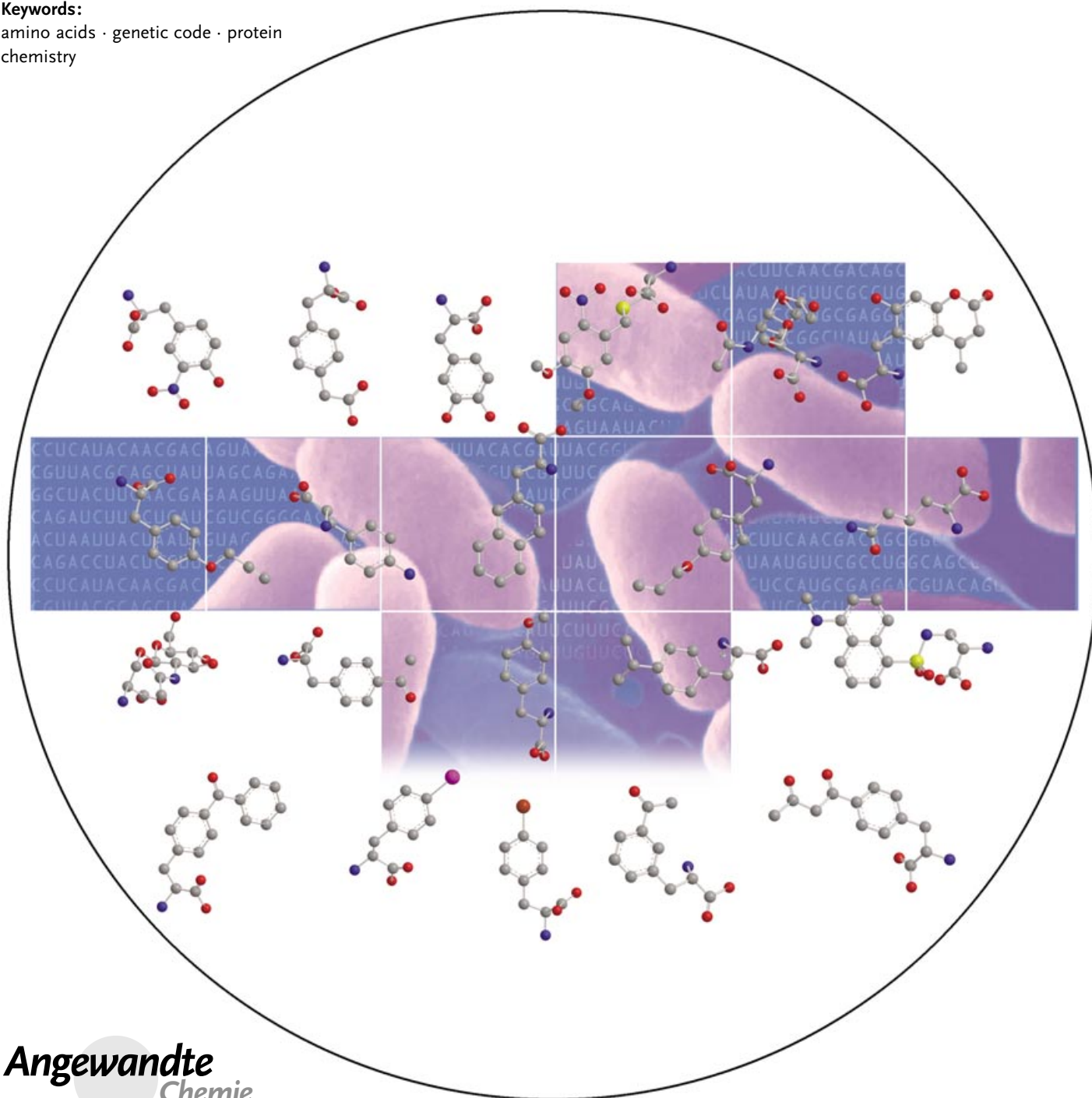


Protein Science

Expanding the Genetic Code

Lei Wang and Peter G. Schultz*

Keywords:
amino acids · genetic code · protein
chemistry



Although chemists can synthesize virtually any small organic molecule, our ability to rationally manipulate the structures of proteins is quite limited, despite their involvement in virtually every life process. For most proteins, modifications are largely restricted to substitutions among the common 20 amino acids. Herein we describe recent advances that make it possible to add new building blocks to the genetic codes of both prokaryotic and eukaryotic organisms. Over 30 novel amino acids have been genetically encoded in response to unique triplet and quadruplet codons including fluorescent, photoreactive, and redox-active amino acids, glycosylated amino acids, and amino acids with keto, azido, acetylenic, and heavy-atom-containing side chains. By removing the limitations imposed by the existing 20 amino acid code, it should be possible to generate proteins and perhaps entire organisms with new or enhanced properties.

1. Introduction

The genetic codes of all known organisms specify the same 20 amino acid building blocks. These building blocks contain a limited number of functional groups including carboxylic acids and amides, a thiol and thiol ether, alcohols, basic amines, and alkyl and aryl groups. Although various arguments have been put forth to explain the nature and number of amino acids in the code,^[1–3] it is clear that proteins require additional chemical groups to carry out their natural functions. These groups are provided through posttranslational modifications including phosphorylation, methylation, acetylation, and hydroxylation; cofactors; and in rare cases, organisms have evolved novel translational machinery to incorporate either selenocysteine or pyrrolysine.^[4,5] The need for additional factors in so many protein-mediated processes suggests that while a code consisting of 20 amino acids is sufficient for life, it may not be ideal. Consequently, the development of a method that makes possible the systematic expansion of the genetic codes of living organisms might enable the evolution of proteins, or even entire organisms, with new or enhanced properties. Moreover, such methodology would dramatically increase our ability to manipulate protein structure and function both *in vitro* and *in vivo*. This in turn would allow a more classical chemical approach to the study of proteins, in which carefully defined changes in the steric or electronic properties of an amino acid are correlated with changes in the properties of proteins.^[6–8] Herein we provide an overview of both chemical and biochemical approaches that have been developed in the past to manipulate protein structure, and review in depth recent advances that have made it possible to add new amino acids to the genetic codes of both prokaryotic and eukaryotic organisms.

2. Chemical Approaches

2.1. Chemical Modification of Proteins

Reagents that allow the selective chemical modification of amino acid side chains have proven quite useful in assigning

functional roles to amino acid residues in proteins. Selectivity depends on the number and reactivity (dependent on both steric and electronic factors) of a particular amino acid side chain. Typical modifications include the oxidation or alkylation of cysteine residues, acylation of either the N-terminal α -amino group or the ϵ -amino group of lysine, and the condensation of amines with the carboxylate groups of aspartate or glutamate. In some cases it is possible to exploit the unique reactivity of active-site residues for selective chemical modification. For example, treatment of the protease subtilisin with phenylmethanesulfonyl fluoride, followed by reaction with thioacetate and hydrolysis, selectively converts the serine residue at the active site into a cysteine residue (Figure 1 a).^[9,10] This mutant thiosubtilisin has diminished proteolytic activity, but still functions as an esterase. Similar approaches have been used to generate a redox-active selenosubtilisin^[11,12] and selenotrypsin^[13] (Figure 1 b).

Chemical modification has also been used to introduce reactive molecules, biophysical probes, tags, and other non-peptidic groups selectively into proteins. Cysteine is the most commonly used residue for this purpose because of its relatively low abundance in proteins and the increased nucleophilicity of the thio group relative to the side chains of other amino acids. Moreover, site-directed mutagenesis makes it possible to selectively introduce a cysteine residue at any desired site in a protein, in many cases without loss of function. This residue can then be modified, typically by either disulfide bond exchange or alkylation reactions.

From the Contents

| | |
|--|----|
| 1. Introduction | 35 |
| 2. Chemical Approaches | 35 |
| 3. In Vitro Biosynthetic Approaches to Protein Mutagenesis | 39 |
| 4. In Vivo Protein Mutagenesis | 43 |
| 5. An Expanded Code | 46 |
| 6. Outlook | 61 |

[*] Prof. P. G. Schultz
 Department of Chemistry and
 The Skaggs Institute for Chemical Biology
 The Scripps Research Institute
 10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)
 Fax: (+1) 858-784-9440
 E-mail: schultz@scripps.edu
 Dr. L. Wang
 Department of Pharmacology, University of California
 San Diego, 9500 Gilman Drive, La Jolla, CA 92093 (USA)

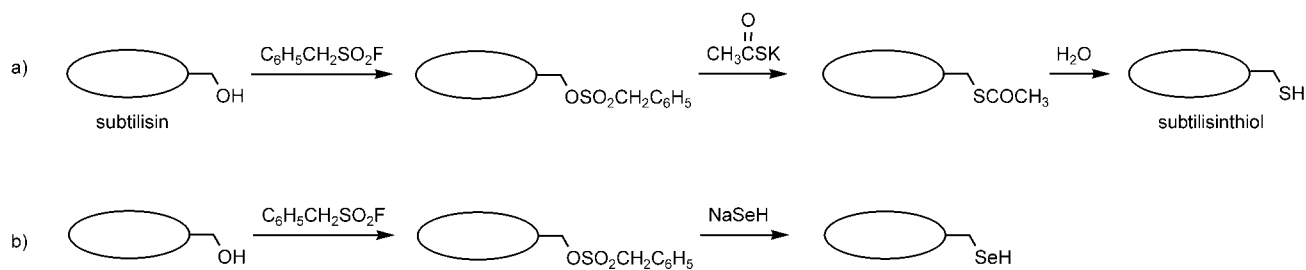


Figure 1. Selective conversion of the hydroxy group of serine to a thio (a) or seleno group (b).

Cysteine modification has been used to alter the catalytic properties of proteins. For example, Kaiser and co-workers conjugated different flavin analogues to the cysteine residue in the active site of papain to generate oxidoreductases that oxidize hydrophobic dihydronicotinamide derivatives (Figure 2a).^[14–16] Conversely, an oligonucleotide binding site was introduced into a Lys116→Cys mutant of the nonspecific deoxyribonuclease staphylococcal nuclease to generate a semisynthetic enzyme that selectively cleaves DNA adjacent to the target sequence (Figure 2b).^[17]

Cysteine has also been used to introduce a variety of biophysical probes into proteins. For example, a photoactivatable cross-linking agent (Figure 2c) was tethered to an Asp161→Cys mutant of catabolite gene activator protein to

probe the interaction of this protein with the α subunit of RNA polymerase during transcriptional activation.^[18] Meares and co-workers attached ferric EDTA to proteins through the Cys residue to generate oxidative cleavage reagents that can be used to map protein–protein and protein–nucleic acid complexes (Figure 2d).^[19–23] Likewise, Harbury and co-workers have created a powerful chemical footprinting technique termed misincorporation proton–alkyl exchange,^[24] which is based upon selective cleavage of the protein backbone at cysteine residues by 2-nitro-5-thiocyanobenzoic acid (Figure 2e).^[25] In the related “substituted cysteine accessibility method” (SCAM) the surface accessibility of residues is inferred by the ability of the corresponding cysteine mutant to react with charged or polymeric reagents.^[26,27] Nitroxide spin

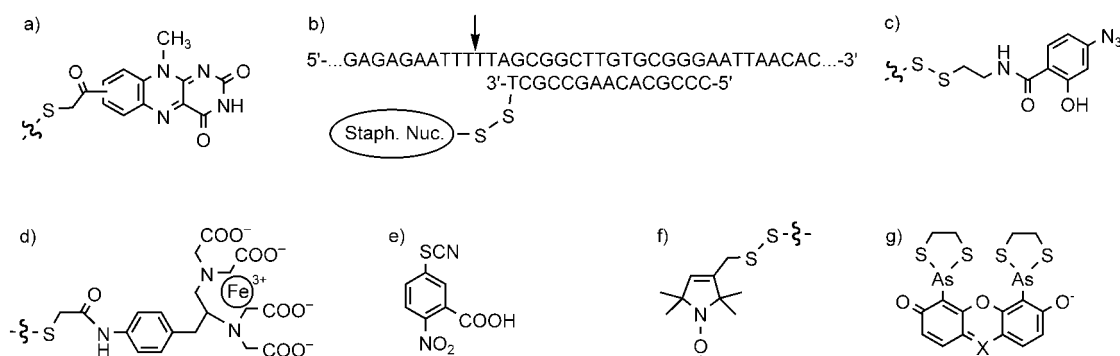
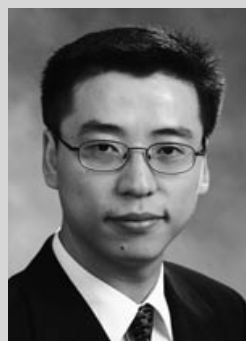


Figure 2. Cysteine residues can be modified with a wide range of nonpeptidic groups: a) flavin, b) oligonucleotide, c) an azido-based photo-cross-linking agent, d) [Fe(babc)], e) 2-nitro-5-thiocyanobenzoic acid, f) a nitroxide spin label, g) biarsenical dyes: FlAsH: X = C₆H₄COOH, ReAsH: X = N.



Peter Schultz received his B.S. and Ph.D. (1984) from Caltech. He joined the Berkeley faculty in 1985 and moved to the Scripps Research Institute in 1999 where he is currently the Scripps Professor of Chemistry and the Director of the Genomics Institute of the Novartis Research Foundation. His interests focus on catalytic antibodies; the development and application of combinatorial methods in biological chemistry, drug discovery, and materials science; and new building blocks for the genetic code. He is a founder of numerous companies, and his awards include the Alan T. Waterman Award (NSF), the Wolf Prize in Chemistry, the Paul Ehrlich Prize, and membership in the National Academy of Sciences and Institute of Medicine.



Lei Wang received both his BS in organic chemistry (1994) and M.S. in physical chemistry (1997) from Peking University, where he conducted research on nanoparticle properties using scanning probe microscopy. From 1997 to 2002, he pursued graduate studies at the University of California, Berkeley under the guidance of Prof. Peter G. Schultz, where he developed a general method for genetically introducing unnatural amino acids into proteins in live cells. He is currently a postdoctoral fellow in Prof. Roger Y. Tsien's group at the University of California, San Diego.

labels (Figure 2 f) and environmentally sensitive fluorophores have also been conjugated to various proteins through the Cys residue to probe side-chain dynamics and backbone fluctuations.^[28–30] Recently, it was shown that biarsenical ligands (FIAsH and ReAsH) can selectively bind to a tetracysteine motif Cys-Cys-Pro-Gly-Cys-Cys in proteins with high affinity and specificity (Figure 2 g).^[31,32] The fluorescence of the biarsenical dyes is greatly enhanced after binding to the target site, thus making possible the selective introduction of a fluorophore into proteins in cells.

Another chemical method that has been used to selectively modify proteins involves the oxidation of N-terminal Ser or Thr residues with periodate to form an aldehyde, which can subsequently react with hydrazide or aminoxy groups to form stable conjugates. In one application the reaction of aminoxyptentane with the N-terminal aldehyde group of the β -chemokine RANTES yielded a potent inhibitor of HIV-1 infection.^[33] Similar chemistry has been used to selectively couple biotin and fluorescent reporters to the N-termini of peptides and proteins,^[34] and for the chemoselective ligation of peptides (see Section 2.3).

Chemical modification is clearly a useful method for altering protein properties, including the synthesis of therapeutic proteins conjugated to polyethylene glycol (PEG; for example, Neulasta, PEG-Intron). Nonetheless, this approach relies on the unique reactivity of an amino acid side chain. Thus, the intrinsic selectivity and overall efficiency of the approach is limited.

2.2. Chemical Synthesis

Stepwise solid-phase peptide synthesis (SPPS),^[35] pioneered by Merrifield in 1963, has greatly facilitated the synthesis of peptides and small proteins (< 100 amino acids) containing unnatural amino acids.^[36] Examples include the substitution of a pyridoxamine phosphate derivative for Phe 8 of the C-peptide to afford a ribonuclease S with transaminase activity,^[37] and the incorporation of iminodiacetic acid into the S-peptide which allows the activity of RNase S to be switched on/off by addition of Fe^{III} ions.^[38] In other examples, D-amino acids and arginine analogues have been incorporated into luteinizing hormone-releasing hormone to improve affinity,^[39–41] and Leu and Val were replaced with their fluorinated analogues (5,5,5-trifluoroleucine and 4,4,4-trifluorovaline, respectively) in the coiled-coil region of the yeast transcription factor GCN4 to increase protein stability.^[42] In addition, unnatural amino acids bearing olefinic side chains have been inserted into the C-peptide of RNase A to form cyclic peptides by subsequent metathesis reactions, which resulted in increased resistance to proteases,^[43] and 5,5-dimethylproline was substituted for Pro 93 in RNase A to lock the peptide bond into a *cis* conformation as a probe of protein folding.^[44]

To overcome the size limitations of SPPS, efficient strategies have been developed to ligate synthetic peptides together to make larger proteins,^[45] including a chemoenzymatic strategy^[46] (see Section 2.3) and chemical ligation.^[47] The latter process takes advantage of chemoselective cou-

pling reactions (for example, thioester,^[47] oxime,^[48] thiazolidine/oxazolidine,^[49,50] thioether,^[51] and disulfide^[52] bond formation) of unique functional groups incorporated into peptides (Figure 3). Template molecules have also been used to bring together the respective C- and N-termini of the two peptide fragments, thus enabling formation of an amide bond by intramolecular acyl transfer reactions (Figure 3 c).^[53] An ingenious strategy termed “native chemical ligation” obviates the need for such a template and allows the direct coupling of peptide fragments to form a native peptide linkage (Figure 3 d).^[54,55] In this method, a peptide with an α -thioester at its C-terminus is first ligated to a second peptide with an N-terminal Cys residue through a transthioesterification reaction. The thioester-linked intermediate then undergoes an irreversible S \rightarrow N acyl rearrangement to form a native peptide bond at the ligation site. This method typically requires an N-terminal Cys residue at the ligation site, but recently it has been shown that an auxiliary sulfhydryl group can be used which is removed after ligation, thus making ligation possible at a non-Cys residue.^[56] Native chemical ligation can be routinely used for the preparation of polypeptides comprising over 100 amino acid residues in length.

The emergence of chemical ligation methods has made it possible to incorporate unnatural amino acids into larger proteins. For example, an HIV-1 protease analogue was synthesized in which the amide CONH linkage between Gly49 and Ile50 was replaced with a COS linkage by thioester ligation to probe the contribution of this backbone hydrogen bond to catalytic activity.^[57] In other studies native chemical ligation has been used to replace Tyr10 of rubredoxin with tyrosine analogues to modulate the redox properties of this protein,^[58] to modify erythropoietin with carbohydrate analogues,^[59] and to incorporate fluorophores into chymotrypsin inhibitor 2 for single-molecule studies of protein folding.^[60] Intramolecular native chemical ligation has also been used to create cyclic peptides and proteins.^[61–63]

Chemical synthesis allows a large variety of unnatural amino acids to be incorporated into peptides and proteins. It is particularly useful for the incorporation of amino acid analogues that are toxic to cells or incompatible with the translational machinery (for example, the synthesis of an all D-amino acid protein^[64,65]). Chemical synthesis also allows isotopic labels (for example, ²H-, ¹³C-, or ¹⁵N-labeled amino acids) to be inserted into proteins at specific locations for spectroscopic studies. On the other hand, chemical synthesis can be problematic with peptides or proteins that have poor solubility, and becomes tedious, low yielding, and expensive when applied to larger proteins.

2.3. Semisynthesis

An alternative approach to the generation of larger proteins containing unnatural amino acids involves semisynthesis in which synthetic peptides are ligated to a truncated native protein generated by chemical cleavage, proteolysis, or recombinant methods.^[66,67] In comparison to SPPS, much larger peptide fragments can be recombinantly expressed or

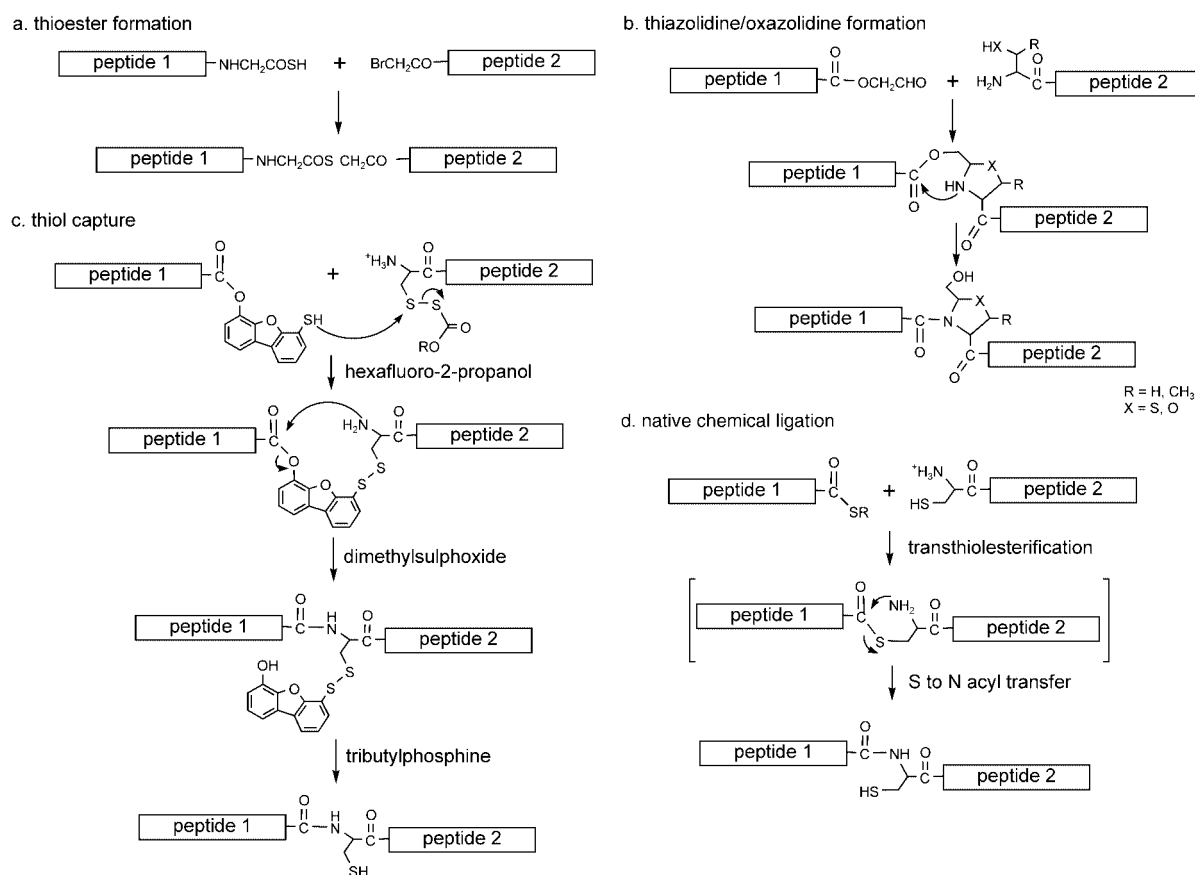


Figure 3. Examples of chemoselective reactions for peptide ligation.

derived from natural proteins, thus allowing access to larger proteins. One such semisynthetic method is based on the autocatalytic re-ligation of peptide fragments generated by cyanogen bromide (CNBr) cleavage of proteins, a phenomenon first noted by Dyckes et al. in 1974.^[68] CNBr cleavage results in a C-terminal homoserine lactone, which reacts with a free N-terminus to reform the full-length protein. This approach has been used to insert a number of unnatural amino acids into cytochrome *c*.^[69–71]

Enzyme-assisted peptide ligation has also been used for protein semisynthesis.^[72] Peptide hydrolysis catalyzed by proteases can be reversed by altering the reaction conditions to favor aminolysis over hydrolysis,^[73–76] for example, by adding organic cosolvents to suppress ionization of the α -carboxylic acid group. Alternatively, a peptide with a C-terminal ester can be condensed with the N-terminal amino group of a second peptide to form a ligated protein. Such methods have been used, for example, to incorporate D-Ala into human growth hormone releasing factor to increase protease resistance.^[77,78] A significant improvement to this method involves the use of a mutant protease which has decreased hydrolytic activity but can still catalyze amide-bond formation with a C-terminal ester.^[79] Kaiser and co-workers were the first to use a subtilisin mutant with a Ser \rightarrow Cys mutation in the active site^[9,10] for peptide ligation.^[79] This mutant was further improved by Wells and co-workers to generate a subtiligase, a subtilisin double mutant (Ser221 \rightarrow

Cys and Pro225 \rightarrow Ala) that has tenfold higher peptide ligase activity than thiosubtilisin.^[80] Subtiligase has been used to assemble a mutant RNase A in which the active-site residues His12 and His119 were independently replaced with 4-fluorohistidine.^[46] Similar approaches have also been used to incorporate biotin, heavy atoms, and polyethylene glycol into proteins.^[81]

Chemoselective ligation can also be used in semisynthesis as long as an orthogonal functional group can be conveniently included at the terminus of the recombinant peptide. For example, Rose and co-workers enzymatically coupled a hydrazide to the C-terminus of a peptide fragment from human granulocyte colony-stimulating factor (G-CSF), and chemically generated an N-terminal aldehyde on the complementary peptide fragment through oxidation with periodate.^[82] These two peptides spontaneously re-ligate to form a G-CSF mutant with a hydrazone linkage.

A technique termed expressed protein ligation (EPL) has greatly facilitated the use of native chemical ligation in protein semisynthesis by generating a recombinant protein with a C-terminal α -thioester for subsequent chemical ligation with a second peptide (Figure 4).^[83–86] The C-terminal thioester intermediate is generated from an intein-mediated protein-splicing reaction in which a mutation in the intein blocks the final splicing step. In EPL, a recombinant protein is expressed in-frame with the mutant intein, and exogenous thiol is added to cleave the recombinant protein from the

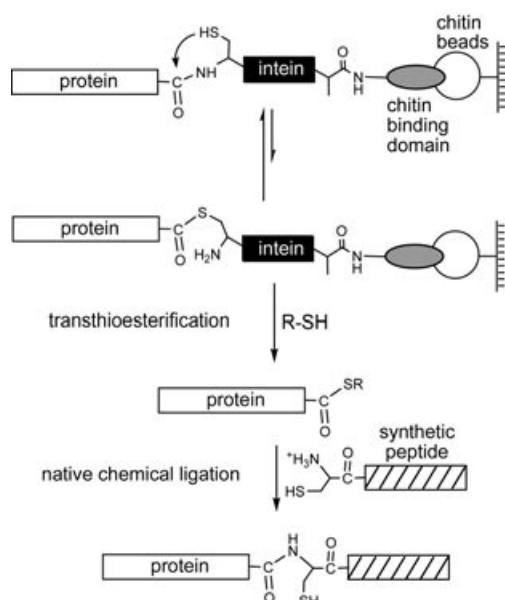


Figure 4. Expressed protein ligation.

partially rearranged splicing intermediate. The resulting α -thioester derivative can subsequently react with an N-terminal Cys peptide (prepared synthetically or enzymatically) to afford the semisynthetic product.^[87–89] In addition to an N-terminal cysteine residue, N-terminal homocysteine^[90,91] and selenocysteine^[92–94] can also be used in native chemical ligation and EPL, thus resulting in the selective incorporation of these amino acids into the semisynthetic protein at the ligation site.

EPL can be used to incorporate fluorophores into the C-terminus,^[83,95,96] middle,^[97,98] or N-terminus^[88] of proteins. In an elegant experiment Muir and co-workers introduced both tetramethylrhodamine and fluorescein into c-Crk-II (a substrate of the Abelson nonreceptor protein tyrosine kinase, c-Abl) to monitor the phosphorylation of c-Crk-II by c-Abl through fluorescence resonance energy transfer (FRET).^[99,100] FRET pairs have also been introduced by EPL and/or Cys labeling into different components of the RNA polymerase holoenzyme and RNA polymerase-promoter open complex of *E. coli*.^[101,102] In another application, a synthetic fragment of eglin which contained either kynurenine or norvaline at position 25 was ligated to a complementary peptide displayed on phage in which two residues making van der Waals contacts with residue 25 were randomized.^[103] After ligation, phage-bearing mutant folded eglins were isolated that accommodate the unnatural side chains in the hydrophobic core. EPL has also been used to introduce posttranslational modifications (for example, glycosylated^[104,105] and lipidated amino acids,^[106] as well as analogues of phosphorylated amino acids^[107–109]) selectively into proteins to study their effects on protein structure and function. Mutations to the protein backbone have also been generated with EPL. For example, Raines and co-workers replaced the Asn113–Pro114 residues of RNase A with a reverse-turn mimic composed of the two β -amino acid residues (*R*)-nipecotic acid–(*S*)-nipecotic acid^[110] to generate a mutant

RNase with the same ribonucleolytic activity as the wild-type enzyme and enhanced thermal stability.

Semisynthesis methods are quite powerful and allow access to larger proteins than does chemical synthesis. The method does, however, require appropriate sites for cleavage and ligation, and becomes cumbersome for internal sites in large proteins or when large amounts of proteins are required. In addition, all of the chemical methods described above are not easily applied to in vivo studies of protein structure and function.

3. In Vitro Biosynthetic Approaches to Protein Mutagenesis

3.1. Methodology

A variety of in vitro (cell-free) methods have been developed to incorporate unnatural amino acids into proteins by using the existing protein biosynthetic machinery of the cell. These methods take advantage of the fact that anticodon–codon recognition between messenger RNA (mRNA) and transfer RNA (tRNA) is largely independent of the structure of the amino acid linked to the 3'-terminus of the acceptor stem of the tRNA. This “adaptor hypothesis”^[111,112] was elegantly demonstrated by showing that a cysteinyl-tRNA^{Cys}, when converted into an alanyl-tRNA^{Cys} by treatment with H₂/Raney nickel, efficiently incorporates alanine into a polypeptide chain in vitro in response to the cysteine codon UGU.^[112] These experiments were extended to derivatives of other common amino acids by the chemical modification of enzymatically aminoacylated tRNAs with synthetic reagents and probes. For example, treatment of Lys-tRNA^{Lys} with *N*-acetoxy succinimide affords (*N*^ε-acetyl-Lys)-tRNA^{Lys}, which can be used to insert *N*^ε-acetyl-Lys into proteins in a cell-free rabbit reticulocyte system.^[113] Unfortunately, these methods lead to incorporation of the modified amino acid at all sites in a protein specified by the codon for the modified tRNA. The derivatized amino acid is also inserted in competition with the underivatized amino acid thus leading to heterogeneous protein product. In addition, substitutions are limited to derivatives of the common amino acids which can be generated under reactive conditions that do not hydrolyze the labile aminoacyl ester linkage between the terminal 3'-adenosine group and the amino acid, and do not chemically inactivate the tRNA.

Extension of this approach to a larger variety of amino acids required more general methods for the selective aminoacylation of tRNAs. Misaminoacylation of the tRNA with unnatural amino acids using aminoacyl-tRNA synthetases is problematic because of the high specificity of these enzymes.^[114,115] Direct chemical acylation of the tRNA is not practical because of the large number of reactive sites in the tRNA. Consequently, semisynthetic methods were developed in which truncated tRNAs (with the 3'-terminal mono- or dinucleotide removed) are enzymatically ligated to chemically aminoacylated mono- and dinucleotides. For example, Hecht and co-workers developed an approach in which the dinucleotide pCpA is chemically acylated with an *N*^ε-

protected amino acid and then enzymatically ligated to a truncated tRNA (tRNA^{-CA}, which is missing the terminal dinucleotide pCpA at the 3'-acceptor stem) with RNA ligase.^[116] A number of unnatural amino acids were incorporated into the first position of dipeptides by this method.^[117–119] However, this approach suffered a number of limitations: yields of the aminoacylation reaction were quite poor and the N-terminal protecting group of the amino acid could not be efficiently removed without hydrolysis of the aminoacyl ester, thereby limiting substitution to P-site donors and, consequently, the peptide N-terminus (Figure 5). Later,

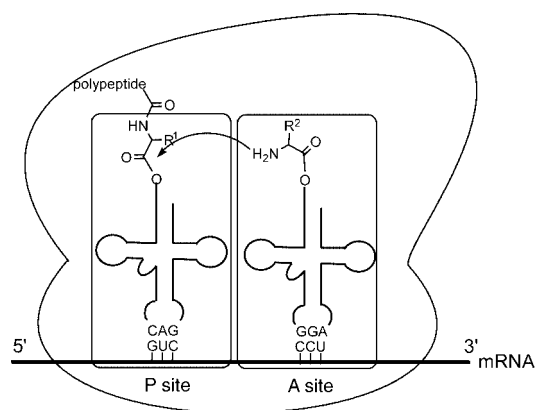


Figure 5. Schematic representation of ribosome and its P site and A site.

Brunner and co-workers developed a method to prepare a fully deprotected aminoacyl-pCpA and ligate it to the truncated tRNA to afford an aminoacyl tRNA which could be used as an A-site donor.^[120] Nonetheless, the yields for the synthesis of the acylated dinucleotide were again poor. This method also does not overcome the above limitations associated with use of an endogenous *E. coli* tRNA for delivery of an amino acid including competition with host amino acids and multisite incorporation. Nonetheless, experiments with chemically aminoacylated tRNAs suggested that the natural translational machinery can accept a wide range of unnatural amino acid side chains.

A general *in vitro* method which allows the site-specific incorporation of a large number of unnatural amino acids into proteins with excellent translational fidelity was reported by us in 1989 (Figure 6).^[121] This approach takes advantage of the degeneracy of the three stop codons UAA, UAG, and UGA (termed nonsense codons). These codons do not encode amino acids, but rather signal termination of polypeptide synthesis by binding release factors.^[122,123] Since only one stop codon is required for the termination of protein synthesis, two “blank” codons exist in the genetic code which can be used to uniquely specify an unnatural amino acid. Indeed, nonsense suppressor tRNAs (tRNAs that insert one of the common amino acids in response to a stop codon) have been shown to incorporate a number of common amino acids efficiently into proteins both *in vivo* and *in vitro*.^[124,125] In particular, experiments using [³H]Phe, and later experiments with α -hydroxy

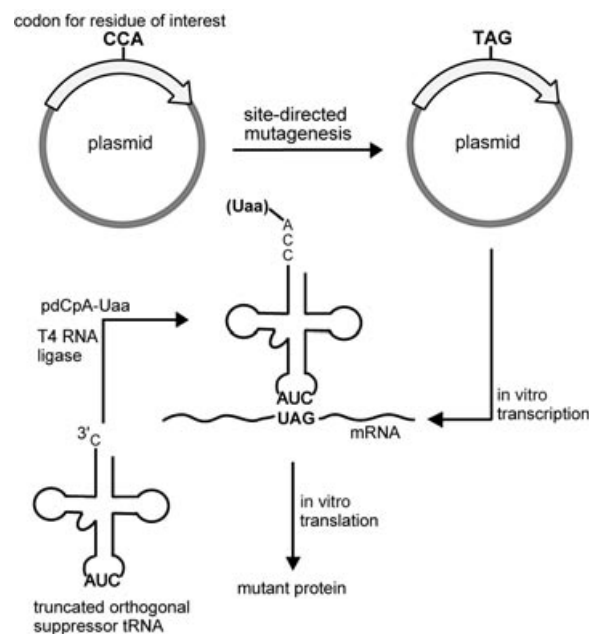


Figure 6. An *in vitro* biosynthetic method for the site-specific incorporation of unnatural amino acids into proteins. Uaa = unnatural amino acid.

acids, demonstrated that the desired amino acid could be selectively incorporated at the position specified by UAG in an *in vitro* translation system.^[121,126]

Another requirement for the selective incorporation of unnatural amino acids into proteins is a tRNA that uniquely recognizes this blank codon and efficiently incorporates its cognate amino acid into the growing polypeptide chain. Importantly, this tRNA must be orthogonal to the endogenous aminoacyl-tRNA synthetases of the host organism from which the *in vitro* translation system is derived, that is, it must not be a substrate for any of the aminoacyl-tRNA synthetases present in the *in vitro* protein-synthesis extract. If the suppressor tRNA were recognized by an endogenous aminoacyl-tRNA synthetase, the tRNA could be subject to proof-reading (deacylation of the noncognate amino acid) and/or re-aminoacylation with the cognate amino acid. This would result in either low suppression efficiency or the incorporation of a common amino acid in competition with the desired unnatural amino acid in response to the nonsense codon.

An orthogonal amber suppressor tRNA derived from yeast phenylalanine-tRNA (tRNA^{Phe}_{CUA}) was constructed for use in an *E. coli* *in vitro* protein transcription–translation system.^[125,127–130] In this tRNA nucleotides 34–37 in the anticodon loop were replaced by 5'-CUAA-3' to create an efficient suppressor tRNA. Runoff transcription provided a functional suppressor tRNA in relatively large quantities.^[131,132] A modification of the two-step method originally developed by Hecht and co-workers^[116] was developed to efficiently aminoacylate the orthogonal tRNA. It was found that the cyanomethyl ester of an *N*^α-protected amino acid selectively monoacylates the 2',3'-hydroxy groups of the dinucleotide to give the desired aminoacylated product in

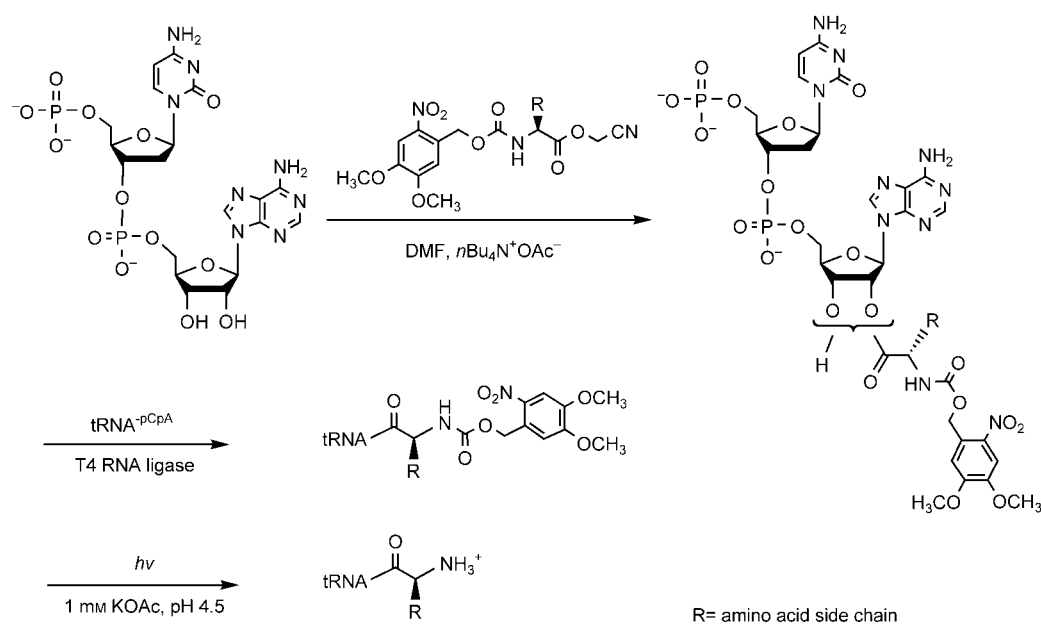


Figure 7. Strategy for the chemical aminoacylation of tRNA^{Phe}_{CUA} using an amino acid cyanomethyl ester protected with a photochemically removable group.

high yield, thereby obviating the need to protect and deprotect the dinucleotide (Figure 7).^[133] Replacement of cytidine with deoxycytidine in pCpA simplified the synthesis and eliminated another reactive 2'-OH group without affecting biological activity. The α -amino group of the amino acid and any reactive side-chain groups are protected as their nitroveratryloxy carbamate, ester, or ether derivatives prior to acylation of the dinucleotide.^[134,135] These protecting groups can be removed photochemically from the intact aminoacyl-tRNA (after ligation) in high yield under mildly acidic conditions which prevent deacylation of the aminoacyl-tRNA.^[133] This aminoacylation protocol is relatively straightforward, proceeds in high yield (both the aminoacylation and ligation reactions), and has been used to incorporate a wide variety of unnatural amino acids into proteins.^[136]

The first demonstration of this general approach for the site-specific incorporation of unnatural amino acids into a protein *in vitro* involved the generation of mutant β -lactamases containing *p*-nitrophenylalanine, *p*-fluorophenylalanine, or homophenylalanine substitutions at Phe66.^[121] Mutagenesis with the chemically acylated suppressor tRNA was carried out with an *E. coli* *in vitro* transcription–translation system consisting of an *E. coli* S-30 extract that contains all of the factors required for transcription and translation, and plasmid DNA containing the gene of interest with a TAG amber codon encoding the unnatural amino acid. Although relatively small quantities of purified protein were obtained (typical yields are 5–30 μ g protein per mL), the specificity and fidelity of incorporation were shown to be very high by peptide mapping experiments. Shortly thereafter, a similar approach was reported by Chamberlin and co-workers, in which a suppressor tRNA derived from *E. coli* tRNA^{Gly} was used to incorporate [¹²⁵I]Tyr site-selectively into a 16-mer peptide.^[137]

Over 50 different amino acids and analogues have been incorporated site-specifically into proteins with high fidelity by the above methodology.^[136] These include conformationally restricted amino acids (for example, methanoproline, cyclopropylglycine, and α -methyl-L-leucine), amino acids with spin labels and photoaffinity labels as side chains, amino acids with altered acidities and hydrogen-bonding properties, photocaged amino acids, α -hydroxy acids, and amino acids with unusual steric properties (for example, *tert*-butylglycine).

3.2. Applications

In vitro methods for unnatural amino acid mutagenesis have been applied to a large number of problems in protein chemistry, including protein folding, enzyme mechanism, and signal transduction. In addition, this methodology has been used to introduce a variety of biophysical probes into proteins.^[138] For example, to determine the degree to which side-chain hydrogen bonding stabilizes the folded state of proteins, Tyr27 in staphylococcal nuclease (SNase) was substituted with several fluorinated tyrosine analogues (Figure 8a).^[139] Denaturation studies of the corresponding mutants revealed a free-energy correlation between stability constant K_{app} and the pK_a value of the Tyr27 hydroxy group (Figure 8b), thus providing strong evidence that intramolecular hydrogen bonds between side chains can preferentially stabilize the folded state of a protein relative to the unfolded state in water. To determine the contribution of the main-chain hydrogen bonds to protein stability, several backbone amide linkages in α -helix 39–50 of T4 lysozyme were replaced with ester linkages by incorporating α -hydroxy acids at the N-terminus, the middle, and C-terminus of the helix, respec-

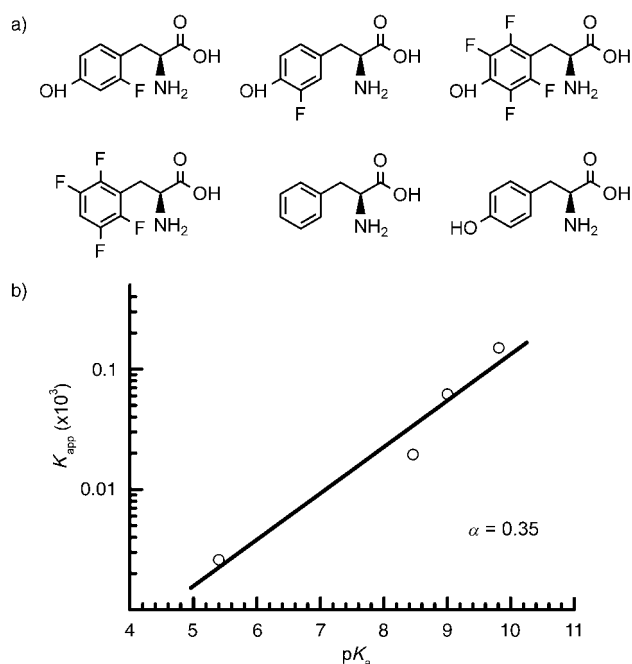


Figure 8. a) Structures of fluorinated amino acids substituted for Tyr27 in SNase. b) Free energy relationship between the pKa value of the Tyr27 mutant and the stability of SNase.

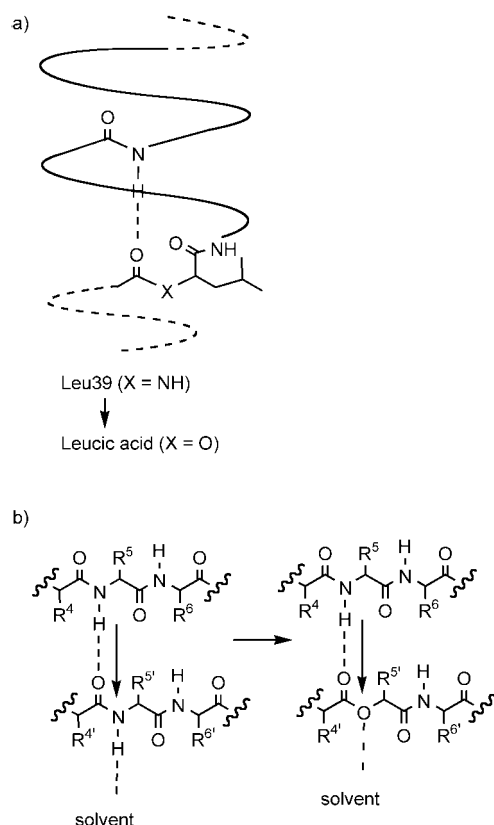


Figure 9. Schematic representation of backbone mutations generated by incorporation of α -hydroxy acids: a) N-terminal mutation of Leu 39 to leucic acid in an α -helix of T4 lysozyme; b) substitution of Leu 14 with leucic acid in a β sheet of SNase.

tively (Figure 9a).^[140] The ester group is a conservative replacement for an amide group since both esters and amides are isosteric and exist mainly in the *trans* conformation.^[141] These mutations effectively substitute a good hydrogen-bond acceptor (the amide carbonyl group) with a considerably weaker hydrogen-bond acceptor (the ester carbonyl group).^[142] It was found that at the N- and C-terminal positions, where only one hydrogen-bonding interaction is perturbed, the ester substitution destabilizes the protein by 0.9 and 0.7 kcal mol⁻¹, respectively. Introduction of ester linkage in the middle of the helix, where two hydrogen-bonding interactions were altered, destabilizes the protein by 1.7 kcal mol⁻¹. Similarly, amide to ester substitutions introduced in antiparallel β sheets were found to decrease the stability by 1.5–2.5 kcal mol⁻¹ (Figure 9b).^[143] These results provide evidence that backbone hydrogen-bonding interactions involving the amide groups, like side-chain hydrogen bonds, contribute significantly to protein stability.

3.3. Extensions and Improvements of Methodology

A number of improvements have been made to increase the levels of in vitro protein expression. For example, an *E. coli* tRNA^{Asn}-derived suppressor led to a significant improvement in suppression efficiency relative to the yeast tRNA^{Phe}_{CUA} for incorporation of polar unnatural amino acids in the *E. coli* in vitro system.^[144] Mild heating of an S-30 extract derived from an *E. coli* release factor 1 (RF1) temperature-sensitive strain increased UAG suppression by decreasing the competition between RF1 and the suppressor tRNA for the amber stop codon.^[145] Optimization of the preparation and composition of the *E. coli* S-30 extract, as well as removal of inhibitory by-products generated during the translation reaction, also greatly improved yields of the protein.^[146]

Efforts have also focused on encoding unnatural amino acids with codons other than nonsense codons. For example, the rare arginine codon AGG has been used to direct the incorporation of several phenylalanine analogues into a polypeptide using a chemically aminoacylated tRNA^{Phe}_{CCU}.^[147] In this case, the naturally occurring cognate tRNA, although present in minor amounts, competes with the synthetic tRNA to incorporate Arg and generate heterogeneous full-length protein (in contrast to truncated protein in the case of the nonsense codon). In the case of the rare AGG codon, lowering the arginine concentration in the in vitro protein synthesis reaction mixture decreased such competition.^[148] Some organisms do not use all of the triplet codons; these noncoding codons can be used to specify additional amino acids. For example, the unassigned codon AGA in *Micrococcus luteus* has been used to insert Phe in an in vitro transcription/translation extract from *M. luteus*.^[149] Recently, a purified in vitro translation system consisting of only ribosomes, initiation factors, and elongation factors was used to simultaneously incorporate several unnatural amino acids into peptides in response to sense codons.^[150] If the efficiency can be improved, this system may ultimately allow the synthesis of peptides and proteins containing multiple unnatural amino acids.

Frameshift suppression of four-base codons can also be used to introduce unnatural amino acids into proteins in vitro. The proper reading frame is maintained only when the extended codon is translated by the suppressor tRNA to afford a full-length protein containing the unnatural amino acid; translation of the triplet codon by an endogenous tRNA results in frameshifting and premature termination. To reduce this competition, the first three nucleotides in most extended codons correspond to either stop or rare codons. In one example, Sisido and co-workers replaced the anticodon of the yeast tRNA^{Phe} with different four-base anticodons, and incorporated a variety of phenylalanine analogues into streptavidin with an in vitro *E. coli* translation system.^[151–156] By using two chemically aminoacylated frameshift suppressor tRNAs that recognize either CGGG or AGGU, they were able to simultaneously incorporate 2-naphthylalanine and a 7-amino-4-nitrobenz-2-oxa-1,3-diazole (NBD) derivative of lysine into streptavidin, although the efficiency was quite low (9%).^[157] Incorporation efficiency was improved to 64% when efficient four-base codons (CGGG and GGGU) were used.^[158] A five-base CGGUA-UACCG codon-anticodon pair was also shown to direct incorporation of unnatural amino acids into streptavidin.^[159]

A codon–anticodon pair containing unnatural bases (iso-C)AG-CU(iso-dG) has been used to incorporate iodotyrosine into a short peptide with the rabbit reticulocyte lysate.^[160] This study convincingly showed that the ribosome tolerates nucleotides with unnatural bases, and that iso-C and iso-dG can direct translation with reasonable efficiency. However, the mRNA and tRNA containing iso-C and iso-dG must be chemically synthesized, thus making this method impractical for most applications.

Unnatural amino acid mutagenesis has been extended to *Xenopus* oocytes by microinjection of the mutant mRNA and a chemically aminoacylated yeast tRNA^{Phe}-derived amber suppressor (Figure 10).^[161] The translational machinery of the oocyte inserts the unnatural amino acid at the position specified by UAG. Later, an amber suppressor derived from *Tetrahymena thermophila* tRNA^{Gln} was found to improve the efficiency and fidelity of unnatural amino acid incorporation.^[162] This method has been used for structure–function studies of integral membrane proteins, which are generally not amenable to in vitro expression systems.^[163,164] For example, fluorinated tryptophan derivatives were incorporated at Trp149 in the binding domain of the nicotinic acetylcholine receptor. A correlation between the receptor activation by acetylcholine (ACh) and the degree of fluorination on tryptophan suggested that the quaternary ammonium center of ACh forms a cation– π interaction with the indole side chain of α Trp149.^[165,166] A significant limitation of this method is that the suppressor tRNA has to be chemically aminoacylated with the unnatural amino acid in vitro, and the acylated tRNA is consumed as a stoichiometric reagent during translation and cannot be regenerated. Therefore, yields of proteins are low and highly sensitive techniques, such as patch clamp, are required to characterize the mutant proteins.

The in vitro mutagenesis methodology described above has been used to site-specifically incorporate a large number

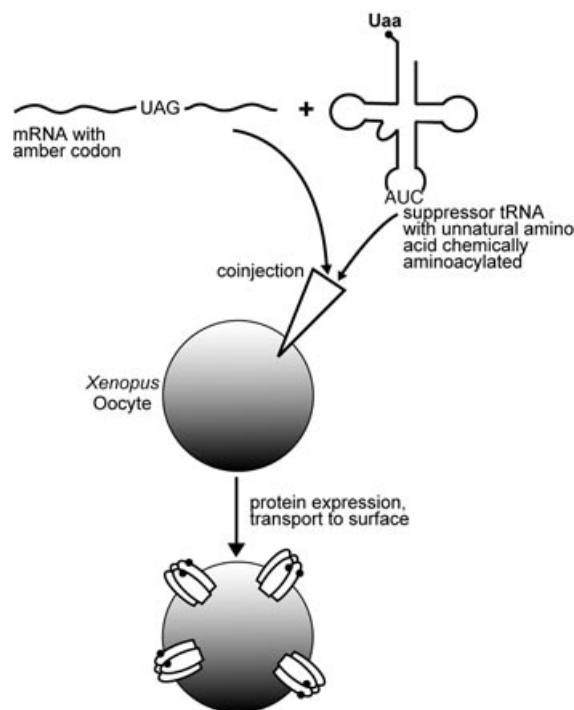


Figure 10. Microinjection of chemically aminoacylated tRNA and mutant mRNA into *Xenopus* oocytes.

of unnatural amino acids into various proteins. Mutagenesis studies using unnatural amino acids have provided important insights into protein structure and function. However, the method is technically demanding, the yields of mutant proteins are quite low, and the study of proteins in cells is limited by the need for microinjection and inability to regenerate the aminoacyl tRNA. Moreover, the nature of the amino acid side chain is limited by the aminoacylation chemistry and the stability of the aminoacyl tRNA linkage.

4. In Vivo Protein Mutagenesis

The ability to incorporate unnatural amino acids directly into proteins in vivo has several advantages over in vitro methods including high yields of mutant proteins, technical ease, and the potential to study proteins in cells or multicellular organisms. The latter can be accomplished either by substitution of one of the common 20 amino acids with an unnatural amino acid, or by augmenting the genetic code of the host organisms with a “21st” unnatural amino acid. In the former case, the unnatural amino acid typically replaces the common amino acid throughout a protein (and the proteome), often in competition with a common amino acid, and results in heterogeneous proteins. In the second approach, unnatural amino acids are added to the genetic repertoire without deletion of a common amino acid, thus resulting in an expanded genetic code and the ability to produce proteins consisting of more than 20 building blocks.

4.1. Site-Directed Mutagenesis

The development of site-directed mutagenesis by Smith and co-workers provided a powerful *in vivo* method to substitute amino acids in proteins with any of the other common 20 amino acids.^[167] Amino acid substitutions are generated by mutating the codon for the amino acid of interest to a codon encoding the desired mutant. This method has dramatically expanded our ability to manipulate protein structure and probe protein structure–function relationships. Notable early examples included studies of β -lactamase,^[168,169] tyrosyl-tRNA synthetase (TyrRS),^[170] dihydrofolate reductase,^[171] trypsin,^[172] and subtilisin.^[173] Mutagenesis studies of *Bacillus stearothermophilus* TyrRS by the Fersht research group were used to quantify the contributions of various amino acid side chains to the binding of substrates, intermediates, and transition states along the reaction pathway of this enzyme.^[174] In other examples, replacement of the complementarity-determining regions of a human antibody with those of a mouse antibody afforded a humanized antibody with the specificity of the latter;^[175] site-directed mutagenesis has also been used to modify the specificity of Zn²⁺ finger proteins.^[176–183]

Many variants of site-directed mutagenesis have been developed. For example, cassette mutagenesis allows for the rapid generation of all possible mutations at a single site.^[184] Homologue- and alanine-scanning mutagenesis provide valuable information about protein function in the absence of structural data.^[185,186] Random mutagenesis by error-prone polymerase chain reaction (PCR) or mutator strains allows for a less biased examination of structure–function relationships over a larger sequence space.^[7,187–189] Genetic recombination methods, such as DNA shuffling,^[190] a recombinative PCR process akin to molecular breeding, can be used to generate libraries of variants from a pool of parent genes. Desired protein properties can be rapidly evolved from these variants with appropriate selections or screens. All of these methods have greatly accelerated our ability to probe protein structure and function and evolve novel protein properties. Unfortunately, all are also limited to substitutions among the common 20 amino acids.

4.2. The Use of Auxotrophic Bacterial Strains To Incorporate Unnatural Amino Acids

More than 50 years ago it was found that many analogues of the common 20 amino acids inhibit the growth of bacteria.^[191–193] This observation led to speculation that these analogues are incorporated into proteins by the protein biosynthetic machinery. To test this possibility, Tarver and co-workers fed rats^[194] and *Tetrahymena*^[195] with the ¹⁴C-labeled methionine analogue ethionine. The detection of radioactivity in the isolated protein and peptides from these organisms strongly suggested that unnatural amino acids could in fact be misincorporated into proteins. To efficiently incorporate analogues of the common 20 amino acids into bacterial proteins, Cohen and co-workers used an auxotrophic bacterial strain that cannot biosynthesize the corresponding common

amino acid. Induction of protein expression in media in which this common amino acid is replaced with its analogue results in the accumulation of a mutant protein containing the unnatural amino acid (Figure 11).^[196]

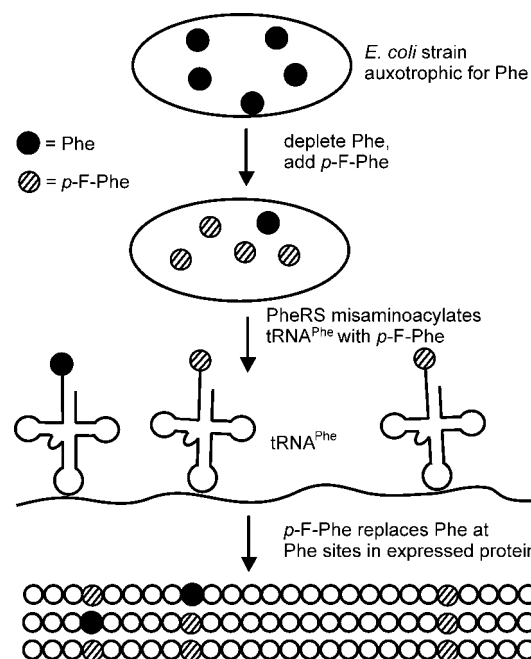


Figure 11. Multiple substitution of phenylalanine by *p*-fluorophenylalanine (*p*-F-Phe) using a phenylalanine auxotrophic strain.

This approach was used to incorporate *p*-fluorophenylalanine into β -galactosidase in an *E. coli* strain auxotrophic for phenylalanine.^[196] Fluorophenylalanine was found to partially replace phenylalanine and tyrosine throughout the protein. Selenomethionine was similarly incorporated into β -galactosidase using a methionine auxotroph.^[197,198] Over 60 analogues of the common 20 amino acids have subsequently been incorporated into proteins by this method.^[193,199] Typical substitutions include replacement of hydrogen with fluorine, methylene with oxygen or sulfur, and ring substitutions. For example, tryptophan has been replaced with tryptazan,^[200] arginine with canavanine,^[201] methionine with norleucine,^[202] leucine with trifluoroleucine,^[203] and the phenyl ring of phenylalanine with thioenyl or furyl rings. Most of the early work focused on studies of the effects of amino acid analogues on bacterial growth and protein synthesis and processing.^[199] More recently these methods have proven valuable for incorporating NMR and X-ray probes into proteins.

4.2.1. Relaxing the Substrate Specificity of Aminoacyl-tRNA Synthetases

Incorporation of amino acid analogues by the above approach relies on the promiscuity of aminoacyl-tRNA synthetases. However, these enzymes have very high substrate specificity to insure high fidelity in protein translation, which limits substitutions to close structural analogues of the

common 20 amino acids. For example, although thiaproline can be incorporated quantitatively into proteins, oxaproline and selenaproline cannot.^[204] To overcome this limitation, a number of research groups have attempted to relax the substrate specificity of the aminoacyl-tRNA synthetases. Hennecke and co-workers first reported that replacement of Ala294 with Gly in *E. coli* phenylalanyl-tRNA synthetase (PheRS) increases the size of the substrate binding pocket and results in the acylation of tRNA^{Phe} with *p*-chlorophenylalanine (*p*-Cl-Phe).^[205] An *E. coli* strain harboring this mutant PheRS allows the substitution of phenylalanine with *p*-Cl-Phe.^[206] Tirrell and co-workers later used this mutant synthetase to incorporate *p*-bromo-, *p*-iodo-, *p*-cyano-, *p*-ethynyl-, and *p*-azidophenylalanine as well as pyridylalanine.^[207,208] Similarly, a point mutation (Phe130→Ser) near the amino acid binding site of *E. coli* tyrosyl-tRNA synthetase allows azatyrosine to be incorporated more efficiently than tyrosine.^[209] In these cases, the isolated protein typically contains a mixture of both the common and unnatural amino acid at the site of substitution.

4.2.2. Attenuating the Proofreading Activity of Aminoacyl-tRNA Synthetases

The fidelity of tRNA aminoacylation is maintained both by substrate discrimination and by proofreading of non-cognate intermediates and products. Therefore, synthetases with impaired proofreading activity can also be used to incorporate unnatural amino acids into proteins by allowing close structural analogues of the cognate amino acid to escape the editing function. This approach was first demonstrated with a valyl-tRNA synthetase (ValRS) from *E. coli* (Figure 12).^[210] ValRS can misaminoacylate tRNA^{Val} with Cys, Thr, or aminobutyrate (Abu), but these noncognate amino acids are subsequently hydrolyzed by the editing domain. After random mutagenesis of the *E. coli* chromosome, a mutant ValRS was selected that has a mutation in the editing domain, which causes charging of tRNA^{Val} with Cys. Since Abu sterically resembles Cys (the thiol group of Cys is replaced with a methyl group in Abu), the mutant ValRS also

incorporates Abu into proteins when this mutant *E. coli* strain is grown in the presence of Abu. Mass spectrometric analysis shows that about 24% of valine is replaced by Abu at each valine position in the native protein. Tirrell and co-workers subsequently used a Thr252→Tyr252 mutant for the partial incorporation of six leucine analogues into proteins.^[211]

4.2.3. Applications

One of the most important applications of the auxotrophic strain method of mutagenesis is the incorporation of amino acid analogues containing heavy atoms for phase determination in X-ray crystallography. Substitution of methionine with selenomethionine is a convenient way to introduce a heavy atom for multiwavelength anomalous diffraction.^[212,213] Similarly, telluromethionine^[214–216] and β -selenolo[3,2- β]pyrrolylalanine^[217] have been incorporated into proteins using methionine and tryptophan auxotrophic *E. coli* strains, respectively.

Isosteric analogues of amino acids have also been incorporated into proteins in *E. coli* to study protein folding, stability, and activity. In one example, norleucine, selenomethionine, and telluromethionine (in which the sulfur atom of methionine is replaced with methylene, Se, and Te, respectively) were incorporated into human annexin V to probe the effects of side-chain packing interactions on protein stability.^[218] Similar studies have been carried out with fluorinated amino acids;^[219–221] for example, incorporation of trifluoroleucine and hexafluoroleucine in place of leucine resulted in increased thermal and chemical stability of a leucine-zipper protein.^[222,223]

Amino acid analogues can also be used as spectroscopic probes. For example, the fluorescence emission maximum and intensity of a barstar mutant containing a tryptophan to aminotryptophan mutation are sensitive to the pH value of the local environment,^[224] and replacement of Tyr66 of enhanced cyan fluorescent protein with 4-aminotryptophan significantly red-shifts the emission peak to produce a “gold” fluorescent protein ($\lambda_{\text{max}} = 574 \text{ nm}$).^[225] Fluorine-substituted amino acids can be used as NMR probes. For example, fluoromethionine has been substituted for methionine in bacteriophage λ -lysozyme to study ligand binding by ¹⁹F NMR spectroscopy,^[226–228] and fluorinated tryptophan has been incorporated for ¹⁹F NMR studies in cyan fluorescent protein^[229] and the chaperone PapD.^[230]

Nonproteinogenic functional groups that provide selective handles for further protein modification have also been introduced into proteins by these methods. Examples include the replacement of phenylalanine with electroactive 3-thienylalanine,^[231] and the substitution of methionine with alkenes and alkynes.^[232–235]

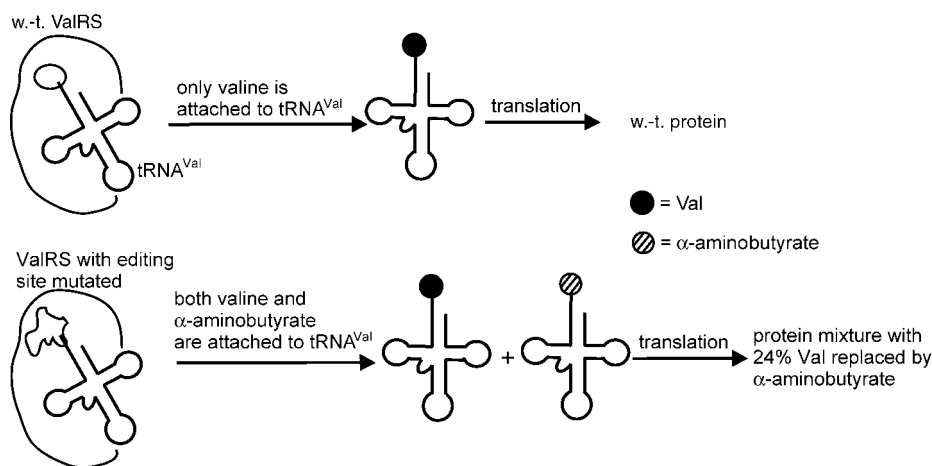


Figure 12. Incorporation of α -aminobutyrate by a ValRS mutant with reduced proofreading activity. w.-t. = wild-type.

Azidohomoalanine has been incorporated using a methionine auxotrophic strain, and can be selectively modified with triarylphosphine reagents by Staudinger ligation,^[236] or with alkyne reagents by a copper(I)-catalyzed cycloaddition reaction.^[237] Finally, a computationally designed *E. coli* PheRS mutant was used to partially incorporate *p*-acetylphenylalanine into proteins; the ketone group was subsequently labeled with biotin hydrazide under mild conditions.^[238]

Though very useful, the use of auxotrophic strains for mutagenesis has a number of limitations. This method is not site-specific—all sites corresponding to a particular amino acid throughout a protein (and the entire proteome) are replaced with the amino acid analogue. The extent of incorporation of the unnatural amino acid may also vary—only in rare cases can quantitative substitution be achieved since it is difficult to completely deplete the endogenous amino acids inside the cell. Finally, this method is applicable in general only to close structural analogues of the common amino acids, since substitutions must be tolerated both by the synthetase and at all nonpermissive sites in the proteome in which a particular amino acid is incorporated.

5. An Expanded Code

Site-directed mutagenesis allows the selective substitution of amino acids in a protein with the other common 20 amino acids. The use of auxotrophic strains allows the substitution of an amino acid with a close structural analogue throughout the proteome, typically in competition with the corresponding common amino acid. Ideally, one would like to be able to carry out the equivalent of site-directed mutagenesis, but with unnatural amino acids. This scenario requires that the unnatural amino acid be genetically encoded by a unique codon in a manner similar to that of the common 20 amino acids.

5.1. Methodology

The “21st” amino acid selenocysteine (Sec) is naturally encoded by an in-frame UGA non-sense codon that is translated by a unique Sec-tRNA^{Sec} (generated by enzymatic modification of seryl-tRNA^{Sec}) in the presence of a special elongation factor and a *cis*-acting element in the mRNA.^[4] However, in theory the only requirements to genetically encode unnatural amino acids *in vivo* are a unique tRNA-codon pair, a corresponding aminoacyl-tRNA synthetase (hereafter referred to as a synthetase), and the unnatural amino acid (Figure 13). The tRNA must be constructed such that it is not recognized by the endogenous aminoacyl-tRNA synthetases of the host, but functions efficiently in translation (an orthogonal tRNA). This tRNA must deliver the novel amino acid in response to a unique codon that does not encode any of the common 20 amino acids. A new aminoacyl-tRNA synthetase (an orthogonal synthetase) is also required that aminoacylates the orthogonal tRNA, but none of the endogenous tRNAs. This synthetase must aminoacylate the tRNA with only the desired unnatural amino acid and no endogenous amino acid. Likewise, the unnatural amino acid

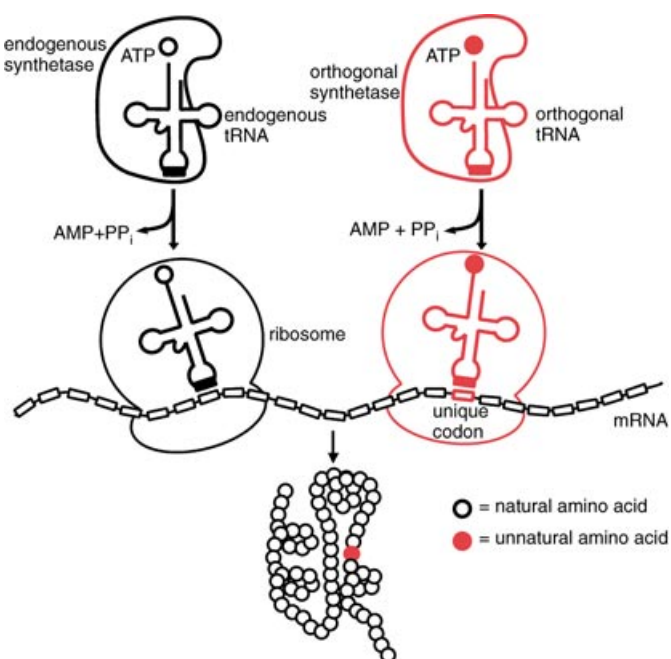


Figure 13. A general approach for the site-specific incorporation of unnatural amino acids into proteins *in vivo*. PP_i = pyrophosphate.

cannot be a substrate for the endogenous synthetases. Finally, the amino acid must be efficiently transported into the cytoplasm when added to the growth medium, or biosynthesized by the host.^[239]

This potentially general approach to mutagenesis with unnatural amino acids was initially developed using *E. coli*, since it is easily genetically manipulated, has a high transformation efficiency, and many established selections and screens already exist. Initially the amber nonsense codon (UAG) was used to encode the unnatural amino acid since it is the least used of the three stop codons in *E. coli* and *S. cerevisiae*. Moreover, some *E. coli* strains contain natural amber suppressor tRNAs that efficiently incorporate common amino acids without significantly affecting growth rates (most genes essential for normal *E. coli* growth do not end in TAG).^[240,241] Amber suppressors can also be engineered, and both natural and engineered amber suppressors have been routinely used for conventional protein mutagenesis in *E. coli*^[124,125,242] as well as the *in vitro* introduction of unnatural amino acids into proteins. In addition, nonsense suppressors have been either identified or generated for use in mammalian cells,^[243,244] yeast,^[245,246] and *C. elegans*.^[247,248] Therefore, it should be possible to use UAG or UGA for the incorporation of unnatural amino acids into proteins in a variety of organisms without adverse effects on the growth of the host organism.

Our initial attempts to generate an orthogonal amber suppressor tRNA/synthetase pair involved modifying the specificity of an existing *E. coli* tRNA/synthetase pair. Key nucleotides at the tRNA/synthetase interface were mutated to reduce the affinity of the tRNA for its cognate synthetase, while maintaining orthogonality to the other endogenous synthetases. A mutant synthetase was then evolved from the

cognate wild-type synthetase that uniquely recognizes the mutant orthogonal tRNA, but at the same time retains its orthogonality with the endogenous tRNAs in *E. coli*. Based on the X-ray crystal structure of the complex of *E. coli* glutamyl-tRNA synthetase (GlnRS) and tRNA^{Gln},^[249] three sites in tRNA^{Gln} which make specific contacts with GlnRS were mutated.^[250,251] These mutant tRNAs were tested individually in in vitro aminoacylation reactions and in vivo amber codon suppression assays, and an efficient orthogonal amber suppressor tRNA was identified. A library of mutant GlnRSs was then generated and mutants were selected based on their ability to aminoacylate the orthogonal tRNA. Although mutant *E. coli* GlnRSs were evolved that charge the orthogonal tRNA, the best mutant aminoacylates the orthogonal tRNA at approximately 10% the rate of wild-type tRNA^{Gln}. An ideal orthogonal synthetase should not acylate any endogenous *E. coli* tRNAs, since even modest misacylation of a wild-type tRNA with an unnatural amino acid will result in misincorporation of the unnatural amino acid at other sites in the protein. The inability to generate an orthogonal glutamyl tRNA/synthetase pair, together with the finding that mutations within the tRNA interact in complicated, non-additive ways,^[250] prompted us to examine alternative strategies.

A second approach to the generation of an orthogonal tRNA/synthetase pair involves the use of a heterologous tRNA/synthetase pair from another organism in *E. coli*. This approach is based on the observation that some prokaryotic tRNA/synthetase pairs do not cross-react to any significant degree with their eukaryotic counterparts, as a result of differences in the tRNA identity elements.^[127,252] For example, it has been shown that *E. coli* GlnRS lacks the N-terminal RNA-binding domain present in *S. cerevisiae* GlnRS and does not aminoacylate yeast tRNA^{Gln}.^[253] We therefore tested and found that the yeast amber suppressor tRNA^{Gln} (*Sc* tRNA^{Gln}_{CUA}) is also not a substrate for *E. coli* GlnRS, yet functions efficiently in translation in bacteria. Moreover, whereas the yeast GlnRS does not aminoacylate any *E. coli* tRNAs, it does aminoacylate the amber suppressor *Sc* tRNA^{Gln}_{CUA} in vitro.^[254] Therefore, the *S. cerevisiae* tRNA^{Gln}_{CUA}/GlnRS is an orthogonal pair in *E. coli*. Unfortunately, no mutant yeast GlnRS could be evolved that aminoacylates the orthogonal tRNA with an unnatural amino acid in *E. coli*, possibly because of the low intrinsic activity of the synthetase in bacteria. Yeast tRNA/synthetase pairs specific for various hydrophobic amino acids were also examined,^[255] and attention was focused on the tRNA^{Tyr}/TyrRS pair, since TyrRS does not have a proofreading function and the prokaryotic and eukaryotic tRNA^{Tyr}s have significant differences (Figure 14). Indeed, in vitro studies have shown that tRNA^{Tyr} of *S. cerevisiae* and *H. sapiens* cannot be aminoacylated by bacterial synthetases, nor do their TyrRS aminoacylate bacterial tRNA.^[256,257] Unfortunately, in vivo amber codon suppression assays showed that the amber suppressor tRNA^{Tyr}_{CUA} derived from either *S. cerevisiae* or *H. sapiens* is not orthogonal in *E. coli*.^[258]

We then focused our attention on archaea as a source of orthogonal tRNA/synthetase pairs for use in *E. coli*. Archaeal aminoacyl-tRNA synthetases are more similar to their

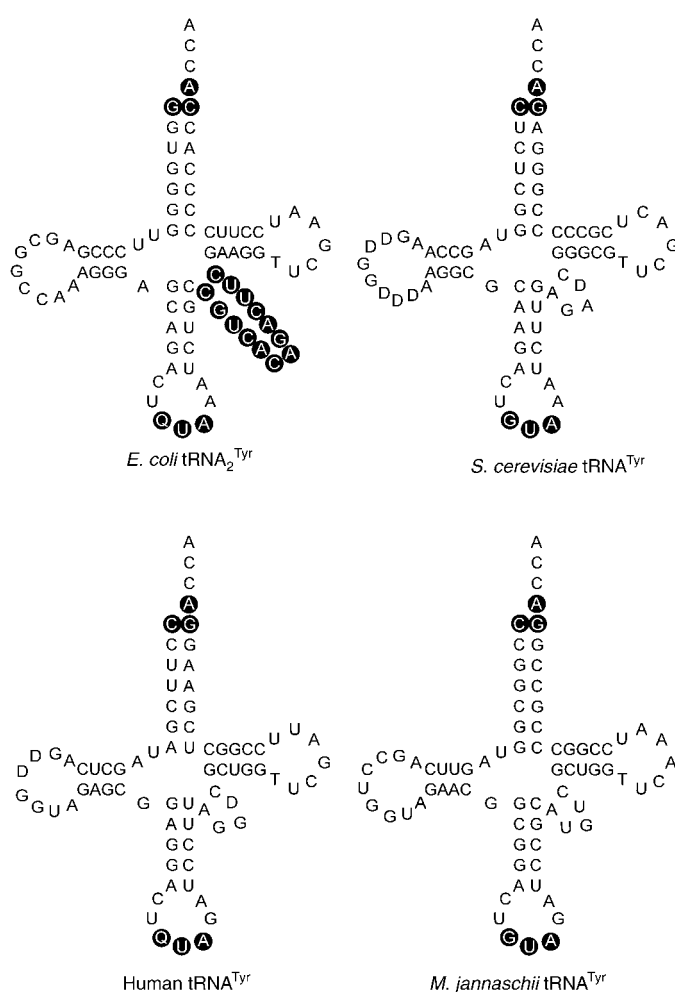


Figure 14. Comparison of tRNA^{Tyr}s from different species. Major identity elements are shaded in black. Prokaryotic tRNA^{Tyr} has a long variable arm and a G1:C72 pair, in contrast to the short arm and C1:G72 pair in eukaryotic and archaeal tRNA^{Tyr}.

eukaryotic than prokaryotic counterparts in terms of homology and tRNA recognition elements.^[115] However, unlike synthetases from eukaryotic cells, which are often expressed poorly or have low activities in *E. coli*, synthetases from archaea can be expressed efficiently in *E. coli* in active form. In addition, the availability of more than 16 archaeal genome sequences, together with the absence of introns in their genomes, facilitates the cloning of archaeal synthetase genes. Finally, it has been shown that almost all the tRNAs of *Halobacterium cutirebrum*, an archaeal halophile, cannot be charged by *E. coli* synthetases.^[127]

The amber suppressor tyrosyl-tRNA/synthetase pair derived from *Methanococcus jannaschii* was the first orthogonal *E. coli* pair generated from archaea.^[258] *M. jannaschii* TyrRS (*Mj*TyrRS) discriminates tRNAs with C1:G72 (an eukaryotic recognition element) from those with G1:C72 (a prokaryotic recognition element), and does not aminoacylate crude tRNA from *E. coli*.^[259] In addition, *Mj*TyrRS is missing most of the C-terminal binding domain for the anticodon loop of its tRNA^{Tyr}, which suggests that identity elements outside

the anticodon region are mainly responsible for recognition by TyrRS and the orthogonality of this pair. Mutation of the anticodon loop of this tRNA should therefore have minimal effect on its recognition by other synthetases and its affinity for *Mj*TyrRS. Indeed, when cells expressing a β -lactamase gene with an amber codon at a permissive site are cotransformed with both a suppressor *M. jannaschii* tRNA^{Tyr}_{CUA} (*MjtRNA*^{Tyr}_{CUA}) and *Mj*TyrRS, they survive in the presence of high concentrations of ampicillin (suppression of the amber mutation results in the production of full-length β -lactamase allowing cells to survive in ampicillin).^[258] Unfortunately, while this result shows that the *MjtRNA*^{Tyr}_{CUA} is an excellent substrate for *Mj*TyrRS in *E. coli*, we did observe a low but significant level of aminoacylation of this tRNA by endogenous synthetases.

A general strategy for the selection of orthogonal tRNAs in *E. coli* was then developed to further reduce recognition of the *MjtRNA*^{Tyr}_{CUA} by endogenous *E. coli* synthetases, while preserving activity with both the cognate synthetase and translational machinery (for example, the ribosomal A site, elongation factor Tu, etc.).^[260] This method consists of a combination of negative and positive selections with a mutant suppressor tRNA library in the absence and presence of the cognate synthetase, respectively (Figure 15). In the negative selection, amber nonsense codons are introduced into the barnase gene at sites permissive to substitution by other amino acids. When a member of the suppressor tRNA library is aminoacylated by an endogenous *E. coli* synthetase (that is, it is not orthogonal to the *E. coli* synthetases), the amber codons are suppressed and the ribonuclease barnase is produced, resulting in cell death. Only cells harboring orthogonal or nonfunctional tRNAs can survive. All surviving clones are then subjected to a positive selection in which an amber codon is placed in the β -lactamase gene at a permissive site. tRNAs that are good substrates for the cognate synthetase are selected based on their ability to suppress the amber codon and produce active β -lactamase. Cells harboring nonfunctional tRNAs, or tRNAs that cannot be recognized by their cognate synthetase, will be sensitive to ampicillin. Therefore, only tRNAs that: 1) are not substrates for

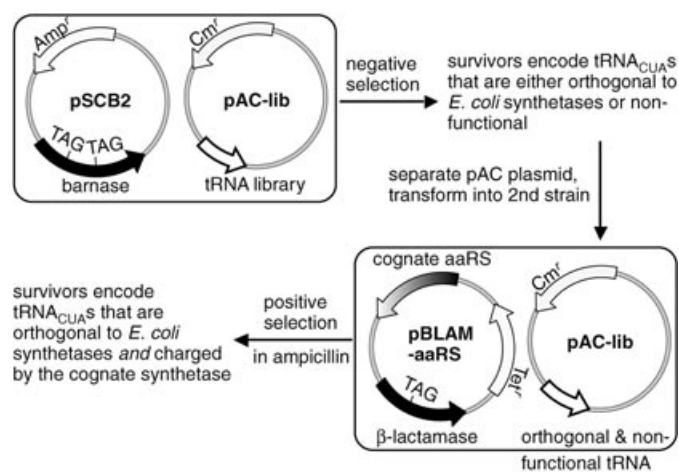
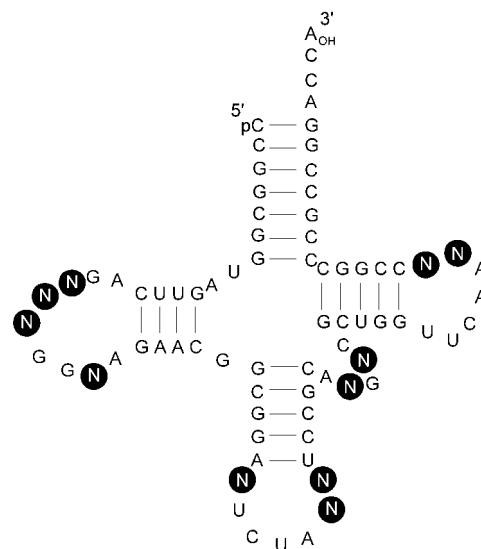


Figure 15. Negative and positive selections used to evolve orthogonal tRNAs.

endogenous *E. coli* synthetases; 2) can be aminoacylated by the synthetase of interest; and 3) function in translation will survive both selections.

Eleven nucleotides of the *MjtRNA*^{Tyr}_{CUA} that do not interact directly with the *Mj*TyrRS were randomly mutated to generate a suppressor tRNA library (Figure 16). This tRNA



| suppressor tRNA | IC ₅₀ (μg mL ⁻¹ of amp) | |
|--------------------------------------|---|-----------------|
| | with pBLAM | with pBLAM-JYRS |
| no tRNA | 10 | 10 |
| w.-t. <i>Mj</i> Tyr | 56 | 1220 |
| mutRNA ^{Tyr} _{CUA} | 12 | 436 |
| mutant 2 | 20 | 632 |
| mutant 3 | 14 | 459 |

Figure 16. A mutant tRNA library based on *M. jannaschii* tRNA^{Tyr} with randomized nucleotides indicated with N. The table lists the suppression efficiencies of several selected mutant tRNAs.

library was passed through rounds of negative and positive selection to afford a mutant tRNA (mutRNA^{Tyr}_{CUA}) that suppresses the β -lactamase amber mutant at a level comparable to the background level in *E. coli*, thus indicating that this mutant tRNA is a poorer substrate for *E. coli* synthetases than the wild-type *MjtRNA*^{Tyr}_{CUA}. When the mutRNA^{Tyr}_{CUA} is coexpressed with *Mj*TyrRS, cells survive at high concentration of ampicillin, which indicates that the tRNA is still aminoacylated by its cognate TyrRS and functions efficiently in translation.^[260] The low background and high activity of this mutRNA^{Tyr}_{CUA}/*Mj*TyrRS pair make it an excellent orthogonal pair in *E. coli*.

Next the substrate specificity of the orthogonal synthetase was altered so that the cognate orthogonal tRNA is charged with only the desired unnatural amino acid and none of the common 20 amino acids. Amino acid promiscuity in the orthogonal synthetase will result in proteins with a mixture of common and unnatural amino acids at the site of mutation. For example, in an attempt to site-specifically incorporate *p*-

fluorophenylalanine (*p*-F-Phe) into proteins, a yeast amber suppressor tRNA^{Phe}_{CUA}/PheRS pair was introduced into an *E. coli* strain auxotrophic for Phe and unable to utilize *p*-F-Phe.^[261] Since yeast PheRS aminoacylates its cognate tRNA with both Phe and *p*-F-Phe, 64–75% *p*-F-Phe is incorporated at the site of mutation; the remainder is Phe and Lys (even in the presence of an excess of *p*-F-Phe added to the growth media). Moreover, 7% *p*-F-Phe is found at sites encoded by the Phe codon, which indicates that the endogenous *E. coli* PheRS incorporates *p*-F-Phe in addition to Phe.

We have developed a systematic and general approach to evolve mutant synthetases that selectively aminoacylate their cognate orthogonal tRNA with unnatural amino acids.^[262] Based on the crystal structure of the homologous TyrRS from *Bacillus stearothermophilus*, five residues (Tyr 32, Asn 123, Asp 158, Ile 159, and Leu 162) in the active site of *M. jannaschii* TyrRS that are within 6.5 Å of the *para* position of the aryl ring of bound tyrosine were randomized to all possible common amino acids (Figure 17). These residues

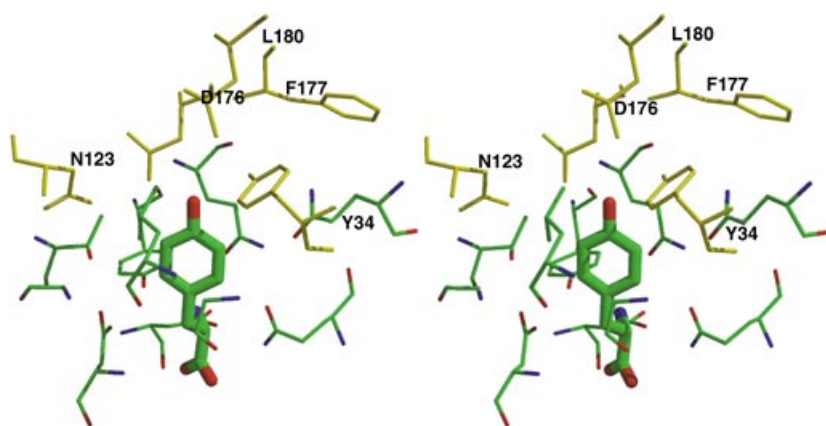


Figure 17. Stereoview of the active site of TyrRS. Residues from *B. stearothermophilus* TyrRS are shown. Corresponding residues from *M. jannaschii* TyrRS are Tyr 32 (Tyr 34), Glu 107 (Asn 123), Asp 158 (Asp 176), Ile 159 (Phe 177), and Leu 162 (Leu 180) with residues from *B. stearothermophilus* TyrRS in parentheses; mutated residues are in yellow.^[262]

were all initially mutated to alanine to avoid a high background of wild-type TyrRS. The resulting inactive Ala₅ TyrRS was used as a template for PCR mutagenesis to generate the TyrRS library by mutation of the five sites to NNK (N = A, T, G, or C and K = G or T). The size of this library was 1.6×10^9 to ensure complete coverage of all possible mutants. A positive selection was then applied which is based on suppression of an amber stop codon in the chloramphenicol acetyltransferase (CAT) gene in the presence of the unnatural amino acid and chloramphenicol (in contrast to β -lactamase which is secreted into the periplasm, CAT localizes in the cytoplasm). Surviving cells were then grown in the presence of chloramphenicol and in the absence of the unnatural amino acid. Those cells that did not survive were isolated from a replica plate supplemented with the unnatural amino acid. The mutant synthetase genes were isolated from these cells, recombined in vitro by DNA shuffling, and transformed back into *E. coli* for further rounds of selection with increasing concentrations of chloramphenicol (Figure 18).

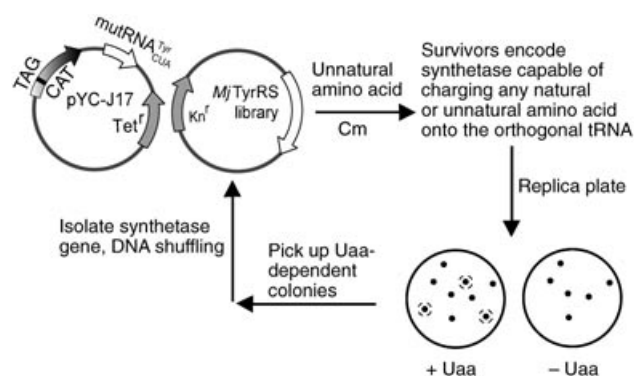


Figure 18. A general positive selection and negative screen scheme for evolving synthetases specific for unnatural amino acids. Cm = chloramphenicol, Uaa = unnatural amino acid.

This selection scheme was first used to generate an *Mj*TyrRS mutant synthetase capable of selectively inserting *O*-methyl-L-tyrosine (OMeTyr) into proteins in response to the amber codon.^[262] OMeTyr is structurally similar to tyrosine and phenylalanine and provided an excellent opportunity to test the degree of translational fidelity that can be achieved by this approach. After two rounds of selection and DNA shuffling, a clone was evolved whose survival in chloramphenicol was highly dependent on the addition of 1 mM OMeTyr to the growth media. Cells survived in the presence of chloramphenicol when OMeTyr was added to the growth media, but died in the absence of the amino acid. An OMeTyr mutant of *E. coli* dihydrofolate reductase (DHFR) was generated and characterized to demonstrate that the observed phenotype arises from the site-specific incorporation of OMeTyr by the mutRNA^{Tyr}_{CUA}/mutant TyrRS pair in response to an amber stop codon. The

third codon of the *E. coli* DHFR gene was mutated to TAG, and a His₆ tag was added to the C-terminus to facilitate protein purification and separation from endogenous DHFR. Full-length DHFR was produced only when the mutant TyrRS, the mutRNA^{Tyr}_{CUA}, and OMeTyr were all present (Figure 19). Insertion of OMeTyr in response to the TAG codon was confirmed by mass analysis of both the intact protein and tryptic fragments. No incorporation of tyrosine or other amino acids at the TAG position was observed, and OMeTyr was incorporated only in response to TAG and not any other sites. The growth rate of *E. coli* was not significantly affected and the protein yield was 2 mg L^{-1} in minimal media, which corresponds to 77% of the yield of wild-type DHFR. These results demonstrated for the first time that one could rationally engineer bacteria that genetically encode an unnatural amino acid with high efficiency and translational fidelity.^[262]

We then attempted to genetically encode L-3-(naphthyl)-alanine, which represents a significant structural perturbation

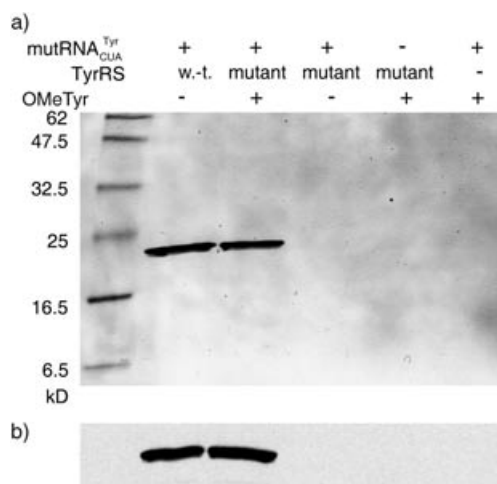


Figure 19. Accumulation of *E. coli* DHFR protein under different conditions. a) Silver-stained sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) of purified DHFR. The left lane is a molecular weight marker. b) Western blot of the protein gel. A penta-His antibody was used to detect the six-histidine tag at the COOH terminus of DHFR.

relative to tyrosine.^[263] A slightly different mutant TyrRS library (in which Tyr32, Asp158, Ile159, Leu162, and Ala167 were randomized) was constructed from which four mutant synthetases were identified that have high activity for this unnatural amino acid. One round of DNA shuffling of these four synthetase genes resulted in a mutant synthetase (SS12) with enhanced activity for L-3-(naphthyl)alanine and greatly reduced activity for endogenous amino acids. An amber codon was introduced in the middle of the gene of mouse DHFR to verify incorporation of the unnatural amino acid. Protein expression and mass-spectrometric analysis confirmed that the mutRNA^{Tyr}_{CUA}/SS12 mutant synthetase selectively incorporates L-3-(naphthyl)alanine into proteins with an efficiency and fidelity rivaling that of the common 20

amino acids.^[263] This result suggested that the above methodology may be applicable to a large variety of unnatural amino acids.

Alternative selection schemes for evolving the amino acid specificity of the orthogonal *Mj*TyrRS have been developed subsequently. One approach involves passing the mutant synthetase library through a positive selection based on the ability of the mutant synthetase to suppress amber mutations introduced at permissive sites in the CAT gene in the presence of an unnatural amino acid as described above. A negative selection is then applied in which those synthetases that recognize endogenous amino acids are eliminated in the absence of the unnatural amino acid. The negative selection is based on suppression of amber nonsense mutations introduced in the barnase gene.^[254] Pilot selections of inactive *Mj*TyrRS and OMeTyr specific mutant synthetases from a large pool of wild-type *Mj*TyrRS were used to optimize the selection conditions for the *M. jannaschii* mutRNA^{Tyr}_{CUA}/TyrRS pair.^[264] A barnase mutant with three amber codons (encoded in pLWJ17B3) was found to provide optimal selection stringency and enrichment efficiency.^[265] Survivors of both the positive and negative selections encode synthetases that charge the orthogonal tRNA with only the unnatural amino acid. More mutations can be introduced by mutagenesis or DNA shuffling into these synthetase genes to generate second-generation synthetase libraries, which can be used for further rounds of selection until a mutant synthetase with a desired activity is evolved. This new scheme (Figure 20) has been used to evolve *M. jannaschii* mutRNA^{Tyr}_{CUA}/TyrRS orthogonal pairs for a wide variety of unnatural amino acids (see Section 5.2). By varying the stringency, order, and number of the positive and negative selection steps one can likely affect both the activity and fidelity of the evolved synthetase.^[265]

A second method for evolving aminoacyl-tRNA synthetase specificity relies on an amplifiable fluorescent reporter of amber suppression.^[266] A T7/GFPuv reporter (pREP) was created in which suppression of amber codons introduced at

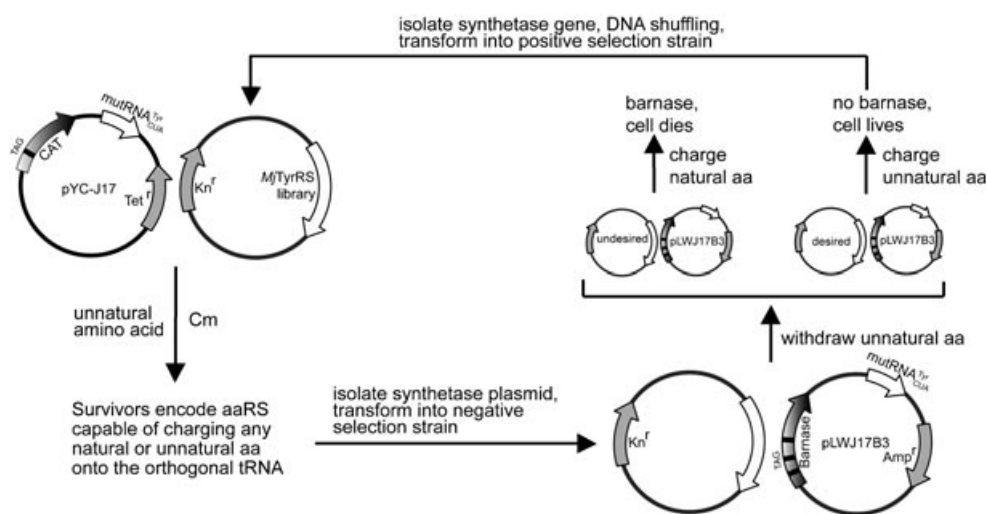


Figure 20. A general positive and negative selection scheme for evolving synthetases specific for unnatural amino acids.

permissive sites in the gene for T7 RNA polymerase produces full-length protein, which in turn drives the expression of green fluorescent protein (GFPuv). This construct was combined with the above CAT reporter (pYC-J17) to yield pREP(2)/YC-JYCUA which contains the orthogonal $\text{mutRNA}_{\text{CUA}}^{\text{Tyr}}$, the amber mutant of CAT, and the T7/GFPuv reporter genes (Figure 21 a).^[266] When *E. coli* containing this reporter are transformed with a mutant *MjTyrRS* library and grown in the presence of an unnatural amino acid, active synthetases can be identified based on fluorescence and/or chloramphenicol resistance. Cells are subsequently grown in the absence of unnatural amino acid and chloramphenicol, and nonfluorescent cells containing the desired mutant synthetase are collected by fluorescence-activated cell sorting (FACS). The advantages of this method are that a single strain of *E. coli* can be used to perform both positive and negative selections, and cells containing the reporter can be replicated in the presence and absence of amino acid and visually inspected for amino acid dependent fluorescence (Figure 21 b).^[266] Alternatively, this chloramphenicol and FACS

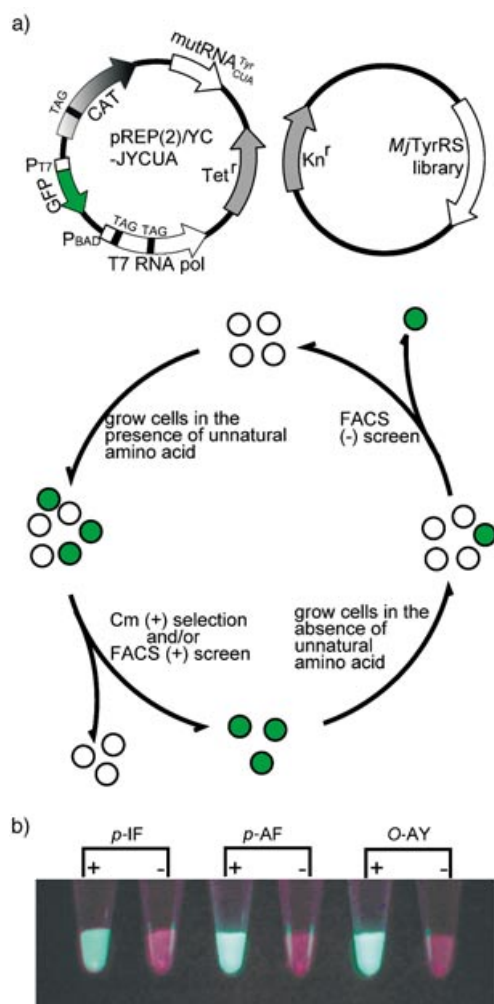


Figure 21. a) A system for evolving synthetase specificity that combines a fluorescent reporter with a chloramphenicol-based selection. b) Green fluorescence can be used to verify unnatural amino acid dependent GFP expression.

double-positive selection can also be followed by the barnase negative selection.^[265]

A direct positive selection has also been developed in which an antibody specific for an unnatural amino acid (or peptide containing the amino acid) is used to enrich phage into which the unnatural amino acid is incorporated (Figure 22).^[267] A C3 peptide with an amber mutation is

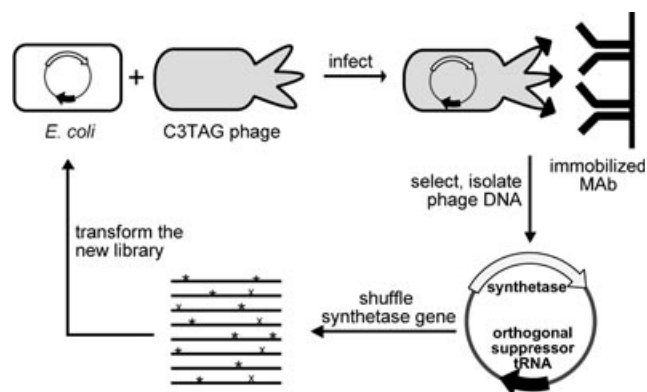


Figure 22. A phage-based selection for the incorporation of unnatural amino acids. *E. coli* are infected by phage with a stop codon in a gene encoding a phage coat protein. Phage containing an active synthetase display the unnatural amino acid on the phage surface and are selected with immobilized monoclonal antibodies.

fused to the N-terminus of the VSCM13 phage coat protein pIII, such that C3TAG phage production requires suppression of the amber stop codon. Cells harboring a phagemid that expresses an orthogonal suppressor tRNA and a synthetase library are infected with the C3TAG phage. An active synthetase results in suppression of C3TAG and display of its cognate amino acid on the phage surface. The phage pool is then incubated with immobilized monoclonal antibodies directed against the unnatural amino acid to isolate only those phage carrying the synthetase specific for the unnatural amino acid. Phage displaying Asp were enriched over 300-fold from a pool of phage displaying Asn in a simulated selection using antibodies raised against the Asp-containing epitope.

5.2. A Survey of Genetically Encoded Unnatural Amino Acids

The above selections have been used to add over 30 novel amino acids to the genetic code of *E. coli* (Figure 23).^[262,263,266,268–275] These include: 1) various substituted tyrosine and phenylalanine analogues such as *O*-methyl-L-tyrosine (**1**),^[262] *p*-amino-L-phenylalanine (**5**),^[266] 3-nitro-L-tyrosine (**15**),^[276] *p*-nitro-L-phenylalanine (**29**),^[276] *m*-methoxy-L-phenylalanine,^[276] and *p*-isopropyl-L-phenylalanine (**4**);^[266] 2) amino acids with aryl azide (**10**) and benzophenone (**11**) groups that can be photo-cross-linked;^[269,270] 3) amino acids that have unique chemical reactivity including *p*-acetyl-L-phenylalanine (**6**) and *m*-acetyl-L-phenylalanine (**9**),^[271,273] *O*-allyl-L-tyrosine (**3**),^[268] *O*-(2-propynyl)-L-tyrosine

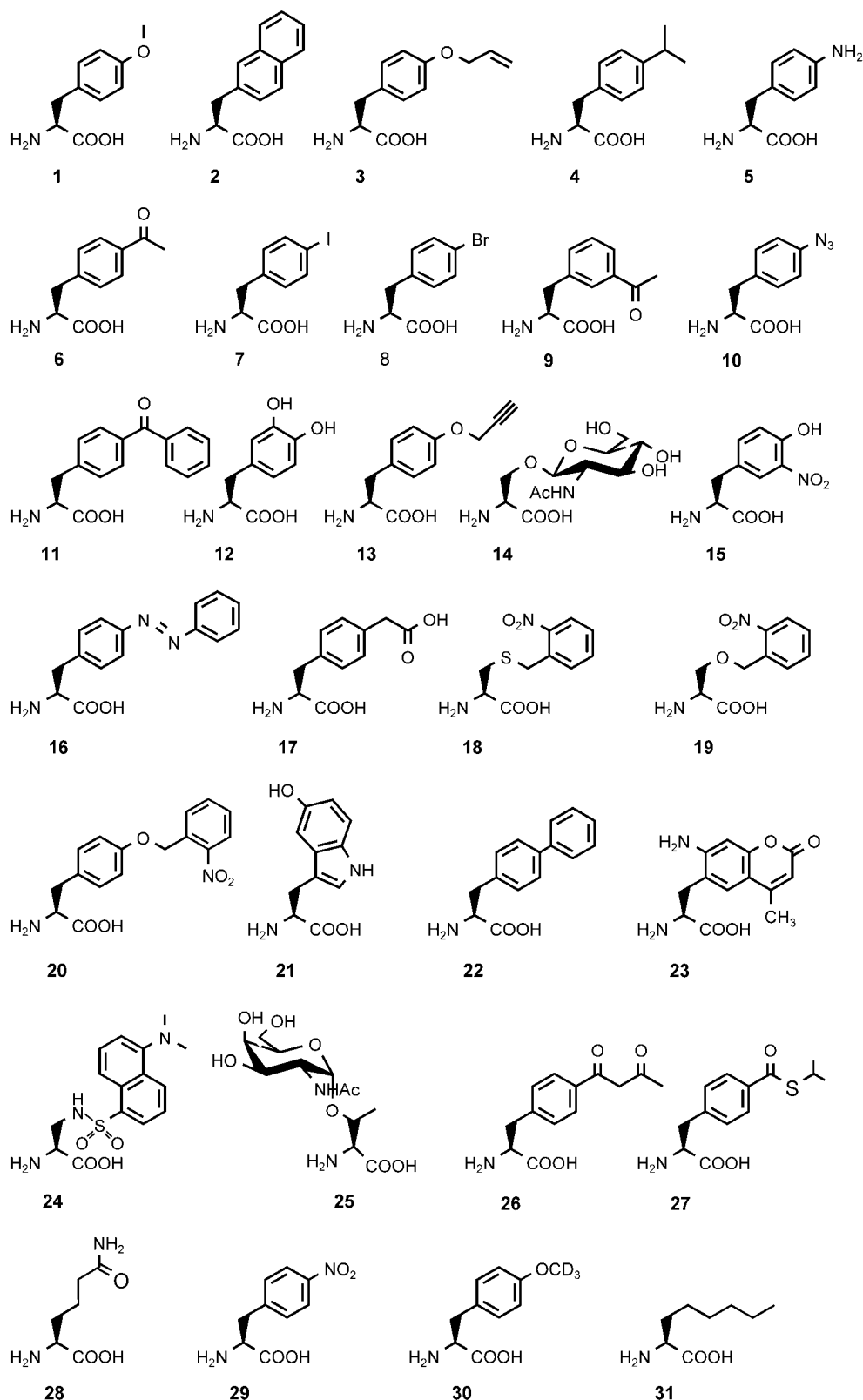


Figure 23. Some structures of unnatural amino acids that have been added to the genetic code of *E. coli*.

(13),^[276] *p*-ethylthiocarbonyl-L-phenylalanine (27), and *p*-(3-oxobutanoyl)-L-phenylalanine (26);^[276] 4) heavy-atom-containing amino acids for phasing in X-ray crystallography

including *p*-iodo (7) and *p*-bromo-L-phenylalanine (8);^[272] 5) the redox-active amino acid dihydroxy-L-phenylalanine (12);^[274] 6) glycosylated amino acids including β -*N*-acetylglu-

cosamine-*O*-serine (**14**) and α -*N*-acetylgalactosamine-*O*-threonine (**25**);^[275] 7) fluorescent amino acids with naphthyl (**2**), dansyl (**24**), and 7-aminocoumarin (**23**) side chains;^[263,276] 8) photocleavable and photoisomerizable amino acids with azobenzene (**16**) and nitrobenzyl Cys, Ser, and Tyr side chains (**18–20**);^[276] 9) the phosphotyrosine mimetic *p*-carboxymethyl-L-phenylalanine (**17**); 10) the glutamine homologue homoglutamine (**28**);^[276] and 11) 2-aminooctanoic acid (**31**). This list suggests that synthetases can be evolved for a large number of structurally diverse amino acid side chains, which range from sugars to extended aromatic groups. Moreover, it appears that the ribosome and elongation factors also tolerate considerable structural diversity, which is consistent with previous *in vitro* mutagenesis studies.

In general these unnatural amino acids are incorporated with excellent fidelity and efficiency by their cognate tRNA/synthetase pairs. Each pair was characterized by their ability to confer chloramphenicol resistance to cells transformed with the CATamber mutant gene in the presence and absence of a particular unnatural amino acid.^[262] Secondly, each orthogonal tRNA/synthetase pair was used to incorporate the corresponding unnatural amino acid into proteins at permissive sites in the genes for *E. coli* or mouse DHFR,^[262,263] the Z domain of protein A,^[268,271] or myoglobin.^[266] Each protein was tagged with hexahistidine and purified. Milligram quantities of the mutant protein could be purified per liter of minimal media by using the unnatural amino acid and the orthogonal tRNA/synthetase pair. Yields were typically 25–50%, and in some cases greater than 75% that obtained with the wild-type protein.^[262,263] High-resolution mass spectrometry was also used to analyze the intact protein and tryptic digests^[271] to confirm that only the unnatural amino acid is incorporated in response to the TAG codon and at no other site in the protein.

Many of these unnatural amino acids have novel properties that are useful for a variety of biochemical and cellular studies. For example, two amino acids (*p*-azidophenylalanine (**10**) and *p*-benzoylphenylalanine (**11**)) have side chains that can be photo-cross-linked both *in vitro* and *in vivo*.^[269,270,277] Benzophenone is a particularly useful photo-cross-linker since it absorbs at relatively long wavelengths (ca. 360 nm), and the excited state inserts efficiently into carbon–hydrogen bonds.^[278] Indeed, a mutant homodimeric glutathione S-transferase with *p*-benzoylphenylalanine substituted at the dimer interface could be crosslinked in greater than 50% yield in the cytoplasm of *E. coli* (Figure 24). These amino acids should be useful for studies of protein–protein and protein–nucleic acid interactions *in vitro* and *in vivo*, for studies of protein structure and dynamics, and for the identification of receptors for orphan ligands. They should also allow the selective immobilization of proteins and cells on surfaces in defined orientations.

The keto-containing amino acids *p*-acetylphenylalanine (**6**; *p*-Ac-Phe) and *m*-acetylphenylalanine (**9**) can be used for the selective chemical modification of proteins, especially those proteins with multiple cysteine residues.^[271,273] The ketone group undergoes a variety of reactions including addition, aldol, transamination, and isomerization reactions. It reacts with hydrazides, alkoxyamines, and semicarbazides

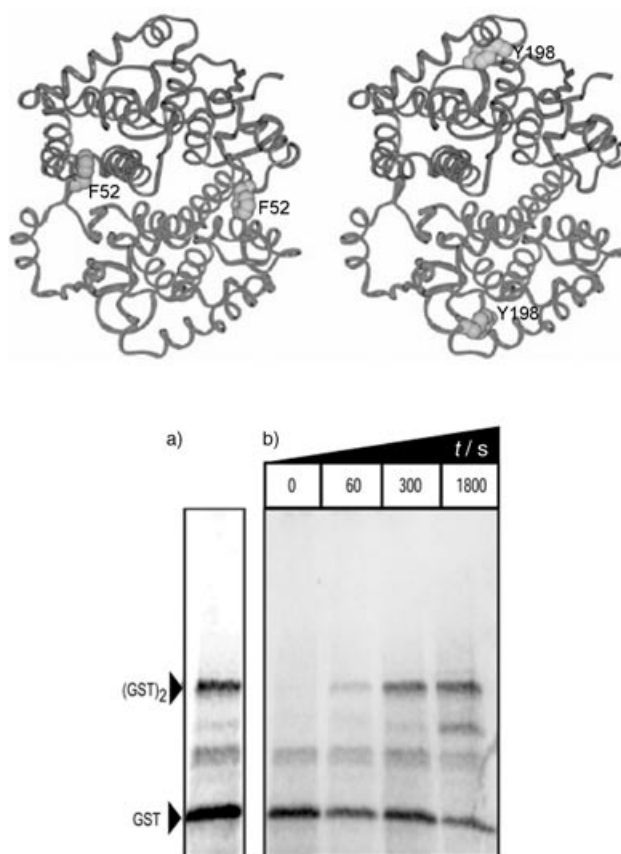


Figure 24. Residues Phe 52 (top left) and Tyr 198 (top right) of GST were mutated to *p*-benzoylphenylalanine (pBpa). a) Purified GST (Phe52pBpa) was cross-linked *in vitro* and detected by Western blot with an anti-GST antibody. b) The covalent dimerization of GST (Phe52-pBpa) in *E. coli* as a function of cellular irradiation at 365 nm, detected as in (a). Tyr 198-pBpa GST could not be cross-linked.

under aqueous, mild conditions to produce hydrazone, oxime, and semicarbazone linkages that are stable under physiological conditions. For example, the *in vitro* reaction of a *p*-Ac-Phe mutant of the Z domain protein with 1 mM fluorescein hydrazide at pH 6.5 led to labeling of the protein with fluorescein in a 1:1 molar ratio in greater than 90% yield (Figure 25a).^[271] Homogenous glycoprotein mimetics were also generated by labeling *p*-Ac-Phe-containing proteins with aminoxy sugars in >90% yield (Figure 25b). Oligosaccharides with defined structures were covalently attached to proteins either by adding additional saccharides to the pendant sugar with glycosyltransferases, or by directly coupling an aminoxy-derivatized glycan in one step.^[279] Similarly, keto-containing mutants were efficiently labeled with hydrazide derivatives of biotin (Figure 25c) and polyethylene glycol. The latter should allow the generation of homogeneous proteins site-specifically derivatized with PEG with very high chemical purity. In addition, *m*-acetylphenylalanine has been used to efficiently label *in vivo* a keto-containing mutant of the *E. coli* membrane protein LamB with any of three water soluble fluorophores (Cascade blue, Alexa 568, or Alexa 647).^[273] In a related experiment, copper-

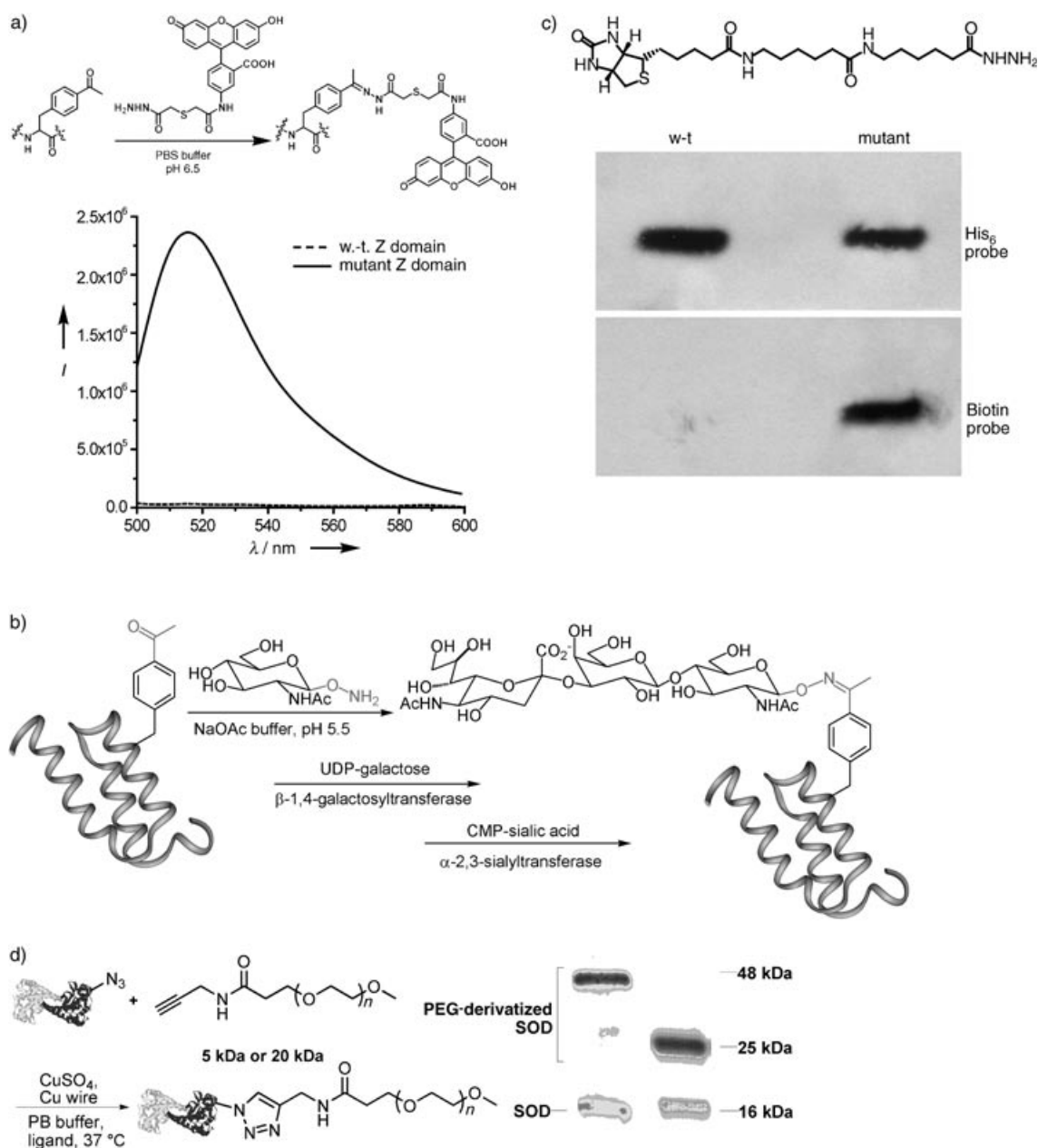


Figure 25. a) In vitro labeling of protein containing *p*-acetyl-L-phenylalanine with fluorescein hydrazide. b) Generation of homogeneous glycoprotein mimetics by conjugating an aminoxy sugar and subsequent extension with glycosyltransferases. c) In vitro labeling of protein containing *p*-acetyl-L-phenylalanine with biotin hydrazide. A His₆ tag was added to the C-terminus of both wild-type and mutant protein and was detected with His₆-specific antibody in Western blot analysis. Only the *p*-acetylphenylalanine-containing protein was labeled with biotin and thus detected by the biotin-specific antibody. d) Protein derivatization with PEG through a copper-mediated cycloaddition reaction and SDS PAGE analysis.

mediated [2+3] cycloaddition reactions between mutant proteins containing either alkyne or azido unnatural amino acids **10** or **13** and a corresponding azide- or alkyne-derivatized chemical label have been used to selectively modify proteins with fluorophores and polyethylene glycol in high yield (> 75 %, Figure 25 d).^[280] In addition, we have recently shown that thioester-containing amino acids can also be used to selectively modify proteins with nonpeptidic groups.^[281]

Protein glycosylation is an essential posttranslational modification in eukaryotic cells, which can modulate protein

folding and secretion, biological activity, serum half-life, and localization.^[282] Although a variety of methods have been developed to obtain pure glycoforms of a protein,^[283–286] most impose severe restrictions on the size, quantity, and/or quality of the glycoprotein produced. In an effort to provide a general solution to this problem, we attempted to alter the specificity of the *Mj*TyrRS to accept β-*N*-acetylglucosamine-*O*-serine (β-GlcNAc-Ser, **14**). The esterified amino acid was used to facilitate transport into the cell, where it is subsequently deacetylated by cytosolic esterases. Mutant synthetases were isolated that site-specifically incorporate the glycosylated

amino acid into myoglobin in response to the amber codon with high translational fidelity and in a yield roughly 20–30% that of wild-type myoglobin. An enzyme-linked immunosorbent assay demonstrated significant binding of the GlcNAc-specific lectin *Bandeiraea simplicifolia* II (BSII) to the β -GlcNAc-Ser mutant myoglobin, whereas no binding was observed for the wild-type myoglobin.^[275] Moreover, the *O*-GlcNAc-serine mutant myoglobin could act as a substrate for galactosyltransferases. Incubation of the myoglobin mutant containing the GlcNAc side chain with radioactive UDP[6-³H]Gal in the presence of β -1,4-galactosyltransferase resulted in radiolabelled product (Figure 26).^[275] This approach has also been used to selectively introduce α -GalNAc-Thr into

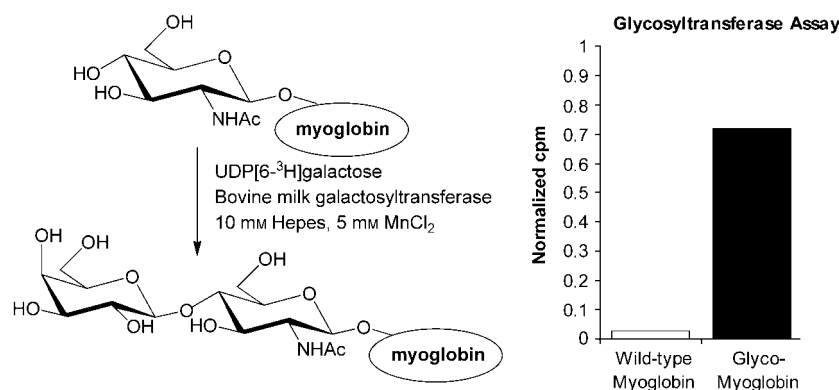


Figure 26. Direct incorporation of β -GlcNAc-Ser into myoglobin in *E. coli* and extension by addition of galactosyltransferase and UDP[6-³H]galactose.

proteins, and is currently being extended to a number of other *O*- and *N*-linked sugars.^[275] We are also attempting to genetically encode other posttranslationally modified amino acids including methylated and acetylated lysine, as well as phosphorylated tyrosine, serine, and threonine. Recently, an *Mj*TyrRS mutant was evolved that selectively incorporates *p*-carboxymethyl-L-phenylalanine (**17**), which is a stable mimetic of phosphotyrosine that can be used for studies of signal transduction. In addition, the incorporation of 2-amino-octanoic acid **31** suggests that it may be possible to incorporate lipid-modified amino acids directly into proteins.

The ability to genetically encode heavy-atom-containing amino acids should facilitate phase determination in X-ray crystallography. To this end, an orthogonal tRNA/synthetase pair has been evolved that selectively incorporates *p*-iodophenylalanine (**7**, *p*-iodo-Phe) in response to the amber codon.^[265] The yield of a mutant *p*-iodo-Phe-containing Z domain protein was around 60% of the wild-type protein (5.5 mg L⁻¹ versus 9.2 mg L⁻¹ in minimal medium). Bacteriophage T4 lysozyme with either residue Ser44 or Phe153 mutated to *p*-iodo-Phe was expressed, purified, and crystallized to determine the utility of *p*-iodophenylalanine for single-wavelength anomalous dispersion (SAD) phasing (the anomalous signal ($\Delta f''$) of iodine at 8 keV is 6.91, six times that of selenium (1.15)). The crystal structure was solved at 1.54-Å resolution by SAD phasing with the incorporated iodine atom using a laboratory Cu_{K α} X-ray source.^[287]

Green fluorescent protein (GFP) is a powerful biological tool, which allows protein expression, localization, dynamics, and interactions to be directly visualized in intact cells.^[288] However, in many cases the size of GFP and the fact that GFP can only be incorporated as an *N*- or *C*-terminal fusion are significant limitations. GFP also requires time for fluorophore maturation. The ability to co-translationally incorporate unnatural amino acids containing relatively small fluorescent side chains would minimize structural perturbations to the host protein, provide a broader range of spectral properties, and allow more control over the location of the fluorescent reporter (namely, at any permissive surface residue, rather than just the *N*- or *C*-terminus). To this end, we have evolved *M. jannaschii* TyrRSs that selectively incorporate 7-aminocoumarin (**23**) and 7-hydroxycoumarin into proteins in *E. coli*,^[289] as well as an orthogonal *E. coli* leucyl synthetase that incorporates the dansyl-containing amino acid **24** (Figure 23) into proteins in yeast.^[290] The environmentally sensitive coumarin-containing amino acid **23** has been used to monitor the unfolding of apo myoglobin and interaction of GroEL with its substrates.

Another novel amino acid that has been incorporated into proteins is dihydroxyphenylalanine (DHP, **12**).^[274] DHP can undergo two-electron oxidation to the corresponding quinone (Figure 27), and should be useful to both probe and manipulate electron-transfer processes in proteins. The redox-active amino acid 5-hydroxytryptophan has also been selectively incorporated into proteins in mammalian cells in response to an opal codon (see below). This amino acid undergoes electrochemical oxidation to form an efficient cross-linking agent in vitro. Unnatural amino acids may also be used to perturb the electronic properties of a protein. For example, introduction of several unnatural amino acids in place of Tyr65 in GFP led to blue shifts in the emission spectra and altered quantum yields.^[272] In addition, deuterated and ¹⁵N-labeled derivatives of OMeTyr have also been site-selectively incorporated into proteins as NMR probes.^[327]

Finally, the azobenzene-containing amino acid **16**^[291] and the photocaged amino acids nitrobenzyl-Cys, -Ser, and -Tyr **18–20**^[292] have recently been incorporated into proteins in *E. coli* and/or yeast. These amino acids make possible the in vivo regulation of protein activity by light. For example, mutation of the active site cysteine residue in the proapoptotic protease caspase 3 to nitrobenzylcysteine led to a catalytically inactive enzyme, whose activity could be restored

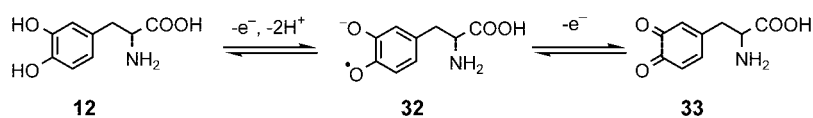


Figure 27. Oxidation of DHP **12** to DHP-semiquinone radical **32**, which is readily oxidized to DHP-quinone **33**.

by photocleavage of the nitrobenzyl protecting group (Figure 28). The azobenzene-containing amino acid can be photochemically converted from the *trans* into the *cis* isomer and therefore provides an opportunity to photoregulate a variety of biological processes *in vivo*. For example, this amino acid has been used to photoregulate the binding of cAMP to the catabolite activator protein of *E. coli*.^[291]

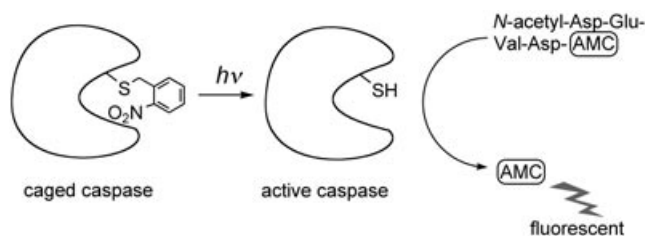


Figure 28. A photocaged caspase with an active-site nitrobenzylcysteine. AMC = 7-amino-4-methylcoumarin.

We have been surprisingly successful in identifying *M. jannaschii* TyrRS mutants with altered specificities, the majority of which are from a library with five active site residues randomized.^[239,262] Crystallographic studies of the wild-type and mutant synthetases provide an opportunity to examine the origins of amino acid specificity in this important class of enzymes. A comparison of the active-site structures of bound and unbound wild-type *Mj*TyrRS indicates that there are significant structural changes that occur upon binding of tyrosine (Figure 29).^[293,294] Asp158 moves closer to the active site to bind the tyrosine hydroxy group, while Tyr32 moves away to accommodate the hydroxy group of the tyrosine residue of the substrate. In addition, the Tyr151 side chain rotates 60° to form a 2.9-Å hydrogen bond with the α -amino group of tyrosine. Recent determination of the X-ray crystal structure of a *p*-bromophenylalanine-specific synthetase suggests that the structural plasticity evident in the wild-type synthetase plays a major role in the generation of the specificity of this mutant enzyme.^[276] This mutant synthetase differs from the wild-type synthetase by five mutations (Y32L, E107S, D158P, I159L, and L162E). Surprisingly, helix α 8 of the mutant is disrupted and residues 158, 159, and 162, which line the active site in the wild-type synthetase,

are flipped out of the active site in the mutant. Residues H160 and Y161, which are exposed to solvent in the wild-type structure, flip in to form a new binding pocket with extensive van der Waals contacts to *p*-bromophenylalanine. We are constructing new libraries to randomize this entire region (residues 158–162) to take advantage of the observed flexibility in this region to further evolve the activity of this synthetase.

Given the above successes and the availability of additional orthogonal tRNA/synthetase pairs, it is likely that many other unnatural amino acids can be genetically encoded. We are currently attempting to evolve orthogonal tRNA synthetase pairs that selectively incorporate amino acids containing spin labels, *p*-cyanophenylalanine (an IR probe), metal-binding amino acids, and backbone-modified amino acids (for example, *N*-methyl amino acids, β -amino acids, and α -hydroxy acids). In addition, a number of approaches are being explored to further increase the yields of the mutant proteins. These include optimizing the expression levels (and induction protocols) of the orthogonal tRNA and synthetase as well as carrying out selections for synthetases in rich versus minimal media. Indeed, it has been shown by Cho and co-workers that such changes can afford up to 750 mg L⁻¹ quantities of mutant proteins.^[328] Synthetase and tRNA mutants with increased activities can also be generated by additional rounds of mutagenesis and selection, or by rational design using X-ray crystal structures of mutant synthetase–amino acid complexes. Indeed, it has recently been shown that the affinities of aminoacylated tRNAs for elongation factor Tu (EF-Tu) and the ribosomal A site vary, depending on the attached amino acid and tRNA.^[295] We are also beginning to examine the role of endogenous factors in determining the efficiency and fidelity of unnatural amino acid incorporation by using transposon and chemical mutagenesis, and complementation with genomic DNA libraries. For example, transposon mutations in the *fts* gene have been identified that improve the efficiency of amber suppression in some cases. Similar approaches may also increase the transport of some unnatural amino acids and lead to further improvements in suppression efficiency by random mutations in the ribosome, release factors, tRNA genes, etc.

The availability of bacteria with additional genetically encoded amino acids allows one to experimentally test

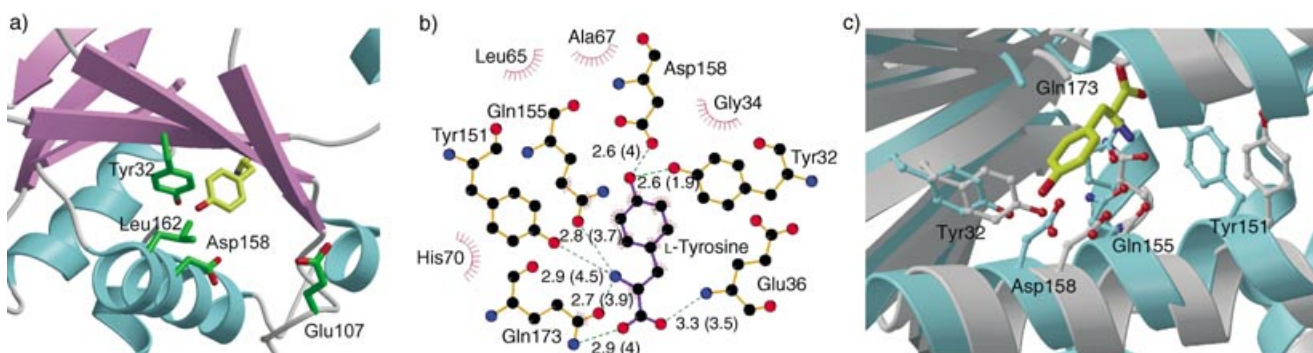


Figure 29. a) The amino acid recognition domain of wild-type *M. jannaschii* TyrRS.^[294] b) Hydrogen-bonding interactions between the substrate tyrosine residue and the interacting residues in the binding pocket. The values correspond to the lengths [Å] of the hydrogen bonds in the TyrRS–Tyr-tRNA^{Tyr} complex structure;^[294] values in parenthesis are analogous distances from the apo structure.^[293] c) Comparison of the *M. jannaschii* TyrRS with (cyan) and without (gray) tyrosine bound in the active site.^[293]

whether additional amino acids can lead to proteins with enhanced chemical or biological properties, or provide an evolutionary advantage for an organism.^[265] A library of unnatural amino acids can be introduced at all sites in a target protein by using an amber codon walking strategy. Alternatively, mutagenesis can be used to introduce amber codons at random sites in the genome, and the resulting strains grown on media containing each of the above unnatural amino acids. These libraries can then be subjected to a variety of selective pressures, and mutations that provide a growth advantage can be mapped and the resulting protein characterized.

In a related experiment we have shown that unnatural amino acids can be efficiently incorporated into polypeptides fused to the pIII protein of a filamentous phage to generate an infectious phage.^[296] This process allows large libraries of phage-displayed peptides and proteins to be created with both natural and unnatural amino acid building blocks, which can subsequently be screened for novel binding and/or catalytic properties.

5.3. Amino Acid Biosynthesis.

In all the above studies the amino acid must be added to the growth media and taken up by the cell in order to be incorporated into a protein in response to the amber codon. There are a large number of amino acid and amine transporters that are relatively nonspecific, and as a consequence, most amino acids we have assayed are transported into *E. coli* cytoplasm from the growth media and accumulate in significant concentrations. In some cases it is necessary to use enzymatically labile derivatives of highly polar amino acids or to mutate/delete enzymes that degrade the amino acid.

An alternative to adding exogenous amino acids to the growth media involves engineering a pathway for the biosynthesis of the unnatural amino acid directly in the host organism. To this end a completely autonomous 21 amino acid bacterium has been generated that contains genes for the biosynthesis of *p*-amino-L-phenylalanine (*p*-AF, **5**) from simple carbon sources,^[297] an aminoacyl-tRNA synthetase^[266] that uses *p*-AF (and no other endogenous amino acids), and a tRNA^[262] that delivers *p*-AF into proteins in response to the amber codon. *p*-AF was biosynthesized from chorismic acid (an intermediate in aromatic amino acid biosynthesis) using

the *papA*, *papB*, and *papC* genes from *S. venezuelae* in combination with a nonspecific *E. coli* transaminase (Figure 30). *E. coli* containing *papA–C* produced *p*-AF at levels comparable to those of the other aromatic amino acids and had normal growth rates. In the presence of a *p*-AF-specific, orthogonal mutRNA^{Tyr}_{CUA}/synthetase pair, *E. coli* transformed with *papA–C* produced mutant proteins containing *p*-AF at sites encoded by the amber codon with excellent yields and fidelity.^[297] In addition to *p*-AF, it should be possible to biosynthesize and genetically encode other amino acids in vivo as well, including methylated, acetylated, and glycosylated amino acids.

5.4. Quadruplet Codons

The ultimate number of unnatural amino acids that can be encoded in any organism is limited by the number of noncoding codons. In addition to nonsense codons, it should be possible to use rare codons to encode unnatural amino acids, although it is likely that some competition with the cognate amino acid will occur. Alternatively, it may be possible to generate a synthetic *E. coli* variant in which rare codons are deleted from the wild-type genome (by replacing them with degenerate codons specifying the same amino acid) and instead used to encode unnatural amino acids.

It should also be possible to use quadruplet codons and cognate suppressor tRNAs with expanded anticodon loops to encode additional amino acids. There are many examples of naturally occurring +1 frameshift suppressors including UAGN (N = A, G, C, or T) suppressors derived from Su7 which encodes glutamine,^[298] *sufJ*-derived suppressors of ACCN codons encoding threonine,^[299] and CAAA suppressors derived from tRNA^{Lys} and tRNA^{Gln}.^[300] Moreover, genetic selections have been used to identify efficient four- and five-base codon suppressor tRNAs from large libraries of mutant tRNAs.^[301,302] In these studies, a tRNA library in which the anticodon loop of tRNA^{Ser} (positions 31–38) was replaced with all possible eight base-pairing sequences was crossed with a library containing all possible four-base codon sequences encoding Ser70 of β -lactamase. Efficient pairs of suppressor and four-base codon were identified by an in vivo selection strategy,^[301] including a mutant tRNA with an anticodon loop sequence of CUUCCUAG that efficiently

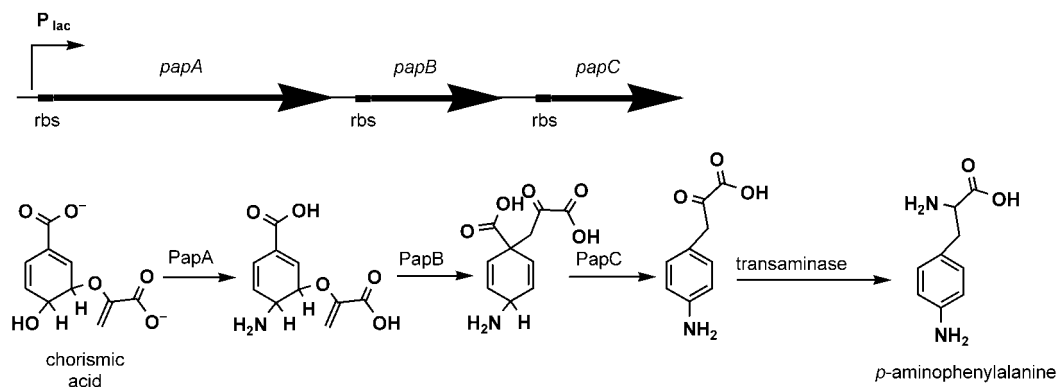


Figure 30. Biosynthesis of *p*-aminophenylalanine in *E. coli* by introducing *papA*, *B*, and *C* genes from *S. venezuelae*.

suppresses an AGGA codon.^[302] In this case, suppression of AGGA is competing with the rare codon AGG, which may account for the efficiency and lack of toxicity of this suppressor tRNA. Frameshift suppressor tRNAs can efficiently incorporate a number of the common 20 amino acids into proteins in vivo,^[303] and as described above, chemically aminoacylated frameshift suppressors have also been used to incorporate unnatural amino acids into proteins in vitro translation systems with four- and five-base codons.

Recently, an orthogonal four-base suppressor tRNA/synthetase pair was generated from archaeal tRNA^{Lys} sequences (see Section 5.5) that efficiently and selectively incorporates the unnatural amino acid homoglutamine into proteins in *E. coli* in response to the quadruplet codon AGGA.^[304] Frameshift suppression with homoglutamine does not significantly affect protein yields or cell growth rates, and is mutually orthogonal with amber suppression. This approach allows the simultaneous incorporation of two unnatural amino acids at distinct sites within proteins. For example, orthogonal AGGA-specific and TAG-specific tRNAs were expressed in the presence of a myoglobin with Gly24→AGGA and Ala75→TAG mutations (Figure 31). The corresponding orthogonal *O*-methyl-L-tyrosine-specific and homoglutamine-specific synthetases were combined in a second plasmid. When cells cotransformed with both plasmids were grown in the presence of the two amino acids, 1.7 mg L⁻¹ of mutant myoglobin was produced. The overall suppression efficiency was approximately 25%. No protein was produced when either of the two unnatural amino acids was excluded. Electrospray mass spectrometric analysis of the full-length protein confirmed that myoglobin contained both unnatural amino acids.^[304] This work suggests that neither the number of available triplet codons nor the translational machinery itself represents a significant barrier to further expansion of the code.

5.5. Other Orthogonal tRNA/Synthetase Pairs in *E. coli*

A “consensus suppressor strategy” has recently been developed for the rational design of additional orthogonal

tRNA/synthetase pairs for use in *E. coli*.^[305] In the first application of this strategy an orthogonal synthetase was generated from the *Methanobacterium thermoautotrophicum* leucyl synthetase. The corresponding suppressor tRNA was based on the consensus sequence of multiple archaeal leucyl tRNAs. The anticodon loop of this consensus tRNA was replaced with the sequence CUXXXAA, where XXX is the reverse complement sequence of distinct nonsense codons. To improve suppression efficiency and further reduce recognition of this tRNA by endogenous *E. coli* synthetases, mutations were introduced into the stems of the consensus tRNA sequence and the resulting mutant tRNA library was subjected to the general negative and positive selections described above.^[260] The two most significant features identified from these studies for an efficient amber orthogonal suppressor tRNA are a CU(X)XXXAA anticodon loop and the lack of noncanonical or mismatched base pairs in the stem regions.^[260,305] An orthogonal tRNA/synthetase pair derived from the archaeal tRNA^{Glu} sequences and the glutamyl tRNA synthetase from *Pyrococcus horikoshii* has also been generated by this strategy (Figure 32).^[306] Both orthogonal pairs have comparable suppression efficiencies to the tyrosyl pair derived from *M. jannaschii*. Other orthogonal pairs have also been reported in the literature for use in *E. coli*. An orthogonal tRNA^{Asp}_{CUA}/AspRS pair was derived from yeast tRNA^{Asp}/AspRS,^[307] in which a yeast AspRS E188K mutant was used to improve recognition for the amber suppressor tRNA^{Asp}_{CUA}.^[307,308] An *E. coli* initiator tRNA^{fMet} amber mutant together with a mutant yeast TyrRS also form an orthogonal pair in *E. coli*.^[309] These synthetases have different active sites and may be useful for evolving specificities for other types of amino acids.

5.6. Genetically Encoding Unnatural Amino Acids in Yeast

The same general strategy developed to genetically encode unnatural amino acids in bacteria should also be applicable to higher organisms. This approach requires the generation of orthogonal tRNA/synthetase pairs for each

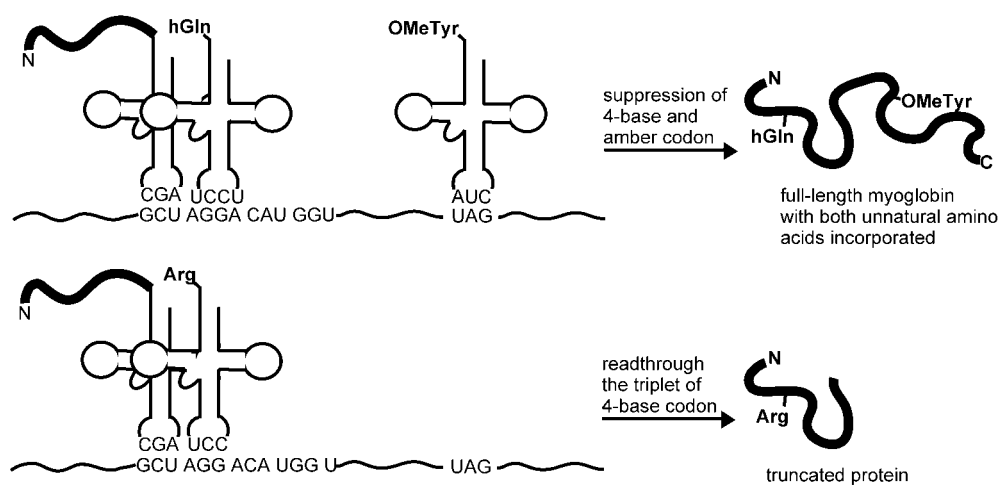


Figure 31. Simultaneous incorporation of the two unnatural amino acids homoglutamine (hGln) and *O*-methyl-L-tyrosine (OMeTyr) into myoglobin by suppression of a four-base codon and an amber nonsense codon.

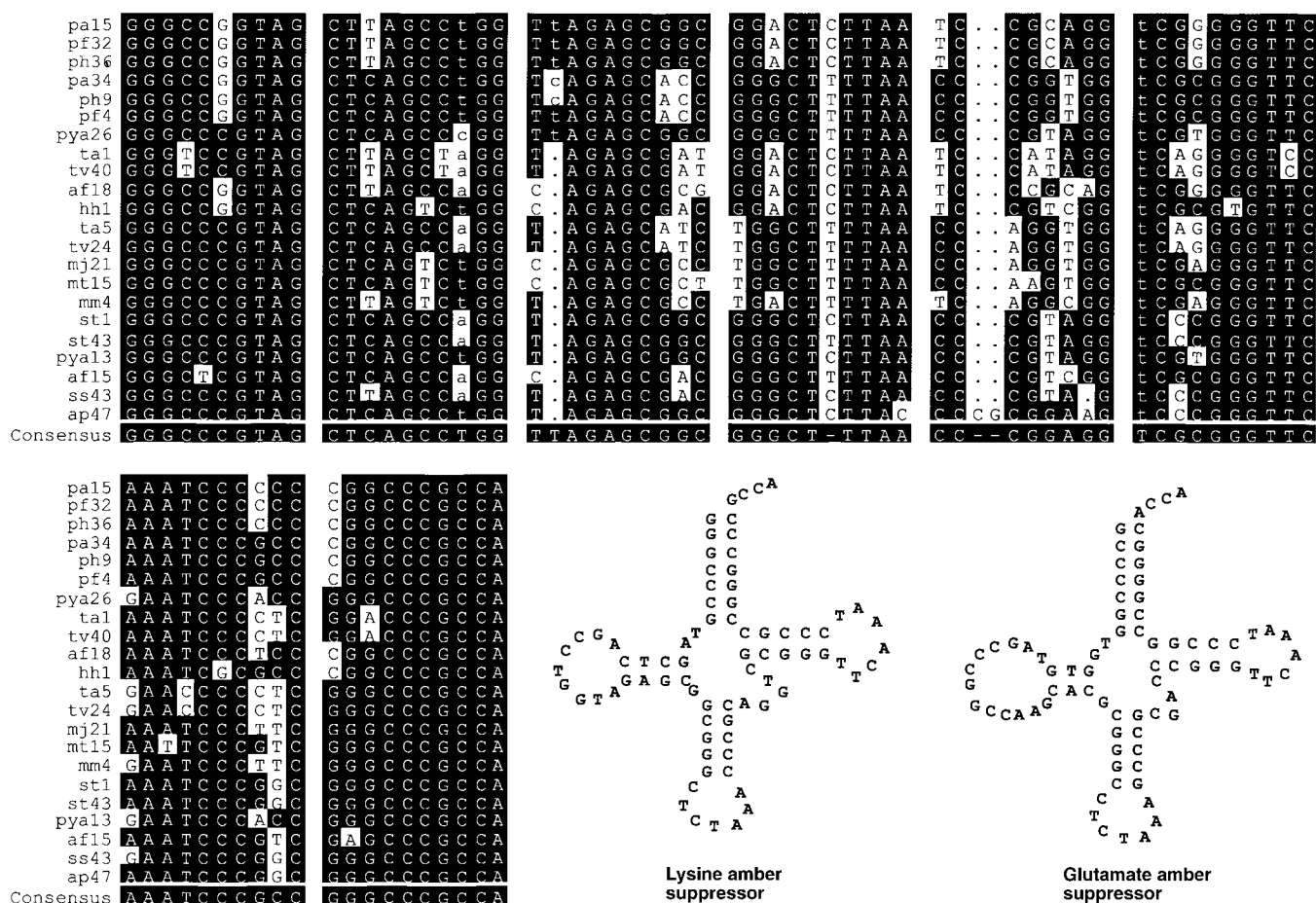


Figure 32. Archaeal lysyl and consensus tRNAs and the derived lysyl and glutamyl amber suppressors.

organism of interest. Most orthogonal tRNA/synthetase pairs that have been reported for eukaryotes are derived from bacterial pairs, reflecting the fact that the divergence of both sequence identity and identity elements of tRNAs from prokaryotes to eukaryotes is greater than that from eukaryotes to archaea and between one eukaryote and another. The transcription of a tRNA gene in eukaryotic cells depends on RNA polymerase III and its associated factors, which recognize A and B box sequences that are internal to the tRNA structural gene.^[310] Some bacterial tRNAs that diverge from the eukaryotic A and B box sequence are not efficiently biosynthesized or processed in eukaryotic cells and their sequences need to be modified accordingly.^[244,311] While all *E. coli* tRNA genes encode full sequences of tRNAs, many bacterial and archaeal and nearly all eukaryotic tRNAs have the 3'-CCA sequence added enzymatically by tRNA nucleotidyltransferase. Thus, it may also be necessary to delete the 3'-CCA in the tRNA gene imported from *E. coli*.^[312] In addition, the generation of a functional tRNA in some cases requires specific 5' and 3' flanking sequences, the removal of introns, posttranscriptional nucleotide modifications, as well as export to the cytoplasm through an exportin-tRNA-dependent process.^[313-315]

A number of orthogonal tRNA/synthetase pairs have been generated for use in eukaryotic organisms. For example,

it has been shown that *E. coli* tyrosyl amber suppressor tRNA^{Tyr}_{CUA} is not aminoacylated by eukaryotic synthetases and acts as an amber suppressor in *S. cerevisiae* in the presence of *E. coli* TyrRS.^[316,317] In addition, *E. coli* TyrRS does not aminoacylate yeast tRNAs.^[318] Thus, *E. coli* tRNA^{Tyr}_{CUA}/TyrRS functions as an orthogonal pair in yeast. A human initiator-tRNA-derived amber suppressor together with *E. coli* GlnRS form another orthogonal pair for use in yeast.^[309] In addition, the *E. coli* tRNA^{Gln}_{CUA}/GlnRS pair has also been used in mammalian cells for efficient suppression of the amber codon.^[244]

To selectively introduce an unnatural amino acid into proteins in eukaryotes, Yokoyama and co-workers screened a small collection of designed active-site variants of *E. coli* TyrRS in a wheat-germ translation system and discovered a mutant synthetase that utilizes 3-iodotyrosine more effectively than tyrosine.^[319] This mutant synthetase was used with the *B. stearrowthermophilus* tRNA^{Tyr}_{CUA} to incorporate 3-iodotyrosine into proteins in mammalian cells.^[311]

To develop a general selection scheme in yeast analogous to that used in *E. coli* for evolving synthetases specific for unnatural amino acids, we created a selection strain of *S. cerevisiae* (MaV203:pGADGAL4 (2 TAG)) that contains the transcriptional activator protein GAL4 in which the codons at two permissive sites were mutated to amber

Cells which incorporate amino acid are blue on X-gal, survive on -URA, and 3-AT.
Cells that do not incorporate an amino acid are white on X-gal and die on -URA or 3-AT.

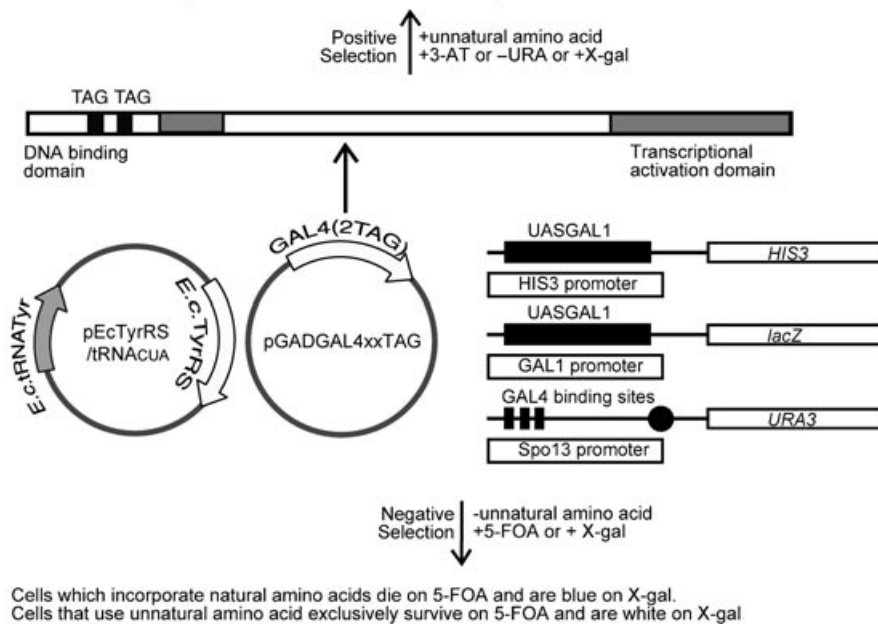


Figure 33. A scheme for selecting mutant synthetases specific for unnatural amino acids in yeast.

nonsense codons (Figure 33). Suppression of these amber codons leads to the production of full-length GAL4, which in turn drives transcription of genomic GAL4 responsive *HIS3*, *URA3*, and *LACZ* reporter genes. Expression of *HIS3* and *URA3* complements the histidine and uracil auxotrophy in this strain and provides a positive selection for clones expressing active tRNA/synthetase pairs. On the other hand, addition of 5-fluoroorotic acid (5-FOA), which is converted into a toxic product by *URA3*, results in the death of cells expressing active tRNA/synthetase pairs. In the absence of the unnatural amino acid, this serves as a negative selection to remove synthetases specific for endogenous amino acids. Like GFP, the *lacZ* reporter can serve as an additional marker to colorimetrically identify active synthetase from inactive ones.

Initially we used the orthogonal *E. coli* tRNA^{Tyr}_{CUA}/TyrRS pair^[316,317] to incorporate unnatural amino acids into proteins in yeast. A mutant *E. coli* TyrRS library (10⁸ in size) was constructed by randomizing five residues in the active site based on the crystal structure of the homologous TyrRS from *Bacillus stearothermophilus* (Figure 17^[262]). By using the selection method described above, aminoacyl-tRNA synthetases were identified that incorporate several unnatural amino acids including *p*-acetyl-L-phenylalanine, *p*-azido-L-phenylalanine, *p*-benzoyl-L-phenylalanine, *O*-methyl-L-tyrosine, *p*-iodo-L-tyrosine, and *O*-(2-propynyl)-L-tyrosine.^[280,320] To characterize the fidelity and efficiency of unnatural amino acid incorporation an amber mutant of human superoxide dismutase was overexpressed in the presence of each tRNA/synthetase pair with or without the cognate unnatural amino acid. In the presence of the unnatural amino acid, superoxide dismutase could be purified (yields are ca. 0.05 mg L⁻¹ corresponding to 20–40% of wild-type protein); no super-

oxide dismutase was produced in the absence of added amino acid.^[320] The purified superoxide dismutase was shown by mass spectrometry to contain the desired unnatural amino acid and no other amino acid at the specified site. A similar approach has been used to evolve orthogonal *E. coli* leucyl tRNA/synthetase pairs that incorporate long-chain, photo-caged, and fluorescent amino acids into proteins in yeast in response to the amber nonsense codon.^[276]

5.7. Genetically Encoding Unnatural Amino Acids in Mammalian Cells

While unnatural amino acids have been incorporated into proteins in mammalian cells in the past,^[321,322] these efforts have relied primarily on transfection of amber suppressor tRNAs that are acylated *in vitro*. This approach severely limits the amount of overall protein that can be produced since the suppressor tRNA must be introduced exogenously and cannot be recycled *in vivo* by an appropriate aminoacyl-tRNA synthetase. Ideally one would like to identify tRNA/synthetase pairs that are orthogonal in mammalian cells, and evolve the substrate specificity of the synthetase such that it can selectively charge a desired unnatural amino acid *in vivo*.

Since the *E. coli* tRNA^{Tyr}_{CUA}/TyrRS pair is orthogonal in mammalian cells (in addition to yeast), and the translational machinery of yeast is similar to that of higher eukaryotes, it should be possible to evolve the specificity of this synthetase in yeast (which is well-suited for genetic selections with large libraries) and transfer the optimized tRNA/synthetase pairs directly to mammalian cells. Unfortunately, *E. coli* tRNA^{Tyr}_{CUA} does not express well in mammalian cells, presumably because of the lack of intact A and B boxes.^[311] However, Yokoyama and co-workers showed that the *B. stearothermophilus* tRNA^{Tyr}_{CUA} can function with *E. coli* TyrRS to suppress the TAG codon with tyrosine *in vivo*. Moreover, this tRNA is not aminoacylated by any endogenous synthetase in mammalian cells.^[311] Indeed, coexpression of an *E. coli* mutant TyrRS led to the incorporation of 3-iodo-L-tyrosine into proteins in response to the TAG codon in both Chinese hamster ovarian cells and human embryonic kidney 293 cells with approximately 95% fidelity.^[311] We have shown that several of the *E. coli* TyrRS mutants evolved in yeast can be used in conjunction with a *B. stearothermophilus* tRNA^{Tyr}_{CUA} to incorporate unnatural amino acids into proteins in mammalian cells.^[323] However, the suppression efficiencies are low, suggesting that improved synthetases or suppressors, or increased expression levels will be required to generate useful amounts of protein.

An orthogonal *B. subtilis* tRNA^{Trp}_{UCA} (*Bst*tRNA^{Trp}_{UCA})/tryptophanyl-tRNA synthetase (TrpRS) has also been developed for use in mammalian cells. Wang and co-workers had previously shown that *B. subtilis* tRNA^{Trp} (*Bst*tRNA^{Trp}) is not

a substrate for TrpRS from yeast and mammalian cells.^[324] To test whether the $BstRNA_{UCA}^{Trp}$ is also orthogonal and can function as an opal suppressor, the codon for Trp68 in a modified bacteriophage T4 fibrin *foldon* gene under control of a CMV promoter^[325] was mutated to TGA. Suppression experiments in human 293T cells together with in vitro aminoacylation assays showed that *B. subtilis* TrpRS ($BsTrpRS$) aminoacylates only $BstRNA_{UCA}^{Trp}$ and no other endogenous mammalian tRNAs, and that the expressed $BstRNA_{UCA}^{Trp}$ is charged only by its cognate $BsTrpRS$ and not by other endogenous mammalian synthetases.^[326]

The crystal structure of the homologous *Bacillus stearothermophilus* TrpRS was then used to rationally design a synthetase mutant (V144P $BsTrpRS$) that selectively charges 5-substituted Trp analogues. Val144 points directly towards C5 of tryptophan, thus leading to unfavorable steric interactions with any tryptophan analogue containing a substituent at the C5-position. Mutation of Val144 to a smaller amino acid might be expected to accommodate 5-substituted tryptophan analogues. This mutant was generated and assayed for its ability to suppress the opal nonsense codon in the *foldon* gene in the presence of 5-hydroxytryptophan (5-HTPP). Indeed, expression of full-length protein was seen only in the presence of 5-HTPP (Figure 34).^[326] No full-length protein was expressed in the absence of 5-HTPP under otherwise identical conditions, which indicates that this mutant synthetase does

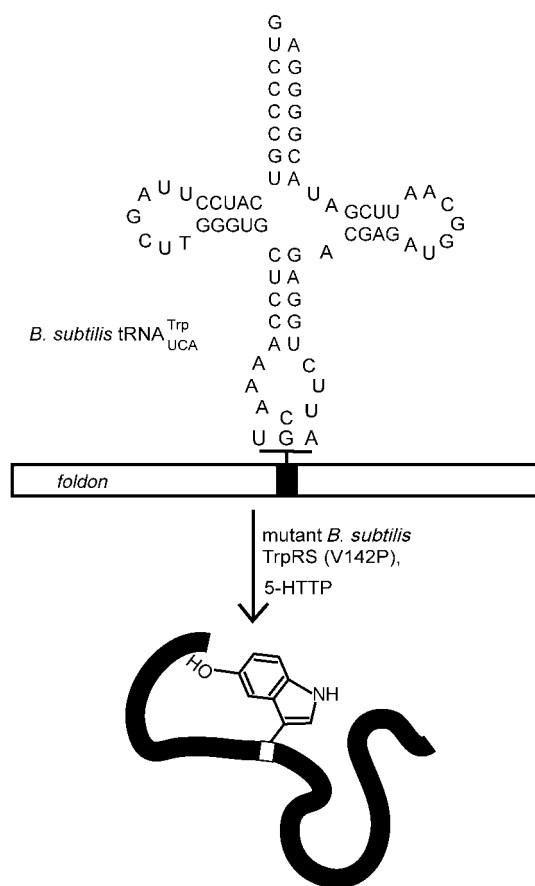


Figure 34. Incorporation of 5-HTPP into the foldon in mammalian cells by suppression of the opal nonsense codon.

not use an endogenous amino acid as a substrate. Electro-spray mass spectrometry of the mutant foldon protein verified site-specific incorporation of 5-HTPP at position 68 with a fidelity of >97%. The yield of the 5-HTPP mutant protein was approximately $100 \mu\text{g L}^{-1}$ of culture, compared to that of about 1 mg L^{-1} for wild-type protein. To evolve the $BsTrpRS$ to charge more complex tryptophan analogues, the $BstRNA_{UCA}^{Trp}/BsTrpRS$ can be moved back into the yeast system and the selection scheme described above can be used to isolate desired synthetase mutants from a large library of active-site mutants.

6. Outlook

The approach described above has proved remarkably effective in allowing us to add a large number of novel amino acids to the genetic codes of both prokaryotic and eukaryotic organisms. Coincidentally, it has recently been shown that nature has evolved a similar strategy (an orthogonal amber suppressor tRNA/synthetase pair) to genetically encode the unnatural amino acid pyrrolysine in *Methanosarcina barkeri*.^[5] Future work in this field will likely focus on expanding the nature and number of amino acids that can be genetically encoded in both prokaryotic and eukaryotic organisms, including multicellular organisms. Additional orthogonal pairs that suppress three- and four-base codons are also being developed. It may even be possible to delete rare redundant codons from the *E. coli* genome and use them instead to encode unnatural amino acids. The ability to genetically encode unnatural amino acids should provide powerful probes, both in vitro and in vivo, of protein structure and function. It may also allow the rational design or evolution of proteins with novel properties. Examples might include homogeneous glycosylated or PEG-derivatized therapeutic proteins with improved pharmacological properties, fluorescent proteins that act as sensors of small molecules and protein-protein interactions in the cell, or proteins whose activity can be photoregulated in vivo. It should also be possible to incorporate non-amino acid building blocks into proteins or perhaps even create biopolymers with entirely unnatural backbones. Finally, the ability to add novel amino acids to the genetic codes of organisms will allow us to experimentally test whether there is an evolutionary advantage for organisms with more than 20 amino acid building blocks.

Received: May 11, 2004

- [1] J. T. Wong, *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 1909.
- [2] M. Di Giulio, M. Medugno, *J. Mol. Evol.* **1999**, *49*, 1.
- [3] T. A. Ronneberg, L. F. Landweber, S. J. Freeland, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13690.
- [4] A. Bock, K. Forchhammer, J. Heider, W. Leinfelder, G. Sawers, B. Veprek, F. Zinoni, *Mol. Microbiol.* **1991**, *5*, 515.
- [5] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459.

- [6] A. R. Fersht, J. P. Shi, A. J. Wilkinson, D. M. Blow, P. Carter, M. M. Y. Waye, G. P. Winter, *Angew. Chem.* **1984**, *96*, 455; *Angew. Chem. Int. Ed. Engl.* **1984**, *23*, 467.
- [7] M. Smith, *Annu. Rev. Genet.* **1985**, *19*, 423.
- [8] J. Pons, A. Rajpal, J. F. Kirsch, *Protein Sci.* **1999**, *8*, 958.
- [9] K. E. Neet, D. E. Koshland, Jr., *Proc. Natl. Acad. Sci. USA* **1966**, *56*, 1606.
- [10] L. Polgar, M. L. Bender, *J. Am. Chem. Soc.* **1966**, *88*, 3153.
- [11] Z. P. Wu, D. Hilvert, *J. Am. Chem. Soc.* **1989**, *111*, 4513.
- [12] Z. P. Wu, D. Hilvert, *J. Am. Chem. Soc.* **1990**, *112*, 5647.
- [13] J. Q. Liu, M. S. Jiang, G. M. Luo, G. L. Yan, J. C. Shen, *Biotechnol. Lett.* **1998**, *20*, 693.
- [14] H. L. Levine, Y. Nakagawa, E. T. Kaiser, *Biochem. Biophys. Res. Commun.* **1977**, *76*, 64.
- [15] H. L. Levine, E. T. Kaiser, *J. Am. Chem. Soc.* **1978**, *100*, 7670.
- [16] E. T. Kaiser, D. S. Lawrence, *Science* **1984**, *226*, 505.
- [17] D. R. Corey, P. G. Schultz, *Science* **1987**, *238*, 1401.
- [18] Y. Chen, Y. W. Ebright, R. H. Ebright, *Science* **1994**, *265*, 90.
- [19] T. M. Rana, C. F. Meares, *J. Am. Chem. Soc.* **1990**, *112*, 2457.
- [20] T. M. Rana, C. F. Meares, *J. Am. Chem. Soc.* **1991**, *113*, 1859.
- [21] J. B. Ghaim, D. P. Greiner, C. F. Meares, R. B. Gennis, *Biochemistry* **1995**, *34*, 11311.
- [22] J. T. Owens, A. Chmura, K. Murakami, N. Fujita, A. Ishihama, C. F. Meares, *Biochemistry* **1998**, *37*, 7670.
- [23] G. M. Culver, H. F. Noller, *Methods Enzymol.* **2000**, *318*, 461.
- [24] J. A. Silverman, P. B. Harbury, *J. Biol. Chem.* **2002**, *277*, 30968.
- [25] G. R. Jacobson, M. H. Schaffer, G. R. Stark, T. C. Vanaman, *J. Biol. Chem.* **1973**, *248*, 6583.
- [26] M. H. Akabas, D. A. Stauffer, M. Xu, A. Karlin, *Science* **1992**, *258*, 307.
- [27] A. Karlin, *J. Gen. Physiol.* **2001**, *117*, 235.
- [28] A. P. Todd, J. Cong, F. Levinthal, C. Levinthal, W. L. Hubbell, *Proteins Struct. Funct. Genet.* **1989**, *6*, 294.
- [29] W. L. Hubbell, C. Altenbach, C. M. Hubbell, H. G. Khorana, *Adv. Protein Chem.* **2003**, *63*, 243.
- [30] L. M. Mannuzzu, M. M. Moronne, E. Y. Isacoff, *Science* **1996**, *271*, 213.
- [31] B. A. Griffin, S. R. Adams, R. Y. Tsien, *Science* **1998**, *281*, 269.
- [32] S. R. Adams, R. E. Campbell, L. A. Gross, B. R. Martin, G. K. Walkup, Y. Yao, J. Llopis, R. Y. Tsien, *J. Am. Chem. Soc.* **2002**, *124*, 6063.
- [33] G. Simmons, P. R. Clapham, L. Picard, R. E. Offord, M. M. Rosenklide, T. W. Schwartz, R. Buser, T. N. C. Wells, A. E. I. Proudfoot, *Science* **1997**, *276*, 276.
- [34] K. F. Geoghegan, J. G. Stroh, *Bioconjugate Chem.* **1992**, *3*, 138.
- [35] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149.
- [36] S. B. Kent, *Annu. Rev. Biochem.* **1988**, *57*, 957.
- [37] R. S. Roy, B. Imperiali, *Protein Eng.* **1997**, *10*, 691.
- [38] I. Hamachi, R. Eboshi, J.-I. Watanabe, S. Shinkai, *J. Am. Chem. Soc.* **2000**, *122*, 4530.
- [39] J. J. Nestor, Jr., T. L. Ho, R. A. Simpson, B. L. Horner, G. H. Jones, G. I. McRae, B. H. Vickery, *J. Med. Chem.* **1982**, *25*, 795.
- [40] J. J. Nestor, Jr., R. Tahilramani, T. L. Ho, G. I. McRae, B. H. Vickery, *J. Med. Chem.* **1988**, *31*, 65.
- [41] K. L. Liu, B. L. He, S. B. Xiao, Q. M. Xia, X. Fang, Z. Y. Wang, *Int. J. Pept. Protein Res.* **1990**, *35*, 157.
- [42] B. Bilgicler, A. Fichera, K. Kumar, *J. Am. Chem. Soc.* **2001**, *123*, 4393.
- [43] C. E. Schafmeister, J. Po, G. L. Verdine, *J. Am. Chem. Soc.* **2000**, *122*, 5891.
- [44] V. Cerovsky, E. Welker, H. A. Scheraga, *J. Pept. Res.* **2003**, *61*, 140.
- [45] E. T. Kaiser, *Acc. Chem. Res.* **1989**, *22*, 47.
- [46] D. Y. Jackson, J. Burnier, C. Quan, M. Stanley, J. Tom, J. A. Wells, *Science* **1994**, *266*, 243.
- [47] M. Schnolzer, S. B. Kent, *Science* **1992**, *256*, 221.
- [48] K. Rose, *J. Am. Chem. Soc.* **1994**, *116*, 30.
- [49] C.-F. Liu, J. P. Tam, *J. Am. Chem. Soc.* **1994**, *116*, 4149.
- [50] C.-F. Liu, J. P. Tam, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 6584.
- [51] D. R. Englebretsen, B. G. Garnham, D. A. Bergman, P. F. Alewood, *Tetrahedron Lett.* **1995**, *36*, 8871.
- [52] M. Baca, T. W. Muir, M. Schnolzer, S. B. H. Kent, *J. Am. Chem. Soc.* **1995**, *117*, 1881.
- [53] D. S. Kemp, R. I. Carey, *J. Org. Chem.* **1993**, *58*, 2216.
- [54] P. E. Dawson, T. W. Muir, I. Clarklewis, S. B. H. Kent, *Science* **1994**, *266*, 776.
- [55] P. E. Dawson, S. B. H. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923.
- [56] L. E. Canne, S. J. Bark, S. B. H. Kent, *J. Am. Chem. Soc.* **1996**, *118*, 5891.
- [57] M. Baca, S. B. Kent, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11638.
- [58] D. W. Low, M. G. Hill, *J. Am. Chem. Soc.* **1998**, *120*, 11536.
- [59] G. G. Kochendoerfer, S. Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. Kent, J. A. Bradburne, *Science* **2003**, *299*, 884.
- [60] A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz, S. Weiss, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5179.
- [61] L. Zhang, J. P. Tam, *J. Am. Chem. Soc.* **1997**, *119*, 2363.
- [62] J. A. Camarero, J. Pavel, T. W. Muir, *Angew. Chem.* **1998**, *110*, 361; *Angew. Chem. Int. Ed.* **1998**, *37*, 347.
- [63] H. Ishida, Z. Qi, M. Sokabe, K. Donowaki, Y. Inoue, *J. Org. Chem.* **2001**, *66*, 2978.
- [64] T. N. Schumacher, L. M. Mayr, D. L. Minor, Jr., M. A. Milholten, M. W. Burgess, P. S. Kim, *Science* **1996**, *271*, 1854.
- [65] D. M. Eckert, V. N. Malashkevich, L. H. Hong, P. A. Carr, P. S. Kim, *Cell* **1999**, *99*, 103.
- [66] R. E. Offord, *Protein Eng.* **1987**, *1*, 151.
- [67] C. J. Wallace, *Curr. Opin. Biotechnol.* **1995**, *6*, 403.
- [68] D. F. Dyckes, T. Creighton, R. C. Sheppard, *Nature* **1974**, *247*, 202.
- [69] C. J. Wallace, I. Clark-Lewis, *J. Biol. Chem.* **1992**, *267*, 3852.
- [70] C. J. Wallace, *FASEB J.* **1993**, *7*, 505.
- [71] D. S. Wuttke, H. B. Gray, S. L. Fisher, B. Imperiali, *J. Am. Chem. Soc.* **1993**, *115*, 8455.
- [72] C. H. Wong, K. T. Wang, *Experientia* **1991**, *47*, 1123.
- [73] G. A. Homandberg, M. Laskowski, Jr., *Biochemistry* **1979**, *18*, 586.
- [74] K. Inouye, K. Watanabe, K. Morihara, Y. Tochino, T. Kanaya, J. Emura, S. Sakakibara, *J. Am. Chem. Soc.* **1979**, *101*, 751.
- [75] H. D. Jakubke, P. Kuhl, A. Koennecke, *Angew. Chem.* **1985**, *97*, 79; *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 85.
- [76] V. Schellenberger, H. D. Jakubke, *Angew. Chem.* **1991**, *103*, 1440; *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 1437.
- [77] J. Bongers, T. Lambros, W. Liu, M. Ahmad, R. M. Campbell, A. M. Felix, E. P. Heimer, *J. Med. Chem.* **1992**, *35*, 3934.
- [78] J. Bongers, W. Liu, T. Lambros, K. Breddam, R. M. Campbell, A. M. Felix, E. P. Heimer, *Int. J. Pept. Protein Res.* **1994**, *44*, 123.
- [79] T. Nakatsuka, T. Sasaki, E. T. Kaiser, *J. Am. Chem. Soc.* **1987**, *109*, 3808.
- [80] L. Abrahmsen, J. Tom, J. Burnier, K. A. Butcher, A. Kossiakov, J. A. Wells, *Biochemistry* **1991**, *30*, 4151.
- [81] T. K. Chang, D. Y. Jackson, J. P. Burnier, J. A. Wells, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 12544.
- [82] H. F. Gaertner, R. E. Offord, R. Cotton, D. Timms, R. Camble, K. Rose, *J. Biol. Chem.* **1994**, *269*, 7224.
- [83] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705.
- [84] K. Severinov, T. W. Muir, *J. Biol. Chem.* **1998**, *273*, 16205.
- [85] T. C. Evans, Jr., J. Benner, M. Q. Xu, *Protein Sci.* **1998**, *7*, 2256.
- [86] T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249.

- [87] F. B. Perler, M. Q. Xu, H. Paulus, *Curr. Opin. Chem. Biol.* **1997**, *1*, 292.
- [88] T. J. Tolbert, C.-H. Wong, *Angew. Chem.* **2002**, *114*, 2275; *Angew. Chem. Int. Ed.* **2002**, *41*, 2171.
- [89] C. J. Noren, J. Wang, F. B. Perler, *Angew. Chem.* **2000**, *112*, 458; *Angew. Chem. Int. Ed.* **2000**, *39*, 450.
- [90] J. P. Tam, Q. Yu, *Biopolymers* **1998**, *46*, 319.
- [91] R. S. Roy, O. Allen, C. T. Walsh, *Chem. Biol.* **1999**, *6*, 789.
- [92] R. J. Hondal, B. L. Nilsson, R. T. Raines, *J. Am. Chem. Soc.* **2001**, *123*, 5140.
- [93] M. D. Gieselman, L. Xie, W. A. van der Donk, *Org. Lett.* **2001**, *3*, 1331.
- [94] R. Quaderer, A. Sewing, D. Hilvert, *Helv. Chim. Acta* **2001**, *84*, 1197.
- [95] J. R. Sydor, M. Mariano, S. Sideris, S. Nock, *Bioconjugate Chem.* **2002**, *13*, 707.
- [96] A. Iakovenko, E. Rostkova, E. Merzlyak, A. M. Hillebrand, N. H. Thoma, R. S. Goody, K. Alexandrov, *FEBS Lett.* **2000**, *468*, 155.
- [97] G. J. Cotton, B. Ayers, R. Xu, T. W. Muir, *J. Am. Chem. Soc.* **1999**, *121*, 1100.
- [98] B. Ayers, U. K. Blaschke, J. A. Camarero, G. J. Cotton, M. Holford, T. W. Muir, *Biopolymers* **2000**, *51*, 343.
- [99] G. J. Cotton, T. W. Muir, *Chem. Biol.* **2000**, *7*, 253.
- [100] R. M. Hofmann, G. J. Cotton, E. J. Chang, E. Vidal, D. Veach, W. Bornmann, T. W. Muir, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 3091.
- [101] J. Mukhopadhyay, A. N. Kapanidis, V. Mekler, E. Kortkhonjia, Y. W. Ebright, R. H. Ebright, *Cell* **2001**, *106*, 453.
- [102] V. Mekler, E. Kortkhonjia, J. Mukhopadhyay, J. Knight, A. Revyakin, A. N. Kapanidis, W. Niu, Y. W. Ebright, R. Levy, R. H. Ebright, *Cell* **2002**, *108*, 599.
- [103] M. A. Dwyer, W. Lu, J. J. Dwyer, A. A. Kossiakoff, *Chem. Biol.* **2000**, *7*, 263.
- [104] D. Macmillan, C. R. Bertozzi, *Tetrahedron* **2000**, *56*, 9515.
- [105] T. J. Tolbert, C.-H. Wong, *J. Am. Chem. Soc.* **2000**, *122*, 5421.
- [106] K. Alexandrov, I. Heinemann, T. Durek, V. Sidorovitch, R. S. Goody, H. Waldmann, *J. Am. Chem. Soc.* **2002**, *124*, 5648.
- [107] M. Huse, M. N. Holford, J. Kuriyan, T. W. Muir, *J. Am. Chem. Soc.* **2000**, *122*, 8337.
- [108] M. Huse, T. W. Muir, L. Xu, Y.-G. Chen, J. Kuriyan, J. Massague, *Mol. Cell* **2001**, *8*, 671.
- [109] W. Lu, D. Gong, D. Bar-Sagi, P. A. Cole, *Mol. Cell* **2001**, *8*, 759.
- [110] U. Arnold, M. P. Hinderaker, B. L. Nilsson, B. R. Huck, S. H. Gellman, R. T. Raines, *J. Am. Chem. Soc.* **2002**, *124*, 8522.
- [111] F. H. C. Crick, *Symp. Soc. Exp. Biol.* **1958**, *12*, 138.
- [112] F. Chapeville, F. Lipmann, G. von Ehrenstein, B. Weisblum, W. J. Ray, Jr., S. Benzer, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 1086.
- [113] A. E. Johnson, W. R. Woodward, E. Herbert, J. R. Menninger, *Biochemistry* **1976**, *15*, 569.
- [114] P. Schimmel, *Annu. Rev. Biochem.* **1987**, *56*, 125.
- [115] M. Ibba, D. Soll, *Annu. Rev. Biochem.* **2000**, *69*, 617.
- [116] T. G. Heckler, L. H. Chang, Y. Zama, T. Naka, M. S. Chorghade, S. M. Hecht, *Biochemistry* **1984**, *23*, 1468.
- [117] T. G. Heckler, Y. Zama, T. Naka, S. M. Hecht, *J. Biol. Chem.* **1983**, *258*, 4492.
- [118] J. R. Roesler, M. S. Chorghade, S. M. Hecht, *Biochemistry* **1986**, *25*, 6361.
- [119] T. G. Heckler, J. R. Roesler, C. Xu, P. I. Chang, S. M. Hecht, *Biochemistry* **1988**, *27*, 7254.
- [120] G. Baldini, B. Martoglio, A. Schachenmann, C. Zugliani, J. Brunner, *Biochemistry* **1988**, *27*, 7951.
- [121] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182.
- [122] F. H. C. Crick, L. Barret, S. Brenner, R. Watts-Tobin, *Nature* **1961**, *192*, 1227.
- [123] S. Brenner, A. O. Stretton, S. Kaplan, *Nature* **1965**, *206*, 994.
- [124] J. H. Miller, D. Ganem, P. Lu, A. Schmitz, *J. Mol. Biol.* **1977**, *109*, 275.
- [125] L. Bossi, J. R. Roth, *Nature* **1980**, *286*, 123.
- [126] J. A. Ellman, D. Mendel, P. G. Schultz, *Science* **1992**, *255*, 197.
- [127] Y. Kwok, J. T. Wong, *Can. J. Biochem.* **1980**, *58*, 213.
- [128] A. G. Bruce, J. F. Atkins, N. Wills, O. Uhlenbeck, R. F. Gesteland, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 7127.
- [129] A. Miller, A. Albertini, *J. Mol. Biol.* **1983**, *164*, 59.
- [130] D. Ayer, M. Yarus, *Science* **1986**, *231*, 393.
- [131] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **1987**, *15*, 8783.
- [132] C. J. Noren, S. J. Anthony-Cahill, D. J. Suich, K. A. Noren, M. C. Griffith, P. G. Schultz, *Nucleic Acids Res.* **1990**, *18*, 83.
- [133] S. A. Robertson, J. A. Ellman, P. G. Schultz, *J. Am. Chem. Soc.* **1991**, *113*, 2722.
- [134] S. R. Adams, J. P. Y. Kao, R. Y. Tsien, *J. Am. Chem. Soc.* **1989**, *111*, 7957.
- [135] A. Patchornik, B. Amit, R. B. Woodward, *J. Am. Chem. Soc.* **1970**, *92*, 6333.
- [136] J. A. Ellman, D. Mendel, S. Anthony-Cahill, C. J. Noren, P. G. Schultz, *Methods Enzymol.* **1992**, *202*, 301.
- [137] J. D. Bain, C. G. Glabe, T. A. Dix, A. R. Chamberlin, E. S. Dials, *J. Am. Chem. Soc.* **1989**, *111*, 8013.
- [138] V. W. Cornish, D. Mendel, P. G. Schultz, *Angew. Chem.* **1995**, *107*, 677; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 621.
- [139] J. S. Thorson, E. Chapman, E. C. Murphy, P. G. Schultz, J. K. Judice, *J. Am. Chem. Soc.* **1995**, *117*, 1157.
- [140] J. T. Koh, V. W. Cornish, P. G. Schultz, *Biochemistry* **1997**, *36*, 11314.
- [141] K. B. Wiberg, K. E. Laidig, *J. Am. Chem. Soc.* **1987**, *109*, 5935.
- [142] M. J. Kamlet, J. L. M. Abboud, M. H. Abraham, R. W. Taft, *J. Org. Chem.* **1983**, *48*, 2877.
- [143] E. Chapman, J. S. Thorson, P. G. Schultz, *J. Am. Chem. Soc.* **1997**, *119*, 7151.
- [144] S. T. Cload, D. R. Liu, W. A. Froland, P. G. Schultz, *Chem. Biol.* **1996**, *3*, 1033.
- [145] G. F. Short, S. Y. Golovine, S. M. Hecht, *Biochemistry* **1999**, *38*, 8808.
- [146] D. M. Kim, T. Kigawa, C. Y. Choi, S. Yokoyama, *Eur. J. Biochem.* **1996**, *239*, 881.
- [147] T. Hoshaka, K. Sato, M. Sisido, K. Takai, S. Yokoyama, *FEBS Lett.* **1994**, *344*, 171.
- [148] C. H. Ma, W. Kudlicki, O. W. Odom, G. Kramer, B. Hardesty, *Biochemistry* **1993**, *32*, 7939.
- [149] A. K. Kowal, J. S. Oliver, *Nucleic Acids Res.* **1997**, *25*, 4685.
- [150] A. C. Forster, Z. Tan, M. N. Nalam, H. Lin, H. Qu, V. W. Cornish, S. C. Blacklow, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 6353.
- [151] T. Hoshaka, Y. Ashizuka, H. Murakami, M. Sisido, *J. Am. Chem. Soc.* **1996**, *118*, 9778.
- [152] H. Murakami, T. Hoshaka, Y. Ashizuka, M. Sisido, *J. Am. Chem. Soc.* **1998**, *120*, 7520.
- [153] T. Hoshaka, D. Kajihara, Y. Ashizuka, H. Murakami, M. Sisido, *J. Am. Chem. Soc.* **1999**, *121*, 34.
- [154] M. Taki, T. Hoshaka, H. Murakami, K. Taira, M. Sisido, *J. Am. Chem. Soc.* **2002**, *124*, 14586.
- [155] M. Taki, T. Hoshaka, H. Murakami, K. Taira, M. Sisido, *J. Am. Chem. Soc.* **2002**, *124*, 14586.
- [156] T. Hoshaka, M. Sisido, *Curr. Opin. Chem. Biol.* **2002**, *6*, 809.
- [157] T. Hoshaka, Y. Ashizuka, H. Sasaki, H. Murakami, M. Sisido, *J. Am. Chem. Soc.* **1999**, *121*, 12194.
- [158] T. Hoshaka, Y. Ashizuka, H. Taira, H. Murakami, M. Sisido, *Biochemistry* **2001**, *40*, 11060.
- [159] T. Hoshaka, Y. Ashizuka, H. Murakami, M. Sisido, *Nucleic Acids Res.* **2001**, *29*, 3646.

- [160] J. D. Bain, C. Switzer, A. R. Chamberlin, S. A. Benner, *Nature* **1992**, 356, 537.
- [161] M. W. Nowak, P. C. Kearney, J. R. Sampson, M. E. Saks, C. G. Labarca, S. K. Silverman, W. G. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G. Schultz, D. A. Dougherty, H. A. Lester, *Science* **1995**, 268, 439.
- [162] M. E. Saks, J. R. Sampson, M. W. Nowak, P. C. Kearney, F. Y. Du, J. N. Abelson, H. A. Lester, D. A. Dougherty, *J. Biol. Chem.* **1996**, 271, 23169.
- [163] D. A. Dougherty, *Curr. Opin. Chem. Biol.* **2000**, 4, 645.
- [164] D. L. Beene, D. A. Dougherty, H. A. Lester, *Curr. Opin. Neurobiol.* **2003**, 13, 264.
- [165] W. G. Zhong, J. P. Gallivan, Y. N. Zhang, L. T. Li, H. A. Lester, D. A. Dougherty, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 12088.
- [166] D. L. Beene, G. S. Brandt, W. Zhong, N. M. Zacharias, H. A. Lester, D. A. Dougherty, *Biochemistry* **2002**, 41, 10262.
- [167] C. A. Hutchison III, S. Phillips, M. H. Edgell, S. Gillam, P. Jahnke, M. Smith, *J. Biol. Chem.* **1978**, 253, 6551.
- [168] G. Dalbadie-McFarland, L. W. Cohen, A. D. Riggs, C. Morin, K. Itakura, J. H. Richards, *Proc. Natl. Acad. Sci. USA* **1982**, 79, 6409.
- [169] I. S. Sigal, B. G. Harwood, R. Arentzen, *Proc. Natl. Acad. Sci. USA* **1982**, 79, 7157.
- [170] G. Winter, A. R. Fersht, A. J. Wilkinson, M. Zoller, M. Smith, *Nature* **1982**, 299, 756.
- [171] J. E. Villafranca, E. E. Howell, D. H. Voet, M. S. Strobel, R. C. Ogden, J. N. Abelson, J. Kraut, *Science* **1983**, 222, 782.
- [172] C. S. Craik, S. Roczniak, C. Largman, W. J. Rutter, *Science* **1987**, 237, 909.
- [173] J. A. Wells, D. A. Estell, *Trends Biochem. Sci.* **1988**, 13, 291.
- [174] A. R. Fersht, *Biochemistry* **1987**, 26, 8031.
- [175] P. T. Jones, P. H. Dear, J. Foote, M. S. Neuberger, G. Winter, *Nature* **1986**, 321, 522.
- [176] J. Nardelli, T. J. Gibson, C. Vesque, P. Charnay, *Nature* **1991**, 349, 175.
- [177] J. R. Desjarlais, J. M. Berg, *Proteins Struct. Funct. Genet.* **1992**, 12, 101.
- [178] J. Nardelli, T. Gibson, P. Charnay, *Nucleic Acids Res.* **1992**, 20, 4137.
- [179] Y. Choo, A. Klug, *Proc. Natl. Acad. Sci. USA* **1994**, 91, 11163.
- [180] M. Isalan, Y. Choo, A. Klug, *Proc. Natl. Acad. Sci. USA* **1997**, 94, 5617.
- [181] M. Isalan, A. Klug, Y. Choo, *Biochemistry* **1998**, 37, 12026.
- [182] Y. Choo, M. Isalan, *Curr. Opin. Struct. Biol.* **2000**, 10, 411.
- [183] S. A. Wolfe, L. Nekludova, C. O. Pabo, *Annu. Rev. Biophys. Biomol. Struct.* **2000**, 29, 183.
- [184] J. A. Wells, M. Vasser, D. B. Powers, *Gene* **1985**, 34, 315.
- [185] B. C. Cunningham, P. Jhurani, P. Ng, J. A. Wells, *Science* **1989**, 243, 1330.
- [186] B. C. Cunningham, J. A. Wells, *Science* **1989**, 244, 1081.
- [187] D. W. Leung, E. Chen, D. V. Goeddel, *Technique* **1989**, 1, 11.
- [188] R. C. Cadwell, G. F. Joyce, *PCR Methods Appl.* **1992**, 2, 28.
- [189] A. Greener, M. Callahan, B. Jerpseth, *Methods Mol. Biol.* **1996**, 57, 375.
- [190] W. P. C. Stemmer, *Nature* **1994**, 370, 389.
- [191] H. K. Mitchell, C. Niemann, *J. Am. Chem. Soc.* **1947**, 69, 1232.
- [192] M. V. Simpson, E. Farber, H. Tarver, *J. Biol. Chem.* **1950**, 182, 81.
- [193] M. H. Richmond, *Bacteriol. Rev.* **1962**, 26, 398.
- [194] M. Levine, H. Tarver, *J. Biol. Chem.* **1951**, 192, 835.
- [195] D. Gross, H. Tarver, *J. Biol. Chem.* **1955**, 217, 169.
- [196] R. Munier, G. N. Cohen, *Biochim. Biophys. Acta* **1956**, 21, 592.
- [197] G. N. Cohen, D. B. Cowie, *C. R. Hebd. Seances Acad. Sci.* **1957**, 244, 680.
- [198] D. B. Cowie, G. N. Cohen, *Biochim. Biophys. Acta* **1957**, 26, 252.
- [199] G. Hortin, I. Boime, *Methods Enzymol.* **1983**, 96, 777.
- [200] G. Brawerman, M. Ycas, *Arch. Biochem. Biophys.* **1957**, 68, 112.
- [201] M. H. Richmond, *Biochem. J.* **1959**, 73, 261.
- [202] D. B. Cowie, G. N. Cohen, E. T. Bolton, H. de Robichon-Szulmajster, *Biochim. Biophys. Acta* **1959**, 34, 39.
- [203] O. M. Rennert, H. S. Anker, *Biochemistry* **1961**, 13, 471.
- [204] N. Budisa, C. Minks, F. J. Medrano, J. Lutz, R. Huber, L. Moroder, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 455.
- [205] M. Ibba, P. Kast, H. Hennecke, *Biochemistry* **1994**, 33, 7107.
- [206] M. Ibba, H. Hennecke, *FEBS Lett.* **1995**, 364, 272.
- [207] N. Sharma, R. Furter, P. Kast, D. A. Tirrell, *FEBS Lett.* **2000**, 467, 37.
- [208] K. Kirshenbaum, S. I. Carrico, A. D. Tirrell, *ChemBioChem* **2002**, 3, 235.
- [209] F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soll, S. Nishimura, *J. Biol. Chem.* **2000**, 275, 40324.
- [210] V. Doring, H. D. Mootz, L. A. Nangle, T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel, P. Marliere, *Science* **2001**, 292, 501.
- [211] Y. Tang, D. A. Tirrell, *Biochemistry* **2002**, 41, 10635.
- [212] W. Yang, W. A. Hendrickson, E. T. Kalman, R. J. Crouch, *J. Biol. Chem.* **1990**, 265, 13553.
- [213] W. Yang, W. A. Hendrickson, R. J. Crouch, Y. Satow, *Science* **1990**, 249, 1398.
- [214] J. O. Boles, K. Lewinski, M. Kunkle, J. D. Odom, B. Dunlap, L. Lebiada, M. Hatada, *Nat. Struct. Biol.* **1994**, 1, 283.
- [215] N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann, R. Huber, *Eur. J. Biochem.* **1995**, 230, 788.
- [216] N. Budisa, W. Karnbrock, S. Steinbacher, A. Humm, L. Prade, T. Neufeind, L. Moroder, R. Huber, *J. Mol. Biol.* **1997**, 270, 616.
- [217] J. H. Bae, S. Alefelder, J. T. Kaiser, R. Friedrich, L. Moroder, R. Huber, N. Budisa, *J. Mol. Biol.* **2001**, 309, 925.
- [218] N. Budisa, R. Huber, R. Golbik, C. Minks, E. Weyher, L. Moroder, *Eur. J. Biochem.* **1998**, 253, 1.
- [219] C. Minks, R. Huber, L. Moroder, N. Budisa, *Biochemistry* **1999**, 38, 10649.
- [220] C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber, L. Moroder, *Angew. Chem.* **2001**, 113, 949; *Angew. Chem. Int. Ed.* **2001**, 40, 923.
- [221] P. Wang, Y. Tang, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, 125, 6900.
- [222] Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado, D. A. Tirrell, *Angew. Chem.* **2001**, 113, 1542; *Angew. Chem. Int. Ed.* **2001**, 40, 1494.
- [223] Y. Tang, D. A. Tirrell, *J. Am. Chem. Soc.* **2001**, 123, 11089.
- [224] N. Budisa, M. Rubini, J. H. Bae, E. Weyher, W. Wenger, R. Golbik, R. Huber, L. Moroder, *Angew. Chem.* **2002**, 114, 4238; *Angew. Chem. Int. Ed.* **2002**, 41, 4066.
- [225] J. H. Bae, M. Rubini, G. Jung, G. Wiegand, M. H. J. Seifert, M. K. Azim, J.-S. Kim, A. Zumbusch, T. A. Holak, L. Moroder, R. Huber, N. Budisa, *J. Mol. Biol.* **2003**, 328, 1071.
- [226] H. Diewel, E. Daub, V. Robinson, J. F. Honek, *Biochemistry* **1997**, 36, 3404.
- [227] M. D. Vaughan, P. Cleve, V. Robinson, H. S. Diewel, J. F. Honek, *J. Am. Chem. Soc.* **1999**, 121, 8475.
- [228] H. S. Diewel, E. Daub, V. Robinson, J. F. Honek, *Biochemistry* **2001**, 40, 13167.
- [229] M. H. J. Seifert, D. Ksiazek, M. K. Azim, P. Smialowski, N. Budisa, T. A. Holak, *J. Am. Chem. Soc.* **2002**, 124, 7932.
- [230] J. G. Bann, J. Pinkner, S. J. Hultgren, C. Frieden, *Proc. Natl. Acad. Sci. USA* **2002**, 99, 709.
- [231] S. Kothakota, T. L. Mason, D. A. Tirrell, M. J. Fournier, *J. Am. Chem. Soc.* **1995**, 117, 536.
- [232] T. J. Deming, M. J. Fournier, T. L. Mason, D. A. Tirrell, *J. Macromol. Sci. Pure Appl. Chem.* **1997**, 34, 2143.
- [233] J. C. M. van Hest, D. A. Tirrell, *FEBS Lett.* **1998**, 428, 68.

- [234] J. C. M. van Hest, K. L. Kiick, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, *122*, 1282.
- [235] K. L. Kiick, J. C. van Hest, D. A. Tirrell, *Angew. Chem.* **2000**, *112*, 2232; *Angew. Chem. Int. Ed.* **2000**, *39*, 2148.
- [236] K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19.
- [237] A. J. Link, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, *125*, 11164.
- [238] D. Datta, P. Wang, I. S. Carrico, S. L. Mayo, D. A. Tirrell, *J. Am. Chem. Soc.* **2002**, *124*, 5652.
- [239] L. Wang, P. G. Schultz, *Chem. Commun.* **2002**, 1.
- [240] S. Benzer, S. P. Champe, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 1114.
- [241] A. Garen, O. Siddiqi, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 1121.
- [242] J. Normanly, L. G. Kleina, J. M. Masson, J. Abelson, J. H. Miller, *J. Mol. Biol.* **1990**, *213*, 719.
- [243] R. M. Hudziak, F. A. Laski, U. L. RajBhandary, P. A. Sharp, M. R. Capecchi, *Cell* **1982**, *31*, 137.
- [244] H. J. Drabkin, H. J. Park, U. L. Rajbhandary, *Mol. Cell. Biol.* **1996**, *16*, 907.
- [245] F. Sherman in *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression* (Eds.: J. N. Strathern, E. W. Jones, J. R. Broach), Cold Spring Harbor Laboratory, Cold Spring Harbor, **1982**, p. 463.
- [246] D. Kim, J. Johnson, *J. Biol. Chem.* **1988**, *263*, 7316.
- [247] J. Kimble, J. Hodgkin, T. Smith, J. Smith, *Nature* **1982**, *299*, 456.
- [248] S. L. Bolten, P. Powell-Abel, D. A. Fischhoff, R. H. Waterston, *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 6784.
- [249] M. A. Rould, J. J. Perona, D. Soll, T. A. Steitz, *Science* **1989**, *246*, 1135.
- [250] D. R. Liu, T. J. Magliery, P. G. Schultz, *Chem. Biol.* **1997**, *4*, 685.
- [251] D. R. Liu, T. J. Magliery, M. Pasternak, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10092.
- [252] B. P. Doctor, J. A. Mudd, *J. Biol. Chem.* **1963**, *238*, 3677.
- [253] E. F. Whelihan, P. Schimmel, *EMBO J.* **1997**, *16*, 2968.
- [254] D. R. Liu, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4780.
- [255] L. Wang, P. G. Schultz, unpublished results **1998**.
- [256] K. Wakasugi, C. L. Quinn, N. Tao, P. Schimmel, *EMBO J.* **1998**, *17*, 297.
- [257] T. A. Kleeman, D. Wei, K. L. Simpson, E. A. First, *J. Biol. Chem.* **1997**, *272*, 14420.
- [258] L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, *J. Am. Chem. Soc.* **2000**, *122*, 5010.
- [259] B. A. Steer, P. Schimmel, *J. Biol. Chem.* **1999**, *274*, 35601.
- [260] L. Wang, P. G. Schultz, *Chem. Biol.* **2001**, *8*, 883.
- [261] R. Furter, *Protein Sci.* **1998**, *7*, 419.
- [262] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498.
- [263] L. Wang, A. Brock, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 1836.
- [264] L. Wang, P. G. Schultz, unpublished results **2001**.
- [265] L. Wang, Ph.D. Dissertation, University of California at Berkeley, **2002**.
- [266] S. W. Santoro, L. Wang, B. Herberich, D. S. King, P. G. Schultz, *Nat. Biotechnol.* **2002**, *20*, 1044.
- [267] M. Pastrnak, P. G. Schultz, *Bioorg. Med. Chem.* **2001**, *9*, 2373.
- [268] Z. Zhang, L. Wang, A. Brock, P. G. Schultz, *Angew. Chem.* **2002**, *114*, 2964; *Angew. Chem. Int. Ed.* **2002**, *41*, 2840.
- [269] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026.
- [270] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020.
- [271] L. Wang, Z. Zhang, A. Brock, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 56.
- [272] L. Wang, J. Xie, A. A. Deniz, P. G. Schultz, *J. Org. Chem.* **2003**, *68*, 174.
- [273] Z. Zhang, A. C. Smith Brian, L. Wang, A. Brock, C. Cho, P. G. Schultz, *Biochemistry* **2003**, *42*, 6735.
- [274] L. Alfonta, Z. Zhang, S. Uryu, J. A. Loo, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 14662.
- [275] Z. Zhang, J. Gildersleeve, Y.-Y. Yang, R. Xu, J. A. Loo, S. Uryu, C.-H. Wong, P. G. Schultz, *Science* **2004**, *303*, 371; R. Xu, S. R. Hanson, Z. Zhang, P. G. Schultz, C.-H. Wong, *J. Am. Chem. Soc.*, in press.
- [276] J. Turner, P. G. Schultz, unpublished results.
- [277] J. W. Chin, P. G. Schultz, *ChemBioChem* **2002**, *3*, 1135.
- [278] R. E. Galardy, L. C. Craig, M. P. Printz, *Nature New Biol.* **1973**, *242*, 127.
- [279] H. Liu, L. Wang, A. Brock, C.-H. Wong, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 1702.
- [280] A. Deiters, T. A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 11782; A. Deiters, T. A. Cropp, M. Mukherji, D. Summerer, P. G. Schultz, *Bioorg. Med. Chem. Lett.*, in press.
- [281] H. Zeng, P. G. Schultz, unpublished results.
- [282] R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683.
- [283] P. Sears, C. H. Wong, *Science* **2001**, *291*, 2344.
- [284] B. G. Davis, *Chem. Rev.* **2002**, *102*, 579.
- [285] M. Wacker, D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam, S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren, M. Aebi, *Science* **2002**, *298*, 1790.
- [286] S. R. Hamilton, P. Bobrowicz, B. Bobrowicz, R. C. Davidson, H. Li, T. Mitchell, J. H. Nett, S. Rausch, T. A. Stadheim, H. Wischniewski, S. Wildt, T. U. Gerngross, *Science* **2003**, *301*, 1244.
- [287] J. Xie, L. Wang, N. Wu, A. Brock, G. Spraggon, P. G. Schultz, *Nature Biotechnol.* **2004**, *22*, 1297.
- [288] R. Y. Tsien, *Annu. Rev. Biochem.* **1998**, *67*, 509.
- [289] P. G. Schultz, unpublished results.
- [290] N. Wu, P. G. Schultz, unpublished results.
- [291] M. Bose, D. Groff, P. G. Schultz, unpublished results.
- [292] N. Wu, A. Dieters, T. A. Cropp, P. G. Schultz, *J. Am. Chem. Soc.* **2004**, *126*, 14306.
- [293] Y. Zhang, L. Wang, I. A. Wilson, P. G. Schultz, unpublished results, **2003**.
- [294] T. Kobayashi, O. Nureki, R. Ishitani, A. Yaremchuk, M. Tukalo, S. Cusack, K. Sakamoto, S. Yokoyama, *Nat. Struct. Biol.* **2003**, *10*, 425.
- [295] F. J. LaRiviere, A. D. Wolfson, O. C. Uhlenbeck, *Science* **2001**, *294*, 165.
- [296] F. Tian, M. Tsao, P. G. Schultz, *J. Am. Chem. Soc.*, in press.
- [297] R. A. Mehl, J. C. Anderson, S. W. Santoro, L. Wang, A. B. Martin, D. S. King, D. M. Horn, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 935.
- [298] J. F. Curran, M. Yarus, *Science* **1987**, *238*, 1545.
- [299] L. Bossi, J. R. Roth, *Cell* **1981**, *25*, 489.
- [300] M. O'Connor, *Nucleic Acids Res.* **2002**, *30*, 1985.
- [301] T. J. Magliery, J. C. Anderson, P. G. Schultz, *J. Mol. Biol.* **2001**, *307*, 755.
- [302] J. C. Anderson, T. J. Magliery, P. G. Schultz, *Chem. Biol.* **2002**, *9*, 237.
- [303] J. F. Atkins, R. B. Weiss, S. Thompson, R. F. Gesteland, *Annu. Rev. Genet.* **1991**, *25*, 201.
- [304] J. C. Anderson, N. Wu, S. W. Santoro, V. Lakshman, G. Schultz Peter, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7566.
- [305] J. C. Anderson, P. G. Schultz, *Biochemistry* **2003**, *42*, 9598.
- [306] S. W. Santoro, J. C. Anderson, V. Lakshman, P. G. Schultz, *Nucleic Acids Res.* **2003**, *31*, 6700.
- [307] M. Pastrnak, T. J. Magliery, P. G. Schultz, *Helv. Chim. Acta* **2000**, *83*, 2277.
- [308] F. Martin, Dissertation, Université Louis Pasteur (Strasbourg, France), **1995**.
- [309] A. K. Kowal, C. Kohrer, U. L. RajBhandary, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2268.

- [310] G. Galli, H. Hofstetter, M. L. Birnstiel, *Nature* **1981**, 294, 626.
- [311] K. Sakamoto, A. Hayashi, A. Sakamoto, D. Kiga, H. Nakayama, A. Soma, T. Kobayashi, M. Kitabatake, K. Takio, K. Saito, M. Shirouzu, I. Hirao, S. Yokoyama, *Nucleic Acids Res.* **2002**, 30, 4692.
- [312] F. Li, Y. Xiong, J. Wang, H. D. Cho, K. Tomita, A. M. Weiner, T. A. Steitz, *Cell* **2002**, 111, 815.
- [313] M. P. Deutscher, *Prog. Nucleic Acid Res. Mol. Biol.* **1990**, 30, 209.
- [314] S. L. Wolin, A. G. Matera, *Genes Dev.* **1999**, 13, 1.
- [315] G. J. Arts, S. Kuersten, P. Romby, B. Ehresmann, I. W. Mattaj, *EMBO J.* **1998**, 17, 7430.
- [316] H. Edwards, P. Schimmel, *Mol. Cell. Biol.* **1990**, 10, 1633.
- [317] H. Edwards, V. Trezeguet, P. Schimmel, *Proc. Natl. Acad. Sci. USA* **1991**, 88, 1153.
- [318] Y. Kwok, J. T. Wong, *Can. J. Biochem.* **1980**, 58, 213.
- [319] D. Kiga, K. Sakamoto, K. Kodama, T. Kigawa, T. Matsuda, T. Yabuki, M. Shirouzu, Y. Harada, H. Nakayama, K. Takio, Y. Hasegawa, Y. Endo, I. Hirao, S. Yokoyama, *Proc. Natl. Acad. Sci. USA* **2002**, 99, 9715.
- [320] J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang, P. G. Schultz, *Science* **2003**, 301, 964.
- [321] C. Kohrer, J. H. Yoo, M. Bennett, J. Schaack, U. L. RajBhandary, *Chem. Biol.* **2003**, 10, 1095.
- [322] S. L. Monahan, H. A. Lester, D. A. Dougherty, *Chem. Biol.* **2003**, 10, 573.
- [323] T. A. Cropp, P. G. Schultz, unpublished results.
- [324] F. Xu, X. Chen, L. Xin, L. Chen, Y. Jin, D. Wang, *Nucleic Acids Res.* **2001**, 29, 4125.
- [325] X. Yang, J. Lee, E. M. Mahony, P. D. Kwong, R. Wyatt, J. Sodroski, *J. Virol.* **2002**, 76, 4634.
- [326] Z. Zhang, L. Alfonta, F. Tian, B. Bursulaya, S. Uryu, D. S. King, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2004**, 101, 8882.
- [327] A. Deiters, P. G. Schultz, *Bioorg. Med. Chem. Lett.*, in press.
- [328] H. Cho et al., unpublished results.
-