Proc. Natl. Acad. Sci. USA
Vol 91, pp. 11283–11286, November 1994

Commentary

Possible role of aminoacyl-RNA complexes in noncoded peptide synthesis and origin of coded synthesis

Paul Schimmel and Barry Henderson

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

The universal system of protein synthesis makes polypeptides of defined sequences from the assembly of aminoacylated tRNAs on a mRNA template embedded in the ribosomal machinery. Protein synthesis was not always done this way—namely, there had to be an earlier system that provided a transition from the RNA theater to the world of proteins. Because they bear the activated amino acids for protein synthesis and have structures resembling those needed for replication of some RNA genomes, tRNAs have been the focus of analysis and speculation on the routes to the development of peptide synthesis (1). Here we suggest one way that aminoacyl-RNA progenitors of tRNAs might combine to assemble peptides without the benefit of the well-honed genetic code and ribosomal machinery that characterize all living systems. This mechanism for assembly of aminoacyl-RNA complexes affords a route to noncoded peptide synthesis that can easily be converted to a system for coded (template-dependent) synthesis.

RNA Oligonucleotides as Substrates for Aminoacylation

Separate Origin for Two Domains of tRNAs. A typical tRNA comprises 75–93 nucleotides that form a cloverleaf secondary structure that in turn is folded in three dimensions into an L-shaped molecule (Fig. 1) (2). Folded in this way, the tRNA has two clear domains which segregate the amino acid attachment site and its flanking nucleotides from the anticodon trinucleotide sequence that, in the algorithm of the genetic code, corresponds to the attached amino acid. These two domains are the acceptor-TψC stem–loop (ψ-pseudouridine) and the dihydrouridine-anticodon stem–biloop, respectively. From independent considerations, three groups (3–6) suggested that these two domains originally existed separately (either as distinct molecules or as discrete domains in a single large RNA molecule) and later were combined into a single two-domain RNA.

Operational RNA Code for Amino Acids. Regardless of whether the two domains of a contemporary tRNA had separate origins, we now know that the tRNA fold *per se* is not required for aminoacylation. For many tRNAs the entire anticodon-containing domain is dispensable for aminoacylation. The specific aminoacylations of RNA oligonucleotides whose sequences are based on tRNA acceptor stems constitute an operational RNA code for aminoacylation that relates sequences/structures in tRNA acceptor stems to specific amino acids (5).

This operational RNA code possibly originated from ribozyme-catalyzed aminoacylations of oligonucleotide substrates. For example, RNA-stimulated aminoacyl-tRNA ester hydrolysis demonstrated the capacity of an RNA molecule to form a transition state for making and breaking an aminoacyl linkage with the 3′-terminal hydroxyl group (7). The recent report of RNA catalysis of the isomerization of a bridged biphenyl further demonstrated the capacity of RNA molecules to catalyze reactions involving not just nucleic acid reactants (8).

The peptides that eventually formed from reactions of early aminoacyl-RNAs could lead to primordial tRNA synthetases, which took over the aminoacylation of RNA molecules and further developed the operational RNA code for amino acids. Contemporary aminoacyl-tRNA synthetases are divided into two distinct classes based on structural motifs that are unique to members of each class (9–11). The overall structures of the enzymes of both classes are roughly organized into two distinct domains that interact with the two domains in a tRNA (5). One structural unit is the class-defining catalytic domain, which contains the site for amino acid activation and for coupling the amino acid to the 3′-end of the tRNA. This domain also has insertions that facilitate interactions with the acceptor stem and provide the basis for "reading" the operational RNA code. This class-defining structural unit of a tRNA synthetase is believed to be the primordial enzyme. Unlike the class-defining catalytic domain, the structures of the second domains of the various synthetases of the same class diverge considerably. In addition, the sequence of this domain for any one particular synthetase is under far less selective pressure than is the class-defining catalytic domain (12). For those synthetases that interact with their anticodons, the second domain provides for those interactions (13–17). This domain is not directly involved in reading the operational RNA code and was probably a much later addition to the synthetases.

Aminoacyl-RNA Complexes for Peptide Bond Formation

Lateral Interactions That Facilitate Formation of Complexes. The transition to tRNA synthetase-like aminoacylation catalysts required a mechanism to make peptides. Aminoacyl-RNAs generated from ribozyme-catalyzed reactions are a potential source of activated amino acids. One problem is how to bring two aminoacyl groups into close proximity so that spontaneous peptide bond formation can occur, owing to the lower energy of the amide versus the ester bond. A second problem is how to make longer oligopeptides, with the assistance of a primordial peptidyltransferase ribozyme. For the sake of concreteness, we attempted to solve these problems in theory by restricting ourselves to consideration of complexes that involve familiar base-pairing and base-stacking interactions between aminoacyl-RNAs themselves or between aminoacyl-RNAs and other RNAs, and that require no specialized three-dimensional matrices to sterically confine the aminoacyl-RNAs in close proximity. We also have focused on a mechanism that could evolve naturally from noncoded to coded (template driven) peptide synthesis or could incorporate elements of both noncoded and coded synthesis at an early stage.

The solutions to these problems are made easier by the removal of the structural constraints imposed by a tRNA-like fold. In particular, hairpin-like aminoacyl-RNA oligonucleotides could bind near each other via hydrogen bonding interactions of the loops to a common template or RNA grid that does not necessarily have any coding capacity (Fig. 2*A*). The many degrees of freedom or "floppiness" of this kind of assembly suggests that the efficiency of peptide bond formation would be low (unless the

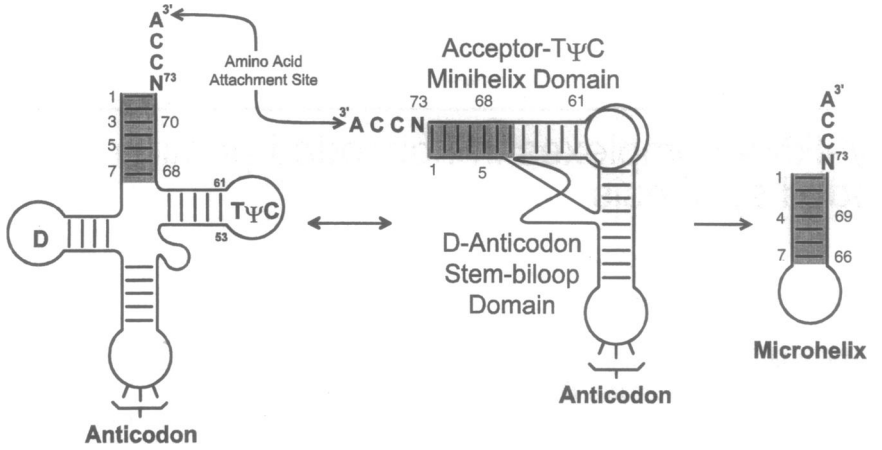

FIG. 1. Basic organization of a tRNA with the two domains formed by the acceptor-TψC stem–loop and the dihydrouridine-anticodon stem–loop structure. A hairpin oligonucleotide based on the acceptor stem including the NCCA single-stranded tetranucleotide is indicated.

assembly is sterically constrained within an RNA matrix that could be, for example, the precursor to a ribosomal RNA). A better solution is for the aminoacyl-RNA oligonucleotides themselves to join together through "side-by-side" or "lateral" complementary base pairing of loop nucleotides. These structures could be considered as intermolecular [as opposed to the more familiar intramolecular (18)] pseudoknots (Fig. 2 *b* and *c*).

Eisinger (19) and Grosjean *et al.* (20) demonstrated the facility for the formation of RNA complexes through loop–loop interactions of tRNAs with complementary anticodons. In their examples, the "head-to-head" interactions of the tRNAs by anti-parallel base pairing of anticodon trinucleotides place the amino acid attachment sites at a maximal separation, where no chemistry involving the 3′ ends can occur. Moreover, when the same element in two different RNAs interacts by complementary pairing, there is no possibility for oligomers greater than dimers to form. Similarly, the novel "kissing" hairpins described by Chang and Tinoco, which are based on the transactivation response element (TAR) of the HIV-1 mRNA, form "head-to-head" hydrogen-bonded complexes with significant separation of the ends of the helical stems associated with the hairpin loops (21).

In contrast, by having "side-by-side" interactions between loops, with the complementarity built into the opposite sides of the loops, the aminoacyl stems are potentially brought into close proximity. In addition, oligomers can in principle build up, provided that the complementary loops are designed with appropriate extra nucleotides that bridge the major and minor grooves of the quasi-continuous helices of the pseudoknots (22).

For each oligonucleotide, the sequence elements needed for specificity are con-

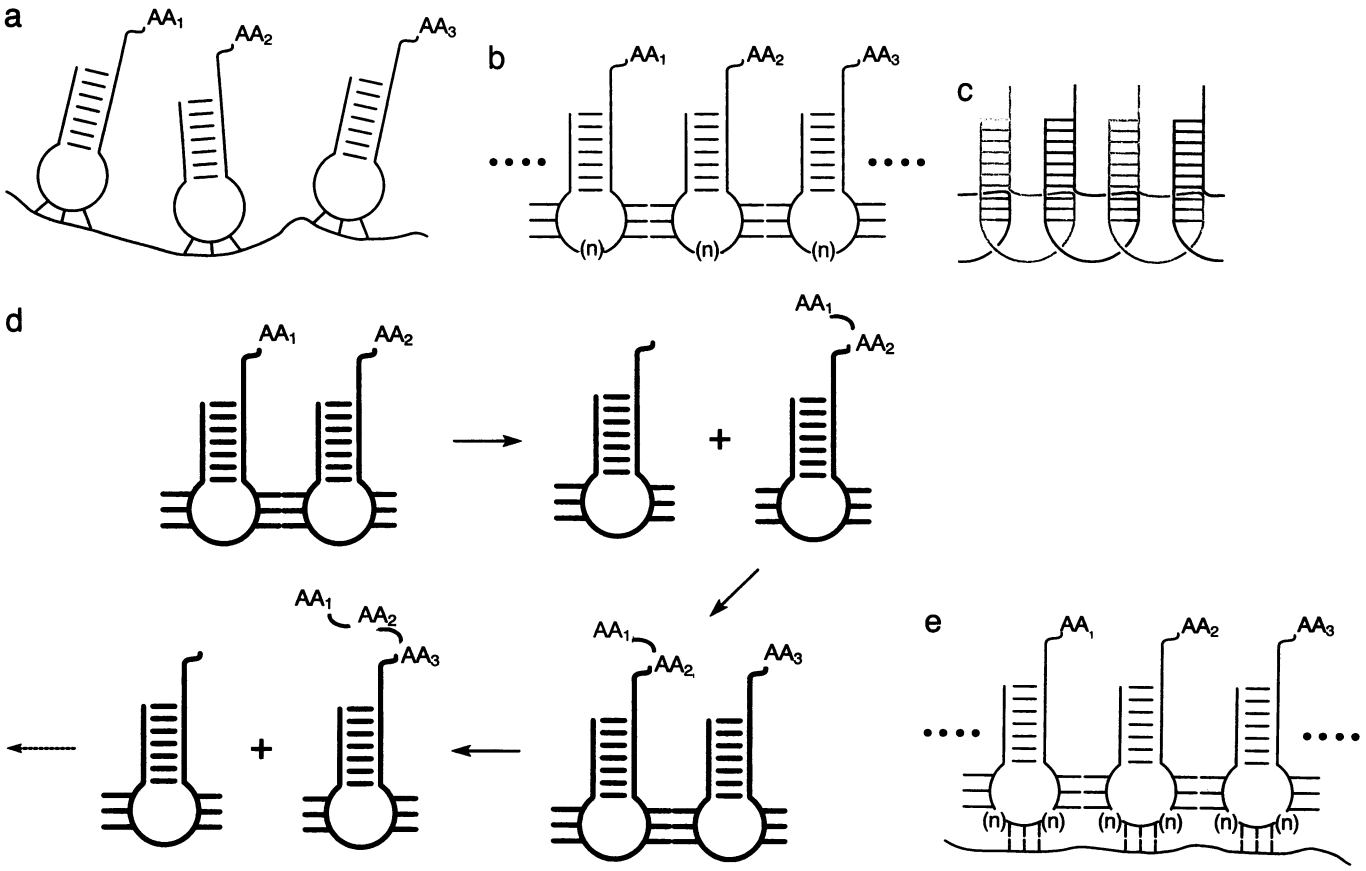

FIG. 2. Aminoacyl-RNA arrays. (*a*) Association of aminocyl-RNA hairpins with a "template" or "grid" to give a relatively unconstrained "floppy" array. (*b*) Association of aminoacyl-RNAs by complementary loop–loop interactions. The assembly of these sorts of structures is sensitive to the size of the loop. (*c*) Schematic representation of the intermolecular pseudoknot structures depicted in *b*. (*d*) Scheme for making peptides from aminoacyl-RNAs that make complementary base pairs by using either side of the loop in an alternating fashion. (*e*) Association of aminoacyl-RNAs by complementary loop–loop interactions and by interactions with a template.

tained in the first few base pairs of the helical stem and the first single-stranded base N (discriminator nucleotide) of the 3'-terminal NCCA tetranucleotide found in all tRNAs (23–27). Thus, the sequence constraints on the loop interactions need not interfere with those that determine aminoacylation specificity and vice versa. Through a system of self-complementary loop–loop interactions, a complex of two or more highly constrained aminoacyl-RNA molecules might assemble. (Computer simulations of assemblies of this sort suggest that an envelope of sterically allowed conformations brings the aminoacyl groups to within 11 angstroms or less of each other (unpublished observations)). Aminoacyl groups could then polymerize into oligopeptides.

Alternatively, with just two aminoacyl-RNAs transiently bound together, the sequential binding and dissociation of different aminoacyl-RNAs could lead to the build-up of larger peptides, initially in the form of peptidyl-RNAs. In this case, opposite sides of the loop are alternately used to build longer peptides (Fig. 2*d*). Thus, the system needs mechanisms to bind and release RNA oligonucleotides. These mechanisms could include exploiting thermal fluctuations in the environment or having conformational changes in the RNA oligonucleotides themselves affect the lifetimes of the complexes.

In all instances, the amino group of one aminoacyl-RNA attacks the carbonyl carbon of its neighbor to generate a peptidyl-RNA. This reaction would be greatly stimulated by a ribozyme peptidyltransferase. The peptidyl group moves to the next aminoacyl-RNA neighbor as the next reaction occurs and so on, somewhat like the contemporary system of protein synthesis where an aminoacyl-tRNA binds to the A-site on the ribosome, and the aminoacyl group then reacts with the carbonyl carbon of the peptidyl group of the peptidyl-tRNA in the adjacent P-site (28). In the depiction of Fig. 2*d*, the opposite sides of each loop serve as an "A" or "P" site.

Noncoded and Coded Peptide Synthesis. If the loop sequence is fixed for each amino acid-specific RNA, then a given aminoacyl-RNA oligonucleotide always has the same neighbor. Thus, peptides might vary in length but not in sequence. This constraint is removed by allowing the same helix-NCCA to be joined with different loop sequences. This possibility is plausible because of the aforementioned independence of the operational RNA code (imbedded in the helical stems and discriminatory bases) from the specific loop sequences. To explore all of sequence space with 20 amino acids, each amino acid needs to be aminoacylated onto a common acceptor stem that is joined to 400 different loop sequences (to enable pairing of each side of the loop with RNAs bearing each of the 20 amino acids). Thus, a total of 8000 different RNA oligonucleotides (400 loop sequences per acceptor stem × 20 different stem sequences) are needed to encode all possible peptides. (A substantial reduction of the required number of RNAs per amino acid would occur in a system that only partially explored all of sequence space by using fewer amino acids.) Contemporary protein synthesis with 20 amino acids can function with as few as 22 tRNAs, as seen in mitochondrial systems (29).

The systems shown in Fig. 2 *b–d* are examples of "noncoded" peptide synthesis that could lead to specific peptide sequences. The transition from noncoded to coded peptide synthesis could occur through copying by a primitive polymerase of "anticodon-like" bases in the RNA oligonucleotides that assembled together as shown in Fig. 2 *c* and *d*. This copying would yield a primitive mRNA that was then used as a template for coded peptide synthesis. The addition of a coding template (Fig. 2*e*) could reduce the number of required RNA oligonucleotides to 20 (1 per amino acid) and would introduce true coding. The reduction in the number of required oligonucleotides assumes that each can be distinguished by the nucleotide sequence that interacts with the template, while at the same time having the same "lateral" sequences hold the RNAs together by complementary loop–loop interactions. [Crick *et al.* (30) have suggested that some sort of RNA–RNA interactions might stabilize tRNA molecules bound to a message in an early system of protein synthesis.] Polypeptide sequence specificity would then be entirely dependent on the incremental stability imparted by loop–template interactions against the background of common loop–loop interactions. However, the additional stacking interactions gained by loop–template pairing (including the possibility that loop–template interactions made base stacking continuous from side-to-side) could give several kilocalories of incremental preference to the loop–template complexes relative to the loop–loop complexes. Similarly, in contemporary protein–nucleic acid complexes, the nonspecific component of the interaction is often relatively large compared to the incremental stability that arises from the specific component.

The operational RNA code for amino acids only requires that an aminoacylation-conferring, specific RNA sequence/structure be placed near the site of amino acid attachment. This requirement may reflect an early development of small RNA and subsequently protein catalysts that could not extend more than a few base pairs beyond the site of aminoacylation. Thus, although Fig. 2 *b–d* suggests how relatively simple aminoacyl-RNA oligonucleotides could be used to achieve a less developed system of protein synthesis, the basic idea of intermolecular pseudoknots as a way to form aminoacyl-RNA arrays is not restricted to simple RNA oligonucleotides. Larger and more complex RNAs may have been precursors to tRNAs. Thus, tRNA-like group I intron analogs (31) that were aminoacylated with specific amino acids could be designed with complementary loop–loop interactions that facilitated self-association and the building up of aminoacyl-RNA complexes. Also, the interactions that bind together the aminoacyl-RNAs in a complex need not involve solely loop–loop interactions but could include base pairing between complementary non-loop single-stranded sections of aminoacyl-RNA neighbors.

Refinement of the System. The lateral interactions that give rise to the aminoacyl-RNA arrays depicted in Fig. 2 *b* and *c* obviate the need in principle for a proto-ribosome to align the charged RNAs and, in this sense, represent a reduction in the number of components needed for peptide synthesis. [Smith and Yarus (32) presented evidence for lateral contacts between anticodon loops in ribosome-bound tRNAs in contemporary systems. These contacts may be remnants in some way of those proposed here to occur between the loops of aminoacyl-RNA molecules.] However, intermolecular interactions between aminoacyl-RNA molecules that took place in isolation (Fig. 2 *b–e*) might be further promoted within the steric confines of an RNA matrix or "aptamer." Matrices of this sort could be generated in the laboratory by RNA selection strategies (33–35). In addition, the greater coupling efficiency within a matrix could be enhanced with a peptidyltransferase ribozyme. Eventually, an RNA aptamer itself might take over completely the function of bringing aminoacyl groups into close proximity and even incorporate a peptidyltransferase ribozyme activity (36). At the same time, the template reading function would eventually be carried out by a separate, second domain that is joined to the operational RNA code at the acceptor end of the molecule to give rise to a tRNA-like molecule (5, 37).

Thus, starting with an operational RNA code for amino acids and RNA oligonucleotide substrates for aminoacylation, a plausible pathway to the contemporary system of protein synthesis can be imagined. While these proposals are speculative, they point out a direction for specific experiments. In particular, laboratory demonstration of "side-by-side" interactions between hairpin loops to give intermolecular pseudoknots of the type shown in Fig. 2 would advance

considerably the plausibility of these ideas.

We thank Profs. Harry Noller, Jack Szostak, and Alan Weiner for suggestions and comments on an earlier draft of this manuscript. We also thank Douglas Auld and Eric Schmidt for suggestions. This work was supported by Grant GM 15539 from the National Institutes of Health and by a grant from the National Foundation for Cancer Research. B.H. is a postdoctoral fellow of the National Institutes of Health.

1. Weiner, A. M. & Maizels, N. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7383–7387.
2. Sprinzl, M., Hartmann, T., Weber, J., Blank, J. & Zeidles, R. (1989) *Nucleic Acids Res.* **17**, r1–r172.
3. Noller, H. F. (1993) in *The RNA World*, eds. Gesteland, R. F. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 137–156.
4. Maizels, N. & Weiner, A. M. (1993) in *The RNA World*, eds. Gesteland, R. F. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 577–602.
5. Schimmel, P., Giege, R., Moras, D. & Yokoyama, S. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 8763–8768.
6. Maizels, N. & Weiner, A. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 6729–6734.
7. Piccirilli, J. A., McConnell, T. S., Zaug, A. J., Noller, H. F. & Cech, T. R. (1992) *Science* **256**, 1420–1424.
8. Prudent, J. R., Uno, T. & Schultz, P. G. (1994) *Science* **264**, 1924–1927.
9. Cusack, S., Hartlein, M. & Leberman, R. (1991) *Nucleic Acids Res.* **19**, 3489–3498.
10. Burbaum, J. J. & Schimmel, P. (1991) *J. Biol. Chem.* **266**, 16965–16968.
11. Moras, D. (1992) *Trends Biochem. Sci.* **17**, 159–164.
12. Shiba, K., Suzuki, N., Shigesada, K., Namba, Y., Schimmel, P. & Noda, T. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 7435–7439.
13. Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1989) *Science* **246**, 1135–1142.
14. Brunie, S., Zelwer, C. & Risler, J. (1990) *J. Mol. Biol.* **216**, 411–424.
15. Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. & Moras, D. (1991) *Science* **252**, 1682–1689.
16. Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1991) *Nature (London)* **352**, 213–218.
17. Cavarelli, J., Rees, B., Ruff, M., Thierry, J.-C. & Moras, D. (1993) *Nature (London)* **362**, 181–184.
18. Dam, t. D., Pleij, K. & Draper, D. (1992) *Biochemistry* **31**, 11665–11676.
19. Eisinger, J. (1971) *Biochem. Biophys. Res. Commun.* **43**, 854–861.
20. Grosjean, H., Söll, D. G. & Crothers, D. M. (1976) *J. Mol. Biol.* **103**, 499–519.
21. Chang, K.-Y. & Tinoco, I., Jr. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 8705–8709.
22. Wyatt, J. R. & Tinoco, I., Jr. (1993) in *The RNA World*, eds. Gesteland, R. F. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 465–496.
23. Francklyn, C., Shi, J.-P. & Schimmel, P. (1992) *Science* **255**, 1121–1125.
24. Musier-Forsyth, K. & Schimmel, P. (1993) *FASEB J.* **7**, 282–289.
25. Nureki, O., Niimi, T., Muto, Y., Kanno, H., Kohno, T., Muramatsu, T., Kawai, G., Miyazawa, T., Giege, R., Florentz, C. & Yokoyama, S. (1993) in *The Translation Apparatus*, eds. Nierhaus, K. H., Franceschi, F., Subramanian, A. R., Erdmann, V. A. & Wittmann-Leibold, B. (Plenum, New York), pp. 59–66.
26. Sampson, J. R. & Saks, M. E. (1993) *Nucleic Acids Res.* **21**, 4467–4475.
27. Frugier, M., Florentz, C. & Giege, R. (1994) *EMBO J.* **13**, 2218–2226.
28. Noller, H. F. (1991) *Annu. Rev. Biochem.* **60**, 191–227.
29. Breitenberger, C. A. & RajBhandary, U. L. (1985) *Trends Biochem. Sci.* **10**, 478–483.
30. Crick, F. H. C., Brenner, S., Klug, A. & Pieczenik, G. (1976) *Orig. Life* **7**, 389–397.
31. Francois, M. & Westhof, E. (1990) *J. Mol. Biol.* **216**, 585–610.
32. Smith, D. & Yarus, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4397–4401.
33. Robertson, D. L. & Joyce, G. F. (1990) *Nature (London)* **344**, 467–468.
34. Ellington, A. D. & Szostak, J. W. (1990) *Nature (London)* **346**, 818–822.
35. Tuerk, C. & Gold, L. (1990) *Science* **249**, 505–510.
36. Noller, H. F., Hoffarth, V. & Zimniak, L. (1992) *Science* **256**, 1416–1419.
37. Buechter, D. D. & Schimmel, P. (1993) *Crit. Rev. Biochem. Mol. Biol.* **28**, 309–322.