# RNA Pseudoknots That Interact with Components of the Translation Apparatus

**Minireview**

Paul Schimmel
Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

It is not a long extrapolation to recognize that the unanticipated structural motifs that were found in tRNA are underpinnings of the RNA pseudoknot. The pseudoknot was proposed as a potentially widespread motif by Pleij et al. (1985) and several pieces of experimental data are consistent with its presence in some messenger, ribosomal, and viral RNAs (see also Rietveld et al., 1983). The present programs for energy minimization of predicted secondary structures exclude the pseudoknot motif (Zucker, 1989), but once it was recognized as at least a formal possibility, examples began to appear. Most recently, two papers describe experiments that not only support the presence of pseudoknot structures in different mRNAs, but also give evidence that these structures are specifically recognized by components of the translation apparatus (Tang and Draper, 1989; Brierley et al., 1989). A third and earlier study proposed that an RNA pseudoknot is recognized by a DNA binding protein that autogenously regulates translation of its mRNA (McPheeters et al., 1988).

The elucidation of the three-dimensional structure of tRNA was a benchmark because, among other contributions, it demonstrated the complexity of molecular form and shape possible with RNA molecules. As anticipated, it confirmed the pattern of hydrogen bonds in the cloverleaf secondary structure that was predicted from the sequence. Unexpectedly, it also revealed that the single-stranded loops in the proposed cloverleaf secondary structure are not passive elements. Instead, they bear nucleotides that are sites for the sophisticated interactions that stabilize the highly differentiated tertiary structure. These interactions include hydrogen bonds that connect the dihydrouridine loop to the TψC and variable loops. The structure also revealed that the four helical stems of the cloverleaf are stacked in pairs to form two longer, continuous helices which are arranged at approximately right angles to form an L-shaped molecule.

Pseudoknots are formed from stem–loop structures in which bases outside of a stem–loop are paired with those in the loop so as to create a second stem (Figure 1A). The second stem can be stacked upon the first to form a continuous coaxial helix. Thus, the hydrogen-bonded articulation of bases in a loop with bases in another part of the RNA, and the coaxial stacking of short helical stems to form a single longer helix, are general themes in tRNAs that are reiterated in pseudoknots. However, the pseudoknot has an added complexity that is not observed in tRNAs: the coaxial stacking of helices requires that single-stranded connecting nucleotides cross the grooves of the RNA double helix (Figure 1B). From a three-dimensional model it is clear that one crossing, over the deep groove, can be accomplished with two nucleotide units. The other crossing, over the shallow groove, requires (for a bridge of two nucleotides) a perturbation of the normal helical and, additionally or alternatively, bridging mononucleotide conformational parameters.

Tinoco's laboratory has done a careful analytical study of the structure and thermodynamic properties in solution of a nonadecanucleotide that has the potential to form a pseudoknot in which stems of 3 and 4 bp combine to make a 7 bp helix (Puglisi et al., 1988). In the proposed structure, three nucleotides bridge across the major groove and two are used to cross the minor groove. Strong evidence for the pseudoknot structure was obtained. The thermodynamic stabilities and hypochromicities of the pseudoknot, of a sequence variant that could not form one of the stems of the pseudoknot, and of the individual stems were investigated. These data showed that the stacking enthalpy was higher for the pseudoknot than for the molecules that contained the individual stems, although not as high as expected for a structure with seven contiguous base pairs. This observation and the reduced hypochromicity of the pseudoknot suggest a distortion in stem and loop portions which could be caused by the presumed two-base crossing of the minor groove.

As presented by Pleij et al. (1985), the pseudoknot proposal developed from structural mapping and modeling of the 3' end of plant viral RNAs that have tRNA-like properties. Thus, tobacco mosaic virus RNA, turnip yellow mosaic virus (TYMV), and brome mosaic virus RNAs are aminoacylated specifically with histidine, valine, and tyrosine, respectively, and each is recognized by the tRNA nucleotidyl transferase, which adds the CCA sequence at
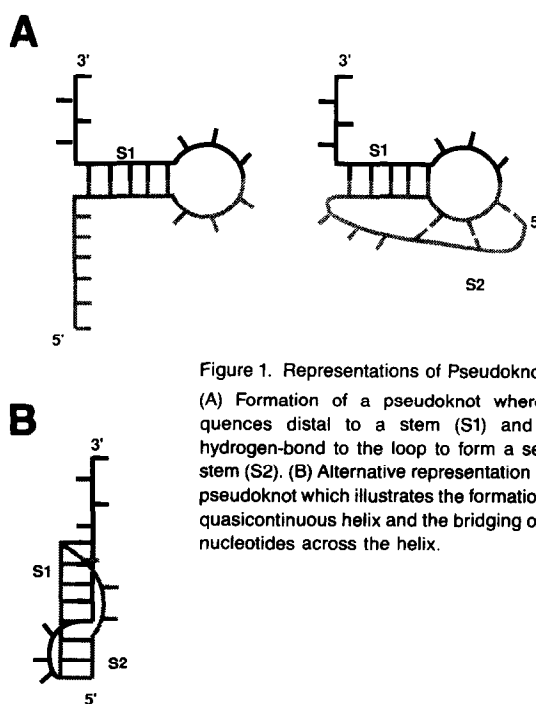
**A**



**B**



Figure 1. Representations of Pseudoknots

(A) Formation of a pseudoknot where sequences distal to a stem (S1) and loop hydrogen-bond to the loop to form a second stem (S2). (B) Alternative representation of the pseudoknot which illustrates the formation of a quasicontinuous helix and the bridging of loop nucleotides across the helix.
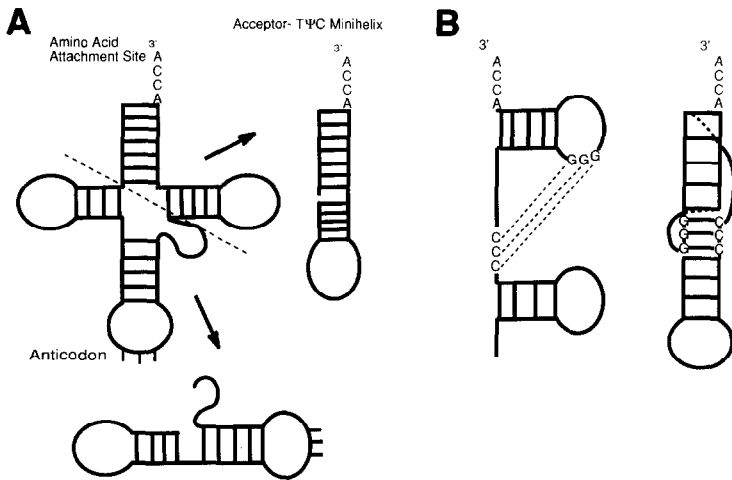
**A**



**B**



Figure 2. The 3' Ends of Plant Viral RNAs Have tRNA-like Properties

(A) Illustration of the way in which the four stems of the cloverleaf secondary structure of a tRNA are combined into two minihelices. In the case of E. coli alanine tRNA, the 12 bp acceptor–TψC minihelix can be efficiently aminoacylated, provided the minihelix encodes a critical G3:U70 base pair (Francklyn and Schimmel, 1989). (B) Generation of a 12 bp minihelix by formation of a pseudoknot. This structure at the 3' end of TYMV RNA can be combined with adjacent sequences to form a complete tRNA-like molecule (see Pleij et al., 1985).

the 3' terminus of all tRNA molecules (reviewed in Haenni et al., 1982).

Figure 2A shows how the four stems of the tRNA cloverleaf are combined into two continuous helices that constitute the two arms of the L-shaped three-dimensional structure. The acceptor–TψC stems form one continuous minihelix of 12 bp. Seven of these pairs are derived from the acceptor and five from the TψC helix. The problem is how to form an equivalent minihelix from sequences, for example, at the 3' terminus of TYMV RNA. A conventional secondary structure analysis suggested that this segment encodes two stem–loop segments with 5 and 4 bp, respectively. However, by implementation of the pseudoknot format, the desired 12 bp minihelix can be constructed (Figure 2B) and combined with adjoining sequences to form a tRNA-like structure (not shown).

Whether the minihelix format is sufficient for aminoacylation of TYMV or other viral RNAs is not known. It is worth noting that, at least for alanine, the acceptor–TψC minihelix and the 7 bp acceptor stem can be charged by the cognate aminoacyl tRNA synthetase. This is because the major determinant for the identity of an alanine tRNA is a single base pair that is located in the acceptor helix, so that the rest of the tRNA structure is dispensable for aminoacylation (Hou and Schimmel, 1988; Francklyn and Schimmel, 1989). The results with alanine tRNA raise the possibility that, for at least some viral RNAs, synthetase recognition requires only a pseudoknot structure that bears resemblance to the 3' helical half (or less) and not to the entire tRNA.

The tRNA-like structures at the 3' ends of plant viral RNAs provide one example where the pseudoknot format may be necessary to form a substrate that is specifically recognized by an enzyme. Dreher and Hall (1988) have shown that a three base substitution that disrupts part of the proposed acceptor stem of brome mosaic virus RNA simultaneously impairs aminoacylation and nucleotidyl transferase activities. To prove the dependence of synthetase recognition on the pseudoknot structure would require evaluation of an ensemble of mutant RNAs (cf. Hou and Schimmel, 1988). These mutants should be designed to determine whether only those which have compensa-

tory base changes that preserve the pseudoknot can be aminoacylated (cf. Dreher and Hall, 1988). This is analogous to the phylogenetic approach, which has been successfully used to test predictions of secondary structure in large RNA molecules (Noller, 1984; James et al., 1988).

In the case of aminoacylation of viral RNAs, however, it is not just a question of formation of the pseudoknot structure, but of pseudoknot-dependent presentation of nucleotide determinants for protein recognition. Consequently, both mutations that disrupt and others that preserve the pseudoknot may inactivate aminoacylation and, therefore, must be classified and studied separately. Obviously, this analysis would be more effective if the sites for synthetase recognition in the associated tRNA were already known, but to date, examples for which the sites for tRNA identity are well defined have no viral RNA counterpart.

The first evidence for a pseudoknot motif in protein binding to an mRNA came from experiments by McPheeters et al. (1988), who dissected the translational regulatory site on bacteriophage T4 gene 32 mRNA. Gene 32 protein binds to single-stranded DNA and has a functional role in T4 DNA transactions; excess amounts of the protein are free to bind to the operator region of its mRNA and thereby block initiation of translation (Gold, 1988). By a combination of methods which were applied to free and complexed RNA, nucleation of binding was suggested to be promoted by a pseudoknot that is approximately 40 nucleotides upstream of the initiation codon. A comparison with the mRNA sequences of the related T2 and T6 phages provided phylogenetic evidence for conservation of the pseudoknot structure (McPheeters et al., 1988). To understand the exact structure required for presentation of the gene 32 protein binding site, direct and quantitative measurements of protein binding to an RNA with the presumed pseudoknot, and to mutants that alter its structure, will have to be carried out.

Two recent studies have used mutational analysis to demonstrate the role of pseudoknots in entirely different systems. One is an investigation of protein recognition of a proposed pseudoknot motif in the 5' region of the E. coli α operon mRNA (Tang and Draper, 1989). This polycistronic mRNA encodes four ribosomal proteins and the α

subunit of RNA polymerase. Ribosomal protein S4 is one of the encoded proteins. The translation of this mRNA is regulated by S4. This and other ribosomal proteins that are known to be translational repressors bind to a specific mRNA structure and also have a binding site on 16S or 23S ribosomal RNA (Lindahl and Zengel, 1986; Thomas et al., 1987).

Previous structural mapping experiments by Draper and co-workers suggested a model for the 5' region of the mRNA which constitutes the S4 binding site; it envisions a hairpin helix and loop that encompasses nucleotides 19 to 72. The loop sequence GGGC at position 49–52 is proposed to pair with the downstream sequence GCCC at position 98–101 so as to create a pseudoknot. The structural model and its relevance to S4 binding was tested by construction of mutants that alternately disrupt and restore (by compensatory changes) the presumed pseudoknot. The mutant RNAs were synthesized by enzymatic methods and assayed directly for S4 binding in vitro (Tang and Draper, 1989).

Nucleotide substitutions that disrupt the proposed pseudoknot also weaken S4 binding. Several compensatory mutations which restore base pairing also restore binding affinity. For example, the $G_{49}G_{50} \rightarrow CC$ and $C_{100}C_{101} \rightarrow GG$ mutants are each recognized by S4 with an 8- to 10-fold lower affinity. However, the double mutant which combines both changes restores the putative pairing and the binding affinity. These and similar data with over 30 mutant RNAs led to a revised and more complex structure which is visualized as a double pseudoknot.

To my knowledge this is the first example of the use of direct protein–RNA binding measurements to define an RNA structure. The changes in affinity that accompany disruption of the proposed structure are in all cases relatively small in terms of binding energy. For example, many of the changes in the S4 association constant are less than 10-fold—a change which itself corresponds to only 1.4 kcal per mole. A change of this magnitude could be due to perturbation of a van der Waals interaction. There are no mutations in the proposed pseudoknot that totally eliminate binding and would, therefore, be analogous to point mutations in a tRNA that eliminate aminoacylation in vitro (Hou and Schimmel, 1988; Schulman and Pelka, 1988).

Possibly the most critical nucleotide determinants for S4 recognition within the pseudoknot have not been identified in the mutational analysis, as they have in the synthetase–tRNA system. Alternatively, the interactions of α operon mRNA with S4 may be distributed over many sites in the structure, so that any given alteration produces only a small change in affinity. It is noteworthy that one of the mutations that disrupts a base pair in the proposed pseudoknot structure cannot be rescued by a second site mutation which restores pairing. This could be a site where S4 interacts directly with a specific base pair, and contributes a small incremental stability to the complex. Ongoing experiments will assess the relationship between alterations that disrupt the pseudoknot-dependent protein–RNA interaction in vitro and the extent of translational repression in vivo. Regardless of the detailed interpretation of the experiments, the analysis of RNA structure by these

methods is instructive and has provided one of the first examples of a functional probe for pseudoknot formation.

The second study provides evidence that pseudoknot formation in a viral mRNA is required for frameshift suppression of a termination codon that, in turn, allows a fusion protein to be synthesized from two overlapping reading frames. Earlier work had demonstrated a role for frameshifting in the production of some retroviral gag-pol or gag-pro-pol fusion proteins (Jacks et al., 1988; Wilson et al., 1988, and references therein). Most commonly, termination occurs at the gag stop codon to yield virus core protein. Occasional "–1" frameshifting and subsequent cleavage of the resulting fusion protein is the mechanism for production of the viral reverse transcriptase (a pol gene product). Frameshift mechanisms may also be operative in the production of reverse transcriptases encoded by retrotransposons in yeast (e.g., Clare et al., 1988) and Drosophila (e.g., Marlor et al., 1986).

Varmus and co-workers established that, in Rous sarcoma virus mRNA, only 147 nucleotides that encode the site for frameshifting are necessary (Jacks et al., 1988). The sequence at the frameshift site is A AAU UUA, where the "0" reading frame is indicated. Operationally there is simultaneous –1 slippage of a UUA-reading tRNA[Leu] (bound to the ribosomal A-site) and an AAU-reading tRNA[Asn] (bound to the P-site), which results in a double frameshift. In the –1 position these tRNAs are proposed to read their respective codons by two instead of three bases. Slippage of these tRNAs is dependent on the formation of a hairpin stem–loop structure that is immediately downstream. Thus, in addition to other possible mechanisms (see Wilson et al., 1988), this work established that –1 frameshifting can be promoted by mRNA secondary structure.

The work of Brierley et al. (1989) is based on studies of a nonretroviral system, avian coronavirus infectious bronchitis virus (IBV). Two long open reading frames overlap by 42 nucleotides, with the second frame shifted by –1 relative to the first. A –1 frameshift results in the production of a fusion protein. An 86 nucleotide element that spans the overlap region is sufficient to promote frameshifting, even in a heterologous context. This element encodes a stem–loop structure that starts six nucleotides downstream from the proposed site of frameshifting, near the end of the first reading frame. Located 30 nucleotides further downstream from the 3' end of the putative stem is a sequence of seven bases that are complementary to the hairpin loop. Thus, the overlap region encodes a sequence that could fold into a pseudoknot.

Support for the pseudoknot structure was sought by analysis of mutants that alternately disrupt and restore the proposed structure. Mutations that disrupt either stem of the pseudoknot severely reduce frameshifting, while compensatory mutations that restore the pseudoknot also restore frameshifting. The pseudoknot inferred by this analysis is a quasicontinuous helix of 16 bp, and the necessary bridging by unpaired bases across the deep and shallow grooves is sterically feasible. Of the retrovirus and related systems suspected to use –1 frameshifting, Brierley et al. (1989) found that over half have the potential for pseudoknot formation immediately downstream of the lo-

cation of the frameshift. This includes three of the four systems where frameshifting has actually been confirmed. It is perhaps significant that Jacks et al. (1988) had shown that frameshifting is attenuated upon deletion of a region downstream of the critical stem in Rous sarcoma virus mRNA.

The mechanism for the pseudoknot-induced frameshift is unknown but there are several points worth noting. Mechanisms for frameshifting with natural as opposed to mutant tRNAs have been studied in bacteria and eukaryotes (reviewed in Roth, 1981; Craigen and Caskey, 1987). In bacteria, frameshift suppression can occur as a result of translational pausing, which is caused, for example, by starvation for an amino acid. In this case, frameshifting results from the use of a surrogate charged tRNA in place of the correct charged species (which is in limiting amounts). Thus, a transiently unoccupied A-site on the ribosome may be "read" by a noncognate tRNA and that event can be accompanied by a frameshift and a missense substitution. The frameshifts that lead to fusion proteins in the retroviral and IBV examples are different in that the double frameshift event is not accompanied by missense substitutions. This suggests that a pseudoknot does not simply produce an unoccupied A-site.

It is not known whether a single hairpin stem equal in length to the elongated pseudoknot structure would be as effective in inducing a frameshift in the system studied by Brierley et al. (1989). The authors show that the upstream stem alone induces frameshifting, but at a much lower efficiency than when the downstream sequences are allowed to form the second stem that results in a pseudoknot minihelix. Thus, the frequency of frameshifting may in principal be fine-tuned by the size and detailed structure of the pseudoknot.

I can suggest one possible advantage to the pseudoknot format over a standard hairpin helix of equivalent size. In the work of Puglisi et al. (1988), the thermal stability of the pseudoknot helix was not greater than that of the most stable of the two stems from which it was assembled. If this is a general principle, then pseudoknots provide a way to generate minihelices that have lower stabilities than their counterparts, which are assembled as unknotted continuous hairpins with the same number of base pairs. The reduction in stability may be critical for efficient movement of ribosomes through an element of secondary structure in an mRNA.

Even the spatial location of the pseudoknot relative to the frameshift site in IBV RNA is sharply constrained—to less than three nucleotides (Brierley et al., 1989). This emphasizes the importance of structural detail and context for the biological function of the pseudoknot in translation. In this and other respects, it has the properties of a substrate that is specifically recognized and acted upon by an enzyme sensitive to details of molecular shape and spacing of functional groups. There is no evidence that a component of the translation apparatus (e.g., ribosomes) performs such a recognition function before triggering the double frameshift, but operationally the result is the same and the possibility has to be at least formally considered.

Before the structure was solved, many predictions were made of the folding of tRNA. None of them correctly predicted the tertiary structure that was elucidated by x-ray diffraction analysis. This structure is now the basis for designing experiments to understand the interactions of tRNAs with proteins (Schimmel, 1989). In the systems described above, the proposed pseudoknot secondary structures are highly schematic. NMR analyses of "simple" pseudoknots are providing some of the structural parameters that will guide model building and design (cf. Wyatt et al., 1989). Further thermodynamic studies may allow more accurate energy estimates that can be the basis for including pseudoknots in programs that compute energy-minimized secondary structures (cf. Zucker, 1989). However, the data of Tang and Draper (1989) suggest a more complex arrangement for the $\alpha$ operon mRNA pseudoknot than the basic motif proposed by Pleij et al. (1985). In general, it is likely that tertiary structural features have an important role in the protein recognition and translational frameshifting that is now associated with RNA pseudoknots. Although in principle the analytical tools are available, it is these tertiary features that will be most difficult to work out.

### References

Brierley, I., Digard, P., and Inglis, S. C. (1989). Cell 57, 537–547.

Clare, J. J., Belcourt, M., and Farabaugh, P. J. (1988). Proc. Natl. Acad. Sci. USA 85, 6816–6820.

Craigen, W. J., and Caskey, C. T. (1987). Cell 50, 1–2.

Dreher, T. W., and Hall, T. C. (1988). J. Mol. Biol. 201, 41–55.

Francklyn, C., and Schimmel, P. (1989). Nature 337, 478–481.

Gold, L. (1988). Annu. Rev. Biochem. 57, 199–233.

Haenni, A.-L., Joshi, S., and Chapeville, F. (1982). Prog. Nucl. Acid Res. Mol. Biol. 27, 85–102.

Hou, Y.-M., and Schimmel, P. (1988). Nature 333, 140–145.

Jacks, T., Madhani, H. D., Masiarz, F. R., and Varmus, H. E. (1988). Cell 55, 447–458.

James, B. D., Olsen, G. J., Liu, J., and Pace, N. R. (1988). Cell 52, 19–26.

Lindahl, L., and Zengel, J. M. (1986). Annu. Rev. Genet. 20, 297–326.

Marlor, R. L., Parkhurst, S. M., and Corces, V. G. (1986). Mol. Cell. Biol. 6, 1129–1134.

McPheeters, D. S., Stormo, G. D., and Gold, L. (1988). J. Mol. Biol. 201, 517–535.

Noller, H. F. (1984). Annu. Rev. Biochem. 53, 119–162.

Pleij, C. W. A., Rietveld, K., and Bosch, L. (1985). Nucl. Acids Res. 13, 1717–1731.

Puglisi, J. D., Wyatt, J. R., and Tinoco, I., Jr. (1988). Nature 331, 283–286.

Rietveld, K., Pleij, C. W. A., and Bosch, L. (1983). EMBO J. 2, 1079–1085.

Roth, J. R. (1981). Cell 24, 601–602.

Schimmel, P. (1989). Biochemistry 28, 2747–2759.

Schulman, L. H., and Pelka, H. (1988). Science 242, 765–768.

Tang, C. K., and Draper, D. E. (1989). Cell 57, 531–536.

Thomas, M., Dedwell, D. M., and Nomura, M. (1987). J. Mol. Biol. 196, 333–345.

Wilson, W., Braddock, M., Adams, S. E., Rathjen, P. D., Kingsman, S. M., and Kingsman, A. J. (1988). Cell 55, 1159–1169.

Wyatt, J. R., Puglisi, J. D., and Tinoco, I., Jr. (1989). In Molecular Biology of RNA, UCLA Symp. Mol. Cell Biol., Volume 94, T. R. Cech, ed. (New York: Alan R. Liss), pp. 25–32.

Zucker, M. (1989). Science 244, 48–52.