Site specific protein labeling by enzymatic posttranslational modification†

Murat Sunbul and Jun Yin*

Received 8th May 2009
First published as an Advance Article on the web 28th July 2009
DOI: 10.1039/b908687k

Site specific protein labeling plays a key role in elucidating the function of the proteins at the molecular level by revealing their locations in the cell, their interaction networks with other cellular components and the dynamic mechanisms of their bio-generation, trafficking and degradation in response to regulatory signals in a biological system. Site specific protein labeling is, in essence, artificial modification of proteins with new chemical entities at the posttranslational stage. Based on the analogy between protein labeling and protein posttranslational modification, enzymatic tools have been developed for site specific and efficient labeling of target proteins with chemical probes of diverse structures and functionalities. This perspective surveys a number of protein labeling methods based on the application of protein posttranslational modification enzymes.

Introduction

Proteins sustain the normal life cycle of a cell by carrying out virtually all aspects of its biological functions. Specific and efficient protein labeling methods combined with advanced fluorescence microscopy have enabled live cell imaging of protein distribution, translocation and interaction with other proteins and biomolecules.[1,2] In this way, actions of the protein under investigation can be visualized in real time for the elucidation of its function in a complex biological network.

For many years, green fluorescent protein (GFP) and its variants have been the default choice to label and image proteins due to their biological compatibility and ability to be expressed as fusion proteins in cellular environments.[3] Much progress has been made to generate fluorescent protein variants with diverse spectra, improved efficiency for chromophore maturation and enhanced fluorescence brightness.[4,5] Compared to fluorescent proteins, small molecule fluorophores provide a broader range

Department of Chemistry, The University of Chicago, 929 East 57th Street, GCIS E505, IL 60637, USA. E-mail: junyin@uchicago.edu; Fax: (+1) 773 702 0805; Tel: (+1) 773 834 5881

† This paper is part of an *Organic & Biomolecular Chemistry* web theme issue on protein labelling and chemical proteomics. Guest editors: Ed Tate and Matthew Bogyo.

Murat Sunbul

Murat Sunbul was born in Turkey where he received his BS degree in Chemistry from the Middle East Technical University (METU) in 2004. He then worked with Professor Cihangir Tanyeli on the synthesis of asymmetric amino alcohol ligands as catalysts for carbon-carbon bond formation reactions. He obtained his MS degree in 2005 from METU. He is currently a PhD student at the University of Chicago. His research interests include developing new protein labeling techniques and directed evolution of posttranslational modification enzymes.

Jun Yin

Jun Yin grew up in Beijing, China. He received his BS in chemistry from Peking University in 1995. He carried out undergraduate research with Professors Bao-Huai Wang, You-Min Zhang and Zhi-Fen Li on characterizing the physical properties of fullerene molecules. He graduated with an MS in chemistry from the Rutgers University in 1997 working with Professor Stephen Anderson on amyloid peptide fibril formation. He received a PhD in chemistry from the University of California at Berkeley in 2003 under the guidance of Professor Peter G. Schultz working on the evolution of catalytic antibodies. He then carried out postdoctoral research with Professor Christopher T. Walsh at Harvard Medical School on phosphopantetheinyl transferase catalyzed protein modification. He began his independent research at the University of Chicago in 2006. His current research is to engineer enzymes for protein posttranslational modification and natural product biosynthesis.

of excitation and emission wavelengths, better photostability and higher fluorescence quantum yield.[6] However, the tool box of small molecule fluorophores has been underused in cell imaging studies due to the difficulties in establishing a specific linkage between the synthetic fluorophores and the target proteins inside the cell. Many labeling experiments have been done with synthetic fluorophore conjugated antibodies that would bind to the target proteins. In these experiments, cells need to be permeabilized and fixed to allow the binding of antibodies to the intracellular proteins, and often, the low specificity of the antibodies may give rise to high background labeling. Thus, an efficient method for site specific attachment of synthetic fluorophores to the target proteins is in high demand for live cell imaging.[7–10]

Current methods for protein labeling with synthetic fluorophores generally fall into two categories. The first category takes advantage of bioorthogonal reactions for the specific attachment of small molecule probes to target proteins. For example, the tetracysteine peptide motif has been used to chelate biarsenical conjugated fluorophores to the target proteins.[11] Bioorthogonal reactions such as Staudinger ligation[12] and Huisgen azide-alkyne cycloaddition[13,14] have been developed for conjugation of proteins with a variety of probes. They have been applied to derivatize azide functionalized cell surface glycans with synthetic fluorophores for cell imaging studies[15–17] or oligonucleotides to induce specific cell-cell interactions.[18,19] The second category of protein labeling reactions is based on enzyme catalyzed protein posttranslational modifications, which is the focus of this article. Besides these methods, specific and high affinity binding between protein and small molecule ligands, such as the binding of dihydrofolate reductase to probe conjugated trimethoprim derivatives, has also been adopted for protein labeling.[20] And at the stage of protein translation, target proteins can be directly labeled with diverse functionalities by incorporation of unnatural amino acids containing fluorophores, bioorthogonal reactive groups, metal chelating groups, photo crosslinking labels and photo switchable groups.[21,22]

There are a few characteristics for a good molecular probe and labeling strategy. The probes preferably should be small in size and chemically stable, with minimal interference of the folding and biological functions of the target protein. The labeling reaction should be highly efficient and adaptable for live cell imaging, preferably establishing a covalent linkage between the synthetic probe and a specific residue in the target protein. According to these criteria, enzyme catalyzed protein labeling methods would have distinctive advantages over non-enzymatic methods. First, enzymatic reactions are generally fast and efficient. Second and perhaps more importantly, enzyme catalyzed reactions are selective, allowing site-specific protein labeling with low background for high resolution imaging of target proteins inside the cell. In fact, protein labeling by various chemical probes is highly analogous to protein posttranslational modification utilized by the cell for diversifying the structure and function of proteins.[23–25] These modifications equip the proteins with new chemical functionalities not provided by the twenty common proteinogenic amino acids so that the modified proteins acquire new biological activities. Thus, enzyme catalyzed protein posttranslational modification provides a diverse and versatile platform for design and development of efficient tools for site specific protein labeling.

Protein labeling catalyzed by posttranslational modification enzymes

1. Transglutaminase

Transglutaminase (TGase) catalyzes protein crosslinking in the cell by the formation of an isopeptide bond between Gln and Lys residues (Fig. 1a).[26] In the reaction catalyzed by TGase, acyl transfer from the γ-carboxamide group of a Gln residue in one protein to the ε-amino group of a Lys residue in another protein establishes a ε-(γ-glutamyl)lysine isopeptide linkage between the proteins. Guinea pig liver TGase (gpTGase) demonstrates high specificity for Gln containing proteins as the acyl donors and low specificity for Lys containing substrates as acyl acceptors.[27] This feature makes it possible to use gpTGase and other TGases in protein labeling with amine functionalized synthetic probes (Fig. 1b).[28–31] It has been shown that biotin-cadaverine and fluorescein-cadaverine can be readily used as the substrates of gpTGase for the modification of short peptide tags (Q-tag, 6-7 aa) with an embedded Gln residue.[32,33] Q-tag has been fused to the N-terminus of the transmembrane (TM) domain of platelet-derived growth factor receptor (PDGF) for cell surface protein labeling catalyzed by gpTGase.[32] A microbial TGase from *Streptomyces mobaraensis* has been used for conjugating DNA oligonucleotides to proteins.[34] In this case, TGase catalyzes isopeptide bond formation between a Gln residue in a tripeptide linked to DNA and a Lys residue in a short peptide tag (K6-tag) fused to the target protein. TGases earn the reputation of "Nature's biological glues" for their common presence in various cell types and wide substrate pool for protein crosslinking including glutathione S-transferases, actin, myosin, β-tubulin, *etc.*[26] Thus, TGases intrinsically have low substrate specificity with Gln or Lys containing peptide tags fused to the target protein, which may cause unfavorable background labeling of other cellular proteins.

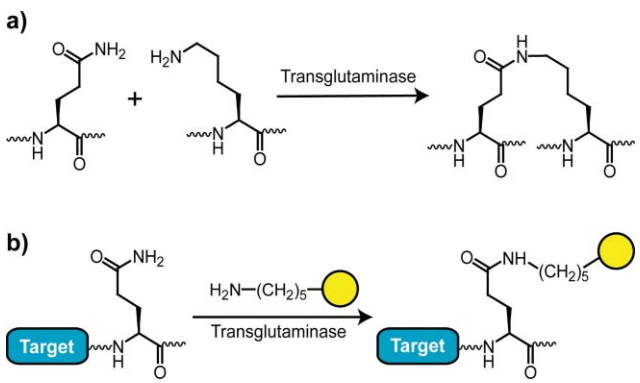

Fig. 1 Transglutaminase (TGase) catalyzed protein modification. (a) The native activity of TGase is to catalyze the transamination reaction between Gln and Lys residues in the modified proteins. (b) In TGase catalyzed protein labeling, the cadaverine (1,5-pentanediamine) functionalized small molecule probe (yellow circle) is transferred to a Gln residue in the Q tag fused to the target protein (blue rectangle). Yellow circles are used to designate small molecule labels and blue rectangles to designate labeled proteins throughout this article.

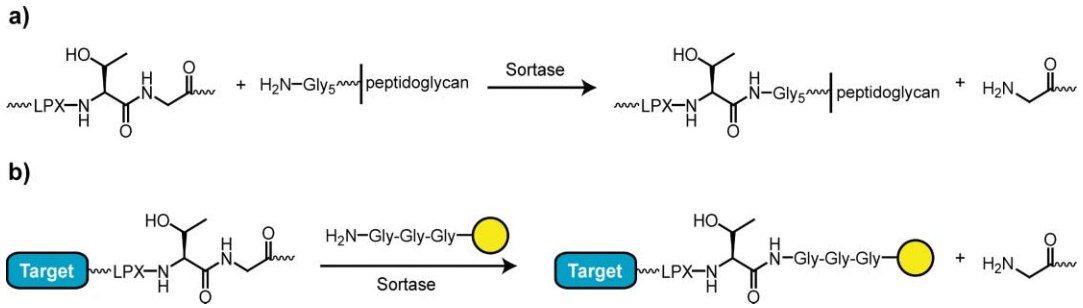

Fig. 2 Sortase catalyzed protein modification. (a) The native activity of sortase is to attach cell surface proteins to peptidoglycans by forming a peptide bond between a Thr and a Gly residue. (b) Target proteins fused to the LPXTG peptide tag can be conjugated with oligo Gly functionalized small molecule probes catalyzed by sortase.

2. Sortase

Sortase, an enzyme with transpeptidase activity, has been used to label proteins. The native function of a sortase is to anchor cell surface proteins to peptidoglycans as part of the cell wall of gram-positive bacteria (Fig. 2a).[35,36] *Staphylococcus aureus* sortase (SrtA) catalyzes the hydrolysis of the peptide bond between Thr and Gly in LPXTG recognition motif in the cell surface protein, and subsequently forms a new peptide bond between the carboxyl group of the exposed Thr and the amino terminal of the pentaglycine cross bridge on the cell surface peptidoglycan.[35] Biotin and fluorescence probes conjugated to peptides with oligo Gly have been demonstrated as the substrate of SrtA for the modification of LPETG peptide tag fused to the C-terminus of the target protein (Fig. 2b).[37–39] For example, oligo Gly peptide conjugated to folate *via* a Lys side chain (G_3K-folate) has been used to label GFP with a C-terminal LPETG tag.[37] It has been shown that mono-, di- and tri-Gly peptides can serve as the substrates of the SrtA, though the rate of the reaction is slower with peptide substrates containing less than three terminal Gly residues.[37] Lately, sortase catalyzed labeling of cell surface proteins has been demonstrated with pentaglycine conjugated fluorophores ligated to a C-terminal LPETG tag on the target protein.[40] N-Terminal protein labeling on the cell surface has also been achieved with sortase. Fluorophores can be attached to the Leu residue in the LPETGG oligopeptide which is ligated with a pentaglycine motif at the N-terminus of the target protein catalyzed by SrtA.[41]

3. Farnesyltransferase

Protein farnesyltransferase (FTase) catalyzes the covalent attachment of 15-carbon farnesyl isoprenoid from farnesyl pyrophosphate to the thiol of a Cys residue within a CaaX motif close to the C-terminus of the modified protein (Fig. 3a).[42] The internal "a" residues in the CaaX motif are aliphatic amino acids and the "X" denotes one of Ala, Ser, Met or Glu residues.[43] The farnesyl group is attached to the modified protein with a thioether bond and acts as a lipid anchor directing subcellular localization of the modified proteins. Protein farnesylation also plays an important role in regulating the activities of many cell signaling proteins including Ras.[44] FTase has been shown to tolerate substitutions of azide, alkyne and diene in farnesyl pyrophosphate and transfers farnesyl analogues to the target protein fused with a C-terminal CaaX tag (Fig. 3b).[45–47] For

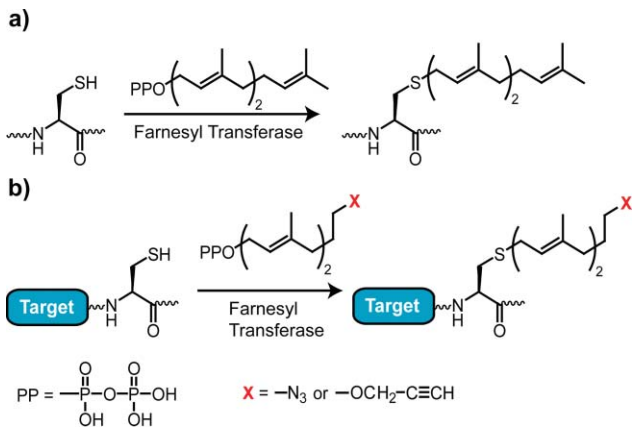

Fig. 3 Farnesyltransferase (FTase) catalyzed protein modification. (a) The native activity of FTase is to attach farnesyl isoprenoid group to a Cys residue of the modified protein. (b) Target proteins fused to the CaaX peptide motif at the C-terminus can be labeled with azide and alkyne functionalized farnesyl isoprenoid analogs.

example, azide functionalized farnesyl diphosphate has been used to label the farnesylated proteins in the cell.[48] After Staudinger ligation for biotin conjugation and enrichment by affinity binding with streptavidin, proteins with azide-farnesyl modification were identified by mass spectrometry. An alkyne functionalized farnesyl group has also been attached to the target protein by FTase and labeled proteins were immobilized on an azide functionalized glass slide.[49] Alkyne, azide or diene functional groups attached to the target proteins by FTase can be further decorated with fluorophores, DNA or other small molecules by Huisgen cycloaddition or Diels–Alder reaction.[45,47,50] FTase has also been engineered to accept biotin conjugated geranylpyrophosphate as the substrate for the labeling and identification of prenylated proteins in the cell proteom.[51] Similar to FTase catalyzed protein labeling, Rab geranylgeranyltransferase (RabGGTase) has been used to label its modification targets with a biotin-geranyl group.[51] FTase catalyzed protein modification would be an efficient method for labeling proteins with bioorthogonal functional groups such alkyne or azide for subsequent chemoselective attachment of small molecule probes. However, since protein farnesylation is a common modification in the cell with nearly 1% mammalian proteins being farnesylated,[52] FTase catalyzed protein labeling would be less useful for cell imaging studies.

4. N-Myristoyl transferase

N-Myristoyl transferase (NMT) covalently attaches myristate (saturated fatty acid with 14-carbons) to the N-terminal Gly of many eukaryotic and viral proteins (Fig. 4a).[53,54] The myristoyl group is transferred from its conjugate with coenzyme A (CoA) to N-terminal Gly, commonly found in a GXXXS motif of the substrate protein (X denotes any amino acid). Protein myristoylation can direct specific protein-protein and protein-membrane interactions and modulate the activity and movement of cell signaling proteins.[44] NMT has been used for protein labeling with myristoyl analogs substituted with terminal azide or alkyne functionalities (Fig. 4b).[55–57] Similar to FTase catalyzed protein labeling, azide or alkyne groups attached to the target protein can be derivatized with secondary labels such as biotin or fluorophores using bioorthogonal Staudinger ligation or Huisgen cycloaddition reactions. Using this method, fluorescent labeling of fatty-acylated proteins has been achieved with permeabilized cells.[58] The relaxed specificity of NMT for protein modification could be advantageous for *in vitro* protein labeling since the only requirement for NMT recognition is the N-terminal GXXXS sequence on the target proteins. However, NMT would be less favorable to be used in cell imaging studies where high specificity in protein labeling is required.

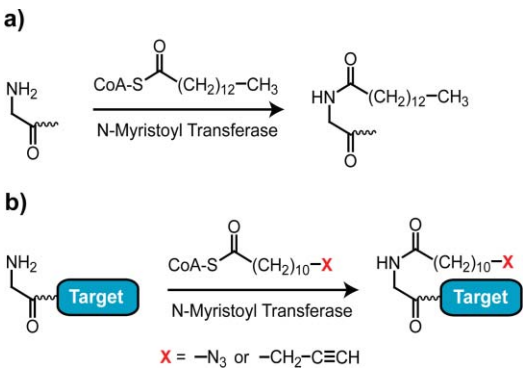

Fig. 4 N-Myristoyl transferase (NMT) catalyzed protein modification. (a) The native activity of NMT is to attach a myristate group to an N-terminal Gly residue of the modified protein. (b) Target proteins fused to the GXXXS peptide motif at the N-terminus can be labeled with azide and alkyne functionalized myristate analogs.

5. Formylglycine-generating enzyme

Formylglycine-generating enzyme (FGE) uses molecular oxygen to oxidize a conserved Cys or Ser residue of sulfatases to form a formylGly residue with an aldehyde functionality (Fig. 5a).[59,60] The formylGly residue at the enzyme active site can then undergo hydration to afford a geminal diol that serves as a nucleophile in sulfate ester hydrolysis catalyzed by the sulfatase enzyme.[61] It was first found that 16-mer peptide derived from arylsulfatase A with an active site Cys can be oxidized by mammalian FGE to generate a formylGly side chain at the site of Cys residue.[60] This result prompted the development of Cys containing peptide tags as short as 6-residue (LCTPSR) that can be posttranslationally modified by FGE and be converted to a peptide with an aldehyde functionality (aldehyde tag).[62,63] The aldehyde tag fused to the target protein can be further derivatized with small molecule

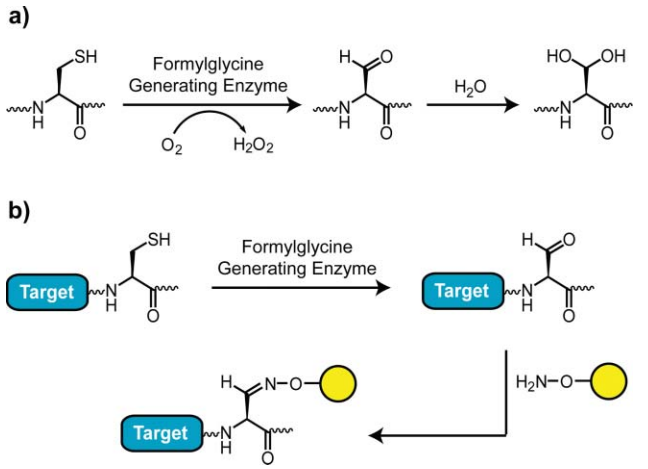

Fig. 5 Formylglycine generating enzyme (FGE) catalyzed protein modification. (a) The native activity of FGE is to convert a Cys side chain into an aldehyde group which forms a geminal diol upon addition of water (b) Target proteins fused to the "aldehyde tag" can be functionalized with aminooxy or hydrazide probes.

probes by hydrazide or aminooxy mediated coupling reactions (Fig. 5b). Using this method, IgG antibody was site specifically conjugated to biotin and fluorophores and cell surface labeling of aldehyde tagged proteins has also been demonstrated.[64] Since FGE is endogenously expressed in mammalian cells, Cys containing peptide tag fused to the target proteins is converted to the aldehyde tag upon expression. Therefore, external addition or coexpression of FGE enzyme is not necessary to initiate the Cys modification.[64]

Protein labeling reactions catalyzed by FTase, NMT and FGE all rely on the enzyme to install a bioorthogonal functional groups such as azide, alkyne or aldehyde on the peptide tag fused to the target protein followed by selective modification of the tag to label the target proteins with small molecule probes. The second step of probe conjugation is not catalyzed by an enzyme. The advantage of such a labeling scheme is that small molecule probes of diverse structures can be attached to the tags by the nonenzymatic coupling reactions since the structure of the probes are not constrained by the substrate specificity of an enzyme. However, at the same time, the nonenzymatic coupling reactions tend to be slow compared to one-step enzyme catalyzed protein labeling reactions. This may limit the use of FTase, NMT and FGE enzymes in the application towards live cell labeling and imaging.

Protein labeling mediated by enzyme self-modification

Enzymes can form covalent adducts with small molecule substrates or mechanism based inhibitors. In these cases, the enzyme itself can serve as a tag for target protein labeling with chemical probes conjugated to enzyme substrate or inhibitor molecules. The advantage of protein labeling with self-modifying enzyme is that the attachment of the chemical probes to the enzyme tag is directly coupled to enzyme catalysis, thus the corresponding labeling reaction is of high efficiency and specificity. The disadvantage of the labeling method is that the enzymes are much larger than a peptide tag. It may be a challenge to express target proteins with a large size tag. In addition, the native activity of the target protein may be affected by the tag. Moreover, the native substrates of

the enzyme tag in the cell may interfere with the protein labeling reaction. In the following sections, we will discuss protein labeling methods based on the self-modification of O^6-alkylguanine-DNA alkyl transferase (hAGT), cutinase, haloalkane dehalogenase and protein self-splicing catalyzed by intein. A β-lactamase mutant has also been used as a tag for the covalent attachment of organic fluorophores that are conjugated to ampicillin.[65]

6. Human O^6-alkylguanine-DNA alkyl transferase (hAGT)

Human O^6-alkylguanine-DNA alkyl transferase (hAGT, 21 kDa) provides an elegant method for protein labeling based on its self-modification with O^6-benzylguanine (BG) conjugated small molecule probes (Fig. 6).[66] In this method, the protein of interest is fused to the enzyme hAGT (SNAP tag) for the transfer of small molecule labels attached to BG to an active site Cys residue in hAGT. Mutants of hAGT have been acquired by directed evolution to increase their activity toward BG functionalized small molecule probes.[67–69] Furthermore, using a combination of yeast surface display and phage display, a hAGT mutant (CLIP tag) has been identified to utilize O^2-benzylcytosine (BC) derivatives for protein labeling.[70] BC conjugated probes are unreactive with hAGT variant SNAP and BG conjugated probes are unreactive with hAGT variant CLIP. Such substrate orthogonality has been applied to simultaneous labeling of different target proteins with distinct fluorophores based on the specific modification of SNAP and CLIP tags.[70]

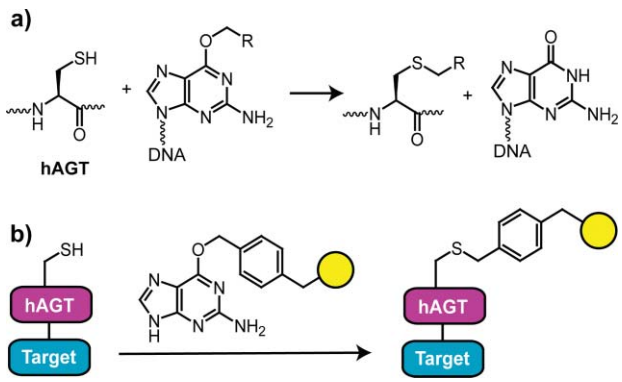

Fig. 6 hAGT catalyzed DNA modification and protein labeling. (a) The native activity of hAGT is to repair alkylated DNA by transferring the alkyl group from O^6-alkylguaninie DNA to an active site Cys residue of the hAGT enzyme. (b) hAGT catalyzed protein labeling by transferring small molecule probes from O^6-benzylguanine (BG) conjugates to the active site Cys residue.

One advantage of the hAGT catalyzed protein labeling method is that the BG and BC conjugated chemical probes are cell membrane permeable, which would allow the labeling of intracellular proteins in live cells. SNAP tags fused to different target proteins have been investigated in different subcellular environments, including β-Gal (cytosolic protein), α-tubulin (cytoskeleton protein), tsVSVG (temperature sensitive protein with a cytoplasmic C-terminus), and G protein-coupled receptor NK_1 with an exoplasmic N-terminus.[71] Calcium and zinc sensors have been attached to SNAP fused to the target proteins.[72,73] BC conjugated small molecule affinity ligands have also been used for the induction of protein dimerization based on SNAP labeling.[74–76] hAGT catalyzed protein labeling has been demonstrated to be an efficient and versatile method for the attachment of small molecule probes of diverse structures and functionalities to the target proteins in the cell. Although the size of hAGT (177 aa, 21 kDa) may be a disadvantage for the construction of fusion proteins, the high labeling efficiency and the membrane permeability of the chemical probes make hAGT a favorable choice for protein labeling and cell imaging studies.

7. Cutinase

Cutinase is a serine esterase of 22 kDa with native activity to hydrolyze cutin, a polyester polymer composed of 16-hydroxy palmitic acid and 18-hydroxy stearic acid (Fig. 7a).[77] Cutinase can also hydrolyze fatty acid esters and triacyl glycerol.[78] Alkyl phosphonate has been found to form a covalent adduct with cutinase by conjugation with a specific Ser residue in the enzyme active site (Fig. 7b).[79] Protein labeling with cutinase has been used to immobilize proteins on a self-assembled monolayer.[80] To demonstrate the application of this method, cutinase was fused to calmodulin (CaM) and the reaction of cutinase-CaM fusion with *p*-nitrophenyl phosphonate ligand presented on an alkanethiolate monolayer resulted in covalent attachment of cutinase-CaM to the alkyl ligand on the monolayer *via* a phosphoester bond.[80] Cutinase has also been used as a tag for the labeling of cell surface proteins with quantum dots (QDs) functionalized with *p*-nitrophenyl phosphonate.[81]

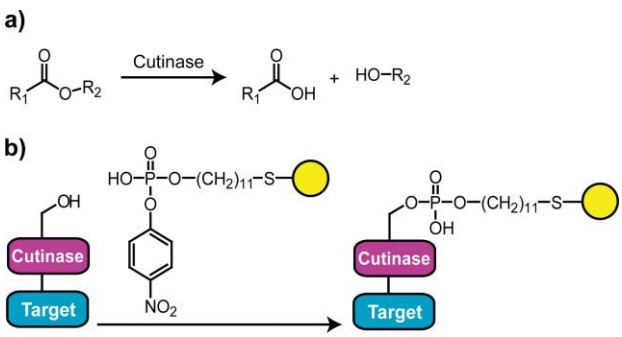

Fig. 7 Cutinase catalyzed reactions. (a) The native activity of cutinase is to hydrolyze the polyester linkages in cutin. (b) In protein labeling catalyzed by cutinase, small molecule probes are transferred from their phosphonate derivatives to a specific Ser residue in the cutinase tag.

8. Haloalkane dehalogenase

The primary function of haloalkane dehalogenase is to remove halides from aliphatic hydrocarbons and afford aliphatic alcohols (Fig. 8a).[82] During enzyme catalyzed dehalogenation, Asp106 in the enzyme active site attacks the halide carbon and nucleophilic substitution of the halide group gives rise to an alkyl ester on the Asp residue. Subsequently, the enzyme alkyl ester intermediate is hydrolyzed by an active site water molecule with the imidazole side chain of His272 acting as a base for water deprotonation.[83] It has been demonstrated that a His272Phe mutant of the dehalogenase enzyme is trapped at the stage of the alkyl ester intermediate.[83] Fusion of His272Phe dehalogenase mutant to the protein of interest would allow the covalent attachment of alkyl halide conjugated chemical probes to the dehalogenase tag (HaloTag) for

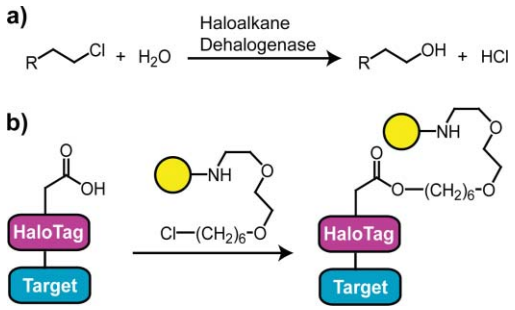

Fig. 8 Haloalkane dehalogenase catalyzed reactions. (a) The native activity of haloalkane dehalogenase is to hydrolyze alkyl halides. (b) In protein labeling reaction, chemical probes linked to an alkyl halide can be covalently conjugated to an active site Asp residue in the HaloTag.

protein labeling (Fig. 8b).[84,85] Chemical probes of diverse structures have been linked to alkyl chloride to serve as the substrates of HaloTag. The HaloTag has also been applied to protein labeling with QDs functionalized with aliphatic chlorides.[86,87] HaloTag is suitable for live cell imaging of proteins inside the mammalian cells due to the membrane permeability of the alkyl halide substrates. Moreover, HaloTag mediated protein labeling is based on the modification of a bacterial enzyme, which diminishes the cross reactivity with the proteins in the mammalian cell. However, the large size of the HaloTag (~33 kDa) may be a burden for the expression of the target protein fusions and may interfere with the native function of the target proteins.

9. Intein mediated protein labeling

Inteins are intercepting protein domains that are expressed in frame with flanking N- and C-terminal polypeptides known as exteins (Fig. 9a).[88] Intein can be self-excised from the precursor protein fusion to afford ligated exteins. Thus, intein can be considered as a single turnover self-modification enzyme for protein splicing.[89] Intein mediated protein labeling is based on the formation of a C-terminal α-thioester linkage between the target protein and intein,[90] and subsequent conjugation of the target protein with Cys functionalized chemical labels in a reaction known as native chemical ligation (Fig. 9b).[91] In this strategy, the target protein is first expressed as a fusion with a C-terminal intein tag. Intein catalyzed self cleavage gives rise to a target protein-intein conjugate linked with an α-thioester bond which undergoes ester exchange with Cys functionalized small molecule probes to release intein and afford target protein conjugated to small molecule probes with a thioester linkage. This intermediate then undergoes a spontaneous rearrangement to give a native peptide bond between the C-terminus of the target protein and Cys conjugated small molecule probes (Fig. 9b).

Intein mediated protein labeling has been used to conjugate biotin and fluorophores to the target protein.[92–94] Protein labeling has also been demonstrated with live bacterial cells expressing intein tagged target proteins.[92] The advantage of intein mediated protein labeling is that the intein tag is self-excised from the target protein upon conjugation with Cys functionalized chemical probes. Thus, the probe is directly attached to the C-terminus of the target protein and the large size of the intein tag would not interfere with the function of the target protein after the labeling reaction. However, intein catalyzed protein splicing is often slow

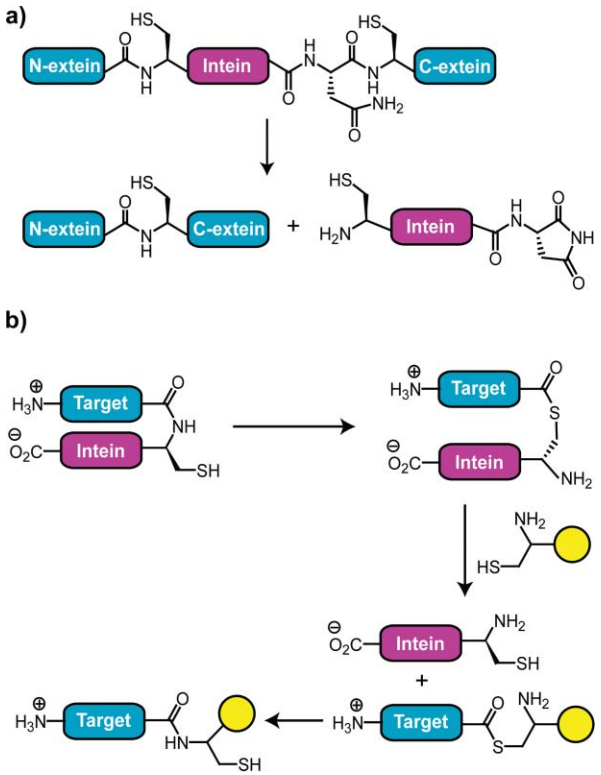

Fig. 9 Intein catalyzed protein modification and labeling with small molecules. (a) Protein splicing by intein. (b) Intein mediated protein labeling with Cys conjugated small molecule probes.

with nonnative exteins and the splicing efficiency may vary with different protein fusions. Consequently, intein catalyzed protein labeling with live cells proceeds slowly compared to other enzyme catalyzed protein labeling reactions and high concentration of Cys functionalized probes is required to increase the labeling yield.[92]

Protein posttranslational modification with swinging arms

Recently, highly efficient and specific methods for protein labeling have been developed based on a family of protein posttranslational modification enzymes that attaches prosthetic "swinging arms" to small protein domains as part of a multi-enzyme complex. These posttranslational modification enzymes include biotin ligase, lipoic acid ligase and phosphopantetheinyl transferase.[95] They load specific protein domains with biotin, lipoyl and phosphopantetheinyl groups as swinging arms that provide the sites of substrate tethering and facilitate the transferring of reaction intermediates between spatially separated enzyme active sites. In this way, multi-step reactions can be streamlined in a designated order within a super assembly of multiple enzyme subunits. By manipulating the catalytic activity of the posttranslational modification enzymes, various chemical probes including fluorophores, quantum dots and photocrosslinkers have been attached to target proteins with a swinging arm linkage for site specific protein labeling.

10. Biotin ligase

The native activity of biotin ligase is to covalently attach biotin to a specific Lys side chain with an amide linkage in biotin carboxyl

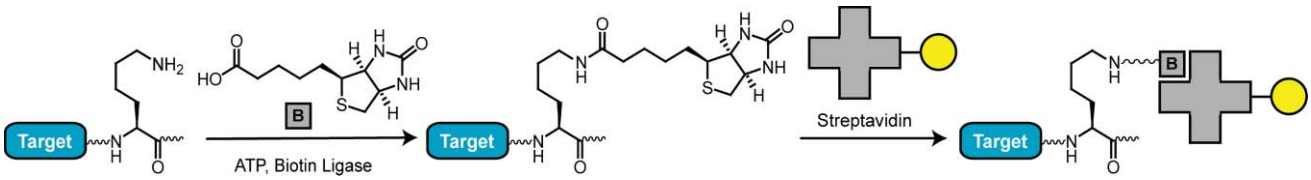

Fig. 10 Biotin ligase catalyzed protein modification with biotin in the native substrate biotin carboxyl carrier protein (BCCP) and in AP peptide tag. Biotinylated proteins can then be labeled with streptavidin conjugated with a variety of chemical probes.

carrier protein (BCCP) in the multi-enzyme complex of acetyl CoA carboxylase (Fig. 10).[96,97] The biotin prosthetic group in BCCP first undergoes N-carboxylation to afford N-carboxybiotin catalyzed by biotin carboxylase. The carboxylated biotin then moves to the active site of carboxyltransferase for the delivery of the carboxyl group to acetyl-CoA to afford malonyl-CoA. Thus, the biotin-lysine conjugate in BCCP functions as a swinging arm for the transfer of activated carboxyl group between the enzyme active sites.

Since biotin has very high affinity for streptavidin,[98] molecular labels of diverse structures and functions can be attached to biotin modified proteins by biotin-streptavidin binding. The development of a short peptide tag (AP tag-15aa) as a surrogate substrate of *E. coli* biotin ligase, BirA, has greatly advanced protein labeling based on biotin modification (Fig. 10).[99] Cell surface receptors has been expressed as a fusion to the AP tag and efficiently labeled with biotin by BirA.[100] Subsequently, the binding of small molecule fluorophore or QD conjugated streptavidin to biotin would allow fluorescence imaging of cell surface proteins.[101]

Native streptavidin functions as a tetramer with four biotin binding sites. The application of multivalent streptavidin to biotinylated cell surface proteins may induce artificial oligomerization of target proteins and interfere with the normal function of the proteins to be imaged.[102] To deal with this problem, streptavidin tetramer has been engineered by mutagenesis so that only one biotin binding site is retained for high affinity binding with biotin.[103] The streptavidin tetramer with monovalent binding stoichiometry to biotin would reduce the level of undesired crosslinking of biotin labeled cell surface proteins upon streptavidin binding. Monovalent streptavidin has also been used for the preparation of QD that is functionalized with only one copy of monovalent streptavidin per QD particle.[104] This would significantly reduce the size of the QD label and prevent artificial crosslinking of biotin labeled proteins induced by the binding of streptavidin conjugated to QD.

Protein labeling by biotin modification would require the binding of streptavidin for the attachment of chemical probes to the labeled protein. The large size of streptavidin tetramer (~60 kDa) may be a burden for the labeled proteins and interfere with their normal biological functions and interactions with other proteins. To overcome this problem, ketone biotin, a biotin isostere with the replacement of ureido nitrogens with methylene groups, was found to be recognized by BirA for AP modification.[100] Ketone biotin attached to AP can then react with hydrazide-probe conjugates for the attachment of synthetic probes to the target proteins. To diversify the peptide sequence that can be modified by biotin ligase, 15 residue peptide yAP has been found as a substrate of yeast biotin ligase by screening a phage displayed peptide library.[105] The yAP peptide is not reactive with *E. coli* biotin ligase BirA and protein fusions of AP and yAP on cell surfaces can be labeled in tandem with biotin and streptavidin conjugated fluorophores by *E. coli* BirA and yeast biotin ligase.

Overall, biotin ligase has been demonstrated to be an attractive method for protein labeling based on the wide availability of streptavidin conjugated chemical probes and high efficiency of enzyme catalyzed AP tag modification. One concern for biotin mediated protein labeling is that biotin modified proteins are naturally present in the cell, and will contribute to the background binding of streptavidin. However, since most biotin labeling experiments are conducted on the cell surface with externally supplied biotin and biotin ligase, the nonspecific binding of streptavidin-probe conjugate to endogenous biotinylated proteins has been minimal.

11. Lipoic acid ligase

Lipoic acid ligase is responsible for ATP dependent modification of a specific lysine residue in the lipoyl domains by lipoic acid conjugation (Fig. 11a).[106,107] The lipoyl domain is involved in the multi-enzyme complexes of 2-oxo acid dehydrogenase for the oxidative decarboxylation of 2-oxo acids such as pyruvate or α-ketoglutarate. The function of the lipoyl group is to uptake activated acyl groups from their conjugates with the thiamine diphosphate (TPP) at the active site of decarboxylase in the enzyme complex. The acyl lipoyl group then swings to the active site of acetyltransferase to deliver the acyl group to CoA to form acyl CoA.[108,109] The lipoyl domain is about 80 residues in length. By truncation and mutagenesis, a 22-residue peptide (LAP) has been identified to be an efficient substrate of *E. coli* lipoic acid ligase, LplA, for lipoyl attachment.[110] Furthermore, LplA has been found to transfer azide conjugated carboxylic acid to the LAP peptide followed by further decoration of the azide

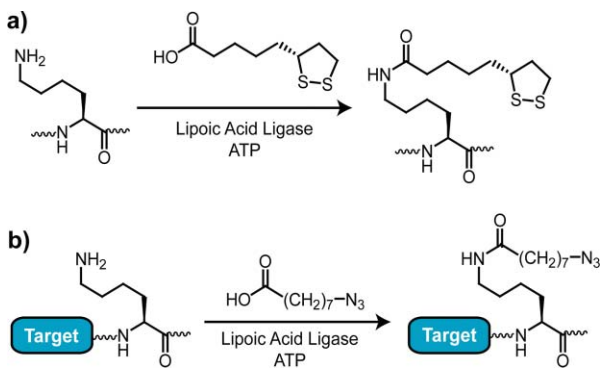

Fig. 11 Lipoic acid ligase catalyzed protein modification. (a) Lipoyl attachment to a specific Lys residue in the lipoyl domain catalyzed by lipoic acid ligase. (b) Protein labeling with lipoic acid analogs.

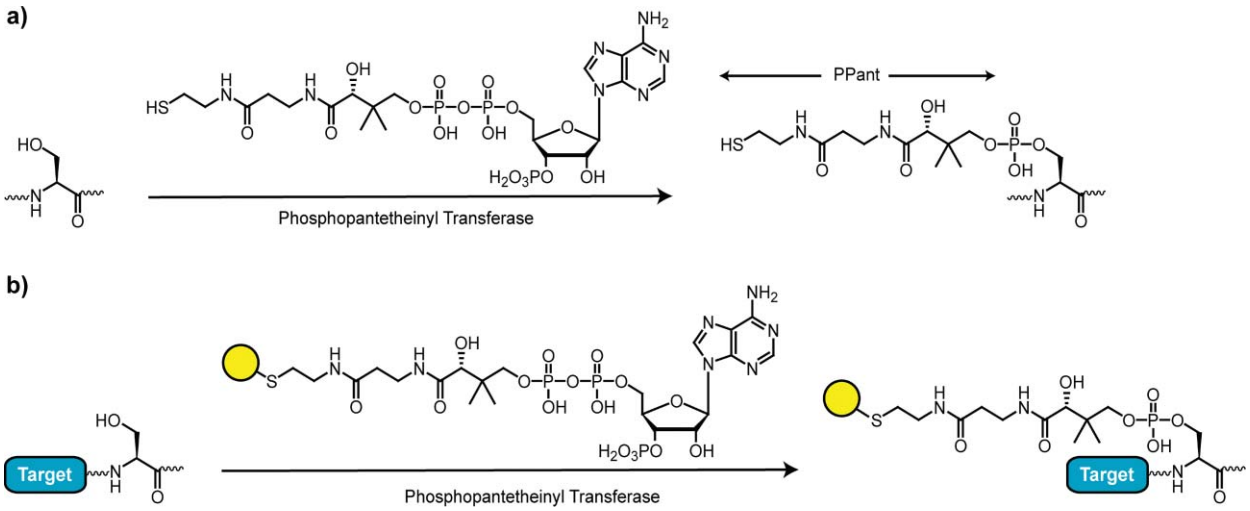

Fig. 12 Phosphopantetheinyl transferase (PPTase) catalyzed protein modification. (a) PPTase catalyzed carrier protein modification with phosphopantetheinyl (Ppant) group derived from CoA. (b) PPTase catalyzed protein labeling with CoA-probe conjugates. The chemical labels are attached to a specific Ser residue in the carrier proteins or short peptide tags with a Ppant linker.

group with various affinity and fluorescence probes by Huisgen azide-alkyne cycloaddition (Fig. 11b).[110] A mutant of LplA has also been found to use a fluorinated aryl azide carboxylate for LAP tag modification in order to label the target protein with a photocrosslinking probe.[111]

12. Phosphopantetheinyl transferases

Phosphopantetheinyl transferases (PPTases) have been recently developed as a useful tool for protein labeling due to their broad substrate specificity with CoA conjugated small molecules.[112–114] The native activity of PPTases is to transfer 4′-phosphopantetheinyl group (Ppant) from CoA to a conserved serine residue of acyl carrier protein (ACP) or peptidyl carrier protein (PCP) domains as a part of fatty acid synthases (FAS), nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) (Fig. 12a).[115–117] NRPS and PKS are multi-modular enzyme complexes devoted to the biosynthesis of secondary metabolites of complex structures. ACP and PCP carrier protein domains are 80–100 residues in length and the Ppant group acts as a swinging arm for anchoring elongating fatty acid, polyketide and nonribosomal peptide chains along the enzymatic assembly lines of FAS, NRPS and PKS.[118]

The crystal structures of *Bacillus subtilis* PPTases Sfp and AcpS show that CoA ligands are bound to the enzyme active sites mainly through the interaction of the 3′-phospho-5′-ADP moiety of CoA with the active site residues whereas the β-mercaptoethylamine group of CoA is exposed outside of the CoA binding pocket.[119,120] This observation provides an explanation of the impressive substrate promiscuity of the Sfp and AcpS enzymes with the chemical functionalities attached to the thiol end of CoA. It has been demonstrated that diverse chemical structures including peptides, fluorophores, carbohydrates and biotin can be conjugated to CoA *via* thioether, thioester or disulfide linkages for the PPTase catalyzed protein labeling of ACP or PCP domains (Fig. 12b).[121–125]

To adopt PPTase catalyzed protein labeling, ACP or PCP domains have been expressed as a fusion to the cell surface receptors such as transferrin receptor, α-glutinin receptor (Aga2p), human G-protein coupled receptor neurokinin-1 (NK_1), *etc*.[126–129] One step reaction with fluorophore-CoA conjugates in the presence of Sfp or AcpS would covalently attach the organic fluorescence probes to the receptor-carrier protein fusion for live cell imaging. Biotin-CoA conjugate can also be used for the labeling of the cell surface proteins followed by the binding of streptavidin-QD conjugate for the attachment of QD to carrier protein tagged receptors.[126] Perhaps a more direct approach for QD labeling is to conjugate CoA to the QD particle and Sfp would readily take QD-CoA as the substrate and covalently attach QD to the target protein *via* a Ppant linker.[130] PPTase catalyzed carrier protein labeling with QD eliminates the use of streptavidin to bridge the linkage between QD and the target protein and provides an easy and efficient one-step protocol for conjugating QD to cell surface receptors. PPTase also catalyzes protein labeling in cell lysates with biotin-CoA conjugates for the immobilization of proteins on streptavidin coated proteins chips.[131] Furthermore CoA has been conjugate to PEGA resin for the subsequent attachment of target proteins to the polymer beads.[132]

Small peptide tags of 11–12 residues in length have also been identified from phage displayed peptide libraries as efficient substrates of *B. subtilis* Sfp and *E. coli* AcpS.[133–135] A1 and S6 peptides have been found to have orthogonal specificities for PPTase modification, with catalytic efficiency (k_{cat}/K_m) of AcpS catalyzed A1 modification 30-fold higher than that of S6 modification and the k_{cat}/K_m of Sfp catalyzed S6 modification 442-fold higher than that of A1 modification.[134] The high substrate specificity of the AcpS/A1 and Sfp/S6 pairs has been used for the tandem labeling of differentially tagged cell surface receptors with different fluorophores.[134] In this way, two target proteins can be imaged simultaneously on the surface of the same cell, which may help to reveal their interactions with each other.

Since CoA is not cell membrane permeable, PPTase catalyzed protein labeling with fluorophore-CoA conjugates has been limited to cell surface protein labeling with exposed peptide or carrier protein tags.[112,114] To adopt PPTase for the labeling of intracellular

proteins, probes are attached to pantothenic acid which can diffuse into the cell for *in vivo* assembly of probe-CoA conjugates by endogenous CoA biosynthetic enzymes. Target proteins fused to the carrier protein tag can then be enzymatically labeled with probe-CoA conjugates formed inside the cell.[136,137]

Sfp mutant R4-4 with substrate specificity for 3′-dephospho CoA (dpCoA) has also been identified by implementing a catalysis based phage selection scheme.[138] R4-4 catalyzes PCP modification with dpCoA at a rate more than 300-fold higher than the wild type Sfp.[138] The mutant also catalyzes carrier protein modification with native CoA at a rate 10-fold higher than that of Sfp. Since R4-4 can recognize dpCoA, structurally simplified CoA analogues can be used for protein labeling catalyzed by PPTase. It has been demonstrated that small molecule probes directly conjugated to ADP can be used as the substrate of R4-4 for site specific protein labeling.[139]

Conclusions

Methods for site specific protein labeling based on a diverse array of protein modification chemistry have been developed in the past few years. At this stage, virtually any chemical entities, ranging from small molecule fluorophores to nano-sized quantum dot crystals can be attached to target proteins for cell imaging applications. Although much progress has been made in establishing the covalent linkage between the target proteins and synthetic probes by bioorthogonal conjugation chemistry or enzyme catalyzed protein modification, challenges still remain on specific labeling of the target proteins with minimal disturbance to their biological functions and cellular environment. For example, almost all protein labeling methods developed so far rely on the construction of fusion proteins and over expression of the fusion proteins in the cell for the subsequent modification of the peptide or protein tags by chemical probes. After the protein labeling reaction, the location, trafficking and interaction of the fusion proteins with other biomolecules can be readily followed by well established biophysical and biochemical methods, such as fluorescence imaging or affinity binding. However there has been little calibration on to what extent the fusion protein recapitulates the functional role of the native protein and how much disturbance of fusion protein expression to the biological fitness of the cell. Thus, despite the development of new enzymes and new chemical reactions for protein labeling, new methods with minimal disturbance to the biological system are still in high demand.

Another challenge for a protein labeling method that is applicable to biological systems is the efficiency and specificity of the labeling reaction. Many protein molecules especially the ones involved in signal transduction pathways are only present at low concentration and often experience dynamic changes regulated by the cell signaling pathways. Over expression of those proteins as fusions to a peptide tag of large size would almost certainly distort the biological role of the target protein and their network of interaction with other partners. Thus, a highly efficient labeling method is needed for the specific chemical modification of target proteins present in low copy numbers in the cell. Posttranslational modification enzymes would have a unique opportunity in this area because their native role in the cell is to modify a small number of target proteins with high catalytic efficiency and fidelity in response to a signal transduction event. A foreseeable challenge in using posttranslational modification enzymes for protein labeling is that the native substrates would compete with the probe functionalized substrate for protein modification. In addition, the probe-substrate conjugate may not be cell membrane permeable, thus they cannot be used for protein labeling inside the cell. In some other cases, protein modification enzymes would have low activity with probe functionalized substrates, preventing the labeling of the target proteins at high efficiency. We envision that protein engineering on posttranslational modification enzymes would generate new enzymatic activities that are tailor-made for protein labeling in a native cellular environment.

Posttranslationally manipulating the function and activity of target proteins by protein labeling/modification would not only reveal the biological activity of the target proteins, but also lead to possibilities for precise control of their biological functions. The second goal is even more challenging but would be approachable with the creative combination of chemical and enzymatic tools that are currently being developed for site specific protein labeling.

Acknowledgements

This work was supported by a lab startup grant from the University of Chicago, the New Faculty Award from the Camille and Henry Dreyfus Foundation, and the NSF MRSEC program (DMR-0213745; J.Y.)

Notes and references

1 M. Fernandez-Suarez and A. Y. Ting, *Nat. Rev. Mol. Cell Biol.*, 2008, **9**, 929–943.
2 R. Y. Tsien, *Nat. Rev. Mol. Cell Biol.*, 2003, **Suppl**, SS16–21.
3 R. Y. Tsien, *Annu. Rev. Biochem.*, 1998, **67**, 509–544.
4 B. N. Giepmans, S. R. Adams, M. H. Ellisman and R. Y. Tsien, *Science*, 2006, **312**, 217–224.
5 R. Y. Tsien, *Keio J. Med.*, 2006, **55**, 127–140.
6 R. P. Haugland, *The Handbook—A Guide to Fluorescent Probes and Labeling Technologie*, Invitrogen, Eugene, OR, USA, 2005.
7 N. Johnsson and K. Johnsson, *ACS Chem. Biol.*, 2007, **2**, 31–38.
8 T. L. Foley and M. D. Burkart, *Curr. Opin. Chem. Biol.*, 2007, **11**, 12–19.
9 M. E. Hahn and T. W. Muir, *Trends Biochem. Sci.*, 2005, **30**, 26–34.
10 J. A. Prescher and C. R. Bertozzi, *Nat. Chem. Biol.*, 2005, **1**, 13–21.
11 B. A. Griffin, S. R. Adams, J. Jones and R. Y. Tsien, *Methods Enzymol.*, 2000, **327**, 565–578.
12 E. Saxon and C. R. Bertozzi, *Science*, 2000, **287**, 2007–2010.
13 H. C. Kolb and K. B. Sharpless, *Drug Discovery Today*, 2003, **8**, 1128–1137.
14 N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo and C. R. Bertozzi, *ACS Chem. Biol.*, 2006, **1**, 644–648.
15 S. T. Laughlin, J. M. Baskin, S. L. Amacher and C. R. Bertozzi, *Science*, 2008, **320**, 664–667.
16 S. T. Laughlin and C. R. Bertozzi, *Nat. Protoc.*, 2007, **2**, 2930–2944.
17 J. A. Prescher, D. H. Dube and C. R. Bertozzi, *Nature*, 2004, **430**, 873–877.
18 R. A. Chandra, E. S. Douglas, R. A. Mathies, C. R. Bertozzi and M. B. Francis, *Angew. Chem., Int. Ed.*, 2006, **45**, 896–901.
19 Z. J. Gartner and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 4606–4610.
20 L. W. Miller, Y. Cai, M. P. Sheetz and V. W. Cornish, *Nat. Methods*, 2005, **2**, 255–257.
21 L. Wang and P. G. Schultz, *Angew. Chem., Int. Ed.*, 2005, **44**, 34–66.
22 J. Xie and P. G. Schultz, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 775–782.
23 C. T. Walsh, S. Garneau-Tsodikova and G. J. Gatto, Jr., *Angew. Chem., Int. Ed.*, 2005, **44**, 7342–7372.
24 N. Khidekel and L. C. Hsieh-Wilson, *Org. Biomol. Chem.*, 2004, **2**, 1–7.
25 J. Reinders and A. Sickmann, *Biomol. Eng.*, 2007, **24**, 169–177.

26 M. Griffin, R. Casadio and C. M. Bergamini, *Biochem. J.*, 2002, **368**, 377–396.
27 H. Sato, M. Ikeda, K. Suzuki and K. Hirayama, *Biochemistry*, 1996, **35**, 13072–13080.
28 M. Taki, M. Shiota and K. Taira, *Protein Eng., Des. Sel.*, 2004, **17**, 119–126.
29 N. Shimba, N. Yamada, K. Yokoyama and E. Suzuki, *Anal. Biochem.*, 2002, **301**, 123–127.
30 M. Jager, E. Nir and S. Weiss, *Protein Sci.*, 2006, **15**, 640–646.
31 A. Josten, L. Haalck, F. Spener and M. Meusel, *J. Immunol. Methods*, 2000, **240**, 47–54.
32 C. W. Lin and A. Y. Ting, *J. Am. Chem. Soc.*, 2006, **128**, 4542–4543.
33 Y. Sugimura, M. Hosono, F. Wada, T. Yoshimura, M. Maki and K. Hitomi, *J. Biol. Chem.*, 2006, **281**, 17699–17706.
34 J. Tominaga, Y. Kemori, Y. Tanaka, T. Maruyama, N. Kamiya and M. Goto, *Chem. Commun.*, 2007, 401–403.
35 A. M. Perry, H. Ton-That, S. K. Mazmanian and O. Schneewind, *J. Biol. Chem.*, 2002, **277**, 16241–16248.
36 S. K. Mazmanian, G. Liu, H. Ton-That and O. Schneewind, *Science*, 1999, **285**, 760–763.
37 H. Mao, S. A. Hart, A. Schink and B. A. Pollok, *J. Am. Chem. Soc.*, 2004, **126**, 2670–2671.
38 R. Parthasarathy, S. Subramanian and E. T. Boder, *Bioconjugate Chem.*, 2007, **18**, 469–476.
39 T. Tanaka, T. Yamamoto, S. Tsukiji and T. Nagamune, *ChemBioChem*, 2008, **9**, 802–807.
40 M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner and H. L. Ploegh, *Nat. Chem. Biol.*, 2007, **3**, 707–708.
41 T. Yamamoto and T. Nagamune, *Chem. Commun.*, 2009, 1022–1024.
42 P. J. Casey and M. C. Seabra, *J. Biol. Chem.*, 1996, **271**, 5289–5292.
43 R. Spence and P. J. Casey, *Mechanism of Catalysis by Protein Farnesyltransferase*, Academic, San Diego, 2000.
44 M. D. Resh, *Nat. Chem. Biol.*, 2006, **2**, 584–590.
45 B. P. Duckworth, Z. Zhang, A. Hosokawa and M. D. Distefano, *ChemBioChem*, 2007, **8**, 98–105.
46 B. E. Dursina, R. Reents, A. Niculae, A. Veligodsky, R. Breitling, K. Pyatkov, H. Waldmann, R. S. Goody and K. Alexandrov, *Protein Expression Purif.*, 2005, **39**, 71–81.
47 U. T. Nguyen, J. Cramer, J. Gomis, R. Reents, M. Gutierrez-Rodriguez, R. S. Goody, K. Alexandrov and H. Waldmann, *ChemBioChem*, 2007, **8**, 408–423.
48 Y. Kho, S. C. Kim, C. Jiang, D. Barma, S. W. Kwon, J. Cheng, J. Jaunbergs, C. Weinbaum, F. Tamanoi, J. Falck and Y. Zhao, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 12479–12484.
49 C. Gauchet, G. R. Labadie and C. D. Poulter, *J. Am. Chem. Soc.*, 2006, **128**, 9274–9275.
50 B. P. Duckworth, Y. Chen, J. W. Wollack, Y. Sham, J. D. Mueller, T. A. Taton and M. D. Distefano, *Angew. Chem., Int. Ed.*, 2007, **46**, 8819–8822.
51 U. T. Nguyen, Z. Guo, C. Delon, Y. Wu, C. Deraeve, B. Franzel, R. S. Bon, W. Blankenfeldt, R. S. Goody, H. Waldmann, D. Wolters and K. Alexandrov, *Nat. Chem. Biol.*, 2009, **5**, 227–235.
52 M. H. Gelb, *Science*, 1997, **275**, 1750–1751.
53 R. S. Bhatnagar, K. Futterer, G. Waksman and J. I. Gordon, *Biochim. Biophys. Acta*, 1999, **1441**, 162–172.
54 T. A. Farazi, G. Waksman and J. I. Gordon, *J. Biol. Chem.*, 2001, **276**, 39501–39504.
55 H. C. Hang, E. J. Geutjes, G. Grotenbreg, A. M. Pollington, M. J. Bijlmakers and H. L. Ploegh, *J. Am. Chem. Soc.*, 2007, **129**, 2744–2745.
56 W. P. Heal, S. R. Wickramasinghe, P. W. Bowyer, A. A. Holder, D. F. Smith, R. J. Leatherbarrow and E. W. Tate, *Chem. Commun.*, 2008, 480–482.
57 W. P. Heal, S. R. Wickramasinghe, R. J. Leatherbarrow and E. W. Tate, *Org. Biomol. Chem.*, 2008, **6**, 2308–2315.
58 G. Charron, M. M. Zhang, J. S. Yount, J. Wilson, A. S. Raghavan, E. Shamir and H. C. Hang, *J. Am. Chem. Soc.*, 2009, **131**, 4967–4975.
59 T. Dierks, A. Dickmanns, A. Preusser-Kunze, B. Schmidt, M. Mariappan, K. von Figura, R. Ficner and M. G. Rudolph, *Cell*, 2005, **121**, 541–552.
60 T. Dierks, B. Schmidt, L. V. Borissenko, J. Peng, A. Preusser, M. Mariappan and K. von Figura, *Cell*, 2003, **113**, 435–444.
61 G. Lukatela, N. Krauss, K. Theis, T. Selmer, V. Gieselmann, K. von Figura and W. Saenger, *Biochemistry*, 1998, **37**, 3654–3664.
62 I. S. Carrico, B. L. Carlson and C. R. Bertozzi, *Nat. Chem. Biol.*, 2007, **3**, 321–322.
63 J. S. Rush and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2008, **130**, 12240–12241.
64 P. Wu, W. Shui, B. L. Carlson, N. Hu, D. Rabuka, J. Lee and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 3000–3005.
65 S. Mizukami, S. Watanabe, Y. Hori and K. Kikuchi, *J. Am. Chem. Soc.*, 2009, **131**, 5016–5017.
66 A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel and K. Johnsson, *Nat. Biotechnol.*, 2003, **21**, 86–89.
67 T. Gronemeyer, C. Chidley, A. Juillerat, C. Heinis and K. Johnsson, *Protein Eng. Des. Sel.*, 2006, **19**, 309–316.
68 A. Juillerat, C. Heinis, I. Sielaff, J. Barnikow, H. Jaccard, B. Kunz, A. Terskikh and K. Johnsson, *ChemBioChem*, 2005, **6**, 1263–1269.
69 A. Juillerat, T. Gronemeyer, A. Keppler, S. Gendreizig, H. Pick, H. Vogel and K. Johnsson, *Chem. Biol.*, 2003, **10**, 313–317.
70 A. Gautier, A. Juillerat, C. Heinis, I. R. Correa, Jr., M. Kindermann, F. Beaufils and K. Johnsson, *Chem. Biol.*, 2008, **15**, 128–136.
71 A. Keppler, H. Pick, C. Arrivoli, H. Vogel and K. Johnsson, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 9955–9959.
72 M. Bannwarth, I. R. Correa, M. Sztretye, S. Pouvreau, C. Fellay, A. Aebischer, L. Royer, E. Rois and K. Johnsson, *ACS Chem. Biol.*, 2009, **4**, 179–190.
73 E. Tomat, E. M. Nolan, J. Jaworski and S. J. Lippard, *J. Am. Chem. Soc.*, 2008, **130**, 15776–15777.
74 C. Chidley, K. Mosiewicz and K. Johnsson, *Bioconjugate Chem.*, 2008, **19**, 1753–1756.
75 S. Gendreizig, M. Kindermann and K. Johnsson, *J. Am. Chem. Soc.*, 2003, **125**, 14970–14971.
76 G. Lemercier, S. Gendreizig, M. Kindermann and K. Johnsson, *Angew. Chem., Int. Ed.*, 2007, **46**, 4281–4284.
77 C. Martinez, P. De Geus, M. Lauwereys, G. Matthyssens and C. Cambillau, *Nature*, 1992, **356**, 615–618.
78 M. L. Mannesse, R. C. Cox, B. C. Koops, H. M. Verheij, G. H. de Haas, M. R. Egmond, H. T. van der Hijden and J. de Vlieg, *Biochemistry*, 1995, **34**, 6400–6407.
79 M. L. Mannesse, J. W. Boots, R. Dijkman, A. J. Slotboom, H. T. van der Hijden, M. R. Egmond, H. M. Verheij and G. H. de Haas, *Biochim. Biophys. Acta*, 1995, **1259**, 56–64.
80 C. D. Hodneland, Y. S. Lee, D. H. Min and M. Mrksich, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 5048–5052.
81 R. Bonasio, C. V. Carman, E. Kim, P. T. Sage, K. R. Love, T. R. Mempel, T. A. Springer and U. H. von Andrian, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 14753–14758.
82 K. H. Verschueren, F. Seljee, H. J. Rozeboom, K. H. Kalk and B. W. Dijkstra, *Nature*, 1993, **363**, 693–698.
83 F. Pries, J. Kingma, G. H. Krooshof, C. M. Jeronimus-Stratingh, A. P. Bruins and D. B. Janssen, *J. Biol. Chem.*, 1995, **270**, 10405–10411.
84 G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit and K. V. Wood, *ACS Chem. Biol.*, 2008, **3**, 373–382.
85 G. V. Los and K. Wood, *Methods Mol. Biol.*, 2007, **356**, 195–208.
86 H. Yao, Y. Zhang, F. Xiao, Z. Xia and J. Rao, *Angew. Chem., Int. Ed.*, 2007, **46**, 4346–4349.
87 Y. Zhang, M. K. So, A. M. Loening, H. Yao, S. S. Gambhir and J. Rao, *Angew. Chem., Int. Ed.*, 2006, **45**, 4936–4940.
88 M. Q. Xu, M. W. Southworth, F. B. Mersha, L. J. Hornstra and F. B. Perler, *Cell*, 1993, **75**, 1371–1377.
89 C. J. Noren, J. Wang and F. B. Perler, *Angew. Chem., Int. Ed.*, 2000, **39**, 450–466.
90 T. W. Muir, *Annu. Rev. Biochem.*, 2003, **72**, 249–289.
91 P. E. Dawson and S. B. Kent, *Annu. Rev. Biochem.*, 2000, **69**, 923–960.
92 S. Chattopadhaya, L. P. Tan and S. Q. Yao, *Nat. Protoc.*, 2006, **1**, 2386–2398.
93 M. L. Lesaicherre, R. Y. Lue, G. Y. Chen, Q. Zhu and S. Q. Yao, *J. Am. Chem. Soc.*, 2002, **124**, 8768–8769.
94 R. Y. Lue, G. Y. Chen, Y. Hu, Q. Zhu and S. Q. Yao, *J. Am. Chem. Soc.*, 2004, **126**, 1055–1062.
95 R. N. Perham, *Annu. Rev. Biochem.*, 2000, **69**, 961–1004.
96 P. Reche and R. N. Perham, *EMBO J.*, 1999, **18**, 2673–2682.
97 A. Chapman-Smith and J. E. Cronan, Jr., *Trends Biochem. Sci.*, 1999, **24**, 359–363.
98 M. Wilchek and E. A. Bayer, *Methods Enzymol.*, 1990, **184**, 5–13.

99 D. Beckett, E. Kovaleva and P. J. Schatz, *Protein Sci.*, 1999, **8**, 921–929.
100 I. Chen, M. Howarth, W. Lin and A. Y. Ting, *Nat. Methods*, 2005, **2**, 99–104.
101 M. Howarth, K. Takao, Y. Hayashi and A. Y. Ting, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 7583–7588.
102 N. M. Green, *Methods Enzymol.*, 1990, **184**, 51–67.
103 M. Howarth, D. J. Chinnapen, K. Gerrow, P. C. Dorrestein, M. R. Grandy, N. L. Kelleher, A. El-Husseini and A. Y. Ting, *Nat. Methods*, 2006, **3**, 267–273.
104 M. Howarth, W. Liu, S. Puthenveetil, Y. Zheng, L. F. Marshall, M. M. Schmidt, K. D. Wittrup, M. G. Bawendi and A. Y. Ting, *Nat. Methods*, 2008, **5**, 397–399.
105 I. Chen, Y. A. Choi and A. Y. Ting, *J. Am. Chem. Soc.*, 2007, **129**, 6619–6625.
106 X. Zhao, J. R. Miller, Y. Jiang, M. A. Marletta and J. E. Cronan, *Chem. Biol.*, 2003, **10**, 1293–1302.
107 T. W. Morris, K. E. Reed and J. E. Cronan, Jr., *J. Bacteriol.*, 1995, **177**, 1–10.
108 X. Y. Pei, C. M. Titman, R. A. Frank, F. J. Leeper and B. F. Luisi, *Structure*, 2008, **16**, 1860–1872.
109 R. A. Frank, J. V. Pratap, X. Y. Pei, R. N. Perham and B. F. Luisi, *Structure*, 2005, **13**, 1119–1130.
110 M. Fernandez-Suarez, H. Baruah, L. Martinez-Hernandez, K. T. Xie, J. M. Baskin, C. R. Bertozzi and A. Y. Ting, *Nat. Biotechnol.*, 2007, **25**, 1483–1487.
111 H. Baruah, S. Puthenveetil, Y. A. Choi, S. Shah and A. Y. Ting, *Angew. Chem., Int. Ed.*, 2008, **47**, 7018–7021.
112 N. Johnsson, N. George and K. Johnsson, *ChemBioChem*, 2005, **6**, 47–52.
113 M. Sunbul, K. Zhang and J. Yin, *Methods Enzymol.*, 2009, **458**, 255–275.
114 J. Yin, A. J. Lin, D. E. Golan and C. T. Walsh, *Nat. Protoc.*, 2006, **1**, 280–285.
115 J. N. Copp and B. A. Neilan, *Appl. Environ. Microbiol.*, 2006, **72**, 2298–2305.
116 R. S. Flugel, Y. Hwangbo, R. H. Lambalot, J. E. Cronan, Jr. and C. T. Walsh, *J. Biol. Chem.*, 2000, **275**, 959–968.
117 R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, *Chem. Biol.*, 1996, **3**, 923–936.
118 A. C. Mercer and M. D. Burkart, *Nat. Prod. Rep.*, 2007, **24**, 750–773.
119 K. D. Parris, L. Lin, A. Tam, R. Mathew, J. Hixon, M. Stahl, C. C. Fritz, J. Seehra and W. S. Somers, *Structure*, 2000, **8**, 883–895.
120 K. Reuter, M. R. Mofid, M. A. Marahiel and R. Ficner, *EMBO J.*, 1999, **18**, 6823–6831.
121 P. J. Belshaw, C. T. Walsh and T. Stachelhaus, *Science*, 1999, **284**, 486–489.
122 A. M. Gehring, R. H. Lambalot, K. W. Vogel, D. G. Drueckhammer and C. T. Walsh, *Chem. Biol.*, 1997, **4**, 17–24.
123 J. J. La Clair, T. L. Foley, T. R. Schegg, C. M. Regan and M. D. Burkart, *Chem. Biol.*, 2004, **11**, 195–201.
124 S. A. Sieber, C. T. Walsh and M. A. Marahiel, *J. Am. Chem. Soc.*, 2003, **125**, 10862–10866.
125 F. Vitali, K. Zerbe and J. A. Robinson, *Chem. Commun.*, 2003, 2718–2719.
126 N. George, H. Pick, H. Vogel, N. Johnsson and K. Johnsson, *J. Am. Chem. Soc.*, 2004, **126**, 8896–8897.
127 B. H. Meyer, J. M. Segura, K. L. Martinez, R. Hovius, N. George, K. Johnsson and H. Vogel, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 2138–2143.
128 L. Vivero-Pol, N. George, H. Krumm, K. Johnsson and N. Johnsson, *J. Am. Chem. Soc.*, 2005, **127**, 12770–12771.
129 J. Yin, A. J. Lin, P. D. Buckett, M. Wessling-Resnick, D. E. Golan and C. T. Walsh, *Chem. Biol.*, 2005, **12**, 999–1006.
130 M. Sunbul, M. Yen, Y. Zou and J. Yin, *Chem. Commun.*, 2008, 5927–5929.
131 J. Yin, F. Liu, X. Li and C. T. Walsh, *J. Am. Chem. Soc.*, 2004, **126**, 7754–7755.
132 L. S. Wong, J. Thirlway and J. Micklefield, *J. Am. Chem. Soc.*, 2008, **130**, 12456–12464.
133 J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15815–15820.
134 Z. Zhou, P. Cironi, A. J. Lin, Y. Xu, S. Hrvatin, D. E. Golan, P. A. Silver, C. T. Walsh and J. Yin, *ACS Chem. Biol.*, 2007, **2**, 337–346.
135 Z. Zhou, A. Koglin, Y. Wang, A. P. McMahon and C. T. Walsh, *J. Am. Chem. Soc.*, 2008, **130**, 9925–9930.
136 K. M. Clarke, A. C. Mercer, J. J. La Clair and M. D. Burkart, *J. Am. Chem. Soc.*, 2005, **127**, 11234–11235.
137 J. L. Meier, A. C. Mercer, H. Rivera, Jr. and M. D. Burkart, *J. Am. Chem. Soc.*, 2006, **128**, 12174–12184.
138 M. Sunbul, N. J. Marshall, Y. Zou, K. Zhang and J. Yin, *J. Mol. Biol.*, 2009, **387**, 883–898.
139 Y. Zou and J. Yin, *J. Am. Chem. Soc.*, 2009, **131**, 7548–7549.