CHAPTER 1

Expanding the Scope of Protein Science and Engineering with Noncanonical Amino Acids

This chapter has been adapted from a review in preparation by James A. Van Deventer, Ying Y. Liu, Jeremiah A. Johnson, and David A. Tirrell.

Abstract

Noncanonical amino acids (ncAAs) have the potential to greatly expand the chemical functionalities available within proteins. Methodologies for the genetic encoding and incorporation of ncAAs into proteins using the protein synthesis machinery from living organisms are now quite common, and applications of proteins containing ncAAs are emerging in a variety of fields. In this review, we highlight the most widely used methodologies for biosynthetic incorporation of ncAAs into proteins. We then discuss applications of ncAA incorporation to areas that include protein structure determination, protein and organism evolution, modulation of the immune system, and proteomics. Numerous successes in these and other fields suggest that biosynthetic ncAA incorporation will continue to be a valuable tool for biological science and engineering in the future. The ease of incorporation and large functional toolkit available with ncAAs will continue to enable researchers from many disciplines to tailor the use of ncAAs to their specific needs.

Introduction

Proteins perform an astonishingly broad range of functions in biological systems, yet they are usually synthesized from no more than twenty amino acid monomers. While the templated synthesis of polypeptide chains provides access to an enormous sequence space (1), and posttranslational modifications significantly expand the available functional space (2), the chemical diversity of natural amino acid side chains is rather limited from a chemist's perspective. Thus, researchers have focused much effort on the development of new methodologies for incorporation of amino acids that are not normally specified by the genetic code, or noncanonical amino acids (ncAAs), into peptides and proteins.

Methods for the incorporation of ncAAs into peptides and proteins are either chemical, biosynthetic, or some combination of the two. In the area of chemical synthesis, solid-phase peptide synthesis has enabled the routine incorporation of virtually any suitably protected amino acid into synthetic peptides (3, 4). Small peptides can then be coupled together to give full-length synthetic proteins via techniques such as native chemical ligation (5, 6). The combination of chemical ligation strategies and biological protein production methods in the form of techniques such as expressed protein ligation (7-9) further enhances researchers' abilities to incorporate new chemical functionalities into proteins. These nontemplated chemical approaches for incorporation of ncAAs into proteins continue to grow in importance to the scientific community.

Despite the successes of chemical methods, the natural protein biosynthetic machinery of living organisms remains unparalleled in its ability to produce complex, genetically templated polypeptide chains in large quantities. For many years scientists have imagined that a reworking of the genetic code or the protein biosynthetic machinery could enable the biological production of proteins containing a nearly endless variety of ncAAs (10) and perhaps even other non-amino acid monomer structures (11). Sequence-specific, monodisperse polymers, which are currently challenging or impossible to synthesize using traditional synthetic techniques, would be expected to display a wide range of chemical and physical properties. Methodological developments over more than fifty years have facilitated the biosynthetic incorporation of an ever-increasing number of ncAAs and other monomers into proteins. These techniques have begun to allow scientists to investigate a vast range of subjects in biology, biological chemistry, and engineering.

In this review, we aim to highlight some of the ways in which scientists have employed biosynthetically produced proteins containing ncAAs to study and manipulate proteins and biological systems. The section "Biosynthetic Methodologies for the Incorporation of Genetically Encoded Noncanonical Amino Acids into Proteins" will summarize some important historical and methodological underpinnings of ncAA incorporation into proteins using enzymatic machinery found in living organisms. In the "Applications" section, we will discuss a number of research topics that have been impacted substantially by the use of ncAAs. The subjects highlighted in this review are meant to emphasize the creative approaches enabled by manipulation of the genetic code and are not intended to cover all areas of application. Finally, in the "Outlook" section, we will briefly speculate about potential future uses of ncAAs.

Biosynthetic Methodologies for the Incorporation of Genetically Encoded Noncanonical Amino Acids into Proteins

Methodologies for biosynthetic incorporation of ncAAs into proteins can be divided into "residue-specific" and "site-specific" strategies. An overview of these strategies is presented in figure 1.1. In a "residue-specific" strategy, a single canonical amino acid is replaced by a ncAA wherever the mRNA encoding the protein specifies the canonical amino acid. Figure 1.1a illustrates the resulting "reprogramming" of protein translation. In contrast, "site-specific" strategies (figure 1.1b) involve the replacement of a single amino acid with a ncAA at a desired location within the polypeptide chain. Usually, this approach utilizes suppression of a stop, or nonsense, codon with an appropriately charged suppressor tRNA, although some additional approaches have also been explored.

Background: key advances that enabled modern ncAA incorporation methodologies. Characterization of the components of the biosynthetic protein translation apparatus has provided insights into how the genetic code functions and how the code can be manipulated to perform unnatural functions. Among the key results in this area are Chapeville and coworkers' "Raney Nickel" experiments (12), which provided strong support for Francis Crick's adaptor hypothesis (13). The adaptor hypothesis proposed that small nucleic acids (now called transfer RNAs, or tRNAs) could provide a means by which a sequence of nucleic acids could be translated into a corresponding polypeptide sequence, with the specific pairing of amino acids and tRNAs ensuring the fidelity of the genetic code. Chapeville and coworkers were able to show for the first time that chemical manipulation of aminoacyl-tRNAs can alter the way in which an RNA transcript is decoded. The researchers treated cysteinyl-$tRNA^{Cys}$ with Raney Nickel to convert cysteinyl-$tRNA^{Cys}$ to alanyl-$tRNA^{Cys}$ and compared the results of polypeptide synthesis after supplementing ribosomal preparations with cysteinyl- or alanyl-$tRNA^{Cys}$ and an RNA transcript coding for cysteine. Polypeptide synthesis was supported in each case, and polypeptides produced in reactions containing cysteinyl-$tRNA^{Cys}$ and alanyl-$tRNA^{Cys}$ contained only cysteine and alanine, respectively. These results showed conclusively that misacylated tRNAs remain substrates of the ribosome and that they support protein synthesis. The ability of the ribosome to use misacylated tRNAs in translation implies that ncAA incorporation into protein is possible in theory. If a ncAA can be attached to a particular tRNA, the protein translation apparatus will likely accept the "misacylated" tRNA as a substrate and place the ncAA into a growing polypeptide chain in response to a particular codon.

The "Raney Nickel" experiments also demonstrated the feasibility of performing chemistry directly on tRNA molecules, and many groups have since extended the types of chemistry that can be performed on tRNAs. In the 1970s, experiments performed in the Menninger group demonstrated that (i) Lys-tRNA^{Lys} could be acetylated specifically at the ε-amino group of lysine, and (ii) AcLys-tRNA^{Lys} could be incorporated into proteins in a cell-free protein synthesis system in response to lysine codons (14). These experiments were the first to show that chemically synthesized aminoacyl-tRNAs containing a ncAA could be utilized by the biosynthetic protein synthesis machinery. Hecht and coworkers expanded on the work of Menninger and coworkers by developing a general route to the chemical acylation of tRNAs (15, 16). Their method involved chemical synthesis of the acylated dinucleotide pCpA, the generation of tRNA missing its last two RNA bases, and enzymatic attachment of the acylated pCpA to the truncated tRNA. Further improvements to chemical acylation techniques were elucidated in the 1980s (17, 18). These synthetic methods have proven crucial in the generation of tRNAs bearing a wide variety of ncAAs for use with site-specific ncAA incorporation into proteins.

The identification and detailed characterization of aminoacyl-tRNA synthetases (aaRSs), the enzymes responsible for joining together amino acids and tRNAs in living organisms, have also been crucial for enabling incorporation of ncAAs into proteins. Early work (19, 20) by Berg and others established that aaRSs are responsible for catalyzing the attachment of tRNA and amino acids using the free energy gained from hydrolysis of adenosine triphosphate (ATP) to adenosine monophosphate (AMP) in a two-step process:

$$\text{Amino Acid (AA)} + \text{ATP} \xrightleftharpoons{\text{aaRS}} \text{AA-AMP} \quad \text{(Activation)} \qquad (1.1)$$

$$\text{AA-AMP} + \text{tRNA}^{AA} \xrightleftharpoons{\text{aaRS}} \text{AA-tRNA}^{AA} + \text{AMP} \quad \text{(Aminoacylation)} \qquad (1.2)$$

Characterizations of the enzymatic activities of aaRSs have established that these enzymes are extremely important for maintaining the fidelity of protein translation (21) and that ncAAs can serve as kinetically efficient substrates of aaRSs (22). These findings imply that understanding how to manipulate the enzymatic activities of aaRSs can facilitate the specific, enzymatic linkage of ncAAs and tRNAs.

Additional characterizations of aaRSs and tRNAs laid the groundwork for adding amino acids to the genetic code by adding new translational components to living cells. One of the important ways in which aaRSs maintain the fidelity of the genetic code is by selectively recognizing and aminoacylating their cognate tRNAs. For example, the methionyl-tRNA synthetase (MetRS) must conjugate methionine specifically to $tRNA^{Met}$ without aminoacylating any other tRNA (21). This specific recognition is critical for maintaining genetic code fidelity, but the mechanisms of specificity are not conserved from species to species. As early as 1963, researchers started to realize that the aaRSs from different species recognize their cognate tRNAs by distinct mechanisms. Doctor and Mudd first observed this phenomenon when they discovered that an aaRS from one species cannot always aminoacylate the corresponding tRNA from another species (23). Smith and coworkers proposed that the structural basis for these observations is aaRS recognition of structural features in tRNA molecules that differ from species to species (24). This hypothesis has since been investigated thoroughly using structural and biochemical experiments, resulting in the detailed characterization of how cognate aaRSs and tRNAs are properly paired in different organisms (25). Another early breakthrough that eventually facilitated adding new amino acids to the genetic code was the discovery and characterization of suppressor mutations in *E. coli* (26-28). These mutations, which result

in the conversion of a "stop" message into a message coding for an amino acid, were found to mutate the tRNA anticodon sequence from nucleotides specifying a sense codon into nucleotides specifying a termination codon (24). These findings suggested that using codons normally reserved for the termination of protein synthesis to instead specify an amino acid could be a way to add an amino acid to the genetic code. The combined exploitation of the species specificity of aaRS-tRNA recognition and suppressor tRNAs eventually enabled researchers to dictate the coding of a 21st amino acid, first in vitro and later in living cells and organisms (see below).

Residue-specific incorporation strategies. ***Conceptual advances.*** Even before the components of the translation apparatus were characterized, scientists recognized that ncAAs could be incorporated into the proteins of multi- and single-celled organisms. These earliest examples employed ncAAs that were isosteric or structurally similar to canonical amino acids. During the 1950s and 1960s, ncAAs were employed extensively in studies that involved microorganisms; this work was reviewed thoroughly in 1962 (29). These early experiments, especially those performed in the laboratory of Georges Cohen, laid the foundation for later advances in the field. Cohen and colleagues were the first to use a medium replacement strategy that we now refer to as a "medium shift." They also demonstrated the utility of auxotrophic strains for achieving high levels of ncAA incorporation into proteins. Furthermore, the Cohen laboratory recognized that, "these analogs [and surrogates] become powerful tools for the study of: (a) the specificity of the protein-forming mechanism and (b) the variation of enzyme activity and affinity that occurs with increasing numbers of incorporated analog molecules" (30). These

observations are still relevant today. We now know with some level of detail that incorporation of a particular amino acid into proteins depends on the ability of the protein translation apparatus of a particular cell or organism to tolerate noncanonical substrates. “Variation of enzyme activity” by the incorporation of ncAAs is still studied today and now forms part of a larger effort to perturb protein structure and function with ncAAs.

After the relatively sparse application of ncAAs to scientific problems in the 1970s and 1980s (31, 32), researchers began seriously revisiting the idea of incorporating ncAAs into proteins residue-specifically in 1990. Hendrickson and coworkers demonstrated the complete replacement of methionine by selenomethionine (**1**, scheme 1.1) in thioredoxin in 1990 (33). Their approach involved growing methionine-auxotrophic bacteria in medium containing a limited supply of methionine but ample amounts of selenomethionine. As the bacteria grow, they use up the supply of methionine, forcing incorporation of selenomethionine in place of methionine. Induction of protein expression after the depletion of methionine in the medium enabled a high level of selenomethionine incorporation into proteins. This method for incorporating selenomethionine into proteins has been adapted for use with a number of other ncAAs by the Budisa laboratory (22). The Tirrell laboratory and others have utilized an alternative approach involving extensive washing of bacterial cells in between growth in a medium containing all canonical amino acids and protein expression in medium containing ncAAs (34). With these protocols, scientists have developed methods for the incorporation of a large set of ncAAs into proteins residue-specifically, primarily by engineering *E. coli* expression hosts and aminoacyl-tRNA synthetases. These contemporary residue-specific incorporation techniques will be described in the following subsection.

Experimental approaches. In most cases, efficient aminoacylation of a tRNA with a noncanonical L-amino acid is sufficient to enable the residue-specific replacement of a canonical amino acid with the ncAA of interest. Thus, the genetic code can be intentionally "reinterpreted" to code for one or more ncAAs by controlling what substrates get attached to specific tRNAs. The conceptual approach is depicted in figure 1.1a, and specific strategies for intentional tRNA misaminoacylation are shown in figure 1.2.

The simplest approach to ncAA incorporation involves replacement of a canonical amino acid with a close structural analog in *E. coli*. With an appropriate analog, a cell strain auxotrophic in the amino acid to be replaced, and a technique for depleting the canonical amino acid from the medium prior to expression of proteins of interest, incorporation of a ncAA at genetically encoded positions can be essentially quantitative (if the ncAA is recognized efficiently by the protein translational machinery of the host). Figure 1.2a illustrates a typical procedure for the incorporation of homopropargylglycine (Hpg, **2**, scheme 1.1) in place of methionine in proteins produced by *E. coli*. Methionine-auxotrophic *E. coli* cells are first grown in minimal media containing all twenty canonical amino acids. Upon reaching a sufficient optical density, cells are washed to remove methionine from the medium and resuspended in the expression medium, which usually contains high concentrations of ncAA (in this case, Hpg). Alternatively, cells can be grown in medium that contains both methionine in small quantities and a ncAA in large quantities. By the time cells have reached an optical density sufficient for protein expression, the concentration of canonical amino acid has been depleted, resulting in the same effect as removal of the canonical amino acid through washing (this is sometimes called selective pressure incorporation). After the medium is depleted of the canonical amino acid to be

replaced and supplemented with large amounts of the ncAA, all proteins synthesized within the cells will contain ncAAs. If a particular protein is to be studied, standard expression, purification, and characterization techniques can be employed to isolate the desired protein. On the other hand, if a proteome-wide response to a particular stress or signal is to be studied, the set of newly synthesized proteins can be isolated and studied using approaches to be described in the "Applications: Proteomics" subsection.

Researchers have employed several variations of medium shifts and selective pressure incorporation in recent years. The techniques have been extended to mammalian cell lines for monitoring the production of newly synthesized proteins (35), and attempts have also been made to extend the procedure to yeast expression systems (36, 37). Recently, reports have indicated that two or three canonical amino acids can be replaced by ncAAs simultaneously in *E. coli*, allowing for more drastic reinterpretations of the genetic code (38, 39).

While the medium shift procedure can enable the incorporation of a number of ncAAs into proteins, further expansion of the number of translationally active amino acids available requires additional engineering of expression hosts. To date, most work in the field has focused on altering aminoacyl-tRNA synthetase (aaRS) activity in *E. coli* to enable additional ncAAs to serve as protein building blocks. These approaches, which are performed in combination with medium shifts, are summarized in figure 1.2b–d. The kinetics of amino acid activation appear to dictate the translational activity of most ncAAs (40). When amino acid activation kinetics are slow, ncAAs cannot be joined to tRNA molecules at a high enough rate to support protein synthesis. Increasing the concentrations of an aaRS by outfitting an *E. coli* expression strain with a plasmid-borne copy of an

endogenous aaRS raises the aminoacylation activity of the host strain, enabling some rather poor aaRS substrates to support protein synthesis in the host. Figure 1.2b illustrates this strategy for the leucine analog hexafluoroleucine (Hfl, **3**) (41). LeuRS overexpression in *E. coli* cells auxotrophic in leucine production enables quantitative replacement of Leu by Hfl.

Sometimes simple overexpression of an aaRS is not enough to overcome poor aaRS activity toward noncanonical substrates. However, the judicious mutation of an aaRS can greatly improve the activation kinetics of nonstandard substrates and thereby lead to more ncAAs that can be quantitatively incorporated into proteins. Computational and screening approaches have proven useful in this regard (42-44). For example, as shown in figure 1.2c, the bulky methionine analog azidonorleucine (Anl, **4**) has only very weak translational activity in an *E. coli* host overexpressing wild-type methionyl-tRNA synthetase (MetRS) (45). However, when *E. coli* is outfitted with one of several MetRS active site mutants identified in high-throughput screens for translational activity in the presence of Anl, the bulky azide amino acid can quantitatively replace Met (46, 47). Kinetic characterizations of the mutant MetRSs revealed vastly improved activation of Anl.

Occasionally, efficient activation of an amino acid substrate is not sufficient to enable the incorporation of a particular ncAA into proteins because of proofreading mechanisms in aaRSs. Several aaRSs have a second active site responsible for ensuring that tRNAs are aminoacylated with their cognate amino acids; these editing active sites can discriminate between amino acids that differ by as little as a methyl group (e.g., valine versus isoleucine) and cleave incorrectly aminoacylated substrates (21). However, attenuation of aaRS editing activity can substantially increase the promiscuity of the enzyme, enabling aminoacylation of a greater pool of substrates. Figure 1.2d portrays such

an approach with leucyl-tRNA synthetase (LeuRS) and the substrate norleucine (Nrl, **5**). Mutation of a critical threonine to a much bulkier tyrosine in the editing active site of LeuRS greatly impairs the editing function of the enzyme, leading to high translational activity of amino acids known to be activated by LeuRS (48-51).

In some ways, manipulating the editing and aminoacylation activities (through overexpression and/or mutation) of aaRSs represent complementary approaches to ncAA incorporation. These techniques for manipulating aaRSs for residue-specific incorporation were all developed in *E. coli*, but work from the Hang laboratory suggests that mutant aaRSs from *E. coli* can be employed in other microbes such as *Salmonella typhimurium* (52). Site-specific incorporation approaches using mutated aaRS (discussed below) have been very successful in mammalian cells, suggesting that residue-specific ncAA incorporation approaches in mammalian cells with mutated aaRSs are also worth investigating. While most aaRS manipulations are performed via the introduction of additional copies of aaRS genes on plasmids, genomic manipulations of aaRSs also hold some promise, as evidenced by the recent report of Abdeljabbar et al. (53). Manipulations of aaRS activities through aaRS overexpression, mutation, and genomic manipulation and the application of these techniques to additional organisms will continue to increase the power of methods for the residue-specific incorporation of ncAAs into proteins.

Combining residue-specific incorporation and in vitro protein synthesis. The key requirement for achieving the global replacement of a canonical amino acid in a protein is usually the efficient aminoacylation of the appropriate tRNA. Several groups have approached this aminoacylation problem in conjunction with cell-free protein synthesis,

forming a powerful combination because of the precise control over the components of the translational machinery possible in cell-free environments. Cell-free protein synthesis with chemically acylated tRNAs (54), aaRS-catalyzed tRNA aminoacylation (55), and ribozyme-catalyzed tRNA aminoacylation (56, 57) have all been demonstrated with residue-specific ncAA incorporation. Several groups have also shown that sense codons can be reassigned to residues containing noncanonical backbones such as α-hydroxy acids, *N*-methyl amino acids, and *N*-substituted glycines (poly *N*-substituted glycines are also called peptoids) (58-63), and the Hecht group has made some progress in engineering *E. coli* ribosomes to accept D-amino acids as translationally active substrates (64, 65). Furthermore, powerful nonribosomal methods to synthesize genetically encoded small molecules and polymers are also emerging (11, 66). All of these approaches enable the production of genetically encoded polymers with compositions that are substantially different from those of naturally occurring proteins, and these polymers may have properties that are vastly different from naturally occurring biopolymers. Future applications of these genetically encoded polymers could prove to be very powerful.

Site-specific incorporation approaches. ***Conceptual advances.*** The first examples of site-specific ncAA incorporations into proteins combined the use of chemical aminoacylation techniques, stop codon suppression, and species-specific recognition of tRNA molecules using cell-free protein synthesis (67, 68). In 1989, the groups of Chamberlain and Schultz each reported strategies for the in vitro incorporation of a single ncAA into polypeptide chains in response to amber codons contained within genes coding for proteins of interest. A key component in each of these systems was the combination of an in vitro translation

system from one species and a suppressor tRNA molecule from another species that was not recognized by the aaRSs of the in vitro translation system. The inability of the aaRSs of the in vitro translation system to recognize the tRNA (chemically acylated with the ncAA of interest) makes the tRNA orthogonal to the translation system, a recurring concept in site-specific incorporation of ncAAs into proteins.

Although powerful and quite general, chemical aminoacylation techniques used with in vitro suppression are limited by the amount of the acylated suppressor tRNA that can be generated in a somewhat technically demanding process. For this reason, researchers in several laboratories initiated research aimed at the development of stop codon (or nonsense) suppression techniques for the incorporation of ncAAs into proteins inside living cells or organisms. The move from in vitro to in vivo suppression systems required the development of additional orthogonal components to ensure both the fidelity of the genetic code and the fidelity of ncAA incorporation at specified locations (69-71). Figure 1.3 illustrates the three key criteria that must be met in order to establish the orthogonality of additional translational components for nonsense suppression in living cells. First, the suppressor tRNA to be added to the cell (figure 1.3a) must not be a substrate for any of the wild-type aaRSs already present in the cell in order to ensure that only the ncAA of interest is used to decode nonsense codons. Next, the aaRS to be added (figure 1.3b) must specifically recognize both its suppressor tRNA and ncAA substrates. Finally, the ncAA (figure 1.3c) to be added to the genetic code must not be a substrate for any of the wild-type aaRSs in the cell. Failure to meet these strict criteria may result in ncAA incorporation in response to sense codons, or canonical amino acid incorporation in

response to stop codons; both of these situations reduce the fidelity of protein translation. A properly functioning orthogonal pair is illustrated in figure 1.3d.

This challenging orthogonality problem was solved partially by a number of groups before an integrated solution was reported. First, the RajBhandary group reported a system in which an orthogonal tRNA-aaRS pair was required for in vivo synthesis of full-length genes containing an amber codon (69). In this work, *E. coli* $tRNA^{Gln}$ and GlnRS were adapted for use in mammalian cell lines as an orthogonal tRNA-aaRS suppressor pair. Differences in tRNA recognition between *E. coli* and mammalian cells enabled *E. coli* GlnRS to selectively aminoacylate *E. coli* $tRNA^{Gln}$ with glutamine and suppress an amber codon in a CAT reporter gene in multiple mammalian cell lines. In 1997, the Schultz group reported the first attempt to evolve an orthogonal tRNA-aaRS pair in *E. coli* (71). They described the rational design of an *E. coli* $tRNA^{Gln}$ variant and directed evolution of a GlnRS mutant with improved selectivity for the $tRNA^{Gln}$ variant compared to wild-type $tRNA^{Gln}$ in *E. coli*. Though they did not find a GlnRS variant that could recognize the mutant $tRNA^{Gln}$ better than the wild-type $tRNA^{Gln}$, their work demonstrated the feasibility of engineering recognition patterns in an aaRS-tRNA pair. Efforts to use an *S. cerevisiae* tRNA-aaRS pair in *E. coli* were also undertaken during this time period (72). While the tRNA-aaRS pair was found to be orthogonal in *E. coli*, no suitable GlnRS mutant capable of selectively charging a noncanonical glutamine analog was identified. Furter demonstrated the first functioning suppression system able to incorporate a ncAA in response to an amber codon in *E. coli* (70). This system utilized a yeast phenylalanine tRNA-aaRS pair to incorporate *p*-fluorophenylalanine (**6**) into proteins site-specifically in response to an amber codon. Although this system was site-specific with respect to the

incorporation of the ncAA, the system fell short of being fully orthogonal because the yeast PheRS still recognized phenylalanine (Phe) and incorporated Phe in place of **6** in approximately thirty percent of the model proteins expressed, and small amounts of **6** were found at other Phe positions within the protein, also.

In the late 1990s and early 2000s, researchers developed additional translationally active tRNA-aaRS pairs by exploiting kingdom-specific tRNA-aaRS recognition elements. While early efforts to establish almost completely orthogonal pairs in mammalian cells were achieved with rational design alone (73, 74), most work in *E. coli* employed directed evolution techniques to improve the orthogonality of existing tRNA-aaRS pairs imported from other organisms (74, 75). The efforts of Schultz and coworkers were particularly important in this regard. Wang and Schultz established a general selection system for isolating tRNA-aaRS pairs orthogonal to the translational machinery of *E. coli* (75). This selection system was used with a nearly orthogonal *Methanococcus jannaschii* $tRNA^{Tyr}$-TyrRS pair to further improve its suppression of amber stop codons in *E. coli*. The Schultz group was also the first to establish high-throughput selection and screening methods for isolating aaRSs with altered amino acid specificities (76). Researchers have performed site-specific incorporation of several dozen ncAAs into proteins produce in *E. coli* via combination of these methods for the generation of orthogonal tRNA-aaRS pairs and aaRS-amino acid pairs. Early work by the group of Yokoyama demonstrated that the "orthogonal pair" strategy could be applied in mammalian cells (77) without the use of evolutionary methods. Additional engineering work has enabled the development of site-specific incorporation methods in yeast (78) and improvement of strategies applicable to mammalian cells (79, 80). These techniques will be described in more detail below.

Experimental approaches. Site-specific incorporation of ncAAs into proteins requires the use of specialized translational components in conjunction with codon suppression in order to "add" an amino acid to the genetic code. Figure 1.4 outlines various ways to achieve this goal; most of these approaches involve a suppressor tRNA that is aminoacylated with a ncAA of choice. This tRNA, which uses a "nonstandard" codon, is appropriately decoded during protein synthesis in either an in vitro or an in vivo translation system.

Chemical acylation provides researchers with the most general strategy for incorporating ncAAs site-specifically into proteins. If an appropriately designed, orthogonal suppressor tRNA is available for a given translation system, it can be chemically acylated with a very broad range of ncAAs. When coupled with a gene containing the codon to be suppressed, nearly any protein can be synthesized containing the ncAA of choice at a specific site. Figure 1.4b depicts this process in an in vitro translation system. The in vitro translation system, gene expression, and protein purification are essentially the same as in systems lacking the chemically acylated suppressor tRNA. The protein yields achievable by using chemically acylated tRNAs in an in vitro translation system are usually in the microgram range. However, a number of scientific questions can be thoroughly studied even with a small amount of protein containing noncanonical amino acids (81). Extensions of chemical acylation to cellular systems have also proven to be quite fruitful. As an example of one elegant approach, the Dougherty and Lester groups have studied ion channels and other membrane proteins in *Xenopus* oocytes at length using chemical acylation techniques. In their system, depicted in figure 1.4c, an mRNA containing a nonsense codon at the site specified for ncAA incorporation and an orthogonal suppressor tRNA acylated with a ncAA of choice are injected into the oocyte.

Electrophysiology and other techniques can be used to interrogate channel function and will be discussed below in the "Applications: Membrane proteins" subsection.

The enzymatic, site-specific incorporation of ncAAs into proteins in living cells or organisms is technically simpler than chemical acylation-based methods if appropriate genetic components are available. Enzymatic approaches in living systems require an orthogonal aaRS-tRNA pair that can be expressed in the host cells or organism of interest. This pair, and the ncAA to be incorporated in response to the "nonsense" codon (or other nonstandard codon), should meet the strict requirements for orthogonality discussed above. The ncAA should also be able to access the cytoplasm of the expression host by means of passive or active cellular transport mechanisms. If these conditions are met, site-specific methodologies become quite powerful. Figure 1.4d depicts the incorporation process in *E. coli*. A plasmid-based system encoding constitutively expressed orthogonal tRNA and aaRS genes is transformed into an appropriate strain along with an inducible gene that codes for the appropriate "nonsense" or other specialized codon. For strategies related to the development of orthogonal tRNA-aaRS pairs the reader is referred to several recent reviews (82-87). The resulting system can then be treated essentially like any bacterial expression system. After cell growth in minimal (76) or rich medium (88), the ncAA is added to the medium and protein expression is induced. Cells are harvested after induction, and the protein of interest can then be isolated using standard purification methods. One of the primary advantages of this method over residue-specific methodologies is that no medium shift is required before expression; the genetic code is manipulated by adding translational machinery to cells rather than by manipulation of the extracellular environment. However, cellular suppression-based incorporation techniques can sometimes

suffer from low protein yields. Incomplete suppression in these systems is very common and leads to a reduced amount of full-length protein compared to similarly expressed genes encoded entirely by sense codons. Typical yields reported in the literature are in the 1–10 mg/L range, although improved expression systems have recently been described (88-93).

Extensions to the basic strategy of employing an orthogonal tRNA-aaRS pair have expanded approaches to incorporating ncAAs into proteins site-specifically. Methods for evolving orthogonal tRNA-aaRS pairs in *Saccharomyces cerevisiae* have been developed and applied to site-specific incorporation in this eukaryotic model organism (78). Subsequent systems have improved on early work by using genes coding for orthogonal tRNAs that incorporate the A and B box elements required for high-level transcription in yeast (94, 95). Recent reports have demonstrated that orthogonal tRNA-aaRS pairs can also be introduced into the organisms *Pichia pastoris* and *Mycobacterium tuberculosis* and used to incorporate ncAAs site-specifically into proteins (96, 97). Similar methodological extensions have been made to mammalian expression systems. While orthogonal tRNA-aaRS pairs have not been directly evolved in mammalian systems, they have been imported and used for nonsense suppression successfully (77, 98, 99). Again, improvement of tRNA expression has facilitated the development of more tRNA-aaRS pairs for incorporation of noncanonical amino acids into proteins in mammalian cells (79, 80).

Researchers have investigated many additional strategies for improving or expanding the scope of site-specific ncAA incorporation into proteins produced in vitro or in living cells or organisms. Extensive work to improve the types of ncAAs that can be incorporated in response to nonsense codons have focused on the development of new aaRS variants, including those based on the recently discovered pyrrolysyl-tRNA

synthetase (100-103) and by employing aaRSs possessing expanded or altered editing capabilities (104-106). A systematic approach to the creation of mutually orthogonal aaRSs has been reported (107), and this methodology may prove especially useful if the genetic code is expanded to include large numbers of ncAAs in addition to the twenty canonical amino acids. Both in vitro and in vivo attempts to use four-base codons to incorporate ncAAs into proteins have proven somewhat successful (108-111), as have attempts to incorporate either two of the same or multiple, chemically distinct ncAAs into a single protein (109, 112-114). However, the limitations of organism survival place constraints on the manipulations that can be made to the protein biosynthesis machinery. The Chin laboratory has recently developed a system in which a large portion of the protein biosynthesis machinery has been relieved of its requirements to support cellular viability (115). Using the same orthogonality concept described above, a set of mRNA-ribosome pairs that are separate from the wild-type ribosome and mRNA pools were isolated. By freeing the ribosome from its usual responsibilities of supporting all cellular protein synthesis, researchers have successfully evolved ribosome variants that enable more facile expansion of the genetic code. The most notable of these variants is a ribosome capable of efficiently decoding quadruplet codons (116). Building on a previously successful improvement of suppression efficiency in orthogonal ribosomes (117), this quadruplet-decoding ribosome is capable of incorporating multiple ncAAs into a single protein with very high efficiency: one amino acid is introduced in response to a four-base codon, and a second amino acid is incorporated in response to an amber codon. Future work in engineering the ribosome appears to be a very promising approach for improving efforts to site-specifically incorporate ncAAs into proteins.

Applications

Scientists have been extremely successful in inventing and refining techniques for genetically encoding ncAAs and incorporating them into proteins. However, developing incorporation strategies does not always answer scientific questions, except possibly in areas directly related to the study of aminoacyl-tRNA synthetases and the protein translation apparatus. In this section of the review, we will describe approaches to protein science and engineering that have either been significantly impacted or that could be significantly impacted in the future by the employment of ncAAs. We will focus on select recent areas of study and refer the reader to earlier reviews (29, 31, 32) and other recent reviews (22, 84-87, 118-121) for additional uses of ncAAs. Scientists have developed applications of genetically encoded ncAAs covering an extremely broad range of topics.

Protein crystallography. The determination of protein crystal structures using X-ray crystallography is an extremely important part of molecular biology and biochemistry. While reflection data can be collected from any high-quality protein crystal, the lack of heavy atoms in naturally occurring proteins prevents researchers from learning phase information from these samples without additional information. This dilemma, known to crystallographers as the phase problem, has been solved using a number of approaches over the years (122, 123). In the past twenty years, one of the methods of choice for resolving the phase problem has become the direct incorporation of heavy atoms into the structure of proteins and the application of multiwavelength anomalous dispersion (MAD). This technique can be used on proteins by residue-specifically replacing of one of the canonical amino acids in the molecule with an amino acid containing a heavy atom. In 1990,

Hendrickson and coworkers rediscovered Cowie and Cohen's 1957 work in which quantitative replacement of methionine by selenomethionine (SeMet, **1**) was reported (30). SeMet is nearly perfect for X-ray crystallography applications because it is almost identical in shape to Met and possesses a heavy atom appropriate for MAD phasing techniques. Using selective pressure incorporation, Hendrickson et al. were able to demonstrate the near-quantitative replacement of Met by SeMet in T4 thioredoxin. Furthermore, reflection data acquired with the SeMet-containing proteins indicated, "…that MAD phasing of prospective selenomethionyl proteins should be readily feasible since diffraction ratios are in excess of those that have proven adequate for related problems" (33). Indeed, the application of MAD to SeMet-containing proteins was used successfully twice in 1990 to solve the structures of previously uncharacterized proteins (124, 125). Extensions to proteins expressed in mammalian (126, 127) and baculovirus expression systems (128, 129), and other improvements to expression procedures over the years have led to the establishment of MAD phasing on SeMet-containing proteins as a method of choice in protein crystallography (130-133). MAD has become an indispensible technique for modern protein crystallography.

Several additional ncAAs containing heavy atoms have been introduced into proteins with the aim of improving crystallographic techniques. Residues **7–10** have been incorporated residue-specifically into proteins with high replacement of the corresponding canonical amino acids. Telluromethionine (**7**) has been shown to enable the multiple isomorphous replacement (MIR) phasing strategy using laboratory radiation sources (134-136), while selenocysteine (**8**) can be used in place of (137) or in combination with **1** to enable MAD phasing (138). The feasibility of using residues **9** and **10** for structure

determination has also been demonstrated (139, 140). Iodinated ncAAs also appear to be promising for aiding structure determination efforts. Proteins containing site-specifically incorporated *p*-iodophenylalanine (**11**) and 3-iodotyrosine (**12**) have been crystallized and used to produce reflection data on laboratory radiation sources (141, 142). Iodine-generated single-wavelength anomalous dispersion (SAD) enabled the data to be properly phased, and high-quality structures have been determined from these crystals. Although these emerging approaches to incorporating heavy atoms into proteins have yet to impact X-ray crystallography in the same way as selenomethionine, they may aid future structure determination efforts, especially those that can be accomplished in the laboratory rather than at a synchrotron.

Nuclear magnetic resonance spectroscopy. Like X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy is a crucial tool for the structural and biophysical characterization of proteins. Using isotopic labeling techniques enabling the incorporation of ^{13}C, ^{15}N, and/or ^{2}H, into all or part of proteins of interest enables a vast range of studies to be performed on proteins of increasing sizes (143-145). These experiments, however, can be quite complex due to the number and similarity of chemical shifts present in a single experiment. Judicious incorporation of ncAAs residue- and site-specifically into protein samples can eliminate some of the complexities inherent in NMR studies of proteins.

The lack of fluorine in most biological systems and the ready availability of fluorinated amino acid analogs makes ^{19}F NMR an attractive option for simplifying some protein NMR experiments (146, 147). Methods for the residue-specific biosynthetic incorporation of fluorinated ncAAs into proteins have been used extensively in the

generation of ^{19}F NMR samples (147). Fluorinated aromatic residues **6**, **13–20** and difluoromethionine (**21**) are examples of some of the probes that have been used in experiments (147-150). Aided by the extreme environmental sensitivity of fluorine, researchers have used proteins containing these residues to explore several protein properties including structure, folding, and ligand binding. Rule, Pratt, Ho, and coworkers performed a number of structural studies on D-lactate dehydrogenase (LDH) protein samples labeled with fluorinated aromatic amino acids in the late 1980s and 1990s (151-153). Changes in the ^{19}F NMR spectra of the protein in the presence of ligand or spin-labeled lipid molecules gave the researchers a better idea of which residues of the protein were sensitive to ligand and which residues were involved in membrane contacts.

Fluorine NMR has been used in conjunction with folding studies on a number of proteins. Ropson and Frieden were the first to report the use of ^{19}F NMR to study protein folding (154). They conducted equilibrium folding experiments on the intestinal fatty acid–binding protein labeled with 6-fluorotryptophan (**15**). Interestingly, their spectra showed the appearance of some peaks only at intermediate urea concentrations, indicative of a previously unrecognized folding intermediate in this biomolecule. Since 1992, this general technique has been used with numerous proteins in the context of protein folding and ligand binding, suggesting that the method is useful for studying proteins with a variety of structures (150, 155-162). An important extension of ^{19}F NMR on proteins has been the combination of this sensitive spectroscopic method with stopped-flow experimental techniques. Hoeltzli and Frieden studied the kinetics of *E. coli* dihydrofolate reductase (DHFR) in the first report of stopped-flow NMR experiments with fluorinated proteins (163). In combination with complementary fluorescence and circular dichroism refolding

experiments, the researchers determined that DHFR appears to unfold through a pathway involving an intermediate state in which individual amino acid side chains gain a large amount of mobility while the majority of secondary structural elements remain intact.

Although residue-specifically fluorinated proteins enable a wide range of experiments, they can sometimes be complicated by sample heterogeneity (fluorination at a particular residue is usually no higher than 95%), difficulties in the peak assignments of individual fluorine atoms, and perturbations of protein structure and/or function (147, 164). Site-specific incorporation of fluorinated amino acids provides an alternative approach to the generation of isotopically labeled NMR samples. Furter was the first to report one such technique using *p*-fluorophenylalanine (**6**), which was substituted site-specifically in dihydrofolate reductase (DHFR) expressed in *E. coli* (70). The Frieden laboratory has employed Furter's method as a complementary approach to global fluorination. Bann and Frieden used a combination of site- and residue-specific incorporation of **6** into the bacterial periplasmic chaperone PapD and used these proteins to identify at least three distinct steps in the folding landscape (165).

Two groups have recently developed very efficient expression systems that improve on Furter's method for protein fluorination at aromatic residues, enabling essentially quantitative site-specific incorporation of fluorinated amino acids **22** and **23** into proteins for NMR studies (91-93). Hammill et al. demonstrated that, like previous ^{19}F NMR work, **22** serves as an extremely sensitive monitor of the environment surrounding the amino acid. Incorporation of the noncanonical residue into nitroreductase and histidinol dehydrogenase enabled both short- and long-range detection of ligand binding events. Cellitti et al. have employed ^{19}F NMR on site-specifically labeled proteins in combination

with other NMR techniques, and in this work the environmental sensitivity of fluorinated amino acids is again evident in ligand binding experiments. Li et al. have demonstrated that the exquisite sensitivity of ^{19}F NMR also enables site-specifically fluorinated proteins to be probed in intact *E. coli* cells (166).

While fluorinated protein samples can help answer many scientific questions, site-specifically labeled, native proteins would provide scientists with the opportunity to investigate a much wider scope of problems. A consortium of scientists at The Scripps Research Institute and The Novartis Research Foundation led by Peter Schultz and Bernard Geierstanger has developed some very promising approaches to introducing site-specific ^{13}C, ^{15}N, or ^{2}H labels into proteins (also discussed by Jones et al. (167)). Deiters et al. first used ^{15}N-labeled **24** to study sperm whale myoglobin, and isotopically labeled derivatives of **24** have since been used to study a thioesterase domain of human fatty acid synthase (FAS-TE) (91) and a cytochrome P450 enzyme (168). Cellitti et al. demonstrated the incorporation of isotopically labeled, photocaged tyrosine **25** into FAS-TE and showed that UV irradiation allowed for essentially complete decaging at multiple amino acid positions (91). The group also went on to label FAS-TE at eleven different amino acid positions with ^{19}F-labeled **23**, ^{15}N- and ^{13}C-labeled **24**, and ^{15}N-labeled, decaged **25** in order to exhaustively study protein-tool ligand interactions. Figure 1.5 depicts a summary of the chemical shift data acquired in these studies, which indicates that a number of residues are affected by the binding of a tool compound, including amino acids located in disordered loops that may have been difficult to identify using other methods. This impressive, detailed work indicates a promising future for biosynthetically produced, site-specifically labeled proteins in NMR studies. The precedent for studying protein folding and ligand

binding has been firmly established using fluorinated amino acids. The advent of improved site-specific ncAA incorporation techniques for insertion of fluorinated and native, isotopically labeled amino acids should allow for many more detailed protein characterizations via NMR in the future.

Protein folding and stability. Proper protein folding and robust maintenance of a specific three-dimensional structure oftentimes dictate a protein's ability to perform additional functions such as catalysis or binding. Understanding and quantifying the phenomena responsible for maintaining proteins in a specified conformation can provide insight into how to stabilize folded structures and how to design and/or engineer proteins with various functions. Because folded proteins are oftentimes only thermodynamically stable by only a few kilocalories per mol, individual noncovalent interactions such as hydrogen bonds and van der Waals contacts can greatly influence whether or not a protein is able to assume a folded conformation and the kinetics that dictate the speed at which this conformation is reached. Conventional mutagenesis allows for the study of many of the noncovalent interactions that govern protein folding and stability, but the limited side chain structures of the canonical amino acids do not always allow for a full exploration of these phenomena. Incorporating ncAAs into proteins allows for more subtle perturbations than many canonical amino acid mutations allow. Researchers have perturbed proteins both locally and globally with ncAAs in order to study factors governing protein folding and stability.

Local perturbations. The subtle mutations possible with ncAAs allow investigators to perturb a single noncovalent interaction within a large protein structure and study the

perturbation's effects on the overall protein characteristics. In the 1990s, the Schultz laboratory published a series of studies demonstrating the utility of in vitro unnatural amino acid mutagenesis for studying local aspects of protein stability. Mendel et al. incorporated structural variants of leucine, including ncAAs **26–31**, at position 133 of T4 phage lysozyme (T4L) in order to systematically study several effects including enlargement and shrinking of the cavity into which leucine 133 points (169). The researchers discovered that sequential removal of methyl groups from position 133 reduced the overall stability of the protein in a nonlinear fashion, a phenomenon that could also be reproduced in computational studies. These and other findings led the authors to suggest that many noncovalent interactions are important in determining protein stability and that a combination of modeling and unnatural amino acid mutagenesis could be used in the future to shed light on the factors leading to stable, folded proteins. Ellman et al. reported a related study examining the effects of backbone mutations at position 82 of T4L (170). In this work, the thermal stabilities of T4L variants containing ncAAs **32–42** at position 82 were measured used CD spectroscopy and heat inactivation assays. The results indicated that both protein backbone conformational restriction and the angle of the restriction are important factors in determining protein stability. Cornish et al. investigated the effects of β-branched amino acids on the stabilities of two α-helices in T4L using ncAAs and computational models (171). Incorporation of β-branched amino acids at two positions in T4L resulted in protein stabilization in one case and substantial destabilization in another. These findings emphasize the fact that context appears to play a role in the effect a β-branched amino acid has on protein stability.

Substantial efforts have also been devoted to the study of hydrogen bonds in proteins using systematic mutations. Thorson et al. used ncAAs **6** and **37–39** in order to study mutations to particular hydrogen bonds in Staphyloccocal nuclease (SNase) (172). Mutations of two glutamates to the weak hydrogen bond acceptor **37** and two tyrosines to weaker hydrogen bond donor **38** and repulsive lone pair interactor **39** confirmed the existence of two glutamate-tyrosine hydrogen bonds. Stability measurements on protein variants confirmed that the hydrogen bonds in question were responsible for one to two kilocalories per mol stabilization of the folded protein state, confirming previous estimates of hydrogen bond strengths. Studies on hydrogen bonds in α-helices, β-sheets, and β-turns with α-hydroxy acid analogs of leucine and isoleucine (**40** and **41**, respectively) demonstrated that hydrogen bonds in key secondary structural features can be investigated by changing the hydrogen bonding character of the backbone (173-175). These studies all demonstrate the utility of backbone variations in assessing the thermodynamic contributions from hydrogen bonds involving polypeptide main chain atoms. Thorson et al. used noncanonical amino acids to perform a linear free energy analysis on a hydrogen bond in SNase (176). The authors focused on the importance of the hydrogen bond mediated by the hydroxyl group of tyrosine 27. Substitution of tyrosine by amino acids **16**, **42**, and **43**, which have increasingly acidic pK_as, confirmed that the side chain of tyrosine 27 serves as a hydrogen bond donor in SNase. The use of a series of tyrosine analogs eliminated context-dependent, confounding factors that oftentimes complicate the direct assessment of the importance of a particular hydrogen bond in a protein.

The above studies illustrate the power of atomic-level mutations in the analysis of protein stability. However, systematic local perturbation of soluble proteins using ncAAs

has become less common in recent years, perhaps because of difficulties in achieving high protein yields using in vitro protein production methods, or the availability of alternative protein production techniques such as protein semisynthetic methods. The recent site-specific incorporation of α-hydroxy acids into proteins produced in *E. coli* may lead to new possibilities for protein stability studies in the future (177). Regardless of the future of local perturbations of globular proteins, the work outlined above was crucial in setting the stage for the study of membrane proteins using extensive unnatural amino acid mutagenesis, a topic to be discussed below.

Global perturbations. Global changes in the amino acid composition of a protein allows for the study of aggregate effects that may be too small to study individually. The fluorination effect has been a heavily studied topic for quite some time (178), including in the context of biological systems (179-181). Coiled coils have been a particularly useful protein-based system for studying the fluorination effect because the residues mediating protein-protein interactions in this system are well defined. Tang et al. have studied variants of one such multimeric α-helical protein, A1, containing different degrees of fluorination within its hydrophobic core. The investigators found that increasing the amount of fluorine-fluorine contacts within the hydrophobic core by substituting leucine with trifluoroleucine (**44**) or hexafluoroleucine (**3**) led to peptides that were increasingly resistant to thermal and chemical denaturation (41, 182). Son et al. found similar trends in a model coiled coil system in which isoleucine residues were replaced with trifluoroisoleucine (**45**) residues or valine residues were replaced with trifluorovaline (**46**) residues, although the degree of stabilization differed depending on whether isoleucine or

valine was converted into its trifluorinated form (183). Montclare et al. also recently reported that stabilization of the model protein A1 is relatively insensitive to the stereochemistry of fluorinated amino acids incorporated into the protein (184). Upon fluorination, amino acid side chain volumes increase by a significant amount, and size increases may be responsible in part for the improved stabilities of fluorinated coiled coils. In order to investigate this possibility in more detail, Van Deventer et al. utilized the amino acid homoisoleucine (Hil, **29**) (185). Hil has a nearly identical side chain molecular surface area to trifluoroleucine but retains the aliphatic character of canonical amino acids. Interestingly, replacing the leucines in A1 with **29** results in stabilization that is equal to or better than the stabilization of A1 that results from replacement of Leu with trifluoroleucine. These results suggest that side chain size does play a role in dictating protein stability in the context of hydrophobic interactions. However, femtosecond timescale experiments performed on proteins containing solvent-exposed fluorinated and aliphatic amino acids indicate that solvation dynamics near a protein surface change drastically depending on the absence or presence of nearby fluorinated groups (186). Future experiments using size-matched aliphatic and fluorinated amino acids may continue to enable the elucidation of the unique properties of fluorinated biomolecules.

Perturbing the amino acid composition of large proteins can alter protein folding and stability significantly, especially when the proteins in question assume more complicated structures than coiled coils. Although sometimes global replacement of a canonical amino acid within a protein can yield biomolecules with reduced thermal stabilities (187, 188) or an increased propensity for aggregation (189), these effects are quite dependent on the identity of the protein and the ncAA substitutions made. Oftentimes,

such substitutions are tolerated quite well in proteins. For example, Wang et al. report that fluorination of murine interleukin-2 (IL-2) at isoleucine residues results in IL-2 molecules that are nearly as effective as their nonfluorinated counterparts in mammalian cell proliferation assays (190). Budisa and coworkers have proposed using ncAAs as general probes of protein folding (191). In one report out of the Budisa laboratory, researchers investigated the effects of incorporating (4*R*)- and (4*S*)-fluoroproline (**47**, **48**) in place of proline in enhanced green fluorescent protein (eGFP) (192). Incorporation of **47** into eGFP resulted in proteins located in inclusion bodies of *E. coli* that could not be refolded. On the other hand, incorporation of **48** into eGFP resulted in a protein with significantly faster refolding kinetics and overall refolding yields than nonfluorinated eGFP, which the authors attributed to **48**'s higher C^{γ}-*endo* puckering and *cis* isomerization preferences compared to proline. This work provides one of the first examples of a protein that has improved folding properties when a ncAA is incorporated throughout its structure. The properties of single-chain Fv fragments of antibody fragments have also been subjected to proline fluorination with **47** and **48** (193). Interestingly, the stability of the scFv was found to be improved when proline was replaced with **47**, the stereoisomer that was found to be detrimental to eGFP folding. Budisa and coworkers recently employed methionine analogs **5** and **49** in the study of protein misfolding leading to prion disease (194). In this work, they used **5** and **49** because of their increased hydrophobicity and hydrophilicity, respectively, relative to methionine. Global replacement of methionine by **5** in recombinant human prion protein ($rhPrP^{C}$) resulted in less protein aggregation than in $rhPrP^{C}$ containing methionine. On the other hand, replacement of methionine by **49** yielded a much more aggregation-prone version $rhPrP^{C}$ than the methionine version of the protein. CD spectroscopy also revealed

significant differences in secondary structure and unfolding behavior as a function of temperature. These results can be seen in figure 1.6. These interesting experiments suggest that the hydrophobicity of the moieties present at positions in the rhPrP^{C} normally occupied by methionine can significantly impact how well the protein maintains its folded state, and that hydrophilic moieties appear to substantially impact the proper folding of the protein in question. Because methionine oxidation results in the formation of more hydrophilic functionalities at methionine positions, this work suggests that oxidative stresses may play a role in the development of prion disease and other diseases caused by protein misfolding. These techniques may also be applicable to studying a number of other cellular proteins that can undergo oxidation at methionine residues (195-197). Studying protein stability using global perturbations introduced with ncAAs appears to be useful in a variety of settings. Both model proteins and more complex proteins can be perturbed in ways such that protein stability is either negatively or positively affected. Observing these changes in stability provides fundamental information regarding how changing molecular properties results in changes to protein properties as a whole and also provides ideas for engineering proteins with altered stabilities. Thus, global perturbations provide a complementary approach to local, site-specific perturbations. These two approaches add substantial capability to researchers' toolkit for assessing factors contributing to protein stability and engineering more stable, faster folding proteins.

Membrane proteins. Membrane proteins play extremely important roles in organisms from all classes of life and may comprise upwards of one quarter of all open reading frames in the genomes of fully sequenced organisms (198). These proteins also form a large

portion of "druggable" protein targets (199-201). However, structural characterization of this important set of proteins has continued to be plagued by problems with overexpression and crystallization, although there have been some recent advances (for example, see (202-204)). The dearth of detailed molecular information has necessitated the adaptation of other experimental techniques for studying membrane proteins, including the very powerful combination of conventional mutagenesis and the patch clamp technique (205). Several laboratories have found that employing ncAAs in studies of receptors and ion channels further augments the power of more traditional biochemical characterization methods. Investigations of topics including channel architecture, the functional significance of individual amino acids and noncovalent interactions, and mechanisms of ligand binding have been aided significantly with the use of ncAAs.

Exploiting unique functionalities of noncanonical amino acids. Several studies of membrane proteins have employed ncAAs to learn about structural and mechanistic features of particular channels and receptors. Surprisingly, many of these investigations have utilized techniques that take advantage of "highly unnatural" features of noncanonical amino acids, an observation that Dougherty made in a recent review (206). Gallivan et al. employed biocytin (**50**) in order to detect surface-exposed residues in muscle-type nicotinic acetylcholine receptor (nAChR) (207). This work used information on protein expression and the binding of streptavidin probes to proteins present in intact *Xenopus* oocytes to reveal the orientations of several amino acids within the main immunogenic region of the nAChR. England et al. employed ultraviolet light-cleavable amino acid **51** and base-cleavable α-hydroxy acids **32**, **40**, and **52** in order to investigate characteristics of a portion

of the muscle-type nAChR known as the Cys-loop (208, 209). These experiments confirmed the functional significance of the Cys-loop in signaling and helped to establish the disulfide bond connectivity within the loop.

Spectroscopically active ncAAs have proven very useful for studying structural features of G protein-coupled receptors (GPCRs). Fluorescent amino acid **53** has been employed in FRET studies of the GPCR tachykinin neurokinin-2 (NK2) in order to study the structure of the protein (210, 211). In vitro preparations of NK2 containing the fluorescent amino acid were exposed to a ligand labeled with a FRET partner. Distance constraints within the receptor were determined from measured FRET effects, leading to a more accurate picture of the general structure of the receptor. Local mechanistic features of the GPCR rhodopsin were recently probed using *p*-azidophenylalanine **54** as a vibrational spectroscopic probe (212). The unique vibrational signature and sensitivity of the azide to changes in the local electrostatic environment enabled Ye et al. to investigate structural changes in the receptor upon light activation. Placement of **54** at different sites within rhodopsin enabled the elucidation of the order of several sequential helix movements as the protein assumes its active conformation, leading to a general model for GPCR activation (213). This work appears to be extremely promising and should be applicable to any number of membrane proteins due to the exquisite sensitivity of azides to local environment.

Noncanonical amino acids of various sizes have also been employed to examine the properties of voltage-gated potassium channels. The mechanism of channel inactivation in potassium channel Kv1.4 has been studied in mammalian cells using amino acids bulkier than tyrosine (80). The authors observe slower channel inactivation times when a naturally

occurring tyrosine residue is replaced by **24** or **55**. They argue that these observations provide support for a mechanism in which the N-terminus of the channel moves through a side pore in the channel to inhibit the flow of ions through the inner pore. A series of phenylalanine analogs of various sizes (**56–59**) was used to examine the importance of a particular Phe residue at position 233 within the *Shaker* voltage-gated potassium channel (214). Interestingly, channels containing the large aliphatic amino acid cyclohexylalanine **60** in place of Phe still retain function, indicating that the size of Phe is the most important functional characteristic of the residue in question. Phe analogs of various sizes also affect channel behavior in a systematic way, supporting the hypothesis that a Phe to Trp mutation at position 233 alters channel function primarily on the basis of its increased bulk. The above examples exploit properties of ncAAs that make these residues distinct from canonical amino acids. The use of these unique functionalities has enabled the study of a wide variety of structural and mechanistic aspects of membrane proteins, and several of these techniques should continue to provide insights into the functions of numerous receptors and channels.

Atomic-level perturbations. In the absence of three-dimensional structural information, meaningful characterization of the role of a particular amino acid or noncovalent interaction within a protein oftentimes requires atom-by-atom perturbations. Since the Dougherty and Lester laboratories first described the incorporation of noncanonical amino acids into membrane proteins in *Xenopus* oocytes (215), much work involving ncAAs in membrane proteins has exploited very subtle perturbations to study aspects of channel and receptor function (206, 216).

Atomic-level mutations have revealed important information about the roles that individual amino acids play in receptor gating. Hanek et al. reported studies on the nAChR investigating the role of a particular valine in the gating of the receptor (217). These studies uncovered a "pin-into-socket" binding mechanism within the channel by examining dose-response curves of channels substituted with amino acids **61–63**. Another investigation authored by Lummis et al. investigated the role of a particular proline in the gating of a 5-hydroxytryptamine type 3 (5-HT_3) receptor (218). The authors hypothesized that a proline residue located in between two transmembrane helices was involved in the gating mechanism of the receptor. Although conventional mutagenesis of the channel resulted in a loss of receptor gating activity, introduction of amino acids **33**, **47**, **48**, and **64–66** in place of proline resulted in functional channels. Interestingly, the effector concentration for half maximal response (EC_{50}) of the resulting channels correlate very strongly with the propensity for each ncAA replacing proline to assume a *cis* protein backbone conformation. Figure 1.7 depicts the linear relationship observed between *cis*-*trans* isomerization and receptor activation. Based on this striking correlation and additional studies, the authors conclude that the relatively facile *cis*-*trans* isomerization of proline, and not any other properties of this cyclic canonical amino acid, is responsible for the gating characteristics of the 5-HT_3 receptor. The data from their studies also led to a proposal of a new model for the gating mechanism of the receptor (figure 1.7c).

Atomic-level mutations have also provided insight into the importance of particular noncovalent interactions within receptors and channels. Hydrogen bonds involving main chain and side chain interactions have been particularly well studied. In 1995, Nowak et al. made mutations to three tyrosine residues in muscle-type nAChR that included analogs of

tyrosine with varying hydroxyl side chain pK_as and other amino acids lacking hydroxyl groups altogether. Based on dose-response curves of the mutant channels, the authors concluded that only one of the residues examined was involved in a hydrogen bonding interaction through the hydroxyl group (215). Beene et al. used similar atomic perturbations to study four different tyrosine residues present in the 5-HT_3 receptor binding site (219), finding that each residue plays a unique functional role within the channel. More recent work with the 5-HT_3 receptor and ncAAs has determined the roles of two highly conserved residues within loop A of the ligand binding site, one of which appears to form a critical hydrogen bond with the ligand (220). Last, the nicotinic pharmacophore of the $\alpha4\beta2$ neuronal nAChR was investigated with the help of **40** (221).

Hydrogen bonds with main-chain atoms can be perturbed effectively with the aid of α-hydroxy amino acids and other ncAAs. Studies of a conserved proline residue within the M1 region of ligand-gated ion channel receptors using amino acids **32**, **33**, **40**, **52**, and **67** have suggested that proline is favored for its hydrogen bonding characteristics (222, 223). Gleitsman et al. used double-mutant cycles with ncAAs to study hydrogen bonding networks in the muscle-type nAChR (224). Their detailed investigations were instrumental in correcting homology models that had incorrectly predicted the hydrogen bonding patterns based on structures of the acetylcholine-binding protein. Similar studies on an aspartate residue participating in several hydrogen bonds (225) and the hydrogen bonds in a beta sheet structure (226) of the muscle nAChR have also provided information regarding the functional significance of particular hydrogen bonding patterns.

Additional ncAAs have been used to examine other important noncovalent interactions that play important roles in agonist and antagonist binding events (227, 228).

Numerous receptors and channels have a very large number of aromatic residues that directly line the binding site, and as early as 1990, aromatic residue-mediated cation-π interactions were proposed in acetylcholine-binding proteins (reviewed in (229, 230)). The use of ncAAs has contributed greatly to the investigation of potential cation-π interactions between ligands and the ligand binding sites of membrane proteins. Zhong et al. first described the use of a series of aromatic ncAAs to study ligand binding in the muscle nAChR in 1998 (231). Experimental measurements of EC_{50} values of channels containing a number of tryptophan analogs at three positions of the ligand binding site showed very little dependence on the tryptophan analog incorporated. However, at the α149 position, the EC_{50} value varied widely depending on the analogs used. Remarkably, comparison of experimentally determined EC_{50} values with fluorinated tryptophan analogs **13–15**, **68–70** and quantum mechanical calculations of cation-π binding strengths of each of these analogs revealed a very strong correlation. This strong relationship between ligand binding and cation-π binding strength at a single amino acid, depicted in figure 1.8, clearly established the presence of a single cation-π interaction in the nAChR. Subsequent studies have continued to reveal important characteristics of numerous cation-π interactions in nAChR binding events, including differences in the binding of nicotine, acetylcholine, and epibatidine to the receptor (232). Cashin et al. compared nicotine and acetylcholine binding to the nAChR with the drug epibatidine and established that epibatidine has a cation-π interaction similar to acetylcholine (233). Xiu et al. investigated differences between the muscle nAChR and the A2B3 form of the α4β2 neuronal nAChR that likely plays a role in nicotine addiction (234). One key finding that they reported was the existence of a strong cation-π interaction between nicotine and a tryptophan residue in α4β2 homologous to

tryptophan α149 in muscle nAChR, which they confirmed with the same series of fluorinated tryptophan analogs described above. This cation-π interaction between nicotine and tryptophan is absent in the muscle nAChR, explaining a longstanding mystery regarding the selectivity of nicotine for neuronal receptors over the muscle nAChR.

In addition to the extensive studies on cation-π interactions in nAChRs, researchers have employed ncAAs to study potential cation-π interactions in other integral membrane proteins. Studies on the Cys-loop family of receptors have revealed that many proteins within the Cys-loop family bind ligands using cation-π interactions, but utilizing aromatic residues at different structural locations (232, 235-238). Studies on the voltage-gated sodium channel $Na_v1.4$ involving ncAAS have also proven fruitful, identifying cation-π interaction binding sites for tetrodotoxin, calcium binding, and local anesthetics (239-241). Use of ncAAs in the voltage-gated potassium channel *Shaker* yielded insights into the location of a cation-π interaction and the structural conformation of the channel, and these findings helped to resolve discrepancies between computational structural predictions and previous experimental work (242). Last, McMenimen et al. have demonstrated that the tryptophan thought to be responsible for cation-π-mediated magnesium ion blockade in the *N*-methyl-D-aspartate receptor is instead favored for its large, hydrophobic, and flat characteristics (243).

Research efforts highlighted above demonstrate that investigations of membrane proteins with ncAAs can provide a wide variety of structural and mechanistic information ranging from atomic to whole-protein distance scales. To date, the majority of investigations in this area have focused on ligand- and voltage-gated ion channels, but recent reports employing ncAAs in the study of GPCRs (212, 244-246) suggest that efforts

to broaden the applicability of ncAAs to a wider range of membrane proteins are underway. As long as three-dimensional structural information about membrane proteins remains scarce, ncAAs should continue to play an important role in the study of these complex biomolecules.

Posttranslational modification mimicry. Although only twenty amino acids are usually genetically encoded in proteins, nature has devised a huge number of posttranslational modifications that modulate the structure, activity, and localization of proteins (2). The precise effects that these modifications have on proteins can be extremely difficult to study due to the dynamic nature of many of these modifications in living cells and the challenges in purifying or preparing uniformly modified proteins to study in vitro. Proteins bearing authentic posttranslational modifications or close structural mimics of these modifications can be prepared in conjunction with ncAAs. Both chemical routes and direct incorporation of amino acids mimicking posttranslational modifications have been explored and applied to the study of posttranslationally modified proteins.

Chemical approaches. Performing chemistry on proteins forms one major approach to generating molecules containing posttranslational modifications found in naturally occurring proteins. Many protein modification techniques have been developed for the chemical transformation of residues found in naturally occurring proteins (247), and some of these methods have been exploited in the generation of structures mimicking posttranslational modifications. For example, the Shokat laboratory has developed an elegant method for modifying cysteine residues in recombinantly produced proteins in such

a way that they resemble methyl-lysines (248). Semisynthetic methods such as native chemical ligation and expressed protein ligation also provide another efficient chemical route to proteins containing posttranslational modifications (5-9). NcAAs provide a complementary approach to these methods because of the improved yields that biosynthetic protein production offers compared to other preparations of posttranslationally modified proteins. Approaches to employing bioorthogonal chemistry for the selective modification of biomolecules have recently been reviewed from the perspectives of the range of bioorthogonal chemistries explored (249), the particular utility of azide-alkyne click chemistry (250), and chemistries that can be performed on ncAAs in proteins (251).

Several groups have utilized chemistries introduced by the incorporation of ncAAs in order to create mimics of posttranslationally modified proteins. Three separate groups have found chemical transformations that lead to dehydroalanine-containing proteins, two of which (Seebeck et al. and Wang et al.) involve the use of ncAAs (252-254). The Schultz laboratory has demonstrated that analogs of methyl- and acetyl-lysine (**71**–**73**) can be chemically installed in proteins by performing chemistry on dehydroalanine-containing proteins (255), and using dehydroalanine (**74**) as a chemical handle has been proposed as a general starting point for protein modification (256). The Chin laboratory has demonstrated the utility of protecting groups in the development of a method for preparing site-specifically and quantitatively methylated histones. They installed dimethyllysine (**75**) by site-specific incorporation of the protected lysine *N*-tert-butyloxycarbonyl-L-lysine (**76**) and an ingenious protection-deprotection scheme (257, 258). Using the same ncAA (**76**), Virdee and coworkers combined genetic code expansion, intein chemistry, and chemoselective ligation to create a method termed GOPAL (genetically encoded

orthogonal protection and activated ligation) to synthesize diubiquitin chains that have not been chemically accessible by previous means of synthesis (259). The resulting proteins were used to solve the structure of a previously uncharacterized diubiquitin conjugate and to study the specificities of human deubiquitinases. Eger et al. have also demonstrated the utility of a combination of site- and residue-specifically incorporated azide and alkyne amino acids **77** and **78** to chemically synthesize diubiquitin molecules using copper-catalyzed azide-alkyne cycloadditions (CuAAC) (260). This method might be simpler and more versatile than the methods described by the Chin laboratory, but the resulting structures contain unnatural triazole ring linkages rather than authentic linkages found in nature.

The complexities of protein glycosylation have inspired a number of approaches to the chemical and enzymatic synthesis of sugar-decorated proteins (recently reviewed by Gamblin et al. (261)). Part of the appeal of chemical approaches to adding glycans to proteins is the ability to precisely define the structure of a glycan prior to attaching it to a protein without having to worry about the enzymatic efficiencies of glycosyltransferases or potential constraints within the protein translation machinery. Davis and coworkers have recently reported two alkene-based routes to glycosylated proteins using chemistry described above. Each of these methods leads to the construction of quantitatively glycosylated model proteins, although they require the use of proteins containing only one cysteine and one methionine each, respectively (254, 262). The Schultz laboratory demonstrated the modification of ketone-containing amino acid **79** in the Z-domain of staphylococcal protein A with an aminooxy analog of *N*-acetylglucosamine (263). After nearly quantitative reaction with the first sugar, the protein could be enzymatically

modified at the glycan with subsequent sugars to produce more complex carbohydrate structures. Chemical attachment of multisugar structures to ketone-containing proteins was also achieved, although not in quantitative yields. Van Kasteren et al. have used azidohomoalanine (**77**) and homopropargylglycine (**2**) to "click" sugar structures to model proteins using CuAAC (264, 265). Unlike the ketone-mediated couplings, CuAAC afforded nearly quantitative yields when complex carbohydrate structures were ligated to proteins. CuAAC-mediated glycosylation was also combined with cysteine modification to generate doubly glycosylated proteins, although this required the use of proteins containing only one methionine and cysteine residue each. Despite these drawbacks, these chemically produced glycan mimics were used to study glycan recognition and as probes of glycan-binding activity in vivo. One especially interesting observation was that neuronal cells in rat brain tissue sections appear to selectively bind proteins containing GlcNAc modifications, while nearby glial cells do not appear to possess the same binding capabilities. Chemical approaches to the generation of posttranslationally modified proteins appear to be especially useful in glycoprotein synthesis, and the generality of this approach should be very useful for generating quantitatively modified structures for studying several outstanding questions in glycobiology.

Direct incorporation approaches. Ribosomal synthesis of proteins containing appropriately designed ncAAs allows for the creation of some authentic, genetically encoded posttranslationally modified proteins and metabolically stable mimics of other modifications. Significant effort has been dedicated to the genetic incorporation of subtle posttranslational modifications into proteins. The Schultz laboratory has biosynthetically

incorporated sulfotyrosine (**80**) into proteins expressed in *E. coli* and used it in a number of contexts. In an initial report, they demonstrated the improved inhibition of thrombin by sulfo-hirudin versus desulfo-hirudin (266). They also solved the X-ray crystal structure of the sulfo-hirudin-thrombin complex using biosynthetically produced sulfo-hirudin (267). In this case, the genetic encoding of the modification enabled the generation of large protein samples quantitatively sulfated at a specific position, a feat difficult to achieve by any other protein production method. More recently, Schultz and coworkers have explored the utility of sulfated antibodies in the context of directed evolution (268, 269); these results will be discussed below in the "Evolution" subsection.

Lysine modifications are particularly important posttranslational modifications. For example, lysine acetylation, which results in the amino acid N^{ε}-acetyllysine (**81**), impacts the function of many proteins. A report has described the biosynthesis of proteins containing this modification, including the naturally lysine-acetylated protein rat mitochondrial manganese superoxide dismutase (100). Chin and coworkers have recently applied this strategy to the study of histone acetylation (270). The preparation of histone H3 quantitatively acetylated at lysine 56 enabled the mechanistic study of a number of previously proposed effects of histone acetylation on histone properties. For example, FRET experiments indicated that reconstituted nucleosomes acetylated at lysine 56 of H3 have DNA breathing increased by seven times over nonacetylated nucleosomes, explaining previously observed changes in gene expression from H3 mutants lacking lysine at position 56. Suga and coworkers have also investigated the effects of lysine acetylation and methylation on heterochromatin protein 1 recognition of histone H3 N-terminal peptides by reprogramming the genetic code to combinatorially incorporate acetylated and methylated

lysine residues into genetically encoded peptides (271). Aside from histone acetylation, cyclophillin A (CypA), a protein with important roles in immunosuppression and viral infection, is also acetylated in human cells. Using an orthogonal acetyllysyl-tRNA synthetase/$tRNA_{CUA}$ pair and amber suppression, Lammers et al. produced homogeneously and site-specifically acetylated recombinant CypA in *E. coli* (272). This approach enabled structural and biophysical analyses on acetylated CypA for the first time and revealed important roles of acetylation on CypA function. These roles include suppression of the protein's catalytic activity, recognition of the HIV-1 capsid, cyclosporine binding, and calcineurin inhibition.

Xie et al. have genetically encoded a phosphotyrosine mimic in proteins (273). While phosphotyrosine is subject to hydrolytic and enzymatic cleavage, amino acid **82** is not due to the more stable linkages involved. The single negative charge on the amino acid is a reasonable substitute for the doubly charged phosphate group present in phosphorylated amino acids. DNA binding studies with a fragment of human signal transducer and activator of translation-1 (STAT1) containing amino acid **82** showed that it had an intermediate apparent affinity for DNA in between analogous phosphorylated STAT1 and unphosphorylated STAT1. These results indicate that the phosphotyrosine mimic may be used to study the effects of phosphorylation at specific protein sites, although the mimic is not a perfect analog of authentic phosphorylation. The amino acid 3-nitrotyrosine (**83**), a residue commonly associated with disease and oxidative damage, has also been incorporated into proteins produced in *E. coli* (274). Studies on the catalytic activity of manganese superoxide dismutase (MnSOD) indicate that nitration of a particular tyrosine in the protein greatly impacts the catalytic activity of MnSOD, suggesting that

future studies of tyrosine nitration in proteins may shed additional light on the molecular and cellular effects of this modification. As stated above, protein modification by glycosylation is an area of widespread interest. However, biosynthetic incorporation of glycans into proteins has been a rather difficult task, and early reports on the subject (275, 276) have since been retracted (277, 278). One report of the successful incorporation of a glycosylated amino acid into proteins using an in vitro translation system has appeared in the literature (279), suggesting that future efforts to directly incorporate glycosylated amino acids into proteins may yet prove fruitful.

More than half a dozen posttranslational modifications or mimics have been incorporated into proteins using biosynthetic approaches. These approaches to incorporating small posttranslational modifications and chemical approaches to incorporating modifications such as glycan structures into proteins appear to be relatively straightforward. The successful creation of uniformly modified proteins and elucidation of many of significantly altered functional properties of these proteins suggests that ncAAs will continue to provide researchers with powerful, specific tools for studying the effects of posttranslational modifications on biological processes.

Immune modulation. Collaborative efforts at The Scripps Research Institute have recently raised the possibility of using ncAAs as aids in the development of vaccines. Many approaches to developing vaccines have previously been described that include the use of viruses, T-helper epitopes, and sophisticated adjuvants (280-282), but difficulties remain in developing robust immune responses against self-proteins. Grünewald et al. first presented the idea of incorporating immunogenic amino acids into self-proteins in 2008 (283). In this

work, researchers incorporated *p*-nitrophenylalanine (pNO_2Phe, **84**) into murine tumor necrosis factor-alpha (mTNF-α) based on previous reports of the highly immunogenic nature of nitroaryl groups. Immunization of mice with purified mTNF-α containing **84** resulted in a very high level of mTNF-α-specific antibodies in the serum of the treated animals. In contrast, immunizations using wild-type mTNF-α or a variant containing a tyrosine to phenylalanine mutation did not result in significant antibody titers. The utility of the immunization procedure was tested using a mouse model of severe endotoxemia. Figure 1.9 depicts the results of challenging mice with bacterial lipopolysaccharides. While the presence of mTNF-α in the serum usually contributes to septic shock and death of the challenged mice, these data indicate that the presence of antibodies against mTNF-α can significantly improve the survival rate of the challenge. A follow-up report on this work expanded upon the above studies (284). In this work, the investigators found that the site of pNO_2Phe within mTNF-α affected the strength of the immune response elicited in the mouse. They also note, somewhat surprisingly, that antibodies isolated from the serum of immunized mice do not necessarily recognize peptide epitopes containing the immunogenic amino acid, although further investigation will be necessary to examine this finding in more detail. Finally, the researchers incorporated pNO_2Phe into murine retinol-binding protein (mRBP4) and immunized mice with the resulting protein, again finding robust immune responses after administration of the proteins containing **84**. In each of these papers, the authors stress that they have produced immunogenic materials based on precise molecular manipulations, an approach not possible with other immunization techniques. These results raise the tantalizing possibility that immunogenic ncAAs may lead to a general approach to developing therapies involving recognition of a self-protein or

other weakly immunogenic protein. While application to treatments for cancer or other diseases may still be quite distant, the initial experiments described here show a great deal of promise.

Evolution. As far as scientific inquiries have been able to establish, the genetic code has remained more or less constant since its establishment in the so-called frozen accident (285). Therefore, evolution has operated with a single genetic code for more than one billion years (with some minor variations arising along the way (286)). However, there is no fundamental barrier to performing evolution experiments with alternative genetic codes. Using ncAAs, researchers have recently combined ncAAs and directed evolution to investigate new approaches to genetic code expansion and protein engineering.

Evolution of organisms with altered genetic codes. Little is known about how the standard genetic code was established. Before becoming fixed, changes to the genetic code likely involved either introduction of a new amino acid into proteins or the substitution of one amino acid for another within proteins (286-288). These significant changes likely required substantial adaptations by the organism in order to accommodate the alterations. Two groups have investigated this adaptation process using laboratory evolution techniques. In 1983, Wong grew tryptophan-auxotrophic *Bacillus subtilis* on solid media supplemented with 4-fluorotryptophan (**13**) in place of tryptophan (Trp) (289). After just two rounds of selection and two more of mutagenesis and selection, Wong isolated a mutant that grew logarithmically in a liquid culture supplemented with **13**, but only linearly when the medium was supplemented with tryptophan. These impressive results demonstrated that it

is possible to change an organism's amino acid preference from a canonical one contained in the genetic code to a noncanonical one. Bacher and Ellington performed similar selection experiments using *E. coli* auxotrophic in Trp production and then identified genetic mutations in the evolved strains (290). Three thousand hours of serial dilutions in liquid cultures containing increasing percentages of **13** in place of Trp resulted in strains showing improved growth rates in medium containing 99.97% **12**, although these strains still showed a growth preference for Trp. Identification and characterization of mutations revealed that one mutation in the gene encoding tryptophanyl-tRNA synthetase enables improved discrimination against **13**, but mutations in other genes conferred improved tolerance of the noncanonical substrate to the organism. Bacher et al. also subjected the phage Qβ to selection in media containing 95% 6-fluorotryptophan (**15**) (291). Two independent phage lines were subjected to twenty-five rounds of selection. Seven mutations became fixed in each of these lines, and, surprisingly, these mutations did not involve the alteration of any tryptophan codons. In the case of both *E. coli* and phage, the evolved organisms retained their abilities to grow on tryptophan while acquiring new growth capabilities. These results suggest that organisms can adapt to tolerate ambiguity within their genetic codes, implying that the "ambiguous intermediate" theory of genetic code expansion is plausible (288). Future studies of the ncAA accommodation process may shed additional light on the mechanisms by which changes to the genetic code are accepted in living creatures.

Directed protein evolution. In the past two decades, scientists have developed a number of high-throughput selections and screens enabling the directed evolution of proteins (1, 292-

295) and higher order protein-based systems (296). These efforts have enabled researchers to study the evolution of individual proteins and engineer proteins with new functions (297). Most work in this area has focused on exploring the sequence space defined by polypeptide chains containing the twenty canonical amino acids. However, two groups have employed ncAAs to explore alternative protein sequence spaces using directed evolution. The Tirrell laboratory has focused on the development of functional proteins fluorinated at leucine positions by substituting trifluoroleucine (**44**) for leucine. Global replacement of leucine in chloramphenicol acetyl transferase (CAT) resulted in a protein with greatly reduced half-life of inactivation at 60 °C (188). Two rounds of error-prone PCR and screening for mutants with increased activity after prolonged incubation at 60 °C yielded a mutant with substantially improved thermostability properties. Interestingly, the nonfluorinated mutant protein retained the thermostability of the parent CAT, indicating that mutations introduced during the course of evolution allowed the protein to improve its properties when fluorinated while retaining its original function. This trend is similar to the trend observed when an amino acid in the genetic code of an organism is replaced globally; both organisms and proteins appear to first adapt to accept an ambiguous genetic code before gaining a preference for the altered genetic code. In another set of evolution experiments, Yoo et al. evolved a GFP variant containing **44** in place of leucine using eleven rounds of random mutagenesis and screening using fluorescence-activated cell sorting (FACS) (189). A summary of the progress observed during the course of the directed evolution experiment is depicted in figure 1.10. The fluorescence of cells expressing various mutants in the presence of leucine or **44** indicates that nonfluorinated GFP mutants retain (and perhaps even slightly improve) their function, similar to the trend

observed in CAT evolution. The authors also found a marked improvement in the folding kinetics of the evolved GFPs in the absence of fluorination. Interestingly, several of the mutations observed in the final isolated mutant were also observed after screening for "folding-enhanced" GFP variants (298), suggesting that evolving proteins containing ncAAs may be a general method for improving the folding properties of proteins.

The Schultz laboratory has recently applied site-specific incorporation methodologies to directed evolution problems. Their work to date has focused on using ncAAs that are known to have affinities for particular molecules to engineer antibody fragments. In an initial report, they described the development of a phage display system that can be used in conjunction with an orthogonal aaRS-tRNA pair to encode for antibody fragments containing twenty-one amino acids (269). In this paper, they demonstrated the utility of encoding sulfotyrosine (**80**) in a saturation mutagenesis library. Despite an expression bias against antibody fragments containing **80**, they were able to isolate variants from a naïve germline library containing sulfotyrosine after panning for binders against the HIV protein gp120. Perhaps these results were not too surprising since naturally occurring high-affinity antibodies against gp120 are known to be sulfated (299, 300), but the demonstration does prove that ncAAs can be used effectively in phage display applications. A follow-up report has shown that functional antibody variants can be isolated from saturation mutagenesis libraries of the tyrosine-sulfated 412d antibody fragment (268), although affinity maturation of the fragment proved challenging given the high affinity of the starting protein. Finally, this technique has also been applied to acyclic sugar binding (301). The amino acid *p*-boronophenylalanine (**85**) was encoded in a saturation mutagenesis library of phage-displayed antibody fragments, and the library was panned

against a glucamine resin. Many of the clones isolated after three rounds of panning were found to contain amino acid **85**. Again, these results might have been expected based on the well-known propensity of boronates and diols to form boronate esters. However, these proof-of-principle experiments suggest that future libraries of antibody fragments containing ncAAs may yield high-affinity binders to antigens that are normally difficult to target. Taken as a whole, directed evolution experiments with proteins containing ncAAs appear to be promising for evolving proteins with properties that are difficult to introduce using only canonical amino acids, including particular molecular recognition and catalysis events. Furthermore, both protein and organism evolution experiments with ncAAs may provide insights into the kinds of adaptations necessary to maintain evolutionary fitness as amino acid compositions are changed or expanded.

Proteomic studies. Labeling and identifying the proteins expressed in biological systems can provide great insights into the inner workings of these systems, including spatial and temporal information about the proteome (302, 303). Recent work with ncAAs has demonstrated that reactive amino acid analogs can function as effective tags for labeling and identifying newly synthesized proteins in a range of biological systems. The versatility of bioorthogonal chemistry enables the use of the same set of ncAAs for visualization and identification of newly synthesized proteins.

Fluorescent labeling of newly synthesized proteins. Although a number of effective genetic, enzymatic, and chemical strategies for fluorescently labeling proteins have been developed (recently reviewed by Sletten and Bertozzi (249)), most of these approaches

have two major shortcomings: the identity of the protein to be labeled must be known in advance of the experiment, and the DNA encoding the protein of interest must be genetically modified in order to enable labeling. Metabolic incorporation of amino acids that can be tagged or visualized in some way provides an approach to labeling newly synthesized proteins without genetically modifying the system in question and without knowing the identities of the proteins to be labeled in advance. A longstanding strategy for labeling and visualization of newly synthesized proteins in living cells and organisms has been the use of [^{35}S]methionine in conjunction with autoradiography (304). This method has enabled the study of a number of systems without genetic manipulations, but experiments involving [^{35}S]methionine must be performed with care due to the inherent dangers of working with radioactivity. Recently described approaches to residue-specifically incorporating reactive amino acids into proteins (for examples, see (46, 47, 305, 306)) have enabled alternative chemical approaches to visualizing newly synthesized proteins in cellular systems. In 2005, Beatty et al. described the chemical modification of proteins produced in *E. coli* containing amino acids **2** (Homopropargylglycine, Hpg) or **86** using a fluorogenic coumarin dye and copper-catalyzed azide-alkyne 1,3 dipolar cycloaddition (CuAAC) to generate fluorescently labeled proteins (307). The researchers found high, specific labeling of *E. coli* only after the cells had been incubated with ncAAs. Gel electrophoresis revealed that both an overexpressed recombinant protein and endogenously expressed *E. coli* proteins were labeled, suggesting that incorporation of ncAAs occurred in all newly synthesized proteins in the bacteria.

The concept of visualizing new protein synthesis proteome-wide using bioorthogonal chemistry has been extended to mammalian cells. Beatty, Liu, and

coworkers demonstrated that amino acid **2** could be used in mammalian cells as a chemical handle for labeling proteins expressed during a specified pulse time of the alkyne-containing amino acid (35). Aided by the methionine auxotrophy of mammalian cells, the technique was applied in several different cell types and studied by microscopy and flow cytometry. Fluorescence quantification by flow cytometry revealed that labeling was selective for cells that had been exposed to the alkyne-containing amino acid by approximately one order of magnitude and that the extent of labeling could be varied by introducing small amounts of methionine along with **2**. Visualization of the proteome has also been extended to labeling multiple populations of newly synthesized proteins using multiple pulses of ncAAs. In this work, proteins synthesized during specific time windows were distinguished by applying sequential pulses of **2** and **77** (Azidohomoalanine, Aha) to mammalian cells in culture. The researchers showed that after the completion of both pulses, proteins containing **2** could be labeled with an azide-containing fluorophore while proteins containing **77** could be labeled with an alkyne-containing fluorophore and visualized simultaneously within the same spatial area (308).

Monitoring the production of newly synthesized proteins within cells and organisms may shed light on a number of problems, including the localized synthesis of new protein populations. One area that has already been investigated using the visualization approaches described above is protein synthesis in neuronal dendrites, a poorly understood and somewhat controversial subject (309). Dieterich et al. designed fluorescent tags for visualization of azide or alkyne-containing newly synthesized neuronal proteins in situ (310). Their CuAAC tagging enabled detection of newly synthesized proteins after as little as 10 minutes of sample exposure to **2** or **77**, and sequential pulses enabled tagging of

multiple time-defined populations of newly synthesized proteins. Furthermore, using strain-promoted azide-alkyne ligation (249), the authors were able to study the diffusion of newly synthesized proteins by appending quantum dots to azide-containing neuronal proteins. The Flanagan laboratory has used ncAAs to aid their studies of the spatial regulation of protein synthesis in neuronal axons and dendrites. Use of **77** and an alkyne fluorophore helped confirm that the transmembrane receptor DCC, which regulates axon growth and guidance, colocalizes with newly synthesized proteins, revealing its previously undiscovered role in translation regulation (311).

Studies of protein *S*-acylation dynamics have also been aided by the visualization of newly synthesized proteins (312). Simultaneous monitoring of *S*-acylation and protein turnover in the protein H-Ras^{G12V} revealed a palmitate half-life of approximately fifty minutes on the protein. The authors note that these results are consistent with palmitate half-life values determined using radioactive compounds and suggest that their nonradioactive approach to monitoring turnover of this posttranslational modification could be applied generally to any cellular *S*-acylated protein. The successful application of ncAAs to study temporal and spatial aspects of protein synthesis and turnover in complex biological systems is a significant accomplishment. These promising results suggest that further employment of ncAAs to visualize the dynamics of the proteome will continue to yield information about where proteins are synthesized within cells and organisms and how these proteins are transported and degraded within these systems.

Future applications of proteome visualization may benefit greatly from recent methodological developments in chemical labeling methods and cell-selective incorporation of ncAAs into proteins. CuAAC ligations for the detection of proteins have

been quite successful in fixed cells, but their application to live cells may not be possible due to the toxic concentrations of copper needed to promote the chemistries involved. Strain-promoted azide-alkyne ligations provide a nontoxic alternative to labeling newly synthesized proteins, as the reaction of strained alkynes with azides does not require copper to promote the labeling chemistry (249, 313). Beatty et al. demonstrated the utility of such an approach by designing a set of membrane-permeant cyclooctynes functionalized with fluorophores and labeling of azide-tagged proteins in live mammalian cells (314). Multiple recent reports of CuAAC using short labeling times and newly developed ligand catalysts have demonstrated that glycans on the surface of live cells can be labeled without affecting cellular viability (315, 316). These results should also be applicable to the labeling of newly synthesized proteins appearing on the exteriors of live cells, but it is unclear whether these findings will enable labeling of proteins appearing in the interiors of cells without affecting cellular viability. The Lin laboratory has recently demonstrated an alternative chemistry for visualization of newly synthesized proteins (317). Use of the methionine analog homoallylglycine (**87**) and an ultraviolet light-promoted reaction with a fluorescent tag enables detection of proteins containing **87** within a precisely defined area. The method appears to be promising, although further application of the approach will likely require improved specificity, and the use of ultraviolet light may limit the method's use to fixed cells.

The use of translationally active methionine analogs in the visualization of newly synthesized proteins enables researchers to study all newly synthesized proteins within a system simultaneously. However, in nature, interactions between different types of cells within the same organism or between the cells of multiple organisms may result in very

different cellular responses in the components of the interacting systems, leading to the problem of determining which cells express particular proteins at a given time. Ngo et al. have recently demonstrated an approach to selectively visualize the newly synthesized proteins from one cell type in a mixed cell population (318). The key component of this method is the use of a mutant methionyl-tRNA synthetase (MetRS) capable of efficiently and selectively charging **4** (Azidonorleucine, Anl) onto $tRNA^{Met}$. When the mutant MetRS (termed NLL-MetRS) is expressed in a particular cell type, proteins expressed in the presence of **4** incorporate this azide-containing amino acid. In the absence of either the mutant synthetase or **4**, cells remain unlabeled. Thus, only proteins synthesized in cells expressing the mutant synthetase will be tagged with azides upon introduction of **4** into the medium. Ngo and coworkers illustrated this concept using a mixed population of *E. coli* cells expressing two different recombinant proteins. Only proteins expressed in the strain containing NLL-MetRS were labeled during CuAAC as determined by Western blotting and microscopy. This technique was also found to be applicable in a mixed population of *E. coli* and mouse alveolar macrophages. Figure 1.11 depicts the results of labeling experiments following the infection of a macrophage culture with various strains of *E. coli* cells. Only *E. coli* cells expressing the NLL-MetRS were labeled with a TAMRA-alkyne probe after fixing the coculture. Macrophages in the same culture were unlabeled, and positive and negative controls further proved that both cellular populations were synthesizing proteins during the course of the infection and that the labeling in the system was very specific. Hang and coworkers have demonstrated the use of cell-selective labeling in cocultures of *Salmonella typhimurium* and mammalian cells (52). Use of Anl or the long-chain alkyne analog 2-aminooctynoic acid (AOA, **88**) enabled selective labeling of *S.*

typhimurium cells with specificities similar to those observed by Ngo et al. These visualization experiments validate the idea of selectively labeling protein populations expressed in a cell type involved in interactions with other cells. The combined selectivity of ncAA incorporation and CuAAC-mediated fluorescence visualization ensures that a minority protein population can be examined in a vast background of other protein populations and should be applicable to a wide range of systems involving cell-cell interactions.

Protein identification. The same concept of metabolically incorporating chemically reactive ncAAs into newly synthesized proteins or more specific protein populations can also be extended to protein identification techniques. Tagging newly synthesized proteins with affinity purification reagents enables selective separation and enrichment of proteins that have been synthesized in a defined temporal or spatial window or within a particular cellular population. The resulting samples can then be analyzed using mass spectrometry identification techniques. Dietrich et al. first described this technique in 2006, naming the resulting approach bioorthogonal noncanonical amino acid tagging (BONCAT) (319, 320). In this work, the researchers demonstrated several important principles in the development of this technique for use in mammalian cell culture. First, the azide-containing noncanonical amino acid azidohomoalanine (**77**) did not perturb cultured mammalian cells significantly. The visible phenotypes of cells remained the same whether or not **77** was added to cultures, and the protein populations present in cells incubated with methionine or **77** were indistinguishable when examined with autoradiograms of one-dimensional gels. Second, as was found in fluorescence visualization studies, the tagging chemistry was very

specific for proteins that had been synthesized in the presence of **77**. Finally, proteins tagged with an alkyne-FLAG tag could be enriched from the much larger overall protein population with a streptavidin column, digested into smaller peptides on the column, and identified using tandem mass spectrometry. Experiments performed in HEK293 cells identified proteins from a wide range of gene ontological categories, showing that a highly diverse set of proteins can be isolated using BONCAT. Numerous improvements to enriching azide- or alkyne-containing proteins have been reported in support of the BONCAT method. All of these approaches have aimed to replace the on-column trypsinization step of the original BONCAT procedure with separate column purification and digestion steps. For example, Kramer et al. reported selective cleavage of newly synthesized proteins containing **77** at the azide side chains of the proteins (321). Other approaches have involved the synthesis of purification tags designed to enable the use of click chemistry to conjugate affinity reagents to azide- or alkyne-containing proteins, and enrich tagged proteins under mild conditions using appropriate columns. Nessen et al. demonstrated the use of strain-promoted click chemistry for the selective enrichment of newly synthesized proteins from *E. coli* (322). In this work, the use of **77** and a cleavable cyclooctyne resin enabled the enrichment and identification of newly synthesized proteins from whole-cell lysates. Szychowski et al. designed and synthesized five biotin-azide probes that can be cleaved over a wide range of conditions including reducing, mild acidic conditions, and ultraviolet irradiation (323). Using a GFP model system, the authors showed that an acid-cleavable tag enables highly selective conjugation with alkyne-containing proteins and leaves a small mass tag on labeled proteins after cleavage, an important consideration for proteomic studies. Hang and coworkers have developed several

tags that can be used for proteomic studies (52, 324). Application of these tags to labeling and mass spectrometric identification of newly synthesized proteins produced in *S. typhimurium* harboring MetRS-NLL in the presence of AOA (**88**) enabled identification of a large number of proteins from complex samples. Furthermore, samples from cocultures were enriched for proteins expressed in *S. typhimurium*, suggesting that the protein purification procedure successfully separated alkyne-containing proteins from unlabeled proteins.

BONCAT represents a potentially significant advance in the field of proteomics. The ability to selectively enrich a set of newly synthesized proteins from a larger proteomic sample may result in more sensitive detection method for protein subsets of interest. Furthermore, as demonstrated in the case of visualizing new protein synthesis, dynamic aspects of the proteome may be studied by varying the pulse time and duration of ncAAs or by pulsing sequentially with multiple ncAAs. The BONCAT technique may also be used to compare proteomes from two or more different cell samples by applying existing techniques such as SILAC (303) or by combining ncAA incorporation techniques with the use of ICAT (325) or iTRAQ (326) reagents. Finally, protein identification techniques appear to be compatible with the selective metabolic labeling strategy of Ngo et al. and Grammel et al. (52, 318). Proteomic studies with ncAAs take advantage of several aspects of ncAAs including global canonical amino acid replacement, bioorthogonal side chain chemistries, and varied aminoacyl-tRNA synthetase selectivities toward these noncanonical substrates. These several features enable the application of noncanonical amino acids in a broad range of proteomic applications, which should lead to new insights into a number of complex biological systems and processes.

Outlook

The future looks bright for applying noncanonical amino acids to problems in protein science and engineering. The question with ncAAs is no longer, "what kinds of noncanonical amino acids can be incorporated into proteins?" The range of chemical structures employed in experiments with ncAAs is quite impressive, ranging from single-atom changes of canonical residues to ncAAs containing functional groups rarely, if ever, seen in nature. Instead, the question is now, "what types of problems can best be solved by incorporating noncanonical amino acids into proteins?" Recent research has hinted at the varied applications possible with ncAAs. Researchers have employed noncanonical amino acids to investigate problems ranging from atomic-level protein structural questions to organism-wide responses to proteins containing noncanonical amino acids, with time scales of individual experiments ranging over several orders of magnitude. The broad range of experiments performed using noncanonical amino acids suggests that these residues should be thought of as possible tools for studying many problems, and not necessarily as essential tools for studying any particular problem or class of problem. Generally speaking, good candidate problems are ones in which ncAAs offer a route to a substantial increase in the quality of information and/or desired protein properties that can be obtained from a particular set of experiments without drastically raising the level of difficulty of performing these experiments. The several application areas discussed in this review have already started to benefit from the use of noncanonical amino acids according to the criteria outlined above. These benefits are perhaps most apparent in the areas of X-ray crystallography and membrane protein studies, in which a very large amount of data has

been acquired using relatively simple approaches that are not possible without the use of noncanonical amino acids.

A few more questions come to mind regarding research with proteins containing noncanonical amino acids. First, aside from the areas discussed above, what additional problems might benefit from the use of ncAAs in the near future? Biophysical characterizations appear to be among the most accessible problems, especially in light of the number of successful uses of ncAAs in X-ray crystallography and NMR studies. Fluorescence spectroscopy can benefit from both the direct incorporation of fluorescent amino acids and the chemical conjugation of fluorescent labels to reactive chemical functionalities in proteins, and proof-of-principle experiments of both of these approaches already exist in the literature (for examples, see (104, 327-329); see Merkel et al. for a review on the subject of intrinsically blue fluorescent amino acids as tools for protein science (330)). Vibrational spectroscopy of proteins performed after the incorporation of bonds with unique infrared or Raman signatures should continue to grow in importance for protein characterizations. The work of Ye et al. on the membrane protein rhodopsin (212, 213) is an early example of the effective exploitation of IR-active amino acids (discussed in the "Applications" section), and some work with IR-active amino acids in globular proteins also shows promise (331-333). Continued application of ncAAs to study the biological functions of proteins in living cells and organisms will also continue to grow in importance. Along with the visualization and identification of proteomic responses to biological stimuli, the precise control of protein function in biological settings is highly desirable. Proteins with activities or locations controlled by light have proven to be very powerful over the years for studying biological systems (334, 335), and some work suggests that using

ncAAs to achieve such photocontrolled systems may yield a number of new tools in this area (336-340). Given the large number of selective chemical conjugation strategies accessible with ncAAs (249, 251), protein therapeutics may also benefit from applications of ncAAs. Several simple strategies for modifying the pharmacochemical properties of proteins with ncAAs and selective chemistries have been reported in the scientific literature, including examples of protein PEGylation (313, 341-344) and viral surface modifications (345, 346). Commercial applications of these conjugation strategies may yield more drug-like bioconjugates for use as therapeutics. Furthermore, many more subtle changes to protein properties possible with ncAAs, such as those described in conjunction with protein stability, membrane proteins, and immune modulation, may facilitate further improvement of protein-based therapeutics.

What will be the role of in vitro protein synthesis in future work with noncanonical amino acids? The ribosomal production of genetically encoded polymers containing multiple ncAAs, ester linkages, *N*-methyl amino acids, and *N*-substituted glycine residues have all been reported in vitro (58-63). Cyclic peptides, which have many advantages over linear peptides as therapeutic entities, can be formed with the use of noncanonical amino acids in in vitro settings (347-349). These are intriguing molecules, but if the yields of such in vitro productions remain low, applications should be considered carefully and should provide a definite advantage over chemical peptide synthesis techniques. One proposed use of this production approach is as a platform for generating genetically encoded libraries of therapeutically relevant peptide drug candidates. These methodologies may complement or improve upon existing approaches to generating DNA-encoded chemical libraries (350). Genetically templated, highly unnatural molecules may also give researchers tools to

examine fundamental aspects of macromolecular folding and evolution. What kinds of three-dimensional structures can nonbiological polymers attain, how do they evolve, and what do these findings tell us about the functional and evolutionary properties of proteins?

Finally, will engineering additional components of the translational machinery make a large contribution to the ribosomal production of polymers containing noncanonical functionalities? Recent work on the ribosome (64, 65, 116, 117) and elongation factor-TU (351) raises the exciting possibility of expanding the translational capabilities of organisms or in vitro translation systems far beyond the capabilities of existing protein translation machinery. However, even the most promising recent experimental results suggest that engineering the translational machinery will require large numbers of incremental improvements to existing biological machinery. The limits to how far the ribosome and other translational components can be evolved away from their natural functions are still unclear, especially when this evolution is performing in living organisms. Performing evolution in vitro may be more appropriate for the creation of translational machinery with highly unnatural functions, but new in vitro protein expression systems and evolution techniques will likely have to be developed in order to enable this challenging type of translational apparatus engineering. Regardless of how the protein synthesis machinery is engineered, any resulting system should ideally be simple enough to use that it could be widely adopted by the scientific community and versatile enough to be employed for the generation of a wide variety of genetically encoded polymers.

In this review, we have highlighted the use of noncanonical amino acids as tools to study scientific problems. With careful experimental design, ncAAs can greatly augment scientists' abilities to address an extremely broad range of questions; the ncAA toolkit is

very large and very effective. Research with noncanonical amino acids is poised to move beyond the methodological development efforts of a select few researchers to the collaborative exploitation of these methods across numerous scientific disciplines.

Acknowledgements

Work at Caltech on noncanonical amino acids is supported by NIH grant GM62523 and by the Institute for Collaborative Biotechnologies under contract W911NF-09-D-0001 with the Army Research Office.

References

1. Tracewell CA & Arnold FH (2009) Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.* **13,** 3-9.
2. Walsh CT, Garneau-Tsodikova S, & Gatto GJ (2005) Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44,** 7342-7372.
3. Chan WC & White PD (2000) *Fmoc Solid Phase Peptide Synthesis: A Practical Approach* (Oxford University Press, New York).
4. Benoiten NL (2005) *Chemistry of Peptide Synthesis* (CRC, Boca Raton, FL, USA) 1st Ed.
5. Dawson PE & Kent SBH (2000) Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **69,** 923-960.
6. Imperiali B & Taylor EV (2009) Native Chemical Ligation: Semisynthesis of Post-translationally Modified Proteins and Biological Probes. *Protein Eng.,* Nucleic

Acids and Molecular Biology, eds Kohrer C & RajBhandary UL (Springer-Verlag, Berlin), pp 65-96.

7. Muir TW (2003) Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **72,** 249-289.

8. Muir TW (2008) Studying protein structure and function using semisynthesis. *Biopolymers* **90,** 743-750.

9. Flavell RR & Muir TW (2009) Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. *Acc. Chem. Res.* **42,** 107-116.

10. Cherfas J (1982) *Man-Made Life: An Overview of the Science, Technology, and Commerce of Genetic Engineering* (Pantheon Books, New York) First American Ed.

11. Brudno Y & Liu DR (2009) Recent progress toward the templated synthesis and directed evolution of sequence-defined synthetic polymers. *Chem. Biol.* **16,** 265-276.

12. Chapeville F*, et al.* (1962) On role of soluble ribonucleic acid in coding for amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **48,** 1086-1092.

13. Crick FHC (1958) On protein synthesis. *Symp. Soc. Exp. Biol.* **12,** 138-163.

14. Johnson AE, Woodward WR, Herbert E, & Menninger JR (1976) N-epsilon acetyllysine transfer ribonucleic acid - biologically active analog of aminoacyl transfer ribonucleic acids. *Biochemistry* **15,** 569-575.

15. Hecht SM, Alford BL, Kuroda Y, & Kitano S (1978) Chemical aminoacylation of transfer-RNAs. *J. Biol. Chem.* **253,** 4517-4520.

16. Heckler TG*, et al.* (1984) T4 RNA ligase mediated preparation of novel chemically misacylated phenylalanine transfer-RNAs. *Biochemistry* **23,** 1468-1473.

17. Baldini G, Martoglio B, Schachenmann A, Zugliani C, & Brunner J (1988) Mischarging escherichia coli phenylalanine transfer RNA with L-4'-[3-(trifluoromethyl)-3H-diazirin-3-yl]phenylalanine, a photoactivatable analog of phenylalanine. *Biochemistry* **27,** 7951-7959.

18. Happ E, Scalfihapp C, & Chladek S (1987) Aminoacyl derivatives of nucleosides nucleotides and polynucleotides 43. New approach to the synthesis of 2'(3')-O-aminoacyl oligoribonucleotides. *J. Org. Chem.* **52,** 5387-5391.

19. Berg P, Bergmann FH, Ofengand EJ, & Dieckmann M (1961) Enzymic synthesis of amino acyl derivatives of ribonucleic acid. *J. Biol. Chem.* **236,** 1726-1734.

20. Berg P & Ofengand EJ (1958) An enzymatic mechanism for linking amino acids to RNA. *Proc. Natl. Acad. Sci. U. S. A.* **44,** 78-86.

21. Ibba M, Francklyn CS, & Cusack S (2005) *The Aminoacyl-tRNA Synthetases* (Landes Bioscience, Austin, TX).

22. Budisa N (2006) *Engineering the Genetic Code: Expanding the Amino Acid Repertoire for the Design of Novel Proteins* (Wiley-VCH, Weinheim, Germany).

23. Doctor BP & Mudd JA (1963) Species specificity of amino acid acceptor ribonucleic acid and aminoacyl soluble ribonucleic acid synthetases. *J. Biol. Chem.* **238,** 3677-3681.

24. Goodman HM, Abelson J, Landy A, Brenner S, & Smith JD (1968) Amber suppression - a nucleotide change in an anticodon of a tyrosine transfer RNA. *Nature* **217,** 1019-1024.

25. Ibba M & Soll D (2000) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69,** 617-650.

26. Capecchi MR & Gussin GN (1965) Suppression in vitro - identification of a serine-sRNA as a nonsense suppressor. *Science* **149,** 417-422.

27. Engelhar Dl, Webster RE, Wilhelm RC, & Zinder ND (1965) In vitro studies on mechanism of suppression of a nonsense suppressor mutation. *Proc. Natl. Acad. Sci. U. S. A.* **54,** 1791-1797.

28. Smith JD, Abelson JN, Clark BFC, Goodman HM, & Brenner S (1966) Studies on amber suppressor tRNA. *Cold Spring Harbor Symp. Quant. Biol.* **31,** 479-485.

29. Richmond MH (1962) Effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriol. Rev.* **26,** 398-420.

30. Cowie DB & Cohen GN (1957) Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. *Biochim. Biophys. Acta* **26,** 252-261.

31. Hortin G & Boime I (1983) Applications of amino acid analogs for studying co-translational and posttranslational modifications of proteins. *Methods Enzymol.* **96,** 777-784.

32. Wilson MJ & Hatfield DL (1984) Incorporation of modified amino acids into proteins in vivo. *Biochim. Biophys. Acta* **781,** 205-215.

33. Hendrickson WA, Horton JR, & Lemaster DM (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD) - a vehicle for direct determination of 3-dimensional structure. *EMBO J.* **9,** 1665-1672.

34. Link AJ & Tirrell DA (2005) Reassignment of sense codons in vivo. *Methods* **36,** 291-298.

35. Beatty KE*, et al.* (2006) Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew. Chem. Int. Ed.* **45,** 7364-7367.

36. Wiltschi B, Wenger W, Nehring S, & Budisa N (2008) Expanding the genetic code of Saccharomyces cerevisiae with methionine analogues. *Yeast* **25,** 775-786.

37. Budisa N, Wenger W, & Wiltschi B (2010) Residue-specific global fluorination of Candida antarctica lipase B in Pichia pastoris. *Mol. Biosyst.* **6,** 1630-1639.

38. Lepthien S, Merkel L, & Budisa N (2010) In vivo double and triple labeling of proteins using synthetic amino acids. *Angew. Chem. Int. Ed.* **49,** 5446-5450.

39. Merkel L, Schauer M, Antranikian G, & Budisa N (2010) Parallel incorporation of different fluorinated amino acids: on the way to "teflon" proteins. *ChemBioChem* **11,** 1505-1507.

40. Kiick KL, Weberskirch R, & Tirrell DA (2001) Identification of an expanded set of translationally active methionine analogues in Escherichia coli. *FEBS Lett.* **502,** 25-30.

41. Tang Y & Tirrell DA (2001) Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host. *J. Am. Chem. Soc.* **123,** 11089-11090.

42. Wang P, Vaidehi N, Tirrell DA, & Goddard WA (2002) Virtual screening for binding of phenylalanine analogues to phenylalanyl-tRNA synthetase. *J. Am. Chem. Soc.* **124,** 14442-14449.

43. Kast P & Hennecke H (1991) Amino acid substrate specificity of Escherichia coli phenylalanyl-transfer RNA-synthetase altered by distinct mutations. *J. Mol. Biol.* **222,** 99-124.

44. Yoo TH & Tirrell DA (2007) High-throughput screening for Methionyl-tRNA synthetases that enable residue-specific incorporation of noncanonical amino acids into recombinant proteins in bacterial cells. *Angew. Chem. Int. Ed.* **46,** 5340-5343.

45. Link AJ, Vink MKS, & Tirrell DA (2004) Presentation and detection of azide functionality in bacterial cell surface proteins. *J. Am. Chem. Soc.* **126,** 10598-10602.

46. Link AJ*, et al.* (2006) Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 10180-10185.

47. Tanrikulu IC, Schmitt E, Mechulam Y, Goddard WA, & Tirrell DA (2009) Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 15285-15290.

48. Mursinna RS, Lee KW, Briggs JM, & Martinis SA (2004) Molecular dissection of a critical specificity determinant within the amino acid editing domain of leucyl-tRNA synthetase. *Biochemistry* **43,** 155-165.

49. Mursinna RS & Martinis SA (2002) Rational design to block amino acid editing of a tRNA synthetase. *J. Am. Chem. Soc.* **124,** 7286-7287.

50. Tang Y & Tirrell DA (2002) Attenuation of the editing activity of the Escherichia coli Leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry* **41,** 10635-10645.

51. Tang Y, Wang P, Van Deventer JA, Link AJ, & Tirrell DA (2009) Introduction of an aliphatic ketone into recombinant proteins in a bacterial strain that overexpresses an editing-impaired leucyl-tRNA synthetase. *ChemBioChem* **10,** 2188-2190.

52. Grammel M, Zhang MZM, & Hang HC (2010) Orthogonal alkynyl amino acid reporter for selective labeling of bacterial proteomes during infection. *Angew. Chem. Int. Ed.* **49,** 5970-5974.

53. Abdeljabbar DM, Klein TJ, Zhang SY, & Link AJ (2009) A single genomic copy of an engineered methionyl-tRNA synthetase enables robust incorporation of azidonorleucine into recombinant proteins in E-coli. *J. Am. Chem. Soc.* **131,** 17078-17079.

54. Forster AC*, et al.* (2003) Programming peptidomimetic syntheses by translating genetic codes designed de novo. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 6353-6357.

55. Josephson K, Hartman MCT, & Szostak JW (2005) Ribosomal synthesis of unnatural peptides. *J. Am. Chem. Soc.* **127,** 11727-11735.

56. Ohta A, Yamagishi Y, & Suga H (2008) Synthesis of biopolymers using genetic code reprogramming. *Curr. Opin. Chem. Biol.* **12,** 159-167.

57. Ohuchi M, Murakami H, & Suga H (2007) The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Curr. Opin. Chem. Biol.* **11,** 537-542.

58. Subtelny AO, Hartman MCT, & Szostak JW (2008) Ribosomal synthesis of n-methyl peptides. *J. Am. Chem. Soc.* **130,** 6131-6136.

59. Tan ZP, Forster AC, Blacklow SC, & Cornish VW (2004) Amino acid backbone specificity of the Escherichia coli translation machinery. *J. Am. Chem. Soc.* **126,** 12752-12753.

60. Zhang BL*, et al.* (2007) Specificity of translation for N-alkyl amino acids. *J. Am. Chem. Soc.* **129,** 11316-11317.

61. Kawakami T, Murakami H, & Suga H (2008) Ribosomal synthesis of polypeptoids and peptoid-peptide hybrids. *J. Am. Chem. Soc.* **130,** 16861-16863.

62. Kawakami T, Murakami H, & Suga H (2008) Messenger RNA-programmed incorporation of multiple N-methyl-amino acids into linear and cyclic peptides. *Chem. Biol.* **15,** 32-42.

63. Ohta A, Murakami H, Higashimura E, & Suga H (2007) Synthesis of polyester by means of genetic code reprogramming. *Chem. Biol.* **14,** 1315-1322.

64. Dedkova LM, Fahmi NE, Golovine SY, & Hecht SM (2003) Enhanced D-amino acid incorporation into protein by modified ribosomes. *J. Am. Chem. Soc.* **125,** 6616-6617.

65. Dedkova LM, Fahmi NE, Golovine SY, & Hecht SM (2006) Construction of modified ribosomes for incorporation of D-amino acids into proteins. *Biochemistry* **45,** 15541-15551.

66. Brudno Y, Birnbaum ME, Kleiner RE, & Liu DR (2010) An in vitro translation, selection and amplification system for peptide nucleic acids. *Nat. Chem. Biol.* **6,** 148-155.

67. Bain JD, Glabe CG, Dix TA, Chamberlin AR, & Diala ES (1989) Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* **111,** 8013-8014.

68. Noren CJ, Anthonycahill SJ, Griffith MC, & Schultz PG (1989) A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244,** 182-188.

69. Drabkin HJ, Park HJ, & RajBhandary UL (1996) Amber suppression in mammalian cells dependent upon expression of an Escherichia coli aminoacyl-tRNA synthetase gene. *Mol. Cell. Biol.* **16,** 907-913.

70. Furter R (1998) Expansion of the genetic code: Site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. *Protein Sci.* **7,** 419-426.

71. Liu DR, Magliery TJ, Pasternak M, & Schultz PG (1997) Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 10092-10097.

72. Liu DR & Schultz PG (1999) Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 4780-4785.

73. Drabkin HJ, Estrella M, & Rajbhandary UL (1998) Initiator-elongator discrimination in vertebrate tRNAs for protein synthesis. *Mol. Cell. Biol.* **18,** 1459-1466.

74. Kowal AK, Kohrer C, & RajBhandary UL (2001) Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 2268-2273.

75. Wang L & Schultz PG (2001) A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **8,** 883-890.

76. Wang L, Brock A, Herberich B, & Schultz PG (2001) Expanding the genetic code of Escherichia coli. *Science* **292,** 498-500.

77. Sakamoto K*, et al.* (2002) Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* **30,** 4692-4699.

78. Chin JW*, et al.* (2003) An expanded eukaryotic genetic code. *Science* **301,** 964-967.

79. Liu WS, Brock A, Chen S, Chen SB, & Schultz PG (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **4,** 239-244.

80. Wang WY*, et al.* (2007) Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat. Neurosci.* **10,** 1063-1072.

81. Cornish VW, Mendel D, & Schultz PG (1995) Probeing protein structure and function with an expanded genetic code. *Angew. Chem. Int. Ed.* **34,** 621-633.

82. Wang L & Schultz PG (2005) Expanding the genetic code. *Angew. Chem. Int. Ed.* **44,** 34-66.

83. Wang L, Xie J, & Schultz PG (2006) Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 225-249.

84. Wang Q, Parrish AR, & Wang L (2009) Expanding the genetic code for biological studies. *Chem. Biol.* **16,** 323-336.

85. Johnson JA, Lu YY, Van Deventer JA, & Tirrell DA (2010) Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Chem. Biol.* **14,** 774-780.

86. Liu CC & Schultz PG (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79,** 413-444.

87. Young TS & Schultz PG (2010) Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J. Biol. Chem.* **285,** 11039-11044.

88. Ryu YH & Schultz PG (2006) Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nat. Methods* **3,** 263-265.

89. Young TS, Ahmad I, Yin JA, & Schultz PG (2010) An enhanced system for unnatural amino acid mutagenesis in E. coli. *J. Mol. Biol.* **395,** 361-374.

90. Guo JT, Melancon CE, Lee HS, Groff D, & Schultz PG (2009) Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew. Chem. Int. Ed.* **48,** 9148-9151.

91. Cellitti SE*, et al.* (2008) In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **130,** 9268-9281.

92. Hammill JT, Miyake-Stoner S, Hazen JL, Jackson JC, & Mehl RA (2007) Preparation of site-specifically labeled fluorinated proteins for F-19-NMR structural characterization. *Nat. Protoc.* **2,** 2601-2607.

93. Jackson JC, Hammill JT, & Mehl RA (2007) Site-specific incorporation of a F-19-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J. Am. Chem. Soc.* **129,** 1160-1166.

94. Chen S, Schultz PG, & Brock A (2007) An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in Saccharomyces cerevisiae. *J. Mol. Biol.* **371,** 112-122.

95. Wang Q & Wang L (2008) New methods enabling efficient incorporation of unnatural amino acids in yeast. *J. Am. Chem. Soc.* **130,** 6066-6067.

96. Young TS, Ahmad I, Brock A, & Schultz PG (2009) Expanding the genetic repertoire of the methylotrophic yeast Pichia pastoris. *Biochemistry* **48,** 2643-2653.

97. Wang F, Robbins S, Guo JT, Shen WJ, & Schultz PG (2010) Genetic incorporation of unnatural amino acids into proteins in Mycobacterium tuberculosis. *PLoS ONE* **5,** e9354.

98. Hino N*, et al.* (2005) Protein photo-crosslinking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat. Methods* **2,** 201-206.

99. Zhang ZW*, et al.* (2004) Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 8882-8887.

100. Neumann H, Peak-Chew SY, & Chin JW (2008) Genetically encoding N-epsilon-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4,** 232-234.

101. Nguyen DP*, et al.* (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNA(CUA) pair and click chemistry. *J. Am. Chem. Soc.* **131,** 8720-8721.

102. Fekner T, Li X, Lee MM, & Chan MK (2009) A pyrrolysine analogue for protein click chemistry. *Angew. Chem. Int. Ed.* **48,** 1633-1635.

103. Chen PR*, et al.* (2009) A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem. Int. Ed.* **48,** 4052-4055.

104. Summerer D*, et al.* (2006) A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9785-9789.

105. Oki K, Sakamoto K, Kobayashi T, Sasaki HM, & Yokoyama S (2008) Transplantation of a tyrosine editing domain into a tyrosyl-tRNA synthetase variant enhances its specificity for a tyrosine analog. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 13298-13303.

106. Ai HW, Shen WJ, Brustad E, & Schultz PG (2010) Genetically encoded alkenes in yeast. *Angew. Chem. Int. Ed.* **49,** 935-937.

107. Neumann H, Slusarczyk AL, & Chin JW (2010) De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs. *J. Am. Chem. Soc.* **132,** 2142-2144.

108. Hohsaka T, Ashizuka Y, Taira H, Murakami H, & Sisido M (2001) Incorporation of nonnatural amino acids into proteins by using various four-base codons in an Escherichia coli in vitro translation system. *Biochemistry* **40,** 11060-11064.

109. Rodriguez EA, Lester HA, & Dougherty DA (2006) In vivo incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 8650-8655.

110. Anderson JC*, et al.* (2004) An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 7566-7571.

111. Taki M, Matsushita J, & Sisido M (2006) Expanding the genetic code in a mammalian cell line by the introduction of four-base codon/anticodon pairs. *ChemBioChem* **7,** 425-428.

112. Hohsaka T, Ashizuka Y, Sasaki H, Murakami H, & Sisido M (1999) Incorporation of two different nonnatural amino acids independently into a single protein through extension of the genetic code. *J. Am. Chem. Soc.* **121,** 12194-12195.

113. Huang Y*, et al.* (2010) A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in Escherichia coli. *Mol. Biosyst.* **6,** 683-686.

114. Wan W*, et al.* (2010) A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angew. Chem. Int. Ed.* **49,** 3211-3214.

115. Rackham O & Chin JW (2005) A network of orthogonal ribosome-mRNA pairs. *Nat. Chem. Biol.* **1,** 159-166.

116. Neumann H, Wang KH, Davis L, Garcia-Alai M, & Chin JW (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464,** 441-444.

117. Wang KH, Neumann H, Peak-Chew SY, & Chin JW (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* **25,** 770-777.

118. Tirrell DA (2009) Reinterpreting the Genetic Code: Implications for Macromolecular Design, Evolution and Analysis. *Physical Biology: From Atoms to Medicine*, ed Zewail AH (Imperial College Press, London), pp 165-188.

119. Connor RE & Tirrell DA (2007) Non-canonical amino acids in protein polymer design. *Polymer Reviews* **47,** 9-28.

120. Beatty KE & Tirrell DA (2009) Noncanonical amino acids in protein science and engineering. *Protein Eng.,* Nucleic Acids and Molecular Biology, eds Kohrer C & RajBhandary UL (Springer, Berlin), pp 127-154.

121. Wu X & Schultz PG (2009) Synthesis at the interface of chemistry and biology. *J. Am. Chem. Soc.* **131,** 12497-12515.

122. McPherson A (2009) *Introduction to Macromolecular Crystallography* (Wiley-Blackwell, Hoboken, NJ) 2nd Ed.

123. Kasai N & Kakudo M (2005) *X-Ray Diffraction By Macromolecules* (Springer-Verlag, Berlin).

124. Graves BJ*, et al.* (1990) Structure of interleukin 1-alpha at 2.7-A resolution. *Biochemistry* **29,** 2679-2684.

125. Yang W, Hendrickson WA, Crouch RJ, & Satow Y (1990) Structure of ribonuclease-H phased at 2-A resolution by MAD analysis of the selenomethionyl protein. *Science* **249,** 1398-1405.

126. Wu H, Lustbader JW, Liu Y, Canfield RE, & Hendrickson WA (1994) Structure of human chorionic-gonadotropin at 2.6-angstrom resolution from MAD analysis of the selenomethionyl protein. *Structure* **2,** 545-558.

127. Lustbader JW*, et al.* (1995) The expression, characterization, and crystallization of wild-type and selenomethionyl human chorionic-gonadotropin. *Endocrinology* **136,** 640-650.

128. Bellizzi JJ, Widom J, Kemp CW, & Clardy J (1999) Producing selenomethionine-labeled proteins with a baculovirus expression vector system. *Structure Fold. Des.* **7,** R263-R267.

129. McWhirter SM*, et al.* (1999) Crystallographic analysis of CD40 recognition and signaling by human TRAF2. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 8408-8413.

130. Rice LM, Earnest TN, & Brunger AT (2000) Single-wavelength anomalous diffraction phasing revisited. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **56,** 1413-1420.

131. Walsh MA, Evans G, Sanishvili R, Dementieva I, & Joachimiak A (1999) MAD data collection - current trends. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **55,** 1726-1732.

132. Hendrickson WA (1999) Maturation of MAD phasing for the determination of macromolecular structures. *J. Synchrotron Rad.* **6,** 845-851.

133. Hendrickson WA & Ogata CM (1997) Phase determination from multiwavelength anomalous diffraction measurements. *Macromolecular Crystallography, Pt A,* Methods in Enzymology, ed Carter CW (Elsevier), Vol 276, pp 494-523.

134. Boles JO*, et al.* (1995) Expression, characterization and crystallographic analysis of telluromethionyl dihydrofolate reductase. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **51,** 731-739.

135. Boles JO*, et al.* (1994) Bio-incorporation of telluromethionine into buried residues of dihydrofolate reductase. *Nat. Struct. Biol.* **1,** 283-284.

136. Budisa N*, et al.* (1995) High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *Eur. J. Biochem.* **230,** 788-796.

137. Sanchez J-F, Hoh F, Strub M-P, Aumelas A, & Dumas C (2002) Structure of the Cathelicidin Motif of Protegrin-3 Precursor: Structural Insights into the Activation Mechanism of an Antimicrobial Protein. *Structure* **10,** 1363-1370.

138. Strub MP*, et al.* (2003) Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing. *Structure* **11,** 1359-1367.

139. Bae JH*, et al.* (2001) Incorporation of beta-selenolo[3,2-b]pyrrolyl-alanine into proteins for phase determination in protein X-ray crystallography. *J. Mol. Biol.* **309,** 925-936.

140. Boles JO, Henderson J, Hatch D, & Silks LA (2002) Synthesis and incorporation of [6,7]-selenatryptophan into dihydrofolate reductase. *Biochem. Biophys. Res. Commun.* **298,** 257-261.

141. Xie JM*, et al.* (2004) The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* **22,** 1297-1301.

142. Sakamoto K*, et al.* (2009) Genetic encoding of 3-iodo-L-tyrosine in Escherichia coli for single-wavelength anomalous dispersion phasing in protein crystallography. *Structure* **17,** 335-344.

143. Hibler DW*, et al.* (1989) [4] Isotopic labeling with hydrogen-2 and carbon-13 to compare conformations of proteins and mutants generated by site-directed mutagenesis, I. *Methods Enzymol.*, (Academic Press), Vol 177, pp 74-86.

144. Muchmore DC*, et al.* (1989) [3] Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol.*, (Academic Press), Vol 177, pp 44-73.

145. Tugarinov V, Hwang PM, & Kay LE (2004) Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu. Rev. Biochem.* **73,** 107-146.

146. Danielson MA & Falke JJ (1996) Use of F-19 NMR to probe protein structure and conformational changes. *Annu. Rev. Biophys. Biomol. Struct.* **25,** 163-195.

147. Frieden C, Hoeltzli SD, & Bann JG (2004) The preparation of F-19-labeled proteins for NMR studies. *Energetics of Biological Macromolecules, Pt E,* Methods in Enzymology, eds Holt JM, Johnson ML, & Ackers GK), Vol 380, pp 400-415.

148. Bann JG, Pinkner J, Hultgren SJ, & Frieden C (2002) Real-time and equilibrium F-19-NMR studies reveal the role of domain-domain interactions in the folding of the chaperone PapD. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 709-714.

149. Cleve P, Robinson V, Duewel HS, & Honek JF (1999) Difluoromethionine as a Novel 19F NMR Structural Probe for Internal Amino Acid Packing in Proteins. *J. Am. Chem. Soc.* **121,** 8475-8478.

150. Eichler JF, Cramer JC, Kirk KL, & Bann JG (2005) Biosynthetic incorporation of fluorohistidine into proteins in E-coli: A new probe of macromolecular structure. *ChemBioChem* **6,** 2170-2173.

151. Rule GS, Pratt EA, Simplaceanu V, & Ho C (1987) Nuclear magnetic resonance and molecular genetic studies of the membrane-bound D-lactate dehydrogenase of Escherichia coli. *Biochemistry* **26,** 549-556.

152. Peersen OB, Pratt EA, Truong HTN, Ho C, & Rule GS (1990) Site-specific incorporation of 5-fluorotryptophan as a probe of the structure and function of the membrane-bound D-lactate dehydrogenase of Escherichia coli - a F-19 nuclear magnetic resonance study. *Biochemistry* **29,** 3256-3262.

153. Sun ZY*, et al.* (1993) A F-19-NMR study of the membrane-binding region of D-lactate dehydrogenase of Escherichia coli. *Protein Sci.* **2,** 1938-1947.

154. Ropson IJ & Frieden C (1992) Dynamic NMR spectral analysis and protein folding - identification of a highly populated folding intermediate of rat intestinal fatty acid-binding protein by F-19 NMR. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 7222-7226.

155. Sun ZY, Pratt EA, Simplaceanu V, & Ho C (1996) A F-19-NMR study of the equilibrium unfolding of membrane-associated d-lactate dehydrogenase of Escherichia coli. *Biochemistry* **35,** 16502-16509.

156. Hoeltzli SD & Frieden C (1994) F-19 NMR spectroscopy of [6-F-19]tryptophan-labeled Escherichia coli dihydrofolate reductase - equilibrium folding and ligand-binding studies. *Biochemistry* **33,** 5502-5509.

157. Shu Q & Frieden C (2004) Urea-dependent unfolding of murine adenosine deaminase: sequential destabilization as measured by F-19 NMR. *Biochemistry* **43,** 1432-1439.

158. Shu Q & Frieden C (2005) Relation of enzyme activity to local/global stability of murine adenosine deaminase: F-19 NMR studies. *J. Mol. Biol.* **345,** 599-610.

159. Li H & Frieden C (2005) Phenylalanine side chain behavior of the intestinal fatty acid-binding protein - The effect of urea on backbone and side chain stability. *J. Biol. Chem.* **280,** 38556-38561.

160. Li H & Frieden C (2006) Fluorine-19 NMR studies on the acid state of the intestinal fatty acid binding protein. *Biochemistry* **45,** 6272-6278.

161. Khan F, Kuprov I, Craggs TD, Hore PJ, & Jackson SE (2006) 19F NMR studies of the native and denatured states of green fluorescent protein. *J. Am. Chem. Soc.* **128,** 10729-10737.

162. Li H & Frieden C (2007) Comparison of C40/82A and P27A C40/82A barstar mutants using F-19 NMR. *Biochemistry* **46,** 4337-4347.

163. Hoeltzli SD & Frieden C (1995) Stopped-flow NMR spectroscopy - real-time unfolding studies of 6-F-19-tryptophan-labeled Escherichia coli dihydrofolate reductase. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 9318-9322.

164. Minks C, Huber R, Moroder L, & Budisa N (1999) Atomic Mutations at the Single Tryptophan Residue of Human Recombinant Annexin V: Effects on Structure, Stability, and Activity. *Biochemistry* **38,** 10649-10659.

165. Bann JG & Frieden C (2004) Folding and domain-domain interactions of the chaperone PapD measured by F-19 NMR. *Biochemistry* **43,** 13775-13786.

166. Li CG*, et al.* (2010) Protein F-19 NMR in Escherichia coli. *J. Am. Chem. Soc.* **132,** 321-327.

167. Jones DH*, et al.* (2010) Site-specific labeling of proteins with NMR-active unnatural amino acids. *J. Biomol. NMR* **46,** 89-100.

168. Lampe JN*, et al.* (2008) Ligand-induced conformational heterogeneity of cytochrome P450 CYP119 identified by 2D NMR spectroscopy with the unnatural amino acid C-13-p-methoxyphenylalanine. *J. Am. Chem. Soc.* **130,** 16168-16169.

169. Mendel D*, et al.* (1992) Probing protein stability with unnatural amino acids. *Science* **256,** 1798-1802.

170. Ellman JA, Mendel D, & Schultz PG (1992) Site-specific incorporation of novel backbone structures into proteins. *Science* **255,** 197-200.

171. Cornish VW, Kaplan MI, Veenstra DL, Kollman PA, & Schultz PG (1994) Stabilizing and destabilizing effects of placing beta-branched amino acids in protein alpha-helices. *Biochemistry* **33,** 12022-12031.

172. Thorson JS, Chapman E, & Schultz PG (1995) Analysis of hydrogen bonding strengths in proteins using unnatural amino acids. *J. Am. Chem. Soc.* **117,** 9361-9362.

173. Shin IJ, Ting AY, & Schultz PG (1997) Analysis of backbone hydrogen bonding in a beta-turn of staphylococcal nuclease. *J. Am. Chem. Soc.* **119,** 12667-12668.

174. Chapman E, Thorson JS, & Schultz PG (1997) Mutational analysis of backbone hydrogen bonds in Staphylococcal nuclease. *J. Am. Chem. Soc.* **119,** 7151-7152.

175. Koh JT, Cornish VW, & Schultz PG (1997) An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. *Biochemistry* **36,** 11314-11322.

176. Thorson JS, Chapman E, Murphy EC, Schultz PG, & Judice JK (1995) Linear free energy analysis of hydrogen bonding in proteins. *J. Am. Chem. Soc.* **117,** 1157-1158.

177. Guo JT, Wang JY, Anderson JC, & Schultz PG (2008) Addition of an alpha-hydroxy acid to the genetic code of bacteria. *Angew. Chem. Int. Ed.* **47,** 722-725.

178. Dunitz JD (2004) Organic fluorine: Odd man out. *ChemBioChem* **5,** 614-621.

179. Marsh ENG (2000) Towards the nonstick egg: designing fluorous proteins. *Chem. Biol.* **7,** R153-R157.

180. Yoder NC & Kumar K (2002) Fluorinated amino acids in protein design and engineering. *Chem. Soc. Rev.* **31,** 335-341.

181. Yoder NC, Yuksel D, Dafik L, & Kumar K (2006) Bioorthogonal noncovalent chemistry: fluorous phases in chemical biology. *Curr. Opin. Chem. Biol.* **10,** 576-583.

182. Tang Y*, et al.* (2001) Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability. *Angew. Chem. Int. Ed.* **40,** 1494-1496.

183. Son S, Tanrikulu IC, & Tirrell DA (2006) Stabilization of bzip peptides through incorporation of fluorinated aliphatic residues. *ChemBioChem* **7,** 1251-1257.

184. Montclare JK, Son S, Clark GA, Kumar K, & Tirrell DA (2009) Biosynthesis and stability of coiled-coil peptides containing (2S,4R)-5,5,5-trifluoroleucine and (2S,4S)-5,5,5-trifluoroleucine. *ChemBioChem* **10,** 84-86.

185. Van Deventer JA, Fisk JD, & Tirrell DA (2011) Homoisoleucine: a translationally active leucine surrogate of expanded hydrophobic surface area. *ChemBioChem* **12,** 700-702.

186. Kwon OH*, et al.* (2010) Hydration dynamics at fluorinated protein surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 17101-17106.

187. Cirino PC, Tang Y, Takahashi K, Tirrell DA, & Arnold FH (2003) Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. *Biotechnol. Bioeng.* **83,** 729-734.

188. Montclare JK & Tirrell DA (2006) Evolving proteins of novel composition. *Angew. Chem. Int. Ed.* **45,** 4518-4521.

189. Yoo TH, Link AJ, & Tirrell DA (2007) Evolution of a fluorinated green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 13887-13890.

190. Wang P, Tang Y, & Tirrell DA (2003) Incorporation of trifluoroisoleucine into proteins in vivo. *J. Am. Chem. Soc.* **125,** 6900-6906.

191. Moroder L & Budisa N (2010) Synthetic biology of protein folding. *Chemphyschem* **11,** 1181-1188.

192. Steiner T*, et al.* (2008) Synthetic biology of proteins: tuning GFPs folding and stability with fluoroproline. *PLoS ONE* **3,** e1680.

193. Edwardraja S, Sriram S, Govindan R, Budisa N, & Lee SG (2011) Enhancing the thermal stability of a single-chain Fv fragment by in vivo global fluorination of the proline residues. *Mol. Biosyst.* **7,** 258-265.

194. Wolschner C*, et al.* (2009) Design of anti- and pro-aggregation variants to assess the effects of methionine oxidation in human prion protein. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 7756-7761.

195. Levine RL, Moskovitz J, & Stadtman ER (2000) Oxidation of methionine in proteins: Roles in antioxidant defense and cellular regulation. *Iubmb Life* **50,** 301-307.

196. Stadtman ER, Moskovitz J, Berlett BS, & Levine RL (2002) Cyclic oxidation and reduction of protein methionine residues is an important antioxidant mechanism. *Mol. Cell. Biochem.* **234,** 3-9.

197. Hoshi T & Heinemann SH (2001) Regulation of cell function by methionine oxidation and reduction. *J. Physiol. (Lond).* **531,** 1-11.

198. Elofsson A & von Heijne G (2007) Membrane protein structure: Prediction versus reality. *Annu. Rev. Biochem.* **76,** 125-140.

199. Hopkins AL & Groom CR (2002) The druggable genome. *Nat. Rev. Drug Discov.* **1,** 727-730.

200. Klabunde T & Hessler G (2002) Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **3,** 929-944.

201. Lagerstrom MC & Schioth HB (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat. Rev. Drug Discov.* **7,** 339-357.

202. Wagner S, Bader ML, Drew D, & de Gier JW (2006) Rationalizing membrane protein overexpression. *Trends Biotechnol.* **24,** 364-371.

203. Link AJ & Georgiou G (2007) Advances and challenges in membrane protein expression. *AICHE J.* **53,** 752-756.

204. Johansson LC, Wohri AB, Katona G, Engstrom S, & Neutze R (2009) Membrane protein crystallization from lipidic phases. *Curr. Opin. Struct. Biol.* **19,** 372-378.

205. Neher E & Sakmann B (1976) Single-channel currents recorded from membrane of denervated frog muscle fibers. *Nature* **260,** 799-802.

206. Dougherty DA (2009) In Vivo Studies of Receptors and Ion Channels with Unnatural Amino Acids. *Protein Eng.,* Nucleic Acids and Molecular Biology, eds Kohrer C & RajBhandary UL (Springer-Verlag, Berlin), Vol 22, pp 231-254.

207. Gallivan JP, Lester HA, & Dougherty DA (1997) Site-specific incorporation of biotinylated amino acids to identify surface-exposed residues in integral membrane proteins. *Chem. Biol.* **4,** 739-749.

208. England PM, Lester HA, Davidson N, & Dougherty DA (1997) Site-specific, photochemical proteolysis applied to ion channels in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 11025-11030.

209. England PM, Lester HA, & Dougherty DA (1999) Mapping disulfide connectivity using backbone ester hydrolysis. *Biochemistry* **38,** 14409-14415.

210. Turcatti G*, et al.* (1997) Fluorescent labeling of NK2 receptor at specific sites in vivo and fluorescence energy transfer analysis of NK2 ligand-receptor complexes. *Receptors Channels*, (Harwood Acad Publ Gmbh), pp 201-207.

211. Turcatti G*, et al.* (1996) Probing the structure and function of the tachykinin neurokinin-2 receptor through biosynthetic incorporation of fluorescent amino acids at specific sites. *J. Biol. Chem.* **271,** 19991-19998.

212. Ye SX, Huber T, Vogel R, & Sakmar TP (2009) FTIR analysis of GPCR activation using azido probes. *Nat. Chem. Biol.* **5,** 397-399.

213. Ye SX*, et al.* (2010) Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. *Nature* **464,** 1386-1389.

214. Tao X, Lee A, Limapichat W, Dougherty DA, & MacKinnon R (2010) A gating charge transfer center in voltage sensors. *Science* **328,** 67-73.

215. Nowak MW*, et al.* (1995) Nicotinic receptor binding site probed with unnatural amino acid incorporation in intact cells. *Science* **268,** 439-442.

216. Dougherty DA (2008) Physical organic chemistry on the brain. *J. Org. Chem.* **73,** 3667-3673.

217. Hanek AP, Lester HA, & Dougherty DA (2008) A stereochemical test of a proposed structural feature of the nicotinic acetylcholine receptor. *J. Am. Chem. Soc.* **130,** 13216-13218.

218. Lummis SCR*, et al.* (2005) Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* **438,** 248-252.

219. Beene DL, Price KL, Lester HA, Dougherty DA, & Lummis SCR (2004) Tyrosine residues that control binding and gating in the 5-hydroxytryptamine(3) receptor revealed by unnatural amino acid mutagenesis. *J. Neurosci.* **24,** 9097-9104.

220. Price KL*, et al.* (2008) A hydrogen bond in loop a is critical for the binding and function of the 5-HT3 receptor. *Biochemistry* **47,** 6370-6377.

221. Blum AP, Lester HA, & Dougherty DA (2010) Nicotinic pharmacophore: The pyridine N of nicotine and carbonyl of acetylcholine hydrogen bond across a subunit interface to a backbone NH. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 13206-13211.

222. England PM, Zhang YN, Dougherty DA, & Lester HA (1999) Backbone mutations in transmembrane domains of a ligand-gated ion channel: Implications for the mechanism of gating. *Cell* **96,** 89-98.

223. Dang H, England PM, Farivar SS, Dougherty DA, & Lester HA (2000) Probing the role of a conserved M1 proline residue in 5-hydroxytryptamine(3) receptor gating. *Mol. Pharmacol.* **57,** 1114-1122.

224. Gleitsman KR, Kedrowski SMA, Lester HA, & Dougherty DA (2008) An intersubunit hydrogen bond in the nicotinic acetylcholine receptor that contributes to channel gating. *J. Biol. Chem.* **283,** 35638-35643.

225. Cashin AL, Torrice MM, McMenimen KA, Lester HA, & Dougherty DA (2007) Chemical-scale studies on the role of a conserved aspartate in preorganizing the agonist binding site of the nicotinic acetylcholine receptor. *Biochemistry* **46,** 630-639.

226. Gleitsman KR, Lester HA, & Dougherty DA (2009) Probing the role of backbone hydrogen bonding in a critical beta sheet of the extracellular domain of a Cys-loop receptor. *ChemBioChem* **10,** 1385-1391.

227. Li LT*, et al.* (2001) The tethered agonist approach to mapping ion channel proteins toward a structural model for the agonist binding site of the nicotinic acetylcholine receptor. *Chem. Biol.* **8,** 47-58.

228. Petersson EJ, Choi A, Dahan DS, Lester HA, & Dougherty DA (2002) A perturbed pK(a) at the binding site of the nicotinic acetylcholine receptor: Implications for nicotine binding. *J. Am. Chem. Soc.* **124,** 12662-12663.

229. Dougherty DA (1996) Cation-pi interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science* **271,** 163-168.

230. Ma JC & Dougherty DA (1997) The cation-pi interaction. *Chem. Rev.* **97,** 1303-1324.

231. Zhong WG*, et al.* (1998) From ab initio quantum mechanics to molecular neurobiology: A cation-pi binding site in the nicotinic receptor. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 12088-12093.

232. Beene DL*, et al.* (2002) Cation-pi interactions in ligand recognition by serotonergic (5-HT3A) and nicotinic acetylcholine receptors: The anomalous binding properties of nicotine. *Biochemistry* **41,** 10262-10269.

233. Cashin AL, Petersson EJ, Lester HA, & Dougherty DA (2005) Using physical chemistry to differentiate nicotinic from cholinergic agonists at the nicotinic acetylcholine receptor. *J. Am. Chem. Soc.* **127,** 350-356.

234. Xiu XA, Puskar NL, Shanata JAP, Lester HA, & Dougherty DA (2009) Nicotine binding to brain receptors requires a strong cation-pi interaction. *Nature* **458,** 534-537.

235. Mu TW, Lester HA, & Dougherty DA (2003) Different binding orientations for the same agonist at homologous receptors: A lock and key or a simple wedge? *J. Am. Chem. Soc.* **125,** 6850-6851.

236. Lummis SCR, Beene DL, Harrison NJ, Lester HA, & Dougherty DA (2005) A cation-pi binding interaction with a tyrosine in the binding site of the GABA(C) receptor. *Chem. Biol.* **12,** 993-997.

237. Padgett CL, Hanek AP, Lester HA, Dougherty DA, & Lummis SCR (2007) Unnatural amino acid mutagenesis of the GABA(A) receptor binding site residues reveals a novel cation-pi interaction between GABA and beta(2)Tyr97. *J. Neurosci.* **27,** 886-892.

238. Pless SA*, et al.* (2008) A cation-pi interaction in the binding site of the glycine receptor is mediated by a phenylalanine residue. *J. Neurosci.* **28,** 10937-10942.

239. Santarelli VP, Eastwood AL, Dougherty DA, Ahern CA, & Horn R (2007) Calcium block of single sodium channels: Role of a pore-lining aromatic residue. *Biophys. J.* **93,** 2341-2349.

240. Santarelli VP, Eastwood AL, Dougherty DA, Horn R, & Ahern CA (2007) A cation-pi interaction discriminates among sodium channels that are either sensitive or resistant to tetrodotoxin block. *J. Biol. Chem.* **282,** 8044-8051.

241. Ahern CA, Eastwood AL, Dougherty DA, & Horn R (2008) Electrostatic contributions of aromatic residues in the local anesthetic receptor of voltage-gated sodium channels. *Circul. Res.* **102,** 86-94.

242. Ahern CA, Eastwood AL, Lester HA, Dougherty DA, & Horn R (2006) A cation-pi interaction between extracellular TEA and an aromatic residue in potassium channels. *J. Gen. Physiol.* **128,** 649-657.

243. McMenimen KA, Petersson EJ, Lester HA, & Dougherty DA (2006) Probing the Mg2+ blockade site of an N-methyl-D-aspartate (NMDA) receptor with unnatural amino acid mutagenesis. *ACS Chem. Biol.* **1,** 227-234.

244. Torrice MM, Bower KS, Lester HA, & Dougherty DA (2009) Probing the role of the cation-pi interaction in the binding sites of GPCRs using unnatural amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 11919-11924.

245. Ye SX*, et al.* (2008) Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis. *J. Biol. Chem.* **283,** 1525-1533.

246. Huang LY*, et al.* (2008) Unnatural amino acid replacement in a yeast G protein-coupled receptor in its native environment. *Biochemistry* **47,** 5638-5648.

247. Hermanson GT (2008) *Bioconjugate Techniques* (Academic Press, London) 2 Ed.

248. Simon MD*, et al.* (2007) The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128,** 1003-1012.

249. Sletten EM & Bertozzi CR (2009) Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* 6974-6998.

250. Best MD (2009) Click chemistry and bioorthogonal reactions: unprecedented selectivity in the labeling of biological molecules. *Biochemistry* **48,** 6571-6584.

251. de Graaf AJ, Kooijman M, Hennink WE, & Mastrobattista E (2009) Nonnatural amino acids for site-specific protein conjugation. *Bioconj. Chem.* **20,** 1281-1295.

252. Seebeck FP & Szostak JW (2006) Ribosomal synthesis of dehydroalanine-containing peptides. *J. Am. Chem. Soc.* **128,** 7150-7151.

253. Wang J, Schiller SM, & Schultz PG (2007) A biosynthetic route to dehydroalanine-containing proteins. *Angew. Chem. Int. Ed.* **46,** 6849-6851.

254. Bernardes GJL, Chalker JM, Errey JC, & Davis BG (2008) Facile conversion of cysteine and alkyl cysteines to dehydroalanine on protein surfaces: Versatile and switchable access to functionalized proteins. *J. Am. Chem. Soc.* **130,** 5052-5053.

255. Guo JT, Wang JY, Lee JS, & Schultz PG (2008) Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. *Angew. Chem. Int. Ed.* **47,** 6399-6401.

256. Chalker JM & Davis BG (2010) Chemical mutagenesis: selective post-expression interconversion of protein amino acid residues. *Curr. Opin. Chem. Biol.* **14,** 781-789.

257. Nguyen DP, Alai MMG, Kapadnis PB, Neumann H, & Chin JW (2009) Genetically encoding N-epsilon-methyl-L-lysine in recombinant histones. *J. Am. Chem. Soc.* **131,** 14194-14195.

258. Nguyen DP, Alai MMG, Virdee S, & Chin JW (2010) Genetically directing epsilon-N, N-dimethyl-L-lysine in recombinant histones. *Chem. Biol.* **17,** 1072-1076.

259. Virdee S, Ye Y, Nguyen DP, Komander D, & Chin JW (2010) Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nat. Chem. Biol.* **6,** 750-757.

260. Eger S, Scheffner M, Marx A, & Rubini M (2010) Synthesis of defined ubiquitin dimers. *J. Am. Chem. Soc.* **132,** 16337-16339.

261. Gamblin DP, Scanlan EM, & Davis BG (2009) Glycoprotein synthesis: an update. *Chem. Rev.* **109,** 131-163.

262. Floyd N, Vijayakrishnan B, Koeppe JR, & Davis BG (2009) Thiyl glycosylation of Olefinic Proteins: S-Linked Glycoconjugate Synthesis13. *Angew. Chem. Int. Ed.* **48,** 7798-7802.

263. Liu HT, Wang L, Brock A, Wong CH, & Schultz PG (2003) A method for the generation of glycoprotein mimetics. *J. Am. Chem. Soc.* **125,** 1702-1703.

264. Van Kasteren SI, Kramer HB, Gamblin DP, & Davis BG (2007) Site-selective glycosylation of proteins: creating synthetic glycoproteins. *Nat. Protoc.* **2,** 3185-3194.

265. van Kasteren SI*, et al.* (2007) Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **446,** 1105-1109.

266. Liu CC & Schultz PG (2006) Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nat. Biotechnol.* **24,** 1436-1440.

267. Liu CC, Brustad E, Liu WS, & Schultz PG (2007) Crystal structure of a biosynthetic sulfo-hirudin complexed to thrombin. *J. Am. Chem. Soc.* **129,** 10648-10649.

268. Liu CC, Choe H, Farzan M, Smider VV, & Schultz PG (2009) Mutagenesis and evolution of sulfated antibodies using an expanded genetic code. *Biochemistry* **48,** 8891-8898.

269. Liu CC*, et al.* (2008) Protein evolution with an expanded genetic code. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 17688-17693.

270. Neumann H*, et al.* (2009) A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol. Cell* **36,** 153-163.

271. Kang TJ, Yuzawa S, & Suga H (2008) Expression of histone H3 tails with combinatorial lysine modifications under the reprogrammed genetic code for the investigation of epigenetic markers. *Chem. Biol.* **15,** 1166-1174.

272. Lammers M, Neumann H, Chin JW, & James LC (2010) Acetylation regulates Cyclophilin A catalysis, immunosuppression and HIV isomerization. *Nat. Chem. Biol.* **6,** 331-337.

273. Xie JM, Supekova L, & Schultz PG (2007) A genetically encoded metabolically stable analogue of phosphotyrosine in Escherichia coli. *ACS Chem. Biol.* **2,** 474-478.

274. Neumann H, Hazen JL, Weinstein J, Mehl RA, & Chin JW (2008) Genetically encoding protein oxidative damage. *J. Am. Chem. Soc.* **130,** 4028-4033.

275. Xu R*, et al.* (2004) Site-specific incorporation of the mucin-type N-acetylgalactosamine-alpha-O-threonine into protein in Escherichia coli. *J. Am. Chem. Soc.* **126,** 15654-15655.

276. Zhang ZW*, et al.* (2004) A new strategy for the synthesis of glycoproteins. *Science* **303,** 371-373.

277. Xu R*, et al.* (2009) Site-specific incorporation of the mucin-type N-acetylgalactosamine-alpha-O-threonine into protein in Escherichia coli. *J. Am. Chem. Soc.* **131,** 13883-13883.

278. Zhang ZW*, et al.* (2009) Retraction. *Science* **326,** 1187.

279. Fahmi NE, Dedkova L, Wang BX, Golovine S, & Hecht SM (2007) Site-specific incorporation of glycosylated serine and tyrosine derivatives into proteins. *J. Am. Chem. Soc.* **129,** 3586-3597.

280. Restifo NP (1996) The new vaccines: building viruses that elicit antitumor immunity. *Curr. Opin. Immunol.* **8,** 658-663.

281. Dalum I*, et al.* (1999) Therapeutic antibodies elicited by immunization against TNF-alpha. *Nat. Biotechnol.* **17,** 666-669.

282. Hackett CJ & Harn DA (2006) *Vaccine Adjuvants: Immunological and Clinical Principles* (Humana Press, Totowa, NJ).

283. Grünewald J*, et al.* (2008) Immunochemical termination of self-tolerance. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 11276-11280.

284. Grünewald J*, et al.* (2009) Mechanistic studies of the immunochemical termination of self-tolerance with unnatural amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 4337-4342.

285. Crick FHC (1968) Origin of genetic code. *J. Mol. Biol.* **38,** 367-&.

286. Knight RD, Freeland SJ, & Landweber LF (2001) Rewiring the keyboard evolvability of the genetic code. *Nat. Rev. Genet.* **2,** 49-58.

287. Wong JTF (2005) Coevolution theory of the genetic code at age thirty. *Bioessays* **27,** 416-425.

288. Bacher JM, Hughes RA, Wong JTF, & Ellington AD (2004) Evolving new genetic codes. *Trends Ecol. Evol.* **19,** 69-75.

289. Wong JTF (1983) Membership mutation of the genetic code - loss of fitness by tryptophan. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 6303-6306.

290. Bacher JM & Ellington AD (2001) Selection and characterization of Escherichia coli variants capable of growth on an otherwise toxic tryptophan analogue. *J. Bacteriol.* **183,** 5414-5425.

291. Bacher JM, Bull JJ, & Ellington AD (2003) Evolution of phage with chemically ambiguous proteomes. *BMC Evol. Biol.* **3,** 5414-5425.

292. Turner NJ (2009) Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5,** 568-574.

293. Kazlauskas RJ & Bornscheuer UT (2009) Finding better protein engineering strategies. *Nat. Chem. Biol.* **5,** 526-529.

294. Shivange AV, Marienhagen J, Mundhada H, Schenk A, & Schwaneberg U (2009) Advances in generating functional diversity for directed protein evolution. *Curr. Opin. Chem. Biol.* **13,** 19-25.

295. Jackel C, Kast P, & Hilvert D (2008) Protein design by directed evolution. *Ann. Rev. Biophys.* **37,** 153-173.

296. Dougherty M & Arnold F (2009) Directed evolution: new parts and optimized function. *Curr. Opin. Biotechnol.* **20,** 486-491.

297. Brustad EM & Arnold FH (2011) Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.* doi:10.1016/j.cbpa.2010.1011.1020.

298. Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, & Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79-88.

299. Choe H*, et al.* (2003) Tyrosine sulfation of human antibodies contributes to recognition of the CCR5 binding region of HIV-1 gp120. *Cell* **114,** 161-170.

300. Huang CC*, et al.* (2007) Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science* **317,** 1930-1934.

301. Liu CC*, et al.* (2009) Evolution of proteins with genetically encoded "chemical warheads." *J. Am. Chem. Soc.* **131,** 9616-9617.

302. Beynon RJ & Pratt JM (2005) Metabolic labeling of proteins for proteomics. *Mol. Cell. Proteomics* **4,** 857-872.

303. Mann M (2006) Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **7,** 952-958.

304. Caro LG, Kolb JA, & Vantubergen RP (1962) High-resolution autoradiography 1. Methods. *J. Cell Biol.* **15,** 173-188.

305. Kiick KL, Saxon E, Tirrell DA, & Bertozzi CR (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 19-24.

306. van Hest JCM, Kiick KL, & Tirrell DA (2000) Efficient incorporation of unsaturated methionine analogues into proteins in vivo. *J. Am. Chem. Soc.* **122,** 1282-1288.

307. Beatty KE, Xie F, Wang Q, & Tirrell DA (2005) Selective dye-labeling of newly synthesized proteins in bacterial cells. *J. Am. Chem. Soc.* **127,** 14150-14151.

308. Beatty KE & Tirrell DA (2008) Two-color labeling of temporally defined protein populations in mammalian cells. *Biorg. Med. Chem. Lett.* **18,** 5995-5999.

309. Sutton MA & Schuman EM (2006) Dendritic protein synthesis, synaptic plasticity, and memory. *Cell* **127,** 49-58.

310. Dieterich DC*, et al.* (2010) In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat. Neurosci.* **13,** 897-905.

311. Tcherkezian J, Brittis PA, Thomas F, Roux PP, & Flanagan JG (2010) Transmembrane receptor DCC associates with protein synthesis machinery and regulates translation. *Cell* **141,** 632-644.

312. Zhang MM, Tsou LK, Charron G, Raghavan AS, & Hang HC (2010) Tandem fluorescence imaging of dynamic S-acylation and protein turnover. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 8627-8632.

313. Debets MF*, et al.* (2010) Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition. *Chem. Commun.* **46,** 97-99.

314. Beatty KE*, et al.* (2010) Live-cell imaging of cellular proteins by a strain-promoted azide-alkyne cycloaddition. *ChemBioChem* **11,** 2092-2095.

315. del Amo DS*, et al.* (2010) Biocompatible copper(I) catalysts for in vivo imaging of glycans. *J. Am. Chem. Soc.* **132,** 16893-16899.

316. Hong V, Steinmetz NF, Manchester M, & Finn MG (2010) Labeling live cells by copper-catalyzed alkyne-azide click chemistry. *Bioconj. Chem.* **21,** 1912-1916.

317. Song WJ*, et al.* (2010) A metabolic alkene reporter for spatiotemporally controlled imaging of newly synthesized proteins in mammalian cells. *ACS Chem. Biol.* **5,** 875-885.

318. Ngo JT*, et al.* (2009) Cell-selective metabolic labeling of proteins. *Nat. Chem. Biol.* **5,** 715-717.

319. Dieterich DC, Link AJ, Graumann J, Tirrell DA, & Schuman EM (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9482-9487.

320. Dieterich DC*, et al.* (2007) Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat. Protoc.* **2,** 532-540.

321. Kramer G*, et al.* (2009) Identification and quantitation of newly synthesized proteins in Escherichia coli by enrichment of azidohomoalanine-labeled peptides with diagonal chromatography. *Mol. Cell. Proteomics* **8,** 1599-1611.

322. Nessen MA*, et al.* (2009) Selective enrichment of azide-containing peptides from complex mixtures. *J. Proteome Res.* **8,** 3702-3711.

323. Szychowski J*, et al.* (2010) Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *J. Am. Chem. Soc.* **132,** 18351-18360.

324. Yang Y-Y, Grammel M, Raghavan AS, Charron G, & Hang HC (2010) Comparative analysis of cleavable azobenzene-based affinity tags for bioorthogonal chemical proteomics. *Chem. Biol.* **17,** 1212-1222.

325. Gygi SP*, et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17,** 994-999.

326. Ross PL*, et al.* (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3,** 1154-1169.

327. Lepthien S, Hoesl MG, Merkel L, & Budisa N (2008) Azatryptophans endow proteins with intrinsic blue fluorescence. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 16095-16100.

328. Brustad EM, Lemke EA, Schultz PG, & Deniz AA (2008) A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J. Am. Chem. Soc.* **130,** 17664-17665.

329. Zhang ZW*, et al.* (2003) A new strategy for the site-specific modification of proteins in vivo. *Biochemistry* **42,** 6735-6746.

330. Merkel L, Hoesl MG, Albrecht M, Schmidt A, & Budisa N (2010) Blue fluorescent amino acids as in vivo building blocks for proteins. *ChemBioChem* **11,** 305-314.

331. Taskent-Sezgin H*, et al.* (2010) Azidohomoalanine: a conformationally sensitive IR probe of protein folding, protein structure, and electrostatics. *Angew. Chem. Int. Ed.* **49,** 7473-7475.

332. Groff D, Thielges MC, Cellitti S, Schultz PG, & Romesberg FE (2009) Efforts toward the direct experimental characterization of enzyme microenvironments: tyrosine 100 in dihydrofolate reductase. *Angew. Chem. Int. Ed.* **48,** 3478-3481.

333. Schultz KC*, et al.* (2006) A genetically encoded infrared probe. *J. Am. Chem. Soc.* **128,** 13984-13985.

334. Lee HM, Larson DR, & Lawrence DS (2009) Illuminating the chemistry of life: design, synthesis, and applications of "caged" and related photoresponsive compounds. *ACS Chem. Biol.* **4,** 409-427.

335. Mayer G & Heckel A (2006) Biologically active molecules with a "light switch." *Angew. Chem. Int. Ed.* **45,** 4900-4921.

336. Chou CJ, Young DD, & Deiters A (2009) A light-activated DNA polymerase. *Angew. Chem. Int. Ed.* **48,** 5950-5953.

337. Edwards WF, Young DD, & Deiters A (2009) Light-activated Cre recombinase as a tool for the spatial and temporal control of gene function in mammalian cells. *ACS Chem. Biol.* **4,** 441-445.

338. Wang JY*, et al.* (2010) A biosynthetic route to photoclick chemistry on proteins. *J. Am. Chem. Soc.* **132,** 14812-14818.

339. Gautier A*, et al.* (2010) Genetically encoded photocontrol of protein localization in mammalian cells. *J. Am. Chem. Soc.* **132,** 4086-4088.

340. Groff D, Wang F, Jockusch S, Turro NJ, & Schultz PG (2010) A new strategy to photoactivate green fluorescent protein. *Angew. Chem. Int. Ed.* **49,** 7677-7679.

341. Cazalis CS, Haller CA, Sease-Cargo L, & Chaikof EL (2004) C-terminal site-specific PEGylation of a truncated thrombomodulin mutant with retention of full bioactivity. *Bioconj. Chem.* **15,** 1005-1009.

342. Deiters A, Cropp TA, Summerer D, Mukherji M, & Schultz PG (2004) Site-specific PEGylation of proteins containing unnatural amino acids. *Biorg. Med. Chem. Lett.* **14,** 5743-5745.

343. Schoffelen S, Lambermon MHL, van Eldijk MB, & van Hest JCM (2008) Site-specific modification of Candida antarctica lipase B via residue-specific incorporation of a non-canonical amino acid. *Bioconj. Chem.* **19,** 1127-1131.

344. Teeuwen RLM*, et al.* (2009) "Clickable" elastins: elastin-like polypeptides functionalized with azide or alkyne groups. *Chem. Commun.* 4022-4024.

345. Strable E*, et al.* (2008) Unnatural amino acid incorporation into virus-like particles. *Bioconj. Chem.* **19,** 866-875.

346. Prasuhn DE*, et al.* (2008) Plasma clearance of bacteriophage Q beta particles as a function of surface charge. *J. Am. Chem. Soc.* **130,** 1328-1334.

347. Goto Y*, et al.* (2008) Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS Chem. Biol.* **3,** 120-129.

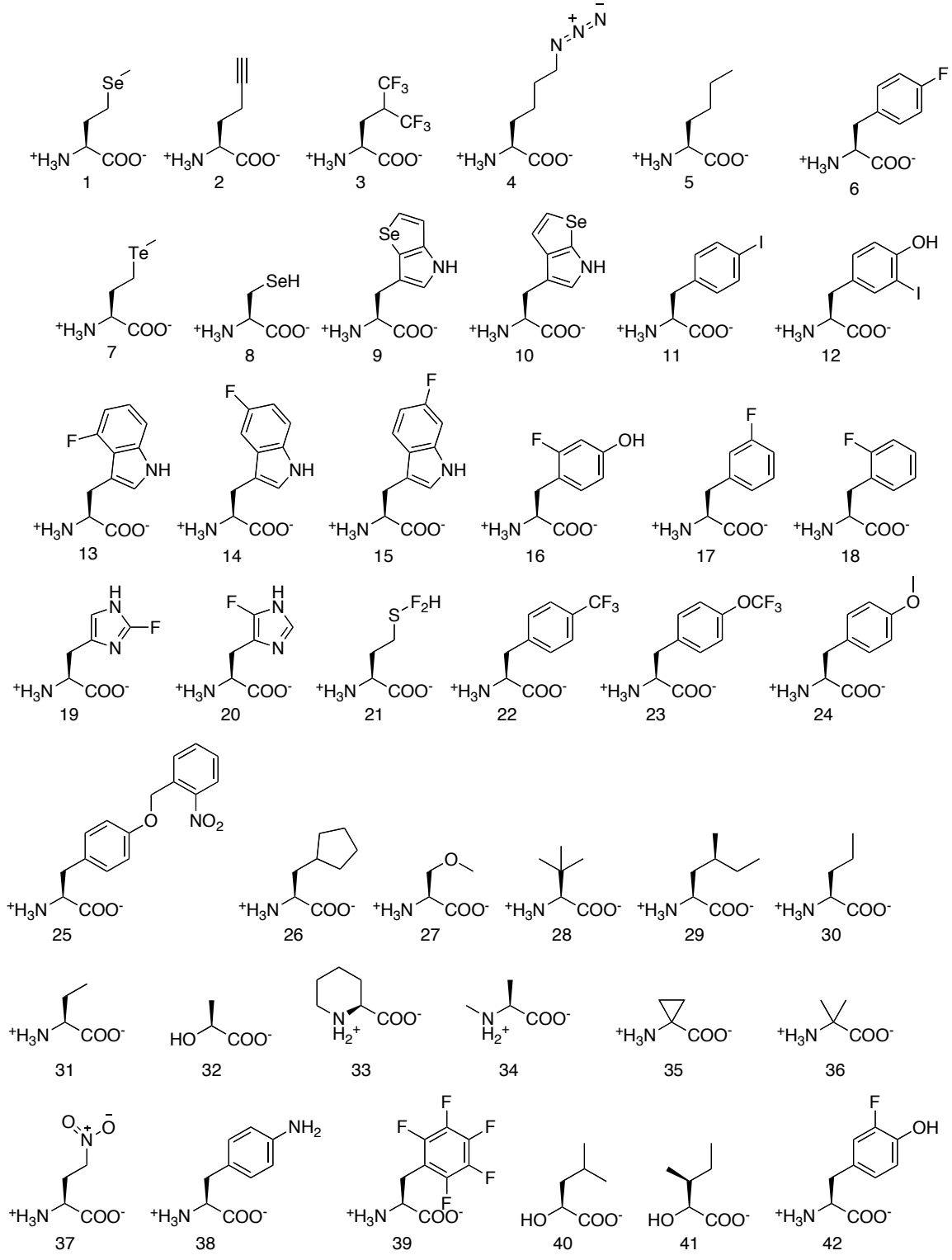
348. Sako Y, Morimoto J, Murakami H, & Suga H (2008) Ribosomal synthesis of bicyclic peptides via two orthogonal inter-side-chain reactions. *J. Am. Chem. Soc.* **130,** 7232-7234.

349. Kawakami T*, et al.* (2009) Diverse backbone-cyclized peptides via codon reprogramming. *Nat. Chem. Biol.* **5,** 888-890.

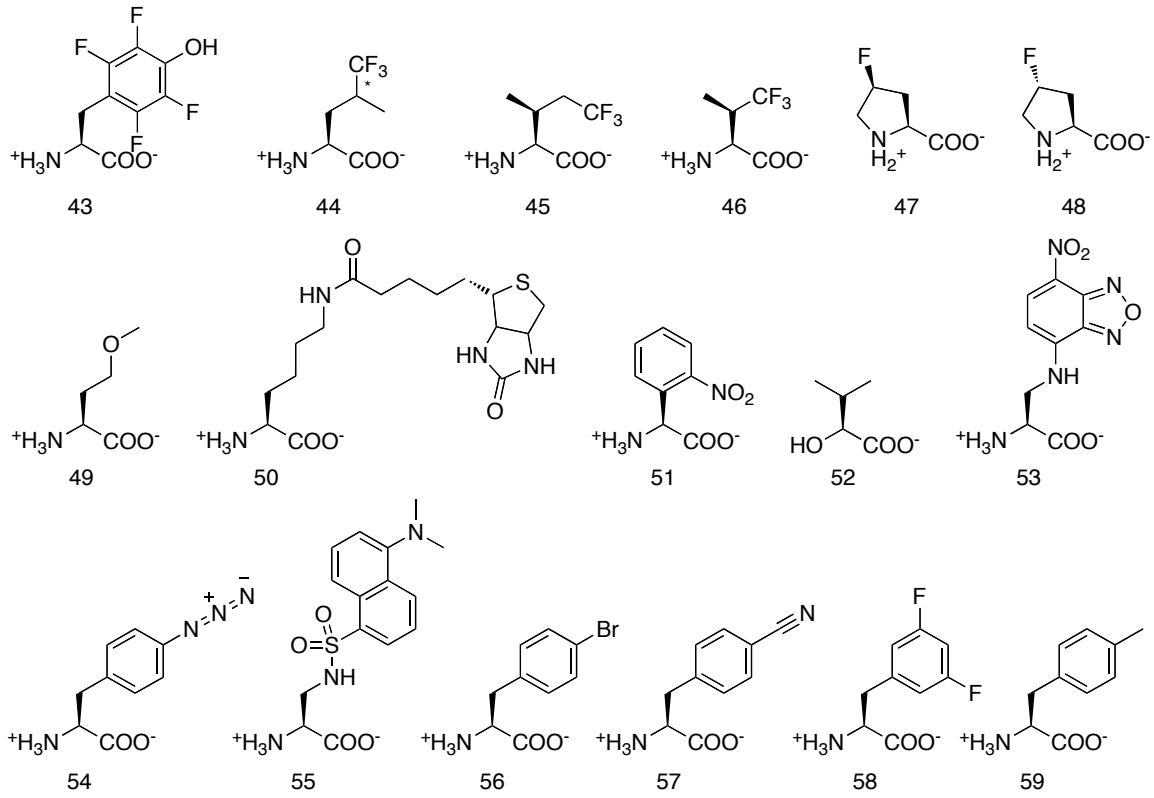
350. Clark MA (2010) Selecting chemicals: the emerging utility of DNA-encoded libraries. *Curr. Opin. Chem. Biol.* **14,** 396-403.

351. Doi Y, Ohtsuki T, Shimizu Y, Ueda T, & Sisido M (2007) Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system. *J. Am. Chem. Soc.* **129,** 14458-14462.

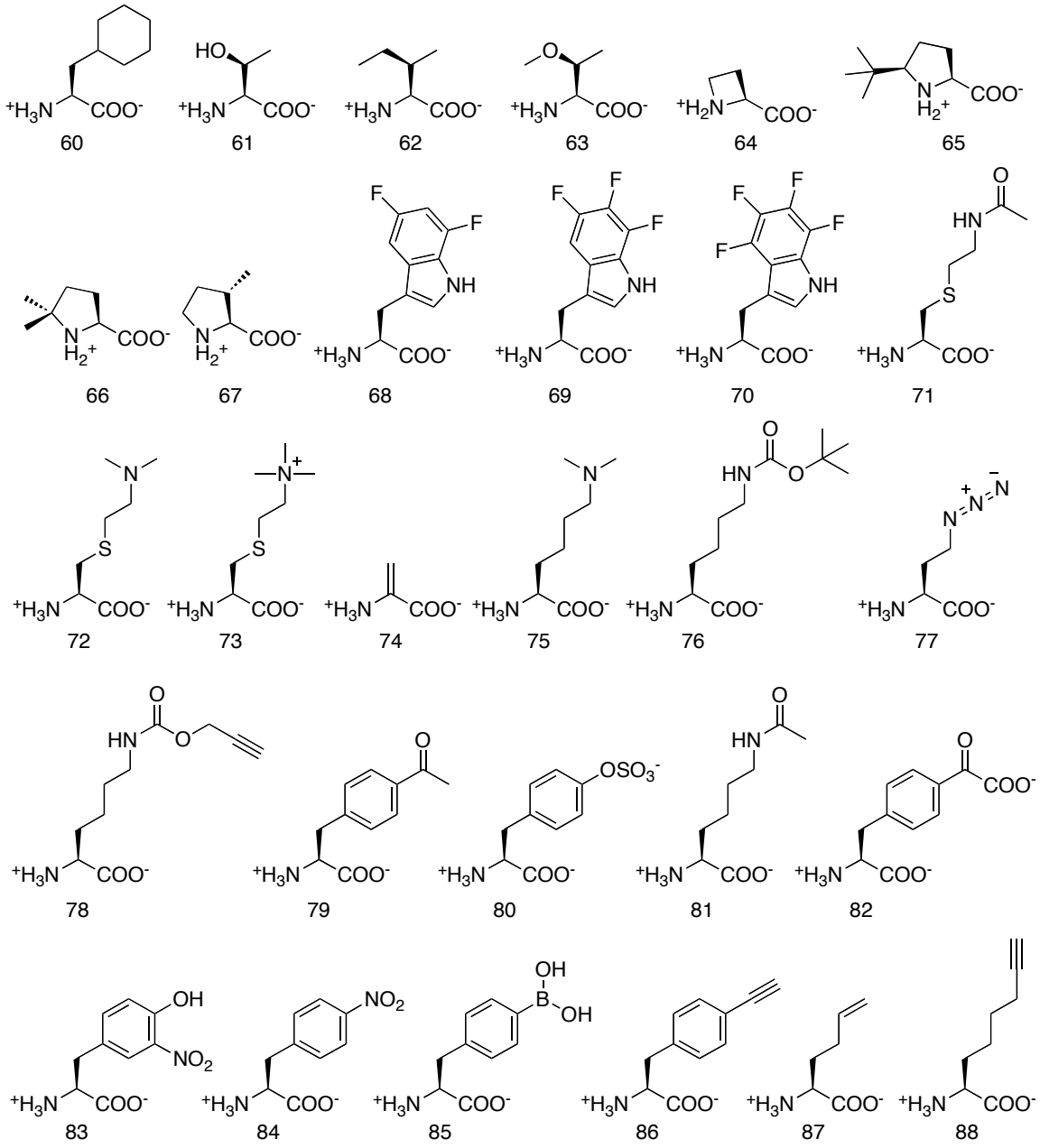
352. Pemble CW, Johnson LC, Kridel SJ, & Lowther WT (2007) Crystal structure of the thioesterase domain of human fatty acid synthase inhibited by Orlistat. *Nat. Struct. Mol. Biol.* **14,** 704-709.



Scheme 1.1. Noncanonical amino acids. Caption to follow.



Scheme 1.1. Noncanonical amino acids, part 2. Caption to follow.



Scheme 1.1. Noncanonical amino acids, part 3. **1**, selenomethionine. **2**, homopropargylglycine. **3**, hexafluoroleucine. **4**, azidonorleucine. **5**, norleucine. **6**, *p*-fluorophenylalanine. **7**, telluromethionine. **8**, selenocysteine. **9**, β-selenolo[3,2-*b*]pyrolyl-alanine **10**, [6,7]selenatryptophan. **11**, *p*-iodophenylalanine. **12**, 3-iodotyrosine. **13**, 4-fluorotryptophan. **14**, 5-fluorotryptophan. **15**, 6-fluorotryptophan. **16**, *m*-fluorotyrosine. **17**, *m*-fluorophenylalanine. **18**, *o*-fluorophenylalanine. **19**, 2-fluorohistidine. **20**,

4-fluorohistidine. **21**, difluoromethionine. **22**, *p*-trifluoromethylphenylalanine. **23**, *p*-trifluoromethoxyphenylalanine. **24**, *p*-methoxyphenylalanine (also called *O*-methyltyrosine). **25**, *O*-nitrobenzyltyrosine. **26**, amino-3-cyclopentylpropanoic acid. **27**, *O*-methylserine. **28**, tert-leucine. **29**, 2-amino-4-methylhexanoic acid (homoisoleucine). **30**, norvaline. **31**, ethylglycine. **32**, α-hydroxyalanine. **33**, pipecolic acid. **34**, *N*-methylalanine. **35**, cyclopropylglycine. **36**, α-aminoisobutyric acid. **37**, γ-nitroglutamate. **38**, *p*-aminophenylalanine. **39**, pentafluorophenylalanine. **40**, α-hydroxyleucine. **41**, α-hydroxyisoleucine. **42**, *p*-fluorotyrosine. **43**, tetrafluorotyrosine. **44**, trifluoroleucine. **45**, trifluoroisoleucine. **46**, trifluorovaline. **47**, 4*R*-fluoroproline. **48**, 4*S*-fluoroproline. **49**, methoxinine. **50**, biocytin. **51**, (2-nitrophenyl)glycine. **52**, α-hydroxyvaline. **53**, 3-*N*-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)-2,3-diaminopropionic acid (NBD-Dap). **54**, *p*-azidophenylalanine. **55**, dansylalanine. **56**, *p*-bromophenylalanine. **57**, *p*-cyanophenylalanine. **58**, 3,5-difluorophenylalanine. **59**, *p*-methylphenylalanine. **60**, cyclohexylalanine, **61**, *allo*-threonine. **62**, *allo*-isoleucine. **63**, *allo*-*O*-methylthreonine. **64**, azetidine-2-carboxylic acid. **65**, 5-tert-butylproline. **66**, 5,5-dimethylproline. **67**, 3*S*-methylproline. **68**, 5,7-difluorotryptophan. **69**, 5,6,7-trifluorotryptophan. **70**, 4,5,6,7-tetrafluorotryptophan. **71**, sulfated acetyllysine. **72**, sulfated dimethyllysine. **73**, sulfated trimethyllysine. **74** dehydroalanine. **75**, dimethyllysine **76**, *N*-tertbutyloxycarbonyllysine. **77**, azidohomoalanine. **78**, alkyne analog of pyrrolysine. **79**, *p*-acetylphenylalanine. **80**, sulfotyrosine. **81**, N^{ϵ}-acetyllysine. **82**, *p*-carboxymethylphenylalanine. **83**, nitrotyrosine. **84**, *p*-nitrophenylalanine. **85**, *p*-boronophenylalanine. **86**, *p*-alkynylphenylalanine. **87**, homoallylglycine. **88**, 2-aminooctynoic acid.

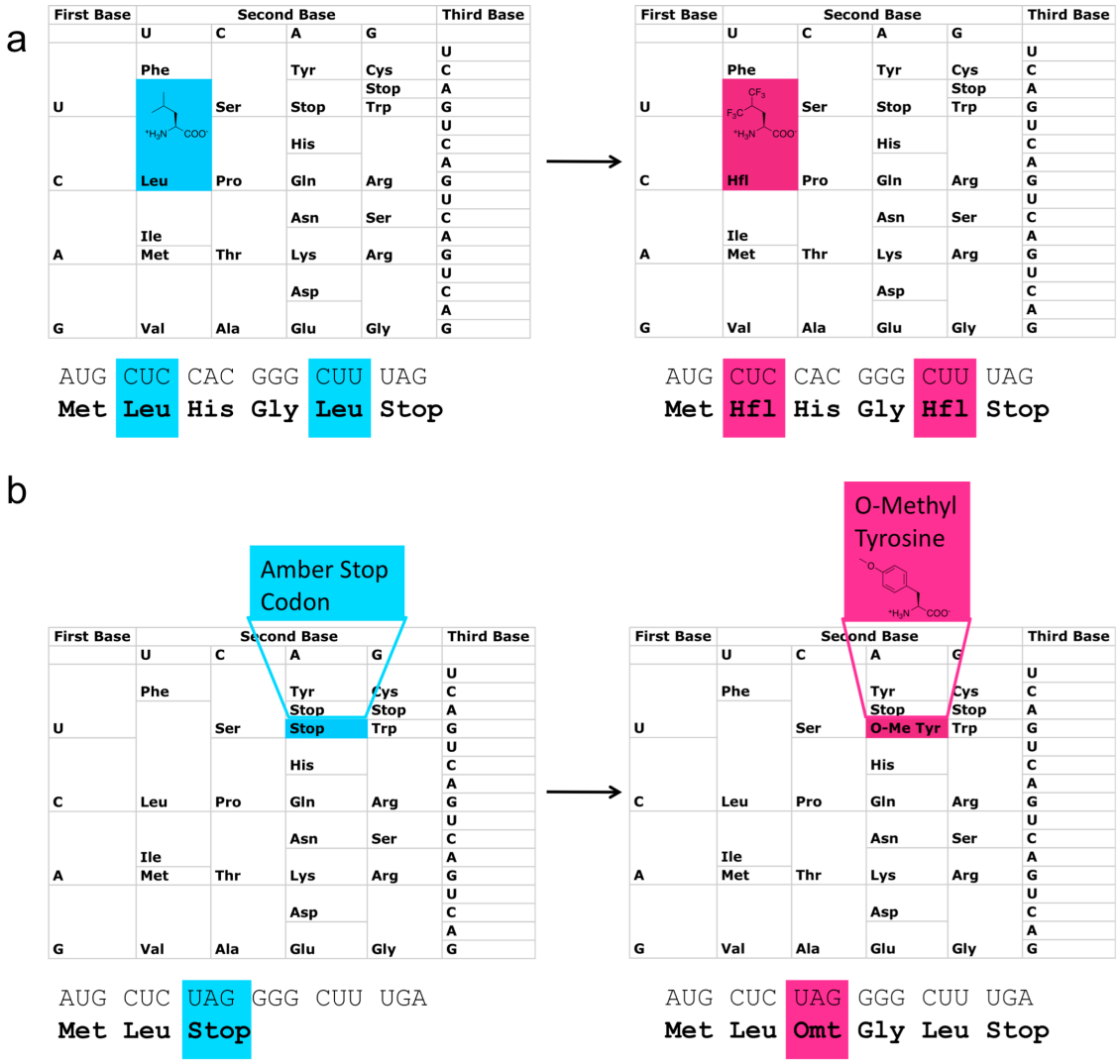


Figure 1.1. Overview of strategies for genetically incorporating noncanonical amino acids into proteins. (*a*) Residue-specific incorporation. A set of codons specifying one of the twenty canonical amino acids is "reprogrammed" to code for a ncAA. In this case, the six leucine (Leu) codons of the *E. coli* genetic code have been reassigned to the fluorinated leucine analog hexafluoroleucine (Hfl, **3**). (*b*) Site-specific incorporation. A stop codon is converted into a sense codon. The depicted example shows the amber stop codon being reprogrammed to code for *O*-methyltyrosine (Omt, **24**).

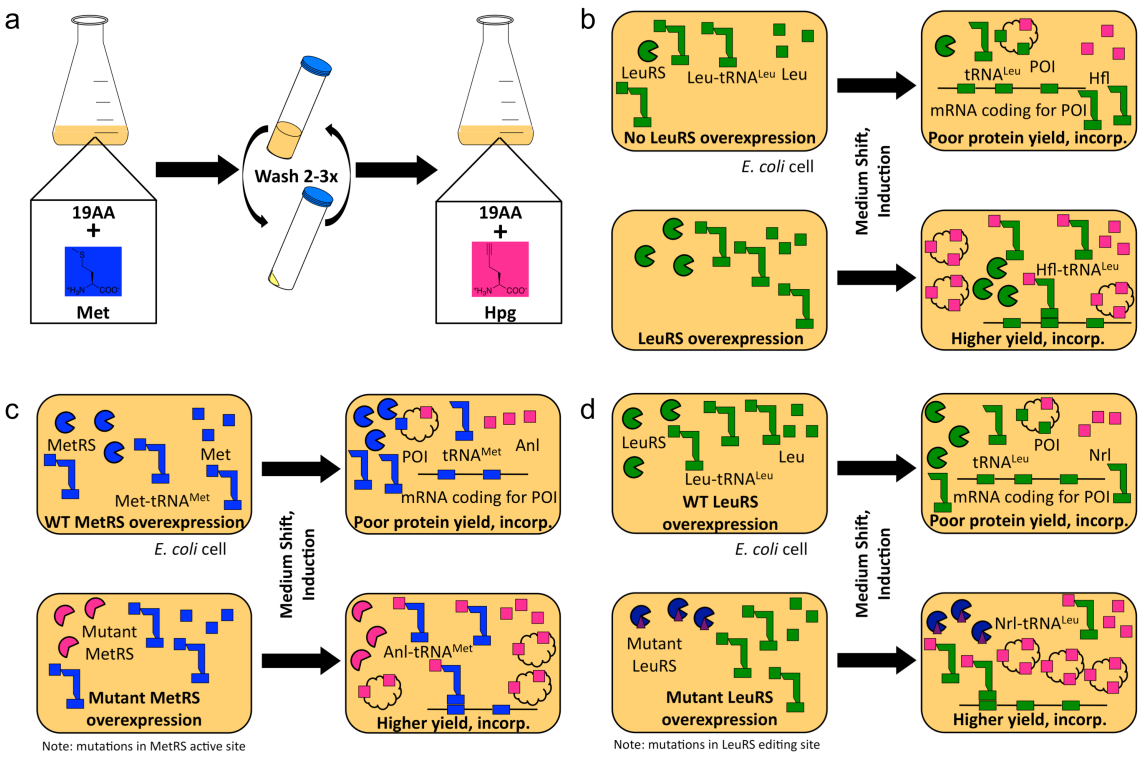


Figure 1.2. Methods for residue-specific incorporation of noncanonical amino acids into proteins. (*a*) Medium shift procedure to produce proteins containing the methionine (Met) analog homopropargylglycine (Hpg, **2**). Methionine-auxotrophic *E. coli* cells are first grown in medium containing twenty canonical amino acids. After reaching a particular optical density, the culture is pelleted and washed two to three times in an isotonic salt solution to remove methionine from the cells. After the final wash, cells are pelleted and resuspended in medium containing nineteen canonical amino acids plus Hpg. If a particular protein of interest is to be overproduced, protein expression can be initiated by the addition of an inducer. (*b–d*) Additional modifications to expression hosts. Engineering auxotrophic *E. coli* strains enables the incorporation of additional ncAAs into proteins. The examples that follow are depicted for the case of overexpressed proteins, but can also be applied in the absence of a particular protein of interest for proteomic applications. (*b*) Overexpression of aminoacyl-tRNA synthetase (aaRS). Noncanonical amino acids that

are poor aaRS substrates do not support protein synthesis unless the *E. coli* strain employed has augmented aaRS activity. In the case shown here, overexpression of leucyl-tRNA synthetase (LeuRS) enables efficient global replacement of hexafluoroleucine (Hfl, **3**) in place of leucine in a protein of interest (POI). (*c*) Overexpression of mutant aaRS. When no wild-type aaRS can activate a particular amino acid analog efficiently, mutant enzymes can be engineered and employed in *E. coli* expression hosts. Overexpression of a mutant methionyl-tRNA synthetase (MetRS) enables the global replacement of Met with azidonorleucine (Anl, **4**). (*d*) Overexpression of editing-deficient mutant aaRS. Some amino acid analogs are kinetically competent aaRS substrates but are subjected to editing after activation or aminoacylation. Altering the editing activity of an aaRS can enable incorporation of amino acid analogs that are proofread by the wild-type aaRS. For example, overexpression of an editing-deficient LeuRS enables norleucine (Nrl, **5**) to replace Leu in proteins.

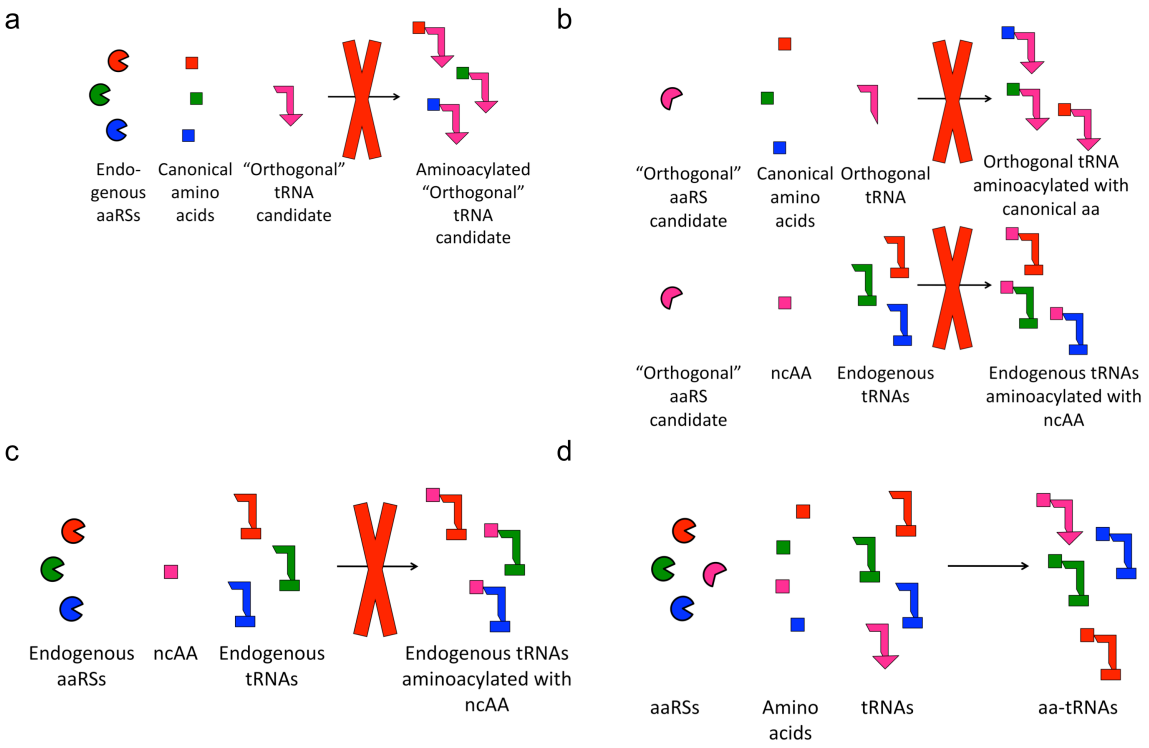


Figure 1.3. Orthogonality requirements for adding tRNAs and aminoacyl-tRNA synthetases (aaRSs) to a translation system. (*a*) Transfer RNA orthogonality. The heterologous tRNA to be introduced should not be a substrate for any of the endogenous aminoacyl-tRNA synthetases in order to avoid the aminoacylation of the tRNA with canonical amino acids. (*b*) AaRS orthogonality. The heterologous aaRS should not aminoacylate the heterologous tRNA with canonical amino acids or aminoacylate endogenous tRNAs with ncAAs. (*c*) Noncanonical amino acid orthogonality. The amino acid to be "added" to the genetic code should not be a substrate for any of the endogenous aaRSs. (*d*) Orthogonal pair. A properly functioning orthogonal aaRS-tRNA pair performs its aminoacylation task efficiently and specifically in the context of the endogenous translational machinery.

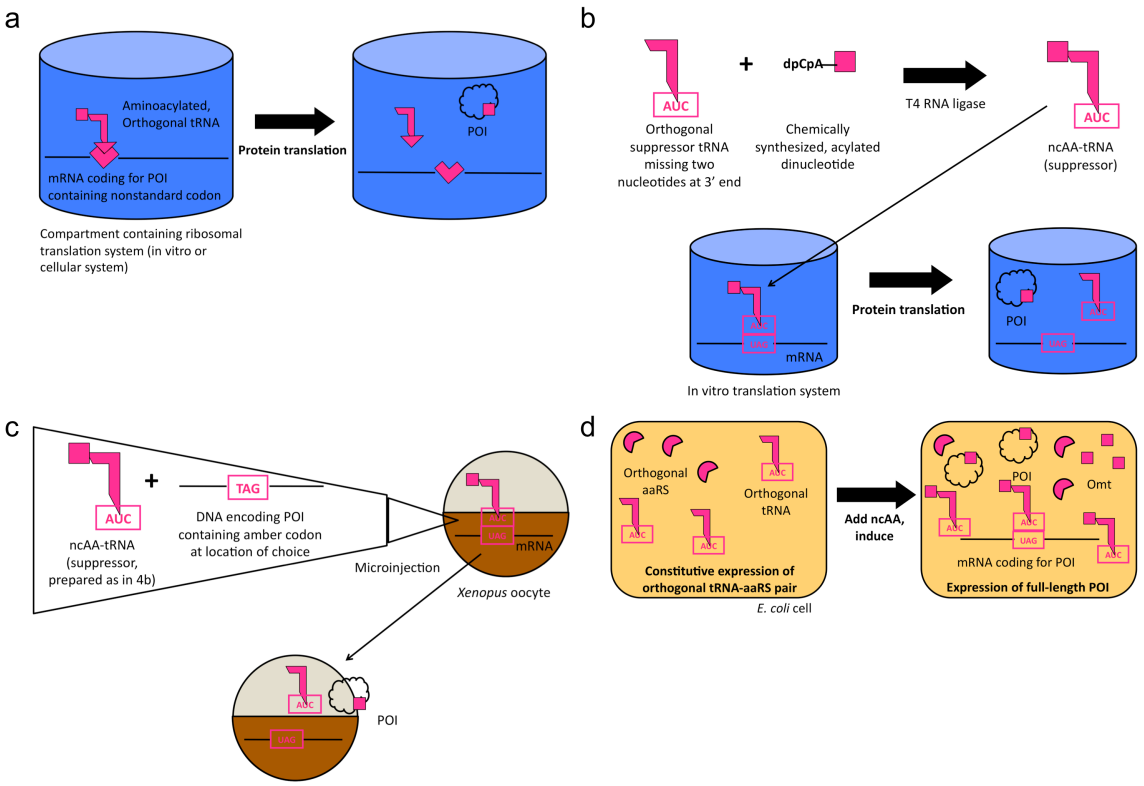


Figure 1.4. Methods for site-specific incorporation of noncanonical amino acids into proteins. (*a*) Suppression of nonstandard codon for site-specific incorporation. An aminoacylated, orthogonal tRNA is introduced into a translation system to decode a nonsense or other nonstandard codon, resulting in a full-length protein containing a noncanonical amino acid at one specified position in the protein. (*b–d*) Systems for employing site-specific incorporation techniques. (*b*) In vitro translation. Chemically synthesized, orthogonal aminoacyl-tRNAs can be employed in conjunction with an in vitro translation system to effect site-specific ncAA incorporation. (*c*) Microinjection into *Xenopus* oocytes. Simultaneous injection of a gene containing a nonstandard codon to be suppressed and an aminoacylated, orthogonal tRNA allows for the synthesis of proteins containing a site-specifically incorporated noncanonical amino acid inside a living cell. Membrane proteins are commonly studied using this technique. (*d*) Use of an orthogonal

pair. *E. coli* cells outfitted with an additional aaRS-tRNA pair can synthesize proteins containing a site-specifically incorporated twentyfirst amino acid. In this case, the aaRS enzymatically aminoacylates the tRNA during the course of the experiment, requiring no chemical synthesis of aminoacyl-tRNAs. Omt, *O*-methyltyrosine (**24**).

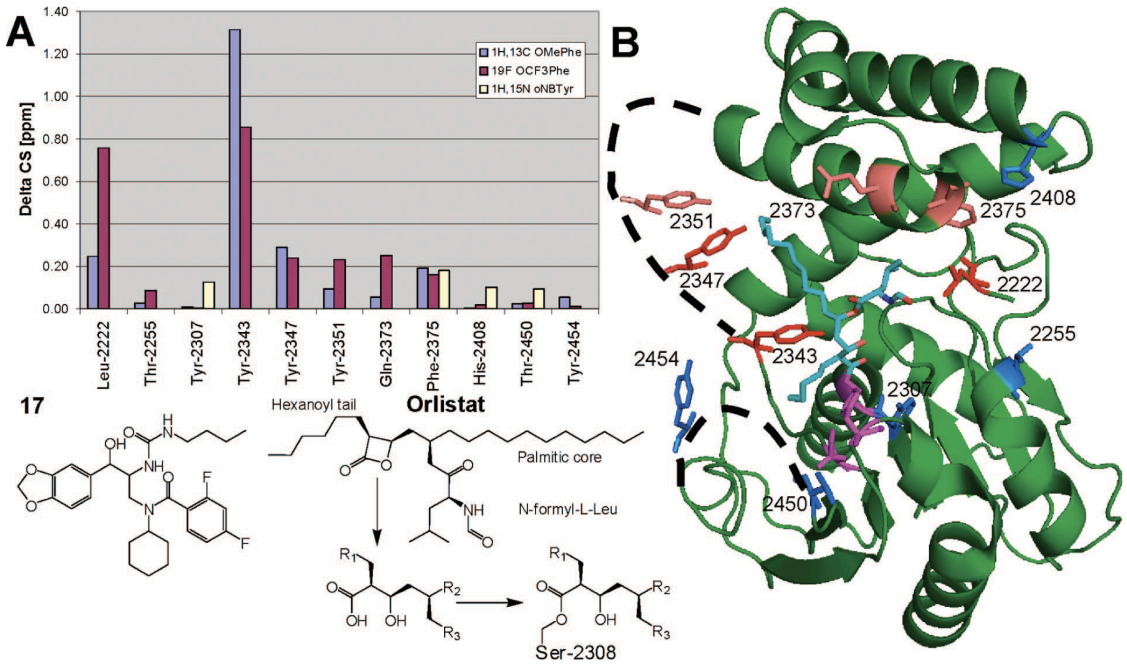


Figure 1.5. NMR studies of ligand binding employing noncanonical amino acids as isotopically labeled probes. (*a*) Combined chemical shift data of tool compound **17** binding to the thioesterase domain of human fatty acid synthase (FAS-TE). Chemicals shift changes ΔCS have been scaled as described by Cellitti et al. in order to enable comparison between ^{19}F, ^{13}C, and ^{15}N NMR experiments (91). Conformational exchange prevented acquisition of chemical shift data on some *o*-nitrobenzyl-tyrosine (oNBTyr, **25**) mutants. OMePhe, *p*-methoxyphenylalanine (**24**). OCF3Phe, *p*-trifluoromethoxylphenylalanine (**23**). (*b*) Structure of FAS-TE covalently modified with orlistat (2PX6.pdb) (352). The average chemical shift changes induced by the binding of **17** are calculated for each ncAA and color-coded for each position: ΔCS < 0.1 ppm, blue; 0.1-0.2 ppm, salmon; > 0.2 ppm, red. Disordered loops are indicated by dashed lines. The active site residues Ser-2308, Asp-2338, and His-2481 are shown in magenta. Adapted from (91) with permission. © 2008 American Chemical Society.

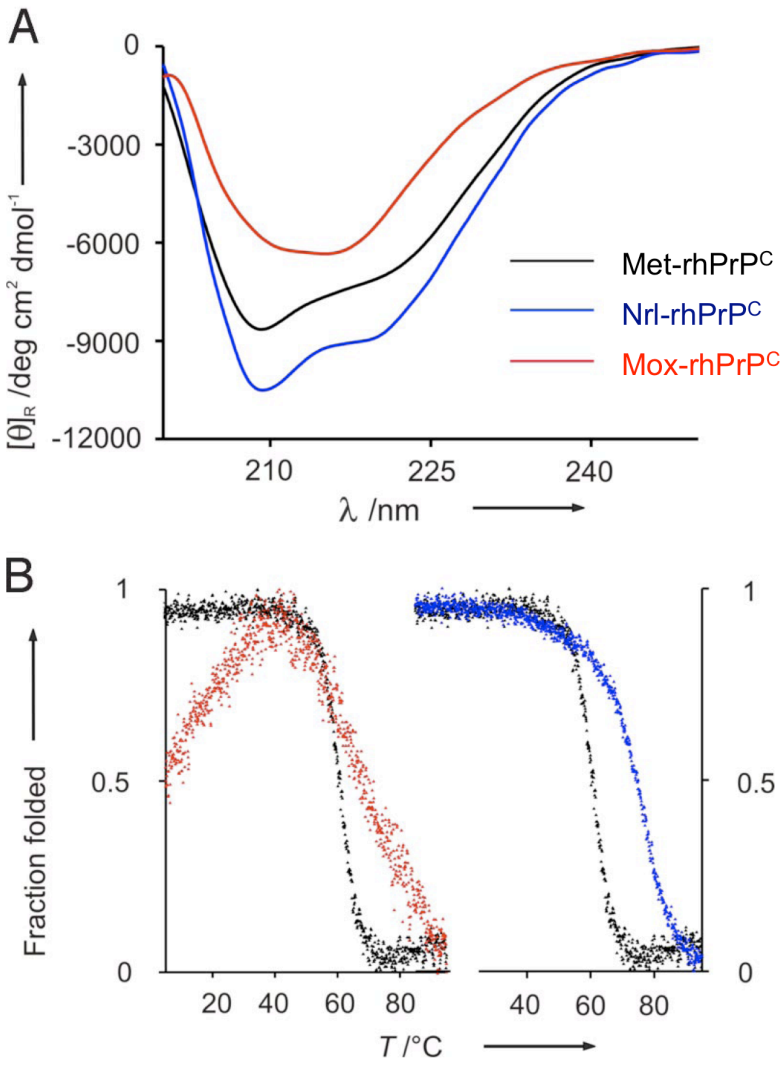


Figure 1.6. Secondary structure analysis and thermal denaturation of recombinant human prion protein (rhPrP^{C}) containing methionine analogs. (*a*) Circular dichroism spectra of Met-rhPrP^{C} and its norleucine (Nrl, **5**) and methoxinine (Mox, **49**) variants at 37 °C and 0.2 mg/mL in 10 mM Mes buffer at pH 6.0. (*b*) Thermal denaturation monitored by the changes of dichroic intensities at 222 nm as a function of temperature. Note that both the secondary structural characteristics and denaturation behaviors are greatly influenced by the hydrophobicities of the amino acids incorporated at the Met positions. Adapted from (194) with permission.

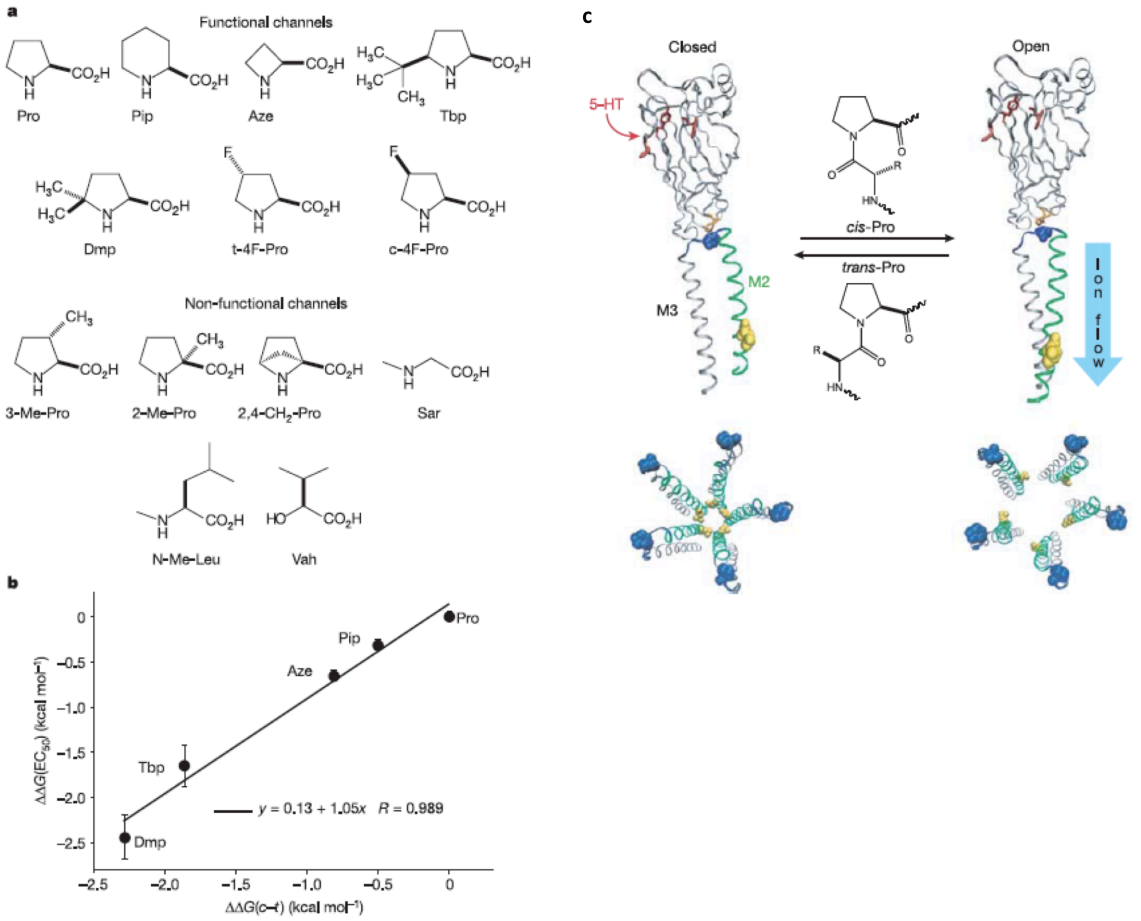


Figure 1.7. Unnatural amino acid mutagenesis on the 5-hydroxytryptamine type 3 receptor leads to a new model for receptor gating. (*a*) NcAAs used in study. These residues were incorporated in place of a key proline residue (Proline 8*) within the receptor. (*b*) The thermodynamics of the *cis–trans* isomerization propensities (ΔΔG(*c*-*t*)) of ncAAs and receptor activation by 5-hydroxytryptophan (ΔΔG(EC_{50})) are strongly correlated, suggesting that the isomerization properties of the residue at the position of interest are critical for forming functional channels. (*c*) Proposed model for receptor gating. Isomerization of the proline residue shown in blue (proline 8*) dictates M2-M3 loop conformation, which in turn controls ion flow through the channel. Adapted from (218) with permission. © 2005 Nature Publishing Group.

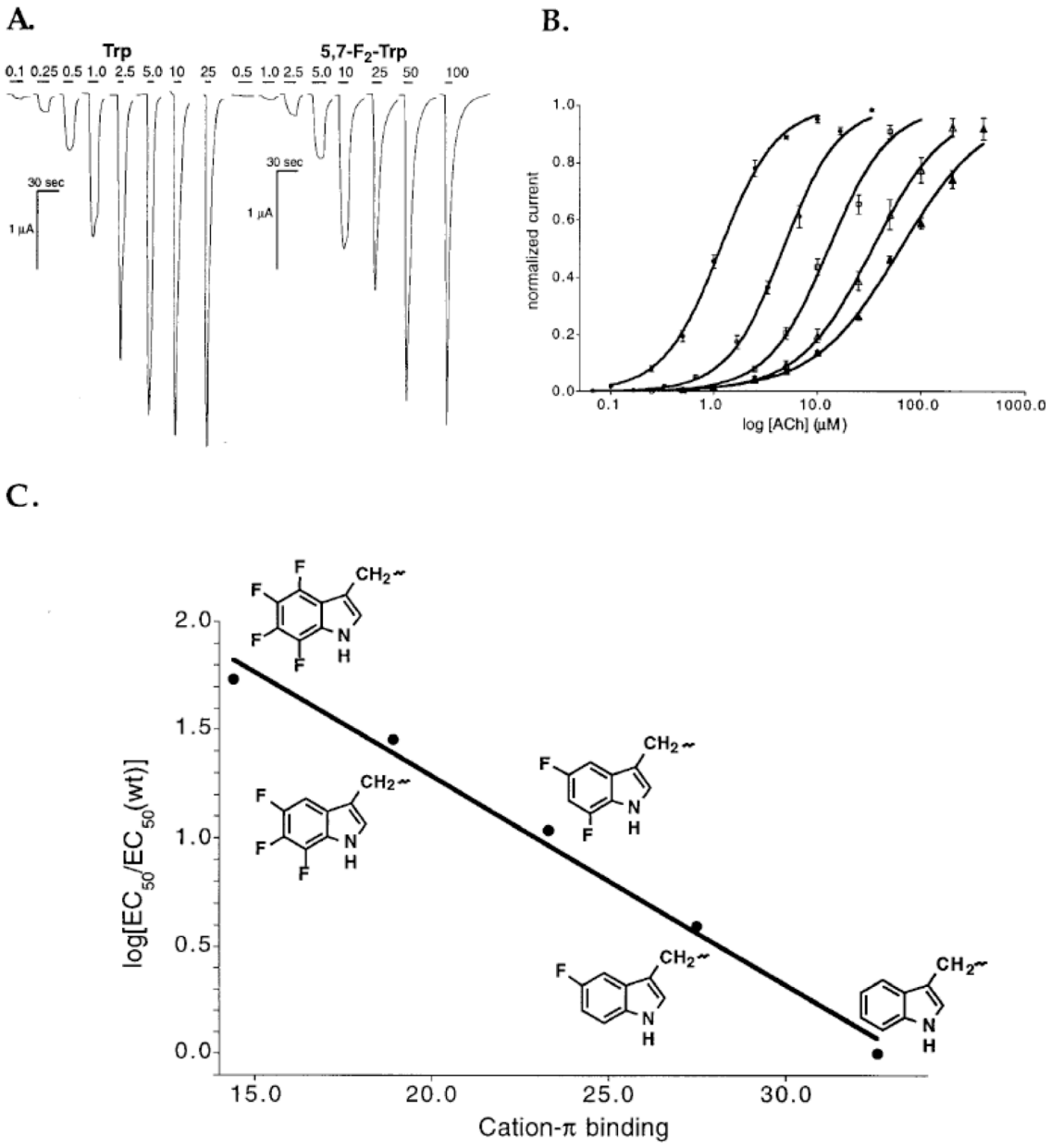


Figure 1.8. Use of noncanonical amino acids to investigate cation-π interactions in the muscle-type nicotinic acetylcholine receptor (nAChR) using patch-clamp experiments. (*a*) Receptor response to increasing doses of acetylcholine as measured by voltage-clamp current traces. The two experiments shown are from oocytes expressing tryptophan (Trp, Left) and 5,7-F_2-Trp (Right, **68**) at α149. Bars represent application of acetylcholine (μM). (*b*) Dose–response relations and fits to the Hill equation for (left to right): Trp; 5-F-Trp

(**14**); 5,7-F_2-Trp (**68**); 5,6,7-F_3-Trp (**69**); and 4,5,6,7-F_4-Trp (**70**). (*c*) Plot of log[EC_{50}/EC_{50}(wild type)] versus quantum mechanically calculated cation–π binding ability at α149 for the same residues as in (*b*). The strong correlation observed confirms that residue α149 binds to acetylcholine through a cation-π interaction. Adapted from (231) with permission.

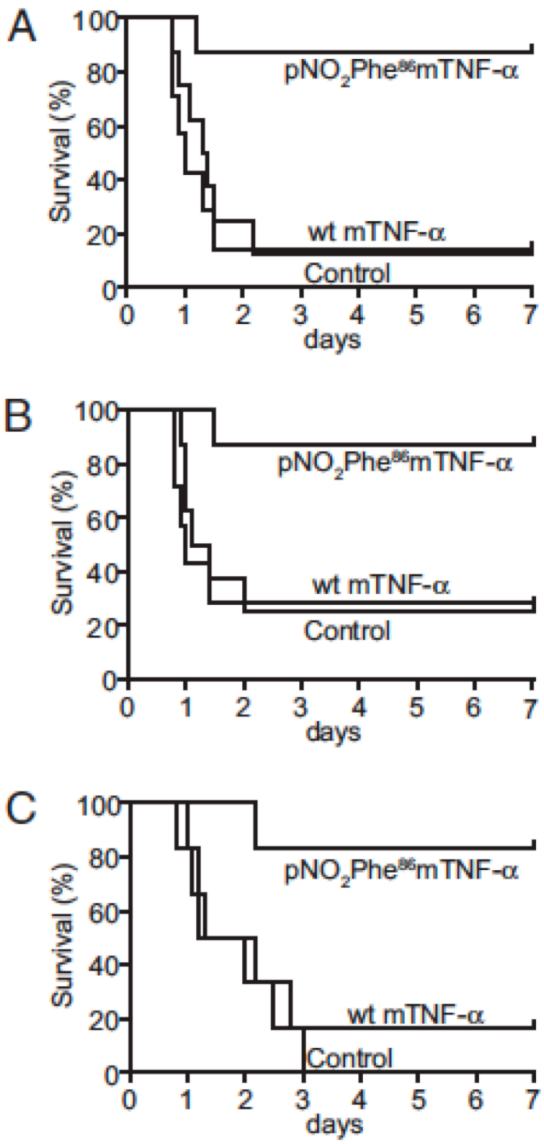


Figure 1.9. Use of ncAAs to breaking immunochemical self-tolerance. Immunization with *p*-nitrophenylalanine (**84**) at position 86 of murine tumor necrosis factor-α (pNO_2Phe^{86}mTNF-α) improves survival of mice in a TNF-α-dependent severe endotoxemia model. Kaplan–Meier survival plots of mice receiving active or passive immunizations are shown. (*a*) Mice (eight per group) immunized with pNO_2Phe86mTNF-α or WT mTNF-α are compared with seven mice receiving sham immunizations. The survival advantage of mice immunized with pNO_2Phe86mTNF-α ($P <$

0.01) versus WT is shown. (*b*–*c*) The survival advantage is preserved when antibodies (*b*) or pooled serum (*c*) from mice immunized with pNO_2Phe86mTNF-α is transferred to other mice. Mice (eight per group) injected with 100 μg of purified IgG from pNO_2Phe86mTNF-α or WT immunized mice were compared with controls receiving saline injection. Survival advantage of mice immunized with pNO_2Phe86mTNF-α ($P <$ 0.01) versus WT is shown. (*c*) Mice (six per group) received 100 μL of pooled serum from mice immunized with pNO_2Phe86mTNF-α or WT mTNF-α. Survival advantage of mice immunized with pNO_2Phe86mTNF-α ($P < 0.01$) versus WT is shown. Adapted from (283) with permission.

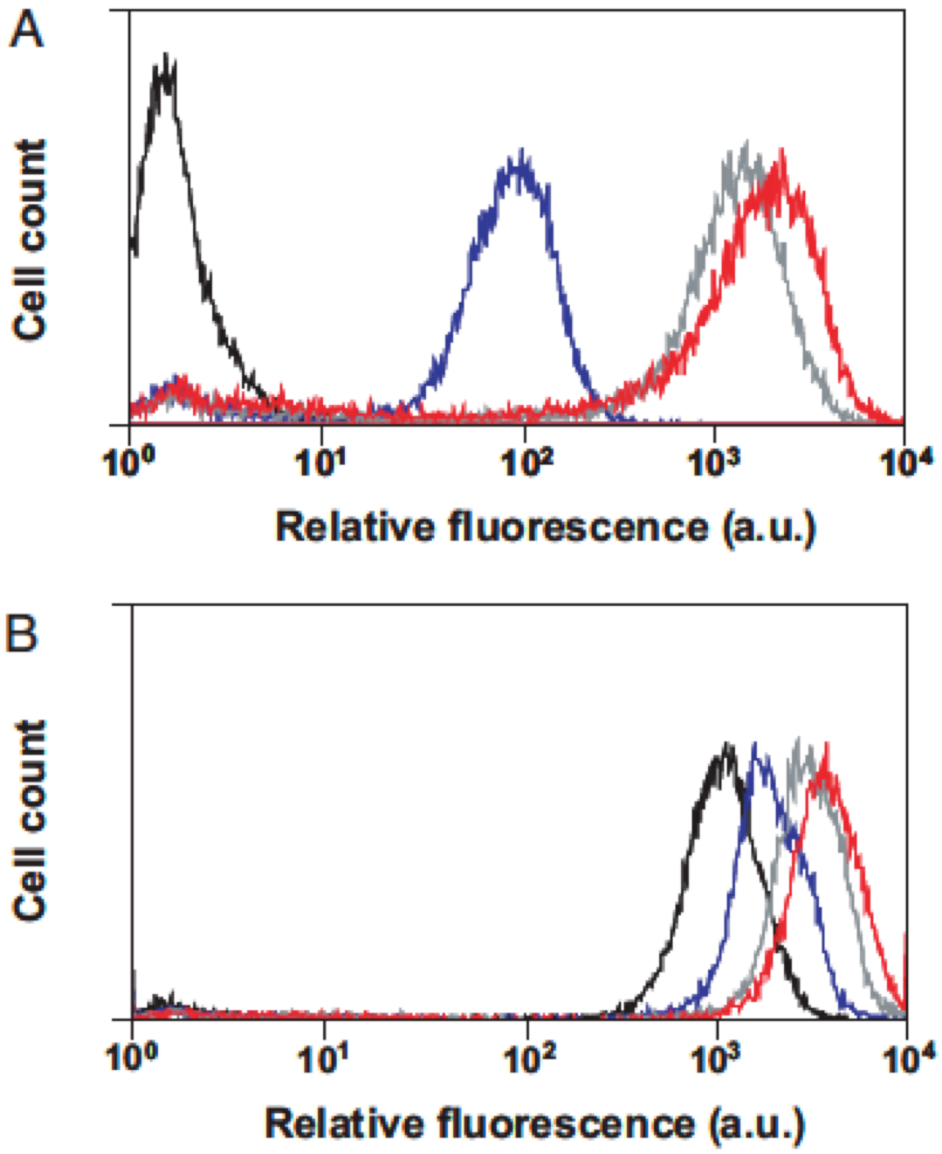


Figure 1.10. Flow cytometric analysis of cells expressing green fluorescent protein (GFP) and GFP variants during the course of evolving fluorinated GFPs. Proteins were expressed in media depleted of Leu and supplemented with trifluoroleucine (Tfl, **44**) (*a*) or in media containing all twenty canonical amino acids (*b*). Black line, GFPm (parent); blue line, 4.2.2 (mutant isolated after sequential construction and enrichment of four libraries); gray line, 8.3.3 (mutant isolated after enrichment of eight libraries); and red line, 11.3.3 (mutant isolated after enrichment of eleven libraries). The evolved, fluorinated variants have

regained fluorescence, likely through improved folding properties. Adapted from (189) with permission.

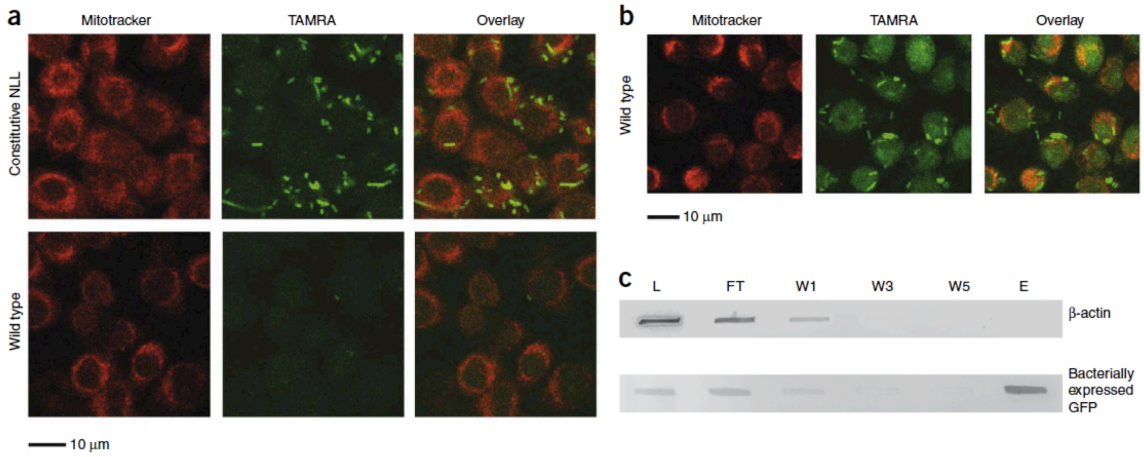


Figure 1.11. Cell-selective labeling in mixtures of bacterial and mammalian cells. (*a*) Fluorescence images of mixed cultures containing bacteria attached to or internalized by mouse alveolar macrophages. Infection was performed in medium containing azidonorleucine (**4**). Bacterial cells constitutively expressing the methionyl tRNA synthetase variant NLL-MetRS were labeled by TAMRA-alkyne (constitutive NLL), whereas cells lacking the NLL-MetRS (wild type) are visible only in the GFP channel (not shown). Macrophages were labeled with Mitotracker Deep Red (Invitrogen) and exhibited very low TAMRA background emission. In all cases, conjugation of TAMRA-alkyne was confined to bacterial cells expressing the NLL-MetRS. (*b*) Fluorescence images of macrophage infection with wild-type bacteria performed in the presence of azidohomoalanine. Protein synthesis by macrophages is indicated by strong TAMRA-alkyne emission from both bacterial cells and macrophages. (*c*) Macrophages were infected with bacterial cells that express GFP under induction with IPTG and that constitutively express the NLL-MetRS. Infection was performed in medium containing IPTG and azidonorleucine to facilitate bacterial synthesis and labeling of GFP. Total cell lysate from the infection was subjected to conjugation with alkyne-functionalized biotin; labeled proteins were enriched with streptavidin avidity. Bacterially expressed GFP and

mammalian β-actin were followed by immunoblots. Analyses of the lysate (L), unbound flow-through (FT), washes (W1, W3, W5) and eluent (E) reveal a separation of bacterial and mammalian representative proteins. Bacterially expressed GFP was labeled with azidonorleucine and thus subject to conjugation to biotin and enrichment with streptavidin. Proteins originating from macrophages, including β-actin, were not labeled with azidonorleucine and therefore were not conjugated to alkyne-functionalized biotin. Adapted from (318) with permission.