished pseudoknots from true knots on a theoretical basis; the latter have not been observed in natural RNAs so far. The first author who described the possibility of base pairing of loops was Richards (1969). Somewhat later Ninio (1971) explored more fully the possibility of loop–loop interactions, although he did not use the term pseudoknot. In fact, throughout the seventies, various proposals were made for secondary structures which according to current terminology contain pseudoknots (Weidner et al., 1977; Kearns et al., 1974), and even in tRNA a pseudoknot structure was proposed on the basis of NMR data (Kearns & Wong, 1974).

It was not until 1982 that pseudoknots were put on an experimental basis when chemical modification, enzymatic digestion, and sequence comparisons led Rietveld and co-workers to propose a pseudoknot structure in order to explain

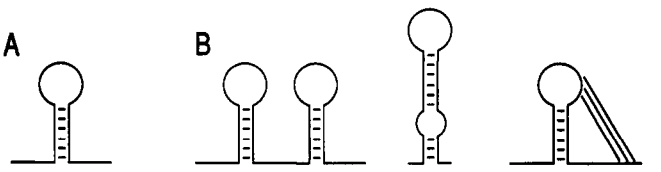


FIGURE 1: Base pair formation in RNA. The first secondary structure element formed from a fully denatured RNA is a hairpin (A). Addition of a second stem can occur in three ways to form (B, from left to right) a separate hairpin, an extension of the hairpin ("classical" secondary structures), or, by involving the loop residues, a pseudoknot.

the tRNA-like properties of the 3′ end of turnip yellow mosaic virus (TYMV) RNA (Rietveld et al., 1982). Presently, the number of pseudoknotted structures described in the literature has grown to such an extent that they seem to occur in virtually all classes of RNA, where they are involved in a number of important functions or are present at functionally strategic positions [for earlier reviews, see Pleij and Bosch (1989), Schimmel (1989), Tinoco et al. (1990), Pleij (1990), and Westhof and Jaeger (1992)]. In this perspective, we first summarized what is known of pseudoknot structure and then discuss the occurrences and possible functional roles of pseudoknots in different classes of RNA.

## *Definitions and Classification*

A pseudoknot is defined as a structure element of RNA formed upon standard base pairing of nucleotides of a loop region with residues outside that loop. A loop region is any single-stranded region in a secondary structure constrained by one or more stem regions (see above). As a consequence, a pseudoknot is always defined by (at least) two stems (Pleij et al., 1985). Which stem is formed first is not relevant for the definition but may have consequences for the natural folding of the RNA.

Figure 2 presents three alternative ways to illustrate the most basic form of a pseudoknot [see also Figure 1B). In a circular representation of an RNA molecule, a pseudoknot

* Correspondence should be addressed to this author.
‡ University of Leiden.
§ Johns Hopkins University.

0006-2960/92/0431-11665$03.00/0 

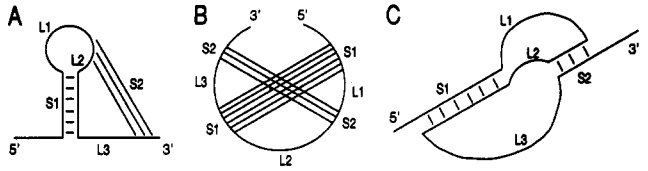


FIGURE 2: Three alternative ways to represent a pseudoknot. (A) Starting from a hairpin, a pseudoknot is formed upon base pairing of loop nucleotides with bases downstream of the hairpin. (B) If the RNA is represented in a circular form, pseudoknotting leads to crossing chords. In panel C, the two stems are aligned. Stem regions (S1, S2) and loops (L1, L2, and L3) are numbered from the 5′ end.

gives rise to crossing chords [see also Zuker and Stiegler (1981)]. Three different single-stranded regions within the pseudoknot, called connecting loops or linkers, become apparent (Mans, 1991; Westhof & Jaeger, 1992). Note that these loops may harbor secondary structure themselves and that the orientation of the stem regions or the conformation of the loops is not relevant here. However, in the case that one of the three loops is reduced to a length of zero nucleotides, some special structural features arise, which are important to the three-dimensional structure of RNA. If L2 is absent, stem regions S1 and S2 become adjacent and can be stacked coaxially to form a quasicontinuous double helix (Pleij et al., 1985) as is the case in the tRNA-like structure of TYMV RNA (see below). This defined orientation of both stems gives rise to what was termed the H(airpin)-type pseudoknot (Pleij & Bosch, 1989), which later was confirmed by NMR data from a model oligonucleotide (Puglisi et al., 1990). It is this type of pseudoknot that has been described in literature most often. Other stacking possibilities arise when either L1 or L3 is shortened to zero nucleotides. In both cases one of the resulting connecting loops is different from those in the classical H-type pseudoknot [see also Abrahams et al. (1990) and Westhof and Jaeger (1992)].

Although theoretically there are 14 different pseudoknot-forming combinations of interactions between loops and single-stranded regions (either unconstrained stretches or loops), they can all be reduced to the same basic pattern shown in Figure 2 (Mans, 1991; Westhof & Jaeger, 1992). L1, L2, and L3 can be very large or harbor secondary structure or even pseudoknots on their own. If, for instance, L1 or L3 is replaced in its entirety by a hairpin, the resulting structure can be considered as a pseudoknot formed on base pairing of a bulge loop with a region outside the bulge.

## *Structural Aspects of Pseudoknots*

*Basic Considerations.* Most of the pseudoknots reported today are of the H-type, which represents the simplest way to form a pseudoknot. In this type, nucleotides from a hairpin loop pair with free single-stranded regions outside the hairpin. It was first found and described for the aminoacylatable 3′ end of some plant viral RNAs (Rietveld et al., 1982, 1984; see Figure 3). On the basis of structure probing experiments, the RNA-A helix geometry, aminoacyl acceptor stem dimensions, and the similarity to the structure of tRNA, a pseudoknot was proposed to form at the 3′ end of TYMV RNA. It was likely that the two stem regions in the pseudoknot would stack upon each other, forming a quasicontinuous double helix resembling the classical aminoacyl acceptor stem. This stacking of the two stems, besides contributing to the overall stability of the structure, imposes constraints on the two connecting loops. From the RNA-A helix geometry, it follows that these connecting loops are not equivalent, crossing the deep and shallow groove of the helix, respectively (see Figure 3). The nucleotides of L1, which crosses the deep groove, can be accommodated in the groove, whereas for the residues of L2 there is no room in the shallow groove, and the bases probably point outward into the solvent. This explains the enhanced sensitivity toward the single-strand-specific chemicals diethyl pyrocarbonate and sodium bisulfite of the A and C residues in the loop crossing the shallow groove (Van Belkum et al., 1989).

*Structure and Thermodynamic Stability.* Nuclear magnetic resonance techniques are a powerful tool to gain insight into the three-dimensional structure of biomacromolecules, including RNA [for reviews, see Van de Ven and Hilbers (1988) and Varani and Tinoco (1991)]. A first attempt to apply this technique to pseudoknotted RNA molecules was made by Van Belkum et al. (1989), but no detailed structural information was obtained. The original proposals about the conformation of pseudoknots made by Pleij et al. (1985) were to a large extent confirmed by studies on short model pseudoknots (Puglisi et al., 1990; Wyatt et al., 1990). By using NMR methods, these authors were able to show that in this H-type pseudoknot the two helical regions were stacked, albeit with minor perturbations at the stem–loop junctions. Both stems indeed adopt A-form geometry, with C-3′ endo sugar puckers. While one side of the quasicontinuous helix shows NOE connectivity from S1 to S2, there is a break on the other side where the two loops emerge from the stem junction (see Figure 4). Model building shows the two loops to be in close proximity here, so that a slight distortion at this point is not surprising.

The thermodynamic stability of the pseudoknots studied by NMR was found to be not much greater than that of either of the two constituent hairpins, the gain in free energy being only 1.5–2 kcal/mol (37 °C), even in the presence of $Mg^{2+}$ (Puglisi et al., 1988; Wyatt et al., 1990). Similar values were found by Mans and colleagues in an indirect way for the pseudoknot at the 3′ end of TYMV RNA, using structure mapping and aminoacylation as a functional assay (Mans et al., 1992). In all H-type pseudoknots studied here, S1 consists of only three base pairs, and this might be the reason for their relatively low stability. Pseudoknots having more base pairs in their stems might be more stable than these relatively small ones. A study of the influence of the length of the loop regions was carried out on a series of potentially pseudoknot-forming RNAs with S1 = 3 bp and S2 = 5 bp (Wyatt et al., 1990). For this particular pseudoknot, the minimum lengths found were three nucleotides for the loop crossing the deep groove and four nucleotides for the loop crossing the shallow groove. Shortening the loops further results in the appearance of conformations other than the pseudoknot, such as the 5′ or 3′ hairpin.

*Modeling.* Much insight may gained from considering RNA molecules as a set of double-helical regions, arranged in space and connected by single-stranded regions. The reality of RNA structure is much more complicated, as we know from tRNA crystal structures [e.g., Kim et al. (1974)] and recent NMR data from stable RNA hairpins (Cheong et al., 1990; Heus et al., 1991). No crystal structures have been resolved of pseudoknot-containing RNAs. Therefore, computer modeling is an alternative to test and visualize how pseudoknotted RNAs may fold. Such a modeling study was done for the pseudoknot-containing tRNA-like structure at the 3′ end of TYMV RNA based on the structure probing data and the $tRNA^{Asp}$ structure (Dumas et al., 1987). In a later study, Cedergren and co-workers used a different algorithm on the TYMV pseudoknot and arrived at a very

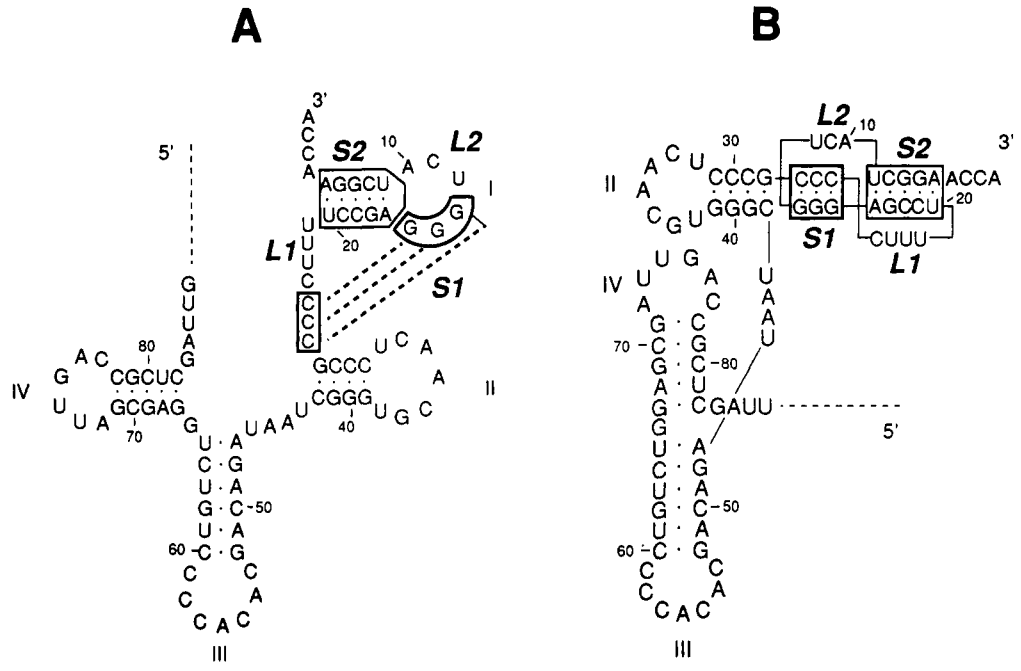


FIGURE 3: tRNA-like structure at the 3′ end of turnip yellow mosaic virus RNA. (A) Secondary structure. (B) L-shape arrangement with two pseudoknot stems aligned to form the aminoacyl acceptor stem. Nucleotides and hairpins (Roman numerals) are numbered from the 3′ end; the pseudoknot is indicated by boxes and dashed lines between them. The stem regions of the pseudoknot, S1 and S2, and the connecting loops, L1 and L2, are indicated.

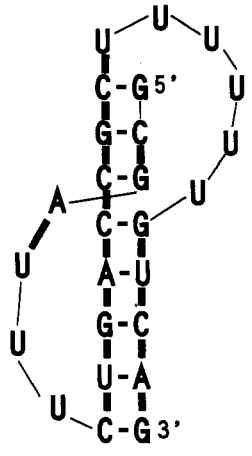


FIGURE 4: Summary of the structural information obtained from the proton NMR study of RNA oligonucleotide PK5 [adapted from Puglisi et al. (1990)]. Internucleotide nuclear Overhauser effects (NOEs) are represented by the thickness of the line connecting two nucleotides. Thick lines indicate strong 3′-H8/H6 to 5′-H2′ NOEs, consistent with A-form stacking. Thin lines indicate no observed NOE interaction. No NOE was observed between the nonexchangeable protons of the bases at the point where the loops emerge from the junction of the two stems.

similar conformation (Major et al., 1991).

*How To Find Pseudoknots in Natural RNAs?*

Since pseudoknots in a way are just an extra helix formed in an otherwise classical secondary structure, it is possible to apply the same tools as when searching for conventional secondary structures. Thus phylogenetic sequence comparisons, computer structure predictions, chemical and enzymatic structure probes, and mutational analysis (complementary and compensatory base changes) are all useful for the detection and confirmation of pseudoknots. However, one has to take into account the steric constraints and thermodynamic rules applying to pseudoknots. For instance, as pointed out above, different length requirements exist for the loop regions, depending on the number of base pairs and the nature of the groove being crossed [for a review, see Pleij and Bosch (1989)].

An interesting new development in the field of RNA structure prediction is the use of circular dichroism (CD). The nature and number of base pairs in short RNAs can be determined by comparing the UV and CD spectra with those of RNAs in a reference set; the method assumes that the helices are A-form and that only nearest-neighbor effects are important. When applied to the PK5 pseudoknotted RNA (Puglisi et al., 1990), results compatible with the NMR data were obtained, although the method needs further refinement (Johnson & Gray, 1992).

A number of efforts have been made over the years to predict the secondary structure of RNA molecules from their sequences (Zuker & Stiegler, 1981; Martinez, 1984; Jacobson et al., 1984, Zuker, 1989). While pseudoknots are now a well-established element of RNA structure, only a few programs are capable of directly predicting pseudoknots (Abrahams et al., 1990; Gultyaev, 1991). A basic problem in the prediction of pseudoknots is the lack of thermodynamic parameters for the connecting loops, but by comparing predictions with known structures while varying yet unknown parameters, improvement of the predictions has been made. In the program of Abrahams and co-workers, for example, a single value of 4.2 kcal/mol is assigned to both loop regions in H-type pseudoknots. Although there is good reason to assume that the two connecting loops are not equivalent and thus will probably make different contributions to the overall thermodynamic stability of the structure, this program was successful in predicting a number of pseudoknotted structures (Clarke et al., 1987; ten Dam et al., 1990). However, the output of RNA structure prediction programs should be treated with caution. These programs still have limitations, and it remains necessary to confirm each predicted structure, including pseudoknots, at least by phylogenetic methods but preferably also by mutational analysis and/or structure probing experiments.

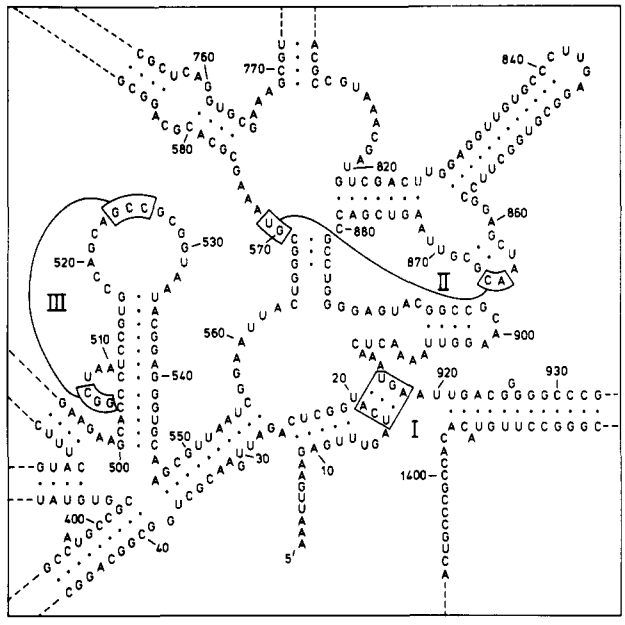

FIGURE 5: Pseudoknots in the small subunit ribosomal RNA. The sequence and numbering is that of the *E. coli* 16S rRNA; only a "core" region is shown. Three base pairings which are phylogenetically supported and create pseudoknots are shown (Roman numerals).

*Ribosomal RNAs*

The availability of a large number of ribosomal RNA sequences from a wide variety of organisms has permitted searches for conserved secondary structures at a rather detailed level. Three phylogenetically conserved pseudoknot foldings have been found in the small subunit ribosomal RNA and are shown in Figure 5. The small subunit ribosomal RNA falls into three main domains of secondary structure which are defined by long-range secondary structures; the proposed three-dimensional models of RNA packing within the subunit preserve these domains as distinct regions [see illustrations in Stern et al. (1988)]. The three pseudoknots are clustered at the junction of these three domains and may be functionally related.

The 570GU–865AC pairing has only phylogenetic evidence in its favor at the moment, though compensating base changes at the 570G–866C pair in both eubacteria and arachaebacteria make a strong case for it (Gutell et al., 1986). The potential A–U pair is universally conserved, and therefore its bonding is only surmised.

The second pseudoknot forms three base pairs between the 530 hairpin loop and a bulge loop at opposite ends of a helical segment (Woese & Gutell, 1989). The extreme conservation of bases in this region makes the evidence for this pseudoknot sparse, and so Powers and Noller (1991) have made mutations in the bulge and hairpin loops and expressed the altered rRNAs in *Escherichia coli*. Disruptions of any one of the three base pairs is either lethal or causes very slow growth, while an A–U substitution for any of the three G–C pairs allows normal growth; these experiments strongly support the existence and importance of the pseudoknot. There is an intriguing connection between this structure and streptomycin, an antibiotic which binds to the ribosome small subunit and causes misreading. Two different G–U substitutions and one A–U substitution conferred weak resistance to streptomycin; in the one case tested in vitro, the mutant ribosomes bound streptomycin about 5-fold more weakly than wild type. The protein S12, in which mutations causing streptomycin resistance have been isolated, appears to stabilize the pseudoknot pairing during ribosome assembly. Most evidence at present suggests that streptomycin does not interact with the 530 loop pseudoknot but in the 900 region of the small subunit rRNA (Gravel et al., 1987; Moazed & Noller, 1987). This implies that there is some allosteric communication between the streptomycin binding site and the pseudoknot and that the mutations conferring streptomycin resistance do so by altering the stability or precise structure of the pseudoknot, rather than by altering contacts with the antibiotic. Evidence that the 530 loop may actually be close to the 900 region has been put forward recently (Rink-Appel et al., 1991), and its positioning within the subunit remains to be resolved. A detailed model of the functional role of this pseudoknot is speculative at this point, but its functional importance is certainly established, and its involvement in conformational changes within the functioning ribosome is strongly hinted.

The third pseudoknot is a pairing of a hairpin near the rRNA 5′ terminus with the 900 region. In the model of Stern et al. (1988), it is coaxially stacked with other helices to form a key structure spanning the waist of the subunit; the structure links the site of codon–anticodon interaction with a region of the ribosome affecting translational accuracy (including the 530 loop pseudoknot described above). It is phylogenetically supported (Gutell et al., 1985), and there is recent direct evidence from mutational studies (M. Brink and H. de Boer, personal communication). The streptomycin binding site is located adjacent to the pseudoknot, as are many mutations affecting streptomycin sensitivity of ribosomes (Harris et al., 1989). The pseudoknot is therefore in a position to mediate interactions between two important regions of the 30S subunit, and two proposals have appeared for the involvement of this pseudoknot in a conformational switch. Kössel et al. (1990) have noticed that an alternate pseudoknot structure is available: UGAUC in the loop of the 5′ hairpin can pair with a sequence just 5′ of the anti-Shine–Dalgarno sequence near the 3′ terminus of the RNA. These RNA segments are probably in close proximity in the intact subunit, so the proposal is physically reasonable, but unfortunately the sequences in question are highly conserved and compensatory base changes to support the alternate pairing are lacking. Kössel et al. suggest that the role of the alternate pseudoknot is to sequester the anti-Shine–Dalgarno sequence from contact with the messenger coding sequences after initiation.

A different conformational switch has been proposed by Leclerc and Brakier-Gingras (1991). This also is a switch between alternate pairings of the 5′ hairpin loop but with a sequence immediately 5′ to the positions paired in the usual secondary structure model. Again, the alternate pairing is highly conserved, and the phylogenetic data are too sparse to support the model. There is direct evidence for a conformational switch in this region; Allen and Noller (1989) have noticed that the reactivity of A908 to dimethylsulfate increases with the translational error frequency of the ribosomes in different streptomycin-resistant and -dependent strains. They propose that the mutations shift a conformational equilibrium between two alternate ribosome structures and that the position of the equilibrium establishes the balance between translational accuracy and speed. The switch proposed by Leclerc and Gingras (1991) potentially accounts for the changes in chemical reactivity seen by Allen and Noller (1989) and the effects of streptomycin on translational accuracy (one of the pseudoknot forms would not bind streptomycin). At this point, the evidence for the pseudoknot involvement in a switch is largely circumstantial, but the proposal makes specific predictions about the effects of mutations in this region on

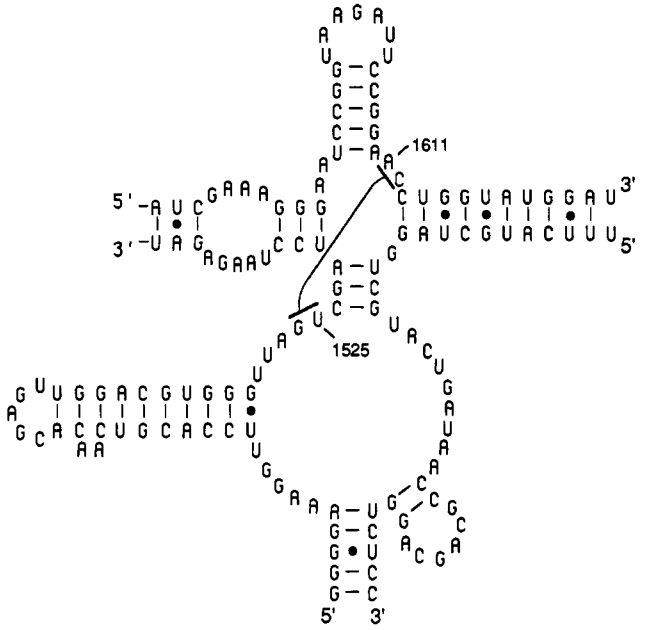


FIGURE 6: L25 protein binding site in the large subunit ribosomal RNA. The sequence and numbering is for *Saccharomyces cerevisiae* 26S rRNA; *E. coli* 23S rRNA has a homologous structure. The base pairing GU1525–AC1612, which is supported by phylogenetic conservation and site-directed mutagenesis, is indicated by lines.

streptomycin binding and translational accuracy which can be tested. The tentative nature of both this proposal and the one from Kössel et al. (1990) is underscored by the fact that both seem very plausible, but it is unlikely that both can be true.

In the large subunit ribosomal RNA, the phylogenetic evidence points to a number of pseudoknot foldings (Gutell & Woese, 1990; Larsen, 1992). There are short helical segments linking hairpin loops as well as more complex pairings within large helix junction structures. One of these, which occurs in the region binding either *E. coli* L23 protein or yeast L25 (El-Baradi et al., 1987), has been tested experimentally. Mutation of GU1525 (yeast numbering) severely reduced L25 binding in vitro, but binding could be rescued by a compensatory mutation in AC1612 (H. Raué and C. Rutgers, personal communication). The latter sequence is in one loop segment formed by the junction of four helices; its complement is at the distal end of one of the stems forming the junction (Figure 6). Formation of this pseudoknot must severely constrain the folding of the L25 recognition site.

The possibility of a pseudoknot folding within an expansion segment of the small subunit rRNA (a structure present only in eukaryotic ribosomes) has been pointed out (Neefs & De Wachter, 1990; Nickrent & Sargent, 1991). Specific roles for this putative structure have not yet been tested by mutagenesis or other means.

### *Catalytic and Self-Splicing RNAs*

*RNase P.* The first phylogenetically derived model for the catalytic RNA component of RNase P contained one pseudoknot structure, a four base pair helix linking a multibranched loop with a bulge loop (James et al., 1988). More recent experiments have obtained compensatory base changes in highly conserved regions, by selection of pseudorevertants (Haas et al., 1991). New eubacterial sequences have also provided more compensatory changes. The new evidence suggests that the original four base pair pseudoknot helix can be extended by at least two base pairs if a single U

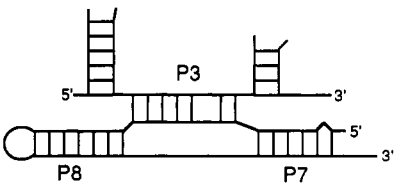


FIGURE 7: Schematic diagram of a conserved pseudoknot in the group I intron core structure.

is bulged; a total of eight base pairs is possible. A new pseudoknot linking a bulge loop and a hairpin loop is also proposed; models of the structure show that two helical segments would be highly constrained to have their axes at approximate right angles. In *Bacillus subtilis*, the hairpin contributing to the pseudoknot is deleted, and there is a small hairpin inserted into the bulge loop. However, models show that a structure very similar to the pseudoknot in three dimensions can still be constructed. This appears to be "compensatory change" involving a significant rearrangement of the secondary structure and is comparable to the original observation of a pseudoknot folding which preserves the essence of tRNA three-dimensional structure despite a major alteration in the secondary structure (Rietveld et al., 1982). In the RNase P case, it seems that the pseudoknot is simply an alternative to standard secondary structures and has no unique functional role.

Neither of the RNase P pseudoknots is absolutely essential for catalytic activity, though the activity is severely diminished when sequences contributing to either are deleted. The precise functional role of the pseudoknots will probably not be understood until substrate binding and catalysis are understood in much greater detail for this ribozyme.

*Group I Introns.* The first phylogenetically derived secondary structures for the self-splicing group I introns contained a pseudoknot structure (Waring & Davies, 1984), though the term pseudoknot was not in use at the time. The structure can be viewed as an interaction between one strand of an internal loop of an extended hairpin (between helices P7 and P8) and a single strand 5′ to this hairpin, forming helix P3 (Figure 7). A requirement for an intact P7 helix was demonstrated by compensatory base changes (Burke et al., 1986). A double mutant substituting G–C for A–U was much less active than wild type in splicing at low temperatures, though in the presence of high $Mg^{2+}$ concentrations even the single base, disruptive mutants were active. The G–C pair in the mutant was supposed to stabilize the P7 helix, and it was suggested that the altered temperature dependence was due to a conformational switch requiring P7 to melt.

More recent work has underscored the importance of the pseudoknot as part of the essential core structure of group I introns. P7, P3, and P8 were modeled as a nearly continuous helix by Kim and Cech (1987), and this model has been extended by Michel and Westhof (1990). These workers added a number of tertiary interactions, supported by phylogenetic covariances, to create a very compact core which contains the binding sites for guanosine cofactor and splicing substrate. The pseudoknot is an essential part of this core, and further site-directed mutagenesis has shown that the guanosine cofactor hydrogen bonds to a G in the major groove of the P7 helix (Michel et al., 1989). The current view of the structure is that it is fairly static; further evidence for denaturation or rearrangement of helix P7 has not materialized.

In some group I intron sequences there is the possibility of a second pseudoknot formed by pairing of one strand of an internal loop with a hairpin loop. This structure has also been

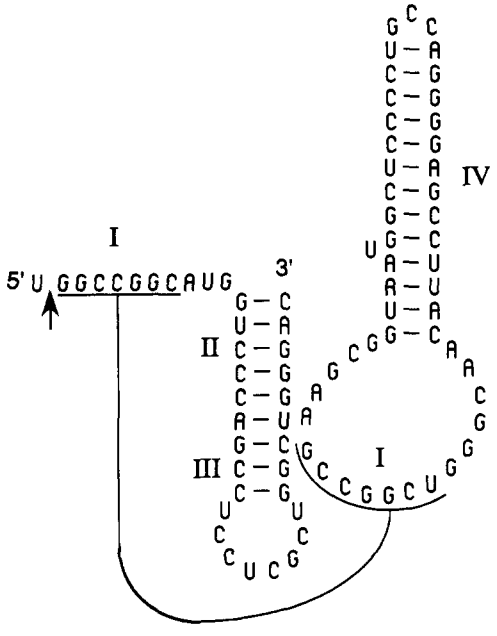


FIGURE 8: Hepatitis delta virus self-cleaving RNA. An arrow indicates the cleavage site, and underlines indicate a helix which creates a pseudoknot. Helices are numbered with Roman numerals. The base pairing is that proposed by Perrotta and Been (1991).

confirmed by compensatory base mutagenesis (Jaeger et al., 1991). The pseudoknot can be accommodated in the intron model of Michel and Westhof (1990), and the model provides some rationalization for how the pseudoknot might increase the stability of the core structure.

*Hepatitis Delta Virus.* Both the genomic and antigenomic RNAs of hepatitis delta virus contain a self-cleavage site (Kuo et al., 1988) formed from about 85 nucleotides (Perrotta & Been, 1990). The extreme stability of the cleavage structure is shown by the fact that strong denaturants (5 M urea, 10 M formamide, or high temperature) accelerate the reaction, presumably by destabilizing competing RNA structures (Rosenstein & Been, 1990). Both of the cleavage sites can be folded as similar pseudoknots. The folding pairs a bulge loop with a sequence at the 5′ terminus (Figure 8), similar to the folding in a tRNA-like structure in TMV RNA (Mans et al., 1991). The base pair stacking arrangement at the junction of helices I, II, and III is not obvious. Both helices I and II defining the pseudoknot have been tested by compensatory base changes and are required for efficient cleavage in denaturing solvents, but helix II is not needed in the absence of denaturants (Perrotta & Been, 1991; Wu et al., 1992). Thus the pseudoknot folding stabilizes the self-cleavage structure but is not essential for its formation.

*Hammerheads.* A variety of viroid and satellite RNAs self-cleave to form monomeric RNAs after replication and have a common motif responsible, the "hammerhead" structure. This consists of three helices forming a junction loop of nearly invariant sequence, within which cleavage occurs (Symons, 1989). In the case of a satellite RNA associated with barley yellow dwarf virus (sBYDV), a pseudoknot can potentially form outside of the conserved hammerhead structure and link the hairpin loop of one of the three helices with one strand of an internal loop adjacent to another of the three helices forming three coaxially stacked helices (Miller et al., 1991). Whether this pseudoknot exists or is functionally important has been tested by appropriate mutants (Miller & Silver, 1991). Compensatory base changes show that the pseudoknot does form and reduces the rate of self-cleavage by about 400-fold; base pair substitutions which increase or decrease the stability of the pseudoknot helix suppress or enhance the cleavage rate, respectively (Miller & Silver, 1991). It thus appears that the pseudoknot destabilizes the hammerhead structure and that cleavage only occurs when the pseudoknot opens transiently.

### *Ribonucleoprotein Complexes*

*Telomerases.* The polymerase which adds repeated DNA sequences to the ends of chromosomes is a ribonucleoprotein complex, telomerase. The RNA component of telomerases contains the template sequence specifying the added DNA repeat (Yu et al., 1990). There is speculation that the telomerase is a ribozyme, though this remains to be proved or disproved. Recent sequencing of a number of tetrahymenine ciliates has led to a proposal for a common, conserved secondary structure for these RNAs (Romero & Blackburn, 1991). Within this structure is an H-type pseudoknot structure supported by compensatory base changes (ten Dam et al., 1991). Another pseudoknot can potentially form between sequences which are conserved among all the telomerase RNAs. This involves a 5′-CAU···CAU sequence adjacent to the template sequence, which could pair with duplicate AUGs if a highly conserved helix

5′-AUG—GAUG

3′-UAC$_{AG}$CUAC

is disrupted. Romero and Blackburn (1991) propose that the AG bulge facilities opening of the hairpin to allow this pairing and that a switch between the two alternate secondary structures is involved in translocation of the telomerase. Since the telomerase is a processive enzyme which must dissociate a nine base pair RNA–DNA hybrid and position itself for another round of polymerization without totally dissociating from the DNA substrate, a substantial conformational rearrangement of the telomerase during this cycle seems likely. The proposed pseudoknot opening and closing is speculative at this point but is an intriguing suggestion.

*7SL RNA.* The signal recognition particle (SRP) is required for translocation of secretory proteins across membranes and consists of a ~300 nucleotide RNA complexed with six proteins. The RNA has been sequenced from a variety of organisms, and a phylogenetically conserved secondary structure has been deduced (Larsen & Zwieb, 1991). In archaebacteria and *B. subtilis*, there is a conserved five base pair interaction between two hairpin loops; a similar pairing has not been conserved in the eukaryotes. An additional pseudoknot pairing has been proposed on the basis of structure mapping experiments and phylogenetic conservation; its formation requires the disruption of a short hairpin stem and part of another helix adjacent to an internal loop in the Larsen and Zwieb (1991) model (Andraezzoli & Gerbi, 1991). Chemical modification evidence supports the existence of this pairing only in deproteinized RNA and not in the protein complex, which has prompted the suggestion that the pseudoknot is involved in assembly of the particle and is rendered unnecessary by the bound proteins (Andreazzoli & Gerbi, 1991). The functional roles of these two pseudoknots remain speculative at this time.

*U2 snRNA.* In splicing of group II introns in yeast, U2 snRNA pairs with pre-mRNA at the branch point and is essential for growth. Phylogenetic comparisons give good evidence for the existence of a pseudoknot structure in the U2 snRNA (Ares & Igel, 1989). However, attempts to construct a number of compensatory changes in the U2 secondary

structure supported the existence of only one of the pseudoknot stems; triple mutations in either strand of the second stem showed an unaltered phenotype in vivo (Ares & Igel, 1990). The pseudoknot is therefore not essential for U2 RNA function, and its role is uncertain; perhaps some assay other than cell viability would reveal a more subtle effect of the U2 pseudoknot on splicing efficiency.

*Messenger RNAs*

*Translationally Regulated Ribosome Binding Sites.* In several translationally regulated mRNAs, the ribosome binding site is positioned within a pseudoknot structure. Two of these are mRNAs which encode ribosomal proteins and are autoregulated by binding of one of the encoded proteins to the pseudoknot structure. The mRNA coding for *E. coli* S15 (*rpsO*) contains an H-type pseudoknot with a 29 base linker sequence, which itself may contain a weak hairpin; evidence for this structure comes from extensive structure mapping experiments and compensatory base changes (Philippe et al., 1990; Portier et al., 1990). The ribosome binding site is located within the pseudoknot structure, and the initiation codon begins two bases 5′ to the second helix. Formation of the fMet-tRNA–ribosome–mRNA initiation complex must disrupt the pseudoknot at least partially. It is proposed that the mRNA is in a conformational equilibrium between the pseudoknot structure and an alternative hairpin structure and that S15 binding stabilizes the pseudoknot conformation (Portier et al., 1990).

The $\alpha$ operon of *E. coli*, which encodes four ribosomal proteins besides the $\alpha$ subunit of RNA polymerase, is translationally repressed by one of the four ribosomal proteins, S4. The S4 protein specifically recognizes a 139 base RNA fragment containing the leader and the first 15 codons of the mRNA (Deckman & Draper, 1985; Deckman et al., 1987; Thomas et al., 1987). Structure mapping experiments and compensatory base changes support the existence of a pseudoknot in this mRNA (Deckman & Draper, 1987; Tang & Draper, 1989). The structure is similar to that of the S15 mRNA, in that the initiation codon is located one nucleotide 5′ to the start of the second helical segment, but the overall structure is more complex. Structure mapping and thermal denaturation experiments indicate that the pseudoknot is very stable at 37 °C (Deckman & Draper, 1987; Spedding et al., 1992).

An allosteric mechanism for translational repression in the $\alpha$ mRNA has been suggested (Tang & Draper, 1990) on the basis of the observation that some mRNA mutations do not affect repressor binding affinity or translation rates but do severely decrease the level of repression. It is postulated that the repressor and ribosome recognize different domains of the pseudoknot and that repressor binding induces an allosteric conformational change in the ribosome binding site. Mutants which reduce repression without affecting repressor or ribosome binding are explained as disruptions of the allosteric communication between the two domains. The mRNA is able to form two different complexes with ribosomes in vitro, only one of which forms ternary $tRNA_f^{met}$ – 30S subunit–mRNA complexes (Spedding et al., 1992). Preliminary data indicate that the other, inactive complex binds S4 (G. Spedding and D. E. Draper, unpublished observations). It thus seems likely that an mRNA conformational switch is involved in the repression mechanism, though the details of the conformations are not yet clear.

A third instance is known in which the ribosome binding site is apparently located within the linker of a pseudoknot structure; this is the *repY* gene of ColIb-P9 plasmid. This is part of a complex system regulating the level of *repZ*, just downstream of *repY*, and involves binding of antisense RNA upstream of *repY* as well as translation of *repY*. Genetic analysis of this system identified compensatory base changes between short complementary sequences 107 bases apart, one upstream of *repY* and one within the *repY* coding region (Asano et al., 1991). The upstream sequence is within the hairpin loop of a predicted very stable hairpin (Hama et al., 1990) and thus potentially forms a pseudoknot; the downstream sequence is just 5′ to the *repZ* Shine–Dalgarno sequence and could alternatively form a stable hairpin with the *repZ* ribosome binding site. Asano et al. (1991) propose that translation of *repY* disrupts the *repZ* ribosome binding site hairpin, and that the pseudoknot then forms as an alternative structure which unmasks the *repZ* initiation site and stimulates its translation.

The gene 32 protein of T4 phage binds to single-stranded nucleic acids with high cooperativity and relatively low specificity, in effect "coating" the strand with bound protein. It can, however, specifically repress its own translation by preferentially binding to its own mRNA (Lemaire et al., 1978). At low concentrations, gene 32 protein binds a site upstream of the ribosome binding site, and with increasing concentrations the mRNA protected from nuclease digestion by bound protein extends downstream from this nucleation site (McPheeters et al., 1988). At sufficiently high protein concentration, the ribosome binding site itself is protected and translation is repressed. These observations imply that there is a particularly high-affinity gene 32 protein binding site in the mRNA leader, and it is binding to this site which nucleates the cooperative binding and leads to repression. An H-type pseudoknot secondary structure is possible at the nucleation site, and its existence is supported by compensatory base changes in other T-even phage sequences (McPheeters et al., 1988). The protein covers about five nucleotides when bound (Kelley et al., 1976) and lengthens the backbone by unstacking bases (Delius et al., 1972; Jensen et al., 1976). An attractive hypothesis is that the pseudoknot single strand, which is five bases and must stretch ~22 Å across the shallow groove of S1 (Pleij et al., 1985), has a distorted backbone conformation which is particularly favorable for protein binding. Physical studies of a gene 32 protein–pseudoknot complex should be able to test this model.

*Frame Shifts and Supressors.* In general, pseudoknots are found in noncoding regions of RNAs. To translate a codon, its bases must be available for the codon–anticodon interaction, and any secondary structure in the mRNA needs to be disrupted before the ribosome can translate it into protein. It was therefore surprising that functional pseudoknots have been found in coding regions of mRNAs.

Translational suppression is used by retroviruses to express the *pol* gene which encodes three essential enzymatic proteins: integrase, protease, and reverse transcriptase. These are expressed as one large fusion protein, together with the internal structural proteins of the virus core (termed the *gag* proteins). However, the structural proteins (core proteins) are needed in much larger amounts than the catalytic proteins (polymerase and protease). To produce these gene products in different amounts, the viruses use ribosomal frame shifts and translational read-through; the emphasis of this discussion is on the role of pseudoknots in both processes. [For recent reviews, see Jacks (1990), Atkins et al. (1990), and Hatfield et al. (1991).]

Ribosomal frame shifting is a process by which a single fusion protein is synthesized from two or more overlapping genes through alteration of the reading frame. This change of frame can occur by movement of the ribosome in either 5′ (−1) or 3′ (+1) direction. Jacks and co-workers proposed a model for −1 ribosomal frame shifting in retroviruses (Jacks et al., 1988a), which also seems to hold for other viruses with similar expression strategies for *pol*-like genes, such as the coronaviruses and a yeast dsRNA virus (Brierley et al., 1987; Dinman et al., 1991). Jacks and co-workers suggested that the frame shift signal of Rous sarcoma virus (RSV) consists of two elements: (1) a heptanucleotide sequence that forms the actual frame shift site and (2) an RNA structure element downstream of the slip site.

The frame shift site consists of a seven nucleotide X XXY YYN motif, where X can be any base, Y is A or U, and N is A, U, or C. This heptanucleotide sequence suggests a mechanism for their function as formulated in the "simultaneous slippage" model. In this model, the two tRNAs in the A and P sites on the ribosome move one base into the 5′ direction at the same time, so that at least two residues of both anticodons will be paired after the slippage has occurred. The protein sequence of transframe products is in agreement with this mechanism: the last two *gag* codons that are translated are coded for by the last six residues of the heptanucleotide sequence, and the first −1 frame codon that is decoded is the one starting at position 7 of the slip site (Jacks et al., 1988b).

Several authors have noticed that stem–loop structures can potentially form downstream of potential −1 frame shift sites (Rice et al., 1985; Sagata et al., 1985; Jacks et al., 1987), and in fact almost all of the potential slip sites in retroviruses are followed within nine nucleotides by stable hairpins. The evidence that this additional structure is needed to obtain efficient frame shifting first came from the examination of the RSV *gag–pol* frame shift site. Jacks et al. (1988a) could show by mutational analysis that a stem region six nucleotides downstream of the slip site was necessary for frame shifting. In addition, they found that an additional 23 nucleotides downstream of this hairpin were also essential for frame shifting, suggesting the formation of a pseudoknot. In a more comprehensive analysis of the orf 1A–1B overlap of the coronavirus IBV, Brierly and colleagues could show that the −1 frame shift is pseudoknot dependent (Brierly et al., 1989). In this study and a later one, mutational analysis showed that both stems forming the pseudoknot were necessary and that mutations in one strand of a stem could be compensated for by complementary mutations in the other strand. Moreover, the pseudoknot could not be replaced by a hairpin of the same length and composition as the two stacked stems forming the pseudoknot, indicating that some specific feature of the pseudoknot is needed (Brierly et al., 1989, 1991).

Computer predictions of structures near −1 overlaps in viral RNAs showed that for a majority of the (potential) frame shift sites it was possible to form a pseudoknot four to eight bases after the slip site (ten Dam et al., 1990). However, for slip sites starting with six identical bases, a pseudoknot could not always be predicted. This is in agreement with observations that for HIV-1, which has the slip site U UUU UUA, no additional structure is necessary (Jacks et al., 1988b; Wilson et al., 1988; Madhani et al., 1988), although a pseudoknot enhances the frame shift efficiency (Brierly et al., 1992). It thus seems that pseudoknot dependence is connected with the composition of the slip site. However, there appeared to be no specific demands in size or composition for the pseudoknots. Spacing between slip site and pseudoknot is more important: pseudoknots are found four to eight bases after the heptanucleotide. Changing the distance from six to three or nine severely lowered frame shift efficiency in IBV (Brierly et al., 1989). Pseudoknot-dependent frame shifting has now been proven for, or is at least shown to be likely for, the coronavirus MHV (Bredenbeek et al., 1990), the torovirus BEV (Snijder et al., 1990), mouse mammary tumor virus (Chamorro et al., 1992), the yeast dsRNA viruses L1 and LA (Dinman et al., 1991; Tzeng et al., 1992), feline immunodeficiency virus (Morikawa & Bishop, 1992), and simian retrovirus 1 (ten Dam et al., manuscript in preparation).

Although some eukaryotic frame shift signals appear to be functional in *E. coli*, the efficiencies and mechanisms used are not identical to those found in eukaryotes. Downstream secondary structure only has a limited effect on frame shifting efficiency in *E. coli* for HIV, MMTV (Weiss et al., 1989), the *E. coli dnaX* gene (Blinkowa et al., 1990; Flower et al., 1990; Tsuchihashi, 1991), and the beet western yellows virus (Garcia et al., manuscript in preparation).

In type C retroviruses the *gag* and *pol* genes are in frame, separated by a single UAG termination codon. This amber stop codon is suppressed to allow the expression of the downstream *pol* reading frame, as first shown for Moloney murine leukemia virus (MoMuLV) (Yoshinaka et al., 1985). The protease is synthesized by suppression of the UAG codon through insertion of a glutamine.

Panganiban (1988) showed that a 30 base pair fragment encompassing the UAG codon was sufficient for giving read-through at wild-type levels in a heterologous context. The amount of amber suppression did not increase in response to viral infection, suggesting that the read-through signal is cis-acting. The analysis by ten Dam et al. (1990) led them to propose the involvement of pseudoknots in translational read-through in type C retroviruses as well. In these viruses, the UAG termination codon and the pseudoknot are arranged in a way very similar to the situation found in the viruses using frame shifting. Honigman and co-workers (1991) found that mutations which would disrupt the pseudoknot as predicted by ten Dam et al. (1990) abolished UAG suppression, but they could not restore it by making the compensatory mutant. The latter mutant, however, has an additional base pair at the bottom of S1, thereby changing the distance between UAG and pseudoknot from eight to seven bases. In all other type C retroviruses, this distance is eight base pairs. A more extensive mutational analysis was done by Wills and colleagues, and in their system read-through is pseudoknot dependent, although the situation is more complicated than in the case of frame shifting. For instance, sequences 5′ of the UAG and in L2 also have an effect on overall read-through efficiency (Wills et al., 1991). In a recent study, Feng and co-workers showed that the pseudoknot is an essential part of the Mo MuLV read-through signal. Interestingly, the identities of six bases of the eight between the UAG and the pseudoknot were important for amber suppression as well (Feng et al., 1992).

The exact role of the pseudoknot in both ribosomal frame shifting and read-through is still unclear, but there are two obvious possibilities. The first is that the pseudoknot is recognized by a protein that upon binding signals the ribosome to change reading frame (or ignore the termination codon). Another one is that the translating ribosome is paused by the structure at a specific site, thereby allowing it more time to interact with the frame shift or read-through signal. The mutational analyses carried out so far show that there is no requirement for specific bases in stems and loops in −1 frame

shifting. This, together with the variety in type and composition of pseudoknots found and predicted, makes it unlikely that there is a single protein factor which binds to the pseudoknot and induces frame shifting. For read-through, this is less evident. The eight viruses examined show moderate homology, so it cannot be excluded that protein binding is needed. Also the mutational studies have not yet been as extensive as in the case of frame shifting. The similarity between the two systems, however, would argue against read-through induced by protein binding.

Several authors have noted pausing or stalling of the ribosome upon encountering a pseudoknot in the mRNA (Tsuchihashi, 1991; Somogyi, Brierly, and Inglis, personal communcation). This slowing down of the translating ribosome would give it more time to respond to the signals to ignore the normal rules of translation and termination, provided that the site of translational repression and the structure are properly spaced. If the ability to do so was a specific property of a pseudoknot, it would explain the need for a pseudoknot at the frame shift site. An intriguing possibility is that the pseudoknot forms a structure more resistant than a simple hairpin to a helicase/unwinding activity on or associated with the translating ribosome (Draper, 1990).

### *Viral RNAs*

The 3′ terminus of a number of plant viral RNAs contains a tRNA-like structure, and most of these are substrates for an aminoacyl-tRNA synthetase. These RNA sequences cannot be folded into a cloverleaf structure, but the three-dimensional structure of tRNA can be mimicked if an H-type pseudoknot is formed (Rietveld et al., 1982, 1984). The occurrence and possible functional roles of tRNA-like structures in viral RNAs have been reviewed extensively (Mans et al., 1991). Site-directed mutagenesis has shown that destabilization of the weaker of the two helices, S2, results in a lower rate of aminoacylation of BMV RNA (Dreher et al., 1988) and TYMV RNA (Mans et al., 1992). Aminoacylation went up to wild-type levels upon restoring base pairing, which means that the pseudoknot is essential for the tRNA-like structure. A gradual replacement of the three G–C pairs with A–U pairs in S2 of TYMV RNA shows that the rate of aminoacylation varies with the calculated stability of the stem (Mans et al., 1992).

Recent data obtained by various groups suggest that biological activities like aminoacylation and replication mainly reside in the aminoacyl acceptor limb of the tRNA-like structure of TYMV RNA. Rudinger et al. (1992) showed that a 42 nucleotide 3′ terminal fragment can be mischarged with histidine, while Gargouri-Bouzid et al. (1991) demonstrated that 38 nucleotide long oligomers are used as templates in vitro by the viral RNA-dependent RNA polymerase. The latter observations is in line with the proposal for the structure at the 3′ terminus of the erysium latent virus RNA, a tymovirus, which seems to lack a valine anticodon-containing stem–loop (Srifah et al., 1992).

An additional set of pseudoknot structures is located just upstream of the tRNA-like structure in plant viral RNAs. Tobacco mosaic virus (TMV) forms a structure with three consecutive pseudoknots; all six helical segments are potentially stacked into one extended helix (Van Belkum et al., 1985). Addition of the entire TMV 3′ domain (204 nucleotides) to other plant or viral RNAs has a dramatic, 100-fold effect on protein expression comparable to that of a poly(A) tail (Gallie & Walbot, 1990). Most of the effect is due to the upstream pseudoknot domain. In this respect, it is noteworthy that almost all nonpolyadenylated plant viral RNAs have one or more (potential) pseudoknots in the 3′ untranslated region just downstream of the stop codon (Pleij et al., 1987; K. Pleij, unpublished observations). A mutational analysis of the three pseudoknots in TMV RNA, including point mutations and progressive deletions for the 5′ side, showed that the double-helical segment just upstream of the tRNA-like structure was essential for viral multiplication (Takamatsu et al., 1990). The precise requirement for pseudoknots or the mechanism by which this domain enhances translation is not known at this time.

*Animal Viruses.* Several pseudoknots have been proposed for functionally important parts of animal viruses, though unequivocal evidence for their existence or importance is lacking. Among these are a pseudoknot in the RNA structure needed for packaging Moloney murine leukemia virus (Alford et al., 1991), a set of four consecutive pseudoknots in the 5′ noncoding region of foot and mouth disease virus (Clarke et al., 1987), and two potential pseudoknots in the 5′ untranslated region of hepatitis A virus (Brown et al., 1991) and encephalomyocarditis virus (Duke et al., 1992), respectively. Recently, a pseudoknotted structure was proposed in the 3′ untranslated region of some picornaviral RNAs like polio virus or Coxsackievirus (Pilipenko et al., 1992).

### *Summary of Pseudoknot Functions*

Since the pseudoknot folding was first defined as a novel RNA folding (Pleij et al., 1985), this structural motif has been found in virtually every class of RNA. Because many of the first described pseudoknots were found either in functionally important locations (e.g., small subunit ribosomal RNAs, self-splicing introns) or in structures associated with regulation of gene expression (autoregulated mRNAs, frame shifting sites), there has been considerable speculation that pseudoknots may have special functional roles which cannot be easily fulfilled by standard hairpin secondary structures. Now that evidence for the specific functions of some pseudoknots is accumulating, and some physical properties of pseudoknots are becoming well defined, it is worth considering whether pseudoknots are the unique or preferred ways to achieve particular classes of structures and functions in RNA molecules. Here we summarize several different ways in which pseudoknots contribute to RNA folding.

*Stabilization of Compact Tertiary Structures.* The tRNA-like structures in viral RNAs and the pseudoknot present in some RNase P RNAs are both instances where a pseudoknot and a regular hairpin secondary structure are apparently alternate ways to construct similar RNAs with similar function. It is therefore difficult to argue, in these cases, that the pseudoknot is a unique solution to a structural problem.

But consider the general problem of how to construct a large RNA with a specific, relatively rigid three-dimensional structure capable of binding substrates or catalyzing a reaction. A large structure based entirely on standard hairpins with loops, bulges, and helix junctions will be a branched, tree-like structure, certainly having the potential for defined three-dimensional structure when unperturbed but showing a great deal of flexibility in even a light "breeze" (i.e., modest thermal energy). Introduction of cross-links by creating loop–loop and loop–single strand base-paired segments (pseudoknots) will tie together the ends of the "tree branches" and impart substantial rigidity to the structure. These "cross-links" may also make it easier to create a compact structure, for instance a substrate-binding pocket. Phosphate–phosphate repulsion works against the formation of globular, protein-like structures

in RNA; by pinning both ends of a single strand to a framework of rigid helices, as happens with the loops of a pseudoknot, several strands of RNA can be forced into close proximity to form a compact tertiary structure. There is some suggestion that the loss of the pseudoknot in RNase P leads to a more flexible structure (Darr et al., 1992).

By this reasoning, it should not be surprising to find pseudoknots at critical points in large RNA tertiary structures, particularly in those which bind substrates. Some of the pseudoknots which have been phylogenetically detected in the large ribosomal RNAs, for instance, greatly constrain the RNA folding; the need to form a fairly rigid G nucleoside binding pocket in group I introns may also be optimally solved by pseudoknot foldings.

Tertiary hydrogen bonding (i.e., non-Watson–Crick interactions) will of course accomplish the same sort of cross-links between loops, as exemplified by the D loop–T loop interaction imparting a fairly rigid "L" shape to transfer RNAs. Whether it is either energetically or evolutionarily "easier" to build a pseudoknot than a set of noncanonical base pairings is difficult to argue at this point, and our ignorance of true tertiary interactions in any RNA other than tRNA limits the discussion. The coaxial helices of many pseudoknots maximize base stacking and thus may be an energetically favorable way to stabilize a desired loop–loop interaction. Protein–RNA interactions can also "cross-link" RNA structures and promote the formation of compact structures by neutralizing phosphates; this is potentially a major role for ribosomal proteins and the protein component of RNase P.

The S4 ribosomal protein may illustrate the structural advantage of pseudoknots. Its binding site in the 16S rRNA is a 460 nucleotide domain, which cannot be much reduced in size without perturbing the specific interaction (Vartikar & Draper, 1989; Sapag et al., 1990) even though the primary interaction site is though to be a compact junction of helices (Stern et al., 1986). The protein probably stabilizes a complex tertiary structure within the domain. S4 almost certainly uses the same binding site to recognize the mRNA pseudoknot (Vartikar & Draper, 1989); in this case a pseudoknot may be the best way to form the compact tertiary structure needed to mimic the 16S rRNA domain in a much smaller (~110 nucleotide) structure.

*Formation of Unique RNA Structures.* A pseudoknot has an unusual arrangement of single strands which distinguish it from hairpin secondary structures. The proposed function of a single-strand linker in the gene 32 mRNA pseudoknot is an illustration. To form a high-affinity nucleation site for the gene 32 protein, several single-strand bases must be unstacked. This is energetically unfavorable, and the unstacking must be paid for by a favorable interaction elsewhere in the RNA structure. In the pseudoknot, the favorable formation of a helix stem can, in principle, be the driving force for unstacking of bases in the linker. Another place where the unique geometry of pseudoknots might be functionally important is in frame shifting, where a pseudoknot is evidently more effective at causing ribosomes to pause than is a hairpin. It has been proposed that the ribosomal machinery for unwinding mRNA secondary structures is jammed by a pseudoknot, since the 5′ and 3′ ends of a pseudoknot are at opposite ends of the structure, rather than adjacent as in a hairpin (Draper, 1990).

*Conformational Switches.* Several of the proposed roles for pseudoknots involve switches between two conformation states of the RNA: suggestions that the pseudoknot structures in the small subunit rRNA, the telomerase RNA, and some of the translationally regulated mRNAs adopt alternate conformations have been put forward. Some evidence for alternate structures is available for the S15 and $\alpha$ mRNAs, but a discussion of pseudoknot switches is primarily speculative at this point. However, some general remarks can be made.

There are two necessary conditions for any kind of macromolecular conformational switch to function. First, the difference in free energy between the two conformations must not be too large, so that some external signal (e.g., ligand binding or GTP hydrolysis) is able to drive the switch in one direction or the other. Second, the kinetics of the transition must be fast compared to the process in which the switch participates. For instance, a round of ribosome elongation takes place in ~0.1 s, so any conformational swtiches associated with the ribosome cycle should have rate constants $>10\ s^{-1}$.

As pointed out in a previous section, H-type pseudoknots are less stable than a standard hairpin of the same number of base pairs and only marginally more stable (~2 kcal) than either one of the two component helices. Thus the input of a small amount of free energy may be sufficient to open a pseudoknot, and one can imagine the cycling of a pseudoknot between "open" and "closed" conformations. The proposal that S15 stabilizes an mRNA pseudoknot, which is otherwise unstable and accessible to ribosomes, therefore seems reasonable. The more complex mRNA pseudoknot recognized by S4 has an overall stability of –7.7 kcal, which is comparable to a moderate-sized hairpin (Spedding et al., 1992).

The only kinetic measurements made so far is for the opening of an H-type pseudoknot to a hairpin; the rates are temperature dependent but are on the order of $1\ s^{-1}$ (Wyatt et al., 1990). The activation enthalpy suggests that one stem must fully dissociate before more base pairs are added onto the remaining helix, hence the slow dissociation rate. This slow rate would rule against pseudoknot switches functioning in ribosomes; however, a range of model pseudoknot structures have not been explored yet, and it is possible that other pseudoknots will have much different kinetics or that bound proteins could accelerate the kinetics.

## ACKNOWLEDGMENT

We thank many colleagues for kindly communicating results prior to publication. Research on pseudoknots in D.E.D.'s laboratory is supported by NIH Grants GM29048 and GM37005.

## REFERENCES

Abrahams, J. P., van den Berg, M., van Batenburg, E., & Pleij, C. (1990) *Nucleic Acids Res. 18*, 3035–3044.

Alford, R. L., Honda, S., Lawrence, C. B., & Belmont, J. W. (1991) *Virology 183*, 611–619.

Allen, P. N., & Noller, H. F. (1989) *J. Mol. Biol. 208*, 457–468.

Andreazzoli, M., & Gerbi, S. A. (1991) *EMBO J. 10*, 767–777.

Ares, M., & Igel, A. H. (1989) in *Molecular Biology of RNA* (Cech, T. R., Ed.) pp 13–23, Alan R. Liss, Inc., New York.

Ares, M., & Igel, A. H. (1990) *Genes Dev. 4*, 2132–2145.

Asano, K., Kato, A., Moriwaki, H., Hama, C., Shiba, K., & Mizobuchi, K. (1991) *J. Biol. Chem. 25*, 3774–3781.

Atkins, J. F., Weiss, R. B., & Gesteland, R. F. (1990) *Cell 62*, 413–423.

Blinkowa, A. L., & Walker, J. R. (1990) *Nucleic Acids Res. 18*, 1725–1729.

Bredenbeek, P. J., Pachuk, C. J., Noten, J. H. F., Charité, J., Luytjes, W., Weiss, S. R., & Spaan, W. J. M. (1990) *Nucleic Acids Res. 18*, 1825–1832.

Brierly, I., Boursnell, M. E. G., Binns, M. M., Bilimora, B., Blok, V. C., Brown, T. D. K., & Inglis, S. C. (1987) *EMBO J. 6*, 3779–3785.
Brierly, I., Digard, P., & Inglis, S. C. (1989) *Cell 57*, 537–547.
Brierly, I., Rolley, N. J., Jenner, A. J., & Inglis, S. C. (1991) *J. Mol. Biol. 229*, 889–902.
Brierly, I., Jenner, A. J., & Inglis, S. C. (1992) *J. Mol. Biol.* (in press).
Brown, E. A., Day, S. P., Jansen, R. W., & Lemon, S. M. (1991) *J. Virol. 65*, 5828–5838.
Burke, J. M., Irvine, K. D., Kaneko, K. J., Kerker, B. J., Oettgen, A. B., Tierney, W. M., Williamson, C. L., Zaug, A. J., & Cech, T. R. (1986) *Cell 45*, 167–176.
Chamorro, M., Parkin, N., & Varmus, H. E. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 713–717.
Cheong, C., Varani, G., & Tinoco, I. (1990) *Nature 346*, 680–682.
Clarke, B. E., Brown, A. L., Currey, K. M., Newton, S. E., Rowlands, D. J., & Carroll, A. R. (1987) *Nucleic Acids Res. 15*, 7067–7078.
Darr, S. C., Zito, K., Smith, D., & Pace, N. R. (1992) *Biochemistry 31*, 328–333.
Deckman, I. C., & Draper, D. E. (1985) *Biochemistry 24*, 7860–7865.
Deckman, I. C., & Draper, D. E. (1987) *J. Mol. Biol. 196*, 323–332.
Deckman, I. C., Draper, D. E., & Thomas, M. S. (1987) *J. Mol. Biol. 196*, 313–322.
Delius, H., Mantell, N. J., & Alberts, B. M. (1972) *J. Mol. Biol. 67*, 341–350.
Dinman, J. D., Icho, T., & Wickner, R. B. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 174–178.
Draper, D. E. (1990) *Curr. Opin. Cell Biol. 2*, 1099–1103.
Dreher, T. W., & Hall, T. C. (1988) *J. Mol. Biol. 201*, 41–55.
Duke, G. M., Hoffmann, M., & Palmenberg, A. C. (1992) *J. Virol. 66*, 1602–1609.
Dumas, P., Moras, D., Florentz, C., Giegé, R., Verlaan, P., Van Belkum, A., & Pleij, C. W. A. (1987) *J. Biomol. Struct. Dyn. 4*, 707–728.
El-Baradi, T. T. A. L., Regt, V. H. C. F. d., Planta, R. J., Nierhaus, K. H., & Raué, H. A. (1987) *Biochemie 69*, 939–948.
Feng, Y., Yuan, H., Rein, A., & Levin, J. G. (1992) *J. Virol. 66*, 5127–5132.
Flower, A. M., & McHenry, C. S. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3713–3717.
Gallie, D. R., & Walbot, V. (1990) *Genes Dev. 4*, 1149–1157.
Gargouri-Bouzid, R., David, C., & Haenni, A.-L. (1991) *FEBS Lett. 294*, 56–58.
Gravel, M., Melançon, P., & Brakier-Gingras, L. (1987) *Biochemistry 26*, 6227–6232.
Gultyaev, A. P. (1991) *Nucleic Acids Res. 19*, 2489–2494.
Gutell, R. R., & Woese, C. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 663–667.
Gutell, R. R., Weiser, B., Woese, C. R., & Noller, H. F. (1985) *Prog. Nucleic Acid Res. Mol. Biol. 32*, 155–215.
Gutell, R. R., Noller, H. F., & Woese, C. R. (1986) *EMBO J. 5*, 1111–1113.
Haas, E. S., Morse, D. P., Brown, J. W., Schmidt, F. J., & Pace, N. R. (1991) *Science 254*, 853–856.
Hama, C., Takiezawa, T., Moriwaki, H., & Mizobuchi, K. (1990) *J. Biol. Chem. 265*, 10666–10673.
Harris, E. H., Burkhart, B. D., Gillham, N. W., & Boynton, J. E. (1989) *Genetics 123*, 281–292.
Hatfield, D. L., Levin, J. G., Rein, A., & Oroszlan, S. (1991) in *Advances in Virus Research* (Maramorosch, F., Murphy, F., & Shatkin, A., Eds.) Academic Press Inc., Orlando, FL.
Heus, H. A., & Pardi, A. (1990) *Science 253*, 191–253.
Honigman, A., Wolf, D., Yaish, S., Falk, H., & Panet, A. (1991) *Virology 183*, 313–319.
Jacks, T. (1990) *Curr. Top. Microbiol. Immunol. 157*, 93–124.
Jacks, T., Townsley, K., Varmus, H. E., & Majors, J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4298–4302.
Jacks, T., Madhani, H. D., Masiarz, F. R., & Varmus, H. E. (1988a) *Cell 55*, 447–458.
Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., & Varmus, H. E. (1988b) *Nature 331*, 280–283.
Jacobson, A. B., Good, L., Simonetti, M., & Zuker, M. (1984) *Nucleic Acids Res. 12*, 45–52.
Jaeger, L., Westhof, E., & Michel, F. (1991) *J. Mol. Biol. 221*, 1153–1164.
James, B. D., Olsen, G. J., Liu, J., & Pace, N. R. (1988) *Cell 52*, 19–26.
Jensen, D. E., Kelley, R. C., & Hippel, P. H. v. (1976) *J. Biol. Chem. 251*, 7215–7228.
Johnson, K. H., & Gray, D. M. (1992) *J. Biomol. Struct. Dyn. 9*, 733–745.
Kearns, D. R., & Wong, Y. P. (1974) *J. Mol. Biol. 87*, 755–774.
Kearns, D. R., Wong, Y. P., Chang, S. H., & Hawkins, E. (1974) *Biochemistry 13*, 4736–4746.
Kelley, R. C., Jensen, D. E., & Hippel, P. H. v. (1976) *J. Biol. Chem. 251*, 7240–7250.
Kim, S.-H., & Cech, T. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8788–8792.
Kim, S.-H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H. J., Seeman, N. C., & Rich, A. (1974) *Science 185*, 435–440.
Kössel, H., Hoch, B., & Zelt, P. (1990) *Nucleic Acids Res. 18*, 4083–4088.
Kuo, M. Y.-P., Sharmeen, L., Dinter-Gottlieb, G., & Taylor, J. (1988) *J. Virol. 62*, 4439–4444.
Larsen, N. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5044–5048.
Larsen, N., & Zwieb, C. (1991) *Nucleic Acids Res. 19*, 209–215.
Leclerc, D., & Brakier-Gingras, L. (1991) *FEBS Lett. 279*, 171–174.
Lemaire, G., Gold, L., & Yarus, M. (1978) *J. Mol. Biol. 126*, 73–90.
Madhani, H. D., Jacks, T., & Varmus, H. E. (1988) in *The Control of Human Retrovirus Gene Expression* (Franz, R., Cullen, B., & Wong-Staal, F., Eds.) pp 119–125, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E., & Cedergren, R. (1991) *Science 253*, 1255–1260.
Mans, R. M. W. (1991) Functional Properties of Plant Viral RNA Pseudoknots, Ph.D. Thesis, University of Leiden, The Netherlands.
Mans, R. M. W., Pleij, C. W. A., & Bosch, L. (1991) *Eur. J. Biochem. 201*, 303–324.
Mans, R. M. W., van Steeg, M. H., Verlaan, P. W. G., Pleig, C. W. A., & Bosch, L. (1992) *J. Mol. Biol. 223*, 221–232.
Martinez, H. M. (1984) *Nucleic Acids Res. 12*, 323–334.
McPheeters, D. S., Stormo, G. D., & Gold, L. (1988) *J. Mol. Biol. 201*, 517–535.
Michel, F., & Westhof, E. (1990) *J. Mol. Biol. 216*, 585–610.
Michel, F., Hanna, M., Green, R., Bartel, D. P., & Szostak, J. W. (1989) *Nature 342*, 391–395.
Miller, W. A., & Silver, S. L. (1991) *Nucleic Acids Res. 19*, 5313–5320.
Miller, W. A., Hercus, T., Waterhouse, P. M., & Gerlach, W. L. (1991) *Virology 183*, 711–720.
Moazed, D., & Noller, H. F. (1987) *Nature 327*, 389–394.
Morikawa, S., & Bishop, D. H. L. (1992) *Virology 186*, 389–397.
Neefs, J.-M., & De Wachter, R. D. (1990) *Nucleic Acids Res. 19*, 5695–5704.
Nickrent, D. L., & Sargent, M. L. (1991) *Nucleic Acids Res. 19*, 227–235.
Ninio, J. (1971) *Biochimie 53*, 485–494.
Panganiban, A. T. (1988) *J. Virol. 62*, 3574–3580.
Perrotta, A. T., & Been, M. D. (1990) *Nucleic Acids Res. 18*, 6821–6827.
Perrotta, A. T., & Been, M. D. (1991) *Nucleic 350*, 434–436.
Philippe, C., Portier, C., Mougel, M., Grunberg-Manago, M., Ebel, J.-P., Ehresmann, B., & Ehresmann, C. (1990) *J. Mol. Biol. 211*, 415–426.

Pilipenko, E. G., Maslova, S. V., Sinyakov, A. N., & Agol, V. I. (1992) *Nucleic Acids Res. 20*, 1739–1745.
Pleij, C. W. A. (1990) *Trends Biochem. Sci. 15*, 143–147.
Pleij, C. W. A., & Bosch, L. (1989) *Methods Enzymol. 180a*, 289–303.
Pleij, C. W. A., Reitveld, K., & Bosch, L. (1985) *Nucleic Acids Res. 13*, 1717–1731.
Pleij, C. W. A., Abrahams, J. P., Van Belkum, A., Reitveld, K., & Bosch, L. (1987) The spatial folding of the 3′ non-coding region of aminoacylatable plant viral RNAs, in *Positive Strand Viruses* (Brinton, M. A., & Rueckert, P., Eds.) pp 299–316, Alan R. Liss, New York.
Portier, C., Philippe, C., Dondon, L., Grunberg-Manago, M., Ebel, J. P., Ehresmann, B., & Ehresmann, C. (1990) *Biochim. Biophys. Acta 1050*, 328–336.
Powers, T., & Noller, H. F. (1991) *EMBO J. 10*, 2203–2214.
Puglisi, J. D., Wyatt, J. R., & Tinoco, I. (1990) *J. Mol. Biol. 214*, 437–453.
Rice, N. R., Stephens, R. M., Burny, A., & Gilden, R. V. (1985) *Virology 142*, 357–377.
Richards, E. G. (1969) *Eur. J. Biochem. 10*, 36–42.
Rietveld, K., van Peolgeest, R., Pleij, C. W. A., van Boom, J. H., & Bosch, L. (1982) *Nucleic Acids Res. 10*, 1929–1946.
Rietveld, K., Linschooten, K., Pleij, C. W. A., & Bosch, L. (1984) *EMBO J. 3*, 2613–2619.
Rink-Appel, J., Jünke, N., Stade, K., & Brimacombe, R. (1991) *EMBO J. 10*, 2195–2202.
Romero, D. P., & Blackburn, E. H. (1991) *Cell 67*, 343–353.
Rosenstein, S. P., & Been, M. D. (1990) *Biochemistry 29*, 8011–8016.
Rudinger, J., Florentz, C., Dreher, T., & Giegé, R. (1992) *Nucleic Acids Res. 20*, 1865–1870.
Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K., Ogawa, Y., & Ikawa, Y. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 677–681.
Sapag, A., Vartikar, J. V., & Draper, D. E. (1990) *Biochim. Biophys. Acta 1050*, 34–37.
Schimmel, P. (1989) *Cell 58*, 9–12.
Snijder, E. J., den Boon, J. A., Bredenbeek, P. J., Horzinek, M. C., Rijnbrand, R., & Spaan, W. J. M. (1990) *Nucleic Acids Res. 18*, 4535–4542.
Spedding, G. S., Gluick, T. C., & Draper, D. E. (1992) *J. Mol. Biol.* (in press).
Srifah, P., Keese, P., Wieler, G., & Gibbs, A. (1992) *J. Gen. Virol. 73*, 1437–1447.
Stern, S., Wilson, R. C., & Noller, H. F. (1986) *J. Mol. Biol. 192*, 101–110.
Stern, S., Weiser, B., & Noller, H. F. (1988) *J. Mol. Biol. 204*, 447–481.
Studnicka, G. M., Rahn, G. M., Cummings, I. W., & Salser, W. A. (1978) *Nucleic Acids Res. 5*, 3365–3387.
Symons, R. H. (1989) *Trends Biochem. Sci. 14*, 445–450.
Takamatsu, N., Watanabe, Y., Meshi, T., & Okada, Y. (1990) *J. Virol. 64*, 3686–3693.
Tang, C. K., & Draper, D. E. (1989) *Cell 57*, 531–536.
Tang, C. K., & Draper, D. E. (1990) *Biochemistry 29*, 4434–4439.
ten Dam, E. B., Pleij, C. W. A., & Bosch, L. (1990) *Virus Genes 4*, 121–136.
ten Dam, E., Belkum, A. v., & Pleij, C. W. A. (1991) *Nucleic Acids Res. 19*, 6951.
Thomas, M., Bedwell, D. M., & Nomura, M. (1987) *J. Mol. Biol. 196*, 333–345.
Tinoco, I., Puglisi, J. D., & Wyatt, J. R. (1990) *Nucleic Acids Mol. Biol. 4*, 205–226.
Tsuchihashi, Z. (1991) *Nucleic Acids Res. 19*, 2457–2462.
Tzeng, T. H., Tu, C. L., & Bruenn, J. A. (1992) *J. Virol. 66*, 999–1006.
Van Belkum, A., Abrahams, J. P., Pleij, C. W. A., & Bosch, L. (1985) *Nucleic Acids Res. 13*, 7673–7686.
Van Belkum, A., Wiersema, P. J., Joordens, J., Pleij, C. W. A., Hilbers, C. W., & Bosch, L. (1989) *Eur. J. Biochem. 183*, 591–601.
Van de Van, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1–38.
Varani, G., & Tinoco, I. (1991) *Q. Rev. Biophys. 24*, 479–532.
Vartikar, J. V., & Draper, D. E. (1989) *J. Mol. Biol. 209*, 221–234.
Waring, R. B., & Davies, R. W. (1984) *Gene 28*, 277–291.
Weidner, H., Yuan, R., & Crothers, D. M. (1977) *Nature 266*, 193–194.
Weiss, R. B., Dunn, D. M., Shuh, M., Atkins, J. F., & Gesteland, R. F. (1989) *New Biol. 1*, 159–169.
Westhof & Jaeger (1992) *Curr. Opin. Struct. Biol.* (in press).
Wills, N. M., Gesteland, R. F., & Atkins, J. F. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 6991–6995.
Wilson, W., Braddock, M., Adams, S. E., Rathjen, P. D., Kingsman, S. M., & Kingsman, A. J. (1988) *Cell 55*, 1159–1169.
Woese, C. R., & Gutell, R. R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3119–3122.
Wu, H.-N., Wang, Y.-J., Hung, C.-F., Lee, H.-J., & Lai, M. M. C. (1992) *J. Mol. Biol. 223*, 233–245.
Wyatt, J. R., Puglisi, J. D., & Tinoco, I. (1990) *J. Mol. Biol. 214*, 455–470.
Yoshinaka, Y., Katoh, I., Copeland, T. D., & Oroszlan, S. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1618–1622.
Yu, G.-L., Bradley, J. D., Attardi, L. D., & Blackburn, E. H. (1990) *Nature 344*, 126–132.
Zuker, M. (1989) *Science 244*, 48–52.
Zuker, M., & Stiegler, T. (1981) *Nucleic Acids Res. 9*, 133–148.