Chapter 5

Breaking the Degeneracy of the Genetic Code

Abstract

A mutant yeast phenylalanine transfer RNA (ytRNA$^{Phe}_{AAA}$) containing a modified (AAA) anticodon was generated to explore the feasibility of breaking the degeneracy of the genetic code in *Escherichia coli*. By using an *E.coli* strain co-transformed with ytRNA$^{Phe}_{AAA}$ and a mutant yeast phenylalanyl-tRNA synthetase, we demonstrate efficient codon-biased replacement of phenylalanine (Phe) by L-3-(2-naphthyl)alanine (2Nal), a non-proteinogenic analog. Site-specific incorporation of 2Nal in response to UUU codons was confirmed by mass spectrometric analysis of recombinant murine dihydrofolate reductase. These results illustrate a general method for increasing the number of distinct, genetically-encoded amino acids available for protein engineering and for exploration of the chemistry and physics of protein-like macromolecules.

Introduction

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 non-canonical amino acids into proteins in response to amber or four-base codons, respectively.[1-10] Such methods have worked well for single-site insertion of novel amino acids; however, their utility in multi-site incorporation is limited by modest (20-60%) suppression efficiencies.[1, 5, 11]

Efficient multi-site incorporation has been accomplished by replacement of natural amino acids in auxotrophic *E. coli* strains,[12-15] and by using aminoacyl-tRNA synthetases with relaxed substrate specificity or attenuated editing activity.[14, 16] Although this method

provides efficient incorporation of analogs 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.

Material and Methods

Materials. All chemical reagents were of analytical grade, obtained from commercial suppliers, and used without further purification unless otherwise noted. L-3-(2-naphthyl)alanine and other natural amino acids were obtained from Sigma (St. Louis, MO).

Strains and Plasmids. An *E. coli* strain XL1-Blue (Stratagene) was used for plasmid propagation and isolation. A Phe auxotrophic strain K10-F6Δ (K10, Hfr(Cavalli) *pheS13rel*-1 *tonA*22 thi $T2^R$ *pheA*18)[4] was a gift from Rolf Furter (University of Massachusetts). Plasmids carrying modified yeast $tRNA^{Phe}$ variants were derived from pRO117[4] in which yeast $tRNA^{Phe}$ expression cassette was inserted at the *Sna*I site of the pREP4 (Qiagen). Plasmids for expression of mDHFR and overexpression of mutant yeast phenylalanyl-tRNA synthetases (yPheRS) were derived from pRO148[4] in which yPheRS genes under constitutive *tac* promoter control was inserted at the *PvuII* site of pQE16 (Qiagen).

Construction of Plasmid Carrying Mutant Yeast $tRNA^{Phe}$. pREP4_ytRNA^{Phe}_AAA containing a mutant yeast $tRNA^{Phe}$ with a modified anticodon (AAA) ($ytRNA^{Phe}_{AAA}$) was generated by PCR mutagenesis kit (Stratagene) using pRO117[4] as a template.

Construction of Plasmid Carrying mDHFR and Mutant Yeast Phenylalanyl-tRNA Synthetase. An intact mDHFR expression cassette was obtained by endonuclease restriction of pQE16 at the *Aat*II and *Nhe*I sites. This expression cassette was inserted between the *Aat*II and *Nhe*I sites of pRO148 to generate pQE16_mDHFR_yPheRS. A mutation of threonine to glycine at the 415th position of α-subunit of yPheRS was performed by PCR mutagenesis kit (Stratagene) using pQE16_mDHFR_yPheRS as a template to generate pQE16_mDHFR_yPheRS (T415G).

Protein Expression and Purification. Cultures of K10-F6Δ outfitted with pREP4_ytRNA_AAA and pQE16_mDHFR_yPheRS (T415G) were grown in M9 minimal medium supplemented with 0.4% (w/v) of glucose, 0.1 mM of $CaCl_2$, 1.0 mM of $MgSO_4$, 35 μg/mL of thiamine, 20 amino acids (25 mg/L), 100 μg/mL of ampicillin, and 35 μg/mL of kanamycin. When an optical density of the culture reached 0.8 to 1.0 at 600nm (OD_{600}), the cultures were centrifuged for 7 min (6,000 rpm) at 4 °C. The cell pellets were washed twice with 0.9% (w/v) of NaCl solution. The cells were resuspended in M9 minimal medium supplemented with 18 amino acids (25 mg/L) and indicated amounts of 2Nal and Phe. After 10 min incubation, 1 mM of isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to induce mDHFR protein expression. The OD_{600} of the culture was measured 4 hrs after induction. Then the cells were harvested by centrifugation for 10 min (6,000 rpm) at 4°C and stored at -70°C. After thawing, mDHFR proteins were purified under denaturing conditions according to manufacturer's

protocol (Qiagen). Protein expression was evaluated by SDS-PAGE with coomassie blue staining. Loading of the gel was normalized for cell densities as determined by OD_{600}.

Amino Acid Analysis. The purified mDHFR solutions were concentrated 10-fold by ultrafiltration (Millipore) followed by a buffer exchange against 0.1% trifluoroacetic acid (TFA) solution. Samples were submitted to the Molecular Structure Facility at the University of California, Davis, for amino acid analysis on a Beckman 6300 instrument (Fullerton, CA).

Matrix-Assisted Laser Desorption Ionization Mass Spectrometry (MALDI–MS) Analysis. The purified mDHFR solutions were desalted by $ZipTip_{C18}$ (Millipore) and eluted with 3 μL of 50% CH_3CN/0.1% TFA. 1 μL was used for matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) analysis with sinapinic acid (30% (v/v) of acetonitrile and 70% (v/v) of 0.1% TFA solution) as the matrix. The analysis was performed on a PerSeptive Biosystems (Framingham, Massachusetts) Voyager DE PRO MALDI-TOF mass spectrometer operating in linear and positive ion modes. For MALDI-MASS analysis of tryptic digests of purified mDHFR, 10 μL of the concentrated protein solution was added to 90 μL of 75 mM ammonium bicarbonate solution. 2 μL of 0.2 g/L of modified trypsin (Promega) was added, and the solution was incubated at 37°C for 2.0 hrs. 12 μl of 5% TFA solution was added to quench the reaction. Chromatography on $Ziptip_{C18}$ columns (Millipore) provided purified peptide samples (1 μL), which were added to 2,5-dihydroxybenzoic acid (DHB) MALDI matrix solution for MALDI-MASS analysis.

Results and Discussion

The genetic code is degenerate, in that the protein biosynthetic machinery utilizes 61 mRNA sense codons to direct the templated polymerization of a set of 20 amino acid monomers.[17] Just two amino acids, methionine and tryptophan, are encoded by unique mRNA triplets. Re-assignment 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 ytRNA, 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.[18] 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^{o}C.[19] 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 ytRNA^{Phe} outfitted with the AAA anticodon (ytRNA$^{Phe}_{AAA}$) might be engineered to read UUU codons faster than wild-type ytRNA^{Phe}. If ytRNA$^{Phe}_{AAA}$ can then be charged selectively with an amino acid analog, one should be able to accomplish codon-biased incorporation of the analog at multiple sites in recombinant proteins. With respect to reading of UUC, an unmodified A in the first position of the anticodon is known to read codons ending with C, as well as U, in the absence of ytRNAs containing G in the

first anticodon position.[20-22] However, the binding of *E. coli* $tRNA^{Phe}_{GAA}$ should dominate the binding of $ytRNA^{Phe}_{AAA}$ owing to differences in the stability of A-C and G-C base pairs (see above).

The approach used here is a modification of the method introduced by Furter for site-specific insertion of amino acid analogs *in vivo*.[4] 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 analog is charged only by the heterologous synthetase, insertion of the analog can be restricted (or at least biased) to sites characterized by productive base-pairing between the heterologous tRNA and the messenger RNA of interest.

In order 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 $ytRNA^{Phe}$ anticodon (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 in order to increase translational efficiency.[10, 23] We believe that charging of yeast $tRNA^{Phe}_{AAA}$ by *E. coli* PheRS can be ignored, because the aminoacylation rate of yeast $tRNA^{Phe}_{AAA}$ by *E. coli* PheRS is known to be 0.1% of that of *E.coli* $tRNA^{Phe}_{GAA}$.[24]

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 in order to accommodate L-3-(2-naphthyl)alanine (2Nal).[10] On the basis of prior work from this laboratory,[16] the mutant yeast PheRS (yPheRS (T415G)) was prepared by introduction of a Thr415Gly mutation in the α-subunit of the synthetase. The kinetics of activation of 2Nal and Phe by yPheRS (T415G) was analyzed *in vitro* via the

pyrophosphate exchange assay. The specificity constant (k_{cat}/K_M) for activation of 2Nal by yPheRS (T415G) was found to be $1.55x10^{-3}$ ($S^{-1}\mu M^{-1}$), 8-fold larger than that for Phe.[10] Therefore, when the ratio of 2Nal to Phe in the culture medium is high, $ytRNA^{Phe}_{AAA}$ should be charged predominantly with 2Nal. Recently, we have shown that 2Nal can be incorporated with better than 95% efficiency via amber suppression in an *E. coli* strain co-transformed with yPheRS (T415G) and $ytRNA^{Phe}_{CUA}$.

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 yPheRS (T415G) 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 2Nal, and 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,[4] a detectable level of mDHFR was expressed even under conditions of nominal depletion of Phe (Figure 2), probably due to release of Phe through turnover of cellular proteins. In negative control experiments, mDHFR was expressed in the absence of either $ytRNA^{Phe}_{AAA}$ or yPheRS (T415G). mDFHR expression levels in these experiments were similar, indicating that neither $ytRNA^{Phe}_{AAA}$ nor yPheRS (T415G) significantly reduces the protein synthesis rate (Figure 2). However, MALDI-MS spectra

and amino acid analyses of purified mDHFRs showed differences among samples prepared under these conditions (Table 1). The molar mass of mDHFR prepared in the absence of 2Nal, ytRNA$^{Phe}_{AAA}$, or yPheRS (T415G) was 23,287 daltons, precisely that calculated for His-tagged mDHFR. However, when ytRNA$^{Phe}_{AAA}$ and yPheRS (T415G) were introduced into the expression strain and 2Nal was added to the culture medium, the observed mass of mDHFR was 23,537 daltons. Because each substitution of 2Nal for Phe leads to a mass increment of 50 daltons, this result is consistent with replacement of five Phe residues by 2Nal. No detectable mass shift was found in the absence of either ytRNA$^{Phe}_{AAA}$ or yPheRS (T415G), confirming that the intact heterologous pair is required for incorporation of 2Nal. For mDHFR isolated from the strain harboring the heterologous pair, amino acid analysis indicated replacement of 4.4 of the 9 Phe residues by 2Nal (Table 1). Without ytRNA$^{Phe}_{AAA}$ or yPheRS (T415G), no incorporation of 2Nal into mDHFR was detected by amino acid analysis.

Because neither MALDI nor amino acid analysis of intact mDHFR shows which Phe residues have been replaced by 2Nal, tryptic digests were analyzed to determine the occupancy of individual Phe sites. Tryptic digestion of mDHFR yields peptide fragments that are readily analyzed by MALDI mass spectrometry as shown in Figure 3. Peptide 1_{UUU} (residues 184-191, YKFEVYEK) contains a Phe residue encoded as UUU, whereas Peptide 2_{UUC} (residues 62-70, KTWFSIPEK) and Peptide 3_{UUC} (residues 26-39, NGDLPWPPLRNEFK) each contain a Phe residue encoded by UUC. In the absence of 2Nal, Peptide 1_{UUU} was detected with a monoisotopic mass of 1105.55 daltons, in accord with its theoretical mass (Figure 3A). However, when 2Nal was added, a strong signal at a mass of 1155.61 daltons was detected, and the 1105.55 signal was greatly reduced in

intensity (Figure 3B). As described earlier, each substitution of 2Nal for Phe leads to a mass increase of 50.06 daltons; the observed shift in the experimental mass is thus consistent with replacement of Phe by 2Nal in response to the UUU codon. Liquid chromatography-tandem mass spectrometry of Peptide 1_{UUU} (Nal) was also carried out in order to determine more directly the origin of the observed increase in mass. The fragment ion masses could be unambiguously assigned as shown in Figure 4, confirming replacement of Phe by 2Nal. The ratio of MALDI signal intensities, though not rigorously related to relative peptide concentrations, suggests that 2Nal incorporation is dominant at the UUU codon.

Similar analyses were conducted for Peptide 2_{UUC} and Peptide 3_{UUC}. In the absence of added 2Nal, the observed masses of Peptides 2_{UUC} and 3_{UUC} are 1135.61 (Figure 3A) and 1682.89 daltons (Figure 3D), respectively. These observed masses match well the corresponding theoretical masses (1135.61 daltons for Peptide 2_{UUC}, and 1682.86 daltons for Peptide 3_{UUC}). Upon addition of 2Nal to the expression medium, the signals at these masses (Figure 3B and 3E) were not substantially reduced, and only very weak signals were observed at masses of 1185.60 and at 1733.03, which would be expected for peptides containing 2Nal in place of Phe. 2Nal incorporation thus appears to be rare at UUC codons under the conditions used here for protein expression. Other peptides containing encoded Phe show similar codon-biased selective incorporation of 2Nal and Phe.

There is at least a formal possibility that the observed codon-biased incorporation of 2Nal might be dependent on codon context rather than - or in addition to - codon identity. In order to test this possibility, mutant mDHFR genes were prepared by

mutating the UUU codon in Peptide 1_{UUU} to UUC, and the UUC codon in Peptide 3_{UUC} to UUU. The resulting peptides were designated Peptide 1_{UUC} and Peptide 3_{UUU}, respectively. In Peptide 1_{UUC}, 2Nal incorporation was greatly reduced (Figure 3C), whereas for Peptide $3_{UUU,}$ 2Nal is readily detected (Figure 3F). 2Nal incorporation is unambiguously codon-biased to UUU.

Conclusions

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*. This method should provide a general strategy for multi-site incorporation of non-canonical amino acids, without the requirement that one of the natural amino acids be excluded. Introduction of more than one set of orthogonal pairs should allow several types of non-canonical amino acids to be incorporated in site-specific fashion. Ongoing experiments address the quantitative selectivity and the generality of the approach demonstrated here.

Acknowledgments

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 Dr. Pin Wang and Dr. Soojin Son for helpful discussions.

References

1. Anderson, R. D.; Zhou, J.; Hecht, S. M. *J. Am. Chem. Soc.* **2002**, *124*, 9674-9675.

2. Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537-539.

3. Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 9026-9027.

4. Furter, R. *Protein Sci.* **1998**, *7*, 419-426.

5. Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M. *Biochemistry* **2001**, *40*, 11060-11064.

6. Magliery, T. J.; Anderson, J. C.; Schultz, P. G. *J. Mol. Biol.* **2001**, *307*, 755-769.

7. Noren, C. J.; Anthonycahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182-188.

8. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498-500.

9. Wang, L.; Brock, A.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 1836-1837.

10. Wang, P., Ph.D. thesis, California Institute of Technology, Pasadena, 2003.

11. Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R.; Diala, E. S. *J. Am. Chem. Soc.* **1989**, *111*, 8013-8014.

12. Kast, P.; Hennecke, H. *J. Mol. Biol.* **1991**, *222*, 99-124.

13. Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A. *FEBS Lett.* **2000**, *467*, 37-40.

14. Tang, Y.; Tirrell, D. A. *Biochemistry* **2002**, *41*, 10635-10645.

15. Wilson, M. J.; Hatfield, D. L. *Biochim. Biophys. Acta* **1984**, *781*, 205-215.

16. Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652-5653.

17. Crick, F. H.; Brenner, S.; Watstobi, R. J.; Barnett, L. *Nature* **1961**, *192*, 1227-1232.

18. Crick, F. H. C. *J. Mol. Biol.* **1966**, *19*, 548-555.

19. Meroueh, M.; Chow, C. S. *Nucleic Acids Res.* **1999**, *27*, 1118-1125.

20. Boren, T.; Elias, P.; Samuelsson, T.; Claesson, C.; Barciszewska, M.; Gehrke, C. W.; Kuo, K. C.; Lustig, F. *J. Mol. Biol.* **1993**, *230*, 739-749.

21. Chen, P.; Qian, Q.; Zhang, S. P.; Isaksson, L. A.; Bjork, G. R. *J. Mol. Biol.* **2002**, *317*, 481-492.

22. Inagaki, Y.; Kojima, A.; Bessho, Y.; Hori, H.; Ohama, T.; Osawa, S. *J. Mol. Biol.* **1995**, *251*, 486-492.

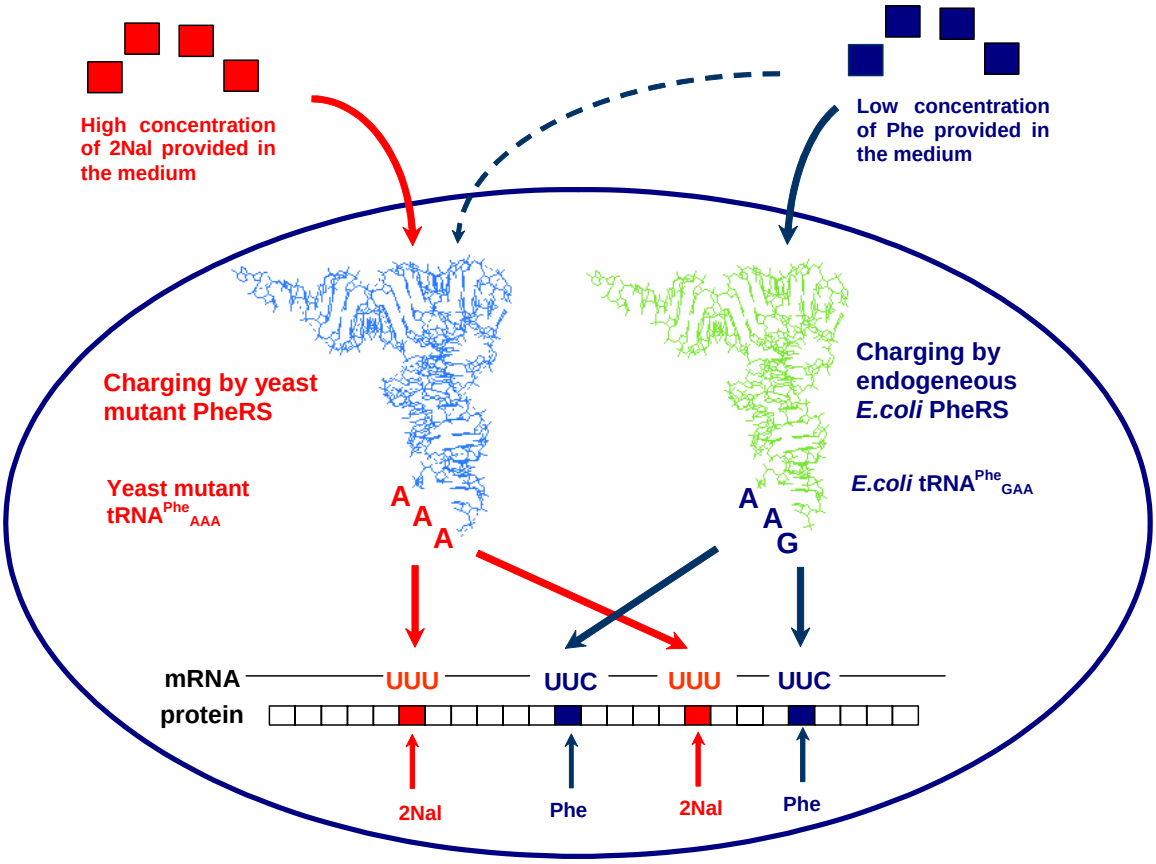
23. Yarus, M. *Science* **1982**, *218*, 646-652.

24. Peterson, E. T.; Uhlenbeck, O. C. *Biochemistry* **1992**, *31*, 10380-10389.

Table 1: Molar masses and numbers of 2Nal residues observed for mDHFR samples prepared under various conditions

$ytRNA^{Phe}_{AAA}$	+	+	-	+
yPheRS (T415G)	+	-	+	+
2Nal (3mM)	-	+	+	+
Phe (3mM)	+	-	-	-
Mass of intact mDHFR	23287	23287	23287	23537
Number of 2Nal residues	ND[a]	ND	ND	4.4

a. Not detected



Phenylalanine auxotrophic *E. coli* strain

Figure 1: A strategy for multi-site incorporation of 2Nal into recombinant proteins by breaking the degeneracy of the phenylalanine codons. ytRNA images were obtained from the Protein Data Bank (www.rcsb.org/pdb/).

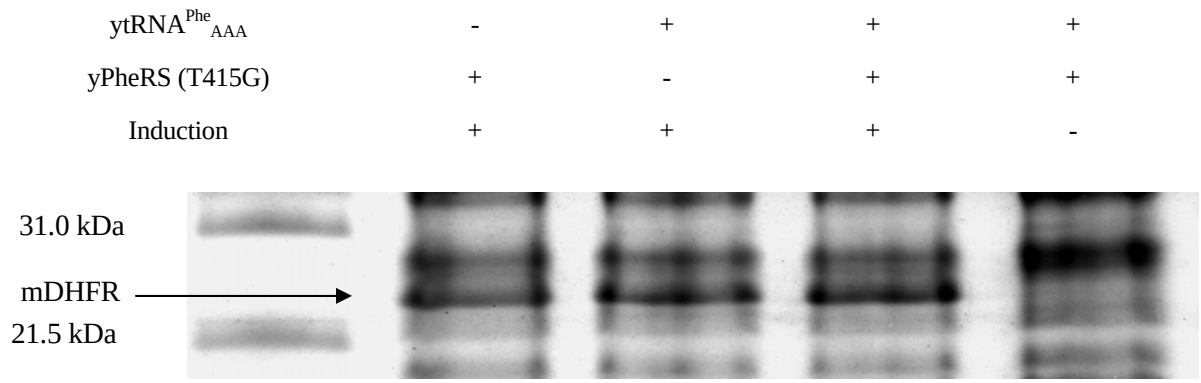


Figure 2: SDS-PAGE analysis of mDHFR prepared in minimal media supplemented with 3 mM 2Nal and free of exogenous Phe. Conditions are noted at the top of each lane. Lane 1 shows molecular weight standards.

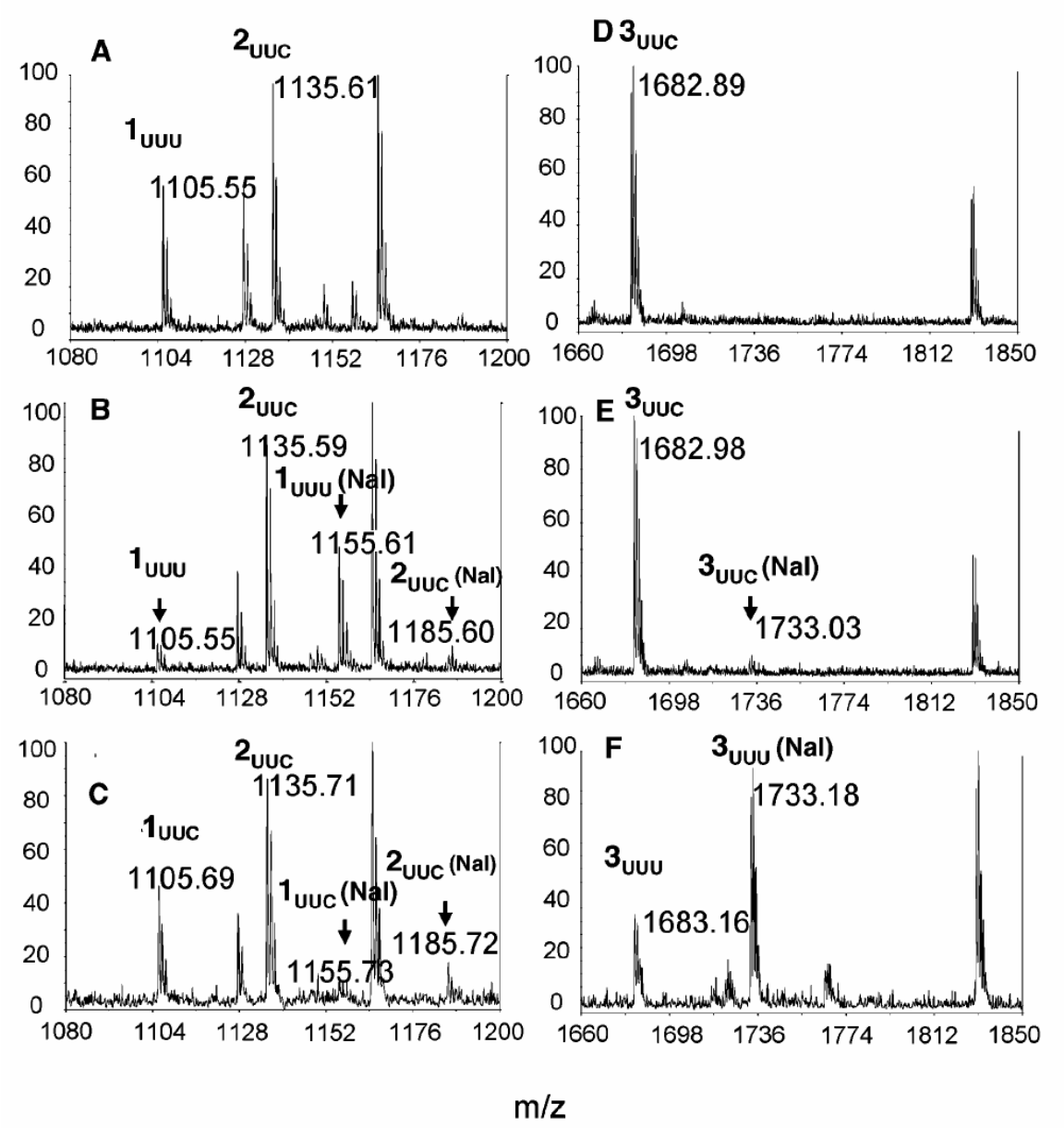

A
2UUC
1135.61
1UUU
1105.55
D 3UUC
1682.89
B
2UUC
1135.59
1UUU (NaI)
1155.61
1UUU
2UUC (NaI)
1105.55
1185.60
E 3UUC
1682.98
3UUC (NaI)
1733.03
C
2UUC
1135.71
1UUC
1105.69
2UUC (NaI)
1UUC (NaI)
1185.72
1155.73
F
3UUU (NaI)
1733.18
3UUU
1683.16
100
80
60
40
20
0
1080
1104
1128
1152
1176
1200
1660
1698
1736
1774
1812
1850
m/z

Figure 3: Replacement of Phe by 2Nal can be detected in MALDI-MS spectra of tryptic fragments of mDHFR. Peptide 1_{UUU} (residues 184-191, YKFEVYEK) contains a Phe (F) residue encoded by UUU, whereas in Peptide 1_{UUC} this codon has been mutated to UUC. Peptides 2 (residues 62-70, KTWFSIPEK) and 3 (residues 26-39, NGDLPWPPLRNEFK) are designated similarly. Peptide 1_{UUU} (Nal) refers to the form of the peptide containing 2Nal in place of Phe. In the absence of 2Nal, Peptide1_{UUU} was detected at a mass of 1105.55 (A). Upon addition of 2Nal, a strong signal corresponding to Peptide 1_{UUU}(Nal) was detected at m/z = 1155.61, and the signal for Peptide 1_{UUU} was greatly reduced (B). For Peptide 1_{UUC}, 2Nal incorporation was much less efficient (C). Signals corresponding to Peptides 2_{UUC} (B) and 3_{UUC} (E) were not substantially reduced upon addition of 2Nal, and only very weak signals for Peptides 2_{UUC} (Nal) and 3_{UUC} (Nal) were detected (B, E). When the UUC codon in Peptide 3_{UUC} is mutated to UUU, a strong signal for Peptide 3_{UUU} (Nal) is detected (F). These data confirm that incorporation of 2Nal is strongly biased to UUU codons.

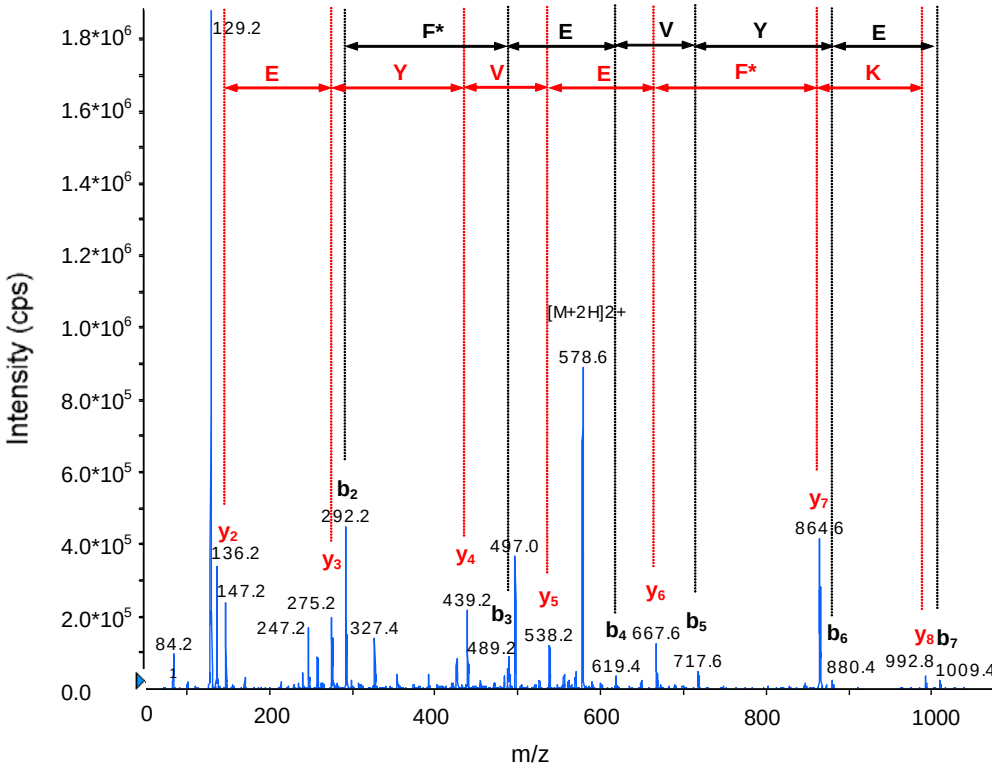


Figure 4: Tandem mass spectrum of Peptide 1_{UUU} (Nal) YKF*EVYEK. The doubly-charged ion at 578.6 daltons was selected and fragmented. The sequence of the peptide containing 2Nal (F*) can be read from the annotated b (black) or y (red) ion series.