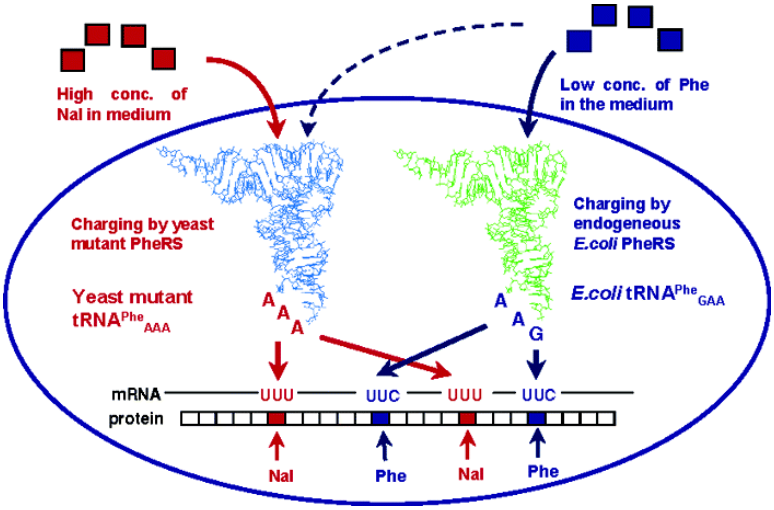
Communication

# Breaking the Degeneracy of the Genetic Code

Inchan Kwon, Kent Kirshenbaum, and David A. Tirrell

*J. Am. Chem. Soc.*, **2003**, 125 (25), 7512-7513• DOI: 10.1021/ja0350076 • Publication Date (Web): 29 May 2003






## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 4 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
COMMUNICATIONS

Published on Web 05/29/2003

# Breaking the Degeneracy of the Genetic Code

Inchan Kwon, Kent Kirshenbaum, and David A. Tirrell*

*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125*

Received March 5, 2003; E-mail: tirrell@caltech.edu

Organisms use a canonical set of 20 amino acids to generate the proteins that sustain the life of the cell. In recent years, several laboratories have pursued an expansion in the number of genetically encoded amino acids, by using either a nonsense suppressor or a frameshift suppressor tRNA to incorporate noncanonical amino acids into proteins in response to amber or four-base codons, respectively.[1] Such methods have worked well for single-site insertion of novel amino acids; however, their utility in multisite incorporation is limited by modest (20−60%) suppression efficiencies.[1e−g]

Efficient multisite incorporation has been accomplished by replacement of natural amino acids in auxotrophic *Escherichia coli* strains, and by using aminoacyl-tRNA synthetases with relaxed substrate specificity or attenuated editing activity.[2] Although this method provides efficient incorporation of analogues at multiple sites, it suffers from the limitation that the novel amino acid must "share" codons with one of the natural amino acids. We present here a potential solution to this coding problem.

The genetic code is degenerate, in that the protein biosynthetic machinery utilizes 61 mRNA sense codons to direct the templated polymerization of the 20 amino acid monomers.[3] Just two amino acids, methionine and tryptophan, are encoded by unique mRNA triplets. Reassignment of degenerate sense codons therefore offers the prospect of a substantially expanded genetic code and a correspondingly enriched set of building blocks for natural and artificial proteins.

As a test case for establishing the feasibility of breaking the degeneracy of the code, we chose the biosynthetic machinery responsible for incorporation of phenylalanine (Phe) into the proteins of *E. coli*. Phe is encoded by two codons, UUC and UUU. Both codons are read by a single tRNA, which is equipped with the anticodon sequence GAA. The UUC codon is therefore recognized through standard Watson−Crick base-pairing between codon and anticodon; UUU is read through a G-U wobble base-pair at the first position of the anticodon.[4] Thermal denaturation of RNA duplexes has yielded estimates of the Gibbs free energies of melting of G-U, G-C, A-U, and A-C base pairs as 4.1, 6.5, 6.3, and 2.6 kcal/mol, respectively, at 37 °C.[5] Thus the wobble base pair, G-U, is less stable than the Watson−Crick base pair, A-U. On this basis, we proposed that a mutant $tRNA^{Phe}$ outfitted with the AAA anticodon ($tRNA^{Phe}_{AAA}$) might be engineered to read UUU codons faster than wild-type $tRNA^{Phe}_{GAA}$. If $tRNA^{Phe}_{AAA}$ can then be charged selectively with an amino acid analogue, one should be able to accomplish codon-biased incorporation of the analogue at multiple sites in recombinant proteins. Although tRNAs bearing unmodified A in the first position of the anticodon are known to read codons ending with C or U,[6] the binding of *E. coli* $tRNA^{Phe}_{GAA}$ at UUC should dominate that of $tRNA^{Phe}_{AAA}$ owing to differences in the stability of A-C and G-C base pairs (see above).

The approach used here builds on the method introduced by Furter for site-specific insertion of amino acid analogues into proteins in vivo.[1c] The method involves introduction into *E. coli* of a heterologous aminoacyl-tRNA synthetase and its cognate tRNA. If cross-charging between the heterologous pair and the translational apparatus of the host is slow or absent and if the analogue is charged only by the heterologous synthetase, insertion of the analogue can be restricted (or at least biased) to sites characterized by the most productive base-pairing between the heterologous tRNA and the messenger RNA of interest.

To test these ideas, we prepared a yeast $tRNA^{Phe}$ ($ytRNA^{Phe}_{AAA}$) with an altered anticodon loop. The first base ($G^{34}$) of the $tRNA^{Phe}_{GAA}$ was replaced with A to provide specific Watson−Crick base-pairing to the UUU codon. Furthermore, $G^{37}$ in the extended anticodon site was replaced with A to increase translational efficiency.[1c] We believe that charging of $ytRNA^{Phe}_{AAA}$ by *E. coli* PheRS can be ignored, because the aminoacylation rate of $ytRNA^{Phe}_{AAA}$ by *E. coli* PheRS is known to be <0.1% of that of *E. coli* $tRNA^{Phe}_{GAA}$.[7]

Since wild-type yeast PheRS does not activate amino acids significantly larger than phenylalanine, a mutant form of the synthetase with relaxed substrate specificity was prepared to accommodate L-3-(2-naphthyl)alanine (Nal).[1j] On the basis of prior work from this laboratory,[2f] the mutant yeast PheRS (mu-yPheRS) was prepared by introduction of a Thr415Gly mutation in the α-subunit of the synthetase. The kinetics of activation of Nal and Phe by mu-yPheRS were analyzed in vitro via the pyrophosphate exchange assay. The specificity constant ($k_{cat}/K_M$) for activation of Nal by mu-yPheRS was found to be $1.55 \times 10^{-3}$ ($s^{-1}\,\mu M^{-1}$), 8-fold larger than that for Phe.[1j] Therefore, when the ratio of Nal to Phe in the culture medium is high, $ytRNA^{Phe}_{AAA}$ should be charged predominantly with Nal.

Murine dihydrofolate reductase (mDHFR), which contains nine Phe residues, was chosen as the test protein. The expression plasmid pQE16 encodes mDHFR under control of a bacteriophage T5 promoter; the protein is outfitted with a C-terminal hexahistidine tag to facilitate purification via immobilized metal affinity chromatography. In this construct, four of the Phe residues of mDHFR are encoded by UUC codons, five by UUU. A full-length copy of the mu-yPheRS gene, under control of a constitutive *tac* promoter, was inserted into pQE16. The gene encoding $ytRNA^{Phe}_{AAA}$ was inserted into the repressor plasmid pREP4 (Qiagen) under control of the constitutive promoter *lpp*. *E. coli* transformants harboring these two plasmids were incubated in Phe-depleted minimal medium supplemented with 3 mM Nal and were then treated with 1 mM IPTG to induce expression of mDHFR. Although the *E. coli* strain (K10-F6Δ) used in this study is a Phe auxotroph,[1c] a detectable level of mDHFR was expressed even under conditions of nominal depletion of Phe, probably because of release of Phe through turnover of cellular proteins. In negative control experiments, mDHFR was expressed in the absence of either $ytRNA^{Phe}_{AAA}$ or mu-yPheRS. The molar mass of mDHFR prepared in the absence of Nal, $ytRNA^{Phe}_{AAA}$, or mu-yPheRS was 23 287 Da, precisely that calculated for His-tagged mDHFR. However, when $ytRNA^{Phe}_{AAA}$ and mu-yPheRS were introduced into the expression strain and Nal

 10.1021/ja0350076 CCC: $25.00 

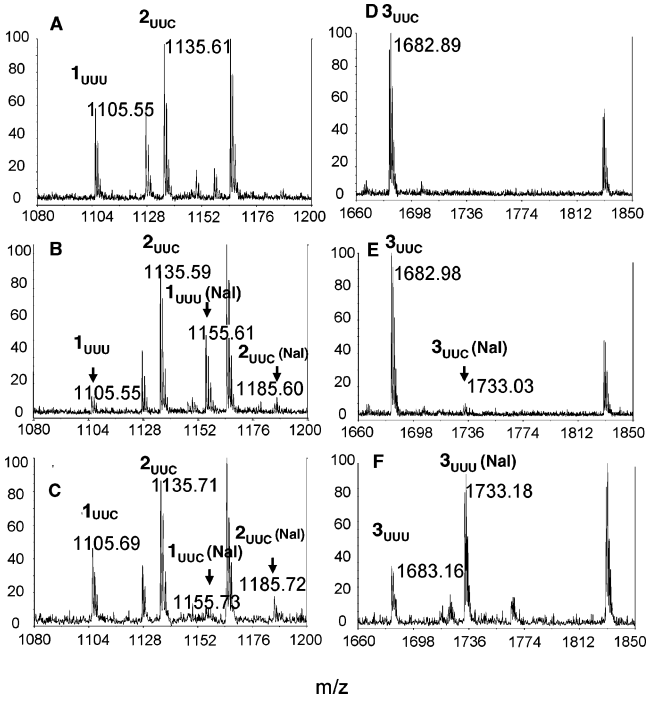



***Figure 1.*** Replacement of Phe by Nal can be detected in MALDI mass spectra of tryptic fragments of mDHFR samples prepared in media supplemented with Phe (A and D) or Nal (B, C, E, and F). Peptide $\mathbf{1}_{UUU}$ contains a Phe residue encoded by UUU, whereas in peptide $\mathbf{1}_{UUC}$ this codon has been mutated to UUC. Peptide $\mathbf{1}_{UUU}$**(Nal)** refers to the form of the peptide containing Nal in place of Phe. Peptides **2** and **3** are designated similarly. See text for details.

was added to the culture medium, the observed mass of mDHFR was 23 537 Da (yield 2.5 mg/L after Ni-affinity chromatography). Because each substitution of Nal for Phe leads to a mass increment of 50 Da, this result is consistent with replacement of five Phe residues by Nal. No detectable mass shift was found in the absence of either ytRNA$^{Phe}_{AAA}$ or mu-yPheRS, confirming that the intact heterologous pair is required for incorporation of Nal. For mDHFR isolated from the strain harboring the heterologous pair, amino acid analysis indicated replacement of 4.4 of the 9 Phe residues by Nal. Without ytRNA$^{Phe}_{AAA}$ or mu-yPheRS, no incorporation of Nal into mDHFR was detected by amino acid analysis.

Tryptic digests of mDHFR were analyzed to determine the occupancy of individual Phe sites. Digestion of mDHFR yields peptide fragments that are readily analyzed by MALDI mass spectrometry as shown in Figure 1. Peptide $\mathbf{1}_{UUU}$ (residues 184−191, YKFEVYEK) contains a Phe residue encoded as UUU, whereas peptides $\mathbf{2}_{UUC}$ (residues 62−70, KTWFSIPEK) and $\mathbf{3}_{UUC}$ (residues 26−39, NGDLP WPPLRNEFK) each contain a Phe residue encoded as UUC. In the absence of Nal, peptide $\mathbf{1}_{UUU}$ was detected with a monoisotopic mass of 1105.55 Da, in accord with its theoretical mass (Figure 1A). However, when Nal was added, a strong signal at a mass of 1155.61 Da was detected, and the 1105.55 was greatly reduced in intensity (Figure 1B). As described earlier, each substitution of Nal for Phe leads to a mass increase of 50.06 Da; the observed shift in mass is thus consistent with replacement of Phe by Nal in response to the UUU codon. Liquid chromatography − tandem mass spectrometry (LC/MS/MS) confirmed this assignment. The ratio of MALDI signal intensities, though not rigorously related to relative peptide concentrations, suggests that Nal incorporation is dominant at the UUU codon.

Similar analyses were conducted for peptides $\mathbf{2}_{UUC}$ and $\mathbf{3}_{UUC}$. In the absence of added Nal, the observed masses of peptides $\mathbf{2}_{UUC}$ and $\mathbf{3}_{UUC}$ are 1135.61 (Figure 1A) and 1682.89 Da (Figure 1D), respectively, as expected. Upon addition of Nal to the expression medium, the 1135.61 signal and 1682.89 signals were not substantially reduced, and only weak signals were observed at masses of 1185.60 and 1733.03 (Figure 1B and 1E), which would be expected for peptides $\mathbf{2}_{UUC}$ and $\mathbf{3}_{UUC}$ containing Nal. Nal incorporation thus appears to be rare at UUC codons under the conditions used here for protein expression.

There is at least a formal possibility that the observed codon-biased incorporation of Nal might be dependent on codon context rather than—or in addition to—codon identity. MALDI sampling errors are also possible. To test these possibilities, a mutant mDHFR gene was prepared by mutating the UUU codon in peptide $\mathbf{1}_{UUU}$ to UUC, and the UUC codon in peptide $\mathbf{3}_{UUC}$ to UUU. In the resulting peptide $\mathbf{1}_{UUC}$, the signal indicating incorporation of Nal was only slightly above background (Figure 1C), whereas for peptide $\mathbf{3}_{UUU}$, Nal is readily detected (Figure 1F). Nal incorporation is unambiguously codon-biased to UUU.

The results described here show conclusively that a heterologous pair comprising a genetically engineered tRNA and cognate aminoacyl-tRNA synthetase can be used to break the degeneracy of the genetic code in *E. coli*. Ongoing experiments address the quantitative selectivity and the generality of the approach demonstrated here.

**Acknowledgment.** This work was supported by NIH GM62523 and by the NSF MRSEC program. We thank Dr. Mona Shahgholi, Dr. Gary M. Hathaway, and Dr. Jie Zhou for mass spectrometry studies. We thank Pin Wang and Soojin Son for helpful discussions.

**Supporting Information Available:** Protein expression gels and LC/MS/MS spectra (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) (a) Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R.; Diala, E. S. *J. Am. Chem. Soc.* **1989**, *111*, 8013. (b) Noren, C. J.; Anthonycahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182. (c) Furter, R. *Protein Sci.* **1998**, *7*, 419. (d) Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 56. (e) Anderson, R. D.; Zhou, J.; Hecht, S. M. *J. Am. Chem. Soc.* **2002**, *124*, 9674. (f) Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537. (g) Hohsaka, T.; Ashizuka, Y.; Murakami, H.; Sisido, M. *Nucleic Acids Res.* **2001**, *29*, 3646. (h) Hohsaka, T.; Sato, K.; Sisido, M.; Takai, K.; Yokoyama, S. *FEBS Lett.* **1994**, *344*, 171. (i) Kowal, A. K.; Oliver, J. S. *Nucleic Acids Res.* **1997**, *25*, 4685. (j) Wang, P.; Tirrell, D. A. Unpublished data.

(2) (a) Wilson, M. J.; Hatfield, D. L. *Biochim. Biophys. Acta* **1984**, *781*, 205. (b) Kast, P.; Hennecke, H. *J. Mol. Biol.* **1991**, *222*, 99. (c) Ibba, M.; Kast, P.; Hennecke, H.; *Biochemistry* **1994**, *33*, 7107. (d) Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A. *FEBS Lett.* **2000**, *467*, 37. (e) Tang, Y.; Tirrell, D. A. *Biochemistry* **2002**, *41*, 10635. (f) Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652. (g) Doring, V.; Mootz, H. D.; Nangle, L. A.; Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P.; Marliere, P. *Science* **2001**, *292*, 501.

(3) Crick, F. H. C.; Barnett, L.; Brenner, S.; Watts-Tobin, R. J. *Nature* **1961**, *192*, 1227.

(4) (a) Crick, F. H. C. *J. Mol. Biol.* **1966**, *19*, 548. (b) Soll, D.; RajBhandary, *J. Mol. Biol.* **1967**, *29*, 113.

(5) Meroueh, M.; Chow, C. S. *Nucleic Acids Res.* **1999**, *27*, 1118.

(6) (a) Inagaki, Y.; Kojima, A.; Bessho, Y.; Hori, H.; Ohama, T.; Osawa, S. *J. Mol. Biol.* **1995**, *251*, 486. (b) Chen, P.; Qian, Q.; Zhang, S. P.; Isaksson, L. A.; Bjork, G. R. *J. Mol. Biol.* **2002**, *317*, 481. (c) Boren, T.; Elias, P.; Samuelsson, T.; Claesson, C.; Barciszewska, M.; Gehrke, C. W.; Kuo, K. C.; Lustig, F. *J. Mol. Biol.* **1993**, *230*, 739.

(7) Peterson, E. C. T.; Uhlenbeck, O. C. *Biochemistry* **1992**, *31*, 10380.

JA0350076