Origins of RNA Catalysis in the Hairpin Ribozyme

David M. J. Lilley*[a]

KEYWORDS:

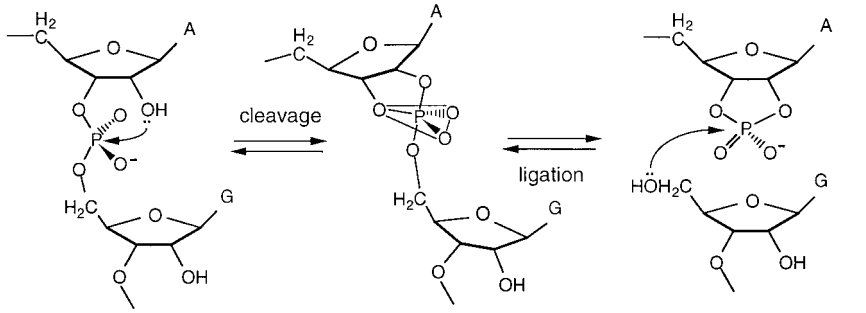
catalysis · ribozymes · RNA structure · structure–activity relationships · structure elucidation

The ribozyme community is living through immensely exciting times. It becomes ever more clear that RNA catalysis is fundamentally important in biological function; the recent demonstration that the peptidyl transferase activity of the ribosome is RNA catalysed[1] means that this ribozyme is ancient and fundamental to life. Data on thio effects in mammalian mRNA splicing[2] make it increasingly probable that this reaction is also RNA catalysed. Despite the undoubted importance of ribozymes however, we find ourselves in a rather similar position to that of the enzyme field around 30 years ago. We have a lot of pointers, but we still do not really understand the origins of chemical catalysis in RNA molecules. This provides an exciting challenge to the chemical biologist. At first sight RNA seems a rather poor prospect for macromolecular catalysis, given its polyelectrolyte character and its unpromising array of functional groups compared to the proteins. RNA is therefore a comparatively "stripped-down" catalyst of meagre resources. Yet the relative simplicity of the ribozymes provides an opportunity to dissect catalysis in different situations, and this could provide new insight into biocatalysis in general. A new crystal structure of the hairpin ribozyme[3] takes us significantly further down that road.

The hairpin ribozyme[4–6] is one of the group of small nucleolytic ribozymes that catalyse site-specific cleavage of the backbone of RNA by a transesterification reaction. Other members of this class include the hammerhead,[7, 8] hepatitis delta virus (HDV)[9] and Varkud satellite (VS)[10] ribozymes, although the reaction was first observed in the lead ion induced cleavage of tRNA.[11] The cleavage reaction proceeds by in-line nucleophilic attack of the 2′-hydroxy group on the adjacent phosphorus atom, which leads to the departure of the 5′-oxygen atom on the adjacent ribose and the formation of a cyclic 2′,3′-phosphate group (Scheme 1, left to right). Inversion of phosphorus chirality indicates that the reaction follows an S_N2 mechanism.[12] The reverse reaction can also be catalysed in some cases, whereby a 5′-oxygen atom attacks a cyclic 2′,3′-phosphate group leading to ligation (Scheme 1, right to left).[4, 13] In the case of the hairpin ribozyme this is especially efficient.[14]

We can identify a number of ways in which the reaction might be catalysed. Firstly, the normal geometry of A-form RNA is not aligned correctly for an S_N2 reaction, and thus altered local conformation could potentially facilitate the trajectory into the required transition state. This is unlikely to accelerate the reaction by more than 100-fold however, and thus could not by itself explain the typically 10^5-fold rate enhancement found for these ribozymes. Secondly, a stronger nucleophile could be obtained by removing the proton from the attacking hydroxy group, to give an alkoxide ion. A protein enzyme would achieve this by general base catalysis, probably by using a histidine sidechain, as in RNaseA for example.[15] Thirdly, the charged transition state could be stabilised by juxtaposition of a positive charge, and lastly, the departing oxyanion could be protonated by a general acid.

There are two clear candidates as participants in the reaction. In principle the nucleic acid bases (nucleobases) could play a direct role in the chemistry, by acting as general acids or bases. Adenine and cytosine can be protonated at the N1 and N3 positions respectively. However, the pK_a values of the free nucleotides (3.5 and 4.1, respectively) are too low to be effective at close to neutral pH values, and so these would have to be elevated in the environment of the folded RNA if they are to play a useful role. There is good evidence for the participation of a cytosine base in the HDV ribozyme, where catalytically defective mutants of the cytosine can be rescued by addition of exogenous imidazole.[16] The other potential players in the catalysis are metal ions, which are likely to have a number of roles. They are almost invariably involved in the



Scheme 1. *The cleavage and ligation reactions of the hairpin ribozyme. Like all the small nucleolytic ribozymes, the cleavage reaction appears to proceed through an S_N2 mechanism involving an in-line attack of the 2′-oxygen atom on the adjacent 3′-phosphorus atom, and departure of the 5′-oxygen atom on the next ribose. The ligation reaction, which is very efficient in the case of the hairpin ribozyme, should be the reverse of this reaction.*

[a] *Prof. Dr. D. M. J. Lilley*
CRC Nucleic Acid Structure Research Group
Department of Biochemistry, MSI/WTB Complex
The University of Dundee
Dundee DD1 5EH (UK)
Fax: (+44) 1382-201063
E-mail: dmjlilley@bad.dundee.ac.uk

folding of the RNA into its active conformation, and the hairpin ribozyme requires the cooperative binding of magnesium ions for folding.[17] Divalent ions may undergo site binding into the folded RNA structure, whereupon they could participate directly in the chemistry. Coordinated water molecules could act in general acid/base catalysis, and the positively charged metal ions could function in electrophilic catalysis, as in acids. There is good evidence for such a role for metal ions in the hammerhead ribozyme,[18] but in the case of the hairpin ribozyme the evidence seems to point in the opposite direction. In contrast to the hammerhead, no effect of phosphorothioate substitution was observed on cleavage rate, which suggests that direct coordination of the metal is not required.[19–21] Additionally, good cleavage rates were obtained in the substitutionally inert hexammine cobalt^{III} ions; this excludes both direct coordination and general acid/base catalysis by the ligand. Moreover, cleavage was observed in high concentrations of monovalent ions,[22] which are unlikely to exhibit tight site-specific binding to the RNA. This suggests that there is a general role for charge stabilisation, but that site binding is not essential.

The hairpin ribozyme is organised around a four-way helical junction (Figure 1). Two adjacent arms have formally unpaired loops (A and B), which include most of the essential nucleotides of the ribozyme[23–25] as well as the cleavage site. It was therefore presumed that interaction between these loops would somehow bring about catalysis of cleavage and ligation, and this was supported by experiments showing that cleavage would proceed when the loop-carrying arms were connected together in a variety of ways.[26–29] Indeed, some activity was even observed in trans, when the loop-carrying duplexes are unconnected.[26, 30] A physical demonstration of an interaction between the loops was finally obtained using fluorescence resonance energy transfer (FRET).[17, 31, 32] Like the majority of RNA species, the hairpin ribozyme undergoes folding induced by the binding of divalent metal ions.[17] In the absence of such ions the A and B loops do not interact, as the axes are held apart at approximately 90°. The ribozyme can be simplified down to the A and B helices, connected by a single-stranded hinge point;[33–35] indeed, much of what we know about this ribozyme comes from extensive studies on this form, particularly in the laboratory of Burke and co-workers. But it is clear that the four-way junction plays an important architectural role in the ribozyme;[31] ion-induced folding is vastly more efficient in this form compared to the simple hinged ribozyme.[36, 37] Interestingly, the ligase activity of the ribozyme is greatly favoured over cleavage in the junction form.[14]

Rupert and Ferré d'Amaré[3] have taken advantage of the greater stability of the natural junction form of the hairpin ribozyme to obtain crystals suitable for structure analysis. They also exploited a trick used in the earlier crystallographic study of the HDV ribozyme,[38] in which a binding loop for the U1A protein is placed at a site remote from the functional activity, and the RNA is cocrystallised with the protein. This aids both the crystallisation, and the solution of the structure by means of multiple wavelength anomalous diffraction (MAD) phasing with selenomethionine-substituted protein. With this approach they have obtained the structure of the ribozyme at a resolution of 2.4 Å.

The overall structure (Figure 2) is very much what was expected. The junction stacks in just the way deduced from FRET and comparative gel electrophoresis,[31] with the A arm stacked on D, and the B arm stacked on C. The junction is antiparallel, which will facilitate the anticipated intimate association between the A and B loops. The junction has one interesting difference from four-way DNA junctions (reviewed in ref. [39])—it is left handed. Whether the left-handed cross of axes is intrinsic to RNA junctions or a consequence of the strong association between the loops will be interesting to discover.

Model helices for the A[40] and B[41] loops had previously been studied by NMR spectroscopy. These helices presumably reflect the conformation of the loops prior to ion-induced interaction, but as one would expect, the interaction substantially alters both structures. There are three noncanonical base pairs in loop A and six in loop B. The geometry of loop B is significantly affected by several extrahelical uridine bases, and a distinctive turn is formed in the backbone of the U-rich strand at one point.

The two loops interact along their distorted minor grooves, to bury 1570 Å^2 of surface. They are stitched together by a network of hydrogen-bonding interactions, which include a ribose zipper

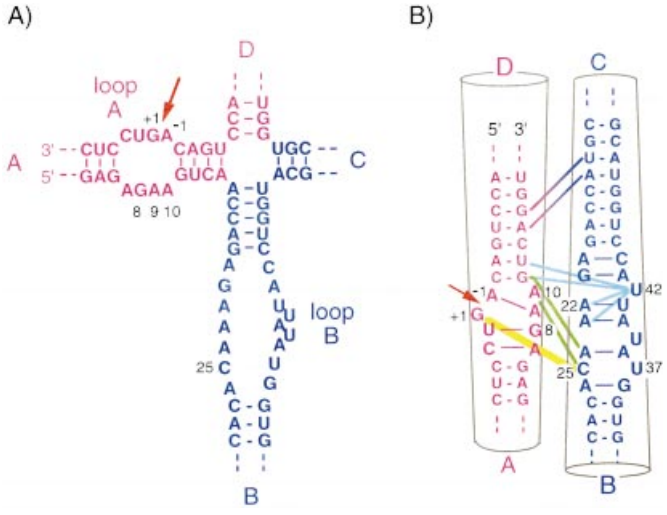


Figure 1. The sequence of the hairpin ribozyme. Site specific cleavage occurs at the position shown by the red arrow. A) The ribozyme is based on four arms (A – D) that are organised by a four-way helical junction. Arms A and B contain formally unpaired loops, which were expected to interact to generate the active conformation. The cleavage and ligation reactions occur in loop A. B) Representation of the sequence and secondary structure as observed in the crystal. The RNA folds by pairwise coaxial stacking of arm A on D (shown in magenta), and arm B on C (shown in blue). Some tertiary interactions are indicated, such as the G+1 ··· C25 base pair (yellow), the ribose zipper (green) and the interactions made by U42 (cyan).

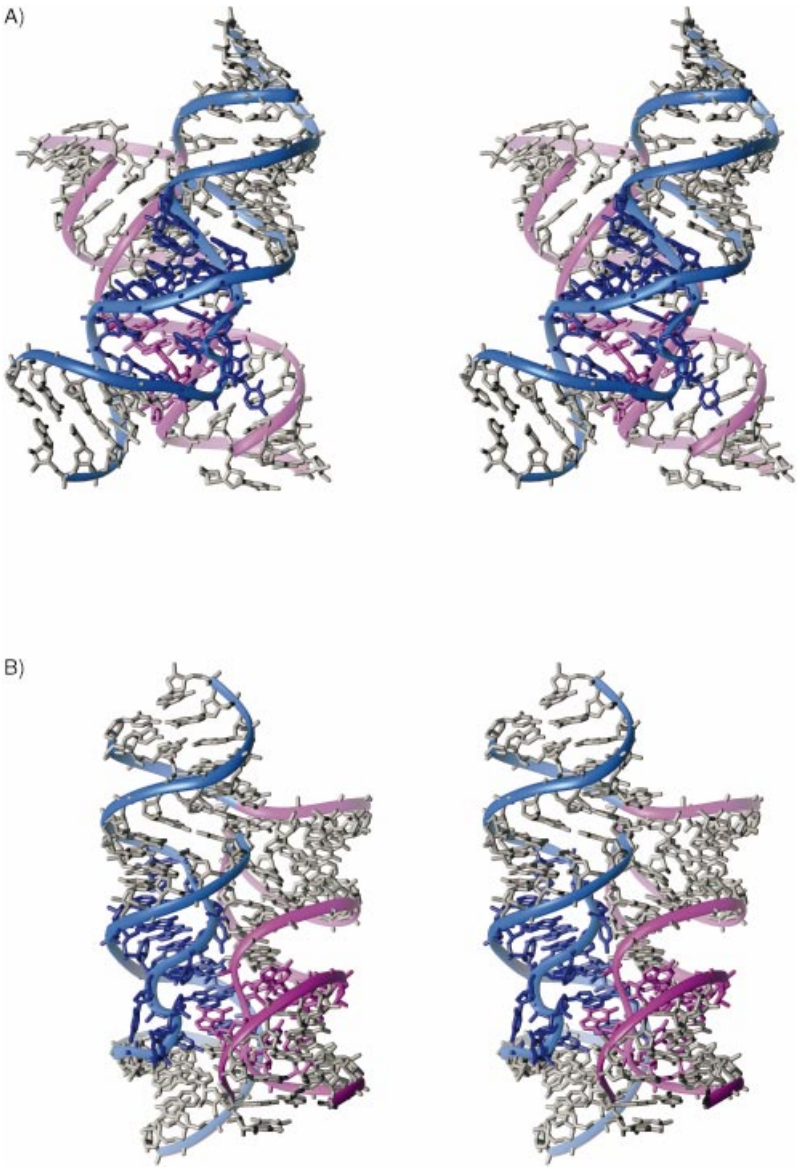


Figure 2. *Stereoviews of the overall structure of the hairpin ribozyme observed in the crystal. The loops and U1A protein have been removed for clarity. The path of the backbone is highlighted by the ribbons. Colour coding: A and D helices = magenta, B and C helices = blue. The nucleotides of the A and B loops are coloured pink and blue respectively, while the remaining nucleotides are grey. In these orientations the four-way helical junction is near the top. A) Face view, with the B and C helices towards the viewer. The distorted trajectory of the backbone of the B loop is clearly seen in this view. B) Side view, with the B and C helices on the left. The close association between the A and B loops is apparent in this view. The insertion of G+1 into the B loop is visible adjacent to the distorted backbone of the B loop. These images were prepared using the coordinates of Rupert and Ferré-d'Amaré.[3]*

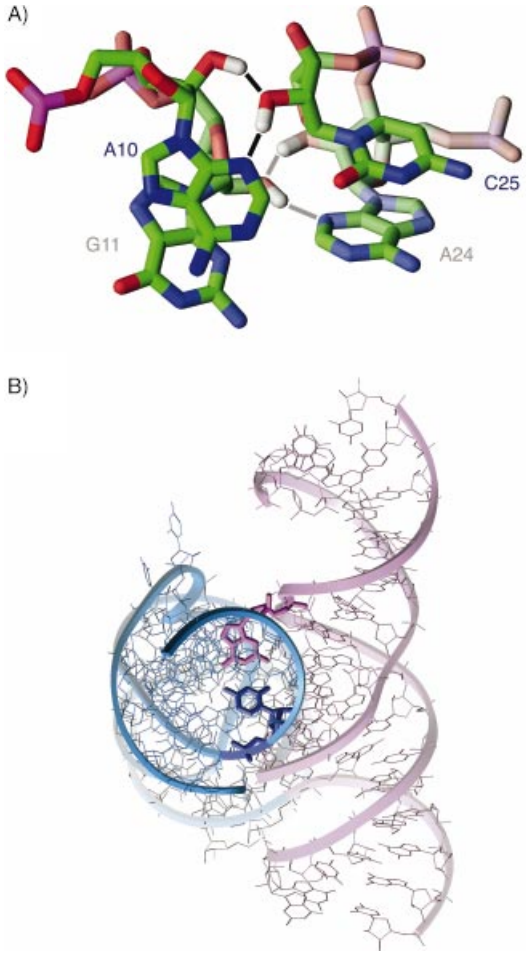


Figure 3. *Tertiary features mediating the interactions of the A and B loops of the hairpin ribozyme. A) The ribose zipper. The hydrogen bonds between the C25 2′-oxygen and the A10 N3 and 2′-oxygen atoms, and between the G11 2′-oxygen and the A24 N3 and 2′-oxygen atoms are indicated. B) The base pair formed between G+1 and C25. The path of the backbone is highlighted by the ribbons. Colour coding: A and D helices = magenta, B and C helices = blue. The view is approximately down the axis of the B helix, and shows the close interaction with the minor groove of the A loop. G+1 is extruded from the A loop and inserted into the B loop where it makes a Watson – Crick basepair with C25. G+1 (magenta) and C25 (blue) are highlighted by a thicker stick size.*

involving the N3 atom of A10 and A24 and the 2′-OH groups of A10, G11, A24 and C25 (Figure 3 A). The zipper was originally proposed by Earnshaw et al.,[29] and its nature was correctly deduced from functional group modification experiments on cleavage activity[42] and folding.[43] One of the extruded uridine bases of loop B (U42) makes extensive hydrogen-bonding interactions with nucleotides in both loops. But the most striking contact between the loops is formed by G+1, adjacent to the cleavage site. The entire nucleoside is extruded from helix A, and inserted into a specific pocket in loop B. The base is stacked between those of A26 and A38, and forms a Watson – Crick base pair with C25 (Figure 3 B). The importance of G+1 has been long established,[35] and the interaction with C25 was correctly predicted by complementation experiments on cleavage activity[44] and supported by corresponding experiments on folding.[43] However, a suggestion of the additional participation of A9 in a triple-base interaction[44] was evidently incorrect; the adenine base is wrongly

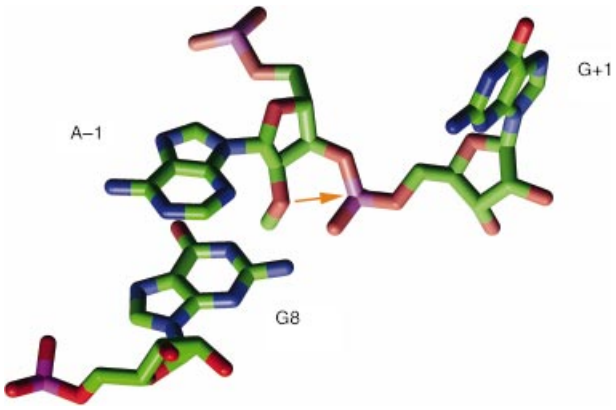


Figure 4. *The local geometry at the active site of the hairpin ribozyme. The scissile phosphate group is between nucleotides G + 1 and A − 1. The 2′-oxygen atom (shown here as a 2′-O-methyl derivative as crystallised) is approximately in line with the P–O5′ bond, indicated by the arrow. The base of G8 is immediately adjacent, and is hydrogen bonded to the attacking nucleophile and the pro-*Rp *oxygen atom.*

oriented in the crystal structure, and is 6 Å away from the G+1 ··· C25 pair.

The heart of the ribozyme is, of course, the active site, where the chemical reaction takes place. This is where we would hope to discover the mechanism of the rate enhancement. The local structure around the scissile bond is shown in Figure 4. No metal ions are resolved at this position, which may be consistent with the lack of thio-substitution effects on cleavage kinetics, as discussed above. However, some caution is required in the interpretation of this negative result; the putative nucleophilic 2′-OH group of A − 1 was replaced by a 2′-*O*-methyl group (to prevent cleavage occurring in the crystal), which might influence local ion binding. The situation in the hairpin structure contrasts with that of the leadzyme, where a barium ion was found bound to the 2′-OH group in the active centre.[45]

The local conformation around the cleavage site of the hairpin ribozyme shows that the 2′-oxygen, 3′-phosphorus and 5′-oxygen atoms are approximately colinear, poised to undergo the S_N2 reaction. The hammerhead ribozyme does not exhibit this alignment in its ground state,[46, 47] although a freeze-trapped, 5′-*C*-methylribose form of the ribozyme underwent substantial distortion in this direction.[48] The leadzyme also exhibited a nearly in-line conformation in the crystal.[45] Apparently the various ribozymes proceed along the path to an in-line transition state to differing extents in their ground states, and the hairpin ribozyme seems to have gone further than most. This is largely achieved by the distortion imposed by the extrusion of G+1 into the pocket in loop B, which leads to an altered pucker of the ribose at the cleavage site.

So finally, what of nucleobase catalysis? The clear candidate in this role is G8, which is hydrogen bonded to the nucleophilic 2′-oxygen atom through the N1 atom, and to the pro-*R*p oxygen atom of the phosphate group through the exocyclic N2 atom. A catalytic role had previously been suggested for G8,[49] and this was consistent with the observation that while a G8U variant was substantially catalytically impaired, its global folding was completely unaffected.[43] Guanine is not the most obvious choice of nucleobase to act as a general base since the N1 atom is normally protonated; either adenine or cytosine would seem to be a more logical choice. It may therefore play a more significant role in the ligation reaction as a general acid. As discussed above, it would take a large change in the pK_a value to make the nucleobase function in acid/base catalysis at neutral pH values. However, the highly charged character of nucleic acids may achieve this in the environment of a catalytic pocket, probably more easily than for proteins. With NMR spectroscopy, altered pK_a values have been measured in the leadzyme, for example.[50] The charged transition state may also be stabilised by the hydrogen bond formed between the phosphate group and G8. Other nearby nucleobases could also play a role in catalysis, including A38, A9 and A10. For example, Rupert and Ferré-d'Amaré[3] suggest that if either A9 or A10 were protonated, they might participate in electrophilic catalysis by stabilising the transition state. However, an A9U variant is not seriously disabled in the cleavage reaction.[43]

In summary, the loops of the ribozyme are organised by the four-way junction to undergo close association, which thereby creates the local environment in which catalysis of cleavage or ligation can proceed. These interactions bring about a marked distortion of the trajectory of the ribose – phosphate backbone at the cleavage site, to give an in-line arrangement of the attacking nucleophile and the scissile bond. Meanwhile, the base of G8 is poised in an adjacent position, ready to participate in the chemistry of the transesterification reactions. Thus, we find ourselves significantly further down the road of understanding RNA catalysis. Of course, a crystal structure is not an endpoint, but a new beginning. As a result of this work we are now much better placed to probe the detailed mechanisms of RNA folding and catalysis, and we await future developments with great anticipation.

I gratefully thank Dr. Adrian Ferré-d'Amaré for provision of coordinates ahead of the release. I thank my colleagues Dr. T. Wilson, Dr. D. Lafontaine and Dr. D. Norman for valuable discussion, and Dr. David Norman for the molecular graphics. This laboratory is supported by the Cancer Research Campaign.

[1] P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, *Science* **2000**, *289*, 920 – 930.
[2] P. M. Gordon, E. J. Sontheimer, J. A. Piccirilli, *RNA* **2000**, *6*, 199 – 205.
[3] P. B. Rupert, A. R. Ferré-d'Amaré, *Nature* **2001**, *410*, 780 – 786.
[4] J. M. Buzayan, W. L. Gerlach, G. Bruening, *Nature* **1986**, *323*, 349 – 353.
[5] P. A. Feldstein, J. M. Buzayan, G. Bruening, *Gene* **1989**, *82*, 53 – 61.
[6] M. J. Fedor, *J. Mol. Biol.* **2000**, *297*, 269 – 291.
[7] A. C. Forster, R. H. Symons, *Cell* **1987**, *49*, 211 – 220.
[8] J. P. Hazeloff, W. L. Gerlach, *Nature* **1988**, *334*, 585 – 591.

[9] L. Sharmeen, M. Y. Kuo, G. Dinter-Gottlieb, J. Taylor, *J. Virol.* **1988**, *62*, 2674 – 2679.
[10] B. J. Saville, R. A. Collins, *Cell* **1990**, *61*, 685 – 696.
[11] R. S. Brown, J. C. Dewan, A. Klug, *Biochemistry* **1985**, *24*, 4785 – 4801.
[12] H. van Tol, J. M. Buzayan, P. A. Feldstein, F. Eckstein, G. Bruening, *Nucleic Acids Res.* **1990**, *18*, 1971 – 1975.
[13] L. A. Hegg, M. J. Fedor, *Biochemistry* **1995**, *34*, 15 813 – 15 828.
[14] M. J. Fedor, *Biochemistry* **1999**, *38*, 11 040 – 11 050.
[15] F. M. Richards, H. W. Wyckoff in *The Enzymes, Vol. 4* (Ed.: P. D. Boyer), Academic Press, New York, **1971**, pp. 647 – 806.
[16] A. T. Perrotta, I. Shih, M. D. Been, *Science* **1999**, *286*, 123 – 126.
[17] F. Walter, A. I. H. Murchie, J. B. Thomson, D. M. J. Lilley, *Biochemistry* **1998**, *37*, 14 195 – 14 203.
[18] E. C. Scott, O. C. Uhlenbeck, *Nucleic Acids Res.* **1999**, *27*, 479 – 484.
[19] A. Hampel, J. A. Cowan, *Chem. Biol.* **1997**, *4*, 513 – 517.
[20] S. Nesbitt, L. A. Hegg, M. J. Fedor, *Chem. Biol.* **1997**, *4*, 619 – 630.
[21] K. J. Young, F. Gill, J. A. Grasby, *Nucleic Acids Res.* **1997**, *25*, 3760 – 3766.
[22] J. B. Murray, A. A. Seyhan, N. G. Walter, J. M. Burke, W. G. Scott, *Chem. Biol.* **1998**, *5*, 587 – 595.
[23] A. Berzal-Herranz, J. Simpson, B. M. Chowrira, S. E. Butcher, J. M. Burke, *EMBO. J.* **1993**, *12*, 2567 – 2574.
[24] P. Anderson, J. Monforte, R. Tritz, S. Nesbitt, J. Hearst, A. Hampel, *Nucleic Acids Res.* **1994**, *22*, 1096 – 1100.
[25] A. Siwkowski, R. Shippy, A. Hampel, *Biochemistry* **1997**, *36*, 3930 – 3940.
[26] S. E. Butcher, J. E. Heckman, J. M. Burke, *J. Biol. Chem.* **1995**, *270*, 29 648 – 29 651.
[27] Y. Komatsu, I. Kanzaki, E. Ohtsuka, *Biochemistry* **1996**, *35*, 9815 – 9820.
[28] Y. Komatsu, I. Kanzaki, M. Shirai, E. Ohtsuka, *Biochemistry* **1997**, *36*, 9935 – 9940.
[29] D. J. Earnshaw, B. Masquida, S. Müller, S. T. Sigurdsson, F. Eckstein, E. Westhof, M. J. Gait, *J. Mol. Biol.* **1997**, *274*, 197 – 212.
[30] C. Shin, J. N. Choi, S. I. Sang, J. T. Song, J. H. Ahn, J. S. Lee, Y. D. Choi, *Nucleic Acids Res.* **1996**, *24*, 2685 – 2689.
[31] A. I. H. Murchie, J. B. Thomson, F. Walter, D. M. J. Lilley, *Mol. Cell* **1998**, *1*, 873 – 881.
[32] N. G. Walter, K. J. Hampel, K. M. Brown, J. M. Burke, *EMBO J.* **1998**, *17*, 2378 – 2391.
[33] A. Hampel, R. Tritz, *Biochemistry* **1989**, *28*, 4929 – 4933.
[34] A. Hampel, R. Tritz, M. Hicks, P. Cruz, *Nucleic Acids Res.* **1990**, *18*, 299 – 304.
[35] B. M. Chowrira, A. Berzal-Herranz, J. M. Burke, *Nature* **1991**, *354*, 320 – 322.
[36] N. G. Walter, J. M. Burke, D. P. Millar, *Nat. Struct. Biol.* **1999**, *6*, 544 – 549.
[37] Z.-Y. Zhao, T. J. Wilson, K. Maxwell, D. M. J. Lilley, *RNA* **2000**, *6*, 1833 – 1846.
[38] A. R. Ferré-d'Amaré, K. Zhou, J. A. Doudna, *Nature* **1998**, *395*, 567 – 574.
[39] D. M. J. Lilley, *Q. Rev. Biophys.* **2000**, *33*, 109 – 159.
[40] Z. P. Cai, I. Tinoco, *Biochemistry* **1996**, *35*, 6026 – 6036.
[41] S. E. Butcher, F. H. Allain, J. Feigon, *Nat. Struct. Biol.* **1999**, *6*, 212 – 216.
[42] S. P. Ryder, S. A. Strobel, *J. Mol. Biol.* **1999**, *291*, 295 – 311.
[43] T. J. Wilson, Z.-Y. Zhao, K. Maxwell, L. Kontogiannis, D. M. J. Lilley, *Biochemistry* **2001**, *40*, 2291 – 2302.
[44] R. Pinard, D. Lambert, N. G. Walter, J. E. Heckman, F. Major, J. M. Burke, *Biochemistry* **1999**, *38*, 16 035 – 16 039.
[45] J. E. Wedekind, D. B. McKay, *Nat. Struct. Biol.* **1999**, *6*, 261 – 268.
[46] H. W. Pley, K. M. Flaherty, D. B. McKay, *Nature* **1994**, *372*, 68 – 74.
[47] W. G. Scott, J. T. Finch, A. Klug, *Cell* **1995**, *81*, 991 – 1002.
[48] J. B. Murray, D. P. Terwey, L. Maloney, A. Karpeisky, N. Usman, L. Beigelman, W. G. Scott, *Cell* **1998**, *92*, 665 – 673.
[49] J. Burke, personal communication.
[50] P. Legault, A. Pardi, *J. Am. Chem. Soc.* **1997**, *119*, 6621 – 6628.