

# A long-range pseudoknot is required for activity of the *Neurospora* VS ribozyme

Toolika Rastogi, Tara L. Beattie, Joan E. Olive and Richard A. Collins<sup>1</sup>

Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8

<sup>1</sup>Corresponding author

**Four small RNA self-cleaving domains, the hammerhead, hairpin, hepatitis delta virus and *Neurospora* VS ribozymes, have been identified previously in naturally occurring RNAs. The secondary structures of these ribozymes are reasonably well understood, but little is known about long-range interactions that form the catalytically active tertiary conformations. Our previous work, which identified several secondary structure elements of the VS ribozyme, also showed that many additional bases were protected by magnesium-dependent interactions, implying that several tertiary contacts remained to be identified. Here we have used site-directed mutagenesis and chemical modification to characterize the first long-range interaction identified in VS RNA. This interaction contains a 3 bp pseudoknot helix that is required for tertiary folding and self-cleavage activity of the VS ribozyme.**

**Keywords:** pseudoknot/ribozyme/RNA structure/tertiary structure

## Introduction

Ribozymes provide good experimental systems with which to study RNA structure–function relationships and RNA folding, because functionally important changes in structure can often be detected by changes in catalytic properties. Several natural RNAs contain short ribozyme sequences that are capable of self-cleavage. Many of these RNAs can form a secondary structure called the hammerhead (for review, see Symons, 1992); however, different sequences and structures are present in hairpin (Burke, 1994), hepatitis delta virus (HDV; Perotta and Been, 1993) and *Neurospora* VS ribozymes (Beattie *et al.*, 1995). Little is known about long-range tertiary interactions that form the catalytically active RNA conformations, although X-ray crystallography of inactive versions of hammerhead ribozymes and chemical cross-linking recently have provided information about tertiary structures in this ribozyme (Pley *et al.*, 1994; Scott *et al.*, 1995; Sigurdsson *et al.*, 1995). Certain tertiary structure models of HDV ribozymes (Perotta and Been, 1993; Tanner *et al.*, 1994) include a long-range pseudoknot that is required for activity in the presence of denaturants. Some tertiary contacts in larger catalytic RNAs have also been identified (reviewed by Cech, 1993); these include group I introns (Michel and Westhof, 1990), group II

introns (Michel and Ferat, 1995) and the RNA component of RNase P (Harris *et al.*, 1994).

We are studying the VS ribozyme, which is found within an abundant non-coding RNA, called VS RNA. VS RNA is present in the mitochondria of certain natural isolates of *Neurospora* (Collins and Saville, 1990) where it is maintained as a satellite RNA that is transcribed from a DNA plasmid by the mitochondrial RNA polymerase and is replicated by a reverse transcriptase encoded by another plasmid (Kennell *et al.*, 1995). VS RNA is capable of self-cleavage and ligation (Saville and Collins, 1990, 1991); the minimal contiguous self-cleaving sequence contains a single nucleotide upstream of the cleavage site and 153 nucleotides downstream (Guo *et al.*, 1993). The products of VS RNA self-cleavage contain 2',3' cyclic phosphate and 5' hydroxyl termini like those of hammerhead, hairpin and HDV ribozymes; however, the secondary structure of VS RNA is distinct from these other self-cleaving domains (Figure 1a and b; Beattie *et al.*, 1995).

Our previous work has provided structural and functional data supporting the existence and importance of stem-loops II–VI in the current secondary structure model of VS RNA (Figure 1b). This portion of VS RNA can also act as a multiple turnover ribozyme catalyzing the cleavage *in trans* of a separate substrate RNA composed of stem-loop I (Guo and Collins, 1995). The low  $K_m$  of this reaction ( $\sim 0.1 \mu\text{M}$ ) indicates a rather strong interaction between stem-loop I and the rest of the ribozyme, comparable with that of some hammerhead and hairpin ribozymes that bind their substrate via multiple Watson–Crick base pairs (Symons, 1992; Joseph *et al.*, 1993); however, the lack of long regions of complementarity between the ribozyme and substrate indicates that extensive intermolecular base pairing is unlikely in the case of the VS ribozyme.

In contrast to the evidence supporting the importance of stem-loops II–VI, the data relating to the structure and function of the sequences that form stem-loop I have been enigmatic. Nuclease probing showed that the short region of VS RNA that serves as a substrate for *trans* cleavage does indeed form the predicted stem-loop structure (Guo and Collins, 1995). However, in larger VS transcripts produced by self-cleavage, chemical modification studies showed that stem-loop I forms only in the absence of magnesium, and that a different structure forms in the presence of magnesium. Mutational data indicate that most of the base pairs in stem I are not essential for self-cleavage, although the identity of a few specific bases is important (Beattie *et al.*, 1995). Taken together, these observations suggest that the region of VS RNA that forms stem-loop I is involved in magnesium-dependent conformational changes.

Here we have used mutational analysis combined with a self-cleavage assay and chemical modification to identify



**Table I.** Relative self-cleavage rates of mutants in the top two potential base pairs of stem I

Clone:	wt	59	30	7	52	12	37	q4	18	49	27
Sequence <sup>a</sup> :	627G-C632	c C	c C	G-C	G-C	c C	c-g	c-g	G g	c-g	G g
	626C-G633	C-G	g-c	C c	g G	g G	C-G	g-c	C-G	g G	C c
$k_{rel}$ <sup>b</sup> :	100	64	156	74	121	145	<0.1	<0.1	<0.1	<0.1	<0.1

<sup>a</sup>Wild-type and mutant bases are shown in upper and lower case, respectively. Where Watson-Crick base pairs are possible, they are indicated by a dash. Bases are numbered as in Figure 1.

<sup>b</sup> $k_{rel}$  is the relative first order rate constant for self-cleavage (cleavage rate of mutant/cleavage rate of wild-type)  $\times 100$ . See Materials and methods for details.

We then attempted to suppress each single base substitution by constructing RNAs with each of the Watson-Crick pairs at each position in the pseudoknot helix (Table II, column 3). Restoration of a substantial amount of activity was possible at each position in the helix by at least one non-wild-type Watson-Crick pair (summarized in Figure 2). This suppression by compensatory base substitutions involving exactly the three adjacent complementary positions in the two distant regions of the VS sequence predicted by the model provides strong genetic evidence that the pseudoknot helix forms and plays an important role in activity, perhaps by facilitating proper folding of the RNA.

#### **Additional constraints on the pseudoknot sequence**

Only certain Watson-Crick combinations in the pseudoknot helix supported activity: this was especially evident at base pairs 630/699 and 631/698 where only the double transition substitutions, but not the double transversions, restored activity (Table II, column 3). Such apparent sequence constraints could simply mean that certain base substitutions cause severe misfolding of the RNA to the extent that the active conformation cannot be restored by the compensatory base substitution. However, functional constraints on the identities of bases in helices in other RNAs have also been observed (discussed by Cech, 1988). Considering the short length of the VS RNA pseudoknot helix, the observed sequence constraints may mean that only certain base pair combinations allow adequate helix stability or proper local helix geometry. It is also possible that bases at certain positions form other interactions in alternative RNA conformations, or additional interactions that contribute to stability of the pseudoknot.

This last possibility makes a testable prediction about one potential additional interaction. The uridine turn, which was first observed in the crystal structure of tRNA pseudouridine and anticodon loops (Quigley and Rich, 1976), and recently in the hammerhead ribozyme (Pley *et al.*, 1994; Scott *et al.*, 1995), is comprised of the sequence motif, UNR (U = uridine, N = any nucleotide, R = purine). This sequence can form a sharp reversal in the direction of the polynucleotide chain stabilized by two hydrogen bonds: one between the N3 proton of the uridine and a phosphate oxygen 3' of the purine; the other between the ribose 2' hydroxyl of the uridine and the N7 of the purine. Our mutational data (Table II) could be interpreted as a requirement for purines at positions 630 and 698, which in both cases are located two nucleotides 3' to uridines in the wild-type sequences of loops I and V. To investigate whether U-turns might be important for activity of VS RNA, we measured the cleavage rates of mutant

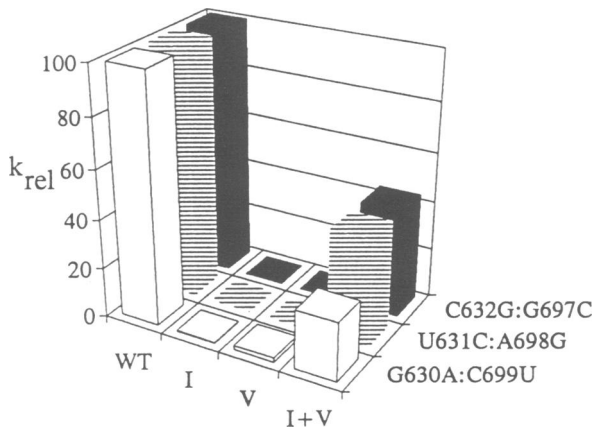
**Table II.** Effects of base substitutions in loops I and V on the rate of self-cleavage

Loop I		Loop V		Loop I+V	
Seq. <sup>a</sup>	$k_{rel}$ <sup>b</sup>	Seq.	$k_{rel}$	Seq.	$k_{rel}$
628	U 100				
	C 3.5				
	G 0.8				
	A 1.4				
629	C 100				
	U 1.5				
	A 6.4				
	G 9.7				
630	G 100	699	C 100	630G:699C	100
	A 0.3		U 1.8	A U	23.8
	<0.1		A 0.6	U A	0.1
	<0.1		G <0.1	C G	0.6
631	U 100	698	A 100	631U:698A	100
	C 0.7		G <0.1	C G	43.5
	G <0.1		C 4.1	G C	0.7
	A <0.1		U <0.1	A U	2.4
632	C 100	697	G 100	632C:697G	100
	U 25.8		A <0.1	U A	23.0
	A <0.1		U 0.1	A U	3.0
	G <0.1		C <0.1	G C	42.6
		696	U 100		
			C 18.3		
			G 2.4		
			A 3.0		

<sup>a</sup>Identifies the base present in single mutants at each position in loop I (column 1), loop V (column 2) or in the double mutants at each base pair in the pseudoknot helix made by combining the mutations listed in loop I and loop V (column 3). In each group, the wild-type sequence at each position is listed first. Bases are numbered as in Figure 1b.

<sup>b</sup> $k_{rel}$  is the relative first order rate constant for self-cleavage (cleavage rate of mutant/cleavage rate of wild-type)  $\times 100$ . The data are averages of at least two independent experiments.

RNAs in which U628 or U696 had been changed to each of the other bases. Consistent with the prediction from the uridine turn model, the data in Table II show that substitution of each uridine with any other base decreased activity, severely so at U628. The base at the N position of a UNR motif is not constrained in identity by its involvement in the uridine turn, although a particular base may be required for other reasons (for example, in the U-turn in the hammerhead ribozyme the N must be guanosine; Ruffner *et al.*, 1990). Similarly, in the VS pseudoknot, G697, which would correspond to the N in the UNR motif of loop V, is constrained in identity by its requirement for forming the bottom base pair of the pseudoknot helix. C629, the N of the UNR sequence in loop I, also appears to be involved in additional interactions, as base substitutions at this position also decreased activity. Model building (not shown) suggests that U-turns



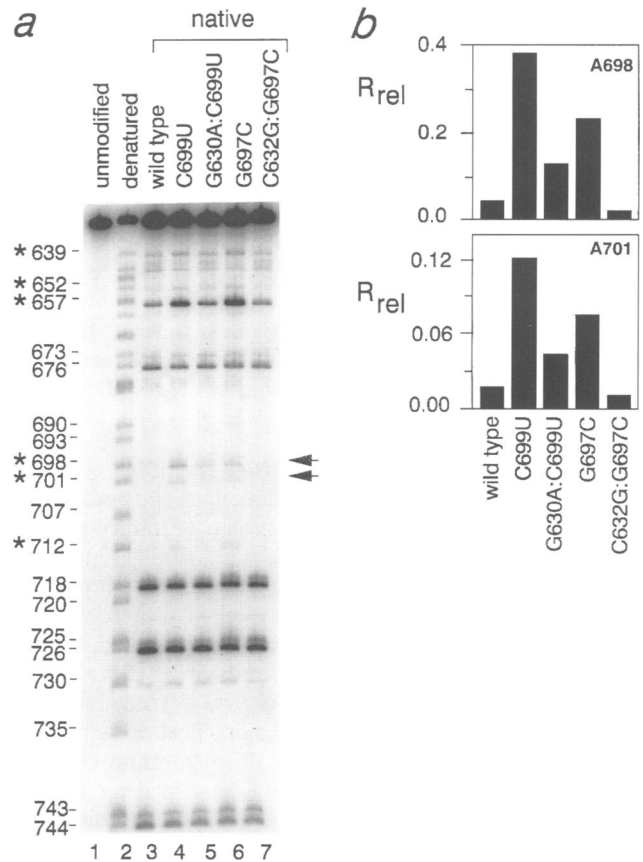
**Fig. 2.** Restoration of Watson–Crick base pairing at each position in the pseudoknot helix restores self-cleavage activity.  $k_{rel}$ : self-cleavage rate relative to wild-type.  $x$ -axis (left-to-right): WT, wild-type; I, substitution in loop I; V, substitution in loop V; I+V, compensatory substitution. The individual base substitutions are listed on the  $z$ -axis; see Materials and methods for details and Table II for additional data.

could co-exist with the pseudoknot helix; however, the mutational data would also be consistent with sequential formation of these interactions in different structures along a folding pathway. Although the available data are insufficient to prove the existence of U-turns, the combined mutational data indicating the importance of U628 and U696 for activity and the requirement for a purine two nucleotides downstream from each uridine are consistent with the mutational signature that would be predicted for uridine turns that are important for activity.

#### Chemical modification supports the existence of the pseudoknot

To obtain further evidence for the formation of the pseudoknot in the wild-type and active double mutant RNAs, we performed structure probing with diethyl pyrocarbonate (DEPC), which modifies the N7 position of adenosines unless the base is stacked in a helix or the N7 is involved in a direct interaction (Ehresmann *et al.*, 1987; Weeks and Crothers, 1993). The chemical modification pattern in native conditions of wild-type D RNA, the downstream cleavage product, has been studied in detail previously (Beattie *et al.* 1995) and differs slightly from that of pre-RNA, the uncleaved precursor (T.L.Beattie and R.A.Collins, unpublished data). To avoid potential interpretation difficulties associated with the mixture of pre- and D RNAs that forms in native conditions during the time required for chemical modification, we examined purified D RNAs from each of the mutant sequences.

In wild-type, we had noticed previously that A698 in loop V was reactive with DEPC in the absence of magnesium but protected in the presence of magnesium (native conditions; Beattie *et al.*, 1995). This protection is consistent with A698 being one member of the middle base pair of the proposed pseudoknot helix that forms only in native conditions. A701, which is at the end of stem V, showed the same pattern of accessibility as A698, as would be expected if it were protected by stacking of the pseudoknot helix on stem V or by a direct interaction of its N7 with an undetermined partner that occurs in native conditions. These two adenosines are therefore



**Fig. 3.** Diethyl pyrocarbonate (DEPC) modification of active and inactive RNAs. (a) Lanes: 1, unmodified wild-type G11 RNA; 2, adenosine ladder from modification of G11 RNA in denaturing conditions (200 mM HEPES pH 8.0, 1 mM EDTA, 90°C); 3–7, modification in native conditions (200 mM HEPES pH 8.0, 50 mM KCl, 10 mM MgCl<sub>2</sub>, 37°C) of the RNAs identified in the figure. Adenosines whose protections parallel the self-cleavage activity of the RNAs are indicated by asterisks; A698 and A701 in loop V are indicated by arrows. (b) Relative reactivities ( $R_{rel}$ ) of bases A698 and A701. Bands in lanes 3–7 were quantitated with a PhosphorImager. To correct for slight differences in loading among lanes, intensities of bands A698 or A701 in each lane were normalized to that of A676 whose reactivity does not appear to change in a variety of conditions (Beattie *et al.*, 1995 and unpublished data).

suitable ‘reporter’ bases for the magnesium-dependent changes that accompany formation of the active RNA structure.

Figure 3a shows that A698 and A701 are accessible to DEPC modification in two inactive single mutants that should disrupt the pseudoknot (G697C or C699U) and protected or much less accessible in the corresponding double mutants that contain compensatory substitutions in loop I that restore activity (C632G:G697C and G630A:C699U). Quantitation of the reactivity of these adenosines relative to bases whose reactivity remains constant, e.g. A676 (or others not shown), confirmed the qualitative impression from the autoradiogram (Figure 3b).

The reactivities of some adenosines in distant parts of the secondary structure (e.g. A639, A652, A657 and A712 indicated by asterisks in Figure 3a), also changed in a correlated fashion, being more reactive in the inactive single mutants and more protected in the active double mutants. This same set of adenosines, all of which are located in bulges or loops in the current secondary structure model, had been found previously to be reactive in the

absence of magnesium but protected in the presence of magnesium (Beattie *et al.*, 1995), suggesting that their protection is due to tertiary folding. Adenosines in standard secondary structure helices remained protected in the single mutants that disrupt the pseudoknot helix, indicating that most or all of the secondary structure remains intact but tertiary folding is destabilized in these mutants. This pattern of increased chemical reactivity of specific bases in inactive RNAs and decreased reactivity in active RNAs suggests that formation of the pseudoknot is important for overall tertiary folding, and provides support for the conclusion from mutation-suppression analyses that formation of the pseudoknot is important for formation of the active structure.

### Implications for VS RNA structure models

Our previous chemical modification data showed that a substantial number of bases are protected from modification in the presence, compared with the absence, of magnesium. Several of the magnesium-dependent protections in loops I and V can now be explained by formation of the pseudoknot described here. The pseudoknot also provides an explanation for our mutational data, which demonstrated that although stem-loops II–VI are important for *cis* cleavage, stem-loop I is not. That is, although stem-loop I forms in the absence of magnesium (Beattie *et al.*, 1995), it is not the stem I base pairing that is required for activity, but rather the pseudoknot pairing.

It is not clear from our available data how much, if any, of VS RNA stem I forms in the presence of magnesium. If none of the stem I base pairs are formed in the active structure, the VS RNA pseudoknot may resemble the typical 'H'-type pseudoknot formed by pairing of loop bases in a hairpin with bases in a single-stranded region elsewhere (reviewed by ten Dam *et al.*, 1992; Pleij, 1994). If stem-loop I is an intermediate in the folding pathway of VS RNA, perhaps only part of stem I must be unwound to allow C632 to pair with G697 in the pseudoknot helix. In this case, VS RNA may contain a structure resembling a loop-loop pseudoknot, such as those involved in regulation of ColE1 plasmid replication in *Escherichia coli* (Eguchi and Tomizawa, 1991) and initiation of dimerization of HIV-1 RNA (Paillart *et al.*, 1994). The details of such structures have been modeled using nuclear magnetic resonance experiments (Chang and Tinoco, 1994; Marino *et al.*, 1995). More complex pseudoknots have also been found. One that may be particularly relevant to our observations has been found in the *E.coli*  $\alpha$  mRNA. Formation of this structure also appears to involve alternative, magnesium-dependent, intermediate secondary structures (Gluck and Draper, 1994).

In conclusion, our mutational and chemical modification data have identified a long-range interaction that is required for proper tertiary folding and self-cleavage activity of VS RNA. The pseudoknot is clearly one important feature of the magnesium-dependent active structure; however, the identities of additional tertiary interactions implied from chemical modification remain to be determined.

## Materials and methods

### Synthesis of RNAs and construction of mutants

Radioactive precursor RNAs (pre-RNAs) were synthesized by transcription with T7 RNA polymerase from plasmid G11 (or its mutant

derivatives) that had been linearized at the *Ssp*I site corresponding to VS nucleotide 783 as described previously (Guo *et al.*, 1993; Collins and Olive, 1993). Pre-RNA contains the minimal contiguous self-cleaving region (nt 620–773) flanked by a few non-essential nucleotides. Self-cleavage was performed as described previously (Collins and Olive, 1993) or in slightly different conditions (30 mM Tris-HCl pH 8.0, 1.5 mM spermidine-HCl, 37.5 mM KCl and 15 mM MgCl<sub>2</sub> at 37°C). No significant difference was observed in the different reaction conditions. The first-order rate constants of self-cleavage were calculated as  $-2.3(m)$ , where  $m$  is the slope of a plot of log fraction uncleaved versus time, and are presented as a percentage of the rate of wild-type G11 RNA. The cleavage rate of wild-type was  $\sim 0.15$  per minute; the range of day to day variation was approximately  $\pm 25\%$ . The limit of detection in our assay is  $\sim 0.1\%$  of the wild-type rate.

Base substitutions were introduced into the G11 sequence by oligonucleotide mutagenesis of plasmid DNAs (Kunkel *et al.*, 1987). Single substitutions were combined using standard recombinant DNA techniques to construct some of the compensatory double mutants. All cloned DNAs were sequenced across the region corresponding to pre-RNA to confirm that only the expected mutations were present.

### DEPC modification

Downstream self-cleavage product RNAs (called D in Figure 1a) were gel-purified after self-cleavage reactions, 3' end-labeled, modified with DEPC, cleaved with aniline and electrophoresed using standard procedures (Bruce and Uhlenbeck, 1978; Peattie, 1979; Peattie and Gilbert, 1980; Ehresmann *et al.*, 1987; Krol and Carbon, 1989) with minor modifications as described previously (Beattie *et al.*, 1995). Because the G697C and C699U mutant RNAs are incapable of self-cleavage, they were cleaved *in trans* by a ribozyme derived from VS RNA (Guo and Collins, 1995) to obtain the equivalent D RNAs, which were gel-purified and treated as above. Band intensities were quantitated with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). To correct for slight differences in loading among lanes, intensities of bands A698 or A701 in each lane were normalized to that of A676 whose reactivity does not appear to change in a variety of solution conditions (Beattie *et al.*, 1995 and unpublished data). A trend similar to that shown in Figure 3b was also observed when normalized to A718, A744 or to the amount of unmodified full-length RNA in each lane (data not shown).

## Acknowledgements

This work was supported by a grant from the Medical Research Council of Canada. R.A.C. is a Fellow of the Canadian Institute for Advanced Research Program in Evolutionary Biology.

## References

- Beattie, T.L., Olive, J.E. and Collins, R.A. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 4686–4690.
- Bruce, A.G. and Uhlenbeck, O.C. (1978) *Nucleic Acids Res.*, **5**, 3665–3677.
- Burke, J. (1994) *Nucleic Acids Mol. Biol.*, **8**, 106–118.
- Cech, T.R. (1988) *Gene*, **73**, 259–271.
- Cech, T.R. (1993) In Gesteland, R.F. and Atkins, J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 239–270.
- Chang, K.-Y. and Tinoco, I. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 8705–8709.
- Collins, R.A. and Olive, J.E. (1993) *Biochemistry*, **32**, 2795–2799.
- Collins, R.A. and Saville, B.J. (1990) *Nature*, **345**, 177–179.
- Eguchi, Y. and Tomizawa, J. (1991) *J. Mol. Biol.*, **220**, 831–842.
- Ehresmann, C., Baudin, F., Mougil, M., Romby, P., Ebel, J.-P. and Ehresmann, B. (1987) *Nucleic Acids Res.*, **15**, 9109–9128.
- Gluck, T.C. and Draper, D.E. (1994) *J. Mol. Biol.*, **241**, 246–262.
- Guo, H.C.-T. and Collins, R.A. (1995) *EMBO J.*, **14**, 368–376.
- Guo, H.C.-T., De Abreu, D.M., Tillier, E.R.M., Saville, B.J., Olive, J.E. and Collins, R.A. (1993) *J. Mol. Biol.*, **232**, 351–361.
- Harris, M.E., Nolan, J.M., Malhotra, A., Brown, J.W., Harvey, S.C. and Pace, N.R. (1994) *EMBO J.*, **13**, 3953–3963.
- Joseph, S., Berzal-Herranz, A., Chowrira, B.M., Butcher, S.E. and Burke, J.M. (1993) *Genes Dev.*, **7**, 130–138.
- Kennell, J.C., Saville, B.J., Mohr, S., Kuiper, M.T.R., Sabourin, J.R., Collins, R.A. and Lambowitz, A.M. (1995) *Genes Dev.*, **9**, 294–303.
- Krol, A. and Carbon, P. (1989) *Methods Enzymol.*, **180**, 212–227.

- Kunkel,T.T., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
- Marino,J.P., Gregorian,R.S., Jr, Csankovszki,G. and Crothers,D.M. (1995) *Science*, **268**, 1448–1454.
- Michel,F. and Ferat,J.-L. (1995) *Annu. Rev. Biochem.* **64**, 435–461.
- Michel,F. and Westhof,E. (1990) *J. Mol. Biol.*, **216**, 585–610.
- Paillart,J.-C., Marquet,R., Skripkin,E., Ehresmann,B. and Ehresmann,C. (1994) *J. Biol. Chem.*, **269**, 27486–27493.
- Peattie,D.A. (1979) *Proc. Natl Acad. Sci. USA*, **76**, 1760–1764.
- Peattie,D.A. and Gilbert,W. (1980) *Proc. Natl Acad. Sci. USA*, **77**, 4679–4682.
- Perrotta,A.T. and Been,M.D. (1993) *Nucleic Acids Res.*, **21**, 3959–3965.
- Pleij,C.W.A. (1994) *Curr. Opin. Struct. Biol.*, **4**, 337–344.
- Pley,H.W., Flaherty,K.M. and McKay,D.B. (1994) *Nature*, **372**, 68–74.
- Quigley,G.J. and Rich,A. (1976) *Science*, **194**, 796–806.
- Ruffner,D.E., Stormo,G. and Uhlenbeck,O.C. (1990) *Biochemistry*, **29**, 10695–10702.
- Saville,B.J. and Collins,R.A. (1990) *Cell*, **61**, 685–696.
- Saville,B.J. and Collins,R.A. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 8826–8830.
- Scott,W.G., Finch,J.T. and Klug,A. (1995) *Cell*, **81**, 991–1002.
- Sigurdsson,S., Tuschl,T. and Eckstein,F. (1995) *RNA*, **1**, 575–583.
- Symons,R.H. (1992) *Annu. Rev. Biochem.*, **61**, 641–671.
- Tanner,N.K., Schaff,S., Thill,G., Petit-Koskas,E., Crain-Denoyelle,A.-M. and Westhof,E. (1994) *Curr. Biol.*, **4**, 488–498.
- ten Dam,E., Pleij,K. and Draper,D. (1992) *Biochemistry*, **31**, 11665–11676.
- Weeks,K.M. and Crothers,D.M. (1993) *Science*, **261**, 1574–1577.

Received on August 30, 1995; revised on January 15, 1996