Chapter 1

Expansion of the Number of the Genetically Encoded Amino Acids in *E. coli*

Abstract

Nonnatural amino acid incorporation has been one of the most important protein engineering platforms. In particular, site-specific incorporation of nonnatural amino acids allows design of proteins containing a nonnatural amino acid with minimal perturbation of native properties. Until now more than 30 nonnatural amino acids have been introduced into proteins in *E. coli* hosts outfitted with orthogonal pairs of cognate tRNAs and aminoacyl-tRNA synthetases (aaRS). Substrate specificity of aaRS has been altered to recognize various nonnatural amino acids through either rational design of the active site of aaRS or high-throughput screening of aaRS library. In this approach, site-specificity has been achieved by assigning a nonnatural amino acid to a stop codon, a degenerate codon, or a frameshift codon for efficient discrimination of the programmed sites from unwanted sites.

Introduction

Genetically encoded nonnatural amino acids have been used to endow recombinant proteins with novel chemical, physical, or biological properties. Several distinct methods have been developed to introduce nonnatural amino acids into recombinant proteins at programmed sites in vivo. Residue-specific incorporation involves the global replacement of a particular natural amino acid with a nonnatural amino acid in a target protein using auxotrophic hosts.[1-14] The strength of this technique lies in efficient protein translation and multi-site incorporation of a nonnatural amino acid, because sense codons are reassigned for a nonnatural amino acid. Residue-specific incorporation of 5’,5’,5’-trifluoroleucine (TFL) into a leucine-zipper protein enhanced its thermal stability.[15] A reactive

phenylalanine analog, *para*-ethynylphenylalanine, has been used for selective dye-labeling of newly synthesized protein in *E. coli*.[16] Newly synthesized proteins in mammalin cells have been selectively indentified using bioorthogonal noncanonical amino acid tagging.[17] Trp analogs containing selenophene and thienyl functional groups have been used in the X-ray crystallography of proteins.[18,19] Aminotryptophan was used to design protein-based pH sensors[3] and greatly changed the spectral properties of fluorescent proteins.[1] Regardless of many successful applications of residue-specific incorporation, its further application to the design of biologically-active macromolecules could be restricted by two limitations. First, reassignment of sense codons for a nonnatural amino acid usually leads to exclusion of one natural amino acid, which is encoded by the sense codons, in protein translation. Second, native properties of a target protein may be impaired by incorporation of a nonnatural amino acid at non-permissive sites of a target protein,[20,21] though the Tirrell laboratory recently showed that the impaired properties may be rescued through directed evolution of the target protein.[21]

Site-Specific Incorporation

Site-specific incorporation involves introduction of a nonnatural amino acid into a target protein at any position selected. Access to all 20 natural amino acids and site-specificity of the method may allow design of proteins containing a nonnatural amino acid with minimal perturbation of native properties. Site-specific incorporation requires outfitting the cell with an "orthogonal pair" comprising a suppressor tRNA and a cognate aminoacyl-tRNA synthetase (aaRS) that operate independently of the endogenous synthetase-tRNA pairs in *E. coli*. The site-specific incorporation of a nonnatural amino acid into recombinant proteins via this strategy was reported by Furter in 1998[22] and the Schultz

laboratory has developed powerful selection methods to identify heterologous synthetases and tRNAs.[23-26] A general strategy for site-specific incorporation consists of four steps and each step will be discussed in more detail in the following sections.

Choice of a Nonnatural Amino Acid

First of all, it is necessary to choose a nonnatural amino acid suitable for goals. Until now more than 30 nonnatural amino acids have been successfully introduced to proteins in a site-specific manner[27,28] (Figure 1) including reactive nonnatural amino acids, such as *para*-acetylphenylalanine **1**, *meta*-acetylphenylalanine **2**, *para*-(3-oxobutanoyl)-L-phenylalanine **3**, *para*-(2-amino-3-hydroxylenthynyl)phenylalanine **4**, *para*-ehthynylthiocarbonyl-phenylalanine **5**, *para*-propargyloxyphenylalanine **6**, *para*-azidophenylalanine **7**, *para*-ethynylphenylalanine **32**, *para*-iodophenylalanine **20**, and *para*-bromophenylalanine **31**. These reactive nonnatural amino acids can be used to modify proteins through bio-orthogonal chemical transformations. For example, the keto and β-diketo functional groups (**1**-**3**) can selectively react with both hydrazides and alkoxyamines.[29-31] Azide functional groups (**7**) can be ligated to alkynes through copper-catalyzed [3+2] cycloaddition.[16,32,33] Aryl halides (**20** and **31**) can be conjugated to terminal alkynes or alkenes via palladium-catalyzed coupling reactions.[34-39] Nonnatural amino acids containing photoactive side chains including *para*-azidophenylalanine **7**,[40,41] *para*-benzoylphenylalanine **9**,[42,43] *O*-(2-nitrobenzyl)tyrosine **12**,[44] and *S*-(2-nitrobenzyl)cysteine **10**[45] have been inserted into proteins.

Genetically-encoded nonnatural amino acids can be used as probes of various kinds. First, nonnatural amino acids containing heavy atoms (I and Se) have been used for X-ray

crystallographic studies of protein structure.[18,46] Second, fluorinated phenylalanine and ^{15}N-labelled methoxy phenylalanine have been used to label proteins for NMR study.[22,47] Third, *para*-L-cyanophenylalanine **22** containing nitrile group has been introduced to proteins as an infrared probe to investigate changes in local environment.[48] Fourth, introduction of a fluorophore into proteins at defined sites can greatly facilitate study of protein function and structure. For example, the two fluorescent amino acids, 2-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamide)propanoic acid (dansylalanine) **18** and L-(7-hydroxycoumarin-4-yl)ethylglycine **17,** have been introduced to proteins to monitor unfolding of proteins.[49,50]

Orthogonal Pairs

Activation of an amino acid followed by charging the activated amino acid into a cognate tRNA by a cognate aminoacyl-tRNA synthetase (aaRS) is a key step to ensure high fidelity in a protein translation process. One orthogonal pair of aaRS/tRNA is assigned to each natural amino acid. Therefore, a new orthogonal aaRS/tRNA pair is necessary for a nonnatural amino acid to be utilized in protein translation machinery. Therefore, the development of orthogonal pairs to each living organism is one of the key steps in achieving site-specific incorporation of a nonnatural amino acid into proteins in vivo.

Modification of Endogenous Orthogonal Pairs. Initially Schultz and colleagues tried to modify an *E. coli* endogenous RNA/aaRS pair to obtain a new orthogonal pair. Three sites in $tRNA^{Gln}_2$ sequence, which are critical for binding with *E. coli* glutamyl-tRNA synthetase (eGlnRS), were mutated to reduce affinity to eGlnRS. Then eGlnRS was evolved to

recognize the mutant $tRNA^{Gln}_2$ and retain its orthogonality with the endogenous tRNAs in *E. coli*. However, the approach was unsuccessful due to poor discrimination of the mutant $tRNA^{Gln}_2$ against wild-type $tRNA^{Gln}_2$ by even the best eGlnRS mutant.[51]

Orthogonal Pairs Derived from Yeast. A second approach to generate an orthogonal pair was the use of a heterologous tRNA/aaRS pair obtained from another organism in *E. coli*. The yeast phenylalanyl-tRNA/yeast phenylalanyl-tRNA synthetase pair was known to be orthogonal to the *E. coli* endogenous system. In vitro aminoacylation studies indicated that there was little cross-talk between *E. coli* and yeast phenylalanyl-tRNA/synthetase pairs.[52,53] Aminoacylation activity of yeast $tRNA^{Phe}$ by *E. coli* PheRS is 10-fold lower than that of *E. coli* $tRNA^{Phe}$.[22] Aminoacylation activity of *E. coli* $tRNA^{Phe}$ by yeast PheRS is 24- or 34-fold lower than that of yeast $tRNA^{Phe}$,[54,55] because U20 in *E. coli* $tRNA^{Phe}$ replaces G20, one of the identity elements of yeast $tRNA^{Phe}$ recognized by yPheRS (Figure 2).[54] A yeast amber suppressor derived from $tRNA^{Phe}_{GAA}$ ($ytRNA^{Phe}_{CUA\_wt}$) has been widely used to study suppression of amber codons, though *E. coli* lysyl-tRNA synthetase (eLysRS) shows a weak activity to aminoacylate $ytRNA^{Phe}_{CUA\_wt}$ with Lys.[22] A wild-type tRNA sequence can be engineered to eliminate the cross-reactivity with *E.coli* synthetase but preserve efficient recognition by the cognate aaRS. When $ytRNA^{Phe}_{CUA\_wt}$ was co-expressed with wild-type yPheRS in an *E. coli* host, 60% of the amber codon was occupied by Lys.[22] The mischarging by Lys was likely due to the presence of U36 and A73 in the $ytRNA^{Phe}_{CUA\_wt}$, which are the main identity elements for eLysRS.[56] However, when G37 was replaced by A37 and $ytRNA^{Phe}_{CUA}$ was co-expressed with wild-type yPheRS, only 5% of the amber codon was occupied by Lys. The remaining 5% of Lys misincorporation was

eliminated by rational design of ytRNA sequence. Disruption of a Watson-Crick base pair between nucleotides 30 and 40 in ytRNA$^{Phe}_{CUA}$ reduced mischarging by the eLysRS. As an alternative orthogonal tRNA/synthetase pair, an amber suppressor derived from yeast tyrosyl-tRNA (ytRNA^{Tyr}) and yeast tyrosyl-tRNA synthetase (yTyrRS) pair was examined.[57,58] RajBhandary and colleagues generated yTyrRS variants using error-prone PCR to enhance discrimination of the suppressor tRNA over *E. coli* proline tRNA by factors of 2.2- to 6.8-fold. Nishikawa and colleagues optimized the sequence of yeast amber suppressor tRNA^{Tyr} to minimize Lys charging by eLysRS.[59] This optimized yeast tRNA^{Tyr}, together with yeast TyrRS, could be used as an orthogonal pair for incorporation of nonnatural amino acids into proteins. Schultz and colleagues developed two orthogonal pairs derived from yeast. An amber suppressor derived from yeast glutamyl-tRNAs (ytRNA$^{Gln}_{2_CUA}$) is not a substrate for any *E. coli* endogenous synthetase. In turn, yeast glutamyl-tRNA synthetase (yGlnRS) aminoacylated ytRNA$^{Gln}_{2}$ in *E. coli* but did not charge *E. coli* tRNA^{Gln}.[60] An amber suppressor derived from yeast aspartidyl-tRNA (ytRNA$^{Asp}_{CUA}$) is not recognized by *E. coli* synthetases. Introduction of E188K mutation in yeast aspartidyl-tRNA synthetase (yAspRS) led to reduction in aminoacylation activity with *E. coli* tRNA^{Asp}.[61] Therefore, both ytRNA$^{Gln}_{2_CUA}$/yGlnRS pair and ytRNA$^{Asp}_{CUA}$/yAspRS (E188K) form orthogonal pairs in *E. coli* suitable for the incorporation of nonnatural amino acids into proteins in vivo.

Orthogonal Pairs Derived from Archaebacteria. Schultz and colleagues developed an orthogonal tyrosyl-tRNA/synthetase pair obtained from archaea, since archeal aminoacyl-tRNA synthetases are more similar to eukaryotic than prokaryotic counterparts. The amber

suppressor tyrosyl-tRNA (*Mj*tRNA^{Tyr})/synthetase (*Mj*TyrRS) pair derived from *Methanococcus jannaschii* was selected and evolved to be orthogonal to the *E. coli* endogenous system.[25,62] The amber suppressor *Mj*tRNA$^{Tyr}_{CUA}$ was efficiently aminoacylated by *Mj*TyrRS in *E. coli*. However, *E. coli* endogenous synthetases showed weak activity towards *Mj*tRNA$^{Tyr}_{CUA}$. Therefore, a selection method to obtain orthogonal tRNAs in *E. coli* has been developed to enhance orthogonality with respect to *E. coli* endogenous synthetases and retain recognition by the cognate synthetase and the protein translational machinery (Figure 3).[25] This method includes positive and negative selection of a mutant tRNA library. In the negative selection, amber codons were introduced into a toxic barnase protein at permissive sites that allow replacement of amino acids. Eleven nucleotides of *Mj*tRNA$^{Tyr}_{CUA}$ that do not bind to *Mj*TyrRS were saturated with all four bases to generate a tRNA library. When the tRNA library was expressed in minimal medium without *Mj*TyrRS, cells containing tRNA variants aminoacylated by *E. coli* endogenous synthetases resulted in cell death. The plasmids were isolated from the cells survived, and co-transformed with a plasmid harboring both a *Mj*TyrRS gene and a β-lactamase gene containing an amber codon at a permissive site into *E. coli* hosts. When cells were cultured in liquid medium in the presence of ampicillin, cells harboring tRNA variants aminoacylated by *Mj*TyrRS survived. After the negative and positive screening of tRNA library, the tRNAs were orthogonal to the *E. coli* endogenous synthetases and retained their affinity to the cognate synthetase. Schultz and colleagues developed another orthogonal pair obtained from Archea. An amber suppressor leucyl-tRNA and a leucyl-tRNA synthetase were derived from *Halobacterium sp. NRS-1* and *Methanobacterium thermoautotrophicum*, respectively. Aminoacylation activity of the suppressor leucyl-

tRNA by the leucyl-tRNA synthetase and its orthogonality with *E. coli* synthetases were significantly enhanced by extensive mutations in the anticodon loop and acceptor stem.[23]

Substrate Specificity Change

Another issue to achieve incorporation of a nonnatural amino acid into protein in vivo is alteration of substrate specificity of an aminoacyl-tRNA synthetase toward the nonnatural amino acid. Aminoacyl-tRNA synthetases ensure the high fidelity of transforming genetic code sequences into biologically functional proteins through a two-step aminoacylation reaction.[63-67] In the first step, the cognate amino acid is activated by aaRS in the presence of ATP to form the amino acid adenylate. Subsequently aaRS catalyzes the esterification reaction to join the amino acid to the 2'- or 3'-OH of the terminal ribonucleotide of its cognate tRNA. Manipulation of such reactions could potentially alter the genetic code to allow incorporation of novel amino acid into proteins in vivo.[68] Some nonnatural amino acids including TFL, azidohomoalanine, and *para*-fluorophenylalanine, are recognized by wild-type cognate aminoacyl-tRNA synthetases.[5,15,22] However, the majority of nonnatural amino acids are very poor substrates for natural aminoacyl-tRNA synthetases. Engineering the new synthetase activity is almost inevitable when one wants to incorporate an amino acid that is not recognized by wild-type enzymes.

Rational Design of the Active Sites of Aminoacyl-tRNA Synthetases. Alteration of substrate specificity of aminoacyl-tRNA synthetases has been achieved in several labs by the rational design of the binding pocket to accommodate nonnatural amino acid. Tirrell

and colleagues have been interested in exploring the possibility of incorporating substituted phenylalanines, such as *p*-bromophenylalanine, *p*-iodophenylalanine, and *p*-azidophenylalanine. On the basis of the crystal structure of the PheRS (tPheRS) from *Thermus thermophilus* (Figure 4a and b), Safro and colleagues proposed that Val 261 and Ala 314 in the amino acid binding pocket of α-subunit of the tPheRS are critical in the discrimination of Phe from its amino acid competitors. Sequence alignment indicates that Ala 314 in α-subunit of the tPheRS corresponds to Ala 294 in α-subunit of the ePheRS (Figure 4c), and the Hennecke group showed that the substrate specificity of the ePheRS can be relaxed by a mutation at Ala 294. The A294G mutant was shown to enable incorporation of *para*-chlorophenylalanine into recombinant proteins.[69] A subsequent computational simulation, consistent with the Safro prediction, identified two cavity-forming mutations (T251G and A294G) in ePheRS binding pocket. These two mutations led to relaxed substrate specificity and efficient in vivo replacement of Phe by *para*-acetylphenylalanine (pAcF).[4] Sequence alignment showed that Thr 415 in yPheRS is equivalent to Thr 251 in ePheRS. Therefore, a T415G mutation was introduced into the yPheRS to enlarge the active site. The yPheRS (T415G) variant showed activity for *p*-bromophenylalanine, *p*-iodophenylalanine, and *p*-azidophenylalanine in ATP-PPi exchange assays in vitro (see Chapters 2 and 4). There was an interesting finding that the yPheRS (T415G) efficiently activated Trp and 2-naphthylalanine as well as Phe analogs. Considering the spectrum of nonnatural amino acids recognized by yPheRS (T415G), Tirrell and colleagues explored whether the yPheRS (T415G) can activate even Trp analogs. In vitro activation studies showed that the yPheRS (T415G) efficiently activated four Trp analogs, 6-chlorotryptophan (6ClW), 6-bromotryptophan (6BrW), and 5-

bromotrytophan (5BrW), which were not utilized by the endogenous *E. coli* translational system (see Chapter 3). A phenylalanine auxotrophic *E. coli* strain transformed with this yPheRS (T415G) and cognate suppressor tRNA enabled the assignment of an amber nonsense codon to Phe analogs and Trp analogs in vivo (see Chapters 2-4). The cavity formed due to the T415G mutation in the binding pocket of yPheRS led to the activation of Trp as well as nonnatural amino acids, which resulted in misincorporation of Trp at the programmed sites. Therefore, the binding site of yPheRS was re-designed to enhance specificity for pBrF. Specifically, Tirrell and colleagues used the T415A variant, which exhibits 5-fold higher activity towards pBrF as compared to Trp in ATP-PPi exchange assays. Use of the yPheRS (T415A) eliminated misincorporation of Trp at programmed sites in proteins.[70]

Nishikawa and colleagues have changed substrate specificity of yeast tyrosyl-tRNA synthetase (yTyrRS) to accommodate tyrosine analogs.[71] Based on the crystal structure of TyrRS obtained from *Bacillus stearothermophilus* TyrRS and sequence alignment of two homologous TyrRSs, active site residues in yTyrRS have been mutated to generate a series of yTyrRS variants. Among the yTyrRS variants, one containing the Tyr43Gly mutation (yTyrRS (Y43G)) was found to show activity for several 3-substituted tyrosine analogs in aminoacylation assays in vitro. yTyrRS (Y43G) efficiently charged ytRNA^{Tyr} with 3-iodo-tyrosine, which was not utilized by wild-type yTyrRS. The yTyrRS (Y43G) showed 400-fold lower activity towards tyrosine than wild-type yTyrRS. Similarly, Yokoyama and colleagues[72] screened *E. coli* TyrRS variants recognizing 3-iodotyrosine via in vitro biochemical assays. The *E. coli* TyrRS variant that has two mutations (Y37V and Q195C) in the active site activated 3-iodotyrosine 10-fold more efficiently than tyrosine.[72]

Combined use of *E. coli* TyrRS (V37C195) and a suppressor tRNA derived from *Bacillus stearothermophilus* allowed incorporation of 3-iodotyrosine into proteins in mammalian cells.[73]

Directed Evolution of Aminoacyl-tRNA Synthetase Substrate Specificity. Aminoacyl-tRNA synthetases are known to be readily evolvable. Schultz and colleagues have developed powerful selection methods to change the substrate specificity of tyrosyl-tRNA synthetase (*Mj*TyrRS) derived from *Methanococcus jannaschii* toward nonnatural amino acids.[23,24,26,74] The selection method consists of a series of positive and negative selections.[68] Based on the crystal structure of the homologous TyrRS from *Bacillus stearothermophilus*, five residues in the active site of *Mj*TyrRS were selected and randomized to generate a *Mj*TyrRS mutant library. In the positive selection, an amber codon was inserted at permissive sites in a chloramphenicol acetyltransferase (CAT) gene. The library was cultured in medium containing chloramphenicol in the presence of a nonnatural amino acid. Only the cells containing *Mj*TyrRS variants allowing suppression of an amber codon in CAT gene survived. Surviving cells were subjected to the negative selection. In the negative selection, the cells were regrown in medium containing chloramphenicol in the absence of the nonnatural amino acid. Those cells that could not survive were isolated from the replica plates containing the nonnatural amino acid. Using this method, substrate specificity of *Mj*TyrRS was changed to selectively recognize *O*-methyl-*L*-tyrosine.[68]

Alternative selection methods for evolving substrate specificity of *Mj*TyrRS towards nonnatural amino acids have been developed. One of them involves a positive

selection similar to the previous one, but a different negative selection.[60] The negative selection is based on suppression of amber codons inserted at permissive sites in the barnase gene that is lethal to *E. coli*. When the cells survived from the positive selection were cultured in the absence of nonnatural amino acids, the cells containing *Mj*TyrRS variants recognizing natural amino acids resulted in cell death due to expression of the barnase protein. Another method for evolving *Mj*TyrRS specificity to nonnatural amino acids involves an amplifiable fluorescent reporter, a green fluorescent protein.[24] Amber codons were introduced at various sites in a T7 RNA polymerase gene that controlled expression of a green fluorescent protein. In the positive screening, the *Mj*TyrRS library cells were cultured in the presence of a nonnatural amino acid. The cells containing *Mj*TyrRS variants that recognized the nonnatural amino acid and suppressed the amber codons in T7 RNA polymerase gene became fluorescent and were collected with a fluorescence-activated cell sorter (FACS). In the negative screening, the cells collected from the positive screening were cultured in the absence of the nonnatural amino acid; the cells containing *Mj*TyrRS variants that did not recognize any natural amino acid were not fluorescent and collected with a FACS. Therefore, the cells collected from the positive and negative screenings contained the *Mj*TyrRS variants that can selectively charge the nonnatural amino acid into the cognate tRNA. One advantage of the latter method is that both reporter genes are encoded in a single plasmid, which eliminates the need for DNA isolation after each selection step. However, the method requires a FACS for high-throughput screening.

Most attempts to evolve aminoacyl-tRNA synthetase substrate specificity have been limited to prokaryotic or archaeal aminoacyl-tRNA synthetases. Recently, Tirrell and

colleagues have evolved a eukaryotic aminoacyl-tRNA synthetase, yPheRS, to change its substrate specificity towards a hydrophobic phenylalanine analog, 2-naphthylalanine (see Chapter 6). The yPheRS library was constructed by randomizing four residues (N412, T415, S418, and S437) that contact the Phe substrate inside the binding pocket (Figure 5a). The screening consisted of a series of positive and negative screenings (Figure 5b), and involved expression of a GFP reporting folding status. The active synthetases would allow incorporation of either 2Nal or other natural amino acids at non-permissive sites of a GFP, which resulted in poorly-folded GFP showing weak fluorescence. Therefore, in the positive screening, the library cells were cultured in the presence of 2Nal and then the cells containing active synthetases were collected by high-throughput screening with a FACS. The yPheRS variants with very low activity toward natural amino acids will not misincorporate any natural amino acid at Phe sites in GFP and so GFP would retain full intensity of fluorescence. In the negative screening, the cells obtained from the positive screening were cultured in the absence of 2Nal and then the cells containing selective synthetases were collected with a FACS.

New Codon-Anticodon Interactions

Among the sixty-four codons, sixty-one (sense codons) are assigned to one of the twenty natural amino acids. The other three codons (nonsense codons) are stop codons that terminate protein translation. In order to incorporate a nonnatural amino acid into proteins in vivo, we need to either re-assign an existing codon or generate a new codon for the nonnatural amino acid. Until now three distinct approaches have been explored, re-

assignment of a stop codon, reassignment of a degenerate codon, and assignment of a frameshift codon.

Nonsense Suppression. In 1989 incorporation of a nonnatural amino acid into proteins in vitro was demonstrated by two independent groups, the Chamberlin group[75] and the Schultz group.[76] Nonsense suppression means that one of the stop codons (nonsense codons) can be read by a suppressor charged with an amino acid. Otherwise protein translation will be terminated at the stop codon. Three stop codons, amber (UAG),[46,68,70,74,77] ochre (UAA),[23,78,79] and opal (UGA) stop codons,[23,78] have been reassigned for nonnatural amino acids. The amber codon has been used most widely,[22,40,42,43,48,70,74,76,80-83] because it is the least common stop codon in *E. coli*, and because several naturally occurring suppressor tRNAs recognize it efficiently.[84,85] Use of the amber codon to encode nonnatural amino acids requires outfitting the cell with an "orthogonal pair" comprising a suppressor tRNA and a cognate aminoacyl-tRNA synthetase (aaRS) that operate independently of the endogenous synthetase-tRNA pairs in *E. coli* as described in previous sections. To date, more than thirty nonnatural amino acids have been incorporated into proteins in response to the amber stop codon with high fidelity.[27,28,70]

Breaking the Degeneracy of the Genetic Code. Suppression of the amber stop codon has worked well for single-site insertion of novel amino acids. However, the utility of this approach is limited by moderate suppression efficiency. Competition between elongation and termination produces mixtures of full-length and truncated chains, and protein yields

are reduced accordingly. Efficient multi-site incorporation has been accomplished by replacement of natural amino acids in auxotrophic *E. coli* strains,[1,6,9-11,86,87] and by using aminoacyl-tRNA synthetases with relaxed substrate specificity[4] or attenuated editing activity.[11] Although this method provides efficient incorporation of analogs at multiple sites, it suffers from the limitation that one of the natural amino acids must be excluded from the engineered protein. Efficient multiple-site-specific incorporation can be achieved by re-assignment of a sense codon, if the sense codon can be efficiently distinguished from other sense codons. Tirrell and colleagues investigated re-assignment of a degenerate codon for a nonnatural amino acid. As a test case for establishing the feasibility of breaking the degeneracy of the code, phenylalanine codons were investigated (see Chapter 5).

Although 2Nal incorporaton was biased to UUU codons in the first study, misincorporation of 2Nal at unwanted sites (UUC codons) could not be avoided due to ambiguity in codon recognition by the AAA anticodon of ytRNA$^{Phe}_{AAA}$. Misincorporation of a nonnatural amino acid at unwanted sites might cause perturbation or loss of function of a target protein.[20,21] Based on experimental findings[88-95] and stereo-chemical modeling[96-99] of codon-anticodon interactions, an expanded wobble rule was proposed by Lim and Curran.[90] The expanded wobble rule indicates that A in the first position of the anticodon can recognize all four bases in the third position of the codon, although its affinity to C is lower than that to U. The observed incomplete discrimination of UUU from UUC codons by the AAA anticodon of ytRNA^{Phe} motivated us to explore a different set of degenerate codons. Several considerations recommend the UUG codon. First, Leu is encoded by six codons: UUA, UUG, CUA, CUG, CUU, and CUC. Discrimination of UUG from CUN (N = A/U/G/C) codons might be highly efficient due to discrimination at the first position in

the codon. Second, our existing yeast orthogonal pair should be readily adapted to the incorporation of Phe analogs in response to UUG codons. Third, according to the expanded wobble rules, C in the first position of the anticodon can recognize only G in the third position of the codon (see Chapter 6).

Frameshift Suppression. The utility of stop or degenerate codons might be restricted by limited number of codons. In order to overcome this limitation, extended codons have been investigated. Although three-base codons are universal for most living organisms, frameshift suppression of a four-base codon occurs is a regulatory mechanism in viruses and bacteria, and is mediated either by slipping of normal tRNAs at the ribosome or by natural frameshift suppressor tRNAs with four-base anticodons. Sisido and colleagues have pioneered frameshift suppression for incorporation of nonnatural amino acids into proteins in *E. coli*.[100-103] In this approach, four-base codons were inserted into a target protein at programmed sites. When the four-base codons are read by a modified tRNA that is charged with nonnatural amino acid and contains the corresponding anticodon, the full length target protein can be synthesized. If the four-base codons are read as three-base codons, truncated proteins will be formed due to early termination of protein translation. Four-base codons were usually derived from rarely used codons to minimize reading as three-base codons by the endogenous tRNAs. The translation efficiencies of some four-base codons were higher than those of three-base stop codons. In particular, a GGGU codon showed 86% translation efficiency,[104] while the suppression efficiencies of an amber codon were ca. 20-60%.[75,105] A number of four-base codons; CGGG, GGGU, CGGU, AGGU, CCCU, and CUCU have been generated.[101-103,106-108] The four-base codon strategy has also been applied to

incorporate two different nonnatural amino acids into two programmed sites in a protein, which showed orthogonality of each four-base codon to other four-base codons as well as three-base codons.[103,104,109] The frameshift suppression method has been extended to five-base codon to incorporate nonnatural amino acids.[110] Recently, Sisido and colleagues have demonstrated frameshift suppression of a four-base codon (UAGN) in mammalian cells using tRNA containing NCUA anticodon.[108]

Schultz and colleagues developed a combinatorial approach to select efficient suppressors of four-base codons; AGGA, UAGA, CCCU, and CUAG.[23,111] This approach has been extended to five- and six-base codons with tRNAs containing 6-10 nucleotides in the anticodon loops.[112] They identified suppressors of several five-base codons, such as AGGAU, CUACU, and CUAGU. However, translation efficiency of five-base codons was ca. 10%, which was lower than that of four-base codons. Suppression of six-base codons was too weak to be detected.[112] Recently, Schultz and colleagues have shown that a four-base codon can be used to incorporate a nonnatural amino acid into proteins in *E. coli* outfitted with a heterologous tRNA/synthetase orthogonal pair. In particular, L-homoglutamine was charged into archeal $tRNA^{Lys}$ by the evolved archeal lysyl-tRNA synthetase, and then was inserted into proteins in response to an AGGA codon in *E. coli*.[113]

Conclusions

In 1989 the concept of site-specific incorporation in vitro was independently demonstrated by the Chamberlin group and the Schultz group. Recently, several groups achieved site-specific incorporation of nonnatural amino acids into proteins in vivo using *E. coli* hosts outfitted with a twenty-first orthogonal pair of tRNA/aaRS. Schultz and

colleagues altered the substrate specificity of TyrRS, which was derived from *M. jannaschii*, to various nonnatural amino acids *via* high-throughput screening of a *Mj*TyrRS library. Tirrell and colleagues altered yPheRS substrate specificity toward Phe analogs, which led to site-specific incorporation of Phe analogs in *E. coli*. The moderate suppression efficiency of amber codons has motivated exploration of alternative codons that encode nonnatural amino acids. Breaking the degeneracy of the genetic code or frameshift suppression greatly expands the theoretical number of codons that can be assigned to nonnatural amino acids.

References

1. Bae, J. H.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M. H. J.; Azim, M. K.; Kim, J. S.; Zumbusch, A.; Holak, T. A.; Moroder, L.; Huber, R.; Budisa, N. *J. Mol. Biol.* **2003**, *328*, 1071-1081.
2. Budisa, N.; Minks, C.; Medrano, F. J.; Lutz, J.; Huber, R.; Moroder, L. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 455-459.
3. Budisa, N.; Rubini, M.; Bae, J. H.; Weyher, E.; Wenger, W.; Golbik, R.; Huber, R.; Moroder, L. *Angew. Chem. Int. Ed.* **2002**, *41*, 4066-4069.
4. Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652-5653.
5. Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 19-24.
6. Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem* **2002**, *3*, 235-237.

7. Link, A. J.; Vink, M. K. S.; Agard, N. J.; Prescher, J. A.; Bertozzi, C. R.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10180-10185.

8. Pratt, E. A.; Ho, C. *Fed. Proc.* **1974**, *33*, 1463-1463.

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

10. Tang, Y.; Tirrell, D. A. *J. Am. Chem. Soc.* **2001**, *123*, 11089-11090.

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

12. Wang, P.; Tang, Y.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 6900-6906.

13. Link, A. J.; Mock, M. L.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2003**, *14*, 603-609.

14. Budisa, N. *Angew. Chem. Int. Ed.* **2004**, *43*, 6426-6463.

15. Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A. *Angew. Chem.Int. Ed.* **2001**, *40*, 1494-1496.

16. Beatty, K. E.; Xie, F.; Wang, Q.; Tirrell, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 14150-14151.

17. Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9482-9487.

18. Bae, J. H.; Alefelder, S.; Kaiser, J. T.; Friedrich, R.; Moroder, L.; Huber, R.; Budisa, N. *J. Mol. Biol.* **2001**, *309*, 925-936.

19. Budisa, N.; Pal, P. P.; Alefelder, S.; Birle, P.; Krywcun, T.; Rubini, M.; Wenger, W.; Bae, J. H.; Steiner, T. *Biol. Chem.* **2004**, *385*, 191-202.

20. Budisa, N.; Pal, P. P. *Biol. Chem.* **2004**, *385*, 893-904.

21. Montclare, J. K.; Tirrell, D. A. *Angew. Chem. Int. Ed.* **2006**, *45*, 4518-4521.

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

23. Anderson, J. C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 9598-9608.

24. Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nat. Biotechnol.* **2002**, *20*, 1044-1048.

25. Wang, L.; Schultz, P. G. *Chem. Biol.* **2001**, *8*, 883-890.

26. Wang, L.; Schultz, P. G. *Chem. Commun.* **2002**, 1-11.

27. Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225-249.

28. Xie, J. M.; Schultz, P. G. *Curr. Opin. Chem. Biol.* **2005**, *9*, 548-554.

29. Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 56-61.

30. Zhang, Z. W.; Smith, B. A. C.; Wang, L.; Brock, A.; Cho, C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 6735-6746.

31. Zeng, H. Q.; Me, J. M.; Schultz, P. G. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5356-5359.

32. Link, A. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 11164-11165.

33. Link, A. J.; Vink, M. K. S.; Tirrell, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 10598-10602.

34. Kodama, K.; Fukuzawa, S.; Nakayama, H.; Kigawa, T.; Sakamoto, K.; Yabuki, T.; Matsuda, N.; Shirouzu, M.; Takio, K.; Tachibana, K.; Yokoyama, S. *ChemBioChem* **2006**, *7*, 134-139.

35. Nguefack, J. F.; Bolitt, V.; Sinou, D. *Tetrahedron Lett.* **1996**, *37*, 5527-5530.

36. Sakakibara, Y.; Ido, Y.; Sasaki, K.; Sakai, M.; Uchino, N. *Bull. Chem. Soc. Jpn.* **1993**, *66*, 2776-2778.

37. Schoenbe, A.; Heck, R. F. *J. Org. Chem.* **1974**, *39*, 3327-3331.

38. Yoneyama, M.; Kakimoto, M.; Imai, Y. *Macromolecules* **1988**, *21*, 1908-1911.

39. Yoneyama, M.; Konishi, T.; Kakimoto, M. A.; Imai, Y. *Makromol. Chem., Rapid Commun.* **1990**, *11*, 381-386.

40. 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.

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

42. Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020-11024.

43. Chin, J. W.; Schultz, P. G. *ChemBioChem* **2002**, *3*, 1135-1137.

44. Deiters, A.; Groff, D.; Ryu, Y. H.; Xie, J. M.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2006**, *45*, 2728-2731.

45. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126*, 14306-14307.

46. Xie, J. M.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G. *Nat. Biotechnol.* **2004**, *22*, 1297-1301.

47. Deiters, A.; Geierstanger, B. H.; Schultz, P. G. *ChemBioChem* **2005**, *6*, 55-58.

48. Schultz, K. C.; Supekova, L.; Ryu, Y. H.; Xie, J. M.; Perera, R.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 13984-13985.

49. Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9785-9789.

50. Wang, J. Y.; Xie, J. M.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 8738-8739.

51. Liu, D. R.; Magliery, T. J.; Pasternak, M.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10092-10097.

52. Doctor, B. P.; Mudd, J. A. *J. Biol. Chem.* **1963**, *238*, 3677-3681.

53. Kwok, Y.; Wong, J. T. F. *Can. J. Biochem.* **1980**, *58*, 213-218.

54. Sampson, J. R.; Direnzo, A. B.; Behlen, L. S.; Uhlenbeck, O. C. *Science* **1989**, *243*, 1363-1366.

55. Roe, B.; Sirover, M.; Dudock, B. *Biochemistry* **1973**, *12*, 4146-4154.

56. Tamura, K.; Himeno, H.; Asahara, H.; Hasegawa, T.; Shimizu, M. *Nucleic Acids Res.* **1992**, *20*, 2335-2339.

57. Kowal, A. K.; Kohrer, C.; RajBhandary, U. L. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2268-2273.

58. Ohno, S.; Yokogawa, T.; Fujii, I.; Asahara, H.; Inokuchi, H.; Nishikawa, K. *J. Biochem. (Tokyo).* **1998**, *124*, 1065-1068.

59. Fukunaga, J.; Yokogawa, T.; Ohno, S.; Nishikawa, K. *J. Biochem. (Tokyo).* **2006**, *139*, 689-696.

60. Liu, D. R.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 4780-4785.

61. Pastrnak, M.; Magliery, T. J.; Schultz, P. G. *Helv. Chim. Acta* **2000**, *83*, 2277-2286.

62. Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 5010-5011.

63. Francklyn, C.; Perona, J. J.; Puetz, J.; Hou, Y. M. *RNA* **2002**, *8*, 1363-1372.

64. de Pouplana, L. R.; Schimmel, P. *Trends Biochem. Sci.* **2001**, *26*, 591-596.

65. Ibba, M.; Soll, D. *Annu. Rev. Biochem.* **2000**, *69*, 617-650.

66. Cusack, S. *Curr. Opin. Struct. Biol.* **1997**, *7*, 881-889.

67. Schimmel, P. *Annu. Rev. Biochem.* **1987**, *56*, 125-158.

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

69. Ibba, M.; Kast, P.; Hennecke, H. *Biochemistry* **1994**, *33*, 7107-7112.

70. Kwon, I.; Wang, P.; Tirrell, D. A. *J. Am. Chem. Soc.* **2006**, *128*, 11778-11783.

71. Ohno, S.; Yokogawa, T.; Nishikawa, K. *J. Biochem. (Tokyo).* **2001**, *130*, 417-423.

72. Kiga, D.; Sakamoto, K.; Kodama, K.; Kigawa, T.; Matsuda, T.; Yabuki, T.; Shirouzu, M.; Harada, Y.; Nakayama, H.; Takio, K.; Hasegawa, Y.; Endo, Y.; Hirao, I.; Yokoyama, S. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 9715-9720.

73. Sakamoto, K.; Hayashi, A.; Sakamoto, A.; Kiga, D.; Nakayama, H.; Soma, A.; Kobayashi, T.; Kitabatake, M.; Takio, K.; Saito, K.; Shirouzu, M.; Hirao, I.; Yokoyama, S. *Nucleic Acids Res.* **2002**, *30*, 4692-4699.

74. Bose, M.; Groff, D.; Xie, J. M.; Brustad, E.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 388-389.

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

76. Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182-188.

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

78. Kohrer, C.; Xie, L.; Kellerer, S.; Varshney, U.; RajBhandary, U. L. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 14310-14315.

79. Kohrer, C.; Yoo, J. H.; Bennett, M.; Schaack, J.; RajBhandary, U. L. *Chem. Biol.* **2003**, *10*, 1095-1102.

80. Alfonta, L.; Zhang, Z. W.; Uryu, S.; Loo, J. A.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 14662-14663.

81. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z. W.; Schultz, P. G. *Science* **2003**, *301*, 964-967.

82. Cornish, V. W.; Benson, D. R.; Altenbach, C. A.; Hideg, K.; Hubbell, W. L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 2910-2914.

83. Mehl, R. A.; Anderson, J. C.; Santoro, S. W.; Wang, L.; Martin, A. B.; King, D. S.; Horn, D. M.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 935-939.

84. Thorbjarnardottir, S.; Uemura, H.; Dingermann, T.; Rafnar, T.; Thorsteinsdottir, S.; Soll, D.; Eggertsson, G. *J. Bacteriol.* **1985**, *161*, 207-211.

85. Yoshimura, M.; Inokuchi, H.; Ozeki, H. *J. Mol. Biol.* **1984**, *177*, 627-644.

86. Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A. *Angew. Chem. Int. Ed.* **2000**, *39*, 2148-2152.

87. Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Lett.* **2001**, *505*, 465-465.

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

89. Watanabe, Y.; Tsurui, H.; Ueda, T.; Furushima, R.; Takamiya, S.; Kita, K.; Nishikawa, K.; Watanabe, K. *Biochim. Biophys. Acta - Gene Struct. Express.* **1997**, *1350*, 119-122.

90. Lim, V. I.; Curran, J. F. *RNA* **2001**, *7*, 942-957.

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

92. Curran, J. F. *Nucleic Acids Res.* **1995**, *23*, 683-688.

93. 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.

94. Osawa, S.; Jukes, T. H.; Watanabe, K.; Muto, A. *Microbiol. Rev.* **1992**, *56*, 229-264.

95. Munz, P.; Leupold, U.; Agris, P.; Kohli, J. *Nature* **1981**, *294*, 187-188.

96. Lim, V. I.; Aglyamova, G. V. *Biol. Chem.* **1998**, *379*, 773-781.

97. Lim, V. I. *J. Mol. Biol.* **1997**, *266*, 877-890.

98. Lim, V. I. *J. Mol. Biol.* **1995**, *252*, 277-282.

99. Lim, V. I.; Venclovas, C. *FEBS Lett.* **1992**, *313*, 133-137.

100. Murakami, H.; Hohsaka, T.; Ashizuka, Y.; Hashimoto, K.; Sisido, M. *Biomacromolecules* **2000**, *1*, 118-125.

101. Murakami, H.; Hohsaka, T.; Ashizuka, Y.; Sisido, M. *J. Am. Chem. Soc.* **1998**, *120*, 7520-7529.

102. Muranaka, N.; Hohsaka, T.; Sisido, M. *FEBS Lett.* **2002**, *510*, 10-12.

103. Taki, M.; Hohsaka, T.; Murakami, H.; Taira, K.; Sisido, M. *J. Am. Chem. Soc.* **2002**, *124*, 14586-14590.

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

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

106. Sisido, M.; Ninomiya, K.; Ohtsuki, T.; Hohsaka, T. *Methods* **2005**, *36*, 270-278.

107. Taki, M.; Hohsaka, T.; Murakami, H.; Taira, K.; Sisido, M. *FEBS Lett.* **2001**, *507*, 35-38.

108. Taki, M.; Matsushita, J.; Sisido, M. *ChemBioChem* **2006**, *7*, 425-428.

109. Hohsaka, T.; Kajihara, D.; Ashizuka, Y.; Murakami, H.; Sisido, M. *J. Am. Chem. Soc.* **1999**, *121*, 34-40.

110. Hohsaka, T.; Ashizuka, Y.; Murakami, H.; Sisido, M. *Nucleic Acids Res.* **2001**, *29*, 3646-3651.

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

112. Anderson, J. C.; Magliery, T. J.; Schultz, P. G. *Chem. Biol.* **2002**, *9*, 237-244.

113. Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7566-7571.

114. Dudock, B. S.; Diperi, C.; Michael, M. S. *J. Biol. Chem.* **1970**, *245*, 2465-2468.

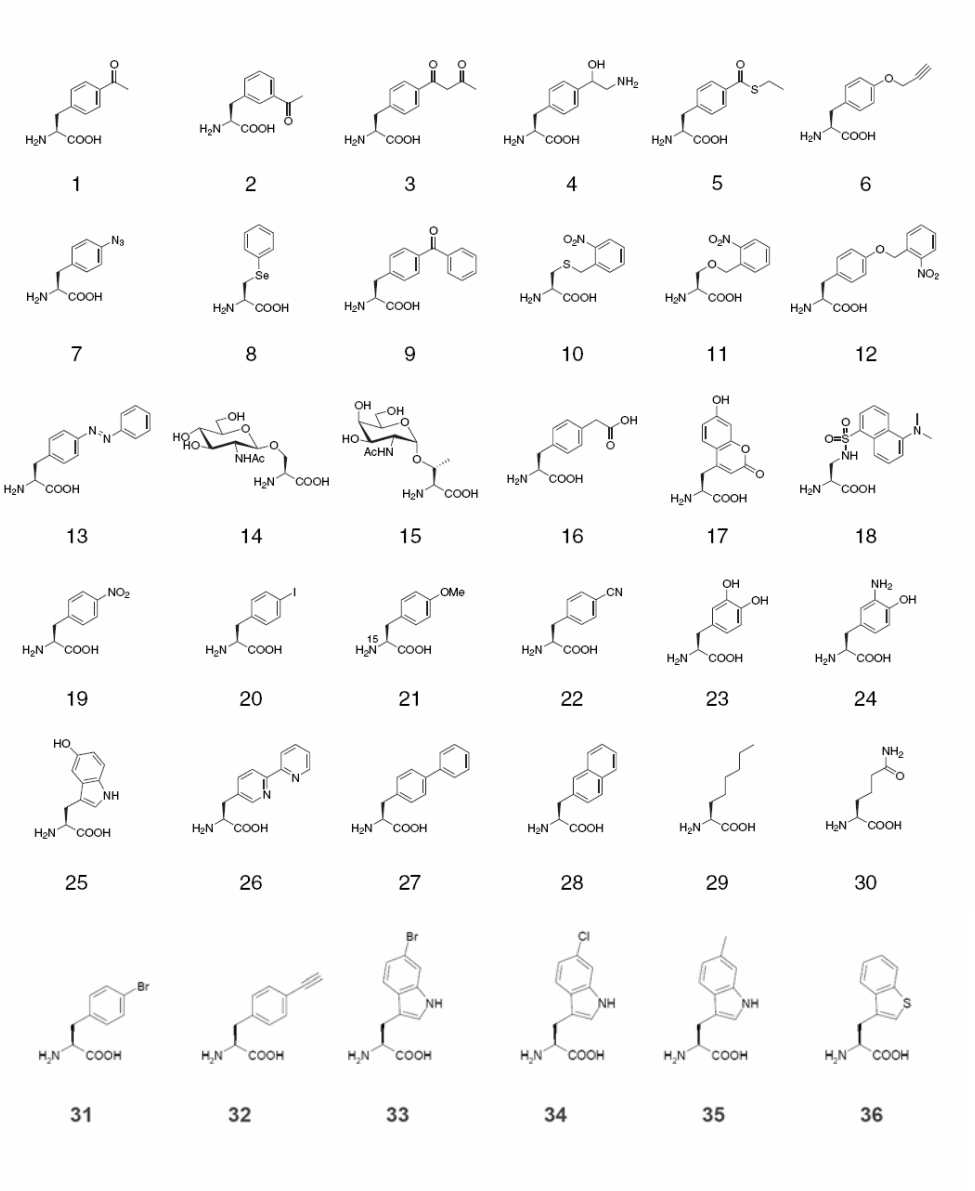


Figure 1: Nonnatural amino acids that have been genetically incorporated into proteins in bacteria, yeast or mammalian cells. This figure is reprinted, with permission, from the *Annual Review of Biophysics and Biomolecular Structure*, Volume **35**, Wang, L.; Xie, J.; Schultz, P. G., Expanding the genetic code, 225,[27] Copyright (2006) by Annual Reviews www.annualreviews.org.

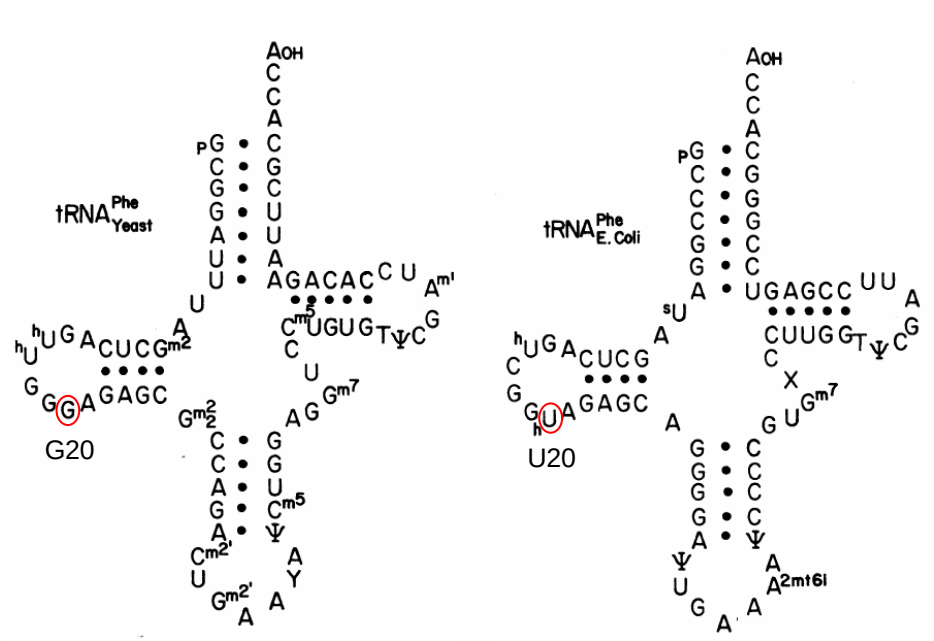
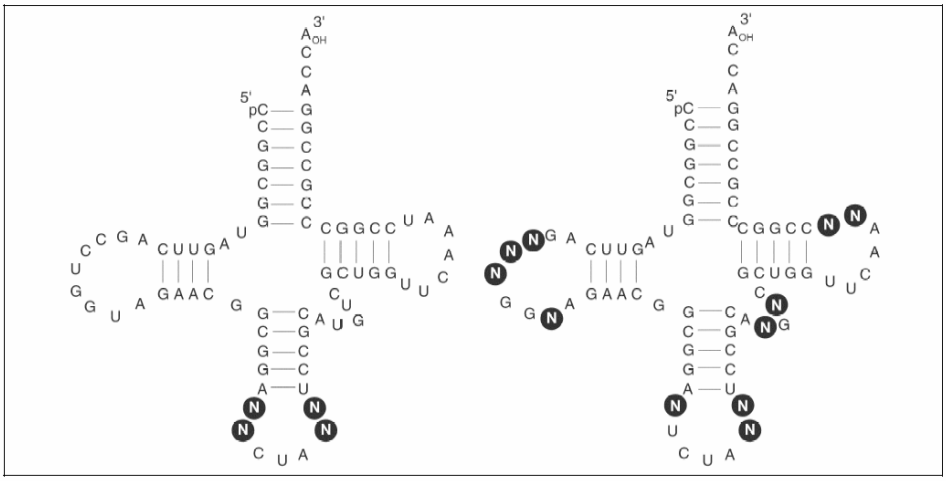


Figure 2: Structures of yeast tRNA^{Phe} (left) and *E. coli* tRNA^{Phe} (right). This figure is reprinted from Dudock, B. S.; Diperi, C.; Michael, M. S. *J. Biol. Chem.* **1970**, *245*, 2465-2468,[114] Copyright 1970 with permission from the American Society for Biochemistry and Molecular Biology.

(a)



(b)

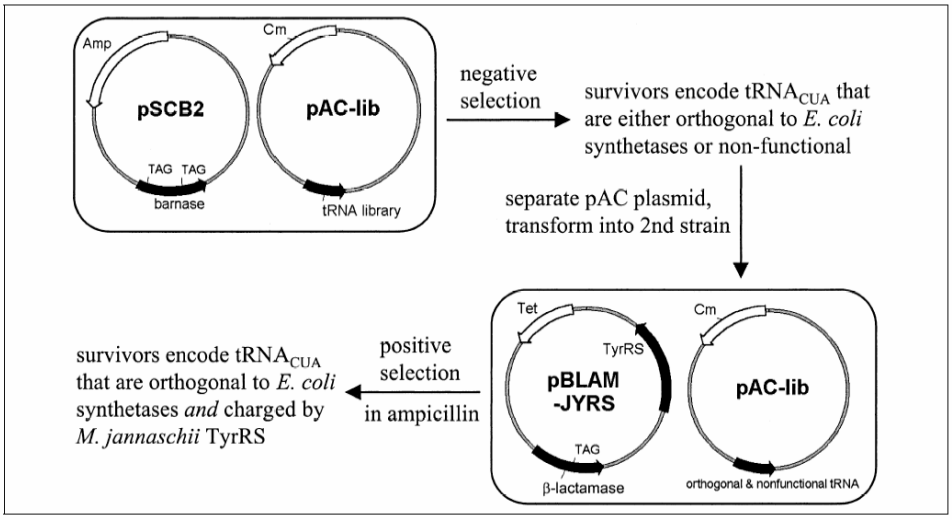
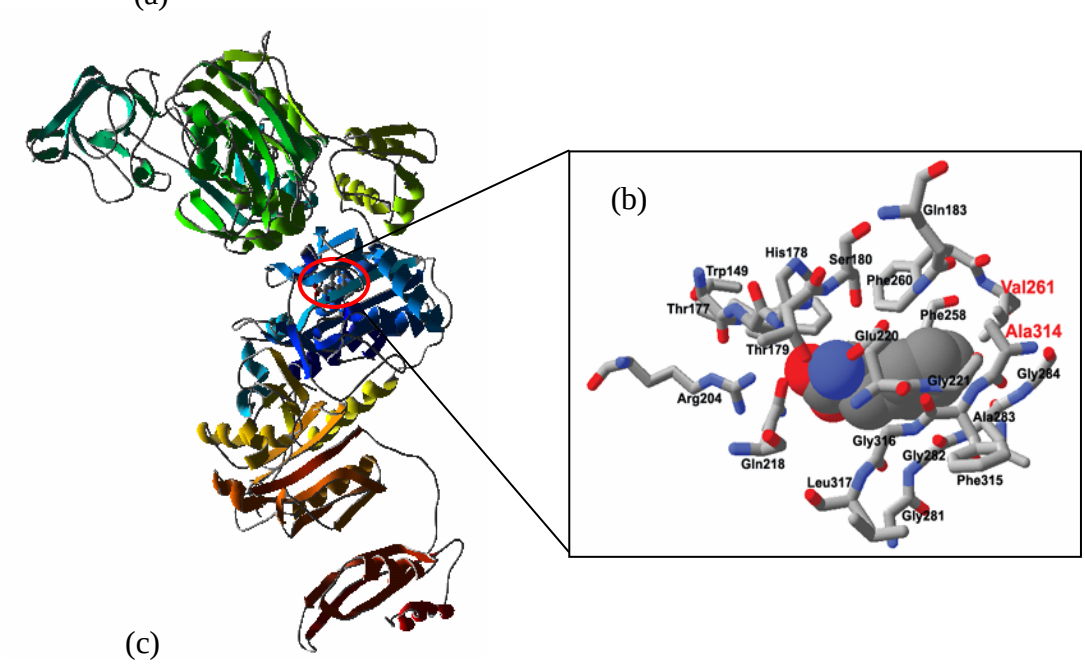


Figure 3: (a) Anticodon-loop tRNA library (left) and all-loop tRNA library (right) derived from *M. jannaschii* tRNA$^{Tyr}_{CUA}$ (N means all four nucleotides (A/U/G/C)). (b) A general selection scheme for suppressor tRNAs that are orthogonal to *the E. coli* endogenous systems and charged efficiently by a cognate synthetase. These figures are reprinted from from *Chem. Biol*., **8**, Wang, L.; Schultz, P.G., A general strategy for orthogonal tRNA, 883,[25] Copyright (2001) with permission from Elsevier.

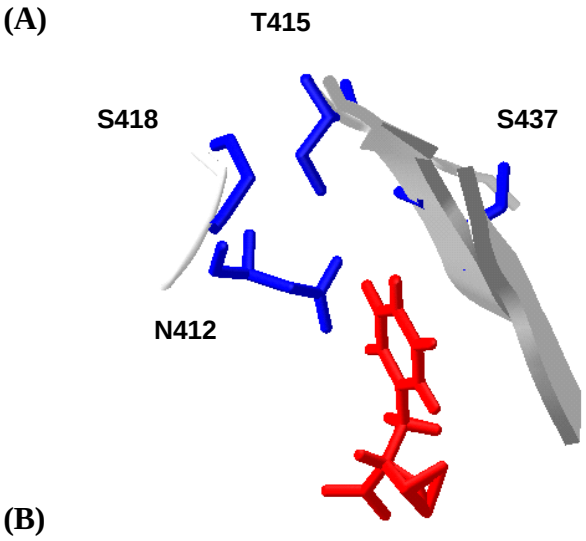
(a)



(c)

T. thermophilus	257YFPFVEP263	·····	313FAFGLGVERLAML325
E. coli	247YFPFTEP253	·····	293FAFGMGMERLTML305
S. cerevisiae	410YNPYTEP416	·····	456LGWGLSLERPTMI468
B. burgdorferi	441YFPFTEP447	·····	484IAWGIGIDRMALM496
B. subtilis	251FFPFTEP257	·····	308FAFGMGVERIAML320
A. aeolicus	247YFPFTEP253	·····	304FAFGMGVERLAML316
H. pylori	240FFPFTEP246	·····	295FAFGMGIERLAML307
H. influenzae	249YFPFTEP255	·····	295FAVGMGVERLTML307
C. trachomatis	252YFPFVEP258	·····	308YALGMGIERLAML320
M. genitalium	249HFPFTEP255	·····	306IAAGIGIERLAML318
H. sapiens	409YNPYTEP415	·····	455IAWGLSLERPTMI467
C. muridarum	251YFPFVEP257	·····	307YALGMGIERLAML319
C. pneumoniae	249YFPFVEP255	·····	305YAVGMGIERLAML317
D. radiodurans	242YYPFVEP248	·····	301FAFGLGLERIAML313
C. jejuni	241FFPFTEP247	·····	296YAFGLGVERFAML308
P. abyssi	420YYPFTEP426	·····	463IAWGIGIDRLAMF475
R. prowazekii	254FFPFTEP260	·····	301FAFGLGVERFAML313
N. meningitidis	249FFPFTEP255	·····	295FAFGIGLDRFAML307
U. urealyticum	243FFPFTEP249	·····	299MAFGVGIDRIAML311
T. pallidum	476YFPFTEP482	·····	519MAWGLGVDRMALL531
T. maritima	246FFPFTEP252	·····	292YAFGMGVERIAML304

Figure 4: (a) Crystal structure of the *T. thermophilus* PheRS (*t*PheRS, pdb 1B70) in ribbon model. Only ($\alpha\beta$) monomeric portion of structure is shown. (b) Active sites of *t*PheRS. Only residues within 6 Å of substrate Phe are shown. Substrate is shown as space-filling model, while residues surrounding the substrate are shown in stick model. (c) Sequence alignment of PheRS from 21 different organisms. Only residues flanking equivalent residues of V261 and A314 in *T. thermophilus* are represented. The shaded residues are the conserved residues equivalent to V261 and A314 from *T. thermophilus*. Sequences adjacent to V261 and A314 are also highly conserved. The sequences are obtained from aminoacyl-tRNA synthetase database (http://rose.man.poznan.pl/aars/). These figures were adapted from Wang's thesis (2003).[41]


(A)
T415
S418
S437
N412
(B)

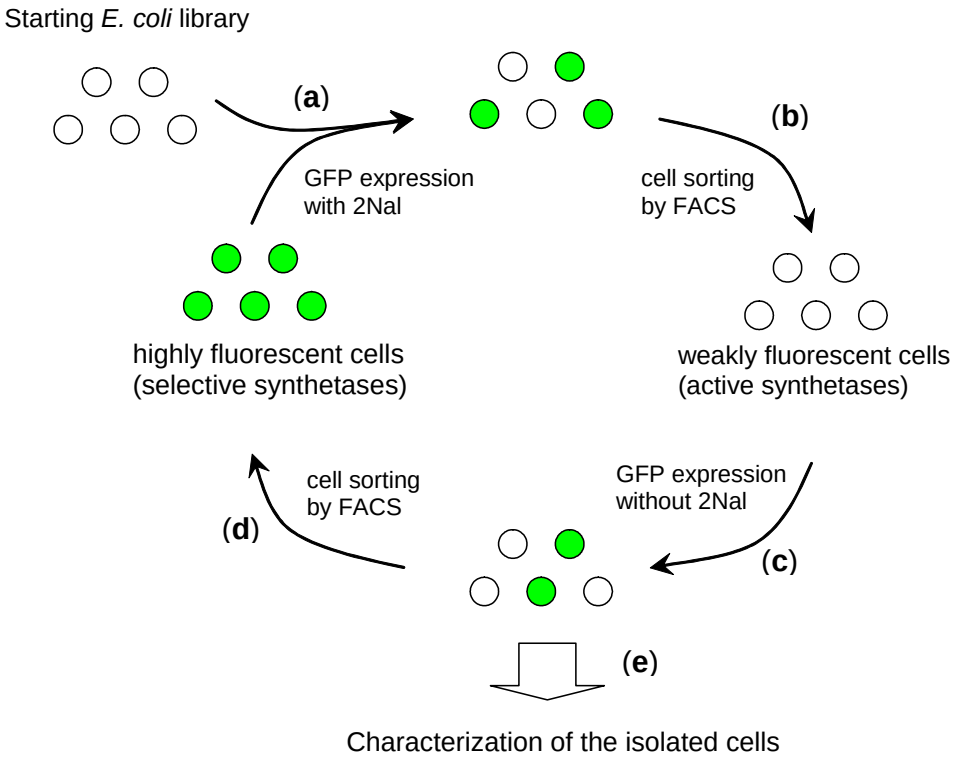

Starting E. coli library
(a)
GFP expression
with 2Nal
(b)
cell sorting
by FACS
highly fluorescent cells
(selective synthetases)
weakly fluorescent cells
(active synthetases)
cell sorting
by FACS
(d)
GFP expression
without 2Nal
(c)
(e)
Characterization of the isolated cells

Figure 5: (**A**) Phe substrate (red) and four residues (blue) within 7 Å of the *para*-position of the phenyl ring of the substrate inside the binding pocket of a homology model of yPheRS. (**B**) A screening scheme for yPheRS library. GFP6 in yPheRS expression library *E. coli* cells outfitted with $ytRNA^{Phe}_{AAA}$ and yPheRS library was expressed in the presence of 2Nal (**a**). Weakly fluorescent cells that contain active yPheRS variants were enriched by FACS (**b**). GFP6 in the collected cells was expressed in the absence of 2Nal (**c**). Highly fluorescent cells that contain selective yPheRS variants were enriched by FACS (**d**). After two rounds of screening, ten colonies were isolated from the enriched cells and characterized (**e**).

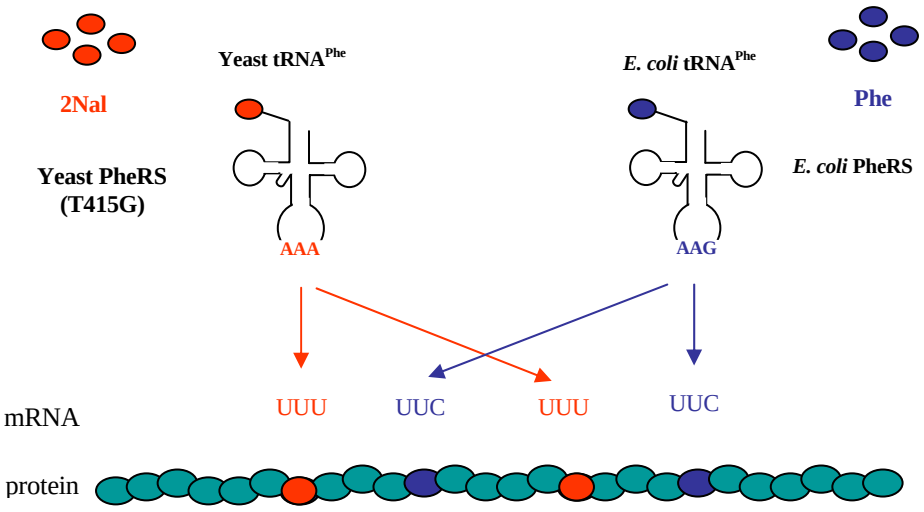


Figure 6: A schematic diagram describing the concept of breaking the degeneracy of the genetic code.