1130–1135 *Nucleic Acids Research, 1997, Vol. 25, No. 6*

An NMR and mutational study of the pseudoknot within the gene *32* mRNA of bacteriophage T2: insights into a family of structurally related RNA pseudoknots

Zhihua Du and David W. Hoffman*

Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX 78712, USA

Received December 31, 1996; Accepted February 4, 1997

ABSTRACT

NMR methods were used to investigate a series of mutants of the pseudoknot within the gene *32* messenger RNA of bacteriophage T2, for the purpose of investigating the range of sequences, stem and loop lengths that can form a similar pseudoknot structure. This information is of particular relevance since the T2 pseudoknot has been considered a representative of a large family of RNA pseudoknots related by a common structural motif, previously referred to as 'common pseudoknot motif 1' or CPK1. In the work presented here, a mutated sequence with the potential to form a pseudoknot with a 6 bp stem2 was shown to adopt a pseudoknot structure similar to that of the wild-type sequence. This result is significant in that it demonstrates that pseudoknots with 6 bp in stem2 and a single nucleotide in loop1 are indeed feasible. Mutated sequences with the potential to form pseudoknots with either 5 or 8 bp in stem2 yielded NMR spectra that could not confirm the formation of a pseudoknot structure. Replacing the adenosine nucleotide in loop1 of the wild-type pseudoknot with any one of G, C or U did not significantly alter the pseudoknot structure. Taken together, the results of this study provide support for the existence of a family of similarly structured pseudoknots with two coaxially stacked stems, either 6 or 7 bp in stem2, and a single nucleotide in loop1. This family includes many of the pseudoknots predicted to occur downstream of the frameshift or readthrough sites in a significant number of viral RNAs.

INTRODUCTION

A pseudoknot is a structural element of RNA formed when a stretch of nucleotides within a single-stranded loop region base pairs with a complementary sequence outside that loop (for reviews, see 1–5). Although 14 types of topologically distinct pseudoknots are possible according to this broad definition (3), most of the pseudoknots documented today are of the so-called H(airpin)-type, in which a stretch of nucleotides in the loop of a hairpin and adjacent to the stem region base pairs with a complementary region outside of that hairpin. Since the studies by Pleij and co-workers proposing the presence of an H-type pseudoknot at the 3′ end of turnip yellow mosaic virus (TYMV) RNA (6–7), a large body of evidence has been accumulated indicating that pseudoknots within messenger RNAs play an important role in a variety of critical biological processes, such as regulation of protein expression in viral systems by ribosomal frameshifting or readthrough. The widespread occurrence of the H-type pseudoknots suggests that they may have common structural features which minimize the energy of their folding, thus forming a basis for their frequent occurrence in natural systems. Coaxial stacking of the two separate helical stems to form a single pseudo-continuous double helix has been proposed as a stabilizing feature in pseudoknots (7). This type of tertiary interaction has been confirmed by NMR studies on a model oligoribonucleotide (8) and more recently on a natural sequence RNA (9). It appears likely, therefore, that coaxial stacking of the two stem regions represents a general recurring theme in the folding of the H-type pseudoknots.

A profound consequence of the coaxial stacking of the two helical stems (S1 and S2) is the generation of two unequal connecting loops (L1 and L2), with L1 crossing the deep major groove of stem S2 and L2 crossing the shallow minor groove of stem S1 (7). This imposes different constraints on the length of each of the two loops. A study of the loop size requirements on a series of model oligoribonucleotides with the potential to form a pseudoknot with a 3 bp stem1 (S1) and 5 bp stem2 (S2) showed that the minimum loop lengths were 3 nt for L1 and 4 nt for L2 (10). However, the minimum lengths of the loops are dependent on the number of base pairs in each of the stems. In our recent NMR study of the pseudoknot from the gene *32* mRNA of bacteriophage T2, we found that loop1 consists of only a single nucleotide, which crosses the major groove of the 7 bp stem2 (9). An analysis of reported pseudoknot-forming sequences revealed that several features of the bacteriophage T2 pseudoknot are frequently repeated in naturally occurring pseudoknots: while the lengths of stem1 and loop2 are varied, the length of stem2 is often 6 or 7 bp, and loop1 often contains only a single adenosine nucleotide (9). It is particularly interesting that in an ideal A-form

*To whom correspondence should be addressed. Tel: +1 512 471 7859; Fax: +1 512 471 8696; Email: dave@noddy.cm.utexas.edu

Figure 1. Models of a standard A-form RNA helix (left) and the T2 wild-type pseudoknot from Du *et al*. (9) (right). In the T2 pseudoknot, stem1 (nucleotides 3–7 and 16–20) is colored yellow, stem2 (nucleotides 9–15 and 28–34) is colored green, loop1 (nucleotide A8) is colored red, loop2 (nucleotides 21–27) is colored purple, and the 5′ and 3′ ends (nucleotides 1–2 and 35–36) are colored blue, with the 5′ end of the molecule on top. The sequence of the double-stranded A-form RNA helix (left) corresponds to the quasi-continuous helix formed by the two coaxially stacked stems of the T2 pseudoknot. Note that an extra U–A base pair is added (colored blue) to illustrate the distance across the major groove for an 8 bp stem. The distances (from an O3′ atom on one strand to a phosphorus atom on the other strand) span the major groove of a 5, 6, 7 or 8 bp stem, illustrated by a red solid bar with white number indicating the number of base pairs being crossed. The distances are 13.43, 11.36, 11.18 and 13.64 Å for 5, 6, 7 and 8 bp, respectively. A comparison of these two representations reveals how it is feasible for the a pseudoknot to use a single nucleotide loop1 to bridge the major grove of a 6 or 7 bp stem2 without substantial distortion of the quasi-continuous A-form helical stems. The figure was made using the program MIDAS (17,18).

RNA helix, the distance between two phosphate groups across the deep major groove reaches a minimum when 6 or 7 bp are bridged (Fig. 1) (7). This 'coincidence' immediately suggests a general theme for the tertiary folding of naturally occurring H-type pseudoknots in which a minimal number of nucleotide(s) is used to span the deep major groove of 6 or 7 bp helical stem.

The purpose of the work presented here is to address several significant issues regarding the proposed CPK1 family of structurally related RNA pseudoknots. First, direct structural (NMR or crystallographic) evidence for the existence of a pseudoknot with a 6 bp stem2 and a single nucleotide loop1 is absent prior to the present work, although a significant number of naturally occurring pseudoknots have sequences that have the potential to conform to this motif (9). Second, although the apparent strong bias for an adenosine nucleotide in loop1 among naturally occurring pseudoknots is intriguing (9), the question as to whether the other 3 nt (G, C and U) can substitute for the wild-type loop1 nucleotide and still maintain the pseudoknot structure requires further investigation. To address these issues, we have systematically mutated the native sequence of the bacteriophage T2 gene *32* mRNA pseudoknot, and the structural consequences of the mutations have been analyzed by one- and two-dimensional NMR methods.

MATERIALS AND METHODS

RNA synthesis and purification

RNA molecules with the sequences shown in Figure 2 were transcribed using T7 RNA polymerase and synthetic DNA templates, as previously described (9). The DNA templates consisted of a double-stranded 18 bp T7 promoter sequence and a single-stranded coding sequence. The RNA was then separated from transcripts of incorrect size by electrophoresis on 20% polyacrylamide gels under denaturing conditions (8 M urea). RNA was visualized by UV shadowing, and removed from the gel using a BioRad model 422 electroeluter. The RNA was further purified by repeated ethanol precipitation, and finally passed through a Sephadex G25 gel filtration column in 1 mM phosphate buffer and lyophilized. A typical yield was 1 mg of purified RNA/10 ml transcription volume.

NMR experiments

NMR spectra of the RNA sequences were collected at 500 MHz using a Varian Unity-Inova spectrometer. Samples typically contained 10–15 mg of RNA dissolved in 550 µl of 10 mM Na/K phosphate buffer in 90% H_2O/10% D_2O at pH 6.8. Two-dimensional nuclear Overhauser effect (NOE) spectra were acquired using the jump and return method (16) with a mixing time of 280 ms. Two-dimensional NMR spectra were acquired in the phase-sensitive mode, typically acquired with 512 blocks of 1024 complex points, using a sweep width of 12 204 Hz in each dimension. Two-dimensional data sets in 90% H_2O/10% D_2O solvent were acquired at 10°C. Spectra were referenced to the solvent water resonance at 4.90 p.p.m. at 10°C.

RESULTS

Variants of the natural sequence of the bacteriophage T2 pseudoknot were prepared and investigated, with sequences illustrated in Figure 2. Nucleotides are numbered to be consistent with the wild-type sequence, and the mutants are named as indicated in the figure. In each case, 'dangling' nucleotides were included on the 3′ and 5′ ends of the pseudoknots. We have found that these nucleotides stabilize the pseudoknots, most likely through base stacking interactions, and should be considered as an integral part of the pseudoknot structure (9,11). Results for each of the RNAs will be discussed in turn.

Six base pairs in stem2

A sequence with the potential to form a pseudoknot with 6 bp in stem2 was prepared by deleting 1 bp from the center of the seven base pair stem2 of the wild-type bacteriophage T2 gene *32* mRNA pseudoknot. This potential pseudoknot-forming sequence was termed PK-STEM6 (Fig. 2). As was the case for the wild-type pseudoknot, this mutant yielded high-quality NMR spectra with sharp and disperse resonances (Fig. 3). Assignment of the imino proton resonances was straightforward using the two-dimensional NOE spectrum (Fig. 4a). Sequential imino-to-imino NOE cross peaks were observed for the imino protons of the Watson–Crick base pairs in the helical stems, including the closing base pair of stem2. Of particular significance is the NOE

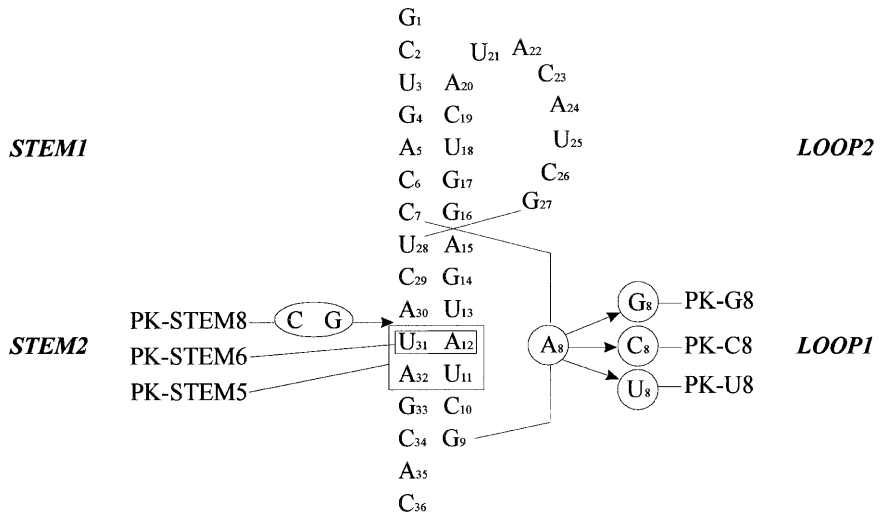

Figure 2. Schematic diagrams of the nucleotide sequences of the wild-type and mutant bacteriophage T2 gene*32* mRNA pseudoknots used in this work. Nucleotides are numbered as in the wild-type sequence. Box, circle and oval indicate deletion, substitution and insertion mutations respectively. In RNA PK-STEM6, the A12–U31 base pair of the wild-type pseudoknot was deleted. In RNA PK-STEM5, the A12–U31 and U11–A32 base pairs were deleted. In RNA PK-STEM8, a C–G base pair was inserted into stem2. In mutants PK-G8, PK-C8 and PK-U8, the A8 nucleotide of the wild-type sequence was changed to G, C or U, respectively. The diagrams are drawn in such a way as to show the base pairing that occurs within the RNAs, and the coaxial stacking of the stems in the wild-type as well as PK-STEM6, PK-G8, PK-C8 and PK-U8 mutant pseudoknots. The pseudoknot from bacteriophage T6 has the same sequence as that from T2.

cross peak across the junction of the two stems (labeled G16–U28). This NOE cross peak, along with the cross peaks between imino protons and amino, ribose H1′, adenosine H2 and pyrimidine H5 protons (data not shown), indicates that the two helical stems are stacked coaxially to form a single pseudo-continuous A-form helix. The striking feature that loop1 consists of only a single nucleotide, A8, is particularly clear since the adjacent nucleotides (C7 and G9) are involved in stable G–C base pairs that are well defined by the NMR data. Although the chemical shifts of some of the protons (especially those belonging to the residues immediately adjacent to the deleted base pair) in the PK-STEM6 sequence are somewhat different from those of the wild-type sequence, the overall NOE connectivities indicate that the PK-STEM6 sequence adopts a pseudoknot structure which is remarkably similar to the wild-type bacteriophage T2 pseudoknot. These NMR results provide the first direct structural evidence for the CPK1 pseudoknots with a 6 bp stem2.

Five or eight base pairs in stem2

Two additional mutants of the wild-type bacteriophage T2 pseudoknot were constructed to further investigate the influence of the length of stem2 on the tertiary folding of the pseudoknot. In mutant PK-STEM5, 2 bp were deleted from the wild-type pseudoknot, and in mutant PK-STEM8, an extra G–C base pair was inserted between U13–A30 and A12–U31 (Fig. 2). The imino proton region of the one-dimensional NMR spectra of each mutant contained resonances with some chemical shifts similar to resonances of stem1 and stem2 of the wild-type pseudoknot (Fig. 3). However, the two-dimensional NOE spectra could not confirm the formation of a pseudoknot structure, and can perhaps be best accounted for by the presence of more than one RNA conformation in solution. Therefore, we are unable to provide clear evidence that a stable pseudoknot can be formed with 5 or 8 bp in stem2 while maintaining a single nucleotide in loop1. This is consistent with the results of previous studies of pseudoknot-

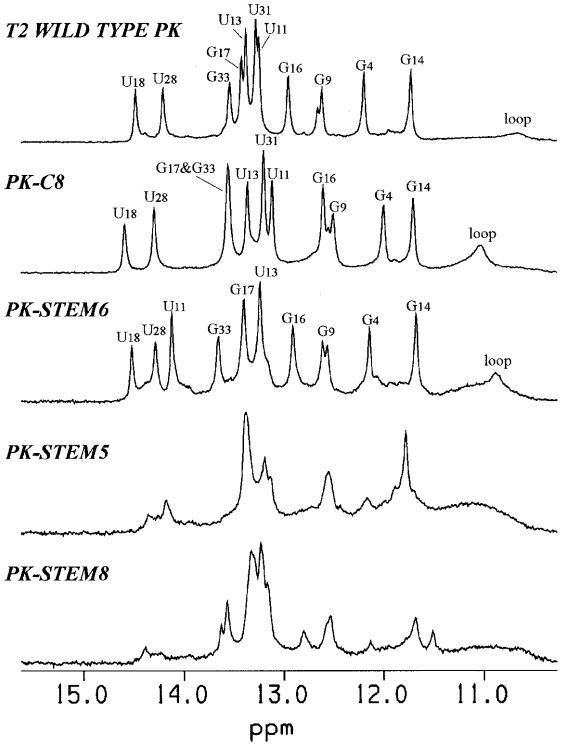

Figure 3. A comparison of the imino protons resonances of the wild-type bacteriophage T2 gene *32* mRNA pseudoknot and some of the mutant pseudoknots investigated in this study. Shown are spectra of the wild-type as well as PK-C8, PK-STEM6, PK-STEM5 and PK-STEM8 mutant pseudoknots. The NMR spectra were recorded at 10°C in 90% H_2O/10% D_2O solvent, 10 mM phosphate buffer, pH 6.8. The broad resonances designated 'Loop' are most likely due to the remaining unassigned imino protons: G1, U3, U21 and/or U25. In each pseudoknot, the imino resonance of G9 appears as a doublet, a feature also observed in the homologous bacteriophage T4 gene *32* mRNA pseudoknot, and perhaps a result of structural heterogeneity at the 3′ end of the molecules (9,11).

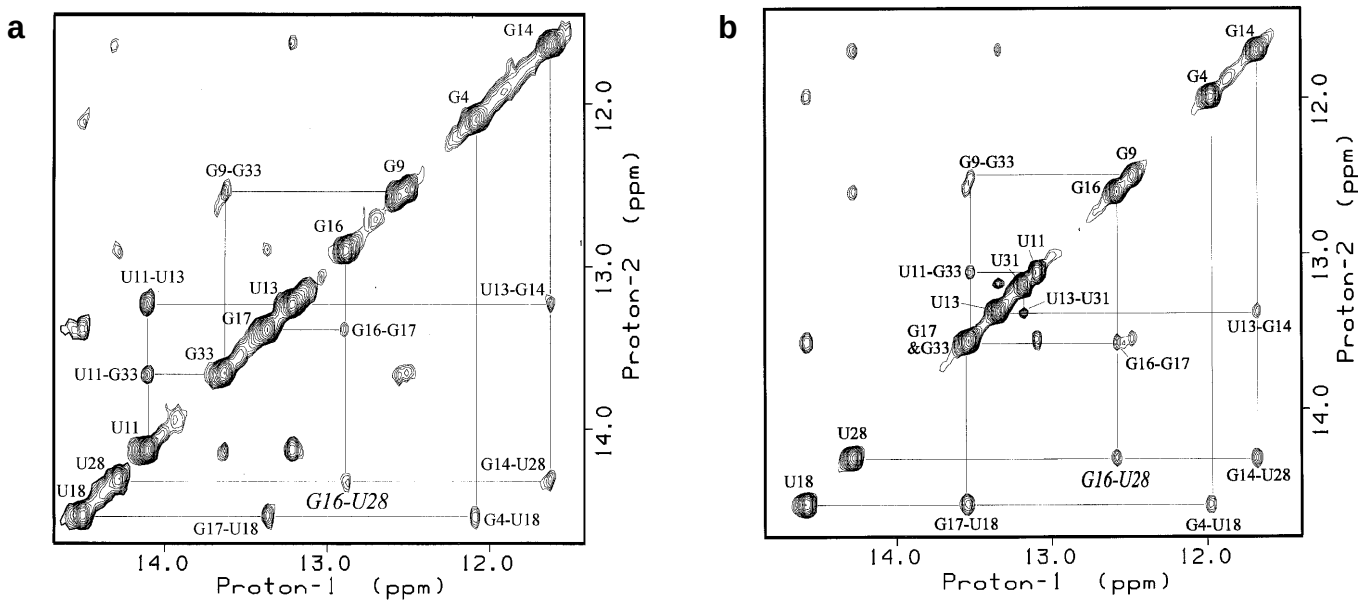

Figure 4. Sections of the two-dimensional NOE spectrum (280 ms mixing time) of the PK-STEM6 (**a**) and PK-C8 (**b**) RNA pseudoknot in 90% H_2O/10% D_2O solvent, 10 mM phosphate buffer, pH 6.8, at 10°C. The solvent resonance was suppressed by the jump and return method. This section of the 2D NOE spectrum shows the cross-peaks between imino protons. The assigned cross-peaks are labeled. Of particular interest is the NOE between the imino protons of G16 and U28, across the junction between the two stems (labeled by italic and larger font), which provides strong evidence that the two stems are coaxially stacked.

forming RNA sequences, in which a minimum of 2 or 3 nt were shown to be required in loop1 when stem2 contained 5 bp (10).

Pseudoknots differing in the identity of the loop1 nucleotide

In our previous work proposing the existence of the CPK1 family of pseudoknots, we noted that a substantial majority of the predicted CPK1 pseudoknots had a single adenosine nucleotide in loop1 (9). This observation raised the possibility that the loop1 adenosine may be conserved for structural or functional reasons. Structurally, an adenosine residue may be favored as the loop1 nucleotide if it participates in a specific tertiary interaction with the residues of the helical stem2. To investigate this possibility, mutants of the wild-type bacteriophage T2 pseudoknot were prepared where the loop1 adenosine was replaced by G, C or U (designated PK-G8, PK-C8 and PK-U8).

The one-dimensional imino proton spectra of the PK-G8, PK-C8 and PK-U8 mutant sequences are similar to the wild-type pseudoknot (Fig. 3, noted that the spectrum of PK-C8 is shown as a representative) and this observation immediately suggests that each of these mutants is similar in structure to the wild-type pseudoknot. The base pairings within the mutant pseudoknots were confirmed using two-dimensional NOE spectra, where the now familiar patterns of sequential imino-to-imino proton NOEs are observed through both stems, and across the junction of the stems (Fig. 4b, only the spectrum of PK-C8 is shown). As was the case in the wild-type and PK-STEM6 pseudoknots, the NOE cross peak between the imino protons of nucleotides G16 and U28 (Fig. 4b), as well as cross peaks to amino, ribose H1′, adenosine H2 and pyrimidine H5 protons, provides strong evidence that the two stems are coaxially stacked. These NMR data indicate that substitution of the loop1 adenosine by any one of the other three nucleotides (G, C or U) is tolerated, without disruption of the overall tertiary folding of the pseudoknot.

CONCLUSIONS

The present study provides further evidence for the existence of a family of similarly structured RNA pseudoknots, containing coaxially stacked stems and a 1 nt loop1 that spans the major groove of a 6 or 7 bp stem2. While pseudoknotted structures could not be unambiguously detected for the PK-STEM5 and PK-STEM8 RNAs, the PK-STEM6 sequence was shown to adopt a structure similar to that of the wild-type T2 bacteriophage pseudoknot. This result is significant in that it demonstrates that the naturally occurring pseudoknots predicted to have a base pairing arrangement of 6 bp in stem2 are indeed feasible. Many of these naturally occurring pseudoknots are of critical importance since they are associated with ribosomal frameshifting processes in retroviruses and other RNA viruses (Fig. 5). One such pseudoknot, located downstream of the *gag–pro* frameshift site in simian retrovirus type-1 (SRV-1), has been proposed to have 6 bp in each of the two stems and a single adenosine nucleotide in loop1 (9,12), a model supported by the results of the present mutational–structural studies, as well as recent NMR studies conducted in our laboratory (manuscript in preparation).

Regarding the apparent common occurrence of an adenosine nucleotide in loop1 of many of the predicted CPK1 pseudoknots (9 and Fig. 5 herein), the present studies show that any 1 of the other 3 nt (G, C or U) can serve as the loop1 nucleotide while preserving the pseudoknot structure. This result argues against the loop1 nucleotide being involved in a tertiary interaction essential for pseudoknot formation; merely embedding the relatively hydrophobic base(s) of the loop1 nucleotide(s) in the major groove of stem2 may provide a substantial stabilizing force for the pseudoknot folding. The present results are consistent with a recent study investigating the role of the pseudoknot component in regulating the SRV-1 *gag–pro* ribosomal frameshifting efficiency, where ten Dam and co-workers found that substitution of the adenosine residue in loop1 of the SRV-1 pseudoknot by any

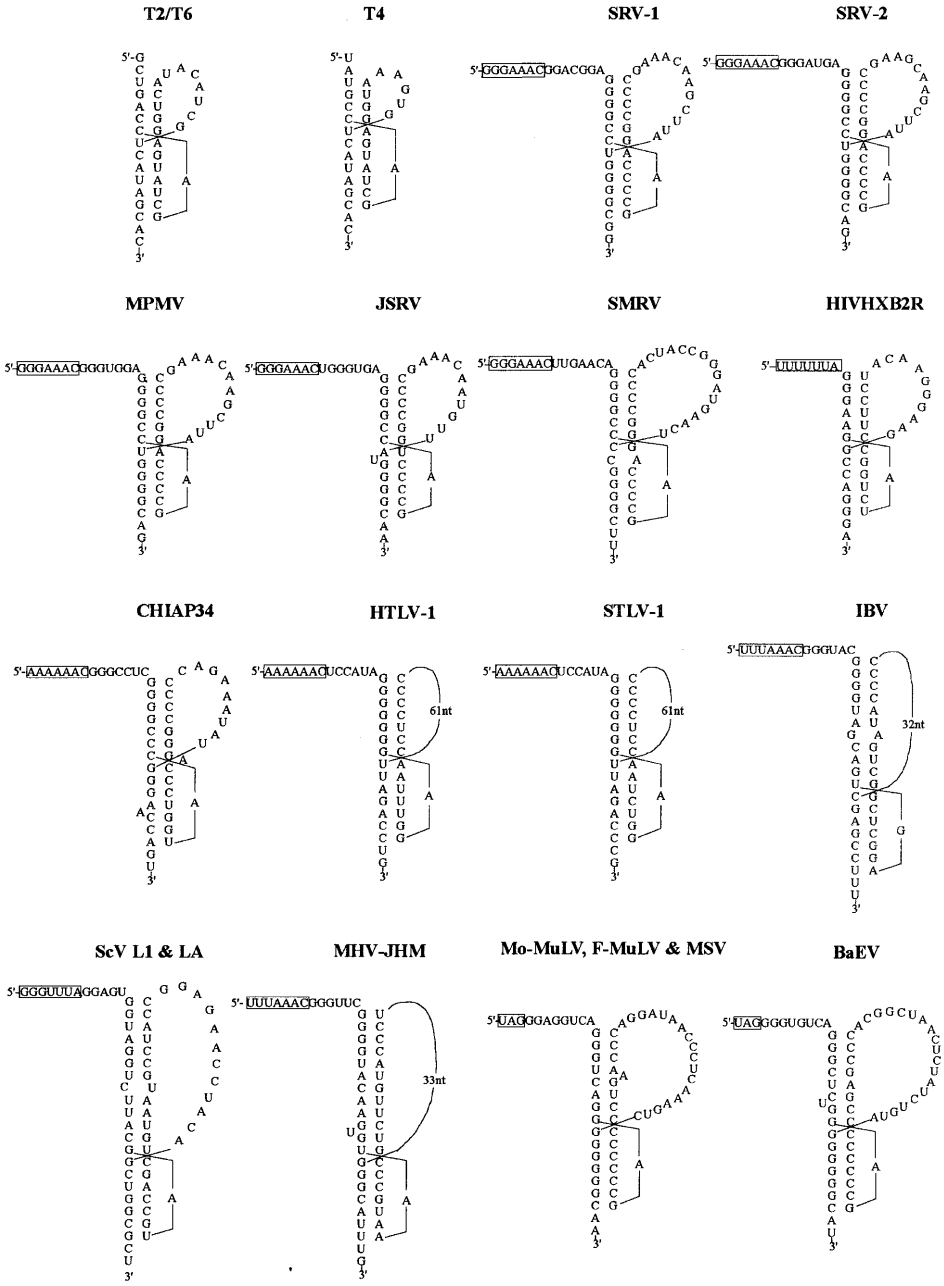

Figure 5. Diagrams of base-pairings for proposed members of the CPK1 family of structurally related RNA pseudoknots. Except for the pseudoknots of gene*32* mRNA of bacteriophages T2, T6 (same sequence) and T4, all of the other pseudoknots are located downstream of the frameshift or readthrough sites (boxed) of viral RNAs. Shown are frameshift-associated pseudoknots of simian retrovirus type-1 (SRV-1) and type-2 (SRV-2), Mason–Pfizer monkey virus (MPMV), Jaagsiekte sheep retrovirus (JSRV), squirrel monkey retrovirus (SMRV), HIV-1 strain HXB2R (HIVHXB2R), Chinese hamster intracisternal type A particle (CHIAP34), human T-cell leukemia virus type-1 (HTLV-1), simian T-cell leukemia virus type-1 (STLV-1), avian coronavirus infectious bronchitis virus (IBV), yeast*Saccharomyces cerevisiae* double-stranded RNA virus L1 and LA (ScV L1 and LA), mouse hepatitis virus strain JHM (MHV-JHM); readthrough-associated pseudoknots of Moloney murine leukemia virus (Mo-MuLV), Friend murine leukemia virus (F-MuLV) and murine sarcoma virus (MSV, same sequence), and baboon endogenous virus (BaEV). The pseudoknots of SRV, MPMV, JSRV, IBV, ScV, MuLV and BaEV were predicted by ten Dam and co-workers (12). Those of HIV-1, HTLV-1, STLV-1 and MHV-JHM are from Du *et al.* (9). The other viral RNA sequences are from GenBank. The possible pseudoknot illustrated for HIV-1 is perhaps more speculative than the other frameshift-associated pseudoknots, since it lacks a spacing sequence between the frameshift site and the pseudoknot.

1 of the other 3 nt (G, C or U) had little effect on the frameshifting efficiency (13). In addition, the naturally occurring frameshift-associated pseudoknot of the coronavirus avian infectious bronchitis virus (IBV) (14–15) is predicted to contain a single guanosine as the loop1 nucleotide (12 and Fig. 5 herein). In light of these previous studies and the present structural investigations, it seems that there is little structural or functional basis for the conservation of adenosine as the loop1 nucleotide in the CPK1 pseudoknots. Two possible explanations that may account for the previously observed 'conservation' of the loop1 adenosine are: (i) lack of sampling; or (ii) the stronger tendency for an adenosine residue to appear in the single-stranded regions of folded RNA

molecules. Of course, it can not be ruled out that an adenosine residue might be favored due to thermodynamic or folding considerations. A systematic investigation of the influence of loop1 nucleotide identity on the pseudoknot stability and unfolding pathway is in progress. We have previously noted, however, that the unfolding of the wild-type bacteriophage gene *32* mRNA pseudoknot is quite complex (9,11).

The combined structural and phylogenetic evidence suggests that the CPK1 motif represents a naturally preferred common structural theme for the tertiary folding of RNA pseudoknots (9 and Fig. 5 herein). It is therefore relevant to further explore the rationale underlying the apparent popularity of this specific motif. The preference of 6 or 7 bp in stem2 most likely has its basis in the fact that the distance across the major groove of the helical stem reaches a minimum when 6 or 7 bp are bridged (Fig. 1). However, this minimum distance only explains why it is *possible* for a single nucleotide to span the major groove of a 6 or 7 bp stem, it does not automatically explain why a 1 nt connection appears to be *preferred*. Perhaps it is the limited space within the major groove of stem2 that restricts the number of nucleotides that are allowed in loop1; the presence of additional loop1 nucleotides may force the relatively hydrophobic bases to be exposed to the solvent. It is also noted that when stem2 contains 6 or 7 bp, corresponding to slightly more than one-half helical turn, the loop1 nucleotide is in a optimum position to embed its base within the deep major groove with the least possibility of interfering with the elements (bases, riboses and phosphates) of stem2 (Fig. 1).

The present query into the underlying rationale of the CPK1 family of pseudoknots reveals that several specific structural features of the A-form RNA helix may work in concert to give rise to the specific folding of this motif. With its distinctive structural features and apparent biological relevance, the CPK1 motif is a significant entry into the rapidly growing structural database of RNA molecules and will provide valuable structural information for related biophysical and biochemical studies in the future.

ACKNOWLEDGEMENTS

We thank Dr David P. Giedroc for many helpful discussions. This work was supported by NIH grant R01-AI40187 (to D.W.H. and D.P.G.), and is in partial fulfillment of the requirements for the Ph.D. degree at the University of Texas at Austin (to Z.D.).

REFERENCES

1 Pleij,C.W.A. and Bosch,L. (1989) *Methods Enzymol.*, **180**, 289–303.
2 Schimmel,P. (1989) *Cell*, **58**, 9–12.
3 Pleij,C.W.A. (1990) *Trends Biochem. Sci.*, **15**, 143–147.
4 Pleij,C.W.A. (1994) *Curr. Opin. Struct. Biol.*, **4,** 337–344.
5 ten Dam,E.B., Pleij,C.W.A. and Draper,D. (1992) *Biochemistry,* **31**, 11665–11676.
6 Rietveld,K., Poelgeest,R.V., Pleij,C.W.A., Van Boom,J.H. and Bosch, L. (1982) *Nucleic Acids Res.*, **10**, 1929–1946.
7 Pleij,C.W.A., Rietveld,K. and Bosch,L. (1985) *Nucleic Acids Res.*, **13**, 1717–1731.
8 Puglisi,J.D., Wyatt,J.R. and Tinoco,I.,Jr (1990) *J. Mol. Biol.*, **214**, 437–453.
9 Du,Z., Giedroc,D.P. and Hoffman,D.W. (1996) *Biochemistry,* **35**, 4187–4198.
10 Wyatt,J.R., Puglisi,J.D. and Tinoco,I.,Jr (1990) *J. Mol. Biol.*, **214**, 455–470.
11 Qiu,H., Kaluarachchi,K., Du,Z., Hoffman,D.W. and Giedroc,D.P. (1996) *Biochemistry,* **35,** 4176–4186.
12 ten Dam,E.B., Pleij,C.W.A. and Bosch,L. (1990) *Virus Genes*, **4**, 121–136.
13 ten Dam,E.B., Verlaan,P.W.G. and Pleij,C.W.A. (1995) *RNA*, **1**, 146–154.
14 Brierley,I., Digard,P. and Inglis,S.C. (1989) *Cell*, **57**, 537–547.
15 Brierley,I., Rolley,N.J., Jenner,A.J. and Inglis,S.C. (1992) *J. Mol. Biol.*, **220**, 889–902.
16 Plateau,P. and Guéron,M. (1982) *J. Am. Chem. Soc.*, **104**, 7310–7311.
17 Ferrin,T.E., Huang,C.C., Jarvis,L.E. and Langridge,R. (1988) *J. Mol. Graphics*, **6**, 13–27.
18 Huang,C.C., Pettersen,E.F., Klein,T.E., Ferrin,T.E. and Langridge,R. (1991) *J. Mol. Graphics*, **9**, 230–236.