Chapter 1

Non-Canonical Amino Acids in Protein Engineering

The text in this chapter is from two review papers: Link, A. J.; Mock, M. L.; Tirrell, D. A. *Current Opinion in Biotechnology* **2003**, *14*, 603-609 and Link, A. J.; Tirrell, D. A. *Methods* **2005**, *36*, 291-298. Reprinted with the permission of the publishers.

Abstract

Methods for engineering proteins that contain non-canonical amino acids have advanced rapidly in the past few years. Novel amino acids can be introduced into recombinant proteins in either a residue-specific or site-specific fashion. The methods are complementary: residue-specific incorporation allows engineering of the overall physical and chemical behavior of proteins and protein-like macromolecules, whereas site-specific methods allow mechanistic questions to be probed in atomistic detail. Challenges remain in the engineering of the translational apparatus and in the design of schemes that can be used to encode both canonical and non-canonical amino acids.

Introduction

The synthesis of proteins on messenger RNA templates represents the most striking example of the power and versatility of templated synthesis in all of chemistry and biology. To be sure, the processes of DNA replication and transcriptional synthesis of RNA exhibit many of the same intriguing features, but the diversity of amino acid structures and the absence of direct structural complementarity between template and product make protein synthesis unique.

The same mechanisms that ensure the fidelity of protein synthesis limit the diversity of protein structure and function. A broad view of protein design would retain the architectural control of protein biosynthesis while embracing an expanded set of amino acid building blocks. Methods for engineering proteins that contain non-canonical amino acid residues have advanced substantially in the last few years and now constitute powerful and versatile tools (Figure 1) for the re-design of natural proteins and for the *de novo* design of artificial protein-like macromolecules.

This is not a new field. Cowie and Cohen showed in the 1950s that selenomethionine serves–with respect to protein synthesis–as an effective surrogate for methionine,[1,2] and over the past decade this substitution has revolutionized protein crystallography.[3] Several excellent reviews of protein engineering with non-canonical amino acids are available;[4-9] the present discussion addresses developments reported in the last five years as well as those to come.

Residue-Specific Incorporation of Non-Canonical Amino Acids

The genetic code is degenerate: sixty-four triplet codons encode only twenty amino acid building blocks. Of the sixty-four codons, sixty-one encode amino acids and are termed "sense" codons. The remaining three "nonsense" or "stop" codons encode termination signals that halt protein synthesis. An old technology that has enjoyed a recent resurgence involves reassignment of sense codons in the *in vivo* expression of recombinant proteins. Through appropriate engineering of the expression host, a sense codon (or a family of sense codons) may be reassigned to a non-canonical amino acid under the conditions used for protein expression. The net result is the generation of proteins that contains nineteen of the canonical amino acids and one non-canonical amino acid. This method is often termed "residue-specific" incorporation of non-canonical amino acids because all of the codons coding for a given canonical amino acid (residue) are reassigned to the non-canonical analog.

The simplest strategy for the production of proteins containing non-canonical amino acids in *E. coli* relies on the use of auxotrophic expression hosts deficient in the biosynthetic machinery needed to produce one or more of the twenty canonical amino acids. By using chemically defined media, such as M9 or M63, in conjunction with such strains, one is able to control the intracellular pools of the amino acids of interest. The success of this strategy was first demonstrated, as mentioned above, nearly fifty years ago in the pioneering work of Cohen and Cowie,[1,2] wherein essentially complete replacement of the natural amino acid methionine by its analog selenomethionine was achieved in an *E. coli* methionine auxotroph. In these early experiments, selenomethionine was presumably incorporated throughout the proteome. More recently, by using standard

recombinant DNA technologies, it has become possible to over-express a single protein under conditions where the non-canonical amino acid is targeted (though surely not exclusively) to the protein of interest.

The metabolic pathways for amino acid synthesis in *E. coli* are thoroughly understood, and amino acid auxotrophs are readily available and easy to generate. Strains containing transposons in the genes for amino acid biosynthesis may be obtained from public repositories such as the Yale *E. coli* Genetic Stock Center (http://cgsc.biology.yale.edu/). These strains are particularly useful in generating amino acid auxotrophs in new genetic backgrounds because of the possibility of transferring the lesion in the biosynthetic pathway via generalized transduction.

Once a suitable auxotroph is obtained or generated, one must consider design of the expression plasmid. Since most non-canonical amino acids will not support bacterial growth, the expression host is grown initially either in a rich medium or in a medium supplemented with all twenty canonical amino acids, and is later shifted to a medium containing the non-canonical amino acid. Since leaky expression in the initial growth medium will lead to a protein product that does not contain the desired non-canonical amino acid, it is important to control protein expression with a strongly repressed promoter. Phage-derived promoters, such as the T5 and T7 promoters found in commercially-available protein expression vectors such as the pQE series from Qiagen or the pET series from Novagen, offer tight control over protein expression while also allowing high expression levels for preparative-scale production of proteins containing non-canonical amino acids. In a typical residue-specific incorporation experiment, a single colony of the amino acid auxotroph transformed with a plasmid encoding the

protein under study is inoculated into minimal medium (M9) containing all twenty canonical amino acids and the appropriate antibiotics. This culture is allowed to grow to late logarithmic phase to amass enough cellular material for protein synthesis. Cells are then sedimented by centrifugation and washed with isotonic saline. The cycle of centrifugation and washing is repeated to remove canonical amino acids. After an additional centrifugation, the cells are resuspended in M9 medium supplemented with nineteen amino acids and lacking the canonical amino acid to be replaced. To further deplete the intracellular pool of the canonical amino acid, cells are incubated at the expression temperature for a short time (10-30 minutes). Protein expression is then initiated by addition of an appropriate inducer. Variants of this procedure have been used in our laboratory and in others to reassign the sense codons of tryptophan,[10,11] methionine,[12-14] leucine,[15] isoleucine,[16] phenylalanine,[17] and proline[18] in recombinant proteins.

Over-Expression of Aminoacyl-tRNA Synthetases (aaRS)

The fidelity of protein synthesis *in vivo* is controlled in large part by the aminoacyl-tRNA synthetases.[19] Most cells contain a distinct aminoacyl-tRNA synthetase (aaRS) for each of the twenty canonical amino acids. These enzymes perform at least two functions: activation of the amino acid (aa) substrate, and ligation of the amino acid to its cognate tRNA ($tRNA^{aa}$). aaRS thus form a pool of aminoacyl-tRNA ($aa\text{-}tRNA^{aa}$) in the cell that is used for protein synthesis. Successful implementation of sense codon reassignment *in vivo* hinges on the ability of the aaRS to activate the non-canonical amino acids of interest. Studies by Kiick *et al.* demonstrated that the *in vitro* activation rates of a series of methionine analogues by the methionyl-tRNA synthetase (MetRS)

correlated strongly with the ability of these amino acids to support protein synthesis *in vivo*.[13,20] As an example, the unsaturated amino acid homoallylglycine (Hag, **2**) is activated approximately 2,000-fold more slowly than the canonical substrate methionine as judged by *in vitro* kinetics experiments with purified MetRS, but Hag still serves as an effective methionine surrogate in experiments of the kind described above. In contrast, *trans*-crotylglycine (Tcg, **3**), which is activated nearly 5,000 times more slowly than methionine, is unable to support protein synthesis (Figure 2). These experiments indicated that the aaRS activity of the expression host is likely to be a critical controlling factor in determining the success or failure of attempts to incorporate non-canonical amino acids into proteins. To test this hypothesis, a methionine auxotrophic strain of *E. coli* carrying a plasmid-borne copy of MetRS was generated. Whole-cell lysates of this strain activated methionine at rates 5- to 10-fold higher than those characteristic of the unmodified expression host, and Tcg was incorporated efficiently into recombinant proteins expressed in the modified strain. This strain was used to effect incorporation of a panel of methionine analogs that are activated even more slowly than Tcg.[21]

Over-expression of other aaRS has allowed reassignment of sense codons other than that of methionine. For example, over-expression of the leucyl-tRNA synthetase (LeuRS) in a leucine auxotroph led to incorporation of hexafluoroleucine (**6**, Figure 2) into an artificial leucine zipper protein.[22] As observed with methionine analogues such as Tcg, elevation of LeuRS activity is required for incorporation of hexafluoroleucine: leucine auxotrophic cells lacking the plasmid-borne copy of LeuRS do not make proteins containing hexafluoroleucine. A striking example of codon reassignment via manipulation of aaRS activity was realized in studies of the non-canonical amino acid

4,4,4-trifluorovaline (Tfv, **8**).[23] Cells disrupted in the *ilv* locus, which renders them auxotrophic for both isoleucine and valine, were evaluated as expression hosts for the generation of proteins containing Tfv. The wild-type aaRS activity in these cells was insufficient for production of proteins containing Tfv, but incorporation of one of the diastereomers of Tfv (*2S,3R*-Tfv) into the test protein dihydrofolate reductase was readily observed upon depletion of valine and over-expression of the valyl-tRNA synthetase (ValRS). Surprisingly, depletion of isoleucine and over-expression of the isoleucyl-tRNA synthetase (IleRS) in these cells also allowed protein synthesis with *2S,3R*-Tfv. Further analysis by mass spectrometry indicated that depletion of valine and over-expression of ValRS led to incorporation of *2S,3R*-Tfv in response to valine codons while depletion of isoleucine and over-expression of IleRS targeted the non-canonical amino acid to isoleucine codons. Thus the same RNA message could be decoded in two different ways depending on the amino acid pools and the aminoacyl-tRNA synthetase activity of the host cells.

Some practical issues arise when considering strategies to increase the aminoacyl-tRNA synthetase activity in *E. coli*. It is important to drive expression of the aaRS either constitutively or with a promoter orthogonal to the promoter used to express the recombinant protein under study (the "test protein"). Since induction of the test protein occurs in medium supplemented with a non-canonical amino acid, non-orthogonal induction schemes will result in incorporation of the non-canonical amino acid into the aaRS and may render the aaRS inactive. In each of the examples described above, a DNA cassette bearing the aaRS gene under control of its endogenous, constitutive promoter was introduced into the expression host via an appropriate plasmid. Recently,

Conticello and co-workers have described a scheme for orthogonal induction of over-expressed prolyl-tRNA synthetases (ProRS).[24]

Engineering of Aminoacyl-tRNA Synthetases

While the translational activity of a non-canonical amino acid may be greatly enhanced by over-expression of the appropriate wild-type aminoacyl-tRNA synthetase, many analogs require further manipulation of the aaRS activity of the cell. The wealth of structural data on the aaRSs[19] provides some insight into the origins of the high fidelity characteristic of protein synthesis. The first level of control of the aaRS occurs at a highly shape-selective binding pocket for the amino acid, which is the site of amino acid activation. Despite the fact that phenylalanine and tyrosine differ only by a single hydroxyl group, the tyrosyl-tRNA synthetase activates tyrosine 10^5-fold more rapidly than phenylalanine.[25] A second level of control involves hydrolysis ("editing") of improperly aminoacylated tRNAs. Editing has been most thoroughly explored in studies of the isoleucyl-tRNA synthetase,[26,27] but is also characteristic of the valyl-,[28] leucyl-,[29] threonyl-,[30] and prolyl-tRNA synthetases.[31] A recent report also describes the editing of tyrosine by the phenylalanyl-tRNA synthetase.[32] The ability of a non-canonical amino acid to infiltrate the genetic code *in vivo* is contingent upon its ability to pass through both of these checkpoints. The availability of high-resolution crystal structures for most of the aaRS allows engineering and redesign of both the synthetic sites (amino acid binding pockets) and the editing sites of these enzymes for the ultimate goal of introducing non-canonical amino acids into proteins.

The specificity of the synthetic site of the *E. coli* phenylalanyl-tRNA synthetase (PheRS) has been relaxed by introduction of a single alanine-to-glycine (A294G) mutation in the phenylalanine binding pocket. Introduction of a plasmid-borne copy of the mutant PheRS into a phenylalanine auxotrophic *E. coli* strain allowed incorporation of the halogenated phenylalanine analogs *para*-chlorophenylalanine[33] and *para*-bromophenylalanine (Figure 3, **10**).[34] Further studies revealed that a set of *para*-substituted phenylalanine analogs containing varied chemical functionality were efficiently incorporated into recombinant proteins by this mutant PheRS.[35] Recently, incorporation of the photoreactive amino acid benzofuranylalanine (**11**) into proteins with this mutant PheRS has also been reported.[36] Additional engineering of PheRS was undertaken with the aid of a computational protein design algorithm to generate a mutant PheRS capable of activating and incorporating *para*-acetylphenylalanine (*p*-AcPhe, **12**). Efficient incorporation of *p*-AcPhe was realized with a doubly-mutated PheRS (T251G, A294G) as predicted by the design algorithm.[37] The net result of these synthetic-site mutations is to create space in the phenylalanine binding pocket in order to accommodate the sterically demanding *para*-substituted analogs. Rational redesign of the synthetic site of prolyl-tRNA synthetase (ProRS) has also been reported. Introduction of a single cysteine-to-glycine (C443G) mutation in the proline binding pocket of ProRS allows incorporation of the ring-expanded proline analog *2S*-pipecolic acid (**14**) into a recombinant elastin-like protein.[24]

Disruption of the editing activity of the leucyl-tRNA synthetase (LeuRS) has led to incorporation of a large set of non-canonical amino acids in response to leucine sense codons. Mursinna and Martinis noted that mutation of a highly-conserved threonine

residue in the editing domain of *E. coli* LeuRS to alanine (T252A) led to greatly enhanced editing of leucine from its cognate tRNA.[38] The authors suggested that the role of T252 is to prevent leucine from entering the editing site, and that substitution of threonine by the smaller alanine residue accounts for the increased editing observed. Replacement of T252 with a larger amino acid was expected to impair or abolish the editing function of LeuRS, and this prediction was borne out by experiment.[39] Tang and Tirrell utilized a plasmid-borne LeuRS variant carrying the T252Y mutation to incorporate several unsaturated, non-canonical amino acids in response to leucine codons.[40]

The promise of aminoacyl-tRNA synthetase engineering for the reassignment of sense codons is clearly demonstrated by the examples described above. Simple changes in the cell's aminoacylation machinery, coupled with judicious control of the amino acid pools of the host, can lead to altered interpretations of the genetic code. Future efforts in aaRS redesign may be aided by implementation of combinatorial mutagenesis approaches analogous to those utilized by the Schultz group in the engineering of aaRS for non-canonical amino acid incorporation at nonsense codons *in vivo.*[41]

Examples of Protein Engineering by Sense Codon Reassignment

The study of sense codon reassignment *in vivo* is of fundamental interest from a biological standpoint because of the insights gained in understanding the malleability of the genetic code. Furthermore, the products of sense codon reassignment experiments, proteins containing multiple copies of non-canonical amino acids, have a host of practical applications. Arguably the most useful and certainly the most prevalent use of sense

codon reassignment in *E. coli* is for the production of proteins containing selenomethionine for crystallographic studies.[3] Several other non-canonical amino acids that can be introduced by sense codon reassignment may also prove useful for multi-wavelength anomalous diffraction studies. The electron-rich halogen atoms found in *para*-bromophenylalanine and *para*-iodophenylalanine[35] make these amino acids attractive candidates for such strategies. The incorporation of the tryptophan isostere β-selenolo[3,2-*b*]pyrrolylalanine into the proteins barstar and annexin V allowed determination of the structures of these proteins.[11] Introduction of this amino acid in response to the tryptophan codon did not affect the structure of either protein.

The unique fluorescence properties of tryptophan make reassignment of the tryptophan codon a particularly promising avenue for the engineering of new spectral properties into proteins. Perhaps the most striking example of this capability was demonstrated by Bae *et al.*[42] Upon replacement of the critical tryptophan residue of enhanced cyan fluorescent protein (eCFP) with the non-canonical amino acid 4-aminotryptophan, a novel fluorescent protein was generated. The emission maximum of this protein, termed gold fluorescent protein (GdFP), is red-shifted nearly 70 nm in comparison to the parent eCFP, with an observed Stokes shift of more than 100 nm. Furthermore, the molar extinction coefficient of GdFP is nearly equivalent to that of eCFP. It has also been demonstrated that 4-aminotryptophan (4-NH_2-Trp) effectively functions as a pH sensor when incorporated into the small protein barstar.[43] While the emission maximum (λ_{max}) of wild-type barstar is nearly pH-independent, λ_{max} of 4-NH_2-Trp substituted barstar shifts from ca. 350 nm at low pH to ca. 330 nm at pH 11.

Non-canonical amino acids bearing fluorine atoms are attractive substrates for sense codon reassignment because of the special chemical character of fluorine. Fluorinated variants of the aliphatic amino acids are nearly isosteric to the parent amino acids, and most are recognized by the essential elements of the protein synthesis machinery. Introduction of fluorinated leucine analogs into leucine zipper peptides leads to retention of helical secondary structure and enhanced stability with respect to thermal and chemical denaturation.[15] Moreover, the observed stability scales with the number of fluorine atoms; coiled-coil proteins containing hexafluoroleucine are more stable than similar proteins containing trifluoroleucine.[22]

One of the major advantages of codon reassignment is the ability to generate proteins with chemical functionality beyond that encoded in the natural proteome (Figure 4). Of particular interest are chemical groups that endow proteins with novel, so-called "bio-orthogonal," reactivity. The aryl halides introduced into proteins upon incorporation of *para*-bromophenylalanine or *para*-iodophenylalanine[34,35] enable palladium-catalyzed coupling procedures such as the Heck and Sonogashira reactions. Similarly, the acetyl group, which can be introduced via *para*-acetylphenylalanine, reacts rapidly and selectively with hydrazides to generate hydrazones.[37] An especially interesting functional group that has been introduced into proteins by sense codon reassignment is the azide group. Azides react via several useful pathways. The aryl azide moiety, which may be introduced into proteins via *para*-azidophenylalanine (**15**, Figure 4), decomposes upon exposure to ultraviolet light, generating highly reactive intermediates that lead to efficient protein crosslinking. Proteins containing the methionine surrogate azidohomoalanine (AHA, **16**) have been selectively tagged via the Staudinger ligation,[14]

and proteins and cell surfaces labeled with azide-bearing amino acids have been covalently modified via the copper-catalyzed [3+2] azide-alkyne ligation.[44,45]

Site-Specific Incorporation of Non-Canonical Amino Acids

Though the global replacement of a given residue in a protein by a novel amino acid is an important technology with far-reaching applications, one can readily envision other applications in which it is desirable to replace only a single residue while retaining access to all twenty natural amino acids. The first examples of this "site-specific" incorporation of non-canonical amino acids in an *in vitro* translation system originated from the laboratories of Schultz[46] and Chamberlin[47] in the late 1980s. These approaches rely on translational read-through of an amber nonsense codon (TAG) by a misacylated suppressor tRNA. To ensure high-level incorporation of the non-natural amino acid, the suppressor tRNA must not be charged with a natural amino acid by any of the aminoacyl-tRNA synthetases present in the *in vitro* translation system.

Misacylated tRNAs have also been used *in vivo* to effect non-natural amino acid mutagenesis of the nicotinic acetylcholine receptor (nAChR) in intact *Xenopus* oocytes.[48] In these experiments, the aminoacyl-tRNA and mRNA are introduced to the oocyte via microinjection. Modified tyrosines have been incorporated into nAChR to generate receptors that have constitutive activity at low pH.[49] Hydroxyacids have been introduced by similar methods.[50] These studies implicated the modified backbone hydrogen bonding of a proline residue (mimicked by the ester linkages formed upon hydroxyacid incorporation) as critical for the voltage gating behavior of nAChR. A recent report

details procedures for expression of receptors containing non-natural amino acids in mammalian cells, including CHO cells and neurons.[51]

Roberts and co-workers have shown that the non-natural amino acid biocytin can be incorporated into puromycin-linked mRNA-peptide hybrids produced in rabbit reticulocyte *in vitro* translation systems.[52] Thus this powerful method for selection of functional peptides need not be limited by the chemistries of the canonical amino acids.

Site-specific incorporation of non-natural amino acids can be realized by using codons other than the amber nonsense codon. Two different four-base codons were used by the Sisido laboratory to incorporate a fluorophore-quencher pair into streptavidin.[53] Streptavidin modified in this way still binds biotin, and efficient quenching of the fluorescent amino acid is observed. A four-base codon and an amber codon were used to incorporate amino acids constituting a FRET pair into a dihydrofolate reductase fusion protein containing an HIV-1 protease cleavage site;[54] the FRET pair allows monitoring of protein cleavage. The Hecht group has designed several amino acid analogs that enable chemospecific backbone cleavage. The protease inhibitor ecotin containing allylglycine was cleaved with iodine,[55] while site-specific cleavage of a DNA polymerase containing aminooxyacetic acid was realized by treatment with zinc and acetic acid.[56]

While the approaches described above are powerful, there are instances in which *in vitro* production of peptides or proteins is impractical. In 1998, Furter implemented a successful *in vivo* strategy for site-specific incorporation of non-canonical amino acids into recombinant proteins.[57] In addition to a yeast suppressor tRNA, a yeast aminoacyl-tRNA synthetase was introduced into an *E. coli* expression host to generate a heterologous "21st" tRNA/synthetase pair. The engineered host was shown to incorporate

p-fluorophenylalanine in response to amber codons with protein yields on the order of ten milligrams per liter.

Schultz and co-workers have advanced this strategy in important ways. Selection schemes[41] have allowed evolution of aminoacyl-tRNA synthetases that charge amino acids with useful functional groups, including acetyl[58] and benzophenone moieties.[59] The acetyl group has been used to label proteins *in vivo*[60] and to afford glycosylated proteins,[61] while proteins containing the benzophenone side chain have been efficiently crosslinked both *in vitro* and *in vivo*.[62] An aminoacyl-tRNA synthetase was evolved to allow incorporation of *p*-aminophenylalanine (*p*-NH_2-Phe) and combined with the biosynthetic genes for *p*-NH_2-Phe, generating an organism that can both synthesize and incorporate twenty-one amino acids.[63] Non-canonical amino acids have also been used to engineer novel variants of GFP.[64] Such variants exhibit absorption and emission maxima that differ by as much as 75 nm from wild-type GFP, though all suffer decreases in quantum yield compared to the wild-type protein.

All of the studies discussed above were performed in *E. coli*, but it has been recently demonstrated that many of the same manipulations may be carried out in *S. cerevisiae*,[65] allowing incorporation of non-canonical amino acids into the proteins of eukaryotes. Incorporation into eukaryotic (CHO) cells has also been accomplished by stable transfection with a plasmid encoding an orthogonal aminoacyl-tRNA synthetase followed by transient transfection with a plasmid-encoded tandemly-repeated suppressor tRNA.[66]

New Methodology

The feasibility of incorporating non-standard amino acids into proteins, both globally and site-specifically, is now firmly established, and the implications of this technology are beginning to emerge. Nevertheless, substantial methodological challenges remain, with respect to both the "hardware" and the "software" of protein synthesis. Hardware challenges include the design and engineering of new aminoacyl-tRNA synthetases and synthetase/tRNA pairs, as well as the development of ribosomes with altered decoding properties. Software considerations address the need for new coding schemes that accommodate both standard and non-standard amino acids.

Synthetases and Synthetase/tRNA Pairs

As mentioned earlier, Schultz and co-workers have reported powerful methods for the selection of aminoacyl-tRNA synthetases that activate non-standard amino acids more efficiently than their standard counterparts. In much of their work, they employ a TyrRS/$tRNA^{Tyr}_{CUA}$ pair derived from *Methanococcus jannaschii*, which is characterized by little cross-charging with the translational machinery of *E. coli*. A library of synthetase mutants is generated and subjected to positive selection based on suppression of an amber codon located at a permissive position in a plasmid-borne gene encoding chloramphenicol acetyl transferase. Mutants that show good suppression in media supplemented with the non-canonical amino acid of interest are then subjected to negative selection in media supplemented only with the canonical amino acids, in order to remove synthetase variants that are effective in charging one or more of the standard

substrates. Negative selections include expression of a toxic protein (e.g., barnase) or cell-sorting on the basis of differential expression of GFP.[67]

Computational methods have also been applied to synthetase design. Datta and co-workers have reported a computationally-designed variant of the *E. coli* PheRS that effects efficient activation of *p*-acetylphenylalanine,[37] and the Goddard laboratory has shown that computational protocols can be used to identify some of the critical sequence changes in the TyrRS mutants selected via the methods described above.[68]

Modified Ribosomes

Although the aminoacyl-tRNA synthetases constitute the most stringent check on the fidelity of protein synthesis *in vivo*, the aminoacyl-tRNAs must also be processed by the ribosome. Early *in vitro* experiments with tRNAs misacylated with D-amino acids showed no evidence of analog incorporation. In an intriguing set of recent experiments, Hecht and co-workers have shown that ribosomes carrying mutations in the peptidyl transferase center and in helix 89 can be used to effect nonsense suppression with D-amino acids at levels of 20-25%.[69] The implications of these results for *in vivo* protein synthesis remain to be established.

Coding Schemes

There is as yet no ideal coding scheme for non-canonical amino acids. Residue-specific replacement precludes incorporation of one of the canonical amino acids, and nonsense suppression and frameshift suppression are characterized by modest

efficiencies.[70] The Sisido and Schultz laboratories have extended suppression methods to five-base and opal codons, but neither method seems to offer improved efficiency.[71-73]

An alternative coding scheme has been reported by Kwon and co-workers, who used a mutant yeast $tRNA^{Phe}$ to re-assign one of the two degenerate phenylalanine codons (UUU) to the non-canonical amino acid 2-naphthylalanine (Figure 5).[74] The method–a form of missense suppression–is complementary to the nonsense and frameshift strategies; re-assignment is still incomplete, but decoding "errors" lead to incorporation of the canonical amino acid rather than to premature chain termination. Frankel and Roberts have described a selection scheme for identifying sense codons that are especially susceptible to suppression.[75]

Conclusions

The prospects for engineering proteins that contain non-canonical amino acids have improved dramatically in the past few years. Engineering of the translational apparatus, especially the aminoacyl-tRNAs and their cognate synthetases, has enabled the introduction of a remarkably diverse set of new chemical functions into natural and artificial proteins. The coding problem remains, however; nonsense and frameshift suppression suffer from low efficiencies, and preliminary attempts to break the degeneracy of the genetic code have yielded less than complete discrimination between degenerate codons. Continuing advances in this field offer the prospect of engineered proteins strikingly different from those prepared to date.

Acknowledgements

Work on codon reassignment at Caltech has been supported by grants from the National Science Foundation (DMR 0080065), the National Institutes of Health (GM 62523), the U.S. Army Research Office (DAAD19-03-D-0004), and the Beckman Institute Endowment.

References

(1) Cohen, G. N.; Cowie, D. B. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* **1957**, *244*, 680-683.

(2) Cowie, D. B.; Cohen, G. N. *Biochimica Et Biophysica Acta* **1957**, *26*, 252-261.

(3) Hendrickson, W. A.; Horton, J. R.; Lemaster, D. M. *Embo J.* **1990**, *9*, 1665-1672.

(4) Dougherty, D. A. *Curr. Opin. Chem. Biol.* **2000**, *4*, 645-652.

(5) Hohsaka, T.; Sisido, M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 809-815.

(6) Link, A. J.; Mock, M. L.; Tirrell, D. A. *Current Opinion in Biotechnology* **2003**, *14*, 603-609.

(7) Budisa, N. *Angew. Chem.-Int. Edit.* **2004**, *43*, 6426-6463.

(8) Wang, L.; Schultz, P. G. *Angew. Chem.-Int. Edit.* **2005**, *44*, 34-66.

(9) Link, A. J.; Tirrell, D. A. *Methods* **2005**, *36*, 291-298.

(10) Budisa, N.; Alefelder, S.; Bae, J. H.; Golbik, R.; Minks, C.; Huber, R.; Moroder, L. *Protein Sci.* **2001**, *10*, 1281-1292.

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

(12) van Hest, J. C. M.; Tirrell, D. A. *FEBS Lett.* **1998**, *428*, 68-70.

(13) Kiick, K. L.; Tirrell, D. A. *Tetrahedron* **2000**, *56*, 9487-9493.

(14) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 19-24.

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

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

(17) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471-5475.

(18) Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L. *Angew. Chem.-Int. Edit.* **2001**, *40*, 923-925.

(19) Ibba, M.; Soll, D. *Annual Review of Biochemistry* **2000**, *69*, 617-650.

(20) Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A. *Angew. Chem.-Int. Edit.* **2000**, *39*, 2148-2152.

(21) Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Lett.* **2001**, *502*, 25-30.

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

(23) Wang, P.; Fichera, A.; Kumar, K.; Tirrell, D. A. *Angew. Chem.-Int. Edit.* **2004**, *43*, 3664-3666.

(24) Kim, W. Y.; George, A.; Evans, M.; Conticello, V. P. *Chembiochem* **2004**, *5*, 928-936.

(25) Fersht, A. R.; Shindler, J. S.; Tsui, W. C. *Biochemistry* **1980**, *19*, 5520-5524.

(26) Fersht, A. R. *Biochemistry* **1977**, *16*, 1025-1030.

(27) Schmidt, E.; Schimmel, P. *Science* **1994**, *264*, 265-267.

(28) Fersht, A. R.; Kaethner, M. M. *Biochemistry* **1976**, *15*, 3342-3346.

(29) Englisch, S.; Englisch, U.; Vonderhaar, F.; Cramer, F. *Nucleic Acids Res.* **1986**, *14*, 7529-7539.

(30) Dock-Bregeon, A. C.; Sankaranarayanan, R.; Romby, P.; Caillet, J.; Springer, M.; Rees, B.; Francklyn, C. S.; Ehresmann, C.; Moras, D. *Cell* **2000**, *103*, 877-884.

(31) Beuning, P. J.; Musier-Forsyth, K. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8916-8920.

(32) Roy, H.; Ling, J.; Irnov, M.; Ibba, M. *The EMBO Journal* **2004**, *23*, 4639-4648.

(33) Ibba, M.; Hennecke, H. *FEBS Lett.* **1995**, *364*, 272-275.

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

(35) Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *Chembiochem* **2002**, *3*, 235-237.

(36) Bentin, T.; Hamzavi, R.; Salomonsson, J.; Roy, H.; Ibba, M.; Nielsen, P. E. *J. Biol. Chem.* **2004**, *279*, 19839-19845.

(37) Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652-5653.

(38) Mursinna, R. S.; Lincecum, T. L.; Martinis, S. A. *Biochemistry* **2001**, *40*, 5376-5381.

(39) Mursinna, R. S.; Martinis, S. A. *J. Am. Chem. Soc.* **2002**, *124*, 7286-7287.

(40) Tang, Y.; Tirrell, D. A. *Biochemistry* **2002**, *41*, 10635-10645.

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

(42) Bae, J. H.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M. H. J.; Azim, M. K.; Kim, J. S.; Zumbusch, A.; Holak, T. A.; Moroder, L.; Huber, R.; Budisa, N. *J. Mol. Biol.* **2003**, *328*, 1071-1081.

(43) Budisa, N.; Rubini, M.; Bae, J. H.; Weyher, E.; Wenger, W.; Golbik, R.; Huber, R.; Moroder, L. *Angew. Chem.-Int. Edit.* **2002**, *41*, 4066-4069.

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

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

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

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

(48) Nowak, M. W.; Kearney, P. C.; Sampson, J. R.; Saks, M. E.; Labarca, C. G.; Silverman, S. K.; Zhong, W.; Thorson, J.; Abelson, J. N.; Davidson, N.; Schultz, P. G.; Dougherty, D. A.; Lester, H. A. *Science* **1995**, *268*, 439-442.

(49) Petersson, E. J.; Choi, A.; Dahan, D. S.; Lester, H. A.; Dougherty, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 12662-12663.

(50) England, P. M.; Zhang, Y. N.; Dougherty, D. A.; Lester, H. A. *Cell* **1999**, *96*, 89-98.

(51) Monahan, S. L.; Lester, H. A.; Dougherty, D. A. *Chem. Biol.* **2003**, *10*, 573-580.

(52) Li, S. W.; Millward, S.; Roberts, R. *J. Am. Chem. Soc.* **2002**, *124*, 9972-9973.

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

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

(55) Wang, B. X.; Brown, K. C.; Lodder, M.; Craik, C. S.; Hecht, S. M. *Biochemistry* **2002**, *41*, 2805-2813.

(56) Eisenhauer, B. M.; Hecht, S. M. *Biochemistry* **2002**, *41*, 11472-11478.

(57) Furter, R. *Protein Sci.* **1998**, *7*, 419-426.

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

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

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

(61) Liu, H. T.; Wang, L.; Brock, A.; Wong, C. H.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 1702-1703.

(62) Chin, J. W.; Schultz, P. G. *Chembiochem* **2002**, *3*, 1135-1137.

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

(64) Wang, L.; Xie, J. M.; Deniz, A. A.; Schultz, P. G. *J. Org. Chem.* **2003**, *68*, 174-176.

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

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

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

(68) Zhang, D. Q.; Vaidehi, N.; Goddard, W. A.; Danzer, J. F.; Debe, D. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 6579-6584.

(69) Dedkova, L. M.; Fahmi, N. E.; Golovine, S. Y.; Hecht, S. M. *J. Am. Chem. Soc.* **2003**, *125*, 6616-6617.

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

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

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

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

(74) Kwon, I.; Kirshenbaum, K.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 7512-7513.

(75) Frankel, A.; Roberts, R. W. *Rna-a Publication of the Rna Society* **2003**, *9*, 780-786.

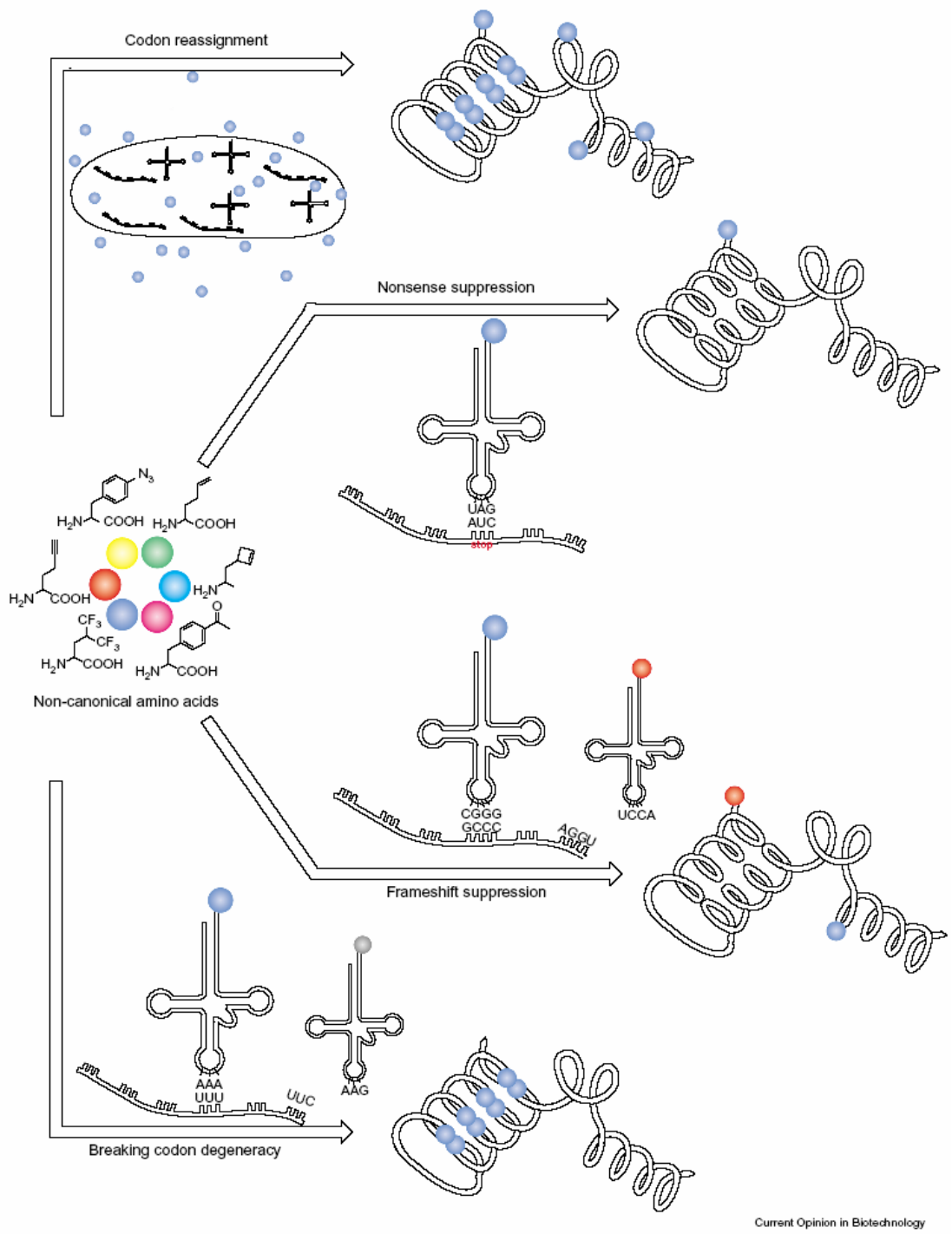

Figure 1: Caption on following page.

Figure 1: Complementary strategies have been developed for the incorporation of non-canonical amino acids into proteins. Global replacement of a natural amino acid with an analog is possible through codon reassignment. Nonsense and frameshift supression strategies permit site-specific incorporation of non-canonical amino acids. Systems that break the degeneracy of the genetic code take advantage of the existence of multiple codons for a single amino acid, assigning an analog to one while incorporating the natural amino acid at others.

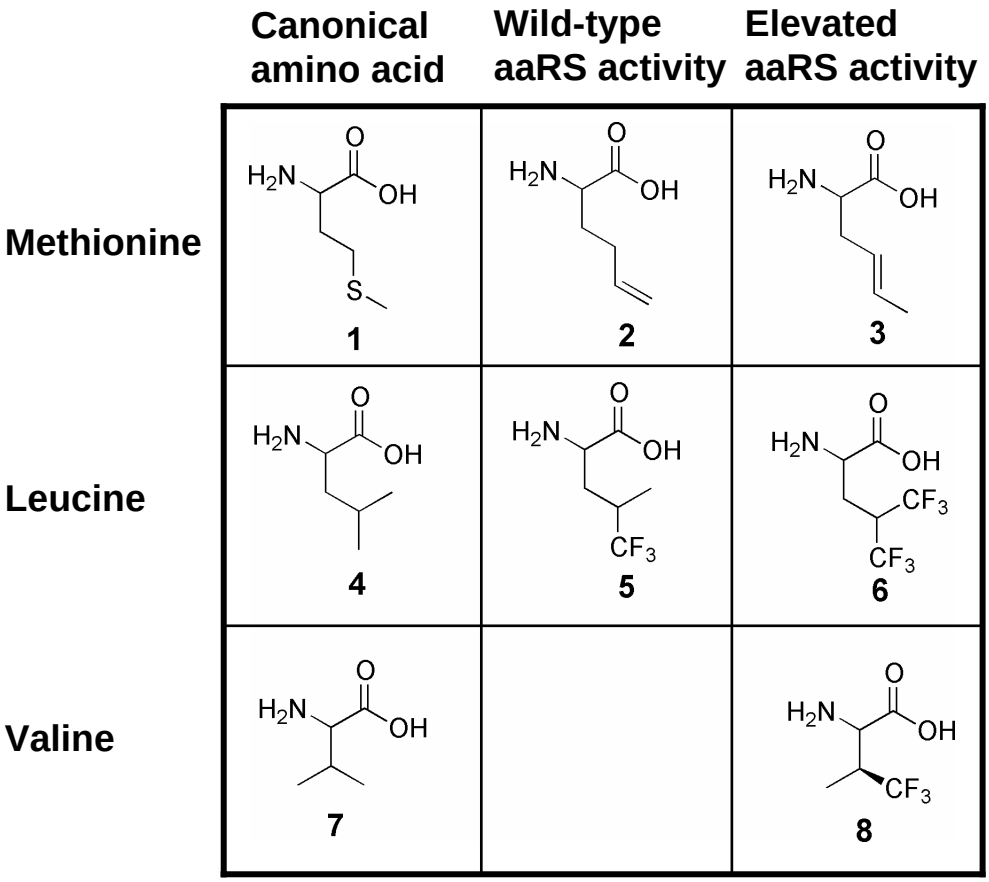

Figure 2: Sense codon reassignment by increasing aminoacyl-tRNA synthetase (aaRS) activity. Some non-canonical amino acids such as homoallylglycine (**2**) and trifluoroleucine (**5**) can replace their canonical counterparts in *E. coli* with wild-type aaRS activity. Elevating the appropriate aaRS activity of the cell results in the ability to incorporate analogues such as *trans*-crotylglycine (**3**), hexafluoroleucine (**6**), and trifluorovaline (**8**).

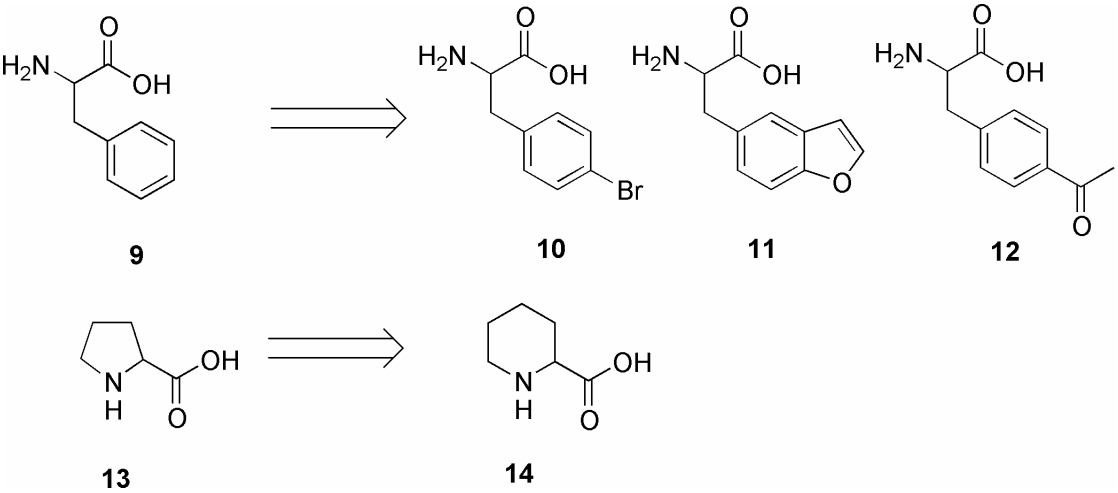

Figure 3: Sense codon reassignment with engineered aminoacyl-tRNA synthetases. Introduction of a single mutation to the synthetic site of phenylalanyl-tRNA synthetase (PheRS) allows replacement of phenylalanine (**9**) with *para*-bromophenylalanine (**10**), benzofuranylalanine (**11**), and several other analogs in recombinant proteins. A doubly-mutated PheRS permits incorporation of *para*-acetylphenylalanine (**12**) into proteins. A synthetic-site mutant of prolyl-tRNA synthetase (ProRS) allows reassignment of proline (**13**) codons to pipecolic acid (**14**).

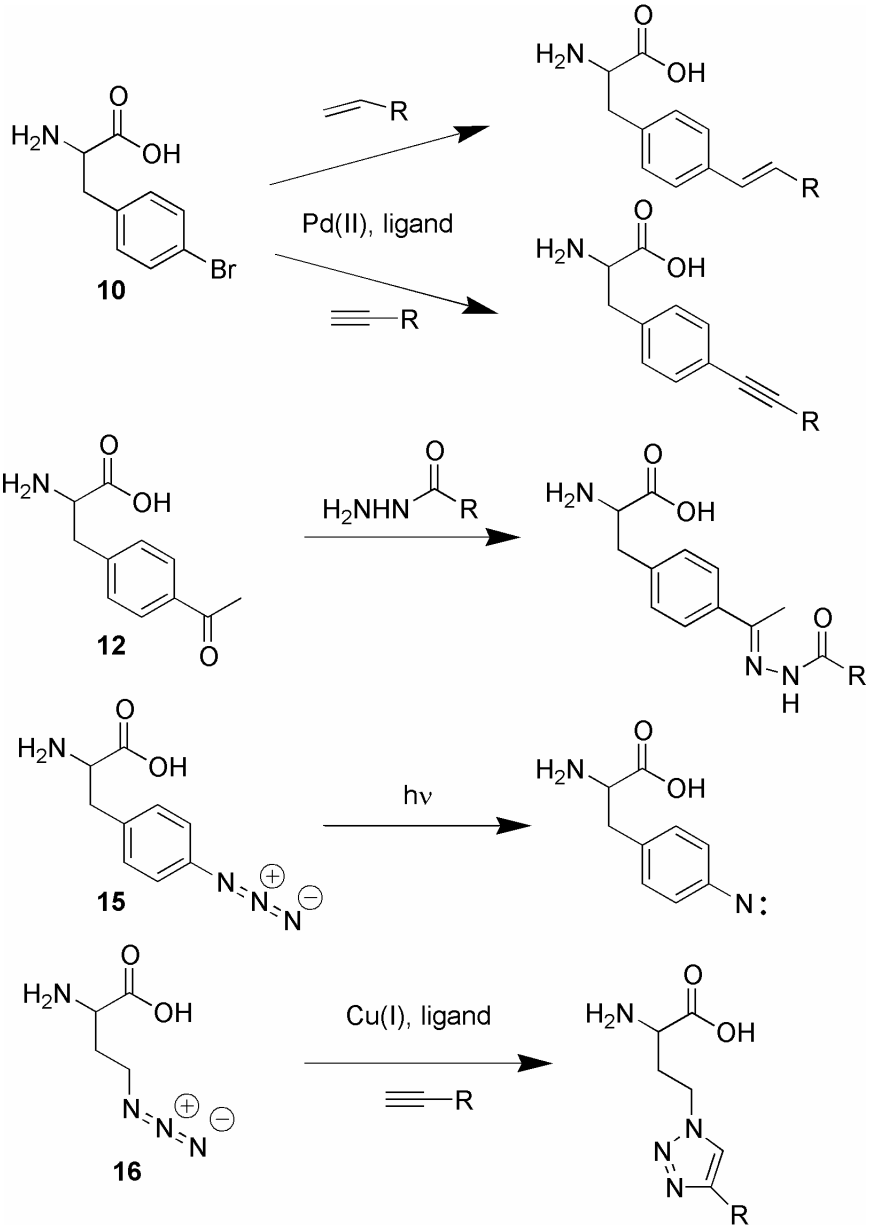

Figure 4: Useful chemistries introduced to proteins by non-canonical amino acids. *para*-Bromophenylalanine can participate in Pd-catalyzed Heck and Sonogashira reactions. *para*-Acetylphenylalanine reacts with hydrazides to form hydrazones. Photodecomposition of *para*-azidophenylalanine and protein crosslinking can be achieved upon irradiation. Proteins and cell surfaces bearing azidohomoalanine have been selectively tagged via Cu-catalyzed azide-alkyne ligation.

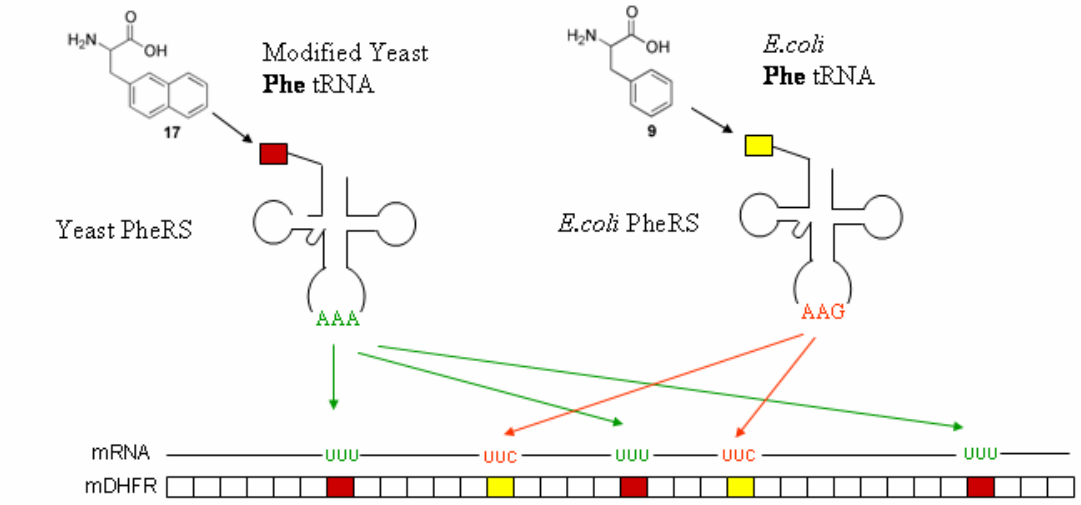

Figure 5: Breaking the degeneracy of the genetic code *in vivo*. *E. coli* equipped with a 21st tRNA/aminoacyl-tRNA synthetase pair can utilize two different amino acids in response to the two degenerate phenylalanine codons. *E. coli* phenylalanyl-tRNA is charged with phenylalanine by the chromosomal PheRS while a heterologous yeast phenylalanyl-tRNA carrying a modified anticodon sequence is charged with naphthylalanine (**17**) by a designed yeast PheRS mutant. Phenylalanine is introduced into proteins primarily in response to UUC codons while naphthylalanine is introduced primarily in response to UUU codons.