Anal Bioanal Chem (2012) 403:2089–2102
DOI 10.1007/s00216-012-5784-2

REVIEW

Biochemical analysis with the expanded genetic lexicon

Hui-wang Ai

Received: 28 November 2011 / Revised: 17 January 2012 / Accepted: 23 January 2012 / Published online: 11 February 2012

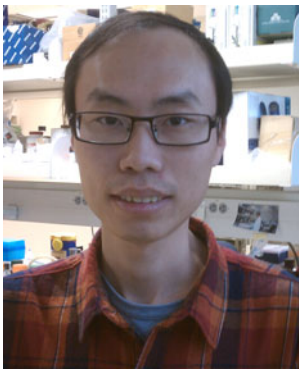
Abstract The information used to build proteins is stored in the genetic material of every organism. In nature, ribosomes use 20 native amino acids to synthesize proteins in most circumstances. However, laboratory efforts to expand the genetic repertoire of living cells and organisms have successfully encoded more than 80 nonnative amino acids in *E. coli*, yeast, and other eukaryotic systems. The selectivity, fidelity, and site-specificity provided by the technology have enabled unprecedented flexibility in manipulating protein sequences and functions in cells. Various biophysical probes can be chemically conjugated or directly incorporated at specific residues in proteins, and corresponding analytical techniques can then be used to answer diverse biological questions. This review summarizes the methodology of genetic code expansion and its recent progress, and discusses the applications of commonly used analytical methods.

Keywords Genetic code · Nonnative amino acids · Bioorthogonal chemistry · Fluorescent probes · Biophysical probes · Photocrosslinking amino acids

Background

Proteins made in different lengths and arrangements from 20 amino acids carry out diverse catalytic, mechanical or structural functions in almost every step of cellular processes. In living cells and organisms, genetic information determining the organization of amino acid residues is stored in the chains of nucleotides, in which every three bases form a codon [1]. There are 61 three-base codons encoding 20 native amino acids and three remaining codons for translation termination [2]. The correspondence between codons and amino acids or translation termination is ensured by cell machineries and pathways, and such connection is nearly universal among all domains of life [3]. In 1976, Cone et al. identified a selenocysteine-containing bacterial protein formate dehydrogenase [4], and in 1986, Chambers et al. and Zinoni et al. independently reported that selenocysteine was encoded by the UGA codon, a commonly used termination codon [5, 6]. In 2002, Srinivasan et al. and Hao et al. reported the 22nd genetically encoded amino acid pyrrolysine, which is used by methanogens in response to the amber UAG codon [7, 8]. The exceptional selenocysteine and pyrrolysine are examples of the natural expansion of genetic codes of organisms for access to ribosomally synthesized unusual proteins. Readers interested in these topics can find useful information in related review papers [9, 10].

Researchers have devoted tremendous efforts to the development of methods that can manipulate protein structures



Hui-wang Ai has been an Assistant Professor of Chemistry at the University of California, Riverside since July 2011. His current research interests are developing biosensors for chemicals of environmental and biological interests, and engineering molecular tools for the interrogation of biochemical networks using light microscopy.

Published in the special issue *Young Investigators in Analytical and Bioanalytical Science* with guest editors S. Daunert, J. Bettmer, T. Hasegawa, Q. Wang and Y. Wei.

H.-w. Ai (✉)
Department of Chemistry, University of California,
510 Big Springs Road,
Riverside, CA 92521, USA
e-mail: huiwang.ai@ucr.edu

and functions in vitro and in living cells [11, 12]. The major driving forces are to generate research tools for understanding of biology and protein reagents with new or improved properties. Proteins can be derivatized through the native cysteine or lysine side chains that are reactive to maleimides or *N*-hydroxysuccinimide (NHS) esters [13]. In addition, chemical and enzymatic transformations have been developed to convert native amino acid residues or peptide termini to reactive handles through which conjugation can be achieved [14, 15]. Solid-phase chemistry has also been used to synthesis peptides containing nonnative building blocks [16]. Although it is not practical to directly synthesize large proteins, a handful of peptide ligation reactions can be used to join multiple chemically synthesized peptides, or chemically synthesized peptides with peptides generated by intein-based technology [17, 18]. A detailed discussion of the abovementioned technology is beyond the scope of this review. Instead, readers can refer to appropriate review articles [13, 16, 17].

Nonnative amino acids can be introduced into proteins by hijacking cell-translation machinery. For example, if a nonnative amino acid that is structurally close to a canonical amino acid is used to culture an auxotrophic bacterial strain, the nonnative amino acid may bypass fidelity checks in the aminoacylation and translation steps and be globally incorporated into proteins in cells [19, 20]. In addition, transfer RNAs (tRNAs) "mischarged" with nonnative amino acids can be prepared in vitro under special reaction conditions, or by ligating truncated tRNAs with chemically aminoacylated nucleotides [21, 22]. In both cases, the resulting aminoacylated tRNAs can be used by ribosomes to synthesize nonnative amino acid-containing proteins. This can be done in cell-free translation systems or in living cells into which nonnative aminoacylated tRNAs are delivered [23, 24]. If the corresponding tRNAs have been engineered to harbor unusual anticodons (e.g. nonsense termination codons or extended four-base codons), these nonnative amino acids can then be incorporated into proteins at particular residue sites whose genetic sequences have been mutated to the corresponding unusual codons [25, 26]. It is worth noting that nonsense suppressors exist naturally in many bacteria, yeast, and other eukaryotic cells, although they typically use native amino acids [27].

All above methods are valuable, but each has profound limitations. The extracellular nature of these methods has restricted the synthesis of proteins difficult to fold and the production of proteins on a large scale. In particular, it is often tedious and sometimes even impossible to use these modified proteins for interrogation of intracellular biology. Additional methods have been developed to genetically encode all major components needed for the incorporation of nonnative amino acids [28]. Orthogonal tRNA/synthetase pairs can be engineered and recombinantly expressed in living cells and organisms, which use unusual codons to integrate nonnative amino acids into proteins. Nonnative amino acids supplied in the culture media for bacterial, yeast, insect, or mammalian cells are then incorporated into cellular proteins by cell machineries. To date, the technology is available for genetic encoding of more than 80 nonnative amino acids harboring various reactive conjugation handles, photo-reactive side chains, pre-installed post-translational modifications (PTMs), fluorescent side chains, metal-chelating functional groups, and other useful side chains [29]. It has enabled the direct and precise manipulation of protein structures and functions in living cells and organisms with unprecedented flexibility. Here, we briefly review the general methodology of expanding the genetic code and its recent advances, and focus on its application to analytical and bioanalytical chemistry. Many important aspects (e.g. those related to cell biology, therapy and vaccine development, protein design and evolution with additional codons) are not covered in this review. Interested readers may refer to other recent reviews [30, 31].

General methodology

To integrate a nonnative amino acid into the genetic code of an organism, one must first identify an aminoacyl-tRNA synthetase that charges the corresponding tRNA with that nonnative amino acid. Importantly, the synthetase should not recognize cell-endogenous tRNAs and amino acids, and the tRNA must not crosstalk with any endogenous synthetases. Such a tRNA/synthetase pair is referred as an orthogonal pair. The orthogonal tRNA must be engineered to respond to an unusual codon (e.g. a nonsense codon or a four-base codon). It is also important that the nonnative amino acid is metabolically stable and can be introduced into cells in adequate quantity. The nonnative amino acid cannot be a substrate of endogenous synthetases, or it will globally replace residues of cellular proteins. Furthermore, the resulting aminoacylated tRNA must be compatible with protein translation machineries, for example elongation factors and the ribosome.

Although it seems a daunting task, researchers have developed procedures integrating positive and negative selection cycles to generate orthogonal tRNA/synthetase pairs for encoding nonnative amino acids [31]. When an orthogonal tRNA/synthetase pair is expressed in cells in the presence of the nonnative amino acid, the nonnative amino acid is specifically acylated to the engineered orthogonal tRNA. The genetic codon of the target protein residue has to be mutated to an unusual codon, so the ribosome will use the aminoacylated nonnative tRNA when reaching the unusual codon in the messenger RNA (mRNA). The nonnative amino acid is then site-specifically introduced into the target protein (Fig. 1).

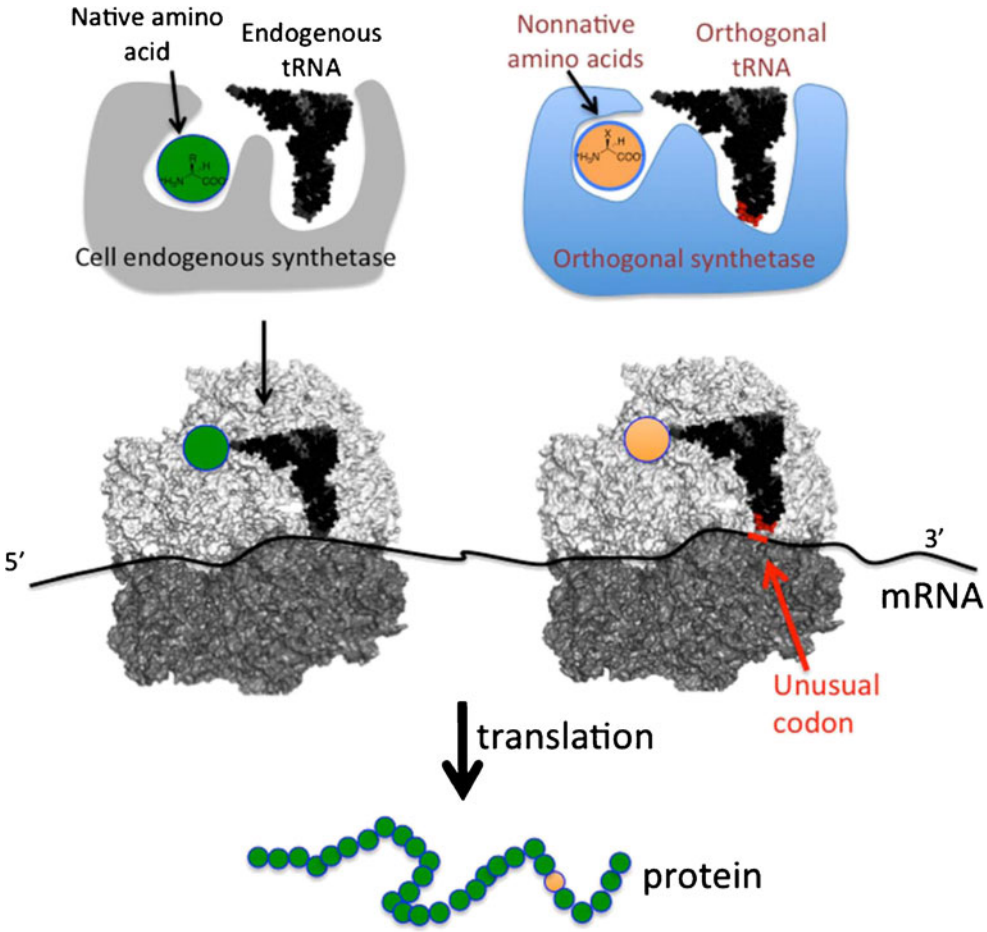


Fig. 1 Schematic diagrams of genetic encoding of nonnative amino acids in living cells

Expansion of the genetic code of *E. coli* to encode *O*-methyltyrosine (Fig. 2 (1)) was reported in 2001 by Wang et al. [32]. The tyrosyl tRNA and tyrosyl-tRNA synthetase pair from the archaebacterium *Methanocaldococcus jannaschii* (*Mj*TyrRS/*Mj*tRNA^{Tyr}) was engineered through multiple rounds of positive and negative selection, leading to improved orthogonality in *E. coli* and the switching of substrate specificity from tyrosine to the nonnative *O*-methyltyrosine. To date, the *Mj*TyrRS/*Mj*tRNA^{Tyr} pair has been extended for the encoding of more than 35 nonnative amino acids in *E. coli* [30]. Another important group of orthogonal tRNA/synthetase pairs are adapted from methanogenic archaebacteria in which the amino acid pyrrolysine is naturally encoded [7]. The pyrrolysyl tRNAs and pyrrolysyl-tRNA synthetases (PylRS/tRNA^{Pyl}) from *Methanosarcina barkeri*, *Methanosarcina mazei*, or *Desulfitobacterium hafniense*, and their engineered mutants have been used to genetically encode at least 25 nonnative amino acids [30, 33]. Furthermore, additional pairs from organisms including *Saccharomyces cerevisiae* and *Pyrococcus horikoshii* have been occasionally used to encode other nonnative amino acids in *E. coli* [34, 35].

In a given experiment, suppression of nonsense or four-base codons competes with cell endogenous pathways (e.g. translation termination or the decoding of the first three bases of a four-base codon). Therefore, the yield of full-length proteins containing nonnative amino acids is reduced because of the use of unusual codons [36]. This problem is amplified when multiple unusual codons are used in a single gene [35]. Recent progress, however, has refined the insertion of multiple nonnative amino acids. In one example, Liu and his coworkers overexpressed the *E. coli* ribosomal protein L11 to reduce RF1 (release factor 1) mediated translation termination, resulting in the production of a protein with three nonnative residues [37]. Wang et al. reported the generation of an autonomous RF-1 deletion *E. coli* strain which, in theory, can decode the UAG codon as efficiently as decoding other natural codons [38]. In addition, *Mj*TyrRS/*Mj*tRNA^{Tyr}-derived tRNA/synthetase pairs and PylRS/tRNA^{Pyl}-derived pairs were used together in the same *E. coli* cell to decode two nonsense codons, so two different nonnative amino acids were inserted into a single protein [39]. Effort has also gone into the development of orthogonal ribosomes that recognize special ribosomal binding sites not existing in endogenous mRNAs [40]. The initial research by *Chin* et al. has enhanced the ability of the orthogonal ribosomes in decoding the nonsense UAG codon and four-base codons [41]. Research in this direction has the potential to enable the biosynthesis of polymers from completely artificial building blocks.

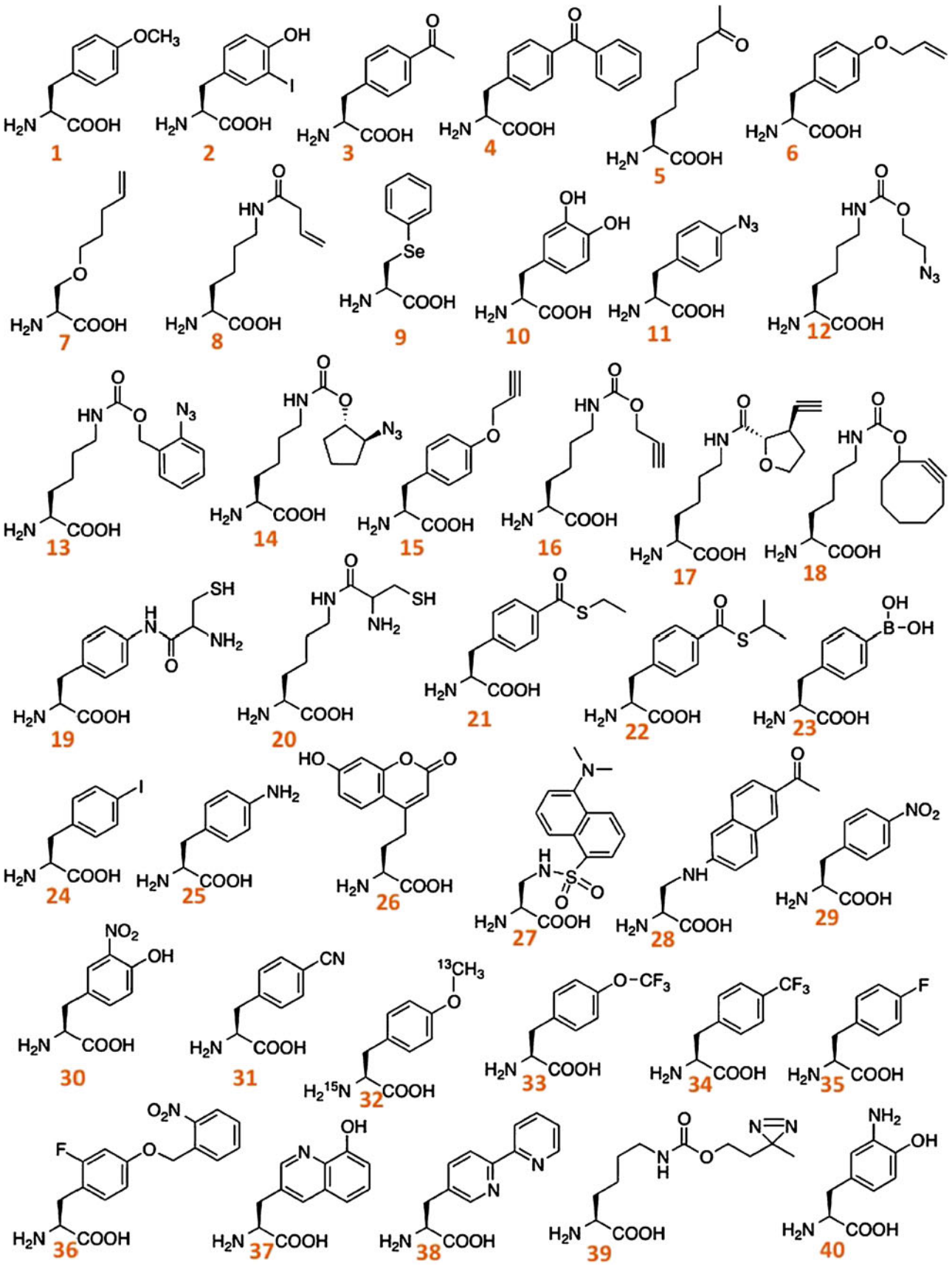


Fig. 2 Chemical structures of nonnative amino acids discussed in the text

Similarly, orthogonal tRNA/synthetase pairs have been developed to expand the genetic code of eukaryotic cells. Yokoyama and his coworkers reported the genetic incorporation of 3-iodotyrosine (Fig. 2 (2)) in response to the UAG codon in mammalian cells by rationally designing a mutant pair based on *E. coli* tyrosyl-tRNA synthetase and *Bacillus stearothermophilus* tyrosyl tRNA (*Ec*TyrRS/*Bst*tRNA^{Tyr}) [42]. To generalize the approach, Schultz and coworkers innovated a positive and negative selection scheme in *S. cerevisiae* (analogous to their selection in *E. coli*) [43]. The group explored both the *Ec*TyrRS/*Ec*tRNA^{Tyr} and *Ec*LeuRS/*Ec*tRNA^{Leu} (*E. coli* leucyl tRNA/synthetases) pairs, and their research has resulted in the addition of ~20 nonnative amino acids to the genetic repertoire of *S. cerevisiae* [30]. Most pairs evolved in *S. cerevisiae* can be directly adapted for expansion of the genetic code of mammalian cells by selecting appropriate mammalian expression promoters [44].

Orthogonal tRNA/synthetase pairs used in a given organism are usually from a different domain of life. For example, *Mj*TyrRS/*Mj*tRNA^{Tyr}-derived pairs are orthogonal in bacteria, and *Ec*TyrRS/*Ec*tRNA^{Tyr} and *Ec*LeuRS/*Ec*tRNA^{Leu}-derived pairs are used in eukaryotic cells. PylRS/tRNA^{Pyl}-derived pairs are unique in this sense: because of their distinctive structural features, they can be used in both prokaryotic and eukaryotic cells for genetic encoding of nonnative amino acids. PylRS/tRNA^{Pyl}-derived pairs engineered in *E. coli* have been directly transported into *S. cerevisiae* and cultured mammalian cells [45–47].

Genetic code expansion technology has been used in a variety of cell types and under different experimental conditions. For example, *E. coli* cells that use 21 amino acids have been used to express bacteriophages with surface displayed peptides and proteins containing nonnative amino acids [48, 49]. This has been used to evolve nonnative amino acid-integrated protein binders, protein folds, and enzyme inhibitors. In addition, nonnative amino acids have been introduced into the bacterial pathogen *Mycobacterium tuberculosis* (MTB) for mechanism studies and vaccine development [50]. Furthermore, these pairs initially developed in *S. cerevisiae* have been tested in the yeast *Pichia pastoris* [51], in primary neurons [52], and in cells of the insect *Drosophila melanogas*ter [53]. Recently, Chin et al. reported the genetic code expansion of the multicellular organism *Caenorhabditis elegans* [54]. Research to extend this into other multicellular organisms (e.g. *Drosophila melanogas*ter and *Mus musculus*) is in progress.

Bioorthogonal labeling of proteins

The native 20 amino acids only have limited biophysical properties that can be tracked by analytical techniques. To more efficiently probe protein structures and functions, analytical and biophysical probes may be used to derivatize proteins through bioorthogonal reactions. Reactions with a native amino acid do not have much selectivity, because such native residues may exist in many proteins. Instead, genetically encoded nonnative amino acids may provide characteristic functional groups not existing in native proteins, and these unique functional groups may be used as reactive handles to conjugate analytical and biophysical probes, assuming bioorthogonal chemical reactions are available for linking probes to the nonnative amino acid side chains.

The keto functional group is not present in the 20 native amino acids. It selectively and efficiently reacts with hydrazides and alkoxyamines. Keto-containing nonnative amino acids, for example *p*-acetylphenylalanine (Fig. 2 (3)), *p*-benzoylphenylalanine (Fig. 2 (4)), and 2-amino-8-oxononanoic acid (Fig. 2 (5)) can be introduced into proteins in *E. coli*, yeast, and mammalian cells [43, 44, 55–58]. Probes containing hydrazides or alkoxyamine functional groups can then be used to label proteins through the keto side chain (Fig. 3a). In one report, T4 lysozyme was mutated to insert a site-specific *p*-acetylphenylalanine residue that was conjugated with the fluorescent Alexa 488 alkoxyamine [59]. Another fluorescent dye Alexa 594, the Förster resonance energy transfer (FRET) acceptor of Alexa 488, was also used to label the cysteine residue of T4 lysozyme by use of maleimide–thiol coupling chemistry. Unfolding of the dual labeled protein was induced, and single-molecule FRET between the two fluorescent dyes was monitored. Two populations of unfolded peptides were observed, supporting a two-state unfolding mechanism of T4 lysozyme. In another example, the reaction was used to site-specifically conjugate a protein with a nitroxide spin label, so that the dynamics of the protein can be followed by electron paramagnetic resonance (EPR) spectroscopy [60]. Despite the success, the reaction between ketones and hydrazides or alkoxyamines is relatively slow (on a timescale of hours to overnight) and needs a weakly acidic environment (pH 4–6.5). In addition, the keto group is abundant in other types of intracellular molecule [61]. These factors have limited the use of this coupling in living cells.

The alkene functionality is present in multiple nonnative amino acids encoded in *E. coli*, yeast, or mammalian cells (Fig. 2 (6–8)) [33, 62–64]. Alkenes can undergo olefin metathesis reactions with another alkene molecule [64, 65], or Mizoroki–Heck reactions with unsaturated halides [66, 67]. Both need appropriate organometallic catalysts, however. Recent research has reported water-soluble catalysts, but their applications to diverse proteins and in living cells remain to be established [60–63]. In addition, Lin and coworkers reported a UV light-induced coupling reaction between olefins and tetrazoles (Fig. 3b) [68]. No catalyst is needed, and it produces fluorescent pyrazoline products.

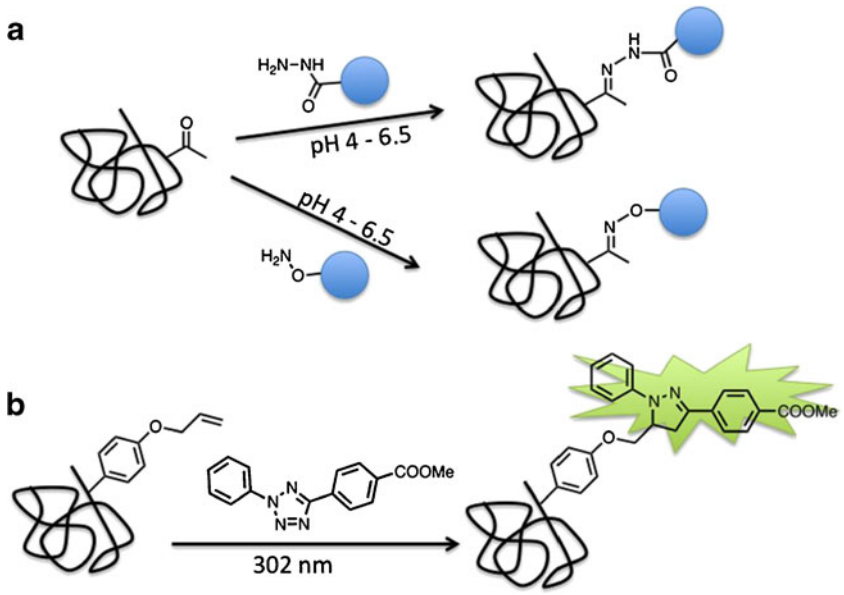


Fig. 3 Protein labeling based on (**a**) the conjugation between the keto group and hydrazides or alkoxyamines, (**b**) the light-induced conjugation between the alkene functional group and a tetrazole

The reaction has been tested for protein labeling in living *E. coli* and mammalian cells, and has much potential for other applications in biological systems [69]. Recently, the tetrazole functional group has also been genetically encoded in *E. coli* [70].

Phenylselenocysteine (Fig. 2 (9)) can undergo oxidative elimination in the presence of H_2O_2 to produce dehydroalanine, reactive to nucleophiles (e.g. thiols) in a Michael addition reaction [71]. Obviously, this method is not applicable in living cells, and cannot be used in vitro to label cysteine and methionine-containing proteins. Similarly, 3,4-dihydroxy-L-phenylalanine (L-DOPA, Fig. 2 (10)) can be oxidized by $NaIO_4$ to form a nucleophile-reactive orthoquinone intermediate [72].

Azide or alkyne functional groups can be inserted into proteins in *E. coli*, yeast, and mammalian cells (Fig. 2 (11–17)) [33, 43, 44, 73–78]. The azide–alkyne Huisgen cycloaddition ("click" chemistry), in particular, has been widely used to label proteins in vitro and in living cells (Fig. 4a) [79]. Cu(I) catalysts are typically used, but not required: cyclic ring-strain alkynes can promote the cycloaddition reaction [80]. Recently, a ring-strain alkyne (Fig. 2 (18)) was genetically encoded, and the labeling was accomplished without use of the Cu(I) catalysts that may be detrimental to living cells (Fig. 4b) [81]. In addition, azides can undergo Staudinger ligation with triarylphosphines under very mild conditions (Fig. 4c) [82, 83].

In addition, 1,2-aminothiols (Fig. 2 (19, 20)) or thioesters (Fig. 2 (21, 22)) can be incorporated into proteins [30, 84]. 1,2-Aminothiols readily react with thioesters at physiological pH (Fig. 5a) [85]. 1,2-Aminothiols also react with cyanobenzothiazole derivatives, which has been reported as a very fast bioorthogonal labeling reaction (Fig. 5b) [86].

Nonnative amino acids containing boronic acid, aryl halide or aniline functional groups (Fig. 2 (23–25)) can also be used for bioorthogonal conjugation reactions [43, 44, 87–89]. Aryl halides can undergo multiple organometal-catalyzed reactions [90, 91]. In particular, aryl halides and boronic acids are reactants of the Suzuki reaction [91]. Anilines and aminophenols can be oxidatively coupled at pH 6.5 in 2–5 min (Fig. 5c) [92].

A large set of genetically encoded reactive handles and bioorthogonal chemical reactions are available for site-specific protein labeling in vitro and in living cells [12, 30]. Proteins can also be immobilized or pulled down from mixtures through these incorporated handles. For example, *N*-methylglucamine resins have been used to purify boronic acid-containing proteins, because of the reversible covalent interaction between boronic acids and polyhydroxy functional groups [87]. Proteins with azide groups can, of course, be immobilized on to alkyne-derivatized beads or surfaces [93].

Although analytical and biophysical probes can be conjugated to proteins through reactive chemical handles, in many cases one would prefer the direct encoding of these probes.

Fluorescent probes

The native amino acid tryptophan is fluorescent under 280-nm UV light. Expanding the genetic code of living cells has led to the site-specific incorporation of commonly used fluorophores. A hydroxycoumarin-derived amino acid (Fig. 2 (26)) has been encoded in *E. coli* [94], and both the dansyl and the prodan fluorophores (Fig. 2 (27, 28)) have been encoded in yeast and mammalian cells [95, 96].

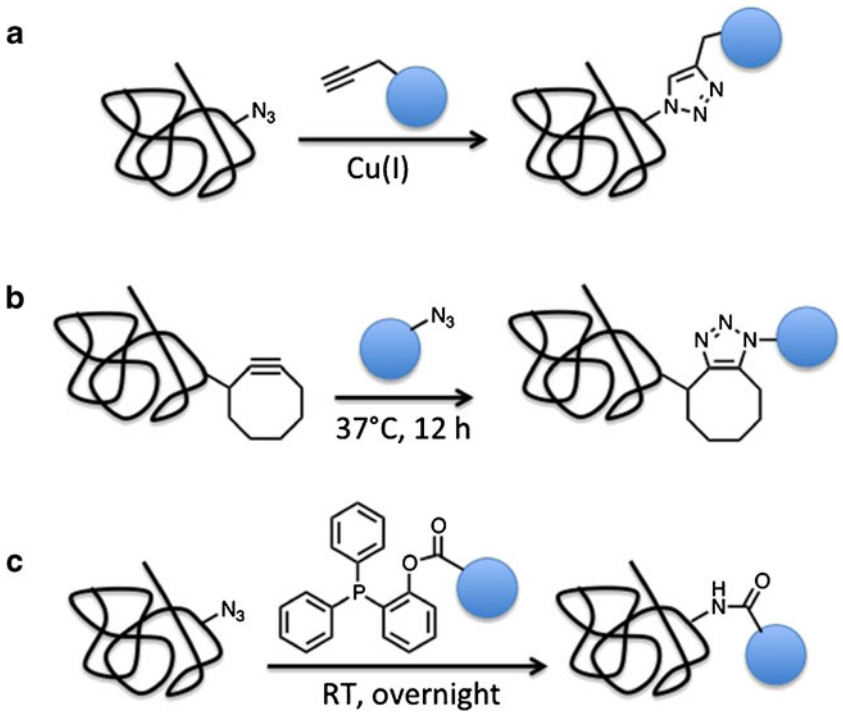


Fig. 4 Protein labeling based on (**a**) the "click" chemistry between the azide functional group and alkynes in the presence of Cu(I) catalysts, (**b**) the copper-free "click" chemistry between a genetically encoded ring-strain alkyne and azides, (**c**) the Staudinger ligation between the azide functional group and aryl phosphines

These genetically encoded fluorescent probes provide the ability to interrogate protein localization and dynamics with little interference.

The simplest application is to label and locate target proteins. In one study, the hydroxycoumarin was inserted into the bacterial tubulin homologue FtsZ protein, and used to track FtsZ location during cell division. It was shown that FtsZ formed a Z-ring at the cleavage furrow [97]. The hydroxycoumarin amino acid has also been used to label newly synthesized peptides in complexes with ribosomes [98]. Similarly, the prodan fluorophore has been inserted into the histone protein in CHO cells, and used to track cell division [30].

A unique application of these fluorescent amino acids is to examine local structural changes in proteins, because their fluorescence is often sensitive to subtle environmental changes. For example, when the prodan fluorophore was incorporated into the binding cleft of the *E. coli* glutamine-binding protein, a dramatic fluorescence emission shift from 480 nm to 430 nm was observed for the protein upon

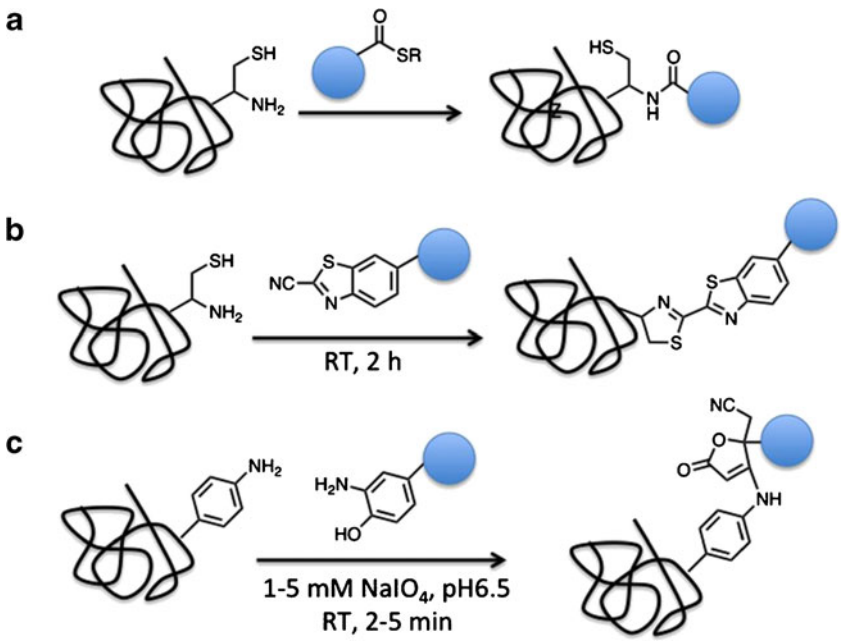


Fig. 5 Protein labeling based on (**a**) the native chemical ligation between 1,2-aminothiols and thioesters, (**b**) the cyanobenzothiazole condensation of 1,2-aminothiols, (**c**) the oxidative coupling of anilines and aminophenols

glutamine binding, a process that induces protein conformation changes [96]. Recently, the hydroxycoumarin amino acid was incorporated into the phosphotyrosine-binding pocket of the protein STAT3. Phosphorylation of STAT3 by upstream enzymes can be monitored, because when the phosphorylated peptide binds to the binding pocket, it affects the fluorescence of hydroxycoumarin in the pocket [99]. In addition, the dansyl fluorophore has been incorporated into the voltage-sensitive domain of *Ciona intestinalis* voltage sensitive phosphatase to monitor membrane depolarization of differentiated neurons from HCN-A94 neural stem cells [52].

Furthermore, genetically encoded *p*-nitrophenylalanine (Fig. 2 (29)) and 3-nitrotyrosine (Fig. 2 (30)) in *E. coli* can be used as fluorescent quenchers for the natural amino acid tryptophan [100, 101]. The quenching process is highly distance-dependent, so it can be used to probe distances related to protein folding and other structural dynamics [102].

Vibrational probes

The infrared (IR) absorption of a protein provides a wealth of information about its structure and dynamics [103]. The bond stretching, bending, and other vibration modes in a protein can be interpreted from its absorption of IR light. Nonnative amino acids with strong characteristic vibration modes have been used to site-specifically label proteins; IR spectroscopy can then be used to probe subtle changes related to that residue.

The C≡N bond is associated with a large dipole moment, so it generates an intense sharp absorption peak at ~2200–2300 cm^{-1}. Importantly, this peak is separated from those resulting from vibrations that occur naturally in proteins. The C≡N vibration is very sensitive to its local environment and routinely used as a vibrational probe in IR spectrometry. *p*-Cyanophenylalanine (Fig. 2 (31)) was incorporated into myoglobin to probe the binding of ligands (e.g. CN^-, NO, O_2, and CO) to the myoglobin iron center [104]. It is worth noting that *p*-cyanophenylalanine is also fluorescent under UV excitation, and its fluorescence is sensitive to its environment [105]. So both fluorescence and IR spectroscopy can be used to interpret the dynamics of *p*-cyanophenylalanine-containing proteins.

Similarly, the anti-symmetric stretch of the azide functional group absorbs at ~2100 cm^{-1}, which is also in the clear spectral window of natural proteins. *p*-Azidophenylalanine (Fig. 2 (11)) has been incorporated into the G-protein coupled receptor rhodopsin to monitor its structural changes induced by light [106, 107]. In other examples, the nitro symmetric stretching of *p*-nitrophenylalanine (Fig. 2 (29)) and 3-nitrotyrosine (Fig. 2 (30)) has been used in IR and Raman spectrometry [108–110].

Residue labeling for nuclear magnetic resonance

Nuclear magnetic resonance (NMR) is a spectroscopic technique which makes use of the magnetic properties of atom nuclei [111]. The resonance of magnetic nuclei is sensitive to the electromagnetic environment. The magnetic nuclei commonly used for protein analysis are ^1H, ^{13}C, and ^{15}N [111]. Traditional isotope enrichment methods often globally introduce magnetic nuclei into proteins. The assignment of NMR peaks can be challenging for large proteins.

Site-specific labeling of proteins can be accomplished by using genetic code expansion technology [112]. For example, $^{13}C/^{15}N$-labeled *p*-methoxyphenylalanine (Fig. 2 (32)), *p*-trifluoromethoxyphenylalanine (Fig. 2 (33)), and *p*-trifluoromethylphenylalanine (Fig. 2 (34)) have been incorporated into proteins by using the corresponding engineered aminoacyl tRNA/synthetase pairs [32, 113, 114]. This strategy can, in theory, be used to isotopically label nonnative amino acid residues for which orthogonal tRNA/synthetase pairs are available.

Fluorine (^{19}F) labeling is attractive to NMR researchers. ^{19}F does not exist in any natural amino acid, and it has 100% natural abundance and high NMR sensitivity [115]. The fluorine atom is a good steric mimic of hydrogen. As noted above, fluorine can be introduced into nonnative amino acids that are structurally distinct from natural amino acids. However, when a single fluorine atom is used to replace one hydrogen atom in a natural amino acid, cell endogenous machinery may still use the resulting amino acid. This leads to cell toxicity and the global replacement of the native amino acid residues. One strategy to circumvent this is to use special host strains. For example, *p*-fluorophenylalanine (Fig. 2 (35)) was site-specifically incorporated into dihydrofolate reductase (DHFR) using a *p*-fluorophenylalanine-resistant phenylalanine-auxotrophic *E. coli* strain [116]. The strain harbors a phenylalanyl-tRNA synthetase that can exclude *p*-fluorophenylalanine, and an exogenously-introduced *S. cerevisiae* phenylalanyl tRNA/synthetase pair recognizing the UAG codon. The cells were cultured in media with a large excess of *p*-fluorophenylalanine and limited phenylalanine. The cell-endogenous phenylalanyl tRNA/synthetase pair is specific to phenylalanine whereas the *S. cerevisiae* phenylalanyl tRNA/synthetase pair can incorporate *p*-fluorophenylalanine into DHFR in response to the UAG codon. Another strategy uses nonnative amino acids that are convertible by mild chemical reactions [47, 64]. For example, *o*-nitrobenzyl-caged fluorotyrosine (Fig. 2 (36)) has been genetically encoded in *E. coli* [117]. Upon UV illumination, it can be converted (Fig. 6a). Photocaged lysine, methyllysine, serine, cysteine, and tyrosine have all been genetically encoded [30]. Therefore, this strategy can be used to generate isotope-labeled residues of many native amino acids. ^{19}F, ^{13}C, and ^{15}N can all be used in this type of experiments.

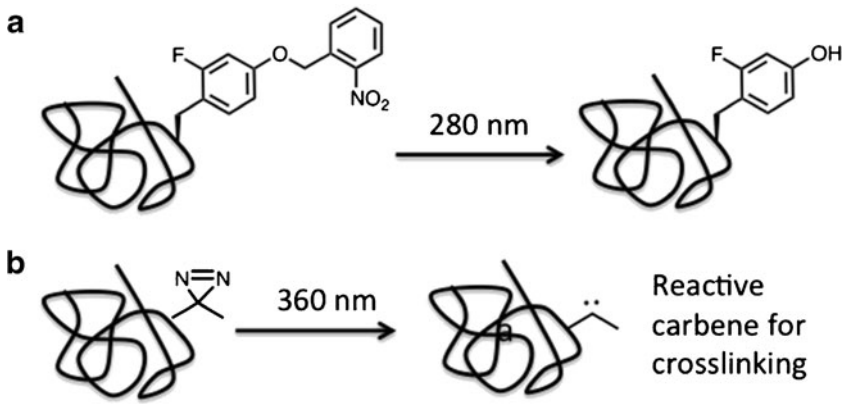


Fig. 6 Schematic diagrams for (**a**) the site-specific incorporation of fluorotyrosine into proteins based on a photocaging reaction, (**b**) the site-specific incorporation of photocrosslinking diazirine and the generation of reactive crosslinking species

Recently, a polyspecific tRNA/synthetase pair that recognizes 3-fluorotyrosine and multiple di and trifluorotyrosines has been reported [118]. It is likely that *E. coli* endogenous machinery was able to exclude di and trifluorotyrosines. However, the specificity for 3-fluorotyrosine remains to be confirmed. Detailed characterization of proteins containing natural tyrosine residues would be needed, because cell endogenous tyrosyl tRNA/synthetase pairs may possibly use 3-fluorotyrosine to replace tyrosine residues.

Further development in this direction may enable NMR analysis of labeled proteins in living cells [112]. Another trend is to develop genetically encoded spin labels which can be used in NMR paramagnetic relaxation enhancement [112]. The interaction between a magnetic nucleus and the unpaired electron of a spin label can provide long-range distance information. Again, the direct encoding of spin labels has been unsuccessful, but reactive chemical handles of nonnative amino acids have been used to conjugate spin labels to proteins [60].

Applications in X-ray crystallography

Introduction of fluorescent, vibrational, and NMR probes at specific sites within proteins is an attractive solution for identifying local dynamic changes in proteins. In contrast, X-ray crystallography provides still images of structural details of proteins. One challenge in X-ray crystallography is to reconstruct atomic structures of proteins from diffraction patterns, and phase information is needed in the process. Single-wavelength anomalous dispersion (SAD) is an often used method [119].

The heavy atom iodine, as *p*-iodophenylalanine (Fig. 2 (24)) or 3-iodotyrosine (Fig. 2 (2)), can be genetically incorporated into proteins in living *E. coli*, yeast, and mammalian cells [42, 44, 88]. As demonstrated, the strong scattering signals from iodine have been used to solve the structures of the bacteriophage T4 lysozyme and the ribosomal protein *N*-acetyltransferase [88, 120]. This method has the advantages that less structural distortion is observed and it requires considerably fewer data than traditional phasing methods [88, 120].

Metal-chelating nonnative amino acids, (8-hydroxyquinolin-3-yl)alanine (Fig. 2 (37)) and (2,2′-bipyridin-5-yl)alanine (Fig. 2 (38)), can be readily incorporated into proteins in *E. coli* [121, 122]. Heavy metal cations can be loaded, and their scattering may also be used to solve the phase. (8-Hydroxyquinolin-3-yl)alanine is preferred for this application because it has less steric hindrance. The structure of TM0665 (*Thermotoga maritima O*-acetylserine sulfhydrylase) was solved by loading Zn^{2+} on to a site-specific (8-hydroxyquinolin-3-yl) alanine [121].

Applications in mass spectrometry

The main application of nonnative amino acids in mass spectrometry (MS) is to stabilize protein–protein or protein–ligand interactions in pull-down assays. A "bait" protein is often immobilized to purify the "prey" (a protein or other type of molecule) that can be identified by one of many MS techniques. When the interaction is weak or transient, crosslinking reagents will be needed to stabilize the interaction. Currently, photocrosslinking nonnative amino acids (Fig. 2 (4, 11, 13, 39)) containing benzophenone, aryl azide and diazirine are available for site-specific protein labeling in *E. coli*, yeast, and mammalian cells [33, 43, 44, 56, 73–75]. Upon UV illumination, side chains of these amino acid form reactive species that can crosslink surrounding interacting molecules (Fig. 6b). Light can be delivered to living cells, resulting in the stabilization of biological interactions in situ.

In one study, *p*-benzoylphenylalanine (Fig. 2 (4)) was incorporated into an endoplasmic reticulum (ER)-associated protein degradation substrate, which was able to crosslink the ubiquitin ligase Hrd1p [123]. Multiple sites in the substrate protein were tested, and crosslinks were observed between the misfolded segment of the substrate and Hrd1p

on the luminal side of the ER membrane. The authors proposed a model for retrotranslocation of the misfolded luminal ER protein by Hrd1p [123].

In another report, *p*-azidophenylalanine (Fig. 2 (11)) was incorporated into different positions of the corticotropin releasing factor receptor type 1 (CRF-R1) [124]. The G-protein coupled receptor CRF-R1 binds peptide hormones to mediate various downstream responses. The cross-linking patterns with multiple peptide ligands were compared, suggesting that these ligands were not tightly bound and different ligands interacted with different regions of the receptor in native complexes.

The aliphatic photocrosslinker 3′-azibutyl-*N*-carbamoyl-lysine (Fig. 2 (39)) has been genetically introduced into the active site of the kinase CDK5 (cyclin-dependent kinase 5) [125]. The downstream kinase PAK1 was identified by crosslinking. A similar diazirine amino acid was used to profile the substrates of the acid-protection chaperone HdeA in *E. coli* periplasm [126]. The flexibility of the long aliphatic chain was important for this particular experiment, because *p*-benzoylphenylalanine (Fig. 2 (4)) did not result in satisfactory crosslinking.

Currently, there are multiple choices of nonnative amino acids for photocrosslinking experiments. The incorporation efficiencies of these amino acids are comparable in *E. coli*. We observed that 3′-azibutyl-*N*-carbamoyl-lysine (Fig. 2 (39)) could be more efficiently incorporated in mammalian cells by use of an engineered tRNA/synthetase pair [125]. In terms of photocrosslinking, *p*-benzoylphenylalanine (Fig. 2 (4)) can be repeatedly excited by 360-nm light, so crosslinking could be increased by increasing illumination [127]. *p*-Azidophenylalanine (Fig. 2 (11)) is structurally similar to tyrosine and phenylalanine, but under the action of light it is known to form a ring-expanded intermediate that can nonspecifically react with nucleophiles [128]. In addition, some *p*-azidophenylalanine residues in proteins have been reported to be unstable (possibly reduced by cell endogenous reductants or enzymes) [44]. When deciding to use a particular photocrosslinking amino acid, one should take the aforementioned aspects into consideration.

Other applications

Nonnative amino acids have been broadly used to probe the roles of individual residues in proteins. For example, the Stubbe group extensively used tyrosine analogous (3-hydroxytyrosine (Fig. 2 (10)), 3-nitrotyrosine (Fig. 2 (30)), 3-aminotyrosine (Fig. 2 (40)) and multiple fluorotyrosines with different redox potentials or pH sensitivity, to elucidate the mechanism of in-protein long-range radical transfer [118, 129, 130]. These residues were used to replace individual tyrosines in proteins, and radical formation and transfer was monitored by use of EPR and other analytical techniques. Comparison of differently modified proteins provided clues about the roles of individual residues.

Nonnative amino acids containing enriched functional groups may be used to generate biosensors. Recently, 3-hydroxytyrosine was inserted into the green fluorescent protein (GFP) to generate a Cu^{2+}-specific biosensor, because of the metal-chelating ability of 3-hydroxytyrosine [131]. Considering the diversity of encoded nonnative amino acids, further investigation of this technique would generate a large group of useful biosensors for a variety of analytes.

Conclusions and perspectives

The fruitful work of expanding the genetic code of living cells and organisms has provided a large family of nonnative amino acids with different chemical, biochemical, and biophysical properties. Application of this technology has advanced a significant field of science.

In general, the incorporation efficiency of nonnative amino acids in yeast and mammalian cells is lower than in *E. coli.* Progress has been made in enhancing tRNAs and synthetases, to identify more strongly expressing promoters, to increase cell uptake of nonnative amino acids, and to reduce nonsense–mediated mRNA decay in cells [30, 31]. Radical efforts are continuing to encode additional nonnative amino acids, to develop novel orthogonal tRNA/synthetase pairs, and to increase efficiency and adapt existing systems in other unicellular and multicellular systems.

Nonnative amino acids have been coupled with analytical chemistry for investigation of many biological problems. Such coupling has driven (and will continue to drive) progress in many fields of research. Further investigation of genetic code-expansion technology will furnish probes with new or improved properties (e.g. genetically encoded spin labels, red-shifted fluorophores, and novel chemical reactive handles with better reactivity and biocompatibility). The utility of these nonnative amino acids has, however, mostly been demonstrated in proof-of-principle experiments, so adoption of the technology to address rather specific questions is expected in the near future.

The technology of genetic code expansion is very versatile and still rapidly evolving. With the set of tools currently available, researchers with different scientific background are able to investigate problems that were not previously accessible.

Acknowledgements I acknowledge support from the University of California, Riverside, and suggestions on the manuscript from Dr Si Chen and Suzan Youssef.

References

1. Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. Microbiol Rev 56:229–264
2. Ambrogelly A, Palioura S, Soll D (2007) Natural expansion of the genetic code. Nat Chem Biol 3:29–35
3. Jukes TH, Osawa S (1993) Evolutionary changes in the genetic code. Comp Biochem Physiol B 106:489–494
4. Cone JE, Del Rio RM, Davis JN, Stadtman TC (1976) Chemical characterization of the selenoprotein component of clostridial glycine reductase: identification of selenocysteine as the organoselenium moiety. Proc Natl Acad Sci U S A 73:2659–2663
5. Chambers I, Frampton J, Goldfarb P, Affara N, McBain W, Harrison PR (1986) The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the 'termination' codon, tga. EMBO J 5:1221–1227
6. Zinoni F, Birkmann A, Stadtman TC, Bock A (1986) Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. Proc Natl Acad Sci U S A 83:4650–4654
7. Srinivasan G, James CM, Krzycki JA (2002) Pyrrolysine encoded by UAG in archaea: charging of a UAG-decoding specialized tRNA. Science 296:1459–1462
8. Hao B, Gong W, Ferguson TK, James CM, Krzycki JA, Chan MK (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science 296:1462–1466
9. Bock A, Forchhammer K, Heider J, Leinfelder W, Sawers G, Veprek B, Zinoni F (1991) Selenocysteine: the 21st amino acid. Mol Microbiol 5:515–520
10. Krzycki JA (2005) The direct genetic encoding of pyrrolysine. Curr Opin Microbiol 8:706–712
11. Boyce M, Bertozzi CR (2011) Bringing chemistry to life. Nat Methods 8:638–642
12. Bertozzi CR (2011) A decade of bioorthogonal chemistry. Acc Chem Res 44:651–653
13. Kaiser ET, Lawrence DS, Rokita SE (1985) The chemical modification of enzymatic specificity. Annu Rev Biochem 54:565–595
14. Witus LS, Moore T, Thuronyi BW, Esser-Kahn AP, Scheck RA, Iavarone AT, Francis MB (2010) Identification of highly reactive sequences for PLP-mediated bioconjugation using a combinatorial peptide library. J Am Chem Soc 132:16812–16817
15. Wu P, Shui W, Carlson BL, Hu N, Rabuka D, Lee J, Bertozzi CR (2009) Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. Proc Natl Acad Sci U S A 106:3000–3005
16. Kimmerlin T, Seebach D (2005) '100 years of peptide synthesis': ligation methods for peptide and protein synthesis with applications to beta-peptide assemblies. J Pept Res 65:229–260
17. Muralidharan V, Muir TW (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. Nat Methods 3:429–438
18. Shang S, Tan Z, Danishefsky SJ (2011) Application of the logic of cysteine-free native chemical ligation to the synthesis of human parathyroid hormone (hPTH). Proc Natl Acad Sci U S A 108:5986–5989
19. 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
20. Hendrickson TL, de Crecy-Lagard V, Schimmel P (2004) Incorporation of nonnatural amino acids into proteins. Annu Rev Biochem 73:147–176
21. Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG (1989) A general method for site-specific incorporation of unnatural amino acids into proteins. Science 244:182–188
22. Hartman MC, Josephson K, Szostak JW (2006) Enzymatic aminoacylation of tRNA with unnatural amino acids. Proc Natl Acad Sci U S A 103:4356–4361
23. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T (2001) Cell-free translation reconstituted with purified components. Nat Biotechnol 19:751–755
24. Saks ME, Sampson JR, Nowak MW, Kearney PC, Du F, Abelson JN, Lester HA, Dougherty DA (1996) An engineered tetrahymena $tRNA^{Gln}$ for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression. J Biol Chem 271:23169–23175
25. Kohrer C, Yoo JH, Bennett M, Schaack J, RajBhandary UL (2003) A possible approach to site-specific insertion of two different unnatural amino acids into proteins in mammalian cells via nonsense suppression. Chem Biol 10:1095–1102
26. 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
27. Murgola EJ (1985) tRNA, suppression, and the code. Annu Rev Genet 19:57–80
28. Wang L, Xie J, Schultz PG (2006) Expanding the genetic code. Annu Rev Biophys Biomol Struct 35:225–249
29. Young TS, Schultz PG (2010) Beyond the canonical 20 amino acids: expanding the genetic lexicon. J Biol Chem 285:11039–11044
30. Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. Annu Rev Biochem 79:413–444
31. Wang Q, Parrish AR, Wang L (2009) Expanding the genetic code for biological studies. Chem Biol 16:323–336
32. Wang L, Brock A, Herberich B, Schultz PG (2001) Expanding the genetic code of *Escherichia coli*. Science 292:498–500
33. Yanagisawa T, Ishii R, Fukunaga R, Kobayashi T, Sakamoto K, Yokoyama S (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode n(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. Chem Biol 15:1187–1197
34. Kwon I, Tirrell DA (2007) Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells. J Am Chem Soc 129:10431–10437
35. Anderson JC, Wu N, Santoro SW, Lakshman V, King DS, Schultz PG (2004) An expanded genetic code with a functional quadruplet codon. Proc Natl Acad Sci U S A 101:7566–7571
36. 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
37. Huang Y, Russell WK, Wan W, Pai PJ, Russell DH, Liu W (2010) A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in *Escherichia coli*. Mol Biosyst 6:683–686
38. Johnson DB, Xu J, Shen Z, Takimoto JK, Schultz MD, Schmitz RJ, Xiang Z, Ecker JR, Briggs SP, Wang L (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol 7:779–786
39. Wan W, Huang Y, Wang Z, Russell WK, Pai PJ, Russell DH, Liu WR (2010) A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. Angew Chem Int Ed Engl 49:3211–3214
40. Wang K, Neumann H, Peak-Chew SY, Chin JW (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nat Biotechnol 25:770–777
41. Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature 464:441–444

42. Sakamoto K, Hayashi A, Sakamoto A, Kiga D, Nakayama H, Soma A, Kobayashi T, Kitabatake M, Takio K, Saito K, Shirouzu M, Hirao I, Yokoyama S (2002) Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. Nucleic Acids Res 30:4692–4699
43. Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, Schultz PG (2003) An expanded eukaryotic genetic code. Science 301:964–967
44. Liu W, Brock A, Chen S, Chen S, Schultz PG (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. Nat Methods 4:239–244
45. Hancock SM, Uprety R, Deiters A, Chin JW (2010) Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. J Am Chem Soc 132:14819–14824
46. Mukai T, Kobayashi T, Hino N, Yanagisawa T, Sakamoto K, Yokoyama S (2008) Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. Biochem Biophys Res Commun 371:818–822
47. Chen PR, Groff D, Guo J, Ou W, Cellitti S, Geierstanger BH, Schultz PG (2009) A facile system for encoding unnatural amino acids in mammalian cells. Angew Chem Int Ed Engl 48:4052–4055
48. Liu CC, Mack AV, Tsao ML, Mills JH, Lee HS, Choe H, Farzan M, Schultz PG, Smider VV (2008) Protein evolution with an expanded genetic code. Proc Natl Acad Sci U S A 105:17688–17693
49. Liu CC, Mack AV, Brustad EM, Mills JH, Groff D, Smider VV, Schultz PG (2009) Evolution of proteins with genetically encoded "chemical warheads". J Am Chem Soc 131:9616–9617
50. Wang F, Robbins S, Guo J, Shen W, Schultz PG (2010) Genetic incorporation of unnatural amino acids into proteins in *Mycobacterium tuberculosis*. PLoS One 5:e9354
51. Young TS, Ahmad I, Brock A, Schultz PG (2009) Expanding the genetic repertoire of the methylotrophic yeast *Pichia pastoris*. Biochemistry 48:2643–2653
52. Shen B, Xiang Z, Miller B, Louie G, Wang W, Noel JP, Gage FH, Wang L (2011) Genetically encoding unnatural amino acids in neural stem cells and optically reporting voltage-sensitive domain changes in differentiated neurons. Stem Cells 29:1231–1240
53. Mukai T, Wakiyama M, Sakamoto K, Yokoyama S (2010) Genetic encoding of non-natural amino acids in Drosophila melanogaster schneider 2 cells. Protein Sci 19:440–448
54. Greiss S, Chin JW (2011) Expanding the genetic code of an animal. J Am Chem Soc 133:14196–14199
55. Wang L, Zhang Z, Brock A, Schultz PG (2003) Addition of the keto functional group to the genetic code of *Escherichia coli*. Proc Natl Acad Sci U S A 100:56–61
56. Chin JW, Martin AB, King DS, Wang L, Schultz PG (2002) Addition of a photocrosslinking amino acid to the genetic code of *Escherichiacoli*. Proc Natl Acad Sci U S A 99:11020–11024
57. Wang W, Takimoto JK, Louie GV, Baiga TJ, Noel JP, Lee KF, Slesinger PA, Wang L (2007) Genetically encoding unnatural amino acids for cellular and neuronal studies. Nat Neurosci 10:1063–1072
58. Huang Y, Wan W, Russell WK, Pai PJ, Wang Z, Russell DH, Liu W (2010) Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications. Bioorg Med Chem Lett 20:878–880
59. 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
60. Fleissner MR, Brustad EM, Kalai T, Altenbach C, Cascio D, Peters FB, Hideg K, Peuker S, Schultz PG, Hubbell WL (2009) Site-directed spin labeling of a genetically encoded unnatural amino acid. Proc Natl Acad Sci U S A 106:21637–21642
61. Prescher JA, Bertozzi CR (2005) Chemistry in living systems. Nat Chem Biol 1:13–21
62. Zhang Z, Wang L, Brock A, Schultz PG (2002) The selective incorporation of alkenes into proteins in *Escherichia coli*. Angew Chem Int Ed Engl 41:2840–2842
63. H-w A, Shen W, Brustad E, Schultz PG (2010) Genetically encoded alkenes in yeast. Angew Chem Int Ed Engl 49:935–937
64. Ai HW, Lee JW, Schultz PG (2010) A method to site-specifically introduce methyllysine into proteins in *E. coli*. Chem Commun 46:5506–5508
65. Kirshenbaum K, Arora PS (2008) Cross-dressing proteins by olefin metathesis. Nat Chem Biol 4:527–528
66. Kodama K, Fukuzawa S, Nakayama H, Kigawa T, Sakamoto K, Yabuki T, Matsuda N, Shirouzu M, Takio K, Tachibana K, Yokoyama S (2006) Regioselective carbon–carbon bond formation in proteins with palladium catalysis; new protein chemistry by organometallic chemistry. ChemBioChem 7:134–139
67. Kodama K, Fukuzawa S, Nakayama H, Sakamoto K, Kigawa T, Yabuki T, Matsuda N, Shirouzu M, Takio K, Yokoyama S, Tachibana K (2007) Site-specific functionalization of proteins by organopalladium reactions. ChemBioChem 8:232–238
68. Song W, Wang Y, Qu J, Lin Q (2008) Selective functionalization of a genetically encoded alkene-containing protein via "photoclick chemistry" in bacterial cells. J Am Chem Soc 130:9654–9655
69. Song W, Wang Y, Yu Z, Vera CI, Qu J, Lin Q (2010) A metabolic alkene reporter for spatiotemporally controlled imaging of newly synthesized proteins in mammalian cells. ACS Chem Biol 5:875–885
70. Wang J, Zhang W, Song W, Wang Y, Yu Z, Li J, Wu M, Wang L, Zang J, Lin Q (2010) A biosynthetic route to photoclick chemistry on proteins. J Am Chem Soc 132:14812–14818
71. Wang J, Schiller SM, Schultz PG (2007) A biosynthetic route to dehydroalanine-containing proteins. Angew Chem 119:6973–6975
72. Ayyadurai N, Prabhu NS, Deepankumar K, Jang YJ, Chitrapriya N, Song E, Lee N, Kim SK, Kim B-G, Soundrarajan N, Lee S, Cha HJ, Budisa N, Yun H (2011) Bioconjugation of L-3,4-dihydroxyphenylalanine containing protein with a polysaccharide. Bioconjug Chem 22:551–555
73. Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG (2002) Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. J Am Chem Soc 124:9026–9027
74. Laughlin ST, Bertozzi CR (2007) Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via staudinger ligation. Nat Protoc 2:2930–2944
75. Nguyen DP, Lusic H, Neumann H, Kapadnis PB, Deiters A, Chin JW (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/$tRNA_{cua}$ pair and click chemistry. J Am Chem Soc 131:8720–8721
76. Fekner T, Li X, Lee MM, Chan MK (2009) A pyrrolysine analogue for protein click chemistry. Angew Chem Int Ed Engl 48:1633–1635
77. Hao Z, Song Y, Lin S, Yang M, Liang Y, Wang J, Chen PR (2011) A readily synthesized cyclic pyrrolysine analogue for site-specific protein "click" labeling. Chem Commun (Camb) 47:4502–4504
78. Deiters A, Schultz PG (2005) In vivo incorporation of an alkyne into proteins in *Escherichia coli*. Bioorg Med Chem Lett 15:1521–1524
79. Hong V, Steinmetz NF, Manchester M, Finn MG (2010) Labeling live cells by copper-catalyzed alkyne–azide click chemistry. Bioconjug Chem 21:1912–1916

80. Baskin JM, Prescher JA, Laughlin ST, Agard NJ, Chang PV, Miller IA, Lo A, Codelli JA, Bertozzi CR (2007) Copper-free click chemistry for dynamic in vivo imaging. Proc Natl Acad Sci U S A 104:16793–16797
81. Plass T, Milles S, Koehler C, Schultz C, Lemke EA (2011) Genetically encoded copper-free click chemistry. Angew Chem Int Ed Engl 50:3878–3881
82. 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
83. Tsao ML, Tian F, Schultz PG (2005) Selective staudinger modification of proteins containing p-azidophenylalanine. ChemBioChem 6:2147–2149
84. Li X, Fekner T, Ottesen JJ, Chan MK (2009) A pyrrolysine analogue for site-specific protein ubiquitination. Angew Chem Int Ed Engl 48:9184–9187
85. Virdee S, Kapadnis PB, Elliott T, Lang K, Madrzak J, Nguyen DP, Riechmann L, Chin JW (2011) Traceless and site-specific ubiquitination of recombinant proteins. J Am Chem Soc 133:10708–10711
86. Nguyen DP, Elliott T, Holt M, Muir TW, Chin JW (2011) Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. J Am Chem Soc 133:11418–11421
87. Brustad E, Bushey ML, Lee JW, Groff D, Liu W, Schultz PG (2008) A genetically encoded boronate-containing amino acid. Angew Chem Int Ed Engl 47:8220–8223
88. Xie J, Wang L, Wu N, Brock A, Spraggon G, Schultz PG (2004) The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. Nat Biotechnol 22:1297–1301
89. Mehl RA, Anderson JC, Santoro SW, Wang L, Martin AB, King DS, Horn DM, Schultz PG (2003) Generation of a bacterium with a 21 amino acid genetic code. J Am Chem Soc 125:935–939
90. Li N, Lim RK, Edwardraja S, Lin Q (2011) Copper-free sonogashira cross-coupling for functionalization of alkyne-encoded proteins in aqueous medium and in bacterial cells. J Am Chem Soc 133:15316–15319
91. Spicer CD, Davis BG (2011) Palladium-mediated site-selective Suzuki-Miyaura protein modification at genetically encoded aryl halides. Chem Commun (Camb) 47:1698–1700
92. Behrens CR, Hooker JM, Obermeyer AC, Romanini DW, Katz EM, Francis MB (2011) Rapid chemoselective bioconjugation through oxidative coupling of anilines and aminophenols. J Am Chem Soc 133:16398–16401
93. Best MD (2009) Click chemistry and bioorthogonal reactions: Unprecedented selectivity in the labeling of biological molecules. Biochemistry 48:6571–6584
94. Wang J, Xie J, Schultz PG (2006) A genetically encoded fluorescent amino acid. J Am Chem Soc 128:8738–8739
95. Summerer D, Chen S, Wu N, Deiters A, Chin JW, Schultz PG (2006) A genetically encoded fluorescent amino acid. Proc Natl Acad Sci U S A 103:9785–9789
96. Lee HS, Guo J, Lemke EA, Dimla RD, Schultz PG (2009) Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in saccharomyces cerevisiae. J Am Chem Soc 131:12921–12923
97. Charbon G, Brustad E, Scott KA, Wang J, Lobner-Olesen A, Schultz PG, Jacobs-Wagner C, Chapman E (2011) Subcellular protein localization by using a genetically encoded fluorescent amino acid. ChemBioChem 12:1818–1821
98. Saraogi I, Zhang D, Chandrasekaran S, Shan SO (2011) Site-specific fluorescent labeling of nascent proteins on the translating ribosome. J Am Chem Soc 133:14936–14939
99. Lacey VK, Parrish AR, Han S, Shen Z, Briggs SP, Ma Y, Wang L (2011) A fluorescent reporter of the phosphorylation status of the substrate protein STAT3. Angew Chem Int Ed Engl 50:8692–8696
100. Tsao ML, Summerer D, Ryu Y, Schultz PG (2006) The genetic incorporation of a distance probe into proteins in *Escherichia coli*. J Am Chem Soc 128:4572–4573
101. Neumann H, Hazen JL, Weinstein J, Mehl RA, Chin JW (2008) Genetically encoding protein oxidative damage. J Am Chem Soc 130:4028–4033
102. Gershkovich AA, Kholodovych VV (1996) Fluorogenic substrates for proteases based on intramolecular fluorescence energy transfer (iFETs). J Biochem Biophys Methods 33:135–162
103. Kong J, Yu S (2007) Fourier transform infrared spectroscopic analysis of protein secondary structures. Acta Biochim Biophys Sin (Shanghai) 39:549–559
104. Schultz KC, Supekova L, Ryu Y, Xie J, Perera R, Schultz PG (2006) A genetically encoded infrared probe. J Am Chem Soc 128:13984–13985
105. Taskent-Sezgin H, Chung J, Patsalo V, Miyake-Stoner SJ, Miller AM, Brewer SH, Mehl RA, Green DF, Raleigh DP, Carrico I (2009) Interpretation of p-cyanophenylalanine fluorescence in proteins in terms of solvent exposure and contribution of side-chain quenchers: a combined fluorescence, IR and molecular dynamics study. Biochemistry 48:9040–9046
106. Ye S, Huber T, Vogel R, Sakmar TP (2009) FTIR analysis of GPCR activation using azido probes. Nat Chem Biol 5:397–399
107. Ye S, Zaitseva E, Caltabiano G, Schertler GF, Sakmar TP, Deupi X, Vogel R (2010) Tracking g-protein-coupled receptor activation using genetically encoded infrared probes. Nature 464:1386–1389
108. Smith EE, Linderman BY, Luskin AC, Brewer SH (2011) Probing local environments with the infrared probe: L-4-nitrophenylalanine. J Phys Chem B 115:2380–2385
109. Izzo GE, Jordan F, Mendelsohn R (1982) Resonance Raman and 500-mHz proton NMR studies of tyrosine modification in hen egg white lysozyme. J Am Chem Soc 104:3178–3182
110. Quaroni L, Smith WE (1999) Nitration of internal tyrosine of cytochrome c probed by resonance raman scattering. Biospectroscopy 5:S71–S76
111. Widmer H, Jahnke W (2004) Protein NMR in biomedical research. Cell Mol Life Sci 61:580–599
112. Jones DH, Cellitti SE, Hao X, Zhang Q, Jahnz M, Summerer D, Schultz PG, Uno T, Geierstanger BH (2010) Site-specific labeling of proteins with NMR-active unnatural amino acids. J Biomol NMR 46:89–100
113. Cellitti SE, Jones DH, Lagpacan L, Hao X, Zhang Q, Hu H, Brittain SM, Brinker A, Caldwell J, Bursulaya B, Spraggon G, Brock A, Ryu Y, Uno T, Schultz PG, Geierstanger BH (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
114. Jackson JC, Hammill JT, Mehl RA (2007) Site-specific incorporation of a (19)F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. J Am Chem Soc 129:1160–1166
115. Suzuki Y, Buer BC, Al-Hashimi HM, Marsh EN (2011) Using fluorine nuclear magnetic resonance to probe changes in the structure and dynamics of membrane-active peptides interacting with lipid bilayers. Biochemistry 50:5979–5987
116. Furter R (1998) Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in *Escherichia coli*. Protein Sci 7:419–426
117. Wilkins BJ, Marionni S, Young DD, Liu J, Wang Y, Di Salvo ML, Deiters A, Cropp TA (2010) Site-specific incorporation of fluorotyrosines into proteins in *Escherichia coli* by photochemical disguise. Biochemistry 49:1557–1559
118. Minnihan EC, Young DD, Schultz PG, Stubbe J (2011) Incorporation of fluorotyrosines into ribonucleotide reductase using an

evolved, polyspecific aminoacyl-tRNA synthetase. J Am Chem Soc 133:15942–15945
119. Brodersen DE, de La Fortelle E, Vonrhein C, Bricogne G, Nyborg J, Kjeldgaard M (2000) Applications of single-wavelength anomalous dispersion at high and atomic resolution. Acta Crystallogr D: Biol Crystallogr 56:431–441
120. Sakamoto K, Murayama K, Oki K, Iraha F, Kato-Murayama M, Takahashi M, Ohtake K, Kobayashi T, Kuramitsu S, Shirouzu M, Yokoyama S (2009) Genetic encoding of 3-iodo-L-tyrosine in *Escherichia coli* for single-wavelength anomalous dispersion phasing in protein crystallography. Structure 17:335–344
121. Lee HS, Spraggon G, Schultz PG, Wang F (2009) Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe. J Am Chem Soc 131:2481–2483
122. Xie J, Liu W, Schultz PG (2007) A genetically encoded bidentate, metal-binding amino acid. Angew Chem Int Ed Engl 46:9239–9242
123. Carvalho P, Stanley AM, Rapoport TA (2010) Retrotranslocation of a misfolded luminal er protein by the ubiquitin-ligase hrd1p. Cell 143:579–591
124. Coin I, Perrin MH, Vale WW, Wang L (2011) Photo-cross-linkers incorporated into g-protein-coupled receptors in mammalian cells: a ligand comparison. Angew Chem Int Ed Engl
125. Ai HW, Shen W, Sagi A, Chen PR, Schultz PG (2011) Probing protein-protein interactions with a genetically encoded photocrosslinking amino acid. ChemBioChem 12:1854–1857
126. Zhang M, Lin S, Song X, Liu J, Fu Y, Ge X, Fu X, Chang Z, Chen PR (2011) A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. Nat Chem Biol 7:671–677
127. Kawamura A, Hindi S, Mihai DM, James L, Aminova O (2008) Binding is not enough: Flexibility is needed for photocrosslinking of Lck kinase by benzophenone photoligands. Bioorg Med Chem 16:8824–8829
128. Baruah H, Puthenveetil S, Choi YA, Shah S, Ting AY (2008) An engineered aryl azide ligase for site-specific mapping of protein–protein interactions through photo-cross-linking. Angew Chem Int Ed Engl 47:7018–7021
129. Minnihan EC, Seyedsayamdost MR, Stubbe J (2009) Use of 3-aminotyrosine to examine the pathway dependence of radical propagation in *Escherichia coli* ribonucleotide reductase. Biochemistry 48:12125–12132
130. Minnihan EC, Seyedsayamdost MR, Uhlin U, Stubbe J (2011) Kinetics of radical intermediate formation and deoxynucleotide production in 3-aminotyrosine-substituted *Escherichia coli* ribonucleotide reductases. J Am Chem Soc 133:9430–9440
131. Ayyadurai N, Saravanan Prabhu N, Deepankumar K, Lee SG, Jeong HH, Lee CS, Yun H (2011) Development of a selective, sensitive, and reversible biosensor by the genetic incorporation of a metal-binding site into green fluorescent protein. Angew Chem Int Ed Engl 50:6534–6537