Adding New Chemistries to the Genetic Code

Chang C. Liu and Peter G. Schultz

The Scripps Research Institute, La Jolla, California 92037; email: ccliu@berkeley.edu, schultz@scripps.edu

Annu. Rev. Biochem. 2010. 79:413–44

First published online as a Review in Advance on March 18, 2010

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

This article's doi: 10.1146/annurev.biochem.052308.105824

0066-4154/10/0707-0413$20.00

Key Words

aminoacyl-tRNA synthetase, protein engineering, protein evolution, translation

Abstract

The development of new orthogonal aminoacyl-tRNA synthetase/tRNA pairs has led to the addition of approximately 70 unnatural amino acids (UAAs) to the genetic codes of *Escherichia coli*, yeast, and mammalian cells. These UAAs represent a wide range of structures and functions not found in the canonical 20 amino acids and thus provide new opportunities to generate proteins with enhanced or novel properties and probes of protein structure and function.

Contents

INTRODUCTION

Proteins carry out a remarkable range of functions—from photosynthesis to transcription and signal transduction—with only 20 amino acids. Nonetheless, proteins often require chemistries beyond those contained in the canonical 20 amino acids for function, including cofactors, such as pyridoxal, thiamine, flavins, and metal ions (1), and posttranslational modifications (PTMs) such as methylation, glycosylation, sulfation, and phosphorylation (2). Moreover, the modification of proteins by chemical methods can lead to new physicochemical, biological, or pharmacological properties (3–8). A number of archaea and eubacteria even encode the noncanonical amino acids selenocysteine or pyrrolysine for added function (9). Thus, the creation of organisms with expanded genetic codes that include amino acids beyond the common 20 building blocks might allow the design of peptides and proteins with enhanced or novel activities. Additional building blocks might also facilitate the incorporation of biophysical probes into proteins for the analysis or control of protein structure and function in vitro and in living cells. Finally, an expanded genetic code may provide an advantage in the evolution of new molecular or organismal function. Here, we describe efforts to augment the genetic codes of both prokaryotic and eukaryotic organisms with unnatural amino acids (UAAs) that have novel properties, and we illustrate the utility of this methodology in exploring protein structure and function and in generating proteins with new and/or enhanced activities.

PTM: posttranslational modification

Unnatural amino acid (UAA): an amino acid not specified by the existing genetic code, which encodes only 20 amino acids

BACKGROUND

Chemists have exquisite control over the structures of small molecules, but the control of macromolecular, especially protein, structure and function still represents a major challenge. A number of approaches are being pursued to address this limitation. The first of these is solid-phase peptide synthesis, which allows chemists to rationally manipulate polypeptide structure using natural and unnatural amino acids (5, 10). Though this method is subject to size limitations (peptides containing over 50–100 amino acids are difficult to synthesize in significant quantities), it can be used in conjunction with semisynthetic methods to produce larger full-length proteins containing UAAs. For example, in native chemical ligation, a peptide containing a C-terminal α-thioester is reacted with a second peptide containing an N-terminal cysteine. The resulting thioester linkage undergoes an acyl rearrangement to form a native peptide bond, joining the two smaller peptides (7). A C-terminal α-thioester leaving group can also be generated by an

intein-mediated reaction arrested before the final splicing step. This method, termed expressed protein ligation (EPL), allows the linkage of a recombinantly expressed protein to a synthetic UAA-containing peptide (8). Solid-phase peptide synthesis and ligation techniques have been used to modify protein backbones (11), make polymer-modified erythropoietin analogs (12, 13), introduce fluorescent probes into peptides and proteins, and produce modified signaling proteins, ion channels, and histones (14–16). However, the general application of these techniques is limited by the need for protecting group chemistry, the restrictions on the sites of ligation, the constraints on protein folding, and the inherent extracellular nature of these synthetic techniques.

Biosynthetic approaches have also been developed for the in vitro synthesis of proteins containing UAAs. In these techniques, truncated tRNAs are enzymatically ligated to chemically aminoacylated nucleotides, effectively decoupling the identity of the tRNA from that of the attached amino acid (17–21). Cell-free translation systems then use these aminoacylated tRNAs for protein synthesis in response to either "blank" (nonsense or frameshift codons that can be used to specify an UAA) or coding codons. Though technically challenging, this method has been very useful in the study of protein structure and function. Examples include the incorporation of amino acids with modified backbones (22, 23), fluorophores (24–26), functional groups corresponding to PTMs (27), photo- and chemically reactive side chains (28), and altered pK_as (29). This method has also been combined with mRNA display to generate selectable peptide libraries that contain UAAs (30, 31). A variation of this approach involves the injection or transfection of chemically or otherwise aminoacylated tRNAs into living cells (32, 33). For example, by chemically acylating an amber suppressor tRNA, such as one derived from *Tetrahymena thermophilia* $tRNA^{Gln}$ (which affords high efficiency and fidelity of UAA incorporation when microinjected into *Xenopus* oocytes) (34), UAAs spanning a range of electronic, structural, and conformational properties have been incorporated into ion channels directly in cells (35–37). This has allowed for detailed structure-function studies by the systematic modification of individual protein residues (38), including the determination of the role of proline isomerization in channel gating (39). Nevertheless, these methods are limited by the accessibility and stability of the aminoacyl-tRNA adducts, the stoichiometric use of aminoacylated tRNAs that cannot be continuously delivered, and the disruptive nature of microinjection and transfection techniques.

Another approach that is applicable to living cells involves the use of wild-type aminoacyl-tRNA synthetases (aaRSs) to incorporate UAAs that are close structural analogs of canonical amino acids (40, 41). In this approach, a strain auxotrophic for one of the common 20 amino acids is used to substitute that amino acid with an UAA analog. Although the resulting wholesale replacement of a common amino acid by an UAA cannot sustain exponential growth, nondividing cells are still viable and are able to overexpress proteins that contain the UAA. In these newly synthesized proteins, the canonical amino acid is efficiently replaced by its UAA analog at all sites (42, 43). The diversity in the range of UAA analogs that can be incorporated using this approach has been increased by aaRS overexpression (44), active-site engineering (45, 46), editing domain mutations (47), and sorting by cell-surface display of the UAA whose incorporation was desired (48). The global incorporation of UAA analogs by this method has a number of useful applications. For example, substitution of methionine with selenomethionine introduces a heavy atom into proteins for crystallographic phasing experiments (49); replacement of methionine by norleucine in cytochrome P450, leucine by 5′,5′,5′-trifluoroleucine in chloramphenicol acetyltransferase, and tryptophan by 4-aminotryptophan in barstar and green fluorescent protein (GFP) yields proteins with new activities and properties (50–53); and timed substitution of methionine or phenylalanine

aaRS: aminoacyl-tRNA synthetase

by alkyne-containing UAA analogs allows the tracking of newly synthesized proteins (54, 55). Such methods, however, are not site-specific (rather, they are residue-specific and result in global replacement), do not usually allow for continuous cell growth, and are only generally applicable to UAAs that are close analogs of canonical amino acids.

Orthogonal aminoacyl-tRNA synthetase/iso-tRNA (aaRS/iso-tRNA) pair: an aaRS/iso-tRNA pair that is specific for its cognate amino acid and does not exhibit cross-reactivity with other aaRS/iso-tRNA pairs

EF: elongation factor

RF: release factor

TRANSLATION WITH NEW AMINO ACIDS

To further enhance our ability to control the structure and properties of proteins, both in vitro and in the context of living cells, we sought to directly create organisms that genetically encode 21 or more amino acids. In this approach, the biological, chemical, or physical properties of new amino acids are precisely defined by the chemist at the bench, but because these UAAs are genetically encoded, their incorporation into proteins should occur with the same fidelity, efficiency, and genetic manipulability of natural protein synthesis. To realize this goal, we developed a general method for the engineering and direct integration of aaRS and tRNA components into the translational machinery.

Translation is a unique biological process in which mRNA templates the assembly of a distinct biopolymer, a polypeptide chain, through a tRNA adapter molecule. This process is different from the synthesis of DNA and RNA where the same basic recognition elements are required in the template and product. As a consequence, translation has a greater intrinsic capacity to be adapted to accommodate new building blocks. The relationship of template (mRNA) to product (polypeptide) is defined by the genetic code, which assigns a specific amino acid to each triplet codon and utilizes aminoacyl-tRNA adapters to establish the map between mRNA and protein sequence. The fidelity and efficiency of translation rely on numerous molecular recognition steps. First, the 20 amino acids are specifically loaded onto 20 isoacceptor tRNA (iso-tRNA) sets by aaRSs, each specific for its own unique amino acid substrate and its own unique set of iso-tRNAs (56–58). This results in a network of orthogonal aaRS/iso-tRNA pairs (that is, a given aaRS, which is specific for its cognate amino acid, will only recognize its cognate iso-tRNA set, which itself is only recognized by the given aaRS) (59). Next, aminoacyl-tRNAs enter the ribosome and recognize the correct mRNA codons. This is accomplished by elongation factor Tu (EF-Tu), which binds and transports the range of aminoacyl-tRNAs into the ribosome, where standard (and sometimes wobble) base pairing mediates the recognition between an incoming aminoacyl-tRNA's anticodon loop and the mRNA codon being read. Correct anticodon-codon pairing results in a peptidyl transfer reaction between the incoming aminoacyl-tRNA and the growing polypeptide chain, a reaction catalyzed by the ribosome's peptidyl transfer center. Finally, translocation and eventually release factor (RF) binding continue and ultimately terminate translation of the desired protein product.

One can expand the genetic code of an organism to include new amino acids by adding new components to this template-directed biosynthetic machinery. These include a cell-permeable or biosynthesized UAA, a unique codon, a corresponding iso-tRNA set (in this case, an iso-tRNA set with cardinality one is simplest), and a cognate aaRS. These components must satisfy a number of criteria: First, the UAA must be metabolically stable and have good cellular bioavailability; it must be tolerated by EF-Tu and the ribosome, but it must not be a substrate for any endogenous aaRSs. Second, the unique codon must be recognized by the new tRNA but not by any endogenous tRNAs. Third, the aaRS/tRNA pair must be specific for the UAA, functional in the host organism, and orthogonal in the context of all endogenous aaRS/tRNA pairs in the organism (60). In general, most UAAs added to the media are taken up by both prokaryotic and eukaryotic cells (exceptions include highly charged amino acids, which can be modified as metabolically labile derivatives or incorporated into dipeptides to increase permeability).

In addition, natural aaRSs have evolved high specificity for their cognate amino acids, and the aminoacyl-binding site of EF-Tu and the ribosome are highly promiscuous (in vitro and in vivo translation experiments have demonstrated that the range of acceptable substrates includes many noncanonical amino acid side chains as well as D-amino acids and even α-hydroxy acids where ester rather than amide backbone bonds are formed) (61). Thus, the first criterion is easily met for most UAAs. Because codon recognition is determined by simple base pairing rules, the second criterion is easily fulfilled by choosing a blank (nonsense, frameshift, or otherwise unused) codon and designing a tRNA with the corresponding anticodon. The real challenge is to fulfill the third criterion, aaRS/tRNA orthogonality and aminoacylation specificity.

Although several strategies to generate orthogonal aaRS/tRNA pairs have been explored, ultimately, the most straightforward solution involves the importation of a heterologous aaRS/tRNA pair from a different domain of life. This is because tRNA recognition by aaRSs can be domain or species specific (62), a feature that can serve as the basis for orthogonality. The anticodon loop of the imported tRNA is then mutated to create a blank codon ($codon_{BL}$) suppressor tRNA ($tRNA_{SB}$), and the orthogonality of the resulting aaRS/$tRNA_{SB}$ pair is assessed. If necessary, the orthogonality of this pair is improved by a two-step process, involving both positive and negative rounds of selection to identify functional optimized $tRNA_{SB}$s that exhibit no cross-reactivity with endogenous aaRSs (60, 63). Finally, structure-based mutagenesis and a similar two-step selection strategy are used to alter the specificity of the heterologous aaRS so that it uniquely recognizes the UAA of interest (60, 64, 65). This process has allowed for the systematic directed evolution of aaRS/tRNA pairs that are specific for a variety of UAAs (**Figure 1**) and are orthogonal in bacteria, yeast, and mammalian cells (**Table 1**). This approach should, in theory, make translation with UAAs accessible in any organism.

Encoding Unnatural Amino Acids in *Escherichia coli*

$Codon_{BL}$: any codon that does not encode a natural amino acid for protein synthesis

$tRNA_{SB}$: a tRNA that suppresses a blank codon

To genetically encode UAAs in *E. coli*, one first imports a heterologous aaRS/tRNA pair from archea or eukaryotes, mutates the anticodon loop to generate $tRNA_{SB}$, and if necessary, uses mutagenesis and selection to improve the orthogonality of the imported pair. Selection is achieved by transforming a library of mutant $tRNA_{SB}$s (which is based on the heterologous $tRNA_{SB}$) into *E. coli* cells that contain the toxic barnase gene with a $codon_{BL}$ at permissive sites. In this negative selection step, only clones containing mutant $tRNA_{SB}$s that are not substrates for endogenous aaRSs grow. The surviving $tRNA_{SB}$s are then transformed into *E. coli* cells that express the heterologous aaRS and have a β-lactamase gene with a $codon_{BL}$ at a permissive site. In this positive selection step, the presence of ampicillin in the growth media kills all clones that contain a nonfunctional $tRNA_{SB}$, leaving only $tRNA_{SB}$ mutants that are aminoacylated by the heterologous aaRS. The result is a $tRNA_{SB}$ that functions with its cognate aaRS as a highly orthogonal pair in *E. coli*.

A similar two-step selection scheme is then used to alter the specificity of the heterologous aaRS so that it uniquely recognizes the UAA of interest (**Figure 2*a***). First, a library of aaRS mutants, containing randomized residues in the amino acid–binding site, is constructed on the basis of available crystal structures. This library is transformed into *E. coli* cells that express $tRNA_{SB}$ and a gene encoding chloramphenicol acetyltransferase (CAT) with a $codon_{BL}$ mutation at a permissive site. In this positive selection step, these cells are grown in the presence of chloramphenicol and the UAA of interest so that only the aaRS mutants capable of aminoacylating $tRNA_{SB}$ with the UAA and/or endogenous amino acids live. Surviving mutants are then transformed into *E. coli* cells that express $tRNA_{SB}$ and the toxic barnase gene with $codon_{BL}$ mutations at permissive sites. In this negative selection step, cells are grown in the absence of the UAA so that all clones whose mutant aaRS aminoacylates endogenous

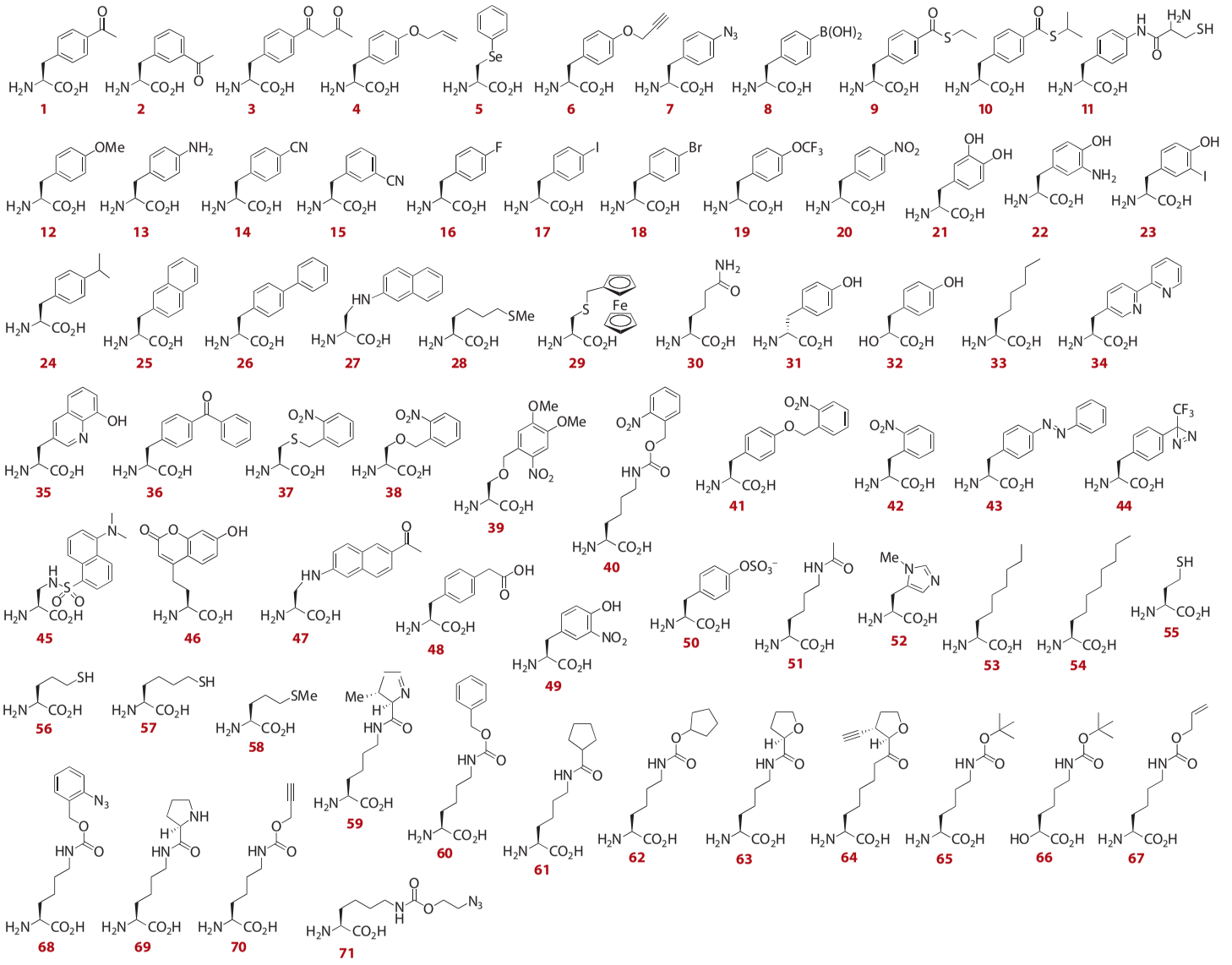

Figure 1

Chemical structures of genetically encoded unnatural amino acids (and hydroxy acids in the case of molecules 32 and 66).

Table 1 Summary of unnatural amino acids incorporated via expanded genetic codes

Unnatural amino acid (UAA)	Common name (if applicable)	Organism(s) in which UAA is encoded[a]	References[b,c] (and notes)
1	*p*-Acetylphenylalanine	*E. coli*, yeast, mammalian cells	66, 123, 131, 133
2	*m*-Acetylphenylalanine	*E. coli*	67
3		*E. coli*	68
4	*O*-allyltyrosine	*E. coli*	69
5	Phenylselenocysteine	*E. coli*	70 (precursor to dehydroalanine)
6	*p*-Propargyloxyphenylalanine	*E. coli*, yeast, mammalian cells	71, 125, 131, 133
7	*p*-Azidophenylalanine	*E. coli*, yeast, mammalian cells	72, 123, 125, 131, 133
8	*p*-Boronophenylalanine	*E. coli*	73
9		*E. coli*	H. Zeng & P.G. Schultz, unpublished
10		*E. coli*	H. Zeng & P.G. Schultz, unpublished
11		*E. coli*	M. Jahnz & P.G. Schultz, unpublished
12	*O*-methyltyrosine	*E. coli*, yeast, mammalian cells	74, 123, 124, 131, 133, 134
13	*p*-Aminophenylalanine	*E. coli*	64, 75
14	*p*-Cyanophenylalanine	*E. coli*	76
15	*m*-Cyanophenylalanine	*E. coli*	J. Chittuluru & P.G. Schultz, unpublished
16	*p*-Fluorophenylalanine	*E. coli*	100 (requires auxotrophic strain)
17	*p*-Iodophenylalanine	*E. coli*, yeast, mammalian cells	78, 123, 131, 133
18	*p*-Bromophenylalanine	*E. coli*	77, 78, 139
19		*E. coli*	79
20	*p*-Nitrophenylalanine	*E. coli*	80
21	L-DOPA	*E. coli*	81
22	3-Aminotyrosine	*E. coli*	82
23	3-Iodotyrosine	*E. coli*, yeast, mammalian cells	83, 132
24	*p*-Isopropylphenylalanine	*E. coli*	64
25	3-(2-Naphthyl)alanine	*E. coli*	84
26	Biphenylalanine	*E. coli*	85
27		Yeast, mammalian cells	126
28		Yeast, mammalian cells	127
29		Yeast, mammalian cells	130
30	Homoglutamine	*E. coli*	104
31	D-tyrosine	*E. coli*	J. Guo & P.G. Schultz, unpublished
32	*p*-Hydroxyphenyllactic acid	*E. coli*	86 (requires a strain with disrupted *tyrB* and *aspC*)
33	2-Aminocaprylic acid	Yeast, mammalian cells	89, 127
34	Bipyridylalanine	*E. coli*	85
35	HQ-alanine	*E. coli*	87
36	*p*-Benzoylphenylalanine	*E. coli*, yeast, mammalian cells	88, 123, 131, 133, 134, 169
37	*o*-Nitrobenzylcysteine	Yeast, mammalian cells	124

(*Continued*)

Table 1 (*Continued*)

Unnatural amino acid (UAA)	Common name (if applicable)	Organism(s) in which UAA is encoded[a]	References[b,c] (and notes)
38	*o*-Nitrobenzylserine	Yeast, mammalian cells	N. Wu & P.G. Schultz, unpublished
39	4,5-Dimethoxy-2-nitrobenzylserine	Yeast, mammalian cells	128
40	*o*-Nitrobenzyllysine	*E. coli*, yeast, mammalian cells	120
41	*o*-Nitrobenzyltyrosine	*E. coli*	89
42	2-Nitrophenylalanine	*E. coli*	90
43		*E. coli*	91
44		*E. coli*	92
45	Dansylalanine	Yeast, mammalian cells	129, 134
46		*E. coli*	93
47		Yeast, mammalian cells	126; J. Guo, H.S. Lee, E.A. Lemke, R.D. Dimla, & P.G. Schultz, unpublished results
48	*p*-Carboxymethylphenylalanine	*E. coli*	94
49	3-Nitrotyrosine	*E. coli*	95
50	Sulfotyrosine	*E. coli*	96
51	Acetyllysine	*E. coli*, yeast, mammalian cells	114, 121
52	Methylhistidine	Yeast, mammalian cells	F. Peters, J. Chittuluru, & P.G. Schultz, unpublished
53	2-Aminononanoic acid	Yeast, mammalian cells	127
54	2-Aminodecanoic acid	Yeast, mammalian cells	127
55		Yeast, mammalian cells	127
56		Yeast, mammalian cells	127
57		Yeast, mammalian cells	127
58		Yeast, mammalian cells	127
59	Pyrrolysine	*E. coli*, yeast, mammalian cells	111 (UAARS from a natural expanded genetic code)
60	Cbz-lysine	*E. coli*, yeast, mammalian cells	112, 114
61		*E. coli*, yeast, mammalian cells	115
62		*E. coli*, yeast, mammalian cells	116, 120
63		*E. coli*, yeast, mammalian cells	115, 117
64		*E. coli*, yeast, mammalian cells	117
65	Boc-lysine	*E. coli*, yeast, mammalian cells	112, 114
66		*E. coli*, yeast, mammalian cells	118
67	Allyloxycarbonyllysine	*E. coli*, yeast, mammalian cells	112
68		*E. coli*, yeast, mammalian cells	112
69		*E. coli*, yeast, mammalian cells	116
70		*E. coli*, yeast, mammalian cells	119
71		*E. coli*, yeast, mammalian cells	119

[a] Underlined font, functionality not experimentally demonstrated but based on parent aminoacyl-tRNA synthetase (aaRS)/tRNA pair orthogonality.
[b] References are those pertinent to the original encoding of the UAA.
[c] Abbreviation: UAARS, UAA-specific mutant aminoacyl-tRNA synthetase.

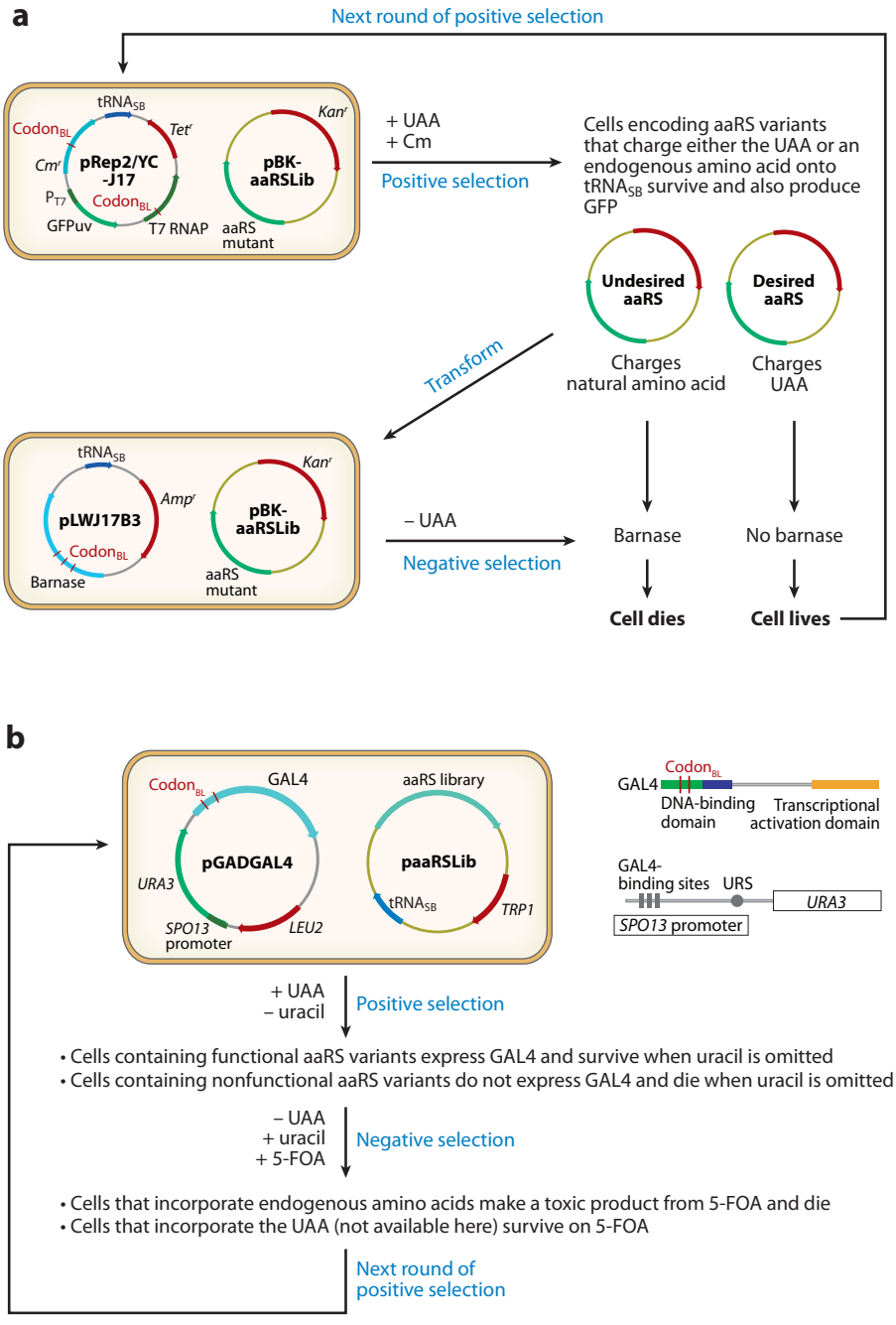

Figure 2

Selection schemes for genetically encoding amino acids in *E. coli* and yeast. (*a*) Two-step selection for the directed evolution of orthogonal UAARS/$tRNA_{SB}$ pairs in *E. coli*. (*b*) Two-step selection for the directed evolution of orthogonal UAARS/$tRNA_{SB}$ pairs in yeast. Abbreviations: aaRS, aminoacyl-tRNA synthetase; $codon_{BL}$, blank codon; 5-FOA, 5-fluoroorotic acid; GFP_{UV}, green fluorescent protein; P_{T7}, bacteriophage T7 promoter; T7 RNAP, bacteriophage T7 RNA polymerase; $tRNA_{SB}$, suppressor tRNA; UAA, unnatural amino acid; UAARS, UAA-specific mutant aaRS; URS, upstream repression sequence.

UAARS: an aminoacyl-tRNA synthetase specific for an unnatural amino acid substrate

amino acids die. This leaves only mutant aaRSs that aminoacylate $tRNA_{SB}$ with the UAA. The two-step selection is usually repeated for two to three additional rounds to yield an UAA-specific mutant aaRS (UAARS). We can assume that the UAARS does not aminoacylate endogenous tRNAs to any appreciable extent as this would lead to missense incorporation throughout the proteome, resulting in cell death. Through this selection scheme and its more facile variants (64, 65), many structurally distinct amino acids have been genetically encoded in *E. coli*, most successfully using the amber (TAG) codon as $codon_{BL}$ and importing the heterologous *Methanocaldococcus jannaschii* $TyrRS/tRNA^{Tyr}$ (*Mj*TyrRS/*Mj*$tRNA^{Tyr}$) pair from which ~38 orthogonal $UAARS/tRNA_{SB}$ pairs specific for chemically distinct UAAs (e.g., UAAs **1–15**, **17–26**, **31**, **32**, **34–36**, **41–44**, **46**, and **48–50**, shown in **Figure 1** and **Table 1**) have been derived (64, 66–96). In general, these UAAs are incorporated with excellent fidelity and yields. Indeed, shake-flask expression yields in the 1–100 mg/L range and/or optimized high-density fermentation yields in the g/L range have been achieved for almost all UAA-containing mutant proteins. However, mRNA context effects, protein folding and stability, and other factors can lead to low incorporation efficiency at some desired sites; these areas remain an opportunity for methodological improvements.

Although evolution of orthogonal $UAARS/tRNA_{SB}$ pairs from the *Mj*TyrRS/*Mj*$tRNA^{Tyr}$ pair has been very successful, the ability to evolve aaRSs specific for any given amino acid depends on structural constraints in the active site. Thus, it may be necessary to consider other aaRS/tRNA pairs with structurally diverse active sites as starting points. Toward this end, several additional orthogonal pairs have been adapted for use in *E. coli*. These include pairs from *Saccharomyces cerevisiae* (e.g., AspRS/$tRNA^{Asp}$, GlnRS/$tRNA^{Gln}$, TyrRS/$tRNA^{Tyr}$, and PheRS/$tRNA^{Phe}$) (97–100), pairs from archea (e.g., *Mj*TyrRS/*Mj*$tRNA^{Tyr}$) (101), and pairs with hybrid or consensus components (e.g., *S. cerevisiae* TyrRS/*E. coli* initiator $tRNA_f^{Met}$, *Methanobacterium thermoautotrophicum* LeuRS/*Halobacterium* sp. *NRC-1* $tRNA^{Leu}$, *Pyrococcus horikoshii* or *Methanosarcina mazei* GluRS/consensus tRNA derived from archael $tRNA^{Glu}$ sequences, and *P. horikoshii* LysRS/consensus tRNA from archael $tRNA^{Lys}$ sequences) (102–105). The utility of each pair is dependent on efficient expression in bacteria, orthogonality, ability to suppress the desired $codon_{BL}$ (once the tRNA's anticodon loop is changed), and ability to evolve the aaRS active site to accommodate novel amino acids. All of these characteristics can be optimized through promoter engineering, rational mutations, and selection. Recently, some pairs have also been imported, without loss of orthogonality, into *Mycobacteria smegmatis*, *bovis*, and *tuberculosis*, thus allowing the genetic incorporation of UAAs into these organisms to study their pathogenesis and produce vaccines (106).

An additional set of heterologous pairs has been added to this list with the recent characterization of naturally occurring 21 amino acid methanogens that genetically encode pyrrolysine (107, 108). These organisms contain a pyrrolysyl-tRNA synthetase (PylRS) that aminoacylates $tRNA^{Pyl}_{CUA}$ with pyrrolysine; pyrrolysyl-$tRNA^{Pyl}_{CUA}$ in turn incorporates this twenty-first amino acid in response to the amber codon. In its native contexts, the PylRS/$tRNA^{Pyl}_{CUA}$ pair functions alongside the 20 canonical aaRS/iso-tRNA set pairs without any cross-reactivity among pairs. Therefore, its importation into other organisms might also preserve orthogonality, especially given that the ancient pre-LUCA PylRS likely underwent significant horizontal gene transfer among other aaRS/tRNA pairs in early life (109, 110). Indeed, this has so far been the case. For instance, importation of the *Methanosarcina barkeri* PylRS/*M. barkeri* $tRNA^{Pyl}$, *M. mazei* PylRS/*M. mazei* $tRNA^{Pyl}$, and *Desulfitobacterium hafniense* PylRS/*D. hafniense* $tRNA^{Pyl}$ pairs (and certain hybrid pair combinations thereof) into *E. coli* and mammalian cells maintains orthogonality without requiring further modification (111–114). This has allowed pyrrolysine (**59**) and

many of its analogs (e.g., UAAs **61–67** and **69–71**) to be genetically encoded in bacteria (111, 112, 114–119). In addition, rational mutagenesis/screening or the general two-step selection scheme has been used to alter the specificity of PylRS/$tRNA^{Pyl}_{CUA}$ pairs to encode structurally distinct amino acids (e.g., UAAs **40**, **51**, **60**, and **68**), including a photocaged lysine and acetyllysine (112, 114, 120, 121).

The identification of multiple heterologous aaRS/tRNA pairs orthogonal in *E. coli* also makes possible the simultaneous encoding of two or more UAAs, which requires two or more different blank codons recognized by two or more mutually orthogonal aaRS/tRNA pairs. An example is the use of both the AGGA frameshift and TAG nonsense codons with corresponding orthogonal UAARS/$tRNA_{SB}$ pairs derived from the *P. horikoshii* LysRS/tRNA (a consensus tRNA from archael $tRNA^{Lys}$ sequences) and the *Mj*TyrRS/*Mj*$tRNA^{Tyr}$ pairs to site specifically incorporate two distinct UAAs (**30** and **12**, respectively) into a single myoglobin protein in *E. coli* (104). In this case, it was not necessary to explicitly engineer mutual orthogonality, probably because the two UAARSs were derived from pairs that specify different natural amino acids. If necessary, one could evolve orthogonality of one pair in the context of the other and vice versa by using the general two-step selection scheme described above.

Encoding Unnatural Amino Acids in Yeast

The processing, expression, and recognition of tRNAs in yeast are distinct from those in *E. coli*. Therefore, new heterologous aaRS/tRNA pairs are needed to expand the genetic code of yeast. Toward this end, several aaRS/tRNA pairs, including the *E. coli* GluRS/human initiator tRNA, the *E. coli* TyrRS/*E. coli* $tRNA^{Tyr}$, the *E. coli* LeuRS/*E. coli* $tRNA^{Leu}$, and the *M. mazei* PylRS/*M. mazei* $tRNA^{Pyl}$ pairs (102, 114, 122–124), have been shown to be orthogonal in *S. cerevisiae*. To change the amino acid specificity of the aaRS/tRNA pairs for incorporation of UAAs in response to $codon_{BL}$, a two-step selection similar to that used in *E. coli* is applied (**Figure 2*b***) (123). First, a library of aaRS mutants is generated and transformed into a *S. cerevisiae* uracil auxotrophic strain containing a GAL4 transcriptional activator with $codon_{BL}$ at two permissive sites, a *URA3* gene (required for uracil synthesis) under the control of GAL4, and the heterologous tRNA with its anticodon mutated to suppress $codon_{BL}$ ($tRNA_{SB}$). In positive selection, these cells are grown in the presence of UAA and in the absence of uracil. Clones expressing aaRS mutants that can aminoacylate $tRNA_{SB}$ with the UAA and/or natural amino acids produce GAL4 by suppressing $codon_{BL}$, which results in uracil synthesis and survival. This is followed by negative selection, where surviving cells are grown in the absence of UAA but in the presence of uracil and 5-fluoroorotic acid (5-FOA). When GAL4 is produced, *URA3* is expressed, which converts 5-FOA into a toxic product, killing the cell. Therefore, those clones containing aaRS mutants that aminoacylate $tRNA_{SB}$ with natural amino acids express *URA3* and die. The clones that survive both steps of selection are ones that aminoacylate $tRNA_{SB}$ exclusively with the desired UAA. This process is often repeated for several rounds to yield orthogonal UAARS/$tRNA_{SB}$ pairs specific for the UAA of interest.

In this way, multiple UAARS/$tRNA_{SB}$ pairs based on both the *E. coli* TyrRS/*E. coli* $tRNA^{Tyr}$ and *E. coli* LeuRS/*E. coli* $tRNA^{Leu}$ pairs have been evolved, resulting in the addition of ~22 UAAs to the yeast genetic code (e.g., UAAs **1**, **6**, **7**, **12**, **17**, **27–29**, **33**, **36–39**, **45**, **47**, and **52–58**) (123–130). Furthermore, these evolved UAARS/$tRNA_{SB}$ pairs retain their activity and orthogonality when imported into the methylotrophic yeast *Pichia pastoris*, a host that facilitates large-scale recombinant protein expression and control over glycosylation patterns. Indeed, expression systems containing optimized transcriptional control elements have been created for the transfer of UAARS/$tRNA_{SB}$ pairs from *S. cerevisiae* to *P. pastoris*. In one case, these systems were

used to produce human serum albumin containing the UAA **1** in shake-flask expression yields of >100 mg/L (131).

Encoding Unnatural Amino Acids in Mammalian Cells

In mammalian cells, directed evolution efforts are severely limited by technical challenges associated with transformation efficiency, slow doubling times, and growth conditions; therefore, the two-step selection strategy is not ideal for evolving aaRSs with altered specificities here. Instead, there are a number of aaRS/tRNA pairs orthogonal both in *E. coli* or *S. cerevisiae* and in mammalian cells. Thus, evolution of UAA specificity can be carried out in *E. coli* or *S. cerevisiae*, where large libraries can be constructed. The resulting UAARS/$tRNA_{SB}$ pairs can then be transferred into mammalian cells without loss of orthogonality. This shuttle approach has been used for several UAARS/$tRNA_{SB}$ pairs, including ones derived from the *M. mazei* PylRS/*M. mazei* $tRNA^{Pyl}$ pair, which is orthogonal in both *E. coli* and mammalian cells (114, 120); and ones derived from the *E. coli* TyrRS/*Bacillus stearothermophilus* $tRNA^{Tyr}$ pair, which is orthogonal in eukaryotes (132, 133). In the latter case, *B. stearothermophilus* $tRNA^{Tyr}$ is used rather than *E. coli* $tRNA^{Tyr}$ because it contains intact internal A- and B-box promoters that drive proper tRNA expression in mammalian cells (132). More recently, use of external H1 and U6 promoters and 3′ flanking regions derived from mammalian sequences has led to the general production of functional heterologous tRNAs (e.g., *E. coli* $tRNA^{Tyr}$ and *E. coli* $tRNA^{Leu}$) in mammalian cells, thus amending the list of shuttle UAARS/$tRNA_{SB}$ pairs with ones derived from the *E. coli* TyrRS/*E. coli* $tRNA^{Tyr}$ and the *E. coli* LeuRS/*E. coli* $tRNA^{Leu}$ pairs (134; W. Liu, J. Guo, & P.G. Schultz, unpublished results). In this way, the genetic codes of mammalian cells including CHO cells, 293T cells, and even primary neurons have been augmented with ~10 structurally diverse UAAs (e.g., UAAs **1**, **6**, **7**, **12**, **17**, **36**, **45**, **47**, and **52**) incorporated by UAARS/$tRNA_{SB}$ pairs evolved in *S. cerevisiae* (133, 134; J. Guo, H.S. Lee, E.A. Lemke, R.D. Dimla, & P.G. Schultz, unpublished results). In addition, several lysine and pyrrolysine derivatives (e.g., UAAs **40**, **51**, **60**, **62**, and **65**) have been incorporated into proteins in mammalian cells by UAARS/$tRNA_{SB}$ pairs derived from the PylRS/$tRNA^{Pyl}$ pairs (114, 120). For these UAAs, laboratory mammalian expression systems result in protein yields in the microgram per 10^7 cells range with no misincorporation detected. These orthogonal pairs are also being transferred to *Caenorhabditis elegans* and *Mus musculus* cells with the aim of creating whole multicellular organisms with expanded genetic codes that allow the cotranslational incorporation of UAAs for in vivo biological studies.

Other Methods for Genetically Encoding Unnatural Amino Acids

Alternative approaches have also been used to genetically encode UAAs. For example, rational mutagenesis of heterologous aaRS/$tRNA_{SB}$ pairs followed by their use in auxotrophic *E. coli* strains has allowed the site-specific incorporation of several UAAs (e.g., tryptophan analogs, UAA **16** and UAA **18**) (100, 135, 136). In this approach, a largely orthogonal yeast aaRS/tRNA pair, modified to suppress a blank codon, is imported into *E. coli*. The yeast aaRS recognizes the UAA of interest but also recognizes one or more of the common 20 amino acids, often owing to limitations on rational design. To favor incorporation of the UAA, an auxotrophic *E. coli* strain that does not biosynthesize the corresponding common amino acid is used so that its concentration can be strictly limited. Even though an auxotrophic strain is necessary, the UAA does not replace a common amino acid throughout the proteome but rather is separately encoded. Another approach uses a rationally designed orthogonal aaRS/tRNA pair that recognizes 3-iodotyrosine (**23**) and site specifically incorporates it into proteins in response to the amber codon in mammalian cells (132). To further favor incorporation of UAA **23** over

tyrosine, the editing domain from PheRS was fused to the mutant aaRS (137). These methods can be quite useful, but they are often limited by moderate aaRS/tRNA pair orthogonality, modest UAA specificity, and difficulties inherent to rational design, especially in addressing UAAs that have structural features significantly different from the canonical amino acids. Therefore, they have mostly been improved or supplanted by directed evolution methods.

Recognition of Unnatural Amino Acids by Evolved Aminoacyl-tRNA Synthetases

To understand how the evolved UAA-specific synthetases recognize their substrates, the X-ray crystal structures of a number of aaRSs have been solved. For UAARSs based on TyrRS, these studies reveal that this class I synthetase contains an amino acid–binding pocket with a high degree of tolerance for substitutions. For the most part, substitutions that arise from the two-step selection scheme alter the side chains that line the amino acid–binding pocket to increase complementarity with UAAs while decreasing complementarity with endogenous amino acids. For example, the UAARS specific for *p*-benzoylphenylalanine (**36**), evolved in *E. coli* from the heterologous *Mj*TyrRS, accommodates the large side chain of UAA **36** primarily through the mutation Tyr32Gly, which deepens the substrate-binding pocket (138). In some cases, substitutions can also change the conformation of the protein backbone, suggesting a significant degree of structural plasticity. For example, in the UAARS specific for *p*-bromophenylalanine (**18**), evolved from *Mj*TyrRS, the Asp158Pro mutation terminates the α_8-helix, flipping residues 158, 159, and 162, which normally line the binding pocket, outward (139). Mutations that accommodate UAAs are often selected together with ones that disfavor binding of natural amino acids. For example, in the UAARS specific for *p*-acetylphenylalanine (**1**), evolved from *Mj*TyrRS, substitutions Tyr32Leu and Asp158Gly have the dual effect of removing hydrogen bonding with tyrosine's phenolic hydroxyl group and opening up the binding pocket to accommodate the acetyl group of UAA **1**. To further increase affinity for UAA **1**, the unmutated residue Gln109 forms a hydrogen bond with the acetyl group (140).

Perhaps the best illustration of aaRS tolerance to substitutions is the selection of an UAARS specific for bipyridylalanine (**34**) (85). In this case, a library of mutants based on *Mj*TyrRS did not directly yield a aaRS specific for bipyridylalanine but did yield one specific for the isostere biphenylalanine (**26**). The crystal structure of the biphenylalanine-specific UAARS bound to UAA **26** was solved, and the structure served as a guide for the design of a new library of aaRS mutants, which was subjected to further selection for incorporation of UAA **34**. Though one may expect that the accumulation of additional substitutions should reduce synthetase stability and activity, this second selection yielded an active bipyridylalanine-specific UAARS, suggesting that the thermophilic *Mj*TyrRS is highly tolerant to mutations in its amino acid–binding domain. This explains, at least in part, why a large number of UAARSs have been successfully evolved from the heterologous *Mj*TyrRS.

Taken together, these structural studies support the use of structure-based library design and selection and suggest that successful selection of new synthetases from libraries based on known structures of *Mj*TyrRS, UAARSs, and the class I *E. coli* LeuRS (from which several UAARSs have been evolved but for which crystal structures are few) (141) will continue. Furthermore, structural and phylogenetic studies on the class II PylRS suggest that its specificity for pyrrolysine is achieved by interactions localized to its side chain–binding pocket (109); therefore, aaRS libraries based on PylRS should also yield new amino acid specificities, a prediction that recent experiments have already begun to confirm. Nonetheless, despite the fact that aaRSs have been engineered to recognize rather large, structurally complex amino acids (including those with ferrocenyl, dimethoxynitrobenzyl seryl, dansyl, and bipyridyl side chains), the

development of strategies and aaRS libraries to accept even greater structural diversity remains a challenge.

Blank Codons

Amber nonsense codons and, to a lesser extent, ochre, opal, or frameshift codons have been used to specify UAAs. In each case, suppression is in competition with other processes, such as RF binding to nonsense codons or recognition of frameshift codons by tRNAs with three-base anticodons, both of which lead to decreased cotranslational incorporation efficiency. Although a number of efforts are employed to mitigate these undesired effects—examples include the overproduction of $tRNA_{SB}$ suppressors, synonymous mutation of codons surrounding $codon_{BL}$, selection of efficient suppressors of frameshift codons that compete well with embedded rare three-base codons, and modulation of RF levels (142–145)—and other efforts have tried to circumvent them—for example, a sense codon was used to encode an UAA in an auxotrophic strain by exploiting wobble position stability differences and limiting endogenous amino acid concentrations (146)—a general solution that involves the conversion of sense codons into blank codons would be ideal.

In theory, such a strategy is feasible with genome synthesis, whereby certain sense codons in an organism are replaced in a wholesale manner through synonymous mutations. Although effects on mRNA folding and translation initiation rates would have to be considered (147, 148), the decreased degeneracy of the code would allow the reassignment of unused sense codons to UAAs. Such an approach is currently being carried out in *S. cerevisiae* and, if successful, would not only avoid the inefficiencies associated with nonsense and frameshift suppression, but would also allow several UAAs to be simultaneously encoded through mutually orthogonal UAARS/$tRNA_{SB}$ pairs. Such strains might allow completely unnatural polypeptides to be made ribosomally. An alternative approach that uses unnatural base pairs in addition to A-T and C-G could also provide new blank codons, in this case by extending the number of possible three-base codons beyond 64 (149–151). Although the inefficiency of unnatural base pairs in replication, transcription, and translation has kept this approach from moving beyond simple in vitro expression systems, the modularity of genetic code elements should allow for the direct integration of unnatural codons once these problems are solved.

Optimized Systems for Protein Expression

The practical utility of organisms with 21 (or greater) amino acid codes is dependent on the efficiency and fidelity with which proteins containing UAAs can be expressed. Therefore, several improvements to the components involved in cotranslational UAA incorporation have been made. The most basic of these involve optimization of expression vectors. For example, in one case, a 20-fold increase in the yield of an UAA-containing myoglobin protein was observed in *E. coli* when the original two-plasmid system encoding *M. jannaschii* $tRNA_{CUA}$ (*Mj*$tRNA_{CUA}$, an amber $tRNA_{SB}$) under the control of the *lpp* promoter and an UAARS derived from *Mj*TyrRS (*Mj*UAARS) under the control of the *glnS* promoter was replaced by a single-plasmid system encoding 3 or 6 copies of *Mj*$tRNA_{CUA}$ under the control of the *ProK* promoter (this promoter drives endogenous $tRNA^{Pro}$ expression in *E. coli*) and the *Mj*UAARS under the control of a strong mutant *glnS* promoter (142). This plasmid design (pSup) was further improved by encoding an additional copy of the *Mj*UAARS under the control of inducible *ara* and *T7* promoters (79). The resulting pSUPAR plasmids afford shake-flask expression yields ranging from ~40 mg/L to ~65 mg/L for UAA containing the FAS-TE (fatty acid synthase thioesterase domain) as well as high yields for many other proteins tested (although yields do vary based on the individual proteins and codon context effects) (152). Plasmid system improvements have also been made for UAA incorporation in *S. cerevisiae*

by increasing promoter strengths, raising plasmid copy number, and using optimized codons (153). For example, a particularly effective optimization effort significantly increased the yield of an UAA-containing protein over earlier systems by placing *E. coli* $tRNA^{Leu}_{CUA}$ (an amber $tRNA_{SB}$ that is part of an orthogonal pair in yeast) under the control of an external promoter containing A- and B-box sequences (154). Expression was conducted in a nonsense-mediated decay-deficient *S. cerevisiae* strain, which prevents the degradation of mRNAs that contain premature stop codons.

Synthetase mutations and tRNA optimization have resulted in further improvements in these expression systems. For example, a T252A mutation in the editing domain of *E. coli* LeuRS variants increases the fidelity of UAA incorporation by reducing net aminoacylation with an undesired common amino acid (129), and a D286R substitution in *Mj*TyrRS variants increases the efficiency of *Mj*UAARS/*Mj*$tRNA_{CUA}$ pairs by mediating better recognition of the CUA anticodon (155). In addition, *Mj*$tRNA_{CUA}$ has been engineered for higher expression levels of UAA-containing proteins by randomizing tRNA T-stem nucleotides at the EF-Tu-binding interface and conducting stringent selection for increased amber suppression efficiency (156). The resulting selected tRNAs are in general less toxic to *E. coli* and in some cases enhance expression efficiency in an UAA-dependent manner. Some of these resulting tRNAs have been combined with the pSUPAR plasmid design to achieve UAA-containing protein overexpression yields that approach wild-type levels (wild-type refers to the analogous proteins where the UAA is replaced by a common amino acid) (157).

Though more challenging, elongation factors and ribosomes themselves can also be optimized for UAA incorporation. For example, the recent development of orthogonal ribosomes that only translate mRNAs containing artificial 5′ sequences allows for ribosome evolution by decoupling mutant and endogenous ribosome action (158). The use of this approach yielded the orthogonal ribosome ribo-X, which contains mutations that likely decrease its interaction with RF-1, thereby preventing early termination and increasing the efficiency of translation with UAAs (159). It is also possible to develop EF-Tu mutants that better tolerate UAAs, especially those with large side chains or altered backbones. This has been demonstrated in an in vitro translation system with rationally designed EF-Tu variants that have mutations in their aminoacyl-binding pockets (160). Although their activity for natural amino acids is lower than that of wild-type EF-Tu, these variants have increased activity for large aromatic UAAs, thus increasing the efficiency of their incorporation. Selection in cells with UAARS/$tRNA_{SB}$ pairs should allow tuning of EF-Tu-binding site variants to better accommodate UAAs. Other efforts, including the selection of optimal nucleotide contexts surrounding blank codons, random genome-wide mutagenesis and complementation to find other factors that affect UAA-incorporation efficiency, and enhancement of cellular uptake of charged UAAs by their delivery as dipeptides or precursors, are currently being explored.

GENETICALLY ENCODED UNNATURAL AMINO ACIDS AND THEIR APPLICATIONS

Clearly, the 20 common amino acids are sufficient for all known forms of life. However, if one assumes a substantial proteomic contribution to fitness during the solidification of the genetic code (Crick's "frozen accident" factor) (161, 162), then one must conclude that the identities of the 20 canonical amino acids are a consequence of factors present during early evolution (163, 164). Therefore, additional genetically encoded amino acids may offer evolutionary advantages to modern organisms and certainly extend our ability to manipulate the physicochemical and biological properties of proteins. To this end, we and others have genetically encoded UAAs representing an extensive range of structural and electronic properties not found in the common 20 amino acids (**Figure 1**).

UAAs with chemically reactive groups can be used as bioorthogonal handles for the site-specific in vitro, and in some cases intracellular, modification of proteins; they can also be used to introduce a new or enhanced catalytic function into proteins. Such amino acids include UAAs **1–11**, **17**, **18**, **64**, **67**, **68**, **70**, and **71**. Proteins containing these UAAs have been selectively modified with nonproteinogenic moieties using a variety of selective chemistries, including oxime condensation reactions, click chemistry, Michael addition reactions, and Suzuki couplings. Members of another class of UAAs represented by UAAs **12–33**, **36**, **44**, **45–47**, **53–63**, **65**, **66**, and **69** contain, in their side chains, in vitro or cellular probes of protein structure and function. These UAAs can be used as IR, NMR, and fluorescent probes, as well as redox-active reagents, heavy atoms for X-ray structure determination, and probes of hydrogen bonding and packing interactions in proteins. This class also includes photocross-linking amino acids, which can be used as in vitro or cellular probes of protein-protein and protein-nucleic acid interactions. UAAs can also be used to directly engineer into proteins new functions that may be challenging to generate with the common 20 amino acids. Such UAAs include metal ion-binding, photoisomerizable, photocaged, and photoreactive amino acids (**34–44**). Another class of UAAs is represented by UAAs **48–52**, which are all free amino acids that correspond to the product of a PTM. These UAAs allow the expression of proteins containing defined PTMs in simple hosts; lift the sequence restrictions to which PTMs must otherwise adhere; and have been used to express sulfated, nitrosylated, and methylated proteins. In addition, *p*-carboxymethylphenylalanine (**48**) was substituted in the transcription factor STAT1 as a stable mimetic of phosphotyrosine and overcame the complexities associated with removal of this PTM by phosphatases (94). Below, we overview a number of examples to illustrate the utility of genetically encoded UAAs in the study and manipulation of protein structure and function.

Probes of Protein Structure and Function

Many biophysical and mechanistic studies require significant quantities of proteins with a probe incorporated at a unique site in a protein. UAA mutagenesis methodology is well suited to many such problems. For example, in a recent application of UAAs as NMR probes, milligram quantities of site specifically labeled FAS-TE were produced from ≤25 mg of either $^{13}C/^{15}N$-labeled *O*-methyltyrosine (**12**), ^{15}N-labeled *o*-nitrobenzyltyrosine (**41**), or OCF_3Phe (**19**) (79). The resulting variants of FAS-TE (33 in total were generated) were then characterized by ^{1}H-^{13}C HSQC, ^{1}H-^{15}N HSQC, and ^{19}F NMR in the absence and presence of a small-molecule inhibitor. Using the isolated NMR signals of the UAA probes, these experiments identified the binding site of the inhibitor and revealed conformation changes of certain residues that occur upon binding. A particularly interesting aspect of these studies was the incorporation of a ^{15}N-labeled variant of the photoreactive UAA **41**, which upon irradiation decages to afford site-specific isotopic labeling at only one of the many tyrosines. When an active-site tyrosine was mutated to a ^{15}N-labeled UAA **41**, chemical shifts corresponding to ligand binding were observed only after the Tyr was decaged.

UAAs with unique IR and X-ray diffraction signatures have also been efficiently incorporated into proteins to study structure and dynamics. For example, *p*-cyanophenylalanine (**14**) contains a cyano group that absorbs in a clear spectral window at ~2200 cm^{-1} and is sensitive to subtle changes in local environment. It was incorporated into myoglobin to probe metal ion and ligand binding (76) and is also used as a Stark effect probe of local protein dipoles. Similarly, *p*-azidophenylalanine (**7**), which absorbs in a clear spectral window at ~2100 cm^{-1} and is sensitive to changes in electrostatic environment, was incorporated site specifically at various sites in rhodopsin. After expression and reconstitution into lipid membranes, changes in the stretch frequency of the

azido group upon light-induced receptor activation were analyzed by Fourier transform infrared (FTIR) difference spectroscopy for several rhodopsin mutants, revealing changes in the local environments of specific amino acids (165). In addition, solution of the X-ray crystal structures of *O*-acetylserine sulfhydrylase, *N*-acetyltransferase, and T4 lysozyme was facilitated by the site-specific incorporation of UAAs containing heavy atoms (Zn^{2+}-bound UAA **35**, **23**, and **17**, respectively) (78, 83, 87). Several other recent studies have also exploited UAA mutagenesis to selectively introduce novel biophysical probes. These include the dissection of radical propagation in ribonucleotide reductase with a redox-active amino acid (82), the exploration of K^+ channel inactivation by perturbing the diameter of the inactivation peptide (134), the mutation of backbone amides to esters in order to evaluate the contribution of hydrogen bonding interactions to protein stability (86, 118), and the study of protein dynamics in dihydrofolate reductase with deuterated amino acids site specifically incorporated by genetically encoded photocaged variants (166). In this last example, FTIR characterization of specifically placed C-D bonds allowed the time-resolved characterization of the microenvironment of Tyr100 upon substrate binding.

Because the genetic encoding of UAAs exploits the cells' own translational machinery, one can use genetically encoded biophysical probes to study processes both in vitro and in living cells. For example, one can genetically encode small fluorescent UAAs at surface sites in a protein with minimal structural perturbation (this is in contrast to traditional fluorescent genetic tags such as GFP and its variants, which are limited by their large size). Recently, such a strategy was used to incorporate UAA **47** into histones, whose nuclear localization in yeast and mammalian cells could then be visualized through fluorescence microscopy (**Figure 3*a***) (J. Guo, H.S. Lee, E.A. Lemke, R.D. Dimla, & P.G. Schultz, unpublished results). Furthermore, because the fluorescence properties of UAA **47** and other fluorescent UAAs are highly sensitive to environmental changes, and these

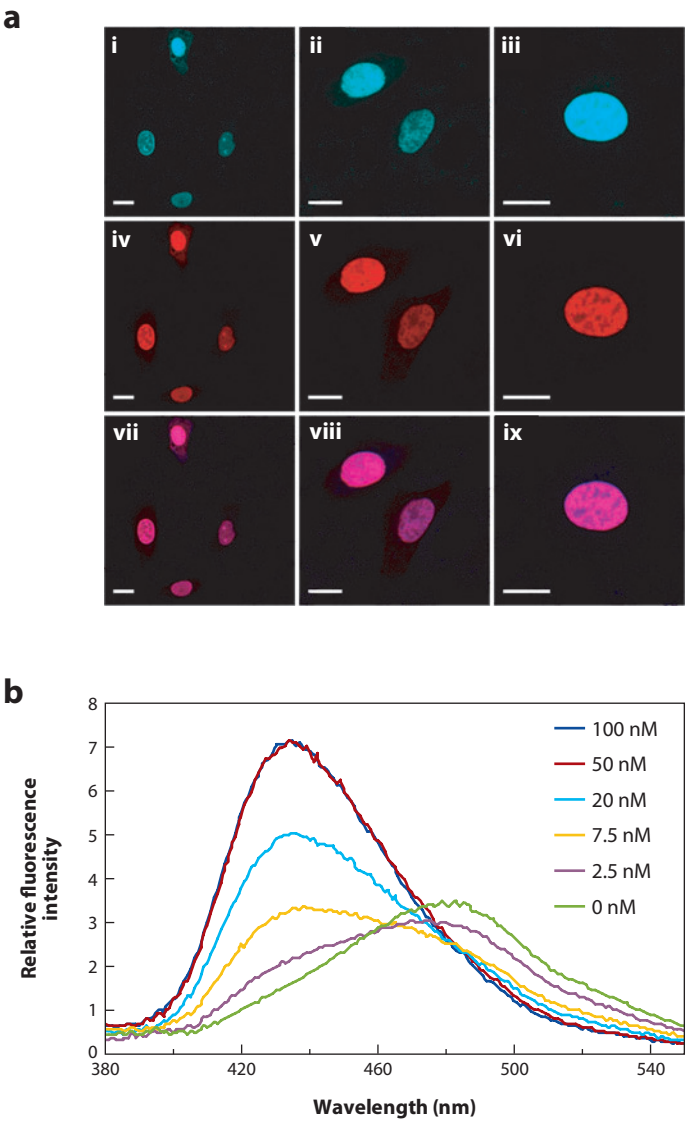

Figure 3

(*a*) Fluorescence images of labeled histone H3 in CHO cells acquired by confocal, two-photon microscopy. Each scale bar is 10 μm. (*i-iii*) CHO cells expressing an mCherry-labeled histone H3 with UAA **47** at position 59 (excitation, 730 nm, two photon; emission, 420–500 nm). (*iv-vi*) CHO cells expressing an mCherry-labeled histone H3 with UAA **47** at position 59 (excitation, 543 nm; emission, 600–700 nm). (*vii*) Composite of images *i* and *iv*. (*viii*) Composite images of *ii* and *v*. (*ix*) Composite of images *iii* and *vi*. (*b*) Fluorescence spectra of glutamine-binding protein with UAA **47** at position 160 and different concentrations of Gln (excitation, 350 nM).

amino acids can be selectively incorporated at or near a site of interest, this labeling method can be used for the characterization of local protein conformation changes, protein folding, and biomolecular interactions (87, 93, 126, 129, 167). Recent examples include the characterization of myoglobin unfolding using UAA **46** (e.g., urea-dependent unfolding of the amino terminus prior to global unfolding), and the detection of ligand binding by glutamine-binding protein (QBP) using UAA **47** (**Figure 3*b***). In the latter case, UAA **47** was placed in a deep cleft that lies between the two domains of QBP, which are connected by two peptide hinges. Upon binding of glutamine, significant conformational changes in this hinge region were directly revealed [without the use of a Förster resonance energy transfer (FRET) pair] by a significant shift in the emission λ_{max} and intensity of UAA **47**. The K_d of the mutant QBP for Gln was determined from these spectral changes and found to be close to that of wild-type protein.

In addition to fluorescent probes, one can use photocaged and photoisomerizable UAAs to study protein function. In vitro these UAAs have been used to photoregulate ligand-protein binding (91), site specifically cleave proteins (90), and probe protein-nucleic acid interactions (168). Such genetically encoded UAAs can also allow the study of cellular processes. For example, 4,5-dimethoxy-2-nitrobenzylserine (**39**), a photocaged serine, was encoded in the transcription factor Pho4 at either Ser114 (termed S2) or Ser128 (termed S3). These variants, fused to GFP, were localized in the nucleus when expressed in *S. cerevisiae* in low-phosphate media. Photolysis with a visible laser pulse (465 nm) decaged S2 or S3, leading to subsequent phosphorylation and export from the nucleus by Msn5 under P_i-rich conditions (**Figure 4*a***). The kinetics of this process were monitored in real time to better understand the activities of differentially phosphorylated Pho4 mutants (**Figure 4*b***) (128).

Another use of UAA incorporation is the characterization of protein-protein and protein-nucleic acid interactions in vitro or in living cells through photocross-linking. In many cases, biomolecular interactions are transient or unstable to in vitro isolation conditions, thus requiring covalent cross-linking of the interacting molecules to isolate relevant complexes. In a proof-of-principle study, *p*-benzoylphenylalanine (**36**), which upon irradiation inserts into nearby C-H bonds (or relaxes back to the ground state), was site specifically incorporated into the adapter protein Grb2. This protein mediates extracellular signaling to the Ras protein by binding various target molecules, including members of the epidermal growth factor (EGF) receptor family. When expressed in CHO cells together with the EGF receptor, in vivo light-induced cross-linking resulted in the identification of covalently linked Grb2-EGF receptor complexes (169). Furthermore, the efficiency of cross-linking varied with the position of *p*-benzoylphenylalanine incorporation. This methodology is now being used in many labs to map cell circuitry and identify orphan ligands and/or receptors.

Protein Therapeutics

UAA mutagenesis is beginning to find many applications in the generation of therapeutic proteins, where the production of large quantities of homogenously modified protein is desired. Three examples illustrate this best. First, is the use of immunogenic amino acids to break immunological tolerance and generate therapeutic vaccines against self-proteins associated with cancer or inflammation. In this approach, incorporation of *p*-nitrophenylalanine (**20**) into a target protein results in immunogenic epitopes that upon immunization induce high-titer, long-lived, and cross-reactive antibodies to native proteins. For example, murine TNF-α mutants containing UAA **20** at position 11 or position 86 induce, in mice, antibodies that are cross-reactive with native murine TNF-α and protect against lipopolysaccharide-induced death (**Figure 5**) (170, 171). This occurs through the stimulation of T cells by epitopes containing UAA **20**, which ultimately leads to activation of autoreactive B cells that produce

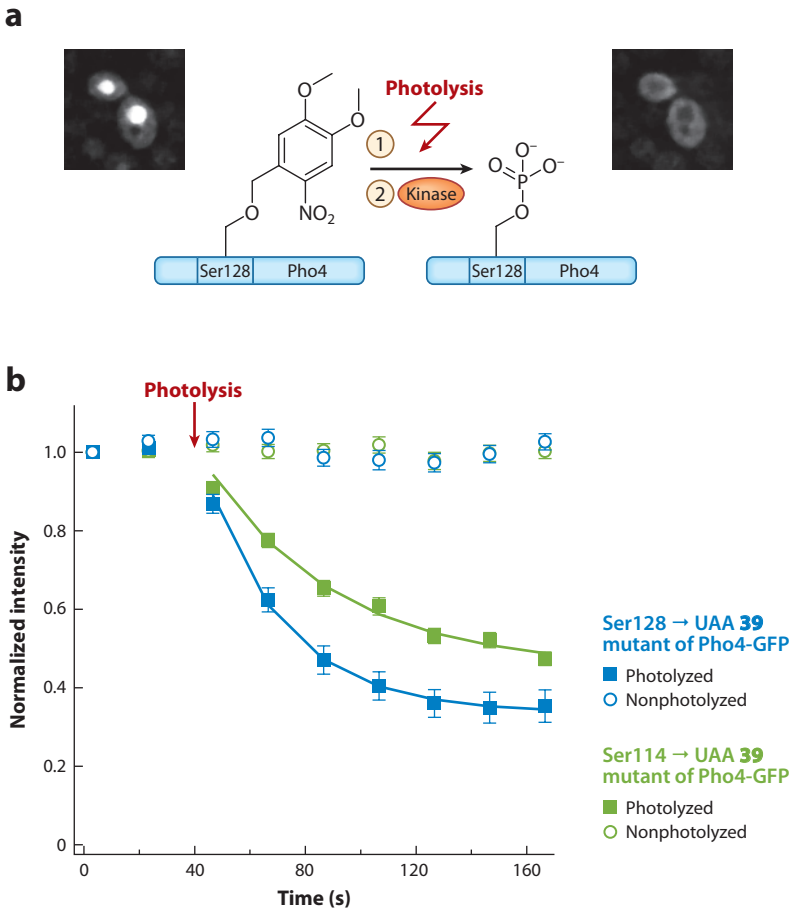

Figure 4

In vivo photolysis and nuclear export of Pho4 in yeast under high-phosphate conditions. (*a*) Magnification of two cells before and after photolysis (Ser128 → UAA **39** mutant of Pho4-GFP), and scheme depicting photolysis followed by serine phosphorylation. (*b*) Real-time analysis of in vivo photolysis of Pho4-GFP mutants. The normalized average fluorescence intensity is plotted as a function of time for the Ser128 → UAA **39** mutant of Pho4-GFP (*blue*) and the Ser114 → UAA **39** mutant of Pho4-GFP (*green*). Panel *b* reprinted with permission from Nature Publishing Group and adapted from Reference 128.

polyclonal antibodies against the native self-protein, a mechanism that should be general to many antigens and epitopes against which a strong immune response is desired. This approach is currently being applied to other self-antigens, such as EGF and HER2 (involved in cancer), PCSK9 (involved in cardiovascular disease), and C5a (involved in inflammation), as well as to conserved epitopes of human immunodeficiency virus (HIV) and malaria that are difficult to target with traditional vaccines. This approach should also prove useful as a tool to knock down levels of secreted proteins through targeted antibody responses. Moreover, the demonstration that a single *p*-nitrophenylalanine residue can break tolerance provides support to the intriguing notion that the enzymatic posttranslational generation of nitrotyrosine in proteins, stimulated by local inflammation and cytokine release, could create immunogenic epitopes that are the underlying initiating event in many autoimmune diseases.

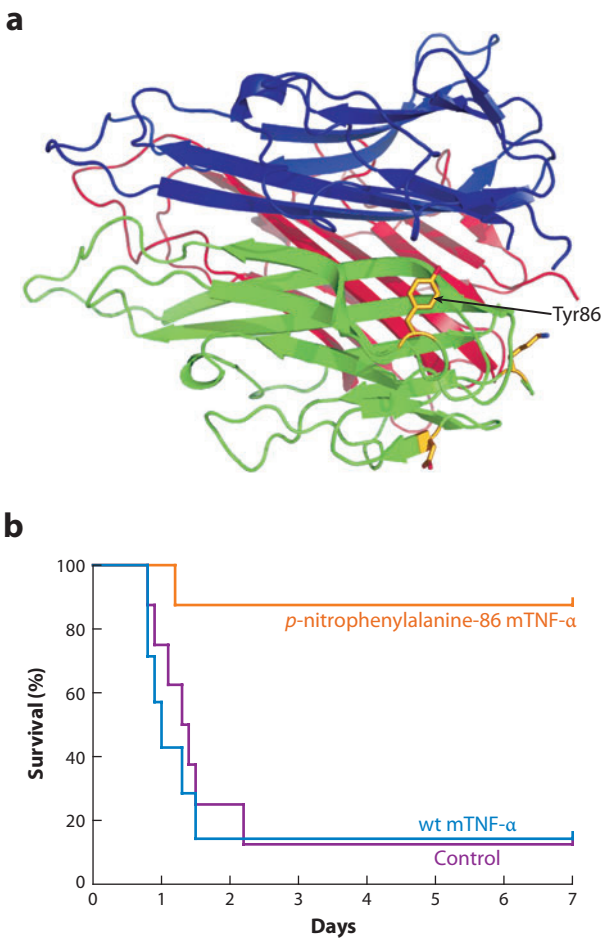

Figure 5

(*a*) X-ray crystal structure of murine tumor necrosis factor alpha (TNF-α) trimer with Tyr86 indicated (Protein Data Bank code 2TNF). (*b*) Substitution of Tyr86 in murine TNF-α with *p*-nitrophenylalanine and subsequent vaccination leads to a robust T cell–driven immune response, which cross-reacts with wild-type (wt) TNF-α and protects mice from lipopolysaccharide (LPS) challenge. Survival of mice vaccinated with the Tyr86 → UAA **20** TNF-α mutant is compared to survival of mice vaccinated with wild-type TNF-α. Survival is plotted as a function of days after LPS challenge. Reprinted with permission from the National Academy of Sciences, U.S.A., copyright © 2008 and adapted from Reference 170.

A second example is the production of the anticoagulant leech protein sulfohirudin. Because natural sulfohirudin is the product of posttranslational tyrosine sulfation, which occurs in higher eukaryotes, it can only be obtained in minute amounts through traditional methods. As such, it is not available in clinically useful quantities; instead, a recombinant hirudin that lacks sulfation is commercially expressed and used as a direct thrombin inhibitor (DTI) with a K_i of ~307 fM. However, by expanding the genetic code of *E. coli* to incorporate sulfotyrosine (**50**), one can access sulfohirudin directly (96). This allows for the expression of the higher-affinity DTI sulfohirudin (K_i = 26 fM) in excellent yields, which also facilitates the crystallization of the sulfohirudin/thrombin complex (**Figure 6**) (172), and should allow clinical application of sulfohirudin.

Third is the production of protein conjugates containing a site specifically attached toxin, radioisotope, polyethylene glycol, or even another protein (creating a bispecific therapeutic). In these cases, an UAA containing a bioorthogonal reactive functional group is incorporated at a specific position, which is then further functionalized in vitro with another

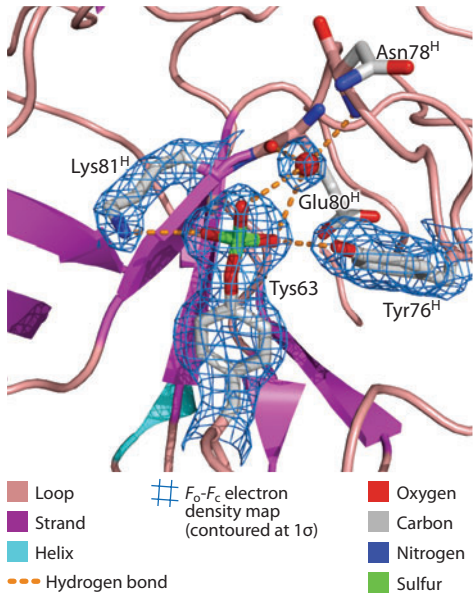

Figure 6

Sulfotyrosine at position 63 (Tys63) in the X-ray crystal structure of the sulfohirudin/thrombin complex. Interactions mediated by the sulfate of UAA **50** are shown. (^{H} refers to the heavy chain of thrombin.) Adapted with permission from Reference 172, copyright © 2007 American Chemical Society.

small molecule or macromolecule of interest. This method overcomes the challenges of nonspecific labeling typically associated with electrophilic reagents or the labeling of cysteine residues that may not be unique or may be involved in folding. In nontherapeutic contexts, this strategy has been used for glycan and lipid conjugation (70, 173), dual-fluorophore labeling for single-molecule FRET studies (174), labeling of G protein–coupled receptors with fluorophores (175), nonnative histone modification (176), and scarless protein purification (73). In therapeutic contexts, this strategy is especially valuable as it allows the creation of homogenous therapeutic agents with defined pharmacological activity and toxicity. Moreover, because the cotranslational incorporation of UAAs is highly efficient, clinically useful quantities of labeled proteins can be produced. Indeed UAA-containing therapeutic candidates can be generated in yields of grams per liter on a >1000 liter fermentation scale in bacteria (high-yielding *Pichia* and mammalian systems are in development). For example, human growth hormone (hGH) containing *p*-acetylphenylalanine (**1**) has been expressed commercially and site specifically PEGylated in kilogram quantities (177). The modified hGH has a significantly improved serum half-life while retaining biological activity and is currently in Phase II clinical trials. Similar approaches are being applied to cytokines, growth factors, and other proteins for clinical use. Various antibodies have also been efficiently expressed with keto amino acid **1** at specific sites for conjugation to therapeutically or diagnostically relevant molecules. For example, in a recent study, *p*-acetylphenylalanine mutants of the α-HER2 antibody trastuzumab were conjugated to the toxins saporin or auristatin, resulting in targeted killing of HER2-positive cells (B. Hutchins, S. Kazane, J. Yin, P.G. Schultz, & V.V. Smider, unpublished results). In addition, the conjugation of antibodies via amino acid **1** to oligonucleotides allows for polymerase chain reaction–based immunodiagnostics as well as the creation of spatially defined arrays of proteins and possibly cells (S. Kazane, B. Hutchins, P.G. Schultz, & V.V. Smider, unpublished results). Efforts are also under way to create structurally defined bispecific antibodies using the additional orthogonal chemistries afforded by UAAs to selectively couple proteins at uniquely defined sites and also to conjugate therapeutic peptides to carrier proteins, such as HSA and antibodies, to improve their half-lives. In all these experiments (and especially in their potential clinical use), large quantities of homogeneous high-molecular-weight UAA-containing proteins are required; thus, the ease and efficiency inherent to translation with expanded genetic codes are highly desirable.

Protein Evolution with an Expanded Genetic Code

It is quite possible that the ability to encode additional amino acids with novel properties would be evolutionarily advantageous, especially since nature's choice of 20 could have been arbitrarily fixed at the point of transition between communal and Darwinian evolution paradigms and subsequently sustained by the code's inertia (162). Furthermore, in the limited scope of laboratory-directed evolution, which concerns only one or few specific functions over a short time rather than general organismal fitness over thousands or millions of years, one can easily envision a selective advantage conferred by additional amino acids. Because the templated assembly of polypeptides from mRNA on the ribosome establishes a direct link between genes (information) and proteins (phenotype), UAA mutagenesis methodology can easily be adapted to the evolution of proteins with novel or enhanced function. As a first step in this direction, robust phage-based systems for use with 21 amino acid organisms were recently developed (178, 179). In one such system, phage-displayed protein libraries are generated in *E. coli* that encode 21 amino acids (X-*E. coli*) such that library members containing a TAG codon in a gene of interest will have a corresponding UAA in the displayed protein. Displayed proteins are then

subjected to selection, and the surviving phage clones are used to reinfect X-*E. coli*. Like traditional phage-display with the canonical 20 amino acids, rounds of mutation, amplification, and selection result in the discovery of new proteins with desired properties that, in this case, may be achieved by UAAs.

After optimizing this system to minimize the expression bias against proteins containing UAAs, several in vitro molecular evolution experiments were conducted. For example, a phage-displayed library of naive antibodies containing six NNK-randomized residues (N = A, C, G, or T; K = G or T) in the V_H CDR3 was produced in X-*E. coli* that encodes the UAA sulfotyrosine (**50**). Selection experiments were then performed on the basis of affinity for the HIV coat protein gp120 (179). After several rounds of amplification and selection, the phage population converged on one antibody clone that contained a sulfotyrosine. The selection of this phage-displayed antibody was based on gp120-binding function, as determined by expression studies, and this function was dependent on UAA **50**, as confirmed by enzyme-linked immunosorbent assays. Another in vitro evolution experiment used X-*E. coli* that encoded the reactive UAA *p*-boronophenylalanine (**8**) to generate an unbiased phage library of naive antibodies containing NNK-randomized residues. The resulting phage particles were exposed to selection for binding to the acyclic glycan glucamine (180). After one round, over 50% of the phage population contained UAA **8**, and after three rounds, over 80% contained UAA **8**. This convergence on UAA-containing sequences was concurrent with functional enrichment for antibodies that bind glucamine, presumably by using the boronate functionality as a "chemical warhead" to react with diols. Furthermore, the function of the selected *p*-boronophenylalanine-containing clones cannot be attributed to UAA **8** alone but requires the contribution of the whole selected sequence, as demonstrated in a series of studies on both phage-displayed and soluble Fab format clones. Several other studies using phage-based evolution with 21 amino acids are currently being conducted, including the use of UAAs to overcome PTM sequence constraints (181), the use of reactive UAAs to evolve inhibitors of glycoproteins and proteases, and the incorporation of metal ion–binding UAAs to evolve metalloproteins with novel activities.

In addition to these phage-based evolution experiments, other efforts to include UAAs in directed evolution experiments are under way. For example, we are currently synthesizing and selecting bioactive cyclic peptides that contain UAAs, generated through the use of split inteins or a recently characterized ribosomal synthesis pathway (182). Furthermore, although there have been several recent successes of protein engineering with UAAs—for example, the generation of a redox-active, bipyridylalanine-containing catabolite activator protein mutant that site specifically cleaves DNA (183)—we expect that protein design alone will not be enough as we move toward the application of UAAs to more complex functions. Therefore, a combination of computational design and molecular evolution techniques are being employed in several new efforts (for example, to engineer metalloproteins that use a metal ion–binding UAA as a structural element or for catalysis). Future studies are also likely to include organismal evolution experiments, including those with an autonomous 21 amino acid *E. coli* that biosynthesizes its own UAA (75).

Taken together, these experiments demonstrate that the "chemical potential" contained in expanded amino acid sets can, through directed evolution, be realized as enhanced protein function and show that an expanded genetic code can confer a selective advantage in protein evolution. In all these cases, the integration of orthogonal UAARS/$tRNA_{SB}$ pairs into the translation paradigm provides the direct link between genotype and phenotype required for protein evolution with UAAs.

CONCLUSION

Genetically encoded UAAs seamlessly integrate new chemistry into biology and offer a number

of advantages over other approaches for chemically manipulating protein structure and function. First, the generality of methods for developing orthogonal UAARS/$tRNA_{SB}$ pairs allows the systematic genetic encoding of new UAAs in vivo to meet new needs in chemistry and biology. Second, the modularity of components involved in this approach allows the combination of multiple blank codons with multiple UAARS/$tRNA_{SB}$ pairs to achieve the biosynthesis of proteins containing many or perhaps exclusively UAAs. Finally, the incorporation of UAAs directly through the translation machinery greatly simplifies the efficient production of homogenous mutant proteins, the development of cell biological probes, and the evolution of new protein function. Thus by removing the constraints imposed by nature on the number and properties of genetically encoded amino acids, exciting new opportunities in basic and applied science research have been (and will continue to be) realized.

SUMMARY POINTS

1. A general method based on the engineering of modular orthogonal unnatural amino acid–specific mutant aminoacyl-tRNA synthetase (UAARS)/tRNA pairs makes possible the genetic encoding of unnatural amino acids (UAAs) in *E. coli*, yeast, and mammalian cells.
2. Approximately 70 UAAs, representing various structures and chemistries distinct from those contained in the 20 natural amino acids, have been genetically encoded in response to blank codons, such as nonsense triplet codons and frameshift quadruplet codons.
3. Structural studies of engineered UAARSs suggest that their actives sites have considerable structural plasticity but that mutations in the amino acid–binding sites of synthetases result in small, local perturbations rather than unpredictable large-scale changes, underscoring the utility of structure-based synthetase library design and selection to manipulate aminoacyl-tRNA synthetase (aaRS) substrate specificity.
4. A two-step selection system, which uses a positive step to select for aaRSs that aminoacylate the cognate tRNA with the desired UAA and a negative step to remove aaRSs that recognize endogenous host amino acids, is highly effective in the evolution of aaRS specificity and orthogonality.
5. The integration of UAARS/tRNA pairs directly into the translation machinery allows for the efficient and selective introduction of UAAs into proteins in living cells.
6. Applications of genetically encoded UAAs are widespread and are already leading to novel protein therapeutics, new tools to study protein structure and function, and the directed evolution of proteins that use UAAs to achieve new function.

FUTURE ISSUES

1. The further development of novel UAARS/tRNA pairs for the incorporation of new, chemically distinct UAAs.
2. The further optimization of the translational machinery for more efficient expression of proteins with UAAs at any defined site.

3. The creation of new codon$_{BL}$s and mutually orthogonal UAARS/tRNA pairs to allow the in vivo expression of proteins containing multiple UAAs or even fully unnatural polypeptide polymers.
4. The further application of UAA mutagenesis methods to the generation of novel protein therapeutics (e.g., vaccines, bispecific proteins, immunotoxins) and to the development of tools (e.g., fluorophores, photoreactive amino acids, amino acids containing modifications normally added posttranslationally) to study protein structure and function in vitro and in living cells.
5. The further use of expanded genetic codes to evolve new proteins with novel physical, biochemical, and catalytic properties that are difficult to produce using only the 20 canonical amino acids.
6. The creation of multicellular organisms that have expanded genetic codes for the in vivo study of protein structure and function.

DISCLOSURE STATEMENT

Some of the work described has been licensed through the Scripps Research Institute. The institution and inventors, including C.C.L. and P.G.S., receive license fees.

ACKNOWLEDGMENTS

C.C.L. is supported by predoctoral fellowships from the Fannie and John Hertz Foundation and the National Science Foundation. P.G.S. is supported by the National Institutes of Health, the U.S. Department of Energy, and the Skaggs Institute of Chemical Biology.

LITERATURE CITED

1. Silverman RB. 2002. *Organic Chemistry of Enzyme-Catalyzed Reactions*. London/San Diego: Academic. 800 pp.
2. Walsh CT. 2005. *Posttranslational Modification of Proteins: Expanding Nature's Inventory*. Englewood, CO: Roberts. 490 pp.
3. Chalker JM, Bernardes GJ, Lin YA, Davis BG. 2009. Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chem. Asian J.* 4:630–40
4. Bernardes GJ, Chalker JM, Errey JC, Davis BG. 2008. Facile conversion of cysteine and alkyl cysteines to dehydroalanine on protein surfaces: versatile and switchable access to functionalized proteins. *J. Am. Chem. Soc.* 130:5052–53
5. Merrifield RB. 1969. Solid-phase peptide synthesis. *Adv. Enzymol. Relat. Areas Mol. Biol.* 32:221–96
6. Marshall GR. 2003. Solid-phase synthesis: a paradigm shift. *J. Pept. Sci.* 9:534–44
7. Dawson PE, Kent SB. 2000. Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* 69:923–60
8. Muir TW. 2003. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* 72:249–89
9. Ambrogelly A, Palioura S, Söll D. 2007. Natural expansion of the genetic code. *Nat. Chem. Biol.* 3:29–35
10. Kimmerlin T, Seebach D. 2005. '100 years of peptide synthesis': ligation methods for peptide and protein synthesis with applications to beta-peptide assemblies. *J. Pept. Res.* 65:229–60
11. Baca M, Kent SB. 1993. Catalytic contribution of flap-substrate hydrogen bonds in "HIV-1 protease" explored by chemical synthesis. *Proc. Natl. Acad. Sci. USA* 90:11638–42

12. Kochendoerfer GG, Chen SY, Mao F, Cressman S, Traviglia S, et al. 2003. Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* 299:884–87
13. Chen SY, Cressman S, Mao F, Shao H, Low DW, et al. 2005. Synthetic erythropoietic proteins: tuning biological performance by site-specific polymer attachment. *Chem. Biol.* 12:371–83
14. Vazquez ME, Nitz M, Stehn J, Yaffe MB, Imperiali B. 2003. Fluorescent caged phosphoserine peptides as probes to investigate phosphorylation-dependent protein associations. *J. Am. Chem. Soc.* 125:10150–51
15. Flavell RR, Muir TW. 2009. Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. *Acc. Chem. Res.* 42:107–16
16. Valiyaveetil FI, Leonetti M, Muir TW, MacKinnon R. 2006. Ion selectivity in a semisynthetic K^+ channel locked in the conductive conformation. *Science* 314:1004–7
17. Hecht SM, Alford BL, Kuroda Y, Kitano S. 1978. "Chemical aminoacylation" of tRNA's. *J. Biol. Chem.* 253:4517–20
18. Heckler TG, Chang LH, Zama Y, Naka T, Chorghade MS, Hecht SM. 1984. T4 RNA ligase mediated preparation of novel "chemically misacylated" $tRNA^{Phe}$ S. *Biochemistry* 23:1468–73
19. Roesser JR, Xu C, Payne RC, Surratt CK, Hecht SM. 1989. Preparation of misacylated aminoacyl-$tRNA^{Phe}$'s useful as probes of the ribosomal acceptor site. *Biochemistry* 28:5185–95
20. Robertson SA, Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG. 1989. The use of 5′-phospho-2 deoxyribocytidylylriboadenosine as a facile route to chemical aminoacylation of tRNA. *Nucleic Acids Res.* 17:9649–60
21. 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
22. Koh JT, Cornish VW, Schultz PG. 1997. An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. *Biochemistry* 36:11314–22
23. Eisenhauer BM, Hecht SM. 2002. Site-specific incorporation of (aminooxy)acetic acid into proteins. *Biochemistry* 41:11472–78
24. Murakami H, Hohsaka T, Ashizuka Y, Hashimoto K, Sisido M. 2000. Site-directed incorporation of fluorescent nonnatural amino acids into streptavidin for highly sensitive detection of biotin. *Biomacromolecules* 1:118–25
25. Taki M, Hohsaka T, Murakami H, Taira K, Sisido M. 2001. A non-natural amino acid for efficient incorporation into proteins as a sensitive fluorescent probe. *FEBS Lett.* 507:35–38
26. Taki M, Hohsaka T, Murakami H, Taira K, Sisido M. 2002. Position-specific incorporation of a fluorophore-quencher pair into a single streptavidin through orthogonal four-base codon/anticodon pairs. *J. Am. Chem. Soc.* 124:14586–90
27. Fahmi NE, Dedkova L, Wang B, Golovine S, Hecht SM. 2007. Site-specific incorporation of glycosylated serine and tyrosine derivatives into proteins. *J. Am. Chem. Soc.* 129:3586–97
28. Kanamori T, Nishikawa S, Shin I, Schultz PG, Endo T. 1997. Probing the environment along the protein import pathways in yeast mitochondria by site-specific photocrosslinking. *Proc. Natl. Acad. Sci. USA* 94:485–90
29. Thorson J, Chapman E, Murphy E, Judice J, Schultz P. 1995. Linear free energy analysis of hydrogen bonding in proteins. *J. Am. Chem. Soc.* 117:1157–58
30. Li S, Millward S, Roberts R. 2002. In vitro selection of mRNA display libraries containing an unnatural amino acid. *J. Am. Chem. Soc.* 124:9972–73
31. Frankel A, Li S, Starck SR, Roberts RW. 2003. Unnatural RNA display libraries. *Curr. Opin. Struct. Biol.* 13:506–12
32. Kohrer C, Yoo JH, Bennett M, Schaack J, RajBhandary UL. 2003. A possible approach to site-specific insertion of two different unnatural amino acids into proteins in mammalian cells via nonsense suppression. *Chem. Biol.* 10:1095–102
33. Rodriguez EA, Lester HA, Dougherty DA. 2006. In vivo incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proc. Natl. Acad. Sci. USA* 103:8650–55
34. Saks ME, Sampson JR, Nowak MW, Kearney PC, Du F, et al. 1996. An engineered *Tetrahymena* $tRNA^{Gln}$ for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression. *J. Biol. Chem.* 271:23169–75

35. Dougherty DA. 2000. Unnatural amino acids as probes of protein structure and function. *Curr. Opin. Chem. Biol.* 4:645–52
36. Beene DL, Dougherty DA, Lester HA. 2003. Unnatural amino acid mutagenesis in mapping ion channel function. *Curr. Opin. Neurobiol.* 13:264–70
37. England PM, Zhang Y, Dougherty DA, Lester HA. 1999. Backbone mutations in transmembrane domains of a ligand-gated ion channel: implications for the mechanism of gating. *Cell* 96:89–98
38. Dougherty DA. 2008. Physical organic chemistry on the brain. *J. Org. Chem.* 73:3667–73
39. Lummis SCR, Beene DL, Lee LW, Lester HA, Broadhurst RW, Dougherty DA. 2005. *Cis-trans* isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* 438:248–52
40. Hendrickson TL, de Crécy-Lagard V, Schimmel P. 2004. Incorporation of nonnatural amino acids into proteins. *Annu. Rev. Biochem.* 73:147–76
41. Link AJ, Mock ML, Tirrell DA. 2003. Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* 14:603–9
42. van Hest JC, Tirrell DA. 1998. Efficient introduction of alkene functionality into proteins in vivo. *FEBS Lett.* 428:68–70
43. Link AJ, Tirrell DA. 2003. Cell surface labeling of *Escherichia coli* via copper(I)-catalyzed [3+2] cycloaddition. *J. Am. Chem. Soc.* 125:11164–65
44. Kiick KL, van Hest JC, Tirrell DA. 2000. Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host. *Angew. Chem. Int. Ed. Engl.* 39:2148–52
45. Datta D, Wang P, Carrico IS, Mayo SL, Tirrell DA. 2002. A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins. *J. Am. Chem. Soc.* 124:5652–53
46. Sharma N, Furter R, Kast P, Tirrell DA. 2000. Efficient introduction of aryl bromide functionality into proteins in vivo. *FEBS Lett.* 467:37–40
47. Döring V, Mootz HD, Nangle LA, Hendrickson TL, de Crécy-Lagard V, et al. 2001. Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway. *Science* 292:501–4
48. Link AJ, Vink MK, Agard NJ, Prescher JA, Bertozzi CR, Tirrell DA. 2006. Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proc. Natl. Acad. Sci. USA* 103:10180–85
49. Yang W, Hendrickson WA, Crouch RJ, Satow Y. 1990. Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. *Science* 249:1398–405
50. Cirino PC, Tang Y, Takahashi K, Tirrell DA, Arnold FH. 2003. Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. *Biotechnol. Bioeng.* 83:729–34
51. Montclare JK, Tirrell DA. 2006. Evolving proteins of novel composition. *Angew. Chem. Int. Ed. Engl.* 45:4518–21
52. Bae JH, Rubini M, Jung G, Wiegand G, Seifert MH, et al. 2003. Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. *J. Mol. Biol.* 328:1071–81
53. Budisa N, Rubini M, Bae JH, Weyher E, Wenger W, et al. 2002. Global replacement of tryptophan with aminotryptophans generates non-invasive protein-based optical pH sensors. *Angew. Chem. Int. Ed. Engl.* 41:4066–69
54. Beatty KE, Xie F, Wang Q, Tirrell DA. 2005. Selective dye-labeling of newly synthesized proteins in bacterial cells. *J. Am. Chem. Soc.* 127:14150–51
55. Beatty KE, Liu JC, Xie F, Dieterich DC, Schuman EM, et al. 2006. Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew. Chem. Int. Ed. Engl.* 45:7364–67
56. Ibba M, Söll D. 2000. Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* 69:617–50
57. Ibba M, Söll D. 2004. Aminoacyl-tRNAs: setting the limits of the genetic code. *Genes Dev.* 18:731–38
58. Giege R, Sissler M, Florentz C. 1998. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* 26:5017–35
59. Chin JW. 2006. Modular approaches to expanding the functions of living matter. *Nat. Chem. Biol.* 2:304–11

60. 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
61. Wang L, Schultz PG. 2004. Expanding the genetic code. *Angew. Chem. Int. Ed. Engl.* 44:34–66
62. 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
63. Wang L, Schultz PG. 2001. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* 8:883–90
64. Santoro SW, Wang L, Herberich B, King DS, Schultz PG. 2002. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* 20:1044–48
65. Melancon CE 3rd, Schultz PG. 2009. One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases. *Bioorg. Med. Chem. Lett.* 19:3845–47
66. 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
67. 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
68. Zeng H, Xie J, Schultz PG. 2006. Genetic introduction of a diketone-containing amino acid into proteins. *Bioorg. Med. Chem. Lett.* 16:5356–59
69. Zhang Z, Wang L, Brock A, Schultz PG. 2002. The selective incorporation of alkenes into proteins in *Escherichia coli*. *Angew. Chem. Int. Ed. Engl.* 41:2840–42
70. Wang J, Schiller SM, Schultz PG. 2007. A biosynthetic route to dehydroalanine-containing proteins. *Angew. Chem. Int. Ed. Engl.* 46:6849–51
71. Deiters A, Schultz PG. 2005. In vivo incorporation of an alkyne into proteins in *Escherichia coli*. *Bioorg. Med. Chem. Lett.* 15:1521–24
72. 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
73. Brustad E, Bushey ML, Lee JW, Groff D, Liu W, Schultz PG. 2008. A genetically encoded boronate-containing amino acid. *Angew. Chem. Int. Ed. Engl.* 47:8220–23
74. Wang L, Brock A, Herberich B, Schultz PG. 2001. Expanding the genetic code of *Escherichia coli*. *Science* 292:498–500
75. 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
76. Schultz KC, Supekova L, Ryu Y, Xie J, Perera R, Schultz PG. 2006. A genetically encoded infrared probe. *J. Am. Chem. Soc.* 128:13984–85
77. Wang L, Xie J, Deniz AA, Schultz PG. 2003. Unnatural amino acid mutagenesis of green fluorescent protein. *J. Org. Chem.* 68:174–76
78. 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
79. Cellitti SE, Jones DH, Lagpacan L, Hao X, Zhang Q, et al. 2008. In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* 130:9268–81
80. Tsao ML, Summerer D, Ryu Y, Schultz PG. 2006. The genetic incorporation of a distance probe into proteins in *Escherichia coli*. *J. Am. Chem. Soc.* 128:4572–73
81. 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
82. Seyedsayamdost MR, Xie J, Chan CT, Schultz PG, Stubbe J. 2007. Site-specific insertion of 3-aminotyrosine into subunit $\alpha 2$ of *E. coli* ribonucleotide reductase: direct evidence for involvement of Y_{730} and Y_{731} in radical propagation. *J. Am. Chem. Soc.* 129:15060–71
83. Sakamoto K, Murayama K, Oki K, Iraha F, Kato-Murayama M, et al. 2009. Genetic encoding of 3-iodo-L-tyrosine in *Escherichia coli* for single-wavelength anomalous dispersion phasing in protein crystallography. *Structure* 17:335–44
84. 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

85. Xie J, Liu W, Schultz PG. 2007. A genetically encoded bidentate, metal-binding amino acid. *Angew. Chem. Int. Ed. Engl.* 46:9239–42
86. Guo J, Wang J, Anderson JC, Schultz PG. 2008. Addition of an alpha-hydroxy acid to the genetic code of bacteria. *Angew. Chem. Int. Ed. Engl.* 47:722–25
87. Lee HS, Spraggon G, Schultz PG, Wang F. 2009. Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe. *J. Am. Chem. Soc.* 131:2481–83
88. 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
89. Deiters A, Groff D, Ryu Y, Xie J, Schultz PG. 2006. A genetically encoded photocaged tyrosine. *Angew. Chem. Int. Ed. Engl.* 45:2728–31
90. Peters FB, Brock A, Wang J, Schultz PG. 2009. Photocleavage of the polypeptide backbone by 2-nitrophenylalanine. *Chem. Biol.* 16:148–52
91. 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
92. Tippmann EM, Liu W, Summerer D, Mack AV, Schultz PG. 2007. A genetically encoded diazirine photocrosslinker in *Escherichia coli*. *ChemBioChem* 8:2210–14
93. Wang J, Xie J, Schultz PG. 2006. A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc.* 128:8738–39
94. Xie J, Supekova L, Schultz PG. 2007. A genetically encoded metabolically stable analogue of phosphotyrosine in *Escherichia coli*. *ACS Chem. Biol.* 2:474–78
95. Neumann H, Hazen JL, Weinstein J, Mehl RA, Chin JW. 2008. Genetically encoding protein oxidative damage. *J. Am. Chem. Soc.* 130:4028–33
96. Liu CC, Schultz PG. 2006. Recombinant expression of selectively sulfated proteins in *Escherichia coli*. *Nat. Biotechnol.* 24:1436–40
97. Pastrnak M, Magliery T, 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
98. 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
99. Ohno S, Yokogawa T, Fujii I, Asahara H, Inokuchi H, Nishikawa K. 1998. Co-expression of yeast amber suppressor tRNATyr and tyrosyl-tRNA synthetase in *Escherichia coli*: possibility to expand the genetic code. *J. Biochem.* 124:1065–68
100. Furter R. 1998. Expansion of the genetic code: site-directed *p*-fluoro-phenylalanine incorporation in *Escherichia coli*. *Protein Sci.* 7:419–26
101. 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
102. 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
103. 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
104. 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
105. 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
106. Wang F, Robbins S, Shen W, Schultz PG. 2010. Genetic incorporation of unnatural amino acids into proteins in *Mycobacterium tuberculosis*. *PLoS ONE* 5(2):e9354
107. Krzycki JA. 2005. The direct genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* 8:706–12
108. Srinivasan G, James CM, Krzycki JA. 2002. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* 296:1459–62

109. Kavran JM, Gundllapalli S, O'Donoghue P, Englert M, Söll D, Steitz TA. 2007. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. USA* 104:11268–73
110. Woese CR, Olsen GJ, Ibba M, Söll D. 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* 64:202–36
111. Blight SK, Larue RC, Mahapatra A, Longstaff DG, Chang E, et al. 2004. Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature* 431:333–35
112. Yanagisawa T, Ishii R, Fukunaga R, Kobayashi T, Sakamoto K, Yokoyama S. 2008. Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* 15:1187–97
113. Nozawa K, O'Donoghue P, Gundllapalli S, Araiso Y, Ishitani R, et al. 2009. Pyrrolysyl-tRNA synthetase-$tRNA^{Pyl}$ structure reveals the molecular basis of orthogonality. *Nature* 457:1163–67
114. Mukai T, Kobayashi T, Hino N, Yanagisawa T, Sakamoto K, Yokoyama S. 2008. Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* 371:818–22
115. Li WT, Mahapatra A, Longstaff DG, Bechtel J, Zhao G, et al. 2009. Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. *J. Mol. Biol.* 385:1156–64
116. Polycarpo CR, Herring S, Berube A, Wood JL, Söll D, Ambrogelly A. 2006. Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* 580:6695–700
117. Fekner T, Li X, Lee MM, Chan MK. 2009. A pyrrolysine analogue for protein click chemistry. *Angew. Chem. Int. Ed. Engl.* 48:1633–35
118. Kobayashi T, Yanagisawa T, Sakamoto K, Yokoyama S. 2009. Recognition of non-alpha-amino substrates by pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* 385:1352–60
119. Nguyen DP, Lusic H, Neumann H, Kapadnis PB, Deiters A, Chin JW. 2009. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNA(CUA) pair and click chemistry. *J. Am. Chem. Soc.* 131:8720–21
120. Chen PR, Groff D, Guo J, Ou W, Cellitti S, et al. 2009. A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem. Int. Ed. Engl.* 48:4052–55
121. Neumann H, Peak-Chew SY, Chin JW. 2008. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* 4:232–34
122. 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
123. Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, Schultz PG. 2003. An expanded eukaryotic genetic code. *Science* 301:964–67
124. Wu N, Deiters A, Cropp TA, King D, Schultz PG. 2004. A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* 126:14306–7
125. 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
126. Lee HS, Guo J, Lemke EA, Dimla RD, Schultz PG. 2009. Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in *Saccharomyces cerevisiae*. *J. Am. Chem. Soc.* 131:12921–23
127. Brustad E, Bushey ML, Brock A, Chittuluru J, Schultz PG. 2008. A promiscuous aminoacyl-tRNA synthetase that incorporates cysteine, methionine, and alanine homologs into proteins. *Bioorg. Med. Chem. Lett.* 18:6004–6
128. Lemke EA, Summerer D, Geierstanger BH, Brittain SM, Schultz PG. 2007. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* 3:769–72
129. Summerer D, Chen S, Wu N, Deiters A, Chin JW, Schultz PG. 2006. A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci. USA* 103:9785–89
130. Tippmann EM, Schultz PG. 2007. A genetically encoded metallocene containing amino acid. *Tetrahedron* 63:6182–84
131. Young TS, Ahmad I, Brock A, Schultz PG. 2009. Expanding the genetic repertoire of the methylotrophic yeast *Pichia pastoris*. *Biochemistry* 48:2643–53

132. 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
133. Liu W, Brock A, Chen S, Chen S, Schultz PG. 2007. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* 4:239–44
134. Wang W, Takimoto JK, Louie GV, Baiga TJ, Noel JP, et al. 2007. Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat. Neurosci.* 10:1063–72
135. Kwon I, Tirrell DA. 2007. Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells. *J. Am. Chem. Soc.* 129:10431–37
136. Kwon I, Wang P, Tirrell DA. 2006. Design of a bacterial host for site-specific incorporation of *p*-bromophenylalanine into recombinant proteins. *J. Am. Chem. Soc.* 128:11778–83
137. Oki K, Sakamoto K, Kobayashi T, Sasaki HM, Yokoyama S. 2008. Transplantation of a tyrosine editing domain into a tyrosyl-tRNA synthetase variant enhances its specificity for a tyrosine analog. *Proc. Natl. Acad. Sci. USA* 105:13298–303
138. Liu W, Alfonta L, Mack AV, Schultz PG. 2007. Structural basis for the recognition of para-benzoyl-L-phenylalanine by evolved aminoacyl-tRNA synthetases. *Angew. Chem. Int. Ed. Engl.* 46:6073–75
139. Turner JM, Graziano J, Spraggon G, Schultz PG. 2006. Structural plasticity of an aminoacyl-tRNA synthetase active site. *Proc. Natl. Acad. Sci. USA* 103:6483–88
140. 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
141. Cusack S, Yaremchuk A, Tukalo M. 2000. The 2 A crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *EMBO J.* 19:2351–61
142. Ryu Y, Schultz PG. 2006. Efficient incorporation of unnatural amino acids into proteins in *Escherichia coli*. *Nat. Methods* 3:263–65
143. Bossi L. 1983. Context effects: translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. *J. Mol. Biol.* 164:73–87
144. Taira H, Hohsaka T, Sisido M. 2006. In vitro selection of tRNAs for efficient four-base decoding to incorporate non-natural amino acids into proteins in an *Escherichia coli* cell-free translation system. *Nucleic Acids Res.* 34:1653–62
145. 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
146. Kwon I, Kirshenbaum K, Tirrell DA. 2003. Breaking the degeneracy of the genetic code. *J. Am. Chem. Soc.* 125:7512–13
147. Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324:255–58
148. Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S. 2008. Virus attenuation by genome-scale changes in codon pair bias. *Science* 320:1784–87
149. Bain JD, Switzer C, Chamberlin AR, Benner SA. 1992. Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature* 356:537–39
150. Hirao I, Ohtsuki T, Fujiwara T, Mitsui T, Yokogawa T, et al. 2002. An unnatural base pair for incorporating amino acid analogs into proteins. *Nat. Biotechnol.* 20:177–82
151. Hirao I, Harada Y, Kimoto M, Mitsui T, Fujiwara T, Yokoyama S. 2004. A two-unnatural-base-pair system toward the expansion of the genetic code. *J. Am. Chem. Soc.* 126:13298–305
152. Liu CC, Cellitti S, Geierstanger BH, Schultz PG. 2009. Efficient expression of tyrosine-sulfated proteins in *E. coli* using an expanded genetic code. *Nat. Protoc.* 4:1784–89
153. Chen S, Schultz PG, Brock A. 2007. An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in *Saccharomyces cerevisiae*. *J. Mol. Biol.* 371:112–22
154. Wang Q, Wang L. 2008. New methods enabling efficient incorporation of unnatural amino acids in yeast. *J. Am. Chem. Soc.* 130:6066–67
155. 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

156. Guo J, Melancon CE, Lee HS, Groff D, Schultz PG. 2009. Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew. Chem. Int. Ed. Engl.* 48:9148–51
157. Young TS, Ahmad I, Yin JA, Schultz PG. 2009. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.* 395:361–74
158. Rackham O, Chin JW. 2005. A network of orthogonal ribosome x mRNA pairs. *Nat. Chem. Biol.* 1:159–66
159. Wang K, Neumann H, Peak-Chew SY, Chin JW. 2007. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* 25:770–77
160. Doi Y, Ohtsuki T, Shimizu Y, Ueda T, Sisido M. 2007. Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system. *J. Am. Chem. Soc.* 129:14458–62
161. Sella G, Ardell DH. 2006. The coevolution of genes and genetic codes: Crick's frozen accident revisited. *J. Mol. Evol.* 63:297–313
162. Vetsigian K, Woese C, Goldenfeld N. 2006. Collective evolution and the genetic code. *Proc. Natl. Acad. Sci. USA* 103:10696–701
163. Lu Y, Freeland S. 2006. On the evolution of the standard amino-acid alphabet. *Genome Biol.* 7:102.1–.6
164. Weber AL, Miller SL. 1981. Reasons for the occurrence of the twenty coded protein amino acids. *J. Mol. Evol.* 17:273–84
165. Ye S, Huber T, Vogel R, Sakmar TP. 2009. FTIR analysis of GPCR activation using azido probes. *Nat. Chem. Biol.* 5:397–99
166. Groff D, Thielges MC, Cellitti S, Schultz PG, Romesberg FE. 2009. Efforts toward the direct experimental characterization of enzyme microenvironments: tyrosine100 in dihydrofolate reductase. *Angew. Chem. Int. Ed. Engl.* 48:3478–81
167. Mills JH, Lee HS, Liu CC, Wang J, Schultz PG. 2009. A genetically encoded direct sensor of antibody-antigen interactions. *ChemBioChem* 10:2162–64
168. Lee HS, Dimla RD, Schultz PG. 2009. Protein-DNA photo-crosslinking with a genetically encoded benzophenone-containing amino acid. *Bioorg. Med. Chem. Lett.* 19:5222–24
169. 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
170. Grunewald J, Tsao ML, Perera R, Dong L, Niessen F, et al. 2008. Immunochemical termination of self-tolerance. *Proc. Natl. Acad. Sci. USA* 105:11276–80
171. Grunewald J, Hunt GS, Dong L, Niessen F, Wen BG, et al. 2009. Mechanistic studies of the immunochemical termination of self-tolerance with unnatural amino acids. *Proc. Natl. Acad. Sci. USA* 106:4337–42
172. Liu CC, Brustad E, Liu W, Schultz PG. 2007. Crystal structure of a biosynthetic sulfo-hirudin complexed to thrombin. *J. Am. Chem. Soc.* 129:10648–49
173. 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
174. Brustad EM, Lemke EA, Schultz PG, Deniz AA. 2008. A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J. Am. Chem. Soc.* 130:17664–65
175. Ye S, Kohrer C, Huber T, Kazmi M, Sachdev P, et al. 2008. Site-specific incorporation of keto amino acids into functional G protein–coupled receptors using unnatural amino acid mutagenesis. *J. Biol. Chem.* 283:1525–33
176. Guo J, Wang J, Lee JS, Schultz PG. 2008. Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. *Angew. Chem. Int. Ed. Engl.* 47:6399–401
177. 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
178. Love KR, Swoboda JG, Noren CJ, Walker S. 2006. Enabling glycosyltransferase evolution: a facile substrate-attachment strategy for phage-display enzyme evolution. *ChemBioChem* 7:753–56
179. Liu CC, Mack AV, Tsao ML, Mills JH, Lee HS, et al. 2008. Protein evolution with an expanded genetic code. *Proc. Natl. Acad. Sci. USA* 105:17688–93
180. Liu CC, Mack AV, Brustad E, Mills JH, Groff D, et al. 2009. Evolution of proteins with genetically encoded "chemical warheads." *J. Am. Chem. Soc.* 131:9616–17

181. Liu CC, Choe H, Farzan M, Schultz PG. 2009. Mutagenesis and evolution of sulfated antibodies using an expanded genetic code. *Biochemistry* 48:8891–98
182. Donia MS, Hathaway BJ, Sudek S, Haygood MG, Rosovitz MJ, et al. 2006. Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat. Chem. Biol.* 2:729–35
183. Lee HS, Schultz PG. 2008. Biosynthesis of a site-specific DNA cleaving protein. *J. Am. Chem. Soc.* 130:13194–95

Annual Review of Biochemistry

Volume 79, 2010

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Biochemistry* articles may be found at http://biochem.annualreviews.org