

Identification in a pseudoknot of a U·G motif essential for the regulation of the expression of ribosomal protein S15

(pseudoknot/U·G base pair/RNA binding)

LIONEL BÉNARD, NATHALIE MATHY, MARIANNE GRUNBERG-MANAGO, BERNARD EHRESMANN*,
CHANTAL EHRESMANN,* AND CLAUDE PORTIER†

Unité de Propre de Recherche 9073 du Centre National de la Recherche Scientifique, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France; and *Unité de Propre de Recherche 9002 du Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, 67084 Strasbourg cedex, France

Contributed by Marianne Grunberg-Manago, December 16, 1997

ABSTRACT The ribosomal protein S15 from *Escherichia coli* binds to a pseudoknot in its own messenger. This interaction is an essential step in the mechanism of S15 translational autoregulation. In a previous study, a recognition determinant for S15 autoregulation, involving a U·G wobble pair, was located in the center of stem I of the pseudoknot. In this study, an extensive mutagenesis analysis has been conducted in and around this U·G pair by comparing the effects of these mutations on the expression level of S15. The results show that the U·G wobble pair cannot be substituted by A·G, C·A, A·C, G·U, or C·G without loss of the autocontrol. In addition, the base pair C·G, adjacent to the 5' side of U, cannot be flipped or changed to another complementary base pair without also inducing derepression of translation. A unique motif, made of only two adjacent base pairs, U·G/C·G, is essential for S15 autoregulation and is presumably involved in direct recognition by the S15 protein.

Although RNA pseudoknots have been implicated in many different regulatory functions (1–8), very little is known about how proteins specifically recognize these structures. In one of the most studied cases, ribosomal frameshifting, protein–RNA recognition studies are difficult because the pseudoknot interacts with a complex protein-synthesizing machinery. A more favorable case is that of the autoregulation of ribosomal protein S15 from *Escherichia coli*. When S15 (10 kDa) is in excess of that required for ribosome assembly, it binds specifically to a pseudoknot structure that forms transiently around the ribosome loading site of its own messenger, thus repressing its own synthesis (9–11). The secondary structure of the pseudoknot was determined *in vitro* by enzymatic and chemical probing of S15 mRNA fragments (12, 13). These studies showed that the pseudoknot is unstable and is in equilibrium with another form consisting of two stem-loops. *In vitro*, S15 binds only to the pseudoknot, as shown by the tight correlation between S15-specific mRNA binding and the ability of the bound RNA to form a pseudoknot. Protection experiments with chemical probes identified two regions shielded by S15 in the pseudoknot (*black dots* in Fig. 1): one located at the distal end of stem 2 and covering loop 1 and the other in stem 1 near the junction of the two stems (13), suggesting, at first glance, that two contact areas in the minor groove might be involved. Protection by S15 against hydroxyl radical attack is restricted to the interaction essentially around the coaxial stack of the two stems (13). In all cases, the bases protected include the U of a U₍₋₄₉₎·G₍₋₃₆₎ wobble base pair present in stem 1. *In vivo*, evidence for pseudoknot formation was obtained by analyzing

the effects of compensatory mutations on the regulatory properties of a translational fusion between *rpsO*, the S15 gene, and *lacZ* (10). An extensive mutagenesis of stem 2 failed to reveal any base involved in autoregulation or S15 binding, suggesting that putative contacts in this region identified by the chemical protection experiments might be unspecific or associated to functional groups belonging to the backbone. On the other hand, mutational analysis of stem 1 showed that a recognition determinant is located in this stem and involves a U·G wobble pair (10). Several examples are known where G·U pairs are involved in RNA recognition (14–17). However, direct recognition of this pair has not been clearly established because it was generally possible (at least *in vivo*) to substitute the wobble pair by some other mispairs, suggesting that access to the functional groups is controlled by the helix geometry. Here it is shown that, *in vivo*, U·G cannot be substituted by other canonical or noncanonical base pairs and that U·G is included in a motif that probably determines a unique configuration of the wobble pair.

MATERIALS AND METHODS

Mutagenesis. Mutations were introduced into the pseudoknot by site-directed mutagenesis (18, 19) of M13 mp8 derivatives carrying a translational fusion between *rpsO* and *lacZ* (10). After mutagenesis, an *EcoRI*–*HindIII* fragment carrying all the regulatory region was sequenced and fused in frame to the distal part of *lacZ*, which is carried by a λ derivative phage (9). Encapsidation, infection of the *E. coli* strain AB5321 (*argG*, *argE*, *his*, *rpsL*, Δ *lacX74*), and plaque screening were carried out as previously described (9, 10). Monolysogens were isolated, and their translational regulatory properties were analyzed by measuring the repression of the β -galactosidase level in the presence of a plasmid overproducing S15 *in trans* (see Table 1).

RESULTS

Previous studies (10) suggested that recognition by S15 requires specific features given by a U·G wobble pair. Recognition may be either direct if S15 interacts precisely with functional groups of the U·G pair or indirect if S15 recognition depends only on a local helical distortion promoted by U·G. Direct recognition may rely on the presence of different accessible functional groups, like the 2-amino group of guanine, which is in an unusual orientation because of the deviation from standard geometry in the helical groove of U and G relative to the bases in C·G or U·A Watson–Crick pairs (20). To distinguish between these hypotheses, autoregulation of S15 was analyzed *in vivo* in the presence of different

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/952564-4\$2.00/0
PNAS is available online at <http://www.pnas.org>.

†To whom reprint requests should be addressed. e-mail: portier@ibpc.fr.

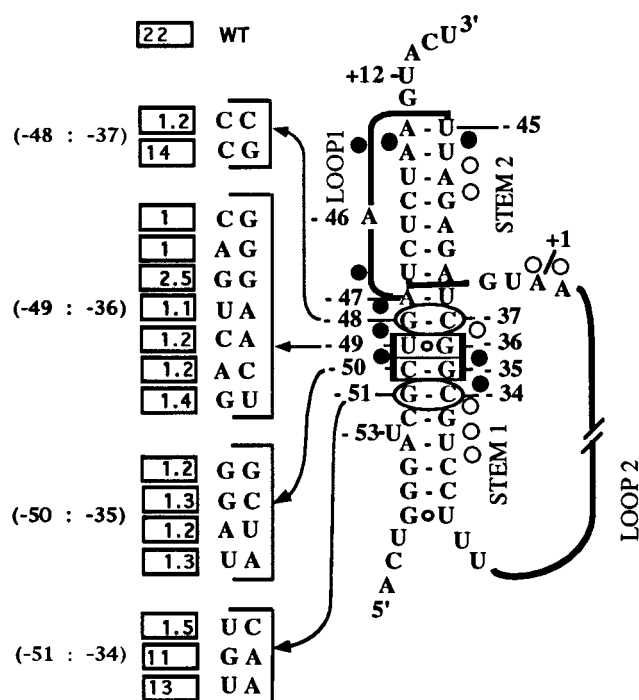


FIG. 1. Pseudoknot structure of the S15 messenger site and regulatory properties of mutations in stem 1. Base pair changes were introduced in a *rpsO-lacZ* translational fusion carried in a λ phage. For each mutation, the effect of overexpression of S15 on the level of β -galactosidase was measured and compared with the value of a strain carrying a wild-type fusion (10). The mutated base pairs indicated are described in Table 1. Repression ratios correspond to the ratio of the β -galactosidase units in the control divided by the value in the mutant and are given in the squares. This ratio is 1 when all control is lost and is 22 for the wild type (WT). The recognition determinant, G-U/G-C, is boxed, and the adjacent base pairs are ovalized. Ribose protection of the messenger binding site against ethylnitrosourea (ENU) in the presence of S15 are taken from ref. 13 and are shown by circles: ●, strong protection; ○, weak protection. +1 corresponds to the translational initiation codon.

mutations affecting the U-G pair. If direct recognition is involved, no substitution will be tolerated, whereas some substitutions might be allowed by an indirect recognition mechanism. Several mutations were introduced into a *rpsO-lacZ* translational fusion carried on the chromosome, and their effect on autoregulation was estimated from their ability to deregulate the β -galactosidase level measured in the presence of a plasmid expressing S15 in trans.

Watson-Crick Pairs Cannot Substitute for the U-G Wobble Pair. When shifting U-G for C-G, the repression ratio, which is 22 in the wild type, drops to 1 (Table 1 and Fig. 1). It has been previously shown that changing U-G for U-A also decreases the repression ratio to 1 (10). These observations clearly indicate that no Watson-Crick pair is tolerated in this position without inducing a total loss of autocontrol. These results do not distinguish between a mechanism involving direct recognition or indirect interaction. The formation of Watson-Crick base pairs presumably does not change the pseudoknot stability but rather removes or modifies the position of some specific functional group or, in another way, induces a change in the helix geometry by removing some local helix distortion.

The U-G Pair Cannot be Replaced by Other Noncanonical or Wobble Pairs. If it is just the presence of a wobble base pair that is needed to properly position the S15 recognition determinants, it should be possible to keep autocontrol *in vivo* when U-G is shifted to C-A. This mispair, which can adopt a wobble configuration when protonated, has often been shown to be able to replace U-G. This is not the case here (Fig. 1).

Table 1. *In vivo* effects of mutations on the autoregulation of a translational fusion *rpsO-lacZ*

Base pair	Mutation	β -galactosidase units in the presence of plasmids			Repression ratio
		Pseudoknot	pBR322 (control)	pBP111 (+ <i>rpsO</i>)	
	Wild type		869 \pm 50	39 \pm 3	22
G-48 C-37	C-C		3,297 \pm 264	2,768 \pm 166	1.2
G-48 C-37	C-G		2,069 \pm 69	184 \pm 14	14
U-49 G-36	C-G		3,873 \pm 160	3,460 \pm 132	1.1
U-49 G-36	A-G		1,927 \pm 199	1,894 \pm 193	1.0
U-49 G-36	G-G		3,399 \pm 198	1,418 \pm 84	2.5
U-49 G-36	U-A		2,467 \pm 56	2,322 \pm 155	1.1
U-49 G-36	C-A		3,188 \pm 81	2,641 \pm 197	1.2
U-49 G-36	A-C		3,967 \pm 565	3,293 \pm 528	1.2
U-49 G-36	G-U		4,387 \pm 119	3,131 \pm 80	1.4
C-50 G-35	G-G		1,421 \pm 94	1,392 \pm 56	1.2
C-50 G-35	G-C		3,963 \pm 66	3,131 \pm 49	1.4
C-50 G-35	A-U		3,466 \pm 281	2,994 \pm 143	1.2
C-50 G-35	U-A		3,769 \pm 468	2,947 \pm 341	1.3
G-51 C-34	G-A		2,137 \pm 68	188 \pm 39	11
G-51 C-34	U-C		2,621 \pm 88	1,689 \pm 245	1.5
G-51 C-34	U-A		1,295 \pm 42	98 \pm 39	13

To analyze the expression of the ribosomal protein S15, a translational in frame fusion between *rpsO*, the gene encoding S15, and *lacZ* was used and the level of β -galactosidase (Miller units) measured in the presence of an overproduction of S15 in trans (pBP111) and in the absence (pBR322). The ratio between the two values (repression ratio) measures the efficiency of the regulation. The higher this value, the stronger the regulation. Conversely, when no control is present, this ratio decreases to 1.

Moreover, other changes like flipping C-A to A-C, U-G to G-U have the same negative effect on regulation (Fig. 1 and Table 1). These observations might mean that the exocyclic amino group of guanosine, which is missing in C-A, is involved in S15 binding. However, replacing U-G by the noncanonical base pair A-G does not restore autocontrol. These experiments demonstrate that the polarity of U-G wobble pair is critical and that no change is tolerated at this position. The only substitution that allows just a very low level of regulation is G:G (repression ratio: 2.5, Table 1), suggesting that some limited specific interaction of S15 is still possible in this case.

U-G Is Part of a Recognition Determinant. If direct recognition is involved, specific functional groups carried by the U-G pair should interact with residues on the protein. The exocyclic amino group of guanosine, which, in a wobble pair, projects in the minor groove, might be involved. However, the proper configuration of bases in the minor groove may also be dependent on neighboring bases and/or include several other functional groups like the 2'-OH of ribose residues (21). If this is the case, flipping the base pair adjacent to the 5' side of the U in the U-G pair is predicted to modify the position of the bases in the helix (22) and hence the position of their functional groups. A total loss of autocontrol was induced when the pair $G_{(-50)}G_{(-35)}$, 5' to the U side, was flipped or replaced by A-U or U-A (Table 1 and Fig. 1). These observations demonstrate that the two stacked base pairs U-G/C-G constitute a motif that seems necessary for the proper configuration of the recognition determinants.

The RNA Determinant Is Constituted by Only Two Base Pairs. To see if the other adjacent base pairs are also involved, the pair $G_{(-48)}C_{(-37)}$ adjacent to the 3' side of the U was flipped. Only a limited effect in autoregulation was observed as shown by the repression ratio, which decreases to 14 (Table 1). Practically the same value is observed when the pair $G_{(-51)}C_{(-34)}$ is changed for U-A. Single changes at positions (-48) or (-51) create the mismatches $C_{(-48)}C$ or $U_{(-51)}C$. These changes, which are known to disrupt base pairing, abolish the

autoregulation. Interestingly, another single change ($A_{(-34)}$), induces only a limited effect on autoregulation. In contrast to the other single changes, a noncanonical base pair (G·A) is probably formed in this case, which would induce only limited structural changes and would not interfere with the access or the position of determinants. These results clearly show that the recognition determinant is constituted by only two base pairs.

DISCUSSION

All of these observations show that the recognition determinant for S15 binding is located in a motif comprising the two stacked base pairs U·G/C·G. The function of this motif, located inside stem 1 of the pseudoknot, is to properly configure the local functional groups interacting with S15. It exhibits specific structural properties linked to the presence of the wobble pair U·G. This pair is a key element that cannot be substituted by other noncanonical pairs. This observation is in contrast with other studies showing that conserved U·G pairs involved in recognition of RNA sites can be successfully substituted by other unusual base pairs. For example, a G·U pair, which constitutes a specific protein determinant in the transcript of yeast ribosomal protein L32 (17), can be shifted to G·A, G·G, U·C, and C·C (but not to C·A) and still give appreciable protein binding *in vitro*. It should be noted though that the G·U pair of the L32 messenger closes an internal loop, which is not the case for the pseudoknot studied here. In the group I intron of *Tetrahymena*, a U·G wobble pair, located near the center of helix P1, is involved in splicing, and the only change giving efficient cleavage *in vitro* (about 50% of the wild-type activity) is a change to C·A (14). Although in this example, only RNA–RNA and not an RNA–protein interaction is involved, there seems to be some similarity in the mode of ligand binding (21, 23). In another case, C·A or G·A can replace G·U in tRNA^{Ala} and still allow efficient tRNA aminoacylation *in vivo* (24) but not *in vitro* (25), suggesting that, *in vivo*, aminoacylation efficiency might be affected by other extrinsic factors. In any case, the high specificity of S15 recognition observed *in vivo* for only U·G stresses the selectivity of the recognition process and suggests that a direct contact might be involved between this base pair and the protein S15.

Another possibility is that the exact U·G configuration is determined by the adjacent base pair located at the 5' side of the U or that S15 interacts with determinants carried by the two base pairs of the motif. It has previously been shown that a U·G pair has a unique stacking configuration, whereby U·G bases stack on the bases to the 3' side of the guanine and unstack from the bases to the 5' side of the guanine (22, 23, 26, 27). In addition, NMR data on the U·G pair embedded in the P1 helix from group I self-splicing introns show that not only the positions of functional groups of the bases but also the local strand conformations are affected by U·G pairs and that these new structural features might be involved in recognition (23). Thus, in addition to the stacking discussed above, an under-twist of the U strand has been shown to be associated with an overwinding of the G strand to the 3' side of guanine. This means that functional groups like the 2'OH of the U·G and groups from adjacent pairs might then adopt a specific configuration dictated by the sequence context. This hypothesis is best illustrated by the comparison between tRNA^{Ala} and the 5' splice site of group I intron. In tRNA^{Ala}, purine nucleotides are located at either side of the guanine, instead of pyrimidines, as in the P1 helix of self-splicing introns. Several structural differences are observed by NMR between the U·G sites of these two RNAs. In tRNA^{Ala}, base stacking is limited (28), whereas it is extensive in the P1 helix so that variations in the helical twist across the base pair are small (23, 28), and backbone angles, α and ζ , instead of being in the canonical *gauche*⁻/*gauche*⁺ conformation, are *trans*/*trans*.

The $U_{(111)}\cdot G_{(9)}$ wobble pair of helix I from the 5 S RNA of *E. coli* exhibits neighboring adjacent base pairs identical to those of the S15 pseudoknot (27). If the sequence context is a primary determinant of double helical geometry, it should be possible to deduce some information on the local structure of the S15 binding site from the 5 S structure. NMR analysis has shown that, in 5 S RNA, the wobble pair stacks with the pair to the 5' side of U, whereas the bases of the pair 3' to the U are completely unstacked. This suggests that the local structure might be similar to what is observed in the P1 helix of the self-splicing intron. However, local detailed RNA structure sufficient to reveal the precise conformation of the bases and of the sugar-phosphate backbone in this region is still lacking, and so the degree of similarity between helix 1 from 5 S RNA and the P1 helix has not been determined. Thus, the extent of local helix distortions induced by the G·U wobble pair in this base pair context is not known. Only the presence of a stacking between the U·G/C·G pairs in the pseudoknot can be deduced from these studies.

The fact that S15 can apparently distinguish between U·G and G·U in an identical base pair context might be interpreted as a functional discrimination between two different angular orientations of the amino exocyclic group of guanine (29). This observation would mean that S15 recognizes specifically a single atomic group carried by the wobble pair, suggesting that the recognition of the operator site by S15 is more likely dependent upon the presence of specific functional groups than upon local helix distortions induced by the G·U base pair in its base pair context. However, other explanations can be proposed. The importance of the $C_{(-50)}\cdot G_{(-35)}$ pair adjacent to the U·G pair indicates that either other functional groups carried by this pair are recognized or that the determinant carried by U·G must be properly positioned to be recognized. In fact, both aspects might be intermingled if modifications to the helix geometry reorientate the RNA determinants, facilitating contacts with specific groups of the S15 protein. Whether these determinants correspond only to the amino exocyclic group of guanine or also to 2' hydroxyl groups (30, 31) or even to another factor (like a divalent metal ion bound near U·G) (32) remains to be demonstrated.

The other base pairs adjacent to the motif, as long as they are engaged in Watson-Crick pairing, have a very limited influence on this motif, as shown by the tolerance of complementary base substitutions in these positions. Even small structural perturbations, presumably induced by the presence of a noncanonical base pair, have practically no influence on S15 binding. Moreover, a deletion of the bulge base $U_{(-53)}$ has been shown to have no drastic effect (10). Thus, the size of the motif seems to be limited to only two base pairs. They form a quite specific structure, embedded in a stable helix, which cannot be modified without losing S15 binding. Consequently, a direct S15–RNA interaction at the U·G motif is strongly suggested and is fully supported by probing experiments with ethylnitrosourea, which showed that this region is shielded by S15 (13).

If the U·G motif is sufficient for S15 binding, a mutant unable to form the pseudoknot but still carrying the U·G motif should nevertheless be able to bind S15 and to induce auto-control. As this is not the case (10, 13), one can imagine that another contact located in another part of the pseudoknot is necessary for S15 binding. In fact, backbone probing experiments show that, in addition to the U·G region, the upper edge of stem 2 and loop 1 are also shielded by S15. Although bases do not seem to be involved in this putative contact(s) as shown by mutational analysis (10), recognition of specific functional groups in the backbone might be involved. Another possibility would be that the protections observed correspond to non-specific backbone interactions stabilizing the complex formed. This would imply that direct and indirect effects are intermingled in recognition and binding. *In vitro* chemical mutagenesis experiments should help to resolve this problem (21, 25).

In conclusion, the U·G motif is a major, if not unique, recognition determinant located in stem 1 of the pseudoknot. Because no change is tolerated at the U·G motif without a complete loss of autoregulation, a direct recognition is probably located in this area. The exocyclic amino group of guanosine, known to be projected in the minor groove, is very likely involved as well as contacts with ribose in the adjacent C₍₋₅₀₎·G₍₋₃₅₎ base pair.

We are very grateful to C. Cachia for preparing the S15 protein and to J. Plumbridge for constructive comments on the manuscript. This work was supported by grants from Centre National de la Recherche Scientifique. L.B. was supported by a doctoral grant from the Ministère de l'Enseignement Supérieur et de la Recherche.

1. Wang, C., Le, S.-Y., Ali, N. & Siddiqui, A. (1995) *RNA* **1**, 526–537.
2. Wills, N. M., Gesteland, R. F. & Atkins, J. F. (1994) *EMBO J.* **13**, 4137–4144.
3. Chen, X., Chamorro, M., Lee, S. I., Shen, L. X., Vines, J. V., Tinoco, I. J. & Varmus, H. E. (1995) *EMBO J.* **14**, 842–852.
4. Takamatsu, N., Watanabe, Y., Meshi, T. & Okada, Y. (1990) *J. Virol.* **64**, 3686–3693.
5. Shamoo, Y., Tam, A., Konigsberg, W. H. & Williams, K. R. (1993) *J. Mol. Biol.* **232**, 89–104.
6. Tang, C. K. & Draper, D. E. (1989) *Cell* **57**, 531–536.
7. Ten Dam, E. B., Verlaan, P. W. G. & Pleij, C. W. A. (1995) *RNA* **1**, 146–154.
8. Asano, K., Kato, A., Moriwaki, H., Hama, C., Shiba, K. & Mizobuchi, K. (1991) *J. Biol. Chem.* **266**, 3774–3781.
9. Portier, C., Dondon, L. & Grunberg-Manago, M. (1990) *J. Mol. Biol.* **211**, 407–414.
10. Bénard, L., Philippe, C., Dondon, L., Grunberg-Manago, M., Ehresmann, B., Ehresmann, C. & Portier, C. (1994) *Mol. Microbiol.* **14**, 31–40.
11. Philippe, C., Eyermann, F., Bénard, L., Portier, C., Ehresmann, B. & Ehresmann, C. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4394–4398.
12. Philippe, C., Portier, C., Mougél, M., Grunberg-Manago, M., Ebel, J. P., Ehresmann, B. & Ehresmann, C. (1990) *J. Mol. Biol.* **211**, 415–426.
13. Philippe, C., Bénard, L., Portier, C., Westhof, E., Ehresmann, B. & Ehresmann, C. (1995) *Nucleic Acids Res.* **23**, 18–28.
14. Doudna, J. A., Cormack, B. P. & Szostak, J. W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7402–7406.
15. McClain, W. H. & Foss, K. (1988) *Science* **240**, 793–796.
16. Hou, Y.-M. & Schimmel, P. (1988) *Nature (London)* **333**, 140–145.
17. White, S. A. & Li, H. (1996) *RNA* **2**, 226–234.
18. Nakamaye, K. L. & Eckstein, F. (1986) *Nucleic Acids Res.* **78**, 9679–9698.
19. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 488–492.
20. Sängér, W. (1984) in *Principles of Nucleic Acid Structure*, ed. Cantor, C. R. (Springer, Heidelberg), pp. 116–158.
21. Strobel, S. A. & Cech, T. R. (1995) *Science* **267**, 675–678.
22. van Knippenberg, P. H., Formenoy, L. J. & Heus, H. A. (1990) *Biochim. Biophys. Acta* **1050**, 14–17.
23. Allain, F. H.-T. & Varani, G. (1995) *J. Mol. Biol.* **250**, 333–353.
24. Gabriel, K., Schneider, J. & McClain, H. (1996) *Science* **271**, 195–197.
25. Beuning, P. J., Yang, F., Schimmel, P. & Musier-Forsyth, K. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 10150–10154.
26. Mizuno, H. & Sundaralingam, M. (1978) *Nucleic Acids Res.* **5**, 4451–4461.
27. White, S. A., Nilges, M., Huang, A., Brünger, A. T. & Moore, P. B. (1992) *Biochemistry* **31**, 1610–1621.
28. Ramos, A. & Varani, G. (1997) *Nucleic Acids Res.* **25**, 2083–2090.
29. Frugier, M. & Schimmel, P. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 11291–11294.
30. Holbrook, S. R., Cheong, C., Tinoco, I. & Kim, S.-H. (1991) *Nature (London)* **353**, 579–581.
31. Musier-Forsyth, K. & Schimmel, P. (1992) *Nature (London)* **357**, 513–515.
32. Allain, F. H.-T. & Varani, G. (1995) *Nucleic Acids Res.* **23**, 341–350.