Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 10007–10009, September 1997

Commentary

When protein engineering confronts the tRNA world

Paul Schimmel and Dieter Söll†‡*

**Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and †Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520*

A long-outstanding goal of protein engineering has run up against the tRNA world, a situation that, in turn, makes the achievement of that goal a formidable challenge. The goal is to expand the genetic code to allow for incorporation of additional, nonnatural, amino acids at predetermined sites in proteins that are synthesized *in vivo*. Achieving this goal would enable a variety of applications, including structure–function analysis of specific sites in proteins by means of probes inserted at defined locations and the creation of proteins with new chemical activities. With protein-based therapeutics now realized with examples such as erythropoietin, growth hormone, and α-interferon, among others, the possibility of therapeutic proteins with novel chemical or biological substituents can also be taken seriously. But to synthesize such proteins requires a major intrusion into the tRNA world and some of the complexities that it contains. A paper by Liu *et al*. (1) in this issue illustrates the point.

The tRNA World

The tRNA world encompasses a large area of biology and chemistry (2). It includes all of the reactions and components of the translation apparatus that have tRNA-dependent interactions—aminoacyl-tRNA synthetases, ribosomal proteins and RNAs, mRNAs, and factors for elongation, initiation, translocation, peptidyl transfer, and peptide release. It also includes examples of biological fine structure recognition manifested by tRNA-dependent editing and recognition reactions (3–5), the specialized splicing and processing systems for tRNA gene transcripts that have contributed much to our understanding of RNA chemistry and ribozyme-based catalysis (6), the role of tRNA structure in intron splicing and as regulatory signal for transcription or translation (7), the many striking examples of sophisticated RNA enzymology as displayed by tRNA base modification reactions (8), and tRNA-dependent amino acid transformations (9). The principle of RNA functional diversity and mimicry can be seen in the use of specific tRNAs as primers for retroviral reverse transcriptases (10) and the mimicry of tRNA structures in introns, respectively (11).

What is now clear is that the tRNA world is ancient. Contemporary translation is thought to have first appeared in a world dominated by RNA chemistry and catalysis (12) where the so-called minihelix domain, which is one of the two structural domains of tRNA, was likely a key player (13). Through billions of years of chemical and biological evolution the contemporary tRNA molecule arose. In the process, it developed a remarkably complex three-dimensional structure built upon a network of secondary and tertiary interactions that utilize metal ions together with each distinct chemical entity in the tRNA primary structure—ribose units, phosphate groups, and bases. It also developed sophistication in the way that each tRNA or tRNA-like structure is distinguished from all others (2). Nowhere is that more evident than in the way that tRNAs interact with components of the translation apparatus to decode genetic information.

A Challenging Goal for Protein Engineering

The genetic code is determined in aminoacylation reactions, whereby each amino acid is attached to the tRNA bearing the anticodon triplet of the codon that corresponds to that amino acid. These reactions are catalyzed by aminoacyl-tRNA synthetases. Typically there is one synthetase for each amino acid, although notable exceptions such as the formation of selenocysteinyl-tRNA (14, 15) and glutaminyl- and asparaginyl-tRNA (16) exist. Aminoacyl-tRNA synthetases have a high specificity for their amino acids and their cognate tRNAs. The enzyme–tRNA partners have also developed so-called "negative" determinants, that is, sites on their surfaces that repel the wrong tRNA from a given synthetase (17). In some cases they also catalyze editing reactions to correct errors of aminoacylation. If misacylations occur that are not corrected, then toxicity results because an amino acid is incorporated into the wrong position in a growing polypeptide chain. Misacylations have been observed when mutations are introduced into either a synthetase or tRNA.

A long-standing challenge for protein engineering is to create a new synthetase–tRNA pair; that is, a synthetase and tRNA specific for a new amino acid and, at the same time, a nucleotide triplet that corresponds to the new amino acid and tRNA. Given that 61 of the 64 nucleotide triplets of the genetic code are assigned to the 20 standard amino acids, and UGA also specifies selenocysteine, only the two remaining stop codons UAA and UAG are desirable candidates for a new codon assignment. The stop codon of choice appears to be the UAG amber codon. Amber suppressors consisting of mutant tRNAs with CUA anticodons have long been known to insert amino acids at premature UAG stop codons in mRNAs, without arresting cell growth (18).

The Use of Amber Suppressors

One approach, therefore, is to charge a tRNA amber suppressor with a nonnatural amino acid and to create a premature stop codon at the desired position in the mRNA of interest. This approach has been used successfully with *in vitro* translation systems, by the laboratories of Hecht (19), Schultz (20), and others (21). In these examples, where collectively many different nonnatural amino acids have now been introduced into a variety of proteins, the tRNA amber suppressor was charged by a combination of chemical and enzymatic methods that omitted an aminoacyl-tRNA synthetase. (One problem is that no naturally occurring tRNA synthetase can activate the virtually unlimited number of nonnatural amino acids that could be introduced into a protein.) The preparation of charged tRNAs *in vitro*, using these specialized methods, precludes doing protein synthesis *in vivo*. But, in spite of

‡To whom reprint requests should be addressed. e-mail: soll@trna.chem.yale.edu.

advances in the efficiency of *in vitro* translation systems, the yields are low while the labor and costs of materials are high, compared with *in vivo* systems. These considerations motivate the effort to develop an *in vivo* aminoacylation system for incorporation of novel amino acids into proteins (22).

The paper by Liu *et al.* (1) describes a first step toward the goal of having a nontoxic *in vivo* aminoacylation system that eventually allows the incorporation of a nonnatural amino acid at a defined position in a protein of choice. The concept is straightforward. A mutant amber suppressor tRNA is created that cannot be charged by any of the naturally occurring 20 aminoacyl-tRNA synthetases. Instead, a 21st synthetase is created that activates the desired amino acid and attaches it to the new tRNA. While the concept is straightforward, its implementation is a great challenge. In the course of their attempts, Liu *et al*. (1) learn a great deal about some of the complexities imposed by that part of the tRNA world that deals with the translation apparatus and the genetic code.

Toward the 21st Synthetase–tRNA Pair

To develop this system, they choose a synthetase–tRNA complex of known three-dimensional structure and manipulate this complex to create a new synthetase–tRNA pair. They choose the glutamine system, where not only has the structure of a cocrystal been solved (23) but also detailed functional analysis has been carried out (24). The idea was to mutate $tRNA^{Gln}_{CUA}$ at locations known to interact with the synthetase, so that the wild-type enzyme will not charge the mutant tRNA and, further, so that no other enzyme will charge it. This mutant can be designated as $^{o}tRNA^{Gln}_{CUA}$, that is, a species which is "orthogonal" to all of the others. The next step is to mutagenize glutaminyl-tRNA synthetase (GlnRS) so that it can charge the orthogonal tRNA ($^{o}tRNA^{Gln}_{CUA}$) and not its normal substrate $tRNA^{Gln}$. This synthetase can be designated as ^{o}GlnRS. Last, ^{o}GlnRS is to be mutagenized further to give ^{oo}GlnRS, which has a new amino acid specificity, directed toward the novel amino acid AA^{N}.

Thus, the basic outline is given below:

$$\text{GlnRS} + \text{Gln} + \text{ATP} + \text{tRNA}^{\text{Gln}} \rightarrow \text{Gln-tRNA}^{\text{Gln}} + \text{AMP} + \text{PP}_i + \text{GlnRS} \quad \textbf{[1]}$$

$$\text{GlnRS} + \text{Gln} + \text{ATP} + {}^{o}\text{tRNA}^{\text{G1n}}_{\text{CUA}} \rightarrow \text{no reaction} \quad \textbf{[2]}$$

$$^{o}\text{GlnRS} + \text{Gln} + \text{ATP} + \text{tRNA}^{\text{Gln}} \rightarrow \text{no reaction} \quad \textbf{[3]}$$

$$^{o}\text{GlnRS} + \text{Gln} + \text{ATP} + {}^{o}\text{tRNA}^{\text{Gln}}_{\text{CUA}} \rightarrow \text{Gln-}{}^{o}\text{tRNA}^{\text{Gln}}_{\text{CUA}} + \text{AMP} + \text{PP}_i + {}^{o}\text{GlnRS} \quad \textbf{[4]}$$

(further mutagenesis of ^{o}GlnRS to give ^{oo}GlnRS)

$$^{oo}\text{GlnRS} + \text{AA}^{\text{N}} + \text{ATP} + {}^{o}\text{tRNA}^{\text{Gln}}_{\text{CUA}} \rightarrow \text{AA}^{\text{N}}\text{-}{}^{o}\text{tRNA}^{\text{Gln}}_{\text{CUA}} + \text{AMP} + \text{PP}_i + {}^{oo}\text{GlnRS} \quad \textbf{[5]}$$

In the work of Liu *et al*. (1), *Escherichia coli* is used as the host organism. Success was achieved in obtaining a $^{o}tRNA^{Gln}_{CUA}$ that is not charged *in vivo* by wild-type GlnRS (Eq. **2**) or by any other wild-type *E. coli* aminoacyl-tRNA synthetase. Subsequent work showed that, when charged, $^{o}tRNA^{Gln}_{CUA}$ was active in translation. Moreover, a ^{o}GlnRS that acylates the orthogonal $^{o}tRNA^{Gln}_{CUA}$ (Eq. **4**) was created, thus giving a new synthetase–tRNA pair.

These achievements required considerable effort and revealed some of the inherent complexities encountered when protein engineering confronts the tRNA world. One example is the creation of $^{o}tRNA^{Gln}_{CUA}$. Three major contact points of $tRNA^{Gln}$ with GlnRS were manipulated, in an effort to eliminate charging with the wild-type enzyme. These contacts are referred to as knobs 1, 2, and 3, and correspond to the G3·C70 base pair in the acceptor helix, the G10·C25 pair in the dihydrouridine stem, and C16 in the dihydrouridine loop. All combinations of specific mutations at these three positions were investigated in the $tRNA^{Gln}_{CUA}$ amber suppressor. Curiously, a double mutant with changes at knobs 2 and 3 was not active as an amber suppressor, even if artificially charged *in vitro* and tested in an *in vitro* translation system. In contrast, when these two mutations were combined with the mutation at knob 1, the artificially charged mutant amber suppressor was active *in vitro*. Thus, one or more components of the translation apparatus (such as elongation factor Tu or another component) other than those associated with aminoacylation were sensitive to the knob 2/knob 3 mutations. This sensitivity was compensated by adding in a knob 1 mutation. The conclusion is that the intricacies of the tRNA world are multifactorial and this feature, in turn, is critical when attempting to create an orthogonal tRNA.

The knob 1/knob 2/knob 3 mutant tRNA was inactive *in vivo* as an amber suppressor, because it was not charged efficiently by GlnRS to give enough glutaminylated tRNA for suppression. (Its charging activity is 13,000-fold less than that of the wild-type $tRNA^{Gln}$.) This variant was thus chosen as $^{o}tRNA^{Gln}_{CUA}$. It was ideal for selecting *in vivo* mutants of GlnRS that can catalyze its aminoacylation. These mutants of GlnRS were generated by an elaborate scheme that allowed for compensatory mutations in the synthetase at the sites that make contacts with the three knobs, in addition to allowing for mutations at other locations in the synthetase structure. Altogether, seven rounds of reiterative mutagenesis and selection were performed. A sampling of the sequences of some of the selected mutant enzymes showed that multiple substitutions had occurred (roughly 15 nonsilent changes on average, in those that were checked).

The best mutant obtained at the end of these selections had an activity on $^{o}tRNA^{Gln}_{CUA}$ of roughly 10% compared with that on wild-type $tRNA^{Gln}$. Thus, while charging of the orthogonal tRNA by the wild-type enzyme was virtually eliminated, the best mutant variant GlnRS still preferentially charged the wild-type tRNA. Moreover, the activity on the wild-type $tRNA^{Gln}$ was substantially reduced, meaning that substitutions that facilitated recognition of $^{o}tRNA^{Gln}_{CUA}$ at the same time disrupted favorable interactions with $tRNA^{Gln}$. This situation underscores the difficulty of making rational manipulations to change specificity and retain or gain activity.

Some of the substitutions in the mutant enzymes that recognize $^{o}tRNA^{Gln}_{CUA}$ removed potential hydrogen bonding interactions with $^{o}tRNA^{Gln}_{CUA}$ at the sites that make contact with the three knobs. In addition, the mutant enzyme with the highest activity on $^{o}tRNA^{Gln}_{CUA}$ had no substitutions at the positions in the protein that interact with the three knobs, even though compensatory mutations at these positions might be expected. Mutations at other sites and indirect conformational effects presumably account for the activities of these mutant enzymes. Thus, aside from the multifactorial nature of the interactions in the parts of the tRNA world that are not involved in aminoacylation, these observations demonstrate the difficulty of rationalizing manipulations of specificity of a synthetase–tRNA complex that itself has multifactorial features.

Conclusions

The last goal (Eq. **5**) of generating a new amino acid specificity remains to be attained. Also to be attained is the isolation of an ^{o}GlnRS which, while charging $^{o}tRNA^{Gln}_{CUA}$, does not acylate the wild-type $tRNA^{Gln}$ (Eq. **3**). Will these goals be achieved?

First, there is always the problem of generating mutant enzymes that have misacylation phenotypes. These enzymes

are automatically eliminated in any selection scheme and yet, in principle, some could have the properties sought for [no activity on $tRNA^{Gln}$ (in Eq. **3**) and a new amino acid specificity (Eq. **5**)]. Any mutagenesis and selection has to yield the right combination of positive and negative determinants for amino acid and tRNA recognition.

Second, even if a new synthetase is generated with just the right combination of positive and negative determinants for a highly specific and novel amino acid and tRNA interaction, there is still the problem of potential interference from interactions with the rest of the tRNA world. For example, mutations in tRNAs or synthetases can generate new, serendipitous positive or negative interactions with other components of the translation apparatus, or with components of nontranslation systems where tRNAs or tRNA-like structures occur. The knob 2/knob 3 mutant $tRNA^{Gln}$ that is not active in protein synthesis [even if aminoacylated (see above)] is but one example where an undefined interaction (that is not part of the aminoacylation system) is defective. Thus, finding a functional 21st synthetase–tRNA pair requires the most sophisticated surgery on a tRNA world that is both vast and intricate. It is for this reason that finding the new pair would be a substantial achievement.

1. Liu, D. R., Magliery, T. J., Pastrnak, M. & Schultz, P. G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10092–10097.
2. Söll, D. & RajBhandary, U. L., eds. (1995) *tRNA: Structure, Biosynthesis, and Function* (Am. Soc. Microbiol., Washington, DC).
3. Baldwin, A. N. & Berg, P. (1966) *J. Biol. Chem.* **241,** 839–845.
4. Hale, S. P., Auld, D. S., Schmidt, E. & Schimmel, P. (1997) *Science* **276,** 1250–1252.
5. Ibba, M., Hong, K.-W., Sherman, J. M., Sever, S. & Söll, D. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 6953–6958.
6. Gopolan, V., Talbot, S. J. & Altman, S. (1994) in *RNA–Protein Interactions*, eds. Nagai, K. & Mattaj, I. (IRL Press at Oxford Univ. Press, Oxford), pp. 103–126.
7. Belfort, M. & Weiner, A. (1997) *Cell* **89,** 1003–1006.
8. Björk, G. R. (1995) *Prog. Nucleic Acid Res. Mol. Biol.* **50,** 263–338.
9. Ibba, M., Curnow, A. W. & Söll, D. (1997) *Trends Biochem. Sci.* **22,** 39–42.
10. Marquet, R., Isel, C., Ehresmann, C. & Ehresmann, B. (1995) *Biochimie* **77,** 113–124.
11. Caprara, M. G., Lehnert, V., Lambowitz, A. M. & Westhof, E. (1996) *Cell* **87,** 1135–1145.
12. Hager, A. J., Pollard, J. D. & Szostak, J. (1996) *Chem. Biol.* **3,** 717–725.
13. Schimmel, P. & Ribas de Pouplana, L. (1995) *Cell* **81,** 983–986.
14. Stadtman, T. C. (1996) *Annu. Rev. Biochem.* **65,** 83–100.
15. Wilting, R., Schorling, S., Persson, B. C. & Böck, A. (1997) *J. Mol. Biol.* **266,** 637–641.
16. Curnow, A., Ibba, M. & Söll, D. (1996) *Nature (London)* **382,** 589–590.
17. Normanly, J. & Abelson, J. (1989) *Annu. Rev. Biochem.* **58,** 1029–1049.
18. Garen, A. (1968) *Science* **160,** 149–159.
19. Heckler, T. G., Chang, L. H., Zama, Y., Naka, T., Chorghade, M. S. & Hecht, S. M. (1984) *Biochemistry* **23,** 1468–1473.
20. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. (1989) *Science* **244,** 182–188.
21. Brunner, J. (1993) *Annu. Rev. Biochem.* **62,** 483–514.
22. Ibba, M. & Hennecke, H. (1994) *Bio/Technology* **12,** 678–682.
23. Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1989) *Science* **246,** 1135–1142.
24. Ibba, M., Hong, K.-W. & Söll, D. (1996) *Genes to Cells* **1,** 429–436.