Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code

J. D. Bain*, Christopher Switzer†, A. Richard Chamberlin* & Steven A. Benner‡

* Department of Chemistry, University of California, Irvine, California 92717, USA
† Department of Chemistry, University of California, Riverside, California 9251, USA
‡ Laboratory for Organic Chemistry, Eidgenössische Technische Hochschule, CH-8092 Zurich, Switzerland

ONE serious limitation facing protein engineers is the availability of only 20 'proteinogenic' amino acids encoded by natural messenger RNA. The lack of structural diversity among these amino acids restricts the mechanistic and structural issues that can be addressed by site-directed mutagenesis. Here we describe a new technology for incorporating non-standard amino acids into polypeptides by ribosome-based translation. In this technology, the genetic code is expanded through the creation of a 65th codon–anticodon pair from unnatural nucleoside bases having non-standard hydrogen-bonding patterns[1,2]. This new codon–anticodon pair efficiently supports translation *in vitro* to yield peptides containing a non-standard amino acid. The versatility of the ribosome as a synthetic tool offers new possibilities for protein engineering, and compares favourably with another recently described approach in which the genetic code is simply rearranged to recruit stop codons to play a coding role[3–9].

To compare the alternative strategies of rearrangement and expansion of the genetic code for increasing the range of amino acids that can be incorporated into proteins by translation, two mRNA molecules were prepared (Fig. 1). Both encode identical hexadecapeptides that are preceded by identical consensus 5′-untranslated sequences[10]. One, however, has the UAG nonsense (stop) codon at position 9 as the signal for the incorporation of the non-standard amino acid L-iodotyrosine, whereas the other uses the novel non-standard (*iso*-C)AG codon (the 65th codon) at this position (Fig. 2) for the same purpose. These mRNA molecules were incubated separately with rabbit reticulocyte lysate containing L-[^{3}H]leucine and L-[^{35}S]methionine. Ribosome-mediated peptide synthesis was evaluated in the presence and absence of the corresponding charged and uncharged transfer RNA molecules incorporating either CUA or CU(*iso*-dG) as the anticodon, where *iso*-G is a non-standard purine complementary to *iso*-C[11,12] (Fig. 2). Translation products were isolated by precipitation (Table 1) or high-performance liquid chromatography (Table 2) and identified by comparison with authentic standards[13].

A 'suppression level' of 63% (Table 1) or 67% (Table 2) was observed for read-through of a UAG nonsense codon in the presence of a semi-synthetic suppressor tRNA incorporating the CUA anticodon and charged with iodotyrosine. In contrast, read-through of the (*iso*-C)AG codon was 90% (Table 1) or 91% (Table 2) in the presence of the corresponding non-standard $tRNA_{CU(iso\text{-}dG)}$ charged with iodotyrosine.

The specificity of translation of the 65th codon is high. Neither semi-synthetic suppressor $tRNA_{CUA}$ nor any natural tRNA allowed reading of the (*iso*-C)AG codon, as shown by the absence of detectable full-length product in translation mixtures containing the non-standard mRNA in the absence of charged $tRNA_{CU(iso\text{-}dG)}$. When the ribosome encountered the (*iso*-C)AG codon in the absence of charged $tRNA_{CU(iso\text{-}dG)}$, the primary outcome was continued translation following a frameshift[14] that skipped the *iso*-C base (Table 2). In contrast, attempted translation of a message containing the UAG nonsense codon in the absence of its corresponding charged $tRNA_{CUA}$ resulted in termination at this position, yielding a truncated peptide; there was no detectable frameshifting.

The data in Tables 1 and 2 illustrate the promise of technology that exploits non-standard nucleosides for expanding the genetic lexicon. In addition, they provide a small but important scientific insight suggesting how translation terminates, one of the least understood parts of the translation process. Our results can be explained most simply by the hypothesis that release factors[15] present in the translation mixture bind to the nonsense codon UAG but not to the 65th codon (*iso*-C)AG (refs 16, 17). Further, the absence of a complementary tRNA is evidently not sufficient for proper termination of translation; apparently, release factors must bind to prevent frameshifting. The relative rates of termination mediated by release factors and frameshifting control the expression of many viral proteins[18], the expression of subunits of DNA polymerase[19], and the natural reorganization of the genetic code away from the 'universal' code (as for example in mitochondria)[20]. The 65th codon offers a new experimental strategy to analyse these systems.

From a technological standpoint, the binding of release factors is also critical. Both the genetic code rearrangement and expansion strategies suffer from an intrinsic disadvantage: the low yields of cell-free translation. The presence of release factors in translation mixtures and their requirement for correct termination create a further limitation. To obtain high yields of

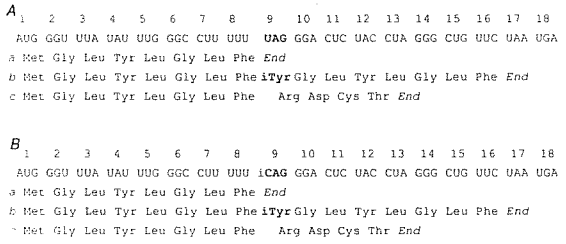

FIG. 1 The mRNA molecules used to compare rearranging (*A*) and expanding (*B*) the genetic code with the 65th codon as alternative strategies for incorporating non-standard amino acids into translated peptides. Shown are the translation products arising in the absence (*a*, octamer) and presence (*b*, hexadecamer) of suppression, and (*c*, a dodecamer) following a frame shift and successful termination at the next stop codon (iTyr, L-3-iodotyrosine).

TABLE 1 Label in precipitated products of translation

	mRNA	tRNA	Charge	^{3}H (d.p.m.)	^{35}S (d.p.m.)	Leu:Met	Suppression (%)
1	UAG	none	—	112,979	57,673	3.0	—
				134,557	67,132		
				106,506	50,561		
2	UAG	CUA	none	130,382	63,814	3.0	—
				128,061	64,307		
				120,183	59,128		
3	UAG	CU(*iso*-dG)	none	115,937	56,437	3.0	—
				126,651	61,518		
				106,834	53,890		
4	UAG	CUA	iodo-Tyr	171,434	52,732	4.9	63
				219,950	65,045		
				158,723	47,313		
5	UAG	CU(*iso*-dG)	iodo-Tyr	141,639	63,821	3.3	10
				123,592	54,144		
				123,500	57,346		
6	(*iso*-C)AG	none	—	30,340	14,994	3.0	—
				31,871	15,405		
				30,670	15,323		
7	(*iso*-C)AG	CUA	none	32,008	16,066	3.0	—
				32,465	15,574		
				34,310	16,607		
8	(*iso*-C)AG	CU(*iso*-dG)	none	30,986	14,723	3.0	—
				30,159	14,849		
				33,115	16,152		
9	(*iso*-C)AG	CUA	iodo-Tyr	35,644	17,363	3.0	—
				29,756	14,181		
				31,143	15,705		
10	(*iso*-C)AG	CU(*iso*-dG)	iodo-Tyr	235,233	60,179	5.7	90
				203,033	52,841		
				202,255	53,087		

A $tRNA^{Gly}_{CUN}$ (ref. 23) altered in sequence in the acceptor stem and anticodon loop regions and charged with iodotyrosine was prepared by ligating a truncated tRNA lacking the 3'-terminal dinucleotide with a dCA dinucleotide 2'(3') acylated with L-iodotyrosine; both of these were prepared by chemical synthesis[24,25]. This particular tRNA was chosen to make unlikely any proof reading by the 'double sieve' mechanism when charged with any non-standard amino acid[25,26]. Messages were prepared by ligation of RNA fragments in the presence of DNA splints. Details will be reported elsewhere (in preparation). Reaction mixtures (10 μl) contained rabbit reticulocyte lysate (9 μl), L-[^{35}S]methionine (15 μCi, 29.7 Ci mmol^{-1}), L-[3,4,5^{3}H]leucine (5 μCi, 19.8 Ci mmol^{-1}), mRNA (2.0 μM), L-3-iodotyrosyl-tRNA (20 μM), and $MgCl_2$ (10 mM). The reaction was quenched with water (1 ml) in the presence of carrier peptides (10 μl of a 0.5 mM solution of each peptide in 77% formic acid), and the precipitated hydrophobic peptide products recovered by centrifugation. Peptides arising from termination have a Leu:Met ratio of 3:1; those derived from complete translation to yield a 16-mer have a Leu:Met ratio of 6:1. The relative amount of terminated and full-length peptides can therefore be calculated from the Leu:Met ratio[8,13]. The hydrophilic peptides resulting from frame shifting are not recovered by this protocol; see Table 2 for their analysis. Codon and anticodon combinations are indicated. Values are from three trials. Counts per minute (c.p.m.) have been converted into disintegrations per minute (d.p.m.) and corrected for background and quenching. Leu:Met ratios were obtained by correcting for specific activities. Suppression efficiencies are obtained from the average suppression of separate runs derived from the ratio of L-[^{35}S]methionine to L-[^{3}H]leucine divided by the [^{35}S]:[^{3}H] ratio obtained from the appropriate control. Data from control experiments with radiolabelled iodotyrosine are not shown. No evidence was found to suggest that the uncharged tRNA (2, 3, 7 and 8) was charged in the lysate during the course of the translation experiment. This is probably because (1) the designed tRNA is structurally altered in the acceptor stem region[13]; (2) it contains no post-transcriptional modifications; and (3) it is derived from an *Escherichia coli* sequence. The possibility that significant amounts of translation products arise from initiation at positions other than the AUG start codon is ruled out by: (1) independent experiments measuring $^{35}S/^{125}I$ ratios in control reactions (not shown); (2) the absence of additional products containing tritium in the HPLC analysis; (3) the Leu:Met ratios of 3.0 seen in 1, 2, 3, 6, 7, 8 and 9; were translation being initiated in frame at codons following the AUG, these ratios would require that a corresponding amount of the translation initiating at the AUG be terminated before the stop codon; and (4) consistent ratios of ^{35}S in data obtained by precipitation (Table 1), where randomly initiated peptides translated in one of the other two reading frames would not precipitate, to those obtained by HPLC, where all peptides are, in principle, observable (Table 2).

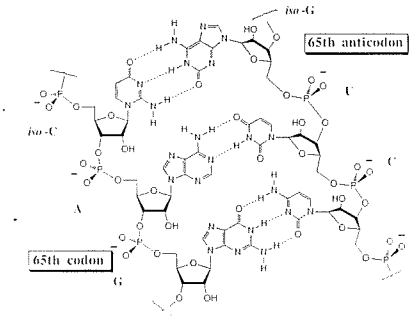

FIG. 2 The 65th codon (incorporating the non-standard nucleoside *iso*-C) and its complementary anticodon (incorporating *iso*-dG) allows the incorporation of non-standard amino acids into proteins synthesized by translation.

TABLE 2 HPLC analysis of the oligopeptides produced by translation

	tRNA	Charge	Full-length products (16-mer) ^{35}S (d.p.m.)	%	Termination products (8-mer) ^{35}S (d.p.m.)	%	Frame shift products (polar) ^{35}S (d.p.m.)	%
			mRNA containing the UAG codon					
1	none	—	799	4	20,548	96	76	—
2	CUA	none	821	4	19,291	96	29	—
3	CU(*iso*-dG)	none	524	3	17,844	97	73	—
4	CUA	iodo-Tyr	15,747	67	7,617	33	90	—
5	CU(*iso*-dG)	iodo-Tyr	1,725	9	18,153	91	47	—
			mRNA containing the (*iso*-C)AG codon					
6	none	—	1,145	3	9,195	25	25,832	71
7	CUA	none	967	4	4,597	17	21,749	80
8	CU(*iso*-dG)	none	1,023	4	3,400	14	19,078	81
9	CUA	iodo-Tyr	853	3	6,722	24	20,984	73
10	CU(*iso*-dG)	iodo-Tyr	17,754	91	1,464	8	246	1

Aliquots obtained directly from incubation mixtures were diluted with carrier peptides and injected onto a Vydac C4 column equilibrated in 0.1% trifluoroacetic acid in H_2O/CH_3CN (3:1). Both hydrophobic and hydrophilic peptide, the latter arising from frameshifting, are quantified in this way. Gradient elution with 0.1% trifluoroacetic acid in CH_3CN (25–55% CH_3CN in 60 min) resolved the 8- and 16-mer peaks, eluting at 30–45% CH_3CN. The low read-through of the UAG stop codon seen with non-standard tRNA is expected from the minor tautomeric form of *iso*-G (ref. 27).

translation products, release factors that bind UAG in competition with the semi-synthetic tRNA must be removed or inactivated. However, this would require that UAG triplet stop signals be removed to avoid continued translation after frameshifting. Although it may be possible to fulfil both of these requirements *in vitro*, they are essentially unattainable *in vivo*. Thus for the rearrangement strategy, the low yields of protein obtained from cell-free translation systems cannot be overcome simply by moving the system into a living cell.

Our demonstration that a ribosome can efficiently translate a 65th codon is the first of three breakthroughs required for incorporation of non-standard amino acids into proteins biosynthesized *in vivo*. The second is that plasmids containing a third base pair must be copied and transcribed *in vivo* with reasonable fidelity; the base pair between diaminopyrimidine and xanthosine[2] may be best suited in this regard. Given faithful replication and transcription, a total of 216 codons will be accessible as a result of the introduction of a third base pair. The third and more demanding requirement is for non-standard aminoacyl tRNA synthetases to be engineered which specifically couple a non-standard amino acid to a non-standard tRNA[21,22] □

Received 7 November 1991; accepted 13 February 1992.

1. Switzer, C. Y., Moroney, S. E. & Benner, S. A. *J. Am. chem. Soc.* **111,** 8322–8323 (1989).
2. Piccirilli, J. A., Krauch, T., Moroney, S. E. & Benner, S. A. *Nature* **343,** 33–37 (1990).
3. Bain, J. D., Glabe, C. G., Dix, T. A., Chamberlin, A. R. & Diala, E. S. *J. Am. chem. Soc.* **111,** 8013–8014 (1989).
4. Bain, J. D., Wacker, D. A., Kuo, E. E. & Chamberlin, A. R. *Tetrahedron* **14/15,** 2389–2400 (1991).
5. Hecht, S. M., Alford, B. L., Kuroda, Y. & Kitano, S. *J. biol. Chem.* **253,** 4517–4520 (1978).
6. Roesser, J. R., Chorghade, M. S. & Hecht, S. M. *Biochemistry* **25,** 6361–6365 (1986).
7. Baldini, G., Martoglio, B., Schachenmann, A., Zugliani, C. & Brunner, J. *Biochemistry* **27,** 7951–7959 (1988).
8. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. *Science* **244,** 182–188 (1989).
9. Robertson, S. A., Ellman, J. A. & Schultz, P. G. *J. Am. chem. Soc.* **113,** 2722–2729 (1991).
10. Kozak, M. *Nature* **308,** 241–246 (1984).
11. Mantsch, H. H. *et al. Biochemistry* **14,** 5593–5601 (1975).
12. McBride, L. J., Kierzek, R., Beaucage, S. L. & Caruthers, M. H. *J. Am. chem. Soc.* **108,** 2040–2048 (1986).
13. Bain, J. D. *et al. Biochemistry* **30,** 5411–5421 (1991).
14. Fluck, M. M., Salser, W. & Epstein, R. H. *Molec gen. Genet.* **151,** 137–149 (1977).
15. Konecki, D. S., Aune, K. C., Tate, W. & Caskey, C. T. *J. biol. Chem.* **252,** 4514–4520 (1977).
16. Smrt, J., Kemper, W., Caskey, T. & Nirenberg, M. *J. biol. Chem.* **245,** 2753–2757 (1970).
17. Valle, R. P. C. & Morch, M.-D. *FEBS Lett.* **235,** 1–15 (1988).
18. Craigen, W. J. & Caskey, C. T. *Cell,* **50,** 1–2 (1987).
19. Tsuchihashi, Z. & Kornberg, A. *Proc. natn. Acad. Sci. U.S.A.* **87,** 2516–2520 (1990).
20. Jukes, T. J., Bessho, Y., Ohama, T. & Osawa, S. *Nature* **352,** 575 (1991).
21. Fersht, A. R. *Biochemistry* **26,** 8031–8037 (1987).
22. Kast, P. thesis ETH, Zurich (1991).
23. Roberts, J. W. & Carbon, J. *J. biol. Chem.* **250,** 5530–5541 (1975).
24. Lyttle, M. H., Wright, P. B., Sinha, N. D., Bain, J. D. & Chamberlin, A. R. *J. org. Chem.* **56,** 4608–4615 (1991).
25. Bain, J. D., Wacker, D. A., Kuo, E. E., Lyttle, M. H. & Chamberlin, A. R. *J. org. Chem.* **56,** 4615–4625 (1991).
26. Fersht, A. R. & Dingwall, C. *Biochemistry* **18,** 2627–2631 (1979).
27. Sepiol, J., Kazimierczuk, Z. & Shugar, D. *Z Naturf.* **31c,** 361–370 (1976).

ACKNOWLEDGEMENTS. We thank S. Hecht and J. Knowles for comments on the manuscript. This work was supported by the Swiss Nationalfond, Sandoz, the ETH and the NIH. C.S. was supported by a fellowship from the NSF.