APPLICATIONS OF ENZYMATIC PROTEIN LABELING BY PROTEIN FARNESYLTRANSFERASE TOWARDS DUAL-PROTEIN MODIFICATION AND CONSTRUCTION OF TARGETED DELIVERY SYSTEMS

A THESIS SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA

BY

YI ZHANG

IN PARTIAL FULFILLMENT OF THEREQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

ADVISOR: DR. MARK D. DISTEFANO

OCTOBER 2018

Acknowledgements

It has been a long journey since I started graduate school in 2012. There have been accomplishments and tears, and I am grateful that I am not alone in the journey. In particular, I would like to thank my advisor, Dr. Mark Distefano. Mark is so knowledgeable and supportive. I sincerely appreciate his advice not only for work but also for life. I am grateful to the all the Distefano lab members for their support and valuable discussions as well as the fun times when we had group activities together. Especially, I would like to thank Dr. Jonathon Dozier, who was a great mentor when I joined the lab.

Being far away from home, I am thankful to my parents for their unconditioned love and support. I am also grateful to my friends, Dr. Tian Qiu, Sifei Fang, Xueyang Pan and Chao Li for the company and support. Last but not the lease, I would like to mention my cat, Chaplin, for bearing with me when I woke him up from sleep and cuddled him without his consent.

Abstract

Over the years, protein-based conjugates have been extensively utilized in a variety of areas, including the study of fundamental biology as well as biotechnological and clinical applications. Compared to the non-selective conjugation methods which result in a mixture of modified products, site-specific modification by enzymatic methods provides homogeneously labeled proteins with identical physical properties and functional activities. In the first chapter of this dissertation, I reviewed the recent development (over the last five years) of enzymatic labeling techniques for a panel of ten different enzymes, including sortase A, subtiligase, microbial transglutaminase, protein farnesyltransferase (PFTase), N-myristoyltransferase, phosphopantetheinyl transferase, tubulin tyrosine ligase, lipoic acid ligase, biotin ligase and formylglycine generating enzyme. It is shown that each enzymatic method possesses its advantages and limitations and there is no single solution for all the applications. Therefore, both the protein substrates and the specific applications need to be considered to choose an optimal modification method. Among the ten enzymes being discussed above, PFTase is highly selective with fast kinetics. The enzyme transfers an isoprenoid moiety from farnesyl diphosphate to a C-terminal cysteine of the protein substrate. Consisted of only four amino acid residues, the enzyme recognition sequence is called CaaX-box, where C is the cysteine being modified. To label a protein of interest, a number of isoprenoid analogues bearing bioorthogonal functional groups have been developed, including azides, alkynes, and aldehydes. In this dissertation, I employed two different prenyltransferases to achieve dual-protein labeling and explored the applicability of using PFTase to modify several alternative binding proteins for the construction of targeted delivery systems, including the Designed Ankyrin Repeat Proteins (DARPins), the 10th extracellular domain of human fibronectin III (Fn3) and the nanobodies (VHH). It was found that all three binding proteins could be modified by PFTase successfully with high efficiency. Notably, the labeled DARPins retained the binding ability and specificity. To avoid complicated chromatographic purification process and streamline the production, a capture and release strategy was employed to create DARPin-fluorophore and DARPin-cytotoxin conjugates. Overall, this dissertation explored the utility of enzymatic labeling by PFTase towards various applications.

Table of Contents

List of Figures

List of Tables

Chapter 1. Review of Recent Progress in Enzymatic Protein Labeling Techniques and Their Applications

Y. Zhang, K. Park, K. F. Suazo and M. D. Distefano, *Chem. Soc. Rev*., 2018, Advance Article, DOI: 10.1039/C8CS00537K - Adapted by permission of The Royal Society of Chemistry

Protein-based conjugates are valuable constructs for a variety of applications. Conjugation of proteins to fluorophores is commonly used to study their cellular localization and the protein-protein interactions. Modification of therapeutic proteins with either polymers or cytotoxic moieties greatly enhances their pharmacokinetics or potency. To label a protein of interest, conventional direct chemical reaction with the side-chains of native amino acids typically yields heterogeneously modified products. This renders their characterization rather complicated, requires difficult separation steps and may impact protein function. Although modification can also be achieved via the insertion of unnatural amino acids bearing bioorthogonal functional groups, these methods can have lower protein expression yields, limiting large scale production. As a site-specific modification method, enzymatic protein labeling is highly efficient and robust under mild reaction conditions. Significant progress has been made over the last five years in modifying proteins using enzymatic methods for numerous applications, including the creation of clinically relevant conjugates with polymers, cytotoxins or imaging agents, fluorescent or affinity probes to study complex protein interaction networks or protein-linked materials for biosensing. This review summarizes developments in enzymatic protein labeling over the last five years for a panel of ten enzymes, including sortase A, subtiligase, microbial transglutaminase, farnesyltransferase, N-myristoyltransferase, phosphopantetheinyl transferases, tubulin tyrosin ligase, lipoic acid ligase, biotin ligase and formylglycine generating enzyme.

1.1 Introduction

The ability to manipulate proteins, from the construction of protein-based conjugates to surface immobilization, is a key contributor to recent rapid advances in the study of fundamental biology as well as in areas of the biotechnology and pharmaceutics. Labeling of proteins with fluorophores aids in identifying their cellular localization.[1] Compared to the green fluorescent protein (GFP), the small size of these organic molecules makes them less likely to perturb the native structure and function of the protein being studied.[2] Synthesis/semisynthesis of proteins with unambiguous post-translational modification states also helps to elucidate the roles of the individual modification, facilitating the study of protein-protein interactions.[3] In the fields of biotechnology, the immobilization of proteins on solid surfaces has been extensively employed for the development of bio-sensors,[4] recyclable catalysts,[5] and protein microarrays.[6] Biomaterials, including protein functionalized nanoparticles, hydrogels and liposomes have also been explored for various applications, including drug delivery systems[7] and bio-responsive materials.[8] As a major component of modern medicine, the use of protein-based therapeutics continues to grow.[9] To achieve improved therapeutic outcomes, enormous efforts have been made in protein engineering. For example, conjugation of polyethylene glycol (PEG) polymer chains to small therapeutic polypeptides significantly improves their pharmacokinetic properties.[10] To augment the efficacy of cancer treatments using antibodies alone, the incorporation of cytotoxic components, such as small molecule drugs[11] and radioactive isotopes[12] make antibody-based therapeutics more potent. Having an imaging agent in place of a toxic drug creates sensitive and specific targeted imaging probes for cancer diagnosis.[13]

Although protein conjugation can be achieved by installing cargos to proteins non-selectively by reacting with functional groups from the side chains of native amino acids, site-specific labeling provides homogeneously modified products. Precise control over the position of modification eliminates potential detrimental effects on the stability and/or function of the protein. For example, almost all PEGylated proteins are less active than

their unmodified versions, in part due to blockage of the functional site by the bulky PEG polymer.[10] In this case, site-specific incorporation of the PEG polymer offers a tremendous advantage over non-selective methods since an optimal position can be chosen to minimize disruption and recover more activity. Compared to early generations of antibody-drug conjugates (ADCs) with variable drug-to-antibody (DAR) ratios and position of attachment, site-specific ADCs are expected to possess better therapeutic index, promoting the development of more efficacious therapeutic agents.[14] To achieve site-specific protein modification, several strategies have been developed, including enzymatic labeling methods. In addition to being highly selective, enzymatic reactions usually take place under mild conditions with fast kinetics and high yield. Accordingly, significant progress has been made in applying enzymatic labeling methods for protein modification in the last decade. Numerous reviews on this topic have been published[15-17] including one by Rashidian et al. in 2013 that summarized several enzymatic labeling strategies and their early applications.[18] In this review, we focus on the developments that have occurred in the last five years (2013-2018). First, protein modification methods based on reactions occurring on either the native functional groups presented by proteins or those inserted via nonsense suppression methods are described in Section 2. The bioorthogonal reactions that are widely used for protein bioconjugation are also included there. In section 3, a panel of 10 enzymes that have been widely explored will be discussed in detail with a focus on applications reported since 2013. These enzymes are organized into four categories, 1) peptidases: sortase A and subtiligase; 2) transferases: microbial transglutaminase, farnesyltransferase, N-myristoyltransferase and phosphopantetheinyl transferase; 3) ligases: tubulin tyrosine ligase, lipoic acid ligase and biotin ligase; 4) oxidoreductases: formylglycine generating enzyme. Finally, some concluding remarks and efforts to compare these different methods are presented. As this area continues to grow, we hope this review article will stimulate additional innovation and developments as well as provide some insights to facilitate the selection of the optimal enzyme to meet specific needs.

1.2 Chemistry on proteins

To label a protein of interest, both chemical and genetic methods have been developed to introduce modifications through either natural amino acid residues or synthetic analogues thereof.[19,20] This has been achieved by exploiting the inherent reactivity of the functional groups present using direct chemical modification, insertion of an unnatural amino acid via nonsense suppression, or utilization of enzymatic labeling strategies. Due to the limited stability of biomolecules, the chemical reactions employed for protein modification must be biocompatible and exhibit rapid kinetics under mild conditions.

1.2.1 Direct chemical modification.

Over the years, a wide variety of chemical reactions have been explored to specifically functionalize the side chains of certain amino acids through direct chemical modification (see Figure 1.1 for some commonly used reactions). In order to be compatible with sensitive proteins, these reactions generally require efficiency at ambient temperatures and aqueous conditions. Cysteine and lysine are the common targets for functionalization. Aromatic residues including tyrosine and tryptophan can also be modified successfully to create protein conjugates with a fluorophore, biotin or cytotoxic drugs.[21] More recently, novel photocatalysis and redox-based strategies for the modification of tryptophan[22] and methionine,[23] respectively, have been reported. Although generally residue-specific, recent efforts have been focused on developing site-specific strategies to target a specific residue in the proteins. For detailed information of these developments, the reader is referred to other comprehensive reviews of this topic.[20,24]

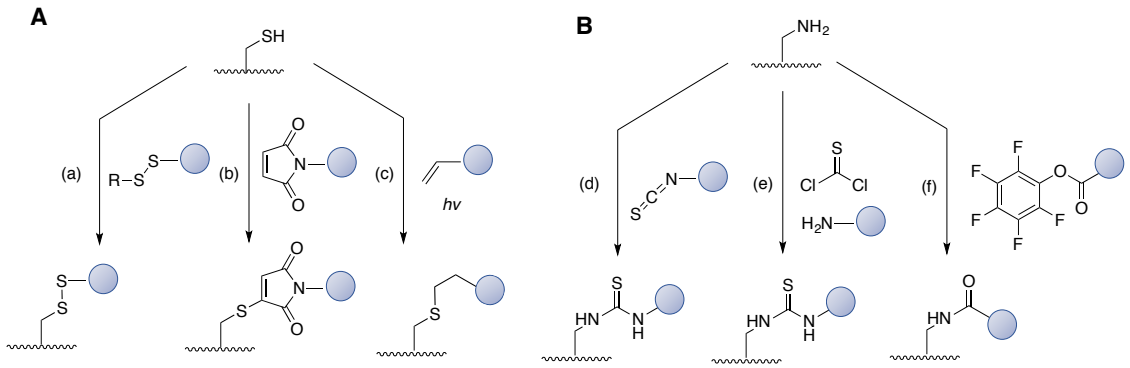


Figure 1.1 Selected direct chemical modifications on amino acids. (A) Cysteine modification via a) disulfide exchange, b) maleimide, or c) photo-catalyzed thiol-ene couplings. (B) Lysine modification via coupling with d) isothiocyanate or activation with e) sulfonyl chloride or f) fluorine-substituted aromatic esters.

Among the 20 canonical amino acids, cysteine has served as the most convenient target owing to its highly nucleophilic sulfhydryl side chain group that offers a distinct reactive site in proteins.[7] Modification can be made based on its ability to react via exchange reactions with disulfide-containing reagents, as well as its ability to undergo alkylation with suitable electrophiles such as maleimides and iodoacetamides.[25] One of the most commonly used cysteine modifications is the reaction with maleimides, which has been applied to modify antibodies with fluorophores, cytotoxic drugs or PEG scaffolds.[25] The frequent use of this reaction has resulted in the commercialization of a wide variety of maleimide derivatives, which in turn has stimulated additional applications. Examples of more recently developed approaches for cysteine alkylation involve photocatalyzed reactions such as the irradiation of 3-(hydroxymethyl)-2-naphthol derivatives (NQMPs) to generate a Michael acceptor intermediate,[26] or the use of alkene- and alkyne-functionalized reagents to initiate radical reactions for thiol-ene[27] and thiol-yne[28] coupling reactions, respectively. While great efforts have been made towards modification of native cysteines, the most critical disadvantage is that most methods lack site-specificity (when multiple cysteine residues are present), which results in heterogeneous product mixtures.[20,24] Moreover, cysteines are relatively uncommon in protein sequences and often buried within the protein structure when involved in disulfide bridges, making them difficult to access.[29] Genetic incorporation of cysteine residues at sites of interest may disrupt protein structure and promote aggregation.

Reactions with the primary amines present on the side chain lysine residues has also been widely explored for protein modification.[30,31] Owing to the strong ionic character of amines (usually present in its protonated form), lysines are frequently present on the surface of proteins and are thereby favorable targets especially for multi-site conjugations. Harder nucleophiles, including NHS esters,[32] isothiocyanates,[33] and sulfonyl chlorides[34] were initially used for lysine labeling. However, such reactions were found to display cross-reactivity with the protein N-terminus. To achieve more selective modification,

kinetically controlled lysine modification using specially designed NHS-activated reagents was developed.[35] Activated esters with fluoro-substituted aromatic leaving groups were also recently harnessed to specifically label an antibody fragment on a specific lysine residue present in a unique chemical environment.[36] Despite potential chemoselectivity, the labeling efficiency is largely determined by the solvent accessibility and the chemical environment of the target lysine. Further improvement is needed.

1.2.2 Modification using unnatural amino acids.

In addition to modifying native amino acids, the introduction of non-natural amino acids into target proteins can also provide reactive handles for protein conjugation. These modified residues often bear bioorthogonal functional groups, which allow for more diverse and more specific subsequent conjugation reactions.[37-40]

To date, two unnatural amino acid (UAA)-incorporation methods have been described to incorporate UAAs. One of them exploits the ability of the existing aminoacyl tRNA synthetase (aaRS) to recognize certain UAA. In this scenario, the UAA will compete with the native amino acid substrates for the aaRS.[41] Using this method, a collection of amino acids bearing bioorthogonal moieties inlcuding azidomethionine, homopropargylalanine, and homoallylyglycine has been successfully incorporated into target proteins using methionyl-tRNA (MetRS).[42-44] However, it should be noted that the MetRS will potentially replace all the methionine residues with the UAA analogue. As a result, a heterogeneously modified protein-conjugates will be obtained after modification, which is not desirable.

A second, more selective approach, involves the reassignment of the stop codons, particularly the amber codon UAG, to insert a UAA.[45] It was found that suppressor tRNAs found in some organisms cause the introduction of amino acids at UAG codons in lieu of stopping translation. Based on this discovery, directed evolution methods have been used to generate aaRS that specifically aminoacylates the UAA and then transfers it to an orthogonal tRNA.[46] This aaRS-tRNA pair is then introduced into a different host, which expresses the target gene with the amber codon at a position of the protein sequence designated for modification. Such orthogonality provides high degree of specificity

without interference from endogenous natural amino acids and their complementary aaRS-tRNA pairs.

This technology has led to the development of a plethora of UAAs with diverse structures and functional groups (see Figure 1.2 for UAA structures). For instance, UAAs equipped with fluorescent reporters,[47] photocross-linkers,[48] or affinity handles[49] have been reported. Functionalization of UAAs with reactive functional groups, which can participate in a wide range of bioorthogonal reactions, is also achievable to allow conjugation of proteins to various cargos. However, several limitations still exist, including the decreased expression yield of the desired full-length UAA inserted proteins compared to the wild-type, the production of truncated protein products that complicate purification, the compatibility of the introduced orthogonal aaRS-tRNA pair to the expression system and the availability of the required plasmids.[50]

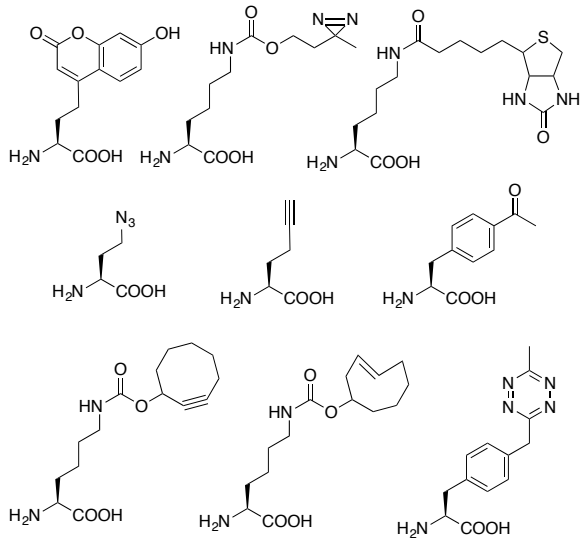


Figure 1.2 Structures of selected UAAs bearing fluorescent, cross-linker, affinity, or bioothogonal reactive moieties.

1.2.3 Bioorthogonal reactions.

Bioorthogonal functional groups refer to functionalities that are not typically present in biological systems, which can undergo reactions that do not occur with natural functionality and are inert within the cellular environment. Incorporating these exogenous functional groups into a target protein, by the insertion of a UAA or enzymatic methods, enables selective modification to produce a homogenous product. In addition, to exploit

their orthogonality, bioorthogonal reactions with high specificity have also been developed and successfully applied in a vast number of applications. In this section, several reactions that are commonly utilized for enzymatic protein labeling will be described. These include aldehyde/ketone condensations, azide-based click reactions and the tetrazine ligation. For a more comprehensive discussion of their mechanisms and applications, other reviews are readily accessible.[51-53]

Aldehydes and ketones are among the first functionalities that have been utilized for bioorthogonal protein labeling.[54] Carbonyl groups can react with α-effect nucleophiles, such as alkoxyamines and hydrazides in aqueous solutions to form oximes and hydrazones (Figure 1.3). The reactions favor acidic conditions and their rates are rather slow at neutral pH. Initially, aniline was employed to accelerate the reactions under both acidic and neutral conditions.[55] Catalysts with enhanced water solubility were later developed to further accelerate the reaction at neutral pH, including 5-methoxyanthranilic acid,[56] *m*-phenylenediamine[57] and *p*-phenylenediamine[58]. Electronic and acid/base effects also greatly influence the reaction rate at biologically relevant pH values, and carbonyl compounds with neighboring acidic/basic groups can form hydrazones at elevated rates[59]. Although oximes are more stable than hydrazones, the C=N bond is still susceptible to hydrolysis[60]. To overcome this limitation, several alternative ligations strategies have been developed to form stable C-C linkages, such as the Pictet-Spengler ligation[61]/hydrazine-Pictet-Spengler ligation (Scheme 1B),[62] and the trapped-knoevenagel ligation,[63] all of which have been applied to construct ADCs.

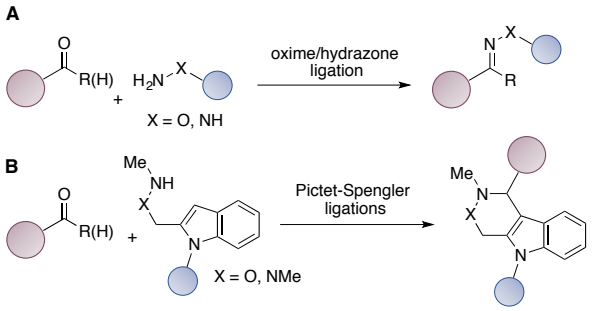


Figure 1.3 Selected bioorthogonal reactions based on aldehyde functionality. (A) Oxime/hydrazone ligation. (B) Pictet-Spengler ligations.

Aldehydes can be easily introduced into proteins of interest by formylglycine generating enzyme, which can convert a cysteine to an aldehyde-bearing formylglycine.[64] Aldehyde and ketone containing substrate analogues for protein farnesyltransferase,[65] N-myristoyltransferase,[66] tubulin tyrosine ligase[67] and lipoic acid ligase[68] have also been developed to facilitate site-specific enzymatic protein labeling. It should be noted that while efficient for *in vitro* protein modification and cell surface labeling,[69] the aldehyde/ketone-based reactions are less suited for *in vivo* applications due to the presence of endogenous carbonyl-containing compounds as well as the potential toxicity of the catalysts required for their efficient modification.

The azide group is essentially absent in biological systems and is generally inert to endogenous functionalities present in the biological milieu, rendering it an ideal candidate for bioorthogonal reactions.[70] Due to its small size, an azide group can be easily incoprporated into various enzyme substrates for protein labeling. Introduction of the reactive partners for azides, including terminal alkynes and strained-alkynes, to make enzyme substrate analogues can also be achieved. Examples for different enzymes are discussed in detail in the following section. Three bioorthogonal reactions involving azides have been extensively explored, including the Staudinger ligation, the copper-catalyzed azide-alkyne cycloaddition (CuAAC) and the strained-promoted azide-alkyne cycloaddition (SPAAC).

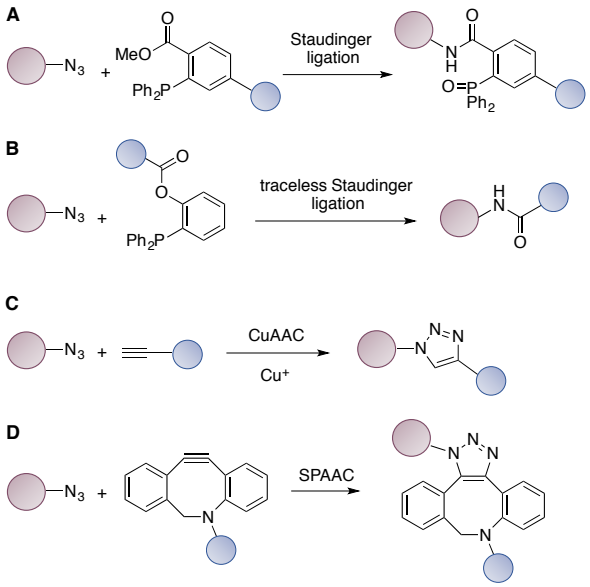


Figure 1.4 Selected bioorthgonal reactions based on azide functionality. (A) The Staudinger ligation. (B) The traceless Staudinger ligation. (C) CuAAC reaction. (D) SPAAC reaction. DBCO is shown as an example for the strained alkyne compound.

The Staudinger ligation (Figure 1.4A), traceless-Staudinger ligation (Figure 1.4B) and Staudinger-phosphite/phosphonite reactions exploit the reaction between azides and trivalent-phosphine reagents to form stable amide bonds.[71] As one of the earliest developed bioorthogonal reactions, the Staudinger ligation has been successfully utilized in both *in vitro* and *in vivo* applications.[72] However, the utility of these reactions is largely limited by their slow reaction rate and the oxidation propensity of the requisite phosphine reagents.[51] CuAAC is a versatile and powerful tool for protein labeling. In the presence of Cu(I) as the catalyst, which is usually generated *in situ* from the reduction of Cu(II) by reducing agents, an azide and a terminal alkyne react to form a stable triazole product (Figure 1.4C). [51,73] To stabilize the Cu(I) ion, a variety of chelating ligands have been synthesized and employed.[74-78] Owing to the fast kinetics and the excellent selectivity, the CuAAC reaction has been used for numerous *in vitro* protein modification applications. A major disadvantage of the CuAAC reaction arises from the cytotoxicity of the Cu(I) catalyst towards living cells, restricting its *in vivo* applications.[79] To reduce toxicity, efforts have been made to decrease the copper concentration while maintaining reaction efficiency by optimizing azide structures[80-82] and designing novel chelating ligands.[76,79]

SPAAC is also known as the copper-free click reaction. In the absence of a catalyst, alkynes can be activated via ring strain allowing them to react with the azides directly (Figure 1.4D), albeit at a much slower rate.[73] Attempts to improve cycloaddition rates by optimizing the structure of strained alkynes have been limited by the inherent instability of these compounds under physiological conditions.[53] Nevertheless, the SPAAC reaction has been widely exploited in bioconjugation with numerous accomplishments, especially for *in vivo* applications.[83,84] The increasing number of commercially available strained-alkyne modified cargos has also facilitated its application. It should be noted that certain limitations are still present, including the hydrophobicity of the aromatic ring-fused strained alkynes and the potential for side-reactions with thiols from cellular proteins.[85] Further tailoring and optimization of the CuAAC and SPAAC reactions will undoubtedly expand their utility for various biological applications.

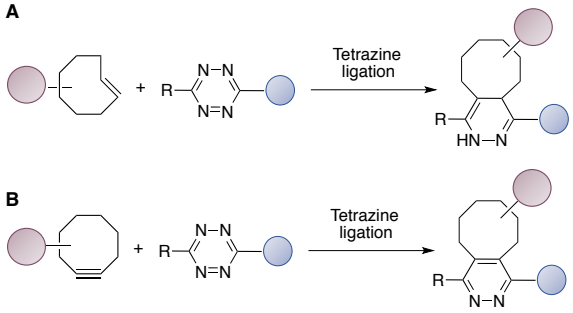


Figure 1.5 Scheme 3 Selected representations for tetrazine ligations. (A) Tetrazine ligation with TCO. (B) Tetrazine ligation with a strained alkyne.

The tetrazine ligation is the most rapid bioorthogonal reactions developed to date.[51] The second order rate constant for the reaction between a tetrazine and an alkene ranges from 1 to 10^6 $M^{-1}s^{-1}$, depending on the structure of the strained alkene (Scheme 3A) or alkyne reagent employed (Scheme 3B) including norbornenes, cyclopropenes, *trans*-cyclo-octenes (TCO) and bicyclononynes.[86] Tetrazines with different substituents also exhibit varied stability and reactivity towards strained alkenes.[53,87] The capability to tune the reaction rate means that specific reagents can be chosen based on the needs dictated by the specific *in vitro* or *in vivo* application. In particular, the rapid kinetics are extremely useful for assembling radioisotope-labeled proteins in cases where very short half-life isotopes are employed.[88] In addition, the extremely fast kinetics and selectivity has allowed the conjugation reaction to be performed *in situ* in live mice for tumor pre-targeting imaging

applications.[89] Importantly, the tetrazine ligation is orthogonal to the CuAAC and SPAAC reactions, enabling simultaneous dual labeling.[90] Introduction of tetrazine moieties into a target protein has been reported using microbial transglutaminase with a tetrazine-amine substrate.[91] As for the strained alkene or alkyne compounds, a TCO-modified triglycine peptide substrate for sortase A has been reported for the construction of immuno-PET reagents.[92] Lipoic acid analogues with TCO[93] and norbornene[94] were also utilized for protein fluorophore labeling in live cells. TCO-bearing analogues have also been developed for protein farnesyltransferase.[95]

1.3 Enzymatic protein labeling strategies

As efficient catalysts for chemical reactions, enzymes have been widely applied in a variety of applications. The high specificity, rapid reaction rates and mild reaction conditions make them excellent choices for protein labeling purposes. Here, recent examples of work performed with enzymes from four classes including peptidases, transferases, ligases and oxidoreductases are described. In addition to those discussed below, other enzymes, including peroxidase,[16] tyrosinase,[96] and the enzymes involved in glycan synthesis,[18] have also been employed on a more limited basis.

Table 1.1 Compatible bioorthogonal functional groups with different enzymatic labeling methods and their applications

Enzyme	Tag sequence	Labeling site	substrates	K_M[a] μM	k_{cat}[a] min^{-1}	k_{cat}/K_M[a] min^{-1} μM^{-1}	Bioorthogonal functional groups	Applications
Sortase A	LPXTG	C-terminus, internal sites	$G_{(n)}$ peptides, primary amines, hydrazides	140	90	0.64[97]	Azide,[98] terminal alkyne,[98] TCO,[92] BCN,[99] DBCO,[100]	One-step labeling with cargos, conjugation to small molecules,[92] polymers[101] and polypeptides,[102] Protein semisynthesis,[3] surface immobilization,[103] *in vivo* labeling,[104] cell-surface labeling[105]
				(GGG peptide)				
	$G_{(n)}$	N-terminus	LPXTG peptides	7600	90	0.012[97]		
Subtiligase	-	N-terminal $-NH_2$	Peptide ester	620	1260	2[106]	Azide[107]	One-step labeling with cargos, conjugation to small molecules,[107] protein semisynthesis[108]
				(s-AAPF-glc-PG-amide peptide)				
Microbial transglutaminase	Q tag (LLQG)	Any site	Lysine peptides, Primary amines	-	-	-	Azide,[109] terminal alkyne,[109] DBCO,[91] tetrazine,[91] triphosphine[91]	One-step labeling with cargos, conjugation to small molecules,[110] polymers[111] and polypeptides,[112] surface immobilization,[113] cell-surface labeling[114]
	K tag (MKHKGS)		Glutamine peptides	52660	2128	0.04[115]		
				(Z-QG)				
Farnesyltransferase	CaaX	C-terminus	Isoprenoid analogues	1.71	31	18[65]	Azide,[116] terminal alkyne,[117] TCO,[95] aldehyde,[65] ketone,[118]	Conjugation to small molecules,[120] polymers,[65] polypeptides[121] and oligonucleotides,[122] surface immobilization[123]
				(FPP)				
N-myristoyltransferase	GNEASYPL	N-terminus	Myristic acid analogues	5[124]	-	-	Azide,[125] terminal alkyne[126]	Conjugation to small molecules,[127] surface immobilization[128], *in vivo* labeling[125]
				(CoA-myristate)				
Phosphopantetheinyl transferase	GDSLSWLLRLLN	Any site	CoA derivatives	60.8	14.7	0.242[129]	-	One-step labeling with cargos, conjugation to small molecules[130] and oligonucleotides,[131] surface immobilization,[129] cell surface labeling[132]
				(CoA-biotin with ybbR tag by Sfp)				
Tyrosine ligase	Tub tag (VDSVEGEGEEEGEE)	C-terminal -COOH	Tyrosine analogues	-	-	-	Terminal alkyne,[133] azide,[67] aryl halides,[133] aldehyde,[133] aryl nitro[133]	Conjugation to small molecules[67,133]
			Functionalized-glycine					
Lipoic acid ligase	GFEIDKVWYDLDA	Any site	Lipoic acid analogues	4.5	15	3.3[134]	Azide,[135] terminal alkyne,[134], norbornene,[94] aldehyde,[68] aryl iodine,[136] TCO[93]	Conjugation to small molecules [68] and polymers,[135] surface immobilization,[135] cell-surface labeling[134]
				(lipoic acid with E2p domain)				
Biotin ligase	GLNDIFEAQKIEWHE	N- or C-terminus	Biotin analogues	0.3	18	60[137]	Ketone,[138] azide,[139] terminal alkyne[139]	Conjugation to small molecules,[140] surface immobilization,[141] cell-surface labeling[138]
				(biotin with BCCP domain by BirA)				
Formylglycine generating enzyme	CXPXR	N- or C-terminus	-	-	-	-	Aldehyde[64]	Conjugation to small molecules[61] and polymers,[64] cell surface labeling,[142] in vivo labeling[142]

[a] The kinetic parameters are values reported for selected "small molecule" substrates from the "substrates" column. In particular, the kinetic parameters for protein substrates with different tag sequences are not provided.

1.3.1 Sortase A.

Sortase A (SrtA) from *Staphylococcus aureus* is a Ca^{2+}-dependent transpeptidase.[143] It recognizes a consensus sequence, LPXTG (known as a sortag) and cleaves the amide bond between the threonine and glycine residues, forming an acyl-enzyme-intermediate. Subsequent attack from the N-terminal amine of an oligoglycine-terminating peptide as the nucleophile yields a ligated product with a new peptide bond (Figure 1.6A).[144] SrtA has been extensively studied for site-specific labeling of peptides and proteins. Protocols have been published describing detailed procedures to label the N-terminus,[145] C-terminus and the internal loops of proteins.[146] Several enzyme variants, including an evolved penta-mutant with enhanced catalytic efficiency[97] and a Ca^{2+}-independent mutant[147] as well as SrtA homologs from other bacterial sources with different recognition sequences have been reported to meet the requirements of various applications.[148-151] Some of the early uses include protein lipid modification,[152] cyclization,[148] and cell-surface labeling,[153] most of which have been summarized in several reviews.[18,154,155] As an active area of research, an enormous range of applications have been reported since 2013, including but not limited to the semisynthesis of proteins with post-translational modifications,[3,156] protein immobilization on solid surfaces,[103,157-161] protein labeling on liposomes,[162,163] virus-like particles[164,165] and hydrogels[166-169] as well as cell surface labeling[105] and *in vivo* protein labeling.[98,104,170] A few of these applications will be discussed in detail below.

In recent years, SrtA has been applied to label full-length antibodies[171,172] and their derivatives[99,173,174] to create homogeneous conjugates. Grawunder and coworkers at NBE-Therapeutics AG disclosed the construction of homogeneous and site-specific counterparts of brentuximab vedotin (Adcetris) and trastuzumab emtansine (Kadcyla), using the evolved SrtA penta-mutant.[171] The sortag was incorporated at the C-termini of both the heavy chain and light chain of the antibody. For the light chain modification, poor labeling was initially obtained due to the inaccessibility of the termini. The insertion of a short peptide spacer between the C-terminus and the sortag was shown to improve the labeling yield substantially. This strategy has also been widely adopted to enhance labeling of sterically hindered or buried protein termini. Cytotoxic drugs were appended to a penta-glycine moiety with the same linkers used in their corresponding commercial counterparts

and subsequently conjugated to the antibodies by SrtA. The enzymatically generated brentuximab vedotin had a DAR of 3.2, indicating conjugation efficiency of 80%. When tested for tumor killing *in vitro* and *in vivo*, similar potencies were demonstrated between the site-specifically conjugated and the chemically conjugated ADCs. Researchers from the same company later described the generation of ADCs bearing novel anthracycline-based cytotoxins by sortase-mediated antibody conjugation (SMAC) technology.[172] They showed that these novel ADCs exhibited potencies exceeding those of Kadcyla and Adcetris, both of which are based on conventional tubulin-targeting payloads. This finding highlighted the importance of exploring alternative toxic payloads for the preparation of highly potent next-generation ADCs.

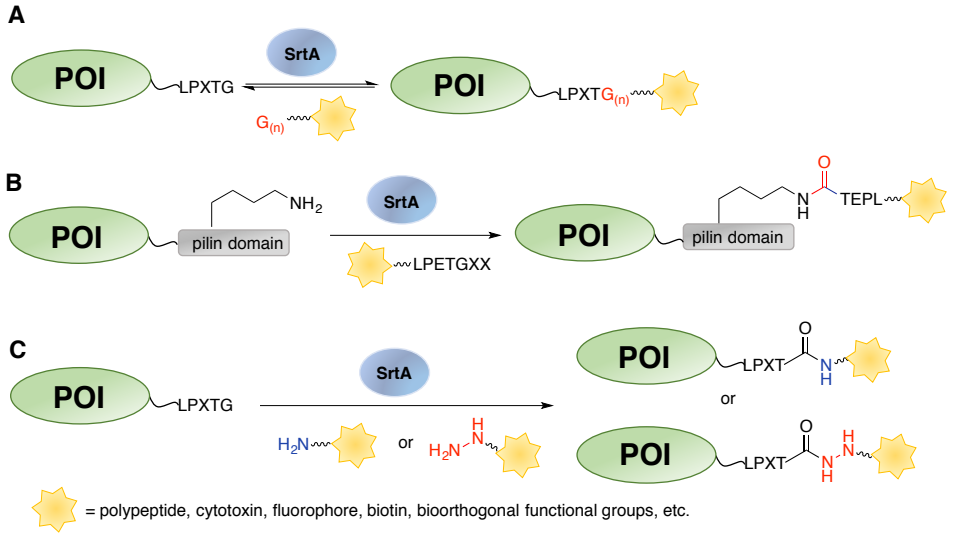


Figure 1.6 Enzymatic labeling by SrtA. (A) Canonical C-terminal labeling catalyzed by SrtA using oligoglycine substrates. (B) Labeling of the lysine (in pilin domain) by SrtA using LPETG peptide substrates. (C) Protein labeling at the C-terminus by SrtA using primary amines or hydrazide substrates. POI: protein of interest.

In addition to regular antibodies, SrtA has also been utilized to label a nanobody,[92,100,175-177] a single-domain antibody (VHH) derived from alpacas and camels. Rashidian et al. used this strategy to label VHH for imaging and radio-diagnostic applications.[92] They incorporated the sortag at the C-terminus of VHHs as this site is positioned away from the target-binding region. Labeling of two VHH constructs, DC8, an anti-Class II MHC nanobody and DC13, targeting CD11b, were evaluated. Dual-functionalized triglycine substrates comprised of two bioorthogonal handles were employed. One of the handles was used for the introduction of a fluorophore or ^{18}F isotope while the other one was conjugated to a PEG moiety or a second VHH protein to tune the

serum half-life or avidity of the conjugates. From PET imaging studies, the bivalent VHH homodimers were found to allow visualization of their targets more effectively than the monomers *in vivo*. Meanwhile, the PEGylated VHHs also displayed improved target staining. All the constructs exhibited high specificity. Apart from the imaging applications, Ploegh and coworkers also made structurally defined homogeneous VHH-drug conjugates.[176] An Class II MHC recognizing VHH7 was modified with a C-terminal sortag and reacted with triglycine peptides containing DM1 as the toxic payload using SrtA. A VHH7-NIR dye conjugate was also prepared to evaluate target binding, cellular internalization and *in vivo* localization of the VHH7. When the VHH7-DM1 was subjected to *in vivo* efficacy tests using highly invasive B cell lymphoma mouse xenografts, the drug-conjugate was shown to decrease tumor size significantly compared to the control group and to reduce metastatic spread.

In the area of creating protein-polymer conjugates, Gao and coworkers reported the *in situ* growth of a polymer chain from the target proteins directly.[178] In this study, a PEG-like polymer, POEGMA, was polymerized from a therapeutically relevant protein, interferon alpha (IFNα), to form a site-specific and stoichiometrically modified IFNα-POEGMA conjugate.[101] Briefly, IFNα was encoded with a C-terminal sortag and labeled with glycine substrates appended with an atom transfer radical polymerization (ATRP) initiator by SrtA. The resulting IFNα-Br acted as a macroinitiator. In the presence of OEGMA monomers under appropriate conditions, the polymer chain was grown from the IFNα protein, forming an IFNα-POEGMA conjugate with a high conversion of 90%. Importantly, when compared to PEGASYS, a commercial non-specifically PEGylated IFNα product, the IFNα-POEGMA exhibited superior pharmacokinetics and showed improved anticancer efficacy over PEGASYS, presumably due to decreased blockage of the binding site using site-specific conjugation strategy. As an alternative method to create protein-polymer conjugates, this site-specific *in situ* growth technology is promising for the development of more effective next-generation protein therapeutics.

In addition to the usage of canonical oligoglycine-containing peptides as substrates, SrtA was shown to have relaxed specificity for the amine nucleophiles.[179,180] When positioned close to the active site of the SrtA, the non-protonated primary amine from a lysine residue can displace the acyl-enzyme intermediate, forming an isopeptide bond

(Figure 1.6B).[181,182] Chilkoti and co-workers took advantage of this reaction and used SrtA to site-specifically conjugate small molecules to lysine residues on proteins.[183] They first examined the labeling reaction between a sortag-containing peptide and a pilin domain (PLN) peptide, which was derived from a natural bacterial protein that can form protein polymers via other related sortase enzymes. The PLN sequence contained a valine at the N-terminus and an internal lysine residue. Successful isopeptide ligation in the presence of SrtA was confirmed by MS. The authors then linked the pilin domain sequence to a fibronectin type III (Fn3) domain protein, an alternative target-binding protein. Fn3 was generated with C-terminal elastin-like polypeptides (ELP) to facilitate expression and purification. Several copies of the pilin domain were inserted between Fn3 and ELP to generate Fn3-PLN_3-ELP. A biotin-appended sortag peptide was incubated with Fn3-PLN_3-ELP overnight with SrtA. The conjugated product was detected by Western blotting against biotin while a control experiment with Fn3-ELP lacking the pilin domain did not yield modified proteins. The site of modification and the formation of the isopeptide bond was further confirmed by LC-MS/MS characterization of the trypsin-digested protein, revealing site-specific modification at the pilin Lys with no off-target labeling. Having established the isopeptide ligation with the Fn3 protein, the authors modified a monoclonal antibody with pilin domain insertion at the C-terminus of the heavy chain using a biotinylated sortag peptide. Approximately 1.8 biotins were incorporated per antibody, and the modified antibody retained its antigen binding capability, as validated by immunofluorescence microscopy. Thus, this isopeptide ligation can conjugate multiple cargos to a protein of interest at either internal or terminal sites, presenting a new bioconjugation technique for protein labeling. Additionally, optimization of the sequence and length of the pilin domain may further improve the reaction kinetics and ligation yield.

Not limited to the ε-amino groups of lysine residues, other primary amine-containing compounds have also been explored as substrates for SrtA (Figure 1.6C).[184,185] Cochran and coworkers reported the usage of a SrtA Hepta-mutant (SrtA7M)[170] to create bioorthogonally tagged proteins of interest directly from *E. coli* using inexpensive, cell permeable and commercially available amine compounds.[98] This SrtA mutant combined the mutations from the evolved enzyme with high catalytic activity and the mutations that abolished Ca^{2+} binding, yielding a Ca^{2+} independent SrtA variant with high reactivity. To

label proteins directly in *E. coli*, SrtA7M was co-expressed with several target proteins engineered with C-terminal sortag respectively. Upon the addition of 3-azido-1-propanamine to the growth culture, they showed that three different target proteins were modified with the azide functional group respectively. Although endogenous bacterial proteins containing an LPXTG sortag were identified from the *E. coli* BL21(DE3) proteome using a RefSeq protein database search, SrtA7M-expressing cell lysate incubated with Cy3-DBCO exhibited minimal background labeling, suggesting the high specificity of the SrtA reaction. By using this technique, large quantities of labeled protein were obtained from cell lysate with one-step purification, greatly facilitating the synthetic process. Interestingly, hydrazine and its derivatives were also shown to be accepted by SrtA as nucleophilic substrates.[186] Liu and coworkers first evaluated the SrtA-mediated hydrazinolysis of a model sortag peptide. A yield of greater than 95% was detected by HPLC while increasing the hydrazine concentration accelerated the reaction. The hydrazinolysis reaction was then applied to protein labeling. Several applications were illustrated including hydrazide-based protein semisynthesis as an alternative to native chemical ligation, installation of alkyne or azide functionalities that were further coupled to polymers to create PEGylated proteins, and construction of protein-fluorophore conjugates using a hydrazide-modified fluorophore. Protein synthesis using SrtA-mediated protein hydrazide ligation was also exemplified in recent work by Li and coworkers, where they synthesized p62 protein with various phosphorylation states and studied their interactions with K63 diubiquitin.[156]

These noncanonical nucleophile substrates can be used to complement the use of oligoglycine compounds, significantly broadening the substrate scope for SrtA. One limitation associated with the oligoglycine substrates is that the ligated product can also be recognized and cleaved by SrtA, resulting in decreased ligation yield. Through the use of these alternative substrates, the ligated protein lacks the glycine residue present in the sortag sequence, which helps to prevent hydrolysis and to improve conversion. However, further engineering of the SrtA to enhance specificity and reactivity towards the noncanonical substrates is highly desirable so that less SrtA and substrates can be used to achieve acceptable yields. To assist the purification of labeled proteins, efforts have been made to develop one-pot labeling and purification techniques.[187,188] As one example,

Tsourkas and coworkers initially described the sortase-tag expressed protein ligation (STEPL) technique.[189] A construct containing the protein of interest followed by the sortag sequence, a $(GGS)_5$ flexible linker, SrtA and a His-tag was designed. The expressed protein fusion was absorbed onto Ni-NTA resin from the cell lysate. Upon the addition of Ca^{2+} and various oligoglycine peptides with different cargos, the target proteins were labeled and eluted, leaving the SrtA still attached to the resin (due to the His-tag). Both the concentrations of the Ca^{2+} and the oligoglycine peptides had an impact on the ligation yield and product purity, which needed to be optimized. Although successfully employed to label a HER2-recognizing affibody, the method was not generally applicable since some proteins expressed poorly once linked to SrtA and in some rare cases the fused SrtA became catalytically inactive. To overcome those limitations, the same group developed a proximity-based sortase-mediated ligation (PBSL) technique.[190] In this system, the target protein was expressed as fusion to the C-terminal sortag, a $(GGS)_5$ linker and a SpyTag peptide. The latter is a short 13-residue peptide that is covalently linked to its partner protein, SpyCatcher, upon binding. An additional construct where SrtA was fused to the N- or C-terminus of a His-tagged form of SpyCatcher was also prepared. The SrtA-SpyCatcher fusion was first immobilized on resin, which was then used to capture eGFP-LPETG-SpyTag from clarified cell lysates. The captured eGFP was modified and released from the resin by the addition of Ca^{2+} and triglycine peptides. Importantly, the cleaved SpyTag remained attached to the SpyCatcher on resin, making PBSL a traceless labeling technique. Peptide concentration and reaction time were optimized to minimize product hydrolysis and to achieve high ligation efficiency. When comparing the ligation yield between PBSL, STEPL and the conventional sortase reaction using an anti-CD3 ScFv protein, STEPL failed to produce the target protein fusion with SrtA while PBSL yielded 2.5-fold more labeled protein than the traditional reaction in a much shorter time. A similar on-resin cleavage and ligation strategy was also reported by Cheng et al. to functionalize peptides.[191] The peptide target with a C-terminal LPXTG sortag was synthesized on PEGA resins that are hydrophilic and permeable to macromolecules up to 35 kDa using standard solid phase peptide synthesis (SPPS). Diglycine compounds with various functional moieties, including biotin, lipid, PEG polymer and polypeptides, were utilized to successfully label and release the ligated peptide from the resin in one-pot using SrtA.

In summary, SrtA has been extensively applied for site-specific protein modification both *in vitro* and *in vivo*. Although the wild-type enzyme is limited by slow reaction kinetics and only moderate ligation yields, methods to improve the reaction have been developed and will continue to be explored. It should be noted that despite the improved catalytic efficacy, the evolved penta-mutant enzyme has a decreased affinity towards the nucleophilic substrate,[97] leading to elevated hydrolysis in some cases.[192] As mentioned above, further engineering of SrtA to improve reaction kinetics while suppressing hydrolysis should allow more efficient labeling using much lower substrate and enzyme concentrations.

1.3.2 Subtiligase.

Subtiligase is an engineered peptide ligase derived from subtilisin BPN', a serine protease obtained from *Bacillus amyloliquefaciens*.[193] It catalyzes the ligation reaction of an acyl-donor peptide ester to the N-terminal α-amine of the acceptor peptide, forming a native peptide bond (Figure 1.7). Compared to the parent subtilisin, the active site S221 was changed to cysteine (S221C) in subtiligase. This mutation shifted the catalytic mechanism to strongly favor aminolysis over hydrolysis.[106] A second mutation of P225 to alanine (P225A) improved the aminolysis efficiency, due to reduced steric crowding in the active site.[106] The double mutant (S221C/P225A) of subtilisin BPN' is known as subtiligase.[194]

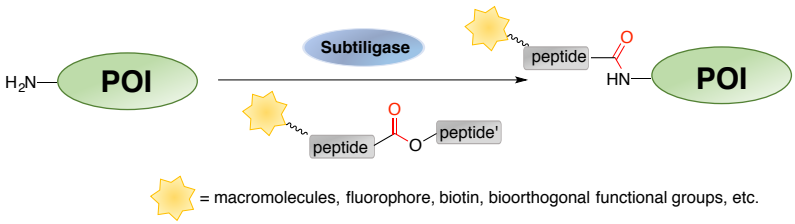


Figure 1.7 Protein N-terminal labeling by Subtiligase using peptide ester substrates.

Capable of catalyzing the formation of a native peptide bond, subtiligase has been used for the synthesis/semisynthesis of large proteins. Notably, using synthetic peptide fragments as building blocks, unnatural amino acids or residues equipped with post-translational modifications can be easily incorporated into the synthetic protein at any

defined position. In an early example, Wells and coworkers performed the total synthesis of Ribonuclease A from 6 peptide fragments using subtiligase.[195] Beyond the native protein, variants containing unnatural catalytic residues were also constructed, which helped to elucidate the catalytic mechanism of the enzyme. It should be noted that one major limitation associated with this total synthesis technique is that it is only applicable to proteins that can be refolded *in vitro*. As to protein semisynthesis, subtiligase is also able to carry out ligations with recombinant protein thioester fragments. As an alternative to the expressed protein ligation (EPL), which requires a cysteine at the ligation junction, careful selection of the conjugation site allowed enzyme-catalyzed expressed protein ligation by subtiligase to be used to synthesize proteins with their precise wild type primary sequence. In 2016, Cole and co-workers demonstrated that phosphorylated PTEN, a tumor suppressor lipid phosphatase, synthesized by subtiligase exhibited a more tightly closed conformation than that of a Y379C mutant generated from EPL, which was found to behave anomalously in cells.[108]

Since the enzyme can distinguish the N-terminal α-amine unambiguously over the lysine ε-amine, subtiligase has also been applied to study apoptosis by probing the unblocked protein N-termini generated from proteolytic cleavage by caspases and caspase-like enzymes.[196] A biotinylated peptide ester tag was designed by Wells and coworkers, which contained a biotin moiety for affinity enrichment and a TEV protease cleavage site to release the captured peptides from beads. To identify the newly exposed N-termini in apoptotic cells, both Jurkat cells treated with or without etoposide, a topoisomerase II inhibitor, were analyzed. Proteins in the cell lysate were labeled with the peptide ester tag by subtiligase. Following trypsin digestion, the biotin-labeled peptides were enriched using avidin immobilized beads and then subjected to LC-MS/MS characterization. The detected N-terminal sequence information was valuable in mapping the cleavage site of the caspase-like enzymes. Interestingly, the identified sequences in this study were generally not predicted by *in vitro* caspase substrate specificity studies, underlining the importance of cellular profiling over *in vitro* assays. The same subtiligase labeling technique was also utilized to monitor the regulation of post-translational modifications occurring at protein N-termini, such as N-α-acetylation.[197]

For protein labeling applications using subtiligase, it is straightforward to envision that peptide ester substrates containing bioorthogonal functional groups can be ligated to the N-terminus of the protein target, enabling subsequent conjugation. The site-specificity relies on the ability of the enzyme to recognize only the N-terminal α-amine rather than a specific peptide recognition tag, which is characteristic of other enzymes, such as SrtA. In this scenario, a broad substrate specificity would be particularly useful so that a variety of proteins can be modified by subtiligase without the effort of genetic modification. Previous qualitative studies revealed that subtiligase exhibited relatively broad reactivity towards α-amine peptides with some sequence preference. The incomplete characterization made it difficult to predict whether a specific protein of interest would be modified by subtiligase or not. Very recently (2018), Weeks et al. employed an approach, called proteomic identification of ligation sites (PILS), to comprehensively characterize the N-termini sequence specificity of subtiligase.[107] To ensure full coverage of the sequence diversity, peptide libraries were generated from protease digestions of the *E. coli* proteome. A similar biotin-containing peptide ester tag as described above was utilized to label substrate peptides, which were then captured on beads, released, and analyzed by LC-MS/MS. It was found that the reactivity of subtiligase was mainly determined by the first two residues (P_1' and P_2') of the N-terminal peptide substrates. Substantial sequence preferences were observed in both P_1' (small amino acids, methionine and arginine) and P_2' (aromatic and large hydrophobic residues) positions. To expand the substrate scope, subtiligase mutants were designed and generated to modify peptide sequences with disfavored residues by the wild-type enzyme. For example, a Y217K mutant was found to improve the reactivity of sequences with an acidic P_1' residue while an F189R mutant increased modification efficiency towards peptides containing an acidic P_2' residue. A free-access web-based tool, α-Amine Ligation Profiling Informing N-terminal Modification Enzyme Selection (ALPINE), was established to help guide the selection of the optimal subtiligase variants for modification of proteins and peptides based on user-defined N-terminal sequences. As a proof of concept, an anti-GFP antibody was modified with an azide-containing peptide at the N-terminus of the heavy chain. Subsequent SPAAC reaction with a DBCO-modified cargo was employed to install either a biotin, a Cy3 fluorophore, an MMAE cytotoxin, a 30-mer oligonucleotide or a PEG polymer to the labeled antibody. Importantly, the Cy3

modified anti-GFP Fab was shown to retaining its target-binding capability. In the cases where the N-terminus of the protein substrate was buried, multiple rounds of labeling by subtiligase or introduction of a short peptide extension were shown to substantially improve the conjugation efficacy.

Overall, subtiligase is a useful tool for protein labeling, whose utility has not been fully explored. The reaction is fast, site-selective and highly efficient. More importantly, the ability to directly modify the native N-terminal sequence of the protein substrates eliminates the need for additional genetic manipulation. This feature combined with the web-based ALPINE tool greatly facilitates the labeling process, which can be readily adapted for a variety of applications.

1.3.3 Microbial transglutaminase.

Transglutaminase catalyzes the amide bond formation between the γ-carboxamide group of a glutamine residue and the ε-amino group of a lysine (Figure 1.8). Microbial transglutaminase (MTG), isolated from *Streptomyces mobaraense*, has been widely utilized in food and textile industries to improve the texture and appearance of the products via protein cross-linking as well as for biotechnological applications to create various protein conjugates.[96,198] Substrate specificity studies reveal that the surface accessibility of glutamine or lysine residues and the backbone flexibility where they are located are crucial factors that dictate enzyme reactivity.[199,200] Additionally, the activity of MTG is also affected by the surrounding residues close to the modification sites.[201] Therefore, even though multiple glutamine or lysine residues are present on a protein surface, only a few of them can be modified by MTG.[199] Concerning the acyl acceptor substrate, both lysine-containing peptides and primary amines with less steric hindrance are found to be recognized by the enzyme.[202] Several peptide sequences containing the glutamine or lysine residue, referred to Q-tag (Figure 1.8A) and K-tag (Figure 1.8B) respectively, have been found to be efficient substrates for MTG and can be engineered into a protein of interest for site-specific modification.[203-205] Importantly, these sequences can be inserted at either the C-terminus,[206] N-terminus[113] or internal loops.[207]

Construction of protein conjugates by MTG can be accomplished in one step using peptide or amine substrates appended to various cargos. Early examples included

fluorophores,[206,208] PEG polymers,[209] metal chelators,[110] or oligonucleotides.[205] Since 2013, two-step approaches consisting of initial protein labeling with a bioorthogonal functional group, such as alkyne,[210] azide,[109,211] or tetrazine[91] have been widely explored. Subsequent conjugation reactions are then performed to install the desired cargos onto the protein of interest. In addition, to avoid tedious purification steps and improve overall yield, one-pot conjugation combining the enzymatic protein labeling and bioorthogonal reactions has also been studied. In 2016, Pelletier and co-workers discovered that the addition of glutathione preserved the activity of MTG in the presence of copper.[15] Since the enzymatic reactions catalyzed by MTG occurred significantly faster than the CuAAC reaction, they were able to conjugate α-lactalbumin, bearing two reactive glutamines, to a Cy5 fluorophore by simply mixing MTG, the amine-substrates with either alkyne or azide, the clickable Cy5 and the CuAAC reagents. To follow up, they expanded the reaction scope to metal-free reactions and demonstrated that this one-pot strategy was also compatible with SPAAC and tetrazine ligation reactions.[91]

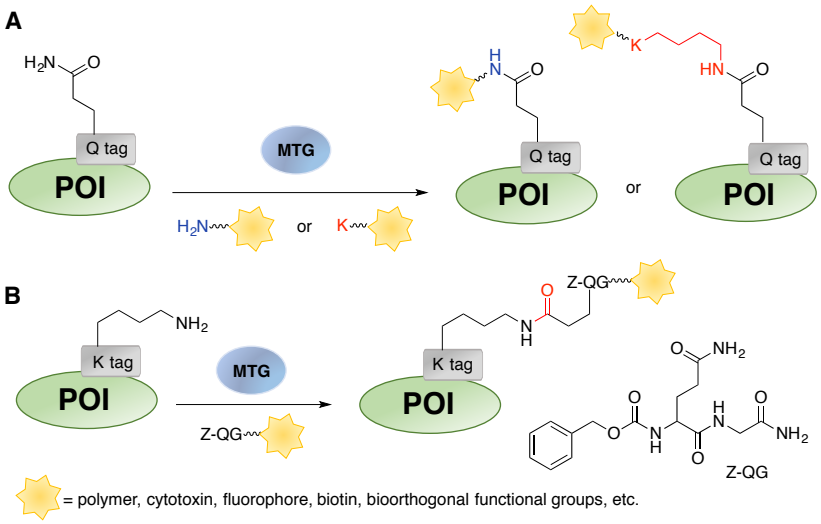


Figure 1.8 Enzymatic protein labeling by MTG. (A) Labeling of a Q tag-inserted protein using lysine or primary amine substrates. (B) Labeling of a K tag-modified protein using ZQG peptide.

Due to the high demand in the manufacturing of homogeneous antibody conjugates, the utility of MTG for such purposes has been extensively explored. Early work by Schibli and co-workers evaluated the labeling of full-length antibodies by MTG using both amine and glutamine substrates.[212] It was found that the glycosylated antibodies were poorly modified by MTG while an unglycosylated antibody could be labeled efficiently with

amine substrates. To explain the difference in reactivity, they identified the modification site to be Gln295, which is spatially close to Asn297.[110] As a conserved glycosylation site, Asn297 could mask the neighboring Gln295 upon glycosylation, rendering it inaccessible to MTG. Based on these findings, the authors developed a method to create antibody conjugates with uniform composition by first deglycosylating the antibodies using PNGase F to expose Gln295. The resulting antibodies were then reacted with MTG to install functional moieties, including metal chelators for imaging purposes[110] and cytotoxins for therapeutic applications.[211] Since Gln295 is preserved in all IgG subtypes, this method can potentially be applied to any given antibody without the need for genetic manipulation.

The integration of Q tags into antibody sequences has also been explored for site-specific modification. Through screening of multiple conjugation sites on an antibody, Rajpal and co-workers demonstrated that the modification location had an impact on the linker stability and the *in vivo* pharmacokinetics of the ADC, creating the opportunity to optimize the therapeutic index of ADCs by tuning the labeling site.[213] Beyond these preliminary studies, in 2016, Strop et al. developed a Trop-2 targeted ADC using site-specific MTG-mediated conjugation and characterized its anti-tumor efficacy and toxicity comprehensively. An LLQGA tag was appended to the C-terminus of the antibody heavy chains and then linked to a lysine-vc-PAB-toxin moiety by MTG.[214] The resulting ADC was highly efficacious in multiple mice xenografts bearing different solid tumors. The safety profile with nonhuman primates indicated on-target epithelial toxicities, which are fully recoverable. These preclinical efficacy and safety results supported the advanced clinical development of the ADC against multiple solid tumor types. Lastly, antibody labeling at native or engineered lysine residues has been reported more recently, allowing the construction of orthogonally dual-labeled antibodies.[215,216]

To streamline the process of antibody conjugation, on-bead modification methods were also developed. In 2016, Graziani and co-workers showed that antibodies immobilized on beads functionalized with protein A could still be modified by MTG with high efficacy.[217] This procedure replaced multiple purification steps by simple washing and could be easily adapted to a 96-well format, allowing rapid screening and optimization of the ADCs. In a related approach, Spycher et al. attached MTG onto glass microbeads.[215] It was shown that the MTG on the solid support retained its catalytic activity to link different

substrates to various proteins, including scFv and full-length antibodies. Notably, higher selectivity was observed for the immobilized MTG in that while two lysine residues on avidin could be modified by MTG in solution, only the one located in a highly-disordered region was labeled using the immobilized MTG. That difference was attributed to the decreased rotational flexibility of the immobilized MTG, which can be potentially beneficial to produce highly homogeneous protein conjugates.

Beside protein modification with small molecules, construction of protein-polymer conjugates using MTG has also been extensively investigated.[209,218] PEGylation of small therapeutic proteins by transferring an amine-functionalized PEG onto intrinsic glutamine residues has been demonstrated. As noted earlier, accessibility and chain flexibility are important determinants of MTG reactivity. Therefore, it is possible to achieve site-selective modification on endogenous glutamine residues without genetic engineering. Such examples include the PEGylation of human growth factor (hGH, Q40 and Q141),[219] human interleukin-2 (hIL-2, Q74),[220] granulocyte colony stimulating factor (GSCF, Q134),[221] and more recently interferon alpha-2b (IFNα-2b, Q101).[222] To further increase the selectivity of MTG, Pasut and co-workers explored the use of organic co-solvents, which was known to influence protein conformation and flexibility.[223] It was found that highly selective conjugation could be achieved when appropriate co-solvents were used. For example, only Q141 was PEGylated when the enzymatic reaction with hGH was performed in 50% ethanol.[224] Different from the widely explored modification on glutamine residues, MTG mediated PEGylation of lysine residues is relatively rare and has only been studied since 2016. In one case, a PEG polymer linked to N-benzyloxycarbonyl-L-glutaminylglycine (ZQG), a well-studied glutamine-bearing dipeptide substrate for MTG. Similar to observations of glutamine labeling, selectivity was also achieved with several proteins. K41 on GSCF[111] and K164 on IFNα-2b[222] were found to be modified by MTG despite the presence of other lysine residues on the proteins.

Protein labeling with oligonucleotides via MTG has also been reported. In one of the earliest examples, Goto and co-workers incorporated ZQG onto the 5'-end of aminated DNA, forming ZQG-DNA that could be recognized by MTG.[205] A K-tag (MKHKGS) was fused to two model proteins, bacterial alkaline phosphatase (AP) and eGFP at either N- or C-terminus. The AP-DNA conjugates were immobilized onto cDNA-displaying

microplates via DNA hybridization. Subsequent addition of a fluorescent AP substrate yielded an intense fluorescence signal in the AP-DNA treated wells, indicating that the biological functions of both the DNA and the AP enzyme were retained. This concept was further expanded to use ZQG-dUTP as building blocks to synthesize DNA/RNA–$(enzyme)_n$ conjugates, which can be applied for sensitive DNA/RNA detection.[225,226] More recently, Takahara et al. designed NH_2-dUTP that was easier to synthesize and could be added to the 3'-end of a DNA aptamer by terminal dexoynucleotidyl transferase (TdT).[227] The NH_2-clustered aptamer was crosslinked to Q-tag (FYPLQMRG) fused eGFP by MTG. Approximately 3 to 6 copies of eGFP were loaded onto each aptamer. Notably, purification of the conjugates from unreacted eGFP was readily achieved by the simple use of a centrifugal filter. The conjugated aptamer retained its binding capability towards cell surface targets as confirmed by confocal imaging. Similarly, a polymer chain was also used to mount multiple proteins. The same group prepared a polymerizable methacrylate-ZQG monomer and polymerized it with acrylamide to form polymer chains.[228] A K-tag (MRHKGS) was introduced to the C-terminus of the protein substrate, in this case, a fusion protein containing a binding domain to the antibody Fc region and a chimera alkaline phosphatase (IPP-PG). The polymer-enzyme conjugates were used in place of the secondary antibody-enzyme constructs for the detection of biomolecules via ELISA. Due to the clustering of proteins in one polymer chain, a maximum of 5-fold increase in signal detection was achieved when compared to that of the free protein, highlighting the advantageous signal amplification effect from multivalency.

Overall, MTG is a versatile and promising tool for protein modification. A number of fruitful applications have been demonstrated, and it is expected that more will be reported in the future. The ability of MTG to label intrinsic residues on the protein substrates greatly simplifies the production process. However, due to its relatively broad substrate specificity, it is difficult to predict whether a given protein of interest will be modified or not. Experiments are required to determine the extent of modification. Extra care must also be taken to examine potential protein-crosslink by-products and to evaluate if the endogenous conjugation site is detrimental to protein function or stability.

1.3.4 Farnesyltransferase.

Protein prenylation is a post-translational lipid modification that can is catalyzed by farnesyltransferase (PFTase). The enzyme recognizes a C-terminal CaaX motif and transfers the isoprenoid moiety from the natural substrate, farnesyl diphosphate (FPP), to the CaaX-box cysteine via a thioether linkage (Figure 1.9); the recognition sequence is denoted as a CaaX box, where C is a cysteine, a represents aliphatic amino acids, and X determines the substrate specificity towards different enzymes in the same family.[229] Extensive studies investigating the enzyme specificity revealed that PFTase is promiscuous towards both the peptide recognition sequence[230-233] and the isoprenoid substrate.[231,234,235] Interestingly, it was discovered very recently (2018) that in addition to the canonical CaaX box, some peptides and proteins bearing C-terminal CaaaX motif can also be accepted by PFTase as substrates.[236] This finding expanded the scope of prenylation within the proteome, implying more proteins may be potentially prenylated, whose cellular functions remained to be explored. For protein labeling applications, a number of isoprenoid analogues have been previously developed that were functionalized with various groups, including fluorophore,[237-239] photoaffinity moieties,[240-242] biotin,[243] azide-,[244-247] alkyne-,[117,248,249] and aldehyde-functionalities[65,250]. It has been shown that these analogues can be incorporated into target proteins bearing engineered C-terminal CaaX-box sequences. A variety of applications were demonstrated, ranging from cellular protein imaging,[251] proteomic analysis,[244,252,253] surface immobilization,[116,246,254] and construction of protein conjugates.[248,255] These early applications were summarized in the review by Rashidian et al.[18] The examples discussed below focus on developments since 2013.

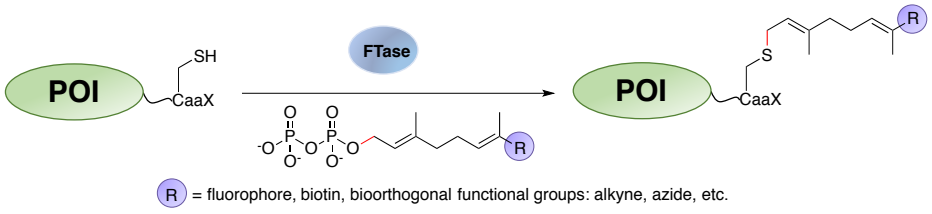


Figure 1.9 Protein labeling at C-terminus by PFTase using isoprenoid analogues bearing bioorthogonal functional groups.

One of the common traits of the isoprenoid analogues described above is that they all contained only one functional moiety, such as a fluorophore/biotin or a reactive handle for subsequent conjugation. To enable protein dual-functional labeling, Distefano and co-

workers designed an analog containing both alkyne and aldehyde functional groups.[120] With these two handles, they were able to install a fluorophore as well as a PEG polymer into proteins of interest. Since the CuAAC and oxime ligation are orthogonal to each other, both conjugation reactions can be performed simultaneously without cross interference. The designated probe is particularly useful to create multi-functional protein conjugates, such as theranostic reagents and PEGylated protein-drug conjugates with improved pharmacokinetic properties. Beyond labeling individual proteins, this multifunctional substrate has been used to assemble nanorings structures bearing multiple copies of a model protein that can be internalized into cells. To expand the scope of compatible bioorthogonoal reactions, Wollack et al. also developed a TCO-functionalized isoprenoid analog and demonstrated that it could be efficiently transferred onto protein substrates by PFTase.[95] Subsequent reactions with tetrazines were also confirmed by MS. Since tetrazine ligation is much faster than the SPAAC reaction without the need of a metal catalyst, this labeling strategy can be potentially useful for site-specific protein labeling inside living cells.

Although PFTase is sufficiently promiscuous to process these more bulky isoprenoid analogues, their catalytic efficiencies is generally lower than that of the FPP. Aiming to address this limitation, Dozier et al. analyzed how different PFTase mutations within the active site could be used to better accommodate these larger analogues.[256] Based on the crystal structures of PFTase and previously reported literature,[257] three key residues were identified that were in close contact with the third isoprenoid unit of FPP. Accordingly, three mutants were prepared, including W102A, Y154A, and Y205A. A panel of four analogues was examined, containing an aryl aldehyde-, a TCO-, a coumarin- or a nitrobenzoxadiazole (NBD) group. It was found that the mutant enzymes catalyzed the reactions more efficiently than the wild-type enzyme to a varying degree. Of particular note, the Y205A mutant exhibited a 300-fold increase in k_{cat}/K_M for the coumarin substrate, manifesting a catalytic efficiency comparable to that of the wild-type enzyme for the natural substrate, FPP. Overall this study established these mutant enzymes as useful tools for protein labeling. Further work demonstrated that it was also feasible to further engineer PFTase to create mutants with substrate specificities orthogonal to the wild-type enzyme, which can be used for simultaneous multi-protein labeling. Thus, it was shown that

concurrent dual-protein labeling could be achieved using PFTase and geranylgeranyltransferase-I (GGTase-I).[121] GGTase-I has a larger active site than that of PFTase. As a result, it catalyzes the transfer of the 20-carbon isoprenoid unit from geranylgeranyl diphosphate (GGPP) to the protein substrates. Different specificities are also observed for the CaaX motif. Combining these features, Zhang et al. showed that by selecting appropriate CaaX sequences and isoprenoid substrates, orthogonal dual protein modification could be achieved. They utilized this strategy to install ketone- and alkyne-functionalities into two model proteins respectively in one-pot without undesired side-products. Based on these results, it should be possible to expand this simultaneous labeling system to include engineered PFTase enzymes for potential *in vivo* multi-protein modification applications.

To implement protein immobilization using PFTase, Poulter and coworkers initially incorporated alkyne- and azide-containing isoprenoid analogues into GFP and GST engineered with a C-terminal CVIA sequence. The modified proteins were then oriented on a solid support using CuAAC and Staudinger ligation reactions.[116] In a follow-up study, they further demonstrated that the regioselectively immobilized GST retained its enzymatic activity.[258] In 2013, they constructed antibody arrays by immobilizing antibody binding proteins A, G and L onto glass slides using a similar strategy,[123,259] enabling antibody capturing for direct and sandwich-type immunofluorescent antigen detection. Besides glass slides, proteins labeled with alkyne isoprenoid analogues can also be deposited onto the azide-modified self-assembled monolayers prepared on gold surfaces using the CuAAC reaction, as illustrated by the Poulter[260] and Maynard[261] groups, respectively.

To capitalize on this highly site-specific protein modification strategy, Tretyakova and Distefano constructed structurally defined DNA-protein conjugates using PFTase.[122] An azide-containing isoprenoid analog was incorporated into GFP engineered with a C-terminal CVIA sequence and then linked to oligonucleotides via alkyne functional groups attached at the 5-postion of thymidine using the CuAAC reaction. The resulting conjugates served as analogues of the naturally occurring DNA-protein cross-links (DPCs) and were then used to investigate the effect of DPCs on the ability of human DNA polymerases to bypass these large lesions for the purpose of repair. It was demonstrated that lesions

containing an entire protein blocked all human polymerases tested while DPCs formed with a short peptide could be bypassed, suggesting that proteolytic degradation to remove the replication block imposed by the DPCs occurs prior to DNA repair.

To explore the therapeutic potential of PFTase-catalyzed protein modification, Kim and co-workers utilized enzymatic prenylation and oxime ligation to synthesize homogeneous protein-toxin conjugates for targeted therapy.[118] In this study, EGFR-binding repebody proteins, engineered with a C-terminal CVIM sequence, were labeled with ketone-containing isoprenoid analogues by PFTase and then conjugated to aminooxy-modified MMAF. Near quantitative labeling and subsequent conjugation were achieved. The resulting repebody-MMAF conjugates (RDCs) displayed much higher cytotoxicity to EGFR-positive cell lines compared to the free drug molecule while the toxic effect on EGFR-negative cells was negligible. A significant tumor regression response was observed when the RDCs were injected into xenograft mice bearing EGFR-positive tumors, emphasizing the therapeutic potential of RDCs. Moreover, this labeling strategy has been developed by LegoChem Biosciences, Inc to establish ConjuALL^{TM} platform technology, which utilized prenyltransferase to make site-specific antibody-drug conjugates, further demonstrating the potential of using PFTase to construct site-specific protein conjugates with therapeutic significance.[262]

Overall, the PFTase labeling reaction is highly selective and efficient. Importantly, the sequence requirement of only four residues minimizes potential perturbation of the target protein structure. It is especially well-suited for labeling of proteins whose C-terminus is far away from the site of action and when high homogeneity of the product is desired.

1.3.5 N-myrsitoytransferase.

N-myristoylation is an important protein-lipid modification in all eukaryotes, which can occur co-translationally or post-translationally.[263,264] The enzyme N-myristoyltransferase (NMT) catalyzes the transfer of myristate from myristoyl-CoA to the N-terminal glycine of protein substrates resulting in the formation of an amide linkage. The enzyme recognition sequence on the protein substrate is usually represented as GXXXS/T, where G is an absolute requirement and X can be a variety of amino acids.[264,265]

Comprehensive substrate specificity studies of yeast NMT conducted by Gordon and co-workers revealed that the enzyme has a relatively conserved myristoyl-CoA binding site, preferring analogues mimicking the structure of myristic acid concerning chain length and flexibility.[124,266,267]

To explore protein labeling using NMT, early work by Tate and co-workers utilized azide and alkyne-bearing CoA-myristic acid analogues to label both a synthetic model peptide and the *P. falciparum* ADP ribosylation factor 1 (*pf*ARF1) protein, which is a natural substrate for NMT.[126,268] Importantly, labeling of the protein substrate was performed in *E. coli* (Figure 1.10), where the target protein was co-expressed with *C. albicans* NMT. The bacterial culture was supplemented with either the azide- or alkyne-myristic acid analogues during protein expression. It should be noted that endogenous enzymatic activities in *E. coli* are critical to convert the analogues into their active CoA forms and to cleave the N-terminal methionine residue on the nascent proteins to expose the N-terminal glycine for enzyme recognition. A detailed protocol for the *in vivo* labeling procedure has been published by Heal et al.[269] Building upon that early work, a variety of applications have been demonstrated since 2013, which are summarized below.

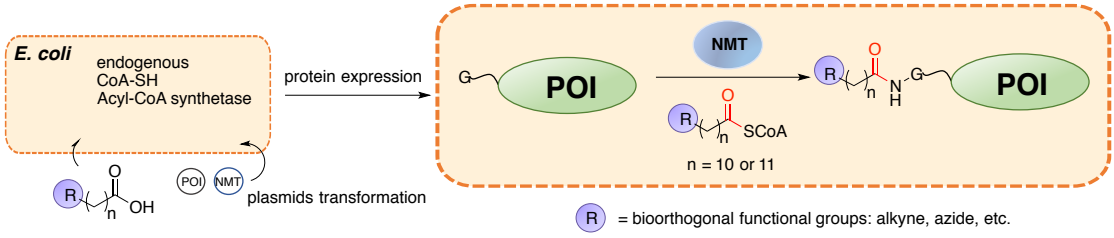


Figure 1.10 Protein labeling by NMT in *E. coli*. The plasmids expressing the POI and NMT are both transformed into *E. coli*. At the time of expression, myristic acid analogues are added to the culture medium, which are then converted to CoA-modified substrates by endogenous enzymatic activities. Proteins are labeled *in vivo* at the N-terminus.

Not limited to the native protein substrates for NMT, Tirrel and co-workers investigated the possibility of labeling recombinant proteins with engineered N-myristoylation motifs.[127] Two engineered GFP constructs, each carrying the N-terminal sequence from a natural myristoylated protein, were co-expressed with human NMT (hNMT) in *E. coli*. The azide-containing mystic acid analog (12-ADA) was added to the culture during protein expression and the resulting cell lysate was collected and incubated

with TAMRA-alkyne. Successful protein-fluorophore conjugation was confirmed using in-gel fluorescence imaging. Additionally, they were able to immobilize the myristoylated GFP onto a DBCO-functionalized solid surface using the SPAAC reaction directly from crude cell lysate. Utilizing the same labeling strategy, Tirrel and Ursem later modified an engineered Ca^{2+} binding protein, calmodulin (CaM), with azide functionality.[128] The engineered and myristoylated CaM maintained its binding capability to Ca^{2+} and was functionally active. In addition to the construction of CaM-TAMRA conjugates from cell lysate, CaM-affinity resins were also prepared by immobilizing CaM onto DBCO-functionalized resins, which demonstrated superior performance in purifying CaM-binding proteins. These examples demonstrate the feasibility of labeling non-natural substrate proteins with engineered N-terminal motifs by NMT.

Taking advantage of the *in situ* myristoylation in engineered *E. coli*, Ho et al. recently developed a facile method for protein imaging in bacteria.[125] A dual-plasmid system, one encoding hNMT and methionyl aminopeptidase and the other for expressing the protein of interest, were transformed into *E. coli*. A total of four target proteins with polar and septal spatial distributions were studied. The azide-containing myristic acid analog 12-ADA was supplied during protein expression. Fluorophore labeling experiments were first conducted in cell lysates where the bacterial cells were lysed and incubated with a BCN-modified BODIPY dye using the SPAAC reaction. Successful conjugation was confirmed by in-gel fluorescence analysis. Next, they performed protein conjugation in fixed cells by incubating the dye to fixed and permeabilized cells. Fluorescence was observed via confocal imaging for each target protein with expected localization . Finally, for protein labeling in live cells, cells were treated with the fluorophore reagent without fixation or permeabilization. The desired localization of each protein was again observed. A slightly more hydrophilic azide-containing myristic acid analog with an oxygen substitution at the C6 position was also studied in the lysate and live cell labeling experiments. Similar results were achieved comparable to those obtained with 12-ADA, suggesting that the toolbox of fatty acid analogues can be expanded by tuning the structure of the acyl chain.

Since the acyl-chain of myristic acid is highly hydrophobic, myristoylation has also been exploited to construct lipid-peptide hybrid biomaterials. Chilkoti and co-worker fused

the NMT recognition sequence to the elastin-like polypeptide (ELP).[270] ELP was co-expressed with yeast NMT in *E. coli* and was myristoylated in the presence of myristic acid. The purified lipid-modified ELP self-assembled into micelles with a hydrophobic core while the unmodified polypeptides remained as soluble unimers. The authors illustrated that the morphology of the self-assembled structures could be tuned by varying the length and composition of the ELP peptide. The presence of the hydrophobic core inside the micelles helped to promote physical encapsulation of hydrophobic drug molecules so that the micelles could serve as a drug delivery system. Although encapsulated doxorubicin (DOX) and paclitaxel showed decreased cytotoxicity in cell cultures compared to the free drugs, *in vivo* pharmacokinetic studies revealed a 6.5-fold higher serum half-life for the encapsulated DOX compared to the free drug. The simple generation of a self-assembled lipidated peptide-polymer drug delivery system directly from *E. coli* without additional labeling and purification steps greatly streamlines the production process. As a delivery system that functions through passive encapsulation, these micelles can be particularly useful for transporting drugs that lack reactive handles or in cases where chemical modification is detrimental to the bioactivity of the drug.

In summary, N-myristoyltransferase is a powerful tool for site-specific protein modification at the N-terminus. The ability to perform myristoylation *in vivo* in engineered *E. coli* without the need to purify the NMT enzyme markedly facilitates the modification process. Due to the lack of endogenous NMT in bacteria, *in situ* generation of modified myristoylated proteins can be especially useful for imaging cellular proteins and studying protein-protein interactions in live bacterial cells.

1.3.6. Phosphopantetheinyl transferases.

Phosphopantetheinylation is a process where a phosphopantetheine (Ppant) group from CoA is transferred to peptidyl (PCP) or acyl carrier protein (ACP) domains of several synthetases by phosphopantetheinyl transferases (PPTase).[271] These domains can be fused to proteins of interest for site-specific labeling. A serine residue in a conserved sequence motif is modified through the formation of a phosphoester bond with the hydroxyl group, along with a concomitant release of ADP (Figure 1.11). Surfactin phosphopantetheinyl transferase (Sfp) is the most utilized PPTase in enzymatic labeling strategies owing to its

known promiscuity in recognizing a wide range of modified phosphopantetheine CoA thioesters onto PCPs.[272] Similarly, acyl carrier protein synthase (AcpS), which recognizes an ACP domain, has also been used to label ACP-fused proteins with fluorescent tags[273] or immobilize them onto resin supports.[274] Although PCP and ACP have served well for labeling proteins as fusion constructs, they have relatively larger sizes (75 to 80 amino acids) compared to the tags employed in other enzymatic labeling approaches. To overcome that size limitation, Walsh and coworkers took advantage of phage display to screen for short peptide tags that can serve as replacements of PCP and ACP.[129] From a genomic library derived from *Bacillus subtilis*, an 11-residue peptide denoted as ybbR (DSLEFIASKLA) was found to be the best substrate for Sfp-catalyzed biotin-Ppant labeling. This tag can be incorporated onto the N- or C-terminus of proteins of interest or inserted into internal flexible loops, enabling efficient labeling in a single step (Figure 1.11B). Later, the same group developed peptide tags S6 (GDSLSWLLRLLN) and A1 (GDSLDMLEWSLM) that manifested higher efficiency and specificity for Sfp- and AcpS-catalyzed protein labeling, respectively.[275] The orthogonality between these two peptide tags towards their corresponding PPTase allowed sequential labeling of two different target proteins performed *in vitro* or in cell surfaces. Furthermore, NMR-based experiments on these short peptide tags identified six key amino acids critical for interaction with the enzyme.[276] That allowed the design of an 8-residue peptide, A4 (DSLDMLEW) containing 5 of the 6 key residues that displayed comparable efficiency to that of A1 in AcpS-mediated protein labeling.

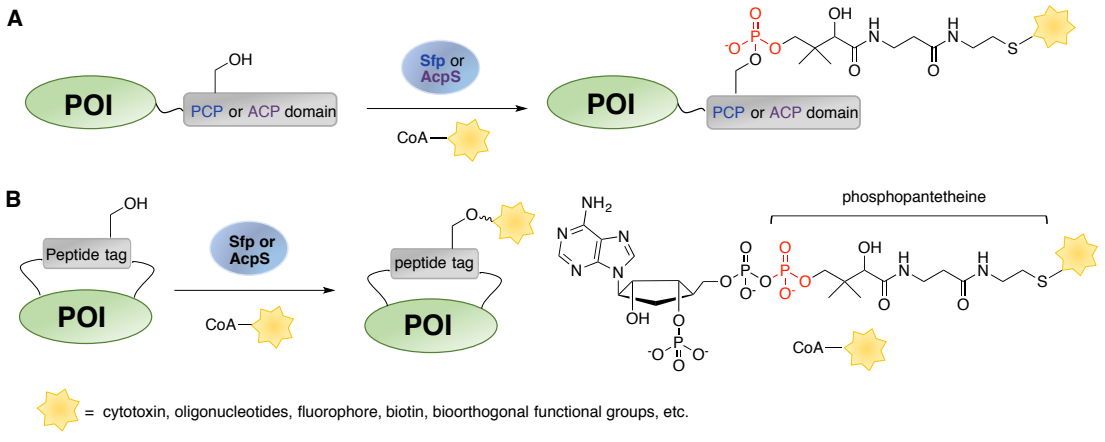


Figure 1.11 PPTase-catalyzed reactions for site-specific enzymatic labeling. (A) Proteins fused with PCP or ACP as recognition domains for modification of CoA derivatives. (B) Shorter tags (ybbR, S6, A1, and A4) can be inserted within exposed loops of the POI for internal modification.

Since 2014, the use of shorter peptide tags in the PPTase-catalyzed protein labeling method has led to the efficient preparation of numerous protein conjugates for biomaterials, therapeutic, and imaging applications. The cargo of interest can be directly coupled with CoA and transferred to target proteins in one-step without the need for secondary conjugation reactions. Thus, direct attachment of fluorophores, affinity handles, drugs or macromolecules to proteins of interest is possible. For example, Gaub and coworkers designed Protein-DNA chimeras through Sfp-catalyzed conjugation of ybbR-tagged GFP using CoA-modified DNA for single molecule analysis, enabling direct assembly of proteins via a Single-Molecule Cut & Paste technique (SMC&P).[131] In SMC&P, the cantilever of an atomic force microscope (AFM) was used to transfer protein-DNA hybrids from a disordered state in the depot site and rearrange them at the target site. The DNA in the protein conjugates serves as an anchor for the cantilever tip.[277] This technique allows for understanding complex networks of proteins that are spatially arranged on a platform. The covalent linkage between target proteins and DNA made from the Sfp-mediated approach improved protein-based SMC&P analyses. As a precise and efficient method, the utility of this protein-DNA coupling technique can be extended to the fields of nanobiotechnology and protein engineering. In the field of ADC construction, insertion of the peptide tags, ybbR and S6, into exposed loops of an antibody was explored to prepare a homogenous product.[130] The cytotoxin auristatin was coupled to CoA through maleimide chemistry and loaded onto trastuzumab via Sfp-mediated reaction. Interestingly, it was found that conjugation on the CH1 domain resulted in favourable pharmacokinetics and minimal drug loss while attachment on the CH2 domain generally produced thermally unstable conjugates that resulted in rapid clearance. When various regions in CH1 were evaluated as modification sites, the resulting ADCs displayed similar *in vitro* potencies. In addition to being highly potent and selective *in vitro*, substantial tumor regression was also observed in a mouse xenograft tumor model. The versatility of this approach provides precise control on the site and stoichiometry of drug attachment in ADCs.

Protein modification with fluorescent reporters or affinity handles have been essential to study PPIs. Handel and coworkers fused an S6 tag to the C-terminus of several chemokines, followed by conjugation to fluorophore-containing CoA derivatives.[278]

Previous experiments using maleimide chemistry to prepare chemokine-fluorophores via genetically introduced cysteine residues led to protein aggregation and thus low yields.[279] In contrast, PPTase-mediated labeling of three chemokines (CXCL12, CCL2, and CCL21) provided conjugates in high yields and allowed their interaction with the respective receptors to be studied using flow cytometry and cellular imaging. Catteneo and coworkers incorporated different PPTase recognition tags into neurotrophin and its receptors.[280] Using a mutagenesis strategy, successful insertions of the A4 tag into nerve growth factor (NGF), as well as A1 and S6 into either tropomyosin kinase receptor A (TrkA) or p75 neurotrophin (P75NTR) receptor were achieved. It was shown that the purified NGF could be biotinylated with CoA-biotin catalyzed by AcpS and remain functional. To image the two different receptors on the surface of a living cell, a sequential dual-colour staining procedure was performed. Cells transfected with the two orthogonally-tagged receptors were first treated with CoA-biotin and AcpS, followed by incubation with streptavidin-quantum dot S-QDot 525. Then a second botinylation using Sfp was performed followed by incubation with S-QDot 655. Distinct colours were observed under confocal imaging, indicating successful labeling. The ability to dually label two cell surface proteins could be potentially applied to study protein-receptor interactions involved in signal transduction pathways.

The combination of PPTase-based labeling with other enzymatic labeling methods has allowed for the development of FRET-based reporters. Schwarzer and coworkers combined PPTase- and SrtA-mediated labeling techniques to simultaneously incorporate GST and mCherry into a synthetic peptide bearing a fluorescein (FRET pair to mCherry).[281] In another dual labeling strategy, Geierstanger and coworkers genetically encoded a C-terminal S6 tag and an internal pyroline-carboxy-lysine (Pcl) residue, in IgE to allow conjugation to different fluorophores for FRET study via enzymatic Sfp labeling and chemical oxime ligation reaction, respectively.[282] As a UAA known to be specifically reactive with a wide range of 2-aminobenzaldehyde-functionalized reagents,[283] Pcl was genetically encoded in place of a surface-exposed proline. Based on the crystal structure of IgE, this site is suitable for FRET-based studies as it is close to the C-terminus in the open state (apo) but moves away upon binding the specific IgE receptor (holo state).[282] After purification, the protein was dually labeled with Alexa Fluor 488 (donor)-CoA and 2-

aminobenzaldyhe-Alexa Fluor 594 (acceptor). Efficient labeling with no cross-reactivity between the enzymatic and chemical modifications was achieved in a one-pot manner under near physiological conditions. The probe acted as a good FRET biosensor for evaluating IgE binding with its high-affinity receptor FcεRI, a key event involved in the allergic response cascade. A decrease in FRET efficiency was observed that correlated with increased intra- and inter-chain distances during the binding event *in vitro*. Since the Pcl and S6 tag modified proteins of interest can be expressed in mammalian cells, this method could be useful for site-specific dual labeling of complex proteins that are resistant to recombinant expression in bacterial hosts.

The development of PPTase-mediated protein labeling in concert with the discovery of shortened enzyme recognition tags has greatly facilitated the preparation of functional protein conjugates. In contrast to other enzymatic labeling strategies that are restricted to either termini of a protein, labeling by PPTase is more versatile with regards to the modification site within the protein sequence. However, limitations on solubility and cell permeability of the synthetic CoA analogues may compromise its application in *in vivo* studies. Chemoenzymatic approaches have been explored in site-specific protein labeling of PCP- and ACP-fused proteins *in vivo* using smaller synthetic substrate analogues.[284,285] In this approach, the analogues are synthesized as D-pantetheine derivatives, which are then metabolically converted to CoA-analogues inside the host, catalyzed by three promiscuous enzymes involved in the CoA biosynthetic pathway: pantothenate kinase (CoAA), phosphopantetheine adenylyltransferase (CoAD), and dephospho-CoA kinase (CoAE).[286] Recently, an Sfp mutant was employed that could accept substrate analogues without the need for CoA phosphorylation by CoAE, which is the problematic step in the CoA biosynthetic pathway.[287] *In vivo* labeling with PPTases through chemoenzymatic methods has thus far been limited to proteins fused with large PCP and ACP domains. Extending this method to labeling proteins inserted with the short peptide tags has yet to be explored but should prove fruitful.

1.3.7. Tubulin tyrosine ligase.

In eukaryotic cells, tubulin tyrosine ligase (TTL) catalyzes the ATP-dependent addition of a tyrosine residue to the C-terminus of α-tubulin forming a native peptide bond

(Figure 1.12A).[288] Together with an uncharacterized carboxypeptidase, the C-terminus of α-tubulin is subjected to a post-translational process where a tyrosine residue can be reversibly added and removed, which is important in regulating microtubule function.[289] TTL suppression has been reported to be associated with progression of neuroblastoma,[290] and even in normal cells, TTL has been shown to play a vital role in controlling neuronal organization.[291]

Soon after the discovery of this enzyme, it was found that TTL-mediated labeling of α-tubulin is not limited to the native tyrosine substrates, but can be extended to functionalized tyrosine analogues, which prompted biotechnological applications of TTL. 3-Fluorotyrosine was one of the first tyrosine-derivatives found to be accepted by the enzyme.[292] Enzymatic incorporation of a fluorine into the target protein allowed the use of ^{19}F NMR to study the PPIs that occur near the C-terminal region of α-tubulin. While studying the effects of nitrotyrosination on tubulin, Arce and coworkers found that nitrotyrosine, which is naturally present in cells, could be reversibly incorporated into tubulin by TTL and removed by the carboxypeptidase without detrimental effects to cells.[293] Years after this report, Bane and coworkers were able to extend the range of tyrosine analogues for protein labeling on live cells.[294] By measuring the degree of inhibition of TTL-mediated [^{3}H]-L-tyrosine incorporation into tubulin, 3-formyltyrosine and 3-azidotyrosine were identified as acceptable substrates with efficiencies lower than the native tyrosine but comparable to that of 3-nitrotyrosine. Since the use of azides often involves the CuAAC reaction which requires the use of copper that may have negative effects on tubulin assembly, initial efforts focused on 3-formyltyrosine. When reacted with a synthetic coumarin-hydrazide, the hydrazone formation reaction with the 3-formyltyrosine molecule was completed within 120 min with a second-order rate constant of 53 $M^{-1}min^{-1}$. Interestingly, hydrazone formation through a covalent linkage led to a red shift of both the absorption and emission maxima of the coumarin, as well as an increased quantum yield, which allowed the conjugates to be detected even in the presence of unreacted fluorophore. This unique bathochromic shift was useful in later cell labeling studies since it obviated the need for medium exchange (to remove unreacted probe) after incubating cells with fluorophore-probes. Importantly, fluorophore labeling using this method did not hinder polymerization of tubulin and was found to be a suitable method for

live cell imaging. Incubation of coumarin hydrazine with CHO cells grown in 3-formyltyrosine containing medium led to strongly fluorescing cells. Western blotting analysis confirmed that the probe was selectively linked to α-tubulin.

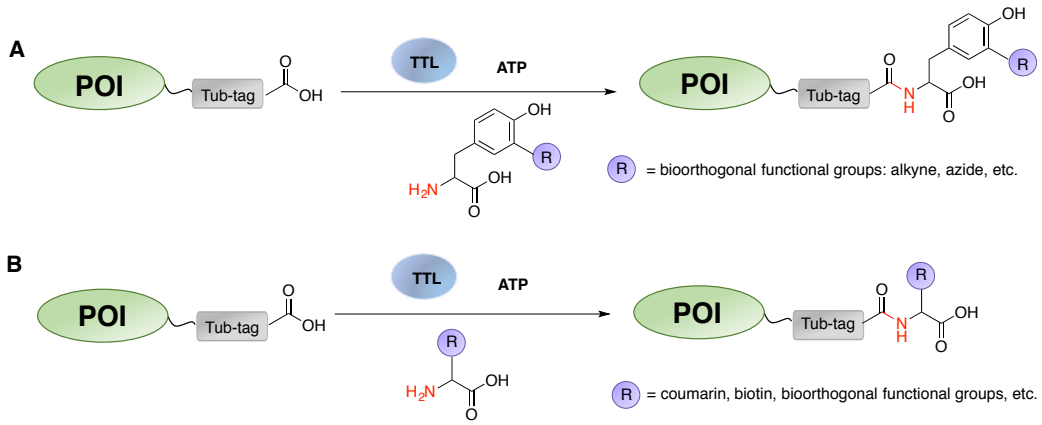


Figure 1.12 TTL-mediated attachment of (A) tyrosine analogues and (B) functionalized glycines to the POI engineered with a C-terminal Tub-tag.

Recently, to expand the scope of TTL-catalyzed labeling of proteins other than tubulin, Leonhardt and coworkers set out to find a sequence that can be efficiently recognized by the enzyme (Tub-tag).[67] A 14 residue Tub-tag peptide which mimics the C-terminal sequence of tubulin was initially discovered to be an acceptable substrate using 3-L-formyltyrosine or 3-N_3-L-tyrosine and a recombinant TTL. For protein labeling studies, two types of GFP-specific nanobodies (GBP1 and GBP4), GFP, and ubiquitin were prepared, all containing the Tub-tag sequence at the C-terminus. Importantly, the insertion of Tub-tag did not affect the fluorescence of GFP, and the engineered nanobodies retained their binding ability, indicating minimal perturbation of the protein structure and function. As for the tyrosine substrate, the azide- and aldehyde-bearing analogues were shown to be incorporated by TTL, which enabled labeling of biotin or fluorophores through bioorthogonal reactions including SPAAC, Staudinger ligation, and hydrazone- and oxime-forming reactions. The two-step modification method was also shown to be highly selective even when the labeling was performed in *E. coli* lysate. To demonstrate the utility of TTL-mediated ligation towards enrichment and isolation of target proteins from crude lysate, GBP1-biotin was prepared by sequential incubation with 3-N_3-L-tyrosine and DBCO-biotin, followed by immobilization on streptavidin-coated magnetic beads. Western blotting was used to confirm specific GFP pulldown from HEK cell lysates.

Applicability for immunostaining was highlighted using GBP1 labeled with Alexa Fluor 594 (GBP1-Alexa594) via TTL-mediated ligation and oxime formation. HeLa cells expressing GFP-LamininB1 that localizes in the inner membrane of the nucleus was fixed, permeablized and stained with GBP1-Alexa594. Super-resolution microscopy showed colocalization of the GFP and Alexa Fluor 594 signals, indicating successful binding of TTL-modified nanobodies to GFP in the intracellular environment.

More recently, Hackenberger and coworkers explored the substrate scope of TTL. Exploiting the plasticity of the TTL active site, they were able to achieve one step protein labeling with biotin and fluorescent coumarin using the wild-type enzyme without any necessary mutation in the active site.[133] Initial labeling efficiencies tested on a carboxyfluorescein-labeled C-terminal Tub-tag peptide showed that while ortho-substituted tyrosine derivatives become fully ligated, substrates with para-substituents including an azide or alkyne are not well-recognized by the TTL. For hydrophobic amino acids, labeling was found to be especially efficient for those bearing aromatic rings. Importantly, the carboxylic acid group of the substrate amino acid was found to be essential for enzyme recognition. Based on these observations, the TTL activity towards functionalized amino acids was tested (Figure 1.12B). Fluorescent molecules including coumarin-functionalized glycine and β-(1-azulenyl)-L-alanine were found to be acceptable substrates. A biotinylated tyrosine that contained an ethylene glycol linker to provide flexibility and liberation from the spatial constraints of the active site, was also found to be successfully ligated. To understand this substrate preference of TTL, two types of computational studies were carried out. Both studies led to a conclusion that interactions in the active site of enzyme stabilized ortho-substituted tyrosines due to spatial conditions as well as hydrophobic molecules with ring structures due to π-interactions. An extended channel extending out of the active site was also observed, which was consistent with the experimental result where labeling was accomplished with the biotin-functionalized tyrosine analogue. Based on this insight, one-step fluorescent labeling was performed on various functional proteins engineered with a Tub-tag at their C-termini. Incubation of cells expressing GFP-fusion proteins with GBP-coumarin showed colocalized signals of GFP and coumarin, demonstrating the utility of the one-step labeling method for cellular imaging. Additionally, annexin V, an apoptosis marker protein, was labeled with coumarin

through the one-step process and showed comparable results for visualizing apoptotic cells to those obtained using commercial probes. This demonstrated the potential of this method to improve the performance of currently available probes, which are partially limited by their product heterogeneity that results from non-specific modification.

In summary, TTL has been applied to site-specifically modify proteins with various functionalities, allowing a broad range of bioorthogonal reactions to be selected from for conjugation to additional cargos. Especially, TTL has started to regain interest due to the recent findings that the wild-type enzyme itself can accept a broad range of substrates with acceptable kinetics.[133] Understanding the mechanism that governs the substrate-enzyme interactions enables one-step TTL-mediated labeling of functional probes to the protein of interest with various functional probes, making it a powerful tool. Combination with enzyme engineering is expected to allow labeling of an even broader range of substrates to facilitate novel applications.

1.3.8. Lipoic acid ligase.

Lipoic acid ligase (LplA) recognizes a specific LplA acceptor peptide (LAP) and catalyzes the attachment of a lipoate moiety to a lysine residue in LAP through an ATP-dependent reaction (Figure 1.13), which was first recognized by Koike and coworkers.[295] LplA was later found to be capable of recognizing lipoate analogues,[296,297] which prompted studies focused on using this enzyme for site-specific labeling. Members of the Ting lab initially showed that azide-containing analogues could be successfully incorporated into the LAP peptide using LplA. Fusion of the LAP peptide to cell-surface proteins, followed by labeling with LplA using the azide analogues and subsequent conjugation with cyclooctyne-probes allowed selective protein imaging of live cells.[134] While initial studies required a 22 amino acid LAP peptide for LplA recognition, further studies employing yeast surface display selection led to the discovery of a 13-residue LAP, GFEIDKVWYDLDA, that has higher catalytic efficiency and reduced capacity for protein structure disruption, which therefore has been used as the standard LAP since then.[298]

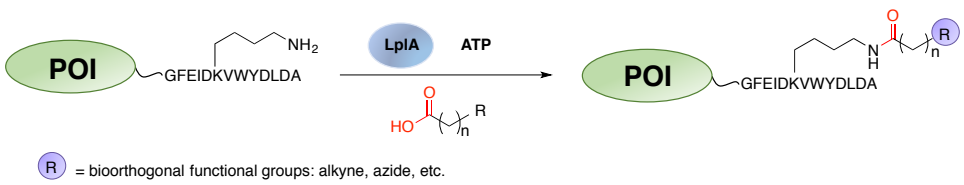


Figure 1.13 LplA-mediated protein labeling using lipoate analogues bearing various bioorthogonal functional groups to a POI containing an engineered C-terminal LplA acceptor peptide.

Since 2013, the scope of lipoate analogues employed has been significantly broadened, leading to a much wider range of compatible biorthogonal reactions that can be used for LplA-mediated labeling. Wombacher and coworkers were able to attach *p*-iodophenyl derivatives to peptides or proteins bearing the LAP motif using LplA, followed by palladium catalyzed Sonogashira cross-coupling between the labeled aryl halide moiety and an alkyne-probe to prepare site-specific protein-fluorophore conjugates.[136] Since these analogues are bulkier than the native lipoate substrate, four mutant LplA constructs were prepared with an enlarged binding pocket. In accordance with previous reports, mutation of W37I was found to be the most crucial mutation for the enzyme to accept analogues of lipoate. Using the LplA W37V mutant ($LplA^{W37V}$), a two-step labeling was performed on dihydrofolate reductase from *E. coli* (eDHFR) which contains an N-terminal LAP to create fluorescently labeled form of DHFR. However, non-specific labeling of the alkyne-probe on halide-free DHFR was also observed, which was likely due to palladium catalyzed thiol-yne coupling. Although this could be prevented by blocking all free thiols with excess N-ethylmaleimide, development of new palladium catalysts with higher selectivity is desirable. More recently, a series of norbornene-bearing lipoate analogues was also developed, which can be used in tetrazine ligation reactions.[94] A small library of analogues was prepared to select the most efficient substrate. Interestingly, overall conformational flexibility was found to be as important as the length of the derivatives in determining enzyme reactivity. Using the $LplA^{W37V}$ mutant enzyme, tetrazine-fluorescein was successfully conjugated to the norbornene-labeled eDHFR containing an N-terminal LAP-tag, in purified form or when diluted into cell lysates. Live-cell labeling was also achieved with HEK293T cells that were transiently transfected with pDisplay-LAP2-CFP-TM fusion protein. After expression, the construct localized to the plasma membrane exposing the extracellular LAP-tag and was then modified by $LplA^{W37V}$. Incubation with tetrazine-

TAMRA showed colocation of the cyan fluorescent protein (CFP) signal and TAMRA fluorescence with minimal background, highlighting the site-specific advantages of LplA-mediated labeling and the copper-free conjugation conditions of tetrazine ligation. More recently, an analogue containing ^{18}F was also developed, enabling one-step modification of an Fab fragment, modified by LAP-insertion, with radio-isotopes that can be potentially useful for diagnostic imaging.[299]

Recently, in addition to modifications of protein termini, labeling by LplA has been developed to functionalize internal sites, even simultaneously at multiple sites.[135] In the first report of exploiting LplA for labeling internal sites by inserting LAP into loop regions, a series of GFP constructs was prepared. When the expression level of different constructs was compared, it was found that the LAP-containing GFPs were expressed to a comparable extent to that of the wild-type GFP, as evaluated by the normalized intrinsic fluorescence of GFP in crude *E. coli* lysate. More importantly, LAP-constructs showed higher expressions than GFP containing two azide-functionalized UAA residues (AzF-GFP). The superior expression levels observed underscores the advantages of protein labeling using enzymatic methods over the incorporation of UAAs, especially when modification of multiple sites is desired. After incorporating the 10-azidodecanoic acid into GFP constructs using $LplA^{W37V}$, subsequent PEGylation was performed by SPAAC using DBCO-PEG (M_w 5 kDa), whose conversion yield, and product homogeneity all exceeded that of non-specifically PEGylated wild-type GFP. The utility of the method was further demonstrated by glycosylation, fatty acid modification, and surface immobilization of GFP. Interestingly, immobilization through the internal site indicated that a unique uniform orientation was obtained, which was different from those labeled at the termini.[135] This opens up the possibility of creating protein chips that require particular orientations of proteins, which are not accessible by labeling through either terminus.

McNaughton and coworkers recently applied LplA-mediated labeling for activation immunotherapeutics by enzymatically linking an antibody recruiting domain (ARD) to a cell binding domain (CBD).[68] An aldehyde-bearing lipoate analogue was incorporated into a HER2 binding nanobody (denoted as 5F7) that contains an N-terminal LAP using $LplA^{W37I}$. Dinitrophenyl (DNP)-functionalized hydrazine was then conjugated to form a chimeric construct where the DNP acts as an ARD, and the 5F7 a CBD. Successful target

binding was confirmed by confocal imaging with SK-BR-3 cells that overexpress HER2. To test the therapeutic ability of DNP-5F7, SK-BR-3 cells were incubated with DNP-5F7, anti-DNP antibodies, and peripheral blood mononuclear cells (PMBCs), which induced cytotoxicity when they interact with antibodies. Cell death was shown to be selectively induced on HER2 expressing cells with good potency (EC_{50} 60 nM).

In summary, a variety of applications have been achieved using LplA-mediated enzymatic protein labeling followed by subsequent bioorthogonal reactions. Particularly, recent developments that allow site-specific labeling of internal regions of a target protein opens up new possibilities for situations where N- or C-terminal labeling is less desirable.

1.3.9. Biotin ligase.

Biotin ligase catalyzes the conjugation of biotin derivatives onto proteins in an ATP-dependent manner. In their seminal report, Ting and coworkers used an *E. coli* biotin ligase (BirA) to label biotin-derivatives to target proteins through the lysine residue of an acceptor peptide (AP), which consists of 15 amino acids (GLNDIFEAQKIEWHE), inserted on to the N- or C-terminus of a protein (Figure 1.14A).[138] By exploiting a ketone moiety in the conjugated biotin, they were able to attach a benzophenone-hydrazide to cyan fluorescent protein-AP using oxime chemistry. Protein labeling on the surface of live cells was also shown, where the biotinylated target proteins were visualized using streptavidin-Alexa Fluor 568. In a subsequent report, Ting and coworkers expanded the range of tolerable substrates by screening various biotin analogues against biotin ligases from nine different microorganisms.[139] *Pyrococcus horikoshii* biotin ligase (*Ph*BL) was found to promote the ligation of azide- and propargyl-functionalized biotin analogues when evaluated on a purified domain of an endogenous biotin acceptor protein (p67). Using the Staudinger ligation, azide-p67 could be conjugated to a phosphine-FLAG peptide. The ability to functionalize target proteins with analogues bearing bioorthogonal moieties highlights the utility of biotin ligase for the construction of protein conjugates.

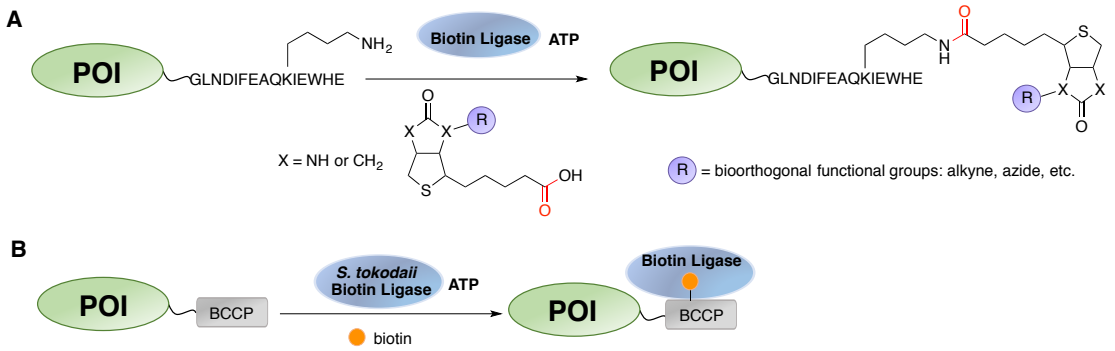


Figure 1.14 Biotin ligase mediated-protein labeling. (A) Protein labeling using biotin ligase with biotin analogues that contain bioorthogonal functional groups. B. Biotin ligase from *S. tokodaii* forms a complex with its biotin-modified protein substrate bearing the BCCP domain.

A unique advantage of biotin incorporation is that it can spontaneously associate with streptavidin with extremely high affinity, which allows conjugation to be performed under protein- and cell-compatible conditions.[300] Taking advantage of this association, the biotin ligase labeling technique has been applied to study PPIs. In work performed by Ting and coworkers, two protein constructs were created, one target protein fused to BirA, and the other one modified with an enzyme recognizable AP. The interaction between these two proteins was gauged by measuring the extent of biotinylation on the AP-protein using a streptavidin-probe.[301] As expected, PPI between a well-studied protein pair of FKGP and FRB, which dimerizes in the presence of rapamycin, was detected by incubation of FKGP-AP and FRB-BirA with biotin and rapamycin, followed by streptavidin blotting. However, due to the high concentrations of FKGP and FRB in live cells that nearly reach the K_M of AP (25 μM) for BirA, PPI was detected independent of rapamycin in live HEK cells. Therefore, the AP was truncated at the C-terminal end to reduce its affinity with BirA so that it would only be modified when in close contact with the enzyme. This strategy led to a successful detection of rapamycin-dependent biotinylation. Evaluation of a phosphorylation-dependent PPI between Cdc25C phosphatase and 14-3-3ε phosphoserine/threonine binding protein was performed to demonstrate the general utility of the method. Incubating mCherry-AP-Cdc25C constructs with BirA-14-3-3ε in the presence of biotin, followed by the addition of streptavidin-Alexa Fluor 647 showed colocalized signals of the two fluorophores, indicating the successful detection of PPI. Mutation of Cdc25C that blocked phosphate binding led to the disappearance of the Alexa

Fluor 647 signal, confirming that the visualized PPI was indeed a result of Cdc25C and 14-3-3ε interaction.

The biotinylation by BirA is a two-step process, in which biotin and ATP first form a biotinoyl-5'-AMP intermediate, which remains inside the enzyme active site until it reacts with the lysine residue of the AP. By mutating BirA so that it releases the highly reactive biotinoyl-AMP species into its surroundings, Burke and coworkers were able to induce proximity-dependent biotinylation of lysine residues on proteins neighbouring the BirA-fused target protein, enabling PPI identification in the native cellular environment.[302] This technique was denoted as BiolD. To illustrate its applicability, a genetic fusion of the mutant BirA to lamin-A protein was created, which allowed identification of proteins in close contact with lamin-A in live HEK cells. More recently, Roux and coworkers improved the proximity-identification method by developing a smaller enzyme that led to more selective and efficient labeling of neighbouring proteins (BiolD2 approach).[303] As the smallest biotin ligase known to date, the enzyme from *Aquifex aeolicus* (*Aa*biotin ligase) was employed. To make it usable for proximity-dependent biotinylation, the enzyme was humanized and mutated within the biotin catalytic domain (R40G) to release the reactive intermediate. The engineered *Aa*biotin ligase mutant was superior to the previously employed BirA enzyme, as it could be used under a wider range of temperatures and detect proteins that were otherwise difficult to identify before. Additionally, it required less biotin substrate due to its higher efficiency. It was also shown that the "biotinylation distance" (the distance between the site where the activated biotin species is generated and the site of lysine modification) could be increased by inserting a flexible peptide linker between the protein being studied and the fused biotin ligase.

Site-specific protein labeling has been demonstrated using biotin ligases functionalized with various probes, which is distinctive from other enzymatic labeling methods in that the enzyme itself becomes covalently attached to the target of interest (Figure 1.14B). Hayashi and coworkers developed a labeling method that involved biotin ligase from *Sulfolobus tokodaii* (*St*biotin ligase), which forms a stable complex with the biotinylated form of biotin carboxyl carrier protein (BCCP). The *St*biotin ligase first catalyzes the biotinylation reaction of BCCP and then associates with the resulting product. Exploiting this unique property, Hayashi and coworkers used *St*biotin ligase for

fluorescence imaging of cells.[140] Fluorophore-labeled biotin ligase was prepared by mutating R152 to a cysteine, followed by reaction with a maleimide-functionalized fluorescein or DyLight549. In the presence of biotin, this modified *St*biotin ligase was then attached to the bradykinin B2 receptor (B2R), a membrane protein expressed in HEK293 cells, that had a truncated form of BCCP (69 residues) appended onto the N-terminus. Visualization of the cell surface was shown by confocal microscopy. To demonstrate the utility of the method for labeling inside cells, B2R that contained a C-terminal BCCP was coexpressed with GFP-*St*biotin ligase in HEK293 cells. GFP signal was observed on the luminal side of the plasma membrane upon biotinylation, which indicated successful labeling.

Recently, the complex formation of *St*biotin ligase and BCCP was utilized to facilitate oriented immobilization of antibodies onto the surface of a sensor chip.[141] A sequential procedure was used, starting with immobilization of *St*biotin ligase onto a gold surface. To make the BCCP-interaction site accessible, N151 and R152 located on the opposite side of biotin ligase were replaced with cysteine, which was adsorbed onto the surface via thiol-gold interactions. Next, either one or two copies of the synthetic IgG-binding domains from protein A, Z-domain, was fused to the N-terminus of BCCP and was immobilized onto the gold surface upon biotinylation. The *St*biotin ligase-BCCP complex positioned between the Z-domain and the surface acted as a rigid spacer to promote antibody capture. The efficiency of the prepared sensor chips to capture antibodies was evaluated by monitoring changes in frequency from the sensor readout as anti-GFP antibodies were introduced followed by GFP. It was shown that the frequency change caused by GFP rose from specific antigen-antibody interactions, which demonstrated the successful capturing of anti-GFP antibodies by the immobilized Z-domain. Along with a dose-dependent response to GFP, the designed sensor showed high sensitivity that allowed detection of concentrations as low as 0.1 nM GFP, which was impossible to achieve when Z-domains were directly attached to the surface via cysteines.

Overall, some features of biotin ligase are unique compared to other enzymatic labeling systems. Instead of only serving as a catalyst, the enzyme can be more involved in the reaction. For example, it can be fused to the protein of interest or form a complex with its biotinylated protein substrate. While most enzymes focus on improving the

substrate specificity, the studies described above have often involved mutating biotin ligase to render it more promiscuous so that labeling is not sequence-specific but distance-dependent for proximity sensing. Given these unique features, it is likely that more applications will be reported in the future.

1.3.10 Formylglycine generating enzyme.

Formylglycine generating enzymes (FGEs) catalyze the post-translational modification of type 1 sulfatases by oxidation and subsequent hydrolysis of the thiol moiety of cysteine within the CXPXR motif to an aldehyde-bearing formylglycine (FGly) (Figure 1.15). The enzyme was first identified by Figura and coworkers when studying diseases related to sulfatase deficiency.[304] Subsequently, Bertozzi and coworkers exploited FGE for site-specific protein modification.[64] In their pioneering work, two types of aldehyde tags were tested on three different recombinant proteins expressed in *E. coli*. One of the tags was a full-length sulfatase motif consisted of 13 amino acids (LCTPSRGSLFTGR) and the other a shorter tag (LCTPSR) which contained only the core conserved residues. Aldehyde formation on the cysteine residue inside the tag was shown to occur regardless of whether the tag was present at either the C- or N-terminus of the protein substrate. Subsequently, site-specific protein modification was performed by reacting an aminooxy-functionalized affinity tag, fluorophore or PEG with the formylglycine residue, demonstrating the generality of the FGE-mediated labeling technique. In a follow-up study, peptide library screening was conducted to identify new aldehyde tag sequences that would minimize any perturbation of the target protein structure.[305] Noncanonical sequences including LCTASR and LCTASA were found to be recognized to a similar degree compared to the canonical sequence of LCTPSR by *Mycobacterium tuberculosis* FGE (*Mt*FGE) *in vitro* and by endogenous FGE-like activities in *E. coli*. Soon after, Bertozzi and coworkers expanded their methodology of protein labeling in prokaryotic cells to mammalian cells,[142] which was an important improvement since many therapeutic proteins are expressed in mammalian cells due to the need for them to be post-translationally modified. They showed that an IgG Fc fragment expressed in CHO cells showed conversion of cysteine to formylglycine by endogenous FGE, with efficiencies ranging from 25-67% by having a 13- (LCTPSRAALLTGR) or 6-residue tag (LCTPSR) appended at either the N- or C-

terminus. The efficiencies were significantly improved by coexpressing human FGE along with the engineered Fc proteins, indicating that insufficient levels of endogenous FGE relative to the overexpressed Fc proteins was a limiting factor. Multiple proteins bearing the aldehyde tag were shown to be modified selectively using this method. Overall, in the early development stages of FGE-mediated labeling, broadening the range of aldehyde tags, exploring different protein expression and labeling systems, testing various protein species, and confirming the selectivity of the subsequent conjugation reactions with the inserted aldehyde moiety were the major focuses, laying a solid foundation for later biotechnological development.

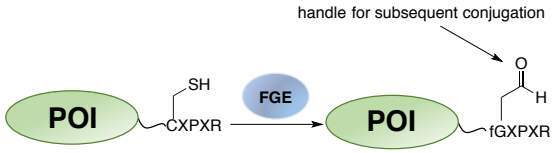


Figure 1.15 FGE catalyzed conversion of a cysteine to an aldehy-bearing formylglycine residue. Labeling occurs in E. coli by coexpression of the tag-fused POI and FGE. The aldehyde functional group serves as a reactive handle for subsequent conjugation.

Site-specific protein glycosylation is important for understanding the role of glycosylation in many key biological processes and developing protein-based therapeutics. Bertozzi and coworkers were able to apply FGE-mediated labeling for site-specific protein glycosylation, by inserting an aldehyde tag into the target protein, followed by oxime formation with complex aminooxy-glycans that were synthesized through a novel synthetic route.[306] A total of three complex glycans were synthesized. To label these complex glycans, a formylglycine containing human growth hormone (hGH) was produced in *E. coli* by coexpression with *Mt*FGE. Slightly increasing the acidity of the reaction buffer was found to improve the oxime ligation yield to allow more efficient coupling of the glycans to the target proteins without noticeable detrimental effects on either reactant. Since glycosylation sites in most proteins are located at internal positions instead of at the termini, Bertozzi and coworkers explored the labeling by FGE on proteins engineered with the internal enzyme recognition motif.[307] Three recognition motifs were evaluated and incorporated into an Fc fragment. Comparing the expression levels of each construct in CHO cells and subsequent conjugation efficiencies with aminooxy-Alexa Fluor 488, the

CTPSR motif was selected as it minimized effects on protein structure and had the highest modification level (76%). After optimizing the reaction conditions using O-benzylhydroxylamine, near quantitative oxime conjugation was obtained.

Over time, improved conjugation chemistry for aldehyde moieties has been developed, thereby increasing the power of FGE-mediated labeling. Oxime and hydrazone chemistries are the most commonly used methods for aldehyde-bearing proteins. However, the resulting C=N linkage favours acidic reaction conditions and is prone to hydrolysis, which limits the utility of these reactions to create conjugates intended to be used in biological systems that require long-term stability at physiological temperatures. Bertozzi and coworkers recently developed a modified Pictet-Spengler reaction (Scheme 1B), which provides ligation products with improved stability over conventional aldehyde-involving linkages.[61] In a canonical Pictet-Spengler reaction, a tryptamine derivative reacts with an aldehyde to form a stable C-C bond. However, the reaction rates were very slow under protein-compatible conditions. To overcome this limitation, the aliphatic amine of tryptamine was replaced with a methylated-aminooxy moiety and was moved to the 2-position of indole to expose the more nucleophilic 3-position for subsequent reaction. The modified reaction, denoted as the Pictet-Spengler ligation, manifests rates that are 4-5 orders of magnitude faster than the conventional method in aqueous media at pH 4.5. The ligated product was shown to be stable after 5 days of incubation under aqueous conditions at pH 4.5 or 5.0. In contrast, the majority of oxime-linked reagents hydrolysed within 2 days. An anti-HER2 human IgG conjugated to Alexa Fluor 488 (AF488-α-HER2) was then prepared by combining FGE-mediated labeling and Pictet-Spengler ligation, resulted in fluorophore-labeled antibodies with preserved affinity for HER2, illustrating the applicability of this chemistry for modifying functional biomolecules.

In a subsequent report, Agarwal et. al introduced the hydrazino-Pictet-Spengler (HIPS) ligation, which was a further improved version of the Pictet-Spengler reaction, displaying fast reaction rates even under neutral conditions with long term hydrolytic stability.[62] In this method, an alkylhydrazine-functionalized indole was used as the nucleophile that reacted with an aldehyde, which allowed rapid reaction rates at near neutral pH (pH 6.0). This new reaction was also utilized to construct antibody-fluorophore conjugates. Notably, when incubated in human plasma, the HIPS-conjugated protein

showed no hydrolysis as verified by monitoring the loss of Alexa Fluor 488 fluorescence, while conjugates employing oxime-linkages deteriorated in a single day. More recently, a new carbonyl ligation method termed the trapped-Knoevengael ligation with faster reaction rates at pH 7 was introduced by Kudirka et al., further expanding the reaction scope that can be coupled to the FGE labeling technique.[63]

The combination of FGE modification and HIPS chemistry was used to study the effect of DAR and the conjugation site on the *in vivo* efficacy and pharmacokinetics of ADCs.[308] Several modification sites were first evaluated, where three of them (the light chain, CH1 domain, and heavy chain C-terminus) were identified to minimize immunogenicity and aggregation. Benchmarking the clinically approved drug T-DMI, the antibody trastuzumab was site-specifically labeled with the cytotoxin DM1 at the three chosen sites using a combination of FGE modification and HIPS chemistry. Interestingly, while the fluorophore conjugates were shown to be stable, the DM1 conjugates were shown to lose their payload over time, to a varying degree depending on the conjugation site. This difference in stability was attributed to the labile ester bond that connected the DM1 payload to the aldehyde-reactive indole. Different efficacy as well as stability was observed when the conjugates were evaluated in tumor-bearing mouse xenografts. These results indicated that the distinct microenvironments at each site have distinct influences on the performance of the corresponding ADC. Importantly, mice treated with the site-specifically created ADCs overall showed better improvements in mortality, body weight loss, liver toxicity and tissue damage in comparison with control constructs prepared by conventional lysine chemistry, emphasizing the value of controlled conjugation methods and linker chemistry on improving ADC effectiveness.

Overall, FGE-mediated protein labeling is a powerful tool. The fact that labeling can be performed by coexpressing the tag-fused POI with FGE obviates the need for additional incubation steps which makes the method convenient while the plethora of bioorthogonal reactions that have been developed for aldehyde conjugation provides a versatile repertoire of chemistries.

1.4 Conclusions

Overall, a wide variety of protein conjugates have been created using enzymatic labeling methods. These conjugates have been applied in a diverse array of applications, including the study of protein sub-cellular localization and PPIs, surface functionalization for biochip and bio-responsive materials development, and the construction of next-generation protein therapeutics with enhanced pharmacokinetics and potency. Table 1 summarizes some of the key characteristics for the enzymatic modification methods discussed in this review and their applications. Amongst all these enzymatic methods, which collectively acts as a powerful tool box, there is no one perfect solution without limitations. Understanding their respective strength and disadvantages can help to select the optimal technique to meet specific requirements. To label a given target protein, it is necessary to consider the effect of the engineered tag sequence and the location of the modification to the structure and function of the protein. Reaction kinetics also play an important role when limited substrate is supplied as well as when the target protein is sensitive to reaction conditions during long incubation times.

In addition, protein conjugation to a desired cargo can be achieved in either one step or two steps using different enzymes. For the one-step approach, the desired cargo is first incorporated into an enzyme substrate and then transferred to the protein of interest by the enzyme in a single step. Examples of this include the preparation of protein-polymer conjugates using MTG or ADCs by SrtA, MTG and PPTase. For smaller functional moieties such as the coumarin fluorophore and biotin, one-step labeling can also be achieved with PFTase, TTL and LplA. In the latter situation, some of the analogues were recognized by the wild-type enzyme, while others required an engineered protein to accommodate these bulkier substrate analogues. In some cases with larger substrates, the labeling yield may vary in a case dependent manner. Optimization of the reaction conditions is needed to achieve the best outcome.

In contrast to the one-step approach, the stepwise conjugation method is applicable to all the enzymes discussed above. The protein of interest is first labeled with a substrate analogue bearing a bioorthogonal functional group by the enzyme. Depending on the functional groups installed, subsequent reaction with their complementary reactive partner appended to cargo yields the desired protein conjugates. The functional groups that have

been incorporated into substrate analogues for each enzyme are summarized in Table 1. Small modifications, such as the azide functionality, are compatible with most enzymes while the bulkier TCO or DBCO functional groups can only be employed by those with a larger substrate binding site. Using this modular approach, a diverse range of cargos can be incorporated into the target protein. More importantly, due to the high utility of the bioorthogonal reactions, a large collection of derivatized cargos are commercially available, making the conjugation process easy to access. However, the step-wise procedure typically requires an additional purification step, which leads to the loss of material. The final conjugation product may also contain multiple unreacted protein species that further complicates the purification procedure.

In addition to the intrinsic properties related to the specific enzymes, purification processes also affect the efficiency of protein conjugate production. For SrtA, NMT, PPTase and FGE, protein modification can be achieved in *E. coli* by coexpression of the target protein and the enzyme. The modified protein can be directly purified from the cell lysate eliminating the effort to prepare the enzyme and the target protein separately by chromatographic methods. It should be noted that for SrtA, a high concentration of the primary amine substrates is often needed to achieve acceptable labeling yields. Taking advantage of the easily separated microbeads, surface immobilized MTG has been developed to facilitate the removal of the catalyst from the protein conjugates. A capture and release strategy based on aldehyde reactivity and PFTase labeling was also reported so that only the conjugated protein is eluted from the capture beads without the need to separate the unlabeled protein by chromatographic methods. Several one-pot procedures enabling site-specific modification and purification of the protein conjugates using SrtA have been also demonstrated, further streamlining the production process.

Significant progress in enzymatic protein modification has been made over the last five years. With the ever-growing demand for protein-based conjugates, the utility of enzymatic labeling methods will continue to be explored and expanded. The ability to combine different enzymatic methods also opens up numerous possibilities to create multi-functional protein constructs. Furthermore, future applications will undoubtedly benefit from the development of new substrate analogues, the capability of enzyme engineering

and the discovery of novel bioorthogonal reactions. As the area grows, novel applications and new ideas will continue to emerge.

Chapter 2. Simultaneous Site-Specific Dual Protein Labeling Using Protein Prenyltransferases

Site-specific protein labeling is an important technique in protein chemistry and is used for diverse applications ranging from creating protein conjugates to protein immobilization. Enzymatic reactions, including protein prenylation, have been widely exploited as methods to accomplish site-specific labeling. Enzymatic prenylation is catalyzed by prenyltransferases, including protein farnesyltransferase (PFTase) and geranylgeranyltransferase type I (GGTase-I), both of which recognize C-terminal CaaX motifs with different specificities and transfer prenyl groups from isoprenoid diphosphates to their respective target proteins. A number of isoprenoid analogues containing bioorthogonal functional groups have been used to label proteins of interest via PFTase-catalyzed reaction. In this study, we sought to expand the scope of prenyltransferase-mediated protein labeling by exploring the utility of rat GGTase-I (rGGTase-I). First, the isoprenoid specificity of rGGTase-I was evaluated by screening eight different analogues and it was found that those with bulky moieties and longer backbone length were recognized by rGGTase-I more efficiently. Taking advantage of the different substrate specificities of rat PFTase (rPFTase) and rGGTase-I, we then developed a simultaneous dual labeling method to selectively label two different proteins by using isoprenoid analogue and CaaX substrate pairs that were specific to only one of the prenyltransferases. Using two model proteins, green fluorescent protein with a C-terminal CVLL sequence (GFP-CVLL) and red fluorescent protein with a C-terminal CVIA sequence (RFP-CVIA), we demonstrated that when incubated together with both prenyltransferases and the selected isoprenoid analogues, GFP-CVLL was specifically modified with a ketone-functionalized analogue by rGGTase-I and RFP-CVIA was selectively labeled with an alkyne-containing analogue by rPFTase. By switching the ketone-containing analogue to an azide-containing analogue, it was possible to create protein tail-to-tail dimers in a one-

pot procedure through the copper (I)-catalyzed alkyne-azide cycloaddition (CuAAC) reaction. Overall, with the flexibility of using different isoprenoid analogues, this system greatly extends the utility of protein labeling using prenyltransferases.

2.1 Introduction

Compared to conventional non-specific modifications of proteins, site-specific protein labeling is becoming increasingly utilized. This labeling method produces homogeneously modified proteins, which can be used in many different applications in chemical biology and biotechnology.[309,310] Site-specific labeling is particularly helpful in creating chimeric molecules including protein-protein conjugates,[311,312] protein-polymer conjugates[313-316] and protein-small molecule conjugates[317,318] with uniform composition, which is highly desirable for many biomedical applications.[319] Moreover, with site-specifically labeled proteins, oriented protein immobilization with uniform surface coverage can also be achieved which is valuable for the development of protein microarrays and for the surface immobilization of enzymes in industrial applications.[254,320,321]

Enzymatic protein labeling has been widely employed to produce site-specifically labeled proteins.[322,323] In this method, a protein of interest is modified with a genetically encoded tag that is recognized by an enzyme, which then transfers its substrate (usually modified with bioorthogonal functional groups) to the tag sequence encoded within the protein. This type of site-specific labeling is widely used due to the fact that it is carried out under mild conditions, can be driven to high levels of conversion and can be used to install a range of different functional groups, allowing a variety of applications.[18]

Protein prenylation, an enzyme-catalyzed post-translational modification, has been exploited as a labeling technique because of its high specificity and its small recognition sequence (four C-terminal residues) on the protein substrate.[65,123,324] In the prenylation reaction, an isoprenoid group is attached to a cysteine residue near the C-terminus of target proteins (Figure 2.1). This reaction is catalyzed by protein prenyltransferases, including protein farnesyltransferase (PFTase) and geranylgeranyltransferase type I (GGTase-I). Both of these enzymes recognize a C-terminal CaaX motif, where the "C" represents a cysteine residue, "a" represents an aliphatic amino acid and the "X" residue determines

whether the protein will be recognized by PFTase and/or GGTase-I.[325,326] These two enzymes also have a different preference regarding the isoprenoid substrates they accept, where PFTase naturally accepts farnesyl diphosphate (FPP), while GGTase-I generally utilizes geranylgeranyl diphosphate (GGPP) due to its larger isoprenoid-binding site.[119,327]

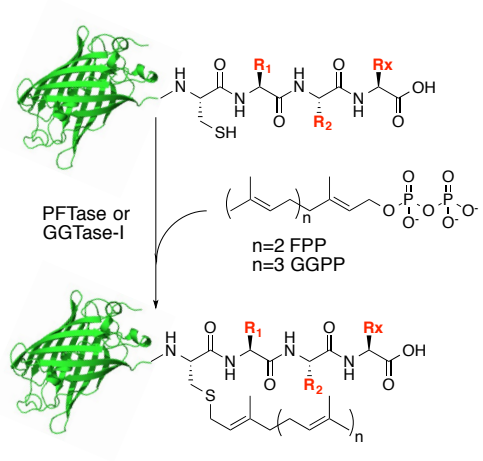


Figure 2.16 The protein prenylation reaction catalyzed by either PFTase or GGTase-I.

In order to site-specifically label a protein bearing a CaaX motif, several isoprenoid analogues containing different bioorthogonal functional groups, including alkynes,[235,253] different aldehydes[65,328] and azides[235,245] have been reported for applications ranging from the synthesis of protein conjugates to probing the prenylome. Substrates for PFTase have been used for site-specific protein labeling to create PEGylated glucose-dependent insulinotropic polypeptide (GIP) as a potential type 2 diabetes treatment by using an aldehyde-containing analogue[65] and to build antibody arrays for multiple antigen detection with an alkyne-containing analogue.[123] A substrate for GGTase-I has been employed to detect and study geranylgeranylated proteins in both cultivated cells and mammalian tissue samples.[329]

While site-specific labeling of one protein is highly useful for a variety of applications and has been explored extensively, the ability to simultaneously label two proteins is also quite useful but has been studied in only a limited number of cases. In one strategy, fusion proteins with intrinsically fluorescent proteins have been used to label multiple protein targets inside cells for visualization and detection.[330] In a different approach, a combination of SNAP and CLIP tagging was used to label two protein targets with fluorophores

simultaneously and selectively inside cells.[331] The insertion of alkyne and azide containing non-natural amino acids into two different proteins has also been utilized to simultaneously label these proteins in fixed cells.[332] Glycans, modified with different bioorthogonal functional groups, have been used for metabolic labeling of different cell surface glycoproteins for two color imaging.[333,334] However, these methods also manifest various limitations. For example protein fusions with intrinsic fluorescent proteins (eg GFP) or large tags (SNAP and CLIP) may interfere with protein function while the introduction of non-natural amino acids may suffer from low incorporation efficiency; metabolic labeling using glycans often occurs at multiple sites. In the work reported here, a simultaneous dual labeling system with high labeling efficiency and minimal tag size was developed. This was achieved by exploiting the different substrate specificities of the prenyltransferases, rat PFTase (rPFTase) and rat GGTase-I (rGGTase-I). By selecting CaaX sequences and isoprenoid analogues that are specific to distinct prenyltransferases, simultaneous selective labeling was accomplished. Using alkyne-containing and azide-containing isoprenoid analogues, protein tail-to-tail dimers were also created in a one-pot procedure.

2.2 Materials and Methods

2.2.1 General.

The synthesis for analogue **2** is provided in the supporting information. The remaining isoprenoid analogues employed here were synthesized as previously reported.[65,117,246,324,329,335] Reagents were obtained from Sigma Aldrich unless otherwise noted. OD_{600} of cells was taken with a DU 720 UV/Vis spectrophotometer (Beckman Coulter). Protein concentration was determined using a standard Bradford assay with a DTX 880 Multimode Detector plate reader (Beckman Coulter). The HPLC-based reaction condition optimization was performed using a Beckman model 125 solvent module coupled to a 305 fluorescence detector (Chrom Tech) and a Varian C_{18} column (Microsorb-MW 5 μm, 4.6 × 250 mm). LC-MS for the study of rGGTase-I isoprenoid substrate specificity was performed on a Synapt G2 mass spectrometer coupled to an Acquity ultra performance LC (Waters). LC-MS analysis for all other reactions was carried out using an Agilent 1100 series LC/MSD trap SL mass spectrometer. FPLC purification of GFP-GFP

dimer was conducted with an AKTApurifier (GE Healthcare) and a HiPrep 26/60 Sephacryl S-100 column (GE Healthcare). Cells and proteins were mixed with 50% glycerol before storage at -80 °C.

2.2.2 GFP-CVIA (10a) expression and purification.

The vector (pJexpress414) expressing GFP-CVIA (**10a**) with a polyhistidine-tag was obtained from DNA 2.0 Inc. and was transformed into BL21(DE3) supercompetent cells **(**Novagen). The detailed sequence information is provided in the supporting information. To express the protein, BL21(DE3) cells were spread on LB-agar plates containing 100 μg/mL ampicillin for overnight incubation at 37 °C. A single colony was used to inoculate 50 mL of LB media containing 100 μg/mL ampicillin and grown overnight at 37 °C, shaking at 250 rpm. 10 mL of solution from the culture was then added to 1 L of LB media containing 100 μg/mL ampicillin, which was then incubated at 37 °C and shaking at 250 rpm. When the OD_{600} of the media reached to 0.6, IPTG was added to a final concentration of 0.5 mM to induce protein expression at 25 °C overnight, shaken at 250 rpm. Cells were harvested by centrifugation at 5,400 g for 10 min and were then stored at -80 °C.

Nickel affinity chromatography was utilized to purify GFP-CVIA (**10a**). The cell pellet from 1 L of LB media was suspended in 50 mL lysis buffer (20 mM NaH_2PO_4, pH 8.0, 500 mM NaCl and 10 mM imidazole). Protease inhibitor cocktail for histidine-tagged proteins (1 mL, Sigma Aldrich) was also added to the buffer. The cells were lysed by sonication for 5 min (10 s on/10 s off) followed by clarification by centrifugation at 13,000 g for 30 min. The resulting clear supernatant was collected and loaded to a column with 25 mL of Ni-NTA resin (Qiagen), pre-equilibrated with wash buffer (20 mM NaH_2PO_4, pH 8.0, 500 mM NaCl and 20 mM imidazole) at a flow rate of 1 mL/min. The column was washed until the A_{280} showed no further decrease. Protein was then eluted with elution buffer (20 mM NaH_2PO_4, pH 8.0, 500 mM NaCl and 250 mM imidazole) and green fractions were collected. All the green fractions were combined and concentrated to 2 mL by centrifugation using an Amicon Ultra-15 centrifugal filter (10 kDa cutoff, Millipore). Buffer exchange was performed by adding Tris·HCl buffer (50 mM, pH 7.3) to the concentrated solution and then concentrating to 2 mL again. The buffer exchange process

was repeated two more times followed by storage at -80 °C in 50% glycerol prior to use. The concentration of GFP-CVIA (**10a**) was determined by its absorbance at 488 nm (ε 55,000 $M^{-1}cm^{-1}$).

2.2.3 Site-directed mutagenesis to produce GFP-CVLL (9a).

GFP-CVLL (**9a**) protein was produced by site-directed mutagenesis of DNA isolated from a clone producing GFP-CVIA (**10a**). The QuikChange II Site Directed Mutagenesis Kit protocol (Agilent, Catalog #200523) was followed and the primer sequences for PCR were determined using the QuikChange primer design tool. The sequence for the forward primer was GAC GAA CTG TAT AAG TGC GTG CTC TTA TAA AAG CTT CCC CCT AGC AT and the reverse primer was ATG CTA GGG GGA AGC TTT TAT AAG AGC ACG CAC TTA TAC AGT TCG TC. PCR amplification, DpnI DNA digestion and *E. coli* transformation were all carried out according to the QuikChange protocol. The plasmid expressing GFP-CVIA (**10a**) was extracted from BL21(DE3) cells and purified using a Wizard *Plus* SV Miniprep DNA Purification Kit (Promega, Catalog #A1460). The presence of the desired GFP-CVLL (**9a**) mutation was confirmed by DNA sequencing and the plasmid was then used to transform BL21(DE3) supercompetent cells (Novagen). The same procedure was employed to express and purify GFP-CVLL (**9a**) as was described above for GFP-CVIA (**10a**).

2.2.4 rPFTase expression and purification.

rPFTase was expressed and purified following a previously reported procedure[336] and is similar to that described for GFP proteins. Streptomycin (50 μg/mL) was used as the selection antibiotic for cell growth. IPTG and $ZnSO_4$ were added, both with final concentration of 0.5 mM to induce protein expression. To purify the enzymes, lysis buffer (50 mM Tris·HCl, pH 7.5, 200 mM NaCl, 5 μM $ZnCl_2$, 5 mM $MgCl_2$, 20 mM imidazole) and elution buffer (50 mM Tris·HCl, pH 7.5, 200 mM NaCl, 5 μM $ZnCl_2$, 5 mM $MgCl_2$, 250 mM imidazole) were used. 1 mM β-mercaptoethanol was added to these two buffers before use. Fractions containing rPFTase were collected and concentrated. Buffer exchange was performed with lysis buffer and the rPFTase was stored in 50% glycerol at -80 °C.

2.2.5 rGGTase-I expression and purification.

rGGTase-I was expressed in BL21(DE3) *E. coli* using a previously described pET23a rGGTase-I vector.[337] Bacterial cells transformed with the rGGTase-I expression vector were grown in autoinduction media (1% tryptone, 0.5% yeast extract, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.2x trace metals (0.01 mM $FeCl_3$, 0.004 mM $CaCl_2$, 0.002 mM $MnCl_2$, 0.002 mM $ZnSO_4$), 0.5% glycerol, 0.05% glucose, 0.2% α-D-Lactose) containing 100 μg/mL ampicillin at 28 °C for 24 hours.[338]

rGGTase-I was purified as previously reported, with dialysis steps replaced by desalting using a HiTrap desalting column (GE Healthcare).[326,337] Expressed rGGTase-I from the bacterial lysate was purified by FPLC (AktaPrime), first using a HiTrap DEAE-Sepharose fast flow anion exchange column (GE Healthcare) over a linear gradient of 0-0.5 M NaCl. The fractions were tested for the presence of rGGTase-I by SDS-PAGE, with fractions containing rGGTase-I pooled and concentrated using Pall Macrosep advanced centrifugal devices (10 kDa molecular weight cut-off). The protein was run through a HiTrap desalting column (GE healthcare) using HTZ buffer (50 mM HEPES, pH 7.8, 1 mM TCEP, 10 μM $ZnCl_2$) then through a HiTrap Q-sepharose HP anion exchange column (GE Healthcare) over a linear gradient of 0-0.5 M NaCl. Fractions containing rGGTase-I were collected and concentrated as before. Purified enzyme was then aliquoted, flash frozen in liquid nitrogen, and stored at -80 °C. Enzyme concentration was measured by active-site titration using dansylated peptides, as described previously.[326,339]

2.2.6 RFP-CVIA (13a) expression and purification.

A C-terminal CVIA motif was appended to TagRFP through one round of PCR using the TagRFP-N vector (Evrogen) as a template, with the CVIA motif and stop codon encoded in the 3' primer upstream of a *Hind*III restriction site. The sequence for the forward primer containing an *EcoR*I restriction site and the 5' end of TagRFP was G CAG AAT TCG ATG GTG TCT AAG GGC GAA G, and the reverse primer containing the CVIA 3' extension and the *Hind*III restriction site was AAA CTG GGG CAC AAA CTT AAT GGT TGC GTA ATC GCT TGA AAG CTT TCG A. DNA primers were

commercially synthesized (Integrative DNA Technologies, Coralville, IA). The PCR reaction (50 μL) contained Standard OneTaq buffer (10 μL), 10mM dNTPs (1 μL), reverse primer (125 ng), forward primer (125 ng), template plasmid (10 ng), and OneTaq DNA polymerase (0.25 μL, 5U/μL; New England BioLabs). PCR was performed on BioRad MyCycler under the following conditions: Initial denaturation (94 °C, 1 min); thirty cycles of denaturation (94 °C, 30s), annealing (56 °C, 1min), and extension (68 °C, 2min); final extension (68 °C, 5min); and a final hold (10 °C, ∞). After PCR, PCR products were purified using EZ-10 Spin Column PCR Purification Kit (BIO Basic Inc.) following the manufacturer's instructions.

A double digestion was then performed for ligation of EcoRI-TagRFP-CVIA-HindIII into CDF-Duet-1 plasmid and consisted of 10x NEB buffer 4 (2 μL), 100x BSA (0.2 μL), CDF-Duet-1 plasmid or PCR product (3 μg), and restriction enzymes (NEB) EcoRI (1 μL, 20 U/μL) and HindIII (1 μL, 20U/μL). Digests were incubated for 2 hours at 37 °C. Digested DNA was then purified using 0.8% agarose gel (1x TAE buffer, 120 V). DNA was excised from the gel and purified using EZ-10 Spin Column DNA Gel Extraction Kit (BIO Basic Inc.) following the manufacturer's instructions.

Ligation reactions were carried out using the Quick Ligation Kit (NEB) following the manufacturer's protocol. An aliquot of the ligation reaction (5 μL) was transformed into chemically competent DH5α *E. coli* cells (Zymo research) before plating on LB-Strep plates and incubating overnight at 37 °C. Colonies were then inoculated into 5 mL cultures of LB media containing 100 μg/mL streptomycin. Cultures were incubated and shaken at 225 rpm overnight at 37 °C. Plasmid DNA was then purified using EZ-10 Spin Column Plasmid DNA Minipreps Kit (BIO Basic Inc.) following the manufacturer's protocol. Successful ligation was confirmed through an analytical digest using EcoRI and HindIII restriction enzymes and the appearance of bands matching the size of the insert (750 bp) on 0.8% agarose gel (1x TAE buffer, 120V).

The CDF-Duet-1 vector containing the gene for RFP-CVIA (**13a**) with an N-terminal polyhistidine-tag was transformed into BL21(DE3) competent cells. The cells were spread on LB-agar plates containing 100 μg/mL streptomycin for overnight incubation at 37 °C. A single colony was used to inoculate 5 mL of LB media containing 100 μg/mL streptomycin and grown overnight at 37 °C, with shaking at 225 rpm. The 5

mL culture was then added to 1 L of LB media containing 100 μg/mL streptomycin, and incubated at 37 °C, shaking at 225 rpm. When the OD_{600} of the media reached to 0.7, IPTG was added to a final concentration of 0.5 mM and protein expression was induced at 28 °C overnight, shaking at 225 rpm. Cells were harvested by centrifugation at 6,000 g for 15 min and then stored at -80 °C.

RFP-CVIA (**13a**) was purified by nickel affinity chromatography. The cell pellet was suspended in 50 mL lysis buffer (20 mM NaH_2PO_4, pH 8.0, 500 mM NaCl and 10 mM imidazole). Benzonase (0.625U/50mL), lysozyme (0.25mg/mL), and PMSF (17μg/mL) were added to the cell suspension and incubated at 37 °C for 45 min. The cells were lysed on ice by sonication for 8 min (30 s on/30 s off) followed by clarification by centrifugation at 13,000 g for 30 min. The resulting clear supernatant was filtered through a 0.45 um filter and loaded onto a HisTrap HP 5 mL column (GE Healthcare), 3 columns in tandem, pre-equilibrated with wash buffer (20 mM NaH_2PO_4, pH 8.0, 500 mM NaCl and 20 mM imidazole) at a flow rate of 5 mL/min. The column was washed for 30 min. Protein was then eluted with elution buffer (20 mM NaH_2PO_4, pH 8.0, 500 mM NaCl and 250 mM imidazole) and red fractions were collected. All the red fractions (6 total, 2.5 mL each) were combined and concentrated to 2 mL by centrifugation using a Pall Macrosep Advance Centrifugal Device (10 kDa molecular weight cut-off). Buffer exchange was performed by adding Tris·HCl buffer (50 mM, pH 7.3) to the concentrated solution and then concentrating to 2 mL again. The buffer exchange process was repeated two more times followed by storage at -80 °C in 50% glycerol prior to use. The concentration of RFP-CVIA (**13a**) was determined by its absorbance at 555 nm (ε 100,000 $M^{-1}cm^{-1}$).

2.2.7 Isoprenoid substrate analysis using rGGTase-I.

A total of eight isoprenoid analogues (compounds **1** to **8**) were examined as substrates for rGGTase-I. Initial efforts using a continuous spectrofluorimetric assay to screen several isoprenoids as possible substrates for rGGTase-I using the dansyl-GCVLL peptide substrate **11a** proved to be difficult due to the lack of changes in dansyl fluorescence intensities for some of the analogues (Table S1). Next, the analogues were evaluated using a full-length protein substrate and LC-MS was employed to detect the prenylated product. The prenylation reaction was carried out for each analogue and for

each reaction, a cocktail (250 μL) was prepared containing Tris·HCl (50 mM, pH 7.5), KCl (20 mM), $MgCl_2$ (10 mM), $ZnCl_2$ (10 μM), DTT (5 mM), and GFP-CVLL (**9a**, 2.5 μM). All these reagents were incubated on ice for 0.5 h. Next, isoprenoid analogue (5 μM, except 20 μM for **8**) and rGGTase-I (200 nM) were added to initiate each reaction which was then allowed to proceed in a 32 °C water bath overnight followed by injection into the LC-MS instrument directly without any purification.

2.2.8 Prenylation reaction for rPFTase using analogue 5 as the isoprenoid substrate.

A reaction cocktail (250 μL) was prepared containing Tris·HCl (50 mM, pH 7.5), KCl (20 mM), $MgCl_2$ (10 mM), $ZnCl_2$ (10 μM), DTT (5 mM), and GFP-CVIA (**10a**, 2.5 μM). All these reagents were incubated on ice for 0.5 h followed by the addition of analogue **5** (15 μM) and rPFTase (150 nM). The reaction was allowed to proceed in a 32 °C water bath overnight followed by injection into the LC-MS instrument directly without any purification.

2.2.9 Docking of isoprenoid analogues to rGGTase-I.

The docking experiments were performed using Maestro (Schrodinger, version 10.1). PDB file 1N4S was used as the rGGTase-I structure. Standard protein preparation was conducted using the default settings in the protein preparation wizard of the software. A receptor grid was generated. Glide docking was performed using the standard precision docking parameters and 10,000 ligand poses were run per docking for each analogue. Analogues **3** and **8** were docked into the substrate-binding site of rGGTase-I respectively and the conformation with the highest docking score was shown for each compound. RMSD was calculated using the RMSD calculation script from Schrodinger.

2.2.10 HPLC-based reaction condition optimization.

Ds-CVLL (**11a**) and Ds-CVIA (**12a**) were synthesized as previously reported.[340] A standard chromatogram for **11a** and **12a** was obtained with 1 mL of solution containing **11a** and **12a** (2.5 μM each), which were incubated with DTT (5 mM) for 1 h at RT before injection to insure complete thiol reduction. For rPFTase prenylation, a reaction cocktail (1 mL) was prepared containing Tris·HCl (50 mM, pH 7.5), KCl (20 mM), $MgCl_2$ (10 mM),

$ZnCl_2$ (10 μM), DTT (5 mM) and was pre-incubated on ice for 0.5 h with **11a** and **12a** (2.5 μM each). Next, **5** (15 μM) and rPFTase (30-150 nM) were added to initiate the prenylation reaction. The reaction mixture was incubated for 6 h in 32 °C water bath before HPLC injection for product analysis. The rGGTase-I prenylation reaction was carried out following the same procedure except that **3** (5 μM) and rGGTase-I (200 nM) were used instead. For the dual labeling experiment, **11a** and **12a** (2.5 μM each), **5** (15 μM), **3** (5 μM), rPFTase (30 nM) and rGGTase-I (200 nM) were added to the prenylation reaction. HPLC mobile phases were buffer A (0.1% aqueous TFA) and buffer B (CH_3CN with 0.1% TFA). A method with a flow rate of 1 mL/min and a gradient of 0-100% buffer B in 100 min was used for separation. Dansyl fluorescence was detected by excitation at 340 nm and emission at 550 nm.

2.2.11 Simultaneous dual protein labeling with 3 and 5.

A reaction cocktail (250 μL) was prepared with Tris·HCl (50 mM, pH 7.5), KCl (20 mM), $MgCl_2$ (10 mM), $ZnCl_2$ (10 μM), DTT (5 mM), GFP-CVLL (**9a**, 2.5 μM) and RFP-CVIA (**13a**, 2.5 μM). The mixture was incubated on ice for 0.5 h and the reaction was initiated by the addition of **3 (**15 μM**)**, **5 (**5 μM**)**, rGGTase-I (200 nM) and rPFTase (30 nM) and allowed to proceed overnight in a 32 °C water bath. Samples were removed and injected into the LC-MS instrument directly without any purification.

2.2.12 Simultaneous dual protein labeling with 3 and 8 and protein dimerization.

The prenylation reaction cocktail (2 mL) was the same as the one described above except that **8** (20 μM) and rPFTase (60 nM) were used. Excess **5** and **8** were removed with a NAP-5 column (Amersham) using phosphate buffer (PB, 100 mM) as the eluent. The resulting solution was concentrated by ultrafilitration (Amicon Ultra-0.5 centrifugal filter, 10kDa cutoff, Millipore) to 50 μL and stored as a stock solution of labeled proteins (100 μM) at 4 °C. The CuAAC reaction (20 μL) was carried out with labeled proteins (50 μM), TBTA (100 μM from a 10 mM stock in 4:1 *t*-butanol:DMSO), TCEP (1 mM from a freshly prepared 50 mM stock) and $CuSO_4$ (1 mM). In the negative control experiment for the CuAAC reaction, an equal volume of the PB (100 mM) was added in lieu of $CuSO_4$. GFP-RFP protein dimers were also prepared using proteins that had been labeled individually

via the same procedure except that each protein was first labeled and purified separately and then added to the CuAAC reaction mixture. The GFP-GFP dimer was purified by size exclusion chromatography using FPLC before LC-MS analysis. The column was pre-equilibrated with 2 column volumes of the buffer solution containing PB (50 mM, pH 7.0), NaCl (150 mM), DTT (5 mM) with a flow rate of 0.5 mL/min. Fractions for each peak were collected and analyzed by SDS-PAGE and LC-MS.

2.3 Results and Discussion

2.3.1 Analysis of isoprenoid analogues as substrates for GGTase-I.

In previous work, we and others have reported that isoprenoid analogues with various functional groups can serve as substrates for PFTase and be successfully incorporated into proteins.[65,117,123,246,324] In contrast, the isoprenoid specificity of GGTase-I had not been explored in detail. Since GGTase-I is an essential enzyme in the prenylation process, the development of isoprenoid analogue substrates for GGTase-I is critical for probing and studying the protein substrates of GGTase-I in cells.[329] Therefore, we first studied the isoprenoid substrate specificity of rGGTase-I using a library of isoprenoid analogues (Figure 2.2). Based on crystallographic data, it is known that rGGTase-I has a larger isoprenoid substrate-binding pocket compared to that of PFTase.[327] Thus, we hypothesized that isoprenoid analogues with bulky moieties would be better recognized by rGGTase-I. Eight isoprenoid analogues that varied in the length of the carbon backbone and the size of additional functionality were screened as substrates for rGGTase-I using a green fluorescent protein modified with the C-terminal sequence, CVLL (GFP-CVLL, **9a**) as the protein substrate; CVLL is a naturally occurring rGGTase-I recognition sequence that has relatively high reaction efficiency compared to other rGGTase-I substrates.[327,339] For the isoprenoid analogues, the syntheses of all except compound **2** have been previously described. The aminooxy-containing analogue (**2**) was prepared from an aldehyde precursor similar to **1** (Scheme S2.1) and was found to be a substrate for yeast PFTase (yPFTase) (Figure S2.1-S2.3). For the other isoprenoids, compounds **1**, **3**, **4** and **7** were previously known to be substrates for yPFTase[65,246,335] and compounds **5** and **6** were known substrates for both yPFTase and rPFTase.[39] Compound **8** was a known substrate for

GGTase-I.[329] LC-MS was used to determine the extent of prenylation of GFP-CVLL catalyzed by rGGTase-I with all these analogues and the reaction efficiency was calculated from the relative deconvoluted mass peak intensities of the prenylated and the unprenylated protein.

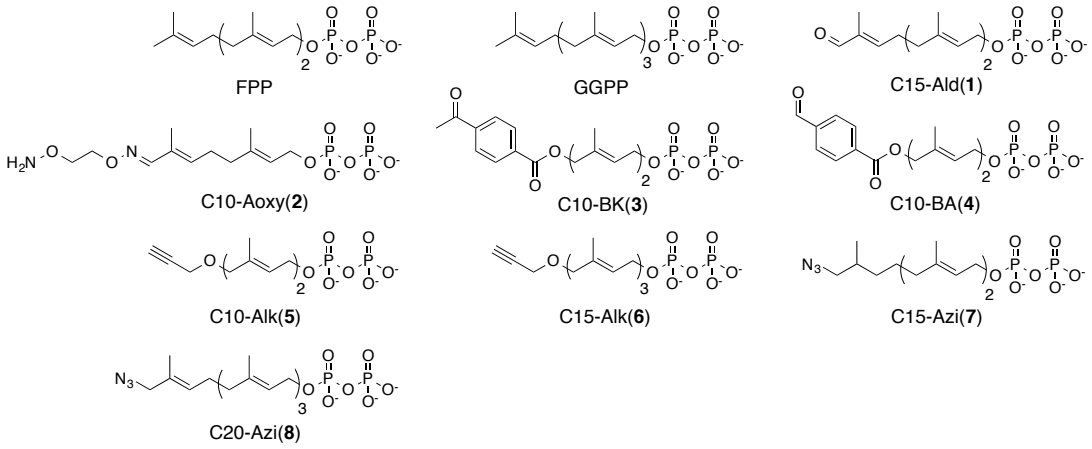


Figure 2.2 Structures of isoprenoid analogues used for the analysis of rGGTase-I substrate specificity.

Based on the mass spectrometry results (Table 2.1), we found that compound **3**, which contained a bulky aromatic moiety, as well as compounds **6** and **8**, which contained a longer isoprenoid backbone, were efficiently incorporated into GFP-CVLL **(9a)** by rGGTase-I. In contrast, analogues that were similar in length to FPP showed little or no product formation. These results were consistent with our hypothesis that the larger substrate-binding site of rGGTase-I would accommodate bulkier isoprenoid substrates.

Table 2.1 Prenylated products detected by LC-MS for the reactions of rGGTase-I with GFP-CVLL (**9a**) and eight isoprenoid analogues.

Substrate	Calculated Mass	Observed Mass	Efficiency
GFP-CVLL (**9a**)	28200	28200	-
FPP	28404	28403	33%
GGPP	28472	28472	100%
C15-Ald (**1**)	28418	N. D.	-
C10-Aoxy (**2**)	28424	N. D.	-
C10-BK (**3**)	28498	28498	84%
C10-BA (**4**)	28484	28481	54%
C10-Alk (**5**)	28390	N. D.	-
C15-Alk (**6**)	28458	28457	77%
C15-Azi (**7**)	28447	28444	50%
C20-Azi (**8**)	28513	28517	83%

To further study the binding of these analogues to rGGTase-I, we also performed computational docking experiments using analogues **3** and **8**. The structure of rGGTase-I bound with both GGPP and a geranylgeranylated peptide (KCVIL) product was used for that work. It was found that the isoprenoid backbone of analogue **8** adopted a similar "L" shape comparable to that of GGPP and that the fourth isoprenoid unit was oriented approximately 90° relative to the axis formed from the remaining isoprenoids, which was also consistent with the GGPP-bound structure (Figure 2.3A). For analogue **3**, due to the rigidity of the benzene ring, it adopted a different conformation than the corresponding isoprenoid units in the GGPP structure (Figure 2.3B). The root-mean-square deviation (RMSD) was calculated to quantify the differences between the conformations of GGPP and the docked analogues when docked into the rGGTase-I active site. The first two isoprenoid units of these compounds were compared because they are closest to the site of reaction and because they have been found to undergo major conformational changes during the prenylation reaction.[257,327,341,342] The RMSD between the first two isoprenoid groups (relative to bound GGPP) was calculated to be 1.09 Å for analogue **8** and 0.73 Å for analogue **3**. These small RMSD values suggest that the isoprenoid groups in these

analogues can adopt similar conformations compared with GGPP and may account for their relatively high incorporation efficiency by rGGTase-I.

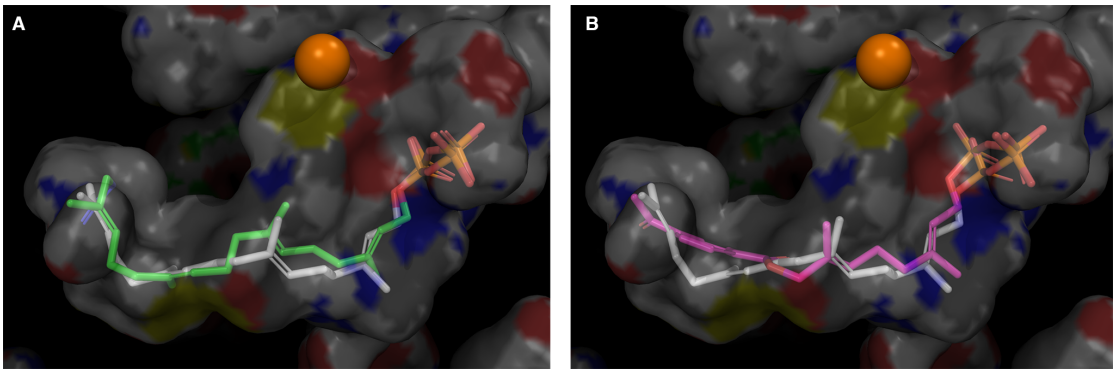


Figure 2.3 Computational modeling of analogues **3** and **8** docked into the isoprenoid substrate-binding site of rGGTase-I. (A) Overlay structure of crystal bound GGPP and the docked analogue **8**. (B) Overlay structure of crystal bound GGPP and the docked analogue **3**. Note the similarity in the conformation of the first two isoprenoid units compared to bound GGPP. Color scheme and representations: the binding pocket of rGGTase-I is presented as a transparent surface. The Zn^{2+} atom is shown as an orange sphere. GGPP is shown in stick format (C and H in white; O in red; P in orange) in both panels. Analogue **8** is presented as green sticks in panel A and analogue **3** is given as pink sticks in panel B. The conformation (pose) with highest docking score is shown for each analogue. The pdb file 1N4S[24] was used to generate these images.

2.3.2. Identification of isoprenoid substrates that can be used for simultaneous dual protein labeling.

Due to the different substrate specificities of rGGTase-I and rPFTase, we hypothesized that these two enzymes could be used to simultaneously label two different proteins (Figure 2.4). To investigate this, we sought to first identify a combination of a CaaX sequence and an isoprenoid analogue, which could be processed selectively by rGGTase-I, even in the presence of rPFTase. Similarly, it was envisioned that another combination of a CaaX sequence and an isoprenoid analogue could serve as a selective rPFTase substrate pair. Thus, when the two proteins with different CaaX motifs and the two isoprenoid analogues were combined in the presence of both enzymes, only two labeled protein products would form, with no cross reactions occurring.

GFP, whose fluorescence can be easily detected, was chosen as a model protein to characterize the dual labeling system. GFP appended with a C-terminal CVIA sequence (GFP-CVIA, **10a**), was used as the substrate for rPFTase since the CVIA motif is

efficiently recognized by rPFTase.[343] GFP-CVLL (**9a**), described above**,** was used for rGGTase-I prenylation since it is also known to be a good substrate. Based on the screening results described above (Table 2.1), compounds **3** and **6** were first chosen as possible candidates for rGGTase-I substrates for the dual labeling system due to their high incorporation efficiency.

In order to determine whether these rGGTase-I isoprenoid substrates would fit the desired criteria of orthogonal reactivity, their inability to be incorporated by rPFTase needed to be ascertained. Therefore prenylation reactions containing analogues **3** and **6** were carried out with rPFTase and GFP-CVIA (**10a**). LC-MS was again employed to detect the presence of putative prenylated products and to quantify the extent of prenylation (Table 2.2). Analogue **6** was found to be incorporated into GFP-CVIA (**10a**) by rPFTase, which was consistent with previously reported results.[343] However, rPFTase was unable to efficiently utilize analogue **3** in the prenylation reaction. Therefore, compound **3**, which was an efficient substrate for rGGTase-I but not for rPFTase was selected to be used as the rGGTase-I complementary substrate in our dual labeling study.

Table 2.2 Prenylated products detected by LC-MS for the reactions of rPFTase with GFP-CVIA (**10a**) using isoprenoid analogues **3** and **6**.

Isoprenoid Substrate	Calculated Mass	Observed Mass	Efficiency
GFP-CVIA (**10a**)	20158	28155	-
C10-BK (**3**)	28456	N. D.	-
C15-Alk (**6**)	28416	28413	23%

To select an isoprenoid analogue for rPFTase prenylation reaction, compound **5**, which is a known rPFTase substrate,[343] was chosen because of its low rGGTase-I prenylation efficiency (Table 2.1) and orthogonal functionality relative to **3**. Reaction of analogue **5** with GFP-CVIA (**10a**) and rPFTase proved to be highly efficient (Figure S2.4), which supported this selection.

After choosing the respective isoprenoid substrates for rGGTase-I and rPFTase in the dual labeling system, it was also necessary to evaluate the cross reactivity of the enzymes to prenylate the non-complimentary CaaX sequence under the reaction conditions.

To study this, four prenylation reactions were performed, including two rGGTase-I reactions with GFP-CVIA (**10a**) using **3** or **5** respectively, and two reactions for rPFTase with GFP-CVLL **(9a)** in the presence of either **3** or **5** (Table S2.2). The reactions with rGGTase-I showed no prenylated product with the noncognate substrate **10a**, confirming the high selectively of rGGTase-I for its CaaX substrate. For the reactions containing rPFTase, it was found that although no prenylated product was produced with GFP-CVLL **(9a)** and **3**, minor amounts of **9a** modified with analogue **5** were detected via LC-MS. These results suggested that if GFP-CVIA (**10a**), GFP-CVLL **(9a)** and **5** were added together with rPFTase using the same reaction conditions as above, it would result in both proteins being modified with **5**, producing a mixture that was not desirable for dual labeling. Therefore, further investigation was needed to optimize the reaction conditions to achieve more specific labeling.

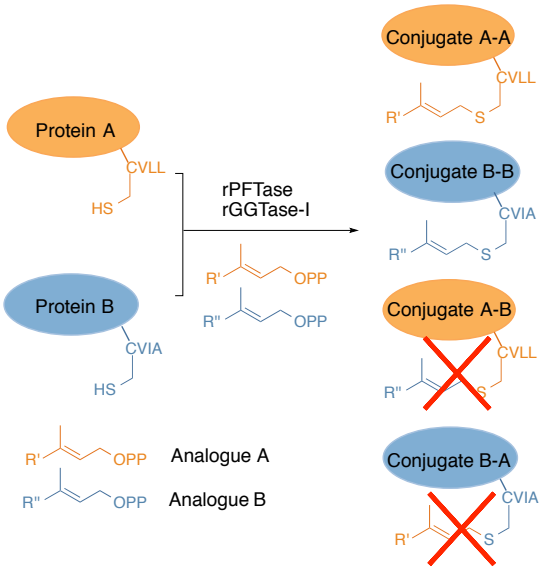


Figure 2.4 Scheme for simultaneous dual labeling using rPFTase and rGGTase-I. Protein A is only modified with analogue A by rGGTase-I and protein B is only labeled with analogue B by rPFTase.

2.3.3. Optimizing reaction conditions to achieve specific labeling.

In order to optimize the prenylation reaction conditions, we employed two dansylated peptide substrates, dansyl-GCVLL (Ds-GCVLL, **11a**) and dansyl-GCVIA (Ds-GCVIA, **12a**) (Figure 2.5A); these peptides and the resulting prenylated products can be readily monitored using reverse phase HPLC with fluorescence detection. The peptides and

their prenylated products also have distinct retention times allowing the respective reaction efficiencies to be easily determined along with any cross-reaction that may occur.

Initial reactions using rPFTase with **11a, 12a** and **5** showed two product peaks corresponding to **11c** and **12b** (Figure S2.5), in concordance with the LC-MS data described above. Previous CaaX substrate screening of rPFTase with peptide libraries have reported that the CVLL motif is a much poorer substrate compared to CVIA.[343] Therefore, it was likely that **11c** was formed here due to a high rPFTase concentration. When the concentration of rPFTase was reduced from 150 nM to 30 nM, only one product peak corresponding to **12b**, was observed (Figure 2.5B, ii). Thus, by reducing the concentration of rPFTase, we were able to selectively label only **12a** in the presence of the other CaaX substrate. For the reaction of rGGTase-I and **3** in the presence of both **11a** and **12a**, only one product, **11b**, was produced (Figure 2.5B, iii), indicating a high degree of selectivity for this reaction.

Finally, an all-in-one reaction was performed containing both prenyltransferases, both peptide substrates and both isoprenoid analogues. Under these optimized conditions, it was found that only the two desired products, **11b** and **12b**, were formed with no signs of cross prenylation (Figure 2.5B, iv). Thus, Ds-GCVIA (**12a**) was only prenylated with **5** by rPFTase and Ds-GCVLL (**11a**) was only labeled with **3** by rGGTase-I, demonstrating that the dual labeling system can be applied to peptide models successfully.

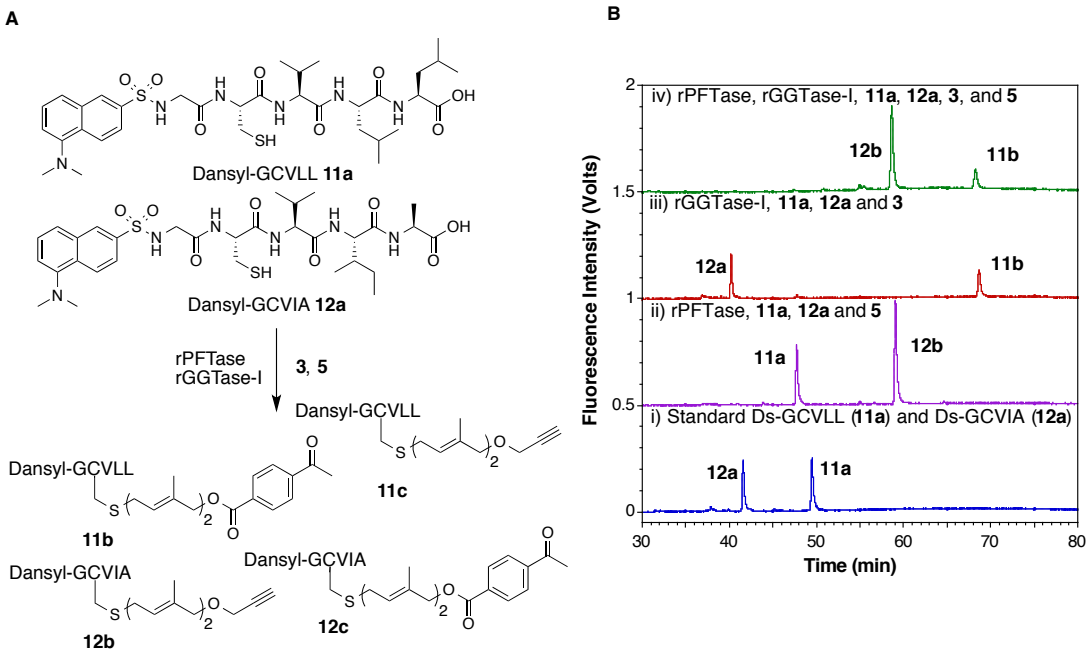


Figure 2.5 Prenylation reactions of rPFTase and rGGTase-I monitored by HPLC. (A) Scheme for the dual labeling reaction of Ds-GCVLL (**11a**) and Ds-GCVIA (**12a**) with isoprenoid analogues **3** and **5** in the presence of rGGTase-I and rPFTase. All the possible prenylated products are shown. (B) HPLC traces of prenylation reactions for rPFTase and rGGTase-I. i). Unmodified **11a** and **12a**. ii). Prenylation reaction of **11a** and **12a** with **5** in the presence of rPFTase. iii). Prenylation reaction for **11a** and **12a** with **3** in the presence of rGGTase-I. iv). All-in-one prenylation reaction with **11a**, **12a**, **3** and **5** in the presence of both rPFTase and rGGTase-I. For each reaction, only the desired products were formed with no signs of cross prenylation.

2.3.4. Simultaneous dual labeling on protein models.

After performing proof-of-concept experiments with peptide substrates, the selective dual labeling strategy was applied to full-length proteins. GFP-CVLL (**9a**) and red fluorescent protein (TagRFP)[344] modified with a C-terminal CVIA motif (RFP-CVIA, **13a**) were chosen as convenient model proteins due to their different fluorescence excitation and emission profiles. Using the optimized conditions obtained from the HPLC experiments described above, an all-in-one reaction with rPFTase, rGGTase-I, **3**, **5**, GFP-CVLL **(9a)** and RFP-CVIA (**13a**) was performed (Figure 2.6). It was found that only two prenylated products, ketone-labeled GFP-CVLL (**9b**) and alkyne-modified RFP-CVIA (**13b**) were produced. Using LC-MS, the labeling efficiency for rPFTase was estimated to be >99% and 52% for the rGGTase-I reaction. Thus, we were able to demonstrate that the dual labeling system can be applied to both large proteins and small peptides. The incorporated alkyne and ketone functionalities are not only bioorthogonal but they can also

participate in different reactions: the copper (I)-catalyzed alkyne-azide cycloaddition (CuAAC) reaction for the alkyne functional group and the oxime ligation for the ketone functional group. Therefore, subsequent ligation reactions can be performed in both parallel and sequential manners, showing significant potential for pulling down multiple proteins simultaneously and for creating patterned protein microarrays. Moreover, with the development of other orthogonal reaction pairs that can be performed simultaneously without cross-reactivity, such as the strain-promoted alkyne-azide cycloaddition (SPAAC) reaction and the tetrazine ligation,[333,345-348] more isoprenoid analogues can be designed and potentially used in the dual labeling system, further expanding its application.

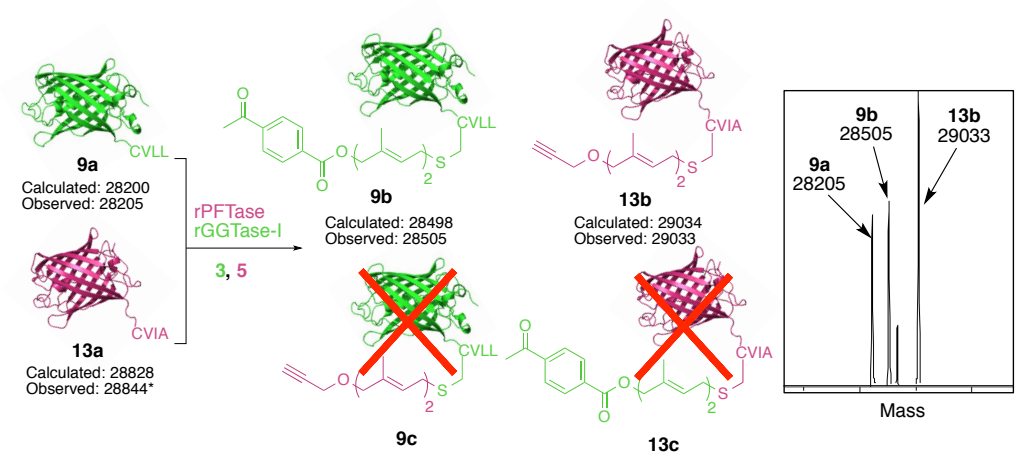


Figure 17 Scheme for the simultaneous labeling of GFP-CVLL (**9a**) and RFP-CVIA (**13a**) with the ketone-containing isoprenoid analogue **3** and the alkyne-containing analogue **5** by rPFTase and rGGTase-I. GFP-CVLL (**9a**) was only modified with the ketone functionality by rGGTase-I and RFP-CVIA (**13a**) was only labeled with the alkyne functional group by rPFTase. *The observed mass for **13a** differs from the calculated mass by 16 units, which is likely due to the methionine oxidation

2.3.5. Protein dimerization using the CuAAC reaction.

As noted above, LC-MS analysis (Table 2.1) showed that analogue **8** was efficiently utilized by rGGTase-I. Further investigation of this compound established that it was not recognized by rPFTase. Therefore, we expected that analogue **8** could be utilized, in place of **3**, in dual labeling experiments. By using analogues **5** and **8**, the two target proteins would be selectively modified with alkyne and azide functionality respectively, which could then be coupled together in a tail-to-tail manner via a one-pot procedure using the CuAAC reaction.

This strategy was first used to prepare a GFP-GFP homodimer (Figure 2.7). A single prenylation reaction was performed containing GFP-CVLL **(9a)**, GFP-CVIA (**10a**), **5**, **8**, rGGTase-I and rPFTase. LC-MS results confirmed that only the azide-appended **9d** and alkyne-modified **10b** were produced. Although no product from cross prenylation was formed, the reaction efficiency for rPFTase-catalyzing reaction was only 36% (Figure S2.6). Since GGPP is known to inhibit the activity of rPFTase,[349] it is likely that compound **8** is also an inhibitor of rPFTase, given their similar structures. To overcome the inhibition effect, the amount of rPFTase used in the reaction was doubled. This led to complete conversion of GFP-CVIA (**10a**) to **10b** without the formation of the cross prenylated product. The labeling efficiency for GFP-CVLL **(9a)** with analogue **8** by rGGTase-I was estimated to be 62%.

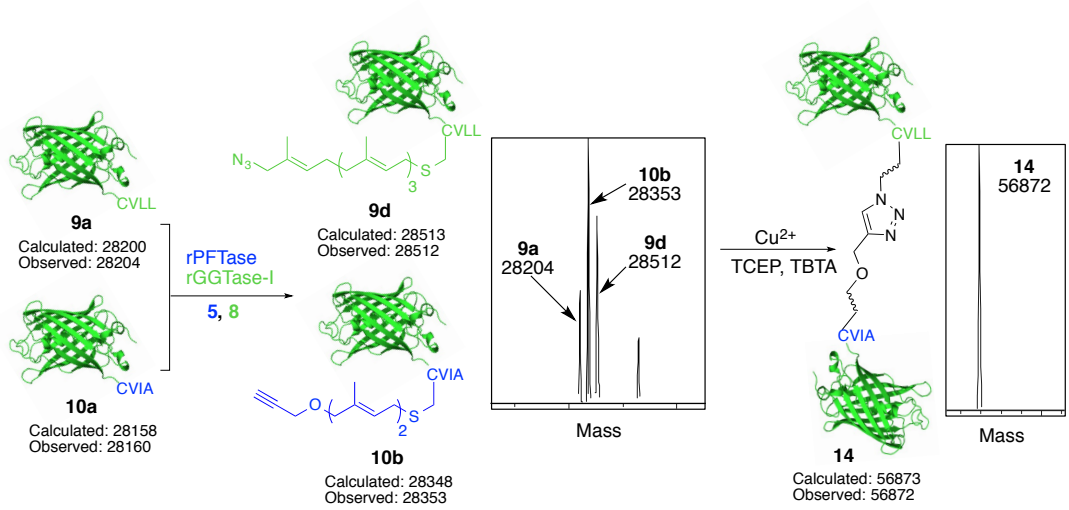


Figure 2.7 Scheme for simultaneous dual labeling of GFP-CVLL (**9a**) and GFP-CVIA (**10a**) with azide-containing isoprenoid analogue **8** and alkyne-containing analogue **5** by rGGTase-I and rPFTase, followed by the CuAAC reaction to produce GFP-GFP homodimer.

Once the specificity of the azide-alkyne dual labeling system was established, the CuAAC reaction was performed to produce GFP-GFP tail-to-tail homodimer. This was accomplished by the simple addition of the reagents ($CuSO_4$, TCEP and TBTA) to the reaction mixture following a desalting step to remove unreacted prenyl donor analogues **5** and **8**. SDS-PAGE and in-gel fluorescence imaging were employed to characterize the dimer (Figure 2.8). Importantly, to prevent the loss of GFP fluorescence due to complete

denaturation, the protein samples were mixed with loading buffer and then applied directly onto an SDS-PAGE gel without heating (Figure S2.7). The appearance of the dimer band, with higher molecular weight, was clearly observed in the CuAAC reaction mixture (Figure 2.8, Lane 4), indicating the successful formation of the GFP-GFP dimer. The reaction efficiency was estimated to be 50% from the gel image (Figure 2.8A, Lane 4). To further confirm that the desired cyclization reaction had occurred, the protein dimer was purified using size exclusion chromatography and characterized by LC-MS, which showed an observed mass of 56873 Da, closely matching the calculated mass of 56872 Da.

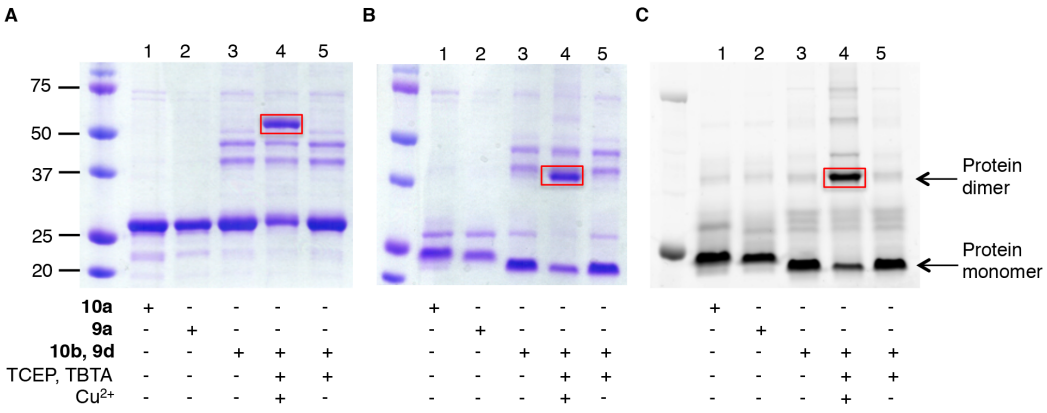


Figure 2.8 SDS-PAGE and in-gel fluorescence imaging analysis of GFP-GFP dimer. (A) Coomassie blue stained gel of protein samples that were heated prior to loading. (B) Coomassie blue stained gel of protein samples that were loaded without heating. (C) In-gel fluorescence imaging of GFP fluorescence. Lane 1: unmodified GFP-CVIA (**10a**). Lane 2: unmodified GFP-CVLL (**9a**). Lane 3: simultaneously labeled **10b** and **9d**. Lane 4: the CuAAC reaction mixture of simultaneously labeled **10b** and **9d** with TBTA (100 μM), TCEP (1 mM) and $CuSO_4$ (1 mM). The dimer band is highlighted in the red box. Lane 5: the CuAAC reaction control for the simultaneously labeled **10b** and **9d** with no $CuSO_4$ added. Densitometry analysis was performed using ImageQuant TL (version 7.0). Samples in panel C were not heated prior to gel loading. Note that the apparent size of GFP changes upon heat denaturation (compare panels A and B).

Beyond the GFP-GFP homodimer described above, we also wanted to demonstrate the generality of the dual labeling method by preparing a heterodimeric construct as well. Thus a chimeric protein incorporating red and green fluorescent proteins was targeted. For this, GFP-CVLL (**9a**), RFP-CVIA (**13a**), **5**, **8** and both prenyltransferases were incubated together. According to the LC-MS results, compound **5** was selectively incorporated onto RFP-CVIA (**13a**) by rPFTase with complete conversion and compound **8** was specifically appended to GFP-CVLL (**9a**) by rGGTase-I with 60% efficiency (Figure S2.8). No cross-reaction was found to occur. After prenylation and removal of unreacted prenyl donors, the

CuAAC reaction was initiated by adding the reagents necessary for cycloaddition ($CuSO_4$, TCEP and TBTA). Protein aggregates were observed, which was probably due to the near neutral isoelectric point of the GFP-RFP dimer, but we were still able to analyze the reaction mixture using both SDS-PAGE and in-gel fluorescence imaging analysis. From the in-gel fluorescence imaging results (Figure 2.9), the fluorescence signal of the dimer band, which also manifested the expected increase in molecular weight, was detected when scanning for both GFP and RFP fluorescence, suggesting the successful dimerization of **9d** and **13b**. The molecular weight of the dimer was estimated to be 57.9 kDa based on SDS-PAGE analysis, which is in good agreement with the calculated mass of 57.5 kDa. The GFP-RFP dimer was also prepared using **9d** and **13b** that had each been generated separately and it was interesting to note that the CuAAC reaction efficiency of the simultaneously labeled **9d** and **13b**, estimated to be 23% based on band intensity (Figure S2.9B, Lane 6), was comparable to the 29% of the separately labeled proteins (Figure S2.9B, Lane 8). Since the yield of dimeric material was comparable whether dimerization was performed via simultaneous labeling (where incomplete prenylation occurred) or separate labeling (where the reactions were driven to completion), these results suggest that other factors such as dimer solubility likely play a role in influencing the yield in this particular case. Overall, with the utilization of the azide-containing and the alkyne-containing isoprenoid analogues, homo- and hetero-dimeric tail-to-tail protein constructs were successfully constructed in a one-pot procedure using the simultaneous dual labeling system.

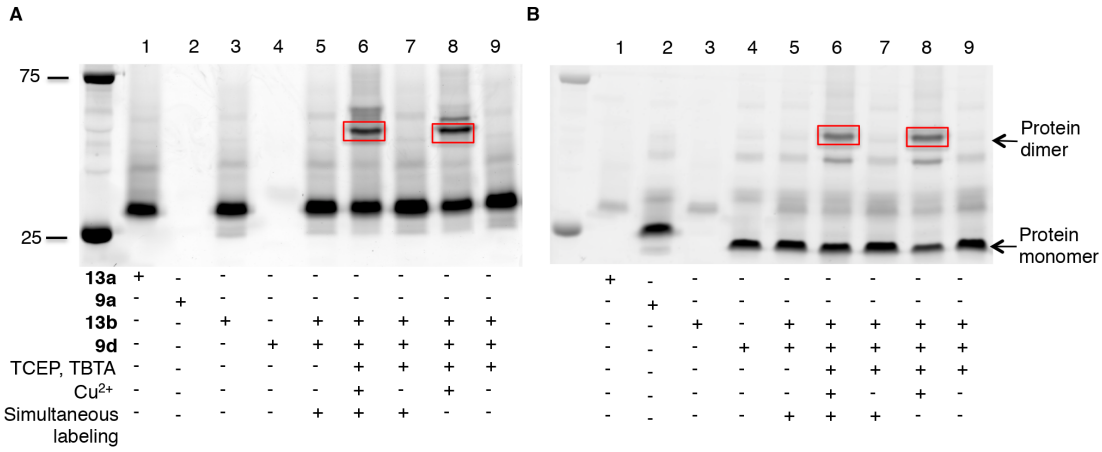


Figure 2.9 In-gel fluorescence imaging analysis of the GFP-RFP dimer. (A) RFP fluorescence imaging. (B) GFP fluorescence imaging. The GFP-RFP dimer exhibited fluorescence in both GFP and RFP fluorescence channels. Lane 1: unmodified RFP-CVIA (**13a**). Lane 2: unmodified GFP-CVLL (**9a**). Lane 3: **13b**, produced from reaction of **13a**, **5** and rPFTase. Lane 4: **9d**, produced from reaction of **9a**, **8** and rGGTase-I. Lane 5: simultaneously labeled **13b** and **9d**. Lane 6: the CuAAC reaction mixture of simultaneously labeled **13b** and **9d** with TBTA (100 μM), TCEP (1 mM) and $CuSO_4$ (1 mM). Lane 7: the CuAAC reaction control for the simultaneously labeled **13b** and **9d** with no $CuSO_4$ added. Lane 8: the CuAAC reaction mixture of separately labeled **13b** and **9d**. Lane 9: the CuAAC reaction control for the separately labeled **13b** and **9d** with no $CuSO_4$ added. The dimer band is highlighted in the red box.

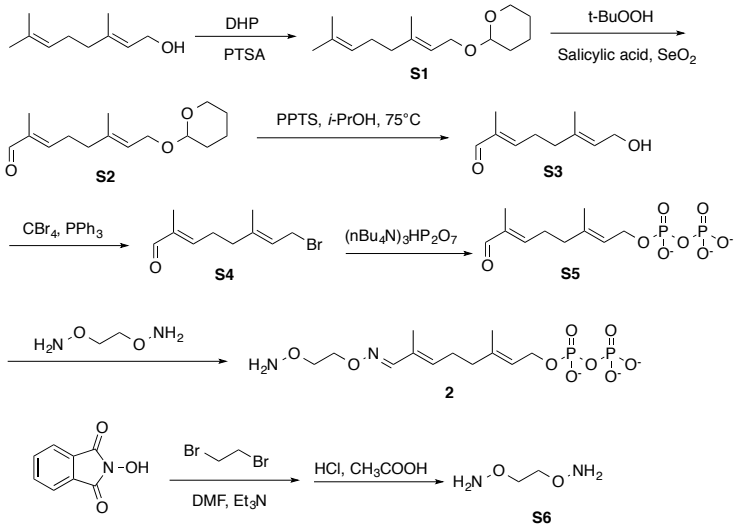
2.4 Conclusion

In this work, we examined the isoprenoid substrate specificity of rGGTase-I using eight different isoprenoid analogues that incorporate various bioorthogonal functionality. In contrast to results obtained using PFTase, it was found that isoprenoid analogues with bulky moieties, such as a benzyl-ketone moiety, and a longer carbon backbone can be efficiently recognized by rGGTase-I. By taking advantage of the differences in both isoprenoid and protein substrates between rGGTase-I and rPFTase, we have successfully developed a simultaneous dual labeling system that can be used to selectively label two proteins in tandem. After optimization of the reaction conditions, it was demonstrated that ketone functionality can be incorporated into GFP-CVLL specifically by rGGTase-I and an alkyne functional group can be attached to RFP-CVIA by rPFTase, with no cross reactivity between the enzymes and their substrates; subsequent work demonstrated that this system is not limited to only one isoprenoid substrate pair. Using azide- and alkyne-

functionalized analogues, tail-to-tail homo- and hetero-dimeric proteins were produced in one-pot labeling reactions via this simple procedure. These *in vitro* biochemical experiments set the stage for future *in vivo* applications including the generation of large multi-domain fusion proteins that may be difficult to express. Moreover, the tail-to-tail protein orientation generated via this strategy is not readily accessible via standard genetic methods. Finally, it should be noted that the ability to simultaneously label PFTase and GGTase-I substrates with orthogonal functionality should also be useful in prenylomic studies since it potentially allows levels of farnesylated and geranylgeranylated proteins to be studied concurrently. Overall, this dual labeling system greatly extends the utility of protein modification using prenyltransferases and opens up numerous avenues for future applications.

2.5 Supplementary Data

2.5.1 Synthesis of C10-Aoxy (2)



Scheme S2.1. Synthesis of C10-Aoxy (**2**)

2.5.2 Synthesis of S3. Compound **S1** and **S2** were prepared as previously described.[65] Protected aldehyde **S2** (0.5 g) was dissolved in *i*-PrOH (20 mL) in a 50 mL flask. Pyridinium p-toluenesulfonate (PPTS, 20 mg) was added as catalyst. The reaction was then refluxed at 75 °C for 4 h, when TLC analysis indicated (near) complete conversion to the product. It was then quenched by adding saturated $NaHCO_3$ (10 mL) and EtOAc (100 mL). The organic layer was then separated and dried over anhydrous $MgSO_4$. The solvent was

removed *in vacuo* and afforded 0.24 g of compound **S3** (72% yield) as a pale yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.30 (s, 1H), 6.42 (t, *J* = 6.8 Hz, 1H), 5.37 (t, *J* = 6.8 Hz, 1H), 4.09 (d, *J* = 7.0 Hz, 2H), 2.43 (q, *J* = 7.5 Hz, 2H), 2.15 (t, *J* = 7.5 Hz, 2H), 1.67 (s, 3H), 1.63 (s, 3H).

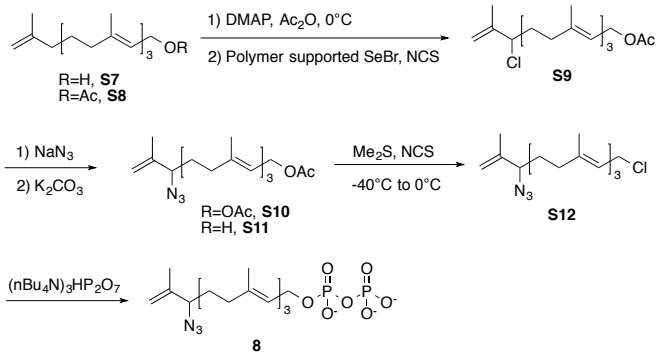
2.5.3 Synthesis of S4. PPh3 (polymer supported beads, 167 mg) was added to a solution of **S3** (77 mg) in CH_2Cl_2 (15 mL), and the reaction was stirred for 20 min to let the beads swell in the solution. CBr_4 (165 mg) was added slowly to the reaction mixture and the reaction was stirred for 15 min when TLC analysis showed most of starting material was converted to product. The reaction mixture was filtered to separate the beads, and the solvent was evaporated *in vacuo*. The product was further purified by silica gel flash chromatography to yield 90 mg of **S4** (85% yield) as a pale yellow liquid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.39 (s, 1H), 6.43 (t, *J* = 7.2 Hz, 1H), 5.57 (t, *J* = 8.5 Hz, 1H), 4.00 (d, *J* = 8.0 Hz, 2H), 2.49 (q, *J* = 7.2 Hz, 2H), 2.25 (t, *J* = 7.5 Hz, 2H), 1.75 (s, 3H), 1.74 (s, 3H).

2.5.4 Synthesis of S5. Compound **S4** (50 mg) and $[(nBu)_4N)]_3HP_2O_7$ (604 mg) were mixed in dry CH_3CN (1 mL) and stirred for 3 h, after which the solvent was removed *in vacuo*. Bio-Rad AG 50WX8 ion exchange resin (100–200 mesh, H^+ form) was used to convert the product to its ammonium form. The resin was packed and washed with three column volumes of H_2O/conc. NH_4OH (2:1, v/v) until the elution solvent became basic. The resin was then equilibrated with four volumes of NH_4HCO_3 (25 mM)/*i*-PrOH (49:1, v/v; solvent A). The crude product was dissolved in 1 mL of solvent A and applied to the column and eluted with additional solvent A (20 mL), lyophilized, and purified by RP-HPLC with a semi-preparative column using the following conditions: detection: 214 nm; flow rate: 5.0 mL/min; 2 mL injection loop; gradient 0–15% solvent B in 30 min, 60–100% in 5 min; solvent A: 25 mM NH_4HCO_3, solvent B: CH_3CN. Compound **S5** eluted from 12–15% solvent B. Fractions containing pure **S5** were collected and lyophilized to yield 18 mg (25% yield) of a white powder. ^{1}H NMR (500 MHz, D_2O) δ 9.22 (s, 1H), 6.74 (t, *J* = 7.2 Hz, 1H), 5.45 (t, *J* = 7.0 Hz, 1H), 4.42 (t, *J* = 6.5 Hz, 2H). 2.53 (q, *J* = 7.2 Hz, 2H), 2.24 (t, *J* = 7.5 Hz, 2H), 1.68 (s, 3H), 1.66 (s, 3H); ^{31}P NMR (121 MHz, D_2O) δ -6.05 (d, *J* = 22.9 Hz, 1P),

-10.00 (d, J = 22.3 Hz, 1P); *m/z* (HR-ESI-MS) found 327.0390 ([M-H]$^-$), calculated 327.0404.

2.5.5 Synthesis of C10-Aoxy (2). Compound **S6** was prepared as previously described.[57] Geranyl aldehyde diphosphate **S5** (1 mL, 10 mM) was mixed with bis-aminooxy compound **S6** (1 mL, 30 mM) in Tris·HCl (50 mM, pH 7.5), and the reaction was allowed to proceed for about 1 h when MS analysis showed (near) complete conversion of **S5** to **2**. Compound **2** was used in enzymatic reactions without further purification. *m/z* (HR-ESI-MS) found 401.0901 ([M-H]$^-$), calculated 401.0884.

2.5.6 Synthesis of C20-Azi (8)



Scheme S2.2. Synthesis of C20-Azi (**8**), based on a previously reported procedure.[329]

2.5.7 Synthesis of S8. Geranylgeraniol **S7** (100 mg) was dissolved in CH_2Cl_2 (3 mL) and added to a stirred solution of 4-dimethylaminopyridine (DMAP, 84 mg) in anhydrous CH_2Cl_2 (4 mL) under nitrogen atmosphere at 0 °C, followed by the addition of Ac_2O (70 µL). Stirring was continued at 0 °C for 1.5 h. After the completion of reaction, which was monitored by TLC, the reaction mixture was diluted with CH_2Cl_2 (20 mL), washed with H_2O (2×10 mL) and brine (1×60 mL). The organic layer was dried over anhydrous $MgSO_4$ and the solvent was removed *in vacuo*. The product was purified by silica column chromatography to afford geranylgeranyl acetate **S8** as a colorless oil (87 mg, 76%).

2.5.8 Synthesis of S9. Geranylgeranyl acetate **S8** was dissolved in anhydrous CH_2Cl_2 (4 mL) and added to a stirred solution of polymer bound selenium bromide (SeBr, 38 mg) in anhydrous CH_2Cl_2 (5 mL) under nitrogen. Stirring was continued and N-chlorosuccinimide (NCS, 38 mg) was added. After stirring the reaction overnight, solvent was removed *in*

vacuo. Et_2O (20 mL) was added to the residue and filtered through cotton wool to remove the polymer-bound SeBr. The filtrate was then washed with H_2O (3×10 mL) and brine (1×10 mL). The solvent was removed *in vacuo* and the crude product was purified by silica column chromatography in EtOAc/hexane to afford 21mg of **S9** (29% yield, brsm, 75% conversion).

2.5.9 Synthesis of S10. NaN_3 (10 mg) was added to a stirred solution of **S9** (27 mg) in DMSO (1.5 mL) at room temperature. The reaction mixture was stirred for 6 h. The reaction mixture was then diluted with H_2O (8 mL) and extracted with Et_2O (5 x 10 mL). Combined organic layer was washed with H_2O (3×10 mL) and brine (1×10 mL), dried over anhydrous $MgSO_4$ and concentrated. Crude product was purified by silica gel column chromatography in EtOAc/hexane to afford 10 mg of **S10** (36% yield).

2.5.10 Synthesis of S11. K_2CO_3 (10 mg) was added to a stirred solution of **S10** (10 mg) dissolved in MeOH (2 mL). The mixture was stirred for 2 h. After the completion of the reaction (monitored by TLC), H_2O (6 mL) was added to the reaction mixture and the aqueous layer was extracted with EtOAc (3×10 mL). The combined organic layers were washed with brine (1x10 mL), dried over anhydrous $MgSO_4$ and concentrated to afford 7 mg of **S11** as yellow oil (85% yield).

2.5.11 Synthesis of S12. Dimethyl sulfide (Me_2S, 20 µL) was added to a stirred solution of NCS (6 mg) in anhydrous CH_2Cl_2 (1.5 mL) under N_2 at -40 °C and the solution was stirred for 5 min. The mixture was warmed to 0 °C over 15 min before cooling it back to -40 °C. A solution of the azido-alcohol **S11** (7 mg) in anhydrous CH_2Cl_2 (1.5 mL) was then added and the solution was gradually warmed to 0 °C over 30 min. Stirring was continued at 0 °C for 1.5 h. Subsequently, the solution was stirred in a room temperature water bath for 15 min before quenching with brine (5 mL). The aqueous layer was extracted with pentane (3x10 mL). The combined organic layer was dried over $MgSO_4$ and the solvent was removed *in vacuo* to afford 4.6 mg of azido-geranylgeranyl chloride **S12** as a yellow oil that was used without further purification.

2.5.12 Synthesis of C20-Azi (8). Compound **S12** (4.6 mg) was dissolved in dry CH_3CN (1.0 mL) and was added drop wise to a stirred solution of $[(nBu)_4N)]_3HP_2O_7$ (45 mg) in anhydrous CH_3CN (1.0 mL) under N_2. The mixture was stirred at room temperature for 18 h. The solvent was removed *in vacuo* at 25 °C. An ion-exchange column with DOWEX 50Wx8 resin (100-200 mesh, H^+ form) was used to convert the product to its ammonium form. The resin was packed and washed using three column volumes of H_2O/conc. NH_4OH (3:1, v/v) until the elution solvent became basic, followed by equilibration with three column volumes of NH_4HCO_3 (25 mM)/*i*-PrOH (49:1, v/v; solvent A). The product was applied to the column dissolved in a minimal amount of solvent A and was eluted with the same solvent. The first fraction collected (yellow) was lyophilized for 15 h and then purified by preparative RP-HPLC using the following conditions: detection: 220 nm; flow rate: 5.0 mL/min; 2 mL injection loop; gradient 0–100% solvent B in 20 min; solvent A: 25 mM NH_4HCO_3, solvent B: CH_3CN. Compound **8** eluted at approximately 23 min. Lyophilization of the pooled fractions of the desired compound yielded a white powder (3.2 mg, 18% from **S11**). ^{1}H NMR (300 MHz, D_2O) δ 5.29 (t, *J* = 6.9 Hz, 1H), 4.89-4.87 (m, 2H), 4.84 (s, 1H), 4.81-4.80 (m, 2H), 4.29 (t, *J* = 6.3 Hz, 2H), 3.79 (dd, *J* = 13.2, 6.6 Hz, 1H), 3.54 (s, 1H), 2.04-1.82 (m, 11H), 1.54 (s, 6H), 1.51 (s, 3H), 1.45 (s, 3H); ^{31}P NMR (121 MHz, D_2O) δ -6.02 (d, *J* = 22.0 Hz, 1P), -9.86 (d, *J* = 22.2 Hz, 1P); *m/z* (ESI-MS) found 490 ($[M-H]^-$), expected 490.

2.5.13 Enzymatic studies of C10-Aoxy (2) using a continuous fluorescence assay. Enzymatic reaction mixtures contained Tris·HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), KCl (20 mM), $ZnCl_2$ (10 μM), DTT (5 mM), 2.4 μM Dansyl-GCVIA (**12a**), 0.04% (w/v) *n*-dodecyl-β-d-maltoside, 80 nM yPFTase, and varying concentrations of **2** (0–50 μM) in a final volume of 250 μL. Reaction mixtures were equilibrated at 30 °C for 5 min, initiated by addition of yPFTase, and monitored for an increase in fluorescence (λ_{ex} = 340 nm, λ_{em} = 505 nm) for approximately 10 min. The initial rates of formation of products were obtained as slopes in IU/min using least-squares analysis. Corrections were applied to all of the rate calculations based on the difference between the fluorescence intensity of the prenylated product and the starting peptide. Assuming 100% conversion, the difference corresponds only to the fluorescence of the total amount of product. The slope was then

divided by the fluorescence difference followed by multiplying by the total concentration of peptide (2.4 μM), which then gave the rate of formation of product in μM/s. It should be noted that the K_M values reported here are actually apparent K_M values, since the measurements were performed at only a single peptide concentration. The data were fit to a Michaelis–Menten model using a nonlinear regression program to determine k_{cat} and K_M.

2.5.14 Enzymatic incorporation of compound 2 into GFP-CVIA (S13). Since GFP-CVIA (**S13**) protein does not have polyhistidine-tag and was prepared as described before.[65] Enzymatic reactions (10 mL) contained Tris·HCl (50 mM, pH 7.0), $MgCl_2$ (10 mM), $ZnCl_2$ (10 μM), DTT (5 mM), GFP-CVIA **(S13**, 2.0 μM), **2** (150 μM), and yPFTase (20 nM). After incubation at 30 °C for 2 h, LC-MS analysis showed (near) complete conversion to the product. Next, the reaction mixture was concentrated by using an Amicon Ultra-15 centrifugal filter (10kDa cutoff, Millipore), and excess **2** was removed using a NAP-5 column (Amersham).

2.5.15 Enzymatic prenylation of glucose-dependent insulinotropic polypeptide modified with C-terminal CVIM (GIP-CVIM, S15) using C10-Aoxy (2). GIP-CVIM (**S15**) was prepared as described before.[65] Enzymatic reactions (2 mL) contained Tris·HCl (50 mM, pH 7.0), $MgCl_2$ (10 mM), $ZnCl_2$ (10 μM), DTT (5 mM), GIP-CVIM (**S15**, 2.0 μM), **2** (150 μM), and yPFTase (20 nM). After incubation at 30 °C for 2 h, LC-MS analysis showed (near) complete conversion to the product. Next, the reaction mixture was concentrated using an Amicon Ultra-15 centrifugal filter (3kDa cutoff, Millipore), and excess **2** was removed using a NAP-5 column (Amersham).

2.5.16 DNA sequence for GFP-CVIA (10a)

ATGGCTAGCCACCACCACCATCATCACGTAAGCAAAGGCGAAGAACTGTTCA
CCGGCGTAGTGCCGATTCTGGTGGAATTGGATGGTGACGTTAACGGTCACAA
GTTTAGCGTTTCGGGTGAGGGCGAGGGTGATGCAACGTATGGTAAACTGACC
TTGAAATTCATTTGTACTACCGGTAAACTGCCGGTCCCGTGGCCGACGCTGGT
TACCACCCTGACCTATGGCGTCCAGTGCTTCAGCCGCTATCCAGACCACATGA
AACAGCATGATTTCTTTAAGTCCGCGATGCCGGAAGGCTACGTGCAAGAACG
CACTATCTTCTTCAAAGATGACGGTAATTACAAGACCCGTGCGGAGGTTAAG
TTTGAGGGCGACACCCTGGTTAATCGTATTGAGCTGAAGGGCATTGATTTCAA
AGAGGACGGCAATATTCTGGGCCATAAACTGGAGTACAATTACAATAGCCAC

AACGTCTACATCATGGCGGATAAGCAGAAAAACGGTATCAAAGTCAACTTTA
AGATCCGTCATAACATCGAAGATGGTTCTGTCCAACTGGCCGACCACTACCA
ACAGAACACCCCTATCGGTGACGGTCCGGTTCTGTTGCCGGACAATCATTATC
TGAGCACGCAAAGCGCCCTGAGCAAGGACCCGAACGAGAAACGTGATCACA
TGGTGCTGCTGGAGTTTGTGACGGCAGCGGGTATTACGCTGGGTATGGACGA
ACTGTATAAGTGCGTGATCGCATAAAAGCTT

2.5.17 DNA sequence for RFP-CVIA (13a)

ATGGTGTCTAAGGGCGAAGAGCTGATTAAGGAGAACATGCACATGAAGCTGT
ACATGGAGGGCACCGTGAACAACCACCACTTCAAGTGCACATCCGAGGGCG
AAGGCAAGCCCTACGAGGGCACCCAGACCATGAGAATCAAGGTGGTCGAGG
GCGGCCCTCTCCCCTTCGCCTTCGACATCCTGGCTACCAGCTTCATGTACGGC
AGCAGAACCTTCATCAACCACACCCAGGGCATCCCCGACTTCTTTAAGCAGT
CCTTCCCTGAGGGCTTCACATGGGAGAGAGTCACCACATACGAAGACGGGGG
CGTGCTGACCGCTACCCAGGACACCAGCCTCCAGGACGGCTGCCTCATCTAC
AACGTCAAGATCAGAGGGGTGAACTTCCCATCCAACGGCCCTGTGATGCAGA
AGAAAACACTCGGCTGGGAGGCCAACACCGAGATGCTGTACCCCGCTGACG
GCGGCCTGGAAGGCAGAAGCGACATGGCCCTGAAGCTCGTGGGCGGGGGCC
ACCTGATCTGCAACTTCAAGACCACATACAGATCCAAGAAACCCGCTAAGAA
CCTCAAGATGCCCGGCGTCTACTATGTGGACCACAGACTGGAAAGAATCAAG
GAGGCCGACAAAGAGACCTACGTCGAGCAGCACGAGGTGGCTGTGGCCAGA
TACTGCGACCTCCCTAGCAAACTGGGGCACAAACTTAATGGTTGCGTAATCG
CTTGA

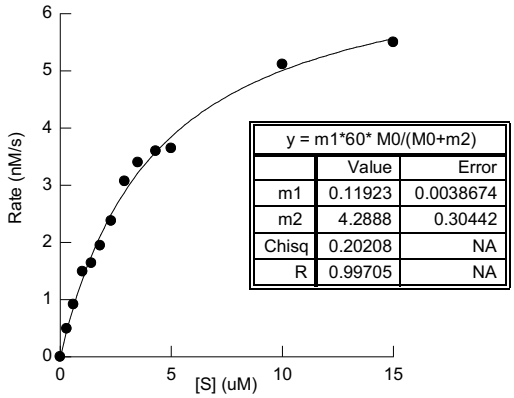


Figure S2.1 Fluorescence-based PFTase enzyme assay for prenylation of model peptide Dansyl-GCVIA (**12a**) using varying concentrations of **2**. The K_M is calculated to be 4.29 μM and the k_{cat} is calculated to be 0.119 s^{-1}, compared to the K_M of 1.71 μM and k_{cat} of 0.52 s^{-1} for FPP.[65]

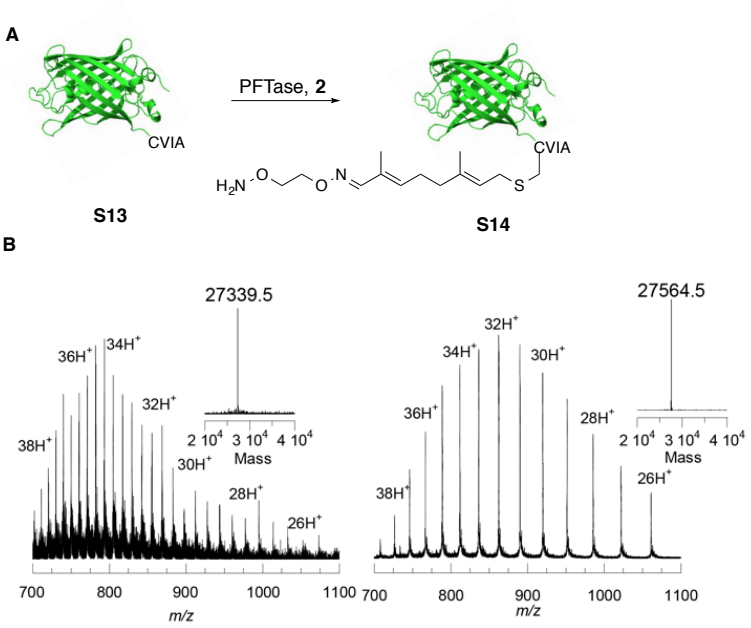


Figure S2.2 (A) Schematic representation of prenylation of GFP-CVIA (**S13**) with **2** by yeast PFTase. (B) ESI-MS spectra of **S13** (left) and aminooxy-functionalized **S14** (right) with the deconvoluted mass spectra shown in the insets.

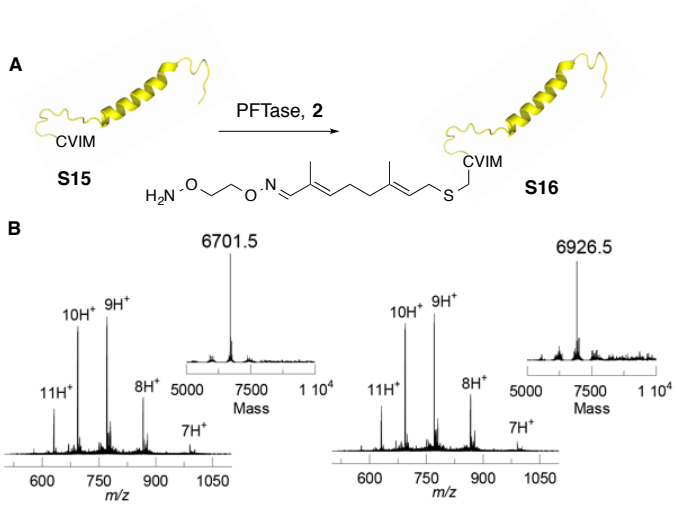


Figure S2.3 (A) Schematic representation of prenylation of GIP-CVIM (**S15**) using compound **2** by yeast PFTase. (B) ESI-MS spectra of **S15** (left) and aminooxy-functionalized (**S16**) (right) with the deconvoluted mass spectra shown in the insets.

Table S2.1 Summary of the results from initial continuous spectrofluorimetric assays to evaluate isoprenoid analogues as possible substrates for rGGTase-I using dansyl-GCVLL (**11a**).

Compound	Results
C10-BK (**3**)	Dansyl fluorescence intensity increased after the addition of rGGTase-I indicating **3** as a substrate for rGGTase-I. However, strong substrate inhibition was observed at high analogue concentration (>5 uM).
C10-Alk (**5**)	No changes in dansyl fluorescence intensity were observed after the addition of rGGTase-I suggesting that **5** is not a good substrate for rGGTase-I. It should be noted that C10-Alk gives a fluorescence increase when PFTase is used with the substrate dansyl-GCVIA suggesting that if it was incorporated into dansyl-GCVLL, a fluorescence increase should occur.
C20-Azi (**8**)	No changes in dansyl fluorescence intensity were observed after the addition of rGGTase-I. However, dansyl-GCVLL modified with **8** was detected using HPLC. That result suggests that the fluorescence increase with this longer isoprenoid may not be as pronounced as that observed with the alkyne-containing analogues.

Table S2.2 Prenylated products detected by LC-MS for the reactions of rPFTase with GFP-CVLL (**9a**) and of rGGTase-I with GFP-CVIA (**10a**) using isoprenoid analogues **3** and **5**.

Enzyme	Protein Substrate	Isoprenoid Substrate	Calculated Mass	Observed Mass
rGGTase-I	GFP-CVIA	C10-BK (**3**)	28456	No product
		C10-Alk (**5**)	28348	No product
rPFTase	GFP-CVLL	C10-BK (**3**)	28498	No product
		C10-Alk (**5**)	28390	28389

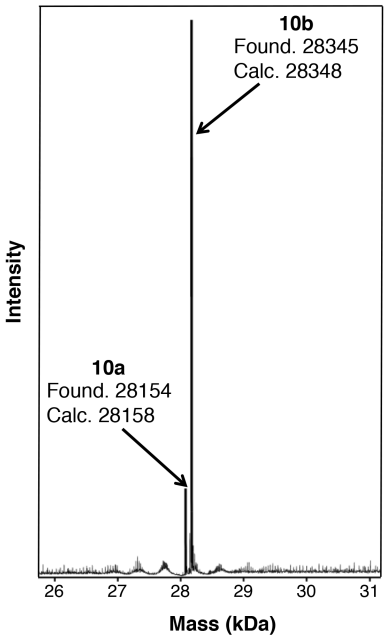


Figure S2.4 LC-MS result of prenylation reaction with GFP-CVIA (**10a**) and C10-Alk (**5**) using rPFTase. The reaction efficiency was estimated to be 86%. It should also be noted that the obtained mass by LC-MS of the same protein samples prepared at different times was varied by 1 to 2 units each time, which is likely limited by the resolution of the instrument.

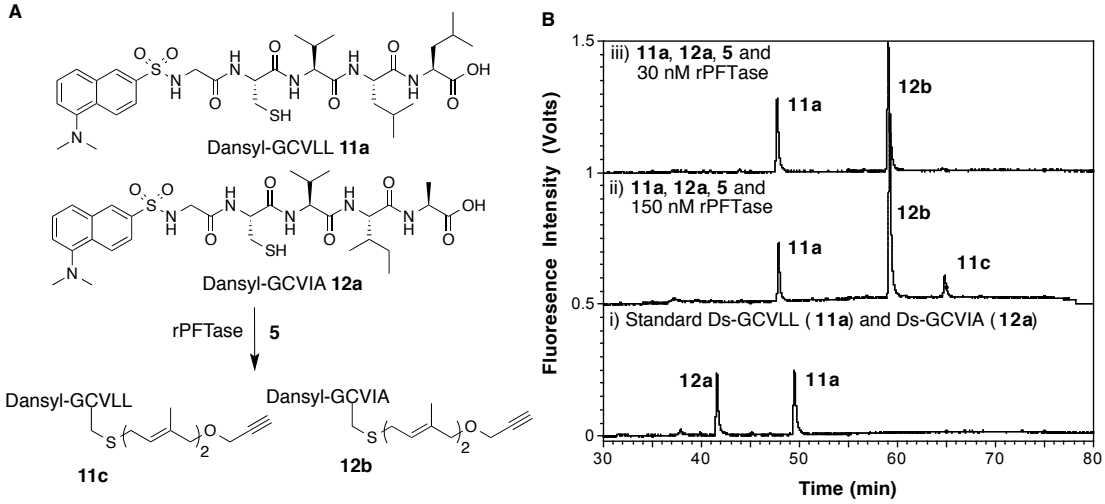


Figure S2.5 Prenylation reactions of rPFTase monitored by HPLC. (A) Scheme for the dual labeling reaction of Ds-GCVLL (**11a**) and Ds-GCVIA (**12a**) with isoprenoid analogue **5** in the presence of rPFTase. All the possible prenylated products are shown. (B) HPLC traces of prenylation reactions for rPFTase. i). Unmodified **11a** and **12a**. ii). Prenylation reaction of **11a** and **12a** with **5** in the presence of 150 nM rPFTase. iii). Prenylation reaction for **11a** and **12a** with **5** in the presence of 30 nM rPFTase. By reducing the concentration of rPFTase, analogue **5** was only incorporated onto Ds-GCVIA (**12a**), with no product formed with Ds-GCVLL (**11a**).

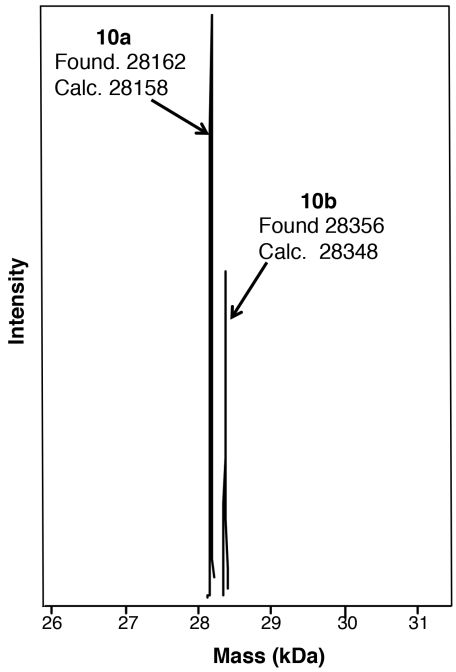


Figure S2.6 LC-MS result of GFP-CVIA (**10a**) labeling with C10-Alk (**5**) by rPFTase in the dual labeling reaction containing GFP-CVLL (**9a**), GFP-CVIA (**10a**), C20-Azi (**8**), C10-Alk (**5**), rGGTase-I and 30 nM rPFTase. The labeling efficiency was estimated to be 36%.

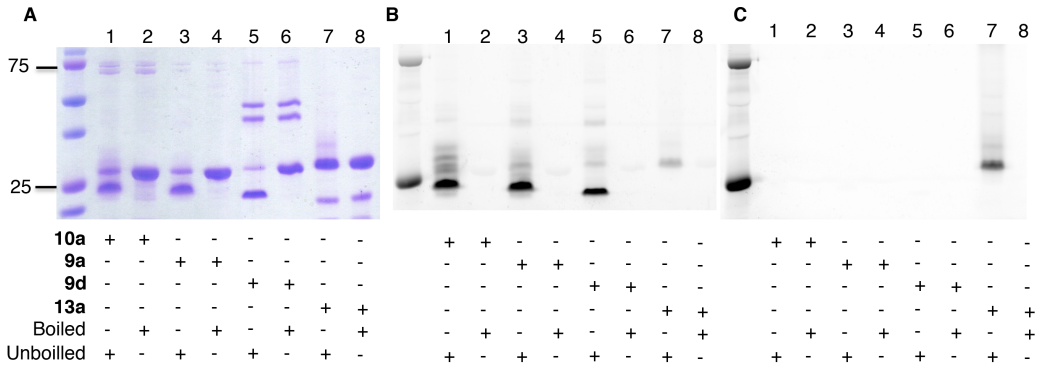


Figure S2.7 SDS-PAGE and in-gel fluorescence imaging analysis of boiled and unboiled protein samples. Protein samples were prepared by mixing with equal volumes of 2X SDS-PAGE loading buffer. (A). Coomassie blue stained gel. (B) In-gel fluorescence imaging for GFP fluorescence. (C) In-gel fluorescence imaging for RFP fluorescence. Fluorescence signals were only detected from unboiled protein samples. Lane 1: unboiled GFP-CVIA (**10a**). Lane 2: boiled GFP-CVIA (**10a**). Lane 3: unboiled GFP-CVLL (**9a**). Lane 4: boiled GFP-CVLL (**9a**). Lane 5: unboiled azide-modified GFP-CVLL (**9d**). Lane 6: boilded azide-modified GFP-CVLL (**9d**). Lane 7: unboiled RFP-CVIA (**13a**). Lane 8: boiled RFP-CVIA (**13a**). It was found that the mobility of boiled and unboiled RFP-CVIA (**13a**) samples was the same but the apparent size of GFP changes upon heat denaturation.

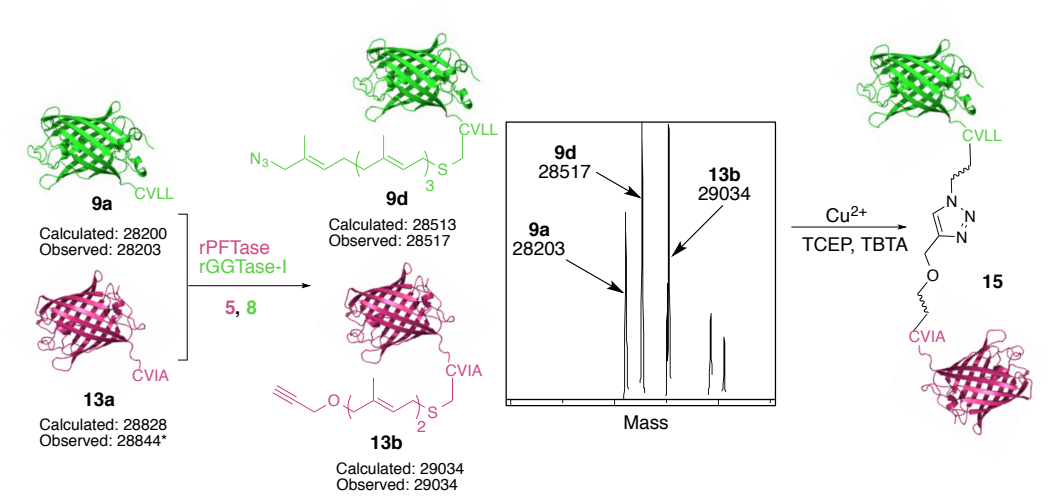


Figure S2.8 Scheme for simultaneous dual labeling of GFP-CVLL (**9a**) and RFP-CVIA (**13a**) with C20-Azi (**8**) and C10-Alk (**5**) using rGGTase-I and rPFTase, followed by the CuAAC reaction to produce GFP-RFP heterodimer. *The mass for **13a** was the observed mass. It should be noted that the observed mass (28844 Da) for **13a** was different from the calculated mass (28828 Da) by 16 units, which is likely due to the oxidation of one of its methionine residues.

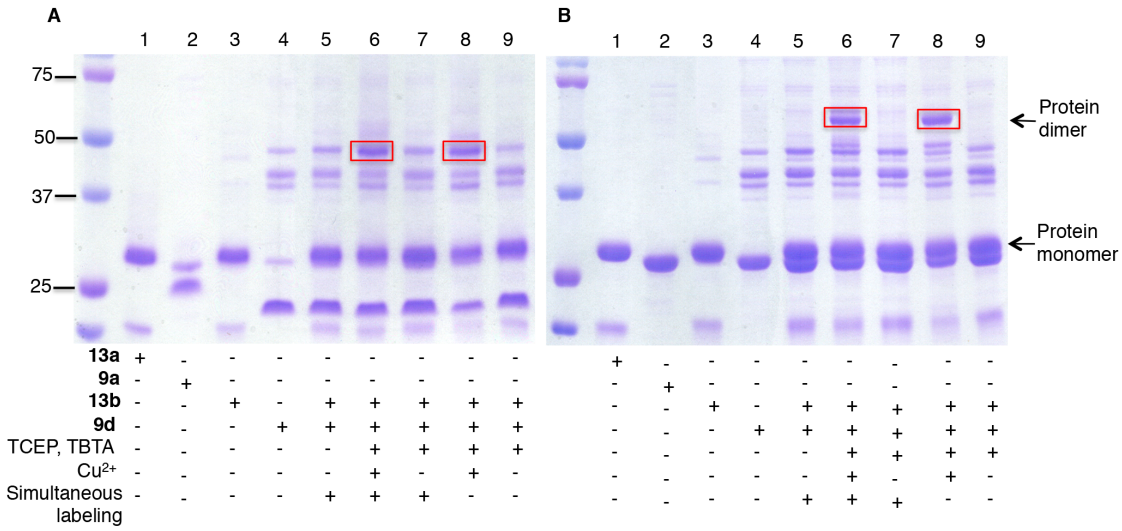


Figure S2.9 SDS-PAGE analysis of GFP-RFP dimer. (A) Coomassie blue stained gel of protein samples without heating. (B) Coomassie blue stained gel of heated protein samples. Lane 1: unmodified RFP-CVIA (**13a**). Lane 2: unmodified GFP-CVLL (**9a**). Lane 3: **13b** (produced from reaction of **13a**, **5** and rPFTase). Lane 4: **9d** (produced from reaction of **9a**, **8** and rGGTase-I). Lane 5: simultaneously labeled **13b** and **9d**. Lane 6: the CuAAC reaction mixture of simultaneously labeled **13b** and **9d** with TBTA (100 μM), TCEP (1 mM) and $CuSO_4$ (1 mM). Lane 7: the CuAAC reaction control for the simultaneously labeled **13b** and **9d** with no $CuSO_4$ added. Lane 8: the CuAAC reaction mixture of separately labeled **13b** and **9d**. Lane 9: the CuAAC reaction control for the separately labeled **13b** and **9d** with no $CuSO_4$ added. The dimer band was highlighted in the red box.

Chapter 3. Enzymatic Construction of DARPin-Based Targeted Delivery Systems Using Protein Farnesyltransferase

The creation of protein-based diagnostic and therapeutic materials is a rapidly growing area of biomedical research. When preparing homogeneous protein conjugates, site-specific modification methods and efficient purification strategies are both critical factors to be considered. As a highly efficient and specific labeling method, enzymatic protein modification has been widely applied to create protein conjugates. In terms of purification, current methods mostly involve complicated chromatographic separations, which require extensive method optimization and result in material loss. Development of a facile conjugation and purification strategy is extraordinary desirable. In an effort to provide a solution to this problem, our lab has previously described a technique to allow one-step preparation and purification of protein-based conjugates. In this technique, the protein of interest is first labeled by the enzyme protein farnesyltransferase (PFTase) with an aldehyde functionality, followed by immobilization onto hydrazide-beads *via* hydrazone bond formation. Simple washing of the beads removes the PFTase enzyme and the unmodified target protein. To create the desired protein conjugates, aminooxy-functionalized cargos, such as fluorophore molecules or PEG polymers are then added to the beads to release the modified protein and form the protein conjugates, enabling one-step conjugation and purification. Here, we apply this technique to create protein conjugates based on the Designed Ankyrin Repeat Proteins (DARPins), which are antigen-binding proteins that can serve as the targeting moieties specific for cancer cells. The DARPins were first engineered with a C-terminal CVIA sequence as the enzyme recognition site and then labeled with an aldehyde functional group by PFTase. Conjugation of the modified DARPins to TAMRA fluorophore was achieved using the capture and release strategy described above starting from purified proteins or from proteins present in crude *E. coli* lysate. Notably, the DARPins retained binding ability and specificity after modification, as confirmed by flow cytometry and confocal imaging analysis. DARPin-drug conjugates were also prepared and their cytotoxicity was evaluated in cell culture. Limited by the commercial availability of the drug-containing reagents, the production process is not efficient enough, and further optimization is in progress. The method described here is broadly applicable to enable streamlined one-step construction and purification of functional protein-based conjugates without complicated chromatographic methods.

3.1 Introduction

In modern medicine, a variety of protein-based conjugates have been created and extensively applied for therapeutic and diagnostic applications. For example, the attachment of a PEG polymer to a therapeutic protein has been widely employed to enhance the pharmacokinetics of the polypeptide.[10] As a promising therapy for cancer treatment, antibody-drug conjugates (ADCs) are created by linking a small-molecule cytotoxin to the antibody so that the toxic drug can be selectively directed to the tumor cells, thereby reducing undesired toxicity to healthy tissues.[11] Similarly, incorporation of a metal chelator into an antibody or a derivative allows specific detection of tumorous tissue when a radio-isotope is loaded.[350] Due to the high demand for protein-based conjugates, a number of conjugation methods have been developed including enzymatic labeling methods.[18] Taking advantage of the specificity of the enzymes, highly homogeneous product can be obtained. However, the introduction of an enzyme into the reaction mixture in addition to the target protein complicates the purification process. Moreover, the conjugation reaction may not achieve complete conversion, making it necessary to purify the protein conjugate from unmodified starting material. In cases where the molecule attached to the protein is small, such as cytotoxic drug molecules, purification can be very challenging since the physical properties of the modified protein differs only marginally compared to the unconjugated one. As a result, sophisticated chromatographic methods are necessary, which requires extensive optimization to achieve optimal separation. In an effort to improve the purification process, solid microbeads were employed and functionalized with microbial transglutaminase (MTG).[215] It was shown that the immobilized MTG was able to catalyze the conjugation of target proteins, which could be easily filtered away from the target protein after reaction. However, this method still does not address the problem of purifying the conjugates from unreacted starting material. To streamline the construction of protein conjugates and improve overall yield, a facile conjugation and purification strategy is highly desirable.

One of the more recent combined conjugation and purification strategies involves the use of sortase A (SrtA) enzyme.[187,190] Such methods utilize a genetic fusion construct consisting of the target protein, followed by a sortag (for SrtA recognition) and an

additional tag for purification, such as a His-tag. When the desired cargo is attached to the target protein by SrtA, the sequence after the sortag is cleaved off. Therefore, the conjugate can be easily separated from the unmodified protein based on the properties of the additional tag that is still attached to the unmodified protein. As a specific example, Policarpo et. al reported a flow-based SrtA ligation technique.[188] A micro-reactor was created by packing Ni-NTA resin pre-loaded with His-tagged SrtA enzyme. When the protein of interest containing C-terminal sortag and a His-tag was added, it would be first absorbed onto the resins through the His-tag. In the presence of the oligoglycine substrates for SrtA, the target protein would be modified and eluted from the resin, leaving the unreacted starting materials still attached to resin. Although the design for the SrtA-based method is highly efficient, the conjugation process is still limited by the slow reaction kinetics and the moderate labeling yield of the wild-type enzyme.[18] Thus, high concentrations of the enzyme or the peptide substrates are usually required to achieve an acceptable yield. Although a penta-mutant variant with enhanced catalytic activity has been evolved,[97] it was shown to be associated with increased hydrolysis in some cases.[192] In addition, when a solid support is utilized, optimization is still needed to minimize the formation of undesired side-products.[189]

As an alternative enzymatic labeling technique, protein farnesyltransferase (PFTase) has also been applied to create site-specific protein conjugates.[351] The enzyme catalyzes the transfer of an isoprenoid group from its native substrate, farnesyl diphosphate (FPP), to a C-terminal cysteine of the protein substrate. The recognition sequence for PFTase, consisting of a minimal four amino acid sequence, is called the CaaX box, where C is the cysteine being modified. It has been shown that the incorporation of a C-terminal CVIA sequence into a given protein of interest makes it a recognizable substrate for the enzyme, enabling efficient modification. In addition, a panel of isoprenoid analogues containing various bioorthogonal functional groups, including azide,[245] alkyne,[117] aldehyde[65] and transcyclooctene[95] groups have been designed and synthesized, which can be used in lieu of FPP for conjugation of the protein. In particular, our lab has previously reported a capture and release strategy based on the aldehyde functionality that can be used to streamline the process of protein conjugation and purification.[65] In this technique, the protein of interest is first labeled with an aldehyde functional group and immobilized onto

hydrazide beads. After washing the beads thoroughly, PFTase enzyme and the unmodified proteins can be removed. Release of the modified protein and conjugation to the desired cargos are achieved by the addition of aminooxy-functionalized cargos, enabling one-step conjugation and purification. Proof-of-concept studies were conducted to modify GFP and a glucose-dependent insulinotropic polypeptide (GIP) without any functional characterization. In this work, we applied the combined PFTase labeling technique and the capture and release strategy to facilitate the construction of protein conjugates based on Designed Ankyrin Repeat Proteins (DARPins). Derived from natural ankyrin repeat proteins, DARPins are alternative binding scaffolds that can be engineered and selected *in vitro* for specific target recognition with high affinity.[352] Compared to conventional antibodies, DARPins are small, highly stable and can be easily obtained from bacterial culture expression in high yield without complex glycosylation states. These favorable binding and physical properties make them attractive for potential targeted drug delivery or imaging applications. In this study, we first applied the PFTase labeling method to modify several DARPin constructs engineered with C-terminal CVIA sequences with an aldehyde functionality. Subsequent conjugation to a fluorophore or cytotoxic drug using the capture and release strategy efficiently produced functional DARPin-conjugates, whose binding ability and specificity were retained. Since both the PFTase labeling technique and the capture and release strategy are compatible with cell lysates, the method described here can be readily adopted for the facile and efficient construction of functional protein-based conjugates avoiding complicated chromatographic methods.

3.2 Materials and Methods

3.2.1 Enzymatic prenylation of DARPins.

Prenylation reactions were performed in buffer containing: 50 mM Tris·HCl, 20 mM KCl, 10 mM $MgCl_2$, 5 mM DTT, 10 μM $ZnCl_2$. The solution was incubated on ice for 0.5 h including 2.5 μM DARPins. After incubation, 15 μM isoprenoid analogues were added to the solution, followed by the addition of 200-500 nM yeast PFTase. The reaction was carried out in a water bath at 32 ºC for 6 h. LCMS characterization was performed with the reaction mixture directly without further purification. To remove excess

isoprenoid analogues, buffer exchange using PBS was conducted three times with Amicon filters (3K cut-off).

3.2.2 DARPin-BA conjugation to TAMRA-aoxy.

Oxime ligation reactions were performed with 20 to 50 μM DARPin-BA and 5 eq of **2** at RT for 6 h on a rotary shaker. No catalyst was employed. A NAP-5 column was used to remove excess **2** from the reaction mixture.

3.2.3 Capture and release strategy to construct DARPin-TAMRA conjugates.

To immobilize the aldehyde-modified DARPins, 20 eq of hydrazide beads (15 μmol/mL), pre-equilibrated with 100 mM PB, pH 6.4, were added to the protein solution. The reaction was carried out in 100 mM PB, pH 6.4 with 100 mM aniline as a catalyst. After 2 h incubation on a rotary shaker at RT, the reaction mixture was centrifuged to remove the supernatant. The beads were washed three times with 300 mM PB, pH 6.4 and three times with 1 M NaCl (10-fold excess of volume to that of the beads). Release of DARPin-TAMRA was conducted in 100 mM PB, pH 6.4, with 10 eq of **2** and 100 mM aniline. The reaction was placed on a rotary shaker for overnight incubation. A NAP-5 column was employed to remove excess TAMRA-aoxy from the conjugates.

3.2.4 Flow cytometry analysis of D1-TAMRA binding to cell surface EpCAM.

PBSA solution was prepared with PBS and 1 mg/mL BSA. Three different cell lines were analyzed including MCF-7 cells, HT-29 cells and U87-MG cells. The cells were harvested and washed twice with PBSA. $1x10^6$ cells/mL of cells (400 μL) were incubated with 100 nM **D1**-TAMRA and **D4**-TAMRA respectively, which were diluted in PBSA, in the dark and left for 45 min at 4 ºC. After incubation, the cells were washed twice with PBSA before analysis. For the competition experiments, the cells were first incubated with 10 μM unlabeled **D1** at 4 ºC for 15 min. Experiments were conducted using a BD LSR II/Fortessa H0081 flow cytometer and $1x10^4$ cells were counted. Data analysis was performed with FlowJo software (v10).

3.2.5 Visualization of D1-TAMRA binding and internalization to MCF-7 cells.

$1x10^5$ MCF-7 cells were seeded onto sterile glass cover slips and incubated for 24 h. The cells were then supplied with fresh DMEM containing 100 nM **D1**-TAMRA or **D4**-TAMRA and incubated for 1 h in the dark at 4 ºC or 37 ºC. After incubation, the cells were washed three times with PBS (three min each time) and then fixed with 4% paraformaldehyde, followed by washing once with PBS. The nuclei of the cells were stained with 0.5 μg/mL Hoechst dye for 10 min at RT. After washing with PBS, the cover slips were mounted on glass slides using SlowFade Diamond Antifade Mountant (Thermo Fisher Scientific) and sealed with nail polish. Images were obtained using an Olympus FluoView FV1000 BX2 Upright Confocal microscope. Analysis were performed with FluoView software.

3.2.6 Serum stability of D1-TAMRA *in vitro*.

Human and mouse plasma (lithium heparin plasma, bioreclamationIVT) were diluted in PBS and filtered through a 0.2 μm filter. **D1**-TAMRA was then added to 50% plasma to a final concentration of 1 μM. Aliquots of 160 μL were made and incubated in a 37 ºC water bath. For the human plasma stability test, incubation times of 0, 2, 4, 8, 24 and 48 h were conducted. To examine the stability in mouse plasma, samples were incubated for 0, 0.5, 1, 2, 4, 12 h. Samples were removed from the water bath and frozen at -20 ºC at the desired time points. Since D7 has a His-tag at the N-terminus, 12 μL Ni-NTA resin, pre-equilibrated with PBS containing 20 mM imidazole, were added to the thawed serum sample to capture the protein. The reaction vials were placed on a rotary shaker for 1 h in the dark at 4 ºC. The supernatant was removed after centrifugation and the beads were washed twice with PBS containing 20 mM imidazole. To release the bound protein, 20 μL of Laemmli buffer with 1 M imidazole was added to the beads. The mixture was incubated at RT for 10 min and then heated at 95 ºC for 10 min. The solution was loaded onto an SDS-PAGE gel and subjected to electrophoresis. Both experiments were performed in triplicate. In-gel fluorescence scanning was performed using a Typhoon FLA 9500 biomolecular imager. The images were analyzed with ImageJ software and plotted in KaleidaGraph.

3.2.7 Construction of DARPin-MMAE conjugates.

The aldehyde modified DARPin proteins were immobilized onto hydrazide beads as described above. After washing, the proteins were eluted from the beads using 10 eq of **3** (BroadPharm) in 100 mM PB, pH 6.4 in the presence of 100 mM aniline to afford the DARPin-oxime-azide product. After overnight release, excess **3** was removed using a PD SpinTrap G-25 column (GE Life Science). PBS was employed to elute the proteins from the column. Protein concentration was determined using a Bradford assay. To perform SPAAC reactions, 10 eq of **4** were added to the DARPin-oxime-azide solution. The reactions were carried out at RT for 6 h on a rotary shaker. A Superdex 75 100/30 GL column (GE Life Science) was used to remove excess **4** from DARPin-MMAE conjugates since simple desalting columns were not able to remove the impurities from **4**. Size exclusion chromatography was performed using a Knauer system employing PBS as the buffer. The flow rate was set as 0.4 mL/min. After purification, fractions containing DARPin-MMAE were concentrated using Amicon filters (3K cut-off). The concentration of the conjugate was determined using a Bradford assay.

3.2.8 DARPin-MMAE cytotoxicity in cell cultures.

To determine the cytotoxicity of DARPin-MMAE conjugates, an XTT assay was utilized. Three different cell lines were tested, including MCF-7, HT-29 and U87-MG. Cells were seeded into 96-well plates at a density of 2,000 cells/well in a volume of 50 µL. After a 24 h incubation, appropriate media containing serial dilutions of **D1**-MMAE or **D4**-MMAE was added so that the final volume in each well was 100 µL. Competition with unmodified **D1** was performed by first incubating the cells with 10-fold excess unmodified DARPins for 0.5 h and then adding **D1**-MMAE. After a 96 h incubation, the media was removed from the plates followed by the addition of 50 µL of XTT reagent (Roche Diagnostic). Cells were incubated for 0.5 to 2 h before analysis. Absorbances at 480 nm and 650 nm (reference wavelength) were measured using a Tecan Infinite M200 PRO plate reader. Readings from untreated cells were used for normalization. The data was analyzed in Excel and plotted in KaleidaGraph.

3.3 Results and Discussion

3.3.1 DARPin modification by PFTase.

Ac2 is a DARPin selected for its ability to bind to epithelial cell adhesion molecule (EpCAM) with low nanomolar affinity (biphasic dissociation was observed with K_{d1} of 2.2nM and K_{d2} of 46 nM).[353] Overexpression of EpCAM is observed on a variety of epithelium-derived tumors, including breast, pancreatic and colorectal carcinoma, making it an attractive target for cancer diagnosis and therapy.[354] To allow for site-specific modification, the C-terminal tetrapeptide sequence, CVIA, was genetically appended to Ac2. To investigate whether additional spacer sequences are needed for enzyme recognition, three constructs were designed and prepared: Ac2-KCVIA (**D1**) with the insertion of a single lysine residue, Ac2-GSGTKCVIA (**D2**) containing a flexible linker, and Ac2-GGKKKKKKTKCVIA (**D3**), whose polylysine spacer was adopted from the C-terminal sequence of the K-Ras protein, a native substrate for PFTase.[355] Prenylation reactions were performed with each of the protein constructs and the natural isoprenoid substrate, farnesyl diphosphate (FPP). The crude reaction mixtures were characterized by LC-MS equipped with a UV-vis detector. Based on the LC-MS results for each reaction, a major product was identified to be the desired modified Ac2 protein as the observed mass matched that of the calculated values (Figure S3.1). Therefore, all three proteins were labeled by PFTase successfully.

In addition to the product peak, a small peak eluted before the modified DARPins was also detected in the LC chromatogram of each reaction. The mass of this small peak could not be deconvoluted due to low signal intensity (Figure S3.2). We hypothesized that this peak belonged to the unmodified starting material, since the introduction of a relatively hydrophobic isoprenoid group typically causes longer retention times of prenylated molecules compared with their unreacted counterparts on a reverse-phase column. To confirm the identity of this unknown peak, the unmodified starting material from the stock solution of each Ac2 construct was spiked into the reaction mixture and analyzed by LCMS. As expected, the intensity for the unknown peak in each reaction increased (Figure S3.2), suggesting that it was indeed the unreacted proteins in the reaction mixture. The separation of the modified Ac2 from the unreacted starting material enables quantification of the

reaction yield by comparing the integrated UV absorbance of both species at 280 nm. It was found that 87% of **D1** was prenylated with FPP, while the yield was 86% for **D2** and 79% for **D3** (Figure S3.3). Therefore, all three constructs were efficiently modified by PFTase. To avoid potential non-specific interactions with the cell membrane from the highly positively charged poly-lysine residues in **D3**, **D1** was chosen for further investigation. As a negative control, E3_5, a stable DARPin that does not bind to any target was also engineered with a C-terminal KCVIA sequence (E3_5-KCVIA, **D4**) and was shown to be labeled with FPP by PFTase with almost complete conversion observed as by LC-MS (Figure S3.4).

3.3.2 Enzymatic incorporation of aldehyde functionality and conjugation to fluorophore.

Once the enzymatic labeling of **D1** and **D4** by PFTase using FPP was validated, they were then modified with aldehyde functionality using BAPP (**1**), an aldehyde-containing FPP analog (Figure 3.1). As shown in Figure 1C, successful labeling was achieved with both **D1** and **D4**. Quantification from the UV absorbance suggested the reaction yielded 88% conversion for **D1** and 95% for **D4**. To create protein-fluorophore conjugates, the modified proteins were reacted with an aminooxy-containing fluorophore, TAMRA-aoxy (**2**) using an oxime ligation reaction. Analysis from SDS-PAGE and in-gel fluorescence scanning showed a strong fluorescent band whose size corresponded to that of the DARPin proteins when screening for TAMRA fluorescence, indicating the desired conjugation of the fluorophore to DARPins (Figure 3.2). Notably, a slight decrease in mobility was observed when the proteins were modified by **1** and again after conjugation to TAMRA. LCMS further validated the mass of the formed DARPin-TAMRA conjugates (Figure 3.2C).

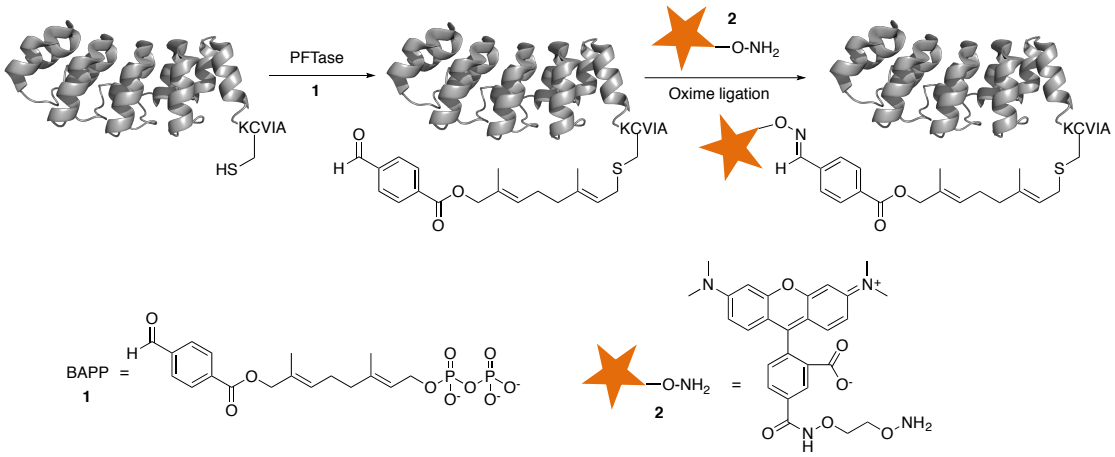


Figure 3.1 Site-specific enzymatic modification of DARPins by PFTase and subsequent conjugation to TAMRA-aoxy using an oxime ligation reaction.

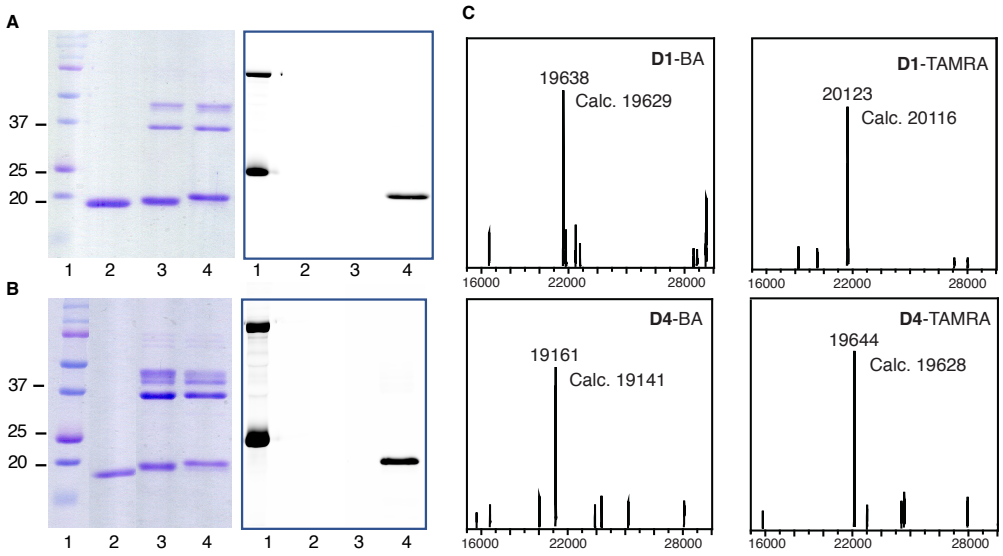


Figure 3.18 SDS-PAGE, in-gel fluorescence imaging and LCMS characterization of DARPin-BA and DARPin-TAMRA conjugates. (A) SDS-PAGE (left) and in-gel fluorescence imaging (right) of **D1**-BA and **D1**-TAMRA. Lane 1: protein ladder; lane 2: unmodified **D1**; lane 3: **D1**-BA reaction mixture; lane 4: **D1**-TAMRA reaction mixture. (B) SDS-PAGE (left) and in-gel fluorescence imaging (right) of **D4**-BA and **D4**-TAMRA. Lane 1: protein ladder; lane 2: unmodified **D4**; lane 3: **D4**-BA reaction mixture; lane 4: **D4**-TAMRA reaction mixture. (C). MS of the modified DARPins.

3.3.3 Capture and release strategy allowed facile construction of DARPin-TAMRA conjugates.

To purify the DARPin-conjugates prepared from oxime ligation reactions, size-exclusive chromatography (SEC) can be employed to remove the larger PFTase enzyme (90 kDa). However, it is challenging to remove any unreacted DARPins from the

conjugates since the mass increase due to fluorophore installation is minimal. To avoid multi-step chromatographic purification and complicated method optimization, a capture and release strategy was employed to create and purify DARPin-TAMRA conjugates simultaneously (Figure 3.3). In this method, the aldehyde modified DARPins were first immobilized onto hydrazide-functionalized beads, forming a covalent hydrazone bond. The beads were then extensively washed to eliminate the PFTase enzyme and the unmodified DARPin proteins. Since the oxime bond is more stable than the hydrazone bond, in the presence of excess amounts of aminooxy-containing compounds, the equilibrium will shift towards the formation of an oxime linkage. Therefore, with the addition of excess **2**, the immobilized DARPins were eluted from the hydrazide beads to form DARPin-TAMRA conjugates. Using this strategy, **D1**-TAMRA and **D4**-TAMRA were constructed successfully, as confirmed by in-gel fluorescence imaging analysis of the eluted material (Figure 3.4). Analysis of the supernatant from the capture step suggested most of the aldehyde-modified DARPins were attached to the beads, however, the release yield of the final DARPin-TAMRA conjugates was around 50%. To optimize the release reaction and improve overall yield, the amount of the added hydrazide beads was first evaluated (Figure S3.5). It was found that a 10-fold excess amount of the hydrazide on the beads compared to the concentration of aldehyde-DARPin was sufficient to capture most of the modified proteins. Additional optimization is currently underway to determine the optimal concentration of compound **2** to release the DARPin-fluorophore conjugates.

To further streamline the process protein conjugate preparation, we applied the capture and release strategy to construct DARPin-TAMRA conjugates directly from crude cell lysate. After the induction of **D1** overexpression in *E. coli*, the bacterial cells were harvested and lysed. The **D1** concentration in the lysate was estimated from Bradford assay and gel-based densitometry. The lysate was then used in a prenylation reaction with **1** and PFTase. Excess amounts of **1** were removed using centrifugal filters, followed by the addition of the hydrazide beads. After immobilization and washing, compound **2** was added to elute the DARPin and create the **D1**-TAMRA conjugate. As shown in Figure 3.4C, **D1**-TAMRA could be successfully produced from cell lysate with high purity, leaving the impurities behind in the supernatant during immobilization.

As a facile method to construct protein-based conjugates, the capture and release strategy eliminates the need of multiple challenging chromatography purification steps to isolate pure protein conjugates from unreacted starting materials. The streamlined process also improves the overall yield. Furthermore, the ability to create protein conjugates directly from bacterial cell lysate further simplifies the production process, which can be easily adapted to large-scale manufacturing.

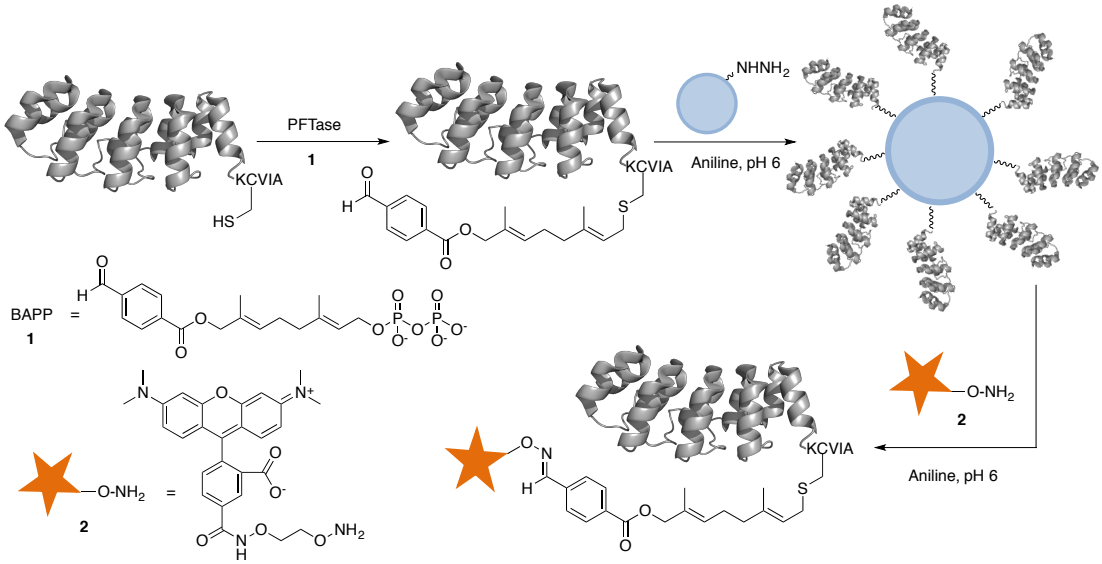


Figure 3.3 Capture and release strategy to construct DARPin-fluorophore conjugates.

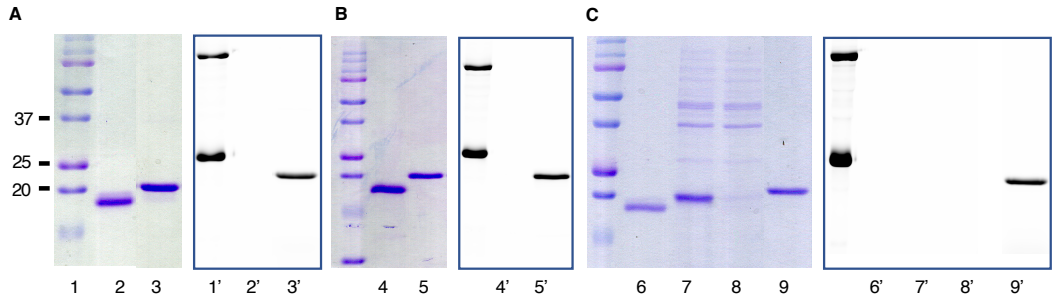


Figure 3.4 SDS-PAGE and in-gel fluorescence characterization of **D1**-TAMRA and **D4**-TAMRA constructed using capture and release strategy. (A) SDS-PAGE (left) and in-gel fluorescence (right) analysis of **D1**-TAMRA. Lane 1: protein ladder; lane 2: unmodified **D1**; lane 3: **D1**-TAMRA released from hydrazide beads. (B) SDS-PAGE (left) and in-gel fluorescence (right) analysis of **D4**-TAMRA. Lane 1: protein ladder; lane 2: unmodified **D4**; lane 3: **D4**-TAMRA released from hydrazide beads. (C) SDS-PAGE (left) and in-gel fluorescence (right) analysis of **D1**-TAMRA from crude cell lysate. Lane 1: protein ladder; lane 2: unmodified **D4**; lane 3: **D4**-BA prenylation mixture in crude lysate; lane 4: supernatant not immobilized onto hydrazide beads; lane 5: **D4**-TAMRA released from the beads.

3.3.4 D1-TAMRA retained specific binding to cell-surface EpCAM.

To investigate whether the modification interfered with the target binding capability of **D1**, flow cytometry experiments were performed using both EpCAM-positive MCF-7 and HT29 cells and EpCAM-negative U87-MG cells. The results showed that **D1**-TAMRA exhibited strong binding to the cell-surface EpCAM while no sign of non-specific absorption was observed when incubated with EpCAM-negative cells (Figure 3.5). Additionally, the binding of **D1**-TAMRA could be completely blocked by pre-incubation of the EpCAM-positive cells with the unmodified protein, further suggesting that the cell surface association was **D1** dependent. As for the non-targeting **D4**-TAMRA, no cell-surface attachment was detected regardless of the EpCAM expression level. Therefore, selective targeted binding of **D1** was not disrupted by the labeling and conjugation to TAMRA. Moreover, these results also demonstrated that the linker used to create the conjugates did not cause undesirable interaction with the cell surface. The binding of **D1**-TAMRA to MCF-7 cells and its subsequent internalization were then visualized by confocal microscopy (Figure 3.6). When incubated with the cells at 4 ºC, **D1**-TAMRA displayed clear cell surface localization. At 37 ºC, the fluorescent signal from **D1**-TAMRA could be observed inside the cells. Instead of a homogeneous cytosolic distribution, a punctate pattern was noted, indicating endosomal localization.

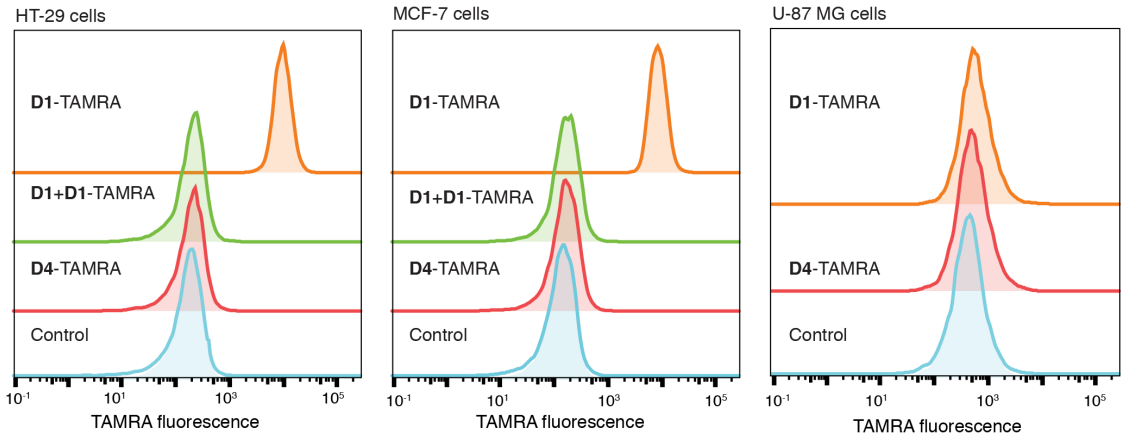


Figure 3.5 Evaluation of **D1**-TAMRA binding to cell-surface EpCAM by flow cytometry. HT29 and MCF-7 cell had EpCAM overexpression while U87-MG cells served as a negative control. **D1**-TAMRA was shown to bind to cell-surface EpCAM specifically, whose binding can be competed off using unmodified **D1**. No signs of non-specific absorption were observed for the non-targeting **D4**-TAMRA.

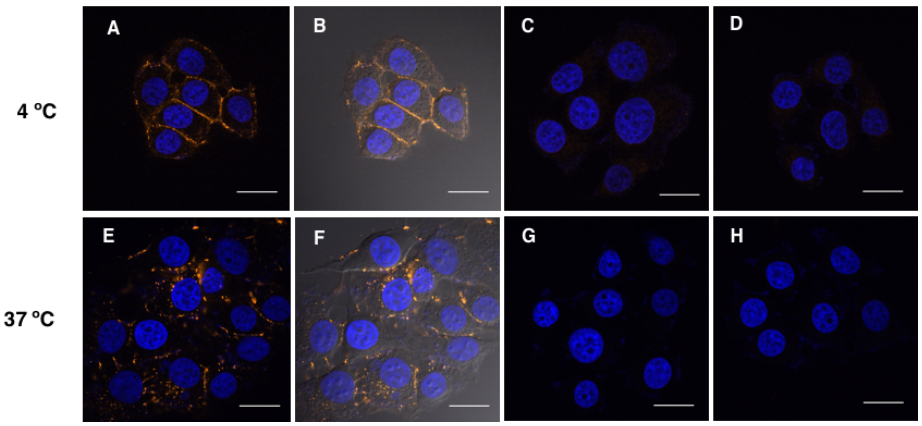


Figure 3.6 Visualization of binding (4 ºC) and internalization (37 ºC) of **D1**-TAMRA to MCF-7 cells by confocal microscopy. Cells were incubated with each construct for 1 h at 4 ºC and 4 h at 37 ºC respectively. (A) and (E) Cells treated with **D1**-TAMRA. (B) and (F) Bright field image of the cells treated with **D1**-TAMRA. (C) and (G) Cells treated with **D4**-TAMRA. (D) and (H) Cells with no treatment.

3.3.5 Application of PFTase labeling to an alternative DARPin protein.

With the successful construction and purification of the EpCAM binding DARPin **D1**, we were interested in evaluating whether the method is applicable to other DARPins. E01, which binds to the epidermal growth factor receptor (EGFR), a clinically validated cancer target, was selected and engineered to contain a C-terminal KCVIA sequence (E01-KCVIA, **D5**). We were able to label the protein with the aldehyde containing isoprenoid analogue **1**, followed by conjugation to a TAMRA fluorophore (Figure S3.6). Due to the inherent stability issue of **D5**, the crude reaction mixture was utilized to evaluate if the conjugated **D5** remained functional. We incubated **D5**-TAMRA with cells displaying various levels of total EGFR expression (MDA-MB-468 ++++, MDA-MB-231 +++ and MCF-7 +)[356] and analyzed the cells by flow cytometry (Figure 3.7). It was found that the fluorescence intensity detected from **D5**-TAMRA bound to cells correlated with their respective EGFR expression level, with the stronger signal corresponding to higher EGFR expression. The negative control, **D4**-TAMRA, did not show any non-specific binding towards any of the cell lines. These results confirmed that **D5**-TAMRA retained its binding specificity towards cell-surface EGFR after labeling by PFTase and conjugation to TAMRA. Overall, the results support the idea that as long as that the binding/active site is not located in close proximity to the C-terminus, proteins engineered with the enzyme recognition sequence can be labeled by PFTase efficiently without affecting their

functional activity.

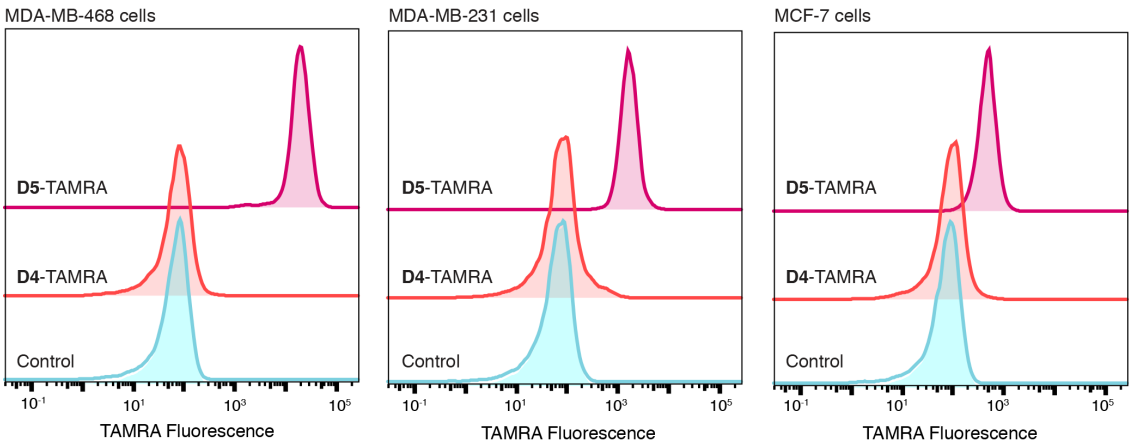


Figure 3.7 Evaluation of **D5**-TAMRA binding to cell-surface EGFR by flow cytometry. MDA-MB-468, MDA-MB-231 and MCF-7 cells displayed varied levels of EGFR expression (MDA-MB-468 ++++, MDA-MB-231 +++ and MCF-7 +). The TAMRA fluorescence detected from **D5**-TAMRA incubated cells correlated with their corresponding EGFR expression. No nonspecific binding was observed with **D4**-TAMRA with any of the cell lines.

3.3.6 Serum stability test of D1-TAMRA.

To create protein conjugates for potential clinical applications, the linkage between the protein and the cargo needs to be highly stable under physiological conditions. We therefore examined the stability of **D1**-TAMRA in the serum plasma *in vitro*. First, **D1**-TAMRA was diluted in human plasma (HP) and incubated at 37 ºC for different periods of time, including 2, 4, 8, and 24 h. At the desired time points, samples were removed from the water bath and flash-frozen. To analyze the integrity of the conjugate, SDS-PAGE and in-gel fluorescence imaging were employed, since the cleavage of the linker would result in the loss of the TAMRA fluorescence signal. As shown in Figure 3.8A, **D1**-TAMRA remained stable in human plasma for 24 h. Based on the results from coomassie blue stain, the presence of one band with molecular weight matching to that of DARPins suggested that the protein was not proteolyzed during incubation. In addition, quantification of the TAMRA fluorescence intensity revealed that 90% of the TAMRA fluorophore was retained in the conjugate after 24 h of incubation (Figure 3.8B). Encouraged by these results, we increased the incubation time to 48 h and found that the conjugate maintained its stability (90%) throughout this period (Figure S3.7). Therefore, the linker used to construct **D1**-TAMRA was stable in human serum plasma for at least 48 h.

Since mouse xenografts are often utilized for preclinical animal studies, incubated **D1**-TAMRA was also treated with mouse plasma at 37 ºC. Unfortunately, it was shown that the fluorescence from TAMRA diminished rapidly (Figure 3.8C). At 0.5 h, only 16% of the fluorescence signal was retained compared to the control D1-TAMRA conjugate. Two bands with sizes matched to that of **D1** were observed in the gel while only one of them was fluorescent. At the 0 h time point, degradation was already visible. Although the sample was flash-frozen immediately after mixing, it was left at 4 ºC for 1 h with Ni-NTA resin to enrich the conjugate for analysis. Therefore, the cleavage occurred even at low temperature. It is known that abundant carboxylesterases are present in mouse serum while minimal esterase activity is detected in human serum.[357-359] Thus, it is likely that the ester bond in the isoprenoid analogue **1** was hydrolyzed by the carboxylesterases in the mouse plasma, leading to the loss of the fluorophore. Accordingly, a more stable isoprenoid analogue needs to be designed for future studies in mouse models. However, the different outcomes in human and mouse sera noted here also highlight a pitfall of using mouse models as the surrogate for the actual clinical target, which should be carefully considered when designing conjugates for potential clinical applications.

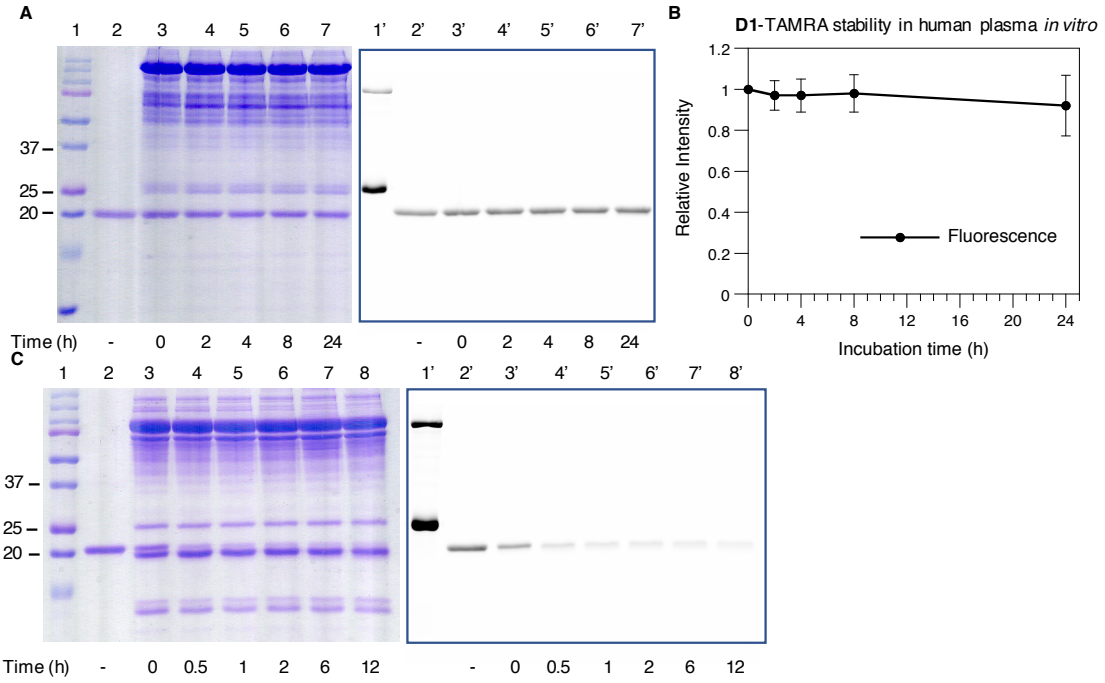


Figure 3.8 D1-TAMRA stability in serum plasma samples *in vitro*. Aliquots of **D1**-TAMRA diluted in 50% serum plasma were incubated at 37 ºC water bath for different duration and then frozen at -20 ºC. All experiments were performed in triplicate, with one representative gel shown above. (A) SDS-PAGE (left) and in-gel fluorescence imaging (right) characterization of **D1**-TAMRA incubated in human plasma. Lane 1: protein ladder; lane 2: **D1**-TAMRA without plasma

incubation; lane 3: 0 h incubation; lane 4: 2 h; lane 5: 4 h; lane 6: 8 h; lane 7: 24 h. (B) Quantification of **D1**-TAMRA fluorescence in each plasma sample. The band intensity was measured using imaging J and plotted in Kaleidagraph. (C) SDS-PAGE (left) and in-gel fluorescence imaging (right) characterization of **D1**-TAMRA incubated in mouse plasma. Lane 1: protein ladder; lane 2: **D1**-TAMRA without plasma incubation; lane 3: 0 h incubation; lane 4: 0.5 h; lane 5: 1 h; lane 6: 2 h; lane 7: 6 h, lane 8: 12 h.

3.3.7 Construction of DARPin-MMAE conjugates and cytotoxicity evaluation.

Since the labeled DARPin proteins maintained their binding specificity and capability, we further constructed DARPin based protein-toxin conjugates and tested their toxicity in cell cultures. Monomethyl auristatin E (MMAE), a well-studied potent cytotoxic compound, was chosen as the warhead. Since there are no commercially available aminooxy-modified MMAE derivatives, a DBCO-functionalized MMAE product (**4**) was utilized instead. Compound **4** contains the same linkage used in Adcetris, a commercial ADC developed by Seattle Genetics, where the Val-Cit dipeptide can be cleaved by lysosomal proteases to release the free MMAE drug.[360] After the aldehyde-modified DARPins **D1** and **D4** were immobilized onto hydrazide beads respectively, they were eluted using an aminooxy-PEG-azide linker (**3**) to form azide-functionalized DARPins. These azide-containing products were then reacted with **4** to afford the DARPin-MMAE conjugates using the strain-promoted azide-alkyne cycloaddition (SPAAC) reaction (Figure 3.9). Simply purification using small desalting columns did not completely remove excess **4**, therefore, SEC was employed. The purified conjugates were analyzed by SDS-PAGE and LCMS (Figure 3.10). Based on the deconvoluted mass of the products, the construction of the **D1**-oxi-SPAAC-MMAE (**D1**-MMAE) and **D4**-oxi-SPAAC-MMAE (**D4**-MMAE) was confirmed. Purity analysis based on the LC chromatogram revealed that **D1**-MMAE was 77% pure with **D1**-oxime-N_3 as the major impurity and small amounts of **D1**-BA, while for **D4**-MMAE, 93% purity was achieved with some unreacted **D4**-oxime-N_3 (Figure S3.8).

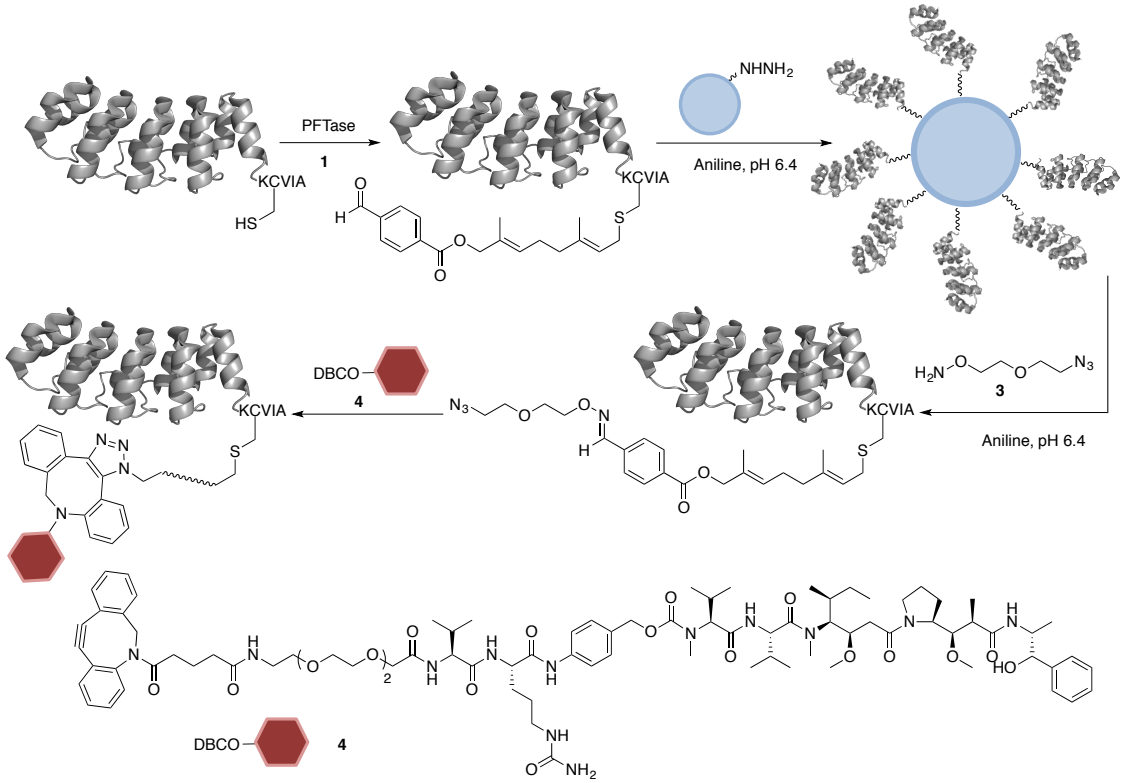


Figure 3.9 Construction of DARPin-MMAE conjugates using a modified capture and release strategy.

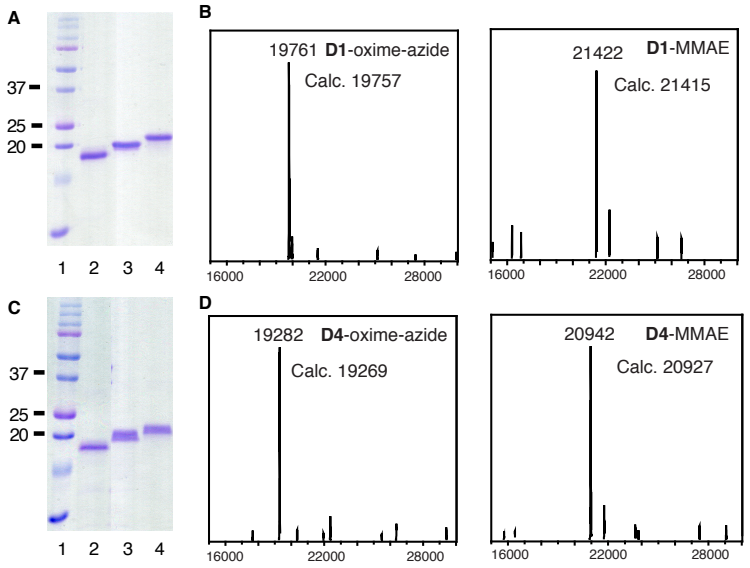


Figure 3.10 SDS-PAGE and LCMS characterization of **D1**-MMAE and **D4**-MMAE. (A) SDS-PAGE of D1-MMAE. Lane 1: protein ladder; lane 2: unmodified D1; lane 3: D1-oxime-azide; lane 4: D1-MMAE. (B) MS of D1-oxime-azide and D1-MMAE. (C) SDS-PAGE of D4-MMAE. Lane 1: protein ladder; lane 2: unmodified D4; lane 3: D4-oxime-azide; lane 4: D4-MMAE. (D) MS of D4-oxime-azide and D4-MMAE.

As a preliminary study, the crude conjugates were examined in cell cultures to test their cytotoxicity. Two cell lines were evaluated, including HT29 cells with EpCAM overexpression and U87-MG cells as a negative control. A serial dilution of **D1**-MMAE and **D4**-MMAE were supplied to the cells and the cytotoxicity was determined using an XTT assay after 96 h incubation. The results are represented in Figure 3.11 and all the IC_{50} values are summarized in Table 3.1. For HT29 cells, although similar potency was observed for **D1**-MMAE and compound **4**, a 2.5-fold enhancement in toxicity was measured when comparing **D1**-MMAE and **D4**-MMAE. Notably, both conjugates were much less potent than **4** in the control U87-MG cells, demonstrating the target specific toxicity of **D1**-MMAE. Furthermore, pre-incubation of HT29 cells with unmodified **D1** before the addition of **D1**-MMAE decreased the toxicity of **D1**-MMAE to the similar level of **D4**-MMAE, further validating that the elevated cytotoxicity of **D1**-MMAE in HT29 cells was resulted from specific **D1** uptake.

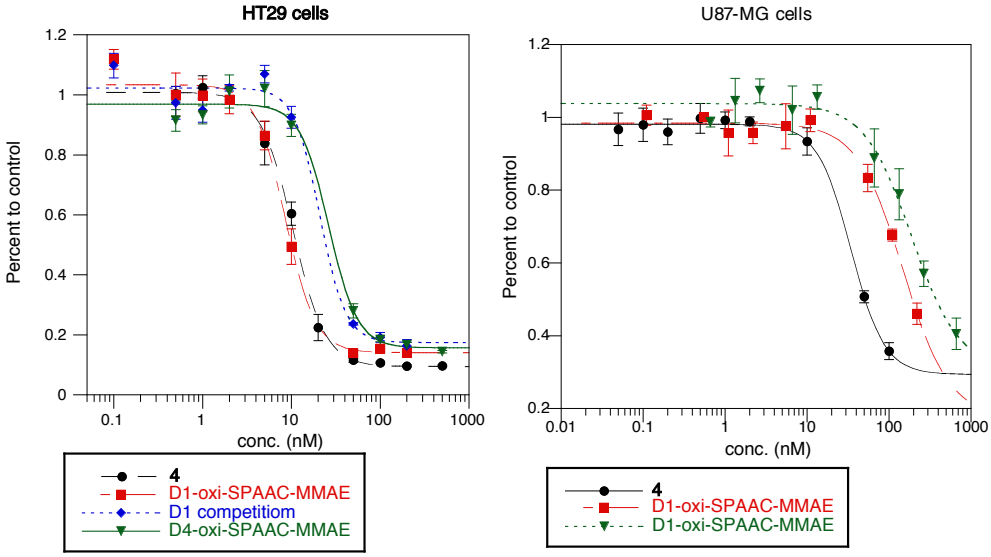


Figure 3.11 Cytotoxicity of **D1**-MMAE and **D4**-MMAE in HT29 (left) and U87-MG cells (right).

Table 3.1 Cytotoxicity of various **D1** and **D4** constructs against HT29 and U87-MG cells.

ENTRY	**IC_{50} (nM)**	
	HT29 cells	U87-MG cells
4	10.4 ± 0.5	34.4 ± 2.9
D1-MMAE	8.5 ± 0.7	149 ± 83.8
D4-MMAE	26.3 ± 5.0	188 ± 48.6
D1 competition	20.9 ± 5.7	n.d.

Although promising, the cytotoxicity profile of **D1**-MMAE was not sufficient to use as a targeted therapeutic. To create a more potent drug conjugate for potential therapeutic applications, we turned to another DARPin protein (Ec1) with higher affinity (370 pM) for cell-surface EpCAM.[353] The C-terminal CVIA sequence was engineered into Ec1 to yield Ec1-KCVIA (**D6**). Efficient labeling of **D6** with the aldehyde-bearing analogue **1** by PFTase was achieved and the modified **D6**-BA was then reacted with **2** to form the **D6**-TAMRA conjugate. Although precipitation formed after conjugation (Figure S3.9A), LCMS of the supernatant from the reaction verified the successful conjugation of **D6**-TAMRA (Figure S3.9B), albeit with a low yield of 16% due to the aggregation issue. The crude **D6**-TAMRA conjugates, containing unreacted **D6**-BA and PFTase were studied directly in flow cytometry experiments to evaluate their target binding capability (Figure S3.10). As expected, **D6**-TAMRA exhibited strong binding to EpCAM-positive HT29 cells with a 4-fold higher fluorescence intensity than that obtained for **D1**-TAMRA. However, a low degree of non-specific absorption of **D6**-TAMRA to U87-MG cells was also detected. Similarly, when HT29 cells were pre-treated with a large excess of unmodified **D6**, the fluorescence from **D6**-TAMRA could not be completely competed off, indicating that the remaining fluorescence signal was from receptor-independent association. Since **D1** and **D4-**TAMRA, constructed using the same method, did not exhibit nonspecific cell-surface binding (Figure 3.5), the observed off-target association of **D6**-TAMRA to both U87-MG and HT29 cells very likely results from the **D6** protein instead of the linkage and the fluorophore. In addition, the propensity of **D6**-TAMRA to precipitate may also contribute to the non-specific absorption onto the cell-surface.

Despite the low level of non-specific association of TAMRA-modified D6, it was desirable to prepare the **D6**-MMAE conjugate and evaluate its cytotoxicity. To avoid the multiple steps involved in the modified capture and release strategy, **D6** was labeled with an azide-containing analogue, C10-azide (**5**) and then conjugated to **4** using the SPAAC reaction to form **D6**-SPAAC-MMAE (Figure 3.12). Similar to the reaction with TAMRA, some of the D6-conjugates precipitated out. The supernatant from the reaction was purified by SEC to remove PFTase and excess reagent **4**. Surprisingly, a peak eluted before PFTase (90 kDa), which turned out to be the size of **D6** (20 kDa) when the fractions were analyzed by SDS-PAGE. This fraction was further analyzed by LCMS, and the desired **D6**-SPAAC-

MMAE conjugate was identified, whose mass closely matched the calculated values (Figure S3.11). A separate peak related to **D6** (identity not confirmed) as well as trace amount of PFTase were also detected in the LC chromatogram. The overall purity of **D6**-SPAAC-MMAE was estimated to be 50% based on gel densitometry. A control **D4**-SPAAC-MMAE conjugate carrying the same linkage was also prepared and verified by LCMS (Figure S3.11). The LC trace suggested the sample purity to be 78% with unreacted **D4**-azide and an unknown species also being present.

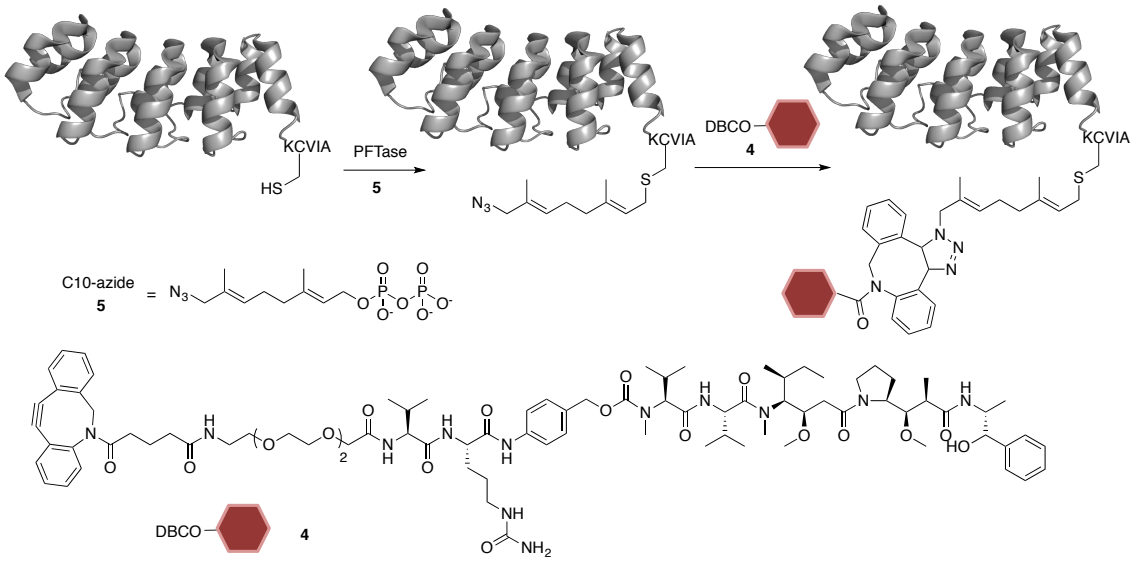


Figure 3.12 Construction of DARPin-MMAE conjugates using azide-containing isoprenoid analogue **5** and conjugation to DBCO-MMAE **4** using SPAAC reaction.

Cytotoxicity of both crude conjugates were evaluated in HT29 and U87-MG cells, using the same assay procedure employed for **D1**-MMAE. As can be seen in Figure 3.13 and Table 3.2, an 8-fold decrease in the IC_{50} in HT29 cells was achieved using **D6**-SPAAC-MMAE compared to that of **D1**-MMAE, however, competition using unmodified **D6** revealed off-target toxicity. **D6**-SPAAC-MMAE was also more toxic to the EpCAM-negative U87-MG cells than compound **4**. Pre-incubation of U87-MG cells with unmodified **D6** did not affect the potency of **D6**-SPAAC-MMAE, confirming that the toxic effect did not rise from **D6** dependent uptake.

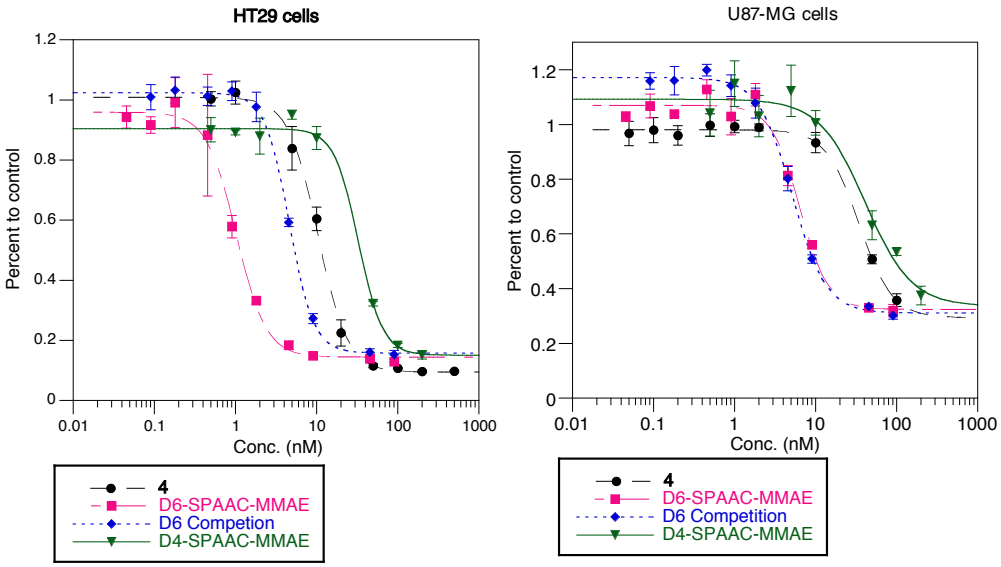


Figure 3.13 Cytotoxicity of **D6**-SPAAC-MMAE and **D4**-SPAAC-MMAE in HT29 (left) and U87-MG cells (right).

Table 3.2 Cytotoxicity of various D6 and D4 constructs against HT29 and U87-MG cells

ENTRY	**IC_{50} (nM)**	
	HT29 cells	U87-MG cells
4	10.4 ± 0.5	34.4 ± 2.9
D6-MMAE	1.0 ± 0.07	6.4 ± 0.7
D4-MMAE	32.9 ± 4.5	41.1 ± 13.9
D6 competition	4.6 ± 0.1	5.2 ± 0.2

To summarize, several DARPin proteins were labeled by PFTase and conjugated to MMAE successfully. Due to the lack of a commercially available aminooxy-functionalized MMAE compound, the aldehyde-modified DARPins were eluted with an azide-containing aminooxy-reagent. An extra SPAAC reaction step was used to provide the final conjugates. To further streamline the conjugation process and improve overall yield, efforts are undergoing to synthesize an aminooxy-coupled MMAE compound so that the toxin conjugates can be directly eluted from the beads. As to the cytotoxicity of the DARPin-MMAE conjugates, **D1**-MMAE showed specific toxicity towards EpCAM positive HT29 cells but was not potent enough. **D6**-SPAAC-MMAE was found to be highly toxic to HT29 cells due to increased binding affinity to the cell-surface EpCAM but suffered from off-target toxicity. These results underlined the fine balance between enhancing targeting binding affinity while suppressing non-specific association.

Additionally, screening and optimization will be needed to create a highly potent protein-drug conjugate suitable for therapeutic applications.

3.4 Conclusion

In conclusion, several DARPin proteins with engineered C-terminal CVIA sequence were successfully labeled by PFTase. The mere requirement of only four amino acid residues as the PFTase recognition sequence minimizes the potential perturbation to the structure and function of the DARPins. The usage of readily synthesized isoprenoid analogues enables incorporation of a variety of bio-orthogonal functional groups into DARPin proteins, which serve as a reactive handle for subsequent conjugation. As a highly efficient enzymatic reaction, excellent labeling yield can be achieved using PFTase.

To facilitate the conjugation and purification process, a capture and release strategy based on the aldehyde functionality was employed. Using this technique, the unmodified proteins and other impurities can be easily removed by simple washing steps, eliminating the need for complicated chromatographic methods. More importantly, this strategy can be expanded to create protein conjugates directly from cell lysates, which further streamlines the production process and can be easily scaled up. Two applications were demonstrated in this study using the capture and release strategy to construct DARPin-based conjugates. First, DARPin-fluorophore conjugates were created, which were shown to be capable of binding to their designated cell-surface targets. Notably, the modification did not cause any non-specific association. The stability of the conjugates was evaluated in serum plasma samples, which were selected to mimic physiological conditions. It was found that the DARPin-fluorophore conjugates were highly stable in human plasma but were hydrolyzed rapidly in mouse serum. Considering the presence of an ester bond in the isoprenoid analogue used to modify the DARPin proteins, the difference in stability was attributed to the different levels of carboxylesterases in these two plasma samples. Therefore, a more stable chemical entity needs to be designed to construct DARPin-conjugates suitable for future pharmacokinetics and efficacy studies in mouse models. As a second example, DARPin-drug conjugates were also prepared and evaluated in cell cultures. An extra reaction step was taken due to the lack of commercially available reagent, however, efforts

to optimize the conjugation procedure are ongoing. The cytotoxicity results suggest that potency and specificity of the conjugates are both crucial factors to be considered when creating protein-drug conjugates for potential therapeutic applications.

Overall, the combination of the enzymatic PFTase labeling method and the capture and release strategy can be broadly applied as a facile technique with improved potential to create various functional protein-based conjugates for a variety of applications.

3.5 Supplemental Data

3.5.1 MS of unmodified and FPP labeled DARPin protein

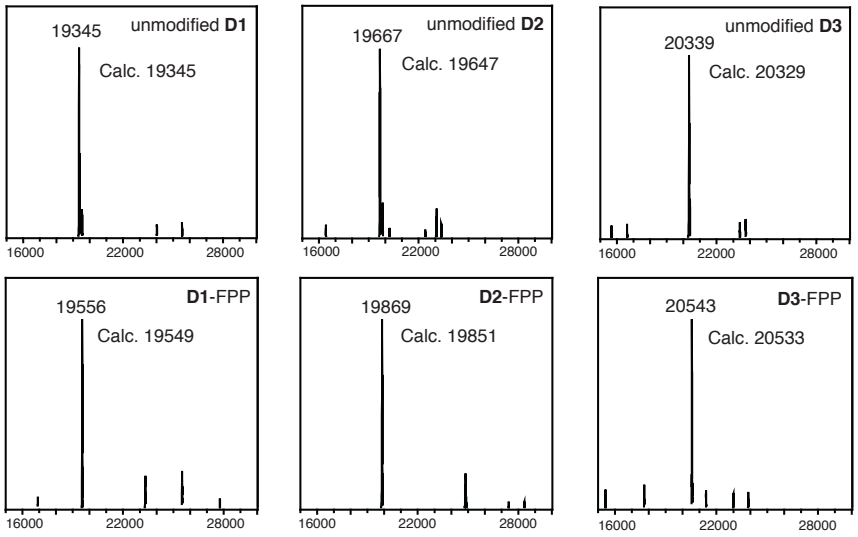


Figure S3.1 MS of unmodified and FPP labeled DARPin proteins.

3.5.2 TIC trace of the FPP prenylation reaction of D1, D2 and D3 from LCMS characterization

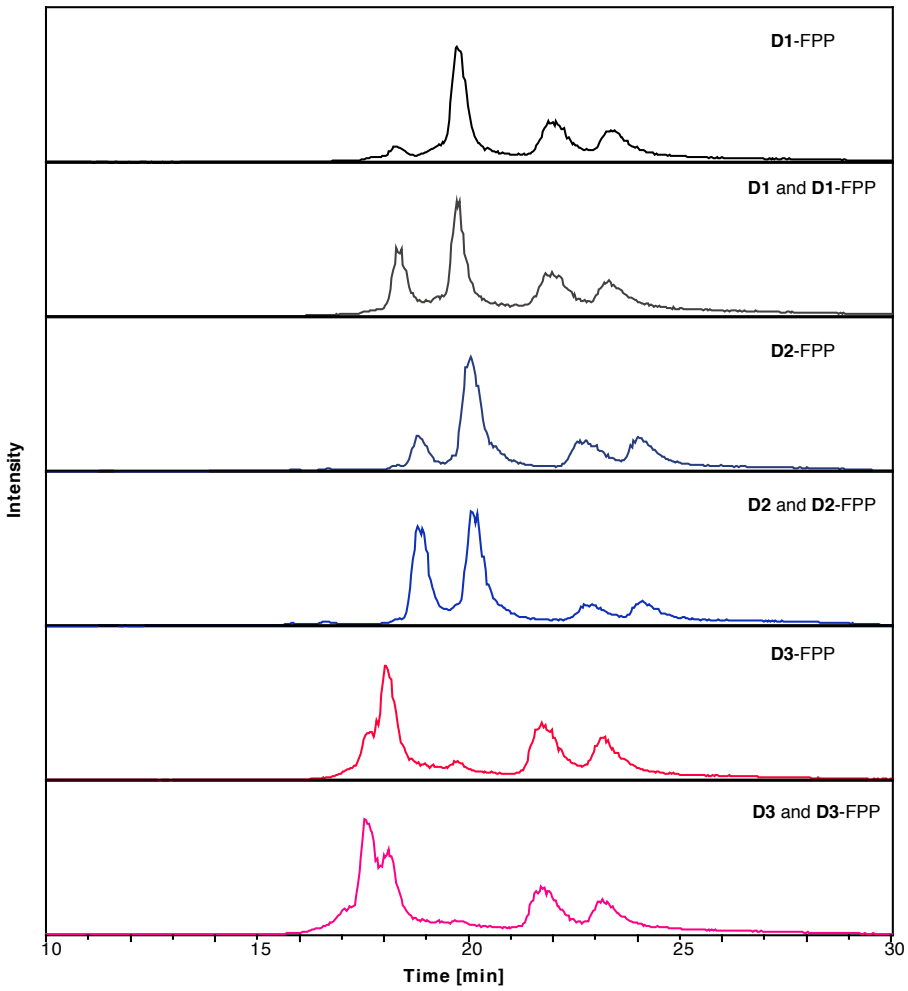


Figure S3.2 TIC trace of the FPP prenylation reaction of **D1**, **D2** and **D3**. Unmodified protein was also spiked in the reaction mixture to help identify the peak for the unmodified protein in the reaction mixture.

3.5.2 UV traces from the LCMS characterization of the FPP prenylation mixture of D1, D2 and D3

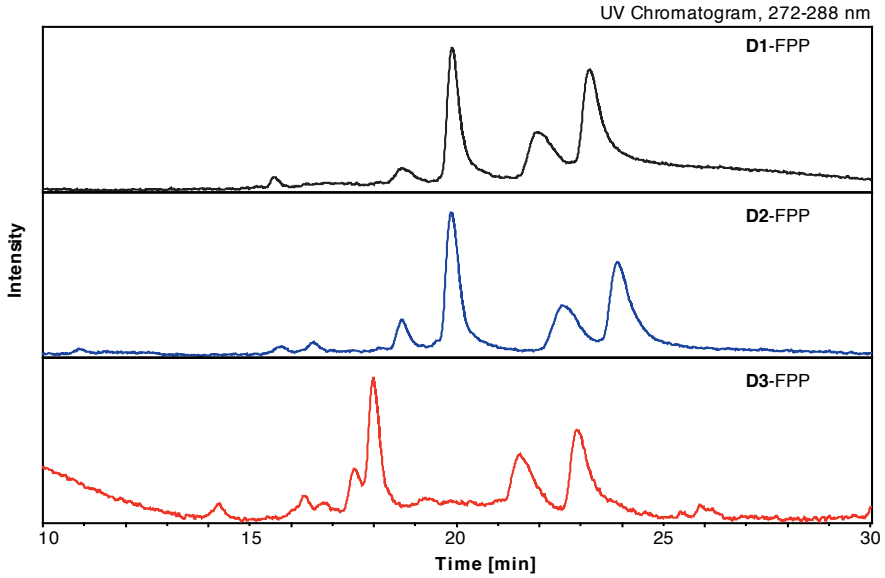


Figure S3.3 UV traces from the LCMS characterization of the FPP prenylation mixture of **D1**, **D2** and **D3**. Absorption at 280 nm was measurement.

3.5.3 LCMS characterization of D4-FPP prenylation reaction

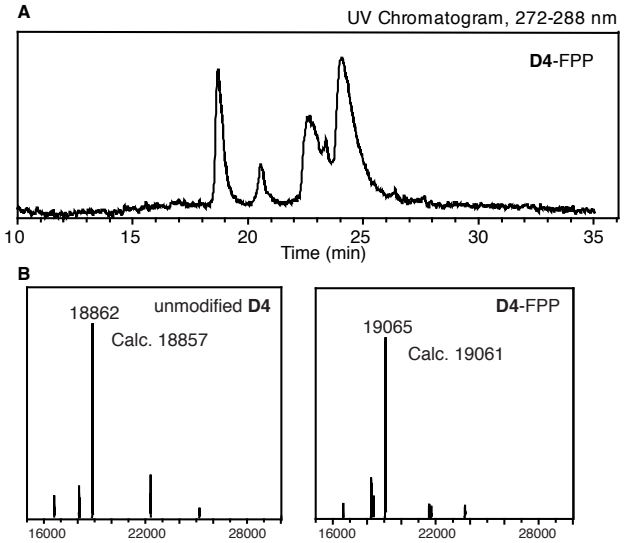


Figure **S3.4** LCMS characterization of **D4**-FPP prenylation reaction. (A) UV trace of the reaction mixture measuring absorption at 280 nm. (B) MS of unmodified **D4** and **D4**-FPP

3.5.4 SDS-PAGE characterization of D4 released with hydroxylamine

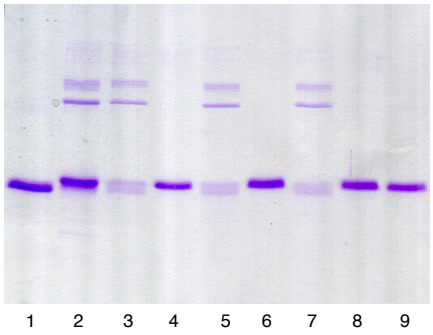


Figure S3.5 SDS-PAGE characterization of **D4** released with hydroxylamine. Lane 1: unmodified **D4**; Lane 2: **D4**-BA prenylation mixture; Lane 3: supernatant from immobilization with 10-fold excess amount of beads; lane 4: **D4** released from beads with hydroxylamine; lane 5: supernatant from immobilization with 20-fold excess amount of beads; lane 6: **D4** released from beads; lane 7: supernatant from immobilization with 50-fold excess amount of beads; lane 8: **D4** released from beads; lane 9: **D4** released from 100-fold excess amount of beads.

3.5.5 SDS-PAGE and in-gel imaging characterization of D5-TAMRA

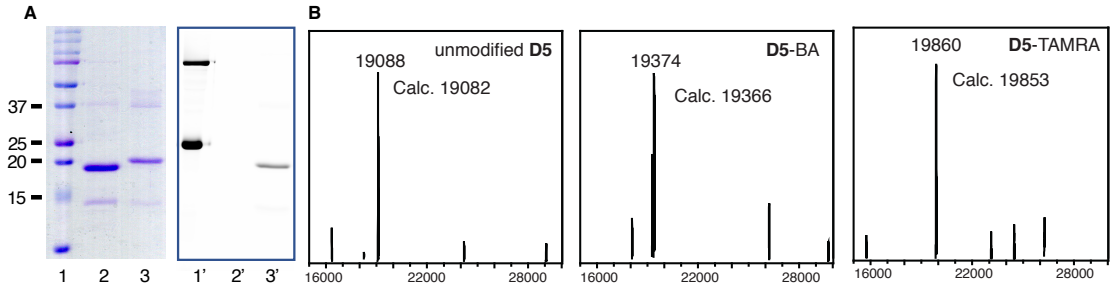


Figure S3.6 (A) SDS-PAGE and in-gel imaging characterization of **D5**-TAMRA. Lane 1: protein ladder; lane 2: unmodified **D5**; lane 3: **D5**-TAMRA. (B) MS of unmodified **D5**, **D5**-BA and **D5**-TAMRA.

3.5.6 Characterization of D1-TAMRA incubated in human plasma *in vitro* for 48 h

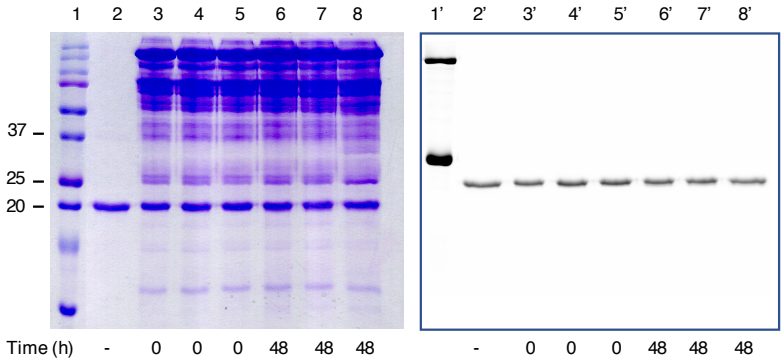


Figure S3.7 SDS-PAGE (left) and in-gel fluorescence imaging (right) characterization of **D1**-TAMRA incubated in human plasma *in vitro* for 48 h. Lane 1: protein ladder; lane 2: **D1**-TAMRA without plasma incubation; lane 3 to 5: **D1**-TAMRA incubated in plasma for 0 h; lane 6 to 8: **D1**-TAMRA incubated in plasma for 48 h.

3.5.7 UV chromatogram from LC-MS characterization of D1-MMAE and D4-MMAE

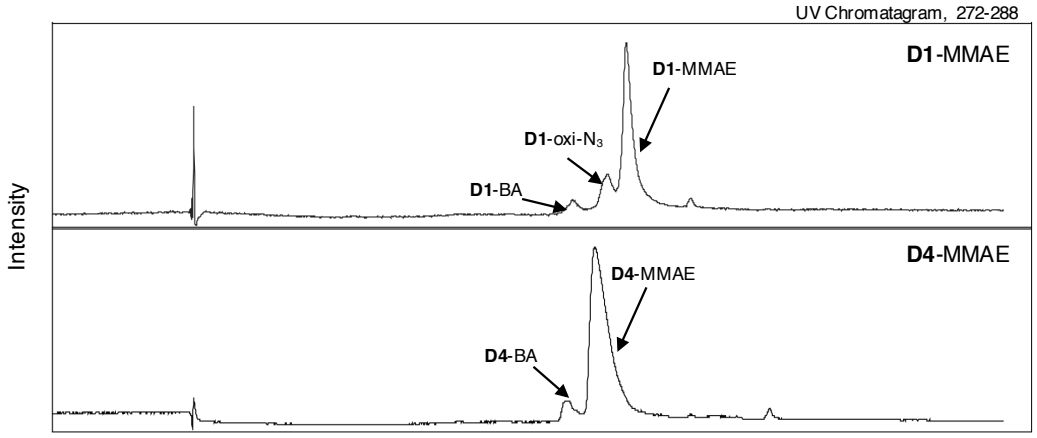


Figure S3.8 UV chromatogram from LC-MS to quantify the purity of **D1**-MMAE (77%) and **D4**-MMAE (93%).

3.5.8 Characterization of D6-TAMRA

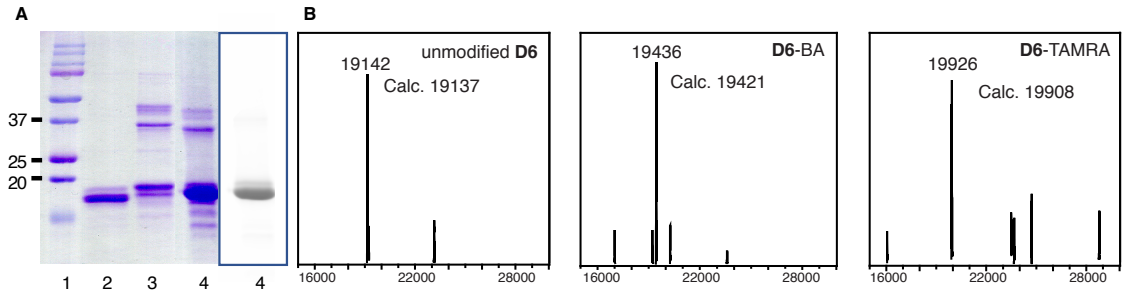


Figure S3.9 A. SDS-PAGE (left) and in-gel fluorescence scanning (right) of D6-TAMRA conjugates. B. MS of various D6 constructs.

3.5.9 Characterization of D6-TAMRA binding to cell-surface EpCAM

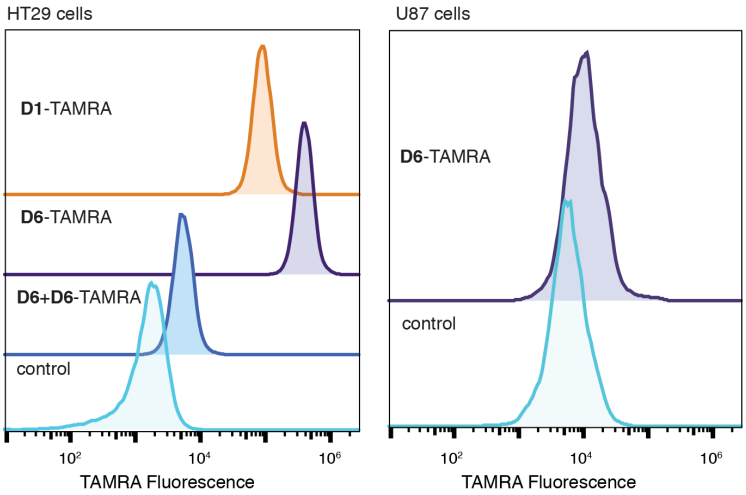


Figure S3.10 Flow cytometry analysis of D6-TAMRA binding to HT29 and U87-MG cells. Strong binding to EpCAM-positive HT29 cells were confirmed, however, a low degree of non-specific absorption to EpCAM-negative U87-MG cells was also detected.

3.5.10 MS of D6- and D4-SPAAC-MMAE

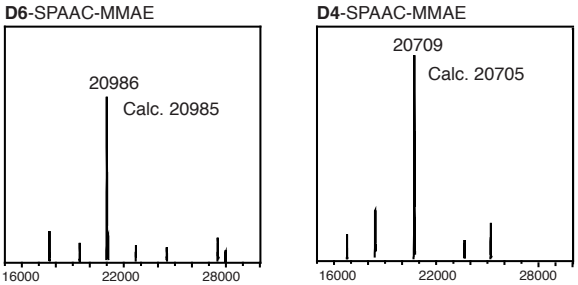


Figure S3.11 Deconvoluted MS of D6-SPAAC-MMAE and D4-SPAAC-MMAE.

Chapter 4. Exploring Enzymatic Labeling of Alternative Binding Proteins Using Protein Farnesyltransferase

Compared to the conventional cancer treatments and diagnosis, targeted therapies and related imaging methods offer tremendous advantages in that they can specifically eliminate or detect tumorous cells without affecting normal tissues. In addition to the well-studied antibodies, small protein scaffolds with similar antigen-binding properties have also been widely explored, including the 10^{th} extracellular domain of human fibronectin III (Fn3) and the single-domain antibody (VHH) proteins. These protein scaffolds can be easily produced from bacterial cells with high homogeneity, while their small size allows them to penetrate deeper into tissues to reach solid tumor sites. These favorable characteristics make them attractive candidates to serve as alternative targeting moieties. To attach drug molecules or imaging agents to the targeting proteins, a variety of methods have been employed. Although direct cysteine modification is an efficient and facile method, it is not broadly applicable due to the lack of selectivity in cases where multiple cysteines are present. Enzymatic protein labeling is a powerful tool for site-specific modification. However, each enzyme manifests its own advantages and limitations and hence selection of an optimal method depends on the protein substrates and the particular application. In this study, we explored the applicability of using protein farnesyltransferase (PFTase) to label Fn3 and VHH proteins. As a highly efficient enzyme with fast kinetics, it requires only four amino acid residues as a recognition sequence. We were able to label several Fn3 and VHH proteins engineered with C-terminal CVIA sequences with a panel of isoprenoid analogues bearing various bioorthogonal functional groups successfully using PFTase, and to construct protein-fluorophore conjugates using CuAAC and oxime ligation reactions. Unfortunately, it was found that all the Fn3 and VHH variants being investigated displayed reduced solubility after modification with the relatively hydrophobic isoprenoid-based substrates. In addition, we also discovered that the oxime ligation reactions produced soluble protein conjugates while all the conjugates precipitated out in the CuAAC reactions. These findings suggest that protein solubility is an important factor to be taken into account when preparing protein-based conjugates and oxime ligation

reactions can be first considered over CuAAC reactions to avoid potential aggregation problems.

4.1 Introduction

Targeted delivery, which can specifically direct cytotoxic molecules or imaging agents to malignant cells, provides tremendous advantages over the conventional non-selective methods in cancer treatment and diagnosis.[11] To achieve selectivity, binding proteins, which can recognize specific molecules that are overexpressed on the surface of tumor cells, have been extensively explored. In addition to conventional antibodies, alternative binding scaffolds with comparable antigen-binding properties have also been widely studied.[361,362] Compared to the large size of an antibody, these proteins are usually small in size, which allows them to penetrate deeper into the tissues.[118] Additionally, without the complex glycosylation modification found on antibodies, alternative binding proteins can be easily produced from bacterial culture in high homogeneity. As one of the promising candidates for targeted delivery, ^{10}Fn3 (Fn3), derived from the 10th extracellular domain of human fibronectin III, contains three solvent-exposed loops that are structurally comparable to the complementarity-determining regions (CDRs) on the heavy chain variable domain (V_H) of IgG.[363] Thus, these three loops can serve as the recognition site for target binding. Another protein scaffold that also utilizes loop elements for binding interaction is the nanobody, which is also called a single-domain antibody (VHH).[364] Discovered from the natural immune systems of camels and alpacas, it is the smallest antibody fragment with target-binding ability. Compared to its counterpart (V_H domain) in IgG, the hydrophobic residues in V_H that are involved in the interaction with V_L are replaced by smaller and/or hydrophilic amino acids in VHH. As a result, VHHs are highly stable in monomeric form. To generate binders towards specific targets, *in vitro* selection is performed from protein libraries containing randomized residues in the loops of Fn3 or the CDRs in VHH and displayed with appropriate methods, such as yeast-surface display or phage display.[365-367] It has been shown that selective binders can be obtained with high affinities comparable to those achieved by the antibodies.[368-370]

To attach cytotoxic drugs or imaging probes to the binding proteins for targeted delivery, one of the commonly used methods is through cysteine conjugation.[29] When there is no endogenous cysteines in the binding protein, genetic introduction of this residue enables site-specific conjugation. However, if there are already cysteine residues present (usually involved in disulfide-bonds), direct chemical modification is not selective and can result in heterogeneously modified protein conjugates that are hard to characterize and may possess different activities. Moreover, VHH proteins contains a disulfide bridge that is structurally important. Interruption of this bond caused by cysteine modification may render the protein unstable. In the case of Fn3, although the wild-type Fn3 does not contain endogenous cysteine, this residue might be evolved from library selection. For example, Wittrup and coworkers generated an Fn3 with 350 pM affinity towards lysozyme, which contained a pair of cysteine residues on adjacent loops that was structurally analogous to that in the VHH.[371] This study highlighted the potential of introducing an inter-loop disulfide bond to generate high-affinity Fn3 binders. Therefore, cysteine modification is not broadly applicable.

Different from the direct chemical modification method, enzymatic protein labeling methods are highly specific. To date, a variety of enzymes have been exploited for protein modification purposes.[18] Each of them possesses its advantages and limitations. For example, sortase A is a versatile enzyme that can accept a wide range of substrates but is also limited by the slow kinetics and moderate labeling yields of the wild-type enzyme.[18] Labeling using microbial transglutaminase will potentially produce crosslinked by-products due to the promiscuous substrate specificity of the enzyme.[222] As a result, both the protein substrates and the specific applications need to be considered to select an optimal method for protein modification.

As an alternative enzymatic labeling technique, our lab has previously utilized protein farnesyltransferase (PFTase), a highly efficient enzyme with fast kinetics, to modify different target proteins, which were then conjugated to various cargos, such as fluorophores,[121] oligonucleotides,[122] PEG polymers[65] and solid surfaces.[246] The enzyme recognizes a C-terminal CaaX motif, where C is the cysteine that is modified and transfers the isoprenoid moiety from farnesyl diphosphate (FPP) to the cysteine via formation of a thiol-ether linkage (Figure 4.1A). It has been previously shown that incorporation of a

CaaX-box, such as CVIA, onto the C-terminus of the target proteins renders them substrates for PFTase. Importantly, the requirement of only four amino acids as the recognition sequence minimizes the potential perturbation to the structure and function of the target proteins. In addition, a variety of isoprenoid analogues bearing bioorthogonal functional groups, including azides,[245] alkynes,[117,324] and aldehydes,[65,250] have been developed (Figure 4.1B). Once installed onto the proteins of interest by PFTase, these functional groups serve as specific reactive handles for subsequent conjugation to additional cargos, including the cytotoxic drugs or imaging agents. In this study, we explored the possibility of using PFTase to label Fn3 and VHH proteins for the construction of targeted delivery systems. Since the C-termini of both Fn3 and VHH is distant from the target-binding site, modification at this position is advantageous since it reduces the potential for interference with their binding ability.

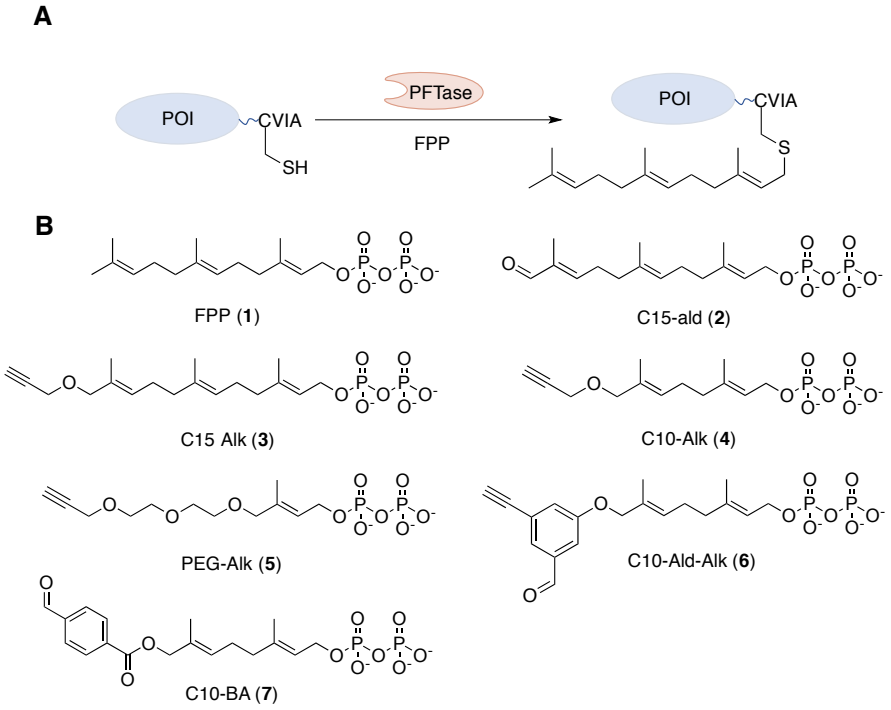


Figure 4.1 (A) Schematic representation of enzymatic labeling by PFTase with a protein of interest (POI) containing a C-terminal CVIA sequence and FPP. (B) Structures of FPP and selected isoprenoid analogues bearing different bioorthogonal functional groups.

4.2 Materials and Methods

4.2.1 Prenylation reactions with EI3.4.3 variants.

F1, **F2** and **F3** proteins were obtained from Austin Battles in Dr. Benjamin Hackel's lab. Prenylation cocktail was prepared with 50 mM Tris·buffer (pH 7.5), 10 mM $MgCl_2$, 50 μM $ZnCl_2$, 15 mM DTT (Gold Biotechnology), 20 mM KCl and 2 μM protein substrate. The reagents were incubated on ice for 50 min. Afterwards, 15 μM isoprenoid analogues and 100 nM yPFTase were added to initiate the prenylation reaction with a final volume of 250 μL. The reaction was placed in a 33 °C water bath for 4 h. For LCMS analysis, the reaction mixture was injected into an Agilent MSD SL Ion Trap instrument directly without further purification. Large scale prenylation reactions were carried out with the same concentration for each reagent but a larger final reaction volume (5 mL).

4.2.2 F1 expression and purification.

The plasmid encoding **F1** was provided by Dr. Larry Stern from Dr. Benjamin Hackel's lab. 1 μL of the plasmid was used to be transformed into 50 μL of BL21(DE3) cells following the instruction manual from the QuickChange II Site-Directed Mutagenesis Kit (Agilent, Catalog #200523). The transformed cells were plated on LB-Agar plates with 50 μg/mL Kanamycin (Kan). After overnight incubation at 37 °C, a single colony was used to inoculate 50 mL of LB media containing 50 μg/mL Kan. The culture was grown overnight at 37 °C, shaking at 250 rpm. 10 mL of solution from the overnight culture was then added to 1 L of LB media containing 50 μg/mL Kan, which was then incubated at 37 °C with shaking at 250 rpm. When the OD_{600} of the media reached to 0.8, IPTG was added to a final concentration of 0.5 mM to induce protein expression at 30 °C for 2 h, shaking at 250 rpm. Cells were harvested by centrifugation at 5,000 g for 30 min and the pellet was then stored at -80 °C if not used immediately. The cell pellet from 1 L of LB media was suspended in 10 mL lysis buffer (50 mM PB pH 7.2, 500 mM NaCl, 0.05 % (v/v) glycerol, 5 mM CHAPS and 25 mM imidazole). The cells were lysed by the freeze-thaw method for four cycles. Clear supernatant was collected by centrifugation at 12,000 g for 10 min and filtered through a 0.45 μm filter. The filtrate was loaded onto a column with 4 mL of Ni-NTA resin (Qiagen), which was pre-equilibrated with 60 mL of equilibration buffer (PBS and 10 mM imidazole). The flow through was collected and loaded onto the column again. The column was then washed with 60 mL of wash buffer

(PBS and 20 mM imidazole). Proteins were then eluted with 8 mL of elution buffer (PBS and 250 mM imidazole). Fractions of 1 mL were collected. Fractions 2 through 6 were concentrated to 500 μL. Buffer exchange was performed using an Amicon Ultra-0.5 centrifugal filter (3k cutoff, Millipore) by adding PBS to the concentrated solution and then concentrating again. The buffer exchange process was repeated two more times. The final solution was flash frozen with liquid nitrogen and stored at -80 °C as a stock. The concentration of the stock was determined by Bradford Assay using a DTX 880 Multimode Detector (Beckman Coulter).

4.2.3 HPLC purification of F1.

To prepare the protein samples for HPLC analysis, 50 μL of TFA was added to 1 mL of the **F1** collected from Ni-NTA column to acidify the solutions which was then filtered through a 0.45 μm filter. A Phenomenex Jupiter C18 column (300 A, 3 microns, 4.6 mm x 150 mm) was used for purification and the buffer system was: buffer A (0.1 % TFA in H_2O) and buffer B (89.9% ACN, 10% H_2O and 0.1% TFA). The method transitioned from 0 to 15% B in 3 min, 15% buffer B to 80% buffer B in 15 min and then to 100% B in 3 min. The flow rate was set at 1 mL/min. Absorbance at 280 and 220 nm was monitored. **F1** eluted around 16 min using this gradient.

4.2.4 F4 and F5 expression and purification.

Plasmids encoding **F4** and **F5** were provided by Dr. Larry Stern from Dr. Benjamin Hackel's lab. The same procedure used to express **F1** (described above) was employed to express and purify **F4** and **F5** from BL21(DE3) *E. coli* cells using Ni-NTA resin. For HPLC purification of **F4** and **F5**, the same method was used except that buffer B was changed to 0.1% TFA in ACN and the gradient was designated to transition from 15% B to 100% B in 22 min. **F4** eluted around 19 min and **F5** eluted around 16 min.

4.2.5 Enzymatic labeling of F4 and F5.

To test whether the various isoprenoid analogues can be incorporated into **F4** and **F5**, small scale reactions (500 μL) were carried out. The same prenylation cocktail was utilized as described for **F1** labeling. For the prenylation reactions of **F4** and **F5** with

analogue **7**, a reaction mixture (14 mL scale) prepared with 50 mM Tris·buffer (pH 7.5), 10 mM $MgCl_2$, 50 μM $ZnCl_2$, 15 mM DTT, 20 mM KCl, 0.04 % w/w DDM and 1 μM protein substrate was used. The mixture was incubated on ice for 0.5 h, followed by the addition of 10 μM **7** and 200 nM yPFTase. The reaction was placed in a 32 °C water bath overnight. Any precipitation from the prenylation reaction was removed by further centrifugation at 13,000 g for 10 min. The supernatant was injected into the LCMS instrument for characterization directly without further purification. To remove excess isoprenoid analogues, the supernatant was concentrated by Amicon Ultra-15 centrifugal filter (3k cutoff, Millipore) to 500 μL and washed with PBS once. The final solution was concentrated to 350 μL.

4.2.6 Construction of Fn3-fluorophore conjugates using CuAAC reactions.

The alkyne modified Fn3 proteins were reacted with a 5-fold excess amount of **8** or **9** in the presence of 0.1 mM TBTA, 1 mM TCEP and 1 mM $CuSO_4$. The reaction was conducted at RT for 2 h. After centrifugation at 10,000 g for 10 min, both the pellet and the supernatant were characterized by SDS-PAGE.

4.2.7 Conjugation of Fn3-2 to TAMRA using oxime ligation.

Oxime ligation reactions (25 μL) contained **F4-2** or **F5-2**, 50 μM **10** (1.8 mM stock in DMSO) and 10 mM mPDA (100 mM stock in H_2O) or 10 mM pPDA (1 M stock in DMSO) as a catalyst. Reactions were performed at RT for 5 h. The reactions were centrifuged at 10,000 g for 10 min to remove any precipitate. Both the precipitate and supernatant of the reaction were analyzed by SDS-PAGE.

4.2.8 Conjugation of Fn3-7 to TAMRA using oxime ligation.

For oxime ligation with the Fn3 proteins modified with analogue **7**, a 5-fold excess of **10** was used with no catalyst. The reaction was carried out at RT overnight. The crude product was purified with a NAP-5 column using 100 mM PB (pH 7.0) as the eluent. The collected colored fraction was characterized by SDS-PAGE, in-gel fluorescence analysis and LCMS and stored in -80 °C freezer. The purity of the Fn3-TAMRA conjugates was determined using the absorption in the LC chromatogram at 555 nm from LCMS

characterization. The concentration of conjugates was calculated using their absorbance at 555 nm (ε = 65,000 M^{-1} cm^{-1}).

4.2.9 Binding analysis of F4-TAMRA to cell-surface EGFR using flow cytometry.

MDA-MB-468 and MCF-7 cells were cultured in DMEM media with 110 mg/mL sodium pyruvate and 10% FBS. For flow cytometry analysis, 80,000 cells were counted and aliquoted into a micro-centrifuge tube. These tubes were kept on ice at all times. After centrifuging the cells at 300 g for 3 min and discarding the media, the cell pellet was gently resuspended in 1 mL of ice-cold PBSA (PBS with 1 g/L BSA). Centrifuge the solution at 300 g for 3 min and discard the PBSA. **F4**-TAMRA or **F5**-TAMRA was diluted in PBSA to a final concentration of 100 nM and 200 μL of the solution was added to the cell pellet. The mixture was incubated in dark for 15 min at 4 °C. The labeling solution was then removed and the cells were washed with PBSA once. The washed cell pellet was resuspended in 200 μL of PBSA for flow cytometry analysis. A total of 6 samples were prepared including: MDA-MB-468 cells treated with only PBSA, **F4**-TAMRA, or **F5**-TAMRA and MCF-7 cells treated with only PBSA, **F4**-TAMRA, or **F5**-TAMRA. Flow cytometry analysis was performed using a BD LSR II/Fortessa H0081 instrument with a 561 nm laser (fluorophore 561-C).

4.3 Results and Discussion

4.3.1. Preliminary studies of Fn3 labeling by PFTase.

To study whether the Fn3 proteins can be labeled by PFTase, EI3.4.3, an Fn3 construct that binds to epidermal growth factor receptor (EGFR) with extremely high affinity (K_d 250 pM) was chosen.[365] Importantly, a disulfide bond is present in this construct which is evolved from library selection and is crucial for target binding. To make EI3.4.3 a substrate for PFTase, the enzyme recognition sequence CVIA was inserted at the C-terminus of the protein. To evaluate if an additional spacer sequence was needed for efficient labeling, three constructs were created with different spacer residues, which included EI3.4.3-GCVIA (**F1**), EI3.4.3-KCVIA (**F2**) and EI3.4.3-SCVIA (**F3**). As a preliminary study, alkyne-containing isoprenoid analogue **3** and aldehyde-bearing

analogue **2** were selected for protein modification since these bioorthogonal functional groups can be used for subsequent conjugation using different chemical reactions. Small-scale reactions were performed with each construct using **2** or **3** by PFTase respectively (Figure 4.2).

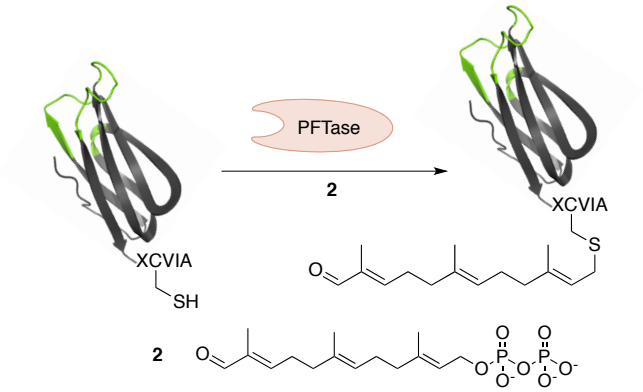


Figure 4.2 Fn3 labeling by PFTase. Analogue **2** is used as an example.

To characterize the reaction product, the crude reaction mixture for each reaction was subjected to LCMS analysis (Figure S4.1). To estimate the yields of the enzymatic labeling, we compared the relative intensity of the modified protein versus the unmodified starting material from MS. Since the size of Fn3 is much larger than that of the added isoprenoid moiety, it is unlikely that the ionization ability of the modified product would be substantially different from the unmodified protein. The LCMS results are summarized in Table 4.1. When analogue **3** was employed, all the three EI3.4.3 variants were modified successfully by PFTase as the detected mass closely matched to those of the calculated values. The yield for **F1** labeling was similar to that of **F2**, while modification of **F3** was much less efficient. In terms of analogue **2**, only **F1** was modified with a moderate yield. Since **F1** can be labeled by both analogues **2** and **3**, it was chosen for further study.

Table 4.1 LC-MS results from prenylation reactions with EI3.4.3 variants using PFTase

Substrate	Calculated Mass	Measured Mass	Efficiency
F1	12406	12404	-
F2	12477	12475	-
F3	12436	12435	-
F1-2	12624	12626	46%

F1-3	12664	12666	91%
F2-3	12735	12737	86%
F3-3	12694	12696	54%

4.3.2. F1 labeling by PFTase.

To prepare sufficient modified proteins for subsequent modification, a large-scale prenylation reaction was performed with **F1** and **3**. Unexpectedly, precipitate was observed after 4 h incubation in a 33°C water bath. Similar results were obtained when switching from **3** to **2** or **4**. A control experiment that combined **F1** and the PFTase in the absence of the isoprenoid analogues revealed minimal pellet formation after the same incubation time. Since the isoprenoid moiety is relatively hydrophobic, it is likely that the modified **F1** aggregated. Compared to the results from small-scale reactions, it is possible that aggregation also formed in previous experiments but was not visible to be observed due to the small scale used for those reactions.

Attempts to prevent precipitation by adding BSA (1 mg/mL) to the prenylation mixture did not appear to be useful. Instead, BSA precipitated out of the solution with the **F1** protein (Data not shown). It was realized later that after storing the unmodified **F1** protein in the cold room for a week, it precipitated out from the stock solution. All these results suggested that the unmodified **F1** protein had a relatively low solubility, which became even lower with the attachment of the hydrophobic isoprenoid analogues. In addition, the protein purification was problematic. The **F1** protein was purified from bacterial cell lysate using Ni-NTA resin. However, the purified product displayed multiple bands on the SDS-PAGE gel (Figure 4.3A). When HPLC was employed to clean up the sample using a reverse-phase column, the lyophilized protein pellet could not be fully re-dissolved. Based on all these results, **F1** was not pursued for additional experiments.

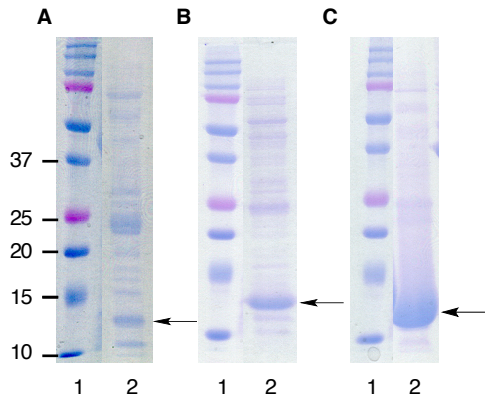


Figure 4.3 SDS-PAGE of **F1** (A), **F4** (B), and **F5** (C) purified from bacterial lysate using Ni-NTA resins. Lane 1: protein ladder; lane 2: purified protein. The black arrow indicates the desired Fn3 proteins.

4.3.3 Design and purification of alternative Fn3 constructs.

Although **F1** exhibited solubility problems when labeled by PFTase, other Fn3 proteins might not have similar issues. Therefore, two alternative Fn3 constructs, E6.2.6 and RDG engineered with C-terminal KCVIA sequences were evaluated. E6.2.6-KCVIA (**F4**) binds to EGFR with sub-nanomolar affinity (260 pM)[365] while RDG-KCVIA (**F5**) does not bind to any target and serves as a negative control. Both proteins were purified from *E. coli* lysate using cobalt affinity resin. The yield was relatively low for **F4** with only 0.2 mg of materials isolated from 1 L of cell culture. When the eluted proteins were characterized by SDS-PAGE, numerous impurities were observed (Figure 4.3). HPLC was initially used to further purify the proteins, but the recovered yield from the lyophilized protein pellet was low. For preliminary studies, it was decided that protein samples collected from the affinity resin would be used directly for modification since only the CVIA-appended **F4** and **F5** could be recognized and modified by PFTase. The mass of each purified protein was confirmed by LCMS (Figure 4.4).

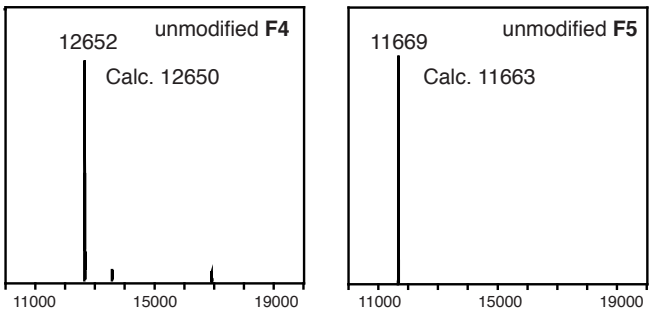


Figure 4.4 MS of unmodified **F4** and **F5**.

4.3.4 Enzymatic labeling of F4 and F5 to construct protein-fluorophore conjugates.

Prenylation reactions were carried out with **F4** and **F5** using a subset of isoprenoid analogues (**2**, **4**, **5** and **6**). These analogues can be categorized into three groups, (1) the alkyne-containing analogues **4** and **5**; (2) the aldehyde-bearing **2** and (3) analogue **6** with both alkyne and aldehyde functionalities. In group 1, Analogue **4** is less hydrophobic than the previously used **3**. Although analogue **5** is quite hydrophilic, it is not well-recognized by PFTase and large amounts of both the analogue and the enzyme are needed to achieve acceptable labeling yields.

Similar to the results obtained with the **F1** protein, precipitate was seen in each reaction. Based on the size of protein pellet, the labeled **F4** proteins were less soluble than the modified **F5**. To try to reduce precipitation, n-dodecyl β-D-maltoside (DDM), a detergent previously applied to improve the solubility of peptide substrates used in PFTase kinetic assays,[245] was added to the prenylaiton reaction mixture. However, this detergent did not appear to be effective in preventing Fn3 proteins from precipitation. LCMS characterization of the supernatant revealed that there was still labeled proteins remained in the solution (Figure S4.2). Therefore, despite the formation of precipitate, soluble modified Fn3 proteins were obtained and could be used to create protein-fluorophore conjugates.

For the Fn3 proteins labeled with alkyne-bearing analogues in group 1, copper-catalyzed alkyne-azide cycloaddition (CuAAC) reaction was employed (Figure 4.5). After 2 h of reaction with TAMRA-azide (**8**) and CuAAC reagents, precipitation was again observed. SDS-PAGE was used to characterize both the precipitate and the supernatant. Detection of a fluorescent band from each reaction when scanning for TAMRA fluorescence indicated the successful formation of the Fn3-TAMRA conjugates (Figure 4.6). However, comparison of the fluorescence intensities between the lane for the precipitate and the one containing the supernatant indicated that all of the **F4**-TAMRA conjugates aggregated while some of the **F5**-TAMRA remained in the solution. Attempts to improve the solubility of the conjugates included switching to a more water-soluble fluorophore Alexa Fluor 488 (**9**), adding 1 M NaCl to the reactions, exploring CH_3CN as

an alternative chelator for Cu^{2+} and using DDM in the reactions. However, none of these methods helped to increase the solubility of the **F4**-based conjugates.

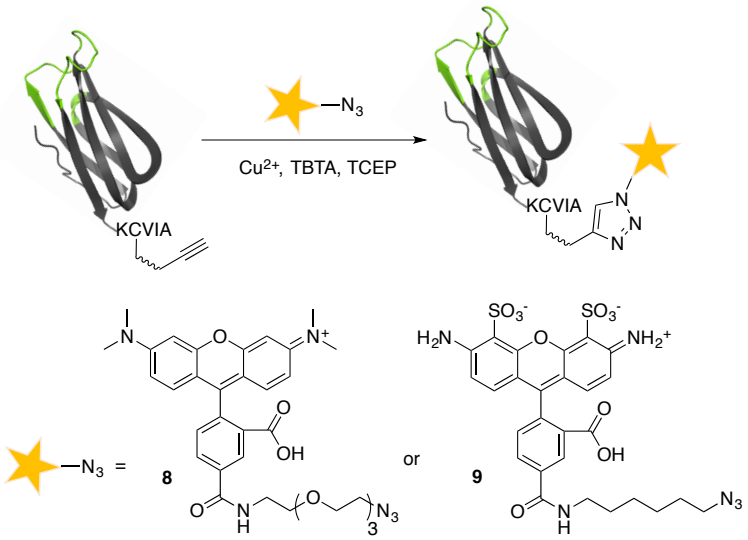


Figure 4.5 Construction of Fn3-fluorophore conjugates using the CuAAC reaction.

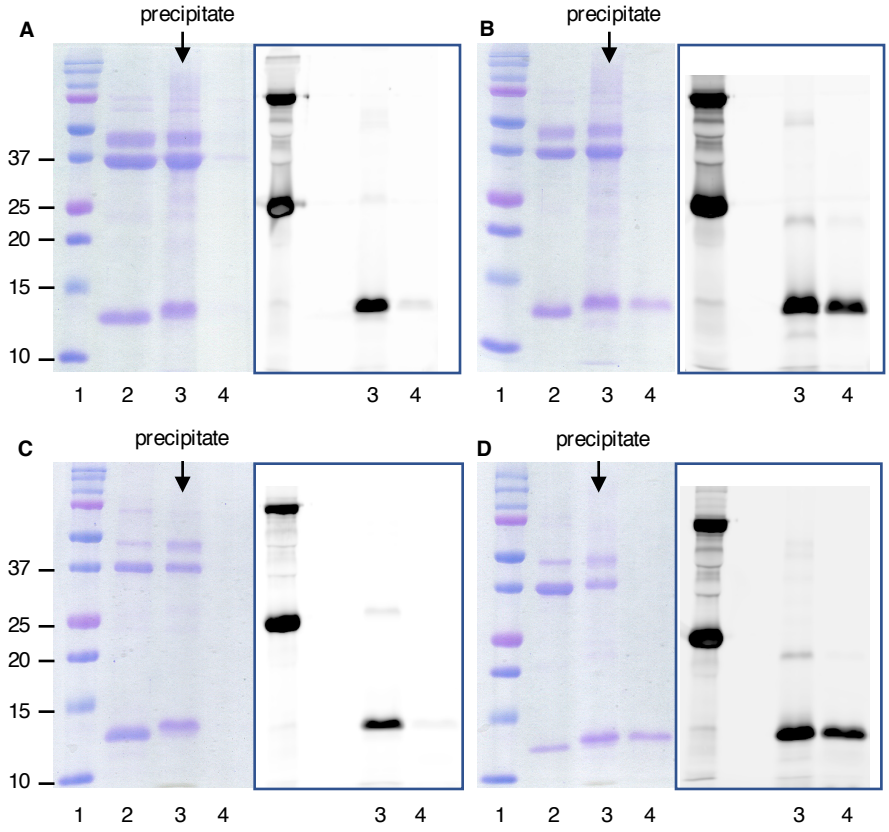


Figure 4.6 SDS-PAGE (left) and in-gel fluorescence analysis (right) of **F4**-TAMRA (A and C) and **F5**-TAMRA (B and D) conjugates constructed with different isoprenoid analogues using CuAAC reactions. (A) **F4-5**-TAMRA conjugates. Lane 1: protein ladder; lane 2: **F4-5**; lane 3: **F4-5**-TAMRA reaction precipitate; lane 4: **F4-5**-TAMRA reaction supernatant; (B) **F5-5**-TAMRA

conjugates. Lane 1: protein ladder; lane 2: **F5-5**; lane 3: **F5-5**-TAMRA reaction precipitate; lane 4: **F5-5**-TAMRA reaction supernatant; (C) **F4-4**-TAMRA conjugates. Lane 1: protein ladder; lane 2: **F4-4**; lane 3: **F4-4**-TAMRA reaction precipitate; lane 4: **F4-4**-TAMRA reaction supernatant; (D) **F5-4**-TAMRA conjugates. Lane 1: protein ladder; lane 2: **F5-4**; lane 3: **F5-4**-TAMRA reaction precipitate; lane 4: **F5-4**-TAMRA reaction supernatant. For **F4**, all the protein-fluorophore conjugates precipitated out while some of the **F5**-fluorophore remained soluble.

For to the **F4** and **F5** proteins modified with the aldehyde-analogue **2** in group 2, oxime ligation reaction was performed with TAMRA-aminooxy analogue (**10**) for 5 h at RT in the presence of *m*-phenylenediamine (mPDA)[57] or *p*-phenylenediamine (pPDA)[58] as a catalyst (Figure 4.7). In contrast to the results from CuAAC reactions, no precipitation was observed. The SDS-PAGE and in-gel fluorescence analysis of the reaction mixtures confirmed the successful formation of both **F4**-TAMRA and **F5**-TAMRA (Figure 4.8). Comparable fluorescence intensities were detected from the reaction catalyzed by mPDA (lane 5) or by pPDA (lane 7), indicating that conjugation similar yields were obtained with these two catalysts under the conditions employed here.

To improve the efficiency of the oxime ligation reaction, we used a different analogue, **7**, to label **F4** and **F5** since the aryl aldehyde in **7** is more reactive than the allylic aldehyde in **2**. After the prenylation reactions, precipitate was removed by centrifugation and the supernatant was used for conjugation with **10** without any catalyst. Formation of the Fn3-TAMRA conjugates were confirmed by SDS-PAGE (Figure 4.9) and LCMS (Figure 4.10) analysis. Comparing the mass intensity of the Fn3-TAMRA and the unreacted Fn3-**7**, the yield was estimated to be 75% for **F4**-TAMRA and 56% for **F5**-TAMRA. It should be noted that an additional impurity was also detected in the **F4**-TAMRA sample, which had a mass of 926 Da and contained the TAMRA fluorophore (absorbed at 555 nm). The structure of this impurity was not assigned.

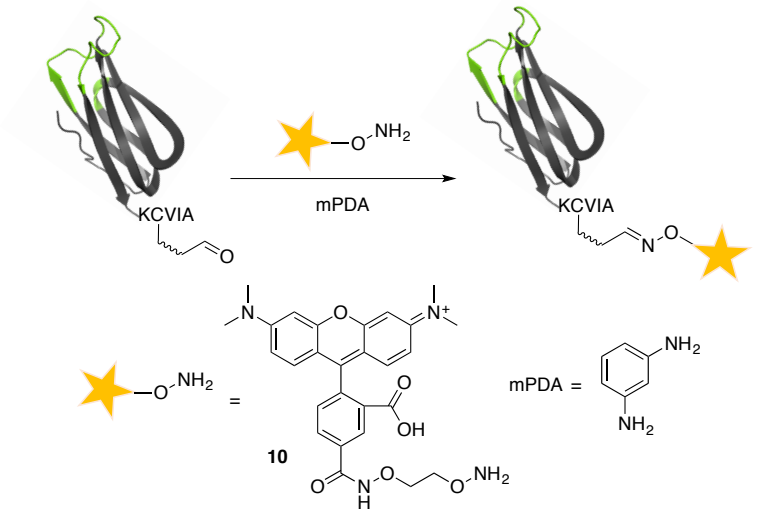


Figure 4.7 Construction of Fn3-fluorophore conjugates using oxime ligation reactions with mPDA as the catalyst.

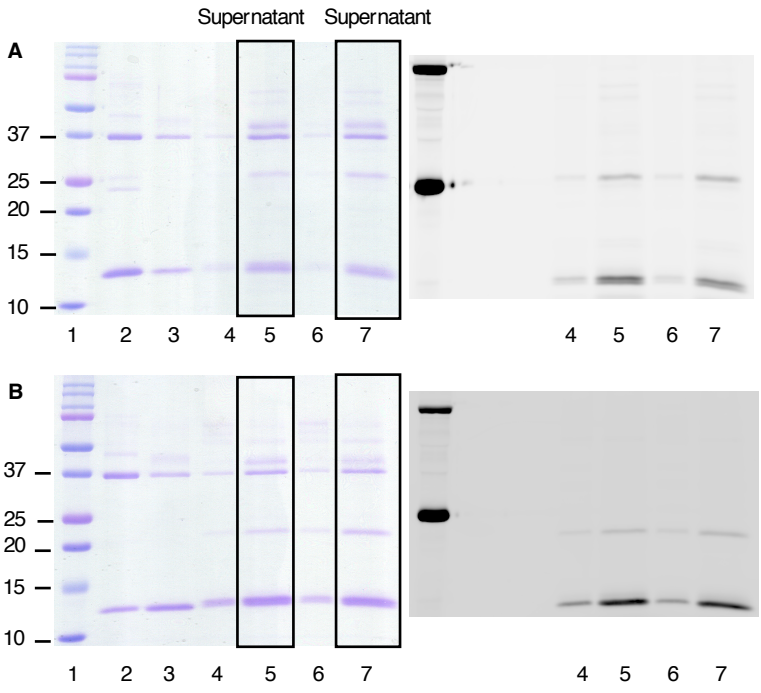


Figure 4.8 SDS (left) and in-gel fluorescence (right) analysis of Fn3-**2**-TAMRA conjugates created using oxime ligation reaction. (A) **F4-2**-TAMRA conjugates. Lane 1 protein ladder; lane 2: precipitate from the **F4-2** prenylation reaction; lane 3: concentrated supernatant from **F4-2** prenylation reaction; lane 4: **F4-2**-TAMRA reaction precipitate with 10 mM mPDA as catalyst; lane 5: **F4-2**-TAMRA supernatant with 10 mM mPDA as catalyst; lane 6: precipitate, 10 mM pPDA as catalyst; lane 7: supernatant, 10 mM pPDA. (B) **F5-2**-TAMRA conjugates. Lane 1: protein ladder; lane 2: precipitate from the **F5-2** prenylation reaction; lane 3: concentrated supernatant from **F5-2** prenylation reaction; lane 4: **F5-2**-TAMRA reaction precipitate with 10 mM mPDA as catalyst; lane 5: **F5-2**-TAMRA supernatant with 10 mM mPDA as catalyst; lane 6: precipitate, 10 mM pPDA as catalyst; lane 7: supernatant, 10 mM pPDA. The reaction efficiency did not appear to be different between mPDA and pPDA catalyzed reactions. Both **F4-2**-TAMRA and **F5-2**-TAMRA conjugates were mostly soluble.

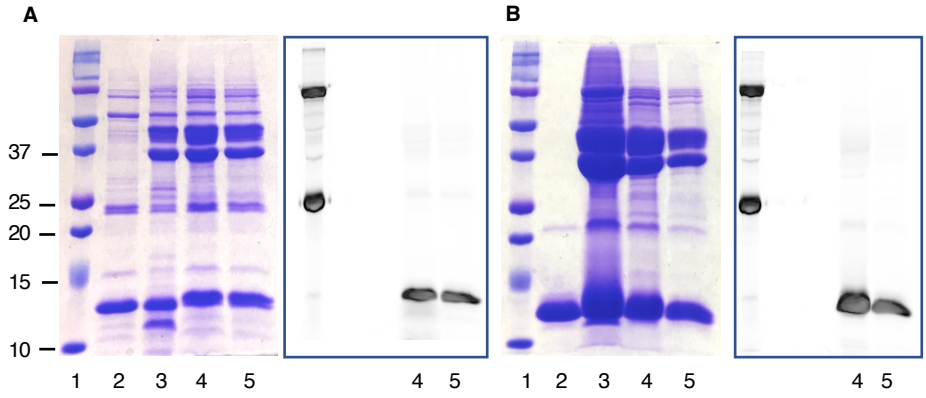


Figure 4.9 SDS-PAGE (left) and in-gel fluorescence analysis (right) of Fn3-**7**-TAMRA conjugates. (A) **F4**-**7**-TAMRA conjugates. Lane 1: protein ladder; lane 2: unmodified **F4**; lane 3: precipitate from **F4**-**7** prenylation reaction; lane 4: **F4**-**7**-TAMRA reaction mixture; lane 5: **F4**-**7**-TAMRA after NAP-5 purification. (B) **F5**-**7**-TAMRA conjugates. Lane 1: protein ladder; lane 2: unmodified **F5**; lane 3: precipitate from **F5**-**7** prenylation reaction; lane 4: **F5**-**7**-TAMRA reaction mixture; lane 5: **F5**-**7**-TAMRA after NAP-5 purification.

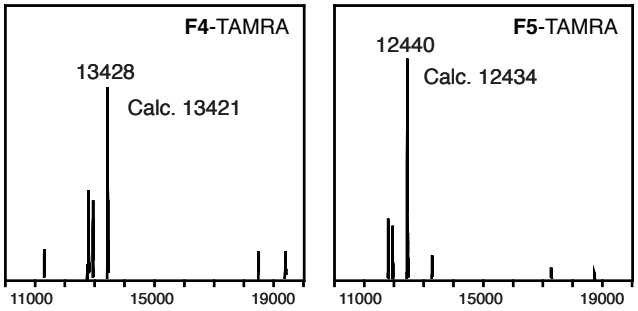


Figure 4.10 Mass of **F4**-TAMRA and **F5**-TAMRA made from Fn3-**7** and **10**

Finally, for the analogue **6** in group 3, the modified Fn3 proteins were first reacted with an aminooxy-containing PEG polymer using the oxime ligation reaction for 2 h at RT, followed by the addition of **8** and CuAAC reagents. Although no visible pellet was observed after the oxime ligation reaction, precipitate was detected following the addition of CuAAC reagents. Examination of the pellet by SDS-PAGE revealed minimal formation of the PEGylated Fn3. The precipitate was mainly **F4**-TAMRA. Therefore, the first oxime ligation step was not efficient and further optimization of the conditions is needed. The solubility of all the Fn3-fluorophore conjugates prepared above is summarized in Table 4.2.

Table 4.1 Solubility of Fn3-fluorophore conjugates.

Analogue	Fluorophore	Additive	Reaction	**F4**	**F5**
5	**8** or **9**	-	CuAAC	L*	M
4	**8** or **9**	-	CuAAC	L	M
4	**8**	DDM	CuAAC	L	M
6	**8**	-	CuAAC	L	M
2	**10**	-	Oxime	H	H
7	**10**	-	Oxime	H	H

* L. low solubility, M. medium solubility, H. high solubility

4.3.5 Characterization of F4-TAMRA binding to cell-surface EGFR.

The binding of **F4**-TAMRA conjugates to cell-surface EGFR was evaluated by flow cytometry. The **F4**-TAMRA sample prepared from analogue **7** modified proteins was utilized, along with the **F5**-TAMRA that served as negative control. We initially tried to use size exclusion chromatography (SEC) to remove the PFTase (90 kDa) from the **F4**-TAMRA sample. However, that attempt was not successful and little amount of purified Fn3-fluorophore conjugates was collected (less than 10% yield). It is not clear whether the conjugates aggregated during the purification process. As a result, the crude samples were utilized to examine the binding capability of **F4**-TAMRA conjugates, since the presence of PFTase should not interfere with the binding of **F4** to its cell-surface target.

Two cell lines were tested, including MDA-MB-468 and MCF-7 cells. It was reported that the EGFR expression for MDA-MB-468 was almost three orders of magnitude higher than that of the MCF-7 cells when comparing their corresponding mRNA levels.[356] Both cells were incubated with 100 nM of **F4**-TAMRA or **F5**-TAMRA for 15 min on ice in the dark and subjected to flow cytometry analysis. The results are provided in Figure 4.11 and summarized in Table 4.3. Examining the fluorescence intensities obtained from MDA-MB-468 cells, only a 2-fold increase was seen for those incubated with **F4**-TAMRA compared with that obtained with **F5**-TAMRA. In addition, similar fluorescence intensities were detected from **F4**-TAMRA treated MDA-MB-468 and MCF-7 cells. All these results indicated that the **F4**-TAMRA in this sample did not bind to cell-surface EGFR efficiently. It is likely that the unconjugated **F4**-**7** partially blocked the binding of **F4**-TAMRA. Moreover, it is also not clear whether the impurity in the **F4**-TAMRA sample interfered with the binding process by interaction with the surface-

exposed antigen recognition site on **F4**. Therefore, a cleaner sample is required to better access the binding of **F4**-TAMRA to its designated target.

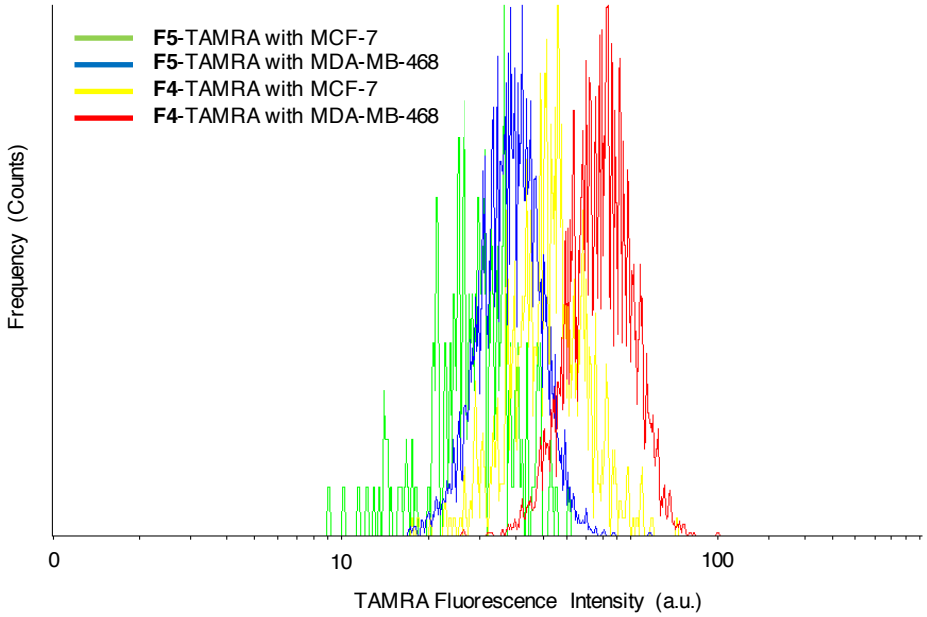


Figure 4.11 Fn3-TAMRA binding to EGFR expressing cells. MDA-MB-468 cells have much higher EGFR expression level than that of the MCF-7 cells. Cells were incubated with 100 nM of **F4**-TAMRA or **F5**-TAMRA for 15 minutes and then washed once with ice cold PBSA.

Table 4.2 Mean TAMRA fluorescence intensities from MDA-MB-468 and MCF-7 cells treated with **F4** and **F5** conjugates.

Conjugates	Mean a. u.	
	MCF-7	MDA-MB-468
F5-TAMRA	35	39.4
F4-TAMRA	60	81

4.3.6. Comparing the expression yield of Fn3 with different C-terminal tags.

One limiting factor in preparing Fn3-based conjugates is the low expression yield of Fn3 proteins from *E. coli*. To investigate whether the addition of the CVIA sequence has a negative impact on protein expression, several Fn3 constructs bearing different C-terminal sequences were created, including EI3.4.3-KGSGK (**F6**), E6.2.6-KGSGK (**F7**) and E6.2.6-no tag (**F8**). These were all transformed into BL21(DE3) *E. coli* cells respectively and the proteins were purified from 1L of culture. It was found that for the

EI3.4.3 variants, the yield for **F6** (5.6 mg) was almost 10-fold higher than that of **F1** (0.6 mg). As to E6.2.6 variants, **F7** had the highest yield of 0.88 mg, followed by **F8** (0.4 mg), with **F4** the lowest (0.18 mg). These results strongly suggested that the small tag sequences at the C-terminus of Fn3 proteins indeed influenced their expression yield. If further engineering of the Fn3 proteins is necessary, the KGSGK sequence can be potentially inserted upstream of the CVIA sequence to improve expression yield.

4.3.7. Labeling of VHH protein by PFTase.

In addition to the Fn3 proteins studied above, we also investigated the labeling of VHH protein, a different protein scaffold, by PFTase. VHH7, a VHH construct targeting the murine Class II MHC molecule,[176] was engineered with a C-terminal GCVIA sequence (VHH7-GCVIA, **V1**). Since there is an internal disulfide bond in the protein construct, periplasmic expression in bacteria was chosen to promote the formation of the correct disulfide bond.[367] Approoximately 6 mg of the protein was obtained after affinity purification using Ni-NTA resins. Prenylation reaction was then performed with **V1** using different isoprenoid analogues (**1**, **5** and **7**) by PFTase. Unfortunately, a large pellet was detected after labeling with **1** and **7**. Since the internal disulfide bond is important for the protein stability, it is possible that the high DTT concentration (5 mM) used in the prenylation buffer completely reduced this disulfide bond, making the protein less stable when an additional hydrophobic moiety is attached. Therefore, different conditions were tested to partially reduce **V1** (using only 1.25 eq of DTT relative to **V1** or by first reducing **V1** with 70 eq TCEP and then removing the reducing agent). DDM was also added to the reaction mixture to help solubilize the prenylated proteins. However, none of these methods resolved the solubility problem. Although less precipitate was observed when **V1** was labeled with analogue **5**, when the modified proteins reacted with **8** using the CuAAC reaction, all the conjugates precipitated.

4.4 Conclusion

In conclusion, two types of alternative binding proteins, Fn3 and VHH proteins, were evaluated in the enzymatic labeling reactions with PFTase. It was shown that both

Fn3 and VHH variants engineered with C-terminal CVIA sequences can be efficiently modified by PFTase. However, all the proteins being studied had solubility problem after the incorporation of the relatively hydrophobic isoprenoid moiety. Multiple attempts were made to prevent precipitate formation with limited success. Due to the aggregation problem, large quantities of starting materials are required to provide enough conjugates for purification and characterization. Unfortunately, the low expression yields of the current Fn3 constructs limit the scale of the reactions. Further optimization of the Fn3 proteins to create more soluble variants is highly desirable to solve the solubility issues. This work also reveals that the insertion of even a small tag sequence may affect the solubility and expression yields of these small proteins, which should be considered when performing protein engineering. Concerning the conjugation reactions with the modified proteins, we found that protein aggregation occurred in the CuAAC reactions, which is likely caused by the Cu^{2+} catalyst. In contrast, protein-fluorophore conjugates constructed using oxime ligation were highly soluble. Therefore, for future protein conjugation applications, oxime ligation and the copper-free strained-promoted alkyne-azide cycloadditon should be first considered to avoid potential precipitation problems.

4.5 Supporting Data

4.5.1. MS of EI3.4.3 variants modified with analogues 2 and 3 by PFTase.

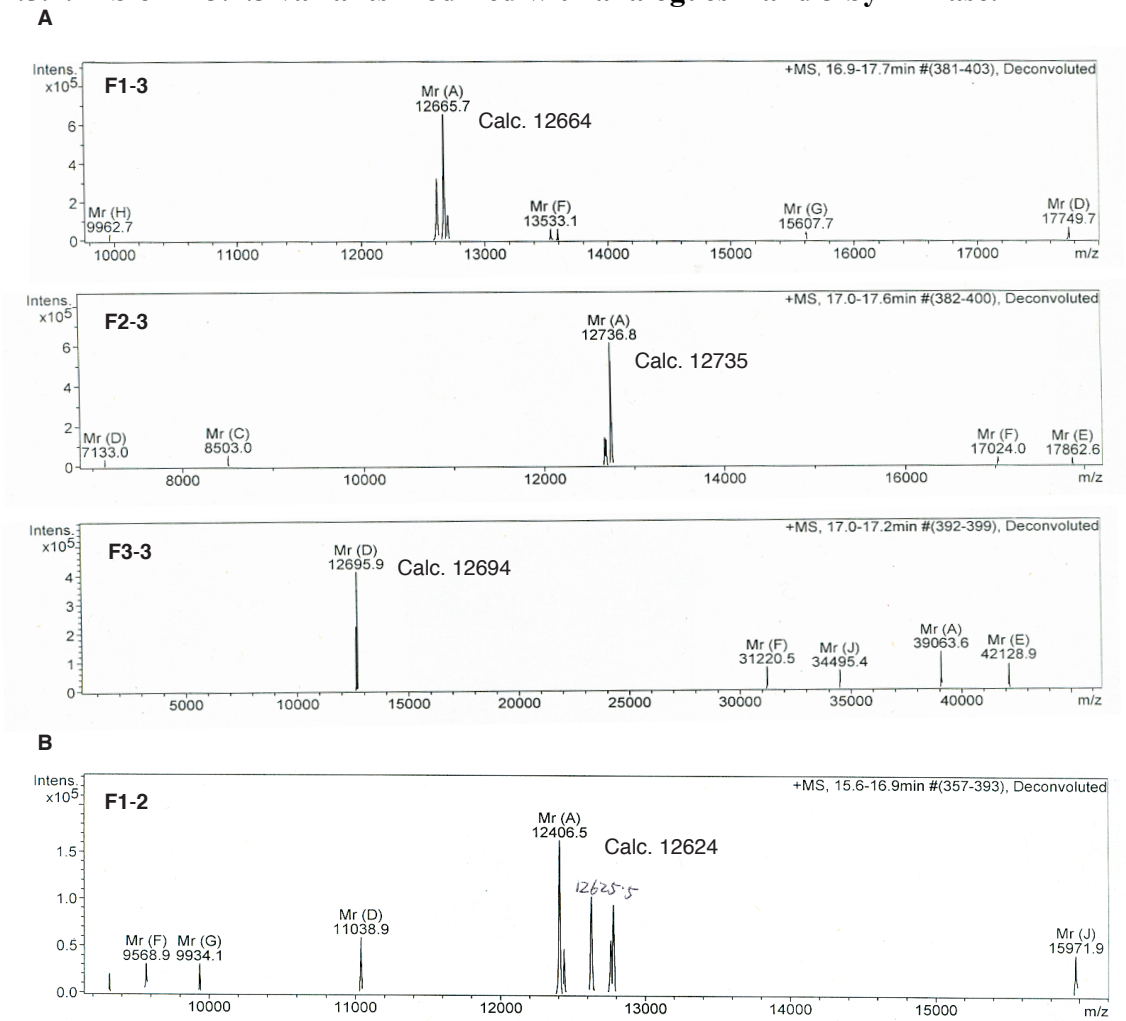


Figure S4.1 MS of EI3.4.3 variants modified with analogues **3** and **2** by PFTase.

4.5.2. MS of selected analogue modified F4 and F5.

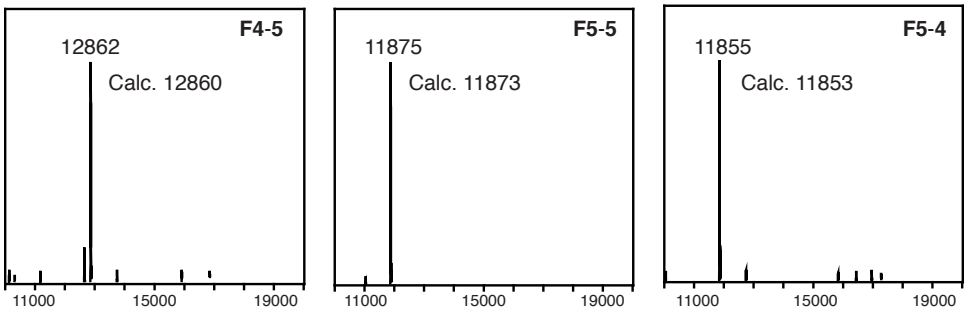


Figure S4.2 MS of **F4-5**, **F5-5** and **F5-4**. The mass for **F4-4** was not tested due to low solubility.

Reference

1. G. Charron, M. M. Li, M. R. MacDonald and H. C. Hang, *Proc. Natl. Acad. Sci. U S A*, 2013, **110**, 11085-11090.
2. E. C. Jensen, *Anat. Rec.* , 2012, **295**, 2031-2036.
3. D. P. Staus, L. M. Wingler, M. Choi, B. Pani, A. Manglik, A. C. Kruse and R. J. Lefkowitz, *Proc. Natl. Acad. Sci. U S A*, 2018, **115**, 3834-3839.
4. A. Ibraheem and R. E. Campbell, *Curr. Opin. Chem. Biol.*, 2010, **14**, 30-36.
5. S. Datta, L. R. Christena and Y. R. Rajaram, *3 Biotech.*, 2013, **3**, 1-9.
6. F. X. R. Sutandy, J. Qian, C.-S. Chen and H. Zhu, *Curr. Protoc. Protein Sci.*, 2013, **72**, 27.21.21-27.21.16.
7. M. Wang and M. Thanou, *Pharmacol. Res.*, 2010, **62**, 90-99.
8. R. V. Ulijn, N. Bibi, V. Jayawarna, P. D. Thornton, S. J. Todd, R. J. Mart, A. M. Smith and J. E. Gough, *Mater. Today*, 2007, **10**, 40-48.
9. H. A. Lagasse, A. Alexaki, V. L. Simhadri, N. H. Katagiri, W. Jankowski, Z. E. Sauna and C. Kimchi-Sarfaty, *F1000Res*, 2017, **6**, 113.
10. J. K. Dozier and M. D. Distefano, *Int. J. Mol. Sci.*, 2015, **16**, 25831-25864.
11. A. Beck, L. Goetsch, C. Dumontet and N. Corvaia, *Nat. Rev. Drug Discov.*, 2017, **16**, 315-337.
12. S. M. Larson, J. A. Carrasquillo, N. K. Cheung and O. W. Press, *Nat. Rev. Cancer*, 2015, **15**, 347-360.
13. G. A. van Dongen, G. W. Visser, M. N. Lub-de Hooge, E. G. de Vries and L. R. Perk, *Oncologist*, 2007, **12**, 1379-1389.
14. D. Schumacher, C. P. Hackenberger, H. Leonhardt and J. Helma, *J. Clin. Immunol.*, 2016, **36 Suppl 1**, 100-107.
15. A. C. Braun, M. Gutmann, T. Luhmann and L. Meinel, *J. Controlled Release*, 2018, **273**, 68-85.
16. S. A. Walper, K. B. Turner and I. L. Medintz, in *Chemoselective and Bioorthogonal Ligation Reactions*, ed. W. R. Algar, Dawson, P. E., Medintz, I. L., 2017, DOI: 10.1002/9783527683451.ch7.
17. J. Lotze, U. Reinhardt, O. Seitz and A. G. Beck-Sickinger, *Mol. Biosyst.*, 2016, **12**, 1731-1745.
18. M. Rashidian, J. K. Dozier and M. D. Distefano, *Bioconjugate Chem.*, 2013, **24**, 1277-1294.
19. N. Stephanopoulos and M. B. Francis, *Nat. Chem. Biol.*, 2011, **7**, 876.
20. N. Krall, F. P. da Cruz, O. Boutureira and G. J. L. Bernardes, *Nat. Chem.*, 2015, **8**, 103.
21. Y. Seki, T. Ishiyama, D. Sasaki, J. Abe, Y. Sohma, K. Oisaki and M. Kanai, *J. Am. Chem. Soc.*, 2016, **138**, 10798-10801.
22. Y. Yu, L. K. Zhang, A. V. Buevich, G. Li, H. Tang, P. Vachal, S. L. Colletti and Z.-C. Shi, *J. Am. Chem. Soc.*, 2018, **140**, 6797-6800.
23. S. Lin, X. Yang, S. Jia, A. M. Weeks, M. Hornsby, P. S. Lee, R. V. Nichiporuk, A. T. Iavarone, J. A. Wells, F. D. Toste and C. J. Chang, *Science*, 2017, **355**, 597.
24. C. D. Spicer and B. G. Davis, *Nat. Commun.*, 2014, **5**, 4740.

25. S. B. Gunnoo and M. Annemieke, *ChemBioChem*, 2016, **17**, 529-553.
26. S. Arumugam, J. Guo, N. E. Mbua, F. Friscourt, N. Lin, E. Nekongo, G.-J. Boons and V. V. Popik, *Chem. Sci.*, 2014, **5**, 1591-1598.
27. E. M. Valkevich, R. G. Guenette, N. A. Sanchez, Y.-c. Chen, Y. Ge and E. R. Strieter, *J. Am. Chem. Soc.*, 2012, **134**, 6916-6919.
28. M. L. Conte, S. Staderini, A. Marra, M. Sanchez-Navarro, B. G. Davis and A. Dondoni, *Chem. Commun.*, 2011, **47**, 11086-11088.
29. J. M. Chalker, G. J. L. Bernardes, Y. A. Lin and B. G. Davis, *Chem. Asian J.* , 2009, **4**, 630-640.
30. E. M. Sletten and C. R. Bertozzi, *Angew. Chem. Int. Ed.*, 2009, **48**, 6974-6998.
31. C. B. Rosen and M. B. Francis, *Nat. Chem. Biol.*, 2017, **13**, 697.
32. S. Kalkhof and A. Sinz, *Anal. Bioanal. Chem.*, 2008, **392**, 305-312.
33. T. Nakamura, Y. Kawai, N. Kitamoto, T. Osawa and Y. Kato, *Chem. Res. Toxicol.*, 2009, **22**, 536-542.
34. M. K. Patel, B. Vijayakrishnan, J. R. Koeppe, J. M. Chalker, K. J. Doores and B. G. Davis, *Chem. Commun.*, 2010, **46**, 9119-9121.
35. X. Chen, K. Muthoosamy, A. Pfisterer, B. Neumann and T. Weil, *Bioconjugate Chem.*, 2012, **23**, 500-508.
36. G. H. Pham, W. Ou, B. Bursulaya, M. DiDonato, A. Herath, Y. Jin, X. Hao, J. Loren, G. Spraggon, A. Brock, T. Uno, B. H. Geierstanger and S. E. Cellitti, *ChemBioChem*, 2018, **19**, 799-804.
37. T. L. Hendrickson, V. de Crécy-Lagard and P. Schimmel, *Annu. Rev. Biochem.*, 2004, **73**, 147-176.
38. C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413-444.
39. N. Heinz, *FEBS Lett.*, 2012, **586**, 2057-2064.
40. K. Lang and J. W. Chin, *Chem. Rev.*, 2014, **114**, 4764-4806.
41. A. J. de Graaf, M. Kooijman, W. E. Hennink and E. Mastrobattista, *Bioconjugate Chem.*, 2009, **20**, 1281-1295.
42. A. J. Link and D. A. Tirrell, *J. Am. Chem. Soc.*, 2003, **125**, 11164-11165.
43. D. C. Dieterich, A. J. Link, J. Graumann, D. A. Tirrell and E. M. Schuman, *Proc. Natl. Acad. Sci. U S A*, 2006, **103**, 9482.
44. W. Song, Y. Wang, Z. Yu, C. I. R. Vera, J. Qu and Q. Lin, *ACS Chem. Biol.*, 2010, **5**, 875-885.
45. K. Wals and H. Ovaa, *Front. Chem.*, 2014, **2**, 15.
46. L. Wang and P. G. Schultz, *Chem. Biol.*, 2001, **8**, 883-890.
47. J. Wang, J. Xie and P. G. Schultz, *J. Am. Chem. Soc.*, 2006, **128**, 8738-8739.
48. H. W. Ai, W. Shen, A. Sagi, P. R. Chen and P. G. Schultz, *ChemBioChem*, 2011, **12**, 1854-1857.
49. J. P. Gallivan, H. A. Lester and D. A. Dougherty, *Chem. Biol.*, 1997, **4**, 739-749.
50. C. H. Kim, J. Y. Axup and P. G. Schultz, *Curr. Opin. Chem. Biol.*, 2013, **17**, 412-419.
51. K. Lang and J. W. Chin, *ACS Chem. Biol.*, 2014, **9**, 16-20.
52. C. S. McKay and M. G. Finn, *Chem. Biol.*, 2014, **21**, 1075-1101.
53. X. Chen and Y. W. Wu, *Org. Biomol. Chem.*, 2016, **14**, 5417-5439.
54. J. A. Prescher and C. R. Bertozzi, *Nat. Chem. Biol.*, 2005, **1**, 13.

55. A. Dirksen, T. M. Hackeng and P. E. Dawson, *Angew. Chem. Int. Ed.* , 2006, **45**, 7581-7584.
56. P. Crisalli and E. T. Kool, *J. Org. Chem.*, 2013, **78**, 1184-1189.
57. M. Rashidian, M. M. Mahmoodi, R. Shah, J. K. Dozier, C. R. Wagner and M. D. Distefano, *Bioconjugate Chem.*, 2013, **24**, 333-342.
58. M. Wendeler, L. Grinberg, X. Wang, P. E. Dawson and M. Baca, *Bioconjugate Chem.*, 2014, **25**, 93-101.
59. E. T. Kool, D. H. Park and P. Crisalli, *J. Am. Chem. Soc.*, 2013, **135**, 17663-17666.
60. J. Kalia and R. T. Raines, *Angew. Chem. Int. Ed.*, 2008, **47**, 7523-7526.
61. P. Agarwal, J. van der Weijden, E. M. Sletten, D. Rabuka and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U S A*, 2013, **110**, 46-51.
62. P. Agarwal, R. Kudirka, A. E. Albers, R. M. Barfield, G. W. de Hart, P. M. Drake, L. C. Jones and D. Rabuka, *Bioconjugate Chem.*, 2013, **24**, 846-851.
63. R. Kudirka, R. M. Barfield, J. McFarland, A. E. Albers, G. W. de Hart, P. M. Drake, P. G. Holder, S. Banas, L. C. Jones, A. W. Garofalo and D. Rabuka, *Chem. Biol.*, 2015, **22**, 293-298.
64. I. S. Carrico, B. L. Carlson and C. R. Bertozzi, *Nat. Chem. Biol.*, 2007, **3**, 321-322.
65. M. Rashidian, J. M. Song, R. E. Pricer and M. D. Distefano, *J. Am. Chem. Soc.*, 2012, **134**, 8455-8467.
66. H. Gao, W. Sun, Z. Song, Y. Yu, L. Wang, X. Chen and Q. Zhang, *Chembiochem*, 2017, **18**, 324-330.
67. D. Schumacher, J. Helma, F. A. Mann, G. Pichler, F. Natale, E. Krause, M. C. Cardoso, C. P. Hackenberger and H. Leonhardt, *Angew. Chem. Int. Ed.*, 2015, **54**, 13787-13791.
68. M. A. Gray, R. N. Tao, S. M. DePorter, D. A. Spiegel and B. R. McNaughton, *Chembiochem*, 2016, **17**, 155-158.
69. Y. Zeng, T. N. Ramya, A. Dirksen, P. E. Dawson and J. C. Paulson, *Nat. Methods*, 2009, **6**, 207-209.
70. M. F. Debets, C. W. van der Doelen, F. P. Rutjes and F. L. van Delft, *Chembiochem*, 2010, **11**, 1168-1184.
71. O. Nieto-García, M. B. Jaffee, M. Mühlberg and C. P. R. Hackenberger, in *Chemoselective and Bioorthogonal Ligation Reactions*, ed. W. R. Algar, Dawson, P. E., Medintz, I. L., 2017, DOI: 10.1002/9783527683451.ch4.
72. Z. P. A. Wang, C. L. Tian and J. S. Zheng, *RSC Adv.*, 2015, **5**, 107192-107199.
73. J. E. Hein, in *In Chemoselective and Bioorthogonal Ligation Reactions*, ed. W. R. Algar, Dawson, P. E., Medintz, I. L., 2017, DOI: 10.1002/9783527683451.ch2.
74. T. R. Chan, R. Hilgraf, K. B. Sharpless and V. V. Fokin, *Org. Lett.*, 2004, **6**, 2853-2855.
75. V. Hong, S. I. Presolski, C. Ma and M. G. Finn, *Angew. Chem. Int. Ed.*, 2009, **48**, 9879-9883.
76. D. S. del Amo, W. Wang, H. Jiang, C. Besanceney, A. C. Yan, M. Levy, Y. Liu, F. L. Marlow and P. Wu, *J. Am. Chem. Soc.*, 2010, **132**, 16893-16899.
77. S. I. Presolski, V. Hong, S. H. Cho and M. G. Finn, *J. Am. Chem. Soc.*, 2010, **132**, 14570-14576.

78. C. Besanceney-Webler, H. Jiang, T. Zheng, L. Feng, D. Soriano del Amo, W. Wang, L. M. Klivansky, F. L. Marlow, Y. Liu and P. Wu, *Angew. Chem. Int. Ed.* , 2011, **50**, 8051-8056.
79. D. C. Kennedy, C. S. McKay, M. C. Legault, D. C. Danielson, J. A. Blake, A. F. Pegoraro, A. Stolow, Z. Mester and J. P. Pezacki, *J. Am. Chem. Soc.*, 2011, **133**, 17993-18001.
80. G. C. Kuang, H. A. Michaels, J. T. Simmons, R. J. Clark and L. Zhu, *J. Org. Chem.*, 2010, **75**, 6540-6548.
81. C. Uttamapinant, A. Tangpeerachaikul, S. Grecian, S. Clarke, U. Singh, P. Slade, K. R. Gee and A. Y. Ting, *Angew. Chem. Int. Ed.*, 2012, **51**, 5852-5856.
82. H. Jiang, T. Zheng, A. Lopez-Aguilar, L. Feng, F. Kopp, F. L. Marlow and P. Wu, *Bioconjugate Chem.*, 2014, **25**, 698-706.
83. P. V. Chang, J. A. Prescher, E. M. Sletten, J. M. Baskin, I. A. Miller, N. J. Agard, A. Lo and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U S A*, 2010, **107**, 1821-1826.
84. K. W. Dehnert, J. M. Baskin, S. T. Laughlin, B. J. Beahm, N. N. Naidu, S. L. Amacher and C. R. Bertozzi, *Chembiochem*, 2012, **13**, 353-357.
85. R. van Geel, G. J. Pruijn, F. L. van Delft and W. C. Boelens, *Bioconjugate Chem.*, 2012, **23**, 392-398.
86. B. L. Oliveira, Z. Guo and G. J. L. Bernardes, *Chem. Soc. Rev.*, 2017, **46**, 4895-4950.
87. M. R. Karver, R. Weissleder and S. A. Hilderbrand, *Bioconjugate Chem.*, 2011, **22**, 2263-2270.
88. T. J. Wades, E. H. Wong, G. R. Weisman and C. J. Anderson, *Chem. Rev.*, 2010, **110**, 2858-2902.
89. R. Rossin, T. Lappchen, S. M. van den Bosch, R. Laforest and M. S. Robillard, *J. Nucl. Med.*, 2013, **54**, 1989-1995.
90. D. M. Patterson and J. A. Prescher, *Curr. Opin. Chem. Biol.*, 2015, **28**, 141-149.
91. N. M. Rachel, J. L. Toulouse and J. N. Pelletier, *Bioconjugate Chem.*, 2017, **28**, 2518-2523.
92. M. Rashidian, L. Wang, J. G. Edens, J. T. Jacobsen, I. Hossain, Q. Wang, G. D. Victora, N. Vasdev, H. Ploegh and S. H. Liang, *Angew. Chem. Int. Ed.*, 2016, **55**, 528-533.
93. D. S. Liu, A. Tangpeerachaikul, R. Selvaraj, M. T. Taylor, J. M. Fox and A. Y. Ting, *J. Am. Chem. Soc.*, 2012, **134**, 792-795.
94. M. Best, A. Degen, M. Baalmann, T. Schmidt Tobias and R. Wombacher, *ChemBioChem*, 2015, **16**, 1158-1162.
95. J. W. Wollack, B. J. Monson, J. K. Dozier, J. J. Dalluge, K. Poss, S. A. Hilderbrand and M. D. Distefano, *Chem. Biol. Drug. Des.*, 2014, **84**, 140-147.
96. E. M. Milczek, *Chem. Rev.*, 2018, **118**, 119-141.
97. I. Chen, B. M. Dorr and D. R. Liu, *Proc. Natl. Acad. Sci. U S A*, 2011, **108**, 11399-11404.
98. J. E. Glasgow, M. L. Salit and J. R. Cochran, *J. Am. Chem. Soc.*, 2016, **138**, 7496-7499.
99. K. Alt, B. M. Paterson, E. Westein, S. E. Rudd, S. S. Poniger, S. Jagdale, K. Ardipradja, T. U. Connell, G. Y. Krippner, A. K. Nair, X. Wang, H. J. Tochon-

Danguy, P. S. Donnelly, K. Peter and C. E. Hagemeyer, *Angew. Chem. Int. Ed.*, 2015, **54**, 7515-7519.
100. S. A. van Lith, S. M. van Duijnhoven, A. C. Navis, W. P. Leenders, E. Dolk, J. W. Wennink, C. F. van Nostrum and J. C. van Hest, *Bioconjugate Chem.*, 2017, **28**, 539-548.
101. J. Hu, G. Wang, W. Zhao, X. Liu, L. Zhang and W. Gao, *Biomaterials*, 2016, **96**, 84-92.
102. M. D. Witte, J. J. Cragnolini, S. K. Dougan, N. C. Yoder, M. W. Popp and H. L. Ploegh, *Proceedings of the National Academy of Sciences*, 2012, **109**, 11993.
103. C. Uth, S. Zielonka, S. Horner, N. Rasche, A. Plog, H. Orelma, O. Avrutina, K. Zhang and H. Kolmar, *Angew. Chem. Int. Ed.*, 2014, **53**, 12618-12623.
104. Q. Wu, H. L. Ploegh and M. C. Truttmann, *ACS Chem. Biol.*, 2017, **12**, 664-673.
105. G. Pasqual, A. Chudnovskiy, J. M. J. Tas, M. Agudelo, L. D. Schweitzer, A. Cui, N. Hacohen and G. D. Victora, *Nature*, 2018, **553**, 496-500.
106. L. Abrahmsen, J. Tom, J. Burnier, K. A. Butcher, A. Kossiakoff and J. A. Wells, *Biochemistry*, 1991, **30**, 4151-4159.
107. A. M. Weeks and J. A. Wells, *Nat. Chem. Biol.*, 2018, **14**, 50-57.
108. S. H. Henager, N. Chu, Z. Chen, D. Bolduc, D. R. Dempsey, Y. Hwang, J. A. Wells and P. A. Cole, *Nat. Methods*, 2016, **13**, 925-927.
109. N. M. Rachel and J. N. Pelletier, *Chem. Commun.*, 2016, **52**, 2541-2544.
110. S. Jeger, K. Zimmermann, A. Blanc, J. Grunberg, M. Honer, P. Hunziker, H. Struthers and R. Schibli, *Angew. Chem. Int. Ed.*, 2010, **49**, 9995-9997.
111. A. Mero, A. Grigoletto, K. Maso, H. Yoshioka, A. Rosato and G. Pasut, *Polym. Chem.*, 2016, **7**, 6545-6553.
112. T. Tanaka, N. Kamiya and T. Nagamune, *FEBS Lett*, 2005, **579**, 2092-2096.
113. J. Tominaga, N. Kamiya, S. Doi, H. Ichinose, T. Maruyama and M. Goto, *Biomacromolecules*, 2005, **6**, 2299-2304.
114. C. W. Lin and A. Y. Ting, *128*, 2006, **4542-4543**.
115. M. T. Yang, C. H. Chang, J. M. Wang, T. K. Wu, Y. K. Wang, C. Y. Chang and T. T. Li, *J Biol Chem*, 2011, **286**, 7301-7307.
116. C. Gauchet, G. R. Labadie and C. D. Poulter, *J. Am. Chem. Soc.*, 2006, **128**, 9274-9275.
117. A. Hosokawa, J. W. Wollack, Z. Zhang, L. Chen, G. Barany and M. D. Distefano, *Int. J. Pept. Res. Ther.*, 2007, **13**, 345-354.
118. J. J. Lee, H. J. Choi, M. Yun, Y. Kang, J. E. Jung, Y. Ryu, T. Y. Kim, Y. J. Cha, H. S. Cho, J. J. Min, C. W. Chung and H. S. Kim, *Angew. Chem. Int. Ed.*, 2015, **54**, 12020-12024.
119. T. C. Turek-Etienne, C. L. Strickland and M. D. Distefano, *Biochemistry*, 2003, **42**.
120. M. Rashidian, S. C. Kumarapperuma, K. Gabrielse, A. Fegan, C. R. Wagner and M. D. Distefano, *J. Am. Chem. Soc.*, 2013, **135**, 16388-16396.
121. Y. Zhang, M. J. Blanden, C. Sudheer, S. A. Gangopadhyay, M. Rashidian, J. L. Hougland and M. D. Distefano, *Bioconjugate Chem.*, 2015, **26**, 2542-2553.
122. J. E. Yeo, S. Wickramaratne, S. Khatwani, Y. C. Wang, J. Vervacke, M. D. Distefano and N. Y. Tretyakova, *ACS Chem. Biol.*, 2014, **9**, 1860-1868.
123. J. S. Seo, S. Lee and C. D. Poulter, *J. Am. Chem. Soc.*, 2013, **135**, 8973-8980.

124. N. S. Kishore, T. B. Lu, L. J. Knoll, A. Katoh, D. A. Rudnick, P. P. Mehta, B. Devadas, M. Huhn, J. L. Atwood, S. P. Adams, G. W. Gokel and J. I. Gordon, *J. Biol. Chem.*, 1991, **266**, 8835-8855.
125. S. H. Ho and D. A. Tirrell, *J. Am. Chem. Soc.*, 2016, **138**, 15098-15101.
126. W. P. Heal, S. R. Wickramasinghe, R. J. Leatherbarrow and E. W. Tate, *Org. Biomol. Chem.*, 2008, **6**, 2308-2315.
127. C. Kulkarni, T. L. Kinzer-Ursem and D. A. Tirrell, *Chembiochem*, 2013, **14**, 1958-1962.
128. C. Kulkarni, M. Lo, J. G. Fraseur, D. A. Tirrell and T. L. Kinzer-Ursem, *Bioconjugate Chem.*, 2015, **26**, 2153-2160.
129. 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.
130. J. Grünewald, H. E. Klock, S. E. Cellitti, B. Bursulaya, D. McMullan, D. H. Jones, H. P. Chiu, X. Wang, P. Patterson, H. Zhou, J. Vance, E. Nigoghossian, H. Tong, D. Daniel, W. Mallet, W. Ou, T. Uno, A. Brock, S. A. Lesley and B. H. Geierstanger, *Bioconjugate Chem.*, 2015, **26**, 2554-2562.
131. D. A. Pippig, F. Baumann, M. Strackharn, D. Aschenbrenner and H. E. Gaub, *ACS Nano*, 2014, **8**, 6551-6555.
132. M. Sunbul, M. Yen, Y. Zou and J. Yin, *Chemical Communications*, 2008, DOI: 10.1039/B812162A, 5927-5929.
133. D. Schumacher, O. Lemke, J. Helma, L. Gerszonowicz, V. Waller, T. Stoschek, P. M. Durkin, N. Budisa, H. Leonhardt, B. G. Keller and C. P. R. Hackenberger, *Chem. Sci.*, 2017, **8**, 3471-3478.
134. M. Fernández-Suárez, H. Baruah, L. Martínez-Hernández, K. T. Xie, J. M. Baskin, C. R. Bertozzi and A. Y. Ting, *Nat. Biotechnol.*, 2007, **25**, 1483.
135. J. G. Plaks, R. Falatach, M. Kastantin, J. A. Berberich and J. L. Kaar, *Bioconjugate Chem.*, 2015, **26**, 1104-1112.
136. S. Hauke, M. Best, T. T. Schmidt, M. Baalmann, A. Krause and R. Wombacher, *Bioconjugate Chem.*, 2014, **25**, 1632-1637.
137. A. Chapman-Smith, T. D. Mulhern, F. Whelan, J. E. Cronan, Jr. and J. C. Wallace, *Protein Sci.*, 2001, **10**, 2608-2617.
138. I. Chen, M. Howarth, W. Lin and A. Y. Ting, *Nat. Methods*, 2005, **2**, 99.
139. S. A. Slavoff, I. Chen, Y.-A. Choi and A. Y. Ting, *J. Am. Chem. Soc.*, 2008, **130**, 1160-1162.
140. S. Sueda, S. Yoneda and H. Hayashi, *ChemBioChem*, 2011, **12**, 1367-1375.
141. H. Miyao, Y. Ikeda, A. Shiraishi, Y. Kawakami and S. Sueda, *Anal. Biochem.*, 2015, **484**, 113-121.
142. 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.
143. U. Ilangovan, H. Ton-That, J. Iwahara, O. Schneewind and R. T. Clubb, *Proc. Natl. Acad. Sci. U S A*, 2001, **98**, 6056-6061.
144. A. W. Jacobitz, M. D. Kattke, J. Wereszczynski and R. T. Clubb, *Adv. Protein Chem. Struct. Biol.*, 2017, **109**, 223-264.
145. C. S. Theile, M. D. Witte, A. E. Blom, L. Kundrat, H. L. Ploegh and C. P. Guimaraes, *Nat. Protoc.*, 2013, **8**, 1800-1807.

146. C. P. Guimaraes, M. D. Witte, C. S. Theile, G. Bozkurt, L. Kundrat, A. E. Blom and H. L. Ploegh, *Nat. Protoc.*, 2013, **8**, 1787-1799.
147. H. Hirakawa, S. Ishikawa and T. Nagamune, *Biotechnol. Bioeng.*, 2012, **109**, 2955-2961.
148. M. W. Popp, S. K. Dougan, T. Y. Chuang, E. Spooner and H. L. Ploegh, *Proc. Natl. Acad. Sci. U S A*, 2011, **108**, 3169-3174.
149. T. Matsumoto, R. Takase, T. Tanaka, H. Fukuda and A. Kondo, *Biotechnol. J.*, 2012, **7**, 642-648.
150. S. Das, V. S. Pawale, V. Dadireddy, A. K. Singh, S. Ramakumar and R. P. Roy, *J. Biol. Chem.*, 2017, **292**, 7244-7257.
151. K. D. Nikghalb, N. M. Horvath, J. L. Prelesnik, O. G. B. Banks, P. A. Filipov, R. D. Row, T. J. Roark and J. M. Antos, *Chembiochem*, 2018, **19**, 185-195.
152. J. M. Antos, G. M. Miller, G. M. Grotenbreg and H. L. Ploegh, *J. Am. Chem. Soc.*, 2008, **130**, 16338-16343.
153. M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner and H. L. Ploegh, *Nat. Chem. Biol.*, 2007, **3**, 707-708.
154. M. W. Popp and H. L. Ploegh, *Angew. Chem. Int. Ed.*, 2011, **50**, 5024-5032.
155. M. Ritzefeld, *Chemistry*, 2014, **20**, 8516-8529.
156. X. L. Tan, M. Pan, Y. Zheng, S. Gao, L. J. Liang and Y. M. Li, *Chem. Sci.*, 2017, **8**, 6881-6887.
157. R. K. Le, M. Raeeszadeh-Sarmazdeh, E. T. Boder and P. D. Frymier, *Langmuir*, 2015, **31**, 1180-1188.
158. S. M. Qafari, G. Ahmadian and M. Mohammadi, *RSC Adv.*, 2017, **7**, 56006-56015.
159. S. Garg, G. S. Singaraju, S. Yengkhom and S. Rakshit, *Bioconjugate Chem.*, 2018, **29**, 1714-1719.
160. Z. Zou, D. M. Mate, K. Rubsam, F. Jakob and U. Schwaneberg, *ACS Comb. Sci.*, 2018, **20**, 203-211.
161. H. O. Ham, Z. Qu, C. A. Haller, B. M. Dorr, E. Dai, W. Kim, D. R. Liu and E. L. Chaikof, *Nat. Commun.*, 2016, **7**, 11140.
162. A. Tabata, Y. Ohkuba, N. Anyoji, K. Hojo, T. Tomoyasu, Y. Tatematsu, K. Ohkura and H. Nagamune, *Anticancer Res.*, 2015, **35**, 4411-4418.
163. J. R. Silvius and R. Leventis, *Bioconjugate Chem.*, 2017, **28**, 1271-1282.
164. L. Schoonen, J. Pille, A. Borrmann, R. J. Nolte and J. C. van Hest, *Bioconjugate Chem.*, 2015, **26**, 2429-2434.
165. D. Patterson, B. Schwarz, J. Avera, B. Western, M. Hicks, P. Krugler, M. Terra, M. Uchida, K. McCoy and T. Douglas, *Bioconjugate Chem.*, 2017, **28**, 2114-2124.
166. E. Cambria, K. Renggli, C. C. Ahrens, C. D. Cook, C. Kroll, A. T. Krueger, B. Imperiali and L. G. Griffith, *Biomacromolecules*, 2015, **16**, 2316-2326.
167. M. R. Arkenberg and C. C. Lin, *Biomater. Sci.*, 2017, **5**, 2231-2240.
168. E. Gau, D. M. Mate, Z. Zou, A. Oppermann, A. Töpel, F. Jakob, D. Wöll, U. Schwaneberg and A. Pich, *Biomacromolecules*, 2017, **18**, 2789-2798.
169. T. Matsumoto, Y. Isogawa, T. Tanaka and A. Kondo, *Biosens. Bioelectron.*, 2018, **99**, 56-61.
170. H. Hirakawa, S. Ishikawa and T. Nagamune, *Biotechnol. J.*, 2015, **10**, 1487-1492.
171. R. R. Beerli, T. Hell, A. S. Merkel and U. Grawunder, *PLoS One*, 2015, **10**, e0131177.

172. N. Stefan, R. Gebleux, L. Waldmeier, T. Hell, M. Escher, F. I. Wolter, U. Grawunder and R. R. Beerli, *Mol. Cancer Ther.*, 2017, **16**, 879-892.
173. A. J. McCluskey and R. J. Collier, *Mol. Cancer Ther.*, 2013, **12**, 2273-2281.
174. C. F. Greineder, C. H. Villa, L. R. Walsh, R. Y. Kiseleva, E. D. Hood, M. Khoshnejad, R. Warden-Rothman, A. Tsourkas and V. R. Muzykantov, *Bioconjugate Chem.*, 2018, **29**, 56-66.
175. M. Rashidian, E. J. Keliher, A. M. Bilate, J. N. Duarte, G. R. Wojtkiewicz, J. T. Jacobsen, J. Cragnolini, L. K. Swee, G. D. Victora, R. Weissleder and H. L. Ploegh, *Proc. Natl. Acad. Sci. U S A*, 2015, **112**, 6146-6151.
176. T. Fang, J. N. Duarte, J. Ling, Z. Li, J. S. Guzman and H. L. Ploegh, *Angew. Chem. Int. Ed.*, 2016, **55**, 2416-2420.
177. S. Massa, N. Vikani, C. Betti, S. Ballet, S. Vanderhaegen, J. Steyaert, B. Descamps, C. Vanhove, A. Bunschoten, F. W. van Leeuwen, S. Hernot, V. Caveliers, T. Lahoutte, S. Muyldermans, C. Xavier and N. Devoogdt, *Contrast Media Mol. Imaging*, 2016, **11**, 328-339.
178. Y. Pang, J. Liu, Y. Qi, X. Li and A. Chilkoti, *Angew. Chem. Int. Ed.*, 2016, **55**, 10296-10300.
179. R. Parthasarathy, S. Subramanian and E. T. Boder, *Bioconjugate Chem.*, 2007, **18**, 469-476.
180. S. Samantaray, U. Marathe, S. Dasgupta, V. K. Nandicoori and R. P. Roy, *J. Am. Chem. Soc.*, 2008, **130**, 2132-2133.
181. S. Mohlmann, C. Mahlert, S. Greven, P. Scholz and A. Harrenga, *Chembiochem*, 2011, **12**, 1774-1780.
182. S. Dasgupta, S. Samantaray, D. Sahal and R. P. Roy, *J. Biol. Chem.*, 2011, **286**, 23996-24006.
183. J. J. Bellucci, J. Bhattacharyya and A. Chilkoti, *Angew. Chem. Int. Ed.*, 2015, **54**, 441-445.
184. T. Ito, R. Sadamoto, K. Naruchi, H. Togame, H. Takemoto, H. Kondo and S. Nishimura, *Biochemistry*, 2010, **49**, 2604-2614.
185. S. Baer, J. Nigro, M. P. Madej, R. M. Nisbet, R. Suryadinata, G. Coia, L. P. Hong, T. E. Adams, C. C. Williams and S. D. Nuttall, *Org. Biomol. Chem.*, 2014, **12**, 2675-2685.
186. Y. M. Li, Y. T. Li, M. Pan, X. Q. Kong, Y. C. Huang, Z. Y. Hong and L. Liu, *Angew. Chem. Int. Ed.*, 2014, **53**, 2198-2202.
187. J. J. Bellucci, M. Amiram, J. Bhattacharyya, D. McCafferty and A. Chilkoti, *Angew. Chem. Int. Ed.* , 2013, **52**, 3703-3708.
188. R. L. Policarpo, H. Kang, X. Liao, A. E. Rabideau, M. D. Simon and B. L. Pentelute, *Angew. Chem. Int. Ed.*, 2014, **53**, 9203-9208.
189. R. Warden-Rothman, I. Caturegli, V. Popik and A. Tsourkas, *Anal. Chem.*, 2013, **85**, 11090-11097.
190. H. H. Wang, B. Altun, K. Nwe and A. Tsourkas, *Angew. Chem. Int. Ed.*, 2017, **56**, 5349-5352.
191. X. Cheng, T. Zhu, H. Hong, Z. Zhou and Z. Wu, *Org. Chem. Front.*, 2017, **4**, 2058-2062.
192. T. Heck, P. H. Pham, A. Yerlikaya, L. Thöny-Meyer and M. Richter, *Catal. Sci. Technol.*, 2014, **4**, 2946-2956.

193. J. A. Wells and D. A. Estell, *Trends Biochem. Sci.*, 1988, **13**, 291-297.
194. T. K. Chang, D. Y. Jackson, J. P. Burnier and J. A. Wells, *Proc. Natl. Acad. Sci. U S A*, 1994, **91**, 12544-12548.
195. D. Y. Jackson, J. Burnier, C. Quan, M. Stanley, J. Tom and J. A. Wells, *Science*, 1994, **266**, 243-247.
196. S. Mahrus, J. C. Trinidad, D. T. Barkan, A. Sali, A. L. Burlingame and J. A. Wells, *Cell*, 2008, **134**, 866-876.
197. C. H. Yi, H. Pan, J. Seebacher, I. H. Jang, S. G. Hyberts, G. J. Heffron, M. G. Vander Heiden, R. Yang, F. Li, J. W. Locasale, H. Sharfi, B. Zhai, R. Rodriguez-Mias, H. Luithardt, L. C. Cantley, G. Q. Daley, J. M. Asara, S. P. Gygi, G. Wagner, C. F. Liu and J. Yuan, *Cell*, 2011, **146**, 607-620.
198. L. Mariniello, R. Porta, A. Sorrentino, C. V. Giosafatto, G. Rossi Marquez, M. Esposito and P. Di Pierro, *Amino Acids*, 2014, **46**, 767-776.
199. B. Spolaore, S. Raboni, A. Ramos Molina, A. Satwekar, N. Damiano and A. Fontana, *Biochemistry*, 2012, **51**, 8679-8689.
200. N. M. Rachel and J. N. Pelletier, *Biomolecules*, 2013, **3**, 870-888.
201. M. Malesevic, A. Migge, T. C. Hertel and M. Pietzsch, *Chembiochem*, 2015, **16**, 1169-1174.
202. M. T. Gundersen, J. W. Keillor and J. N. Pelletier, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 219-230.
203. Y. Sugimura, K. Yokoyama, N. Nio, M. Maki and K. Hitomi, *Arch. Biochem. Biophys.*, 2008, **477**, 379-383.
204. J. H. Lee, C. Song, D. H. Kim, I. H. Park, S. G. Lee, Y. S. Lee and B. G. Kim, *Biotechnol. Bioeng.*, 2013, **110**, 353-362.
205. J. Tominaga, Y. Kemori, Y. Tanaka, T. Maruyama, N. Kamiya and M. Goto, *Chem. Commun.*, 2007, 401-403.
206. M. Taki, M. Shiota and K. Taira, *Protein Eng., Des. Sel.*, 2004, **17**, 119-126.
207. Y. Mori, M. Goto and N. Kamiya, *Biochem. Biophys. Res. Commun.*, 2011, **410**, 829-833.
208. N. Kamiya, H. Abe, M. Goto, Y. Tsuji and H. Jikuya, *Org. Biomol. Chem.*, 2009, **7**, 3407-3412.
209. A. Fontana, B. Spolaore, A. Mero and F. M. Veronese, *Adv. Drug Deliv. Rev.*, 2008, **60**, 13-28.
210. S. K. Oteng-Pabi, C. Pardin, M. Stoica and J. W. Keillor, *Chem. Commun.*, 2014, **50**, 6604-6606.
211. P. Dennler, A. Chiotellis, E. Fischer, D. Bregeon, C. Belmant, L. Gauthier, F. Lhospice, F. Romagne and R. Schibli, *Bioconjugate Chem.*, 2014, **25**, 569-578.
212. T. L. Mindt, V. Jungi, S. Wyss, A. Friedli, G. Pla, I. Novak-Hofer, J. Grünberg and R. Schibli, *Bioconjugate Chem.*, 2008, **19**, 271-278.
213. P. Strop, S. H. Liu, M. Dorywalska, K. Delaria, R. G. Dushin, T. T. Tran, W. H. Ho, S. Farias, M. G. Casas, Y. Abdiche, D. Zhou, R. Chandrasekaran, C. Samain, C. Loo, A. Rossi, M. Rickert, S. Krimm, T. Wong, S. M. Chin, J. Yu, J. Dilley, J. Chaparro-Riggers, G. F. Filzen, C. J. O'Donnell, F. Wang, J. S. Myers, J. Pons, D. L. Shelton and A. Rajpal, *Chem. Biol.*, 2013, **20**, 161-167.
214. P. Strop, T. T. Tran, M. Dorywalska, K. Delaria, R. Dushin, O. K. Wong, W. H. Ho, D. Zhou, A. Wu, E. Kraynov, L. Aschenbrenner, B. Han, C. J. O'Donnell, J.

Pons, A. Rajpal, D. L. Shelton and S. H. Liu, *Mol. Cancer Ther.*, 2016, **15**, 2698-2708.
215. P. R. Spycher, C. A. Amann, J. E. Wehrmüller, D. R. Hurwitz, O. Kreis, D. Messmer, A. Ritler, A. Küchler, A. Blanc, M. Béhé, P. Walde and R. Schibli, *Chembiochem*, 2017, **18**, 1923-1927.
216. J. L. Spidel, B. Vaessen, E. F. Albone, X. Cheng, A. Verdi and J. B. Kline, *Bioconjugate Chem.*, 2017, **28**, 2471-2484.
217. S. Puthenveetil, S. Musto, F. Loganzo, L. N. Tumey, C. J. O'Donnell and E. Graziani, *Bioconjugate Chem.*, 2016, **27**, 1030-1039.
218. K. Maso, A. Grigoletto and G. Pasut, *Adv. Protein Chem. Struct. Biol.*, 2018, **112**, 123-142.
219. A. Mero, B. Spolaore, F. M. Veronese and A. Fontana, *Bioconjugate Chem.*, 2009, **20**, 384-389.
220. H. Sato, E. Hayashi, N. Yamada, M. Yatagai and Y. Takahara, *Bioconjugate Chem.*, 2001, **12**, 701-710.
221. S. Scaramuzza, G. Tonon, A. Olianas, I. Messana, R. Schrepfer, G. Orsini and P. Caliceti, *J. Controlled Release*, 2012, **164**, 355-363.
222. B. Spolaore, S. Raboni, A. A. Satwekar, A. Grigoletto, A. Mero, I. M. Montagner, A. Rosato, G. Pasut and A. Fontana, *Bioconjugate Chem.*, 2016, **27**, 2695-2706.
223. A. Mero, M. Schiavon, F. M. Veronese and G. Pasut, *J. Controlled Release*, 2011, **154**, 27-34.
224. D. da Silva Freitas, A. Mero and G. Pasut, *Bioconjugate Chem.*, 2013, **24**, 456-463.
225. M. Kitaoka, Y. Tsuruda, Y. Tanaka, M. Goto, M. Mitsumori, K. Hayashi, Y. Hiraishi, K. Miyawaki, S. Noji and N. Kamiya, *Chem. Eur. J.*, 2011, **17**, 5387-5392.
226. M. Kitaoka, M. Mitsumori, K. Hayashi, Y. Hiraishi, H. Yoshinaga, K. Nakano, K. Miyawaki, S. Noji, M. Goto and N. Kamiya, *Anal. Chem.*, 2012, **84**, 5885-5891.
227. M. Takahara, R. Wakabayashi, K. Minamihata, M. Goto and N. Kamiya, *Bioconjugate Chem.*, 2017, **28**, 2954-2961.
228. R. Wakabayashi, K. Yahiro, K. Hayashi, M. Goto and N. Kamiya, *Biomacromolecules*, 2017, **18**, 422-430.
229. H. Jiang, X. Zhang, X. Chen, P. Aramsangtienchai, Z. Tong and H. Lin, *Chem. Rev.*, 2018, **118**, 919-988.
230. H. L. Hartman, K. A. Hicks and C. A. Fierke, *Biochemistry*, 2005, **44**, 15314-15324.
231. A. J. Krzysiak, S. A. Scott, K. A. Hicks, C. A. Fierke and R. A. Gibbs, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 5548-5551.
232. J. L. Hougland, C. L. Lamphear, S. A. Scott, R. A. Gibbs and C. A. Fierke, *Biochemistry*, 2009, **48**, 1691-1701.
233. Y. C. Wang, J. K. Dozier, L. S. Beese and M. D. Distefano, *ACS Chem Biol*, 2014, **9**, 1726-1735.
234. S. A. Reigard, T. J. Zahn, K. B. Haworth, K. A. Hicks, C. A. Fierke and R. A. Gibbs, *Biochemistry*, 2005, **44**, 11214-11223.
235. G. R. Labadie, R. Viswanathan and C. D. Poulter, *J. Org. Chem.*, 2007, **72**, 9291-9297.
236. M. J. Blanden, K. F. Suazo, E. R. Hildebrandt, D. S. Hardgrove, M. Patel, W. P. Saunders, M. D. Distefano, W. K. Schmidt and J. L. Hougland, *J. Biol. Chem.*, 2018, **293**, 2770-2785.

237. D. J. Owen, K. Alexandrov, E. Rostkova, A. J. Scheidig, R. S. Goody and H. Waldmann, *Angew. Chem. Int. Ed.*, 1999, **38**, 509-512.
238. A. P. C. Chen, Y. H. Chen, H. P. Liu, Y. C. Li, C. T. Chen and P. H. Liang, *J. Am. Chem. Soc.*, 2002, **124**, 15217-15224.
239. B. Dursina, R. Reents, C. Delon, Y. Wu, M. Kulharia, M. Thutewohl, A. Veligodsky, A. Kalinin, V. Evstifeev, D. Ciobanu, S. Szedlacsek, H. Waldmann, R. S. Goody and K. Alexandrov, *J Am Chem Soc*, 2006, **128**, 2822-2835.
240. T. C. Turek, I. Gaon and M. D. Distefano, *J Org Chem*, 2001, **66**, 3253-3264.
241. K. A. H. Chehade, K. Kiegiel, R. J. Isaacs, J. S. Pickett, K. E. Bowers, C. A. Fierke, D. A. Andres and H. P. Spielmann, *J. Am. Chem. Soc.*, 2002, **124**, 8206-8219.
242. T. C. Turek-Etienne, C. L. Strickland and M. D. Distefano, *Biochemistry*, 2003, **42**, 3716-3724.
243. U. T. T. Nguyen, Z. Guo, C. Delon, F. Wu, C. Deraeve, B. Fränzel, R. S. Robin, W. Blankenfeldt, R. S. Goody, H. Waldmann, D. Wolters and K. Alexandrov, *Nat Chem Biol*, 2009, **5**, 227-235.
244. 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.
245. M. W. Rose, N. D. Rose, J. Boggs, S. Lenevich, J. Xu, G. Barany and M. D. Distefano, *J. Pept. Res.*, 2005, **65**, 529-537.
246. B. P. Duckworth, J. Xu, T. A. Taton, A. Guo and M. D. Distefano, *Bioconjugate Chem.*, 2006, **17**, 967-974.
247. 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.
248. 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.
249. D. Das, Z. Tnimov, U. T. Nguyen, G. Thimmaiah, H. Lo, D. Abankwa, Y. Wu, R. S. Goody, H. Waldmann and K. Alexandrov, *Chembiochem*, 2012, **13**, 674-683.
250. M. Rashidian, J. K. Dozier, S. Lenevich and M. D. Distefano, *Chem. Commun.*, 2010, **46**, 8998-9000.
251. M. Kim, T. S. Kleckley, A. J. Wiemer, S. A. Holstein, R. J. Hohl and D. F. Wiemer, *J. Org. Chem.*, 2004, **69**, 8186-8193.
252. U. T. Nguyen, Z. Guo, C. Delon, Y. Wu, C. Deraeve, B. Fränzel, R. S. Bon, W. Blankenfeldt, R. S. Goody, H. Waldmann, D. Wolters and K. Alexandrov, *Nat. Chem. Biol.*, 2009, **5**, 227-235.
253. A. J. DeGraw, C. Palsuledesai, J. D. Ochocki, J. K. Dozier, S. Lenevich, M. Rashidian and M. D. Distefano, *Chem. Biol. Drug. Des.*, 2010, **76**, 460-471.
254. D. Weinrich, P. C. Lin, P. Jonkheijm, U. T. Nguyen, H. Schroder, C. M. Niemeyer, K. Alexandrov, R. Goody and H. Waldmann, *Angew. Chem. Int. Ed.*, 2010, **49**, 1252-1257.
255. S. L. Khatwani, J. S. Kang, D. G. Mullen, M. A. Hast, L. S. Beese, M. D. Distefano and T. A. Taton, *Bioorg. Med. Chem.*, 2012, **20**, 4532-4539.
256. J. K. Dozier, S. L. Khatwani, J. W. Wollack, Y. C. Wang, C. Schmidt-Dannert and M. D. Distefano, *Bioconjugate Chem.*, 2014, **25**, 1203-1212.
257. J. S. Pickett, K. E. Bowers, H. L. Hartman, H. W. Fu, A. C. Embry, P. J. Casey and C. A. Fierke, *Biochemistry*, 2003, **42**, 9741-9748.

258. R. Viswanathan, G. R. Labadie and C. D. Poulter, *Bioconjugate Chem.*, 2013, **24**, 571-577.
259. J. S. Seo and C. D. Poulter, *Langmuir*, 2014, **30**, 6629-6635.
260. S. R. Choi, J. S. Seo, R. F. Bohaty and C. D. Poulter, *Bioconjugate Chem.*, 2014, **25**, 269-275.
261. Z. P. Tolstyka, W. Richardson, E. Bat, C. J. Stevens, D. P. Parra, J. K. Dozier, M. D. Distefano, B. Dunn and H. D. Maynard, *Chembiochem*, 2013, **14**, 2464-2471.
262. Y. Kim, T. Park, S. Woo, H. Lee, S. Kim, J. Cho, D. Jung, Y. Kim, H. Kwon and K. Oh*, U. S. Pat.,* 13/466,875, 2012.
263. C. Giglione, S. Fieulaine and T. Meinnel, *Biochimie*, 2015, **114**, 134-146.
264. D. D. Martin, E. Beauchamp and L. G. Berthiaume, *Biochimie*, 2011, **93**, 18-31.
265. D. A. Towler, S. P. Adams, S. R. Eubanks, D. S. Towery, E. Jackson-Machelski, L. Glaser and J. I. Gordon, *Proc. Natl. Acad. Sci. U S A*, 1987, **84**, 2708-2712.
266. B. Devadas, T. Lu, A. Katoh, N. S. Kishore, A. C. Wade, P. P. Mehta, D. A. Rudnick, M. L. Bryant, S. P. Adams, Q. Li, G. W. Gokel and J. I. Gordon, *J. Biol. Chem.*, 1992, **267**, 7224-7239.
267. T. Lu, Q. Li, A. Katoh, J. Hernandez, K. Duffin, E. Jackson-Machelski, L. J. Knoll, G. W. Gokel and J. I. Gordon, *J. Biol. Chem.*, 1994, **269**, 5346-5357.
268. 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.
269. W. P. Heal, M. H. Wright, E. Thinon and E. W. Tate, *Nat. Protoc.*, 2011, **7**, 105-117.
270. K. M. Luginbuhl, D. Mozhdehi, M. Dzuricky, P. Yousefpour, F. C. Huang, N. R. Mayne, K. L. Buehne and A. Chilkoti, *Angew. Chem. Int. Ed.*, 2017, **56**, 13979-13984.
271. J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel and M. D. Burkart, *Nat. Prod. Rep.*, 2014, **31**, 61-108.
272. J. Yin, F. Liu, X. Li and C. T. Walsh, *J. Am. Chem. Soc.*, 2004, **126**, 7754-7755.
273. N. M. Kosa, R. W. Haushalter, A. R. Smith and M. D. Burkart, *Nat. Methods*, 2012, **9**, 981.
274. M. Rothmann, N. M. Kosa and M. D. Burkart, *RSC adv.*, 2014, **4**, 9092-9097.
275. 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.
276. Z. Zhou, A. Koglin, Y. Wang, A. P. McMahon and C. T. Walsh, *J. Am. Chem. Soc.*, 2008, **130**, 9925-9930.
277. M. Strackharn, D. A. Pippig, P. Meyer, S. W. Stahl and H. E. Gaub, *J. Am. Chem. Soc.*, 2012, **134**, 15193-15196.
278. T. Kawamura, B. Stephens, L. Qin, X. Yin, M. R. Dores, T. H. Smith, N. Grimsey, R. Abagyan, J. Trejo, I. Kufareva, M. M. Fuster, C. L. Salanga and T. M. Handel, *PLoS ONE*, 2014, **9**, e81454.
279. S. J. Allen, D. J. Hamel and T. M. Handel, *Cytokine*, 2011, **55**, 168-173.
280. L. Marchetti, T. De Nadai, F. Bonsignore, M. Calvello, G. Signore, A. Viegi, F. Beltram, S. Luin and A. Cattaneo, *PLoS ONE*, 2014, **9**, e113708.
281. L. Schmohl and D. Schwarzer, *J. Pept. Sci.*, 2014, **20**, 145-151.

282. J. Grünewald, H. Jones David, A. Brock, H. P. Chiu, B. Bursulaya, K. Ng, T. Vo, P. Patterson, T. Uno, J. Hunt, G. Spraggon and B. H. Geierstanger, *ChemBioChem*, 2014, **15**, 1787-1791.
283. W. Ou, T. Uno, H.-P. Chiu, J. Grünewald, S. E. Cellitti, T. Crossgrove, X. Hao, Q. Fan, L. L. Quinn, P. Patterson, L. Okach, D. H. Jones, S. A. Lesley, A. Brock and B. H. Geierstanger, *Proc. Natl. Acad. Sci. U S A*, 2011, **108**, 10437.
284. K. M. Clarke, A. C. Mercer, J. J. La Clair and M. D. Burkart, *J. Am. Chem. Soc.*, 2005, **127**, 11234-11235.
285. A. S. Worthington and M. D. Burkart, *Org. Biomol. Chem.*, 2006, **4**, 44-46.
286. I. Nazi, K. P. Koteva and G. D. Wright, *Anal. Biochem.*, 2004, **324**, 100-105.
287. T. Kittilä and M. J. Cryle, *Biochem. Cell Biol.*, 2017, **96**, 372-379.
288. K. Ersfeld, J. Wehland, U. Plessmann, H. Dodemont, V. Gerke and K. Weber, *J. Cell Biol.*, 1993, **120**, 725.
289. C. Janke, *J. Cell Biol.*, 2014, **206**, 461.
290. C. Kato, K. Miyazaki, A. Nakagawa, M. Ohira, Y. Nakamura, T. Ozaki, T. Imai and A. Nakagawara, *Int. J. Cancer*, 2004, **112**, 365-375.
291. C. Erck, L. Peris, A. Andrieux, C. Meissirel, A. D. Gruber, M. Vernet, A. Schweitzer, Y. Saoudi, H. Pointu, C. Bosc, P. A. Salin, D. Job and J. Wehland, *Proc. Natl. Acad. Sci. U S A*, 2005, **102**, 7853.
292. O. Monasterio, E. Nova, A. López-Brauet and R. Lagos, *FEBS Letters*, 1995, **374**, 165-168.
293. C. G. Bisig, A. Purro Silvia, A. Contín María, S. Barra Héctor and A. Arce Carlos, *Eur. J. Biochem.*, 2002, **269**, 5037-5045.
294. A. Banerjee, T. D. Panosian, K. Mukherjee, R. Ravindra, S. Gal, D. L. Sackett and S. Bane, *ACS Chem. Biol.*, 2010, **5**, 777-785.
295. L. J. Reed, F. R. Leach and M. Koike, *J. Biol. Chem.*, 1958, **232**, 123-142.
296. D. E. Brookfield, J. Green, S. T. Ali, R. S. Machado and J. R. Guest, *FEBS Lett.*, 1991, **295**, 13-16.
297. T. W. Morris, K. E. Reed and J. E. Cronan, Jr., *J. Biol. Chem.*, 1994, **269**, 16091-16100.
298. S. Puthenveetil, D. S. Liu, K. A. White, S. Thompson and A. Y. Ting, *J. Am. Chem. Soc.*, 2009, **131**, 16430-16438.
299. C. R. Drake, N. Sevillano, C. Truillet, C. S. Craik, H. F. VanBrocklin and M. J. Evans, *ACS Chem. Biol.*, 2016, **11**, 1587-1594.
300. W. Liu, S. K. Samanta, B. D. Smith and L. Isaacs, *Chem. Soc. Rev.*, 2017, **46**, 2391-2403.
301. M. Fernández-Suárez, T. S. Chen and A. Y. Ting, *J. Am. Chem. Soc.*, 2008, **130**, 9251-9253.
302. K. J. Roux, D. I. Kim, M. Raida and B. Burke, *J. Cell Biol.*, 2012, **196**, 801.
303. D. I. Kim, S. C. Jensen, K. A. Noble, B. Kc, K. H. Roux, K. Motamedchaboki, K. J. Roux and Y. Zheng, *Mol. Biol. Cell*, 2016, **27**, 1188-1196.
304. T. Dierks, B. Schmidt, L. V. Borissenko, J. Peng, A. Preusser, M. Mariappan and K. von Figura, *Cell*, 2003, **113**, 435-444.
305. J. S. Rush and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2008, **130**, 12240-12241.
306. J. E. Hudak, H. H. Yu and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2011, **133**, 16127-16135.

307. E. L. Smith, J. P. Giddens, A. T. Iavarone, K. Godula, L. X. Wang and C. R. Bertozzi, *Bioconjugate Chem.*, 2014, **25**, 788-795.
308. P. M. Drake, A. E. Albers, J. Baker, S. Banas, R. M. Barfield, A. S. Bhat, G. W. de Hart, A. W. Garofalo, P. Holder, L. C. Jones, R. Kudirka, J. McFarland, W. Zmolek and D. Rabuka, *Bioconjugate Chem.*, 2014, **25**, 1331-1341.
309. T. L. Foley and M. D. Burkart, *Curr. Opin. Chem. Biol.*, 2007, **11**, 12-19.
310. I. Chen and A. Y. Ting, *Curr. Opin. Biotechnol.*, 2005, **16**, 35-40.
311. E. Hudak Jason, M. Barfield Robyn, W. de Hart Gregory, P. Grob, E. Nogales, R. Bertozzi Carolyn and D. Rabuka, *Angew. Chem. Int. Ed.*, 2012, **51**, 4161-4165.
312. B. C. Bundy and J. R. Swartz, *Bioconjugate Chem.*, 2010, **21**, 255-263.
313. R. Haag and F. Kratz, *Angew. Chem. Int. Ed.*, 2006, **45**, 1198-1215.
314. M. A. Gauthier and H.-A. Klok, *Chem. Commun.*, 2008, DOI: 10.1039/B719689J, 2591-2611.
315. S. Jevševar, M. Kunstelj and G. Porekar Vladka, *Biotechnol. J.*, 2010, **5**, 113-128.
316. N. G. Yabbarov, G. A. Posypanova, E. A. Vorontsov, S. I. Obydenny and E. S. Severin, *J. Controlled Release*, 2013, **168**, 135-141.
317. E. L. Sievers and P. D. Senter, *Annu. Rev. Med.*, 2013, **64**, 15-29.
318. A. Keppler, H. Pick, C. Arrivoli, H. Vogel and K. Johnsson, *Proc. Natl. Acad. Sci. U S A*, 2004, **101**, 9955.
319. A. M. Wu and P. D. Senter, *Nat. Biotechnol.*, 2005, **23**, 1137.
320. E. Steen Redeker, D. T. Ta, D. Cortens, B. Billen, W. Guedens and P. Adriaensens, *Bioconjugate Chem.*, 2013, **24**, 1761-1777.
321. H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein and M. Snyder, *Science*, 2001, **293**, 2101.
322. D. Rabuka, *Curr. Opin. Chem. Biol.*, 2010, **14**, 790-796.
323. M. Sunbul and J. Yin, *Org. Biomol. Chem.*, 2009, **7**, 3361-3371.
324. B. P. Duckworth, Z. Zhang, A. Hosokawa and M. D. Distefano, *ChemBioChem*, 2007, **8**, 98-105.
325. T. Scott Reid, K. L. Terry, P. J. Casey and L. S. Beese, *J. Mol. Biol.*, 2004, **343**, 417-433.
326. H. L. Hartman, K. A. Hicks and C. A. Fierke, *Biochemistry*, 2005, **44**.
327. J. S. Taylor, T. Reid, K. L. Terry, P. J. Casey and L. S. Beese, *EMBO J.*, 2003, **22**, 5963-5974.
328. M. Rashidian, S. C. Kumarapperuma, K. Gabrielse, A. Fegan, C. R. Wagner and M. D. Distefano, *J. Am. Chem. Soc.*, 2013, **135**, 16388-16396.
329. F. H. Berry Alexandra, P. Heal William, K. Tarafder Abul, T. Tolmachova, A. Baron Rudi, C. Seabra Miguel and W. Tate Edward, *ChemBioChem*, 2010, **11**, 771-773.
330. N. C. Shaner, P. A. Steinbach and R. Y. Tsien, *Nat. Methods*, 2005, **2**, 905.
331. A. Gautier, A. Juillerat, C. Heinis, I. R. Corrêa, M. Kindermann, F. Beaufils and K. Johnsson, *Chem. Biol.*, 2008, **15**, 128-136.
332. K. E. Beatty and D. A. Tirrell, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 5995-5999.
333. A. Niederwieser, A. K. Späte, D. Nguyen Long, C. Jüngst, W. Reutter and V. Wittmann, *Angew. Chem. Int. Ed.*, 2013, **52**, 4265-4268.

334. P. V. Chang, J. A. Prescher, M. J. Hangauer and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2007, **129**, 8400-8401.
335. M. M. Mahmoodi, M. Rashidian, Y. Zhang and M. D. Distefano, *Curr. Protoc. Protein Sci.*, 2015, **79**, 15.14.11-15.14.28.
336. J. K. Dozier and M. D. Distefano, *Anal. Biochem.*, 2012, **421**, 158-163.
337. H. L. Hartman, K. E. Bowers and C. A. Fierke, *J. Biol. Chem.*, 2004, **279**, 30546-30553.
338. F. W. Studier, *Protein Expression Purif.*, 2005, **41**, 207-234.
339. S. A. Gangopadhyay, E. L. Losito and J. L. Hougland, *Biochemistry*, 2014, **53**, 434-446.
340. W. Wollack James, M. Silverman Julie, J. Petzold Christopher, D. Mougous Joseph and D. Distefano Mark, *ChemBioChem*, 2009, **10**, 2934-2943.
341. G. Cui and K. M. Merz, *Biochemistry*, 2007, **46**, 12375-12381.
342. D. K. Chakravorty and K. M. Merz, *Acc. Chem. Res.*, 2015, **48**, 439-448.
343. Y. C. Wang, J. K. Dozier, L. S. Beese and M. D. Distefano, *ACS Chem. Biol.*, 2014, **9**, 1726-1735.
344. E. M. Merzlyak, J. Goedhart, D. Shcherbo, M. E. Bulina, A. S. Shcheglov, A. F. Fradkov, A. Gaintzeva, K. A. Lukyanov, S. Lukyanov, T. W. J. Gadella and D. M. Chudakov, *Nat. Methods*, 2007, **4**, 555.
345. A. Sachdeva, K. Wang, T. Elliott and J. W. Chin, *J. Am. Chem. Soc.*, 2014, **136**, 7785-7788.
346. I. Willems Lianne, N. Li, I. Florea Bogdan, M. Ruben, A. van der Marel Gijsbert and S. Overkleeft Herman, *Angew. Chem. Int. Ed.*, 2012, **51**, 4431-4434.
347. B. Wu, Z. Wang, Y. Huang and W. R. Liu, *Chembiochem*, 2012, **13**, 1405-1408.
348. R. Karver Mark, R. Weissleder and A. Hilderbrand Scott, *Angew. Chem. Int. Ed.*, 2011, **51**, 920-922.
349. Y. Reiss, M. S. Brown and J. L. Goldstein, *J. Biol. Chem.*, 1992, **267**, 6403-6408.
350. S. Kaur, G. Venktaraman, M. Jain, S. Senapati, P. K. Garg and S. K. Batra, *Cancer Lett.*, 2012, **315**, 97-111.
351. C. C. Palsuledesai and M. D. Distefano, *ACS Chem. Biol.*, 2015, **10**, 51-62.
352. R. Tamaskovic, M. Simon, N. Stefan, M. Schwill and A. Pluckthun, *Methods Enzymol.*, 2012, **503**, 101-134.
353. N. Stefan, P. Martin-Killias, S. Wyss-Stoeckle, A. Honegger, U. Zangemeister-Wittke and A. Pluckthun, *J. Mol. Biol.*, 2011, **413**, 826-843.
354. C. Patriarca, R. M. Macchi, A. K. Marschner and H. Mellstedt, *Cancer Treat. Rev.*, 2012, **38**, 68-75.
355. K. L. Bryant, J. D. Mancias, A. C. Kimmelman and C. J. Der, *Trends Biochem. Sci.*, 2014, **39**, 91-100.
356. J. M. Rae, J. O. Scheys, K. M. Clark, R. B. Chadwick, M. C. Kiefer and M. E. Lippman, *Breast Cancer Res. Treat.*, 2004, **87**, 87-95.
357. B. Li, M. Sedlacek, I. Manoharan, R. Boopathy, E. G. Duysen, P. Masson and O. Lockridge, *Biochem. Pharmacol.*, 2005, **70**, 1673-1684.
358. T. Satoh and M. Hosokawa, *Chem.-Biol. Interact.*, 2006, **162**, 195-211.
359. F. G. Bahar, K. Ohura, T. Ogihara and T. Imai, *J. Pharm. Sci.*, 2012, **101**, 3979-3988.

360. S. O. Doronina, B. E. Toki, M. Y. Torgov, B. A. Mendelsohn, C. G. Cerveny, D. F. Chace, R. L. DeBlanc, R. P. Gearing, T. D. Bovee, C. B. Siegall, J. A. Francisco, A. F. Wahl, D. L. Meyer and P. D. Senter, *Nat. Biotechnol.*, 2003, **21**, 778.
361. P.-Å. Nygren and A. Skerra, *J. Immunol. Methods*, 2004, **290**, 3-28.
362. R. Vazquez-Lombardi, T. G. Phan, C. Zimmermann, D. Lowe, L. Jermutus and D. Christ, *Drug Discovery Today*, 2015, **20**, 1271-1283.
363. D. Lipovšek, *Protein Eng., Des. Sel.*, 2011, **24**, 3-9.
364. S. Muyldermans, *Annu. Rev. Biochem.*, 2013, **82**, 775-797.
365. B. J. Hackel, M. E. Ackerman, S. W. Howland and K. D. Wittrup, *J. Mol. Biol.*, 2010, **401**, 84-96.
366. L. A. Stern, C. M. Csizmar, D. R. Woldring, C. R. Wagner and B. J. Hackel, *ACS Comb. Sci.*, 2017, **19**, 315-323.
367. E. Pardon, T. Laeremans, S. Triest, S. G. F. Rasmussen, A. Wohlkönig, A. Ruf, S. Muyldermans, W. G. J. Hol, B. K. Kobilka and J. Steyaert, *Nat. Protoc.*, 2014, **9**, 674.
368. B. J. Hackel, A. Kapila and K. Dane Wittrup, *J. Mol. Biol.*, 2008, **381**, 1238-1252.
369. M. H. Parker, Y. Chen, F. Danehy, K. Dufu, J. Ekstrom, E. Getmanova, J. Gokemeijer, L. Xu and D. Lipovsek, *Protein Eng., Des. Sel.*, 2005, **18**, 435-444.
370. A. Koide, V. Tereshko, S. Uysal, K. Margalef, A. A. Kossiakoff and S. Koide, *J. Mol. Biol.*, 2007, **373**, 941-953.
371. D. Lipovšek, S. M. Lippow, B. J. Hackel, M. W. Gregson, P. Cheng, A. Kapila and K. D. Wittrup, *J. Mol. Biol.*, 2007, **368**, 1024-1041.