Expanding the Genetic Code

Lei Wang,[1] Jianming Xie,[2] and Peter G. Schultz[2]

[1]The Jack H. Skirball Center for Chemical Biology & Proteomics, The Salk Institute for Biological Studies, La Jolla, California 92037

[2]Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037; email: schultz@scripps.edu

Annu. Rev. Biophys. Biomol. Struct.
2006. 35:225–49

First published online as a
Review in Advance on
January 13, 2006

The *Annual Review of
Biophysics and Biomolecular
Structure* is online at
biophys.annualreviews.org

doi: 10.1146/
annurev.biophys.35.101105.121507

1056-8700/06/0609-
0225$20.00

Key Words

amino acid, directed evolution, protein engineering, aminoacyl-tRNA synthetase

Abstract

Recently, a general method was developed that makes it possible to genetically encode unnatural amino acids with diverse physical, chemical, or biological properties in *Escherichia coli*, yeast, and mammalian cells. More than 30 unnatural amino acids have been incorporated into proteins with high fidelity and efficiency by means of a unique codon and corresponding tRNA/aminoacyl-tRNA synthetase pair. These include fluorescent, glycosylated, metal-ion-binding, and redox-active amino acids, as well as amino acids with unique chemical and photochemical reactivity. This methodology provides a powerful tool both for exploring protein structure and function in vitro and in vivo and for generating proteins with new or enhanced properties.

Contents

INTRODUCTION

Genetic code: consists of 64 nucleotide triplets (A, U, C, and G) that encode the 20 amino acids used in protein translation, and is universal in all known organisms

Unnatural amino acids: amino acids not specified by the existing genetic code for protein synthesis

With the rare exceptions of selenocysteine (9) and pyrrolysine (72), the genetic codes of all known organisms specify the same 20 amino acid building blocks. Although a number of arguments have been put forth to explain the nature and number of amino acids in the code, it is clear that proteins require many additional chemistries, beyond the limited number of functional groups contained in the 20 amino acids, to carry out their natural functions. Thus while a 20 amino acid code is sufficient for life, it may by no means be ideal. The development of a method that allows one to genetically encode additional amino acids might enable the evolution of proteins, or even entire organisms, with new or enhanced properties.

Such methodology would also provide powerful new tools to probe protein structure and function both in vitro and in vivo. Although conventional site-directed mutagenesis has dramatically expanded our ability to manipulate protein structure, the number of substitutions that can be made remains limited. Ideally, one would like to be able to make precise changes in the steric and electronic properties of an amino acid (e.g., acidity, size, redox potential, nucleophilicity, and polarity) or introduce spectroscopic probes, posttranslational modifications, metal chelators, photoaffinity labels, and other chemical moieties at unique sites in a protein. Here, we describe an approach that makes it possible for the first time to uniquely encode additional amino acids with novel physical, chemical, or biological properties in both prokaryotic and eukaryotic organisms with high fidelity and efficiency.

BACKGROUND

Both chemical and biosynthetic strategies have been developed to incorporate unnatural amino acids into proteins. The former are simple and straightforward but are often limited by the homogeneity and/or size of the protein that can be synthesized. For example, chemical modification of amino acid side chains (13) can lead to nonselective and nonquantitative derivatization. Moreover, only a limited number of residues can be chemically modified with exogenous agents. Solid-phase peptide synthesis allows a large number of modifications to be made to protein structures but is generally limited to peptides and smaller proteins owing to the decreased yield and purity associated with the synthesis of larger proteins (46). The recent development of chemical (21) and intein-mediated (31, 36) peptide ligation allows the semisynthesis of larger proteins, but substitutions are confined largely to the N or C terminus.

A general in vitro biosynthetic method for incorporating unnatural amino acids into proteins uses nonsense or frameshift suppressor tRNAs that are chemically misacylated with the amino acid (6, 40, 62). This method can be used to synthesize proteins of virtually any size with the novel amino acid located at any site designated by the corresponding codon. Indeed, more than 80 novel amino

acids have been incorporated into proteins by this methodology (18). However, the protein in cell-free translation systems yields are low and the generation of aminoacyl-tRNA is relatively complex.

The ability to incorporate unnatural amino acids directly into proteins in vivo offers considerable advantages over both chemical and in vitro biosynthetic strategies, including higher yields, fidelity, and technical ease; it also allows one to study protein structure and function both in vitro and in vivo. One such method involves growing bacteria in media in which a close structural analogue is substituted for the corresponding genetically encoded amino acid. For example, replacement of methionine with selenomethionine (38), or a common amino acid with its ^{15}N-labeled analogue (60), has been used extensively in crystallographic and NMR studies of protein structure, respectively. However, nonquantitative substitution and substitution at multiple sites throughout the protein (and proteome) can limit the utility of this technique. The use of strains that are auxotrophic in a particular amino acid minimizes the competition between the unnatural amino acid and its natural counterpart (42), but the former must nonetheless be a close analogue of the common amino acid. In some cases it is also possible to relax the substrate specificity of the aminoacyl-tRNA synthetases through active-site mutations (44) or to attenuate the proofreading activity of the synthetase (26, 74). Finally, microinjection of a chemically misacylated amber suppressor tRNA and the corresponding mutant mRNA into *Xenopus laevis* oocytes has led to the selective incorporation of unnatural amino acids into proteins (7). Unfortunately, protein yields are again low because the tRNA is chemically acylated in vitro and cannot be reacylated in vivo.

METHODOLOGY

To cotranslationally introduce an unnatural amino acid at a defined site in a protein directly in a living organism, one requires a unique tRNA-codon pair, a corresponding aminoacyl-tRNA synthetase, and significant intracellular levels of the unnatural amino acid. To ensure that the unnatural amino acid is incorporated uniquely at the site specified by its codon, the tRNA must be constructed such that it is not recognized by the endogenous aminoacyl-tRNA synthetases of the host, but functions efficiently in translation (an orthogonal tRNA). Moreover, this tRNA must deliver the novel amino acid in response to a unique codon that does not encode any of the common 20 amino acids. Another requirement for high fidelity is that the cognate aminoacyl-tRNA synthetase (an orthogonal synthetase) aminoacylates the orthogonal tRNA but none of the endogenous tRNAs. In addition, this synthetase must aminoacylate the tRNA with only the desired unnatural amino acid and no endogenous amino acids. Likewise, the unnatural amino acid cannot be a substrate for the endogenous synthetases if it is to be incorporated uniquely in response to its cognate codon. Finally, the amino acid must be transported efficiently into the cytoplasm when added to the growth medium, or biosynthesized by the host, and be stable to endogenous metabolic enzymes.

Aminoacyl-tRNA synthetase: an enzyme that catalyzes the attachment of a specific amino acid to the acceptor stem at the 3′ end of the cognate tRNA

An Orthogonal tRNA-Codon Pair

Efforts to develop such an approach focused first on *Escherichia coli* because of its ease of genetic manipulation and the large body of knowledge around its translational machinery. To uniquely specify an unnatural amino acid, one can use either nonsense (triplet) codons or frameshift (quadruplet) codons, or even construct a bacterium in which redundant codons and their corresponding tRNAs are deleted from the genome. Initially the amber nonsense codon (UAG) was used to specify the unnatural amino acid because it is the least-used stop codon in *E. coli* and *Saccharomyces cerevisiae*. Moreover, some *E. coli* strains contain natural amber suppressor tRNAs that efficiently incorporate common amino acids without significantly affecting growth rates

Orthogonal tRNA/aminoacyl-tRNA synthetase pair: the tRNA is not recognized by the endogenous aminoacyl-tRNA synthetases of the host but functions efficiently in translation; the cognate aminoacyl-tRNA synthetase aminoacylates the orthogonal tRNA but none of the endogenous tRNAs

***Mj*:** *Methanococcus jannaschii*

TyrRS: tyrosyl-tRNA synthetase

(8, 35). Amber suppressors can also be engineered, and natural or engineered amber suppressors have been used routinely for conventional protein mutagenesis in *E. coli* (63) and for the in vitro introduction of unnatural amino acids into proteins using chemically aminoacylated suppressor tRNAs. In addition, nonsense suppressors also exist in mammalian cells (27, 43) and yeast (48, 71).

In theory it should be possible to evolve an orthogonal tRNA/aminoacyl-tRNA synthetase pair from an existing bacterial pair for use in *E. coli*. This was first attempted by mutating an *E. coli* glutaminyl tRNA to a translationally competent suppressor tRNA that is no longer recognized by its cognate synthetase. Although an orthogonal tRNA was successfully generated (56), we were not able to alter the specificity of the corresponding glutaminyl-tRNA synthetase to selectively recognize this new suppressor tRNA and not the wild-type tRNA (which would result in misincorporation of the unnatural amino acid at glutamine sites) (55). An alternative strategy for generating orthogonal tRNA/aminoacyl-tRNA synthetase pairs in bacteria makes use of orthologues from archaeal bacteria and eukaryotic organisms. Some prokaryotic tRNA/aminoacyl-tRNA synthetase pairs do not cross-react to any significant degree with their eukaryotic counterparts, as a result of differences in tRNA identity elements, primarily in the acceptor stem and variable arm (25, 53). Therefore, it should be possible to generate an orthogonal tRNA/synthetase pair for *E. coli* by importing a pair from an eukaryotic organism. Although we found that the yeast amber suppressor $tRNA^{Gln}_{CUA}$ (*Sc* $tRNA^{Gln}_{CUA}$) and cognate glutaminyl-tRNA synthetase are orthogonal in *E. coli* (57), no mutant yeast GlnRS could be evolved that aminoacylates the *Sc* $tRNA^{Gln}_{CUA}$ with an unnatural amino acid in *E. coli* (possibly owing to the low intrinsic expression and/or activity of the synthetase in bacteria). Efforts to use yeast tRNA/synthetase pairs specific for various hydrophobic amino acids, as well as pairs from *Homo sapiens*, were also unsuccessful (L. Wang & P. G. Schultz, unpublished results).

Attention was then focused on archaea as a source of orthogonal tRNA/synthetase pairs for use in *E. coli*. Archaeal aminoacyl-tRNA synthetases are more similar to their eukaryotic than prokaryotic counterparts (45, 53), but unlike synthetases from eukaryotic cells, which often express poorly or have low activities in *E. coli*, synthetases from archaea can be expressed efficiently in *E. coli* in their active forms. Moreover, early work (53) indicated that most tRNAs from the halophile *Halobacterium cutirebrum* cannot be charged by *E. coli* aminoacyl-tRNA synthetases. The first orthogonal *E. coli* tRNA/synthetase pair to be generated from archaeal bacteria was derived from the tyrosyl pair from *Methanococcus jannaschii* (79). Previous experiments showed that the major recognition elements of *M. jannaschii* $tRNA^{Tyr}$ include the discriminator base A73 and the first base pair, C1-G72, in the acceptor stem; the anticodon triplet participates only weakly in identity determination (32). This pattern is the same as that of yeast $tRNA^{Tyr}$ but differs from that of *E. coli* $tRNA^{Tyr}$; the latter uses A73, G1-C72, a long variable arm, and the anticodon as identity elements (**Figure 1*a***). The *M. jannaschii* tyrosyl-tRNA synthetase (*Mj* TyrRS) also has a minimalist anticodon loop binding domain (73), which should make it possible to change the anticodon loop of its cognate tRNA to CUA with little loss in affinity by the synthetase. Finally, this aminoacyl-tRNA synthetase does not have an editing mechanism that could deacylate the attached unnatural amino acid. Indeed, an amber suppressor *M. jannaschii* $tRNA^{Tyr}_{CUA}$ (*Mj* $tRNA^{Tyr}_{CUA}$) and cognate *Mj* TyrRS were shown to function efficiently in *E. coli*, but unfortunately some degree of aminoacylation of this tRNA by endogenous *E. coli* synthetases was observed (81).

A general strategy for the evolution of orthogonal tRNAs in *E. coli* from heterologous precursors was therefore developed (82). This method consists of a combination of negative

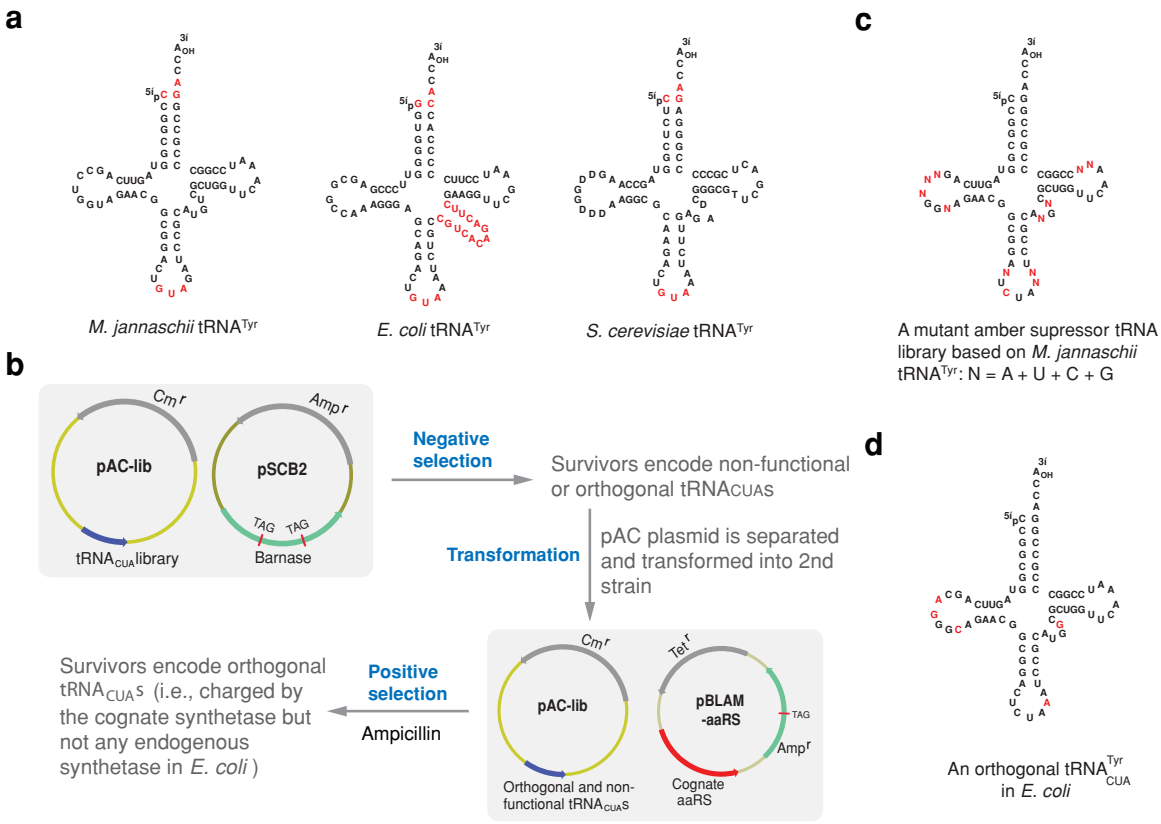

Figure 1

(*a*) Comparison of sequences of $tRNA^{Tyr}$ from three different species. The major identity elements are in red. (*b*) Directed evolution of orthogonal amber suppressor $tRNA^{Tyr}_{CUA}$ in *E. coli* by alternating negative and positive selections. (*c*) A library of amber suppressor $tRNA^{Tyr}$ was generated by randomizing 11 nucleotides (*red*) of *M. jannaschii* $tRNA^{Tyr}_{CUA}$ that do not interact directly with the cognate TyrRS. (*d*) The orthogonal amber suppressor *Mj* $tRNA^{Tyr}_{CUA}$ (changed nucleotides are in *red*).

and positive selections with a library of tRNA mutants (derived from a heterologous suppressor tRNA) in the absence and presence of the cognate synthetase, respectively (**Figure 1*b***). In the negative selection, the tRNA library is introduced into *E. coli* along with a mutant barnase gene in which amber nonsense codons are introduced at sites permissive to substitution by other amino acids. When a member of the suppressor tRNA library is aminoacylated by an endogenous *E. coli* synthetase (i.e., it is not orthogonal to the *E. coli* synthetases), the amber codons are suppressed and the ribonuclease barnase is produced, resulting in cell death. Only cells harboring orthogonal or nonfunctional tRNAs can survive. All tRNAs from surviving clones are then subjected to a positive selection in the presence of the cognate heterologous synthetase and a β-lactamase gene with an amber codon at a permissive site. tRNAs that can function in translation and are good substrates for the cognate heterologous synthetase are selected on the basis of their ability to suppress the amber codon and produce active β-lactamase. Therefore, only tRNAs that

CAT: chloramphenicol acetyl transferase

OMeTyr: *O*-methyl-L-tyrosine

(*a*) are not substrates for endogenous *E. coli* synthetases, (*b*) can be aminoacylated by the synthetase of interest, and (*c*) function in translation will survive both selections.

This approach was applied to the *Mj* $tRNA^{Tyr}_{CUA}$ to further reduce recognition of this tRNA by endogenous *E. coli* synthetases, while preserving activity with both the cognate synthetase and translational machinery (82). Eleven nucleotides of *Mj* $tRNA^{Tyr}_{CUA}$ that do not interact directly with the *Mj* TyrRS were randomly mutated (based on consensus analysis of archaeal $tRNA^{Tyr}$ sequences) to generate a suppressor tRNA library (**Figure 1*c***). This tRNA library was passed through rounds of negative and positive selections to afford a functional, orthogonal tRNA ($mutRNA^{Tyr}_{CUA}$) that functions efficiently with *Mj* TyrRS to translate the amber codon (**Figure 1*d***). Additional orthogonal tRNA/synthetase pairs have since been generated and include a $tRNA^{Asp}_{CUA}$/AspRS pair derived from yeast (65) and an *E. coli* initiator $tRNA^{fMet}_{CUA}$/yeast TyrRS pair (52). Orthogonal suppressor tRNAs can also be derived from consensus sequences of multiple archaeal tRNAs and then improved with the above selections (4). This approach has been used to evolve an orthogonal *Methanococcus thermoautotrophicum* $tRNA^{Leu}_{CUA}$/LeuRS pair (3), an orthogonal *Methanosarcina mazei* $tRNA^{Glu}_{CUA}$/GluRS pair (68), and an orthogonal *Pyrococcus horikoshii* $tRNA^{Lys}_{CUA}$/LysRS pair (4).

An Orthogonal Aminoacyl-tRNA Synthetase

Next, it was necessary to alter the substrate specificity of the orthogonal *Mj* TyrRS to charge its cognate orthogonal tRNA with only the desired unnatural amino acid and none of the common 20 amino acids. Our goal was to develop a general scheme for evolving the specificity of aminoacyl-tRNA synthetases that is independent of the structure of the amino acid of interest. One such approach involves generating a focused library of synthetase active-site mutants and again using a combination of positive and negative selections to evolve synthetases that aminoacylate the cognate tRNA with the unnatural and no endogenous amino acids (79, 83). Libraries of synthetase variants were generated by randomizing five or six residues in the substrate binding pocket of the synthetase, based on an analysis of the X-ray crystal structure of the synthetase (or a homologue) complexed with its cognate amino acid or aminoacyl adenylate (**Figure 2*a***). To identify synthetase variants that specifically recognize the unnatural amino acid and no endogenous host amino acid, the synthetase libraries were subjected to alternating rounds of positive and negative selections (**Figure 2*b***). In *E. coli*, the positive selection is based on resistance to chloramphenicol conferred by suppression of an amber mutation at a permissive site in the chloramphenicol acetyl transferase (CAT) gene; the negative selection uses the barnase gene with amber mutations at permissive sites. When the library of synthetase mutants is passed through the positive selection in the presence of the unnatural amino acid, those cells with mutant synthetases that can acylate the tRNA with either the unnatural amino acid or an endogenous amino acid survive. Plasmids encoding active mutant synthetases are then transformed into the negative selection strain, and selections are carried out in the absence of the unnatural amino acid. Those cells containing mutant synthetases that recognize endogenous amino acids incorporate the latter in response to the amber codons in the barnase gene and die. Repeated rounds of positive and negative selections lead to the isolation of mutant synthetases that can specifically incorporate the unnatural amino acid in response to the amber codon.

This selection scheme was first used to evolve a *Mj* TyrRS mutant capable of selectively inserting *O*-methyl-L-tyrosine (OMeTyr) into proteins in *E. coli* in response to the amber codon (79). A library of synthetase mutants was generated by randomizing the five active-site residues Tyr-32,

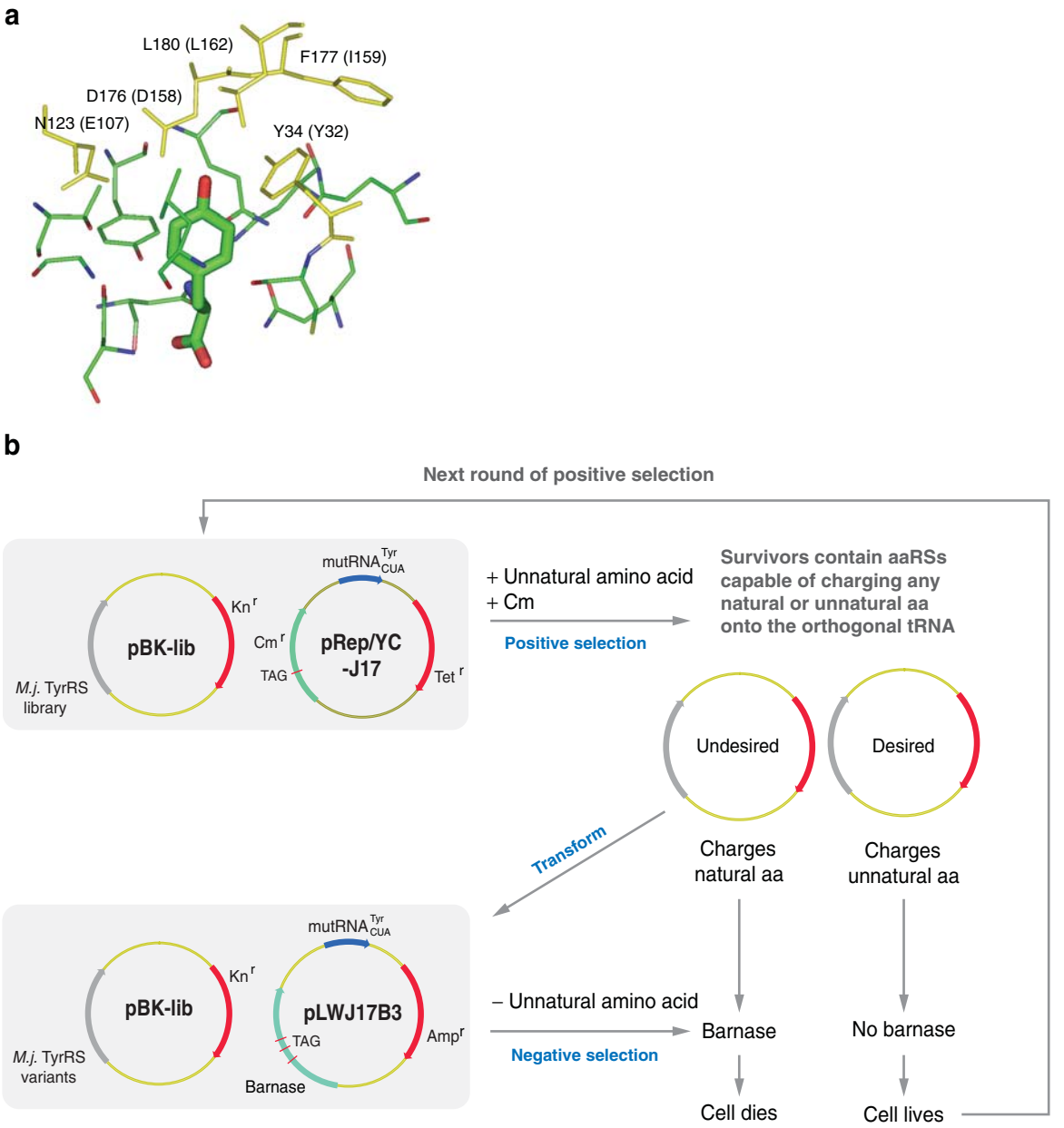

Figure 2

Modification of the amino acid specificity of an orthogonal *M. jannaschii* TyrRS (*Mj* TyrRS) in *E. coli*. (*a*) A library of *Mj* TyrRS mutants was generated by randomizing five residues (in parentheses) in the tyrosine binding site. (*b*) A general positive and negative selection scheme for evolving synthetase variants specific for an unnatural amino acid in *E. coli*. Cm, chloramphenicol.

GFP: green fluorescent protein

***p*AcPhe:** *p*-acetylphenylalanine

Glu-107, Asp-158, Ile-159, and Leu-162 based on the crystal structure of the homologous *Bacillus stearothermophilus* TyrRS–tyrosyl adenylate complex (**Figure 2*a***) [more recently, the X-ray structure of *Mj* TyrRS itself was solved (49, 90), allowing the generation of improved libraries]. This library was then subjected to rounds of positive and negative selection to afford clones that survived at high chloramphenicol concentrations in the presence of OMeTyr and at low chloramphenicol concentrations in its absence. Incorporation of OMeTyr by the evolved *Mj*tRNA$_{CUA}$/synthetase pair was assayed directly by suppression of amber codons in *E. coli* dihydrofolate reductase (DHFR). In the presence of OMeTyr and the mutant *Mj* tRNA$_{CUA}$/synthetase pair, 5 to 10 mg/liter of protein was purified from minimal media. In the absence of any one component (synthetase, tRNA, or amino acid) no protein could be detected by Western blot analysis, silver stain, or coomassie stain. High-resolution mass analysis of intact protein and tryptic digests confirmed that only OMeTyr is incorporated in response to the amber codon, and OMeTyr is incorporated at no other site in the protein.

An alternative selection scheme makes use of an amber-T7/GFPuv (a type of green fluorescent protein) instead of an amber-barnase reporter in the negative selection (69). Suppression of amber codons introduced at permissive sites in T7 RNA polymerase produces full-length T7 RNA polymerase, which drives the expression of GFPuv. In this approach both the amber-CAT reporter and amber-T7/GFPuv reporter are encoded in a single plasmid. After positive selection, surviving cells are grown in the absence of both the unnatural amino acid and chloramphenicol. Cells containing mutant synthetases that can acylate the tRNA with any of the 20 common amino acids express GFPuv, whereas cells containing mutant synthetases that can acylate the tRNA only with the unnatural amino acid do not. These nonfluorescent cells are sorted using fluorescence-activated cell-sorting. One advantage of this latter method is that both reporters are contained within a single genetic construct, eliminating the need for plasmid shuttling between positive and negative selections. Other selection schemes have been pursued, including cell surface and phage display systems, but these are less general (i.e., require capture reagents specific for the amino acid of interest) or not as efficient (66).

More than 30 unnatural amino acids have been incorporated into proteins in *E. coli*. In general, for most unnatural amino acids, suppression efficiencies range from 25% to 75% of wild-type protein and translational fidelity is >99%. In an optimized system, approximately 1 g/liter of mutant protein containing the unnatural amino acid *p*-acetylphenylalanine (*p*AcPhe) was produced (H. Cho & T. Daniel, unpublished results).

Encoding Unnatural Amino Acids in Eukaryotic Cells

In addition to having distinct identity elements, eukaryotic tRNAs differ from bacterial and archaeal tRNAs in transcription and modification. Eukaryotic tRNAs have internal A- and B-box sequences required for transcription, and the 3′-CCA is added enzymatically rather than encoded in the gene as in *E. coli* (34, 54). In addition, eukaryotic tRNAs are transcribed in the nucleus and must be exported to the cytoplasm via an exportin-tRNA-dependent process (5). Consequently, a new family of orthogonal tRNA/synthetase pairs was generated to genetically encode unnatural amino acids in eukaryotic organisms. These include an *E. coli* $tRNA^{Tyr}_{CUA}$/TyrRS pair (29, 30, 53) and a human initiator $tRNA^{fMet}_{CUA}$/*E. coli* GlnRS pair (52) that are orthogonal in yeast. In addition, a modified *Bacillus subtilis* $tRNA^{Trp}_{CUA}$/TrpRS (91) and a *B. stearothermophilus* $tRNA^{Tyr}_{CUA}$/*E. coli* TyrRS pair (67) were found to be orthogonal opal and amber suppressors, respectively, in mammalian cells.

To selectively introduce an unnatural amino acid into proteins in eukaryotes, Yokoyama and coworkers (47) screened a collection of designed active-site variants of *E. coli* TyrRS in a wheat germ translation system and discovered a mutant synthetase that uses 3-iodotyrosine more effectively than it uses tyrosine. This mutant synthetase was used with the *B. stearothermophilus* $tRNA^{Tyr}_{CUA}$ to incorporate 3-iodotyrosine into proteins in mammalian cells (67). Similarly, an orthogonal *B. subtilis* $tRNA^{Trp}_{UCA}$/TrpRS pair has been used to selectively introduce 5-hydroxytryptophan (5-HTTP) into proteins in 293T cells (91). On the basis of the crystal structure of the homologous *B. stearothermophilus* TrpRS, a synthetase mutant was generated that selectively charges 5-HTPP. This mutant synthetase and its cognate orthogonal tRNA were used to suppress the opal nonsense codon in the *foldon* gene in the presence of 5-HTPP (91). Indeed, expression of full-length protein was seen only in the presence of 5-HTPP. Electrospray mass spectrometry of the mutant foldon protein verified site-specific incorporation of 5-HTPP with a fidelity of >97%. The yield of the 5-HTPP mutant protein was approximately 100 μg/liter of culture, compared with that of about 1 mg/liter for wild-type protein.

To develop a general selection scheme in yeast to evolve synthetases specific for unnatural amino acids (analogous to those used in *E. coli*), a selection strain of *S. cerevisiae* that contains the transcriptional activator protein GAL4 was created in which codons at two permissive sites were mutated to amber nonsense codons (14) (**Figure 3**). Suppression of these amber codons leads to the production of full-length GAL4, which in turn drives transcription of genomic GAL4-responsive *his3*, *ura3*, and *lacZ* reporter genes. Expression of *HIS3* and *URA3* complements the histidine and uracil auxotrophy in this strain and provides a positive selection for clones expressing active tRNA/synthetase pairs. On the other hand, addition of 5-fluoroorotic acid (5-FOA), which is converted to a toxic product by URA3, results in the death of cells expressing active tRNA/synthetase pairs. In the absence of the unnatural amino acid, this serves as a negative selection to remove synthetases specific to endogenous amino acids. Like GFP, the *lacZ* reporter can serve as an additional chromogenic marker to identify active synthetase from inactive ones.

FOA: 5-fluoroorotic acid

This selection scheme has allowed us to evolve orthogonal *E. coli* $tRNA^{Tyr}_{CUA}$/TyrRS pairs that have been used to incorporate more than 10 unnatural amino acids into proteins in yeast (14, 22). A synthetase library (10^8 in size) was similarly constructed by randomizing five active-site residues in *E. coli* TyrRS. Mutant synthetases were identified after several rounds of positive and negative selection that incorporate a number of unnatural amino acids into proteins, with yields corresponding to 20% to 40% of those of wild-type protein and expression levels up to 75 mg/liter. A similar approach has been used to evolve orthogonal *E. coli* leucyl $tRNA_{CUA}$/LeuRS pairs that selectively incorporate photochromic and fluorescent amino acids into proteins in yeast (86). These same mutant aminoacyl-tRNA synthetases that were evolved in yeast (to accept *p*-azidophenylalanine, *p*-benzoyl-L-phenylalanine, *p*-iodophenylalanine, *p*-acetylphenylalanine, and *p*-methoxyphenylalanine) have been used together with a *B. stearothermophilus* amber suppressor $tRNA^{Tyr}_{CUA}$ to selectively insert unnatural amino acids into proteins in mammalian cells (P. G. Schultz, unpublished results), albeit currently in low yield.

Additional Codons to Specify Unnatural Amino Acids

It should also be possible to use quadruplet codons and cognate suppressor tRNAs with expanded anticodon loops to specify additional amino acids. There are many examples of naturally occurring +1 frameshift suppressors (11, 19, 64). Moreover, genetic selections have been used to identify efficient four- and five-base codon suppressor tRNAs from

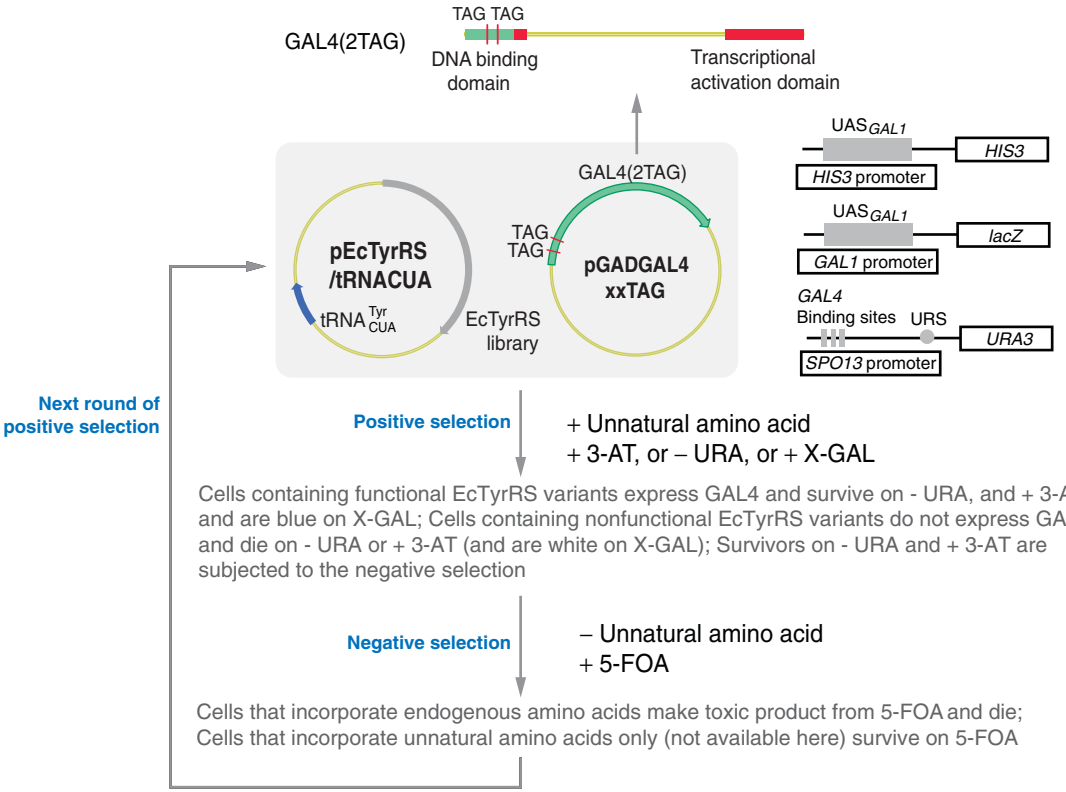

Figure 3

A general positive and negative selection scheme for evolving synthetase variants specific for unnatural amino acids in yeast. 5-FOA, 5-fluoroorotic acid.

mutant tRNA libraries (2, 59). The generation of an orthogonal tRNA/synthetase pair that decodes four-base codons requires either that the tRNA anticodon loop is not involved in tRNA/synthetase recognition or that the corresponding anticodon binding site of the synthetase can be modified to recognize a specific four-nucleotide anticodon sequence. Recently, an orthogonal four-base suppressor tRNA-synthetase pair was generated from $tRNA^{Lys}_{UCCU}$/LysRS of *Pyrococcus horikoshii*. The resultant orthogonal pair efficiently and selectively incorporates the unnatural amino acid homoglutamine into proteins in *E. coli* in response to the quadruplet codon AGGA (4). Frameshift suppression with homoglutamine does not significantly affect protein yields or cell growth rates, and it is mutually orthogonal with amber suppression. This has allowed the simultaneous incorporation of two unnatural amino acids at distinct sites within a single protein, suggesting that neither the number of available triplet codons nor the translational machinery itself represents a significant barrier to further expansion of the code.

An alternative approach that can be used to generate additional codons that uniquely encode unnatural amino acids involves eliminating degenerate codon-tRNA pairs from the *E. coli* genome. To this end we have shown that it is possible to remove four codons from a number of genes, including genes essential for growth, without affecting *E. coli* growth rates.

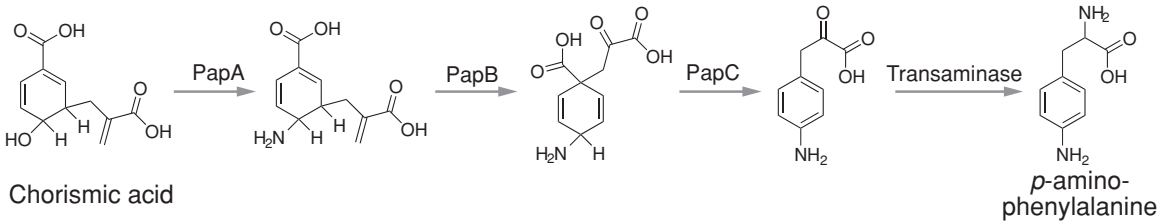

Figure 4

The biosynthesis of an unnatural amino acid, *p*-aminophenylalanine, in *E. coli*. Proteins PapA, PapB, and PapC convert chorismate to *p*-aminophenylpyruvic acid, and *E. coli* aromatic aminotransferase completes the biosynthesis to afford *p*-aminophenylalanine.

We are currently evaluating methods for the efficient construction of a "codon-deleted" *E. coli* genome.

Amino Acid Transport and Biosynthesis

We have found from measurements of cytoplasmic levels of amino acids added to the growth medium that a large number of amino acids are efficiently transported to the *E. coli* cytoplasm in millimolar concentrations. Highly charged or hydrophilic amino acids may require derivatization (e.g., esterification, acylation) with groups that hydrolyze in the cytoplasm. Metabolically labile amino acids or analogues (e.g., α-hydroxy acids, *N*-methyl amino acids) may require strains in which specific metabolic enzymes are deleted.

An alternative to adding exogenous amino acids to the growth media involves engineering a pathway for the biosynthesis of the unnatural amino acid directly in the host organism. For example, a completely autonomous 21-amino-acid bacterium has been generated that contains genes for the biosynthesis of *p*-amino-L-phenylalanine (*p*AF) from simple carbon sources, an aminoacyl-tRNA synthetase that uses *p*AF (and no other endogenous amino acids), and a tRNA that delivers *p*AF into proteins in response to the amber codon (61). *p*AF was biosynthesized from the metabolic intermediate, chorismic acid, using the *papA*, *papB*, and *papC* genes from *Streptomyces venezuelae* in combination with a nonspecific *E. coli* transaminase (**Figure 4**). *E. coli* containing these genes produced *p*AF at levels comparable to those of the other aromatic amino acids and had normal growth rates. In the presence of a *p*AF-specific, orthogonal mutRNA$^{Tyr}_{CUA}$/TyrRS pair, *E. coli* transformed with *papA-C* produced mutant proteins containing *p*AF at sites encoded by the amber codon with excellent yield and fidelity (61). In addition to *p*AF, it should be possible to biosynthesize and genetically encode other amino acids in vivo as well, including methylated, acetylated, and glycosylated amino acids.

UNNATURAL AMINO ACIDS AND APPLICATIONS

The above methodology has been used successfully to add more than 30 unnatural amino acids to the genetic codes of *E. coli*, yeast, or mammalian cells (**Figure 5**). Many of these unnatural amino acids have novel properties that are useful for a variety of biochemical and cellular studies of protein structure and function. For example, unnatural amino acids with chemically reactive groups have been genetically encoded in bacteria and/or yeast (16, 23, 85, 93), including *p*-acetylphenylalanine **1** and *m*-acetylphenylalanine **2**, *p*-(3-oxobutanoyl)-L-phenylalanine **3**, *p*-(2-amino-3-hydroxyethyl)phenylalanine **4**,

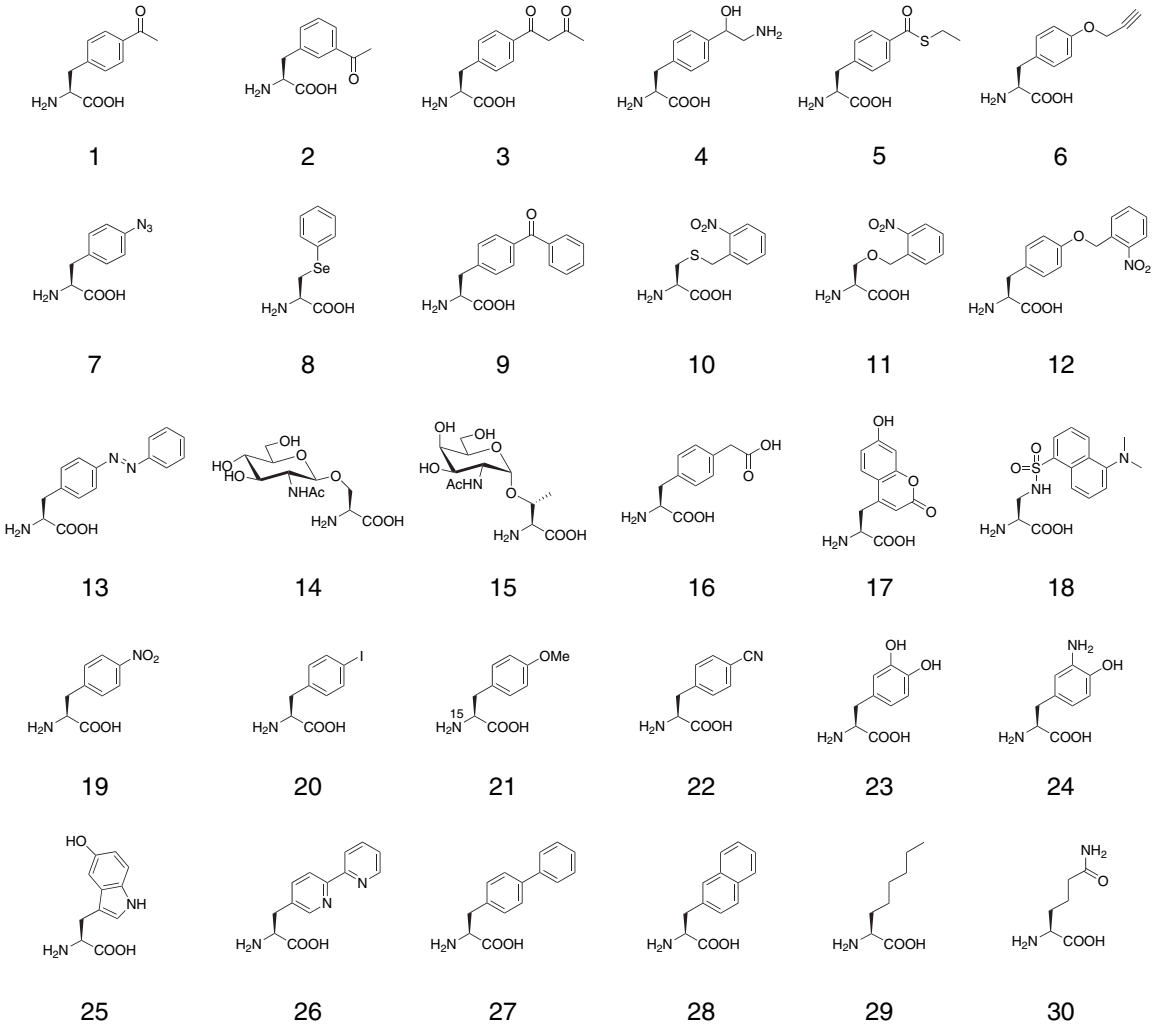

Figure 5

A list of representative unnatural amino acids that have been added to the genetic codes of *E. coli*, yeast, or mammalian cells.

p-ethylthiocarbonyl-phenylalanine **5**, *p*-propargyloxyphenylalanine **6**, and *p*-azidophenylalanine **7**. By virtue of their unique reactivity, these amino acids can be used to selectively modify native proteins under mild conditions with a variety of reagents in the absence of protecting groups. For example, the keto and β-diketo moieties selectively react with both hydrazides and alkoxyamines (85, 93); the azide can be selectively modified by a copper(I)-catalyzed [3+2] cycloaddition reaction with an alkyne derivative (23) or by a Staudinger conjugation reaction with a phosphine derivative (76); and the thioester group can be used in a modified chemical ligation reaction to derivatize protein side chains (and may allow the facile generation of cyclic and branched proteins and peptides). Similarly, a phenylselenide containing amino acid **8** has been introduced selectively into proteins,

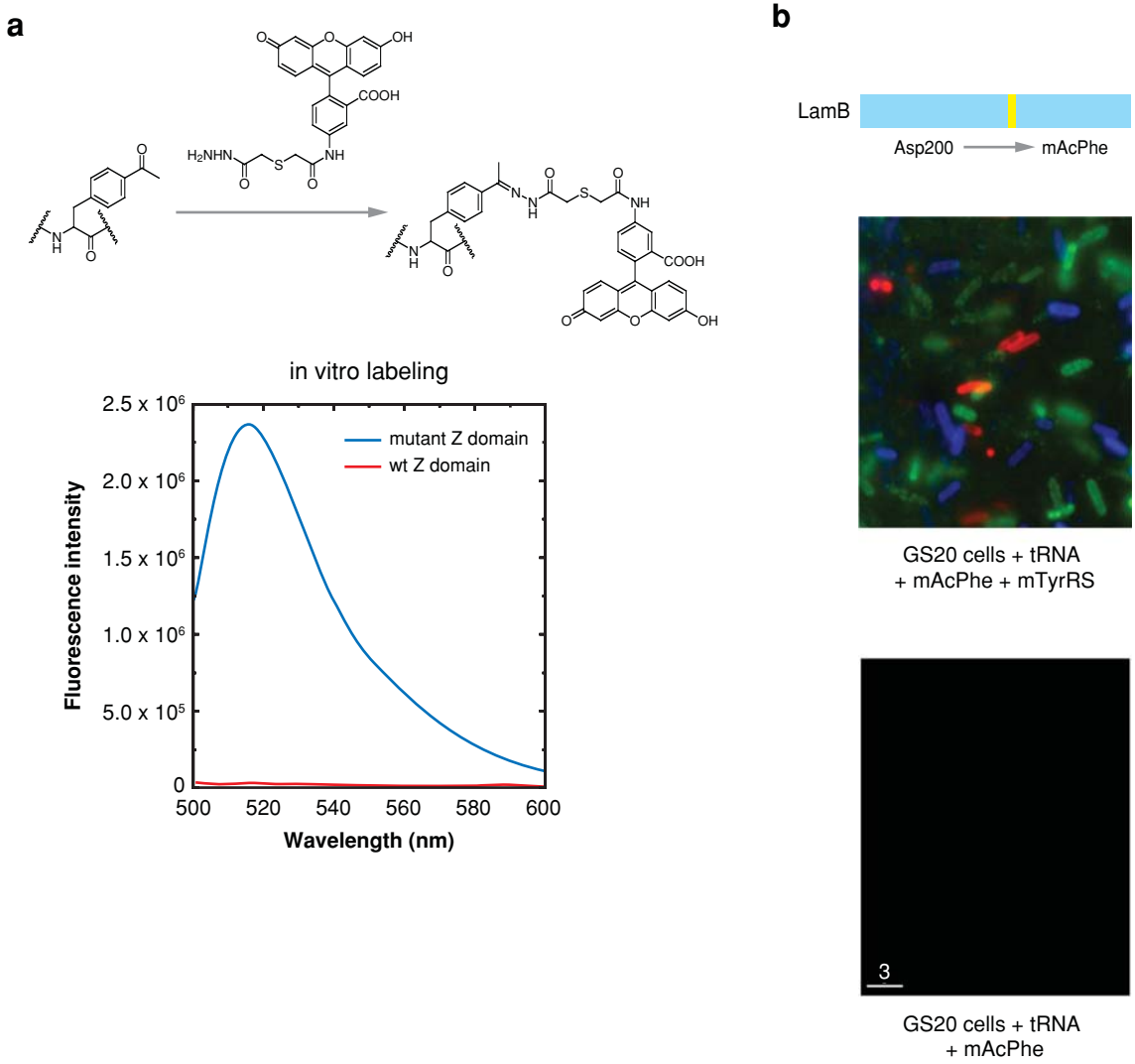

Figure 6

Site-specific modification of proteins containing a keto amino acid. (*a*) In vitro labeling of mutant Z domain that contains *p*-acetyl-L-phenylalanine with fluorescein hydrazide. (*b*) In vivo labeling of *E. coli* surface-expressed LamB mutant that contains *m*-acetyl-L-phenylalanine (*m*AcPhe) with the nonmembrane-permeable fluorescent dyes Cascade blue hydrazide (*blue*), Alexa568 hydrazide (*red*), and Alexa647 hydrazide (*green*).

which, after oxidation and β-elimination, allows intramolecular or intermolecular thiol conjugation reactions (P. G. Schultz, unpublished results).

These orthogonal chemistries have been used to selectively modify proteins with a number of fluorophores (85, 93) (**Figure 6*a***), tags, and other exogenous reagents (23, 85, 93). In one example, a mutant human growth hormone was site-specifically modified in high yield with polyethylene glycol (PEG) to afford a protein that retained wild-type activity but had a considerably improved half-life in serum

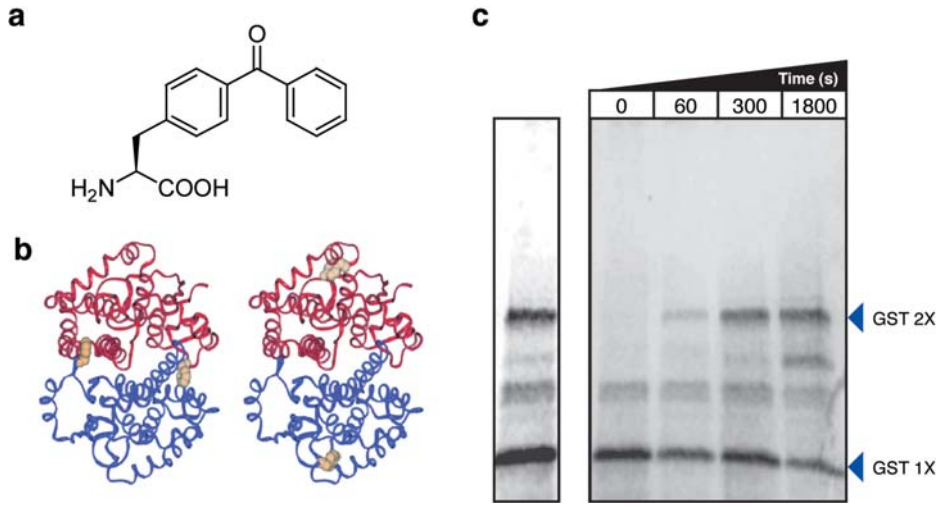

Figure 7

Site-specific incorporation of a photocrosslinker into proteins for mapping protein-protein interactions in vivo. (*a*) The chemical structure of *p*-benzoyl-L-phenylalanine (*p*Bpa). (*b*) Residue Phe-52 or Tyr-198 in *Schistosoma japonica* glutathione S-transferase (SjGST) was substituted by *p*Bpa. Monomers of the dimer are shown in blue and red. The side chain of Phe-52 is shown in orange for each monomer (*left*). The side chain of residue Tyr-198 is shown in orange (*right*). (*c*) The covalent dimerization of SjGST (Phe52*p*Bpa) in vivo upon irradiation at 365 nm.

PEG: polyethylene glycol

Grb2: growth factor receptor-bound protein 2

(H. Cho & T. Daniel, unpublished results). This work is being extended to other therapeutic proteins, as well as the generation of homo- and heterodimeric PEG-linked Fab fragments, and may allow the synthesis of chemically modified proteins and peptides with unprecedented control over selectivity, homogeneity, and chemical structure. In a second example a series of fluorescent dyes were selectively introduced at one site of a *m*-acetylphenylalanine mutant of the membrane protein lamB in live bacteria (**Figure 6*b***). In a third example of generating homogeneous glycoprotein mimetics, amino-oxy-containing sugars were selectively introduced at defined sites in proteins that were subsequently elaborated by adding additional saccharides to the pendant sugar with glycosyltransferases (58).

Two unnatural amino acids, *p*-azidophenylalanine **7** and *p*-benzoylphenylalanine **9**, have side chains that can be photocrosslinked both in vitro and in vivo (15–17). Benzophenone is a particularly useful photocrosslinker because it absorbs at relatively long wavelengths (~360 nm), and the excited state inserts efficiently into carbon-hydrogen bonds (33). Indeed, a mutant homodimeric glutathione S-transferase with *p*-benzoylphenylalanine substituted site-specifically at the dimer interface could be cross-linked in greater than 50% yield in the cytoplasm of *E. coli* (**Figure 7**). More recently, Yokoyama and coworkers showed that *p*-benzoylphenylalanine can be selectively incorporated into human Grb2 protein in mammalian Chinese hamster ovary cells and cross-linked with the epidermal growth factor receptor upon exposure of cells to 365-nm light (39). These amino acids should be useful for in vitro and in vivo probes of protein-protein and protein–nucleic acid interactions, studies of protein structure and dynamics, and the identification of receptors for orphan ligands.

Photocaged cysteine, serine, and tyrosine amino acids **10–12** have also been site-specifically introduced into proteins using this

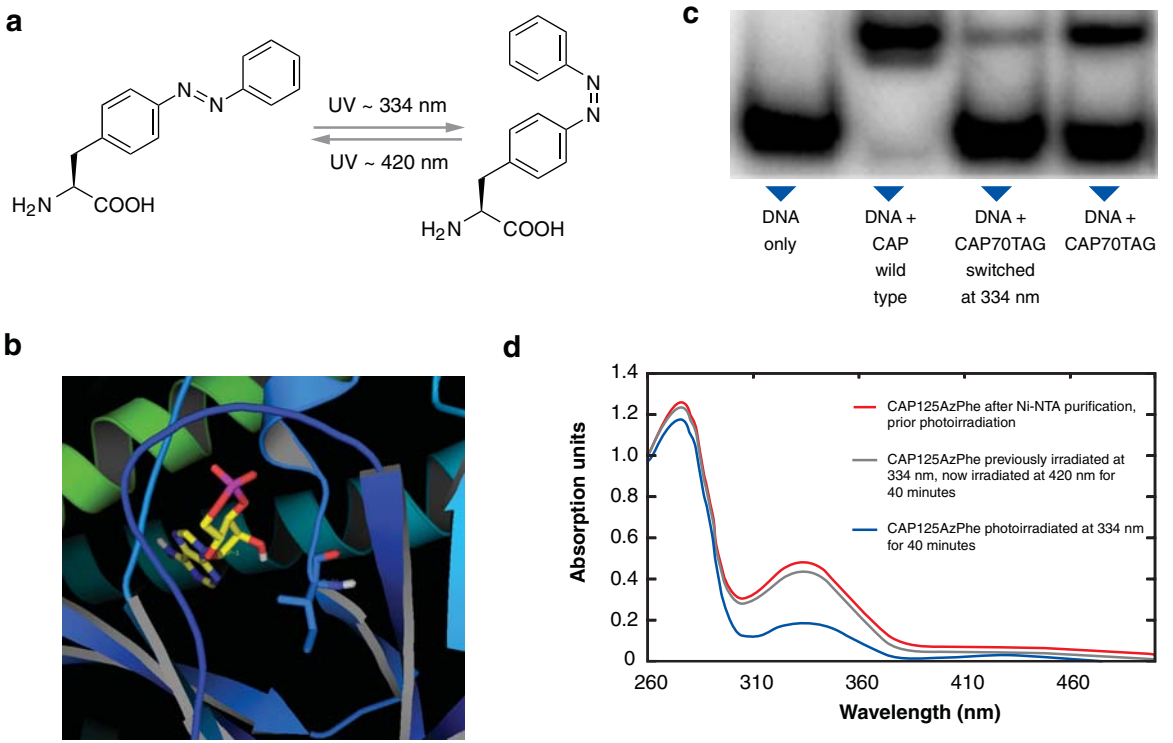

Figure 8

Photomodulation of mutant *E. coli* transcriptional factor catabolite activator protein (CAP)–DNA interactions. (*a*) The photochromic amino acid *p*-azophenyl-phenylalanine. (*b*) Ile-70 in the cAMP binding site of CAP was mutated to *p*-azophenyl-phenylalanine. (*c*) The affinity of mutant CAP for its promoter was regulated by photo-isomerizing *trans p*-azophenyl-phenylalanine to the *cis* isomer. (*d*) Absorption spectra of the mutant CAP (CAP125TAG or CAP125AzPhe; 50 μM) after purification and upon irradiation.

methodology (86). The side chain hydroxy or thiol groups of these amino acids are blocked by nitrobenzyl groups, which can be photochemically removed either in vivo or in vitro upon irradiation with 330 nm light. In one example the active-site cysteine of the proapoptotic cysteine protease caspase 3 was substituted with nitrobenzyl cysteine to afford inactive protein, which could be photoactivated with greater than 70% efficiency upon photolysis. In a related experiment the photochromic amino acid, *p*-azophenylphenylalanine **13**, was site-specifically introduced into the cAMP binding site of the *E. coli* transcription factor catabolite activator protein (CAP) (10). Irradiation of this amino acid at 334-nm light converts the *trans* form of the amino acid to predominantly the *cis* form. Because the two isomers differ significantly in structure and dipole, the *cis* and *trans* **13** mutants have different affinities for cAMP, and as a result, the affinity of this mutant CAP for its promoter can be photoregulated (**Figure 8**). Similarly, it should be possible to photomodulate the activity of other enzymes (e.g., kinase, phosphatases) as well as receptors, ion channels, and transcription factors either in vitro or in vivo.

CAP: transcription factor catabolite activator protein

This methodology has also been used to incorporate glycosylated amino acids into proteins in *E. coli* to produce homogeneous glycoproteins (**Figure 9**). Protein

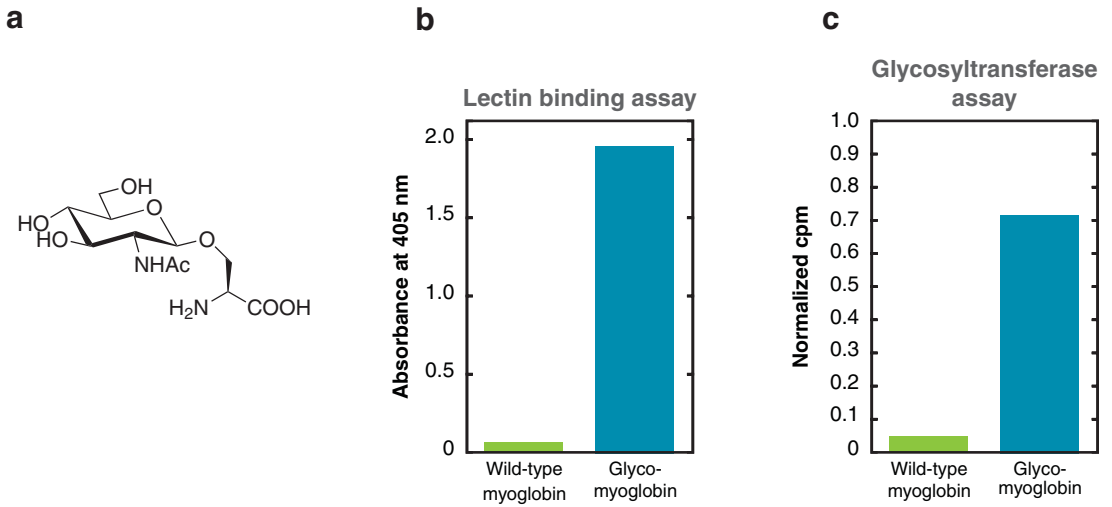

Figure 9

Cotranslational incorporation of a glycosylated amino acid into proteins in *E. coli*. (*a*) The chemical structure of β-*N*-acetylglucosamine-*O*-serine (β-GlcNAc-Ser). (*b*) Lectin binding assay with wild-type myoglobin and mutant myoglobin containing β-GlcNAc-Ser. (*c*) Analysis of the galactosyltransferase reaction with wild-type myoglobin and mutant myoglobin containing β-GlcNAc-Ser. UDP-[H^3]-galactose was used as the glycosyl donor.

glycosylation is an essential posttranslational modification in eukaryotic cells (28), but existing methods for generating glycoproteins often impose restrictions on the size, quantity, and/or purity of the glycoproteins produced (20, 37, 70, 78). To provide a general solution to this problem, mutant synthetases were evolved that site-specifically incorporate β-*N*-acetylglucosamine-*O*-serine (β-GlcNAc-Ser) **14** into proteins in *E. coli* with high translational fidelity (92). An *O*-GlcNAc-Ser mutant myoglobin generated by this method could be elaborated further to form more complex saccharides with galactosyltransferases (**Figure 9*c***). A similar approach was used to selectively introduce α-GalNAc-Thr **15** into proteins (89), and it is currently being extended to a number of other *O*- and *N*-linked sugars. This methodology should make it possible to genetically encode other posttranslationally modified amino acids including methylated and acetylated lysine, and phosphorylated tyrosine, serine, and threonine. Indeed, *p*-carboxylmethyl-L-phenylalanine **16** has recently been successfully introduced into proteins in *E. coli* as a stable mimetic of phosphotyrosine (P. G. Schultz, unpublished results).

Unnatural amino acids that can serve as probes of protein structure and function both in vivo and in vitro have also been genetically encoded in *E. coli* and yeast. For example, fluorescent amino acids with 7-hydroxycoumarin **17** and dansyl side chains **18** have been selectively incorporated into proteins, providing small fluorescent probes for the direct visualization of protein conformational changes, localization, and intermolecular interactions. These amino acids may offer an advantage over GFP and its derivatives by virtue of their smaller size and the fact that they can be introduced throughout a protein at any site. For example, the dansyl containing amino acid **18** was used to spectroscopically monitor the unfolding of superoxide dismutase. A combination of fluorescent amino acids and/or GFPs may also facilitate

FRET measurements in vitro or in vivo. Indeed, nitrophenylalanine **19** can be selectively introduced into proteins and used to quench tryptophan fluorescence. In another example the heavy-atom-containing amino acid *p*-iodo-L-phenylalanine **20** was genetically encoded in both *E. coli* and yeast and used for single-wavelength anomalous dispersion phasing in protein crystallography (with a laboratory CuKα X-ray source). This amino acid caused little structural perturbation when substituted for Phe in the core of T4 lysozyme (88). In yet another example, ^{15}N-labeled *O*-methyltyrosine **21** has been selectively incorporated into proteins as a site-selective NMR probe (24). Finally, *p*-cyanophenylalanine **22** was recently successfully incorporated into proteins (P. G. Schultz, unpublished results). Because the nitrile group has a distinct vibrational frequency, it should be a useful infrared probe of local environment and protein dynamics.

Another novel amino acid that has been incorporated into proteins is dihydroxyphenylalanine (DHP) **23** (1). DHP can undergo 2-electron oxidation to the corresponding quinone and can be used to both probe and manipulate electron transfer processes in proteins. Similarly, a second redox-active amino acid, 3-amino-L-tyrosine **24**, was selectively incorporated into proteins in *E. coli*. This amino acid can act as a radical trap owing to the stability of its oxidized semiquinone form or can serve as a unique handle for chemical modification of proteins (41). The redox-active amino acid 5-HTTP **25** has also been selectively incorporated into proteins in mammalian cells in response to an opal codon (91). This amino acid undergoes electrochemical oxidation to form an efficient protein cross-linking agent. Unnatural amino acids may also be used to perturb the electronic properties of a protein. For example, introduction of several unnatural amino acids in place of Tyr-65 in GFP led to changes in absorbance and fluorescence spectra and altered quantum yields (84).

In another series of experiments, a bipyridyl amino acid **26** has been selectively incorporated into proteins (P. G. Schultz, unpublished results). In this case we could not evolve a synthetase specific for **26** directly, but rather first evolved a synthetase specific for the biphenyl analogue **27**. Generation of a second-generation library from the latter in conjunction with subsequent rounds of positive and negative selections afforded a synthetase specific for **26**. This amino acid can be used to (*a*) introduce redox-active or electrophilic metal ions into proteins, (*b*) form fluorescent metal ion complexes, or (*c*) mediate the metal-ion-dependent dimerization of proteins containing **26**. Other representative unnatural amino acids that have been added to the genetic code of *E. coli* or yeast include L-3-(2-naphthyl)alanine **28** (80), α-aminocaprylic acid **29** (86), and L-homoglutamine **30** (4). The large number of structurally diverse amino acids that have been genetically encoded to date suggest that the translational machinery is rather tolerant of side chain structure. Thus, it should be possible to add additional building blocks to the genetic code including spin labels, electron transfer mediators, infrared and near-infrared probes, α-hydroxy acids, and *N*-alkyl amino acids.

STRUCTURAL STUDIES OF MUTANT SYNTHETASES

We have been surprisingly successful in identifying mutant aminoacyl-tRNA synthetases with altered specificities (79, 83, 87). Crystallographic studies of these mutant synthetases provide an opportunity to examine the origins and evolution of amino acid specificity in this important class of enzymes. For example, in the case of an *Mj* TyrRS mutant specific for *p*AcPhe (**Figure 10*a***), there are significant changes in the active site that lead to altered hydrogen bonding and packing interactions that favor binding of the unnatural amino acid relative to tyrosine (77). Active-site D158G and Y32L mutations remove two hydrogen bonds to the tyrosine side chain hydroxyl group, which would be expected to

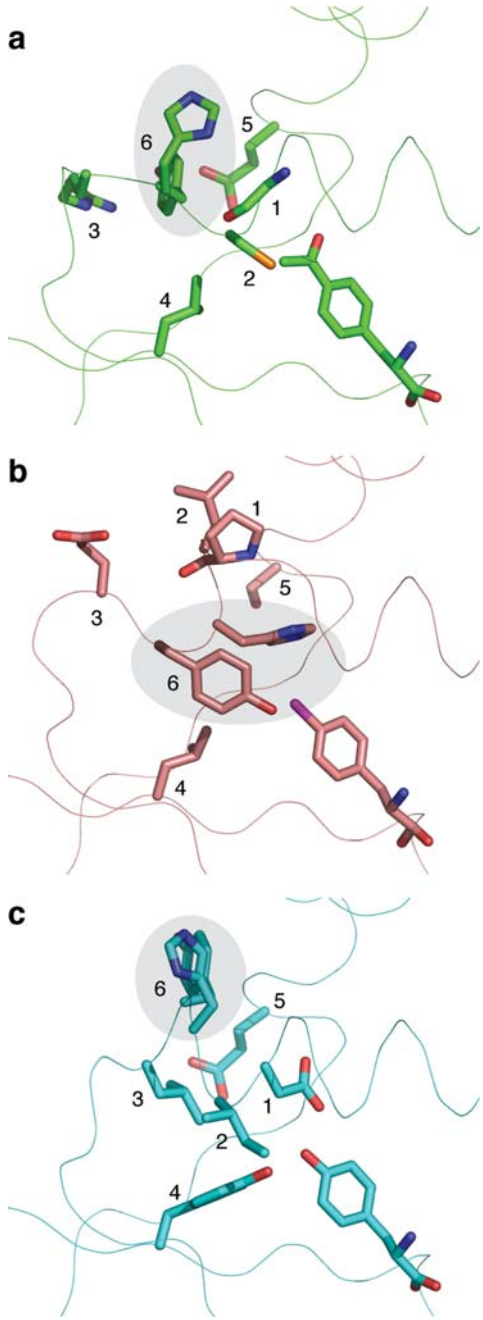

Figure 10

Structure of wild-type and mutant *Mj* TyrRSs bound to cognate amino acids. (*a*) *p*AcPhe synthetase bound to *p*-acetylphenylalanine (*p*AcPhe). (*b*) *p*BrPhe synthetase bound to *p*-bromophenylalanine. (*c*) Wild-type *Mj* TyrRS bound to tyrosine. 1: residue 158; 2: residue 159; 3: residue 162; 4: residue 32; 5: residue 107; 6: residues 160 and 161.

dramatically reduce binding of the natural substrate to the enzyme. The D158G mutation also deepens the binding pocket to accommodate the *para* substituent of *p*AcPhe, while Y32L forms a suitable hydrophobic packing surface for the acetyl methyl group. In addition, the side chain carbonyl oxygen of *p*AcPhe forms a hydrogen bond to Gln-109. In the wild-type *Mj* TyrRS structure (49, 90), the side chain of Asp-158 disrupts this hydrogen bond vector; in the *p*AcPhe-specific synthetase the D158G mutation removes this intervening side chain. The D158G mutation also alters the backbone structure of helix α8. The glycine residue truncates the C terminus of helix α8 by four residues, resulting in a new C-terminal cap. Several intrahelical main chain hydrogen bonds are broken by premature termination of helix α8; however, a short 3_{10}-helix formed in the new structure is stabilized by side chain/main chain and main chain/main chain hydrogen bonds. This altered helix α8 conformation facilitates reconfiguration of active side chains to selectively bind *p*AcPhe.

Similar side chain and backbone conformational changes were observed in the structure of a mutant *Mj* TyrRS specific for *p*BrPhe (which has the mutations Y32L, E107S, D158P, I159L, and L162E) (P. G. Schultz, unpublished results) (**Figure 10*b***). The D158P and Y32L mutations remove two hydrogen bonds to the hydroxyl oxygen of tyrosine. The Y32L mutation also increases the size of the active-site cavity and results in van der Waals interactions between the alkyl side chain and the bromine atom of the bound *p*BrPhe. The D158P mutation also terminates helix α8 and, as a result, induces formation of a subsequent short 3_{10}-helix by residues 157–161, which again is stabilized by side chain and main chain hydrogen bonds. As a result of the D158P mutation and the resulting 3_{10}-helix, significant translational and rotational movements of several active residues expand the active-site cavity [compared with the wild-type structure (49, 90)] to form new van der Waals interactions with the *p*BrPhe side chain. Preliminary results in the case of an *Mj* TyrRS mutant specific for naphthylalanine show that similar backbone conformational changes again reconfigure the active site to bind the unnatural amino acid.

In both of these mutant enzymes, a handful of random mutations alter the substrate specificity of the *Mj* TyrRS by changing the pattern of hydrogen bonding and packing interactions with bound substrate. These mutations lead to changes in both side chain and backbone conformations, indicating a high degree of structural plasticity in the active site of the enzyme. The structural differences between these mutant and wild-type TyrRS are more pronounced than those found in previously described mutant aminoacyl-tRNA synthetases that recognize unnatural or unusual amino acids (12, 50, 51). Indeed, mutations that alter enzyme specificity generally have little effect on backbone configuration. The structural changes observed in these mutant synthetases are likely due to the sites of mutation and the method used to generate the libraries. Structural changes in the solvent-exposed α8 helix minimize the number of required compensatory structural changes on the exposed face, as bulk water can easily reorganize to solvate the exposed residues. In addition, the use of a focused library of mutated residues in close proximity to each other increases the likelihood of selecting compensatory mutations. In contrast, mutations that arise from DNA shuffling are recombined from independent selection experiments, and the resulting set of mutations may not be cooperative. Random-point mutations also have a low probability of being cooperative owing to the limited size of sequence space that can be explored.

CONCLUSIONS AND PERSPECTIVES

The approach described above has proven remarkably effective in allowing us to add a large number of structurally diverse amino acids to the genetic codes of both prokaryotic

and eukaryotic organisms. Coincidently, it has recently been shown that nature has evolved a similar strategy using an orthogonal amber suppressor tRNA/synthetase pair derived from lysine to genetically encode the unnatural amino acid pyrrolysine in *Methanosarcina barkeri* (72). Future work in this field will likely focus on expanding the nature and number of amino acids that can be genetically encoded in both prokaryotic and eukaryotic organisms, including multicellular organisms. Additional orthogonal pairs that suppress three- and four-base codons are also being developed. It may even be possible to delete rare redundant codons from the *E. coli* genome and use them instead to encode unnatural amino acids. Considerable improvements in the yields of mutant proteins have been realized in *E. coli* and yeast, but further modifications to the structures of the tRNAs and synthetases, optimization of expression levels, and genomic mutations to the host organism (e.g., ribosome, transporters) are likely to further increase yields, especially in mammalian systems.

The ability to genetically encode unnatural amino acids should provide powerful probes, both in vitro and in vivo, of protein structure and function. It may also allow the design or evolution of proteins with novel properties. Examples might include the rational design of glycosylated or PEGylated therapeutic proteins with improved pharmacological properties, fluorescent proteins that act as sensors of small molecules and protein-protein interactions in the cell, or proteins whose activity can be photoregulated in vivo. One may be able to select for peptides and proteins with enhanced function from libraries of unnatural amino acid mutants. For example, we recently showed that it is possible to incorporate unnatural amino acids into phage-displayed peptides (75); a peptide with enhanced affinity for streptavidin was isolated that contained an unnatural amino acid. It should also be possible to incorporate nonamino acid building blocks into proteins or perhaps even create biopolymers with entirely unnatural backbones. Finally, the ability to add novel amino acids to the genetic codes of organisms should allow us to test experimentally whether there is an evolutionary advantage for organisms with more than the 20 genetically encoded amino acids.

SUMMARY POINTS

1. A general method was developed that makes it possible to genetically encode unnatural amino acids in *E. coli*, yeast, and mammalian cells.
2. More than 30 unnatural amino acids have been cotranslationally incorporated into proteins with high fidelity and efficiency by means of unique nonsense (triplet) or frameshift (quadruplet) codons and the corresponding tRNA/aminoacyl-tRNA synthetase pairs.
3. Unnatural amino acids that have been genetically encoded include those containing spectroscopic probes, posttranslational modifications, metal chelators, photoaffinity labels, functional groups with unique reactivity, and other chemical moieties.
4. Orthogonal tRNA/aminoacyl-tRNA synthetase pairs were derived from heterologous pairs with distinct tRNA identity elements.
5. The amino acid substrate specificity of the orthogonal aminoacyl-tRNA synthetase was altered by alternating positive and negative genetic selections using a large library of active-site mutants of the synthetase.

6. The ability to incorporate amino acids with defined steric and electronic properties at unique sites into proteins will provide powerful new tools for exploring protein structure and function both in vitro and in vivo, and for generating proteins with novel properties.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant GM-62159 and Department of Energy grant DE-FG03-00ER46051 to P.G.S.

LITERATURE CITED

1. Alfonta L, Zhang Z, Uryu S, Loo JA, Schultz PG. 2003. Site-specific incorporation of a redox-active amino acid into proteins. *J. Am. Chem. Soc.* 125:14662–63
2. Anderson JC, Magliery TJ, Schultz PG. 2002. Exploring the limits of codon and anticodon size. *Chem. Biol.* 9:237–44
3. Anderson JC, Schultz PG. 2003. Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry* 42:9598–608
4. **Anderson JC, Wu N, Santoro SW, Lakshman V, King DS, Schultz PG. 2004. An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. USA* 101:7566–71**
5. Arts GJ, Kuersten S, Romby P, Ehresmann B, Mattaj IW. 1998. The role of exportin-t in selective nuclear export of mature tRNAs. *EMBO J.* 17:7430–41
6. Bain JD, Glabe CG, Dix TA, Chamberlin AR, Diala ES. 1989. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* 111:8013–14
7. Beene DL, Dougherty DA, Lester HA. 2003. Unnatural amino acid mutagenesis in mapping ion channel function. *Curr. Opin. Neurobiol.* 13:264–70
8. Benzer S, Champe SP. 1962. A change from nonsense to sense in the genetic code. *Proc. Natl. Acad. Sci. USA* 48:1114–21
9. Bock A, Forchhammer K, Heider J, Leinfelder W, Sawers G, et al. 1991. Selenocysteine: the 21st amino acid. *Mol. Microbiol.* 5:515–20
10. Bose M, Groff D, Xie J, Brustad E, Schultz PG. 2006. The incorporation of a photoisomerizable amino acid into proteins in *E. coli*. *J. Am. Chem. Soc.* 128:388–89
11. Bossi L, Roth JR. 1981. Four-base codons ACCA, ACCU and ACCC are recognized by frameshift suppressor sufJ. *Cell* 25:489–96
12. Buddha MR, Crane BR. 2005. Structure and activity of an aminoacyl-tRNA synthetase that charges tRNA with nitro-tryptophan. *Nat. Struct. Mol. Biol.* 12:274–75
13. Carne AF. 1994. Chemical modification of proteins. *Methods Mol. Biol.* 32:311–20
14. **Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, Schultz PG. 2003. An expanded eukaryotic genetic code. *Science* 301:964–67**
15. **Chin JW, Martin AB, King DS, Wang L, Schultz PG. 2002. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 99:11020–24**
16. Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG. 2002. Addition of *p*-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc.* 124:9026–27

4. A combination of amber (TAG) and frameshift (AGGA) suppression was used to incorporate two unnatural amino acids into a protein simultaneously with high fidelity.

14. This report expands this methodology to the genetic code of yeast. Five unnatural amino acids were genetically encoded in yeast in response to the amber nonsense codon.

15. A photocrosslinking amino acid, *p*Bpa, was incorporated into proteins in *E. coli*. It should be useful in probing protein-protein and protein-nucleic acid interactions both in vitro and in vivo.

17. Chin JW, Schultz PG. 2002. In vivo photocrosslinking with unnatural amino acid mutagenesis. *Chembiochem* 3:1135–37
18. Cornish VW, Mendel D, Schultz PG. 1995. Probing protein structure and function with an expanded genetic code. *Angew Chem. Int. Ed. Engl.* 34:621–33
19. Curran JF, Yarus M. 1987. Reading frame selection and transfer RNA anticodon loop stacking. *Science* 238:1545–50
20. Davis BG. 2002. Synthesis of glycoproteins. *Chem. Rev.* 102:579–602
21. Dawson PE, Kent SB. 2000. Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* 69:923–60
22. Deiters A, Cropp TA, Mukherji M, Chin JW, Anderson JC, Schultz PG. 2003. Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*. *J. Am. Chem. Soc.* 125:11782–83
23. Deiters A, Cropp TA, Summerer D, Mukherji M, Schultz PG. 2004. Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorg. Med. Chem. Lett.* 14:5743–45
24. Deiters A, Geierstanger BH, Schultz PG. 2005. Site-specific in vivo labeling of proteins for NMR studies. *Chembiochem* 6:55–58
25. Doctor BP, Mudd JA. 1963. Species specificity of amino acid acceptor ribonucleic acid and aminoacyl soluble ribonucleic acid synthetases. *J. Biol. Chem.* 238:3677–81
26. Doring V, Mootz HD, Nangle LA, Hendrickson TL, de Crecy-Lagard V, et al. 2001. Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway. *Science* 292:501–4
27. Drabkin HJ, Park HJ, RajBhandary UL. 1996. Amber suppression in mammalian cells dependent upon expression of an *Escherichia coli* aminoacyl-tRNA synthetase gene. *Mol. Cell. Biol.* 16:907–13
28. Dwek RA. 1996. Glycobiology: toward understanding the function of sugars. *Chem. Rev.* 96:683–720
29. Edwards H, Schimmel P. 1990. A bacterial amber suppressor in *Saccharomyces cerevisiae* is selectively recognized by a bacterial aminoacyl-tRNA synthetase. *Mol. Cell. Biol.* 10:1633–41
30. Edwards H, Trezeguet V, Schimmel P. 1991. An *Escherichia coli* tyrosine transfer RNA is a leucine-specific transfer RNA in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 88:1153–56
31. Evans TC Jr, Xu MQ. 1999. Intein-mediated protein ligation: harnessing nature's escape artists. *Biopolymers* 51:333–42
32. Fechter P, Rudinger-Thirion J, Tukalo M, Giege R. 2001. Major tyrosine identity determinants in *Methanococcus jannaschii* and *Saccharomyces cerevisiae* tRNA(Tyr) are conserved but expressed differently. *Eur. J. Biochem.* 268:761–67
33. Galardy RE, Craig LC, Printz MP. 1973. Benzophenone triplet: a new photochemical probe of biological ligand-receptor interactions. *Nat. New Biol.* 242:127–28
34. Galli G, Hofstetter H, Birnstiel ML. 1981. Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature* 294:626–31
35. Garen A, Siddiqi O. 1962. Suppression of mutations in the alkaline phosphatase structural cistron of *E. coli*. *Proc. Natl. Acad. Sci. USA* 48:1121–27
36. Giriat I, Muir TW. 2003. Protein semi-synthesis in living cells. *J. Am. Chem. Soc.* 125:7180–81
37. Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, et al. 2003. Production of complex human glycoproteins in yeast. *Science* 301:1244–46

38. Hendrickson WA, Horton JR, LeMaster DM. 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* 9:1665–72
39. Hino N, Okazaki Y, Kobayashi T, Hayashi A, Sakamoto K, Yokoyama S. 2005. Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat. Methods* 2:201–6
40. Hohsaka T, Kajihara D, Ashizuka Y, Murakami H, Sisido M. 1999. Efficient incorporation of nonnatural amino acids with large aromatic groups into streptavidin in in vitro protein synthesizing systems. *J. Am. Chem. Soc.* 121:34–40
41. Hooker JM, Kovacs EW, Francis MB. 2004. Interior surface modification of bacteriophage MS2. *J. Am. Chem. Soc.* 126:3718–19
42. Hortin G, Boime I. 1983. Applications of amino acid analogs for studying co- and post-translational modifications of proteins. *Methods Enzymol.* 96:777–84
43. Hudziak RM, Laski FA, RajBhandary UL, Sharp PA, Capecchi MR. 1982. Establishment of mammalian cell lines containing multiple nonsense mutations and functional suppressor tRNA genes. *Cell* 31:137–46
44. Ibba M, Hennecke H. 1995. Relaxing the substrate specificity of an aminoacyl-tRNA synthetase allows in vitro and in vivo synthesis of proteins containing unnatural amino acids. *FEBS Lett.* 364:272–75
45. Ibba M, Soll D. 2000. Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* 69:617–50
46. Kent SB. 1988. Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem.* 57:957–89
47. Kiga D, Sakamoto K, Kodama K, Kigawa T, Matsuda T, et al. 2002. An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. *Proc. Natl. Acad. Sci. USA* 99:9715–20
48. Kim D, Johnson J. 1988. Construction, expression, and function of a new yeast amber suppressor, tRNATrpA. *J. Biol. Chem.* 263:7316–21
49. Kobayashi T, Nureki O, Ishitani R, Yaremchuk A, Tukalo M, et al. 2003. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol.* 10:425–32
50. Kobayashi T, Sakamoto K, Takimura T, Sekine R, Kelly VP, et al. 2005. Structural basis of nonnatural amino acid recognition by an engineered aminoacyl-tRNA synthetase for genetic code expansion. *Proc. Natl. Acad. Sci. USA* 102:1366–71
51. Kobayashi T, Takimura T, Sekine R, Vincent K, Kamata K, et al. 2005. Structural snapshots of the KMSKS loop rearrangement for amino acid activation by bacterial tyrosyl-tRNA synthetase. *J. Mol. Biol.* 346:105–17
52. Kowal AK, Kohrer C, RajBhandary UL. 2001. Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. *Proc. Natl. Acad. Sci. USA* 98:2268–73
53. Kwok Y, Wong JT. 1980. Evolutionary relationship between *Halobacterium cutirubrum* and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes. *Can. J. Biochem.* 58:213–18
54. Li F, Xiong Y, Wang J, Cho HD, Tomita K, et al. 2002. Crystal structures of the *Bacillus stearothermophilus* CCA-adding enzyme and its complexes with ATP or CTP. *Cell* 111:815–24
55. Liu DR, Magliery TJ, Pastrnak M, Schultz PG. 1997. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci. USA* 94:10092–97

39. A photocrosslinking amino acid was incorporated into proteins in mammalian cells.

56. Liu DR, Magliery TJ, Schultz PG. 1997. Characterization of an 'orthogonal' suppressor tRNA derived from *E. coli* tRNA2(Gln). *Chem. Biol.* 4:685–91
57. Liu DR, Schultz PG. 1999. Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. USA* 96:4780–85
58. Liu H, Wang L, Brock A, Wong CH, Schultz PG. 2003. A method for the generation of glycoprotein mimetics. *J. Am. Chem. Soc.* 125:1702–3
59. Magliery TJ, Anderson JC, Schultz PG. 2001. Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in *Escherichia coli*. *J. Mol. Biol.* 307:755–69
60. McIntosh LP, Dahlquist FW. 1990. Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q. Rev. Biophys.* 23:1–38
61. Mehl RA, Anderson JC, Santoro SW, Wang L, Martin AB, et al. 2003. Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* 125:935–39
62. Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG. 1989. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* 244:182–88
63. Normanly J, Kleina LG, Masson JM, Abelson J, Miller JH. 1990. Construction of *Escherichia coli* amber suppressor tRNA genes. III. Determination of tRNA specificity. *J. Mol. Biol.* 213:719–26
64. O'Connor M. 2002. Insertions in the anticodon loop of tRNA1Gln(sufG) and tRNA(Lys) promote quadruplet decoding of CAAA. *Nucleic Acids Res.* 30:1985–90
65. Pastrnak M, Magliery TJ, Schultz PG. 2000. A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code. *Helv. Chim. Acta* 83:2277–86
66. Pastrnak M, Schultz PG. 2001. Phage selection for site-specific incorporation of unnatural amino acids into proteins in vivo. *Bioorg. Med. Chem.* 9:2373–79
67. Sakamoto K, Hayashi A, Sakamoto A, Kiga D, Nakayama H, et al. 2002. Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* 30:4692–99
68. Santoro SW, Anderson JC, Lakshman V, Schultz PG. 2003. An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in *Escherichia coli*. *Nucleic Acids Res.* 31:6700–9
69. Santoro SW, Wang L, Herberich B, King DS, Schultz PG. 2002. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* 20:1044–50
70. Sears P, Wong CH. 2001. Toward automated synthesis of oligosaccharides and glycoproteins. *Science* 291:2344–50
71. Sherman F. 1982. Suppression in the yeast *Saccharomyces cerevisiae*. In *The Molecular Biology of the Yeast* Saccharomyces: *Metabolism and Gene Expression*, ed. JN Strathern, EW Jones, JR Broach, pp. 463–86. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory
72. Srinivasan G, James CM, Krzycki JA. 2002. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* 296:1459–62
73. Steer BA, Schimmel P. 1999. Major anticodon-binding region missing from an archaebacterial tRNA synthetase. *J. Biol. Chem.* 274:35601–6
74. Tang Y, Tirrell DA. 2002. Attenuation of the editing activity of the *Escherichia coli* leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry* 41:10635–45
75. Tian F, Tsao ML, Schultz PG. 2004. A phage display system with unnatural amino acids. *J. Am. Chem. Soc.* 126:15962–63

61. Describes the generation of a completely autonomous bacterium with a 21-amino-acid genetic code that can biosynthesize a nonstandard amino acid (*p*AF) and selectively incorporate it into proteins.

77. X-ray crystallographic studies of a mutant aminoacyl-tRNA synthetase showed significant structural plasticity in the amino acid side chain and protein backbone conformations in the amino acid binding site.

76. Tsao M, Tian F, Schultz PG. 2005. The Staudinger modification of proteins containing azido amino acids. *Chembiochem* 6:2147–49
77. Turner JM, Graziano J, Spraggon G, Schultz PG. 2005. Structural characterization of a *p*-acetylphenylalanyl aminoacyl-tRNA synthetase. *J. Am. Chem. Soc.* 127:14976–77
78. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, et al. 2002. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* 298:1790–3
79. Wang L, Brock A, Herberich B, Schultz PG. 2001. Expanding the genetic code of *Escherichia coli*. *Science* 292:498–500
80. Wang L, Brock A, Schultz PG. 2002. Adding L-3-(2-Naphthyl)alanine to the genetic code of *E. coli*. *J. Am. Chem. Soc.* 124:1836–37
81. Wang L, Magliery TJ, Liu DR, Schultz PG. 2000. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. *J. Am. Chem. Soc.* 122:5010–11
82. Wang L, Schultz PG. 2001. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* 8:883–90
83. Wang L, Schultz PG. 2004. Expanding the genetic code. *Angew Chem. Int. Ed. Engl.* 44:34–66
84. Wang L, Xie J, Deniz AA, Schultz PG. 2003. Unnatural amino acid mutagenesis of green fluorescent protein. *J. Org. Chem.* 68:174–76
85. Wang L, Zhang Z, Brock A, Schultz PG. 2003. Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 100:56–61
86. Wu N, Deiters A, Cropp TA, King D, Schultz PG. 2004. A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* 126:14306–7
87. Xie J, Schultz PG. 2005. An expanding genetic code. *Methods* 36:227–38
88. Xie J, Wang L, Wu N, Brock A, Spraggon G, Schultz PG. 2004. The site-specific incorporation of *p*-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* 22:1297–301
89. Xu R, Hanson SR, Zhang Z, Yang YY, Schultz PG, Wong CH. 2004. Site-specific incorporation of the mucin-type *N*-acetylgalactosamine-alpha-*O*-threonine into protein in *Escherichia coli*. *J. Am. Chem. Soc.* 126:15654–55
90. Zhang Y, Wang L, Schultz PG, Wilson IA. 2005. Crystal structures of apo wild-type *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for *O*-methyl-L-tyrosine. *Protein Sci.* 14:1340–49
91. Zhang Z, Alfonta L, Tian F, Bursulaya B, Uryu S, et al. 2004. Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc. Natl. Acad. Sci. USA* 101:8882–87
92. Zhang Z, Gildersleeve J, Yang YY, Xu R, Loo JA, et al. 2004. A new strategy for the synthesis of glycoproteins. *Science* 303:371–73
93. Zhang Z, Smith BA, Wang L, Brock A, Cho C, Schultz PG. 2003. A new strategy for the site-specific modification of proteins in vivo. *Biochemistry* 42:6735–46

79. The first report that an unnatural amino acid (OMeTyr) could be successfully added to the genetic code of *E. coli* by means of amber-suppression and an engineered orthogonal *M. jannaschii* TyrRS-tRNA pair.

85. A keto amino acid, *p*AcPhe, was incorporated into proteins. The unique reactivity of the introduced keto group allows the selective modification of proteins.

88. A heavy-atom-containing amino acid, iodophenylalanine, was selectively incorporated into proteins, affording a reliable method to prepare iodinated proteins to facilitate SAD phasing in crystallography.

92. The incorporation of a glycosylated amino acid directly into proteins provides a promising tool for the synthesis of homogenous glycoproteins.

Annual Review of Biophysics and Biomolecular Structure

Volume 35, 2006

Contents