# Site-selective Protein Modification by Chemical Design: Towards Multifunctional Bioconjugates

Kumulative Dissertation zur Erlangung des Doktorgrades Doctor rerum naturalium (Dr. rer.nat.) der Fakultät für Naturwissenschaften der Universität Ulm

vorgelegt von

## *Lujuan Xu*

aus Henan, China

2022

**Dissertation der Universität Ulm**

Amtierender Dekan: Prof. Dr. Thorsten Bernhardt

    1. Gutachter: Prof. Dr. Tanja Weil

    2. Gutachter: Prof. Dr. Carsten Streb

Tag der Promotion: December 13, 2021

Universität Ulm, Fakultät für Naturwissenschaften

# Site-selective Protein Modification by Chemical Design: Towards Multifunctional Bioconjugates

Cumulative Dissertation to obtain the degree Doctor rerum naturalium (Dr. rer.nat.) of the Faculty of Natural Sciences of the University of Ulm

submitted by

## *Lujuan Xu*

from Henan, China

2022

**<u>Dissertation of Ulm University</u>**

Dean of the Faculty: Prof. Dr. Thorsten Bernhardt

1. Reviewer: Prof. Dr. Tanja Weil

2. Reviewer: Prof. Dr. Carsten Streb

Day of the Promotion: December 13, 2021

Faculty of Natural Sciences, Ulm University

# Abstract

Proteins are considered as Nature's most versatile and fascinating biomolecules due to their exceptional functional and structural properties. Molecular precision and defined geometry are the central hallmarks of proteins, which are considered as prerequisites to their biological functions. On the other hand, the flexibility and repertoire of functions that modern synthetic chemistry can provide beyond Nature's machinery are also undeniable. In this context, the integration of the protein with synthetic functionalities could bridge these two worlds so that the advantages of both components could be fully exploited to provide a new perspective for devising new functional materials. Therefore, there is a pressing need to develop efficient methodologies that can introduce the desired synthetic moieties to proteins in a site-selective fashion without interfering their structure and function integrity.

In this regard, this thesis focuses on developing new bioconjugation chemistries for protein modification by chemical design, which have high yields, excellent labeling efficiencies, good operational simplicities, and offer stable bioconjugates. In the first work, an inverse electron demand Diels-Alder reaction (iEEDA) between tetrazine and *trans*-cyclooctene (TCO) was employed for protein modification at the disulfide site in a highly selective manner. The tetrazine functionalized peptide or the antigen-binding fragment (Fab) of IgG antibody were conjugated with a series of TCO-modified functionalities, such as dyes, polymers, or proteins/enzymes offering a small library of stable bioconjugates within a short reaction time in quantitative conversions. This strategy would enable the preparation of the desired bioconjugates "on-site", e.g. in hospitals, which can be applied to the patients immediately after preparation, preventing the loss of the final conjugates after long-term storage, quality control, and batch-to-batch variations.

In order to further improve the modification parameters above, the second work aims at developing new strategies for cysteine and disulfide modification which offers higher modification efficiency and operational simplicity compared to the existing techniques. In this work, a chloromethyl acryl scaffold was developed to serve as a versatile platform for synthesizing acrylamide or acrylate derivatives by coupling with different chemical end-group (amino group or hydroxyl group) via a one-pot reaction. In this way, the chemical properties can be tailored by chemical design so that their inherent reactivity can be tuned to achieve selective protein modification at cysteine or disulfide sites on demand. It was demonstrated that chloromethyl acrylamide reagents favor selective modification at the cysteine site while chloromethyl acrylate reagents undergo two successive Michael reactions to achieve site-selective disulfide modification. Both reagents exhibit high efficiency and the resultant bioconjugates fully preserved their bioactivity and exhibit superior stability compared to those derived from the existing methodologies.

In summary, this thesis aims to develop elegant approaches for site-selective protein modification at disulfide and cysteine sites with high labeling efficiency and specificity. The approaches developed herein overcome drawbacks of the existing methodologies in the literature

and serve as valuable additions to the current bioconjugation toolbox. The bioconjugation techniques developed in this thesis hold immense potential for the design and preparation of precision protein conjugates for sophisticated biological applications, such as bioimaging, biosensing, or theranostics.

# Table of Contents

# 1. Introduction

Proteins are large macromolecules with defined 3D structures and a broad range of bioactivities, which are considered as the most versatile and fascinating biomolecules in Nature. They regulate virtually all biological processes in living systems, such as information transport, energy conversion, inter- and intra-molecular signaling as well as catalyzing biological reactions.[1] Recently, protein therapeutics have gained tremendous attention since the first protein therapeutic - human insulin was approved in 1982.[2] Until now, 130 different proteins or peptides have been approved by the US Food and Drug Administration (FDA) for clinical use and many more are still in development.[3] Even though attractive, unfavorable physicochemical properties often hinder the efficacy of these therapeutic proteins to be fully exerted *in vivo*.[3] For example, these therapeutic proteins usually have short half-lives and can be quickly cleared by the kidneys shortly after the administration. In addition, the unwanted immune reactions originated from the protein itself may also reduce their therapeutic effects and even cause life-threatening complications.[4]

Nature solves this problem by using post-translation modifications (PTMs), where proteins are covalently modified with various groups, such as carbohydrate, phosphate groups, or methyl groups, to tune the physiochemical properties of proteins, modulate protein trafficking, protein-ligand interactions, protein-protein recognitions, and among others.[1] For example, the introduction of carbohydrates to proteins can greatly improve their solubilities in biological environments and modulate their interactions with cell membrane or extracellular matrix. These examples demonstrate how Nature re-design the structure and function of the proteins to diversify their biological functions. Inspired by Nature's elegance, scientists are also strived to modify the proteins with various synthetic functionalities to improve the efficacy, delivery, safety as well as to reduce the immunogenicity.[5] These protein bioconjugates have contributed to the emergence of a new family of protein-based hybrid materials that are applied in various fields including cancer therapy, bioimaging, biosensing, and materials science.[6-8] One of the most classical examples is to conjugate the poly(ethylene glycol) (PEG) to proteins (so-called "PEGylation") to increase their hydrodynamic radius, prevent rapid renal clearance, and thereby increase the serum half-life.[4, 9] In addition, these polymer chains can also protect the protein conjugates from the degradation by proteases and recognition by the immune system.[10]

Furthermore, antibody-based therapeutics have arisen as the largest and fastest-growing class of protein therapeutics due to their highly specific targeting capabilities to the specific antigens.[11-12] It is reported that five of the ten top-selling protein therapeutics in 2009 were antibody-based therapeutics. Owing to their superior specificity, antibodies can be modified with desired synthetic functionalities, such as cytotoxic drugs, for the preparation of various antibody-based protein conjugates to combine the merits from both components.[11] Among these, antibody-drug

conjugates (ADCs) have emerged as a new class of highly potent pharmaceutical drugs, where the cytotoxic drug can be selectively delivered to cancer cells without causing severe side effects to the healthy cells.[13] Central to the success of this concept is the proper modification of the therapeutic antibodies, which enables the desired payloads to be successfully incorporated to facilitate the ADC's mechanism of action.[14]

In addition, protein modification can also provide an opportunity to introduce the PTMs into the proteins so that their exert biological roles can be elucidated in greater detail.[6] Incorporation of fluorescent dyes into proteins allows for the real-time monitor of the trafficking, localization, or interactions of the labeled protein in their native environments.[15] All these examples highlight the significance of proper modifications of proteins and how the properties and applications of proteins can be influenced, modulated, and further expanded by attaching synthetic functionalities, which greatly facilitates numerous developments at the interfaces of chemistry and biology, with applications in cancer therapy, nanomedicine, bioimaging, and advanced materials.

## 1.1 Site-selective protein modification

Since proteins are fragile 3D biomolecules and contain a lot of different functionalities on their surfaces, there are certain requirements for a reaction that can be utilized for protein modification. First, the reactions must be bioorthogonal, which means the reactions need to proceed rapidly and selectively in biological environments without cross-reactivity with endogenous functional groups.[16] In addition, this selectivity should be also realized in biocompatible conditions, for example, aqueous media with as little organic solvent as possible, close to neutral pH and ambient temperature, to prevent protein denaturation.[17] Furthermore, since proteins are usually available in limited quantities and low concentrations, the reactions to modify proteins should be efficient to achieve high conversation in a reasonable timescale.[16] These requirements together present significant challenges to find the suitable chemical reactions to achieve protein modification.

Canonical amino acids (AAs) are the standard 20 AAs encoded directly by the codons of the genetic code. Initial attempts to achieve protein modification mainly rely on the statistical modifications of the reactive canonical AAs residues on the protein surface. The most popular strategy to achieve such modifications is the application of N-hydroxysuccinimidyl (NHS) esters that can react with accessible amino groups of the lysine side chains on the protein surface.[18] Even though simple, given the high abundance of lysine residues on the protein surface, it will lead to random modification and heterogenous product formation, which has been associated with the loss of the structural and functional integrity of proteins. In order to circumvent these limitations, site-selective protein modification has emerged as the most promising strategy as it enables the selective installation of desired functionalities on the protein surface at pre-defined sites with high chemo- and regio-selectivity thus providing easy access to homogeneous protein conjugates

without interfering their structural and functional integrity (Fig. 1).[5-6]

Consequently, site-selective modification of proteins has already become an invaluable tool for the precise analysis of proteins structures and functions in complicated cellular environments and served as a driving force for the innovation of protein therapeutics with improved features.[19] The past few decades witnessed significant progress in the development of new methodologies for site-selective protein modification, ranging from the development of novel chemistry to modify canonical AAs residues to the incorporation of non-canonical AAs (ncAAs, which are AAs that are often synthesized, non-proteinogenic, and can be incorporated into protein sequence in a residue- or site-specific manner.[6, 20] In the subsequent section, well-established bioorthogonal reactions will firstly be introduced, which are considered as the basis for protein modification. Next, contemporary strategies for site-selective protein modification are also briefly reviewed to provide an overview of this burgeoning field.
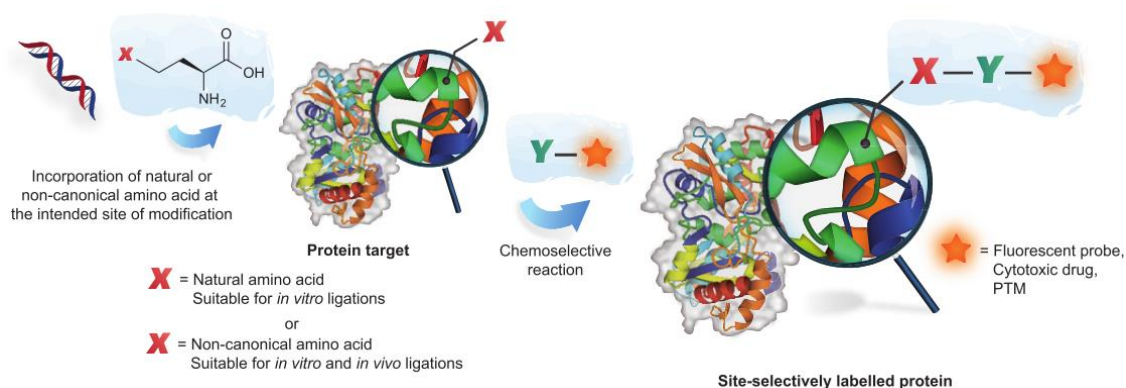


Fig. 1.1 Conceptual strategy to achieve site-selective protein-modification using biocompatible chemical reactions *in vitro* and *in vivo*. Typically, site-selective protein modification can be achieved either by (1) the modification on canonical amino acid residues which have distinct reactivity on the protein surface or (2) the incorporation of non-canonical amino acids, which bear bioorthogonal tags, e.g. alkyne, azide, or tetrazine, allowing subsequent functionalization. For the reaction at natural amino acid residues, the modification is limited to the purified protein in the test tube, as many other biomacromolecules also contain the same functionalities in complex cellular environments. In contrast, the modification of certain ncAAs residues can be achieved in complex mixtures and even in living organisms since the bioorthogonal tags in the ncAAs are not present in living systems. Adapted with permission from reference 3. Copyright (2016), Springer Nature.

## 1.2 Bioorthogonal chemistry

The philosophy of click chemistry was introduced by Sharpless in 2001, which describes the reactions that are modular, simple to perform, broad in scope, high yielding, stereospecific, and result in no or only inert byproducts.[21] The concept of bioorthogonal chemistry was introduced by Bertozzi and coworkers in 2003 and has a significant overlap with click chemistry.[22] In general, bioorthogonal chemistry describes the reactions that can proceed efficiently in a complex biological environment under physiological conditions (ambient temperature, close to neutral pH, aqueous

media) without cross reactions with the plethora of functional groups found in living systems. In addition, they also do not interfere with biological processes.[16, 23-24] However, click chemistry also includes reactions that are not strictly bioorthogonal. For example, biological thiols could interfere with thiol-ene reactions.

In the past decades, bioorthogonal chemistry has emerged as a general and powerful tool to label biomolecules for various applications and triggered a monumental transformation in the field of chemical biology.[25] Until now, a wide variety of bioorthogonal reactions with good biocompatibility and selectivity have been developed.[26] Among these, two types of chemical reactions, the 1,3-dipolar cycloadditions and the Diels–Alder reactions, have gained particular attention. The 1,3-dipolar cycloaddition between terminal alkynes and azides to form 1,2,3-triazoles was firstly reported by Huisgen fifty years ago.[27] However, this reaction requires high temperature or pressure to overcome the reaction barrier to afford products in reasonable yields, which in turn make it not feasible to be utilized in biological systems.[27] A breakthrough was made in 2002 by Sharpless *et. al* and Meldal *et. al* independently when catalytic amounts of Cu(I) salts were found to greatly accelerate the reaction.[28-29] After its discovery, this copper-catalyzed azide-alkyne cycloaddition (CuAAC) has rapidly gained a lot of interest in the field of chemical biology (Fig. 1.2).

However, a primary drawback of CuAAC originated from the inherent toxicity of Cu(I) salts, which promotes the generation of reactive oxygen species (ROS) from $O_2$ causing adverse effects to the structure and function of biomolecules and preventing bioconjugation reactions inside living systems.[30] To circumvent this hurdle, an elegant strategy to improve the biocompatibility of the azide-alkyne cycloaddition was based on the introduction of ring strain to the alkyne derivatives, which is termed as "strain-promoted azide-alkyne cycloaddition (SPAAC)" developed by the Bertozzi group in 2004 (Fig. 1.2).[31] The prototype of SPAAC was based on the highly selective reaction of cyclooctynes with azides in acetonitrile with relatively slow reaction kinetics (second-order rate constant of 0.00012 $M^{-1}s^{-1}$).[31] With the efforts to further improve the reaction kinetics, other more reactive strained cyclooctyne analogs, including difluorocyclooctyne (DIFO)[32], bicyclo[6.1.0]nonyne (BCN)[33], dibenzocyclooctyne (DIBO)[34] and aza-dibenzocyclooctyne (DIBAC)[35], have been developed and they are now widely used in living biological systems, including reactions at the cell surface[36], inside mammalian cells[31] or in zebrafish embryos[37]. In this case, one can envision that the biomacromolecules in the living systems, which play important roles in various cellular processes, can be monitored in situ in their native environments. In addition, pharmaceuticals can also be synthesized or released in situ inside humans to achieve high therapeutic effects with minimal off-target effects. Nowadays, several strained cyclooctyne derivatives have become commercially available, which greatly facilitates their applications in the field of chemical biology. Even though continuous efforts are devoted to develop various strained alkynes with improved reactivity, several major drawbacks of SPAAC still remained such as the

limited water solubility of strained alkynes and their slow reaction kinetics (the most reactive strained alkyne shows the rate constant up to 1 $M^{-1}s^{-1}$).[38]

Besides the azide group, strained cyclooctyne can also react with other 1,3-dipoles, e.g. nitrones.[39] This reaction is termed as "strain-promoted alkyne-nitrone cycloaddition (SPANC)", which was reported by van Delft and coworkers in 2010 by using DIBO (Fig. 1.2).[39] SPANC reactions are fast proceeding with second-order rate constants up to 39 $M^{-1} s^{-1}$ and they are about 30 times faster than the SPAAC reactions. However, the fast reactivity is associated with the instability of the reactive nitrones, which are prone to hydrolysis in aqueous media. As a result, more efforts are required to solve the stability problems so that this reaction could be applied in more complex media or for in vivo applications.

Instead of using alkynes as reactive partner for 1,3- dipolar cycloadditions, the azide group, due to its small size and kinetic stability under physiological conditions, has been in the focus due to its reaction with phosphines to generate primary amines, which was firstly reported by Hermann Staudinger in 1919.[40] Based on this work, Staudinger ligation was introduced by Bertozzi and co-workers in 2000, in which a specific phosphine reagent was designed so that the aza-ylide intermediate could be trapped by an adjacent electrophilic carbonyl group which upon hydrolysis forms a stable amide bond.[41] Shortly after the discovery of the Staudinger ligation, a modified version termed as "traceless Staudinger ligation" was described, in which the oxidized phosphine motif was liberated from the formed conjugate.[42] Nonetheless, the major limitations of Staudinger ligation are the slow reaction kinetics (second-order rate constant around $10^{-3}$ $M^{-1}s^{-1}$) and oxidation of phosphines.[43]

Another well-known 1,3-dipolar reaction is a cycloaddition between nitrile imines generated by photo-irradiation of tetrazole and a terminal alkene (Fig. 1.2).[40] This photoclick reaction was firstly reported by Lin and coworkers in 2008, which offers the advantage of operational simplicity, fast reaction kinetics as well as spatial and temporal control by using light rather than potentially toxic catalysts.[44] However, recent evidence indicates that the so-called "photoclick reactions" suffer from potential cross-reactivity with e.g. amine and carboxylic acid residues, which are abundant at the surface of proteins. Therefore, the term bioorthogonal remains controversial for these reactions.

Among all the existing biorthogonal reactions, the inverse electron demand Diels–Alder reaction (IEDDA) stands out because of the fast reaction kinetics (rate constant of up to $10^6$ $M^{-1}s^{-1}$), excellent bioorthogonality, and good biocompatibility without the need to use toxic catalyst. (Fig. 1.2).[45] The iEDDA reaction describes the reaction between a diene (e.g. 1,2,4,5-tetrazines) and a dienophile (alkene or alkyne) to form a cycloadduct, which subsequently undergoes a retro Diels-Alder reaction to form the final product. Because of its favorable reaction kinetics, it revolutionized our ability to explore the reactions in living systems, in which the rate constant of the reaction is very important as the labeling processes occur with a low concentration of reactants in living cells,

which is known as crowded with organelles, biomembranes, macromolecules and actin filaments. In addition to fast reaction kinetics, another main merit of this catalyst-free reaction is that tetrazines exhibit an absorption at around 500–530 nm, which can serve as a fluorescence quencher when linked with a proper-designed chromophore before iEDDA reactions. Upon reaction, the fluorescence of the fluorophore will be restored showing a "turn-on" effect, which is advantageous to report successful *in vivo* labeling with minimal background noise.[46-48] However, the high reactivity of the diene and dienophile could also limit the stability of the *trans*-cyclooctyne (TCO), which readily isomerizes to the nonreactive configuration "*cis*-cyclooctyne". As a result, many ongoing synthetic efforts are focused on the development of rather stable yet reactive substrates for iEDDA reactions.[49-51]
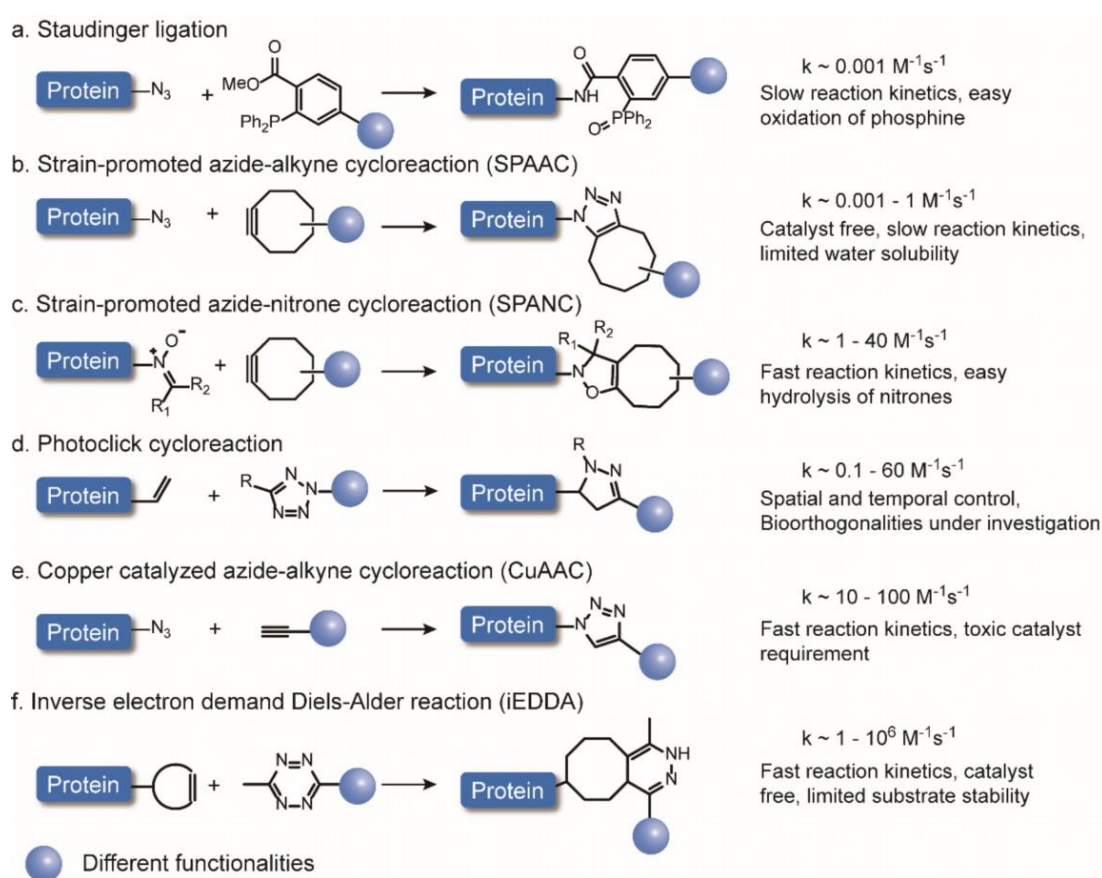


Fig. 1.2 Commonly used bioorthogonal reactions for protein modification and summary of some selected features. Adapted with the permission from reference 20. Copyright (2020), John Wiley and Sons.

## 1.3 Site-selective single modification of proteins

In the past decades, site-selective single modification of proteins has witnessed significant progress and various kinds of methodologies have been developed in order to improve the modification efficiency, specificity as well as the stability of the resulting bioconjugates, offering an

ever-growing methodology toolbox for different intended applications. These strategies either take advantage of the novel chemistry via reacting with the natural AAs on the protein surface that bear distinct reactivity profiles, such as cysteine, or focus on the incorporation of ncAAs via genetic code expansion in combination with bioorthogonal chemistries. In this section, these two different strategies for site-selective single modification of proteins are summarized.

## 1.3.1 Chemical methods for site-selective single modification of proteins

With the restrictions of the high abundance of unprotected chemical entities on the native protein surface, it is usually difficult to achieve protein modification in a site-selective manner by modifying the canonical AAs on the protein surface. However, due to the easy accessibility of native proteins with sufficient quantity, it is still arguably an ideal method to modify native protein by chemical approaches in a site-selective manner.[26, 52-53] In addition, there is no need to conduct genetic engineering, which is often tedious and sometimes causes negative effects on the folding and functions of proteins. Chemical methods for site-selective protein modification are based on the direct modification of a selected, canonical AAs residue on the surface of proteins, e.g. cysteine, disulfide bond, tyrosine, tryptophan and so on.[53-54]

### 1.3.1.1 Single modification at cysteines

Owing to the unique nucleophilicity and versatile reactivity profile of thiol groups, solvent-accessible cysteines, which have a low abundance on the protein surface, are the most popular targets for the site-selective protein modification.[55]

Maleimide compounds are the most commonly used cysteine-modification reagents, which have high selectivity for cysteines and fast reaction kinetics (rate constant up to $10\text{-}10^4$ $M^{-1}$ $s^{-1}$).[56] In addition, a variety of maleimide derivatives containing different functionalities, e.g. dyes, affinity tags, and cross-linkers, are commercially available, which make them easily accessible also for the non-chemist. Nonetheless, maleimide conjugates reveal limited stability as the newly formed thioether bonds can be cleaved in plasma (Fig. 1.3a).[57] In addition, maleimide bioconjugates can undergo retro-Michael reactions to liberate the attached functionality and subsequent exchange reactions with abundant thiol-containing molecules can occur, e.g. human serum albumin (HSA) or glutathione, resulting in off-target effects in *in vivo* applications (Fig. 1.3a).[57] On the other hand, maleimides are also easy to hydrolyze to form the unreactive maleic amides (Fig. 1.3a).[58] Therefore, there are continuing efforts to develop new bioconjugation strategies, which can provide the protein bioconjugates that are stable during blood circulation before reaching their targets to eliminate the premature release. On the other hand, these protein conjugates should be able to dissociate to release the attached cargo in response to certain stimuli after reaching their targeted areas, e.g. cancer cells.

One strategy is to increase the thiosuccinimide hydrolysis rate to produce stable bioconjugate. For example, Dimasi and coworkers demonstrated that thiosuccinimide hydrolysis can be accelerated by utilizing N-alkyl maleimide substituents that contain an aromatic ring to increase the imide electrophilicity (Fig. 1.3b). A different strategy was employed by Setter and co-workers, where an adjacent basic amino group was introduced to promote intramolecular base-catalyzed hydrolysis in order to accelerate the rate of thiosuccinimide hydrolysis (Fig. 1.3b). It was shown that the one-carbon side-chain offered the most effective geometry for rapid and quantitative post-conjugation thiosuccinimide hydrolysis. Instead of accelerating the rate of thiosuccinimide hydrolysis, some exocyclic maleimide derivatives, e.g. 3-substituted exocyclic maleimide, have been developed to improve the stability of the formed thiolsuccinimide conjugate (Fig. 1.3b).[59] Besides improved stability, the conjugation reaction with 3-substituted exocyclic maleimides also demonstrated better chemo-selectivity towards cysteine residues and negligible thiol exchange between the formed bioconjugate and bovine serum albumin (BSA). Alternatively, a chemo-selective adduct can be obtained by employing the exocyclic bond in 5-methylene pyrrolones reacting with cysteines, which also showed better stability than the conventional maleimide conjugation (Fig. 1.3b).[60]

Halogenated maleimides and their analogs are considered as another class of Michael acceptors for cysteine modification through $S_N2$ reactions. Bromomaleimides, including mono-bromomaleimide and dibromomaleimides, developed by Baker and coworkers are among the most popular ones, which allows selective modification of cysteines over other reactive AAs residues, e.g. lysines or arginines. Mono-bromomaleimides have been used to achieve exclusion modification of cysteine residues of the SH2 domain of the growth factor receptor-bound protein 2 protein (single point mutation L111C) and THIOMAB trastuzumab, both of which contain multiple accessible lysine residues on the protein surface.[61] Compared to the classical maleimides, mono-bromomaleimides provide access to the well-defined bioconjugates with improved stability, as the hydrolysis of the thiomaleimide ring is accelerated presumably because of the presence of local cationic charges next to the bioconjugation site. In addition, the resultant conjugates can also undergo thiol exchange reactions to release the payload in the presence of excess tris(2-carboxyethyl)phosphine (TCEP) or Dithiothreitol (DTT), which provide great opportunity to modulate the proteins' function or achieve controlled release of attached cargos, such as drug molecules in a revisable fashion.[62] Furthermore, the resultant thiolmaleimides still contain a Michael acceptor, which allows the incorporation of a second functionality to achieve dual functionalization.[61] With a similar conceptual strategy, dibromomaleimides have also been developed and revealed a similar reactivity profile as mono-bromomaleimide.[63] However, it was shown that the cargo release rate of thiolmaleimide is faster than that of thiolsuccinimide, presenting a first-order kinetic dependence.
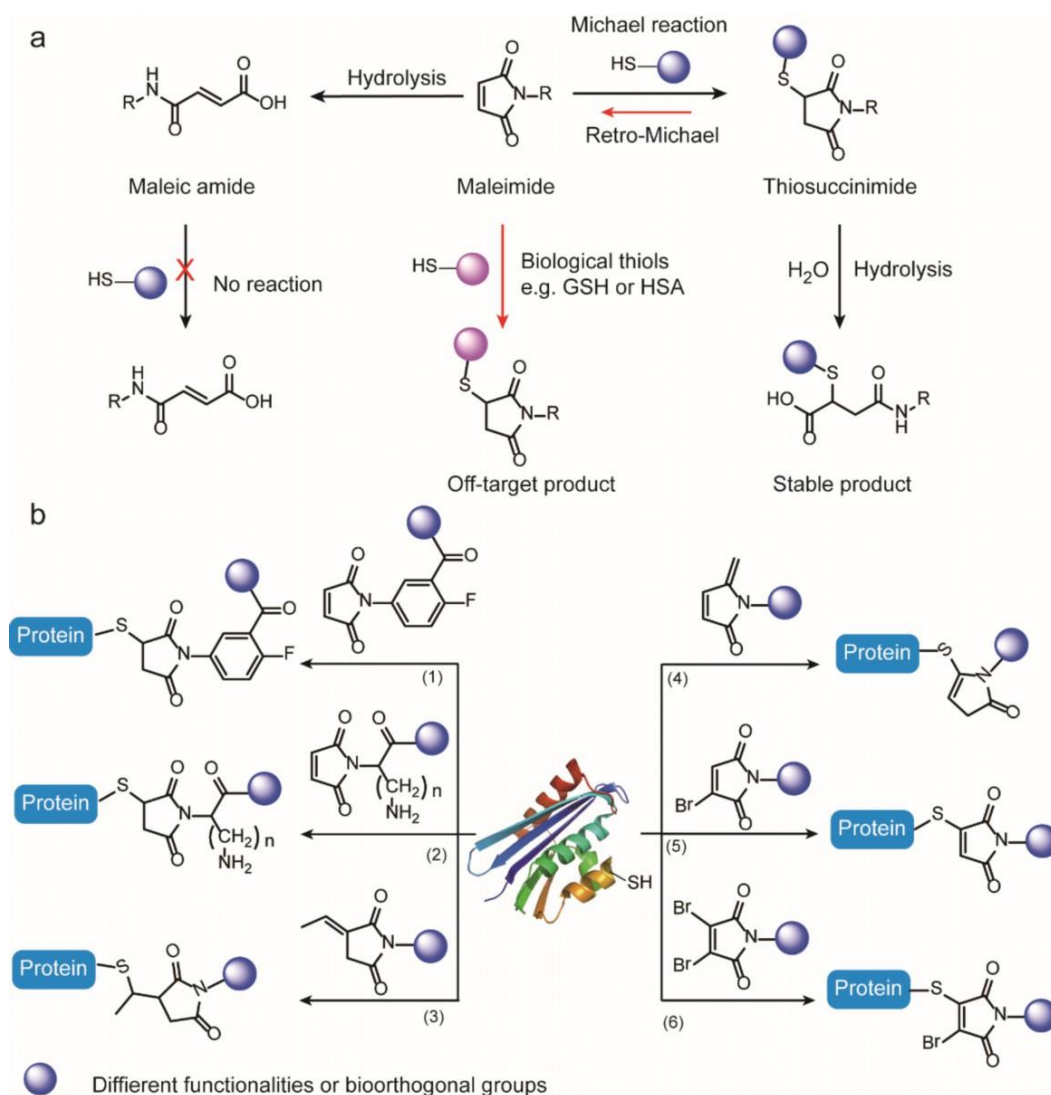
Fig. 1.3 (a) Maleimide conjugations and potential hydrolytic pathways of maleimide and thiosuccinimides (b) Different maleimide derivatives for site-selective protein modification at cysteine site

Besides the classical maleimides and their derivatives, sulfones are another class of Michael acceptors for efficient cysteine modification. For example, N-methyl-N-phenyl vinylsulfonamides allow cysteine modification at neutral pH in an aqueous buffer (Fig. 1.4a).[64] The authors demonstrated the successful conjugation of the cytotoxic agent monomethyl auristatin E (MMAE) to trastuzumab to obtain the well-defined antibody-drug conjugates (ADCs) with retained binding affinity. Alternatively, azabicyclic vinyl sulfones have also been demonstrated for efficient cysteine modification, which also introduces an azabicyclic dienophile that is suitable for further bioorthogonal derivatization by iEDDAs to achieve dual functionalization (Fig. 1.4b).[65] Even though the resulting conjugates are stable against thiol exchange, they are prone to retro Diels–Alder degradation reactions, causing off-target effects. In the following up work, a variation of the azabicyclic bromovinyl sulfone scaffold revealed efficient cysteine modification with dipyridyl

tetrazines leading to bond-cleavage and the formation of a pyrrole-linked conjugate (Fig. 1.4c).[66]

Furthermore, structurally diverse electrophiles have emerged for cysteine modification in recent years. Bernander and coworkers have reported the site-selective modification of native proteins and antibodies with carbonylacrylic derivatives (Fig. 1.4d)[67], 1,3-bromo-oxetane (Fig. 1.4e)[68], and quaternizated vinylpyridine reagents (Fig. 1.4f)[69] at cysteine sites in aqueous conditions. Hackenberger and coworkers have developed the electrophilic ethynylphosphonamidates (Fig. 1.4g)[70] and vinylphosphonothiolate (Fig. 1.4h)[71] for cysteine modification of peptides and proteins. In addition, 1,4-dinitroimidazoles (DNIms) (Fig. 1.4i) have also been shown as highly efficient cysteine bioconjugation reagents developed by Wang and coworkers.[72] Depending on the reaction conditions, DNIms revealed distinct specificity for either cysteine or lysine residues, in which 1,4-DNIms showed exclusively thiol selectivity on proteins in aqueous conditions while demonstrated high lysine selectivity in organic solvents in the presence of weak bases.
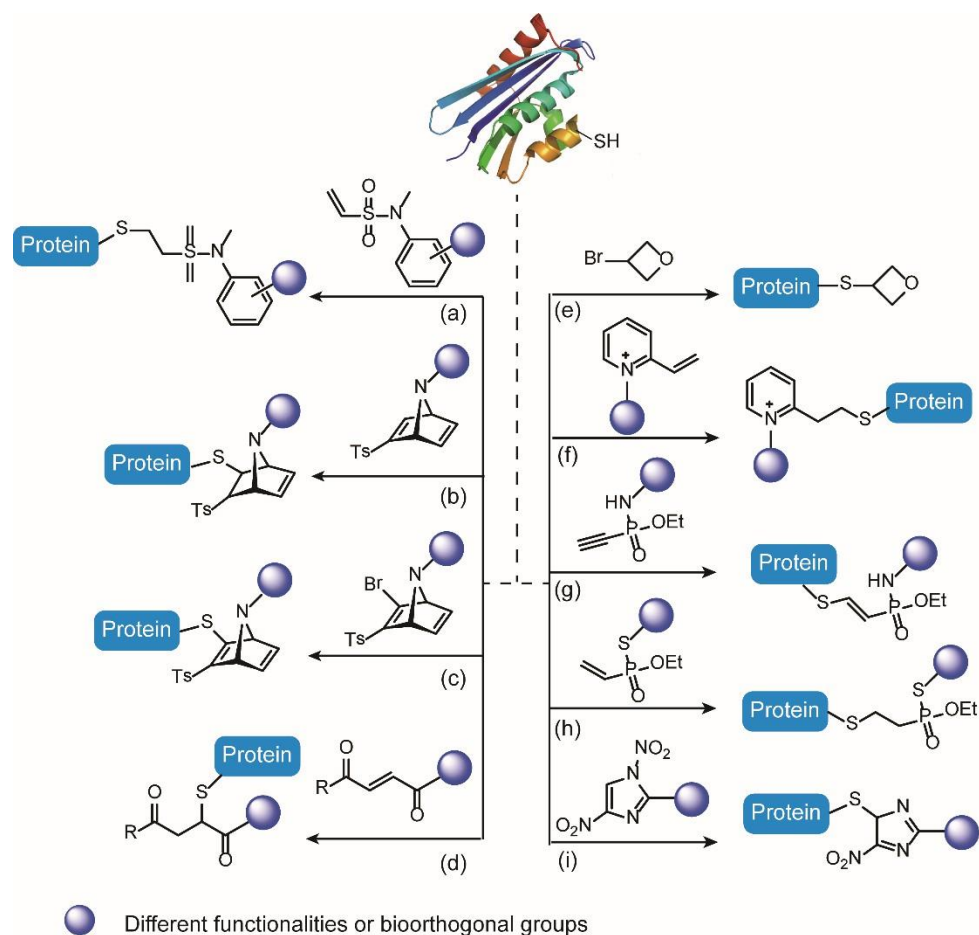


Fig. 1.4 Chemical methods for site-selective protein modification at cysteine sites.

In addition, thiol-ene reactions also provide site-selective protein modifications under mild reaction conditions, in which the thiyl radicals are generated upon irradiation and are subsequently

conjugated to the alkene group (Fig. 1.5a).[73] While thiol−ene coupling reactions enable the incorporation of a single functionality, thiol-yne reactions can allow the introduction of two different functionalities into proteins to achieve dual functionalization (Fig. 1.5b).[74] It includes a first radical addition forming a stable vinyl sulfide moiety that undergoes a second radical addition reaction with a different thiol-containing functionality forming a 1,2-modified thioether compound under mild conditions.[54] Thiol-ene/yne reactions are considered as "click" reactions with the advantages of good bioorthogonality, fast reaction kinetics, no organic catalyst and good biocompatibility in the presence of water and oxygen. However, these reactions also suffer from the photodamage to the integrity of proteins structure, thus leading to the protein degradation, which in turn greatly hampers its broader applicability.
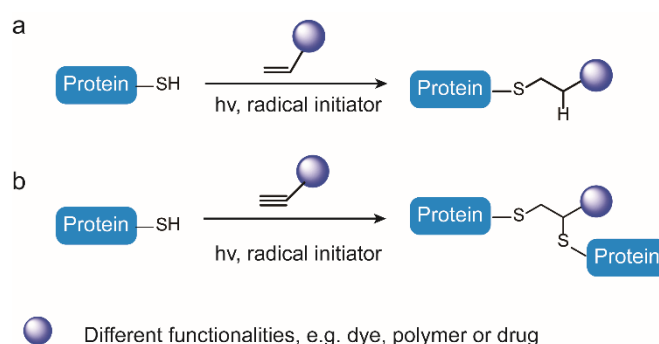


Fig. 1.5 General scheme for (a) Thiol-ene reaction (b) Thiol-yne reaction.

### 1.3.1.2 Single modification at disulfide bonds

Even though cysteine residues are considered as the most popular targets for protein modification, very few native proteins contain unpaired cysteine residues and recombinant techniques are often required to introduce the cysteine mutations to the protein surface for further modification.[75] However, this strategy clearly has limitations if the protein itself consists of critical native cysteines that play important structural or functional roles. Furthermore, free cysteine residues also suffer from a high tendency to form disulfide bonds and the reaction reversibility of thiol-maleimide conjugations. In this context, solvent-accessible disulfide bonds can serve as potentially valuable targets to achieve site-selective protein modification on naturally occurring amino acid residues.[75] In addition, many therapeutic relevant proteins, e.g. antibody or Fab fragment, contain at least one disulfide bond that is solvent accessible.[12] Therefore, disulfide modification to incorporate the tailored functionalities into these therapeutic relevant proteins can open up new prospects for the advancements in therapeutics, diagnostics, and fundamental science.[11]

Bissulfones and dibromomaleimides represent the most well-known disulfide rebridging reagents. The bissulfone reagents developed by Brocchini can be obtained by a four-step synthesis starting from *p*-acetylbenzoic with an overall yield of 28%. Generally, it rebridges the disulfide bonds

by a sequential addition-elimination process, in which a mono-sulfone is obtained from the respective bis-sulfone upon the elimination of the *p*-toluene sulfinic acid, reacts with the first thiol group to yield another Michael system (Fig. 1.6a).[76] Thereafter, the second thiol group undergoes a subsequent Michael addition to rebridge the disulfide bond. One representative example using bis-sulfone reagents for protein disulfide modification is to incorporate hydrophilic PEG chains to an immunotherapeutic protein: interferon.[77] In addition, humanized antibody trastuzumab and its Fab fragment have also been successfully functionalized with bis-sulfone reagents to introduce anticancer drugs for ADCs with high selectivity and efficiency without any interference to their biological and structural integrity.[78] However, side reactions will occur between bis-sulfone reagents and water at basic pH (pH>8) in the absence of thiol groups.[79] Furthermore, poor solubility of bis-sulfone reagents in aqueous media due to a large number of hydrophobic aromatic groups in their scaffolds is also a limitation for efficient bioconjugation. The relatively high amount of organic solvent required for bioconjugation can also lead to the denaturation or aggregation issue, thus greatly diminishing the modification efficiency.

Recently, our group developed the allyl sulfone reagents as efficient disulfide rebridging regents with improved reactivity, high water solubility, and good site selectivity (Fig. 1.6b).[80] Compared to the bis-sulfone reagents, ally sulfone reagents have two fewer hydrophobic benzene groups, thus offering lower n-octanol-water partition coefficients which indicates higher water solubility. Compared to bis-sulfone reagents, the synthesis for allyl sulfone reagents is much more simplified and therefore provides easy access to different disulfide rebridging reagents with various bioorthogonal tags. In addition, allyl sulfone reagents don't need any in-situ elimination to obtain the reactive intermediate for disulfide modification which is usually not quantitative, thus offering higher labeling efficiency. The broad applicability of this approach was demonstrated by the successful functionalization of several therapeutic relevant peptides and proteins, e.g. somatostatin (SST), octreotide, insulin, and lysozyme. Moreover, ally sulfone reagents are also capable to incorporate two different functionalities into proteins at cysteine residues to achieve dual functionalization of proteins, demonstrating its great potential for the preparation of multifunctional protein conjugates for more sophisticated biological study.

Dibromomaleimides developed by Caddick, Baker, and coworkers are considered as another well-known disulfide rebridging reagents (Fig. 1.6c). They contain two bromo groups in the 3- and 4-position, which serve as good leaving groups when reacting with the two thiol groups from the reduced disulfide bonds to offer the thiolmaleimide adducts. The obtained thiolmaleimide ring can undergo hydrolysis to provide stable bioconjugates (< 1h, pH 8). In the presence of other thiol-containing reagents, such as glutathione, a thiol release reaction can take place to regenerate the native protein and liberate the attached functionality, thus offering great prospects for revisable protein functionalization. In addition, dibromomaleimides crosslinkers have also been explored for

the creation of more complex protein conjugates, for example, bispecific conjugates, but with relatively low yields. In order to circumvent this limitation, diiodinemaleimides have been developed with a slower rate of maleimide hydrolysis but a higher rate of disulfide rebridging. It was reported that human serum albumin-Fab conjugate can be obtained by using diiodinemaleimide in quantitative yield. Furthermore, more complex bioconjugates, for example, a heterodimer Fab-scFV and even a triabodies, have also been successfully obtained via this approach demonstrating its great potential for the creation of well-defined protein conjugates with a higher level of complexity. Despite dibromomaleimides demonstrate efficient disulfide modification efficiency, they also suffer from the cross-reactivity with some reducing reagents, e.g. TCEP. In order to overcome this limitation, diarylthiomaleimides reagents have emerged as dihalogenated maleimide replacements so that cross-reactivity with TCEP will be eliminated, thus allowing the one-pot rebridging of disulfide bonds.[81]

Another conceptually similar disulfide rebridging reagent dibromopyridazinediones was developed by Caddick and Chudasama (Fig. 1.6d).[82] This reagent can offer serum stable bioconjugates after disulfide functionalization, which avoids the necessity of additional hydrolysis for stabilization. This reagent has been intensively employed for the generation of homogeneous ADCs, protein-protein conjugates, and targeted nanoparticles. In addition, site-selective dual functionalization of the disulfide-containing proteins can also be achieved by incorporating two different but compatible bioorthogonal tags into dibromopyridazinediones. This paves the way for more advanced and sophisticated biological applications, such as theranostics or dual-modality imaging.[83]

Besides the well-known sulfone and maleimide derivatives for disulfide rebridging, other reagents have also emerged in recent years, which serve as valuable additions to the current disulfide modification toolbox. Bernardes and coworkers reported the oxetane reagents for site-selective bis-alkylation of a range of biologically relevant peptides and proteins, such as SST, antibodies, and Fab fragment, under mild and biocompatible conditions (Fig. 1.6e).[84] It was shown that the stability and activity of the resultant bioconjugates were enhanced via the installation of the oxetane graft. However, this method only allows the oxetane stapling of the disulfide bonds with no possibilities for further functionalization to introduce tailored functionalities, e.g. drug, imaging reagent, or polymers, which greatly hampers its further applications. Another novel disulfide rebridging reagent divinylpyrimidine was developed by Spring and coworkers for selective functionalization of antibodies with high specificity (Fig. 1.6f).[85] Trastuzumab antibody was successfully functionalized with variable payloads with precise drug-to-antibody ratio, offering highly stable and functional antibody conjugates with high efficiency.
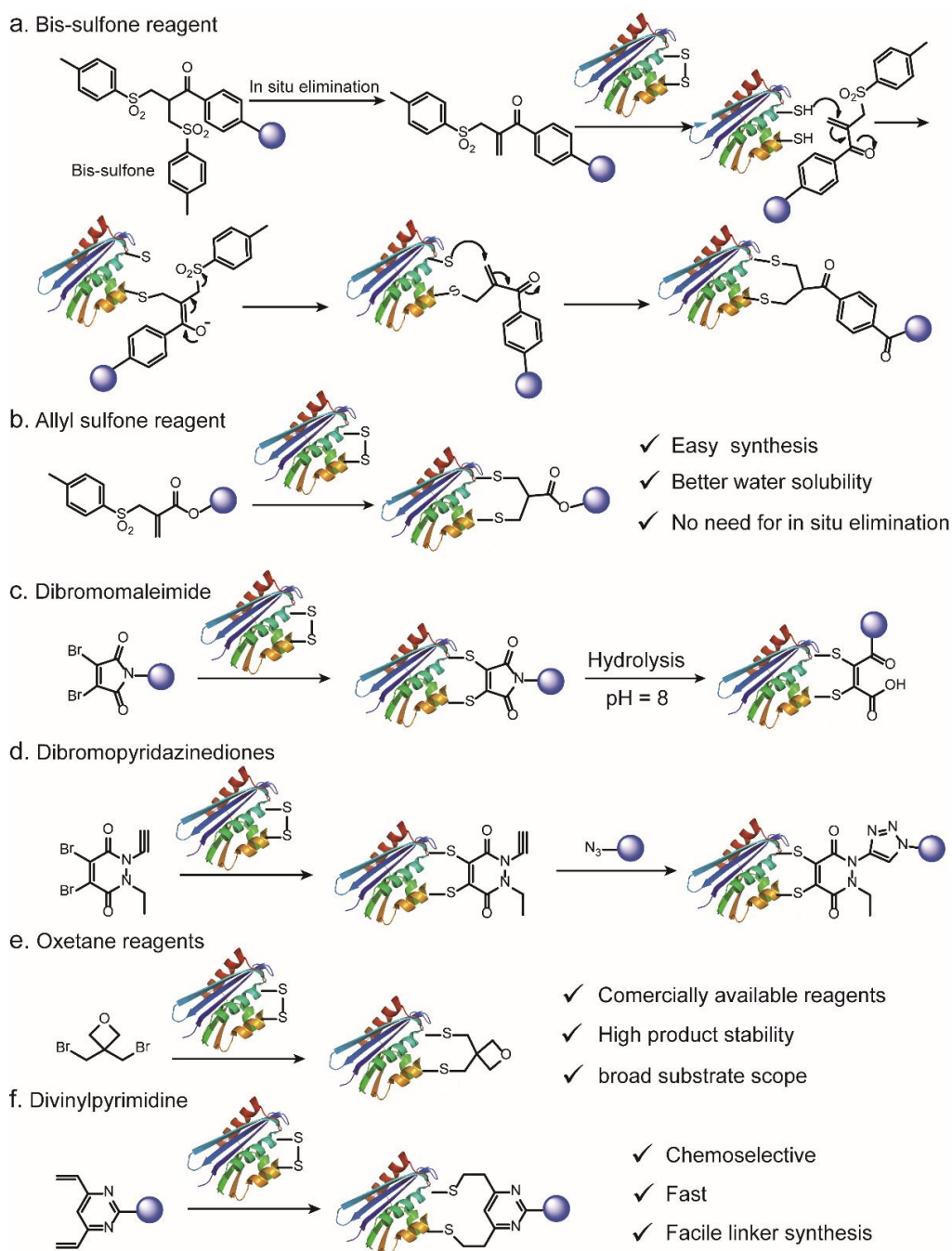
Fig. 1.6 Chemical methods for site-selective protein modification at disulfide site

### 1.3.1.3. Single modification at other amino acid residues

Besides the widely-employed cysteine residues and disulfide bonds for site-selective protein modification, other amino acid residues, for example, N-terminus, tyrosines, and tryptophans can also serve as potential targets for functionalization due to their unique reactivities. Among these, the α-amine of the N terminus stands out as a unique site for modification due to its distinct features compared to other amino acid residues, such as the high accessibilities for modification, minimal

disruption to the integrity of proteins structure and functions, and low basicity.[86] For example, the low pKa of N-terminal α-amine (pKa = 6-8) compared to other aliphatic amines (pKa 10.5) makes the modification at N-terminus modification possible with NHS ester compounds by adjusting the pH (acidic pH required). Even though straightforward, this pH-controlled modification method generally requires a lower pH (usually lower than pH 4.5), which also leads to the decreased conjugation yields with potential interference to proteins structure and function.

Besides the pH-dependent conjugation methods, N-terminus modification can also be successfully achieved with the aid of its side chain through the formation of cyclic intermediates and products. One well-known example is the native chemical ligation (NCL), in which an N-terminal cysteine can react with a thioester to offer an amide bond, making it a powerful approach for site-selective N-terminal modification.[87-88] Besides that, the aldehyde derivatives, e.g. benzaldehydes carrying an ortho-boronic acid functionality[89-90], and 2-cyanobenzothiazole (CBT)[91] were also reported for modification at N-terminal cysteine residues via a condensation reaction offering a cyclic product. Beyond employing the cysteine side chain, other N-terminal amino acid residues, such as tryptophan, serine, and threonine, have also been reported for N-terminal modification owing to the particular features offered by the respective side chain. It was reported that the Pictet-Spengler reaction allows for the cyclic condensation between the aldehyde group, α-amine, and the indole group from the tryptophan side chain to offer a stable cyclic bioconjugate.[92] Furthermore, N-terminal transamination and subsequent oxime ligation are considered as another powerful approach for the functionalization at N-termini. For example, Pyridoxal-5′-phosphate (PLP), an enzyme cofactor, can help to convert the amino groups of the N-terminus to a ketone or an aldehyde, which can further react with an alkoxyamine probe to form an oxime-linked protein bioconjugate.[93] In addition to PLP, Rapoport's salt (RS) is also reported to perform transamination reactions especially with glutamate-terminal proteins, thus offering a powerful method for modification of wide-type IgG1 antibodies.[94] Besides the two-step approach including the transamination and subsequent oxime conjugation, a one-step strategy using 2-pyridinecarboxaldehyde (2PCA) has recently been reported for N-terminal modification, which bears the advantage of being general and straightforward. Enzyme-mediated N-terminal modification also emerges as an attractive methodology over the past decades due to its high selectivity and biocompatibility. Sortase A, N-myristoyltransferase (NMT), Subtiligase and Butelase 1 have all been reported for N-terminal modification with high selectivity and efficiency, serving as valuable additions for the N-terminal modification toolbox.[86, 95-97]

Compared to its cysteine analog, the modification at methionine residue remains unexplored due to its poor solvent accessibility and redox sensitivity. Even though challenging, several remarkable works have been reported for selective methionine modification by utilizing its high
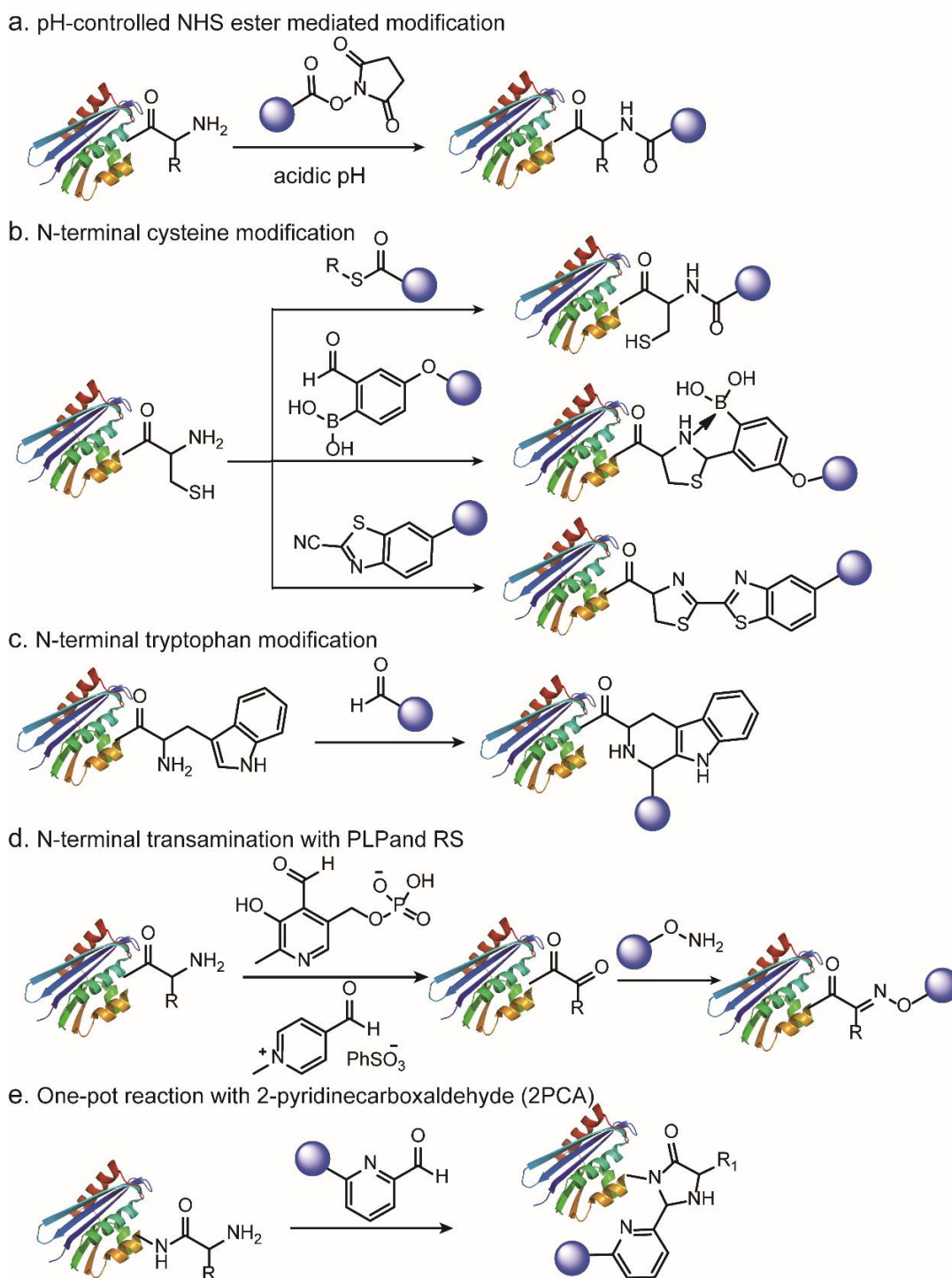
Fig. 1.7 Site-selective protein modification at N-terminal residue

intrinsic feature. In a recent effort, Chang and coworkers developed a redox-activated chemical tagging for methionine residues in native proteins by utilizing the well-designed oxaziridine reagents to form functionalized sulfimide.[98] Another remarkable approach was reported by Gaunt and coworkers by using the hypervalent iodine reagents for methionine modification without any cross-reaction with other amino acid residues.[99]

Amino acid residues with aromatic side chains have also been often reported for modification. For example, Francis and coworkers demonstrated the Rh-carbenoid mediated chemoselective modification of tryptophan with high efficiency.[100] Besides that, TIPS-EBX or 1-[(triisopropylsilyl)ethynyl]-1,2-benziodoxol-3(1H)-one have also been developed for tryptophan modification, leading to the C2-ethylation of indole group in tryptophan residue through the Au(I) mediated transformation. As for tyrosine, the phenolic side chain dominates the reaction landscape offering the potential for chemoselective modification. For example, the aromatic carbon ortho to the hydroxyl group can undergo ene-type reactions[101], diazonium couplings[102], and Mannich-type reaction[103] under acidic or neutral conditions. The histidine residue has an imidazole group which offers a moderate nucleophilicity (pKa = 6). Therefore, the modification at histidine residue suffers from the cross-reaction with cysteine, lysine, and N-terminal amine. For example, iodoalkane is well-known for histidine modification but the side reactions towards other nucleophiles are considered as a major limitation for further applications.[53] Even though challenging, other reagents, such as thiophosphorodichloridate and 2-cyclohexenone, have been developed to enable the single site labelling of histidine with satisfied yields.[104-105]
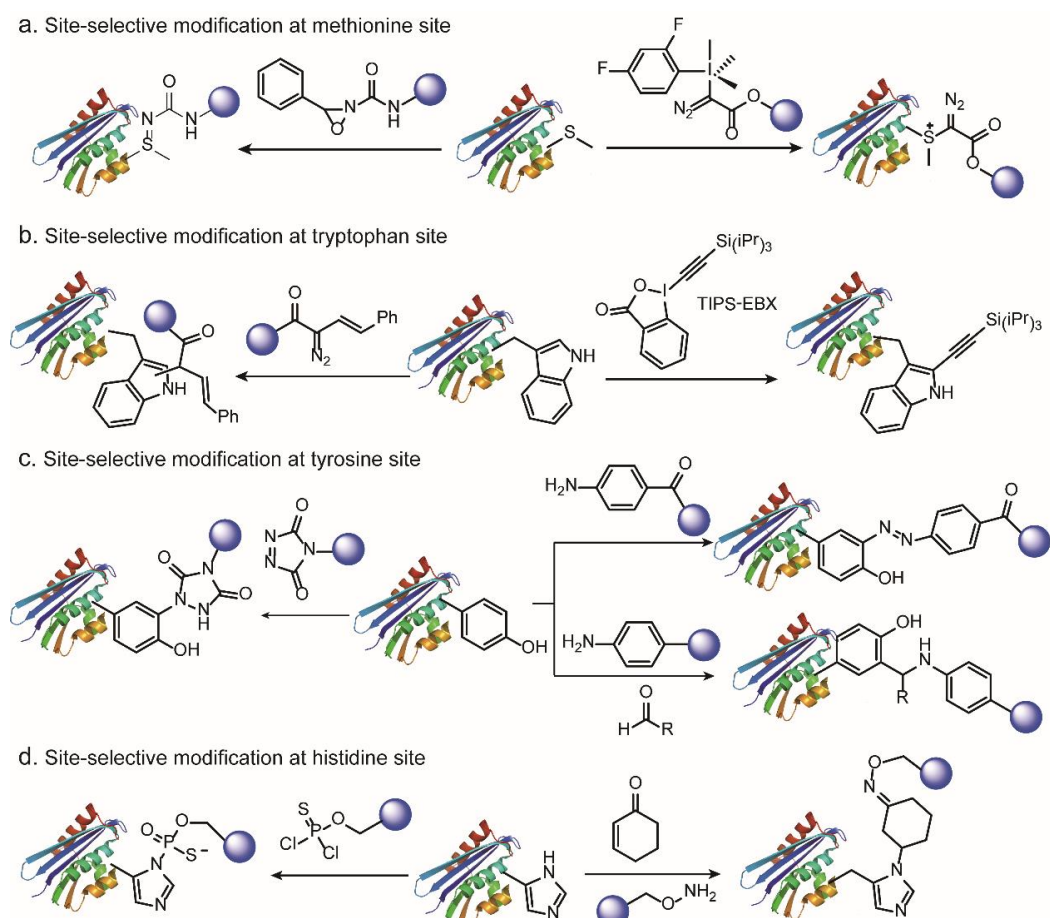


Fig. 1.8 Site-selective protein modification at (a) methionine residue (b) tryptophan residues (c) tyrosine residues (d) histidine residue

## 1.3.2 Genetic methods for site-selective single modification of proteins

Although chemical methods for site-selective protein modification bear the advantages of simplicity and straightforwardness, it is greatly restricted by the limited type of functionalities offered by the 20 canonical AA residues on the protein surface. To expand the reactivity beyond what is offered by nature, the incorporation of new AAs (canonical or noncanonical ones) at distinct sites by genetic methods has become an essential tool for site-selective protein modification. Owing to the development in synthetic biology, the incorporation of canonical AAs into proteins has become well-established and straightforward. Among all the canonical AAs, cysteine residues remain the most popular AA candidate for genetic insertion owing to the high nucleophilicity of the thiol group, relatively low abundance on the protein surface as well as the ever-expanding chemical toolbox at the cysteine site. As a consequence, early attempts to achieve site-selective protein modification focus on the introduction of a cysteine mutation to the target protein at distinct site, which can then be functionalized by choosing the respective method from the available cysteine-modification toolkit.[106]

Despite the simplicity of introducing cysteine mutations as the functional sites for site-selective modification, it is problematic if the protein itself consists of critical native cysteines that play important structural or functional roles. Hence, the most general and reliable strategy is to incorporate ncAAs bearing bioorthogonal tags, which can react with the respective bioorthogonal groups without cross-reaction with other functionalities on the protein surface. Along with the significant progress in the genetic engineering field, manipulating the cellular biosynthetic machinery to incorporate ncAAs into proteins at pre-determined sites has flourished and paved the way for labeling and manipulation of biomolecules with unprecedented molecular precision.[107] Until now, an incredibly diverse range of ncAAs bearing unique functionalities or bioorthogonal handles has been reported for genetic incorporation into proteins either in a residue- or site-specific manner.[108] Residue specific strategy to incorporate the unnatural amino acids is based on the addition of AA analog which can compete as a substrate for a specific endogenous aminoacyl-tRNA synthetase, allowing for the replacement of the natural AAs in auxotrophic strain.[109] In general, azide- or alkyne-containing methionine analogs have been incorporated into proteins in methionine-auxotrophic *E. coli* strain using this strategy. Even though simple, the library of the noncanonical AAs, which can be incorporated into proteins using natural synthetases are rather limited, thereby lacking general applicability.

Alternatively, the site-specific introduction of noncanonical AAs can also make use of the genetic code expansion technique, which in theory can insert uncanonical AA at virtually any desired site, thus presenting a much higher level of flexibility and spatial precision.[106] It is often accomplished by using the orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pair, charging a
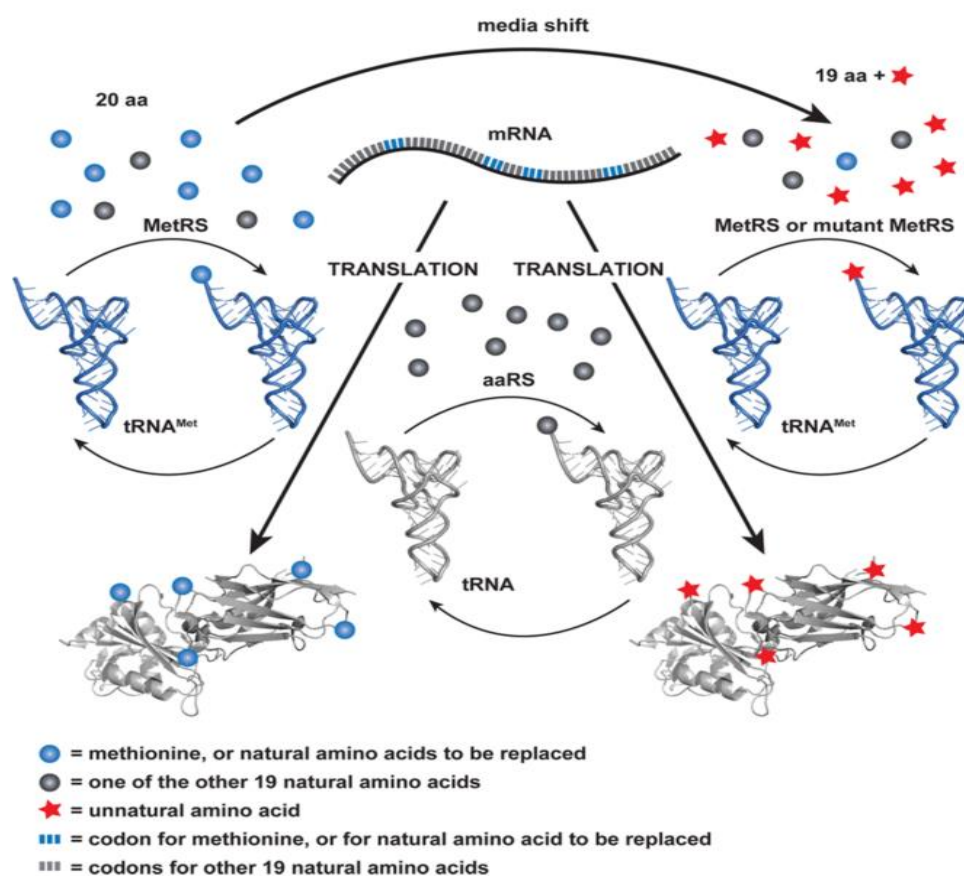
Fig. 1.9 Selective pressure, residue-specific incorporation of unnatural amino acids into proteins. A natural mRNA contains codons for the 20 natural amino acids (gray). Methionine (blue sphere, corresponding blue codons on mRNA) is replaced with an unnatural amino acid (red star). A medium shift removes methionine, in an auxotrophic strain, and introduces the unnatural amino acid (red star) together with the other 19 natural amino acids (gray spheres). The unnatural amino acid is charged to tRNAMet, and the unnatural amino acid is incorporated into proteins in place of methionine. Adapted with permission from reference 12. Copyright (2014), American Chemical Society

designed noncanonical AA in response to a nonsense codon, such as amber stop codon (UAG), due to their minimal occurrence in most organisms. Within this approach, the orthogonal aminoacyl-tRNA synthetase will only aminoacylate its cognate orthogonal tRNA without cross-reaction with any endogenous tRNAs in the host cells but its cognate orthogonal tRNA. Similarly, the orthogonal synthetase is only the substrate for the orthogonal tRNA, which will then specifically direct the target unnatural AAs into proteins at the pre-selected sites.[108] There are four aminoacyl-tRNA synthetases commonly employed in the literature for introducing unnatural AAs into proteins: the *Methanococcus jannaschii* Tyrosyl-tRNA synthetase (*Mj*TyrRS)/tRNACUA pair, the *E. coli* Tyrosyl-tRNA synthetase (*Ec*TyrRS)/tRNACUA pair, the *E. coli* Leucyl-tRNA synthetase (*Ec*LeuRS)/tRNACUA pair, and the Pyrrolysyl-tRNA synthetase (PylRS)/tRNACUA pairs from *Methanosarcinae*.[110] Until now, more than 50 uncanonical AAs containing various reactive handles and functionalities have been reported to be genetic encoded into the proteins with this strategy.

Over the past few decades, genetic code expansion has witnessed incredible achievements, which greatly promotes and revolutionizes the field of site-selective protein modification. Compared to other methods, this strategy bears significant advantages, such as the small size of the noncanonical AAs, flexible incorporation sites but high site-selectivity, as well as diverse reactive tags available for incorporation.[111] Despite these exciting and outstanding advancements, significant hurdles remain in this rapidly evolving field, e.g. low catalytic efficiency of engineered aminoacyl-tRNA synthetases, low expression yields, tedious optimization of protocols for effective protein engineering, the scope and compatibilities of the functional groups capable to be inserted. Therefore, further efforts need to be devoted to improve the encoding efficiency, expand the current orthogonal aaRS/tRNA pair toolbox as well as increase the expression yield.
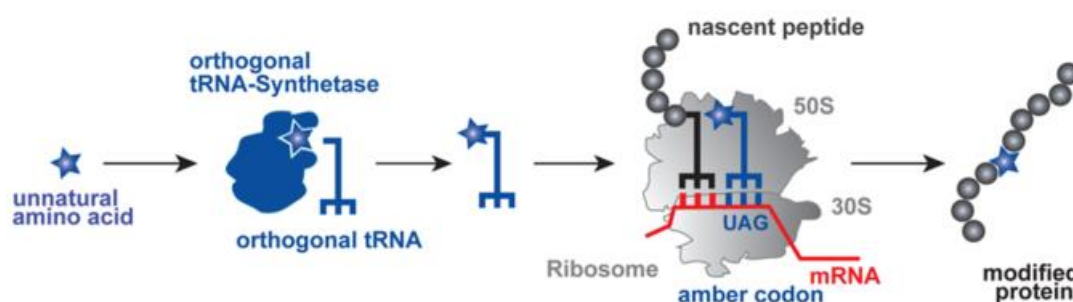


Fig. 1.10 Expanding the genetic code. An unnatural amino acid (blue star), added to the cell growth medium, is specifically recognized by an orthogonal aminoacyl tRNA synthetase and attached to an orthogonal amber suppressor tRNA, which is decoded by the ribosome in response to an amber codon (UAG) introduced into the gene of interest, allowing the synthesis of a protein with a site-specifically introduced unnatural amino acid. Adapted with permission from reference 12. Copyright (2014), American Chemical Society

## 1.4 Site-selective dual modification of proteins

**This section has been already published in the following journal:**
**L. Xu, S. L. Kuan, T. Weil. Contemporary Approaches for Site-selective Dual Functionalization of Proteins. Angew. Chem. 2021, 133, 13874 – 13894.**
**Copyright 2020 The Authors. Published by Wiley Distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, https://creativecommons.org/licenses/by/4.0/**

The past decades revealed significant progress in the development of new methodologies for site-selective protein functionalization to install the single functionalities at pre-defined sites. These well-defined protein conjugates offer great prospects for a wide range of fields including biomedicine, bioimaging, biosensing and materials science. Nevertheless, there is an increasing demand for multifunctional bioconjugates to perform more sophisticated biological studies *in vitro* as well as *in vivo*, appending only one type of functionality to proteins is often insufficient to

customize proteins for the desired applications. For example, despite the clinical success of ADCs, classical ADCs equipped with a single type of anticancer drug could still suffer from low efficacy, drug resistance, unfavorable pharmacokinetics, immunogenicity, and the inherent hydrophobicity of the drug, which greatly hampers their further *in vivo* applications. In addition, there is also an increasing demand for new therapeutic strategies which combine imaging agents and drug molecules within the protein, e.g. antibody, for real-time monitoring during the treatment.[112] In another example, a protein was functionalized with three copies of cell-targeting somatostatin peptide and an enzyme. Remarkably, the resultant bioconjugate inhibited tumor growth already at 100-fold lower concentration than a clinically approved antibody acting via a similar mode of action. Furthermore, co-administration of this protein bioconjugate with an approved anti-cancer drug, doxorubicin, boosted its antitumor activity in a combination therapy approach.[113] However, these multifunctional protein conjugates have been mainly achieved by statistical modification on the protein surface, which results in heterogeneous mixtures with reduced protein activity as well as batch-to-batch variations. In view of biosafety, these limitations significantly hamper their further developments. Consequently, there is a pressing need to devise new strategies to generate protein bioconjugates that exhibit higher order of structural and functional complexity but retaining structural perfection. In this regard, site-selective dual functionalization of proteins has emerged as a promising strategy to customize proteins for the respective applications.

In this review, dual functionalization is defined as the incorporation of two different functionalities into proteins in a site-selective fashion, which is accomplished either at two different amino acids (AA) sites or at a single AA residue at the protein surface (Fig. 1). In both cases, the reagents as well as the sequence of the bioconjugation reactions need to be carefully considered. Dual-modified protein bioconjugates can harness the function and properties imparted by the respective payloads, which complement the capabilities of biomolecules and significantly expand their functional arsenal.[114] For example, dual functionalization with two chromophores at distinct sites allows real-time monitoring of changes in protein conformations and dynamics upon ligand binding in native environments or in response to certain stimuli by Förster Resonance Energy Transfer (FRET) measurements. This cannot be achieved by the mono-functionalization of proteins. Moreover, a new generation of protein therapeutics/diagnostics can be derived by incorporating two different functionalities, as exemplified by integrating a drug and an imaging agent into one platform for simultaneous therapy and diagnostics applications (theranostics) or inserting two different drugs into an antibody for combination therapy.[112, 115-116]

Owing to the emerging interest for addressing previously "unanswered" fundamental scientific questions as well as offering practical solutions for biomedical applications, site-selective dual functionalization of proteins has seen rapid developments in the past ten years. In this review, we first summarize the synthetic strategies for protein dual functionalization to provide a timely

overview of this burgeoning field, which serves as a guideline for the selection of the most appropriate method for the envisaged applications (Figure 2). In addition, the rationale and principles behind these synthetic strategies as well as the strength and inherent limitations are discussed. In the last section, some representative applications of dual-functionalized protein conjugates benefitting from the additive or even synergistic features are highlighted.

## 1.4.1 Chemical methods for site-selective dual modification of proteins

Most synthetic strategies for protein dual modification target canonical AAs exposed at the protein surface. These residues are immediately accessible without the need for tedious genetic engineering of recombinant protein variants and also mitigate the risk of negative effects on proteins folding and function. Chemical methods for dual functionalization of proteins can be achieved at two different sites as well as at a single site on the protein surface.

### 1.4.1.1 Dual modification at two different sites

Direct modification of two rare AA residues on the protein surface is a straightforward approach to achieve dual modification of proteins. The combination of two different mono-functionalization methods requires the stringent selection to ensure orthogonality, compatibility and preferably high modification efficiency. For example, Gaunt and coworkers developed a method based on chemoselective labeling of a single methionine residue with a hypervalent iodine reagent.[99] This hypervalent iodine reagent selectively reacted with the moderate nucleophilic methionine residue in the presence of other competitive nucleophilic AA residues, which make it compatible and complementary to other bioconjugation strategies targeting other AA residues. This has been demonstrated by first modifying the unpaired cysteine residue of GTP-binding protein fragment Gα with a maleimide derivative in a Michael reaction to form a thioether bond. Subsequent modification with hypervalent iodine reagent selectively addresses the thioether in a methionine residue yielding the dual modified protein bioconjugate (Fig. 5a). Notably, the methionine modification showed high site selectivity without any cross-reactivity with the thiol-maleimide conjugation. Besides that, Paavola and coworkers have also achieved dual functionalization by combining cysteine modification and an N-terminal transamination reaction mediated by pyridoxal 5-phosphate.[117] The periplasmic glutamine binding protein was site-selectively functionalized by a FRET pair, in which the ligand-induced conformational movements were monitored via changes in FRET efficiency.

An alternative strategy for site-selective protein dual modification is based on the differences in the reactivity of cysteines in their free (thiol) and oxidized (disulfide) forms.[118] Disulfide bonds could be considered as protected thiols, which require activation to form the reduced free thiols for subsequent functionalization. Therefore, site-selective dual modification of native proteins proceeds stepwise by modifying an unpaired cysteine with a maleimide reagent as the initial step,

followed by the disulfide reduction to liberate two additional free thiols, which will react with a disulfide rebridging reagent, e.g. an allyl sulfone (Fig.5b). It is essential that the thiol-maleimide reaction should be applied first to functionalize the unpaired cysteine residue, followed by the disulfide functionalization using allyl sulfone reagents. Otherwise, the allyl sulfone reagents will also react with the unpaired cysteine, resulting in a heterogeneous product mixture.
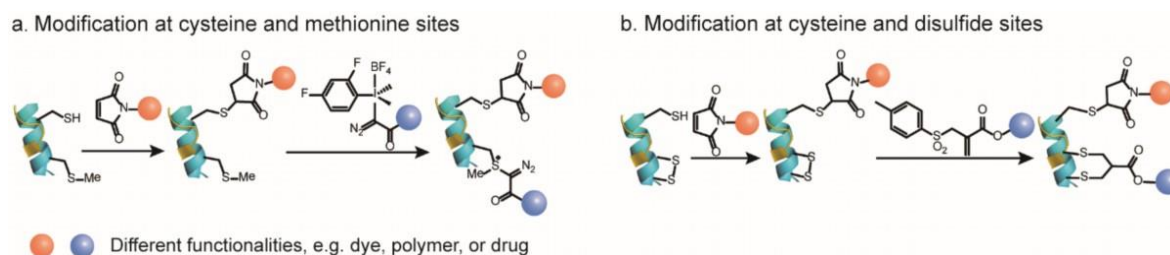


Fig. 1.11 Site-selective protein dual modification at different AA residues. (a) Cysteine and methionine sites.59 (b) Cysteine and disulfide sites. Adapted with permission from reference 20. Copyright (2020), John Wiley and Sons.

**1.4.1.1 Dual modification at a single site**

Despite the simplicity of dual modification at two distinct sites, the availability of proteins with two different AA residues with orthogonal reactivities is rather limited. Therefore, alternative methods have been developed to target one specific AA residue with a multifunctional bioconjugation reagent.

Baker, Caddick, and coworkers have demonstrated protein dual modification using mono- and dibromomaleimide reagents.[61] The first functionality is introduced by reacting these reagents with an accessible thiol group through an addition-elimination reaction. Subsequently, an additional thiol conjugation introduces the second functionality (Fig.6a). A conceptually similar strategy utilizing dibromopyridazinediones is depicted in Fig. 6a.[82] Interestingly, the native protein, e.g. Grb2 adaptor protein, could be regenerated from the dual-modified conjugate after the addition of phosphine or a large excess of thiols, which opens access to the reversible modulation of proteins function or controlled release of the attached cargos, such as drug molecules.[82] In a recent example, another maleimide analog, 3-bromo-5-methylene pyrrolones (3Br-5MPs), was reported for cysteine-specific dual modification of proteins, which has comparable modification efficiency but higher cysteine specificity than the traditional maleimide reagents (Fig. 6b).[119] The dual modification was achieved by two sequential Michael reactions. First, a Michael reaction of cysteine and 3Br-5MPs generated the bioconjugate that is amenable to a second Michael addition with another thiol, allowing protein dual functionalization at a cysteine site. Due to the slow release of the second functionality, a reducing reagent, for example NaBH$_4$, was required to retard the elimination reaction to generate a stable and bioactive conjugate for subsequent applications.[119]

Alternatively, vinyl sulfones are also commonly explored Michael acceptors for protein modification due to their high electrophilic properties that enable their reaction with nucleophiles on the protein surface. Nevertheless, their application for dual functionalization was hampered by the cross-reactivity with e.g., amino or imidazole groups generating heterogeneous products.[120] Recently, Bernardes and coworkers combined the strained [2.2.1]bicyclic systems with the vinyl sulfone systems and developed the azabicyclic vinyl sulfone reagents for dual functionalization (Fig. 6c).[65] Such combination results in a fast chemoselective protein modification at the cysteine site, while the dienophile in the azabicyclic strained moiety concomitantly offers an opportunity for further bioorthogonal modification via iEDDA to liberate the energy stored in the strain systems. The second functionalization could even proceed inside living cells for selective apoptosis imaging. Besides vinyl sulfone, allyl sulfone reagents with enhanced water solubility and higher reactivity have also been proposed as a viable strategy for dual modification in a stepwise fashion.[80] By simply adjusting the pH, allyl sulfone reagents first reacted in a Michael reaction at pH 6 to attach the first functionality, yielding a conjugated ester system that reacted with the second thiol-containing moiety at pH 8 to achieve dual functionalization of proteins (as depicted in Fig. 6d).[80]

In addition to maleimide analogs and sulfone derivatives, other strategies have also been developed for dual-modification. For example, Goncalves and coworkers reported that dichloro-1,2,4,5-tetrazine can undergo two successive nucleophilic aromatic substitutions to introduce the thiol-containing payload at a cysteine residue with excellent selectivity (Fig. 6e).[121] The tetrazine linkage could serve as the second handle for subsequent bioorthogonal iEDDA reaction, allowing the preparation of site-selective dual-modified protein conjugates. The feasibility of this strategy has been shown by the dual labeling of the human serum albumin with a macrocyclic chelator for nuclear imaging and a fluorescent probe for fluorescence imaging.[121] Despite the simplicity of this method, it could suffer from lower yield if bulky and hydrophobic functionalities need to be incorporated. Furthermore, a different strategy utilizing the inherent reactivity of the hypervalent bond was also reported. Waser and coworkers showed the dual modification of proteins with ethynylbenziodoxolones (EBXs) in high efficiency and chemoselectivity by introducing two reactive groups, i.e., an azide and a hypervalent iodine (Fig. 6f).[122] Dual modification was achieved via a strain-release-driven cycloaddition and Suzuki–Miyaura cross-coupling of the vinyl hypervalent iodine bond with using palladium diacetate complex as catalyst.

Due to the emergence of ADCs for targeted cancer therapy, dual functionalization at the disulfide site has also gained growing interest because of the presence of accessible disulfide bonds in antibodies and the antigen-binding fragment (Fab). Previously, dibromopyridazinediones have been extensively employed as versatile reagents for dual modification at the single cysteine site.[82] Further reports revealed that it can also serve as a disulfide rebridging reagent to introduce two bioorthogonal tags into disulfide-containing proteins, for example, antibodies or antibody
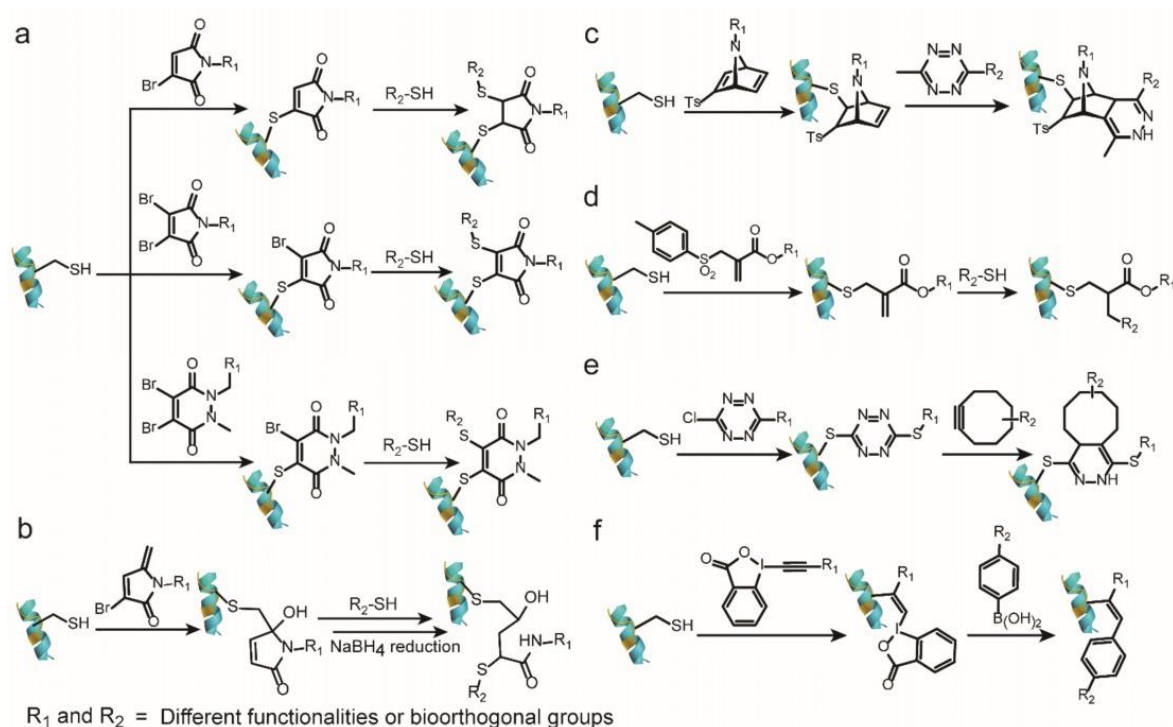
Fig. 1.12 Site-selective protein dual modification at a single cysteine site to introduce multifunctional bioconjugation reagents. (a) Mono and dibromomaleimide, dibromopyridazinediones (from top to down) (b) 3-Bromo-5-methylene pyrrolones (3Br-5MPs) (c) Azabicyclic vinyl sulfone (d) Allyl sulfone (e) Dichloro-1,2,4,5-tetrazine (f) Ethynylbenziodoxolones (EBXs). Adapted with permission from reference 20. Copyright (2020), John Wiley and Sons.

fragments. Chudasama *et al.* exploited the insertion of dibromopyridazinediones bearing two bioorthogonal tags into the disulfide bonds in full antibody and antibody Fab fragments (the antigen-binding fragment) without perturbing the internal disulfide bonds that are vital for activity (Fig. 7). Such a plug-and-play platform allowed the introduction of two functionalities via two sequential bioorthogonal reactions in a modular and efficient way, paving the way for the next-generation ADCs.
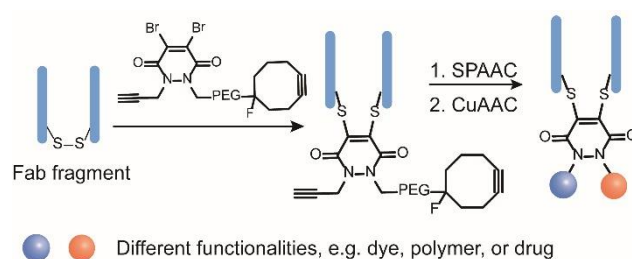


Fig. 1.13 Site-selective dual modification of proteins at the disulfide site of antibody Fab fragment. Adapted with permission from reference 20. Copyright (2020), John Wiley and Sons.

Besides addressing cysteines and disulfide bonds, other modification strategies at less-explored AA residues have also been reported to expand the existing protein functionalization toolkit. For example, the hypervalent iodine reagents developed by Gaunt and coworkers can be

combined with maleimide reagents to achieve dual functionalization at cysteine and methionine sites, which is described in Fig. 5a. In addition, the hypervalent iodine reagents have also been demonstrated to show multifaceted reactivity.[99] The electrophilicity of the diazo sulfonium conjugate enables a photoredox radical cross-coupling reaction with C-4 benzylated Hantzsch ester derivatives to attach the second functionality yielding dual functionalized conjugates with high conversion.[99]

Achieving dual modification at one single AA site has less restriction in terms of the choice of reagents compared to the functionalization at two AA sites. One possible limitation of the single-site strategy is that the two orthogonal groups can be sterically hindered due to close proximity on a relatively small multifunctional linker. This could prevent bulkier groups such as polymers to be attached to the protein. In addition, modification besides cysteine residues is relatively unexplored, thus this field would greatly benefit from further investigations of modification strategies at other low abundant AA residues such as tyrosine, serine or the N-terminus.

**1.4.2 Genetic methods for site-selective dual modification of proteins**

To expand the reactivity beyond what is offered in native proteins, reactive groups for functionalization through genetic incorporation of new AAs (canonical or noncanonical ones) have emerged as a valuable tool for site-selective protein dual functionalization. In this section, genetic methods to integrate new reactive canonical AAs, ncAAs or peptides tags for dual protein functionalization are summarized.

**1.4.2.1 Incorporation of canonical amino acids**

The expression of recombinant proteins containing one or more point mutations is a straightforward and well-established technology.[108] Early attempts to achieve dual functionalization of proteins via genetic methods focused on the preparation of dual cysteine mutants of the target protein. Two cysteine mutations were incorporated at two distinct locations and both thiols exhibited different reactivity toward different thiol-reactive reagents.[123] For instance, Caddick *et al*. reported the two-cysteine insertions to genetically engineered antibody mimetic proteins, so-called designed ankyrin repeat proteins (DARPins).[123] Both thiols are carefully selected and revealed different nucleophilicity, which may originate from their different solvent accessibility. This allowed for the dual functionalization to be executed in a stepwise fashion with high overall labeling efficiency (Fig. 8a).[123] The authors proposed that the less reactive reagent (bromoacetamide) was reacted first with the more nucleophilic cysteine and subsequently, the more reactive reagent (maleimide) was applied to the second, less nucleophilic cysteine. They successfully demonstrated that the differences in thiol nucleophilicity afforded a homogeneous product with quantitative conversion at each reaction step, and no purification was needed. Although this concept for dual modification of

proteins is very elegant, the delicate balance of thiol reactivity and the selection of the most appropriate cysteine mutation site to prevent heterogeneous product formation (single and dual modified products) could be very challenging. Caddick and coworkers also succeeded in protein dual modification by reacting two cysteines with the same thiol-reactive reagent generating two identical sulfoniums.[124] Due to the different accessibilities of the α-protons at the two cysteine mutations, one of the sulfonium groups, which had a good solvent-accessible α-proton on adjacent position, underwent a β-elimination reaction affording dehydroalanine (Fig. 8b). In contract, the other sulfonium group, which has a shielded α-protons next to it, remained intact because of the different protein local microenvironment. The functionalities were incorporated via two different chemoselective reactions, offering the site-selective dual-modified protein conjugate in high yield. The examples mentioned above took advantage of the different local microenvironments offered by the native protein. Inspired by Nature's elegance, Pentelute and coworkers have developed a strategy to create a specific local chemical environment for the cysteine residue by a newly developed, fine-tuned four-amino-acid peptide sequence (FCPF), which is termed as "π-clamp" (Fig. 8c).[125] This π-clamp enables the conjugation exclusively at this cysteine site with perfluoroaromatic reagents with almost quantitative conversion. The reaction proceeds even in the presence of other competing thiols thus rendering this approach compatible and complementary to other thiol-conjugation strategies.[125] Dual functionalization has developed a strategy to create a specific local chemical environment for the cysteine residue by a newly developed, fine-tuned four-amino-acid peptide sequence (FCPF), which is termed as "π-clamp" (Fig. 8c).[125] This π-clamp enables the site-selective demonstrated on a model protein substrate bearing a cysteine and a π-clamp mutation, which was functionalized with a perfluoroaryl probe first based on the π-clamp-mediated conjugation and followed by the thiol-maleimide conjugation reaction.[125]
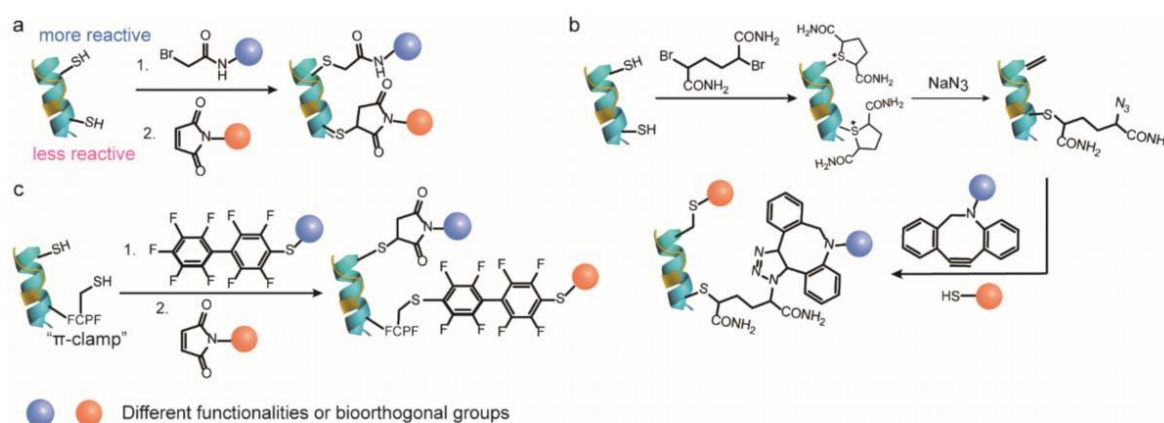


Fig. 1.14 Genetic encoding of two cysteine mutations with different nucleophilicity. (b) Genetic encoding of two cysteine mutations possessing different protein local microenvironment. (c) Genetic encoding of a cysteine mutation and a "π-clamp" FCPF peptide sequence creates a new microenvironment for the cysteine residue. Adapted with permission from reference 20. Copyright (2020), John Wiley and Sons.

## 1.4.2.2 Incorporation of noncanonical amino acids

The introduction of point mutations has certain limitations as the available functionalities and their respective reactivities could only be selected from the pool of the 20 canonical AAs. However, the cellular biosynthetic machinery can be manipulated to incorporate ncAAs that often represent structurally similar derivatives of the canonical AAs. These ncAAs allow protein labeling with unprecedented molecular precision.[126] To date, a diverse set of ncAAs with various functionalities or bioorthogonal groups has been reported for genetic incorporation into proteins. In the auxotrophic strain, an organism such as *E. coli* is applied that is not able to synthesize a certain amino acid required for its growth. The ncAAs need to structurally resemble the natural AA to allow binding to the respective endogenous aminoacyl-tRNA synthetase and to replace the natural AA in the polypeptide sequence.[108] This strategy has been mainly employed to introduce azide- or alkyne-containing methionine analogs in a methionine-auxotrophic *E. coli* strain.[108] For example, the Davis group incorporated the methionine analog, azidohomoalanine (AHA), and a cysteine mutation to the target protein based on the combination of site-directed gene mutagenesis and the residue-specific replacement of methionine by its analogs (Fig. 9a).[127] Dual modification was accomplished through the initial cysteine conjugation with methanethiosulphonates derivatives, followed by the CuAAC reaction of the azido group in AHA and ethynyl functionalities. This approach offers the benefit of the established AHA incorporation and its efficient modification by cycloaddition reactions. However, the set of available ncAA is limited as they have to bind to their respective tRNA synthetase efficiently, using auxotrophic strains and all the AA residues within a sequence will be replaced by the respective ncAA analogue.

The genetic code expansion technique has been developed as another technique that allows the insertion of a broad variety of ncAAs with spatial precision at virtually any desired site.[126] It is accomplished by using an orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pair, which is capable of charging a designed ncAA in response to a nonsense codon, such as the amber stop codon (UAG), due to their minimal occurrence in most organisms. This strategy allows the genetic encoding of more than 150 ncAAs containing various reactive handles and functionalities.[108] Therefore, the incorporation of ncAA via genetic code expansion in combination with the site-directed canonical AA mutation provides a versatile strategy for protein dual functionalization. Representative work was reported by Deniz and coworkers, in which they utilized an engineered Tyrosyl-tRNA synthetase (MjTyrRS)/tRNA$_{CUA}$ pair (*Mj*TyrRS/tRNA$_{CUA}$ pair) derived from *Methanococcus jannaschi* in response to the amber (UAG) stop codon to insert *p*-acetylphenylalanine, a ketone bearing ncAA, into T4 lysozyme.[128] In combination with a single cysteine mutation, dual labeling of the T4 lysozyme mutant was demonstrated by modification with a FRET dye pair through the thiol-maleimide reaction and oxime ligation (Fig. 9b).[128]

Despite the simplicity of introducing a cysteine mutation as one of the target sites for protein dual modification, this approach could become problematic if the protein consists of native cysteines that play important structural or functional roles. Furthermore, cysteine residues could form disulfide bonds, which could limit expression yields and the reaction reversibility of the thiol-maleimide conjugation could also cause stability problems of the resulting protein bioconjugates.[129] Hence, there have been many efforts to introduce two different ncAAs with bioorthogonal tags that could be functionalized independently. Liu and coworkers applied two mutually orthogonal aaRS/tRNA pairs in response to two blank codons.[130] They mutated the Pyrrolysyl-tRNA synthetase (PylRS)/tRNA$_{CUA}$ pair (PylRS/tRNA$_{CUA}$ pair) to produce a new variant, PylRS/tRNA$_{UUA}$, which can suppress the ochre (UAA) stop codon. In combination with the evolved $Mj$TyrRS/tRNA$_{CUA}$ pair, these two orthogonal aaRS/tRNA pairs were capable of recognizing and inserting two ncAAs, p-azido-L-phenylalanine, and N$^\varepsilon$-propargyloxycarbonyl-L-lysine, into a glutamine-binding protein in response to the amber UAG codon and the ochre UAA codon (Fig. 9c).[130] Two sequential CuAAC reactions were employed to install two chromophores for the protein dual modification. In a subsequent report, the authors incorporated the azide- and ketone-bearing ncAAs into GFP with the use of an evolved $Mj$TyrRS (AzFRS)/tRNA$_{CUA}$ pair and PylRS (AcKRS)/tRNA$_{UUA}$ pair, to eliminate protein aggregation and oxidation induced by the copper catalyst.[131] Notably, dual functionalization was accomplished by SPAAC and an oxime ligation in a one-pot and catalyst-free fashion. In addition, besides in $E.\ coli$, Schultz and coworkers have also successfully incorporated two different ncAAs, which contained azido and ketone groups, in mammalian cells by utilizing the two orthogonal $Methanosarcina\ barkeri$ pyrrolysyl-tRNA synthetase (MbPylRS)/$Methanosarcina\ mazei$ pyrrolysyl tRNA ($Mb$PylRS/$Mm$tRNA$_{UUA}$ pair) and Tyrosyl-tRNA synthetase (EcTyrRS)/tRNA$_{CUA}$ pair ($Ec$TyrRS/tRNA$_{CUA}$ pair).[132] The dual-tagged antibody was subsequently functionalized with a toxic drug payload and a fluorophore with high conversion.

Alternatively, instead of reassignment of the triplet stop codons, Chin and coworkers have exploited the two orthogonal aaRS/tRNA pairs in response to a quadruplet blank codon (four-base codon) and a stop codon for the incorporation of two ncAAs.[133] However, natural ribosomes suffer from very low efficiency in decoding the quadruplet codon. In this context, Chin $et\ al$. have synthetically evolved an orthogonal lysozyme (ribo-Q1), which was not responsible for synthesizing the proteome as natural lysozyme, for selectively decoding the quadruplet codon. By the combination of ribo-Q1 with two orthogonal aaRS/tRNA pairs, AzPheRS*/tRNA$_{UCCU}$ (a derivative of $Mj$TyrRS/tRNA) and PylRS/tRNA$_{CUA}$, two ncAAs were site-selectively introduced into Calmodulin, forming a triazole intramolecular crosslink through the subsequent CuAAC reaction. Nonetheless, the major drawback of this system lies in the low efficiency and specificity of the AzPheRS*/tRNA$_{UCCU}$ pair in directing the corresponding ncAA incorporation. The efficiency of the
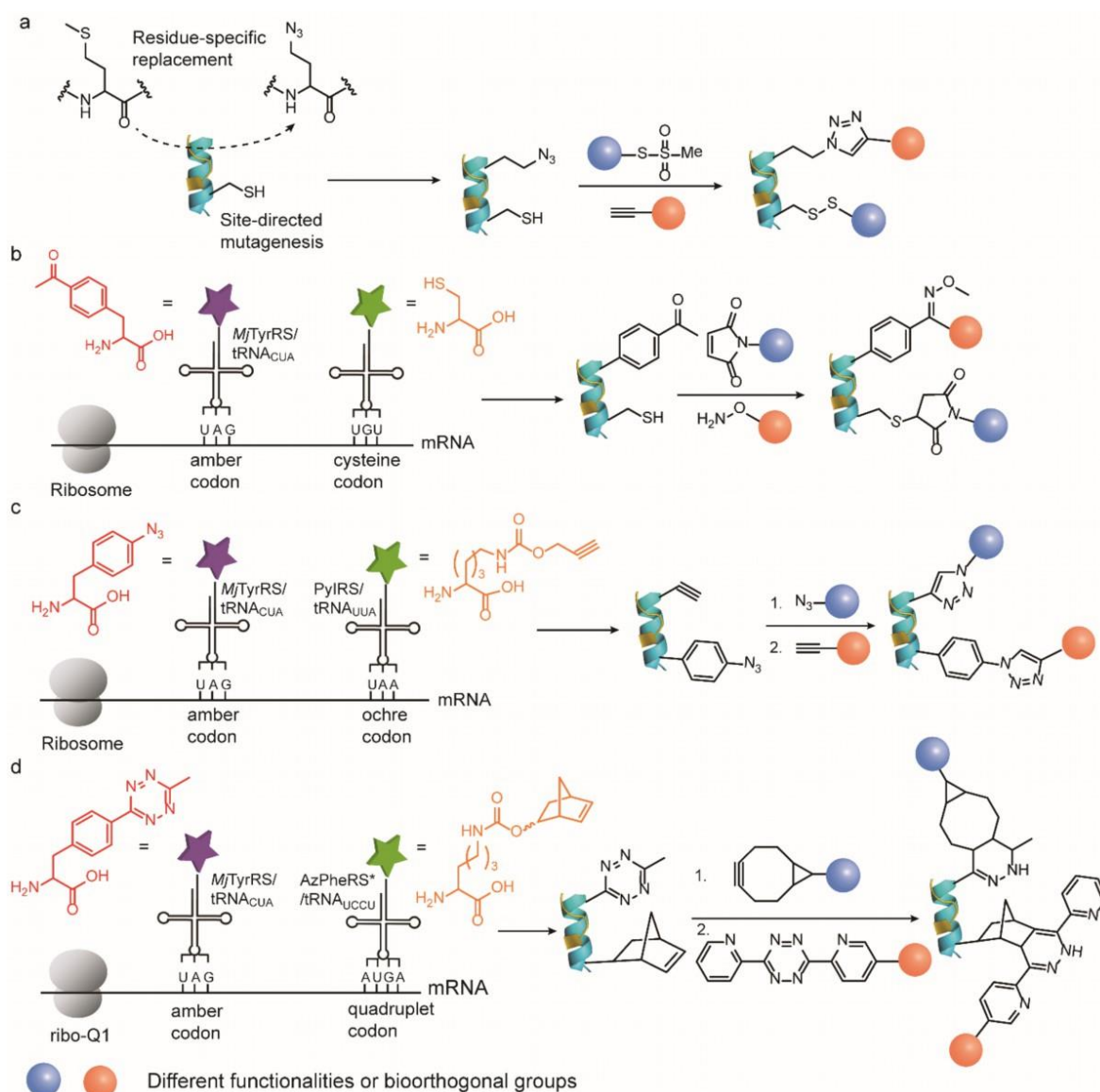
Fig. 1.15 (a) Site-selective incorporation of one cysteine mutation via site-directed mutagenesis and one methionine analogue via residue-specific replacement experiment. (b) Genetic encoding of one cysteine mutation via site-directed mutagenesis and one ncAA via genetic code expansion in response to UAG codon. (c) Genetic encoding of two noncanonical AAs via genetic code expansion in response to UAG and UAA codon. (d) Genetic encoding of two noncanonical AAs via genetic code expansion in response to UAG and AUGA codon. Adapted with permission from reference 20. Copyright (2020), John Wiley and Sons.

original system was substantially improved based on the evolution of the PylRS/tRNA$_{CUA}$ to obtain an optimized quadruplet decoding variant, PylRS/tRNA$_{UACU}$ pair. As a proof-of-concept, the evolved PylRS/tRNA$_{UACU}$ pair was combined with the AzPheRS*/tRNA$_{CUA}$ pair to site-selectively introduce two bioorthogonal tags, norbornene and tetrazine, into Calmodulin (Fig. 9d). Dual modification was successfully accomplished via two sequential iEDDA reactions. With the established technologies, in the latter example, the authors incorporated the alkyne and cyclopropane handles into

Calmodulin by using the orthogonal aaRS-tRNA pair described above, permitting the simultaneous dual functionalization in one pot.[134]

Genetic code expansion has witnessed incredible achievements in recent years, which greatly promotes and revolutionizes the field of site-selective protein dual functionalization. The advantage of this strategy lies in the small size of the ncAAs, flexible incorporation sites but high site-selectivity, and the diverse reactive tags available for incorporation. However, some important challenges remain in this rapidly evolving field, e.g. low catalytic efficiency of engineered aaRS which require tedious evaluation and optimization to improve their performance, relative low expression yields, repetitive optimization of the protocols for effective protein engineering, the scope, and compatibilities of the functional groups capable to be inserted.[108] Besides developing the recombinant engineering techniques, the bioorthogonal chemistry toolbox can also be expanded so that there are more choices for compatible bioorthogonal reaction pairs in terms of reaction rate, catalyst type, and substrate solubility for their intended applications.

**1.5 Higher level of modifications of proteins**

Driven by exciting developments in the discovery of new synthetic procedures as well as a scientific curiosity, achieving a higher level of protein modification in a site-selective manner is considered crucial for advancing different fields in chemistry, biology, and material science. Compared to mono- and dual functionalization of proteins, triple functionalization strategies allow the incorporation of three different functionalities into proteins, thus offering well-defined protein bioconjugates with further expanded structural and functional diversities that are beyond what current synthetic strategies can accomplish. These endeavors will greatly boost our capability to investigate, modulate and re-design the chemical and physical properties of proteins.

However, triple functionalization of proteins is even more demanding than mono- and dual functionalization as the combination of three orthogonal chemistries have to be applied in an aqueous solution and the presence of many reactive proteinogenic groups. Furthermore, by increasing the number of functionalities attached to the protein surface, it may lead to potential detrimental influence on their functional and structural integrity. In addition, the selection of the reactive groups requires a more deliberate chemical design as the increase in the hydrophobicity

of the bioconjugation reagents may cause protein aggregation. The labeling efficiency may also be compromised due to the steric hindrance from the insertion of three different functionalities.

Recently there are some examples demonstrating successful triple functionalization of proteins. Chatterjee and coworkers have reported the introduction of three ncAAs into proteins via genetic code expansion. In this work, they assigned the EcTrp, *Mj*Tyr, and Pyl pairs to suppress UGA, UAG and UAA codons for the incorporation of three ncAAs, 5-hydroxytryptophan *p*-azidophenylalanine, and cyclopropene-lysine, in the engineered *E. coli* strain ATMW1 (Fig. 11a).[135] The triple functionalization of the protein was successfully achieved by three mutually compatible reactions, SPAAC, iEDDA, and Chemoselective Rapid Azo-Coupling Reaction (CRACR), in which the electron-rich 5-hydroxyindole ring in 5-hydroxytryptopan reacts with electron-deficient aromatic diazonium ions with fast reaction kinetics and high conversion[136]. This is the first example of triple modification of proteins where three ncAAs were incorporated into the target protein in living cells, which further expands the simultaneous ncAAs coding capacity via genetic code expansion. However, this strategy utilized three different aaRS/tRNA pairs and three stop codons, which leaves no codon for the termination of endogenous genes and therefore greatly interferes with the translation termination inside cells. Shortly after, Chin and coworkers have demonstrated the genetic encoding of three distinct ncAAs into proteins with three aaRS/tRNA pairs in a different strategy.[137] In their work, they identified new ΔNPylRS/$^{ΔNPyl}$tRNA pairs, which lack the N-terminal domains, and revealed that they can be assigned into two classes, class A and B, based on their sequences. Specifically, class A ΔNPylRS preferentially function with class A $^{ΔNPyl}$tRNA and vice versa. Next, they discovered a *Mm*PylRS/*Spe*$^{Pyl}$tRNA pair, in which *Spe*$^{Pyl}$tRNA is orthogonal to both class A and B ΔNPylRS. In this context, the triply orthogonal aaRS/tRNA pairs were identified, which contained the *Mm*PylRS/*Spe*$^{Pyl}$tRNA pair, an evolved class A ΔNPylRS/$^{ΔNPyl}$tRNA pair and an evolved class B ΔNPylRS/$^{ΔNPyl}$tRNA pair. Three ncAAs, Nε-((tert-butoxy)carbonyl)-l-lysine, 3-methyl-L-histidine and Nε-(carbobenzyloxy)-l-lysine, were genetically encoded into proteins in response to the UAG, AGGA and AGUA codons. However, subsequent functionalizations have not been demonstrated yet. Compared to the reassignment of three stop codons, this work utilizes the combination of a stop codon and quadruplet codons, which do not interfere with the translation termination and serve as a more general and applicable strategy to achieve higher-level protein functionalization. Even though there is no application shown and there are still obvious limitations, e.g. low encoding efficiency, these examples represent the first proof-of-concept for protein triple modifications. These examples still represent significant technological advancements for the development of multifunctional protein conjugates, which holds immense potential for drug delivery, bioimaging, and material science.

Besides increasing the number of functionalities that can be incorporated into proteins, more complex protein bioconjugates have also been reported. For example, Chudasama *et al.* described

an efficient and modular strategy for the generation of bispecific antibodies as well as the dual functionalization of the resulting bioconjugate (Fig. 11b).[138] By fine-tuning the bioorthogonal chemistry employed, two dual-modified antibodies were first conjugated together via the iEDDA reaction of the two bioorthogonal handles from each antibody (Fig. 11b). Subsequently, two other bioorthogonal handles from each antibody allow for the dual functionalization of the chemically constructed bispecific antibodies. Despite these seminal studies, the preparation of protein conjugates with a higher level of modification or structural complexity remains relatively unexplored. We envision that with the technical breakthroughs in methodology, exciting yet unexplored applications will be discovered, which will undoubtedly revolutionize fundamental studies and applications of proteins.
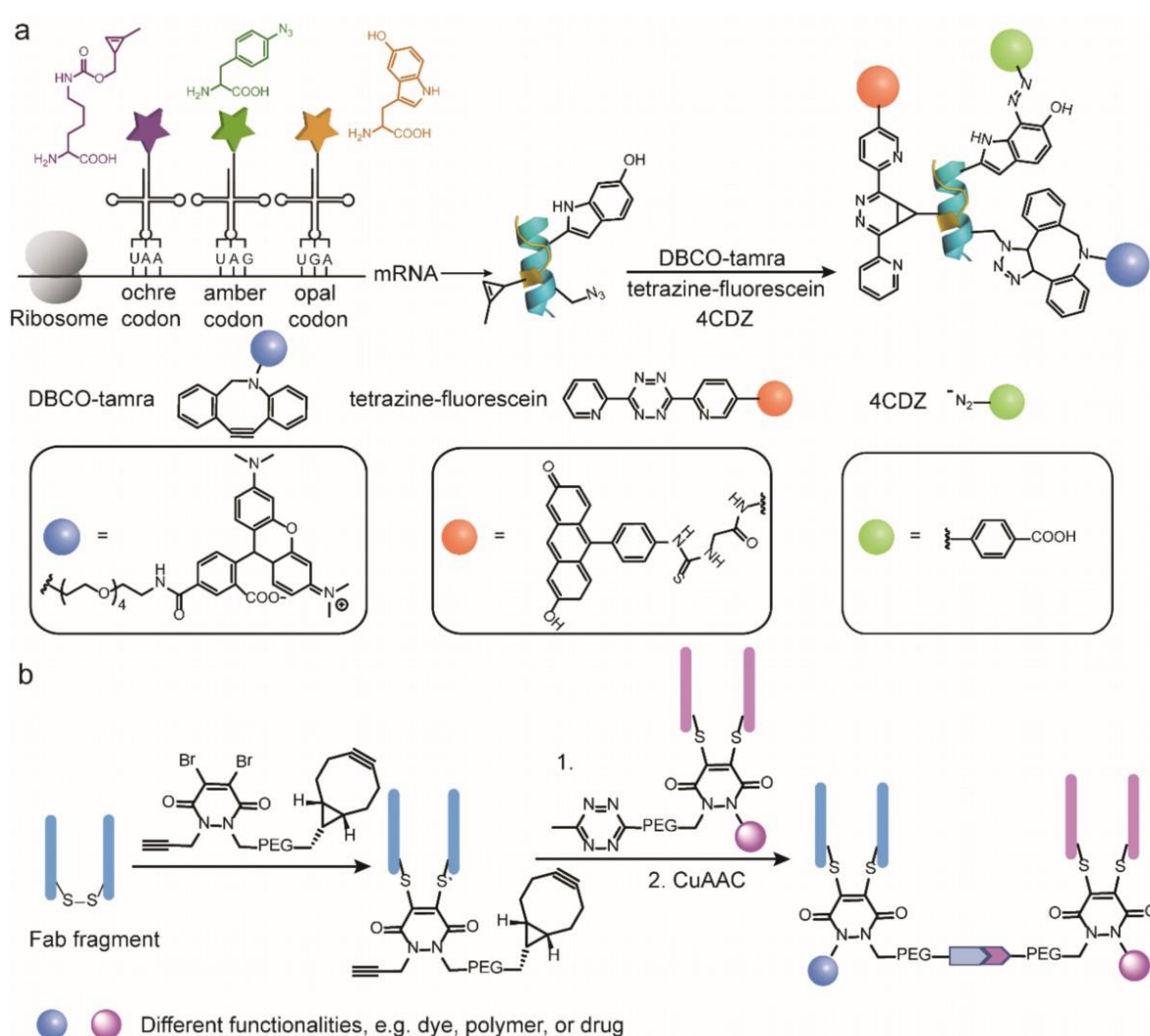


Fig. 1.16 (a) Genetic incorporation of three ncAAc into protein via genetic code expansion in response to the UAA, UAG and UGA codons for tri-functionalization of proteins. (b) Preparation of antibody bispecifics by the conjugation of two dual-modified antibody Fab fragments. Adapted with permission from reference 20. Copyright (2020), John Wiley and Sons.

## 1.6 References

1. Walsh, C. T., Garneau-Tsodikova, S.and Gatto, G. J., Jr., Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **2005**, *44*, 7342-7372.

2. Benjamin Leader, Q. J. B. a. D. E. G., Protein therapeutics: a summary and pharmacological classification. *Nat Rev Drug Discov* **2008**, *7*, 21-39.

3. Carter, P. J., Introduction to current and future protein therapeutics: a protein engineering perspective. *Exp Cell Res* **2011**, *317*, 1261-1269.

4. Jevsevar, S., Kunstelj, M.and Porekar, V. G., PEGylation of therapeutic proteins. *Biotechnol J* **2010**, *5*, 113-128.

5. Hoyt, E. A., Cal, P. M. S. D., Oliveira, B. L.and Bernardes, G. J. L., Contemporary approaches to site-selective protein modification. *Nat. Rev. Chem.* **2019**, *3*, 147-171.

6. Krall, N., da Cruz, F. P., Boutureira, O.and Bernardes, G. J. L., Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **2016**, *8*, 103-113.

7. Shadish, J. A.and DeForest, C. A., Site-Selective Protein Modification: From Functionalized Proteins to Functional Biomaterials. *Matter* **2020**, *2*, 50-77.

8. Witus., L. S.and Francis., M. B., Using Synthetically Modified Proteins to Make New Materials. *Acc. Chem. Res.* **2011**, *44*, 774-783.

9. Dozier, J. K.and Distefano, M. D., Site-Specific PEGylation of Therapeutic Proteins. *Int J Mol Sci* **2015**, *16*, 25831-25864.

10. Vladka Gaberc-Porekar, I. Z., Barbara Podobnik and Viktor Menart, Obstacles and pitfalls in the PEGylation of therapeutic proteins. *Current Opinion in Drug Discovery & Development* **2008**, *11*, 242-250.

11. Chudasama, V., Maruani, A.and Caddick, S., Recent advances in the construction of antibody–drug conjugates. *Nat. Chem.* **2016**, *8*, 114-119.

12. Akkapeddi, P., Azizi, S. A., Freedy, A. M., Cal, P., Gois, P. M. P.and Bernardes, G. J. L., Construction of homogeneous antibody-drug conjugates using site-selective protein chemistry. *Chem. Sci.* **2016**, *7*, 2954-2963.

13. Zolot, R. S., Basu, S.and Million, R. P., Antibody-drug conjugates. *Nat Rev Drug Discov* **2013**, *12*, 259-260.

14. Bargh, J. D., Isidro-Llobet, A., Parker, J. S.and Spring, D. R., Cleavable linkers in antibody-drug conjugates. *Chem. Soc. Rev.* **2019**, *48*, 4361-4374.

15. Zheng, S., Zhang, G., Li, J.and Chen, P. R., Monitoring Endocytic Trafficking of Anthrax Lethal Factor by Precise and Quantitative Protein Labeling. *Angew. Chem. Int. Ed.* **2014**, *53*, 6449-6453.

16. Chemistry in living systems. *Nat. Chem. Biol.* **2005**, *1*, 13-21.

17. Lang, K.and Chin, J. W., Bioorthogonal reactions for labeling proteins. *ACS. Chem. Biol.* **2014**, *9*, 16-20.

18. Stephanopoulos, N.and Francis, M. B., Choosing an effective protein bioconjugation strategy. *Nat.*

*Chem. Biol.* **2011**, *7*, 876-884.

19. Ng, D. Y., Wu, Y., Kuan, S. L.and Weil, T., Programming supramolecular biohybrids as precision therapeutics. *Acc. Chem. Res.* **2014**, *47*, 3471-3480.

20. Spicer, C. D.and Davis, B. G., Selective chemical protein modification. *Nat. Commun.* **2014**, *5*, 4740.

21. Hartmuth C. Kolb, M. G. F. a. K. B. S., Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001** *40*, 2004-2021.

22. Sletten, E. M.and Bertozzi, C. R., Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem. Int. Ed.* **2009**, *48*, 6974-6998.

23. Borrmann, A.and van Hest, J. C. M., Bioorthogonal chemistry in living organisms. *Chem. Sci.* **2014**, *5*, 2123-2134.

24. Devaraj, N. K., The Future of Bioorthogonal Chemistry. *ACS. Cent. Sci.* **2018**, *4*, 952-959.

25. Bertozzi, C. R., A Decade of Bioorthogonal Chemistry. *Acc. Chem. Res.* **2011**, *44*, 651-653.

26. Chen, X.and Wu, Y. W., Selective chemical labeling of proteins. *Org. Biomol. Chem.* **2016**, *14*, 5417-39.

27. Huisgen, R., Kinetics and Mehanism of 1,3-Dipolar Cycloadditions. *Angew. Chem. Int. Ed.* **1963**, *2*, 633-645.

28. Rostovtsev., V. V., Green., L. G., Fokin., V. V.and Sharpless., K. B., A Stepwise Huisgen Cycloaddition Process: Copper-Catalyzed Regioselective Ligation of Azides and Terminal Alkynes. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596-2599.

29. Christian W. Tornøe, C. C., and Morten Meldal, Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* **2002**, *67*, 3057-3064.

30. David C. Kennedy, R. K. L., and John Paul Pezacki, Cellular Lipid Metabolism Is Influenced by the Coordination Environment of Copper. *J. Am. Chem. Soc.* **2009**, *131*, 2444-2445.

31. Nicholas J. Agard, J. A. P., and Carolyn R. Bertozzi, A Strain-Promoted [3 + 2] Azide-Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **2004**, *126*, 15046-15047.

32. Julian A. Codelli, J. M. B., Nicholas J. Agard, and Carolyn R. Bertozzi, Second-Generation Difluorinated Cyclooctynes for Copper-Free Click Chemistry. *J. Am. Chem. Soc.* **2008**, *130*, 11486-11493.

33. Dommerholt, J., Schmidt, S., Temming, R., Hendriks, L. J., Rutjes, F. P., van Hest, J. C., Lefeber, D. J., Friedl, P.and van Delft, F. L., Readily accessible bicyclononynes for bioorthogonal labeling and three-dimensional imaging of living cells. *Angew. Chem. Int. Ed.* **2010**, *49*, 9422-9425.

34. Ning, X., Guo, J., Wolfert, M. A.and Boons, G.-J., Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-Free and Fast Huisgen Cycloadditions. *Angew. Chem. Int. Ed.* **2008**, *47*, 2253-2255.

35. Debets, M. F., van Berkel, S. S., Schoffelen, S., Rutjes, F. P., van Hest, J. C.and van Delft, F. L., Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition.

*Chem. Commun.* **2010**, *46*, 97-9.

36. Baskin, J. M., Prescher, J. A., Laughlin, S. T., Agard, N. J., Chang, P. V., Miller, I. A., Lo, A., Codelli, J. A.and Bertozzi, C. R., Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci.* **2007**, *104*, 16793-16797.

37. Scott T. Laughlin, J. M. B., Sharon L. Amacher and Carolyn R. Bertozzi, In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science.* **2008**, *320*, 664-667.

38. Kim, E.and Koo, H., Biomedical applications of copper-free click chemistry: in vitro, in vivo, and ex vivo. *Chem. Sci.* **2019,** 7835-7851.

39. Ning, X., Temming, R. P., Dommerholt, J., Guo, J., Ania, D. B., Debets, M. F., Wolfert, M. A., Boons, G. J.and van Delft, F. L., Protein modification by strain-promoted alkyne-nitrone cycloaddition. *Angew. Chem. Int. Ed.* **2010**, *49*, 3065-3068.

40. Lim, R. K.and Lin, Q., Bioorthogonal chemistry: recent progress and future directions. *Chem. Commun.* **2010**, *46*, 1589-600.

41. Bertozzi, E. S. a. C. R., Cell Surface Engineering by a Modified Staudinger Reaction. *Science.* **2000**, 2007-2010.

42. Eliana Saxon, J. I. A., and Carolyn R. Bertozzi, A "Traceless" Staudinger Ligation for the Chemoselective Synthesis of Amide Bonds. *Org. Lett.* **2000**, *2*, 2141-2143.

43. Patterson, D. M., Nazarova, L. A.and Prescher, J. A., Finding the Right (Bioorthogonal) Chemistry. *ACS Chem. Biol.* **2014**, *9*, 592-605.

44. Song, W., Wang, Y., Qu, J., Madden, M. M.and Lin, Q., A Photoinducible 1,3-Dipolar Cycloaddition Reaction for Rapid, Selective Modification of Tetrazole-Containing Proteins. *Angew. Chem. Int. Ed.* **2008**, *120*, 2874-2877.

45. Oliveira, B. L., Guo, Z.and Bernardes, G. J. L., Inverse electron demand Diels–Alder reactions in chemical biology. *Chem. Soc. Rev.* **2017**, *46*, 4895-4950.

46. Meimetis, L. G., Carlson, J. C., Giedt, R. J., Kohler, R. H.and Weissleder, R., Ultrafluorogenic coumarin-tetrazine probes for real-time biological imaging. *Angew. Chem. Int. Ed.* **2014**, *53*, 7531-7534.

47. Carlson, J. C., Meimetis, L. G., Hilderbrand, S. A.and Weissleder, R., BODIPY-tetrazine derivatives as superbright bioorthogonal turn-on probes. *Angew. Chem. Int. Ed.* **2013**, *52*, 6917-6920.

48. Devaraj, N. K., Hilderbrand, S., Upadhyay, R., Mazitschek, R.and Weissleder, R., Bioorthogonal turn-on probes for imaging small molecules inside living cells. *Angew. Chem. Int. Ed.* **2010**, *49*, 2869-2872.

49. Darko, A., Wallace, S., Dmitrenko, O., Machovina, M. M., Mehl, R. A., Chin, J. W.and Fox, J. M., Conformationally Strained trans-Cyclooctene with Improved Stability and Excellent Reactivity in Tetrazine Ligation. *Chem. Sci.* **2014**, *5*, 3770-3776.

50. Chen, W., Wang, D., Dai, C., Hamelberg, D.and Wang, B., Clicking 1,2,4,5-tetrazine and cyclooctynes with tunable reaction rates. *Chem. Commun.* **2012**, *48*, 1736-1738.

51. Rossin, R., van den Bosch, S. M., ten Hoeve, W., Carvelli, M., Versteegen, R. M., Lub, J.and Robillard, M. S., Highly Reactive trans-Cyclooctene Tags with Improved Stability for Diels–Alder

Chemistry in Living Systems. *Bioconjugate Chem.* **2013**, *24*, 1210-1217.

52. Reddy, N. C., Kumar, M., Molla, R.and Rai, V., Chemical methods for modification of proteins. *Org. Biomol. Chem.* **2020**, *18*, 4669-4691.

53. Rawale, D. G., Thakur, K., Adusumalli, S. R.and Rai, V., Chemical Methods for Selective Labeling of Proteins. *Eur. J. Org. Chem.* **2019**, *2019*, 6749-6763.

54. Gunnoo, S. B.and Madder, A., Chemical Protein Modification through Cysteine. *ChemBioChem* **2016**, *17*, 529-553.

55. Ochtrop, P.and Hackenberger, C. P. R., Recent advances of thiol-selective bioconjugation reactions. *Curr. Opin. Chem. Biol.* **2020**, *58*, 28-36.

56. Northrop, B. H., Frayne, S. H.and Choudhary, U., Thiol–maleimide "click" chemistry: evaluating the influence of solvent, initiator, and thiol on the reaction mechanism, kinetics, and selectivity. *Polym. Chem.* **2015**, *6*, 3415-3430.

57. Fontaine, S. D., Reid, R., Robinson, L., Ashley, G. W.and Santi, D. V., Long-term stabilization of maleimide-thiol conjugates. *Bioconjugate Chem.* **2015**, *26*, 145-152.

58. Christie, R. J., Fleming, R., Bezabeh, B., Woods, R., Mao, S., Harper, J., Joseph, A., Wang, Q., Xu, Z. Q., Wu, H., Gao, C.and Dimasi, N., Stabilization of cysteine-linked antibody drug conjugates with N-aryl maleimides. *J Control Release* **2015**, *220*, 660-670.

59. Kalia, D., Malekar, P. V.and Parthasarathy, M., Exocyclic Olefinic Maleimides: Synthesis and Application for Stable and Thiol-Selective Bioconjugation. *Angew. Chem. Int. Ed.* **2016**, *55*, 1432-1435.

60. Zhang, Y., Zhou, X., Xie, Y., Greenberg, M. M., Xi, Z.and Zhou, C., Thiol Specific and Tracelessly Removable Bioconjugation via Michael Addition to 5-Methylene Pyrrolones. *J. Am. Chem. Soc.* **2017**, 139, 6146-6151.

61. Smith, M. E. B., Schumacher, F. F., Ryan, C. P., Tedaldi, L. M., Papaioannou, D., Waksman, G., Caddick, S.and Baker, J. R., Protein Modification, Bioconjugation, and Disulfide Bridging Using Bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132*, 1960-1965.

62. Nathani, R. I., Chudasama, V., Ryan, C. P., Moody, P. R., Morgan, R. E., Fitzmaurice, R. J., Smith, M. E., Baker, J. R.and Caddick, S., Reversible protein affinity-labelling using bromomaleimide-based reagents. *Org. Biomol. Chem.* **2013**, *11*, 2408-2411.

63. Ryan, C. P., Smith, M. E., Schumacher, F. F., Grohmann, D., Papaioannou, D., Waksman, G., Werner, F., Baker, J. R.and Caddick, S., Tunable reagents for multi-functional bioconjugation: reversible or permanent chemical modification of proteins and peptides by control of maleimide hydrolysis. *Chem. Commun.* **2011**, *47*, 5452-5454.

64. Huang, R., Li, Z., Sheng, Y., Yu, J., Wu, Y., Zhan, Y., Chen, H.and Jiang, B., N-Methyl- N-phenylvinylsulfonamides for Cysteine-Selective Conjugation. *Org. Lett.* **2018**, *20*, 6526-6529.

65. Gil de Montes, E., Jiménez-Moreno, E., Oliveira, B., Navo, C. D., Cal, P. M. S. D., Jiménez-Osés, G., Robina, I., Moreno-Vargas, A. J.and Bernardes, G. J. L., Azabicyclic Vinyl Sulfones for Residue-specific Dual Protein Labelling. *Chem. Sci.* **2019**, *10*, 4515-4522.

66. Gil de Montes, E., Istrate, A., Navo, C., Jimenez-Moreno, E., Hoyt, E., Corzana, F., Robina, I.,

Jimenez-Oses, G., Moreno-Vargas, A.and Bernardes, G. J. L., Stable pyrrole-linked bioconjugates through tetrazine-triggered azanorbornadiene fragmentation. *Angew. Chem. Int. Ed.* **2020**, 59, 6196-6200.

67. Bernardim, B., Cal, P. M., Matos, M. J., Oliveira, B. L., Martinez-Saez, N., Albuquerque, I. S., Perkins, E., Corzana, F., Burtoloso, A. C., Jimenez-Oses, G.and Bernardes, G. J., Stoichiometric and irreversible cysteine-selective protein modification using carbonylacrylic reagents. *Nat. Commun.* **2016**, *7*, 13128.

68. Boutureira, O., Martinez-Saez, N., Brindle, K. M., Neves, A. A., Corzana, F.and Bernardes, G. J. L., Site-Selective Modification of Proteins with Oxetanes. *Chem. Eur. J.* **2017**, *23*, 6483-6489.

69. Matos, M. J., Navo, C. D., Hakala, T., Ferhati, X., Guerreiro, A., Hartmann, D., Bernardim, B., Saar, K. L., Companon, I., Corzana, F., Knowles, T. P. J., Jimenez-Oses, G.and Bernardes, G. J. L., Quaternization of Vinyl/Alkynyl Pyridine Enables Ultrafast Cysteine-Selective Protein Modification and Charge Modulation. *Angew. Chem. Int. Ed.* **2019**, *58*, 6640-6644.

70. Kasper, M. A., Stengl, A., Ochtrop, P., Gerlach, M., Stoschek, T., Schumacher, D., Helma, J., Penkert, M., Krause, E., Leonhardt, H.and Hackenberger, C. P. R., Ethynylphosphonamidates for the Rapid and Cysteine-Selective Generation of Efficacious Antibody-Drug Conjugates. *Angew. Chem. Int. Ed.* **2019**, *58*, 11631-11636.

71. Baumann, A. L., Schwagerus, S., Broi, K., Kemnitz-Hassanin, K., Stieger, C. E., Trieloff, N., Schmieder, P.and Hackenberger, C. P. R., Chemically Induced Vinylphosphonothiolate Electrophiles for Thiol-Thiol Bioconjugations. *J. Am. Chem. Soc.* **2020**, *142*, 9544-9552.

72. Luo, Q., Tao, Y., Sheng, W., Lu, J.and Wang, H., Dinitroimidazoles as bifunctional bioconjugation reagents for protein functionalization and peptide macrocyclization. *Nat. Commun.* **2019**, *10*, 142.

73. Hoyle, C. E.and Bowman, C. N., Thiol-ene click chemistry. *Angew. Chem. Int. Ed.* **2010**, *49*, 1540-1573.

74. Lowe, A. B., Thiol-yne 'click'/coupling chemistry and recent applications in polymer and materials synthesis and modification. *Polymer,* **2014**, *55*, 5517-5549.

75. Kuan, S. L., Wang, T.and Weil, T., Site-Selective Disulfide Modification of Proteins: Expanding Diversity beyond the Proteome. *Chem. Eur. J.* **2016**, *22*, 17112-17129.

76. Sibu Balan, J.-w. C., Antony Godwin, Ian Teo, Carlos M. Laborde, Sibylle Heidelberger, Mire Zloh, Sunil Shaunak, and Steve Brocchini, Site-Specific PEGylation of Protein Disulfide Bonds Using a Three-Carbon Bridge. *Bioconjugate Chem.* **2007**, *18*, 61-76.

77. Shaunak, S., Godwin, A., Choi, J.-W., Balan, S., Pedone, E., Vijayarangam, D., Heidelberger, S., Teo, I., Zloh, M.and Brocchini, S., Site-specific PEGylation of native disulfide bonds in therapeutic proteins. *Nat. Chem. Biol.* **2006**, *2*, 312-313.

78. Badescu, G., Bryant, P., Bird, M., Henseleit, K., Swierkosz, J., Parekh, V., Tommasi, R., Pawlisz, E., Jurlewicz, K., Farys, M., Camper, N., Sheng, X., Fisher, M., Grygorash, R., Kyle, A., Abhilash, A., Frigerio, M., Edwards, J.and Godwin, A., Bridging disulfides for stable and defined antibody drug conjugates. *Bioconjugate Chem.* **2014**, *25*, 1124-36.

79. Wang, T., Pfisterer, A., Kuan, S. L., Wu, Y., Dumele, O., Lamla, M., Müllen, K.and Weil, T., Cross-conjugation of DNA, proteins and peptides via a pH switch. *Chem. Sci.* **2013**, *4*, 1889-1894.

80. Wang, T., Riegger, A., Lamla, M., Wiese, S., Oeckl, P., Otto, M., Wu, Y., Fischer, S., Barth, H., Kuan, S. L.and Weil, T., Water-soluble allyl sulfones for dual site-specific labelling of proteins and cyclic peptides. *Chem. Sci.* **2016**, *7*, 3234-3239.

81. Schumacher, F. F., Nobles, M., Ryan, C. P., Smith, M. E., Tinker, A., Caddick, S.and Baker, J. R., In situ maleimide bridging of disulfides and a new approach to protein PEGylation. *Bioconjugate Chem.* **2011**, *22*, 132-136.

82. Chudasama, V., Smith, M. E., Schumacher, F. F., Papaioannou, D., Waksman, G., Baker, J. R.and Caddick, S., Bromopyridazinedione-mediated protein and peptide bioconjugation. *Chem. Commun.* **2011**, *47*, 8781-8783.

83. Maruani, A., Smith, M. E., Miranda, E., Chester, K. A., Chudasama, V.and Caddick, S., A plug-and-play approach to antibody-based therapeutics via a chemoselective dual click strategy. *Nat. Commun.* **2015**, *6*, 6645.

84. Martinez-Saez, N., Sun, S., Oldrini, D., Sormanni, P., Boutureira, O., Carboni, F., Companon, I., Deery, M. J., Vendruscolo, M., Corzana, F., Adamo, R.and Bernardes, G. J. L., Oxetane Grafts Installed Site-Selectively on Native Disulfides to Enhance Protein Stability and Activity In Vivo. *Angew. Chem. Int. Ed.* **2017**, *56*, 14963-14967.

85. Walsh, S. J., Omarjee, S., Galloway, W., Kwan, T. T., Sore, H. F., Parker, J. S., Hyvonen, M., Carroll, J. S.and Spring, D. R., A general approach for the site-selective modification of native proteins, enabling the generation of stable and functional antibody-drug conjugates. *Chem. Sci.* **2019**, *10*, 694-700.

86. Rosen, C. B.and Francis, M. B., Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol.* **2017**, *13*, 697-705.

87. Philip E. Dawson, T. W. M., Ian Clark-Lewis and Stephen B. H. Kent, Synthesis of Proteins by Native Chemical Ligation. *Science.* **1994**, *266*, 776-779.

88. Busch, G. K., Tate, E. W., Gaffney, P. R., Rosivatz, E., Woscholski, R.and Leatherbarrow, R. J., Specific N-terminal protein labelling: use of FMDV 3C pro protease and native chemical ligation. *Chem. Commun.* **2008**, 3369-3371.

89. Bandyopadhyay, A., Cambray, S.and Gao, J., Fast and selective labeling of N-terminal cysteines at neutral pH via thiazolidino boronate formation. *Chem. Sci.* **2016**, *7*, 4589-4593.

90. Faustino, H., Silva, M., Veiros, L. F., Bernardes, G. J. L.and Gois, P. M. P., Iminoboronates are efficient intermediates for selective, rapid and reversible N-terminal cysteine functionalisation. *Chem. Sci.* **2016**, *7*, 5052-5058.

91. Ren, H., Xiao, F., Zhan, K., Kim, Y. P., Xie, H., Xia, Z.and Rao, J., A biocompatible condensation reaction for the labeling of terminal cysteine residues on proteins. *Angew. Chem. Int. Ed.* **2009**, *48*, 9658-9662.

92. Xianfeng Li, L. Z., Steven E. Hall, and James P. Tam, A new ligation method for N-terminal tryptophan-containing peptides using the Pictet-Spengler reaction. *Tetrahedron Letters.* **2000**, *41*, 4069-

4073.

93. Gilmore, J. M., Scheck, R. A., Esser-Kahn, A. P., Joshi, N. S.and Francis, M. B., N-terminal protein modification through a biomimetic transamination reaction. *Angew. Chem. Int. Ed.* **2006**, *45*, 5307-5311.

94. Witus, L. S., Netirojjanakul, C., Palla, K. S., Muehl, E. M., Weng, C. H., Iavarone, A. T.and Francis, M. B., Site-specific protein transamination using N-methylpyridinium-4-carboxaldehyde. *J. Am. Chem. Soc.* **2013**, *135*, 17223-17229.

95. Williamson, D. J., Fascione, M. A., Webb, M. E.and Turnbull, W. B., Efficient N-terminal labeling of proteins by use of sortase. *Angew. Chem. Int. Ed.* **2012**, *51*, 9377-9380.

96. Schoonen, L., Pille, J., Borrmann, A., Nolte, R. J.and van Hest, J. C., Sortase A-Mediated N-Terminal Modification of Cowpea Chlorotic Mottle Virus for Highly Efficient Cargo Loading. *Bioconjugate Chem.* **2015**, *26*, 2429-2434.

97. Giang K. T. Nguyen, Y. C., Wei Wang, Chuan Fa Liu, and James P. Tam, Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide. *Angew. Chem. Int. Ed.* **2015**, *127*, 15920-15924.

98. Shixian Lin, X. Y., Shang Jia, AmyM.Weeks, Michael Hornsby, Peter S. Lee, Rita V. Nichiporuk, Anthony T. Iavarone, James A. Wells, F. Dean Toste and Christopher J. Chang, Redox-based reagents for chemoselective methionine bioconjugation. *Science.* **2017**, *355*, 597-602.

99. Taylor, M. T., Nelson, J. E., Suero, M. G.and Gaunt, M. J., A protein functionalization platform based on selective reactions at methionine residues. *Nature.* **2018**, *562*, 563-568.

100. John M. Antos, J. M. M., Anthony T. Iavarone, and Matthew B. Francis, Chemoselective Tryptophan Labeling with Rhodium Carbenoids at Mild pH. *J. Am. Chem. Soc.* **2009**, *131*, 6301-6308.

101. Hitoshi Ban, J. G., and Carlos F. Barbas, III, Tyrosine Bioconjugation through Aqueous Ene-Type Reactions: A Click-Like Reaction for Tyrosine. *J. Am. Chem. Soc.* **2010**, *132*, 1523-1525.

102. Gavrilyuk, J., Ban, H., Nagano, M., Hakamata, W.and Barbas, C. F., 3rd, Formylbenzene diazonium hexafluorophosphate reagent for tyrosine-selective modification of proteins and the introduction of a bioorthogonal aldehyde. *Bioconjugate Chem.* **2012**, *23*, 2321-2328.

103. Neel S. Joshi, L. R. W., and Matthew B. Francis, A Three-Component Mannich-Type Reaction for Selective Tyrosine Bioconjugation. *J. Am. Chem. Soc.* **2004**, *126*, 15942-15943.

104. STADTMAN, K. U. A. E. R., Modification of histidine residues in proteins by reaction with 4-hydroxynonenal. *Proc. Natl. Acad. Sci.* **1992**, *89*, 4544-4548.

105. Joshi, P. N.and Rai, V., Single-site labeling of histidine in proteins, on-demand reversibility, and traceless metal-free protein purification. *Chem. Commun.* **2019**, *55*, 1100-1103.

106. Wang, L.and Schultz, P. G., Expanding the genetic code. *Angew. Chem. Int. Ed.* **2004**, *44*, 34-66.

107. Davis, L.and Chin, J. W., Designer proteins: applications of genetic code expansion in cell biology. *Nat Rev Mol Cell Biol.* **2012**, *13*, 168-82.

108. Lang, K.and Chin, J. W., Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* **2014**, *114*, 4764-4806.

109. Tirrell, J. T. N. a. D. A., Noncanonical Amino Acids in the Interrogation of Cellular Protein Synthesis.

*Acc. Chem. Res.* **2011**, *44*, 677-685.

110. Neumann, H., Peak-Chew, S. Y.and Chin, J. W., Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **2008**, *4*, 232-234.

111. Wang, Q., Parrish, A. R.and Wang, L., Expanding the genetic code for biological studies. *Chem. Biol.* **2009**, *16*, 323-336.

112. Crawley, N., Thompson, M.and Romaschin, A., Theranostics in the growing field of personalized medicine: an analytical chemistry perspective. *Anal. Chem.* **2014**, *86*, 130-160.

113. Kuan, S. L., Fischer, S., Hafner, S., Wang, T., Syrovets, T., Liu, W., Tokura, Y., Ng, D. Y. W., Riegger, A., Förtsch, C., Jäger, D., Barth, T. F. E., Simmet, T., Barth, H.and Weil, T., Boosting Antitumor Drug Efficacy with Chemically Engineered Multidomain Proteins. *Adv. Sci.* **2018**, 1701036.

114. Maruani, A., Richards, D. A.and Chudasama, V., Dual modification of biomolecules. *Org. Biomol. Chem.* **2016**, *14*, 6165-6178.

115. Adumeau, P., Vivier, D., Sharma, S. K., Wang, J., Zhang, T., Chen, A., Agnew, B. J.and Zeglis, B. M., Site-Specifically Labeled Antibody–Drug Conjugate for Simultaneous Therapy and ImmunoPET. *Mol. Pharmaceutics.* **2018**, *15*, 892-898.

116. Levengood., M. R., Zhang., X., Hunter., J. H., Emmerton., K. K., Miyamoto., J. B., Lewis., T. S.and Senter., P. D., Orthogonal CysteineProtection Enables Homogeneous Multi-DrugAntibody–Drug conjugates. *Angew. Chem. Int. Ed.* **2017**, *56*, 733-737.

117. Crochet, A. P., Kabir, M. M., Francis, M. B.and Paavola, C. D., Site-selective dual modification of periplasmic binding proteins for sensing applications. *Biosens. Bioelectron.* **2010**, *26*, 55-61.

118. Agrawalla, B. K., Wang, T., Riegger, A., Domogalla, M. P., Steinbrink, K., Dorfler, T., Chen, X., Boldt, F., Lamla, M., Michaelis, J., Kuan, S. L.and Weil, T., Chemoselective Dual Labeling of Native and Recombinant Proteins. *Bioconjugate Chem.* **2018**, *29*, 29-34.

119. Zhang, Y., Zang, C., An, G., Shang, M., Cui, Z., Chen, G., Xi, Z.and Zhou, C., Cysteine-specific protein multi-functionalization and disulfide bridging using 3-bromo-5-methylene pyrrolones. *Nat. Commun.* **2020**, *11*, 1015.

120. Morales-Sanfrutos, J., Lopez-Jaramillo, F. J., Hernandez-Mateo, F.and Santoyo-Gonzalez, F., Vinyl sulfone bifunctional tag reagents for single-point modification of proteins. *J. Org. Chem.* **2010**, *75*, 4039-4047.

121. Canovas, C., Moreau, M., Bernhard, C., Oudot, A., Guillemin, M., Denat, F.and Goncalves, V., Site-Specific Dual Labeling of Proteins on Cysteine Residues with Chlorotetrazines. *Angew. Chem. Int. Ed.* **2018**, *130*, 10806-10810.

122. Tessier, R., Ceballos, J., Guidotti, N., Simonet-Davin, R., Fierz, B.and Waser, J., "Doubly Orthogonal" Labeling of Peptides and Proteins. *Chem.* **2019**, *5*, 2243-2263.

123. Moody, P., Chudasama, V., Nathani, R. I., Maruani, A., Martin, S., Smith, M. E.and Caddick, S., A rapid, site-selective and efficient route to the dual modification of DARPins. *Chem. Commun.* **2014**, *50*, 4898-4900.

124. Nathani, R. I., Moody, P., Chudasama, V., Smith, M. E., Fitzmaurice, R. J.and Caddick, S., A novel

approach to the site-selective dual labelling of a protein via chemoselective cysteine modification. *Chem. Sci.* **2013**, *4*, 3455-3458.

125. Zhang, C., Welborn, M., Zhu, T., Yang, N. J., Santos, M. S., Van Voorhis, T.and Pentelute, B. L., Pi-Clamp-mediated cysteine conjugation. *Nat. Chem.* **2016**, *8*, 120-128.

126. Chin, J. W., Expanding and reprogramming the genetic code. *Nature.* **2017**, *550*, 53-60.

127. van Kasteren, S. I., Kramer, H. B., Jensen, H. H., Campbell, S. J., Kirkpatrick, J., Oldham, N. J., Anthony, D. C.and Davis, B. G., Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature.* **2007**, *446*, 1105-1109.

128. Eric M. Brustad, E. A. L., Peter G. Schultz, and Ashok A. Deniz, A General and Efficient Method for the Site-Specific Dual-Labeling of Proteins for Single Molecule Fluorescence Resonance Energy Transfer. *J. Am. Chem. Soc.* **2008**, *130*, 17664-17665.

129. Ravasco, J., Faustino, H., Trindade, A.and Gois, P. M. P., Bioconjugation with Maleimides: A Useful Tool for Chemical Biology. *Chem. Eur. J.* **2019**, *25*, 43-59.

130. Wan, W., Huang, Y., Wang, Z., Russell, W. K., Pai, P. J., Russell, D. H.and Liu, W. R., A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angew. Chem. Int. Ed.* **2010**, *49*, 3211-3214.

131. Wu, B., Wang, Z., Huang, Y.and Liu, W. R., Catalyst-free and site-specific one-pot dual-labeling of a protein directed by two genetically incorporated noncanonical amino acids. *ChemBioChem.* **2012**, *13*, 1405-1408.

132. Xiao, H., Chatterjee, A., Choi, S. H., Bajjuri, K. M., Sinha, S. C.and Schultz, P. G., Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. *Angew. Chem. Int. Ed.* **2013**, *52*, 14080-14083.

133. Wang, K., Schmied, W. H.and Chin, J. W., Reprogramming the genetic code: from triplet to quadruplet codes. *Angew. Chem. Int. Ed.* **2012**, *51*, 2288-2297.

134. Sachdeva, A., Wang, K., Elliott, T.and Chin, J. W., Concerted, rapid, quantitative, and site-specific dual labeling of proteins. *J. Am. Chem. Soc.* **2014**, *136*, 7785-7788.

135. Italia, J. S., Addy, P. S., Erickson, S. B., Peeler, J. C., Weerapana, E.and Chatterjee, A., Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites. *J. Am. Chem. Soc.* **2019**, *141*, 6204-6212.

136. Addy, P. S., Erickson, S. B., Italia, J. S.and Chatterjee, A., A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. *J. Am. Chem. Soc.* **2017**, *139*, 11670-11673.

137. Dunkelmann, D. L., Willis, J. C. W., Beattie, A. T.and Chin, J. W., Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nat. Chem.* **2020**, *12*, 535-544.

138. Maruani, A., Szijj, P. A., Bahou, C., Nogueira, J. C. F., Caddick, S., Baker, J. R.and Chudasama, V., A Plug-and-Play Approach for the De Novo Generation of Dually Functionalized Bispecifics. *Bioconjugate Chem.* **2020**, *31*, 520-529.

## 2. Motivation and conceptual design

Proteins are an emerging class of biotherapeutics with high target affinity and specificity, as exemplified by monoclonal antibodies that are applied for treating several severe diseases such as multiple sclerosis or cancer. In order to improve their properties as therapeutics, synthetics functionalities, such as polymers, drugs, or imaging reagents, need to be incorporated to improve their efficacy, delivery, safety as well as reduce the immunogenicity for the in vivo applications. In this context, site-selective protein modification stands out, as it enables to selectively install desired functionalities on the protein surface at pre-defined sites with exquisite chemo- and regio-selectivity, providing easy access for homogeneous, well-defined protein conjugates with fully preserved structure and function integrity.

Among the 20 natural AAs, cysteine and disulfide bonds are still among the most popular targets for modification as they have a low abundance on the protein surface and thiol groups are good nucleophiles. In addition, a lot of therapeutic peptides and proteins contain solvent-accessible cysteines or disulfide bonds. Therefore, developing novel chemical methods to achieve modification at cysteine or disulfide sites is still highly desirable, which holds great potential for advancements in therapeutics, diagnostics, and fundamental science. Even though significant progress has been made in the development of new methodologies for protein modification at cysteine or disulfide sites, many of these strategies are still limited by complicated linker synthesis, poor water-solubility of the conjugation reagents, low labeling efficiency, and the instability of the resultant conjugates. These drawbacks in turn greatly restrict their scalability and general accessibility for further applications. On the other hand, the methods developed so far are mainly designed to target a single specific residue, for example cysteine site. There are rare examples of bioconjugation strategies that modify cysteine or disulfide sites independently in a user-defined fashion by chemical design. Therefore, the development of such bioconjugation strategies, in which the methodologies are simple but highly efficient, offering the formation of stable bioconjugates and, in addition, can be tailored to achieve cysteine or disulfide modification on demand would be highly advantageous to enrich the existing bioconjugation toolbox.

In this regard, this thesis is focused on the development of novel bioconjugation chemistry by chemical design for site-selective protein modification with high efficiency and specificity. First, an inverse electron demand Diels-Alder reaction (iEEDA) between tetrazine and trans-cyclooctene (TCO) was employed for protein modification at the disulfide site in a selective manner. Owing to the characteristic features of iEDDA reactions, homogeneous protein bioconjugates can be obtained in a straightforward fashion by conjugating tetrazine modified proteins and TCO-modified synthetic moieties within minutes in a quantitative conversion without any need for further purification. In this way, the desired bioconjugates could be applied to the patients immediately

after the preparation "on site", e.g. in hospitals, thus greatly eliminating the risk of long-term storage, back-to-batch variations, and loss of bioactivity of the final conjugates.

However, the allyl sulfone disulfide rebridging reagents used in the first work still suffer from drawbacks in terms of multi-step linker synthesis, low modification efficiency, and high hydrophobicity. In this context, the second part of this thesis is focused on the development of novel strategies for site-selective protein modification with improved efficiency. Chloromethyl acrylate and acrylamide reagents, which bear the same chloromethyl acyl scaffold but different linkages (ester or amide bonds), can be easily synthesized in a one-pot reaction. According to their intrinsic reactivity landscape dominated by the ester or amide linkage, chloromethyl acrylamide reagents allow selective modification of cysteines with similar efficiency to maleimide conjugation but showing better stability of the resultant bioconjugates. In contrast, chloromethyl acrylate reagents can be used for efficient disulfide modification, which bears the advantage of facile linker synthesis, high modification efficiency, and low hydrophobicity. In this work, the reaction activity can be finely tuned by synthetic customization to achieve cysteine or disulfide modification on demand, which serves as valuable additions to the current bioconjugation toolbox. In addition, this new approach offers the possibility for dual modification of proteins by capitalizing on the reactivity difference of the 2-chloromethyl acrylamide and acrylate compounds. In this way, we could envision protein dual functionalization at cysteine residues and disulfide bonds can be achieved in a stepwise fashion within one system

In summary, this thesis will contribute to the development of novel chemical strategies for site-selective protein modification of single cysteines and disulfides. The methodologies presented here have been developed so that various therapeutic relevant protein conjugates could be realized to address previously "unanswered" fundamental scientific questions as well as provide practical solutions for the emerging healthcare problems.
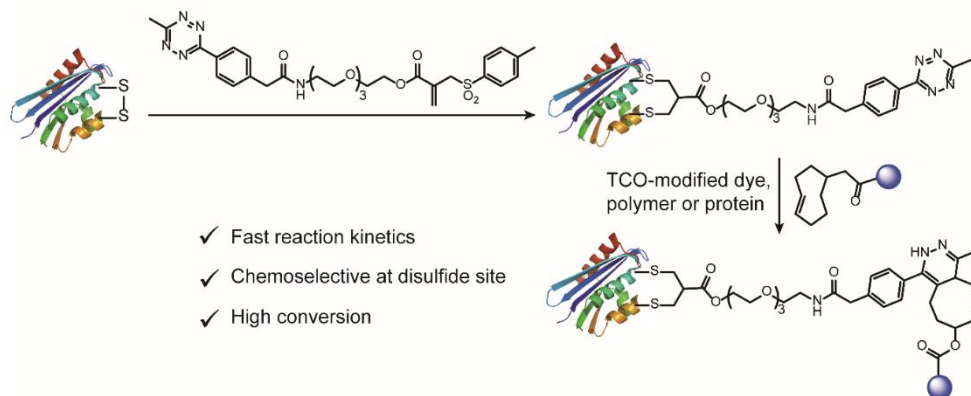
Fig. 2.1 Overview of bioconjugation strategies developed in this thesis. (a) Disulfide modification of protein with ally sulfone derivatives to introduce the tetrazine tag for the subsequent iEDDA reaction. (b) Chloromethyl derivatives for selective cysteine and disulfide modification on demand by synthetic customization.

## 3. Results and discussion

### 3.1 Site-selective protein modification via disulfide rebridging for fast tetrazine/trans-cyclooctene bioconjugation
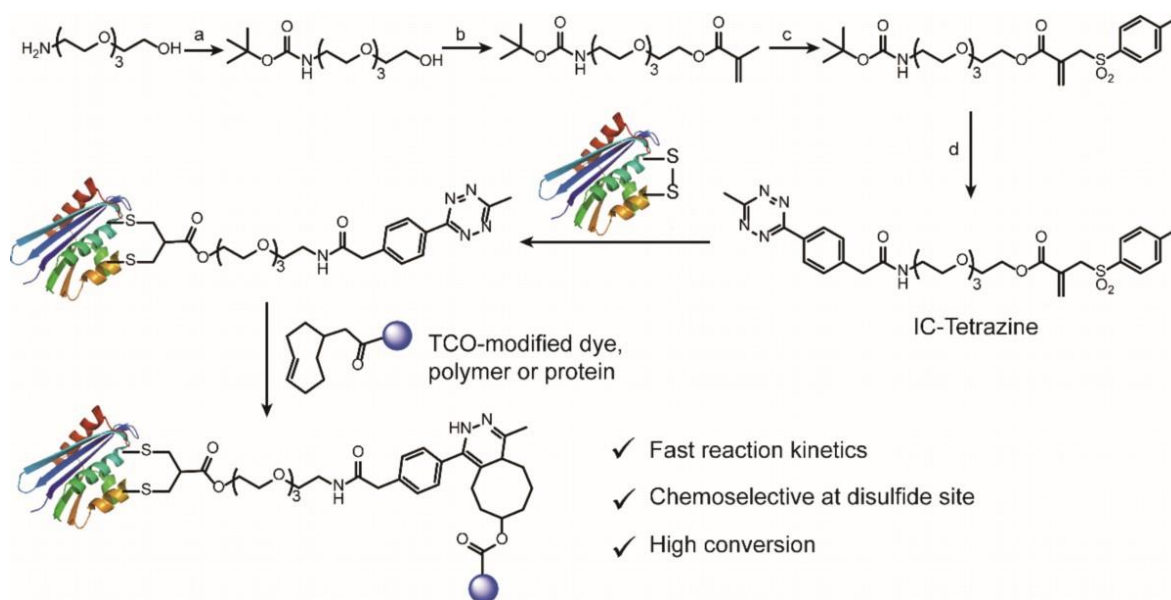


Fig. 3.1 Scheme overview of site-selective incorporation of a reactive tetrazine tag into disulfide-containing proteins (a cell-targeting peptide: somatostatin and antibody fragment: Fab fragment). The tetrazine-modified peptide/protein was used for post-functionalization to prepare a small set of protein bioconjugates with different functionalities in a fast and highly selective fashion with high conversion (one isomer is shown as a representation). The ally sulfone disulfide rebridging reagent (IC-Tetrazine) was successfully obtained via a four-step synthesis (reaction conditions: (a) Di-tert-butyl dicarbonate, $CH_2Cl_2$, overnight. (b) Methacryloyl chloride, $Et_3N$, $CH_2Cl_2$, 90%. (c) 1. I2, sodium p-toluene-sulfinate, $CH_2Cl_2$, 3 days. 2. $Et_3N$, $CH_2Cl_2$, overnight. 3. $Et_3N$, ethyl acetate, 95 °C, overnight, final yield: 60%. (d) Trifluoroacetic acid (TFA), $CH_2Cl_2$, 98%. (e) Methyltetrazine NHS ester, $Et_3N$, $CH_2Cl_2$, overnight,40%.)[1]

Site-selective modification of the therapeutically relevant peptides and proteins has already become an invaluable tool to expand the features and functions of the existing proteins to address the challenges in various biological and medical applications.[2-3] Recently, protein modification with bioorthogonal tags has emerged as an elegant approach to achieve post-functionalization of proteins in an "on-demand" fashion, as it enables the properties and functions of the protein therapeutics to be easily programmed and customized via the subsequent bioorthogonal reactions.[4-5] Among the existing bioorthogonal reactions, the iEDDA reaction between 1,2,4,5-tetrazine with trans-cyclooctene (TCO) stands out, providing fast reaction kinetics (rate constant of up to 106 M−1 s−1), excellent orthogonality, catalyst-free conditions, and good biocompatibility, and is arguably the most appropriate bioorthogonal reactions up to date.[6] However, utilizing iEDDA reaction for protein modification mainly relies on the statical modification of the amino groups on the protein surface via using NHS ester chemistry.[7-9] Although it was shown that the tetrazine or TCO groups were incorporated into proteins in a site-selective fashion, this method actually suffers

from the low yields and tedious optimization of the respective tRNA synthetase/tRNA pair for efficient incorporation of the ncAAs.[10-13] This in turn will limit its scalability and accessibility for further applications. Therefore, there is still a high demand to develop more general and robust methods to employ iEDDA reaction for fast and efficient protein modification.

In this work, we present a simple and straightforward method for the introduction of tetrazine groups to disulfide-containing proteins for the preparation of well-defined protein conjugates via the iEDDA reaction with three different TCO-containing functionalities. First, a tetrazine-based allyl sulfone disulfide rebridging reagent was designed and readily obtained via a four-step synthesis. Compared to the well-known bis-sulfone disulfide rebridging reagents, ally sulfone reagents bear the intrinsic advantages of higher water solubility and no need for in-situ activation, thus greatly facilitating the efficiency for subsequent modification.

Two model substrates (SST and IgG Fab) were selected for site-selective disulfide modification with IC-Tetrazine. SST is a cyclic peptide containing a single accessible disulfide bond in its structure, which targets the five G protein-coupled receptors (SSTR) that are overexpressed in high levels in various kinds of cancer cells and tumor blood vessels.[14] After the reduction of the disulfide bond by TCEP, tetrazine groups were successfully introduced to the disulfide site yielding the SST-Tetrazine after HPLC purification, which was demonstrated by the MALTOF-Tof-MS confirming the correct MS increase. In addition, the antibody fragment from human immune globulin G (IgG), which retains the antigen-binding region for active binding but circumvents non-specific binding of the Fc region of the antibodies[15], was also successfully modified with tetrazine tag after disulfide reduction. The successful modification was demonstrated by MALDI-Tof-MS characterization showing an increase of ~ 500 Da as well as an intact band at ~ 48 kDa on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) in the presence of 100 equivalents of TCEP.

A set of well-defined proteins conjugates was successfully obtained via the conjugation of tetrazine-modified SST and IgG Fab with a series of TCO-modified functionalities, such as a fluorophore (cyanine-5, Cy5) and a PEG chain. Owing to the fast reaction kinetics and excellent conversion of iEDDA reaction, the bioconjugation reaction can be finished within 30 minutes with almost quantitative yields. In addition, SST-Tetrazine was also conjugated to a protein enzyme (cytochrome C, CytC) to form a peptide-protein conjugate with high conversion. However, MALDI-Tof-MS data showed that no conjugation product was observed in the reaction mixture, presumably due to the steric hindrance of the two bulky macromolecules.

The stability of the resultant bioconjugates in cellular environments is also an important consideration for further applications. Herein, the stability of the obtained bioconjugates was also evaluated in glutathione (GSH) and 1% fetal bovine serum (FBS). LC-MS data showed that SST-Tetrazine remained intact after 24 hours of incubation and no degradation products were detected in the SIM profile. In-gel fluorescence of Fab-Cy5 demonstrated that the intensity of the

fluorescence band has no change when incubated with FBS for over 24 hours. Besides the stability study, the investigation of the structural and functional change of the resultant bioconjugates has also been performed through circular dichroism (CD) and enzyme-linked immunosorbent assay (ELISA) experiments. CD data demonstrated that the SST, SST-Tetrazine, and SST-PEG conjugates showed typical random coil structures with a negative band at 201 nm. Compared to the native IgG Fab, Fab-Tetrazine and Fab-PEG conjugate still posse well-defined antiparallel β-pleated sheets confirmed by the peaks at 218 nm and 202 nm consistent with the literature report, proving that the secondary structures of the Fab conjugates were preserved after modification and conjugation. The binding ability of modified IgG Fab to protein L was also evaluated by EISA under different concentrations, suggesting the comparable recognition capability between IgG Fab and Fab-Tetrazine.

In summary, this work has contributed a straightforward and broadly applicable method for protein modification and conjugation via the iEDDA reaction. Notably, the modified proteins exhibited good stability under biologically relevant conditions and their secondary structure was preserved after modification and bioconjugation. The approach presented herein is a valuable addition to the chemical toolbox for site-selective protein labeling and manipulation, offering a fast, robust and straightforward method to be easily adapted in any laboratory.

These results are presented comprehensively in section 5.2.

## 3.2 Chemoselective cysteine or disulfide modification via single atom substitution in chloromethyl acryl reagents
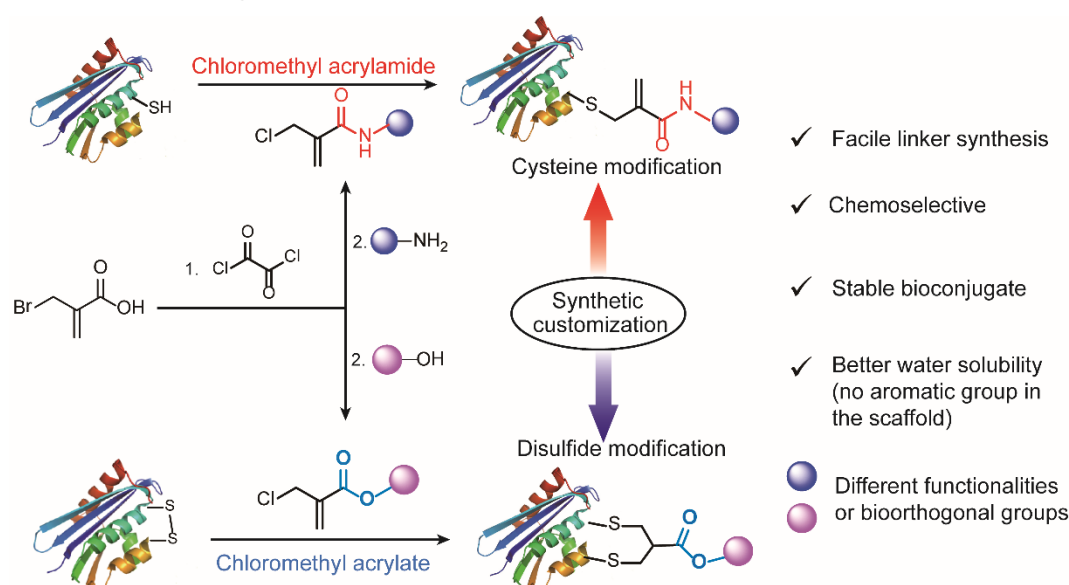


Fig. 3.2 General scheme for chloromethyl acrylamide and acrylate for site-selective protein modification at cysteine and disulfide sites.

In the past decades, significant progress has been made in developing new bioconjugation methodologies for site-selective protein modification. In particular, a variety of structurally diverse reagents have been reported for the selective modification of cysteine residues.[16-18] However, synthesis strategies for disulfide modification have been less explored and the current toolset is still limited to five to six conjugation methods available in the literature.[19] The stark contrast in the availability of disulfide functionalization strategies, compared to free cysteine modifications, stems from the stringent requirement to simultaneously react with two cysteine residues using a single reagent in order to preserve bioactivity. Even though significant progress has been made in the development of new methodologies for protein modification at cysteine or disulfide sites, many of these strategies are still limited by complicated synthesis, poor biocompatibility of the conjugation reagents, or the instability of the resultant conjugates. Consequently, the scalability and general accessibility for further applications are severely restricted. On the other hand, the methods developed to date are mainly designed to target a single specific residue, for example cysteine or disulfide site. Besides (bromo)maleimides[20], 3-bromo-5-methylene pyrrolones[21] and diethynyl phosphinates[22], there are rare examples of bioconjugation strategies that can provide a broad spectrum scaffold to address either cysteine or disulfide sites in a user-defined fashion by synthetic customization. Such a strategy is more advantageous compared to reinvent a different scaffold for every single purpose. In this way, the synthetic path is more straightforward if both the cysteine and disulfide modification need to be achieved for certain biological applications.

Therefore, the development of a novel bioconjugation strategy where the respective properties

of the bioconjugation reagent can be tailored to achieve cysteine or disulfide modification on demand would be highly advantageous to enrich the existing toolbox. This contribution has introduced for the first time chloromethyl acryl reagents that serve as general yet versatile scaffolds for synthesizing the respective acrylamide or acrylate derivatives by conjugation with different end-group functionalities (amino group or hydroxyl group) in a one-pot reaction (Fig. 1). A toolbox containing different functionalities, e.g. dye or bioorthogonal groups, was prepared as demonstrated in Scheme 1 showing the broad applicability of this method. Compared to other disulfide- and cysteine-modification reagents, which require multiple-step synthesis (e.g. allyl sulfone requires five-step synthesis[23]), the chloromethyl acryl derivatives are readily available through a straightforward one-pot synthesis from commercially available 2-(bromomethyl)acrylic acid precursors. The simplicity of the synthesis provides greater access to introducing a broad spectrum of functional features for subsequent protein conjugations.

Model reactions with two exemplary amino acid substrates (Boc-Cys-OMe and Boc-Ly-OH) demonstrated that both chloromethyl acrylamide and acylates compounds revealed excellent chemoselectivity towards thiol groups over amino groups. However, chloromethyl acrylamide compounds can only react with one equivalent of the thiol group while acrylate compounds can readily undergo two successive Michael reactions to react with two equivalents of thiol groups. The preliminary studies established on the model reaction level clearly indicated the difference in reactivity between the chloromethyl acrylamide and acrylate reagents, presumably originating from the amide or ester linkage. It was reported that catalysts and high temperatures are required for thiol addition while utilizing α, β-unsaturated amide as Michael acceptor.[24-25] Therefore, we introduce chloromethyl acrylamides for the modification of free cysteines with high efficiency and excellent chemoselectivity without further additional reaction with excess thiols. In contrast, chloromethyl acrylate compounds can undergo two Michael reactions in a successive manner, thereby making them suitable candidates to achieve protein modification at disulfide sites.

Cysteine modification using chloromethyl acrylamide compounds was further investigated based on the model reaction with Boc-Cys-OMe. Conversion rates reached 80% in less than two hours and near quantitative conversion was achieved in less than six hours with the second-order rate constant that was determined to 1.17 $M^{-1}s^{-1}$. Next, chloromethyl acryl compounds were applied for cysteine modification of bioactive peptide substrates. It was shown that WSC02 peptide, an endogenous peptide inhibitor of CXCR4 receptor which is highly relevant for anti-infectivity in viral infection and anti-migratory effect in cancer[26], can be efficiently modified with 95% yield in less than four hours. In addition, if the WSC02 peptide was firstly masked with a thiol-reactive reagent 4,4'-dithiodipyridine (4-DPS), no further modification was observed with chloromethyl acrylamide compounds. In addition to WSC02, four other peptides, including RGDC, CEIE, PC-8 and EK-1 peptides, have also been successfully modified. The broad range of substrates used here clearly

demonstrates the general applicability of chloromethyl acrylamide compounds for site-selective cysteine modification. After showing the successful modification of model amino acids and peptides, we proceeded to modify a more complex protein substrate. Ubiquitin, a protein that contains a cysteine mutation at its K63 position, was selected for modification. It was shown that ubiquitin can be successfully modified with ten equivalents of the respective chloromethyl acrylamide compounds which was characterized by MALDI-Tof-MS with the correct molecule weight increase. In addition, the resultant protein conjugate demonstrated good stability under three different pH (pH = 6, 7 and 8) and the secondary structure of the ubiquitin conjugate was well-preserved after modification.

Thereafter, the feasibility of chloromethyl acrylates for disulfide modification was evaluated on both peptide and protein substrates. A cyclic peptide hormone somatostatin (SST), which plays a key role in regulating the endocrine system and contains an accessible disulfide bond in the sequence[14], was selected as a model peptide for modification. HPLC data indicated that good modification efficiency was achieved by using 1.1 eq of chloromethyl acrylate compounds with negligible side product observed. Furthermore, the disulfide modification efficiency of SST with chloromethyl acrylate developed in this paper and the allyl sulfone reagents developed by our group before were also evaluated. The data demonstrated that disulfide modification with chloromethyl acrylate compound requires less organic solvents to soluble the modification reagent (88% based on the quantification of HPLC peak for the reaction mixture) and the modification efficiency was also significantly higher than that using ally sulfone reagents (67%). Furthermore, the disulfide modification strategy was also evaluated on a much more complex protein substrate, lysozyme (from hen egg white). Three different reagents were applied to successfully modify the lysozyme with the isolated yield of 28 %, 22 % and 24 % respectively. More importantly, it was shown that the site-selectively modified lysozyme retained retains 86% of the activity. In contrast, statistical modification of lysine residues of lysozyme via using tetrazine N-hydroxysuccinimide compounds resulted in total loss of its catalytic activity. Therefore, disulfide modification of proteins with chloromethyl acrylate compounds represents an attractive approach to functionalize enzymatic proteins with well-preserved activity.

In summary, this study has contributed a simple yet versatile class of chloromethyl acryl derivatives to achieve selective protein modification of cysteines or disulfides on demand. The reported chloromethyl acrylate and acrylamide compounds can be synthesized in a simple one-pot reaction using widely accessible starting materials with moderate to good yields. The simple and straightforward synthesis highlights its high practicality, which in turn makes it easy to be adapted in any lab. According to the further intended purposes, the reactivity profile of the conjugation reagents can be pre-determined in a user-defined fashion by choosing different end-group functionalities (either the amino or hydroxyl groups) to obtain the respective chloromethyl acrylate

and acrylamide compounds. The strategy presented here greatly enrich the currently available methodology toolbox for cysteine and disulfide modification, providing easy access to the design and preparation of advanced protein conjugates for various sophisticated biological applications

These results are presented comprehensively in section 5.3.

## 3.3 References

1. Xu, L., Raabe, M., Zegota, M. M., Nogueira, J. C. F., Chudasama, V., Kuan, S. L.and Weil, T., Site-selective protein modification via disulfide rebridging for fast tetrazine/trans-cyclooctene bioconjugation. *Org. Biomol. Chem.* **2020**, *18*, 1140-1147.

2. Krall, N., da Cruz, F. P., Boutureira, O.and Bernardes, G. J. L., Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **2016**, *8*, 103-113.

3. Hu, Q. Y., Berti, F.and Adamo, R., Towards the next generation of biomedicines by site-selective conjugation. *Chem. Soc. Rev.* **2016**, *45*, 1691-719.

4. Shadish, J. A.and DeForest, C. A., Site-Selective Protein Modification: From Functionalized Proteins to Functional Biomaterials. *Matter* **2020**, *2*, 50-77.

5. Hoyt, E. A., Cal, P. M. S. D., Oliveira, B. L.and Bernardes, G. J. L., Contemporary approaches to site-selective protein modification. *Nat. Rev. Chem.* **2019**, *3*, 147-171.

6. Oliveira, B. L., Guo, Z.and Bernardes, G. J. L., Inverse electron demand Diels–Alder reactions in chemical biology. *Chem. Soc. Rev.* **2017**, *46*, 4895-4950.

7. Devaraj, N. K., Weissleder, R.and Hilderbrand, S. A., Tetrazine-Based Cycloadditions: Application to Pretargeted Live Cell Imaging. *Bioconjugate Chem.* **2008**, *19*, 2297-2299.

8. Devaraj, N. K.and Weissleder, R., Biomedical Applications of Tetrazine Cycloadditions. *Acc. Chem. Res.* **2011**, *44*, 816-827.

9. Devaraj, N. K., Upadhyay, R., Haun, J. B., Hilderbrand, S. A.and Weissleder, R., Fast and sensitive pretargeted labeling of cancer cells through a tetrazine/trans-cyclooctene cycloaddition. *Angew. Chem. Int. Ed.* **2009**, *48*, 7013-7016.

10. Seitchik, J. L., Peeler, J. C., Taylor, M. T., Blackman, M. L., Rhoads, T. W., Cooley, R. B., Refakis, C., Fox, J. M.and Mehl, R. A., Genetically Encoded Tetrazine Amino Acid Directs Rapid Site-Specific in Vivo Bioorthogonal Ligation with trans-Cyclooctenes. *J. Am. Chem. Soc.* **2012**, *134*, 2898-2901.

11. Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D. J., Blackman, M. L., Fox, J. M.and Chin, J. W., Genetic Encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. *J. Am. Chem. Soc.* **2012**, *134*, 10317-10320.

12. Lang, K., Davis, L., Torres-Kolbus, J., Chou, C., Deiters, A.and Chin, J. W., Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat. Chem.* **2012**, *4*, 298-304.

13. Machida, T., Lang, K., Xue, L., Chin, J. W.and Winssinger, N., Site-Specific Glycoconjugation of Protein via Bioorthogonal Tetrazine Cycloaddition with a Genetically Encoded trans-Cyclooctene or Bicyclononyne. *Bioconjugate Chem.* **2015**, *26*, 802-806.

14. Kuan, S. L., Fischer, S., Hafner, S., Wang, T., Syrovets, T., Liu, W., Tokura, Y., Ng, D. Y. W., Riegger, A., Förtsch, C., Jäger, D., Barth, T. F. E., Simmet, T., Barth, H.and Weil, T., Boosting Antitumor Drug Efficacy with Chemically Engineered Multidomain Proteins. *Adv. Sci.* **2018**, 1701036.

15. Richards, D. A., Maruani, A.and Chudasama, V., Antibody fragments as nanoparticle targeting

ligands: a step in the right direction. *Chem. Sci.* **2017**, *8*, 63-77.

16. Ochtrop, P.and Hackenberger, C. P. R., Recent advances of thiol-selective bioconjugation reactions. *Curr. Opin. Chem. Biol.* **2020**, *58*, 28-36.

17. Reddy, N. C., Kumar, M., Molla, R.and Rai, V., Chemical methods for modification of proteins. *Org. Biomol. Chem.* **2020**, *18*, 4669-4691.

18. Gunnoo, S. B.and Madder, A., Chemical Protein Modification through Cysteine. *ChemBioChem* **2016**, *17*, 529-553.

19. Kuan, S. L., Wang, T.and Weil, T., Site-Selective Disulfide Modification of Proteins: Expanding Diversity beyond the Proteome. *Chem. Eur. J.* **2016**, *22*, 17112-17129.

20. Smith, M. E. B., Schumacher, F. F., Ryan, C. P., Tedaldi, L. M., Papaioannou, D., Waksman, G., Caddick, S.and Baker, J. R., Protein Modification, Bioconjugation, and Disulfide Bridging Using Bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132*, 1960-1965.

21. Zhang, Y., Zang, C., An, G., Shang, M., Cui, Z., Chen, G., Xi, Z.and Zhou, C., Cysteine-specific protein multi-functionalization and disulfide bridging using 3-bromo-5-methylene pyrrolones. *Nat. Commun.* **2020**, *11*, 1015.

22. Stieger, C. E., Franz, L., Korlin, F.and Hackenberger, C. P. R., Diethynyl Phosphinates for Cysteine-Selective Protein Labeling and Disulfide Rebridging. *Angew. Chem. Int. Ed.* **2021**, *60*, 1-7.

23. Wang, T., Riegger, A., Lamla, M., Wiese, S., Oeckl, P., Otto, M., Wu, Y., Fischer, S., Barth, H., Kuan, S. L.and Weil, T., Water-soluble allyl sulfones for dual site-specific labelling of proteins and cyclic peptides. *Chem. Sci.* **2016**, *7*, 3234-3239.

24. Liu, Y., Lai, Z., Yang, P., Xu, Y., Zhang, W., Liu, B., Lu, M., Chang, H., Ding, T.and Xu, H., Thio-Michael addition of α,β-unsaturated amides catalyzed by Nmm-based ionic liquids. *RSC Advances* **2017**, *7*, 43104-43113.

25. Alexander Bohme, D. T., Albrecht Paschke, and Gerrit Schrmann, Kinetic Glutathione Chemoassay To Quantify Thiol Reactivity of Organic Electrophiles-Application to - r,Unsaturated Ketones, Acrylates, and Propiolates. *Chem. Res. Toxicol.* **2009**, *22*, 742–750.

26. Zirafi, O., Kim, K. A., Standker, L., Mohr, K. B., Sauter, D., Heigele, A., Kluge, S. F., Wiercinska, E., Chudziak, D., Richter, R., Moepps, B., Gierschik, P., Vas, V., Geiger, H., Lamla, M., Weil, T., Burster, T., Zgraja, A., Daubeuf, F., Frossard, N., Hachet-Haas, M., Heunisch, F., Reichetzeder, C., Galzi, J. L., Perez-Castells, J., Canales-Mayordomo, A., Jimenez-Barbero, J., Gimenez-Gallego, G., Schneider, M., Shorter, J., Telenti, A., Hocher, B., Forssmann, W. G., Bonig, H., Kirchhoff, F.and Munch, J., Discovery and characterization of an endogenous CXCR4 antagonist. *Cell Rep* **2015**, *11*, 737-747.

## 4. Conclusion and outlook

Site-selective protein modification has already emerged as an avant-garde approach to prepare precision protein conjugates with unprecedented functional and structural diversities by integrating the advantages from both the synthetic world and Nature in a synergistic fashion. Cysteines and their oxidized form as "disulfide bond" represent widely applied targets for modification because of their low abundance on protein surfaces, high therapeutic relevance since a lot of therapeutic proteins and peptides contain disulfide bonds and versatile reactivity profile of the thiol group. Despite the substantial achievements made in this field, the current bioconjugation methods still suffer from some inherent drawbacks, such as the low labeling efficiency, tedious linker synthesis, low water solubility and the instability of the resultant conjugates. Therefore, this thesis has focused on the development of novel bioconjugation approaches based on the chemical design for site-selective protein modification at disulfides or cysteine residues, which address many of the aforementioned limitations of current strategies in the literature. At the same time, the approaches devised within this thesis provide facile linker synthesis, highly labeling efficiency, and specificity, excellent operational simplicity to afford stable bioconjugates.

In the first work, the ultrafast click reaction "iEDDA reaction", which demonstrated fast reaction kinetics, excellent orthogonality, and good biocompatibility, was utilized for disulfide modification to prepare well-defined protein conjugates in high yield. Herein, the tetrazine groups were site-selectively introduced into proteins at the disulfide site via a newly designed reagent based on the allyl sulfone scaffold. Two model substrates, SST and IgG Fab, have been successfully modified with the tetrazine tag, which subsequently reacted with a series of TCO-containing functionalities via the ultrafast iEDDA reaction to construct a small library of well-defined protein conjugates in high yield. The two-step approach provides convenient access to the introduction of bulky functionalities and would enable the preparation of the desired conjugates "on-site", thus the final conjugates can be applied to the patient immediately after preparation, preventing long-term storage, quality control, and loss of bioactivity during storage.

Although allyl sulfone reagents allow site-selective protein disulfide modification, they still suffer from some inherent limitations, such as low water solubility, low modification efficiency and relatively slow reaction kinetics. In the second project, simple yet versatile reagents were designed, in which a commercially available starting material was coupled with the functionalities containing either an amino or hydroxyl end-group to synthesize the respective chloromethyl acrylamide or acrylate compounds. Based on the reactivity profile dominated by the amide or ester linkage, chloromethyl acrylamide compounds can modify the protein at cysteine residue, while acrylate derivates allow for the modification at the disulfide site via two successive addition-elimination reactions with the two thiol groups. Compared to the current methodologies existing in the literature,

the strategy presented bears the advantages of facile linker synthesis, high labeling efficiency, good water solubility and excellent stability of the resulting conjugates. All these features make this approach a simple yet broadly applicable tool towards diverse functional protein conjugates with well-defined structures and preserved functions.
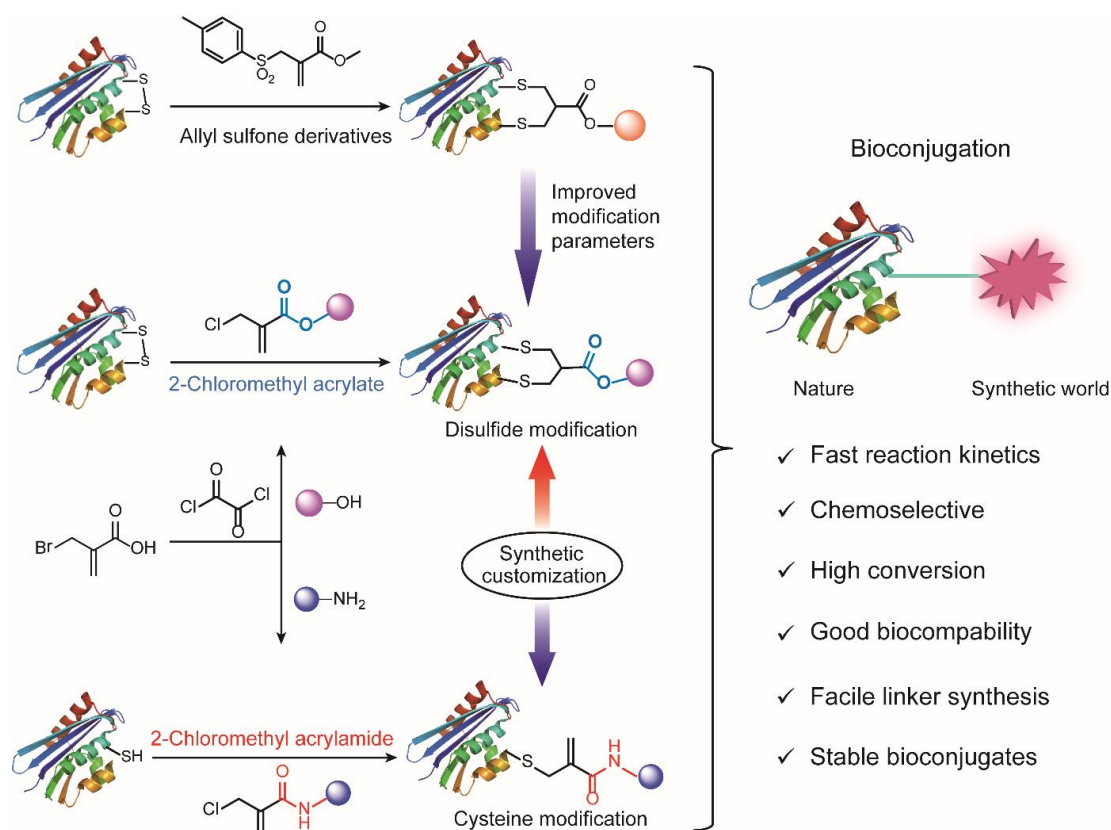


Fig. 4.1 Overview of the bioconjugation methodologies developed in this thesis.

By overcoming the significant limitations of the existing bioconjugation methodologies for single modification of proteins, the results of the current thesis contribute to the development of how proteins and synthetic chemistry can be intertwined to create innovative functional protein conjugates to solve therapeutic challenges. Moreover, the technologies developed here offer versatile reactivity profiles so that they can be combined with themselves or other complementary approaches to achieve dual modification of proteins in view of meeting the needs and demand for multifunctional protein conjugates for contemporary biomedicine and personalized therapy. For example, the efficiencies of traditional ADCs, which only contain one type of drug, are still hampered by the high hydrophobicity and drug resistances. In addition, state-of-art therapy requires the integration of imaging and therapy agents into a single targeting moiety, e.g. antibody, for real-time monitoring during the treatment, which is beyond the scope of protein single functionalization. In this context, dual modification of proteins, which permit the incorporation of two

different types of functionalities, has emerged as a valuable strategy to solve these problems, as exemplified by the incorporation of two different anticancer drugs to antibodies to overcome the drug resistance, decorating the traditional ADCs with additional PEG chain to increase the hydrophilicity as well as extend the circulation life, integration of a toxic drug and an imaging agent into one platform for theranostic applications, and so forth. These multifunctional protein conjugates will undoubtedly provide new insights for the preparation of innovative protein hybrid materials including exceptional features or properties to address the great challenges in biomedical research.

Furthermore, due to the great flexibility and the breadth that modern chemical design offer, one could envision that the incorporated component can be tailored to impart a certain feature that is individualized to each patient to maximize the in vivo efficacy. This will pave the way towards "smart" and "intelligent" protein conjugates that can respond or be activated under specific stimuli, such as the biomarker in cancer cells, to optimize the biological profile and the efficiency of the resultant bioconjugates. In addition, with the help of modern chemical tools, it would be highly attractive that the protein conjugates can self-adapt to the microenvironment of diseases, communicate with each other and provide a self-feedback loop to achieve an output or decision, e.g. termination/activation, to solve a targeted biological problem by themselves. In this way, current bottlenecks and challenges in therapeutic applications can be addressed with a completely new perspective through rationale chemical design.

# 5. Publications

## 5.1 Contemporary approaches for site-selective dual functionalization of proteins

Lujuan Xu, Seah Ling Kuan* and Tanja Weil*

**Contribution of the respective authors:**

**Lujuan Xu:** Discussion of the scope of the review, writing the review and design of the figures

**Seah Ling Kuan:** Discussion of the scientific scope of the review, correcting the manuscript

**Tanja Weil:** Acquiring funding for the project, discussion of the scientific scope of the review, correcting the manuscript

Protein Functionalization

# Contemporary Approaches for Site-Selective Dual Functionalization of Proteins

*Lujuan Xu, Seah Ling Kuan\* and Tanja Weil\**

59
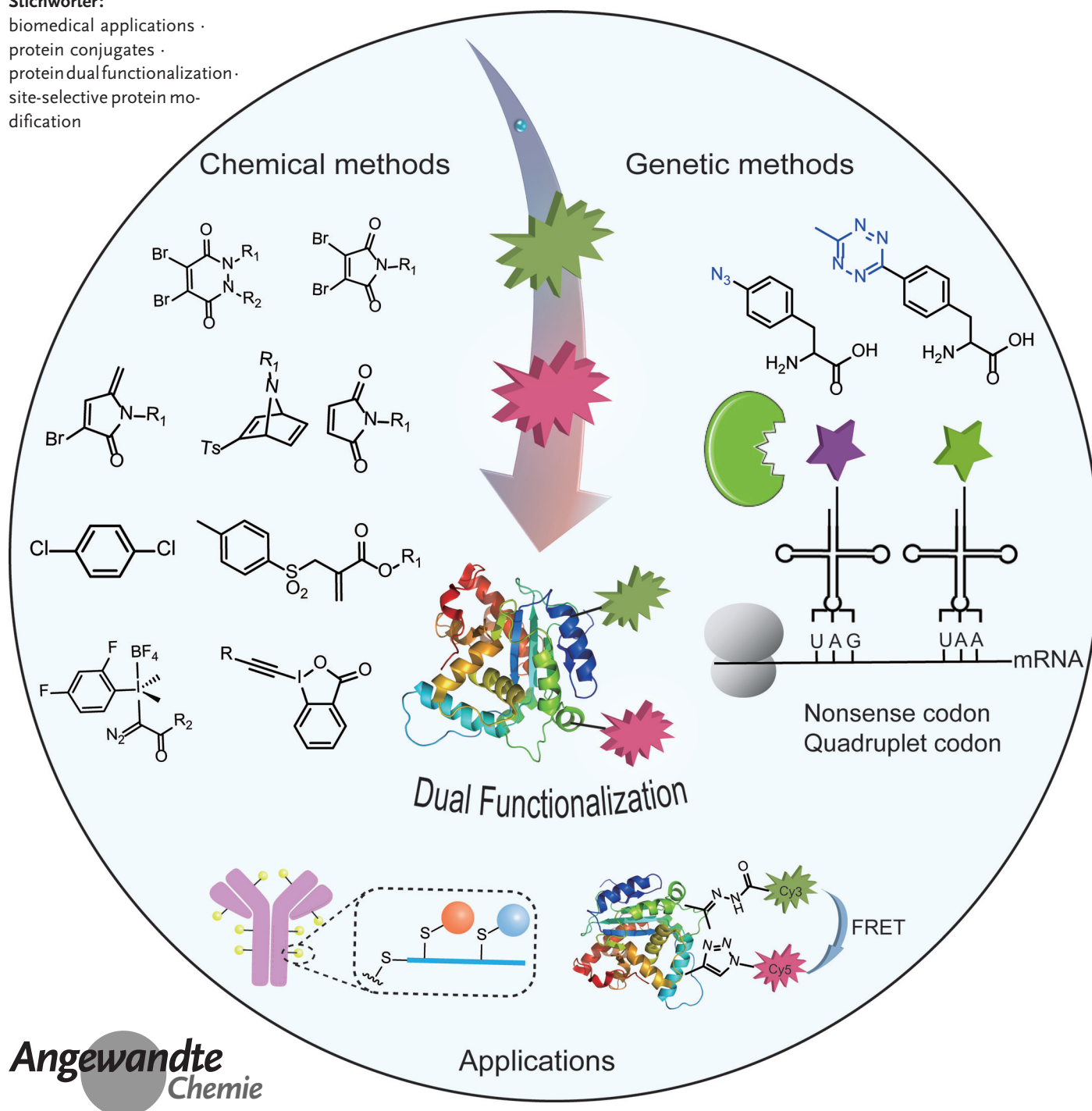
**S**ite-selective protein functionalization serves as an invaluable tool for investigating protein structures and functions in complicated cellular environments and accomplishing semi-synthetic protein conjugates such as traceable therapeutics with improved features. Dual functionalization of proteins allows the incorporation of two different types of functionalities at distinct location(s), which greatly expands the features of native proteins. The attachment and crosstalk of a fluorescence donor and an acceptor dye provides fundamental insights into the folding and structural changes of proteins upon ligand binding in their native cellular environments. Moreover, the combination of drug molecules with different modes of action, imaging agents or stabilizing polymers provides new avenues to design precision protein therapeutics in a reproducible and well-characterizable fashion. This review aims to give a timely overview of the recent advancements and a future perspective of this relatively new research area. First, the chemical toolbox for dual functionalization of proteins is discussed and compared. The strengths and limitations of each strategy are summarized in order to enable readers to select the most appropriate method for their envisaged applications. Thereafter, representative applications of these dual-modified protein bioconjugates benefiting from the synergistic/additive properties of the two synthetic moieties are highlighted.

## 1. Introduction

Proteins are ubiquitous in Nature, serving as the basic building blocks of life. They play many essential roles in a myriad of biological processes, such as molecular transport, energy conversion, inter- and intramolecular signaling.[1,2] Nature expands protein diversity by post-translational modifications (PTMs) after their biosynthesis in the ribosome, thus vastly enlarging their structural and functional repertoire by up to two orders of magnitude.[3,4] Inspired by Nature's elegance, scientists strived to modify proteins with diverse synthetic moieties, allowing for the creation of bioconjugates with high degree of structural perfection and new functional characteristics.[5]

In this context, the past decades revealed significant progress in the development of new methodologies for site-selective protein functionalization to install the desired functionalities at pre-defined sites.[3,6–10] These well-defined protein conjugates offer great prospects for a wide range of fields including biomedicine, bioimaging, biosensing and materials science.[5,11–17] Some promising examples include the conjugation of synthetic polymers to therapeutic proteins to improve their solubility and extend their plasma circulation half-life[13] or the attachment of anticancer drugs to antibodies forming antibody-drug conjugates (ADCs) for cell-targeted cancer therapy.[11,12]

Nevertheless, along with the increasing demand for multifunctional bioconjugates to perform more sophisticated biological studies in vitro as well as in vivo, appending only one type of functionality to proteins is often insufficient to customize proteins for the desired applications. For example, despite the clinical success of ADCs, classical ADCs equipped with a single type of anticancer drug could still suffer from low efficacy, drug resistance, unfavorable pharmacokinetics, immunogenicity, and the inherent hydrophobicity of the drug, which greatly hampers their further in vivo applications.[18,19] In addition, there is also an increasing demand for new therapeutic strategies which combine imaging agents and drug molecules within the protein, for example, antibody, for real-time monitoring during the treatment.[20] In another example, a protein was functionalized with three copies of cell-targeting somatostatin peptide and an enzyme. Remarkably, the resultant bioconjugate inhibited tumor growth already at 100-fold lower concentration than a clinically approved antibody acting via a similar mode of action. Furthermore, co-administration of this protein bioconjugate with an approved anticancer drug, doxorubicin, boosted its antitumor activity in a combination therapy approach.[21] However, these multifunctional protein

[*] L. Xu, Dr. S. L. Kuan, Prof. Dr. T. Weil
Max Planck Institute for Polymer Research
Ackermannweg 10, 55128 Mainz (Germany)
and
Institute of Inorganic Chemistry I, Ulm University
Albert-Einstein-Allee 11, 89081 Ulm (Germany)
E-Mail: kuan@mpip-mainz.mpg.de
weil@mpip-mainz.mpg.de

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/anie.202012034.

60

conjugates have been mainly achieved by statistical modification on the protein surface,[22,23] which results in heterogeneous mixtures with reduced protein activity as well as batch-to-batch variations. In view of biosafety, these limitations significantly hamper their further developments. Consequently, there is a pressing need to devise new strategies to generate protein bioconjugates that exhibit higher order of structural and functional complexity but retaining structural perfection. In this regard, site-selective dual functionalization of proteins has emerged as a promising strategy to customize proteins for the respective applications.

In this review, dual functionalization is defined as the incorporation of two different functionalities into proteins in a site-selective fashion, which is accomplished either at two different amino acids (AA) sites or at a single AA residue at the protein surface (Figure 1). In both cases, the reagents as well as the sequence of the bioconjugation reactions need to be carefully considered. Dual-modified protein bioconjugates can harness the function and properties imparted by the respective payloads, which complement the capabilities of biomolecules and significantly expand their functional arsenal.[24] For example, dual functionalization with two chromophores at distinct sites allows real-time monitoring of changes in protein conformations and dynamics upon ligand binding in native environments or in response to certain stimuli by Förster Resonance Energy Transfer (FRET) measurements.[25] This cannot be achieved by mono-functionalization of proteins. Moreover, a new generation of protein therapeutics/diagnostics can be derived by incorporating two different functionalities, as exemplified by integrating a drug and an imaging agent into one platform for simultaneous therapy and diagnostics applications (theranos-
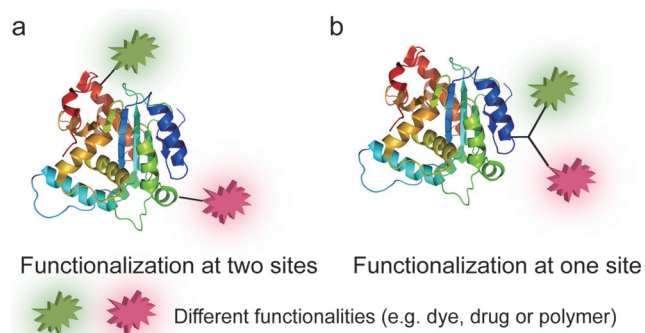


**Figure 1.** Two different approaches for dual functionalization of proteins (a) at two different sites and (b) at one single site with a multifunctional linker.

tics) or inserting two different drugs into an antibody for combination therapy.[20,26,27]

Owing to the emerging interest for addressing previously „unanswered" fundamental scientific questions as well as offering practical solutions for biomedical applications, site-selective dual functionalization of proteins has seen rapid developments in the past ten years. In this review, we first summarize the synthetic strategies for protein dual functionalization to provide a timely overview of this burgeoning field, which serves as a guideline for the selection of the most appropriate method for the envisaged applications (Figure 2). In addition, the rationale and principles behind these synthetic strategies as well as the strength and inherent limitations are discussed. In the last section, some representative applications of dual-functionalized protein conjugates

*Lujuan Xu studied chemistry at Zhengzhou University, where she received her Bachelor's (2014) and Master's (2016) degrees. From 2014 to 2016, she conducted her Master's thesis on the synthesis of small organic molecules for optoelectronic devices in the group of Prof. Jian Pei at Peking University. In 2017 she received a CSC scholarship to undertake her PhD with Tanja Weil at the MPIP. Her research focuses on the development of new chemical methodologies for site-selective modification of proteins.*

*Tanja Weil received her PhD from Mainz University in 2002. She held several leading positions at Merz Pharmaceuticals GmbH (Frankfurt, 2002–2008). She was an Associate Professor at the National University of Singapore (NUS, 2008–2010) and since 2010, director at the Institute of Organic Chemistry III (OC III) at Ulm University (Uulm). In 2016, she was appointed as director of the Department of Synthesis of Macromolecules at the Max Planck Institute for Polymer Research (MPIP) in Mainz, Germany. She has received an ERC Synergy Grant. Her scientific interests focus on innovative synthesis concepts to achieve functional macromolecules, hybrid materials and life-like systems to solve current challenges in biomedicine and material science.*

*Seah Ling Kuan received her B.Sc (Hons) in 2003 and completed her PhD (Chemistry) at the NUS in 2009. She joined Prof. Tanja Weil in NUS from 2008–2010 as a research fellow. In 2011, she began her independent research as an Alexander von Humboldt fellow in UUlm. She was a group leader in the Institute of Organic Chemistry III at UUlm from 2013–2016. She joined the MPIP in 2016 as a group leader (Protein Therapeutics). Her research focuses on the development of chemical approaches for the innovation of precision protein therapeutics to address devastating diseases such as cancer, inflammation or infections.*

**Figure 2.** Overview of dual functionalization methodologies for proteins and biomedical applications of the resultant protein conjugates (protein dual functionalization at two different amino acid residues is selected as an example).

benefitting from the additive or even synergistic features are highlighted.

## 2. Chemical toolbox for site-selective functionalization of proteins

Site-selective protein functionalization either relies on the functionalization of canonical AAs on the protein surface or genetically encoded noncanonical AA (ncAA) bearing the appropriate bioorthogonal groups.[28] Different methodologies have been reported to achieve site-selective mono-functionalization of proteins with high efficiency in a residue specific manner. Bioconjugations have been accomplished at the N-terminus, tyrosine, cysteine or serine residues, to list just a few examples. A brief summary of the commonly used chemical methods for mono-functionalization of natural AA residues is given in Table 1.[3,6,7,29–32] For more detailed discussions on protein mono-functionalization, we refer the readers to other excellent reviews on this topic.[3,6,7,29–31] Among the 20 AAs, unpaired cysteines have become the

primary choice for functionalization owing to their low abundance at the protein surface as well as the unique nucleophilicity and versatile reactivity profile of the thiol groups.[7,33–35] Besides the unpaired cysteine residues, targeting other side chain residues, such as disulfide bonds and methionine, have also emerged as valuable alternatives to the more commonly used cysteine-based strategies for protein functionalization (Table 1). All these available methods potentially offer a versatile bioconjugation toolbox to achieve dual functionalization of proteins at two different AA sites. For instance, cysteine and methionine residues are exploited for dual modification with maleimide and hypervalent iodine reagents that proceed in a sequential manner without cross reactivity.[59] Despite the simplicity and convenience to achieve dual modification at two different AA residues, there are still only few reports that address two different AA residues at the protein surface. Furthermore, judicious selection and combination of the two mono-functionalization methods is crucial as the bioconjugation reactions need to be orthogonal and also compatible with each other, to ensure high modification efficiency. Therefore, although other methods involving serine, tyrosine or tryptophan potentially offer a versatile bioconjugation toolbox that is in principle suitable for protein dual functionalization at two different AA residues, these reactions have not been reported yet in this context.

Alternatively, the attachment of a multifunctional linker containing two reactive orthogonal groups to a single site on the protein surface, usually an exposed cysteine or a disulfide bond, also provides straightforward access for the incorporation of two different functionalities. These bioconjugation reagents include maleimide derivatives,[47] sulfone derivatives[37,38] and other novel reagents,[44,46] which are discussed in Section 3.

The recent advances in bioorthogonal chemistry have greatly promoted the progress of site-selective protein

**Tabelle 1:** Brief summary of the currently available chemical toolbox for site-selective protein functionalization.

| AA residues | Methods |
| --- | --- |
| Cysteine | Halocarbonyl,[9] Maleimide derivatives,[33, 36] Sulfone derivatives,[37,38] Metal-mediated arylation[39–41] Thiol–ene/yne reactions,[42, 43] Dichlorotetrazine[44] Vinylphosphonothiolate,[45] Ethynylbenziodoxolones[46] |
| Disulfide | Allyl sulfone,[38] Dibromomaleimide,[47] Oxetane,[48] Divinylpyridine,[49] Dibromopyridazinediones[50] |
| N-terminus | 2-Formylphenylboronic derivatives,[51, 52] 2-Pyridinecarboxaldehyde,[53] 2-Ethynylbenzaldehydes,[54] o-Aminophenols,[55] 2-Cyanobenzothiazole,[56] Pictet–Spenger reaction,[57] Ketenes[58] |
| Methionine | Hypervalent iodine reagent,[59] Oxaziridines,[60] Epoxide[61] |
| Tyrosine | Diazonium,[62] Mannich-type reaction,[63] Pd-mediated alkylation,[64] Trizoline-diones[65] |
| Tryptophan | Rhodium-carbenoid,[66] Keto-ABNO,[67] Metal-mediated arylation/alkynylation[68–71] |
| Serine | Salicylaldehyde ester[72] Phosphorus–sulfur incorporation reagents[73] |
| Arginine | Glyoxal reagents[74] |

modification. Nowadays, various biorthogonal reactions have been reported with optimized reaction parameters, such as reaction rate, catalyst type, and substrate stability, to impart the desired functionalities for the envisaged applications.[75,76] Some commonly used bioorthogonal chemistries as well as selected characteristics are summarized in Figure 3. The copper-catalyzed azide-alkyne cycloaddition (CuAAC) bears the advantage of fast reaction kinetics but suffers from the usage of toxic copper catalyst.[77] Strain-promoted azide-alkyne cycloaddition (SPAAC) utilizes the strained alkyne derivatives as reactive partner to avoid the toxic catalyst.[78] Nonetheless, the major drawbacks of SPAAC are the limited water solubility of strained alkyne and slow reaction kinetics.[83] Strained cyclooctyne can also react with other 1,3-dipoles, for example, nitrones, which is termed as „strain-promoted alkyne-nitrone cycloaddition (SPANC)". SPANC reactions proceed rapidly with second order rate constants of up to 39 $M^{-1}s^{-1}$, which is about 30 times faster than the SPAAC reaction.[84] However, the fast reactivity is associated with the instability of the reactive nitrones, which are prone to hydrolysis in aqueous media.[84] Photoclick reactions offer the advantages of operational simplicity as well as spatial and temporal control due to the usage of light.[85] But recent evidences have shown that photoclick reactions could be limited by potential cross reactivity with, for example, amine residues, and its bioorthogonality still remains controversial.[86] Among all the existing bioorthogonal reactions, the



a. Staudinger ligation
k ~ 0.001 $M^{-1}s^{-1}$
Slow reaction kinetics, easy oxidation of phosphine
(Ref: *J. Org. Chem.* **2013**, *78*, 1184–1189)

b. Strain-promoted azide-alkyne cycloreaction (SPAAC)
k ~ 0.001 - 1 $M^{-1}s^{-1}$
Catalyst free, slow reaction kinetics, limited water solubility
(Ref: *PNAS.* **2007**, *104*, 16973-16797)

c. Strain-promoted azide-nitrone cycloreaction (SPANC)
k ~ 1 - 40 $M^{-1}s^{-1}$
Fast reaction kinetics, easy hydrolysis of nitrones
(Ref: *Angew. Chem.* **2010**, *122*, 3129–3132)

d. Photoclick cycloreaction
k ~ 0.1 - 60 $M^{-1}s^{-1}$
Spatial and temporal control, Bioorthogonalities under investigation
(Ref: *Curr Opin Chem Biol.* **2014**, *21*, 89–95)

e. Copper catalyzed azide-alkyne cycloreaction (CuAAC)
k ~ 10 - 100 $M^{-1}s^{-1}$
Fast reaction kinetics, toxic catalyst requirement
(Ref: *J. Am. Chem. Soc.* **2010**, *132*, 14570–14576)

f. Inverse electron demand Diels–Alder reaction (iEDDA)
k ~ 1 - $10^6$ $M^{-1}s^{-1}$
Fast reaction kinetics, catalyst free, limited substrate stability
(Ref: *Chem. Sci.,* **2016**, *7*, 1257–1261)

Different functionalities

**Figure 3.** Commonly used bioorthogonal reactions for protein modification and summary of some selected features.

inverse electron demand Diels–Alder reaction (iEDDA) stands out because of the fast reaction kinetics (rate constant of up to $10^6 \, M^{-1} s^{-1}$), high bioorthogonality, catalyst-free conditions and good biocompatibility.[87] Because of its very fast reaction kinetics, it revolutionized bioconjugation of various biomolecules and stimulated labeling in living systems with rate constants comparable to biological reactions.[87] The currently available toolbox of diverse bioorthogonal reactions provides the basis for protein dual functionalization ensuring no cross-reactivity and high yields of the bioconjugates. Figure 4 summarizes the different combinations of the bioorthogonal reactions that are commonly employed in the literature. Dual modification proceeds either in a sequential or simultaneous fashion. For example, if two identical CuAAC (or iEDDA) reactions are combined, protein dual modification has to be executed in a sequential manner to prevent cross-reactions, which otherwise would result in a mixture of products. In contrast, the combination of CuAAC and iEDDA introduces two bioorthogonal tags on proteins simultaneously. Since both reactions allow quantitative conversions, dual-modified bioconjugates can be obtained in a one-pot reaction without the need for purification of the single modified product.[82]

## 3. Synthetic strategies for dual functionalization of proteins

Most synthetic strategies for protein dual modification target canonical AAs exposed at the protein surface. These residues are immediately accessible without the need for tedious genetic engineering of recombinant protein variants and also mitigate the risk of negative effects on proteins folding and function. In this section, dual functionalization at two different sites as well as at a single site is discussed.

### 3.1. Dual functionalization at two different sites

Direct modification of two rare AA residues on the protein surface is a straightforward approach to achieve dual modification of proteins. The combination of two different mono-functionalization methods requires stringent selection to ensure orthogonality, compatibility and preferably high modification efficiency. For example, Gaunt and co-workers developed a method based on chemoselective labeling of a single methionine residue with a hypervalent iodine reagent.[59] This hypervalent iodine reagent selectively reacted with the moderate nucleophilic methionine residue in the presence of other competitive nucleophilic AA residues, which make it compatible and complementary to other bioconjugation strategies targeting other AA residues. This has been demonstrated by first modifying the unpaired cysteine residue of GTP-binding protein fragment Gα with a maleimide derivative in a Michael reaction to form a thioether bond. Subsequent modification with hypervalent iodine reagent selectively addresses the thioether in a methionine residue yielding the dual modified protein bioconjugate (Figure 5a). Notably, the methionine modification showed high site selectivity without any cross reactivity with the thiol–maleimide conjugation. Besides that, Paavola and co-workers have also achieved dual functionalization by combining cysteine modification and an N-terminal transamination



**Figure 4.** Bioorthogonal chemistry combinations for dual functionalization of proteins (a) CuAAC + CuAAC.[79] (b) CuAAC + SPAAC.[50] (c) CuAAC + iEDDA.[25] (d) SPAAC + iEDDA.[80] (e) CuAAC + Oxime ligation.[81] (f) CuAAC + iEDDA.[82] (Solid line refers to two moieties reacting with each other; dashed line refers to their orthogonal reactivities).
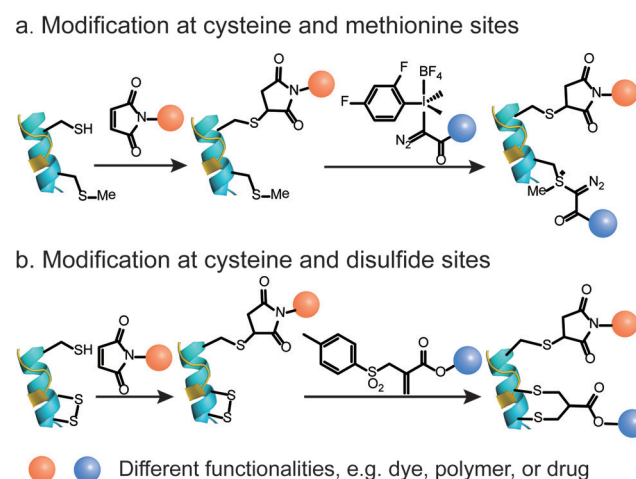


**Figure 5.** Site-selective protein dual modification at different AA residues. (a) Cysteine and methionine sites.[59] (b) Cysteine and disulfide sites.[89]

reaction mediated by pyridoxal 5-phosphate.[88] The periplasmic glutamine binding protein was site-selectively functionalized by a FRET pair, in which the ligand-induced conformational movements were monitored via changes in FRET efficiency.

An alternative strategy for site-selective protein dual modification is based on the differences in the reactivity of cysteines in their free (thiol) and oxidized (disulfide) forms.[89] Disulfide bonds could be considered as protected thiols, which require activation to form the reduced free thiols for subsequent functionalization. Therefore, site-selective dual modification of native proteins proceeds stepwise by modifying an unpaired cysteine with a maleimide reagent as the initial step, followed by the disulfide reduction to liberate two additional free thiols, which will react with a disulfide rebridging reagent, for example, an allyl sulfone (Figure 5b). It is essential that the thiol–maleimide reaction should be applied first to functionalize the unpaired cysteine residue, followed by the disulfide functionalization using allyl sulfone reagents. Otherwise, the allyl sulfone reagents will also react with the unpaired cysteine, resulting in a heterogeneous product mixture.

### 3.2. Dual functionalization at a single site

Despite the simplicity of dual modification at two distinct sites, the availability of proteins with two different AA

residues with orthogonal reactivities is rather limited. Therefore, alternative methods have been developed to target one specific AA residue with a multifunctional bioconjugation reagent.

Baker, Caddick, and co-workers have demonstrated protein dual modification using mono- and dibromomaleimide reagents.[47] The first functionality is introduced by reacting these reagents with an accessible thiol group through an addition-elimination reaction. Subsequently, an additional thiol conjugation introduces the second functionality (Figure 6a). A conceptually similar strategy utilizing dibromopyridazinediones is depicted in Figure 6a.[90] Interestingly, the native protein, for example, Grb2 adaptor protein, could be regenerated from the dual-modified conjugate after addition of phosphine or a large excess of thiols, which opens access to the reversible modulation of proteins function or controlled release of the attached cargos, such as drug molecules.[90] In a recent example, another maleimide analogue, 3-bromo-5-methylene pyrrolones (3Br-5MPs), was reported for cysteine-specific dual modification of proteins, which has comparable modification efficiency but higher cysteine specificity than the traditional maleimide reagents (Figure 6b).[91] The dual modification was achieved by two sequential Michael reactions. First, a Michael reaction of cysteine and 3Br-5MPs generated the bioconjugate that is amenable to a second Michael addition with another thiol, allowing protein dual functionalization at a cysteine site. Due to the slow release of the second functionality, a reducing reagent, for example
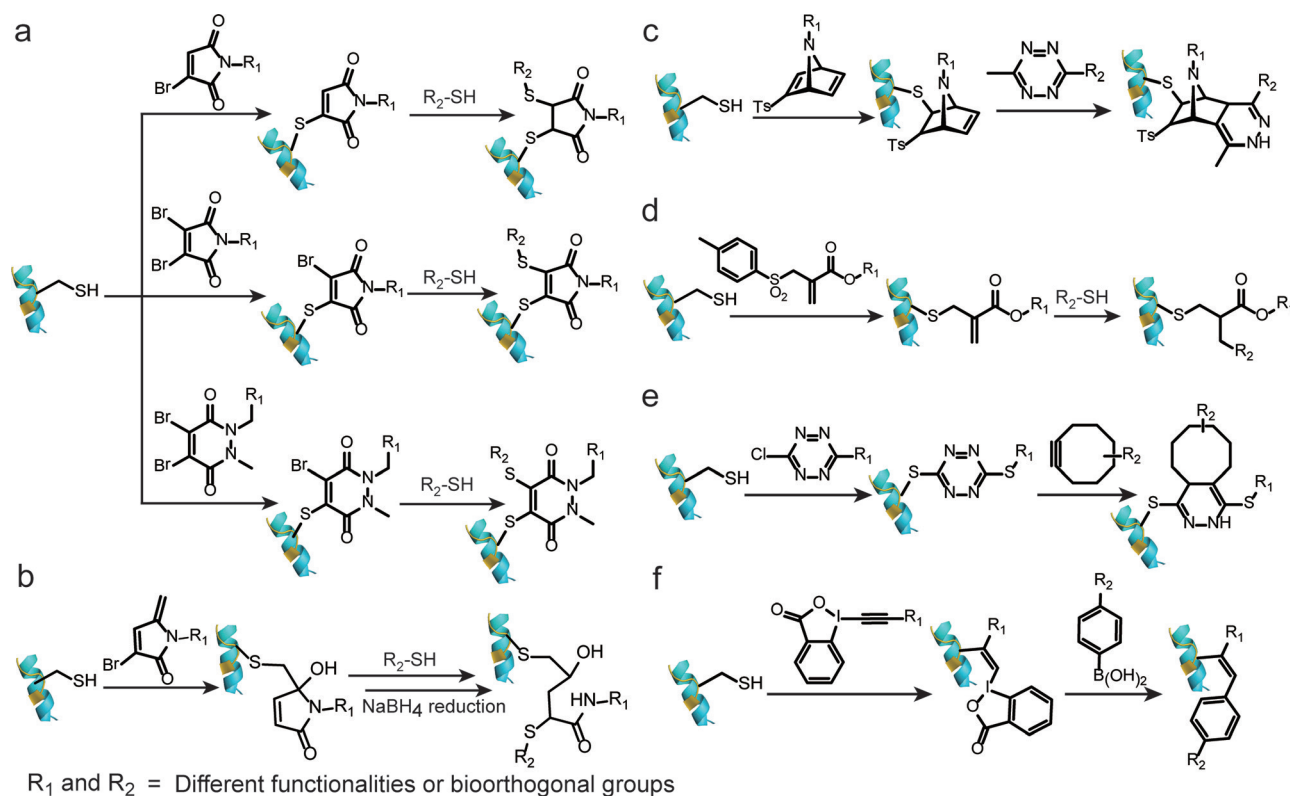


**Figure 6.** Site-selective protein dual modification at a single cysteine site to introduce multifunctional bioconjugation reagents. (a) Mono and dibromomaleimide, dibromopyridazinediones (from top to down)[47,90] (b) 3-Bbromo-5-methylene pyrrolones (3Br-5MPs)[91] (c) Azabicyclic vinyl sulfone[37] (d) Allyl sulfone[38] (e) Dichloro-1,2,4,5-tetrazine[44] (f) Ethynylbenziodoxolones (EBXs).[46]

$NaBH_4$, was required to retard the elimination reaction to generate a stable and bioactive conjugate for subsequent applications.[91]

Alternatively, vinyl sulfones are also commonly explored Michael acceptors for protein modification due to its high electrophilic properties that enable their reaction with nucleophiles on the protein surface.[92] Nevertheless, their application for dual functionalization was hampered by the cross-reactivity with e.g., amino or imidazole groups generating heterogeneous products.[93] Recently, Bernardes and co-workers combined the strained [2.2.1]bicyclic systems with the vinyl sulfone systems and developed the azabicyclic vinyl sulfone reagents for dual functionalization (Figure 6c).[37] Such combination results in a fast chemoselective protein modification at the cysteine site, while the dienophile in the azabicyclic strained moiety concomitantly offers an opportunity for further bioorthogonal modification via iEDDA to liberate the energy stored in the strain systems. The second functionalization could even proceed inside living cells for selective apoptosis imaging. Besides vinyl sulfone, allyl sulfone reagents with enhanced water solubility and higher reactivity have also been proposed as a viable strategy for dual modification in a stepwise fashion.[38] By simply adjusting the pH, allyl sulfone reagents first reacted in a Michael reaction at pH 6 to attach the first functionality, yielding a conjugated ester system that reacted with the second thiol-containing moiety at pH 8 to achieve dual functionalization of proteins (Figure 6d).[38]

In addition to maleimide analogues and sulfone derivatives, other strategies have also been developed for dual-modification. For example, Goncalves and co-workers reported that dichloro-1,2,4,5-tetrazine can undergo two successive nucleophilic aromatic substitutions to introduce the thiol-containing payload at a cysteine residue with excellent selectivity (Figure 6e).[44] The tetrazine linkage could serve as the second handle for subsequent bioorthgonal iEDDA reaction, allowing the preparation of site-selective dual-modified protein conjugates. The feasibility of this strategy has been shown by the dual labeling of the human serum albumin with a macrocyclic chelator for nuclear imaging and a fluorescent probe for fluorescence imaging.[44] Despite the simplicity of this method, it could suffer from lower yield if bulky and hydrophobic functionalities needs to be incorporated. Furthermore, a different strategy utilizing the inherent reactivity of the hypervalent bond was also reported. Waser and co-workers showed the dual modification of proteins with ethynylbenziodoxolones (EBXs) in high efficiency and chemoselectivity by introducing two reactive groups, i.e., an azide and a hypervalent iodine (Figure 6f).[46] Dual modification was achieved via a strain-release-driven cycloaddition and Suzuki–Miyaura cross-coupling of the vinyl hypervalent iodine bond with using palladium diacetate complex as catalyst.

Due to the emergence of ADCs for targeted cancer therapy, dual functionalization at disulfide site has also gained growing interest because of the presence of accessible disulfide bonds in antibodies and the antigen-binding fragment (Fab).[11,12] Previously, dibromopyridazinediones have been extensively employed as versatile reagents for dual

modification at the single cysteine site.[90] Further reports revealed that it can also serve as a disulfide rebridging reagent to introduce two bioorthogonal tags into disulfide-containing proteins, for example, antibodies or antibody fragments. Chudasama et al. exploited the insertion of dibromopyridazinediones bearing two bioorthogonal tags into the disulfide bonds in full antibody and antibody Fab fragments (the antigen-binding fragment) without perturbing the internal disulfide bonds that are vital for activity (Figure 7).[50] Such a plug-and-play platform allowed the introduction of two functionalities via two sequential bioorthogonal reactions in a modular and efficient way, paving the way for the next-generation ADCs.

Besides addressing cysteines and disulfide bonds, other modification strategies at less-explored AA residues have also been reported to expand the existing protein functionalization toolkit. For example, the hypervalent iodine reagents developed by Gaunt and co-workers can be combined with maleimide reagents to achieve dual functionalization at cysteine and methionine sites, which is described in Figure 5a. In addition, the hypervalent iodine reagents have also been demonstrated to show multifaceted reactivity.[59] The electrophilicity of the diazo sulfonium conjugate enables a photoredox radical cross-coupling reaction with C-4 benzylated Hantzsch ester derivatives to attach the second functionality yielding dual functionalized conjugates with high conversion.[59]

Achieving dual modification at one single AA site has less restriction in terms of the choice of reagents compared to the functionalization at two AA sites. One possible limitation of the single site strategy is that the two orthogonal groups can be sterically hindered due to close proximity on a relatively small multifunctional linker. This could prevent bulkier groups such as polymers to be attached onto the protein. In addition, modification besides cysteine residues is relatively unexplored, thus this field would greatly benefit from further investigations of modification strategies at other low abundant AA residues such as tyrosine, serine or the N-termininus.
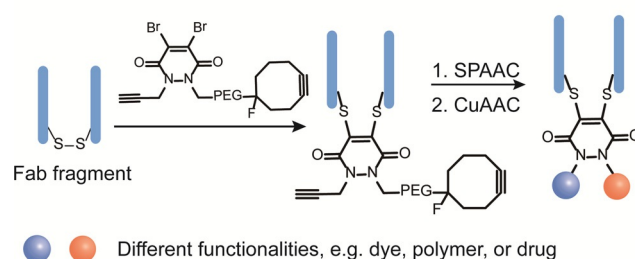


**Figure 7.** Site-selective dual modification of proteins at the disulfide site of antibody Fab fragment.[50]

## 4. Genetic engineering for dual functionalization of proteins

To expand the reactivity beyond what is offered in native proteins, reactive groups for functionalization through genetic incorporation of new AAs (canonical or noncanonical ones)

66

has emerged as an indispensable tool for site-selective protein dual functionalization. In this section, genetic methods to integrate new reactive canonical AAs, ncAAs or peptides tags for protein dual functionalization are summarized.

### 4.1. Incorporation of canonical AAs

The expression of recombinant proteins containing one or more point mutations is a straightforward and well-established technology.[84] Early attempts to achieve dual functionalization of proteins via genetic methods focused on the preparation of dual cysteine mutants of the target protein. Two cysteine mutations were incorporated at two distinct locations, and both thiols exhibited different reactivity toward different thiol-reactive reagents.[94] For instance, Caddick et al. reported the two-cysteine insertions to genetically engineered antibody mimetic proteins, so-called designed ankyrin repeat proteins (DARPins).[94] Both thiols are carefully selected and revealed different nucleophilicity, which may origin from their different solvent accessibility. This allowed for the dual functionalization to be executed in a stepwise fashion with high overall labeling efficiency (Figure 8 a).[94] The authors proposed that the less reactive reagent (bromoacetamide) was reacted first with the more nucleophilic cysteine and subsequently, the more reactive reagent (maleimide) was applied to the second, less nucleophilic cysteine. They successfully demonstrated that the
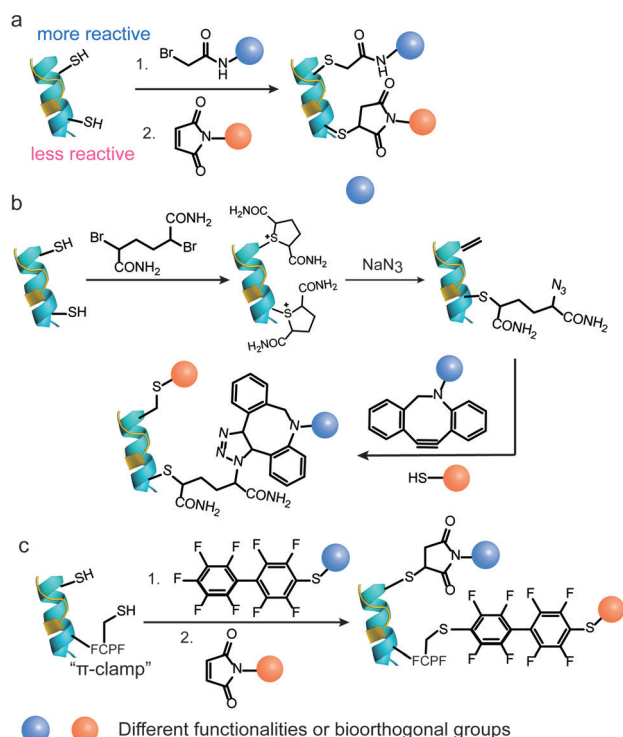


**Figure 8.** (a) Genetic encoding of two cysteine mutations with different nucleophilicity.[94] (b) Genetic encoding of two cysteine mutations possessing different protein local microenvironment.[95] (c) Genetic encoding of a cysteine mutation and a „π-clamp" FCPF peptide sequence creates a new microenvironment for the cysteine residue.[96]

differences in thiol nucleophilicity afforded a homogeneous product with quantitative conversion at each reaction step, and no purification was needed. Although this concept for dual modification of proteins is very elegant, the delicate balance of thiol reactivity and the selection of the most appropriate cysteine mutation site to prevent heterogeneous product formation (single and dual modified products) could be very challenging. Caddick and co-workers also succeeded in protein dual modification by reacting two cysteines with the same thiol-reactive reagent generating two identical sulfoniums.[95] Due to the different accessibilities of the α-protons at the two cysteine mutations, one of the sulfonium groups, which had a good solvent accessible α-proton on adjacent position, underwent a β-elimination reaction affording dehydroalanine (Figure 8 b). In contrast, the other sulfonium group, which has a shielded α-proton next to it, remained intact because of the different protein local microenvironment. The functionalities were incorporated via two different chemoselective reactions, offering the site-selective dual-modified protein conjugate in high yield.

The examples mentioned above took advantage of the different local microenvironment offered by the native protein. Inspired by Nature's elegance, Pentelute and co-workers have developed a strategy to create a specific local chemical environment for the cysteine residue by a newly developed, fine-tuned four-amino-acid peptide sequence (FCPF), which is termed as „π-clamp" (Figure 8 c).[96] This π-clamp enables the conjugation exclusively at this cysteine site with perfluoroaromatic reagents with almost quantitative conversion. The reaction proceeds even in the presence of other competing thiols, thus rendering this approach compatible and complementary to other thiol-conjugation strategies.[96] Dual functionalization was demonstrated on a model protein substrate bearing a cysteine and a π-clamp mutation, which was functionalized with a perfluoroaryl probe first based on the π-clamp-mediated conjugation and followed by the thiol–maleimide conjugation reaction.[96]

### 4.2. Incorporation of ncAAs

The introduction of point mutations has certain limitations as the available functionalities and their respective reactivities could only be selected from the pool of the 20 canonical AAs. However, the cellular biosynthetic machinery can be manipulated to incorporate ncAAs that often represent structurally similar derivatives of the canonical AAs. These ncAAs allow protein labeling with unprecedented molecular precision.[97] To date, a diverse set of ncAAs with various functionalities or bioorthogonal groups has been reported for genetic incorporation into proteins.[3,84] In the auxotrophic strain, organism such as *E. coli* are applied that are not able to synthesize a certain amino acid required for its growth. The ncAAs need to structurally resemble the natural AA to allow binding to the respective endogenous aminoacyl-tRNA synthetase and to replace the natural AA in the polypeptide sequence.[84] This strategy has been mainly employed to introduce azide- or alkyne-containing methionine analogues in a methionine-auxotrophic *E. coli* strain.[84]

For example, the Davis group incorporated the methionine analogue, azidohomoalanine (AHA), and a cysteine mutation to the target protein based on the combination of site-directed gene mutagenesis and the residue-specific replacement of methionine by its analogues (Figure 9 a).[98] Dual modification was accomplished through the initial cysteine conjugation with methanethiosulfonates derivatives, followed by the CuAAC reaction of the azido group in AHA and ethynyl functionalities. This approach offers the benefit of the established AHA incorporation and its efficient modification by cycloaddition reactions. However, the set of available ncAA is limited as they have to bind to their respective tRNA synthetase efficiently, using auxotrophic strains and all the AA residues within a sequence will be replaced by the respective ncAA analogue.

The genetic code expansion technique has been developed as another technique that allows the insertion of a broad variety of ncAAs with spatial precision at virtually any desired site.[97] It is accomplished by using an orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pair, which is capable of charging a designed ncAA in response to a nonsense codon, such as the amber stop codon (UAG), due to
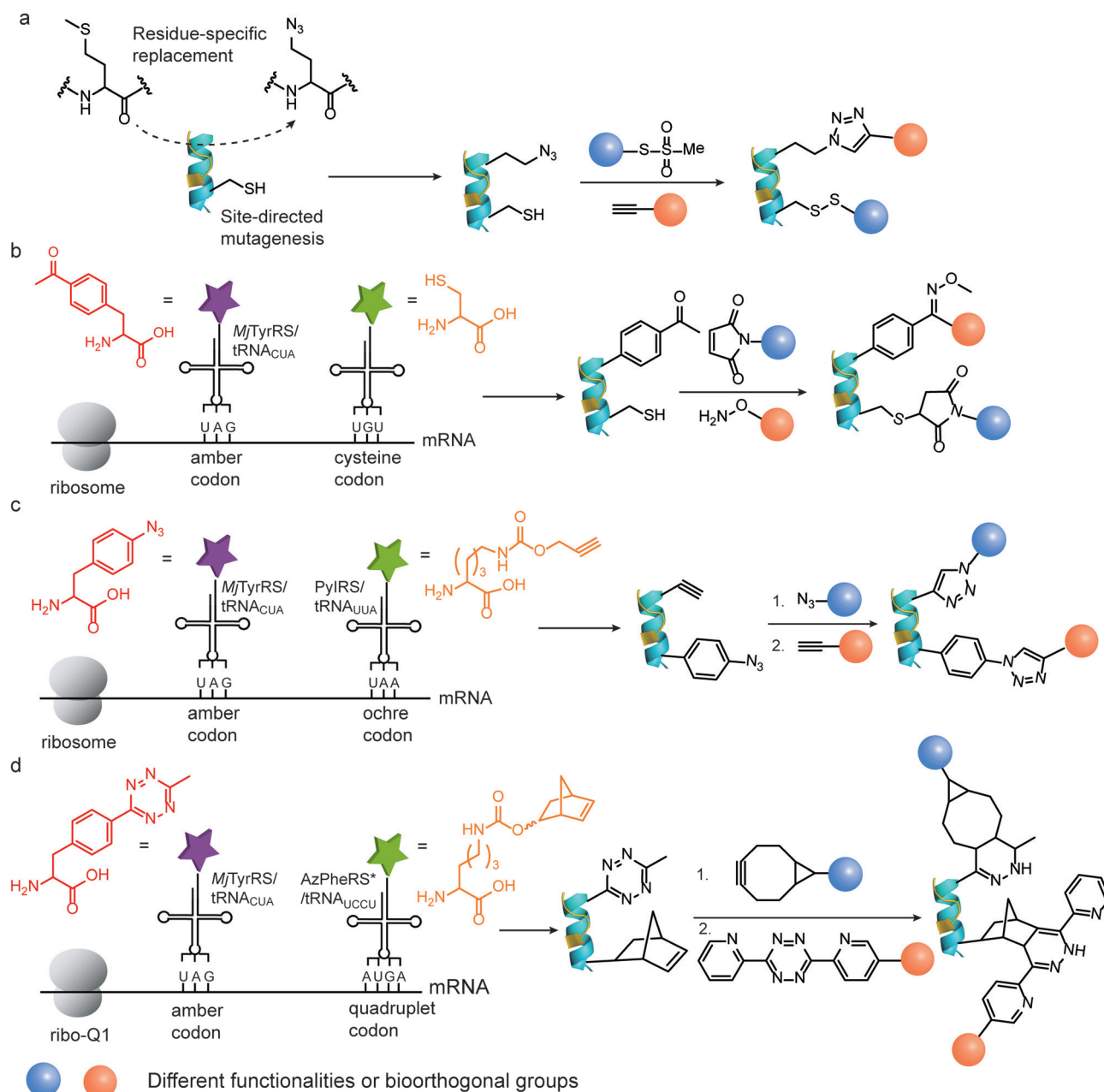


***Figure 9.*** (a) Site-selective incorporation of one cysteine mutation via site-directed mutagenesis and one methionine analogue via residue-specific replacement experiment.[98] (b) Genetic encoding of one cysteine mutation via site-directed mutagenesis and one ncAA via genetic code expansion in response to UAG codon.[60] (c) Genetic encoding of two noncanonical AAs via genetic code expansion in response to UAG and UAA codon.[100] (d) Genetic encoding of two noncanonical AAs via genetic code expansion in response to UAG and AUGA codon.[25]

their minimal occurrence in most organisms.[97,99] This strategy allows the genetic encoding of more than 150 ncAAs containing various reactive handles and functionalities.[84] Therefore, the incorporation of ncAA via genetic code expansion in combination with the site-directed canonical AA mutation provides a versatile strategy for protein dual functionalization. Representative work was reported by Deniz and co-workers, in which they utilized an engineered tyrosyl-tRNA synthetase (MjTyrRS)/tRNA$_{CUA}$ pair (*Mj*TyrRS/tRNA$_{CUA}$ pair) derived from *Methanococcus jannaschi* in response to the amber (UAG) stop codon to insert *p*-acetylphenylalanine, a ketone bearing ncAA, into T4 lysozyme.[101] In combination with a single cysteine mutation, dual labeling of the T4 lysozyme mutant was demonstrated by modification with a FRET dye pair through the thiol–maleimide reaction and oxime ligation (Figure 9b).[101]

Despite the simplicity of introducing a cysteine mutation as one of the target sites for protein dual modification, this approach could become problematic if the protein consists of native cysteines that play important structural or functional roles. Furthermore, cysteine residues could form disulfide bonds, which could limit expression yields and the reaction reversibility of the thiol–maleimide conjugation could also cause stability problems of the resulting protein bioconjugates.[36] Hence, there have been much efforts to introduce two different ncAAs with bioothogonal tags that could be functionalized independently. Liu and co-workers applied two mutually orthogonal aaRS/tRNA pairs in response to two blank codons.[100] They mutated the pyrrolysyl-tRNA synthetase (PylRS)/tRNA$_{CUA}$ pair (PylRS/tRNA$_{CUA}$ pair) to produce a new variant, PyIRS/tRNA$_{UUA}$, which can suppress the ochre (UAA) stop codon. In combination with the evolved *Mj*TyrRS/tRNA$_{CUA}$ pair, these two orthogonal aaRS/tRNA pairs were capable of recognizing and inserting two ncAAs, *p*-azido-ʟ-phenylalanine and N$^\varepsilon$-propargyloxycarbonyl-ʟ-lysine, into a glutamine-binding protein in response to the amber UAG codon and the ochre UAA codon (Figure 9c).[100] Two sequential CuAAC reactions were employed to install two chromophores for the protein dual modification. In a subsequent report, the authors incorporated the azide- and ketone-bearing ncAAs into GFP with the use of an evolved *Mj*TyrRS (AzFRS)/tRNA$_{CUA}$ pair and PyIRS (AcKRS)/tRNA$_{UUA}$ pair, to eliminate protein aggregation and oxidation induced by the copper catalyst.[81] Notably, dual functionalization was accomplished by SPAAC and an oxime ligation in one-pot and catalyst-free fashion. In addition, besides in *E. coli*, Schultz and co-workers have also successfully incorporated two different ncAAs, which contained azido and ketone groups, in mammalian cells by utilizing the two orthogonal *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase (MbPylRS)/*Methanosarcina mazei* pyrrolysyl tRNA (*Mb*PylRS/*Mm*tRNA$_{UUA}$ pair) and tyrosyl-tRNA synthetase (EcTyrRS)/tRNA$_{CUA}$ pair (*Ec*TyrRS/tRNA$_{CUA}$ pair).[102] The dual-tagged antibody was subsequently functionalized with a toxic drug payload and a fluorophore with high conversion.

Alternatively, instead of reassignment of the triplet stop codons, Chin and co-workers have exploited the two orthogonal aaRS/tRNA pairs in response to a quadruplet blank codon (four-base codon) and a stop codon for the incorpo-

ration of two ncAAs.[103] However, natural ribosomes suffer from very low efficiency in decoding the quadruplet codon. In this context, Chin et al. have synthetically evolved an orthogonal lysozyme (ribo-Q1), which was not responsible for synthesizing the proteome as natural lysozyme, for selectively decoding the quadruplet codon.[79] By the combination of ribo-Q1 with two orthogonal aaRS/tRNA pairs, AzPheRS*/tRNA$_{UCCU}$ (a derivative of *Mj*TyrRS/tRNA) and PylRS/tRNA$_{CUA}$, two ncAAs were site-selectively introduced into Calmodulin, forming a triazole intramolecular crosslink through the subsequent CuAAC reaction.[79] Nonetheless, the major drawback of this system lies in the low efficiency and specificity of the AzPheRS*/tRNA$_{UCCU}$ pair in directing the corresponding ncAA incorporation. The efficiency of the original system was substantially improved based on the evolution of the PylRS/tRNA$_{CUA}$ to obtain an optimized quadruplet decoding variant, PylRS/tRNA$_{UACU}$ pair.[25] As a proof-of-concept, the evolved PylRS/tRNA$_{UACU}$ pair was combined with the AzPheRS*/tRNA$_{CUA}$ pair to site-selectively introduce two bioorthogonal tags, norbornene and tetrazine, into Calmodulin (Figure 9d).[25] Dual modification was successfully accomplished via two sequential iEDDA reactions.[25] With the established technologies, in the latter example, the authors incorporated the alkyne and cyclopropane handles into Calmodulin by using the orthogonal aaRS-tRNA pair described above, permitting the simultaneous dual functionalization in one pot.[82]

Genetic code expansion has witnessed incredible achievements in recent years, which greatly promotes and revolutionizes the field of site-selective protein dual functionalization. The advantages of this strategy lie in the small size of the ncAAs, flexible incorporation sites but high site-selectivity, and the diverse reactive tags available for incorporation. However, some important challenges still remain in this rapidly evolving field, for example, low catalytic efficiency of engineered aaRS which require tedious evaluation and optimization to improve their performance, relative low expression yields, repetitive optimization of the protocols for effective protein engineering, the scope and compatibilities of the functional groups capable to be inserted.[84] Besides developing the recombinant engineering techniques, the bioorthogonal chemistry toolbox can also be expanded so that there are more choices for compatible bioorthogonal reaction pairs in terms of reaction rate, catalyst type, and substrate solubility for their intended applications.

### 4.3. Incorporation of peptide tags

Besides the incorporation of ncAAs bearing bioorthogonal tags for dual modification of proteins, the insertion of two artificial peptide sequences that can be recognized by two distinct enzymes for the subsequent covalent labeling with user-defined probes also serves as a straightforward approach for dual modification of proteins. This enzyme-mediated peptide labeling combines the advantage of high specificity of the enzyme towards the peptide tags and excellent labeling efficiency with minimal perturbation to proteins structure and function.[104] A brief summary of the commonly used peptide

tags and their respective enzymes is given in Table 2 and relevant examples for dual modification are further discussed in this section. In a recent example, two distinct enzymes, Sortase A and Butelase 1, which demonstrate orthogonal specificity to LPXTGG and HNV motif, respectively, have been combined for dual modification of IgG antibody (Figure 10a).[118] Notably, a simple centrifugation process is sufficient to obtain the pure dual-modified protein conjugate, demonstrating the high selectivity and excellent conversion of the enzyme-mediated labeling approach. In addition, peptide tags have also been utilized in combination with ncAAc[105] or „π-clamp" sequence[106] for site-selective dual-modification of proteins. For example, Chen and co-workers have demonstrated the incorporation of a LAP peptide (LplA acceptor

**Tabelle 2:** Commonly used peptide tags and their respective enzymes.

| Tag sequence[a] | Enzyme |
| --- | --- |
| GLNDIFEAQKIEWHE[108] | Biotin ligase[108] |
| LPXTGG[109] | Sortase A[109] |
| NHV[110] | Butelase A[110] |
| CaaX[b][111] | Farmesyl transferase[111] |
| LLQG[112] | Microbial transglutaminase[112] |
| GFEIDKVWYDLDA[113] | Lipoate acid ligase A[113] |
| CXPXR[114] | Formylglycine Generating Enzyme[114] |
| GDSLSWLLRLLN[115] | Phosphopantetheinyl transferase[115] |
| VDSVEGEGEEEGEE[116] | Tubulin tyrosine ligase[116] |
| AHIVMVDAYKPTK(Spytag)[117] | SpyCatcher[117] |

[a] The capital letters are the abbreviation of different amino acids. „X" represents any amino acids. [b] „a" represents small aliphatic amino acids and „X" denotes one of Ala, Ser, Met or Glu residues amino acids in CaaX motif.

peptide), which can be recognized by a lipoic acid ligase (LpIA) to ligate a lipoic acid derivative, and an pyrrolysine analogues which bear an azido group for the site-selective dual labeling of epidermal growth factor receptor (EGFR) on living cells.[105]

Despite the simplicity of a one-step introduction of the desired functionalities by two independent reactions, the labeling efficiency may be greatly influenced by the steric effects and hydrophobicity of the introduced probes. This limitation can be overcome by the incorporation of a multifunctional scaffold possessing the bioorthogonal tags. Distefano et al. used farnesyl transferase, which can catalyze the transfer of an isoprenoid group from farnesyl diphosphate to the cysteine site of a tetrapeptide (CVIA, the letter code is the abbreviation of a specific AA), to introduce two reactive tags to the model protein, GFP (Figure 10b).[107] The bifunctional alkyne-aldehyde modified protein can undergo two independent bioorthogonal reactions simultaneously, offering the dual-modified conjugate in good conversion.

In some instances, the enzyme-mediated dual labeling of proteins can also be achieved on the native protein surface in a site-selective manner without the need to insert artificial peptide tags. For example, in Alabi and cowokers' work, microbial transglutaminase, which catalyzes the formation of interprotein isopeptide bonds between glutamine and lysine residues, can recognize the glutamine 295 on the aglycosylated human IgGs and thus allowed the incorporation of two bioorthogonal tags, azido group and methyltetrazine group (Figure 10c).[80] The combination of SPAAC and iEDDA enables the generation of a dual-modified antibody therapeutic containing a cytotoxic payload along with a hydro-



a. Combination of Sortase A and Butelase 1

b. Farmesyl transferase

c. Microbial transglutaminase

Different functionalities, e.g. dye, polymer, or durg

**Figure 10.** Dual modification of proteins by using (a) the combination of Sortase A and Butelase A.[118] (b) Farmesyl transferase through the incorporation of a dual-functional scaffold bearing an alkyne and an aldehyde tag.[107] (c) Microbial transglutaminase through the incorporation of a dual-functional scaffold bearing a tetrazine and an azido tag.[80]

phobicity-masking PEG side chain in a mix-and-match manner. Nevertheless, this strategy is greatly limited by the protein substrate, which requires the recognition motif to be present on the native protein surface at a specific site.

## 5. Higher-level functionalization of proteins

Driven by exciting developments in the discovery of new synthetic procedures as well as scientific curiosity, achieving higher level of protein modification in a site-selective manner is considered crucial for advancing different fields in chemistry, biology and material science. Compared to mono- and dual functionalization of proteins, triple functionalization strategies allow the incorporation of three different functionalities into proteins, thus offering well-defined protein bioconjugates with further expanded structural and functional diversities that is beyond what current synthetic strategies can accomplish. These endeavors will greatly boost our capability to investigate, modulate and re-design the chemical and physical properties of proteins.

However, triple functionalization of proteins is even more demanding than mono- and dual functionalization as the combination of three orthogonal chemistries have to be applied in aqueous solution and in the presence of many reactive proteinogenic groups. Furthermore, by increasing the number of functionalities attached to the protein surface, it may lead to potential detrimental influence to their functional and structural integrity. In addition, the selection of the reactive groups requires more deliberate chemical design as the increase in the hydrophobicity of the bioconjugation reagents may cause protein aggregation. The labeling efficiency may also be compromised due to the steric hindrance from the insertion of three different functionalities.

Recent examples have demonstrated the successful triple functionalization of proteins. Chatterjee and co-workers have reported the introduction of three ncAAs into proteins via genetic code expansion. In this work, they assigned the EcTrp, *Mj*Tyr, and Pyl pairs to suppress UGA, UAG and UAA codons for the incorporation of three ncAAs, 5-hydroxytryptophan, *p*-azidophenylalanine, and cyclopropene-lysine, in the engineered *E. coli* strain ATMW1 (Figure 11 a).[119] The triple functionalization of the protein was successfully achieved by three mutually compatible reactions, SPAAC, iEDDA, and chemoselective rapid azo-coupling reaction (CRACR), in which the electron-rich 5-hydroxyindole ring in 5-hydroxytryptopan reacts with electron-deficient aromatic diazonium ions with fast reaction kinetics and high conversion.[120] This is the first example of triple modification of proteins where three ncAAs were incorporated into the target protein in living cells, which further expands the simultaneous ncAAs coding capacity via genetic code expansion. However, this strategy utilized three different aaRS/tRNA pairs and three stop codons, which leaves no codon for termination of endogenous genes and therefore greatly interferes with the translation termination inside cells. Shortly after, Chin and co-workers have demonstrated the genetic encoding of three distinct ncAAs into proteins with three aaRS/tRNA pairs in a different strategy.[121] In their work, they identified new

ΔNPyIRS/^ΔNPyI^tRNA pairs, which lack the N-terminal domains, and revealed that they can be assigned into two classes, class A and B, based on their sequences. Specifically, class A ΔNPyIRS preferentially function with class A ^ΔNPyI^tRNA and vice versa. Next, they discovered a *Mm*PyIRS/*Spe* ^PyI^tRNA pair, in which *Spe* ^PyI^tRNA is orthogonal to both class A and B ΔNPyIRS. In this context, the triply orthogonal aaRS/tRNA pairs were identified, which contained the *Mm*PyIRS/*Spe* ^PyI^tRNA pair, an evolved class A ΔNPyIRS/^ΔNPyI^tRNA pair and an evolved class B ΔNPyIRS/^ΔNPyI^tRNA pair. Three ncAAs, Nε-((tert-butoxy)-carbonyl)-L-lysine, 3-methyl-L-histidine and Nε-(carbobenzy-loxy)-L-lysine, were genetically encoded into proteins in response to the UAG, AGGA and AGUA codons. However, subsequent functionalizations have not demonstrated yet. Compared to the reassignment of three stop codons, this work utilizes the combination of a stop codon and quadruplet codons, which do not interfere with the translation termination and serve as a more general and applicable strategy to achieve higher-level protein functionalization. Even though there is no application shown and there are still obvious limitations, for example, low encoding efficiency, these examples represent the first proof-of-concept for protein triple modifications. They still represent significant technological advancements for the development of multifunctional protein conjugates, which holds immense potential for drug delivery, bioimaging, and material science.

Besides increasing the number of functionalities that can be incorporated into proteins, more complex protein bioconjugates have also been reported. For example, Chudasama et al. described an efficient and modular strategy for the generation of a bispecific antibodies as well as the dual functionalization of the resulting bioconjugate (Figure 11 b).[122] By fine-tuning the bioorthogonal chemistry employed, two dual-modified antibodies were first conjugated together via the iEDDA reaction of the two bioorthogonal handles from each antibody (Figure 11 b). Subsequently, two other bioorthogonal handles from each antibody allow for the dual functionalization of the chemically constructed bispecific antibodies. Despite these seminal studies, the preparation of protein conjugates with higher level of modification or structural complexity remains relatively unexplored. We envision that with technical breakthrough in methodology, exciting yet unexplored applications will be discovered, which will undoubtedly revolutionize fundamental studies and applications of proteins.

## 6. Applications

With the emergence of protein dual functionalization techniques, homogeneous protein bioconjugates with remarkable functional complexity have been achieved and studied for diverse applications. In the following, three major fields of applications are highlighted that have greatly benefited from these novel synthesis opportunities.
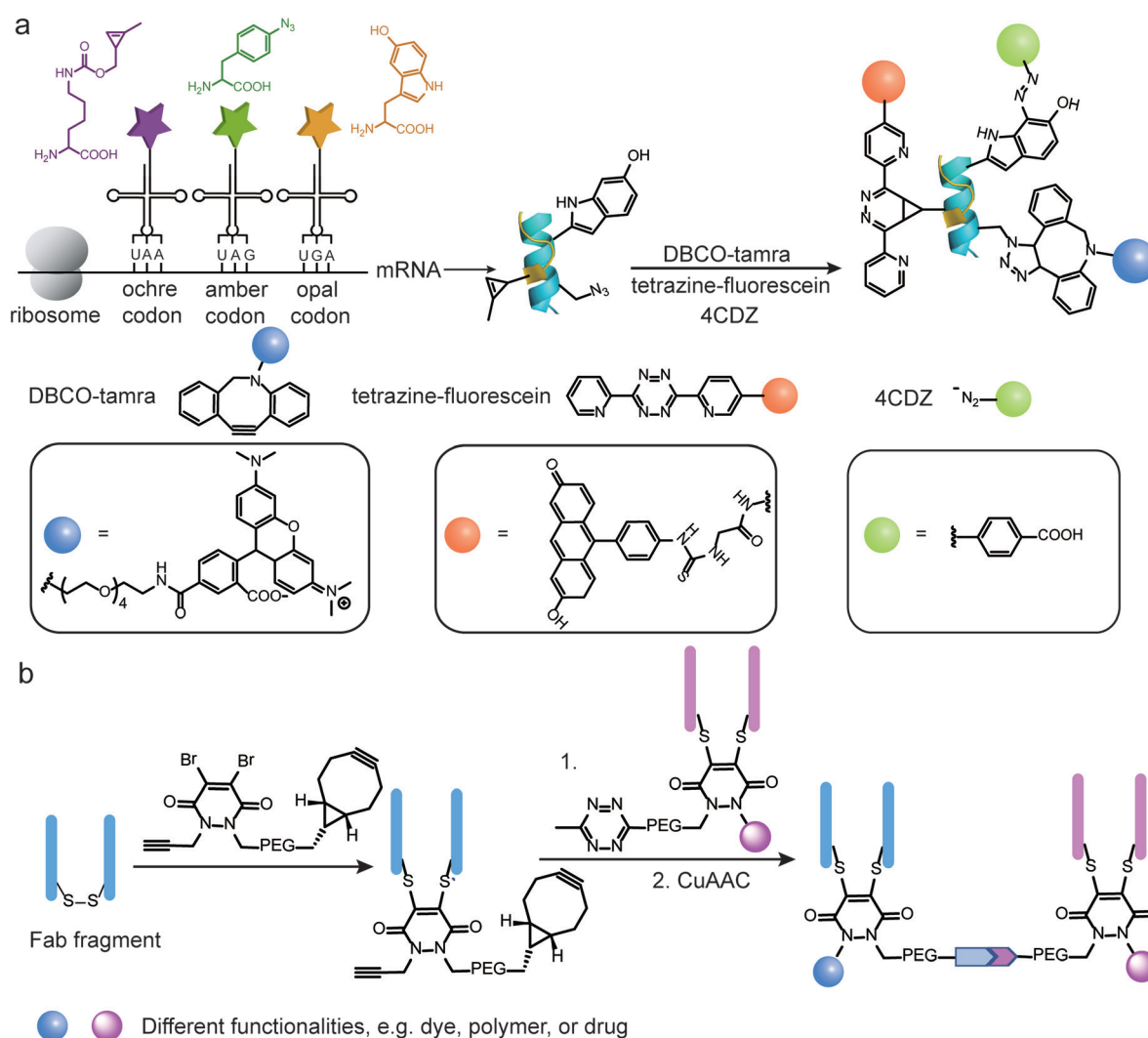
**Figure 11.** (a) Genetic incorporation of three ncAAc into protein via genetic code expansion in response to the UAA, UAG and UGA codons for tri-functionalization of proteins.[119] (b) Preparation of antibody bispecifics by the conjugation of two dual-modified antibody Fab fragments.[122]

### 6.1. Probing protein dynamics through Förster resonance energy transfer

The excitation energy transfer from a donor to an acceptor chromophore, also termed Förster resonance energy transfer (FRET), allows assessing important parameters such as chromophore distances and their orientations in real time and in complex biological environments. Although cryo-electron microscopy is becoming a popular technique to determine structural snapshots of biomacromolecules at atomic resolution, samples need to be frozen and studies cannot be carried out in native environment. In contrast, FRET bears the advantage of easy accessibility and enables the real-time monitoring of the dynamics and conformational changes of proteins in their native environment, thus serving as an important complementary methodology for elucidating proteins structural dynamics.[123] A critical requirement for FRET studies is the site-selective incorporation of the respective donor and acceptor chromophores at pre-defined

locations on the protein surface without interfering with the proteins' structure and function.[123]

Fluorescent proteins (FPs), for example, cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP), can be genetically encoded into the target proteins and are widely utilized for the investigation of proteins inside living systems.[15] For example, agonist-induced activation of G-protein coupled receptor (GPCRs) is thought to cause a conformational rearrangement of their seven transmembrane α-helices.[124] The insertion of CFP and YFP to the G-protein coupled receptors (GPCR) offered the dual tagged protein bioconjugate, which allows real-time monitoring of the activation switch of GPCRs in living cells.[125] However, due to the large size of FPs (molecular weight close to 30 kDa), the CFP-YFP FRET pair was found to perturb the proteins structure and function and impair the downstream signal transduction.[125] In order to overcome these limitations, the self-labeling peptide tags have been developed, which can bind or react with small-molecule reagents to attach the fluorophore to target proteins. For example, tetracysteine-

72
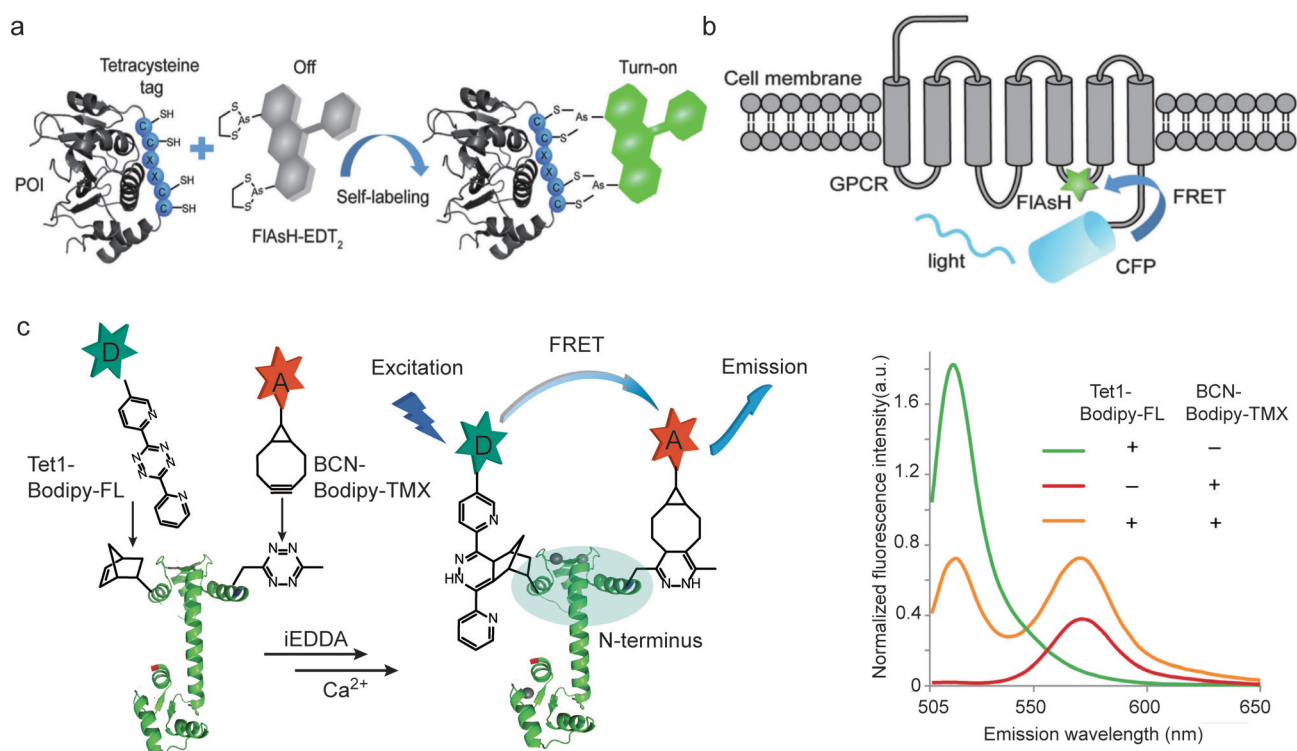
**Figure 12.** (a) The self-labeling peptide of tetracysteine reacting with FlAsH forming fluorescent conjugate. Adapted with permission from ref. [15]. Copyright (2015) Royal Society of Chemistry (b) Incorporating the chromophore and CFP as FRET pair to GPCR to probe protein dynamics. Adapted with permission from ref. [15]. Copyright (2015) Royal Society of Chemistry (c) Incorporation of FRET pair into Calmodulin (CaM) via genetic code expansion technique and the corresponding fluorescence spectra after addition of $Ca^{2+}$. Adapted with permission from ref. [25]. Copyright (2014) Springer Nature.

containing motif CCXXCC (C is the abbreviation of cysteine and XX can be virtually any AA sequence), can selectively react with fluorescein arsenical hairpin binder (FlAsH) to form fluorescent product (Figure 12a).[126] Lohse and co-workers used the FlAsH-tetracysteine system to label the human adenosine $A_{2A}$ receptor in a site-specific fashion, which, in conjunction with a CFP, generated a FRET construct ($A_{2A}$-FlAsH-CFP) that exhibited a 5-fold enhanced agonist-triggered FRET signal compared to the dual-tagged CFP and YFP conjugate (Figure 12b).[127]

In comparison to FPs, small molecule chromophores bear the advantage of small size, excellent photostability and high quantum yield.[128] So the incorporation of two chromophores into target proteins as FRET pair have represented as a more advantageous strategy to probe protein dynamics without interfering their structure and functions. One representative example is the introduction of two fluorophore into a calcium-binding protein, Calmodulin (CaM), via genetic code expansion to study its conformation changes as a function of the calcium ion ($Ca^{2+}$) concentration. CaM consists of two domains, the C-terminal and N-terminal domain that contain two EF-hand motifs to bind $Ca^{2+}$.[129] Upon $Ca^{2+}$binding, the conformation of CaM changes, which affects its binding to different target enzymes.[129]

The structures of CaM with four $Ca^{2+}$ and without $Ca^{2+}$ have been resolved in detail via NMR or X-ray crystallography. However, the transition states that dynamically occur during the $Ca^{2+}$ binding processes and that play an important role in modulating the interaction of CaM with its binding partners could not be resolved with conventional structure analysis by NMR and X-Ray.[130] Therefore, dual-tagged CaM was obtained by genetic code expansion and conjugated with the FRET pairs, BODIPY-TMR-X (Bodipy-tetramethylrhod-amine, acceptor) and BODIPY-FL (4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-propionic acid, donor), respectively, to investigate the local conformation changes occurring in the N-terminal domain of CaM in response to the $Ca^{2+}$ concentration (Figure 12c).[25] After excitation at 485 nm, the dual chromophore tagged CaM reveals two distinct signal corresponding to the emission of the BODIPY-FL donor at 515 nm and the BODIPY-TMR-X acceptor at 570 nm. With increasing $Ca^{2+}$ concentrations, the emission intensity of the acceptor dye decreased while the emission of donor dye increased indicating that the FRET pair at the N-terminal domain moved apart as a response of $Ca^{2+}$ binding. The current applications utilizing FRET to follow protein dynamics are mainly applying a certain stimulus in test tubes rather than measurements inside cells. In the future, continuous advancements in dual modification of proteins could provide valuable tools to shed light on the dynamics of proteins in living cells or even in organisms. There is much interest and progress in probing protein labeling reactions proceeding in living environments. How-

ever, no dual labeling reactions have been achieved inside cells that would allow FRET studies until now.

### 6.2. Combination therapy

ADCs have emerged as one of the most powerful and promising strategies for targeted cancer therapy.[12] This new class of biopharmaceuticals combines the exquisite target-specificity of antibodies with a highly potent cytotoxic payload via various conjugation technologies and enables the selective delivery of the payload to cancer cells with minimal off-target effects.[11,12] Owing to the elegant concept, there are already four ADCs approved by the US Food and Drug Administration on the market and over 60 ADCs are in clinical trials, demonstrating its great prospect for targeted therapeutic applications.[12]

Despite the clinical success, most ADCs are still restricted to the conjugation with a single type of payload, for example a certain anticancer drug, to target a specific type of cancer cells. In fact, this kind of single-functionalized ADCs often suffer from the unfavorable pharmacokinetics and the inherent hydrophobicity of the chemotherapeutic drugs, which could significantly limit the in vivo therapeutic efficiency.[11] Dual modification of antibodies can serve as an elegant strategy to alleviate these drawbacks, allowing for the incorporation of two complementary modalities to achieve synergistic effects. This can be exemplified by the studies on extending the plasma circulation half-life and attenuating immunogenicity of proteins through the attachment of hydrophilic polyethylene glycol (PEG) chain to proteins. The hydrodynamic radius of the biomolecule also increases after attaching a long PEG chain, thereby mitigating

the renal filtration and circumventing its degradation by proteases and recognition by the immune system, which would otherwise lead to faster plasma clearance.[131] In this context, dual modification of antibodies allows the simultaneous introduction of a cytotoxic drug molecule and a PEG chain in a site-selective manner to afford well-defined dual-functional ADCs with improved pharmacokinetics. Senter and co-workers have reported that the accelerated plasma clearance of ADCs originated from their hydrophobicity, which could be reduced by introducing a PEG chain with different configuration.[132] As such, three different drug linkers were designed, in which the first drug linker **4** had no PEG chain while both the drug linker **5** and **6** consisted of a unfunctionalized PEG chain ($PEG_{24}$, containing 24 repeating units) but with different configuration (Figure 13a).[132] Specially, drug linker **6** incorporated a branched scaffold. The cAC10 antibody (CD30-directed antibody) was conjugated with the three different drug linkers via thiol–maleimide chemistry offering three different ADCs, cAC10-4 (without PEG24), cAC10-5 (with PEG24) and cAC10-6 (with branched PEG24) as depicted in Figure 13b.[132] Both the in vitro and in vivo experiments revealed an inverse correlation between hydrophobicity and tumor volume (Hodgkin's lymphoma), namely that cAC10-6 exhibited slower plasma clearance and much slower tumor growth compared to cAC10-4 and cAC10-5 (Figure 13c,d).[132] In particular, the improved pharmacokinetics and therapeutic efficiency of the branched conjugate cAC10-6 compared to the linear conjugate cAC10-5 highlights the importance of the hydrophilic linkers for optimizing the pharmacokinetic parameters.

On the other hand, the therapeutic efficacy ADCs could also suffer from drug resistance after multiple treatments due to inherent and acquired resistance or tumor heterogenic-



**Figure 13.** (a) Chemical structure of the three different drug linkers (**4**, **5** and **6**). (b) Preparation of the corresponding ADCs (cAC10-4, cAC10-5 and cAC10-6). (c) Plasma pharmacokinetics in rats of cAC10 ADCs (cAC10-4, cAC10-5 and cAC10-6), (data are given as means ± standard deviations.; $n = 3$ animals per group). (d) Antitumor activity of cAC10 ADCs in the Hodgkin's lymphoma model L540cy (data are given as means; $n = 6$ animals per group) at a single dose of 1 mg kg$^{-1}$ (cAC10-4 and −6) or 2 mg kg$^{-1}$ (cAC10-5). (c, d) Adapted with permission from ref. [132]. Copyright (2015) Springer Nature.

74

**Figure 14.** (a) The chemical structure of two different drug linker (mc-Val-Cit-PABC-MMAF and mc-Val-Cit-PABC-MMAE). (b) ADCs with the incorporation of two different drug molecules: MMAE and MMAF. (c) *In vitro* data of antitumor activity of cAC10-(2+3) on MDR(+)DEL-BVR cell lines. The values are reported as $IC_{50ng}$/mLof ADC. ALCL=anaplasticlarge cell lymphoma. MDR=multidrug resistance. DEL-BVR is a cAC10-vc-MMAE resistant cell line, which is generated after prolonged exposure of the DEL anaplastic large cell lymphoma (ALCL) cell line to cAC10-3 (d) In vivo data of antitumor activity of cAC10-(2+3) on MDR(+)DEL-BVR cells lines. (c, d) Adapted with permission from ref. [27] under the Creative Commons CC BY-NC license. Copyright (2017) Wiley-VCH.

ity.[12] Although the exact mechanism of drug resistance is still under much investigation, current available clinical data indicate that malignant cells, which are resistant to a particular drug, could still respond to other drugs.[19] Therefore, the attachment of two drugs with different modes of action to an antibody can afford more complex ADCs. This could be co-delivered to cancer cells to overcome drug resistance,which represents an emerging field in targeted cancer therapy.[18] For example, Levengood and co-workers have reported a dual-conjugation strategy for the preparation of novel ADCs including two complementary drugs.[27] In their work, two different auristatin molecules, monomethyl auristatin E-(MMAE) and monomethyl auristatin F (MMAF), were selected to conjugate to the cAC10 antibody due to their complementary physicochemical properties and anticancer activities (Figure 14 a,b).[27] Specifically, MMAE is cell-per-meable and exhibits bystander activity capable of killing neighboring antigen-negative cells.[133] However, MMAE is a substrate for multidrug efflux pump exporters, hence showing diminished activity on cells with high pump expression.[134] In contrast, MMAF is susceptible to drug export, but not cell-permeable and exhibits negligible bystander effect.[133] In vitro and in vivo experiments demonstrated that the dual-auristatin ADCs were active on cells and tumors that were refractory to treatment with either of the individual component drugs (Figure 14 c,d). This work highlights the potential of dual modification of antibody for delivering two synergistic and complementary drug payloads for improved antitumor activities, which represents a notable advancement of the ADCs technology.

Furthermore, dual modification of antibodies also provides great opportunities for simultaneous therapy and diagnostic applications, so-called theranostics, in which the antibody is conjugated to an anticancer drug and a contrast

agent to allow tumor diagnosis and therapy at the same time. In one representative example, Zeglis and co-workers reported the dual labeling of a HER2-targeting trastuzumab with a toxin, MMAE, and a positron-emitting radiometal 89Zr for theranostic applications (Figure 15 a).[26] The in vivo experiment indicated that the resulting 89Zr-trastuzumab-MMAE bioconjugate demonstrated excellent tumor targeting and therapeutic efficacy (Figure 15 b). Importantly, the dual labeled ADCs represents a targeted drug delivery system that could be tracked in vivo using PET providing information of the in vivo biodistribution and real-time drug doses during the treatment (Figure 15 c).

### 6.3. Dual-modality imaging

Molecular imaging is a powerful and invaluable tool for noninvasive visualization of physiological processes that occur in living organisms at cellular levels.[5] Until now, various modern imaging technologies, including optical imaging (OI), magnetic resonance imaging (MRI), positron emission tomography (PET) and computed tomography (CT), have been developed and widely used to monitor the structural, functional and dynamic changes in cancer tissues.[135] Each imaging modality has its own unique strength and intrinsic limitations. Consequently, combining two modal imaging methods has received increasing attention, as it allows the collection of complementary imaging data, thereby improving the reliability and accuracy of the diagnosis. For example, PET can provide real-time images of tumor lesions as well as monitor their whole-body distribution and migration, while optical imaging can provide high-resolution imaging to support surgeons in identifying tumor margins during surgical resection.[136, 137] By combining both modalities

75
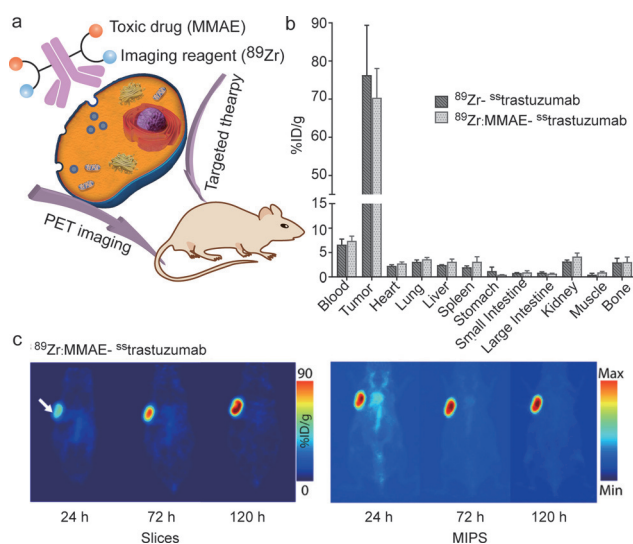
**Figure 15.** (a) Incorporation of a toxic drug (MMAE) and imaging reagent ($^{89}$Zr) for targeted theranostic applications; (b) Biodistribution data of athymic nude mice bearing HER2-expressing BT474 breast cancer xenografts after 120 h of administration of the corresponding bioconjugate; (c) PET images of athymic nude mice bearing HER2-expressing BT474 breast cancer xenografts after the injection of $^{89}$Zr-trastuzumab-MMAE bioconjugate. (b, c) Adapted with permission from ref. [26]. Copyright (2018) American Chemical Society.



**Figure 16.** Incorporation of radiometal and $^{18}$F and sCy5 into A2cDb diabodies for PET imaging and optical imaging. Adapted with permission from ref. [138]. Copyright (2019) Society of Nuclear Medicine and Molecular Imaging.

into a single imaging agent, doctors are able to assess the extent of the disease before and after surgery by PET imaging, meanwhile fluorescence imaging can be utilized for image-guided surgery.[135] Site-selective dual modification of monoclonal antibodies provides an elegant chemical platform to implement the combination of a radionuclide and a fluorescent dye for dual modality PET and fluorescence imaging. One representative example focused on the conjugation of $^{18}$F and far-red dye sulfonate cyanine 5 (sCy5) to anti-prostate stem cell antigen (PSCA) cysteine mutated diabody A2 via a dual-modality linker, offering the dual-modified imaging probe ($^{18}$F- sCy5-A2cDb) (Figure 16).[138] $^{18}$F-immuno-PET showed fast and specific tumor uptake of prostate cancer xenografts, suggesting high-contrast whole-body images with organ-level biodistribution as early as 1 h after injection (Figure 16).[138] Postmortem optical imaging confirmed high-contrast fluorescence in the PSCA-expressing tumors and excellent delineation of cancerous cellsfrom surrounding tissue (Figure 16). The dual-modality imaging provides complemental data, significantly contributing to the reliable and accurate diagnostic applications.

## 7. Conclusion and Outlook

The technical breakthroughs in protein bioconjugation chemistry has served as a major driving force for the elucidation of protein trafficking and interactions in living cells as well as the emergence of protein therapeutics. With the increasing need for personalized treatments to address great challenges in biomedical research such as efficient cell or organ targeting, overcoming drug resistance and reducing

systemic toxicity, there is a high demand for protein therapeutics with improved features. Multifunctional protein conjugates provide more than one functionality, and they are expected to enhance in vivo performance of therapeutic proteins by combining different functional groups possessing therapy, diagnostics and imaging properties in a single protein system that is tailored to the requirements of the patient. Nevertheless, dual functionalization of proteins remains much more challenging relative to single functionalization owing to the plethora of reactive functional groups on the protein surface and the requirement for the optimal combination of different orthogonal reactions with high efficiencies. In this review, we have summarized the remarkable progress in both synthetic and genetic engineering strategies that have overcome these significant hurdles and thus allow covalent functionalization of proteins with two different functionalities at distinct sites.

Undoubtedly, dual functionalization of proteins has witnessed significant advancements over the past ten years which offers a new arsenal of functional protein conjugates with advantages over singly-functionalized proteins in probing protein dynamics, combination therapy and bioimaging. We envision that tremendous efforts will continue to be devoted to enriching the current methodology toolkit by exploring new chemistries to improve the specificity as well as efficiency of dual functionalization or to achieve a higher level of functionalization, for example, triple functionalization. Moreover, the proper combination of different sophisticated strategies, for example, chemical methods, genetic methods, to capitalize the advantage of different approaches

from the currently expanding toolbox will provide new insights for the preparation of a myriad of functional nanomaterials. Ultimately, we believe that this will pave the way towards „smart" and „intelligent" protein conjugates that can self-adapt to the microenvironment of diseases and provide a self-feedback loop to achieve an output or decision, for example, termination/activation. In this way, current bottlenecks and challenges in therapeutic applications can be addressed with a completely new perspective through rationale chemical design.

## Conflict of interest

The authors declare no conflict of interest.

[1] M. Karplus, J. Kuriyan, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6679–6685.

[2] O. Shlyk-Kerner, I. Samish, D. Kaftan, N. Holland, P. S. Sai, H. Kless, A. Scherz, *Nature* **2006**, *442*, 827–830.

[3] C. D. Spicer, B. G. Davis, *Nat. Commun.* **2014**, *5*, 4740.

[4] C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto, Jr., *Angew. Chem. Int. Ed.* **2005**, *44*, 7342–7372; *Angew. Chem.* **2005**, *117*, 7508–7539.

[5] N. Krall, F. P. da Cruz, O. Boutureira, G. J. L. Bernardes, *Nat. Chem.* **2016**, *8*, 103–113.

[6] E. A. Hoyt, P. M. S. D. Cal, B. L. Oliveira, G. J. L. Bernardes, *Nat. Rev. Chem.* **2019**, *3*, 147–171.

[7] O. Boutureira, G. J. L. Bernardes, *Chem. Rev.* **2015**, *115*, 2174–2195.

[8] X. Chen, Y. W. Wu, *Org. Biomol. Chem.* **2016**, *14*, 5417–5439.

[9] N. C. Reddy, M. Kumar, R. Molla, V. Rai, *Org. Biomol. Chem.* **2020**, *18*, 4669–4691.

[10] S. L. Kuan, M. Raabe, *ChemMedChem* **2021**, *16*, 94–104.

[11] P. Akkapeddi, S. A. Azizi, A. M. Freedy, P. Cal, P. M. P. Gois, G. J. L. Bernardes, *Chem. Sci.* **2016**, *7*, 2954–2963.

[12] V. Chudasama, A. Maruani, S. Caddick, *Nat. Chem.* **2016**, *8*, 114–119.

[13] C. Chen, D. Y. W. Ng, T. Weil, *Prog. Polym. Sci.* **2020**, *105*, 101241.

[14] I. Cobo, M. Li, B. S. Sumerlin, S. Perrier, *Nat. Mater.* **2015**, *14*, 143–159.

[15] G. Zhang, S. Zheng, H. Liu, P. R. Chen, *Chem. Soc. Rev.* **2015**, *44*, 3405–3417.

[16] Q. Y. Hu, F. Berti, R. Adamo, *Chem. Soc. Rev.* **2016**, *45*, 1691–1719.

[17] L. S. Witus, M. B. Francis, *Acc. Chem. Res.* **2011**, *44*, 774–783.

[18] R. X. Zhang, H. L. Wong, H. Y. Xue, J. Y. Eoh, X. Y. Wu, *J. Controlled Release* **2016**, *240*, 489–503.

[19] S. García-Alonso, A. Ocaña, A. Pandiella, *Cancer Res.* **2018**, *78*, 2159–2165.

[20] N. Crawley, M. Thompson, A. Romaschin, *Anal. Chem.* **2014**, *86*, 130–160.

[21] S. L. Kuan, S. Fischer, S. Hafner, T. Wang, T. Syrovets, W. Liu, Y. Tokura, D. Y. W. Ng, A. Riegger, C. Förtsch, D. Jäger, T. F. E. Barth, T. Simmet, H. Barth, T. Weil, *Adv. Sci.* **2018**, *5*, 1701036.

[22] S. Chakrabortty, B. K. Agrawalla, A. Stumper, N. M. Vegi, S. Fischer, C. Reichardt, M. Kogler, B. Dietzek, M. Feuring-Buske, C. Buske, S. Rau, T. Weil, *J. Am. Chem. Soc.* **2017**, *139*, 2512–2519.

[23] S. Hafner, M. Raabe, Y. Wu, T. Wang, Z. Zuo, V. Rasche, T. Syrovets, T. Weil, T. Simmet, *Adv. Ther.* **2019**, *2*, 1900084.

[24] A. Maruani, D. A. Richards, V. Chudasama, *Org. Biomol. Chem.* **2016**, *14*, 6165–6178.

[25] K. Wang, A. Sachdeva, D. J. Cox, N. M. Wilf, K. Lang, S. Wallace, R. A. Mehl, J. W. Chin, *Nat. Chem.* **2014**, *6*, 393–403.

[26] P. Adumeau, D. Vivier, S. K. Sharma, J. Wang, T. Zhang, A. Chen, B. J. Agnew, B. M. Zeglis, *Mol. Pharm.* **2018**, *15*, 892–898.

[27] M. R. Levengood, X. Zhang, J. H. Hunter, K. K. Emmerton, J. B. Miyamoto, T. S. Lewis, P. D. Senter, *Angew. Chem. Int. Ed.* **2017**, *56*, 733–737; *Angew. Chem.* **2017**, *129*, 751–755.

[28] N. Stephanopoulos, M. B. Francis, *Nat. Chem. Biol.* **2011**, *7*, 876–884.

[29] J. N. deGruyter, L. R. Malins, P. S. Baran, *Biochemistry* **2017**, *56*, 3863–3873.

[30] C. B. Rosen, M. B. Francis, *Nat. Chem. Biol.* **2017**, *13*, 697–705.

[31] D. G. Rawale, K. Thakur, S. R. Adusumalli, V. Rai, *Eur. J. Org. Chem.* **2019**, 6749–6763.

[32] L. Xu, M. Raabe, M. M. Zegota, J. C. F. Nogueira, V. Chudasama, S. L. Kuan, T. Weil, *Org. Biomol. Chem.* **2020**, *18*, 1140–1147.

[33] P. Ochtrop, C. P. R. Hackenberger, *Curr. Opin. Chem. Biol.* **2020**, *58*, 28–36.

[34] J. M. Chalker, G. J. Bernardes, Y. A. Lin, B. G. Davis, *Chem. Asian J.* **2009**, *4*, 630–640.

[35] S. B. Gunnoo, A. Madder, *ChemBioChem* **2016**, *17*, 529–553.

[36] J. Ravasco, H. Faustino, A. Trindade, P. M. P. Gois, *Chem. Eur. J.* **2019**, *25*, 43–59.

[37] E. Gil de Montes, E. Jiménez-Moreno, B. Oliveira, C. D. Navo, P. M. S. D. Cal, G. Jiménez-Osés, I. Robina, A. J. Moreno-Vargas, G. J. L. Bernardes, *Chem. Sci.* **2019**, *10*, 4515–4522.

[38] T. Wang, A. Riegger, M. Lamla, S. Wiese, P. Oeckl, M. Otto, Y. Wu, S. Fischer, H. Barth, S. L. Kuan, T. Weil, *Chem. Sci.* **2016**, *7*, 3234–3239.

[39] E. V. Vinogradova, C. Zhang, A. M. Spokoyny, B. L. Pentelute, S. L. Buchwald, *Nature* **2015**, *526*, 687–691.

[40] J. M. Antos, M. B. Francis, *Curr. Opin. Chem. Biol.* **2006**, *10*, 253–262.

[41] C. Zhang, E. V. Vinogradova, A. M. Spokoyny, S. L. Buchwald, B. L. Pentelute, *Angew. Chem. Int. Ed.* **2019**, *58*, 4810–4839; *Angew. Chem.* **2019**, *131*, 4860–4892.

[42] C. E. Hoyle, C. N. Bowman, *Angew. Chem. Int. Ed.* **2010**, *49*, 1540–1573; *Angew. Chem.* **2010**, *122*, 1584–1617.

[43] C. E. Hoyle, A. B. Lowe, C. N. Bowman, *Chem. Soc. Rev.* **2010**, *39*, 1355–1387.

[44] C. Canovas, M. Moreau, C. Bernhard, A. Oudot, M. Guillemin, F. Denat, V. Goncalves, *Angew. Chem. Int. Ed.* **2018**, *57*, 10646–10650; *Angew. Chem.* **2018**, *130*, 10806–10810.

[45] A. L. Baumann, S. Schwagerus, K. Broi, K. Kemnitz-Hassanin, C. E. Stieger, N. Trieloff, P. Schmieder, C. P. R. Hackenberger, *J. Am. Chem. Soc.* **2020**, *142*, 9544–9552.

[46] R. Tessier, J. Ceballos, N. Guidotti, R. Simonet-Davin, B. Fierz, J. Waser, *Chem* **2019**, *5*, 2243–2263.

[47] M. E. B. Smith, F. F. Schumacher, C. P. Ryan, L. M. Tedaldi, D. Papaioannou, G. Waksman, S. Caddick, J. R. Baker, *J. Am. Chem. Soc.* **2010**, *132*, 1960–1965.

[48] N. Martinez-Sáez, S. Sun, D. Oldrini, P. Sormanni, O. Boutureira, F. Carboni, I. Compañón, M. J. Deery, M. Vendruscolo, F. Corzana, R. Adamo, G. J. L. Bernardes, *Angew. Chem. Int. Ed.* **2017**, *56*, 14963–14967; *Angew. Chem.* **2017**, *129*, 15159–15163.

[49] S. J. Walsh, S. Omarjee, W. Galloway, T. T. Kwan, H. F. Sore, J. S. Parker, M. Hyvonen, J. S. Carroll, D. R. Spring, *Chem. Sci.* **2019**, *10*, 694–700.

[50] A. Maruani, M. E. Smith, E. Miranda, K. A. Chester, V. Chudasama, S. Caddick, *Nat. Commun.* **2015**, *6*, 6645.

[51] K. Li, W. Wang, J. Gao, *Angew. Chem. Int. Ed.* **2020**, *59*, 14246–14250; *Angew. Chem.* **2020**, *132*, 14352–14356.

[52] H. Faustino, M. Silva, L. F. Veiros, G. J. L. Bernardes, P. M. P. Gois, *Chem. Sci.* **2016**, *7*, 5052–5058.

[53] J. I. MacDonald, H. K. Munch, T. Moore, M. B. Francis, *Nat. Chem. Biol.* **2015**, *11*, 326–331.

[54] J.-R. Deng, N. C.-H. Lai, K. K.-Y. Kung, B. Yang, S.-F. Chung, A. S.-L. Leung, M.-C. Choi, Y.-C. Leung, M.-K. Wong, *Commun. Chem.* **2020**, *3*, 67.

[55] A. C. Obermeyer, J. B. Jarman, M. B. Francis, *J. Am. Chem. Soc.* **2014**, *136*, 9572–9579.

[56] G. Liang, H. Ren, J. Rao, *Nat. Chem.* **2010**, *2*, 54–60.

[57] S. Pomplun, M. Y. H. Mohamed, T. Oelschlaegel, C. Wellner, F. Bergmann, *Angew. Chem. Int. Ed.* **2019**, *58*, 3542–3547; *Angew. Chem.* **2019**, *131*, 3580–3585.

[58] A. O. Chan, C. M. Ho, H. C. Chong, Y. C. Leung, J. S. Huang, M. K. Wong, C. M. Che, *J. Am. Chem. Soc.* **2012**, *134*, 2589–2598.

[59] M. T. Taylor, J. E. Nelson, M. G. Suero, M. J. Gaunt, *Nature* **2018**, *562*, 563–568.

[60] A. H. Christian, S. Jia, W. Cao, P. Zhang, A. T. Meza, M. S. Sigman, C. J. Chang, F. D. Toste, *J. Am. Chem. Soc.* **2019**, *141*, 12657–12662.

[61] J. R. Kramer, T. J. Deming, *Biomacromolecules* **2012**, *13*, 1719–1723.

[62] M. W. Jones, G. Mantovani, C. A. Blindauer, S. M. Ryan, X. Wang, D. J. Brayden, D. M. Haddleton, *J. Am. Chem. Soc.* **2012**, *134*, 7406–7413.

[63] D. W. Romanini, M. B. Francis, *Bioconjugate Chem.* **2008**, *19*, 153–157.

[64] S. D. Tilley, M. B. Francis, *J. Am. Chem. Soc.* **2006**, *128*, 1080–1081.

[65] H. Ban, M. Nagano, J. Gavrilyuk, W. Hakamata, T. Inokuma, C. F. Barbas 3rd, *Bioconjugate Chem.* **2013**, *24*, 520–532.

[66] J. M. Antos, A. T. Iavarone, M. B. Francis, *J. Am. Chem. Soc.* **2009**, *131*, 6301–6308.

[67] Y. Seki, T. Ishiyama, D. Sasaki, J. Abe, Y. Sohma, K. Oisaki, M. Kanai, *J. Am. Chem. Soc.* **2016**, *138*, 10798–10801.

[68] A. J. Reay, T. J. Williams, I. J. Fairlamb, *Org. Biomol. Chem.* **2015**, *13*, 8298–8309.

[69] H. R. Alexandra Schischko, N. Kaplaneris, L. Ackermann, *Angew. Chem. Int. Ed.* **2017**, *56*, 1576–1580; *Angew. Chem.* **2017**, *129*, 1598–1602.

[70] Z. Ruan, N. Sauermann, E. Manoni, L. Ackermann, *Angew. Chem. Int. Ed.* **2017**, *56*, 3172–3176; *Angew. Chem.* **2017**, *129*, 3220–3224.

[71] M. B. Hansen, F. Hubalek, T. Skrydstrup, T. Hoeg-Jensen, *Chem. Eur. J.* **2016**, *22*, 1572–1576.

[72] X. Li, H. Y. Lam, Y. Zhang, C. K. Chan, *Org. Lett.* **2010**, *12*, 1724–1727.

[73] J. C. Vantourout, S. R. Adusumalli, K. W. Knouse, D. T. Flood, A. Ramirez, N. M. Padial, A. Istrate, K. Maziarz, J. N. deGruyter, R. R. Merchant, J. X. Qiao, M. A. Schmidt, M. J. Deery, M. D. Eastgate, P. E. Dawson, G. J. L. Bernardes, P. S. Baran, *J. Am. Chem. Soc.* **2020**, *142*, 17236–17242.

[74] Y. Gong, D. Andina, S. Nahar, J. C. Leroux, M. A. Gauthier, *Chem. Sci.* **2017**, *8*, 4082–4086.

[75] C. R. Bertozzi, *Acc. Chem. Res.* **2011**, *44*, 651–653.

[76] E. M. Sletten, C. R. Bertozzi, *Angew. Chem. Int. Ed.* **2009**, *48*, 6974–6998; *Angew. Chem.* **2009**, *121*, 7108–7133.

[77] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599; *Angew. Chem.* **2002**, *114*, 2708–2711.

[78] J. C. Jewett, C. R. Bertozzi, *Chem. Soc. Rev.* **2010**, *39*, 1272–1279.

[79] H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, *Nature* **2010**, *464*, 441–444.

[80] J. A. Walker, J. J. Bohn, F. Ledesma, M. R. Sorkin, S. R. Kabaria, D. N. Thornlow, C. A. Alabi, *Bioconjugate Chem.* **2019**, *30*, 2452–2457.

[81] B. Wu, Z. Wang, Y. Huang, W. R. Liu, *ChemBioChem* **2012**, *13*, 1405–1408.

[82] A. Sachdeva, K. Wang, T. Elliott, J. W. Chin, *J. Am. Chem. Soc.* **2014**, *136*, 7785–7788.

[83] R. K. Lim, Q. Lin, *Chem. Commun.* **2010**, *46*, 1589–1600.

[84] K. Lang, J. W. Chin, *Chem. Rev.* **2014**, *114*, 4764–4806.

[85] C. P. Ramil, Q. Lin, *Curr. Opin. Chem. Biol.* **2014**, *21*, 89–95.

[86] Z. Li, L. Qian, L. Li, J. C. Bernhammer, H. V. Huynh, J. S. Lee, S. Q. Yao, *Angew. Chem. Int. Ed.* **2016**, *55*, 2002–2006; *Angew. Chem.* **2016**, *128*, 2042–2046.

[87] B. L. Oliveira, Z. Guo, G. J. L. Bernardes, *Chem. Soc. Rev.* **2017**, *46*, 4895–4950.

[88] A. P. Crochet, M. M. Kabir, M. B. Francis, C. D. Paavola, *Biosens. Bioelectron.* **2010**, *26*, 55–61.

[89] B. K. Agrawalla, T. Wang, A. Riegger, M. P. Domogalla, K. Steinbrink, T. Dorfler, X. Chen, F. Boldt, M. Lamla, J. Michaelis, S. L. Kuan, T. Weil, *Bioconjugate Chem.* **2018**, *29*, 29–34.

[90] V. Chudasama, M. E. Smith, F. F. Schumacher, D. Papaioannou, G. Waksman, J. R. Baker, S. Caddick, *Chem. Commun.* **2011**, *47*, 8781–8783.

[91] Y. Zhang, C. Zang, G. An, M. Shang, Z. Cui, G. Chen, Z. Xi, C. Zhou, *Nat. Commun.* **2020**, *11*, 1015.

[92] J. Morales-Sanfrutos, J. Lopez-Jaramillo, M. Ortega-Munoz, A. Megia-Fernandez, F. Perez-Balderas, F. Hernandez-Mateo, F. Santoyo-Gonzalez, *Org. Biomol. Chem.* **2010**, *8*, 667–675.

[93] J. Morales-Sanfrutos, F. J. Lopez-Jaramillo, F. Hernandez-Mateo, F. Santoyo-Gonzalez, *J. Org. Chem.* **2010**, *75*, 4039–4047.

[94] P. Moody, V. Chudasama, R. I. Nathani, A. Maruani, S. Martin, M. E. Smith, S. Caddick, *Chem. Commun.* **2014**, *50*, 4898–4900.

[95] R. I. Nathani, P. Moody, V. Chudasama, M. E. Smith, R. J. Fitzmaurice, S. Caddick, *Chem. Sci.* **2013**, *4*, 3455–3458.

[96] C. Zhang, M. Welborn, T. Zhu, N. J. Yang, M. S. Santos, T. Van Voorhis, B. L. Pentelute, *Nat. Chem.* **2016**, *8*, 120–128.

[97] J. W. Chin, *Nature* **2017**, *550*, 53–60.

[98] S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony, B. G. Davis, *Nature* **2007**, *446*, 1105–1109.

[99] A. R. Nodling, L. A. Spear, T. L. Williams, L. Y. P. Luk, Y. H. Tsai, *Essays Biochem.* **2019**, *63*, 237–266.

[100] W. Wan, Y. Huang, Z. Wang, W. K. Russell, P. J. Pai, D. H. Russell, W. R. Liu, *Angew. Chem. Int. Ed.* **2010**, *49*, 3211–3214; *Angew. Chem.* **2010**, *122*, 3279–3282.

[101] E. A. L. Eric, M. Brustad, P. G. Schultz, A. A. Deniz, *J. Am. Chem. Soc.* **2008**, *130*, 17664–17665.

[102] H. Xiao, A. Chatterjee, S. H. Choi, K. M. Bajjuri, S. C. Sinha, P. G. Schultz, *Angew. Chem. Int. Ed.* **2013**, *52*, 14080–14083; *Angew. Chem.* **2013**, *125*, 14330–14333.

[103] K. Wang, W. H. Schmied, J. W. Chin, *Angew. Chem. Int. Ed.* **2012**, *51*, 2288–2297; *Angew. Chem.* **2012**, *124*, 2334–2344.

[104] E. M. Milczek, *Chem. Rev.* **2018**, *118*, 119–141.

[105] Y. Yang, S. Lin, W. Lin, P. R. Chen, *ChemBioChem* **2014**, *15*, 1738–1743.

[106] M. D. Lee, W. Y. Tong, T. Nebl, L. A. Pearce, T. M. Pham, A. Golbaz-Hagh, S. Puttick, S. Rose, T. E. Adams, C. C. Williams, *Bioconjugate Chem.* **2019**, *30*, 2539–2543.

[107] M. Rashidian, S. C. Kumarapperuma, K. Gabrielse, A. Fegan, C. R. Wagner, M. D. Distefano, *J. Am. Chem. Soc.* **2013**, *135*, 16388–16396.

[108] I. Chen, M. Howarth, W. Lin, A. Y. Ting, *Nat. Methods* **2005**, *2*, 99–104.

[109] L. Schmohl, D. Schwarzer, *Curr. Opin. Chem. Biol.* **2014**, *22*, 122–128.

[110] G. K. T. Nguyen, Y. Cao, W. Wang, C. F. Liu, J. P. Tam, *Angew. Chem. Int. Ed.* **2015**, *54*, 15920–15924; *Angew. Chem.* **2015**, *127*, 15920–15924.

[111] Y. Zhang, M. J. Blanden, C. Sudheer, S. A. Gangopadhyay, M. Rashidian, J. L. Hougland, M. D. Distefano, *Bioconjugate Chem.* **2015**, *26*, 2542–2553.

[112] L. Deweid, L. Neureiter, S. Englert, H. Schneider, J. Deweid, D. Yanakieva, J. Sturm, S. Bitsch, A. Christmann, O. Avrutina, H. L. Fuchsbauer, H. Kolmar, *Chem. Eur. J.* **2018**, *24*, 15195–15200.

[113] J. G. Plaks, R. Falatach, M. Kastantin, J. A. Berberich, J. L. Kaar, *Bioconjugate Chem.* **2015**, *26*, 1104–1112.

[114] P. Wu, W. Shui, B. L. Carlson, N. Hu, D. Rabuka, J. Lee, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 3000–3005.

[115] 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, B. H. Geierstanger, *Bioconjugate Chem.* **2015**, *26*, 2554–2562.

[116] D. Schumacher, J. Helma, F. A. Mann, G. Pichler, F. Natale, E. Krause, M. C. Cardoso, C. P. Hackenberger, H. Leonhardt, *Angew. Chem. Int. Ed.* **2015**, *54*, 13787–13791; *Angew. Chem.* **2015**, *127*, 13992–13996.

[117] S. C. Reddington, M. Howarth, *Curr. Opin. Chem. Biol.* **2015**, *29*, 94–99.

[118] T. J. Harmand, D. Bousbaine, A. Chan, X. Zhang, D. R. Liu, J. P. Tam, H. L. Ploegh, *Bioconjugate Chem.* **2018**, *29*, 3245–3249.

[119] J. S. Italia, P. S. Addy, S. B. Erickson, J. C. Peeler, E. Weerapana, A. Chatterjee, *J. Am. Chem. Soc.* **2019**, *141*, 6204–6212.

[120] P. S. Addy, S. B. Erickson, J. S. Italia, A. Chatterjee, *J. Am. Chem. Soc.* **2017**, *139*, 11670–11673.

[121] D. L. Dunkelmann, J. C. W. Willis, A. T. Beattie, J. W. Chin, *Nat. Chem.* **2020**, *12*, 535–544.

[122] A. Maruani, P. A. Szijj, C. Bahou, J. C. F. Nogueira, S. Caddick, J. R. Baker, V. Chudasama, *Bioconjugate Chem.* **2020**, *31*, 520–529.

[123] B. Schuler, W. A. Eaton, *Curr. Opin. Struct. Biol.* **2008**, *18*, 16–26.

[124] K. L. Pierce, R. T. Premont, R. J. Lefkowitz, *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 639–650.

[125] J.-P. Vilardaga, M. Bunemann, C. Krasel, M. Castro, M. J. Lohse, *Nat. Biotechnol.* **2003**, *21*, 807–812.

[126] B. A. Griffin, R. Y. Tsien, *Science* **1998**, *281*, 269–271.

[127] C. Hoffmann, G. Gaietta, M. Bunemann, S. R. Adams, S. Oberdorff-Maass, B. Behr, J. P. Vilardaga, R. Y. Tsien, M. H. Ellisman, M. J. Lohse, *Nat. Methods* **2005**, *2*, 171–176.

[128] L. D. Lavis, R. T. Raines, *ACS Chem. Biol.* **2014**, *9*, 855–866.

[129] J. L. Fallon, F. A. Quiocho, *Structure* **2003**, *11*, 1303–1307.

[130] K. P. Hoeflich, M. Ikura, *Cell* **2002**, *108*, 739–742.

[131] G. Pasut, F. M. Veronese, *Adv. Drug Delivery Rev.* **2009**, *61*, 1177–1188.

[132] R. P. Lyon, T. D. Bovee, S. O. Doronina, P. J. Burke, J. H. Hunter, H. D. Neff-LaFord, M. Jonas, M. E. Anderson, J. R. Setter, P. D. Senter, *Nat. Biotechnol.* **2015**, *33*, 733–735.

[133] F. Li, K. K. Emmerton, M. Jonas, X. Zhang, J. B. Miyamoto, J. R. Setter, N. D. Nicholas, N. M. Okeley, R. P. Lyon, D. R. Benjamin, C. L. Law, *Cancer Res.* **2016**, *76*, 2710–2719.

[134] R. Chen, J. Hou, E. Newman, Y. Kim, C. Donohue, X. Liu, S. H. Thomas, S. J. Forman, S. E. Kane, *Mol. Cancer Ther.* **2015**, *14*, 1376–1384.

[135] M. Wu, J. Shu, *Contrast Media Mol. Imaging* **2018**, *2018*, 1382183.

[136] S. Hernandez Vargas, S. C. Ghosh, A. Azhdarinia, *J. Nucl. Med.* **2019**, *60*, 459–465.

[137] B. M. Zeglis, C. B. Davis, D. Abdel-Atti, S. D. Carlin, A. Chen, R. Aggeler, B. J. Agnew, J. S. Lewis, *Bioconjugate Chem.* **2014**, *25*, 2123–2128.

[138] K. A. Zettlitz, C. M. Waldmann, W. K. Tsai, R. Tavare, J. Collins, J. M. Murphy, A. M. Wu, *J. Nucl. Med.* **2019**, *60*, 1467–1473.

79

## 5.2 Site-selective protein modification via disulfide rebridging for fast tetrazine/trans-cyclooctene bioconjugation

Lujuan Xu, Marco Raabe, Maksymilian M. Zegota, João C. F. Nogueira, Vijay Chudasama, Seah Ling Kuan*, and Tanja Weil*

**Contribution of the respective authors:**

**Lujuan Xu:** Design of the experiments, the molecules and bioconjugates. Preparation of all the organic molecules and protein bioconjugates with full characterization. Writing and correcting the manuscript.

**Marco Raabe:** Preparation of some protein conjugates. Giving scientific advice

**Maksymilian Marek Zegota:** Interpreting mass spectrometry data

**João C. F. Nogueira**: Transferring the technology of Fab fragment generation.

**Vijay Chudasama:** Giving scientific advice. Correcting the manuscript.

**Seah Ling Kuan:** Design of the project, discussing results, and supervising the project. Writing and correcting the manuscript.

**Tanja Weil:** Design and discussion the scientific concept and the obtained results. Acquiring funding for the project. Writing and correcting the manuscript.

ROYAL SOCIETY OF CHEMISTRY

Check for updates

# Site-selective protein modification *via* disulfide rebridging for fast tetrazine/*trans*-cyclooctene bioconjugation†

Lujuan Xu,[a,b] Marco Raabe, [ID][a,b] Maksymilian M. Zegota,[a,b] João C. F. Nogueira,[c] Vijay Chudasama, [ID][c] Seah Ling Kuan [ID] *[a,b] and Tanja Weil [ID] *[a,b]

An inverse electron demand Diels–Alder reaction between tetrazine and *trans*-cyclooctene (TCO) holds great promise for protein modification and manipulation. Herein, we report the design and synthesis of a tetrazine-based disulfide rebridging reagent, which allows the site-selective installation of a tetrazine group into disulfide-containing peptides and proteins such as the hormone somatostatin (SST) and the antigen binding fragment (Fab) of human immunoglobulin G (IgG). The fast and efficient conjugation of the tetrazine modified proteins with three different TCO-containing substrates to form a set of bioconjugates in a site-selective manner was successfully demonstrated for the first time. Homogeneous, well-defined bioconjugates were obtained underlining the great potential of our method for fast bioconjugation in emerging protein therapeutics. The formed bioconjugates were stable against glutathione and in serum, and they maintained their secondary structure. With this work, we broaden the scope of tetrazine chemistry for site-selective protein modification to prepare well-defined SST and Fab conjugates with preserved structures and good stability under biologically relevant conditions.

## Introduction

Peptides and proteins are emerging as powerful treatment options, as exemplified by the application of antibodies and antibody fragments (Fab) in immunotherapy and antibody–drug conjugates in targeted cancer therapy.[1] Nevertheless, protein stability, immunogenicity and the time required to engineer recombinant antibodies could limit their development for *in vivo* studies.[2] In this regard, nature has evolved an optimal synthetic factory in the form of posttranslational processes, in which diverse functionalities can be attached to proteins in a site-directed fashion.[3] Inspired by this, chemists strive to develop methodologies to impart diverse functionalities to native proteins with similar levels of precision.

Site-selective modification of therapeutically relevant peptides and proteins to introduce reactive bioorthogonal handles has emerged for post-modification in an "on-demand" fashion to expand the features and functions of proteins to address the challenges in fundamental biological and medical applications.[4] In this manner, protein therapeutics can be synthetically customized to program their properties for envisaged applications.

Over the past decades, several bioorthogonal reactions, including the [3 + 2] azide–alkyne cycloaddition, photoclick 1,3-dipolar cycloaddition and the inverse electron demand Diels–Alder (IEDDA) reactions, have been developed and applied for protein modifications.[5] Among these, the IEDDA reaction of 1,2,4,5-tetrazine with *trans*-cyclooctene (TCO) stands out, providing fast reaction kinetics (rate constant of up to $10^6$ $M^{-1}$ $s^{-1}$), excellent orthogonality, catalyst-free conditions and good biocompatibility, and is a widely employed bioorthogonal approach in chemical biology.[6] Previously, the IEDDA reaction has been applied for the preparation of antibody conjugates, mainly *via* statistical modifications using *N*-hydroxysuccinimide (NHS) ester chemistry for cell imaging,[7] nanoparticle functionalization,[8] and antibody targeted therapy.[6a,9] In the literature, two common approaches to site-selectively modify proteins utilizing the IEDDA reaction are, the introduction of a diene or dienophile by genetic code expansion methods *via* the incorporation of noncanonical amino acids[10] and tag-based posttranslational attachment strategies using enzymes (*e.g.* lipoic acid protein ligase A).[11] Nevertheless, the genetic code expansion method suffers from low yields and tedious synthesis. This in turn limits its scal-

*[a]Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany. E-mail: weil@mpip-mainz.mpg.de, kuan@mpip-mainz.mpg.de*
*[b]Institute of Inorganic Chemistry I, University of Ulm, Albert-Einstein-Allee 11, 89081 Ulm, Germany*
*[c]Department of Chemistry, University College London, London, UK*

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c9ob02687h

1140 | *Org. Biomol. Chem.*, 2020, **18**, 1140–1147

81

This journal is © The Royal Society of Chemistry 2020

ability and accessibility for further applications, while tag-based posttranslational attachment strategies first require the genetic fusion of a specific peptide sequence (*e.g.* lipoate acceptor peptide) to the protein of interest.[11] A single tetrazine or TCO group has also been introduced site-selectively into peptides or proteins at an unpaired cysteine site *via* the thiol-maleimide reaction.[6a,12]

However, very few native proteins contain free cysteine residues.[13] Furthermore, thiol–maleimide chemistry easily undergoes the retro-Michael reaction resulting in a maleimide exchange with other reactive thiols present, for example glutathione, under physiological conditions causing off-target effects.[4a] Therefore, other methods, which can introduce a tetrazine or TCO group in a site-selective fashion into proteins for subsequent IEDDA reactions are highly desirable to expand the scope of their applications. Since many therapeutically relevant peptides and proteins contain at least one disulfide bond close to the protein surface, the disulfide rebridging strategy provides a versatile technique to modify the solvent accessible disulfide bonds on these proteins and peptides to install suitable bioorthogonal tags.[13,14] Smith and coworkers presented an elegant approach to introduce the tetrazine group at the disulfide site of proteins/peptides *via* dichlorotetrazine.[15] However, a significantly reduced activity of the protein was reported after the modification. In this context, disulfide modification based on bis-alkylating reagents to form bisthioether conjugates gained wide applications for site-selective protein modifications without compromising their bioactivity. Furthermore, it has been reported that bisthioether conjugates formed by disulfide rebridging are more stable than thiol-maleimide conjugates[16] suggesting that it could be a viable strategy for the installation of IEEDA reaction handles. Herein, we present a simple and straightforward method for the site-selective incorporation of a tetrazine group into native peptides and proteins through disulfide rebridging with the allyl sulfone scaffold reported previously.[17] In comparison with known bis-sulfone disulfide rebridging reagents, allyl sulfone reagents provide improved reactivity and higher water solubility, which greatly facilitates protein bioconjugations.[17] The disulfide rebridging reactions of allyl sulfones proceed *in situ* without side reactions *e.g.* with reducing agents such as tris(2-carboxyethyl)-phosphine hydrochloride (TCEP). In comparison, disubstituted maleimide disulfide rebridging reagents, such as 3,4-dibromomaleimide, show side reactions with TCEP, presumably due to the nucleophilic addition of

phosphine to the electron-poor alkene moiety.[18] The versatility of the method was demonstrated by modifying two model substrates: SST and IgG Fab under mild conditions which yielded a set of well-defined protein conjugates (Fig. 1). The formed protein conjugates showed good stability and maintained their secondary structures after bioconjugation. The reported methodology provides a rapid, robust and straightforward strategy for site-selective protein labeling and expands the scope of using IEDDA chemistry in the functionalization of therapeutically relevant peptides and proteins in a precise manner.

## Results and discussion

### Synthesis and characterization of the tetrazine-based disulfide rebridging reagent

For the site-selective installation of bioorthogonal handles on the proteins, a longer incubation time (usually overnight) is often required.[13] Therefore, the diene "6-methyl-1,2,4,5-tetrazine" instead of the dienophile "TCO" was introduced into the protein. This strategy fully utilizes the fast kinetics of the IEDDA reaction but minimizes the inherent isomerization issues of TCO to nonreactive *cis*-cyclooctene, which occurs at a longer incubation time or in the presence of thiols.[4c,19] In this context, a tetrazine-based allylsulfone disulfide rebridging reagent (IC-Tetrazine, Scheme 1) was designed and readily
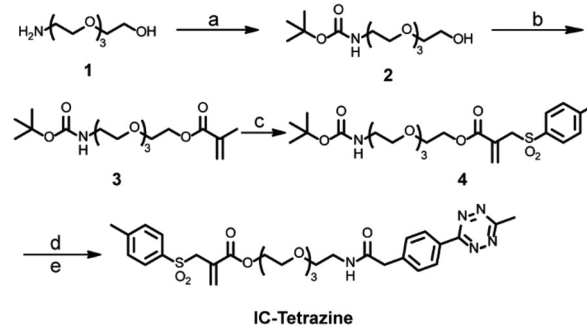
**Scheme 1** Synthesis of the tetrazine-containing disulfide rebridging reagent (IC-Tetrazine). (a) Di-*tert*-butyl dicarbonate, CH$_2$Cl$_2$, overnight. (b) Methacryloyl chloride, Et$_3$N, CH$_2$Cl$_2$, 90%. (c) 1. I$_2$, sodium *p*-toluenesulfinate, CH$_2$Cl$_2$, 3 days. 2. Et$_3$N, CH$_2$Cl$_2$, overnight. 3. Et$_3$N, ethyl acetate, 95 °C, overnight, final yield: 60%. (d) Trifluoroacetic acid (TFA), CH$_2$Cl$_2$, 98%. (e) Methyltetrazine NHS ester, Et$_3$N, CH$_2$Cl$_2$, overnight, 40%.

**Fig. 1** Schematic overview of site-selective incorporation of a reactive tetrazine tag into a solvent-accessible disulfide site of a cell-targeting peptide or antibody fragment. The tetrazine-modified peptide/protein is used for post-functionalization to construct a small set of protein bioconjugates with a (1) dye, (2) polymer and (3) protein in a fast and highly selective manner with high conversions (one isomer is shown as a representation).

obtained through a four-step synthesis. As shown in Scheme 1, tetraethylene glycol monoamine (**1**) was protected to afford 2-(2-boc-aminoethoxy)ethanol (**2**) that underwent condensation with methacryloyl chloride yielding the corresponding methacrylate derivative (**3**). After a tandem iodosulfonylation–dehydroiodination reaction, the tosyl group was introduced to form allyl-sulfone (**4**). The *tert*-butoxycarbonyl group was subsequently removed in an acidic medium to afford the primary amino group, which then reacted with methyl-tetrazine NHS ester to afford the IC-Tetrazine reagent.

## Site-selective protein modification

Subsequently, two model substrates (somatostatin and human IgG Fab) were selected for site-selective modification with IC-Tetrazine. Somatostatin (SST) regulates the endocrine system and mediates signal transduction *via* five G protein-coupled receptors (SSTR) that are overexpressed in high levels in various kinds of cancer cells and tumor blood vessels.[20] SST conjugates have been widely investigated for targeted drug delivery into SSTR-expressing cancer cells.[20,21] IC-Tetrazine was reacted with SST after its single accessible disulfide bond was reduced to generate free thiol groups by the addition of two equivalents of TCEP at room temperature (Fig. 2a). The product (SST-Tetrazine) was purified by high-performance liquid chromatography (HPLC) yielding SST-Tetrazine in 30% yield, which is comparable to that reported in the literature (20% to 40%).[13] Characterization by matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) revealed two signals corresponding to the desired SST-Tetrazine ($m/z$ = 2110.95 $[M + H]^+$, Fig. 2b, the molecular weight of native SST is 1637 g mol$^{-1}$) and the laser fragmentation species of SST-Tetrazine ($m/z$ = 2041.93, the

chemical structure of the fragmentation species is shown in Fig. S3†), respectively. Notably, SST-Tetrazine is stable for more than one year when stored as a solid at −20 °C (Fig. S4†).

Next, we applied the modification method to an antibody fragment (Fab) from human immunoglobulin G (IgG). Fab fragments have been successfully used for constructing antibody–drug and antibody–nanoparticle conjugates providing several advantages compared to the complete antibody.[22] The conjugate retains the antigen-binding region which is crucial for active targeting, but circumvents the non-specific binding of the Fc region of antibodies.[22a] Since the interchain disulfides are much more exposed to the solvent and less stable relative to the intrachain disulfides that are buried between the two layers of anti-parallel β-sheet structures,[23] IgG Fab was incubated with 50 equivalents of TCEP (optimization of the amounts of TCEP is shown in Fig. S5†) to reduce the solvent accessible interchain disulfide bonds and subsequently reacted with IC-Tetrazine overnight (Fig. 2a). The tetrazine modified Fab fragment (Fab-Tetrazine) was purified by ultrafiltration to remove residual TCEP and IC-Tetrazine. The MALDI-TOF-MS analysis of Fab-Tetrazine revealed a molecular weight of 48.5 kDa (Fig. 2c), showing an increase of ~500 Da compared to that of the native Fab (48.0 kDa, Fig. S5B†) demonstrating its successful site-selective modification. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of Fab-Tetrazine revealed a single band at ~48 kDa showing that only a slight reduction occurred after site-selective modification even in the presence of 100 equivalents of TCEP after one hour (Fig. 2d, band 2). In contrast, under the same conditions, the native IgG Fab was almost completely reduced exhibiting a single band appearing at ~24 kDa on SDS-PAGE (Fig. 2d, band 1). Based on the concentration calcu-



**Fig. 2** (a) Site-selective modification of SST and IgG Fab with IC-Tetrazine. (b) MALDI–TOF–MS spectrum of SST–Tetrazine (found: 2110.95 $[M + H]^+$, calculated: 2110.95 $[M + H]^+$). The $m/z$ peak at 2041.93 corresponds to the fragmentation of SST–Tetrazine. (c) MALDI–TOF–MS spectrum of Fab-Tetrazine (Fab-Tetrazine: found: 48.5k $[M + H]^+$, calculated: 48.5k $[M + H]^+$). (d) SDS-PAGE analysis of native Fab and Fab-Tetrazine (M: Applichem Protein Marker VI, 1: Fab (native) + 100 eq. TCEP incubation for one hour, and 2: Fab-Tetrazine + 100 eq. TCEP incubation for one hour).

**1142** | *Org. Biomol. Chem.*, 2020, **18**, 1140–1147

83

This journal is © The Royal Society of Chemistry 2020

lated from gel densitometry (Fig. 2d), about 80% of the Fab fragment was successfully modified. A thiol quantification experiment using the thiol reagent 4,4′-dithiodipyridine was performed and the result showed that about 73% of IgG Fab were successfully modified, which is consistent with the result obtained from gel densitometry (details are shown in the ESI, Fig. S5c†). Taken together, the results clearly indicated that the disulfide bonds in IgG Fab were successfully rebridged using IC-Tetrazine.

### Bioconjugation with a chromophore, PEG and enzyme

The versatility of the tetrazine-modified SST and IgG Fab for the construction of well-defined bioconjugates was demonstrated through post-modification with a fluorophore (cyanine-5, Cy5) and a PEG chain. In addition, SST-Tetrazine was also conjugated to a protein enzyme (cytochrome C, CytC) to form a peptide–protein conjugate with high conversion. A TCO-terminated PEG chain of a precise chain length with 12 repeating units (TCO-PEG12) was selected to facilitate the characterization by MALDI-TOF-MS. TCO-PEG12 was incubated with SST-Tetrazine in phosphate buffer (PB, 50 mM, pH = 7.4) and the pink color of the tetrazine group disappeared immediately after mixing, indicating that the reaction took place instantaneously. The bioconjugation reaction was carried out for 30 minutes to ensure completion. Thereafter, the reaction mixture was injected into the HPLC column and a predominant peak of SST-PEG12 was obtained (Fig. S7, ESI†). SST-PEG12 was isolated in nearly quantitative yield (95%) and characterized by MALDI-TOF-MS (Fig. 3b). SST-Tetrazine was also conjugated to TCO-Cy5 under similar conditions yielding

the desired conjugate SST-Cy5 in 90% yield. MALDI-TOF-MS showed the $m/z$ signal at 3041.17 (Fig. 3c).

Due to the much larger molecular weight of Fab-Tetrazine compared to SST-Tetrazine, a longer PEG chain with a molecular weight of 5000 Da containing 113 repeating units on average (TCO-PEG113) was incubated with Fab-Tetrazine in phosphate buffer (pH = 7.4) for 30 minutes to afford Fab-PEG113. The characteristic band of Fab-Tetrazine at ~48 kDa almost entirely disappeared after PEGylation and a new band emerged, which was broader due to the polydispersity of the PEG chain (Fig. 3d). A control experiment of mixing native Fab and TCO-PEG113 did not show any unspecific absorption on SDS-PAGE (Fig. 3d). MALDI-TOF-MS spectra revealed a signal at 53.5 kDa indicating the successful conjugation (Fig. S11D†). Under similar conditions, the widely applied chromophore Cy5 was also attached to Fab-Tetrazine to afford Fab-Cy5 (molecular weight: 49.5 kDa, Fig. 3e) in 73% yield (the calculation is shown in the ESI†). Successful conjugation was confirmed by an increase in the molecular weight of ~1.5 kDa in the MALDI-TOF-MS spectra in comparison with those of native Fab (molecular weight: 48 kDa, Fig. 3e).

CytC is a 12 kDa hemoprotein typically located in the mitochondria of viable cells.[24] Iso-yeast CytC has a single thiol group on its surface for modification *via* thiol-maleimide chemistry.[25] TCO-PEG3-maleimide was reacted with CytC to introduce the TCO group *via* thiol-maleimide chemistry affording CytC-PEG3-TCO (Scheme S4†). MALDI-TOF-MS data revealed the successful modification of CytC (Fig. S6, ESI†). In addition, a tetrazine-based dye (tetrazine-5-fluorescein, Tetrazine-FAM, excitation wavelength: 492 nm and emission
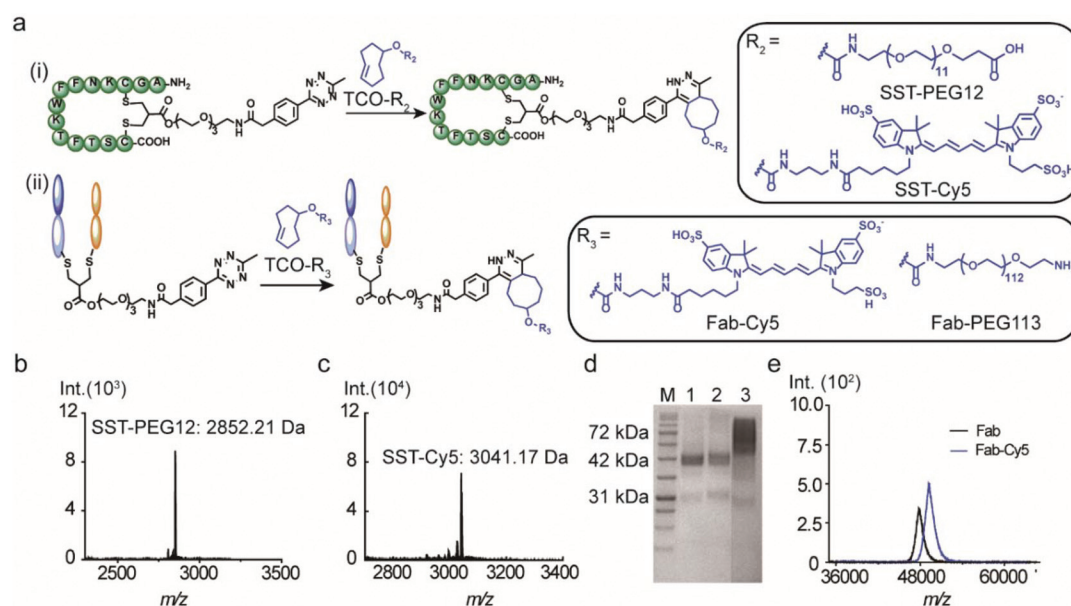


**Fig. 3** (a) Bioconjugation between (i) SST-Tetrazine with TCO-PEG12 and TCO-Cy5. (ii) Fab-Tetrazine with TCO-Cy5 and TCO-PEG113 *via* a fast click reaction. (b) MALDI-TOF-MS spectrum of SST-PEG12 (found: 2852.21 [M + H]⁺, calculated: 2852.28 [M + H]⁺). (c) MALDI-TOF-MS spectrum of SST-Cy5 (found: 3041.17 [M + H]⁺, calculated: 3041.28 [M + H]⁺). Sinapic acid was used as the matrix for all MALDI-TOF-MS measurements. (d) SDS-PAGE analysis of Fab-PEG113 (M: Applichem Protein Marker VI, 1: Fab-Tetrazine, and 2: Fab + TCO-PEG113, 3: Fab-PEG113). (e) MALDI-TOF-MS spectra of Fab-Cy5 (found: 49.5k [M + H]⁺, calculated: 49.5k [M + H]⁺) and native Fab (found: 48.0k [M + H]⁺, calculated: 48.0k [M + H]⁺).

This journal is © The Royal Society of Chemistry 2020

84

*Org. Biomol. Chem.*, 2020, **18**, 1140–1147 | **1143**

wavelength: 517 nm) was used to react with CytC-PEG3-TCO to assess the modification yield of CytC. Based on the absorbance of CytC at 280 nm and the FAM dye at 490 nm, about 90% of CytC was successfully modified with the TCO group (experimental details are shown in the ESI†). Thereafter, the targeting peptide SST-Tetrazine was conjugated to CytC-PEG3-TCO by mixing and shaking for 30 minutes.

After ultrafiltration to remove the residual SST-Tetrazine, SST-PEG3-CytC was obtained and characterized by MALDI-TOF-MS and SDS-PAGE. In the MALDI-TOF-MS spectra, the signal at 15 315 Da corresponds to the desired SST-PEG3-CytC conjugate (Fig. 4b), whereas the signal for CytC-PEG3-TCO (13 233 Da) completely disappeared. On SDS-PAGE, a slight band shift was observed indicating the successful conjugation between SST-Tetrazine and CytC-PEG3-TCO (Fig. 4c). The conjugation of Fab-Tetrazine with CytC-PEG3-TCO was

also performed. However, MALDI-Tof-MS data showed almost no conjugation product, presumably due to the steric hindrance of the two bulky macromolecules.

### Stability studies of SST-Tetrazine, Fab-Tetrazine and Fab-Cy5

The stability of the bioconjugates after disulfide modification is an important consideration for their intended applications in cellular environments. Therefore, the stability of the modified protein with the newly synthesized disulfide rebridging reagent in glutathione (GSH) was investigated using liquid chromatography mass spectrometry (LC-MS). SST-Tetrazine (10 µM) was incubated with two-fold excess of GSH at biologically relevant concentrations (20 µM)[26] at 37 °C and SST-Tetrazine and the degradation product (GSH-Tetrazine, 1085 $[M + H]^+$; the chemical structure is shown in Scheme S9†) were identified by simultaneous detection by UV-Vis spectroscopy at an absorption wavelength of 254 nm and selective ion monitoring (SIM). The amount of SST-Tetrazine in each sample was determined as a ratio of the integration of the chromatogram at 254 nm of SST-Tetrazine to the standard (Fmoc-L-phenylalanine) (Fig. 5a and b). LC-MS data showed that SST-Tetrazine remained intact after 24 hours of incubation and no degradation products were detected in the SIM profile (Fig. S12C†). Next, the stability of the Fab conjugates was investigated based on the literature protocol.[22b] Fab-Tetrazine and Fab-Cy5 were incubated with 1% fetal bovine serum (FBS) at 37 °C and monitored using SDS-PAGE. SDS-PAGE data revealed no change in Fab-Tetrazine after 24 hours, indicating no obvious degradation (Fig. 5c). The stability of Fab-Cy5 was also investigated based on the quantification of the fluorescence of the conjugate (shown in the ESI,† page 15). The data indicated a good stability of Fab-Cy5 after incubation with FBS for 24 hours (Fig. 5d). Taken together, these experiments showed that the resultant peptide and protein bioconjugates formed by the disulfide rebridging strategy and the subsequent IEDDA reaction remained stable under physiological conditions, e.g. in the presence of GSH or in media containing serum proteins.
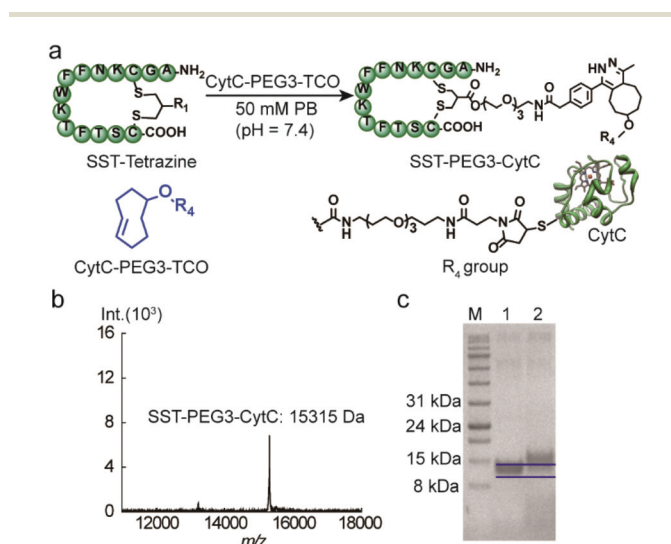


**Fig. 4** (a) Bioconjugation between SST-Tetrazine and CytC-PEG3-TCO. (b) MALDI-TOF-MS spectrum of SST-PEG3-CytC (found: 15 315 [M + H]⁺, calculated: 15 314 [M + H]⁺) matrix: sinapic acid. (c) SDS-PAGE analysis of SST-PEG3-CytC (M: Applichem Protein Marker VI, 1: CytC-PEG3-TCO, and 2: SST-PEG3-CytC).
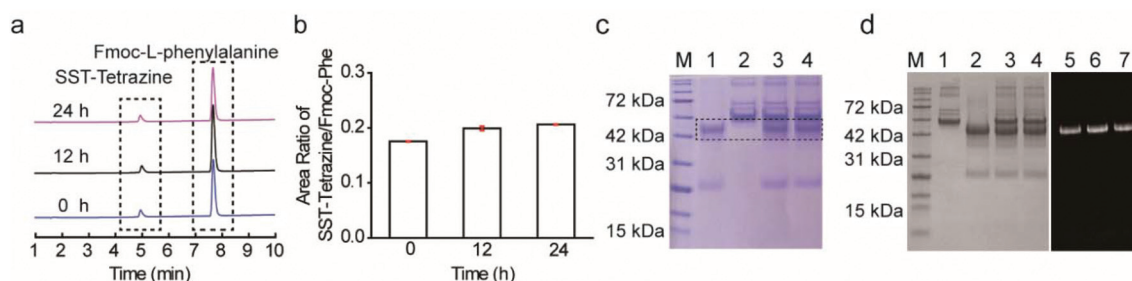


**Fig. 5** (a) Stability study of SST-Tetrazine via LC-MS. (b) Area ratio of SST-Tetrazine at 254 nm compared to Fmoc-L-phenylalanine (standard) at three time points (n = 3, values are given as mean ± SD). (c) Stability study of Fab-Tetrazine by SDS-PAGE (M: Applichem Protein Marker VI, 1: Fab-Tetrazine, 2: 1% FBS, 3: Fab-Tetrazine + 1% FBS incubated for 12 hours, and 4: Fab-Tetrazine + 1% FBS incubated for 24 hours), (d) Stability study of Fab-Cy5 by SDS-PAGE with (left) and without (right) Coomassie blue (M: Applichem Protein Marker VI, 1: 1% FBS, 2: Fab-Cy5, 3: Fab-Cy5 + 1% FBS incubated for 12 h, 4: Fab-Cy5 + 1% FBS incubated for 24 h, 5: fluorescence of Fab-Cy5, 6: fluorescence of Fab-Cy5 after incubation with 1% FBS for 12 h, and 7: fluorescence of Fab-Cy5 after incubation with 1% FBS for 24 h).

**1144** | *Org. Biomol. Chem.*, 2020, **18**, 1140–1147

85

This journal is © The Royal Society of Chemistry 2020
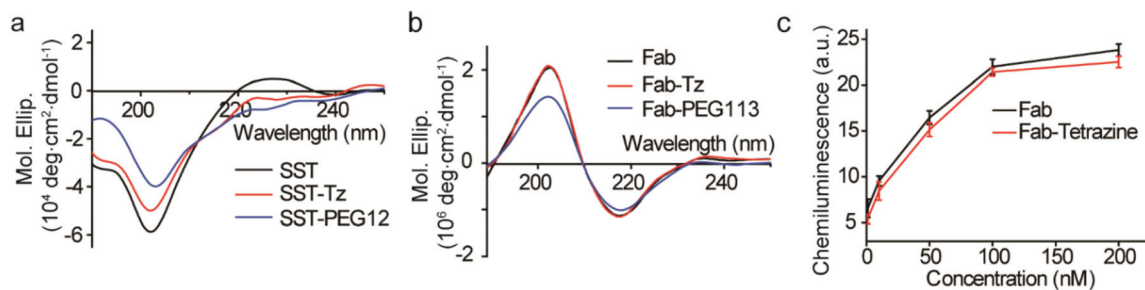
**Fig. 6** (a) CD spectra of SST, SST-Tetrazine and SST-PEG12 (SST, SST-Tetrazine and SST-PEG12 were diluted to 0.1 mg mL$^{-1}$ in PBS, the data are the average of three runs). (b) CD spectra of Fab, Fab-Tetrazine and Fab-PEG113 (Fab, Fab-Tetrazine and Fab-PEG113 were measured at 0.1 mg mL$^{-1}$ in PBS, the data are the average of three runs). (c) ELISA data of Fab and Fab-Tetrazine (five different concentration points: 1 nM, 10 nM, 50 nM, 100 nM, and 200 nM). Data are plotted as mean $\pm$ SD ($n$ = 3).

### Structural and functional studies

Circular dichroism (CD) is a versatile technique to determine the secondary and tertiary structures of proteins. According to Fig. 6a, SST, SST-Tetrazine and SST-PEG12 showed typical random coil structures with a negative band at 201 nm. The peak at 225 nm of SST is attributed to the presence of a higher degree of polyproline (PPII) conformation.[27]

Both native IgG Fab, Fab-Tetrazine and Fab-PEG113 conjugates, possess well-defined antiparallel β-pleated sheets confirmed by the peaks at 218 nm and 202 nm consistent with the literature report,[23] which confirmed that the secondary structure of Fab was preserved after modification (Fig. 6b). Next, we assessed whether the modification of IgG Fab affects its function to bind to protein L, a bacterial protein known to interact with Fab fragments of immunoglobulins.[28] The binding and recognition of native Fab and Fab-Tetrazine were thus investigated by enzyme-linked immunosorbent assay (ELISA). Different concentrations (1 nM, 10 nM, 50 nM, 100 nM, and 200 nM) of IgG Fab and Fab-Tetrazine were added to the protein L coated well plate and incubated for one hour. After removing the unbound protein, anti-human IgG (Fab specific)-peroxidase antibody was added to the well plate and incubated for one hour. Enhanced chemiluminescence solution was added to the well plate after washing three times with PBS. As shown in Fig. 6c, the binding affinities between IgG Fab and Fab-Tetrazine are comparable, suggesting that the binding of Fab-Tetrazine to protein L was preserved.

## Conclusions

Herein, we report a straightforward and broadly applicable bioconjugation method for introducing a single tetrazine group into a targeting peptide and antibody fragment at a predefined, distinct site in aqueous media through disulfide modification. The tetrazine functionalized peptide or Fab was further conjugated with a series of TCO-modified functionalities, such as dyes, polymers, or proteins/enzymes yielding a small library of stable bioconjugates within a short reaction time. Such bulky functionalities could only be introduced by a two-step procedure as the disulfide rebridging reaction is greatly limited by the steric demand of the reagent. Bulky substituents attached to the rebridging reagent such as polymers, proteins or drug molecules have only limited access to the reduced disulfide residues resulting in complex product mixtures with large amounts of unreacted starting materials and very low product yield.[29] Therefore, the two-step approach presented herein based on the site-selective installation of the tetrazine tag first followed by the IEDDA reaction provides convenient access to compound libraries, in which a broad range of small and bulky TCO-modified functionalities could be attached to the corresponding tetrazine-modified proteins in a convenient, fast, and efficient way.

Notably, the modified proteins exhibited good stability under biologically relevant conditions and their secondary structure was preserved after modification. We envision that this method would enable the preparation of the desired bioconjugates "on site", *e.g.* in hospitals for the desired applications. In this way, the final conjugate would be applied to the patient immediately after preparation, preventing long term storage, quality control and loss of bioactivity of the final conjugate. The approach presented herein is a valuable addition to the chemical toolbox for site-selective protein labelling and manipulation, offering a fast, robust and straightforward method to be easily adapted in any laboratory.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

## Notes and references

1 L. M. Weiner, R. Surana and S. Wang, *Nat. Rev. Immunol.*, 2010, **10**, 317.

2 V. Chudasama, A. Maruani and S. Caddick, *Nat. Chem.*, 2016, **8**, 114.

3 J. M. Chalker, G. J. Bernardes, Y. A. Lin and B. G. Davis, *Chem. – Asian J.*, 2009, **4**, 630–640.

4 (*a*) N. Krall, F. P. da Cruz, O. Boutureira and G. J. Bernardes, *Nat. Chem.*, 2016, **8**, 103–113; (*b*) O. Boutureira and G. J. L. Bernardes, *Chem. Rev.*, 2015, **115**, 2174–2195; (*c*) C. D. Spicer and B. G. Davis, *Nat. Commun.*, 2014, **5**, 4740; (*d*) G. Zhang, S. Zheng, H. Liu and P. R. Chen, *Chem. Soc. Rev.*, 2015, **44**, 3405–3417; (*e*) T. B. Nielsen, R. P. Thomsen, M. R. Mortensen, J. Kjems, P. F. Nielsen, T. E. Nielsen, A. L. B. Kodal, E. Cló and K. V. Gothelf, *Angew. Chem., Int. Ed.*, 2019, **58**, 9068–9072; (*f*) J. B. Trads, T. Tørring and K. V. Gothelf, *Acc. Chem. Res.*, 2017, **50**, 1367–1374; (*g*) C. D. Spicer, E. T. Pashuck and M. M. Stevens, *Chem. Rev.*, 2018, **118**, 7702–7743; (*h*) C. D. Spicer, T. Triemer and B. G. Davis, *J. Am. Chem. Soc.*, 2012, **134**, 800–803; (*i*) C. D. Spicer, C. Jumeaux, B. Gupta and M. M. Stevens, *Chem. Soc. Rev.*, 2018, **47**, 3574–3620.

5 (*a*) A.-C. Knall and C. Slugovc, *Chem. Soc. Rev.*, 2013, **42**, 5131–5142; (*b*) E. M. Sletten and C. R. Bertozzi, *Angew. Chem., Int. Ed.*, 2009, **48**, 6974–6998.

6 (*a*) B. L. Oliveira, Z. Guo and G. J. L. Bernardes, *Chem. Soc. Rev.*, 2017, **46**, 4895–4950; (*b*) A. C. Knall and C. Slugovc, *Chem. Soc. Rev.*, 2013, **42**, 5131–5142.

7 (*a*) N. K. Devaraj, R. Weissleder and S. A. Hilderbrand, *Bioconjugate Chem.*, 2008, **19**, 2297–2299; (*b*) N. K. Devaraj and R. Weissleder, *Acc. Chem. Res.*, 2011, **44**, 816–827; (*c*) N. K. Devaraj, R. Upadhyay, J. B. Haun, S. A. Hilderbrand and R. Weissleder, *Angew. Chem., Int. Ed.*, 2009, **48**, 7013–7016.

8 (*a*) J. B. Haun, N. K. Devaraj, S. A. Hilderbrand, H. Lee and R. Weissleder, *Nat. Nanotechnol.*, 2010, **5**, 660; (*b*) M. K. Rahim, R. Kota, S. Lee and J. B. Haun, *Nanotechnol. Rev.*, 2013, **2**, 215–227.

9 B. Oller-Salvia, G. Kym and J. W. Chin, *Angew. Chem., Int. Ed.*, 2018, **57**, 2831–2834.

10 (*a*) J. L. Seitchik, J. C. Peeler, M. T. Taylor, M. L. Blackman, T. W. Rhoads, R. B. Cooley, C. Refakis, J. M. Fox and R. A. Mehl, *J. Am. Chem. Soc.*, 2012, **134**, 2898–2901; (*b*) K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox and J. W. Chin, *J. Am. Chem. Soc.*, 2012, **134**, 10317–10320; (*c*) K. Lang, L. Davis, J. Torres-Kolbus, C. Chou, A. Deiters and J. W. Chin, *Nat. Chem.*, 2012, **4**, 298; (*d*) R. J. Blizzard, D. R. Backus, W. Brown, C. G. Bazewicz, Y. Li and R. A. Mehl, *J. Am. Chem. Soc.*, 2015, **137**, 10044–10047; (*e*) T. Machida, K. Lang, L. Xue, J. W. Chin and N. Winssinger, *Bioconjugate Chem.*, 2015, **26**, 802–806; (*f*) Y. Kurra, K. A. Odoi, Y.-J. Lee, Y. Yang, T. Lu, S. E. Wheeler, J. Torres-Kolbus, A. Deiters and W. R. Liu, *Bioconjugate Chem.*, 2014, **25**, 1730–1738.

11 (*a*) M. Baalmann, M. J. Ziegler, P. Werther, J. Wilhelm and R. Wombacher, *Bioconjugate Chem.*, 2019, **30**, 1405–1414; (*b*) 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.

12 (*a*) M. L. Blackman, M. Royzen and J. M. Fox, *J. Am. Chem. Soc.*, 2008, **130**, 13518–13519; (*b*) C. Canovas, M. Moreau, C. Bernhard, A. Oudot, M. Guillemin, F. Denat and V. Goncalves, *Angew. Chem., Int. Ed.*, 2018, **57**, 10646–10650.

13 S. L. Kuan, T. Wang and T. Weil, *Chem. – Eur. J.*, 2016, **22**, 17112–17129.

14 (*a*) S. Brocchini, S. Balan, A. Godwin, J.-W. Choi, M. Zloh and S. Shaunak, *Nat. Protoc.*, 2006, **1**, 2241–2252; (*b*) S. Shaunak, A. Godwin, J.-W. Choi, S. Balan, E. Pedone, D. Vijayarangam, S. Heidelberger, I. Teo, M. Zloh and S. Brocchini, *Nat. Chem. Biol.*, 2006, **2**, 312–313.

15 S. P. Brown and A. B. Smith, *J. Am. Chem. Soc.*, 2015, **137**, 4034–4037.

16 G. Badescu, P. Bryant, M. Bird, K. Henseleit, J. Swierkosz, V. Parekh, R. Tommasi, E. Pawlisz, K. Jurlewicz, M. Farys, N. Camper, X. Sheng, M. Fisher, R. Grygorash, A. Kyle, A. Abhilash, M. Frigerio, J. Edwards and A. Godwin, *Bioconjugate Chem.*, 2014, **25**, 1124–1136.

17 T. Wang, A. Riegger, M. Lamla, S. Wiese, P. Oeckl, M. Otto, Y. Wu, S. Fischer, H. Barth, S. L. Kuan and T. Weil, *Chem. Sci.*, 2016, **7**, 3234–3239.

18 (*a*) F. F. Schumacher, M. Nobles, C. P. Ryan, M. E. B. Smith, A. Tinker, S. Caddick and J. R. Baker, *Bioconjugate Chem.*, 2011, **22**, 132–136; (*b*) M. Henkel, N. Röckendorf and A. Frey, *Bioconjugate Chem.*, 2016, **27**, 2260–2265.

19 R. Rossin, S. M. van den Bosch, W. ten Hoeve, M. Carvelli, R. M. Versteegen, J. Lub and M. S. Robillard, *Bioconjugate Chem.*, 2013, **24**, 1210–1217.

20 S. L. Kuan, S. Fischer, S. Hafner, T. Wang, T. Syrovets, W. Liu, Y. Tokura, D. Y. W. Ng, A. Riegger, C. Förtsch, D. Jäger, T. F. E. Barth, T. Simmet, H. Barth and T. Weil, *Adv. Sci.*, 2018, 1701036.

21 T. Wang, N. Zabarska, Y. Wu, M. Lamla, S. Fischer, K. Monczak, D. Y. Ng, S. Rau and T. Weil, *Chem. Commun.*, 2015, **51**, 12552–12555.

22 (*a*) D. A. Richards, A. Maruani and V. Chudasama, *Chem. Sci.*, 2017, **8**, 63–77; (*b*) A. Maruani, M. E. Smith, E. Miranda, K. A. Chester, V. Chudasama and S. Caddick, *Nat. Commun.*, 2015, **6**, 6645.

1146 | *Org. Biomol. Chem.*, 2020, **18**, 1140–1147

87

This journal is © The Royal Society of Chemistry 2020

23 S. Sun, P. Akkapeddi, M. C. Marques, N. Martinez-Saez, V. M. Torres, C. Cordeiro, O. Boutureira and G. J. L. Bernardes, *Org. Biomol. Chem.*, 2019, **17**, 2005–2012.

24 F. R. Salemme, *Annu. Rev. Biochem.*, 1977, **46**, 299–330.

25 K. Qu, J. L. Vaughn, A. Sienkiewicz, C. P. Scholes and J. S. Fetrow, *Biochemistry*, 1997, **36**, 2884–2897.

26 T. Wang, D. Y. W. Ng, Y. Wu, J. Thomas, T. TamTran and T. Weil, *Chem. Commun.*, 2014, **50**, 1116–1118.

27 (*a*) Y. Tachibana, G. L. Fletcher, N. Fujitani, S. Tsuda, K. Monde and S.-I. Nishimura, *Angew. Chem., Int. Ed.*, 2004, **43**, 856–862; (*b*) S. Sun, I. Compañón, N. Martínez-Sáez, J. D. Seixas, O. Boutureira, F. Corzana and G. J. L. Bernardes, *ChemBioChem*, 2018, **19**, 48–52.

28 A. C. A. Roque, M.Â Taipa and C. R. Lowe, *J. Chromatogr. A*, 2005, **1064**, 157–167.

29 T. Wang, Y. Wu, S. L. Kuan, O. Dumele, M. Lamla, D. Y. W. Ng, M. Arzt, J. Thomas, J. O. Mueller, C. Barner-Kowollik and T. Weil, *Chem. Eur. J.*, 2015, **21**, 228–238.

This journal is © The Royal Society of Chemistry 2020

88

*Org. Biomol. Chem.*, 2020, **18**, 1140–1147 | **1147**

# Supporting Information for

# Site-selective protein modification via disulfide rebridging for fast tetrazine/*trans*-cyclooctene bioconjugation

Lujuan Xu[a,b], Marco Raabe[a, b], Maksymilian Marek Zegota,[a, b] João C. F. Nogueira[c,], Vijay Chudasama[c], Seah Ling Kuan*[a,b], Tanja Weil*[a,b]

a. Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

b. Institute of Inorganic Chemistry I, University of Ulm, Albert-Einstein-Allee 11, 89081 Ulm, Germany

c. Department of Chemistry, University College London, London, UK

**1. General experimental details**

**2. Synthetic Protocol of IC-Tetrazine**

**3. Synthesis of tetrazine modified somatostatin (SST-Tetrazine)**

**4. Synthesis of tetrazine modified Fab fragment (Fab-Tetrazine)**

**5. Synthesis and characterization of CytC-PEG3-TCO**

**6. SST-Tetrazine conjugation with TCO-PEG12, TCO-Cy5 and CytC-PEG3-TCO**

**7. Fab-Tetrazine conjugation with TCO-Cy5, TCO-PEG113**

**8. Stability study of SST-Tetrazine, Fab-Tetrazine and Fab-Cy5**

**9. CD experiment of SST-Tetrazine and Fab-Tetrazine**

**10. Enzyme-linked immunosorbent assay (ELISA) experiment**

**11. Reference**

## 1. General experimental details

Unless otherwise stated, all the operations were performed without taking precautions to exclude moisture and air. All the organic solvents ($CH_2Cl_2$, $CHCl_3$, Ethyl Acetate (EA), acetonitrile (ACN)) were bought from commercial sources (Merck, Sigma Aldrich and so on) and are used directly without further purification. IgG Fab Fragment from Human (polyclonal) was obtained from Dianova GmbH and used as bought. Methyltetrazine *N*-hydroxysuccinimide (NHS) ester, *trans*-cyclooctene (TCO)-PEG3-maleimide were bought from Click Chemistry Tools. Sulfo-Cy5-TCO was bought from Santa Cruz Biotechnology. Tetrazine-5-FAM were purchased from Jena Bioscience. TCO-PEG113 was obtained from NANOCS. TCO-PEG12 was bought from Broadpharm. Iso-yeast Cytochrome C was purchased from Sigma-Aldrich. Glutathione was bought from Sigma-Aldrich. Protein L coated well plate, BCA assay and enhanced chemiluminescence (ECL) solution were purchased from Thermo Fisher Scientific. 4,4′-dithiodipyridine was bought from ACROS organics. Zeba spin desalting column (7K MWCO) was obtained from Thermo Fisher Scientific. $H_2O$ used for the reactions was obtained from the Millipore purification system. Vivaspin sample concentrators were purchased from GE Healthcare. Reaction progress was monitored by thin layer chromatography (TLC) using Merck 60 $F_{254}$ pre-coated silica gel plates and visualized under ultraviolet lamp (254 nm) or using appropriate staining solution ($KMnO_4$, ninhydrin, iodine). Flash column chromatography was carried out using Merck silica gel 60 mesh. High performance liquid chromatography (HPLC) was carried out using Shimadzu Analytical HPLC system. NMR spectra were recorded on Bruker Avance 300 NMR spectrometer and the chemical shifts ($\delta$) were reported as parts per million (ppm) referenced with respect to the residual solvent peaks. MALDI-TOF-MS spectra were acquired on a Bruker Time-of-flight MS rapifleX. Fluorescence (or absorbance) spectra or intensities were measured on microplate reader (Tecan Spark 20M). The chemiluminescence was measured by GloMax Discover Microplate Reader.

## 2. Synthetic Protocol of IC-Tetrazine
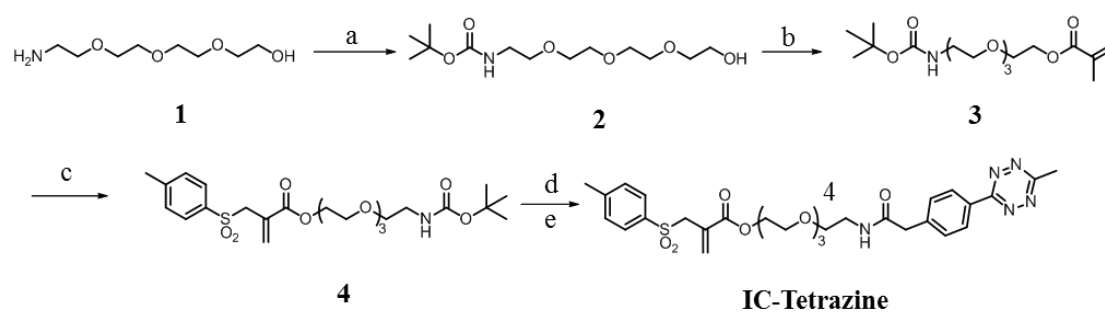


Scheme 1 Synthesis of tetrazine-containing disulfide rebridging reagent (IC-Tetrazine). (a) Di-*tert*-butyl dicarbonate, $CH_2Cl_2$, overnight. (b) methacrylchloride, $Et_3N$, $CH_2Cl_2$, 90%. (c)

1. $I_2$, sodium *p*-toluenesulfinate, $CH_2Cl_2$, 3 days. 2. $Et_3N$, $CH_2Cl_2$, overnight. 3. $Et_3N$, ethyl acetate, 95 °C, overnight, yield after all three steps: 60%. (d) Trifluoroacetic acid (TFA), $CH_2Cl_2$, 98%. (e) methyltetrazine NHS ester, $Et_3N$, $CH_2Cl_2$, overnight, 40%.

## Synthesis of 2,2-dimethyl-4-oxo-3,8,11,14-tetraoxa-5-azahexadecan-16-yl methacrylate (compound 3)

**Step a:** 2-(2-(2-(2-aminoethoxy)ethoxy)ethoxy)ethanol (**1**, 0.50 g, 2.59 mmol) was dissolved in 6 mL $CH_2Cl_2$. Di-*tert*-butyl dicarbonate (0.68 g, 3.10 mmol) was also dissolved in 6 mL $CH_2Cl_2$ and then was added to the former solution dropwise with a syringe pump. The reaction mixture was stirred at room temperature overnight and the solvent was removed under vacuum. The compound **2** was got as a colorless oil and used directly for the next step without further purification.

**Step b:** To a 10 mL $CH_2Cl_2$ solution of crude product from step a (200 mg, 0.68 mmol) was added $Et_3N$ (82.9 mg, 114 μL, 0.82 mmol) followed by methacryloyl chloride (85.7 mg, 0.82 mmol) at 0 °C. The mixture was stirred at room temperature overnight. The solvent was evaporated under high vacuum and the crude product was dissolved in 30 mL $CH_2Cl_2$. Subsequently, the crude product was washed with 1 M HCl and brine solution. The organic layer was dried over anhydrous $Na_2SO_4$ and the solvent was removed under vacuum. The crude product was purified by column chromatography (EA : hexane = 1:1) to afford the compound **3** as colorless oil in 90% yield.

**$^1$H NMR** (300 MHz, chloroform-d) δ 6.10 (s, 1H), 5.57–5.53 (s, 1H), 4.32 – 4.23 (m, 2H), 3.77 – 3.54 (m, 12H), 2.47 (t, J = 6.6 Hz, 2H), 1.92 (s, 3H), 1.42 (s, 9H).

**$^{13}$C NMR** (75 MHz, chloroform-d) δ 167.09, 155.92, 135.98, 125.83, 70.58, 70.24, 69.16, 63.86, 28.42, 18.31.

**LC-MS**: *m/z* = 362 [M+H]$^+$, 384 [M+Na]$^+$ (calculated exact mass: 362.21 [M+H]$^+$, 384.21 [M+Na]$^+$, formula:$C_{17}H_{31}NO_7$ )

## Synthesis of 2,2-dimethyl-4-oxo-3,8,11,14-tetraoxa-5-azahexadecan-16-yl 2-(tosylmethyl)-acrylate (compound 4)

Compound **3** (211 mg, 0.58 mmol) was dissolved in 5 mL $CH_2Cl_2$ followed by adding the sodium *p*-toluenesulfinate (156 mg, 0.87 mmol) and $I_2$ (222 mg, 0.87 mmol) sequentially. The reaction mixture was stirred at room temperature for 3 days. Next, $Et_3N$ (176 mg, 242 μL, 1.74 mmol) was added to the mixture and stirred overnight. Then the organic layer was washed with 1 M HCl solution, saturated $NaHCO_3$ solution, $Na_2S_2O_3$ solution and brine solution. The organic layer was dried over $MgSO_4$ and the solvent was removed by high vacuum. The residue was dissolved in 10 mL ethyl acetate and $Et_3N$ (176 mg, 242 μL, 1.74 mmol) was added dropwise at 0 ℃. The mixture was refluxed overnight. After that, the solvent was evaporated

and the crude product was purified by column chromatography (EA : hexane = 3:1) to afford compound **4** as slight yellow oil in 60% yield.

**¹H NMR** (300 MHz, chloroform-*d*) δ 7.72 (d, *J* = 8.3 Hz, 2H), 7.31 (d, *J* = 8.0 Hz, 2H), 6.50 (s, 1H), 5.86 (s, 1H), 4.17 – 4.11 (m, 4H), 3.68 – 3.58 (m, 10H), 3.52 (t, *J* = 5.2 Hz, 2H), 3.28 (t, *J* = 4.8 Hz, 2H), 2.42 (s, 3H), 1.42 (s, 9H).

**¹³C NMR** (75 MHz, chloroform-*d*) δ 164.82, 155.98, 144.87, 135.37, 133.44, 129.61, 128.67, 79.25, 70.49, 70.23, 68.81, 64.50, 57.57, 40.52, 28.42, 21.54.

**LC-MS**: *m/z* = 516 [M+H]$^+$, 538 [M+Na]$^+$ (calculated exact mass: 516.22 [M+H]$^+$, 538.22 [M+Na]$^+$, formula:$C_{24}H_{37}NO_9S$ )

**Synthesis of 1-(4-(6-methyl-1,2,4,5-tetrazin-3-yl)phenyl)-2-oxo-6,9,12-trioxa-3-azatetradecan-14-yl 2-(tosylmethyl)acrylate (IC-Tetrazine)**

**Step d:** In a 10 mL flask, compound **4** (15.1 mg, 0.03 mmol) was dissolved in 3 mL $CH_2Cl_2$. Trifluoroacetic acid (TFA, 171.2 mg, 1.54 mmol) was also added. The resulting mixture was stirred at room temperature overnight. The solvent and TFA was removed in vacuum to obtain 11.9 mg boc-deprotected product in 98%.

**Step e:** Boc-deprotected product (11.9 mg, 0.03 mmol) was dissolved in 2 mL $CH_2Cl_2$ followed by adding *N,N*-diisopropylethylamine (DIEA, 7.5 mg, 0.06 mmol) to ensure a basic condition. Methyltetrazine-NHS ester (11.4 mg, 0.03 mmol) was dissolved in 1 mL $CH_2Cl_2$ and added to the aforementioned mixture above. The reaction mixture was stirred at room temperature overnight and the solvent was evaporated under high vacuum. The residue was dissolved in 20 mL $CH_2Cl_2$ and was washed with brine solution. The organic layer was dried over anhydrous $MgSO_4$ and the solvent was removed under high vacuum. The residue was purified by column chromatography ($CH_2Cl_2$ : MeOH = 20:1) to afford compound **IC-Tetrazine** as red oil in 40% yield. The purify of IC-Tetrazine is 98% based on the quantification of the peak at 214 nm in LC-MS.

**¹H NMR** (300 MHz, chloroform-*d*) δ 8.54 (d, *J* = 7.5 Hz, 2H), 7.72 (d, *J* = 7.6 Hz, 2H), 7.51 (d, *J* = 7.8 Hz, 2H), 7.33 (d, *J* = 7.5 Hz, 2H), 6.50 (s, 1H), 5.83 (s, 1H), 4.25 – 4.08 (m, 4H), 3.75 – 3.37 (m, 16H), 3.09 (s, 3H), 2.44 (s, 3H).

**¹³C NMR** (75 MHz, CDCl₃) δ 170.10, 167.24, 164.70, 163.85, 145.37, 144.96, 140.00, 135.50, 135.21, 133.46, 130.23, 129.73, 128.93, 128.73, 128.28, 70.62, 70.49, 70.25, 69.77, 68.81, 64.48, 57.60, 55.28, 43.48, 39.53, 35.78, 21.72, 21.12.

**LC-MS**: *m/z* = 628 [M+H]$^+$, 650 [M+Na]$^+$(calculated exact mass: 628.24 [M+H]$^+$, 650.24 [M+Na]$^+$, formula:$C_{30}H_{37}N_5O_8S$ )
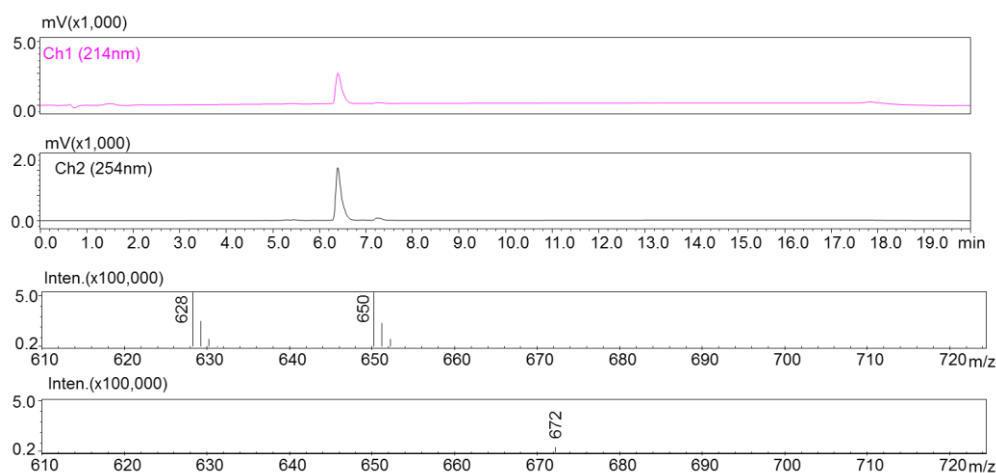
Figure S1 LC-MS data of IC-Tetrazine

## 3. Synthesis of tetrazine modified somatostatin (SST-Tetrazine)
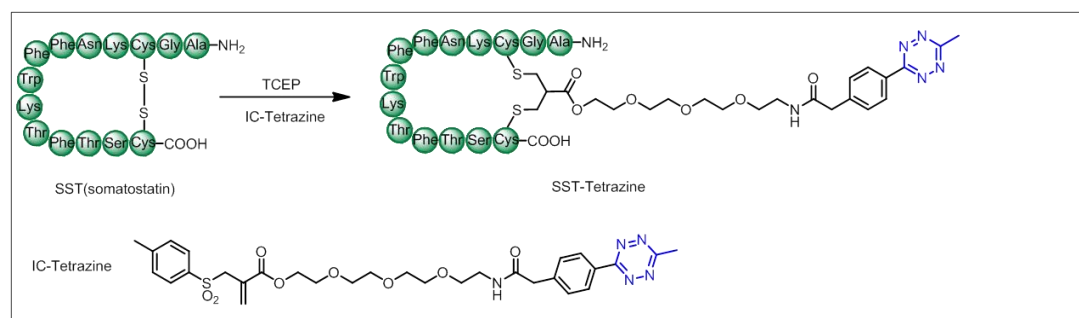


Scheme S2 Synthesis of SST-Tetrazine

Somatostatin (SST, 2 mg, 1.22 μmol) was dissolved in 1 mL ACN/phosphate buffer (PB, 50 mM, pH = 7.8) mixture (v/v = 2:3). Tris(2-carboxyethyl)phosphine (TCEP, 0.7 mg, 2.44 μmol) was dissolved in 100 μL ACN/PB mixture and added to the somatostatin solution. The mixture was incubated at room temperature for 30 min. Afterwards, IC-Tetrazine (1.9 mg, 3.05 μmol) was dissolved in 1 mL ACN/PB mixture as well which was added to the SST and TCEP solution. The reaction mixture was degassed for several minutes to exclude air in the solvent and gently shaken at room temperature overnight. The crude product was purified by Prep HPLC using Atlantis Prep OBD T3 Column (19×100 mm, 5 μm) with the mobile phase starting from 100% solvent A (0.1% TFA in MilliQ water) and 0% solvent B (0.1% TFA in ACN), reaching 45% B at 35 min and finally reaching 100% B at 42 min with a flow rate of 10 mL/min. The absorbance was monitored at 280 nm, 254 nm and 515 nm. The retention time for SST-Tetrazine was 34 min. About 0.8 mg SST-Tetrazine was obtained after lyophilisation (30% yield).

**LC-MS**: $m/z$ = 1056 [M+2H]$^{2+}$, 705 [M+3H]$^{3+}$, 1054 [M-2H]$^{2-}$

**MALDI-TOF-MS** (matrix: sinapinic acid): $m/z = 2110.95$ [M+H]$^+$ (calculated exact mass: 2110.95 [M+H]$^+$, formula: $C_{99}H_{135}N_{23}O_{25}S_2$)
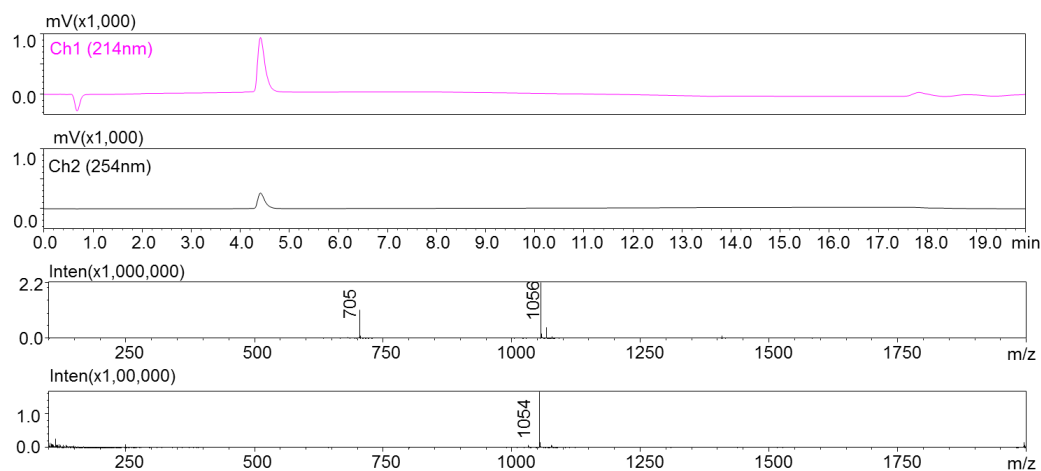


Figure S2 LC-MS data of purified SST-Tetrazine
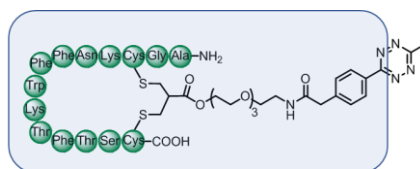


Fragmentation of SST-Tetrazine
Mr = 2041.93 Da

Figure S3 The chemical structure of the fragmentation part (in the frame) of SST-Tetrazine in MALDI-TOF-MS spectrum

Furthermore, SST-Tetrazine is also showing no obvious degradation after one year storage in the freezer based on the LC-MS data.
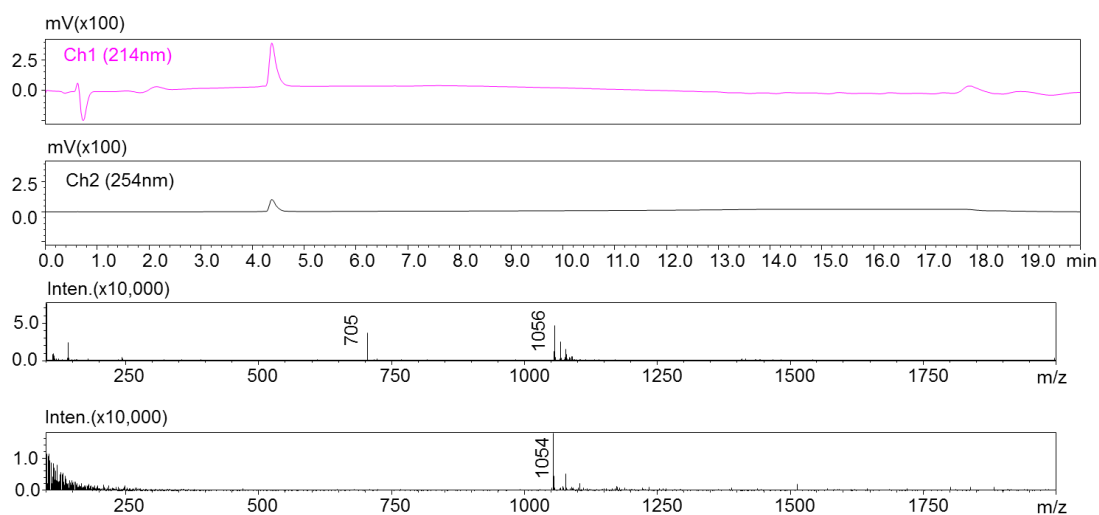


Figure S4 LC-MS data of SST-Tetrazine after one year storage in the freezer ($-20\,°C$)

## 4. Synthesis of tetrazine modified Fab fragment (Fab-Tetrazine)
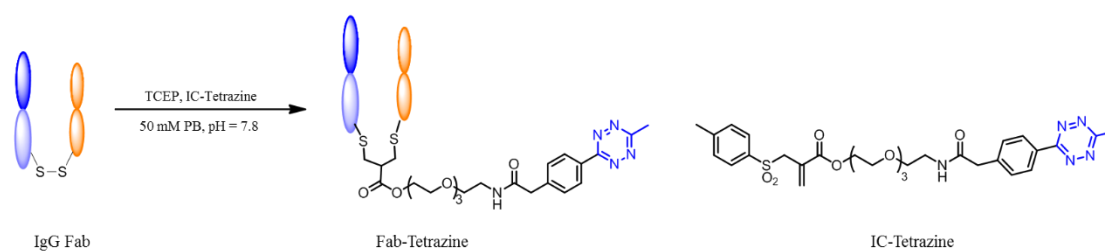


Scheme S3 Synthesis of Fab-Tetrazine

First, native IgG Fab was incubated with different amounts of TCEP to determine how much TCEP is necessary to totally reduce the interchain disulfide bond of IgG Fab. IgG Fab (3.2 μg, 0.067 nmol, 4.8 mg/mL, 1 eq) was mixed with 1 eq, 5 eq, 10 eq, 25 eq, 50 eq, 100 eq TCEP for incubation for 1 hour. Then, the mixture was loaded to the SDS-PAGE. SDS-PAGE data (Figure S5A) showed that 50 eq TCEP is necessary to fully reduce the interchain disulfide bond of IgG Fab.

Second, the buffer of the Fab solution (100 μg, 4.8 mg/mL, 2.08 nmol) was changed to PB (50 mM, pH = 7.8) by using 10 kDa MWCO ultrafiltration tube and diluted to 1 mg/mL with PB (50 mM, pH = 7.8). TCEP (30 μg, 0.10 μmol) was dissolved in 10 μL PB (50 mM, pH = 7.8) and added to the Fab solution. Next the mixture was incubated at room temperature for 1 hour. IC-Tetrazine (65.5 μg, 0.10 μmol, 20 mg/mL in dimethyl sulfoxide (DMSO)) was first dissolved in 5.5 μL DMSO and then added to the Fab solution. The mixture was incubated at room temperature overnight. Excess reagents were removed by repeated ultrafiltration in water using Vivaspin sample concentrator (10 kDa MWCO). The sample was analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). Based on the concentration calculated from gel densitometry, about 80% of the Fab fragment was successfully modified (SDS-PAGE is shown in Figure S5B).

SDS gel electrophoresis: 10% separating gel was prepared in the following order: 2.5 mL 40% acrylamide solution, 5 mL water, 2.5 mL 4× Bis-Tris gel buffer, 100 μL of 10% sodium dodecyl sulfate buffer, 50 μL ammonium persulfate (v/w) solution, 5 μL tetramethylethylenediamine. The mixed gel solution was added to the Bio-Rad Mini-Gel apparatus and stand for 15 min for polymerization. The isopropanol was added above the separating gel to exclude air. Next, 6% stacking gel was prepared in the following order: 0.75 mL 40% acrylamide solution, 3 mL water, 1.25 mL 4× Bis-Tris gel buffer, 30 μL 10% sodium dodecyl sulfate buffer, 25 μL ammonium persulfate (v/w) solution, 2.5 μL tetramethylethylenediamine. The stacking gel was also added above the separating gel after totally removing the isopropanol. A 10-well comb was inserted and the stacking gel was polymerized for another 10 min. For the sample loading to the SDS-PAGE, 16 μL protein solution was mixed with 6 μL loading dye and 1 μL TCEP solution. The mixture was incubated for one hour and then loaded to the gel.

On the other hand, a thiol-quantification experiment was performed to quantify how much thiol groups are still left in Fab-Tetrazine sample. 4,4′-Dithiodipyridine is a disulfide-containing reagent which reacts with protein thiols to form stoichiometric amounts of the chromogenic compound 4-thiopyridone (4-TP), absorbing at 324 nm. In this experiment, cysteine was selected as a standard compound. The procedure is described below:

1. 30 μL of 4,4′-dithiodipyridine (2 mM in 100 mM citrate buffer, pH = 4.5) incubated with 30 μL cysteine solution (concentration: 0.01 mM, 0.05 mM, 0.1 mM, 0.25 mM, 0.5 mM, dissolved in 100 mM citrate buffer, pH = 4.5) for 15 minutes at room temperature. 50 μL of cysteine and 4,4′-dithiodipyridine mixture were taken in 384-well plate. Absorbance was monitored at 324 nm and the standard curve is shown in Figure S5C.

2. 50 μL of Fab-Tetrazine sample was mixed with 100 eq TCEP (100 mM PB, pH = 7.8) and incubated for 1 hour to tally reduce the unmodified disulfide bonds.

3. Zeba spin desalting columns (7K MWCO) was buffer exchanged to the 100 mM citrate buffer first following the protocol provided by the supplier. Next, the Fab-Tetrazine solution was loaded to the column and the desalting steps was repeated six to eight times to tally remove the residue TCEP. After desalting, the protein concentration was determined to be 3.21 mg/mL based on the BCA assay following the protocol provided by the supplier.

4. Next, 30 μL of 4,4′-dithiodipyridine (2 mM in 100 mM citrate buffer, pH = 4.5) was incubated with 30 μL the Fab-Tetrazine solution at room temperature for 15 minutes. 50 μL of the protein and 4,4′-dithiodipyridine mixture were taken in 384-well plate. Absorbance was monitored at 324 nm and 0.51 was obtained for the Fab-Tetrazine sample.

Based on the standard curve, 0.51 indicated the concentration of thiol groups in 60 μL protein and 4,4′-dithiodipyridine mixture is 0.018 mM. The concentration of unmodified disulfide bond is 0.009 mM. The concentration of Fab-Tetrazine is 1.6 mg/mL (0.033 mM) in 60 μL protein and 4,4′-dithiodipyridine mixture. From this readout, about 73% of IgG Fab was successfully modified.
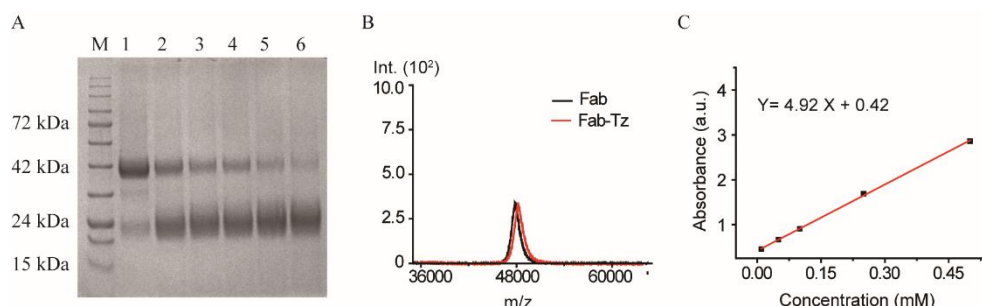


Figure S5 (A) SDS-PAGE analysis of Fab fragment which was incubated with different amounts of TCEP (M: Protein Marker, 1: Fab + 1 eq TCEP, 2: Fab + 5 eq TCEP, 3: Fab + 10 eq TCEP, 4: Fab + 20 eq TCEP, 5: Fab + 40 eq TCEP, 6: Fab + 50 eq TCEP ) (B) MALDI-

TOF-MS of native IgG Fab (found: 48.0k [M+H]$^+$, calculated: 48.0k [M+H]$^+$) and Fab-Tetrazine (found: 48.5k [M+H]$^+$, calculated: 48.5k [M+H]$^+$) (C) standard curve based on checking the absorbance of cysteine and 4,4′-dithiodipyridine solution at 324 nm

## 5. Modification and characterization of Cytochrome C (CytC) with *trans*-cyclooctene (TCO) group
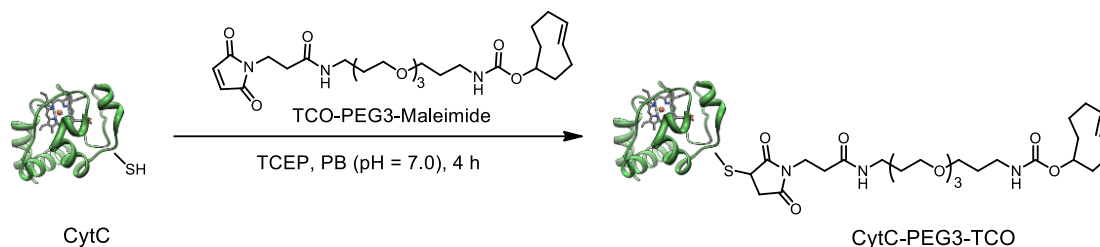


Scheme S4 Modification of Cytochrome C with TCO-PEG3-Maleimide

CytC (212 µg, 0.02 µmol) was dissolved in 212 µL PB (50 mM, pH = 7.0) and mixed with TCEP (9.7 µg, 0.03 µmol). The mixture was shaken at room temperature for 1 hour. TCO-PEG3-Maleimide (267 µg, 0.51 µmol) was added to the mixture above and shaken at room temperature for 4 hours. Afterwards, the crude product was purified by repeated ultrafiltration in water using Vivaspin sample concentrator (5 kDa MWCO). The MALDI-TOF-MS of the modified product (CytC-PEG3-TCO) is shown in Figure S6B. Compared to the native CytC (Figure S6A), the molecular weight of CytC-PEG3-TCO increased by 525 Da which corresponds to the molecular weight of TCO-PEG3-Maleimide (523.62 Da).
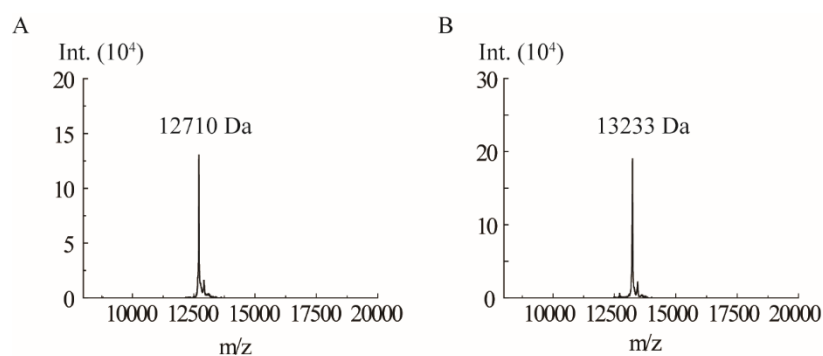


Figure S6 MALDI-TOF-MS of (A) native CytC (found: 12710 [M+H]$^+$, calculated: 12588 [M+H]$^+$) and (B) CytC-PEG3-TCO (found: 13233 [M+H]$^+$, calculated: 13235 [M+H]$^+$)
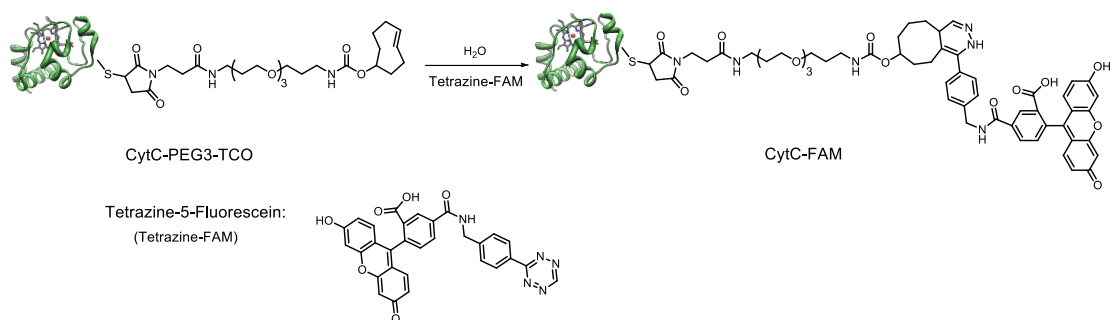
Scheme S5 Reaction between CytC-PEG3-TCO and Tetrazine-FAM

In order to check the modification yield of CytC with the TCO group, CytC-PEG3-TCO was used to react with a dye (Tetrazine-FAM) in $H_2O$. After reaction, the modification yield was estimated based on the absorbance of CytC at 280 nm and the FAM dye at 490 nm.

CytC-PEG3-TCO (70 μg, 5.29 nmol) was mixed with 2.27 μL Tetrazine-FAM (25.7 μg, 0.05 μmol, 10 mg/mL stock solution in DMSO) in $H_2O$. The mixture was shaken at room temperature for 1 hour. After that, the crude product was purified by repeated ultrafiltration in water using Vivaspin sample concentrator (5 kDa MWCO) to get the purified CytC-PEG3-FAM. The degree of labelling (DOL) was calculated based on the following equation:

$$DOL = \frac{A_{max} \times \varepsilon_{280} \, (protein)}{(A_{280} - A_{max} \times CF) \times \varepsilon_{max}}$$

We check the absorbance of CytC-PEG3-TCO at 280 nm and 490 nm, the values are listed below:

$A_{280}$ (the absorbance of the sample at 280 nm) = 0.42,

$A_{490}$ (the absorbance of the sample at 490 nm) = 0.89

$\varepsilon_{280}$ (the extinction coefficient of CytC at 280 nm) = 13075 $M^{-1}cm^{-1}$

$\varepsilon_{max}$ (the extinction coefficient of FAM dye at 490 nm) = 83000 $M^{-1}cm^{-1}$

CF (correction factor) = 0.3

Based on the calculation, the DOL is 90%.

Absorbance at 280 nm ($A_{280}$) is used to determine the protein concentration in a sample. However, fluorescent dyes also absorb at 280 nm. Therefore, a correction factor must be used to adjust for $A_{280}$ contributed by the dye. The correction factor (CF) equals the $A_{280}$ of the dye divided by the $A_{max}$ of the dye as shown in the equation above.

## 6. SST-Tetrazine conjugation with TCO-PEG12, TCO-Cy5 and CytC-PEG3-TCO
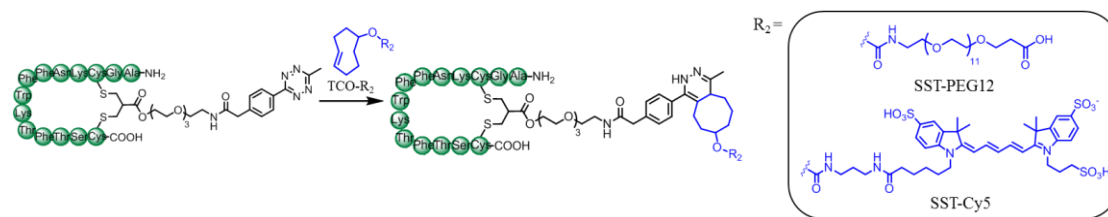
Scheme S6 Bioconjugation between SST-Tetrazine, TCO-PEG12 and TCO-Cy5

SST-Tetrazine (2 mg, 0.95 μmol) was dissolved in 1.332 mL PB (50 mM, pH = 7.4). TCO-PEG12 (730 μg, 1.00 μmol) was dissolved in 667 μL PB (50 mM, pH = 7.4) and added to the SST-Tetrazine solution. The reaction mixture was shaken for 30 min. After mixing the two components, the pink color originating from the tetrazine group disappeared immediately. The crude product was purified by semi-preparative HPLC using Agilent Eclipse XDB-C18 column (9.4 × 250 mm, 5 μm) with the mobile phase starting from 100% solvent A (0.1% TFA in MilliQ water) and 0% solvent B (0.1% TFA in ACN), reaching 43% B at 35 min and finally reaching 100% B at 38 min with a flow rate of 4 mL/min. The absorbance was monitored at 280 nm and 254 nm. The retention time for SST-PEG12 was 32 min. 2.56 mg SST-PEG12 was obtained (yield: 95%).

LC-MS: $m/z$ = 718 [M+Na+3H]$^{4+}$, 959 [M+Na+2H]$^{3+}$, 966 [M+2Na+H]$^{3+}$, 1438 [M+Na+H]$^{2+}$, 1424 [M-H]$^{-}$

MALDI-TOF-MS (matrix: sinapinic acid): $m/z$ = 2852.2102 [M+H]$^{+}$ (calculated exact mass: 2852.3873 [M+H]$^{+}$, formula: $C_{135}H_{202}N_{22}O_{41}S_{2}$).
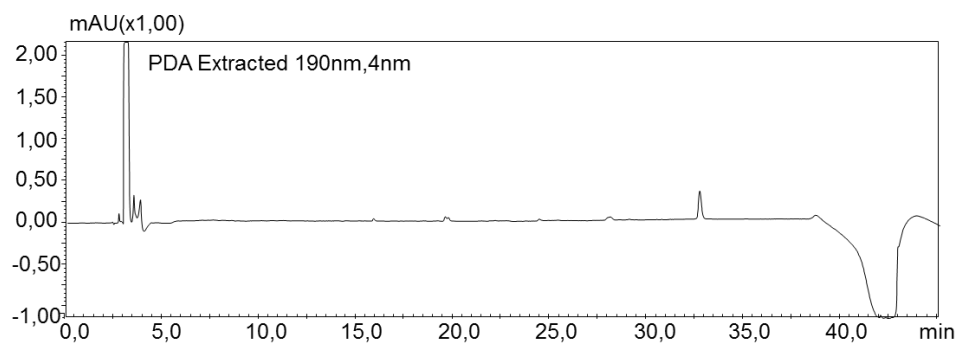


Figure S7 HPLC chromatogram (280 nm detection wavelength) of SST-PEG12
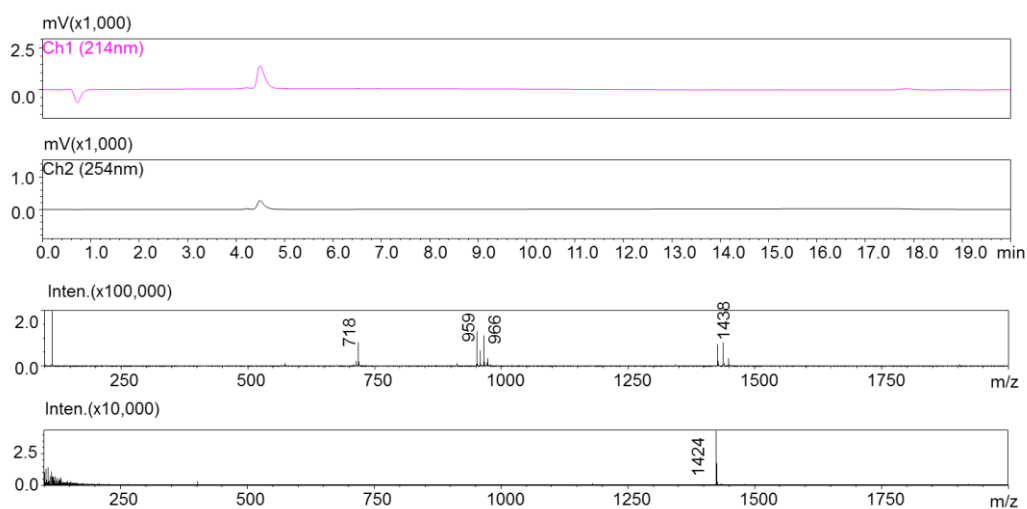
Figure S8 LC-MS data of SST-PEG12

SST-Tetrazine (180 mg, 0.85 μmol) was dissolved in 1.26 mL PB (50 mM, pH = 7.4). TCO-Cy5 (1.03 μg, 0.9 μmol) was dissolved in 252 μL DMSO and added to the SST-Tetrazine solution. The reaction mixture was stirred for 30 min. The crude product was purified by Semipreparative HPLC using Agilent Eclipse XDB-C18 column (9.4×250 mm, 5 μm) with the mobile phase starting from 100% solvent A (0.1% TFA in MilliQ water) and 0% solvent B (0.1% TFA in ACN), reaching 43% B at 35 min and finally reaching 100% B at 38 min with a flow rate of 4 mL/min. The absorbance was monitored at 280 nm and 254 nm. The retention time for SST-Cy5 was 28.5 min. 2.3 mg SST-Cy5 was obtained (yield: 90%).

LC-MS: $m/z$ = 1014 [M+3H]$^{3+}$, 1520 [M-2H]$^{2+}$, 1012 [M-3H]$^{3-}$

MALDI-TOF-MS (matrix: sinapinic acid): $m/z$ = 3041.1742 [M+H]$^+$ (calculated exact mass: 3041.2939 [M+H]$^+$, formula: $C_{145}H_{197}N_{25}O_{37}S_5$)
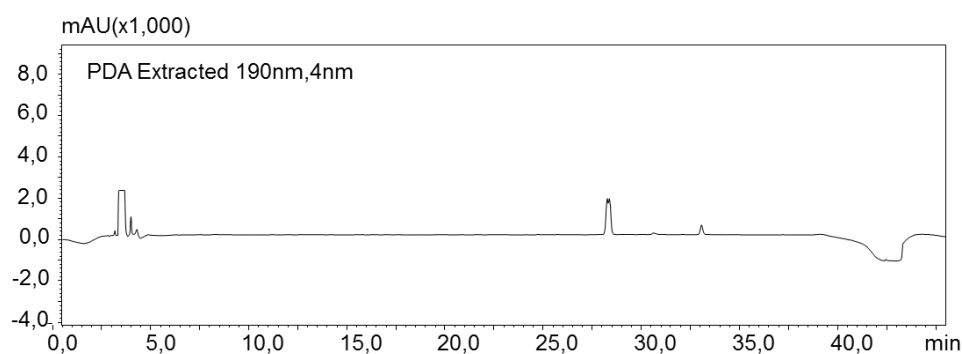


Figure S9 HPLC chromatogram (280 nm detection wavelength) of SST-Cy5
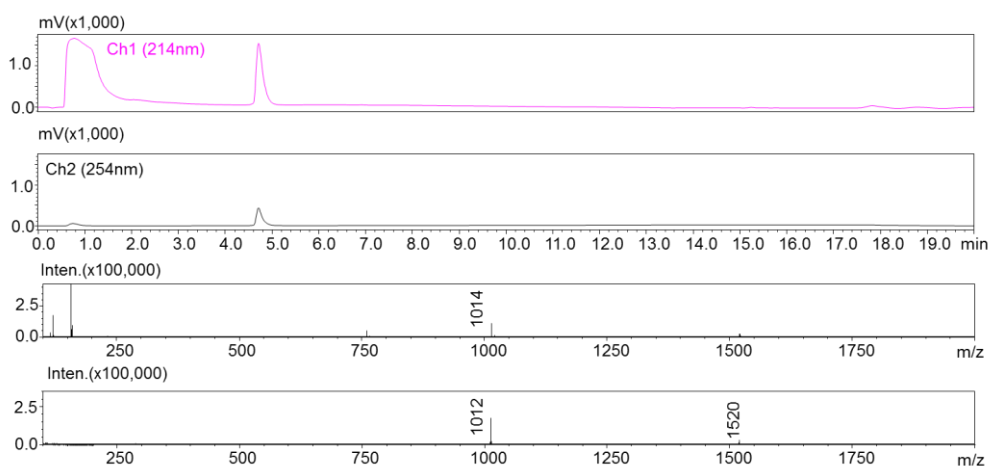
Figure S10 LC-MS data of SST-Cy5
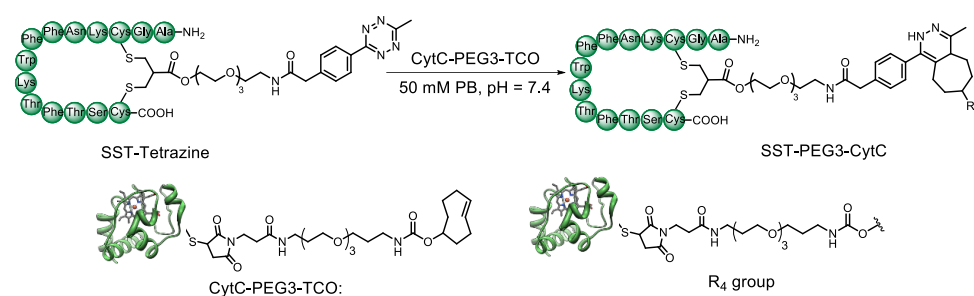


Scheme S7 Bioconjugation between SST-Tetrazine and CytC-PEG3-TCO

SST-Tetrazine (32 μg, 0.015 μmol) was mixed with CytC-PEG3-TCO (132 μg, 0.010 μmol) in PB (50 mM, pH = 7.4). The mixture was shaken for 30 min. Afterwards, the reaction mixture was purified by repeating ultrafiltration in water using Vivaspin sample concentrator (5 kDa MWCO) to get the purified SST-PEG3-CytC. The desired SST-PEG3-CytC was characterized by MALDI-TOF-MS and SDS-PAGE.

## 7. Fab-Tetrazine conjugation with TCO-Cy5 and TCO-PEG113
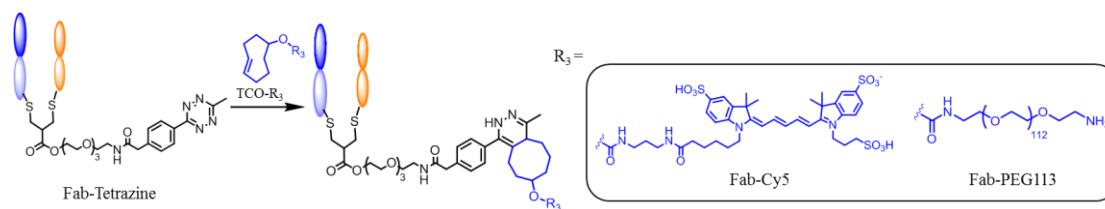


Scheme S8 Bioconjugation between Fab-Tetrazine, TCO-Cy5 and TCO-PEG113

Fab-Tetrazine (48 μg, 1.00 nmol) was dissolved in 85 μL PB (50 mM, pH = 7.4). TCO-Cy5 (5.7 μg, 5.00 nmol, 20 mg/mL in DMSO stock solution) was added to the Fab-Tetrazine solution. The reaction mixture was incubated at room temperature for 30 min. Afterwards, the crude product was purified by repeated ultrafiltration in water using Vivaspin sample

concentrator (10 kDa MWCO). Based on Figure S11A, Fab-Cy5 showed obvious fluorescence compared to the Fab-Tetrazine indicating the successful conjugation.

The degree of labelling of Fab-Cy5 was calculated using two different methods.
One method to determine the degree of labeling of Fab-Cy5 is based on the following equation:

$$DOL = \frac{A_{max} \times \varepsilon_{280}\,(protein)}{(A_{280} - A_{max} \times CF) \times \varepsilon_{max}}$$

After extensive desalting steps using Zeba spin desalting columns (repeated six to eight times to totally remove the free dye), we check the absorbance of Fab-Cy5 at 280 nm and 650 nm, the values are listed below:

$A_{280}$ (the absorbance of the sample at 280 nm) = 0.23

$A_{max}$ (the absorbance of the sample at 650 nm) = 0.56

$\varepsilon_{280}$ (the extinction coefficient of native Fab at 280 nm) = 70000 $M^{-1}cm^{-1}$

$\varepsilon_{max}$ (the extinction coefficient of Cy5 dye at 650 nm) = 250000 $M^{-1}cm^{-1}$

CF (correction factor) = 0.03

Based on the calculation, the DOL is 73%.

The other method to determine the degree of labelling is based on the gel densitometry. The absorbance of Fab-Cy5 at 650 nm is 0.56. The Cy5 concentration was 13.3 μg/mL (13.8 μM) calculated using the calibration curve (Figure 11C). Fab-Cy5 solution was diluted 10 times and then loaded to the SDS-PAGE. Based on the gel densitometry (FigureS11B), the concentration of Fab-Cy5 (1:10 dilution) is 0.094 mg/mL (1.96 μM). So the concentration of the original Fab-Cy5 (no dilution) is 0.94 mg/mL (19.6 μM). Therefore, the labelling efficiency was 70% (= 13.8/19.6×100%).
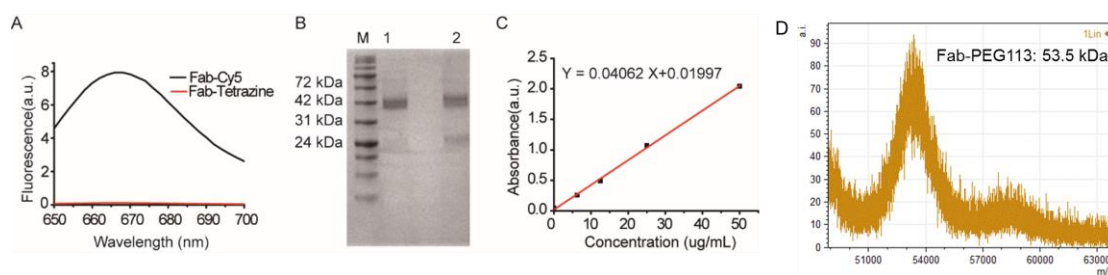


Figure S11 (A) Fluorescence spectra of Fab-Cy5 and Fab-Tetrazine (B) SDS-PAGE of IgG Fab and Fab-Cy5 (M: Protein marker, 1: Fab-Cy5 (based on the gel densitometry, the concentration of Fab-Cy5 is 0.094 mg/mL), 2: IgG Fab (0.1 mg/mL) (C) calibration curve of TCO-Cy5 based on the absorbance of 650 nm (D) MALDI-Tof-MS of Fab-PEG113 showing a signal at 53.5 kDa (found: 53.5k [M+H]$^+$, calculated: 53.5k [M+H]$^+$)

TCO-PEG113 (molecular weight: 5000 Da containing 113 repeating units on average) (25 μg, 5.00 nmol) was dissolved in PB (50 mM, pH = 7.4) and added to the Fab-Tetrazine (48 μg, 1.00

nmol) solution. The reaction mixture was stirred at room temperature for 30 min. After the reaction, the crude product was purified by repeated ultrafiltration in water using Vivaspin sample concentrator (10 kDa MWCO). The sample was analyzed by SDS-PAGE. Based on the gel, the band of Fab-Tetrazine totally disappeared and shifted to higher molecular weight (Fig. 3d in the main text). Because TCO-PEG113 is not monodisperse, the band of Fab-PEG113 also became broaden after conjugation. MALDI-Tof-MS of Fab-PEG113 also showed a signal at 53.55 kDa indicating the successful modification (Fig. S11D).

## 8. Stability test of SST-Tetrazine and Fab-Tetrazine

0.01 mM SST-Tetrazine solution was incubated with 20 μM Glutathione at 37 °C for 0 hours, 12 hours and 24 hours. For each sample, 10 μL of 10 mg/mL Fmoc-L-phenylalanine was added as internal standard. After incubation, the samples were injected into LC-MS. Identification of SST-Tetrazine product was performed simultaneously by UV-VIS detection at 254 nm and 214 nm and selective ion monitoring (SIM). The amount of SST-Tetrazine in each sample was determined as a ratio of the integration of the chromatogram at 254 nm of SST-Tetrazine to the internal standard.
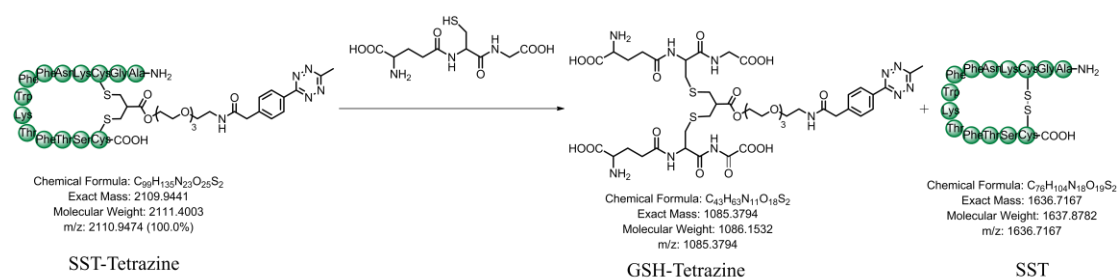


Scheme S9 Reaction of GSH-mediated cleavage of SST-Tetrazine
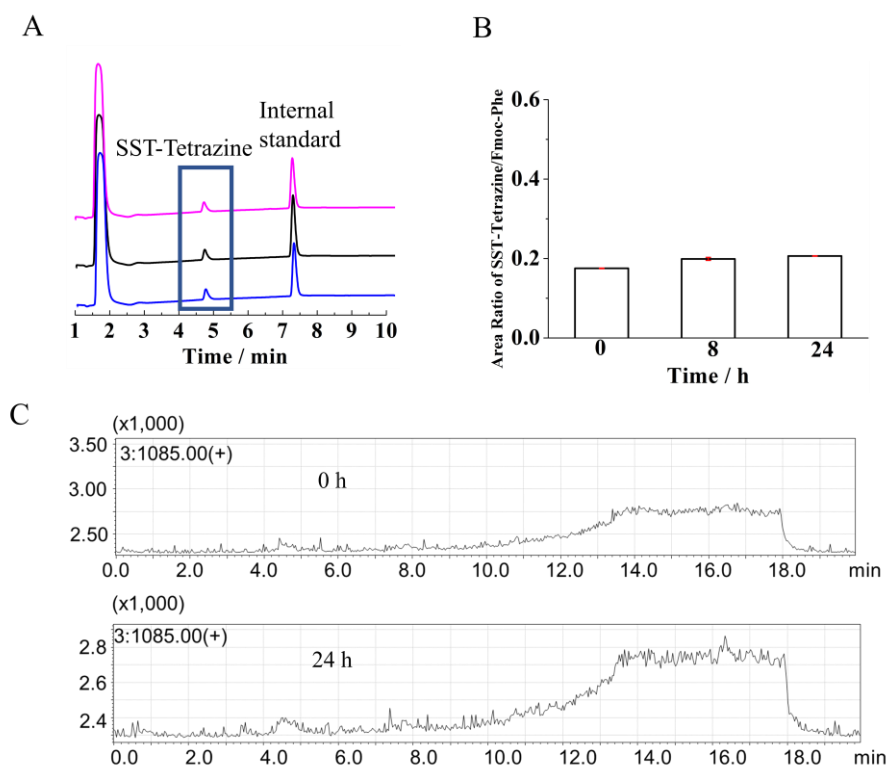
Figure S12 (A) LC-MS chromatogram of SST-Tetrazine incubating with GSH at 214 nm (the peak at 2 min is from DMSO). (B) The area ratio of SST-Tetrazine at 214 nm compared to internal standard. (C) selective ion monitoring profile of m/z 1085 (GSH-Tetrazine) at 0 h and 24 h showing no degradation products were detected.

As suggested in the reference[1], the stability of Fab-Tetrazine and Fab-Cy5 were investigated through incubation with fetal bovine serum (FBS) at 37 °C. 0.01 mM Fab-Tetrazine and Fab-Cy5 were incubated with 1% FBS for 0 hour, 12 hours and 24 hours at 37 °C. After incubation, the mixture was analyzed by SDS-PAGE as described in the protocol in the section above. From the SDS-PAGE data, the protein band of Fab-Tetrazine did not have obvious change showing the good stability of Fab-Tetrazine. The stability of Fab-Cy5 was also investigated based on quantifying the fluorescence of the conjugate. The intensity of the fluorescence band of lane 5–7 is 23226.886, 24341.886 and 23008.664 using Image J. The data indicated good stability of Fab-Cy5 after incubation with FBS for 24 hours (Fig. 5d)

## 9. CD experiment of SST-Tetrazine and Fab-Tetrazine

Native SST, SST-Tetrazine and SST-PEG12 were dissolved in PBS to prepare 0.1 mg/mL solution. CD measurements were performed on JASCO J-1500 spectrometer in a 1 mm High Precision Cell by Hellma Analytics from 260 to 190 nm with a bandwidth of 1 nm. The data pitch was set to 0.2 nm, while the scanning speed was 10 nm/min. Each sample was measured

three times and the signal from the buffer blank was subtracted from the sample scan. The data was processed in the software Spectra Analysis and CD Multivariate SSE by JASCO.

Native IgG Fab, Fab-Tetrazine and Fab-PEG113 conjugate were dissolved in PBS at a final concentration of 0.1 mg/mL. CD experiment was performed as SST and SST-Tetrazine using the same machine with the same procedure. Each sample was measured three times and the signal from the buffer blank was subtracted from the sample scan. The data was processed in the softwares Spectra Analysis and CD Multivariate SSE by JASCO.

## 10. Enzyme-Linked immunosorbent assay (ELISA) experiment[2]

Protein L coated 96-well plate was washed with 100 µL wash buffer (PBS with added 0.05% Tween-20 detergent). Then dilution buffer (2% BSA in washing buffer) was used to dilute the native IgG Fab and Fab-Tetrazine to the following concentration: 1 nM, 10 nM, 50 nM, 100 nM, 200 nM and 100 µL of dilution solution were added to the corresponding well incubating for 2 hours. After 2 hours, the dilution solution was removed and each well was washed three times with wash buffer. Next, 100 µL of anti-human IgG, Fab-specific-HRP solution (prepared by taking 4 µL of a 1:5000 diluted solution and further diluting with 20 mL of PBS) was added to each well and incubate for 1 hour. Next, the solution was removed and washed with wash buffer 3 times (200 µL). For each well, 100 µL enhanced chemiluminescence solution (ECL solution, from Thermo Fisher Scientific) was added to each well and chemiluminescence was measured after 5 minutes.

## 11. Reference

1. A. Maruani, M. E. Smith, E. Miranda, K. A. Chester, V. Chudasama and S. Caddick, *Nat. Commun.*, 2015, **6**, 6645.
2. M. K. Greene, D. A. Richards, J. C. F. Nogueira, K. Campbell, P. Smyth, M. Fernández, C. J. Scott and V. Chudasama, *Chem. Sci.*, 2018, **9**, 79-87.

¹H NMR of compound **3**
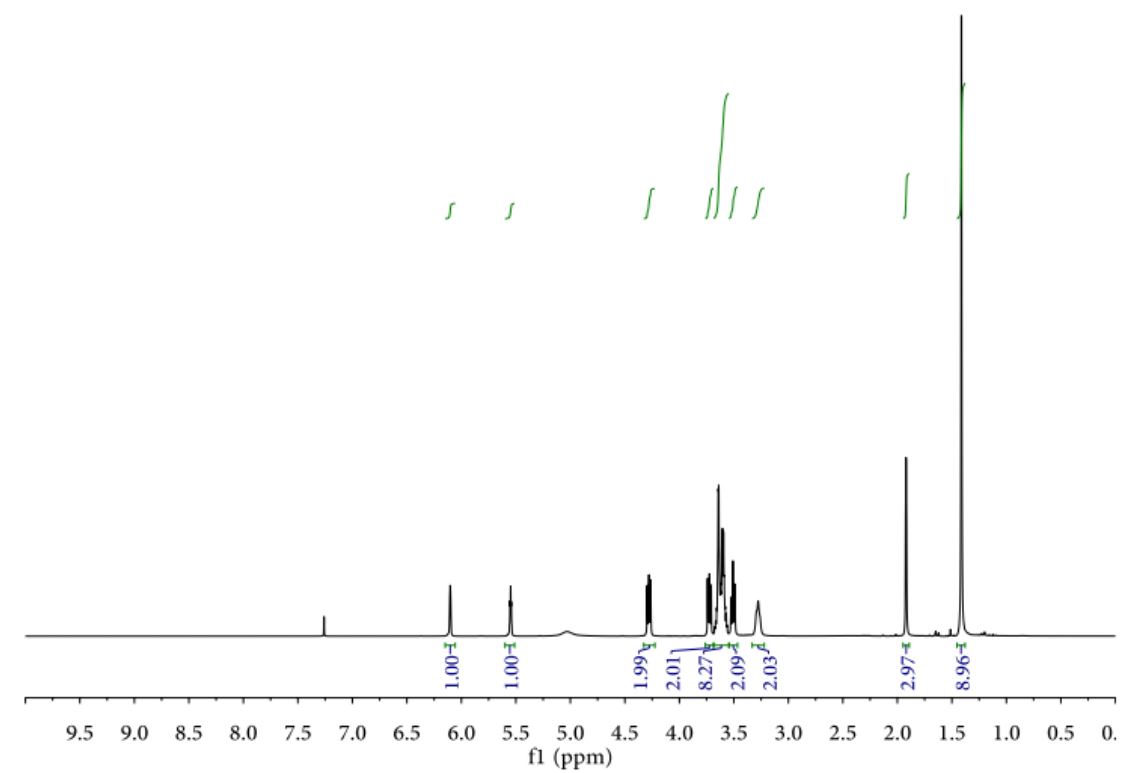


¹³C NMR of compound **3**
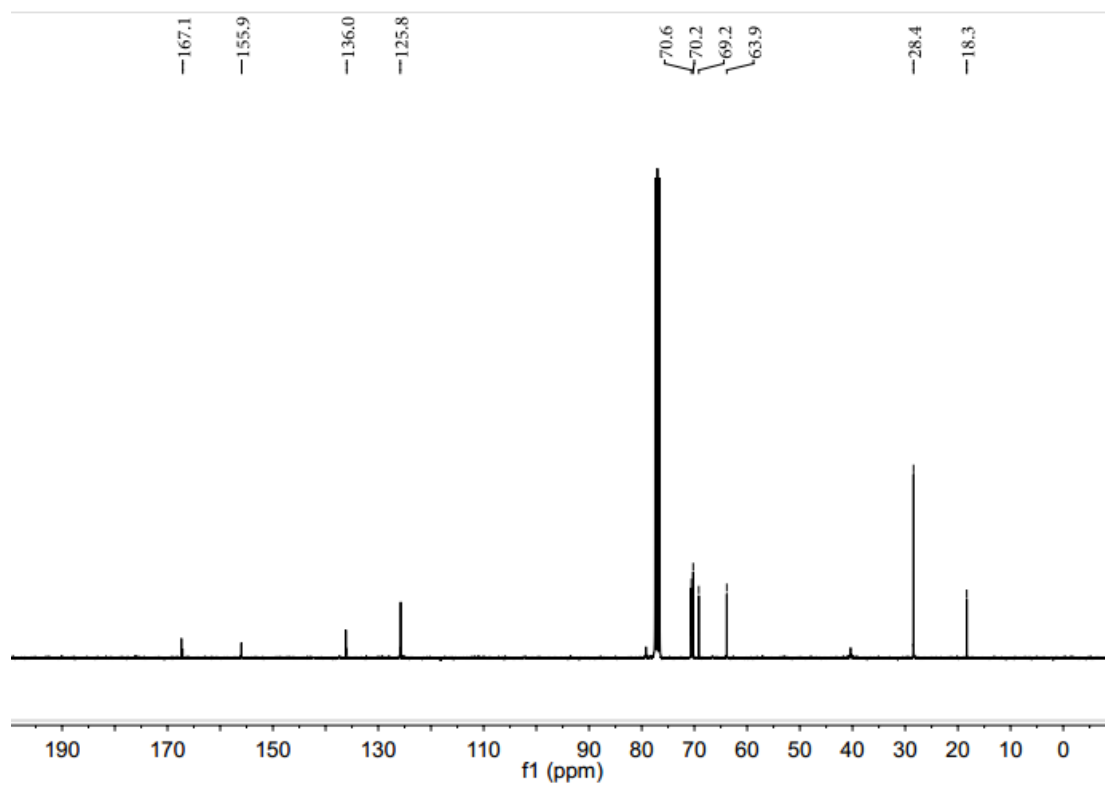
¹H NMR of compound **4**
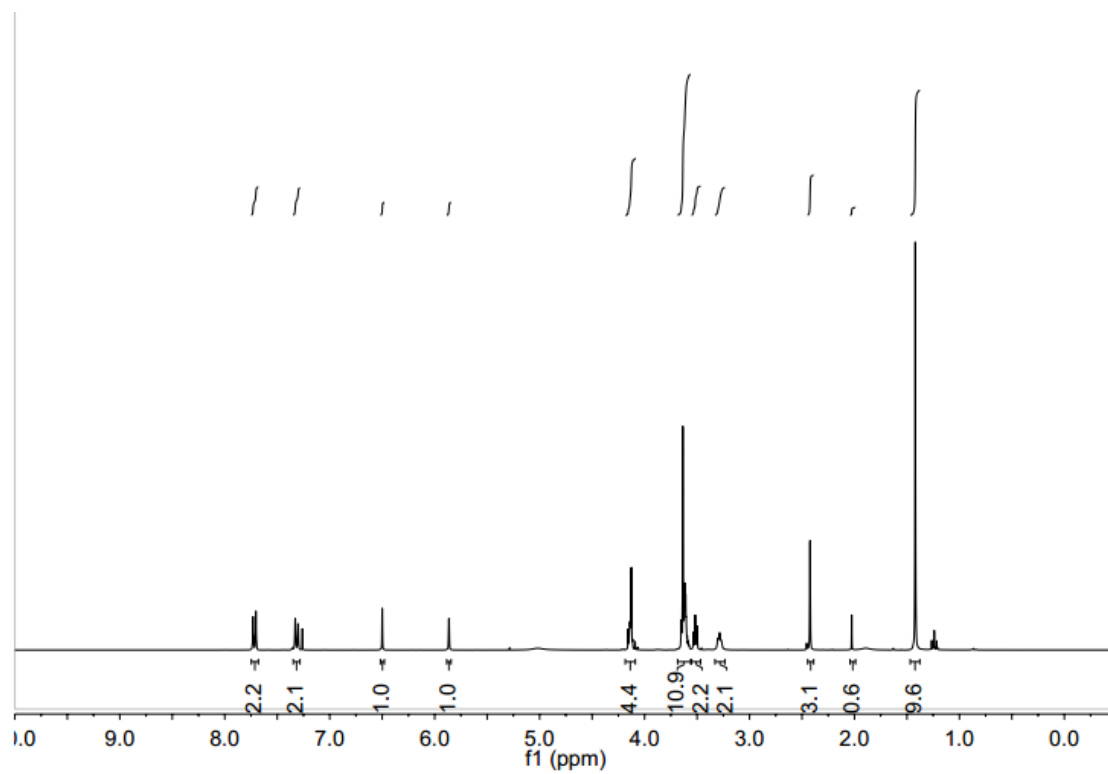


¹³C NMR of compound **4**
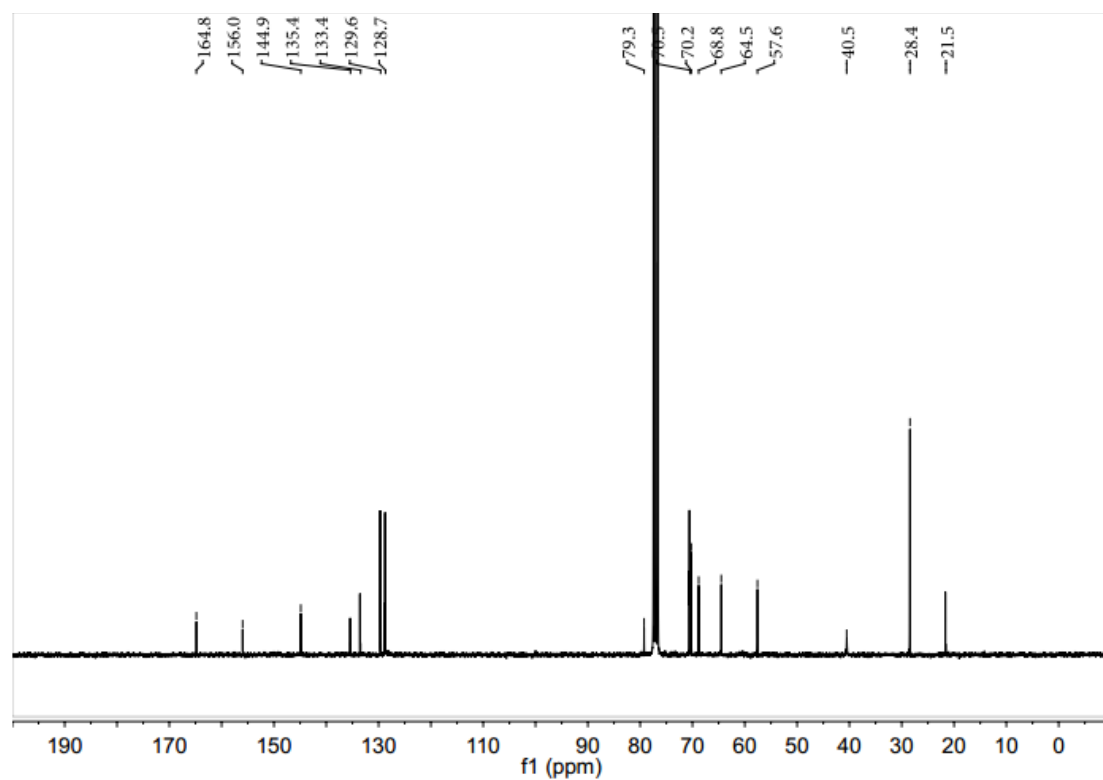
<sup>1</sup>H NMR of IC-Tetrazine
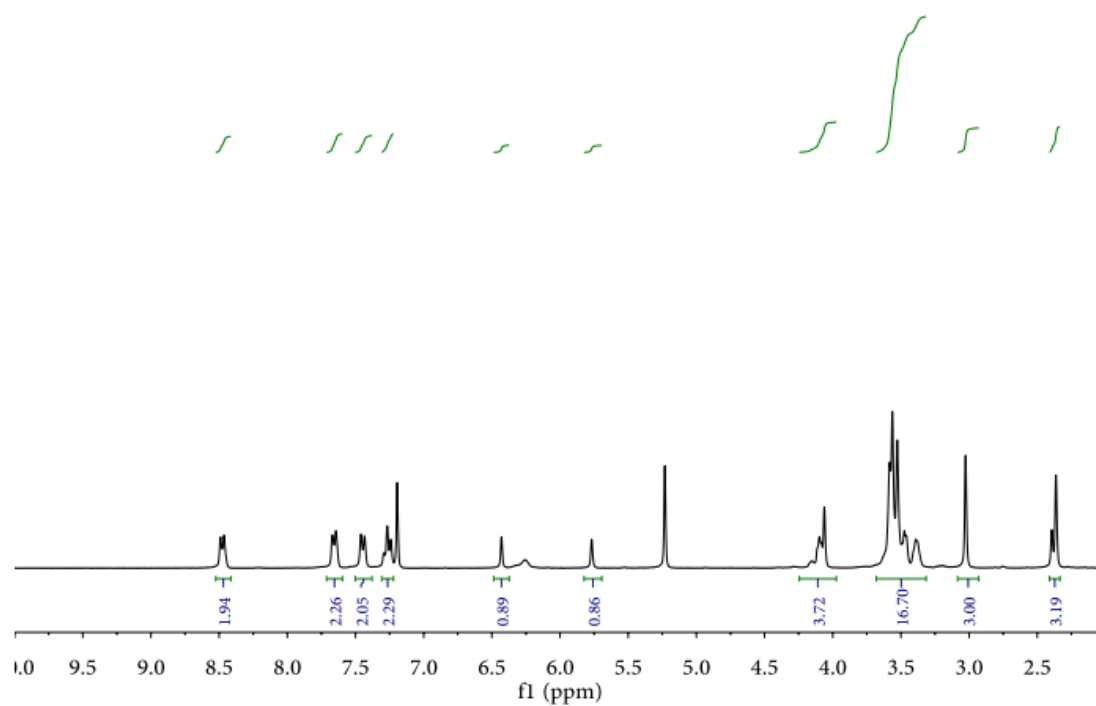


<sup>13</sup>C NMR of IC-Tetrazine
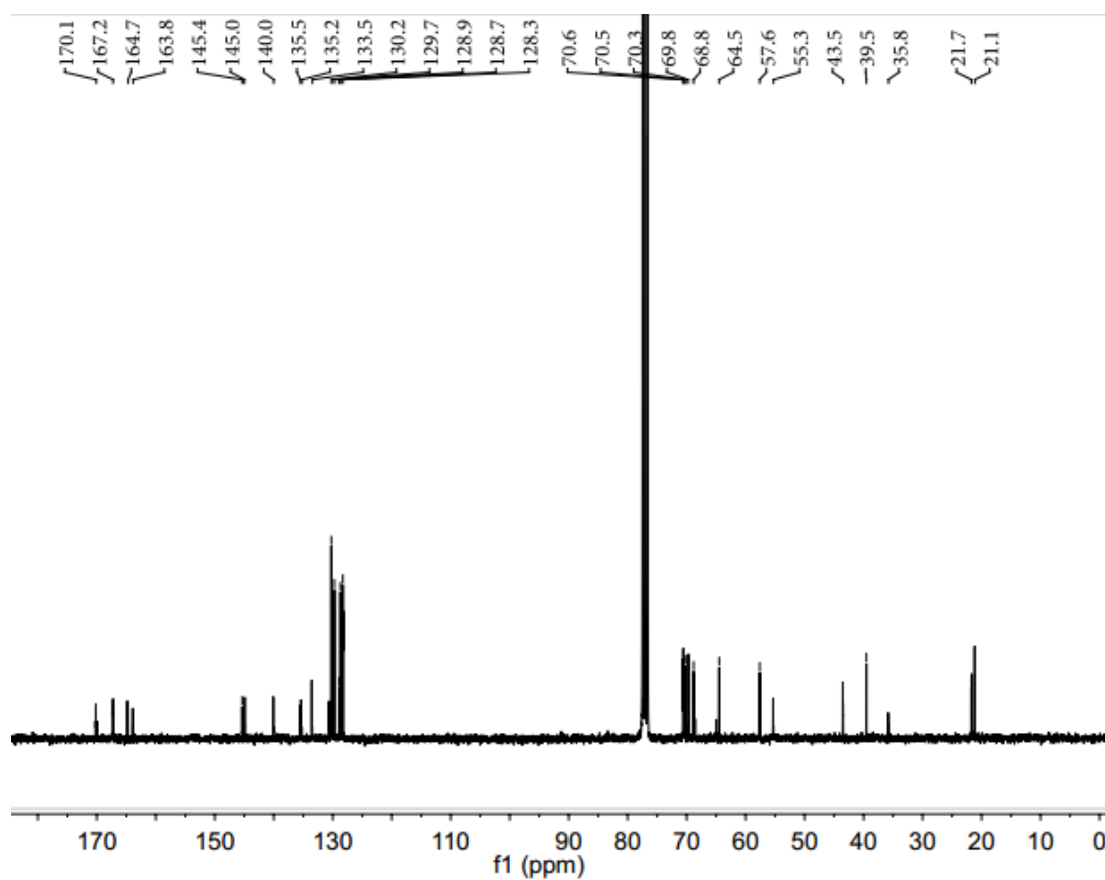
## 5.3 Chemoselective cysteine or disulfide modification via single atom substitution in chloromethyl acryl reagents

Lujuan Xu, Maria J. S. A. Silva, Pedro M. P. Gois, Seah Ling Kuan* and Tanja Weil*

\* Corresponding author

**Contribution of the respective authors:**

**Lujuan Xu:** Designed and performed the experiments, wrote and corrected the manuscript.

**Maria J. S. A. Silva and Pedro M. P. Gois**: Contributed with MS studies to elucidate the structure of several bioconjugates present in this manuscript.

**Seah Ling Kuan:** Supervised the project, corrected the manuscript.

**Tanja Weil:** Conceptualized and supervised the project. Wrote and corrected the manuscript.

## EDGE ARTICLE

Check for updates

# Chemoselective cysteine or disulfide modification *via* single atom substitution in chloromethyl acryl reagents†

Lujuan Xu,[ab] Maria J. S. A. Silva, [iD][c] Pedro M. P. Gois, [iD][c] Seah Ling Kuan [iD]*[ab]
and Tanja Weil [iD]*[ab]

The development of bioconjugation chemistry has enabled the combination of various synthetic functionalities to proteins, giving rise to new classes of protein conjugates with functions well beyond what Nature can provide. Despite the progress in bioconjugation chemistry, there are no reagents developed to date where the reactivity can be tuned in a user-defined fashion to address different amino acid residues in proteins. Here, we report that 2-chloromethyl acryl reagents can serve as a simple yet versatile platform for selective protein modification at cysteine or disulfide sites by tuning their inherent electronic properties through the amide or ester linkage. Specifically, the 2-chloromethyl derivatives (acrylamide or acrylate) can be obtained *via* a simple and easily implemented one-pot reaction based on the coupling reaction between commercially available starting materials with different end-group functionalities (amino group or hydroxyl group). 2-Chloromethyl acrylamide reagents with an amide linkage favor selective modification at the cysteine site with fast reaction kinetics and near quantitative conversations. In contrast, 2-chloromethyl acrylate reagents bearing an ester linkage can undergo two successive Michael reactions, allowing the selective modification of disulfides bonds with high labeling efficiency and good conjugate stability.

## 1. Introduction

Proteins are an emerging class of biotherapeutics with high target affinity and specificity.[1–3] Site-selective modification of proteins enables the incorporation of desired synthetic functionalities into proteins at distinct sites, which combine the advantages from both the synthetic world and Nature for the construction of protein bioconjugates with novel functional characteristics.[4–10] Chemical approaches for protein modification allow the straightforward attachment of the desired functionalities at natural amino acid residues on the protein surface, thereby eliminating the need for tedious genetic engineering.[11] Among these, unpaired cysteines are considered the most sought-after targets owing to the high nucleophilicity and versatile chemistry landscapes of thiol groups.[12–15] In addition, disulfide bonds have also emerged as attractive modification sites to incorporate tailored functionalities, as a lot of

therapeutic relevant proteins or peptides, *e.g.* antibodies or their antigen-binding fragments, contain at least one solvent-accessible disulfide bond.[16,17]

Maleimides constitute a group of widely-used cysteine bioconjugation reagents due to their fast and efficient reactions with thiols.[12] Besides that, a variety of structurally diverse reagents have also been reported for cysteine modification in order to improve the stability of the resultant bioconjugates as well as retaining similar reaction kinetics.[18,19] However, the strategies for disulfide modification are much less explored and the current toolset is limited to five to six conjugation methods available in the literature.[20–26] Moreover, the reagents developed to date mainly target a single amino acid residue, for example a cysteine residue or a disulfide bond. Besides the (bromo) maleimides,[23] 3-bromo-5-methylene pyrrolones[27] and diethynyl phosphinates,[28] there are only a few bioconjugation reagents that can provide a broad spectrum scaffold to address both cysteines and disulfides with high labeling efficiency. Such a strategy is more advantageous compared to reinventing a novel scaffold for every single purpose. Therefore, the development of such a bioconjugation approach, which enables the selective modification at target amino acid residues in a user-defined fashion with great ease, would be highly advantageous to enrich the existing toolbox and also to enable non-experts to conduct such protein labeling reactions.

*[a]Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany. E-mail: weil@mpip-mainz.mpg.de*

*[b]Institute of Inorganic Chemistry I, Ulm University, Albert-Einstein-Allee 11, 89081, Ulm, Germany*

*[c]Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, 1649-003 Lisbon, Portugal*

This prompts us to rethink the strategies and enormous possibilities offered by synthetic chemistry. In fact, modern synthetic technologies provide immense flexibility and potential to access structurally diverse reagents, which allows for the customization of their reactivities at the atomic level. From this perspective, we envisioned that multifunctional bioconjugation reagents, which are capable of targeting the specific residues on demand, can be designed by finely tuning their chemoselectivities with the aid of synthetic chemistry. Inspired by the inherent features of the electron-deficient systems serving as good Michael acceptors for the reactions with nucleophiles on the protein surface,[12] we proposed 2-halomethyl acryl derivatives (acrylamide or acrylate) as an appropriate option for reactions with thiol groups to accomplish the chemoselective modification of cysteine residues. In addition, considering the different electron-withdrawing properties of the ester and amide bond, we further speculated that a single atom substitution in the acryl position of chloromethyl acryl reagents would influence their reactivity profiles as electrophiles for the second Michael reaction. This, in turn, will allow the customization of their properties to achieve selective modification at either cysteine or disulfide sites.

Herein, we reported the convenient, one-pot synthesis of 2-chloromethyl derivatives (acrylamide and acrylate) via coupling reactions between commercially available 2-(bromomethyl) acrylic acid with different end-group functionalities (amino group or hydroxyl group) (Fig. 1). The inherent chemoselectivity of 2-chloromethyl acrylamide and acrylate are influenced by the different electron-withdrawing properties of the amide and ester linkage, which render them suitable for protein modification at either cysteine or disulfide site. Specifically, we showed that 2-chloromethyl acrylamide compounds containing an amide bond in the scaffold can react with proteins containing a free thiol group via a single Michael reaction with near quantitative conversions. By replacing the amide with an ester

linkage yielding the respective 2-chloromethyl acrylate reagents, site-selective disulfide modification can be achieved as exemplified by successful modification of three disulfide-containing substrates. In addition, the bioconjugation reagents reported herein are characterized by facile linker synthesis, high water solubility as well as good labeling efficiency.

## 2. Results and discussion

### Synthesis of 2-chloromethyl acrylamide and acrylate compounds

We initiated our study by using the commercially available compound, 2-(bromomethyl)acrylic acid, as starting material to synthesize both 2-halomethyl acrylamide and acrylate bioconjugation reagents (Scheme 1 and S1†). First, 2-(bromomethyl)acrylic acid reacted with oxalyl chloride to convert the carboxylic acid group to acid chloride in situ. Thereafter, different end-group nucleophiles, amino or alcohol groups (usually 1.5 to 2 equiv.) were added under basic conditions for further reactions. The respective 2-chloromethyl acrylamide and acrylate were subsequently purified and isolated in moderate yields (Scheme 1). Mass spectrometry (MS) data demonstrated that the bromine atom is completely replaced by the chlorine atom affording the 2-chloromethyl acryl compounds (Fig. S69–S74†). A toolbox containing different functionalities, e.g. dye or bioorthogonal groups, was obtained as demonstrated in Scheme 1 underlining the broad applicability of this method. Compared to other disulfide- and cysteine-modification reagents, which require multiple-step synthesis (e.g., allyl sulfones require four-step synthesis[20]), the 2-chloromethyl acryl derivatives are readily available through a straightforward one-pot synthesis from commercially available 2-(bromomethyl)acrylic acid precursors. The simplicity of the synthesis provides fast and efficient access to a broad spectrum of functionalities that are of great interest for bioconjugation.
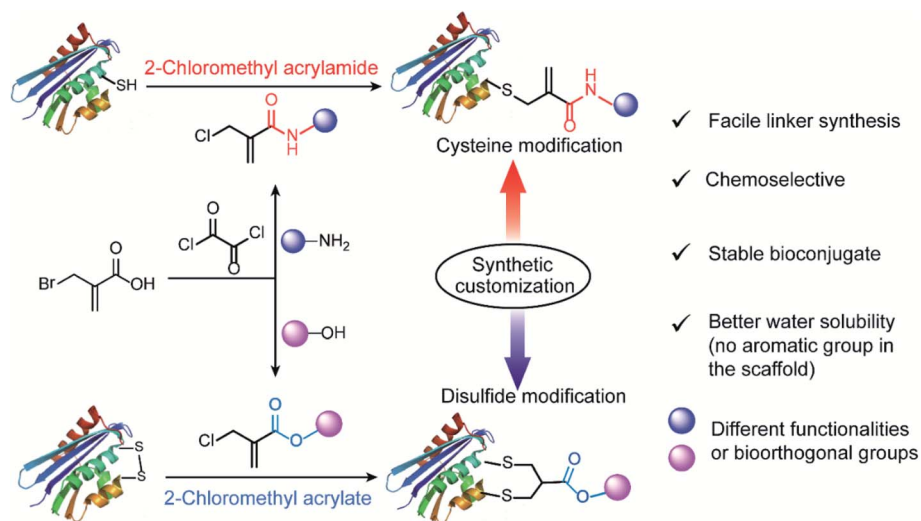


Fig. 1 General scheme for 2-chloromethyl acrylamide and acrylate compounds for site-selective protein modification at cysteine or disulfide sites via synthetic customization.
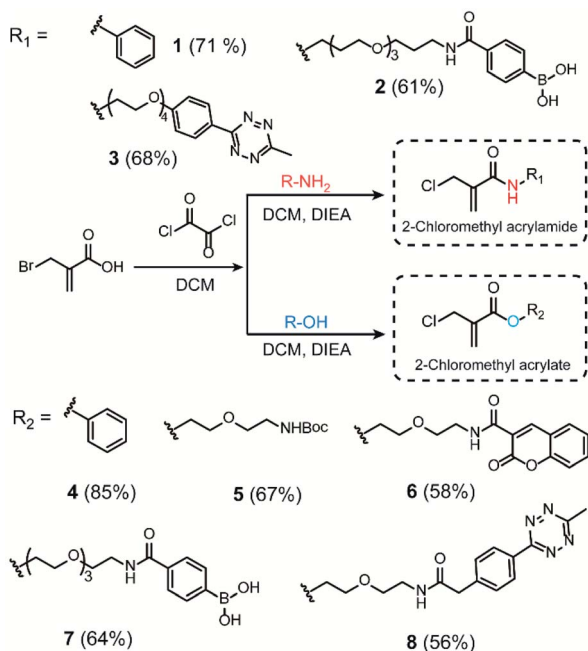
**Scheme 1** Synthesis route of 2-chloromethyl acrylamide and acrylate derivatives containing different functionalities.

In addition, compared to the reported cysteine and disulfide modification reagents, *e.g.* carbonylacrylic reagent[29] or allyl sulfone reagents[20] that contain a hydrophobic phenyl group, the 2-chloromethyl acryl derivatives do not contain any aromatic group in the scaffold, where they are estimated to have lower partition coefficients (*n*-octanol to water, log $P_{o/w}$) (Fig. S2 and S3†) indicating improved water solubility. Furthermore, the stability of the bioconjugation reagents in different aqueous environment represents an important consideration for their subsequent usage. The stability of the 2-chloromethyl acrylamide and acrylate compounds was evaluated by incubating compound **3** and compound **4** at three different pH (pH 6, 7 and 8), and the HPLC data indicated that they remained stable over a time course of 36 hours without any degradation (Fig. S4–S9†). In contrast, maleimides reagents, which are the most commonly used bioconjugation reagents for cysteine functionalization, easily hydrolyze to nonreactive maleic amides, especially at basic pH ($t_{1/2}$ < two hours) (Fig. S10–S12†).

## Chemoselectivity of 2-chloromethyl acrylamide and acrylate towards thiol groups

The reactivity and selectivity of 2-chloromethyl acrylamides and acrylates towards two model amino acids: Boc-Cys-OMe and Boc-Lys-OH (Fig. 2) were evaluated. For 2-chloromethyl acrylamide, compound **1** was incubated with both Boc-Cys-OMe and Boc-Lys-OH in acetonitrile (ACN)/phosphate buffer (PB, pH 7) for four hours (Fig. 2a). Liquid chromatography (LC) data indicated quantitative conversion to the cysteine-modified compound **10** while lysine-modified compound **9** was not observed, which clearly demonstrated its excellent
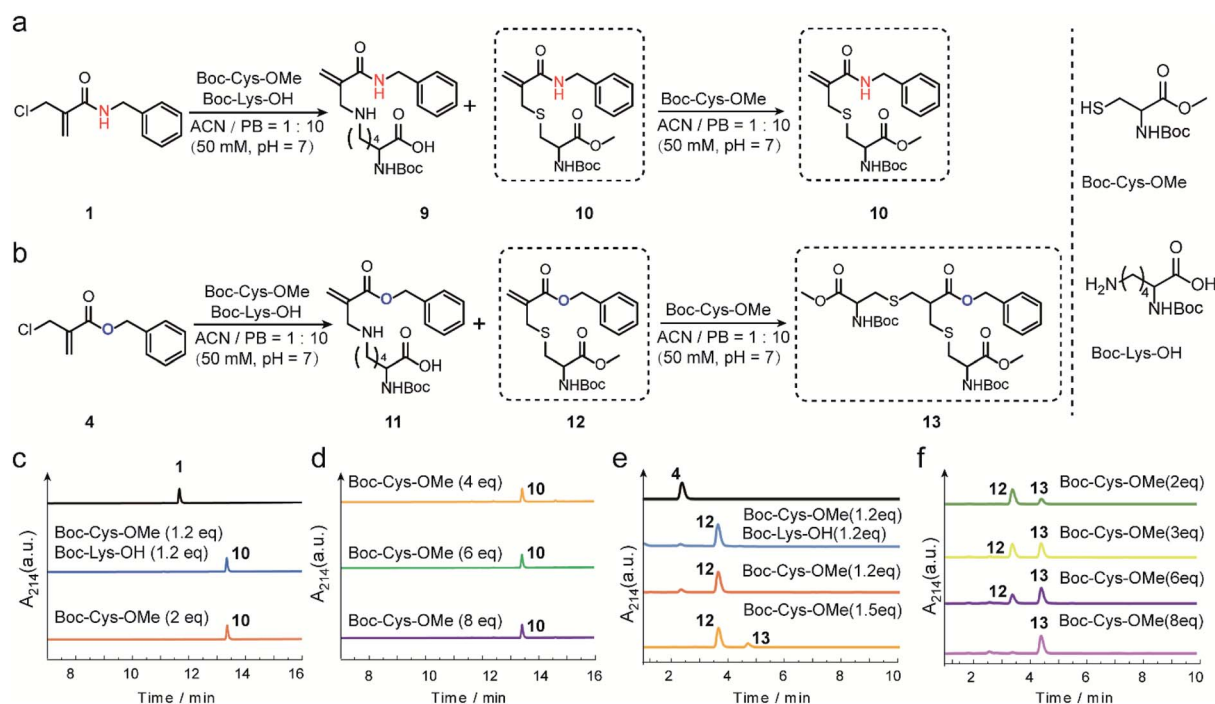


**Fig. 2** (a) Reaction scheme between 2-chloromethyl acrylamide and Boc-Cys-OMe (and Boc-Lys-OH). (b) Reaction scheme between 2-chloromethyl acrylate and Boc-Cys-OMe (and Boc-Lys-OH). (c) LC trace of the reaction between 2-chloromethyl acrylamide and Boc-Cys-OMe (and Boc-Ly-OH). (d) LC trace of 2-chloromethyl acrylamide with Boc-Cys-OMe (from 4 equiv. to 8 equiv.). (e) LC trace of the reaction between 2-chloromethyl acrylate and Boc-Cys-OMe (and Boc-Lys-OH) (f) LC trace of 2-chloromethyl acrylate and Boc-Cys-OMe (from 2 equiv. to 8 equiv.).

112

chemoselectivity towards cysteine over lysine residues (Fig. 2c). In addition, the absence of the side reaction with benzylamine during the synthesis of compound 1 also indicated there were no cross-reactions with amino groups (page 5 in ESI†). With increasing Boc-Cys-OMe to 8 equiv., only compound **10** was found in the LC without the observation of the further addition products (Fig. 2d). Further studies demonstrated that compound **10** does not react with other nucleophiles, such as hydroxyl or amino groups, even when used in 30 equiv. excess at 37 °C (Scheme S3 in ESI†). However, if a second thiol functionality was given in very large excess (for example 30 equiv. of compound **18**, Scheme S3 in ESI†), the first thiol functionality was eliminated affording compound **15**, presumably due to an addition–elimination reaction (Fig. S14 in ESI†).

For 2-chloromethyl acrylate, compound **4** was incubated with both Boc-Cys-OMe and Boc-Lys-OH under the same conditions used for compound **1** (Fig. 2b). The LC trace also revealed the excellent chemoselectivity towards thiol groups as lysine-modified compound **11** was also not observed in the mixture (Fig. 2e). However, in contrast to the reaction with acrylamides, the peak for compound **12** decreased while the signal for compound **13** increased (Fig. 2f), with increasing amounts of Boc-Cys-OMe used. After adding eight equivalents of Boc-Cys-OMe, compound **4** was fully converted to compound **13** with negligible side product formation (Fig. 2f).

These model reactions clearly indicated a pronounced difference in the reactivity of the 2-chloromethyl acrylamide *versus* the acrylate reagents, presumably originating from the amide or ester linkage. The observed reactivity of 2-chloromethyl acrylamide is consistent with literature where it was reported that catalysts and high temperature are required for thiol addition with α,β-unsaturated amides as Michael acceptors.[30,31] Therefore, we speculate that the second Michael reaction of the 2-chloromethyl acrylamide did not proceed due to the relatively weak electron-withdrawing property of the amide bond, which rendered the α,β-unsaturated amide a poor Michael acceptor.[27] Taken together, these results demonstrated that 2-chloromethyl acrylamides allow straightforward modification of free cysteines with high efficiency and excellent chemoselectivity. On the other hand, 2-chloromethyl acrylates can undergo two Michael reactions in a successive manner, thereby making them suitable candidates to achieve protein modification at the disulfide sites.

### 2-Chloromethyl acrylamide reagents for cysteine modification

Next, the reaction kinetics were first studied using a model reaction between compound **3** and Boc-Cys-OMe (Fig. 3a). Compound **3** (1 mM) and Boc-Cys-OMe (1 mM) were incubated in ACN/PB (pH 7) mixture (volume ratio: 1 : 10) using Fmoc-Phe-OH (Scheme S4†) as internal standard. At different time intervals, the reaction was monitored by high-performance liquid chromatography (HPLC) (Fig. S17†), and quantification of compounds **3** and **14** overtime was plotted with reference to the internal standard (Fig. 3b). HPLC analysis indicated that 80% of compound **3** was converted to the cysteine-modified compound **14** in less than two hours and near quantitative conversion was
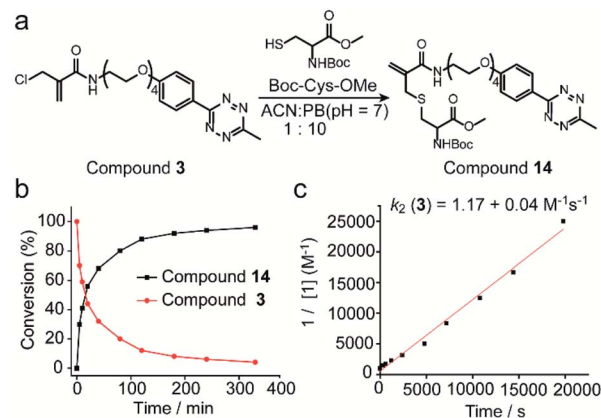


Fig. 3 (a) Reaction scheme of compound **3** reacting with Boc-Cys-OMe to form compound **14**. (b) Percentage of compound **3** and **14** as determined by the integration of the HPLC peak in comparison to the internal standard at different time points. (c) Experimental determination of the second-order rate constant of the model reaction between compound **3** (1 mM) and Boc-Cys-OMe (1 mM). Details about the calculation and kinetics data were demonstrated in Section 4 in ESI.†

achieved in less than six hours (Fig. S17†). The second-order rate constant was determined to be 1.17 $M^{-1}$ $s^{-1}$ with the concentration of compound **3** at 1 mM (Fig. 3c). Although this reaction is slower than the maleimide conjugation (10–1000 $M^{-1}$ $s^{-1}$),[32] it is still comparable to or even faster than some of the recently reported bioconjugation reagents, *e.g.* ethynyl-phosphonamidates (0.62 $M^{-1}$ $s^{-1}$),[33] diethynyl phosphinate (0.47 $M^{-1}$ $s^{-1}$) (Table S1 in ESI†), and some other conventional bioconjugation methods such as oxime ligation (0.001 $M^{-1}$ $s^{-1}$),[34,35] Pictet–Spengler ligation (0.015 $M^{-1}$ $s^{-1}$)[32] and strain-promoted azide–alkyne reaction (0.9 $M^{-1}$ $s^{-1}$).[36]

Thereafter, 2-chloromethyl acrylamide derivatives were applied for cysteine modification on peptide substrates using compound **3** (Fig. 4a). First, the known WSCO2 peptide (sequence: IVRWSKKVCQVS), an endogenous peptide inhibitor of the chemokine CXCR4 receptor that is highly relevant for anti-infectivity in viral infection and anti-migratory effect in cancer,[37] was selected as bioactive substrate (Fig. 4b). In ACN/PB mixture (1 : 10), one equivalent WSCO2 peptide was incubated with 1.1 equivalents of compound **3** for four hours. HPLC analysis of the crude reaction mixture indicated that more than 95% conversion to the desired modified product (WSCO2-PEG4-Tz) was achieved (Fig. 4c). As a control, the thiol-reactive reagent 4,4′-dithiodipyridine (4-DPS), which is often used for free thiol quantification on proteins *via* a thiol–disulfide exchange reaction (the reaction mechanism is shown in Scheme S6†),[38] was used to mask the cysteine residue. In this case, no further reaction was observed in the HPLC chromatogram in the presence of compound **3** under the same reaction conditions (Fig. 4c). Taken together, these data clearly indicated that the 2-chloromethyl acrylamide compounds exhibit excellent chemoselectivity in combination with excellent modification efficiency. In addition to WSCO2, five other peptides, including RGDC, CEIE, PC-8, Tet, and EK-1 peptides (sequences and MS of

113
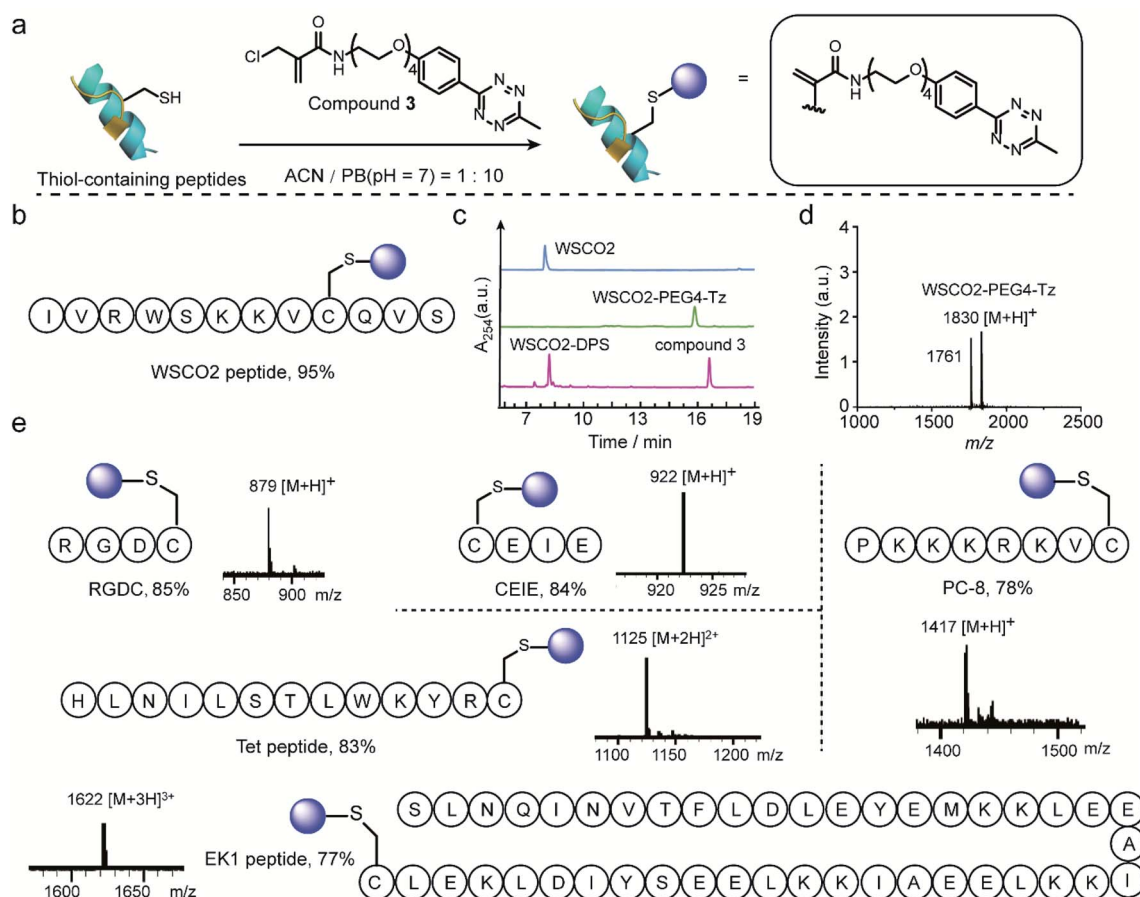
Fig. 4   (a) General scheme for the chemoselective modification of thiol-containing peptides with compound **3** in ACN/PB (pH 7) mixture. (b) WSCO2 peptide was selected as model substrate for modification at cysteine site with compound **3**. (c) HPLC trace of WSCO2 peptide, crude reaction mixture between WSCO2 and compound **3**, WSCO2-DPS and compound **3** (from top to down) which demonstrated the efficient modification efficiency and good chemoselectivity. (d) MALDI-Tof-MS of modified WSCO2 peptide showing a signal at 1830 which is attributed to the WSCO2-PEG4-Tz (calculated: 1830 [M + H]$^+$, found: 1830 [M + H]$^+$). The signal at 1761 corresponded to the fragmentation product with the double bond breaking at tetrazine moiety, the chemical structure is shown in ESI.† (e) Site-selective modification of different thiol-containing peptides including RGDC (calculated: 879 [M + H]$^+$, found: 879 [M + H]$^+$), CEIE (calculated: 922 [M + H]$^+$, found: 922 [M + H]$^+$), PC-8 (calculated: 1417 [M + H]$^+$, found: 1417 [M + H]$^+$), Tet peptide (calculated: 1125 [M + 2H]$^{2+}$, found: 1125 [M + 2H]$^{2+}$) and EK1 peptide (calculated: 1622 [M + 3H]$^{3+}$, found: 1622 [M + 3H]$^{3+}$) with a tetrazine group.

the modified peptides were shown in Fig. 4e and S24–S28†), have also been successfully modified with compound **3**. The broad range of substrates used here clearly demonstrates the general applicability of 2-chloromethyl acrylamide compounds for chemoselective modification at cysteine residues.

After demonstrating the successful modification of the model peptides, we proceeded to functionalize the more complex substrates, *i.e.* proteins. The protein ubiquitin that plays an important role in protein degradation by the proteasome, which contains a cysteine mutation at its K63 position, was selected (Fig. 5a). After incubation of one equivalent ubiquitin with ten equivalents of two different 2-chloromethyl acrylamide derivatives respectively, the desired bioconjugates were obtained. The successful modification was confirmed with the expected *m/z* in the MS shown in Fig. 5b. Similarly, if 4-DPS was used to mask the accessible cysteine residue on the protein surface, no reaction was observed even in the presence of ten

equivalents of the 2-chloromethyl acrylamides (Fig. S29†). Besides ubiquitin, a single-chain V$_H$H antibody domain with specific binding activity against the green fluorescent protein (anti-GFP nanobody) has also been successfully modified with 2-chloromethyl acrylamide derivatives (Fig. 5c). The MALDI-Tof-MS characterization clearly indicated the successful modification with the expected *m/z* shown in Fig. 5d.

## 2-Chloromethyl acrylate reagents for disulfide modification

Next, the feasibility of 2-chloromethyl acrylate for disulfide bond modification was evaluated on both peptide and protein substrates. The cyclic peptide hormone somatostatin (SST), which plays a key role in regulating the endocrine system and contains an accessible disulfide bond in its sequence,[39] was selected (Fig. 6a). The disulfide bond in SST was first reduced by two equivalents of tris(2-carboxyethyl)phosphine (TCEP) to generate the two free
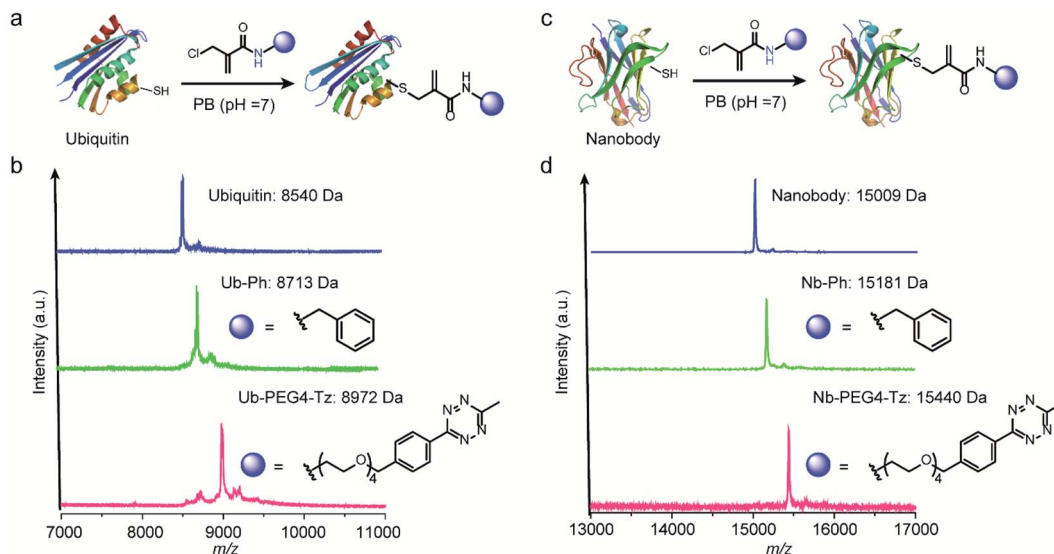
**Fig. 5** (a) Chemoselective modification of ubiquitin at cysteine site with two different functionalities: a phenyl and a tetrazine group. (b) The successful modification was proved by the MALDI-Tof-MS characterization with a peak at 8713 for Ub-Ph (calculated: 8714 Da, found: 8713 Da) and 8972 for Ub-PEG4-Tz (calculated: 8970 Da, found: 8972 Da). (c) Chemoselective modification of anti-GFP nanobody at cysteine site with two different functionalities: a phenyl and a tetrazine group (d) MALDI-Tof-MS of the modified nanobody with a peak at 15 181 for Nb-Ph (calculated: 15 183 Da, found: 15 181 Da) and 15 440 for Nb-PEG4-Tz (calculated: 15 438 Da, found: 15 440 Da).

thiol groups in ACN/PB mixture (1 : 10) at pH 7, followed by incubation with 1.1 equivalents of compound **4** for overnight in one-pot. HPLC of the crude reaction mixture revealed good

modification efficiency (91% based on HPLC quantification) (Fig. 6b). The isolated SST-Ph conjugate was also characterized by MALDI-Tof-MS showing successful functionalization
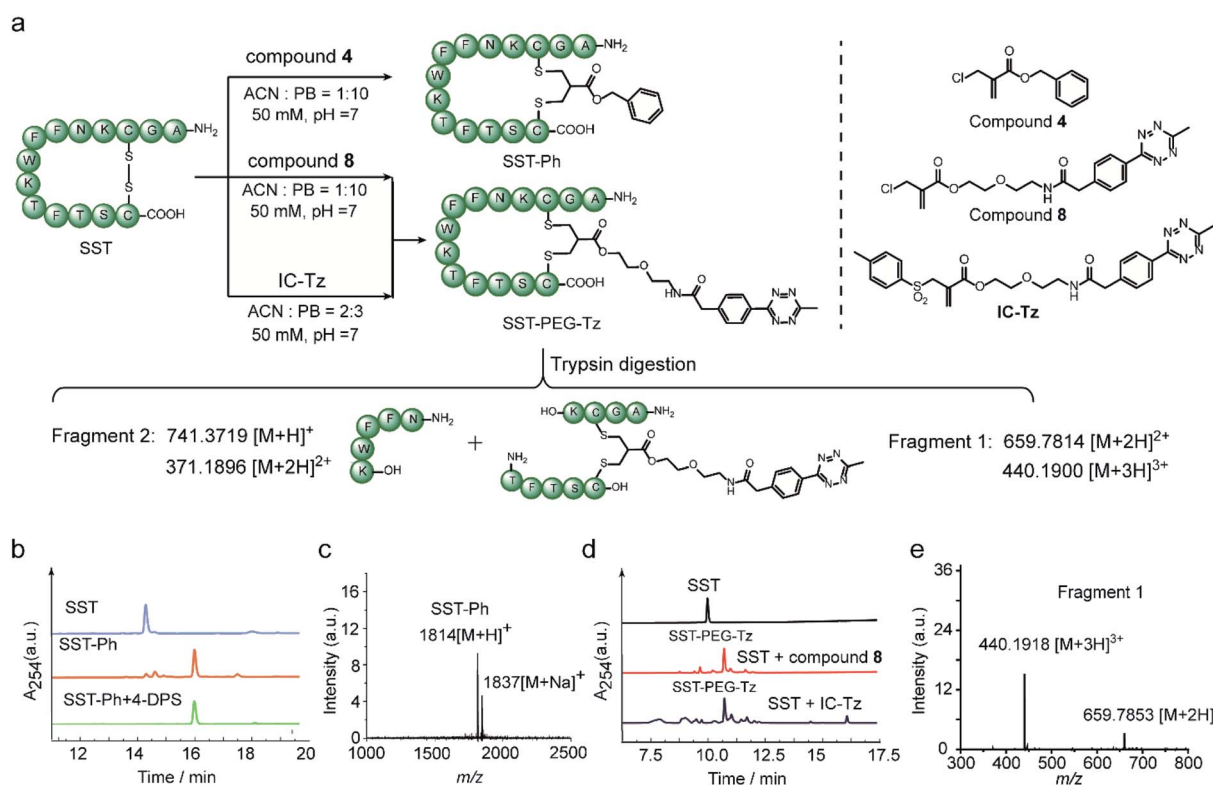


**Fig. 6** (a) Modification of SST with compound **4**, compound **8**, and an allyl sulfone reagent (IC-Tz). (b) HPLC trace of the crude reaction mixture between SST and compound **4** as well as the mixture of SST-Ph and 4-DPS; (c) MALDI-Tof-MS of the modified SST (SST-Ph) (calculated: 1814 [M + H]$^+$, found: 1814 [M + H]$^+$, 1837 [M + Na]$^+$); (d) HPLC trace of the crude reaction mixture between SST and compound **8** (or IC-Tz); (e) ESI-MS of fragment 1 after trypsin digestion (calculated: 659.7814 [M + 2H]$^{2+}$, 440.1900 [M + 3H]$^{3+}$, found: 659.7853 [M + 2H]$^{2+}$, 440.1918 [M + 3H]$^{3+}$).

(Fig. 6c). SST-Ph was further incubated with TCEP before the subsequent addition of the thiol-reactive reagent, 4,4′-dithiodipyridine (4-DPS). HPLC analysis showed that SST-Ph remained intact without any observation of side reactions occurring with 4-DPS. Native SST was used as a control and the LC showed that a new peak was formed when 4-DPS and TCEP were used (Fig. S32†). These results taken together indicated the complete modification at the disulfide site (Fig. 6b). Since the 2-chloromethyl acrylate compounds do not contain aromatic groups in their scaffold, they have a rather low log $P_{o/w}$ and thus provide better water-solubility than the reagents that contain phenyl groups, such as allyl sulfone reagents. This is particularly advantageous when modifying some therapeutic relevant proteins, which will suffer from aggregation issues if a large amount of organic solvent is needed during the modification process. Therefore, the disulfide modification efficiency of SST with 2-chloromethyl acrylate and allyl sulfone reagents was evaluated and compared with using compound 8 and an allyl sulfone reagent (denoted as "IC-tetrazine", which was developed by our group before[22]) (Fig. 6a). Disulfide modification of SST with IC-tetrazine required 40% ACN for solubilization, whereas less than 10% of ACN was needed to dissolve compound 8. More importantly, the modification efficiency of compound 8 was considerably higher (83% based on the quantification of HPLC peak for the reaction mixture) compared to IC-tetrazine (67%) (Fig. 6e). For definitive confirmation of the modification site, SST-PEG-Tz was selected for LC-MS/MS analysis. After trypsin digestion, a fragment with $m/z$ at 659.7814 $[M + 2H]^{2+}$ was observed corresponding to fragment 1 (Fig. 6e). The expected modified

sequence was detected by MS/MS, thus confirming that the modification occurred at the disulfide site (Tables S2–S5 and Fig. S40–S44†). Besides SST, another therapeutic relevant cyclic peptide octreotide, an analog of somatostatin with a longer biological half-life that is often applied in cancer diagnostics,[40] was also successfully functionalized with a coumarin motif under the similar reaction conditions mentioned above (Scheme S11†). The MALDI-Tof-MS data confirmed the successful functionalization with a peak at 1364 corresponding to $[M + H]^+$ (Fig. S46†).

Subsequently, this new disulfide modification strategy was also evaluated on a more complex substrate, the protein enzyme lysozyme (from hen egg white), in which the disulfide at C6–C127 is predicted to be solvent-accessible among the four available disulfide bonds.[20,41] To test the applicability of the 2-chloromethyl acrylate compounds for disulfide modification, different functionalities were incorporated into lysozyme, such as a phenyl group, a fluorescent dye (coumarin), or a bio-orthogonal tag (tetrazine group) (Fig. 7a). After adding 1.2 equivalents of TCEP, the 2-chloromethyl acrylate derivatives were also added in one pot, and the reaction mixture was incubated at 50 mM PB (pH 7) overnight. Some precipitates were observed after incubation overnight, presumably due to the aggregation of reduced lysozyme despite the mild conditions employed.[42] Thereafter, the modified lysozyme derivatives (Ly-Ph, Ly-PEG-Cou, and Ly-PEG-Tz) were purified by using Hi Trap hydrophobic interaction column with the isolated yields of 28%, 22%, and 24%, respectively. MALDI-Tof-MS data of the three modified lysozymes derivatives confirmed their successful functionalization (Fig. 7b). The yields are higher than our previous report where lysozyme was modified with allyl sulfone
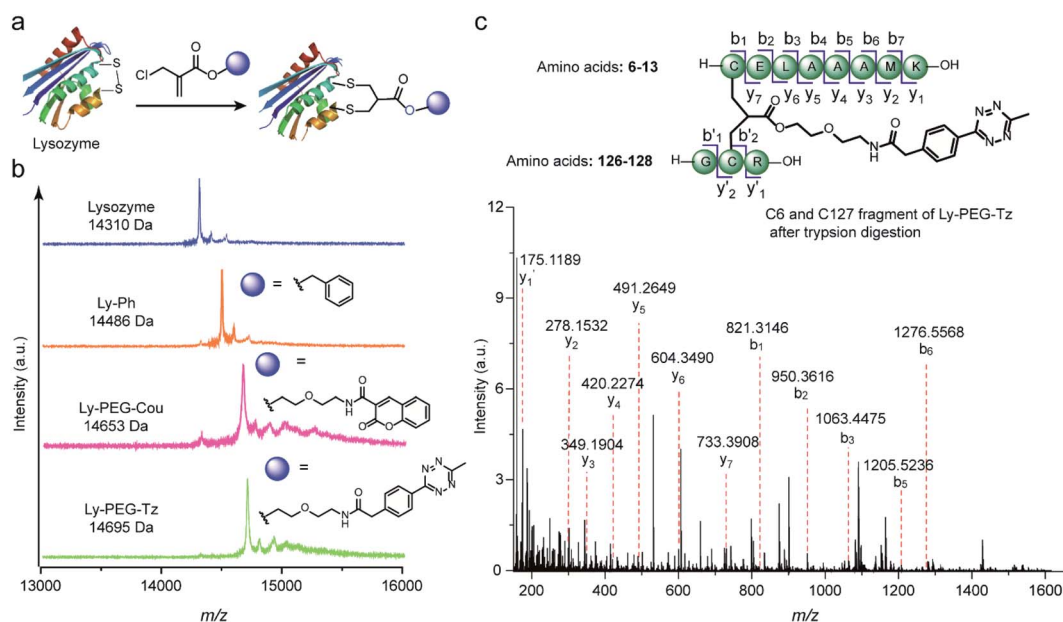


Fig. 7 (a) Site-selective disulfide modification of lysozyme with different functionalities. (b) MALDI-Tof-MS of modified lysozyme with different functionalities: Ly-Ph (calculated: 14 485 $[M + H]^+$, found: 14 486 $[M + H]^+$), Ly-PEG-Cou (calculated: 14 654 $[M + H]^+$, found: 14 653 $[M + H]^+$), Ly-PEG-Tz (calculated: 14 694 $[M + H]^+$, found: 14 695 $[M + H]^+$) (c) MS/MS analysis of the C6 and C127 fragment of Ly-PEG-Tz after trypsin digestion. The expected y and b ions and the zoom-in spectra of the respective fragment ions are given in Section 8.2 in ESI.†
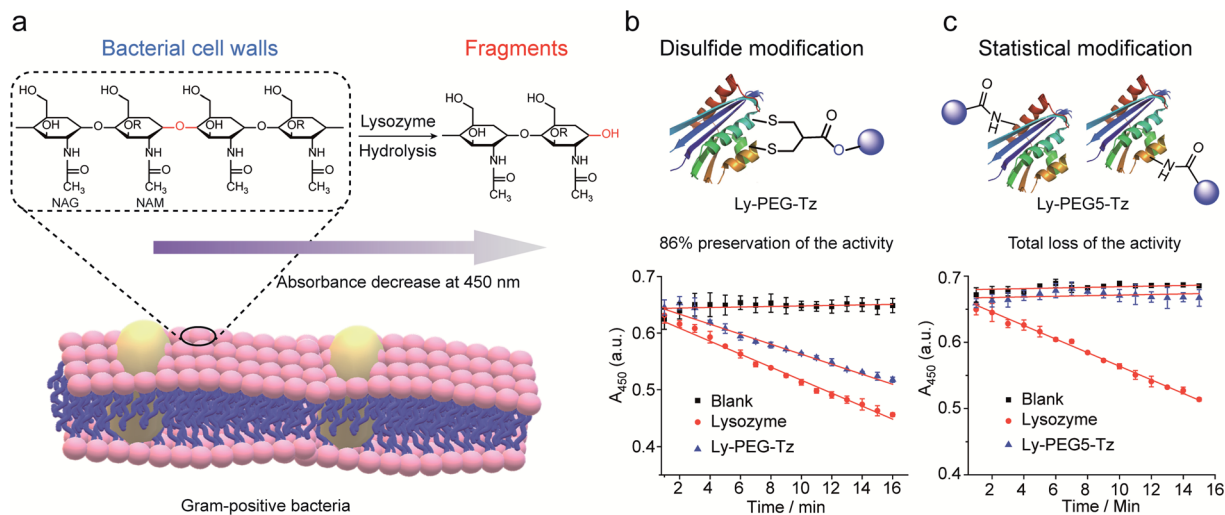
**Fig. 8** (a) The hydrolysis of 1,4-beta-linkages between *N*-acetyl-D-glucosamine (NAG) and *N*-acetylmuramic acid (NAM) residues in peptido-glycans of Gram-positive bacterial cell walls. The absorbance of the cell suspension at 450 nm was monitored to evaluate their catalytic activity and the rate of absorbance decrease was proportional to its activity. (b) Compared to native lysozyme (red), disulfide-modified lysozyme retained 86% activity (blue). The calculation details are shown in Section 8.3 of the ESI.† (c) Statistical modified lysozyme based on NHS ester chemistry resulted in total loss of the catalytic activity. The calculation details were shown in ESI.†

reagent (19% isolated yield, Table S12†) and comparable to that where cysteine in human serum albumin was modified with maleimide (∼30%).[20,43,44] Notably, around 25–30% of native lysozyme was recovered after the purification, which can be recycled for modification. In order to identify the modification site, Ly-PEG-Tz was analyzed by LC-MS/MS. After trypsin digestion, only the fragment containing C6–C127 disulfide bonds was observed with an addition of PEG-Tz functionality ($m/z$ 1553.7021 $[M + H]^+$, Fig. S49–S54†). The fragments showing modification at other disulfide bonds were not observed in the analysis (Table S7 in ESI†). Further MS/MS analysis confirms the expected sequence and demonstrates the site-selective modification at the disulfide site (Fig. 7c). The expected y and b ions and the zoom-in spectra of the respective fragment ions are shown in Section 8.2 in ESI.†

Lysozyme is an antimicrobial enzyme that is capable of hydrolyzing the 1,4-beta-linkages in the peptidoglycan of Gram-positive bacterial cell walls, thus leading to the lysis of bacteria (Fig. 8a). Therefore, the catalytic activity of modified lysozyme was assessed by investigation of the absorbance change at 450 nm of *Micrococcus lysodeikticus* lyophilized cell suspensions over time, where the activity of the modified lysozyme is proportional to their capability to hydrolyze the bacterial cell walls.[45] In comparison to native lysozyme, the disulfide-modified lysozyme Ly-PEG-Tz retained 86% of its activity (Fig. 8b, calculation details shown in Section 8.3 in ESI†). In contrast, statistical modification of lysine residues of lysozyme using tetrazine *N*-hydroxysuccinimide compounds (Scheme S13†), which gave a heterogeneous mixture according to the MS data (Fig. S67†), resulted in total loss of its catalytic activity (Fig. 8c, calculation details shown in Section 8.3 in ESI†). Hence, disulfide modification of proteins with 2-chloromethyl acrylate compounds represents an attractive approach to functionalize enzymatic proteins at distinct sites to preserve their catalytic activity.

## 3. Conclusion

In conclusion, we report that single atom substitution in 2-chloromethyl acryl reagents can achieve selective protein modification at cysteine or disulfide sites on demand. The reactivity profile of the prepared bioconjugation reagents can be customized by simply selecting different end-group functionalities (either amino or hydroxyl groups) to obtain the respective 2-chloromethyl acrylamide and acrylate compounds. Notably, the synthesis of the reported 2-chloromethyl acrylamide and acrylate compounds proceeds *via* a simple and easily implemented one-pot reaction based on easily accessible starting materials. We anticipate that the synthetic approach presented herein can be easily adapted in any laboratory for a broader scientific community.

Excellent labeling efficiency and high chemoselectivity of the 2-chloromethyl acrylamide compounds were demonstrated by the chemoselective modification of cysteine residues in several model peptides as well as proteins. In contrast, 2-chloromethyl acrylate regents allow modification of disulfide-containing peptides and proteins, such as SST, octreotide, and lysozyme. In addition, our new approach could offer the possibility for the dual modification of proteins by capitalizing on the reactivity difference of the 2-chloromethyl acrylamide and acrylate compounds. In this way, one could envision protein dual functionalization at cysteine residues and disulfide bonds can be achieved in a stepwise fashion within one system. We believe that the strategy presented herein offers an entirely new and elegant chemical approach to chemists and biologists to greatly enrich the currently available methodology toolbox for cysteine and disulfide modification. In this way, such progressive technologies will provide easy access to the broader scientific community in the design and preparation of advanced protein conjugates for various biological, biophysical, and medicinal applications.

117

## Data availability

Supporting data for this article is presented in the ESI.†

## Author contributions

T. W. conceptualized and supervised the project; S. L. K. supervised the project. L. X. designed and performed the experiments. M. J. S. A. S. and P. M. P. G. contributed with MS studies to elucidate the structure of several bioconjugates present in this manuscript. L. X. wrote the original draft of the paper; M. J. S. A. S., P. M. P. G., S. L. K. and T. W. reviewed and edited the paper.

## Conflicts of interest

The authors declare no conflict of interest.

## Acknowledgements

## References

1  N. Krall, F. P. da Cruz, O. Boutureira and G. J. L. Bernardes, *Nat. Chem.*, 2016, **8**, 103–113.

2  P. Akkapeddi, S. A. Azizi, A. M. Freedy, P. Cal, P. M. P. Gois and G. J. L. Bernardes, *Chem. Sci.*, 2016, **7**, 2954–2963.

3  P. M. Cal, G. J. Bernardes and P. M. Gois, *Angew. Chem., Int. Ed.*, 2014, **53**, 10585–10587.

4  C. D. Spicer and B. G. Davis, *Nat. Commun.*, 2014, **5**, 4740.

5  N. Stephanopoulos and M. B. Francis, *Nat. Chem. Biol.*, 2011, **7**, 876–884.

6  N. C. Reddy, M. Kumar, R. Molla and V. Rai, *Org. Biomol. Chem.*, 2020, **18**, 4669–4691.

7  G. Zhang, S. Zheng, H. Liu and P. R. Chen, *Chem. Soc. Rev.*, 2015, **44**, 3405–3417.

8  E. A. Hoyt, P. M. S. D. Cal, B. L. Oliveira and G. J. L. Bernardes, *Nat. Rev. Chem.*, 2019, **3**, 147–171.

9  L. Xu, S. L. Kuan and T. Weil, *Angew. Chem., Int. Ed.*, 2020, **60**, 13757–13777.

10  Y. Hou and H. Lu, *Bioconjugate Chem.*, 2019, **30**, 1604–1616.

11  O. Boutureira and G. J. L. Bernardes, *Chem. Rev.*, 2015, **115**, 2174–2195.

12  J. Ravasco, H. Faustino, A. Trindade and P. M. P. Gois, *Chem.–Eur. J.*, 2019, **25**, 43–59.

13  J. M. Chalker, G. J. Bernardes, Y. A. Lin and B. G. Davis, *Chem.–Asian J.*, 2009, **4**, 630–640.

14  P. Ochtrop and C. P. R. Hackenberger, *Curr. Opin. Chem. Biol.*, 2020, **58**, 28–36.

15  J. Lu, H. Wang, Z. Tian, Y. Hou and H. Lu, *J. Am. Chem. Soc.*, 2020, **142**, 1217–1221.

16  S. L. Kuan, T. Wang and T. Weil, *Chem.–Eur. J.*, 2016, **22**, 17112–17129.

17  V. Chudasama, A. Maruani and S. Caddick, *Nat. Chem.*, 2016, **8**, 114–119.

18  C. Zhang, P. Dai, A. A. Vinogradov, Z. P. Gates and B. L. Pentelute, *Angew. Chem., Int. Ed.*, 2018, **57**, 6459–6463.

19  H. Choi, M. Kim, J. Jang and S. Hong, *Angew. Chem., Int. Ed.*, 2020, **59**, 22514–22522.

20  T. Wang, A. Riegger, M. Lamla, S. Wiese, P. Oeckl, M. Otto, Y. Wu, S. Fischer, H. Barth, S. L. Kuan and T. Weil, *Chem. Sci.*, 2016, **7**, 3234–3239.

21  S. Shaunak, A. Godwin, J.-W. Choi, S. Balan, E. Pedone, D. Vijayarangam, S. Heidelberger, I. Teo, M. Zloh and S. Brocchini, *Nat. Chem. Biol.*, 2006, **2**, 312–313.

22  L. Xu, M. Raabe, M. M. Zegota, J. C. F. Nogueira, V. Chudasama, S. L. Kuan and T. Weil, *Org. Biomol. Chem.*, 2020, **18**, 1140–1147.

23  M. E. B. Smith, F. F. Schumacher, C. P. Ryan, L. M. Tedaldi, D. Papaioannou, G. Waksman, S. Caddick and J. R. Baker, *J. Am. Chem. Soc.*, 2010, **132**, 1960–1965.

24  A. Maruani, M. E. Smith, E. Miranda, K. A. Chester, V. Chudasama and S. Caddick, *Nat. Commun.*, 2015, **6**, 6645.

25  S. J. Walsh, S. Omarjee, W. Galloway, T. T. Kwan, H. F. Sore, J. S. Parker, M. Hyvonen, J. S. Carroll and D. R. Spring, *Chem. Sci.*, 2019, **10**, 694–700.

26  N. Martinez-Saez, S. Sun, D. Oldrini, P. Sormanni, O. Boutureira, F. Carboni, I. Companon, M. J. Deery, M. Vendruscolo, F. Corzana, R. Adamo and G. J. L. Bernardes, *Angew. Chem., Int. Ed.*, 2017, **56**, 14963–14967.

27  Y. Zhang, C. Zang, G. An, M. Shang, Z. Cui, G. Chen, Z. Xi and C. Zhou, *Nat. Commun.*, 2020, **11**, 1015.

28  C. E. Stieger, L. Franz, F. Korlin and C. P. R. Hackenberger, *Angew. Chem., Int. Ed.*, 2021, **60**, 15359–15364.

29  B. Bernardim, P. M. Cal, M. J. Matos, B. L. Oliveira, N. Martinez-Saez, I. S. Albuquerque, E. Perkins, F. Corzana, A. C. Burtoloso, G. Jimenez-Oses and G. J. Bernardes, *Nat. Commun.*, 2016, **7**, 13128.

30  Y. Liu, Z. Lai, P. Yang, Y. Xu, W. Zhang, B. Liu, M. Lu, H. Chang, T. Ding and H. Xu, *RSC Adv.*, 2017, **7**, 43104–43113.

31  D. T. Alexander Bohme, P. Albrecht and G. Schrmann, *Chem. Res. Toxicol.*, 2009, **22**, 742–750.

32  F. Saito, H. Noda and J. W. Bode, *ACS Chem. Biol.*, 2015, **10**, 1026–1033.

33  M. A. Kasper, A. Stengl, P. Ochtrop, M. Gerlach, T. Stoschek, D. Schumacher, J. Helma, M. Penkert, E. Krause, H. Leonhardt and C. P. R. Hackenberger, *Angew. Chem., Int. Ed.*, 2019, **58**, 11631–11636.

34 S. Ulrich, D. Boturyn, A. Marra, O. Renaudet and P. Dumy, *Chem.–Eur. J.*, 2014, **20**, 34–41.

35 A. Dirksen, T. M. Hackeng and P. E. Dawson, *Angew. Chem., Int. Ed.*, 2006, **45**, 7581–7584.

36 C. John, E. M. S. Jewett and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2010, **132**, 3688–3690.

37 O. Zirafi, K. A. Kim, L. Standker, K. B. Mohr, D. Sauter, A. Heigele, S. F. Kluge, E. Wiercinska, D. Chudziak, R. Richter, B. Moepps, P. Gierschik, V. Vas, H. Geiger, M. Lamla, T. Weil, T. Burster, A. Zgraja, F. Daubeuf, N. Frossard, M. Hachet-Haas, F. Heunisch, C. Reichetzeder, J. L. Galzi, J. Perez-Castells, A. Canales-Mayordomo, J. Jimenez-Barbero, G. Gimenez-Gallego, M. Schneider, J. Shorter, A. Telenti, B. Hocher, W. G. Forssmann, H. Bonig, F. Kirchhoff and J. Munch, *Cell Rep.*, 2015, **11**, 737–747.

38 R. E. Hansen, H. Østergaard, P. Nørgaard and J. R. Winther, *Anal. Biochem.*, 2007, **363**, 77–82.

39 S. L. Kuan, S. Fischer, S. Hafner, T. Wang, T. Syrovets, W. Liu, Y. Tokura, D. Y. W. Ng, A. Riegger, C. Förtsch, D. Jäger, T. F. E. Barth, T. Simmet, H. Barth and T. Weil, *Adv. Sci.*, 2018, 1701036.

40 S. W. J. Lamberts and L. J. Hofland, *Eur. J. Endocrinol.*, 2019, **181**, R173–R183.

41 M. Zegota, T. Wang, C. Seidler, D. Y. W. Ng, S. L. Kuan and T. Weil, *Bioconjugate Chem.*, 2018, **29**, 2665–2670.

42 R. Swaminathan, V. K. Ravi, S. Kumar, M. V. Kumar and N. Chandra, *Adv. Protein Chem. Struct. Biol.*, 2011, **84**, 63–111.

43 P. Zhang, H. Huang, S. Banerjee, G. J. Clarkson, C. Ge, C. Imberti and P. J. Sadler, *Angew. Chem., Int. Ed.*, 2019, **58**, 2350–2354.

44 A. J. Stewart, C. A. Blindauer, S. Berezenko, D. Sleep, D. Tooth and P. J. Sadler, *FEBS J.*, 2005, **272**, 353–362.

45 D. Shugar, *Biochim. Biophys. Acta*, 1952, **8**, 302–309.

119

# Supporting Information

**Chemoselective cysteine or disulfide modification via single atom substitution in chloromethyl acryl reagents**

Lujuan Xu[a,b], Maria J. S. A. Silva[c], Pedro M. P. Gois[c], Seah Ling Kuan[a,b*], Tanja Weil[a,b*]

[a]Max-Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz (Germany)

[b]Institute of Inorganic Chemistry I, Ulm University Albert-Einstein-Allee 11, 89081, Ulm (Germany)

[c] Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, 1649-003 Lisbon (Portugal)

* Corresponding authors

## 1. Materials and methods

### 1.1 Materials

Unless otherwise stated, all solvents and reagents are purchased from commercial sources (Merck, Sigma Aldrich) and are used directly without further purification. High-performance liquid chromatography (HPLC) was performed using acetonitrile (ACN, HPLC grade) and water (obtained from a Millipore purification system), both containing 0.1 % trifluoroacetic acid (TFA). The reactions were monitored by thin-layer chromatography (TLC) with Macherey-Nagel Alugram Sil G/UV254 plates at 254 nm or with proper stains ($KMnO_4$ solutions). Flash column chromatography is carried out with silica gel (0.04 mm–0.063 mm, 60 Å).

### 1.2 Methods and Instruments

1.2.1 Nuclear Magnetic Resonance Spectroscopy (NMR)

The NMR spectra are recorded using Bruker Avance 300 NMR fourier transform spectrometer using $CDCl_3$, $CD_2Cl_2$ or $CD_3OD$. The chemical shifts ($\delta$) were reported as parts per million (ppm) referenced with respect to the residual solvent peaks.

1.2.2 High-Performance Liquid Chromatography (HPLC)

Preparative HPLC was performed using Shimadzu LC-20AP system. Either Atlantis T3 Prep OBDTM 5 µm, 19 × 150 mm column (with a flowrate of 10 mL/min) or a Phenomenex Gemini 5 µm NX-C18 110 Å 150 × 30 mm (with a flowrate of 25 mL/min) was used. The gradient started with 5% ACN, which was linearly increased to 100% ACN within 20 minutes.

Analytical HPLC was performed using Shimadzu LC-20AT system. Atlantis T3 column (4.6 × 100 mm, 5 µm) with a flowrate of 4 mL/min was used. The gradient started with 5% ACN, which was increased to 100% ACN within 20 minutes or 26 minutes depending on the compounds.

2

For both preparative and analytical HPLC, ACN and water, which contains 0.1% TFA, were used as eluting solvent. The absorbance was recorded at 190 nm, 214 nm, 254 nm and 520 nm. The software LabSolutions by Shimadzu and Powerpoint were used to process all HPLC spectra.

1.2.3 Liquid Chromatography - Mass Spectrometry (LC-MS)

LC-MS(ESI) was measured using Shimadzu LCMS2020 with a Kinetex 2.6 μm EVO C18 100 Å LC 50 × 2.1 mm column, an electrospray ionization source, a SPD-20A UV-Vis detector. MilliQ water and ACN, both of which contain 0.1% formic acid, were used as the eluting solvents for all the measurements. The solvent gradients start with 5% ACN and 95% water, reach 100% ACN at 16 minute, and then go back to 5% ACN and 95% water at 20 minute. Data were processed with LabSolutions provided by Shimadzu.

1.2.4 Matrix-Assisted Laser Desorption/Ionisation - Time of Flight Mass Spectrometry (MALDI-Tof-MS)

All the MALDI-Tof-MS spectra were obtained from rapifleX MALDI-TOF/TOF from Bruker. Protein samples were mixed with a saturated solution with sinapinic acid and peptide samples were mixed with α-cyano-4-hydroxycinnamic acid (CHCA) in water/ACN 1/1 + 0.1% TFA. Data processing was performed in mMass.

1.2.5 Circular Dichroism Spectroscopy (CD)

Circular dichroism was measured at room temperature from 260 to 190 nm (a bandwith of 1 nm) with the sample concentration of 0.1 mg/mL in PBS buffer (pH 7.4). The data pitch was set to 0.2 nm with the scanning speed 5 nm/min. Each sample was measured three times and the signal from the buffer blank was subtracted from the sample scan. The obtained data was processed in the software Spectra Analysis and CD Multivariate SSE by JASCO

3

## 2. Synthesis of the 2-chloromethyl acrylamide and acrylate compounds

General procedure for one-pot synthesis of compound **1** to **8**
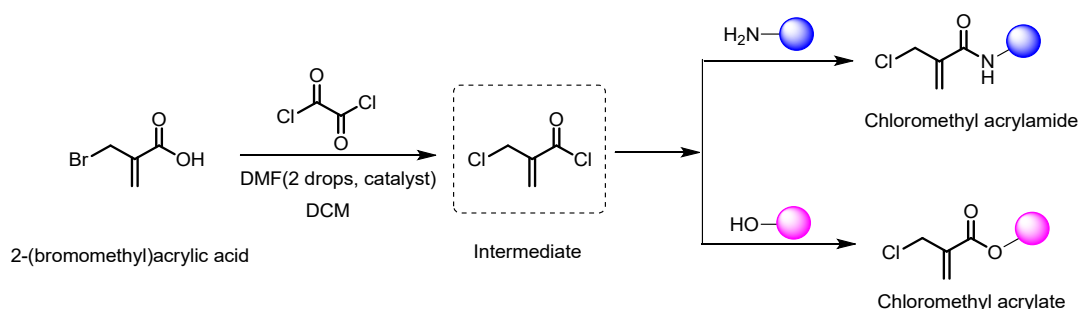


**Scheme S1** The general synthetic scheme of the respective 2-chloromethyl acrylamide and acrylate derivatives with using the commercial available 2-(bromomethyl)acrylic acid as starting material.

Commercial available compound 2-(bromomethyl)acrylic acid (200 mg, 1.2 mmol, 1 eq) was added to a dry round-bottom with 10 mL anhydrous dichloromethane (DCM) under argon. Next, oxalyl chloride (1.2 g, 9.6 mmol, 8 eq) was added dropwise at 0°C. Then, several drops of anhydrous dimethylformamide (DMF) were also added via a syringe. The mixture was stirred under argon for 4 h. After that, the solvent was removed in vacuo and a white-to-yellow solid was obtained in the bottle. Without any further purification, the obtained intermediate, 2-(chloromethyl)acryloyl chloride, was dissolved in 10 mL anhydrous DCM and used directly for the next step. Thereafter, *N,N*- diisopropylethylamine (DIEA, 2 eq) and the amine or alcohol substrate (1 eq) were added to the acid chloride solution sequentially at 0°C. The resulting reaction mixture was stirred at room temperature overnight. After that, the mixture was washed with 1 M NaCl solution and the organic layer was dried over anhydrous MgSO$_4$. The solvent was evaporated under reduced pressure and the crude product was purified either with HPLC or flash column chromatography.

2.1 Synthesis of compound **1**
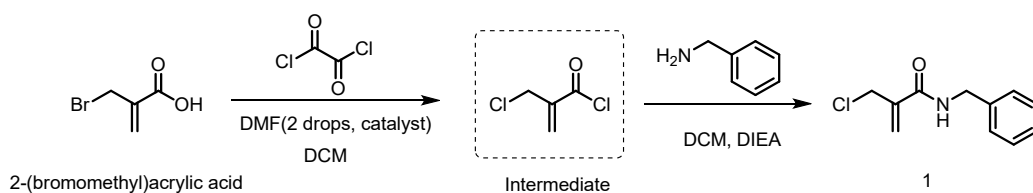
Compound **1** was synthesized according to the procedure described above. The crude product was purified

with flash column chromatography (DCM:MeOH = 25:1) and compound **1** was obtained as white solid with

71% yield.

^{1}H NMR (300 MHz, CDCl$_3$) δ 7.44–7.28 (m, 5H), 5.86 (s, 1H), 5.70 (s, 1H), 4.54 (d, $J$ = 5.7 Hz, 2H), 4.35 (s,

2H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 166.04, 141.12, 137.84, 128.81, 127.78, 127.68, 122.05, 43.86,43.55

## 2.2 Synthesis of Compound **2**
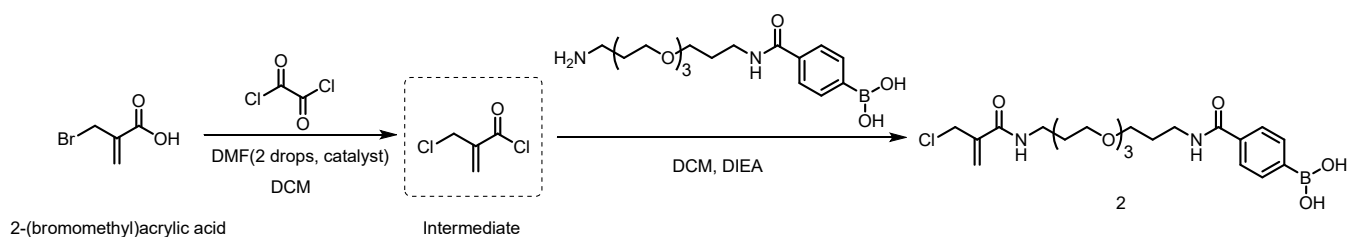


Amine-PEG3-B(OH) was synthesized based on the methods published from our group before.[1-3] Compound

**2** was synthesized according to the general procedure described above and was purified with preparative

HPLC using the general methods described above to offer a colorless liquid with 61% yield.

^{1}H NMR (300 MHz, CDCl$_3$) δ 7.81 (dd, $J$ = 17.8, 7.6 Hz, 4H), 5.86 (s, 1H), 5.68 (s, 1H), 4.31 (s, 2H), 3.74 –

3.54 (m, 12H), 3.51 – 3.45 (m, 4H), 3.42 – 3.35 (m, 2H), 1.96 – 1.85 (m, 2H), 1.84 – 1.69 (m, 2H).^{13}C NMR

(75 MHz, CD$_2$Cl$_2$) δ 168.75, 143.06, 134.69, 127.16, 122.70, 71.55, 71.53, 71.31, 71.25, 70.31, 70.04, 44.32,

38.75, 38.28, 30.42, 30.28. LC-MS: m/z = 471 [M+H]$^+$, 493 [M+Na]$^+$

## 2.3 Synthesis of Compound **3**

5
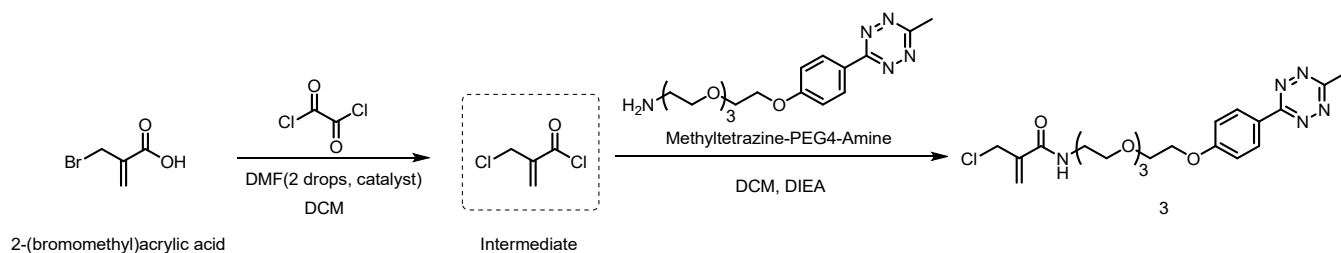
Compound **3** was synthesized according to the procedure described above. The crude product was purified with flash column chromatography (DCM:MeOH = 20:1) to obtain compound **3** as a pink to purple solid with 68% yield.

$^1$H NMR (300 MHz, CDCl$_3$) δ 8.53 (d, $J$ = 8.8 Hz, 2H), 7.09 (d, $J$ = 8.8 Hz, 2H), 5.86 (s, 1H), 5.67 (s, 1H), 4.31 (s, 2H), 4.24 (t, 2H), 3.90 (t, 2H), 3.79 – 3.58 (m, 11H), 3.57 – 3.51 (m, 2H), 3.06 (s, 3H).$^{13}$C NMR (75 MHz, CDCl$_3$) δ 166.63, 166.21, 163.71, 162.43, 141.16, 129.72, 124.41, 121.86, 115.25, 70.93, 70.61, 70.34, 69.62, 67.65, 53.43, 43.54, 39.56, 29.78, 21.06.LC-MS: m/z = 466 [M+H]$^+$, 488 [M+Na]$^+$

2.4 Synthesis of compound **4**
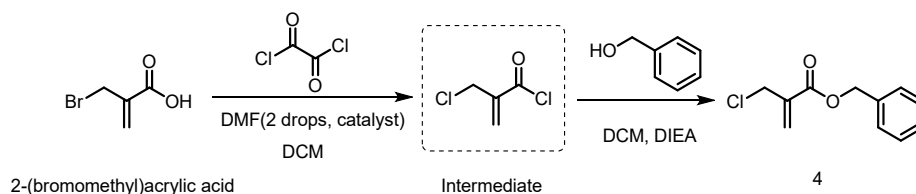


Compound **4** was synthesized according to the procedure described above. The crude product was purified with flash column chromatography (DCM:MeOH = 40:1) and compound **4** was obtained as white solid with 85% yield.

$^1$H NMR (300 MHz, CDCl$_3$) δ 7.49 – 7.29 (m, 5H), 6.44 (s, 1H), 6.01 (s, 1H), 5.25 (s, 2H), 4.31 (s, 2H).$^{13}$C NMR (75 MHz, CDCl$_3$) δ 164.88 (s), 136.81 (s), 135.56 (s), 129.05 (s), 128.63 (s), 128.38 (s), 128.17 (s), 66.97 (s), 42.53 (s).

2.5 Synthesis of compound **5**

6

2-(bromomethyl)acrylic acid → Intermediate → 5

DMF(2 drops, catalyst) DCM

DCM, DIEA
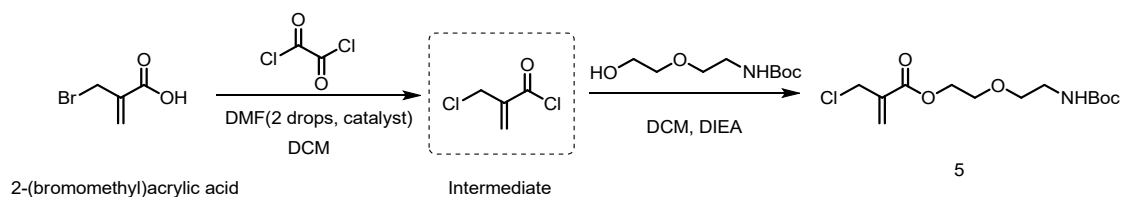
Compound **5** was synthesized according to the procedure described above. The crude product was purified with flash column chromatography (DCM:MeOH = 40:1) and compound **5** was obtained as white liquid with 67% yield.

$^1$H NMR (300 MHz, CDCl$_3$) δ 6.40 (s, 1H), 5.99 (s, 1H), 4.34 (t, 2H), 4.28 (s, 2H), 3.70 (t, 2H), 3.53 (t, $J$ = 5.1 Hz, 2H), 3.36 – 3.24 (m, 2H), 1.42 (s, 10H). $^{13}$C NMR (75 MHz, CD$_2$Cl$_2$) δ 165.17, 156.12, 137.26, 129.11, 79.24, 70.48, 69.05, 64.58, 43.07, 40.71, 28.48. LC-MS: m/z = 207 (Boc-deprotection under conditions used for LC-MS), 308 [M+H]$^+$, [330 M+Na]$^+$

2.6 Synthesis of compound **6**



2-(bromomethyl)acrylic acid → Intermediate → 6

DMF(2 drops, catalyst) DCM

DCM, DIEA
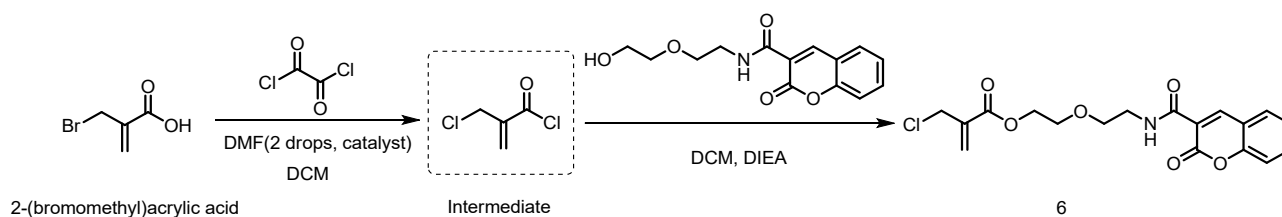
Compound **6** was synthesized according to the procedure described above. The crude product was purified with preparative HPLC via using the general method described above and compound **6** was obtained as yellow liquid with 58% yield.

$^1$H NMR (300 MHz, CDCl$_3$) δ 7.76 – 7.62 (m, 2H), 7.46 – 7.31 (m, 2H), 6.44 (s, 1H), 5.99 (s, 1H),   4.38 (t, 2H), 4.30 (s, 2H), 3.77 (t, 3H), 3.69 (s, 4H). $^{13}$C NMR (75 MHz, CDCl$_3$) δ 164.97, 161.67, 161.24, 154.52, 148.31, 136.65, 134.05, 129.80, 129.11, 125.26, 118.68, 118.45, 116.65, 69.66, 68.85, 64.25, 42.57, 39.69. LC-MS: m/z = 380 [M+H]$^+$, 402 [M+Na]$^+$

2.7 Synthesis of compound **7**

7
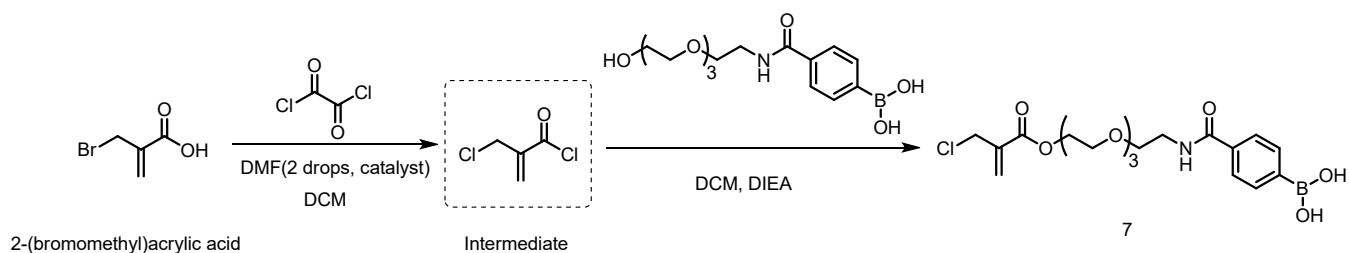
Compound **7** was synthesized according to the procedure described above.[1-3] The crude product was purified with preparative HPLC via using the general methods described above and compound **7** was obtained as white liquid with 64% yield.

$^1$H NMR (300 MHz, CD$_2$Cl$_2$) δ 7.89 (dd, $J$ = 22.9, 7.5 Hz, 4H), 6.53 (s, 1H), 6.13 (s, 1H), 4.56 – 4.37 (m, 4H), 3.97 – 3.69 (m, 12H). $^{13}$C NMR (75 MHz, CD$_2$Cl$_2$) δ 168.19, 165.36, 137.28, 134.60, 129.33, 126.47, 70.75, 70.61, 70.48, 70.38, 69.29, 64.67, 43.06, 40.26, 25.14. LC-MS: m/z = 444 [M+H]$^+$, 466 [M+Na]$^+$

2.8 Synthesis of compound **8**
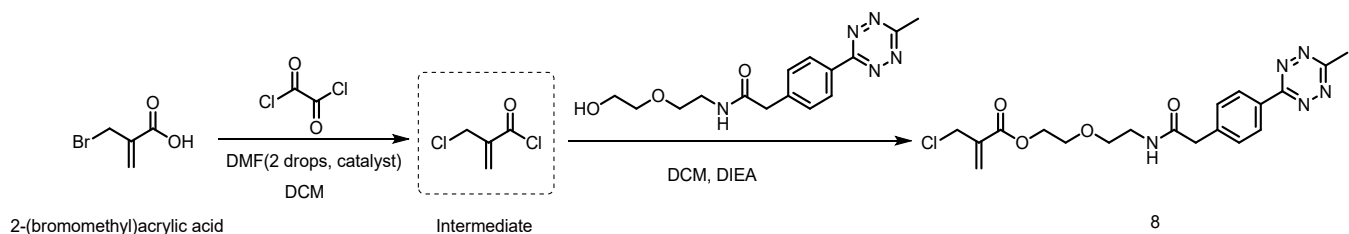


Compound **8** was synthesized according to the procedure described above The crude product was purified with flash column chromatography (DCM:MeOH = 25:1) and compound **8** was obtained as pink solid with 56% yield.

$^1$H NMR (300 MHz, CDCl$_3$) δ 8.46 (d, $J$ = 8.1 Hz, 2H), 7.44 (d, $J$ = 8.1 Hz, 2H), 6.28 (s, 1H), 5.91 (s, 1H), 4.31 – 4.14 (m, 4H), 3.60 (t, 2H), 3.56 (s, 2H), 3.47 (t, $J$ = 5.0 Hz, 2H), 3.34 (dd, $J$ = 10.2, 5.1 Hz, 2H), 2.99 (s, 1H). $^{13}$C NMR (75 MHz, CDCl$_3$): δ 170.16, 167.30, 164.80, 163.85, 139.78, 136.72, 130.80, 130.31, 129.25, 128.38, 69.61, 68.84, 64.00, 43.53, 42.70, 39.50, 21.17.LC-MS: m/z = 420 [M+H]$^+$, 442 [M+Na]$^+$

8

## 2.9. Proposed mechanism
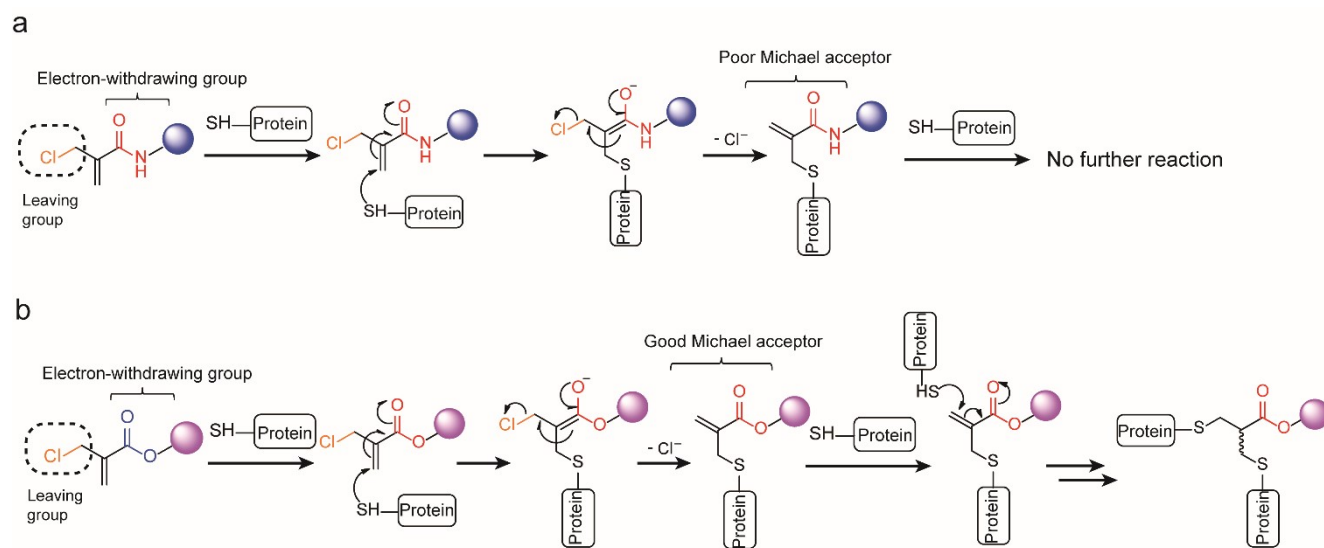


Fig. S1 Proposed mechanism for the reaction of 2-chloromethyl acryl derivatives and proteins which contain surface-exposed cysteines.

## 2.10 Log P (*n*-octanol to water) prediction of 2-chloromethyl acrylamide and acrylate derivative based on Chembiodraw 2016

Partition coefficient (P) refer to the ratio of the concentration of a compound in two immiscible solvents in equilibrium, such as *n*-octanol and water. The logarithm of the ratio is log P, which is often used in the literature and considered as a parameter to estimate how hydrophilic or hydrophobic a compound is. The smaller the log P is, the more hydrophilic a compound should be and vice versa. The log P of the 2-chloromethyl acrylamide and acrylate derivatives are summarized below. The log P data suggested that the 2-chloromethyl acrylamide and acrylate compounds \ designed in this paper have lower log P, thus they should be more hydrophilic than the corresponding carbonylacrylic and allyl sulfone derivatives developed by Bernardes[4] and our group[5] respectively. Higher hydrophilicity of the conjugation reagent should facilitate the subsequent reactions as less organic solvent is required to eliminate the risk for protein degradation or

9

aggregation.
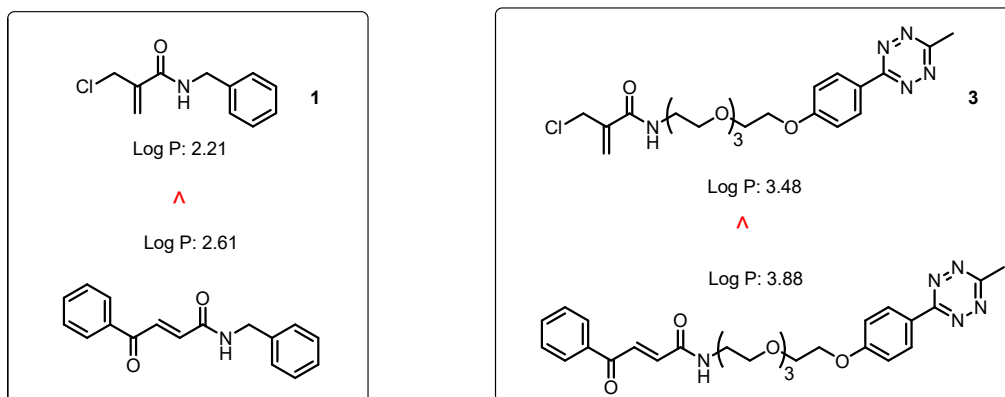


Fig. S2 The log P value of the 2-chloromethyl acrylamide derivatives (compound **1** and **3**) and the corresponding carbonylacrylic derivatives (developed by Bernardes and coworkers[4]) predicted from Chemdraw
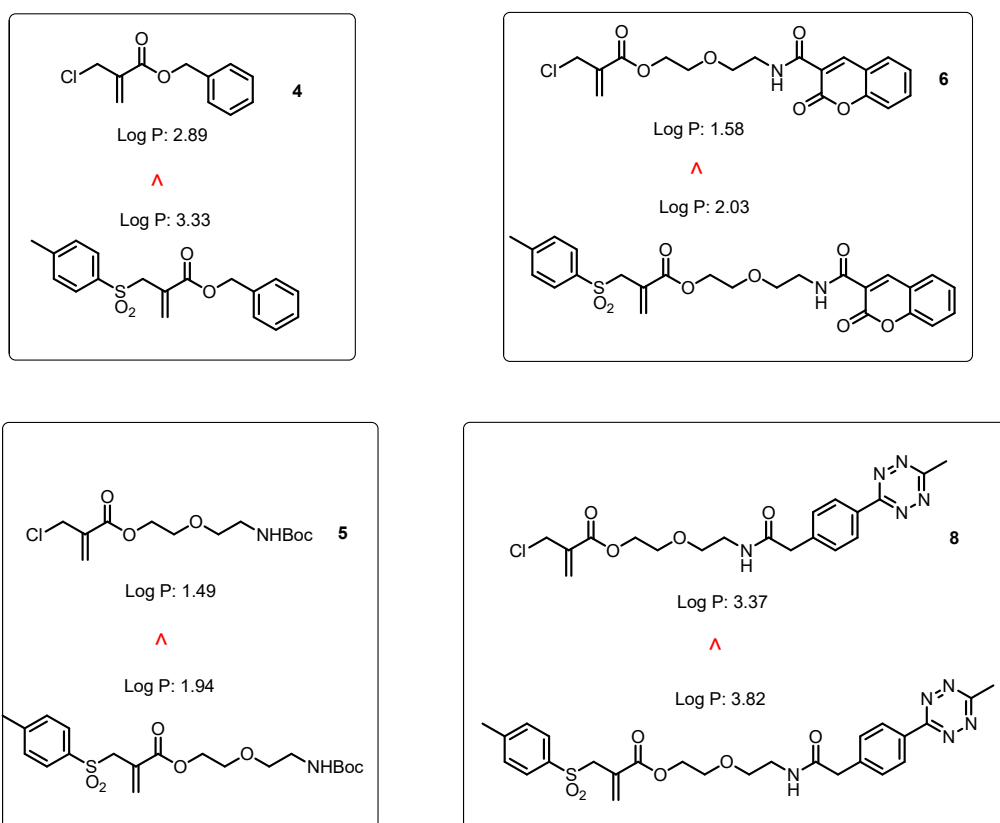


Fig. S3 The log P value of the 2-chloromethyl acrylate derivatives (compound **4**, **5**, **6** and **8**) and the corresponding allyl sulfone derivatives (developed by our group before[5]) predicted from Chemdraw.

2.10 Stability study of the 2-chloromethyl acrylamide and acrylate reagents

Compound **3** and compound **4** were selected for stability studies. In general, compound **3** (or **4**) was dissolved in 50 mM PB buffers at three different pH (pH 6, 7 or 8) with a concentration of 1 mg/mL. The mixed solutions at three different pH were incubated for 12, 24 and 36 hours. An aliquot of 10 µL of each solution was taken at each time point, mixed with 200 µL methanol and injected into HPLC to evaluate their stability.
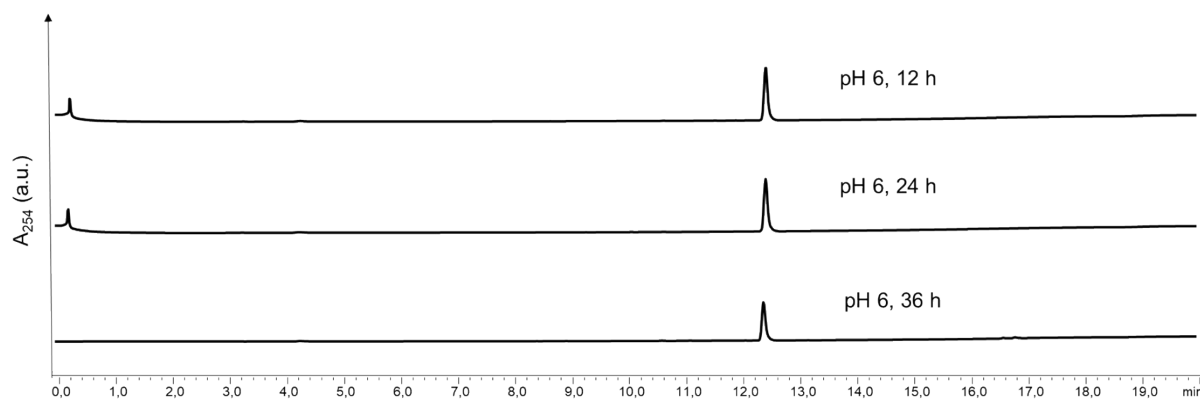


Fig. S4 Stability study of compound **3** at pH 6.
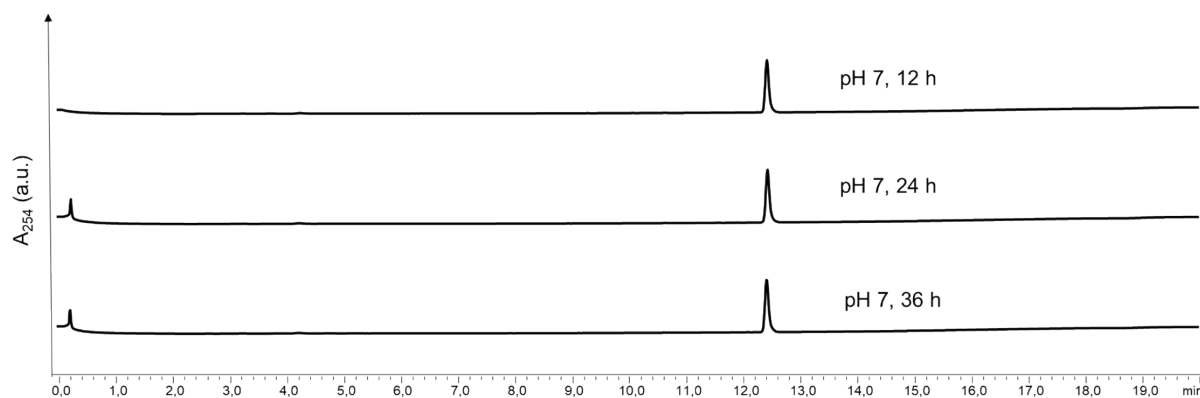


Fig. S5 Stability study of compound **3** at pH 7.
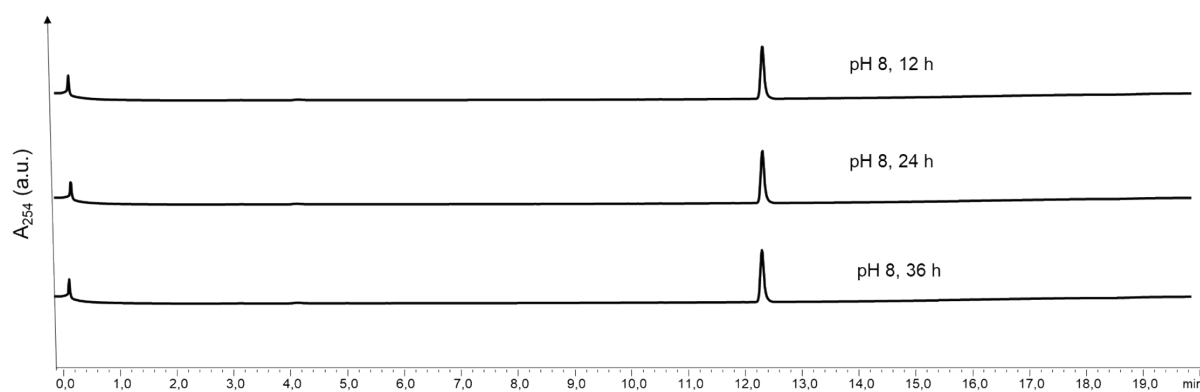


Fig. S6 Stability of compound **3** at pH 8.

11
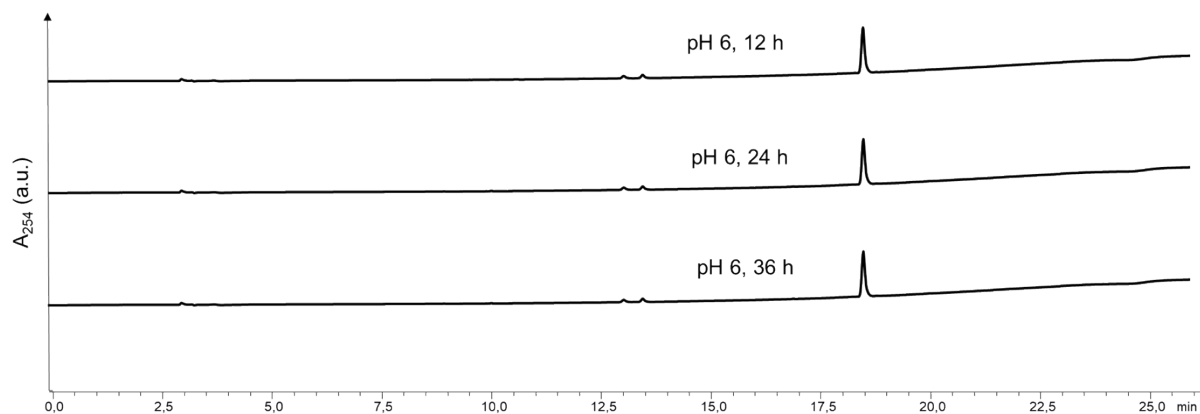
Fig. S7 Stability of compound **4** at pH 6.
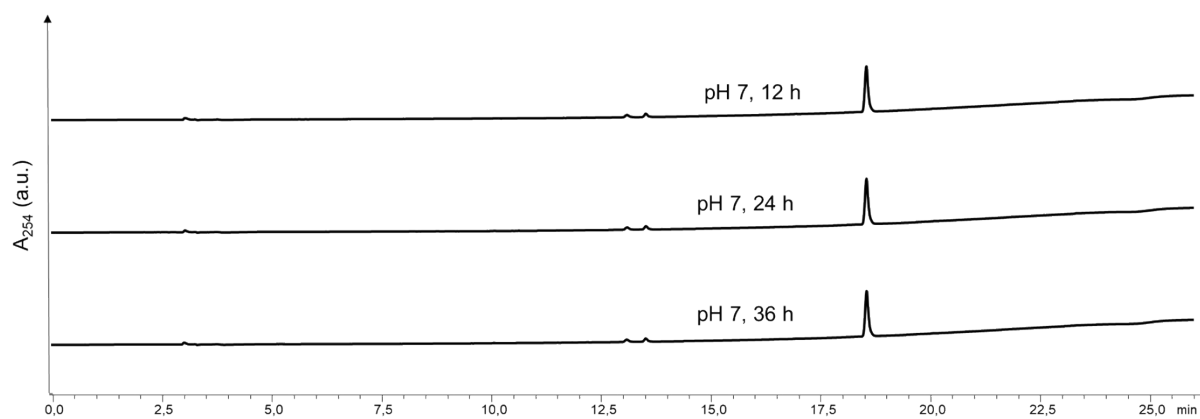


Fig. S8 Stability of compound **4** at pH 7.
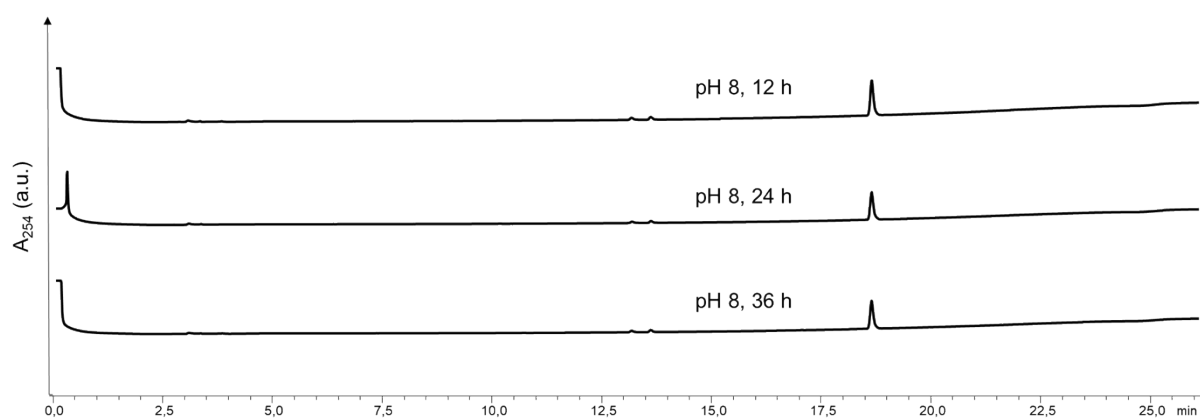


Fig. S9 Stability of compound **4** at pH 8.

2.11 Stability studies of the maleimide reagent

Maleimide compounds are the most commonly applied cysteine-modification reagent in the literature.

12

However, it is easy to undergo hydrolysis to obtain the ring-opening compound.[6] The hydrolytic pathway is shown below. The stability of the maleimide reagents was also evaluated at three different pH (pH 6, 7 and 8) and the results are shown below.
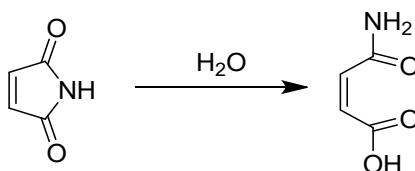


Scheme S2 General hydrolytic pathway of maleimide compounds.
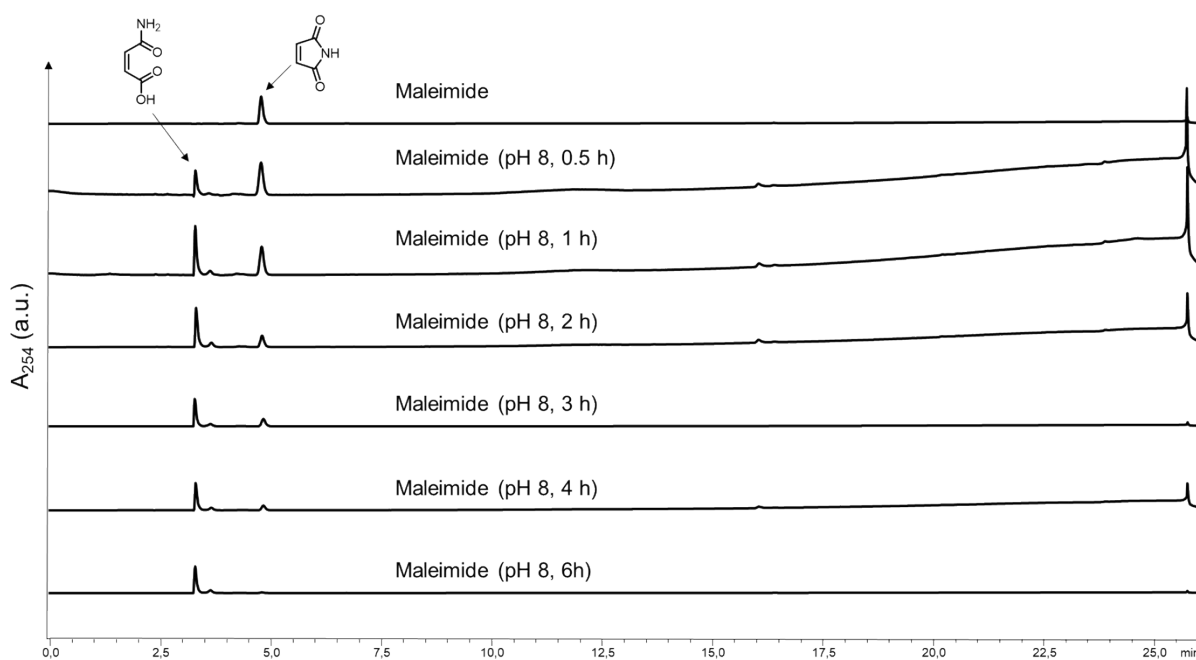


Fig. S10 Stability of maleimide reagent at pH 8.
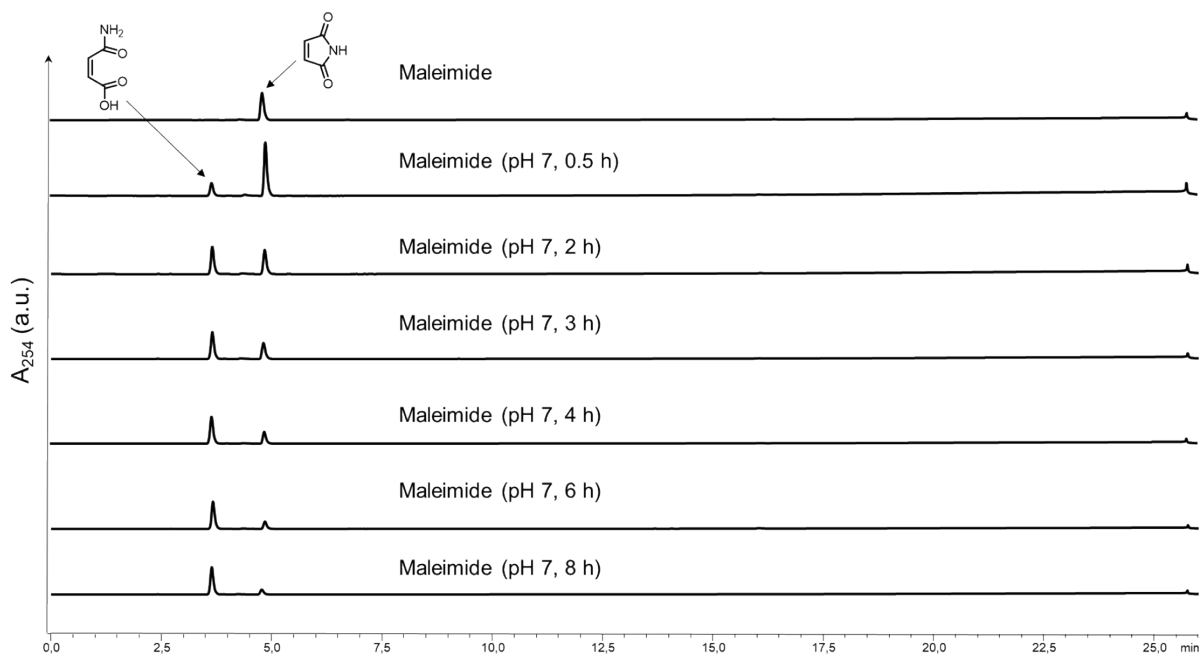
Fig. S11 Stability of maleimide reagents at pH 7.
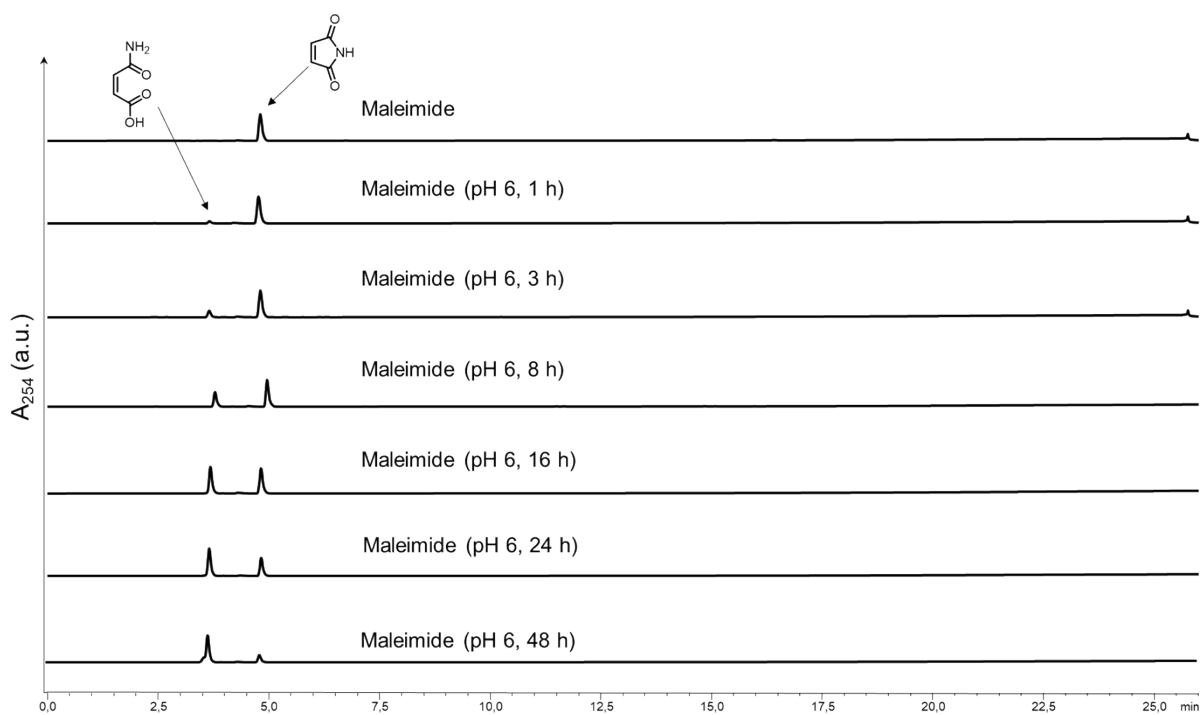


Fig. S12 Stability of maleimide reagent at pH 6.

14

## 3. Procedure for conducting model reactions to check chemoselectivity

To evaluate the chemoselectivity of the reagents towards thiol and amino groups, two model amino acids, Boc-Cys-OMe and Boc-Lys-OH, were selected to react with compound **1** or compound **4**. For the reaction between 2-chloromethyl acrylamide compound **1** and Boc-Cys-OMe (and Boc-Ly-OH), compound **1** (50 μg, 1 eq, 0.24 μmol) was mixed with Boc-Cys-OH (67 μg ,1.2 eq, 0.29 μmol, 50 mg/mL in DMF stock solution) and Boc-Ly-OH (70 μg, 1.2 eq, 0.29 μmol, 50 mg/mL in DMF stock solution) in ACN:PB (pH = 7) = 1:10. The mixture was incubated at room temperature for four hours. After that, 10 μL of the reaction mixture was mixed with 200 μL methanol and injected to LC-MS to check the product. By using the same procedure, the ratio of Boc-Cys-OMe was gradually increased to 2 eq, 4 eq, 6 eq, and 8 eq. After 4 h, the mixture was again diluted with methanol and injected to LC-MS to check the product.
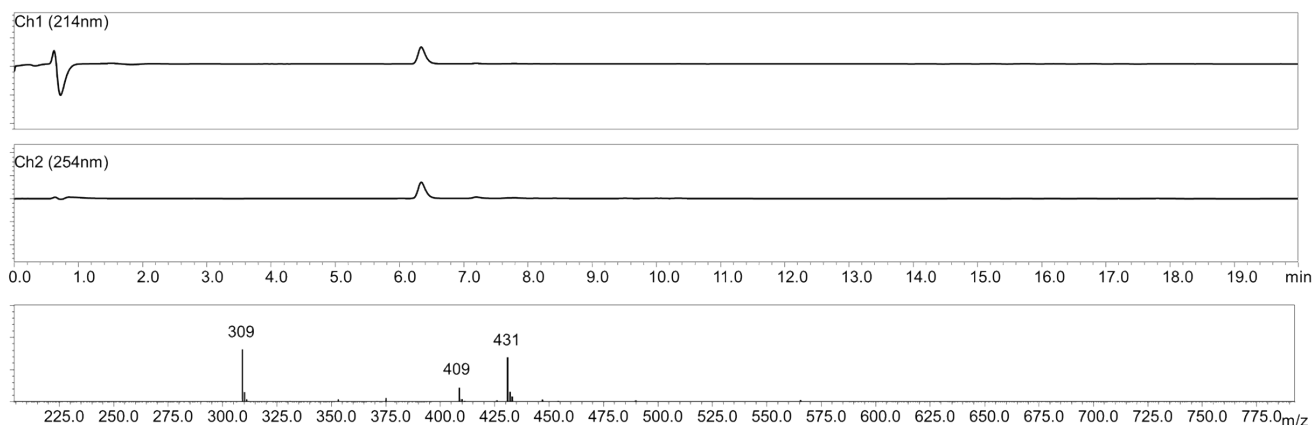


Fig. S13 LC-MS of compound **10** (calculated: 409 [M+H]$^+$, 431 [M+Na]$^+$, found: 409 [M+H]$^+$, 431 [M+Na]$^+$)

**Reactivity studies of compound 10 with different nucleophiles**

15
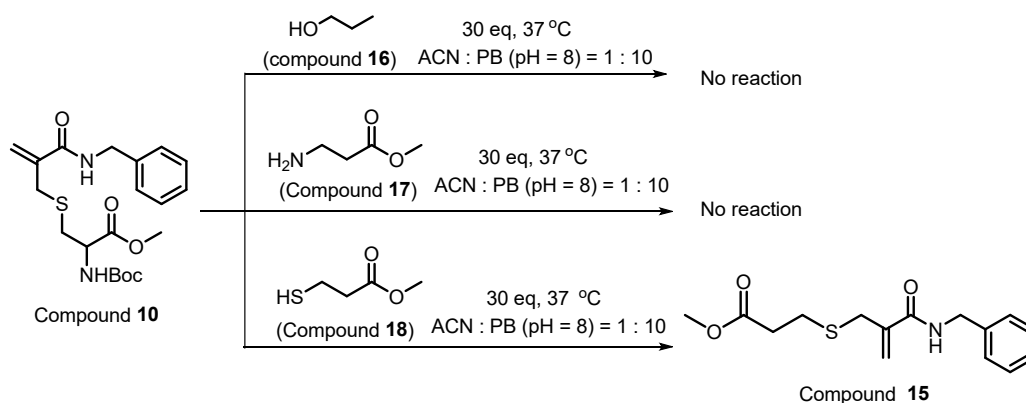
Scheme S3. Assessment of the reactivity of compound 10 with different nucleophiles (hydroxyl group, amino group and thiol group).

In order to assess the reactivity of compound **10** with different nucleophiles (hydroxyl group, amino group and thiol group), compound **10** (50 µg, 1 eq, 0.12 µmol) was incubated with a large excess (30 eqiv.) of compound **16**, **17** or **18** in ACN / PB (pH = 8) at 37 °C for overnight. No further reaction was observed between compound **10** and compound **16** or **17** on the LC-MS trace. However, compound **15** was formed after incubating compound **10** with compound **18** overnight at 37 °C, in which the first thiol-functionality was eliminated presumably due to an addition-elimination reaction. The reactions were monitored using LC and MS (Figure S14).
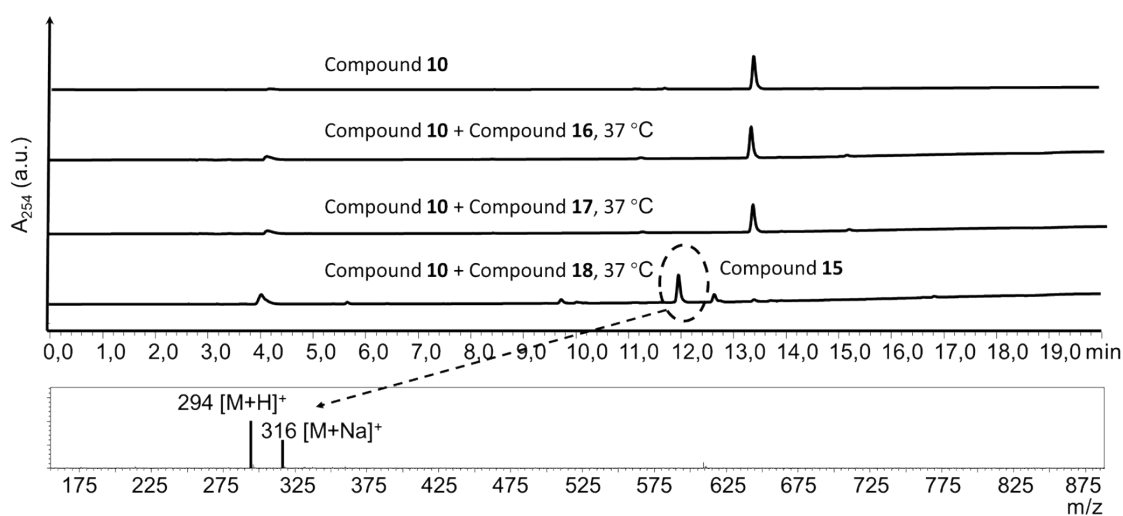


Fig. S14 HPLC analysis of the reaction between compound **10** and three different nucleophiles (hydroxyl, amino and thiol groups)

16

For the reaction between 2-chloromethyl acrylate compound **4** and Boc-Cys-OMe (and Boc-Ly-OH), compound **4** (50 µg, 1 eq, 0.24 µmol, 40 mg/mL in DMSO stock solution) was mixed with 67 µg ,1.2 eq, 0.29 µmol, 50 mg/mL in DMF stock solution) and Boc-Ly-OH (70 µg, 1.2 eq, 0.29 µmol, 50 mg/mL in DMF stock solution) in ACN:PB (pH = 7) = 2:3. After that, 10 µL of the reaction mixture was diluted with 200 µL methanol and injected to LC-MS to check the product. Following the same procedure, the ratio of Boc-Cys-OMe increased from 1.2 eq to 2 eq, 3 eq, 6 eq, and finally to 8 eq. For each reaction, 10 µL of reaction mixture was diluted with 200 µL methanol and LC-MS was used to evaluate the reaction progress.
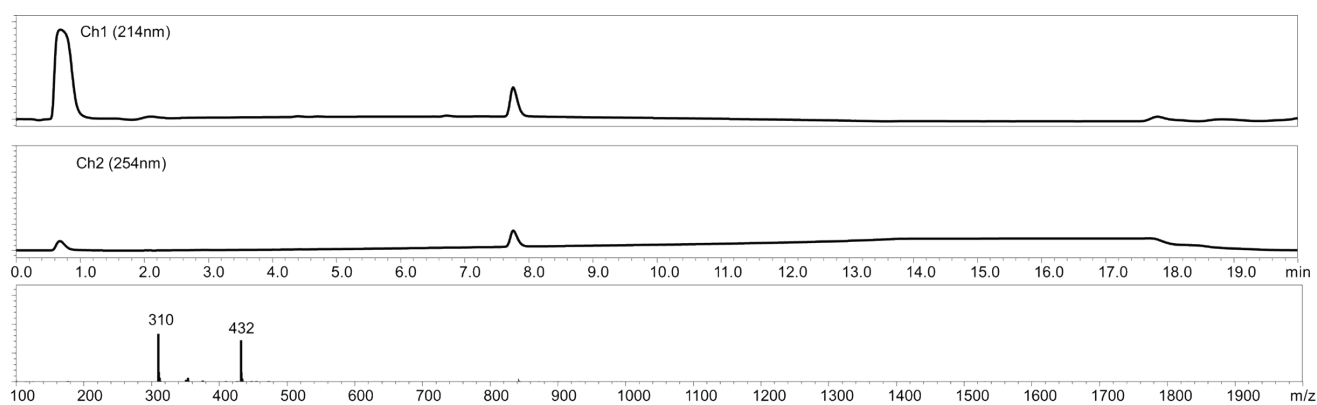


Fig. S15 LC-MS of compound **12** (calculated: 410 [M+H]+ , found: 432 [M+Na]+, 310 [boc-deprotected product]+)
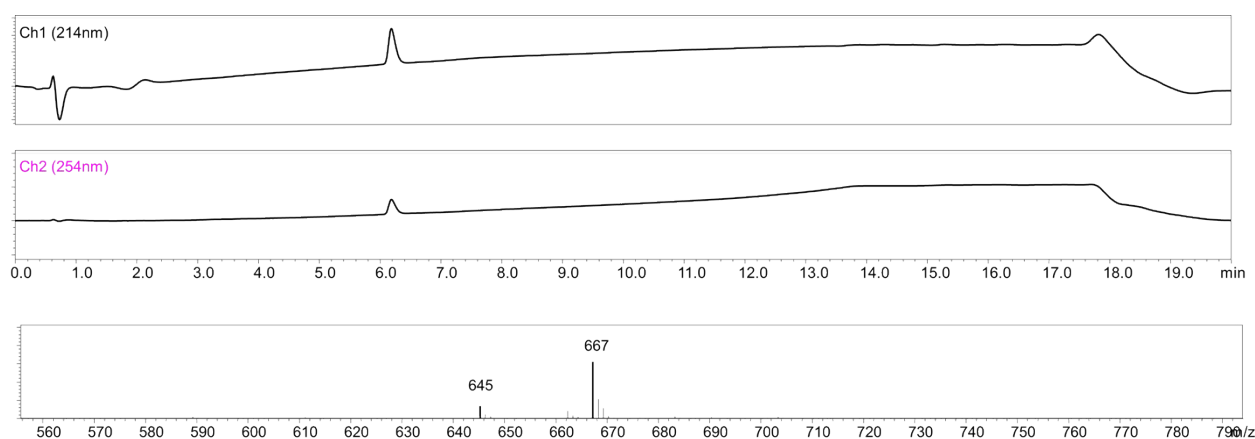


Fig. S16 LC-MS of compound **13** (calculated: 645 [M+H]+, 667 [M+Na]+, found: 645 [M+H]+, 667 [M+Na]+)

## 4. Kinetic study
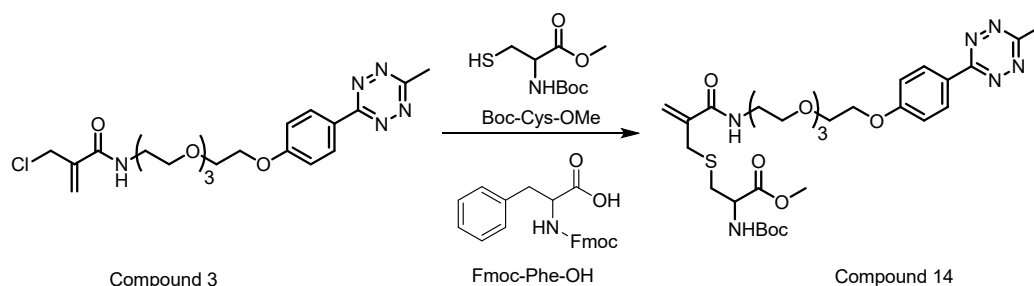


Scheme S4 Reaction scheme between compound **3** and Boc-Cys-OMe with Fmoc-Phe-OH as internal standard

For the kinetic study, compound **3** (1 mM, 20 mg/mL in DMF stock solution), Boc-Cys-OMe (1 mM, 50 mg/mL in DMF stock solution) and internal standard Fmoc-Phe-OH (0.2 mM, 40 mg/mL in DMF) were mixed in 200 μL ACN : PB (pH 7) mixture in 1.5 mL eppendorf. At certain time point (e.g. 15 min, 30 min, 60 min, 90 min, 120 min, 180 min, 240 min, 270 min, 360 min), 10 μL of reaction mixture was withdrawn from the eppendorf and the reaction was immediately quenched by 200 μL methanol which contained 10 μL trifluoroacetic acid. After quenching the reaction, the mixture was frozen before injecting to HPLC. HPLC data clearly indicated that the peak for compound **3** was gradually decreased while the peak for compound **14** increased. The quantification was obtained based on the integration of the absorbance peak of compound **3** and compound **14** in comparison to the internal standard Fmoc-Phe-OH, which correspond to the conversion during the reaction progress.

The rate constant of the reaction between compound **3** and Boc-Cys-OMe was obtained based on the second-order reaction kinetics. The second-order reaction equation: v = k × [A] [B]

If [A] = [B], $-\dfrac{d[A]}{dt} = k \times [A]^2$, so $\dfrac{d[A]}{[A]^2} = -k \times t$,

$$\int_{[A_0]}^{[A]} \frac{d[A]}{[A]^2} = -k \int_{0}^{t} dt$$

$$\frac{1}{[A]} = \frac{1}{[A_0]} + kt$$

18

[A]: real-time concentration of compound **3**, [B]: real-time concentration of Boc-Cys-OMe

[$A_0$]: initial concentration of compound **3** (1 mM), k: second-order rate constant.

Based on the HPLC data, $\dfrac{1}{[A]}$ over time was linearly-plotted which is shown in Fig. 3 and the slope is the second-order rate constant ($1.17 \pm 0.01\ M^{-1}s^{-1}$).
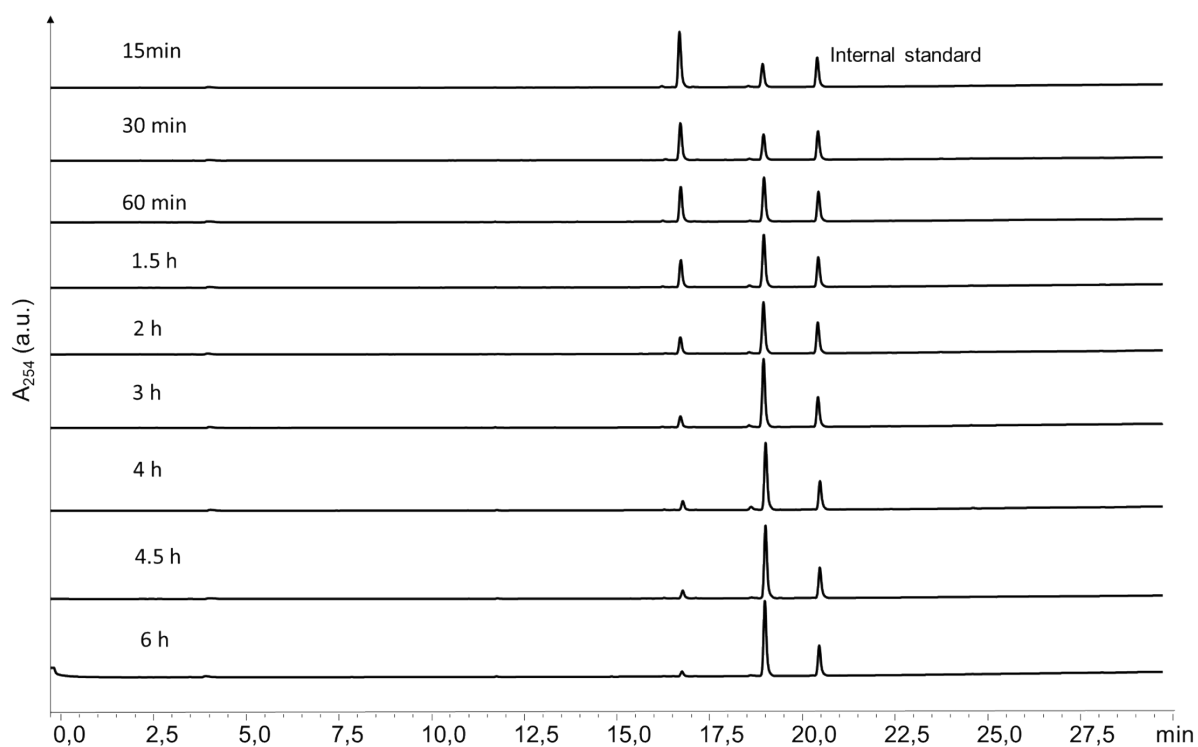


Fig. S17 HPLC trace of the reaction progress between compound **3** and Boc-Cys-OMe with Fmoc-Phe-OH as internal standard at different time points.
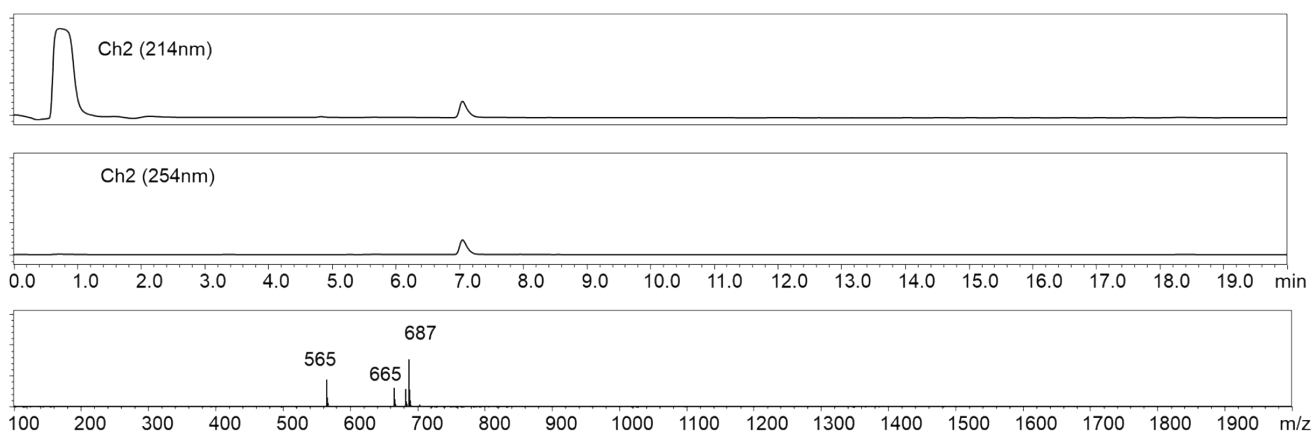


Fig. S18 LC-MS of compound **14**. (calculated: 664 [M+H]$^+$, found: 664 [M+H]$^+$. 687 [M+Na]$^+$, 565: Boc-deprotected product)

19
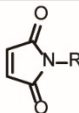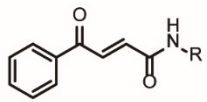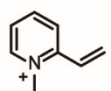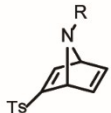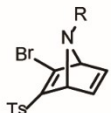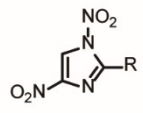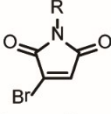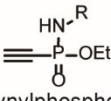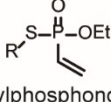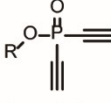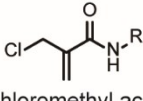
| Reagents | Rate constant ($M^{-1}s^{-1}$) | Reference |
|---|---|---|
|  Maleimide | 10 - 1000 | *ACS Chem. Biol.,* **2015**, *10*, 1026−1033 |
|  Carbonylacrylic reagents | 10.9 | *Nat. Comm.,* **2016**, *7*,13128 |
|  Quaternized vinyl pyridine | 200 | *Angew. Chem. Int. Ed.,* **2019**, *58*, 6640 −6644 |
|  Azabicyclic vinyl sulfones | Not reported | *Chem. Sci.,* **2019**, *10*, 4515–4522 |
|  Azanorbornadiene bromovinyl sulfones | Not reported | *Angew. Chem. Int. Ed.,* **2020**, *132*, 6255 −6259 |
|  1,4-Dinitroimidazoles | Not reported | *Nat. Commun.* **2019**, *10*, 142 |
|  3-Bromo-5-methylene | Not reported | *Nat. Commun.,* **2020**, *11*, 1015 |
|  Ethynylphosphoamidate | 0.62 | *Angew. Chem. Int. Ed.,* **2019**, *58*,11625 −11630 |
|  Vinylphosphonothiolate | 0.0021 | *J. Am. Chem. Soc.* **2020**, *142*, 9544–9552 |
|  Diethynyl phosphinate | 0.47 | *Angew. Chem. Int. Ed.,* **2021**, *60*, 15359 −15364 |
|  2-Chloromethyl acrylamide | 1.17 | This work |

Table S1 A brief summary of the reaction rate of representative cysteine modification strategies developed in the literature and in this work.

## 5. Peptide modification with 2-chloromethyl acrylamide compounds

20

## 5.1 Cysteine modification of WSC02 peptide
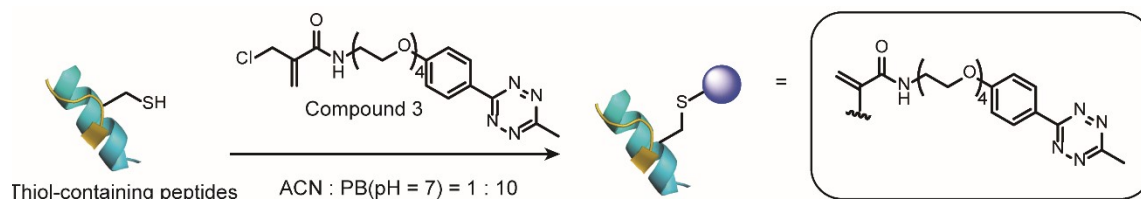


Scheme S5 Site-selective modification of different thiol-containing peptides with compound 3

WSC02 peptide (sequence: IVRWSKKVCQVS) was selected as model substrate for modification. WSCO2 peptide (1 mg, 1 eq, 0.71 μmol) was dissolved in 1 mL ACN: PB (pH = 7) = 1:10 mixture. Next, compound **3** (36 μg, 1.1 eq, 0.78 μmol, 20 mg/mL in DMF stock solution) was also added to the WSC02 peptide solution. The mixture was incubated at room temperature for 4 h. Next, 10 μL of mixture was diluted with 200 μL methanol and injected to LC-MS to check if any product was formed. The HPLC trace of the crude reaction mixture was shown in Fig. 4c showing excellent modification efficiency. Modified WSC02 peptide was purified by analytical HPLC and WSC02-PEG4-Tz was obtained as pink powder with the yield of 88%. The modified WSC02 peptide was also characterized with MALDI-Tof-MS with sinapinic acid as matrix showing the correct molecular weight at 1830 ([M+H]$^+$). The peak at 1761 in MALDI-Tof-MS spectra is the fragmentation product, which is shown below.



Fragmentation of WSC02-PEG4-Tz (1761)
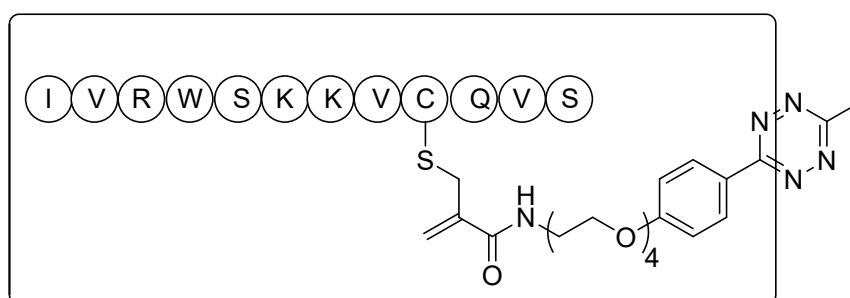
Fig. S19 The chemical structure of the fragmentation part (in the frame) of WSC02-PEG4-Tz in MALDI-Tof-MS spectra.

4,4'-Dithiodipyridine is a disulfide-containing compound which is often used for thiol quantification.[7] It

reacts with thiols quantitatively in a thiol-disulfide exchange reaction. The reaction mechanism is shown in Scheme S5.
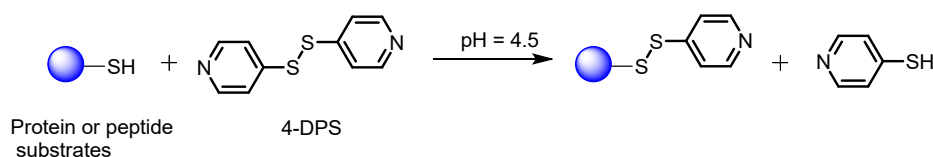


Scheme S6 Reaction mechanism of thiol-containing substrates with 4-DPS.

In order to check if the reaction is chemoselctive to thiol groups, the thiol group of native WSCO2 was firstly masked with 4-DPS. WSC02 peptide (100 µg, 1 eq, 0.071 µmol) was dissolved in 100 µL ACN : citrate buffer (pH = 4.5) = 1:10 mixture. 4-DPS (126 µg, 8 eq, 0.57 µmol) was added. The mixture was incubated at room temperature for 4 h. After that, the pH of the reaction mixture was adjusted to pH 7 and compound **3** (39.9 µg, 1.2 eq, 0.085 µmol) was added and the resultant mixture was incubated for another 4 hours. After that, 10 µL mixture was withdrawn and diluted with 200 µL methanol. LC-MS data indicated that there is no further reaction observed after the masking of the thiol groups (shown in Fig. 4c).
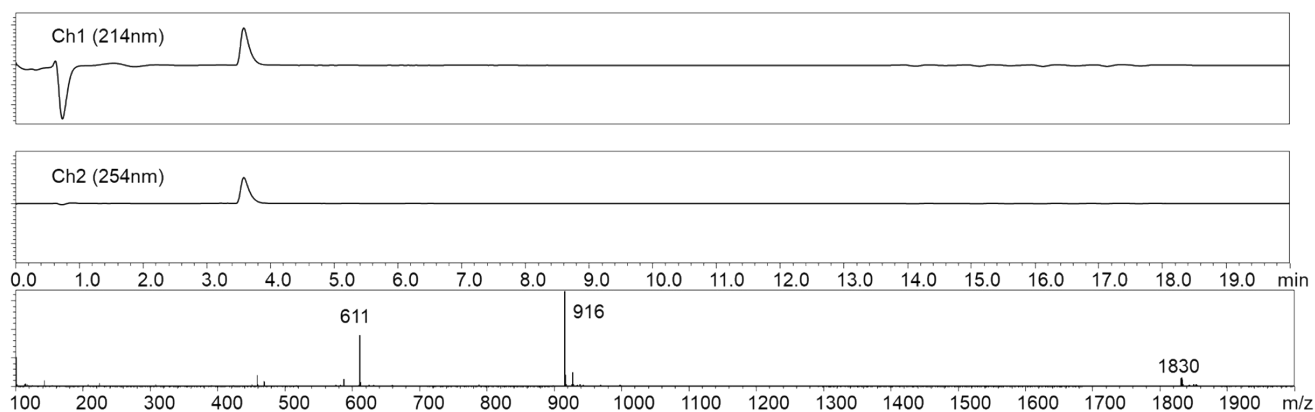


Fig. S20 LC-MS of the modified WSC02 peptide (WSC02-PEG4-Tz) (calculate: 1830 $[M+H]^+$, found: 1830 $[M+H]^+$. 916 $[M+2H]^{2+}$, 611 $[M+3H]^{3+}$)

5.2 Stability study of the tetrazine-modified WSC02 peptide (WSC02-PEG4-Tz).

It was reported that thiol-maleimide conjugates are prone to decompose via hydrolysis and/or retro-

Michael reaction.[6, 8] Therefore, the stability of WSC02-PEG4-Tz was evaluated at three different pH (pH 6, 7 and 8). As shown below, there is no obvious decomposition of WSC02-PEG4-Tz in the tested pH for 36 hours.
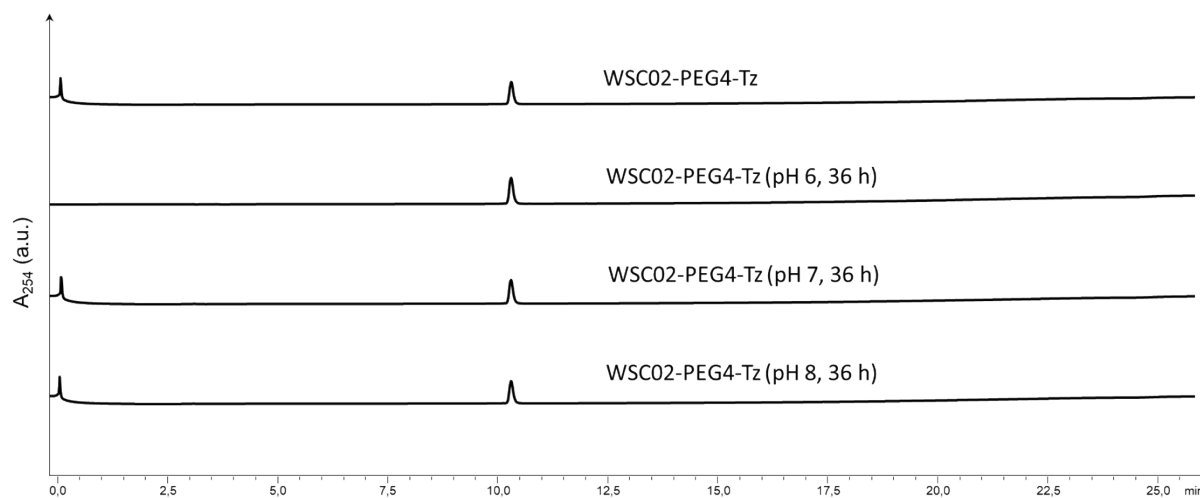


Fig. S21 Stability study of WSC02-PEG4-Tz at three different pH (pH 6, 7 and 8) for 36 hours.
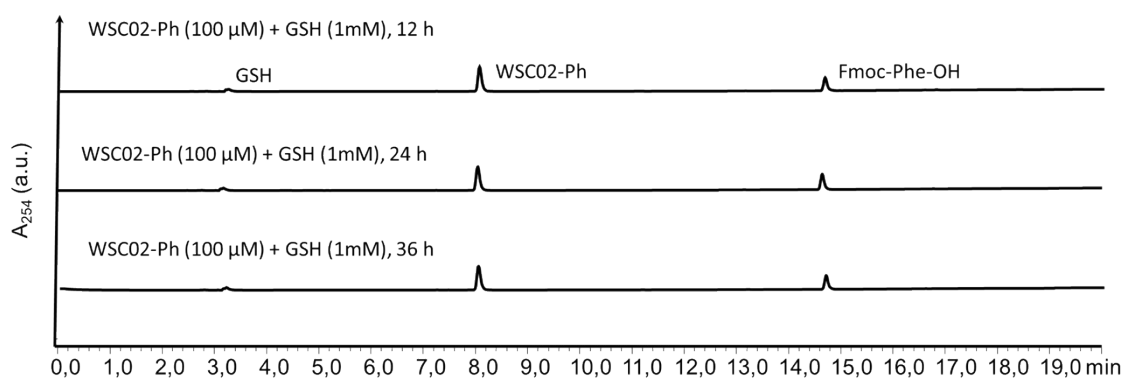


Fig. S22 Stability study of WSC02-Ph (100 µM) under 1 mM GSH for different time (12 h, 24 h, and 36 h).
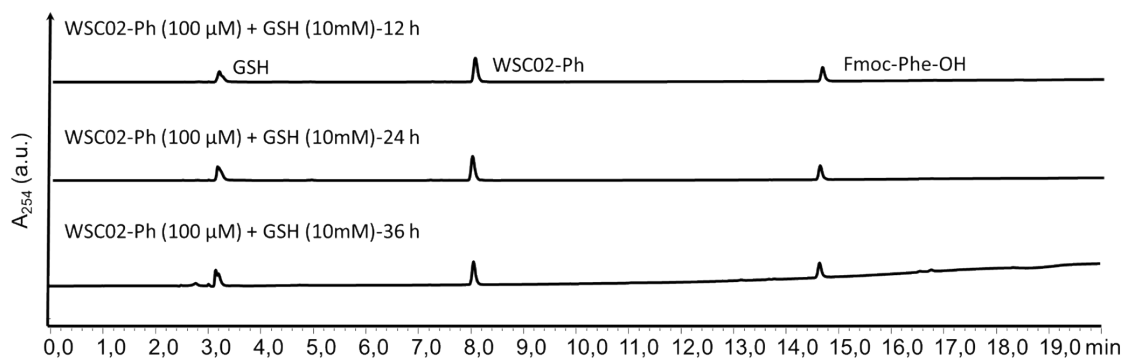


Fig. S23 Stability study of WSC02-Ph (100 µM) under 10 mM GSH for different time (12 h, 24 h, and 36 h).

23

## 5.3 Cysteine modification of other thiol-containing peptides

Besides WSC02 peptides, five other peptide substrates (RGDC, CEIE, PC-8, Tet and EK1 peptide) were also successfully modified following the similar procedures described above. Briefly, the peptide substrate (1 eq) was dissolved in ACN : PB (pH = 7) mixture with a concentration of 1 mg/mL. Compound **3** (1.2 eq, 20 mg/mL in DMF stock solution) was added and the mixture was incubated at room temperature for 4 hours. Next, 10 μL of mixture was took and diluted with 200 μL methanol. The modified peptide conjugates were purified by analytical HPLC with using the general methods described above and the isolated yield for each conjugate are 85% (RGDC), 84% (CEIE), 78% (PC-8), 83% (Tet peptide) and 77% (EK1 peptide). The LC-MS of the modified peptide conjugate was shown below.
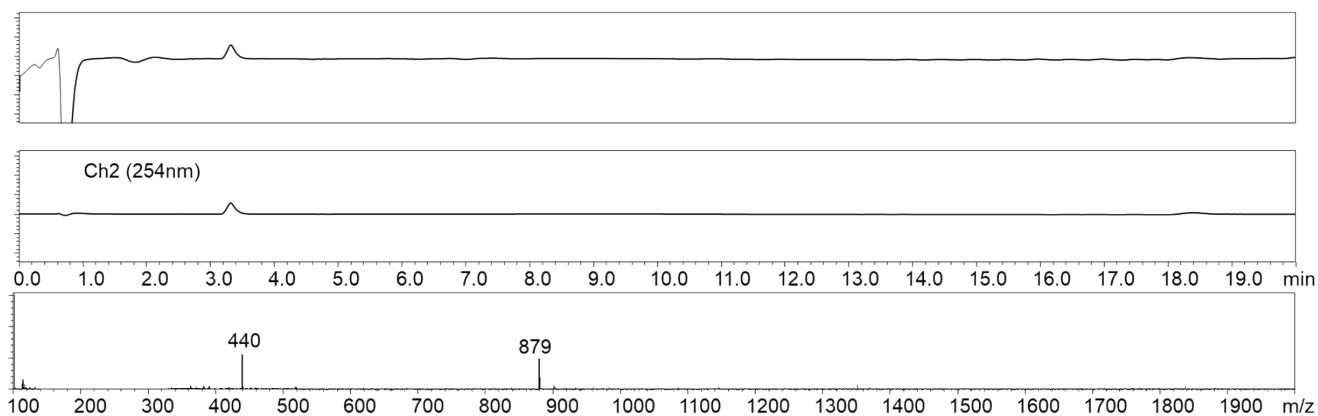


Fig. S24 LC-MS of modified peptide RGDC (calculated: 440 [M+2H]$^{2+}$, 879 [M+H]$^{+}$, found: 879 [M+H]$^{+}$, 440 [M+2H]$^{2+}$)

24
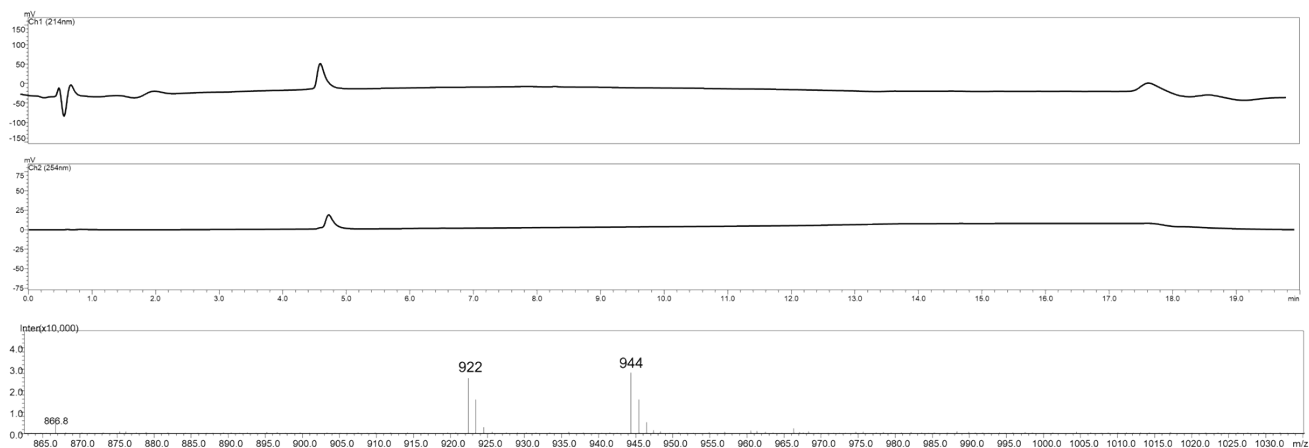
Fig. S25 LC-MS of modified peptide CEIE (calculated: 922 [M+H]$^+$, 944 [M+2H]$^{2+}$, found: 922 [M+H]$^+$, 944 [M+2H]$^{2+}$)
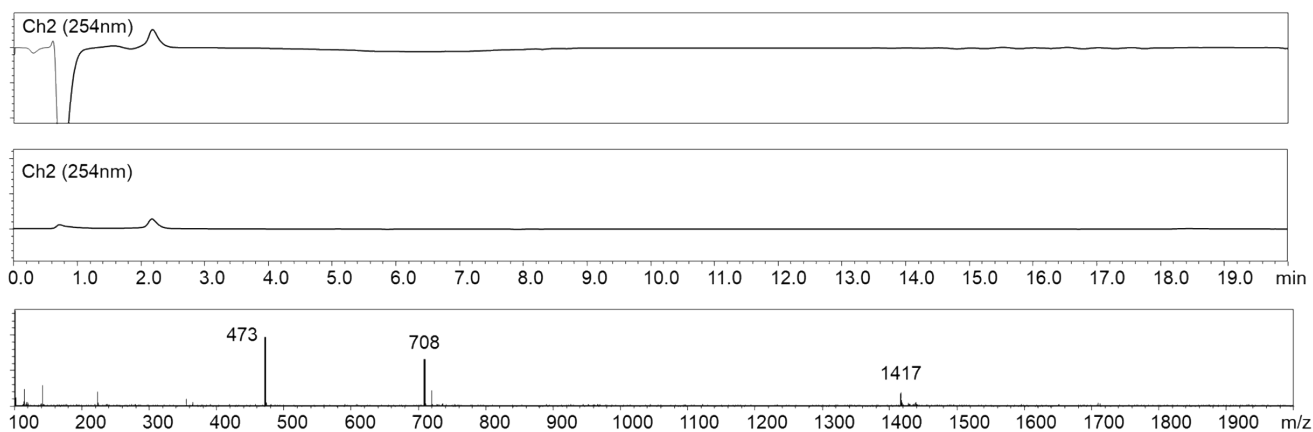


Fig. S26 LC-MS of modified peptide PC-8 (calculated: 473 [M+3H]$^{2+}$, 708 [M+2H]$^{2+}$, 1417 [M+H]$^+$, found:473 [M+3H]$^{2+}$, 708 [M+2H]$^{2+}$, 1417 [M+H]$^+$)
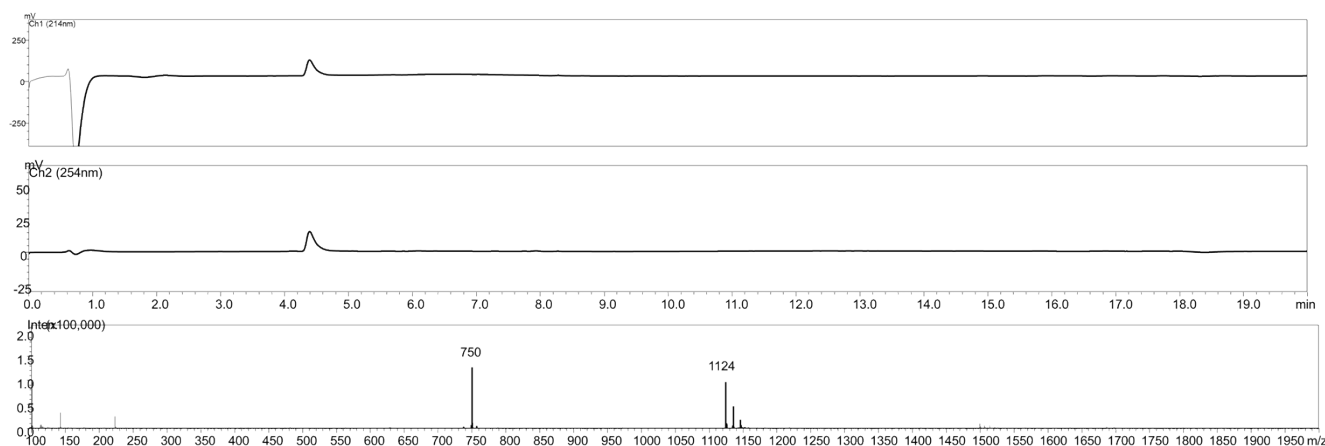


Fig. S27 LC-MS of modified Tet peptide (calculated: 750 [M+3H]$^{3+}$, 1124 [M+2H]$^{2+}$, found: 750 [M+3H]$^{3+}$,

25

1124 [M+2H]$^{2+}$)
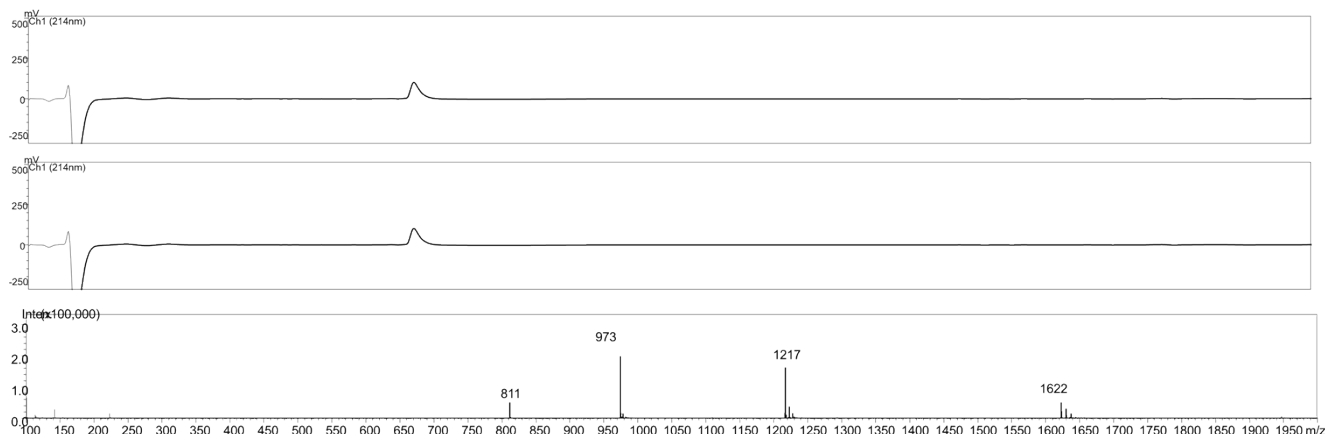


Fig. S28 LC-MS of modified EK1 peptide (calculated: 811[M+6H]$^{6+}$, 973 [M+5H]$^{5+}$, 1217 [M+4H]$^{4+}$, 1622 [M+3H]$^{3+}$, found: 811[M+6H]$^{6+}$, 973 [M+5H]$^{5+}$, 1217 [M+4H]$^{4+}$, 1622 [M+3H]$^{3+}$)

## 6. Ubiquitin modification with 2-chloromethyl acrylamide compounds

6.1 Cysteine modification of ubiquitin with 2-chloromethyl acrylamide compounds

(a).
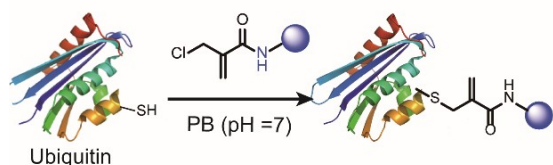
(b).
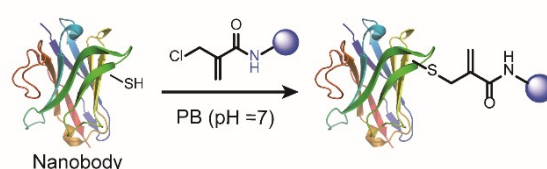


Scheme S7 Site-selective modification of (a) ubiquitin and (b) anti-GFP nanobody with 2-chloromethyl acrylamide compounds at cysteine site

Ubiquitin-K63C (Ub-K63C), in which lysine 63 has been mutated to a cysteine residue, was selected as model substrate to evaluate the modification efficiency on the protein level. Ub-K63C was purchased from the company UbiQ as lyophilized powder. The sequence of Ub-K63C: MQIFVKTLTGKTITLEVEPSDTIENVKAKI QDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQ**C**ESTLHLVLRLRGG. Its molecular weight is 8540 Da which was confirmed by MALDI-Tof-MS.

The general procedure to modify the ubiquitin substrate is as follows. First, ubiquitin was dissolved in

26

DMSO to prepare the stock solution with the concentration of 40 mg/mL. To an Eppendorf containing 8 µL

PB (pH = 7) buffer, 8 µg Ub-K63C (1 eq) was added and the resulting mixture was vortexed for 20 seconds.

Tris(2-carboxyethyl)phosphine (TCEP) (2eq, dissolved in PB buffer) was added to the ubiquitin solution and

incubated for 30 minutes. Afterwards, compound **1** (10 eq, 20 mg/mL stock solution in DMF) or compound **3**

(10 eq, 40 mg/mL stock solution in DMF) was added and the resulting mixture was incubated for another 4

hours at room temperature. Next, 5 µL aliquot was analyzed by MALDI-Tof-MS. The peak for Ub-K63C totally

disappeared in MALDI-Tof-MS spectra while the product peaks (observed MS 8713 for Ub-Ph and 8970 for

Ub-PEG4-Tz) were observed indicating the very excellent modification efficiency.

Similarly, if Ub-K63C reacted with 4-DPS first to mask the free cysteine residue on the protein surface,

no further reaction was observed when adding compound **3** into it using the same conditions described above
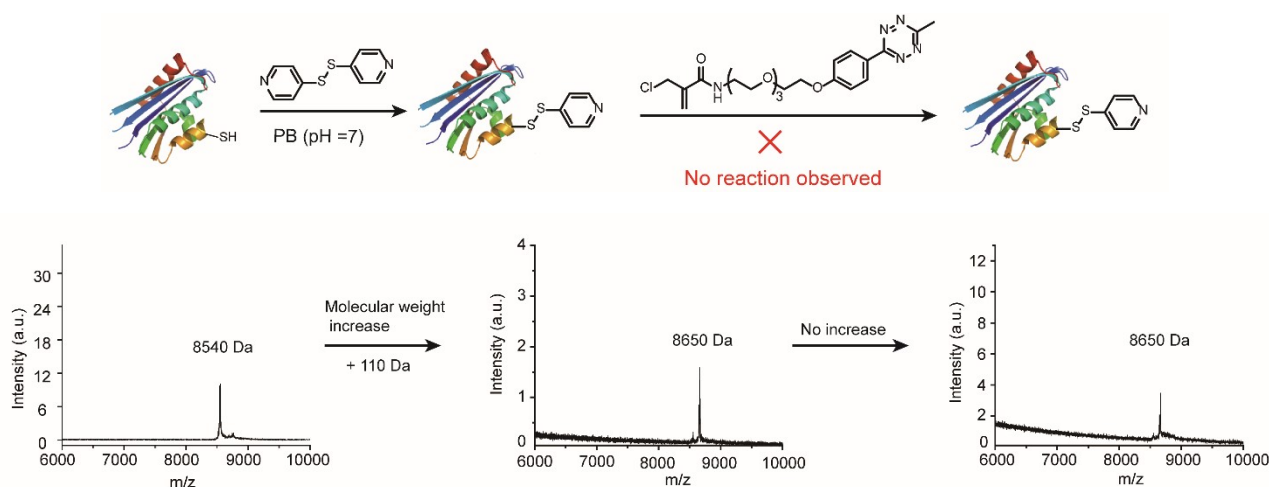


Fig. S29 Modification of ubiquitin with 4-DPS first to mask the accessible cysteine residue which followed the

addition of compound **3**.

Nanobodies are small, antigen-binding, single-domain polypeptides derived from the variable heavy

chain, which are considered as potent alternatives to conventional antibodies with enhanced stability and

reduced size but similar antigen-binding characteristics.[9-10] The anti-GFP nanobody used in this study was

purchased from NanoTag Biotechnologies, which contains a solvent accessible mutated cysteine residue.

The sequence of anti-GFP nanobody: MADVQLVESGGALVQPGGSLRLSCAASGFPVNRYSMRWYRQAP

GKEREWAVGMSSAGDRSSYEDSVKGRFTISRDDARNTVYLQMNSLKPEDTAVYYCNVNVGFEYWGQGT

QVTVSSADPNSSSVDKLACALEHHHHHH. The modification of anti-GFP nanobody at cysteine site with 2-

chloromethyl acrylamide compounds was following the same procedures described for ubiquitin. After the

reaction, the reaction mixture was characterized by MALDI-Tof-MS as sinapinic acid as matrix. The results

are shown in Fig. 5b of the main text.

6.3 CD spectra of ubiquitin and Ub-PEG4-Tz
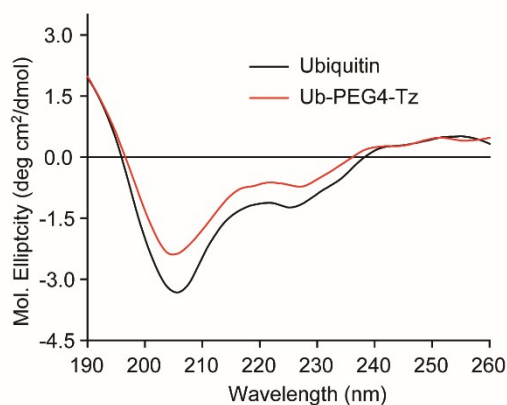


Fig. S30 CD spectra of ubiquitin and Ub-PEG4-Tz (0.1 mg/mL in PB (pH 7.4)).


## 7. Disulfide modification of somatostatin and Octreotide

7.1 Disulfide modification of somatostatin
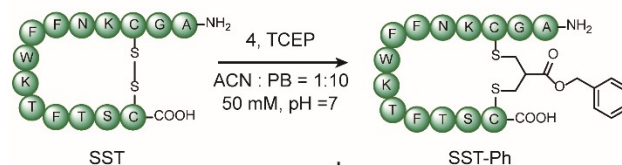


Scheme S8 Disulfide modification of SST with compound **4**

Somatostatin (SST, 1 mg, 1 eq, 0.61 μmol) was dissolved in 1 mL ACN / PB (50 mM, pH = 7) (v/v = 1:10).

TCEP (0.35 mg, 2eq, 1.22 μmol) was dissolved in 10 μL ACN/PB (50 mM, pH = 7) (v/v = 1:10) and added to

the SST solution. The mixture was incubated at room temperature for 30 minutes. Thereafter, compound **4**

(0.14 mg, 1.1 eq, 0.67 µmol, 40 mg/mL in DMSO stock solution) was also added in one-pot and the resulting mixture was gently shaken at room temperature overnight. 10 µL crude mixture was diluted with 200 µL methanol and injected to LC-MS directly to check the modification efficiency. The rest of the mixture was purified with analytical HPLC by using the general methods described above.

Next, 4-DPS was used to investigate if there are any free thiol groups in the modified SST-Ph. If the modification did not take place at the disulfide site, after the reduction of TCEP, the free thiol groups would still be able to react with 4-DPS, leading to a clearly detectable molecular weight increase. Consequently, purified SST-Ph (94 µg, 1 eq, 0.052 µmol) was dissolved in 94 µL citrate buffer and TCEP (59 µg, 4 eq, 0.20 µmol) was added. The mixed solution was incubated at room temperature for 60 minutes to ensure that TCEP reduced the disulfide bond if they are existing in the SST-Ph compound. Thereafter, 4-DPS (57 µg, 5 eq, 0.26 µmol) was added and the mixture was incubated for another 4 h. 10 µL of the obtained mixture was diluted with 200 µL methanol and injected to LC-MS to check if there is any further reaction between SST-Ph and 4-DPS. From the LC-MS data, no reaction was observed as the retention time of the SST-Ph was not changed and the molecular weight remained the same. As a control, native SST (94 µg, 1 eq, 0.057 µmol) was also incubated with TCEP (59 µg, 4 eq, 0.20 µmol), followed by the addition of 4-DPS (57 µg, 5 eq, 0.26 µmol). HPLC data clearly showed the occurrence of a new peak with a different retention time (Fig. S32).
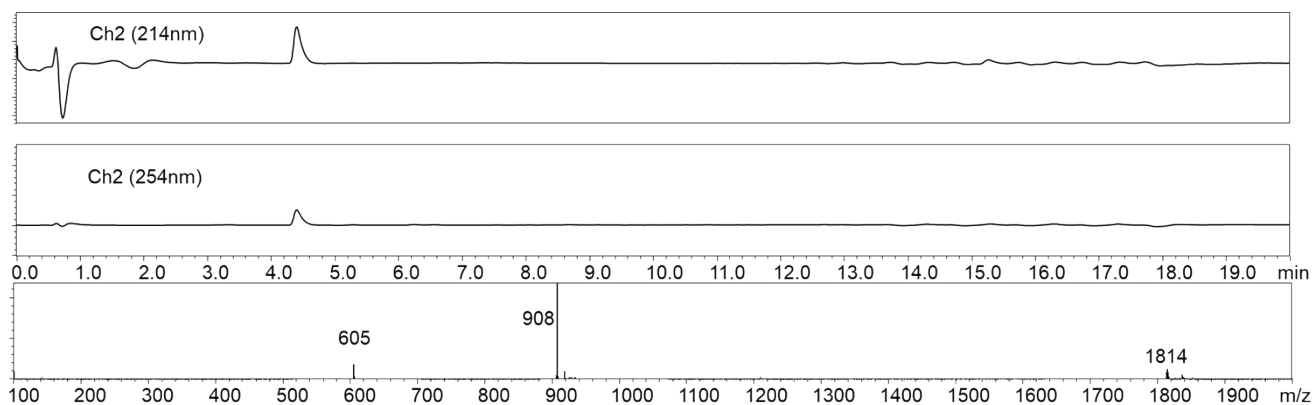


Fig S31 LC-MS of the purified SST-Ph (calculated: 604 [M+3H]$^{3+}$, 908 [M+2H]$^{2+}$, 1814 [M+H]$^{+}$, found: 605 [M+3H]$^{3+}$, 908 [M+2H]$^{2+}$, 1814 [M+H]$^{+}$)
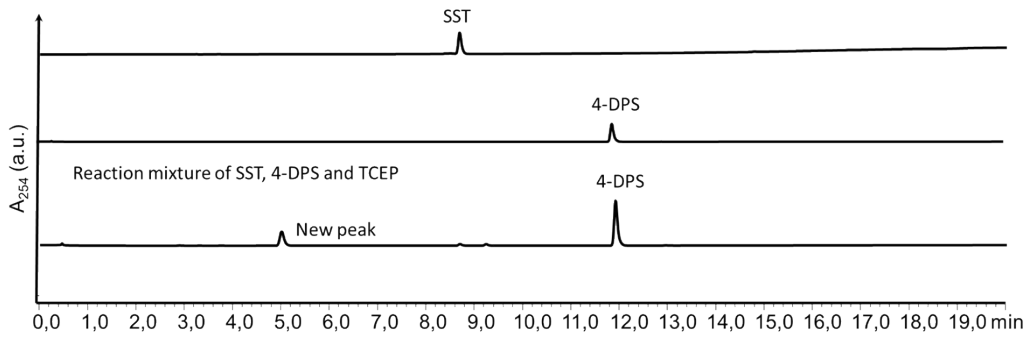
Fig. S32 HPLC analysis of the reaction between native somatostatin (SST) that contains a disulfide bond and 4-DPS. A new peak was formed after the reaction of SST with 4-DPS. However, the peak of SST-Ph, which was functionalized at the disulfide site, remained intact in the presence of 4-DPS indicating that the modification took place at the disulfide site.

The stability of the disulfide modified SST was also evaluated in pH 6, 7, and 8 as well as against GSH, which is depicted below.
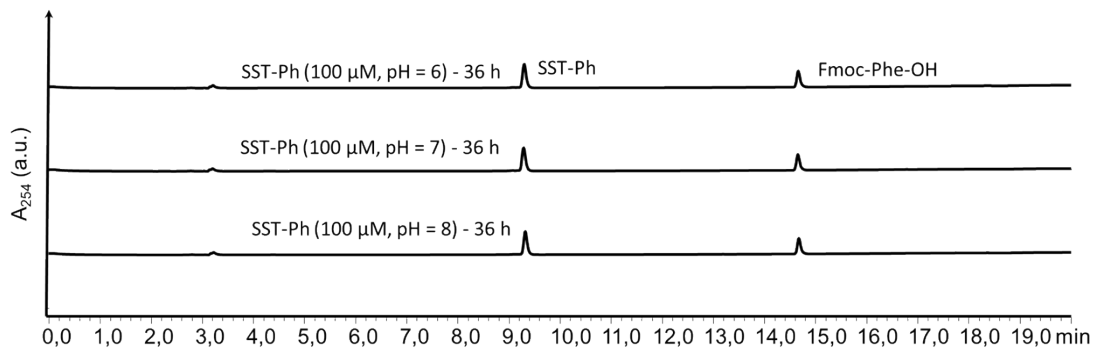


Fig. S33. Stability study of SST-Ph (100 µM) under three different pH for 36 h.
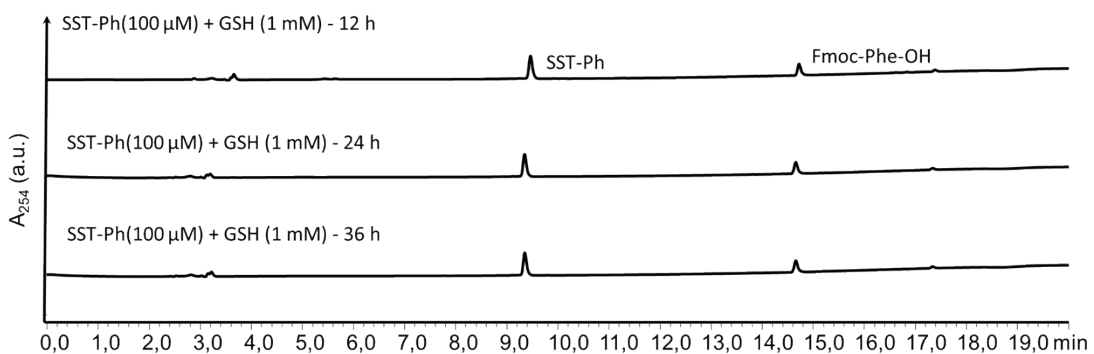


Fig. S34 Stability study of SST-Ph (100 µM) under 1 mM GSH for different time (12 h, 24 h, and 36 h).

30
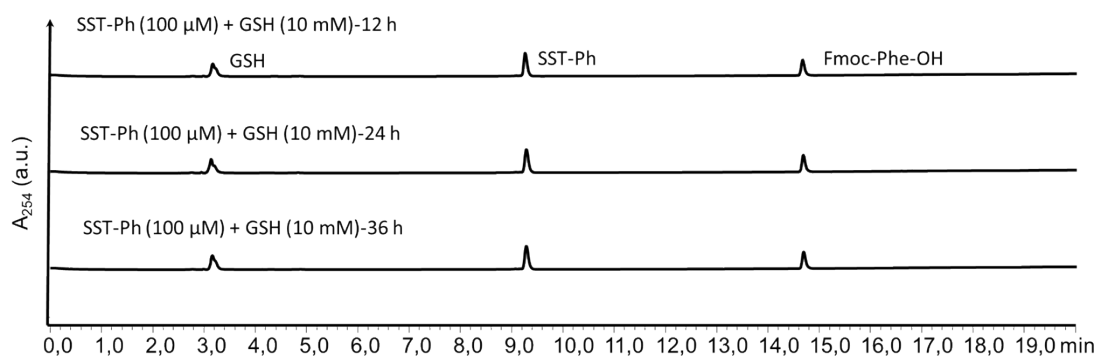
Fig. S35 Stability study of SST-Ph (100 μM) under 10 mM GSH for different time (12 h, 24 h, and 36 h)

CD spectra of native SST and SST-Ph
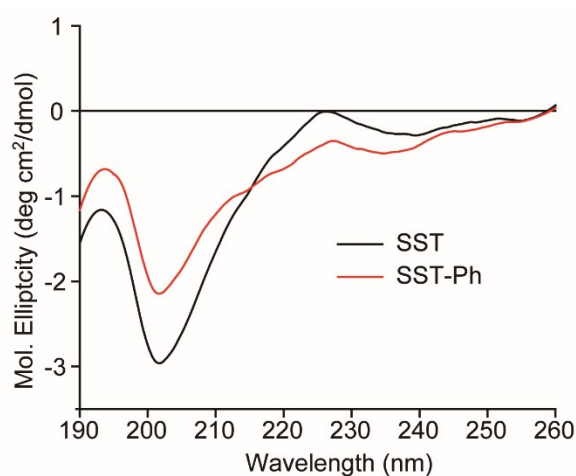


Fig. S36 CD spectra of native SST and modified one (SST-Ph)

7.2 Disulfide modification of somatostatin with compound **8** and IC-Tz

In order to compare the disulfide modification efficiency of the 2-chloromethyl acrylate compounds developed in this paper and the allyl sulfone reagents reported previously by our group, an allyl sulfone reagent (denoted as "IC-Tz"), which contained the same PEG chain length as compound **8**, was synthesized by following the protocol published before.[11] SST modification with compound **8** was following the procedure described above. For the modification with IC-Tz, SST (1 mg, 1 eq, 0.61 μmol) was dissolved in ACN: PB (pH =7.8) (v/v = 2:3). TCEP (0.35 mg, 2 eq,1.2 μmol) was added. The mixture was incubated at room temperature for 30 minutes. Afterwards, IC-Tz (0.66 mg, 2 eq, 1.2 μmol) was also added and the resulting mixture was gently shaken at room temperature overnight. 10 μL crude mixture was diluted with 200 μL methanol and

31

injected to HPLC directly to check the modification efficiency. The result is shown in Fig 6e of the main text.
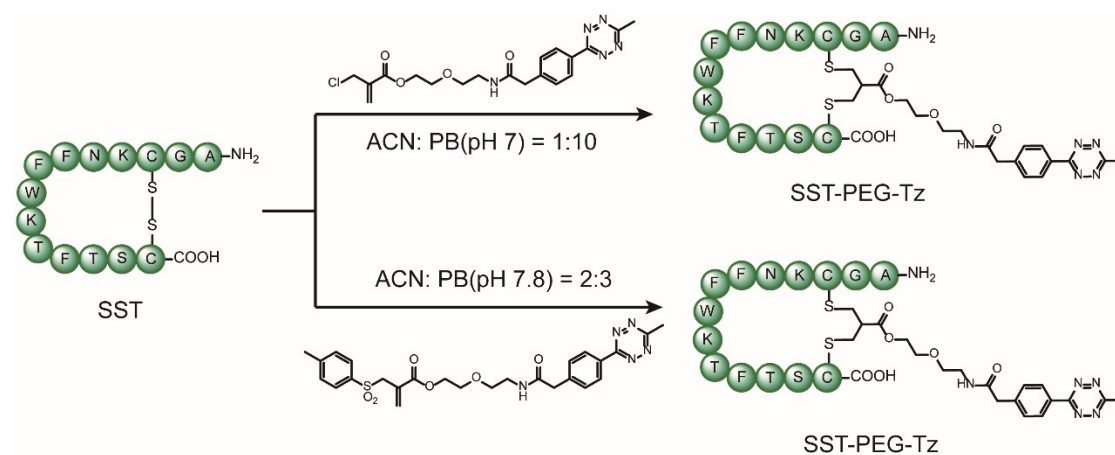


Scheme S9 Somatostatin modification with two different disulfide rebridging reagents.
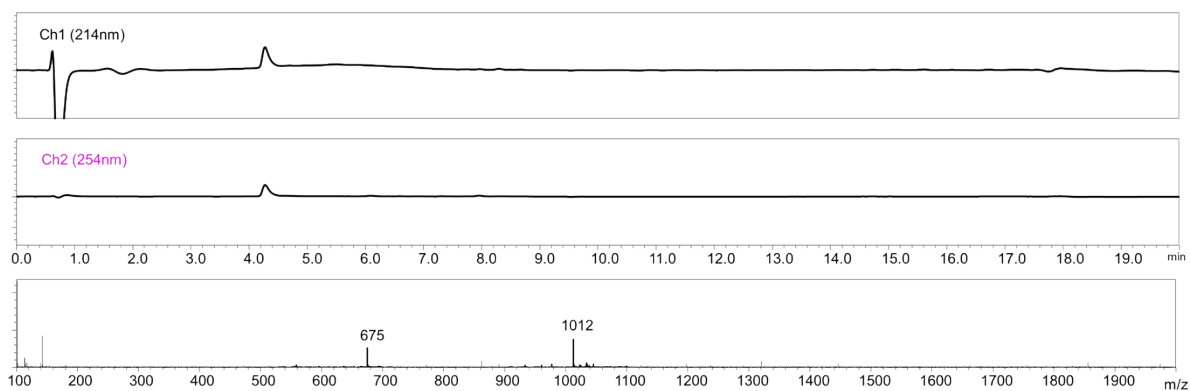


Fig. S37 LC-MS of SST-PEG-Tz (calculated: 1011 [M+2H]$^{2+}$, 674 [M+3H]$^{3+}$, found: 1012 [M+2H]$^{2+}$, 675 [M+3H]$^{3+}$

## 7.3 Trypsin digestion of SST-PEG-Tz

SST-PEG-Tz (0.83 µg/µL; 0.45 mM in water) (50 µL, 0.023 µmol) was added to 62.6 µL ammonium bicarbonate (200 mM, 16 mg/mL), followed by the addition of trypsin (0.1 µg/µL) (10.4 µL, 0.56 nmol). The resultant mixture was incubated for 2 h at 37 °C. Samples for LC-MS analysis were diluted 4 times in ammonium acetate solution (20 mM, pH 7.0).

Liquid chromatography–mass spectrometry (LC-MS) measurements were conducted using a Dionex Ultimate 3000 UHPLC+ system equipped with a Multiple-Wavelength detector, an imChem Surf C18 TriF 100 Å 3 µm 100 x 2.1mm column connected to Thermo Scientific Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Scientific™ Q Exactive™ Plus). Tryptic peptides were separated with a 0.2 mL/min and mixture of water with 0.1% formic acid (buffer A) acetonitrile with 0.1% formic acid (buffer B), according to the following table:

| Time (min) | A (%) | B (%) | curve |
|---|---|---|---|
| 0-5 | 97 → 97 | 3 → 3 | 5 |
| 5-75 | 97 → 5 | 5 → 95 | 7 |
| 75-85 | 5 → 5 | 95 → 95 | 5 |
| 85-88 | 5 → 97 | 95 → 3 | 5 |
| 88-90 | 97 → 97 | 3 → 3 | 3 |

The exactive mass spectrometer was operated in the positive ion mode with alternating MS scans of the precursor ions and AIF (all ion fragmentation) scans in which the peptides were fragmented by SID (Surface-Induced Dissociation) of 50 eV. Both scan types were performed with 70,000 resolution (at $m/z$ 200) with each scan taking 0.05 s, and the maximum fill time was set to 1 s as well. The $m/z$ range for the MS scans was selected between 300–1600, and the $m/z$ range for AIF scans was chosen between 150–1600. The target value for the MS scans was $10^6$ ions, and the target value for the AIF scans was 3 x $10^6$ ions.

In order to confirm the structure of the double modification, trypsin digestion was performed. Taking in consideration that trypsin will cut the peptide at Lys (K) and Arg (R), the extracted ion chromatogram (EIC) within a 6 ppm range charged by 2-3 H⁺ was searched for the fragments (Figures S36-44). A fragment was observed with single modification at the C-terminal Cys but was determined to occur via a retro-Michael reaction during the digestion process (data not shown).

33
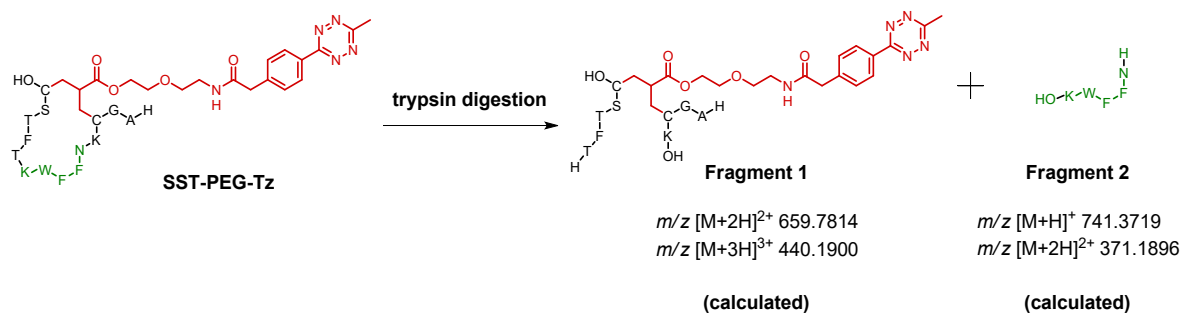
Scheme S10 Fragment 1 and Fragment 2 formed from trypsin digestion of SST-PEG-Tz



Fig. 38 ESI-MS of fragment 1

34

b₁ b₂ b₃ labels, peptide structure
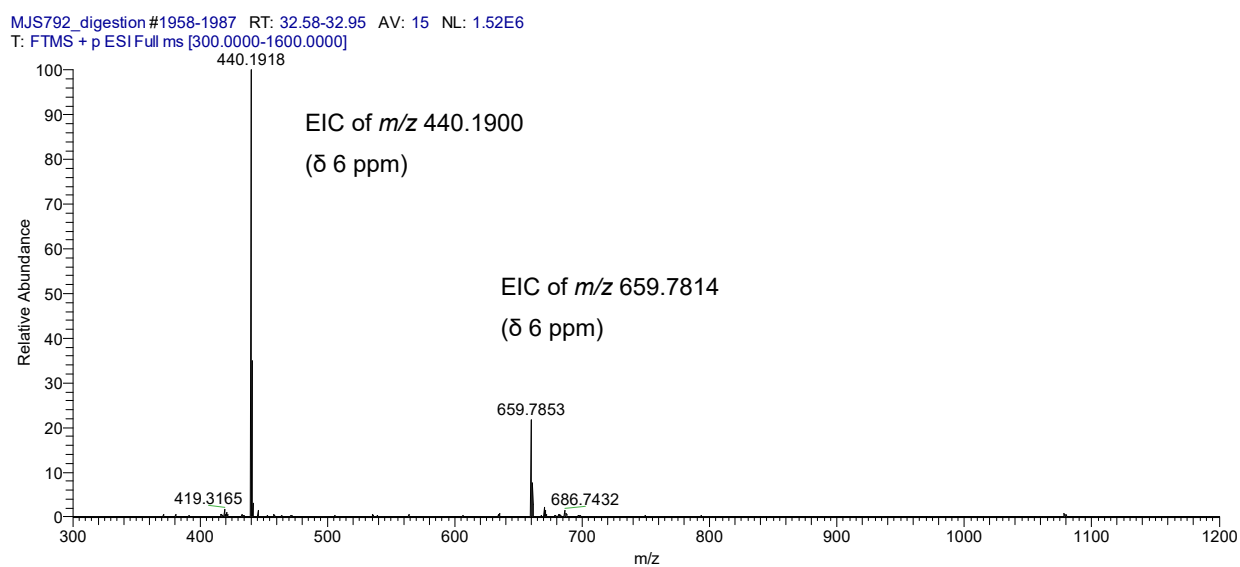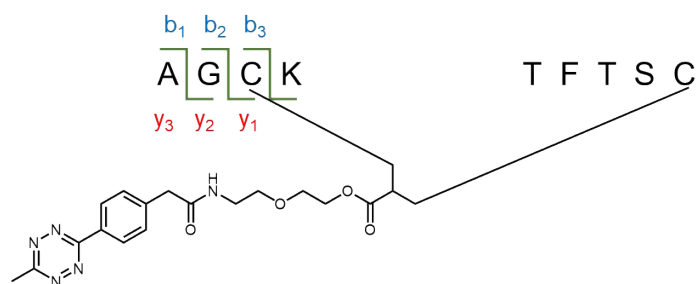


| Lost fragment | #b | b⁺ | b²⁺ | Seq | y⁺ | y²⁺ | #y | Lost fragment |
|---|---|---|---|---|---|---|---|---|
| A | 1 | 72.0444 | 36.02222 | A | 1318.5555 | 659.78135 | 3 | - |
| AG | 2 | 129.0659 | 64.53295 | G | 1247.51826 | 623.75913 | 2 | A |
| - | 3* | - | - | C | 1190.49680 | 595.2484 | 1* | AG |

Table S2 Expected b and y ions from AIF (ions that are found are highlighted in red)
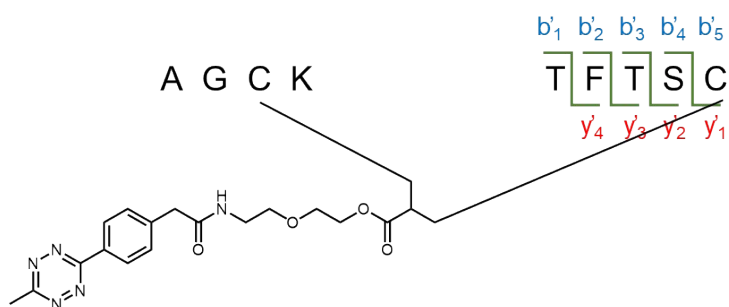


| Lost fragment | #b' | b⁺ | b²⁺ | Seq | y⁺ | y²⁺ | #y' | Lost fragment |
|---|---|---|---|---|---|---|---|---|
| H₂O | 5* | 1301.5521 | 650.7761 | T | 18.01056 | 9.00528 | 1* | H₂O |
| T | 4 | 1217.5077 | 608.7539 | F | 102.055 | 51.0275 | 2 | T |
| TF | 3 | 1070.4393 | 535.2196 | T | 249.12341 | 124.561705 | 3 | TF |
| TFT | 2 | 969.3916 | 484.6958 | S | 350.17109 | 175.085545 | 4 | TFT |

Table S3 - Expected b and y ions from AIF (ions that are found are highlighted blue and red)

b''₁ b''₂ b''₃ → $b''_1$ $b''_2$ $b''_3$

A G C K       T F T S C

$y''_3$ $y''_2$ $y''_1$

PEG-Tz
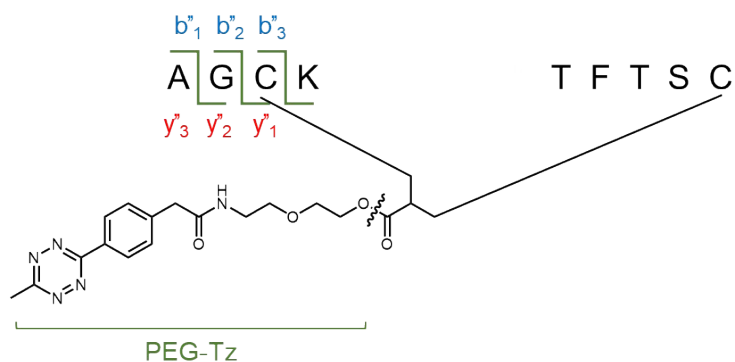
| Lost fragment | #b'' | b⁺ | b²⁺ | Seq | y⁺ | y²⁺ | #y'' | Lost fragment |
|---|---|---|---|---|---|---|---|---|
| A | 1 | 72.0444 | 36.02222 | A | 930.3696 | 465.1848 | 3 | A |
| AG | 2 | 129.0659 | 64.53295 | G | 873.3481 | 436.6741 | 2 | AG |

Table S4. Expected b and y ions from AIF (ions that are found are highlighted in red)

$b'''_1$ $b'''_2$ $b'''_3$ $b'''_4$ $b'''_5$

A G C K       T F T S C

$y'''_5$ $y'''_4$ $y'''_3$ $y'''_2$ $y'''_1$

PEG-Tz
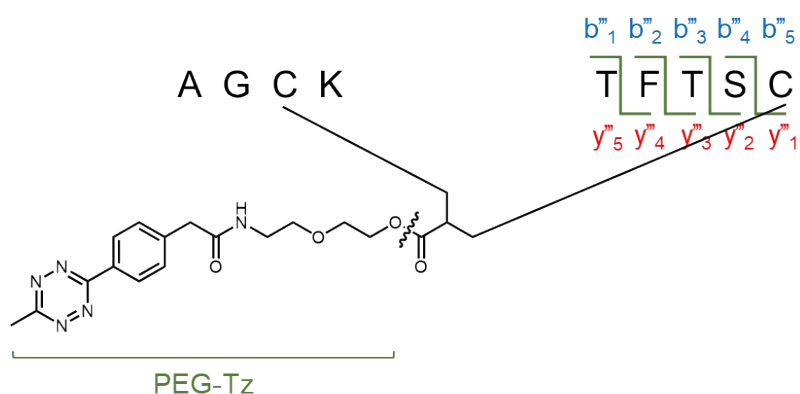
| Lost fragment | #b''' | b⁺ | b²⁺ | Seq | y⁺ | y²⁺ | #y''' | Lost fragment |
|---|---|---|---|---|---|---|---|---|
| H₂O | 5* | 18.01056 | 9.00528 | T | 984.4034 | 492.2017 | 1* | H₂O |
| T | 4 | 102.0550 | 51.0275 | F | 900.3590 | 450.1795 | 2 | T |
| TF | 3 | 249.1234 | 124.5617 | T | 753.2906 | 376.6453 | 3 | TF |
| TFT | 2 | 350.1711 | 175.0855 | S | 652.2429 | 326.1215 | 4 | TFT |

Table S5. Expected b and y ions from AIF (ions that are found are highlighted in red and blue)
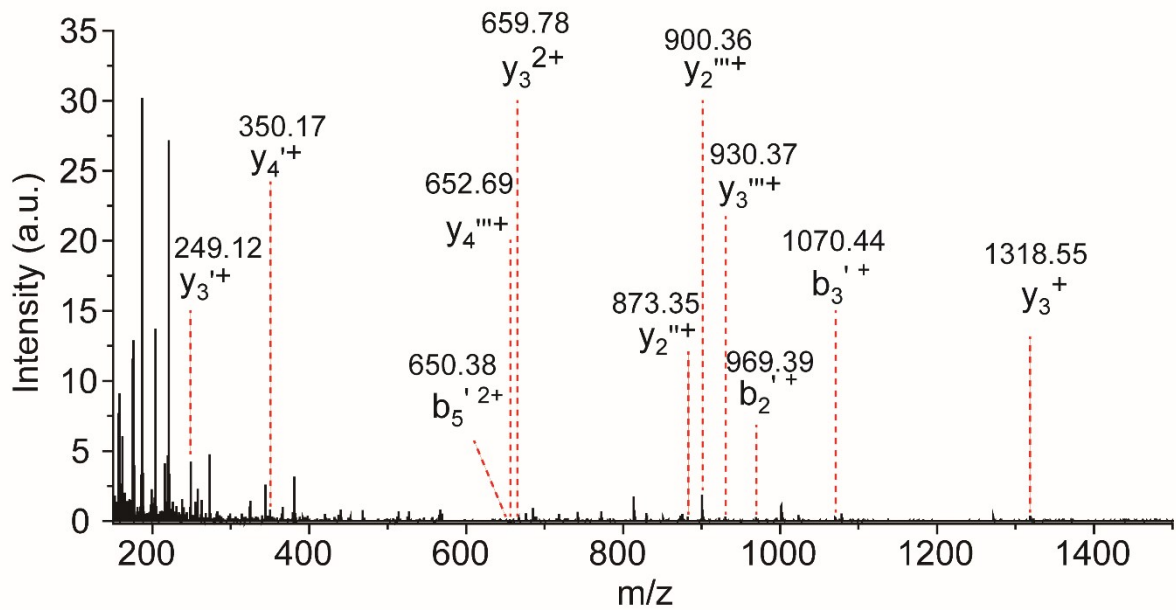
Fig. 39 Full spectrum of fragment 1 using SID 50 eV

**Spectra zoom-in:**
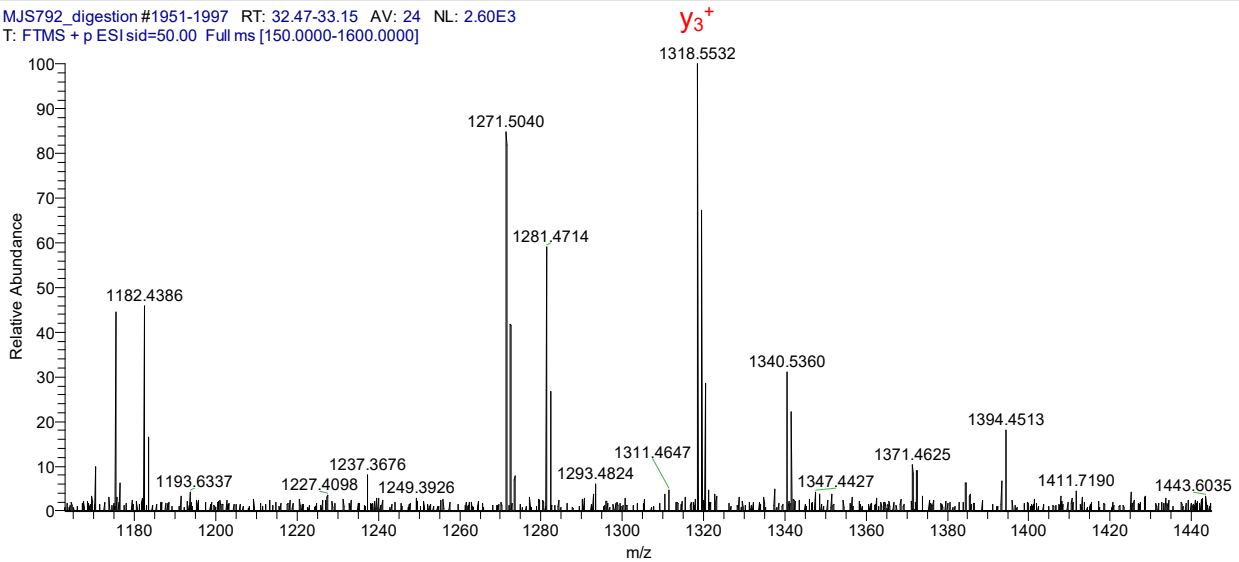


Fig S40. AIF spectrum of peak at 32.8 min using SID 50 eV between *m/z* 1150-1450 range.

37

MJS792_digestion #1951-1997  RT: 32.47-33.15  AV: 24  NL: 7.88E2
T: FTMS + p ESI sid=50.00  Full ms [150.0000-1600.0000]
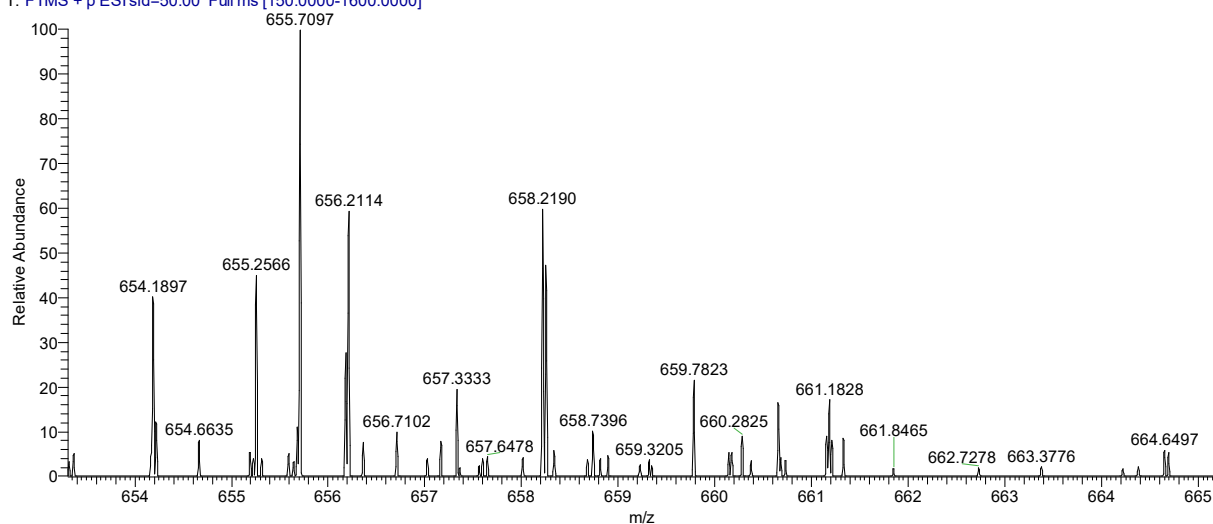


Fig S41. AIF spectrum of peak at 32.8 min using SID 50 eV between *m/z* 653-665 range.

MJS792_digestion #1951-1997  RT: 32.47-33.15  AV: 24  NL: 2.26E2
T: FTMS + p ESI sid=50.00  Full ms [150.0000-1600.0000]
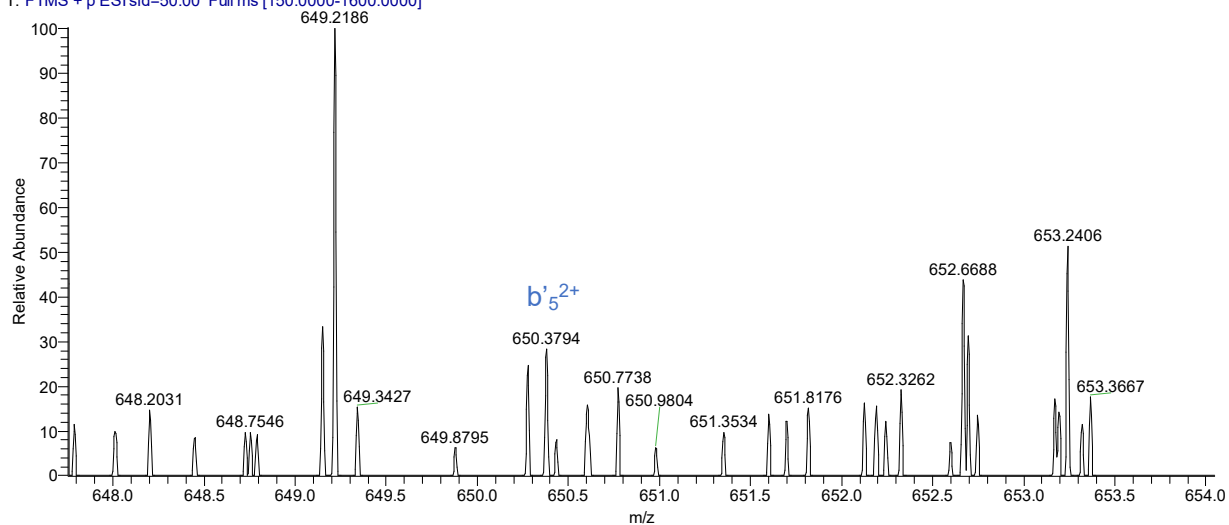


Fig S42. AIF spectrum of peak at 32.8 min using SID 50 eV between *m/z* 648-654 range.

38

MJS792_digestion #1951-1997  RT: 32.47-33.15  AV: 24  NL: 5.07E3
T: FTMS + p ESI sid=50.00  Full ms [150.0000-1600.0000]
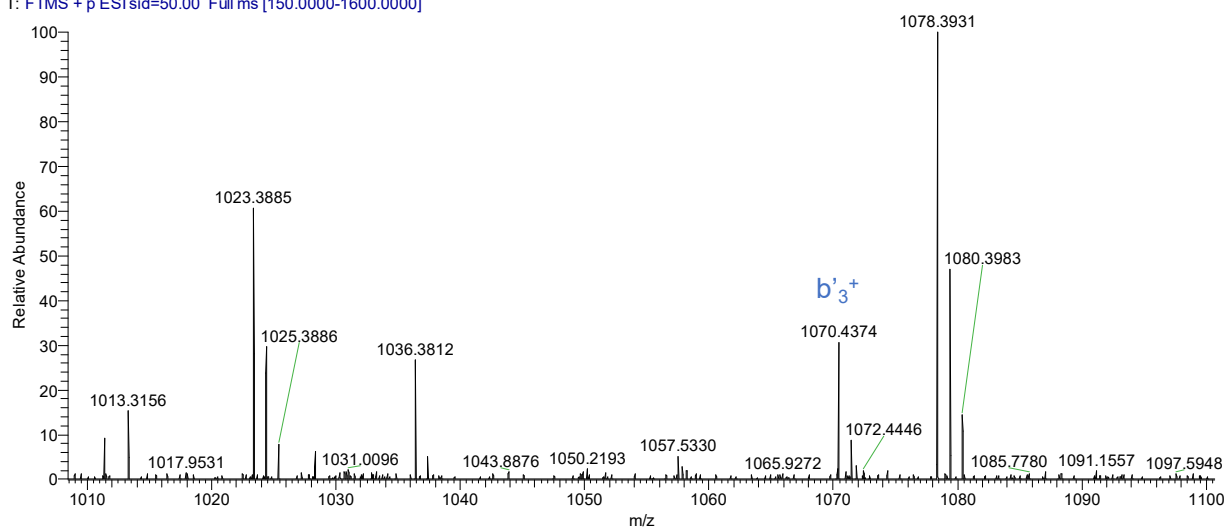


Fig S43. AIF spectrum of peak at 32.8 min using SID 50 eV between *m/z* 1010-1100 range.

MJS792_digestion #1951-1999  RT: 32.47-33.18  AV: 25  NL: 1.30E4
T: FTMS + p ESI sid=50.00  Full ms [150.0000-1600.0000]
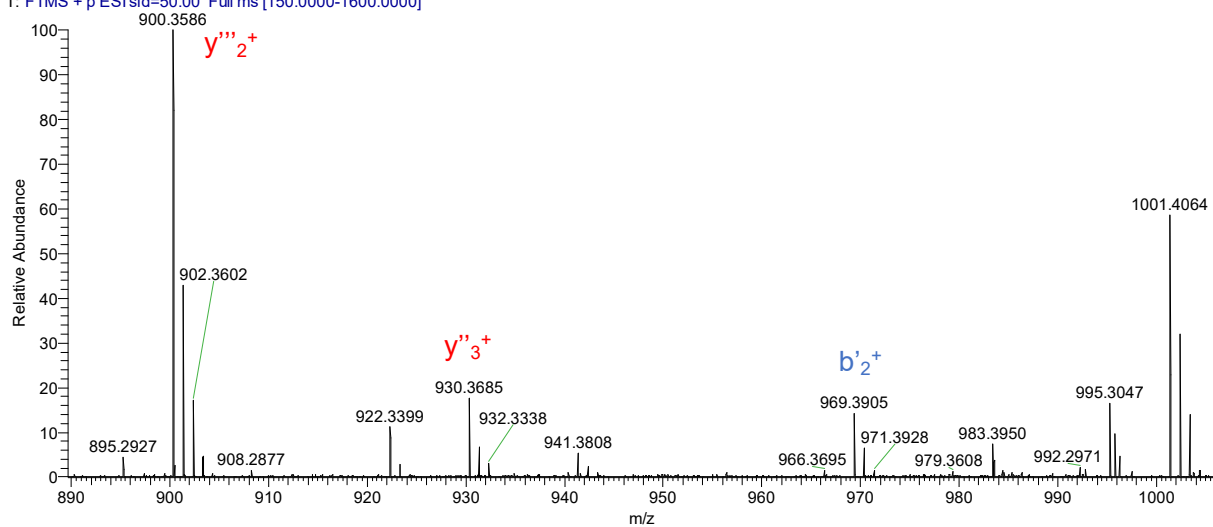


Fig S44. AIF spectrum of peak at 32.8 min using SID 50 eV between *m/z* 890-1005 range.

39

## 7.4 Disulfide modification of octreotide



Scheme S11 Disulfide modification of octreotide with compound **6**

The modification of octreotide followed the similar procedure as the modification with SST. Briefly, octreotide (200 µg, 1 eq, 0.20 µmol) was dissolved in ACN/PB (50 mM, pH = 7) (v/v = 1:10) with the concentration of 1 mg/mL. TCEP (112 µg, 2 eq, 0.40 µmol) was also added and the resulting mixture was incubated at room temperature for 30 minutes to fully break the disulfide bond. Afterwards, compound **6** (89 µg, 1.2 eq, 0.24 µmol) was added to the aforementioned mixture above and incubated for overnight. After that, the crude mixture was purified with analytical HPLC to get a while solid. MALDI-Tof-MS data demonstrated the successful modification
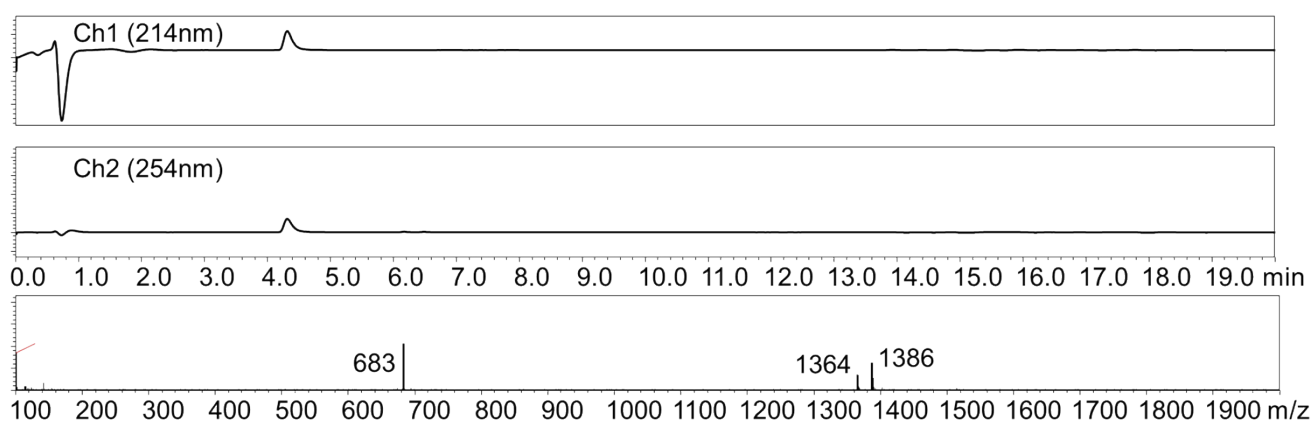


Fig. S45 LC-MS of modified octreotide with coumarin group (Oc-Coumarin) (calculated: 1364 [M+H]⁺, found: 1364 [M+H]⁺, 1386 [M+Na]⁺, 683 [M+2H]⁺)
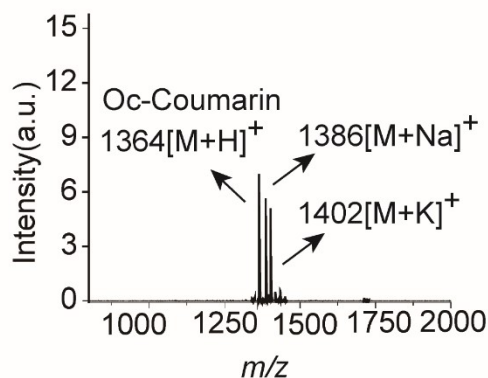


Fig. S46 MALDI-Tof-MS of Oc-Coumarin (calculated: 1364 [M+H]⁺, 1386 [M+Na]⁺, 1402 [M+K]⁺, found: 1364

[M+H]+, 1386 [M+Na]+, 1402 [M+K]+)

CD spectra of native octreotide and Oc-Cou
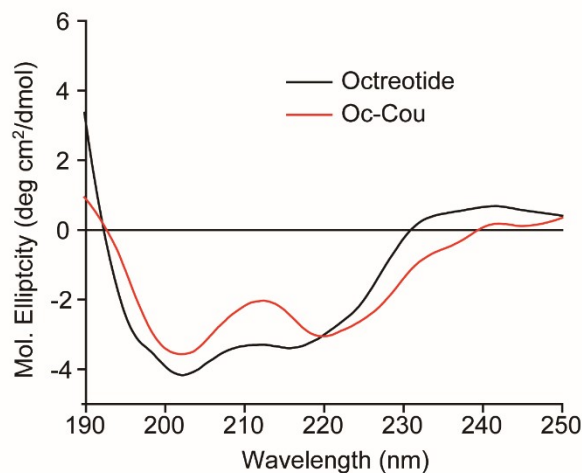


Fig. S47 CD spectra of native octreotide and Oc-Cou (0.1 mg/mL in PB (pH 7.4)).

## 8. Lysozyme modification

8.1 Disulfide modification of lysozyme
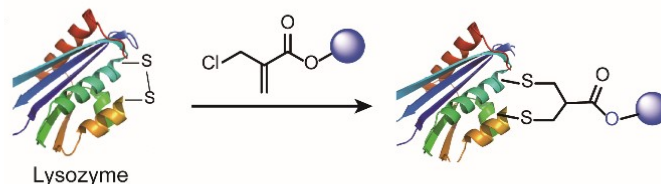


Scheme S12 Disulfide modification of lysozyme with 2-chloromethyl acrylate compounds

Lysozyme (500 µg, 1 eq, 0.035 µmol) was dissolved in 500 µL PB buffer (50 mM, pH = 7). TCEP (12 µg, 1.2 eq, 0.042 µmol) was dissolved in 10 µL PB buffer (50 mM, pH = 7) and added to the lysozyme solution. Subsequently, the respective 2-chloromethyl acrylate compounds, for example compound **8** stock solution (23 µg, 1.1 eq, 0.042 µmol) was added to the aforementioned lysozyme mixture in one-pot. The mixed solution was incubated for overnight at room temperature and purified with FPLC (ÄKTA from GE) by using Hi Trap phenyl HP column (1 mL, from GE) based on the different hydrophobicity of native and modified lysozyme. The collected fractions were concentrated with ultrafiltration tube and the solvent was removed with freeze dryer. The obtained modified lysozyme was characterized by MALDI-Tof-MS showing the correct MS.

## 8.2 Trypsin digestion of Ly-PEG-Tz

Lysozyme-PEG-Tz protein solution (9.6 µL, 100 µM) in ammonium acetate (20 mM, pH 7.0) was added to a solution of ammonium bicarbonate (14.4 µL, 16 mg/mL), followed by activated trypsin (7 µL, 0.1 µg/µL in ammonium bicarbonate (16 mg/mL) solution. The resultant mixture was incubated at 37 °C for 20 h. Thereafter, the reaction was quenched with formic acid (5 µL), vortexed briefly and then centrifuged.

## LC-MS method

Liquid chromatography–mass spectrometry (LC-MS) runs were realized using a Dionex Ultimate 3000 UHPLC+ system equipped with a Multiple-Wavelength detector, an imChem Surf C18 TriF 100 Å, 3 µm 100x2,1 mm column connected to Thermo Scientific Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Scientific™ Q Exactive™ Plus). Tryptic peptides were separated with a 0.2 mL/min and mixture of water with 0.1% formic acid (buffer A) acetonitrile with 0.1% formic acid (buffer B), according to the following table:

| Time (min) | A (%) | B (%) | curve |
|---|---|---|---|
| 0-5 | 97 → 97 | 3 → 3 | 5 |
| 5-75 | 97 → 5 | 5 → 95 | 7 |
| 75-85 | 5 → 5 | 95 → 95 | 5 |
| 85-88 | 5 → 97 | 95 → 3 | 5 |
| 88-90 | 97 → 97 | 3 → 3 | 3 |

Table S6 HPLC method used for the analysis of Trypsin digestion of Ly-PEG-Tz

The exactive mass spectrometer was operated in the positive ion mode with alternating MS scans of the precursor ions and AIF (all ion fragmentation) scans in which the peptides were fragmented by SID (Surface-Induced Dissociation) with 50 eV. Both scan types were performed with 70,000 resolution (at $m/z$ 200) with each scan taking 0.05 s, and the maximal fill time was set to 1 s as well. The $m/z$ range for MS scans was 300–1600, and the $m/z$ range for AIF scans was 150–1600. The target value for the MS scans was $10^6$ ions, and the target value for the AIF scans was 3 x $10^6$ ions.
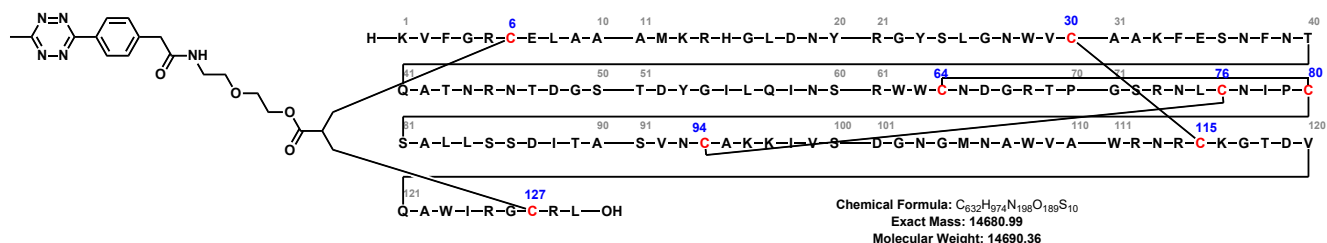


Fig. S48 Sequence profile and expected modification site of Ly-PEG-Tz.

## Possible combinations for PEG-Tz crosslink in Lysozyme protein

To validate the site-selectivity of PEG-Tz crosslink for C6 and C127 of Lysozyme C (P00698), all possible S-S rebridging positions were considered. Skyline software[12] was used to generate the list of Lysozyme digestion fragments from *in silico* Trypsin digestion.[12] The simulation of protein digestion also considered zero missed cleavages, no Methionine oxidation and all Cys in reduced form. All generated fragments were combined and added PEG-Tz crosslink as a structural modification variable on Cys residues. For the analysis, the transition list was used to check Extracted Ion chromatogram (EIC) within a 6 ppm range for all possible fragments that included precursor charges of 1 to 6, within 300-1600 *m/z* range.

| Peptide A | Peptide B | Precursor *m/z* | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1+ | 2+ | 3+ | 4+ | 5+ | 6+ |
| C$^6$ELAAAMK | GYSLGNWVC$^{30}$AAK | 1294.0733 | 1244.0845 | 829.7254 | 622.5459 | 498.2382 | 415.3664 |
| C$^6$ELAAAMK | WWC$^{64}$NDGR | - | 1077.9690 | 718.9817 | 539.4881 | 431.7920 | 359.9945 |
| C$^6$ELAAAMK | NLC$^{76}$NIPC$^{80}$SALLSSDITASVNC$^{94}$AK | - | - | 1185.8973 | 889.6748 | 711.9413 | 593.4523 |
| C$^6$ELAAAMK | C$^{115}$K | 1468.6745 | 734.8409 | 490.2297 | 367.9241 | - | - |
| C$^6$ELAAAMK | GC$^{127}$R | **1553.7022** | **777.3547** | **518.5722** | **389.1810** | 311.5463 | - |
| GYSLGNWVC$^{30}$AAK | WWC$^{64}$NDGR | - | 1294.0733 | 863.0513 | 647.5403 | 518.2337 | 432.0293 |
| GYSLGNWVC$^{30}$AAK | NLC$^{76}$NIPC$^{80}$SALLSSDITASVNC$^{94}$AK | - | - | 1329.966854 | 997.72696 | 798.383023 | 665.4871 |
| GYSLGNWVC$^{30}$AAK | C$^{115}$K | - | 950.9453 | 634.2993 | 475.9763 | 380.9825 | 317.6533 |
| GYSLGNWVC$^{30}$AAK | GC$^{127}$R | - | 993.4591 | 662.6418 | 497.2332 | 397.9880 | 331.8245 |
| WWC$^{64}$NDGR | NLC$^{76}$NIPC$^{80}$SALLSSDITASVNC$^{94}$AK | - | - | 1219.2231 | 914.6692 | 731.9368 | 610.1152 |
| WWC$^{64}$NDGR | C$^{115}$K | 1568.6522 | 784.8297 | 523.5556 | 392.9185 | 314.5363 | - |
| WWC$^{64}$NDGR | GC$^{127}$R | - | 827.3435 | 551.8981 | 414.1754 | 331.5418 | - |
| NLC$^{76}$NIPC$^{80}$SALLSSDITASVNC$^{94}$AK | C$^{115}$K | - | 1485.2030 | 990.4711 | 743.1052 | 594.6856 | 495.7392 |
| NLC$^{76}$NIPC$^{80}$SALLSSDITASVNC$^{94}$AK | GC$^{127}$R | - | 1527.7168 | 1018.8136 | 764.3621 | 611.6911 | 509.9105 |
| C$^{115}$K | GC$^{127}$R | 967.4237 | 484.2155 | 323.1461 | - | - | - |

*The ions found in the Trypsin digestion LC-MS are highlighted in green.

Table S7 Possible fragment combinations of Lysozyme protein with PEG-Tz crosslink modification from Trypsin digestion

# Site-selective crosslink for C6 and C127



**Fragment 1**

C6 and C127 fragment
of Lysozyme-PEG-Tz

$m/z$ [M+2H]$^{2+}$ 777.3547
$m/z$ [M+3H]$^{3+}$ 518.5722
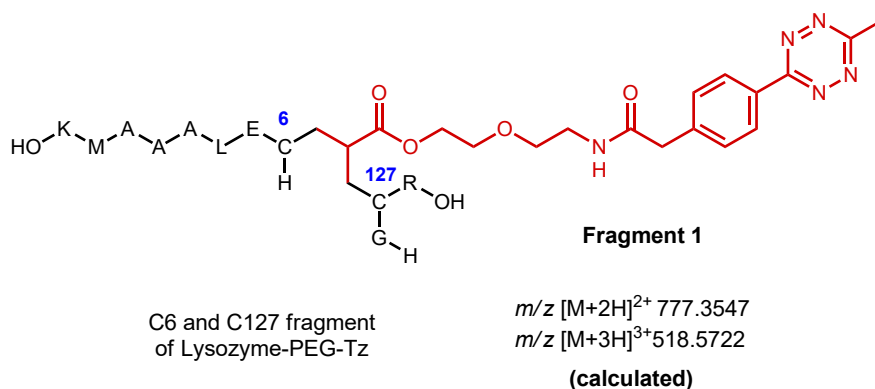
**(calculated)**

Fig. S49 Chemical structure of the C6 and C127 fragment.

| Peptide Modified Sequence | Charge | Calculated $m/z$ | Observed $m/z$ | δ ppm |
|---|---|---|---|---|
| **C$^6$ELAAAMK - PEG-Tz - GC$^{127}$R**<br><br>Fragment 1<br>**[M+383.1594+H$_2$O]** | 1+ | 1553.7022 | 1553.7021 | -0.06 |
| | 2+ | 777.3547 | 777.3577 | 3.86 |
| | 3+ | 518.5722 | 518.5723 | 0.19 |
| | 4+ | 389.1810 | 389.1821 | 2.83 |
| | 5+ | 311.5463 | NF | - |

*NF: not found

Table S8 Calculated and observed $m/z$ values for Fragment 1 (highlighted in green)
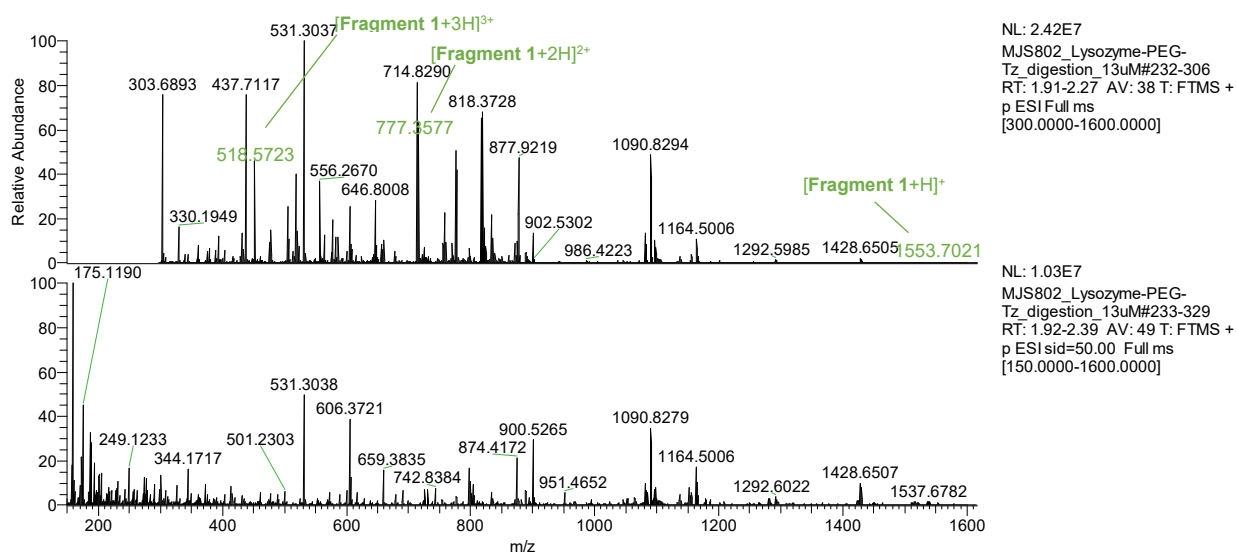


Fig. S50 Full spectrum of peak at 2.0 min, ions found are highlighted in green (top) and full spectrum from AIF spectrum of peak at 2.0 min using SID 50 eV (bottom).

44

163

MJS802_Lysozyme-PEG-Tz_digestion_13uM #231-277  RT: 1.91-2.13  AV: 24  NL: 2.35E5
T: FTMS + p ESI sid=50.00 Full ms [150.0000-1600.0000]

**[Fragment 1**+H]⁺
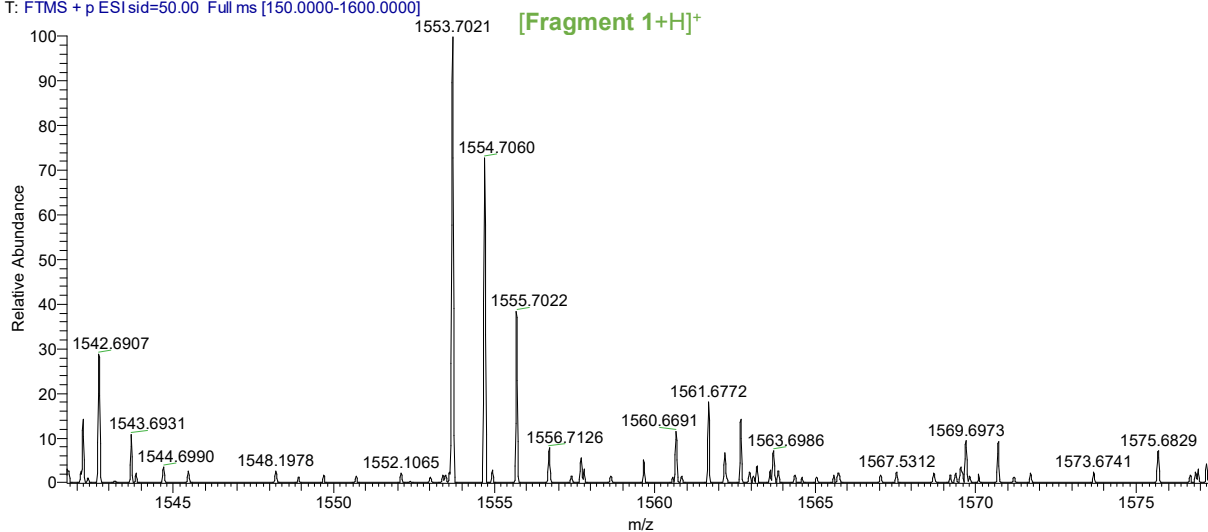


Fig. S51 Spectrum of peak at 2.0 min between 1540-1575 *m/z* range.

MJS802_Lysozyme-PEG-Tz_digestion_13uM #232-296  RT: 1.91-2.22  AV: 33  NL: 1.37E7
T: FTMS + p ESI Full ms [300.0000-1600.0000]
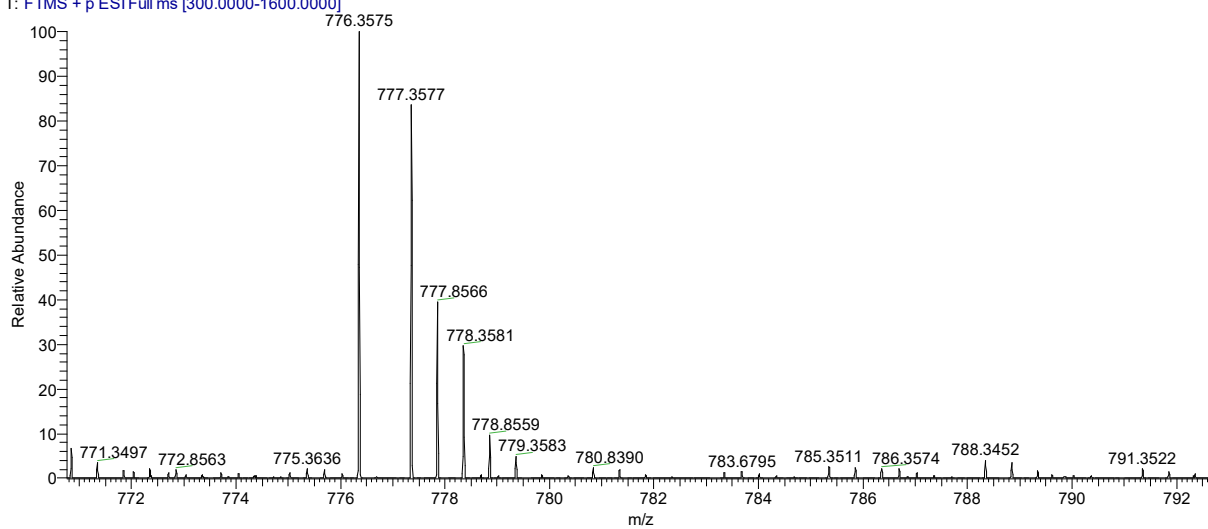


Fig. S52 Spectrum of peak at 2.0 min between 770-792 *m/z* range.

45

MJS802_Lysozyme-PEG-Tz_digestion_13uM #234-300   RT: 1.92-2.24   AV: 34   NL: 1.02E7
T: FTMS + p ESI Full ms [300.0000-1600.0000]
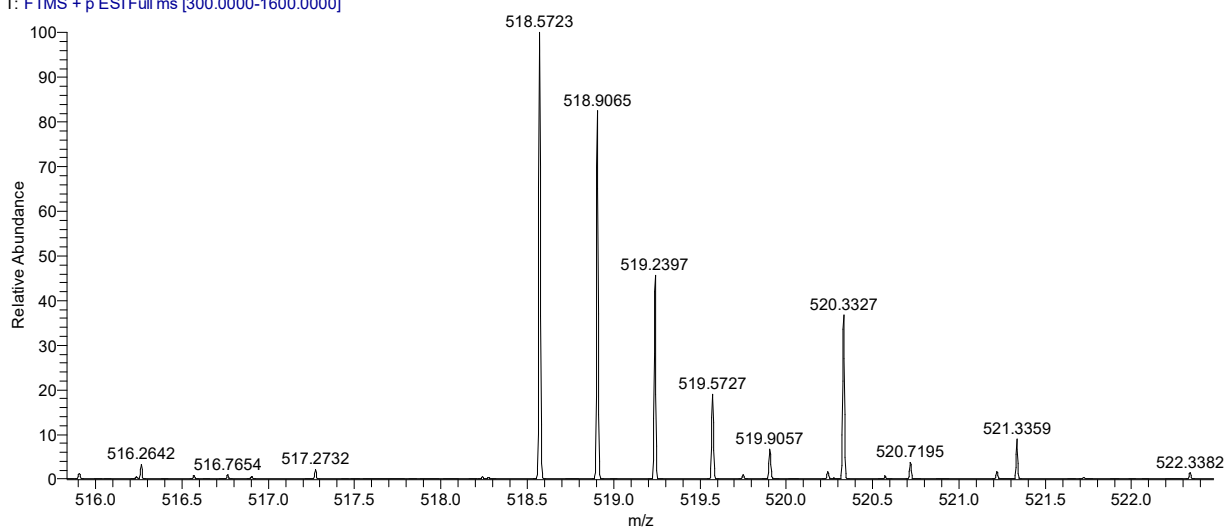


Fig. S53 Spectrum of peak at 2.0 min between 516-522 *m/z* range.

MJS802_Lysozyme-PEG-Tz_digestion_13uM #234-312   RT: 1.92-2.30   AV: 40   NL: 2.76E6
T: FTMS + p ESI Full ms [300.0000-1600.0000]
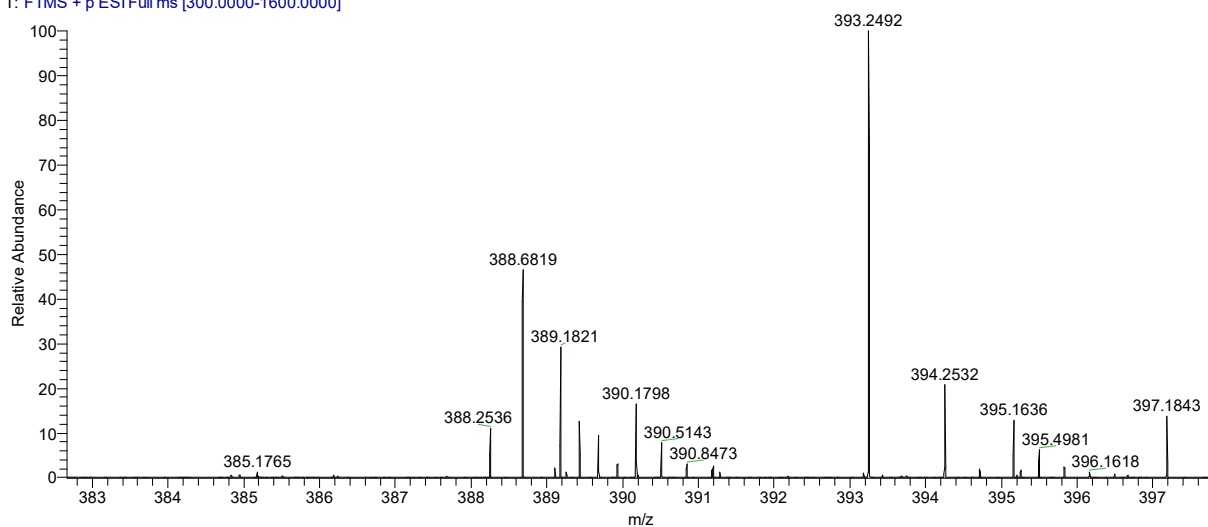


Fig. S54 Spectrum of peak at 2.0 min between 383-397 *m/z* range

46

## LC–MS/MS analysis crosslink for C6 and C127

For the fragment sequence analysis five different sets of b and y ions were considered and searched in the SID 50eV spectrum of peak at 2.0 min. Found b ions are highlighted in blue, y ions in red and PEG-Tz fragment ions in green.
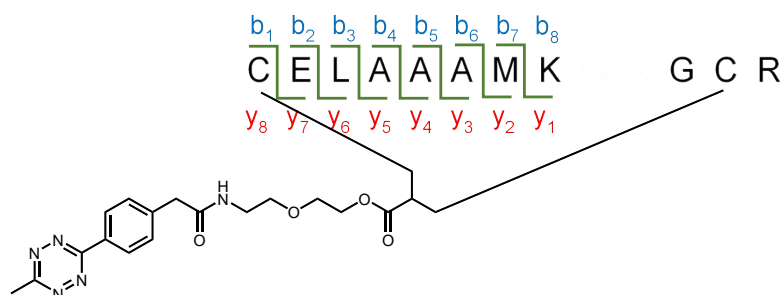


| Lost fragment | #b | b+ | Seq | y+ | #y | Lost fragment |
|---|---|---|---|---|---|---|
| ELAAAMK | 1* | 821.3187 | C | 733.3913 | 7 | C |
| LAAAMK | 2 | 950.3613 | E | 604.3487 | 6 | CE |
| AAAMK | 3 | 1063.4453 | L | 491.2646 | 5 | CEL |
| AAMK | 4 | 1134.4825 | A | 420.2275 | 4 | CELA |
| AMK | 5 | 1205.5196 | A | 349.1904 | 3 | CELAA |
| MK | 6 | 1276.5567 | A | 278.1533 | 2 | CELAAA |
| K | 7 | 1407.5972 | M | 147.1128 | 1* | CELAAAM |

Table S9 Expected b and y ions from AIF
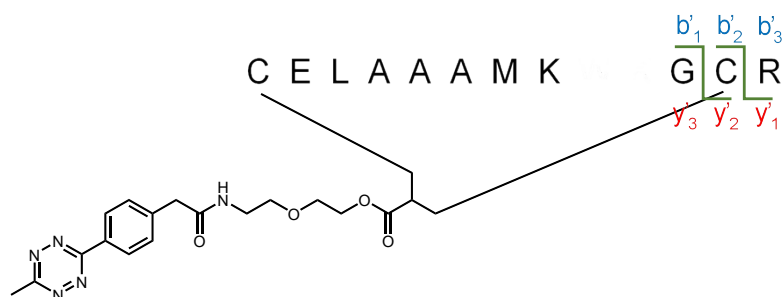


| Lost fragment | #b' | b+ | Seq | y+ | #y' | Lost fragment |
|---|---|---|---|---|---|---|
| G | 1 | 58.0287 | G | 1496.6807 | 2* | G |
| R | 2* | 1380.5978 | C | 175.1190 | 1 | R |

Table S10 Expected b' and y' ions from AIF

MJS802_Lysozyme-PEG-Tz_digestion_13uM #234-328  RT: 1.93-2.38  AV: 47  NL: 1.55E5
T: FTMS + p ESI sid=50.00  Full ms [150.0000-1600.0000]
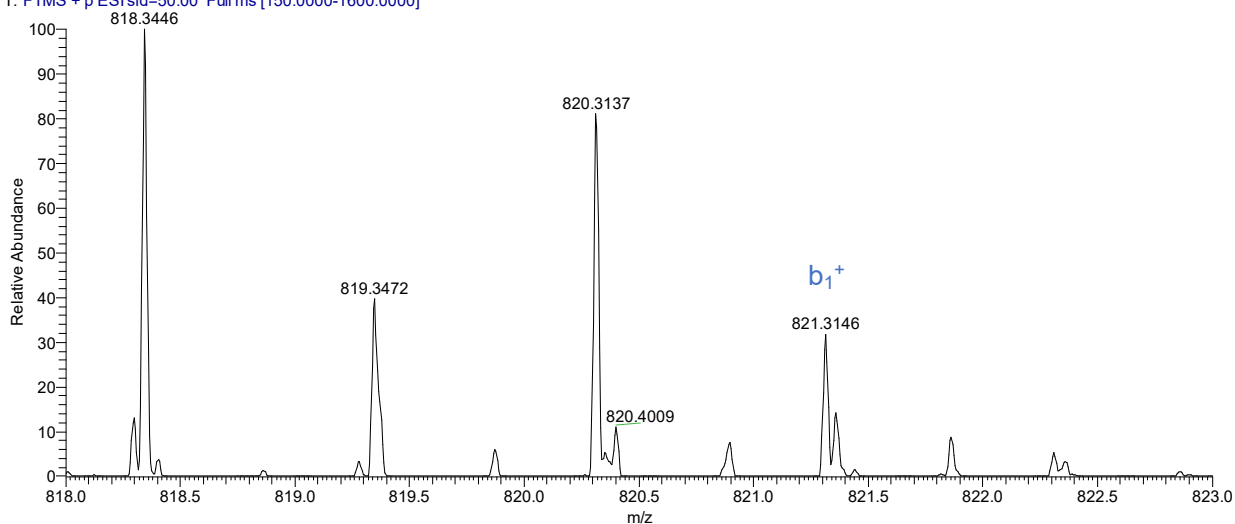
Fig. S55 AIF spectrum of peak at 2.0 min using SID 50 eV between 818-823 *m/z* range.

MJS802_Lysozyme-PEG-Tz_digestion_13uM #234-328  RT: 1.93-2.38  AV: 47  NL: 5.94E5
T: FTMS + p ESI sid=50.00  Full ms [150.0000-1600.0000]
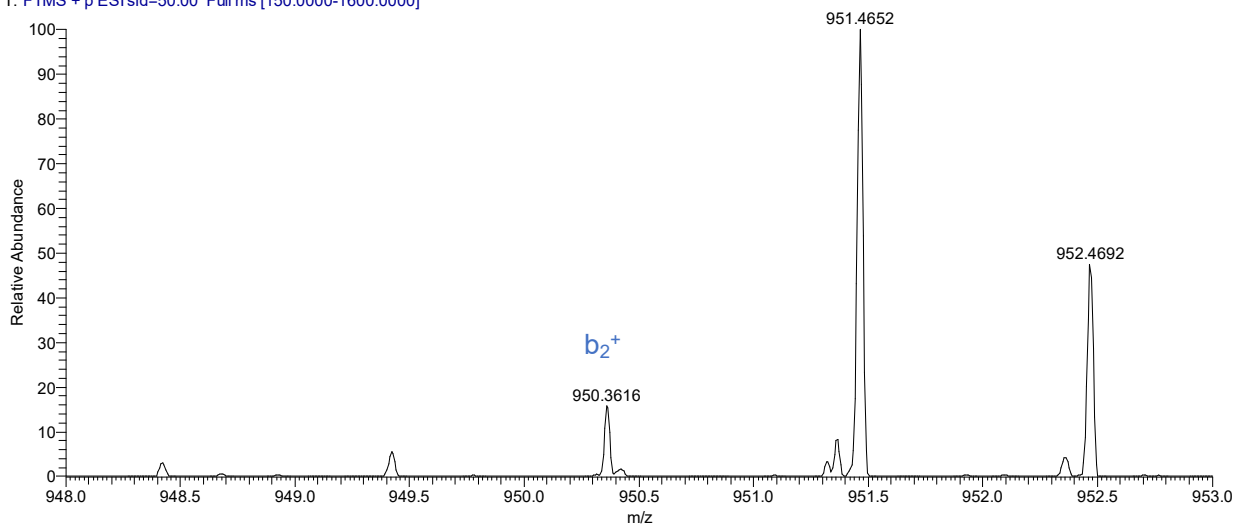
Fig. S56 AIF spectrum of peak at 2.0 min using SID 50 eV between 948-953 *m/z* range.

48
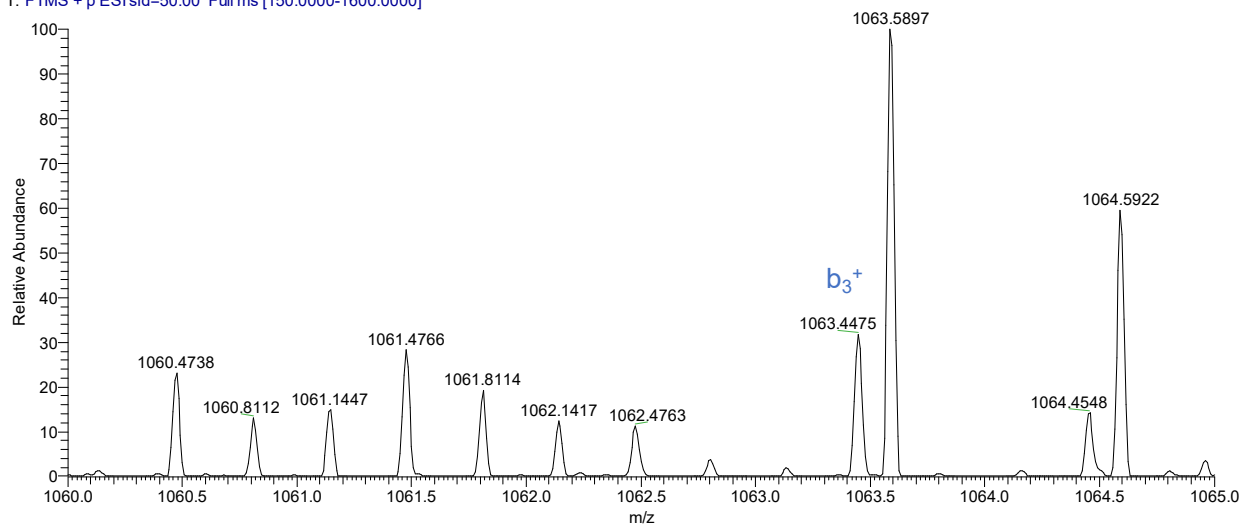
Fig. S57 AIF spectrum of peak at 2.0 min using SID 50 eV between 1060-1065 *m/z* range.
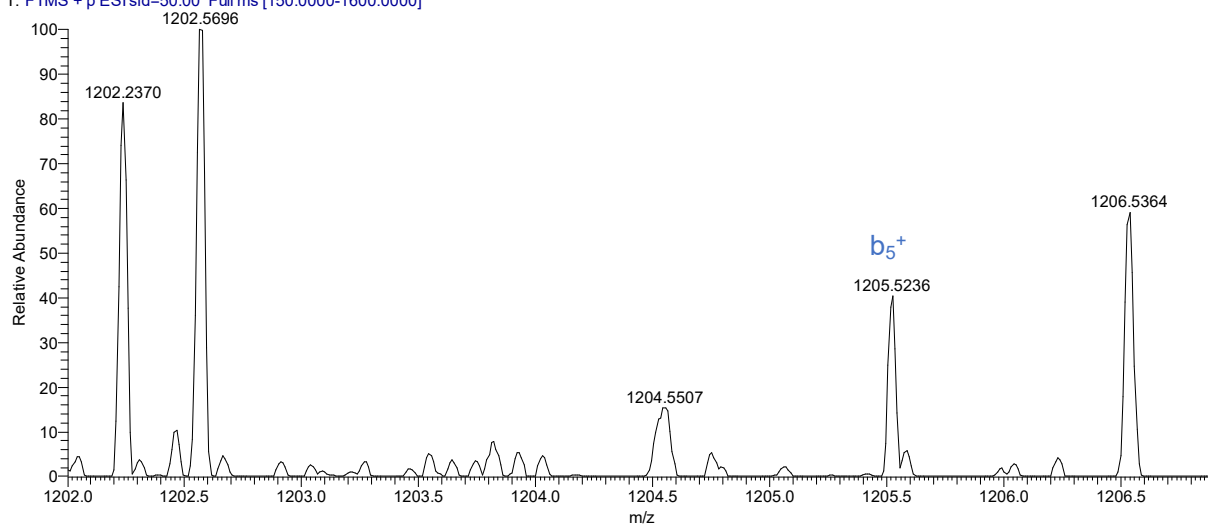
Fig. S58 AIF spectrum of peak at 2.0 min using SID 50 eV between 1202-1206.9 *m/z* range.

49
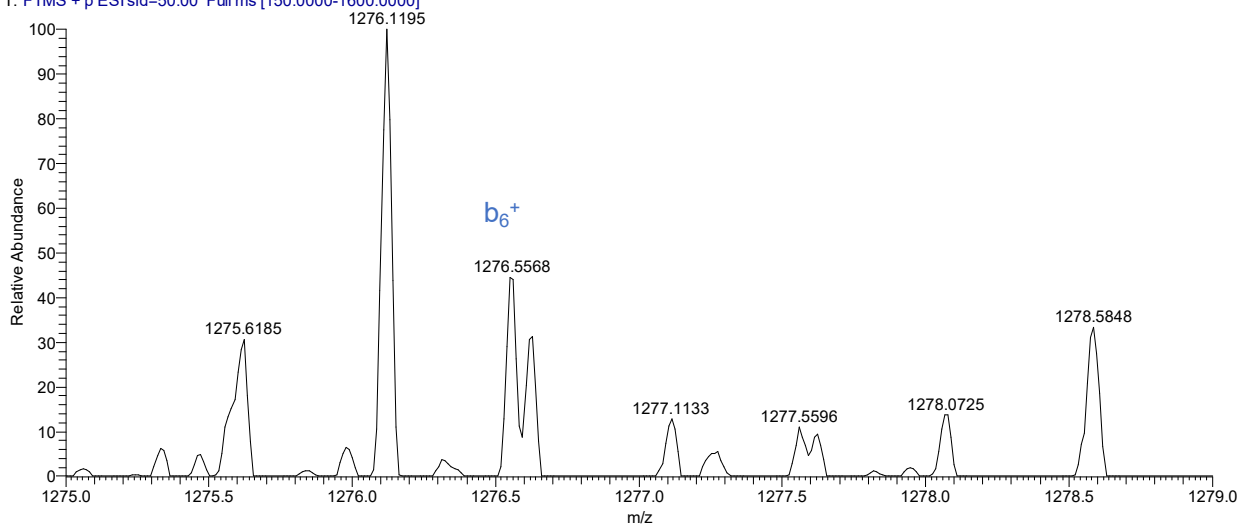
Fig. S59 AIF spectrum of peak at 2.0 min using SID 50 eV between 1275-1279 *m/z* range.
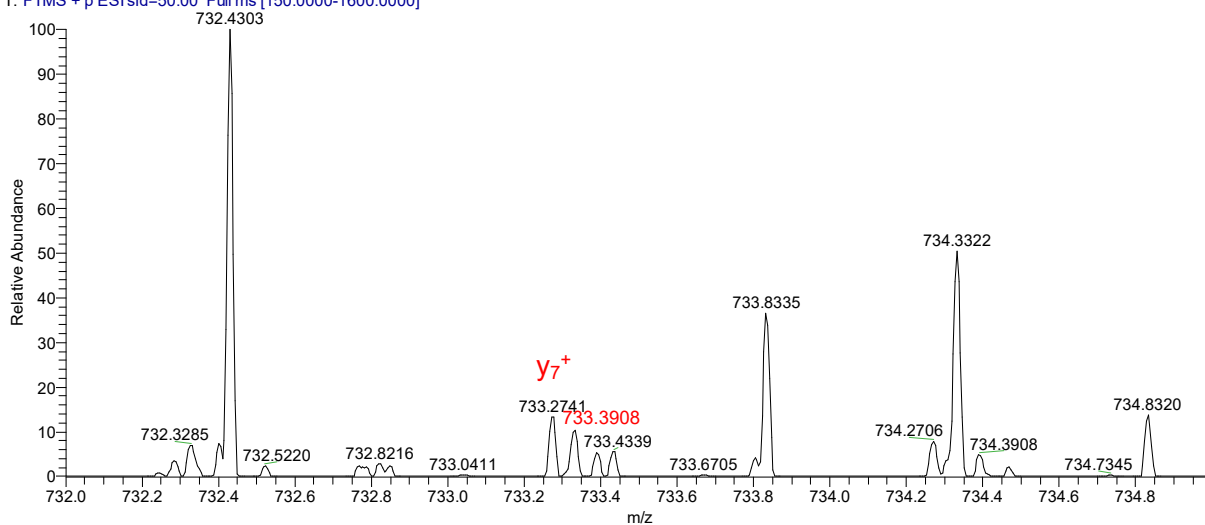
Fig. S60 AIF spectrum of peak at 2.0 min using SID 50 eV between 732-735 *m/z* range.

50

MJS802_Lysozyme-PEG-Tz_digestion_13uM #234-328  RT: 1.93-2.38  AV: 47  NL: 9.18E4
T: FTMS + p ESI sid=50.00 Full ms [150.0000-1600.0000]
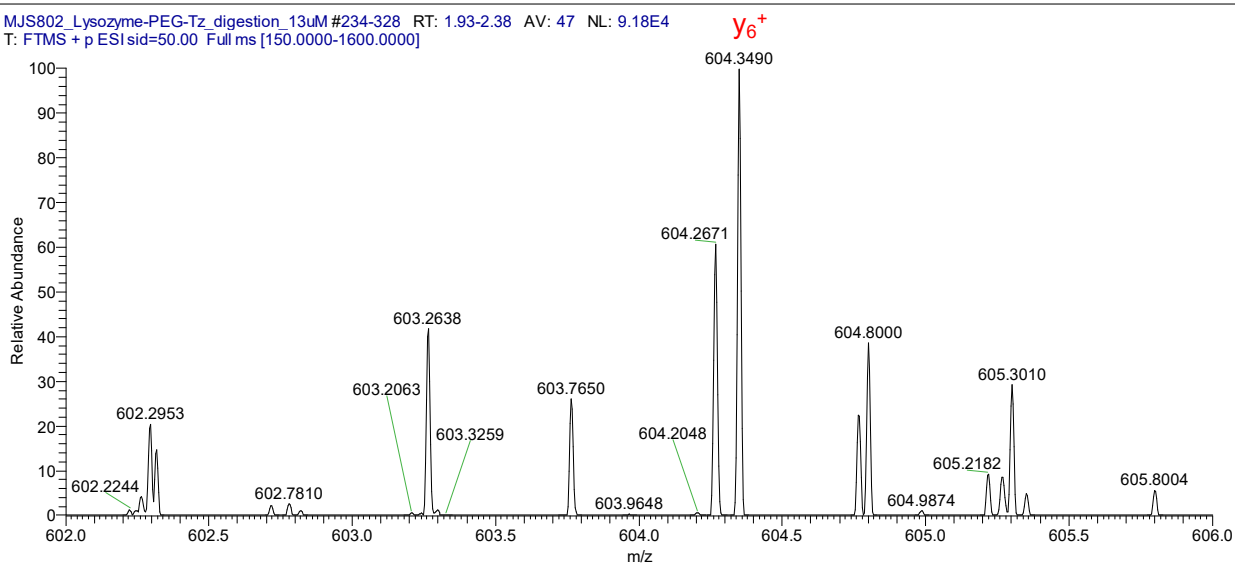


Fig. S61 AIF spectrum of peak at 2.0 min using SID 50 eV between 602-606 *m/z* range.

MJS802_Lysozyme-PEG-Tz_digestion_13uM #234-328  RT: 1.93-2.38  AV: 47  NL: 2.87E5
T: FTMS + p ESI sid=50.00 Full ms [150.0000-1600.0000]
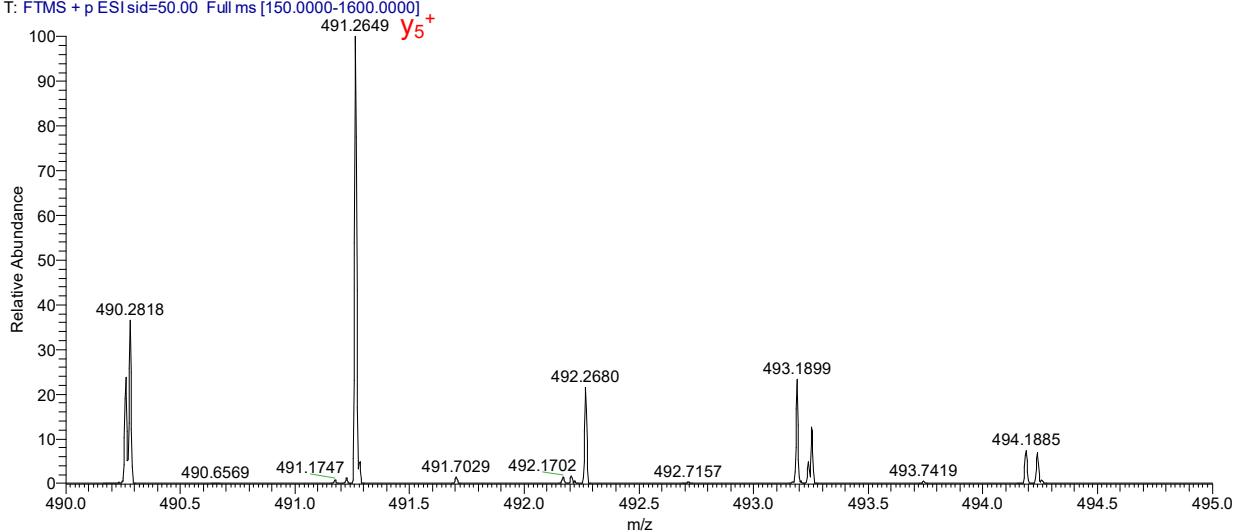


Fig. S62 AIF spectrum of peak at 2.0 min using SID 50 eV between 490-495 *m/z* range.

51
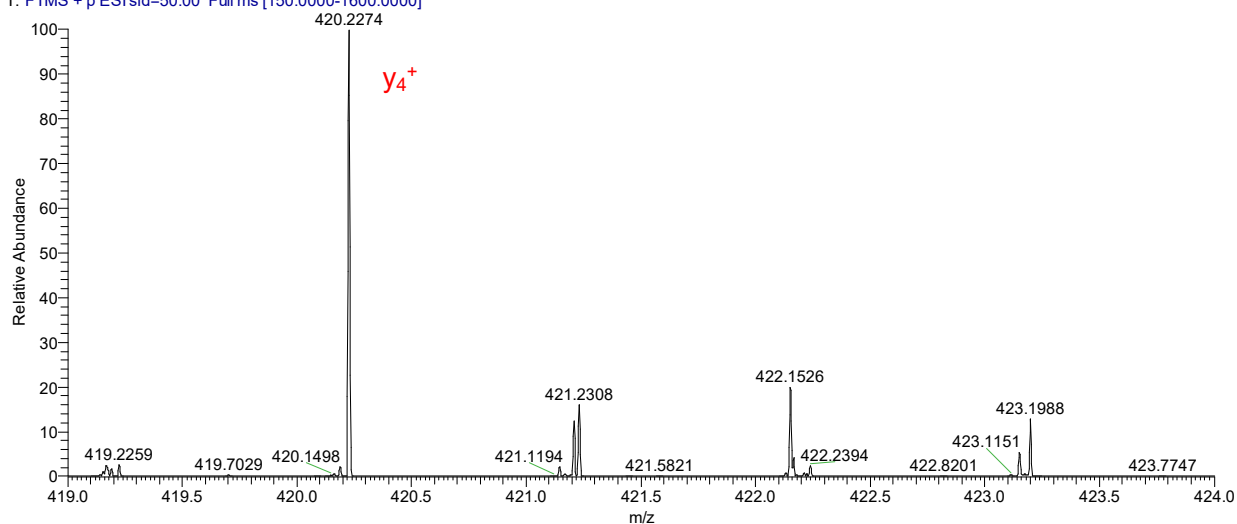
Fig. S63 AIF spectrum of peak at 2.0 min using SID 50 eV between 419-424 *m/z* range.
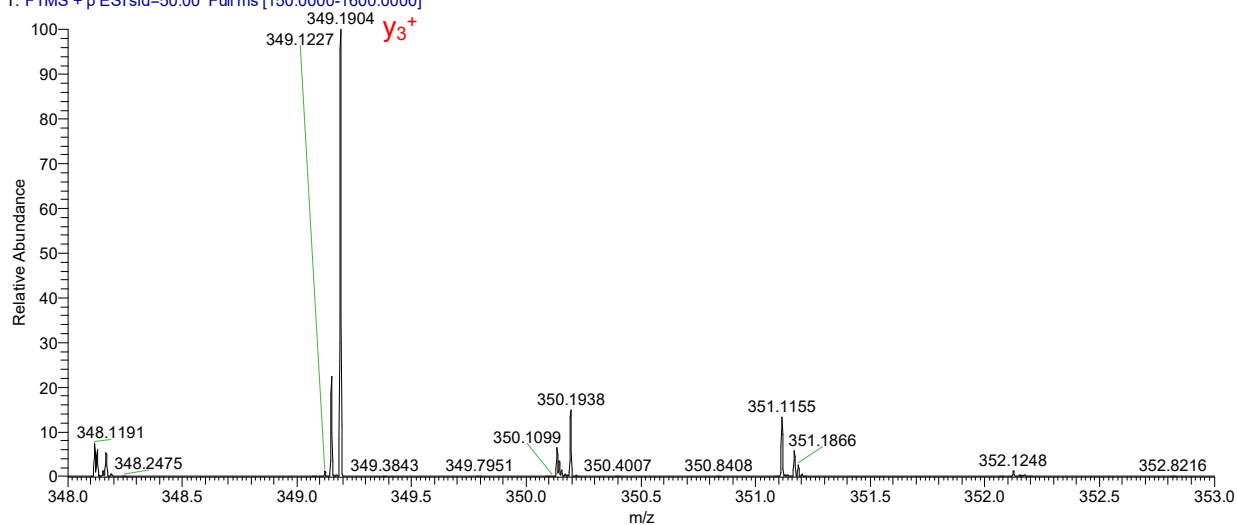
Fig. S64 AIF spectrum of peak at 2.0 min using SID 50 eV between 348-353 *m/z* range.
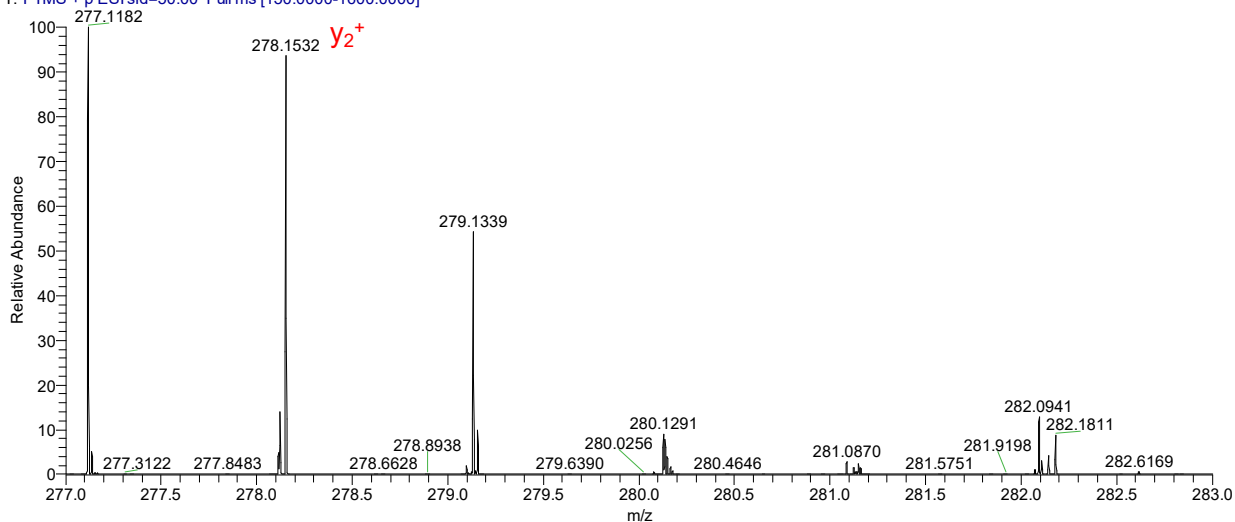
Fig. S65 AIF spectrum of peak at 2.0 min using SID 50 eV between 277-283 *m/z* range.
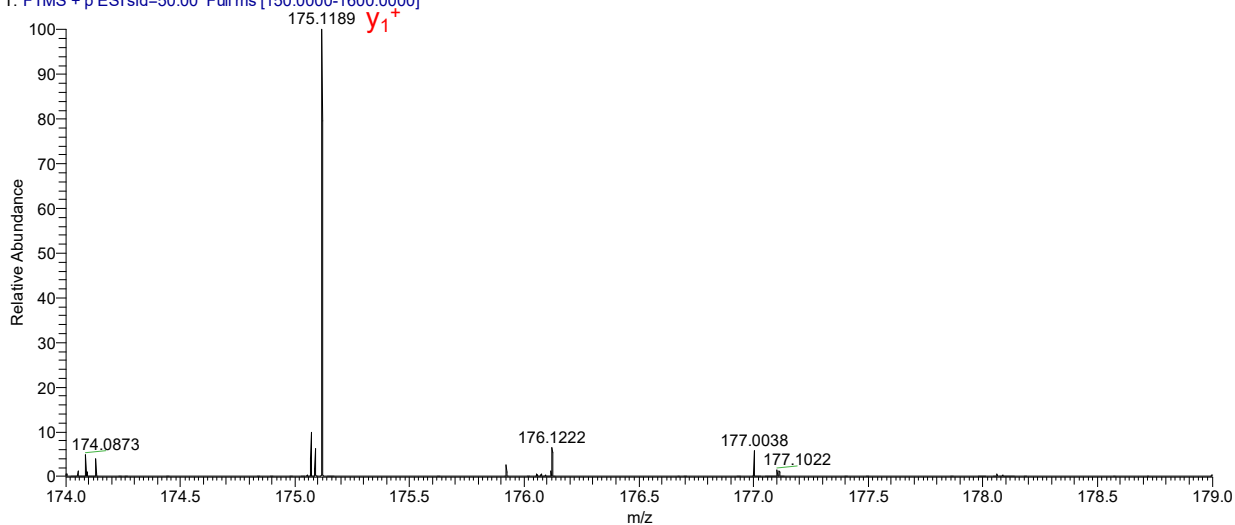
Fig. S66 AIF spectrum of peak at 2.0 min using SID 50 eV between 174-179 *m/z* range.

53
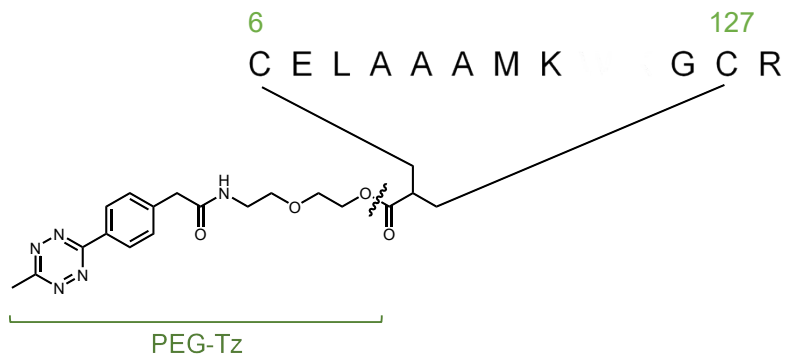
| Fragment | Calculated m/z | Observed m/z |
|---|---|---|
| PEG-Tz | 317.1482 | NF |
| CELAAAMK + GCR + $[(CH_2)_2CHCO^{·+}]$ | 1236.5534 | 1236.5533 |

Table S11 Expected ions after losing the PEG-Tz
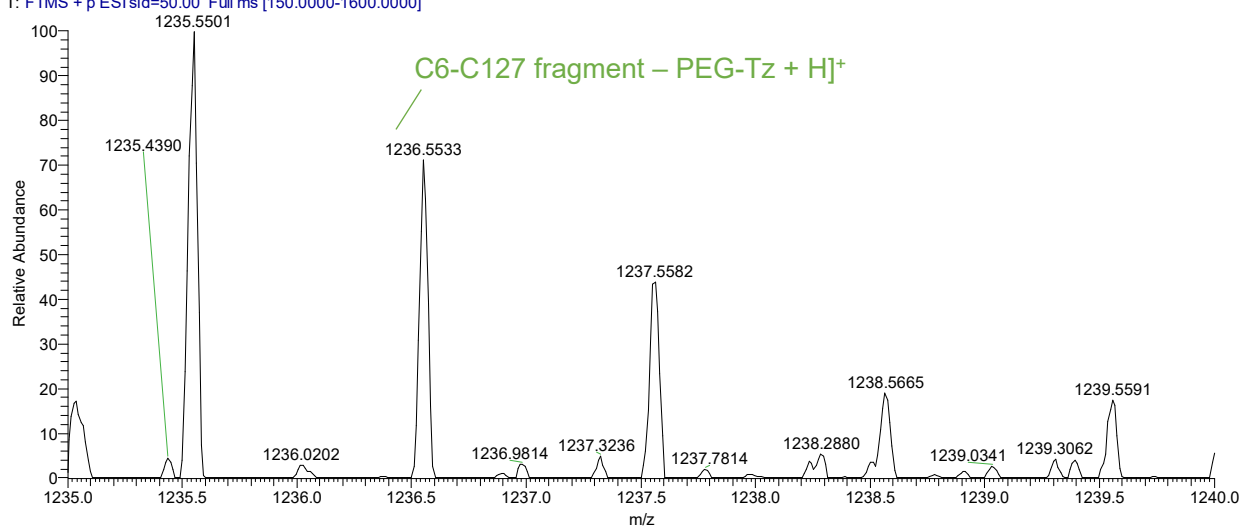


Fig. S67 AIF spectrum of peak at 2.0 min using SID 50 eV between 1235-1240 m/z range.

54

173

## 8.3 Statistical modification of lysozyme
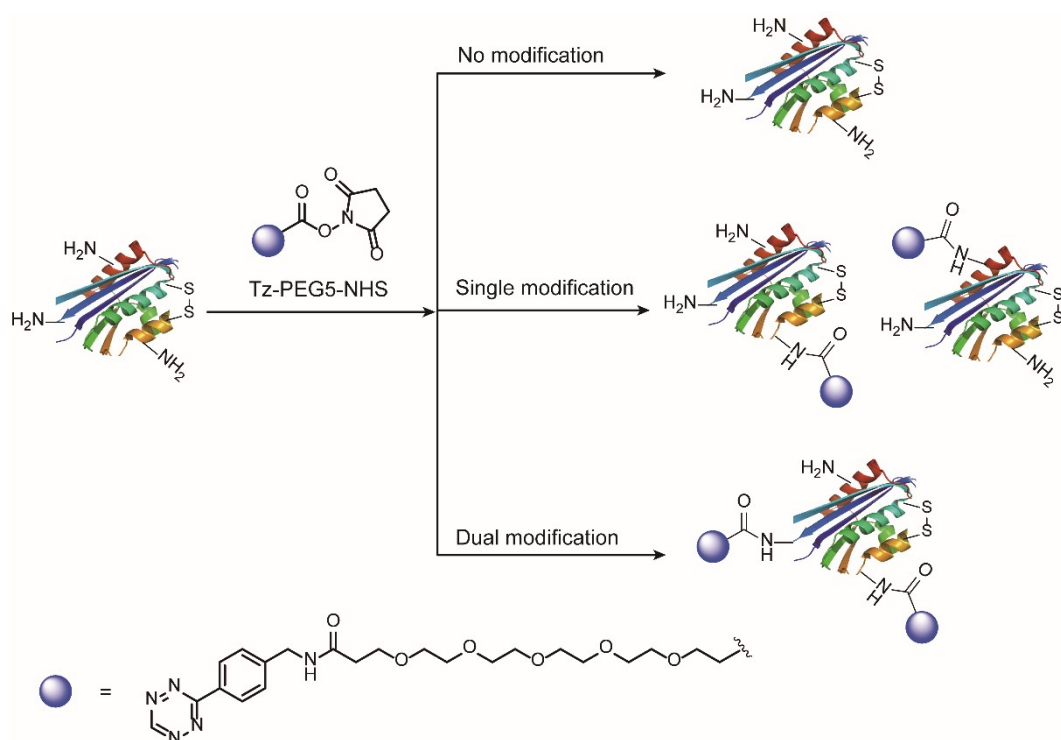


Scheme S13 Statistical modification of lysozyme with NHS ester chemistry.

Lysozyme (500 µg, 1 eq, 0.035 µmol) was dissolved in 500 µL PB buffer (50 mM, pH = 7.4). Next, Tz-PEG5-NHS (42 µg, 2 eq,.0.070 µmol) was added to the lysozyme solution and the mixture was stirred at room temperature for 4 h. Thereafter, the mixture was purified by ultrafiltration tube (cutoff: 5 kDa) to fully remove the DMSO and the excess organic molecules and change the solvent to water. The obtained solution was frozen with liquid nitrogen and water was removed with freeze dryer to afford a pink powder. The pink powder (Ly-PEG5-Tz) was characterized by MALDI-Tof-MS (Fig. S36). The MS data demonstrated that the statistical modification resulted in a mixture, which included the lysozyme with no modification, single and dual-modification.

55
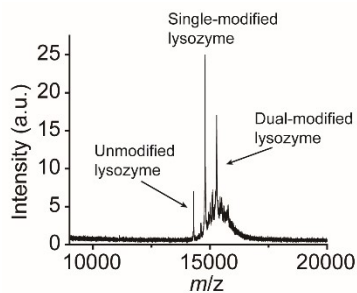
Fig. S67 MALDI-Tof-MS of Ly-PEG5-Tz after statistical modification based on the NHS ester chemistry.

8.4 Lysozyme enzymatic estimation

The lysozyme enzymatic assay was performed according to the protocol form Sigma Aldrich's technical bulletin. *Micrococcus lysodeikticus lyophilized cells* was suspended in 66 mM phosphate buffer with the pH at 6.2 with different concentrations, e.g. 0.5 mg/mL, 0.4 mg/mL, 0.3 mg/mL or 0.2 mg/mL. 250 μL of the cell suspensions was added to the 96-well plate and the absorbance was measured at 450 nm. According to the obtained absorbance, the concentration was carefully adjusted to ensure the absorbance at 450 nm between 0.6 and 0.7. In the experiment performed in this paper, the concentration was determined to be 0.32 mg/mL with the absorbance of around 0.65 at 450 nm. Ly-PEG-Tz was selected as a representative one to estimate its observed activity after modification. 250 μL of the cell suspensions was put in 3 wells of a 96-well plate. For the first well, 10 μL of phosphate buffer (66 mM, pH = 6.2) was added as a blank. To the second and third well, 10 μL of native lysozyme (0.01 mg/mL) and Ly-PEG-Tz (0.01 mg/mL) were added. The mixtures were pipetted up and down to fully mix the suspensions. Then the absorbance of each mixture was measured by Microplate reader (Tecan Spark 20 M) every minute for 15 minutes. The plotted graph was shown in Fig. 8b of the main text. The preserved activity of the modified lysozyme was estimated by comparing the slope of the plotted linear functions. The percent of the preserved activity of modified was calculated by dividing ( $\frac{\Delta A_{450}}{min}$ (Ly-PEG-Tz) - $\frac{\Delta A_{450}}{min}$ (blank)) by ( $\frac{\Delta A_{450}}{min}$ (native Lysozyme) - $\frac{\Delta A_{450}}{min}$ (blank)).

56

| Samples | Blank | Native lysozyme | Ly-PEG-Tz |
|---|---|---|---|
| Slopes of the linear function | 0.00046 | -0.00985 | -0.01148 |
| $\frac{\Delta A_{450}}{min}$ (protein) - $\frac{\Delta A_{450}}{min}$ (blank) | | 0.01031 | 0.01194 |
| Preserved enzymatic activity | | | 86% |

Similarly, the enzyme activity of the statistical modified lysozyme was also evaluated following the same protocol described above. The absorbance of each mixture was also measured by Microplate reader (Tecan Spark 20 M) every minute for 15 minutes. The preserved activity of the modified lysozyme was estimated by comparing the slope of the plotted linear functions. The plotted graph was shown in Fig. 8c. The percent of the preserved activity of modified was calculated by dividing ($\frac{\Delta A_{450}}{min}$ (Ly-PEG-Tz) - $\frac{\Delta A_{450}}{min}$ (blank)) by ($\frac{\Delta A_{450}}{min}$ (native Lysozyme) - $\frac{\Delta A_{450}}{min}$ (blank)).

| Samples | Blank | Native lysozyme | Ly-PEG5-Tz |
|---|---|---|---|
| Slopes of the linear function | 0.00047 | -0.01027 | 0.00045 |
| $\frac{\Delta A_{450}}{min}$ (protein) - $\frac{\Delta A_{450}}{min}$ (blank) | | 0.01074 | -0.00002 |
| Preserved enzymatic activity | | | ~ 0% |

8.5 Circular dichroism

57
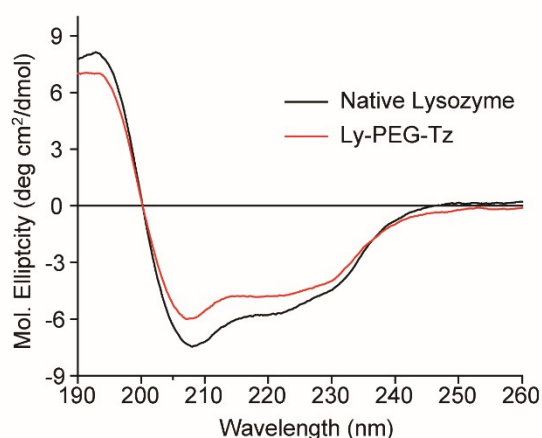
Fig. S68 CD spectra of native lysozyme and modified lysozyme from 190 nm to 260 nm.

## 9. Summary and comparison the disulfide modification strategy
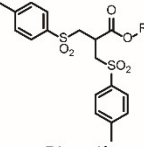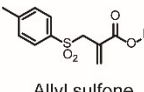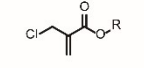
| Reagents | Synthetic steps | Conjugation Yield | Reaction time | log P | Reaction conditions |
|---|---|---|---|---|---|
| Dibromomaleimide | Two steps | >80 (Determined by SDS-PAGE and MS spectrum) | 4 h | 0.1 | Borate buffered saline (50 mM, pH = 7 ) 2 eq conjutation reagent |
| Divinylpyrimidine | Three steps | Good conversion (Determined by MS spectrum) | 2 h | 3.13 | Tris buffer (25 mM, pH = 8) 15 eq conjutation reagent |
| Oxetane | Commerical (No possibility for post-functionalization) | > 95% (Dertermined by MS spectrum) | 24 h | 1.63 | Phosphate buffer (100 mM, pH = 9) 20 eq conjugation reagent |
| Bis-sulfone | Three steps | around 20% Isolated yield | Overnight | 2.4 | Phosphate buffer (100 mM, pH = 7.8) 2 eq conjugation reagent |
| Allyl sulfone | Four steps | 19% Isolated yield | Overnight | 1.6 | Phosphate buffer (50 mM, pH = 7.8) 2 eq conjugation reagent |
| 2-Chrlomethyl acrylate | One step | 30% Isolated yield | Overnight | 1.15 | Phosphate buffer (50 mM, pH = 7) 1.2 eq conjugation reagent |

Table S12 A brief summary of the current available disulfide modification strategies reported in the literature with respect to the synthetic steps, conjugation yield, reaction time, log P as well as reaction conditions.

As shown in Table S12, the disulfide strategy developed in this work bears the advantages of easy preparation (one-step synthesis), good water solubility (rather low log P) and relatively efficient conjugation with using a

58

little bit excess of rebridging reagent.

## 10. LC-MS of the synthesized compounds
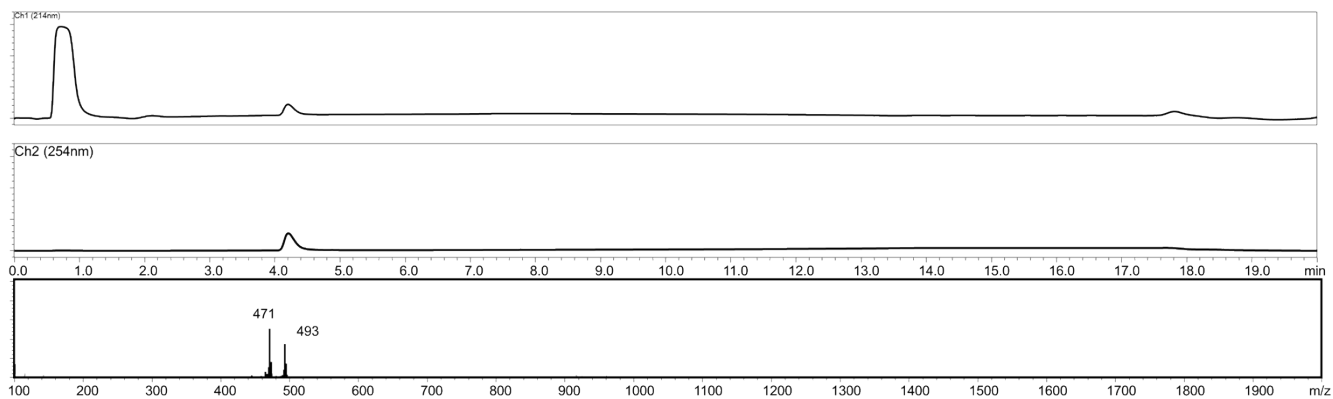


Fig. S69 LC-MS of compound **2** (calculated: 471[M+H]$^+$, 493 [M+Na]$^+$; found: 471 [M+H]$^+$, 493 [M+Na]$^+$)
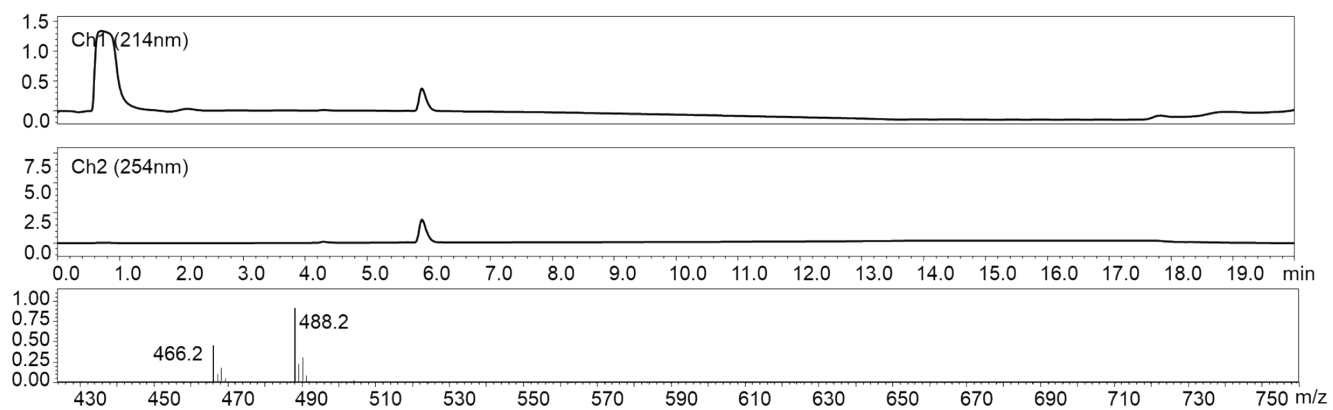


Fig. S70 LC-MS of compound **3** (calculated: 466 [M+H]$^+$, 488 [M+Na]$^+$; founded: 466 [M+H]$^+$, 488 [M+Na]$^+$)



59
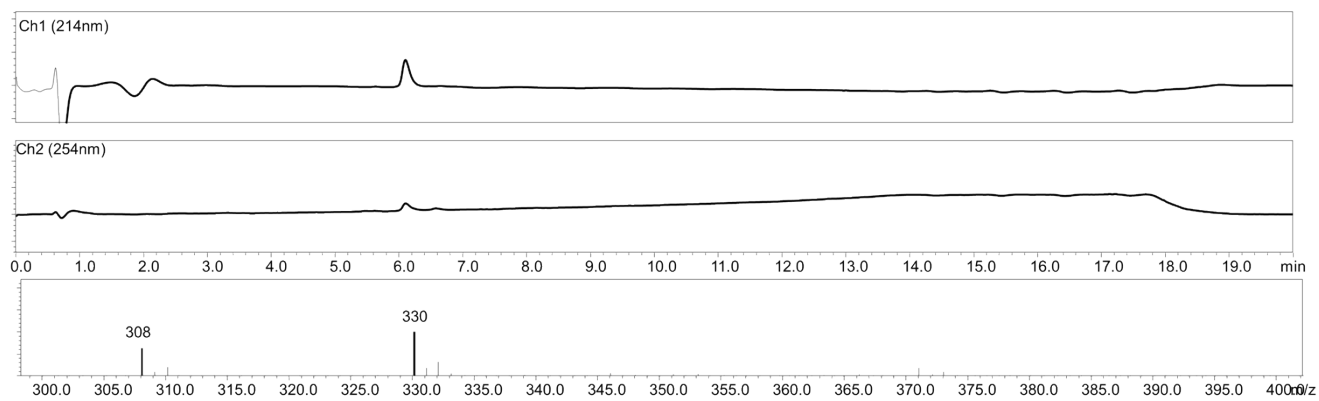
Fig. S71 LC-MS of compound **5** (calculated: 308 [M+H]⁺, 330 [M+Na]⁺; found: 308 [M+H]⁺, 330 [M+Na]⁺)
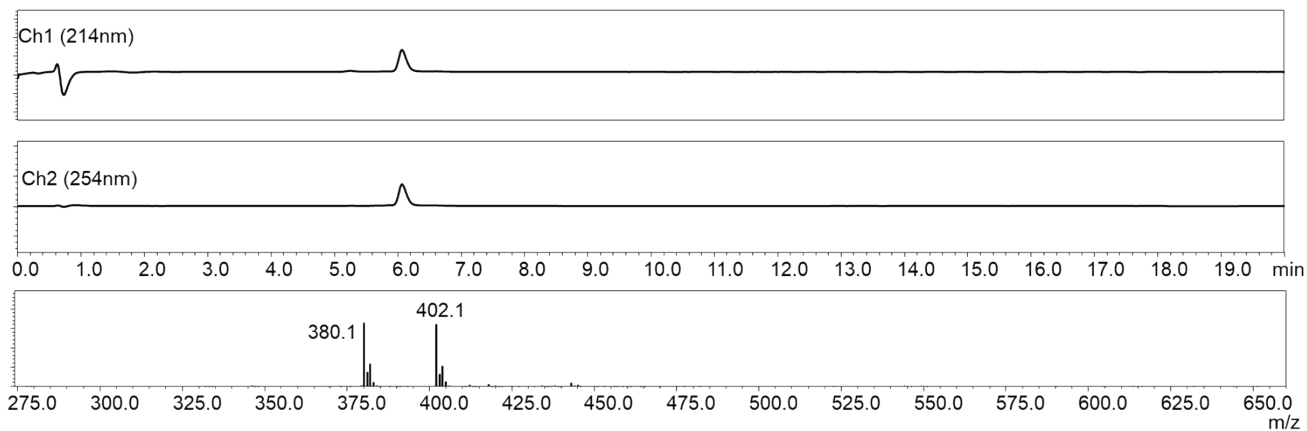


Fig. S72 LC-MS of compound **6** (calculated: 380 [M+H]⁺, 402 [M+Na]⁺; found: 380 [M+H]⁺, 402 [M+Na]⁺)
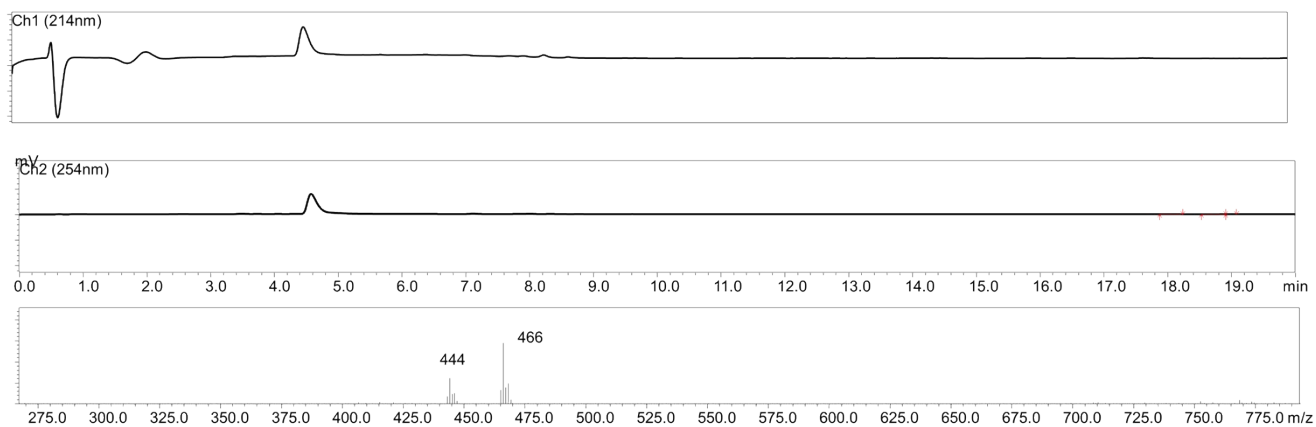


Fig. S73 LC-MS of compound **7** (calculated: 444 [M+H]⁺, 466 [M+Na]⁺; found: 444 [M+H]⁺, 466 [M+Na]⁺)
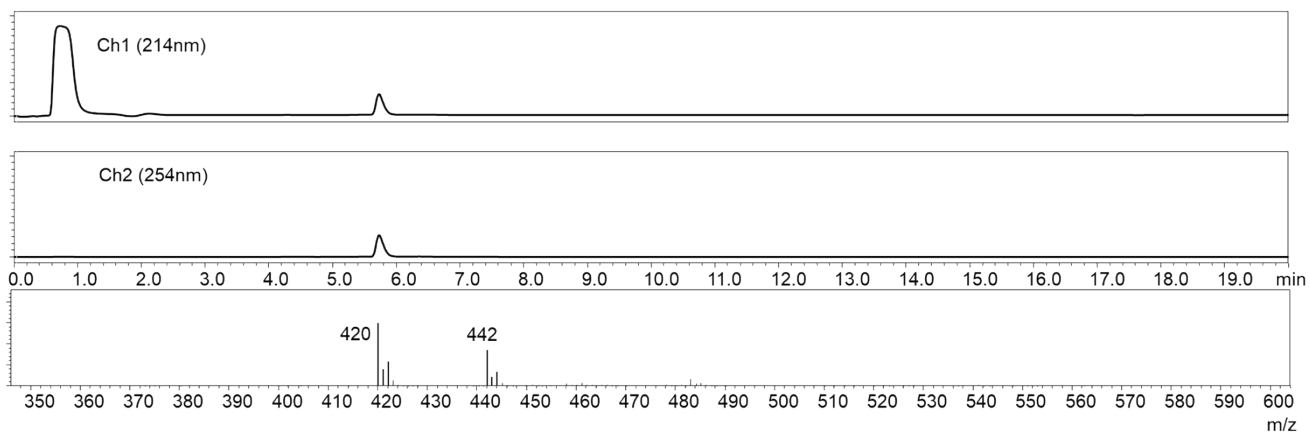


Fig. S74 LC-MS of compound **8** (calculated: 420 [M+H]⁺, 442 [M+Na]⁺; found: 420 [M+H]⁺, 442 [M+Na]⁺)
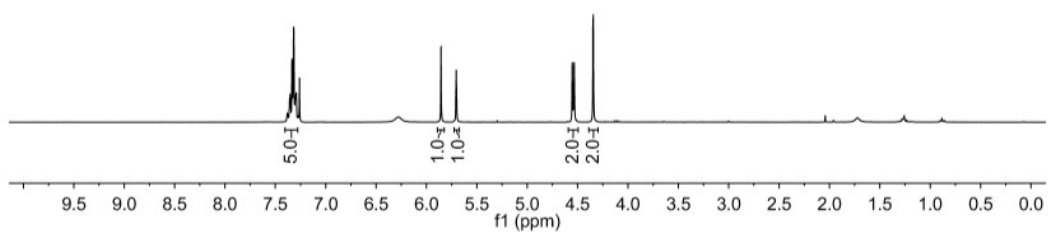
Fig. S75 $^{1}$H NMR of compound **1** (CDCl$_3$, 298K, 300 MHz)
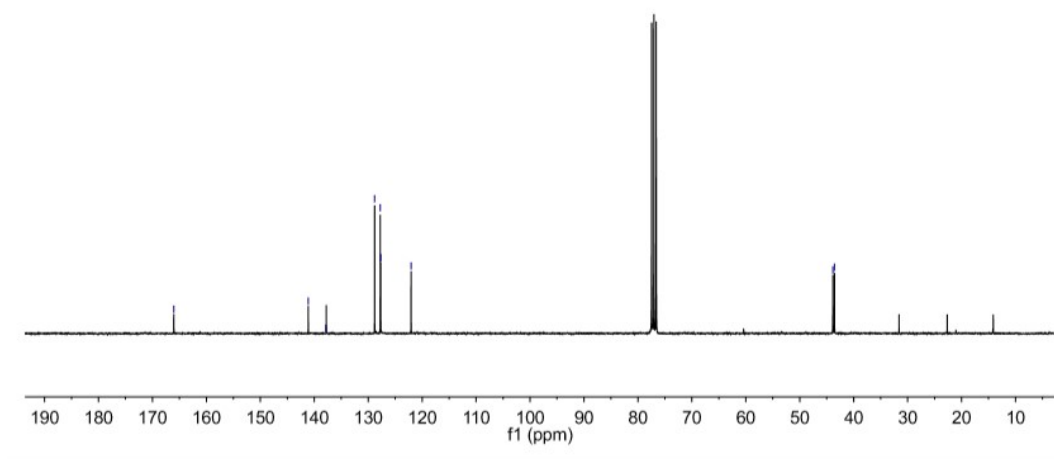


Fig. S76 $^{13}$C NMR of Compound **1**[1] (CDCl$_3$, 298K, 75 MHz)

61
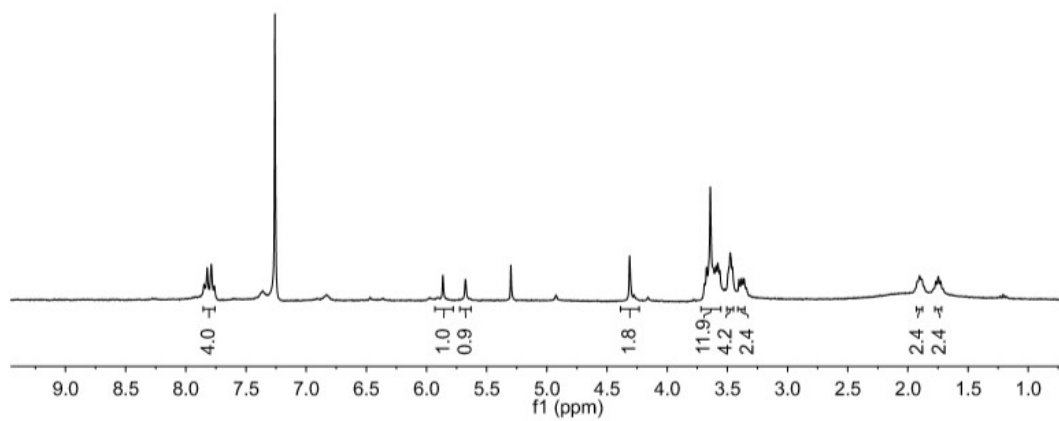
Fig. S77 $^1$H NMR of compound **2** (CDCl$_3$, 298K, 300 MHz)
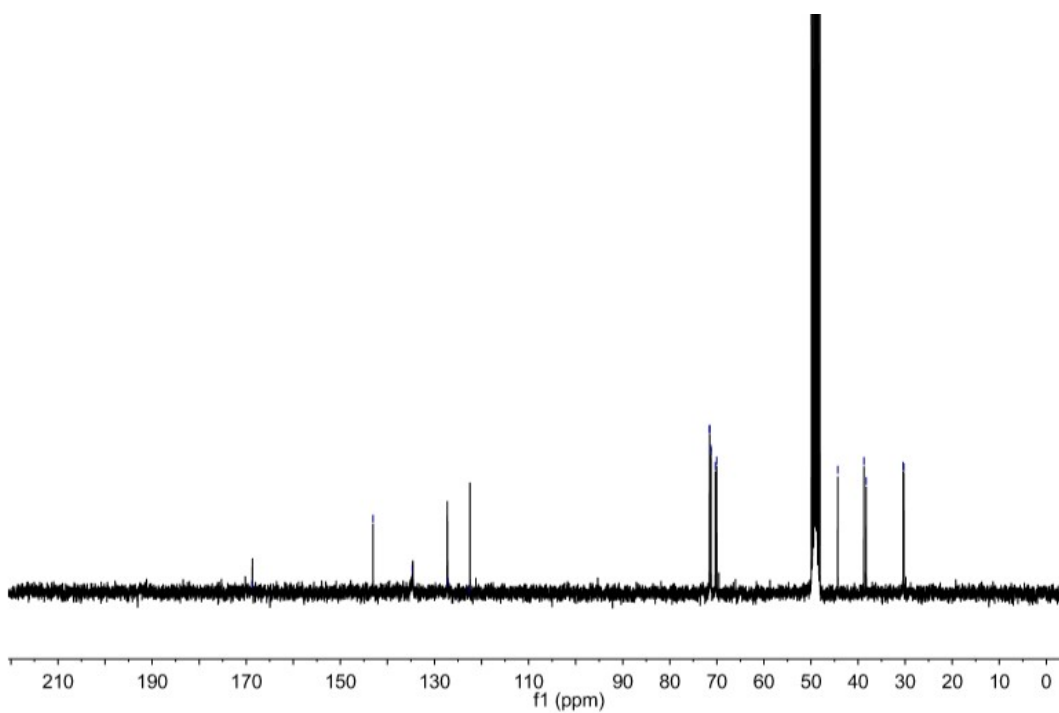


Fig. S78 $^{13}$C NMR of compound **2** (CD$_2$Cl$_2$, 298K, 75 MHz)

62
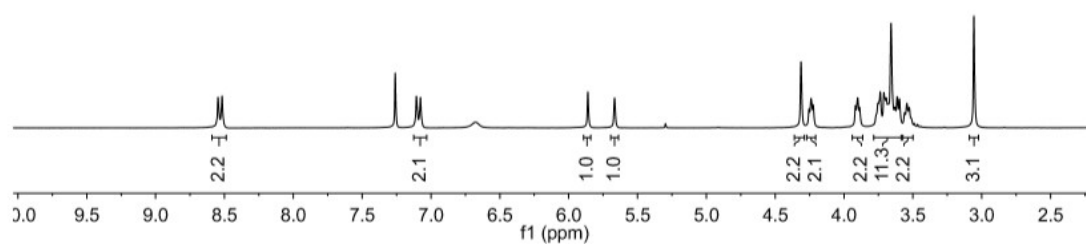
Fig. S79 $^1$H NMR of compound **3** (CDCl$_3$, 298K, 300 MHz)
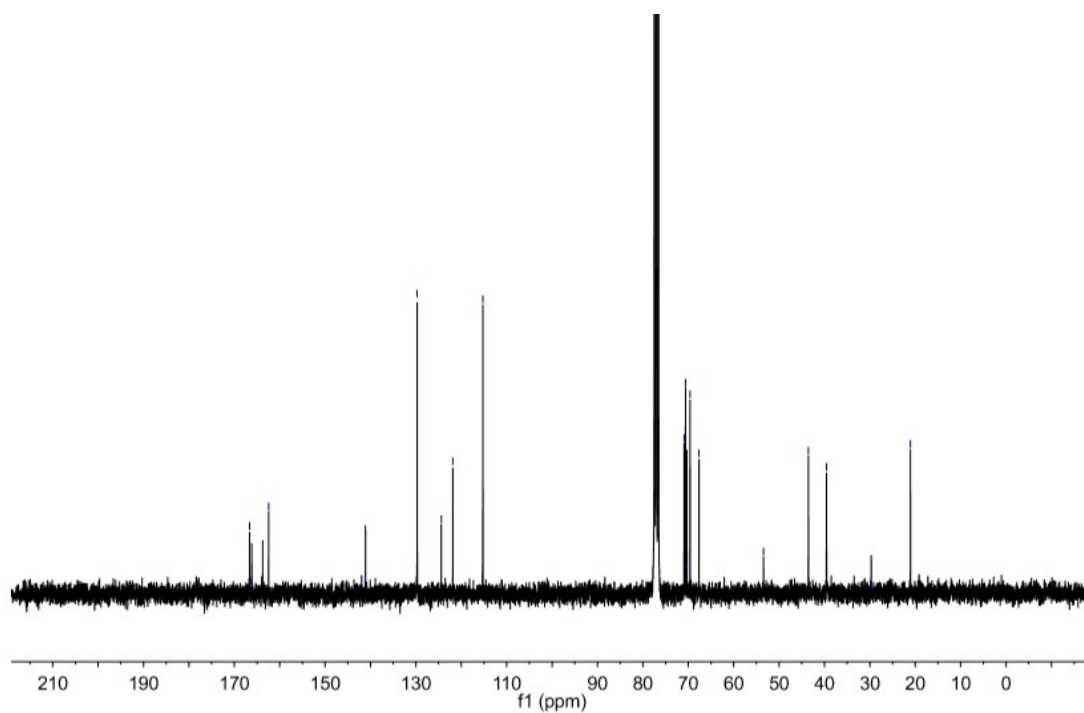


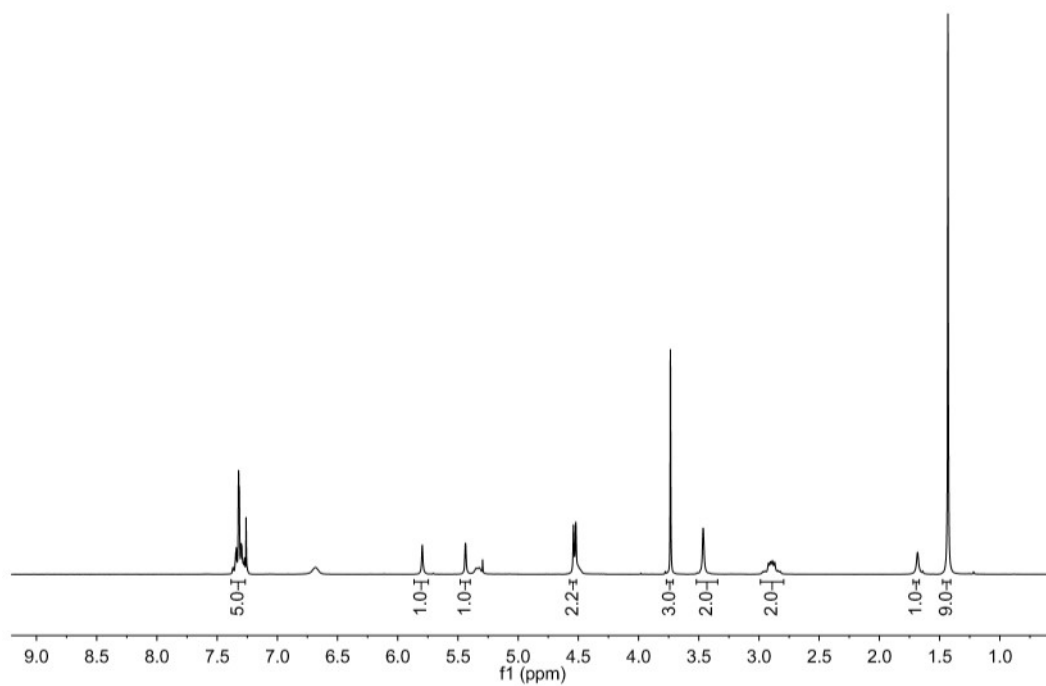Fig. S80 $^{13}$C NMR of compound **3** (CDCl$_3$, 298K, 75 MHz)

63
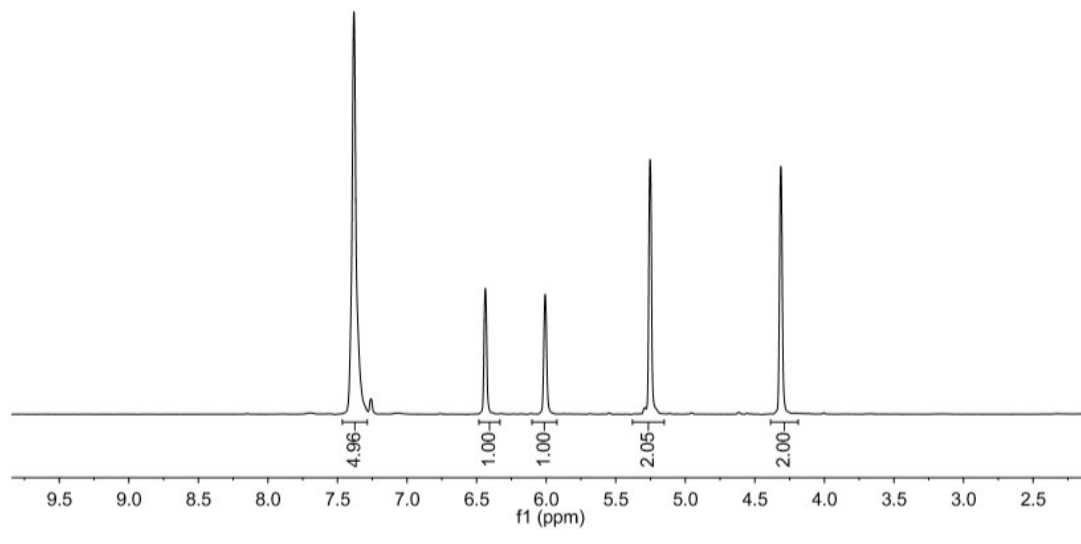
Fig. S81 $^1$H NMR of compound **4** (CDCl$_3$, 298K, 300 MHz)
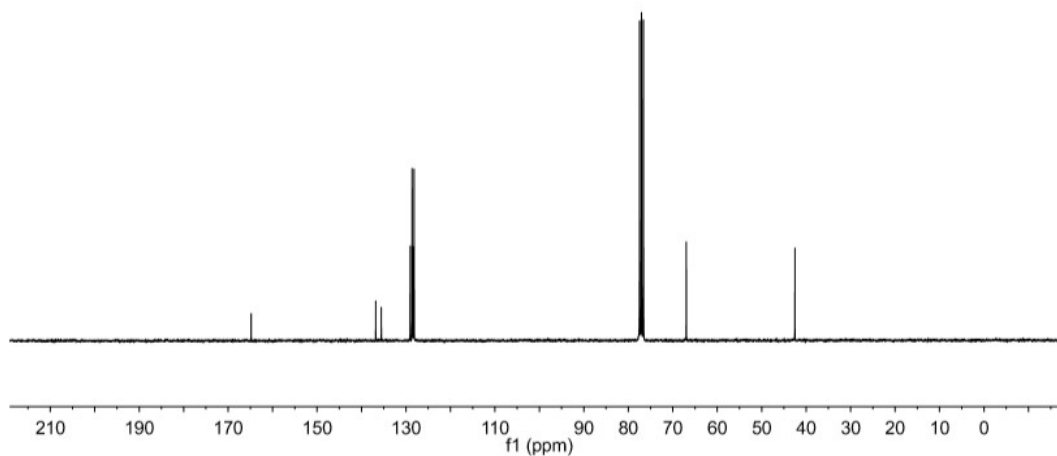


Fig. S82 $^{13}$C NMR of compound **4** (CDCl$_3$, 298K, 75 MHz)

64
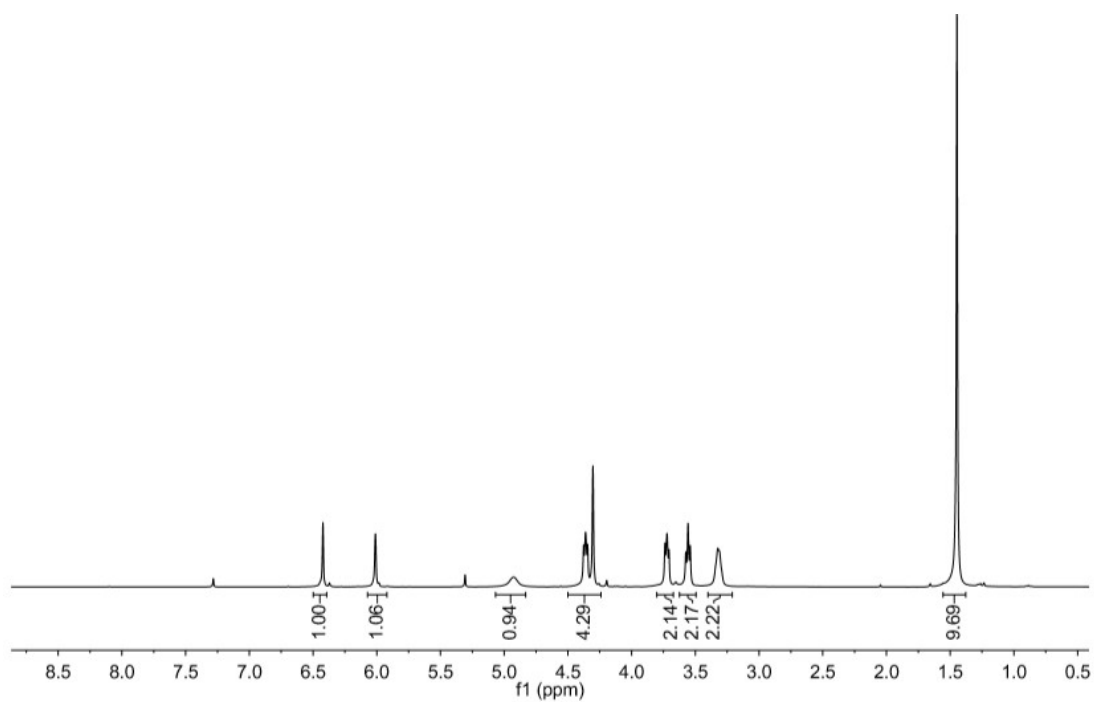
Fig. S83 $^1$H NMR of compound **5** (CDCl$_3$, 298K, 300 MHz)
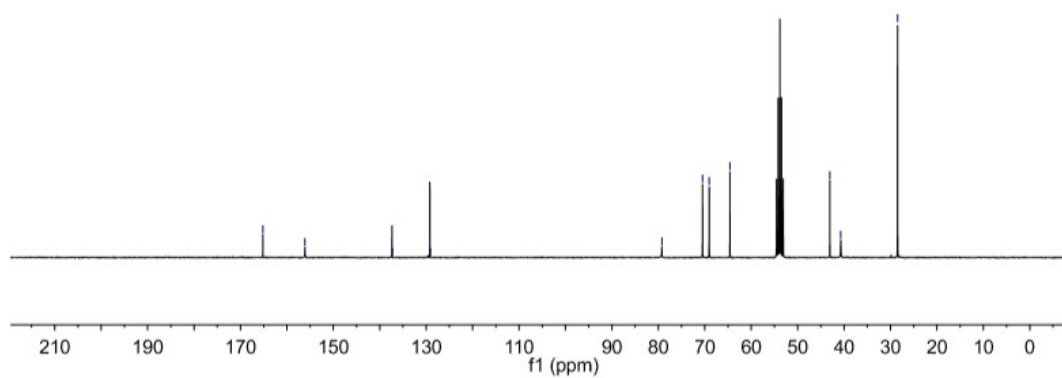


Fig. S84 $^{13}$C NMR of compound **5** (CD$_2$Cl$_2$, 298K, 75 MHz)

65
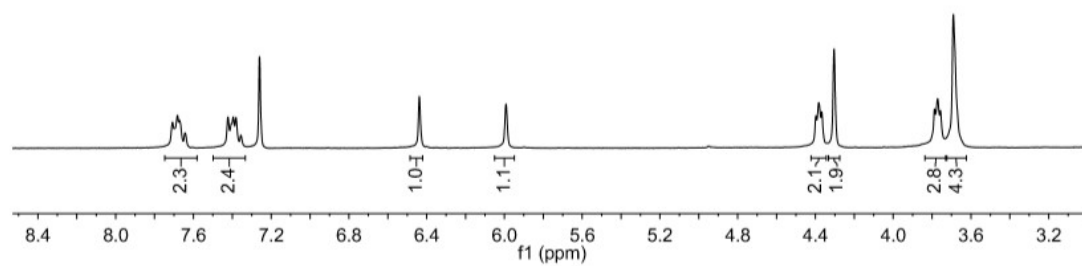
Fig. S85 $^1$H NMR of compound **6** (CDCl$_3$, 298K, 300 MHz)
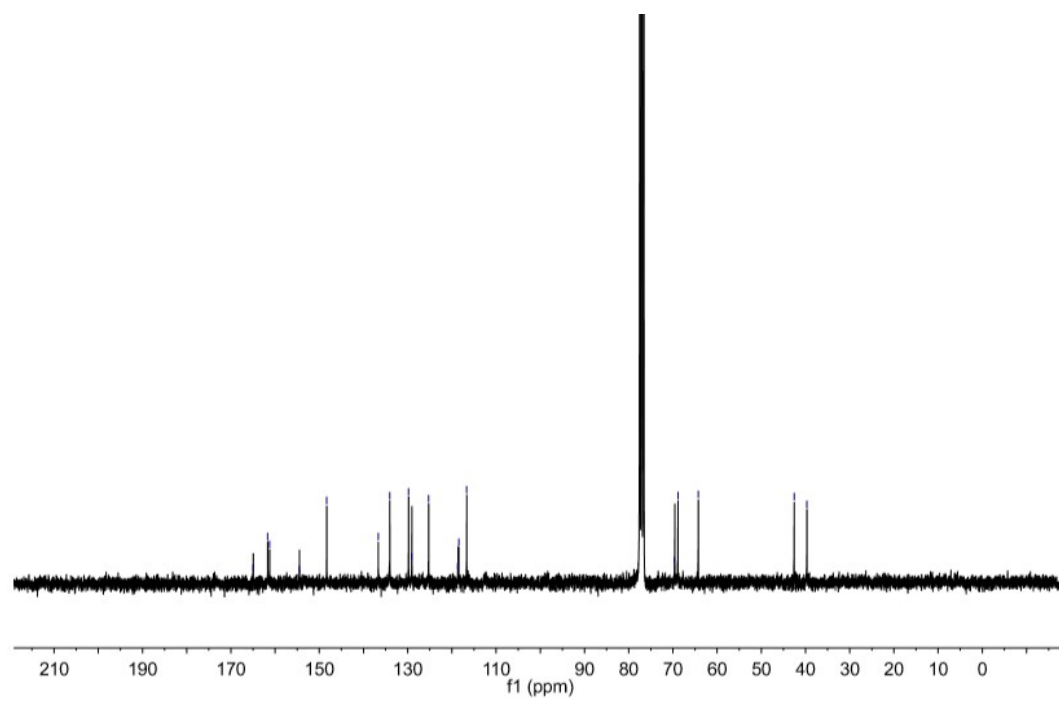


Fig. S86 $^{13}$C NMR of compound **6** (CDCl$_3$, 298K, 75 MHz)

66
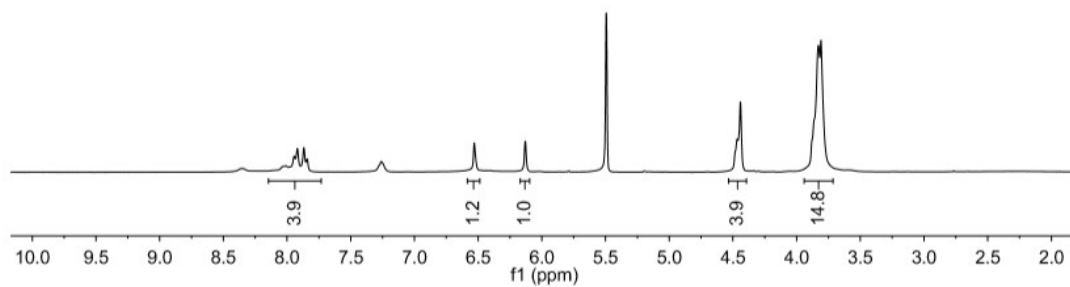
Fig. S87 $^1$H NMR of compound **7** (CD$_2$Cl$_2$, 298K, 300 MHz)
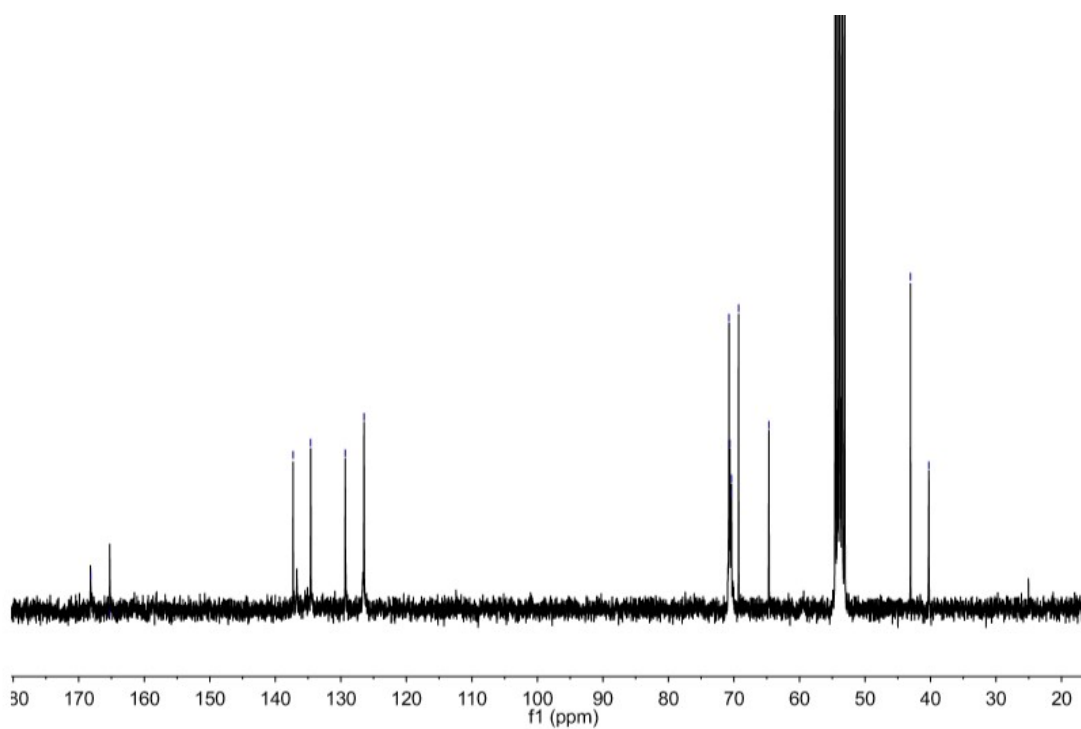


Fig. S88 $^{13}$C NMR of compound **7** (CD$_2$Cl$_2$, 298K, 75 MHz)

67
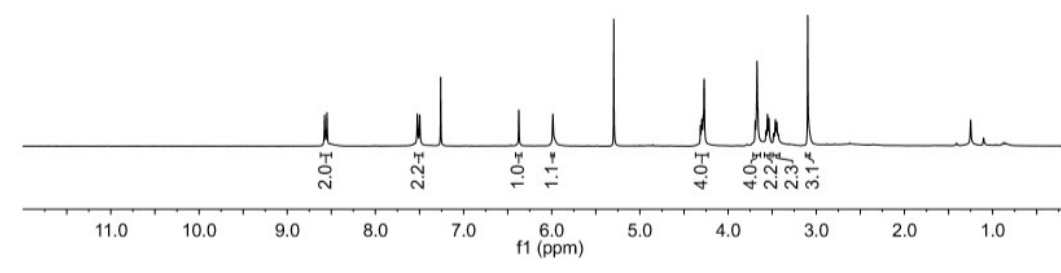
Fig. S89 $^1$H NMR of compound **8** (CDCl$_3$, 298K, 300 MHz)
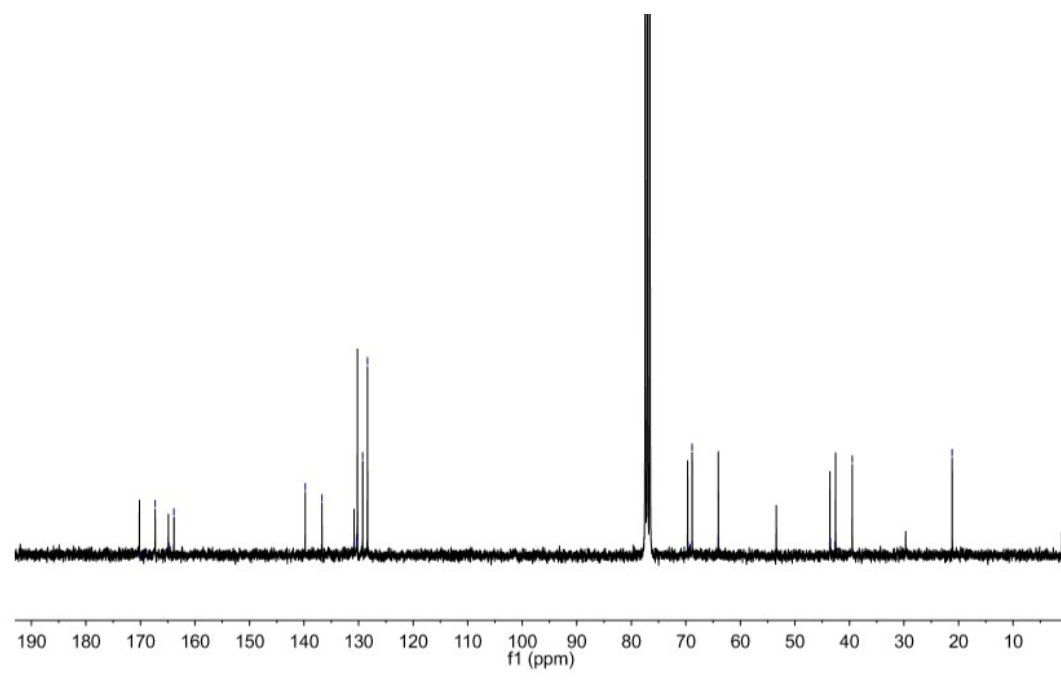


Fig. S90 $^{13}$C NMR of compound **8** (CDCl$_3$, 298K, 75 MHz)

68

Fig. S91 $^1$H NMR of compound **10** (CDCl$_3$, 298K, 300 MHz)
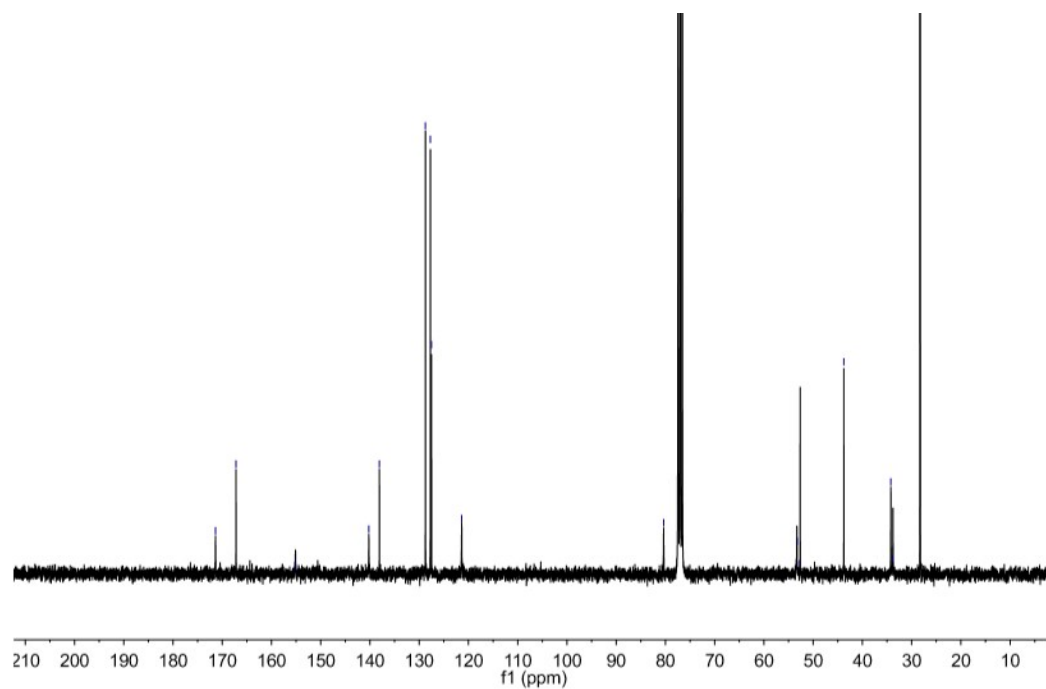


Fig. S92 $^{13}$C NMR of compound **10** (CDCl$_3$, 298K, 75 MHz)

69
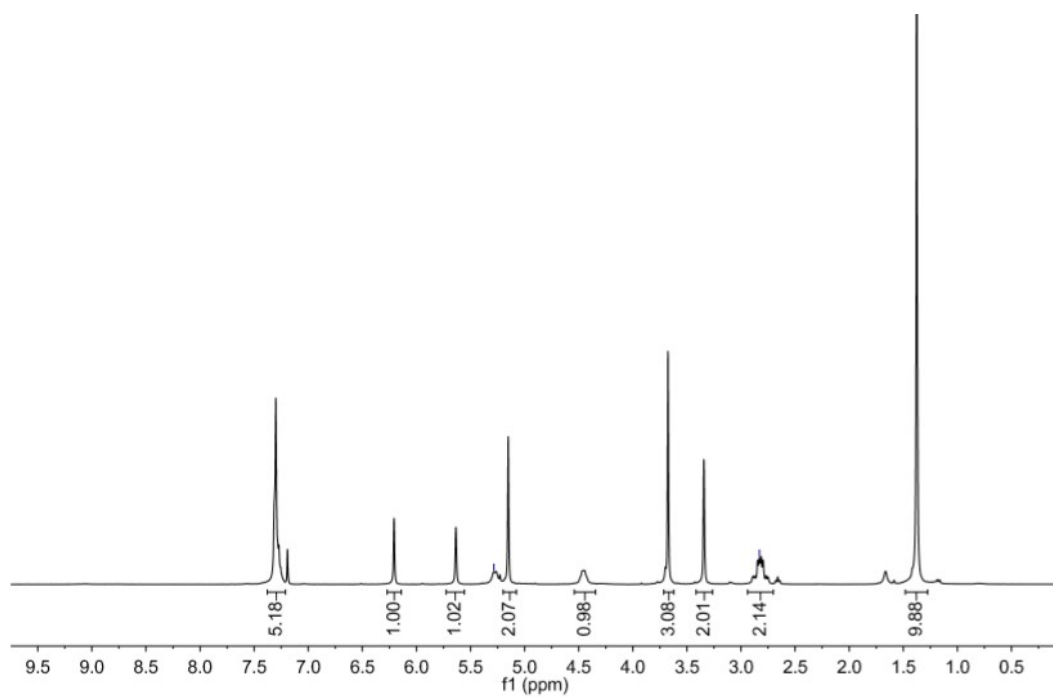
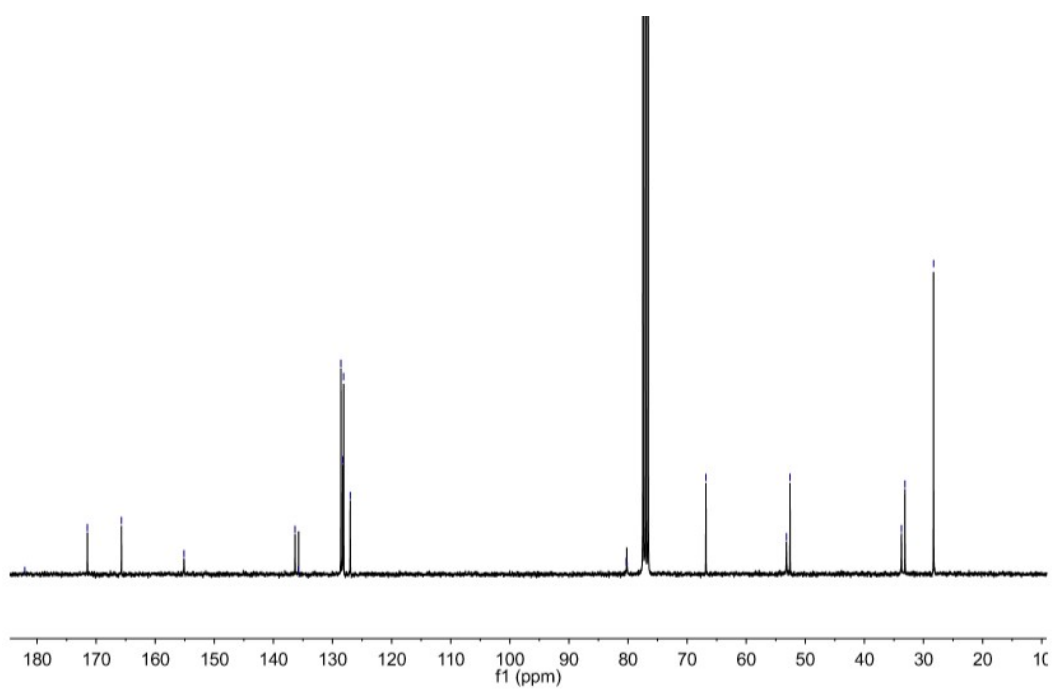Fig. S93 $^1$H NMR of compound **12** (CDCl$_3$, 298K, 300 MHz)



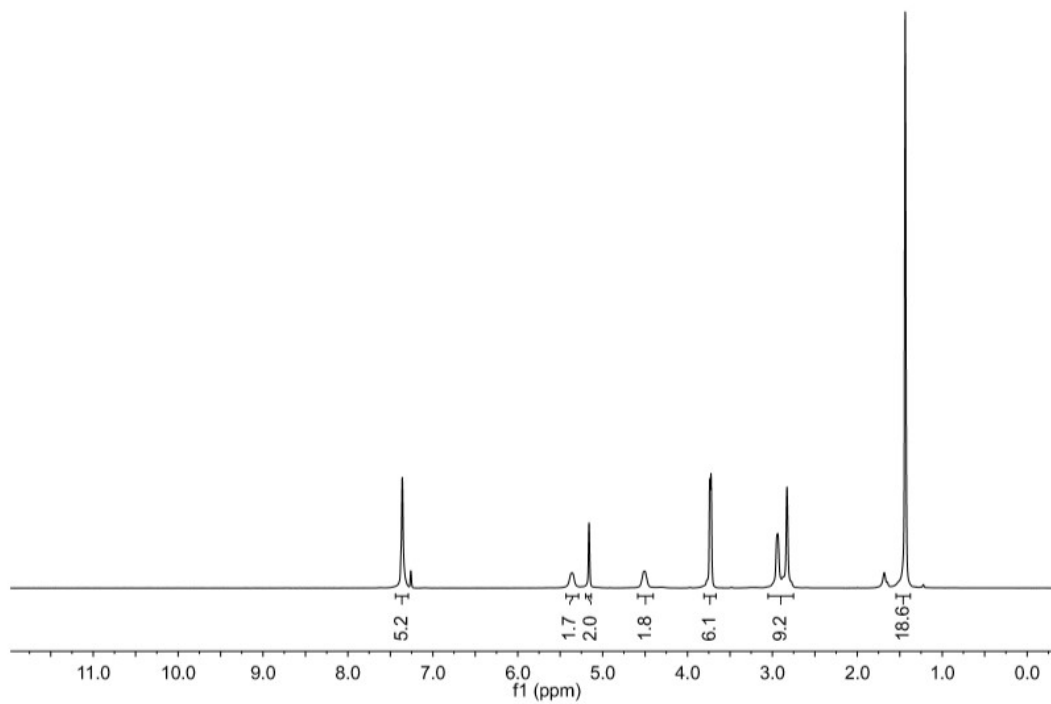Fig. S94 $^{13}$C NMR of compound **12** (CDCl$_3$, 298K, 75 MHz)

70

Fig. S95 $^1$H NMR of compound **13** (CDCl$_3$, 298K, 300 MHz)
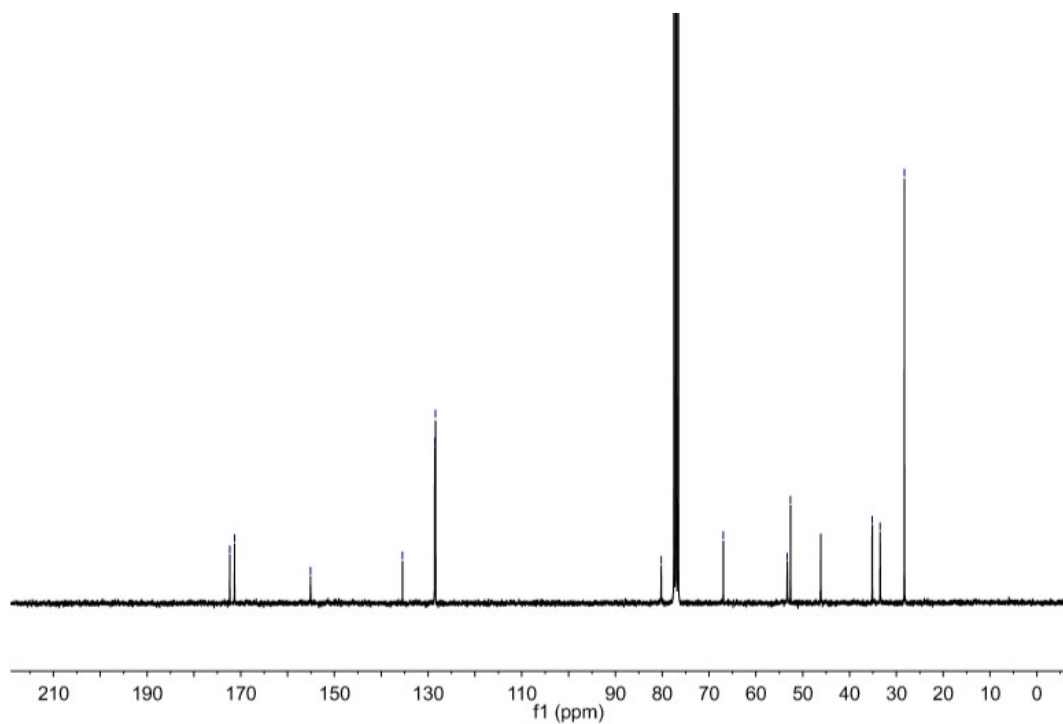


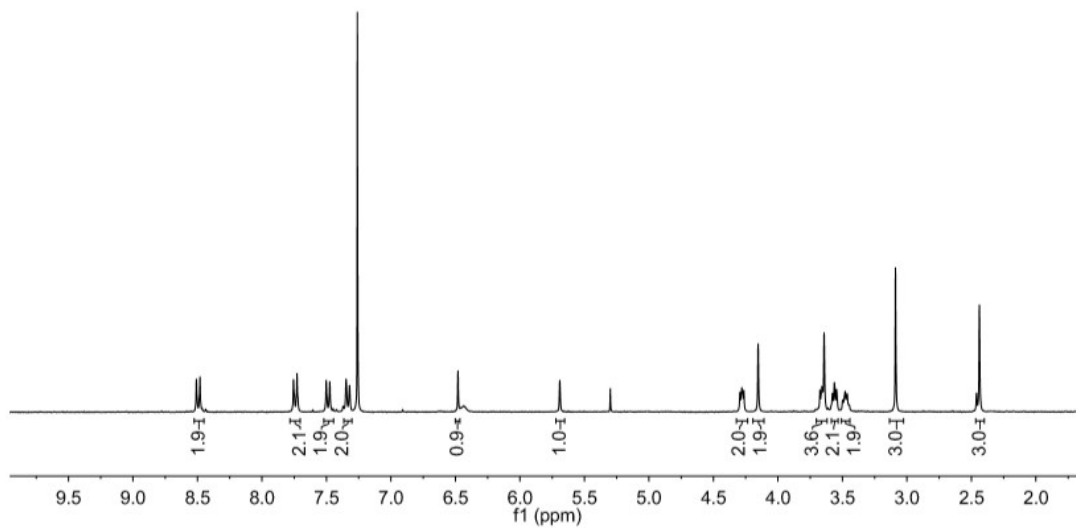Fig. S96 $^{13}$C NMR of compound **13** (CDCl$_3$, 298K, 75 MHz)

71

Fig. S97 ¹H NMR of compound **IC-Tz** (CDCl₃, 298K, 300 MHz)
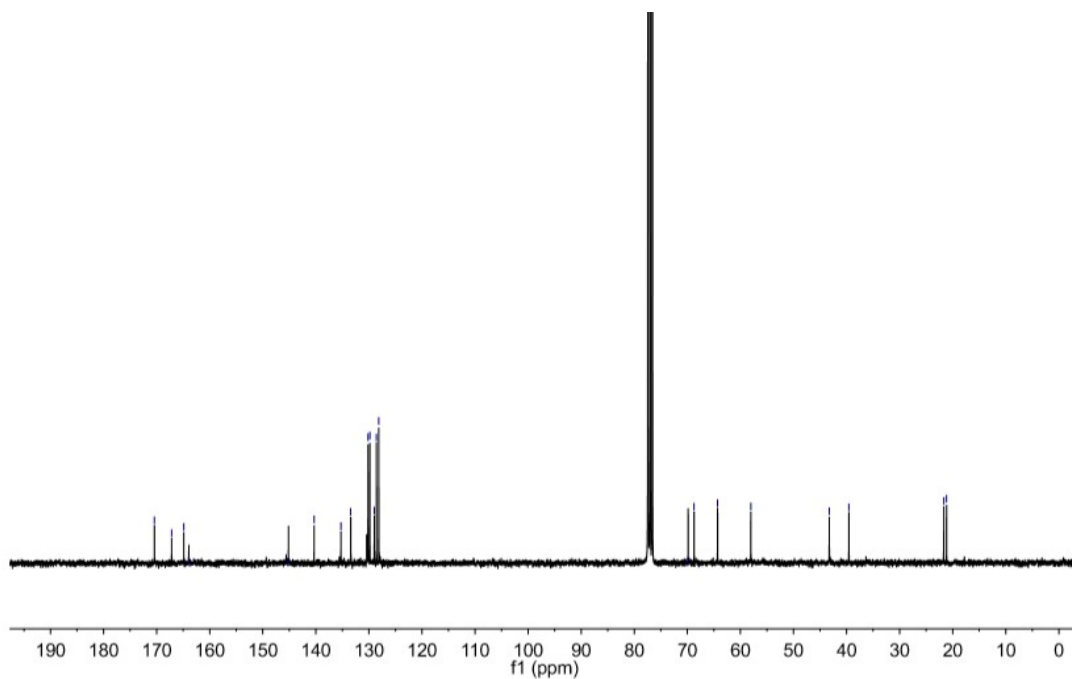


Fig. S98 ¹³C NMR of compound **IC-Tz** (CDCl₃, 298K, 75 MHz)

72

# References

1. Zegota, M., Wang, T., Seidler, C., Ng, D. Y. W., Kuan, S. L. and Weil, T., *Bioconjugate Chem.*, 2018, **29***,* 2665-2670.

2. Wang, T., Wu, Y., Kuan, S. L., Dumele, O., Lamla, M., Ng, D. Y., Arzt, M., Thomas, J., Mueller, J. O., Barner-Kowollik, C. and Weil, T., *Chem. Eur. J.*, 2015, **21***,* 228-238.

3. Ng, D. Y., Arzt, M., Wu, Y., Kuan, S. L., Lamla, M. and Weil, T., *Angew. Chem. Int. Ed.*, 2014, **53***,* 324-328.

4. Bernardim, B., Cal, P. M., Matos, M. J., Oliveira, B. L., Martinez-Saez, N., Albuquerque, I. S., Perkins, E., Corzana, F., Burtoloso, A. C., Jimenez-Oses, G. and Bernardes, G. J., *Nat. Commun.*, 2016, **7***,* 13128.

5. Wang, T., Riegger, A., Lamla, M., Wiese, S., Oeckl, P., Otto, M., Wu, Y., Fischer, S., Barth, H., Kuan, S. L. and Weil, T., *Chem. Sci.*, 2016, **7***,* 3234-3239.

6. Ravasco, J., Faustino, H., Trindade, A. and Gois, P. M. P., *Chem. Eur. J.*, 2019, **25***,* 43-59.

7. Hansen, R. E., Østergaard, H., Nørgaard, P. and Winther, J. R., *Anal. Biochem.*, 2007, **363***,* 77-82.

8. Fontaine, S. D., Reid, R., Robinson, L., Ashley, G. W. and Santi, D. V., *Bioconjugate Chem.*, 2015, **26***,* 145-152.

9. Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M., Muyldermans, S., Casas-Delucchi, C. S., Cardoso, M. C., Leonhardt, H., Hopfner, K. P. and Rothbauer, U., *Nat. Struct. Mol. Biol.*, 2010, **17***,* 133-139.

10. C. Hamers.Casterman, T. A., S. Muyldermans, G. Robinson, C. Hamers, E. Bajyana Songa, N. Bendahman and R. Hamerst, *Nature*. 1993, **363***,* 446-448.

11. Xu, L., Raabe, M., Zegota, M. M., Nogueira, J. C. F., Chudasama, V., Kuan, S. L. and Weil, T., *Org. Biomol. Chem.*, 2020, **18***,* 1140-1147.

12. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C. and MacCoss, M. J., *Bioinformatics*. 2010, **26***,* 966-968.

## 6. Declaration of originality

Ich versichere hiermit, daß ich die Arbeit selbständig angefertigt habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht habe.

Ulm, den

..........................

(Lujuan Xu)

I hereby declare that this thesis and the work reported herein was composed by and originated entirely from me. Information derived from the published and unpublished work of others has been acknowledged in the text and references are given in the list of sources.

Ulm, den

..........................

(Lujuan Xu)

# 7. Curriculum Vitae

The content on this page has been removed

according to personal data protection rules!

## 8. Acknowledgements

The content on this page has been removed

according to personal data protection rules!