Expanding the genetic code

Lei Wang[a] and Peter G. Schultz*[b]

[a] Department of Chemistry, University of California at Berkeley, Berkeley, CA 94720 USA
[b] Department of Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037 USA. E-mail: schultz@scripps.edu

The ability to incorporate unnatural amino acids into proteins directly in living cells will provide new tools to study protein and cellular function, and may generate proteins or even organisms with enhanced properties. Due to the limited promiscuity of some synthetases, natural amino acids can be substituted with close analogs at multiple sites using auxotrophic strains. Alternatively, this can be achieved by deactivating the editing function of some synthetases. The addition of new amino acids to the genetic code, however, requires additional components of the protein biosynthetic machinery including a novel tRNA-codon pair, an aminoacyl-tRNA synthetase, and an amino acid. This new set of components functions orthogonally to the counterparts of the common 20 amino acids, *i.e.*, the orthogonal synthetase (and only this synthetase) aminoacylates the orthogonal tRNA (and only this tRNA) with the unnatural amino acid only, and the resulting acylated tRNA inserts the unnatural amino acid only in response to the unique codon. Using this strategy, the genetic code of *Escherichia coli* has been expanded to incorporate unnatural amino acids with a fidelity rivaling that of natural amino acids. This methodology is being applied to other cell types and unnatural analogs with a variety of functionalities.

Professor Peter G. Schultz did his undergraduate and graduate work at the California Institute of Technology. His thesis work resulted in the first synthetic molecules that sequence-selectively cleave DNA. In 1985, after postdoctoral studies at the Massachusetts Institute of Technology, he joined the faculty of the University of California at Berkeley, where he was Professor of Chemistry, a Principal Investigator at Lawrence Berkeley National Laboratory and an Investigator of the Howard Hughes Medical Institute. His research—which spans the interface of biology, chemistry, and materials science — includes (a) the development of catalytic antibodies; (b) the application of molecular diversity to problems in biomolecular recognition and catalysis, drug discovery, and materials science; (c) the development of methods for incorporating unnatural amino acids selectively into proteins; (d) single-molecule biological imaging; and most recently (e) studies in functional genomics. He was a founding scientist of Affymax Research Institute and is a cofounder of Symyx Technologies, Syrrx, Phenomix and Kalypsys. His awards include the Waterman Award of the National Science Foundation, membership in the National Academy of Sciences, and the 1994 Wolf Prize in Chemistry. He is currently a Professor of Chemistry at the Scripps Research Institute, and the director of the new Genomics Institute of Novartis Research Foundation in San Diego.

Lei Wang received his B. S. in organic chemistry and M. S. in physical chemistry from Peking University in 1994 and 1997 respectively, where he conducted research on nanoparticle properties using Scanning Probe Microscopy in Center for Intelligent Material Research. Presently he is a graduate student in the Department of Chemistry at the University of California, Berkeley, under the guidance of Professor Peter G. Schultz, and is investigating in vivo *methods for the site-specific incorporation of unnatural amino acids into proteins.*

Introduction

Proteins carry out virtually all of the complex processes of life, from photosynthesis to signal transduction and the immune response. To understand and control these intricate activities, we need to better understand the relationship between the structure and function of proteins. Both site-directed and random mutagenesis, in which specific amino acids in a protein can be replaced with any of the other nineteen common amino acids, have become important tools for this purpose. These methodologies have made possible the generation of proteins with enhanced properties including stability, catalytic activity and binding specificity. Nevertheless, changes in proteins are limited to the 20 common amino acids, most of which have simple functional groups. The ability to include unnatural amino acids with various sizes, acidities, nucleophilicities, hydrophobicities, and other properties into proteins would greatly expand our ability to rationally and systematically manipulate the structures of proteins, both to probe protein function and create new proteins with novel properties.

Several strategies have been employed to introduce unnatural amino acids into proteins. The first experiments involved the derivatization of amino acids with reactive side-chains such as Lys, Cys and Tyr, for example, the conversion of lysine to N^ε-acetyllysine. Chemical synthesis also provides a straightforward method to incorporate unnatural amino acids, but routine solid-phase peptide synthesis is generally limited to small peptides or proteins with less than 100 residues. With the recent development of enzymatic ligation and native chemical ligation of peptide fragments, it is possible now to make larger proteins.[1] A general *in vitro* biosynthetic method, in which a suppressor tRNA chemically acylated with the desired unnatural amino acid is added to an *in vitro* extract capable of supporting protein biosynthesis, has been used to site-specifically incorporate over 100 unnatural amino acids into a variety of proteins of virtually any size.[2] A broad range of functional groups have been introduced into proteins for studies of protein stability, protein folding, enzyme mechanism, and signal transduction. Although these studies demonstrate that the protein biosynthetic machinery tolerates a wide variety of amino acid side chains, the method is technically demanding, and yields of mutant proteins are low.

The ability to incorporate unnatural amino acids directly into proteins *in vivo* offers the advantages of high yields of mutant proteins, technical ease, and the potential to study the mutant proteins in cells or possibly in living organisms. *In vivo* approaches can be generally categorized into multisite substitu-

tion and site-specific incorporation. This article will give a brief review of different *in vivo* strategies, and focus on our efforts to develop a general *in vivo* site-specific mutagenesis method.

Multisite substitution

Over 50 years ago, it was found that many analogs of natural amino acids inhibit the growth of bacteria. Analysis of the proteins produced in the presence of these amino acid analogs revealed that they had been substituted for their natural counterparts to various extents.[3] This occurs because the aminoacyl-tRNA synthetase, the enzyme responsible for the attachment of the correct amino acid to its cognate tRNA, cannot rigorously distinguish the analog from the corresponding natural amino acid. For instance, norleucine is charged by methionyl-tRNA synthetase,[4] and *p*-fluorophenylalanine is charged by phenylalanine-tRNA synthetase.[5]

An *in vivo* method termed selective pressure incorporation[6] was later developed to exploit the promiscuity of wild-type synthetases. An auxotrophic strain, in which the relevant metabolic pathway supplying the cell with a particular natural amino acid is switched off, is grown in minimal media containing limited concentrations of the natural amino acid while transcription of the target gene is repressed. At the onset of stationary growth phase, the natural amino acid is depleted and replaced with the unnatural amino acid analog. Induction of expression of the recombinant protein results in the accumulation of a protein containing the unnatural analog. For example, using this strategy, *o*, *m* and *p*-fluorophenylalanines have been incorporated into proteins, and exhibit two characteristic shoulders in the UV spectrum which can be easily identified;[7] trifluoromethionine has been used to replace methionine in bacteriophage λ lysozyme to study its interaction with chit-ooligosaccharide ligands by ^{19}F NMR;[8] and trifluoroleucine has been inserted in place of leucine, resulting in increased thermal and chemical stability of a leucine-zipper protein.[9] Moreover, selenomethionine and telluromethionine are incorporated into various recombinant proteins to facilitate the solution of phases in X-ray crystallography.[10–13] Methionine analogs with alkene or alkyne functionalities have also been inserted efficiently, allowing for additional modification of proteins by chemical means.[14–16]

The success of this method depends on the recognition of the unnatural amino acid analogs by aminoacyl-tRNA synthetases, which in general require high selectivity to insure the fidelity of protein translation. Therefore, the range of chemical functionality accessible *via* this route is limited. For instance, although thiaproline can be incorporated quantitatively into proteins, oxaproline and selenaproline cannot.[17] One way to expand the scope of this method is to relax the substrate specificity of aminoacyl-tRNA synthetases, which has been achieved in a limited number of cases. For example, it was found that replacement of Ala294 by Gly in *E. coli* phenylalanyl-tRNA synthetase (PheRS) increases the size of substrate binding pocket, and results in the acylation of tRNA[Phe] by *p*-Cl-phenylalanine (*p*-Cl-Phe).[18] An *E. coli* strain harboring this mutant PheRS allows the incorporation of *p*-Cl-phenylalanine or *p*-Br-phenylalanine in place of phenylalanine.[19,20] Similarly, a point mutation Phe130Ser near the amino acid binding site of *E. coli* tyrosyl-tRNA synthetase was shown to allow azatyrosine to be incorporated more efficiently than tyrosine.[21]

The fidelity of aminoacylation is maintained both at the level of substrate discrimination and proofreading of non-cognate intermediates and products. Therefore, an alternative strategy to incorporate unnatural amino acids into proteins *in vivo* is to modify synthetases that have proofreading mechanisms. These synthetases cannot discriminate and therefore activate amino acids that are structurally similar to the cognate natural amino acids. This error is corrected at a separate site, which deacylates the mischarged amino acid from the tRNA to maintain the fidelity of protein translation. If the proofreading activity of the synthetase is disabled, structural analogs that are misactivated may escape the editing function and be incorporated. This approach has been demonstrated recently with the valyl-tRNA synthetase (ValRS).[22] ValRS can misaminoacylate tRNA[Val] with Cys, Thr, or aminobutyrate (Abu); these non-cognate amino acids are subsequently hydrolyzed by the editing domain. After random mutagenesis of the *E. coli* chromosome, a mutant *E. coli* strain was selected that has a mutation in the editing site of ValRS. This edit-defective ValRS incorrectly charges tRNA[Val] with Cys. Because Abu sterically resembles Cys (–SH group of Cys is replaced with –CH$_3$ in Abu), the mutant ValRS also incorporates Abu into proteins when this mutant *E. coli* strain is grown in the presence of Abu. Mass spectrometric analysis shows that about 24% of valine are replaced by Abu at each valine position in the native protein.

In general, *in vivo* multisite substitution methods are relatively simple to carry out, and can provide large quantities of engineered proteins. Proteins with modifications at multiple sites are useful for crystallographic studies and NMR determination of protein structure. However, a major limitation is that all sites corresponding to a particular natural amino acid throughout the protein are replaced. The extent of incorporation of the natural and unnatural amino acid may also vary—only in rare cases can quantitative substitution be achieved since it is difficult to completely deplete the cognate natural amino acid inside the cell. Another limitation is that these strategies make it difficult to study the mutant protein in living cells, because the multisite incorporation of analogs often results in toxicity. Finally, this method is applicable in general only to close structural analogs of the common amino acids, again because substitutions must be tolerated at all sites in the genome.

Microinjection of aminoacylated tRNAs into cells

The *in vivo* site-specific incorporation of unnatural amino acids into proteins was first realized by an extension of the *in vitro* biosynthetic methodology using microinjection techniques.[23,24] A *Xenopus* oocyte was coinjected with two RNA species made *in vitro*: an mRNA encoding the target protein with a UAG stop codon at the amino acid position of interest and an amber suppressor tRNA aminoacylated with the desired unnatural amino acid. The translational machinery of the oocyte then inserts the unnatural amino acid at the position specified by UAG. This method has allowed *in vivo* structure-function studies of integral membrane proteins, which are generally not amenable to *in vitro* expression systems. Examples include the incorporation of a fluorescent amino acid into tachykinin neurokinin-2 receptor to measure distances by fluorescence resonance energy transfer;[25] the incorporation of biotinylated amino acids to identify surface-exposed residues in ion channels;[26] the use of caged tyrosine analogs to monitor conformational changes in an ion channel in real time;[27] and the use of α-hydroxy amino acids to change ion channel backbones for probing their gating mechanisms.[28,29]

One limitation is inevitably inherited from the *in vitro* biosynthetic method: the suppressor tRNA has to be chemically aminoacylated with the unnatural amino acid *in vitro*, and the acylated tRNA is consumed as a stoichiometric reagent during translation and cannot be regenerated. This limitation results in poor suppression efficiency and low protein yields, necessitating highly sensitive techniques to assay the mutant protein such as electrophysiological measurements. Moreover, this method is only applicable to cells that can be microinjected.

A general site-specific incorporation methodology

We have undertaken to develop a general approach for the site-specific incorporation of unnatural amino acids directly into proteins *in vivo*. Importantly, the unnatural amino acid is added to the genetic repertoire, rather than substituting for one of the common 20 amino acids. This method should (i) allow the site-selective insertion of one or more unnatural amino acids at any desired position of any protein, (ii) be applicable to both prokaryotic and eukaryotic cells, (iii) enable *in vivo* studies of mutant proteins in addition to the generation of large quantities of purified mutant proteins, and (iv) be adaptable to a large variety of amino acid side chains. Once established, the methodology should be generally accessible to most chemistry and biology laboratories.

This approach requires the addition of new components to the biosynthetic machinery including a new tRNA-codon pair, an aminoacyl-tRNA synthetase, and an unnatural amino acid (Fig. 1). A new tRNA must be constructed that is not recognized by the endogenous aminoacyl-tRNA synthetases of the host organism, but functions efficiently in translation (an orthogonal tRNA). This tRNA must deliver the novel amino acid in response to a codon that does not encode any of the common 20 amino acids (a unique codon). A new aminoacyl-tRNA synthetase (an orthogonal synthetase) is also required that aminoacylates the orthogonal tRNA, but does not recognize any of the endogenous tRNAs. This synthetase must aminoacylate the tRNA with only the desired unnatural amino acid but none of the common 20 amino acids. Likewise, the unnatural amino acid cannot be a substrate for the endogenous synthetases.

Finally, the amino acid, when added to the growth medium, must be efficiently transported into the cytoplasm. It may also be possible to synthesize the unnatural amino acid biosynthetically or by other modifications to the cellular machinery.

Developing unique genetic codons for unnatural amino acids

The 64 genetic codons code for 20 amino acids and 3 stop signals. Because only one stop codon is needed for translational termination, the other two can in principle be used to encode nonproteinogenic amino acids. The amber stop codon, UAG, has been successfully used in our *in vitro* biosynthetic system and in *Xenopus* oocytes to direct the incorporation of unnatural amino acids. Among the 3 stop codons, UAG is the least used stop codon in *E. coli*. Some *E. coli* strains contain natural suppressor tRNAs, which recognize UAG and insert a natural amino acid. In addition, these amber suppressor tRNAs have been used in conventional protein mutagenesis. Therefore, it should be possible to use UAG for the incorporation of unnatural amino acids *in vivo* without significant perturbation of the host *E. coli*. Because the suppression efficiency for the UAG codon depends upon the competition between the amber suppressor tRNA and the release factor 1 (RF1) (which binds to the UAG codon and initiates release of the growing peptide from the ribosome), it should be possible to modulate the suppression efficiency by either increasing the expression level of the suppressor tRNA or using an RF1 deficient strain. Indeed,
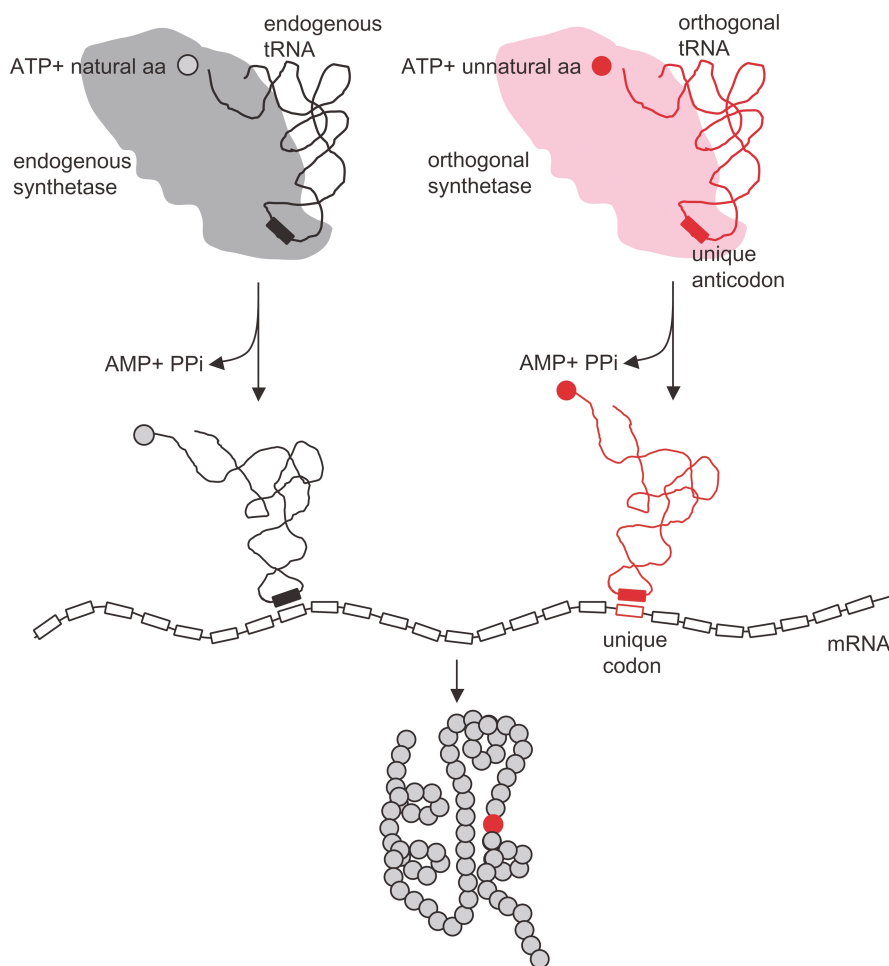

Fig. 1 A general approach for the site-specific incorporation of unnatural amino acids into proteins *in vivo*. The orthogonal aminoacyl-tRNA synthetase acylates the orthogonal tRNA with an unnatural amino acid. The acylated orthogonal tRNA inserts the unnatural amino acid at the position specified by the unique codon, which is introduced into the gene encoding the protein of interest.

several amber suppressor orthogonal tRNAs for use in *E. coli* and eukaryotic cells have been developed in our group and by others (*vide infra*).

It should also be possible to encode unnatural amino acids with rare codons. For example, when the arginine concentration in an *in vitro* protein synthesis reaction is reduced, the rare arginine codon, AGG, has proven to be efficient for insertion of Ala by a synthetic tRNA acylated with alanine.[30] In this case, the synthetic tRNA competes with the naturally occurring tRNA^Arg, which exists as a minor species in *E. coli*. A limitation of using either an amber codon or rare codon as the signal is that only one unnatural amino acid can be encoded at a time. Interestingly, some organisms do not use all triplet codons. An unassigned codon AGA in *Micrococcus luteus* has been utilized for insertion of amino acids in an *in vitro* transcription–translation extract.[31]

An alternative approach is to use extended codons based on frameshift suppression. Four base codons have the potential for insertion of multiple unnatural amino acids into the same protein. For example, four-base codons have been used to incorporate unnatural amino acids into proteins using *in vitro* biosynthetic methods.[30,32] CGGG and AGGU were used to simultaneously incorporate 2-naphthylalanine and an NBD derivative of lysine into streptavidin *in vitro* with two chemically acylated frameshift suppressor tRNAs.[33] In an *in vivo* study, Moore *et al.* examined the ability of tRNA^Leu derivatives with NCUA anticodons to suppress UAGN codons (N can be U, A, G, or C), and found that the quadruplet UAGA can be decoded by a tRNA^Leu with a UCUA anticodon with an efficiency of 13 to 26% with little decoding in the 0 or −1 frame.[34]

We have used a combinatorial approach to exhaustively identify tRNAs that efficiently suppress four-base codons.[35] A reporter library was constructed in which a serine codon in the β-lactamase gene was replaced by four random nucleotides. A tRNA suppressor library was then generated that consists of derivatives of *E. coli* tRNA$_2^{Ser}$ with the anticodon loop (7 nt) replaced with eight or nine random nucleotides. When these two libraries are crossed, an appropriate frameshift suppressor tRNA that decodes the four-base sequence as a single codon results in translation of full-length β-lactamase, rendering the cells resistant to ampicillin. Survival at higher concentrations of ampicillin indicates that the corresponding tRNA has higher suppression efficiency for the four-base codon. Using this selection, four quadruplet codons AGGA, CUAG, UAGA, and CCCU and their cognate suppressor tRNAs were identified that decode only the canonical four-base codon with efficiencies close to that of natural triplet codon suppressors. Novel five- and six-base codon suppressors have also been selected using this strategy.[36] These extended codons, some of which are newly identified, should be useful for the incorporation of multiple unnatural amino acids *in vitro*. Efforts to include them for *in vivo* protein mutagenesis are also underway. Extended codons also have potential problems: in-frame readthrough of the first three bases as a triplet in the extended codon competes with the overall frameshift suppression. As a result, extended codons based on rare codons or nonsense codons may reduce missense readthrough and frameshift suppression at other unwanted sites.

The ultimate solution for generating unique codons is to develop additional unnatural base pairs, *i.e.*, to expand the existing genetic alphabet. One extra base pair would increase the number of triplet codons from 64 to 125. Essential requirements for third base pair candidates include stable and selective base pairing, efficient enzymatic incorporation into DNA with high fidelity by a polymerase, and the efficient continued primer extension after synthesis of the nascent unnatural base pair. For *in vivo* usage, the unnatural nucleoside must be membrane permeable and be phosphorylated to form the corresponding triphosphate. In addition, the increased genetic information must be stable and not destroyed by cellular enzymes. Previous efforts by Benner and others took advantage of hydrogen bonding patterns that are different from those in canonical Watson–Crick pairs, the most noteworthy example of which is the *iso*-C:*iso*-G pair.[37–39] These bases in general mispair to some degree with natural bases and cannot be enzymatically replicated. Kool and co-workers demonstrated that hydrophobic packing interactions between bases can replace hydrogen bonding to drive the formation of base pair.[39,40] In an effort to develop an unnatural base pair satisfying all the above requirements, Schultz, Romesberg and co-workers have systematically synthesized and studied a series of unnatural hydrophobic bases. To date, the PICS:PICS (Scheme 1) self-pair is found to be more stable than natural base pairs,
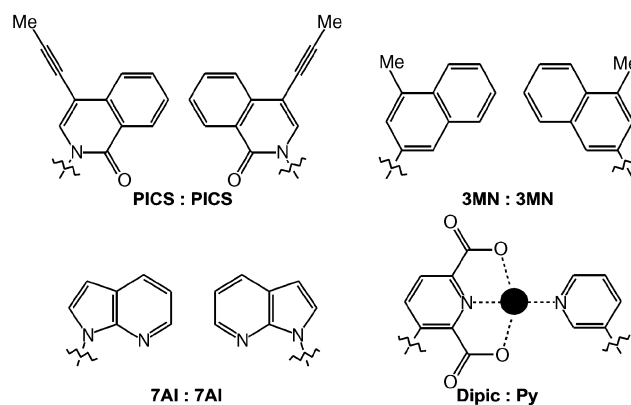
Scheme 1 Structures of unnatural base pairs driven by forces other than hydrogen bonding.

and can be efficiently incorporated into DNA by Klenow fragment of *E. coli* DNA polymerase I (KF).[41,42] A 3MN:3MN self-pair can be synthesized by KF with efficiency and selectivity sufficient for biological function.[43] However, both bases act as a chain terminator for further replication. A mutant DNA polymerase has been recently evolved that can be used to replicate the PICS self pair.[36] In addition, a 7AI self pair can be replicated using a combination of KF and pol β polymerase.[44] More unnatural base pairs, including those with changes to both the base and ribose subunit, are currently investigated. For instance, a novel metallobase pair, Dipic:Py, has been developed, which forms a stable pair upon binding Cu(II).[45]

Generation of orthogonal tRNA–aminoacyl-tRNA synthetase pairs

Extended codons and unnatural base codons have not yet reached the stage of immediate usage *in vivo*. Therefore, we initially focused on developing orthogonal tRNAs that are amber suppressors for use with the UAG codon. However, the strategies described here should be useful to generate orthogonal tRNAs decoding other codons. Because extended codons and unnatural codons are intrinsically orthogonal to natural codons, it may be relatively easier to develop orthogonal tRNAs for them. We chose *E. coli* as the host organism to develop this mutagenesis methodology at the outset due to the ease of genetic manipulation, high transformation efficiency, and the availability of established selections and screens.

One can attempt to generate an orthogonal tRNA–synthetase pair from an existing *E. coli* tRNA–synthetase pair. Specifically, the tRNA's affinity toward its cognate synthetase is

eliminated by mutating nucleotides at the tRNA-synthetase interface while preserving its orthogonality to other synthetases and its ability to function in translation. Using the cognate wild-type synthetase as the starting template, a mutant synthetase is then evolved that uniquely recognizes the engineered orthogonal tRNA. Based on an analysis of the X-ray crystal structure of *E. coli* glutaminyl-tRNA synthetase (GlnRS) complexed with $tRNA_2^{Gln}$, three sites ('knobs') in $tRNA_2^{Gln}$ were identified which make specific contacts with GlnRS.[46,47] These sites were mutated in the tRNA, and mutant suppressor tRNAs containing all possible combinations of knobs 1, 2, and 3 were generated and tested individually by *in vitro* aminoacylation with GlnRS and *in vitro* suppression of amber mutants of chorismate mutase. A mutant tRNA (O-tRNA) bearing all three-knob mutations was shown to be orthogonal to all endogenous *E. coli* synthetases and competent in translation. Next, multiple rounds of DNA shuffling together with oligonucleotide-directed mutagenesis were used to generate libraries of mutant GlnRS's. These mutant enzymes were selected for their ability to acylate the O-tRNA *in vivo* using *E. coli* strain BT235. Only if a mutant GlnRS charges the O-tRNA with glutamine can the genomic amber codon in *lacZ* be suppressed, enabling BT235 cells to grow on lactose minimal media. Several mutant synthetases surviving each round of selection were purified and assayed *in vitro*. The ratio of wild-type (wt) $tRNA^{Gln}$ acylation to O-tRNA acylation by mutant synthetase decreased significantly upon multiple rounds of selection. However, no mutant *E. coli* GlnRS's have been evolved that charge the O-tRNA more efficiently than wild-type *E. coli* $tRNA_2^{Gln}$: the best mutant evolved after seven rounds of DNA shuffling and selection acylates the O-tRNA at only one-ninth the rate of wt $tRNA^{Gln}$. Devising a negative selection against wt tRNA recognition may eventually lead to mutant synthetases with reversed specificities toward the wt- and O-tRNAs. An ideal orthogonal synthetase should not acylate any wt tRNA, since even modest misacylation of a wt tRNA with an unnatural amino acid will likely result in a lethal phenotype. The lack of such a synthetase candidate, together with the finding that mutations within the tRNA interact in complicated, non-additive ways with respect to both aminoacylation and translation,[46] prompted us to examine alternative strategies.

A second strategy for generating an orthogonal tRNA–synthetase pair involves importing a tRNA–synthetase pair from another organism into *E. coli*. The heterologous synthetase candidate should not charge any *E. coli* tRNA, and the heterologous tRNA candidate should not be acylated by any *E. coli* synthetase. In addition, the suppressor tRNA derived from the heterologous tRNA should be orthogonal to all *E. coli* synthetases. Schimmel *et al.* reported that *E. coli* GlnRS (*Ec*GlnRS) does not acylate *Saccharomyces cerevisiae* $tRNA^{Gln}$ (*Ec*GlnRS lacks an N-terminal RNA-binding domain possessed by *S. cerevisiae* GlnRS (*Sc*GlnRS)).[48] This finding prompted us to determine whether the *S. cerevisiae* amber suppressor $tRNA^{Gln}$ (*Sc*$tRNA_{CUA}^{Gln}$) is also not a substrate for *Ec*GlnRS. *In vitro* aminoacylation assays showed this to be the case; and *in vitro* suppression studies show that the *Sc*$tRNA_{CUA}^{Gln}$ is competent in translation.[49] We further showed that *Sc*GlnRS does not acylate any *E. coli* tRNA, only the *Sc*$tRNA_{CUA}^{Gln}$ *in vitro*. The degree to which *Sc*GlnRS is able to aminoacylate the *Sc*$tRNA_{CUA}^{Gln}$ in *E. coli* was also evaluated using an *in vivo* complementation assay. An amber nonsense mutation was introduced at a permissive site in the β-lactamase gene. Suppression of the mutation by an amber suppressor tRNA should produce full-length β-lactamase and confer ampicillin resistance to the cell. When only *Sc*$tRNA_{CUA}^{Gln}$ is expressed, cells exhibit an IC_{50} of 20 μg mL^{-1} ampicillin, indicating virtually no acylation by endogenous *E. coli* synthetases; when *Sc*$tRNA_{CUA}^{Gln}$ is coexpressed with *Sc*GlnRS, cells acquire an IC_{50}

of about 500 μg mL^{-1} ampicillin, demonstrating that *Sc*GlnRS acylates *Sc*$tRNA_{CUA}^{Gln}$ efficiently in *E. coli*.[49] Therefore, the *S. cerevisiae* $tRNA_{CUA}^{Gln}$–GlnRS constitutes an orthogonal pair in *E. coli*.

This strategy was later applied to a $tRNA^{Asp}$–AspRS system. *S. cerevisiae* $tRNA^{Asp}$ is known to be orthogonal to *E. coli* synthetases.[50,51] We demonstrated that an amber suppressor tRNA derived from it (*Sc*$tRNA_{CUA}^{Asp}$) is also orthogonal in *E. coli* using the aforementioned *in vivo* β-lactamase assay. However, the anticodon of $tRNA^{Asp}$ is a critical recognition element of AspRS,[52] and mutation of the anticodon to CUA results in a loss of affinity of the suppressor for AspRS. Fortunately, an *E. coli* AspRS E93K mutant has been shown to recognize *E. coli* amber suppressor $tRNA_{CUA}^{Asp}$ about an order of magnitude better than wt AspRS.[53] It was speculated that introduction of the related mutation in *S. cerevisiae* AspRS (E188K) might restore its affinity for *Sc*$tRNA_{CUA}^{Asp}$. Indeed, the *S. cerevisiae* AspRS(E188K) mutant does not acylate *E. coli* tRNAs, but charges *Sc*$tRNA_{CUA}^{Asp}$ with moderate efficiency as shown by *in vitro* aminoacylation experiments.[54] Hence the *Sc*$tRNA_{CUA}^{Asp}$–*Sc*AspRS(E188K) serves as another orthogonal pair in *E. coli*, albeit with weak activity. To make this pair valuable for later evolution of a synthetase with unnatural amino acid specificity, expression levels of the synthetase and the tRNA were adjusted, and an RF1 deficient strain was employed, resulting in better selection sensitivity.

A similar approach involves the use of a heterologous synthetase as the orthogonal synthetase but a mutant initiator tRNA of the same organism or a related organism as the orthogonal tRNA. RajBhandary and coworkers found that an amber mutant of human initiator $tRNA^{fMet}$ is acylated by *E. coli* GlnRS and acts as an amber suppressor in yeast cells only when *Ec*GlnRS is coexpressed.[55] This pair thus represents an orthogonal pair for use in yeast. Also, an *E. coli* initiator $tRNA^{fMet}$ amber mutant was found that is inactive toward any *E. coli* synthetases. A mutant yeast TyrRS was selected that charges this mutant tRNA, resulting in an orthogonal pair in *E. coli*.[55]

The development of multiple orthogonal tRNA–synthetase pairs may allow the simultaneous incorporation of multiple unnatural amino acids using different codons. Moreover, different aminoacyl-tRNA synthetase may be better starting points for generating active sites with particular side chain specificities. Therefore we also turned to tRNA–synthetase pairs that deliver hydrophobic amino acids. A hydrophobic synthetase active site may be more amenable to engineering new substrate specificities for hydrophobic unnatural amino acids such as fluorophores and affinity labels. Attention was focused on $tRNA^{Tyr}$–TyrRS pair since prokaryotic and eukaryotic $tRNA^{Tyr}$–TyrRS have significant differences: the identity elements of prokaryotic $tRNA^{Tyr}$ include a long variable arm in contrast to the short arm of eukaryotic $tRNA^{Tyr}$. In addition, eukaryotic $tRNA^{Tyr}$ contains a C1:G72 positive recognition element whereas prokaryotic $tRNA^{Tyr}$ has no such consensus base pair. *In vitro* studies have also shown that $tRNA^{Tyr}$ of *S. cerevisiae* and *H. sapiens* cannot be aminoacylated by bacterial synthetases, nor do their TyrRS aminoacylate bacterial $tRNA$.[56,57] In spite of all these promising features for orthogonality, *in vivo* β-lactamase complementation assays showed that the amber suppressor $tRNA_{CUA}^{Tyr}$ derived from both *S. cerevisiae* and *H. sapiens* are not orthogonal in *E. coli*.[58]

The susceptibility of the suppressor tRNA to acylation by *E. coli* synthetases is due to the change of one single nucleotide in the anticodon (G34 to C34). Therefore, a tRNA–synthetase pair with identity elements outside of the anticodon, in particular, the $tRNA^{Tyr}$–TyrRS pair from the archaebacterial *Methanococcus jannaschii* was chosen. This TyrRS is missing most of the non-conserved domain binding for the anticodon loop of its

tRNA^{Tyr}, but can discriminate tRNA with C1:G72 from that with G1:C72. Thus, the *M. jannaschii* TyrRS (*Mj*TyrRS) aminoacylates *S. cerevisiae* but not *E. coli* crude tRNA.[59] Using the *in vivo* complementation assay, we showed that cells expressing the *M. jannaschii* tRNA$^{Tyr}_{CUA}$ (*Mj*tRNA$^{Tyr}_{CUA}$) alone survive to an IC$_{50}$ of 55 µg mL^{-1} ampicillin; cells coexpressing *Mj*tRNA$^{Tyr}_{CUA}$ with its TyrRS survive to an IC$_{50}$ of 1220 µg mL^{-1} ampicillin.[58] This result demonstrates that *M. jannaschii* tRNA$^{Tyr}_{CUA}$–TyrRS is a good candidate for an orthogonal pair in *E. coli*. Although *Mj*tRNA$^{Tyr}_{CUA}$ is less orthogonal in *E. coli* than the *Sc*tRNA$^{Gln}_{CUA}$(IC$_{50}$ 20 µg mL^{-1}), the *Mj*TyrRS has higher aminoacylation activity toward its cognate amber suppressor tRNA.

A general approach was then developed to improve the orthogonality of this *Mj*tRNA$^{Tyr}_{CUA}$ while preserving its affinity toward *Mj*TyrRS.[60] This method consists of a combination of negative and positive selections with a mutant suppressor tRNA library in the absence and presence of the cognate synthetase, respectively. In the negative selection, amber nonsense codon(s) are introduced in a toxic gene at a nonessential position. When a member of the suppressor tRNA library is aminoacylated by endogenous *E. coli* synthetases (*i.e.*, it is not orthogonal to the *E. coli* synthetases), the amber codon is suppressed and the toxic gene product produced leads to cell death. Only cells harboring orthogonal tRNAs or non-functional tRNAs can survive. All survivors are then subjected to a positive selection in which an amber codon is placed in a drug resistance gene at a nonessential position. tRNAs are then selected for their ability to be aminoacylated by the coexpressed cognate synthetase and to insert an amino acid in response to this amber codon. Cells harboring non-functional tRNAs, or tRNAs that cannot be recognized by the synthetase of interest will be sensitive to antibiotic. Therefore, only tRNAs that: (i) are not substrates for endogenous *E. coli* synthetases; (ii) can be aminoacylated by the synthetase of interest; and (iii) are functional in translation will survive both selections.

Eleven nucleotides of the *Mj*tRNA$^{Tyr}_{CUA}$ that do not interact directly with the TyrRS were randomly mutated to generate a suppressor tRNA library. This tRNA library was passed through a negative selection using the ribonuclease barnase (expression of barnase is toxic to cells) and then a positive selection based on suppression of an amber codon in β-lactamase gene. The best mutant tRNA (mutRNA$^{Tyr}_{CUA}$) selected confers cells an IC$_{50}$ of 12 µg mL^{-1} ampicillin, a value similar to that of background readthrough of the amber codon, indicating this mutant tRNA is a much poorer substrate for *E. coli* synthetase than the wt *Mj*tRNA$^{Tyr}_{CUA}$. When the mutRNA$^{Tyr}_{CUA}$ is coexpressed with *Mj*TyrRS, cells survive to an IC$_{50}$ of 440 µg mL^{-1} ampicillin, showing the tRNA is still aminoacylated efficiently by the TyrRS. The low background and high activity of this mutRNA-$^{Tyr}_{CUA}$–*Mj*TyrRS pair make it an excellent candidate orthogonal pair. This *in vivo* double selection strategy should be generalizable to additional tRNA–synthetase pairs. As long as an orthogonal synthetase has been identified, in principle it is possible to select an orthogonal suppressor tRNA that can be charged by the synthetase.

Engineering a synthetase with unnatural amino acid specificity

At this stage, the orthogonal synthetase acylates the orthogonal tRNA with a natural amino acid. The substrate specificity of the synthetase must be altered so that only the desired unnatural amino acid, but not any common 20 amino acids are charged. Promiscuity in the orthogonal synthetase will result in mutant proteins with a mixture of natural and unnatural amino acids at the target position. For instance, in an attempt to site-specifically incorporate *p*-F-Phe, a yeast amber suppressor tRNA$^{Phe}_{CUA}$–phenylalanyl-tRNA synthetase pair was used in a *p*-F-Phe resistant, Phe auxotrophic *E. coli* strain.[61] Because yeast PheRS does not have high substrate specificity for *p*-F-Phe, the mutagenesis site is translated with 64–75% *p*-F-Phe and the remainder as Phe and Lys even in the excess of *p*-F-Phe added to the growth media. Also, at the Phe codon positions, 7% *p*-F-Phe is found, indicating that the endogenous *E. coli* PheRS incorporates *p*-F-Phe in addition to Phe. Besides its translational infidelity, this approach is not generally applicable to other unnatural amino acids.

Modification of the substrate specificity of a synthetase is expected to be difficult due to the high intrinsic fidelity of the natural synthetases and the fact that unnatural amino acids are not required for any cellular function. We have pursued a combinatorial approach to this problem, in which a pool of mutant synthetases is generated from the framework of a wild-type synthetase, and then selected based on their specificity for an unnatural amino acid relative to the common twenty. To isolate such a synthetase, the selection method should be: (i) sensitive, as the activity of 'hits' from the initial rounds can be low and the population small; (ii) tunable, since it would be desirable to vary the selection stringency at different selection rounds; (iii) general, so that it can be used for different unnatural amino acids. We describe here several approaches developed in our laboratory to tackle this problem.

A general *in vivo* selection/screen strategy was developed that is based on the combination of a positive selection followed by a negative selection (Fig. 2). In the positive selection, suppression of the amber codon introduced at nonessential position(s) of the positive marker will allow cells to survive under positive selection pressure. In the presence of both natural and unnatural amino acids, survivors thus encode active synthetases charging the orthogonal suppressor tRNA with either a natural or unnatural amino acid. In the negative selection, those synthetases with specificities for natural amino acids charge the orthogonal tRNA, resulting in suppression of an amber codon in the negative marker and cell death. Since no unnatural amino acid is added, synthetases with specificities for the unnatural amino acid will survive. Survivors passing both selection/screen therefore must encode synthetases charging the orthogonal tRNA with an unnatural amino acid. More mutations are introduced by mutagenesis or DNA shuffling into these synthetase genes to generate a second generation synthetase library, which is used for further rounds of selection until a mutant synthetase with desired activity is evolved.

Our general selection scheme involves a positive selection based on suppression of an amber stop codon at nonessential position in the β-lactamase gene, rendering cells ampicillin resistance; and a negative selection using the ribonuclease barnase as the negative marker.[49] An alternative is to replace β-lactamase gene with the chloramphenicol acetyltransferase (CAT) gene, so that chloramphenicol can be applied as the positive selection pressure.[54] In contrast to β-lactamase which is secreted into the periplasm, CAT localizes in the cytoplasm; moreover, ampicillin is bacteriocidal while chloramphenicol is bacteriostatic. Because barnase is an extremely toxic protein, it was necessary to control the stringency of the negative selection by introducing a different number of amber codons into the barnase gene. A direct replica plate method was also employed:[62] after passing the positive selection, cells are grown in the presence of the either ampicillin or chloramphenicol and the absence of the unnatural amino acid. Those cells that do not survive are isolated from a replica plate supplemented with the unnatural amino acid. No transformation into a second negative selection strain is needed and the phenotype is clear-cut. A limitation of this approach is that the number of colonies that can be handled using replica plates is relatively small.
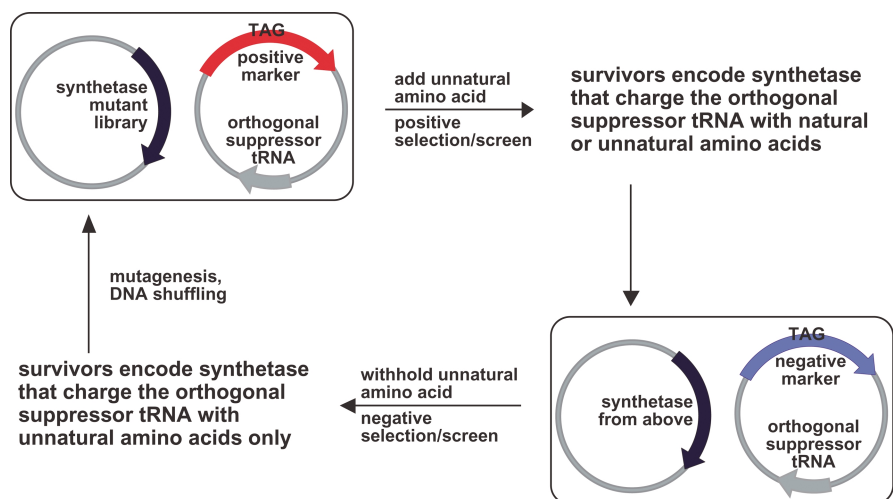
Fig. 2 Schematic illustration of the general selection/screen for aminoacyl-tRNA synthetases with unnatural amino acid specificities. In the positive selection, active synthetases with either natural or unnatural amino acid specificities are identified; in the negative selection, synthetases with natural amino acid specificities are eliminated. Therefore, only synthetases charging the orthogonal tRNA with the unnatural amino acid can survive both selections/screens.

Compared to other potential selection markers, a positive selection based on antibiotic resistance offers the ability to tune selection stringency by varying the concentration of the antibiotic, and to compare the suppression efficiency by monitoring the highest antibiotic concentration cells can survive. More importantly, the growth process is also an enrichment procedure. This can lead to quick accumulation of the desired phenotype; yet can be potentially skewed by other factors affecting growth rates, such as reporter gene reversion.

We have also developed a general fluorescence-activated cell sorting (FACS) based screen with green fluorescent protein (GFP) as the reporter. A T7 RNA polymerase containing amber mutations, together with GFP under control of a T7 promoter are used. Only when the amber codons are suppressed can cells produce functional T7 RNA polymerase and express GFP, rendering cells fluorescence. In the positive screen, fluorescent cells are collected which encode active synthetases charging the orthogonal tRNA with either natural or unnatural amino acids. The selected cells are then diluted and grown in the absence of the unnatural amino acid, and then sorted by FACS for cells without fluorescence, *i.e.*, that express synthetases with specificities for unnatural amino acids only. By setting the collection threshold of the fluorescence intensity, the stringency of both positive and negative screen can be conveniently controlled.[36]

A direct positive selection specific for a particular unnatural amino acid has also been developed which exploits the high affinity of a monoclonal antibody for an unnatural amino acid displayed on a phage surface (Fig. 3).[63] A C3 peptide with an amber mutation is fused to the N-terminus of VSCM13 phage coat protein pIII, such that phage production requires suppression of the amber stop codon. Cells harboring a phagemid that expresses an orthogonal suppressor tRNA and a synthetase library are infected with the C3TAG phage. An active synthetase results in suppression of C3TAG and display of its cognate amino acid on the phage surface. The phage pool is then incubated with immobilized monoclonal antibodies directed against the unnatural amino acid to isolate only those phage carrying the synthetase specific for the unnatural amino acid. In a simulated selection, phage displaying Asp were enriched over 300-fold from a pool of phage displaying Asn using antibodies raised against the Asp-containing epitope.

Several *in vitro* screen methods are also currently being developed. In one such method, a library of mutant synthetases is displayed on the phage, and the phage particles are panned against immobilized sulfamoyl analogs of the aminoacyl adenylate intermediate (Scheme 2). In a preliminary test, *M. jannaschii* TyrRS was fused to the pIII coat protein of M13 phage. This phage was enriched 1000-fold over a control phage displaying an unrelated antibody after panning against the
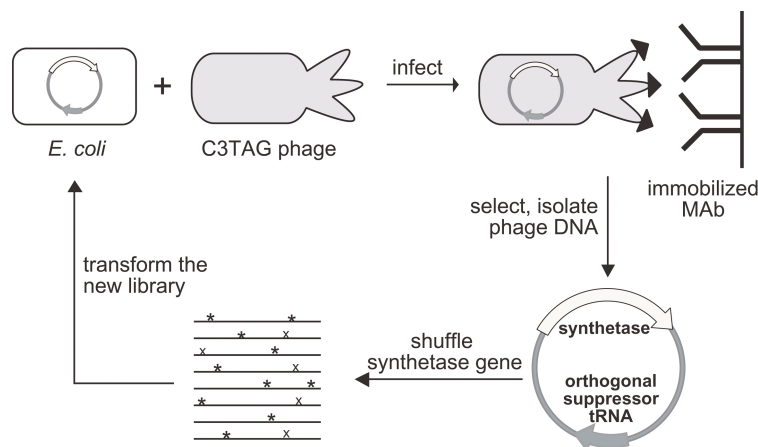

Fig. 3 Phage-based selection for the incorporation of unnatural amino acids into a surface epitope. *E. coli* carrying the mutant synthetase library are infected by phage with a stop codon in a gene encoding a surface protein. Phage containing an active synthetase display the unnatural amino acid on the phage surface and are selected with immobilized monoclonal antibodies.
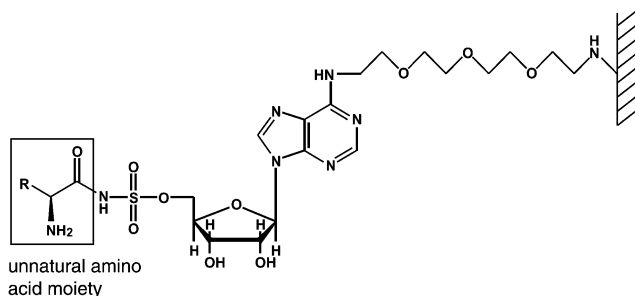
Scheme 2 The immobilized sulfamoyl analog of the aminoacyl adenylate intermediate used to screen phage-displayed synthetases with unnatural amino acid specificity.

sulfamoyl analog of tyrosyl adenylate.[36] Given that only 0.1 to 1% of the starting TyrRS phage population displays the TyrRS protein, the actual enrichment factor can be as high as 10^5 to 10^6.

Uptake of unnatural amino acids

A prerequisite for *in vivo* incorporation of unnatural amino acids is that unnatural analogs must be efficiently uptaken into the cell from the growth media. A rapid screen was developed to assess which unnatural amino acids can be uptaken by cells.[49] For some toxic unnatural amino acids, the presence of excess of natural amino acids rescues the ability of cells to grow in the presence of the toxin. These toxic unnatural amino acids are assigned as 'lethal alleles'. Complementation of the toxic allele, evidenced by the restoration of cell growth, suggests that the nontoxic unnatural amino acids are either uptaken by the cell or competitively inhibit transport of the toxic amino acid. Using this screen, it was found 13 out of 22 analogs of Glu and Gln are likely taken up by the cell, indicating that the *E. coli* Glu and Gln transport pathways may tolerate significant perturbations in amino acid structure. Recent results also show that a variety of tyrosine analogs can be efficiently uptaken by *E. coli*.

Some specific unnatural analogs with interesting properties may not be uptaken by the cell. Thus it may be necessary to develop methods for transport of unnatural amino acids. One approach takes advantage of peptide permeases, which transport dipeptides and tripeptides across the bacterial inner cell membrane. Peptide permeases are not very side-chain specific, and the K_d values for their substrates are comparable to K_d values of amino acid permeases (0.1–10 µM). The desired

unnatural amino acid can be conjugated to a natural amino acid, and the resulting dipeptide is fed to a strain of *E. coli* deficient in the biosynthesis and uptake of the natural amino acid. The survival of this strain in minimal media would be dependent on the uptake of the dipeptide, and efficient cytoplasmic hydrolysis of the dipeptide to supply the natural amino acid, thus releasing the unnatural amino acid. One can also envisage the use of molecular transporters to actively deliver unnatural amino acids into cells. A candidate transporter is the polyguanidine peptoid derivative, which exhibits significantly enhanced cellular uptake compared to the natural subunit Tat.[64] Unnatural amino acids can be covalently attached to this substrate *via* a native peptide bond, and released by protease cleavage after being translocated into cells by the conjugate.

The first 'unnatural' organism

Recently, we have successfully expanded the number of genetically encoded amino acids in *E. coli*.[62] A unique tRNA–aminoacyl-tRNA synthetase pair was generated that, when introduced into *E. coli*, led to the site-specific incorporation of *O*-methyl-L-tyrosine (*O*-met-Tyr) into proteins in response to an amber nonsense codon with a fidelity rivaling that of natural amino acids. The mutRNA$_{CUA}^{Tyr}$ derived from *M. jannaschii* tRNA^{Tyr} (see above) was used as the orthogonal suppressor tRNA; a mutant TyrRS that uniquely charges this tRNA with *O*-met-Tyr only was evolved from a library of *M. jannaschii* TyrRS mutants. Five residues in the active site of *M. jannaschii* TyrRS were all initially mutated to alanine to eliminate wild-type TyrRS contamination (Fig. 4). The resulting inactive Ala$_5$ TyrRS was used as a template for PCR random mutagenesis to generate the mutant TyrRS library with the five residues randomized. A positive selection was applied that is based on suppression of an amber stop codon in the chloramphenicol acetyltransferase (CAT) gene in the presence of the unnatural amino acid. Cells surviving on chloramphenicol were then grown in the presence of chloramphenicol and in the absence of the unnatural amino acid. Those cells that did not survive were isolated from a replica plate supplemented with the unnatural amino acid. The mutant TyrRS genes were isolated from these cells, recombined *in vitro* by DNA shuffling, and transformed back into *E. coli* for further rounds of selection with increasing concentrations of chloramphenicol. After two rounds of selection and DNA shuffling, a clone was evolved whose survival in chloramphenicol was dependent on the addition of 1 mM *O*-methyl-L-tyrosine to the growth media.
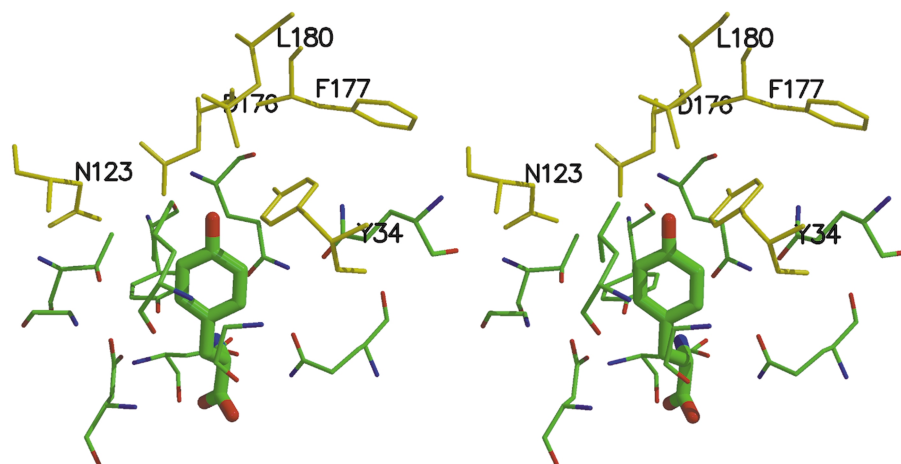

Fig. 4 Stereoview of the active site of TyrRS. Residues from *B. stearothermophilus* TyrRS are shown in the figure. Corresponding residues from *M. jannaschii* TyrRS are Tyr32 (Tyr34), Glu107 (Asn123), Asp158 (Asp176), Ile159 (Phe177), and Leu162 (Leu180) with residues from *B. stearothermophilus* TyrRS in parenthesis. Mutated residues are in yellow.

To demonstrate that the observed phenotype is due to the site-specific incorporation of *O*-methyl-L-tyrosine by the mutRNA$_{\text{CUA}}^{\text{Tyr}}$–mutant TyrRS pair in response to an amber stop codon, an *O*-met-Tyr mutant of dihydrofolate reductase (DHFR) was generated and characterized. The third codon of the *E. coli* DHFR gene was mutated to TAG. When the mutant TyrRS was expressed in the presence of tRNA$_{\text{CUA}}^{\text{Tyr}}$ and 1 mM *O*-met-Tyr, full length DHFR was produced. In the absence of either *O*-met-Tyr, tRNA$_{\text{CUA}}^{\text{Tyr}}$ or mutant TyrRS, no DHFR ($<0.1\%$ by densitometry) was observed by analysis with silver-stained SDS-PAGE gel and Western blot (Fig. 5). The identity of the

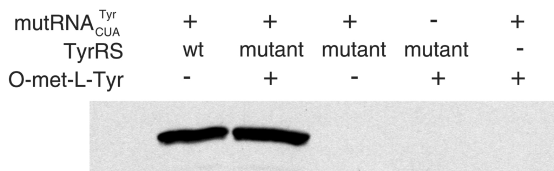| mutRNA$_{\text{CUA}}^{\text{Tyr}}$	+	+	+	-	+
TyrRS	wt	mutant	mutant	mutant	-
O-met-L-Tyr	-	+	-	+	+

Fig. 5 Western blot analysis of the accumulation of *E. coli* DHFR protein under different conditions. A six-histidine tag was added to the COOH terminus of DHFR, and protein was purified by immobilized metal affinity chromatography. A penta-His antibody was used to detect the six-histidine tag. Expression conditions are notated at the top of each lane.

amino acid inserted in response to the TAG codon was confirmed to be *O*-met-Tyr by mass analysis of both the intact protein and tryptic fragments (Fig. 6). No indication of the incorporation of tyrosine or other amino acids at that position was observed. Analysis of the sequence of the mutant TyrRS revealed the following mutations: Tyr32 → Gln, Asp158 → Ala, Glu107 → Thr, and Leu162 → Pro (Fig. 4). Kinetics of
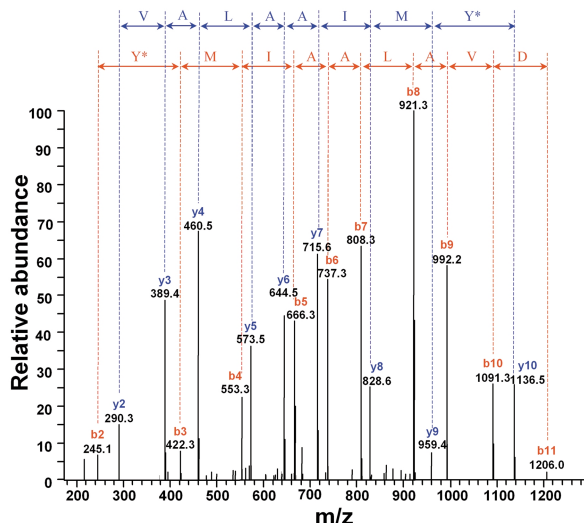

Fig. 6 Tandem mass spectrum of the NH$_2$-terminal peptide MIY*MIAA-LAVDR from the mutant DHFR protein. The annotated b (red) or y (blue) ion series confirmed the identity of the residue (Y*) inserted at the TAG codon to be *O*-methyl-L-tyrosine. The observed value for the monoisotopic mass for the intact mutant protein by FT-ICR MS was 18 096.002 daltons, which is within 5 ppm of the theoretical mass of 18 095.908 daltons.

adenylate formation of *O*-met-Tyr and tyrosine with ATP catalyzed by the mutant TyrRS was analyzed *in vitro* using a pyrophosphate-exchange assay. The value of $k_{\text{cat}}/K_{\text{m}}$ of the mutant TyrRS for *O*-methyl-L-tyrosine is about 100 fold higher than that of tyrosine.
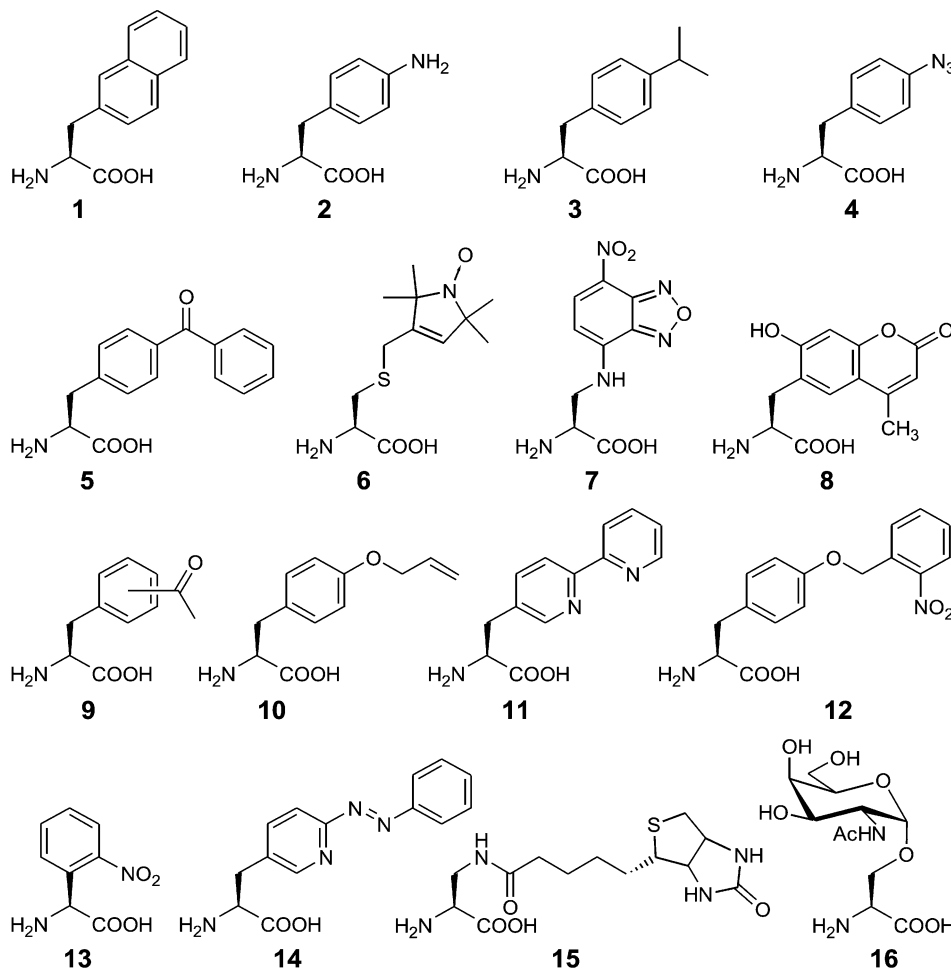
Scheme 3 Structures of unnatural amino acids discussed in the text.

Future directions

To demonstrate the generality of this methodology, we are currently evolving more mutant TyrRS with specificities for different tyrosine analogs.[36] To date, we have isolated another mutant TyrRS that incorporates L-3-(2-naphthyl)alanine (**1**) site-specifically (Scheme 3). Promising hits that encode mutant TyrRS's charging *p*-amino-L-phenylalanine (**2**) and *p*-iso-propyl-L-phenylalanine (**3**) have also been selected. These results suggest that the TyrRS active site is surprisingly amenable to modifications of its specificity.

In the next stage, a variety of amino acids with novel structural, chemical, and physical properties will be investigated.[36] For instance, mutant TyrRS's with specificities for photoactivatable cross-linkers **4** and **5** have been selected. We are also evolving synthetases with specificities for spin-labeled (**6**) and fluorescent amino acids (**7, 8**), amino acids with novel functional groups (**9, 10**), metal binding amino acids (**11**), photocaged (**12, 13**) and photoisomerizable amino acids (**14**), and biotinylated (**15**) and glycosylated amino acids (**16**). To increase the diversity of unnatural amino acids that can be incorporated, a new orthogonal tRNA^Leu–LeuRS is also being developed.[36] The active site of this synthetase should accommodate large amino acid side chains without movement of the polypeptide backbone.

We are also developing orthogonal tRNA^Tyr–TyrRS and tRNA^Trp–TrpRS pairs for use in mammalian cells.[36] Since it is more difficult to generate large libraries in mammalian cells for selection and screening, the active site of mutant TyrRS evolved in *E. coli* will be transplanted into mammalian cells. Alternatively, a mutant synthetase can be selected using *in vitro* methods, such as the direct affinity binding approach.

To generate new proteins with novel properties, a 'random unnatural amino acid mutagenesis' method is being developed, in which amber stop codons will randomly replace any codon in the gene of a target protein.[36] Incorporation of an unnatural amino acid at the TAG codon will result in an unnatural mutant protein library. It is also possible to enable cells to generate unnatural amino acids on their own. We are modifying amino acid biosynthetic pathways or adding new enzymes that convert natural amino acids or metabolic intermediates into unnatural amino acids, thus obviating the addition and uptake of unnatural amino acids.

Conclusion

We have shown it is possible to augment the protein biosynthetic machinery of *E. coli* to accommodate additional genetically encoded amino acids. Additional orthogonal tRNA–synthetase pairs, as well as new codons, may further expand the number and scope of amino acids that can be incorporated. The ability to introduce novel amino acids into proteins directly in living cells will provide new tools for studies of protein and cellular function and may lead to generation of proteins with enhanced properties.

Acknowledgements

Notes and references

1 P. E. Dawson and S. B. H. Kent, *Annu. Rev. Biochem.*, 2000, **69**, 923.
2 V. W. Cornish, D. Mendel and P. G. Schultz, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 621.
3 M. H. Richmond, *Bacteriol. Rev.*, 1962, **26**, 398.
4 D. B. Cowie, G. N. Cohen, E. T. Bolton and H. DeRrobinchon-Szulmajst, *Biochim. Biophys. Acta*, 1959, **34**, 39.
5 R. Munier and G. N. Cohen, *Biochim. Biophys. Acta*, 1959, **31**, 378.
6 N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. M. Dong, L. Moroder and R. Huber, *FASEB J.*, 1999, **13**, 41.
7 C. Minks, R. Huber, L. Moroder and N. Budisa, *Anal. Biochem.*, 2000, **284**, 29.
8 H. Duewel, E. Daub, V. Robinson and J. F. Honek, *Biochemistry*, 1997, **36**, 3404.
9 Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado and D. A. Tirrell, *Angew. Chem., Int. Ed. Engl.*, 2001, **40**, 1494.
10 W. A. Hendrickson, J. R. Horton and D. M. Lemaster, *EMBO J.*, 1990, **9**, 1665.
11 J. O. Boles, K. Lewinski, M. Kunkle, J. D. Odom, B. Dunlap, L. Lebioda and M. Hatada, *Nat. Struct. Biol.*, 1994, **1**, 283.
12 N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann and R. Huber, *Eur. J. Biochem.*, 1995, **230**, 788.
13 N. Budisa, W. Karnbrock, S. Steinbacher, A. Humm, L. Prade, T. Neuefeind, L. Moroder and R. Huber, *J. Mol. Biol.*, 1997, **270**, 616.
14 J. C. M. vanHest and D. A. Tirrell, *FEBS Lett.*, 1998, **428**, 68.
15 J. C. M. van Hest, K. L. Kiick and D. A. Tirrell, *J. Am. Chem. Soc.*, 2000, **122**, 1282.
16 K. L. Kiick and D. A. Tirrell, *Tetrahedron*, 2000, **56**, 9487.
17 N. Budisa, C. Minks, F. J. Medrano, J. Lutz, R. Huber and L. Moroder, *Proc. Natl. Acad. Sci. U.S.A.*, 1998, **95**, 455.
18 M. Ibba, P. Kast and H. Hennecke, *Biochemistry*, 1994, **33**, 7107.
19 M. Ibba and H. Hennecke, *FEBS Lett.*, 1995, **364**, 272.
20 N. Sharma, R. Furter, P. Kast and D. A. Tirrell, *FEBS Lett.*, 2000, **467**, 37.
21 F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soll and S. Nishimura, *J. Biol. Chem.*, 2000, **275**, 40 324.
22 V. Doring, H. D. Mootz, L. A. Nangle, T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel and P. Marliere, *Science*, 2001, **292**, 501.
23 M. W. Nowak, P. C. Kearney, J. R. Sampson, M. E. Saks, C. G. Labarca, S. K. Silverman, W. G. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G. Schultz, D. A. Dougherty and H. A. Lester, *Science*, 1995, **268**, 439.
24 D. A. Dougherty, *Curr. Opin. Chem. Biol.*, 2000, **4**, 645.
25 G. Turcatti, K. Nemeth, M. D. Edgerton, U. Meseth, F. Talabot, M. Peitsch, J. Knowles, H. Vogel and A. Chollet, *J. Biol. Chem.*, 1996, **271**, 19991.
26 J. P. Gallivan, H. A. Lester and D. A. Dougherty, *Chem. Biol.*, 1997, **4**, 739.
27 J. C. Miller, S. K. Silverman, P. M. England, D. A. Dougherty and H. A. Lester, *Neuron*, 1998, **20**, 619.
28 P. M. England, Y. Zhang, D. A. Dougherty and H. A. Lester, *Cell*, 1999, **96**, 89.
29 T. Lu, A. Y. Ting, J. Mainland, L. Y. Jan, P. G. Schultz and J. Yang, *Nat. Neurosci.*, 2001, **4**, 239.
30 C. H. Ma, W. Kudlicki, O. W. Odom, G. Kramer and B. Hardesty, *Biochemistry*, 1993, **32**, 7939.
31 A. K. Kowal and J. S. Oliver, *Nucl. Acid. Res.*, 1997, **25**, 4685.
32 T. Hohsaka, D. Kajihara, Y. Ashizuka, H. Murakami and M. Sisido, *J. Am. Chem. Soc.*, 1999, **121**, 34.
33 T. Hohsaka, Y. Ashizuka, H. Sasaki, H. Murakami and M. Sisido, *J. Am. Chem. Soc.*, 1999, **121**, 12194.
34 B. Moore, B. C. Persson, C. C. Nelson, R. F. Gesteland and J. F. Atkins, *J. Mol. Biol.*, 2000, **298**, 195.
35 T. J. Magliery, J. C. Anderson and P. G. Schultz, *J. Mol. Biol.*, 2001, **307**, 755.
36 P. G. Schultz *et al.*, unpublished results.
37 C. Switzer, S. E. Moroney and S. A. Benner, *J. Am. Chem. Soc.*, 1989, **111**, 8322.
38 J. A. Piccirilli, T. Krauch, S. E. Moroney and S. A. Benner, *Nature*, 1990, **343**, 33.
39 E. T. Kool, *Curr. Opin. Chem. Biol.*, 2000, **4**, 602.
40 K. M. Guckian and E. T. Kool, *Angew. Chem., Int. Ed. Engl.*, 1998, **36**, 2825.
41 D. L. McMinn, A. K. Ogawa, Y. Q. Wu, J. Q. Liu, P. G. Schultz and F. E. Romesberg, *J. Am. Chem. Soc.*, 1999, **121**, 11 586.
42 A. K. Ogawa, Y. Q. Wu, D. L. McMinn, J. Q. Liu, P. G. Schultz and F. E. Romesberg, *J. Am. Chem. Soc.*, 2000, **122**, 3274.
43 A. K. Ogawa, Y. Q. Wu, M. Berger, P. G. Schultz and F. E. Romesberg, *J. Am. Chem. Soc.*, 2000, **122**, 8803.
44 E. J. L. Tae, Y. Q. Wu, G. Xia, P. G. Schultz and F. E. Romesberg, *J. Am. Chem. Soc.*, 2001, **123**, 7439.
45 E. Meggers, P. L. Holland, W. B. Tolman, F. E. Romesberg and P. G. Schultz, *J. Am. Chem. Soc.*, 2000, **122**, 10714.
46 D. R. Liu, T. J. Magliery and P. G. Schultz, *Chem. Biol.*, 1997, **4**, 685.

47 D. R. Liu, T. J. Magliery, M. Pasternak and P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.*, 1997, **94**, 10092.
48 E. F. Whelihan and P. Schimmel, *EMBO J.*, 1997, **16**, 2968.
49 D. R. Liu and P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.*, 1999, **96**, 4780.
50 B. P. Doctor and J. A. Mudd, *J. Biol. Chem.*, 1963, **238**, 3677.
51 Y. Kwok and J. T. Wong, *Can. J. Biochem.*, 1980, **58**, 213.
52 R. Giege, C. Florentz, D. Kern, J. Gangloff, G. Eriani and D. Moras, *Biochimie*, 1996, **78**, 605.
53 F. Martin, 'Thesis', Universite Louis Pasteur, Strasbourg, France, 1995.
54 M. Pastrnak, T. J. Magliery and P. G. Schultz, *Helv. Chim. Acta*, 2000, **83**, 2277.
55 A. K. Kowal, C. Kohrer and U. L. RajBhandary, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 2268.
56 K. Wakasugi, C. L. Quinn, N. Tao and P. Schimmel, *EMBO J.*, 1998, **17**, 297.
57 T. A. Kleeman, D. Wei, K. L. Simpson and E. A. First, *J. Biol. Chem.*, 1997, **272**, 14420.
58 L. Wang, T. J. Magliery, D. R. Liu and P. G. Schultz, *J. Am. Chem. Soc.*, 2000, **122**, 5010.
59 B. A. Steer and P. Schimmel, *J. Biol. Chem.*, 1999, **274**, 35601.
60 L. Wang and P. G. Schultz, *Chem. Biol.*, 2001, **8**, 883.
61 R. Furter, *Protein Sci.*, 1998, **7**, 419.
62 L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498.
63 M. Pastrnak and P. G. Schultz, *Bioorg. Med. Chem.*, 2001, **9**, 2373.
64 P. A. Wender, D. J. Mitchell, K. Pattabiraman, E. T. Pelkey, L. Steinman and J. B. Rothbard, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 13003.